

# A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta

Paolo Rossi,<sup>1</sup> Lei Shi,<sup>2</sup> Gaohua Liu,<sup>1</sup> Christopher M. Barbieri,<sup>1</sup> Hsiau-Wei Lee,<sup>3,4</sup> Thomas D. Grant,<sup>5,6</sup> Joseph R. Luft,<sup>5,6</sup> Rong Xiao,<sup>1</sup> Thomas B. Acton,<sup>1</sup> Edward H. Snell,<sup>5,6</sup> Gaetano T. Montelione,<sup>1,7</sup> David Baker,<sup>2,8</sup> Oliver F. Lange,<sup>2</sup> and Nikolaos G. Sgourakis<sup>2,9\*</sup>

<sup>1</sup> Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, and Northeast Structural Genomics Consortium, Rutgers University, Piscataway, New Jersey 08854

<sup>2</sup> Department of Biochemistry, University of Washington, Seattle, Washington 98195

<sup>3</sup> Department of Chemistry, Complex Carbohydrate Research Center, and Northeast Structural Genomics Consortium, University of Georgia, Athens, Georgia 30602

<sup>4</sup> Department of Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, and Northeast Structural Genomics Consortium, University of Georgia, Athens, Georgia 30602

<sup>5</sup> Hauptman–Woodward Medical Research Institute, and Northeast Structural Genomics Consortium, Buffalo, New York 14203

<sup>6</sup> SUNY Buffalo Department of Structural Biology, Buffalo, New York 14203

<sup>7</sup> Department of Biochemistry, Robert Wood Johnson Medical School, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854

<sup>8</sup> Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195

<sup>9</sup> Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

## ABSTRACT

Oligomeric proteins are important targets for structure determination in solution. While in most cases the fold of individual subunits can be determined experimentally, or predicted by homology-based methods, protein–protein interfaces are challenging to determine *de novo* using conventional NMR structure determination protocols. Here we focus on a member of the bet-V1 superfamily, Aha1 from *Colwellia psychrerythraea*. This family displays a broad range of crystallographic interfaces none of which can be reconciled with the NMR and SAXS data collected for Aha1. Unlike conventional methods relying on a dense network of experimental restraints, the sparse data are used to limit conformational search during optimization of a physically realistic energy function. This work highlights a new approach for studying minor conformational changes due to structural plasticity within a single dimeric interface in solution.

Proteins 2015; 83:309–317.  
© 2014 Wiley Periodicals, Inc.

**Key words:** protein complex; nuclear magnetic resonance spectroscopy; CS-Rosetta modeling; residual dipolar coupling; small-angle X-ray scattering.

## INTRODUCTION

Oligomerization plays an important role in the function and activity of proteins in signaling pathways and enzymes.<sup>1</sup> X-ray crystallography has been the workhorse in this endeavor but its application requires the availability of diffraction quality crystals. More importantly, proteins might adopt drastically different oligomeric structures in the crystal due to crystal lattice interactions that have no bearing on the biologically significant structure.<sup>2,3</sup> NMR is a powerful technique for probing the physiologically relevant solution state and, despite complications that depend on the dimer association affinity, the combined size of the molecule and conformational

Additional Supporting Information may be found in the online version of this article.

Abbreviations: CS, chemical shift; ILV, Isoleucine, Leucine, and Valine; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; RASREC, resolution-adapted structural recombination; RDC, residual dipolar coupling; SAXS, small angle X-ray scattering.

Grant sponsor: NIH; Grant numbers: GM76222 (to D.B.) and GM088396 (to E.H.S., T.D.G., and J.R.L.); Grant sponsor: National Institute of General Medical Sciences Protein Structure Initiative; Grant number: U54-GM094597 (to G.T.M.); Grant sponsor: DTRA; Grant number: HDTRA1-10-c-0057 (to E.H.S.); Grant sponsors: HHMI (to D.B.) and Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (to N.G.S.).

\*Correspondence to: Nikolaos G. Sgourakis, Laboratory of Chemical Physics N.I.D.D.K, N.I.H. 9000 Rockville Pike, Bldg. 5, Rm.B1-31 Bethesda, MD 20892. E-mail: sgourakis@mail.nih.gov

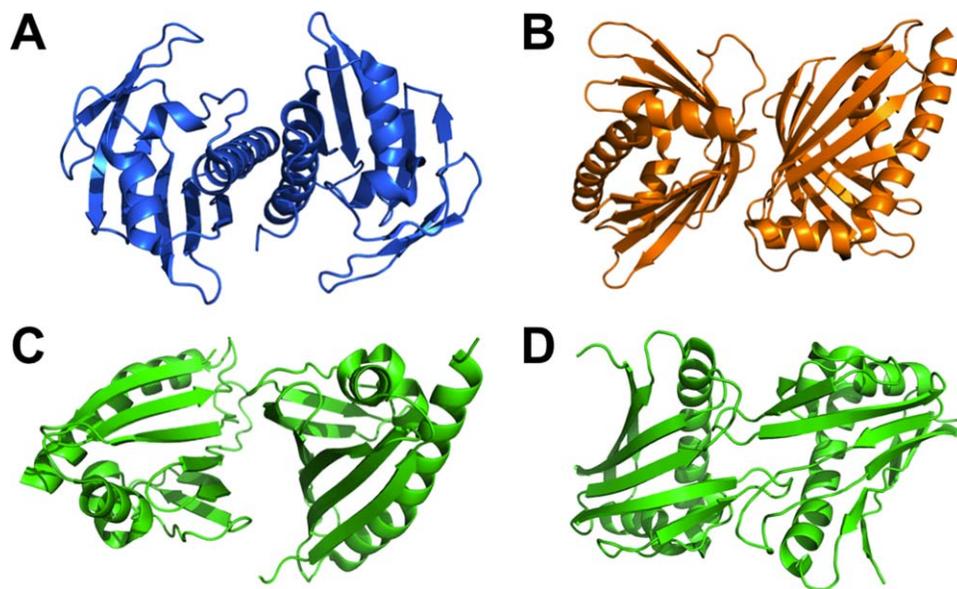
Received 23 July 2014; Revised 10 October 2014; Accepted 29 October 2014  
Published online 12 November 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24719

exchange processes,<sup>4</sup> NMR continues to provide key contributions in the understanding of protein–protein<sup>5</sup> and protein–nucleic acid<sup>6</sup> interactions. These effects tend to broaden NMR cross-peaks beyond detection and require specialized isotopic labeling schemes, including the labeling of distinct residues on individual chains to overcome symmetry ambiguities in the spectrum.<sup>7</sup> These approaches are generally more expensive and time-consuming, and yield less accurate structures due to the sparse nature of the data.<sup>8</sup> Methods to determine the orientation of individual subunits relative to the diffusion tensor using <sup>15</sup>N backbone relaxation rate ratios ( $R_2/R_1$ ),<sup>9</sup> in combination with sparse NOEs and shape restraints (SAXS), or, alternatively using SAXS and ion mobility measurements from mass-spectrometry,<sup>10</sup> have been recently introduced in the popular software packages Xplor-NIH<sup>11</sup> and HADDOCK<sup>12</sup>. However, the sophisticated diffusion tensor analysis is limited to well-behaved systems that show no aggregation or transient dimer formation, while the HADDOCK approach works well when supplied with pre-existing models of the unbound subunits and chemical shift perturbation data reporting on the mapping of the dimerization interface that are difficult to obtain for obligate complexes (nM range  $K_D$ ). The symmetry-ADR method,<sup>13</sup> involving the use of extensive sets of ambiguous NOE distance restraints (intra and inter-subunit), is an important advance toward this goal, but the calculation of symmetric oligomers by NMR remains challenging.<sup>14</sup> In particular, further development is needed to accurately determine the structures of larger

sized dimers or higher order oligomers in the absence of any starting models of the monomeric subunit.

We have recently developed a series of computational methods that model larger monomeric protein targets,<sup>15</sup> small-sized protein oligomers,<sup>16</sup> and complex macromolecular assemblies<sup>17</sup> within the CS (chemical shift)-Rosetta framework.<sup>18</sup> These methods effectively overcome low restraint count by advanced conformational sampling algorithms and refinement using a physically realistic all-atom energy function to yield accurate structures.<sup>19</sup> In particular, sparse constraints from deuterated samples can be used to guide structure determination with RASREC-Rosetta.<sup>20</sup> This advanced sampling algorithm has allowed structure determination of challenging targets from a limited set of data, relative to conventional protocols.<sup>15,21–23</sup> RosettaOligomers allows structure determination of symmetric protein assemblies using chemical shifts and RDC data.<sup>24</sup> This approach has been shown to yield highly accurate (<2 Å backbone RMSD relative to the X-ray structure) oligomeric structures consisting of subunits with both independently folded monomers and domain-swapped topologies in the absence of inter-subunit NOE restraints.<sup>16</sup> This is primarily due to an advanced Rosetta symmetric modeling framework<sup>25</sup> that allows efficient conformational sampling and energy scoring by using an explicitly symmetric representation of the system.

Here we integrate the powerful RASREC-Rosetta and RosettaOligomers methods, and illustrate a practical application of the combined approach in determining the structure of the 33 kDa dimer Aha1 domain from



**Figure 1**

Examples of variability in the bet-V1 clan dimer interfaces. Aggregation screening was conducted prior to structure determination by NESG and the proteins were found to be dimeric under the crystallization conditions: **A**) SSP2350 (PDB ID 3Q6A). **B**) MM0500 (PDB ID 1XUV). **C**, **D**) Two plausible crystallographic dimer interfaces observed for MLL2253 (PDB ID 3Q63).

*Colwellia psychrerythraea* member of the AHSA1 family (PF08327) and bet-V1 clan. Aha1 (NESG target CsR251) was selected for structural analysis as part of a broader program on structural coverage of the bet-V1 superfamily. Proteins of the bet-V1 superfamily span diverse molecular functions including small molecule and protein transport as in the case of *Bacillus subtilis* Yndb<sup>26</sup> and stimulation of Hsp90s ATPase activity in Human (Hch1) and yeast.<sup>27</sup> These proteins are known to dimerize in a variety of different orientations in the crystalline phase [Fig. 1(A–D)] that poses a challenge to conventional structure determination methods indicating a potential role of crystal lattice contacts in perturbing the functional dimerization mode [Fig. 1(C,D)]. The final Hybrid Rosetta/NMR/SAXS ensemble shows a unique binding interface that deviates from previous bet-V1 structures, suggesting a functional adaptation involving a higher-order variation on a basic structural theme.

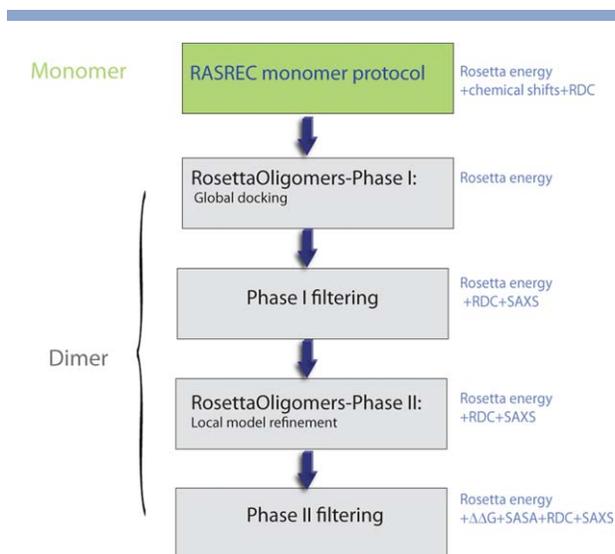
## MATERIALS AND METHODS

### NMR sample preparation and experimental assignment procedure

Aha1 samples for SAXS, NMR, and complementary biochemical techniques were obtained as described in Supporting Information Methods. Aha1 forms a dimer in solution ( $K_D$  of 80 nM) as shown by biochemical characterization (Supporting Information Figs. S1–S3). SAXS analysis further confirms that Aha1 is monodisperse and has a compact fold (Supporting Information Fig. S4 and Table S1). In order to gain insight into the dimeric structure by NMR we produced ILV({[U-<sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N]; Ile $\delta$ 1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>]}) as well as [U-<sup>13</sup>C, <sup>15</sup>N]-labeled Aha1 samples. Deuteration and selective Ile, Leu, and Val (ILV) methyl protonation were required to obtain backbone and methyl chemical shift (CS) assignments (see Supporting Information Methods and Figs. S5 and S6). Interestingly, while signal was barely observable for triple resonance experiments involving magnetization transfer through <sup>13</sup>C $\alpha$ , due to the relatively short (~35 ms) backbone <sup>15</sup>N T<sub>2</sub> relaxation time (Supporting Information Fig. S7), the [U-<sup>13</sup>C, <sup>15</sup>N] sample gave excellent three-dimensional (3D) <sup>15</sup>N and <sup>13</sup>C edited-NOESYs spectra. Extensive sidechain CS assignments that included complete methyl assignments for Thr, Ala, and Met, and aromatics were obtained manually using the CH<sub>3</sub>-CH<sub>3</sub> and H<sup>N</sup>-CH<sub>3</sub> NOEs from the [U-<sup>13</sup>C, <sup>15</sup>N] sample and a starting set of H<sup>N</sup>, ILV methyl, <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts.<sup>28</sup> In addition, we recorded <sup>1</sup>H-<sup>15</sup>N RDCs using samples aligned in PEG<sup>29</sup> and Pfl Phage<sup>30</sup> media. Finally, we augmented the NMR dataset with SAXS data,<sup>31</sup> collected as described in Supporting Information Methods.

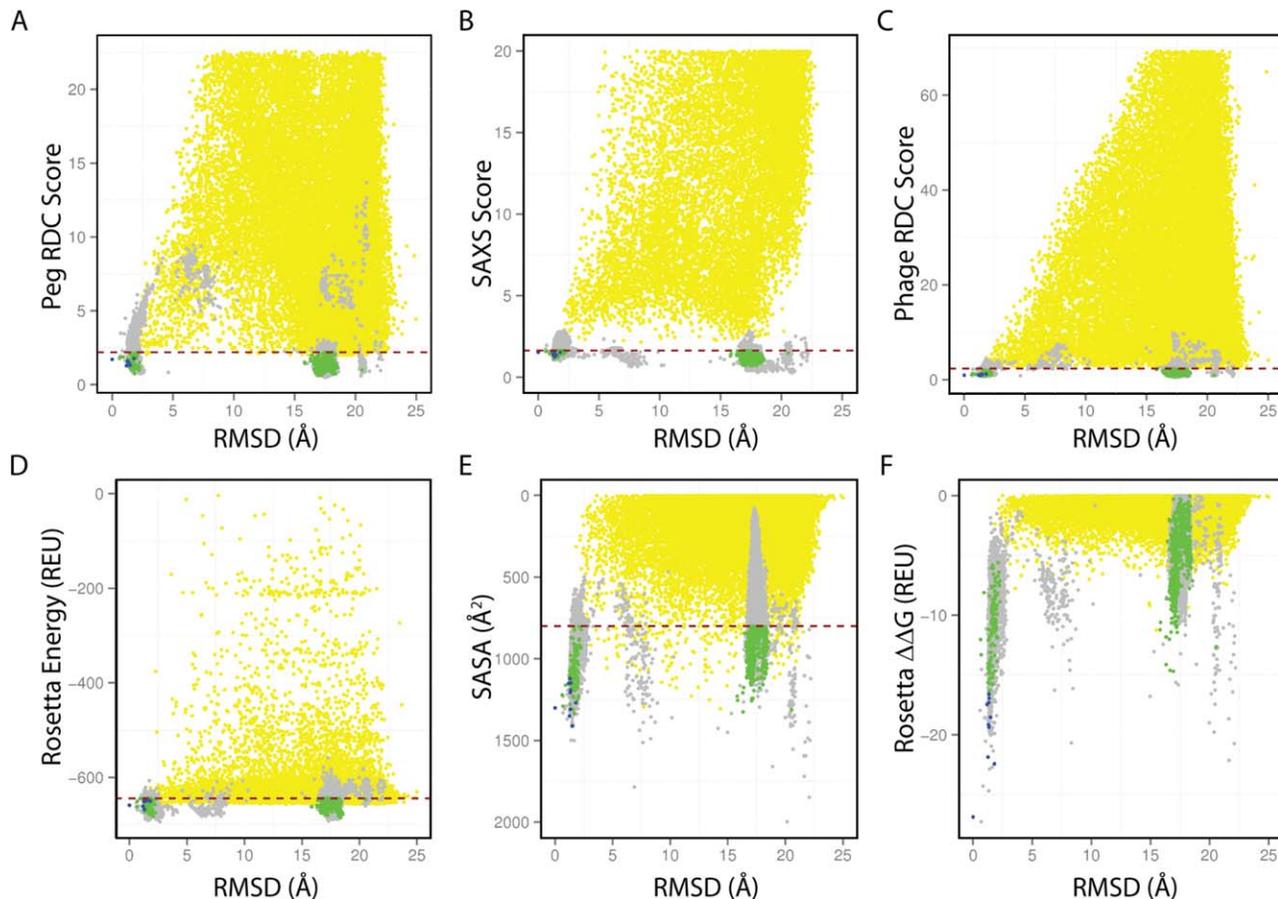
### Rosetta structure calculations

Dimer structure determination was carried out in a two-step manner (as highlighted in the workflow diagram of Fig. 2). First, we determined a structural ensemble of the monomeric subunit using a combination of backbone CS, H<sup>N</sup>-H<sup>N</sup>, H<sup>N</sup>-CH<sub>3</sub> and CH<sub>3</sub>-CH<sub>3</sub> NOEs and amide RDCs using RASREC-Rosetta (as described in Supporting Information Methods). Then, we performed symmetric docking calculations starting from the 10 lowest-energy monomer structures using the two sets of RDCs and SAXS data. The combination of RDCs with SAXS data was crucial, as each data type by itself was insufficient to obtain convergence within a single dimeric binding mode (Fig. 3). The employed hybrid approach aims to combine data reporting at different levels of structural precision to obtain a highly converged solution of the dimeric complex. Specifically, RDC and SAXS data that can discriminate between different overall dimer topologies are combined with the detailed Rosetta all-atom energy,<sup>19</sup> which is sensitive to the local sidechain packing and hydrogen bonding details of the binding interface. To account for these differences in the various sources of structural information employed, we developed a two-tier refinement strategy, as outlined in detail in Supporting Information Methods. First, we performed global docking calculations using the Rosetta energy function alone (Phase I). The motivation for this approach is to sample the docking energy landscape globally, without introducing additional experimental



**Figure 2**

Workflow of the Integrative Rosetta modeling approach. The different steps of RASREC monomer determination followed by Phase I and Phase II docking, filtering and final model selection are outlined as boxes. The type of experimental/Rosetta energy information and structure quality factors used at each step are indicated on the right of each box. Inter-chain NOE data were not employed in the structure determination process, but were instead used to validate the final dimer structure.



**Figure 3**

Experimental score, structure quality terms, and Rosetta energy distributions for all sampled dimer models. As a reference structure for backbone RMSD calculations ( $x$ -axis), we used the model with the lowest interface interaction score (1st model in the submitted PDB ensemble, residues 1–131). All sampled Phase-I (global docking) models are shown in yellow (40,000 points), while Phase II models (local perturbation) in gray, (14,000 models). Phase II conformations within the lowest 75 percentile (red dashed lines) of Rosetta all atom energy, RDC penalty, SAXS penalty and with solvent exposed surface area (SASA) greater than  $800 \text{ \AA}^2$  were kept for further analysis (green, 900 points). From these 900 conformations, the 10 with lowest predicted  $\Delta\Delta G$  values were selected as the final ensemble (blue). In detail: **A**) PEG RDC score computed as  $\text{RMS}(D_{\text{EXP}} - D_{\text{CALC}}) / D^{\text{A}}$ , where  $D^{\text{A}}$  is the alignment tensor magnitude. **B**) SAXS score, computed as  $\text{RMS}(I_{\text{EXP}} - I_{\text{CALC}})$ . **C**) Phage RDC score, computed in a similar manner. **D**) Rosetta score12 all-atom energy, in Rosetta Energy Units (REU). **E**) SASA ( $\text{\AA}^2$ ). **F**) Interface free energy (ddg or  $\Delta\Delta G$ ) defined as:  $\Delta G_{\text{dimer}} - 2 * \Delta G_{\text{monomer}}$

biases in order to consider the maximum number of dimer conformations (Fig. 3, yellow points). Instead, the experimental data were used in a filtering step to select Phase I dimer models sampled by Rosetta that also showed reasonable agreement with the SAXS and RDC data. The selected models were further refined using small-amplitude perturbations of the 4 rigid body degrees of freedom defining the dimer orientation, followed by backbone minimization and sidechain repacking (Phase II), using both sets of RDCs and SAXS as additional scoring terms to the Rosetta energy (Fig. 3, gray points). From this refined set of structures we considered the top 75%, ranked according to Rosetta energy, RDC and SAXS score, which buried at least  $800 \text{ \AA}^2$  of solvent exposed surface area (SASA) in the interface (Fig. 3 and Supporting Information Fig. S8, green points). We finally selected an ensemble of 10 conforma-

tions showing the lowest interface energies ( $\Delta\Delta G$ ) (Fig. 3 and Supporting Information Fig. S8, blue points). This highly converged ensemble (within  $0.7 \text{ \AA}$  heavy-atom RMSD relative to the average structure) was deposited in the PDB (PDB ID 2M89).

### Selection of experimental weights and filtering parameters

For Phase II sample, we use a hybrid energy function of the form:

$$E_{\text{Tot}} = E_{\text{Rosetta}} + w_{\text{EXP}} * (w_1 E_{\text{RDC1}} + w_2 E_{\text{RDC2}} + w_3 E_{\text{SAXS}})$$

where  $E_{\text{Rosetta}}$  is the default Rosetta3 score function (score12),  $E_{\text{RDC1}}$ ,  $E_{\text{RDC2}}$  measure the RMSD between experimental and calculated RDCs after non-linear fitting of the five alignment

tensor parameters and  $E_{\text{SAXS}}$  is proportional to the RMSD between the experimental and calculated SAXS profiles (as outlined in Supporting Information Methods). In this approach, the experimental scoring terms are used as a filter after Phases I, II and as calculation restraints in Phase II (Fig. 2) to eliminate dimer arrangements that are largely inconsistent with the data. Accordingly, the overall weights of the experimental score terms ( $w_{\text{EXP}}$ ) were optimized in a series of preliminary calculations using a grid search, and the weights that give the lowest distribution of Rosetta energies ( $E_{\text{Rosetta}}$ ) were selected for the final calculations. While the relative weights between different RDC datasets ( $w_1, w_2$ ) were scaled according to the inverse magnitude of the corresponding alignment tensors, the SAXS weight ( $w_3$ ) was adjusted such that a similar dynamic range of SAXS scores as the combined RDC scores is sampled in the calculations.

Although the method presented here can be readily applied to a wide range of oligomers, several parameters must be optimized on a case-by-case basis according to the size and complexity of each target. While the filters based on the fits to the experimental data and Rosetta energy are shown here as a percentile over the total number of sampled models (Fig. 3), the exact  $\Delta\Delta G$  cut-off is related to the type of interface (according to size, secondary structure, and type of interactions).<sup>32</sup> In the absence of large conformational changes upon dimer formation this value can be estimated based on the statistics of observed protein–protein interfaces in the PDB, as reviewed recently,<sup>33</sup> by monitoring the sampled SASA values in preliminary structure calculations. In the Phase II filtering step, we consider dimer conformations that bury at least 800 Å<sup>2</sup> of SASA, based on the distribution of SASA in crystallographic interfaces in the PDB.<sup>2</sup>

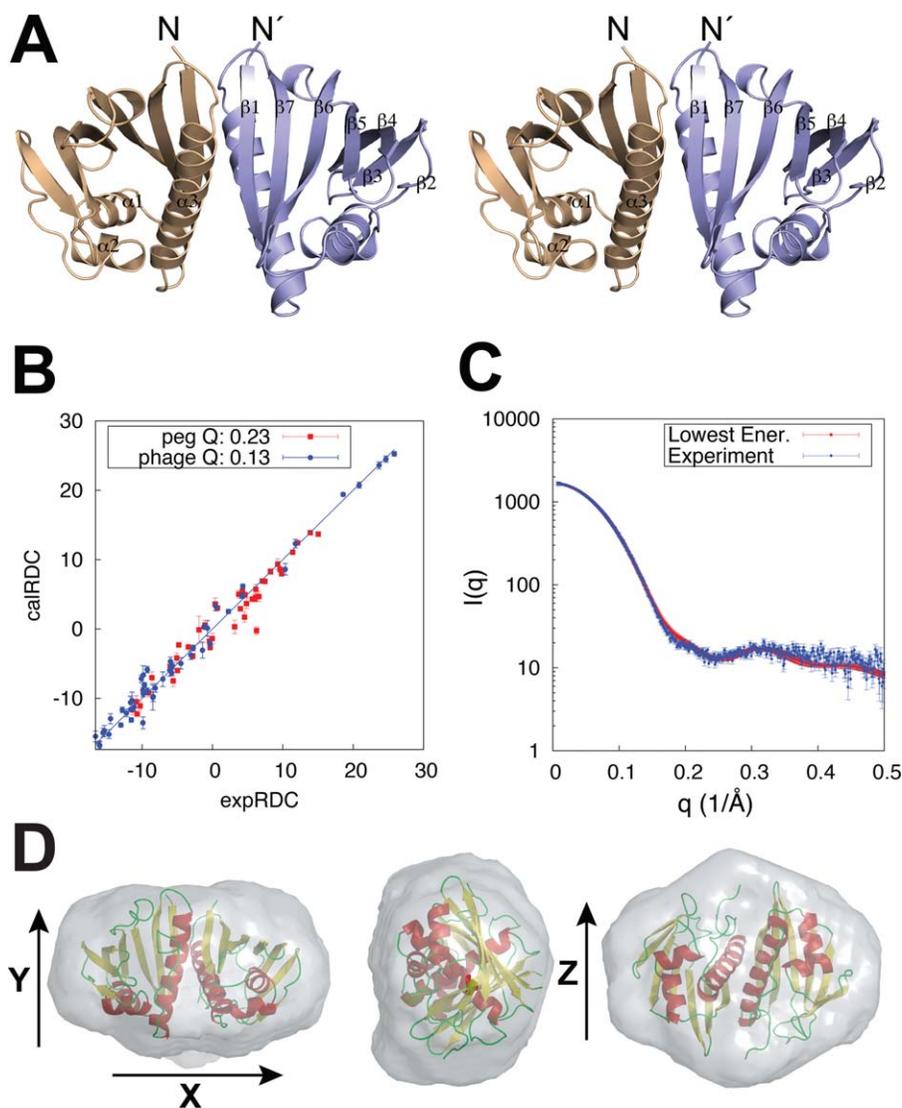
## RESULTS AND DISCUSSION

The protomer structure in the final ensemble (Fig. 4 and Supporting Information Fig. S9) is characteristic of the bet-V1 superfamily fold with seven antiparallel  $\beta$ -strands flanked by two small contiguous  $\alpha$ -helices ( $\alpha 1$ ,  $\alpha 2$ ) following  $\beta 1$  and a long carboxyl-terminal helix ( $\alpha 3$ ). The computed monomer structure falls within 2.0 Å C $\alpha$  RMSD to the closest homolog structure MLL2253 (PDB ID 3Q63) at 29% sequence identity. The dimer interface is formed by a unique edge strand-to-helix interaction involving  $\beta 1$  and  $\alpha 3$  and buries a large solvent-accessible surface area (SASA), of approximately 1300 Å<sup>2</sup> [Fig. 3(E)]. Unlike the flavonoid-binding members of the bet-V1 clan that can accommodate a ligand at the core,<sup>26</sup> the Aha1 core is tightly packed suggesting a structural adaptation to a distinct molecular function. The structure is in good agreement with the experimental RDCs, fits well the SAXS-derived *ab initio* molecular envelope, and is independently validated by inter-chain NOEs that were not

used in determining the structure (Fig. 4; Fig. 5 and discussion below).

The combination of RDCs and SAXS with the Rosetta Energy function and structure quality metrics (SASA,  $\Delta\Delta G$ ) allows identifying the correct dimeric binding mode with high accuracy, according to our previous benchmarks.<sup>16,34</sup> The RDCs provide domain orientational information, yet are insensitive to rigid-body translations of one monomer relative to the other, where SAXS data provide an overall envelope of the dimeric assembly that is sensitive to both rotations and translations, albeit with much lower precision. Therefore, both RDCs and SAXS are highly sensitive to different sources of structural “noise” and the calculations do not converge on a single dimer structure when repeating the entire Phase I/Phase II protocol by using either the RDCs or SAXS data alone. To illustrate the discriminating power of different experimental and structure quality scoring terms in deriving the final ensemble of models [Supporting Information Fig. S8(G), “cluster A”], we consider an alternative cluster of conformations sampled in Phase II [Supporting Information Fig. S8(H), “cluster B”], that differs significantly in the domain orientation of the two monomers [Fig. 5(A)]. While there is certainly a noticeable difference in the SAXS profiles between the two ensembles (cluster A/panel I vs. cluster B/panel J) for  $Q$  values  $>0.35\text{\AA}^{-1}$  (a part of the profile that also has increased experimental errors and was therefore not included in structure refinement), the  $\Delta\Delta G$  values show a much sharper trend that clearly demarcates cluster A as the correct structure over the alternative cluster B sampled in Phase II [Supporting Information Fig. S8(F)]. However, it is the inclusion of the experimental data as a filter after Phase I that helps enhance sampling of conformations in the vicinity of the native dimer structure by eliminating additional “decoy” conformations. For example, if the calculated  $\Delta\Delta G$  values were used for Phase I filtering instead [Fig. 3(F)—yellow points], we would obtain a much larger number of false positives for Phase II refinement (i.e., dimers that still show realistic interface features using an incorrect binding surface), thus limiting the efficiency and convergence of the method on a unique binding mode.

In contrast to conventional *de novo* structure determination methods that rely heavily on an extensive network of experimental restraints to achieve structural convergence, our method utilizes a physically realistic energy function<sup>19</sup> that is enhanced by the intersection of all sources of experimental information as a means to limit the search of conformational space. As a result of the chosen approach, the final structures display a well-packed interface with good structural statistics (as assessed by MOLPROBITY<sup>35</sup>) while using a minimal set of experimental restraints to guide the search (Supporting Information Table S2). The weakly restraining strategy adopted here prevents over-fitting without the need



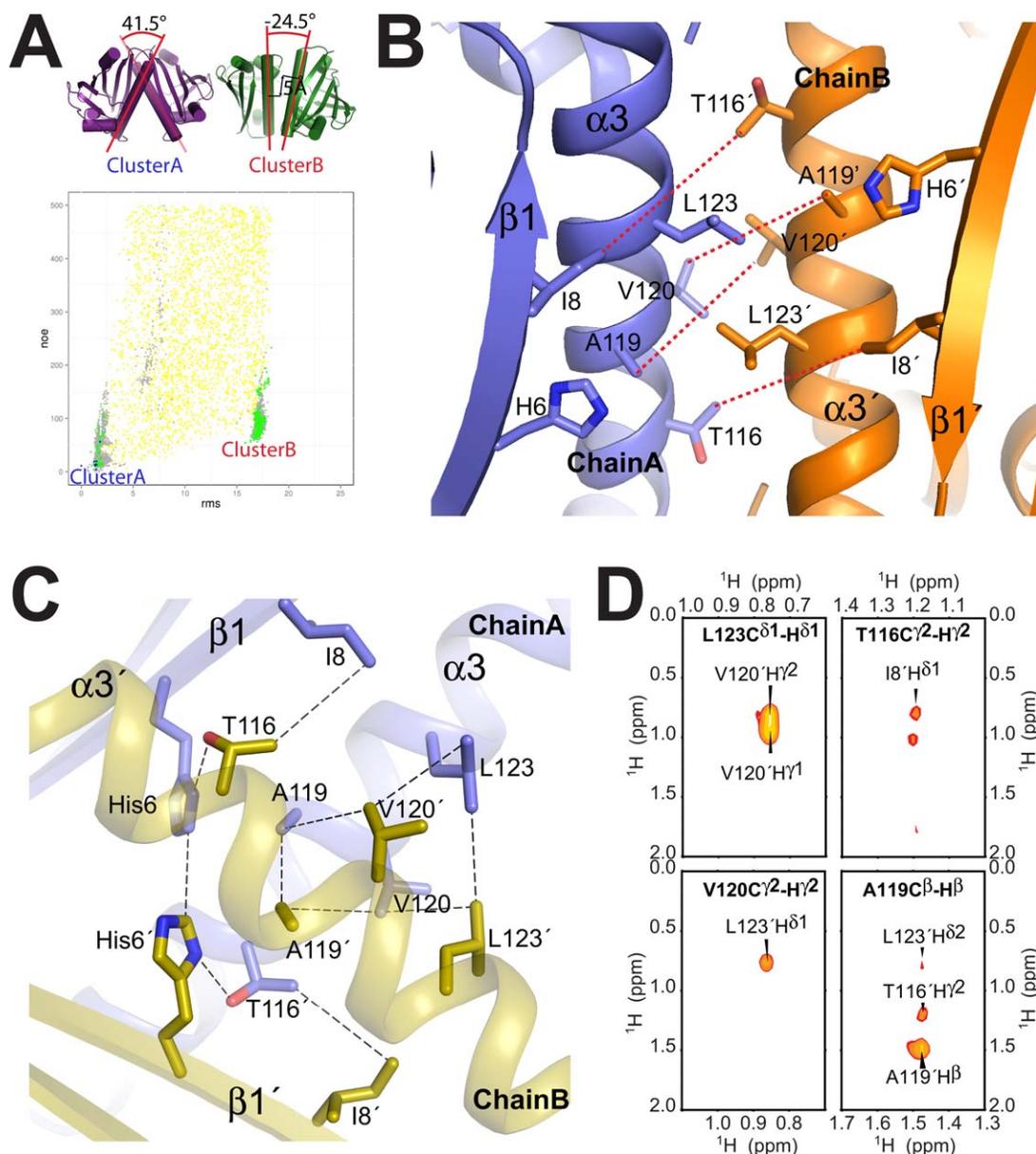
**Figure 4**

Final model fit to the RDC and SAXS data. **A)** Stereo view of the Aha1 symmetric dimer structure with secondary structure elements indicated for a single chain. **B)** Calculated versus experimental  $^1\text{H}$ - $^{15}\text{N}$  RDCs showing the agreement of the minimally restrained CS-Rosetta lowest energy structure to the experimental data. **C)** SAXS data and corresponding fit obtained for the lowest energy structure ( $\chi^2 = 1.71$ ). **D)** *Ab initio* molecular envelope from a consensus model of 20 individual reconstructions shown along  $x$ ,  $y$ , and  $z$  vectors. Spatial discrepancy of 0.540 and variation of 0.023 were obtained in the final fit. The final structural ensemble was selected using energy-based criteria (Fig. 3 and Supporting Information Fig. S8) and further validated using an independent dataset of inter-molecular NOEs [Fig. 5(B)].

for data-intensive cross-validation methods to adjust experimental weights,<sup>36</sup> since the total energy of the final models is mostly (>95%) determined by the Rosetta energy function. As a result of these features, the new approach provides improved convergence compared to models defined by the available experimental data alone using conventional structure determination protocols.

In order to experimentally validate the identified dimer interface, we recorded additional 3D  $^{13}\text{C}$ ,  $^{15}\text{N}$ -filtered- $^{13}\text{C}$ ,  $^{15}\text{N}$ -edited NOESY spectra<sup>37</sup> on a mixed [ $\text{U-}^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled/fully protonated sample as an independent dataset. Not surprisingly, the methyl region of the spec-

trum yielded important inter-subunit restraints (Supporting Information Table S2). Specifically, 23 cross-peaks between Ala119, V120, Leu123, Thr116, and Ile8 were well resolved and unambiguously assigned (Supporting Information Table S3). As Ala119, V120, Leu123, Thr116, and Ile8 are located in the  $\alpha 3$  region of the dimer interface region, these 23 cross-peaks in the double-half filtered NOE data unequivocally confirm the placement of the key inter-chain methyl contacts observed in the Rosetta models [Fig. 5(C,D)]. Notably, the lowest-energy models have significantly lower NOE scores than all other dimer conformations sampled in the Rosetta docking calculations



**Figure 5**

Validation using an independent set of interchain NOEs. **A**) Agreement of double half-filtered intermolecular NOEs (not used as structure restraints in Phase I/II calculations or filtering steps shown in Figure 2) vs. RMSD to the lowest-energy structure. For clarity, only models with below 500 NOE penalty score (using a flat-bottom scoring term with a 5.5Å upper limit and an exponential penalty function) are shown. Color scheme is the same as in Figure 3: Phase I/global docking (yellow), Phase II/local perturbation (grey), below post-Phase II filter (green) and lowest 10  $\Delta\Delta G$  scores (final structures, blue). Two most prominent clusters of filtered models (“Cluster A”, “Cluster B”) are indicated, and their differences in terms of the dimer orientation are highlighted on the lowest-energy representative structural models on the top left. **B**) Lowest-energy representative structure from cluster B shows a number of intermolecular NOE violations, between Ile8, Thr116, Ala119 and V120, as indicated by red dashed lines. **C**) Close up view of the dimer interface in the final model from Cluster A (1st member of the final ensemble). Dashed lines indicate the intermolecular contacts identified in the NMR spectra (3D double half-filtered  $^{13}\text{C}$ -edited NOESY in black, 3D- $^{13}\text{C}$ -edited NOESY in green). **D**) Assigned X-filtered NOESY cross-peaks used to validate the Aha1 structure. In bold is the assignment of the  $^{13}\text{C}$  labeled direct observed  $^1\text{H}$  atom and label on the peak is the interchain  $^{12}\text{C}$  labeled  $^1\text{H}$ . Further interchain cross-peaks were assigned in the 3D NOESY spectra that connect T116 H $\gamma$ 2 to His6 H $\alpha$  unambiguously (Supporting Information Table S3).

[Fig. 5(A)], including the alternative binding mode (cluster B) sampled in Phase II, that shows multiple NOE violations [Fig. 5(B)]. Subsequently, as a further validation test, the cross-filtered NOE data alone were used as structural

restraints in new docking calculations (as outlined in Supporting Information Methods). The NOE-driven calculations converged to the same binding mode as the original calculations using RDCs and SAXS. Conversely, the

intermolecular NOE contacts are inconsistent with all interfaces observed in previous crystal structures of known homologs (Fig. 1). Taken together with the SAXS, RDC fits and energetic criteria defined above, these results unequivocally confirm the validity of the determined Rosetta models.

In summary, we present a hybrid approach for determining high-quality structures of protein dimers that uses NMR chemical shifts, sparse distance restraints (NOEs) obtained in uniformly and selectively methyl-labeled samples, and global orientational and shape restraints (RDCs and SAXS data). We demonstrate the discriminating power of the new approach by determining the dimeric structure of a representative of the variable bet-V1 superfamily, Aha1, and validate our models using an independent NOE restraint dataset. While the presence of a high-affinity ( $K_D$  80 nm) dimer allows the measurement of interchain NOEs in this study, in other cases the presence of an obligate oligomeric species may prevent the exchange of labeled / unlabeled chains towards obtaining mixed samples for such measurements. The new approach can provide a high-accuracy structure using RDCs and SAXS alone, thus allowing these NOEs, when available, to be used for independent validation. Notably, the Aha1 interface identified here is clearly distinct from the structures of other bet-V1 superfamily members. The widely varying dimerization modes observed within a single protein family could arise from perturbations of the domain arrangement in the crystal lattice or may reflect species-specific adaptation to the functional role contended by different family members. Further structural studies using the new solution-based approach focusing on the other Aha1 homologs are needed to address this important question.

## REFERENCES

- Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 2000;29:105–153.
- Janin J. Specific versus non-specific contacts in protein crystals. *Nat Struct Biol* 1997;4:973–974.
- Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein-protein interfaces. *J Mol Biol* 2004;336:943–955.
- Xu Y, Matthews S. TROSY NMR spectroscopy of large soluble proteins. *Top Curr Chem* 2011;335:97–119.
- Ihms EC, Zhou MW, Zhang Y, Kleckner IR, McElroy CA, Wysocki VH, Gollnick P, Foster MP. Gene regulation by substoichiometric heterocomplex formation of undecameric TRAP and trimeric anti-TRAP. *Proc Natl Acad Sci USA* 2014;111:3442–3447.
- Varani L, Gunderson SI, Mattaj JW, Kay LE, Neuhäus D, Varani G. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat Struct Biol* 2000;7:329–335.
- Traaseth NJ, Verardi R, Veglia G. Asymmetric methyl group labeling as a probe of membrane protein homo-oligomers by NMR spectroscopy. *J Am Chem Soc* 2008;130:2400–2401.
- Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, Kay LE. Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J Mol Biol* 2000;300:197–212.
- Ryabov Y, Clore GM, Schwieters CD. Direct use of N-15 relaxation rates as experimental restraints on molecular shape and orientation for docking of protein-protein complexes. *J Am Chem Soc* 2010;132:5987–5989.
- Karaca E, Bonvin AMJJ. On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr D* 2013;69:683–694.
- Schwieters CD, Kuszewski JJ, Clore GM. Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Mag Res Sp* 2006;48:47–62.
- Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731–1737.
- Nilges M. A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins* 1993;17:297–309.
- O'Donoghue SI, Nilges M. Calculation of symmetric oligomer structures from NMR data. In: Krishna NR, Berliner L, editors. *Biological magnetic resonance*. US: Biological Magnetic Resonance: Springer; 2002. Vol. 17, pp 131–161.
- Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci USA* 2012;109:10873–10878.
- Sgourakis NG, Lange OF, DiMaio F, Andre I, Fitzkee NC, Rossi P, Montelione GT, Bax A, Baker D. Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. *J Am Chem Soc* 2011;133:6288–6298.
- Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle. *Nature* 2012;486:276–279.
- Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 2008;105:4685–4690.
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YEA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Method Enzymol* 2011;545–574.
- Lange OF, Baker D. Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation. *Proteins* 2012;80:884–895.
- Rao T, Lubin JW, Armstrong GS, Tucey TM, Lundblad V, Wuttke DS. Structure of Est3 reveals a bimodal surface with differential roles in telomere replication. *Proc Natl Acad Sci USA* 2014;111:214–218.
- Sgourakis NG, Natarajan K, Ying J, Vogeli B, Boyd LE, Margulies DH, Bax A. The structure of mouse cytomegalovirus m04 protein obtained from sparse NMR data reveals a conserved fold of the m02-m06 viral immune modulator family. *Structure* 2014;22:1263–1273.
- Lloyd NR, Wuttke DS. Less is more: structures of difficult targets with minimal constraints. *Structure* 2014;22:1223–1224.
- Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science* 2005;309:1868–1871.
- DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 2011;6.
- Stark JL, Mercier KA, Mueller GA, Acton TB, Xiao R, Montelione GT, Powers R. Solution structure and function of YndB, an AHS1 protein from *Bacillus subtilis*. *Proteins* 2010;78:3328–3340.

27. Panaretou B, Siligardi G, Meyer P, Maloney A, Sullivan JK, Singh S, Millson SH, Clarke PA, Naaby-Hansen S, Stein R, Cramer R, Mollapour M, Workman P, Piper PW, Pearl LH, Prodromou C. Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. *Mol Cell* 2002;10:1307–1318.
28. Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci USA* 2005;102:10487–10492.
29. Ruckert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 2000;122:7793–7797.
30. Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of Pfl bacteriophage. *J Biomol NMR* 2001;20:365–377.
31. Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, Snell EH. Small angle X-ray scattering as a complementary tool for high-throughput structural studies. *Biopolymers* 2011;95:517–530.
32. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999;285:2177–2198.
33. Kastiris PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure-based benchmark for protein-protein binding affinity. *Protein Sci* 2011;20:482–491.
34. Parsons LM, Grishaev A, Bax A. The periplasmic domain of To1R from *Haemophilus influenzae* forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data. *Biochemistry* 2008;47:3131–3142.
35. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* 2010;66:12–21.
36. Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM. Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat Struct Biol* 1999;6:166–173.
37. Zwahlen C, Legault P, Vincent SJ, Greenblatt J, Konrat R, Kay LE. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage  $\lambda$  N-peptide/boxB RNA complex. *J Am Chem Soc* 1997;119:6711–6721.