Extending the applicability of P3D for structure determination of small molecules

Alain Ibáñez de Opakua[1] and Markus Zweckstetter[1,2]

[1]German Center for Neurodegenerative Diseases (DZNE), Von-Siebold-Str. 3a, 37075 Göttingen, Germany.
5 [2]Department for NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany.

Correspondence to: Alain Ibáñez de Opakua (Alain.Ibanez-de-Opakua@dzne.de) and Markus Zweckstetter (Markus.Zweckstetter@dzne.de)

Abstract. The application of anisotropic NMR parameters for the correct structural assignment of small molecules requires
10 the use of partially ordered media. Previously we demonstrated that the use of P3D simulations using PBLG as alignment medium allows the determination of the correct diastereomer from extremely sparse NMR data. Through the analysis of the structural characteristics of small molecules in different alignment media we here show that when steric or electrostatic factors dominate the alignment, P3D-PBLG retains its diastereomer discrimination power. We also demonstrate that P3D simulations can define the different conformations of a flexible small molecule from sparse NMR data.

1 Introduction

Complete structure determination of small molecules, including stereochemistry, is a challenging task and an important step in organic chemistry and drug discovery. In the last years anisotropic NMR parameters, especially residual dipolar couplings (RDCs), have been used to determine the conformation of small organic molecules in organic solvents (Li et al., 2018). RDCs provide a spatial view of the relative orientations of bonds irrespective of internuclear distances being a good complement for
20 conventional NMR restraints such as NOE distances (Anet and Bourn, 1965) and dihedral angles from 3J spin-spin coupling constants (Haasnoot et al., 1980). However, RDCs cannot be measured in isotropic conditions because of the averaging of anisotropic NMR parameters by uniform molecular tumbling (Luy and Kessler, 2006). Thus, access to anisotropic NMR parameters requires the generation of anisotropic environments in solution and this is achieved by using an alignment medium (Canet et al., 1995; Lesot et al., 1995, 1996a, 1996b; Tjandra and Bax, 1997).

25

There are two major types of alignment media with different inherent mechanism to orient tumbling molecules: lyotropic liquid crystalline (LLC) phases and strain induced alignment in a gel (SAG) (Tycko et al., 2000; Sass et al., 2000; Bax, 2003; Böttcher and Thiele, 2012; Canales et al., 2012; Leyendecker et al., 2017; Schmidts, 2017; Lesot et al., 2019; Krupp et al., 2020). The LLC phases spontaneously align in the presence of strong external magnetic fields because of their large magnetic
30 susceptibility anisotropy. This alignment is then partially transmitted to the solvent and the molecules in solution with the

degree and characteristics of the alignment depending on the concentration, temperature and other parameters of the sample (Krupp and Reggelin, 2012). Alignment requires a minimum concentration of lyotropic medium and often aligns strongly at this concentration, which limits the tunability of the alignment strength. The degree of orientation in these media is therefore sometimes too large, complicating the extraction of RDCs. The strain induced alignment in a gel (SAG) method generates the anisotropy mechanically, by compressing or stretching the gels. The alignment is then independent of the magnetic field and scalable over a wide range, making it easier to tune the alignment strength. However, sample preparation can be more difficult in the case of compressed and stretched gels. In addition, it often takes several days for the solute molecule to properly diffuse into the gel (Li et al., 2018).

Each alignment medium has its own degree of induced order and orientation of alignment. These properties can be influenced by experimental conditions and the properties of the small molecule to be aligned. How this happens is largely unknown. To fill this gap, we recently developed a three-dimensional molecular alignment model termed P3D (Ibáñez de Opakua et al., 2020). P3D allows the establishment of a quantitative correlation between the atomic structure of the alignment medium, the molecular structure of the small molecule and molecule-specific anisotropic NMR parameters. For the implementation of the model we selected poly(γ-benzyl-L-glutamate) (PBLG) as alignment medium. PBLG forms a LLC phase and has a well-defined α-helical structure (Doty et al., 1954; Marx and Thiele, 2009). The P3D simulation uses a combination of steric obstruction and continuum electrostatics. Analysis of several small molecules demonstrated that the P3D model reliably discriminates between different relative configurations of small molecules dissolved in PBLG (Ibáñez de Opakua et al., 2020). To gain insight into the applicability of P3D-based enantiomer discrimination when other alignment media are used, we here investigate how different the alignment of solutes dissolved in other alignment media are and how these differences are tuned by the type of solute. A better understanding of these questions will allow the application of molecular alignment simulations performed for one alignment media to other media if the alignment is expected to be similar. In addition, an informed choice of the alignment medium might become possible when different alignments are required to discriminate between candidate structures (Ramirez and Bax, 1998).

The discrimination between different structures of a molecule is a major application of anisotropic NMR parameters (Zweckstetter and Bax, 2000). In the case of small molecules, the determination of the constitution by NMR spectroscopy is often a standard procedure, while the determination of its conformation and configuration can be more challenging. Relative configuration determination has been improved in the last years (Li et al., 2018; Liu et al., 2018; Ibáñez de Opakua et al., 2020), in particular through the use of RDCs and residual chemical shift anisotropies (RCSAs). In addition, a number of different chiral alignment media, which possess enantiodifferentiating properties, were developed (Marx et al., 2009; Arnold et al., 2010; Krupp and Reggelin, 2012; Meyer et al., 2012; Hansmann et al., 2016; Reller et al., 2017). However, it is so far not possible to reliably predict the enantiodifferentiating properties from the molecular properties of the chiral alignment medium and the small molecule, leaving the determination of the absolute configuration on the basis of anisotropic NMR

65 parameters an unsolved problem (Berger et al., 2012). In addition, flexible small molecules with multiple conformations complicate the use of least-square methods to fit the experimental RDC/RCSA values to an ensemble of structures, especially if the conformers align with different alignment tensors. Here molecular alignment simulations might help to identify those structures that fit better to the experimental RDCs. We therefore also investigate in the current study, if P3D simulations are not only beneficial for the analysis of the relative configuration of small molecules, but also for their conformational analysis.

2 Methods

Strychnine and IPC structures were built as described previously (Strychnine: Bifulco et al., 2013; IPC: Ibáñez de Opakua et al., 2020). RDCs of the molecules in the different alignment media were obtained from the references of Table 1. Sucrose conformer structures and RDCs were obtained from (Ndukwe et al., 2019; supplemental information).

75 Molecular alignment simulations (P3D), implemented in the software PALES (Zweckstetter, 2008), were performed as described previously (Ibáñez de Opakua et al., 2020). During the simulation the solute molecule is moved in steps on a three-dimensional grid that covers the central part of PBLG. At each step, an uniform distribution of different solute molecule orientations is sampled. The spacing of the three-dimensional grid was set to 0.4 Å and the number of sampled orientations to 1,800 (100 orientations on the unit sphere and 18 in the third dimension). The interaction energy between the solute molecule

80 and the PBLG particle is then calculated for each orientation/grid position on the basis of the precomputed potential file of the PBLG particle and the charges of the solute molecule. The interaction energy is converted into a Boltzmann weighing factor and the RDCs are calculated. Charges of the small molecules were calculated using AtomicChargeCalculator server (Ionescu et al., 2015) via the electronegativity equalization method based on a common charge calculation scheme (atoms-in-molecules) and a robust quantum mechanical approach (HF/6-311G). Therefore, the PALES command used to run the simulations is as

85 follows:

```
pales -elPales -3D -pot3D PBLG.dx -lcS 0.8 -maxPot 2 -z1 150 -zN 250 -nX 129 -nY 129 -nZ 385 -dX 0.4 -dY 0.4 -dZ 0.4 -
H -nosurf -pdb SmallMolecule.pdb -inD RDCs.tbl -wv 0.12 -rM 8 -pka charges.pka -outD output.out
```

90 where PBLG.dx is the potential file of PBLG obtained from APBS, SmallMolecule.pdb is the PDB file of the small molecule, RDCs.tbl is the list of experimental RDCs, charges.pka is the list of atomic charges obtained from AtomicChargeCalculator server as explained, and output.out is the output file with the final results.

The alignment tensors for strychnine and (-)-IPC in the different alignment media and for the different conformers of sucrose

95 were calculated by best-fitting experimental RDCs to the respective structures using singular value decomposition (SVD) as implemented in the software PALES (Zweckstetter, 2008). To evaluate the correlation between experimental and

fitted/simulated RDCs, R (Pearson's correlation coefficient) and RQ parameters are used. The quality parameter RQ is defined as $(R+1)^2/Q_S$, where Q_S is the RDC quality factor $Q = rms(D^{exp}-D^{P3D})/rms(D^{exp})$ scaled by the slope of the D^{exp} vs D^{P3D} fitting (Ibáñez de Opakua et al., 2020). Variations in the SVD-derived quality measures R and RQ were evaluated using a Monte Carlo noise method (Zweckstetter and Bax, 2002), in which random noise was added to the experimental RDCs according to their estimated accuracy.

In order to calculate the relative populations of each sucrose conformer from the P3D-predicted RDCs, RQ values were maximized by a grid search over the conformer populations using steps of 1 %.

3 Evaluation of the partial alignment of P3D-PBLG and different alignment media with two different solutes

Several different alignment media for organic solvents are nowadays available. In order to evaluate the alignment properties of these media, we selected the two most widely studied molecules in the field: strychnine and isopinocampheol (IPC). Strychnine is a reference compound for relative configuration determination because of the high number of chiral centers (six chiral centers generating 13 diastereomers). In addition, the low flexibility of strychnine minimizes contributions from different conformations. IPC is also a rigid molecule, with little overlap in the two-dimensional proton-carbon correlation spectrum and with both enantiomers available. Because of these favorable properties, the groups of Reggelin and Thiele used IPC for the development of alignment media with enantiodifferentiation capabilities. As we are not focusing here on the absolute configuration problem, we selected the enantiomer (-)-IPC for the analysis. We further examined the literature and extracted a set of nine alignment media for each of the compounds from which five are shared. Table 1 shows the list of the 12 selected alignment media together with the references from where the data were taken. The structures of the basic units of the alignment media are displayed in Figure 1.

Table 1: Selected alignment media for which experimental RDCs of strychnine and (-)-IPC were reported.

	Strychnine	
PBLG	poly(γ-benzyl-L-glutamate)	Liu et al., 2018
PELG	poly(γ-ethyl-L-glutamate)	Thiele, 2004
PMMA	poly(methyl methacrylate)	Nath et al., 2015
PIAF	poly(L-isocyanoalanyl-L-phenylalanine benzyl ester)	Li et al., 2017
PS	Cross-linked polystyrene	Luy et al., 2004
PA1	Poly-A-1/L-alanine derived polyacetylene	Dama and Berger, 2012b
PL1	Poly-L-1/poly(phenylisocyanide)	Dama and Berger, 2012a
PALV	L-valine derived polyacetylene	Meyer et al., 2012
PADV	D-valine derived polyacetylene	Meyer et al., 2012
	(-)-IPC (isopinocampheol)	
PBLG	poly(γ-benzyl-L-glutamate)	Marx et al., 2009
PBDG	poly(γ-benzyl-D-glutamate)	Marx et al., 2009
PELG	poly(γ-ethyl-L-glutamate)	Hansmann et al., 2016
PALF300	L-phenylalanine derived polyacetylene at 300 K	Krupp and Reggelin, 2012
PALF316	L-phenylalanine derived polyacetylene at 316 K	Krupp and Reggelin, 2012
PL1	Poly-L-1/poly(phenylisocyanide)	Reller et al., 2017
PALV	L-valine derived polyacetylene	Meyer et al., 2012
PADV	D-valine derived polyacetylene	Meyer et al., 2012
PPEMG	poly(N-methyl-N'-((R)-1-phenylethyl)guanidine)	Arnold et al., 2010

140

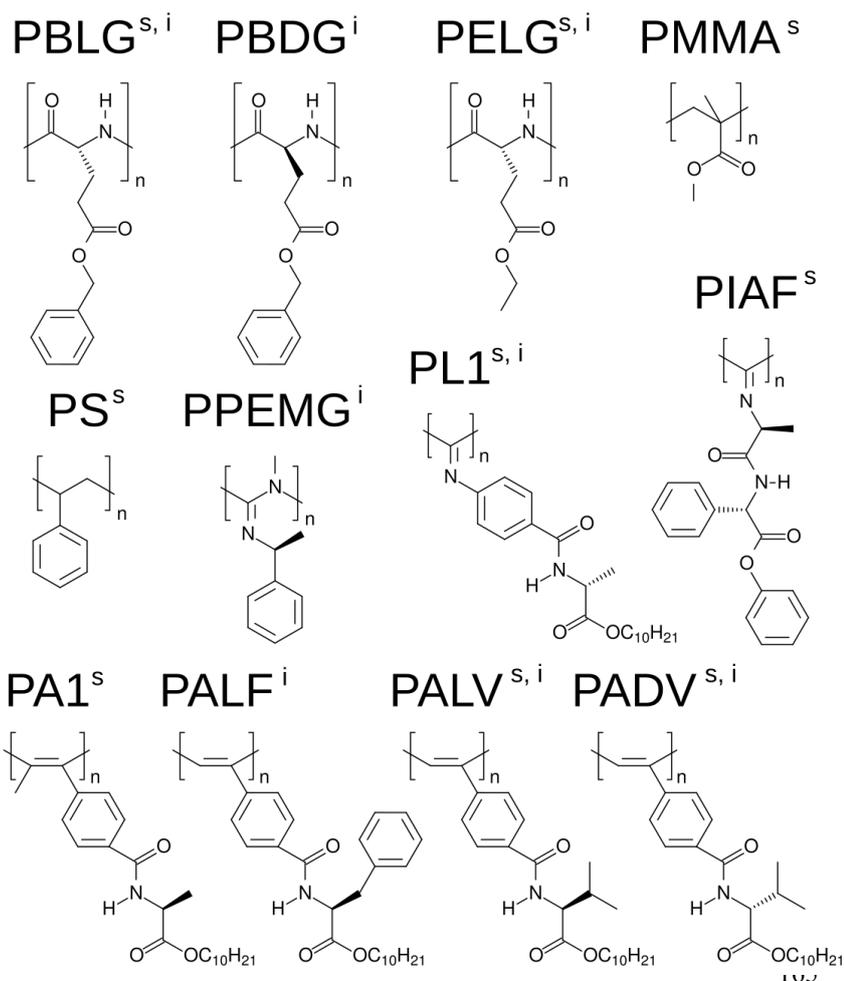



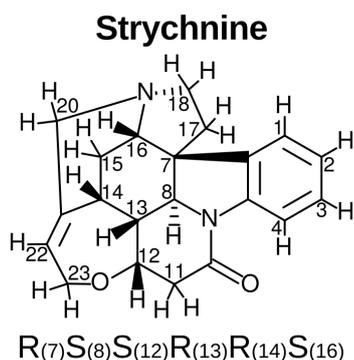
Figure 1: Representation of the basic units of the anisotropic media used for the alignment of strychnine (s) and (-)-IPC (i).

From the 12 selected alignment media, ten form LLC phases. PMMA and PS were used as compressed and stretched gels, respectively. The bias towards LLC phases likely arises, because a major objective of the current development of new alignment media is on enantiodifferentiation where helical chiral nonracemic polymers, capable of forming LLC phases, are good candidates. The LLC-forming polymers include polyglutamates, polyisocyanates, polyacetylenes, polyisocyanides and polyguanidines.

The experimental one-bond CH RDCs observed for strychnine and (-)-IPC in these alignment media were compared both among each other and with the RDCs calculated by the P3D simulation using the PBLG model (Figs. 2 and 3). In the case of strychnine, half of the alignment media besides PBLG have a good correlation with the RDCs simulated by P3D on the basis of the PBLG alignment model (Fig. 3a). These alignment media are PELG, PMMA, PIAF and PS. The experimental RDCs in

the second group of alignment media (PA1, PL1, PALV, PADV) also largely correlate among each other, but deviate more from the P3D-calculated RDCs (Fig. 3a). This is confirmed by inspection of the alignment tensors (Fig. 3b): the orientation of the z axis of the P3D-calculated alignment tensor is similar to those derived by singular-value decomposition (SVD) from the experimental RDCs in many of the alignment media, with the exception of PA1 and PL1. Notably, a smaller amount of RDCs were reported for PA1 and PL1 (five and six RDCs, respectively), which makes SVD-derived tensor orientations sensitive to the exact CH bond orientations in the employed structural models (Zweckstetter and Bax, 2002). We also point out that in some cases alignment tensor axes were swapped, e.g. the y-axis is positioned where in other alignment media the z axis is found (Fig. 3b). This can occur when two consecutives axes/eigenvalues have similar magnitude such that inaccuracies in experimental RDCs or molecular alignment simulation result in an exchange/relabeling of these axes, which however has only little influence on the back-calculated RDCs.

Strychnine

$R_{(7)}S_{(8)}S_{(12)}R_{(13)}R_{(14)}S_{(16)}$

	RDCs (Hz)									
	P3D PBLG	PBLG	PELG	PMMA	PIAF	PS	PA1	PL1	PALV	PADV
CH1	-114.9	-178.9	-43.5	37.0	19.6	-9.3	8.5	35.0	27.9	19.9
CH2	-118.1	-36.9	-15.3	22.8	-3.3	0.8		-0.5	-14.9	-11.7
CH3	-185.5	-69.8	-12.8	15.7	5.6		-16.6	-18.4	-11.8	-5.2
CH4	-115.7	-182.4	-43.5	37.8	21.2	-10.5	10.6	39.5	28.0	21.5
CH8	84.5	87.1	24.6	-17.5	-11.0	3.6		-5.7		
CH12	200.1	146.8	32.4	-40.0	-10.0	11.4			-4.1	-2.6
CH13	15.6	48.9	12.2	-11.0	-5.7	-3.6	3.6		-5.8	-4.8
CH14	184.1	151.7	31.5	-38.7	-11.5	12.0			-8.4	-5.4
CH16	-101.1	-49.6	-14.8	25.1	-1.9	1.8		-6.4		
CH22	-34.1	-1.1	3.0	-2.8	2.5	-6.6	6.0		1.0	0.0

(-)-IPC

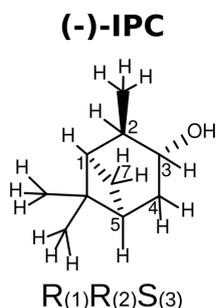
$R_{(1)}R_{(2)}S_{(3)}$

	RDCs (Hz)									
	P3D PBLG	PBLG	PBDG	PELG	PALF300	PALF316	PL1	PALV	PADV	PPEMG
CH1	17.8	13.3	14.5	9.2	6.9	2.7	9.8	-23.6	7.4	9.0
CH2	2.9	-7.4	-5.2	-12.1	5.0	4.7	34.2	13.8	1.0	12.0
CH3	3.7	11.1	13.1	2.9	6.0	-2.7	21.6	11.6	5.3	64.0
CH5	0.7	-5.5	-5.3	-7.3	-10.2	0.1	-25.2	-10.3	-14.0	52.0
$CH_2$4	1.1	-4.0	-4.4	-3.4	-7.6	-1.3	-19.7	4.6	-2.9	-13.0
	10.8	10.9	14.3	-2.4	19.0	5.5	56.2	-3.8	10.0	40.0
$CH_2$7	-27.5	-16.4	-17.4	-9.8	-4.8	2.2	-0.5	15.1	-1.9	
	-1.5	4.1	1.9	13.6	1.2	-5.5	-4.7	9.5	6.9	

Figure 2: Strychnine and (-)-IPC structures together with the CH carbon labels and the correct configuration (left), as well as the respective lists of RDCs in the different alignment media (right).

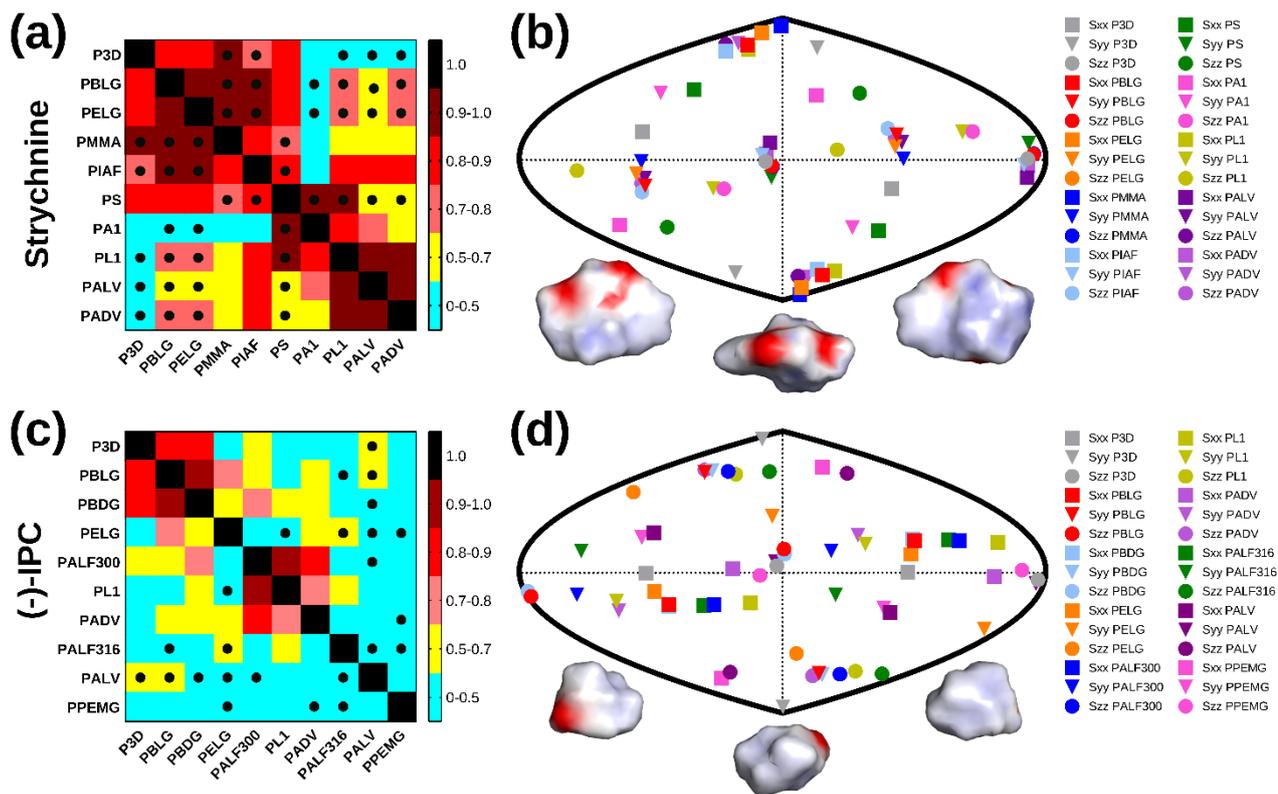



Figure 3: RDCs and alignment comparison of different alignment media and P3D for strychnine and (-)-IPC. (a, c) Matrices of Pearson's correlation R between P3D-calculated RDCs and the experimental RDCs in different alignment media for strychnine (a) and (-)-IPC (c). Dots mark negative correlations. (b, d) Comparison of the orientation of the P3D-predicted alignment tensor (grey) with alignment tensors derived by SVD from experimental RDCs of the different alignment media for strychnine (b) and (-)-IPC (d). The orientation of the three axes corresponding to the eigenvalues Szz (circle), Syy (triangle) and Sxx (square) of the diagonalized alignment tensor are projected onto a two-dimensional world map. Different orientations of the charged surface of strychnine (b) and (-)-IPC (d) are shown below.

The alignment media that correlate better with the P3D prediction were further analyzed in Figure 4. Comparison of P3D-predicted and experimental RDCs (Fig. 4a), as well as the alignment tensors projected onto a two-dimensional world map (Fig. 4c) and the aligned strychnine structures (Fig. 4d), show that the partial ordering of strychnine in these alignment media are very similar and well predicted by P3D. Experimental RDCs observed in PMMA have a very good correlation with the P3D-calculated RDCs (Fig. 3a, 4a) and also a very similar alignment tensor (Fig. 4c), but the correlation is negative (Fig 3a, dots). The negative slope indicates that the major alignment axis in PMMA is oriented orthogonal to the field, while PBLG aligns with its helix axis parallel to the magnetic field. Indeed, the PMMA gel was compressed, while the stretched PS gel displayed

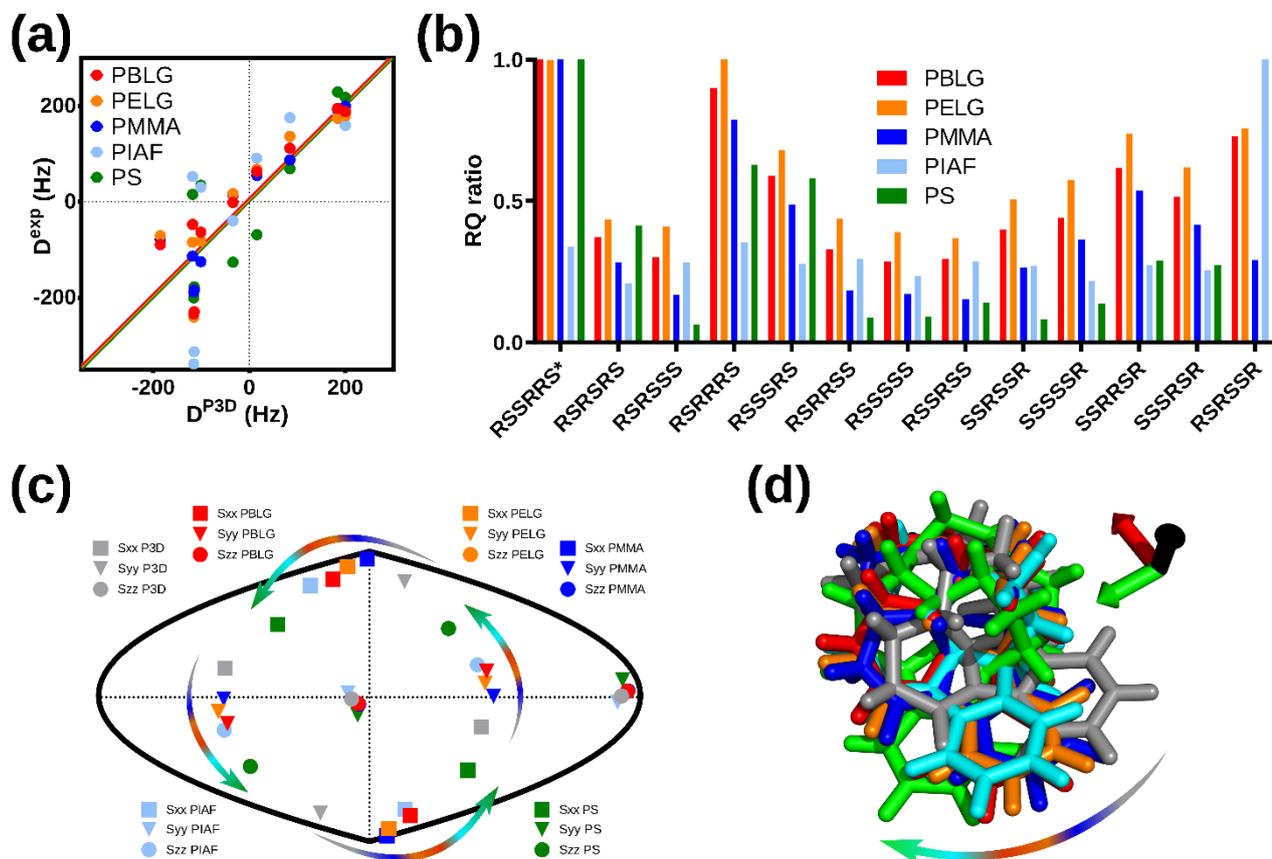
a positive correlation with the P3D-calculated RDCs. In other words, strychnine has in PMMA a highly similar alignment tensor as in PBLG/PS but with an opposite sign of the axial component of the alignment (D_a). For these reasons, when different alignment media are being compared, we use here the absolute value of the Pearson's correlation coefficient R.

250

Figure 4: P3D and the alignment media with most similar alignment properties for strychnine. (a) Correlation between P3D-simulated RDCs (D^{P3D}) and experimental RDCs (D^{exp}) with different alignment media for strychnine. Experimental values are normalized by the slope of the linear fitting. (b) Diastereomer discrimination power of different alignment media for strychnine based on the P3D simulation and using the RQ ratio for judging the quality of correlation. (c) Comparison of the orientation of the P3D-predicted alignment tensor (grey) of strychnine with alignment tensors derived by SVD from the experimental RDCs using PALES (Zweckstetter, 2008) observed in different alignment media. The orientation of the three axes corresponding to the eigenvalues Szz (circle), Syy (triangle) and Sxx (square) of the diagonalized alignment tensor are projected onto a two-dimensional world map. (d) Oriented structures of strychnine according to the PBLG-based P3D simulation and the different experimentally analyzed alignment media. Axes colors are black (z), red (x) and green (y), and are rotated when the axes of the alignment tensor are swapped because of a similar magnitude of the corresponding eigenvalues. The arrows in (c) and (d) illustrate the change in the orientation of the x and y axes in different alignment media.

We then investigated the diastereomer discrimination power of P3D-PBLG when using experimental RDCs observed in strychnine dissolved in the five different alignment media (Fig. 4b). To enhance the discrimination power we use the quality parameter RQ. The ability of P3D-PBLG to select the correct diastereomer is retained for alignment of strychnine in PELG, PMMA and PS. This is not the case for PIAF, which has a smaller Pearson's R (0.71) when compared to PELG/PMMA/PS (all over 0.8) (Fig. 3a). This suggests that R values larger than 0.8 are needed to identify the correct diastereomer, in agreement with the previously published P3D-based diastereomer discrimination analysis for six different small molecules (Ibáñez de Opakua et al., 2020).

Next, we performed the same P3D-based analysis for (-)-IPC, which has different alignment properties when compared to strychnine. The correlation matrix for (-)-IPC only shows a strong correlation of P3D with PBLG and PBDG (Fig. 3c). In agreement with a weaker enantiodiscrimination power of PBLG when compared to PELG (Hansmann et al., 2016), or the other helical chiral nonracemic polymers shown here, the experimental RDCs observed for (-)-IPC in PBLG and PELG differ. On the other hand, PBLG and PBDG induce similar alignment such that the discrimination of different diastereomers of (-)-IPC was retained for PBDG. Notably, a change in the enantiomer of the alignment medium (e. g. from PBLG to PBDG) has the same effect as changing the enantiomer of the solute (e.g. from (-)-IPC to (+)-IPC) (Marx et al., 2009).

As far as correlations between experimental RDCs in different alignment media are concerned, only few media induce similar alignment of (-)-IPC (Fig. 3c). Only PL1, PALF300 and PADV form a small cluster in the correlation matrix. A correlation between experimental RDCs induced by PADV and PL1 was present for both (-)-IPC and strychnine with R values of 0.71 and 0.92, respectively. The pronounced differences in the alignment of (-)-IPC in different alignment media is also evident from the comparison of the respective alignment tensors: the projected axes orientations do not cluster in certain regions (Fig. 3d), in contrast to the alignment tensors of strychnine (Fig. 3b). Further notable are the RDC differences when (-)-IPC is aligned in PALF300 and PALF316 (R=0.45). This L-phenylalanine derived polyacetylene forms different LLC phases at different temperatures, with the helical structure severely disrupted at 316 K (Krupp and Reggelin, 2012). The pronounced differences in the alignment of (-)-IPC in PALF300 and PALF316 indicates that for certain molecules/alignment media fine structural details of the alignment medium are critical for enantiodiscrimination.

In order to rationalize the distinct alignment properties of strychnine and IPC, we analyzed the structural properties of the two molecules (Fig. 3b,d). While strychnine has an oval disc-like shape, the shape of IPC is quite spherical, and both molecules have asymmetric charge distributions (Fig. 3b,d). Comparison of the correlation coefficients of the experimental RDCs with RDCs predicted by molecular alignment simulation using P3D or only steric interactions (1D obstruction model; Zweckstetter and Bax, 2000) suggested that electrostatic interactions are more important for the alignment of strychnine: R values dropped from 0.88 to 0.64 in the case of strychnine and from 0.84 to 0.68 in the case of (-)-IPC, when replacing P3D simulations by 1D obstruction model simulations.

The excellent correlation between P3D-predicted and experimental RDCs of strychnine in PBLG indicates that the alignment of strychnine in PBLG is dominated by steric and electrostatic factors (Fig. 3, 4). At the same time, the fine structural details of the alignment media appear to be less important, which results in similar alignment of strychnine in PBLG, PELG, PMMA and PS (Fig. 4). On the other hand, the quite spherical shape of IPC suggests that steric obstruction is less important for its molecular alignment. Instead, specific molecular interactions between IPC and the alignment medium become relevant and are responsible for the differences in RDC values observed in different alignment media. The difference in the alignment of IPC in PBLG and PELG might be correlated with the stronger enantiodifferentiating power of PELG, which has been linked to the change in the bulkiness and mobility of the lateral side chain (Hansmann et al., 2016). Due to these differences, IPC can have more and stronger diastereomorphous interactions with the chiral helical backbone of PELG. The importance of fine structural details for the alignment of IPC also provides a rationale for why IPC is an excellent test molecule to study the enantiodifferentiation properties of alignment media.

4 Analysis of conformational ensembles using P3D

Because strychnine and IPC are rigid molecules, a single alignment tensor accurately describes their weak LLC/gel-induced alignment. However, for more flexible molecules it is necessary to determine alignment tensors for all the conformers or independently for every flexible part of the molecule (Thiele and Berger, 2003). In order to simplify this problem, linearly independent alignment media would be needed (Ramirez and Bax, 1998), which, as shown in Fig. 3, is not always easy to achieve. When only one alignment medium is available, selection of energetically more promising structures and back-calculation of anisotropic NMR parameters to best-fit experimental values might be used (Tzvetkova et al., 2019). The latter approach, however, requires a large amount of anisotropic NMR parameters and becomes difficult when the alignment tensors of the conformers are different.

We previously developed the P3D alignment simulation to solve the relative configuration problem, demonstrating that P3D can identify the correct diastereoisomer from a very small number of RDCs, even with less than five RDCs, the minimum number of RDCs required for SVD. Here we now use P3D to address the problem of conformation. To this end, we selected sucrose (Fig. 5a), which has recently been analyzed by anisotropic NMR in PBLG (Ndukwe et al., 2019). Studies based on molecular dynamics (MD) simulations and solution NMR suggested the presence of multiple sucrose conformations (Delaglio et al., 2005; Xia and Case, 2012). On the basis of 11 RDCs and 12 RCSAs, the conformational ensemble of sucrose in CDCl$_3$/DMSO (70:30) was best described with three conformers, which were selected from a set of low energy DFT structures (Ndukwe et al., 2019). The respective ΔG of conformers 1, 2 and 3 were 0, 2.04 and 2.72 kcal/mol, with conformer 3 being highly similar to the crystal structure of sucrose (Russo et al., 2013).

Instead of the 23 anisotropic NMR parameters used by Ndukwe and colleagues, we use here only the eight one-bond CH RDCs

330 (Fig. 5a,b). Following the same rationale as before (Ibáñez de Opakua et al., 2020), we selected the one-bond CH RDCs (Fig. 5b) because they are the largest RDCs in small molecules, i.e. can be measured with high accuracy, and there is less ambiguity in the assignment. We then subjected the three conformers of sucrose to P3D alignment simulation. P3D-simulated RDCs were averaged over the three-member ensemble and compared with the experimental RDCs (Fig. 5b,c). The results indicate that the three-member ensemble of conformers improved the correlation reaching a R value of 0.975 (Fig. 5c). The RQ of the weighted

335 average is also significantly larger than the RQ values for any of the three individual conformers (Fig. 5d).

(a) Sucrose structure with CH carbons labeled (10, 11, 9, 5, 1, 2, 3, 4)

(b)

	RDCs (Hz)							
	CH1	CH2	CH3	CH4	CH5	CH9	CH10	CH11
Conformer 1 (33%)	34.53	34.65	36.78	34.34	43.22	-13.40	-11.59	-16.83
Conformer 2 (35%)	53.20	48.10	51.00	49.61	42.23	-25.31	52.39	-8.60
Conformer 3 (32%)	-0.53	1.27	0.60	-1.66	59.48	-27.44	-10.79	-19.65
Weighted average	29.85	28.68	30.18	28.17	48.08	-22.06	11.06	-14.85
Experimental	22.23	17.53	23.32	21.3	52.25	-19.79	3.12	-17.74

(c) Weighted average R=0.975 Q_S=0.266; Conformer 1 R=0.899 Q_S=0.465; Conformer 2 R=0.760 Q_S=1.079; Conformer 3 R=0.906 Q_S=0.638

(d) RQ ratio

(e) Conformer 1 Sxx, Conformer 1 Syy, Conformer 1 Szz, Conformer 2 Sxx, Conformer 2 Syy, Conformer 2 Szz, Conformer 3 Sxx, Conformer 3 Syy, Conformer 3 Szz

Figure 5: Validation of the conformational ensemble of sucrose using P3D. (a) Sucrose structure with the CH carbons labeled. **(b)** List of P3D-simulated RDCs for the 3 conformers together with the average weighted RDCs (from Ndukwe et al. populations) and the experimental RDCs. **(c)** Correlations between the experimental RDCs (D^{exp}) and the P3D-simulated RDCs for the 3 conformers and the weighted average. **(d)** RQ ratios of the 3 different conformers in reference to the weighted average. Error bars are calculated as the propagation of R and Q_S errors and these errors are calculated from the std of 100 repetitions including noise in the RDCs.

(e) Comparison of the orientation of the P3D-predicted alignment tensors for the 3 conformers. The orientation of the three axes corresponding to the eigenvalues Szz (circle), Syy (triangle) and Sxx (square) of the diagonalized alignment tensor are projected
345 **onto a two-dimensional world map.**

The structures of the three conformers were aligned before the simulation in order to have the same molecular frame and be able to compare the alignment tensors (Fig. 5e). The result shows that all the conformers have a similar but not identical alignment, indicating that the differences in RDCs come mainly from the structural differences, in agreement with the use of

350 the variable-weight single-tensor SVD method to solve the conformational structure of the molecule (Ndukwe et al., 2019).

While the variable-weight single-tensor SVD method strongly relies on the assumption that different conformers have similar alignment tensors, this is not required for the P3D-based conformational analysis. We therefore determined the relative amounts of conformers by maximizing the P3D-based RQ parameter (Fig. 6a,b). The comparison of the RQ ratios of the RQ

355 maximized 3-conformer ensemble and the RQ maximized 2-conformer ensembles (Fig. 6c) shows that three is the minimum number of conformers to get an almost perfect fit (R=0.996; Q_S=0.076). In addition, the contribution of conformer 3 was increased to 49 % (32% in (Ndukwe et al., 2019)) in the refined 3-conformer ensemble (Fig. 6a,b). Notably, the most populated conformer (conformer 3) is closest to the crystal structure of sucrose.

360

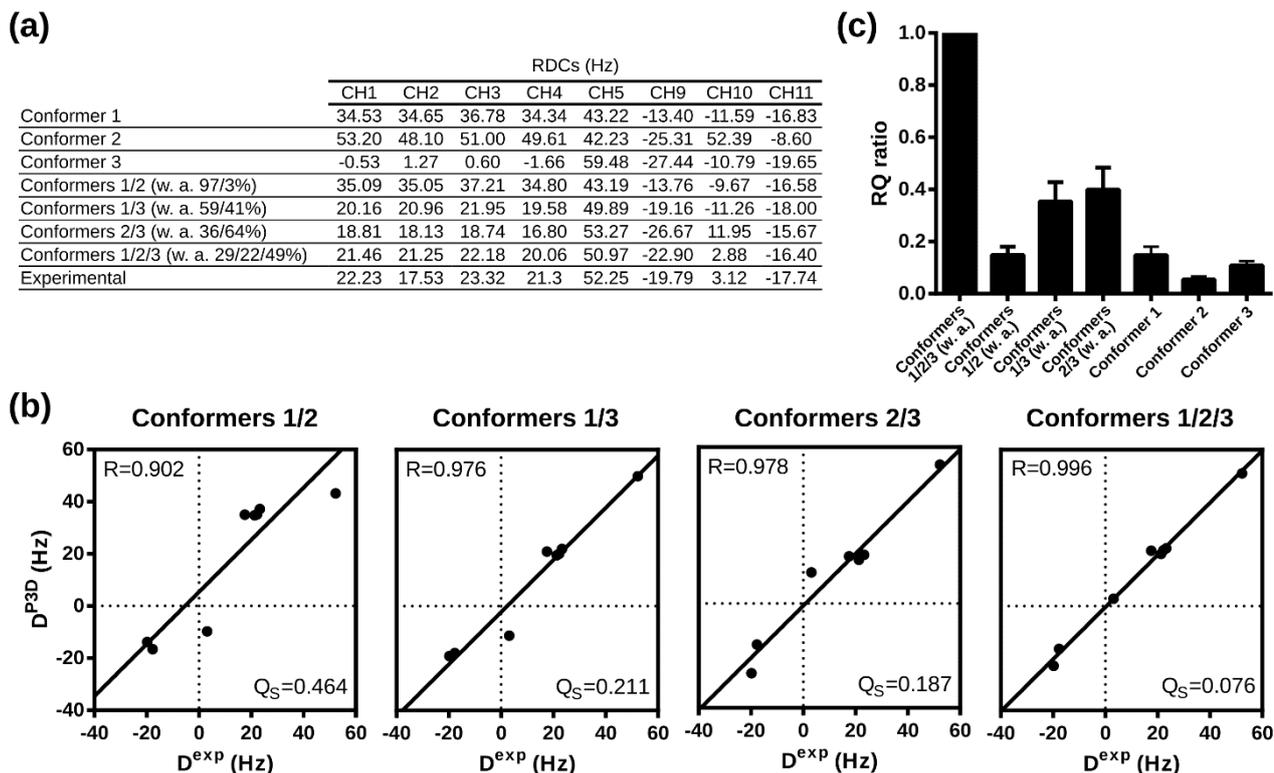



Figure 6: P3D-based refinement of the conformational ensemble of sucrose. (a) List of P3D-calculated RDCs for the 3 conformers together with the average weighted RDCs from the RQ maximization for 2- and 3-conformer ensembles and the experimental RDCs. (b) Correlation between the experimental RDCs. When all RDCs are considered we get R=0.926 for the average with 0.806, 0.615 and 0.838 respectively for each conformer. The obtained populations (26/20/54%) are very similar to the ones obtained with the one-bond CH RDCs only. (D^{exp}) and the P3D-calculated RDCs for the weighted average of the RQ maximization for 2- and 3-conformer ensembles. (c) RQ ratios of the three different conformers in reference to their weighted average as 2- and 3-conformer ensembles. Error bars were calculated as the propagation of R and Q_S errors and these errors are calculated from the std of 100 repetitions including noise in the RDCs.

5 Conclusions

The current study highlights the important role of molecular alignment simulations for the structural analysis of small molecules. In agreement with previous data (Ibáñez de Opakua et al., 2020), the new analysis supports the applicability of P3D simulations for the determination of the relative configurations, but also extends it to the analysis of conformational ensembles of small molecules. In addition, molecular alignment simulations might – with further improvements – become crucial for the determination of the absolute configuration. While much progress has been made in the development of powerful chiral alignment media, this is restricted to the differentiation between enantiomers, similar to exposing small molecules to polarized

light. To determine the absolute configuration, atomistic descriptions are required that link the NMR anisotropic parameters obtained from chiral alignment media with the correct enantiomer. A next step towards this goal could be the inclusion of specific interactions between the solute and the alignment medium, for example salt bridges, into molecular alignment simulations and the consideration of thermodynamic and kinetic contributions.

Towards this next step, it is important to define an alignment medium, which has a structure amenable to structural modeling and strong enantiodiscrimination capabilities. We therefore analyzed different alignment media and compared them with our P3D alignment simulation model of PBLG. Literature search identified only two molecules, strychnine and IPC, for which RDCs in several different alignment media had been reported. The results of our analysis suggested that the weak alignment induced by LLC phases and gels critically depends on both the molecular properties of the alignment medium and the small molecule. The comparison further showed that the alignment of IPC varies more strongly across the available alignment media when compared to strychnine (Fig. 3). We interpret this as a consequence of the more symmetrical/spherical shape of IPC such that more specific interactions with the alignment medium more strongly contribute to the alignment process. In the case of strychnine, on the other hand, steric obstruction together with electrostatic interactions dominate molecular ordering. This leads to less variability in the partial alignment of the small molecule allowing the application of the P3D-PBLG model to other alignment media (Fig. 4). The choice of alignment medium should therefore take into account the structural properties of the small molecule of interest, especially its shape as well as the charge distribution.

We also investigated the applicability of the P3D-PBLG simulation approach to the challenge of determining conformational ensembles of flexible small molecules. On the example of sucrose, we showed that P3D can be used to determine the populations of different conformers in an ensemble, with the advantage that it can work even when individual conformers have different alignment tensors. Through the P3D-based analysis we optimized the population of each conformer in the ensemble. The analysis resulted in an ensemble in which the population of conformer 3, which is closest to the crystal structure of sucrose, was increased to almost 50 % (Fig. 6). The lower population of conformer 3 in the SVD-based ensemble might arise from the slightly different alignment tensors of the three conformers (Fig. 5e).

In summary, P3D alignment simulations establish a quantitative connection between the alignment medium, the molecular structure of small molecules and anisotropy-based NMR parameters. P3D can therefore predict RDCs for different alignment media depending on the structural details of both the alignment medium and the small molecule and thus enables the determination of conformational ensembles of flexible small molecules.

Data availability. The Linux version of the PALES software with the P3D algorithm used in this work can be downloaded from the PALES webpage (https://www3.mpibpc.mpg.de/groups/zweckstetter/_links/software_pales.htm). All other data that support the findings of this study are available from the corresponding authors upon reasonable request.

Author contribution. AIdeO and MZ designed the project and wrote the paper. AIdeO conducted the data analysis.

415 **Competing interests.** The authors declare that they have no conflict of interest.

Financial support. M.Z. was supported by the Helmholtz society through the Impuls- und Vernetzungsfond 'Multiscale Bioimaging: From Molecular Machines to Networks of Excitable Cells', and by the advanced grant '787679 - LLPS-NMR' of the European Research Council.

420 **References**

Anet, F. A. L. and Bourn, A. J. R.: Nuclear magnetic resonance spectral assignments from nuclear Overhauser effects. J. Am. Chem. Soc., 87, 5250-5251, doi:10.1021/ja00950a048, 1965.

Arnold, L, Marx, A., Thiele, C. M. and Reggelin, M.: Polyguanidines as chiral orienting media for organic compounds. Chemistry, 16, 10342-6, doi:10.1002/chem.201000940, 2010.

425 Bax, A.: Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci., 12, 1-16, doi: 10.1110/ps.0233303, 2003.

Berger, R., Courtieu, J., Gil R. R., Griesinger, C., Köck, M., Lesot, P., Luy, B., Merlet, D., Navarro-Vázquez, A., Reggelin, M., Reinscheid, U. M., Thiele, C. M., and Zweckstetter, M.: Is enantiomer assignment possible by NMR spectroscopy using residual dipolar couplings from chiral nonracemic alignment media?--A critical assessment. Angew. Chem. Int. Ed. Engl., 51,
430 8388-91, doi:10.1002/anie.201107626, 2012.

Bifulco, G., Riccio, R., Martin, G.E., Buevich, A.V. and Williamson, R.T.: Quantum chemical calculations of 1J(CC) coupling constants for the stereochemical determination of organic compounds. Org Lett 15, 654-7, doi:10.1021/ol3034803, 2013.

Böttcher, B. and Thiele, C. M.: Determining the Stereochemistry of Molecules from Residual Dipolar Couplings (RDCs). eMagRes, 1, 169-180, doi:10.1002/9780470034590.emrstm1194, 2012.

435 Canales, A., Jiménez-Barbero, J. and Martín-Pastor, M.: Review: Use of residual dipolar couplings to determine the structure of carbohydrates. Magn. Reson. Chem., 50, S80-S85, doi:10.1002/mrc.3888, 2012.

Canet, I., Courtieu, J., Loewenstein, A., Meddour, A. and Pechine, J.M.: Enantiomeric analysis in a polypeptide lyotropic liquid crystal by deuterium NMR. J. Am. Chem. Soc., 117, 6520-6526, doi: 10.1021/ja00129a015, 1995.

Dama, M. and Berger, S.: Polyisocyanides as a new alignment medium to measure residual dipolar couplings for small organic
440 molecules. Org. Lett., 14, 241-3, doi:10.1021/ol202547y, 2012a.

Dama, M. and Berger, S.: Polyacetylenes as a new alignment medium to measure residual dipolar couplings for chiral organic molecules. Tetrahedron Letters, 53, 6439-42, doi:10.1016/j.tetlet.2012.09.060, 2012b.

Doty, P., Holtzer, A. M., Bradbury, J. H. and Blout, E. R.: Polypeptides. II. The Configuration of Polymers of Gamma-Benzyl-L-Glutamate in Solution. J. Am. Chem. Soc., 76, 4493-4494, doi: 10.1021/ja01646a079, 1954.

445 Haasnoot, C. A. G., de Leeuw, F. A. A. M. and Altona, C.: The relationship between proton-proton NMR coupling constants and substituent electronegativities—I: An empirical generalization of the karplus equation. Tetrahedron, 36, 2783-2792, doi:10.1016/0040-4020(80)80155-4, 1980.

Hansmann, S., Larem, T. and Thiele, C.M.: Enantiodifferentiating Properties of the Alignment Media PELG and PBLG – A Comparison. European Journal of Organic Chemistry, 7, 1324-9, doi:10.1002/ejoc.201501410, 2016.

450 Ibáñez de Opakua, A., Klama, F, Ndukwe, I. E., Martin, G. E., Williamson, R. T. and Zweckstetter, M.: Determination of Complex Small-Molecule Structures Using Molecular Alignment Simulation. Angew. Chem. Int. Ed. Engl., 59, 6172-6176, doi:10.1002/anie.202000311, 2020.

Ionescu, CM., Sehnal, D., Falginella, F.L., Pant, P., Pravda, L., Bouchal, T., Vařeková, R. S., Geidl, S. and Koča, J.: AtomicChargeCalculator: interactive web-based calculation of atomic charges in large biomolecular complexes and drug-like 455 molecules. J. Cheminform. 7, doi:10.1186/s13321-015-0099-x, 2015.

Krupp, A. and Reggelin, M.: Phenylalanine-based polyarylacetylenes as enantiomer-differentiating alignment media. Magn. Reson. Chem., 50, S45-52, doi:10.1002/mrc.3894, 2012.

Krupp, A., Noll, M. and Reggelin, M.: Valine derived poly (acetylenes) as versatile chiral lyotropic liquid crystalline alignment media for RDC-based structure elucidations. Magn. Reson. Chem., 1-10, doi: 10.1002/mrc.5003, 2020.

460 Lesot, P., Merlet, D., Meddour, A., Courtieu, J. and Loewenstein, A.: Visualization of Enantiomers in a Polypeptide Liquid-Crystal Solvent through C-13 Nmr-Spectroscopy. J. Chem. Soc., Faraday Trans., 91, 1371-1375, doi: 10.1039/ft9959101371, 1995.

Lesot, P., Merlet, D., Courtieu, J. and Emsley, J. W.: Discrimination and analysis of the NMR spectra of enantiomers dissolved in chiral liquid crystal solvents through 2D correlation experiments. Liq. Cryst., 21, 427-435, 10.1080/02678299608032851, 465 1996a.

Lesot, P., Gounelle, Y., Merlet, D., Loewenstein, A. and Courtieu, J.: Measurement and Analysis of the Molecular Ordering Tensors of 2 Enantiomers Oriented in a Polypeptide Liquid-Crystalline System. J. Phys. Chem. 99, 14871-14875, doi: 10.1021/jp961819v, 1996b.

Lesot, P., Berdague, P., Meddour, A., Kreiter, A., Noll, M. and Reggelin, M. H-2 and C-13 NMR-Based Enantiodetection 470 Using C4 Polyacetylene versus Polypeptide Aligning Media: Versatile and Complementary Tools for Chemists. Chempluschem, 84, 144-153, doi: 10.1002/cplu.201800493, 2019.

Leyendecker, M., Meyer, N. C. and Thiele, C. M.: Development of New Supramolecular Lyotropic Liquid Crystals and Their Application as Alignment Media for Organic Compounds. Angew. Chem. Int. Ed., 56, 11471-11474. doi: 10.1002/anie.201705642, 2017.

475 Li, G. W., Cao, J. M., Zong, W., Hu, L., Hu, M. L., Lei, X., Sun, H. and Tan, R. X.: Helical Polyisocyanopeptides as Lyotropic Liquid Crystals for Measuring Residual Dipolar Couplings. Chemistry, 23, 7653-6, doi:10.1002/chem.201700539, 2017.

Li, G. W., Liu, H., Qiu, F., Wang, X. J. and Lei, X. X.: Residual Dipolar Couplings in Structure Determination of Natural Products. Review Nat. Prod. Bioprospect., 8, 279-295, doi:10.1007/s13659-018-0174-x, 2018.

Liu, Y., Cohen, R. D., Martin, G. E. and Williamson, R. T.: A practical strategy for the accurate measurement of residual dipolar couplings in strongly aligned small molecules. J. Magn. Reson. 291, 63-72, doi:10.1016/j.jmr.2018.04.002, 2018.

Lorieau, J., Yao, L. and Bax, A.: Liquid Crystalline Phase of G-Tetrad DNA for NMR Study of Detergent-Solubilized Proteins. J. Am. Chem. Soc., 130, 7536-7537, doi:10.1021/ja801729f, 2008.

Luy, B., Kobzar, K. and Kessler, H.: An easy and scalable method for the partial alignment of organic molecules for measuring residual dipolar couplings. Angew. Chem. Int. Ed. Engl., 43, 1092-4, doi:10.1002/anie.200352860, 2004.

Luy, B. and Kessler, H.: Partial Alignment for Structure Determination of Organic Molecules, chapt. 138 (Springer International Publishing, Modern Magnetic Resonance), 1279–1285, doi:10.1007/1-4020-3910-7_138, 2006.

Nath, N., d'Auvergne, E. J. and Griesinger, C.: Long-Range Residual Dipolar Couplings: A Tool for Determining the Configuration of Small Molecules. Angew. Chem. Int. Ed. Engl., 54, 12706-10, doi:10.1002/anie.201504432, 2015.

Marx, A. and Thiele, C. M.: Orientational properties of PBLG: Influence of molecular weight and solvent on order parameters of the solute. Chem. Eur. J., 15, 254-260, doi:10.1002/chem.200801147, 2009.

Marx, A., Schmidts, V. and Thiele, C. M.: How different are diastereomorphous orientations of enantiomers in the liquid crystalline phases of PBLG and PBDG: a case study. Magn. Reson. Chem., 47, 734-40, doi:10.1002/mrc.2454, 2009.

Meyer, N.C., Krupp, A., Schmidts, V., Thiele, C.M. and Reggelin, M.: Polyacetylenes as enantiodifferentiating alignment media. Angew. Chem. Int. Ed. Engl., 51, 8334-8, doi:10.1002/anie.201201891, 2012.

Ndukwe, I. E., Wang, X., Pelczer, I., Reibarkh, M., Williamson, R. T., Liu, Y. and Martin, G. E.: PBLG as a Versatile Liquid Crystalline Medium for Anisotropic NMR Data Acquisition. Chem. Comm., 30, doi:10.1039/C9CC01130G, 2019.

Ramirez, B. E. and Bax, A.: Modulation of the Alignment Tensor of Macromolecules Dissolved in a Dilute Liquid Crystalline Medium. J. Am. Chem. Soc., 120, 9106-9107, doi:10.1021/ja982310b, 1998.

Reller, M., Wesp, S., Koos, M. R. M., Reggelin, M. and Luy B.: Biphasic Liquid Crystal and the Simultaneous Measurement of Isotropic and Anisotropic Parameters by Spatially Resolved NMR Spectroscopy. Chemistry, 23, 13351-9, doi:10.1002/chem.201702126, 2017.

Russo, L., Grazulis, S. and Bagdziunas, G.: Structure of Sucrose. Personal communication to COD, http://www.crystallography.net/cod/3500015.html, 2013.

Sass, H. J., Musco, G., J. Stahl, S., T. Wingfield, P. and Grzesiek, S.: Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J. Biomol. NMR 18, 303-309. doi:10.1023/A:1026703605147, 2000.

Schmidts, V.: Perspectives in the application of residual dipolar couplings in the structure elucidation of weakly aligned small molecules. Magn. Reson. Chem., 55, 54-60, doi:10.1002/mrc.4543, 2017.

Thiele, C. M. and Berger, S.: Probing the Diastereotopicity of Methylene Protons in Strychnine Using Residual Dipolar Couplings. Org. Lett., 5, 705-708, doi:10.1021/ol0275163, 2003.

Thiele, C.M.: Simultaneous assignment of all diastereotopic protons in strychnine using RDCs: PELG as alignment medium for organic molecules. J. Org. Chem. 69, 7403-13, doi:10.1021/jo049867w, 2004.

Tzvetkova, P., Sternberg, U., Gloge T., Navarro-Vázquez, A. and Luy, B.: Configuration determination by residual dipolar couplings: accessing the full conformational space by molecular dynamics with tensorial constraints. Chem. Sci., 10, 8774-8791, doi:10.1039/C9SC01084J, 2019.

Tjandra, N. and Bax, A.: Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science, 278, 1111-4, doi: 10.1126/science.278.5340.1111, 1997.

Tycko, R., Blanco, F.J. and Ishii, Y.: Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole−Dipole Couplings in High-Resolution Biomolecular NMR. J. Am. Chem. Soc., 122, 38, 9340-9341. doi: 10.1021/ja002133q, 2000.

Venable, R. M., Delaglio, F., Norris, S. E. and Freedberg, D. I.: The utility of residual dipolar couplings in detecting motion in carbohydrates: application to sucrose. Carbohydr. Res., 340, 863-74, doi:10.1016/j.carres.2005.01.025, 2005.

Xia, J. and Case, D. A.: Sucrose in aqueous solution revisited. Biopolymers, 97, 276–288, doi:10.1002/bip.22017, 2012.

Zweckstetter, M. and Bax, A.: Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. J. Am. Chem. Soc., 122, 3791-3792, doi:10.1021/ja0000908, 2000.

Zweckstetter, M. and Bax, A.: Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. J. Biomol. NMR. 23, 127-137. doi:10.1023/A:1016316415261, 2002.

Zweckstetter, M.: NMR: prediction of molecular alignment from structure using the PALES software. Nat. Protoc., 3, 679-90, doi:10.1038/nprot.2008.36, 2008.