J|A|C|S
ARTICLES

Using the Experimentally Determined Components of the Overall Rotational Diffusion Tensor To Restrain Molecular Shape and Size in NMR Structure Determination of Globular Proteins and Protein−Protein Complexes

Yaroslav Ryabov,[†] Jeong-Yong Suh,[‡] Alexander Grishaev,[‡] 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 24, 2009; E-mail: mariusc@mail.nih.gov; charles.schwieters@nih.gov

Abstract: This paper describes an approach for making use of the components of the experimentally determined rotational diffusion tensor derived from NMR relaxation measurements in macomolecular structure determination. The parameters of the rotational diffusion tensor describe the shape and size of the macromolecule or macromolecular complex, and are therefore complementary to traditional NMR restraints. The structural information contained in the rotational diffusion tensor is not dissimilar to that present in the small-angle region of solution X-ray scattering profiles. We demonstrate the utility of rotational diffusion tensor restraints for protein structure refinement using the N-terminal domain of enzyme I (EIN) as an example and validate the results by solution small-angle X-ray scattering. We also show how rotational diffusion tensor restraints can be used for docking complexes using the dimeric HIV-1 protease and the EIN-HPr complexes as examples. In the former case, the rotational diffusion tensor restraints are sufficient in their own right to determine the position of one subunit relative to another. In the latter case, rotational diffusion tensor restraints complemented by highly ambiguous distance restraints derived from chemical shift perturbation mapping and a hydrophobic contact potential are sufficient to correctly dock EIN to HPr. In each case, the cluster containing the lowest-energy structure corresponds to the correct solution.

Introduction

NMR structure determination is primarily based upon short-range interactions in the form of nuclear Overhauser enhancement (NOE)-derived short (<6 Å) interproton distance restraints, supplemented by torsion angle restraints derived from ^3J couplings and chemical shift restraints.[1−3] Residual dipolar couplings measured in weakly aligning media can be used to obtain orientational information,[4,5] and paramagnetic NMR in the form of either paramagnetic relaxation enhancement or pseudocontact shifts can be used to derive long-range (15−35 Å) distance restraints between protons and a paramagnetic center.[6,7] The target function that is minimized includes not only terms describing the experimental restraints but also terms defining the covalent geometry, stereochemistry and nonbonded interactions.[3] The latter can range from a simple hard-sphere van der Waals repulsion term to prevent atomic overlap[8] to a full-scale empirical energy function including a Lennard-Jones potential, electrostatic contributions, and explicit solvent molecules.[9,10] These terms can also be supplemented by various knowledge-based potentials of mean force, including database-derived multidimensional torsion angle[11,12] and hydrogen bonding[13] potentials, as well as more global packing potentials describing, for example, the radius of gyration[14] or gyration volume.[15] As the systems studied become larger and more complex, the relative number of restraints that can be extracted

† Center for Information Technology.
‡ National Institutes of Diabetes and Digestive and Kidney Diseases.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(2) Clore, G. M.; Gronenborn, A. M. *Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 479–564.
(3) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891–5898.
(4) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127–174.
(5) Prestegard, J. H.; al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371–424.
(6) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 5879–5896.
(7) Pintacuda, G.; John, M.; Su, X. C.; Otting, G. *Acc. Chem. Res.* **2007**, *40*, 206–212.
(8) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(10) Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins: Struct., Funct. Genet.* **2003**, *50*, 496–506.
(11) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *Protein Sci.* **1996**, *5*, 1067–1080.
(12) Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc.* **2002**, *124*, 2866–2867.
(13) Grishaev, A.; Bax, A. *J. Am. Chem. Soc.* **2004**, *126*, 7281–7292.
(14) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337–2338.
(15) Schwieters, C. D.; Clore, G. M. *J. Phys. Chem. B* **2008**, *112*, 6070–6073.

 This article not subject to U.S. Copyright. Published XXXX by the American Chemical Society

from NOE spectra tends to become sparser on account of extensive spectral overlap and line broadening. Under these circumstances, additional information on the overall shape and size of the system under consideration can become invaluable.

Low-resolution information on the overall molecular shape of macromolecules and their complexes is directly encoded in small-angle X-ray scattering (SAXS) profiles,[16,17] and indeed direct incorporation of SAXS data into NMR structure calculations has proved extremely useful.[18−22] An alternative source of molecular shape information can be obtained by using the components of the overall rotational diffusion tensor as global shape restraints. The principal values of the rotational diffusion tensor describe rates of anisotropic tumbling of a protein or protein complex in solution and depend upon protein size and shape: larger proteins tumble more slowly than smaller ones, while asymmetrically shaped proteins tumble around the different rotation axes with different rates.[23−26] The components of the rotational diffusion tensor are readily accessible from ^{15}N R_1/R_2 ratios obtained from ^{15}N relaxation measurements, in conjunction with an available structure.[27−29] The latter can be a preliminary NMR structure obtained by conventional means or the known structures (either NMR or X-ray) of the free proteins in the context of a structure determination of a protein−protein complex.

The recent development of a fast and reliable method for calculation of the rotational diffusion tensor of a protein[30] makes it possible to incorporate the information encoded in the diffusion tensor directly into NMR structure calculations. This information can be used either for refining structures of globular proteins or for assembling structures of multidomain proteins and protein−protein complexes. Initial work using this approach described only a search algorithm for protein domain positioning with translational degrees of freedom when the domain orientations were already known from independent measurements.[31]

In this paper we present the implementation of a pseudopotential term to minimize the difference between calculated and target values of the components of the rotational diffusion tensor within the Xplor-NIH NMR structure determination package,[8,32] and we describe simulated annealing protocols for protein structure refinement and for docking of complexes where the structures of the individual components of the complex are known. We illustrate the performance of the rotational diffusion tensor potential with respect to refinement of the structure of the N-terminal domain of enzyme I (EIN) in conjunction with NOE data, to the docking of the subunits of the HIV-1 protease dimer in the absence of any additional experimental data, and to the docking of EIN and the histidine phosphocarrier protein HPr in conjunction with NMR chemical shift perturbation data.

Theory and Computational Methods

Rotational Diffusion Tensor Potential. Restraints on the protein rotational diffusion tensor were implemented in a potential energy term E_{diff} designed to minimize the sum of squares difference between the experimentally determined components of the protein rotational diffusion tensor, D_{ij}^{exp}, and those calculated from the molecular structure, D_{ij}^{calc}:

$$E_{\mathrm{diff}} = k_{\mathrm{diff}} \sum_{i,j} (D_{ij}^{\mathrm{exp}} - D_{ij}^{\mathrm{calc}})^2 \tag{1}$$

where k_{diff} is a force constant and the indices i and j range over the six unique components of the two symmetrical 3×3 rotational diffusion tensors.

To calculate the diffusion tensor given a molecular structure, we represent the surface of the protein by an equivalent ellipsoid[30] and then apply Perrin's equations[23,24] to calculate the diffusion tensor using the ellipsoid's dimensions and orientation. The key computational feature of this approach lies in an efficient method[33] for mapping the protein's surface using a fast triangulation algorithm, which is then used for building the equivalent ellipsoid. Once the coordinates of the surface are available, the calculation of the principal values of the rotational diffusion tensor, D_l^{calc}, can be expressed in closed form as

$$D_l^{\mathrm{calc}} = k_{\mathrm{B}} T_{\mathrm{diff}} / f_l \tag{2}$$

where k_{B} is the Boltzmann constant; $l = x, y, z$; T_{diff} is absolute temperature in kelvin, and

$$f_x = 16\pi\eta(a_y^2 + a_z^2)/[3(a_y^2 P_y + a_z^2 P_z)] \tag{3}$$

f_y and f_z are obtained from eq 3 by appropriate permutation of the indices. In eq 3, η is the solvent viscosity, and P_l are parameters describing the protein geometry, given by

$$P_x = \int_0^\infty \frac{\mathrm{d}s}{[(a_x^2 + s)^3 (a_y^2 + s)(a_z^2 + s)]^{1/2}} \tag{4}$$

with P_y and P_z given by cyclic permutations of the indices. The semi-axes of the equivalent ellipsoid, a_x, a_y, and a_z (in general $a_x \neq a_y \neq a_z$) in eqs 3 and 4 can be obtained directly from the eigenvalues, E_l, of the covariance matrix:

$$\mathrm{Cov}_{l,k} = \frac{1}{N_{\mathrm{p}}} \sum_{j=1}^{N_{\mathrm{p}}} (X_j^l - \langle X_j^l \rangle)(X_j^k - \langle X_j^k \rangle), \quad \langle X_j^l \rangle = \frac{1}{N_{\mathrm{p}}} \sum_{j=1}^{N_{\mathrm{p}}} X_j^l \tag{5}$$

where X_j^l are the coordinates of the protein surface representation; $l, k = x, y, z$; and N_{p} is the number of points used to represent the protein surface. The values of the semi-axes of the equivalent ellipsoid are then given by

(16) Svergun, D.; Barberato, C.; Koch, M. H. J. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
(17) Svergun, D. I.; Koch, M. H. *Curr. Opin. Struct. Biol.* **2002**, *12*, 654–660.
(18) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621–16628.
(19) Schwieters, C. D.; Clore, G. M. *Biochemistry* **2007**, *46*, 1152–1166.
(20) Grishaev, A.; Ying, J.; Canny, M. D.; Pardi, A.; Bax, A. *J. Biomol. NMR* **2008**, *42*, 99–109.
(21) Grishaev, A.; Tugarinov, V.; Kay, L. E.; Trewhella, J.; Bax, A. *J. Biomol. NMR* **2008**, *40*, 95–106.
(22) Gabel, F.; Simon, B.; Nilges, M.; Petoukhov, M.; Svergun, D.; Sattler, M. *J. Biomol. NMR* **2008**, *41*, 199–208.
(23) Perrin, F. *J. Phys. Radium* **1934**, *5*, 497–511.
(24) Perrin, F. *J. Phys. Radium* **1936**, *7*, 1–11.
(25) Favro, D. L. *Phys. Rev.* **1960**, *119*, 53–62.
(26) Woessner, D. E. *J. Chem. Phys.* **1962**, *37*, 647–654.
(27) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.
(28) Fushman, D.; Xu, R.; Cowburn, D. *Biochemistry* **1999**, *38*, 10225–10230.
(29) Dosset, P.; Hus, J. C.; Blackledge, M.; Marion, D. *J. Biomol. NMR* **2000**, *16*, 23–28.
(30) Ryabov, Y. E.; Geraghty, C.; Varshney, A.; Fushman, D. *J. Am. Chem. Soc.* **2006**, *128*, 15432–15444.
(31) Ryabov, Y.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 7894–7902.
(32) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.
(33) Varshney, A.; Brooks, F. P.; Wright, W. V. *IEEE Comp. Graphics App.* **1994**, *14*, 19–25.

$$a_l = \sqrt{3E_l} \tag{6}$$

Accordingly, the principal vectors of the diffusion tensor coincide with the corresponding eigenvectors of the covariance matrix, v_x, v_y, and v_z. Thus, given D_l^{calc} and v_l, the full rotational diffusion tensor can be reconstructed.

The gradients of E_{diff} with respect to all atomic displacements can be evaluated in closed form. However, a finite atomic displacement requires recalculation of the protein surface, which is computationally expensive. So, in practice, we recalculate the protein surface when the backbone atomic root-mean-square (rms) difference between the previously triangulated and current protein structure is greater than 0.5 Å, or after 30 gradient evaluations, whichever comes first. By default we assume that the protein is tumbling in water. Consequently, we use a series approximation derived from the tabulated values of the water viscosity η given in the *CRC Handbook of Chemistry and Physics*,[34] given by $\eta(T_{\text{diff}}) = 1.7753 - 0.0565(T_{\text{diff}} - 273) + (1.0751 \times 10^{-3})(T_{\text{diff}} - 273)^2 - (9.2222 \times 10^{-6})(T_{\text{diff}} - 273)^3$. However, other values of η can be specified explicitly (e.g. calculated using the program SEDNTERP 1.1, http://www.jphilo, mailway.com, or obtained experimentally). Two other critical parameters that affect the calculation of D_{ij}^{calc} and, therefore, the value of E_{diff} are the temperature, T_{diff}, and the thickness of the hydration layer used to construct the solvated protein. For the latter parameter, we assume a uniform value of 2.8 Å, which resembles a monolayer of water molecules.[30]

In many real experimental situations, the viscosity and temperature of the sample may not be known precisely. For example, there may be sample heating as a consequence of the application of strong radiofrequency pulses during the NMR experiment. In addition, the use of concentrated (>0.5 mM) solutions of large proteins may markedly change the viscosity of the sample. Further, the details of the hydration layer are expected to change from protein to protein such that, for example, it is common to assign a fitting parameter to this layer when comparing calculated and experimental SAXS data.[16] Uncertainties in the experimental viscosity and temperature coupled with approximations inherent in the hydration layer model require a procedure to account for these effects. Since the primary result of uncertainties from these three sources is a scaling of the diffusion tensor D^{calc}, we chose to compensate for these three factors by adjusting only the apparent diffusion tensor temperature $T_{\text{diff}}^{\text{app}}$. We reiterate that the value of $T_{\text{diff}}^{\text{app}}$ is no longer a physical temperature, but rather a fitting parameter to collect errors in temperature, viscosity, and hydration layer description. In this work we calculated structures using a series of different $T_{\text{diff}}^{\text{app}}$ values. The value of $T_{\text{diff}}^{\text{app}}$ that results in the lowest values of the total target energy is then chosen as the best match for the experimental and sample conditions. We have also implemented a procedure whereby $T_{\text{diff}}^{\text{app}}$ is optimized during the course of the calculations (see below).

Simulated Annealing Protocols. Two different protocols were employed, one for refinement of globular proteins and the other for docking protein−protein complexes, as described below and in the Supporting Information.

In the refinement protocol, the rotational diffusion tensor potential, E_{diff}, was simply added to the standard Xplor-NIH[8] structure refinement protocol in torsion angle space.[35] In addition to the E_{diff} term, the minimized target energy function comprises square-well potentials for the experimental NMR restraints, harmonic potentials for the covalent geometry (bonds, angles, and improper torsions), a quartic van der Waals repulsion term to prevent atomic overlap,[36] and a multidimensional torsion angle database potential of mean force.[12] Details of the protocol are provided in the Supporting Information. The experimental NMR data comprise NOE-derived approximate interproton distance restraints and backbone and side-chain torsion angle restraints for the N-terminal domain of EIN.[37] The components of the diffusion tensor were determined from the published ^{15}N R_1 and R_2 data for EIN[38] in conjunction with the NMR coordinates of EIN (PDB code 1EZA)[37] using well-established procedures.[25,26]

The refinement protocol can be used in two modes. In the first mode the value of $T_{\text{diff}}^{\text{app}}$ used to calculate the components of the rotational diffusion tensor is held fixed. In the second mode, the value of $T_{\text{diff}}^{\text{app}}$ is optimized during the course of the calculation. The latter is carried out by introducing three pseudoatoms, O, X, and Y, with the XOY angle, θ_{XOY}, directly mapped onto $T_{\text{diff}}^{\text{app}}$:

$$T_{\text{diff}}^{\text{app}} = T_{\text{diff}}^{0} + \Delta T_{\text{diff}} \cos \theta_{\text{XOY}} \tag{7}$$

where T_{diff}^{0} is the nominal experimental temperature, and the value of $T_{\text{diff}}^{\text{app}}$ can vary within a range of $T_{\text{diff}}^{0} \pm \Delta T_{\text{diff}}$. Aside from E_{diff}, the three pseudoatoms are coupled to the rest of the protein structure only through interaction with the thermal bath. The same approach was previously used to optimize the magnitude of the dipolar coupling alignment tensor.[39]

The docking protocol involves the application of conjoined rigid body/torsion angle dynamics and minimization[35] and was designed for protein−protein docking where the backbone structures of the individual proteins or subunits are known. Conceptually, this protocol consists of two parts: rough positioning of the proteins within the complex, followed by simulated annealing refinement of the complex. Specifically, the protocol starts with one protein or subunit held fixed and the position and orientation of the second protein randomized around the first. The second protein is then translated and rotated as a rigid body such that the energies corresponding to the experimental NMR restraints (i.e., the rotational diffusion tensor and, if available, supplementary data such as highly ambiguous intermolecular distance restraints from chemical shift perturbation mapping[40]) and the van der Waals repulsion term are minimized using conjugate gradient minimization. This procedure is repeated 10 times, and the resulting lowest-energy structure is used as the starting point for molecular dynamics simulated annealing optimization. Simulated annealing starts at 500 K and then performs relatively short runs of conjoined rigid body/torsion angle dynamics in 10 K steps down to a temperature of 10 K with the backbone of the first protein held fixed, the backbone of the second protein allowed to rotate and translate as a rigid body, and the surface-exposed side chains of both proteins given torsional degrees of freedom. (Note that in the case of a symmetric homodimer for which C_2 symmetry restraints are applied, both subunits are allowed to rotate and translate during simulated annealing.) In addition to terms representing the experimental restraints and the quartic van der Waals repulsion term, the target energy includes a term for the radius of gyration (calculated from the total number of residues, N, in the complex given by $R_{\text{gyr}} = 2.2N^{0.38}$)[14] to ensure reasonable intermolecular packing density (included only for the EIN/HPr docking calculations), a multidimensional torsion angle database potential of mean force to ensure that side-chain torsion angles populate physically realistic rotamers,[12] and a new hydrophobic contact potential described below. The force constants for most of the potential terms are set low at the higher temperatures and gradually increased during the course

(34) Weast, R. C. *CRC Handbook of Chemistry and Physics*, 1st student ed.; CRC Press: Boca Raton, FL, 1988.
(35) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *152*, 288–302.
(36) Nilges, M.; Gronenborn, A. M.; Brunger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27–38.
(37) Garrett, D. S.; Seok, Y. J.; Liao, D. I.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1997**, *36*, 2517–2530.
(38) Tjandra, N.; Garrett, D. S.; Gronenborn, A. M.; Bax, A.; Clore, G. M. *Nat. Struct. Biol.* **1997**, *4*, 443–449.
(39) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938.
(40) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2003**, *125*, 2902–2912.

of simulated annealing as the temperature is lowered. Further details of the docking protocol are provided in the Supporting Information.

The value of $T_{\mathrm{diff}}^{\mathrm{app}}$ can also be optimized in the docking protocol. However, for the docking protocol, we found that the optimization strategy described above for the refinement protocol was unsuccessful. This is largely due to the fact that variations in the value of $T_{\mathrm{diff}}^{\mathrm{app}}$ correspond essentially to changes in the trace of the calculated diffusion tensor and, therefore, to alterations in the overall size of the complex. As a result, variation of $T_{\mathrm{diff}}^{\mathrm{app}}$ during the course of the docking calculation can lead to instabilities that severely reduce the convergence properties of the protocol. We therefore employed an alternative strategy based on random sampling: for every run of initial rigid body domain positioning (by minimization), $T_{\mathrm{diff}}^{\mathrm{app}}$ was fixed to a random value within a specified range (usually ±5 K), and the number of repetitions was increased from 10 to 50. The lowest-energy structure was then used as the starting coordinates for the simulated annealing phase of the docking protocol.

Hydrophobic Contact Potential. A knowledge-based, low-resolution hydrophobic contact potential comprising interaction strengths calibrated using known protein structures[41] was added to Xplor-NIH and found to be useful in the docking calculations. Unlike a full, realistic nonbonded description including electrostatics and discrete solvent, this contact term is computationally inexpensive. Further, unlike the all-atom description, the contact potential is quite insensitive to side-chain conformations, information about which is usually lacking in docking calculations.

The contact potential depends linearly on the distance d_{ij} between residues i and j, defined as

$$d_{ij} = [\sum_{k_i,k_j} |q_{k_i} - q_{k_j}|^{-6}]^{-1/6} \tag{8}$$

where there are sums over q_{k_i}, the positions of the atoms in residue i, and over q_{k_j}, the positions of the atoms in residue j. The contact energy term, E_{contact}, is given by

$$E_{\mathrm{contact}} = k_{\mathrm{contact}} \sum_{i,j} M_{ij} V_p(d_{ij}) \tag{9}$$

where k_{contact} is a force constant, the sum is over all pairs of residues, M_{ij} is the interaction strength between the residue types as determined by Miyazawa et al.,[41] and V_{p} is the following continuous piecewise linear potential:

$$V_{\mathrm{p}}(d) = \begin{cases} 1, & d < d_< \\ 1 - \gamma(d - d_<)^2 + \delta(d - d_<)^3, & d_< < d < d_< + d_{\mathrm{sw}} \\ (d_> - d)/(d_> - d_<), & d_< + d_{\mathrm{sw}} < d < d_> - d_{\mathrm{sw}} \\ \gamma(d_> - d)^2 - \delta(d_> - d)^3, & d_> - d_{\mathrm{sw}} < d < d_> \\ 0, & d > d_> \end{cases} \tag{10}$$

where $\gamma = 2/[d_{\mathrm{sw}}(d_> - d_<)]$ and $\delta = [d_{\mathrm{sw}}^2(d_> - d_<)]^{-1}$. In this work we chose $d_< = 0$ Å, $d_> = 10$ Å, and $d_{\mathrm{sw}} = 0.5$ Å. Thus, the E_{contact} potential is fully enabled when the residues are touching and disabled for residues separated by more than 10 Å. The E_{contact} potential proved useful in discriminating against docked structures containing bad hydrophobic interactions which otherwise met the various other restraints.

Cluster Analysis. For the docking calculations, 512 independent structure calculations were carried out. To characterize the efficacy of the structure determination algorithm, it is essential to group the calculated structures together in clusters that are similar to one another by an appropriate metric. For this purpose we used the agglomerative algorithm outlined by Ward.[42] Initially, each structure is placed in a separate cluster, and the intercluster distance is computed as the rms distance between the Cα atoms in any two clusters. The clusters are then combined as follows: (a) the two closest clusters are found and combined; (b) a new distance matrix is generated in which the distance between clusters is the shortest structure−structure distance for these clusters; (c) this process is repeated until the shortest distance is larger than a specified tolerance, which in this work was set to 1 Å. This algorithm detects clusters of structures that are well-separated from each other in atomic rms displacement space but does not guarantee that structures within a given cluster are similar. For each cluster, a measure of this similarity can be calculated as the Cα atomic rms difference from the mean structure after the clusters have been determined. This clustering algorithm is available as the 'findClusters' helper script included in the Xplor-NIH package (version 2.22 or later).

Experimental Methods

Sample Preparation. EIN and HPr from *Escherichia coli* were expressed and purified as described previously.[43,44] The EIN construct employed in the current work consists of residues 1−249 of EIN. The original NMR[37] and X-ray[45] work on EIN made use of a construct comprising residues 1−259 of EIN. However, residues 250−259 of EIN are disordered in solution and invisible in the X-ray-derived electron density map. NMR samples were prepared in 20 mM Tris buffer, pH 7.4, 90% H_2O/10% D_2O (v/v). Two samples were used for ^{15}N relaxation measurements: the first comprised 0.5 mM U-[^{15}N/^{2}H]-EIN and 0.8 mM HPr at natural isotopic abundance; the second consisted of 0.5 mM U-[^{15}N/^{2}H]-HPr and 0.8 mM EIN at natural isotopic abundance.

NMR Spectroscopy. NMR spectra were recorded at 313 K on a DRX Bruker 600 MHz spectrometer equipped with a z-shielded gradient triple-resonance cryoprobe. ^{15}N R_1 and $R_1\rho$ relaxation measurements were carried out using pulse schemes described previously[46,47] and recorded in an interleaved manner using seven different relaxation delays. Delays of 103.3, 151.3, 391.3, 691.3, 1291.3, 1891.3, and 2491.3 ms were used for the R_1 relaxation measurements, and delays of 6.7, 16.3, 25.9, 32.3, 48.3, 64.3, and 96.3 ms were used for $R_{1\rho}$ relaxation measurement, with a recycle delay of 5 s for both measurements. $R_1\rho$ relaxation was obtained using a ^{15}N spin-lock field strength of 1.8 kHz. Spectra were processed using NMRPipe,[48] and peak intensity values were fit to a single-exponential decay. ^{15}N R_2 relaxation rates were calculated from the ^{15}N R_1 and ^{15}N $R_{1\rho}$ relaxation rates using the following equation:[49]

$$R_2 = (R_1 \sin^2\theta - R_{1\rho})/\cos^2\theta \tag{11}$$

where $\theta = \arctan(\Omega_{\mathrm{N}}/\gamma_{\mathrm{N}}B_1)$, Ω_{N} is the resonance offset, and $\gamma_{\mathrm{N}}B_1$ is the spin-lock field strength.

SAXS Measurements. SAXS data for EIN were collected on a SAXSess instrument from Anton Paar, which is configured as a Kratky camera coupled with high-flux multilayer monochromator optics. X-ray radiation from a sealed fine-focus tube source (Princeton Instruments), operating at 40 kV and 50 mA, was monochromated at the Cu Kα wavelength (1.542 Å) and incident on the sample in a 1-mm inner diameter quartz capillary of 24 mm

(41) Miyazawa, S.; Jernigan, R. L. *Proteins: Struct., Funct. Genet.* **1999**, *34*, 49–68.
(42) Ward, J. H. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
(43) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1997**, *36*, 4393–4398.
(44) Suh, J. Y.; Cai, M. L.; Clore, G. M. *J. Biol. Chem.* **2008**, *283*, 18980–18989.
(45) Liao, D. I.; Silverton, E.; Seok, Y. J.; Lee, B. R.; Peterkofsky, A.; Davies, D. R. *Structure* **1996**, *4*, 861–872.
(46) Farrow, N. A.; Zhang, O. W.; Forman-Kay, J. D.; Kay, L. E. *J. Biomol. NMR* **1994**, *4*, 727–734.
(47) Tjandra, N.; Wingfield, P.; Stahl, S.; Bax, A. *J. Biomol. NMR* **1996**, *8*, 273–284.
(48) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(49) Davis, D. G.; Perlman, M. E.; London, R. E. *J. Magn. Reson. Ser. B* **1994**, *104*, 266–275.

length, thermostatted at 298 K. A line-shaped X-ray beam 20 mm in length was used to maximize the incident flux. Sample buffer conditions were the same as those used for NMR measurements described above, except that 150 mM NaCl was used to suppress the effects of interparticle correlations (structure factor). Data were collected as a series of sequential 1 h acquisitions with the protein sample, followed immediately by the dialysis buffer. Due to signal relaxation, the imaging plates were read out with a 5 min delay at the end of each acquisition session. Data at 50% dilution were collected to investigate the magnitude of the interparticle structure factor. Wide-angle scattering data were collected within a q-range of ~0.02 to ~2.80 Å^{-1}. Here, $q = 4\pi \sin \theta/\lambda$, where 2θ is the scattering angle and λ the wavelength of the incident radiation. The recorded two-dimensional (2D) images were converted to one-dimensional (1D) scattering profiles by radial integration within 15 mm strips aligned at the center of the incident beam. 1D profiles were then mapped onto the q-axis by reference to the position of the primary beam attenuated by the semitransparent beamstop of the instrument. The converted profiles were corrected for the readout noise of the imaging plate scanner and normalized to the recorded intensities of the transmitted primary beam. The scattering curves from the buffer were then subtracted from the scattering curve of the protein sample. The final scattering data were averaged over three independent sample/buffer data acquisitions of 1 h each. The line-collimation 1D profiles were desmeared using the GNOM software,[50] taking into account the recorded length profile of the incident beam. A maximum dimension of 70 Å was obtained by application of GNOM's regularized Fourier transform procedure. The resulting point collimation-like data were used for the subsequent structural analysis in the q interval from 0.03 to 0.60 Å^{-1} (crystallographic resolutions between ~300 and ~10 Å). Evaluation of the quality of the fit of the scattering data to the various structural models was carried out with CRYSOL version 2.5.[16]

Results and Discussion

Refinement of a Globular Protein Structure. To demonstrate the impact of rotational diffusion tensor restraints on structure refinement of globular proteins, we made use of the N-terminal domain of enyzme I (EIN). EIN is an elongated molecule of approximately 30 kDa with a ratio of 3:3:1 for the principal components of the inertia tensor.[37,38] The initial structure used for refinement is the NMR structure (specifically the restrained regularized mean coordinates; PDB code 1EZA) based on NOE, torsion angle, $^3J_{HN\alpha}$ coupling constant, and $^{13}C\alpha/^{13}C\beta$ restraints.[37] The target values for the tensor components of the rotational diffusion tensor were obtained by least-squares minimization between the observed R_1/R_2 ratios (recorded at 313 K on a sample containing 1.1 mM EIN)[38] and those calculated from the 1EZA coordinates (see Supporting Information). Only data in regular secondary structure elements were employed to exclude potential errors arising from any large-scale motions in the picosecond to nanosecond time scale or from exchange line broadening and to minimize the impact of inaccuracies in the coordinates. The resulting tensor components for an axially symmetric diffusion model (see Supporting Information) were then used as target values in a simulated annealing refinement protocol incorporating the rotational diffusion tensor potential E_{diff}. The other experimental NMR data included in the refinement were a set of distance restraints derived from NOE measurements (2818 interproton distance restraints) and backbone hydrogen bond analysis (230 distance restraints for 115 backbone hydrogen bonds) and 571 torsion angle restraints.[37] The target function also included a quartic van der Waals repulsion term and a multidimensional torsion angle database potential of mean force.[12] The latter is an improved version relative to that used in the original NMR structure determination.[11]

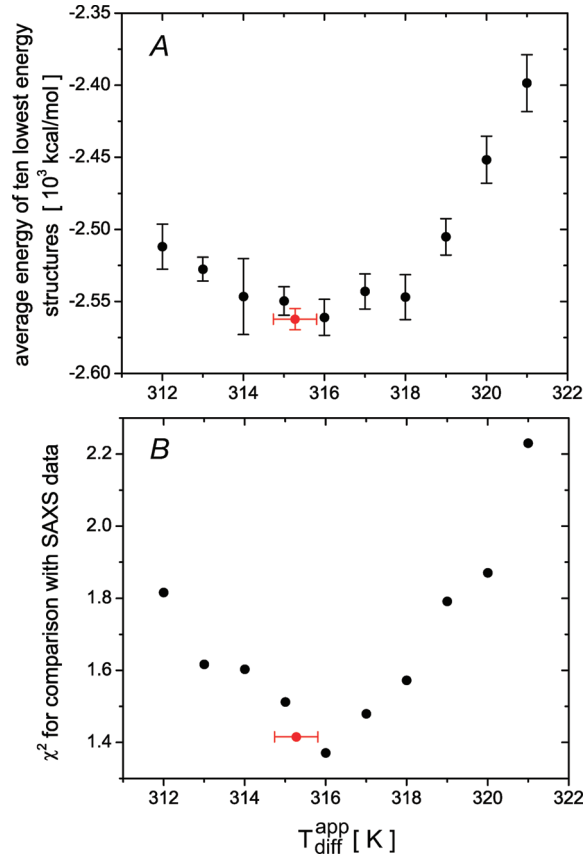

Figure 1. Results of refinement of the structure of EIN with rotational diffusion tensor restraints as a function of the T_{diff}^{app} value used in the calculations. (A) Total Xplor-NIH energies, averaged over the 10 lowest-energy structures (errors bars, 1 SD). (B) χ^2-fit[16] of the calculated structures to the SAXS data. The structures used for fitting the SAXS data are the restrained regularized averages over the 10 lowest-energy structures calculated at each value of T_{diff}^{app}. The black symbols correspond to the structures obtained using a grid search for the T_{diff}^{app} settings; the red symbols are the results of automated optimization of T_{diff}^{app}.

The values of the total Xplor-NIH energies averaged over the 10 lowest-energy structures obtained at a series of T_{diff}^{app} values are plotted in Figure 1A. The minimum in the Xplor-NIH energy is observed at $T_{diff}^{app} = 316$ K, compared to a nominal experimental temperature of 313 K. Independent validation of these results is obtained from the corresponding agreement of these structures with experimental SAXS data, where the minimum χ^2 value is also found for the structures calculated at $T_{diff}^{app} = 316$ K (Figure 1B). Thus, the agreement between the locations of the minima in the T_{diff}^{app} dependence of the total Xplor-NIH energy and the independently validated χ^2-fit to the SAXS data strongly suggest that the use of the rotational diffusion tensor for protein structure refinement provides structures that are consistent not only with the experimental NMR restraints used in the refinement but also with other independent experimental information. When the force constant for the diffusion tensor potential is increased 5-fold, the minimum in the total Xplor-NIH energy is found at $T_{diff}^{app} = 315$ K compared to a minimum at 316 K for the χ^2-fit to the SAXS data (Figure S1, Supporting Information). This difference is sufficiently small that the

(50) Semenyuk, A. V.; Svergun, D. I. *J. Appl. Crystallogr.* **1991**, *24*, 537–540.

Table 1. Cα Atomic Root-Mean-Square Differences for the Refined Enzyme I Structures

	Cα rms difference (Å)	
	refinement with E_{diff}	refinement without E_{diff}
precision[a]	0.93	1.02
vs 1EZA NMR structure[b]	1.26	1.21
vs 1ZYM X-ray structure[b]	1.37	1.36
vs each other[b]	0.43	

[a] Defined as the Cα rms difference from the mean coordinates averaged over the 10 lowest-energy structures. [b] Cα atomic rms difference from the mean coordinates averaged over the 10 lowest-energy structures.

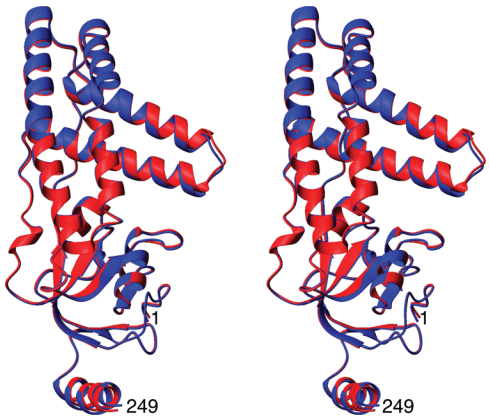

Figure 2. Stereoview showing a best-fit superposition of the two restrained regularized average structures obtained from refinement with (red) and without (blue) the rotational diffusion tensor restraints. The value of T_{diff}^{app} was 316 K.

Table 2. Comparison of EIN Structures with SAXS data[a]

	SAXS χ^2
1EZA (original NMR)	2.03
1ZYM (X-ray)	1.38
refinement without E_{diff}	1.58
refinement with E_{diff}	1.37

[a] Analysis of the SAXS data was performed using CRYSOL version 2.5.[16] The SAXS data were acquired on EIN(1−249), and therefore the analysis was performed using residues 1−249. The NMR and X-ray data were acquired on EIN(1−259). However, residues 250−259 are disordered in solution[37,38] and invisible in the electron density map.[45]

procedure of choosing the T_{diff}^{app} value that minimizes the overall Xplor-NIH energy seems to be a reasonable strategy and leads to a degree of self-consistency. Automated optimization of T_{diff}^{app} during simulated annealing (within a range of 313 ± 5 K) also indicates that the lowest-energy structures correspond to the optimal structures at 315.3 ± 0.5 K (see Figure 1).

Comparison of previously determined X-ray (1ZYM[45]) and NMR (1EZA[37]) structures with the structures refined using the current refinement protocol with and without inclusion of the E_{diff} term is presented in Table 1. Although the two sets of structures calculated with the current protocol are quite similar to each other (Cα atomic rms difference of 0.4 Å; Figure 2) and exhibit approximately the same Cα atomic rms difference from the 1ZYM and 1EZA structures, the addition of the E_{diff} term does lead to a significant improvement in agreement with the SAXS data (Table 2 and Figure S2, Supporting Information). Indeed, the χ^2-fit to the SAXS data is as good for the current NMR structure refined with the E_{diff} term as for the 1ZYM X-ray structure.

Protein−Protein Docking. In the two following sections we examine the impact of rotational diffusion tensor restraints on

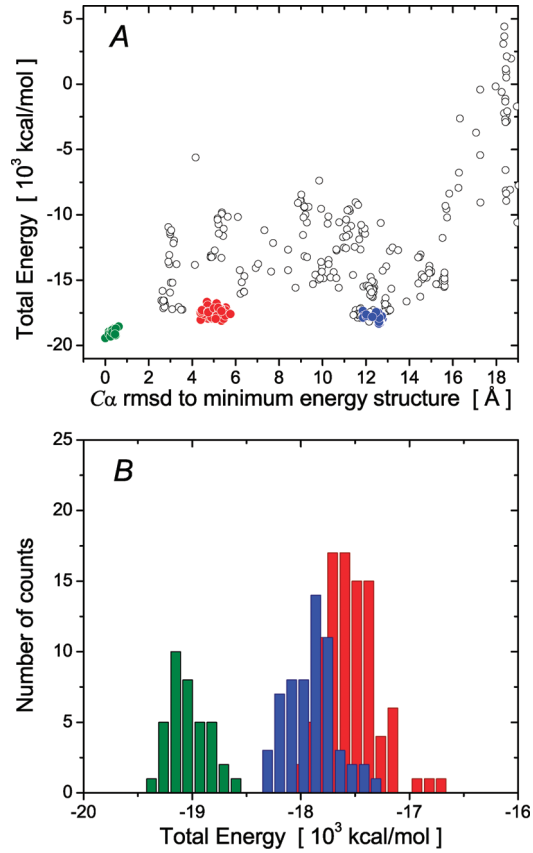

Figure 3. Statistical properties of the largest clusters of docking solutions for HIV-1 protease. (A) Dependence of the total Xplor-NIH energy on the Cα rms difference from the minimum energy structure. Structures in the first, second, and third largest clusters are indicated by red, blue, and green filled-in circles, respectively. The remaining structures, the majority of which are in single-member clusters, are shown as open circles. (B) Histograms of the distribution of total Xplor-NIH energy values with the same color-coding scheme as in panel A. The value of T_{diff}^{app} was 303 K.

protein−protein docking using two extreme examples. The first consists of assembling the two identical subunits of HIV-1 protease into the homodimer on the basis of only rotational diffusion tensor restraints. The second involves assembly of the EIN−HPr complex from the X-ray structures of the free proteins based upon the rotational diffusion tensor restraints combined with highly ambiguous distance restraints derived from chemical shift perturbation mapping. In most instances additional experimental data would be available in the form of intermolecular distance restraints (e.g., from NOE, paramagnetic relaxation enhancement, FRET, or DEER measurements), possibly supplemented by orientational restraints (from dipolar couplings or paramagnetic-induced pseudocontact shifts).

Application to the HIV-1 Protease Dimer. HIV-1 protease is a symmetric dimer of identical subunits. Here we demonstrate the utility of rotational diffusion tensor restraints alone to correctly assemble a homodimer. In the context of an NMR structure determination of a homodimer, this type of docking calculation would be useful in cases where the structure of an individual subunit is determined with confidence but no reliable intersubunit NOE information could be obtained (e.g., due to spectral overlap, exchange line broadening of interfacial residues, inability to make mixed isotopically labeled samples, etc.).

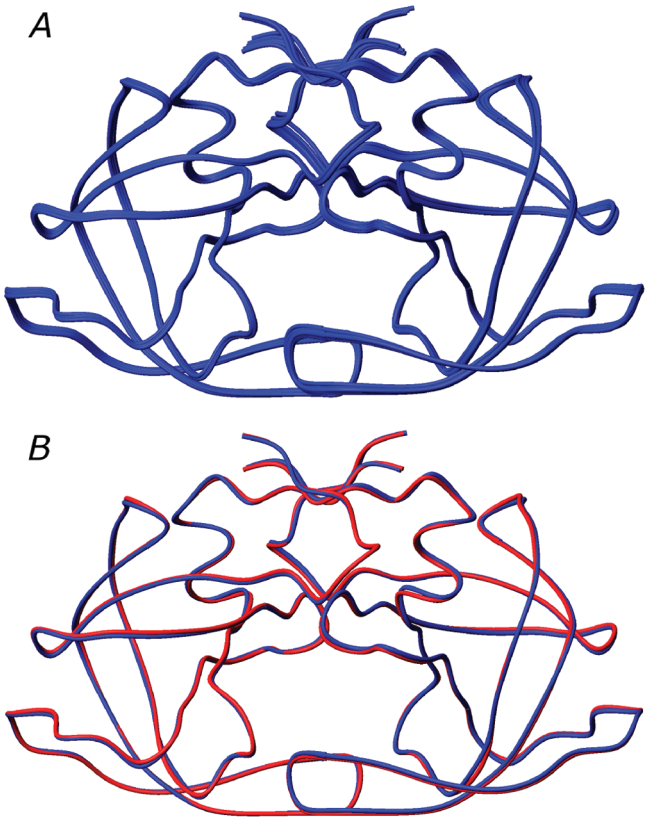

Figure 4. Lowest-energy HIV-1 protease dimer structures. (A) Best-fit superposition of the 10 lowest-energy structures from the cluster containing the lowest-energy structure. (B) Comparison of the restrained regularized mean structure (blue), derived from the 10 lowest-energy structures, with the subunit arrangement in the X-ray structure[51] (red). The value of $T_{\mathrm{diff}}^{\mathrm{app}}$ was 303 K.

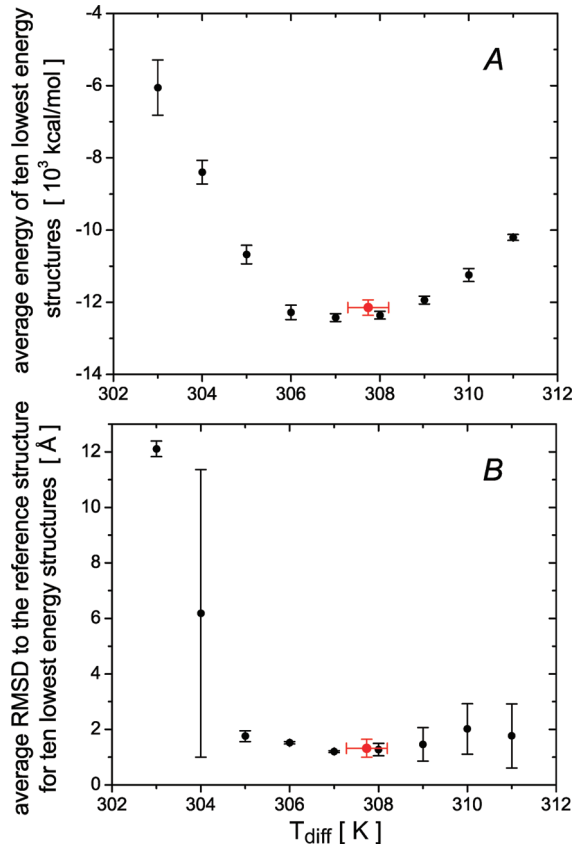

Figure 5. Results of several runs of the docking protocol for the EIN-HPr complex using different values of $T_{\mathrm{diff}}^{\mathrm{app}}$. The data for the two panels are averaged over the 10 lowest-energy structures (errors bars = 1 SD): (A) total target energy and (B) Cα atomic rms difference from the reference structure. In both panels, the black symbols correspond to the grid search results obtained with a fixed value of $T_{\mathrm{diff}}^{\mathrm{app}}$; the red symbols are the results using automated optimization of $T_{\mathrm{diff}}^{\mathrm{app}}$.

Calculations were performed using the X-ray coordinates for the individual subunits (PDB code 2NPH[51]). The only experimental restraints employed in the docking protocol were the components of the rotational diffusion tensor calculated from ^{15}N relaxation data acquired at 300 K (see Supporting Information).[47] More specifically, we used the experimental data recorded at a spectrometer frequency of 600 MHz and derived the diffusion tensor using the coordinates of the first subunit of the 2NPH structure. Although a fully anisotropic diffusion tensor model yields a lower χ^2 than the axially symmetric model, the statistical F-test criterion does not justify the use of the fully anisotropic model (see Supporting Information for details). Trials with D^{calc} computed using different $T_{\mathrm{diff}}^{\mathrm{app}}$ values show that a value of 303 K results in structures with the lowest total Xplor-NIH energies (Figure S3, Supporting Information).

Since HIV-1 protease is a natural homodimer, only the coordinates of one of the subunits of 2NPH[51] were used in the calculations. One copy of this subunit was held fixed during the initial stage of rigid body minimization, while the position and orientation of the second copy were randomized around the center of gravity of the first copy. In the crystal structure, the side chains at the dimer interface are in their optimal configurations to yield good packing. To mimic the more general situation in which the docking protocol starts with undocked subunits or proteins, the side-chain conformations of the individual subunits were randomized by running 1 ps of torsion angle molecular dynamics at 500 K with the backbone coordinates held fixed. The resulting structures were then subjected to the docking protocol described in the Computational Methods section and Supporting Information.

Table 3. Clustering Statistics and Cα Atomic Rms Differences from the Reference Structure for the EIN-HPr Docking Calculations

relaxation data	% of structures in cluster containing lowest-energy structure	Cα rms difference from reference structure (Å)[a] overall	HPr with EIN superimposed
experimental data	76	1.20 ± 0.03	2.79 ± 0.12
synthetic data			
0% noise[b]	83	0.51 ± 0.27	1.32 ± 0.76
0% noise	83	0.48 ± 0.19	1.35 ± 0.39
1% noise	80	0.59 ± 0.15	1.65 ± 0.32
2% noise	89	0.79 ± 0.17	2.04 ± 0.37
5% noise	82	0.48 ± 0.15	1.27 ± 0.43
10% noise	72	1.82 ± 0.35	4.08 ± 1.02

[a] The reference structure of the complex was generated by best-fitting the X-ray structures of free EIN (1ZYM)[45] and HPr (1POH)[54] to the NMR structure of the EIN-HPr complex (3EZA).[52] This ensures that the reported rms differences do not reflect internal coordinate differences between the NMR and X-ray coordinates of EIN and HPr.[40] [b] The hydrophobic contact potential was not included in these calculations.

The results of cluster analysis of the 512 calculated structures are shown in Figure 3. There are three major clusters of solutions, comprising 19, 11, and 7% of the structures,

(51) Das, A.; Prashar, V.; Mahale, S.; Serre, L.; Ferrer, J. L.; Hosur, M. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18464–18469.

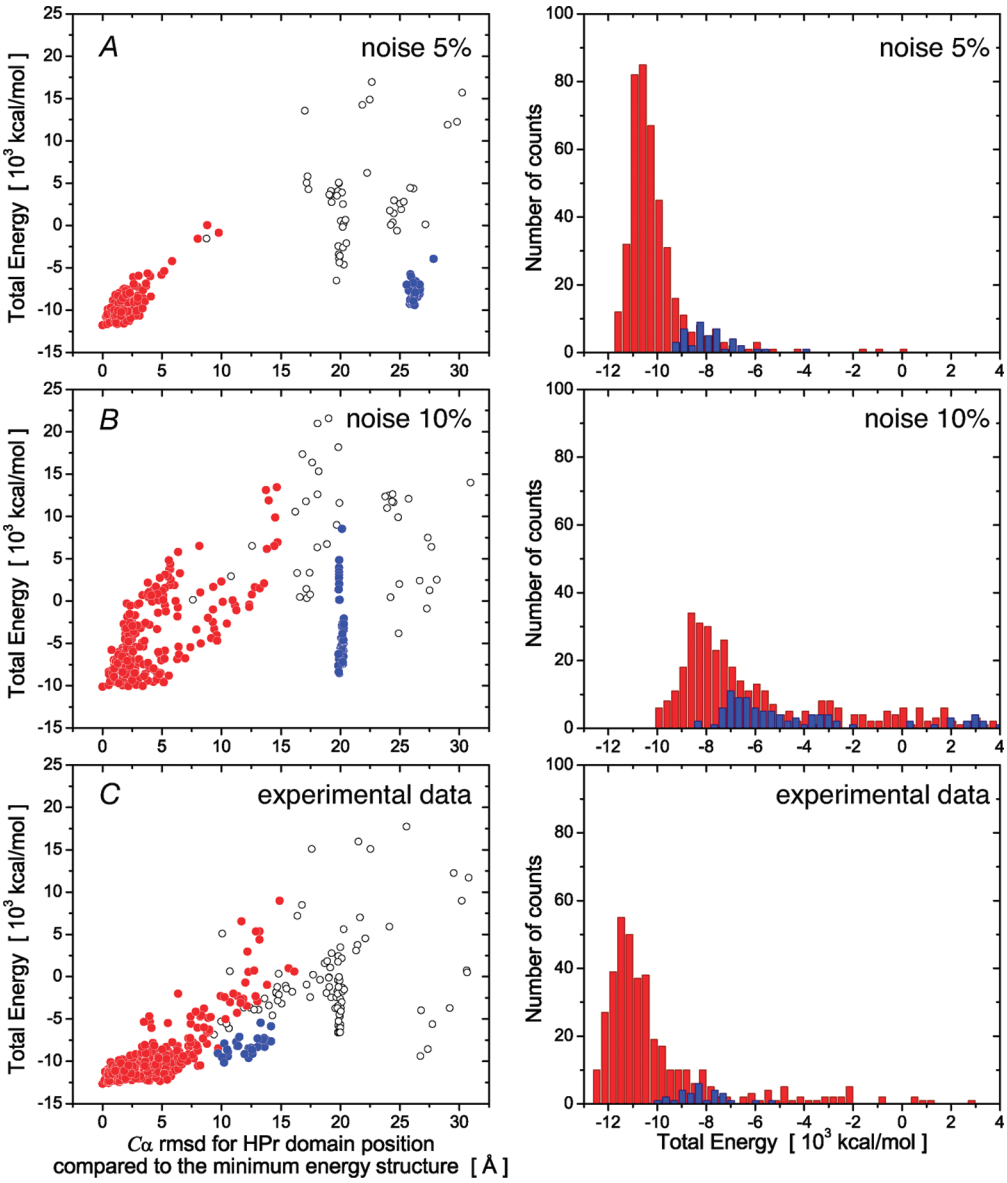

Figure 6. Statistical properties of the two largest clusters of docking solutions for the EIN-HPr complex using rotational diffusion tensors derived from simulated (5 and 10% added noise) and experimental ^{15}N R_1 and R_2 relaxation data: (A,B) synthetic diffusion tensor derived from simulated relaxation data with 5 and 10% added Gaussian noise, respectively; (C) experimentally determined diffusion tensor. The left-hand panels display the Cα atomic rms difference of HPr relative to the minimum energy structure with the coordinates of EIN superimposed; the right-hand panels present a histogram of the distribution of total Xplor-NIH energies for the two largest clusters of calculated structures. The filled-in red and blue circles represent the first and second largest clusters of docked structures, respectively; the open circles in the left-hand panels correspond to structures in the remaining smaller clusters.

respectively. The third cluster contains the lowest-energy structures. The rest of the structures are distributed among even smaller clusters, the majority of which include only a single member. We compared the 10 lowest-energy structures (Figure 4A), all of which are members of the same cluster, to the reference X-ray structure. (Note, in the reference structure the two subunits are identical and the dimer coordinates were created by superimposing the coordinates of the first subunit onto the second subunit of the 2NPH structure.) The average Cα rms difference beween the 10 lowest-energy dimer structures and the reference structure is 0.35 ± 0.09 Å; the Cα rms difference between the regularized mean coordinates and the reference structure is 0.31 Å (Figure 4B). Thus, the rotational diffusion tensor-based docking protocol coupled with clustering analysis is sufficient in its own right to define the subunit arrangement of the HIV-1 protease homodimer.

Application to the EIN-HPr Complex. The structure of the 40 kDa EIN-HPr complex was previously determined by NMR based on NOE and residual dipolar coupling data[52] and has been used as a test system for docking based on highly ambiguous intermolecular distances derived from chemical shift perturbation mapping either alone[53] or in combination with residual dipolar coupling data.[40] Here we demonstrate the use of rotational diffusion tensor restraints in conjunction with highly ambiguous intermolecular distance restraints to dock the EIN-HPr complex using the crystal structures of free EIN (1ZYM[45]) and HPr (1POH[54]) as starting coordinates.

(52) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166–173.
(53) Dominguez, C.; Boelens, R.; Bonvin, A. M. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.

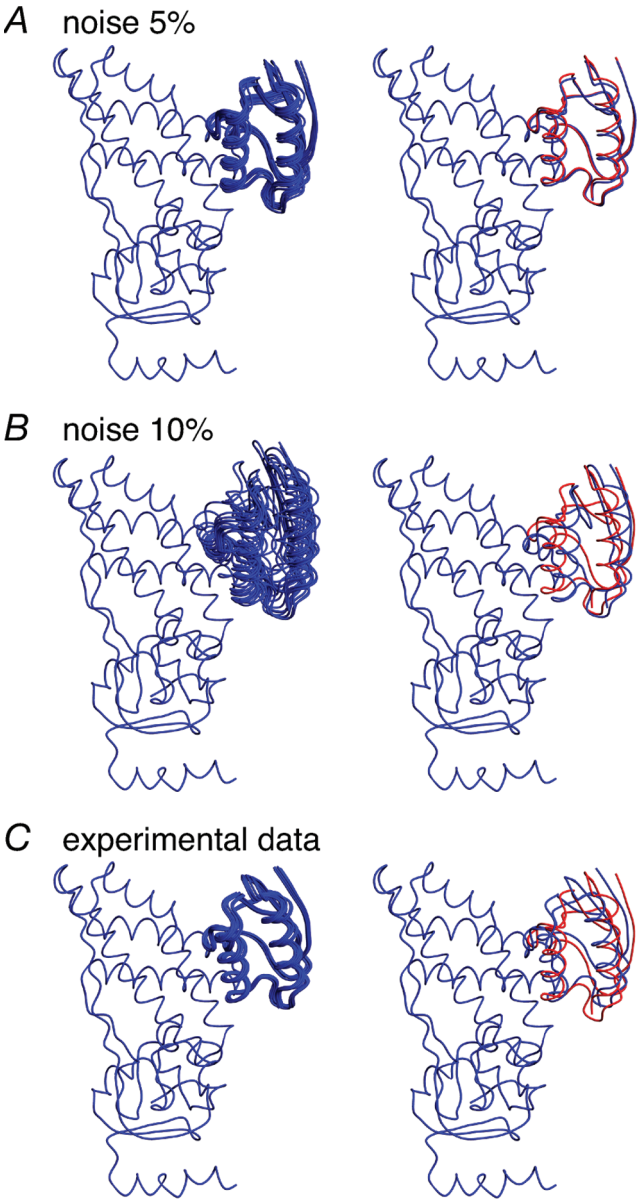

Figure 7. Lowest-energy docked structures obtained from rotational diffusion tensors derived from (A, B) simulated (5 and 10% added Gaussian noise) and (C) experimental ^{15}N R_1 and R_2 relaxation data. In all panels the coordinates of EIN are superimposed. The left-hand panels show a superposition of the 10 lowest-energy structures; the right-hand panels show the comparison of the restrained regularized mean structures (blue) derived from the 10 lowest energy structures with the reference structure (red).

The ^{15}N R_1 and R_2 relaxation rates for the EIN-HPr complex were acquired as described in the Experimental Methods section and used to calculate the components of the rotational diffusion tensor making use of only the coordinates of N−H bond vectors in regions of regular secondary structure. The eigenvalues of the diffusion tensor were assumed to be the same for the EIN and HPr partners in the complex. However, the orientations of the principal axis frame (PAF) of the diffusion tensor with respect to the molecular frames of 1ZYM[45] and 1POH[54] were fit independently. The relaxation data were analyzed using two different models for the rotation diffusion tensor: axially symmetric and fully anisotropic. Comparison of the normalized χ^2 values for these models using the F-test[55] criterion (see Supporting Information) indicates that the fully anisotropic model does not provide a statistically significant improvement in the description of the relaxation data. We therefore used the tensor components of the diffusion tensor obtained from the axially symmetric model in the docking calculations. In addition to the rotational diffusion tensors, the docking protocol used highly ambiguous distance restraints obtained from chemical shift perturbation mapping,[40] a hydrophobic contact potential (see Computational Methods section), a radius of gyration (R_{gyr}) potential with the target R_{gyr} value set to 20 Å,[14] and a multidimensional torsion angle database potential of mean force for the side chains.[12]

Unlike the identical subunits of the HIV-1 protease homodimer, EIN (249 residues) and HPr (85 residues) are markedly different in size and shape and thus represent a more general situation. In the generic simulated annealing docking protocol described in the Computational Methods section, one of the partners of the complex is held fixed in space together with the corresponding orientation of the diffusion tensor. Note that the diffusion tensor not only provides shape restraints but also contains information about its orientation with respect to a molecular reference frame. With HPr held fixed and EIN free to rotate and translate, convergence to the correct structure is achieved. If, on the other hand, EIN is held fixed and HPr is free to rotate and translate, HPr is placed at the correct binding site, but the correct orientation of HPr cannot be determined. This is because reorientation of HPr does not significantly affect the overall shape of the complex, and hence the calculated diffusion tensor, owing to the relatively small size and almost spherical shape of HPr. When HPr, on the other hand, is held fixed and EIN is allowed to rotate and translate, the simulated annealing docking protocol correctly determines the structure of the complex since the structure of EIN is quite asymmetric and much larger than that of HPr.

In addition to calculations with the experimental ^{15}N relaxation data, calculations with several sets of synthetic relaxation data were carried out to assess the robustness of the docking protocol. To this end we calculated the diffusion tensor for a synthetic reference structure assembled from the X-ray coordinates of EIN (1ZYM[45]) and HPr (1POH[54]) superimposed onto the NMR structure of the EIN-HPr complex (3EZA[52]). This diffusion tensor, together with the structures of EIN and HPr, was used to calculate simulated ^{15}N R_1 and R_2 relaxation rates to which random Gaussian noise ranging from 0 to 10% was added. The simulated data were then processed to calculate synthetic diffusion tensors.

Trials with different T_{diff}^{app} values using the diffusion tensor derived from the experimental ^{15}N relaxation data yielded an optimal value of $T_{diff}^{app} = 307$ K (compared to the nominal experimental temperature of 313 K) resulting in both the lowest overall Xplor-NIH target energy and the smallest Cα atomic rms difference from the reference structure (Figure 5). Automated optimization of T_{diff}^{app} (within a range 313 ± 8 K) leads to a very similar value of 307.7 ± 0.5 K (Figure 5).

The performance of the diffusion tensor-based docking protocol for the EIN-HPr complex is summarized in Table 3 and Figures 6 and 7. For both simulated and experimental target diffusion tensors, the resulting sets of calculated structures always exhibit significant clustering, with the cluster containing

(54) Jia, Z. C.; Quail, J. W.; Waygood, E. B.; Delbaere, L. T. J. *J. Biol. Chem.* **1993**, *268*, 22490–22501.

(55) Snedecor, G. W.; Cochran, W. G. *Statistical methods*, 8th ed.; Iowa State University Press: Ames, IA, 1989.

the lowest-energy structure comprising more than 75% of all calculated structures. The trials with simulated diffusion tensors presented in Table 3 suggest that the docking protocol demonstrates reasonable accuracy and precision up to a level of 10% noise in the simulated R_1 and R_2 relaxation data. Inclusion of the hydrophobic contact potential energy term does not significantly affect the accuracy of the computed structures but improves both convergence and precision. Using the experimentally derived components of the rotational diffusion tensor yields good agreement between the docked and reference structures (Table 3 and Figures 6 and 7). Figure 6 presents more detailed statistics for the distributions of Xplor-NIH energy values obtained for the docking calculations with both simulated (with 5 and 10% Gaussian noise) and experimental data. The corresponding lowest-energy structures calculated from these input data sets are displayed and compared to the reference structure in Figure 7.

In the above calculations, the diffusion tensor was derived from relaxation data acquired for both components of the EIN-HPr complex. In principle, however, relaxation data need only be acquired for one of the components of a complex, since the resulting diffusion tensor will contain all the relevant information on the rotational diffusion for the entire complex. (Note that a reduction in the number of R_1/R_2 data points will lead to a decrease in the accuracy with which the components of the diffusion tensor are determined.[56]) Such an approach could be especially useful for complexes in which one of the components is large, and hence its data are more difficult to analyze due to spectral overlap. The use of relaxation data for the smaller component may then still be sufficient to permit reasonably accurate docking of the complex. In the case of the EIN-HPr complex, using only the relaxation data from HPr to derive the components of the rotational diffusion tensor results in an average Cα rms difference between the 10 lowest-energy docked structures and the reference structure of 2.12 ± 0.28 Å, compared to 1.20 ± 0.03 and 1.82 ± 0.35 Å when the experimental and synthetic (with 10% added noise) relaxation data, respectively, for both EIN and HPr are employed.

Concluding Remarks

In this paper we have shown that the components of the rotational diffusion tensor provide effective restraints on overall molecular shape and size. The incorporation of rotational diffusion restraints into protein structure refinement leads to increased accuracy in the resulting shape and size, as judged by improvements in the agreement with independent SAXS data. The rotational diffusion tensor restraints can also be used to efficiently assemble complexes comprising proteins of known structure either in their own right or in conjunction with minimal additional experimental information, such as highly ambiguous intermolecular distance restraints derived from chemical shift perturbation mapping. The latter are required when the shape of one or more components of the complex is close to spherical or when one of the components of the complex is significantly smaller than the other. Although there are uncertainties in sample temperature, viscosity, and the representation of the protein hydration shell, these can be readily compensated for by optimization of the total target function energy with respect to $T_{\mathrm{diff}}^{\mathrm{app}}$, either using a grid search or by optimizing $T_{\mathrm{diff}}^{\mathrm{app}}$ during the course of the calculations. The fact that the optimal value of $T_{\mathrm{diff}}^{\mathrm{app}}$ is close to the nominal experimental temperature indicates that the approximations used in the representation of the protein hydration shell (possibly the largest source of uncertainty, the details of which are expected to vary from protein to protein),[16] including the assumption of a uniform thickness of 2.8 Å, are remarkably good, and similarly the estimate of sample viscosity based on the viscosity of water as a function of temperature[34] is reasonably accurate.

Acknowledgment. We thank Dan Garrett and John Kuszewski for useful discussions. Y.R. acknowledges a National Research Council Research Associateship (Award No. 0710430). This work was supported by the NIDDK Intramural Research Program of the NIH (to G.M.C.), the AIDS Targeted Antiviral Program of the Office of the Director of the NIH (to G.M.C.), and the CIT intramural research program of the NIH (to C.D.S.)

Supporting Information Available: Simulated annealing protocols; experimentally derived diffusion tensors used as target parameters in simulated annealing; validation of the refined EIN structure with SAXS data; and dependence of various parameters on $T_{\mathrm{diff}}^{\mathrm{app}}$ in the docking calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA902336C

(56) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127–137.