

Article

iDC: A comprehensive toolkit for the analysis of residual dipolar couplings for macromolecular structure determination

Yufeng Wei & Milton H. Werner*

Laboratory of Molecular Biophysics, Rockefeller University, 1230 York Avenue, New York, NY, Box 4210021, USA

Received 24 January 2006; Accepted 7 March 2006

Key words: alignment tensor, format converter, residual dipolar coupling, software program, tool kit

Abstract

Measurement of residual dipolar couplings (RDCs) has become an important method for the determination and validation of protein or nucleic acid structures by NMR spectroscopy. A number of toolkits have been devised for the handling of RDC data which run in the Linux/Unix operating environment and require specifically formatted input files. The outputs from these programs, while informative, require format modification prior to the incorporation of this data into commonly used personal computer programs for manuscript preparation. To bridge the gap between analysis and publication, an easy-to-use, comprehensive toolkit for RDC analysis has been created, iDC. iDC is written for the WaveMetrics Igor Pro mathematics program, a widely used graphing and data analysis software program that runs on both Windows PC and Mac OS X computers. Experimental RDC values can be loaded into iDC using simple data formats accessible to Igor's tabular data function. The program can perform most useful RDC analyses, including alignment tensor estimation from a histogram of RDC occurrence versus values and order tensor analysis by singular value decomposition (SVD). SVD analysis can be performed on an entire structure family at once, a feature missing in other applications of this kind. iDC can also import from and export to several different commonly used programs for the analysis of RDC data (DC, PALES, REDCAT) and can prepare formatted files for RDC-based refinement of macromolecular structures using XPLOR-NIH, CNS and ARIA. The graphical user interface provides an easy-to-use I/O for data, structures and formatted outputs.

Introduction

The measurement of residual dipolar couplings (RDCs) has become nearly universal for the determination and validation of macromolecular structures determined in solution by NMR (Prestegard et al., 2000; Bax et al., 2001; Blackledge, 2005). The global, orientational constraints provided by RDCs complement local NMR constraints (scalar couplings and nOes), making

RDCs extremely useful for the refinement of macromolecular structures as well as for the determination of relative domain orientation in multi-domain or multi-protein assemblies. In order to represent the orientational dependence of RDCs, a user-friendly analysis and visualization tool has been created that is particularly useful for the presentation of such data to the structural biology community.

Currently, there are several analysis tools for RDCs, REDCAT (Valafar and Prestegard, 2004), Pales (Zweckstetter and Bax 2000), DC (part of the *nmrPipe* (Delaglio et al., 1995) distribution)

*To whom correspondence should be addressed.
E-mail: mwerner@portugal.rockefeller.edu

and TENSOR (Dosset et al., 2000) are probably the most popular. All these programs are Linux/Unix based, requiring specifically formatted input files for analysis. Pales and DC are command-line driven while REDCAT has a graphical user interface (GUI) that can be ported to Windows PCs running Cygwin. However, porting REDCAT is technically cumbersome and requires some programming expertise. The output files and plots from many RDC analysis programs can be hard to understand for the non-expert and can be difficult to incorporate into other personal computer programs for publication quality output. Hence, there is a need for a format-free RDC toolkit that perform many different types of analyses which could be appended to common publication and analysis software programs in the personal computing environment.

iDC addresses this need. iDC is a comprehensive RDC toolkit running under the WaveMetrics Igor Pro program (<http://www.wavemetrics.com/>) that performs most common RDC analyses. iDC is user-friendly and does not require any special formatting for programmatic input. iDC is capable of making plots that are readily incorporated into other graphic and presentation programs suitable for publication. The iDC program is distributed as a procedure file for Igor Pro, making installation easy using the Igor procedure file GUI. iDC accepts many different types of RDC measurements, which can be loaded into separate data inputs (referred to as waves) and iDC can import and analyze multiple coordinate files in PDB format. The iDC program produces various types of output which can be saved in typical high resolution formats (EPS, PNG, JPEG, TIFF) for presentation and publication. iDC has been targeted to the structural biology community, so most, but not all detailed analyses of RDC data are provided.

User interface

Once the iDC procedure file is opened into Igor Pro and compiled, a new menu item named “iDC” will appear on the menu bar (Figure 1). All the functions provided by iDC will be activated through this menu. This is a user-friendly GUI with popup menus for input of parameters, so no command line input is required. The results are

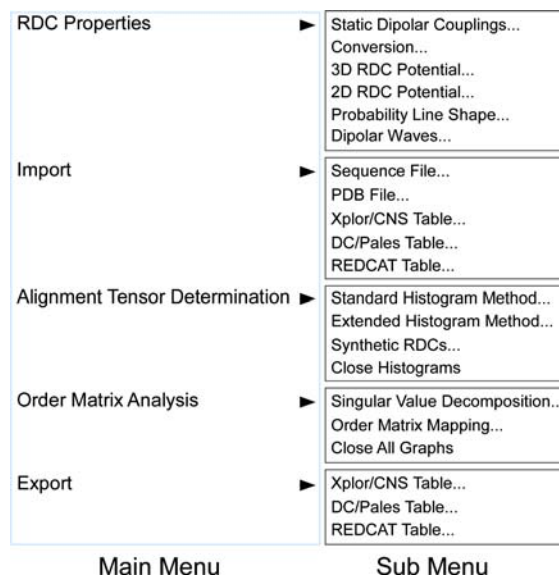


Figure 1. iDC menu structure in the Igor Pro program.

normally presented as easy-to-understand graphs using Igor Pro’s graphing functions, therefore eliminating the need for separate plotting software. In the following sections, we will illustrate the functions built into the program.

Residual dipolar coupling simulations

Dipolar couplings arise from the interaction of two magnetically active nuclei (such as ^1H and ^{15}N) in the presence of the external magnetic field. Equation (1) describes the dipolar interaction between a pair of spin 1/2 nuclei:

$$D_{\max}^{\text{AB}} = -\frac{\mu_0}{4\pi^2} \frac{\gamma_A \gamma_B \hbar}{r_{\text{AB}}^3} \quad (1)$$

Where D_{\max}^{AB} is the static dipole–dipole interaction between nuclei A and B, γ_A and γ_B are gyromagnetic ratios for nuclei A and B, respectively, r_{AB} is the internuclear distance (assumed fixed or calculated from PDB coordinates), \hbar is Planck’s constant divided by 2π , and μ_0 is the magnetic moment. For the application of RDCs in structural biology, it is convenient to normalize all observed dipolar couplings to the N–H dipolar coupling (Bax et al., 2001). The normalization factors (λ_{AB}) for various one-bond and two-bond dipolar couplings can be readily calculated using

the static dipolar couplings, as described in Equation (2)

$$\lambda_{AB} = \frac{\gamma_A \gamma_B r_{NH}^3}{\gamma_N \gamma_H r_{AB}^3} \quad (2)$$

The static dipolar couplings and normalization factors of a few commonly observed spin pairs as calculated from iDC are listed in Table 1.

Dipolar interactions can be readily measured in solid-state NMR using separation of local field (SLF) experiments, such as PISEMA (Rama-moorthy et al., 2004). However, the fast isotropic molecular tumbling motion in solution averages the dipolar couplings to zero as expressed in Equation (3)

$$D^{AB} = -\frac{\mu_0 \gamma_A \gamma_B \hbar}{4\pi^2 r_{AB}^3} \left\langle \frac{3 \cos^2(\theta_{AB}(t)) - 1}{2} \right\rangle \quad (3)$$

Where $\theta_{AB}(t)$ is the time dependent angle of the internuclear vector with respect to the external magnetic field. The time averaging of isotropic random samples of all possible angles reduces the dipolar couplings to zero. Re-introduction of anisotropy by alignment in a dilute liquid crystal results in measurable dipolar couplings with significantly reduced magnitude, indicative of the average orientation of an internuclear vector.

Assuming a rigid molecule, the orientation of the alignment can be described by a 3×3 order matrix, which is real and symmetric. Under this alignment tensor, the residual dipolar couplings then can be written as

$$D^{AB}(\theta, \phi) = D_a \left[(3 \cos^2 \theta - 1) + \frac{3}{2} R \sin^2 \theta \cos 2\phi \right] \quad (4)$$

Where D_a is the axial component of the molecular alignment tensor, $R = D_a/D_r$ is the rhombicity of the tensor, and θ and ϕ are the polar and azimuthal angles that describe the orientation of the A–B vector with respect to the alignment frame. The RDC distribution (ca. potential function) can be plotted as a function of θ and ϕ angles for given D_a and R values as either 3D contour or 2D curve (Figure 2). It is clearly shown in the distribution map (Figure 2A) that the extreme D^{AB} values are found at $\theta = 0^\circ$ and $\theta = 90^\circ$, $\phi = 90^\circ$, which correspond to the z and y axes of the alignment frame. Assuming uniformly and isotropically distributed

Table 1. Dipolar coupling magnitude and normalization factors

Parameter	Internuclear distance (Å)	Dipolar coupling magnitude ^a (Hz)	Normalization factor ^b (calculated)
$^1\text{H}-^{15}\text{N}$	1.04	10826.6	1.00
$^1\text{H}_\alpha-^{13}\text{C}_\alpha$	1.08	23995.3	2.0909
$^{13}\text{C}_\alpha-^{13}\text{C}'$	1.525	2143.5	0.1868
$^{13}\text{C}'-^{15}\text{N}$	1.329	1304.8	0.1137
$^1\text{H}-^{13}\text{C}'$	2.07	3407.9	0.297
$^{15}\text{N}-^{13}\text{C}_\alpha$	1.458	988.2	0.0861
$^1\text{H}-^{13}\text{C}_\alpha$	2.106	3236.1	0.282

^aComputed according to Equation (1). ^bComputed according to Equation (2).

A–B vectors, the probability of finding D^{AB} values will have the same pattern as the chemical shift anisotropy (CSA) tensor (Figure 3), with the highest probability D^{AB} values the same in magnitude as the A–B vector along the x axis ($\theta = 90^\circ$, $\phi = 90^\circ$). The probability pattern of residual dipolar couplings is the basis of the histogram method for the determination of the alignment tensor magnitude and rhombicity (Clare et al., 1998).

Normalization of residual dipolar couplings

The prevalence of the HSQC for the representation of a protein NMR spectrum made it logical that the most commonly measured dipolar coupling is that of the backbone amide $^1\text{H}-^{15}\text{N}$ vectors. However, many other couplings can be measured in the backbone, such as $^{15}\text{N}-^{13}\text{C}'$, $^{15}\text{N}-^{13}\text{C}_\alpha$, $^1\text{HN}-^{13}\text{C}'$, $^1\text{HN}-^{13}\text{C}_\alpha$, $^{13}\text{C}'-^{13}\text{C}_\alpha$, and $^1\text{H}_\alpha-^{13}\text{C}_\alpha$. Because RDCs have different intrinsic (or static) dipolar coupling constants due to the combination of gyromagnetic ratios and internuclear distances, the apparent RDC values for different types of couplings can vary greatly for the same sample and in the same alignment medium. For example, the magnitude of $^1\text{H}_\alpha-^{13}\text{C}_\alpha$ RDC values can be twice as much as $^1\text{H}-^{15}\text{N}$ values, while $^{13}\text{C}'-^{15}\text{N}$ magnitude is only 1/8 of $^1\text{H}-^{15}\text{N}$ values. To place a measured coupling into a single frame of reference, a common practice is to normalize the observed coupling to that measured for $^1\text{H}-^{15}\text{N}$, using normalization factors (λ_{AB}) calcu-

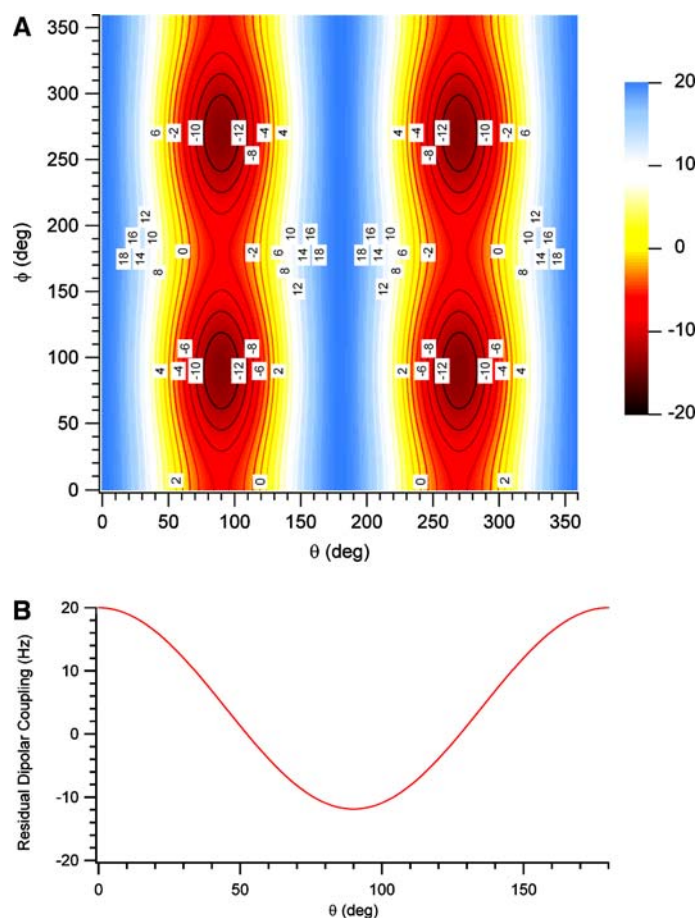


Figure 2. Dipolar coupling potential calculated from Equation 4 using $D_a = 10$ Hz, and $R = 0.25$: (A) 3D contour plot. The maximum is at $\theta = 0^\circ$ (z-axis), and the minimum is at $\theta = 90^\circ$ and $\phi = 90^\circ$ (y-axis). (B) 2D plot for $\phi = 60^\circ$.

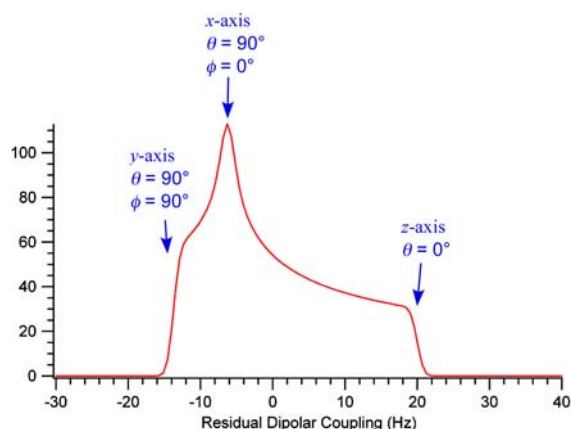


Figure 3. The distribution of residual dipolar couplings with $D_a = 10$ Hz and $R = 0.25$. The two extrema are along the y-axis ($\theta = 90^\circ$ and $\phi = 90^\circ$) and z-axis ($\theta = 0^\circ$), while the most probable D^{AB} values are along the x-axis ($\theta = 90^\circ$, $\phi = 0^\circ$).

lated from Equation (2) (see Table 1). For any type of dipolar couplings other than $^1\text{H}-^{15}\text{N}$, the apparent values should be divided by the normalization factor for that type, before they can be used in further analysis. Normalization is done on-the-fly in iDC, thus the user need only input and track the raw experimental values throughout an analysis session.

Determination of the Fully Asymmetric Alignment Tensor

Residual dipolar coupling data measured from partially aligned samples can provide unique long-range structural information. The axial and rhombic components of the molecular alignment tensor must be known in order to make use of the

RDC data in macromolecular structural determination. Clore and co-workers proposed that a histogram of the ensemble of normalized residual dipolar couplings can be used to extract the magnitude of axial and rhombic components of the alignment tensor in the absence of any prior structural information (Clore et al., 1998). A histogram, the plot of occurrence versus residual dipolar coupling values, provides a distribution of dipolar couplings along all orientations. The largest absolute value corresponds to z -axis of the alignment tensor, D_{zz} , while the other extremum provides the alignment on the y -axis, D_{yy} . The RDC value with highest occurrence is that on the x -axis, D_{xx} . As previously discussed, the histogram resembles a CSA powder pattern, and the magnitude of the alignment tensor, therefore, can be obtained by reading the extrema and highest occurrence values from the histogram, with, $D_a = \frac{1}{2}D_{zz}$ and $R = (D_{xx} - D_{yy})/3D_a$. In practice, however, the highest occurrence value, D_{xx} , cannot be easily obtained with precision from the histogram. Fortunately, the RDC tensor components fit the relation that $D_{xx} + D_{yy} + D_{zz} = 0$. Therefore, both D_a and R values can be obtained from the extrema, D_{zz} and D_{xx} , with $D_a = \frac{1}{2}D_{zz}$, and $R = -\frac{2}{3}(1 + D_{yy}/D_a)$. Similar to the CSA tensor, this histogram pattern provides only the magnitude of the alignment tensor. The orientation of the alignment tensor can be obtained from singular value decomposition (SVD) analysis once the structure of the macromolecule is known (see below). In most NMR structural refinement programs (such as XPLOR-NIH, Schweiters et al., 2003), however, only the magnitudes of the alignment tensor are required, and the orientations of the tensor will be optimized during the calculation.

Bryce and Bax (2004) extended this histogram approach by generating a continuous distribution of synthetic dipolar couplings using available $^1\text{H}-^{15}\text{N}$, $^{13}\text{C}'-^{13}\text{C}_\alpha$, and $^{15}\text{N}-^{13}\text{C}'$ dipolar couplings. This extended histogram method (EHM) takes advantage of the symmetry of the dipolar coupling tensor and the correlated property of protein backbone vectors. Since normalized dipolar coupling values depend on the orientation, but not on the position, of the nuclear spin vector, one can correlate all available types of dipolar couplings on the same peptide plane, and calculate the values for other orientations on that plane. In the EHM application, the $^1\text{H}-^{15}\text{N}$ vector is set as a

reference, and assumed to be 0° on the plane. Then the $^{13}\text{C}'-^{13}\text{C}_\alpha$ vector is along the angle of 56.76° , and the $^{15}\text{N}-^{13}\text{C}'$ vector is oriented at 120° angle. Due to the intrinsic symmetry of the dipolar coupling, the value at angle θ is the same as the value at $180^\circ + \theta$ on the same plane. From these three points, the dipolar coupling values at all other angles on the same plane can be calculated in a correlated manner. Those calculated dipolar couplings at angles where no actual bond vectors exist are termed synthetic dipolar couplings (SDCs) by Bryce and Bax (2004). Practically, one can generate hundreds of SDCs from the three measured dipolar couplings for each peptide plane, greatly increasing the sampling size of the histogram. As a result, a more consistent distribution curve can be obtained, allowing a more accurate analysis of the tensor magnitude.

iDC implements both the standard and extended histogram methods. The EHM is applicable when the $^1\text{H}-^{15}\text{N}$, $^{13}\text{C}'-^{13}\text{C}_\alpha$, and $^{15}\text{N}-^{13}\text{C}'$ dipolar couplings are available; otherwise, the standard method should be used. Once either of the methods is called from an iDC pull-down menu, a pop-up window will allow users to select RDC waves for each type of dipolar coupling. The program will automatically generate the histogram and calculate the D_a and R values based on the extrema (D_{yy} and D_{zz}) of the corresponding histogram. If the number of measured dipolar couplings is reasonably large, the results from the standard and extended histogram methods should be very similar (Figure 4A and B); however, if the available dipolar couplings are limited, the EHM can give much more accurate results. The program can also display the synthetic dipolar couplings for each peptide plane when all three types of couplings are available (Figure 4C).

Order Matrix Analysis

Use of singular value decomposition (SVD) to analyze the order tensor of residual dipolar couplings is an important step in refinement and validation of NMR structures (Losonczi et al., 1999). It allows the calculation of the full alignment tensor, magnitudes and orientations, from a structure and a set of dipolar coupling data. SVD analysis can also back-calculate the dipolar coupling data once the alignment tensor is obtained.

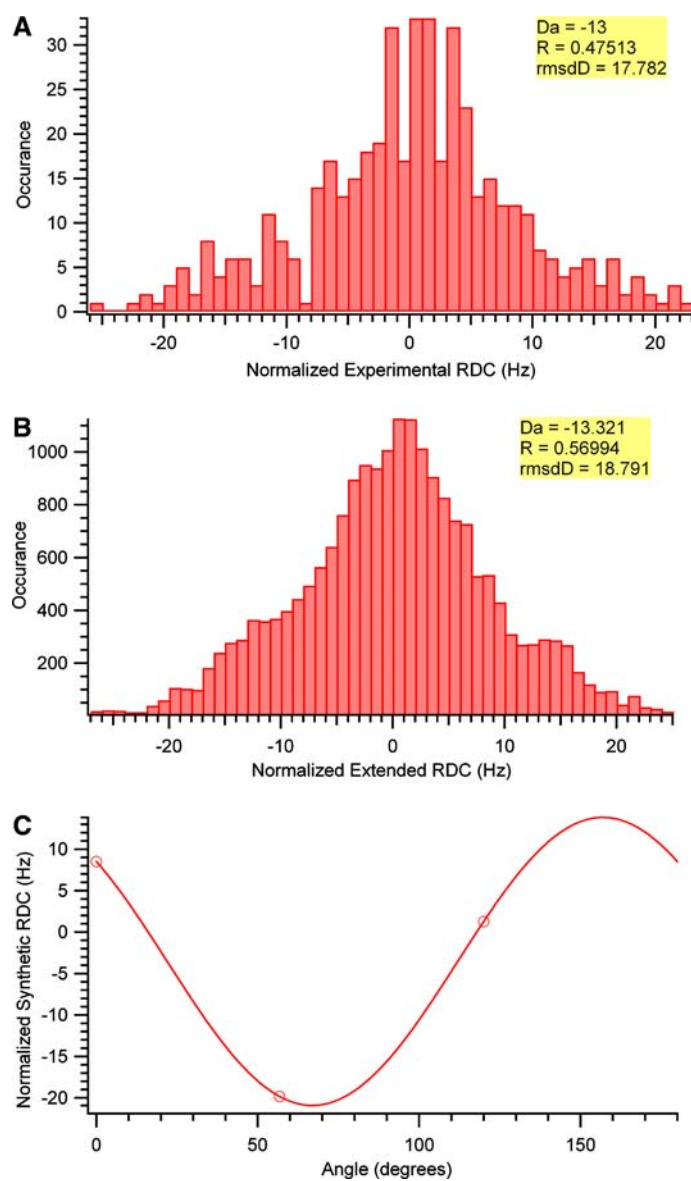


Figure 4. Determination of the magnitude of the fully asymmetric alignment tensor from histogram methods, using D^{N-H} , $\mathbf{D}^{C-\alpha}$, and D^{C-N} RDCs: (A) standard histogram method (Clare et al., 1998); (B) extended histogram method (EHM) (Bryce and Bax, 2004); (C) the synthetic dipolar coupling curve. The D_a and R values determined from these two methods are very close, with much larger distribution from the EHM plot. In the histograms, $\text{rmsdD} = [2\langle D^2 \rangle_{\text{random}}]^{1/2} = [2D_a^2(4 + 3R^2)/5]^{1/2}$ (Clare and Garrett 1999).

One can compare the experimental and back-calculated RDC values, and assess the quality of the structure refinement. The SVD method can also be easily called from the pull-down menu in iDC; RDC waves and types can be selected from the popup window. The SVD analysis in practice is to set-up and solve a system of linear equations of the form $\mathbf{Ax} = \mathbf{b}$. In the case of RDC analysis, the

matrix \mathbf{A} is composed of the direction cosines of the internuclear vectors calculated from the PDB file; the vector \mathbf{x} contains the five independent components of Saupe tensor (two magnitude components and three orientational components); the vector \mathbf{b} consists of the experimentally measured RDC data properly scaled (or normalized) according to the type of RDCs. Theoretically, only

five RDC data points will uniquely determine the full Saupe tensor; in practice, however, the number of RDC data are far more than five, and hence the SVD analysis can be considered as a best fit of the Saupe tensor with respect to the experimental RDCs. The RMSD between the experimental and back-calculated RDC ($\mathbf{Ax} - \mathbf{b}$) provides a criterion on how well the refined structure fits the measured RDC data. iDC will also calculate the R -factor (Equation 5) which can be used for cross-validation during NMR structure refinement (Clare and Garrett 1999).

$$\langle D^2 \rangle_{\text{random}} = D_a^2(4 + 3R^2)/5 \quad (5a)$$

$$R_{\text{dip}} = \left[\frac{\langle (D_{\text{obs}} - D_{\text{calc}})^2 \rangle}{2\langle D^2 \rangle_{\text{random}}} \right]^{1/2} \quad (5b)$$

The outputs of the SVD analysis include plots of experimental and back-calculated RDCs vs. residue number and back-calculated vs. experimental RDCs, with detailed tensor information in a separate window. If the mapping option is selected, the orientations of the alignment tensor can be plotted in the Sauson-Flamsteed projection with error-space sampling and null-space sampling as described for REDCAT (Valfar and Prestegard, 2004).

Unlike other RDC analysis packages, which can only use one set of PDB coordinates to back calculate RDCs using the SVD method, the iDC toolkit can read in an ensemble of NMR structures, and perform the SVD analysis on one or all of the structures in the ensemble. If the whole structure ensemble is selected, the error bar representing the simulation error between the structures will be given in the output. In practice, in order to refine an unknown protein structure with RDCs, one can estimate the magnitude of the alignment tensor, D_a and R , using one of the histogram methods, and feed them into a structure calculation protocol. After an ensemble of structures is obtained, SVD analysis can be performed on the ensemble to obtain a new set of D_a and R values. This new set of D_a and R can subsequently be fed back into the next round of refinement, until a good fit between experimental and back-calculated RDCs is achieved. SVD

analysis for the C-terminal domain of the bacterial RNA polymerase promoter recruitment factor σ^{70} in a complex with the T4 anti-sigma factor AsiA (Lambert et al., 2004) illustrates the iDC output in Figure 5. The alignment tensor as a Sauson-Flamsteed projection calculated with iDC is shown in Figure 6.

Import/Export and Format Conversion

iDC is a format-free package, which does not require any propriety format as input file. The RDC data is loaded as Igor Pro waves. However, iDC provides the convenience to import RDC data from data formats found in XPLOR-NIH, CNS (Brunger et al., 1998), DC, Pales or REDCAT. Similarly, iDC can also generate files for input into these programs, so one can use the output in structure calculations or perform further analysis using these program packages. iDC can therefore be used as a format converter in addition to RDC analysis.

Summary

An RDC toolkit, iDC, has been created that runs in the personal computing environment. iDC is a convenient add-on to the family of available programs for RDC analysis and presentation. Its interactive user interface, cross-platform, format-free, and graphical output features make it easy to install and use. It also combines the most useful tools for calculating the alignment tensor, such as the histogram method and SVD analysis, into a single program. The outputs of the toolkit are graphically presented, which can be subsequently saved in any Igor Pro supported graphical format (TIFF, PNG, JPEG and EPS). Thus, we believe iDC will facilitate RDC analysis by providing intuitive, easy-to-use tools to connect RDC restraints with macromolecular structures. The iDC toolkit can be downloaded for free at <http://www.rockefeller.edu/labheads/werner/index.php>, or by request to weiy@rockefeller.edu. If using an email request, use 'iDC' as the only text in the subject line and an autoreply will be given with the download information. iDC has been tested for Igor Pro Versions 4 and 5.

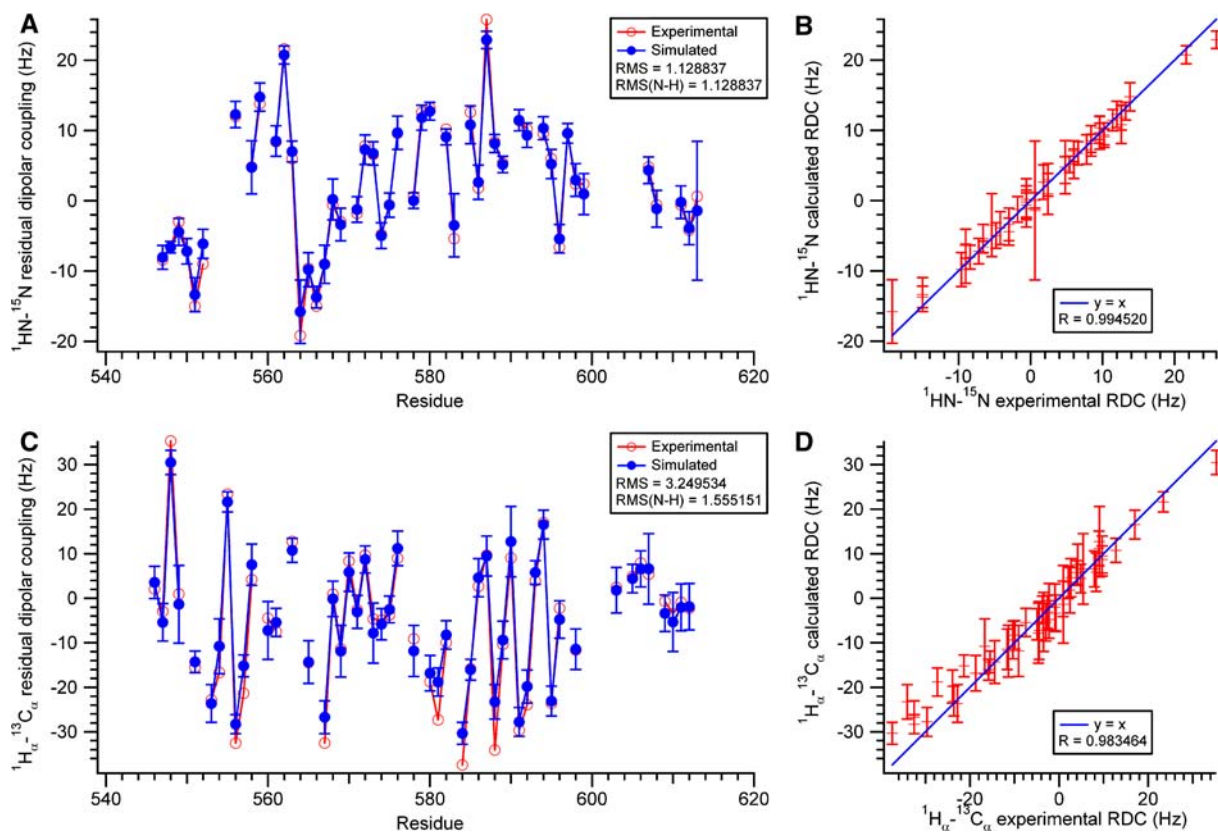


Figure 5. SVD analysis for σ^{70} region 4 in the complex with AsiA: (A) Experimental and simulated N-H dipolar couplings plotted as magnitude vs. residue number; (B) The fit between experimental and simulated N-H dipolar couplings; (C) Experimental and simulated $\text{C}_\alpha\text{-H}_\alpha$ dipolar couplings plotted as magnitude vs. residue number; (D) The fit between experimental and simulated $\text{C}_\alpha\text{-H}_\alpha$ dipolar couplings. The error bars on the simulated data represent the standard deviation among the cluster of NMR ensembles.

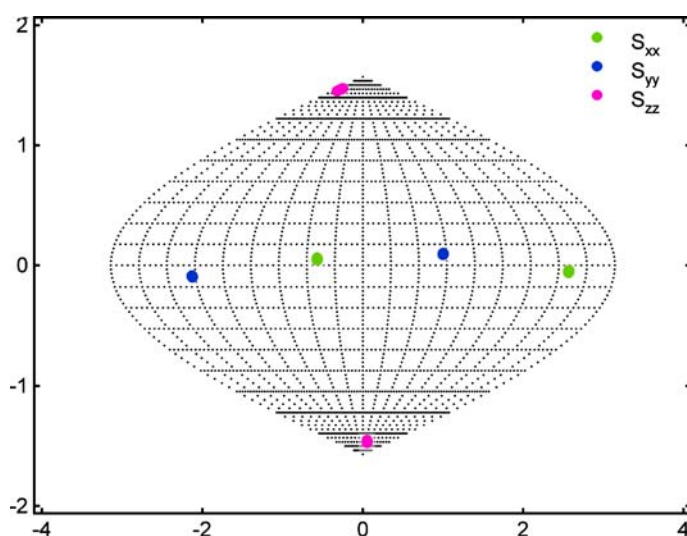


Figure 6. The alignment tensor for σ^{70} region 4 in complex with AsiA (Lambert et al., 2004) is represented as a Sauson-Flamsteed projection.

References

- Bax, A., Kontaxis, G. and Tjandra, N. (2001) *Methods Enzymol.*, **339**, 127–174.
- Blackledge, M. (2005) *Prog. Nucl. Magn. Reson. Spectrosc.*, **46**, 23–61.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, **54**, 905–921.
- Bryce, D.L. and Bax, A. (2004) *J. Biomol. NMR*, **28**, 273–287.
- Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.
- Clore, G.M. and Garrett, D.S. (1999) *J. Am. Chem. Soc.*, **121**, 9008–9012.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
- Dosset, P., Hus, J.C., Blackledge, M. and Marion, D. (2000) *J. Biomol NMR*, **16**, 23–28.
- Lambert, L.J., Wei, Y.F., Schirf, V., Demeler, B. and Werner, M.H. (2004) *Embo J.*, **23**, 2952–2962.
- Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.
- Prestegard, J.H., Al-Hashimi, H.M. and Tolman, J.R. (2000) *Q. Rev. Biophys.*, **33**, 371–424.
- Ramamoorthy, A., Wei, Y.F. and Lee, D.K. (2004) *Annu. Reports Nmr Spectrosc.*, **52**, 1–52.
- Schwieters, C.D., Kuszaewski, J.J., Tjandra, N. and Clore, G.M. (2003) *J. Magn. Reson.*, **160**, 66–73.
- Valafar, H. and Prestegard, J.H. (2004) *J. Magn. Reson.*, **167**, 228–241.
- Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.