



Cite this: *Phys. Chem. Chem. Phys.*,  
2020, 22, 17375

# The simulation of NMR data of flexible molecules: sagittamide A as an example for MD simulations with orientational constraints†

Ulrich Sternberg, \*<sup>ab</sup> Pavleta Tzvetkova <sup>c</sup> and Claudia Muhle-Goll<sup>c</sup>

The recently developed MDOC (Molecular Dynamics with Orientational Constraints) simulation is applied for the first time to a fully flexible molecule. MDOC simulations aim to single out the naturally existing configuration of molecules and to elucidate conformer populations. The performance of the method was first demonstrated on a well-studied test case, the five-membered ring lactone ( $\alpha$ -methylene- $\gamma$ -butyrolactone). In the case of sagittamide A, one-bond  $^1\text{H}$ – $^{13}\text{C}$  residual dipolar couplings (RDC) are used as orientational constraints that reorient the molecule or parts of it. In addition, NOE distances and  $^3\text{J}$  scalar couplings are used as constraints. Five possible configurations of sagittamide A (labelled a to e) are considered. One experimental RDC value per flexible unit was available and this was not sufficient to single out one valid configuration. The problem could be solved by including NOE distances as well as  $^3\text{J}$  couplings as complementary constraints into the MDOC simulations. In accordance with former investigations, we confirmed the configuration a for the natural product. A detailed analysis of conformers of the central chain of 6 chiral carbon atoms could be given by inspecting the MDOC trajectory. The relative abundance of these conformers is crucial in fulfilling all three sets of constraints.

Received 8th April 2020,  
Accepted 7th July 2020

DOI: 10.1039/d0cp01905d

rsc.li/pccp

## Introduction

To study molecules that form several rotamers in solutions, molecular dynamics (MD) simulation is the method of choice. However, if one intends to simulate NMR data used in structure determination, traditional all-atom MD simulations cannot bridge the gap of several orders of magnitude between the MD time-scale (up to  $\mu\text{s}$ ) and that of the equilibria governing NMR (up to ms). It will be shown that this time-scale problem can be solved by restrained MD simulation methods that are guided by NMR parameters. However, there is a distinct difference between NMR parameters like NOE distances or  $^3\text{J}$  couplings, which are scalars, and NMR interaction tensors: scalar quantities do not change under coordinate transformations but tensors are transformed by proper transformation matrices. To compare calculated interaction tensors such as chemical shift tensors, quadrupolar and dipolar coupling tensors with NMR measurements, they have to be transformed from their local molecular coordinate systems to the common

laboratory frame or coordinate system of the external magnetic field. The NMR interaction tensors depend, therefore, on molecular orientations with respect to the outer magnetic field, and are averaged by fast molecular motions in a specific way; if a molecule rotates fast about a fixed axis, all off diagonal elements of the interaction tensors will vanish. The basic idea of MDOC (Molecular Dynamics with tensorial Orientational Constraints) is to use NMR interaction tensors as constraints to force molecular rotations. In NMR experiments of uniaxial oriented media, such as stretched gels or bio-membranes, the fixed axis or vector of orientation (director<sup>1</sup>) is mostly oriented parallel to the outer magnetic field and the interaction tensors of rapid moving molecules become diagonal in this orientation. Dipolar interactions measured in oriented media<sup>2</sup> are called residual dipolar couplings (RDCs) and are used in MDOC to introduce orientational tensorial constraints. In most traditional structure investigations,<sup>3</sup> order parameters or more general order tensors are introduced to account for the motional and ensemble averaging of the dipolar interaction. In MDOC, order tensors are avoided and the average is performed by rotating the molecules or parts of them using RDC-derived tensors as constraints. Scalar NMR parameters can be added as constraints to adjust distances and angles if only sparse tensor data are available. NOE distances<sup>4</sup> and  $^3\text{J}$  couplings<sup>5</sup> put more realistic weights on conformers that are generated during MDOC simulations.

<sup>a</sup> Research Partner of Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.  
E-mail: ulrich.sternberg@partner.kit.edu

<sup>b</sup> COSMOS-Software, Jena, Germany

<sup>c</sup> Institute of Biological Interfaces 4, Karlsruhe Institute of Technology (KIT),  
Karlsruhe, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp01905d



For rigid molecular systems, it is convenient to describe the time and ensemble average observed in NMR experiments by introducing an extra entity, the alignment, or order tensor. This tensor describes the downscaling of tensorial interactions by motions and distributions of orientations. The exploration of the RDC is hampered by the fact that the preferred average orientation of the molecules is not known *a priori* and there is no straightforward way to extract the alignment tensor from the experimental data. For a given rigid molecule, the five independent elements of the alignment tensor can be derived from a system of linear equations with the experimental RDC as the solution vector (see *e.g.* Losonszki *et al.*<sup>6</sup>). For molecular systems with several possible conformers, this method is not practicable.

So far, MDOC was applied to membrane-bound molecules with relatively rigid backbones (Sternberg *et al.*<sup>7,8</sup>) and to helical peptides (Sternberg *et al.*<sup>9</sup>). These first investigations used mainly deuterium quadrupolar splitting tensors as orientational constraints. An MDOC investigation using <sup>1</sup>H–<sup>13</sup>C RDC as constraints was performed by Tzvetkova *et al.*<sup>10</sup> who applied the method to small molecules with a rigid backbone or moderate flexibility. Di Pietro *et al.*<sup>11</sup> performed similar investigations but additionally used <sup>13</sup>C–<sup>13</sup>C RDC to study the conformers and chirality of cellobiose. A simulation on a flexible macrolide system by Farès *et al.*<sup>12</sup> combined RDC data with NOE distances and <sup>3</sup>J couplings as constraints in MDOC. Here, the authors determined a high number of NMR data, including even prochiral hydrogen atoms; thus, MDOC has already proven its suitability for partially flexible molecules.

In this paper, the MDOC investigation of two molecules is presented: the five-membered ring lactone ( $\alpha$ -methylene- $\gamma$ -butyrolactone) and sagittamide A, which is produced by sea organisms called tunicates. The well-studied lactone provided the necessary initial tuning for the application of the MDOC simulations to molecules with several rotamers. The first investigation of the relative configuration and the conformers of the lactone using RDC was performed by Thiele *et al.*<sup>13</sup> and the MDOC simulations confirmed their outcome. The MDOC simulations on sagittamide A are the first calculations on a fully flexible molecule with a long chain linker. The chiral part consists of eight stereocenters, and six of them are located in a per-acetylated 1,2,3,4,5,6-hexahydroxyhexane unit. The high mobility caused by the single bonds connecting the carbons gives rise to a multitude of rotamers at ambient temperature. Since the measurements of Schuetz *et al.*<sup>14</sup> provided only 8 <sup>13</sup>C–<sup>1</sup>H RDC values, it seems impossible to single out the correct configuration from RDC alone since we had to cope with several conformers with individual orientations and exchange equilibria. In this paper, it is shown that in the case of sparse RDC data, the incorporation of additional constraints like NOE distances and <sup>3</sup>J couplings into MDOC simulations can provide definitive conclusions on chirality and conformer equilibria.

## Methods

### Calculation of pseudo-forces in MDOC simulations

To perform MDOC simulation, pseudo-energy terms were added to the energy expression of the COSMOS-NMR force field<sup>15</sup>  $E^{\text{FF}}$ .

The pseudo-energies contain expressions depending on the differences between the calculated and experimental dipolar tensors or other NMR parameters such as NOE distances and <sup>3</sup>J couplings.

$$E = E^{\text{FF}} + \frac{k_{\text{RDC}}}{2} \sum_k \sum_{\alpha\beta} (D_{\alpha\beta}^{\text{calc}_k} - D_{\alpha\beta}^{\text{exp}_k})^2 + \frac{k_{\text{R}}}{2} \sum_l (R_l^{\text{calc}} - R_l^{\text{exp}})^2 + \frac{k_{\text{J}}}{2} \sum_m (J_m^{\text{calc}} - J_m^{\text{exp}})^2 \quad (1)$$

The pseudo-energies are defined using harmonic potentials for the difference of experimental and calculated constraints. The forces acting on all atoms that are needed in MD simulations are calculated as Cartesian derivatives of the force field energies  $E^{\text{FF}}$  and pseudo-energies:

$$\vec{F}^i E = \frac{\partial^i}{\partial x_\gamma} E^{\text{FF}} + k_{\text{RDC}} f(\Delta D) \sum_k \sum_{\alpha\beta} (D_{\alpha\beta}^{\text{calc}_k} - D_{\alpha\beta}^{\text{exp}_k}) \frac{\partial^i}{\partial x_\gamma} D_{\alpha\beta}^{\text{calc}_k} + k_{\text{R}} f(\Delta R) \sum_l (R_l^{\text{calc}} - R_l^{\text{exp}}) \frac{\partial^i}{\partial x_\gamma} R_l^{\text{calc}} + k_{\text{J}} f(\Delta J) \sum_m (J_m^{\text{calc}} - J_m^{\text{exp}}) \frac{\partial^i}{\partial x_\gamma} J_m^{\text{calc}} \quad (2)$$

$\vec{F}^i$  is the vector of Cartesian coordinate derivatives  $\partial^i/\partial x_\gamma$ ,  $\gamma = \{x, y, z\}$  with respect to the coordinates of all atoms *i*. In geometry optimizations with NMR constraints, the harmonic form of the pseudo-energy as given in eqn (1) is useful, especially if we are near to a pseudo-energy minimum. For the situation in MD simulations far from the minima, the harmonic potential causes constantly growing forces that lead to unrealistic structures and motions. These unrealistically large pseudo-forces are avoided by multiplying their values by weighting factors  $f(\Delta q)$  ( $q$  stands for  $D$ ,  $R$  or  $J$ ), which gives rise to nearly constant forces if the difference,  $q^{\text{calc}} - q^{\text{exp}}$ , exceeds a threshold  $\Delta q$  (see Witter *et al.*<sup>16</sup>):

$$f(\Delta q) = \frac{p - p^{-1}}{p + p^{-1}}, \quad p = e^{\frac{q^{\text{calc}} - q^{\text{exp}}}{\Delta q}} \quad (3)$$

For  $q^{\text{calc}} - q^{\text{exp}} < \Delta q$ , this factor behaves in a similar manner to the derivative of the original energy expression (1). It is possible to choose different force constants  $k_q$  (eqn (1)) and width parameters  $\Delta q$  for each type of NMR property. Practically,  $\Delta q$  is chosen in the order of magnitude of the experimental error (see Tables S2 and S6 in the ESI†).

The second term in eqn (2) represents the orientational pseudo-forces containing the derivatives of all tensor elements of the calculated dipolar tensor  $\mathbf{D}^{\text{calc}}$ . These orientational pseudo-forces are evaluated as Cartesian derivatives of the transformation matrices that transform the bond dipolar tensors into the laboratory system:

$$F_{X(A_i B_i)}^i = k_{\text{RDC}} f(\Delta D) \sum_{\alpha\beta} (D_{\alpha\beta}^{L_i} - D_{\alpha\beta}^{\text{exp}_i}) \frac{\partial}{\partial X(A_i B_i)} D_{\alpha\beta}^{\text{calc}_i} \frac{\partial}{\partial X(A_i B_i)} D_{\alpha\beta}^{L_i} = \sum_{\alpha'\beta'} \left( T_{\beta\beta'}^i \frac{\partial}{\partial X(A_i B_i)} T_{\alpha\alpha'}^i + T_{\alpha\alpha'}^i \frac{\partial}{\partial X(A_i B_i)} T_{\beta\beta'}^i \right) D_{\alpha'\beta'}^{\text{calc}_i} \quad (4)$$



The transformation matrices  $\mathbf{T}^i$  are constructed from the unit vectors that define the actual orientation of the dipolar A-B systems: one unit vector,  $\mathbf{e}_3$ , in the A-B direction and two,  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , perpendicular to the other unit vectors. The transformation matrix  $\mathbf{T}$  is then constructed with the unit vectors as columns  $\mathbf{T} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ . The pseudo-forces in the Cartesian directions  $x, y$  and  $z$  (denoted with the index  $x_{(A,B,i)}$ ) are acting on the atom pairs A and B (denoted with  $i$ ; for the form of the derivatives see Sternberg *et al.*<sup>7</sup>).

### Calculation of the dipolar coupling tensors

For the dipolar coupling of two magnetic nuclei, A-B, the interaction tensor  $\mathbf{D}^{AB}$  (A and B denote the coupling atoms) will be diagonal and axial within the coordinate system connecting the two nuclei and we can easily assign one principal value of  $\mathbf{D}^{AB}$  to the axis parallel to the connection line and the other two values perpendicular to it:

$$\mathbf{D}^{AB} = \begin{pmatrix} -D^{AB}/2 & 0 & 0 \\ 0 & -D^{AB}/2 & 0 \\ 0 & 0 & D^{AB} \end{pmatrix} \quad (5)$$

In eqn (5),  $D^{AB}$  is the dipolar coupling constant depending on the gyromagnetic ratios of nuclei A and B and the distance between them.

NMR measurements of uniaxially oriented media with the director of the oriented sample oriented parallel to the magnetic field of the spectrometer the dipolar tensor within this laboratory frame will have the same form as eqn (5), but  $D^{AB}$  has to be replaced by the RDC value. Since the rotational diffusion around the director is isotropic, the time-averaged tensor becomes diagonal and rotationally symmetric. This must be true for every measured site and, therefore, it is of great value to use the whole tensors as constraints and not only the RDC value. The off-diagonal zeros of the RDC tensor indicate rotations and if used as constraints, molecular rotations are driven by differences to the calculated tensors.

The tensor components of  $\mathbf{D}^{AB}$  in the molecular frame can be transformed by proper rotation matrices to yield  $\mathbf{D}^L$  in the common laboratory frame according to

$$D_{\alpha\beta}^{L,i} = \sum_{\alpha'\beta'} T_{\alpha\alpha'}^i T_{\beta\beta'}^i D_{\alpha'\beta'}^i, \quad (6)$$

where  $\mathbf{T}^i$  is the transformation matrix that connects the local A-B coordinate system with the common laboratory frame (for the calculation of the transformation matrices see Sternberg *et al.*<sup>7</sup>). For every investigated site, we get separate transformation matrices  $\mathbf{T}^i$  where  $i$  is the index of the atom pair A-B under study. In NMR experiments, a time and ensemble average was observed and we obtained the time average of  $\mathbf{D}^{\text{calc}}$  using eqn (6):

$$\langle D_{\alpha\beta}^{L,i} \rangle_t = \sum_{\alpha'\beta'} \langle T_{\alpha\alpha'}^i T_{\beta\beta'}^i \rangle_t D_{\alpha'\beta'}^i \quad (7)$$

In the case of bond dipolar couplings tensor  $\mathbf{D}^i$ , it can be assumed that the tensors do not change with time (bond

vibrations are neglected) and we have to perform time averaging only over the products of the transformation matrix elements  $T_{\alpha\alpha'} T_{\beta\beta'}$ . In the NMR experiments used in this paper, the director is oriented into the  $z$ -direction of the laboratory system and the observed dipolar splittings can then be directly calculated from the  $zz$ -component of the time-averaged dipolar tensor:

$$\Delta\nu_i = 2S_{\text{am}} \langle D_{33}^{L,i} \rangle \quad (8)$$

In eqn (8), we introduced in analogy to Marsan *et al.*<sup>17</sup> an order parameter of the alignment medium  $S_{\text{am}}$  (this equation is only valid for the case of weak coupling). The experimental values represent an ensemble average over a large number of molecules and a time average. The relevant time scale for averaging is given by the inverse of the frequency range, which results from the anisotropy of the considered interaction. With motions faster than this time scale, only the time average of the NMR parameters is observed. As Torda and van Gunsteren<sup>4</sup> pointed out, the arithmetic mean value over the ever-growing period of the MD simulation becomes progressively less sensitive to changes within the system. Therefore, the time average should be performed using an exponential memory function:

$$\begin{aligned} \mathbf{D}^i(t) &= \frac{1}{N(t)} \int_{t'=t_0}^t e^{-\frac{(t-t')}{\tau}} \mathbf{D}^i(t') dt' \\ N(t) &= \int_{t'=t_0}^t e^{-\frac{(t-t')}{\tau}} dt' \end{aligned} \quad (9)$$

The memory time constant is denoted by  $\tau$  and  $N$  is the norm of the integral. This memory function will ensure that contributions from the beginning of the MDOC calculation will be exponentially scaled down. Since in MD simulations the equations of motion are integrated in finite time steps  $\Delta t$ , the integral in eqn (9) is converted into a discrete sum (see ESI†). During the MD simulation, the sum  $\langle \mathbf{D}^i \rangle_{(k+1)\Delta t}$  of the time step  $t = (k+1)\Delta t$  is calculated from the previous time step  $\langle \mathbf{D}^i \rangle_{k\Delta t}$  in a recursive manner (the index  $i$  for the site is dropped, see also ESI†):

$$\begin{aligned} \langle \mathbf{D} \rangle_{t=0} &= \mathbf{D}_0 \\ N_0 &= 1 \\ N_{k+1} &= N_k e^{-\frac{\Delta t}{\tau}} + 1 \\ \langle \mathbf{D} \rangle_{(k+1)\Delta t} &= \frac{1}{N_{k+1}} \left( \mathbf{D}_{k+1} + N_k \langle \mathbf{D} \rangle_{k\Delta t} e^{-\frac{\Delta t}{\tau}} \right) \end{aligned} \quad (10)$$

In the course of the MDOC simulation,  $\langle \mathbf{D}^i \rangle_t$  contains the exponentially weighted time average of the property  $\mathbf{D}^i$  and in the case of a tensor, the time averaging is performed with all components. The type of mean values as calculated using eqn (9) and (10) are still functions of time and consequently, the final result of the MDOC simulations has to be obtained by an arithmetic average over the  $\langle \mathbf{D}^i \rangle_t$  values of the trajectory.

The introduction of the time average with exponential memory according to eqn (9) effectively introduces a new timescale for rotational reorientations and fluctuations. This new characteristic time  $\tau$  can be regarded as the lifetime of a



molecular orientation. Using eqn (9), we have only to wait for a time period in the order of the memory time  $\tau$  until a difference between constraint and time mean value builds up and the orientational pseudo-forces are flipping the system into another state. Therefore,  $\tau$  is now the timescale of our accelerated MDOC simulation. This mode of time averaging, eqn (9) and (10), is also used for all types of properties where running mean values are needed to define pseudo-forces in constraint MDOC simulations, especially for NOE distances and  $^3J$  couplings.

The magnitude of the pseudo-forces is governed by the differences between experimental values and mean values calculated according to eqn (9) or (10) but the direction of the orientational pseudo-forces is calculated from derivatives of the actual transformation matrices (see eqn (4)). In this way, any mismatch in the orientation triggers an immediate response proportional to the mismatch. In the first timespan of MDOC simulation, no mean values developed and the resulting strong pseudo-forces could distort the molecule. Therefore, we switched on the pseudo-forces in exponential fashion using a pre-factor  $1 - e^{-t/\rho}$  with  $\rho$  the time constant for the rise of the pseudo-forces. This value was set in our simulation to the same value as the memory time  $\tau$  (see eqn (9) or (10)).

### Calculation of NOE distances

In the calculation of NOE distances during an MD simulation, one has to take into account motional effects influencing the mean values. This can be performed by so-called  $r^{-6}$  averaging (see Vögel<sup>18</sup>). For the case where the internal distance between the nuclei  $i$  and  $j$  changes slower than the overall tumbling of the molecule, we obtain

$$\langle r_{ij}(t) \rangle = \left\langle \frac{1}{r_{ij}^6} \right\rangle^{-1/6} \quad (11)$$

Combination of this mean value with the time average as given in eqn (9) results in the NOE mean value as defined by Torda and van Gunsteren:<sup>4</sup>

$$\langle r_{ij}(t) \rangle = \left( \frac{1}{N(t)} \int_{t'=t_0}^t e^{-\frac{(t-t')}{\tau}} \frac{1}{r_{ij}(t')^6} dt' \right)^{-\frac{1}{6}} \quad (12)$$

For the case of a rapid rotating  $\text{CH}_3$ , it is assumed that the distances to the protons of the group fluctuate faster than the overall tumbling. In MDOC simulations, by default, eqn (11) and (12) are used for calculating NOE distances; however, for the distances to the protons of  $\text{CH}_3$  groups, we used, in accordance with Vögel<sup>18</sup>, the  $r^{-3}$  averaging and a modified eqn (12).

### Calculation of $^3J$ couplings

To determine torsion angles around carbon-carbon single bonds, Karplus<sup>19</sup> developed an equation describing the dependence of  $^3J_{\text{HH}}$  couplings from the torsion angle  $\phi$ , introducing three empirical parameters  $P_1$ ,  $P_2$  and  $P_3$ . To account for the dependence of the  $^3J_{\text{HH}}$  couplings on influences of substituents in the surroundings of the bond being studied, Haasnoot *et al.*<sup>20</sup>

introduced correction terms into the Karplus equation and this approach is called the Altona equation:

$$^3J_{\text{HH}} = P_1 \cos^2 \phi + P_2 \cos \phi + P_3 + \sum_i \Delta\chi_i \{ P_4 + P_5 \cos^2(\xi_i \phi + P_6 |\Delta\chi_i|) \} \quad (13)$$

The polarising influence of the neighbors of the coupling protons is described by terms with electronegativity differences  $\Delta\chi$  and additional empirical parameters  $P_4$ ,  $P_5$  and  $P_6$  (for the sign  $\xi$  see ref. 20).

It was shown by Palermo *et al.*<sup>21</sup> that a similar expression to eqn (13) can also be derived for  $^3J_{\text{CH}}$  couplings and this expression contains an extra electronegativity term with the parameters  $P_4'$ ,  $P_5'$  and  $P_6'$ . The total number of 9 empirical parameters was determined in the case of Palermo *et al.*<sup>21</sup> by quantum chemical calculations of molecular models. These augmented Karplus expressions allow the predictions of  $^3J_{\text{HH}}$  and  $^3J_{\text{CH}}$  couplings with an error below 1 Hz and both dependences were implemented in the COSMOS-NMR force field. Additionally, we introduced Cartesian coordinate derivatives of the  $^3J$  couplings as prerequisites for the calculation of pseudo-forces in the MDOC simulations. The functional form of the pseudo-forces of the  $^3J$  couplings with the pre-factor  $f(\Delta f)$  is given in eqn (2).

### Data evaluation

All possible configurations of the molecule under study were calculated in MDOC simulations with the same experimental data and MDOC conditions. To measure which of the configurations best fulfills the constraints, we used the quality criterion introduced by Intelmann *et al.*<sup>22</sup> The quality  $Q = n/\chi^2$  is calculated from the  $\chi^2$  value:

$$\chi^2 = \sum_i^n \left( \frac{q_i^{\text{exp}} - q_i^{\text{calc}}}{\text{Error}_i} \right)^2 \quad (14)$$

The quality  $Q$  should be larger than one ( $n$  being the total number of experimental data) if the calculated data are, on average, within the experimental error bounds. The advantage of the quality parameter  $Q$  is that it includes the experimental or estimated errors and different types of experimental data can be easily compared.

## Results and discussion

### MDOC simulations of the five-membered ring lactone

The lactone molecule ( $\alpha$ -methylene- $\gamma$ -butyrolactone) serves as a simple model since it has one flexible five-membered ring and a rotating  $\text{CH}_3$  group at ambient temperature. Thiele *et al.*<sup>13</sup> studied this example and constructed conformer models of the ring with *RS* (C2-*R* and C3-*S*) and *SS* configuration. The geometries of the models were optimized on the DFT and MP2 level of theory and the resulting dipolar couplings were investigated using alignment tensors. The relative configuration was determined to be C2-*R* and C3-*S* (see left panel of Fig. 1).





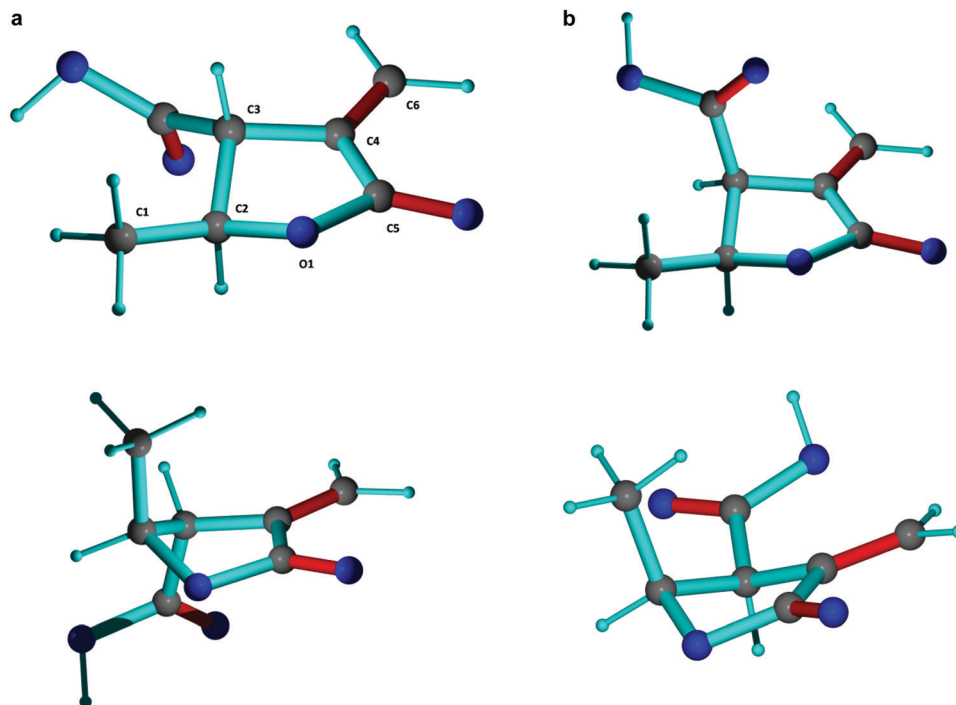


Fig. 1 (a) *RS* (C2-*R* and C3-*S*) configuration in two conformations of the 5-membered ring (upper and lower panel – denoted as *trans*-1a and *trans*-2b in ref. 26 and **2trans1** and **2trans2** in ref. 24). (b) *SS* (C2-*S* and C3-*S*) configuration in two conformations of the 5-membered ring (upper and lower panel – denoted as *cis*-1a and *cis*-2a in ref. 26).

In a subsequent investigation by Thiele *et al.*<sup>23</sup> it was demonstrated that in addition to the relative configuration, the population of the two ring conformers (upper and lower panels of Fig. 1) can also be determined. The possibility of measuring highly precise NOE distances provided the necessary data for the study of the ring conformers. Kolmer *et al.*<sup>24</sup> used newly determined NOE distances to re-investigate the population of the conformers of the lactone.

We used the lactone example taking the NMR data as published by Thiele *et al.*,<sup>13</sup> and the NOE distances as given by Kolmer *et al.*,<sup>24</sup> as constraints for the MDOC simulations of the *RS* and the *SS* forms. For the evaluation of the quality of the MDOC result, we used the errors as given in the supplements of the paper,<sup>24</sup> with one exception: the errors of the NOE distances were estimated according to the rules as given by Butts *et al.*<sup>25</sup> Distances lower than 2.8 Å can be determined with an error 0.05 Å and longer distances up to 4.5 Å with an error of 0.11 Å.

The MDOC method is nearly independent of the conformer at the start and this feature was demonstrated by using both configurations with different conformers as the initial geometry (see Fig. 1). The final results should only depend on the configuration, providing the simulation is sufficiently long. In the course of the MDOC simulation, all possible conformers are generated and their populations depend on the constraints. Since all internal and total molecular rotations are driven by orientational pseudo-forces, these orientational constraints are necessary to generate the conformers and even CH<sub>3</sub>-groups will rotate if proper CH one-bond RDCs are present. NOE distances

to the protons of CH<sub>3</sub> groups can be used in a general way since their fast rotation simulates experimental mean distances.

For the lactone structures as shown in Fig. 1a and b, 80 ns MDOC simulations at 300 K were performed (for the details of the simulations see Tables S1 and S2 in the ESI†). In a preliminary simulation, only the 8 RDC constraints with experimentally-determined signs of the couplings were used as constraints and the other 8 RDC without experimental signs were predicted. For the latter values, an RMS deviation of 0.67 Hz was obtained and the signs as predicted by Thiele *et al.*<sup>13</sup> were confirmed (see Table S4 in the ESI†).

The results of the final MDOC simulations of the five-membered lactone structures are demonstrated in Fig. 2 and as expected, the two simulations with two different conformers (Fig. 1a) developed into highly similar trajectories with almost identical final results. A clear difference can be observed between the two configurations 1 and 2 (see Fig. 1a and b). The quality values for the one bond RDC differ strongly with respect to the configurations. This confirms the results of Thiele *et al.*,<sup>13</sup> that the relative configuration is (C2-*R*, C3-*S*) or **2trans**.

The quality criteria  $n/\chi^2 > 1$  for the valid configuration indicate that most calculated values are within the error bounds of the constraints (see also Tables S4–S6 in ESI†) but for the NOE distances, an  $n/\chi^2$  of only 0.5 could be reached. This was caused by the calculated distances of H2 to the CH<sub>3</sub> protons deviating by 0.1 Å from the constraints (the error was estimated to 0.05 Å). A critical outlier is also the difference between the experimental and calculated H2 H3 distances of 0.186 Å.



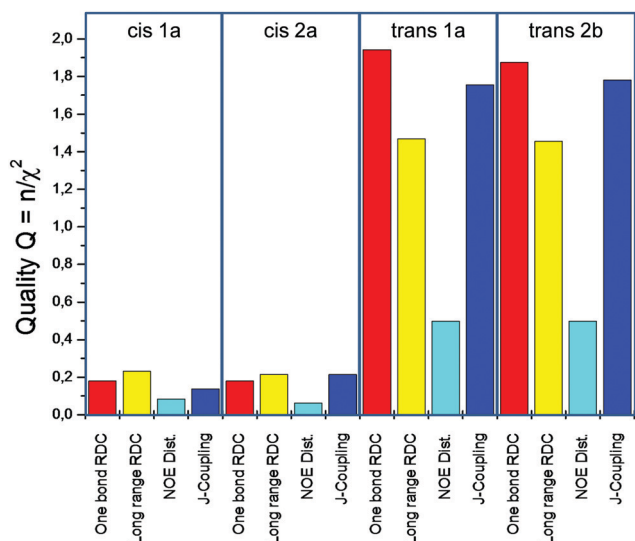


Fig. 2 Quality statistics of the four MDOC simulations of the five-membered ring lactone (see eqn (14)). The values were calculated as mean values over 2000 data snapshots of time-dependent values obtained with the exponential memory function.

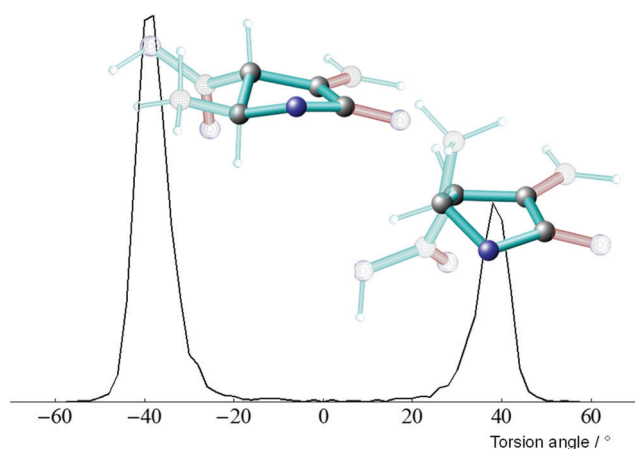


Fig. 3 Torsion statistics of the 5 ring of the lactone configuration *RS* (the conformers denoted **2trans1** and **2trans2**) around the C2–O1 bond (C3–C2–O1–C5). The ratios of (*gauche* – **2trans1**, *gauche* + **2trans2**) are (0.667, 0.333).

As can be seen in Fig. 3, the five-membered ring is mostly in the conformation with a (C3–C2–O1–C5) torsion angle of  $-40^\circ$ . This conformation, as shown in Fig. 3, leads to a *trans*-conformation of the hydrogen atoms H2 and H3. During the MDOC, it jumps into the other ring conformation with torsion angles (C3–C2–O1–C5) around  $40^\circ$ , which leads to a contribution of this conformation of 33%. Based on the RDC analysis using only one alignment tensor (MCST method), Thiele *et al.*<sup>23</sup> determined a population of  $(60 \pm 20\%)$  for the **2trans1** conformer. The improved analysis with two alignment tensors (MCMT method) led to a population of  $65 \pm 7\%$ . Using selected NOE distances, a population of 71% for **2trans1** was determined.<sup>24</sup> From the sampling, the dihedrals of the geometry snapshots

generated within the MDOC trajectory a population of 67% for **2trans1** was reached. This value is in perfect agreement with previous studies using different methods.

### MDOC simulations of sagittamide A

In the previous subsection, it was demonstrated on a simple example that MDOC simulations can be applied for the simultaneous determination of the configuration and conformation of flexible fragments in organic molecules. Sagittamide A is much more challenging since this molecule has a long flexible chain linker with a number of chiral centres. NMR data of sagittamide A are available for the central part of the molecule denoted C4 to C10, containing 6 CH–O–C(CH<sub>3</sub>)=O groups linked by C–C $\sigma$  bonds (see Fig. 4). Schuetz *et al.*<sup>14</sup> determined for every group one  $^1\text{H}$ – $^{13}\text{C}$  RDC coupling and two values were obtained for the H(a,b)–C4 group. This means for the MDOC simulations that there is only one RDC coupling per flexible unit (except for C4). These data are not sufficient for a complete configurational and conformational study. Using the NMR data, Schuetz *et al.*<sup>14</sup> singled out 5 possible configurations (a–e) for the acetylated units (see Fig. 4). A multitude of conformers are possible and the RDC values alone are not sufficient to assign one configuration.

In addition to the RDC values, the authors<sup>14</sup> provided 6  $^3J_{\text{CH}}$  and 7  $^3J_{\text{HH}}$  couplings and 7 precise NOE distances of the central part of the molecule (see ESI†). At least two of the  $^3J$  couplings have intermediate averaged values, suggesting torsional equilibria for the involved bonds (C5–C6 and C6–C7). It will be demonstrated that the MDOC trajectories display equilibria of all conformers including their orientation and this is only possible if MDOC simulations reach the NMR average. All further conclusions on conformer equilibria can only be derived for the central part of the molecule were constraints are given.

The weight factors for the MDOC pseudo-forces were estimated in short preliminary simulations. It turned out that  $^3J$  constraints and distance constraints are both necessary since the orientational constraints are by far not sufficient to predict the orientations and conformational equilibria of this molecule. If one set of constraints was left out, most distances and  $^3J$ -couplings were far off the error limits. It is not possible to discriminate between the 5 configurations a–e using the RDC values alone. The criterion must be that all NMR data are reasonably represented. This is not the case for the  $^3J$ -couplings nor the NOE distances if the data are not used as constraints. Simulations that used the simpler Karplus equation instead of the Altona<sup>20</sup> or Palermo<sup>21</sup> refined versions failed to discriminate the 5 configurations.

To exclude possible dependences of the results on the simulation time, one should use the final duration of the MDOC simulation much larger than  $5\tau = 1.0$  ns; therefore, we chose 80 ns as the final duration. As can be seen from the trajectories of the calculated properties (see *e.g.* Fig. 5) for MDOC simulations, this simulation time is sufficient to give stable values for all simulated properties. All 5 configurations of sagittamide A run parallel on the HPC cluster of KIT and lasted 57 h and 25 min (total duration).



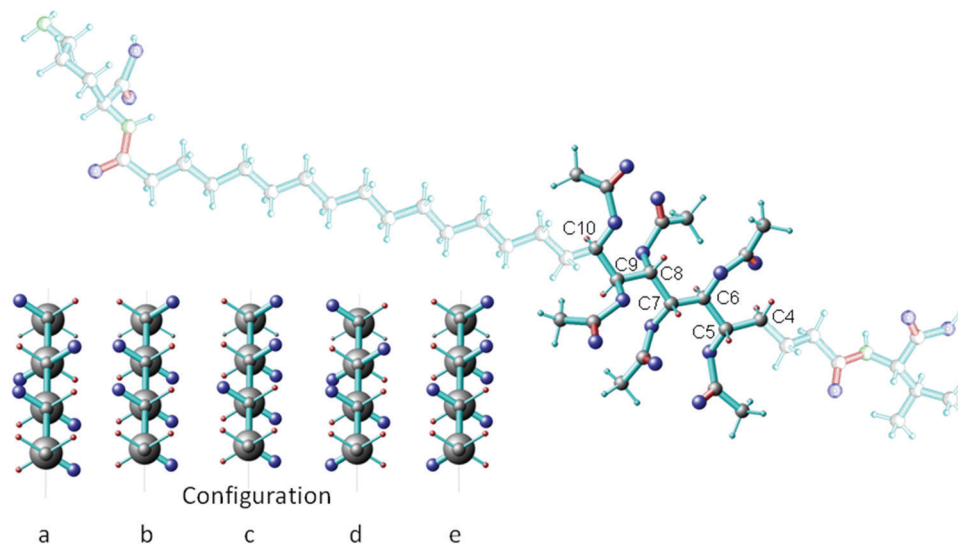


Fig. 4 The upper panel shows the molecular model of sagittamide A in configuration **a** with the numbering of the carbon atoms, where NMR data are taken into account. The hydrogen atoms for which  $^1\text{H}$ – $^{13}\text{C}$  dipolar couplings are measured<sup>14</sup> are shown in red. The lower 5 panels display the investigated part of the molecule in configurations **a** to **e**. Only C4 to C11 are displayed and the position of the acetyl group is shown in dark blue. The view of the projections is oriented with the C4–C5 bond (and additionally the C6–C7, C8–C9 and C10–C11 bonds) pointing backwards.

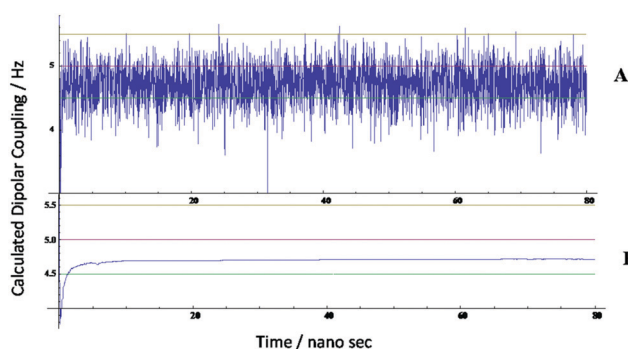


Fig. 5 The 80 ns trajectory of the dipolar coupling frequency of the  $^{13}\text{C}$ – $^1\text{H}$  coupling of the C4–H4a bond of sagittamide A. Large values at the beginning of the trajectories are cut off. The dipolar frequencies are scaled down by the order parameter of the alignment medium ( $S_{\text{am}} = 0.004$ ). Red line: Experimental value, green line: lower limit and yellow line: upper limit. Panel A: Mean value with exponential memory time (see eqn (9) and (10)) calculated at every step of the simulation. Panel B: Running arithmetic mean value (blue) of the data of panel A.

In panel A of Fig. 5, the typical trajectory of the  $zz$ -component of the dipolar tensor of the C4–H4a bond is displayed. The upper panel A displays a trajectory of the mean value with exponential memory as calculated using eqn (9). The memory time constant was  $\tau = 200$  ps and the actual coupling values were scaled down using an order parameter of the alignment medium of  $S_{\text{am}} = 0.004$ . In most of the simulation time we observed fluctuations in the calculated dipolar coupling of the order of magnitude of the experimental error. In the ideal case the value should fluctuate around the measured mean value (shown in red). In the lower panel B, the running arithmetic mean value of the data of panel A is displayed. After about 10 ns, we reached a stable mean value in a range within the experimental error

bounds. This means that in this case, after 10 ns, the molecule had sampled all global reorientations and nearly all conformational rotations were performed. Though these reorientations are clearly visible in panel A, throughout the simulation time there is hardly any detectable contribution to the final mean value. This behavior should be observed in a successful MDOC simulation and should be tested for all sites including  $^3J$ -couplings and NOE-distances.

Fig. 6 displays the quality criteria of the 5 configurations of sagittamide A calculated from 2000 coordinate snapshots of the 80 ns trajectories. The values in Fig. 6 were obtained by calculating the  $n/\chi^2$  values from averages over 2000 MDOC snapshots (for example, see the final value of trajectory B in Fig. 5).

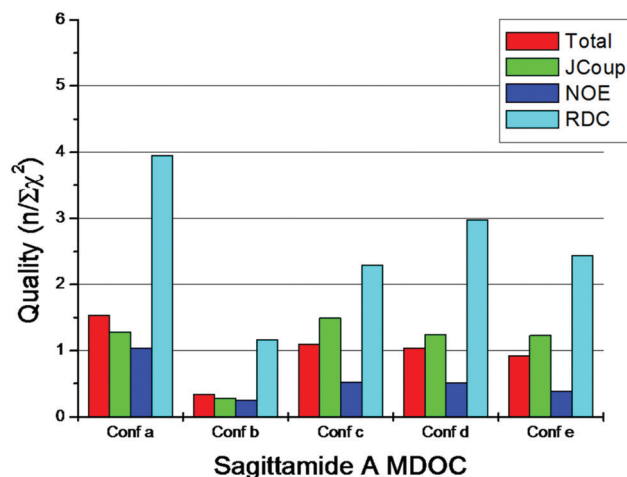


Fig. 6 Comparison of the quality  $Q$  (see eqn (13)) of the calculated data of an MDOC simulation with respect to the NMR experiment for the configurations **a**–**e** of sagittamide A.



A quality of  $n/\chi^2 > 1.0$  indicates that the calculated values are mostly within the experimental error bounds (see eqn (14)). As per the definition,  $\chi^2$  has no dimension; the sum in eqn (14) can be extended over different types of parameters, giving a total  $\chi^2$  and quality. Only configuration **a** fulfills the quality criterion for all three types of parameters and gives the highest total quality. For the quality criterion, the estimation of the experimental error should be performed with care. For the RDC values, all errors were set to 0.5 Hz. As can be seen from Fig. 6, all 5 simulations fulfilled the quality criterion for the RDC and we could not distinguish the configurations using the RDC alone; as discussed before, this is a consequence of the ambiguities in dipolar tensors. NOE distances had no ambiguities and their errors were estimated to 0.2 Å (except distance H8–H6 with the error of 0.4 Å, see ESI†). Only configuration **a** could fulfill the NOE quality criterion and this result emphasizes the need for precise NOE distances for flexible systems.

The  $^3J$  couplings are ambiguous and one would obtain the same value for the coupling constant for two torsion angles, but the *gauche*- and *trans*-states can clearly be discriminated. To estimate the error, we have to take into account the experimental error and the uncertainty of the prediction using the Altona equation. For  $^3J_{\text{HH}}$  couplings, the experimental error was estimated to be 0.6 Hz, and 1 Hz for the  $^3J_{\text{CH}}$  couplings (see Schuetz *et al.*<sup>14</sup>). Haasnoot *et al.*<sup>20</sup> obtained an RMS deviation of 0.65 Hz for the prediction of  $^3J_{\text{HH}}$  couplings and Palermo *et al.*<sup>21</sup> got a similar value of 0.662 Hz for the RMS deviation for DFT calculated coupling constants. Taking these errors together, the  $^3J$  couplings seem to be slightly weaker constraints than NOE distances; however, since these couplings depend directly on the torsion angles, the pseudo-forces derived from  $^3J$  couplings have a big influence on the population of rotamers.

### Analysis of the backbone dynamics of sagittamide A

The analysis of the dynamics in the long free chain is a crucial part of the elucidation of the configuration and conformation of the studied molecules. The best way to analyze the dihedral rotations that occur in the course of a MDOC trajectory is to look at the carbon atoms C4 to C10 and the 4 torsion angles in between (rotations about the bonds between C5–C6, C6–C7, C7–C8 and C8–C9). In the course of MDOC simulation for configuration **a**, the system performs numerous jumps around the bonds between C5 to C10 with a frequency in the order of magnitude of the memory time constant  $\tau$  (see eqn (6)). Since for the rest of the molecule no orientational constraints are given, the torsions around other bonds could not be investigated.

To analyze these trajectories, we counted the occurrence of angles for the whole trajectory and the relative ratios for the 4 torsion angles are given in Fig. 7. There was an obvious trend detectable in the data, namely, growing *trans* contributions when going from C5–C6 to C9–C10.

On analyzing the data, we can also ask whether some torsion states are coupled or independent. In Table 1 we give the relative populations of torsion states that add up to 77.7%. The trend that can be seen in Fig. 7 is also documented in Table 1: in 69% of all torsion states, the first two bonds after the

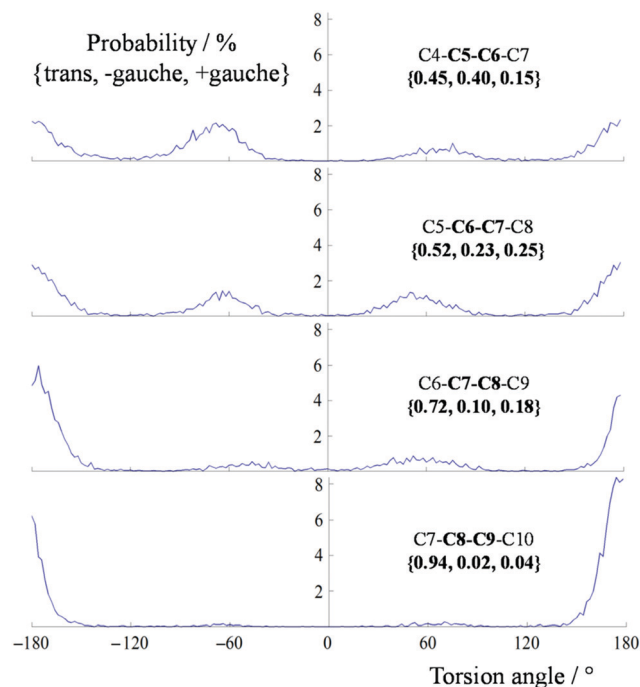


Fig. 7 Torsion angle distributions from C4 to C10 obtained from the MDOC simulation of sagittamide A in configuration **a**. For the atom labels, see Fig. 4.

aliphatic chain starting with C10 are in the *trans* orientation and in 94%, the first bond {C7–C8–C9–C10} is oriented *trans* (the complete analysis is given in the ESI†, Table S10).

In the MDOC simulations performed in this investigation,  $^3J_{\text{HH}}$  and  $^3J_{\text{CH}}$  are used as constraints and the calculated values gave a maximum deviation from the experiment of about 2 Hz only in the cases of the couplings C8–H10 and H9–H10 (see ESI†). Constraints like NOE distances and  $^3J$  couplings are necessary in nearly all the cases to gain information on the secondary structure of molecules.

From the analysis of couplings  $^3J_{\text{H5–H6}} = 4.8$  Hz and  $^3J_{\text{H6–H7}} = 7.1$  Hz, Schuetz *et al.*<sup>14</sup> concluded that these values were caused by conformational averaging around the C5–C6 and the C6–C7 bond, respectively. The C7–C8 bond was assumed to be fixed. Taking this data together with the H5 to H8 NOE distance of  $2.5 \pm 0.2$  Å, Schuetz *et al.*<sup>14</sup> estimated the population of H5–C5–C6–H6 and the H6–C6–C7–H7 rotamers, and this population was the prerequisite of the PALES<sup>27</sup> RDC analyses,<sup>14</sup> which indicated that the configuration **a** was the most probable. Finally, Seike *et al.*<sup>28</sup> confirmed the configuration **a** for the natural product of sagittamide A by total synthesis in combination with the analysis of the  $^3J_{\text{HH}}$  couplings.

Table 1 Highly populated combinations of torsion angles of the bonds C5 to C9 {C7–C8–C9–C10, C6–C7–C8–C9, C5–C6–C7–C8, C4–C5–C6–C7}

Probability/%	Torsion angle combination
22.0	{ <i>trans</i> , <i>trans</i> , <i>trans</i> , <i>gauche</i> }
18.0	{ <i>trans</i> , <i>trans</i> , <i>gauche</i> , <i>trans</i> }
14.6	{ <i>trans</i> , <i>trans</i> , <i>trans</i> , <i>trans</i> }
14.4	{ <i>trans</i> , <i>trans</i> , <i>gauche</i> , <i>gauche</i> }
8.7	{ <i>trans</i> , <i>gauche</i> , <i>trans</i> , <i>gauche</i> }





## Conclusions

It was demonstrated that orientational constraints in MDOC simulations generate a multitude of conformers of flexible molecules. In the case of sparse RDC data, pseudo-forces from scalar constraints as NOE distances or  $^3J$  couplings are necessary to adjust the population of these conformers. The time mean values of the RDC of these conformers fulfilled the measured RDC values within the error bounds. These conformers did not have to be inferred from other NMR data but are a direct result of the action of orientational pseudo-forces that reorient the whole molecule or parts of it. For the conformer population of the five-membered ring lactone, the MDOC simulations are in agreement with the results of Kolmer *et al.*<sup>24</sup> In contrast to the former investigations, no refined structures were necessary and alignment tensors were completely avoided.

In the case of sagittamide A, we faced a bigger challenge since this molecule can form a multitude of rotamers in solution. Nevertheless, the application of the MDOC simulations indicated that configuration **a** fulfills the RDC constraints at best, but only when all constraints are taken together did a clear decision for configuration **a** emerge. This result is in accordance with the previous studies by Schütz *et al.*<sup>14</sup> and Seike *et al.*<sup>28</sup> The advantage of MDOC comes from the possibility of blending several NMR data into a coherent picture of the structure and dynamics of the molecular system.

The most encouraging feature of MDOC simulations is the possibility to derive the occurrence of conformers from NMR measurements and to analyze their relative content. It should be mentioned that MDOC simulations are not dependent on the quality of the initial coordinates. The method opens up future prospects for tackling such questions as, for instance, the reaction of target molecules with flexible protein or nucleic acid receptors.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

We thank Burkhard Luy (KIT Karlsruhe) for promoting this work and for many inspiring discussions. We also thank Anne Schuetz for providing additional NOE data on sagittamide A. This work was performed thanks to the support of Karlsruhe Institute of Technology (KIT) and its computational resource bwUniCluster funded by the Ministry of Science, Research and Arts and the universities of the State of Baden-Württemberg, Germany.

## References

- 1 F. Separovitch, R. Pax and B. Cornell, NMR order parameter analysis of a peptide plane aligned in a lyotropic liquid crystal, *Mol. Phys.*, 1993, **78**, 357–369.
- 2 E. Brunner, Residual Dipolar Couplings in Protein NMR, *Concepts Magn. Reson.*, 2001, **13**, 238–259.
- 3 R. S. Lipsitz and N. Tjandra, Residual Dipolar Couplings in NMR Structure Analysis, *Annu. Rev. Biomol. Struct.*, 2004, **33**, 387–413.
- 4 A. E. Torda and W. F. van Gunsteren, The refinement of NMR structures by molecular dynamics simulation, *Comput. Phys. Commun.*, 1991, **62**, 289–296.
- 5 P. L. R. Markwick, S. A. Showalter, G. Bouvignies, R. Brüschweiler and M. Blackledge, Structural dynamics of protein backbone  $\phi$  angles: extended molecular dynamics simulations versus experimental  $^3J$  scalar couplings, *J. Biomol. NMR*, 2009, **45**, 17–21.
- 6 J. A. Losonszki, M. Andrec, M. W. F. Fischer and J. H. Prestegard, Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition, *J. Magn. Reson.*, 1999, **138**, 334–342.
- 7 U. Sternberg, R. Witter and A. S. Ulrich, All-atom molecular dynamics simulations using orientational constraints, *J. Biomol. NMR*, 2007, **38**, 23–39.
- 8 U. Sternberg and R. Witter, Molecular dynamics simulations on PGLa using NMR orientational constraints, *J. Biomol. NMR*, 2015, **63**, 265–274.
- 9 U. Sternberg, M. Klipfel, S. L. Grage, R. Witter and A. S. Ulrich, Calculation of Fluorine Chemical Shift Tensors for the Interpretation of Oriented  $^{19}\text{F}$ -NMR Spectra of Gramicidin A in Membranes, *Phys. Chem. Chem. Phys.*, 2009, **11**, 7048–7060.
- 10 P. Tzvetkova, U. Sternberg, T. Gloge, A. Navarro-Vázquez and B. Luy, Configuration Determination by Residual Dipolar Couplings: Accessing the Full Conformational Space by Molecular Dynamics with Tensorial Constraints, *Chem. Sci.*, 2019, **10**, 8774–8791.
- 11 M. E. Di Pietro, U. Sternberg and B. Luy, Molecular Dynamics with Orientational Tensorial Constraints: A New Approach to Probe the Torsional Angle Distributions of Small Rotationally Flexible Molecules, *J. Phys. Chem. B*, 2019, **123**(40), 8480–8491.
- 12 C. Farès, J. B. Lingnau, C. Wirtz and U. Sternberg, Conformational Investigations in Flexible Molecules using Orientational NMR Constraints in Combination with  $^3J$ -Couplings and NOE Distances, *Molecules*, 2019, **24**, 4417, DOI: 10.3390/molecules24234417.
- 13 C. M. Thiele, A. Marx, R. Berger, J. Fischer, M. Biel and A. Giannis, Determination of the Relative Configuration of a Five-Membered Lactone from Residual Dipolar Couplings, *Angew. Chem. Int. Ed.*, 2006, **45**, 4455–4460.
- 14 A. Schuetz, J. Junker, A. Leonov, O. F. Lange, T. F. Molinski and C. Griesinger, Stereochemistry of sagittamide A from Residual Dipolar Coupling Enhanced NMR, *J. Am. Chem. Soc.*, 2007, **129**, 15114–15115.
- 15 M. Möllhoff and U. Sternberg, Molecular mechanics with fluctuating atomic charges – a new force field with a semi-empirical charge calculation, *J. Mol. Model.*, 2001, **7**, 90–102.
- 16 R. Witter, W. Prieß and U. Sternberg, Chemical Shift Driven Geometry Optimization, *J. Comput. Chem.*, 2001, **23**, 298–305.



- 17 M. P. Marsan, I. Muller, C. Ramos, F. Rodriguez, E. J. Dufourc, J. Czaplicki and A. Milon, Cholesterol orientation and dynamics in dimyristoylphosphatidylcholine bilayers: a solid state deuterium NMR analysis, *Biophys. J.*, 1999, **76**, 351–359.
- 18 B. Vögeli, The nuclear Overhauser effect from a quantitative perspective, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2014, **78**, 1–46.
- 19 M. Karplus, Contact Electron-Spin Coupling of Nuclear Magnetic Moments, *J. Chem. Phys.*, 1959, **30**, 11–15.
- 20 C. A. G. Haasnoot, F. A. A. M. de Leew and C. Altona, The Relationship between Proton-Proton NMR Coupling Constants and Substituent Electronegativities, *Tetrahedron*, 1980, **36**, 2783–2792.
- 21 G. Palermo, R. Riccio and G. Bifulco, Effect of Electronegative Substituents and Angular Dependence on the Heteronuclear Spin-Spin Coupling Constant  $^3J_{C-H}$ : An Empirical Prediction Equation Derived by Density Functional Theory Calculations, *J. Org. Chem.*, 2010, **75**, 1982–1991.
- 22 D. Intelmann, G. Kummerlöwe, G. Haseleu, N. Desmer, K. Schulze, R. Fröhlich, O. Frank, B. Luy and T. Hofmann, *Chemistry*, 2009, **15**, 13047–13058.
- 23 C. M. Thiele, V. Schmidts, B. Böttcher, I. Louzao, R. Berger, A. Maliniak and B. Stevensson, On the Treatment of Conformational Flexibility when Using Residual Dipolar Couplings for Structure Determination, *Angew. Chem., Int. Ed.*, 2009, **48**, 6708–6712.
- 24 A. Kolmer, L. J. Edwards, I. Kuprov and C. M. Thiele, Conformational analysis of small organic molecules using NOE and RDC data: A discussion of strychnine and  $\alpha$ -methylene- $\gamma$ -butyrolactone, *J. Magn. Reson.*, 2015, **261**, 101–109.
- 25 C. P. Butts, C. R. Jones, E. C. Towers, J. L. Flynn, L. Appleby and N. J. Barron, Interproton distance determinations by NOE – surprising accuracy and precision in a rigid organic molecule, *Org. Biomol. Chem.*, 2011, **9**, 177–184.
- 26 C. M. Thiele, A. Marx, R. Berger, J. Fischer, M. Biel and A. Giannis, Determination of the Relative Configuration of a Five-Membered Lactone from Residual Dipolar Couplings, *Angew. Chem. Int. Ed.*, 2006, **45**, 4455–4460.
- 27 M. Zweckstetter and A. Bax, Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR, *J. Am. Chem. Soc.*, 2000, **122**, 3791–3792.
- 28 H. Seike, I. Ghosh and Y. Kishi, Stereochemistry of sagittamide A: Prediction and Confirmation, *Org. Lett.*, 2006, **8**, 3865–3868.

