Published in final edited form as:

Science. 2017 April 07; 356(6333): . doi:10.1126/science.aam5349.

Unequivocal Orthogonal Definition of Complex Molecular Structures Using Anisotropic NMR Measurements

Yizhou Liu¹, Josep Sauri², Emily Mevers³, Mark W. Pecuh⁴, Henk Hiemstra⁵, Jon Clardy³, Gary E. Martin^{6,*}, and R. Thomas Williamson^{6,*}

¹Merck & Co., Inc., Process & Analytical Research and Development, Structure Elucidation, 2000 Galloping Hill Road, Kenilworth, NJ 07033, USA ²Merck & Co., Inc., Process & Analytical Research and Development, Structure Elucidation, 33 Avenue Louis Pasteur, Boston, MA 02115, USA ³Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115 USA ⁴Department of Chemistry, University of Connecticut, 55 N. Eagleville Road, U3060, Storrs, CT 06269 USA ⁵Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands ⁶Merck & Co., Inc., Process & Analytical Research and Development, Structure Elucidation, 126 E. Lincoln Ave., Rahway, NJ 07065

Abstract

Assignment of complex molecular structures from nuclear magnetic resonance (NMR) data can be prone to interpretational mistakes. Residual dipolar couplings (RDCs) and residual chemical shift anisotropy (RCSA) provide a spatial view of the relative orientations between bonds and chemical shift shielding tensors respectively, regardless of separation. Consequently, these data constitute an excellent reporter of global structural validity. Anisotropic NMR parameters can be used to evaluate investigators' structure proposals or structures generated by Computer-Aided Structure Elucidation (CASE). Application of the method to several complex structure assignment problems shows promising results that signal a potential paradigm shift from conventional NMR data interpretation, of particular utility for compounds not amenable to x-ray crystallography.

ONE PAGE SUMMARY

INTRODUCTION: Single crystal X-ray diffraction (SCXRD) studies represent the gold standard for unequivocal establishment of molecular structure and configuration. For molecules that will not crystallize or that form poorly diffracting crystals, nuclear magnetic resonance (NMR) spectroscopy methods provide the primary viable alternative means to determine molecular structures. However, misinterpretation of NMR data either as a result of poor data quality, inappropriate experiment selection, or investigator bias has led to burgeoning numbers of structure revision reports. Clearly, the development of a method to more effectively use NMR data and simultaneously quell reports of incorrect structures would be highly beneficial.

*Correspondence to: gary.martin2@merck.com and robert.williamson2@merck.com.

SUPPLEMENTARY MATERIAL

Figures S1 to S10

Tables S1 to S9

RATIONALE: Combining computer-assisted structure elucidation (CASE) algorithms and density functional theory (DFT) calculations with measured anisotropic NMR parameters, specifically residual dipolar coupling (RDC) and residual chemical shift anisotropy (RCSA), holds strong promise as an effective alternative means of assigning three-dimensional molecular structures. Anisotropic NMR data provide a spatial view of the relative orientations between bonds (RDCs) and chemical shift shielding tensors (RCSAs) regardless of the separation between the bonds and atoms. Hence, these data are sensitive reporters of global structural validity.

RESULTS: Anisotropic NMR data can be used directly to evaluate the validity of proposed structures or they can be combined with structures generated by a CASE program in conjunction with DFT calculation of the expected RDC and RCSA data for a proposed structure or structures. The RDC data are used for C-H bond vectors while the RCSA data report on the chemical shift tensor of both protonated and non-protonated carbons, the latter only accessible by long-range RDC data that are difficult to measure and interpret. These data are evaluated for a given structure based on the agreement between the experimentally measured data and theoretical values calculated for the proposed structure using DFT methods. When structures generated by a CASE program are being considered, the method only requires a multi-dimensional NMR data set of sufficient quality and sophistication to allow the CASE program to generate a set of valid structural proposals; the molecules being studied should also be amenable to modern DFT calculations for three-dimensional model building. Results obtained using the proposed method demonstrate its applicability to a diverse range of complex molecules, each of which challenged the investigators originally reporting the structures.

CONCLUSIONS: The technique described here represents a potential paradigm shift from conventional NMR data interpretation and can provide an unequivocal and unbiased confirmation of interatomic connectivity and relative configuration for organic and natural product structures.

Graphical Abstract

The principle of RDC-based model differentiation is shown using aquatolide as an example. The four methine groups in aquatolide are shown in different colors. In the "revised" and the "original" structures, these C-H vectors clearly have different relative orientations, as is evident after translating them to the same origin. Theoretical RDC values associated with these or other vectors

can be calculated for each model based on the alignment tensor that can be determined by singular value decomposition (SVD). A correlation plot is used to assess the agreement between calculated and experimental RDC values. Here, correlation data are shown only for the four CH groups, although the alignment tensor was actually determined using all available data. RCSA data can be utilized in a similar fashion though chemical shift shielding tensors are more difficult to illustrate graphically.

One Sentence Summary:

Anisotropic NMR data can be employed to provide an orthogonal and unequivocal confirmation of complex molecular structures and to avoid incorrect structural assignments.

Single-crystal X-ray diffraction (SCXRD) is unquestionably the gold standard in structure elucidation. However, a crystal that diffracts well is not always easily obtained, and many compounds, especially natural products, are isolated and purified in quantities too small for conventional crystallization screening. For some molecules that will not crystallize directly, crystalline sponges offers an exciting possibility, although currently, the sample soaking step requires careful optimization and serendipity on a per molecule basis (1). On a different front, atomic force microscopy (AFM) with atomic resolution (2) has reached the breakthrough point of being able to visualize individual bonds but the current state of the art requires that the molecule under study adopt a planar structure for quality resolution.(3,4)

In the realm of spectroscopic methods, nuclear magnetic resonance (NMR) is the primary technique for full structure elucidation. However, conventional NMR has inherent limitations subject to the interpretational biases of the investigator. A recent SciFinder™ search under the keywords “structure revision” revealed >1200 reports, including 39 in 2016, with five of these examples appearing in a single week in *Organic Letters* and the *Journal of Organic Chemistry*.(5–9) Clearly, a more objective and robust protocol for NMR structural determination not hampered by interpretational difficulties is highly desirable.

Traditional NMR structural elucidation is essentially a reverse engineering process, i.e., one that deduces the actual structure starting from various pieces of experimental data in a puzzle-solving fashion. For example, proton and carbon chemical shifts are first measured, providing information on the numbers of atoms and types of functional groups. Next, experiments based on scalar (*J*) couplings, such as homonuclear correlation spectroscopy (COSY) and heteronuclear multiple bond correlation (HMBC), are used to establish connectivities between these groups. Distance constraints afforded by nuclear Overhauser effect spectroscopy (NOESY) and rotating frame Overhauser effect spectroscopy (ROESY), can also be employed to further facilitate the assembly of connectivity networks and to define a three-dimensional configuration. Unfortunately, prejudices of the investigator can intrude at any point in this process.

Here, we describe an orthogonal check on the validity of molecular structure assignments based on anisotropic NMR parameters, namely residual dipolar couplings (RDC) (10–12) and residual chemical shift anisotropy (RCSA) (13–16). These data provide three-dimensional information on relative orientations of different bonds and chemical shielding

tensors in the molecule. Because RDC and RSCA data are insensitive to the distances between the bonds and atoms, they reveal angular relationships of various structural elements from all positions of a molecule, and therefore reflect the correct overall structure without being subject to investigator bias.

Fundamental principles of combining CASE and DFT calculations with RDC and RSCA measurements

Over the past decade, advances in computer algorithms (17–19) and quantum-chemistry computational methods (20–23) have underpinned a trend in structural elucidation that could be termed forward engineering. Rather than striving to assemble the correct structure in a single stroke using all available information, all possible structures consistent with the ensemble of available experimental data are first assembled by a computer-assisted structure elucidation (CASE) algorithm using homo- and heteronuclear shift correlation data.(17–19) Next, theoretical values associated with critical measurements (i.e. ^1H and ^{13}C NMR chemical shifts) are calculated from each of the proposed structural candidates through density functional theory (DFT). Structures generated by the program are finally sorted on the basis of the congruence of the experimental and calculated ^{13}C chemical shift data, leading to the selection of the best candidate or candidates.

Clearly, the CASE software undertakes some of the reverse engineering task in its black-box suite of algorithms, but in the approach presented here, we are not so concerned with obtaining the exact structure, but more with covering sufficient chemical space so that the correct structure is contained in the ensemble of structural proposals generated by the program. Inclusion of additional long-range ^1H - ^{13}C correlations absent in conventional HMBC experiments but available from newer NMR experiments such as LR-HSQMBC(24), and 1/n-bond ^{13}C - ^{13}C correlations available from 1,1/n-HD-ADEQUATE experiments(25–27), can greatly facilitate candidate generation or narrow down the range of candidates in the CASE output.(28,29)

As an alternative to using a CASE algorithm, candidate structures deduced by an investigator can also be interrogated using anisotropic NMR data. Examples of both CASE-generated structures and those deduced by an investigator will be discussed.

Typically, the top candidates from the CASE program or those deduced by the investigator are subjected to DFT geometry optimization in order to obtain three-dimensional structures. These calculations can be a rate-limiting step if too many candidates are to be evaluated; however on modern computational clusters, they can often be completed in a matter of hours, and optimization with an extensive conformer search can be completed for a dozen candidates within 2 to 3 days. With the three-dimensional models in hand, experimentally observable parameters such as chemical shifts, scalar (J) couplings, NOE/ROE patterns, and VCD/IR spectra can be calculated. Although these data can potentially serve as critical measurements, all have well-known limitations. For example, structural information from chemical shifts, scalar (J) couplings, and NOE/ROE correlations are inherently local, as these phenomena arise mostly through short-range interactions; the readout from VCD is a

cumulative effect from all stereogenic centers in a molecule so that diastereomeric differentiation can be very challenging for compounds possessing multiple chiral centers.

For reasons above, we propose here an alternative/orthogonal method that provides a check on the global structure in an objective manner. Anisotropic NMR parameters, RDCs(10–12) and RCSAs(13–16), used here for critical structure/configuration assessment arise from well-known NMR phenomena. RDCs and RCSAs are observed only when a molecule is partially aligned, usually with the aid of an alignment medium such as a constrained polymeric gel. A schematic of how a molecule might orient with a polymeric gel stretched along the B_0 axis of the magnetic field is shown in Figure 1. Experimentally, RDC values are extracted from the difference of two measurements that yield the heteronuclear $^1J_{CH}$ coupling constant in isotropic solution and the total coupling constant $^1J_{CH} + ^1D_{CH}$ in an anisotropic medium, respectively. The $^1D_{CH}$ component of the total coupling constant measured in an anisotropic medium is the residual dipolar coupling or RDC, which can be either positive or negative, *i.e.* the total coupling constant measured in anisotropic media can be either larger or smaller than the value of $^1J_{CH}$. The commonly measured one-bond ^{13}C - 1H RDC data, for example, report on the relative orientations of different C-H bonds in the molecule. This information forms the basis of RDC-based structural differentiation. The corresponding bond/bond orientation relationships are invariably different between the correct and incorrect structures. Whereas the theoretical RDC values based on a correct structure must agree with experimental measurements, the same is not true for an incorrect structure. Such agreement can be qualitatively assessed by a theoretical vs. experiment correlation plot, or numerically measured by the quality (Q) factor(30). A low Q -factor indicates good agreement between theory and experiment.

A complementary anisotropic NMR parameter, residual chemical shift anisotropy or RCSA, has the advantage of providing structural information for carbons that are not bonded to hydrogen. Structural information for these carbons was previously only obtainable with challenging-to-acquire long-range RDC measurements.(31–33). Due to difficulties in the elimination of isotropic contributions to the RCSA measurement and other issues, the RCSA measurements were only recently applied to small molecules.(13–16) In the most current report, two alternative methods for successfully measuring RCSAs in stretched and compressed gels have been reported along with the means of eliminating any isotropic chemical shift change that otherwise contaminates RCSA measurement.(16) As the measurement of RCSA is operationally somewhat more demanding than RDCs, any satisfactory RCSA-enabling techniques usually facilitate RDC measurements as well, although the reverse is not always true. As RCSA and RDC can provide complementary structural information, the measurement of both is advisable when RCSA data are being acquired. Unlike RDCs, which inform on relative bond orientations, RCSAs report on the relative orientations of different chemical shielding tensors in the molecular structure. In terms of their application in structure elucidation, RCSAs share a similar utility as RDCs.

The principle for structural differentiation based on RCSAs illustrated in Figure 2, and is comparable to RDC-based analysis, except that the bond vectors are now replaced by chemical shielding tensors that can be accurately calculated by gauge invariant atomic orbital (GIAO) DFT.(20–23) Simultaneously using both RDC and RCSA data also leads to

more reliable determination of the alignment tensor, consequently enhancing the robustness of overall structure differentiation.

Recently, we reported the application of a combination of RDC and RCSA measurements in establishing the stereochemistry of the novel natural product homodimericin A.(29) During the course of that work, it became obvious that the combination of RDCs and RCSAs provides a powerful, and orthogonal means of confirming not only relative configuration of a given stereocenter, but also the overall molecular structure and atomic connectivity of the molecule under study. Combined analysis of RDC and RCSA data provides an independent means of confirming or refuting a proposed structure or choosing among alternative structures generated by a CASE program. The confluence of capabilities embodied by computer-assisted structure elucidation (CASE) methods, DFT calculations, and now the relatively facile measurement of anisotropic NMR parameters has facilitated the development of a generally applicable method for the definition of molecular structure and configuration that should help to address the growing and general problem of structural mischaracterization.

Application to cryptospirolepine

The structure proposed for cryptospirolepine, **1**, first reported in 1993, was based solely on COSY, ^1H - ^{13}C HMQC and optimized 8 Hz ^1H - ^{13}C HMBC data.(34) In 2002, the original NMR sample was examined chromatographically and found to contain 26 components with none of **1** remaining. The two major components, **2** and **3**, were isolated and fully characterized.(35) Identification of the major degradants strongly suggested that the original structure report was incorrect (Fig. 3). Finally, in 2015, utilizing 1.7 mm MicroCryoProbe™ technology and the newly developed 1,1-HD-ADEQUATE experiment(27) it was possible to revise the structure to **4**.

Although cryptospirolepine, **4**, is racemic, the molecule provides a useful sample for assessing the structural validation capability of RDC/RCSA data, which are measured in achiral alignment media that are insensitive to absolute configuration. The comparison is presented in Figure 4. Even a cursory inspection reveals that the correlation between theoretical and experimental values for the revised structure, **4** (panel B) is considerably stronger than for the original structure, **1** (panel A). This difference is also reflected in a much lower Q -value of 0.122 for **4** in comparison with the value of 0.245 for **1**. There are three strongly coupled pairs of resonances in the 500 MHz ^1H NMR spectrum of **4** that can potentially undermine RDC measurement accuracy.(36) After removing the corresponding RDC data, the Q -value for **4** decreased to 0.082 while that for **1** was 0.217, still providing ample basis for differentiating the correct structure, **4**, from the erroneous original structure, **1** (see supplementary material).

In this particularly case (Fig. 5), the bottom half of the structure (blue) is actually identical in structures **1** and **4** (correlation Q -factors of 0.16 and 0.13, respectively), but the top half (red) is dramatically different (correlation Q -factors of 0.29 and 0.12 for **1** and **4**, respectively). The fact that the poor agreement in **1** is localized to the top half further confirms the location of the structural assignment mistake.

Application to spiroketal rearrangement products

We recently collaborated on a study that involved the rearrangement of a spiroketal triggered by an enol-ether epoxidation.(37) Because of the indeterminate nature of the number of bonds spanned by HMBC correlations, it was not possible to assemble the structure from the normal ensemble of NMR data (COSY, HSQC, HMBC). Resorting to computer-assisted structure elucidation reduced the number of structures consistent with the data, to the two best choices (based on cumulative error between the experimental vs. calculated ^{13}C shift data) represented by **5** and **6**. The correct structure was readily identified as **5** following the acquisition of a 40 Hz optimized 1,1-HD-ADEQUATE spectrum.(27) In the event that an investigator either didn't have a working knowledge of the ADEQUATE experiments or access to a spectrometer equipped with a cryoprobe with which to acquire those data, the differentiation of **5** from **6** afforded another opportunity for structure differentiation using RDC/RCSA data.

DFT calculations performed on both **5** and **6** suggested multiple conformers of comparable thermal energies, but the RDC/RCSA data instead suggested a single major conformer (see supplementary material). Plotting the experimental vs. the back-calculated RDC/RCSA data as above gave the results shown in Figure 6. Although it was readily possible to identify the correct structure between the two choices using the 1,1-HD-ADEQUATE data, clearly the RDC/RCSA data independently identify the correct choice as **5** as well, based on the significant difference in the $Q = 0.20$ for **5** vs. $Q = 0.55$ for **6**.

Application to aquatolide

For a final example, we chose the natural product aquatolide.(38–42) The originally proposed structure, **7**, incorporated a very rare ladderane moiety in the molecular framework.(38) Reisolation followed by extensive NMR, quantum mechanical calculations, and X-ray crystallography led to the revision of the structure, **8**,(39) which was followed in 2015 by the total synthesis reported by Hiemstra and co-workers.(41) A more recent study by Buevich and Elyashberg examined the use of DFT calculation of chemical shifts as a means of choosing between alternative structures suggested by a CASE program.(42) In addition to the revised structure **8**, the CASE program output contained two additional structures, **9** and **10**, that were considered to be potentially viable alternatives based on the cumulative error between the experimentally observed and calculated ^{13}C chemical shifts for the structures.

As is readily apparent from the RDC/RCSA data for aquatolide in DMSO- d_6 plotted in Figure 7, the best fit is obtained for the revised structure(39,41) for which the data are shown in panel A, $Q = 0.12$. In stark contrast, the fit for the originally proposed structure with the ladderane moiety is extremely poor, with $Q = 0.72$. The alternative structures, **9** and **10**, had successively poorer correlation plots with $Q = 0.23$ and 0.59 , respectively. Hence, as is shown in the two previous examples, a correlation plot between the experimental and back-calculated RDC/RCSA data for the model structures readily establishes which structure is in best agreement with the data.

Outlook

The suite of model compounds in this report, all of which have occasioned either incorrect structure reports or structural assignment difficulties, illustrate the power of combining modern CASE algorithms and DFT calculations with RDC and RCSA data to simultaneously define chemical connectivity and configuration as well as three-dimensional structure. The best candidates in our examples have RDC/RCSA-combined Q -values of 0.1~0.2. A Q -value in this range is typically associated with high-resolution crystallographic analysis.⁽⁴³⁾ A three-dimensional structure of this accuracy provides valuable insights into understanding drug-target interactions and structure-activity relationships (SAR). On the other hand, we should point out that a successful execution of this approach has some prerequisites when a CASE algorithm is employed to generate structures. First, the initial input for the program must contain sufficient information to allow for candidate generation within a reasonable amount of time. In many cases, proton/carbon chemical shifts and HMBC correlations suffice, but in more challenging cases, such as that of proton-deficient compounds like homodimericin-A (29), long-range H-C(24) correlation and C-C correlation data will be required. Fortunately, these data are now available even for sub-milligram samples owing to new NMR pulse sequences and development in spectrometer hardware. (27) Second, the candidates generated by CASE must contain the correct structure. Our experience with the CASE program indicates that this requirement is met in most cases. In fact, a good computer algorithm has been reported to outperform a human expert with respect to deep exploration of all structural possibilities.^(17–19) Third, the proposed chemical structures should be approachable by DFT geometry optimization. In this work, DFT geometry optimization was conducted in *Gaussian* 09 at the B3LYP/6–31G(d,p) level. B3LYP works well with a small basis set such as that used in this work, and employing a large basis set causes virtually no difference in the final structure in our tests. All DFT calculations were performed in vacuum without solvent modeling, although the actual samples were analyzed in DMSO. Our previous experience shows that including solvent effects minimally changes the optimized structure, but does alter its calculated energy, which should be considered when accurate Boltzmann distribution weighting is needed among multiple rotamers of comparable energies. For compounds in this study, only a single lowest-energy rotamer was predicted, except for the spiroketal molecule. In that case, as described in the supplementary material, anisotropic NMR data agreed best with a major rotamer that is actually not the lowest energy rotamer according to DFT in vacuum. This discrepancy may reflect an effect of DMSO that was not taken into account, although including the implicit effect of DMSO by polarizable continuum model (PCM) did not lower the relative energy of this rotamer. It should be noted that while for some compounds even simple molecular mechanics calculations can yield three-dimensional structures of high accuracy, DFT calculation can still be quite challenging for other compounds, e.g. those whose structures are stabilized mostly through intra- or intermolecular hydrogen bonds, such as polypeptides. However, there are other structure prediction tools better tailored to these needs, such as CS-ROSETTA.⁽³⁰⁾ For GIAO chemical shielding calculation, the mpw1pw91 functional and the 6–31G(d,p) basis set were used, which consistently gave slightly better RCSA Q -factors than the B3LYP/6–31G(d,p) combination in all tested cases. It is noteworthy that the RCSA-based analysis is more robust against GIAO-DFT inaccuracy

than the chemical shift based analysis, because for the former only the relative size of the prediction error to the overall shielding actually affects the analysis, while for the latter the absolute prediction error directly impacts the analysis. Once a reasonable three-dimensional model associated with each candidate is generated, whether via computational methods, or investigator deduction, and chemical shielding tensors are calculated by DFT based on this three-dimensional model, RDC and RCSA data can be employed as a sensitive critical measure to evaluate the validity of the structural assignment. The possibility of a false positive determination, i.e. agreement of RDC and RCSA data with an incorrect structure, is significantly lower (44) than that in an analysis using only conventional NMR data, especially when both RDC and RCSA are jointly used. These data can serve as a convenient NMR litmus test of structure and stereochemical validity. As such, we feel that using the method described in this work has considerable potential to be widely applied and that doing so could be help to quell the flow of incorrect structures appearing in the literature.

Materials and methods

Preparation of Poly-(2-hydroxyethyl methacrylate) (poly-HEMA) Gel

The preparation of EGDMA (ethylene glycol dimethylacrylate) cross-linked poly-HEMA gel followed a published protocol (45), but the HEMA monomer concentration during polymerization and the cross-linking ratio were optimized specifically for use with the gel-stretching device to final values of 60% (v/v) and 0.07% (v/v), respectively. Polymerization was carried out in a 1/8" ID FEP tubing (Cole-Parmer) at 50°C with 0.06% (v/v) V70 (Wako Chemicals USA, Inc) as the radical initiator. The polymerized gel was cut into 2 cm segments and washed three times in methanol over a period of two days. The gel sticks were dried on a glass surface prior to use.

NMR Sample Preparation and Experiments

The NMR experiments for RDC and RCSA measurements are relatively straightforward and well-documented (16,29). The experimental time closely depends on the material availability. For example, *J*-resolved HSQC and $^{13}\text{C}\{^1\text{H}\}$ spectra of excellent quality can be obtained in about 15 hours each, for 1 mg of aquatolide in an unstretched HEMA gel on a 500 MHz magnet equipped with a Prodigy probe. Under the gel-stretched state, longer experiments with 50% more transients are advised to account for the reduction in active sample volume in the narrower segment of the tube, as previously described (16,29). The compounds used in our current work, which ranged from 1 to 3 mg in quantity, took 2~3.5 days of analysis time per sample. It is worth noting that the DMSO-compatible gel has an advantage over the chloroform compatible gel for dilute samples in that longer experiments can be run without solvent evaporation.

The samples used for this study were either isolated and purified or synthesized as described in the primary references for the individual compounds. Resonance assignments of cryptospirolepine (34) were adopted from a previous publication (27). Samples of the spiroketal (1 mg) (37), and aquatolide (1 mg) (41) were first dissolved in 150 μL DMSO- d_6 for resonance assignment prior to RDC and RCSA measurements, as the original chemical shift assignments were carried out in different solvents. NMR resonance assignments were

completed using a combination of ^1H , ^{13}C , COSY, ^1H - ^{13}C HSQC, and ^1H - ^{13}C HMBC experiments.

To prepare the gel sample, the test compound was dissolved in 350 μL DMSO- d_6 with the addition of 5 μL tetramethylsilane (TMS) for carbon chemical shift referencing, and a gel stick was added in a horizontal position to swell for a period of 3 days. The fully swollen gel was then transferred to a 4.2/3.2mm gel-stretching device as described previously (16). For RDC measurements, the J -resolved BIRD-HSQC experiment with F2 homonuclear decoupling (HD- J -HSQC) (46) was utilized with a F1-acquisition time ranging from 256–312 ms; an F2-acquisition time of 120 ms; and a recycling delay of 1.5 s. For the spiroketal and aquatolide, some signals overlapped with signals from the gel polymer, so the F2-coupled CLIP-HSQC experiment (47) was employed to obtain coupling data for those overlapped resonances, as well as couplings of individual C-H vectors of methylene groups (Note that the HD- J -HSQC experiment only provides the sum of two C-H couplings for anisochronous protons of a methylene group). An F1-acquisition time of 12 ms, a F2-acquisition time of 500 ms, and a recycling delay of 1.5 s were used in the F2-coupled CLIP-HSQC experiment; the aromatic regions were folded in the F1-dimension to conserve spectrometer time. All NMR experiments were conducted at 25°C on a Bruker 500 MHz spectrometer equipped with a ProdigyTM probe.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was funded, in part, by NIH GM086258. Experimental data are available in supplementary material. The authors thank Dr. Katrina Lexa for providing the coordinates of 100,000 peptide conformers for the false positive rate estimation (44) and Figure S10.

References and Notes

1. Inokuma Y, Yoshioka S, Ariyoshi J, Arai T, Hitora Y, Takena K, Matsunaga S, Rissanen K, Fujita M, X-ray analysis on the nanogram to microgram scale using porous complexes *Nature*, 495, 461–466 (2013); DOI: 10.1038/nature11990. [PubMed: 23538828]
2. Gross L, Mohn F, Moll N, Liljeroth P, Meyer G, The chemical structure of a molecule resolved by atomic force microscopy, *Science*, 325, 1110–1114 (2009); DOI: 10.1126/science.1176210. [PubMed: 19713523]
3. Hansen KØ, Schuler B, Williams AJ, Demissie TB, Hansen E, Andersen JH, Svenson J, Blinov K, Repisky M, Mohn F, Meyer G, Svendsen J-S, Ruud K, Elyashberg M, Gross L, Jaspars M, Isaksson J, A combined atomic force microscopy and computational approach for the structural elucidation of breitfussin A and B: highly modified halogenated dipeptides from *Thuria breitfussi*, *Angew. Chem. Int. Ed.*, 51, 12238–12241 (2012); DOI: 10.1002/anie.201203960.
4. Gross L, Mohn F, Moll N, Meyer G, Ebel R, Abdel-Mageed WM, Jaspars M, Organic structure determination using atomic-resolution scanning probe microscopy *Nature Chemistry*, 2, 821–825 (2010); DOI: 10.1038/nchem.765.
5. Reddy DS, Kutateladze AG, Structure revision of an acorane sesquiterpene cordycepol A, *Org. Lett.*, 18, 4860–4863 (2016); DOI: 10.1021/acs.orglett.6b02341 [PubMed: 27611774]
6. Ribaucourt A, Hodgson DM, Total synthesis and structural revision of the cytotoxin aruncin B, *Org. Lett.*, 18, 4364–4367 (2016); DOI: 10.1021/acs.orglett.6b02120. [PubMed: 27529438]

7. Shin I, Lee D, Kim H, Substrate-controlled asymmetric total synthesis and structure revision of (–)-bisezakynine A, *Org. Lett* 18, 4420–4423 (2016); DOI: 10.1021/acs.orglett.6b02239. [PubMed: 27551943]
8. Inman M, Carvalho C, Lewis W, Moody CJ, Synthesis of the reported pyranonaphthoquinone structure of the indoleamine-2,3-dioxygenase inhibitor annulin B by Regioselective Diels–Alder reaction, *J. Org. Chem*, 81, 7924–7930 (2016); DOI: 10.1021/acs.joc.6b01622. [PubMed: 27513176]
9. Gosh A, Bainbridge RT, Stanley LM, Enantioselective model synthesis and progress toward the putative structure of yuremamine J. *Org. Chem*, 81, 7945–7951 (2016); DOI: 10.1021/acs.joc.6b01730. [PubMed: 27494137]
10. Tjandra N, Bax A, Direct Measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science*, 278, 1111–1114 (1997); DOI: 10.1126/science.278.5340.1111. [PubMed: 9353189]
11. Kummerlöwe G, Luy B, Residual dipolar coupling for the configurational and conformation analysis of organic molecules, *Ann. Rep. Nucl. Magn. Reson. Spectrosc*, 68, 193–232 (2009); DOI: [10.1016/S0066-4103\(09\)06804-5](https://doi.org/10.1016/S0066-4103(09)06804-5).
12. Gil RR, Griesinger C, Navarro-Vázquez A, A.; Sun H, in *Structure Elucidation in Organic Chemistry: The Search for the Right Tools*; M.-M. Cid, Bravo J, J., Eds.; Wiley-VCH: New York, 2015; pp. 279–323.
13. Hallwass F, Schmidt M, Sun H, Mazur A, Kummerlöwe G, Luy B, Navarro-Vázquez A, Griesinger C, Reinscheid UM, Residual chemical shift anisotropy (RCSA): A tool for the analysis of the configuration of small molecules, *Angew. Chem. Intl. Ed*, 50, 9487–9490 (2011); DOI: 10.1002/anie.201101784.
14. Kummerlöwe G, Grage SL, Thiele CM, Kuprov I, Ulrich AS, Luy B, Variable angle NMR spectroscopy and its application to the measurement of residual chemical shift anisotropy, *J. Magn. Reson*, 209, 19–30 (2011); DOI: 10.1016/j.jmr.2010.11.019. [PubMed: 21256060]
15. Liu YZ, Prestegard JH, J. H., A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins, *J. Biomol. NMR*, 47, 249–258 (2010); DOI: 10.1007/s10858-010-9427-7. [PubMed: 20506033]
16. Nath N, Schmidt M, Gil RR, Williamson RT, Martin GE, Navarro-Vázquez A, Griesinger C, Liu Y, Determination of relative configuration from residual chemical shift anisotropy, *J. Am. Chem. Soc*, 138, 9548–9556 (2016); DOI: 10.1021/jacs.6b04082. [PubMed: 27294984]
17. Elyashberg ME, Williams AJ, Martin GE, Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation, *Prog. NMR Spectrosc*, 53, 1–104 (2008); DOI: 10.1016/j.pnmrs.2007.04.003.
18. Elyashberg M, Williams A, Blinov K, Contemporary computer-assisted approaches to molecular structure elucidation, RSC Publishing, Cambridge, UK, 2012 pp 482 + xxii.
19. Elyashberg ME, Williams AJ, Computer-based structure elucidation from spectral data, Springer-Verlag, Heidelberg, Germany, 2015, pp 447 + xvi.
20. Schreckenbach G, Ziegler T, Density functional calculations of NMR chemical shifts and ESR *g*-tensors, *Theor. Chem. Acc*, 99, 71–82 (1998).
21. Tantillo DJ, Walking in the woods with quantum chemistry – applications of quantum chemical calculations in natural products research, *Nat. Prod. Rep*, 50, 1079–1086 (2013); DOI: 10.1039/C3NP70028C.
22. Grimblat N, Sarotti AM, Computational chemistry to the rescue: Modern toolboxes for the assignment of complex molecules by GIAO NMR calculations, *Chem. Eur. J*, 22, 12246–12261 (2016); DOI: 10.1002/chem.201601150. [PubMed: 27405775]
23. Navarro-Vázquez A, State of the art and perspectives in the application of quantum chemical prediction of ^1H and ^{13}C shifts and scalar couplings for structural elucidation of organic compounds, *Magn. Reson. Chem*, 55, 29–32 (2016); DOI: 10.1002/mrc.4502. [PubMed: 27531665]
24. Williamson RT, Buevich AV, Martin GE, Parella T, LR-HSQMBC: A sensitive NMR technique to probe very long-range heteronuclear coupling pathways, *J. Org. Chem*, 79, 3887–3894 (2014); DOI: 10.1021/jo500333u. [PubMed: 24708226]

25. Martin GE, Using 1,1- and 1,n-ADEQUATE 2D NMR data in structure elucidation protocols, *Ann. Rep. Nucl. Magn. Reson. Spectrosc.*, 74, 215–291 (2011); DOI: 10.1016/B978-0-08-097072-1.00005-4.
26. Martin GE, Reibarkh M, Buevich AV, Blinov KA, Williamson RT, Development of 1,n-ADEQUATE and modified variants and their application to structure elucidation and spectral assignment problems,” *eMagRes*, 3, 215–234 (2014); DOI: 10.1002/9780470034590.emrstm1370.
27. Saurí J, Bermel W, Buevich AV, Sharaf MHM, Schiff PL Jr., Parella T, Williamson RT, Martin GE, Homodecoupled 1,1- and 1,n-ADEQUATE: Pivotal NMR experiments for the structure revision of cryptospirolepine, *Angew. Chem. Intl. Ed.*, 54, 10160–10164 (2015); DOI: 10.1002/anie.201502540.
28. Blinov KA, Buevich AV, Williamson RT, Martin GE, The impact of LR-HSQMBC very long-range heteronuclear correlation data on computer-assisted structure elucidation, *Org. Biomol. Chem.*, 12, 9505–9509; DOI: 10.1039/C4OB01418A.
29. Mevers E, Saurí J, Liu Y, Moser A, Ramadhar TR, Varlan M, Williamson RT, Martin GE, Clardy J, Homodimericin A: A complex hexacyclic fungal metabolite, *J. Am. Chem. Soc.*, 138 12324–12327 (2016); DOI: 10.1021/jacs.6b07588. [PubMed: 27608853]
30. Shen Y, Vernon R, Baker D, Bax A, *De novo* protein structure generation from incomplete chemical shift assignments, *J. Biomol. NMR*, 43, 63–78 (2009); DOI: 10.1007/s10858-008-9288-5. [PubMed: 19034676]
31. Nath N, d’Auvergne EJ, Griesinger C, Long-range residual dipolar couplings: A tool for determining the configuration of small molecules, *Angew. Chem. Intl. Ed.*, 54, 12706–12710 (2015); DOI: 10.1002/anie.201504432.
32. Trigo-Mouriño P, de la Fuente MC, Gil RR, Sánchez-Pedregal VM, Navarro-Vázquez A, Conformational analysis of the anti-obesity drug Loracaserin in water: How to take advantage of long-range residual dipolar couplings, *Chem. Eur. J.*, 19, 14989–14997 (2013); DOI: 10.1002/chem.201202509. [PubMed: 24105936]
33. Molinski TF, Morinaka BI, Integrated approaches to the configurational assignment of marine natural products, *Tetrahedron* 68, 9307–9343 (2014); DOI: 10.1016/j.tet.2011.12.070.
34. Tackie AN, Boye GL, Sharaf MHM, Schiff PL Jr.; Crouch RC, Spitzer TD, Johnson RL, Dunn J, Minick D, Martin GE, Cryptospirolepine, a unique spiro-nonacyclic alkaloid isolated from *Cryptolepis sanguinolenta*, *J. Nat. Prod.*, 56, 653–670 (1993); DOI: 10.1021/np50095a001. DOI:
35. Martin GE, Hadden CE, Russell DJ, Kaluzny BD, Guido JE, Duholke WK, Stiemsma BA, Thamann TJ, Crouch RC, Blinov KA, Elyashberg ME, Martirosian ER, Molodtsov SG, Williams AJ, Schiff PL Jr., Identification of degradants of a complex alkaloid using NMR cryoprobe technology and ACD/Structure Elucidator, *J. Heterocycl. Chem.*, 39, 1241–1250 (2002); DOI: 10.1002/jhet.5570390619.
36. Yu B, van Ingen H, Vivekanandan S, Rademacher C, Norris SE, Freedberg DI, More accurate $^1\text{J}_{\text{CH}}$ coupling measurement in the presence of $^3\text{J}_{\text{HH}}$ strong coupling in natural abundance, *J. Magn. Reson.*, 215, 10–22 (2012); DOI: 10.1016/j.jmr.2011.09.037. [PubMed: 22227287]
37. Lorenc C, Saurí J, Moser A, Buevich AV, Williams AJ, Williamson RT, Martin GE, Peczu MW, Turning spiroketals inside out: A rearrangement triggered by an enol ether epoxidation, *ChemistryOpen*, 4, 577–580 (2015); DOI: 10.1002/open.201500122. [PubMed: 26491634]
38. San Feliciano A, Merdard M, Miguel del Corral JM, Aramburu A, Gordaliz M, Barrero AF, Aquatolide. A new type of humulane-related sesquiterpene lactone, *Tet. Lett.*, 30, 2851–2854 (1989).
39. Lodewyk MW, Soldi C, Jones PB, Olmstead MM, Rita J, Shaw JT, Tantillo DJ, The correct structure of aquatolide – experimental validation of a theoretically-predicted structure revision, *J. Am. Chem. Soc.*, 134, 18550–18553 (2012); DOI: 10.1021/ja3089394. [PubMed: 23101682]
40. Sarotti AM, Successful combination of computationally inexpensive GIAO ^{13}C NMR calculations and artificial neural network pattern recognition: a new strategy for simple and rapid detection of structural misassignments, *Org. Biomol. Chem.*, 11, 4847–4859 (2013); DOI: 10.1039/C3OB40843D. [PubMed: 23779148]

41. Saya JM, Vos K, Kleinnijenhuis RA, van Maarseveen JH, Ingemann S, Hiemstra H, Total synthesis of aquatolide, *Org. Lett* 17, 3892–3894 (2015); DOI: 10.1021/acs.orglett.5b01888. [PubMed: 26186274]
42. Buevich AV, Elyashberg ME, Synnergistic combination of CASE algorithms and DFT chemical shift predictions: A powerful approach for structure elucidation ,verification, and revision, *J. Nat. Prod*, 79, 3104–3116 (2016); DOI: 10.1021/acs.jnatprod.6b00799.
43. Bax A, Weak alignment offers new NMR opportunities to study protein structure and dynamics, *Protein Sci*, 12, 1–16 (2003); DOI: 10.1110/ps.0233303. [PubMed: 12493823]
44. In an unpublished study on a nine residue cyclized peptide, out of 100,000 rotamers generated by a diversity-orientated conformer search, three incorrectly folded structures were identified that had Q-factors of 0.33, 0.33, and 0.35, respectively. In contrast, comparison of 26 RDCs with the single crystal X-ray structure provided a significantly better Q-factor of 0.23.. The average Q-factor in the test set was 0.81 with a standard deviation of 0.10. Therefore, a false positive rate in this case is estimated to be ~30 per million even with a loose Q-factor cutoff of 0.35. This rate is expected to drop orders of magnitude further if carbonyl RCSA data are also included.
45. Gil-Silva LF; Santamaría-Fernández R; Navarro-Vázquez A; Gil RR, Collection of Scalar and Residual Dipolar Couplings Using a Single Experiment, *Chemistry Eur. J.*, , 472–476 (2016); DOI: 10.1002/chem.201503449.
46. Castañar L; García M; Hellemann E; Nolis P; Gil RR; Parella T, One-Shot Determination of Residual Dipolar Couplings: Application to the Structural Discrimination of Small Molecules Containing Multiple Stereocenters, *J. Org. Chem*,81, 11126–11131 (2016); DOI: 10.1021/acs.joc.6b02103. [PubMed: 27709936]
47. Enthart A; Freudenberger JC; Furrer J; Kessler H; Luy B, The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings, *J. Magn. Reson*, 192, 314–322 (2008); DOI.org/10.1016/j.jmr.2008.03.009. [PubMed: 18411067]
48. Ottiger M; Bax A, How tetrahedral are methyl groups in proteins? A liquid crystal NMR study, *J. Am. Chem. Soc*, 121, 4690–4695 (1999); DOI: 10.1021/ja984484z.
49. Sánchez-Pedregal VM; Santamaría-Fernández R; Navarro-Vázquez A, Residual dipolar couplings of freely rotating groups in small molecules. Stereochemical assignment and side-chain conformation of 8-phenylmenthol, *Org Lett*, 11, 1471–1474 (2009); DOI: 10.1021/ol8029565. [PubMed: 19320500]

Figure 1.

Illustration of partial molecular ordering at the gel polymer surface. The polymer structure (left) represents a poly(methylmethacrylate) (PMMA) filament. The filament is displayed along B_0 , demonstrating the excess projection along this direction in the actual NMR experimental setup. The rotation of the analyte (center) is restricted at the polymer surface due to steric occlusion, leading to partial molecular alignment. The experimentally determined principal axis frame of the alignment tensor, a particular molecular frame in which the alignment tensor is diagonalized, is displayed as the yellow dashed frame on the far right. The most ordered axis, P_z , can frequently be understood intuitively based on the molecular shape. For example, in aquatolide, the most ordered molecular axis P_z is, as expected, roughly perpendicular to the planar envelope of this nearly flat molecule.

Figure 2.

The principle of RCSA-based model differentiation shown using aquatolide as an example. The chemical shielding anisotropy tensors are shown in magenta and green for two sp^2 carbons for which RDC data are unavailable. In the “revised” and the “original” structures, these two anisotropy tensors have different orientations relative to each other as clearly seen here, as well as relative to anisotropy tensors of the other carbons which are not displayed for visual clarity. The “revised” structure is clearly favored by much better agreement between experimental RCSAs and calculated values predicted using alignment tensor parameters determined from all data including both RCSA and RDC. Correlations for carbons whose chemical shielding anisotropy tensors are not displayed are shown in tan.

Figure 3.

Cryptospirolepine, originally reported as 1,(34) was found to degrade to two major compounds, 2 and 3, in 2002.(35) Although the formation of 2 could be mechanistically rationalized, the formation of 3 could not, suggesting that the originally reported structure 1, was likely incorrect. Using 1.7 mm MicroCryoProbe™ capabilities in conjunction with the recently reported 1,1-HD-ADEQUATE experiment allowed revision of the structural assignment to 4.(27) The formation of both 2 and 3 can be mechanistically rationalized from 4.

Figure 4. RDC/RCSA analysis for structural assignment of cryptospirolepine.

A) Plot of the experimental vs. calculated RDC (red) and RCSA (blue) data for the incorrect structure of cryptospirolepine reported in 1993.(34) $Q = 0.245$. B) Plot of the experimental vs. calculated RDC/RCSA data for the revised structure of cryptospirolepine (4) reported in 2015.(27) $Q = 0.122$.

Figure 5.

Division of RDC and RCSA data based on their structural locations narrows the structural error in **1** to the top half. The correlation associated with the common green part is nearly identical between **1** and **4**, whereas the correlation associated with the diverging red part is a substantially poorer fit in **1** than in **4**, indicating the red part in **1** contains the structural error.

Figure 6. RDC/RCSA analysis for spiroketal rearrangement.

A) Plot of the RDC (red data points) and RCSA (blue data points) for the structure of the enol-epoxide rearrangement product confirmed with 1,1-HD-ADEQUATE data(27) as 5. The Q-value was 0.20. B) Correlation plot of the RDC/RCSA data for the alternative structure, 6, suggested by CASE, associated with a Q-value of 0.55. RDC/RCSA data alone can clearly differentiate between the two structures.

Figure 7.

Plots of the calculated vs. experiment RDC (red)/RCSA (blue) data for A) the revised structure of aquatolide, 8; B) the originally proposed incorrect structure of aquatolide (7) and C and D two alternative structures, 9 and 10, respectively. Clearly there is a vast difference in the Q-value for the correct structure, 8, and the originally reported structure of aquatolide, 7. The alternative structures generated by CASE, 9 and 10, had intermediary Q-values of 0.23 and 0.59, respectively. A two-fold difference in the Q-value between the correct structure, 8, and the best alternative structure from the CASE program, 9, still allowed an ample basis for choosing between the structures.