Direct Use of $^{15}$N Relaxation Rates as Experimental Restraints on Molecular Shape and Orientation for Docking of Protein–Protein Complexes

Yaroslav Ryabov,† G. Marius Clore,*‡ and Charles D. Schwieters*†

Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624, and, Laboratory of Chemical Physics, Building 5, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Received March 4, 2010; E-mail: mariusc@mail.nih.gov; charles.schwieters@nih.gov

Protein–protein interactions play a central role in a multitude of biological processes. While full structure determination of protein–protein complexes by crystallography or NMR spectroscopy is clearly desirable, it may not always be possible. In the case of NMR spectroscopy, it is also the case that the larger the system, the more difficult it is to acquire and analyze intermolecular nuclear Overhauser enhancement data to derive interproton distance restraints that are the mainstay of conventional NMR structure determination.1 While technical improvements have been made in ab initio protein docking, it is still not particularly reliable.2 For this reason, recent efforts have been focused on developing methods that can make use of sparse solution experimental data.3–4 For example, incorporation of highly ambiguous distance restraints derived from chemical shift perturbation mapping can facilitate protein docking,3,4 and reliability can be further improved by the addition of orientational restraints derived from residual dipolar couplings (RDCs).4

Supplementary structural information also resides in $^{15}$N transverse ($R_2$) and longitudinal ($R_1$) relaxation rates. In the absence of large scale internal motions or conformational exchange line broadening, $^{15}$N $R_2/R_1$ ratios depend on only two factors: the rotational diffusion tensor which is related to the size and shape of the complex5 and the orientation of $N$–H bond vectors relative to the axes of the diffusion tensor.6–8 In initial work,9 $^{15}$N $R_2/R_1$ ratios were employed in a manner analogous to RDCs9 to provide $N$–H bond orientational restraints with predetermined, approximate values of the magnitude of the diffusion tensor estimated directly from the distribution of measured $R_2/R_1$ ratios.9 Recently, fast methods for computing the rotational diffusion tensor based on molecular structure11 have permitted direct refinement against the components of the rotational diffusion tensor,12 thereby incorporating information on molecular shape and size from relaxation measurements that is not dissimilar to that afforded by solution small-angle X-ray scattering. Here we describe a new approach in which $^{15}$N $R_2/R_1$ relaxation data are used to simultaneously provide information on both molecular shape and $N$–H bond orientations without requiring prior information to be extracted from either the data or a known molecular structure. The application to protein–protein docking by conjoined rigid body/torsion angle simulated annealing13 is illustrated using the complex of the N-terminal domain of enzyme I (EIN) and HPr,12,14 and the HIV-1 protease dimer7,15 as examples.

Conceptually, the elements of the rotational diffusion tensor are calculated from the atomic coordinates at each step of the simulated annealing protocol and then used together with the $N$–H bond vector orientations to compute the $^{15}$N $R_2/R_1$ ratios. The pseudo-potential energy term, $E_{relax}$, that is minimized in the Xplor-NIH (v2.25) structure determination package16 is given by

$$E_{relax} = k_{relax} \sum_{i=1}^{N} (\rho_i^{calc} - \rho_i^{obs})^2 / \sigma_i^2$$

where $\rho_i^{calc}$ and $\rho_i^{obs}$ are the calculated and observed $R_2/R_1$ ratios, respectively, for residue $i$; $\sigma_i$ is the corresponding experimental error in the measurements; $N$ is the number of data points; and $k_{relax}$ is a force constant. $\rho_i^{calc}$ is obtained from the spectral density function, $J_\omega(\omega)$ (ignoring the very small contribution from $^{15}$N chemical shift anisotropy).17

$$\rho_i^{calc} = \frac{4J(0) + 6J(\omega_H - \omega_N) + J(\omega_H + \omega_N) + 6J(\omega_H) + 3J(2\omega_H)}{2[6J(\omega_H - \omega_N) + J(\omega_H + \omega_N) + 3J(2\omega_H)]]}$$

where $\omega_H$ and $\omega_N$ are the $^1$H and $^{15}$N nuclear Lamor frequencies. $J_\omega(\omega)$ is expressed as

$$J_\omega(\omega) = \frac{2}{5} \sum_{r=2}^{3} \frac{E_r}{E_r^2 + \omega^2} F_r([D_x, D_y, D_z], \Omega_{D-A})$$

where the values of $E_r$ are frequencies depending on the three eigenvalues, $D_x$, $D_y$, and $D_z$, of the rotational diffusion tensor $D$; and the function $F_r$ depends on the same set of eigenvalues and on the orientation of the $N$–H bond of residue $A_i$ relative to the principal axis of $D$ (see Supporting Information for details). $D$ is calculated from the atomic coordinates by representing the surface of the protein by an equivalent ellipsoid and then applying Perrin’s equations18 to calculate $D$ from the dimensions and orientation of the ellipsoid.12 The gradients of $E_{relax}$ with respect to all atomic displacements are evaluated in closed form and used to calculate the atomic forces during gradient minimization and molecular dynamics. In addition to molecular size and shape, $D$ also depends on the solvent viscosity and temperature. To account for uncertainties in the latter, the apparent diffusion tensor temperature (which is not a physical temperature but a fitting parameter) is also optimized during minimization and simulated annealing.12

In the context of docking, highly ambiguous distance restraints ($E_{CSHgap}$) derived from chemical shift perturbation mapping serve to delineate the interaction surface,3,4 while the incorporation of the $E_{relax}$ potential provides simultaneous restraints on the molecular shape and size of the complex and on the relative orientation of the proteins within the complex. The other terms included in the target function are covalent geometry terms, a knowledge-based low-resolution hydrophobic contact potential,12 a multidimensional torsion angle database potential of mean force,19 and a quartic van der Waals repulsion term ($E_{vdW}$).20 Briefly, the protocol consists of initial rigid body minimization starting from randomized positions

---

1 Center for Information Technology.
2 National Institutes of Diabetes and Digestive and Kidney Diseases.
and orientations of the two proteins subject to only $E_{\text{CSMap}}$, $E_{\text{relax}}$, and $E_{\text{repel}}$ (applied to only the C$\alpha$ atoms at this initial stage). This is followed by conjoined rigid body/torsion angle dynamics simulated annealing in which the backbone and internal side chains of each protein are treated as rigid bodies with translational and rotational degrees of freedom and the surface side chains are given full torsional degrees of freedom. The force constants for the various terms in the target function are progressively ramped up as the temperature is decreased from 1000 to 10 K. A full description is provided in Supporting Information.

The $^{15}$N $R_2/R_1$ ratios are subject to uncertainties arising from both errors in the protein backbone coordinates (that are treated as rigid bodies) and the presence of significant local motions (either on the sub-nanosecond or $\mu$s-ns time scales which decrease and increase the $R_2/R_1$ ratios, respectively). Since EIN and hence the EIN–HPr complex are highly anisotropic, anomalous $^{15}$N $R_2/R_1$ data were identified in an iterative fashion; the experimental $R_2/R_1$ data were fit to the X-ray coordinates of free EIN and HPr separately by optimizing the respective diffusion tensors (while restraining the anisotropy and rhombicity to be the same) and excluding from further analysis data for those residues with relative deviations larger than 1.5σ between observed and calculated $^{15}$N $R_2/R_1$ ratios (this corresponds to excluding 13% of the data; see Supporting Information). The starting coordinates were the X-ray structures of free EIN and HPr with the torsion angles of the surface side chains partially randomized by a 1 ns run of torsion angle dynamics at 3000 K with the backbone fixed. The reference structure was generated by best-fitting the backbone of the X-ray coordinates of the free structures onto the NMR structure of the complex, so that the calculated atomic rms differences reflect only differences in the positions of the two proteins relative to one another in the complex. A total of 512 structures were calculated.

Figure 1 compares the dependence of the total energy on the C$\alpha$ rms difference between the calculated structures, without (Figure 1A and B) and with (Figure 1C and D) the inclusion of the $E_{\text{relax}}$

![Figure 1. Docking of the EIN–HPr complex using $^{15}$N $R_2/R_1$ relaxation data. Dependence of the total energy on the C$\alpha$ rms difference from the minimum energy docked structures (left panels) and the reference structure (right panels) for calculations based on $E_{\text{CSMap}}$ alone (A and B) and in combination with $E_{\text{relax}}$ (C and D). Structures located in the cluster that includes the minimum energy structure are indicated by the red filled-in circles. The inset in the lower right panel shows a comparison of the location of HPr in the restrained regularized mean docked structure derived from the lowest 10 energy structures calculated using both $E_{\text{CSMap}}$ and $E_{\text{relax}}$ (red) with the position of HPr in the reference structure (green); the backbone of EIN is shown in blue. $T_{\text{rms}}^{\text{EIN}}$ was optimized independently for EIN and HPr within a range of 313 ± 10 K, and the nominal experimental temperature was 313 K.](image1)

---

### Table 1. Impact of $E_{\text{relax}}$ on Docking Accuracy of the EIN/HPr Complex

<table>
<thead>
<tr>
<th></th>
<th>$E_{\text{CSMap}}$</th>
<th>$E_{\text{relax}} + E_{\text{CSMap}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>1.8 ± 2.7</td>
<td>0.3 ± 0.2</td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.2 ± 2.3</td>
<td>1.6 ± 0.1</td>
</tr>
<tr>
<td>Mean coord.</td>
<td>2.4</td>
<td>1.6</td>
</tr>
<tr>
<td>$\langle</td>
<td>\rho_{\text{calc}}^{\text{EIN}} - \rho_{\text{ref}}^{\text{EIN}}</td>
<td>\rangle$</td>
</tr>
</tbody>
</table>

$^a$ C$\alpha$ rmsd from the mean coordinates averaged over the 10 lowest energy structures. $^b$ C$\alpha$ rmsd from reference coordinates averaged over the 10 lowest energy structures. $^c$ C$\alpha$ rmsd between restrained regularized mean and reference structure. $^d$ This is the rms deviation averaged over all $R_2/R_1$ experimental data. The value for the reference structure is 2.4, and the values of $\rho$ span from $\sim$20 to $\sim$40.

potential, and the minimum energy (left panels) or reference (right panels) structure. Cluster analysis relative to the respective minimum energy structures shows that 40% of the structures lie within the cluster containing the minimum energy structure when both $E_{\text{CSMap}}$ and $E_{\text{relax}}$ potentials are used compared to 36% when $E_{\text{CSMap}}$ alone is used. The inclusion of $E_{\text{relax}}$, however, results in better discrimination (cf. compare Figure 1A with Figure 1C, and Figure 1B with Figure 1D): the 10 lowest energy structures have a backbone accuracy of $\sim$1.6 Å, and the data clearly show that the introduction of $E_{\text{relax}}$ results in an approximately 2-fold improvement in both backbone accuracy and precision, compared to structures calculated with only the $E_{\text{CSMap}}$ term (Table 1).

We also carried out calculations using only the $E_{\text{relax}}$ term in the absence of the $E_{\text{CSMap}}$ restraints. For the EIN/HPr complex docking with only the $E_{\text{relax}}$ term generated several clusters of solutions due to degeneracy in the orientations of the principal axis frame of the diffusion tensors (Figure 2). Three of the clusters (denoted as A, B, and C and comprising 5, 16, and 1% of the calculated structures, respectively) have approximately the same minimum energy which, most probably, reflects the fact that for those clusters the nearly spherical shape of HPr does not provide sufficient differentiation between the alternate HPr orientations and locations that satisfy the orientational restraints from the $E_{\text{relax}}$ term. However, the clusters

![Figure 2. Docking of the EIN/HPr complex based only on the $E_{\text{relax}}$ potential for the $^{15}$N $R_2/R_1$ relaxation data. (A) Dependence of the total energy on the C$\alpha$ rms difference from the minimum energy docked structure (left panel) and the reference structure (right panel). There are three main clusters, A, B, and C, denoted by red, green, and blue circles. (B) Comparison of the location of HPr in the restrained regularized mean docked EIN/HPr complexes (red) with that in the reference structure (green). The backbone of EIN is shown in blue; the active site histidines are depicted as space-filling models in the same color as the corresponding backbone.](image2)
can readily be differentiated and the correct solution ascertained by rescoring the results based on the number of $E_{\text{CSMap}}$ distance violations or using other experimental information (Table 2). For example, in the case of the EIN/HPt complex, the distance between the Cα atoms of His-189 of EIN and His-15 of HPt must be within 11–16 Å to permit phosphoryl transfer to occur between EIN and HPt. On the basis of either criterion, cluster B, which is ∼1.7 Å from the reference structure, can be unambiguously identified as the correct solution. This is further confirmed by validation against backbone amide RDCs which provide independent orientational information.

Inclusion of RDCs in the docking calculations may increase discrimination and convergence by reducing orientational degeneracy. Thus when $E_{\text{relax}}$ is supplemented by an RDC potential term for backbone amide RDCs (measured in a charged liquid crystalline medium), but omitting the $E_{\text{CSMap}}$ term, the cluster containing the minimum energy structure is clearly distinct from other clusters and corresponds to the correct solution with a mean coordinate accuracy of 1.5 Å for the 10 lowest energy structures.

**Figure 3.** Docking of the symmetric HIV-1 protease dimer using the $E_{\text{relax}}$ potential for the $^{13}$N R/R relaxation data. Dependence of the total energy on the Cα rms difference from the minimum energy docked structure (left) and the symmetrized reference X-ray structure (right). The inset in the right panel shows a backbone superposition (displayed as tubes) of the restrained regularized mean docked structure (red) and the crystal structure (blue). $T_{2}^{\text{app}}$ was optimized within the range 300 ± 5 K, and the nominal experimental temperature was 300 K.

Similar docking calculations were also carried out for the symmetric HIV-1 protease dimer, using the $E_{\text{relax}}$ potential but, in addition, included two symmetry restraints, one to confer C2 and the other to ensure that the coordinates of the surface side chains remain identical for both subunits. The 10 lowest energy structures converge to the correct solution with a backbone coordinate accuracy of ∼0.3 Å relative to the X-ray coordinates (Figure 3).

In conclusion, we have shown that the shape and orientational information afforded by directly using $^{13}$N R/R relaxation data in structure calculations without the need for any prior assumptions, combined with minimal additional information in the form of highly ambiguous distance restraints derived from chemical shift perturbation mapping, additional biochemical data, or, in the case of dimers, symmetry restraints, provides a powerful tool to facilitate reliable docking of protein–protein complexes.

**Acknowledgment.** Y.R. acknowledges a National Research Council Associateship. This work was supported by the NIH Intramural Research Programs of NIDDK (G.M.C.) and CIT (C.D.S.).

**Supporting Information Available:** Details of the algorithm, docking protocol, data treatment, Xplor-NIH scripts, and experimental restraints files. This material is available free of charge via the Internet at http://pubs.acs.org.