



Cite this: *Phys. Chem. Chem. Phys.*, 2024, 26, 2016

# Exploring the impact of alignment media on RDC analysis of phosphorus-containing compounds: a molecular docking approach†

Markéta Christou Tichotová, <sup>ab</sup> Lucie Tučková, <sup>a</sup> Hugo Kocek, <sup>a</sup> Aleš Růžička, <sup>c</sup> Michal Straka <sup>a</sup> and Eliška Procházková <sup>\*a</sup>

Residual dipolar couplings (RDCs) are employed in NMR analysis when conventional methods, such as *J*-couplings and nuclear Overhauser effects (NOEs) fail. Low-energy (optimized) conformers are often used as input structures in RDC analysis programs. However, these low-energy structures do not necessarily resemble conformations found in anisotropic environments due to interactions with the alignment medium, especially if the analyte molecules are flexible. Considering interactions with alignment media in RDC analysis, we developed and evaluated a molecular docking-based approach to generate more accurate conformer ensembles for compounds in the presence of the poly- $\gamma$ -benzyl-L-glutamate alignment medium. We designed chiral phosphorus-containing compounds that enabled us to utilize <sup>31</sup>P NMR parameters for the stereochemical analysis. Using P3D/PALES software to evaluate diastereomer discrimination, we found that our conformer ensembles outperform moderately the standard, low-energy conformers in RDC analysis. To further improve our results, we (i) averaged the experimental values of the molecular docking-based conformers by applying the Boltzmann distribution and (ii) optimized the structures through normal mode relaxation, thereby enhancing the Pearson correlation factor *R* and even diastereomer discrimination in some cases. Nevertheless, we presume that significant differences between *J*-couplings in isotropic and in anisotropic environments may preclude RDC measurements for flexible molecules. Therefore, generating conformer ensembles based on molecular docking enhances RDC analysis for mildly flexible systems while flexible molecules may require applying more advanced approaches, in particular approaches including dynamical effects.

Received 25th August 2023,  
 Accepted 12th December 2023

DOI: 10.1039/d3cp04099b

rsc.li/pccp

## Introduction

Conformation and configuration are key features of molecules, affecting their physical, chemical, and biological properties. Relative configuration can be determined by X-ray diffraction (XRD). However, XRD requires a single crystal, which is often difficult to obtain, particularly for biologically active compounds. Moreover, conformations in solid crystals do not necessarily match conformations in solution. In solution, molecular structures are typically analyzed by nuclear magnetic resonance (NMR) spectroscopy. Standard stereochemical analysis based on NMR

relies on vicinal <sup>1</sup>H–<sup>1</sup>H <sup>3</sup>*J*-couplings, which are correlated to dihedral angles in a molecule through the Karplus equation,<sup>1</sup> or nuclear Overhauser effect<sup>2,3</sup> (NOE). When both *J*-couplings and NOEs fail, molecular stereochemistry may be analyzed using residual dipolar couplings (RDCs).

Residual dipolar couplings can be detected when molecular tumbling is reduced in an anisotropic environment.<sup>4,5</sup> The anisotropic environment is induced in alignment media,<sup>6</sup> such as poly- $\beta$ -phenethylaspartates,<sup>7</sup> stretched polyvinyl acetate gels,<sup>8</sup> or poly- $\gamma$ -benzyl-L-glutamate<sup>9</sup> (PBLG). For structural determination, experimental RDC values are correlated with theoretical RDCs, which are back-calculated using alignment tensors of optimized molecular structures (conformers) of isomers studied.

Several program packages based on different principles are available to back-calculate theoretical RDCs (MSpin,<sup>10</sup> RDC@hotFCHT,<sup>11</sup> MDOC,<sup>12</sup> REDCAT,<sup>13</sup> and DipoCoup<sup>14</sup>). However, these approaches usually disregard specific analyte interactions with the alignment medium. To take these interactions into consideration, we must know the 3D structure of both interacting partners. However, to the best of our knowledge, the exact molecular structure of some

<sup>a</sup> Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 160 00 Prague, Czech Republic. E-mail: prochazkova@uochb.cas.cz

<sup>b</sup> Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, 116 28 Prague, Czech Republic

<sup>c</sup> Department of General and Inorganic Chemistry, Faculty of Chemical Technology, University of Pardubice, Pardubice 532 10, Czech Republic

† Electronic supplementary information (ESI) available. CCDC 2277412–2277415. For ESI and crystallographic data in CIF or other electronic format see DOI: <https://doi.org/10.1039/d3cp04099b>



alignment media remains beyond the reach of contemporary experimentation.

P3D/PALES<sup>15–17</sup> was the first method to consider the atomic structure of the alignment medium in determining RDCs of small molecules. P3D/PALES has been adapted to a model of PBLG, which has a well-defined helical structure in chloroform.<sup>18–20</sup> The simulations use molecular structures of low-energy conformers as inputs (henceforth referred to as the low-energy approach). However, utilizing this approach in our previous work did not lead to the desired diastereomer discrimination, especially for flexible molecules.<sup>21</sup> The potential energy surface (PES) of flexible molecules is usually shallow and may contain several minima (corresponding to conformers) within a few tenths of kcal mol<sup>-1</sup> from the global minimum (GM). Importantly, PES may be strongly affected by the environment, *i.e.*, by the presence of the alignment medium. Thus, low-energy structures provided for P3D/PALES simulations may significantly differ from those present in the experiment.

So far, several methods have been proposed for studying non-rigid molecules using RDCs. The most common RDC analysis relies on the alignment tensor fit; this has been adapted either as a single-alignment-tensor<sup>22,23</sup> or multiple-alignment-tensor<sup>24,25</sup> fitting approach to an optimised conformation ensemble. However, the idea of a single-alignment-tensor fitting to an ensemble of conformers is not generally applicable. The latter approach requires a large amount of anisotropic data, which are not always available. Another approach, usually used for macromolecules, employs the so-called reweight models<sup>26–29</sup> (maximum entropy, maximum parsimony, maximum allowed probability, *etc.*). However, these methods are computationally demanding and fail for more flexible systems. To overcome problems connected to the generation of structures, molecular-dynamics based MDOC<sup>12,30,31</sup> or multi-alignment-media based TITANIA<sup>32</sup> are an option. However, the main drawback of these methods is, again, the limited availability of larger amounts of experimental data.

The most widely used low-energy approach may not result in a correct back-calculation of theoretical RDCs of non-rigid molecules because a change in analyte conformation affects the RDC values. Based on this assumption, we raised the following research questions: (i) does the alignment medium significantly affect the analyte conformation? (ii) How large can the energy penalty paid by the analyte for the transition to the conformation constrained by the alignment medium be? We propose that conformer ensembles generated *via* molecular docking may provide a more accurate representation of the reality of the RDC experiment, consequently aiding in answering the posed questions.

Molecular docking is primarily used in drug design<sup>33–35</sup> to predict the binding pose (pose = conformation, position, and orientation) of a docked substrate (usually a small molecule = ligand) within a protein binding site. Nevertheless, by treating an analyte as the ligand and an alignment medium as the binding site, molecular docking may generate a more accurate conformer ensemble than the standard low-energy approach. Accordingly, we can consider the low-energy approach to generate conformers in isotropic solution and the molecular

docking procedure to generate conformers in the presence of the alignment medium in RDC experiments.

In this study, we developed and evaluated a molecular docking-based approach to generate more accurate conformer ensembles for compounds in the poly- $\gamma$ -benzyl-L-glutamate (PBLG)<sup>9,36–40</sup> alignment medium environment. The model compounds used in this work feature a chiral phosphorus atom. We had shown in our previous work<sup>21</sup> that <sup>31</sup>P NMR parameters provide structural insights that aid the stereochemical analysis of phosphorus-containing compounds. We presume that the conformer ensembles generated *via* docking may provide a better approximation to the reality of RDC experiments than the conventional low-energy approach. Conformers generated by docking were used as inputs for RDC analysis. With the obtained back-calculated (theoretical) RDC datasets, we created Boltzmann distribution-weighted average RDC datasets which were then correlated with experimental RDC values. In addition, we relaxed the resulting geometries by constrained optimization in internal coordinates (henceforth referred to as *normal mode relaxation*). This approach preserves the overall molecular shape predicted by docking while relaxing the bond lengths and angles to obtain even more accurate input structures. The applicability of our molecular docking-based approach was evaluated by the P3D/PALES method. Finally, we also provide a comprehensive discussion of the underlying reasons behind the obtained results.

## Results and discussion

### Design and characterization of model compounds

We designed three classes of model compounds (Fig. 1) differing in molecular flexibility, as indicated by the  $n\text{Conf}_{20}$  parameter, which have been devised to quantify the flexibility of a molecule<sup>41</sup> (details are provided in the Experimental Section). As the most rigid compounds, we prepared phosphorylated derivatives of isopinocampheol (IPC; **1**,  $n\text{Conf}_{20} = 0$ ), which is often used as a model compound in RDC studies.<sup>9,42–44</sup> We synthesized all four isomers of **1** to probe the robustness of the proposed docking-based approach. The absolute configuration of **1** was determined by X-ray diffraction (more details in ESI†). Subsequently, we applied the docking approach to mildly (**2**,  $n\text{Conf}_{20} = 2$ ) and highly flexible compounds (**3**,  $n\text{Conf}_{20} = 8$ ) that have been studied in our previous work.<sup>21</sup> Both **2** and **3** were characterized in our previous study<sup>21</sup> by means of NMR spectroscopy. Therein, also the absolute configuration of **2-SR** and **2-RR** was determined by X-ray diffraction. So far, we have been unable to prepare suitable crystals of **3** for X-ray diffraction analysis. This reluctance to crystallize has been previously associated with high molecular flexibility.<sup>45</sup>

### Molecular docking of **1** (rigid structures)

Standard conformational analysis revealed only one low-energy conformer for each stereoisomer of **1**. Conversely, by molecular docking we found four unique conformers of **1-SR**, one of **1-SS**, three of **1-RS**, and three of **1-RR**. The DFT-calculated single



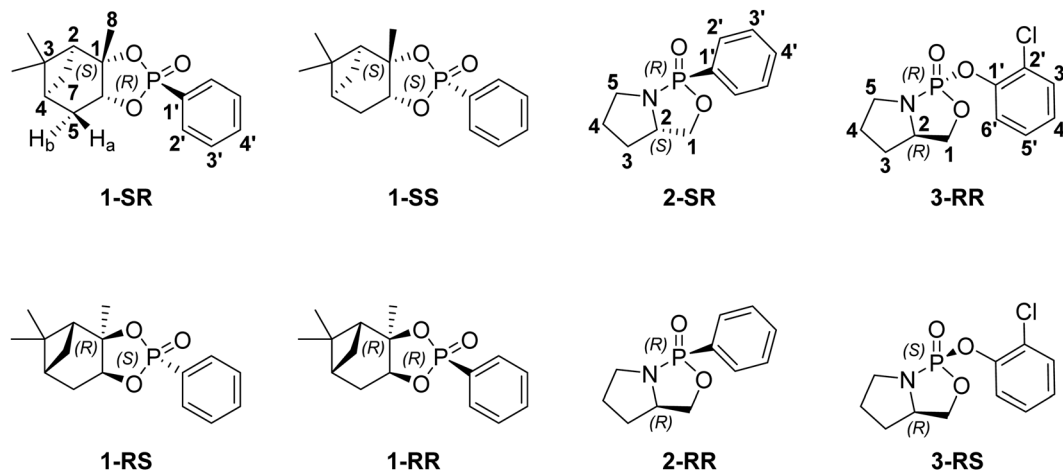


Fig. 1 Chemical structures and atom numbering of the compounds studied. Absolute configuration of compounds **1** and **2** was determined by X-ray diffraction. Stereochemistry is denoted only for the relevant chiral centers (*i.e.*, phosphorus and carbon C1 for **1** or C2 for **2** and **3**).

point energies of the **1-SR** and **1-SS** conformers generated by docking were located approximately 6 kcal mol<sup>-1</sup> above the individual global minima (GMs), whereas those of **1-RS** and **1-RR** were located more than 13 kcal mol<sup>-1</sup> above GMs (Table S2 in ESI<sup>†</sup>).

The different energy distributions of unique conformations of individual isomers may be explained by differences in their diastereotopic interactions with the alignment medium. In particular, the energy penalty seems to be associated with the configuration on carbon C1 (Fig. 1). As shown in Fig. 2, the detailed microscopic view of poses obtained from molecular docking reveals that the (*SX*)-isomers (*X* = *R* or *S*; Fig. 2(a) and (b)) fit into the groove that winds around the PBLG polymer chain. By contrast, the (*RX*)-isomers (Fig. 2(c) and (d)) do not fit into this groove due to the shape of the bicyclic part of the molecule. The  $\pi$ - $\pi$  stacking between the phenyl groups of the polymer and the analyte was observed

throughout majority of the poses. The distance between the stacked rings was  $\sim 4$  Å. Table S2 (ESI<sup>†</sup>) lists the DFT single-point energies of all unique conformers generated by docking and calculated relatively to the individual GMs ( $\Delta E = E(\text{conf.}) - E(\text{GM})$ , where *E* is the electronic energy of the given system).

To address the conformational tumbling of the molecule in the environment of the alignment medium, we calculated the Boltzmann distribution-weighted average of the RDCs of conformers obtained from docking. These RDCs (Unrelaxed. Boltz.) were correlated with all the experimental datasets of each isomer, as shown in Fig. 3 (the RDC values of each conformer are provided in Fig. S8–S23 in ESI<sup>†</sup>). The Pearson correlation coefficients (*R*) of the GM theoretical RDCs are also shown in Fig. 3.

For **1-SR** and **1-RS**, only the correct RDC datasets yields high correlation coefficients (*R* > 0.8); for **1-SS** and **1-RR**, both the

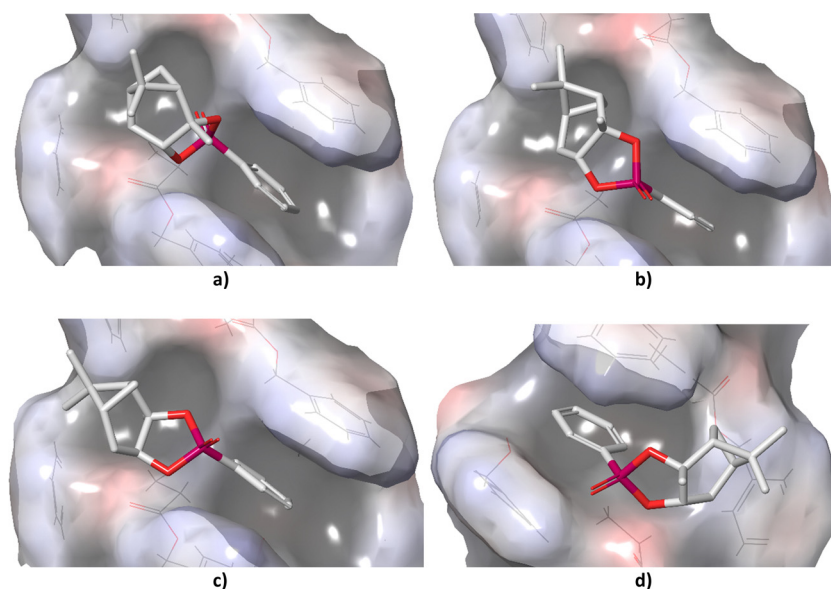


Fig. 2 Docked poses of (a) **1-SR**, (b) **1-SS**, (c) **1-RS**, and (d) **1-RR** in the helical structure of PBLG obtained from molecular docking. Carbon atoms are in white, oxygen in red, phosphorus in purple; hydrogens were omitted for clarity.



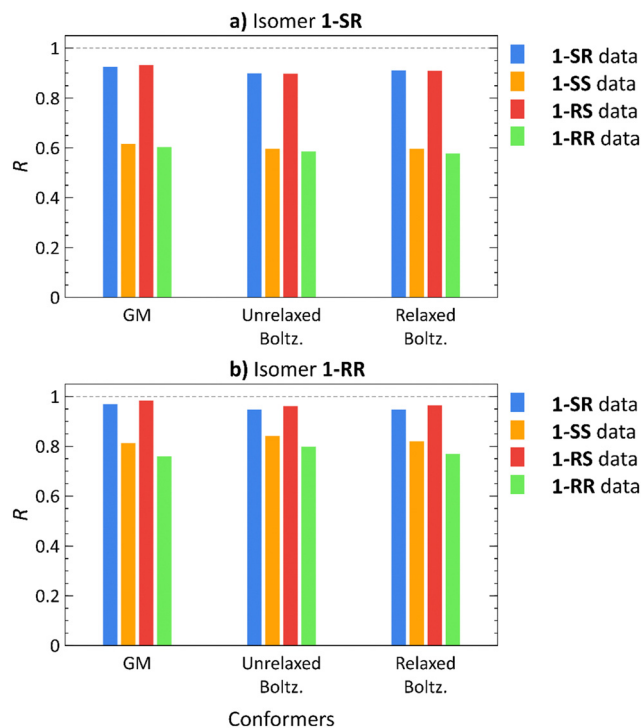


Fig. 3 Pearson correlation coefficients ( $R$ ) of RDCs calculated using global minimum (GM) conformers, Boltzmann distribution-weighted theoretical RDCs of conformers obtained from molecular docking (Unrelaxed Boltz.), and the same conformers after normal mode relaxation (Relaxed Boltz.) of isomers **1-SR** (a) and **1-RR** (b) correlated with **1-SR** (blue), **1-SS** (orange), **1-RS** (red), and **1-RR** (green) RDC experimental datasets.

correct and the incorrect datasets lead to high correlation coefficients (Fig. 3). Nevertheless, the correct **1-SS** and **1-RR** datasets show a stronger correlation with the (correct) **1-SS** and **1-RR** structures, than with **1-SR** and **1-RS**. Therefore, diastereomer discrimination is not unequivocal although we are able to unilaterally determine the configuration of the enantiomer pairs of the compounds under study.

To improve the diastereomer discrimination, we relaxed the structures generated by docking using the normal mode relaxation method developed by Bouř *et al.*<sup>46,47</sup> (computational details are provided in the Experimental Section). We assumed that density functional theory (DFT)-aided relaxation of the conformers obtained from docking would lead to more accurate conformer ensembles. The relaxed conformers were found approximately 1 kcal mol<sup>-1</sup> for **1-SR** and **1-SS** and 2–4 kcal mol<sup>-1</sup> for **1-RS** and **1-RR** above the corresponding GM (Table S3, ESI<sup>†</sup>). Normal mode relaxation slightly improved the diastereomer distinction, albeit with the same trends as those observed in unrelaxed conformers and GM structures. Thus, implementing molecular docking does not appear to be beneficial for RDC-aided structural elucidation of structures as rigid as isomers of **1** ( $n\text{Conf}_{20} = 0$ ).

### Molecular docking of **2** (mildly flexible molecules)

The molecules **2-SR** and **2-RR** are more flexible than the isomers of **1**, as shown by the number of conformations (see below) as well as by  $n\text{Conf}_{20}$  parameter of **2**, but still partly rigid

due to their two interconnected five-membered rings. Using the low-energy approach, we have previously identified three non-redundant low-energy conformers for each diastereomer of isolated system **2**.<sup>21</sup> In turn, the docking procedure, which mimics the anisotropic environment, found tens of docking poses for each diastereomer of **2**. Again,  $\pi$ - $\pi$  stacking was observed for the majority of the discovered poses. However, many of the resulting conformers were nearly identical; hence, we had to apply a method for eliminating redundant conformers.

In our previous study, we have employed an approach based on dihedral angle analysis,<sup>21</sup> ultimately identifying three unique conformers for both isomers. In this work, we used a more complex machine learning (ML)-based algorithm for eliminating redundant conformers (henceforth referred to as ML elimination, see Experimental Section). As a result, we found 22 and 21 unique conformers for **2-SR** and **2-RR**, respectively. According to the DFT single-point energies, all of these conformers are located 4–10 kcal mol<sup>-1</sup> above the corresponding GMs (Table S4, ESI<sup>†</sup>). The structures generated by docking were then relaxed by normal mode relaxation. The relaxed conformers lie 1–6 kcal mol<sup>-1</sup> above the corresponding GMs (Table S5, ESI<sup>†</sup>).

Employing the Boltzmann distribution-weighted theoretical RDCs, we calculated the  $R$  values against the sets of experimental RDCs. The results shown in Fig. 4 demonstrate that the overall trends match those observed in our previous study<sup>21</sup> (also included in Fig. 4 for a direct comparison: conformers A, B, C).

The **2-RR** RDC dataset provides the best diastereomer discrimination (Tables S57 and S59, ESI<sup>†</sup>), whereas the **2-SR** dataset yields a good fit ( $R > 0.8$ ) for both **2-SR** and **2-RR** structures (Tables S56 and S58, ESI<sup>†</sup>). However, the **2-SR** structures show a higher overall correlation for conformers from molecular docking as compared to the low-energy approach. Furthermore, the additional normal mode relaxation increases the  $R$  values in all cases, except for the **2-SR** RDC dataset with **2-RR** conformers, where the  $R$  is slightly lower ( $R = 0.988$  for the unrelaxed *versus* 0.985 for relaxed structures).

For the **2-SR** isomer, the GM structure yields  $R = 0.785$  using the **2-SR** experimental RDCs,<sup>21</sup> Fig. 4. Employing conformers generated by molecular docking increases the **2-SR** dataset  $R$  value to 0.815. Subsequent normal mode relaxation further improves the correlation ( $R = 0.871$ ), albeit for both the correct and the incorrect datasets. Using conformers of **2-RR**, diastereomer discrimination was not possible, neither using the low-energy nor the docking approach. Although the **2-SR** dataset produces  $R > 0.8$  for both **2-RR** and **2-SR** molecular structures, the **2-RR** RDC dataset unambiguously yields a weak correlation for **2-SR** ( $R = 0.578$  after normal mode relaxation) and a strong correlation for **2-RR** structures ( $R = 0.964$  after normal mode relaxation). Thus, we have been able to achieve only unilateral diastereomer discrimination.

Notwithstanding the unilateral diastereomer discrimination, the structures can be assigned to datasets at a specific level of probability. Overall, the trends of **1** and **2** are rather similar. For visual comparison, the lowest energy conformers obtained in the previous study (A,<sup>21</sup> red), and the lowest energy



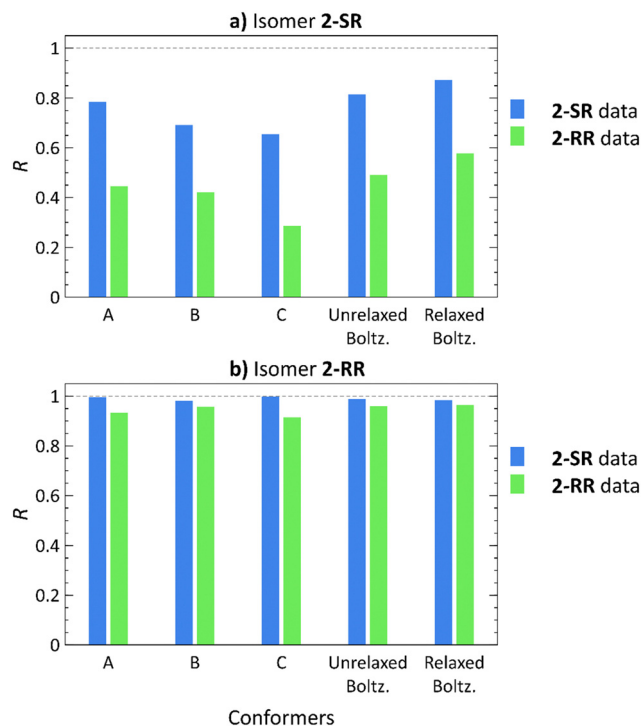


Fig. 4 Pearson correlation coefficients ( $R$ ) of RDCs calculated using low-energy conformers<sup>21</sup> (A)–(C), Boltzmann distribution-weighted theoretical RDCs of conformers obtained from molecular docking (Unrelaxed Boltz.), and the same conformers after relaxation by normal mode relaxation (Relaxed Boltz.) of structures **2-SR** (a) and **2-RR** (b) correlated with the **2-SR** (blue) and **2-RR** (green) RDC experimental datasets.

conformers generated by molecular docking with (green) and without (blue) subsequent normal mode relaxation are shown in Fig. 5.

### Molecular docking of 3 (highly flexible molecules)

**3-RS** and **3-RR** are the most flexible molecules in this study. In our previous work,<sup>21</sup> conformational analysis of **3** with subsequent elimination of redundant structures produced 9 and 17 unique low-energy conformers for **3-RS** and **3-RR**, respectively. The docking approach with ML elimination of redundant conformers identified 23 unique conformers for **3-RS** and 34 conformers for **3-RR** isomer. The DFT-calculated single-point energies revealed that these conformers are located 6–15 kcal mol<sup>-1</sup> above the

individual GMs, and 2–11 kcal mol<sup>-1</sup> above the GMs after the subsequent normal mode optimization (details in Tables S6 and S7, ESI<sup>†</sup>). Unfortunately, we could not assign the experimental RDC datasets to the two isomers of **3**, as we were unable to prepare crystals of **3** suitable for X-ray diffraction analysis (presumably because of the high flexibility of these molecules as indicated above). For this reason, the two measured RDC datasets are denoted as **3-A** and **3-B** in the following.

The **3-A** RDC dataset yields higher correlation coefficients with both **3-RS** and **3-RR** structures than the **3-B** dataset, Fig. 6. The best correlation coefficients ( $R > 0.8$ ) were obtained for the normal mode-relaxed **3-RS** conformers with the **3-A** dataset. However, the **3-B** dataset fits on **3-RS** structures as well, although the absolute  $R$  values are smaller than for the **3-A** dataset ( $R = 0.802$  and  $0.590$  for **3-A** and **3-B** RDC datasets, respectively). Thus, whether any degree of diastereomer discrimination can be inferred is at best questionable.

The trend, in which one dataset yields good correlation for both diastereomers was observed for all of the studied compounds, albeit less markedly for **3**. So, does this trend have a physical meaning? We speculate that a possible explanation is that the isomers with an ‘all-fitting’ RDC dataset (**1-SR**, **1-RS**, **2-SR**, **3-A**) are moving more freely in the alignment medium environment. Thus, the induced RDCs are averaged over a higher number of geometries, accounting the goodness-of-fit of the experimental RDCs and the calculated RDCs of various conformers and isomers. Conversely, the isomers with the ‘non-fitting’ RDC dataset (**1-SS**, **1-RR**, **2-RR**, **3-B**) may be involved in isomer-specific interactions with the alignment medium, which cannot be adequately modelled by molecular docking. For **3**, another possible explanation may be the higher variability of the RDC values of the **3-A** dataset (Tables S32–S39, ESI<sup>†</sup>). Consequently, the **3-A** dataset has higher  $R$  values than the **3-B** dataset.

Fig. 6 shows yet another trend: using normal mode relaxation significantly improves the correlation of the worst-fitting RDC dataset **3-B** with **3-RS** structures and worsens the correlation with **3-RR** isomer structures. A similar trend is observed in Fig. 4, where the **2-SR** RDC dataset shows a slightly lower  $R$  with the incorrect **2-RR** structures after normal mode relaxation, suggesting an incorrect structure/dataset match. However, without the XRD structure, we cannot ascertain whether this trend can help us to identify which dataset belongs to which isomer with a high level of certainty.

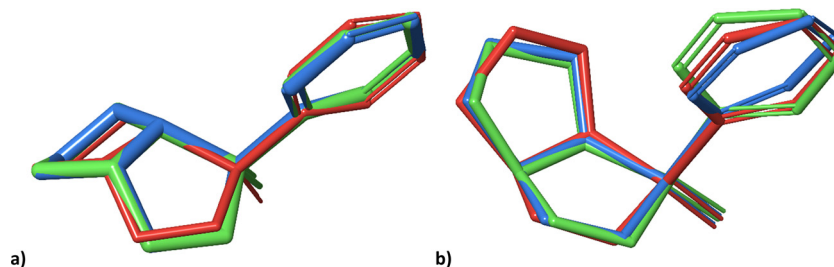


Fig. 5 Geometries of the lowest energy conformers of **2-SR** (a) and **2-RR** (b) from the previous study<sup>21</sup> (red), and the lowest energy conformers obtained from molecular docking into PLBG before (blue) and after normal mode relaxation (green).



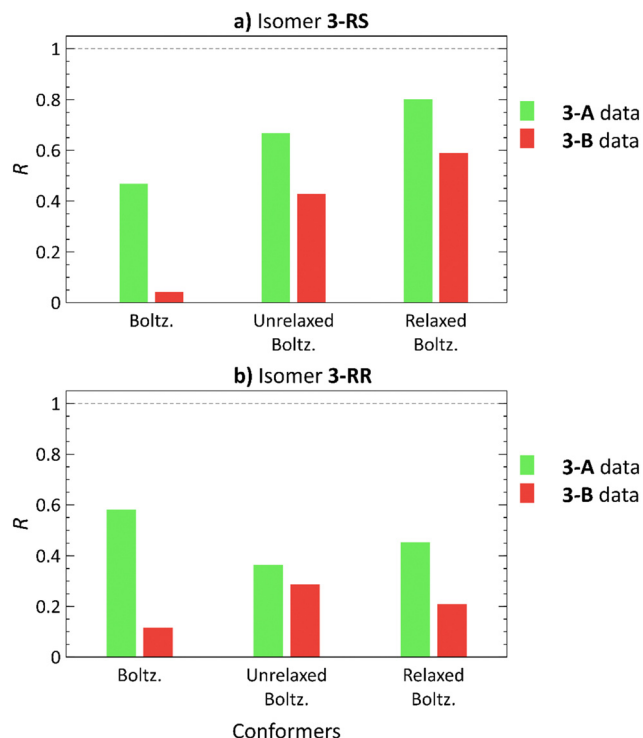


Fig. 6 Pearson correlation coefficients ( $R$ ) of Boltzmann distribution-weighted average RDCs calculated using theoretical RDCs of low-energy conformers<sup>21</sup> (Boltz.), conformers obtained from molecular docking (unrelaxed Boltz.) and these conformers after normal mode relaxation (relaxed Boltz.) of **3-RS** (a) and **3-RR** (b) correlated with the **3-A** (green) and **3-B** (red) RDC experimental datasets.

Furthermore, for highly flexible molecules such as **3**, the conventional approach for the experimental determination of RDCs may fail.<sup>12,24,32,48,49</sup> Typically, the  $J$ -couplings extracted from experiment in the isotropic environment (*i.e.*, isotropic  $J$ -couplings) are subtracted from the total  $T$ -couplings measured in the anisotropic environment to yield RDCs (eqn (1) in Experimental Section). However, the isotropic  $J$ -couplings may significantly differ from the real anisotropic  $J$ -couplings in the alignment medium, as indicated by the different conformer populations in isotropic and in anisotropic environments.

The differences between  $J$ -couplings calculated employing different conformer ensembles are clear and support our hypothesis: as shown in Table 1, we calculated  $J$ -couplings in the isotropic (presented by the low-energy conformer ensemble) and in the anisotropic environment (presented by the docking-obtained conformer ensemble). As an example, the calculated isotropic and anisotropic  $J_{C3-P}$ -couplings of the **3-RR** isomer are 3.8 and 1.0 Hz, respectively, resulting in a difference of 2.8 Hz between the isotropic and anisotropic value. For RDC calculated according to eqn (1) in Experimental Section, this difference would lead to an error of 2000% of the experimental RDC value determined using the isotropic  $J$ -coupling. More examples to show the difference between  $J$ -couplings calculated using different conformer ensembles can be found in Table 1. A more detailed study on this subject is currently underway.

Table 1 Comparison between experimental  $J$ -couplings of **3-A** and **3-B** measured in chloroform<sup>21</sup> (exp.  $J$ ) and Boltzmann distribution-averaged  $J$ -couplings of **3-RS** and **3-RR** calculated at the B3LYP-D3/IGLO-III DFT level with PCM = chloroform of low-energy conformers (calc.  $J$  low-energy ensemble),<sup>21</sup> and relaxed conformers obtained from molecular docking (calc.  $J$  relaxed docking ensemble). Only calculated  $J$ -couplings of isomers with >2% relative abundance were used for the averaging

Interaction	Exp. $J$ (Hz) <sup>21</sup>		Calc. $J$ low-energy ensemble (Hz) <sup>21</sup>		Calc. $J$ relaxed docking ensemble (Hz)	
	3-A	3-B	3-RS	3-RR	3-RS	3-RR
C1-P	4.8	4.6	6.8	5.6	6.9	5.1
C2-P	13.0	12.9	12.2	13.1	10.7	12.0
C3-P	4.1	3.6	3.1	3.8	0.4	1.0
C4-P	3.4	3.1	2.0	2.7	-0.7	0.8
C5-P	3.2	1.4	0.4	2.4	-0.5	1.4
C1'-P	6.1	8.9	-10.8	-10.7	-11.7	-7.7
C2'-P	7.6	8.9	8.7	5.9	10.9	7.0
C3'-P	2.9	2.6	-0.8	-2.3	0.0	-1.2
C4'-P	1.8	1.4	1.2	2.6	0.5	1.7
C5'-P	1.7	1.3	-1.5	-2.6	-0.8	-1.9
C6'-P	1.3	0.8	2.1	3.2	2.9	2.2
C2-H2	151.4	150.3	147.3	141.4	145.5	134.3
C3'-H3'	163.6	163.9	162.9	156.7	162.2	150.3
C4'-H4'	162.6	162.5	160.6	154.4	160.0	148.4
C5'-H5'	165.4	163.0	159.6	153.9	159.4	147.6
C6'-H6'	165.6	165.3	164.7	156.8	162.3	149.9

## Experimental section

### Computational methodology

**Conformational sampling and molecular docking.** As an input for molecular docking, structures generated *via* initial conformational sampling were used. The conformational sampling in MacroModel v13.1 (Schrödinger 2022-1 suite)<sup>50</sup> was performed with the following set of parameters: OPLS4 force field,<sup>51</sup> chloroform solvent, mixed torsional/low-mode sampling method with a maximum of 1000 steps, 40 kJ mol<sup>-1</sup> energy window for saving structures, and 0.75 Å maximum atom deviation cut-off. Molecular docking was performed using Glide (as implemented in Schrödinger 2022-1 suite)<sup>52-54</sup> in flexible XP mode using grids of various sizes (10 Å, 12 Å, 15 Å, 18 Å, and 20 Å) to sample all possible positions of the analyte molecules. Aromatic hydrogen atoms were included as donors (0.0 partial charge cutoff), and chlorine was included as both the acceptor and the donor in separate searches with each grid for **3**. The model of PBLG used in docking was provided to us by Alain Ibáñez de Opakua.<sup>15</sup>

**Density functional theory (DFT) calculations.** DFT calculations were performed using the Gaussian 16 program package.<sup>55</sup> Single-point energies of the conformers generated by molecular docking were calculated using the B3LYP functional<sup>56-59</sup> with empirical dispersion correction D3 by Grimme<sup>60</sup> and the 6-31+G(d,p)<sup>61</sup> basis set. DFT calculations were performed for single molecule *in vacuo* at 0 K.  $J$ -Couplings were calculated using the B3LYP functional and a basis set optimized for <sup>31</sup>P parameters (IGLO-III)<sup>62-64</sup> within the polarizable continuum model (PCM)<sup>65</sup> using chloroform as solvent.

**Machine learning-aided elimination of redundant structures.** Unique conformers were identified using the machine learning (ML)-based mean-shift algorithm<sup>66</sup> as implemented in



the Scikit-learn Python package (scikit-learn.org). This approach was previously adapted for identification of non-redundant conformers in a series of dipeptides.<sup>67</sup> During the ML-aided elimination procedure, the conformers were sorted into clusters based on selected dihedral angles. Each cluster contained all structurally similar conformers. The mean shift clustering algorithm specifies the cluster size using a bandwidth parameter. The value of this parameter determines whether each data point belongs to a given cluster, *i.e.*, if a given set of conformers is structurally similar. This parameter is proportional to the size of a given cluster. The optimal values for each structure were decided as described in the ESI† (numbers are provided in Table S64). For each cluster, only the lowest energy conformer was selected as the unique non-redundant conformer.

**Normal mode relaxation.** Normal mode optimization has been used to refine molecular structures, namely bond lengths and angles, while keeping the overall shape of the molecule, *i.e.*, by fixing the lowest vibrational modes.<sup>46,47</sup> In this procedure, modes up to 300 cm<sup>-1</sup> were fixed, and the QGRAD program was used for normal mode optimization. The accompanying DFT computations were performed at the B3LYP-D3/DGDZVP<sup>68</sup> level using the Gaussian 16 software suite.<sup>55</sup> The final energies of the relaxed conformers were recalculated at the B3LYP-D3/6-31+G(d,p) level.

**P3D/PALES.** We performed molecular alignment simulations using P3D<sup>15</sup> as implemented in the PALES<sup>17</sup> software, as recommended.<sup>16</sup> The following command was used to run the simulation:

```
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 Molecule.pdb -inD RDCs.tbl -wv
0.12 -rM 8 -pka charges.pka -outD output.out,
```

where PBLG.dx is the potential file of PBLG, Molecule.pdb is the PDB file of the studied molecule, RDCs.tbl is the list of experimental RDCs, charges.pka is the list of atomic charges obtained from AtomicChargeCalculator II,<sup>69</sup> and output.out is the final output file. Atomic charges were calculated using the electronegativity equalization method (EEM)<sup>70</sup> based on the atoms in molecules (AIM) calculation scheme at the B3LYP/6-311G level of theory.<sup>71</sup>

**Calculation of  $n\text{Conf}_{20}$ .** We calculated the values of the  $n\text{Conf}_{20}$  flexibility descriptor using the code provided in the original article by Wicker and Cooper.<sup>41</sup> Given that a more detailed scale was necessary for the studied compounds, we set the atom root-mean-square (RMS) distance threshold for removing duplicate conformers to 0.5 Å. The number of generated conformers was set to 10 000 to mitigate the influence of stochasticity in the process of structure generation.

## Experimental methodology

**Sample preparation.** The quantity of the alignment media was derived from the quantities of the solvent and analyte to acquire approximately 6.1–7.9 weight% of the alignment medium in the sample. This number was determined based on our

experience<sup>21,40,72</sup> and on the premeasurement screening. The calculated amount of the alignment medium was mixed with the respective amounts of the solvent and the analyte (detailed sample compositions are provided in ESI†), and the final mixture was left standing to dissolve overnight.

Due to the high viscosity of the solution, the sample was homogenized using a manual centrifuge. Sample homogeneity was then monitored by <sup>2</sup>H image experiments, and the alignment order was determined by measuring the quadrupolar splitting of the solvent in <sup>2</sup>H NMR spectra.

**Extraction of experimental RDCs and their evaluation.** Residual dipolar couplings were acquired using F1-coupled HSQC (one-bond <sup>13</sup>C–<sup>1</sup>H couplings) or APT (<sup>13</sup>C–<sup>31</sup>P couplings) experiments. The F1-coupled HSQC sequence was *J*-scaled, and the separation of the components was 8-times enhanced with respect to the <sup>13</sup>C chemical shift evolution.<sup>44</sup> The one-bond <sup>13</sup>C–<sup>1</sup>H and <sup>13</sup>C–<sup>31</sup>P couplings were obtained using the following equation:

$${}^nD_{C-Y} = \frac{{}^nT_{C-Y} - {}^nJ_{C-Y}}{2}, \quad (1)$$

where *J* stands for the scalar coupling (obtained from an isotropic experiment – measured in CDCl<sub>3</sub>), *T* is the total coupling (collected from an anisotropic experiment measured in the solvent CDCl<sub>3</sub> and in the alignment medium PBLG), *D* is the residual dipolar coupling, index *n* indicates the number of bonds, and C–Y stands for <sup>13</sup>C–<sup>1</sup>H or <sup>13</sup>C–<sup>31</sup>P coupling. The values corresponding to the rotating *t*-butyl groups were converted into one-bond C–<sup>1</sup>D<sub>CC</sub> – the RDC between methyl and tertiary carbon nuclei.<sup>73</sup>

The RDC data were evaluated using Pearson's correlation coefficient *R* and quality factor *Q*.

Pearson's correlation coefficient *R* is defined as follows:

$$R = \frac{n \sum D_{\text{exp}} D_{\text{calc}} - \sum D_{\text{exp}} \sum D_{\text{calc}}}{\sqrt{n \sum D_{\text{exp}}^2 - (\sum D_{\text{exp}})^2} \sqrt{n \sum D_{\text{calc}}^2 - (\sum D_{\text{calc}})^2}}, \quad (2)$$

where *n* is the number of experimental RDC values, *D*<sub>exp</sub> are the experimental RDCs, and *D*<sub>calc</sub> are the theoretical RDC values.

The quality factor is given by the following equation:

$$Q = \sqrt{\frac{\frac{1}{n} \sum (D_{\text{exp}}^2 - D_{\text{calc}}^2)}{\frac{1}{n} \sum D_{\text{exp}}^2}}, \quad (3)$$

where *D*<sub>exp</sub> are the experimental RDC values, *D*<sub>calc</sub> are the calculated RDCs, and *n* is the number of experimental RDCs.

## Conclusions

In this work, we investigated whether the involvement of alignment media in generation of conformer ensembles *via* molecular docking improves the performance of RDC-aided stereochemical analysis. We have found that it depends on the flexibility of the molecules under study. The docking approach does not markedly affect the conformation space of rigid molecules such as **1**, neither it significantly improves the quality of RDC analysis. However, the different conformer ensembles do affect the results, indicating that the alignment medium influences the conformation space of



the analyte molecules. Generally, a non-negligible energetic penalty for higher-lying conformers may be compensated for by the analyte–medium interactions which may extend/alter the conformer ensembles in the anisotropic environment in comparison with the isotropic solution. This energy penalty is diastereotopic (isomer-dependent), this effect is most profound in the case of rigid molecules, such as isomers of **1** that are not as capable of adapting to the chiral environment of PBLG. For mildly flexible systems, such as **2**, the ensembles of conformers generated by molecular docking clearly enhance the accuracy of RDC-aided stereochemical analysis. Additional DFT optimization of conformers obtained from molecular docking by constrained normal mode relaxation further improves the accuracy of RDC analysis and brings in a clearer diastereomer discrimination. The docking approach does not improve the results of RDC analysis for highly flexible structures, such as **3**; we speculate that this is due to the possibly incorrect experimental determination of RDCs using  $J$ -couplings obtained from experiments in isotropic solutions. The total coupling  $T$  of an anisotropic conformer ensemble consists of RDCs and  $J$ -couplings, however, the difference between anisotropic and isotropic  $J$ -couplings may increase with molecular flexibility due to the fact that  $J$ -couplings are rather sensitive to changes in molecular geometries. This difference may preclude RDC determination for highly flexible molecules because – like RDCs – anisotropic  $J$ -couplings cannot be measured directly. To answer the questions laid in the Introduction, the consideration of alignment medium does have an effect on the conformation space of analyte and thus the quality of RDC analysis. Even more accurate conformer ensembles may be generated using advanced simulation methods (employing, e.g., QM/MM or molecular dynamics). Furthermore, the simple model containing “static snapshots”, i.e., singular conformers of the studied systems, is perhaps insufficient for the correct description of the entire conformational space. Methods addressing fast molecular motion and dynamic effects may provide more precise results, however, at a larger computational cost. From experimental point of view, the effect of different alignment media on the RDC values should be compared because specific analyte–medium interactions likely induce different conformer ensembles, which are associated with different sets of RDCs.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This research was funded by the Czech Science Foundation (21-23014S). We thank Dr Jakub Kaminský for his help with the normal mode relaxation method, Kvetoslava Kertisová for HR-MS analysis, Alain Ibáñez de Opakua for the model structure of PBLG, and Dr Carlos V. Melo for editing the manuscript.

## References

- M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870–2871.
- R. A. Bell and J. K. Saunders, *Can. J. Chem.*, 1970, **48**, 1114–1122.
- A. W. Overhauser, *Phys. Rev.*, 1953, **92**, 411–415.
- F. Kramer, M. V. Deshmukh, H. Kessler and S. J. Glaser, *Concepts Magn. Reson., Part A*, 2004, **21A**, 10–21.
- C. M. Thiele, *Eur. J. Org. Chem.*, 2008, 5673–5685.
- G. Kummerlöwe and B. Luy, *TrAC, Trends Anal. Chem.*, 2009, **28**, 483–493.
- M. Schwab, D. Herold and C. M. Thiele, *Chem. – Eur. J.*, 2017, **23**, 14576–14584.
- J. C. Freudenberger, S. Knor, K. Kobzar, D. Heckmann, T. Paululat, H. Kessler and B. Luy, *Angew. Chem., Int. Ed.*, 2005, **44**, 423–426.
- A. Marx and C. Thiele, *Chem. – Eur. J.*, 2009, **15**, 254–260.
- A. Navarro-Vázquez, *Magn. Reson. Chem.*, 2012, **50**, S73–S79.
- V. Schmidts, *PhD thesis*, Technische Universität Darmstadt, 2013.
- M. Di Pietro, U. Sternberg and B. Luy, *J. Phys. Chem. B*, 2019, **123**, 8480–8491.
- H. Valafar and J. H. Prestegard, *J. Magn. Reson.*, 2004, **167**, 228–241.
- J. Meiler, W. Peti and C. Griesinger, *J. Biomol. NMR*, 2000, **17**, 283–294.
- A. Ibáñez de Opakua, F. Klama, I. E. Ndukwe, G. E. Martin, R. T. Williamson and M. Zweckstetter, *Angew. Chem., Int. Ed.*, 2020, **59**, 6172–6176.
- A. Ibáñez de Opakua and M. Zweckstetter, *Magn. Reson.*, 2021, **2**, 105–116.
- M. Zweckstetter, *Nat. Protoc.*, 2008, **3**, 679–690.
- J. R. Kim and T. Ree, *J. Polym. Sci., Polym. Chem. Ed.*, 1985, **23**, 215–221.
- P. Doty, J. H. Bradbury and A. M. Holtzer, *J. Am. Chem. Soc.*, 1956, **78**, 947–954.
- M. Sarfati, P. Lesot, D. Merlet and J. Courtieu, *Chem. Commun.*, 2000, 2069–2081.
- M. Tichotová, A. Ešnerová, L. Tučková, L. Bednářová, I. Císařová, O. Baszczyński and E. Procházková, *J. Magn. Reson.*, 2022, **336**, 107149.
- A. Schuetz, J. Junker, A. Leonov, O. F. Lange, T. F. Molinski and C. Griesinger, *J. Am. Chem. Soc.*, 2007, **129**, 15114–15115.
- P. Trigo-Mouriño, R. Santamaria-Fernández, V. M. Sánchez-Pedregal and A. Navarro-Vázquez, *J. Org. Chem.*, 2010, **75**, 3101–3104.
- C. M. Thiele, V. Schmidts, B. Böttcher, I. Louzao, R. Berger, A. Maliniak and B. Stevansson, *Angew. Chem., Int. Ed.*, 2009, **48**, 6708–6712.
- C. M. Thiele, A. Maliniak and B. Stevansson, *J. Am. Chem. Soc.*, 2009, **131**, 12878–12879.
- D. Selegato, C. Bracco, C. Giannelli, G. Parigi, C. Luchinat, L. Sgheri and E. Ravera, *ChemPhysChem*, 2021, **22**, 127–138.
- E. Ravera, L. Sgheri, G. Parigi and C. Luchinat, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5686–5701.
- S. Bottaro, T. Bengtsen and K. Lindorff-Larsen, *Methods Mol. Biol.*, 2020, **2112**, 219–240.
- I. Bertini, Y. K. Gupta, C. Luchinat, G. Parigi, M. Peana, L. Sgheri and J. Yuan, *J. Am. Chem. Soc.*, 2007, **129**, 12786–12794.





- 30 U. Sternberg, P. Tzvetkova and C. Muhle-Goll, *Phys. Chem. Chem. Phys.*, 2020, **22**, 17375–17384.
- 31 P. Tzvetkova, U. Sternberg, T. Gloge, A. Vazquez and B. Luy, *Chem. Sci.*, 2019, **10**, 8774–8791.
- 32 F. A. Roth, V. Schmidts, J. Rettig and C. M. Thiele, *Phys. Chem. Chem. Phys.*, 2022, **24**, 281–286.
- 33 F. Stanzione, I. Giangreco and J. C. Cole, *Prog. Med. Chem.*, 2021, **60**, 273–343.
- 34 N. S. Pagadala, K. Syed and J. Tuszynski, *Biophys. Rev.*, 2017, **9**, 91–102.
- 35 K. A. Peele, C. Potla Durthi, T. Srihansa, S. Krupanidhi, V. S. Ayyagari, D. J. Babu, M. Indira, A. R. Reddy and T. C. Venkateswarulu, *Inform. Med. Unlocked*, 2020, **19**, 100345.
- 36 V. M. Marathias, G. J. Tawa, I. Goljer and A. C. Bach, *Chirality*, 2007, **19**, 741–750.
- 37 A. Marx, V. Schmidts and C. M. Thiele, *Magn. Reson. Chem.*, 2009, **47**, 734–740.
- 38 A. Marx, B. Böttcher and C. M. Thiele, *Chem. – Eur. J.*, 2010, **16**, 1656–1663.
- 39 L. Vrzal, M. Kratochvilova-Simanova, T. Landovsky, K. Polivkova, J. Budka, H. Dvorakova and P. Lhotak, *Org. Biomol. Chem.*, 2015, **13**, 9610–9618.
- 40 T. Landovský, M. Tichotová, L. Vrzal, J. Budka, V. Eigner, H. Dvořáková and P. Lhoták, *Tetrahedron*, 2018, **74**, 902–907.
- 41 J. G. P. Wicker and R. I. Cooper, *J. Chem. Inf. Model.*, 2016, **56**, 2347–2352.
- 42 S. Hansmann, T. Larem and C. M. Thiele, *Eur. J. Org. Chem.*, 2016, 1324–1329.
- 43 T. Montag and C. M. Thiele, *Chem. – Eur. J.*, 2013, **19**, 2271–2274.
- 44 S. Jeziorowski and C. Thiele, *Chem. – Eur. J.*, 2018, **24**, 15631–15637.
- 45 L. Yu, S. M. Reutzel-Edens and C. A. Mitchell, *Org. Process Res. Dev.*, 2000, **4**, 396–402.
- 46 P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126–4132.
- 47 P. Bouř, *Collect. Czech. Chem. Commun.*, 2005, **70**, 1315–1340.
- 48 A. Canales, J. Jiménez-Barbero and M. Martín-Pastor, *Magn. Reson. Chem.*, 2012, **50**, S80–S85.
- 49 M. E. Di Pietro, G. De Luca, G. Celebre, D. Merlet and C. Aroulanda, *Mol. Cryst. Liq. Cryst.*, 2015, **614**, 39–53.
- 50 *Schrödinger Release 2022-1 ed.*, Schrödinger, LLC, New York, NY, 2022.
- 51 C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G. A. Ross, M. K. Dahlgren, E. Russell, C. D. Von Bargen, R. Abel, R. A. Friesner and E. D. Harder, *J. Chem. Theory Comput.*, 2021, **17**, 4291–4300.
- 52 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177–6196.
- 53 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J. Med. Chem.*, 2004, **47**, 1750–1759.
- 54 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
- 55 *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
- 56 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 57 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 58 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.
- 59 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 60 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 61 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- 62 L. Benda, Z. Sochorová Vokáčová, M. Straka and V. Sychrovský, *J. Phys. Chem. B*, 2012, **116**, 3823–3833.
- 63 J. Fukal, O. Páv, M. Buděšínský, I. Rosenberg, J. Šebera and V. Sychrovský, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9924–9934.
- 64 J. Fukal, O. Páv, M. Buděšínský, J. Šebera and V. Sychrovský, *Phys. Chem. Chem. Phys.*, 2017, **19**, 31830–31841.
- 65 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.
- 66 D. Comaniciu and P. Meer, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, 603–619.
- 67 T. Kalvoda, M. Culka, L. Rulišek and E. Andris, *J. Phys. Chem. B*, 2022, **126**, 5949–5958.
- 68 N. Godbout, D. Salahub, J. Andzelm and E. Wimmer, *Can. J. Chem.*, 2011, **70**, 560–571.
- 69 T. Raček, O. Schindler, D. Toušek, V. Horský, K. Berka, J. Koča and R. Svobodová, *Nucleic Acids Res.*, 2020, **48**, W591–W596.
- 70 W. J. Mortier, S. K. Ghosh and S. Shankar, *J. Am. Chem. Soc.*, 1986, **108**, 4315–4320.
- 71 S. Geidl, T. Bouchal, T. Raček, R. Svobodová Vařeková, V. Hejret, A. Křenek, R. Abagyan and J. Koča, *J. Cheminformatics*, 2015, **7**, 59.
- 72 M. Tichotová, T. Landovský, J. Lang, S. Jeziorowski, V. Schmidts, M. Kohout, M. Babor, P. Lhoták, C. M. Thiele and H. Dvořáková, *J. Org. Chem.*, 2023, DOI: [10.1021/acs.joc.2c02594](https://doi.org/10.1021/acs.joc.2c02594).
- 73 R. S. Stoll, M. V. Peters, A. Kuhn, S. Heiles, R. Goddard, M. Bühl, C. M. Thiele and S. Hecht, *J. Am. Chem. Soc.*, 2009, **131**, 357–367.

