Article

Determination of Relative Configuration from Residual Chemical Shift Anisotropy

Nilamoni Nath, Manuel Schmidt, Roberto R Gil, R. Thomas Williamson, Gary E Martin, Armando Navarro-Vázquez, Christian Griesinger, and Yizhou Liu

J. Am. Chem. Soc., **Just Accepted Manuscript** • Publication Date (Web): 13 Jun 2016

Determination of Relative Configuration from Residual Chemical Shift Anisotropy

Nilamoni Nath†[1], Manuel Schmidt†[1], Roberto R. Gil[2], R. Thomas Williamson[3], Gary E. Martin[3], Armando Navarro-Vázquez*[4,5], Christian Griesinger*[1] and Yizhou Liu†*[3]

[1] Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

[2] Department of Chemistry, Carnegie Mellon University, 4400 Fifth Ave, Pittsburgh, PA 15213 USA

[3] Process Research and Development, NMR Structure Elucidation Group, Merck & Co., Inc., Mail Stop RY800-C163, 126 E Lincoln Ave, Rahway, NJ 07065 USA

[4] Departamento de Química Fundamental, Universidade Federal de Pernambuco, Cidade Universitária, Recife, PE 50740-560, Brazil,

[5] Institute of Organic Chemistry and Institute for Biological Interfaces Karlsruhe, Institute of Technology (KIT), Fritz-Haber-Weg 6, 76131 Karlsruhe, Germany

Abstract

Determination of relative configuration is frequently a rate-limiting step in the characterization of small organic molecules. Solution NMR–based Nuclear Overhauser Effect and scalar *J*-coupling constants can provide useful spatial information but often fail when stereocenters are separated by more than 4-5 Å. Residual dipolar couplings (RDCs) can provide a means of assigning relative configuration without limits of distance between stereocenters. However, sensitivity limits their application. Chemical shift is the most readily measured NMR parameter and partial molecular alignment can reveal the anisotropic component of the chemical shift tensor, manifested as residual chemical shift anisotropy (RCSA). Hence, ^{13}C RCSAs provide information on the relative orientations of specific structural moieties including non-protonated carbons and can be used for stereochemical assignment. Herein, we present two robust and sensitive methods to accurately measure and apply ^{13}C RCSAs for stereochemical assignment. The complementary techniques are demonstrated with five molecules representing differing structural classes.

Introduction

Small molecule structural diversity is rapidly expanding through traditional natural products and medicinal chemistry, CH-bond functionalization chemistry, and diversity–oriented organic and biomolecular syntheses. This rapid expansion of chemical space is steadily increasing the demand and time pressure to efficiently and accurately elucidate structures. In this respect, assigning relative configuration continues to be one of the most time-consuming and error-prone[1,2] hurdles despite a growing list of approaches which includes NOE, *J*-based configuration analysis (JBCA), and correlation of chemical shifts and scalar coupling constants with DFT calculations, among other methods. Hence, robust, broadly applicable and accurate new methods are in demand to facilitate configuration analysis and avoid incorrect assignments.[1] Among the most recently developed techniques are those that involve the NMR analysis of molecules weakly aligned in anisotropic media. Partial molecular alignment in these media re-introduces measureable anisotropic NMR phenomena such as dipolar couplings and chemical shift anisotropy for molecules in solution. These partially averaged anisotropic interactions, commonly referred to as residual dipolar couplings (RDCs)[3] and residual chemical shift anisotropy (RCSAs),[4-6] carry rich structural information. Both RDCs and RCSAs were recognized and described decades ago but utilization was largely limited to molecules strongly aligned when dissolved in thermotropic liquid crystals.[7-9] The break-through in this field was the concept of using an alignment medium to transfer weak molecular ordering to arbitrary molecules of interest. Indeed, development of weak alignment media and measurement methods over nearly two decades has led to a boom in the utilization of anisotropic NMR data, particularly RDCs, as a generally applicable NMR parameter for small molecule applications[10-15] and for structural and dynamics studies of bio-molecules.[3,16-19]

As noted above, a critical component of the RDC/RCSA technology is the alignment medium used for molecular ordering. In addition to liquid crystal media,[20-24] one of the most versatile alignment media developed thus far is the so-called “Strain-induced Alignment in a Gel” (SAG) system that was initially introduced for NMR analyses in water[25-27] and further modified for utilization with organic solvents.[28-31] The alignment gel is made of a highly stable, cross-linked polymer, whose chemical composition can be altered such that it swells in a wide range of NMR

solvents. Molecular alignment in these gels is induced by compressing or stretching the gel matrix within the NMR tube in an adjustable fashion.[27,28]

The aforementioned ^{1}H-^{13}C RDCs, either one-bond[10-15] or long-range,[32] report on the relative orientations of inter-atomic vectors in the molecule. In contrast, RCSAs report on the relative orientations of chemical shielding tensors of individual atoms. RCSAs, since they are reflected in the chemical shift, are measured with high sensitivity thus facilitating measurement of minute samples for which RDC measurement would either be much more challenging or impossible.[33] Both RDCs and RCSAs provide orientational restraints relative to a global reference frame and are therefore complementary to conventional, distance-constrained NMR measurements of NOE (or ROE) or *J*-couplings. In comparison to RDCs, RCSAs have the further advantage of providing orientation information on quaternary carbons, which otherwise can only be reached through less sensitive long-range RDC measurements.[32] The ability to address the orientation of quaternary carbons is especially important for proton-deficient, drug-like molecules where NOE, *J*-coupling, and one-bond RDC data may be insufficient. Furthermore, the chemical shift tensor information necessary for RCSA analyses can be obtained with sufficient accuracy at low computational cost using density functional theory (DFT) calculations.[34] However, until now, RCSAs have only occasionally been used for the validation[35] and refinement of the structure of proteins[36] and nucleic acids,[37,38] and their application to the analysis of small molecules is even less frequent with only a scant few reports.[4,5] In the current work, we have removed the primary obstacles that previously impeded the use of RCSAs in a convenient, robust, and routine manner.

Superficially, RCSA data are more attractive to use than RDCs because chemical shift is the simplest NMR parameter to measure. The practical challenge, however, is to reliably eliminate isotropic chemical shift changes during molecular alignment. Any viable method must be able to distinguish RCSA shifts due to weak alignment from isotropic shifts due to chemical environment variations between different alignment conditions. The key to accurately measure RCSAs is to collect the data without altering environmental conditions. This goal was previously achieved by using a rubber-based stretching apparatus in Kuchel's group[39] and later improved by Luy and coworkers.[40] Unfortunately, these rubber-based devices impose experimental limitations. For example, chloroform, a commonly utilized solvent for small molecule NMR studies, cannot be employed because it degrades the rubber used in the stretching device. Other

methods that have been evaluated to measure RCSA's included variable-angle sample spinning (VASS)[41,42] and variable-angle NMR spectroscopy,[5] but the specialized hardware required is not routinely available to most chemists. Clearly, generally applicable methods for RCSA measurement are needed.

In this work, we introduce two robust and complementary methods to measure RCSAs and RDCs. The first relies on a two-stage NMR tube originally designed for biomolecules aligned in polymeric gels. By stretching the gel inside each stage, two different alignment strengths are obtained.[6] Using this method, the composition of the solvent, gel, and analyte remain constant between the maximum and minimum alignment thereby avoiding problems associated with isotropic chemical shifts changes. The only drawback to this approach is that the strong alignment condition is achieved in a smaller diameter section of the tube leading to lower signal-to-noise as already noted for Kuchel's method.[39] Our second method achieves two alignment strengths using a compression device. In the compression implementation, the same active volume is used for strong and weak alignment, yielding similar signal-to-noise for both conditions. However, due to solvent and gel composition changes experienced under the different alignment strengths, the isotropic shift must be compensated. This adjustment can be achieved by a robust post acquisition correction as discussed in the method section. For this study, both methods were tested with small molecules aligned in chloroform-compatible poly(methyl methacrylate) (PMMA) gels,[28] but the same tools can also be applied to constrain gels compatible with other solvents.

Strychnine was chosen as the model compound for the initial demonstration of these new methods for two reasons. First, data quality and accuracy can be conveniently evaluated by direct comparison to published values.[5] Second, the utility of RDCs and RCSAs in the determination of relative configuration can be assessed for strychnine by comparing the anisotropic data to previous reports using other NMR-based parameters.[43] The robustness of both methods was also tested on estrone, mefloquine, retrorsine, and menthol. For all molecules tested, we were able to determine the relative configuration using RCSAs alone. We also introduce a modified quality factor Q_{CSA} which is based on the conventional *Q*-factor, but takes into account the significantly varying chemical shift anisotropy values of different carbon atoms, and therefore offers enhanced capabilities to distinguish different relative configurations based on RCSA data.

Method section

(a) Stretching and Compressing Devices:

For both devices, we used PMMA gels to measure RCSAs from two different alignment conditions, labeled "max" (alignment tensor $\hat{A}^{max}$) and "min" ($\hat{A}^{min}$) for maximum and minimum alignment, respectively. The first method employs a stretching device, consisting of an NMR tube with a "wide-bore" section of 4.2 mm inner diameter (ID) connected to a "narrow-bore" section of 3.2 mm ID. The "min" data are collected first with the gel situated in the 4.2 mm ID section before the gel was transferred to and stretched into the narrow section to acquire the "max" data (Figure 1a-e). Alternatively, in the compression method, gel compression was achieved inside a standard 4.2 mm ID tube by extending or retracting the piston of a compression device manufactured by New Era Enterprises, Inc.[28]

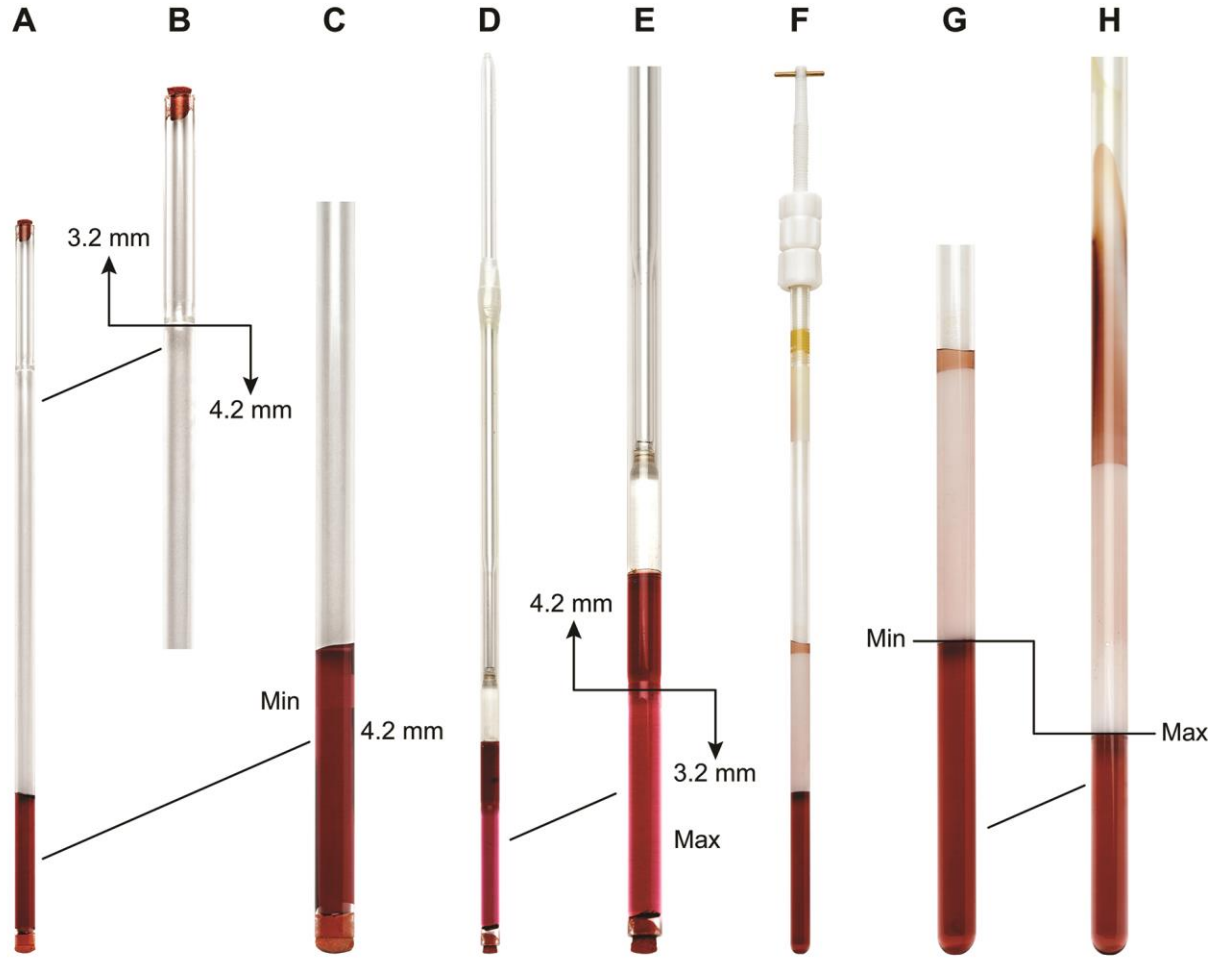

Figure 1. Photographs of the stretching and compression devices. The PMMA gel used in this figure is colored for visualization with the violet colored alkaloid cyptolepine dissolved in deuterochloroform. Panels A-E pertain to the tube designed for stretching the gel. Panels F-H show the apparatus for compressing the gel used for the RCSA measurements. A.) Tube with the PMMA gel positioned in the 4.2 mm diameter section of the NMR tube. Rubber stoppers are used to close the top and bottom of the tube. B.) Expansion showing a close-up of the top of the tube from panel A. The line indicates the location where the inner diameter of the tube changes from 3.2 to 4.2 mm. The 3.2 mm diameter section of the tube is used for maximum stretching while the 4.2 mm section of the tube give minimum stretching of the gel. C.) Close-up of the tube with the gel in the 4.2 mm diameter section of the tube. D.) Full tube with the gel stretched in the 3.2 mm diameter segment of the tube. The difference in color defines where the inner diameter changes from 3.2 to 4.2 mm. The narrower segment is lighter in color than the larger ID segment and is centered on the radiofrequency coil of the probe. E.) Close-up of the segment of the tube with the gel in the 3.2 mm diameter section. The difference in color where the inner diameter of the tube changes is more readily visible in this panel. F.) Photograph of the full assembly used for compressed gel sample measurements. G.) Close-up showing the gel in the minimum compression condition. H.) Close-up showing the gel in the maximum compression condition. Panels G.) and H.) are scaled identically. The difference in the vertical dimension corresponds to the difference in compression of the gel in the minimum and maximum compression conditions.

Formulations of RCSAs and RDCs have been extensively described in literature.[6,35,44,45] In any arbitrary molecular frame, the RCSA of any nuclei *i* is given by:[44]

$$RCSA^{i}(ppm) = \sum_{\alpha,\beta=x,y,z} A_{\alpha\beta}\, CSA^{i}_{\alpha\beta} \qquad \text{Eq (1)}$$

where $A_{\alpha\beta}$ (equals to $\frac{2}{3}S_{\alpha\beta}$, representing the Saupe order matrix) are the molecular alignment tensor matrix elements. We notated the chemical shift tensor elements as $CSA_{\alpha\beta}$ instead of δ to avoid confusion since δ symbol is reserved for the chemical shifts observed in the spectra. Similarly, the RDC between two nuclear spins *i* and *j* is given by:

$$RDC_{ij} = -\frac{3\mu_0 h\gamma_i\gamma_j}{16\pi^3 r_{ij}^5}\sum_{\alpha,\beta=x,y,z} A_{\alpha\beta} r_{ij}^{\alpha} r_{ij}^{\beta} \qquad \text{Eq (2)}$$

where r_{ij}^{α} and r_{ij}^{β}represent the projections of the internuclear vector $\vec{r}_{ij}$ along the α and β axes, i.e. x,y,z axes, respectively.

The RCSA of a nucleus *i* was determined as the difference in chemical shift measured under the "max" and "min" alignment conditions (Eq. (3)) with respect to a chosen reference nucleus (δ_{ref}).[4] For the stretched gel, we used tetramethylsilane (TMS) for internal chemical shift

referencing. Referencing removes factors that impact all chemical shifts equally, including potential changes of the overall effective field due to susceptibility changes as well as mis-referencing due to locking ambiguity associated with the split ^{2}H signal of $CDCl_3$. Due to the difference character of the measurement, RCSA is manifested as Δ*RCSA* between two alignment conditions as shown in Eq.(3).

$$\Delta RCSA_i = \left(\delta_i^{aniso} - \delta_{ref}^{aniso}\right)^{max} - \left(\delta_i^{aniso} - \delta_{ref}^{aniso}\right)^{min}$$

$$= \sum_{\alpha,\beta=x,y,z}\left[\left(A_{\alpha\beta}^{max} - A_{\alpha\beta}^{min}\right)CSA_{i,\alpha\beta} - \left(A_{TMS,\alpha\beta}^{max} - A_{TMS,\alpha\beta}^{min}\right)CSA_{TMS,\alpha\beta}\right]$$

Eq. (3)

Because TMS is highly symmetric, it has negligible alignment under the conditions used here and therefore $\left(\hat{A}_{TMS}^{max} - \hat{A}_{TMS}^{min}\right) \approx 0$. Furthermore the carbon of TMS has small chemical shielding anisotropy, *i.e.* $\widehat{CSA}_{TMS} \approx 0$. Consequently, the product $(\hat{A}_{TMS}^{max} - \hat{A}_{TMS}^{min})\widehat{CSA}_{TMS}$ is negligibly small, and hence Eq. (3) simplifies to the following form:

$$\Delta RCSA_i = \sum_{\alpha,\beta=x,y,z}\left(A_{\alpha\beta}^{max} - A_{\alpha\beta}^{min}\right)CSA_{i,\alpha\beta} = \Delta\delta_i \qquad \text{Eq. (4)}$$

For the stretching apparatus, molecular alignment is induced by elongating the gel through mechanical force. Analyte concentration is kept constant and thus no concentration-related correction is needed. Therefore, the experimentally determined chemical shift referenced to TMS, $\Delta\delta_i$, is identical with the desired $\Delta RCSA_i$, as expressed in Eq. (4).

To evaluate the reliability of the ΔRCSA-derived alignment tensors, one-bond ^{13}C-^{1}H ΔRDCs were also measured as differences in ^{13}C-^{1}H total couplings (T_{ij}) between the two stages of alignment for comparison.

$$\Delta RDC_{ij} = T_{ij}^{max} - T_{ij}^{min} = -\frac{3h\mu_0\gamma_i\gamma_j}{16\pi^3\left(r_i - r_j\right)^5}\overrightarrow{\left(r_i - r_j\right)}^T\left(A^{\max} - A^{\min}\right)\overrightarrow{\left(r_i - r_j\right)} \qquad \text{Eq. (5)}$$

All equations above are expressed in an arbitrary molecular frame so that DFT structure and tensor outputs can be directly used following these equations for the prediction of the RCSAs and RDCs.

(b) **Isotropic Shift Compensation in the Compression Device:**

Our initial data from compression device, measured as the chemical shift differences between minimal and maximal alignment after proper referencing, agreed poorly with predicted RCSA values, indicating the presence of a compression-related isotropic chemical shift change that was not accounted for (see results). Unlike the stretching device where the gel can freely elongate through the tube's open end while being stretched, the gel in the compression device is restricted in all directions when maximally compressed so that a small fraction of liquid is most probably expelled from the gel under pressure. Liquid leaving the gel increases the gel-to-analyte ratio inside the gel matrix causing an isotropic change, represented by $\Delta\delta_i^{iso}$ in Eq(6), in addition to the anisotropic RCSA shift. To eliminate this isotropic contribution, we derived a robust post-acquisition correction procedure starting from first principles (see supporting information) that eliminates the isotropic contribution and produces RCSA values in satisfactory agreement with predictions (see results). The correction method capitalizes on the fact that a certain amount of isotropic analyte is always present in the compression device as the gel does not fill the entire sample space even under maximal compression (Figure 2), whose NMR signals are used to internally calibrate the correction equation. In practice, the isotropic signals are easily distinguishable from the corresponding anisotropic signals as their intensities decrease upon compression (Figure 2e-h). In addition, the isotropic signals of the analyte overlap very well when two spectra between two alignment conditions are referenced with respect to the isotropic chloroform signal. Here we also used an alternative but equally valid chemical shift referencing method. *In lieu* of referencing to TMS as in the stretching device, we instead set the analyte carbon with the smallest computed CSA, e.g. C8 of estrone in Figure 2e, as the reference nucleus.

$$\Delta\delta_i = \Delta RCSA_i + \Delta\delta_i^{iso} = \sum_{\alpha,\beta=x,y,z}\left[\left(A_{\alpha\beta}^{max} - A_{\alpha\beta}^{min}\right)\left(CSA_{i,\alpha\beta} - CSA_{ref,\alpha\beta}\right)\right] + \Delta\delta_i^{iso} \qquad \text{Eq. (6)}$$

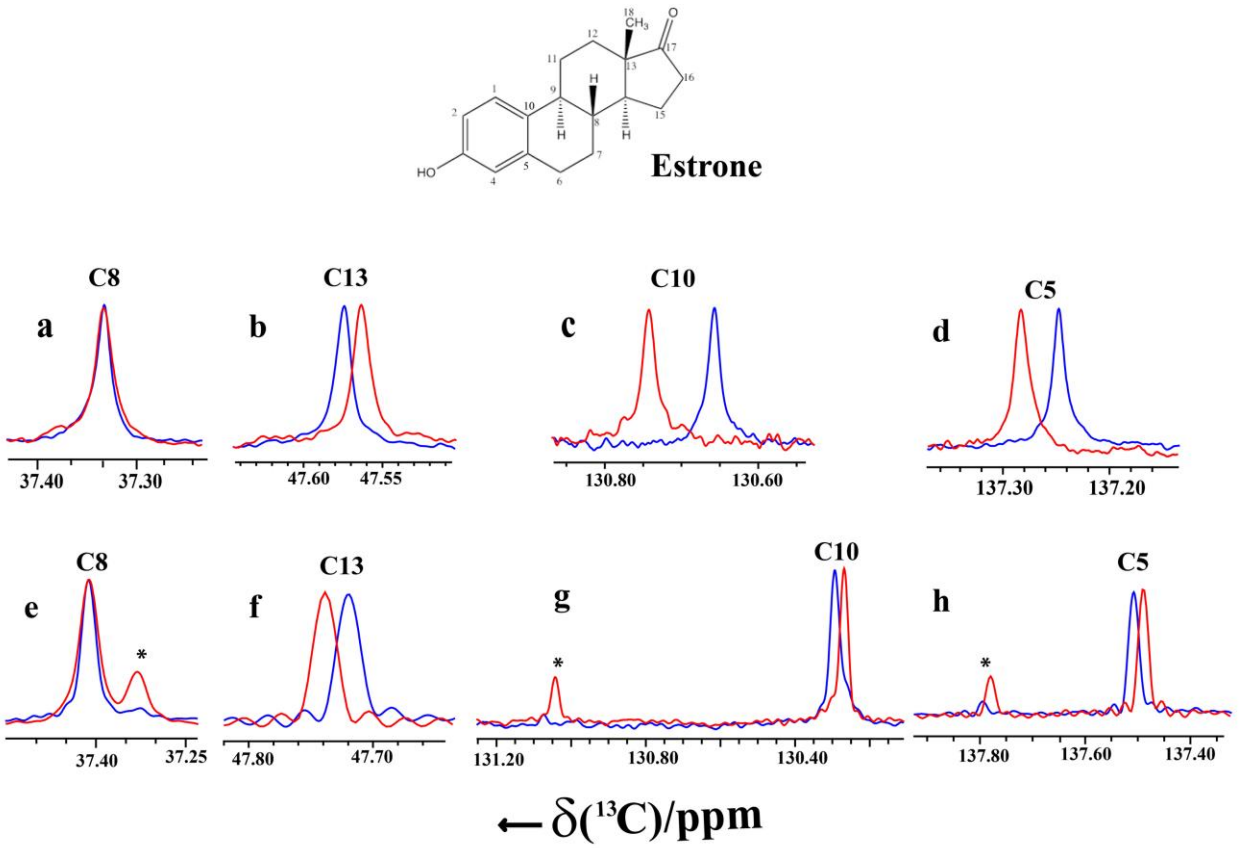

Figure 2. ^{13}C RCSAs obtained with stretching and compression devices**.** Stretching device: Panels a-d show three resonances extracted from the ^{13}C-{^{1}H} 150 MHz NMR spectra of estrone in the narrow (blue color)) and wide (red) sections. Compression device: Panels e-g show resonances from the ^{13}C-{^{1}H} 225 MHz NMR spectra from estrone observed under minimum (red) and maximum (blue) compression. The C8 resonance shown in panel (e) was used as the reference resonance. Note the presence of both isotropic (marked with an asterisk) and anisotropic signals for some carbons. Spectra recorded with minimum alignment was recorded under almost complete relaxation of the PMMA gel ($CDCl_3$ $\Delta\nu_Q$ = 5.5 Hz) and had approximately 67% of the analyte in the gel. Maximum compression ($\Delta\nu_Q$ = 48 Hz) increased the molar fraction of the gel-residing population to 15%. It is evident that the RCSA shifts the resonance in opposite directions for stretched and compressed gels due to their inherent anti-correlation, although in the latter case the presence of isotropic contribution may violate this relationship.

(c) Inter-tensor angle

As RDC and RCSA depend on the same alignment tensor, theoretically an inter-tensor angle[46] (θ) of 0° is expected between tensors derived from RCSA and RDC data. To check this consistency, the angle θ was calculated from the normalized scalar product of their matrix representations in a common frame of reference with the equation:[45]

$$\cos(\theta) = \frac{\langle A^{RCSA}|A^{RDC}\rangle}{|A^{RCSA}||A^{RDC}|} = \frac{\sum_{\alpha\beta=x,y,z} A_{\alpha\beta}^{RCSA} A_{\alpha\beta}^{RDC}}{\sqrt{\sum_{\alpha\beta=x,y,z}(A_{\alpha\beta}^{RCSA})^2}\sqrt{\sum_{\alpha\beta=x,y,z}(A_{\alpha\beta}^{RDC})^2}} \qquad \text{Eq. (7)}$$

(d) A CSA-size based *Q*-factor for RCSA-based stereochemical analysis: Q_{CSA}

One-bond ^{1}H-^{13}C bond-length is very stable and therefore dipolar coupling is indifferent to the specific carbon involved and different ^{1}H-^{13}C bond vectors are equally weighted in the *Q*-factor based analysis. In contrast, CSA tensors vary widely, particularly between sp^3 and sp^2 carbons: according to DFT calculations for the molecules examined in this work, CSAs of sp^2 carbons are 4 to 10 times larger than those of sp^3 carbons. Since the configuration of a molecule is encoded in the network of all carbons, the ideal *Q* value should scale the CSA values to the same size so that only orientation information is reflected. Towards this goal, we employed the following procedure: First the alignment tensor is derived by fitting all ΔRCSAs to the DFT-computed CSA tensor through the Singular Value Decomposition (SVD) method.[44] Then a new quality factor, Q_{CSA}, is calculated by scaling both experimental and back-calculated ΔRCSAs by corresponding atom's chemical shift anisotropies, using the formula given below, where $CSA_{i,ax}$ equals $\sigma_{33}-(\sigma_{22}+\sigma_{11})/2$ and the chemical shielding Eigen-values $\sigma_{11\text{-}33}$ are obtained from DFT.

$$Q_{CSA} = \sqrt{\frac{\sum\left((\Delta RCSA_{i,ax}^{exp} - \Delta RCSA_{i}^{theo})/CSA_{i,ax}\right)^2}{\sum\left(\Delta RCSA_{i}^{exp}/CSA_{i,ax}\right)^2}} \qquad \text{Eq. (8)}$$

Results and Discussion

(a) Strychnine:

As an initial proof of concept, we collected ΔRCSA and ΔRDC data for strychnine using both the stretching and compression devices. Twenty ^{13}C ΔRCSAs were measured from the stretching device through chemical shift differences in 1D ^{13}C-{^{1}H} 150 MHz spectra collected under two alignment conditions (Figure. S1). The range of ΔRCSA shifts of the resonances ranged between -1.1 to 10.5 Hz (Table S1). As expected, aromatic and carbonyl carbons tended to exhibit larger ΔRCSAs than aliphatic carbons, reflecting their respective differences in the CSA tensors. To further validate the ΔRCSA data, we also collected sixteen ^{1}H-^{13}C ΔRDCs under the same alignment conditions. We performed alignment tensor analysis using ΔRCSA and ΔRDC data by SVD[44] with an in-house written computer program. ΔRDCs ranged between approximately ±18

Hz, a span that is approximately four times that of RCSAs at 150MHz (Table S2). The utility of these data in the assignment of configuration was tested on thirteen different strychnine structures generated from DFT-optimized geometries. Due to the structural constraints imposed by the polycyclic system, only thirteen plausible structures were generated from the thirty-two possible diastereoisomers. The diastereomers were labelled via the *R* or *S* configuration of carbons C7, C8, C12, C13, C14, and C16, respectively. Hence, *RSSRRS* represents the true configuration of *7R,8S,12S,13R,14R,16S*. As shown in Figure S4, the alignment tensors derived from ΔRCSAs and ΔRDCs are similar with an inter-tensor angle[45,46] of 14°, corresponding to a normalized inner product of 0.97, and simultaneously fitting both ΔRCSAs and ΔRDCs only slightly raises the ΔRDC *Q*-factor. Discrepancy between RDC and RCSA-derived alignment tensors may reflect errors in tensor determination with a relatively small pool of data of limited orientation sampling, and potentially inaccuracies in DFT-derived structure models and chemical shielding tensors. Discrimination based on ΔRCSA-derived *Q* and Q_{CSA} values is presented in Figure 3(b), in good agreement with results obtained from ΔRDC-derived *Q* values as well as an analysis utilizing $^{1}J_{CC}$ scalar coupling constants (see the Figure S5 in the supplement).[43] The correct configuration *RSSRRS* has a sufficiently lower *Q* (Q_{CSA}) factor of 0.050 (0.174) to differentiate it from other incorrect configurations with the next closest $Q = 0.100$ ($Q_{CSA} = 0.399$) corresponding to C12 epimer *RSRRRS.*

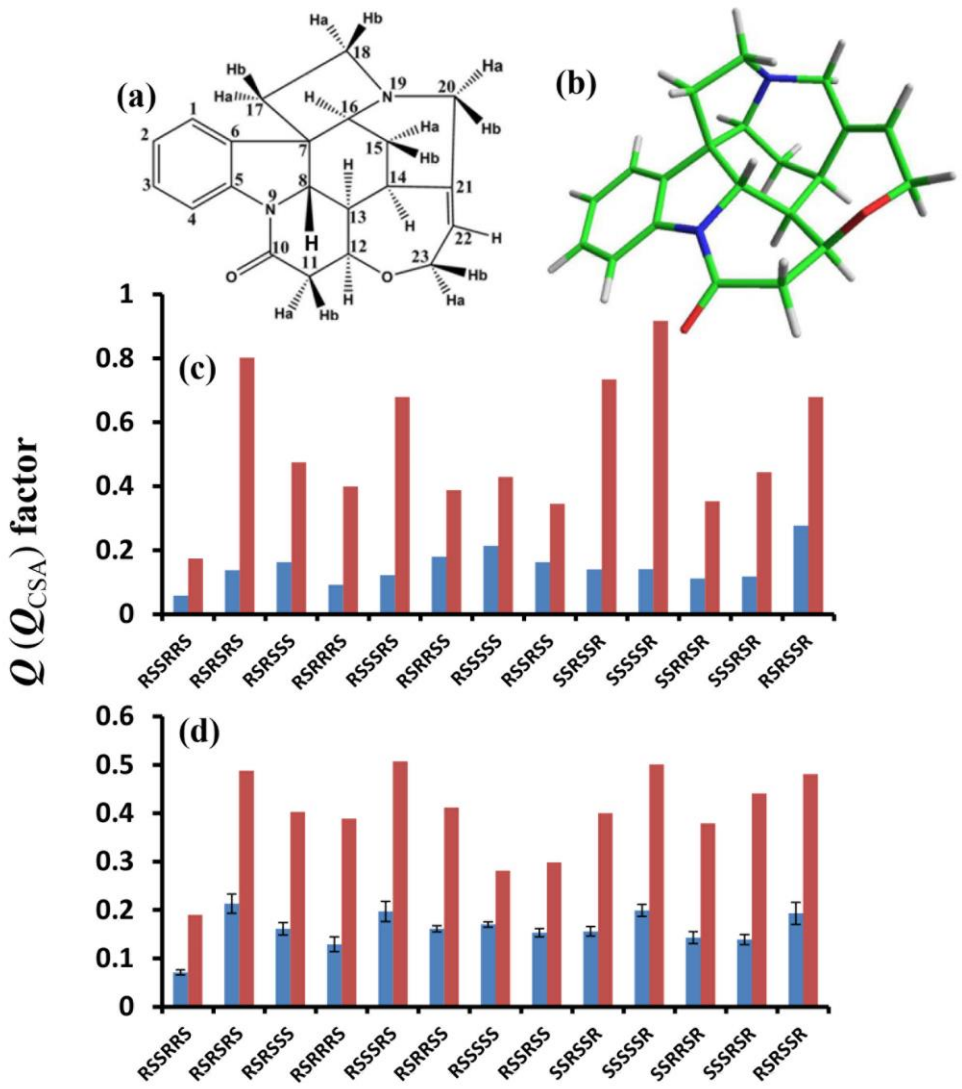

Figure 3. ΔRCSA-based stereochemical analysis of strychnine. a-b) Molecular structure of strychnine with atomic numbering and its 3D structure. The quality of the fit is expressed as a quality factor (*Q*) where the correct structure is expected to have the lowest *Q* value. (c) Results from the stretching device: *Q* (blue) and Q_{CSA} (red) factors calculated for lowest-energy structures of different diastereoisomers from DFT calculation using only ΔRCSAs. (d) Results from the compression device: *Q* factors obtained from the analysis of experimental ΔRCSAs and the 13 possible configurations. Statistical error is calculated by removing 10% of the data randomly and calculating the *Q* values 10 times.

For the compression device, ΔRCSAs ranged from -29.2 to 2.9 Hz at a ^{13}C observation frequency of 225 MHz. As expected, the largest ΔRCSAs values were again observed for the aromatic and

carbonyl carbons. As an example, the ΔRCSA of the aromatic carbon atom (C1) was -29.2 Hz and that of carbonyl atom (C10) was -16.4 Hz. Altogether, twenty ΔRCSAs and twenty ΔRDCs were measured. The *Q*-factor based stereochemical differentiation was performed through SVD method with the MSpin program[47] using only ΔRCSA data for the same thirteen strychnine diastereomers as used with the stretched ΔRCSA data. The ΔRDC data were used only to validate the ΔRCSA-derived alignment tensor. The correct configuration, *RSSRRS*, was found to have the lowest *Q* (Q_{CSA}) factor of 0.071 (0.190). The next best configuration, in this case the epimer at C12 (*RSRRRS*), had a *Q* (Q_{CSA}) factor of 0.129 (0.389); *Q*-factors for other configurations were larger (See bar diagram with their corresponding errors in Figure 3(c)). Given uncertainties ±0.005 for the *Q* values (see caption of Figure 3) for the first two closest structures, the *Q*-factor difference of 0.058 is significant enough for assignment of configuration. For strychnine we found very small differences between the isotropic and minimum-alignment anisotropic chemical shifts and consequently, no isotropic correction was performed. The reliability of the alignment tensor derived from uncorrected ΔRCSA was validated through its good agreement with the ΔRDC-derived alignment tensor as indicated by an inter-tensor angle[45,46] of only 7.5°.

(b) Estrone:

In the second example, we collected ΔRCSA data for estrone using both the compression and stretching methods. This molecule was previously the subject of a ΔRCSA analysis using data obtained with the Kuchel device.[39,40] In that study, the discrimination of configuration was not possible from the ΔRCSAs alone but rather required combining both ΔRCSA and ΔRDC data.[4] Here using the compression device, we experimentally measured fifteen $\Delta\delta_i$ values. First, we neglected isotropic contributions by assuming that $\Delta\Delta\delta_i = \Delta RCSA_i$ and SVD-fitted all experimental ΔRCSAs to the DFT-calculated CSA tensors. A very poor fit was obtained for both isomers [*Q* (Q_{CSA}) = 0.510 (0.420) for estrone and *Q* (Q_{CSA}) = 0.653 (0.853) for 18-*epi*-estrone] with particularly large mismatches for some of the carbons (Figures 4b and c). However, it was observed that the outliers in Figure 4b corresponded to carbons exhibiting large chemical shift differences between anisotropic and isotropic signals under minimum alignment, indicating these

carbons might have been affected by strong isotropic contributions $\Delta\delta_i^{iso}$ (blue and red data points; refer also to Figure 2). Indeed, after isotropic correction (Eq S11), significant improvement was observed. The Q (Q_{CSA})-factor dropped to 0.114 (0.347) for estrone (Figure 4d) but remained high, Q (Q_{CSA}) =0.441 (0.973), for *epi*-estrone (Figure 4e). As expected, the greatest improvement was seen for carbons that are most affected by large isotropic contributions (blue and red data points in Figure 4d, whereas the correlation of ΔRCSAs for *epi*-estrone was significantly worse (Figure 4e). Hence, the distinction between estrone and *epi*-estrone is now possible using ΔRCSA data alone with this method. The robustness of the ΔRCSA data was further confirmed by a small inter-tensor angle of 7.5º between RCSA- and RDC-derived alignment tensors. The improved discrimination resulted from higher data accuracy of the current method as indicated by a Q value that is 1/3 of that observed previously for data obtained with the Kuchel device.[4] In addition, the alignment tensors are orientated differently between the compression and Kuchel devices, with an inter-tensor angle of 75°, which also accounts for differences in the effectiveness of differentiation. The 75° does reflect linearly independent alignment tensors in these two studies, which will be reported in more details in follow-up works.

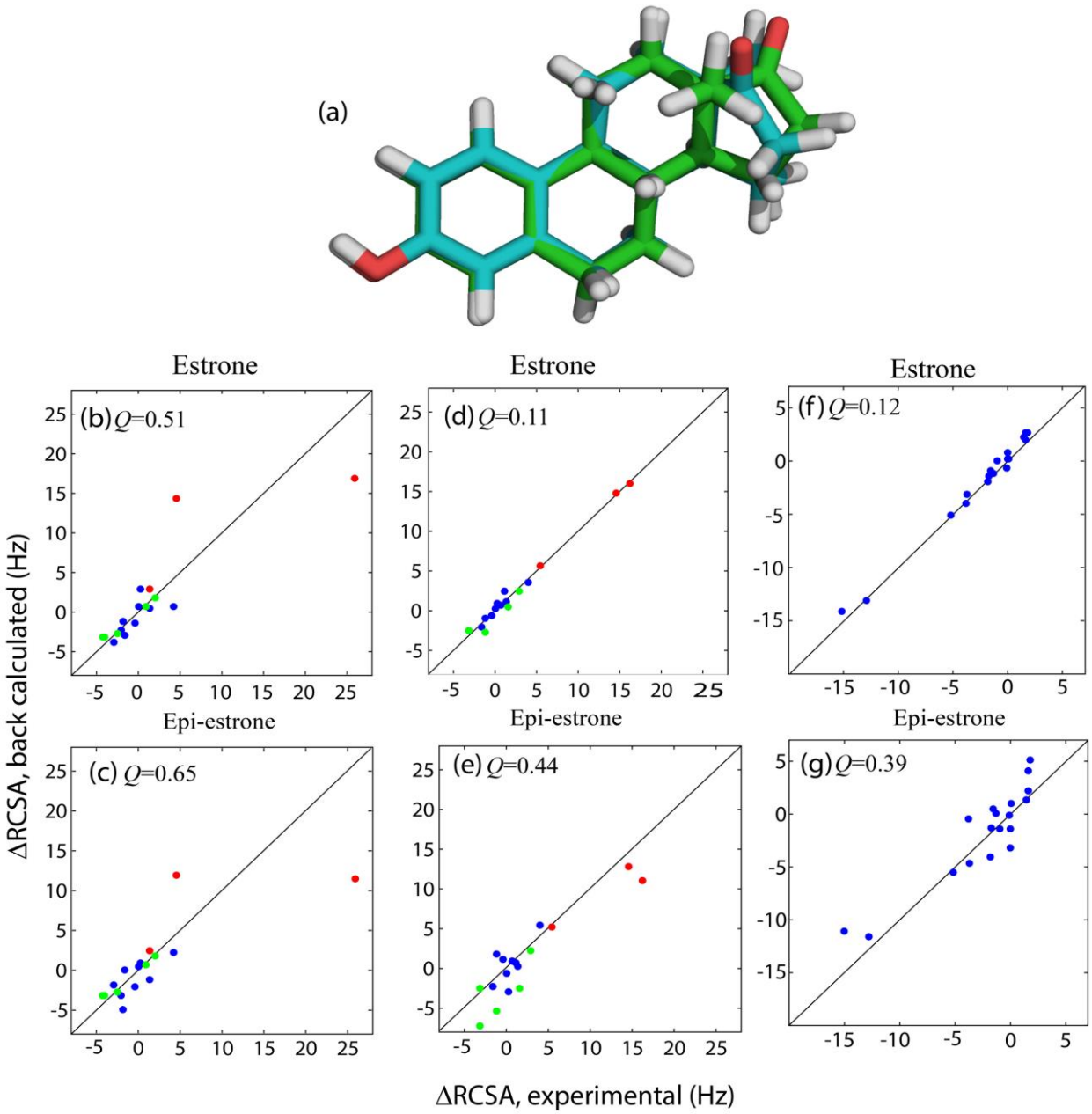

Figure 4: Comparison of experimental and back-calculated ΔRCSAs for estrone and 18-epi-estrone using compression (b-e) and stretching (f-g) devices. Panel (a) shows the 3D structural overlay of estrone and *epi*-estrone. Panels (b) and (c) show ΔRCSAs without isotropic correction while panels (d) and (e) show ΔRCSAs from the compression device after isotropic correction. a) Result for estrone: a correlation was observed between the goodness of fit and chemical shift difference between anisotropic and isotropic signals ($\Delta\delta_{i,iso}^{min}$) under minimum alignment. Specifically, ΔRCSAs matched very well (green dots) for carbons with $\Delta\delta_{i,iso}^{min} = 0$, but poorly for carbons with significant $\Delta\delta_{i,iso}^{min}$ (carbons with $\Delta\delta_{i,iso}^{min} < 50\ Hz$ are represented by blue dots; carbons with $\Delta\delta_{i,iso}^{min} > 50\ Hz$ denoted by red dots). b) Result for *epi*-estrone: the same trend was seen between the goodness of fit and the

size of $\Delta\delta_{i,iso}^{min}$, although the agreement of ΔRCSAs was even worse. Between (a) and (b), the *Q* values do favor the correct configuration (*Q* = 0.51 for estrone vs. *Q* = 0.653 for 18-*epi*-estrone) but the *Q* values are both quite large. c) Correction of the isotropic shift leads to substantially better differentiation with *Q* values of 0.11 for estrone and 0.44 for *epi*-estrone. Panels (f-g) show the agreement between experimental and back-calculated ΔRCSAs for estrone and *epi*-estrone.

In parallel, we also collected eighteen ΔRCSAs for estrone using the stretching device and obtained *Q* (Q_{CSA}) factors of 0.12 (0.347) and 0.39 (1.135) for estrone and 18-*epi*-estrone, respectively, which clearly allows the correct configuration assignment. Finally, it is noteworthy that estrone aligns in a rather peculiar way in both compressing and stretching conditions perhaps due to its low solubility in chloroform (see supporting information).

(c) Mefloquine:

As a third example, we chose mefloquine to explore the feasibility of using ΔRCSAs to differentiate two diastereomers, *erythro*-mefloquine and *threo*-mefloquine. For the compression device, the uncorrected RCSAs lead to poor *Q*-factors with significant fitting outliers as seen in the estrone example (see the supporting information) and a large inter-tensor angle of 58.6° between the ΔRCSA and ΔRDC derived tensors. Satisfactory agreement between the experimental and back-calculated ΔRCSAs was only obtained after isotropic shift correction. As in the estrone case, improved agreement was indicated by a significant decrease in *Q* value (Figures 5(a) and (b)) as well as a small inter-tensor angle of 6.8° between ΔRCSA- and ΔRDC-derived alignment tensors. When using the standard *Q*-factor, non-discriminating values of 0.049 and 0.053 for the *erythro* and *threo* configurations were obtained whereas good discrimination was observed when, Q_{CSA} factors were employed with values of 0.181 for the *erythro* and 0.417 for the *threo*-configurations, respectively. The significantly improved discrimination by Q_{CSA} was expected from the distribution of sp^3 and sp^2 carbons in the molecule.

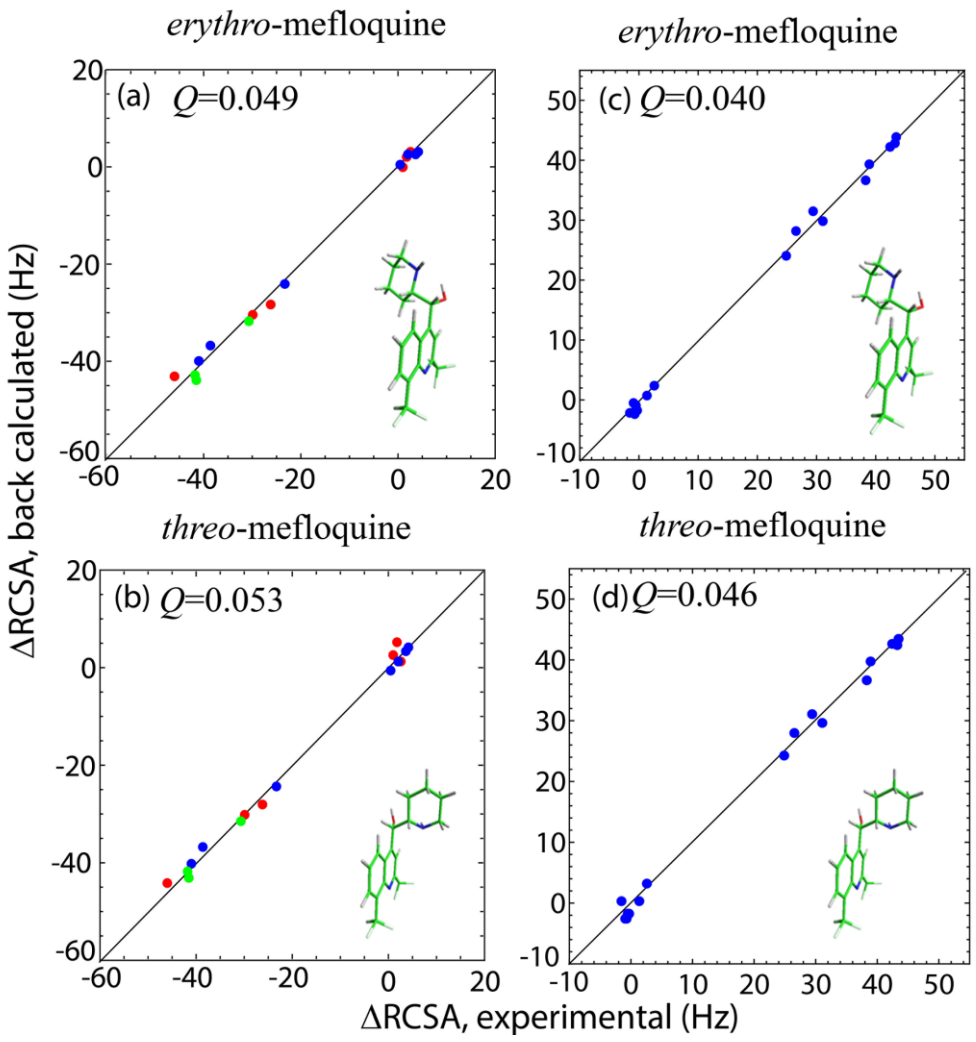

Figure 5: Correlation of experimental and back calculated ΔRCSAs for mefloquine. The data collected with the compression device are presented in panels (a-d) . The data point color-encoding in (a-b) is based on the same criteria as in Figure 4. Low *Q* values indicate satisfactory matching of the ΔRCSAs after isotropic correction. Data acquired using the stretching device is shown in panels (e-h): Note that the correct configuration (*erythro*) produced slightly lower *Q* factor than the incorrect one (*threo*).

From the stretching device, seventeen ΔRCSAs were measured, and *Q*-factors of 0.040 and 0.046 were obtained for the correct *erythro*-mefloquine and the incorrect *threo* form, respectively, while Q_{CSA} factors of 0.290and 0.650 are clearly more discriminating.

(d) Retrorsine:

As a fourth example, we chose the alkaloid retrorsine, a compound examined in a number of earlier NMR studies.[48] It is noteworthy that a previous attempt of RDC-based configuration

assignment was unsuccessful.[48] Using the stretching device, we collected sixteen ΔRCSAs to test whether ΔRCSAs could assign the correct configuration in contrast to the RDC results. Retrorsine is challenging, since some of the configurations occur as several conformers of very similar energy, as determined from molecular modeling calculations. Conformers were generated using the molecular mechanics force-field (MMFF94) in Macromodel with an energy threshold of 9 kJ/mol. Therefore we used these conformations and optimized their populations such that the ensemble of conformations optimally reproduced the ΔRCSA values in a single–tensor fit. In single alignment fit, the alignment tensor is treated as identical for all conformers of the ensemble whose carbon atoms are superimposed. This procedure is implemented in the MSpin program.[47] The configurations were labelled via the *R* or *S* configuration of carbons C2, C3, C10, and C11 respectively, by which convention *RRRS* represents the correct retrorsine configuration. As mentioned, ΔRDC data yielded a slightly higher *Q* factor (0.074) for the correct *RRRS* configuration than for the incorrect *RRRR* configuration (0.066). The difficulty with one-bond RDCs arises since the key carbon, C-11, at which isomer *RRRS* and *RRRR* are epimers, is quaternary. Clearly, RCSA data provide information on quaternary carbons, and indeed the results from ΔRCSA-based ensemble fitting allowed differentiation of all eight possible relative configurations. *Q* (Q_{CSA}) factors of 0.184 (0.370) were obtained for the correct *RRRS* configuration whereas the *RRRR* epimer, had *Q* (Q_{CSA}) factors of 0.215 (0.429). The other six configurations have even higher *Q* (Q_{CSA}) factors and can be more easily ruled out. When fitting the correct *RRRS* configuration, it is worth noting that, despite starting from an ensemble of four conformations, the optimal fit found a 100% population of a single conformer. This conformation closely resembles the X-ray determined structure.[49] The result for retrorsine is a remarkable demonstration of the power of RCSA in proton-deficient molecules, as the same stereochemistry problem could not be solved with single bond RDCs. It should also be mentioned here that the flexibility of retrorsine added significantly to the challenge of assigning the correct diastereomer but was effectively overcome with a combination of modern DFT computational methods and ΔRCSA measurements.

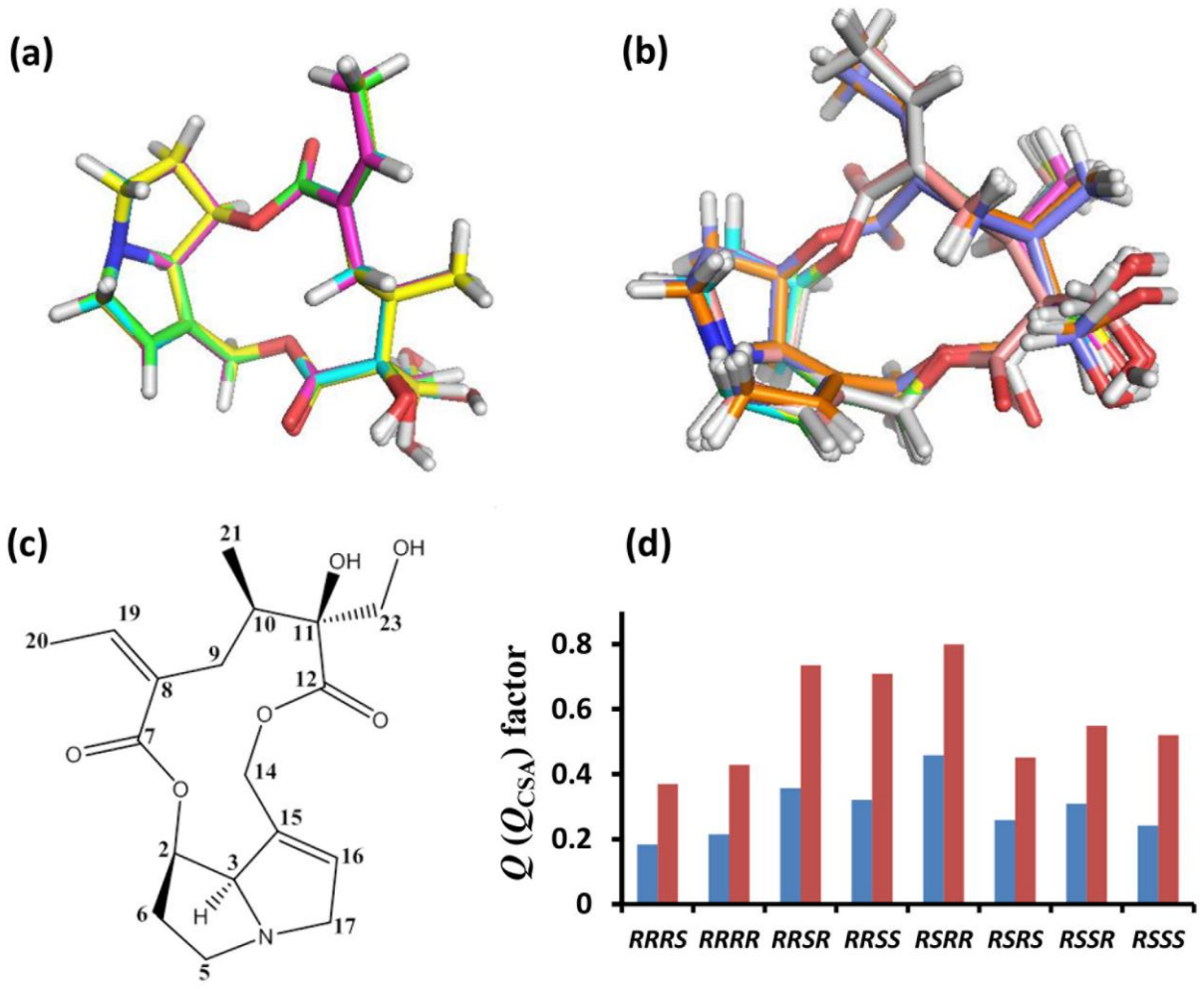

Figure 6: Analysis of retrorsine by an ensemble fit to a unique alignment tensor. Overlay of the conformers for the rigid *RRRS* (a) and flexible *RRRR* (b) configurations. Panel (c) shows the molecular structure of retrorsine with carbon numbering. Panel (d) shows the Q (blue) (Q_{CSA}: red) factors for retrorsine and seven rigid and flexible diastereomers. Inversion of the configuration at two chiral centers in the rings maintains structural rigidity while inversion of the other two chiral centers leads to structural flexibility. Out of eight configurations for retrorsine, *RRRS and RSRS* are structurally rigid diastereomers. Bootstrapping analysis of first two close configurations (*RRRS* and *RRRR*) is provided in the supplemental information.

(e) Menthol:

As a final example, we performed the configuration analysis for menthol. Menthol represents another challenging case: firstly it contains only sp^3 carbons and therefore has a rather small RCSA data span that makes it more susceptible to measurement inaccuracy; secondly menthol only has nine carbons for RCSA readout from which we must determine both alignment tensor

parameters and stereochemistry. Despite these difficulties, we were again able to unambiguously assign the correct relative configuration using ΔRCSAs from the nine carbons. The results are detailed in the supporting information.

Although for all molecules studied here, one alignment condition was sufficient to assign the configuration, it is well known that complicated cases may profit from the availability of more than one alignment conditions. It is worth noting that stretching and compression methods yield highly correlated data (see Figure S9) as theoretically expected, and therefore do not provide independent alignments.

Conclusions

Two complementary methods are introduced that provide a robust means of measuring ΔRCSAs through stretched and compressed PMMA gels. Accurate ΔRCSAs complement conventional *J*-couplings, NOEs, and one-bond or long-range RDC data for analysis of relative configurations. ΔRCSA measurements have the potential to be more appealing to organic chemists since they are simply measured by standard 1D ^{13}C-NMR spectrum acquired with proton decoupling. The experimental simplicity of ΔRCSA measurement is also accompanied by higher sensitivity when compared to the measurement of heteronuclear couplings, particularly when the molecular alignment becomes strong causing broadened resonances and a broad dynamic range of couplings that make magnetization transfer inefficient during coupling constant measurements. For example, signal-to-noise ratios ranging from 56:1 to 176:1 were obtained for 520 µg of strychnine aligned in PMMA gel using the compression device at 900 MHz in 18 hours in a 5 mm cryo-probe. Extrapolating from this result, it is expected that 125 µg for the anisotropic condition would suffice in a 5 mm cryo-probe. This amount could be further reduced to 25 µg when using a 1.7 mm NMR cryo-probe. Alignment with gels was shown to be possible without problems of gel homogeneity in 1.7 mm tubes[50] and the 1.7 mm probe has a 5 times larger signal-to-noise ratio for equal sample mass than the 5 mm probe.[50-52]

Through several examples, we have demonstrated that RCSAs can provide similar (strychnine and mefloquine) and sometimes even better (retrorsine) stereochemical differentiation in comparison to RDCs especially when the chiral center in question is a quaternary carbon. The new analysis parameter introduced here, namely the Q_{CSA} factor, provides enhanced

stereochemical discrimination in cases where conventional *Q*-factor calculations give ambiguous conclusions. We expect that the combination of experimental and analysis method described in this report will offer new avenues for addressing challenging small molecule structural and stereochemical characterization problems in general, and with particular emphasis on proton-deficient structures of natural products of limited abundance and pharmaceutically relevant molecules.

Supporting Information: Experimental details, alignment tensors, RCSA and RDC data for all investigated molecules, a description about gel preparation, derivation of isotropic shift correction formula, comparison of RCSA data with stretching and compression devices, structural co-ordinates and chemical shift tensors.

Acknowledgements: C.G. thanks the Max Planck Society and the DFG (Forschergruppe FOR 934). RRG gratefully acknowledges support from the NSF (CHE-1111684). ANV thanks KIT for a Gastwissenschaftler-Stipendium and UFPE for a visiting professorship as well as the HGF programme BIFTM, the DFG (instrumentation facility Pro2NMR, LU 835/11) and the FACEPE (APQ-0507-1.06/15) for financial support.

Authors information

† Have contributed equally

*To whom the correspondence should be addressed.

Armando Navarro-Vázquez

Tel: +55812126. 8440, Fax: +55812126. 8442; email: armando.navarro@ufpe.br

Yizhou Liu

Tel: +1 732 594 2881, Fax: +1 732 594 9456; email: yizhou.liu@merck.com

Christian Griesinger

Tel: +49 551 201 2201, Fax: +49 551 201 2202; email: cigr@nmr.mpibpc.mpg.de

References:

(1) Nicolaou, K. C.; Snyder, S. A. *Angew. Chem., Int. Ed.* **2005**, *44*, 1012.
(2) Suyama, T. L.; Gerwick, W. H.; McPhail, K. L. *Bioorg. Med. Chem.* **2011**, *19*, 6675.
(3) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1697.
(4) Hallwass, F.; Schmidt, M.; Sun, H.; Mazur, A.; Kummerlowe, G.; Luy, B.; Navarro-Vázquez, A.; Griesinger, C.; Reinscheid, U. M. *Angew. Chem., Int. Ed.* **2011**, *50*, 9487.
(5) Kummerlowe, G.; Grage, S. L.; Thiele, C. M.; Kuprov, I.; Ulrich, A. S.; Luy, B. *J. Magn. Reson.* **2011**, *209*, 19.
(6) Liu, Y. Z.; Prestegard, J. H. *J. Biomol. NMR* **2010**, *47*, 249.
(7) Saupe, A.; Englert, G. *Phys. Rev. Lett.* **1963**, *11*, 462.
(8) Rowell, J. C.; Phillips, W. D.; Melby, L. R.; Panar, M. *J. Chem. Phys.* **1965**, *43*, 3442.
(9) Panar, M.; Phillips, W. D. *J. Am. Chem. Soc.* **1968**, *90*, 3880.
(10) Klages, J.; Neubauer, C.; Coles, M.; Kessler, H.; Luy, B. *Chembiochem* **2005**, *6*, 1672.
(11) Schuetz, A.; Junker, J.; Leonov, A.; Lange, O. F.; Molinski, T. F.; Griesinger, C. *J. Am. Chem. Soc.* **2007**, *129*, 15114.
(12) Thiele, C. M.; Maliniak, A.; Stevensson, B. *J. Am. Chem. Soc.* **2009**, *131*, 12878.
(13) Kummerlowe, G.; Crone, B.; Kretschmer, M.; Kirsch, S. F.; Luy, B. *Angew. Chem., Int. Ed.* **2011**, *50*, 2643.
(14) Trigo-Mouriño, P.; Navarro-Vázquez, A.; Ying, J. F.; Gil, R. R.; Bax, A. *Angew. Chem., Int. Ed.* **2011**, *50*, 7576.
(15) Yan, J. L.; Kline, A. D.; Mo, H. P.; Shapiro, M. J.; Zartler, E. R. *J. Org. Chem.* **2003**, *68*, 1786.
(16) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371.
(17) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127.
(18) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020.
(19) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098.
(20) Czarniecka, K.; Samulski, E. T. *Mol. Cryst. Liq. Cryst.* **1981**, *63*, 205.
(21) Meyer, N. C.; Krupp, A.; Schmidts, V.; Thiele, C. M.; Reggelin, M. *Angew. Chem., Int. Ed.* **2012**, *51*, 8334.
(22) Lei, X. X.; Xu, Z.; Sun, H.; Wang, S.; Griesinger, C.; Peng, L.; Gao, C.; Tan, R. X. *J. Am. Chem. Soc.* **2014**, *136*, 11280.
(23) Sau, S. P.; Ramanathan, K. V. *J. Phys. Chem. B* **2009**, *113*, 1530.
(24) Zong, W.; Li, G. W.; Cao, J. M.; Lei, X. X.; Hu, M. L.; Sun, H.; Griesinger, C.; Tan, R. X. *Angew. Chem., Int. Ed.* **2016**, *55*, 3690.
(25) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340.
(26) Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303.
(27) Ishii, Y.; Markus, M. A.; Tycko, R. *J. Biomol. NMR* **2001**, *21*, 141.
(28) Gayathri, C.; Tsarevsky, N. V.; Gil, R. R. *Chem.-Eur. J.* **2010**, *16*, 3622.
(29) Freudenberger, J. C.; Spiteller, P.; Bauer, R.; Kessler, H.; Luy, B. *J. Am. Chem. Soc.* **2004**, *126*, 14690.
(30) Kummerlowe, G.; Auernheimer, J.; Lendlein, A.; Luy, B. *J. Am. Chem. Soc.***2007**, *129*, 6080.
(31) Haberz, P.; Farjon, J.; Griesinger, C. *Angew. Chem., Int. Ed.* **2005**, *44*, 427.
(32) Nath, N.; d'Auvergne, E. J.; Griesinger, C. *Angew. Chem., Int. Ed.* **2015**, *54*, 12706.
(33) Wolkenstein, K.; Sun, H.; Falk, H.; Griesinger, C. *J. Am. Chem. Soc.* **2015**, *137*, 13460.

(34) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839.
(35) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836.
(36) Choy, W. Y.; Tollinger, M.; Mueller, G. A.; Kay, L. E. *J. Biomol. NMR* **2001**, *21*, 31.
(37) Tolbert, B. S.; Miyazaki, Y.; Barton, S.; Kinde, B.; Starck, P.; Singh, R.; Bax, A.; Case, D. A.; Summers, M. F. *J. Biomol. NMR* **2010**, *47*, 205.
(38) Grishaev, A.; Ying, J. F.; Bax, A. *J. Am. Chem. Soc.* **2006**, *128*, 10010.
(39) Kuchel, P. W.; Chapman, B. E.; Muller, N.; Bubb, W. A.; Philp, D. J.; Torres, A. M. *J. Magn. Reson.* **2006**, *180*, 256.
(40) Kummerlowe, G.; McCord, E. F.; Cheatham, S. F.; Niss, S.; Schnell, R. W.; Luy, B. *Chem. Eur. J.* **2010**, *16*, 7087.
(41) Courtieu, J.; Bayle, J. P.; Fung, B. M. *Prog. Nucl. Mag. Res. Spectrosc.* **1994**, *26*, 141.
(42) Grishaev, A.; Yao, L. S.; Ying, J. F.; Pardi, A.; Bax, A. *J. Am. Chem. Soc.* **2009**, *131*, 9490.
(43) Bifulco, G.; Riccio, R.; Martin, G. E.; Buevich, A. V.; Williamson, R. T. *Org. Lett.* **2013**, *15*, 654.
(44) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **2011 1999**, *138*, 334.
(45) Kramer, F.; Deshmukh, M. V.; Kessler, H.; Glaser, S. J. *Concepts Magn. Reson. Part A* **2004**, *21A*, 10.
(46) Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J. G.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047.
(47) Navarro-Vázquez, A. *Magn. Reson. Chem.* **2012**, *50*, S73.
(48) García, M. E., Woodruff, S. R., Hellemann, E., Tsarevsky, N. V., Gil, R. R. *Magn. Reson. Chem.* **2016** (doi: 10.1002/mrc.4400).
(49) Coleman, P. C.; Coucourakis, E. D.; Pretorius, J. A. *S. Afr. J. Chem.* **1980**, *33*, 116.
(50) Ge, H. M.; Sun, H.; Jiang, N.; Qin, Y. H.; Dou, H.; Yan, T.; Hou, Y. Y.; Griesinger, C.; Tan, R. X. *Chem. Eur. J.* **2012**, *18*, 5213.
(51) Hilton, B. D.; Martin, G. E. *J. Nat. Prod.* **2010**, *73*, 1465.
(52) Dalisay, D. S.; Molinski, T. F. *J. Nat. Prod.* **2010**, *73*, 679.

Graphical abstract