

## COMMUNICATION

# Incomplete Data Sets in the Model Free Analysis of Experimental Residual Dipolar Couplings in Small Organic Compounds

Received 00th January 20xx,  
Accepted 00th January 20xx

Felix A. Roth, Volker Schmidts, Jan Rettig and Christina M. Thiele\*

DOI: 10.1039/x0xx00000x

Residual dipolar couplings (RDCs) contain information on the relative arrangement and dynamics of internuclear spin vectors in chemical compounds. Classically, RDC data is analyzed by fitting to structure models, while model-free approaches (MFA) directly relate RDCs to the corresponding internuclear vectors. The recently introduced software TITANIA implements the MFA and extracts structure and dynamics parameters directly from experimental RDCs to facilitate *de novo* structure refinement for small organic compounds. Herein, we focus on the challenges faced with experimentally inaccessible data often encountered for small organic compounds leading to voids in the RDC matrix and the concomitant effects on the structure refinement. It is shown that RDC data sets obtained experimentally from currently available alignment media and measurement methods are of sufficient quality to allow relative configuration determination even when the relative configuration of the analyte is completely unknown.

Anisotropic NMR parameters such as RDCs can be used in modern NMR spectroscopy to elucidate three-dimensional structures.<sup>1–3</sup> For small organic compounds, this can generally be done *via* structure validation,<sup>4–6</sup> by using the relationship between the RDCs **D** and the structure matrix **B** (direction cosine matrix)<sup>8</sup> to discriminate structure proposals in a fitting procedure.

$$\mathbf{D} = \text{diag}[D_{\max}] \mathbf{B} \cdot \mathbf{A} \quad (1)$$

The alignment matrix **A** is determined by SVD (singular value decomposition) and used to back-calculate RDCs with equation (1), where  $D_{\max}$  is the maximum static dipolar coupling constant for each interacting spin pair. The accordance of the structure proposal and the RDCs is described by the quality factor *Q* (or analogous expressions).<sup>9</sup> The disadvantage of this approach is that the description of structure and dynamics are

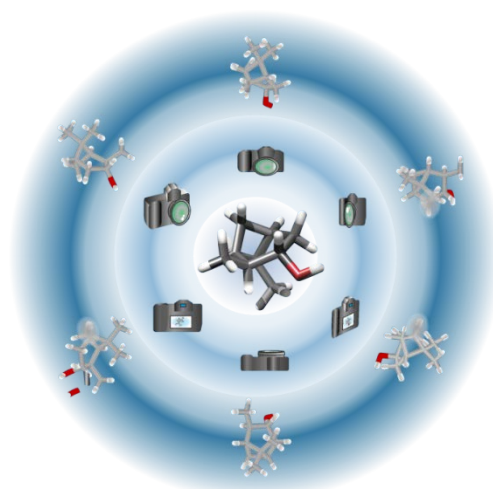


Figure 1: By measuring (camera) RDCs in an alignment medium (perspective), a projection (outer structures) of the molecule can be obtained. This result itself does not contain all 3D structural information needed. If five (or more) sufficiently different (linearly independent) projections are combined, the 3D structure (central structure) can be reconstructed and its dynamics determined. Non-measurable RDCs lead to missing information (shadows on the projections), which complicate the reconstruction.

strongly limited by the choice of structural proposals and mathematical models used to describe conformational flexibility.

Part of this problem is addressed by modern structure optimization algorithms,<sup>10–14</sup> as structure proposals are refined based on experimental data. However, to the best of our knowledge there are no implementations that determine vector orientations directly from RDCs and yield local order parameters for small organic molecules at the same time.

These local order parameters are the result of the model free approach (MFA),<sup>15, 16</sup> which was originally developed for the analysis of relaxation data. In order to apply the MFA to RDC data, RDC data sets from at least five alignment media with linear independence are required.<sup>17–19</sup> The working principle of the MFA is shown in Figure 1. The acquisition (the camera) of an RDC data set in an alignment medium (the perspective) leads to a "projection" of the measured structure (outer planar structures). The individual projections can be combined into a

<sup>a</sup> Clemens-Schöpf-Institut für Organische Chemie und Biochemie, Technical University of Darmstadt, Alarich-Weiss-Str. 16, 64287 Darmstadt

Electronic Supplementary Information (ESI) available: One pdf file containing all runs and their discussion and two zip archive as supplementary material, which contain all input / output files and all experimental spectra. See DOI: 10.1039/x0xx00000x

3D structure (middle structure in Figure 1) provided that the projections are based on sufficiently different viewpoints (multiple alignment conditions), as in the MFA. In practice, this is accomplished by determination of the direction vectors for the RDC spin pairs from iterative application of the MFA. For this purpose, equation (1) is recast into an expression where the structure is represented by spherical harmonics  $Y_{2,m}$  and the orientation tensors by the corresponding Wigner rotations.

$$\tilde{\mathbf{D}} = \mathbf{F} \cdot \mathbf{Y} \quad (2)$$

In equation (2),  $\tilde{\mathbf{D}}$  is the transposed, normalized (to  $D_{\max}$  and  $A_a$ , the axial component of the alignment tensor) RDC matrix and  $\mathbf{F}$  the matrix containing the Wigner rotation elements corresponding to the orientation parameters of the individual alignment media. For a detailed description of the mathematics, please refer to the ESI. The advantage of the MFA is that structure and dynamics information are computed from the RDCs simultaneously, without relying on a pre-defined structure proposal. Only the connectivity needs to be known.

Recently, we introduced the software TITANIA (TITANIA performs iterative analysis of independent alignments),<sup>20</sup> which enables MFA in the field of NMR spectroscopy of small organic compounds for the first time. The software can determine the structure of small organic molecules based on any proposed structure by an iterative algorithm. As a first step, we demonstrated that the relative configuration of small organic compounds can be determined using *simulated* RDCs derived from experimental orientation parameters. In addition, random Gaussian error was added to simulate the applicability of TITANIA under practical conditions. We were able to show that it was not necessary to define stereo information of any kind as the starting point of the optimization.

Modern pulse sequences<sup>21–31</sup>, developed in recent years, have simplified access to long-range RDCs (we focus mainly on  $^nD_{HH}$ )<sup>29, 30</sup> and new alignment media<sup>32–40</sup> allow the measurement of potentially different alignments of the same molecule in different (co-)solvents or, for example, at different temperatures. This progress has now made it possible to obtain sufficiently large and different RDC data sets and apply the MFA to small organic compounds as demonstrated here.

In experimental spectra not all theoretically possible couplings are always accessible. This is typically the case due to broadened lines (elicited by the large number of dipolar couplings), strong or too large coupling (if a too high degree of orientation is present) or simply obfuscation by spectral overlap. These inaccessible RDCs lead either to the reduction of the total data set, since the corresponding rows (RDC vectors) or columns (alignment media) have to be removed from the RDC matrix, or to undefined elements in this very matrix. In such a case, the presented matrix equations (1) and (2) cannot be applied without further modification. If such undefined RDCs would be set to 0.0 Hz, this would introduce wrong structure or dynamics information in the analysis (see ESI).

Here we present two algorithms that reduce the impact of undefined RDCs (voids in  $\mathbf{D}$ ) in the MFA. These are called *recalculation* and *weighting scheme*, respectively. The recalculation scheme takes advantage of the iterative nature of TITANIA. In each iteration step a refined structure is generated, on the basis of which the equations (1) and (2) are solved. Thus, at the beginning of every iteration, undefined RDCs are back-calculated. The weighting scheme decomposes the matrix equations into vector equations. Thereby only one alignment medium (adjusted equation 1) and RDC vector (adjusted equation 2), respectively, is considered at a time. In this scheme, the undefined RDCs are removed from the equations by weighting factors prior to SVD (for more information see ESI).

By back-calculating the RDCs the TITANIA optimization is expected to proceed more slowly, since structural errors affect not only the current iteration, but also each subsequent step. On the other hand, if RDCs are removed, especially if the number of alignment media is small, errors due to the reduction of matrix rank in the formation of pseudoinverses by SVD have to be expected. This data reduction is in competition with the requirement of linear independence of the orientations, which may be more challenging to achieve for small organic compounds due to the intrinsically smaller amount of data as compared to biomacromolecules. For a more detailed description of the algorithms, please refer to Chapter 1 of the ESI.

The applicability of TITANIA, also regarding the problem of missing elements in the RDC matrix, will be demonstrated on the NMR spectroscopically well studied compound isopinocampheol **1** (IPC). For this purpose, we measured 38 RDCs (11  $^1D_{CH}$ <sup>24, 26</sup> or derived  $^1D_{CC}$ <sup>41</sup> and 27 long-range  $^nD_{HH}$ <sup>29</sup> RDCs) in six alignment conditions<sup>7, 36, 42</sup> (hereafter called *set*). This results in an RDC matrix with up to 228 elements, with a total of 23 undefined RDCs. To investigate the influence of the latter, four RDC setups were generated from the described matrix. These have different amounts and clusters of undefined RDCs. The process of generating the setups is described in the ESI (see chapter 7). Setup **A-6** uses all (38) RDCs with 23 undefined elements (228 RDCs / 10.0%). **B-6** with a maximum of 33 RDCs per set and a total of 12 undefined elements (198 RDCs / 6.1%) was generated by removing as many RDCs with multiple undefined elements as possible. **C-6** with a maximum of 28 RDCs per set and 8 undefined elements (168 RDCs / 4.8%) was created by having as few undefined elements per set as possible. The completely reduced setup **D-6** with 24 RDCs has no undefined elements (144 RDCs / 0.0%). The assignment of the media and RDC pairs can be found in the ESI.

First, the setups are analysed using SECONDIA (self-consistency of dipolar couplings analysis),<sup>43</sup> which is used in biomolecular NMR spectroscopy to analyse heterogeneities between RDC datasets. In addition, as a principal component analysis (PCA), SECONDIA collates information on the linear independence of the RDC datasets. As discussed above, SECONDIA is also sensitive to voids in the RDC matrix, requiring data reduction or back-calculation<sup>43</sup> or adjustments to the

underlying equations. The adjustments we made and the resulting data are summarized in the ESI. The key result of the SECONDA analysis is that the data sets contain at least three strongly different (linearly independent) orientations. This is especially clear when looking at the cumulative sum of the principle variances. For all setups, the largest three principle variances ( $\lambda_1$ – $\lambda_3$ ) represent >98% of the total variance. The following two principle variances ( $\lambda_4$  and  $\lambda_5$ ) are greater than  $1e-1$ . Thus, mathematically, a rank of 5 is always achieved. As a result of the analysis, the optimization of the rigid molecule IPC should succeed when using the four RDC setups within TITANIA.

Four optimizations are performed for each setup, which differ in the choice of the algorithm for handling the voids and the choice of the starting structure. Thus, both schemes were started from the C3 epimer to ensure comparability of the optimizations. In addition, a randomly generated geometry was used. The trajectories of the chiral volumes of the three centres C4 (two  $^1D_{CH}$  RDCs), C3 (one  $^1D_{CH}$  RDC), and C6 (two  $^1D_{CC}$  RDCs) in the course of the optimizations of setup A-6 (23 undefined / 228 total RDCs) are shown in Figure 2, where the distributions of the chiral volumes are shown as a histogram on the y-axis.

Panels a and d show how most trajectories are expected to progress: many inversion steps sample the relevant geometries during the initial stages with only a few changes quickly reverting to the stable configuration in the latter stages, once convergence of the full trajectory is reached. With only two exceptions, the configuration of the centres can be unambiguously determined when considering the final structure and the distribution of the chiral volumes.

The first exception is C6 in the recalculation scheme with C3 epimer as starting structure (green line in panel c). Here, just before reaching convergence, centre C6 is inverted to the wrong configuration. The statistical distribution, on the other hand, shows the correct configuration. In the Cartesian

coordinates, the final structure has no reasonable geometry at the corresponding centre. Due to the small number and magnitude of RDCs at a quaternary centre, its correct configuration is sometimes difficult to determine in TITANIA optimizations. For the case described here, however, an interchange of the C6–C8 and C6–C9 bond vectors would have little effect on the structure determination, since these are diastereotopic methyl groups.

The second exception is C3 in the weighting scheme with random starting structure (blue line in panel b). Here, the final structure is stable in the correct relative configuration. The distribution, on the other hand, does not show a single maximum as would be expected (a more detailed discussion is given in the ESI). However, when examining the Cartesian coordinates' trajectory, it becomes obvious that C3 is much more stable in the final structure, while the incorrect, early structures show a distorted geometry. The ambiguous distribution thus arises from the comparatively fast convergence which complicates the statistical evaluation of the chiral volume distribution.

Figure 2 shows that TITANIA can handle undefined elements in the experimental RDC matrix for both algorithms. Chiral centres involved in multiple RDCs (e.g. C4) show very narrow distributions in all cases and a very stable chiral volume in late iterations. This stability decreases with lower numbers of RDCs. The discussed ambiguities can be identified and interpreted by close inspection of the Cartesian coordinates. Additional indications for incorrect solutions are bond lengths that deviate strongly from literature data (*vide infra*) or inconsistent configurations when starting from different structures. Thus, changing the starting structure for inconclusive runs is a tool for confirming relative configurations.

In addition to the previously discussed trajectories of setup A-

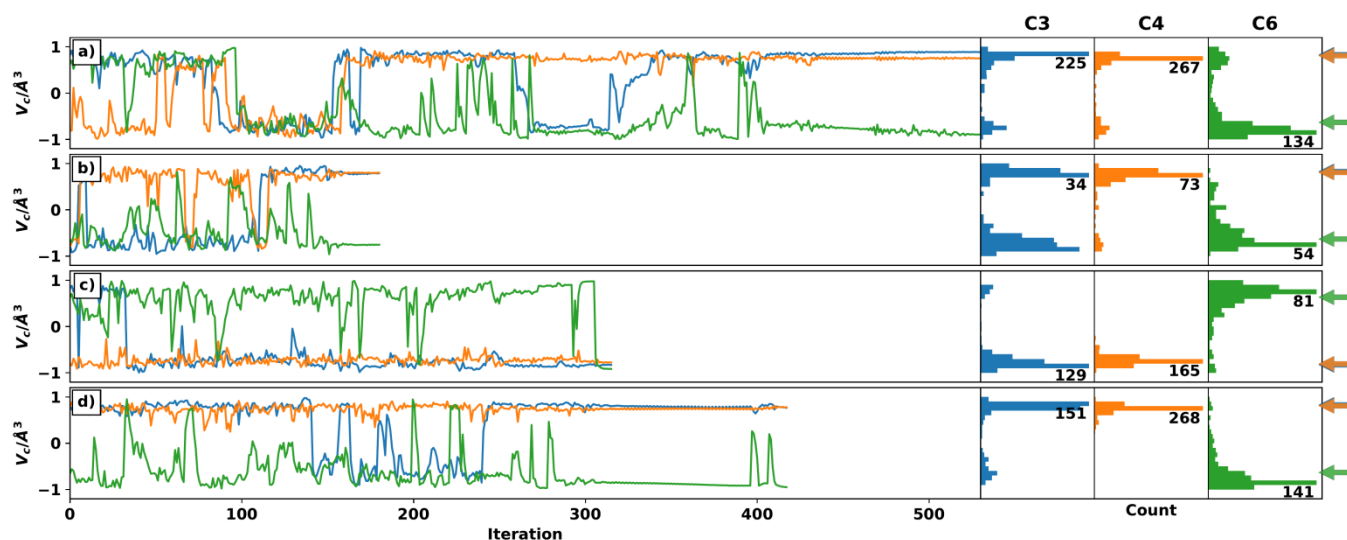
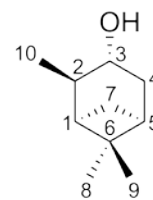


Figure 2: Trajectories of the chiral volumes of setup A-6 (all RDC data, 23 undefined elements) using the weighting scheme (panel a/b) and the recalculation scheme (panel c/d), respectively. For better comparability, the C3 epimer was used as the starting structure in panel a/c. The starting structures in panel b/d were chosen randomly. The arrows (blue and orange are almost identical) on the right side show the correct chiral volumes of the respective enantiomer obtained. Only C6 in panel c has the incorrect relative configuration in the final structure. It can be assigned correctly by statistical analysis of the distribution of chiral volumes shown as histograms on the right.

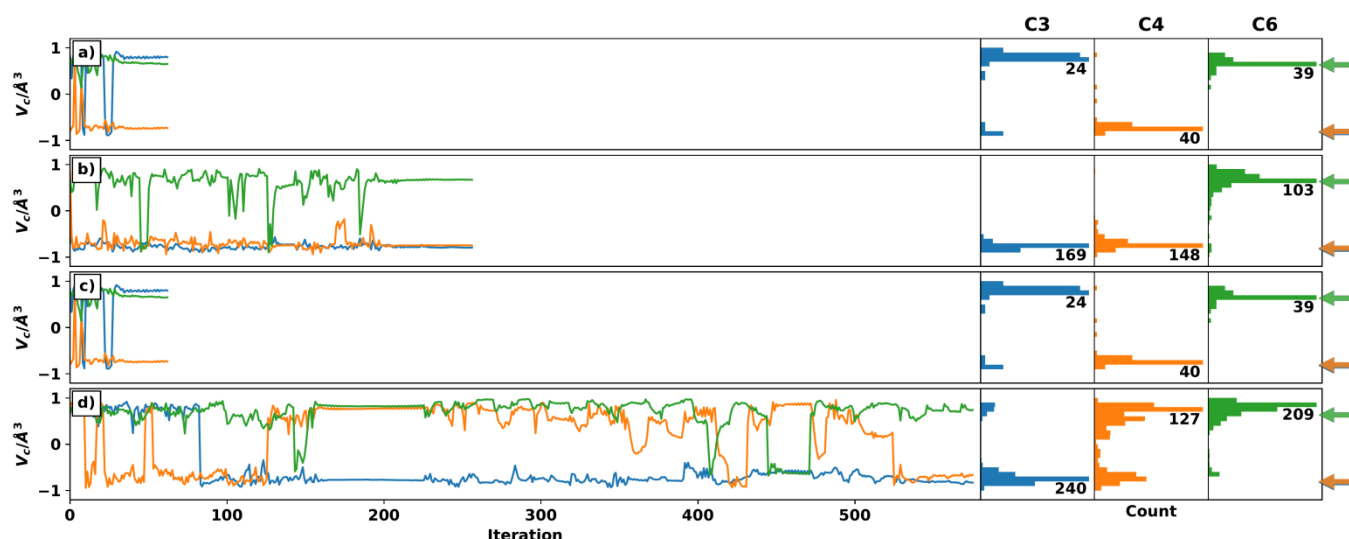


Figure 3: Trajectories of the chiral volumes of setup **D-6** (reduced RDC data, no undefined elements) using the weighting scheme (panel **a/b**) and the recalculation scheme (panel **c/d**), respectively. For better comparability, the C3 epimer was used as the starting structure in panel **a/c**. The starting structures in panel **b/d** were chosen randomly. The arrows (blue and orange are almost identical) on the right side show the correct chiral volumes of the respective enantiomer obtained. The panels **a** and **c** are identical, due to the fact that no RDCs are missing. The wrong configuration of C3 is discussed in the main text and SI.

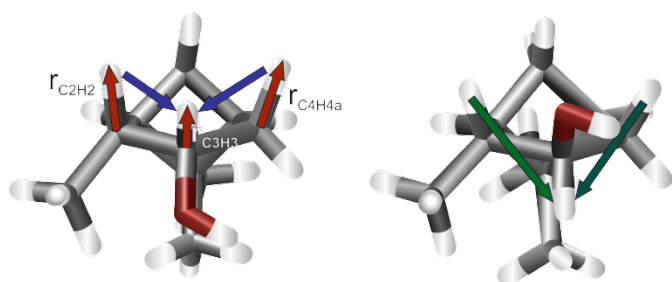


Figure 4: Final structure of **D-6** starting from *epi*-C3 (left) and the reference structure from literature.<sup>7</sup> The red arrows show the bond lengths, that are distorted to compensate the difference in the long-range  $^nD_{HH}$  coupling vectors (blue and green). Starting the optimization from a different geometry, e.g. random coordinates, alleviates this problem.

**6**, the chiral volume trajectories and Monte-Carlo data of all other setups are given in the ESI. The optimization of setups with undefined elements in the RDC matrix should be accompanied by the statistical consideration of the Cartesian coordinate trajectories for reliable structure elucidation. The weighting scheme is superior to the recalculation scheme, especially for smaller data sets with voids (**B-6** and **C-6**, see ESI). Additionally, the setup with the smallest data set (**D-6**) but without undefined elements will be discussed here (see Figure 3).

The optimizations with setup **D-6** were performed using both schemes, although there are no voids in the RDC data set. On the one hand, it can be shown that panels **a** and **c** have the same trajectory, since the same starting structure was used. Thus the choice of the different schemes has no effect for complete datasets. On the other hand, setup **D-6** shows the possibility of confirming configurations by changing the starting structure (panel **b** and **d**). This has proven to be important for C3 and C4. Centre C4 shows an ambiguous assignment in panel **d** based on the distribution of chiral volumes. The reason for this is the late inversion to the correct

configuration, which can be detected by checking the Cartesian coordinates as described above.

Similarly, the wrong configuration of C3 in the identical trajectories of panels **a** and **c** can be detected by the distorted bond lengths at C3 and the neighbouring carbons C2 and C4 of the final structure. For example, the bond length  $r_{C3H3} = 0.82 \text{ \AA}$  is exceptionally short, which is typical for stereogenic centres in the wrong configuration. In addition, the bonds *vicinal* to this are also significantly elongated with  $r_{C2H2} = 1.24 \text{ \AA}$  and  $r_{C4H4a} = 1.29 \text{ \AA}$ . The reason for this becomes clear when assessing the final (left) and reference structure (right) in Figure 4. The aforementioned bonds (red arrows) are shortened or lengthened to optimize the RDC vectors (blue arrows) to fit better to the actual orientation (green arrows). These distorted geometries for wrongly configured stereogenic centres are often elicited by the inverse vector solution.<sup>20</sup> In this special case, however, it is the result of too fast convergence. In the ESI, two additional trajectories are shown, in which either the weighting of the bond lengths was reduced by a factor of 5 or the number of iteration steps was increased by more stringent convergence criteria. Both options lead to the correct configuration at all centres.

## Conclusions

With the presented data it was shown that TITANIA is able to perform a *de novo* structure determination of small organic compounds with experimental RDCs obtaining the correct configuration in most cases. In contrast to previous approaches the new quality of information, namely *direct vector information*, accessible from multiple alignment data sets was used to achieve this goal. In the discussed borderline cases that did not lead to the correct configuration of the final structures in the first try, these errors could be detected by assessing the geometry in detail or the statistical processing of

the data. Here the correct structure was additionally obtained by changing the starting structure or adjusting the settings. The selection of different alignment media used here is limited to polypeptide-based alignment media. While the alignments proved sufficiently linearly independent, we expect a further broadening of the applications when employing the full range of all recently developed alignment media.<sup>5, 32-40, 42</sup> Further improvements are expected upon the inclusion of more difficult to access long-range  $^nD_{CH}$  or  $^{1/n}D_{CC}$  couplings. The use of additional complementary data is expected to significantly improve the determination of flexibility or reduce the influence of experimental errors and even missing RDCs. Moreover, additional couplings<sup>21-31</sup> would drastically simplify the determination of challenging centres (such as C6 in IPC). Herein we were able to demonstrate the capability of TITANIA to determine the relative configuration of IPC using an incomplete, experimental RDC dataset. This revolutionizes the structure elucidation of small organic compounds. By the continuous developments of new (stimuli responsive) alignment media and modern pulse sequences the full investigation of flexible molecules, including the determination and analysis of local order parameters, is currently under investigation.

## Conflicts of interest

There are no conflicts of interest to declare.

## Notes and references

The source code for the TITANIA program as well as precompiled binaries are available from the authors upon request.

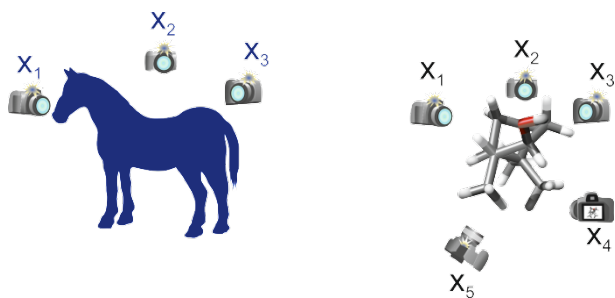
1. B. Böttcher and C. M. Thiele, *Encyclopedia of Magnetic Resonance*, 2012, DOI: 10.1002/9780470034590.emrstm1194.
2. R. R. Gil, *Angew. Chem. Int. Ed.*, 2011, **50**, 7222-7224.
3. G. Kummerlöwe and B. Luy, in *Annu. Rep. NMR Spectrosc.*, ed. G. A. Webb, Academic Press, 2009, vol. 68, pp. 193-232.
4. J. A. Losonczy, M. Andrec, M. W. F. Fischer and J. H. Prestegard, *J. Magn. Reson.*, 1999, **138**, 334-342.
5. V. Schmidts, *Magn. Reson. Chem.*, 2017, **55**, 54-60.
6. F. Kramer, M. V. Deshmukh, H. Kessler and S. J. Glaser, *Conc. Magn. Reson. A*, 2004, **21A**, 10-21.
7. A. Marx, V. Schmidts and C. M. Thiele, *Magn. Reson. Chem.*, 2009, **47**, 734-740.
8. J. R. Tolman, *J. Am. Chem. Soc.*, 2002, **124**, 12020-12030.
9. G. Cornilescu, J. L. Marquardt, M. Ottiger and A. Bax, *J. Am. Chem. Soc.*, 1998, **120**, 6836-6837.
10. D. Ban, T. M. Sabo, C. Griesinger and D. Lee, *Molecules*, 2013, **18**, 11904-11937.
11. K. Chen and N. Tjandra, *Top. Curr. Chem.*, 2012, **326**, 47-67.
12. G. Cornilescu, R. F. Ramos Alvarenga, T. P. Wyche, T. S. Bugni, R. R. Gil, C. C. Cornilescu, W. M. Westler, J. L. Markley and C. D. Schwieters, *ACS Chem. Biol.*, 2017, **12**, 2157-2163.
13. S. Immel, M. Köck and M. Reggelin, *Chem. Eur. J.*, 2018, **24**, 13918-13930.
14. P. Tzvetkova, U. Sternberg, T. Gloge, A. Navarro-Vázquez and B. Luy, *Chem. Sci.*, 2019, **10**, 8774-8791.
15. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 1982, **104**, 4546-4559.
16. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 1982, **104**, 4559-4570.
17. J. Meiler, J. J. Prompers, W. Peti, C. Griesinger and R. Brüschweiler, *J. Am. Chem. Soc.*, 2001, **123**, 6098-6107.
18. W. Peti, J. Meiler, R. Brüschweiler and C. Griesinger, *J. Am. Chem. Soc.*, 2002, **124**, 5822-5833.
19. N.-A. Lakomek, K. F. A. Walter, C. Farès, O. F. Lange, B. L. de Groot, H. Grubmüller, R. Brüschweiler, A. Munk, S. Becker, J. Meiler and C. Griesinger, *J. Biomol. NMR*, 2008, **41**, 139.
20. F. A. Roth, V. Schmidts and C. M. Thiele, *under review*, 2021.
21. M. Kurz, P. Schmieder and H. Kessler, *Angew. Chem. Int. Ed.*, 1991, **30**, 1329-1331.
22. D. Uhrin, G. Batta, V. J. Hruby, P. N. Barlow and K. E. Kövér, *J. Magn. Reson.*, 1998, **130**, 155-161.
23. A. Meissner and O. W. Sørensen, *Magn. Reson. Chem.*, 2001, **39**, 49-52.
24. K. Fehér, S. Berger and K. E. Kövér, *J. Magn. Reson.*, 2003, **163**, 340-346.
25. K. E. Kövér and K. Fehér, *J. Magn. Reson.*, 2004, **168**, 307-313.
26. A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler and B. Luy, *J. Magn. Reson.*, 2008, **192**, 314-322.
27. C. M. Thiele and W. Bermel, *J. Magn. Reson.*, 2012, **216**, 134-143.
28. L. Castañar, J. Sauri, R. T. Williamson, A. Virgili and T. Parella, *Angew. Chem. Int. Ed.*, 2014, **53**, 8379-8382.
29. D. Sinnaeve, J. Ilgen, M. E. Di Pietro, J. J. Primozic, V. Schmidts, C. M. Thiele and B. Luy, *Angew. Chem. Int. Ed.*, 2020, **59**, 5316-5320.
30. D. Sinnaeve, M. Foroozandeh, M. Nilsson and G. A. Morris, *Angew. Chem. Int. Ed.*, 2016, **55**, 1090-1093.
31. W. Koźmiński and D. Nanz, *J. Magn. Reson.*, 2000, **142**, 294-299.
32. A. Marx, B. Böttcher and C. M. Thiele, *Chem. Eur. J.*, 2010, **16**, 1656-1663.
33. N.-C. Meyer, A. Krupp, V. Schmidts, C. M. Thiele and M. Reggelin, *Angew. Chem. Int. Ed.*, 2012, **51**, 8334-8338.
34. C. Merle, G. Kummerlöwe, J. C. Freudenberger, F. Halbach, W. Stöwer, C. L. v. Gostomski, J. Höpfner, T. Beskers, M. Wilhelm and B. Luy, *Angew. Chem. Int. Ed.*, 2013, **52**, 10309-10312.
35. X. Lei, Z. Xu, H. Sun, S. Wang, C. Griesinger, L. Peng, C. Gao and R. X. Tan, *J. Am. Chem. Soc.*, 2014, **136**, 11280-11283.
36. S. Jeziorowski and C. M. Thiele, *Chem. Eur. J.*, 2018, **24**, 15631-15637.
37. M. Schwab, V. Schmidts and C. M. Thiele, *Chem. Eur. J.*, 2018, **24**, 14373-14377.
38. M. Schwab, D. Herold and C. M. Thiele, *Chem. Eur. J.*, 2017, **23**, 14576-14584.
39. L. F. Gil-Silva, R. Santamaría-Fernández, A. Navarro-Vázquez and R. R. Gil, *Chem. Eur. J.*, 2016, **22**, 472-476.

## COMMUNICATION

Journal Name

40. M. Reller, S. Wesp, M. R. M. Koos, M. Reggelin and B. Luy, *Chem. Eur. J.*, 2017, **23**, 13351-13359.
41. L. Verdier, P. Sakhaei, M. Zweckstetter and C. Griesinger, *J. Magn. Reson.*, 2003, **163**, 353-359.
42. M. Hirschmann, D. S. Schirra and C. M. Thiele, *Macromolecules*, 2021, **54**, 1648-1656.
43. J.-C. Hus, W. Peti, C. Griesinger and R. Brüschweiler, *J. Am. Chem. Soc.*, 2003, **125**, 5596-5597.

## COMMUNICATION



Combining anisotropic NMR data from multiple alignments is like combining photos into a 3D image and enables de novo structure determination of small organic compounds.

# Incomplete Data Sets in the Model Free Analysis of Experimental Residual Dipolar Couplings in Small Organic Molecules

Felix A. Roth, Volker Schmidts, Jan Rettig, Christina M. Thiele\*

## Contents

<b>1</b>	<b>Theoretical background</b>	<b>2</b>
1.1	Information from RDC data . . . . .	2
1.2	Structure generation from spherical coordinates . . . . .	4
<b>2</b>	<b>Handling undefined RDCs</b>	<b>6</b>
<b>3</b>	<b>Sample preparation</b>	<b>8</b>
<b>4</b>	<b>NMR-experiment conditions</b>	<b>9</b>
<b>5</b>	<b>List of RDCs</b>	<b>11</b>
<b>6</b>	<b>Orientations of IPC</b>	<b>12</b>
<b>7</b>	<b>RDC setups</b>	<b>13</b>
<b>8</b>	<b>SECONDA</b>	<b>18</b>
<b>9</b>	<b>Trajectories</b>	<b>22</b>
<b>10</b>	<b>Chiral volume distributions</b>	<b>29</b>
<b>11</b>	<b>Adapted runs of setup D-6</b>	<b>37</b>
<b>12</b>	<b>Content of Supplementary Material</b>	<b>39</b>

---

\* *Clemens-Schöpf-Institut für Organische Chemie und Biochemie, Technical University of Darmstadt, Alarich-Weiss-Str. 16, 64287 Darmstadt (Germany).*



# 1 Theoretical background

## 1.1 Information from RDC data

Recently we presented the TITANIA protocol, explaining the fundamentals of the approach and applying it to simulated RDCs.<sup>[1]</sup> This chapter will summarize the framework introduced, which is described in more depth elsewhere.<sup>[2,3]</sup>

The basic definition of RDCs  $D$  uses the averaged angle  $\Theta$  between the inter spin vector and the magnetic field.

$$D = D_{\max} \left\langle \frac{3 \cos^2 \Theta - 1}{2} \right\rangle \quad (1)$$

Here  $D_{\max}$  is the maximum dipolar coupling and the angular brackets  $\langle \rangle$  denote the time and ensemble averaging. A more common expression of RDCs used for structure validation uses a formalism introduced by SAUPE.<sup>[4]</sup> A generalized formulation of the Saupe approach uses the normalized cosine matrix  $\mathbf{B}$  and the alignment matrix  $\mathbf{A}$ .<sup>[5]</sup> These matrices contain the independent elements  $T_{ij}$  of the corresponding second rank tensor  $\mathbf{T}$ .

$$\mathbf{T}^{(2)} = \left[ T_{zz}, \frac{1}{\sqrt{3}} (T_{xx} - T_{yy}), \frac{2}{\sqrt{3}} T_{xy}, \frac{2}{\sqrt{3}} T_{xz}, \frac{2}{\sqrt{3}} T_{yz} \right] \quad (2a)$$

$$A_{ij} = \left\langle \frac{1}{2} (3 \cos \alpha_i \cos \alpha_j - \delta_{ij}) \right\rangle \quad (2b)$$

$$B_{ij} = \left\langle \frac{1}{2} (3 \cos \beta_i \cos \beta_j - \delta_{ij}) \right\rangle \quad (2c)$$

$$\mathbf{D} = \text{diag} [D_{\max}] \mathbf{B} \mathbf{A} \quad (2d)$$

$$\mathbf{B}^+ \text{diag} [D_{\max}^{-1}] \mathbf{D} = \mathbf{A} \quad (2e)$$

Here  $\alpha$  is the angle between the external magnetic field and the axes  $i \in \{x, y, z\}$  of the molecular frame,  $\beta$  is the angle between the internuclear spin vector and  $i$ ,  $D_{\max}$  is the maximum dipolar coupling of the respective spin pair and  $\mathbf{B}^+$  is the pseudo-inverse of  $\mathbf{B}$  obtained from singular value decomposition (SVD). Alternatively, as known from literature,<sup>[6]</sup> RDCs can be expressed in terms of dynamically averaged spherical harmonics according to

$$D = A_a D_{\max} \sqrt{\frac{4}{5\pi}} \left[ \left\langle Y_{2,0}^{(2)} \right\rangle + \sqrt{\frac{3}{8}} R \left( \left\langle Y_{2,2}^{(2)} \right\rangle + \left\langle Y_{2,2}^{*(2)} \right\rangle \right) \right] \quad (3)$$

where  $A_a$  is the axial component and  $R$  the rhombicity of the alignment tensor and  $Y_{2,m}$  are the second rank spherical harmonics of a spin pair. Eq. 3 can be expressed in any reference frame when utilizing the WIGNER rotation elements  $D_{M,M'}^{(2)}$ .

$$D = A_a D_{\max} \sqrt{\frac{4}{5\pi}} \left[ \sum_{M=-2}^2 \left\langle D_{M,0}^{(2)} \right\rangle \left\langle Y_{2,M}^{(2)} \right\rangle + \sqrt{\frac{3}{8}} R \left( \sum_{M=-2}^2 \left\langle D_{M,2}^{(2)} \right\rangle \left\langle Y_{2,M}^{(2)} \right\rangle + \sum_{M=-2}^2 \left\langle D_{M,-2}^{(2)} \right\rangle \left\langle Y_{2,M}^{(2)} \right\rangle \right) \right] \quad (4)$$

Combining the WIGNER elements and  $R$  in the matrix  $\mathbf{F}$  leads to the matrix equation:

$$\tilde{\mathbf{D}} = \mathbf{F}\mathbf{Y} \quad (5a)$$

$$\tilde{\mathbf{D}} = \text{diag} [A_a^{-1}] \mathbf{D}^t \text{diag} [D_{\max}^{-1}] \quad (5b)$$

TITANIA uses eq. (2d) and eq. (5a) in an iterative fashion to update the alignment, structure and dynamics information in an alternating fashion. The sequence for one iterative cycle is i) calculate the alignment information using eq. (2e), ii) calculate the spherical harmonics using eq. (6) and iii) convert the spherical harmonics to a new structure, which is used to restart this cycle. The structure information is calculated by

$$\langle \mathbf{Y}_{\text{ref}} \rangle = \mathbf{F}^+ \tilde{\mathbf{D}} \quad (6)$$

From these refined spherical harmonics the structure parameter can be extracted by maximizing  $Y_{2,0}$ , which essentially is a transformation to the respective vector frame (VF). In this frame the RDC vector is parallel to the z-axis.

$$\max \left( Y_{2,0}^{(2)} (\theta^{\text{VF}}, \phi^{\text{VF}}) \right) = \sum_{M=-2}^2 D_{M,0}^{(2)} (\phi_{\text{av}}^{\text{MF}}, \theta_{\text{av}}^{\text{MF}}, 0) \left\langle Y_{2,M}^{(2)} (\theta^{\text{MF}}, \phi^{\text{MF}}) \right\rangle \quad (7)$$

Additionally the dynamics parameters  $S_{\text{RDC}}^2$  (local order parameter),  $\eta$  (asymmetry parameter) and  $\varphi$  (direction of anisotropic motion) can be extracted:

$$S_{\text{RDC}}^2 = \frac{4\pi}{5} \sum_{m=-2}^2 \left\langle Y_{2,m}^{(2)} \right\rangle \left\langle Y_{2,m}^{*(2)} \right\rangle \quad (8)$$

$$\eta = \sqrt{\frac{\sum_{M=-2,2} \langle Y_{2,M}(\Theta^{\text{VF}}) \rangle \langle Y_{2,-M}(\Theta^{\text{VF}}) \rangle}{\sum_{M=-2}^2 \langle Y_{2,M}(\Theta^{\text{VF}}) \rangle \langle Y_{2,-M}(\Theta^{\text{VF}}) \rangle}} \quad (9)$$

$$\varphi = \frac{1}{2} \text{atan} \frac{\langle Y_{2,2}(\Theta^{\text{VF}}) \rangle - \langle Y_{2,-2}(\Theta^{\text{VF}}) \rangle}{i (\langle Y_{2,2}(\Theta^{\text{VF}}) \rangle + \langle Y_{2,-2}(\Theta^{\text{VF}}) \rangle)} \quad (10)$$

To properly extract these parameters (at least) five alignment media with sufficient degree of linear independence are essential.

$\mathbf{A}$  absorbs a part of internal structure dynamics due to the fact that it is calculated from a rigid model. As a consequence the absolute value of  $S_{\text{RDC}}^2$  does not reflect RDC motion without further scaling. PETI proposed different strategies to determine a meaningful scaling factor  $S_{\text{overall}}$ .<sup>[7]</sup> We decided to use the approach that is based on simple considerations of  $S_{\text{RDC}}^2$  limits and does not need any further investigations. The maximum value of  $S_{\text{RDC}}^2$  is 1.0 for an RDC that does not show any internal motion. Therefore  $S_{\text{overall}}$  can be obtained from the relation of the maximum value and the largest  $S_{\text{RDC,max}}^2$ :

$$S_{\text{overall}} = \sqrt{\frac{1}{S_{\text{RDC,max}}^2}} \quad (11)$$

## 1.2 Structure generation from spherical coordinates

Peng *et al.*<sup>[8]</sup> and Bakken *et al.*<sup>[9]</sup> proposed a solution to transform an overdetermined set of internal coordinates (bond lengths, bond angles, dihedral angles, ...) into a Cartesian base. This transformation uses the WILSON matrix  $\mathbf{B}_w$  with the individual elements:

$$B_w(q_i, x_j) = \frac{\partial q_i}{\partial x_j} \quad (12)$$

Assuming small changes in the internal coordinates the Cartesian displacement vector  $\Delta \vec{\mathbf{X}}$  can be formulated using the internal displacement vector  $\Delta \vec{\mathbf{q}}$ :

$$\Delta \vec{\mathbf{X}} = \mathbf{B}_w^+ \Delta \vec{\mathbf{q}} \quad (13)$$

The power of this algorithm is to combine the holonomic terms (like the bond lengths and angles) with experimental data (RDC vector orientations) or structural restraints (minimum distances or planarity). The downside of this algorithm is the iterative fashion and computationally demanding SVD, especially for large WILSON matrices. This iterative algorithm uses eq. (13) to update the Cartesian coordinates  $\vec{\mathbf{X}}$ .

In a previous publication<sup>[1]</sup> we introduced two types of damping terms added to the algorithm. The first one is an iteration-step-independent weighting factor  $w$ , that is applied to a specific type of internal coordinates (for example the bond lengths), to lower their contribution to the Cartesian displacement vector  $\Delta \vec{\mathbf{X}}$ . The second term is an iteration-step-dependent global damping factor, that is ramped up to 1.0 in a sigmoidal shape when reaching the maximum number of iteration steps:

$$D = \frac{\exp\left(3.5 \frac{iter}{maxiter}\right)^2}{\delta + \exp\left(3.5 \frac{iter}{maxiter}\right)^2} \frac{\delta + \exp(3.5)^2}{\exp(3.5)^2} \quad (14)$$

Here  $\delta$  is the user defined damping constant that determines the growth of the damping factor  $D$  and 3.5 is an empirically determined factor, that allows a smooth manipulation of the sigmoidal shape.

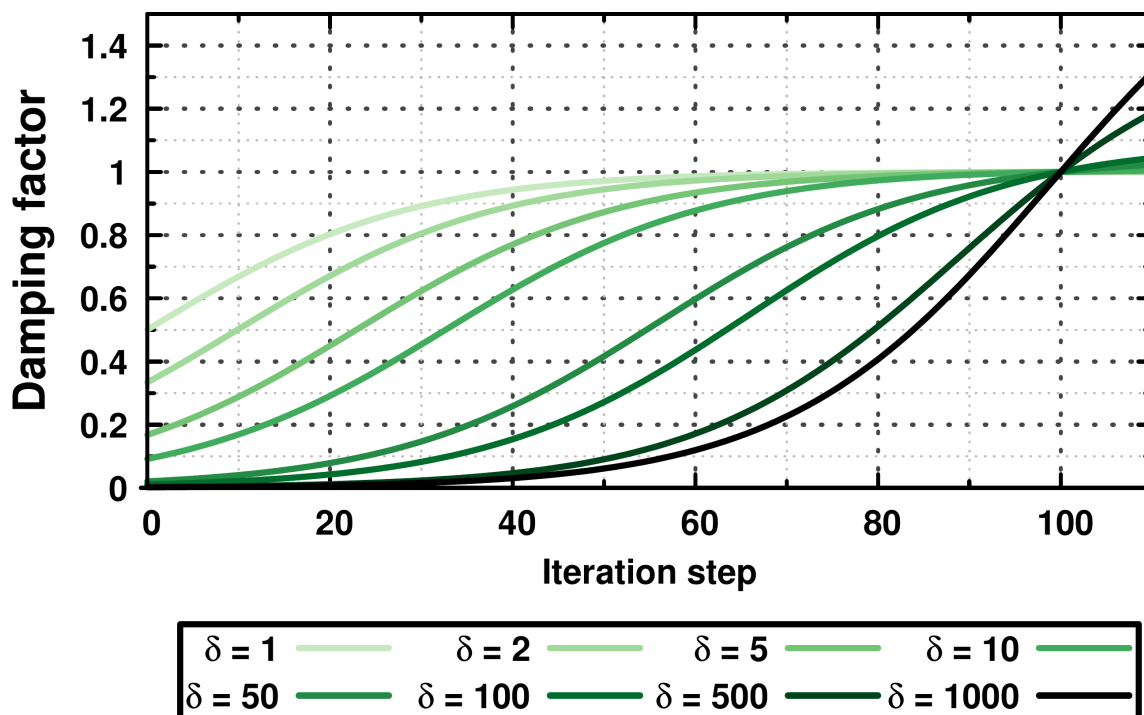


Figure 1: Damping factors  $D$  for a set of damping constants  $\delta$  as shown in the TITANIA concept paper. For the calculations a maximum of 100 iterations was assumed. Reproduced from Roth *et al.*<sup>[1]</sup>

This default algorithm is accompanied by a second protocol in TITANIA: the vector addition algorithm, which generally is less flexible but more robust. The description and implementation of this algorithm is described elsewhere<sup>[1]</sup> since it is not used in the context of this communication.

### MMFF94 and holonomic terms

TITANIA implements the MMFF94<sup>[10–14]</sup> force field to use the equilibrium constants as holonomic terms in the redundant internal coordinates algorithm. Additionally it is used to update atoms not defined by RDCs (e.g. methyl protons) when no redundant internal coordinates are used for the structure generation. The third usage is activated if the user submits a planar structure to TITANIA, where an MMFF94 minimization is used to generate a random starting geometry.

The underlying algorithm aligns a planar input molecule in an arbitrary plane of the Cartesian reference frame (e.g. the x,y plane). The perpendicular coordinates (the z coordinates in the x,y plane example) are randomized and a short MMFF94 force-field optimization is performed to apply holonomic restraints and obtain a chemically meaningful, non-planar geometry. Due to the randomized positioning of the respective coordinates perpendicular to the initial molecular plane, the subsequent optimization leads to a random configuration for the stereogenic center from which the TITANIA optimization is started. By this simple method, restarting calculations from planar geometries serves as a method to validate runs. A stable result should converge to the same relative configuration irrespective of the (randomized) starting coordinates.

## 2 Handling undefined RDCs

Ideally the matrix  $\mathbf{D}$  contains the same number of RDCs for each alignment medium. This is, however, hardly ever possible with real experimental data. Thus these missing data have to be dealt with. At first it seems intuitive to define missing RDCs in  $\mathbf{D}$  as zero, since basic linear algebra cannot handle non-numeric elements. From the basic definition of RDCs (eq. (1)) it follows, that 0.0 Hz would carry false information for the direction  $\Theta = \arccos \sqrt{\frac{1}{3}}$ . Alternatively the RDC would be 0.0 Hz due to isotropic tumbling resulting in an order parameter  $S_{\text{RDC}}^2 = 0$ . Therefore undefined RDCs have to be handled in another way.

A simple/straightforward algorithm is made accessible by the iterative nature of the TITANIA optimization. As described in previous publication,<sup>[1]</sup> TITANIA updates the  $D_{\text{max}}$  values at the beginning of every iteration, to properly consider structural changes. At the same time, undefined RDCs are back-calculated using eq. (2d) (if the user defined the flag `recalculateRDCs=1`) and the respective RDC values are updated in the  $\mathbf{D}$  matrix. To distinguish back-calculated RDCs, that are applied in the structure validation, from the back-calculated RDCs mentioned above, we refer to the latter ones as *recalculated* RDCs since TITANIA updates them in every step. This update of missing RDCs is performed in every iteration step prior to the calculation of the alignment tensors.

From the updated  $\mathbf{D}$  matrix the alignment tensors  $\mathbf{A}$  and refined spherical harmonics  $\mathbf{Y}$  are calculated. This means that the RDCs are always biased by the structure model of the previous step. This may help guide the optimization or – if the structure model of the previous step has errors, e.g. is present in a wrong relative configuration – this bias may lead to instabilities in the optimization. In the following this algorithm will be referred to as *recalculation scheme*.

A more complex algorithm, hereafter referred to as *weighting scheme*, is independent of the previous structure. This algorithm introduces a pseudo-dimension in eq. (2d) and (5a) by splitting the RDC matrices in vectors containing individual RDC sets (to determine  $\mathbf{A}$ ) and RDC vectors (to determine  $\mathbf{Y}_{\text{ref}}$ ) respectively. The graphical illustration of this idea is shown in fig. 2.

By this, the second dimension of  $\mathbf{A}$  and  $\mathbf{Y}_{\text{ref}}$  are reduced from  $M$  and  $N$  respectively to 1, leading to the definition of individual vector equations:

$$\vec{\mathbf{A}}_{\mathbf{m}} = (\text{diag}[w_i]\mathbf{B})^+ \text{diag}[w_i]\vec{\mathbf{D}}_{\mathbf{m}} \quad (15a)$$

$$\vec{\mathbf{Y}}_{\mathbf{n}} = (\text{diag}[w_j]\mathbf{F})^+ \text{diag}[w_j]\vec{\mathbf{D}}_{\mathbf{n}} \quad (15b)$$

In eq. (15a) all RDCs defined in medium  $m$  are used to calculate the corresponding alignment tensor. Additionally a diagonal matrix with weights  $w_i$  of the individual RDC vectors  $i$  are introduced. These weights allow to remove missing RDCs from the system of equations, by setting them to 0, without changing the shape of  $\mathbf{D}$  and  $\mathbf{B}$  for the individual media. The same is true for eq. (15b), where all RDCs of one spin pair  $n$  are used to calculate the spherical harmonics for the respective pair, utilizing the corresponding weighting factors  $w_j$  of the individual media  $j$ .

A problem of the weighting scheme can arise if the rank of  $\mathbf{B}$  or  $\mathbf{F}$  drops below 5 by removing rows. This is rather unlikely for  $\mathbf{B}$  since the number of RDCs per set is rather large to allow the application of TITANIA in this context. For  $\mathbf{F}$  this reduction can be a realistic scenario, since finding/measuring new alignment conditions is one of the challenging aspects for the successful applications of the MFA. In such a case the MOORE-PENROSE inverse of  $\mathbf{F}$  will produce equally good refined spherical harmonics  $\vec{\mathbf{Y}}$  as solution of the implemented formula (black part) for any vector  $\vec{\mathbf{Z}}$  of dimension 5 added to it (grey part):<sup>[15,16]</sup>

$$\vec{\mathbf{Y}}_{\mathbf{n}} = \mathbf{F}^+ \vec{\mathbf{D}}_{\mathbf{n}} + (\mathbf{I} - \mathbf{F}^+ \mathbf{F}) \vec{\mathbf{Z}} \quad (16)$$

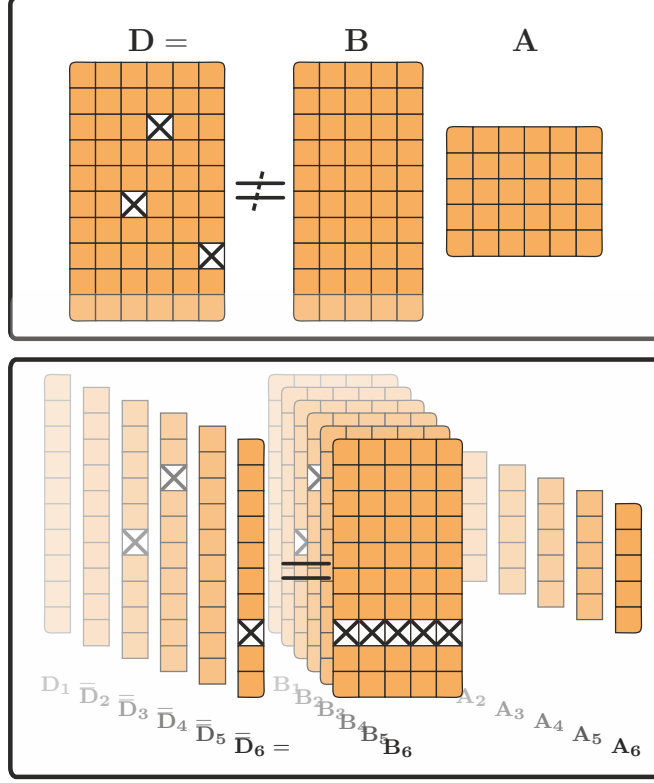


Figure 2: Shown are graphical representations of matrix eq. (2d). The orange filled parts of the matrices represent defined elements. The white parts marked with a  $\times$  are undefined RDCs. In the upper panel the equation is not correct for a matrix multiplication without further adjustments. The lower panel uses the fact, that a matrix equation can be separated in multiple equations of lower dimension. These individual equations form an additional pseudo-dimension, illustrated by shift and fading. The individual matrices  $\mathbf{B}$  and  $\mathbf{A}$  are obtained by weighting as described in eq. (15a). Equivalently this can be transferred to eq. (15b).

A result is, that a part of  $\vec{\mathbf{Y}}$  is constructed randomly from the nullspace of  $\mathbf{F}$  if its rank  $R(\mathbf{F}) < 5$  (grey part of eq. (16)). Accordingly the solution will be unique if and only if  $\mathbf{F}$  is full rank ( $R(\mathbf{F}) = 5$ ). This behavior has to be taken into account when undefined RDCs are clustering on a specific vector. Thus the user should consider to remove an RDC vector from the analysis, if it is defined in too few independent media to reduce the probability of reaching a rank below 5.

Both, the weighting and recalculation scheme, are used to interpret the experimental data to show the applicability in the TITANIA context.

The weighting scheme enables, additionally to the discussed application, the incorporation of experimental uncertainties in the TITANIA scheme. This can be done by weighting based on the user defined RDC errors and the following equation:

$$w = w_{\text{user}} / \Delta D \quad (17)$$

where  $w_{\text{user}}$  is the weighting factor and  $\Delta D$  the experimental error, which are both defined by the user in the RDC input. Both inputs are only considered in the weighting scheme and additionally the error weighting (`errorWeightInSVD=1`) has to be activated by the user. Otherwise errors are just used to generate the RDC matrices in the Monte-Carlo bootstrap. This additional weighting is not used in the context of this paper to facilitate the comparison of the two schemes.

### 3 Sample preparation

All samples were prepared in 5 mm NMR tubes. The individual preparation protocols were performed as reported in literature (for references see table 1). To check the spatial homogeneity of the samples slice selective  $^2\text{H}$ -NMR spectra were used.<sup>[17]</sup>

Table 1: Summary of the samples used for the datasets. The choice of the enantiomers of IPC does not influence the determination of the relative configuration. Set 01 and 02 were measured on the same thermoresponsive sample at different temperatures.

Set	Polymer	Analyte	$w(\text{Polymer})^{\text{a)}$	Solvent	Literature
#01	<i>co</i> -PPLA-PBLA (305 K)	(−)-IPC	12.5%	TCE- $d_2$	[18]
#02	<i>co</i> -PPLA-PBLA (309 K)	(−)-IPC	12.5%	TCE- $d_2$	[18]
#03	PBLG	(+)-IPC	7.9%	TCE- $d_2$	[19]
#04	PBDG	(+)-IPC	14.0%	DCM- $d_2$	[19]
#05	PBPMLG-C2	(+)-IPC	16.5%	$\text{CDCl}_3$	unpublished
#06	PBPMLG- $d_{1\alpha}$	(+)-IPC	16.5%	$\text{CDCl}_3$	[20]

a)  $w(\text{Polymer}) = \frac{w(\text{Polymer})}{w(\text{Polymer}) + w(\text{Analyte}) + w(\text{Solvent})}$

All polymers used belong to the group of (homo)-poly amino acids (aspartates and glutamates). The choice of using in-house polymers only is to demonstrate the applicability of TITANIA to small organic compounds without exhausting the toolbox of NMR spectroscopists. Nevertheless we plan to extend this approach using other alignment media to achieve superior complementarity of the RDC sets. This is especially a goal when investigating compounds with higher flexibility.

## 4 NMR-experiment conditions

All isotropic and anisotropic IPC spectra were recorded on a 700 MHz spectrometer (Bruker AVANCE III HD, equipped with a QCI cryo probe ( $^1\text{H}/^{19}\text{F}$ -31P/ $^{13}\text{C}/^{15}\text{N}/^2\text{H}$ ) with z-gradient). The individual solvents and additives used are summarized in table 1. The total coupling constants ( $^1T$ ) and scalar coupling constants were extracted using the CLIP-HSQC<sup>[21]</sup> and perfect CLIP-HSQC<sup>[22]</sup> experiment (INEPT delay of 145 Hz) with a total of 8k data points in the direct and 256 data points in the indirect dimension (NS = 4, DS = 32). The spectra were processed using the `qsine` window function. A zero filling to 1k in the indirect dimension was applied. For each coupling the corresponding trace (row) of the CLIP-HSQC spectrum was extracted.

To reduce the impact of measurement errors a protocol similar to the one conducted in literature<sup>[23]</sup> was used. In this protocol a trace of the 2D spectrum is duplicated to extract four values for one coupling constant. The left and right flank of the corresponding multiplet parts are aligned to extract two values (see fig. 3 a and b). The third value is extracted by minimizing the difference between the respective parts of the doublet. The last value is simply obtained by peak picking. The RDC is obtained by averaging these values, and the corresponding standard deviation is used to estimate the experimental error. The minimum experimental error was set to the spectral resolution after processing.

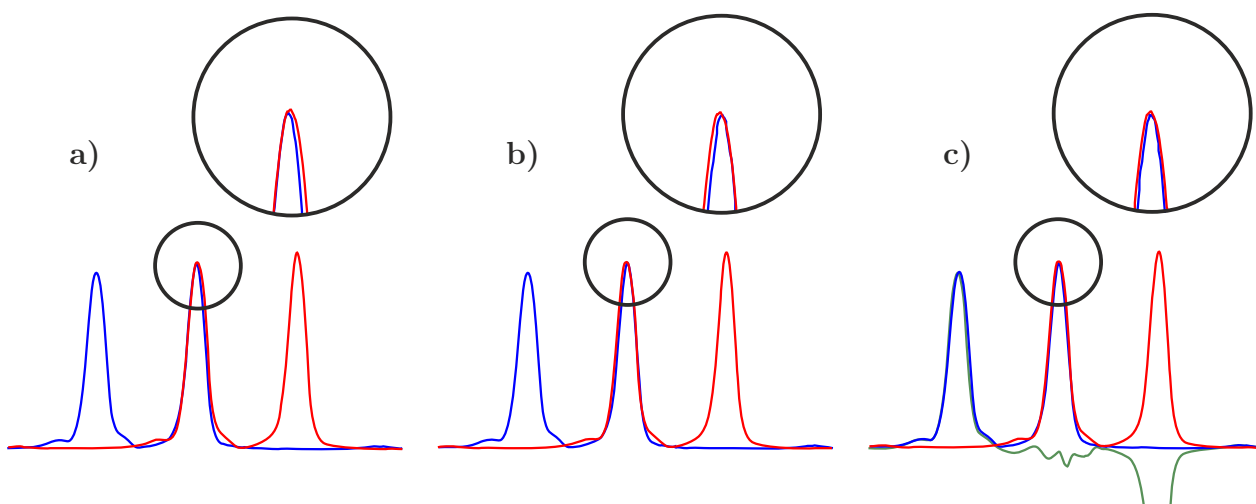


Figure 3: Representation of the protocol used for RDC extraction and estimation of the experimental uncertainty by using one trace from a 2D spectrum. Subplot a) aligns the left flank of the multiplets, b) the right flank and c) minimizes the difference (green) between the two doublet sides. Additionally a peak picking, which is not shown here, was performed to obtain a fourth measure.

For set 04 not all couplings were available from the CLIP-HSQC spectrum. Therefore an F1-coupled HSQC<sup>[24]</sup> spectrum was recorded with a total of 1398 data points in the direct and 4k in the indirect dimension (NS = 4, DS = 32). The spectra were processed using the `qsine` window function. A zero filling to 4k in the direct and 8k in the indirect dimension was applied. The couplings were extracted as described above, but instead of extracting traces from rows the respective columns were used.



The total coupling constants ( ${}^nT$ ) and scalar coupling constants ( ${}^nJ$ ) of HH pairs were extracted from TSE-PSYCHEDELIC<sup>[25]</sup> spectra. In table 2 the parameters are summarized. The spectra were processed using the `qsine` window function. A zero filling to 32k in the direct and 512 in the indirect dimension was applied. The couplings were extracted as described above using the combination of multiple traces from the columns.

Table 2: Experimental settings for the TSE-PSYCHEDELIC spectra of the individual datasets used in the optimization.

<b>Parameters</b>	<b>set 01</b>	<b>set 02</b>	<b>set 03</b>	<b>set 04</b>	<b>set 05</b>	<b>set 06</b>
points (F2/F1)	64k / 256	64k / 256	64k / 256	16k / 128	16k / 128	64k / 128
NS / DS	4 / 16	4 / 16	4 / 16	4 / 8	4 / 32	4 / 32

## 5 List of RDCs

All RDCs obtained are summarized in table 3.

Table 3: All RDCs extracted from the experiments described above (section 4) are summarized with the corresponding uncertainties. Missing RDCs are marked by a "-".

coupling pair	${}^nD_{\text{XH}}^{\text{a)}} / \text{Hz}$					
	set 01	set 02	set 03	set 04	set 05	set 06
C1-H1	-7.89( 68)	14.53( 68)	15.01( 73)	25.60( 81)	-5.78(124)	-0.55( 34)
C2-H2	4.59( 68)	-7.96(173)	0.09( 68)	-12.21( 48)	-6.61(139)	-8.76( 37)
C3-H3	12.12( 68)	-21.07( 68)	16.37( 68)	14.38( 34)	-29.50( 71)	-16.51( 34)
C4-H4a	2.41( 68)	-4.08( 68)	-3.63( 68)	-2.40( 83)	6.48( 25)	2.65( 34)
C4-H4s	2.95( 68)	-5.34( 68)	17.62( 68)	15.24( 76)	-28.64(101)	-17.84( 34)
C5-H5	-2.34( 68)	4.10( 68)	-7.29( 68)	-15.32( 34)	-12.80( 31)	-5.80( 34)
C7-H7a	-1.29( 68)	5.65( 68)	-17.76( 68)	-24.49( 34)	35.63( 95)	16.58( 36)
C7-H7s	1.04( 68)	-3.56( 68)	-1.40( 68)	4.08(121)	14.03(103)	9.29( 34)
C6-C8	-0.52( 68)	0.74( 68)	-1.52( 68)	-0.60( 32)	4.06( 27)	3.82( 34)
C6-C9	1.05( 68)	-1.92( 68)	1.35( 68)	1.26( 84)	-2.46( 14)	-1.45( 34)
C2-C10	0.49( 68)	-1.14( 68)	0.17( 68)	0.40( 43)	-1.85( 14)	-0.93( 34)
H1-H2	-2.56( 4)	5.05( 7)	-6.56( 7)	-5.11( 14)	-	9.77( 11)
H1-H3	0.00( 4)	0.52( 4)	-	2.08( 50)	0.00( 50)	-
H1-H4a	-0.46( 4)	0.93( 4)	-	0.00( 50)	-	-
H1-H4s	0.00( 4)	0.67( 4)	-	1.64( 4)	0.00( 50)	-
H1-H5	-0.52( 4)	0.85( 4)	0.56( 4)	-	-	-0.87( 15)
H1-H7a	-0.26( 4)	1.31( 4)	0.26( 4)	-0.26( 50)	-	0.28( 10)
H1-H7s	1.10( 47)	-2.94( 5)	0.45( 4)	0.56( 4)	-7.42( 9)	-3.14( 10)
H2-H3	0.92( 18)	-1.88( 7)	4.66( 4)	4.37( 10)	-6.83( 4)	-3.54( 40)
H2-H4a	-0.83( 4)	1.60( 4)	1.11( 4)	0.77( 50)	-	-1.40( 10)
H2-H4s	-0.07( 51)	0.18( 9)	0.72( 6)	1.84( 50)	-2.05( 6)	-1.27( 15)
H2-H7a	1.64( 6)	-4.07( 9)	-1.78( 4)	-3.20(200)	-	-2.90( 10)
H2-H7s	0.00( 4)	0.00( 10)	-1.59( 4)	-2.15( 5)	2.19( 50)	1.43( 10)
H3-H4a	1.43( 46)	-3.23( 9)	0.92( 4)	0.45( 50)	-5.68( 5)	-2.62( 49)
H3-H4s	-0.68( 4)	1.16( 4)	-2.41( 4)	-4.66( 50)	-7.43( 5)	-3.94(144)
H3-H5	0.00( 4)	-0.59( 10)	-	0.00( 50)	0.00( 50)	-
H3-H7a	0.48( 96)	-1.94( 4)	1.48( 4)	1.47( 50)	-2.50( 16)	-1.48( 10)
H3-H7s	0.00( 4)	-0.68( 10)	-	0.00( 50)	0.00( 50)	-
H4a-H4s	8.03( 20)	-14.97( 11)	5.49( 7)	-2.06( 7)	-12.80( 12)	-9.93( 10)
H4a-H5	-2.08( 4)	4.28( 5)	-7.63( 5)	-8.14( 7)	-	9.38( 10)
H4a-H7a	-1.05( 7)	1.59( 9)	6.80( 4)	10.23( 11)	-	-2.65( 10)
H4a-H7s	-0.19( 39)	1.19( 4)	0.91( 4)	2.87( 8)	1.93( 11)	1.44( 10)
H4s-H5	0.81( 27)	-1.79( 6)	0.48( 6)	2.26(100)	4.22(100)	2.74( 10)
H4s-H7a	-0.26( 52)	-0.86( 4)	2.40( 4)	-	-	-1.78( 10)
H4s-H7s	0.07( 4)	-0.25( 19)	1.33( 6)	2.07( 8)	-2.56( 50)	-0.89( 53)
H5-H7a	0.00( 4)	0.99( 4)	-0.70( 4)	-	-	-1.29( 10)
H5-H7s	-0.39( 23)	0.94( 6)	6.40( 5)	6.82( 4)	-9.05(100)	-4.89( 10)
H7a-H7s	-5.29(170)	12.34( 5)	-19.52( 7)	-19.15(200)	50.85(103)	26.89( 10)

a)  $n = 1$  for  $X = \text{C}$  and  $n \geq 2$  for  $X = \text{H}$ . Note that  ${}^1D_{\text{CH}}$  couplings of methyl groups were converted to the respective  ${}^1D_{\text{CC}}$  RDCs.<sup>[26]</sup>

## 6 Orientations of IPC

The orientations used for the setups, reported in the main text, are summarized in table 4. All parameters were calculated using the in-house RDC module (**RDC@hotFCHT**) of the **hotFCHT** software.<sup>[27]</sup> The structure model used to determine the orientations of IPC was taken from literature.<sup>[19]</sup>

Table 4: Orientations of the individual RDC sets used for the setups in the main text.

<b>Alignment parameters</b>	<b>set 01</b>	<b>set 02</b>	<b>set 03</b>	<b>set 04</b>	<b>set 05</b>	<b>set 06</b>
$A_a$	-5.275e-04	1.033e-03	1.139e-03	1.564e-03	-2.384e-03	-1.270e-03
$R$	2.652e-01	2.494e-01	5.002e-01	4.737e-01	8.722e-02	1.839e-01
RMSD	3.288e-01	4.156e-01	5.418e-01	6.342e-01	8.255e-01	5.642e-01
Q-Factor	1.131e-01	7.601e-02	7.258e-02	7.085e-02	5.406e-02	6.910e-02
Cond. number <sup>a</sup>	1.468	1.468	1.630	1.605	1.753	1.630
$\alpha/^\circ$	34.54	23.86	99.41	113.39	82.00	86.23
$\beta/^\circ$	38.90	40.66	73.13	64.28	84.27	88.39
$\gamma/^\circ$	116.71	120.01	7.33	1.02	23.53	24.19

<sup>a)</sup> Condition number of the normalized cosine matrix **B**.

## 7 RDC setups

The full RDC matrix and the corresponding experimental uncertainties are summarized in table 3. To investigate the impact of undefined elements we adjusted the set sizes (removed rows from the matrix). The first setup (**A-6**) is not reduced in size and contains 23 undefined RDCs, which show clustering on proton H1 (11 undefined RDCs). The RDCs of H1 show two RDC-vectors which are undefined in three sets ( ${}^4D_{\text{H1H3}}$ ,  ${}^4D_{\text{H1H5}}$  and  ${}^5D_{\text{H1H4s}}$ ) and even one RDC pair which is undefined in three media ( ${}^5D_{\text{H1H4a}}$ ).

We used two of the measured media as a template to reduce the number of undefined RDCs in the matrix. For setup **B-6** these two media were **03** and **06**. Since both media had the same undefined RDCs they were the obvious choice to remove the maximum number of rows that have two or more undefined RDCs (see tables on the following pages). This leads to a matrix with 12 undefined RDCs, with H7a showing seven undefined RDCs (twice undefined in two media:  ${}^4D_{\text{H7aH4s}}$ ,  ${}^3D_{\text{H7aH5}}$ ).

The second template used for setup **C-6** was set **05**, since it had the highest number of undefined RDCs of all media. As a result only eight RDCs are undefined, but all of them are undefined in two media. Three of these four RDC-vectors involve H3. Additionally five out of the seven possible H7a RDCs were removed from the set.

In the last and most obvious setup **D-6** we removed all RDC-vectors that were undefined in any of the media, leaving 13 long range  ${}^nD_{\text{HH}}$  RDCs for the analysis. This drastically reduced the number of RDCs of H1 (only  ${}^3D_{\text{H1H7s}}$ ) and H7a (only  ${}^2D_{\text{H7aH7s}}$  and  ${}^5D_{\text{H7aH3}}$ ). All results are summarized in the following sections. The tables on the following pages show the setups generated, where orange cells (with an **x**) represent available RDCs used for the analysis. White cells (with a **-**) are undefined RDCs kept in the analysis. Grey rows (with a **-**) were removed from the data sets.

Table 5: The RDCs used in the setup **A-6** are shown below. An x shows that the RDC is available and a - that the RDC is undefined (white cell). This setup has 23 undefined RDCs.

RDC	set 01	set 02	set 03	set 04	set 05	set 06	index
C1-H1	x	x	x	x	x	x	1
C2-H2	x	x	x	x	x	x	2
C3-H3	x	x	x	x	x	x	3
C4-H4a	x	x	x	x	x	x	4
C4-H4s	x	x	x	x	x	x	5
C5-H5	x	x	x	x	x	x	6
C7-H7a	x	x	x	x	x	x	7
C7-H7s	x	x	x	x	x	x	8
C6-C8	x	x	x	x	x	x	9
C6-C9	x	x	x	x	x	x	10
C2-C10	x	x	x	x	x	x	11
H1-H2	x	x	x	x	-	x	12
H1-H3	x	x	-	x	x	-	13
H1-H4a	x	x	-	x	-	-	14
H1-H4s	x	x	-	x	x	-	15
H1-H5	x	x	x	-	-	x	16
H1-H7a	x	x	x	x	-	x	17
H1-H7s	x	x	x	x	x	x	18
H2-H3	x	x	x	x	x	x	19
H2-H4a	x	x	x	x	-	x	20
H2-H4s	x	x	x	x	x	x	21
H2-H7a	x	x	x	x	-	x	22
H2-H7s	x	x	x	x	x	x	23
H3-H4a	x	x	x	x	x	x	24
H3-H4s	x	x	x	x	x	x	25
H3-H5	x	x	-	x	x	-	26
H3-H7a	x	x	x	x	x	x	27
H3-H7s	x	x	-	x	x	-	28
H4a-H4s	x	x	x	x	x	x	29
H4a-H5	x	x	x	x	-	x	30
H4a-H7a	x	x	x	x	-	x	31
H4a-H7s	x	x	x	x	x	x	32
H4s-H5	x	x	x	x	x	x	33
H4s-H7a	x	x	x	-	-	x	34
H4s-H7s	x	x	x	x	x	x	35
H5-H7a	x	x	x	-	-	x	36
H5-H7s	x	x	x	x	x	x	37
H7a-H7s	x	x	x	x	x	x	38

Table 6: The RDCs used in the setup **B-6** are shown below. An x shows that the RDC is available and a - that the RDC is undefined (white cell) or removed (grey row) based on the undefined RDCs of **set 03/06**. This setup has 12 undefined RDCs.

RDC	set 01	set 02	set 03	set 04	set 05	set 06	index
C1-H1	x	x	x	x	x	x	1
C2-H2	x	x	x	x	x	x	2
C3-H3	x	x	x	x	x	x	3
C4-H4a	x	x	x	x	x	x	4
C4-H4s	x	x	x	x	x	x	5
C5-H5	x	x	x	x	x	x	6
C7-H7a	x	x	x	x	x	x	7
C7-H7s	x	x	x	x	x	x	8
C6-C8	x	x	x	x	x	x	9
C6-C9	x	x	x	x	x	x	10
C2-C10	x	x	x	x	x	x	11
H1-H2	x	x	x	x	-	x	12
H1-H3	-	-	-	-	-	-	-
H1-H4a	-	-	-	-	-	-	-
H1-H4s	-	-	-	-	-	-	-
H1-H5	x	x	x	-	-	x	13
H1-H7a	x	x	x	x	-	x	14
H1-H7s	x	x	x	x	x	x	15
H2-H3	x	x	x	x	x	x	16
H2-H4a	x	x	x	x	-	x	17
H2-H4s	x	x	x	x	x	x	18
H2-H7a	x	x	x	x	-	x	19
H2-H7s	x	x	x	x	x	x	20
H3-H4a	x	x	x	x	x	x	21
H3-H4s	x	x	x	x	x	x	22
H3-H5	-	-	-	-	-	-	-
H3-H7a	x	x	x	x	x	x	23
H3-H7s	-	-	-	-	-	-	-
H4a-H4s	x	x	x	x	x	x	24
H4a-H5	x	x	x	x	-	x	25
H4a-H7a	x	x	x	x	-	x	26
H4a-H7s	x	x	x	x	x	x	27
H4s-H5	x	x	x	x	x	x	28
H4s-H7a	x	x	x	-	-	x	29
H4s-H7s	x	x	x	x	x	x	30
H5-H7a	x	x	x	-	-	x	31
H5-H7s	x	x	x	x	x	x	32
H7a-H7s	x	x	x	x	x	x	33

Table 7: The RDCs used in the setup **C-6** are shown below. An x shows that the RDC is available and a - that the RDC is undefined (white cell) or removed (grey row) based on the undefined RDCs of **set 05**. This setup has 8 undefined RDCs.

RDC	set 01	set 02	set 03	set 04	set 05	set 06	index
C1-H1	x	x	x	x	x	x	1
C2-H2	x	x	x	x	x	x	2
C3-H3	x	x	x	x	x	x	3
C4-H4a	x	x	x	x	x	x	4
C4-H4s	x	x	x	x	x	x	5
C5-H5	x	x	x	x	x	x	6
C7-H7a	x	x	x	x	x	x	7
C7-H7s	x	x	x	x	x	x	8
C6-C8	x	x	x	x	x	x	9
C6-C9	x	x	x	x	x	x	10
C2-C10	x	x	x	x	x	x	11
H1-H2	-	-	-	-	-	-	-
H1-H3	x	x	-	x	x	-	12
H1-H4a	-	-	-	-	-	-	-
H1-H4s	x	x	-	x	x	-	13
H1-H5	-	-	-	-	-	-	-
H1-H7a	-	-	-	-	-	-	-
H1-H7s	x	x	x	x	x	x	14
H2-H3	x	x	x	x	x	x	15
H2-H4a	-	-	-	-	-	-	-
H2-H4s	x	x	x	x	x	x	16
H2-H7a	-	-	-	-	-	-	-
H2-H7s	x	x	x	x	x	x	17
H3-H4a	x	x	x	x	x	x	18
H3-H4s	x	x	x	x	x	x	19
H3-H5	x	x	-	x	x	-	20
H3-H7a	x	x	x	x	x	x	21
H3-H7s	x	x	-	x	x	-	22
H4a-H4s	x	x	x	x	x	x	23
H4a-H5	-	-	-	-	-	-	-
H4a-H7a	-	-	-	-	-	-	-
H4a-H7s	x	x	x	x	x	x	24
H4s-H5	x	x	x	x	x	x	25
H4s-H7a	-	-	-	-	-	-	-
H4s-H7s	x	x	x	x	x	x	26
H5-H7a	-	-	-	-	-	-	-
H5-H7s	x	x	x	x	x	x	27
H7a-H7s	x	x	x	x	x	x	28

Table 8: The RDCs used in the setup **D-6** are shown below. An x shows that the RDC is available and a - that the RDC is removed (grey row). This setup contains no undefined RDCs.

RDC	set 01	set 02	set 03	set 04	set 05	set 06	index
C1-H1	x	x	x	x	x	x	1
C2-H2	x	x	x	x	x	x	2
C3-H3	x	x	x	x	x	x	3
C4-H4a	x	x	x	x	x	x	4
C4-H4s	x	x	x	x	x	x	5
C5-H5	x	x	x	x	x	x	6
C7-H7a	x	x	x	x	x	x	7
C7-H7s	x	x	x	x	x	x	8
C6-C8	x	x	x	x	x	x	9
C6-C9	x	x	x	x	x	x	10
C2-C10	x	x	x	x	x	x	11
H1-H2	-	-	-	-	-	-	-
H1-H3	-	-	-	-	-	-	-
H1-H4a	-	-	-	-	-	-	-
H1-H4s	-	-	-	-	-	-	-
H1-H5	-	-	-	-	-	-	-
H1-H7a	-	-	-	-	-	-	-
H1-H7s	x	x	x	x	x	x	12
H2-H3	x	x	x	x	x	x	13
H2-H4a	-	-	-	-	-	-	-
H2-H4s	x	x	x	x	x	x	14
H2-H7a	-	-	-	-	-	-	-
H2-H7s	x	x	x	x	x	x	15
H3-H4a	x	x	x	x	x	x	16
H3-H4s	x	x	x	x	x	x	17
H3-H5	-	-	-	-	-	-	-
H3-H7a	x	x	x	x	x	x	18
H3-H7s	-	-	-	-	-	-	-
H4a-H4s	x	x	x	x	x	x	19
H4a-H5	-	-	-	-	-	-	-
H4a-H7a	-	-	-	-	-	-	-
H4a-H7s	x	x	x	x	x	x	20
H4s-H5	x	x	x	x	x	x	21
H4s-H7a	-	-	-	-	-	-	-
H4s-H7s	x	x	x	x	x	x	22
H5-H7a	-	-	-	-	-	-	-
H5-H7s	x	x	x	x	x	x	23
H7a-H7s	x	x	x	x	x	x	24



## 8 SECONDA

SECONDA is a method to analyse RDC data matrices reported by HUS and BRÜSCHWEILER.<sup>[28]</sup> It utilizes the weighted ( $w$ ) covariance matrix  $\mathbf{Cov}$  for a principle component analysis (PCA). The results are the principle variances (eigenvalues  $\lambda$  of  $\mathbf{Cov}$ ) and eigenmodes (eigenvectors  $|q\rangle$  of  $\mathbf{Cov}$ ) which are primarily used to gain information on the homo- and heterogeneity of the RDC data. In this interpretation the principle variances contain information on the amplitude of differences in the alignment conditions. The corresponding eigenmodes carry information on the localization of these amplitudes.

Our main goal is to find and interpret the rank of the RDC matrix to gain information if the experimental data is of sufficient quality (also in terms of independence of alignments induced) to allow for a structure optimization in TITANIA.

$$Cov_{ij} = \frac{1}{M-1} \sum_{m \notin (\mathbf{I} \cup \mathbf{J})}^M w_m (D_i^m - \langle D_i \rangle) (D_j^m - \langle D_j \rangle) \quad (18a)$$

$$w = \frac{1}{\frac{1}{N-1} \sum_n^N (D_n^m - \langle D^m \rangle)^2} \quad (18b)$$

BRÜSCHWEILER proposed to use the collectivity  $\kappa$ ,<sup>[29]</sup> which reduces the eigenmodes to a single value to judge whether the amplitude appears in a global or localized fashion. The range of  $\kappa$  is  $\frac{100}{N}\%$  to 100%, where  $N$  is the number of RDC vectors. This allows for the representation of the eigenvalue and eigenvector pairs in two scalars  $\lambda$  and  $\kappa$ . Typically these values are used to generate an  $\kappa = f(\lambda)$  plot. When discussing the eigenvalues  $\lambda_i$  they are ordered in descending order.

$$\kappa_q = \frac{1}{N} \exp \left[ - \sum_i^N ||q\rangle_i|^2 \log ||q\rangle_i|^2 \right] \cdot 100\% \quad (19)$$

As reported in literature<sup>[30]</sup> SECONDA is sensitive to undefined RDCs. HUS handled undefined RDCs in an RDC matrix of 58 ubiquitin residues in 10 media in three approaches. The easiest was to remove every residue from the analysis, that is undefined in any of the 10 media, leaving 32 residues. Alternatively, the media with several undefined RDCs were removed, followed by removing the residues that were undefined in the remaining 8 media. By this 41 residues were analyzed for 8 media. The last approach used a rigid structure model, which was used to back-calculate the missing RDCs.

Here we propose an additional method to handle missing RDCs. The weighting factors  $w$  and the covariances  $Cov$  are adapted to neglect RDC vectors for the media in which they are undefined.

$$Cov_{ij} = \frac{1}{M-1-|\mathbf{I} \cup \mathbf{J}|} \sum_{m \notin (\mathbf{I} \cup \mathbf{J})}^M w_m (D_i^m - \langle D_i \rangle) (D_j^m - \langle D_j \rangle) \quad (20a)$$

$$w = \frac{1}{\frac{1}{N-K-1} \sum_{n \neq k}^N (D_N^m - \langle D^m \rangle)^2} \quad (20b)$$

where  $N$  is the number of RDCs (including undefined ones),  $M$  is the number of alignment media,  $K$  is the number of undefined RDCs (with the index  $k$ ) in the medium  $m$ . The mean RDC  $\langle D^m \rangle$  treats the undefined RDCs by removing them from the sum and the denominator ( $N-K$ ) in

the same manner.  $\mathbf{I}$  is the (mathematical) set of media, in which the RDC  $i$  is undefined,  $\mathbf{I} \cup \mathbf{J}$  is the union of the respective (mathematical) sets for  $i$  and/or  $j$ ,  $|\mathbf{X}|$  is the cardinality (= number of members) in  $\mathbf{X}$  and  $\langle D_i \rangle$  is the mean RDC of the vector  $i$  in all media (neglecting undefined RDCs). Using eq. (18a) and (18b) allows to construct a weighted covariance matrix, that neglects undefined RDCs in a proper manner. This matrix will be referred to as *reduced* covariance matrix. It is noteworthy that in the sum of the covariance matrix elements the whole medium  $m$  has to be dropped if one RDC  $i$  or  $j$  is undefined.

This approach is compared to the standard SECONDA analysis, either keeping RDCs as 0.0 in the data matrix (referred to as *sparse*) or back-calculating them using the correct structure model (referred to as *filled*).

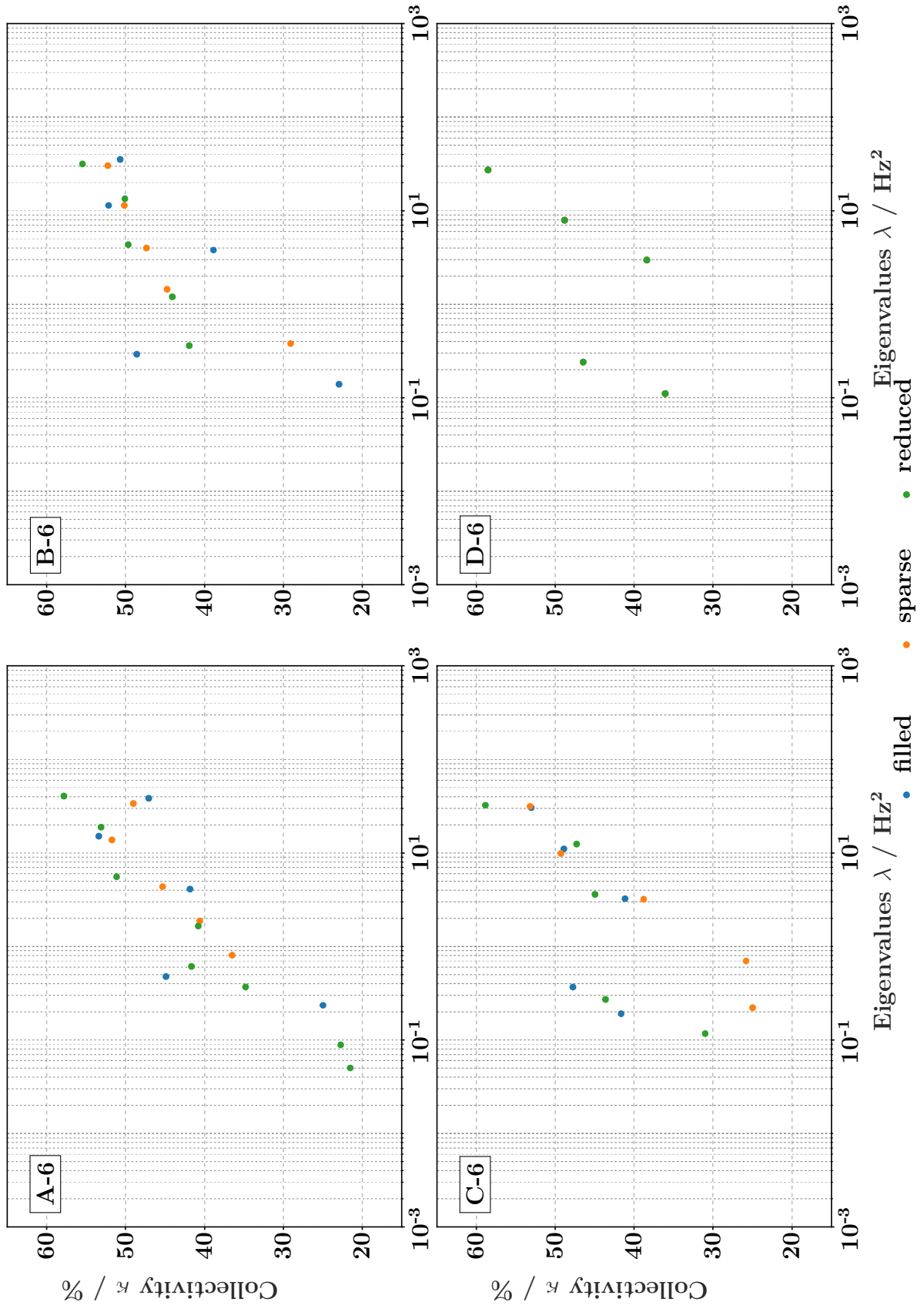


Figure 4: SECONDNA plots of all setups summarized in section 7.

For all plots the filled (with back-calculated values) data sets are considered to represent the results closest to the correct solution.

**A-6** shows a complex pattern in the SECONDA plots and therefore needs the highest attention. While the different approaches exhibit similar magnitudes  $\lambda$  for the first three eigenvalues, the respective collectivities  $\kappa$  vary strongly. This indicates that these eigenvalues are independent of experimental errors, here in the form of undefined RDCs. This was reported in literature<sup>[28]</sup> for linearly independent alignment conditions. The eigenvalues  $\lambda_4$  and  $\lambda_5$  obtained by the default implementation (eq. (18a), used in combination with the sparse data matrix) and the adapted implementation (eq. (20a), resulting in a reduced covariance matrix) both are shifted to higher values (up to 5-fold increase), compared to the eigenvalues of the filled RDC matrix. This directly shows that undefined RDCs induce heterogeneities in the datasets. The reduced covariance matrix is not capable of fully removing these heterogeneities, but the respective difference to the results of the matrix filled with back-calculated RDCs is smaller than for the scheme setting undefined RDCs to zero.

Here the most conspicuous point is that the covariance matrix obtained from eq. (20a) is the only matrix (of all SECONDA results shown here) resulting in more than five non-zero eigenvalues (threshold implemented to be interpreted as zero is  $1e-9$ ). We did not observe a direct correlation of the eigenvalues  $\lambda_n$  with  $n > 5$  and undefined RDCs in the eigenmodes (for more information see `.trj` output files in supplementary material).

When reducing the number of undefined RDCs to 12 (**B-6**) these additional non-zero eigenvalues vanish and only five eigenvalues can be observed. Just like for **A-6** the eigenvalues  $\lambda_4$  and  $\lambda_5$  of the reduced covariance matrix (eq. (20a)) are closer to the ones obtained from the filled RDC matrix. This small trend becomes clearer when the undefined RDCs are reduced even further (**C-6**, 8 undefined RDCs). Like before the eigenvalues  $\lambda_1$  to  $\lambda_3$  remain similar, but the principle variances  $\lambda_4$  and  $\lambda_5$  are now more similar for the back-calculated and reduced covariance matrix. This demonstrates that the reduction of the covariance matrix is capable of reducing the impact of undefined RDCs without fully removing media or RDC vectors from the PCA. The limits of this new approach introduced here were not investigated in depth. Thus the question whether data matrices with higher amounts of alignment media are tolerant to more undefined RDCs remains open.

The last SECONDA plot (**D-6**, no undefined RDCs) just confirms that the implementation of the reduced covariance matrix is self-consistent. All RDCs are defined, so the (mathematical) sets **I** and **J** in eq. (20a) and (20b) are the empty set  $\emptyset$  and the reduced approach results in the literature implementation.

From the data presented here we can see that for all setups **A-6** to **D-6** the first three eigenvalues  $\lambda_1$  to  $\lambda_3$  remain nearly unchanged for the different approaches to handle the undefined RDCs. This clearly shows, that these eigenvalues are insensitive to the algorithm used and therefore to heterogeneity. As a result we conclude that at minimum three strongly different alignment media are available for the TITANIA optimization.<sup>[28]</sup> This becomes even clearer when investigating the cumulative sums (see output files in supplementary material) of the eigenvalues for  $\lambda_3$  which are  $>98\%$  of the filled RDC matrix. Since we are investigating a rigid structure the data are sufficient for the protocol used. The eigenvalues  $\lambda_4$  and  $\lambda_5$  show a higher dependency on the treatment of undefined RDCs putting TITANIA to the test.

## 9 Trajectories

TITANIA uses a Monte-Carlo bootstrap to estimate the uncertainties of the structure and orientation parameters based on the user defined experimental error of the RDCs. Additionally the results of the bootstrap are used to check for convergence in the iterative TITANIA cycle. For this reason the *rmsds* of the reduced SAUPE tensors  $\mathbf{S}$ , the spherical coordinates  $p$  and their respective standard deviations  $\sigma$  are monitored. In addition the mean RDC vector lengths  $R_s$  (obtained by averaging all orientation vectors obtained from the Monte-Carlo Bootstrap for one RDC spin pair) are used.

The trajectory of the Monte-Carlo parameters,  $S_{\text{overall}}$  and the chiral volumes  $V_c$  of the independent centers (C2, C3, C4, C6, C7) in IPC are plotted to visualize the progression. The chiral volumes used do not follow the Cahn-Ingold-Prelog convention and thereby cannot be translated into absolute configurations of the respective stereocenters.

All available setups (**A-6**, **B-6**, **C-6**, **D-6**) using the recalculation and weighting scheme to handle missing RDCs are plotted using different starting structures (*epi*-C3 and random configuration).

The trajectories are rather complex for **B-6** and **C-6** and the direct results cannot be read from the trajectories. If using statistics of the chiral volumes, however, interpretation is simplified drastically (see section 10).

For the following trajectories the panels are all arranged in the same manner:

- a) C3-epimer starting structure using the weighting scheme.
- b) C3-epimer starting structure using the recalculation scheme.
- c) starting structure with random configuration using the weighting scheme.
- d) starting structure with random configuration using the recalculation scheme.

At the right side of the chiral volume trajectory the arrows indicate the correct stereo chemistry of the corresponding centers. As the runs may converge to different enantiomeric forms, this helps to discriminate the cases of correct vs incorrect relative configurations.

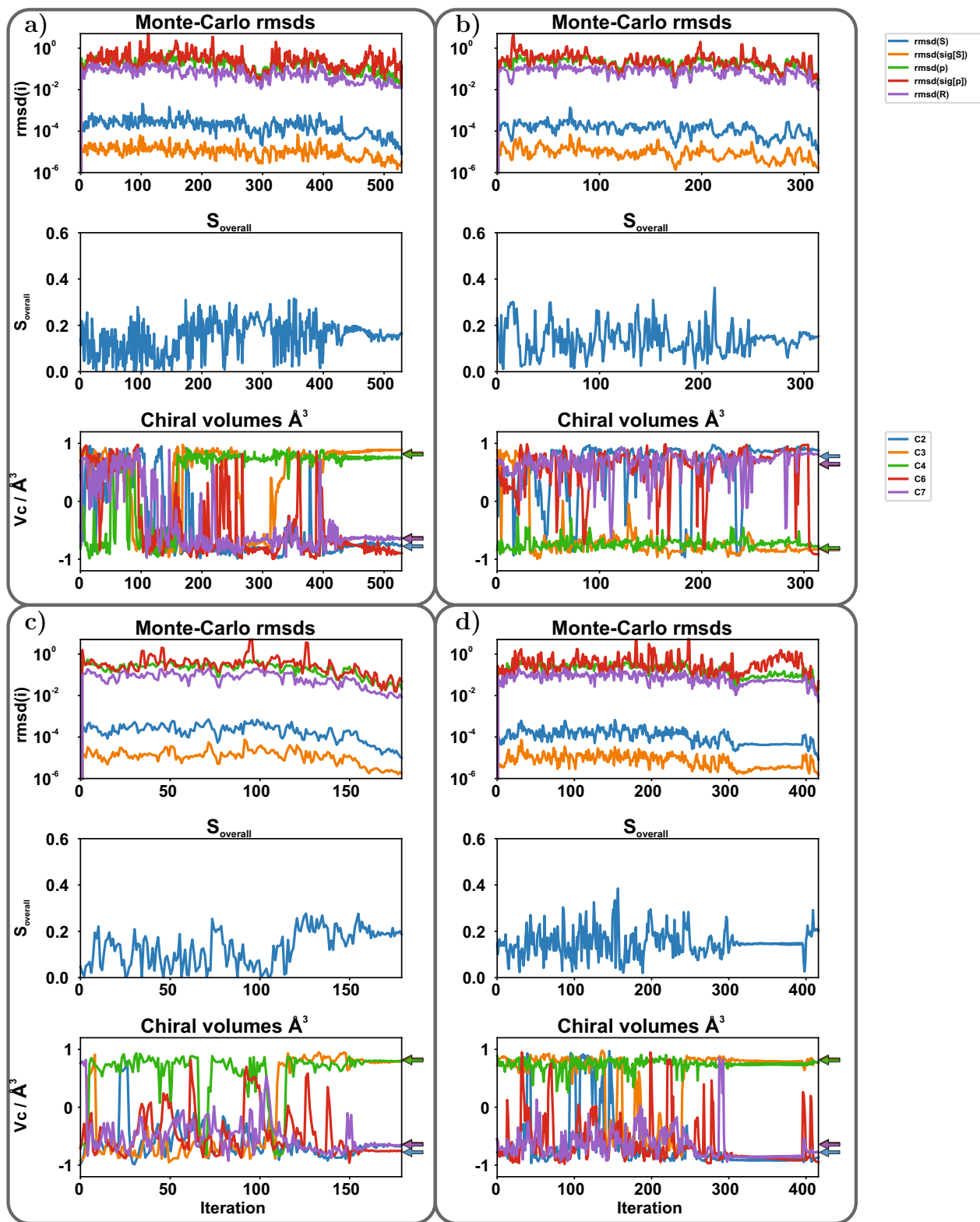


Figure 5: Trajectories for setup **A-6**. The upper panels show the runs using C3-epimer as starting structure using the weighting (**a**) and recalculation scheme (**b**). The lower panels started from a diastereomer of IPC with random configurations (**c**: weighting scheme, **d**: recalculation scheme). Note that the pairs of reference chiral volumes C3 (golden) / C4 (green) and C6 (red) / C7 (purple) coincide.

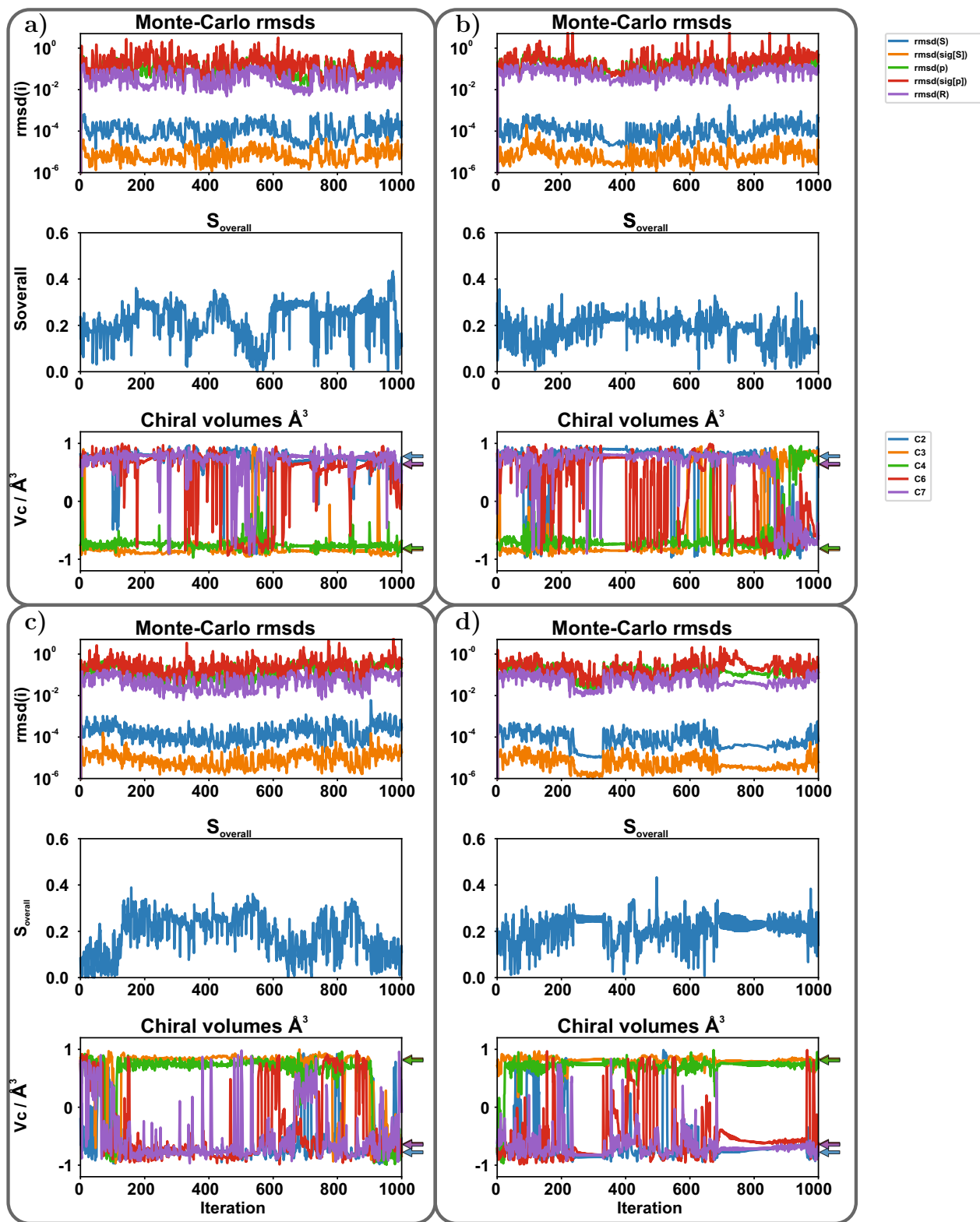


Figure 6: Trajectories for setup **B-6**. The upper panels show the runs using C3-epimer as starting structure using the weighting (a) and recalculation scheme (b). The lower panels started from a diastereomer of IPC with random configurations (c: weighting scheme, d: recalculation scheme). Note that the pairs of reference chiral volumes C3 (golden) / C4 (green) and C6 (red) / C7 (purple) coincide.

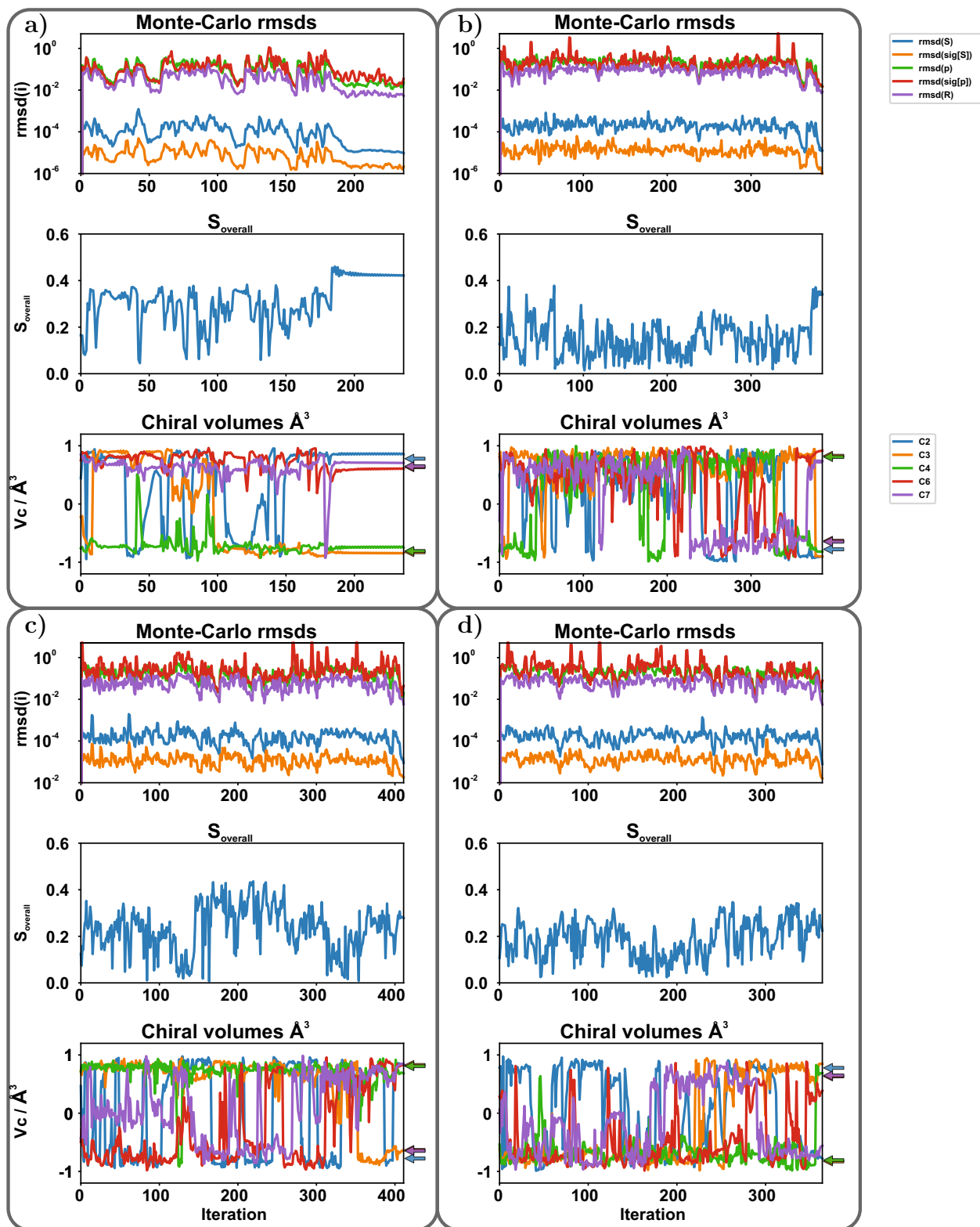


Figure 7: Trajectories for setup **C-6**. The upper panels show the runs using C3-epimer as starting structure using the weighting (**a**) and recalculation scheme (**b**). The lower panels started from a diastereomer of IPC with random configurations (**c**: weighting scheme, **d**: recalculation scheme). Note that the pairs of reference chiral volumes C3 (golden) / C4 (green) and C6 (red) / C7 (purple) coincide.



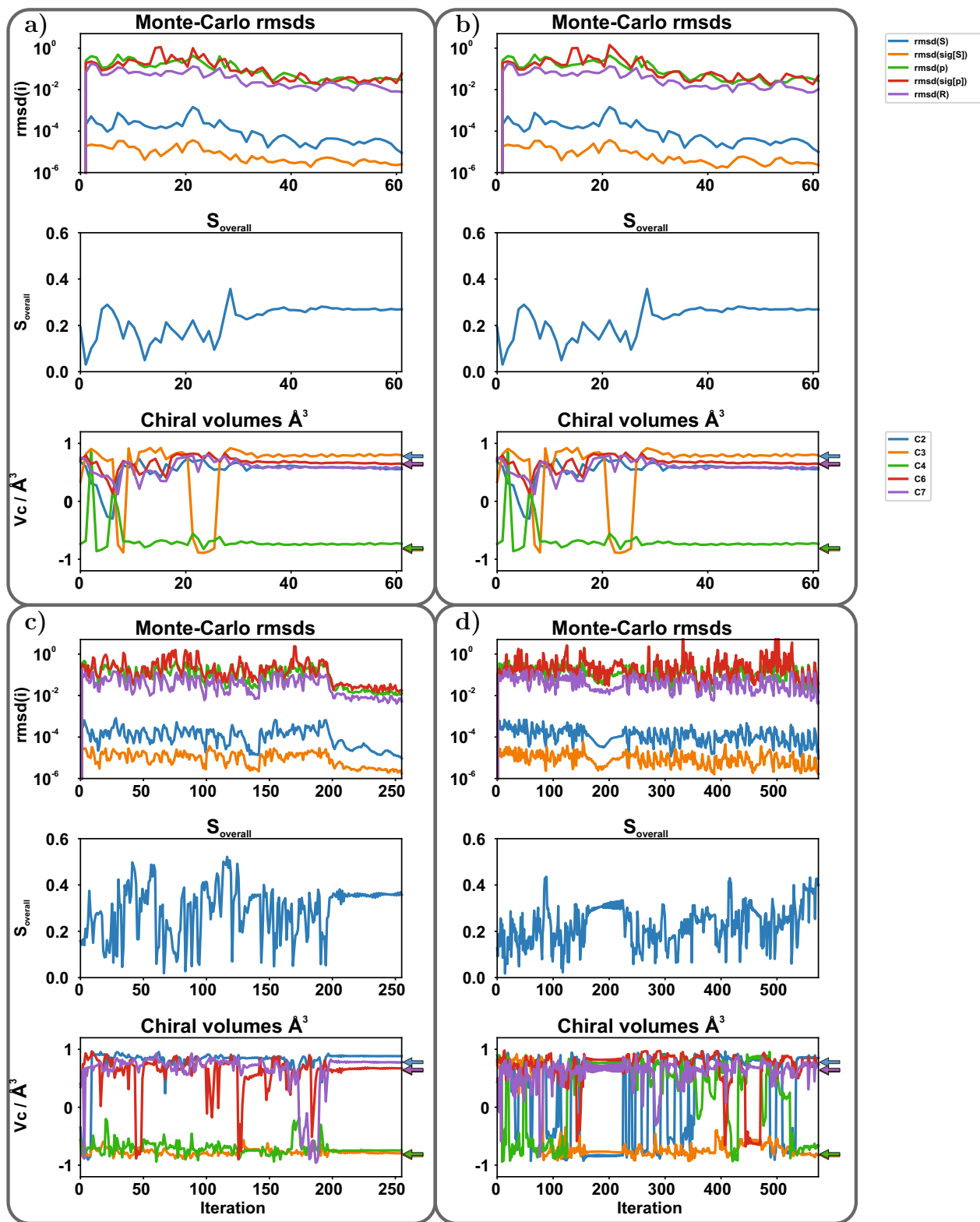


Figure 8: Trajectories for setup **D-6**. The upper panels show the runs using C3-epimer as starting structure using the weighting (**a**) and recalculation scheme (**b**). The lower panels started from a diastereomer of IPC with random configurations (**c**: weighting scheme, **d**: recalculation scheme). Note that the pairs of reference chiral volumes C3 (golden) / C4 (green) and C6 (red) / C7 (purple) coincide.

### A-6 (fig. 5):

Fig. 5 shows the trajectories of all four **A-6** runs performed. All of these converge before reaching the defined maximum iterations (1000). Especially the weighting scheme runs (panels **a** and **c**) show highly stable chiral volumes prior to the convergence. These runs result in the correct relative configuration. A close look at C7 reveals a slightly longer bond length of C7-H7a and C6-C8 (see `.xyz` files in supplementary material). This is a direct consequence of the large amount of missing RDCs, especially for H7a.

The only run not converging into the correct stereo chemistry is shown in panel **b** (recalculation scheme with *epi*-C3 starting structure). Here C6 shows an inversion a few iterations before the convergence. When taking a close look at the `.xyz` outputs it can be seen that actually stereo information is lost at that point and a strongly distorted structure (with C6 in close to square planar coordination) is obtained.

For most of the data in fig. 5 the trajectories are of high enough quality to determine the correct configurations. Nevertheless, as detailed for C7 above, a critical assessment regarding the RDC data and the `xyz` files is always advisable. If the trajectories are not of sufficient quality (as for panel **b**) an assessment of the chiral volume distribution allows for more detailed insights (see section 10).

### B-6 (fig. 6):

Fig. 6 shows the trajectories of setup **B-6** which still misses 12 RDCs (**A-6**: 23) while reducing the set size by 5 RDCs compared to **A-6** (see table 5 and 6). These 12 missing RDCs are mainly in set 05 (9 missing RDCs).

The first difference to **A-6** is that the optimizations do not converge within 1000 steps making it harder to analyse them on the first sight. A close look shows that panel **a** (weighting scheme with *epi*-C3 starting structure) and **d** (recalculation scheme with random starting structure) reveal the correct relative configuration in the end.

The other two (**b** and **c**) have incorrect and somehow random-like final configurations. This changes when comparing the whole trajectories, instead of the final configuration, to the correct relative configuration (arrows on the right, see also chiral volume distributions in section 10). Here mainly C6 (see panel **b**) is unstable, which is a behaviour already known from the previous publication and is the result of too few RDCs for C6.<sup>[1]</sup>

A more in-depth discussion based on the distributions of the chiral volumes is presented in the next section.

### C-6 (fig. 7):

**C-6** does not leave much space for a discussion of the structural information from the chiral volume plots. Panel **a** (weighting scheme with *epi*-C3 starting structure) clearly converges into the correct relative configuration. All other trajectories show no correlation on the first sight. Therefore the discussion of the structure information obtained by this setup necessitates using the chiral volume distributions in the next section.

### D-6 (fig. 8):

The setup **D-6** does not contain any missing RDCs and therefore shows the same trajectory for the weighting and recalculation scheme. This can even be seen in the Monte-Carlo values where only very small changes due to the principle of the algorithm can be found. All other values, which do not depend on random numbers, are equal (see  $S_{\text{overall}}$  and chiral volume panels). However, fast convergence like in panel **a** and **b** does not necessarily mean the correct relative configuration is found. This is seen for C3 (orange) which is obtained in the wrong configuration. The user can

recognize this when taking a critical look at the structure. The C3-H3 bond is extraordinarily short (0.82 Å) while the C2-H2 (1.24 Å) and C4-H4a (1.29 Å) bonds are very long (as discussed in the main text). This is a typical sign that no stable configuration is found here. We used two simple methods to demonstrate how to overcome this problem. The first way is to reduce all convergence criteria by 50% to run the optimization for more iteration steps. The second option is to adapt the weighting factor for bond lengths in the redundant internal coordinates. The results of these options are shown in section 11.

When using random starting structures (panel **c** and **d**) the correct relative configuration is found. Here the trajectories do not have to be the same since the starting geometries were generated randomly. This is an example how to verify a TITANIA solution. Using different starting geometries will lead to the same solution for a solid RDC base. If the structures (or single centers) do not converge into the same configuration, the alignment conditions are not independent enough, the RDCs have too large heterogeneities (e.g. experimental errors) or they are just simply too few to allow the use of this approach.

## 10 Chiral volume distributions

As demonstrated in section 9 the trajectories did not converge in all cases and complex plots of the chiral volumes are obtained. Hence we show and discuss the distribution of  $V_c$  in the following. RDCs are not capable of determining the absolute configuration of organic compounds without additional information.<sup>[31]</sup> That is why both enantiomers of IPC were reached in the optimization.

TITANIA does not implement the full Cahn–Ingold–Prelog priority rules. Therefore the sign of the chiral volumes might differ from those determined by software specialized on this task. However, here we only compare  $V_c$  within TITANIA. The reference values calculated by TITANIA for IPC are: C2:  $0.778 \text{ \AA}^3$ , C3:  $-0.828 \text{ \AA}^3$ , C4:  $-0.807 \text{ \AA}^3$ , C6:  $0.632 \text{ \AA}^3$ , C7:  $0.648 \text{ \AA}^3$  which were already used in the previous section. An equally good solution is obtained if all chiral volumes show the inverse sign corresponding to the inverse absolute but same relative configuration (= enantiomers).

In addition to the distribution of the chiral volumes, the plots contain the final chiral volume with the *rmsd* (of all possible vector permutations used in eq. (22)) as uncertainty, the mean chiral volume  $\langle V_c \rangle$  with the standard deviation as uncertainty and the median of the distribution. We assumed a normal distribution to calculate the standard deviation. While this might not be a proper statistical model, it needs to be mentioned that we do not try to extract statistical information on the probabilities of the distribution, but rather want to get an estimate of the data dispersion.  $\langle V_c \rangle$  is thus taken as assessment on the distribution of the chiral volumes in the course of the optimization. To take the trajectory into account an iteration-weighted mean  $\langle V_{c,w} \rangle$  and the respective standard deviation  $\sigma_w$  are calculated:

$$\langle V_{c,w} \rangle = \frac{\sum_{it}^{It} it \cdot V_{c,it}}{\sum_{it}^{It} it} \quad (21a)$$

$$\sigma_w = \sqrt{\frac{\sum_{it}^{It} it (V_c - \langle V_{c,w} \rangle)^2}{\sum_{it}^{It} it}} \quad (21b)$$

$$V_c = \vec{r}_1 \cdot (\vec{r}_2 \times \vec{r}_3) \quad (22)$$

where *It* is the number of iterations performed. Using the unweighted and weighted mean allows for a better interpretation of the data. In the perfect case of (close to) normal distributed data the median, weighted and unweighted mean will be very similar and unequal zero for non-planar centers. The corresponding standard deviations will be small.

If the median and the two mean values are close to zero with large standard deviations a determination of the configuration is not possible due to rapid inversions of the chiral volumes. In such a scenario the final chiral volume is not to be relied on. The weighted mean allows a differentiation of random inversions from equally distributed chiral volumes which are caused by an iterative improvement of the structure.

Even though the first case (all values match and are unequal zero) is the perfect scenario, it needs a very high quality of data (small heterogeneities, low linear dependence of the sets and large set sizes). The most common case is that the median, weighted and unweighted mean values do not match. Here different patterns in the distribution and the values discussed are possible. If the median and weighted mean are close to the final chiral volume the optimization retains the respective configuration in the late iteration steps. This is the most common scenario for well defined centers and an example is discussed below. If the median and unweighted mean are close to each other, but the final value and weighted mean are different as compared to the latter, the trajectory becomes unstable in the late iteration steps. In this case a close inspection of the

trajectory of the 3D-structure will give better insights on how to properly adapt TITANIA to solve this problem. A longer optimization or adapting the weight in the redundant internal coordinates might help driving converge towards the correct solution. The last possible pattern would be an apparently random order of the median, weighted and unweighted chiral volume, where the distribution itself might help in the interpretation. For the IPC examples discussed below, the last two cases were especially found for centres with only few RDCs. In each case the `xyz` file should be investigated in more detail, checking for centers which are highly unstable. Such centers can affect the whole optimization.

An example which shows how the correct solution is populated more in the later steps of the optimization is fig. 9. Here the mean value (red line) is around  $-0.5 \text{ \AA}^3$ . This is clearly lower than the expected value of  $-0.82 \text{ \AA}^3$  (red arrow). Since the optimization tends towards the correct configuration in the later steps the weighted mean (blue) is closer to the correct solution and very similar to the median (green). Also the standard deviation is reduced, since the incorrect configuration ( $0.82 \text{ \AA}^3$ ) has a very low weighting. The final value (grey line) is even closer to the reference value. This might be due to the fact, that this example was calculated using a damping of the redundant internal coordinates in TITANIA.

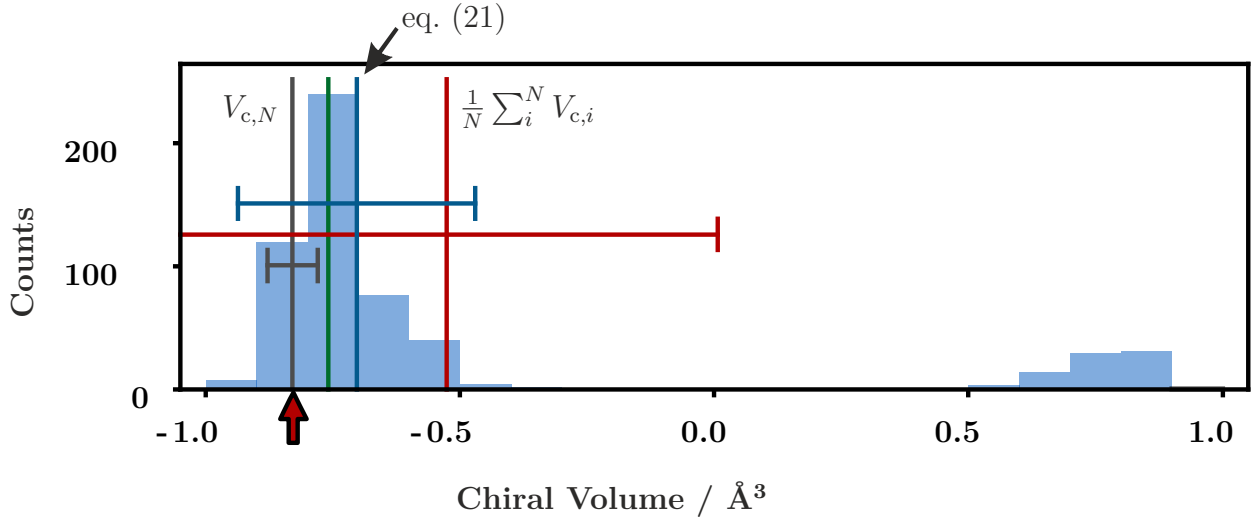


Figure 9: Exemplary distribution of chiral volumes with the following statistical parameters: the median (green), the final chiral volume  $V_c$  (grey) with the *rmsd* of all possible permutations for the  $\vec{r}_i$  (see eq. (22)), the mean chiral volume  $\langle V_c \rangle$  (red) and the weighted mean  $\langle V_{c,w} \rangle$  (blue) with their respective standard deviations. On the bottom the reference chiral volume is indicated by the red arrow.

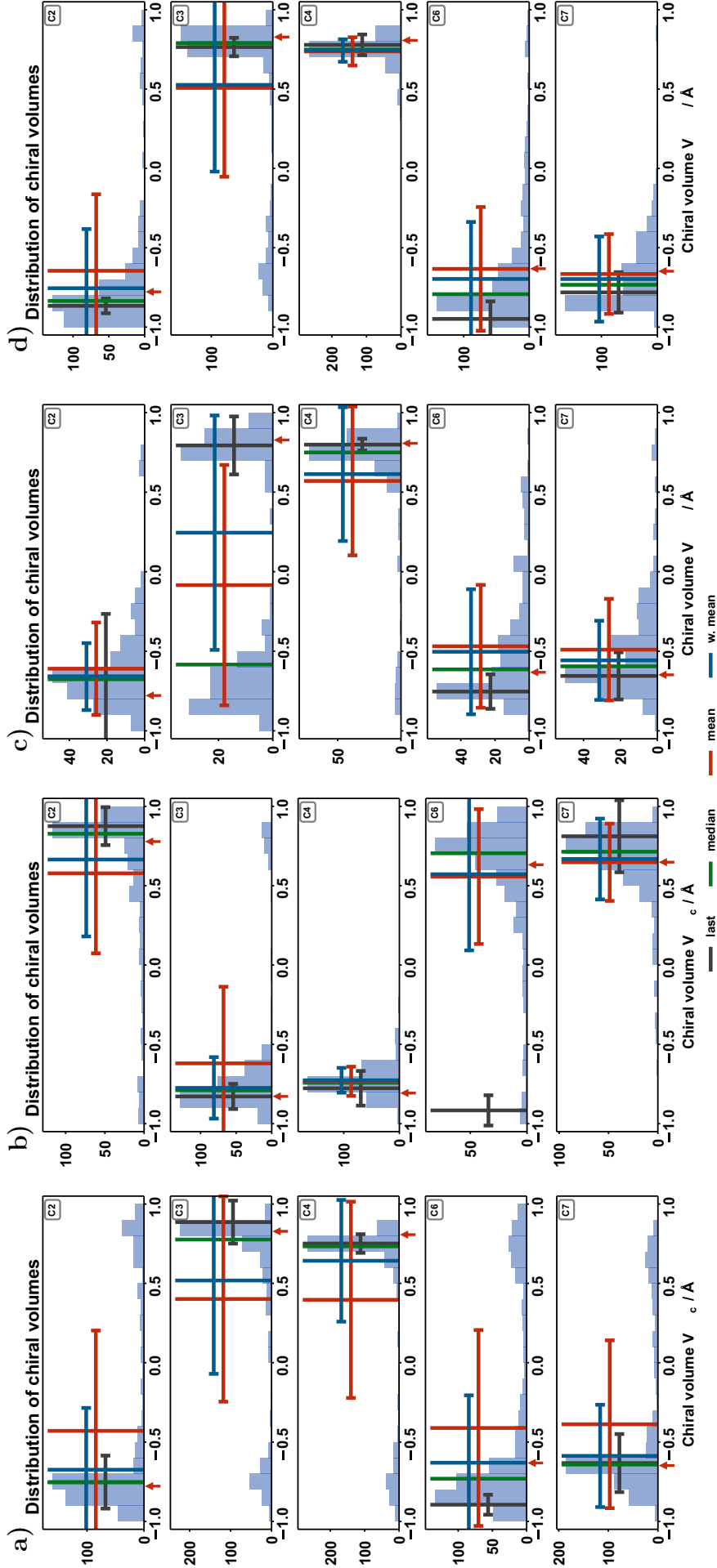


Figure 10: Chiral volume distribution of A-6 (see section 7). For more information on the plotted values see fig. 9 in the beginning of this section. Panel a (weighting scheme) and b (recalculation scheme) shows the data obtained when starting from the C3-epimer. Panel c (weighting scheme) and d (recalculation scheme) shows the corresponding data for the starting structure with a random configuration.

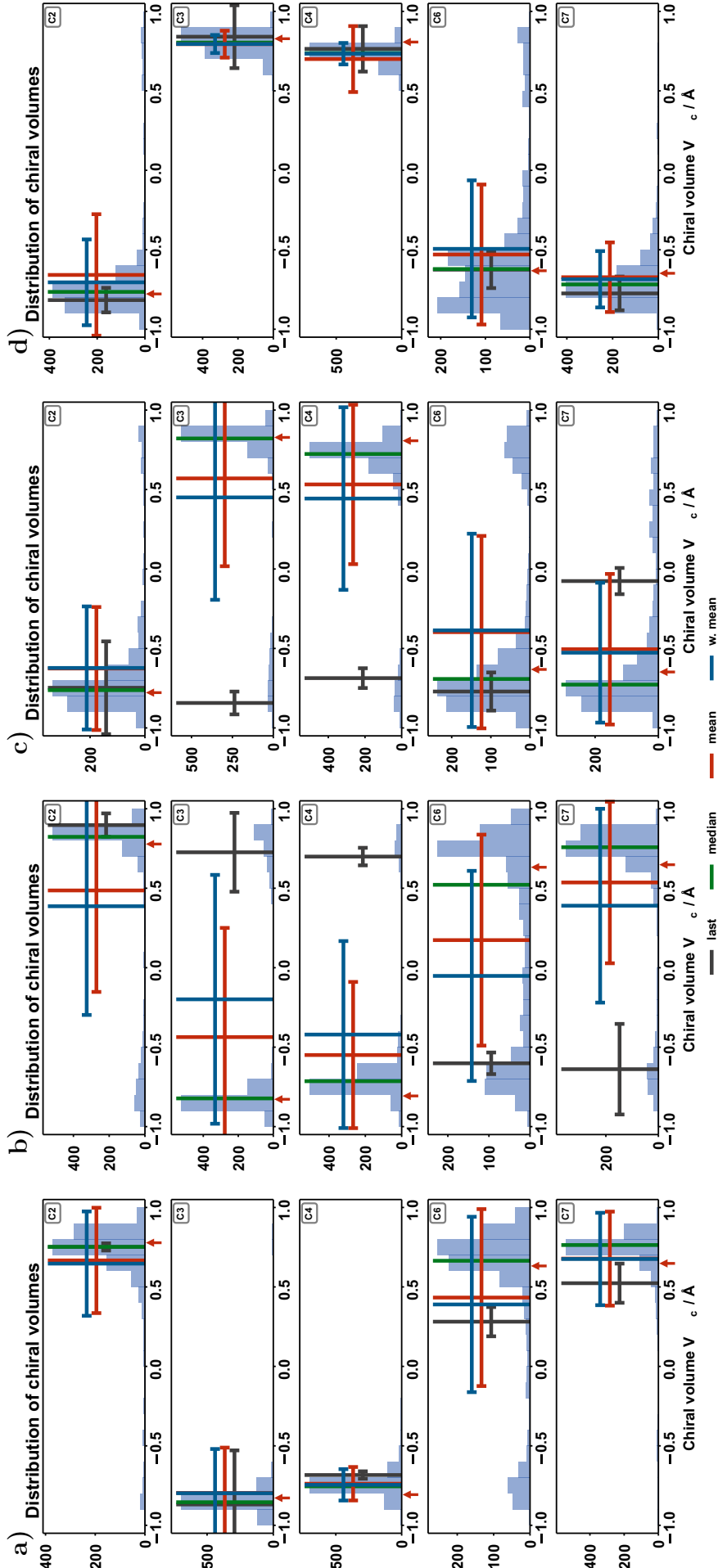


Figure 11: Chiral volume distribution of **B-6** (see section 7). For more information on the plotted values see fig. 9 in the beginning of this section. Panel **a** (weighting scheme) and **b** (recalculation scheme) shows the data obtained when starting from the C3-epimer. Panel **c** (weighting scheme) and **d** (recalculation scheme) shows the corresponding data for the starting structure with a random configuration.

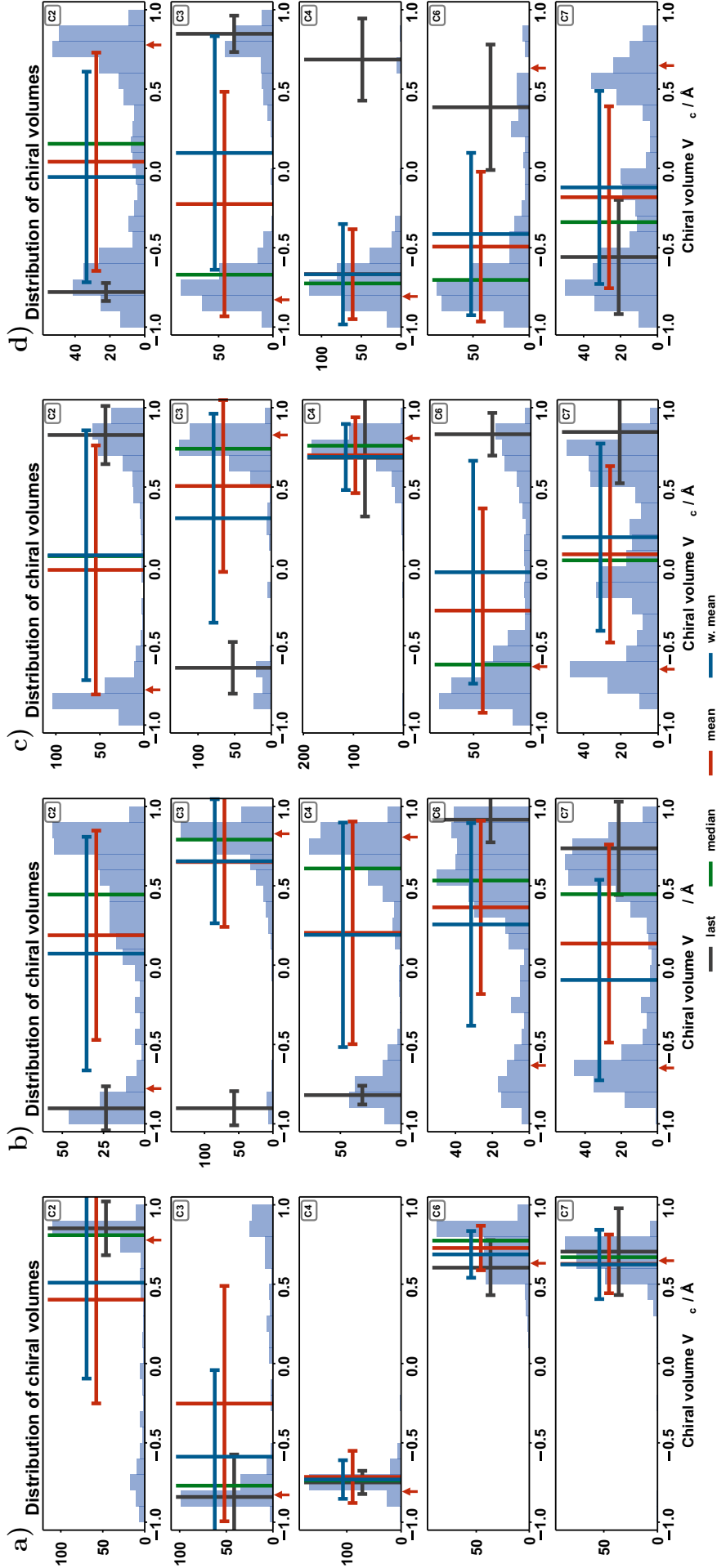


Figure 12: Chiral volume distribution of C-6 (see section 7). For more information on the plotted values see fig. 9 in the beginning of this section. Panel a (weighting scheme) and b (recalculation scheme) shows the data obtained when starting from the C3-epimer. Panel c (weighting scheme) and d (recalculation scheme) shows the corresponding data for the starting structure with a random configuration.



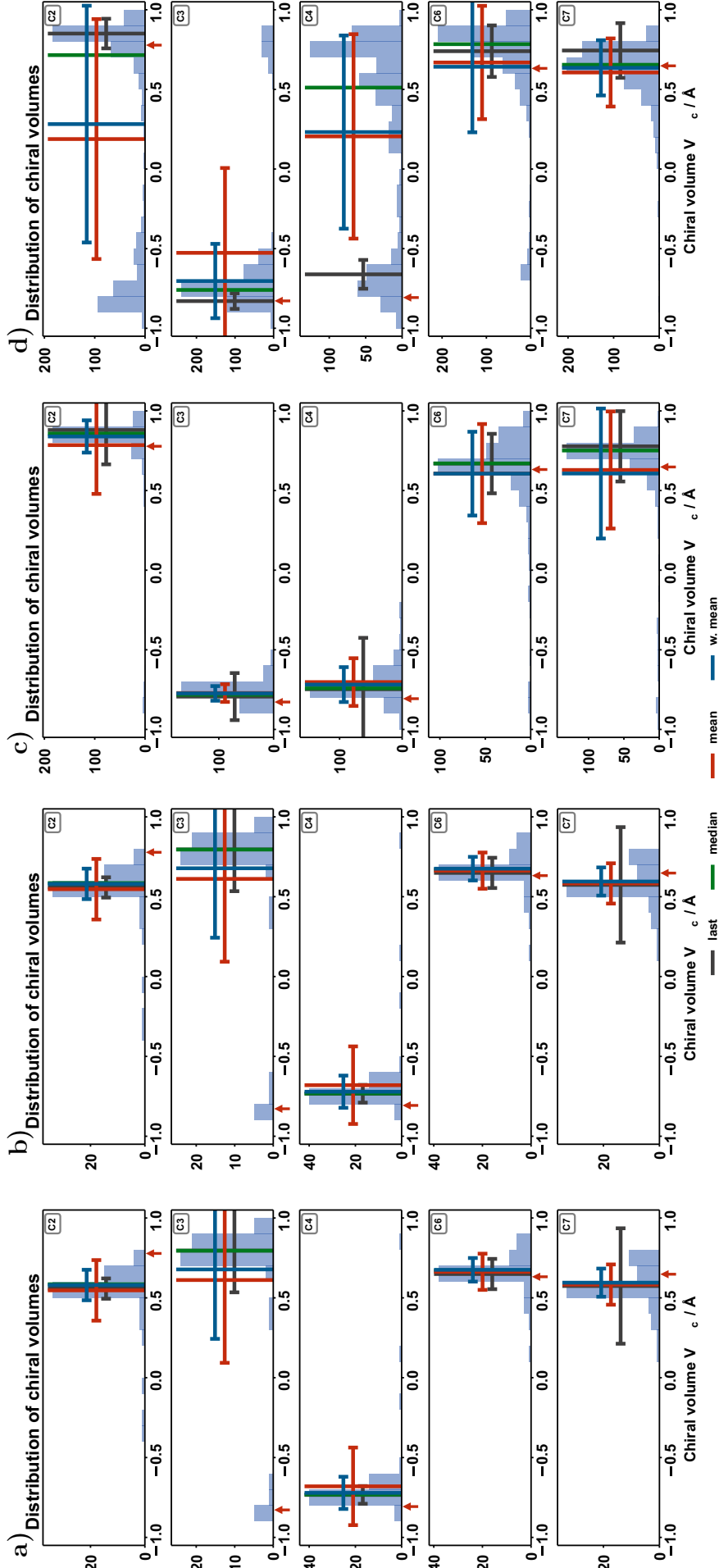


Figure 13: Chiral volume distribution of **D-6** (see section 7). For more information on the plotted values see fig. 9 in the beginning of this section. Panel **a** (weighting scheme) and **b** (recalculation scheme) shows the data obtained when starting from the C3-epimer. Panel **c** (weighting scheme) and **d** (recalculation scheme) shows the corresponding data for the starting structure with a random configuration.

### A-6 (fig. 10):

Like in the trajectories discussed above we see that **A-6** shows good results in terms of relative configurations. Panel **a** and **d** have rather sharp distributions of the chiral volumes and the values discussed cluster near the correct chiral volume. The only wrong configuration is in panel **b** (recalculation scheme with *epi*-C3 starting structure) for the center C6. When investigating this plot it becomes clear that both mean values and the median suggests the correct configuration. This was also seen in the trajectory, where the wrong configuration was obtained within the very last steps. As stated above determining the correct configuration of center C6 with higher certainty is hard due to the low amount of experimentally (easily) accessible RDCs.

The center C3 in panel **c** (weighting scheme with random starting structure) is an example which is hard to interpret. It shows that in the early iterations the wrong configuration for C3 was present and later the correct configuration was found. This interpretation is based on the fact, that the median and unweighted mean have a negative magnitude, while the weighted mean is positive. This optimization is a perfect example for the power of using different starting structures when a center is not defined properly in a single run, since all other panels verify the correct chiral volume of C3. All other centers are properly defined by each single run as can be seen from the distribution of the parameters discussed.

### B-6 (fig. 11):

**B-6** contains trajectories that are very hard to interpret and do not converge within 1000 steps. As read from the trajectories the panels **a** (weighting scheme with *epi*-C3 starting structure) and **d** (recalculation scheme with random starting structure) have the correct relative configuration for the final structure. This configuration is confirmed by the statistical interpretation. As before, panels **b** (recalculation scheme with *epi*-C3 starting structure) and **c** (weighting scheme with random starting structure) are harder to interpret. In **b** the correct configuration of C2 is represented by all statistical parameters shown. C3, C4 and C7 show good agreement between the statistic parameters and the correct configuration, but the final chiral volume has the wrong sign. C6 does not show an unambiguous configuration. A close look at the the `.xyz` file reveals the origin of this strange behaviour. Between iteration 860 and 920 an inversion of the dependent centers C1 and C5 occurs, which leads to a slow inversion of the whole structure (resulting in the enantiomer). The center C2, which does not invert until iteration 1000, shows the inverse solution of the C2-C10 vector. With the correct solution for this vector the enantiomer would have been present. The same behaviour can be observed in the `.xyz` file for panel **c**, where a full inversion of the structure happens late in the trajectory which is partially reverted.

### C-6 (fig. 12):

The trajectories for **C-6**, apart from panel **a** of fig. 7, are too complex to allow any reasonable discussion. This is also reflected in the distributions of the chiral volumes. Panel **a** shows the only run that results in a clear relative configuration of IPC.

Neglecting C6 the majority of the centers in panel **b** to **d** show either the correct or an undefined configuration when checking the statistic parameters. In the concept paper a similar behaviour was observed. There the runs with medium set size (23 RDCs) with experimental error **1-D23** and **1-E23** showed the most unstable trajectories of the whole comparison.<sup>[1]</sup> In this example (**C-6**) the reason is most likely the significantly reduced number of RDCs (28 RDCs) while four long-range RDCs still are undefined in two media. These two points might influence each other negatively. The difference to **D-6** are only these four RDC vectors, but the optimization of **D-6** shows much better results.

### D-6 (fig. 13):

As discussed in the previous chapter (trajectories) panel **a** and **b** are exactly the same and only show the wrong configuration on C3, which can be identified in the `.xyz` file. Adapting the settings of TITANIA (see section 11) results in the correct configuration of this center. All other centers in **a** and **b**, like the whole panel **c**, show the correct results with high accuracy. Panel **c** shows two centers that have to be discussed: C2 and C4. C2 shows low mean values that are slightly shifted to the correct configuration. This is an example of a late inversion, explaining why the unweighted mean is close to  $0.0 \text{ \AA}^3$ . The weighted mean is slightly shifted towards the median and the final value. This indicates the correct solution. As stated before, the correct configuration can be confirmed when using other starting geometries (panels **a** to **c**).

Harder is C4, which cannot be interpreted without the `.xyz` data. The trajectory of the Cartesian coordinates of this center show a trend towards the inversion early on in the trajectory. This can be concluded from the distorted bond lengths at C4 and the neighbouring C3 and C4 (this was discussed for C3 **D-6** in section 9). Additionally the bond angles of the H4 protons are very conspicuous. Only when the correct (final) configuration is reached, the bond angles and lengths reach reasonable values.

The data shown demonstrate that TITANIA is capable of optimizing structures of small, rigid organic compounds based solely on experimental RDCs. This confirms the findings utilizing simulated RDCs in the concept paper. We also were able to show that missing RDCs can be handled in the five alignment media approach using different schemes. For the cases that did not provide a result with sufficiently high certainty (e.g. C2 in fig. 13 panel **d**) a change in the input structure was a reasonable approach to verify the results.

## 11 Adapted runs of setup D-6

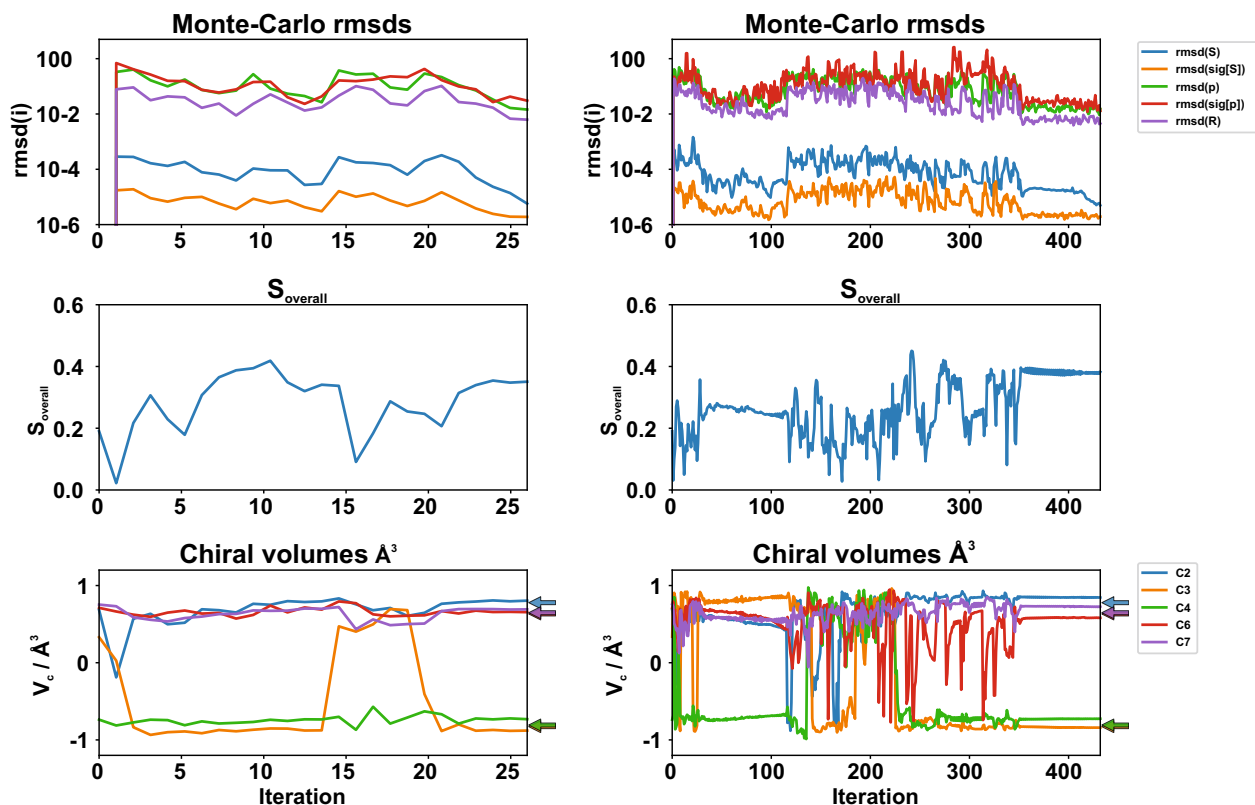


Figure 14: Trajectories for **D-6** with reduced weighting of the bond lengths (left) and tighter stop criteria (right).

The trajectory using reduced weighting of the bond lengths shows a very fast convergence into the correct relative configuration. In a practical structure determination this trajectory is too short for a doubtless assignment of the correct configuration. Here it is used as an example of how the weighting of one type of internal coordinates can change the behaviour of the optimization. The second approach, lowering the stop criteria to enforce a longer trajectory, results in the correct configurations while achieving a superior sampling. This allows for a reliable determination of the structure.

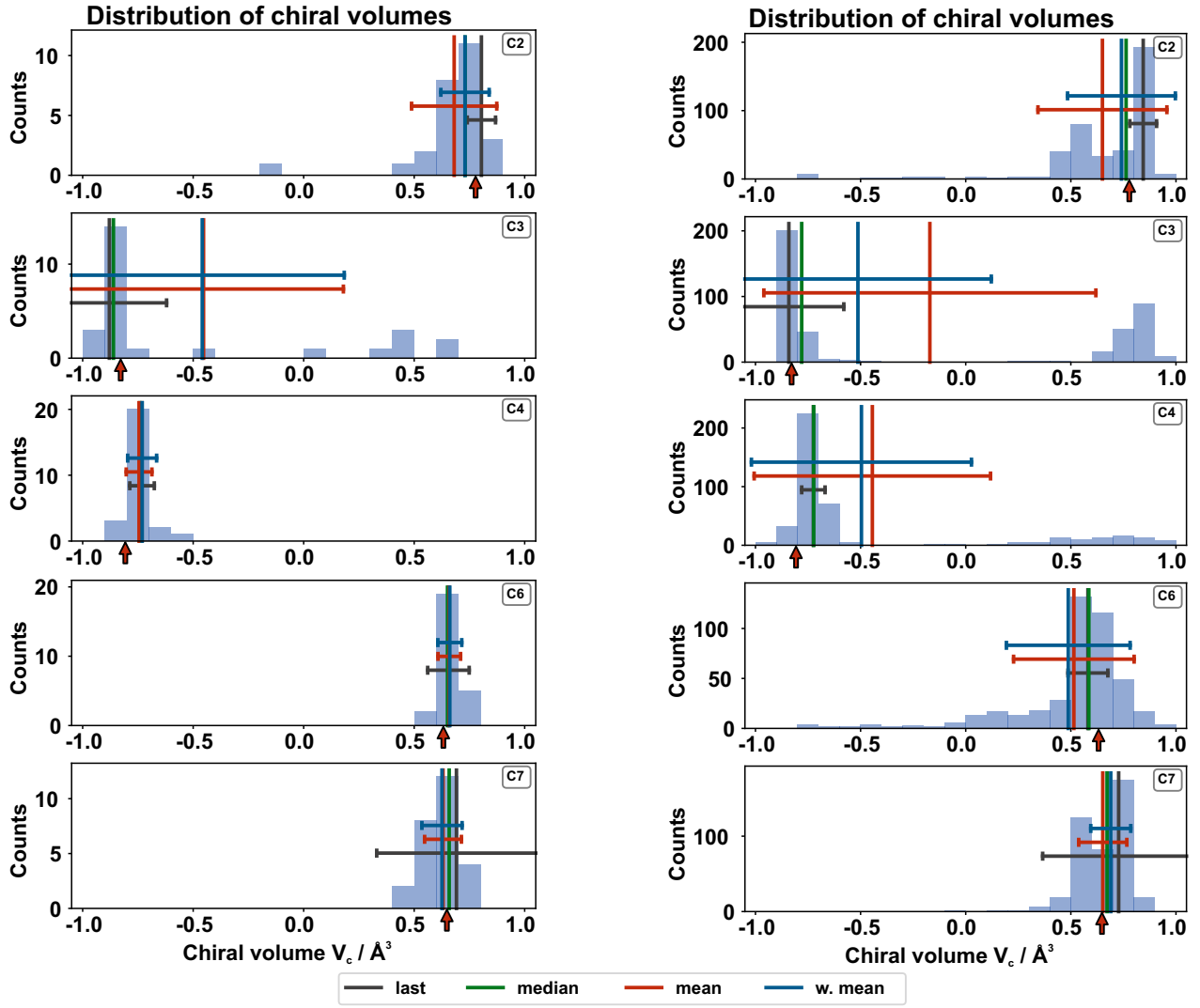


Figure 15: Chiral volume distributions for **D-6** with reduced weighting of the bond lengths (left) and tighter stop criteria (right).

The chiral volume distributions of **D-6** with adapted parameters show the results expected from the trajectories. All centers show well defined configurations which match the correct values. As stated for the trajectories, the left plot (reduced bond weighting) is not resilient for an unambiguous structure determination (as it is too short) and is reported to demonstrate the behavior of the optimization.

## 12 Content of Supplementary Material

The supplementary material is available as two zip-archives which contain all information used to perform the TITANIA runs with the respective outputs (TITANIAdata.zip) and the NMR-spectra (NMRspectra.zip). A README.md file was added to the TITANIAdata.zip archive using the markdown syntax. This is also added here:

```
# Incomplete Data Sets in the Model Free Analysis of Experimental  
Residual Dipolar Couplings in Small Organic Compounds  
## Felix A. Roth, Volker Schmidts, Jan Rettig  
and Christina M. Thiele
```

```
## Top Level Directories  
- weight_C3Start    (—> weighting scheme starting from C3-epimer)  
- weight_RandConf   (—> weighting scheme starting from random  
    configurations)  
- recalc_C3Start     (—> recalculation scheme starting from  
    C3-epimer)  
- recalc_RandConf    (—> recalculation scheme starting from  
    random configurations)
```

All directories listed above contain subdirectories for the individual runs A-6 to D-6.

recalc\_C3Start additionally contains the adapted runs for setup D-6.

Some individual runs use different keywords, RDC data and structures. These are located in the remaining directories:

```
- keywords  
- rdcs  
- structures
```

## References

- [1] F. A. Roth, V. Schmidts, C. M. Thiele, *under review* **2021**.
- [2] J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, R. Brüschweiler, *JACS* **2001**, *123*, 6098–6107.
- [3] N.-A. Lakomek, K. F. A. Walter, C. Farès, O. F. Lange, B. L. de Groot, H. Grubmüller, R. Brüschweiler, A. Munk, S. Becker, J. Meiler, C. Griesinger, *J. Biomol. NMR* **2008**, *41*, 139.
- [4] A. Saupe, G. Englert, *Phys. Rev. Lett.* **1963**, *11*, 462–464.
- [5] F. Kramer, M. V. Deshmukh, H. Kessler, S. J. Glaser, *Conc. Magn. Reson. A* **2004**, *21A*, 10–21.
- [6] N. R. Skrynnikov, N. K. Goto, D. Yang, W.-Y. Choy, J. R. Tolman, G. A. Mueller, L. E. Kay, *J. Mol. Biol.* **2000**, *295*, 1265–1273.
- [7] W. Peti, J. Meiler, R. Brüschweiler, C. Griesinger, *JACS* **2002**, *124*, 5822–5833.
- [8] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **1996**, *17*, 49–56.
- [9] V. Bakken, T. Helgaker, *J. Chem. Phys.* **2002**, *117*, 9160–9174.
- [10] T. A. Halgren, *Journal of Computational Chemistry* **1996**, *17*, 490–519.
- [11] T. A. Halgren, *Journal of Computational Chemistry* **1996**, *17*, 520–552.
- [12] T. A. Halgren, *Journal of Computational Chemistry* **1996**, *17*, 553–586.
- [13] T. A. Halgren, R. B. Nachbar, *Journal of Computational Chemistry* **1996**, *17*, 587–615.
- [14] T. A. Halgren, *Journal of Computational Chemistry* **1996**, *17*, 616–641.
- [15] A. Ben-Israel, T. N. E. Greville in *Generalized Inverses: Theory and Application*, Springer-Verlag New York, **2003**.
- [16] J. R. Tolman, *JACS* **2002**, *124*, 12020–12030.
- [17] P. Trigo-Mouriño, C. Merle, M. R. M. Koos, B. Luy, R. R. Gil, *Chem. Eur. J.* **2013**, *19*, 7013–7019.
- [18] M. Hirschmann, D. S. Schirra, C. M. Thiele, *Macromolecules* **2021**, *54*, 1648–1656.
- [19] A. Marx, V. Schmidts, C. M. Thiele, *Magn. Reson. Chem.* **2009**, *47*, 734–740.
- [20] S. Jeziorowski, C. M. Thiele, *Chem. Eur. J.* **2018**, *24*, 15631–15637.
- [21] A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler, B. Luy, *J. Magn. Reson.* **2008**, *192*, 314–22.
- [22] L. Castañar, E. Sistaré, A. Virgili, R. T. Williamson, T. Parella, *Magn. Reson. Chem.* **2015**, *53*, 115–119.
- [23] G. Kummerlöwe, S. Schmitt, B. Luy, *The Open Spectroscopy Journal* **2010**, *4*, 16–27.
- [24] K. Fehér, S. Berger, K. E. Kövér, *J. Magn. Reson.* **2003**, *163*, 340–346.
- [25] D. Sinnaeve, M. Foroozandeh, M. Nilsson, G. A. Morris, *Angew. Chem. Int. Ed.* **2016**, *55*, 1090–1093.
- [26] L. Verdier, P. Sakhaii, M. Zweckstetter, C. Griesinger, *J. Magn. Reson.* **2003**, *163*, 353–9.
- [27] R. Berger, C. Fischer, M. Klessinger, *J. Phys. Chem. A* **1998**, *102*, 7157–7167.

- [28] J.-C. Hus, R. Brüschweiler, *J. Biomol. NMR* **2002**, *24*, 123–132.
- [29] R. Brüschweiler, D. A. Case, *Phys. Rev. Lett.* **1994**, *72*, 940–943.
- [30] J.-C. Hus, W. Peti, C. Griesinger, R. Brüschweiler, *JACS* **2003**, *125*, 5596–5597.
- [31] R. Berger, J. Courtieu, R. R. Gil, C. Griesinger, M. Köck, P. Lesot, B. Luy, D. Merlet, A. Navarro-Vázquez, M. Reggelin, U. M. Reinscheid, C. M. Thiele, M. Zweckstetter, *Angew. Chem. Int. Ed.* **2012**, *51*, 8388–8391.