**SUPPLEMENTARY INFORMATION**

# 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,[1] Jeong-Yong Suh,[2] Alexander Grishaev,[2] G. Marius Clore,[2,*] and Charles D. Schwieters[1,*]*

[1]Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624,

[2]Laboratory of Chemical Physics, Building 5, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520.

## 1. Simulated annealing protocols.

### 1.1. Protocol for structure refinement

- *Step 1*: High temperature dynamics in torsion angle space (3000 K, 10 ps or 5000 steps whichever comes first).
- *Step 2*: Simulated annealing in torsion angle space (3000 to 25 K in 12.5 K increments, with 0.2 ps or 100 steps, whichever comes first, at every temperature).
- *Step 3*: Gradient minimization in torsion angle space.
- *Step 4*: Final gradient minimization in Cartesian space.

The protocol can be used in two modes: either with a fixed value of $T_{diff}^{app}$, or optimizing $T_{diff}^{app}$ during the course of simulated annealing. In the latter case $T_{diff}^{app}$ is optimized by introducing three pseudo atoms, O, X and Y with the XOY angle directly mapped onto $T_{diff}^{app}$ (see Eq. 7, main text). The three pseudoatoms only interact with one another and, aside from $E_{diff}$, are coupled to the atoms of the macromolecule through interaction with the thermal bath. Optimization of the OX and OY orientations during the course of simulated annealing leads to optimization of the $T_{diff}^{app}$ setting by minimizing the diffusion potential term $E_{diff}$.

**Table S1** Potential terms used in the structure refinement protocol.

| | force constant | | |
|---|---|---|---|
| Potential terms | high temperature | simulated annealing | final minimization |
| bonds (kcal.mol$^{-1}$.Å$^{-2}$)[a] | 1000 | 1000 | 1000 |
| angles (kcal.mol$^{-1}$.rad$^{-2}$)[a] | 400 | 400→500 | 500 |
| impropers (kcal.mol$^{-1}$.rad$^{-2}$)[a] | 100 | 100→500 | 500 |
| NOE (kcal.mol$^{-1}$.Å$^{-2}$) | 0.1 | 2→30 | 30 |
| torsion angles (kcal.mol$^{-1}$.rad$^{-2}$) | 10 | 200 | 200 |
| diffusion tensor[b] | 1 | 1 | 1 |
| repulsive vdw (kcal.mol$^{-1}$.Å$^{-4}$) | 0.004 | 0.004→4 | 4 |
| torsion angle database potential | 0.002 | 0.002→1 | 1 |

[a]Note that during torsion angle dynamics, the bond, angle and improper terms are only applied to bonds within closed rings structures (e.g. one of the bonds of a proline ring) or closed loops (e.g. a disulphide bridge) that are broken to permit the appropriate tree topology to be created.

[b]The values shown represent the scale factor $s_{diff}$; the actual force constant, $k_{diff}$, for the diffusion tensor restraints is obtained by multiplying $s_{diff}$ by 200x($N$/855) where $N$ is the number of atoms. This multiplicative factor was calibrated using the GB1 domain of streptococcal protein G (Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337-2338).

**1.2 Simulated annealing docking protocol using diffusion tensor restraints**

- *Step 1*: Randomization of positions and orientations of proteins relative to one another.[a]
- *Step 2*: Initial rigid body gradient minimization with experimental restraints.[b]
- *Step 3*: Initial rigid body gradient minimization with experimental restraints and van der Waals repulsion term.[b]
    - For calculations with a fixed $T_{diff}^{app}$, steps 1-3 are repeated 10 times for every run of the protocol. The structure with the lowest energy is then used for subsequent steps of the protocol.
    - For calculations in which $T_{diff}^{app}$ is optimized, Steps 1-3 are repeated 50 times for every run of the protocol, using a random value of $T_{diff}^{app}$ within a defined range. The structure with the lowest energy is then used for subsequent steps of the protocol.
- *Step 4*: Conjoined rigid body/torsion angle dynamics simulated annealing (from 500 to 10 K in 10 K increments with 1 ps or 500 steps, whichever comes first, at every temperature in the case of the HIV-1 protease calculations; and 0.5ps or 300 ps, whichever comes first, at every temperature for the EIN-HPr calculations).[c]
- Step 5: Final conjoined rigid body/torsion angle minimization.[c]

A total of 512 structures were calculated.

---

[a] For HIV-1 protease the position and orientation of second subunit was randomized within a cube 30x30x30 Å around the center of gravity of the first subunit; for the EIN-HPR complex the position and orientation of EIN was randomized within a cube 45x45x45 Å around the center of gravity of HPr.

[b] For HIV-1 protease, the only experimental restraints used are components of rotational diffusion tensor; for the EIN-HPr complex the experimental restraints consist of the components of the rotational diffusion tensor together with ambiguous distance restraints derived from chemical shift perturbation mapping (see Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc*. 2003, *125*, 2902-2912).

[c] The backbones are treated as rigid bodies, while the sidechains are given torsional degrees of freedom. A multidimensional torsion angle database potential of mean force is used to ensure that the sidechain torsion angles adopt physically realistic conformations (Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc*. 2002, *124*, 2866-2867)

**Table S2**. Potential terms used in the simulated annealing docking protocol.

| Potential term | force constant | | |
|---|---|---|---|
| | initial minimization | simulated annealing | final minimization |
| diffusion tensor[a] | 100 | 100 | 100 |
| NOE (kcal.mol$^{-1}$.Å$^{-2}$)[b,c] | 0.3 | 0.3→60 | 60 |
| repulsive vdw (kcal.mol$^{-1}$.Å$^{-4}$) | 0.01 | 0.004→4 | 4 |
| hydrophobic contact potential | 0 | 1→50 | 50 |
| $R_{gyr}$ potential (kcal.mol$^{-1}$.Å$^{-2}$)[c] | 0 | 0.05→500 | 500 |
| torsion angle database potential | 0.002 | 0.002→1 | 1 |
| $C_2$ symmetry (kcal.mol$^{-1}$.Å$^{-2}$)[d] | 10 | 10 | 10 |

[a]The values shown represent the scale factor $s_{diff}$; the actual force constant, $k_{diff}$, for the diffusion tensor restraints is obtained by multiplying $s_{diff}$ by 200x($N$/855) where $N$ is the number of atoms.

[b]The NOE potential is used for the highly ambiguous $\Sigma(r^{-6})^{-1/6}$ distance restraints derived from chemical shift perturbation mapping (see Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2003**, *125*, 2902-2912).

[c]Used only for the EIN-HPr docking calculations.

[d]$C_2$ symmetry restraints were *only* used for the HIV-1 protease docking calculations, and are only relevant to homodimers.

## 2. Experimentally derived diffusion tensors used as target parameters in simulated annealing.

### 2.1 Definitions

The diffusion tensor can be represented by the following terms:

Overall rotational correlation time: $\tau_c = 0.5(D_x + D_y + D_z)^{-1}$

Anisotropy: $A = 2D_z/(D_x + D_y)$

Rhombicity: $\eta = 1.5(D_y - D_x)/[D_z - 0.5(D_x + D_y)]$

The expressions for $A$ and $\eta$ assume $D_x \le D_y \le D_z$. The components of the diffusion tensors presented below in matrix form and their eigenvalues, $D_x$, $D_y$ and $D_z$, are in units of [$10^{-7}$/s].

For diffusion tensors calculated from NMR relaxation we also report the values of the normalized $\chi^2$ function $\chi^2/D_f$ where $D_f$ is the number of degrees of freedom. $\chi^2 = \sum_i (\rho_i^{\exp} - \rho_i^{calc})^2 / {\sigma_i}^2$ where $\rho_i^{\exp}$ and $\rho_i^{calc}$ are the experimental and calculated $^{15}$N $R_2/R_1$ ratios, respectively; $\sigma_i$ are the experimental errors; and index $i$ enumerates amino acids in the protein sequence.

**2.2 EIN**

The experimental NMR relaxation rates are taken from Tjandra et al. (Tjandra, N.; Garrett, D. S.; Gronenborn, A. M.; Bax, A.; Clore, G. M. *Nature Struct. Biol.* **1997**, *4*, 443-449).

Using an axially symmetric model for the diffusion tensor we obtain:

$D_x = 1.33 \quad D_y = 1.33 \quad D_z = 2.0$

$\tau_c = 10.6$ ns; $A = 1.54$; $\eta = 0$; $\chi^2/D_f = 72.8$; $D_f = 113$

The components of the diffusion tensor in the molecular frame of the 1EZA structure that were used as target parameters for simulated annealing refinement are:

| | *x* | *y* | *z* |
|---|---|---|---|
| *x* | 2.0434 | -0.0764 | -0.0421 |
| *y* | -0.0764 | 1.3403 | 0.0045 |
| *z* | -0.0421 | 0.0045 | 1.3346 |

$\chi^2/D_f$ for a fully asymmetric model of the diffusion tensor is 73.9, larger than that for the axially symmetric model, thus justifying the use of the latter in the simulated annealing calculations.

**2.3 HIV-1 protease**

To derive the components of the rotational diffusion tensor for HIV-1 protease we used the experimental data recorded at a spectrometer frequency of 600 MHz from Tjandra et al. (Tjandra, N.; Wingfield, P.; Stahl, S.; Bax, A. *J. Biomol. NMR* **1996**, *8*, 273-284). Because HIV-1 protease is a homodimer of identical subunits, only a single set of $R_1$ and $R_2$ data are observed for the dimer. These comprised 78 pairs of $R_1$ and $R_2$ data; eliminating data for all residues with low $^1$H-{$^{15}$N} heteronuclear NOE values, suspected conformational exchange, chemical shift overlap in the spectrum, and weak cross-peak intensities, left 64 pairs of $R_1$ and $R_2$ data points for analysis (Tjandra et al.).

To mimic a real situation where the true subunit arrangement is unknown, we made use of the coordinates of only one subunit to process the relaxation data and obtain the rotational diffusion tensor. In the 2NPH crystal structure, the coordinates of the two subunits are not identical. The fitting results for the two subunits, however, are virtually identical with a difference in $\chi^2$ of less than 2%. Since the first subunit gave fits with the lower $\chi^2$, we used the diffusion tensor calculated from these coordinates.

Both axially symmetric and fully anisotropic models of the rotational diffusion tensor were tested and yielded $\chi^2$ values of 1.45 and 1.34, respectively. The difference is not statistically significant as judged by the *F*-test, and therefore does not justify the use of the fully anisotropic model. Thus, for the docking calculations, we used the diffusion tensor for the axially symmetric model.

Using an axially symmetric model for the diffusion tensor we obtain:

$D_x = 1.37 \quad D_y = 1.37 \quad D_z = 1.92$

$\tau_c = 10.71$ ns; $A = 1.40$; $\eta = 0$; $\chi^2/D_f = 1.45$; $D_f = 60$

The components of the diffusion tensor in the molecular frame of the first subunit of the 2NPH crystal structure that were used as target parameters for simulated annealing refinement are:

| | *x* | *y* | *z* |
|---|---|---|---|
| *x* | 1.4467 | 0.0616 | 0.1744 |
| *y* | 0.0616 | 1.4269 | 0.1487 |
| *z* | 0.1744 | 0.1487 | 1.7955 |

**2.4 The EIN-HPr complex**

The experimental $^{15}$N relaxation data were recorded as described in the Experimental Methods section of the main paper. The diffusion tensor which was extracted from experimental data for both components of the complex using an axially symmetric model for the diffusion tensor, is as follows:

$D_x = 0.87 \quad D_y = 0.87 \quad D_z = 1.19$

$\tau_c = 17.02$ ns; $A = 1.36$; $\eta = 0$; $\chi^2/D_f = 32.7$; $D_f = 132$

The components of the diffusion tensor in the molecular frame of the reference structure that were used as target parameters for the simulated annealing docking calculations are:

| | *x* | *y* | *z* |
|---|---|---|---|
| *x* | 1.0215 | -0.0797 | 0.1344 |
| *y* | -0.0797 | 0.9182 | -0.0730 |
| *z* | 0.1344 | -0.0730 | 0.9981 |

[Note the reference structure was generated by fitting the X-ray coordinates of free EIN (1ZYM) and HPr (1POH) onto the NMR coordinates of the EIN-HPr complex (3EZA).] The small improvement in $\chi^2/D_f$ (28.3) for a fully asymmetric model of the diffusion tensor is not statistically significant as judged by the *F*-test (Snedecor, G. W.; Cochran, W. G. *Statistical methods*; 8th ed.; Iowa State University Press: Ames, 1989). The expected ratio of the normalized $\chi^2$ values for the axially symmetric to fully asymmetric models at the 5% confidence level is 1.34 compared to an observed value of only 1.15.

## 3. Validation of the refined EIN structure with SAXS data.

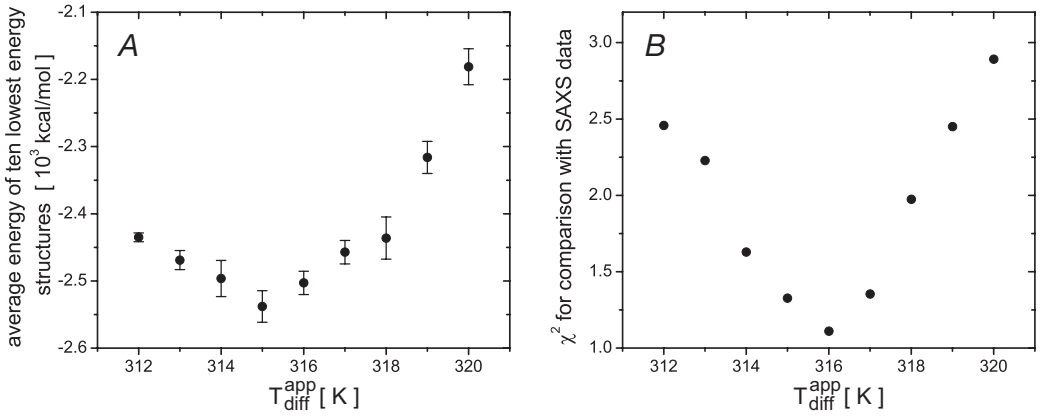


**Figure S1.** Results of several runs of EIN refinement with different values of $T_{diff}^{app}$ using a value of 5 for the diffusion tensor scale factor $s_{diff}$ The results are very similar to those shown in Fig. 1 of the main text obtained with a value of $s_{diff}$ = 1. (A) Total Xplor-NIH energies, averaged over the ten lowest energy structures (errors bars, 1 s.d). (B) $\chi^2$-fit of the calculated structures to the SAXS data. The structures used for fitting the SAXS data are the restrained regularized averages over the ten lowest energy structures calculated at each value of $T_{diff}^{app}$ .

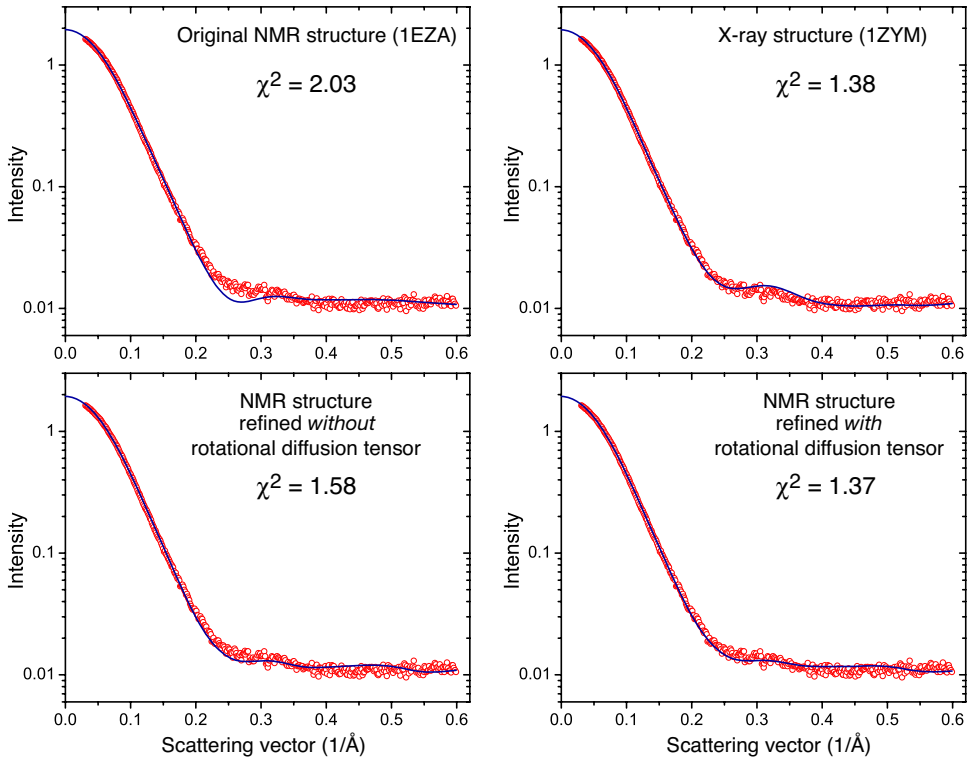


**Figure S2.** Comparison of experimental (red dots) and calculated (blue lines) SAXS profiles obtained with the program CRYSOL.

**Table S3.** Best-fit parameters for various EIN structures obtained with CRYSOL version 2.5 (Svergun, D.; Barberato, C.; Koch, M. H. *J. Appl. Crystallogr.* **1995**, *28*, 768-773).

| Structure | contrast of the hydration shell (e/Å$^3$) | average displaced solvent radius (Å) | radius of gyration experimental / theoretical (Å) | total displaced solvent volume (Å$^3$) | $\chi^2$ |
|---|---|---|---|---|---|
| NMR (1EZA)[a] | 0.013 | 1.80 | 20.90 / 21.62 | 33714 | 2.03 |
| X-ray (1ZYM) | 0.022 | 1.68 | 20.93 / 21.48 | 35579 | 1.38 |
| refinement *without* diffusion tensor[b] | 0.018 | 1.78 | 20.92 / 21.47 | 33714 | 1.58 |
| refinement *with* diffusion tensor[b] | 0.015 | 1.76 | 20.92 / 21.45 | 33543 | 1.37 |

[a]Residues 250-259 of 1EZA were deleted since the present construct consisted of residues 1-249 of enzyme I.
[b]Restrained regularized average structure derived from the 10 lowest energy structures.

## 4. Dependence of various parameters on $T_{diff}^{app}$ in the docking calculations.

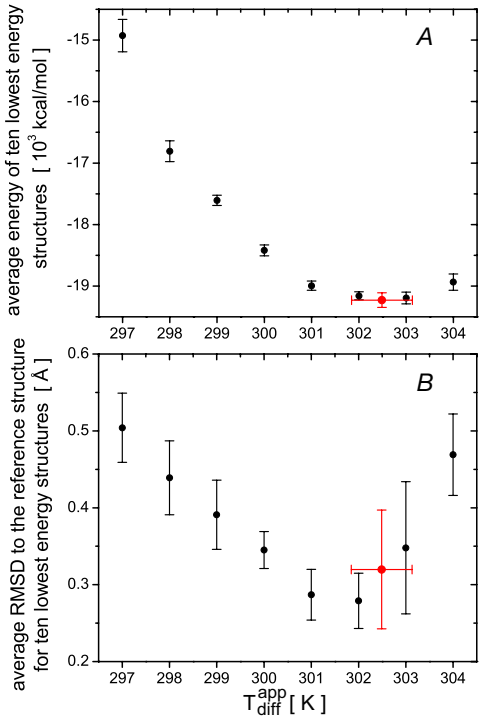


**Figure S3.** Dependence of (A) the average Xplor-NIH energy and (B) the averaged Cα atomic rms difference to the reference structure on $T_{diff}^{app}$ for the HIV-1 protease docking calculations starting from the NMR and X-ray subunit coordinates. The values are averaged over the ten lowest energy structures (error bars, 1 s.d.). The black symbols correspond to the structures obtained using a grid search for $T_{diff}^{app}$; the red symbols are the results obtained with automated optimization of $T_{diff}^{app}$.