Nick Patrick

1 Objective

NVR correlates experimentally determined RDCs to backcalculated RDCs to determine probabilistic constraints on assignment. For NVR to make accurate assignments, the algorithm must be able to accurately compute alignment tensors and backcompute RDCs.

NVR was originally written to use two sets of N-H RDCs in two different media. Two different alignment tensors are calculated, one for each medium. However, NVR could be modified to use one set of N-H RDCs and one set of C_α-H_α RDCs, both in the same medium. One alignment tensor must be calculated and used with both sets of RDCs.

The short-term objective for this project was to ensure that C_α-H_α RDCs are scaled properly and used correctly in tensor computation and RDC backcomputation with N-H RDCs. This was done independently of most of the NVR source code, and stands accomplished as of this distribution. The long-term objective of this project is to modify NVR to use both N-H and C_α-H_α RDCs to compute resonance assignments when the correct assignments are not known. The short-term objective is necessarily a prerequisite to completing the long-term objective.

This project was completed by Nick Patrick in Fall 2007 as part of an independent study in the Donald Lab at Duke University.

2 Background

Each RDC D can be backcalculated using

$$D = D_{max}\mathbf{v}^T\mathbf{S}\mathbf{v},$$

where D_{max} is a constant, $\mathbf{v}$ is the internuclear vector, and $\mathbf{S}$ is the Saupe alignment tensor, which has five degrees of freedom. In NVR, the elements of $\mathbf{S}$ are calculated using singular value decomposition (SVD). Given at least five RDCs and the corresponding internuclear vectors, one can solve for the elements of $\mathbf{S}$.

To use C_α-H_α RDCs in calculations with N-H RDCs, C_α-H_α RDCs must be scaled relative to N-H RDCs. A different expression describing the RDC D between two nuclei a and b is

$$D = -\frac{\gamma_a\gamma_b h}{2\pi^2\langle r_{ab}\rangle^3}\langle\frac{3\cos^2\Theta-1}{2}\rangle,$$

where γ_a and γ_b are the gyromagnetic ratios of nuclei a and b, h is Planck's constant, $\langle r_{ab}\rangle$ is the internuclear distance between nuclei a and b, and Θ is the angle between the internuclear vector and the applied magnetic field B_0.

Hence, C_α-H_α RDCs are scaled to N-H RDCs by the ratio

$$\frac{\gamma_{^{15}\mathrm{N}}/\langle r_{\mathrm{NH}}\rangle^3}{\gamma_{^{13}\mathrm{C}}/\langle r_{\mathrm{C}_\alpha\mathrm{H}_\alpha}\rangle^3},$$

as all other terms in the coefficient cancel. Using the values[1] $\gamma_{^{13}\mathrm{C}} = 6.73$, $\gamma_{^{15}\mathrm{N}} = 2.71$, $\langle r_{\mathrm{C}_\alpha\mathrm{H}_\alpha}\rangle = 1.090$, and $\langle r_{\mathrm{NH}}\rangle = 1.020$,

$$\frac{\gamma_{^{15}\mathrm{N}}/\langle r_{\mathrm{NH}}\rangle^3}{\gamma_{^{13}\mathrm{C}}/\langle r_{\mathrm{C}_\alpha\mathrm{H}_\alpha}\rangle^3} \approx 0.491.$$

[1] from `Const.java` by Lincong Wang, included in archive

3 Source

This section describes the source files in the archive. Only Perl scripts used for parsing and MATLAB functions used in the testing of RDC backcomputation are included. The rest of the NVR source code is not, as it is not required by these test functions. The rest of the NVR source code will be correctly modified to use C_α-H_α RDCs and included a future archive.

MATLAB functions

- `test2media.m`: Tests tensor calculation and RDC backcalculation for two sets of the same type of RDCs in different media (e.g. two sets N-H RDCs in different media). Uses `updateTen.m` for tensor computation. Outputs per residue data (experimental RDC, backcomputed RDC) to `out1.txt` and `out2.txt`.
- `test1medium.m`: Tests tensor calculation and RDC backcalculation for two sets of different types of RDCs in the same medium (e.g. one set of N-H RDCs and one set of C_α-H_α RDCs in the same medium). Uses `updateTen_CH.m` for tensor computation. Outputs per residue data (experimental RDC, backcomputed RDC) to `out1.txt` and `out2.txt`.
- `updateTen.m`: From the original NVR distribution used with this project. Computes the elements of the Saupe alignment tensor **S** using SVD, matching bond vectors to RDCs using a master assignment table. In the test code, the correct assignments are used (the master assignment table is an identity matrix).
- `updateTen_CH.m`: A modified version of `updateTen.m`. When two sets of RDCs in the same medium are used, both sets of RDCs must be used in the computation of the alignment tensor. This modified function allows callers to pass in two sets of RDCs and two sets of vectors for use in tensor computation.

Perl scripts

- `parse_vectors_N-H.pl`: Parses N-H bond vectors from a PDB file, normalizing the vectors. If a vector is not found for a given residue (e.g. for proline), the value $(-999, -999, -999)$ is used. This will be changed for use with the full NVR algorithm.
- `parse_vectors_C-H.pl`: Parses C_α-H_α bond vectors from a PDB file, normalizing the vectors. If two vectors are found for a given residue (e.g. for glycine), the value $(-999, -999, -999)$ is used. This will be changed for use with the full NVR algorithm.
- `parse_RDCs_C-H_N-H.pl`: Parses N-H and C_α-H_α RDCs from an MR file. C_α-H_α RDC values are scaled by the factor 0.491, as described in the Background. If no RDC is found for a given residue, the value -999 is used.

4 Future Work

The long-term objective of this project is to modify NVR to use N-H and C_α-H_α RDCs in the same medium. The first step was to ensure that C_α-H_α RDCs can be scaled properly and used correctly in tensor computation and RDC backcomputation, given the correct assignments. That has been accomplished in this distribution. In the next distribution, the NVR algorithm will be modified to use N-H and C_α-H_α RDCs to accurately compute resonance assignments when the correct assignments are not known. Additional efforts will be made to clean up the code, perform more testing, and write more documentation.