

Fitting alignment tensor components to experimental RDCs, CSAs and RQCs

Lukas N. Wirz · Jane R. Allison

Received: 18 November 2014 / Accepted: 27 January 2015 / Published online: 5 February 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Residual dipolar couplings, chemical shift anisotropies and quadrupolar couplings provide information about the orientation of inter-spin vectors and the anisotropic contribution of the local environment to the chemical shifts of nuclei, respectively. Structural interpretation of these observables requires parameterization of their angular dependence in terms of an alignment tensor. We compare and evaluate two algorithms for generating the optimal alignment tensor for a given molecular structure and set of experimental data, namely SVD (Losonczi et al. in *J Magn Reson* 138(2):334–342, 1999), which scales as $\mathcal{O}(n^2)$, and the linear least squares algorithm (Press et al. in *Numerical recipes in C. The art of scientific computing*, 2nd edn. Cambridge University Press, Cambridge, 1997), which scales as $\mathcal{O}(n)$.

Keywords Residual dipolar coupling · Chemical shift anisotropy · Residual quadrupolar coupling · Alignment tensor · Nuclear magnetic resonance

Introduction

The measurement of residual dipolar couplings (RDCs; Lipsitz and Tjandra 2004; Prestegard et al. 2004; Blackledge 2005) by solution state nuclear magnetic resonance

(NMR) spectroscopy gives information about the alignment of the vectors linking pairs of nuclei with non-zero spin relative to the magnetic field. The measured RDC value D_k is a function of $\langle \cos^2(\theta(\mathbf{r}_k, \mathbf{r}_H)) \rangle$, where $\theta(\mathbf{r}_k, \mathbf{r}_H)$ is the angle between the inter-spin vector \mathbf{r}_k and the magnetic field director \mathbf{r}_H . Specifically, D_k is given as

$$D_k(\mathbf{r}_k, \mathbf{r}_H) = -\frac{\gamma_{k_1}\gamma_{k_2}\mu_0 h}{8\pi^3} \left\langle \frac{3 \cos^2(\theta(\mathbf{r}_k, \mathbf{r}_H)) - 1}{2r_k^3} \right\rangle, \quad (1)$$

where γ_{k_1} and γ_{k_2} are the gyromagnetic ratios of the two spins, μ_0 is the magnetic permittivity of vacuum, h is Planck's constant, and r_k is the distance between the two spins. The averaging in Eq. 1 encompasses ensemble as well as time averaging.

Chemical shift anisotropies (CSAs; Mehring 1983; Mason 1993) and residual quadrupolar couplings (RQCs; Mehring 1983; Moltke and Grzesiek 1999) are other types of NMR observables that yield information about the global alignment of atoms and their local environment in a molecule relative to a magnetic field and, therefore, indirectly relative to each other. The measured anisotropic contribution δ_k^{an} of the local environment to the chemical shift δ_k of a nucleus and the residual quadrupolar interaction of nuclei of angular momentum $I > \frac{1}{2}$ with their electronic environment have the same angle dependence as RDCs, and can therefore be expressed by the same alignment tensor (Mehring 1983; Moltke and Grzesiek 1999).

Theory

Description of alignment

In order to extract structural information from RDCs it is necessary to introduce a parametrization of the averaged

L. N. Wirz
Centre for Theoretical Chemistry and Physics, The New Zealand
Institute for Advanced Study, Massey University Auckland,
Private Bag 102904, 0632 Auckland, New Zealand

J. R. Allison (✉)
Centre for Theoretical Chemistry and Physics, Institute of
Natural and Mathematical Sciences, Massey University
Auckland, Private Bag 102904, 0632 Auckland, New Zealand
e-mail: j.allison@massey.ac.nz

angle dependence in Eq. 1. This is achieved by defining first, a coordinate system (molecular frame), and then, relative to this, the orientation of the magnetic field vector \mathbf{r}_H and the inter-spin vector \mathbf{r}_k (Fig. 1). The alignment of the molecule relative to the magnetic field can be then described using an alignment tensor [Saupe matrix (Saupe 1964)], leading to the concise formula

$$D_k(\mathbf{r}_k, \mathbf{S}) = D_k^{\max} \sum_{i,j} S_{ij} \Delta_{kij}(\mathbf{r}_k), \quad i, j \in x, y, z, \quad (2)$$

where $D_k^{\max} = -\frac{\gamma_{k_1} \gamma_{k_2} \mu_0 h}{8\pi^3} \frac{1}{r_k^3}$ is the maximum possible RDC for a pair of spins k_1 and k_2 at a given distance r_k , $S_{ij} = \frac{3}{2} \langle \cos \beta_i \cos \beta_j \rangle - \frac{1}{2} \delta_{ij}$ is the (i, j) th component of the alignment tensor and $\Delta_{kij} = \cos \alpha_{ki} \cos \alpha_{kj}$ expresses the (static) orientation of the k th inter spin vector relative to the molecular frame.

CSAs and RQCs allow for an equivalent parametrization in terms of an alignment tensor \mathbf{S} ; the respective equations can be found in, for example, Moltke and Grzesiek (1999) and Losonczi et al. (1999).

Equation 2 is based on three approximations: First, the length of the inter spin vector \mathbf{r}_k , which is generally between a bonded pair of atoms, is assumed to be constant. Second, the orientation of \mathbf{r}_k in the molecular frame is treated as fixed, i.e., all \mathbf{r}_k that are represented by the same \mathbf{S} have a fixed alignment relative to each other, which is commonly summarized as assuming the molecule to be rigid. Third, it is assumed that neither the length nor the alignment of \mathbf{r}_k in the molecular frame affects the alignment of the molecular frame with the magnetic field. An alternative, which leads to analogous equations, is to treat the orientation of the molecular frame with respect to the magnetic field as constant and the alignment of the inter spin vector within the molecule as variable (Hess and Scheek 2003).

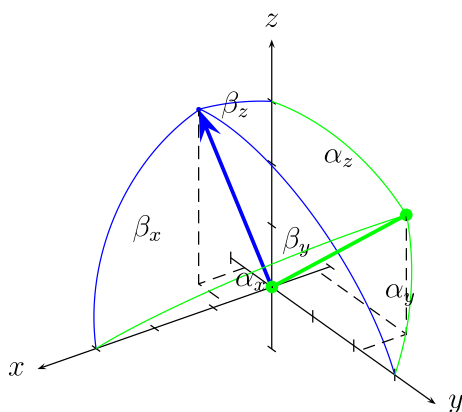


Fig. 1 The inter-spin vector \mathbf{r}_k (green) and magnetic field vector \mathbf{r}_H (blue) with respect to the molecular frame; angles between the inter spin vector and the axes of the molecular frame are denoted as $\alpha_{\{x,y,z\}}$; angles between the magnetic field vector and the axes of the molecular frame are denoted as $\beta_{\{x,y,z\}}$

The alignment tensor \mathbf{S} that describes the weak alignment of the molecular frame in a magnetic field is traceless and symmetric and can therefore be reduced to five independent components, which allows Eq. 2 to be simplified to

$$D_k(\mathbf{r}_k, \mathbf{a}) = D_k^{\max} \sum_h^5 a_h C_{k,h}(\mathbf{r}_k), \quad (3)$$

with

$$\begin{aligned} a_1 &= \frac{3}{2} \langle \cos^2 \beta_x \rangle - \frac{1}{2}, \\ a_2 &= \frac{3}{2} \langle \cos^2 \beta_y \rangle - \frac{1}{2}, \\ a_3 &= \frac{3}{2} \langle \cos \beta_x \cos \beta_y \rangle, \\ a_4 &= \frac{3}{2} \langle \cos \beta_x \cos \beta_z \rangle, \\ a_5 &= \frac{3}{2} \langle \cos \beta_y \cos \beta_z \rangle, \end{aligned} \quad (4)$$

and

$$\begin{aligned} C_{k,1}^{\text{RDC}} &= \cos^2 \alpha_x - \cos^2 \alpha_z, \\ C_{k,2}^{\text{RDC}} &= \cos^2 \alpha_y - \cos^2 \alpha_z, \\ C_{k,3}^{\text{RDC}} &= 2 \cos \alpha_x \cos \alpha_y, \\ C_{k,4}^{\text{RDC}} &= 2 \cos \alpha_x \cos \alpha_z, \\ C_{k,5}^{\text{RDC}} &= 2 \cos \alpha_y \cos \alpha_z. \end{aligned} \quad (5)$$

Analogous expansions of the CSA and RQC can be derived from their respective Hamiltonians (Mehring 1983; Moltke and Grzesiek 1999) to obtain C_k^{CSA} and C_k^{RQC} . Dividing by the respective numerical interaction constant provides the reduced interaction and allows for fitting to an alignment tensor.

In the following, we limit ourselves to describing RDCs for simplicity. However, all sums over RDC interactions can be extended to mixed sums over RDC, CSA and/or RQC interactions with the appropriate substitutions of D_k , $C_{k,h}^{\text{RDC}}$, etc, to obtain a common alignment representation, analogous to the joint evaluation of RDCs and CSAs by Losonczi et al. (1999).

Efficient determination of the optimal alignment tensor

The parametrization described above leads to the following question: given a molecular topology and a number of experimentally measured RDCs, how can the five independent components of the alignment tensor be determined that best describe the molecular orientation, and therefore also best reproduce the experimental results?

Solving this problem is relevant for both assessing ensembles of molecular structures to ascertain how well they agree with experimental data, and biasing simulations

to agree with experimental values, because in either case, fitting of the alignment tensor **S** is required either for each structure of the ensemble or at each cycle [e.g. integration time step, in the case of molecular dynamics (MD) simulations].

Losonczi et al. (1999) proposed an elegant solution, in which an equation $\mathbf{A}\mathbf{a} = \mathbf{b}$ is constructed, where **A** is an $n \times 5$ matrix that holds the five $C_{k,h}$ (Eq. 5) for each interaction, **b** is a vector of size n that holds the reduced experimentally measured RDCs, **a** (**x** in Losonczi et al. 1999) is a vector of size 5 that holds the five independent components of the alignment tensor (Eq. 4) and n is the number of experimentally observed interactions. Diagonalizing **A** by singular value decomposition (SVD) into **U**, **Σ**, and **V**^T such that $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, inverting **Σ**, and solving for **a** gives the exact result for $n = 5$ and the least squares fit for more than five independent components. SVD is numerically very stable in cases where **Σ** has near-zero eigenvalues and even allows for handling underdetermined systems, a property that makes this method very reliable and suitable for systems with few, possibly dependent, interactions, such as parallel inter spin vectors.

For the SVD of a matrix **A** of size $n \times m$ there are two algorithms available: the Golub–Reinsch algorithm, which has a computational cost of $4n^2m + 8nm^2 + 9m^3$ flops, and the R-SVD algorithm, which costs $4n^2m + 22m^3$ flops if all three resulting matrices **U**, **V** and **Σ** are required (Golub and Loan 2012), which is the case for the specific task of solving a system of equations by SVD (Losonczi et al. 1999). For fitting to RDCs, n is the number of interactions (≥ 5) and m is the number of independent alignment tensor components ($=5$). As m is constant in these applications, the time complexity of both algorithms is $\mathcal{O}(n^2)$, that is, they scale quadratically with the number of interactions.

In systems with a large number of observables ($n \gg 5$) that can be expressed in terms of an alignment tensor, underdetermined matrices **A** that lead to zero and near-zero eigenvalues of **Σ** are a negligible risk, but speed becomes relevant, particularly when the fitting needs to be done for every structure in a large ensemble or long trajectory, or at every (2 fs) integration step in a long (μs – ms) MD simulation.

A faster algorithm for obtaining the alignment tensor is linear least squares (LLS; Press et al. 1997; Sass et al. 1999), which scales linearly with the number of RDCs. LLS can be derived by defining a potential

$$V(\mathbf{D}; \mathbf{D}^{\text{exp}}, \mathbf{D}^{\text{max}}, \mathbf{w}) = \sum_k^{\text{RDC}} V_k(D_k; D_k^{\text{exp}}, D_k^{\text{max}}, w_k) \quad (6)$$

that penalizes the deviation of calculated from experimental RDCs. Individual weight factors w_k specific to each interaction allow, for instance, measurements made

with greater confidence to be assigned higher weight. In order to give equal weight to different types of interactions, i.e., RDCs, CSAs and RQCs, as well as RDCs between different types of spins, the potential can be formulated in terms of reduced interactions, e.g., for RDCs, $D_k^{\text{red,calc}} = D_k/D_k^{\text{max}}$ and $D_k^{\text{red,exp}} = D_k^{\text{exp}}/D_k^{\text{max}}$. The harmonic potentials V_k are defined as:

$$V_k = \sum_k^{\text{RDC}} \frac{1}{2} w_k \left(D_k^{\text{red,calc}}(\mathbf{r}_k, \mathbf{a}) - D_k^{\text{red,exp}} \right)^2 \quad (7)$$

The use of reduced RDCs differentiates Eq. 7 from the similarly defined restraining potential by Hess and Scheek (2003) in that it minimizes $\|\mathbf{D}^{\text{red,calc}} - \mathbf{D}^{\text{red,exp}}\|$ instead of $\|\mathbf{D}^{\text{calc}} - \mathbf{D}^{\text{exp}}\|$.

The **a** that minimizes $\|\mathbf{D}^{\text{red,calc}} - \mathbf{D}^{\text{red,exp}}\|$ can then be found by deriving V (Eq. 7) according to **a**,

$$\mathbf{0} = \frac{\partial V}{\partial \mathbf{a}} = \sum_k^{\text{RDC}} w_k (\mathbf{a} \cdot \mathbf{C}_k(\mathbf{r}_k)) \mathbf{C}_k(\mathbf{r}_k) - \sum_k^{\text{RDC}} w_k D_k^{\text{red,exp}} \mathbf{C}_k(\mathbf{r}_k), \quad (8)$$

which can be rewritten as a system of five linear equations $\mathbf{A}\mathbf{a} = \mathbf{f}$, where

$$A_{h,h'} = \sum_k^{\text{RDC}} w_k C_{k,h} C_{k,h'} \quad \text{and} \quad (9)$$

$$f_{h'} = \sum_k^{\text{RDC}} w_k D_k^{\text{red,exp}} C_{k,h'}, \quad (10)$$

and solved by standard techniques such as lower upper (LU) decomposition (Press et al. 1997) to yield **a**.

The time limiting step of this computation is the summation over all interactions to obtain the 25 elements of **A** and the five elements of **f**, which takes a time that is proportional to the number of interactions n . The matrix diagonalization takes constant time because **A** has constant size. Therefore, this algorithm scales as $\mathcal{O}(n)$, that is, linearly with the number of interactions.

Results

Calculation time

To explore how these two algorithms perform in practise, we fitted an alignment tensor to differently sized sets of RDCs measured for human ubiquitin in uncharged bicelles (Ottiger and Bax 1998) using each algorithm and measured the calculation time (Fig. 2). The smallest two sets consist of the 68 N–H^N RDCs and a subset thereof of size 30. The 66 C^α–H^α, 64 C^α–C' and 66 C'–N RDCs were added incrementally to generate sets of increasing size.

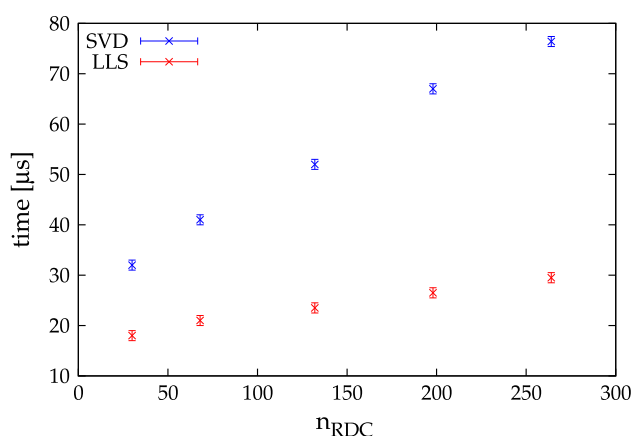


Fig. 2 Time required to obtain a least squares fitted alignment tensor using the SVD or LLS algorithm as a function of the number of RDCs

The calculation times shown in Fig. 2 are for the single function that takes an $n \times 5$ matrix (as an array) of $C_{k,h}$, as well as weights and experimental values of the RDCs, and returns the five independent alignment tensor components. The functions for least squares fitting using SVD or LLS are part of the `gromos++` suite of programs for preparing and analyzing biomolecular simulations (Eichenberger et al. 2011). The SVD and matrix diagonalization are implemented using GSL (Galassi et al. 2009), which uses the Golub–Reinsch SVD algorithm and solves the system of linear equations via LU decomposition. Times were measured using `gettimeofday()`, which has a resolution of $1 \mu\text{s}$, on a single core of a 2.67 GHz Intel/Nehelem processor.

As can be seen in Fig. 2, the LLS algorithm vastly improves the speed of the tensor fitting procedure. However, the quadratic behavior of the SVD algorithm (Galassi et al. 2009; Golub and Loan 2012) is not observed for the numbers of RDCs examined here because the prefactor of the linear term dominates for low numbers of RDCs. Furthermore, there is a constant term in both algorithms which contributes strongly for small n but becomes less relevant for larger n .

Stability

We also sought to explore how robust each algorithm is, as LLS is known to be less stable in the case of near-singular \mathbf{A} (Press et al. 1997). We constructed a case with only five RDCs, in which two of the corresponding inter-nuclear vectors are manipulated so as to become increasingly parallel. We find that for two vectors that differ by an angle of at least 10^{-5} rad, the calculated RDCs are equal to the input on at least the first 7 (11) significant digits using LLS (SVD). When the two vectors differ by an angle of 10^{-7} rad, the calculated RDCs remain equal to the input on

the first 3 digits for LLS and the first 10 digits for SVD. Finally, when two of the inter-nuclear vectors differ by an angle of just 10^{-9} rad, some of the LLS back-calculated RDCs differ from the input values on the first digit (but the relative difference is below 10 %) while SVD still yields values that are correct on the first 8 significant digits.

In the extreme case of inter spin vectors that are parallel to within numerical noise, the matrix diagonalization in LLS may fail. A simple means of allowing a RDC-restrained MD simulation to proceed is to selectively catch the GSL error code and omit the tensor update for that MD step.

We emphasize that the situation described here is rather implausible: not only is it unlikely that any two inter-nuclear vectors in a real molecule will be parallel to such high precision, but most experimentally-measured datasets contain far more than five RDCs.

Other than for the cases described above where one or both of the fitting algorithms become unstable, the calculated RDCs as well as the fitted tensor components are identical up to more than 10 significant figures, which is larger than the precision with which RDCs are typically measured experimentally, thus the results of using either algorithm can be treated as equal.

Availability

The LLS algorithm for obtaining the least squares fitted alignment tensor has been implemented in the GROMOS biomolecular simulation software (Schmid et al. 2011, 2012; Kunz et al. 2012) and the related GROMOS++ suite of analysis programs (program `fit_rdc`) (Eichenberger et al. 2011), all of which are written in C++.

Conclusions

We compare two least squares fitting algorithms for obtaining the alignment tensor that describes the best fit between a molecular structure and a set of experimental RDC, CSA or RQC values. The SVD fit algorithm computes the five independent alignment tensor components \mathbf{a} in a time that scales as $\mathcal{O}(n^2)$, whereas the LLS algorithm scales as $\mathcal{O}(n)$. While the calculation time required for SVD is dominated by the prefactor of the linear term, and thus does not scale as predicted for small n , LLS is still significantly faster for small n . Importantly, we find LLS to also be rather robust even in the limiting case of five RDCs and near-parallel inter nuclear vectors, so that it is unlikely to fail under more realistic conditions. We therefore recommend the use of LLS during MD simulations, as the faster computation speed of this algorithm will be advantageous for restraining simulations to fit with large sets of

NMR observables that may combine RDCs, CSAs and RQCs in the longer simulations that are now possible due to improved parallelization of code and increasing high performance computing resources.

Acknowledgments The authors wish to thank Wilfred van Gunsteren for very helpful discussions on the calculation and interpretation of RDCs. This work was financially supported by a Marsden Fast-Start Award (13-MAU-039, J.R.A.) and a Massey University Doctoral Scholarship (L.N.W.).

References

- Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog Nucl Magn Reson Spectrosc* 46(1):23–61. doi:10.1016/j.pnmrs.2004.11.002
- Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang D, van Gunsteren WF (2011) The gromos++ software for the analysis of biomolecular simulation trajectories. *J Chem Theory Comput* 7:3379–3390. doi:10.1021/ct2003622
- Galassi M, Davies J, Theiler J, Gough B, Jungman G, Alken P, Booth M, Rossi F, Ulerich R (2009) GNU scientific library reference manual, 3rd edn. Network Theory Ltd
- Golub GH, van Loan CF (2012) Matrix computations, 4th edn. Johns Hopkins University Press, Baltimore
- Hess B, Scheek RM (2003) Orientation restraints in molecular dynamics simulations using time and ensemble averaging. *J Magn Reson* 164(1):19–27. doi:10.1016/S1090-7807(03)00178-2
- Kunz APE, Allison JR, Geerke DP, Horta BAC, Hünenberger PH, Riniker S, Schmid N (2012) New functionalities in the gromos biomolecular simulation software. *J Comput Chem* 33:340–353. doi:10.1002/jcc.21954
- Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in nmr structure analysis. *Ann Rev Biophys Biomol Struct* 33(1):387–413. doi:10.1146/annurev.biophys.33.110502.140306
- Losonczi JA, Andrec M, Fischer MW, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 138(2):334–342. doi:10.1006/jmre.1999.1754
- Mason J (1993) Conventions for the reporting of nuclear magnetic shielding (or shift) tensors suggested by participants in the NATO ARW on NMR shielding constants at the university of maryland, college park, July 1992. *Solid State Nucl Magn Reson* 2(5):285–288. doi:10.1016/0926-2040(93)90010-K
- Mehring M (1983) Principles of high resolution NMR in solids, 2nd edn. Springer, Berlin. doi:10.1007/978-3-642-68756-3
- Moltke S, Grzesiek S (1999) Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. *J Biomol NMR* doi:10.1023/A:1008309630377
- Ottiger M, Bax A (1998) Determination of relative N–H^N, N–C', C^α–C', and C^α–H^α effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J Am Chem Soc* 120(12):334–341. doi:10.1021/ja9826791
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1997) Numerical recipes in C. The art of scientific computing, 2nd edn. Cambridge, Cambridge University Press
- Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* 104(8):3519–3540. doi:10.1021/cr030419i
- Sass J, Cordier F, Hoffmann A, Rogowski M, Cousin A, Omichinski JG, Löwen H, Grzesiek S (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. *J Am Chem Soc* 121(10):2047–2055. doi:10.1021/ja983887w
- Saupe A (1964) Kernresonanzen in kristallinen Flüssigkeiten und in kristallinflüssigen Lösungen. *Zeitung für Naturforschung* 19a:161–171
- Schmid N, Allison JR, Dolenc J, Eichenberger AP, Kunz APE, van Gunsteren WF (2011) Biomolecular structure refinement using the gromos simulation software. *J Biomol NMR* 51:265–281. doi:10.1007/s10858-011-9534-0
- Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF (2012) Architecture, implementation and parallelisation of the gromos software for biomolecular simulation. *Comput Phys Commun* 183:890–903