

Docking of Protein–Protein Complexes on the Basis of Highly Ambiguous Intermolecular Distance Restraints Derived from $^1\text{H}_\text{N}/^{15}\text{N}$ Chemical Shift Mapping and Backbone ^{15}N – ^1H Residual Dipolar Couplings Using Conjoined Rigid Body/Torsion Angle Dynamics

G. Marius Clore*[†] and Charles D. Schwieters[‡]

Contribution from the Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0510, and Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624

Received October 10, 2002; E-mail: mariusc@intra.niddk.nih.gov

Abstract: A simple and reliable method for docking protein–protein complexes using $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping and backbone ^{15}N – ^1H residual dipolar couplings is presented and illustrated with three complexes (EIN-HPr, IIA^{Glc}-HPr, and IIA^{Mil}-HPr) of known structure. The $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping data are transformed into a set of highly ambiguous, intermolecular distance restraints (comprising between 400 and 3000 individual distances) with translational and some degree of orientational information content, while the dipolar couplings provide information on relative protein–protein orientation. The optimization protocol employs conjoined rigid body/torsion angle dynamics in simulated annealing calculations. The target function also comprises three nonbonded interactions terms: a van der Waals repulsion term to prevent atomic overlap, a radius of gyration term (E_{gyr}) to avoid expansion at the protein–protein interface, and a torsion angle database potential of mean force to bias interfacial side chain conformations toward physically allowed rotamers. For the EIN-HPr and IIA^{Glc}-HPr complexes, all structures satisfying the experimental restraints (i.e., both the ambiguous intermolecular distance restraints and the dipolar couplings) converge to a single cluster with mean backbone coordinate accuracies of 0.7–1.5 Å. For the IIA^{Mil}-HPr complex, twofold degeneracy remains, and the structures cluster into two distinct solutions differing by a 180° rotation about the z axis of the alignment tensor. The correct and incorrect solutions which have mean backbone coordinate accuracies of ~0.5 and ~10.5 Å, respectively, can readily be distinguished using a variety of criteria: (a) examination of the overall $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift perturbation map (because the incorrect cluster predicts the presence of residues at the interface that experience only minimal chemical shift perturbations; this information is readily incorporated into the calculations in the form of ambiguous intermolecular repulsion restraints); (b) back-calculation of dipolar couplings on the basis of molecular shape; or (c) the E_{gyr} distribution which, because of its global nature, directly reflects the interfacial packing quality. This methodology should be particularly useful for high throughput, NMR-based, structural proteomics.

Introduction

Protein–protein complexes represent the central theme of regulatory pathways, and knowledge of their structure is critical for an understanding of function. Despite recent advances,¹ ab initio docking from structures of free proteins is still in its infancy and remains problematic.² Experimental determination of the atomic structures of protein–protein complexes, either

by crystallography or NMR, is therefore still the method of choice. Solving such structures using conventional NMR methodology presents a considerable technical challenge and is highly time-consuming.³ If the structures of the free proteins are already known at high resolution, and conformational changes upon complexation are either minimal or localized, it is possible to use conjoined rigid body/torsion angle dynamics^{4,5} to solve the structure of the complex based solely on intermolecular interproton distance restraints, derived from isotope-

[†] National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

[‡] Center for Information Technology, National Institutes of Health.

(1) (a) Fahmy, A.; Wagner, G. *J. Am. Chem. Soc.* **2002**, *124*, 1241–1250. (b) Muñoz-Carpio, C. A. D.; Ichiishi, E.; Yoshimori, A.; Yoshikawa, T. *Proteins* **2002**, *48*, 696–732. (c) Chen, R.; Weng, Z. *Proteins* **2002**, *47*, 281–294. (d) Fernandez-Recio, J.; Totrov, M.; Abagyan, R. *Protein Sci.* **2002**, *11*, 280–291. (e) Lorber, D. M.; Udo, M. K.; Shoichet, B. K. *Protein Sci.* **2002**, *11*, 1393–1408. (f) Tovchirechko, A.; Wells, C. A.; Vakser, I. A. *Protein Sci.* **2002**, *11*, 1888–1896.

(2) Smith, G. R.; Sternberg, M. J. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28–35.

(3) (a) Clore, G. M.; Gronenborn, A. M. *Trends Biotechnol.* **1998**, *16*, 34. (b) Walters, K. J.; Ferentz, A. E.; Hare, B. J.; Hidalgo, P.; Jasanoff, A.; Matsuo, H.; Wagner, G. *Methods Enzymol.* **2001**, *339*, 238–258. (c) Zuiderweg, E. R. P. *Biochemistry* **2002**, *41*, 1–7.

(4) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *152*, 288–302.

(5) Clore, G. M.; Bewley, C. A. *J. Magn. Reson.* **2002**, *154*, 329–335.

filtered/edited nuclear Overhauser enhancement (NOE) measurements, and orientational restraints from residual dipolar couplings.^{4,6–8} Nevertheless, unambiguous assignment of intermolecular NOEs is still difficult and time-consuming, particularly for larger complexes.³ In contrast, mapping of interaction surfaces by $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift perturbation is a simple and rapid procedure.³ Likewise, measurement of backbone ^{15}N – ^1H residual dipolar couplings ($^1\text{D}_\text{NH}$) is entirely straightforward and fast.⁹ Here, we demonstrate with three examples of protein–protein complexes previously solved in our laboratory by NMR^{7,8,10} that it is possible to reliably dock such complexes based on highly ambiguous intermolecular distance restraints derived from $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping, in conjunction with backbone $^1\text{D}_\text{NH}$ dipolar couplings, using conjoined rigid body/torsion angle dynamics.

Methods

$^1\text{D}_\text{NH}$ dipolar coupling data were taken from the papers describing the structure determinations of the EIN-HPr,¹⁰ IIA^{Glc}-HPr,⁷ and IIA^{Mtl}-HPr⁸ complexes. $^1\text{D}_\text{NH}$ dipolar couplings for EIN-HPr (126 for EIN, 75 for HPr), IIA^{Glc}-HPr (118 for IIA^{Glc}, 75 for HPr), and IIA^{Mtl}-HPr (114 for IIA^{Mtl}, 71 for HPr) were measured in liquid crystalline media of phage fd^{11a} (negatively charged rod), tobacco mosaic virus^{11a} (negatively charged rod), and poly(ethylene glycol)/hexanol^{11b} (neutral), respectively. $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping data for EIN-HPr were taken from ref 12, while those for the IIA^{Glc}-HPr and IIA^{Mtl}-HPr complexes were derived from unpublished data obtained at the time we solved their structures.^{7,8}

All minimization and dynamics calculations were carried out using the program Xplor-NIH.¹³ Rigid body/torsion angle dynamics was carried out using a sixth-order predictor-corrector integrator with automatic time step selection.⁴ Residue accessible surface area, ASA (expressed as a percentage of that residue's surface accessibility in an extended Gly-X-Gly tripeptide), was calculated using the program GETAREA.¹⁴ Molecular structure figures were made with the programs GRASP¹⁵ and VMD-XPLOR.¹⁶ The ensemble distributions of the docked structures are depicted by isosurfaces of the reweighted atomic density maps¹⁷ drawn at a value of 20% of the maximum with a uniform atomic radius of 1 Å.

The calculations made use of the X-ray coordinates for free HPr (PDB code IPOH),^{18a} EIN (PDB code IZYM),^{18b} IIA^{Glc} (PDB code 2F3G, molecule 2),^{18c} and IIA^{Mtl} (PDB code 1A3A, molecule D).^{18d} The experimental structures of the EIN-HPr (PDB code 3EZE),¹⁰ IIA^{Glc}-

HPr (PDB code 1GGR),⁷ and IIA^{Mtl}-HPr (PDB code 1J6T)⁸ complexes were determined by NMR, and the restrained regularized mean structures are used as the reference structures in the present study. The experimental IIA^{Glc}-HPr⁷ and IIA^{Mtl}-HPr⁸ structures were determined using conjoined rigid body/torsion angle dynamics on the basis of intermolecular NOE data and residual dipolar couplings;⁴ hence, the backbone coordinates of the individual proteins in the complexes are identical to those of the free crystal structures (with the exception of four short loop regions in the case of IIA^{Mtl} which undergo small atomic shifts upon complex formation⁸). The NMR structure of the EIN-HPr complex,¹⁰ on the other hand, was determined conventionally using a full complement of intramolecular NOEs, in addition to intermolecular NOEs and residual dipolar couplings, and while the coordinates of the individual components are close to those of the corresponding free X-ray structures (~ 1 Å for EIN and ~ 0.6 Å for HPr), they are not identical. Consequently, the reference structure for the EIN-HPr complex was obtained by best-fitting the X-ray coordinates of free EIN and HPr to the restrained regularized mean NMR structure of the EIN-HPr complex. This ensures that the backbone root-mean-square (rms) difference between the calculated docked structures and the reference structure does not reflect internal coordinate differences between the X-ray and NMR structures.

Back-calculation of dipolar couplings based on molecular shape using a steric obstruction model was carried out using the program SSIA.¹⁹

Results and Discussion

Converting $^1\text{H}_\text{N}/^{15}\text{N}$ Chemical Shift Maps into Highly Ambiguous Distance Restraints. Backbone $^1\text{H}_\text{N}$ and ^{15}N chemical shifts are highly sensitive to environment and have been extensively used to map interaction surfaces on proteins.³ Perturbations in backbone $^1\text{H}_\text{N}$ and ^{15}N chemical shifts are mainly influenced by electronic effects and, in the case of $^1\text{H}_\text{N}$ chemical shifts, by ring current effects as well. (Note that ring current effects arising from aromatic residues are local and generally extend out to only 3–4 Å from the aromatic ring.) It has to be noted, however, that chemical shift perturbation is subject to indirect effects, and hence some degree of common sense and caution are always required to appropriately map a protein–protein interface in this manner. For example, the backbone $^1\text{H}_\text{N}/^{15}\text{N}$ shifts of an internal residue can be significantly perturbed as a consequence of intramolecular interactions with a residue located at the protein–protein interface. Similarly, in cases where the thickness of the molecule at the interface is small (comprising, for example, only two elements of structure such as two helices, or a sheet and a helix), it is possible that sizable backbone $^1\text{H}/^{15}\text{N}$ shifts can be manifested by a residue that is only one layer (i.e., element of structure) removed from the interface and yet whose side chain may be exposed on a surface that is actually located opposite of the interaction surface. Thus, irrespective of the specific details used to select interfacial residues on the basis of chemical shift mapping, a selected residue should satisfy three criteria: (a) significant chemical shift perturbation upon complex formation, (b) at least one or two atoms of the residue should be readily visible on the surface in a space-filling representation of the molecule, and (c) the selected residue should constitute part of a cluster of residues that can form a contiguous, single binding surface. Both (b) and (c) are readily assessed by visual inspection using an appropriate molecular graphics program.

The first step in our procedure is to convert the $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift perturbation maps into a set of highly ambiguous

- (6) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021–9025.
 (7) Wang, G.; Louie, J. M.; Sondej, M.; Seok, Y.-J.; Peterkofsky, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 5635–5649.
 (8) Comilescu, G.; Lee, B. R.; Comilescu, C. C.; Wang, G.; Peterkofsky, A.; Clore, G. M. *J. Biol. Chem.* **2002**, *277*, 42289–42298.
 (9) (a) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127–174. (b) Prestegard, J. H.; Kishore, A. I. *Curr. Opin. Chem. Biol.* **2001**, *5*, 584–590.
 (10) Garrett, D. S.; Seok, Y.-J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166–173.
 (11) (a) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571–10572. (b) Rückert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.
 (12) Garrett, D. S.; Seok, Y.-J.; Peterkofsky, A.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1997**, *36*, 4393–4398.
 (13) (a) Schwieters, C. D.; Kuszewski, J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 66–74. (b) Xplor-NIH is available on-line at <http://nmr.cit.nih.gov/xplor-nih>.
 (14) Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319–333.
 (15) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281–296.
 (16) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *149*, 239–244.
 (17) Schwieters, C. D.; Clore, G. M. *J. Biomol. NMR* **2002**, *23*, 221–225.
 (18) (a) Jia, Z.; Quail, J. W.; Waygood, E. B.; Delbaere, L. T. *J. Biol. Chem.* **1993**, *268*, 22940–22501. (b) Liao, D.-I.; Silverton, E.; Seok, Y.-J.; Lee, B. R.; Peterkofsky, A.; Davies, D. R. *Structure* **1996**, *4*, 861–872. (c) Feese, M. D.; Comolli, L.; Meadow, N. D.; Roseman, S.; Remington, S. J. *Biochemistry* **1997**, *36*, 16087–16096. (d) van Montfort, R. L.; Pijning, T.; Kalk, K. H.; Hangy, I.; Kouwijzer, M. L. C. E.; Robillard, G. T.; Dijkstra, B. W. *Structure* **1998**, *6*, 377–388.

(19) Zwecketter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.

intermolecular distance restraints. The representation that we have chosen makes use of the so-called “ r^{-6} -summed” distance that is generally used to interpret ambiguous NOE assignments,²⁰ with the key difference that in the latter case only a small number of interproton distances are involved. Given N_a residues on protein *A* and N_b residues on protein *B* that have been localized to the protein–protein interface by chemical shift mapping, we derive a set of $(N_a + N_b)$ ambiguous distance restraints (d_{aB} and d_{bA}) between all hydrogen, nitrogen, and oxygen atoms (*i*) of each residue *a* on protein *A* and all hydrogen, nitrogen, and oxygen atoms (*j*) of all residues *b* on protein *B*, and vice versa:

$$d_{aB} = \left(\sum_b \sum_{ij} r_{ai,bj}^{-6} \right)^{-1/6} \quad \text{and} \quad d_{bA} = \left(\sum_a \sum_{ij} r_{ai,bj}^{-6} \right)^{-1/6} \quad (1)$$

where $r_{ai,bj}$ is the distance between atom *i* of residue *a* of protein *A* and atom *j* of residue *b* of protein *B*. The number of atoms per residue range from 5 for Gly to 18 for Arg. Each d_{aB} restraint therefore comprises a set of $r_{ai,bj}$ distances involving 5–18 atoms of residue *a*, depending on the nature of residue *a*, and anywhere between 50 and 250 atoms from protein *B*, depending on the number and type of selected interfacial residues *b* on protein *B*. In the examples presented in this paper, the number of $r_{ai,bj}$ distances encompassed in a single ambiguous distance restraint ranges from 400 to 3000. Each d_{aB} and d_{bA} ambiguous distance restraint is given an upper bound of 5 Å. It should be emphasized that this does not imply that any individual $r_{ai,bj}$ distance is 5 Å or less because d_{aB} is always smaller than the shortest $r_{ai,bj}$ distance. Moreover, a cutoff of 5 Å is actually quite generous because of the nature of the ambiguous distance restraints. Thus, for example, if a given d_{aB} ambiguous distance restraint is made up of 20 individual $r_{ai,bj}$ distances, each 10 Å in length, the value of d_{aB} is 6 Å.

The Simulated Annealing Docking Protocol. The potential surface generated by such highly ambiguous intermolecular distance restraints (eq 1) is rough, and there are many false local minima on the path to the global minimum region of the target function. We have therefore designed a powerful simulated annealing protocol, implemented in Xplor-NIH,¹³ to overcome these barriers and reach the global minimum region of the target function.²¹ The protocol combines both rigid body minimization and conjoined rigid body/torsion angle dynamics in which the interfacial side chains are given their full torsional degrees of freedom.^{4,8} The target function comprises two experimental terms: a quadratic square-well potential for the ambiguous distance restraints²¹ and a harmonic potential for the ¹D_{NH} dipolar couplings.²² (Note that because the backbone is treated as a rigid body, no additional information is gained by measuring other backbone dipolar couplings.) In addition, three terms are used to represent the nonbonded interactions: a quartic van der Waals repulsion term (E_{vdw}) to prevent atomic overlap,²¹ a radius of gyration term (E_{rgyr}),²³ and a side chain torsion angle database potential of mean force (E_{db}) to bias the interfacial side chain

conformations toward those rotamer conformations observed in very high-resolution protein crystal structures.²⁴

Because only upper bounds are employed for the ambiguous distance restraints and because the van der Waals term does not contain an attractive component, the radius of gyration term, E_{rgyr} , is absolutely essential to avoid expansion at the protein–protein interface. Expansion arises because there are many more expanded structures that can satisfy the restraints than compacted ones which can only be attained by a more limited set of side chain configurations. The target value for the radius of gyration, R_{gyr}^{target} , is given by $2.2N^{0.38}$ where N is the number of residues in the calculated complex.²³ The calculated value of R_{gyr}^{target} tends to underestimate the true value of R_{gyr} (R_{gyr}^{true}). The exact value of R_{gyr}^{target} , in the context of the present calculations, however, is not critical providing $R_{gyr}^{target} \leq R_{gyr}^{true}$ because the E_{rgyr} potential is a global soft packing potential and the quartic van der Waals repulsion term prevents atomic overlap. For EIN-HPr, where the value of R_{gyr} for the experimentally determined structure is ~ 22.6 Å, for example, essentially identical results are obtained for R_{gyr}^{target} values of 20 and 22 Å. However, if R_{gyr}^{target} were significantly larger than R_{gyr}^{true} , the E_{rgyr} potential would allow expansion to occur. The R_{gyr}^{target} values employed are 20.0 Å for the EIN-HPr complex (residues 2–249 of EIN + 1–85 of HPr), 17.5 Å for the IIA^{Glc}-HPr complex (residues 19–168 of IIA^{Glc} and 1–85 of HPr), and 17.4 Å for the IIA^{Mtl}-HPr complex (residues 4–147 of IIA^{Mtl} + 1–85 of HPr). Because the backbone and noninterfacial side chains are treated as rigid bodies throughout, E_{rgyr} , in effect, directly reflects the packing quality at the protein–protein interface.

The force constants for the distance and dipolar coupling restraints, and the quartic van der Waals repulsion, radius of gyration, and torsion angle database terms are denoted as k_{dist} , k_{dip} , k_{vdw} , k_{rgyr} , and k_{db} , respectively. In addition, the van der Waals repulsion term also includes a van der Waals radius scale factor s_{vdw} .²¹ To maintain computational efficiency during simulated annealing, the masses of all protein atoms are set to 100 amu, while those of the four atoms of the dipolar coupling alignment axis are set to 5000 amu. This ensures that the moments of inertia of the three rigid bodies (i.e., the two proteins and the axis of the alignment tensor) are comparable, thereby making the time scale of their motion similar. The complete simulated annealing protocol is as follows: (i) Rigid body minimization with one of the two proteins fixed using only the ambiguous distance restraints ($k_{dist} = 0.01$ kcal mol⁻¹ Å⁻²) and the van der Waals repulsion term ($k_{vdw} = 4$ kcal mol⁻¹ Å⁻², $s_{vdw} = 0.8$). (ii) Rigid body dynamics with one protein held fixed using only the ambiguous distance restraints and the van der Waals repulsion term: the temperature is slowly decreased over 40 cycles (60 ps/cycle with the integration time step ranging from 15 fs to 4.6 ps and averaging 1.5 ± 1.1 ps) from 1500 to 500 K in increments of 25 K, while k_{dist} and k_{vdw} are increased geometrically from 0.01 to 30 kcal mol⁻¹ Å⁻² and 0.004 to 1 kcal mol⁻¹ Å⁻⁴, respectively, and s_{vdw} is decreased from 0.9 to 0.75. (iii) Conjoined rigid body/torsion angle dynamics with both proteins free to rotate and translate and with the interfacial side chains (as defined by the ¹H_N/¹⁵N chemical shift mapping) given their full torsional degrees of freedom: all five terms of

(20) (a) Nilges, M. *Proteins* **1993**, *17*, 297–309. (b) Nilges, M. *J. Mol. Biol.* **1995**, *245*, 645–660.

(21) (a) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *12*, 27–38. (b) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1993**, *229*, 317–324.

(22) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159–162.

(23) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337–2338.

(24) Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc.* **2002**, *124*, 2866–2867.

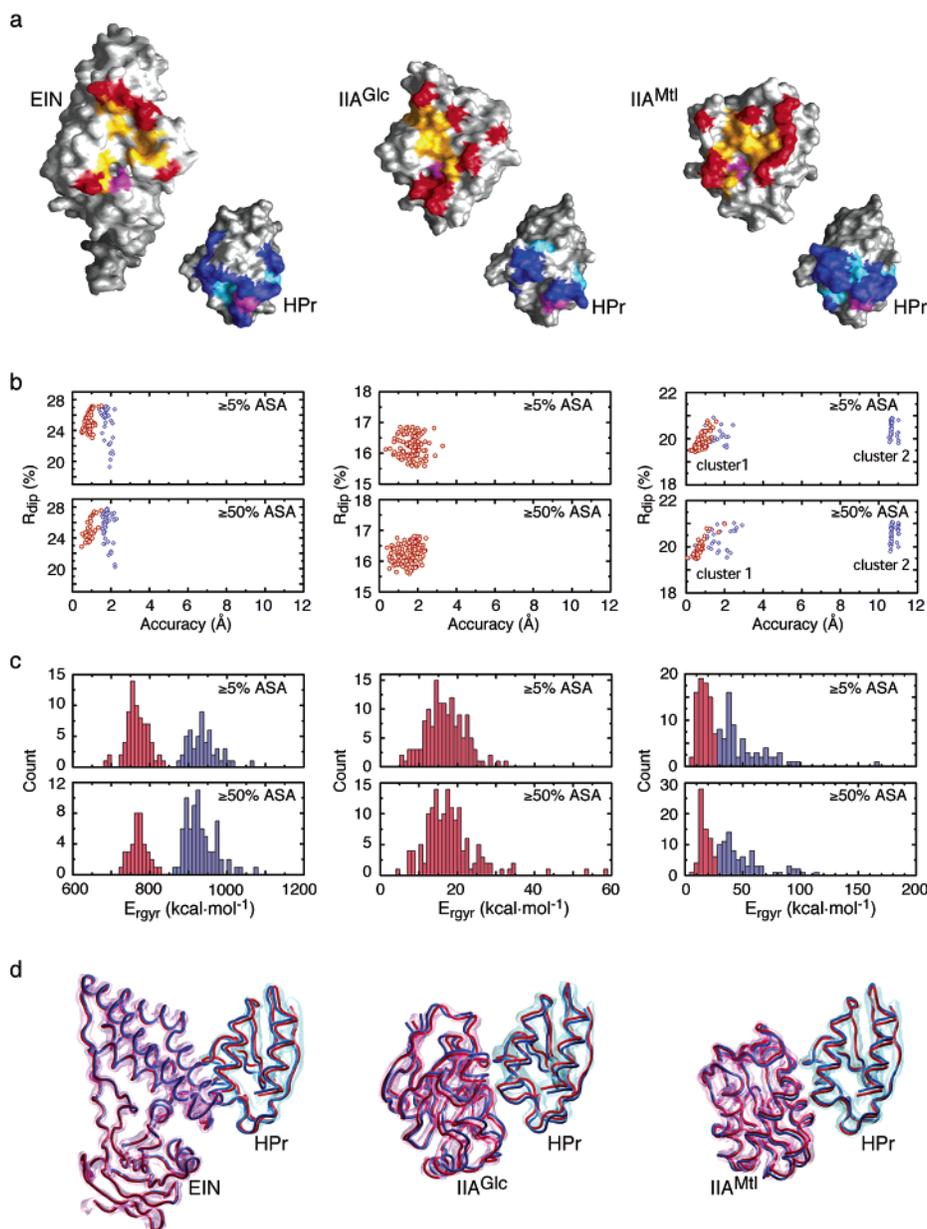


Figure 1. Results of docking calculations for the EIN-HPr (left), IIA^{Glc}-HPr (middle), and IIA^{Mtl}-HPr (right) complexes. (a) Interfacial residues (blue/cyan for HPr, red/orange for the three enzymes, and purple for the active site histidines) identified by ¹H/¹⁵N chemical shift perturbation mapping (see text) displayed on a molecular surface representation of the proteins. (The blue and red colored interfacial residues indicate residues with an accessible surface area (ASA) in the free proteins $\geq 50\%$ of that in an extended Gly-X-Gly peptide; the cyan and orange colored residues indicate interfacial residues in the free proteins with $5\% \leq \text{ASA} < 50\%$.) (b) Plots of the dipolar coupling R -factor (R_{dip}) versus accuracy for the converged structures characterized by no violations $>0.5 \text{ \AA}$ in the highly ambiguous intermolecular distance and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$. In the case of the EIN-HPr (left panel) and IIA^{Mtl}-HPr complexes (right panel), the red circles and blue diamonds indicate structures in the lower and higher energy populations, respectively, of the E_{rgyr} distribution. (c) Histograms of the E_{rgyr} distributions for the converged structures. The E_{rgyr} distribution is unimodal for the IIA^{Glc}-HPr complex (middle), but bimodal for the EIN-HPr (left) and IIA^{Mtl}-HPr (right) complexes. For the bimodal distributions, the lower and higher energy E_{rgyr} populations are colored red and blue, respectively. Note that in the case of the IIA^{Mtl}-HPr complex, all of the structures in the lower energy E_{rgyr} population reside in the correct cluster 1 ensemble; all of the structures in the incorrect cluster 2 ensemble reside in the higher energy E_{rgyr} population. (d) Backbone (depicted as tubes) best-fit superpositions of the average coordinates (red) of the converged structures on the previously determined NMR structures (blue) solved on the basis of intermolecular NOEs and residual dipolar couplings.^{7,8,10} In the case of the IIA^{Mtl}-HPr complex, the mean coordinates are derived from the cluster 1 ensemble. The ensemble distributions of the docked structures are depicted by isosurfaces of the reweighted atomic density maps. Residues 1–85 of HPr, 2–230 of EIN, 19–168 of IIA^{Glc} and 4–147 of IIA^{Mtl} are displayed in (d).

the target function are employed, the temperature is slowly decreased over 59 cycles (3.25 ps/cycle with the integration time step ranging from 1.5 to 80 fs and averaging 20 ± 14 fs) from 1500 to 25 K in increments of 25 K, and k_{dist} , k_{dip} , k_{vdw} , k_{rgyr} , and k_{db} are increased geometrically from 1 to 30 kcal mol⁻¹ Å⁻², 0.001 to 0.01 kcal mol⁻¹ Hz⁻², 0.1 to 1.0 kcal mol⁻¹ Å⁻⁴, 0.01 to 100 kcal mol⁻¹ Å⁻², and 0.002 to 1, respectively, and

s_{vdw} is decreased from 0.78 to 0.75. (iv) Conjoined rigid body/torsion angle minimization with the force constants unchanged except for $k_{\text{vdw}} = 3 \text{ kcal mol}^{-1} \text{ \AA}^{-4}$ and $s_{\text{vdw}} = 0.78$. A complete set of Xplor-NIH input files for the docking protocol is available on-line at <http://spin.niddk.nih.gov/clore>.

Application to the EIN-HPr, IIA^{Glc}-HPr, and IIA^{Mtl}-HPr Complexes. Figure 1 summarizes the results obtained for the

Table 1. Number of Interfacial Residues Used To Generate the Ambiguous Intermolecular Distance Restraints^a

	number of residues used for ambiguous distance restraints					
	EIN-HPr		IIA ^{Glc} -HPr		IIA ^{Mtl} -HPr	
	EIN	HPr	IIA ^{Glc}	HPr	IIA ^{Mtl}	HPr
≥5% ASA ^b	14	15	16	9	14	16
≥50% ASA ^c	6	9	7	6	9	9

^a The selected residues are characterized by $\Delta_{H/N} = [(\Delta\delta^1H_N)^2 + (\Delta\delta^{15}N)^2]^{1/2} \geq 90$ Hz (at a spectrometer frequency of 600 MHz) upon complexation for IIA^{Glc}-HPr and IIA^{Mtl}-HPr and ≥ 60 Hz upon complexation for EIN-HPr (see main text). ^b Only interfacial residues with an accessible surface area (ASA) in the free proteins of $\geq 5\%$ of that in an extended Gly-X-Gly tripeptide are included. The identity of the residues is as follows. For EIN-HPr: residues 68, 69, 72, 79, 82–85, 110, 111, 115, 120, 123, and 126 of EIN, and residues 12–17, 21, 24, 43, 49, and 51–55 of HPr. Note that the active site histidine (His189) of EIN, which is actually located at the very edge of the protein–protein interface,¹⁰ is not included because its backbone amide is only minimally perturbed upon complexation ($\Delta_{H/N} \approx 20$ Hz). For IIA^{Glc}-HPr: residues 42, 46–48, 69, 71, 86, 88, 90, 91, 96, 97, 110, 141, 143, 144 for IIA^{Glc}, and 15–17, 21, 22, 46, 48, 51, and 54 for HPr. For IIA^{Mtl}-HPr: residues 49, 52–54, 93, 112, 115, 116, 119, 120, 123, and 124 for IIA^{Mtl}, and residues 13, 15–17, 19–21, 47–52, and 54–56 for HPr. Note that the active site histidine (His65) of IIA^{Mtl} is not included because its $\Delta_{H/N}$ shift was slightly less than the cutoff value of 90 Hz. ^c Only interfacial residues with an ASA in the free proteins of $\geq 50\%$ of that in an extended Gly-X-Gly tripeptide are included. The identity of the residues is as follows. For EIN-HPr: residues 68, 83, 84, 110, 111, and 120 for EIN, and residues 12, 15–17, 24, 49, 51, 52, and 54 for HPr. For IIA^{Glc}-HPr: residues 69, 86, 94, 96, 97, 110, and 144 for IIA^{Glc}, and residues 15–17, 48, 51, and 54 for HPr. For IIA^{Mtl}-HPr: residues 52, 53, 59, 93, 112, 116, 120, 123, and 124 for IIA^{Mtl}, and residues 15–17, 20, 48, 49, 51, 52, and 56 for HPr.

40 kDa EIN-HPr (left-hand panels), the 30 kDa IIA^{Glc}-HPr (middle panels), and the 30 kDa IIA^{Mtl}-HPr (right-hand panels) complexes. Although enzymes EIN, IIA^{Glc}, and IIA^{Mtl} interact with the same target protein, HPr, they bear no similarity in either primary amino acid sequence or tertiary structure.^{7,8,10}

In this instance, the initial criteria used to select the interfacial residues were based on the combined $^1H_N/^{15}N$ shift perturbation upon complexation, $\Delta_{H/N}$ (in Hz) given by $[(\Delta\delta^1H)^2 + (\Delta\delta^{15}N)^2]^{1/2}$. For the IIA^{Glc}-HPr and IIA^{Mtl}-HPr complexes, residues with $\Delta_{H/N} \geq 90$ Hz (at 600 MHz) were included. For the EIN-HPr complex, the overall shift perturbations were smaller so a cutoff of 60 Hz was employed.²⁵ The second criterion was based on surface accessibility. Two sets of calculations were carried out. In the first set, residues with an accessible surface area (ASA) in the free protein $\geq 5\%$ of that in an extended Gly-X-Gly peptide were employed. In the second set, only residues with an ASA $\geq 50\%$ in the free protein were used. The average number of interfacial residues per protein was ~ 14 for ASA $\geq 5\%$ and ~ 8 for ASA $\geq 50\%$. The locations of these residues on the surfaces of the proteins are displayed in Figure 1a, and the exact number of interfacial residues employed in each case is summarized in Table 1. The ASA cutoffs of $\geq 5\%$ and $\geq 50\%$ represent the extremes for the choice of interfacial residues, and the results below indicate that

between these limits the choice is not critical. With an ASA cutoff greater than 50%, the likelihood of defining a representative interaction surface is low; for an ASA cutoff less than 5%, too many internal residues are likely to be included.

In the case of the IIA^{Glc}-HPr and IIA^{Mtl}-HPr complexes, all of the residues selected in this manner are located on a single face of the molecule and clearly constitute a contiguous interaction surface (Figure 1a, middle and right panels). In the case of the EIN-HPr complex, on the other hand, there were two additional surface accessible residues (Arg131 and Lys135) that experience significant chemical shift perturbations ($\Delta_{H/N}$ of ~ 95 and ~ 130 Hz, respectively) but whose exposed side chains (ASA $> 50\%$) are located on the face opposite of the interaction face (i.e., the backside of the molecule), and hence are not visible in the view shown in Figure 1a (left-hand panel). This phenomenon is readily explained. Both Arg131 and Lys135 are located in helix 4; one exposed face of helix 4 constitutes part of the binding surface located in the front of the molecule (in the view shown in Figure 1a), while the other exposed face of helix 4 is directed toward the backside of the molecule. This is an example of indirect, short range, effects resulting in chemical shift perturbation of residues outside of the interaction surface. Clearly, neither Arg131 nor Lys135 are part of a cluster of perturbed residues that can form a contiguous binding surface. For consistency with the guidelines put forward in the preceding section dealing with the conversion of $^1H_N/^{15}N$ chemical shift maps into ambiguous distance restraints, Arg131 and Lys135 were therefore excluded from the calculations. However, test calculations showed that the inclusion of Arg131 and Lys135 in the ambiguous distance restraints has absolutely no impact on the results. The reason for this lies in the very generous nature of the ambiguous distance restraints defined by eq 1 with an upper bound of 5 Å; thus, examination of the experimentally determined NMR structure of the EIN-HPr complex¹⁰ reveals that $d_{\text{Arg131,HPr}}$ is less than 5 Å and $d_{\text{Lys135,HPr}}$ is only 5.5 Å. The latter can readily be reduced to ≤ 5 Å by minor alterations in the side chain conformation of Lys135 without having any impact on either translation or orientation of HPr relative to EIN.

For each set of calculations, 300 simulated annealing structures were computed with randomly assigned initial velocities starting with the X-ray coordinates of the proteins¹⁸ placed 50–100 Å away from each other, in four random orientations and directions (i.e., 75 structures were calculated per orientation). Because of the complexity of the energy landscape on the path to the global minimum region, not all calculated structures converged. Structure selection from the ensemble of 300 calculated structures was therefore carried out using a simple two-step procedure based on the ambiguous intermolecular distance restraints and dipolar couplings. In the first step, structures with ambiguous intermolecular distance restraint violations greater than 0.5 Å were excluded from further consideration. For the calculations using ambiguous intermolecular distance restraints derived from residues with ASA $\geq 5\%$, 251, 286, and all 300 structures for the EIN-HPr, IIA^{Glc}-HPr, and IIA^{Mtl}-HPr complexes, respectively, converged with no ambiguous distance restraint violations > 0.5 Å. In the second step, only those structures with residual dipolar coupling R -factors, R_{dip} ,²⁶ in the first half of the R_{dip} distribution (i.e., $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$) were retained. The choice of the median as a

(25) Calculations for the EIN-HPr complex were also carried out using a $\Delta_{H/N}$ cutoff of 80 Hz. For calculations using interfacial residues in the free proteins with ASA $\geq 5\%$, there were 9 residues for EIN and 13 for HPr, and the results (backbone mean coordinate accuracy of ~ 1 Å for the structures with no ambiguous intermolecular distance violations > 0.5 Å and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$) are very similar to those obtained using a $\Delta_{H/N}$ cutoff of 60 Hz. For calculations using interfacial residues with ASA $\geq 50\%$ in the free proteins, however, there are only 4 residues for EIN and 8 residues for HPr. The number of interfacial residues for EIN in this instance is clearly too small to define a representative binding surface, and consequently the backbone mean coordinate accuracy (~ 1.6 Å) of the converged structures is a little lower.

Table 2. Statistics of Structural Convergence and Selection^a

	EIN-HPr	IIA ^{Glc} -HPr	IIA ^{Mtl} -HPr
(1) Number of Structures with No Distance Violations >0.5 Å ^b			
≥5% ASA	251 (83.7%)	286 (95.3%)	300 (100%)
≥50% ASA	230 (76.7%)	296 (98.7%)	300 (100%)
(2) Number of Structures from (1) with $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ c,d			
≥5% ASA ^c	126 (42.0%)	143 (47.7%)	150 (50%)
≥50% ASA ^d	115 (38.3%)	148 (49.3%)	150 (50%)

^a The total number of structures calculated in each case is 300. Two sets of calculations were carried out for each complex using interfacial residues with either ASA ≥ 5% or ≥50% in the free proteins to generate the ambiguous intermolecular distance restraints (see Table 1). The percentage of retained structures (out of the total of 300 calculated) is given in parentheses. ^b Although selection was based on a distance violation cutoff of 0.5 Å, in fact none of the converged structures exhibited distance violations >0.1 Å. ^c The ranges for the first and second halves of the R_{dip} distribution after the first selection stage, based on violations of ambiguous intermolecular distance restraints, are 19.3–27.2% and 27.2–63.9% for EIN-HPr; 15.6–16.9% and 16.9–39.1% for the IIA^{Glc}-HPr; 19.4–20.9% and 20.9–34.7% for the IIA^{Mtl}-HPr. ^d The ranges for the first and second halves of the R_{dip} distribution after the first selection stage, based on violations of ambiguous intermolecular distance restraints, are 20.2–27.7% and 27.8–63.1% for EIN-HPr; 15.6–16.8% and 16.8–38.1% for the IIA^{Glc}-HPr; 19.5–21.1% and 21.1–31.4% for the IIA^{Mtl}-HPr.

cutoff is based on the observation that the R_{dip} distribution is highly skewed (see footnotes c and d to Table 2). Thus, after the R_{dip} -based selection, there are 126, 143, and 150 structures for the EIN-HPr, IIA^{Glc}-HPr, and IIA^{Mtl}-HPr complexes, respectively. The corresponding numbers of converged structures for the calculations employing ambiguous intermolecular distance restraints derived from interfacial residues with only ASA ≥ 50% are comparable and are summarized in Table 2.

Figure 1b displays plots of R_{dip} , versus coordinate accuracy, defined as the backbone atomic rms difference between the simulated annealing structures remaining after the second stage of the selection procedure and the experimental coordinates determined previously on the basis of a full complement of intermolecular NOEs and residual dipolar couplings (see Methods section). For the correct solution, the value of R_{dip} for the complex should be approximately comparable to that obtained when best-fitting the two proteins of the complex individually to separate alignment tensors. The range of R_{dip} values for $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ is narrow and satisfies this criteria (see footnotes b and c to Table 2). Note, however, that there are uncertainties in the determination of the magnitude and orientation of the alignment tensor as a consequence of noise generated from errors in the X-ray coordinates of the individual proteins, as well as uncertainties in the experimental measurement of the ¹D_{NH} dipolar couplings themselves.²⁷ Thus, it is important to stress that while the weighted mean of the R_{dip} values for the individual proteins represents an absolute lower limit of R_{dip} for the complex (because the backbone coordinates are held rigid), R_{dip} values for the complex that are somewhat higher are still acceptable.

In the case of both the EIN-HPr and the IIA^{Glc}-HPr complexes, all of the structures that exhibit no distance violations ≥0.5 Å with $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ converge to the same region of conformational space (Figure 1b and Table 3). The backbone

Table 3. Backbone Ensemble Precision and Coordinate Accuracy for Converged Structures with No Distance Violations >0.5 Å and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ a

	EIN-HPr	IIA ^{Glc} -HPr	IIA ^{Mtl} -HPr ^b	
			cluster 1	cluster 2
≥5% ASA				
$\langle R_{\text{dip}} \rangle$ (%) ^c	25.3 ± 1.5	16.2 ± 0.3	19.9 ± 0.3	20.5 ± 0.3
ensemble precision (Å)	1.11 ± 0.28	0.90 ± 0.39	0.74 ± 0.55	0.76 ± 0.68
ensemble accuracy (Å)	1.26 ± 0.48	1.70 ± 0.46	0.96 ± 0.46	10.68 ± 0.11
mean coordinate accuracy (Å)	0.71	1.47	0.52	10.66
≥50% ASA				
$\langle R_{\text{dip}} \rangle$ (%) ^c	25.8 ± 1.6	16.3 ± 0.3	20.0 ± 0.4	20.6 ± 0.3
ensemble precision (Å)	1.12 ± 0.38	0.75 ± 0.34	0.84 ± 0.63	0.87 ± 0.64
ensemble accuracy (Å)	1.52 ± 0.48	1.54 ± 0.52	0.95 ± 0.62	10.72 ± 0.15
mean coordinate accuracy (Å)	1.06	1.40	0.41	10.68

^a Two sets of calculations were carried out for each complex using interfacial residues with either ASA ≥ 5% or ≥50% to generate the ambiguous intermolecular distance restraints (see Table 1). Backbone ensemble precision is defined as the average backbone (N, Cα, C') atomic rms difference between the individual simulated annealing structures and the mean coordinates of the ensemble (obtained after best-fitting the individual simulated annealing structures to the backbone atoms of all residues of the complex); ensemble accuracy is the average backbone atomic rms difference between the individual simulated annealing structures and the coordinates of the experimentally determined structures derived from the full complement of intermolecular NOE and dipolar coupling data (see Methods section). The mean backbone coordinate accuracy is the backbone atomic rms difference between the mean coordinates of the ensemble of simulated annealing structures and the experimental coordinates. ^b There are two clusters of structures for the IIA^{Mtl}-HPr complex. The first cluster represents the correct solution, while the second cluster represents an alternate incorrect solution. The ratio of the number of structures in the first cluster to the number in the second cluster is ~2. For the calculations using interfacial residues with ASA ≥ 5%, there are 102 structures in the first correct cluster and 48 structures in the second incorrect cluster. The corresponding numbers for the calculations using interfacial residues with ASA ≥ 50% are 99 and 51, respectively. ^c The definition of the dipolar coupling R -factor is given in ref 26b.

accuracy of the mean coordinates ranges from ~0.7 to ~1.5 Å and is comparable to backbone coordinate precision which ranges from ~0.8 to 1.1 Å (Table 3). In addition, the results obtained with ambiguous intermolecular distance restraints derived using interfacial residues with ASA ≥ 5% and ≥50% in the free proteins are comparable, illustrating the robustness of the protocol (Table 3).

Satisfying the ambiguous intermolecular distance restraints and the dipolar couplings is not necessarily sufficient to ensure a single correct solution. Thus, in the case of the IIA^{Mtl}-HPr complex, there are two distinct clusters in which the orientation of HPr relative to IIA^{Mtl} differs by a 180° rotation about the z axis of the alignment tensor (Figure 2). The presence of two distinct structural solutions in this instance arises from an unfavorable combination of the orientation of the alignment tensor and the intrinsic degeneracy of the ambiguous distance restraints. The mean coordinates of the first cluster are 0.5–0.7 Å away from the correct solution, whereas those of the second cluster are ~10.7 Å away (Table 3). The number of structures in the first cluster is about double that in the second. Although the average R_{dip} value for the first cluster (19.9 ± 0.3%) is a little lower than that for the second (20.5 ± 0.3%), it is clear that no distinction between the two solutions can be made on the basis of R_{dip} alone.

(26) (a) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008–9012.(b) The dipolar coupling R -factor R_{dip} is given by the ratio of the of the rms deviation between observed and calculated values to the expected rms deviation if the vectors were randomly oriented, given by $[2D_a^2(4 + 3\eta^2)/5]^{1/2}$, where D_a is the magnitude of the axial component of the alignment tensor, and η is the rhombicity.(27) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127–137.

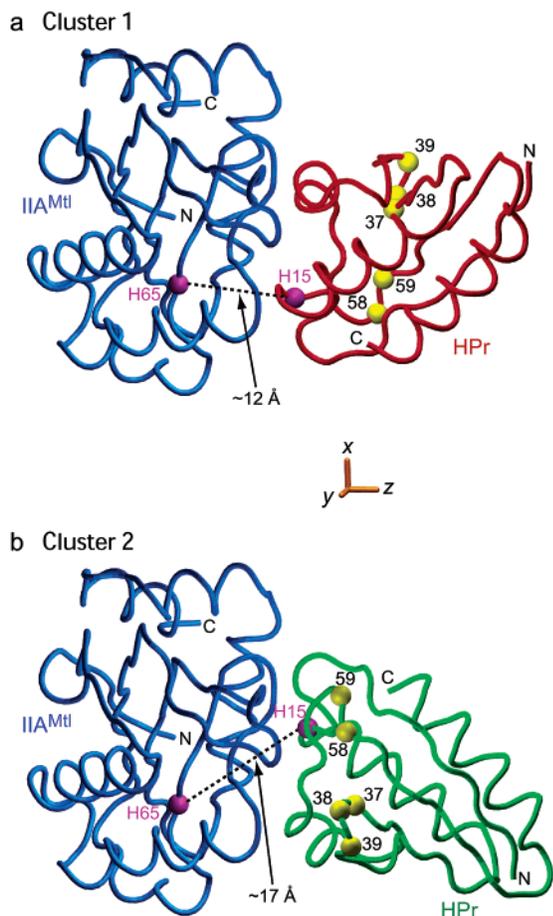


Figure 2. Comparison of the mean coordinates of the correct cluster 1 (a) and incorrect cluster 2 (b) ensembles of structures obtained for the IIA^{Mtl}-HPr complex. The backbone coordinates are displayed as a tube. IIA^{Mtl} is shown in blue in the same orientation in (a) and (b). The orientation of HPr in the two clusters (red for cluster 1 and green for cluster 2) differs by a 180° rotation about the z axis of the alignment tensor (displayed in orange). Note that the x and y axes of the alignment tensor lie in the plane of the protein–protein interface so that a 180° rotation about the z axis of the alignment tensor does not result in steric clash. In this particular case, the ambiguous intermolecular distance restraints cannot distinguish between the two alternative orientations. The C α atoms of the active site histidines (His65 of IIA^{Mtl} and His15 of HPr) are shown as purple spheres. In the correct cluster 1 ensemble, the C α –C α separation between the active site histidines is 12.3 ± 0.7 Å, fully consistent with the formation of a pentacoordinate phosphoryl transfer intermediate that has been established by biochemical methods.³¹ The C α atoms of five residues of HPr (Ser37, Asn38, Gly39, Gly58, and Thr59) that exhibit minimal ¹H_N/¹⁵N chemical shift perturbation upon complexation but are predicted to be at the protein–protein interface in the cluster 2 ensemble are depicted as yellow spheres; these residues are far from the protein–protein interface in the correct cluster 1 ensemble.

Distinguishing between Alternate Solutions. For an asymmetric alignment tensor, the ¹D_{NH} dipolar couplings measured in a single alignment medium are consistent with four possible relative protein–protein orientations, two of which differ by a 180° rotation about the z axis of the alignment tensor, and the other two by a 180° rotation about the y axis of the alignment tensor. In most instances, exemplified by the EIN-HPr and IIA^{Glc}-HPr complexes, the ambiguous intermolecular distance restraints derived from ¹H_N/¹⁵N chemical shift mapping resolve the fourfold degeneracy such that only a single orientation is consistent with both the ambiguous intermolecular distance restraints and the ¹D_{NH} dipolar couplings.

In unfavorable cases, such as the IIA^{Mtl}-HPr complex, the ambiguous intermolecular distance restraints only reduce the number of solutions to two. The 2-fold reduction in degeneracy is achieved because the ambiguous intermolecular distance restraints ensure that the two binding surfaces are opposed, and interpenetration of the two molecules is prohibited by the van der Waals repulsion term. In the case of the IIA^{Mtl}-HPr complex, the persistence of twofold degeneracy arises from the fact that the x and y axes of the alignment tensor lie in the plane of the protein–protein interface, such that a 180° rotation about the z axis can occur without interpenetration of the two molecules (Figure 2).

In those instances where twofold degeneracy remains resulting in two alternative protein–protein orientations, a variety of experimental, computational, and empirical approaches can be used to distinguish the correct solution from the incorrect one. These are summarized below.

The simplest initial approach is to reexamine the ¹H_N/¹⁵N chemical shift perturbation maps in the light of the calculated structures and assess whether these maps can permit one to distinguish between the two alternative solutions. In the case of the IIA^{Mtl}-HPr complex, this qualitative approach permits a straightforward discrimination between the two structural solutions. Thus, while both cluster 1 (Figure 2a) and 2 (Figure 2b) solutions are consistent with the ambiguous intermolecular distance restraints derived from residues that exhibit significant ¹H_N/¹⁵N chemical shift perturbation upon complexation, the cluster 2 solution is not fully consistent with the observed ¹H_N/¹⁵N chemical shift perturbation map. Specifically, there are five surface exposed residues of HPr (Ser37, Asn38, Gly39, Gly58, and Thr59) that are present at the interface in the cluster 2 solution and yet exhibit only minimal $\Delta_{H/N}$ (5–20 Hz) shifts upon complexation (Figure 2b). In contrast, in the cluster 1 solution, these five residues are far from the interface (Figure 2a). A simple method for incorporating this type information directly into the calculations is to introduce repulsive ambiguous intermolecular distance restraints with $d_{ab} \geq 0.5$ Å between each residue a on protein A with a minimal $\Delta_{H/N}$ to all residues with significant $\Delta_{H/N}$ on protein B. The results of such calculations are shown in Figure 3. The twofold degeneracy is completely lifted, and all converged structures (i.e., no distance violations ≥ 0.5 Å and $R_{dip} \leq R_{dip}^{median}$) now reside in the cluster 1 ensemble. Clearly, in this instance, the ambiguous repulsive restraints were added in an ad hoc manner after visual inspection of the structures. However, the calculations suggest that automatic introduction of repulsive restraints is readily feasible.

If the liquid crystalline medium employed is neutral and orientational order is governed by steric interactions between the liquid crystalline medium and the complex, one can differentiate between alternate solutions on the basis of molecular shape using a steric obstruction model to back-calculate the alignment tensor and residual dipolar couplings.^{19,28} For the IIA^{Mtl}-HPr complex,⁸ the ¹D_{NH} dipolar couplings were measured in a neutral poly(ethylene glycol)/hexanol liquid crystalline medium.^{11b} The experimental values of the axial component of the alignment tensor (D_a^{NH}) and the rhombicity (η) are 12.1 Hz and 0.42, respectively, and the average value of R_{dip} for all 150 structures with $R_{dip} \leq R_{dip}^{median}$ (i.e., clusters 1 and 2 combined) is $20.2 \pm 0.5\%$. The predicted values of D_a^{NH} and η using the

(28) Bewley, C. A.; Clore, G. M. *J. Am. Chem. Soc.* **2000**, *122*, 6009–6016.

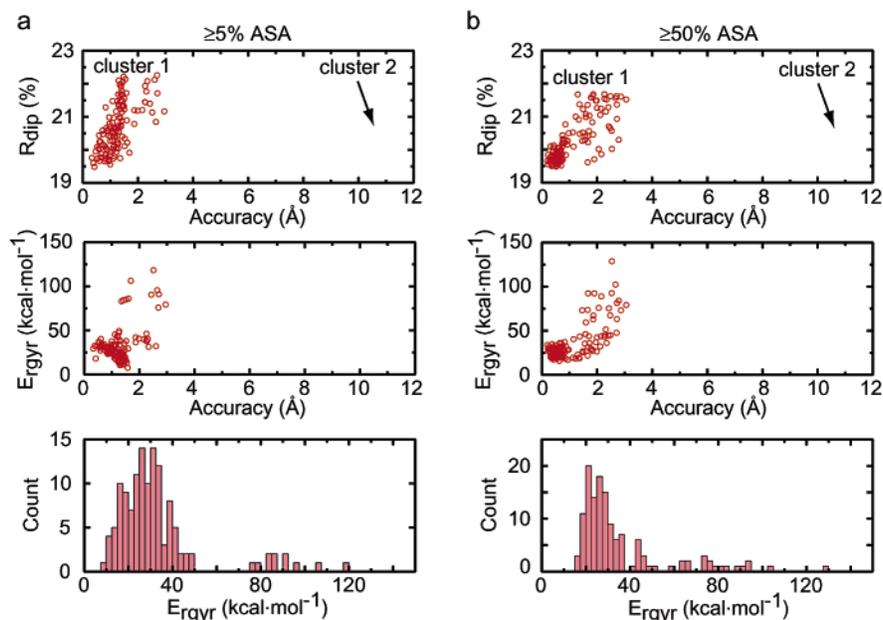


Figure 3. Results of docking calculations for the IIA^{Mtl}-HPr complex using both attractive ($d_{AB} \leq 5 \text{ \AA}$) and repulsive ($d_{AB} \geq 5 \text{ \AA}$) ambiguous intermolecular distance restraints. The repulsive restraints involve five surface exposed residues of HPr (Ser37, Asn38, Gly39, Gly58, and Thr59) that exhibit only minimal Δ_{HN} (5–20 Hz) shifts upon complexation (cf. Figure 2). Incorporation of the repulsive intermolecular distance restraints resolves the 2-fold degeneracy (cf. Figure 1b, right-hand panel, and Figure 2), and all structures with no violations $>0.5 \text{ \AA}$ in the ambiguous intermolecular distance restraints and $R_{\text{dip}} \geq R_{\text{dip}}^{\text{median}}$ converge to the correct cluster 1 solution. Structures were calculated with attractive ambiguous intermolecular distance restraints derived from interfacial residues with (a) ASA $\geq 5\%$ and (b) ASA $\geq 50\%$ (cf. Table 1). A total of 300 simulated annealing structures was calculated in each case, and the number of structures with no distance violations $>0.5 \text{ \AA}$ and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ was 130 and 134, respectively. The top panels show plots of the dipolar coupling R -factor, R_{dip} , versus accuracy. The values of $\langle R_{\text{dip}} \rangle$, the ensemble precision, ensemble accuracy, and mean coordinate accuracy for the structures in (a) are $20.7 \pm 0.7\%$, $0.83 \pm 0.53 \text{ \AA}$, $1.18 \pm 0.43 \text{ \AA}$, and 0.77 \AA , respectively; the corresponding values for the structures in (b) are $20.2 \pm 0.7\%$, $0.99 \pm 0.74 \text{ \AA}$, $1.07 \pm 0.81 \text{ \AA}$, and 0.46 \AA , respectively. The middle panels show plots of E_{rgyr} versus accuracy, and the lower panels show histograms of the E_{rgyr} distribution. The E_{rgyr} distribution is unimodal but highly skewed with a mode at $\sim 27 \text{ kcal mol}^{-1}$ in (a) and 24 kcal mol^{-1} in (b). The individual simulated annealing structures in the tails of the E_{rgyr} distribution ($E_{\text{rgyr}} \geq 55 \text{ kcal mol}^{-1}$) are of lower accuracy.

steric obstruction model for the cluster 1 structures are $12.7 \pm 0.8 \text{ Hz}$ and 0.26 ± 0.04 , respectively, with a back-calculated R_{dip} value of $30 \pm 4\%$ and a correlation coefficient of 0.91 ± 0.02 . For the cluster 2 structures, on the other hand, the predicted values of D_a^{NH} and η are $16.1 \pm 0.6 \text{ Hz}$ and 0.22 ± 0.04 , respectively, with a back-calculated R_{dip} value of $49.2 \pm 3.4\%$ and a correlation coefficient of 0.73 ± 0.06 . One can therefore conclude that the measured dipolar couplings are fully consistent with the molecular shape of the cluster 1 structures but incompatible with that of the cluster 2 structures.

From an experimental NMR standpoint, the two alternate solutions can also be readily distinguished by measuring a second set of $^1\text{D}_{\text{NH}}$ dipolar couplings in an alternate liquid crystalline medium characterized by a significantly different alignment tensor (e.g., charged versus uncharged alignment media).^{5,29} While simple in principle, this may not always be possible in practice because of the presence of interactions between the complex and the liquid crystalline medium which preclude the measurement of residual dipolar couplings. For example, both the EIN-HPr and IIA^{Glc}-HPr complexes bind to bicelle-based³⁰ liquid crystalline media.

Prior biochemical information, when available, can also be employed. For the IIA^{Mtl}-HPr complex, an upper bound of 14 \AA for the $\text{C}\alpha$ – $\text{C}\alpha$ distance between the active site histidines (His65 of IIA^{Mtl} and His15 of HPr) is readily calculated because phosphoryl transfer between the active site histidines is known, from isotope studies, to occur with inversion of the configuration

of the phosphorus, indicative of a transition state with a pentacoordinate phosphoryl group.^{31a} Hence, in a dissociative transition state complex, the distance between the $\text{N}\epsilon 2$ atom of His65^{31b} and the $\text{N}\delta 1$ of His15^{31c} would be $\sim 6 \text{ \AA}$, and in an associative ($\text{S}_{\text{N}}2$) transition state, this distance would be reduced to $\sim 4 \text{ \AA}$. The $\text{C}\alpha$ – $\text{C}\alpha$ distance between the active site histidines in the cluster 1 structures is $12.3 \pm 0.7 \text{ \AA}$, while in the cluster 2 structures it is $\sim 17.2 \pm 1.6 \text{ \AA}$. Thus, biochemical and functional considerations permit one to eliminate the cluster 2 structures. Obviously, in other cases, such clear-cut prior biochemical data may not be available, in which case one would have to resort to making a few selected site-specific mutations and ascertaining their effect in an appropriate binding or functional assay.

Finally, an empirical method based on an effective packing score, in the form of the radius of gyration term, E_{rgyr} , can also be used. This is discussed in detail in the following section.

Discrimination Using E_{rgyr} as a Measure of Packing Quality. Of the three nonbonded terms in the target function, the E_{rgyr} term most directly reflects the backbone accuracy metric as a consequence of its global nature. A change in E_{rgyr} corresponds to a relative displacement of all inter-protein atoms and thus is directly associated with a change in backbone accuracy. E_{rgyr} is a soft packing potential that reflects the overall packing density, and hence surface complementarity, at the

(31) (a) Mueller, E. G.; Khandekar, S. S.; Knowles, J. R.; Jacobson, G. R. *Biochemistry* **1990**, *29*, 6892–6896. (b) Van Dijk, A. A.; de Lange, L. C. M.; Bachovchin, W. W.; Robillard, G. T. *Biochemistry* **1992**, *31*, 5552–5556. (c) Weigel, N.; Powers, D. A.; Roseman, S. *J. Biol. Chem.* **1982**, *257*, 14499–14509.

(29) Ramirez, B.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106–9107.

(30) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.

protein–protein interface (because the backbone and noninterfacial side chains are treated as rigid bodies) and is therefore not particularly influenced by details of local interactions. E_{vdw} and E_{db} , on the other hand, are heavily influenced by local interactions, and large changes in their values (due, for example, to a single bad nonbonded contact or one poor rotamer for a side chain at the interface) can readily be associated with very small changes in backbone positions. Indeed, backbone accuracy for the structures with $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$ is not at all correlated with E_{vdw} or E_{db} . Consequently, we have found examination of the E_{rgyr} distribution to be useful.

Figure 1c displays histograms of E_{rgyr} for those structures with no ambiguous intermolecular distance violations $>0.5 \text{ \AA}$ and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$. For the IIA^{Glc}-HPr complex, the E_{rgyr} distribution is unimodal and approximately normal (Figure 1c, middle panel), and only a single cluster of structures is observed (Figure 1b, middle panel). For the IIA^{Mtl}-HPr and EIN-HPr complexes, on the other hand, a clear-cut bimodal E_{rgyr} distribution is observed (Figure 1c, right and left panels, respectively). The presence of bimodality indicates the presence of two structure populations characterized by different overall dimensions and shape, with the higher energy E_{rgyr} population being more expanded (i.e., larger R_{gyr}). The results obtained using ambiguous intermolecular distance restraints derived from either interfacial residues with $\text{ASA} \geq 5\%$ or $\geq 50\%$ in the free proteins are very similar with the exception that the occupancy of the lower E_{rgyr} population is slightly reduced for the latter (Figure 1c, left and right panels; Table 4).

Examination of the plot of R_{dip} versus accuracy for the IIA^{Mtl}-HPr complex (with the lower energy E_{rgyr} population colored in red, and the higher energy E_{rgyr} population in blue) reveals that all of the structures in the lower E_{rgyr} energy population but none of the structures in the higher E_{rgyr} energy population reside in the correct cluster 1 ensemble. Thus, all of the structures in the incorrect cluster 2 ensemble reside in the higher E_{rgyr} energy population (Figure 2b, right panel).

The overall distribution of E_{rgyr} for the IIA^{Mtl}-HPr cluster 1 ensemble, while unimodal, is in fact highly skewed (cf. Figure 3c for the structures calculated with additional ambiguous intermolecular repulsive restraints), and the presence of higher energy (more loosely packed) E_{rgyr} structures within cluster 1 corresponds to the tail of the cluster 1 E_{rgyr} distribution. While the structures in the tail of the E_{rgyr} distribution are both less precise and less accurate, the accuracies of their mean backbone coordinates ($\sim 0.7 \text{ \AA}$) are only slightly worse than those of either the structures in the main envelope of the E_{rgyr} distribution ($\sim 0.5\text{--}0.6 \text{ \AA}$) or of all of the cluster 1 structures combined ($\sim 0.4\text{--}0.5 \text{ \AA}$) (Tables 3 and 4).

For the EIN-HPr complex, although all converged structures fall into a single cluster in Figure 1b (left), it is evident from the multicolored plots of R_{dip} versus accuracy that there are two subpopulations of structures within this cluster with a boundary of $\sim 1.5 \text{ \AA}$ in backbone coordinate accuracy. These two subpopulations reflect two distinct populations in the E_{rgyr} distribution (Figure 1c, left), with the lower energy E_{rgyr} population (red) corresponding to the more accurate structures (Table 4 and Figure 1b, left). The difference between the two subpopulations is primarily a translational one which is reflected in E_{rgyr} : the $\text{C}\alpha\text{--C}\alpha$ separation between the two active site histidines (His189 of EIN and His15 of HPr) is $13.9 \pm 0.5 \text{ \AA}$

Table 4. Discrimination of Subpopulations of Structures on the Basis of the E_{rgyr} Distribution for the EIN-HPr and IIA^{Mtl}-HPr (Cluster 1) Complexes^a

	EIN-HPr		IIA ^{Glc} -HPr cluster 1 ^b	
	lower E_{rgyr}	higher E_{rgyr}	lower E_{rgyr}	higher E_{rgyr}
	$\geq 5\% \text{ ASA}^c$			
occupancy (%)	57.6%	42.4%	77.0%	23.0%
$\langle R_{\text{dip}} \rangle$ (%)	25.3 ± 1.1	25.4 ± 2.0	19.9 ± 0.3	20.1 ± 0.34
E_{rgyr} range (kcal mol ⁻¹)	685–838	871–1065	6.3–26.2	28.1–98.9
ensemble precision (\AA) ^d	0.50 ± 0.20	0.68 ± 0.32	0.47 ± 0.22	1.39 ± 0.40
ensemble accuracy (\AA) ^d	0.88 ± 0.19	1.78 ± 0.16	0.78 ± 0.27	1.53 ± 0.52
mean coordinate accuracy (\AA) ^d	0.72	1.62	0.64	0.71
	$\geq 50\% \text{ ASA}^c$			
occupancy (%)	33.3%	66.7%	64.3%	35.4%
$\langle R_{\text{dip}} \rangle$ (%)	25.1 ± 1.3	26.1 ± 1.7	19.9 ± 0.3	20.2 ± 0.5
E_{rgyr} range (kcal mol ⁻¹)	728–829	870–1071	7.6–26.8	27.8–102
ensemble precision (\AA) ^d	0.53 ± 0.21	0.63 ± 0.35	0.41 ± 0.29	1.42 ± 0.48
ensemble accuracy (\AA) ^d	0.93 ± 0.21	1.80 ± 0.16	0.67 ± 0.28	1.47 ± 0.73
mean coordinate accuracy (\AA) ^d	0.76	1.65	0.48	0.67

^a The boundary between the lower and higher E_{rgyr} energy populations is at $\sim 840 \text{ kcal mol}^{-1}$ for the EIN-HPr complex (Figure 1c, left panel) and at $\sim 27 \text{ kcal mol}^{-1}$ for the IIA^{Mtl}-HPr complex (Figure 1c, right panel). Note that the higher values of E_{rgyr} for the EIN-HPr complex are simply due to the fact that the value of 20 \AA for $R_{\text{gyr}}^{\text{target}}$ calculated using the empirical relationship $2.2N^{0.38}$ (where N is the number of atoms) is a little underestimated, and the quartic van der Waals repulsion term imposes a hard lower limit for R_{gyr} of $\sim 22.5 \text{ \AA}$. Repeating the EIN-HPr calculations using a value of 22 \AA yields essentially the same results in terms of coordinate precision and accuracy, except that E_{rgyr} spans a range from ~ 61 to $\sim 250 \text{ kcal mol}^{-1}$, and the boundary between the lower and higher energy E_{rgyr} populations is $\sim 110 \text{ kcal mol}^{-1}$. ^b The structures for the IIA^{Mtl}-HPr complex correspond to those calculated in Figure 1 (right-hand panels) using attractive ambiguous intermolecular restraints and do not include the use of ambiguous intermolecular repulsive restraints. ^c Two sets of calculations were carried out for each complex using interfacial residues with either $\text{ASA} \geq 5\%$ or $\geq 50\%$ in the free proteins to generate the ambiguous intermolecular distance restraints (see Table 1). ^d Backbone ensemble precision, ensemble accuracy, and mean coordinate accuracy are defined in footnote a of Table 3.

for the lower energy E_{rgyr} population, but $15.5 \pm 0.5 \text{ \AA}$ for the higher energy E_{rgyr} population. Thus, the $\text{C}\alpha\text{--C}\alpha$ distance in the higher energy E_{rgyr} population is a little long to permit phosphoryl transfer to occur. Despite the presence of these two subpopulations with substantially different accuracies (both in terms of the ensembles and their respective mean coordinates, Table 4), it is worth noting that the backbone accuracy of the mean coordinates of the overall ensemble (Table 3) is comparable to that of the mean coordinates for the lower energy E_{rgyr} population (Table 4).

Comparison with Docking Based on Dipolar Couplings and Ring Current Shift Calculations. Our approach differs significantly from one proposed recently in which the difference between experimental $^1\text{H}_\text{N}$ chemical shift perturbations and those calculated from ring current shifts is used to locate the proteins, previously oriented by residual dipolar couplings, on a pre-defined grid.³² The latter approach, which was illustrated for the EIN-HPr complex, depends critically on the orientation of aromatic side chains at the interface. Phe48 of HPr plays a

(32) McCoy, M. A.; Wyss, D. F. *J. Am. Chem. Soc.* **2002**, *124*, 2104–2105.

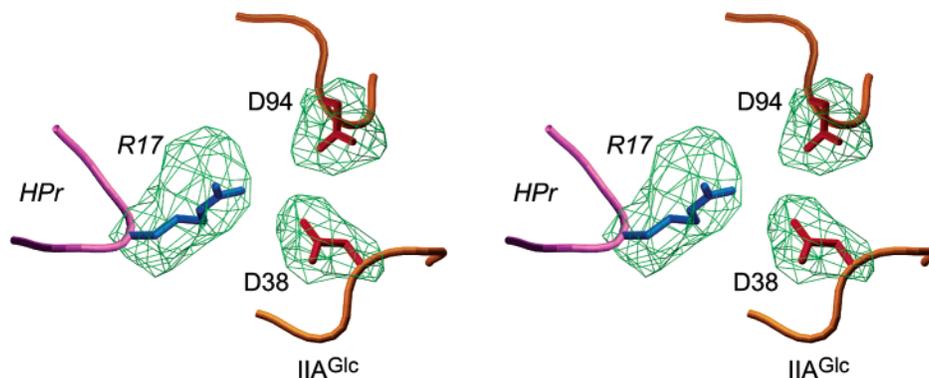


Figure 4. Stereoview illustrating the interaction of Arg17 of HPr with Asp38 and Asp94 of IIA^{Glc} in the docked IIA^{Glc}-HPr complex. The backbone of the mean coordinates is shown as tubes (purple for HPr, orange for IIA^{Glc}). The isosurface of the reweighted atomic density map (contoured at 20% of the maximum value) for Arg17, Asp38, and Asp94, calculated from the ensemble of 143 converged structures (no violations >0.5 Å in ambiguous distance restraints and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$; cf. Table 2), is shown in green. It is readily apparent from the atomic density map that Arg17 can form salt bridges with both Asp38 and Asp94. To guide the eye, the side chain of Arg17 (blue) of HPr has been fitted into the atomic density map with $\chi_1/\chi_2/\chi_3$ side chain torsion angles in a $g^-/t/g^-$ conformation; the side chains of Asp38 and Asp94 (orange) of IIA^{Glc} have likewise been fitted into the atomic density map with χ_1 angles in the t and g^- rotamers, respectively. (The ensemble of docked structures used to generate the figure was obtained from the calculations using interfacial residues with ASA $\geq 5\%$ to derive the ambiguous intermolecular distance restraints, cf. Tables 1 and 2.)

critical role in the interface of all three complexes^{7,8,10} described in the present paper. In free solution, Phe48 is rotamer averaged.³³ In the crystal structure of free HPr,^{18a} the χ_1 angle of Phe48 is in the g^- rotamer. This rotamer is preserved in the EIN-HPr complex¹⁰ so that an approach based on ring current shifts could be successfully applied (although details of selection criteria were not provided, so it is difficult to ascertain the robustness of the method). However, in the case of the IIA^{Glc}-HPr and IIA^{Mtl}-HPr complexes,^{7,8} the χ_1 angle of Phe48 adopts a trans conformation which would completely preclude any attempt at correct ring current shift predictions based upon a g^- rotamer. Such conformational plasticity of surface side chains is a very common feature of protein–protein interactions.^{7,8,10} In the case of the present approach, however, the exact placement of side chains, including those with large rigid groups, such as aromatic rings, is not at all critical. Thus, while the torsion angle database potential biases the side chain conformations toward physically allowed rotamers,²⁴ all three χ_1 rotamers of Phe48 are in fact populated (albeit with a predominance of the trans rotamer) in the converged structures for all three complexes. This being said, ring current effects could readily be incorporated in further refinement³⁴ and could potentially increase the coordinate accuracy of some interfacial side chains.

Side Chains in the Docked Complexes. The experimental information used to dock protein–protein complexes in the present work relates principally to the backbone in the form of $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift perturbations and $^1\text{D}_\text{NH}$ dipolar couplings. Although the ambiguous intermolecular distance restraints employ all hydrogen, oxygen, and nitrogen atoms of a residue, and hence include many side chain atoms, this information is insufficient to uniquely define side chain conformations. Thus, the conformational space sampled by the interfacial side chains is to a large extent influenced by the torsion angle database potential of mean force²⁴ (which comprises two-, three-, and four-dimensional correlations relating backbone ϕ, ψ and side chain torsion angles) and is further limited by the van der Waals

repulsion term which prevents both intermolecular and intramolecular atomic overlap of side chain atoms. Nevertheless, there are many side chain rotamer combinations that are compatible with the relative orientation and translation of the proteins determined from the ambiguous intermolecular distance restraints and $^1\text{D}_\text{NH}$ dipolar couplings. Indeed, as discussed in the section above, this is highly advantageous because accurate side chain conformations are not at all required to obtain correct docking using the present procedure.

While the protein–protein complexes docked using the present method therefore do not permit a detailed analysis of the geometry of intermolecular side chain interactions, they are still more than sufficient to ascertain the correct identity of pairwise intermolecular side chain interactions. This is illustrated in Figure 4 with regard to the IIA^{Glc}-HPr complex. Arg17 of HPr is critical for phosphoryl transfer to IIA^{Glc};³⁵ the role of Arg17 is to neutralize the negatively charged carboxylate groups of Asp38 and Asp94 of IIA^{Glc} close to the active site by forming bifurcated salt bridges.⁷ Figure 4 displays a reweighted atomic probability density map of Arg17, Asp38, and Asp94 representing the ensemble of 143 converged structures with no violations >0.5 Å in the ambiguous intermolecular distance restraints and $R_{\text{dip}} \leq R_{\text{dip}}^{\text{median}}$. It is evident from the probability map that Arg17 does indeed interact with Asp38 and Asp94. In addition, it is readily possible to fit allowed side chain rotamer combinations within the map that permit good salt bridges to be formed.

Concluding Remarks

In this paper, we have provided a simple method for reliably docking protein–protein complexes on the basis of easily measured $^1\text{D}_\text{NH}$ dipolar couplings and highly ambiguous intermolecular distance restraints derived from $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping (cf. eq 1), combined with a powerful simulated annealing rigid body/torsion angle dynamics protocol. While the interaction surfaces in the present study were derived from $^1\text{H}_\text{N}/^{15}\text{N}$ chemical shift mapping, a number of other simple NMR and biochemical methods could also be employed. These include

(33) van Nuland, N. A. J.; Boelens, R.; Scheek, R. M.; Robillard, G. T. *J. Mol. Biol.* **1995**, *246*, 180–193.

(34) (a) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1995**, *Ser. B* *107*, 293–297. (b) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1996**, *112*, 79–81.

(35) (a) Anderson, J. W.; Pullen, K.; Georges, F.; Kleiv, R. E.; Waygood, E. B. *J. Biol. Chem.* **1993**, *268*, 12323–12333. (b) Kruse, R.; Hengstenber, W.; Benecke, W.; Kalbitzer, H. R. *Protein Eng.* **1993**, *6*, 417–423.

NMR cross-relaxation measurements^{36,37} and protein modification by either site-directed mutagenesis or alanine scanning mutagenesis coupled with an appropriate binding or functional assay to assess the effect of the mutations.³⁸ The methodology presented here should provide a powerful tool in high throughput structural proteomics and, moreover, should greatly accelerate

- (36) (a) Takahashi, H.; Nakanishi, T.; Kami, K.; Arata, Y.; Shimada, I. *Nat. Struct. Biol.* **2000**, *7*, 220–223. (b) Nakanishi, T.; Miyazawa, M.; Sakakura, M.; Terasawa, H.; Takahashi, H.; Shimada, I. *J. Mol. Biol.* **2002**, *318*, 245–249.
- (37) The delineation of interaction surfaces by cross-saturation³³ is potentially more precise than that obtained by $^1\text{H}/^{15}\text{N}$ chemical shift mapping, because cross-saturation is directly dependent on the close spatial proximity (≤ 7 Å) between the backbone amides of one partner and the aliphatic protons of the other. However, experimentally, cross-saturation necessitates the preparation of two highly deuterated ($>98\%$) ^{15}N -labeled samples in which nonexchangeable protons of the ^{15}N -labeled partner in the complex are replaced by deuterons (i.e., $\text{U}[^{15}\text{N}/^2\text{H}]$ -protein A + $\text{U}[^{14}\text{N}/^1\text{H}]$ -protein B, and vice versa). Such samples are expensive to make because, in addition to growing the bacteria in D_2O , d_7 -glucose must also be employed to ensure a very high level of deuteration. Given the nature of the cross-saturation experiment, an ambiguous distance restraint d_{aB} would comprise distances from the amide proton of a mapped residue a on protein A to all of the protons of all of the mapped residues on protein B.

the determination of higher accuracy NMR structures of complexes (including the detailed placement of interfacial side chains) by providing a good starting point for the assignment of intermolecular NOE data.

Acknowledgment. This study utilized the high performance computational capabilities of the Biowulf/LoBoS3 cluster at the National Institutes of Health. This work was in part supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

Note Added after ASAP Publication: The version published on the Web 2/15/2003 contained errors in the ASA concentrations in Table 4. The final Web version published 2/20/2003 and the print version are correct.

JA028893D

- (38) Jin, L.; Wells, J. A. In *Structure of Antigens*; Van Regenmortel, M. H. V., Ed.; CRC Press: New York, 1996; Vol. 3, pp 21–36. (b) DeLano, W. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20.