Published in final edited form as:
*J Biomol NMR*. 2013 December ; 57(4): 333–352. doi:10.1007/s10858-013-9791-1.

# Deriving Quantitative Dynamics Information for Proteins and RNAs using ROTDIF with a Graphical User Interface

**Konstantin Berlin**,
Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA, Tel.: 1-301-405-3165 (TKD), 1-301-405-3461 (DF) Fax.: 1-301-314-0386. Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

**Andrew Longhini**,
Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA, Tel.: 1-301-405-3165 (TKD), 1-301-405-3461 (DF) Fax.: 1-301-314-0386

**T. Kwaku Dayie**, and
Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA, Tel.: 1-301-405-3165 (TKD), 1-301-405-3461 (DF) Fax.: 1-301-314-0386

**David Fushman**
Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA, Tel.: 1-301-405-3165 (TKD), 1-301-405-3461 (DF) Fax.: 1-301-314-0386. Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

## Abstract

To facilitate rigorous analysis of molecular motions in proteins, DNA, and RNA, we present a new version of ROTDIF, a program for determining the overall rotational diffusion tensor from single-or multiple-field Nuclear Magnetic Resonance (NMR) relaxation data. We introduce four major features that expand the program's versatility and usability. The first feature is the ability to analyze, separately or together, $^{13}$C and/or $^{15}$N relaxation data collected at a single or multiple fields. A significant improvement in the accuracy compared to direct analysis of $R_2/R_1$ ratios, especially critical for analysis of $^{13}$C relaxation data, is achieved by subtracting high-frequency contributions to relaxation rates. The second new feature is an improved method for computing the rotational diffusion tensor in the presence of biased errors, such as large conformational exchange contributions, that significantly enhances the accuracy of the computation. The third new feature is the integration of the domain alignment and docking module for relaxation-based structure determination of multi-domain systems. Finally, to improve accessibility to all the program features, we introduced a graphical user interface (GUI) that simplifies and speeds up the analysis of the data. Written in Java, the new ROTDIF can run on virtually any computer platform. In addition, the new ROTDIF achieves an order of magnitude speedup over the previous version by implementing a more efficient deterministic minimization algorithm. We not only demonstrate the improvement in accuracy and speed of the new algorithm for synthetic and experimental $^{13}$C and $^{15}$N relaxation data for several proteins and nucleic acids, but also show that careful analysis

7 Supporting Information
Spectral density functions corresponding to various models for the overall rotational diffusion tensor and detailed description of the corresponding optimization algorithms, CSA values for $^{15}$N and $^{15}$C in various bonds in proteins and RNA.

required especially for characterizing RNA dynamics allowed us to uncover subtle conformational changes in RNA as a function of temperature that were opaque to previous analysis.

**Keywords**

rotational diffusion tensor; ELM; ELMDOCK; ROTDIF

## 1 Introduction

Proteins and nucleic acids are two biopolymers at the center of numerous cellular functions that can adopt intricate three-dimensional architectures, and for the most part remodel those pliant structures to accommodate ligands and other signaling cues (Mittermaier and Kay, 2006; Peng, 2012; Dayie, 2013; Dethoff et al, 2012). They can remodel their three-dimensional architectures either by induced fit or conformational capture or both (Pozzi et al, 2012). In the former case the nature of the binding event drives the free state to the bound structure, whereas in the latter case the ligand binding selects or captures the competent conformation out of a population of pre-existing conformations (Leulliot and Varani, 2001). Given the primal roles that structure and dynamics play, analysis of the molecular motions of these biopolymers is critical for advancing our understanding of the interplay of structure, dynamics, and function. At the heart of those types of analyses is the ability to accurately quantify the overall rotational diffusion tensor from NMR relaxation data. (Fushman et al, 1994; Bruschweiler et al, 1995; Mandel et al, 1995; Tjandra et al, 1995; Peng and Wagner, 1995; Tjandra et al, 1997; Fushman et al, 1997; Akke et al, 1997; Fushman and Cowburn, 1998b; Fushman et al, 1999b; Hall and Fushman, 2003; Fushman, 2012). Recently progress has been made on a number of fronts: (i) the improved ability to refine the structures of complexes and quantify molecular motion; (ii) the use of the diffusion tensor as a long-range orientational restraint for the structural characterization of multidomain systems (Fushman et al, 1999b; Ghose et al, 2001; Fushman and Cowburn, 2002; Fushman et al, 2004; Ryabov and Fushman, 2006, 2007a) and for analysis of dynamics of mostly the backbone of proteins (Hall and Fushman, 2003; Fushman et al, 2004; Hall and Fushman, 2006), and rarely for nucleic acid bases and ribose moieties (Akke et al, 1997; Legault et al, 1998; Hoogstraten et al, 2000; Dayie et al, 2002; Boisbouvier et al, 2003; Duchardt and Schwalbe, 2005; Eldho and Dayie, 2007) and (iii) introduction of rotational diffusion tensor as a translational restraint in rigid-body docking of multi-domain proteins and protein-protein complexes (Ryabov and Fushman, 2007b; Ryabov et al, 2010; Berlin et al, 2011).

All these approaches depend on fast and accurate determination of the overall rotational diffusion tensor. To directly obtain the fully anisotropic rotational diffusion tensor from spin-relaxation rates using the model-free approach(Lipari and Szabo, 1982) data at multiple fields need to be collected(Fushman and Cowburn, 2001) and subsequently a nonlinear, non-convex least-squares analysis performed (d'Auvergne and Gooley, 2008). The difficulty of obtaining experimental relaxation data at several fields, the algorithmic complexity of the computation, as well as uncertainties in obtaining very accurate measures of both the magnitudes and orientation of the chemical shift anisotropy (CSA) tensors have all limited the adaptation of the fully anisotropic diffusion tensor for structure analysis.

To obtain accurate estimate of the fully anisotropic rotational diffusion tensor from spin-relaxation measurements, the analysis is usually limited to those regions of the macromolecules that are structured. There the local motions are characterized by relatively high order parameters ($S^2 \approx 1$) and fast internal correlation times ( $\tau_{loc} \leq 100$ ps). In this case it is possible to extract the rotational diffusion tensor from ratios of relaxation rates (e.g., $R_2/R_1$) instead of the relaxation data directly (Tjandra et al, 1995; Fushman et al, 1999b),

because the relaxation-rates ratio cancels out the unknown factors such as the dipolar and CSA terms and order parameters (Fushman et al, 1999b; Fushman and Cowburn, 2002; Fushman et al, 2004), thus allowing a direct determination of the rotational diffusion tensor without the need for any additional site-specific characteristics except for the orientation of the relaxation-relevant internuclear vectors (see e.g., (Fushman et al, 1999b)). Several publicly available packages have been developed to derive the anisotropic rotational diffusion tensor from $^{15}$N relaxation rate ratios, including TENSOR (Dosset et al, 2000), DIFFTENS (Ghose et al, 2001), ROTDIF (Walker et al, 2004), and a $^{13}$C adaptation of the latter program, ROTDIF_RNA (Eldho and Dayie, 2007).

Here we introduce a new, redesigned version of the ROTDIF software with numerous features that significantly improve and expand the functionality and performance of the program. The new version, ROTDIF 3, allows analysis of $^{15}$N and $^{13}$C relaxation data (separately or together) at a single or at multiple magnetic fields, and significantly improves the computational performance of the previous version by introducing a new multi-start convex optimization algorithm. ROTDIF 3 is an order of magnitude faster than the previous version, and includes an option for robust regression that increases the accuracy of computation in the presence of outliers, such as those arising from conformational exchange contributions, highly flexible residues, and other contributions not captured by the models used in the existing software.

As part of the current package, we integrate the new ROTDIF module with the *ab initio* diffusion tensor predictor ELM (Ryabov et al, 2006) and the alignment and translational docking modules developed in ELMDOCK (Berlin et al, 2011). The updated domain-alignment method extends the previous eigenvector-based domain-alignment approach by now computing the globally optimal orientational alignment(Fushman et al, 1999b, 2004; Berlin et al, 2011). The extended alignment approach yields improved solutions when the principal values of the anisotropic rotational diffusion tensors measured for the two domains are not identical. These new modules are designed for quantitative analysis and interpretation of relaxation data in terms of structural change.

All the modules are tightly integrated into a Graphical User Interface (GUI), which replaces the previously developed (MATLAB) command-line interface with a more intuitive visual interface. Users can now rapidly compute, analyze, and refine their diffusion tensor results, as well as instantly compute an aligned and docked structure of a two-domain system. Importantly, the new ROTDIF package (and the associated ELM and ELMDOCK modules) is written in Java, runs on any system with a Java 6+ virtual machine, and requires no installation or any adjustable parameters.

Finally, we apply the new package to synthetic data as well as published relaxation data for two proteins (GB3 and ubiquitin) and several nucleic acids (a Dikerson DNA dodecamer, a fragment of RNA enzyme (D5), and UUCG tetraloop capped RNA element). We show that careful analysis of relaxation data, especially for nucleic acids, is key for making meaningful conclusions about macromolecular structure and function.

## 2 Method

The rotational diffusion tensor **D** is a symmetric positive definite 3 × 3 matrix that characterizes the (generally) anisotropic overall random reorientation (tumbling) of a molecule in a solvent (Woessner, 1962; Bruschweiler et al, 1995). Anisotropy applies when the tumbling rates around various directions in a molecule are different. We label the sorted eigenvalues (principal components) of **D** as $D_x \leq D_y \leq D_z$, and define the orientation of the

tensor through the associated rotation matrix of a sorted eigendecomposition. We also define the overall rotational correlation time as $\tau_c = 1/[2 \times \mathrm{tr}(\mathbf{D})]$, where tr(**D**) is the trace of **D**.

The overall molecular tumbling causes spin relaxation of a nucleus P by modulating various interactions, including the interaction with the external magnetic field and dipolar couplings with other nuclei. For an isolated pair of spin-1/2 nuclei P and Q (where, e.g., P is $^{15}$N or $^{13}$C and Q is $^{1}$H), the rates of longitudinal ($R_1$) and transverse ($R_2$) spin-relaxation and the steady-state nuclear Overhauser enhancement (NOE) of nucleus P are related to the rotational diffusion tensor of the molecule via the following equations (e.g. (Dayie et al, 1996; Fushman and Cowburn, 2001)).

$$\begin{aligned} R_1 &= 3(d^2+c^2)J(\omega_P) + d^2[J(\omega_Q-\omega_P) + 6J(\omega_Q+\omega_P)], \\ R_2 &= \tfrac{1}{2}(d^2+c^2)[4J(0) + 3J(\omega_P)] + \tfrac{1}{2}d^2[J(\omega_Q-\omega_P) + 6J(\omega_Q+\omega_P) + 6J(\omega_Q)] + R_{ex}, \\ R_3 &= 1 + \frac{\gamma_Q}{\gamma_P} d^2[6J(\omega_Q+\omega_P) - J(\omega_Q-\omega_P)]/R_1, \end{aligned} \quad (1)$$

where $J(\omega) = J(\omega, \mathbf{v}, \mathbf{D})$ is the spectral density of reorientational motion for the PQ bond, **v** is the bond's orientation, $\omega$ is the Larmor precession frequency, $R_{ex}$ is the conformational exchange contribution to $R_2$, $d = -\mu_0\gamma_P\,\gamma_Q h/(16\pi^2 r^3)$ and $c = -\omega_P$ CSA/3 are constants representing the magnitude of the dipolar and chemical shift anisotropy (CSA) interactions, $r$ is the length of the PQ bond, $h$ is the Planck's constant, and $\mu_0$ is the vacuum permeability. Here we refer to the steady state NOE as $R_3$ throughout the manuscript, and define it as $R_3 = I_{sat}/I_{eq}$ where $I_{sat}$ and $I_{eq}$ are signal intensities of nucleus P measured when the nucleus Q is in the saturated and in the equilibrium states, respectively. The equations for $R_1$ and $R_2$ assume that the chemical shift tensor of nucleus P is axially symmetric and approximately oriented along the PQ bond (see (Fushman and Cowburn, 1999) for corrections when this assumption is not valid). We assume the model-free or extended model-free forms of the spectral density functions $J(\omega)$ (Lipari and Szabo, 1982; Clore et al, 1990). Expressions for $J(\omega)$ for the isotropic, axially-symmetric, and the fully anisotropic diffusion models are given in Supporting Information. Since $J(\omega)$'s dependence on the diffusion tensor is nonlinear, solving Eq. (1) becomes a nonlinear optimization problem.

Given $R_1$, $R_2$, and $R_3$ at several fields, Eq. (1) can potentially be solved for the overall rotational diffusion tensor (together with the microdynamic parameters, e.g. $S^2$ and $\tau_{loc}$, and the dipolar and CSA terms) without making any assumptions other than the Lipari-Szabo model for $J(\omega)$ (Lipari and Szabo, 1982; Fushman and Cowburn, 2001). However, when $R_1$, $R_2$, and $R_3$ are only available at a single field the general solution of Eq. (1) for **D** is ill-posed (i.e. there are multiple solutions, as the number of unknowns exceeds the number of available experimental parameters), and hence impossible to solve without making *a priori* assumptions about some of the variables (see e.g. (Fushman and Cowburn, 2001)).

In our approach we introduce an *a priori* restraint by limiting our input to only those bonds that are in the structurally well-defined ("rigid") parts of the molecule and were $R_{ex} \ll R_2$. It is possible to justify these assumptions, for example, by selecting bonds from secondary structure elements and assessing $R_{ex}$ contributions from comparison of the transverse auto- and cross-correlation rates or from analysis of relaxation data at multiple fields (Fushman and Cowburn, 1998a; Fushman et al, 1999a; Kroenke et al, 1998; Fushman and Cowburn, 2001).

Our analysis focuses on the ratio $\rho$ of spectral density components at $\omega = 0$ and $\omega = \omega_P$. Based on Eq. (1) and assuming that $R_{ex} = 0$, $\rho$ is directly related to the modified ratio of spin-relaxation rates

$$\rho \equiv \frac{4}{3}\frac{J(0)}{J(\omega_P)} \approx \frac{2R_2 - R_1 - HF_2}{R_1 - HF_1}, \quad (2)$$

in which we subtracted the contributions from high-frequency components of the spectral density, $HF_1$ and $HF_2$, defined as

$$\begin{aligned} HF_1 &= d^2[J(\omega_Q - \omega_P) + 6J(\omega_Q + \omega_P)], \\ HF_2 &= d^2 6J(\omega_Q). \end{aligned} \quad (3)$$

As shown earlier (Fushman et al, 1999b; Fushman and Cowburn, 2001, 2002), for "rigid" PQ bonds the reformulation of the relaxation rates via a single $\rho$ value allows one to quantify spin-relaxation parameters for each individual PQ pair via a single value that depends only on $\omega$, **v**, and **D** and is independent of the CSA, dipolar term, and order parameter values. This approach, therefore, reduces significantly the number of model assumptions needed for forward prediction of the relaxation rates. Note that the definition of $\rho$ in our new version is the inverse of the definition in the previous ROTDIF.

Based on the above definitions, we can approximate the high-frequency contributions as

$$\begin{aligned} HF_1 &\approx w_1 f(R_1, R_3), \\ HF_2 &\approx w_2 f(R_1, R_3), \end{aligned} \quad (4)$$

where

$$f(R_1, R_3) = \frac{\gamma_P}{\gamma_Q} R_1(R_3 - 1) = d^2[6J(\omega_Q + \omega_P) - J(\omega_Q - \omega_P)], \quad (5)$$

and the coefficients $w_1$ and $w_2$ are estimated(Fushman and Cowburn, 1998a; Fushman et al, 1999a; Eldho and Dayie, 2007) using reduced spectral density approach (Ishima and Nagayama, 1995; Farrow et al, 1995; Peng and Wagner, 1995), and are usually assumed to be constant. For the typical values of the overall rotational correlation time $\tau_c$ of macromolecules ($\tau_c > 4$ ns) the values of $\omega_Q$, $\omega_Q \pm \omega_P$ are at the high-frequency tail of $J(\omega)$ ($\tau_c\omega \gg 1$); therefore the coefficients $w_i$ are nearly independent of $\tau_c$, if $S^2 \approx 1$. However, in our algorithm, instead of using pre-defined constant $w_1$, $w_2$ values, as has been previously done(Fushman and Cowburn, 1998a; Fushman et al, 1999a; Eldho and Dayie, 2007), we compute $w_1$ and $w_2$ numerically for each bond (see next section).

Given the above definitions, the experimental ratio can be estimated from the measured relaxation rates as

$$\rho^{exp} \approx \frac{2R_2 - R_1 - w_2 f(R_1, R_3)}{R_1 - w_1 f(R_1, R_3)}, \quad (6)$$

and the predicted ratio can be simply computed as

$$\rho^{pred} = \frac{4}{3}\frac{J(0)}{J(\omega_P)}. \quad (7)$$

The problem of finding the rotational diffusion tensor can now be expressed as a least-squares optimization problem:

$$\mathbf{D}_{exp}=\arg\min_{\mathbf{D}}\chi^2(\mathbf{v},\mathbf{D}), \quad (8)$$

where

$$\chi^2(\mathbf{v},\mathbf{D})=\sum_{i=1}^{N}\left[\frac{\rho_i^{pred}(\mathbf{v}^i,\mathbf{D})-\rho_i^{exp}}{\sigma_i}\right]^2, \quad (9)$$

$N$ is the number of PQ bonds in the molecule, $\mathbf{v}^i=\left[v_x^i,v_y^i,v_z^i\right]$ is the unit vector in the direction of the $i$-th PQ bond, $\rho_i^{exp}$ is the ratio, given by Eq. (6), of experimentally measured transverse and longitudinal spin-relaxation rates (with high-frequency correction) for nucleus P in the PQ bond $i$, $\rho_i^{pred}(\mathbf{v}^i,\mathbf{D})$ is given by Eq. (7), and $\sigma_i$ is the experimental error in $\rho_i^{exp}$. By selecting only relatively rigid bonds for the analysis we assured that $\rho^{pred}$ depends only on **D** and known values of **v** and $\omega$ (Fushman et al, 1999b; Fushman and Cowburn, 2002), thus making the new formulation well-posed.

### 2.1 Initial Estimate of Parameters

Before performing global minimization of $\chi^2$ to determine the overall rotational diffusion tensor, we directly fit the relaxation data, separately for each bond, in order to estimate bond-specific $\tau_c$, CSA, $S^2$, and $R_{ex}$. Here $\tau_c$ has the meaning of an effective (isotropic) overall rotational correlation time. These values are not directly used in the computation, but are instead used to provide warnings for users when the estimated $S^2$ values are too low, or $R_{ex}$ is too high. If $R_3$ is not provided, $R_{ex}$ is assumed to be 0, $S^2$ is set to 0.9, and only $\tau_c$ is estimated. When $R_3$ is provided, both $\tau_c$ and CSA values are estimated. If relaxation data at only one field are available, we set $R_{ex}$ = 0 and $S^2$ = 0.9. For relaxation data at multiple fields all four variables are computed. Initial estimates for CSA values are taken from (Stueber and Grant, 2002; Bryce et al, 2005; Ying et al, 2006; Hansen and Al-Hashimi, 2006; Fushman and Cowburn, 2001) (see Table S1), the CSA values are allowed to vary by 40 ppm, if they are fitted. The bond lengths are assumed to be 1.02 Å and 1.09 Å, respectively, for NH and CH bonds (Case, 1999; Ferner et al, 2008).

### 2.2 Subtracting High-Frequency Contributions from Relaxation Rates

Accurate subtraction of the high-frequency contributions from the experimental rates in Eq. (2) is fundamental to our $\chi^2$ minimization method. The accuracy of the subtraction depends on finding the proper $w_1$ and $w_2$ values. We provide two methods for computing $w_1$ and $w_2$ depending on whether the $R_3$ data are available or not.

For the first method, when $R_3$ data are available, the $w_1$ and $w_2$ are approximated from Eq. (3), Eq. (4), and Eq. (5), by solving for $w_1$ and $w_2$ the following equations

$$\begin{aligned} J(\omega_Q-\omega_P)+6J(\omega_Q+\omega_P)&=w_1[6J(\omega_Q+\omega_P)-J(\omega_Q-\omega_P)],\\ 6J(\omega_Q)&=w_2[6J(\omega_Q+\omega_P)-J(\omega_Q-\omega_P)], \end{aligned} \quad (10)$$

based on the theoretical definition of $J(\omega)$ (see Supporting Information) and using our initial estimate for $\tau_c$. The values of $w_1$ and $w_2$ vary only slightly for different $\tau_c$ values (as $S^2$

cancels out in the equations), providing a highly stable estimate for the constants, even when our $\tau_c$ estimate significantly deviates from the true value.

For the second method, when $R_3$ data are not available, we approximate the $R_3$ value from Eq. (1) using the experimental $R_1$ value and computing the spectral densities using our estimate for $\tau_c$. Since the initial $\tau_c$ value estimated from just one bond can be unreliable, we refine the $w_1$, $w_2$, and $\rho^{exp}$ by solving Eq. (8) for $\mathbf{D}_{exp}$ using the initial $\rho$ estimates, and then updating each $\rho^{exp}$ based on the more stable estimate of $\tau_c$. These refined $\rho^{exp}$ values are then used to solve Eq. (8) for the final value of $\mathbf{D}_{exp}$.

Fig. 1 shows that we are able to accurately compute $\rho$ from synthetically generated $^{15}$N or $^{13}$C relaxation data by subtracting high-frequency contributions using Eq. (6) and Eq. (10). The accuracy is measured by the relative error of the recovered $\tau_c$ that we extract from the computed relaxation ratio $\rho$. The synthetic relaxation data were created from randomly sampled inputs of $\tau_c$ = [5, 20] ns, $S^2$ = [0.8, 1], $\tau_{loc}$ = [0.001, 0.1] ns, and a range of CSA and $r$ values of [120, 220] ppm and [0.98, 1.05] Å for N-H, [5, 60] ppm and [1.07, 1.10] Å for C1′-H1′, and [150, 230] ppm and [1.07, 1.10] Å for C6-H6, assuming isotropic overall rotational diffusion. The spectrometer frequency was set to 600 MHz, 3% Gaussian error was added to the $R_1$ and $R_2$ data and 5% error was added to $R_3$. In addition, we also evaluated the accuracy of the high-frequency contribution subtraction approach in the case when $R_3$ is unknown, and of the direct $R_2/R_1$ ratio approach, computed using the assumption of $S^2$ = 1. Note that by including C1′-H1′ and C6-H6 in our results, we tested our method for both low and high $^{13}$C CSA values.

The results in Fig. 1 show that for majority of the input $^{13}$C data the errors in $\tau_c$ are smaller than 6% (< 2% for $^{15}$N). Furthermore, high-frequency-contribution subtraction provides a significant improvement in accuracy in the $^{13}$C case, compared to the direct $R_2/R_1$ method, while removing the dependence on CSA, dipolar coupling, $S^2$, and providing a significant reduction in the computational complexity. The values of $w_1$ and $w_2$ fluctuate by about 0.01% for $^{15}$N and 0.1% for $^{13}$C, for the synthetic data ranges used here (see above).

### 2.3 Algorithms for Computing the Rotational Diffusion Tensor

Having derived $\rho^{exp}$, we now present algorithms for solving Eq. (8). As mentioned above, there are three rotational diffusion tensor models that are commonly used to model $J(\omega)$. The most general is the *fully anisotropic* model, where all three eigenvalues of the rotational diffusion tensor are unique. In the case when two eigenvalues of the tensor are equal, the fully anisotropic model can be simplified to an *axially-symmetric* model. Finally, when all three eigenvalues are equal, a simple *isotropic* model is used. The axially-symmetric and isotropic models can be directly derived from the fully anisotropic model. See Supporting Information for the definition of $J(\omega)$ for the three diffusion tensor models and their derivations.

First, we note that for the isotropic diffusion tensor model, Eq. (8) can be analytically solved for **D** (see Supporting Information),

$$\tau_c = \left| \frac{1}{\omega_P} \sqrt{\frac{3}{4}\rho^{exp} - 1} \right|, \quad (11)$$

where $D_{iso} = 1/(6\tau_c)$ is the eigenvalue of **D**. This solution will serve as an initial estimate for the eigen-values of the axially-symmetric and the fully anisotropic models. The isotropic-model algorithm is given in Supporting Information.

We then use the isotropic solution as an initial guess for the axially-symmetric $\mathbf{D}_{exp}$. This eliminates the need for a user-defined range of eigenvalues, required in the previous ROTDIF version (Walker et al, 2004), and at the same time speeds up convergence of the iterative minimization. **V**, the orthonormal matrix of the eigenvectors of the diffusion tensor, can be expressed using three Euler angles $\alpha$, $\beta$, and $\gamma$. Since two of the eigenvalues are equal in the case of the axially-symmetric model, the orientation of the diffusion tensor can be described by the orientation of the unique eigenvalue, $D_{\|}$. Therefore, we can express this orientation using only $\alpha$ and $\beta$ angles and set $\gamma = 0$.

Due to the periodicity of the diffusion tensor in spherical coordinate space (see e.g. (Ghose et al, 2001)), Eq. (9) is $\pi$ periodic in the two Euler angles (Walker et al, 2004). We take a similar approach to minimizing our $\chi^2$ as before (Walker et al, 2004), but instead of randomly sampling a large number of angles for initial guesses to the nonlinear least-squares solver, we only make four initial guesses for $\alpha$ and $\beta$: [0, 0], [0, $\pi$/2], [$\pi$/2, 0], and [$\pi$/2, $\pi$/2]. Additionally, we alternate between the last and the first two eigenvalues being equal to handle the prolate ($Dx \approx Dy < Dz$) and the oblate ($Dx < Dy \approx Dz$) cases. The alternating is needed when using an axially-symmetric diffusion tensor model to analyze relaxation data for a molecule with full rotational anisotropy, since both a prolate and an oblate tensor solution exist in this case.(Blackledge et al, 1998) We therefore perform nonlinear least-squares minimization for eight initial guesses only. The complete algorithm is shown in Supporting Information. By applying the algorithm to real and randomly generated synthetic data, we empirically confirmed that our method was able to find the true minimizer every time.

Finally, we solve for the fully anisotropic diffusion tensor $\mathbf{D}_{exp}$. Again, we use the solution from the isotropic model as an initial guess. We make an observation, similar to that for the axially-symmetric case, that $\rho^{pred}$ is $\pi$ periodic for $\alpha$, $\beta$, and $\gamma$. We therefore take eight initial guesses for the Euler angles: [0, 0, 0], [0, 0, $\pi$/2], [0, $\pi$/2, 0], [0, $\pi$/2, $\pi$/2], [$\pi$/2, $\pi$/2, 0], and [$\pi$/2, $\pi$/2, $\pi$/2]. The complete algorithm is provided in Supporting Information. Again, by applying the algorithm to real and randomly generated synthetic data, we empirically confirmed that our method was able to find the true minimizer every time.

### 2.4 Robust Least-Squares

In Eq. (8) we defined the experimental diffusion tensor as the minimizer of $\chi^2$. In expressing our problem as the minimization of the sum of squares of weighted residuals we made an implicit assumption that the weighted prediction model (values of $\rho_i^{pred}/\sigma_i$) is an unbiased estimator. However, when equating $\rho^{exp}$ to $\rho^{pred}$ in Eq. (2) we assumed that $R_{ex}$ is negligible. For some sites in a molecule $R_{ex}$ can be a significant contributor to the $R_2$ relaxation rate. Since $R_{ex}$ is always positive, our $\rho^{pred}$ values are biased estimators of $\rho^{exp}$, violating the basic assumption of $\chi^2$-minimization. We therefore introduce a robust regression method for estimating the experimental diffusion tensor that should partially compensate for this bias. The method is deterministic and makes an assumption that it is possible to detect major outliers based on the initial solution, given by equation Eq. (8).

We redefine our target function $\chi^2$ by replacing it with the outlier-damping version, $\chi_T^2$, such that

$$\chi_T^2(\mathbf{v},\mathbf{D}) \sum_{i=1}^{n} \min\left( [\Delta + \ln(|\delta_i - \Delta| + 1)]^2, \delta_i^2 \right), \quad \delta_i = \frac{\rho_i^{pred}(\mathbf{v}^i,\mathbf{D}) - \rho_i^{exp}}{\sigma_i}, \tag{12}$$

where $\Delta = 3.0$. Setting $\Delta = 3.0$ significantly dampens the contribution of residuals that are greater than $3\sigma_i$ (above 99.7 percentile). The initial guess into the least-squares method is now given by Eq. (8).

## 3 Molecular Alignment, Docking, and Dynamics

The major goal of ROTDIF is to allow users to accurately and efficiently process relaxation data and understand the results in an intuitive way. Though the diffusion tensor provides important information about the overall shape of the molecular system under observation, it is difficult to relate the quantitative values of the diffusion tensor to actual structural models.

In case of bimolecular systems, we have previously demonstrated that rigid-body alignment and docking of the molecules based on relaxation data alone can provide fairly accurate models for inter-molecular interactions (Ryabov and Fushman, 2007b; Berlin et al, 2011). In ROTIDF 3 we provide a built-in interface, described in the next section, that allows the user to seamlessly switch from simple derivation of the diffusion tensor to immediate quantification of the results (using ELMDOCK(Berlin et al, 2011)) in terms of intermolecular orientation and positioning, without requiring any additional input data or program parameters. These new functionalities, described below, have an advantage of providing immediate quantitative/structural information and its visualization.

First, the computationally improved version of the *ab initio* prediction of the rotational diffusion tensor (ELM) developed earlier (Ryabov et al, 2006) can be directly accessed via a button on the main screen. ELM is based on an ellipsoid representation of the molecular shape. This component has been completely rewritten to provide significant improvement in speed and stability, and can be accessed as a standalone component or through the provided Application Programming Interface (API). Instead of fully computing the solvent-accessible surface, as implemented in the original version, the surface is approximated by placing multiple uniformly distributed spheres (representing water molecules) around each atom. Each sphere is checked for collision against all atoms of the molecule in $O(\log N)$ time by use of an octree, and the colliding spheres are removed from the computation. The contact points between the remaining spheres and the atoms are used to compute the best-fit ellipsoid, and the associated diffusion tensor of the ellipsoid is computed, as previously described (Ryabov et al, 2006). The *ab initio* predicted diffusion tensor is computed in under a second, and can be immediately compared to the tensor derived directly from the experimental data.

Second, our diffusion-tensor-guided docking program ELMDOCK (Berlin et al, 2011) is also directly integrated into ROTDIF 3. The docked model based on the relaxation data (currently limited to two molecules/domains) is computed within a few seconds, and the resulting structure can be saved directly to a PDB-format coordinate file. The integrated version of ELMDOCK includes a new global alignment algorithm that improves upon the eigendecomposition-based alignment method used previously(Berlin et al, 2011). The eigendecomposition method is now used as an initial guess for a convex non-linear least-squares optimizer that further refines the Euler angles, so either Eq. (9) or Eq. (12) is now minimized.

The new alignment algorithm and the *ab initio* diffusion tensor predictor ELM are combined to create a rigid-body alignment and docking algorithm. The algorithm uses the BOBYQA derivative-free minimizer (Powell, 2009), and follows the basic algorithm described previously (Berlin et al, 2011), but now the overall $\chi^2$ (Eq. (9)) is directly minimized. All the docking components can be accessed separately through the provided API or directly from

the main ROTDIF 3 interface. All docking parameters are automatically selected by the program, and no problem-specific adjustment is ever required.

Finally, model-free analysis of local bond dynamics, based on the DYNAMICS(Fushman et al, 1997; Hall and Fushman, 2003; Fushman, 2012) program, has been integrated into ROTDIF 3. The model selection is now performed using Akaike information criterion (AIC), instead of the F-test, as suggested elsewhere(d'Auvergne and Gooley, 2003).

## 4 User Interface

The most visible change in the new version of ROTDIF is the introduction of a Graphical User Interface (GUI). The interface ties together all the features of ROTDIF in a simple main window, instead of the previous command-line interface. The concept behind the interface is to allow the user to leverage the fast computation time of our algorithm in order to quickly and intuitively understand and refine the experimental data. The status of parsing an input file is immediately displayed once the file has been selected, and the progress of computation is shown via a progress bar. Once the diffusion tensor is computed, its principal axes can be overlaid on top of the molecule and visualized in PYMOL (DeLano, 2002) via an automatically generated python script. The sample screenshots of the interface are given in Fig. 2.

The interface also provides error-dialog screens to allow the user to clearly understand where the error is (e.g., wrong variable format, wrong variable names, mismatched information, etc), and how it should be corrected. All of the plots available in the previous ROTDIF version (Walker et al, 2004) can now be accessed through individual buttons on the main screen. The plots also provide interactive access to additional information: individual data points in the plots can now be identified and detailed information displayed when mousing over them.

## 5 Results

To quantify the speed and accuracy of our algorithms, we tested ROTDIF 3 on various synthetic and real experimental data. The computation and timing were performed on a Macbook Pro laptop with a 2.66 GHz Intel Core i7 processor. Apache Math Commons 3.2 library was used for non-linear minimizations.

### 5.1 Analysis of Synthetic Data

To demonstrate the speedup of ROTDIF 3 we compared the computation time of our new algorithm against the direct analysis of $R_2/R_1$ ratios, also using our newly developed deterministic minimization algorithm, and the previous version of ROTDIF. The timing results for variously sized synthetic input data (Fig. 3) demonstrate orders of magnitude improvement in computational performance relative to the previous version. This improvement is due to the new deterministic sampling of starting points during minimization, efficient caching techniques for quick recomputation of $J(\omega)$, and the superior performance of Java 6 compiler and virtual machine relative to MATLAB's virtual machine (previous ROTDIF). The results also show that using the $\rho$ values (Eq. (6)) instead of the $R_2/R_1$ ratios, gives about a two-fold speedup in computation.

We further demonstrate the accuracy of diffusion tensor recovery from randomly generated data for the input parameters range used for Fig. 1, where we used relaxation data for a single bond-vector, but now we test for accuracy using an aggregate of 100 randomly orientated bond-vectors which share the same rotational diffusion tensor. For the isotropic case the diffusion tensor was chosen to have uniformly distributed random $\tau_c$ values in the

range between 5 ns and 20 ns. For the anisotropic tensor the principal values of the diffusion tensor were set to 1.0, 2.0, and $3.0 \times 10^{-7}$ $s^{-1}$. The average errors in the recovered diffusion tensor characteristics are shown in Fig. 4 in comparison with the direct analysis of $R_2/R_1$ .

As in our single-bond simulations (see Fig. 1), by subtracting high-frequency contributions from relaxation rates we are able to accurately recover the full rotational diffusion tensor in the presence of noise, in the absence of conformational exchange (see Fig. 4A–C and the 0% values in the plots in Fig. 4D–I). While our approach provides only a minor improvement in the accuracy over the $R_2/R_1$ method for $^{15}N$ relaxation and $R_{ex} = 0$, the accuracy is significantly improved in the $^{13}C$ case, while not making any assumptions about the CSA values.

As mentioned above, $R_2$ relaxation rates could contain an additional positive contribution from $R_{ex}$, which introduces a bias in our model. This bias breaks the basic assumption of $\chi^2$ minimization, and can introduce significant errors in the solution. Therefore, we analyzed the accuracy of our method under increasing amounts of error introduced by adding large $R_{ex}$ values to some percentage of residues. $R_{ex}$ values were chosen from a uniform distribution, with $5 < R_{ex} < 10$ $s^{-1}$ for $^{15}N$, and $25 < R_{ex} < 50$ $s^{-1}$ for $^{13}C$. The results of the simulation, shown in Fig. 4D–I, clearly demonstrate that the robust regression method can significantly reduce errors in the derived rotational diffusion tensor in those cases where conformational exchange contributions are present. Even in case of significant overall $R_{ex}$ contributions, or potentially other outliers that do not fit the model-free assumption, the robust regression method is able to provide an accurate estimate of the anisotropic tensor. This ability to accurately estimate the diffusion tensor tensor can be used to quickly identify outliers, orient multi-domain molecules during molecular docking, or when computing diffusion tensors for large uncurated datasets in high-throughput applications.

### 5.2 Application to Experimental Relaxation Data

We now analyze experimental data for various systems to demonstrate that our method successfully reproduces previously published diffusion tensors, while all of our results were computed without any input-specific adjustment, selection of initial values (e.g. the upper and lower bounds on the **D** eigenvalues), any manual curation of local minima, or any other non-default option. The comparison is done using previously published results for the B3 domain of protein G (GB3)(Hall and Fushman, 2003), ubiquitin (Sheppard et al, 2009, 2010), as well as for several RNA and DNA constructs for which relaxation data are available (Boisbouvier et al, 2003; Eldho and Dayie, 2007; Duchardt and Schwalbe, 2005; Akke et al, 1997; Ferner et al, 2008). We use the first model from the PDB coordinates of each system for all analyses, and the PDB-data coordinate frame for our reported results. We define the anisotropy of the diffusion tensor as (Fushman et al, 2004)

$$\zeta = \begin{cases} 2D_z/(D_x+D_y) & \text{if } |D_x-D_y| < |D_y-D_z|, \\ 2D_x/(D_y+D_z) & \text{otherwise,} \end{cases} \quad (13)$$

and the rhombicity as

$$\eta = \begin{cases} 3(D_y-D_x)/(2D_z-D_x-D_y) & \text{if } |D_x-D_y| < |D_y-D_z|, \\ 3(D_y-D_z)/(2D_x-D_y-D_z) & \text{otherwise.} \end{cases} \quad (14)$$

For GB3 we analyzed an extensive set of $^{15}N$ relaxation data (see Fig. 5) measured at five different magnetic fields (corresponding to $^{1}H$ frequency of 400, 500, 600, 700, and 800

MHz) at 24°C (Hall and Fushman, 2003, 2006). Residues 2, 11, 12, 39, 40, 41, 47, 48, 49, 50, and 56 were excluded from the analysis because of their signal overlap, location in the flexible loops or tail, mutations that deviated from the original sequence/structure, and residues that were identified to undergo conformational exchange(Hall and Fushman, 2003). Additionally, for the 800 MHz and 500 MHz data, residues 10 and 38 were extreme outliers and therefore were excluded as well.

The results in Table 1 demonstrate that the old and the new versions of ROTDIF yield almost identical results for the GB3 dataset (in fact, the two versions give essentially identical results for any $^{15}$N dataset that includes $R_1$, $R_2$, $R_3$). Moreover, when performing a "blind" computation on all the residues, not just the core, our new robust regression algorithm provided much better result than the standard optimization algorithm, and was able to converge to the expected solution, as given by the "core" residues. In all cases, the angles between the computed diffusion tensor, the inertia tensor, and ELM-predicted diffusion tensor largest principal axes are around 5°, a good match to the theoretically expected results. Since robust regression provides a close solution to that for manually curated data, we conclude that it can be used to effectively identify significant outliers, as well as be used directly in blind high-throughput analyses of relaxation data.

We now focus on three new features of ROTDIF 3: deriving rotational diffusion tensor from $^{13}$C data, from combined $^{15}$N and $^{13}$C relaxation data, and validation of the results based on *ab initio* prediction of the rotational diffusion tensor using our built-in ELM module. Combined analysis of $^{15}$N and $^{13}$C relaxation data for a protein, using the axially-symmetric model, has been demonstrated before (Lee et al, 1997). However, accurate analysis of $^{13}$C or both $^{15}$N and $^{13}$C relaxation data using fully anisotropic diffusion tensor model, while particularly important for DNA and RNA, has been lacking; yet their function rely heavily on dynamics, and thus accurate characterization of the overall rotational diffusion tensor is essential (Dethoff et al, 2012; Shajani and Varani, 2007; Rinnenthal et al, 2011).

We start with the combined analysis of published $^{15}$N and $^{13}$C relaxation data (Sheppard et al, 2009, 2010) for a well-characterized protein ubiquitin (PDB ID: 1D3Z). In our approach we combine the carbon and nitrogen $\rho$ ratios, and fit them directly. The data were collected at 26°C on a spectrometer with $^{1}$H frequency of 600 MHz. The $^{13}$C relaxation data were obtained using a ubiquitin sample selectively $^{13}$C labeled at $C_\alpha$ positions; this removed $^{13}$C-$^{13}$C contributions to the measured relaxation rates. Based on the secondary structure and previous studies of ubiquitin, we defined residues 1–6, 12–14, 16, 17, 22, 25, 27–29, 31, 34, 39–45, 57–59, 64, 66–69 as the "core" rigid residues in this protein. The results of our analyses are shown in Table 2 together with the previously published data.

For the combined data, based on both the Akaike Information Criterion (AIC) (d'Auvergne and Gooley, 2003) (AIC= 97 for axially-symmetric vs. 57 for fully anisotropic) and the statistical F-test ($F$ = 27), the fully anisotropic diffusion model provides the best fit to the experimental data. However, due to the higher level of experimental noise in the reported $^{13}$C data, from the $^{13}$C data alone it was only possible to reliably determine the isotropic diffusion tensor. The axially-symmetric solution is also shown in this case, for completeness.

The ROTDIF 3 results in Table 2 closely match the published results for this dataset (Sheppard et al, 2009) and the results of previous studies of $^{15}$N relaxation in ubiquitin (Tjandra et al, 1995). Note that the previously reported analysis(Sheppard et al, 2009) of the more problematic $^{13}$C data used independently-derived *residue-specific* $^{13}$C CSA values (Tjandra and Bax, 1997), while our ROTDIF 3 computation was performed with no prior

information and using only default settings (see Supporting Information). The results demonstrate that our method for subtracting high-frequency contributions from $^{13}C$ relaxation data is effective and consistent with the more established approach to $^{15}N$ data analysis (Fushman et al, 1999b,a; Fushman and Cowburn, 2002; Walker et al, 2004; Fushman et al, 2004). Note that, like for GB3, the robust-regression analysis, which blindly selects all the residues, yielded a solution similar to that for "core" residues.

We now focus of analysis of relaxation data for nucleic acids. The first nucleic acid relaxation data set analyzed is for the Dickerson DNA dodecamer, $d(CGCGAATTCGCG)_2$ (Boisbouvier et al, 2003). This molecule was the first complete turn of B-DNA to be successfully crystallized and has been used as a model system since. Due to the palindromic nature of this DNA, the measured data correspond to both strands, which improves the orientational sampling of bond vectors. Indeed, we noticed a decrease in the orientational sampling parameter ($\Xi$) (Fushman et al, 2000) from 0.138, for just one-strand vectors, to 0.108 for the double-strand vectors. The improved sampling of bond orientations can also be easily visualized using the new ROTDIF 3 GUI. The importance of orientational sampling for accurate determination of the diffusion tensor was discussed earlier (Fushman et al, 2000; Boisbouvier et al, 2003). The experimental $^{13}C$ $R_1$ and $R_2$ data taken from published data (Boisbouvier et al, 2003) were obtained at 35°C and a $^1H$ frequency of 600 MHz. To avoid large CSA contributions to the relaxation rates, data were only measured for the ribose carbons, (C1′, C3′, and C4′), where the carbon CSA is small (Boisbouvier et al, 2003). No $R_3$ values were available with these data, so we estimated $R_3$ as described above. Nucleotides C1 and G12, corresponding to the flexible tail of the DNA fragment, were excluded from the definition of "core" residues due to their rapid internal motions. The bond vector orientations were obtained from the highly-refined NMR structure (PDB ID 1NAJ) (Wu et al, 2003). The results are shown in Table 3.

When the nineteen $\rho^{exp}$ values were assigned to only one strand of the DNA palindrome, a $\tau_c$ of 3.22 ± 0.01 ns was determined for the axially-symmetric model, shown in Table 3. The fully anisotropic solution also closely matches in magnitude and orientation the axially-symmetric solution, though the improvement in the fit is not statistically significant.

Assigning the nineteen $\rho^{exp}$ values to both strands of the palindromic DNA improved the sampling of bond orientations and yielded a $\tau_c$ of 3.29 ± 0.01 ns (Table 3). This matches the published value of 3.35 ± 0.03 ns, as well as the theoretical value predicted by ELM. Note that ROTDIF 3 automatically selected the prolate solution in the axially-symmetric case, since our algorithm was able to detect also the local minimum corresponding to the oblate solution (see e.g. (Blackledge et al, 1998)), without the need for any manual curation of the two cases. The orientation of the axially-symmetric diffusion tensor deviates only by 3° from the theoretical ELM solution (shown in Fig. 7), and by 5° from the inertia tensor. The quality of axially-symmetric tensor fit is shown in Fig. 8, and demonstrates that all the residuals fall within $2\sigma$ of the predicted values. Robust regression was not run because of high $\sigma$ error estimates, such that all residuals for the regular regression were already below $3\sigma$.

The next set of relaxation data analyzed here is for the D5 RNA fragment from the group 2 intron ribozyme (Eldho and Dayie, 2007) (Table 4), to illustrate the difficulty of correctly analyzing relaxation data from a very flexible RNA. The data were taken at 25° C at a $^1H$ frequency of 499.95 MHz. The solution structure (PDB ID 2F88) of D5 was used for orientation of the CH bond vectors (Seetharaman et al, 2006). The original publication(Eldho and Dayie, 2007) used a previous version of modified ROTDIF to obtain a diffusion tensor using carbon relaxation data, which yielded a $\tau_c$ of 6.0 ± 0.3 ns. One of the great challenges of relaxation data analysis for RNAs with flexible components is which

"flexible" nucleotides to exclude from the diffusion tensor determination. In the original publication, 16 out of the 25 bond vectors were excluded (Eldho and Dayie, 2007). These vectors corresponded to the highly flexible stem loop (nucleotides 15, 16, 17, and 18), internal bulge (nucleotides 24, 25, 26, and 28), and catalytic triad (nucleotides 2, 3, and 4). Disregarding this large number of data points prompted us to re-investigate how the results for D5 depend on which nucleotides are excluded. Table 4 shows diffusion tensor results for various combinations of the flexible regions removed. One can see that, as more nucleotides are added back to the data pool the $\tau_c$ value gradually increases but remains within the error bounds of the original publication. However, the estimate of the diffusion tensor orientation is what is most affected. This illustrates the need to be extremely cautious in excluding flexible nucleotides from analysis. Introducing robust analysis is one significant improvement over the previous iteration of the ROTDIF software for RNA analysis.

In this regard, it is also important to note that exclusion of nucleotides/vectors can greatly reduce the orientational sampling of the data. When all nucleotides are included, the orientational sampling parameter is $\Xi = 0.195$. Upon removal of the stem loop, internal bulge, and catalytic triad the $\Xi$ value raises to 0.280 and triggers a warning message. This case illustrates an important caveat when analyzing data sets with a limited number of vectors. It is not always immediately apparent which flexible vectors to exclude while retaining a sufficient orientational sampling in such cases. However the GUI and fast computational speed of ROTDIF 3 allow the user to quickly perform extensive analysis of various possible combinations of included vectors. While this does not answer the question which of the possible combinations most closely resembles the physical reality, it does provide the necessary metrics to arrive at reasonable conclusions.

To illustrate the utility of simultaneously analyzing both $^{15}$N and $^{13}$C relaxation data, not just for proteins but also for DNA and RNA, we now demonstrate ROTDIF 3 analysis on the well-studied cUUCGg tetraloop motif. For this cUUCGg tetraloop motif there is a wealth of relaxation data available, including both $^{15}$N and $^{13}$C. In particular, we analyzed $^{15}$N relaxation data taken at 0° C ("Akke et al.")(Akke et al, 1997), $^{15}$N and $^{13}$C data taken at 25° C ("Duchardt et al.")(Duchardt and Schwalbe, 2005), and just $^{13}$C data taken at 44° C ("Ferner et al.")(Ferner et al, 2008). All these data were collected at 600 MHz $^{1}$H frequency. All the data were fit using a recently obtained high-resolution structure of the tetraloop (PDB ID 2KOC) (Nozinovic et al, 2010). The characteristics of the diffusion tensor for the three datasets are presented in Table 5. Overall, our analysis reproduced the previously reported diffusion tensor parameters. However, we also discovered previously unreported temperature-driven structural transitions in the tetraloop.

The Akke et al. dataset, collected at 0° C, contains imino nitrogen $R_1$, $R_2$, and $R_3$ measurements. Out of the seven NH bond vectors available, G1 was excluded due to large internal motions. This left only six "core" bond vectors, which prevented us from using the fully anisotropic diffusion tensor model. Nevertheless, the axially-symmetric model fits experimental data well (see Fig. 9), and gives a $\tau_c$ of 5.35 ± 0.24 ns, in agreement with the previously reported value of 5.4 ± 0.10 ns (Akke et al, 1997). However, these results need to be interpreted conservatively since the resulting uncertainties are large due to the small amount of available data, for example, yielding an estimated 54° uncertainty in the orientation of the unique principal axis of the diffusion tensor.

The Duchardt et al. dataset includes both $^{15}$N and $^{13}$C $R_1$, $R_2$, and $R_3$ data measured at 25° C. In the original paper, the carbon and nitrogen needed to be analyzed separately, and the fully anisotropic diffusion model was not analyzed due to software limitation of ModelFree (Mandel et al, 1995). We analyzed all data simultaneously, with only nucleotides G1, U7, and G12 excluded. ROTDIF 3 allows selection of the appropriate rotational diffusion model

based on two criteria: the F-test and the AIC. Historically the F-test has been predominantly used for model selection; however, recently the AIC has been suggested as a potentially more accurate alternative (d'Auvergne and Gooley, 2003). According to AIC, the fully anisotropic model (AIC=357) provides the best fit to the Duchardt et al. dataset compared to the isotropic (AIC=1232) and axially-symmetric models (AIC=377), while the F-test suggests the axially-symmetric model ($F$ = 0.72). The results for the two models are similar, with the angle between the unique axis of the axially-symmetric tensor and the $D_z$ axis of the fully anisotropic tensor of only 1°, and only a small difference in the tensors eigenvalues. The $\tau_c$ of 2.33 ± 0.01 ns for both models is in line with the ELM prediction, and the angle between the principal axes of the derived and *ab initio* predicted anisotropic tensors is 8°. The anisotropic tensor's $\tau_c$ is also comparable to the previously published value of 2.31 ± 0.13 ns. The residuals of fit are shown in Fig. 10, and demonstrate a consistent simultaneous fit for both carbon and nitrogen data. The robust fit provides slightly different principal values, but the $\tau_c$ and the overall orientation of the tensor remained the same.

The Ferner et al. data set, collected at 44° C, contains significant outliers. The most common way for dealing with the outliers would be curation of the data set, i.e. removal of problematic data points/nucleotides that have elevated $R_2$ (possibly due to conformational exchange) and/or low $R_3$ (increased local dynamics) values. When we employed this approach, we ended up excluding G2, A4, U7, G12, and C14 (we define the remaining nucleotides as the "core"). As an alternative approach, we used robust regression, a built-in feature of ROTDIF 3, which yielded an almost identical result without any manual curation of the input data. The results for both methods are shown in Table 5, and the quality of fit of the axially-symmetric diffusion model for the core nucleotide dataset is shown in Fig. 11.

Interestingly, when analyzing the cUUCGg tetraloop data at 44° C, our axially-symmetric solution deviated from the expected prolate-tensor solution. Our results are consistent for both regular and robust fit, as well as with those for fully anisotropic tensor model (which only deviates by 2° from axially-symmetric tensor orientation), shown in Table 5). Further analysis indicated that an oblate tensor with an anisotropy of 0.59 ± 0.10 fits significantly better than a prolate tensor. The best prolate-tensor solution has a $\chi^2$ of 98, which is significantly higher than $\chi^2$ of 33 for the best oblate solution.

To visually demonstrate the difference in the orientation of the diffusion tensor at 25° and 44° C, in Fig. 12 we show the orientations of the principal $D_z$ axes of the anisotropic diffusion tensors for Duchardt et al. and Ferner et al. datasets, as well as of the theoretically predicted diffusion tensor. As previously discussed, the Akke et al. dataset contains a large uncertainty in the orientation of its diffusion tensor, therefore it is not shown. The orientation of the rotational diffusion tensor measured at 25° C is consistent with the theoretical prediction based on the shape of the RNA molecule (about 9° difference). However, at higher temperature, the $D_z$ axis changes its orientation significantly, while the overall rotational diffusion tensor changes from prolate to oblate. A subtle conformational switch can change the inertia tensor, and by association the diffusion tensor, from being along axis *A* to that of *C*. Most analysis of relaxation data assume that the structures at the different temperatures are identical and therefore readily applicable. Our current analysis suggests an important caution: Either the structure at 44° C is different from that at 25° C, or the structures remain the same and the molecule undergoes a temperature dependent conformational switch, or both. We cannot decide in favor of any of the three scenarios. What is clear however is that the data at the three temperatures are not identical.

Assuming the three-dimensional structures are identical at low and high temperatures, the above analysis suggests that at higher temperatures the tetraloop may tumble differently than at the lower temperatures. During our analysis, as noted above, we do in fact find both the

prolate and oblate solution. This is not unexpected, since relaxation data analysis using axially-symmetric tensor model is known to potentially find two minima (Blackledge et al, 1998), one for the prolate solution and one for the oblate solution. It is physically reasonable that a short RNA can undergo thermally induced fluctuations and fraying at higher temperatures, and may in fact tumble as an oblate ellipsoid. The consistently lower $\chi^2$ for the oblate solution for the axially-symmetric model, combined with the fact that the fully anisotropic model provides similar results, suggests that at a certain temperature a conformational change in the cUUCGg tetraloop could mediate the prolate to oblate transition. However in the absence of an actual three-dimensional structure at 44° C, this observation remains purely speculative but a testable hypothesis.

Another interesting observation is that the overall tumbling of cUUCGg measured at different temperatures generally follows the Stokes-Einstein-Debye relationship ( $\tau_c \propto$ viscosity/$T$ ). If we rescale all the $\tau_c$ values obtained using ROTDIF 3 to 25° C, we get $\tau_c$ values of 2.88, 2.35, 2.44 ns for Akke et al., Duchardt et al., and Ferner et al., respectively. Thus, even though the RNA potentially undergoes a conformational change or adopts a different three-dimensional structure or both, this would not be detected by a simple isotropic tumbling model. Detecting such a conformational change might require a more complex analysis using non-isotropic rotational diffusion models, which, as demonstrated above, is quite straightforward even for RNA, with the new ROTDIF 3 program.

## 6 Conclusions

We have introduced a major revision to the ROTDIF program that includes a combination of several important and unique features that are not found in any other published software package and should be of broad use to the NMR community. The features are implemented using new state-of-the-art algorithms, object-oriented design, and caching schemes, that provide orders of magnitude increase in computational speed compared to previous methods, contains complete Monte Carlo error analysis, and allows API access to the underlying computational engine. The API can be directly accessed in Java, MATLAB, and several other languages.

This is the first publicly available program that combines anisotropic rotational diffusion tensor analysis of relaxation data for $^{15}$N with $^{13}$C at multiple fields. This combination allows relaxation analysis of protein and/or RNA or DNA complexes using a unified program, for example to orient and position individual components of such complexes based on the overall rotational diffusion tensor. The experimental rotational diffusion tensor can be derived in seconds even for very large datasets, and a robust regression method can be used to compensate for unexpected contributions from conformational exchange and other factors.

The module is integrated with an updated version of the *ab initio* diffusion tensor predictor ELM to allow fast global alignment and rigid-body docking of molecules based solely on the relaxation data. We demonstrate that the ELM predictions match closely the experimentally obtained tensors for the tested DNA and RNA datasets. The ELM predictor is easily accessible through GUI and provides an independent validation method for the derived experimental results. To our knowledge this is the first and only package that integrates both the experimental and *ab initio* methods for relaxation data analysis, and allows rigid-body docking using relaxation data.

These features are easily accessed through our new multi-threaded GUI that works on virtually all modern computer platforms. The GUI includes interactive help, robust error checking and messages, and button access to interactive plots of the results. This tight visual

integration of features and robust feedback significantly lowers the technical barrier and reduces the time needed to perform relaxation analysis. Finally, the new version of ROTDIF is integrated into the new ARMOR package that includes a similar toolbox for data analysis and docking based on residual dipolar couplings (RDCs) (Berlin et al, 2009, 2010).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by the NIH grants GM095755 and GM065334 to DF and GM077326 to TKD, and Maryland Biotechnology award to TKD. We thank Vitali Tugarinov for providing relaxation data for ubiquitin.

ROTDIF, including the source code, can be downloaded from https://bitbucket.org/kberlin/armor or http://gandalf.umd.edu/FushmanLab/pdsw.

## References

Akke M, Fiala R, Jiang F, Patel D, Palmer A. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. RNA. 1997; 3(7): 702–709. [PubMed: 9214654]

Berlin K, O'Leary DP, Fushman D. Improvement and analysis of computational methods for prediction of residual dipolar couplings. Journal of Magnetic Resonance. 2009; 201(1):25–33. [PubMed: 19700353]

Berlin K, O'Leary DP, Fushman. Structural assembly of molecular complexes based on residual dipolar couplings. Journal of American Chemical Society. 2010; 132(26):8961–8972.

Berlin K, O'Leary DP, Fushman D. Fast approximations of the rotational diffusion tensor and their application to structural assembly of molecular complexes. Proteins: Structure, Function, and Bioinformatics. 2011; 79(7):2268–2281.

Blackledge M, Cordier F, Dosset P, Marion D. Precision and uncertainty in the characterization of anisotropic rotational diffusion by 15N relaxation. Journal of the American Chemical Society. 1998; 120(18):4538–4539.

Boisbouvier J, Wu Z, Ono A, Kainosho M, Bax A. Rotational diffusion tensor of nucleic acids from 13C NMR relaxation. Journal of Biomolecular NMR. 2003; 27:133–142. [PubMed: 12913409]

Bruschweiler R, Liao X, Wright P. Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. Science. 1995; 268(5212):886–889. [PubMed: 7754375]

Bryce DL, Grishaev A, Bax A. Measurement of ribose carbon chemical shift tensors for A-form RNA by liquid crystal NMR spectroscopy. Journal of the American Chemical Society. 2005; 127(20): 7387–7396. [PubMed: 15898787]

Case DA. Calculations of NMR dipolar coupling strengths in model peptides. Journal of Biomolecular NMR. 1999; 15(2):95–102. [PubMed: 10605083]

Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. Journal of the American Chemical Society. 1990; 112(12):4989–4991.

d'Auvergne EJ, Gooley PR. The use of model selection in the model-free analysis of protein dynamics. Journal of Biomolecular NMR. 2003; 25:25–39. [PubMed: 12566997]

d'Auvergne EJ, Gooley PR. Optimisation of nmr dynamic models ii. a new methodology for the dual optimisation of the model-free parameters and the brownian rotational diffusion tensor. Journal of Biomolecular NMR. 2008; 40(2):121–133. [PubMed: 18085411]

Dayie KT, Wagner G, Lefevre JF. Theory and practice of nuclear spin relaxation in proteins. Annual Review of Physical Chemistry. 1996; 47(1):243–282.

Dayie KT, Brodsky AS, Williamson JR. Base flexibility in HIV-2 TAR RNA mapped by solution 15N, 13C NMR relaxation. Journal of Molecular Biology. 2002; 317(2):263–278. [PubMed: 11902842]

Dayie, TK. Nucleic Acids: Dynamics Studies by Solution NMR. John Wiley & Sons, Ltd; 2013.

DeLano, W. The PyMOL user's manual. DeLano Scientific; San Carlos, CA: 2002. p. 382

Dethoff E, Chugh J, Mustroe A, Al-Hashimi H. Functional complexity and regulation through RNA dynamics. Nature. 2012; 482:322–330. [PubMed: 22337051]

Dosset P, Hus JC, Blackledge M, Marion D. Effcient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. Journal of Biomolecular NMR. 2000; 16(1):23–28. [PubMed: 10718609]

Duchardt E, Schwalbe H. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR 13C relaxation. Journal of Biomolecular NMR. 2005; 32(4):295–308. [PubMed: 16211483]

Eldho N, Dayie K. Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: Implications for catalysis. Journal of Molecular Biology. 2007; 365(4):930–944. [PubMed: 17098254]

Farrow N, Zhang O, Szabo A, Torchia D, Kay L. Spectral density function mapping using 15N relaxation data exclusively. Journal of Biomolecular NMR. 1995; 6(2):153–162. [PubMed: 8589604]

Ferner J, Villa A, Duchardt E, Widjajakusuma E, Wohnert J, Stock G, Schwalbe H. Nmr and md studies of the temperature-dependent dynamics of rna ynmg-tetraloops. Nucleic Acid Research. 2008; 36:1928–1940.

Fushman, D. Protein NMR Techniques, Methods in Molecular Biology. Vol. 831. Humana Press; 2012. Determining protein dynamics from 15N relaxation data by using DYNAMICS; p. 485-511.

Fushman D, Cowburn D. Model-independent analysis of 15N chemical shift anisotropy from NMR relaxation data. ubiquitin as a test example. Journal of the American Chemical Society. 1998a; 120(28):7109–7110.

Fushman D, Cowburn D. Studying protein dynamics with NMR relaxation. Journal of Biomolecular Structure and Dynamics. 1998b; 16(1):63–74.

Fushman D, Cowburn D. The effect of noncollinearity of 15N-1H dipolar and 15N CSA tensors and rotational anisotropy on 15N relaxation, CSA/dipolar cross correlation, and TROSY. Journal of Biomolecular NMR. 1999; 13(2):139–147. [PubMed: 10070755]

Fushman D, Cowburn D. Nuclear magnetic resonance relaxation in determination of residue-specific 15N chemical shift tensors in proteins in solution: Protein dynamics, structure, and applications of transverse relaxation optimized spectroscopy. Methods in Enzymology. 2001; 339:109–122. [PubMed: 11462809]

Fushman, D.; Cowburn, D. Protein NMR for the Millenium. In: Rama Krishna, N.; Berliner, Lawrence J., editors. Biological Magnetic Resonance. Vol. 20. Kluwer; 2002.

Fushman D, Weisemann R, Thüring H, Rüterjans H. Backbone dynamics of ribonuclease T1 and its complex with 2′ GMP studied by two-dimensional heteronuclear NMR spectroscopy. Journal of Biomolecular NMR. 1994; 4(1):61–78. [PubMed: 22911159]

Fushman D, Cahill S, Cowburn D. The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration. Journal of Molecular Biology. 1997; 266(1):173–194. [PubMed: 9054979]

Fushman D, Tjandra N, Cowburn D. An approach to direct determination of protein dynamics from 15N nmr relaxation at multiple fields, independent of variable 15N chemical shift anisotropy and chemical exchange contributions. Journal of the American Chemical Society. 1999a; 121(37): 8577–8582.

Fushman D, Xu R, Cowburn D. Direct Determination of Changes of Interdomain Orientation on Ligation: Use of the Orientational Dependence of 15N NMR Relaxation in Abl SH (32). Biochemistry. 1999b; 38(32):10,225–10,230.

Fushman D, Ghose R, Cowburn D. The effect of finite sampling on the determination of orientational properties: A theoretical treatment with application to interatomic vectors in proteins. Journal of the American Chemical Society. 2000; 122(43):10,640–10,649.

Fushman D, Varadan R, Assfalg M, Walker O. Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. Progress in Nuclear Magnetic Resonance Spectroscopy. 2004; 44(3–4):189–214.

Ghose R, Fushman D, Cowburn D. Determination of the rotational diffusion tensor of macromolecules in solution from NMR relaxation data with a combination of exact and approximate methods–application to the determination of interdomain orientation in multidomain proteins. Journal of Magnetic Resonance. 2001; 149(2):204–217. [PubMed: 11318619]

Hall JB, Fushman D. Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. Journal of Biomolecular NMR. 2003; 27:261–275. [PubMed: 12975584]

Hall JB, Fushman D. Variability of the 15N chemical shielding tensors in the B3 domain of protein G from 15N relaxation measurements at several fields. implications for backbone order parameters. Journal of the American Chemical Society. 2006; 128(24):7855–7870. [PubMed: 16771499]

Hansen AL, Al-Hashimi HM. Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: Towards new long-range orientational constraints. Journal of Magnetic Resonance. 2006; 179(2):299–307. [PubMed: 16431143]

Hoogstraten CG, Wank JR, Pardi A. Active site dynamics in the lead-dependent ribozyme. Biochemistry. 2000; 39(32):9951–9958. [PubMed: 10933815]

Ishima R, Nagayama K. Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. Biochemistry. 1995; 34(10):3162–3171. [PubMed: 7880811]

Kroenke CD, Loria JP, Lee LK, Rance M, Palmer AG. Longitudinal and transverse 1H-15N dipolar/15N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. Journal of the American Chemical Society. 1998; 120(31):7905–7915.

Lee L, Rance M, Chazin W, Palmer A. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C$\alpha$ nuclear spin relaxation. Journal of Biomolecular NMR. 1997; 9(3):287–298. [PubMed: 9204557]

Legault P, Hoogstraten CG, Metlitzky E, Pardi A. Order, dynamics and metal-binding in the lead-dependent ribozyme. Journal of Molecular Biology. 1998; 284(2):325–335. [PubMed: 9813121]

Leulliot N, Varani G. Current topics in rna-protein recognition, control of specificity and biological function through induced fit and conformational capture. Biochemistry. 2001; 40:7947–7956. [PubMed: 11434763]

Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. Journal of the American Chemical Society. 1982; 104(17):4546–4559.

Mandel AM, Akke M, Palmer AG III. Backbone dynamics of escherichia coli ribonuclease HI: Correlations with structure and function in an active enzyme. Journal of Molecular Biology. 1995; 246(1):144–163. [PubMed: 7531772]

Mittermaier A, Kay LE. New tools provide new insights in NMR studies of protein dynamics. Science. 2006; 312(5771):224–228. [PubMed: 16614210]

Nozinovic S, Fürtig B, Jonker H, Richter C, Schwalbe H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. Nucleic Acids Research. 2010; 38(2):683–694. [PubMed: 19906714]

Peng J. Exposing the moving parts of proteins with nmr spectroscopy. The Journal of Physical Chemistry Letters. 2012; 3(8):1039–1051. [PubMed: 22545175]

Peng JW, Wagner G. Frequency spectrum of NH bonds in Eglin C from spectral density mapping at multiple fields. Biochemistry. 1995; 34(51):16,733–16,752.

Powell, MJD. Tech. Rep. NA2009/06. Department of Applied Mathematics and Theoretical Physics, Cambridge University; Cambridge, England: 2009. The BOBYQA algorithm for bound constrained optimization without derivatives. technical report na2009/06.

Pozzi N, Vogt AD, Gohara DW, Di Cera E. Conformational selection in trypsin-like proteases. Current Opinion in Structural Biology. 2012; 22(4):421–431. [PubMed: 22664096]

Rinnenthal J, Buck J, Ferner J, Wacker A, Furthig B, Schwalbe H. Mapping the landscape of RNA dynamics with NMR spectroscopy. Accounts of Chemical Research. 2011; 44:1292–1301. [PubMed: 21894962]

Ryabov Y, Fushman D. Interdomain mobility in di-ubiquitin revealed by nmr. Proteins: Structure, Function, and Bioinformatics. 2006; 63(4):787–796.

Ryabov Y, Fushman D. A model of interdomain mobility in a multidomain protein. Journal of the American Chemical Society. 2007a; 129(11):3315–3327. [PubMed: 17319663]

Ryabov Y, Fushman D. Structural assembly of multidomain proteins and protein complexes guided by the overall rotational diffusion tensor. Journal of the American Chemical Society. 2007b; 129(25): 7894–7902. [PubMed: 17550252]

Ryabov Y, Geraghty C, Varshney A, Fushman D. An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. Journal of the American Chemical Society. 2006; 128(48):15,432–15,444.

Ryabov Y, Clore G, Schwieters C. Direct use of 15N relaxation rates as experimental restraints on molecular shape and orientation for docking of protein- protein complexes. Journal of the American Chemical Society. 2010; 132(17):5987–5989. [PubMed: 20392103]

Seetharaman M, Eldho N, Padgett R, Dayie K. Structure of a self-splicing group II intron catalytic effector domain 5: Parallels with spliceosomal u6 RNA. RNA. 2006; 12(2):235. [PubMed: 16428604]

Shajani Z, Varani G. Nmr studies of dynamics in rna and dna by 13C relaxation. Biopolymers. 2007; 86:348–359. [PubMed: 17154290]

Sheppard D, Li DW, Brüschweiler R, Tugarinov V. Deuterium spin probes of backbone order in proteins: 2H NMR relaxation study of deuterated carbon $\alpha$ sites. Journal of the American Chemical Society. 2009; 131(43):15,853–15,865.

Sheppard D, Li DW, Godoy-Ruiz R, Brüschweiler R, Tugarinov V. Variation in quadrupole couplings of $\alpha$ deuterons in ubiquitin suggests the presence of $C^{\alpha}$-$H^{\alpha}$…O=C hydrogen bonds. Journal of the American Chemical Society. 2010; 132(22):7709–7719. [PubMed: 20476744]

Stueber D, Grant DM. 13C and 15N chemical shift tensors in adenosine, guanosine dihydrate, 2′-deoxythymidine, and cytidine. Journal of the American Chemical Society. 2002; 124(35):10,539–10,551. [PubMed: 11772048]

Tjandra N, Bax A. Large variations in 13C$\alpha$ chemical shift anisotropy in proteins correlate with secondary structure. Journal of the American Chemical Society. 1997; 119(40):9576–9577.

Tjandra N, Feller S, Pastor R, Bax A. Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. Journal of the American Chemical Society. 1995; 117(50):12,562–12,566.

Tjandra N, Garrett D, Gronenborn A, Bax A, Clore G. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. Nature Structural & Molecular Biology. 1997; 4(6):443–449.

Walker O, Varadan R, Fushman D. Effcient and accurate determination of the overall rotational diffusion tensor of a molecule from 15N relaxation data using computer program ROTDIF. Journal of Magnetic Resonance. 2004; 168:336–345. [PubMed: 15140445]

Woessner D. Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. The Journal of Chemical Physics. 1962; 37:647.

Wu Z, Delaglio F, Tjandra N, Zhurkin V, Bax A. Overall structure and sugar dynamics of a DNA dodecamer from homo-and heteronuclear dipolar couplings and 31 P chemical shift anisotropy. Journal of Biomolecular NMR. 2003; 26(4):297–315. [PubMed: 12815257]

Ying J, Grishaev A, Bryce DL, Bax A. Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. Journal of the American Chemical Society. 2006; 128(35):11,443–11,454.

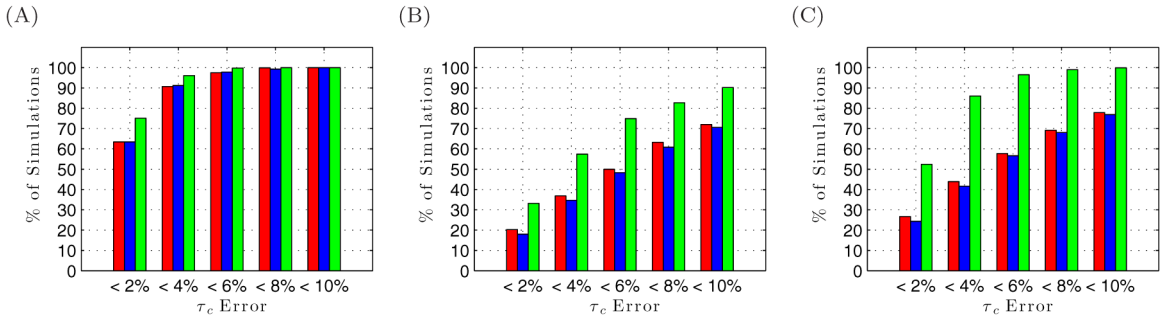


**Fig. 1.**

The percent of simulations, from 1000 independent runs, in which the relative error in the recovered $\tau_c$ was below the listed thresholds. The $\tau_c$ values were derived from the generated synthetic relaxation data using the following methods: direct analysis of the $R_2/R_1$ ratios (red bars, left) or analysis of the $\rho$ values (Eq. (6)) with high-frequency contributions subtracted using known $R_3$ values (green bars, right) or using predicted $R_3$ values, assuming that the measured $R_3$ values are not available (blue bars, middle). (A) The results for $^{15}$N in N-H bonds in a protein. (B) The results for $^{13}$C in C1′-H1′ bonds in RNA. (C) The results for $^{13}$C in C6-H6 bonds in RNA.

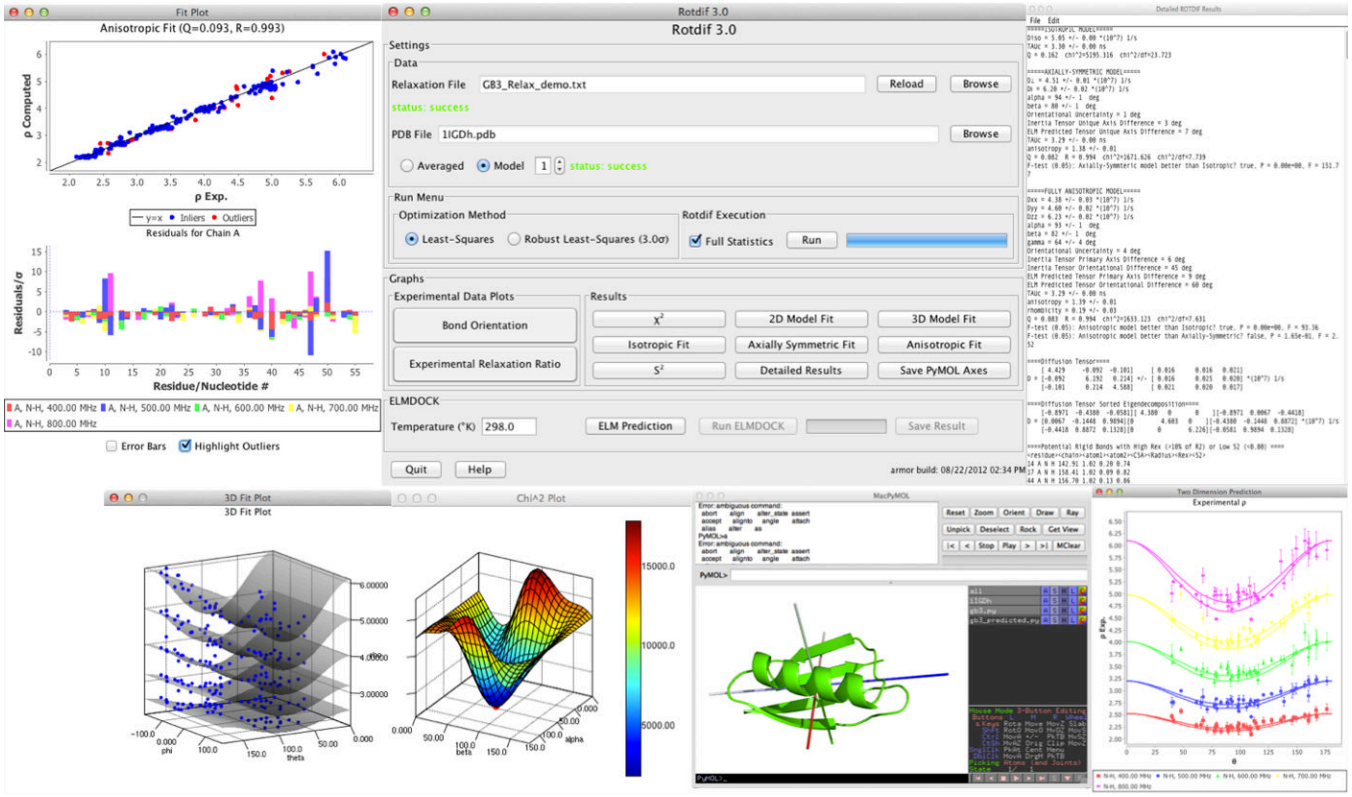


**Fig. 2.**

Demo screenshot of ROTDIF's Graphical User Interface and an overlay of the rotational diffusion tensor axes onto a protein structure in PYMOL (DeLano, 2002) via a ROTDIF-generated script.

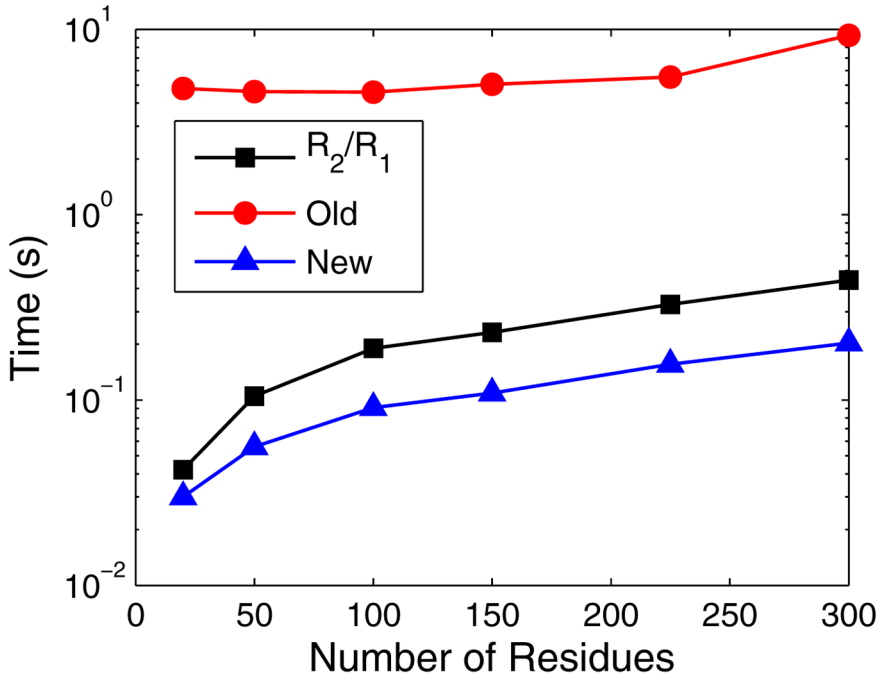


**Fig. 3.**

Timing results for computation of the anisotropic rotational diffusion tensor for randomly generated data of various sizes. The black line (squares) corresponds to direct analysis of $R_2/R_1$, achieved using our new deterministic initial sampling approach. The blue line (triangles) shows timing for the new deterministic high-frequency subtraction algorithm. The red line (circles) represents the previous version of ROTDIF, which uses a stochastic initial sampling algorithm. Note that both $R_2/R_1$ and ROTDIF 3 are implemented in Java, while the old ROTDIF runs in Matlab.

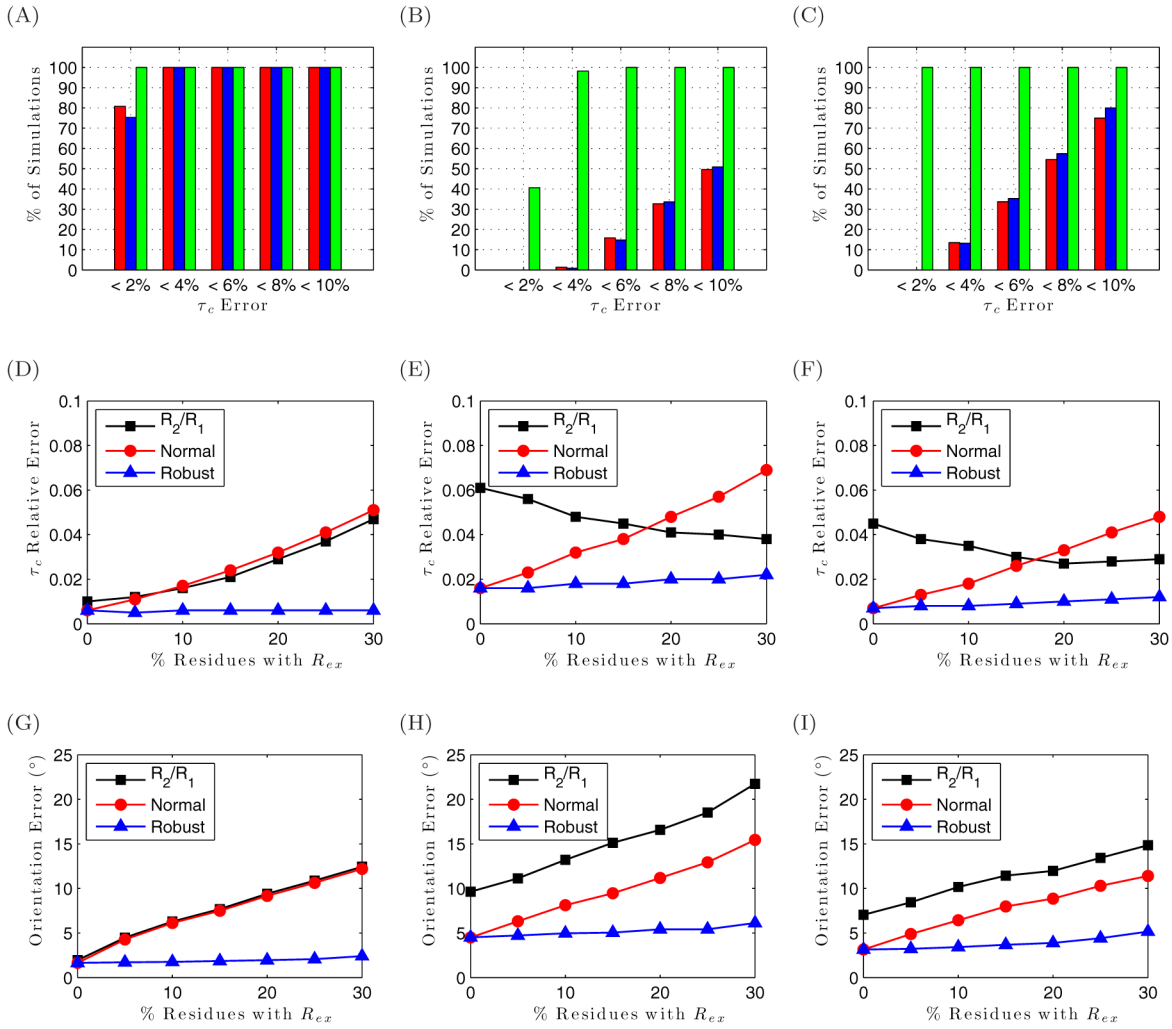


**Fig. 4.**

Simulation results for a set of 100 uniformly oriented PQ vectors, based on relaxation data for $^{15}$N in N-H bonds in a protein (left column), $^{13}$C in C1′-H1′ bonds in RNA (middle column), and $^{13}$C in C6-H6 bonds in RNA (right column). All simulations were performed for 1000 independent runs. (A–C) The percent of simulations in which the relative error in the recovered $\tau_c$ was below the listed thresholds, for isotropic diffusion tensor model. The $\tau_c$ values were computed using the following methods: direct analysis of the $R_2/R_1$ ratios (red bars) or analysis of the $\rho$ values with high-frequency contributions subtracted using known $R_3$ values (green bars) or using predicted $R_3$ values, assuming that the measured $R_3$ values are not available (blue bars). (D–I) Errors in the computed diffusion tensor ($\mathbf{D}_{pred}$) relative to the input tensor ($\mathbf{D}_{exp}$) for the anisotropic diffusion tensor model ($D_x = 1 \times 10^{-7}$ s$^{-1}$, $D_y = 2 \times 10^{-7}$ s$^{-1}$, $D_z = 3 \times 10^{-7}$ s$^{-1}$). The $x$-axis shows the percentage of residues with $R_{ex} > 0$. Shown are errors in the magnitude (D–F) and orientation (G–I) of the tensor.

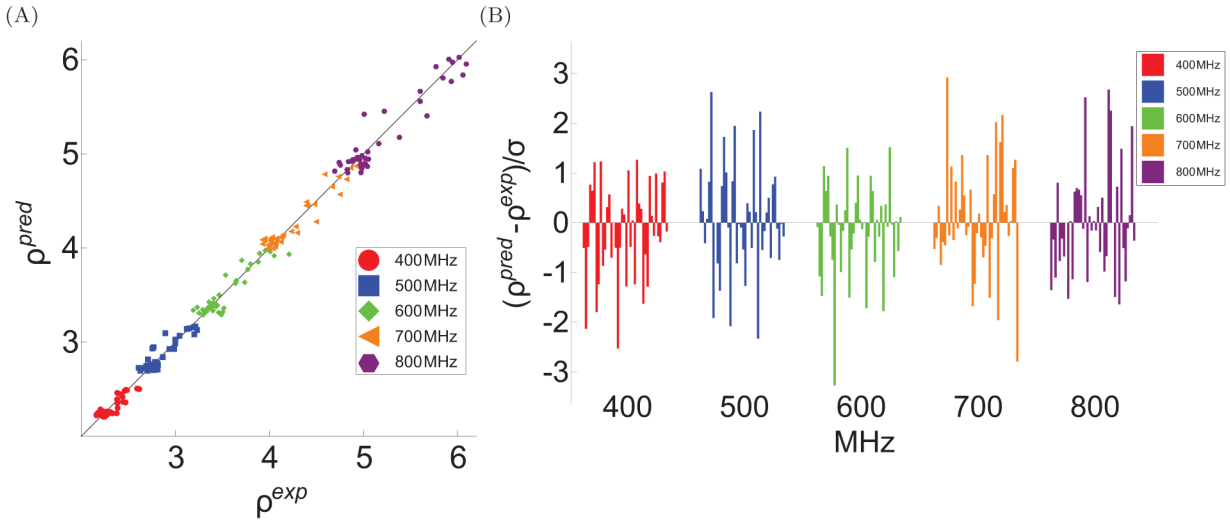


**Fig. 5.**

The agreement between the experimental and back-calculated $^{15}N$ relaxation data for GB3 at five magnetic fields for the fully anisotropic diffusion tensor model. (A) The correlation plot of the experimental vs. back-calculated $\rho$ values. (B) Fit residuals for individual residues, scaled by their associated standard deviations.

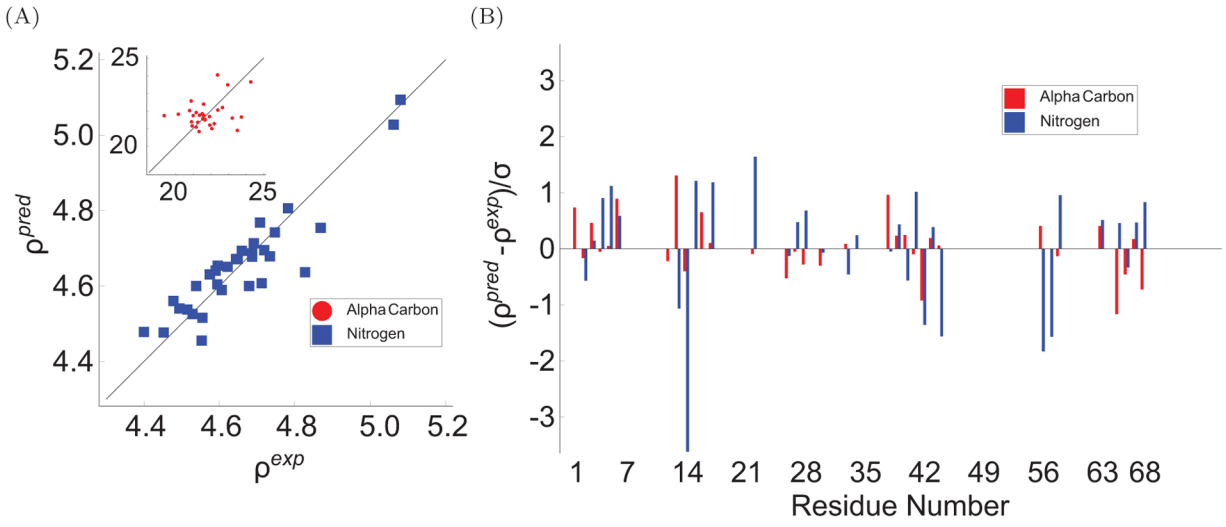


**Fig. 6.**

Combined fit of the $^{15}N$ and $^{13}C$ (inset) relaxation data for the core residues in ubiquitin to the fully anisotropic diffusion tensor model. (A) The agreement between the experimental and back-calculated $\rho$ values. (B) The residuals of fit for individual residues, scaled by their associated standard deviations.

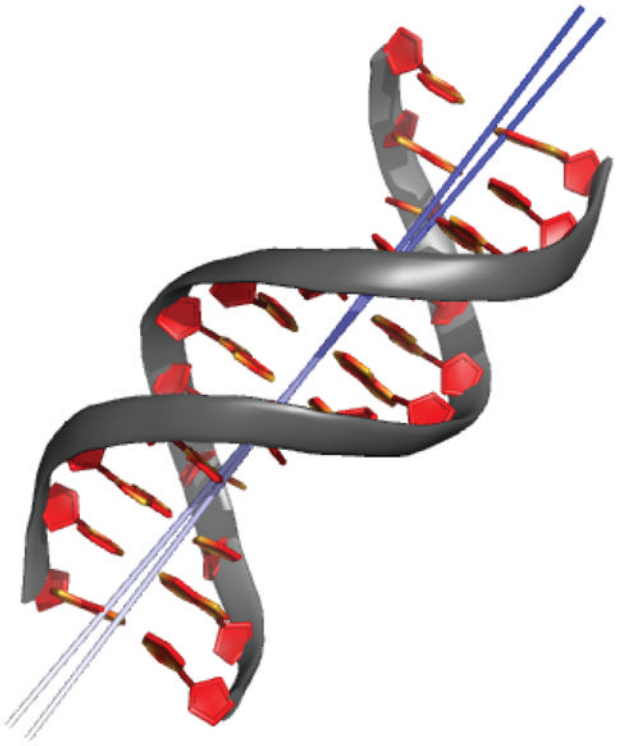

**Fig. 7.**

The principal axes (z) corresponding to the $D_z$ component of the determined (axially-symmetric) and ELM-predicted rotational diffusion tensors, overlaid on top of the cartoon representation of the Dickerson DNA dodecamer, d(CGCGAATTCGCG)$_2$. There is a 1° difference in the orientation of the two axes. The PyMOL script for drawing the axes was automatically generated by ROTDIF.

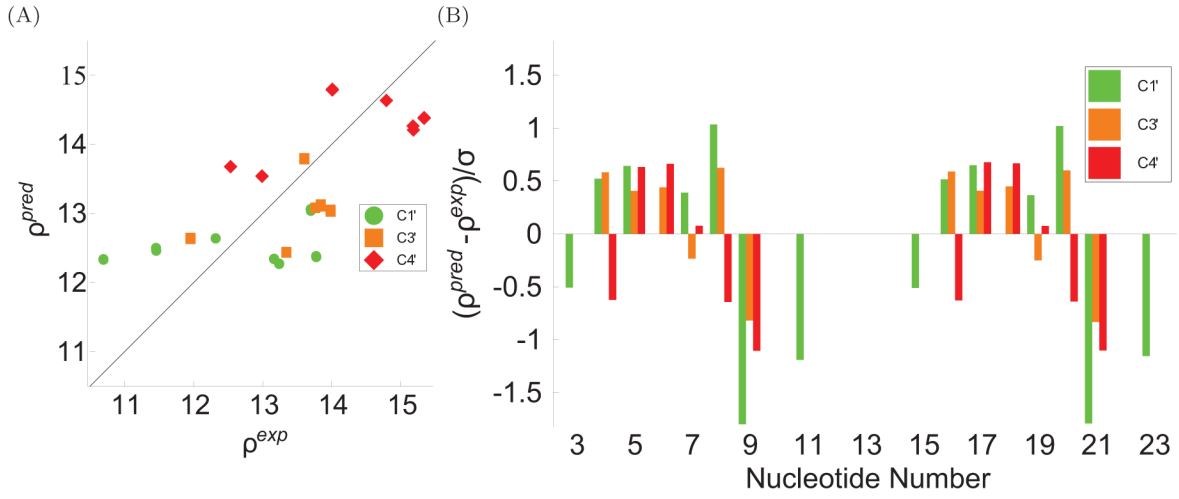


**Fig. 8.**

The fit of the $^{13}C$ relaxation data for the Dickerson DNA dodecamer using the axially-symmetric diffusion tensor model. (A) The agreement between the experimental and back-calculated $\rho$ values. (B) The residuals of fit, scaled by their associated standard deviations.

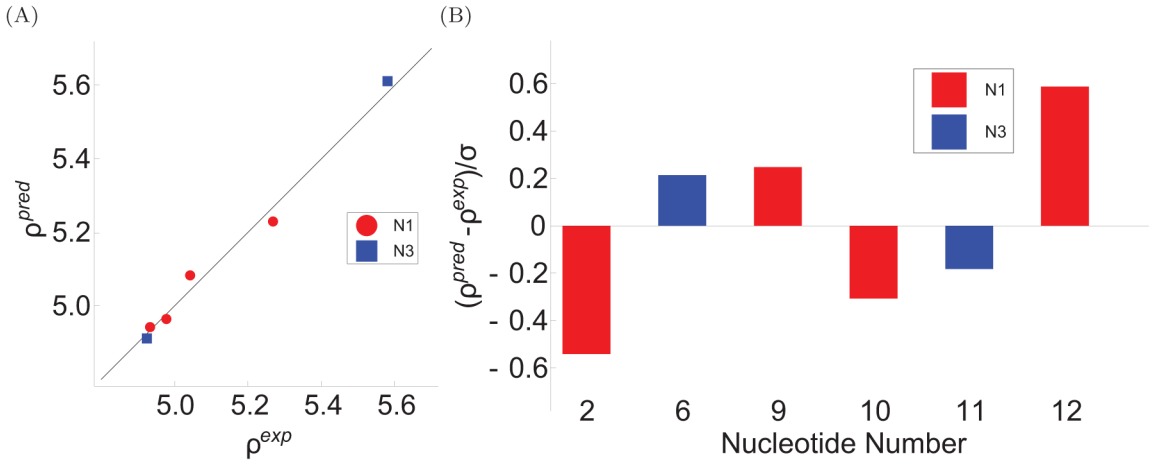


**Fig. 9.**

The fit of the $^{15}N$ relaxation data from the Akke et al. dataset (Akke et al, 1997) for the core nucleotides in cUUCGg tetraloop to the axially-symmetric diffusion tensor model. (A) The agreement between the experimental and back-calculated $\rho$ values. (B) The residuals of fit, scaled by their associated standard deviations.

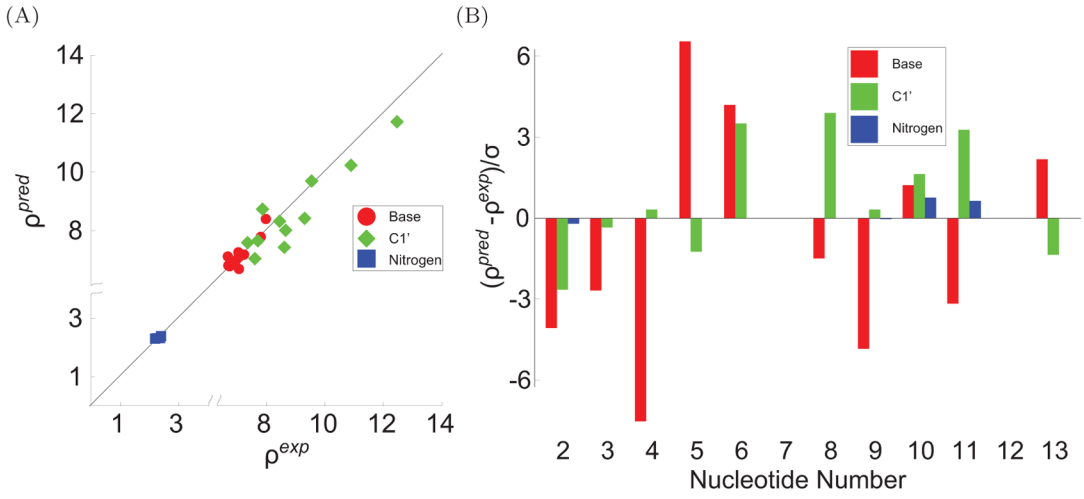


**Fig. 10.**

Simultaneous fit of the $^{15}N$ and $^{13}C$ relaxation data from the Duchardt et al. dataset (Duchardt and Schwalbe, 2005) for the core nucleotides in cUUCGg tetraloop to the fully anisotropic diffusion tensor model. (A) The agreement between the experimental and back-calculated $\rho$ values. (B) The residuals of fit scaled by their associated standard deviations.

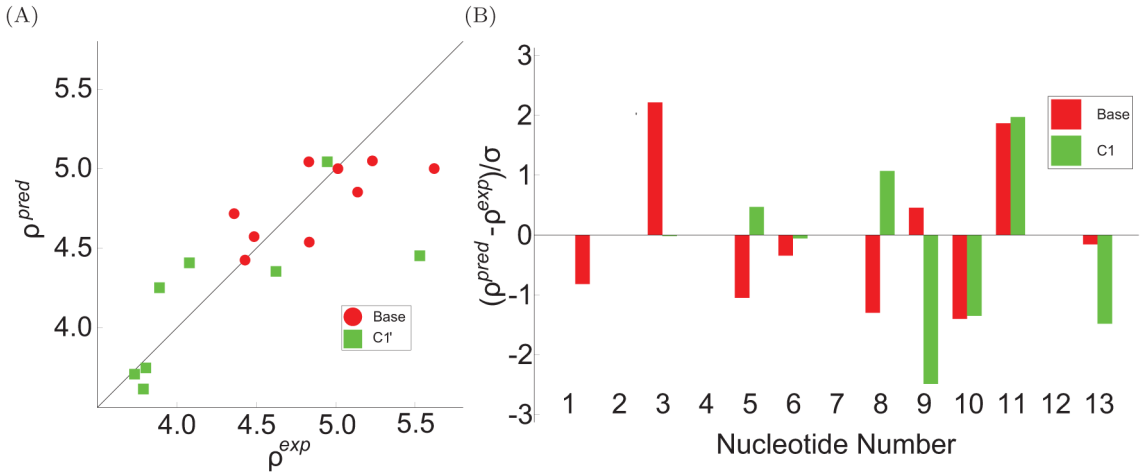


**Fig. 11.**

The fit of the $^{13}C$ relaxation data from the Ferner et al. dataset (Ferner et al, 2008) for the core nucleotides in cUUCGg tetraloop to the axially-symmetric diffusion tensor model. (A) The agreement between the experimental and back-calculated $\rho$ values. (B) The residuals of fit scaled by their associated standard deviations.

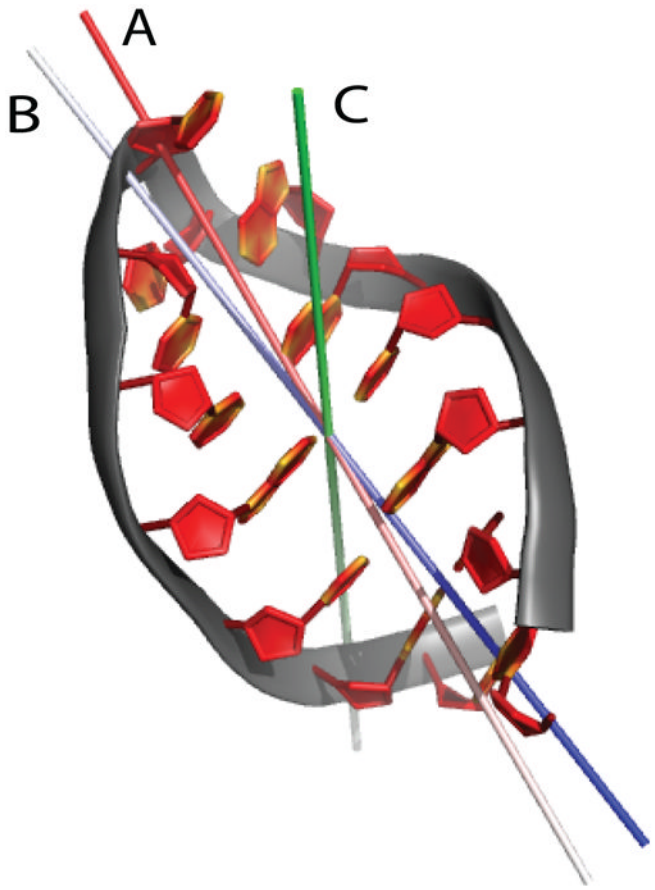


**Fig. 12.**

The principal $D_z$ axes of the anisotropic tensors from ELM prediction (A, red) and derived from Duchardt et al. (B, blue), and Ferner et al. (C, green) datasets, overlaid on top of the cartoon of cUUCGg. The PyMOL script for drawing the axes was automatically generated by ROTDIF 3.

**Table 1**

Characteristics of the rotational diffusion tensor of GB3 derived by ROTDIF from $^{15}$N relaxation data for the fully anisotropic diffusion model.

| Method[a] | Residues[b] | $D_x$ * | $D_y$ | $D_z$ | $\alpha$* | $\beta$ | $\gamma$ | $\tau_c$ * | $\zeta$[c] | $\eta$[d] |
|---|---|---|---|---|---|---|---|---|---|---|
| Old | core | 4.43 ± 0.16 | 4.57 ± 0.21 | 5.97 ± 0.39 | 93 ± 9 | 77 ± 7 | 56 ± 32 | 3.34 ± 0.10 | 1.33 ± 0.09 | 0.14 ± 0.02 |
| ROTDIF 3 | core | 4.43 ± 0.07 | 4.57 ± 0.07 | 5.97 ± 0.09 | 93 ± 3 | 77 ± 3 | 59 ± 29 | 3.34 ± 0.01 | 1.33 ± 0.03 | 0.14 ± 0.10 |
| ROTDIF 3 | all | 4.39 ± 0.04 | 4.49 ± 0.04 | 6.05 ± 0.06 | 90 ± 1 | 68 ± 2 | 116 ± 31 | 3.35 ± 0.01 | 1.36 ± 0.02 | 0.10 ± 0.06 |
| ROTDIF 3 (Robust) | all | 4.44 ± 0.05 | 4.58 ± 0.07 | 5.98 ± 0.08 | 96 ± 2 | 71 ± 3 | 81 ± 24 | 3.34 ± 0.01 | 1.33 ± 0.03 | 0.14 ± 0.11 |
| ROTDIF 3 | core (5 fields) | 4.41 ± 0.03 | 4.62 ± 0.03 | 6.03 ± 0.03 | 97 ± 1 | 75 ± 1 | 60 ± 4 | 3.32 ± 0.00 | 1.34 ± 0.01 | 0.21 ± 0.05 |
| ROTDIF 3 (Robust) | all (5 fields) | 4.39 ± 0.03 | 4.65 ± 0.03 | 6.05 ± 0.03 | 98 ± 1 | 75 ± 1 | 59 ± 3 | 3.31 ± 0.00 | 1.34 ± 0.01 | 0.26 ± 0.06 |

[a] The method used for the computation of the diffusion tensor. “Old” is the previously published version of ROTDIF, “ROTDIF 3” is our new ROTDIF 3, and “ROTDIF 3 Robust” is the robust regression algorithm in ROTDIF 3.

[b] Set of bond vectors selected for diffusion tensor analysis. “Core” residues are defined as 3–10,13–38,42–46,51–55 in GB3. 600 MHz data were used, unless specified otherwise.

[c] Anisotropy, see Eq. (13).

[d] Rhombicity, see Eq. (14).

* The units are $10^7 s^{-1}$ for $D_X$, $D_Y$, $D_Z$, ns for $\tau_C$, and degrees for $\alpha$, $\beta$, and $\gamma$ angles.

**Table 2**

Ubiquitin diffusion tensor values computed by ROTDIF for the fully anisotropic model derived from $^{15}$N (NH bonds) and $^{13}$C ($C_\alpha H_\alpha$ bonds) relaxation data.

| Method[a] | Residues | $D_x$ * | $D_y$ | $D_z$ | $\alpha$* | $\beta$ | $\gamma$ | $\tau_c$ * | $\zeta$ | $\eta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Published[b] | $^{15}$N | 3.58 | 3.58 | 4.22 | - | - | - | 4.16 ± 0.01 | 1.18 ± 0.02 | - |
| Published[c] | $^{13}$C | 3.60 | 3.60 | 4.28 | - | - | - | 4.11 ± 0.03 | 1.19 ± 0.02 | - |
| ROTDIF 3[e] | $^{15}$N core | 3.54 ±0.03 | 3.81 ±0.04 | 4.53 ±0.04 | 97 ±6 | 156 ±2 | 148 ±7 | 4.21 ±0.01 | 1.23 ±0.02 | 0.46 ±0.07 |
| ROTDIF 3 | $^{13}$C core | 4.04 ±0.06 | 4.04 ±0.06 | 4.04 ±0.06 | - - | - - | - - | 4.12 ±0.06 | - - | - - |
| ROTDIF 3[d] | $^{13}$C core | 3.75 ±0.13 | 3.75 ±0.13 | 4.45 ±0.26 | 125 ±83 | 145 ±17 | - - | 4.18 ±0.07 | 1.19 ±0.17 | - - |
| ROTDIF 3[e] | $^{13}$C&$^{15}$N core | 3.55 ±0.03 | 3.81 ±0.04 | 4.53 ±0.04 | 98 ±6 | 156 ±2 | 150 ±7 | 4.21 ±0.01 | 1.23 ±0.02 | 0.45 ±0.07 |
| ROTDIF 3[e] (Robust) | $^{13}$C&$^{15}$N all | 3.60 ±0.02 | 3.82 ±0.02 | 4.48 ±0.02 | 102 ±4 | 155 ±1 | 164 ±5 | 4.20 ±0.01 | 1.21 ±0.01 | 0.43 ±0.05 |

[a] The method used for the computation of the diffusion tensor. "Published" are the results taken from the corresponding publications. See Table 1 for other definitions.

[b] See Devon et al. (Sheppard et al, 2009) for the list of selected residues.

[c] See Devon et al. (Sheppard et al, 2010) for the list of selected residues.

[d] Axially-symmetric diffusion tensor model.

[e] Anistropic diffusion tensor model.

* See Table 1 for definition of the columns. The units are $10^7 s^{-1}$ for $D_X$, $D_Y$, $D_Z$, ns for $\tau_C$, and degrees for $\alpha$, $\beta$, and $\gamma$ angles.

**Table 3**

Rotational diffusion tensor for the Dickerson DNA dodecamer, d(CGCGAATTCGCG)$_2$, determined by ROTDIF from ribose CH $^{13}$C relaxation data. First model of 1NAJ was used for all computations.

| Method | Nucleotides | $D_x$ * | $D_y$ | $D_z$ | $\alpha$* | $\beta$ | $\gamma$ | $\tau_c$ * | $\zeta$ | $\eta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Published (Boisbouvier et al, 2003) | core[b] | 3.64<br>- | 3.64<br>- | 7.64<br>- | -<br>- | -<br>- | -<br>- | 3.35 ± 0.03 | 2.1 ± 0.04 | -<br>- |
| ROTDIF 3 | core[a] | 4.30 ± 0.10 | 4.30 ± 0.10 | 6.93 ± 0.15 | 94 ± 1 | 85 ± 2 | -<br>- | 3.22 ± 0.01 | 1.61 ± 0.07 | -<br>- |
| ROTDIF 3 | core[b] | 3.77 ± 0.09 | 3.77 ± 0.09 | 7.66 ± 0.16 | 95 ± 1 | 97 ± 1 | -<br>- | 3.29 ± 0.01 | 2.03 ± 0.09 | -<br>- |
| ROTDIF 3[c] | core[b] | 3.65 ± 0.10 | 3.82 ± 0.09 | 7.73 ± 0.17 | 95 ± 1 | 97 ± 1 | 92 ± 8 | 3.29 ± 0.01 | 2.07 ± 0.10 | 0.06 ± 0.01 |
| ELM[d] | - | 3.35 | 4.49 | 6.94 | 97 | 93 | 98 | 3.38 | 1.77 | 0.57 |

[a] Single-stranded, where the relaxation data were assigned only to strand A.

[b] Double-stranded, where due to the palindromic nature of the DNA, the same relaxation data were assigned to strands A and B.

[c] Anisotropic rotational diffusion model was used.

[d] Results from our built in *ab initio* diffusion tensor predictor ELM.

* The units are $10^7 s^{-1}$ for $D_X$, $D_Y$, $D_Z$, ns for $\tau_C$, and degrees for $\alpha$, $\beta$, and $\gamma$ angles.

**Table 4**

Rotational diffusion tensor characteristics for D5 derived using ROTDIF from A and G base CH $^{13}$C relaxation data for axially-symmetric diffusion tensor model.

| Method | Nucleotides | $D_x$ * | $D_y$ | $D_z$ | $\alpha$* | $\beta$ | $\gamma$ | $\tau_c$ * | $\zeta$ | $\eta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Published (Eldho and Dayie, 2007) | -SL[a], -IB[b], -CT[c] | 2.5 ±0.1 | 2.5 ±0.1 | 3.4 ±0.2 | 69 ±7 | 98 ±15 | -<br>- | 6.0 ±0.3 | 1.4 ±0.2 | -<br>- |
| ROTDIF 3 | -SL, -IB, -CT | 1.81 ±0.05 | 3.11 ±0.04 | 3.11 ±0.04 | 126 ±42 | 3 ±2 | -<br>- | 6.23 ±0.04 | 0.58 ±0.02 | -<br>- |
| ROTDIF 3 | -SL, -IB | 1.78 ±0.04 | 3.12 ±0.05 | 3.12 ±0.05 | 128 ±24 | 5 ±2 | -<br>- | 6.23 ±0.05 | 0.57 ±0.02 | -<br>- |
| ROTDIF 3 | -SL, -CT | 1.99 ±0.02 | 2.84 ±0.02 | 2.84 ±0.02 | 103 ±2 | 18 ±2 | -<br>- | 6.51 ±0.03 | 0.70 ±0.01 | -<br>- |
| ROTDIF 3 | -IB, -CT | 2.01 ±0.02 | 2.93 ±0.03 | 2.93 ±0.03 | 101 ±43 | 2 ±1 | -<br>- | 6.35 ±0.03 | 0.68 ±0.01 | -<br>- |
| ROTDIF 3 | -SL | 2.01 ±0.02 | 2.83 ±0.02 | 2.83 ±0.02 | 108 ±2 | 19 ±2 | -<br>- | 6.52 ±0.02 | 0.71 ±0.01 | -<br>- |
| ROTDIF 3 | -IB | 1.99 ±0.03 | 2.92 ±0.03 | 2.92 ±0.03 | 105 ±2 | 6 ±2 | -<br>- | 6.38 ±0.03 | 0.68 ±0.02 | -<br>- |
| ROTDIF 3 | -CT | 2.02 ±0.02 | 2.86 ±0.02 | 2.86 ±0.02 | 95 ±2 | 12 ±2 | -<br>- | 6.46 ±0.02 | 0.70 ±0.01 | -<br>- |
| ROTDIF 3 (Robust) | all | 2.02 ±0.44 | 2.84 ±0.75 | 2.84 ±0.75 | 96 ±117 | 16 ±47 | -<br>- | 6.50 ±0.25 | 0.71 ±0.91 | -<br>- |
| ELM | - | 1.57 | 2.33 | 4.53 | 135 | 77 | 109 | 5.94 | 2.32 | 0.44 |

[a] Without Stem Loop

[b] Without Internal Bulge

[c] Without Catalytic Triad

* The units are $10^7 s^{-1}$ for $D_X$, $D_Y$, $D_Z$, ns for $\tau_C$, and degrees for $\alpha$, $\beta$, and $\gamma$ angles.

**Table 5**

Rotational diffusion tensor characteristics for cUUCGg tetraloop.

| Method | Nucleotides/Type | Dataset[c] | $D_x$ * | $D_y$ | $D_z$ | $\alpha$* | $\beta$ | $\gamma$ | $\tau_c$ * | $\zeta$ | $\eta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Published (Akke et al, 1997) | core/NH | 0°C | 2.77<br>- | 2.77<br>- | 3.71<br>- | -<br>- | -<br>- | -<br>- | 5.4 ±0.1 | 1.34 ±0.12 | -<br>- |
| ROTDIF 3[a] | core/NH | 0°C | 2.82 ±0.56 | 2.82 ±0.56 | 3.71 ±1.32 | 18 ±40 | 133 ±19 | -<br>- | 5.35 ±0.20 | 1.31 ±1.33 | -<br>- |
| ELM | - | 0°C | 2.14 | 2.67 | 3.66 | 1 | 91 | 42 | 5.90 | 1.52 | 0.63 |
| Published (Duchardt and Schwalbe, 2005) | core/NH&CH | 25°C | 6.46<br>- | 6.46<br>- | 8.72<br>- | -<br>- | -<br>- | -<br>- | 2.31 ±0.13 | 1.35 ±0.02 | -<br>- |
| ROTDIF 3[a] | core/NH&CH | 25°C | 5.29 ±0.07 | 5.29 ±0.07 | 10.91 ±0.16 | 173 ±1 | 85 ±1 | -<br>- | 2.33 ±0.01 | 2.06 ±0.06 | -<br>- |
| ROTDIF 3[b] | core/NH&CH | 25°C | 4.97 ±0.08 | 5.53 ±0.08 | 10.97 ±0.16 | 175 ±1 | 85 ±1 | 35 ±6 | 2.33 ±0.01 | 2.09 ±0.06 | 0.15 ±0.02 |
| ROTDIF 3[b] (Robust) | all/NH&CH | 25°C | 4.73 ±0.13 | 5.24 ±0.13 | 11.56 ±0.26 | 174 ±2 | 86 ±1 | 20 ±9 | 2.32 ±0.01 | 2.32 ±0.10 | 0.12 ±0.03 |
| ELM | - | 25°C | 4.67 | 5.81 | 7.97 | 1 | 91 | 42 | 2.71 | 1.52 | 0.63 |
| Published (Ferner et al, 2008) | core/CH | 44°C | 9.42 | 9.42 | 13.84 | - | - | - | 1.53 | 1.47 | - |
| ROTDIF 3[a] | core/CH | 44°C | 7.14 ±0.26 | 12.19 ±0.25 | 12.19 ±0.25 | 131 ±4 | 162 ±4 | -<br>- | 1.59 ±0.02 | 0.59 ±0.03 | -<br>- |
| ROTDIF 3[b] | core/CH | 44°C | 6.96 ±0.27 | 11.50 ±0.40 | 13.11 ±0.61 | 50 ±20 | 116 ±17 | 107 ±38 | 1.58 ±0.02 | 0.57 ±0.24 | 0.43 ±0.23 |
| ROTIDF 3[a] (Robust) | all/CH | 44°C | 7.31 ±0.39 | 12.15 ±0.52 | 12.15 ±0.52 | 138 ±13 | 166 ±9 | -<br>- | 1.58 ±0.03 | 0.60 ±0.15 | -<br>- |
| ELM | - | 44°C | 7.55 | 9.41 | 12.90 | 1 | 91 | 42 | 1.67 | 1.52 | 0.63 |

[a] Axially-symmetric model was used.

[b] Anisotropic model was used.

[c] Akke et al. collected at 0°C, Duchardt et al. collected at 25°C, and Ferner et al. collected at 44°C.

* The units are $10^7 s^{-1}$ for $D_X$, $D_Y$, $D_Z$, ns for $\tau_C$, and degrees for $\alpha$, $\beta$, and $\gamma$ angles.