Docking of Protein–Protein Complexes on the Basis of Highly Ambiguous Intermolecular Distance Restraints Derived from $^{1}H_{N}/^{15}N$ Chemical Shift Mapping and Backbone $^{15}N–^{1}H$ 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 $^{1}H_{N}/^{15}N$ chemical shift mapping and backbone $^{15}N–^{1}H$ residual dipolar couplings is presented and illustrated with three complexes (EIN-HPr, IIA$^{\text{Glc}}$-HPr, and IIA$^{\text{Mtl}}$-HPr) of known structure. The $^{1}H_{N}/^{15}N$ 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 to prevent atomic overlap, a radius of gyration term ($E_{rgy}$) 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$^{\text{Glc}}$-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$^{\text{Mtl}}$-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 $^{1}H_{N}/^{15}N$ 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 $E_{rgy}$ 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, ab initio docking from structures of free proteins is still in its infancy and remains problematic. 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. 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 to solve the structure of the complex based solely on intermolecular interproton distance restraints, derived from isotope-
filtered/edited nuclear Overhauser enhancement (NOE) measurements, and orientational restraints from residual dipolar couplings. Nevertheless, unambiguous assignment of intermolecular NOEs is still difficult and time-consuming, particularly for larger complexes. In contrast, mapping of interaction surfaces by \( \text{H}_2\text{N}/\text{N} \) chemical shift perturbation is a simple and rapid procedure. Likewise, measurement of backbone \( ^{15}\text{N}/\text{H} \) residual dipolar couplings (\( ^{1}\text{D}_{\text{iso}} \)) is entirely straightforward and fast. Here, we demonstrate with three examples of protein–protein complexes previously solved in our laboratory by NMR that it is possible to reliably dock such complexes based on highly ambiguous intermolecular distance restraints derived from \( ^{1}\text{H}/^{15}\text{N} \) chemical shift mapping, in conjunction with backbone \( ^{1}\text{HN} \) dipolar couplings, using conjoined rigid body/torsion angle dynamics.

Methods

\( ^{1}\text{D}_{\text{iso}} \) dipolar coupling data were taken from the papers describing the structure determinations of the EIN-HPr,\( ^{9} \) IIA\(^{\text{GH}}\)-HPr,\(^{9} \) and IIA\(^{\text{HH}}\)-HPr\(^{4} \) complexes. \( ^{1}\text{D}_{\text{iso}} \) dipolar couplings for EIN-HPr (126 for EIN, 75 for HPr), IIA\(^{\text{GH}}\)-HPr (118 for IIA\(^{\text{GH}}\), 75 for HPr), and IIA\(^{\text{HH}}\)-HPr (114 for IIA\(^{\text{HH}}\), 71 for HPr) were measured in liquid crystalline media of phage fd\(^{11} \) (negatively charged rod), tobacco mosaic virus\(^{11} \) (negatively charged rod), and poly(ethylene glycol)/hexanol\(^{11} \) (neutral), respectively. \( ^{1}\text{H}/^{15}\text{N} \) chemical shift mapping data for EIN-HPr were taken from ref 12, while those for the IIA\(^{\text{GH}}\)-HPr and IIA\(^{\text{HH}}\)-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.\(^3 \) 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 GRASP\(^15 \) and VMD-XPLOR.\(^16 \) The ensemble distributions of the docked structures are depicted by isosurfaces of the reweighted atomic density maps\(^7 \) 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 1POH),\(^18 \) EIN (PDB code 1ZYM),\(^18 \) IIA\(^{\text{GH}} \) (PDB code 2F3G, molecule 2),\(^16 \) and IIA\(^{\text{HH}} \) (PDB code 1A3A, molecule D).\(^18 \) The experimental structures of the EIN-HPr (PDB code 3EZE),\(^10 \) IIA\(^{\text{GH}} \).

Effects of chemical shift perturbation in protein complexes

The first step in our procedure is to convert the \( ^{1}\text{H}/^{15}\text{N} \) chemical shift perturbation upon complex formation, (b) at least one or two atoms of the 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 \( ^{1}\text{H}/^{15}\text{N} \) chemical shift perturbation maps into a set of highly ambiguous

Results and Discussion

Converting \( ^{1}\text{H}/^{15}\text{N} \) Chemical Shift Maps into Highly Ambiguous Distance Restraints. Backbone \( ^{1}\text{H} \) and \( ^{15}\text{N} \) chemical shifts are highly sensitive to environment and have been extensively used to map interaction surfaces on proteins. Perturbations in backbone \( ^{1}\text{H} \) and \( ^{15}\text{N} \) chemical shifts are mainly influenced by electronic effects and, in the case of \( ^{1}\text{H} \) 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 \( ^{1}\text{H}/^{15}\text{N} \) 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 \( ^{1}\text{H}/^{15}\text{N} \) chemical shift perturbation maps into a set of highly ambiguous
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:

$$d_{ab} = \left( \sum_b \sum_j r_{ai,bj}^{-6} \right)^{-1/6} \quad \text{and} \quad d_{BA} = \left( \sum_a \sum_j r_{ai,bj}^{-6} \right)^{-1/6}$$

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.

### 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 distance restraints, and a harmonic potential for the $1^\text{DNH}$ 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,23 a radius of gyration term ($E_{rgyr}$),24 and a side chain torsion angle database potential of mean force ($E_{gb}$) 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, $R_{gyr \text{ target}}$, is given by $2.2N^{0.38}$ where $N$ is the number of residues in the calculated complex.21 The calculated value of $R_{gyr \text{ target}}$ tends to underestimate the true value of $R_{gyr}$ ($R_{gyr \text{ true}}$). The exact value of $R_{gyr \text{ target}}$, in the context of the present calculations, however, is not critical providing $R_{gyr \text{ target}} \leq R_{gyr \text{ true}}$ 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 $R_{gyr \text{ target}}$ values of 20 and 22 Å. However, if $R_{gyr \text{ target}}$ were significantly larger than $R_{gyr \text{ true}}$, the $E_{rgyr}$ potential would allow expansion to occur. The $R_{gyr \text{ target}}$ 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$^\text{Glc}$-HPr complex (residues 19–168 of IIA$^\text{Glc}$ and 1–85 of HPr), and 17.4 Å for the IIA$^\text{Mil}$-HPr complex (residues 4–147 of IIA$^\text{Mil}$ + 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 $k_{gb}$, 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 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 $k_{dip}$ 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 $1^\text{Hs}^{15}\text{N}$ chemical shift mapping) given their full torsional degrees of freedom: all five terms of


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 14 fs from 1500 to 25 K in increments of 25 K, and $k_{\text{dip}}, k_{\text{vdw}}, k_{\text{egyr}},$ and $k_{\text{db}}$ 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 $s_{\text{vdw}}$ is decreased from 0.78 to 0.75. (iv) Conjoined rigid body/torsion angle minimization with the force constants unchanged except for $k_{\text{vdw}} = 3$ kcal mol$^{-1}$ Å$^{-2}$ 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$^{\text{Glc}}$-HPr, and IIA$^{\text{Mtl}}$-HPr Complexes. Figure 1 summarizes the results obtained for the
Calculations for the EIN-HPr complex were also carried out using a 40 kDa EIN-HPr (left-hand panels), the 30 kDa IIA Glc -HPr complexes. Although enzymes EIN, IIA Glc , and IIA Mtl interact with the EIN-HPr complex, the overall shift perturbations were too small to define a representative binding surface, and consequently the results (backbone mean coordinate accuracy of 0.5 Å) of the converged structures are 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 (ΔH/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 1 H/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 complex reveals that d_{arg131,hpr} is less than 5 Å and d_{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 proteins 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 ≥ 5%, 15%, 25%, 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_{dip} ≤ R_{dip,median}) were retained. The choice of the median as a
Table 2. Statistics of Structural Convergence and Selection

<table>
<thead>
<tr>
<th></th>
<th>EIN-HPr</th>
<th>IIAGlc-HPr</th>
<th>IIAMtl-HPr</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Number of Structures with No Distance Violations &gt; 0.5 Å</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 5% ASA</td>
<td>251 (83.7%)</td>
<td>286 (95.3%)</td>
<td>300 (100%)</td>
</tr>
<tr>
<td>50% ASA</td>
<td>230 (76.7%)</td>
<td>296 (98.7%)</td>
<td>300 (100%)</td>
</tr>
</tbody>
</table>

| (2) Number of Structures from (1) with R_dip ≤ R_dip\(\text{median}\) |         |            |            |
| ≥ 5% ASA      | 126 (42.0%) | 143 (47.7%) | 150 (50%)  |
| 50% ASA       | 115 (38.3%) | 148 (49.3%) | 150 (50%)  |

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 percent of retained structures (out of the total of 300 calculated) is given in parentheses. Although selection was based on a distance cutoff of 0.5 Å, in none of the converged structures exhibited distance violations > 0.1 Å. a 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. b 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. c d e

Table 3. Backbone Ensemble Precision and Coordinate Accuracy for Converged Structures with No Distance Violations > 0.5 Å and R_dip ≤ R_dip\(\text{median}\)

<table>
<thead>
<tr>
<th></th>
<th>EIN-HPr</th>
<th>IIAGlc-HPr</th>
<th>IIAMtl-HPr</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_dip) (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cluster 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean coordinate accuracy (Å)</td>
<td>0.71</td>
<td>1.47</td>
<td>0.52</td>
</tr>
<tr>
<td>precision (Å)</td>
<td>1.26 ± 0.48</td>
<td>1.70 ± 0.46</td>
<td>0.96 ± 0.46</td>
</tr>
<tr>
<td>cluster 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean coordinate accuracy (Å)</td>
<td>1.06</td>
<td>1.40</td>
<td>0.41</td>
</tr>
<tr>
<td>precision (Å)</td>
<td>1.52 ± 0.48</td>
<td>1.54 ± 0.52</td>
<td>0.95 ± 0.62</td>
</tr>
</tbody>
</table>

* 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. a 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. b The definition of the dipolar coupling R-factor is given in ref 26b.


The backbone atomic rms difference between the ensembles of the two correct solutions is given by the following expression:

\[
\langle R_{\text{dip}} \rangle = \frac{\langle R_{\text{dip}}^2 \rangle}{\langle R_{\text{dip}}^2 \rangle + \langle \text{J}^2 \rangle^{1/2}}
\]

where \(R_{\text{dip}}\) is the backbone atomic rms difference between the mean coordinates of the ensemble of the experimentally determined structures derived from the full complement of intermolecular NOE and dipolar coupling data, and \(\text{J}^2\) is the quantum mechanical coupling constant.
In unfavorable cases, such as the IIAMtl-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 IIAMtl-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/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 IIAMtl-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/15N chemical shift perturbation upon complexation, the cluster 2 solution is not fully consistent with the observed 1H/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 ΔH/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 dIPR ≥ 0.5 Å between each residue on protein A with a minimal ΔH/N to all residues with significant ΔH/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 RIPR ≤ RIPR 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. For the IIAMtl-HPr complex, the 1DNH dipolar couplings were measured in a neutral poly(ethylene glycol)/hexanol liquid crystalline medium. The experimental values of the axial component of the alignment tensor (D2) and the rhombicity (η) are 12.1 Hz and 0.42, respectively, and the average value of RIPR for all 150 structures with RIPR ≤ RIPR median (i.e., clusters 1 and 2 combined) is 20.2 ± 0.5%. The predicted values of D2 and η using the
Figure 3. Results of docking calculations for the IIA\textsuperscript{ML}-HPr complex using both attractive \((d_{ab} \leq 5\ \text{Å})\) and repulsive \((d_{ab} \geq 5\ \text{Å})\) ambiguous intermolecular distance restraints. The repulsive restraints involve five surface exposed residues of HPr (Ser37, Asn38, Gly39, Gly58, and Thr59) that exhibit only minimal \(\Delta_{\text{sox}}\) \((5–20\ \text{Hz})\) 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 \(>0.5\ \text{Å}\) in the ambiguous intermolecular distance restraints and \(R_{\text{gyr}} \leq 4\ \text{Å}\) 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 \(>0.5\ \text{Å}\) and \(R_{\text{gyr}} \leq 4\ \text{Å}\) was 130 and 134, respectively. The top panels show plots of the dipolar coupling \(R\)-factor \(R_{\text{dip}}\) versus accuracy. The values of \(R_{\text{dip}}\), 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 Å, 1.18 \(\pm\) 0.43 Å, and 0.77 Å, respectively; the corresponding values for the structures in (b) are 20.2 \(\pm\) 0.7\%, 0.99 \(\pm\) 0.74 Å, 1.07 \(\pm\) 0.81 Å, and 0.46 Å, respectively. The middle panels show plots of \(E_{\text{rgyr}}\) versus accuracy, and the lower panels show histograms of the \(E_{\text{rgyr}}\) distribution. The \(E_{\text{rgyr}}\) distribution is unimodal but highly skewed with a mode at \(\sim 27\ \text{kcal mol}^{-1}\) in (a) and \(24\ \text{kcal mol}^{-1}\) in (b). The individual simulated annealing structures in the tails of the \(E_{\text{rgyr}}\) distribution \((E_{\text{rgyr}} \geq 55\ \text{kcal mol}^{-1})\) are of lower accuracy.


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. \(E_{\text{dw}}\) and \(E_{\text{db}}\), 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 \(R_{\text{dp}} \leq R_{\text{dp,median}}\) is not at all correlated with \(E_{\text{dw}}\) or \(E_{\text{db}}\). Consequently, we have found examination of the \(E_{\text{rgyr}}\) distribution to be useful.

Figure 1c displays histograms of \(E_{\text{rgyr}}\) for those structures with no ambiguous intermolecular distance violations >0.5 Å and \(R_{\text{dp}} \leq R_{\text{dp,median}}\). For the IIA\(^{\text{Glc-HPr}}\) complex, the \(E_{\text{rgyr}}\) 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\(^{\text{Mtl-HPr}}\) and EIN-HPr complexes, on the other hand, a clear-cut bimodal \(E_{\text{rgyr}}\) 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 \(E_{\text{rgyr}}\) population being more expanded (i.e., larger \(R_{\text{rgyr}}\)). The results obtained using ambiguous intermolecular distance restraints derived from either interfacial residues with ASA ≥5% or ≥50% in the free proteins are very similar with the exception that the occupancy of the lower \(E_{\text{rgyr}}\) population is slightly reduced for the latter (Figure 1c, left and right panels; Table 4).

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

The overall distribution of \(E_{\text{rgyr}}\) for the IIA\(^{\text{Mtl-HPr}}\) 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) \(E_{\text{rgyr}}\) structures within cluster 1 corresponds to the tail of the cluster 1 \(E_{\text{rgyr}}\) distribution. While the structures in the tail of the \(E_{\text{rgyr}}\) distribution are both less precise and less accurate, the accuracies of their mean backbone coordinates (≈0.7 Å) are only slightly worse than those of either the structures in the main envelope of the \(E_{\text{rgyr}}\) distribution (≈0.5–0.6 Å) or of all of the cluster 1 structures combined (≈0.4–0.5 Å) (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 \(R_{\text{dp}}\) versus accuracy that there are two subpopulations of structures within this cluster with a boundary of ≈1.5 Å in backbone coordinate accuracy. These two subpopulations reflect two distinct populations in the \(E_{\text{rgyr}}\) distribution (Figure 1c, left), with the lower energy \(E_{\text{rgyr}}\) 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 \(E_{\text{rgyr}}\); the \(\text{C}α\)–\(\text{C}α\) separation between the two active site histidines (His189 of EIN and His15 of HPr) is 13.9 ± 0.5 Å for the lower energy \(E_{\text{rgyr}}\) population, but 15.5 ± 0.5 Å for the higher energy \(E_{\text{rgyr}}\) population. Thus, the \(\text{C}α\)–\(\text{C}α\) distance in the higher energy \(E_{\text{rgyr}}\) 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 \(E_{\text{rgyr}}\) 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 \(^1\text{H}\) chemical shift perturbations and those calculated from ring current shifts is used to locate the proteins, previously oriented by residual dipolar couplings, on a predefined 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

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

<table>
<thead>
<tr>
<th>EIN-HPr</th>
<th>IIA\text{Mtl-HPr} cluster 1b</th>
<th>lower (E_{\text{rgyr}})</th>
<th>higher (E_{\text{rgyr}})</th>
<th>lower (E_{\text{rgyr}})</th>
<th>higher (E_{\text{rgyr}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>occupancy (%)</td>
<td>57.6%</td>
<td>42.4%</td>
<td>77.0%</td>
<td>23.0%</td>
<td>0.21</td>
</tr>
<tr>
<td>((R_{\text{dp}})) (%)</td>
<td>25.3 ± 1.1</td>
<td>25.4 ± 2.0</td>
<td>19.9 ± 0.3</td>
<td>20.1 ± 0.3</td>
<td>5% or 50% ASA</td>
</tr>
<tr>
<td>(E_{\text{rgyr}}) range (kcal mol(^{-1}))</td>
<td>685–838</td>
<td>871–1065</td>
<td>6.3–26.2</td>
<td>28.1–98.9</td>
<td>0.50 ± 0.20</td>
</tr>
<tr>
<td>ensemble precision (Å) (\Delta)</td>
<td>0.72</td>
<td>1.62</td>
<td>0.64</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td>ensemble accuracy (Å) (\Delta)</td>
<td>0.72</td>
<td>1.62</td>
<td>0.64</td>
<td>0.71</td>
<td>0.76</td>
</tr>
</tbody>
</table>

\(\Delta\) The boundary between the lower and higher \(E_{\text{rgyr}}\) energy populations is at ~840 kcal mol\(^{-1}\) for the EIN-HPr complex (Figure 1c, left panel) and at ~27 kcal mol\(^{-1}\) for the IIA\text{Mtl-HPr} complex (Figure 1c, right panel). Note that the higher values of \(E_{\text{rgyr}}\) for the EIN-HPr complex are simply due to the fact that the value of 20 Å for \(R_{\text{rgyr}}\)\(^{\text{target}}\) calculated using the empirical relationship 2.2 Å \(N\) (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 \(R_{\text{rgyr}}\) of ~22.5 Å. Repeating the EIN-HPr calculations using a value of 22 Å yields essentially the same results in terms of coordinate precision and accuracy, except that \(E_{\text{rgyr}}\) spans a range from ~61 to ~250 kcal mol\(^{-1}\), and the boundary between the lower and higher energy \(E_{\text{rgyr}}\) populations is ~110 kcal mol\(^{-1}\). The structures for the IIA\text{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. 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). Backbone ensemble precision, ensemble accuracy, and mean coordinate accuracy are defined in footnote a of Table 3.

critical role in the interface of all three complexes described in the present paper. In free solution, Phe48 is rotamer averaged. In the crystal structure of free HPt, the χ₁ angle of Phe48 is in the g⁻ rotamer. This rotamer is preserved in the EIN-HPt complex 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 IIAGlct-HPt and IIAMo-HPt complexes, the χ₁ 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. 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, all three χ₁ 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 refinement and could potentially increase the coordinate accuracy of some interface 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 ¹H/¹5N chemical shift perturbations and ¹D 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 interface side chains is to a large extent influenced by the torsion angle database potential of mean force (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 ¹D 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 IIAGlct-HPt complex. Arg17 of HPt is critical for phosphor transfer to IIAGlct; the role of Arg17 is to neutralize the negatively charged carboxylate groups of Asp38 and Asp94 of IIAGlct close to the active site by forming bifurcated salt bridges. 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 $R_{ap} \leq R_{ap median}$. 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 ¹D NH dipolar couplings and highly ambiguous intermolecular distance restraints derived from ¹H/¹5N 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 ¹H/¹5N chemical shift mapping, a number of other simple NMR and biochemical methods could also be employed. These include

---


---

**Figure 4.** Stereoview illustrating the interaction of Arg17 of HPt with Asp38 and Asp94 of IIAGlct in the docked IIAGlct-HPt complex. The backbone of the mean coordinates is shown as tubes (purple for HPt, orange for IIAGlct). 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 $R_{ap} \leq R_{ap median}$; 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 HPt has been fitted into the atomic density map with the relative orientation and translation of the proteins determined from the ambiguous intermolecular distance restraints and ¹D 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 IIAGlct-HPt complex. Arg17 of HPt is critical for phosphoryl transfer to IIAGlct; the role of Arg17 is to neutralize the negatively charged carboxylate groups of Asp38 and Asp94 of IIAGlct close to the active site by forming bifurcated salt bridges. 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 $R_{ap} \leq R_{ap median}$. 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.
NMR cross-relaxation measurements 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. The methodology presented here should provide a powerful tool in high throughput structural proteomics and, moreover, should greatly accelerate 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


(37) The delineation of interaction surfaces by cross-saturation is potentially more precise than that obtained by $^1$H$^1$N chemical shift mapping, because cross-saturation is directly dependent on the close spatial proximity ($\leq 7$ 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., U[$^{15}$N/$^2$H]-protein A + U[$^{14}$N/$^1$H]-protein B, and vice versa). Such samples are expensive to make because, in addition to growing the bacteria in D$_2$O, $d_7$-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 $B$.