Karlsruher Institut für Technologie

Prediction of Residual Dipolar Couplings in analyte molecules aligned by Poly-γ-benzyl-L-glutamate

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

David Elsing

Tag der mündlichen Prüfung: 20.12.2024
Referent: Prof. Dr. Wolfgang Wenzel
Korreferenten: Prof. Dr. Burkhard Luy und Dr. Mariana Kozlowska

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is an essential tool for the determination of molecular structures. However, due to the identical chemical environment of enantiomers and to a lesser extent diastereomers, the determination of relative and especially absolute configuration remains challenging. A promising method is the measurement of residual dipolar couplings (RDCs), which arise due to the partial alignment of analyte molecule orientations by an alignment medium. If the alignment of chiral analyte molecules is caused by a chiral alignment medium, the measured RDCs between the enantiomers differ. Previous studies showed that an approximate prediction of RDCs suitable for the determination of relative configuration is possible in some cases. In the case of the absolute configuration however, only tentative evidence exists that the assignment of enantiomers based on the prediction of RDCs is feasible.

In this thesis, I use atomistic approaches for the development of computational schemes to predict the RDCs in small analyte molecules aligned by the poly-γ-benzyl-L-glutamate (PBLG) alignment medium in (deuterated) chloroform. This includes the detailed description of the interactions between the analyte molecules and PBLG, which lead to partial alignment. One approach are Molecular Dynamics (MD) simulations with explicit solvent, which provide detailed insights into the alignment interactions and, most importantly, indeed succeed in the correct assignment of enantiomers for molecules with hydrogen bond donors by comparing the differences between RDCs. For these molecules, hydrogen bonds to PBLG are essential for the alignment and the enantiodiscrimination. However, there are molecules whose RDCs by themselves do not agree with experiment and the MD simulations are computationally expensive. For this reason, I study the binding free energies to PBLG and evaluate different interaction models, including implicit solvation models. Using them in Monte Carlo (MC) calculations showed a pure Coulomb potential between PBLG and the analyte molecule to result in the best agreement with experiment. For the analyte molecules with O – H hydrogen bond donor, the agreement is also significantly better than with the MD simulations, although the strength of the alignment is severely overestimated. The difference between enantiomers in the MC simulations is very small however, showing the full MD simulations to be necessary for the determination of the absolute configuration.

The methods developed in this thesis are thus a significant step for the prediction of RDCs, which in the future may be used for the design of new alignment media. Additionally, the demonstrated theoretical approach can be already applied for the distinction of enantiomers if the differences of the RDCs are large enough.

Contents

1. Introduction

Insights into the structure of organic molecules is indispensable for a large variety of applications, e.g. drug design, chemical synthesis, material design or understanding biological processes. One of the most important experimental techniques for this purpose is nuclear magnetic resonance (NMR) spectroscopy, which allows for the non-destructive probing of nuclear spins inside molecules. For nuclei in different chemical environments, the magnetic shielding by the electrons varies, resulting in different resonance frequencies. The situation is more challenging when attempting to distinguish stereoisomers, where the chemical environment is almost identical, especially if done in a non-destructive way. However, the chirality of organic molecules plays an important role in many chemical reactions and biological processes. In particular, many drug molecules are chiral and their enantiomers can have different biological activity, e.g. by anesthetic/analgesic effects on pregnant women (by R-thalidomide) to teratogenic malformations of born children (by S-thalidomide) [114], although interconversion occurs in vivo in this case. Therefore, the determination of absolute configuration of chiral organic compounds became a crucial task in chemistry and it is extremely challenging up to now. Several methods for enantiomer differentiation or even measuring enantiomeric excess exist, with different limitations [81, 124]: If the compound studied can be purified and crystallized, X-ray diffraction is often used for distinguishing enantiomers [19]. However, the crystallization of many molecules is cumbersome or may not be possible. Additionally, for molecules with only "light" atoms, the necessary accuracy to determine significant diffraction intensity differences needed cannot be achieved [50]. Optical methods, such as electronic or vibrational circular dichroism (CD spectroscopy), are known to well distinguish enantiomers [16, 31], but multiple stereocenters in larger molecules often do not allow a sufficiently specific interpretation of spectra using state-of-the-art predictions. Using molecules in the gas-phase, spectroscopic methods such as Coulomb explosion imaging or microwave three-wave mixing may be also applied to determine their absolute configuration [122]. However, the former relies on the nearly complete fragmentation at the stereogenic center and is mostly suitable for very small molecules, while the latter is limited to molecules which have a significant dipole moment and can be vaporized.

NMR spectroscopy has been found as a valuable and less time-consuming method for determination of molecular configuration, including absolute configuration [146]. This makes it very important for pharmaceutical applications, including structure determination [103] or drug discovery [123]. Using chiral derivatizing agents (CDA), such as a mandelate, O-methylmandelate or α-methoxy-α-trifluoromethylphenylacetate (in Mosher ester analysis) [40], it can be determined from the chemical shifts of nuclei near the chiral center. However, the disadvantage of such an approach is that a derivatization with the

CDA (e.g. by the covalent bond formation) is required, which is only possible for certain classes of molecules. At the same time, the alignment of the chiral analytes by a (chiral) alignment medium in NMR allows the measurement of anisotropic NMR parameters [141, 142]. In a large number of studies, the importance of such parameters for the determination of relative configuration has been demonstrated [21, 84, 145, 164, 165, 116, 160, 159, 109]. For interacting nuclei whose movement is isotropic (such as free molecules in solution), the couplings depending on the anisotropy of their environment average out to zero, such as the chemical shift anisotropy, dipolar couplings or (electric) quadrupolar couplings (the latter for nuclei with spin ≥ 1). If the movement has however an isotropic component, such as for large biomolecules [22, 128], nematic liquid crystals [141, 161] or the alignment of molecules by an alignment medium [83, 90], these couplings do not average to zero and the effects of residual chemical shift anisotropy [59], residual dipolar couplings (RDCs) [91] and residual quadrupolar couplings can be observed in the NMR spectra.

Dipolar couplings are magnetic interactions between nearby nuclei, which depend on the distance between two nuclear spins i and j and the angle θ_{ij} of the internuclear vector $\vec{r}_{ij}$ to the static magnetic field (see Figure 1.1). RDCs are then averaged dipolar couplings and therefore proportional to $\langle(3\cos^2(\theta_{ij}) - 1)/r_{ij}^3\rangle$, where the average goes over a full experimental free induction decay (FID), which is typically larger than 100 ms. As mentioned, RDCs average to zero under isotropic movement and can therefore only be measured upon the introduction of anisotropy. While RDCs are used for structure prediction of proteins [72, 10] and other macromolecules [43], it is also possible to determine RDCs for analyte molecules, where the anisotropy is induced by the interaction with an alignment medium.

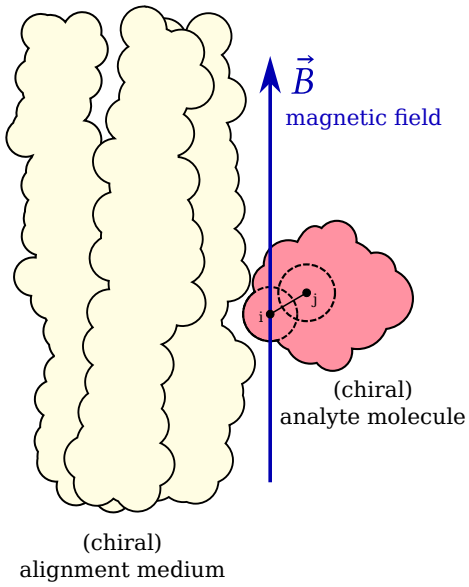


Figure 1.1.: Scheme of the partial alignment of an analyte molecules by the interactions with an alignment medium. The alignment medium is itself partially aligned along the magnetic field $\vec{B}$ and induces anisotropic motion of the analyte molecule, giving rise to residual dipolar couplings. For chiral alignment media and analyte molecules, differences between enantiomers can be observed.

It is in particular also well-established that RDCs upon alignment by a chiral alignment medium may, in principle, allow for the determination of the absolute configuration of the analyte [135, 134, 14], without the requirement of covalent bond formation. Moreover, such measurements are performed for dissolved molecules at temperatures close to room temperature[25], making the method usable also for compounds which cannot be crystallized or dissolved in water. For the same alignment medium and analyte molecule at the same temperature, the RDC values should be reproducible up to a constant scaling factor [158], making calculated and measured data reusable and not specific to the experimental setup. In addition, this NMR method was demonstrated to be used for measuring enantiomeric excess [49] and for the special case that similar molecules with known chirality are available, even the absolute configuration could be determined [176] using cross-fitting [85]. This alone is however not sufficient to determine the absolute configuration of a compound for which it is unknown, as a comparison to a previous measurement of both enantiomers (or with the other enantiomer of the alignment medium) would be required. This problem could be solved by a theoretical prediction for both enantiomers which can distinguish between them. While a theoretical predictions suitable for the distinction of diastereomers has been demonstrated using RDCs [116], the reliable assignment of the absolute configuration in this manner has not been shown before. Therefore, a more sophisticated theoretical treatment of and a detailed understanding of the alignment is necessary, which is the main objective for this thesis.

RDCs can be calculated by modeling the average orientation of $\vec{r}_{ij}$ depending on the molecular motion. For analyte molecules partially aligned by an alignment medium, this requires the description of the interactions with the alignment medium. Simple methods for the prediction of RDCs by an alignment medium include the averaging of the alignment by steric exclusion and electrostatic interaction with a (charged) wall or cylinder ("Prediction of alignment from structure", PALES) [179, 178] or approximations thereof [7, 3, 174]. For the alignment of analyte organic molecules in a nematic solvent, Pizzirusso et al. [127] used all-atom molecular dynamics (MD) simulations to predict the order parameter $\langle P_2 \rangle$, which was demonstrated to correlate with measured dipolar couplings in NMR. Frank et al. [55] have shown stochastic MD simulations using Langevin dynamics of strychnine aligned by a polystyrene/$CDCl_3$ gel to predict RDCs. They have shown a good correlation between simulated and experimental values, although the binding to PBLG may have been overestimated due to the lack of explicit solvent, which nonetheless led to a reasonable result for strychnine. Opakua et al. [116] demonstrated a relatively simple model (called P3D) for the estimation of RDCs for diastereomers of rigid molecules (e.g. 4,6-diacetylhygro-phorone A, strychnine or isopinocampheol) [116, 74]. There, the correct diastereomer used in experiment could be assigned using PBLG as alignment medium. This was accomplished by using a precomputed electrostatic potential of PBLG based on the Poisson equation and steric exclusion by modeling atoms as hard spheres. However, theoretical methods reported have not shown conclusive differences in RDCs of enantiomers usable for the assignment of absolute configuration for multiple molecules [14, 5], although preliminary studies have shown some success in limited cases [102, 136]. Even though the experimental RDCs of the enantiomers can, in principle, be reliably distinguished [104], a theoretical prediction in silico is necessary to determine the absolute

configuration of new compounds where measurement data for both enantiomers does not already exist. Moreover, it may permit an in-depth understanding of molecular bases of alignment and designing rules for the development of target functional alignment media for more conclusive measurements. However, there are several difficulties for theoretical predictions of RDCs, including the weakness of the alignment, i.e. the lack of specific interactions between the chiral molecule and the alignment medium [140], and the accuracy required to resolve the small differences in alignment between enantiomers [104, 136].

Although a large number of chiral alignment media is known [25, 6, 97], the most important one is the liquid crystalline phase poly-γ-benzyl-L-glutamate (PBLG), which formed the basis for most developments regarding the distinction of enantiomers [104]. It has the advantage that there is no lower limit to the alignment strength, as is the case for liquid crystals [96]. In this thesis, I thus focus on RDCs of small analyte molecules aligned by poly-γ-benzyl-L-glutamate (PBLG) in (deuterated) chloroform solution. It is a chiral alignment medium and forms a lyotropic liquid crystal with a stable α-helix [9, 138], such that its molecular structure for the use in simulations is known in detail [67]. My main goal is the understanding of the alignment in several analyte molecules by PBLG and the prediction of RDCs by simulating the alignment. Additionally, I show the applicability of these RDCs for the assignment of absolute configuration for analyte molecules with hydrogen bond donor.

The thesis is organized into three parts: In the first part, I explain the theoretical background relevant for the thesis, beginning with the basis of NMR spectroscopy, how RDCs influence NMR spectra and can be calculated from theory. This is followed by an introduction of molecular simulation methods, such as molecular dynamics (MD) simulations and Monte Carlo (MC) methods, as well as implicit solvation models. In the second part, I describe the methodology used to obtain the results. This includes the new MD simulations performed using an automated workflow for several pairs of enantiomers of analyte molecules, an implicit solvation model for the calculation of binding free energies, as well as an MC approach suitable for calculating RDCs with different models for the binding between the analyte molecules and PBLG. In the final part, the results from the MD simulations, the implicit solvation model and the MC calculations are presented. In particular, I compare the calculated RDCs to experimental data, and show the applicability of the MD simulations to the determination of the absolute configuration. Additionally, I show the impact of water molecules on the alignment of and the RDCs in isopinocampheol (IPC). For the binding free energies, I discuss the fit results of the implicit solvation model to the values obtained from MD free energy calculations. Lastly, I describe the MC results and compare the RDCs obtained with different binding models among each other and to RDCs from MD and experiment.

2. Theoretical background

In the following, I give an overview over residual dipolar couplings and how they can be obtained from NMR spectra. Afterwards, I describe the MD simulation techniques used for the simulations and free energy calculations. This is followed by an overview of the implicit solvation models I adapted for the approximation of binding free energies. The last part describes the Monte Carlo methods for the calculation of the average alignment without explicit solvent.

2.1. NMR concepts

As all the RDCs in this work are between ^{13}C and ^{1}H nuclei, both of which have spin $\frac{1}{2}$, I will only consider that case here. For spins $\frac{1}{2}$, the electric quadrupolar interaction is zero, such that the most relevant interactions are the chemical shift and the J- and dipolar couplings explained below.

For a single spin $\frac{1}{2}$ nucleus in a static external magnetic field $\vec{B}^{\text{ext}} = B^{\text{ext}}\vec{e}_z$, the Hamiltonian describing the magnetic interaction with the field is [89]

$$\hat{H}^{\text{Zeeman}} = -\hbar\gamma(\mathbb{1} - \boldsymbol{\delta})\vec{B}^{\text{ext}} \cdot \hat{\vec{I}}, \tag{2.1}$$

with the gyromagnetic ratio γ of the nucleus and the chemical shift tensor $\boldsymbol{\delta}$ describing the local environment of the nucleus. Here, $\hat{\vec{I}} = \sum_k \hat{I}_k \vec{e}_k$ with $k \in \{x, y, z\}$ is the spin operator of the nucleus. For spin $\frac{1}{2}$ nuclei, $\hat{I}_k = \frac{1}{2}\hat{\sigma}_k$ with the Pauli matrices $\hat{\sigma}_k$.

Because the movement of the molecules considered here is nearly isotropic, I will additionally assume that $\boldsymbol{\delta}$ is isotropic, i.e. $\boldsymbol{\delta} \approx \delta\mathbb{1}$, such that

$$\hat{H}^{\text{Zeeman}} = -\hbar\omega\hat{I}_z. \tag{2.2}$$

with the Larmor frequency $\omega = \gamma(1 - \delta)B^{\text{ext}}$.

Ensemble of spins

In a usual NMR sample, there are many identical molecules whose nuclei all contribute to the NMR signal. Additionally, the interactions between spins only act locally (such that interactions between spins in distant molecules can be neglected), while the relevant external interactions are the static magnetic field and radio frequency pulses.

The average behavior of the spin system can be described by the density operator $\hat{\rho}$, which describes a quantum mechanical ensemble. Like in the classical case, this means that each state $|\psi_i>$ (an element of a basis of the Hilbert space) is assigned an ensemble probability p_i, e.g. the Boltzmann distribution for the canonical ensemble. The density operator is then defined as $\sum_i p_i |\psi_i><\psi_i|$ and is therefore self-adjoint and fulfills $\mathrm{Tr}(\hat{\rho}) = 1$. Thus, its diagonal elements describe how the states are populated, while the off-diagonal elements describe the correlations between the states.

$\hat{\rho}$ is related to the (macroscopic) magnetization $\vec{M}$ by [89]

$$M_k = \frac{2}{c} \mathrm{Tr}(\hat{\rho}\hat{I}_k) \tag{2.3}$$

and, the other way around,

$$\hat{\rho} = \frac{1}{2}\hat{\mathbb{1}} + c\left(\sum_k M_k \hat{I}_k\right) \tag{2.4}$$

where $k \in \{x, y, z\}$ and c is a constant.

In thermal equilibrium density operator at a temperature T follows the canonical ensemble, where the population of each state is proportional to the Boltzmann factor [78]:

$$\hat{\rho}_{\mathrm{eq}} = \frac{e^{-\beta\hat{H}}}{\mathrm{Tr}\left(e^{-\beta\hat{H}}\right)}, \tag{2.5}$$

with the inverse temperature $\beta = 1/(k_{\mathrm{B}}T)$. This means that the equilibrium magnetization is aligned along the external magnetic field. Neglecting other interactions and the chemical shift anisotropy, the Hamiltonian is given by Equation (2.2). If the temperature is additionally high enough (e.g. at room temperature), $\hbar\omega \ll k_{\mathrm{B}}T$ and we can approximate

$$\hat{\rho}_{\mathrm{eq}} \approx \frac{1}{2}\hat{\mathbb{1}} + \frac{1}{2}\beta\hbar\omega\hat{I}_z. \tag{2.6}$$

This is (usually) the starting point of NMR experiments, on which radio frequency pulse sequences are applied to rotate the magnetization.

2.1.1. Couplings between bonded atoms

For a system consisting of pairs of spins (i, j), the corresponding Hilbert space is formed by the tensor product of the Hilbert spaces of single spins. The thermal equilibrium density operator for high temperatures is therefore

$$\hat{\rho}_{\mathrm{eq}} = \frac{1}{4}\hat{\mathbb{1}}_i \otimes \hat{\mathbb{1}}_j + \frac{1}{4}\hbar\omega_i \hat{I}_{i,z} \otimes \hat{\mathbb{1}}_j + \frac{1}{4}\hbar\omega_j \hat{\mathbb{1}}_j \otimes \hat{I}_{j,z} \tag{2.7}$$

and the Hamiltonian for uncoupled spins is given by

$$\hat{H} = \hbar\omega_i \hat{I}_{i,z} \otimes \hat{\mathbb{1}}_j + \hbar\omega_j \hat{\mathbb{1}}_i \otimes \hat{I}_{j,z}. \tag{2.8}$$

In the following, the tensor product sign $\otimes$ is omitted.

Dipolar coupling

The dipolar coupling between two spins can be described by the potential energy between two magnetic dipoles $\gamma_i \hbar \hat{\vec{I}}_i$ [89], i.e.

$$\hat{H}_{\text{dipolar}} = \hbar \underbrace{\left(-\frac{\hbar \gamma_i \gamma_j \mu_0}{4\pi r_{ij}^3} \right)}_{=:b_{ij}} \left(3(\hat{\vec{I}}_i \cdot \vec{e}_{ij})(\hat{\vec{I}}_j \cdot \vec{e}_{ij}) - \hat{\vec{I}}_i \cdot \hat{\vec{I}}_j \right), \tag{2.9}$$

where r_{ij} is the distance between the dipoles and the unit-vector $\vec{e}_{ij}$ their relative orientation. Its time average can be written as

$$\left\langle \hat{H}_{\text{dipolar}} \right\rangle_t = 3\hbar b_{ij} \left(\hat{\vec{I}}_i^{\top} \underbrace{\left(\left\langle \vec{e}_{ij} \vec{e}_{ij}^{\top} \right\rangle_t - \frac{1}{3} \mathbb{1} \right)}_{=:\boldsymbol{A}_{ij}} \hat{\vec{I}}_j \right), \tag{2.10}$$

where $\boldsymbol{A}_{ij}$ is an alignment tensor for the spin pair (i, j) (which is different from the alignment tensor $\boldsymbol{A}$ for a (rigid) molecule described below).

J-coupling

Due to the hyperfine interaction between the nuclei and the electrons in a chemical bond, there is a contribution to the Hamiltonian which depends on the relative orientation of the nuclear spins [78]. When the partial alignment is small, we can again neglect anisotropic contributions to the J-coupling and only consider the isotropic part:

$$\left\langle \hat{H}_{\text{J}} \right\rangle_t = \hbar J_{ij} \hat{\vec{I}}_i \cdot \hat{\vec{I}}_j. \tag{2.11}$$

Here, J_{ij} can be treated as constant and does not depend on the magnetic field or the chemical shift. J_{ij} does not depend on the magnetic field (or the chemical shift), nor on $\boldsymbol{A}_{ij}$. Its sign is the relative sign of the gyromagnetic ratios of the two nuclei. As I only use C – H couplings in this work, this means that the sign of J_{ij} is always positive.

2.1.2. Residual Dipolar Couplings

Secular approximation for weak coupling

If the off-diagonal elements of the total Hamiltonian $\hat{H}^{\text{Zeeman}} + \langle \hat{H}_J + \hat{H}_{\text{dipolar}} \rangle_t$ are small compared to the absolute difference between the eigenvalues of the eigenstates they

couple, they can be neglected in the case that a time-independent treatment is valid (*secular approximation*). This condition is the *weak coupling* [89] defined as

$$\frac{1}{2}|J_{ij} - 2D_{ij}| \ll |\omega_i - \omega_j|, \tag{2.12}$$

where D_{ij} is the residual dipolar coupling explained below. This is always the case for a pair of different nuclei, where ω_i and ω_j are very different, while for nuclei of the same isotope, they only differ due to the different chemical shift.

Thus we have the diagonal time-averaged secularly approximated Hamiltonian as an approximation to the total Hamiltonian for a pair of (weakly aligned) spins:

$$\hat{H} = \hbar \left[\omega_i \hat{I}_{i,z} + \omega_j \hat{I}_{j,z} + 2\pi(J_{ij} + 2D_{ij})\hat{I}_{i,z}\hat{I}_{j,z}\right]. \tag{2.13}$$

The last term is off-diagonal and leads to a splitting of the NMR spectrum, which I show in the following.

The RDC D_{ij} is defined as

$$D_{ij} = -\left\langle \frac{3\gamma_i\gamma_j\mu_0\hbar}{16\pi^2 r_{ij}^3}\left(\cos^2(\theta_{ij}) - \frac{1}{3}\right)\right\rangle_t, \tag{2.14}$$

where θ_{ij} is the angle between the internuclear vector $\vec{r}_{ij}$ and the static magnetic field $\vec{B}^{\mathrm{ext}}$ as illustrated in Figure 2.1. In a simulation, all that needs to be described for the prediction of an RDC is therefore $\vec{r}_{\mathrm{ij}}$ and its time-averaged relation to $\vec{B}^{\mathrm{ext}}$.

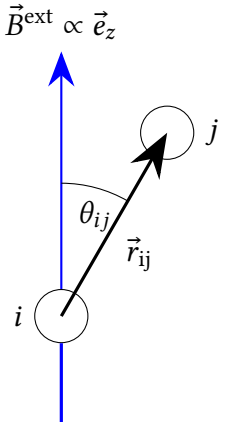


Figure 2.1.: Definition of the angle θ_{ij} used for the calculation of RDCs. $\vec{B}^{\mathrm{ext}}$ is the static magnetic field in NMR and $\vec{r}_{ij}$ is the distance between nuclei i and j.

Alignment tensor

Assuming that r_{ij} in Equation 2.14 is independent of the molecular orientation, which can be assumed if atoms i and j are bonded (i.e. D_{ij} is a ${}^1D$ coupling), it can be replaced by the

average bond length, which is approximately the bond length in the optimized geometry. If the analyte molecule is rigid, all θ_{ij} depend only on the molecular orientation and can be described by a common alignment tensor $\boldsymbol{A}$:

$$\langle \cos^2(\theta_{ij}) \rangle_t - \frac{1}{3} = \vec{r}_{ij}^{\top} \boldsymbol{A} \vec{r}_{ij}. \tag{2.15}$$

$\boldsymbol{A}$ is a symmetric and traceless tensor of order 2 and therefore has 5 independent components. With the rotation matrix $\boldsymbol{R}_t$ transforming the reference structure of the rigid molecule to the orientation at time t, it can be calculated as

$$\boldsymbol{A} = \frac{1}{3} \langle \boldsymbol{R}_t^{\top} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \boldsymbol{R}_t \rangle_t. \tag{2.16}$$

CH_3 groups

Even in relatively rigid molecules, the hydrogen atoms in methyl (CH_3) groups exhibit a fast rotation around the carbon atom [68]. This rotation is on a smaller timescale than the spectral timescale or the orientation of the molecule and can be expected to be almost independent from it. Therefore, the C – H couplings are the same for all hydrogen atoms, which cannot be distinguished in the NMR spectrum. Averaging the $\cos^2(\theta_{ij}) - 1/3$ term in Equation (2.14) over this rotation leads to a scaling by a factor of $\frac{1}{2}\left(3\cos^2(\varphi) - 1\right)$ compared to a C – H bond with the orientation of the X – C bond [139], where X is the atom the CH_3 group is connected to (see Appendix A.3 for a derivation) and ϕ is the X – C – H bond angle. With $\varphi = \arccos(-1/3)$ (the bond angle for an ideal tetrahedron), this results in a factor of $-1/3$. In reality, the bond angle may be slightly different, but the difference was shown to be small [117], especially compared to the relative experimental error of many CH_3 RDCs.

2.1.3. Influence of $J_{ij} + 2D_{ij}$ on the NMR spectrum

Time evolution of uncoupled spins $\frac{1}{2}$

The time evolution of $\hat{\rho}$ is described by the Liouville-von Neumann equation

$$i\hbar\dot{\hat{\rho}} = [\hat{H}, \hat{\rho}], \tag{2.17}$$

where $[\hat{H}, \hat{\rho}]$ is the commutator of $\hat{H}$ and $\hat{\rho}$. If $\hat{H}$ is time-independent, this is solved by

$$\hat{\rho}(t) = e^{-i(\hat{H}/\hbar)t} \hat{\rho}(t=0) e^{i(\hat{H}/\hbar)t}. \tag{2.18}$$

Starting from a transverse magnetization $\vec{M}(t=0) = M_x \vec{e}_x$ with $\hat{\rho}(t=0) = \frac{1}{2}\hat{\mathbb{1}} + c\hat{I_x}$, the time evolution under the static Hamiltonian (Equation (2.2)) is

$$\hat{\rho}(t) = \frac{1}{2}\hat{\mathbb{1}} + c(\cos(\omega t)\hat{I_x} + \sin(\omega t)\hat{I_y}), \tag{2.19}$$

i.e. a precession around the static magnetic field $\vec{B}$.

Due to (thermal) interactions with the environment, the longitudinal component of $\vec{M}$ returns to the thermal equilibrium over time. This process is called *spin-lattice* relaxation and is approximately exponential, with a time constant T_1 [89].

The transverse magnetization also decays over time due to a loss of coherence as e^{-t/T_2} with a time constant T_2 [89]. Because of small differences in the magnetic environment of different nuclei, the transverse component actually decays much faster because of the precession. However, this effect can be counteracted using spin-echo pulse sequences, where the magnetization is inverted, leading the transverse magnetization to increase again after some time, because the difference of the rotation between the spins was reversed.

Pulse sequences of radio frequency pulses are the basis of NMR spectroscopy, as they are the primary way to externally modify $\hat{\rho}$ in a variety of ways. In particular, they can achieve transverse magnetization, i.e. magnetization not aligned along the z-axis. This allows for an observable time evolution of $\hat{\rho}$ (by measuring the radio frequency signal due to the changing magnetization), which depends on the environment of and the interactions between nuclear spins.

When measuring a NMR spectrum, the signal is split into two parts at different phases using quadrature detection [8], which effectively results in a measurement of the magnetization in two direction, e.g. M_x and M_y. They can be combined into a complex signal [8], which for the static Hamiltonian is

$$M_x(t) + iM_y(t) \propto e^{-t/T_2} e^{i\omega t}. \tag{2.20}$$

Calculating the Fourier transform to obtain the NMR spectrum, the real part of this signal is a Lorentzian, which can be observed in the spectrum.

Time evolution of coupled spins $\frac{1}{2}$

The coupling terms leads to a transfer of transverse magnetization between the two spins during time evolution. To understand their effect on the time evolution of $\hat{\rho}$, it is useful to apply Equation (2.18) to the operators contributing to the initial $\hat{\rho}$ (see Equation (2.4)) [78]. To see this, consider as an example a contribution to the density operator $\hat{I}_{i,x}$ with the Hamiltonian in Equation (2.13). It evolves in time as [78]

$$\begin{aligned} \hat{I}_{i,x} \rightarrow e^{-t/T_2}\big(&\cos(\omega_i t)\cos(\frac{1}{2}T_{ij}t)\hat{I}_{i,x} + \cos(\omega_i t)\sin(\frac{1}{2}T_{ij}t)\hat{I}_{i,y}\hat{I}_{j,z} \\ &+ \sin(\omega_i t)\cos(\frac{1}{2}T_{ij}t)\hat{I}_{i,y} - \sin(\omega_i t)\sin(\frac{1}{2}T_{ij}t)\hat{I}_{i,x}\hat{I}_{j,z}\big), \end{aligned} \tag{2.21}$$

where $T_{ij} = J_{ij} + 2D_{ij}$ is the total splitting. Because $\cos(\alpha)\cos(\beta) = \frac{1}{2}(\cos(\alpha+\beta)+\cos(\alpha-\beta))$, this results in two peaks in the spectrum separated by a difference in frequency T_{ij}. The same holds for an initial term $\hat{I}_{i,y}$ or with spins i and j exchanged. Thus, T_{ij} leads to splitting of the NMR spectra and can be obtained from the spectra of different pulse

sequences [144, 125, 46, 44]. This splitting is depicted in Figure 2.2, where the Lorentzian of the Hamiltonian without couplings is split into two Lorentzians separated by T_{ij}.

In the case of more than one coupled spin, such as in CH_2 or CH_3 groups, the couplings lead to the splitting into a multiplet, but the total sum of T_{ij} can still be determined [118].

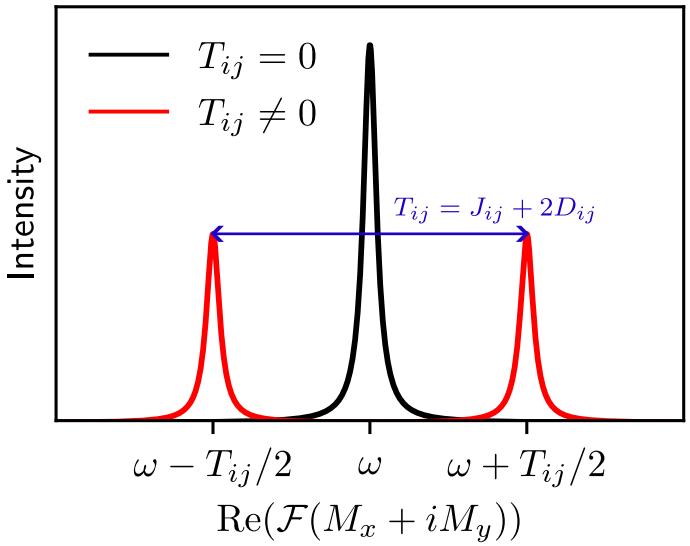


Figure 2.2.: Idealized NMR spectrum with and without the influence of the total coupling $T_{ij} = J_{ij} + 2D_{ij}$. Due to the T_2 relaxation in Equation (2.21), the peaks are drawn as Lorentzian.

2.1.4. Alignment Media

To induce a partial alignment of analyte molecule used in NMR experiments, different alignment media are used [128]. One class of alignment media are liquid crystals whose anisotropic magnetic susceptibility causes the liquid crystal to be aligned along the NMR magnetic field $\vec{B}^{\text{ext}}$ [119]. The most prominent example is poly-γ-benzyl-L-glutamate (PBLG), which form a stable α helix in many organic solvents [9, 67]. Compared to liquid crystal alignment media, stretched gels have the advantage that the alignment is variable very easily by changing the amount of stretching [97, 98, 56, 131]. This allows for the precise control of the alignment even without variable angle sample spinning [36] in order to make it strong enough for the RDCs to be measurable and weak enough such that the Hamiltonian in Equation (2.13) is a good approximation and the spectra do not become too complex [128]. The original alignment medium used for the determination of RDCs in biomolecules such as proteins was a nematic liquid crystal consisting of bicelles in aqueous solution [161], which are useful for structure prediction.

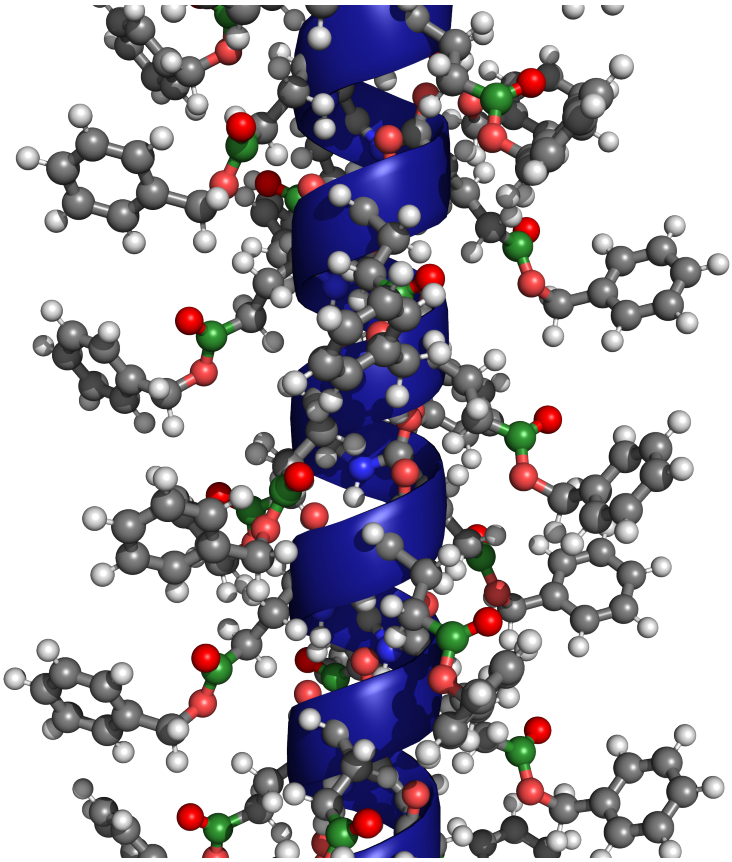


Figure 2.3.: Structure of PBLG used in the simulations. The backbone 18/5 α-helix is indicated in blue. In red, the oxygen atoms acting as hydrogen bond acceptors are shown for the C – O – C and C = O groups. The neighboring carbon atom is shown in green.

2.1.5. Scaling of the alignment

Even though the chains in the PBLG liquid crystal are mostly aligned along the z-axis, this alignment cannot be assumed to be perfect. This is especially relevant for other alignment media such as stretched gels, where the deviations might be even larger. Under the assumption however that the only anisotropy of the alignment medium is between the z-axis (along $\vec{B}^{\text{ext}}$) and the xy-plane and the analyte molecules interact with a large enough part of the alignment medium, the only effect is a scaling of the RDCs as shown in Appendix A.2. For PBLG, this is the case because the liquid crystal is oriented along the z-axis by the magnetic field, while there is no preferred direction in the xy-plane for the distribution of the PBLG molecules. In the case of a stretched gel, this assumption is also given as long as the stretching is oriented along the z-axis. Notably, the scaling factor is *positive* if the orientation of the alignment medium is stronger along the z-axis than in the xy-plane (see Appendix A.2). This is very important when comparing predicted RDCs to measured values, as it means that the cosine similarity is +1 when the RDCs agree up to a constant positive factor.

Quadrupolar coupling of deuterium

For nuclei with spins $I > \frac{1}{2}$, the electric quadrupole moment is nonzero. This leads to a contribution to the Hamiltonian in the presence of an electric field gradient. In the secular approximation, it has the form [89]

$$\hat{H}_{\text{quad.}} = \frac{1}{6}\hbar\langle\Delta\nu\rangle_t \left(3\hat{I}_z^2 - I(I+1)\hat{\mathbb{1}}\right), \tag{2.22}$$

where $\Delta\nu \propto \left\langle \partial_z^2\varphi \right\rangle_t$ is the first-order residual quadrupolar coupling and φ the electrostatic potential. In a frame rotating with the molecule, in which the electrostatic potential is approximately time-independent, we can write

$$\left\langle \partial_z^2\varphi \right\rangle_t = \tilde{\vec{e}}_z^{\top} H\varphi \tilde{\vec{e}}_z \tag{2.23}$$

with the Hessian matrix $H\varphi$ and the magnetic field direction $\tilde{\vec{e}}_z$ in the molecule frame. Because $\text{Tr}\,(H\varphi) = \Delta\varphi = 0$ (the Poisson equation) as we only consider the electric field gradient caused by the charge distribution external to the nucleus, we have the same situation as in Equation (2.15) and Equation (2.16).

This is important for the following reason: One commonly used application of the quadrupolar coupling for measurements in deuterated chloroform (with $I = 1$) is the quantification of the alignment strength of the alignment medium [82, 86]. It is based on the partial alignment of the chloroform molecules by the alignment medium. Here, the dependence of the alignment of chloroform on the concentration and average alignment of the alignment medium should be the same as for the alignment of the analyte molecules. In particular, given the assumptions of Appendix A.2, the same scaling with $\frac{1}{2}\left(3\cos^2(\alpha) - 1\right)$ still holds like for the analyte molecules. Therefore, the RDCs obtained from different probes with different alignment strengths should be proportional to the first-order quadrupolar couplings [86].

This scaling is especially relevant for the comparison of RDCs between enantiomers [105, 136, 137], as a preparation of the exact same conditions in two different experimental samples is not feasible in practice, leading to different alignment strengths for each experiment. The approximate proportionality of the RDCs to the quadrupolar splitting is additionally used in the thesis of Sager [136] to determine the sign of $J_{ij} + 2D_{ij}$ for HMIP, which I describe further in the results.

2.2. Atomistic molecular dynamics simulations

Bonded interactions

In classical force fields, bonded interactions are usually described as harmonic potentials for bond lengths and bond angles, while periodic potentials including dihedrals are described by low-order terms of a cosine transform [172]. As the analyte molecules in this works

are mostly rigid and the alignment comes from the nonbonded interactions between the analyte molecules and the alignment medium, these interactions play a secondary role, except possible for the movement of the PBLG sidechains and thus the O – H hydrogen bond donors.

Nonbonded interactions

Nonbonded interactions are usually described by two contributions: The Lennard-Jones potential and the Coulomb potential [172]. The Lennard-Jones (LJ) potential describes both the attractive van der Waals forces (such as London dispersion forces) (with the term $\propto -r^{-6}$) and a repulsive force for close atom (with the term $\propto r^{-12}$), modelling the Pauli exclusion principle. The potential between two atoms i and j can thus be written as

$$V_{ij}^{\mathrm{LJ}} = 4\epsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right). \tag{2.24}$$

How σ_{ij} and ϵ_{ij} are determined depends on the force field. They are frequently determined by a combining rule from parameters σ_i and ϵ_i of the individual atoms, which are determined for different atom types to reproduce experimental properties of liquids. For the AMBER ("Assisted Model Building with Energy Refinement") force field [173, 33], the Lorentz-Berthelot rules [94, 17] are used, where $\sigma_{ij} = 0.5(\sigma_i + \sigma_j)$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$. The Coulomb potential describes the electrostatic interactions between the atoms. In classical force fields, the (electronic) charge distribution is approximated by point charges centered on the atoms [172], although more detailed alternatives such as polarizable force fields [75] or QM/MM approaches [38] (based e.g. on Density Functional Theory (DFT) or semiempirical methods) exist.

Because the chemical bonds between atoms lead to a different electronic structures than for atoms belonging to a different molecule or a different part of a molecule, the nonbonded interactions are not applied for atoms connected by one or two bonds. Depending on the force field, the nonbonded potentials for atoms connected over three bonds (called 1-4 interactions) are scaled, e.g. 0.5 for the LJ potential and ~ 0.83 for the Coulomb potential in the case of AMBER. [173, 33]

In the AMBER family of force fields, charges are usually obtained by doing a restrained fit to the electrostatic potential of the molecule (RESP) [11, 34, 33, 32, 54]. The electrostatic potential is calculated using Hartree-Fock and the 6-31G* basis set [42, 61, 65], as it consistently overestimates the polarity of molecules, which is argued to fit the conditions in solution better [129, 32]. Due to the periodic boundary conditions used in MD simulations especially with solvent, the nonbonded interactions become an infinite sum over all periodic images. As the Lennard-Jones potential falls off with r^{-6}, it can be approximated by a simple cutoff and long-range dispersion correction for the total energy and pressure [2]. The Coulomb potential however scales with r^{-1} and the electrostatic potential of an uncharged, bounded object with a nonzero dipole moment with r^{-2}, ignoring the screening by the solvent. Thus, the long-range part needs to be treated more accurately. In the

particle mesh Ewald method [41], this is achieved by splitting the potential into a part described by atoms within a cutoff radius of each other, whose contribution is calculated directly as a pairwise sum, and a long-range part treated as a continuum. The latter contribution is calculated by applying Ewald summation in Fourier space [47].

Integration of the equations of motion

In usual MD simulations, the Newtonian equations of motion based on the force field are integrated in discrete time steps. For all-atom simulations, the time steps are typicall of the order of 1 fs. Especially suited for this case are symplectic integrators (such as Leapfrog), as they respect energy conservation for separable Hamiltonians [28]. In simulations with a constant temperature, the sampling from the canonical ensemble is usually achieved by coupling the system to a heat bath or adding stochastic noise to the equations of motion. As the motion of hydrogen atoms is relatively fast, the use of a time step ≤ 0.5 fs is required. If the precise dynamics of the hydrogen are not important however, is is possible to use a larger timestep while restraining the bond length for the bonds involving hydrogen atoms, using an algorithm such as SHAKE [133] or LINCS [69].

The Berendsen thermostat [13] couples the system to a heat bath with a constant temperature. It works by rescaling the velocities at every timestep to restore the total kinetic energy. This however has a severe limitation: Most importantly, the absolute motion (translation and rotation) of the system (or larger parts of the system) also contributes to the total kinetic energy, such that the kinetic energy attributed to random motion in the system (i.e. corresponding to the actual temperature) is reduced. Langevin dynamics [88] on the other hand introduces both fricion and random accelerations for each particle, which together keep the system at an (on average) constant temperature. The equations of motion for each particle with mass m are given by

$$m\ddot{\vec{x}} = -m\gamma\dot{\vec{x}} + \vec{F}(\vec{x}) + \sqrt{2m\gamma k_\mathrm{B}T}\vec{R}(t), \tag{2.25}$$

where the velocity-dependent friction is described by the frictuion constant γ, $\vec{F}$ is the force determined by the force field (dependent on the atom positions) and the third term introduces random noise. Here, $\vec{R}(t)$ is drawn from a 3-dimensional standard normal distribution independently at every time step (i.e. the stochastic process is δ-distributed). The resulting ensemble is the correct canonical ensemble. However, the equations of motions are no longer physical and dynamic properties and time scales might therefore not be described correctly. For free energy calculations, this does not matter though and in fact, it is especially suited for this case, because additional force is applied independently on each particle and therefore also works if parts of the system are completely decoupled from the rest. The velocity-rescaling thermostat [24] works by rescaling the velocity of all particles by a constant factor α at each timestep. α is chosen in a way such that when equilibrium is reached, the proper canonical ensemble is sampled for the given target temperature. This is achieved by using a stochastic process for the total kinetic energy $K = \frac{1}{2}\sum_i m_i\vec{v}_i^2$. In contrast to the Berendsen thermostat, it results in the correct sampling of the canonical ensemble.

Like with a thermostat, sampling from a NPT ensemble (with constant pressure) can be achieved by coupling the simulation volume to the equations of motion. One such approach is the Parrinello-Rahman barostat [120], where the triclinic simulation box vectors are coupled to the equations of motion by including their "kinetic" movement and the pressure-volume work into the system Hamiltonian. This is especially relevant for simulations with explicit solvent, where the pressure influences the movement of the solvent molecules.

Static dielectric constant in the MD simulations

The static dielectric constant of a solvent model used in MD simulations can be determined from the dipole fluctuation of the solvent during the simulation [150, 48]. Given the total dipole moment $\vec{M}$ of a sample, the Kirkwood G factor is defined proportional to its average squared fluctuation $G = \frac{4\pi\beta\langle\vec{M}^2\rangle}{3V}$, with inverse temperature β and volume V of the sample. If the system is isotropic, $\langle\vec{M}\rangle = 0$. Given a large volume V within an infinite (neutral) dielectric with static dielectric constant ϵ, it was shown in Kirkwood [80] and Fröhlich [57] that ϵ is related to the fluctuations of the total dipole moment $\vec{M}$ inside V by

$$(\epsilon - 1)(2\epsilon + 1) = \frac{4\pi\beta\langle\vec{M}^2\rangle}{V}. \tag{2.26}$$

In a MD simulation with periodic boundary conditions, the situation is somewhat different due to the correlations between the periodic images. However, the relation in this case actually simplifies to

$$\epsilon - 1 = \frac{4\pi\beta\langle\vec{M}^2\rangle}{V}. \tag{2.27}$$

For calculating $\langle\vec{M}^2\rangle$, the simulation time should be much longer than the autocorrelation time $\tau_{\vec{M}}$. For non-polarizable force fields, the ϵ calculated in this manner can be an underestimation of the real static dielectric constant, even if the effect is roughly included implicitly in the force field parameters. [48]

2.3. Binding free energy calculation

To calculate the free energy difference of a system described by the canonical ensemble with inverse temperature β between two thermodynamic states with potential energies $U_1(q)$ and $U_2(q)$, free energy perturbation can be used. Here, $q \in \Gamma$ describes the configuration of the system. The free energy difference $F_2 - F_1$ is given by the Zwanzig equation [177]

$$\beta(F_2 - F_1) = -\ln\left\langle e^{-\beta(U_2(q)-U_1(q))}\right\rangle_1, \tag{2.28}$$

where the average is calculated for state 1. The effect of the pressure p_i or chemical potential μ_i can be be included by replacing the potential by $U_i(q) \longrightarrow U_i(q) + p_i V(q) + \mu_i n(q)$, with the volume $V(q)$ and particle number $n(q)$ [147].

When the sampling is done in both states, the Bennett Acceptance Ratio (BAR) method or its generalization to multiple states (multistate BAR (MBAR) [147]) can be used to obtain a relation between the free energies between the states. The free energies F_i are then obtained by solving the equation(s) numerically. For the MBAR method with K states and N_i samples for state i, the relation is

$$F_i = -\ln\left[\sum_{j=1}^{K}\sum_{n=1}^{N_j}\frac{e^{-U_i(x_{jn})}}{\sum_{k=1}^{K} N_k e^{F_k - U_k(x_{jn})}}\right]. \tag{2.29}$$

These equations need to be solved iteratively until a self-consistent solution is found. This results in a unique set of free energies F_i up to an additive constant [147]. The BAR method [12] is the special case of $K = 2$, for which Equation (2.29) only needs to be solved for the single free energy difference $F_2 - F_1$. The main benefit of MBAR over the successive application of the BAR method is that the sampling done in one state contributes to the information about all other states (although the contribution is small if the states are too dissimilar), thus requiring less sampling. This is particularly relevant for many intermediate states, where each individual state is similar to multiple other states.

Another alternative for calculating free energy differences is thermodynamic integration, provided the intermediate states are differentiable with respect to a continuus parameter $\lambda \in [0, 1]$. The free energy difference can then be calculated by [79]

$$F_2 - F_1 = \int_0^1 \left\langle \frac{\partial U(\lambda, q)}{\partial \lambda} \right\rangle_\lambda d\lambda, \tag{2.30}$$

which is usually approximated as a discrete sum over the intermediate states. For all methods mentioned above, the intermediate states do not need to be "physical" and their choice only influences the statistical error of the final free energy difference.

In the case of solvation free energies, the nonbonded interactions between the solvent and the solutes need to be decoupled. For the Coulomb potential, this can be achieved simply by scaling the charges $q_k \to \lambda q_k$ of all atoms k. For the Lennard-Jones potential, the divergence with r^{-12} as $r \to 0$ makes this approach problematic however. A viable alternative are soft-core potentials [18], which gradually switch off the Lennard-Jones potential and remove the divergence at $r = 0$ for $\lambda < 1$.

2.4. Free energies of a subsystem

Given the total state space Γ and an observable $A(q)$, its average under the canonical ensemble at inverse temperature β with energy $E(q)$ is given by

$$\langle A \rangle = \frac{1}{N}\int_\Gamma A(q) e^{-\beta E(q)}\, dq, \tag{2.31}$$

with the normalization $N = \int_\Gamma e^{-\beta E(q)}\, dq$. An example for q are the positions of the atoms in a molecular system.

If the state space is split into two subspaces, i.e. $\Gamma = \Gamma_1 \times \Gamma_2$, and the observable A only depends on variables $q_1 \in \Gamma_1$, Equation (2.31) can be rewritten as

$$\begin{aligned}\langle A \rangle &= \frac{1}{N} \int_{\Gamma_1} \int_{\Gamma_2} A(q_1) e^{-\beta E(q_1,q_2)}\, dq_2\, dq_1 \\ &= \frac{1}{N} \int_{\Gamma_1} A(q_1) \int_{\Gamma_2} \int_{\Gamma_1} e^{-\beta E(q_1',q_2)} \delta(q_1' - q_1)\, dq_1'\, dq_2\, dq_1 \\ &= \frac{1}{N} \int_{\Gamma_1} A(q_1) \int_{\Gamma} e^{-\beta E(q_1',q_2)} \delta(q_1' - q_1)\, d(q_1', q_2)\, dq_1,\end{aligned} \tag{2.32}$$

where δ is the Dirac delta function.

This leads to the definition of the free energy $F(q_1)$ for state $q_1 \in \Gamma_1$

$$e^{-\beta F(q_1)} \coloneqq \frac{1}{N_2(q_1)} \int_{\Gamma} e^{-\beta E(q_1',q_2)} \delta(q_1' - q_1)\, d(q_1', q_2) \tag{2.33}$$

with a normalization constant $N_2(q_1)$ and therefore

$$\langle A \rangle = \frac{1}{N_1} \int_{\Gamma_1} A(q_1) e^{-\beta F(q_1)}\, dq_1. \tag{2.34}$$

with another normalization constant N_1.

In MD simulations for the calculation of binding free energies, $\delta(q_1' - q_1)$ can be approximated by restraints using a harmonic potential. Apart from this approximation, Equation (2.34) is exact and using the calculated $F(q_1)$ should in principle lead to the same result as regular MD simulations without restraints. Obviously, for calculations of RDCs, this would be many orders of magnitude more computationally expensive than doing direct simulations of the alignment.

2.5. Implicit Solvation Models

For the MD simulations, using explicit solvent molecules is generally preferable for the accuracy of the simulations [175, 151]. However, having many solvent molecules increases the computational cost of the simulations, and the large number of moving atoms makes Monte Carlo schemes challenging [111]. An alternative is the treatment of the solvent effects using implicit solvation, where they are treated in an averaged fashion which only depends on the solute coordinates. Especially for rigid molecules, this makes the application of Monte Carlo methods feasible. Implicit solvation models are commonly split into a nonpolar and polar part [130]. The nonpolar part describes the interactions of chargeless solutes with the solvent, while the polar part describes the free energy change upon restoring the solute charges. In the following, I describe both the nonpolar and polar solvation models that I adjusted to the approximate calculation of binding free energies and test on MD data of analytes interacting with PBLG.

2.5.1. Nonpolar solvation

A common approach for the nonpolar solvation is to split it into two parts: An attractive part ΔG_{att} representing the dispersion interaction between the solutes and the solvent, and a repulsive part ΔG_{rep} describing the cavity formation and surface tension of the solvent.

For the attractive part, the approach in Wagoner and Baker [167] is to treat the solvent as a continuum and splitting the LJ potential into an attractive and repulsive part. ΔG_{att} is then calculated as the integral over the attractive contribution to the LJ potential between the solute and the solvent:

$$\Delta G_{\text{att}} = \bar{\rho} \sum_i \int_\Omega V_{\text{LJ,att},i}(\vec{x}_i, \vec{y}) g(\vec{y})\, d^3y. \tag{2.35}$$

Here, $\rho =: \bar{\rho} g$ is the solvent number density around the solutes and $\bar{\rho}$ the bulk solvent number density. The solvent distribution function g is commonly approximated as a step function, where the atom radii are extended by a constant probe radius. The integral is calculated over the volume Ω outside the solutes, where the atoms are modeled as spheres whose radii are extended by a probe radius representing an effective solvent radius.

With the Weeks-Chandler-Andersen (WCA) decomposition [169], the split between the attractive and repulsive part is made at the minimum of the LJ potential (i.e. at $2^{1/6}\sigma$), such that the attractive part is given by

$$V^{\text{WCA}}_{\text{LJ,att},i}(\vec{x}_i - \vec{y}) = \begin{cases} -\sqrt{\epsilon_i \epsilon_{\text{solv}}}, & |\vec{x}_i - \vec{y}| < 2^{1/6}(\sigma_i + \sigma_{\text{solv}}) \\ V^{i,\text{solv}}_{\text{LJ}}(|\vec{x}_i - \vec{y}|), & |\vec{x}_i - \vec{y}| \geq 2^{1/6}(\sigma_i + \sigma_{\text{solv}}) \end{cases}, \tag{2.36}$$

with the Lennard-Jones parameters σ_{solv} and ϵ_{solv} for the solvent. For the σ decomposition [157] on the other hand, the split is made at the root of the LJ potential (i.e. at $|\vec{x}_i - \vec{y}| = \sigma$). The attractive part is therefore

$$V^{\sigma}_{\text{LJ,att},i}(\vec{x}_i - \vec{y}) = \begin{cases} 0, & |\vec{x}_i - \vec{y}| < \sigma_i + \sigma_{\text{solv}} \\ V^{i,\text{solv}}_{\text{LJ}}(|\vec{x}_i - \vec{y}|), & |\vec{x}_i - \vec{y}| \geq \sigma_i + \sigma_{\text{solv}} \end{cases}. \tag{2.37}$$

In the case of approximating g as a step function, Equation (2.35) can be calculated by a surface integral over the surface $\partial\Omega$ of the solutes by applying the divergence theorem [51, 157]. Given a vector field $\vec{A}_i$ with $\vec{\nabla}\vec{A} = V_{\text{LJ,att},i}$,

$$\Delta G_{\text{att}} = \bar{\rho} \sum_i \int_{\partial\Omega} \vec{A}_i(\vec{y}) \cdot \vec{n}_y\, d^2\sigma_y, \tag{2.38}$$

where $\vec{y}$ is a point on the surface and $\vec{n}_y$ the corresponding normal vector (pointing outward the solute). Due to the radial symmetry of $V_{\text{LJ,att},i}$, the vector field $\vec{A}(\vec{r})$ can be determined by setting $\vec{A} = f(r)\vec{r}$ [51] with

$$f(r) = \frac{1}{r^3} \int_0^r x^2 V_{\text{LJ,att},i}(x)\, dx. \tag{2.39}$$

The description of the repulsive part in Wagoner and Baker [167] is motivated by scaled particle theory [155, 126]. Here, the contribution due to the surface tension of the solvent and its collisions with the solutes is assumed to be proportional to the (solvent-accessible) surface area of the solutes (SASA), while the contribution due to cavity formation is described by the solvent-accessible volume (SAV). The repulsive part is therefore given by a linear combination of both,

$$\Delta G_{\text{rep}} = \gamma \text{SASA} + p\text{SAV}, \tag{2.40}$$

where γ is the surface tension parameter and p is the pressure parameter.

2.5.2. Polar solvation

A commonly used implicit solvent model for the polar solvation is based on the Poisson-Boltzmann equation, which described the average electrostatic potential Φ with a spatial variation of the dielectric constant and possibly the average distribution of ions [53]. In nonpolar solvents such as chloroform, no ions exist, such that only the Poisson equation remains:

$$\vec{\nabla}(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})) = -\frac{1}{4\pi\epsilon_0\epsilon(\vec{r})}\sum_i q_i\delta(\vec{r}-\vec{r}_i). \tag{2.41}$$

Here, r_i are the positions of point charges, which are typically the atom positions in the solute molecules. The dielectric constant $\epsilon(\vec{r})$ can be approximated by dividing the into the volume into two parts: The solvent volume with dielectric constant ϵ_{out} and the solute volume with dielectric constant ϵ_{in} [166]. For the latter, each atom is modelled as a sphere whose radius is the sum of its van der Waals radius and the solvent radius. As solving Equation (2.41) is computationally expensive, it is commonly approximated using a Generalized Born (GB) model [154]. It is an extension of the Born equation, which describes the free energy of solvation of an ion with charge q in a medium with dielectric constant ϵ_{out}, modelled as a sphere with radius a and dielectric constant. In this case, the electrostatic part of the free energy of solvation described by Equation (2.41) can be expressed analytically as

$$\Delta G^{\text{Born}} = -\frac{q^2}{4\pi\epsilon_0 a}\left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}}\right). \tag{2.42}$$

In the most commonly used formulation of the GB model by Still et al. [154], the sphere approximation is modified to reflect non-spherical geometries. Additionally, the assumption $\epsilon_{\text{out}} \gg \epsilon_{\text{in}}$ is made, which is quite reasonable for biomolecules in water, but less so for solvents with small ϵ_{out}. These changes result in

$$\Delta G^{\text{GB}} = \sum_{ij}\Delta G_{ij}^{\text{GB}} = -\frac{1}{8\pi\epsilon_0}\left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}}\right)\sum_{ij} q_i q_j \frac{1}{f_{ij}} \tag{2.43}$$

with $f_{ij} = \sqrt{r_{ij}^2 + a_i a_j \exp(-r_{ij}^2/(4a_i a_j))}$.

A modification of the usual GB model which is also valid for solvents with lower dielectric constant was developed by Sigalov, Fenley, and Onufriev [148]. Because it is derived as an approximation to the linarized Poisson-Boltzmann equation, they call it the "analytical linearized Poisson-Boltzmann" approach (ALPB). There, the assumption that $\epsilon_{\text{out}} \gg \epsilon_{\text{in}}$ is no longer made, but an additional parameter A describing the effective size of the molecule is introduced:

$$\Delta G^{\text{ALPB}} = \sum_{ij} \Delta G_{ij}^{\text{ALPB}} = \underbrace{-\frac{1}{8\pi\epsilon_0}\left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{1+\alpha\beta}}_{=:c_{\text{ALPB}}} \sum_{ij} q_i q_j \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A}\right). \tag{2.44}$$

For uncharged molecules (like in this work), $\sum_{ij} q_i q_j = 0$ and thus ΔG^{ALPB} does not depend on A and can be still be regarded as a variant of the original GB model.

Both in the original GB model and the ALPB model, the Born radius of each atom is required. A very accurate, but expensive method to determine them are "perfect Born radii" [115]. They are calculated by solving the Poisson equation for each individual atoms, where the charge for the atom itself is kept, while the charges of the other atoms are set to zero, keeping only their shape and internal dielectric constant ϵ_{in}. To make the assumption $\epsilon_{\text{out}} \gg \epsilon_{\text{in}}$ for Equation (2.43) valid, the dielectric ϵ_{in} is set to 1, while ϵ_{out} is chosen very large, e.g. 1000. The Born radius a_i of the atom can then be calculated from Equation (2.43), because $q_j = 0$ for $j \neq i$ and $f_{ii} = a_i$.

A relatively simple model for the approximate calculation of Born radii is the HCT model described in Hawkins, Cramer, and Truhlar [63]. It is based on an analytical expression of the overlap between two spheres, while the "double counting" if there are additional overlapping spheres is described by fitted scaling factors S_i for each atom. The Born radii are then given by

$$\frac{1}{a_i} = \frac{1}{R_i} - \sum_j H(R_i, S_j(R_j), r_{ij}), \tag{2.45}$$

where $H(R_1, R_2, r)$ is an analytical expression for two spheres with radii R_1 and R_2 and distance r.

The HCT approximation is adapted in the ΔHCT model [92] for the application of binding energies. There, the "perfect" Born radii (as described above) of the isolated solutes are modified for a complex according to the HCT model:

$$\frac{1}{\tilde{a}_i} = \frac{1}{a_i} - \sum_k H(R_i + r_{\text{ext}}, S'_k(R'_k + r_{\text{ext}}), r_{ik}), \tag{2.46}$$

In Liu, Grinter, and Zou [92], the atoms of a complex are split into "primary" (close to the respective other solute) and "secondary" (further away) atoms and the ΔHCT is only applied on the "secondary" atoms, while the Born radii for the "primary" atoms are calculated with a more accurate method.

2.6. Monte Carlo Calculations

Monte Carlo integration is the numerical calculation of integrals using random samples. In the simplest case, the (bounded) integration domain Ω is sampled uniformly and the average over a function $f : \Omega \to \mathbb{R}^n$ is given by

$$\bar{X} \approx \frac{1}{N} \sum_i f(X_i), \tag{2.47}$$

where N is the number of samples. If the samples are independent and the variance of f is finite,

$$\mathrm{Var}(\bar{X}) = \frac{1}{N} \mathrm{Var}(f), \tag{2.48}$$

whose scaling with N is independent from the type of Ω, such as its dimensionality if it is a subset of $\mathbb{R}^n$. If f has deep extrema or discontinuities, this approach can however be very inefficient, because most of the samples are wasted on "uninteresting" regions of Ω.

There are several modifications to reduce the variance, such as stratified sampling [112] or importance sampling [163]. Importance sampling is especially relevant if $f \geq 0$ and $\int_\Omega f < \infty$, such that it is proportional to a probability density p. If we now sample from a probability density $\tilde{p}$, the average becomes

$$\bar{X} \approx \frac{1}{N} \sum_i \frac{f(X_i)}{\tilde{p}(X_i)}, \tag{2.49}$$

where $X_i \sim \tilde{\mathbb{P}}$ (the corresponding probability distribution). The variance of the average can be shown to be minimal for $\tilde{p} \propto p \propto f$ [132].

Another class of Monte Carlo algorithms which may be used to produce samples from a given probability distribution are Markov Chain Monte Carlo (MCMC) methods. The most prominent and straightforward example is the Metropolis-Hastings algorithm [106, 62]. It works by iteratively proposing a new sample X_j depending on the last sample X_i with a proposal probability density p_{prop}, with which the entire state space needs to be accessible in a finite amount of steps. The new sample is the accepted with the probability

$$p_{\mathrm{acc}}(X_i \to X_j) = \min\left(\frac{p(X_j) p_{\mathrm{prop}}(X_j \to X_i)}{p(X_i) p_{\mathrm{prop}}(X_i \to X_j)}\right), \tag{2.50}$$

otherwise X_i is appended again to the resulting Markov chain. A significant benefit is that p only needs to be known up to a constant factor. Additionally, the samples can be used to calculate averages over arbitrary functions $f(X_i)$ over the sampled distribution, which can be much more efficient than calculating separate integrals when $\|f(X_i)\| p(X_i)$ is not too dissimilar from $p(X_i)$. The Metropolis algorithm is the special case where the proposal distribution is symmetric. Thus, the acceptance criterion only depends on the target distribution, simplifying the implementation. For the canonical ensemble, using the Metropolis-Hastings algorithms means that the Boltzmann factors $e^{-\beta E_i}$ can be used directly without knowing the partition function $Z = \sum_i e^{-\beta E_i}$.

In the case of the canonical ensemble, $p(X_j)/p(X_i) = e^{-\beta(E_j - E_i)}$. This however means that deep energy minima are a problem: If E_i is such a minimum, $\beta(E_j - E_i) \gg 1$ and $p(X_j)/p(X_i) \ll 1$ for most $j \neq i$. To overcome this problem even in the case of multiple deep minima, different Monte Carlo methods are required to still allow efficient sampling of the entire sample space. As mentioned above, to calculate an average of a function f over p, using samples from p is nearly optimal when the variance of f itself is not too high. For RDCs D_{ij} (Equation (2.14)) and the alignment tensor $\boldsymbol{A}$ (Equation (2.16)), this is certainly the case, especially for ${}^1D$ couplings, where r_{ij} does not change significantly.

When calculating averages from a MCMC simulation, the correlations between the samples in the chain needs to be taken into account. If the sample size is large anough, a Markov chain central limit theorem [76] can be assumed to apply and thus the uncertainty of the average approximately follows a normal distribution. Methods to estimate the statistical uncertainty of the average include e.g. dividing the samples into "blocks" or by estimating the correlation function of the samples [52]. One estimator based on the correlation function is the adjusted initial sequence estimator ("mISadj") described in Dai and Jones [39]. It has the benefit of always producing a positive definite covariance matrix, at the cost of a possible (slight) overestimation of the variances. The estimator is based on calculating the autocorrelation function γ_t, where t is the lag between samples. For a single Markov chain with samples X_i,

$$\hat{\gamma}_t = \frac{1}{n} \sum_{i=1}^{n-t} (X_i - \hat{\mu}_n)(X_{i+1} - \hat{\mu}_n)^\top. \tag{2.51}$$

The estimated covariance matrix $\Sigma = \operatorname{Cov} X$ of the probability distribution sampled by the Markov chain is then [39]

$$\hat{\Sigma} = -\gamma_0 + 2 \sum_{i=0}^{\infty} (\gamma_t + \gamma_{-t}), \tag{2.52}$$

In a classical MD simulation, the next step usually depends only on the current positions and velocities, which can also be considered a Markov chain. If the snapshots of a MD trajectory are additionally spaced far enough apart such that the velocities of the are uncorrelated and the Hamiltonian behavior is averaged out sufficiently, the trajectory approaches a random walk in the position part of the state space. Thus, only the positions also approximately form a Markov chain and Equation (2.52) is also applicable. I use this to calculate covariance matrices of MD averages.

2.6.1. Population Annealing

Population annealing [71, 99, 100] is a Monte Carlo method for calculating averages in the canonical ensemble. It is based on earlier approaches of population-based Monte Carlo methods [73] and was originally motivated was the application to spin glass models. There, the free energy landscape has many metastable states at lower temperatures, making the application of the Metropolis algrithm very inefficient. The aim of population annealing is therefore to find a population of states sampled from the correct canonical ensemble. It is

based on the observation that at high temperatures, the sampling becomes much easier and the population at a lower temperature can be achieved by resampling the population. This resampling however increases the correlations between the samples (because some are repeated), which is why their correlations need to be reduced by performing regular MCMC updates on them. The procedure is thus the following:

1. Generate the population at a small β_0 with another method. For $\beta = 0$, the canonical ensemble is just a uniform distribution and therefore often trivial to sample.
2. Resample the population from $\beta_i \longrightarrow \beta_{i+1} = \beta_i + \Delta\beta_i$.
3. Reduce the correlations between the samples by performing (short) independent MCMC simulations.
4. Repeat steps 2 and 3 until the final β is reached.

It allows for several variations: The size of the initial population and of populations after each resampling step, the method used for resampling, the choices of the inverse temperatures β_i and the MCMC method. Instead of the MCMC method, it is also possible to use other sampling methods, such as MD simulations [30].

Different resampling schemes for a constant population size N are described in Gessert, Janke, and Weigel [58], where the "systematic resampling" is recommended for a constant population size. Given the required number of copies of sample j in the new population $\tau_j = Ne^{-\Delta\beta_i E_j} / \sum_{j=1}^{N} e^{-\Delta\beta_i E_j}$, it works by assigning sample j to the index i in the new population if $\sum_{k=1}^{j-1} \tau_k <= i + r < \sum_{k=1}^{j} \tau_k$, where $r \in [0, 1)$ is a single random number. For the choice of $\Delta\beta_i$, it was argued in Gessert, Janke, and Weigel [58] that the "replica-averaged squared family size" ρ_t should be minimized in order to improve the accuracy of population averages, which means minimizing $\Delta\rho_t / \Delta\beta_i$ at each resampling step [58]. ρ_t is defined as $\sum_{j=1}^{N} N_j^2 / N$, where N_j is the number of samples originating from sample j in the original population at β_0. This is because samples belonging to different families are by definition uncorrelated, while the correlation between samples of the same family are only reduced by the MCMC simulations done after they originated as copies of the same sample.

Another method to overcome free energy minima during sampling is parallel tempering (or replica-exchange MCMC), where single MCMC simulations are simulated simulateneously at different temperatures and samples at neighboring temperatures are exchanged regularly. The main benefit of population annealing is its parallelizability [99, 170]: Only the resampling step, which does not depend on any description of the system except for the energies of the samples in the current population, requires the consideration of multiple samples, while the generation of the population at $\beta = 0$ and the MCMC steps are embarassingly parallel. In this work, I only use the variant with a constant population size and systematic resampling, while the MCMC steps are done with the Metropolis algorithm.

2.6.2. “Cylinder” (PALES) model

A rough approximation of the alignment medium can be achieved by replacing it with a cylinder. For large enough molecules where the alignment is determined by steric interactions, this can be expected to be a reasonable approach. It is used in the PALES (“Prediction of alignment from structure”) model [178], where the alignment is determined by integrating over all rotations and positions of the analyte molecule close to the cylinder. There, the integration is done by simple quadrature weighted by the Boltzmann factor, but the integral can be equivalently calculated using Monte Carlo methods as described above. Due to symmetry, the positions need only be varied in one dimension, the distance from the cylinder. Additionally, the electrostatic interaction between the analyte and the (uniformly charged) cylinder may also be included by weighting the poses with the corresponding Boltzmann factor. However, this model by definition does not take into account any specific interactions between the analyte and the alignment medium and yields exactly the same results for two enantiomers. Apart from its limited accuracy, this means that the enantiomer differentiation by a chiral alignment medium can only be described by a model which is detailed enough to include interactions between the analyte and the alignment medium that depend on their chirality.

Further approximations have been reported which depend only on the structure of the analyte, arguing that its shape depends the steric alignment, which can be expressed by the inertia tensor [7] or the gyration tensor [3, 174]. However, they can be considered approximations of the “cylinder” (PALES) model, so I use only the latter for comparison.

2.6.3. P3D model

In the P3D model [116, 74], the alignment is calculated by treating the atoms in PBLG and the analyte molecules as hard spheres, in addition to the interaction with a precalculated electrostatic potential around the alignment medium. Like for the PALES approach, the integral is calculated by simple quadrature in Opakua et al. [116], but now the cylindrical symmetry is not given anymore and the integral needs to be calculated in all three spatial dimensions around the alignment medium. The electrostatic potential is calculated by solving the Poisson equation using the dielectric constant of chloroform outside the alignment medium. For the different poses of the analyte, it is then treated as test particle inside the unchanged electrostatic potential. Other interactions between the analyte and the alignment medium are not included. Despite this, they demonstrate good agreement with experiment for several analyte molecules. However, except for sucrose in Ibáñez de Opakua and Zweckstetter [74], only a single structure was used for the analyte. While in Opakua et al. [116], it was shown that the influence of the force field with which the optimization was done, the tested structures were still very similar. In this work, I test this assumption and show that it is not valid for molecules with O – H hydrogen bond donor or even for some molecules where the largest flexibility are CH_3 groups.

3. Developed methodology

3.1. Residual Dipolar Couplings from Molecular Dynamics Simulations

3.1.1. Molecular system

The samples used in experiment consist of a lyotropic PBLG/$CDCl_3$ liquid crystal aligned along the magnetic field, while the analyte molecule are dissolved in the deuterated chloroform. As long as the liquid crystal is intact, different PBLG concentrations can be expected to only scale the RDCs by a constant (positive) factor. Likewise, in Appendix A.2, I show that any imperfect alignment of the PBLG chains would also scale the RDCs only by a constant, positive factor. It is therefore reasonable to do the simulations using only one PBLG chain aligned along the z-axis.

Due to experimental evidence that PBLG forms a 18/5 α-helix in organic solvents [9, 67] with a length of 2.7 nm for 18 monomers, I followed the approach described in Helfrich, Hentschke, and Apel [67] and Helfrich and Hentschke [66] to describe the PBLG molecule as an α-helix connected periodically along the z-axis. The box size is fixed in the z-direction with a length of 2.7 nm, while the barostat is applied only in the xy-plane. The periodic boundary conditions already ensure that the PBLG molecule remains a stable α-helix aligned along the z-direction, such that no restraints are necessary for the simulations.

In the xy-plane, I chose the box size as small as possible, but large enough such that there is enough space between the PBLG chains not to have simultaneous interaction of the analyte with two periodic images of PBLG. Additionally, it needs to be large enough for the barostat to remain stable and the fluctuation of the box size not too large. Larger box sizes are problematic because they increase the space where the analyte molecules move almost isotropically. If we scale the box to reduce the RDCs by a factor of s, it is reasonable to assume the variance of the dipolar couplings to be roughly constant regardless of the alignment, because the instantaneous dipolar couplings only depend on the orientation to the magnetic field and do not change depending on the closeness to the alignment medium. The statistical uncertainty of the couplings ΔD is then unchanged, such that the relative uncertainty is scaled by a factor of $1/s$. This means that to reach the same relative uncertainty without the scaling, the amount of sampling is increased by a factor of $1/s^2$. The simulation box is shown in Figure 3.1 with 1/3 of the actual chloroform molecules. Its box vectors in the xy-plane have a length of around 3.6 nm after the NPT equilibration (described below).

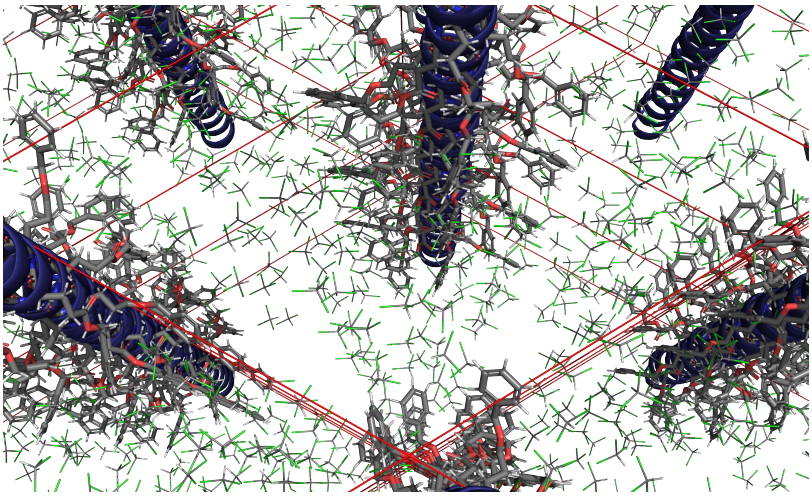


Figure 3.1.: Simulation box used for the MD simulations to calculate RDCs (shown without the analyte molecule). For the visualization, only every third chloroform molecule is included. The simulation box is shown in red and the backbone α-helix of PBLG, which is aligned along the z-axis, is indicated as a blue spiral.

To build the α-helix of PBLG, I started with the capped monomer and optimized it using the B3LYP functional [153] and def2-SVP basis set [143] using `PySCF` [156] with the `XCFun` [45] and the `geomeTRIC` [168] libraries. The resulting structure is shown in Figure A.1a. Afterwards, I adjusted the bond lengths, angles and dihedrals of the backbone to fit the experimentally determined structure of an 18/5 α-helix of α-poly-L-alanine [4], with the result shown in Figure A.1b. The α-helix can then be built by removing the caps and rotating and translating copies of the monomer to match the bond lengths, angles and dihedrals between them to an α-helix. Here, the exact structure is not important, as I later relaxed the molecule in MD using the force field. In Figure 2.3, the resulting PBLG structure used for the simulations is shown.

For the force field parameters (except for the charges), I used `Ambertools 21` [27] for the automatic assignment with the `antechamber`, `parmchk2` and `tleap` programs. For PBLG, each monomer should have the same parameters. Therefore, I did the assignment of the force field parameters for three joined monomers capped with $COCH_3$ (ACE) and $NHCH_3$ (NME) groups as shown in Figure A.1c, using the force field parameters involving the atoms of the central monomer. As the maximum number of neighboring atoms included in force field terms is 4 (the torsion terms), this suffices to correctly assign all parameters. For constructing non-periodic PBLG molecules, the parameters involving atoms in the caps are likewise taken from the caps in the molecule with 3 monomers.

As γ-benzyl-L-glutamic acid is partially similar in structure to proteinogenic amino acids like glutamic acid and phenylalanine, I used the ff14SBonlysc [113] force field parameters, which is an AMBER force field for proteins using ff99SB [70] backbone parameters with updated side chain parameters. Unlike ff14SB [101], it does not include an empirical correction for the backbone parameters derived from experimental data in water, which would not have much benefit in my case, as PBLG forms a stable α-helix in chloroform.

For chloroform, I used the non-charge parameters from Fox and Kollman [54], with a mass of 2 Da for the hydrogen atom (due to the constraints involving bonds with hydrogen atoms, the influence of the hydrogen mass should be negligible however).

In accordance with the AMBER methodology, the charges of the molecules are determined by a RESP fit of the optimized structures to an electrostatic potential determined with Hartree-Fock using the 6-31G* basis set [42, 61, 65]. For PBLG, I used a single monomer capped with NME and ACE caps for the charge fitting as shown in Figure A.1a, while restraining the fit to keep the charges of the monomer and the two caps zero, respectively.

In the simulations with a few water molecules in the simulation box, I use the TIP3P water model with rigid molecules [77], because the time step in the MD simulations is 2 fs.

3.1.2. Analyte molecules used in the MD simulations

I performed the MD simulations with 7 pairs of enantiomers for which experimental data is available. The analyte molecules isopinocampheol (IPC), quinuclidinol, borneol and camphor are shown in Figure 3.2 and indolophenanthridine (HMIP), ibuprofen and 1-phenylethanol in Figure 3.3. There, the 3D structures are the DFT-optimized reference structures described below in Section 3.1.4. Also shown are the atom numberings used to label the $^1D_{\mathrm{C-H}}$ couplings.

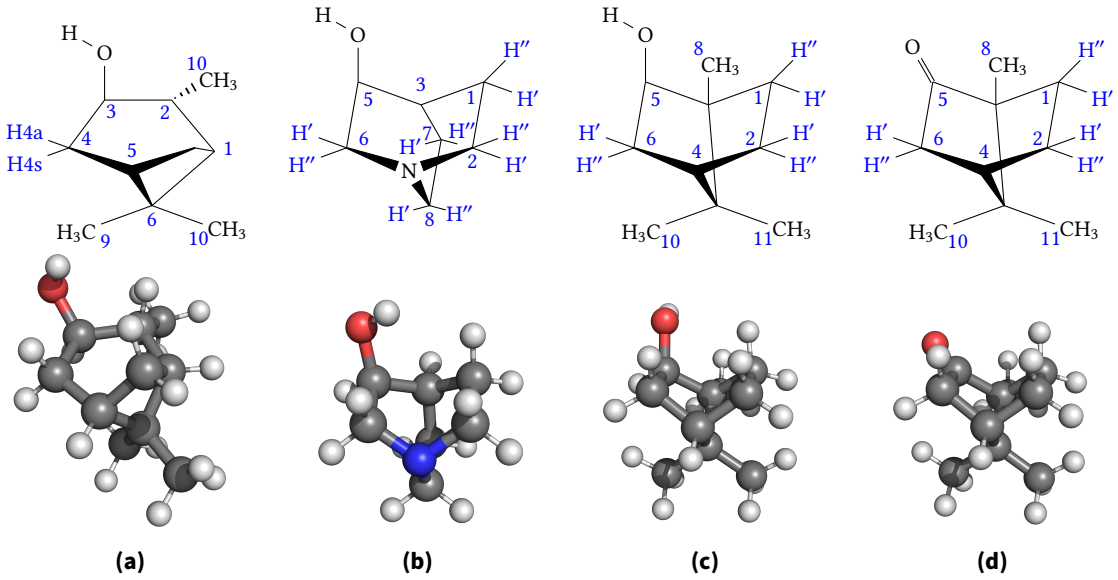


Figure 3.2.: Molecule structures with the used atom numbering for **(a)** (-)-IPC, **(b)** *S*-(+)-quinuclidinol, **(c)** *S*-(-)-borneol and **(d)** *S*-(-)-camphor.

The used $^1D_{\mathrm{C-H}}$ couplings for these molecules and their experimental values are specified in Tables A.1 to A.4. Due to symmetry and the rotatable phenyl ring, some couplings in ibuprofen and 1-phenylethanol need to be averages. Additionally, some experimental CH_2 RDCs in HMIP and ibuprofen had unclear assignment [136, 102], so they are described by their average. For isopinocampheol (IPC), the measurements were reported in Marx

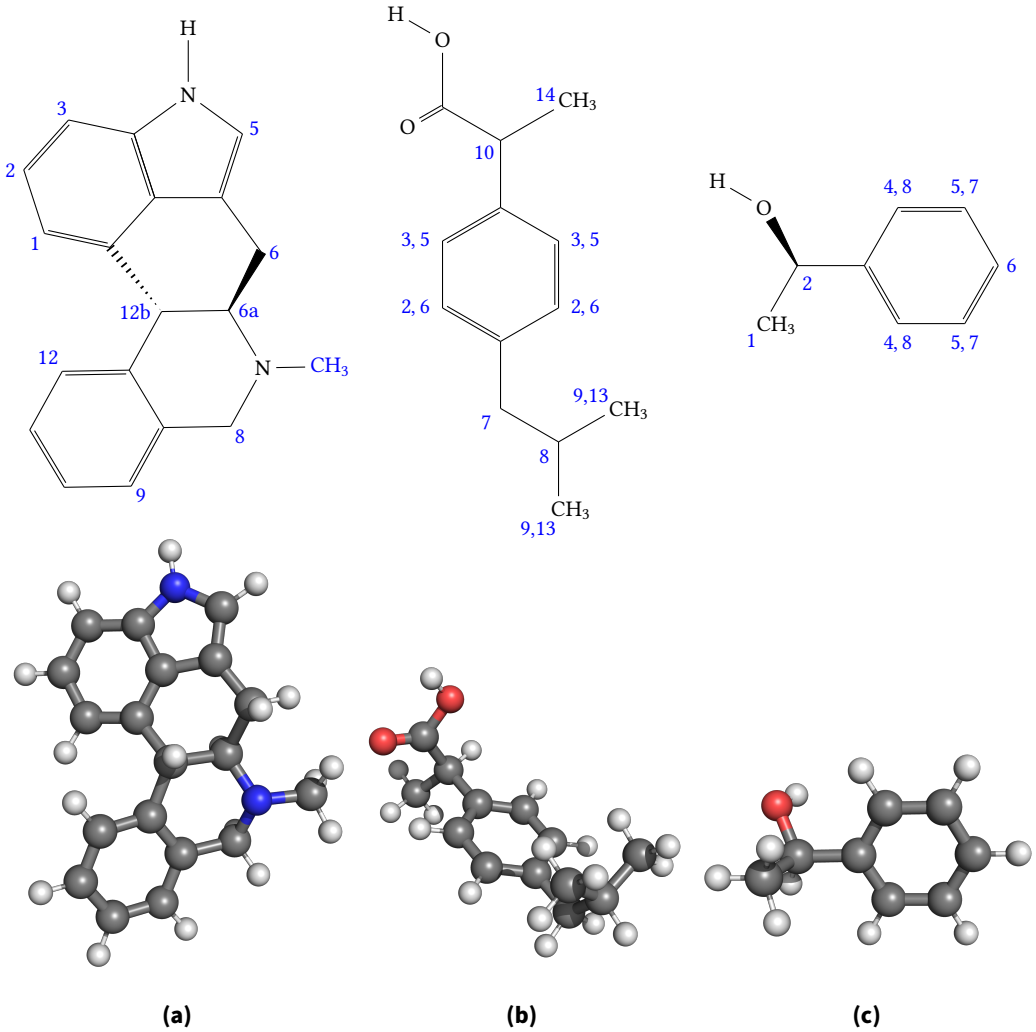


Figure 3.3.: Molecule structures with the used atom numbering for **(a)** *R*-HMIP, **(b)** *R*-ibuprofen and **(c)** *R*-1-phenylethanol.

and Thiele [105]. However, there is a discrepancy between the measurement in Marx and Thiele [105] and Sinnaeve et al. [149] for the C8-C6 and C9-C6 couplings in (+)-IPC (named here CH_3-8 and CH_3-8), whose assignment is swapped. An alignment tensor fit to the DFT-optimized reference structures (the ones depicted in Figure 3.2) clearly favors the assignment in Sinnaeve et al. [149], so I used this assignment instead. As the other RDCs of (-)-IPC are close to the RDCs in (+)-IPC, I also swapped the two couplings there. The structure of (+)-IPC and the atom numbering for the couplings is shown in Figure 3.2a.

The measurements for quinuclidinol, borneol and camphor were reported in Sager et al. [137] and previously in the thesis of Sager [136]. Like for (+)-IPC, their structures and atom numberings are shown in Figures 3.2b to 3.2d. Like IPC, they have a spherical-like shape and are almost rigid, apart from the O – H and CH_3 groups. Also, the O – H hydroxy group in IPC, quinuclidinol and borneol is a hydrogen bond donor, while camphor has the same structure as borneol with the hydroxy group replaced by a carbonyl group. HMIP is rather flat and rigid, with a N – H hydrogen bond donor exposed on one side. For ibuprofen, I

used the RDCs reported in Marathias et al. [102], while the RDCs for 1-phenylethanol were measured by Dr. Emine Sağer and Dr. Pavleta Tzvetkova and are listed in Table A.4. Unlike the other molecules, ibuprofen and 1-phenylethanol are more flexible and possess a phenyl ring, which can rotate with respect to the rest of the molecule. Both also have an O – H hydrogen bond donor.

As described in Section 2.1.2, the ${}^1D_{\mathrm{C-H}}$ RDCs are the same for the three hydrogen atoms in a CH_3 groups and they are averaged over the rotation around the CH_3 group. Therefore, I calculate them from the angle between the A – C bond and the z-axis, where A is the atom the CH_3 group is attached to.

Signs of the RDCs for HMIP

$|J_{ij} + 2D_{ij}|$ can be measured from the NMR spectra and not D_{ij} directly, there are two possibilities for the sign, i.e. $J_{ij} + 2D_{ij} = \pm|J_{ij} + 2D_{ij}|$. If the RDCs are small enough that the assumption $|2D_{ij}| < J_{ij}$ is correct, this is unambiguous because $J_{ij} > 0$. For HMIP, this was not the case however [137]. As the RDCs should be approximately proportional to the quadrupolar coupling of deuterium in deuterated chloroform (see Section 2.1.5), $|J_{ij} + 2D_{ij}|$ was measured for several samples with different quadrupolar coupling in Sager et al. [137]. As shown in Figure A.2, the choice is not clear in many cases. Therefore, I decided to do a fit of the alignment tensor for all combinations of the sign choices instead. This approach should be reliable in this case, because HMIP is very rigid. Additionally, I fixed the signs of the RDCs where the assignment was relatively clear from the quadrupolar coupling: CH-6a, CH-12b, CH-2 and CH-3 for *R*-HMIP and CH_2-6 and CH-6a for *S*-HMIP. Then, I assigned the remaining signs with the alignment tensor fit like before. This resulted in the same signs as when I determined all signs using the alignment tensor fit.

Compared the signs determined in Sager et al. [137], only the sign of CH-5 differs in my assignment for both enantiomers. For *R*-HMIP, the comparison with the quadrupolar couplings in Figure A.2 is not clear for this coupling, while the comparison for *S*-HMIP in Figure A.2 favors the opposite sign of $J_{ij} + 2D_{ij}$. However, the alignment tensor fit gave the same result for both enantiomers (whose RDCs are similar for HMIP). Additionally, this is much more in line with the result from the MD simulations shown below in Figure 4.7, which also has very good agreement of the other RDCs with the experimental values.

For ibuprofen, the sign of some $J_{ij} + 2D_{ij}$ couplings was also ambiguous. However, Marathias et al. [102] report that it could be determined by comparing the measurement between different magnetic field strength, which scales the RDCs linearly (see Equation (2.14)), but leaves the J-couplings unchanged.

3.1.3. Simulations

To prepare the MD simulations, I first equilibrated a cubic box of (deuterated) chloroform with a size of $5.12 \times 5.12 \times 5.12\mathrm{nm}^3$ using the experimental density of $1.48\,\mathrm{g\,cm^{-3}}$ [64].

For this, I used the usual equilibration procedure, with an initial energy minimization to a maximum force of 1000 kJ mol^{-1} nm^{-1}, followed by an NVT equilibration over 100 ps. For the initial NPT equilibration, I first did 200 ps with the stochastic cell rescaling barostat [15], using a relaxation time constant of 1 ps. This was followed by the final equilibration with the Parrinello-Rahman barostat [121] for 1 ns, using a relaxation time constant of 5 ps (which I also used for all other NPT simulations). In both cases, I used a compressibility of 1.086×10^{-4} bar^{-1} and a reference pressure of 1 bar. The final density was 1.49 g cm^{-3}, very close to the experimental value.

Afterwards, I prepared the PBLG system similarly as in Helfrich, Hentschke, and Apel [67] from the 18/5 α-helix. It continues periodically after 18 monomers and 2.7 nm, such that an α-helix consisting of 18 monomers continues unbroken when aligning it along the z-axis and using a periodic box length of 2.7 nm in z-direction. The periodic connection is achieved by adding the respective bonded force field terms between the atoms at the ends of the 18 monomer chain to the topology. Due to the periodic PBLG molecule in the z-direction, I kept the box size fixed in this direction only applied the pressure coupling to the box vectors in the xy-plane, where its direct effect on the PBLG molecule is much smaller due to the solvent. The equilibration of the PBLG system was done in a similar way, with energy minimization to a maximum force of 1000 kJ mol^{-1} nm^{-1}, followed by NVT equilibration over 100 ps and NPT equilibration over 1 ns. For both, the thermostat was separately applied to the solute (i.e. PBLG) and the solvent. The inserted solvent is already sufficiently close to the equilibrated density, such that the Parrinello-Rahman barostat [121] could be applied immediately. No restraints are applied on the PBLG molecule (see Figure 3.1), such that it can freely move in the xy-plane, but the alignment along the z-axis is ensured by the periodic continuation of the α-helix. The system for the actual simulations to calculate RDCs is derived from this box by removing the chloroform molecules in the above PBLG box, inserting the analyte molecule(s) and solvating the box again with chloroform. This is explained further below.

To test whether multiple interacting PBLG chains affect the RDCs, I also did the above procedure with two PBLG chains in the simulation box, which are prepared in the same way. Due to the additional space occupied by the second PBLG molecule, I increased the length of the box vectors in xy-direction to ~ 4.4 nm.

3.1.4. Calculation of residual dipolar couplings

The RDCs are calculated for each snapshot according to Equation 2.14, separately for each molecule in the simulations with multiple molecules. In order to test the validity of taking the bond lengths or the relative orientations of the couplings in the molecule (or both, i.e. using the alignment tensor) from the reference structure, I also fitted the orientation by minimizing the root mean square deviation (RMSD) of the atom positions between the rotated (and translated) reference structure and the structures in the MD trajectory, ignoring the hydrogen atoms due to the flexibility especially of the CH_3 groups. The reference structures are optimized with the DFT implementation in `PySCF` [156], using the B3LYP functional [153] and the def2-SVP basis set [143]. Like described above in

Section 2.1.2, the ${}^1D_{\mathrm{C-H}}$ couplings in a CH_3 group should be treated specially due to the fast rotation of the hydrogen atoms. To calculate the RDCs for these coupling, I therefore used the angle of the C – C bond to the z-axis instead and scaled the result by $-1/3$, while still using the bond lengths of the C – H bonds.

The alignment tensor A can be calculated using Equation (2.16), where the rotation matrices R_t are fit to overlap the reference structure to the structure at time t. For the experimental alignment tensor, I used the optimized reference structure to do a linear least squares fit the alignment tensor to fit the experimental RDCs [95, 21].

As described in Section 2.1.5, the RDCs are only modeled up to a positive constant factor, which mainly depends on the concentration of the alignment medium and its alignment to the magnetic field. Additionally, the strength of the alignment may depend on interactions between the analyte molecules and possibly clustering of the alignment medium molecules. The prediction of the scaling factor would therefore be very challenging, although it can be assumed to be positive. Instead, two sets of RDCs $\vec{D}^1$ and $\vec{D}^2$ can be compared by the cosine similarity

$$S = \frac{\vec{D}^1 \cdot \vec{D}^2}{\|D^1\|\|D^2\|}, \tag{3.1}$$

which is independent of the relative scaling (if it is positive). The often used Pearson correlation coefficient is not entirely adequate, because it is also unchanged by a constant offset and thus can overestimate the agreement, especially for a small number of couplings.

3.1.5. Convergence of the MD average

Due to the weakness of the alignment, the RDCs are usually around 2 to 3 orders of magnitude smaller than the maximum dipolar coupling (46 kHz for a ${}^1D_{\mathrm{C-H}}$ coupling). Therefore, the average needs to be determined to a very high accuracy, requiring very long simulation times. Additionally, the saved trajectories require significant disk space. Because of the temporal correlations, where further apart samples are less correlated, there is a tradeoff between the simulation time and the required disk space. For the time between snapshots, I chose 10 ps, allowing the molecule to undergo a significant rotation between frames.

As there are no large conformational changes of PBLG or the analyte molecules, this average can be calculated from many independent MD simulations, provided that their starting points are (approximately) sampled from the correct distribution. Because of this, I decided to do parallel MD simulations with a length of 100 ns each. To fulfill the above conditions, the analyte molecule is placed randomly in the simulation box with PBLG, after which the solvent is added, taking care that each simulation contains the same number of chloroform molecules. This is achieved by trying different scales for the van der Waals radii of the molecules until the desired number of solvent molecule is inserted. If no such scale is found to match the number of solvent molecules, the largest scale for which too many solvent molecules are inserted is chosen to create the solvated system, from which the additional solvent molecules are then removed. This ensures that the equilibration

works reliably. After that, I equilibrated the systems again like for the PBLG/chloroform system only, with a NPT equilibration time of 1 ns.

For the four spherical-like molecules (IPC, quinuclidinol, borneol and camphor), the alignment is especially weak. Therefore, I decided to do 1000 parallel simulations, with a total simulation time of 100 µs for each molecule. For the other molecules, I chose a simulation time of 30 µs. Another way to achieve more sampling is to do the simulations with multiple analyte molecules the simulation box. However, the interactions between the analyte molecules may influence the alignment and thus the results cannot be expected to be identical. Additionally, the concentration of the analyte molecules in experiment differs for each probe [136], which might influence the RDCs due to interactions between the analyte molecules. Therefore, I tested the simulations with both 1 and 8 analyte molecules in the simulation box and compared the results in as explained in Section 4.1.5.

The different RDCs in a molecule and the dipolar couplings of snapshots in the MD trajectory which are close in time are both strongly correlated. Thus, for a careful comparison with experiment, the covariance matrix of the RDCs needs to be correctly estimated, which can be done using Equation (2.52) [39]. In my case, there are however many independent trajectories. To extend the estimation of γ_t in Equation (2.51) for this case, I make the assumption that the trajectories start from the target distribution. Due to the cylindrical symmetry of the averaged PBLG chain (apart from the backbone chirality), this means that the equilibration time and chain length only need to be large enough for the conformation of the PBLG sidechains and the orientation and distance of the analyte molecules to be independent from the start. The trajectories themselves also need to be long enough to reach a sufficiently large t, which is clearly given for a chain length of 100 ns. γ_t can then be estimated by

$$\hat{\gamma}_t = \frac{1}{\sum_{k=1}^{N} n_k} \sum_{i=1}^{N} \sum_{i=1}^{n_k - t} (X_{k,i} - \hat{\mu}_{k,n_k})(X_{k,i+1} - \hat{\mu}_{k,n_k})^\top, \tag{3.2}$$

where $X_{k,i}$ are the n_k samples of chain k and $\hat{\mu}_{k,n_k}$ is their estimated average as in Equation (2.51).

For the MD simulations shown below in Section 4.1.3 (with the molecules listed in Section 3.1.2), the maximal relative uncertainty $\frac{\max_i \mathrm{Cov}\left[\vec{D}^{T_\mathrm{sim}}\right]_{ii}}{\max_i |D_i^{T_\mathrm{full}}|}$ at different total simulations times is shown in Figure 3.4. Here, $D_i^{T_\mathrm{sim}}$ is the average of coupling i for a total simulation time of T_sim, while T_full is the longest total simulation time for the analyte molecule. The covariance matrix $\mathrm{Cov}\left[\vec{D}^{T_\mathrm{sim}}\right]$ is calculated using Equation (2.52), divided by the number of samples n.

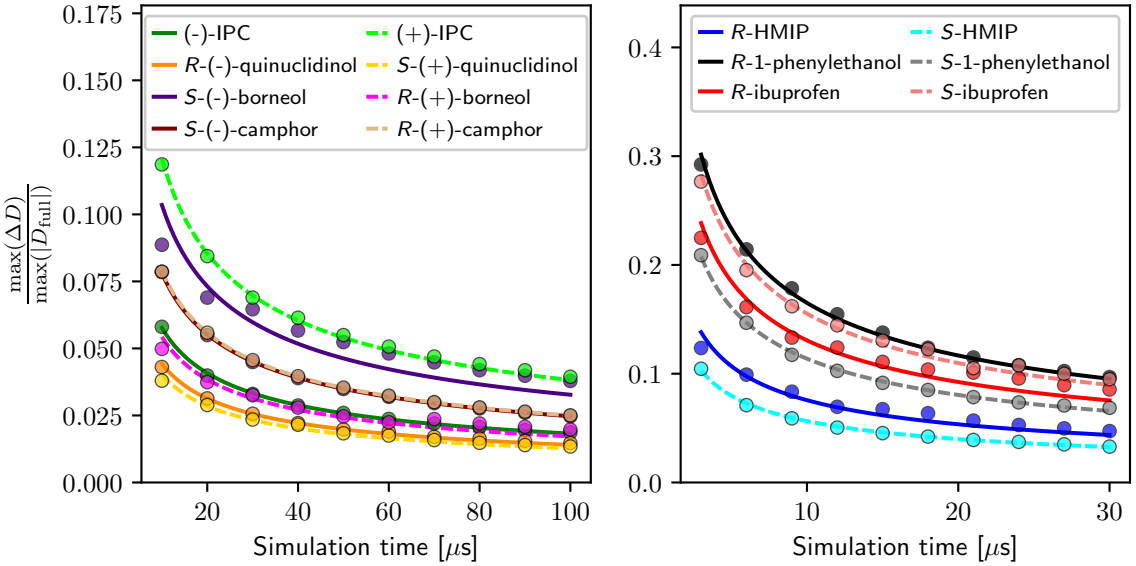


Figure 3.4.: Maximal uncertainty of the RDC average relative to the maximum absolute value of the RDCs for different simulation times. The simulation time indicates the total length of the subset of the full 100 µs or 30 µs trajectory used for the respective average. Under a Markov chain central limit theorem, the uncertainties should behave as $1/\sqrt{N_{\text{samples}}}$; this is confirmed by the solid lines which show a fitted inverse square root law.

3.1.6. Automated workflow

For the 14 molecules treated in this work (see Section 3.1.2), I ran 9800 independent simulations in total, each of which is split into 10 parts of 10 ns. To achieve this and for reproducibility of all calculations, I did all simulations in an automatic workflow using the `Snakemake` workflow management system [108]. A scheme of the steps in the workflow are shown in Figure A.4. All software is managed using the `GNU Guix` package manager [35]. This ensures that the software is almost bit-by-bit reproducible, while the exact source code for all transitive dependencies is known and can still be modified later.

In the workflow, all steps to prepare, run and analyze the simulations are fully automated, making its application for additional molecules straightforward. This includes the preparation of force field parameters using `Ambertools` and the calculation of the ESP for the fitting of RESP charges. It is applicable for an arbitrary number of (the same) analyte molecules in the simulation box, which is taken into account for the analysis of the MD trajectories. For validation, I additionally generalized the simulations to include two PBLG chains in the simulation box, as well as PBLG in vacuum. To test the influence of water molecules, I additionally made a variant for an arbitrary number of water and (the same) analyte molecules in the simulation box. Additionally, all steps to prepare the PBLG molecule as described in Section 3.1.1 are fully automated. The workflow is available under `https://gitlab.kit.edu/david.elsing/RDCMD`.

3.1.7. Analysis of intermolecular interactions between PBLG and the analyte molecules

To calculate the distance between the center of mass (COM) of the analyte molecule and the center axis of the PBLG molecule, the periodic boundary conditions have to be taken into account. Additionally, the PBLG molecule might not be perfectly aligned along the z-axis. Therefore, I first determined the center axis of the PBLG molecule by calculating the eigenvector of the inertia tensor of the backbone atoms corresponding to the smallest eigenvalue. Afterwards, I minimized the distance of the analyte molecule to the determined center axis over the 27 surrounding unit cells. This approach is justified because the α-helix is very stable and the periodic chain of 18 monomers is small enough such that the backbone atoms form a helix whose center lies in an almost straight line.

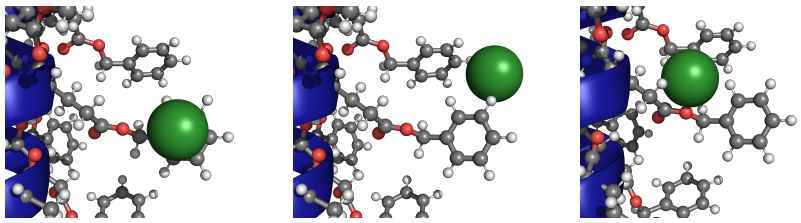


Figure 3.5.: Types of contacts with the PBLG sidechains used for the analysis of the PBLG-analyte interactions. **(a)** Center of the phenyl ring. **(b)** Hydrogen atoms on the side of the phenyl ring. **(c)** The sidechain without the phenyl rings.

For a further understanding of the alignment of the analyte molecules by PBLG, I determined which parts of the PBLG sidechain any part of the analyte molecule is in contact with. The types of contact are shown in Figure 3.5 and defined as follows: Here, a "ring side" contact is counted when an atom of the analyte molecule is within 3 Å from the center of the phenyl ring in the PBLG sidechain. The other contact types "hydrogens of ring" and "sidechain without ring" are defined respectively as any atom j of the analyte molecules being within $1.5\sigma_{ij}$ of a hydrogen atom i in the phenyl ring or of any atom i in the remaining part of the sidechain, where σ_{ij} is the corresponding LJ parameter. Like for the COM distances, all distances between the molecules are calculated taking the periodic boundary conditions into account.

3.2. Adaption of implicit solvation models for the prediction of binding free energies

In order to better understand the energetic interactions leading to the alignment, I analyzed the binding free energies of different poses near PBLG in chloroform calculated using MD simulations. These are then used to fit model parameters of an implicit solvent model. As test molecules, I chose a flat and a spherical-like molecule both for the nonpolar solvation and polar solvation: They are shown in Figure 3.6. Because the nonpolar

contribution describes the solvent interaction with chargeless molecules, I chose coronene and bicyclo[2.2.2]octane (BCO). For the polar solvation however, the hydrogen bonds to PBLG are important, which is why I chose 2-naphthol and bicyclo[2.2.2]octan-1-ol (BCOO), as they both have an O – H hydrogen bond donor.

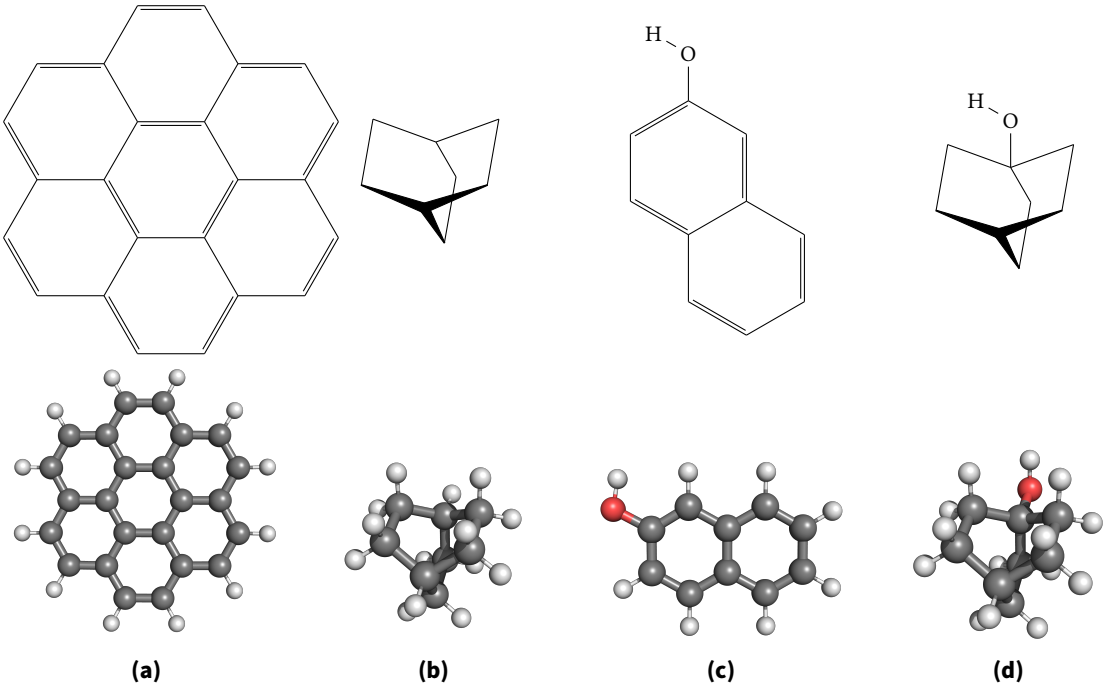


Figure 3.6.: Molecules used for the fit and test of the implicit solvation models. For the nonpolar fit: **(a)** coronene and **(b)** bicyclo[2.2.2]octane (BCO). For the polar fit: **(c)** 2-naphthol and **(d)** bicyclo[2.2.2]octan-1-ol (BCOO).

As the alignment only happens near the PBLG molecule, I'm mostly interested in poses with close contact to PBLG. To generate them, I first generated 1000 poses of the analyte with a center of mass within 14 Å of the center axis of PBLG and with a negative value of the LJ interaction with PBLG. From these poses, I selected 40 final test poses according to equally spaced values of the LJ interaction between the minimum value and 0. These poses are denoted as "LJ" in the following.

Additionally, the molecules which form hydrogen bonds with PBLG are very important for the alignment. For the polar solvation, I therefore used the 20 lowest "LJ" poses per molecule in addition to 20 poses with hydrogen bond. To generate the latter poses, I first drew a random distance between the hydrogen atom and the C = O hydrogen bond acceptor in PBLG from a uniform distribution between the sum of the van der Waals radii of the molecules and 3 Å. Afterwards, I randomly placed the hydrogen atom within the drawn distance from the acceptor and rotate the rest of the molecule around it randomly. If the hydrogen bond angle for the generate pose is $< 140^\circ$ or the atoms of PBLG and the analyte overlap with their van der Waals radii, the pose is rejected. This is repeated until 20 valid poses are generated.

Thus, there are in total 80 different poses for the nonpolar, as well as the polar solvation, for which I calculated the binding free energy using MD simulations. Some of the poses are shown in Figure 3.7.

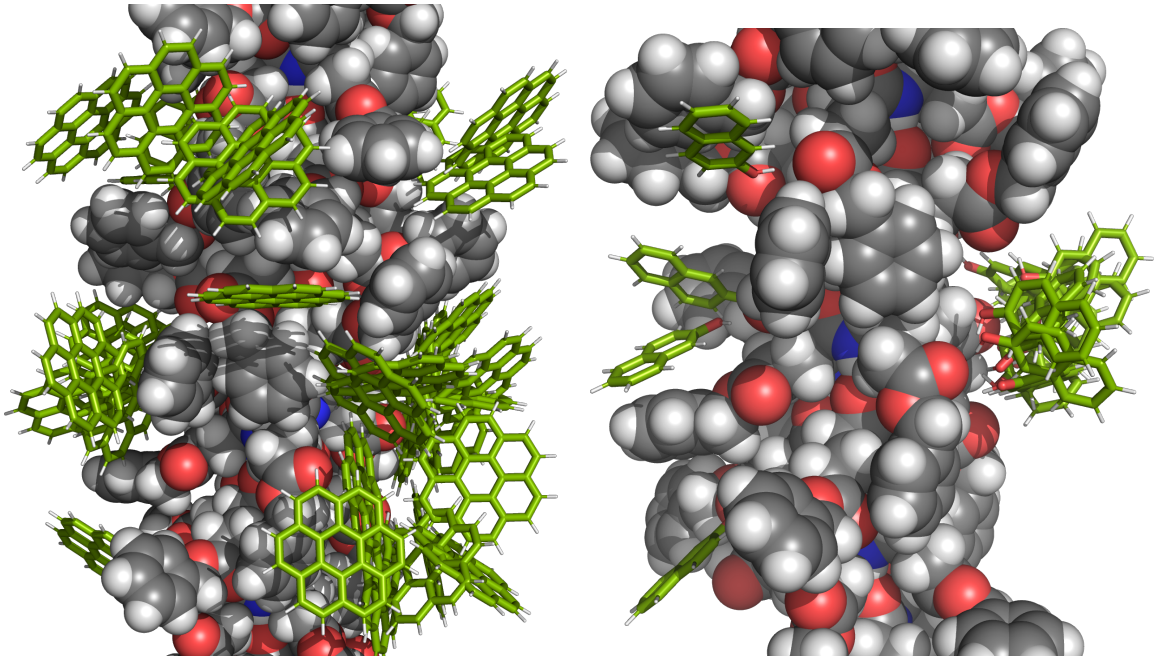


Figure 3.7.: Poses of the analyte molecule near the PBLG molecule used for the fitting of the binding free energies as explained in the main text and in Fig. 3.11. The PBLG model is shown by its vdW surface, while the analyte molecules are displayed in green. In (a), the "LJ" poses for coronene and in (b), the "HB" poses for 2-naphthol are shown.

3.2.1. Free energy perturbation MD simulations

To determine the binding free energies of the test poses from MD simulations, I used alchemical free energy perturbation for the different complexes with restrained solutes. As the force field parameters for chloroform and PBLG are the same as in the MD simulations and determined in the same manner for the test molecules, this provides further understanding of the alignment of analytes by PBLG in the MD simulations.

Nonpolar solvation

The nonpolar solvation free energies are calculated with the charges of the solutes (PBLG + test molecule) set to 0. For this, I wanted to calculate the free energy difference between the restrained complex and the restrained isolated solutes. As the solutes are restrained and only the free energy change due to the solvent is of interest, this can be achieved by calculating the free energy change when decoupling the analyte both in the respective pose near PBLG and without PBLG. Thus, the free energy change of decoupling the PBLG

molecule from the solvent does not need to be calculated, which would be considerably more computationally expensive for the required accuracy due to its much larger size compared to the test molecule.

As the charges are set to 0 in this case and the positions of the solutes are restrained, the decoupling only involves gradually setting the Lennard-Jones interaction between the analyte molecule and the solvent to 0. This can be achieved by using a soft-core potential [18], which gradually switches off the full LJ potential between atoms i and j dependent on the coupling coefficient $\lambda \in [0, 1]$. At $\lambda = 1$, the potential is the original LJ potential, while at $\lambda = 0$, it is set to 0. Using the soft-core potential removes removing the discontinuity at $r_{ij} = 0$ for values of λ in between.

Let $G_{\mathrm{x},\lambda}$ be the free energy of solute x (the complex or the analyte) and λ the coupling coefficient of the analyte. The binding free energy contribution due to the nonpolar solvation is then

$$\Delta\Delta G_{\mathrm{nonpolar}} = G_{\mathrm{complex},\lambda=0} - G_{\mathrm{complex},\lambda=1} - (G_{\mathrm{analyte},\lambda=1} - G_{\mathrm{analyte},\lambda=0}). \tag{3.3}$$

The differences between the $G_{\mathrm{x},\lambda}$ can be calculated from the simulations at the different λ using the Multistate Bennett Acceptance Ratio (MBAR) method [147]. For this, I used the implementation in the `pymbar` [29] and `alchemlyb` [1] libraries. To estimate the error of the calculation, the trajectory was subsampled such that the remaining samples are effectively uncorrelated. For the approach described here, the average uncertainties are shown in Table 4.2. To determine $\Delta\Delta G_{\mathrm{att}}$, different simulations of the separate PBLG and analyte molecules, as well as of their complex for the test poses are required. These calculation are set up in the same way as for the calculation of $\Delta\Delta G_{\mathrm{nonpolar}}$, but only with the full LJ potential between the solutes and the solvents. For this reason, a longer simulation time is required: I chose 50 independent simulations for each configuration, with a simulation time of 1 ns each, where the trajectory is again saved every 1 ps. $\Delta\Delta G_{\mathrm{rep}}$ is then determined simply as the difference between $\Delta\Delta G_{\mathrm{nonpolar}}$ and $\Delta\Delta G_{\mathrm{att}}$.

Polar solvation

Unlike for the implicit solvation model, where the solvation free energies are split into a nonpolar and a polar part, the total free energy difference for solutes with partial charges determined from MD simulations is calculated directly. The "polar binding free energy contribution" is then defined in accordance with the implicit solvation model as the difference between the total solvation contribution and the nonpolar contribution; it does not correspond to an actual binding free energy difference between two states in the MD simulations.

To calculate the total total binding free energies, I used a similar procedure as for the nonpolar solvation, where the solutes are treated as rigid. For this however, the charges of the solutes are nonzero and the Coulomb potential between the test molecule needs to be switched off gradually as well. As suggested in Beutler et al. [18], I first switched

off the Coulomb potential with a coupling coefficient λ_C and then the LJ potential with a coupling coefficient λ_{LJ}, where the test molecule charges are set to 0 (corresponding to $\lambda_C = 0$). The latter part is different from the calculation of $\Delta\Delta G_{\text{nonpolar}}$, because here, the PBLG charges are still present, which influences the solvent and should be included in the polar solvation model. The total solvation contribution to the binding free energy is therefore

$$\begin{aligned}\Delta\Delta G_{\text{total solv}} = \\ G_{\text{complex},\lambda_C=0,\lambda_{LJ}=0} \underbrace{-G_{\text{complex},\lambda_C=1,\lambda_{LJ}=0} + G_{\text{complex},\lambda_C=1,\lambda_{LJ}=0}}_{=0} - G_{\text{complex},\lambda_C=1,\lambda_{LJ}=1} \\ - (G_{\text{analyte},\lambda_C=0,\lambda_{LJ}=0} \underbrace{-G_{\text{analyte},\lambda_C=1,\lambda_{LJ}=0} + G_{\text{analyte},\lambda_C=1,\lambda_{LJ}=0}}_{=0} - G_{\text{analyte},\lambda_C=1,\lambda_{LJ}=1}),\end{aligned} \tag{3.4}$$

where the free energy differences of G_{complex} and G_{analyte} are again calculated with the MBAR method. For each structure, this is split into the two parts where λ_C or λ_{LJ} are switched off.

3.2.2. Implicit solvation model for binding free energy estimation

The implicit solvation model is split into a nonpolar and polar part [130], both of which change if the analyte is near the macromolecule. Therefore, the total binding free energy can be calculated as

$$\Delta\Delta G_{\text{total}} = \Delta\Delta G_{\text{nonpolar}} + \Delta\Delta G_{\text{polar}} + V_{LJ} + V_{\text{Coulomb}}/\epsilon_{\text{in}}, \tag{3.5}$$

where $\Delta\Delta G_{\text{nonpolar}}$ and $\Delta\Delta G_{\text{polar}}$ are the changes in the nonpolar and the polar part, respectively. V_{Coulomb} and V_{LJ} are the Coulomb and Lennard-Jones potential between the analyte and macromolecule; they are both (apart from the periodic boundary conditions) identical as in the MD simulations.

Nonpolar solvation

The change of the attractive part of the nonpolar solvation (see Equation (2.35)) upon binding $\Delta\Delta G_{\text{att}}$ is the integral of $\bar{\rho} V_{LJ,\text{att}}$ over the volume where the solvent is excluded due to the respective other solute. It can thus be calculated as

$$\Delta\Delta G_{\text{att}} = -\bar{\rho}\left[\sum_{i\in\text{macrom.}} \int_{\Omega_{i,\text{macrom.,lig}}} V_{LJ,\text{att},i}(\vec{x}_i, \vec{y})\, d^3y + \sum_{i\in\text{lig.}} \int_{\Omega_{i,\text{lig.,macrom.}}} V_{LJ,\text{att},i}(\vec{x}_i, \vec{y})\, d^3y\right]. \tag{3.6}$$

Here, $\Omega_{i,a,b}$ is the volume outside solute a occupied by the other solute b, where the atoms of the solutes are represented as balls $B_r(\vec{x})$ with their radii r being the sum of the atom

radius $0.5\sigma_k$ and a constant extension representing the probe radius of the solvent Thus, $\Omega_{i,a,b}$ is defined as

$$\Omega_{i,a,b} = \bigcup_{\substack{j \in b \\ |\vec{r}_i - \vec{r}_j| < r_{\text{cutoff}}}} B_{0.5\sigma_j + R_{\text{other}}}(\vec{x}_j) \setminus \bigcup_{i' \in a} B_{0.5\sigma_{i'} + R_{\text{self}}}(\vec{x}_{i'}). \tag{3.7}$$

Because the solvent density is approximated to be constant outside and zero inside the solute, the radii R_{other} and R_{self} are effective radii and need not be equal. I therefore determine them as fit parameters for the nonpolar model. r_{cutoff} is a numerical cutoff radius, which should be chosen large enough for sufficient accuracy.

For the repulsive part of the nonpolar solvation, the contribution to the binding free energy is simply determined in the change of the SASA and SAV:

$$\Delta\Delta G_{\text{rep}} = \gamma \Delta\text{SASA} + p\Delta\text{SAV} \tag{3.8}$$

with a separate probe radius R_{rep}, which is also a fit parameters.

The different radii used for the nonpolar solvation model are visualized in Figure 3.8.

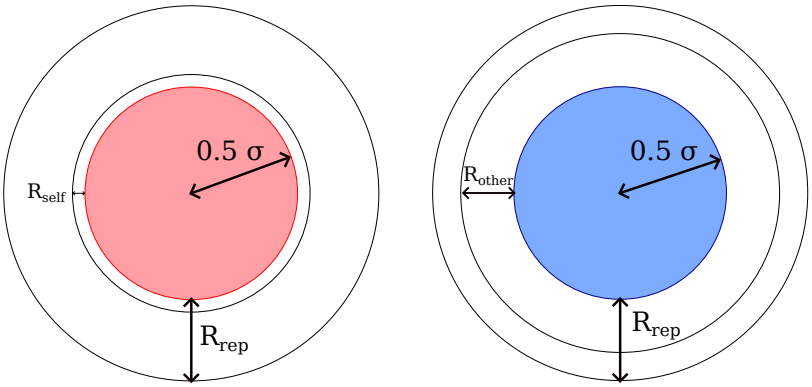


Figure 3.8.: The different radii used for $\Delta\Delta G_{\text{nonpolar}}$. Here, R_{self} is the probe radius for the same solute that the atom belongs to for which the contribution to $\Delta\Delta G_{\text{att}}$ is calculated (marked in red), while R_{other} is the probe radius of the respective other solute. When the contribution for an atom of the other solute is calculated, their role changes. R_{rep} is the probe radius used for $\Delta\Delta G_{\text{rep}}$.

In the above model for $\Delta\Delta G_{\text{att}}$, the density of chloroform is assumed to be uniform everywhere outside the solutes and zero inside them, with effective probe radii R_{self} and R_{other} used to determine the excluded volume due to the solutes. Here, the question arises whether this approximation is reasonable. To investigate its importance, I modified the integrals in Equation (3.6) by multiplying the integrand with the radial distribution function (RDF) $g_i^{\text{rad}}(r)$ of the atom i to the center of mass in chloroform. This means that in Equation (2.35), the approximation $g(\vec{y}) \approx g_i^{\text{rad}}(\|\vec{x}_i - \vec{y}\|)$ is made for atom i, where $r = \|\vec{x}_i - \vec{y}\|$ is the distance of the solvent center of mass to the atom i. $g_i^{\text{rad}}(r)$ can be calculated by MD simulations for single, uncharged atom with the Lennard-Jones parameters of atom i; this is explained in more detail in Section 4.2.2. Additionally, the volume used for the integral needs to include the atom i itself: This means replacing "$i' \in a$" by "$i' \in a \setminus \{i\}$"

in Equation (3.7). It does not lead to a divergent integral, because $g_i^{\mathrm{rad}}(r) = 0$ for small r, which however needs to be treated carefully in the numerical implementation.

In the following, using the uniform density is referred to as "step" RDF, while the use of the RDF determined from MD is referred to as "calc." RDF.

Polar solvation

For the polar solvation, I used the ALPB model [148] (Equation (2.44)) together with the ΔHCT Born radii [92] described above in Section 2.5.2. In the following, this approach is called ALPB-ΔHCT. When the solutes get close, the Born radii of the atoms change, as well as the effective molecule size for the ALPB model changes. However, I neglect the second change for simplicity and assume it to be constant, such that only the contributions between the solutes and due to the change of the Born radii need to be considered.

Because the sum in Equation (2.44) also goes over pairs of atoms within a solute, they need to be recalculated as well. Additionally, the $1/f_{ij}$ term is a long-range interaction and cannot be described by a simple cutoff. Fortunately however, $f_{ij}(r) \approx r$ for large r and the other terms do not depend on the Born radii, such that the change of ΔG^{ALPB} can be described by only local interactions.

With $\Delta G_{ij}^{\mathrm{ALPB}}$ being the ALPB contribution with the Born radii of the isolated molecules and $\Delta \tilde{G}_{ij}^{\mathrm{ALPB}}$ the contribution using the Born radii for the complex, the change of ΔG^{ALPB} is then

$$\begin{aligned}
\Delta\Delta G_{\mathrm{polar}} = & \sum_{\substack{i\in\text{analyte},\\ j\in\text{macromolecule}}} \Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} \\
& + \sum_{\substack{i,j\in\text{analyte},\\ j\le i}} (\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} - \Delta G_{i,j}^{\mathrm{ALPB}}) + \sum_{\substack{i,j\in\text{macromolecule},\\ j\le i}} (\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} - \Delta G_{i,j}^{\mathrm{ALPB}}) \\
\approx & - c_{\mathrm{ALPB}} V_{\mathrm{Coulomb}} + \sum_{i\in\text{analyte}} \sum_{\substack{j\in\text{macromolecule},\\ |\vec{r}_i-\vec{r}_j|<r_{\min}}} \left(\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} + \frac{c_{\mathrm{ALPB}} q_i q_j}{4\pi\epsilon_0 r_{ij}}\right) \\
& + \sum_{\substack{i,j\in\text{analyte},\\ |\vec{r}_i-\vec{r}_j|<r_{\min}}} (\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} - \Delta G_{i,j}^{\mathrm{ALPB}}) + \sum_{\substack{i,j\in\text{macromolecule},\\ |\vec{r}_i-\vec{r}_j|<r_{\min}}} (\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} - \Delta G_{i,j}^{\mathrm{ALPB}}),
\end{aligned} \tag{3.9}$$

where c_{ALPB} is defined in Eq. 2.44. $r_{\min}$ is a cutoff radius beyond which the approximation $f_{ij} \approx r_{ij}$ is made. There, $\Delta\tilde{G}_{i,j}^{\mathrm{ALPB}} \approx \Delta G_{i,j}^{\mathrm{ALPB}} \approx -c_{\mathrm{ALPB}} V_{\mathrm{Coulomb}}$. As described in Section 2.5.2, the Born radii for $\tilde{G}^{\mathrm{ALPB}}$ can be approximated by the ΔHCT model [92]. However, I found that the ΔHCT method was accurate enough for my purposes even for the "primary atoms", such that a more exact calculation for closer atoms is not required.

3.2.3. Model implementation

All discussed models are implemented in the `MRAE` library available under `https://gitlab.kit.edu/david.elsing/mrae`. For all models used here (except the electrostatic term of P3D), including the determination of overlaps of the SAV or the atoms by the van der Waals radii, it is for each pose required to find the PBLG atoms which are inside a given radius for each analyte atom. As the PBLG molecule is rigid for the MC integration, I implemented this by precomputing cells lists around it by creating a grid of cells, for each of which the indices of the PBLG atoms within the cutoff of the cell are stored. For each position of a analyte atom, only the PBLG atoms corresponding to the indices of the cell the atom is in have to be checked.

The individual contributions are then calculated as described in the following. Even though motivated by describing the interactions between an analyte molecule and PBLG, the implementation works for any two rigid solutes.

Implementation of pairwise contributions

The pairwise interactions V_{LJ} and V_{Coulomb} can be determined numerically rather easily by precalculating a grid around PBLG for the long-range contributions, which is then linearly interpolated between the 8 neighboring grid points. For better accuracy, the short-range contributions are calculated directly by a pairwise sum of nearby atoms. As both potentials diverge for $r \to 0$ and the long-range values are determined by linear interpolation, I actually extended the potential linearly for small $r < r_{\min}$ and later correct it for the short-range contributions closer than the cutoff $r_{\min}$. The grid potential for the respective pairwise potential V_{ij} is defined as

$$V_{\mathrm{grid}}(\vec{r}_i) = \sum_{j \in \mathrm{macromolecule}} \begin{cases} V_{ij}(r_{ij}), & r_{ij} >= r_{\min} \\ V_{ij}(r_{\min}) - (r_{\min} - r) V'_{ij}(r_{\min}), & r_{ij} < r_{\min}, \end{cases} \tag{3.10}$$

where $r_{ij} = |\vec{r}_i - \vec{r}_j|$. Given the linearly interpolated value $\tilde{V}_{\mathrm{grid}}$, the contributions of the atom pairs (i, j) which are closer than $r_{\min}$ are corrected in the following manner, yielding the total energy

$$V \approx \tilde{V}_{\mathrm{grid}}(\vec{r_{ij}}) - \sum_{\substack{j, \\ r_{ij} \leq r_{\min}}} \left(V_{\mathrm{grid}}(r_{ij}) + V_{ij}(r_{ij}) \right). \tag{3.11}$$

Polar solvation

For the P3D model explained in Section 2.6.3, the precomputed potential grid with a grid spacing of 0.39 Å is simply linearly interpolated. I implemented the ALPB-ΔHCT model according to Equation (3.9). For the contribution between the solutes, this again entails a sum over nearby atoms within $r_{\min}$. V_{Coulomb}, which is already calculated with the

grid-based approach described above, can be reused for the long-range part. The modified Born radii are calculated with the ΔHCT model [92] (Equation (2.45)), which is again a sum over nearby atoms within $r_{\Delta\mathrm{HCT}}$, using the precomputed Born radii of the isolated molecules as well as the HCT structure factors. For the isolated molecules, the perfect Born radii are compared to the Born radii from the HCT model whose scaling factors were fit to the perfect Born radii in Figure A.34. For the test molecules, the fit is quite good, while there are larger deviations for the PBLG atoms, especially for larger Born radii (where the atoms are more buried). As I show in the results, using the ΔHCT model for all atoms close to the other solute is however sufficiently accurate in my case.

For the contribution within a solute, I again used the treatment of the solutes as rigid to precompute the indices of the nearby atoms (within r_{min}) in the same solute. When evaluating the sums in Equation (3.9), the sum needs to be applied for all atom pairs where at least one atom of the pair has a modified Born radii due to the ΔHCT model.

Nonpolar solvation

To calculate the repulsive part of the nonpolar solvation contribution to the binding free energy, the volume and surface of the SAV overlap between the two solutes is required (see Equation (3.8)). As I already need to calculate surface integrals over an overlap of the solutes (with different atom radii) for the attractive part anyway (see below), I decided to apply the same methodology for the repulsive part as well.

For the surface integrals, an efficient method with variable accuracy is required. A commonly used method for surface integrals on a three-dimensional sphere in quantum chemistry programs is Lebedev quadrature [87], which is based on points on a sphere which are symmetric under octahedral rotation and inversion and which integrate polynomials of a given degree exactly. Additionally, precomputed grids exist for many different numbers of points [23]. As all atoms are treated as spheres with different radii, I therefore precalculate Lebedev grid points on the atoms which lie outside the molecule. For each analyte pose, I then first determine the atoms which overlap with the other solute, and afterwards the surface points of these atoms which are inside the other respective solute.

Given the Lebedev quadrature points $\vec{x}_{ik}$ with weights w_k (normalized such that $\sum_k w_k = 1$), the change of the SASA because of the overlap is given by

$$\Delta\mathrm{SASA} = -4\pi \left[\sum_{i \in \mathrm{macrom.}} \sum_{\substack{k \\ \vec{x_{ik}} \text{ inside lig.}}} w_k (0.5\sigma_i + R_{\mathrm{rep}})^2 + \sum_{i \in \mathrm{lig.}} \sum_{\substack{k \\ \vec{x_{ik}} \text{ inside macrom.}}} w_k (0.5\sigma_i + R_{\mathrm{rep}})^2 \right]. \tag{3.12}$$

To calculate the volume of the SAV overlap as a surface integral as well, the divergence theorem can be used by observing that $\vec{\nabla}(\vec{x} - \vec{x}_{\text{ref}})/3 = 1$, such that

$$\begin{aligned}\Delta\text{SAV} = -\frac{4}{3}\pi\Bigg[&\sum_{i\in\text{macrom.}} \sum_{\substack{k \\ \vec{x_{ik}} \text{ inside lig.}}} w_k(0.5\sigma_i + R_{\text{rep}})^2(\vec{x}_{ik} - \vec{x}_{\text{ref}}) \cdot \vec{n}_{ik} \\ &+ \sum_{i\in\text{lig.}} \sum_{\substack{k \\ \vec{x_{ik}} \text{ inside macrom.}}} w_k(0.5\sigma_i + R_{\text{rep}})^2(\vec{x}_{ik} - \vec{x}_{\text{ref}}) \cdot \vec{n}_{ik}\Bigg].\end{aligned} \tag{3.13}$$

where $\vec{n}_{ik}$ is the unit normal vector of point $\vec{x}_{ik}$, pointing outward from atom k. To improve the numerical accuracy, $\vec{x}_{\text{ref}}$ should be chosen near the surface so that the individual terms do not become too large. I therefore put it at the average position of the atoms involved in the surface integral.

For the attractive part of the nonpolar solvation, I need to calculate Equation (3.6) for each analyte pose, which entails an integral for all atoms near the respective other solute. This is much more computationally expensive than the other contributions, so I decompose it into two parts for each atom: The contribution if the solute the atom belongs to only consists of the atom itself, as well as the correction due to the other atoms in the same solute:

$$\Delta\Delta G_{\text{att}} = \underbrace{\sum_{i\in\text{macrom.}} \Delta\Delta G_{i,\text{ single}} + \sum_{i\in\text{lig.}} \Delta\Delta G_{i,\text{ single}}}_{\text{calc. with precomputed grid}} + \underbrace{\sum_{i\in\text{macrom.}} \Delta\Delta G_{i,\text{ corr.}} + \sum_{i\in\text{lig.}} \Delta\Delta G_{i,\text{ corr.}}}_{\text{calc. with surface integral}}. \tag{3.14}$$

The first part only depends on the position of the single atom around the other solute, which can be again determined by linear interpolation of a precomputed grid. This is visualized in Figure 3.9a, where the integration is done over the volume indicated by the dots, which lie outside the single atom, but within the other solute and the numerical cutoff.

The second part still needs to be calculated directly, but only in the volume where the solutes overlap (with the atom radii extended by R_{other} or R_{self}), which is much more local than the total cutoff r_{cutoff} in Equation (3.6). As mentioned in Floris and Tomasi [51] and Tan, Tan, and Luo [157] and like for Equations (3.12) and (3.13), these kinds of integrals can be calculated as surface integrals using the divergence theorem, which is especially suited here due to the limited spatial extent of the overlap volume, making the accurate numerical calculation of the surface integrals feasible.

To determine the integrand for the surface integral in Equation (2.38), the vector field $\vec{A}_i$ is required for each atom i. For the WCA and σ decompositions with a step function for g, the analytical expressions are given in Tan, Tan, and Luo [157], which is what I use for the "step" RDF.

For the "calc." RDF however, $\vec{A}_i$ needs to be calculated numerically using Equation (2.39) [51], with $V_{\mathrm{LJ,att},i}$ modified by $g_i^{\mathrm{rad}}(x)$, such that

$$f(r) = \frac{1}{r^3} \int_0^r x^2 g_i^{\mathrm{rad}}(x) V_{\mathrm{LJ,att},i}(x)\, dx. \tag{3.15}$$

Because here, the integration volume may include $r = 0$, it is important that the boundary conditions for $\vec{A}$ (which do not matter for the surface integral) are chosen such that $f(r = 0) = 0$. This is different from the choice in Tan, Tan, and Luo [157], where the integral is calculated for the entire volume *outside* the solutes.

More difficult is the determination of the surface points to do the integration numerically. With the "step" RDF, the corresponding volume used for the integral for atom i is the intersection of the solute volume outside atom i with the volume of the other solute. This is visualized by the stripes in Figure 3.9b and the corresponding surface by the green points in Figure 3.10a. For the "calc." RDF, the volume is the same as for the "step" RDF, but including the volume of atom i which lies both inside the other atoms in the same solute and the other solute. This is shown by the dots in Figure 3.9b and again the corresponding surface in Figure 3.10b.

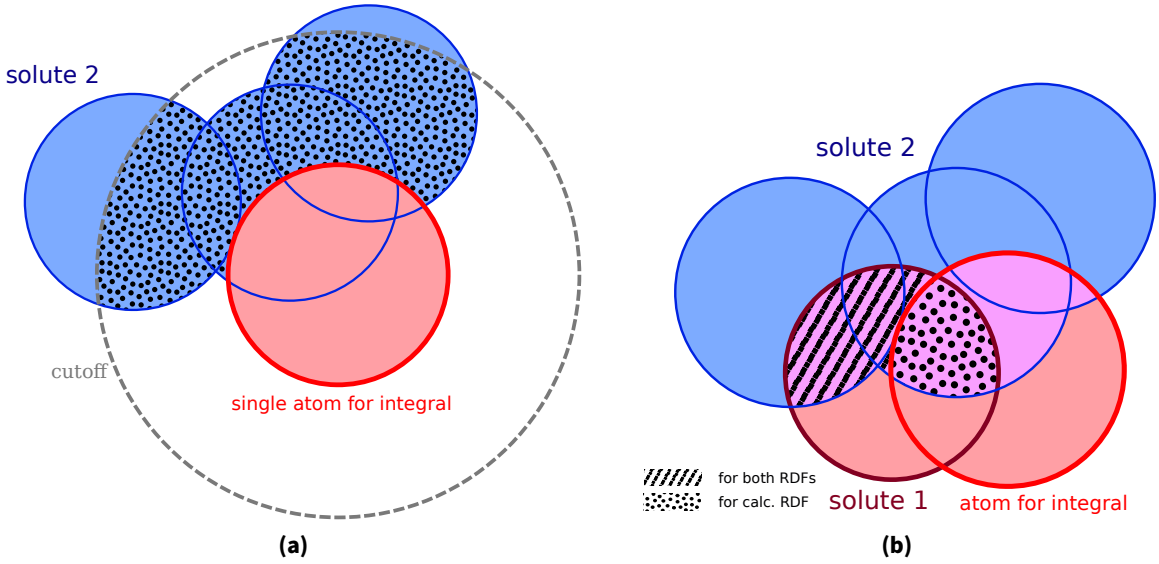
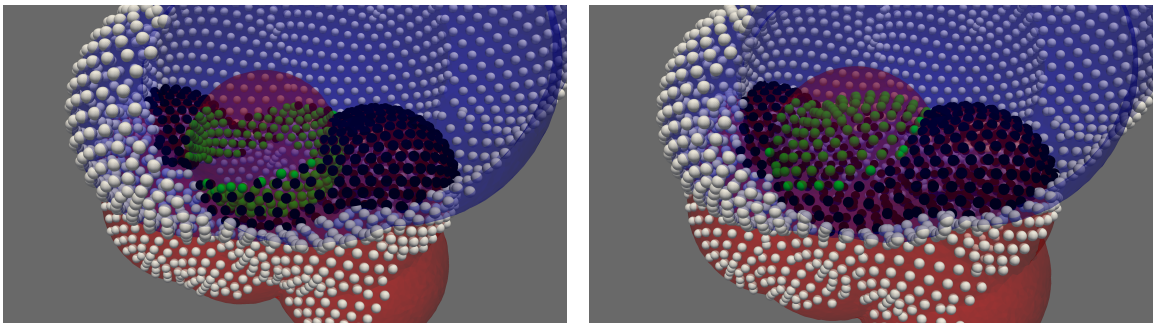


Figure 3.9.: (a) Visualization of the domain of the integral (in blue) specified in Equation 3.6 for a single atom of one of the solutes (in red). The dots indicate the surface of solute 2 and the cutoff radius for the integral is shown in gray. (b) Visualization of the correction for the calculation of the attractive part of the nonpolar solvation for multiple atoms in solute 1. It is added to the contribution for the single atom represented in (a) and is evaluated by a surface integral. For the "step" RDF option, the correction integral goes over the volume marked by stripes, while for the "calc." RDF option, it includes the volume marked by dots. For a corresponding 3D visualization, see Fig. S1.

Like for $\Delta\Delta G_{\mathrm{rep}}$, I calculate the surface integrals by Lebedev quadrature points on the atoms. For the "step" RDF, this requires the points of the molecule surface with the atom radii being extended by both R_{self} and R_{other} for the two parts in Equation (3.6). For the "calc." RDF, the situation is again a bit more complicated as shown in Figure 3.9b and



(a) (b)

Figure 3.10.: Visualization of the points for the surface integral used for the numerical calculation of $\Delta\Delta G_{\text{att}}$, corresponding to Fig. 3 in the main text. Here, the points used for the contribution of the atom in the center of the images (colored in red) are shown. The integral goes over the volume enclosed by the surface indicated by the black and green points. In (a), the surface for the "step" RDF method is shown and in (b) for the "calc." method.

Figure 3.10b. Here, for each atom i, the surface formed by all other atoms except i which lies inside atom i needs to be calculated for the atom radii extended by R_{self}. Because the solutes are treated as rigid, these surface points can be precalculated and are transformed by the offset and rotation with the molecule itself during the evaluation in different poses. As for $\Delta\Delta G_{\text{rep}}$, I first determine the atoms which overlap with the other solute (with R_{self} for one solute and R_{other} for the other, and vice versa) and the surface points for the atoms with overlap.

The influence of the numerical parameters for the evaluation of the LJ and Coulomb potentials and the implicit solvation models, such as the grid spacings or cutoffs, is shown in Appendix A.13 together with the chosen values. As shown in Figures A.28b, A.31, A.33b and A.35b, they chosen such that the numerical error is negligible in all cases.

3.2.4. Fitting procedure

To determine the model parameters, I fitted them to the binding free energies calculated from MD as explained above. For the nonpolar solvation, I fitted $\Delta\Delta G_{\text{nonpolar}}$ for the 80 test poses of coronene and BCO, while for the polar solvation, I did the fit of $\Delta\Delta G_{\text{polar}}$ for the 80 test poses of 2-naphthol and BCOO. A scheme of the fitting procedure is shown in the upper part of Figure 3.11. Like the MD simulations for the calculations of RDCs, it is automated in a workflow available under `https://gitlab.kit.edu/david.elsing/pblg_tcm_ddg`.

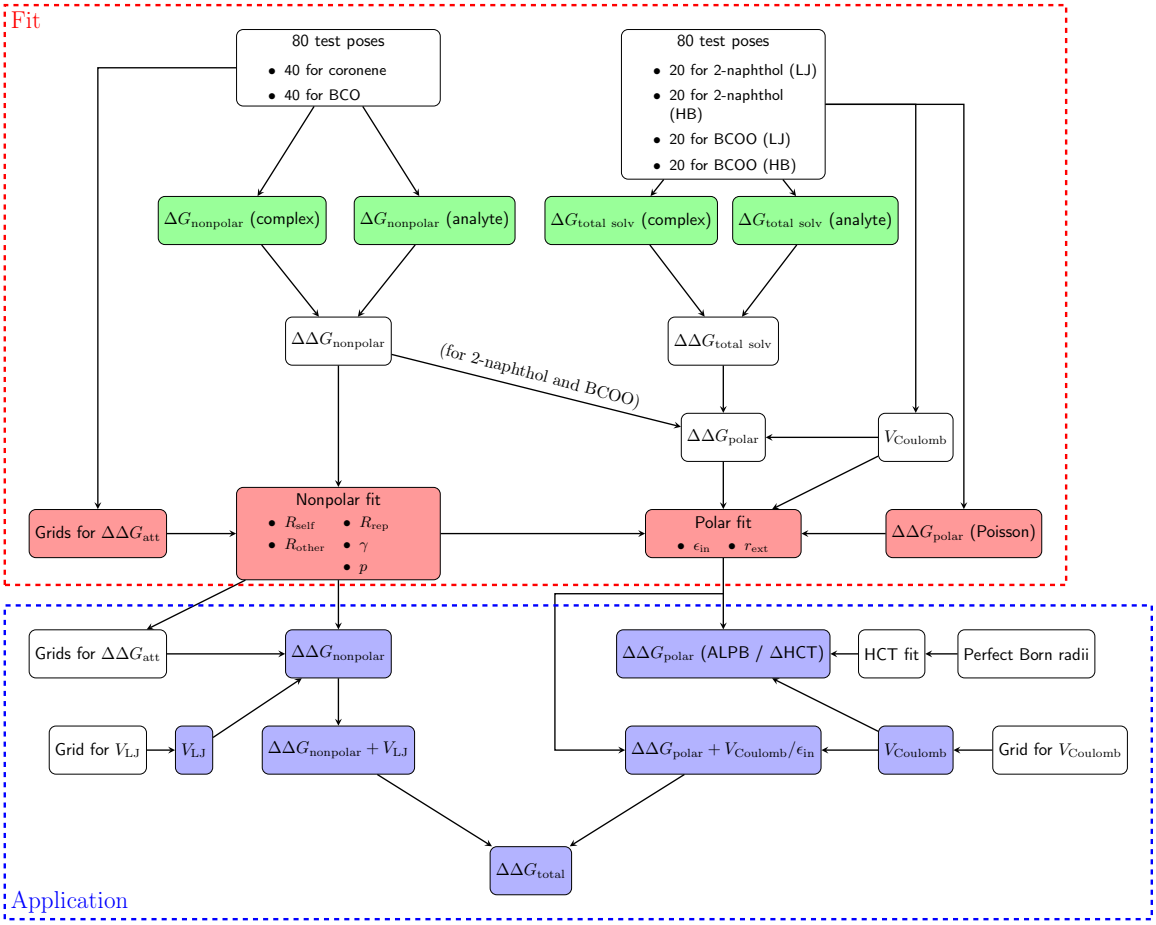


Figure 3.11.: Fitting procedure and steps for the evaluation of the binding free energies. In green, the MD simulations for calculating alchemical binding free energies are shown. The grids for $\Delta\Delta G_{\text{att}}$ (in red) are calculated for all tested R_{self} and R_{other}, while $\Delta\Delta G_{\text{polar}}$ (Poisson) (also in red) is calculated for each pose and for all tested ϵ_{in} and r_{ext}. For the application of the model, the terms in blue are evaluated for each pose, while the other parts can be precomputed.

Nonpolar fit

In the fit of $\Delta\Delta G_{\text{nonpolar}}$, there are 5 fit parameters: The probe radii R_{self} and R_{other} for $\Delta\Delta G_{\text{att}}$ and R_{rep}, γ and p for $\Delta\Delta G_{\text{rep}}$. Additionally, I did the fit separately for the 4 combinations of the WCA and σ decompositions, as well as the "step" and "calc." RDFs.

For the $\Delta\Delta G_{\text{rep}}$ contribution, calculating the values for the different model parameters only requires calculating the surface integrals for each tested R_{rep}. As it depends linearly on γ and p, they can be determined by linear least squares. With $\Delta\Delta G_{\text{att}}$, the fitting procedure is more computationally expensive due to the required precomputed grids, which need to be calculated for each set of the relevant model parameters for both PBLG and the analytes. For the "step" RDF, the grids depend on both R_{self} and R_{other}, while for the "calc." RDF, they only depend on R_{other}. Finally, the optimal parameters for $\Delta\Delta G_{\text{nonpolar}}$ are determined by minimizing all combinations of the probe radii, and doing the linear least squares fit of γ and p for each combination.

Polar fit

The polar model has 2 fit parameters used for $\Delta\Delta G_{\text{polar}}$ calculated with the Poisson equation: The probe radius r_{ext} and the internal dielectric constant of the solutes ϵ_{in}. V_{Coulomb} is the same for the MD and the model values (up to the numerical accuracy) and so the fit is equivalent if done for $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$. Therefore, I calculated $\Delta\Delta G_{\text{polar}}$ for different combinations of r_{ext} and ϵ_{in} and choose the values with the lowest root mean square deviation (RMSD) compared to the MD values. To find suitable values, the fit is done in two stages: first with r_{ext} between 0.30 Å to 0.70 Å in steps of 0.05 Å and ϵ_{in} between 1.0 to 1.3 in steps of 0.1. As I show below, this resulted in $\epsilon_{\text{in}} = 1.0$, which is expected, as the MD simulations are done without polarization. In the second stage, I thus minimize the RMSD to the MD values only for r_{ext} between 0.25 Å to 0.45 Å in steps of 0.02 Å.

3.3. Monte Carlo Methods for the calculation of RDCs

Due to the uniform nature of the PBLG helix, it seems reasonable to expect that the RDCs can be calculated using a sufficiently long PBLG chain, which is kept completely rigid. If the solvent is neglected as well, this would make the calculation of RDCs by Monte Carlo integration, where the average is calculated over all possible orientations around the rigid PBLG molecule. Indeed, it was already demonstrated with the P3D method [116] that for different PBLG chains with a length in the order of 10 nm, the RDCs are very similar.

In the following, I show the validity of this approach: To evaluate Equation (2.34), the generalized coordinates $q_1 \in \Gamma_1$ can be split into three parts: $q_{\text{AM}} \in \Gamma_{\text{AM}}$ for the configuration of the alignment medium, $q_{\text{L}} \in \Gamma_{\text{L}}$ for the analyte molecule structure and $q_{\text{COM}} \in \Gamma_{\text{COM}}$ for the analyte molecule center of mass position and its rotation. This decomposition is not unique, but this only affects the normalization constant in Equation (2.34). If the alignment medium and the analyte molecule are infinitely separated, the free energies $F(q_1)$ can be decomposed into two additive parts (up to a constant), i.e. $F_{\text{sep}}(q_{\text{AM}}, q_{\text{L}}) = F_{\text{AM}}(q_{\text{AM}}) + F_{\text{L}}(q_{\text{L}})$. The binding free energy can then be defined as

$$F_{\text{binding}} \coloneqq F(q_{\text{COM}}, q_{\text{AM}}, q_{\text{L}}) - F_{\text{sep}}(q_{\text{COM}}, q_{\text{AM}}, q_{\text{L}}). \tag{3.16}$$

Thus, the average in Equation (2.34) becomes

$$\begin{aligned}\langle A \rangle &= \frac{1}{N_1} \int_{\Gamma_1} A(q_1) e^{-\beta(F_{\text{binding}}(q_{\text{COM}}, q_{\text{AM}}, q_{\text{L}}) + F_{\text{AM}}(q_{\text{AM}}) + F_{\text{L}}(q_{\text{L}}))} \, dq_{\text{L}} \, dq_{\text{AM}} \, dq_{\text{COM}} \\ &= \frac{1}{N_1} \int_{\Gamma_{\text{AM}}} e^{-\beta F_{\text{AM}}(q_{\text{AM}})} \int_{\Gamma_{\text{L}}} e^{-\beta F_{\text{L}}(q_{\text{L}})} \int_{\Gamma_{\text{COM}}} A(q_1) e^{-\beta F_{\text{binding}}(q_{\text{COM}}, q_{\text{AM}}, q_{\text{L}})} \, dq_{\text{COM}} \, dq_{\text{L}} \, dq_{\text{AM}}.\end{aligned} \tag{3.17}$$

If the assumptions are made that the alignment medium is large enough such that the dependence of the average on q_{AM} is negligible, and the analyte molecule is rigid such that the dependence of q_{L} is negligible, the average becomes

$$\langle A \rangle = \frac{1}{\tilde{N}} \int_{\Gamma_{\text{COM}}} A(q_{\text{COM}}, \tilde{q}_{\text{AM}}, \tilde{q}_{\text{L}}) e^{-\beta F_{\text{binding}}(q_{\text{COM}}, \tilde{q}_{\text{AM}}, \tilde{q}_{\text{L}})} \, dq_{\text{COM}}, \tag{3.18}$$

where $\tilde{q}_{\mathrm{AM}}$ and $\tilde{q}_{\mathrm{L}}$ are the chosen alignment medium and analyte structures and $\tilde{N}$ is another normalization constant.

If the assumption that a single analyte structure suffices is not true, the integration over q_{L} also needs to be included. I discuss this below in Section 3.3.4.

With the MD simulations for the alchemical free energy perturbation calculations, the free energy difference F_{binding} can be determined quite accurately for the given MD force field and parameters (not necessary "real" values as in NMR experiments). If they could be approximated with reasonable accuracy in a computationally efficient way, e.g. by an implicit solvation model if Γ_2 includes the coordinates of the solvent, Equation (2.34) can be used to approximate averages. How accurate this approximation needs to be to determine the alignment of analyte molecules is one of the central questions of this thesis and I tested the approach for different models of F_{binding}. When F_{binding} describes the electrostatic potential of the alignment medium and exclusion by hard spheres, this is exactly the P3D approach.

Using $F_{\mathrm{binding}} \approx V_{\mathrm{LJ}}$ alone is not feasible, as it leads to very deep energy minima, making the calculation of an average very difficult with MC integration. I also showed this with the MD simulations in vacuum, where the analytes stick very strongly to PBLG (see Section 4.1.5). My original motivation for the $\Delta\Delta G_{\mathrm{nonpolar}}$ model was to include $V_{\mathrm{LJ}} + \Delta\Delta G_{\mathrm{nonpolar}}$ to test the validity of the hard spheres approach used by P3D. However, as preliminary tests showed (described in the results), it still leads to deep minima likely caused by outliers of the model, which make the convergence of MC averages difficult and which do not appear in the MD simulations in any case.

Therefore, I decided to keep the steric exclusion by hard spheres and calculate the binding (free) energies between the analytes and PBLG using the following six different models for F_{binding}:

- "V_{Coulomb}"
- "P3D, unopt" (with $\epsilon_{\mathrm{in}} = 1$ and $r_{\mathrm{ext}} = 0$)
- "P3D, fit" (with $\epsilon_{\mathrm{in}} = 1$ and $r_{\mathrm{ext}} = 0.86\,\text{Å}$)
- "GB" (the ALPB-ΔHCT model with $\epsilon_{\mathrm{in}} = 1$ and $r_{\mathrm{ext}} = 0.39\,\text{Å}$)
- "Hard spheres" ($\Delta\Delta G_{\mathrm{total}} = 0$ without overlap)
- "Cylinder" ($\Delta\Delta G_{\mathrm{total}} = 0$ without overlap)

For all models except the "cylinder" model, the PBLG and analyte atoms are represented by hard spheres with radii $0.5\sigma_i$ with the LJ parameter σ_i for atom i. With the "cylinder" model, PBLG is instead replaced by a cylinder with radius 10 Å, as in the PALES model without charges [178] (see Section 2.6.2).

3.3.1. Structure of the PBLG molecule

Like for the P3D method [116], the integration is done with a long PBLG molecule aligned along the z-axis. To do this, I constructed a PBLG molecule of 140 monomers (ca. 21 nm long), aligned the eigenvector corresponding to the smallest eigenvalue of the inertia tensor along the z-axis and relaxed the structure with a MD simulation in chloroform. This simulation is set up like for the MD simulations for the calculation of RDCs, but without periodic boundary conditions. The simulation box has a similar shape as for the periodic simulations (i.e. forming a hexagonal lattice in the xy-plane), but with a length in the z-direction of $\sim$ 25.7 nm to cover the entire PBLG molecule. To keep the PBLG molecule aligned along the z-axis, I applied cylindrical flat-bottom restraints on the backbone atoms, keeping them within a radius of 4 Å from the z-axis.

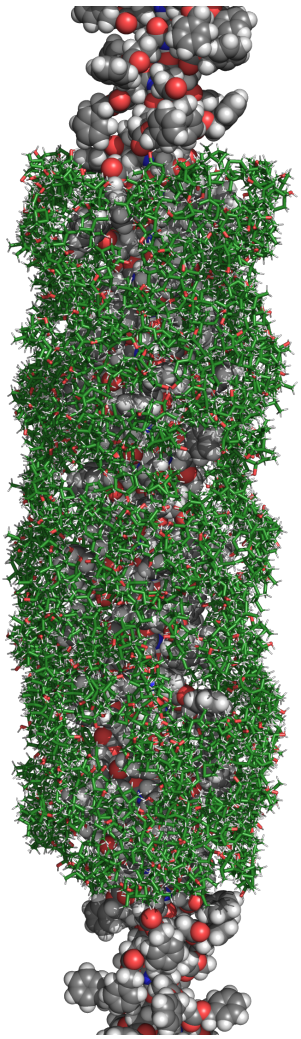


Figure 3.12.: Visualization of the integration volume used for the Monte Carlo calculations. Shown are 1000 random poses of (-)-IPC around PBLG.

For the calculation of RDCs, I then used snapshots of the PBLG chain from this simulation, using the central part of length 10 nm for the calculation. The snapshots are saved every

100 ps to allow the sidechains to reorganize. As described above, adding free space where the movement of the analyte molecule is isotropic increases the required amount of sampling dramatically. Therefore, I performed the integration over a cylindrical domain around the PBLG molecule. An example for (-)IPC is shown in Figure 3.12. In order not to introduce artificial alignment by the cylinder, the cylindrical domain used applied for the center of the analyte molecule, around which all rotations are sampled. This means that without the alignment medium, the rotations are still isotropic. The cylinder radius is determined by the sum of the maximal distance of the PBLG atoms from the z-axis, the maximum distance of the atoms in the analyte molecules from the center of mass, and a constant buffer distance of 2 Å.

3.3.2. Analyte molecules used in the MC simulations

For the MC calculations, I chose the 4 pairs of enantiomers of rigid molecules shown in Figure 3.2 (IPC, quinuclidinol, borneol and camphor), which I also used for the MD simulations. Additionally, I did the calculations with both enantiomers of β-pinene, (-)-caulamidine A and (-)-strychnine, see Figure 3.13. Like for the other molecules, the ${}^1D_{\mathrm{C-H}}$ couplings and the experimental values are listed in Tables A.5 and A.6. The measurements for β-pinene were reported in Hansmann, Larem, and Thiele [60], for (-)-caulamidine A in Milanowski et al. [107] (together with its structure) and for (-)-strychnine in Liu et al. [93]. Again, the CH_3 groups are treated like in the MD simulations, which is especially relevant as there is not as much sampling of the CH_3 rotation.

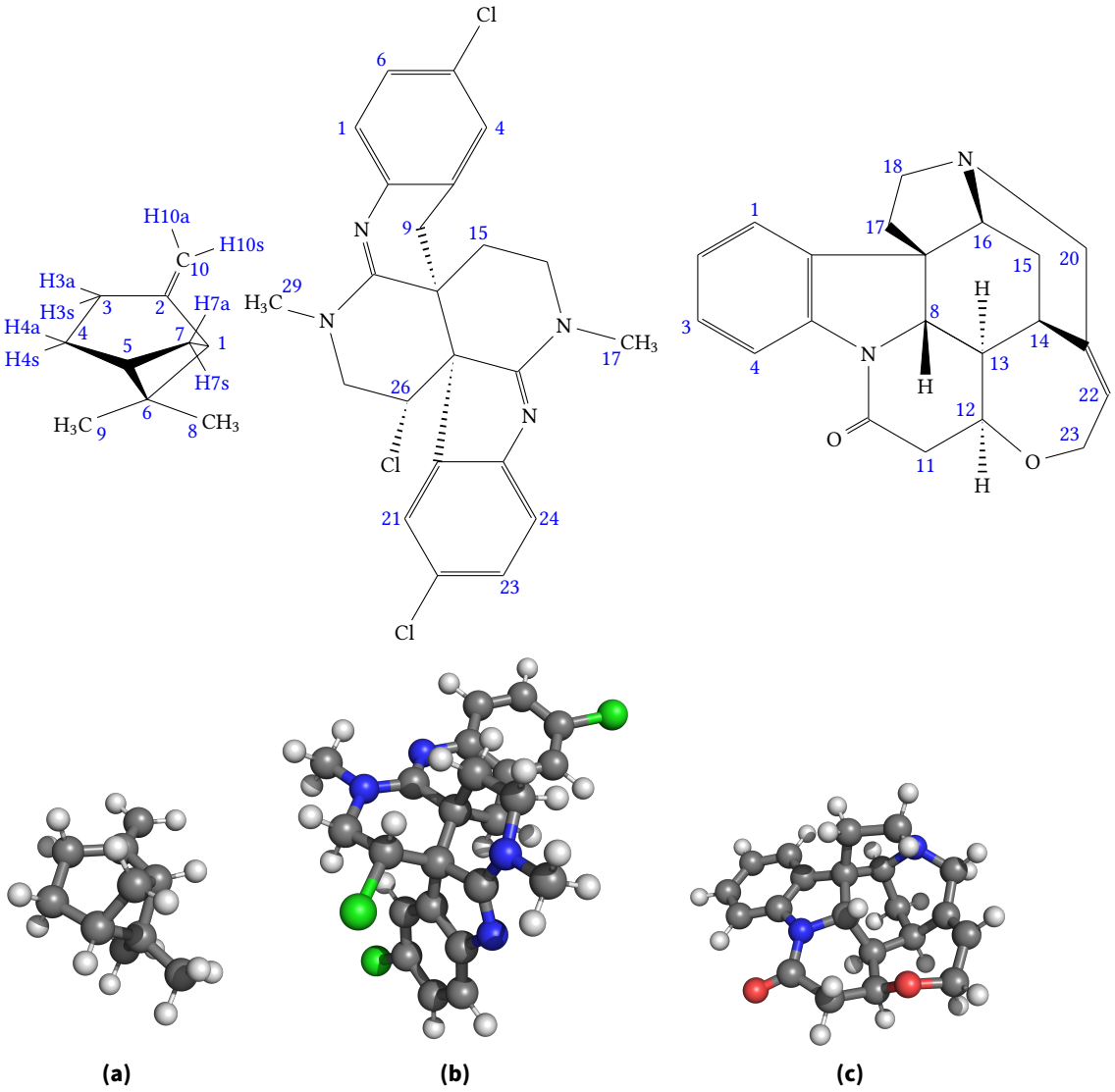


Figure 3.13.: Molecule structures with the used atom numbering for **(a)** (+)-β-pinene, **(b)** (-)-caulamidine A and **(c)** (-)-strychnine.

3.3.3. Application of population annealing

In the presence of strong binding energy minima, such as with specific interactions between the analyte molecule and the alignment medium like hydrogen bonds, the canonical ensemble distribution is very different from a uniform distribution. This makes naive Monte Carlo integration or quadrature rules (as used in the original PALES and P3D methods [116, 178]) very inefficient, as unimportant samples with high energies are treated as being equally important as samples with low energy and therefore large Boltzmann factors. In my case of calculating RDCs, this means that much of the samples are wasted on poses far away from the alignment medium, where the energy is near zero and the (unnormalized) Boltzmann factors are near 1.

The equilibrium distribution of a MCMC simulation matches the target distribution exactly. For a distribution with strong maxima, the simulation can be stuck at a maximum and thus reaching the equilibrium distribution can take a very long time. Here, a parallel approach is required, where the individual are weighted accordingly. In the case of a canonical ensemble distribution, two options are parallel tempering and population annealing. The latter is however much more suited for a very large number of parallel simulations [170], as for each inverse temperature β_i, all MCMC simulations are independent and the only serial step is the resampling, which does not involve the simulated system at all and only consists of reorganizing existing samples.

For the resampling, I used the "systematic resampling" approach with an adaptive schedule described in Gessert, Janke, and Weigel [58], see Section 2.6.1. A direct application of minimizing $\Delta\rho_t/\Delta\beta$ did not prove practical however, because it is both minimal for small values of $\Delta\beta$ and no longer sufficiently smooth for very small values, at least for the chosen population size. I therefore determine the minimum value of $\Delta\rho_t/\Delta\beta$ for $\Delta\beta \geq 50$ a.u. and find the maximum $\Delta\beta$ where $\Delta\rho_t/\Delta\beta$ does not exceed 1.5 times the minimal values instead. Naturally, $\Delta\beta$ is chosen in the last step such that the final β_i corresponds exactly to the target temperature (which in this case is 300 K).

Another significant benefit of a method like population annealing over calculating weighted averages is that like the MD simulations, it results in (correlated) samples from the distribution of the analyte orientations (and structures), which can then be analyzed afterwards to calculate different properties of the distribution. I use this for the same kinds of analysis like for the MD simulations, such as the distances to PBLG or the hydrogen bonds.

3.3.4. Analyte molecule flexibility

The available molecules I chose to test the MC calculations with (shown below in Section 3.3.2) are all almost rigid. However, due to the flexibility of the hydrogen atom attached to the O – H hydrogen bond donors and the rotating CH_3 groups, it can still be necessary to calculate an average over different molecule structures.

As the implementation of the models is based on rigid molecules and the sampling of molecule structures in the presence of explicit solvent is more challenging in MC calculations, I instead used an ensemble of precomputed analyte structures from MD simulations in chloroform. In the MC calculations, these structures are then used as separate states with no additional weighting required, apart from the model for the binding (free) energy of a given analyte structure. This means approximating the state space Ω_L by a set of fixed analyte structures $q_\mathrm{L}^i \sim P_\mathrm{L}$, where P_L is the (canonical) probability distribution of the states $q_\mathrm{L} \in \Omega_\mathrm{L}$.

The alignment medium is still assumed to be large enough for a single structure $\tilde{q}_{\mathrm{AM}}$ to suffice. Equation (3.17) can thus be approximated numerically by

$$\begin{aligned}\langle A\rangle &\approx \frac{1}{N_1}\int_{\Gamma_{\mathrm{L}}} e^{-\beta F_{\mathrm{L}}(q_{\mathrm{L}})}\int_{\Gamma_{\mathrm{COM}}} A(q_1)e^{-\beta F_{\mathrm{binding}}(q_{\mathrm{COM}},\tilde{q}_{\mathrm{AM}},q_{\mathrm{L}})}\,dq_{\mathrm{COM}}\,dq_{\mathrm{L}}\\ &\approx \frac{1}{N_{\mathrm{L}}}\sum_{i=1}^{N_{\mathrm{L}}}\frac{1}{N_{\mathrm{COM}}(q_{\mathrm{L}}^i)}\int_{\Gamma_{\mathrm{COM}}} A(q_1)e^{-\beta F_{\mathrm{binding}}(q_{\mathrm{COM}},\tilde{q}_{\mathrm{AM}},q_{\mathrm{L}}^i)}\,dq_{\mathrm{COM}}\end{aligned} \tag{3.19}$$

with other normalizations $N_{\mathrm{COM}}(q_{\mathrm{L}}^i)$. To reduce the computational cost, the sum over q_{L}^i is included together with the integration over Γ_{COM} in the population annealing procedure. In principle, the same approach is still possible for Ω_{AM}, but as I show in the results, the chosen size of the PBLG chain is large enough for the approximation by a single chain to be sufficient.

The MD simulations for the generation of the analyte structures are set up in the same way as the PBLG simulations, but with a cubic box with periodic boundary conditions in all directions and a box size determined by the analyte molecule size with a buffer of 1.5 nm, resulting in side lengths of $\sim$ 3.4 nm to 4 nm. To generate the analyte structures, I simulate independent chains of 1 ns length, each equilibrated with a different random seed using the same procedure as for the MD simulations with PBLG for calculating RDCs. This is sufficient, as all tested analyte molecules do not have conformers separated by a large energy barrier. The structures are then collected every 10 ps from as many chains as necessary.

For the MCMC steps in the population annealing method, sampling from a proposal distribution is required. Similar like for the proposals for the position and rotation, I used a normal distribution for the proposal distribution of the structure index, wrapped periodically at the ends for simplicity, making it symmetric. For the initial distribution at $\beta = 0$, the structures are uniformly sampled. To increase the acceptance rate of the proposals in the Metropolis algorithm, the proposal distribution should propose similar structures, which means that structures with nearby index should be similar. Therefore, they need to be ordered in a way to minimize the change between neighboring structures. I achieved this by first moving the center of mass of each structure to the origin and then calculating the pairwise symmetry-corrected RMSD [37] between all analyte structures. Finding a reordering which minimizes the sum of the RMSDs of adjacent structure (with the last structure neighboring the first structure) is exactly the metric traveling salesman problem, because aligned RMSDs form a metric space [152]. For this, many approximate methods exist, but the quality of the solution does not need to be very good, just good enough for some Metropolis proposals to be accepted to reduce the bias of the final population annealing distribution. Afterwards, I aligned adjacent structures to minimize the same RMSD. As the hydrogen bonds to PBLG are very important, the oxygen and hydrogen atoms of the O – H hydrogen bond donors are weighted by a factor of 5 and 10 for all RMSD calculations, respectively.

3.3.5. Adjustments for calculating RDCs with the "hard spheres" and "cylinder" models

When calculating the RDCs based on steric exclusion by treating the atoms as hard spheres with no other interaction, the problem is that the alignment is very weak for spherical-like molecules, which would require a lot more samples to obtain a reasonable relative uncertainty (e.g. at least of the order of $\sim 10^8$ for quinuclidinol in my case). However, by observing that for analyte positions where there is no overlap between the molecules regardless of the analyte rotation, the orientation is exactly isotropic and the contribution to the RDCs is 0. This scales the RDCs by a constant factor > 1 compared to the full average, decreasing the relative uncertainty considerably.

Likewise, for the "cylinder" model, the squared distance from the z-axis is uniformly sampled between the smallest distance where a rotation exists for which the molecule lies entirely outside the cylinder and the largest distance for which a rotation exists where the molecule overlaps with the cylinder.

3.3.6. Implementation

Like for the other parts, I automated the MC calculations and their preparation with a `Snakemake` workflow. It is available under `https://gitlab.kit.edu/david.elsing/RDCMC`. The force field parameters are obtained in the same manner as before, taking care that they are exactly the same as for the MD simulations for the molecules tested with both methods. Also, the model contributions, such as the Born radii or the different grids, are calculated using the numerical parameters explained above for the implicit solvation models.

The actual population annealing procedure consists of generating the population at $\beta = 0$, followed by alternating resampling and MCMC steps. For the resampling steps, the existing poses from the current population are only reordered using the already calculated binding (free) energies, such that their computational cost is negligible compared to the MCMC steps. On the other hand, the MCMC calculations are completely independent and thus directly parallelizable. As the evaluation time of the binding (free) energies for the different models varies considerably for different poses, the parallelization should be implemented with a work stealing approach [20]. Compared to the MD simulations, the necessary disk space to store the results is much lower, because the PBLG structure is kept rigid. Also, for each pose of the analytes, only five things need to be stored: The index of the precalculated analyte structure, the position and rotation around the center of mass, the family index of the PA method and the binding (free) energy.

4. Results

4.1. RDCs calculated with Molecular Dynamics Simulations

In the following, the results from the MD simulations using the computational setup described in Section 3.1.2 of the alignment of the analyte molecules by PBLG in chloroform are described. In particular, I investigated the interactions between the analyte molecules with PBLG, which includes the distance distribution from PBLG, the contacts with the different parts of the PBLG sidechains and the hydrogen bonds between the analyte molecules and PBLG. Most importantly, I calculated the RDCs in the analyte from the MD trajectories and compared them to experimental values, including the differences between enantiomers. Using them, I show the applicability of the MD simulations for the determination of the absolute configuration.

4.1.1. Interactions between chiral analytes and PBLG

To see where the alignment of the analyte molecules occurs, I calculated the distance of their center of mass from the central axis of PBLG. The distribution of the COM distances is shown in Figure 4.1, where clear differences between the molecules are visible: All molecules with hydrogen bond donor exhibit a large peak between 5 Å to 10 Å. As explained below (see Figures 4.4 and 4.5), this is because hydrogen bonds with PBLG are formed, allowing the analyte molecules to stay in between the PBLG sidechains. For camphor, which does not have a hydrogen bond donor, this is not the case: There is still a small peak near 7.5 Å, which can be explained by the other interactions with PBLG, but the most occurring distances are further away than $\geq$ 15 Å, showing that there is no strong binding to PBLG. For the spherical-like molecules IPC, quinuclidinol, borneol and camphor, there is no significant difference between the enantiomers, while there are small, but noticeable differences at the first peak (i.e. the poses with hydrogen bonds) for the larger and more flexible molecules.

In Figure A.5, the contribution to the RDC averages (described below) of the different COM distances is shown. Here, the behavior for camphor is clearly different from the other molecules, as most of the alignment occurs much further from the PBLG center than for the other molecules with hydrogen bond donor. For IPC, quinuclidinol and borneol, the contributions are distributed very similarly, while it differs more strongly for HMIP, 1-phenylethanol and ibuprofen due to their different shape (and flexibility of the latter two). However, all molecules with hydrogen bond donor show two distinct peaks. As

the poses with hydrogen bonds explain most of the alignment (which I explain below), a possible explanation is that the closer peak corresponds to hydrogen bonds with the analyte molecule being far inside the PBLG sidechains, while the outer peak corresponds to the C = O acceptor pointing away from the PBLG center (see Figure 2.3).

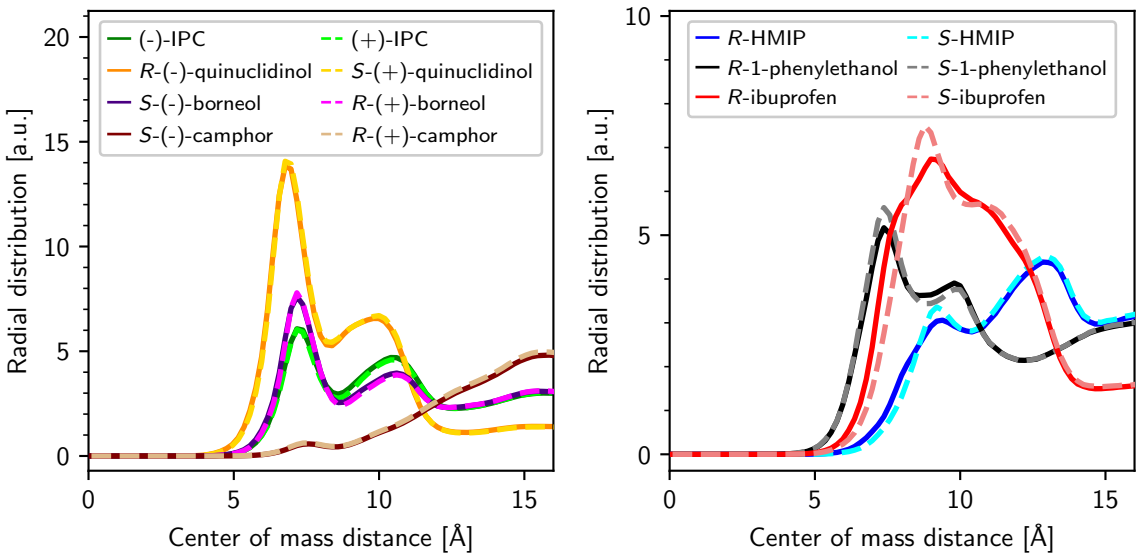


Figure 4.1.: Distribution of the center of mass distance of the analyte molecules to the PBLG center.

For the contact types defined in Figure 3.5, I investigated the interactions between the analyte molecules and PBLG by determining which parts of the PBLG molecule the analyte molecules are in contact with. The results are shown in Figure 4.2, where for each contact type, the fraction of simulation time such a contact is formed for at least one, two, three or four different sidechains simultaneously. Also shown is the fraction of simulation time for which there is any interaction with a sidechain, which is defined in the same manner as for the "hydrogens of ring" and "sidechain without ring" contacts.

Camphor does not form a hydrogen bond with PBLG and therefore has the lowest probability for the interactions with PBLG, but there are still contacts ~ 50 % of the time. Here, the contacts with more than two different sidechains at once are less important, while the other molecules interact with at least three sidechains more than 30 % of the time.

As the hydrogen atoms of the phenyl ring in PBLG are the most exposed to the analyte molecules, there are "hydrogens of ring" contacts almost the entire time there is a contact at all. Depending on the strength of the hydrogen bond in the MD simulation, the "sidechain without ring" contacts are present in a large fraction of all contacts, but are more frequent than the "ring side" contacts even for camphor. The latter appear to only play a relatively minor role, not exceeding 15 % of the simulation time in all cases.

Remarkably, the fractions of simulation time with contacts are almost identical between enantiomers for all molecules. This confirms that the difference of the RDCs between enantiomers for the tested molecules does not come from a significant change in the interactions with PBLG, but because they are oriented differently in similar poses.

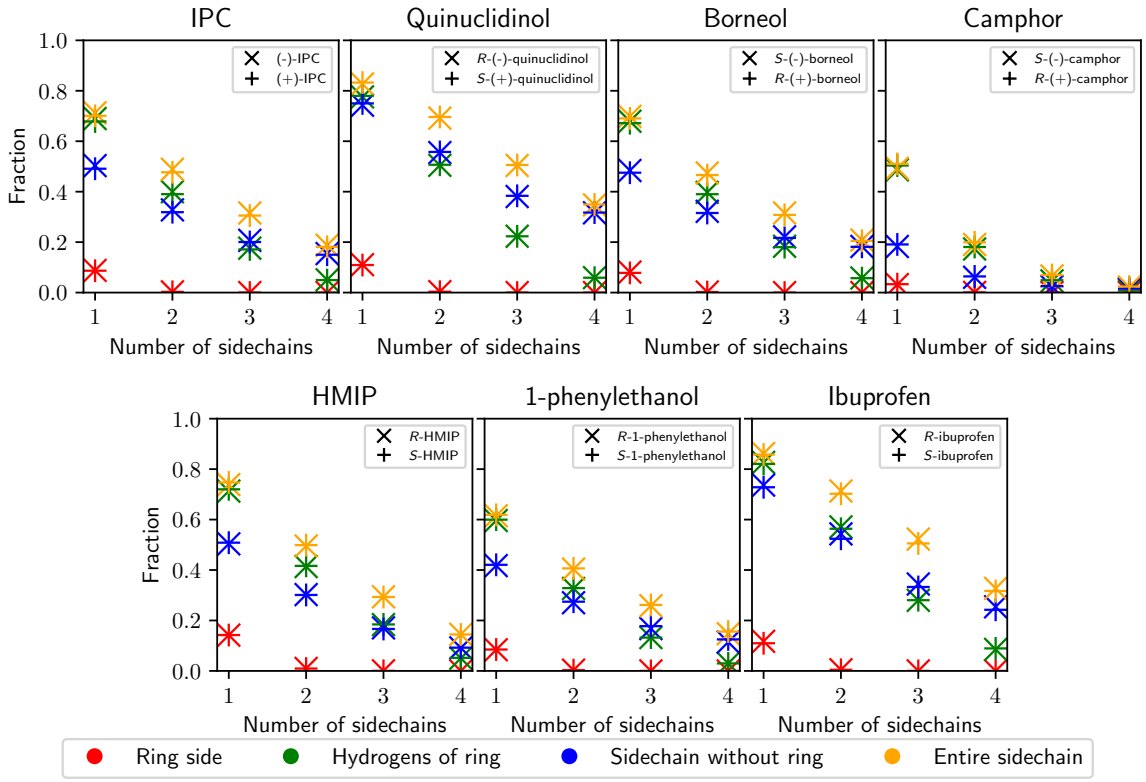


Figure 4.2.: Fraction of the simulation time with contact between the analyte molecule with different parts of the PBLG sidechains for the 7 pairs of enantiomers. The "number of sidechains" indicates the number of different repeat units of PBLG, for which the analyte molecule is simultaneously in contact with the respective part. The kinds of contacts are shown in Figure 3.5.

The RDCs and the alignment tensor (described below) only give limited information about the average orientation of the couplings or of the entire molecule. To further understand the alignment, I also calculated the distribution of $\cos(\theta)$ for the molecule axes defined in Figure A.6, where θ is the angle of the respective axis to the z-axis in the simulation. This shows how the orientation of the molecules differs from isotropic movement, where $\cos(\theta)$ would be uniformly distributed. The results obtained from the MD simulations with one analyte molecule are shown in Figure 4.3. In order to keep the orientation of the axes defined as eigenvectors of the inertia consistent, the axes of the reference structure are rotated instead of recalculating the inertia tensor for each pose, like for the calculation of the alignment tensor with Equation (2.16). There is a clear contrast between the molecules with hydrogen bond donor and camphor: For the latter, the distribution is almost isotropic and nearly unimodal. In contrast, for the molecules with O – H donor, the behavior is much more complex and anisotropic. For HMIP, which has a N – H donor, the distribution is also very anisotropic, but has a smoother shape. This is probably because it is rather flat and rigid and the N – H hydrogen bond donor is very directed in the same plane, restricting the possible orientations while a hydrogen bond with PBLG is formed. For the molecules with O – H hydrogen bond donor except 1-phenylethanol, the distributions are very different between the enantiomers. For HMIP and 1-phenylethanol, the differences are smaller, but still significant. For camphor, the differences are very small and only visible near

$\cos(\theta) = \pm 1$, but those might be numerical artifacts due to the limited precision with which the positions in the trajectories are stored (0.01 Å).

Due to the flat shape of HMIP, its orientation between the PBLG sidechains when a hydrogen bond is form is limited, while IPC, quinuclidinol and borneol are rather spherical and can rotate easier. For 1-phenylethanol and ibuprofen, the rotatability of the phenyl ring is an additional complication, making the accurate description in the simulation more difficult.

For comparison, the $\cos(\theta_{ij})$ distribution of the used couplings (which are not averaged) are shown in Figure A.7 for HMIP. Here, θ_{ij} is the angle of the $\vec{r}_{ij}$ to the static magnetic field as in Equation (2.14). Like for the axes in Figure 4.3, there is a noticeable difference between the enantiomers for some couplings, depending on the orientation of $\vec{r}_{ij}$ in the molecule.

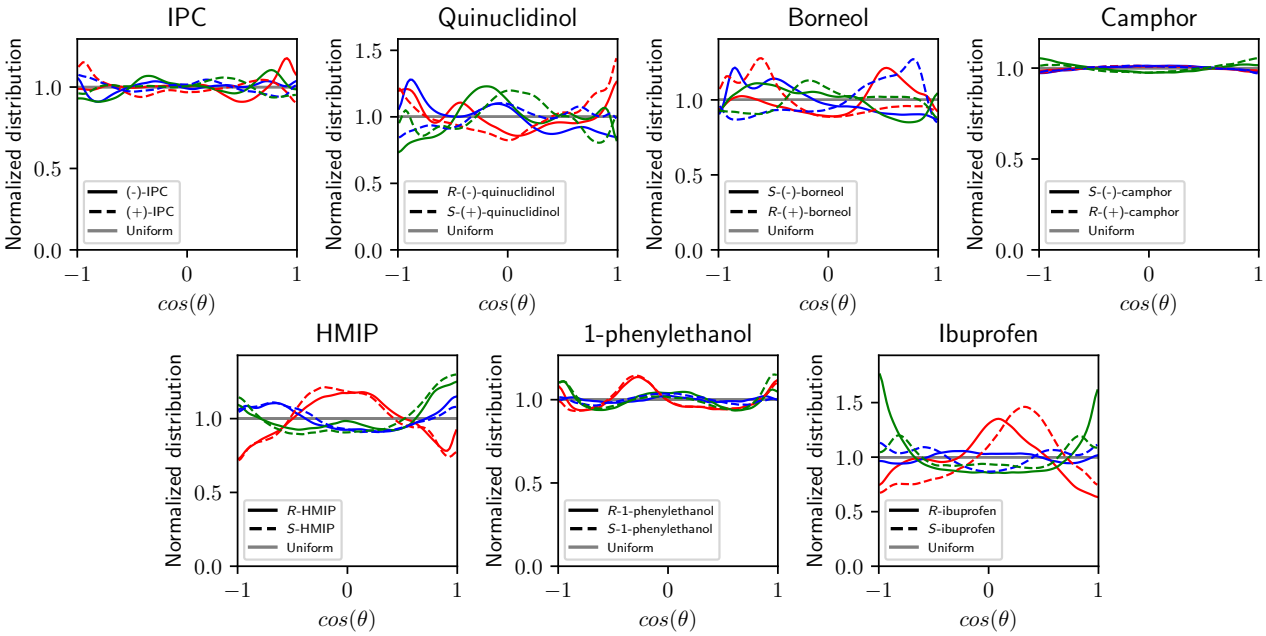


Figure 4.3.: Distribution of $\cos(\theta)$ (where θ is the angle to the z-axis) for the molecule axes defined in Figure A.6 of the 7 pairs of enantiomers used in the MD simulations. The colors of the axes are the same as in Figure A.6. In gray, a uniform distribution is shown, which is the distribution for isotropic movement.

Because the PBLG molecule is oriented along the z-axis and is on average symmetric (but still chiral!) when rotated around the z-axis, the distribution of the angle in the xy-plane should be uniformly distributed. Indeed, this is confirmed in Figure A.8, where the distribution is approximately uniform for all molecules.

4.1.2. Hydrogen bonds

In addition to the contacts between the analyte molecules and PBLG, I investigated the hydrogen bonds formed in the simulation. For determining whether a pose has a hydrogen

bond to PBLG, I chose a rather broad definition, with a minimal X – H ⋯ O angle (where X is the hydrogen bond donor atom) of 150° and a maximal distance of 3 Å. The reason for this choice is that the hydrogen bonds are responsible for significant (free) energy minima and a significant fraction of the poses do have a hydrogen bond to PBLG. Thus, the contribution of similar poses is expected to be rather small, while the broad definition does not remove poses which can still be counted as possessing a hydrogen bond.

In Figure 4.4, the fraction of simulation time during which a hydrogen bond is formed and the average lifetime of the hydrogen bonds is shown for both the C – O – C and C = O acceptors in PBLG (see Figure 2.3). The lifetimes are approximated by counting the number of consecutive frames where a hydrogen bond is formed with the above definition. It should be noted that the temporal resolution is the time between snapshots in the saved MD trajectory, i.e. 10 ps, which causes the average lifetimes to be biased to larger values. Clearly, the C – O – C acceptor is entirely irrelevant, because the hydrogen bonds with it as acceptor occur during a negligible fraction of the simulation time. Their average lifetimes is also only near the temporal resolution of the snapshots. The C = O acceptor is very important on the other hand: There is a hydrogen bond more than 20 % of the time for all molecules with hydrogen bond donor. Like for the other contacts with the PBLG sidechains, the occurrence and lifetimes of the hydrogen bonds are nearly identical for both enantiomers. They are also very similar for borneol and IPC, probably because the molecular environment of the O – H donor is similar for both cases, with a nearby CH_3 group. Quinuclidinol and ibuprofen lack such a nearby CH_3 group, making the hydrogen bond donor more exposed (see Figures 3.2 and 3.3) and thus the occurrence of hydrogen bonds higher. For HMIP, the hydrogen bonds also occur less often, which can be explained by the lower electronegativity of nitrogen compared to oxygen, as well as the lower flexibility of the N-H donor compared to the O – H donors.

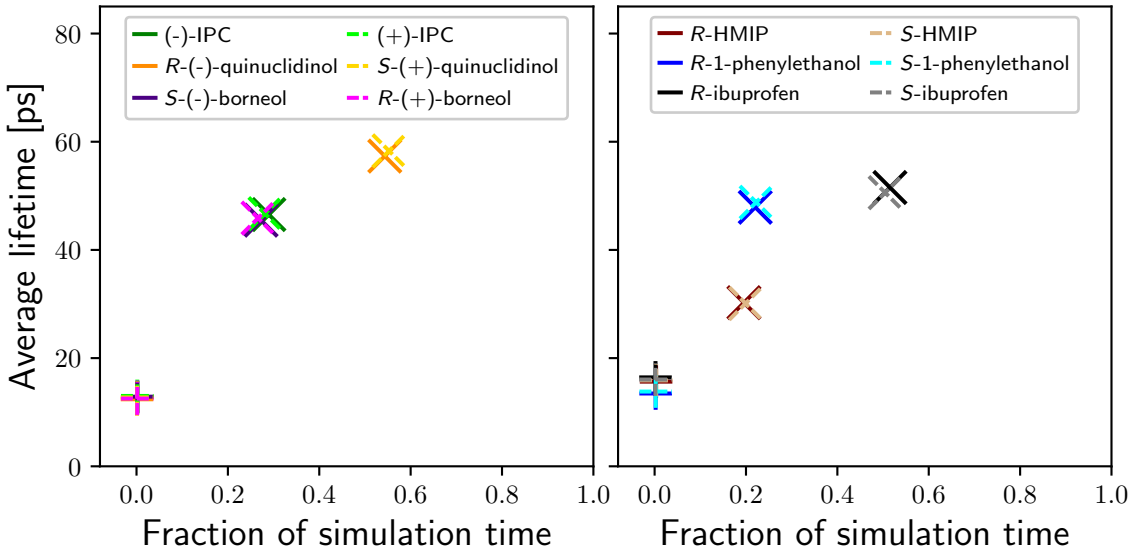


Figure 4.4.: Fraction of simulation time with hydrogen bond to the C = O (marked by ×) and C – O – C (marked by +) acceptors in PBLG and the average lifetime of the hydrogen bonds. Note that the MD snapshots were saved every 10 ps, such that smaller lifetimes are not resolved.

To better understand the nature of the hydrogen bonds in the MD simulations (where they are only described by point charges), I calculated the distribution of the X – H ⋯ O distances and angles. The results are shown in Figure 4.5. Note that a uniform distribution would not appear uniform in the plot, but rather proportional to $r^2 \sin(\theta)$. Most notable are the significantly stronger hydrogen bonds for quinuclidinol and ibuprofen as already seen in Figure 4.4. The shape of the distributions is very similar for all molecules with O – H hydrogen bond donor, while it is shifted to larger distances and lower angles for HMIP. For the molecules with O – H hydrogen bond donor, the hydrogen bonds are least populated for 1-phenylethanol. Here, the O – H group is directly next to a phenyl ring and a CH_3 group, limiting the ways it fits between the PBLG sidechains. Nonetheless, as Figure 4.4 shows, the hydrogen bonds still make up a significant part of the simulation time in all cases.

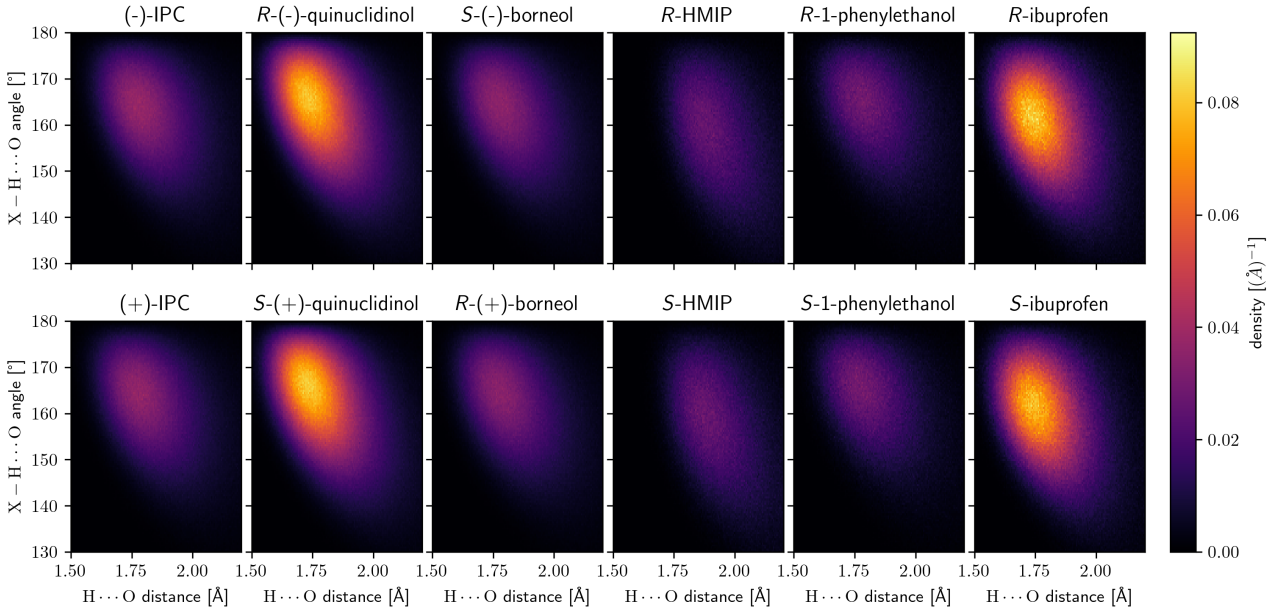


Figure 4.5.: Distribution of the O – H ⋯ O hydrogen bond distances and angles in the MD simulations for the 6 pairs of enantiomers with hydrogen bond donor.

4.1.3. Residual dipolar couplings and alignment tensors

To verify that the average converged sufficiently, the statistical uncertainty determined by Equation (2.52) relative to the maximum RDC value (determined from the full average) is shown for different subsets of all simulations in Figure 3.4.

The calculated RDCs for all 14 molecules are shown in Figures 4.6 and 4.7. To compare them, the calculated RDCs $\vec{D}^{\text{sim}}$ are scaled to match the experimental values $\vec{D}^{\text{exp}}$ by $\|\vec{D}^{\text{exp}}\|/\|\vec{D}^{\text{sim}}\|$. This ensures a positive scaling (which leaves the cosine similarity intact), while not giving misleading results if the agreement is low. For HMIP, the calculated RDCs agree almost perfectly with experiment, while for camphor and (+)-IPC, the agreement is still good for most of the couplings. With the other molecules however, the RDCs differ

strongly from the experimental data, although there are clear differences between the enantiomers.

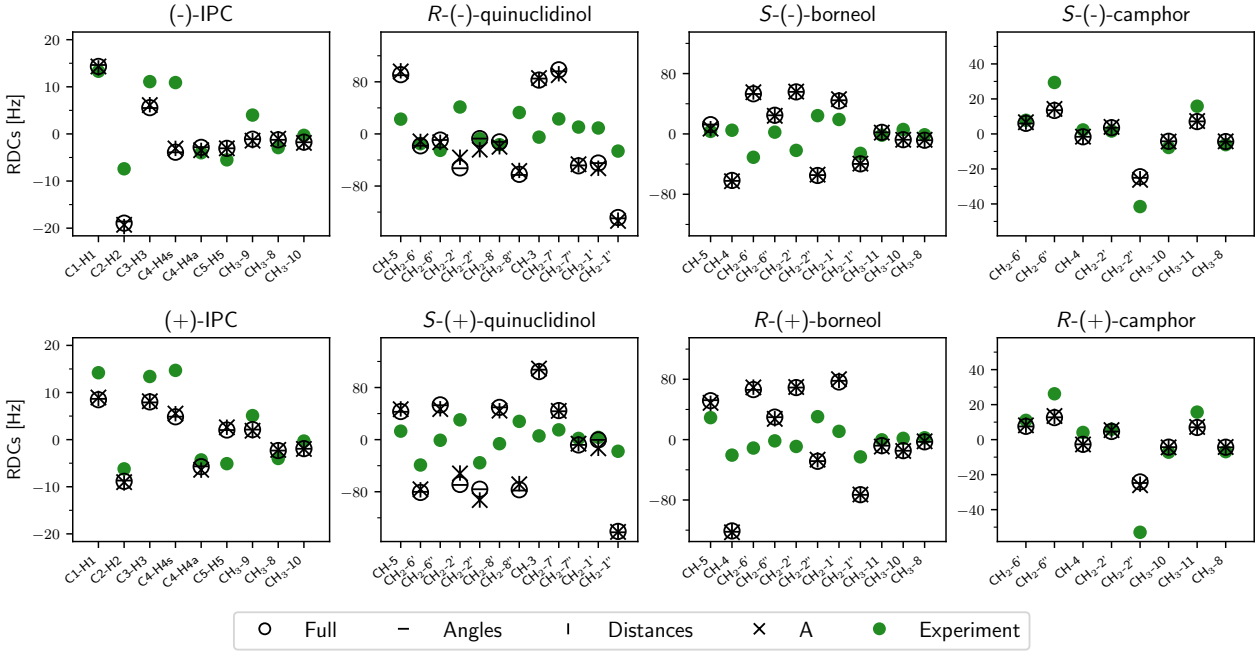


Figure 4.6.: RDCs ($2D_{ij}$) calculated from MD for both enantiomers of IPC, quinuclidinol, borneol and camphor. The different ways the RDCs were calculated ("Full", "Angles", "Distances" and "A") are described in this section (Section 4.1.3). In green, the experimental RDCs are shown, to which the RDCs from MD were scaled (with the same scaling for all couplings, see Section 3.1.4) by the ratio of the sum of the squared RDCs, where the scaling for each enantiomer pair was averaged.

Besides the full average ("Full") calculated with Equation (2.14), I also calculated the averages using the DFT-optimized reference structure for the rigid molecules to calculate the bond distances, while the angles θ_{ij} are calculated directly (denoted as "Angles" in the figures). The inverse case of using the bond distances from each simulated pose is denoted as "Distances", where the angles θ_{ij} are determined by fitting a rotation matrix which rotates the reference structure to each pose, from which θ_{ij} is then calculated. This is equivalent to calculating the alignment tensor with Equation (2.16) and with it the angles using Equation (2.15). Therefore, the case of using both bond distances and angles θ_{ij} from the reference structure is the same as calculating the RDCs from the calculated alignment tensor, which is denoted as "A" in the figures. They are not shown for ibuprofen and phenylethanol due to the rotatable phenyl ring, where a single structure is not enough to describe the RDCs with an alignment tensor. In all other cases, the "Full" and "Angles" values agree almost perfectly (as well as "Distances" and "A"), because the bond length fluctuations are not important for ${}^1D_{\mathrm{C-H}}$. As the molecules are almost rigid, the differences between "Full" compared to "Distances" and "A" are not large either, but significant is some cases, with the largest difference being the CH_2-2' and CH_2-2" RDCs in quinuclidinol.

This can also be seen be looking at the cosine similarities between the experimental and calculated RDCs. To confirm that the convergence is sufficient for a comparison with

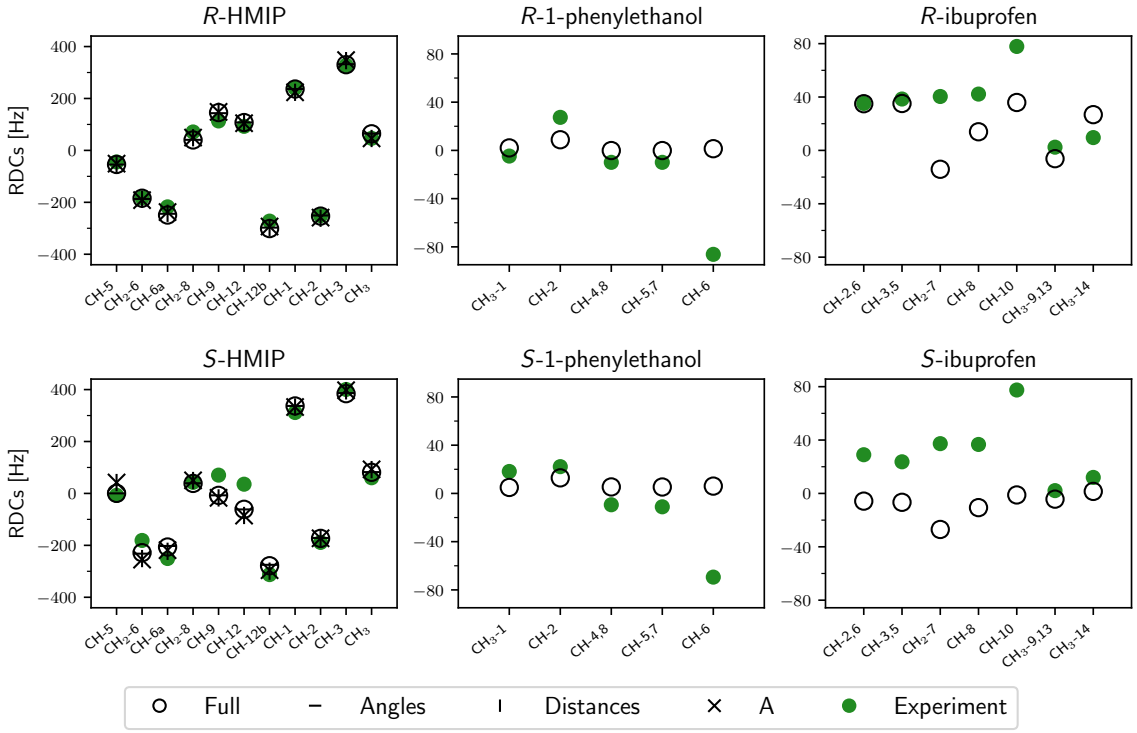


Figure 4.7.: RDCs ($2D_{ij}$) calculated with MD for both enantiomers of HMIP, 1-phenylethanol and ibuprofen. See Figure 4.6 for an explanation.

experiment, I additionally calculated the probability distribution of the cosine similarity between the RDCs, taking into account the estimated covariance matrix of the MD average and the experimental uncertainties. The results are shown in Figures 4.8 and 4.9. Additionally, the cosine similarities between the averages using only the poses with hydrogen bond (using the definition described above) and the experimental RDCs are shown in red. In all cases, they do not differ strongly from the full average.

In Figures 4.10 and 4.11, the calculated alignment tensors from MD, the ones with the average only over the poses with hydrogen bonds to PBLG and the fit to the experimental RDCs are shown. Like for the RDCs, the alignment tensors for HMIP agree very well. Even though the RDCs have good agreement for camphor, the alignment tensors are more distinct compared to experiment, probably because there are not many experimental RDCs available, which means that the alignment tensor fit is not unambiguous. As seen before, for all molecule with hydrogen bond donor, the average over only the poses with hydrogen bond to PBLG gives very similar results to the full average. Notable is also the difference between enantiomers both for the experimental and calculated alignment tensors for IPC, quinuclidinol and borneol. For camphor, the calculated alignment tensors are almost identical and the experimental ones are still similar, showing that there is almost no enantiodiscrimination. Likely due to the flat shape of HMIP, the differences between enantiomers are much smaller than for IPC, quinuclidinol and borneol, but there are still differences present both in experiment and simulation, which allow the distinction of enantiomers.

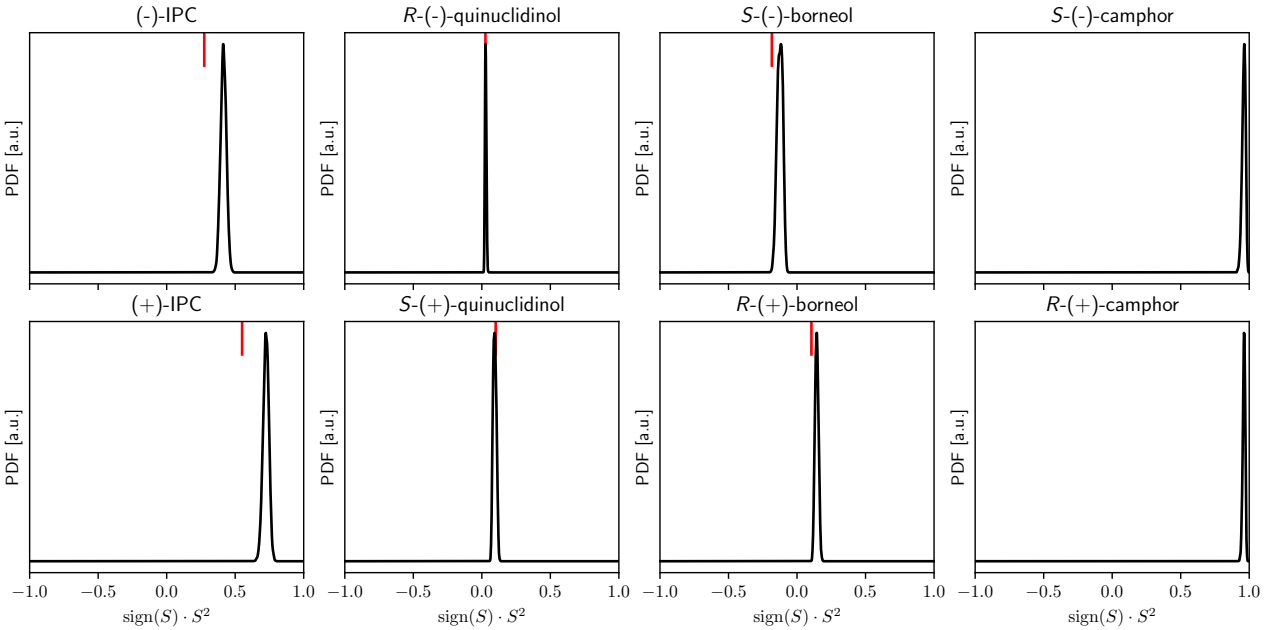


Figure 4.8.: Distribution of the cosine similarity between the experimental RDCs and the MD average for both enantiomer of IPC, quinuclidinol, borneol and camphor. For the experimental RDCs, a uniform distribution is assumed, while for the MD average, the distribution is a multivariate normal distribution with the estimated covariance matrix of the average. In red, the cosine similarity using only poses with hydrogen bonds to the C = O acceptor is shown.

Looking back at Figure 4.3, the distribution of the molecules with O – H hydrogen bond donors appears more complex than for HMIP or camphor, where the MD averages agree very well with the experimental values. This might explain why the agreement with experiment is harder to reach for these molecules. Using the MC calculations, I studied this in more detail, which is shown below in Section 4.3.4.

In Table A.7, the averages *excluding* the poses with hydrogen bond to PBLG using the abovementioned definition are compared. Surprisingly, they agree almost perfectly with the full MD average in all cases. Also, for all molecules except 1-phenylethanol, they are also almost identical to the average using only the poses *with* hydrogen bond. As poses are counted as having a hydrogen bond for O ⋯ H distances ≤ 3 Å and X – H ⋯ O angles ≥ 150°, the likely explanation is that the alignment of the poses with angles ≤ 150° (at any distance) is very similar to the poses with hydrogen bonds. This would mean that while the poses with hydrogen bond explain most of the alignment, there are additionally many similar poses where the hydrogen atom has moved somewhat, which does not strongly influence the orientation of the entire molecule.

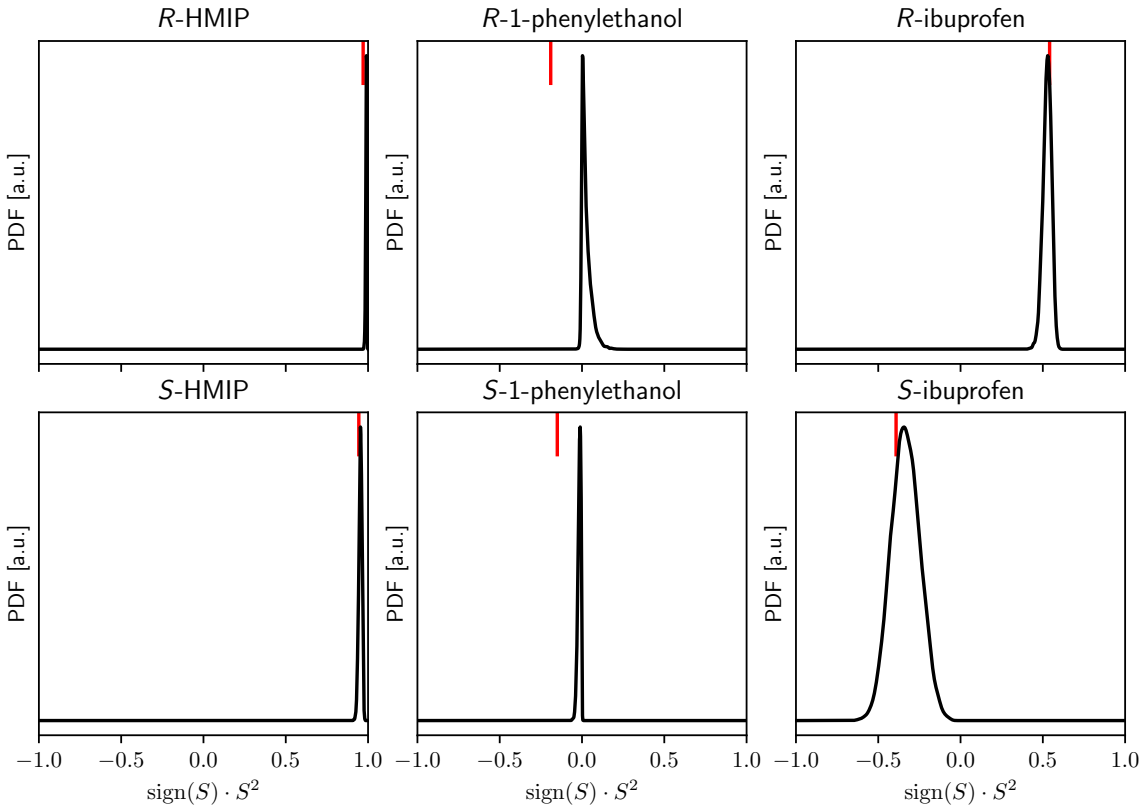


Figure 4.9.: Distribution of the cosine similarity between the experimental RDCs and the MD average for both enantiomer of HMIP, 1-phenylethanol and ibuprofen. See Figure 4.8 for an explanation.

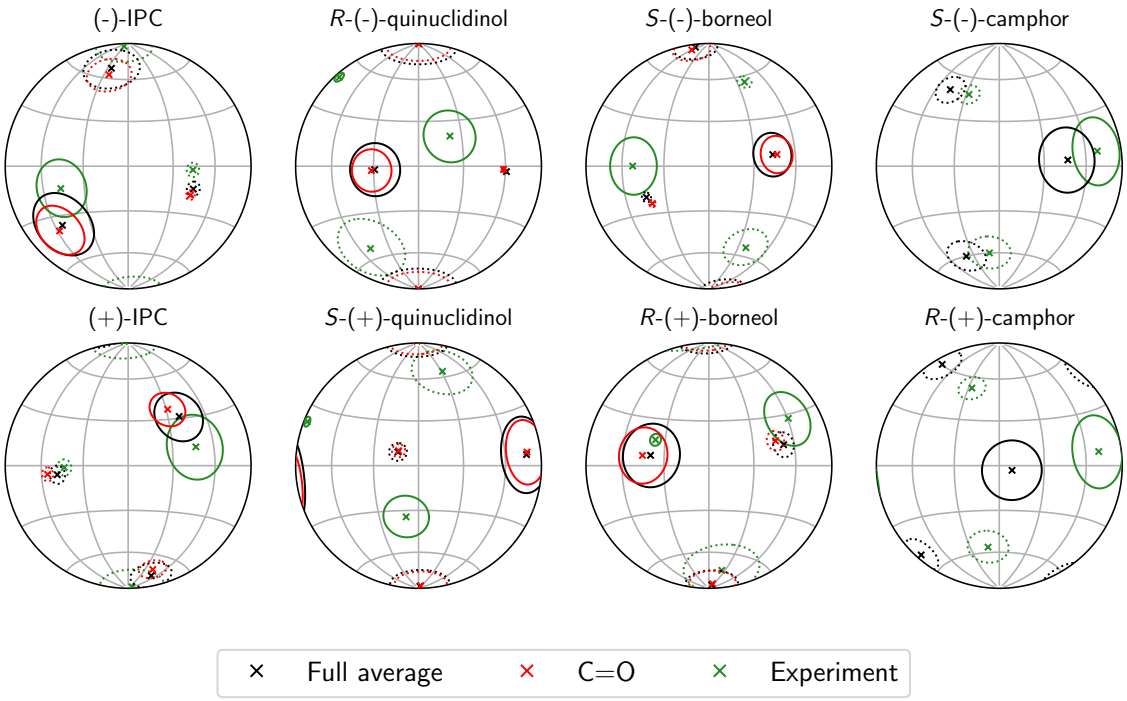


Figure 4.10.: Alignment tensors of both enantiomers of IPC, quinuclidinol, borneol and camphor represented by their eigenvectors and -values on a hemisphere. The eigenvectors are indicated by the × markers, while their corresponding eigenvalues are represented by the circles around them, whose area is proportional to the absolute value of the eigenvalue. Here, dashed lines indicate a negative sign and solid a positive sign of the eigenvalue. In black, the full average is shown and in red, the average over only the poses with hydrogen bond to the C = O acceptor in PBLG. The fitted experimental values are shown in green.

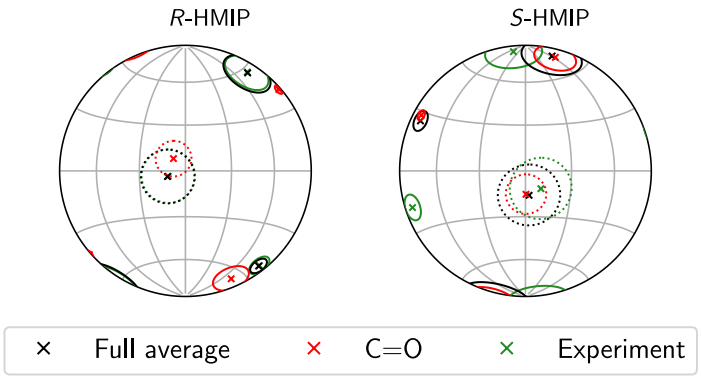


Figure 4.11.: Alignment tensors of both enantiomers of HMIP represented by their eigenvectors and -values on a hemisphere. Refer to Figure 4.8 for an explanation.

4.1.4. Comparison of RDCs between enantiomers

To see how much of the difference between the enantiomers comes from the hydrogen bonds with PBLG and how the enantiomers differ in simulation and experiment, I compared the experimental and calculated (the full average and only poses with hydrogen bond) RDCs for both enantiomers as shown in Figure 4.12. In all cases, the difference between enantiomers is similar for the full MD average and the average using only the poses with hydrogen bond. This means that the enantiodiscrimination is almost exclusively provided due to the hydrogen bonds as specific action near to the chiral backbone of PBLG. Indeed, for camphor, which does not have such a specific interaction with PBLG, the cosine similarity between enantiomers is 0.997 for the simulated RDCs. Nonetheless, the experimental RDCs for IPC, HMIP and 1-phenylethanol are almost identical, while the differences are larger for the MD simulations.

Regarding the RDCs themselves, the agreement of the full MD average and only the hydrogen bonds is ≥ 0.98 for all molecules except for 1-phenylethanol (see also Figures 4.10 and 4.11 for the corresponding alignment tensors of the rigid molecules). This indicates that the contribution to the alignment of the poses without hydrogen bonds to PBLG is more important than for the other molecules.

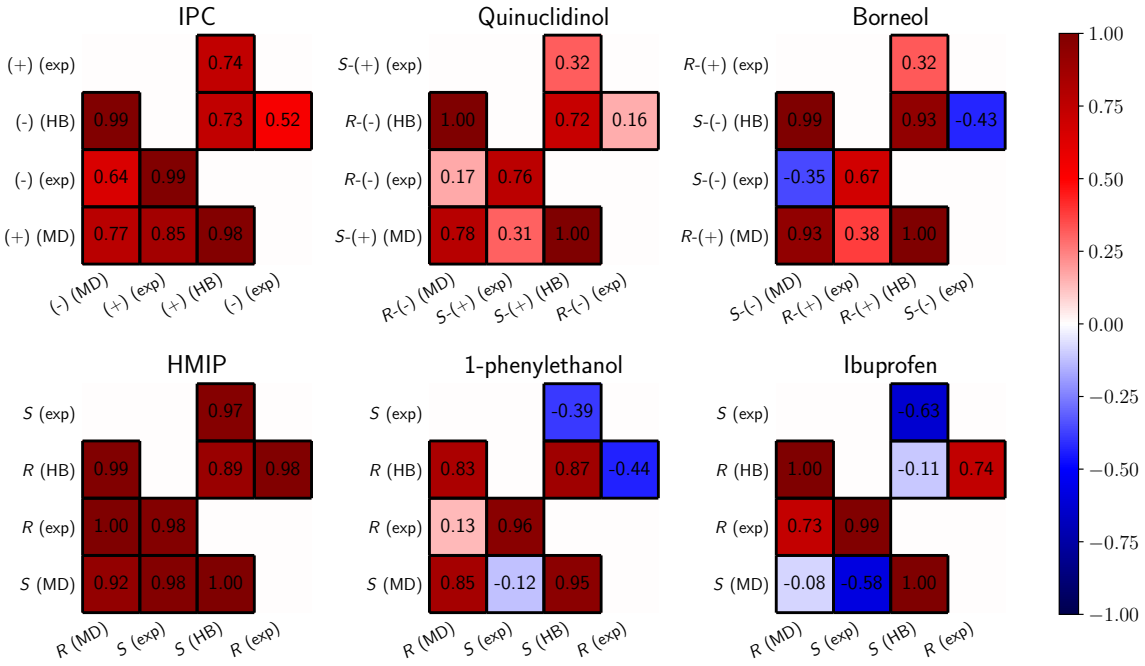


Figure 4.12.: Cosine similarities between different RDCs in both enantiomers of the six analyte molecules with hydrogen bond donor used in the MD simulations. Compared are the MD average, the MD average using only poses with hydrogen bond between the analyte and PBLG, as well as the experimental RDCs. Additionally, the different averages or measured values are compared between enantiomers.

4.1.5. Multiple PBLG chains and analyte molecules in the simulation box

To validate the use of a single PBLG chain and one analyte molecule in the simulation box, I did additional simulations for IPC, quinuclidinol, borneol and camphor with 8 analyte molecules in the simulation box, as well as with 8 analyte molecules and 2 PBLG chains. Furthermore, I did simulations with 1 analyte molecule and a PBLG chain without solvent (i.e. in vacuum) to see its influence on the RDCs. Here, the pressure coupling was turned off. These simulations were done for 15 µs, which is sufficient for 8 analyte molecules in the simulation box, as the RDCs are averaged over all of them. For the vacuum simulations, the alignment is much stronger, such that less sampling is required as well.

Comparing the center of mass distributions in Figures 4.1 and A.11, the interactions between the analyte molecules indeed lower the probability of analyte positions near the PBLG chain. This can be explained in part because the analyte molecules block each other from accessing the hydrogen bond acceptors of PBLG. However, when analyzing the MD trajectories, I noticed that the analyte molecules with O – H hydrogen bond donor tend to form small clusters, which additionally hinders the interactions with PBLG. Especially for quinuclidinol, the poses with close distance to PBLG (with hydrogen bond) are strongly reduced. For camphor, HMIP and 1-phenylethanol, the distribution does not change much. This could be explained by the reduced formation of dimers, as camphor and HMIP do not have an O – H donor and in 1-phenylethanol, it is next to both a CH_3 group and a phenyl ring. The influence of the interactions between analyte molecules can thus also be seen in the hydrogen bonds to PBLG: In Figures A.12 and A.13, the hydrogen bond counts and distributions are shown for the simulations with 8 analyte molecules. Notably, the hydrogen bonds to PBLG are strongly reduced in all cases. For the molecules with O – H donor, the oxygen atom can also function as a hydrogen bond acceptor, while C = O group in ibuprofen and the nitrogen atom in quinuclidinol and HMIP can act as additional acceptors. This is presumably why now, the hydrogen bonds occur the most for borneol. The shape of the hydrogen bond angle and distance distributions in Figure A.13 however did not significantly change compared to the simulations with one analyte molecule (in Figure A.13).

The resulting RDCs are compared in Figures A.14 to A.21. Using 1 or 2 PBLG chains does not make a significant difference, with the cosine similarity being ~ 1.00 in all cases. For the choice between 1 and 8 analyte molecules, the difference is larger, but still acceptable, with the lowest cosine similarity being 0.93 for (-)-IPC. Additionally, the cosine similarity of the RDCs calculated using only the poses with hydrogen bond to PBLG between 1 and 8 analyte molecules is at least 0.97 and in all cases higher than for the comparison of the full averages, which indicates that the change when using multiple molecules comes mostly due to the alignment of (temporary) aggregates of the analyte molecules. Thus, doing the simulations with multiple analyte molecules is a very valid way to reduce the required simulation time, even if the analyte concentration is higher than in experiment.

The RDCs calculated from the vacuum simulations are significantly different from the RDCs calculated with solvent: They show low agreement for camphor, IPC and quinuclidinol, while the cosine similarities are ≥ 0.85 for borneol. Most of the vacuum RDCs also do

not agree with experiment at all. This is because the LJ potential between the solutes is very strong, but is in balance with the LJ potential to the chloroform molecules in the simulations with solvent.

4.1.6. Chirality-driven difference of alignment

Following the original motivation in the thesis of Sager [136], the intended application of the prediction of RDCs in my work is the distinction of enantiomers. Even though for the tested molecules, there is hardly any difference in the interactions with PBLG between enantiomers (see Figures 4.1, 4.2 and 4.4), the average orientation and RDCs differ substantially in the molecules with hydrogen bond to PBLG. Like in the thesis of Sager [136], where it was originally proposed by Prof. Dr. Luy, I refer to this difference between enantiomers as "chirality-driven difference in alignment" (CDDA). The idea for the application of distinguishing enantiomers is to compare the CDDA of the calculated and experimental RDCs. Using only PBLG, this would require measurements with both enantiomers, which severely hinders the application of this method. However, as confirmed experimentally for IPC in Marx, Schmidts, and Thiele [104], the RDCs of one enantiomer aligned by PBLG are the same as when using the other enantiomer with PBDG (Poly-γ-benzyl-D-glutamate, the mirror image of PBLG). If therefore both PBLG and PBDG are available, only one enantiomer of the analyte needs to be available for the experiments to determine its absolute configuration.

Due to differences of the PBLG concentration and alignment in experimental samples [136], the scaling of the alignment differs for measurements with different analyte molecules. To calculate the difference between RDCs, this needs to be compensated for by scaling the RDCs of one enantiomer. This is commonly done by scaling the RDCs by the fraction of the quadrupolar couplings of deuterium in deuterated chloroform, which is also aligned by PBLG and should scale in the same manner as the RDCs (see Section 2.1.5). In fact, this scaling has already been applied on the RDCs in Marx, Schmidts, and Thiele [104], Sager [136], and Sager et al. [137]. For the MD simulations, no such scaling is required, because the simulation box is the same for both enantiomers. Under the assumption that the errors in the alignment made in MD are mostly achiral, the comparison of the CDDA should still give the same result. This nonetheless requires a very high statistical accuracy of the calculated RDCs in order not to introduce random differences between the enantiomers due to insufficient sampling. For camphor, the difference of the RDCs between enantiomers is too small to make a reliable comparison between the measured and the calculated values, as it does not form hydrogen bonds with PBLG and the alignment happens at the ends of the PBLG sidechains, as seen in Figure A.5.

In the thesis of Sager [136], the determination of absolute configuration in this way was demonstrated for HMIP, where the agreement of the RDCs themselves is also rather good. Here, I confirm this observation for HMIP with a higher accuracy of the MD simulations, but also for the other molecules with hydrogen bond to PBLG, where the agreement of the RDCs themselves with experiment is lower. The CDDA results are shown in Figure 4.13, where the differences are compared by the cosine similarity, which is clearly positive in all

cases, confirming the correct assignment of enantiomers. In addition, using only poses with hydrogen bond to PBLG gives similar results, confirming that this is indeed the relevant source of the difference in alignment. Even though the cosine similarity to experiment of the RDCs themselves is near 0 for 1-phenylethanol, borneol, quinuclidinol and ibuprofen (see Figures 4.8 and 4.9), the CDDA is also unambiguous for these cases. Additionally, using multiple analyte molecules as described in Section 4.1.5 does not significantly change the differences of the RDCs between the enantiomers: Comparing Figures 4.13 and A.10, the agreement of the CDDA is very similar to the simulations with one analyte molecule.

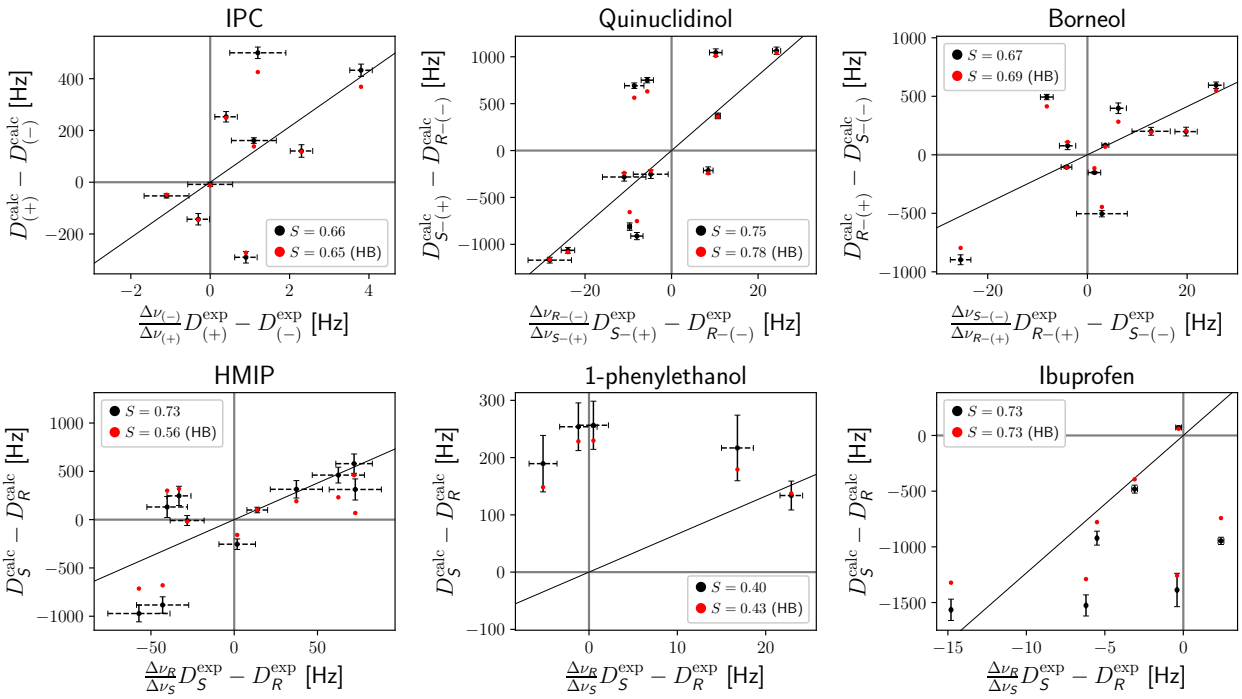


Figure 4.13.: Comparison of the CDDA (difference between RDCs) obtained from the experimental RDCs and the RDCs obtained from the MD simulations for the 6 analyte molecules with hydrogen bond donor. The full average is shown in black together with the experimental uncertainties (assumed to be a triangular distribution) and the standard deviation of the (normally distributed) MD average. In red, the MD average using only the poses with hydrogen bond to PBLG are shown. The experimental RDCs of one enantiomer is scaled to match the quadrupolar coupling of deuterium (in deuterated chloroform) in the measurement of the respective other enantiomer.

To see how the statistical uncertainties from simulation and the experimental errors influence the CDDA cosine similarities, I calculated their 95 % confidence intervals shown in Table 4.1. Like for Figures 4.8 and 4.9, I assumed a normal distribution for the MD average with the estimated covariance matrix, as well as pairwise independent uniform distributions for the individual RDCs from experiment. Of the molecule with hydrogen bond donor, 1-phenylethanol has the worst agreement with experiment, but still a 95 % confidence interval with clearly positive values ranging from 0.27 to 0.51 and thus the correct assignment of the absolute configuration. For the other molecules with hydrogen bond donor, the lower bound of the 95 % confidence interval is significantly higher, even if the RDCs themselves do not agree. The CDDA for camphor is also shown in Table 4.1, but

Table 4.1.: Cosine similarities and their 95 % confidence intervals of the CDDA (difference between RDCs) between experiment and simulation for the different molecules.

Molecule	$S_{\mathbf{CDDA}}$	95 % **CI**
IPC	0.66	[0.56, 0.73]
quinuclidinol	0.75	[0.72, 0.77]
borneol	0.67	[0.61, 0.72]
HMIP	0.73	[0.59, 0.83]
1-phenylethanol	0.40	[0.27, 0.51]
ibuprofen	0.73	[0.72, 0.74]
camphor	0.23	[-0.52, 0.67]

the experimental and statistical uncertainties are too large for a conclusive comparison, with the 95 % confidence interval including a large range of negative and positive values. As mentioned earlier, this confirms that for the assignment of the absolute configuration using RDCs of analyte molecules aligned by PBLG, the hydrogen bonds to PBLG are necessary.

4.1.7. Impact of water molecules on the RDCs

When the probes with the PBLG/chloroform liquid crystal and the dissolved analyte molecules are prepared, it could be possible that they obtain some amount of water due to humidity in the air. So far, it is unclear whether and how much this influences the RDCs, especially for the molecules with hydrogen bond donor.

Therefore, I did additional simulations for (-)- and (+)-IPC with different amounts of water molecules: With 1 IPC molecule in the simulation box, I performed simulations for 30 µs with 1, 2, 4 and 8 water molecules. Additionally, I did simulations with 8 IPC molecules and 4, 8, 12 and 16 water molecules.

The impact of the water molecules on the COM distance of the analytes to the PBLG center axis are shown in Figure 4.14. Both with 1 and 8 analyte molecules, increasing the number of water molecules decreases the number of poses at the first maximum around $\sim$ 7 Å (which are determined by the hydrogen bond).

The influence of the water molecules on the hydrogen bonds of IPC to PBLG can also be seen in Figures 4.15, A.22 and A.23. Increasing the number of water molecules decreases the number of hydrogen bonds, while their average lifetime is only slightly affected. However, the influence does not appear to be as strong as in the case of multiple analyte molecules. Additionally, in the case of 8 analyte molecules in the simulations, the influence of the water molecules on the amount of hydrogen bonding to PBLG is smaller and adding more than 8 water molecules only reduces it slightly. As even in the case of 16 water molecules, there are now only 2 water molecules for each analyte, this is not very surprising. The distribution of the hydrogen bond distances and angles in Figures A.22 and A.23 does not

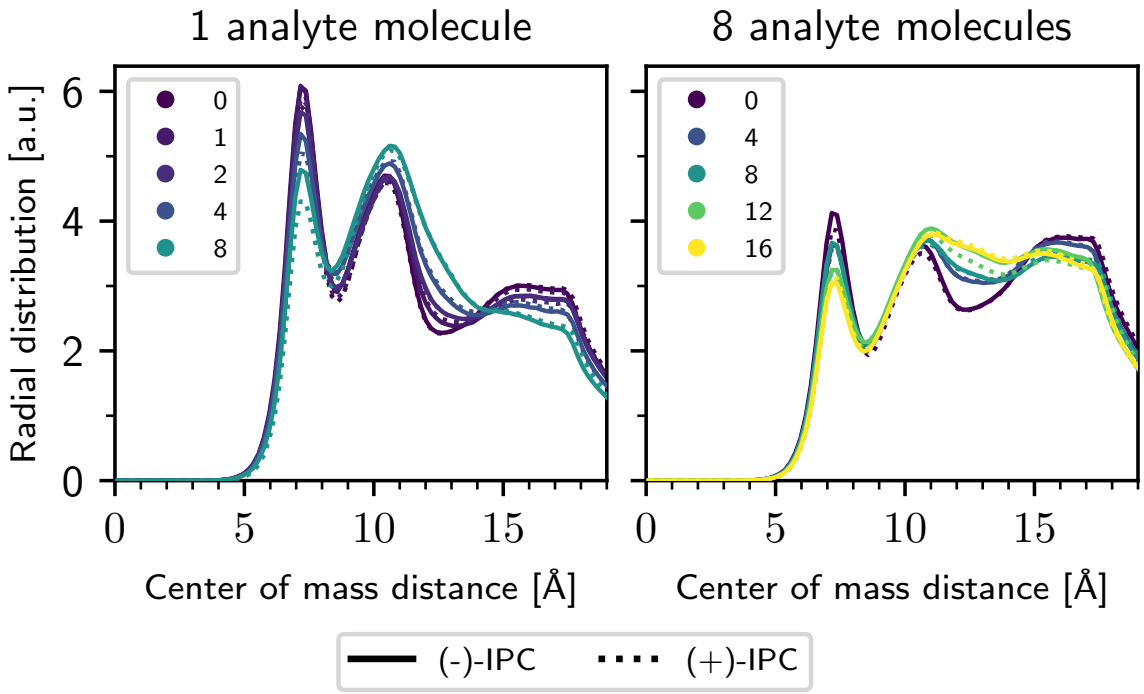


Figure 4.14.: Distribution of the center of mass distance of (-)- and (+)-IPC to the PBLG center axis for different numbers of water molecules in the simulation box indicated in the legend. On the left, the results for the simulations with 1 IPC molecule in the simulations box are shown; and with 8 IPC molecules on the right. Compare Figures 4.1 and A.11 for the distributions of all tested molecules without water.

visibly change apart from being reduced overall due to less hydrogen bonds forming. In all cases, the behavior of the hydrogen bonds is nearly identical between enantiomers, like for the COM distances.

For the RDCs on the other hand, the impact of the water molecules is significant. As Figure 4.16 shows, they change significantly when more water molecules are added, both for 1 and 8 analyte molecules. If the same holds in experiment, this would have practical implications for the measurements with hygroscopic analytes: When preparing the probes, water from the air could mix with the samples, affecting the measured RDCs. If the amount of water is however small compared to the number of analyte molecules, the simulations with IPC indicate that its influence should not cause large changes in the RDCs. The same observation can be made for the CDDA, shown in Figure 4.17: Like for the RDCs, it changes depending on the amount of water molecules, although in all tested cases, the cosine similarity remains clearly positive.

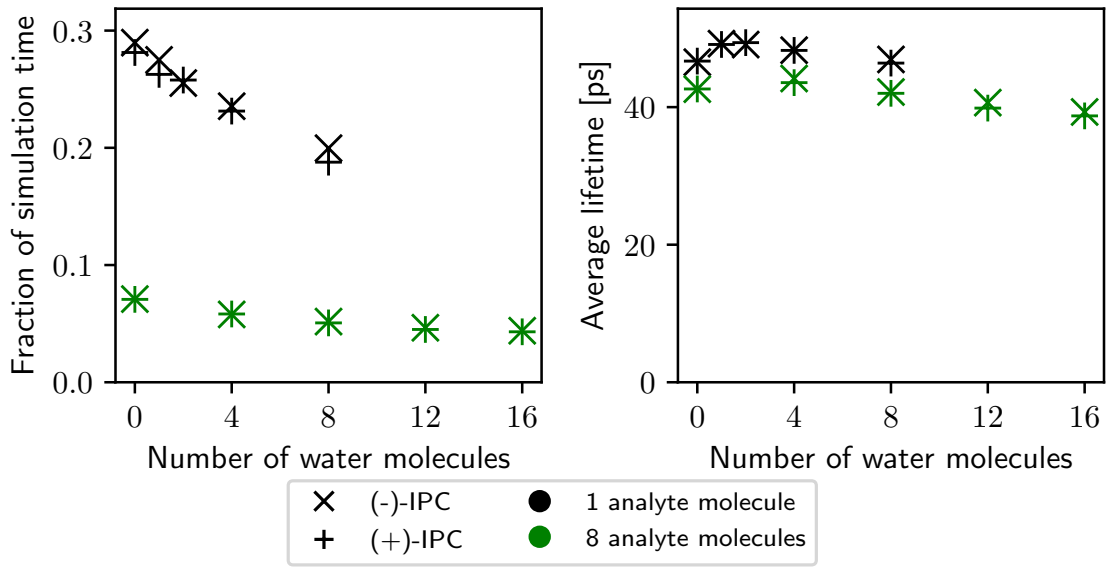


Figure 4.15.: Fraction of simulation time (left) with hydrogen bond to the C = O acceptor in PBLG and their average lifetime (right) for (-)- and (+)-IPC with different numbers of water molecules in the simulations box. The simulations with 1 IPC molecule are shown in black; and with 8 IPC molecules in green. Compare Figures 4.4 and A.12 for the values of all tested molecules without water.

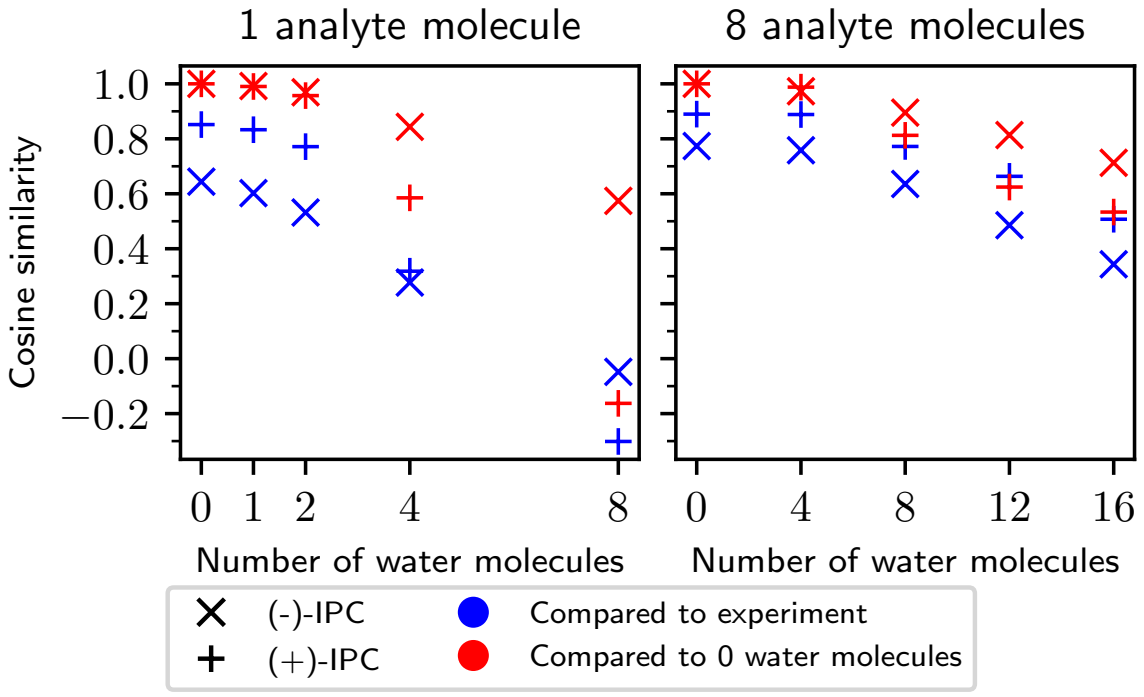


Figure 4.16.: Cosine similarity of the RDCs of (-)-IPC ('×') and (+)-IPC ('+') with different numbers of water molecules in the simulation box compared to experiment (blue) and the simulations with 0 water molecules (red). On the left, the results for the simulations with 1 IPC molecule in the simulations box are shown; and with 8 IPC molecules on the right.

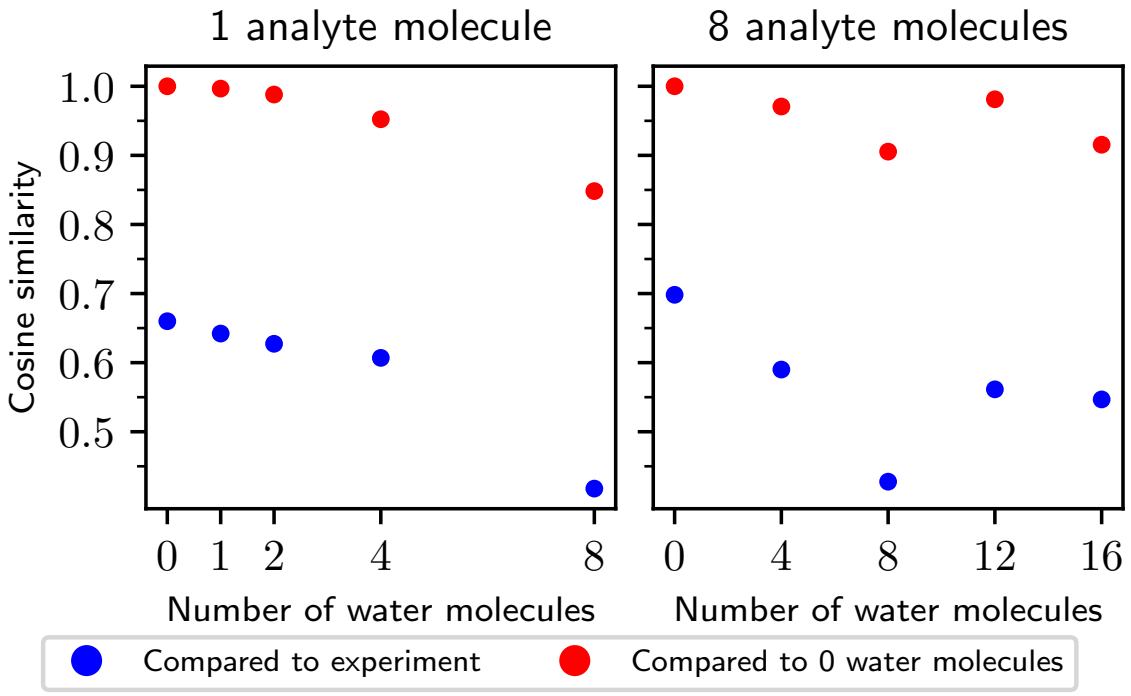


Figure 4.17.: Cosine similarity of the CDDA for IPC with different numbers of water molecules in the simulation box compared to experiment (blue) and the simulations with 0 water molecules (red). On the left, the results for the simulations with 1 IPC molecule in the simulations box are shown; and with 8 IPC molecules on the right.

4.2. Calculation of binding free energies in implicit chloroform

As the results in the previous section show, obtaining RDCs using MD simulations with explicit solvent is computationally expensive. Additionally, the binding free energies calculated using the approach described in Section 3.2.1 illustrate the importance of the different contributions to the binding free energies, which have an approximate physical interpretation with the implicit solvation models shown in Section 3.2. Below, the results of the binding free energies and the fitting of the implicit solvation model parameters are given.

4.2.1. Binding free energy for the test poses calculated with MD

In Figure 4.18, the different nonpolar contributions to the nonpolar binding free energies calculated with MD of the "LJ" test poses are shown for coronene and BCO. The test poses and the calculation of the binding free energies with MD are described in Section 3.2. Especially notable is the strong negative correlation between $\Delta\Delta G_{\text{att}}$ and $\Delta\Delta G_{\text{rep}}$, with a correlation coefficient of -0.989. This is to be expected, because $\Delta\Delta G_{\text{att}}$ describes the missing attractive potential between the solutes and the solvent where the latter is blocked

by the solvent, while $\Delta\Delta G_{\text{rep}}$ describes the missing repulsive interaction between them through the $\gamma\Delta$SASA term and the decrease of the cavity size when the solutes are close through $p\Delta$SAV (see Section 2.5.1). However, it also means that even if both contributions can be determined with relatively good accuracy, the relative accuracy of their sum is lower due to the partial cancellation and is the reason why I fit $\Delta\Delta G_{\text{nonpolar}}$ directly instead of the two contributions. On top of that, the correlation between $\Delta\Delta G_{\text{nonpolar}}$ and V_{LJ} is also very negative, with a correlation coefficient of -0.903. Therefore, the accuracy which can be expected from the nonpolar solvation model is limited, as it has to accurately predict two components which partially cancel each other and whose sum is again partially canceled by V_{LJ}. Nonetheless and due to the lack of an computationally cheap alternative, I proceeded with the model in order to see how accurate it and its components actually are and whether it is applicable for the calculation of RDCs. Interestingly, $\Delta\Delta G_{\text{nonpolar}}$ and $\Delta\Delta G_{\text{att}}$ itself have a strong correlation as well (0.969), but the former is roughly reduced in magnitude by a factor of 2.

A similar comparison is done for the nonpolar, polar and total binding free energies of the test poses for 2-naphthol and BCOO in Figure 4.19. The correlation of $\Delta\Delta G_{\text{total}}$ and $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ is rather high (0.739), especially for the poses with hydrogen bond, as they can be expected to influence the binding energy the most when present. Surprisingly, the correlation between $\Delta\Delta G_{\text{total}}$ and $\Delta\Delta G_{\text{nonpolar}}+V_{\text{LJ}}$ is lower than with $\Delta\Delta G_{\text{nonpolar}}$ alone, although the values are systematically larger and only positive. This might be explained by the strong negative correlation between V_{LJ} and $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ (-0.727).

To confirm that the MD simulations for the calculation of the binding free energies were long enough, the average statistical uncertainties are shown in Table 4.2. For $\Delta\Delta G_{\text{nonpolar}}$ and $\Delta\Delta G_{\text{total}}$, they are the corresponding uncertainties of the MBAR estimator [147] (with subsampling) and lie in the range $0.07\,\text{kcal}\,\text{mol}^{-1}$ to $0.10\,\text{kcal}\,\text{mol}^{-1}$, much lower than the absolute values of $\Delta\Delta G_{\text{total}}$ and $\Delta\Delta G_{\text{total}}$. For $\Delta\Delta G_{\text{att}}$ (and thus $\Delta\Delta G_{\text{rep}} = \Delta\Delta G_{\text{nonpolar}} - \Delta\Delta G_{\text{att}}$), the uncertainties are somewhat higher, ranging from $0.20\,\text{kcal}\,\text{mol}^{-1}$ to $0.25\,\text{kcal}\,\text{mol}^{-1}$.

Table 4.2.: Average statistical uncertainties for the free energies differences from MD, obtained with the MBAR method. All values are in $\text{kcal}\,\text{mol}^{-1}$.

Molecule	**Decomposition**	$\Delta\Delta G_{\text{total}}$	$\Delta\Delta G_{\text{nonpolar}}$	$\Delta\Delta G_{\text{att}}$	$\Delta\Delta G_{\text{rep}}$
coronene	WCA		0.10	0.23	0.25
coronene	σ		0.10	0.21	0.24
BCO	WCA		0.08	0.22	0.23
BCO	σ		0.08	0.20	0.22
2-naphthol		0.09	0.07		
BCOO		0.10	0.08		

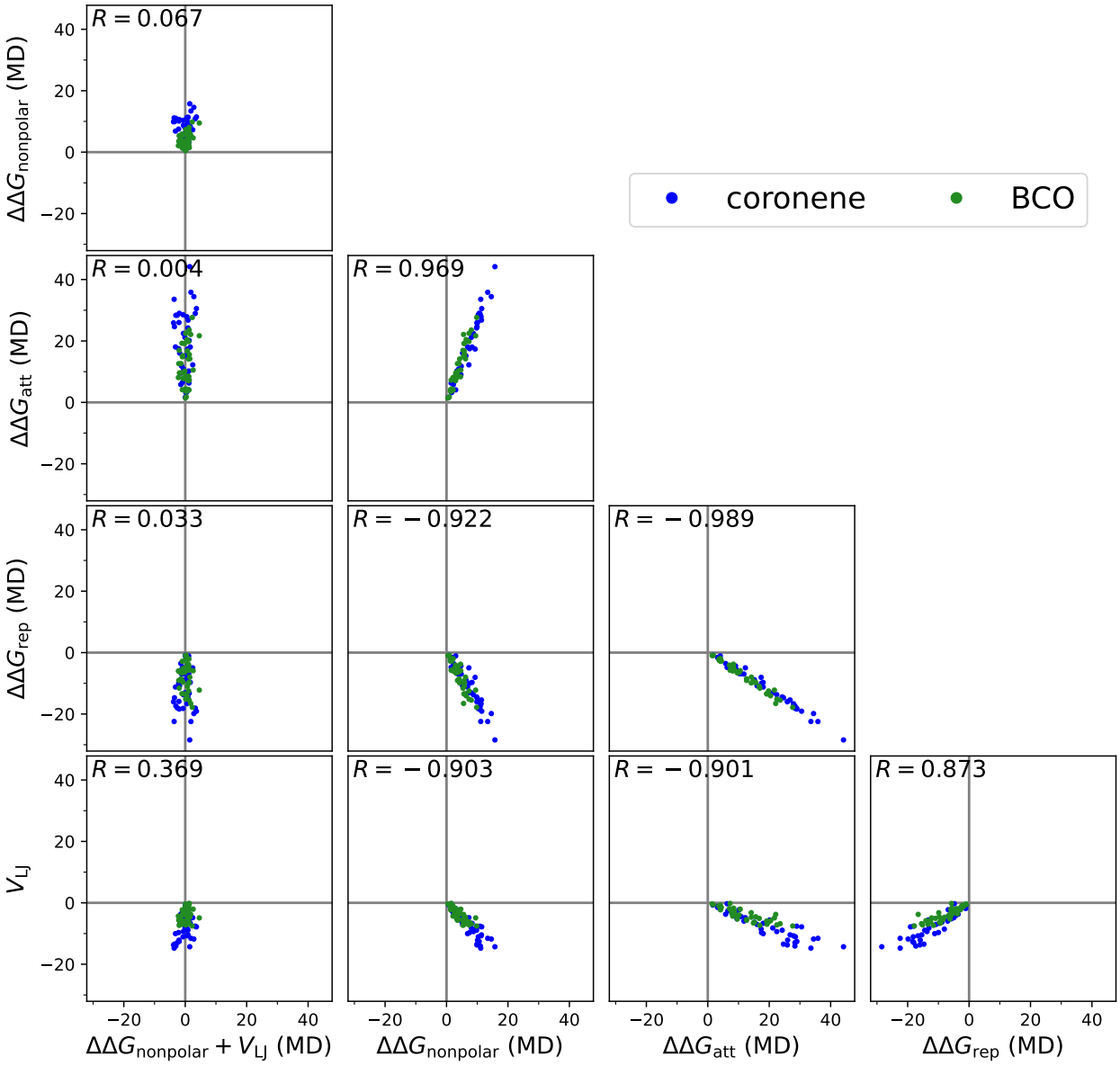


Figure 4.18.: Correlations between the nonpolar binding free energies and their contributions of coronene and BCO calculated with MD simulations. Here, $\Delta\Delta G_{\text{att}}$ is calculated with the WCA decomposition. All (free) energy values are given in kcal mol^{-1}.

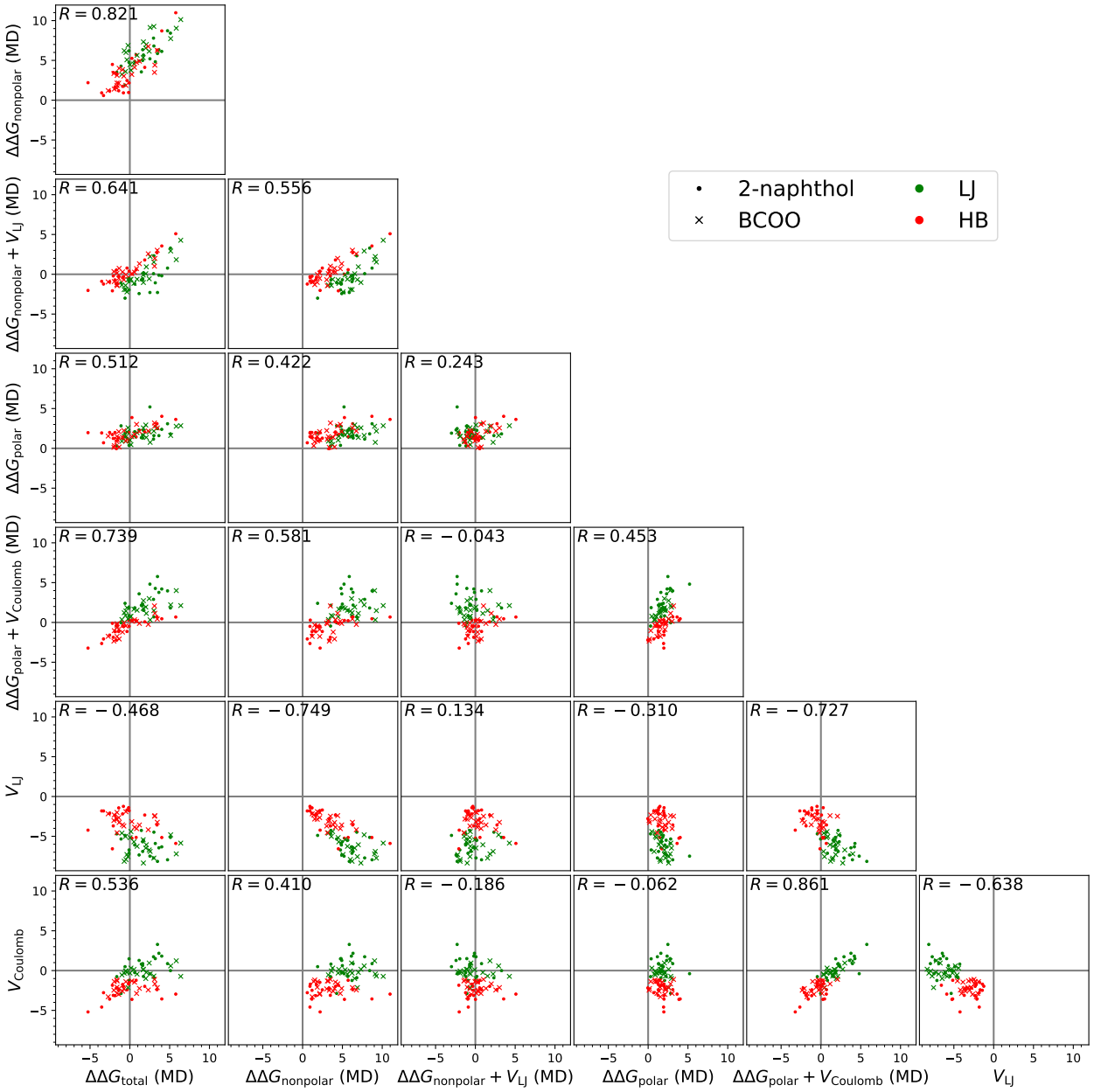


Figure 4.19.: Correlations between the total binding free energies calculated with MD simulations their different contributions for the test poses of 2-naphthol and BCOO. All (free) energy values are given in kcal mol^{-1}.

4.2.2. Nonpolar solvation model

Effective solvent parameters

In the attractive part of the nonpolar solvation (Equation (3.14)), the effective Lennard-Jones parameters of the solvent are used. To determine them, I did MD simulations with single, uncharged atoms in explicit chloroform with different atom types ('H', 'C' and 'O') in a cubic box with side length of 4.5 nm. After running the same equilibration procedure as for the chloroform box (see Section 3.1.3), I performed the simulation for 1 ns, saving the trajectory every ps. From this, I obtained a histogram of $\sum_j V_{\mathrm{LJ}}^{ij}$ over the distance r between the center of mass of chloroform and the single uncharged atom, where i is the single uncharged atom and j are the chloroform atoms.

With this data, the effective Lennard-Jones parameters for chloroform represented by a single "bead" can be determined, replacing the atoms j by a single "bead". As $\sum_j V_{\mathrm{LJ}}^{ij}$ (or actually $V_{\mathrm{LJ,att},i}$, see Equation (3.6)) is not used directly, but instead 3D integrals over the volume of the respective other solute, I do the fitting to $-\int_{\mathbb{R}^3 \setminus B_r} V_{\mathrm{LJ}}(\vec{r}')\, d^3r'$ for different distances r instead, where B_r is a ball of radius r centered around the single uncharged atom i. The values from MD are obtained by integration of the abovementioned histogram.

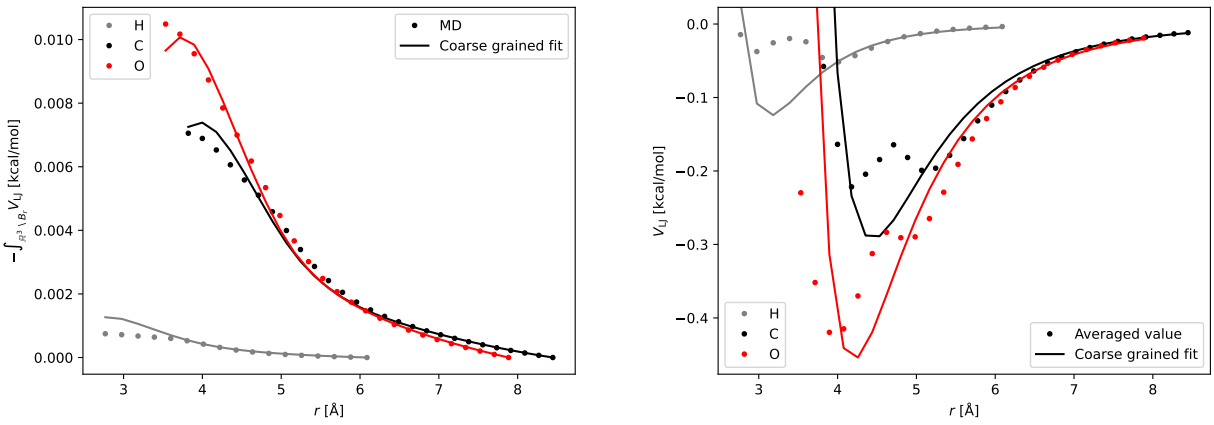


Figure 4.20.: Fit results for the effective Lennard-Jones parameters σ and ϵ of chloroform. **(a)** Fitted values of $-\int_{\mathbb{R}^3 \setminus B_r} V_{\mathrm{LJ}}(\vec{r}')\, d^3r'$ for three atom types compared to the MD averages at different distances r. **(b)** The corresponding comparison for $V_{\mathrm{LJ}}(r)$.

In Figure 4.20a, the fit results are shown for the 3 atom types. As expected, there are no large differences for larger distances, while they are only a few percent for the 'C' and 'O' atom types also at smaller distances. For the 'H' atom type, the total contribution is much smaller, while the error is similar to the other atom types (but occurring at significantly smaller distances due to the smaller atom radius of hydrogen). For comparison, the fitted effective potential is also compared in Figure 4.20b. Here, the differences are larger because of a local maximum in the MD values (as the distance of the individual chloroform atoms

can vary for the same COM distance), which cannot be represented by the coarse-grained model, but the location of the minimum and the repulsive part are still represented as expected. Even so, the agreement in Figure 4.20a is good enough that the approximation with effective LJ parameters appears reasonable for calculating $\Delta\Delta G_{\text{att}}$.

Radial distribution function of single, uncharged atoms to chloroform

To obtain the radial distribution function of single, uncharged atoms used for the "calc." RDF, I did simulations like for the determination of effective Lennard-Jones parameters of chloroform described in the previous subsection. This time, the simulations are needed for every unique combination of the LJ parameters σ and ϵ used for the atoms in PBLG and the analyte molecule. As before, I ran the simulations for 1 ns and determined the distances between the single atom and the center of mass of all surrounding chloroform molecules.

For the AMBER atom types 'HC', 'CT' and 'O', the resulting RDF is shown in Figure 4.21. As expected with the LJ potential, there is a large peak between 4 Å to 5 Å and a minimum around 6 Å to 7 Å, very different from a step function.

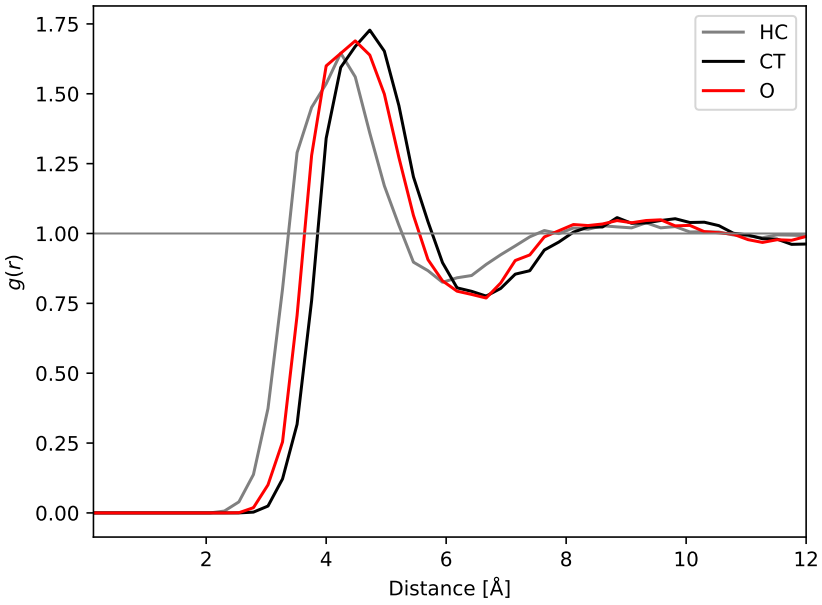


Figure 4.21.: Radial distribution obtained from MD of chloroform to the single uncharged solute atoms with the LJ parameters from the AMBER atom types 'HC' (hydrogen attached to carbon), 'CT' (sp^3 carbon) and 'O' (carbonyl oxygen).

Fit results

To fit the model parameters R_{self}, R_{other}, R_{rep}, γ and p specified in Section 3.2.2, I used the fitting procedure described above in Section 3.2.4 for the decompositions and RDCs also defined in Section 3.2.2. The results are shown in Table 4.3, with the different components shown in Figure 4.22 for the WCA decomposition with the "step" RDF and in Figures A.24 to A.26 for the WCA decomposition with the "calc." RDF and for the σ decomposition with the "step" and the "calc." RDF, respectively.

Table 4.3.: Fitted parameters of the nonpolar solvation model, for each combination of the WCA or the σ decomposition and the two RDFs. R_{self} and R_{other} are the probe radii for $\Delta\Delta G_{\text{att}}$; R_{rep}, γ and p are the probe radius, the surface tension parameter and the pressure parameter for $\Delta\Delta G_{\text{rep}}$.

Dec.	RDF	R_{self} [Å]	R_{other} [Å]	R_{rep} [Å]	γ [kcal mol^{-1} Å^{-2}]	p [kcal mol^{-1} Å^{-3}]
WCA	step	0.20	0.85	1.30	0.0256	0.0385
WCA	calc.	0.40	0.75	1.30	0.0270	0.0358
σ	step	0.00	1.25	1.86	0.0026	0.0310
σ	calc.	0.00	0.95	1.34	0.0256	0.0279

The best agreement of $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ is for the WCA decomposition with the "calc." RDF (Figure A.24), where the RMSD compared to MD is 1.13 kcal mol^{-1} and the correlation coefficient 0.765. However, with the "step" RDF (Figure 4.22), the agreement is not much worse, with a RMSD of 1.24 kcal mol^{-1} and a correlation coefficient of 0.720. Unlike for the results with water of $\Delta G_{\text{nonpolar}}$ reported in Tan, Tan, and Luo [157], the agreement with the σ decomposition is worse. For the "calc." RDF, it is close to the WCA decomposition (see Figure A.26), but for the "step" RDF (Figure A.25), the agreement is significantly worse, with a RMSD of 1.51 kcal mol^{-1} and a correlation coefficient of 0.674.

Considering the fitted model parameters in Table 4.3, the results for the WCA decomposition are mostly similar between the RDCs, with the largest difference being $R_{\text{self}} = 0.2$ Å for "step" and $R_{\text{self}} = 0.4$ Å for "calc.". The fit values of the σ decomposition with the "calc." are also similar, except for $R_{\text{self}} = 0$. However, all fit results for the σ decomposition with the "step" RDF are very different except p: Here, R_{self} is also zero, but to compensate, $R_{\text{other}} = 1.25$ Å is significantly larger, as is R_{rep}. Most surprisingly, $\gamma \approx 0$ in this case (see below).

These results show that the effective description of the "calc." RDF by the "step." RDF is a reasonable approximation for the WCA decomposition, while for a relatively large solvent such as chloroform, it does not work for the σ decomposition. There, the contribution for small radii $< \sigma$ (see Equation (2.37)) is 0, which is probably why the missing first peak of the RDF (see Figure 4.21) cannot be sufficiently compensated by a smaller probe radius R_{self} and a larger probe radius R_{other}.

As ΔSASA and ΔSAV can be expected to be highly correlated, I compare the fitted $\gamma\Delta$SASA and $p\Delta$SAV in Figure A.27. The SASA and SAV values and thus the correlation coefficient do not depend on γ and p, only on the probe radii. It therefore does not significantly change

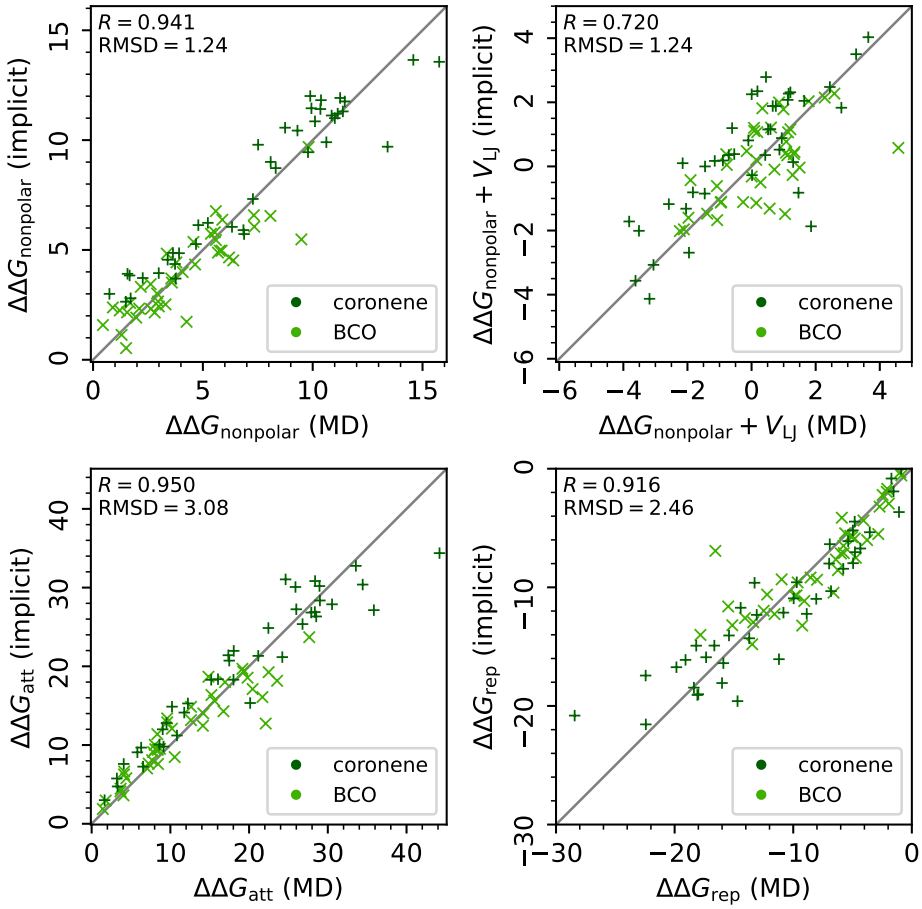


Figure 4.22.: Nonpolar fit results for $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ and its contributions with the WCA decomposition and the "step" RDF. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

between the decompositions and RDFs, being ~ 0.975 in all cases. For the "calc." RDF, the $p\Delta$SAV values are systematically less in magnitude than the $\gamma\Delta$SASA contribution. Most noticeable however is the very different behavior for the "step." RDF with the σ decomposition: Here, the $\gamma\Delta$SASA values are much smaller than the $p\Delta$SAV contribution, while in the other cases, they contribute approximately equally. This again shows that the σ decomposition with the "step." RDF is a less realistic description of $\Delta\Delta G_{\text{nonpolar}}$ and its components compared to the other choices. Therefore, I used the WCA decomposition together with the "step." for the comparison of the total binding free energies discussed in the following.

4.2.3. Polar solvation model

Dielectric constant of chloroform in the MD simulation

To determine how the static dielectric constant of chloroform differs in the MD simulations compared to the experimental values, I calculated it using Equation (2.27) from an MD simulation of pure chloroform. Fitting an exponential to the autocorrelation function of the dipole moment of the chloroform molecules gave an autocorrelation time of $\tau_{\vec{M}} \sim 4\,\text{ps}$. As the statistical error scales with $\sqrt{\tau_{\vec{M}}/t}$, a relatively long simulation time t is required and I did a simulation of the chloroform box over 10 ns such that the statistical error of $|\vec{M}^2|$ was around 2 %. Calculating ϵ_{out} with Equation (2.27) resulted in a dielectric constant of $\epsilon_{\text{out}} = 4.55$, which is consistent with underestimations for other solvents using non-polarizable force fields [48]. As the fitted "GB" and P3D models (described below) however show, the choice of ϵ_{out} on the binding free energies does not matter: In Figure A.36, there are no significant differences between $\epsilon_{\text{out}} = 4.55$ and $\epsilon_{\text{out}} = 4.8$.

Polar fit results

Like for $\Delta\Delta G_{\text{nonpolar}}$, I obtained the model parameters for the "GB" and the "P3D, fit" models to the MD data using the fitting procedure described in Section 3.2.4. For both, the fit was done with $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$, where the fit parameters are the probe radius r_{ext} and the solute dielectric constant ϵ_{in}. Because the "GB" model is an approximation to the Poisson equation (Equation (2.41)) and has the same model parameters, I do the fit to the latter.

The fit results for $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$ are shown in Figure 4.23, where the fit results are $r_{\text{ext}} = 0.39\,\text{Å}$ and $\epsilon_{\text{in}} = 1$. Due to the size of the chloroform molecule and the approximation of the solute-solvent boundary by a step function (with the atoms modeled as spheres), the agreement of $\Delta\Delta G_{\text{polar}}$ alone is not that good, with a correlation coefficient of only 0.394. As the values of $\Delta\Delta G_{\text{polar}}$ are not large however, $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$ has much better agreement, with a correlation coefficient of 0.883 and an RMSD of $1.10\,\text{kcal mol}^{-1}$. Using only V_{Coulomb} on the other hand systematically underestimates the polar contribution to the binding free energy, with a similar correlation coefficient but an RMSD of $2.05\,\text{kcal mol}^{-1}$. Thus, even though the electrostatic model of $\Delta\Delta G_{\text{polar}}$ is not very accurate compared to MD, it is sufficient and less problematic than the nonpolar solvation.

The "GB" model is compared to the Poisson equation (with the same model parameters r_{ext} and ϵ_{in}) and the MD values in Figure 4.24. For poses with low $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$, the agreement with the Poisson equation is high, but it is lower for poses with larger $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$. The reason for this might be that the ΔHCT model for the Born radii becomes less accurate when the atoms of the two solutes are closer. In the plot for $\Delta\Delta G_{\text{polar}}$ compared to the Poisson equation, it appears that there is a relatively small, but systematic overestimation for the "LJ" poses. However, this overestimation actually leads to a better agreement with the MD values, reducing the RMSD to $0.92\,\text{kcal mol}^{-1}$. The cosine similarity of $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ is 0.898 for the "GB" model, higher than with the

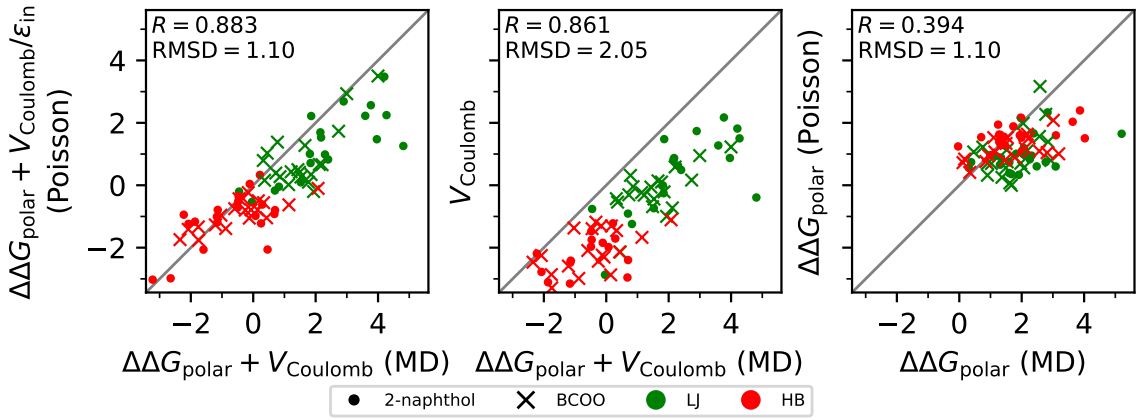


Figure 4.23.: Fit results for $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$ calculated with the Poisson equation to the MD values for 2-naphthol and BCOO. Shown are the results for $r_{\text{ext}} = 0.39$ Å and $\epsilon_{\text{in}} = 1$. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

Poisson equation and much higher than for the nonpolar solvation model. This means that the modeling of the polar solvation contribution to the binding free energy appears to be less challenging than of the nonpolar solvation contribution shown above.

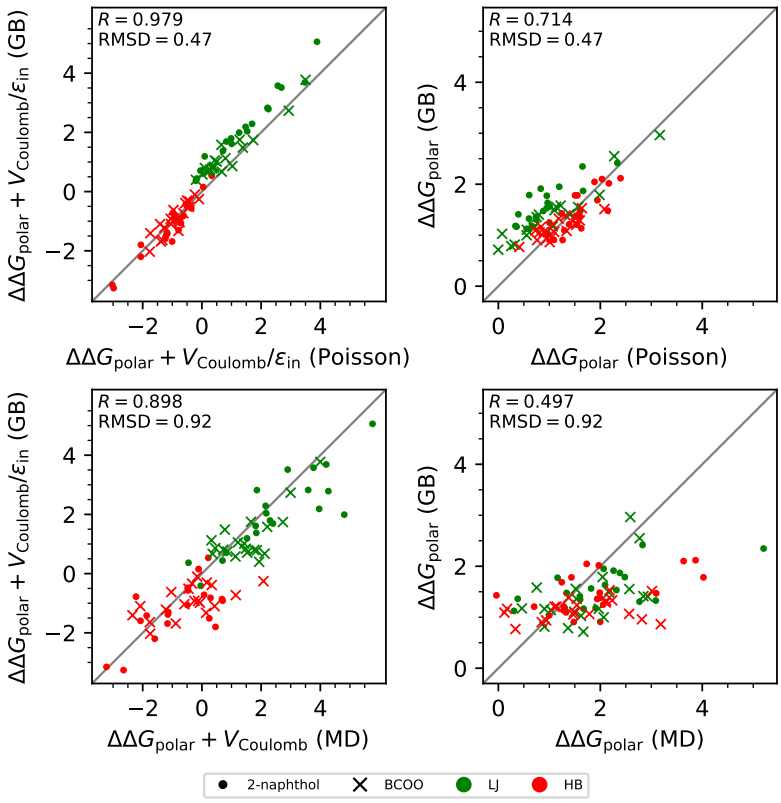


Figure 4.24.: Comparison of $\Delta\Delta G_{\text{polar}}$ and $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}/\epsilon_{\text{in}}$ calculated with the "GB" model to the results from the Poisson equation (fitted to MD data) and the MD data itself. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

4.2.4. Comparison of implicit solvation models

The previously discussed electrostatic models are compared to the MD values in Figure 4.25. For the "P3D, unopt" model, the strength of the electrostatic interaction is severely underestimated, where the attractive potential of the poses with hydrogen bonds is too weak, while there is hardly any repulsion for the close poses with large positive values in MD. With the "P3D, fit" model, some hydrogen bond poses are better represented, but the repulsive contributions are still much too weak. As seen above in Figure 4.19, V_{Coulomb} alone without any implicit solvation shows a large negative systematic shift of the contribution to the binding free energies compared to the MD results with (explicit) solvation. The correlation coefficient is however still high (0.863). The best agreement is for the "GB" model, where there are no significant systematic errors anymore, which was already shown in Figure 4.24.

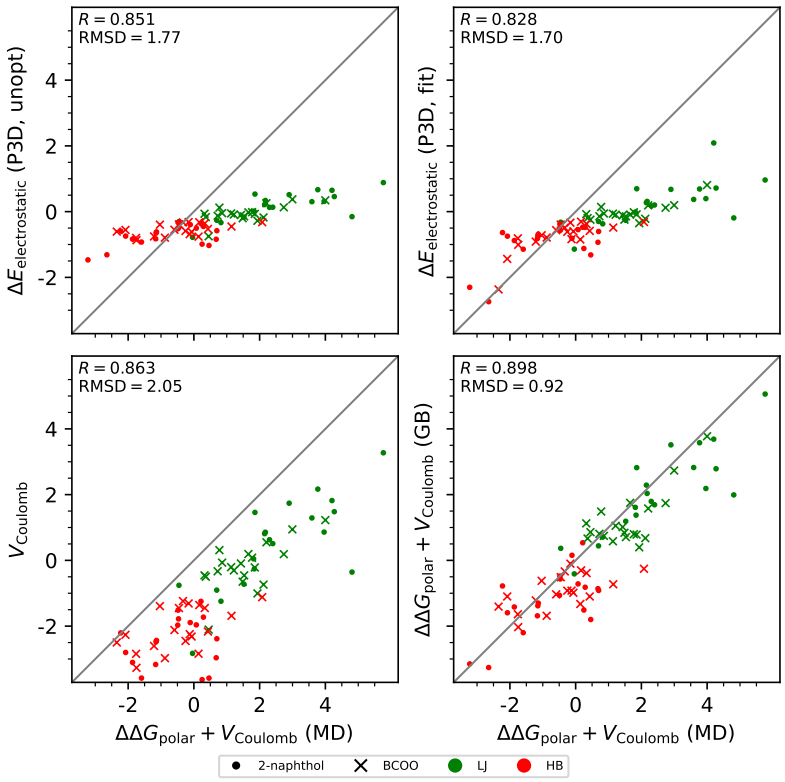


Figure 4.25.: Comparison of $\Delta E_{\mathrm{electrostatic}}$ from the "P3D, unopt" and "P3D, fit" models, V_{Coulomb} and $\Delta\Delta G_{\mathrm{polar}} + V_{\mathrm{Coulomb}}$ from the "GB" model to $\Delta\Delta G_{\mathrm{polar}} + V_{\mathrm{Coulomb}}$ calculated with the MD simulations.

The comparison of the total binding free energies to the MD simulations is done in Figure 4.26 for the different models, including only the polar ones. As expected, the agreement is best for $\Delta\Delta G_{\mathrm{total}}$, which includes the nonpolar implicit solvation model and the "GB" model for the polar part. However, there are still several significant outliers, especially with too negative values, leading to an RMSD of 1.84 kcal mol^{-1} and a correlation coefficient of only 0.641. Using only the polar "GB" model by itself only results in a small reduction of the agreement, although for the test pose with the lowest $\Delta\Delta G_{\mathrm{total}}$ in MD,

the binding strength is significantly underestimated. Like for $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ in Figure 4.25, using only V_{Coulomb} leads to stronger binding of almost all poses, but the correlation coefficient is similar to $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ from the "GB" model. For the P3D models, the situation is also like for $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ alone, where the strength of the interactions is systematically underestimated. The correlation coefficient of "P3D, fit" is however close to the "GB" model.

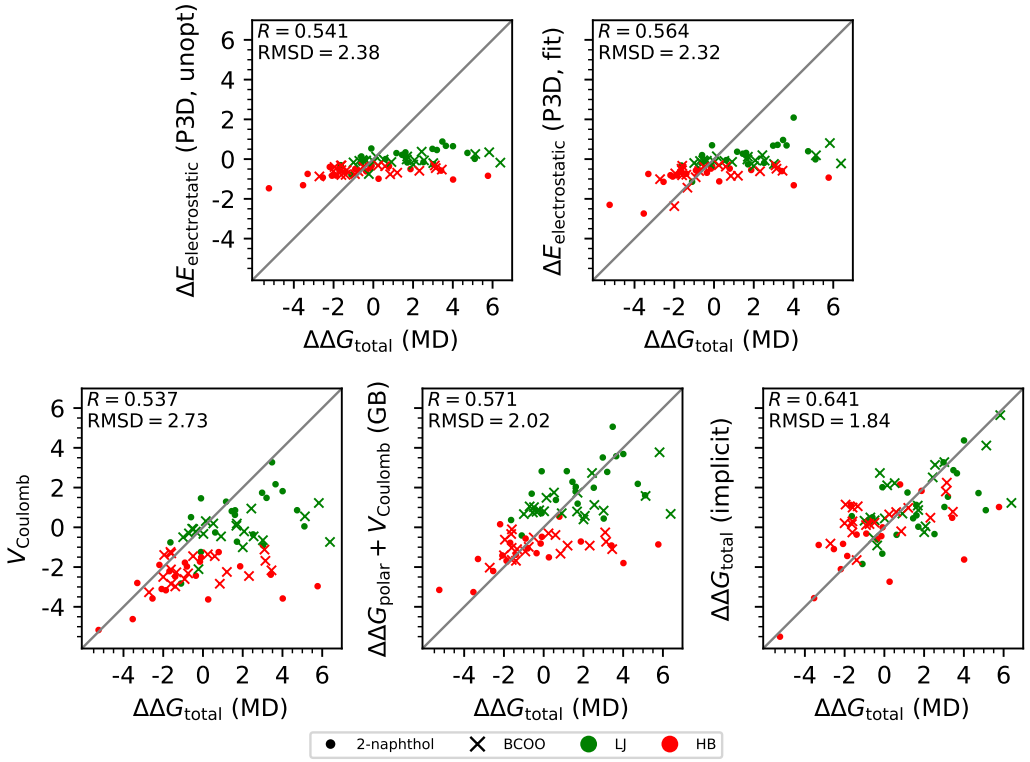


Figure 4.26.: Comparison of $\Delta E_{\text{electrostatic}}$ from the "P3D, unopt" and "P3D, fit" models, V_{Coulomb} and $\Delta\Delta G_{\text{polar}} + V_{\text{Coulomb}}$ from the "GB" model and $\Delta\Delta G_{\text{total}}$ from the implicit solvation models to $\Delta\Delta G_{\text{total}}$ calculated with the MD simulations.

Overall, this shows that even though the polar solvation model is not very accurate by itself, the electrostatic interaction is described reasonably well and the split into the nonpolar and polar contribution is not unfounded. The more problematic part for the description of total binding free energies is the nonpolar solvation model, mostly due to its composition of several contributions with opposite sign.

4.3. RDCs calculated with Monte Carlo

The original motivation for the full implicit solvation model for $\Delta\Delta G_{\text{total}}$ developed in this work (which is described in the previous section) was to understand the importance of the different contributions to the binding free energy on the RDCs. However, in a preliminary test with population annealing for (-)-IPC (using a single optimized structure), I found

a pose with $\Delta\Delta G_{\text{total}} = -12.8\,\text{kcal}\,\text{mol}^{-1}$. Even for strychnine, there were poses with $\Delta\Delta G_{\text{total}} = -11.5\,\text{kcal}\,\text{mol}^{-1}$, even though it has no hydrogen bond donor. This indicates that the nonpolar solvation model is, at least in its current form, too inaccurate for the use in MC calculation of RDCs in analyte molecules aligned by PBLG. Therefore, I keep the hard spheres approach as in the P3D method [116] and use only the models listed in Section 3.3 for the MC calculations in the following.

In the case of the "hard spheres" and the "cylinder" models, no resampling is required and all samples are independent and unbiased, such that the statistical uncertainty can be calculated reliably. With 10^7 samples as for the other energy methods and the calculation procedure described in Section 3.3.5, the largest relative uncertainty (compare Figure 3.4 for the MD data) was around 3.6 % for the "hard spheres" model and 2.5 % for the "cylinder" model.

For the other models, where population annealing is used, the situation appears to be more challenging: Due to the resampling, the samples are no longer uncorrelated and there is a bias due to the finite sample size [171]. Therefore, I opted to determine whether the population size is sufficient by instead doing two independent runs with different random number generator seeds and compare the calculated RDCs between them.

Using "V_{Coulomb}", the number of resampling steps was between 1 (for pinene) and 7 (for quinuclidinol), with the other molecules lying in between. As the minima with hydrogen bond to PBLG are much weaker for the "P3D, unopt" model, only one resampling step was done, but with "P3D, fit", it was up to 5 steps (for quinuclidinol).

4.3.1. Center of mass distances between PBLG and the analyte molecules

Like for the MD simulations, I analyzed the distribution of the distance between the center of mass of the analyte molecules and the center axis of PBLG for the different models of F_{binding}. As the PBLG molecule is perfectly aligned along the z-axis in this case, this is just the distance to the z-axis. For the "P3D, unopt", "P3D, fit", "V_{Coulomb}" and "GB" models, the integration volume was the same and so the distributions can be compared directly. As described in Section 3.3.5, I only do the integration with the "cylinder" and "hard spheres" models for center of mass positions of the analyte where the average over all rotations is not isotropic, which scales the RDCs by a constant factor. However, the RDCs themselves agree very well between the two models (see below), so I scaled the "hard spheres" distribution to match the RDCs of the "cylinder" model for the sake of better visualization.

The resulting distributions are shown in Figure 4.27 (for IPC, quinuclidinol, borneol and camphor) and Section 4.3.1 (for (-)-caulamidine A, (-)-strychnine and β-pinene), with the "cylinder" and "hard spheres" models shown on the respective right using a different scale for the distribution. The corresponding distributions from the MD simulations of the molecules in Figure 4.27 are shown above in Figure 4.1. Here, it is evident that only for "V_{Coulomb}", the shape of the distribution of the three analytes with hydrogen bond donor

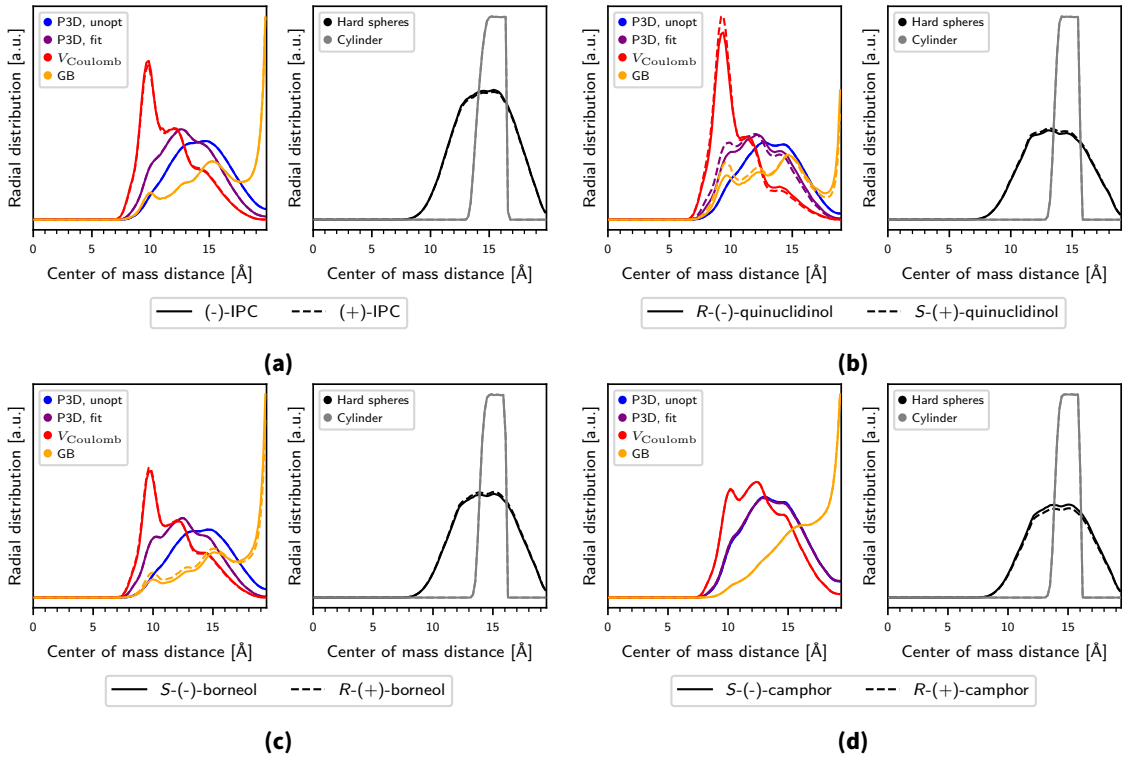


Figure 4.27.: Distribution of the center of mass distance of the analyte molecules to the PBLG center obtained from the MC calculations for **(a)** IPC, **(b)** quinuclidinol, **(c)** borneol and **(d)** camphor, using the different models for the binding (free) energies listed in Section 3.3. The normalization of the respective left part of the figures (in color) are the same, while the distribution of the "hard spheres" model was scaled to match the RDCs of the "cylinder" model. Note that in some cases, the curves of both enantiomers overlap.

(IPC, quinuclidinol and borneol) is similar to the MD values, with the largest peak clearly representing the hydrogen bonds to the C=O acceptor.

Most notable is the "GB" model: Here, there is a large peak at the furthest distance for which the integration is done. This is the case for all molecules, even for β-pinene. This is in spite of being calculated as $V_{\mathrm{Coulomb}} + \Delta\Delta G_{\mathrm{polar}}$. As shown in Figures 4.19 and 4.24, $\Delta\Delta G_{\mathrm{polar}}$ is positive for all test poses of 2-naphthol and BCOO, both for the values determined by the MD simulations and the "GB" model, because the dielectric constant within the solutes is smaller than for the solvent. In the actual MD-simulations however, also the Lennard-Jones potential is present, complicating the situation. This is probably the reason why the significant repulsive effect of $\Delta\Delta G_{\mathrm{polar}}$ is not observed there. It is also hinted at by looking at the hydrogen bonds with PBLG: $V_{\mathrm{Coulomb}} + \Delta\Delta G_{\mathrm{polar}}$ is not strongly overestimated for 2-naphthol and BCOO compared to MD in Figure 4.25. Looking at the poses with $\Delta\Delta G_{\mathrm{total}}$ in Figure 4.26 however, $V_{\mathrm{Coulomb}} + \Delta\Delta G_{\mathrm{polar}}$ from the "GB" model is an overestimation to $\Delta\Delta G_{\mathrm{total}}$ from MD in all cases. For the molecules with hydrogen bond donor, there nonetheless still exists a peak near the distance corresponding to the C=O acceptor, but much reduced in magnitude and located a little further away than for "V_{Coulomb}" alone. The distribution for the "P3D, unopt" and "P3D, fit" models lies in between, with the "P3D, fit" model being shifted to smaller distances. This is because there, r_{ext} is > 0, extending

the volume of PBLG (with assumed internal dielectric constant of 1) for the calculation of the electrostatic potential, such that the strength of the electrostatic interaction between the analytes and PBLG is increased for short distances. Likewise, the "hard spheres" model, where the binding (free) energy is set to 0 everywhere, has almost an identical distribution compared to "P3D, unopt".

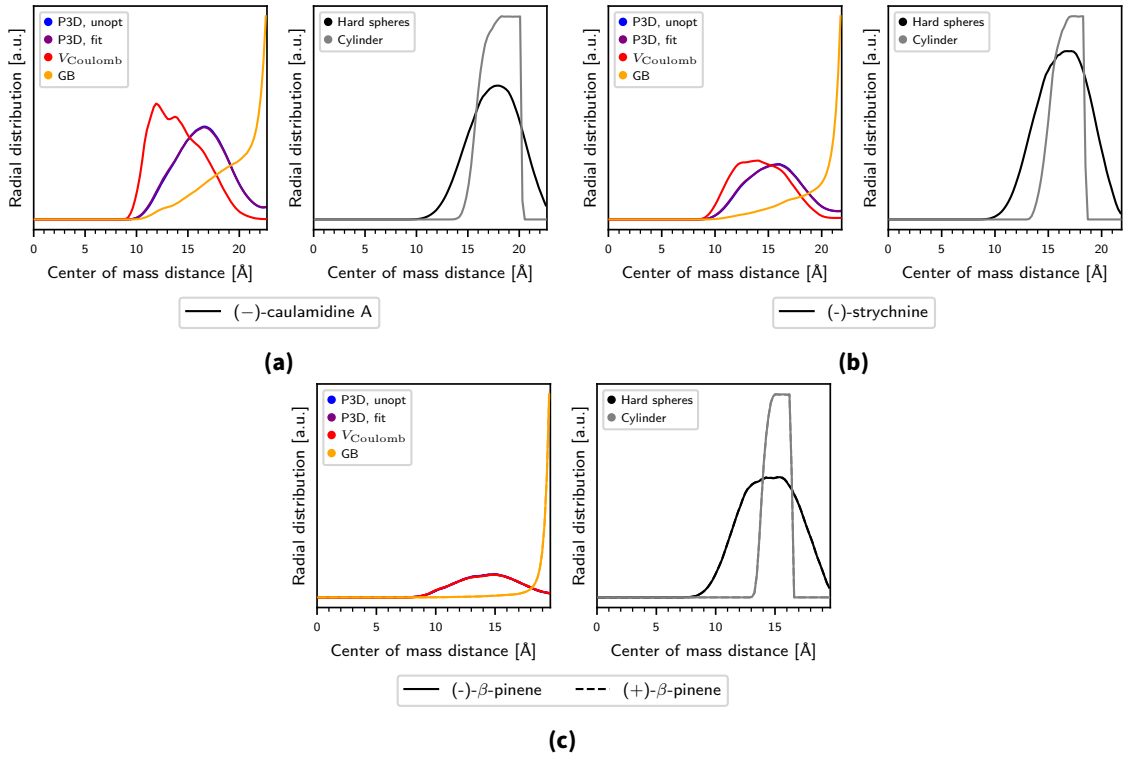


Figure 4.28.: Distribution of the center of mass distance of the analyte molecules to the PBLG center obtained from the different MC calculations for **(a)** (-)-caulamidine A, **(b)** (-)-strychnine and **(c)** (-) and (+)-β-pinene. The normalization of the respective left part of the figures (in color) are the same, while the distribution of the "hard spheres" model was scaled to match the RDCs of the "cylinder" model. For (-)-caulamidine A and (-)-strychnine, the curves of "P3D, unopt" and "P3D, fit" overlap and in the case of pinene, also of "V_{Coulomb}". Additionally, there are no visible differences between enantiomers for pinene.

Noteworthy is also that for the MD results of quinuclidinol in Figure 4.1, the closest distances which appear in the simulation are at around 5 Å. For the MC simulations however, it lies around 7 Å. There are three possibilities to explain this: In the MD simulations, two atoms i and j can come closer than $0.5(\sigma_i + \sigma_j)$, which is prohibited in the MC simulations. Also, the solvent effects are not represented accurately in the MC simulations. Additionally, it could be that (small) changes in the PBLG structure where enough space exists for the analyte molecules to be a little closer are unlikely for the isolated PBLG molecule, but appear when a hydrogen bond to the analyte molecule is formed. Given a perfectly accurate implicit solvation model and using the LJ potential between solutes instead of the hard spheres approach, this discrepancy would not be present for an infinite PBLG chain where the whole configuration space of the sidechains is sampled. Here, I only use a chain of 10 nm length, so this possibility cannot be excluded

however, although the RDCs do not change in most cases when using a different PBLG structure and the hard spheres approach seems likely to be the largest factor.

For the "cylinder" model, the distance distribution is as expected from the procedure described in Section 3.3.5: There is a step function at the outer radius, where the analyte molecule has no overlap with the cylinder for any rotation. Because the molecules are not perfect spheres, the transition to the minimal distance, where the analyte molecule overlaps with the cylinder, is continuous.

β-pinene only consists of carbon and hydrogen atoms and has a very small dipole moment [110]. Therefore, "P3D, unopt", "P3D, fit", "V_{Coulomb}" and "hard spheres" have almost identical center of mass distributions, considering that the electrostatic interaction is very weak and the integration is done over the exact same volume. As discussed above, the repulsive effect of the lower dielectric constant assumed for the solutes compared to the solvent in the "GB" model causes the distribution to be very different, with the largest value at the furthest distance.

For the z-distribution of the center of mass, the analogous plots are given in Figures A.37 and A.38. By definition, the distribution is uniform for the "cylinder" model, while it is less uniform for the "hard spheres" model due to the geometry of the sidechains in the used PBLG structures. Nonetheless, the resulting RDCs are almost identical (see below). With the electrostatic models, the distribution for the molecules with hydrogen bond donor is mainly determined by the hydrogen bonds. Especially for "V_{Coulomb}", there are several large peaks corresponding to PBLG sidechains which are oriented such that the hydrogen bond acceptor is exposed. However, even for camphor, (-)-caulamidine A and (-)-strychnine, there are larger peaks than for the "hard spheres" model due to their polar groups. Only for β-pinene, the peaks are significantly less pronounced, albeit positioned differently than for the hard spheres model.

Another way to compare the models used for the MC calculations is by looking at the distribution of the binding (free) energies to PBLG. It is shown in Figure A.39 for all tested molecules. For "V_{Coulomb}", the binding is vastly more strong, with energy minima up to $\approx -8\,\text{kcal}\,\text{mol}^{-1}$ for quinuclidinol, although the poses with binding energy $\geq -3\,\text{kcal}\,\text{mol}^{-1}$ are still in the majority for all molecules. As described earlier, the "GB" model alone leads to a repulsive force between PBLG and the analyte molecule. In contrast to the other models, there is now a significant fraction of poses with positive binding free energies. The free energies minima are significantly less deep than with "V_{Coulomb}" alone, such that no significant amount of poses with binding free energies $\leq -0.8\,\text{kcal}\,\text{mol}^{-1}$ remains. This shows that compared to MD, the electrostatic interaction and the hydrogen bonds are significantly overestimated for "V_{Coulomb}". Nevertheless, the agreement of the RDCs with experiment is however *better* in most cases as I show below. Unlike the two models above, the P3D models do not have a noticeable contribution at lower energies, which likely corresponds to poses with hydrogen bond. As "P3D, unopt" underestimates the electrostatic interaction compared to MD, the "P3D, fit" model has unsurprisingly somewhat lower binding (free) energies. Due to the apolar nature of β-pinene, all binding (free) energies are close to 0. As with the COM distances and z coordinate distributions, there are no significant differences between enantiomers in any case, confirming the results from the

MD simulations, where the binding behavior was essentially the same for all pairs of enantiomers.

4.3.2. Hydrogen bonds to PBLG

Like for the MD simulations in Figure 4.5, I calculated the distribution of the hydrogen bond distances and angles between the analyte molecules for the different MC models. They are shown in Figures 4.29 to 4.31 for IPC, quinuclidinol and borneol, respectively. Due to the treatment of the atoms as hard spheres, the distribution is zero below a distance of 2 Å. Here, it is important to note that the `HO` atom type in the AMBER force field has LJ parameters $\sigma_i = 0$ and $\epsilon_i = 0$, i.e. only the attached oxygen atom interacts with other atoms via the LJ potential. The minimal possible hydrogen bond distance is therefore increased due to the hard spheres of the oxygen atom and the other atoms, but not due to the hydrogen atom itself. Apart from being at a larger distances, the hydrogen bond distributions with "V_{Coulomb}" and "GB" resemble the MD results in Figure 4.5 most closely, especially considering the angular distribution. Because of the modeling by hard spheres and possibly the rigid PBLG molecule, they are however also less populated compared to MD, despite the overestimation of the electrostatic interactions by "V_{Coulomb}". For the "P3D, unopt" model, there is no significant peak and the distribution is more uniform (as mentioned above in Section 4.1.2, a uniform density does not appear uniform in the plots). With the "P3D, fit" model, the distribution lies in between.

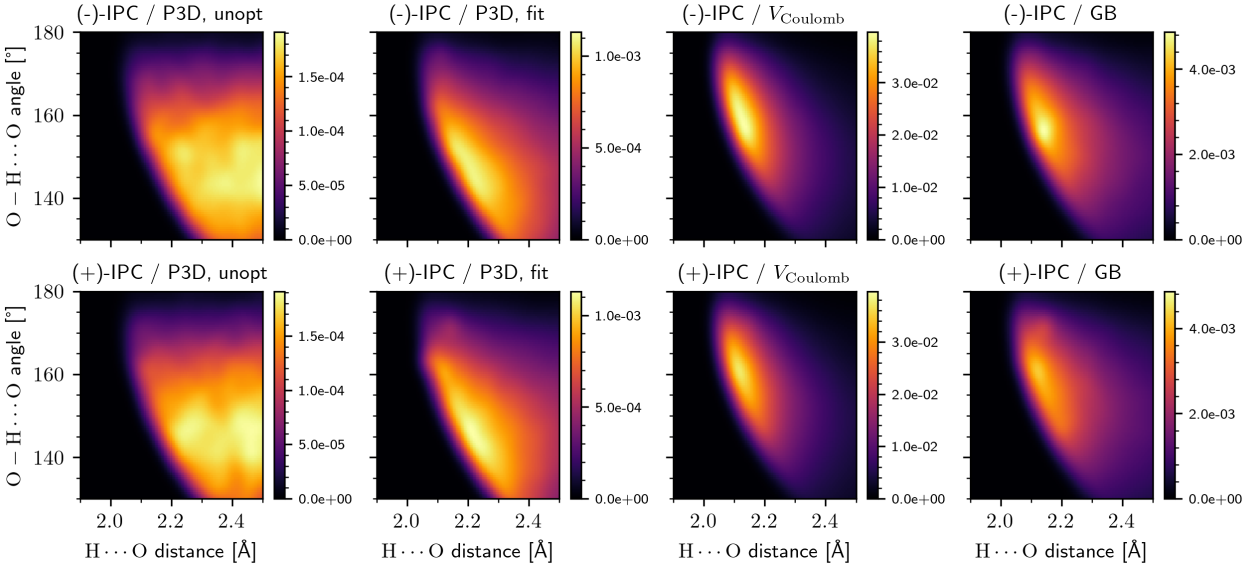


Figure 4.29.: Distribution of the O – H · · · O hydrogen bond distances and angles obtained from the different MC calculations for (-)- and (+)-IPC.

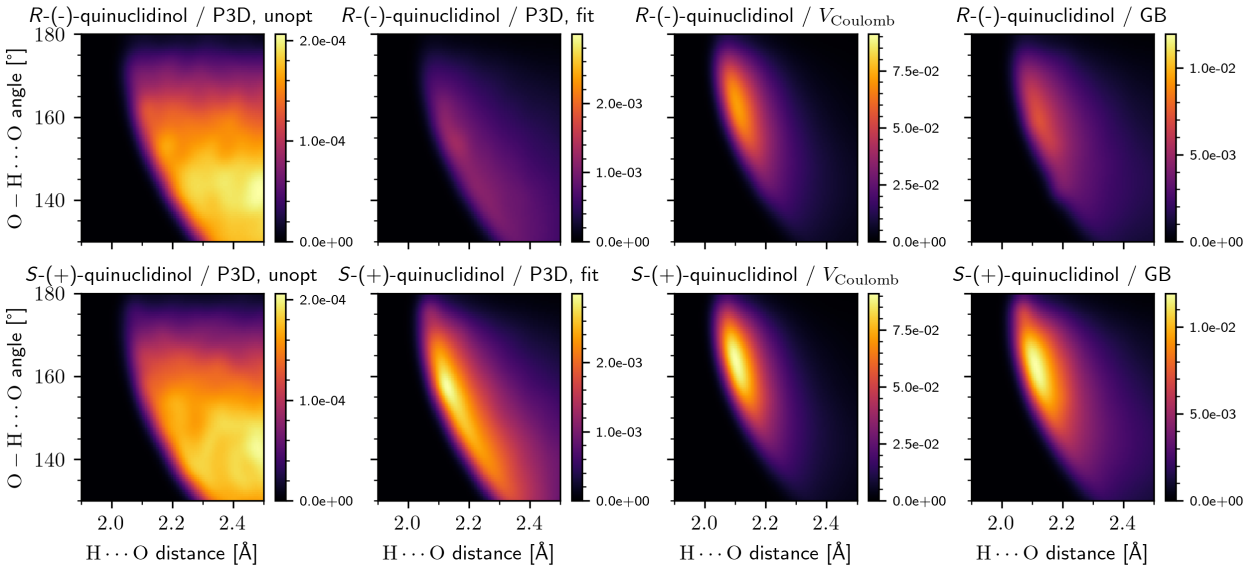


Figure 4.30.: Distribution of the O – H · · · O hydrogen bond distances and angles obtained from the different MC calculations for *R*-(-)- and *S*-(+)-quinuclidinol.

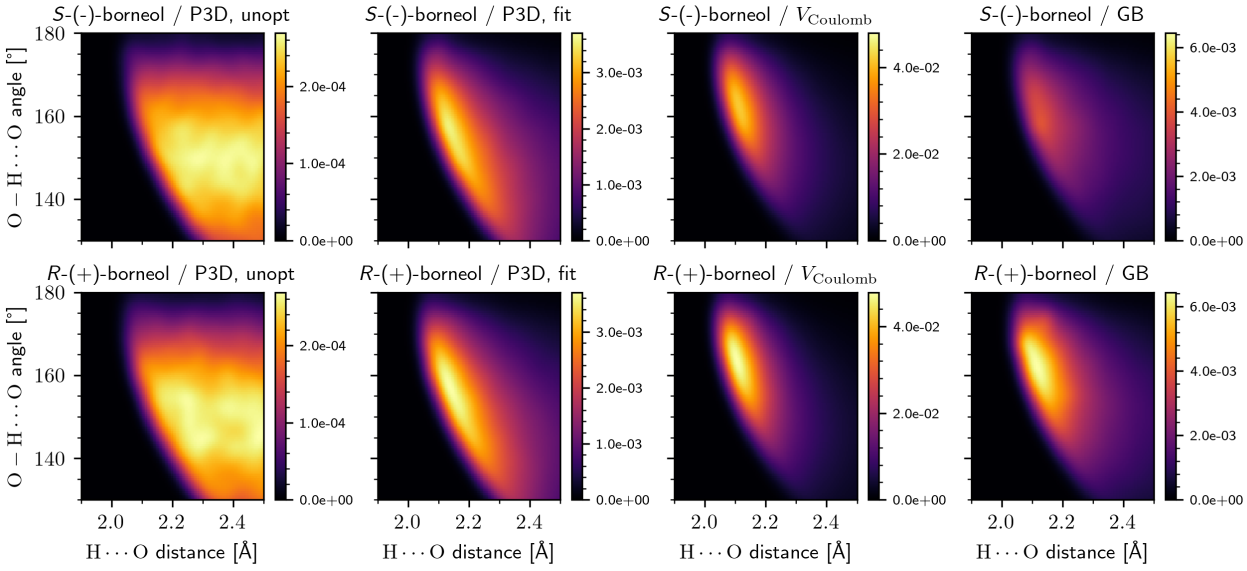


Figure 4.31.: Distribution of the O – H · · · O hydrogen bond distances and angles obtained from the different MC calculations for *S*-(-)- and *R*-(+)-borneol.

4.3.3. Convergence

In the population annealing approach described in Sections 3.3.3 and 3.3.4, there are four numerical parameters which need to be chosen large enough for sufficient convergence: The population size (10^7); the number of MCMC steps (30 for each resampling step); the length of the PBLG snapshot within the integration volume (10 nm); and the number of analyte molecule structures (1000). In Tables 4.4 and 4.5, those choices are confirmed by respectively comparing the RDCs between two different calculations for the “P3D, unopt”,

"V_{Coulomb}" and "hard spheres" methods. The other electrostatic methods can be assumed to lie in between "P3D, unopt" and "V_{Coulomb}", while the convergence of the "cylinder" model should not be significantly slower compared to "hard spheres" due to the minimal choice of the radius of the cylindrical integration volume.

To see that the length of the integration volume along the z-axis is sufficient, the RDCs are compared to calculations with a different PBLG structure in Table 4.4. For "P3D, unopt" and "hard spheres", the cosine similarity of the RDCs is ≥ 0.97 and even for "V_{Coulomb}" it is ≥ 0.94 except for (-)-caulamidine A. With the latter, the agreement between using different PBLG structures is much lower (0.81). Also, increasing the box used for the integration to a z-size of 18 nm (the PBLG molecule used here is $\sim$ 20 nm long) did not increase the agreement between snapshots for "V_{Coulomb}". The problem with using "V_{Coulomb}" for (-)-caulamidine A is probably because of its larger size compared to the other molecules. Thus, the results for (-)-caulamidine A with "V_{Coulomb}" (and possibly with "GB" and "P3D, fit" as well) shouldn't be taken as conclusive.

For the convergence of the population annealing average, I did separate calculations with a different random generator seed and compared the parameters in Table 4.5. In all cases, the cosine similarity to the previous calculation was ≥ 0.98, with most values ~ 1.00. This confirms that the approach works sufficiently well and can even deal with the deep energy minima of "V_{Coulomb}" reliably.

Regarding the analyte structures, the necessity of using multiple structures is demonstrated in Table 4.4, where the average calculated with 1000 analyte structures is compared to the average with the DFT-optimized reference structure. With "V_{Coulomb}", the agreement is only relatively large for (-)-caulamidine A and (-)-strychnine (0.89 and 0.97), but ≤ 0.5 for all other cases, even for the relatively rigid molecules like IPC and quinuclidinol, where the cosine similarity is large for the "P3D, unopt" and "hard spheres" models. As I show below, this is to a large part because of the O – H hydrogen bond donor, where the rotation of the hydrogen atom around the C – O axis is essential for the alignment with hydrogen bonds. Another important source of flexibility are CH_3 groups: Even though the prediction is largely identical between the "hard spheres" and "cylinder" models (see below), the results with the "hard spheres" model for borneol, camphor and pinene are very different with one or many analyte structures. To confirm that the 1000 analyte structures are enough, I repeated the same calculation with 2000 analyte structures. As shown in Table 4.5, the lowest cosine similarity of the RDCs compared to 1000 structures was 0.98 for R-(-)-quinuclidinol and "V_{Coulomb}".

Therefore, the numerical parameters for the MC simulations using population annealing (i.e. the population size and the $\Delta\beta$-steps for the resampling) are sufficient to compare the RDCs conclusively in all cases. The same can be said about the number of analyte structures and the length of the integration volume along the PBLG molecule, with the exception of (-)-caulamidine with "V_{Coulomb}" and to a lesser extent quinuclidinol and strychnine with "V_{Coulomb}".

Table 4.4.: Cosine similarities between the MC calculations with 1000 analyte structures and calculations with a different PBLG snapshot or a single analyte structure.

Molecule	Different PBLG snapshot			Single analyte structure		
	P3D, unopt	V_{Coulomb}	Hard spheres	P3D, unopt	V_{Coulomb}	Hard spheres
(-)-IPC	1.00	1.00	1.00	0.83	-0.16	1.00
(+)-IPC	1.00	1.00	1.00	0.83	-0.13	1.00
S-(-)-borneol	1.00	1.00	1.00	0.02	0.10	-0.81
R-(+)-borneol	1.00	1.00	1.00	0.05	0.11	-0.81
S-(-)-camphor	0.98	0.98	1.00	0.14	0.50	0.22
R-(+)-camphor	0.99	0.96	0.99	-0.06	0.28	0.27
R-(-)-quinuclidinol	0.99	0.99	1.00	0.96	0.12	0.97
S-(+)-quinuclidinol	1.00	0.94	1.00	0.96	-0.07	0.98
(-)-caulamidine A	0.97	0.81	1.00	0.95	0.89	0.97
(-)-strychnine	0.97	0.96	1.00	0.82	0.97	1.00
(-)-β-pinene	0.99	1.00	1.00	0.00	0.16	-0.59
(+)-β-pinene	1.00	0.99	1.00	0.11	0.22	-0.59

Table 4.5.: Cosine similarities of the RDCs between the MC calculations with 1000 different analyte structures and calculations with a different random seed or 2000 different analyte structures.

Molecule	Different random seed			2000 analyte structures		
	P3D, unopt	V_{Coulomb}	Hard spheres	P3D, unopt	V_{Coulomb}	Hard spheres
(-)-IPC	1.00	1.00	1.00	1.00	1.00	1.00
(+)-IPC	1.00	1.00	1.00	1.00	1.00	1.00
S-(-)-borneol	1.00	1.00	1.00	0.99	1.00	1.00
R-(+)-borneol	1.00	1.00	1.00	0.99	1.00	1.00
S-(-)-camphor	1.00	1.00	1.00	1.00	1.00	1.00
R-(+)-camphor	0.99	1.00	1.00	1.00	1.00	1.00
R-(-)-quinuclidinol	1.00	0.98	1.00	1.00	0.98	1.00
S-(+)-quinuclidinol	1.00	0.99	1.00	1.00	0.99	1.00
(-)-caulamidine A	1.00	1.00	1.00	1.00	1.00	1.00
(-)-strychnine	0.99	1.00	1.00	1.00	1.00	1.00
(-)-β-pinene	1.00	1.00	1.00	1.00	1.00	1.00
(+)-β-pinene	1.00	1.00	1.00	1.00	1.00	1.00

4.3.4. Multiple analyte structures

As shown in Table 4.5, the use of a single analyte structure for the molecules with O – H hydrogen bond donor (i.e. IPC, borneol and quinuclidinol) yields very different RDCs if "V_{Coulomb}" is used, also for IPC and quinuclidinol, which are otherwise nearly rigid. Therefore, it stands to reason that the alignment when a hydrogen bond is formed depends

strongly on the position of the hydrogen atom, as it can rotate around the C – O bond rather freely.

To confirm this, I calculated the distribution of C – C – O – H dihedral angles (see Figure 3.2) for the MD simulations and the different MC calculations. For the first carbon atom, I chose atom number 4 for IPC and atom number 6 for quinuclidinol and borneol, using the numbering in Figure 3.2. From the MD simulations to calculate RDCs, I calculated the distribution of the poses with hydrogen bonds. As the 1000 analyte structures used for the MC calculations are obtained from (unbiased) MD simulations of only the analytes in chloroform, I used them to represent the distribution without interactions with PBLG. Like for the MD simulations, I additionally calculated the distributions from the poses with hydrogen bond obtained from the MC results for the different models. To make the distributions better visible, I smoothed them by kernel density estimation using Gaussians with a width of 10° and periodic boundaries at 0 and 360°. As the for the same distribution, the dihedral angles of the opposite enantiomer have opposite sign, the x-axis for one enantiomer is reversed in the figure.

The results for IPC, quinuclidinol and borneol are shown in Figures 4.32 to 4.34. There are three large and very broad peaks (with a width of $\sim 100°$) for all three molecules. This means that using a single structure is indeed not sufficient and even using three different analyte structures would be inaccurate. Interestingly, the poses with hydrogen bond in MD have a very different distribution from the free analyte, where the relative size of the three peaks changes considerably. The distributions from the poses with hydrogen bond for different MC models are closer to the MD average than the distribution of the free analyte for IPC and quinuclidinol, but significant differences remain for all molecules. Like for the other properties except the RDCs, the distributions are almost identical between enantiomers (with opposite sign of the dihedral angles), also in the MD simulations. For the 1000 analyte structures, the distributions are the same by definition, and I use the mirrored structures of one enantiomer for the respective other enantiomer.

Regarding the discrepancy between the RDCs from MD and experiment for the molecules with O – H hydrogen bond donor, this dihedral distribution might be a possible explanation: If it is not represented correctly by the force field in the MD simulations, the MC results with one (optimized) structure clearly indicate that this would strongly affect the RDCs.

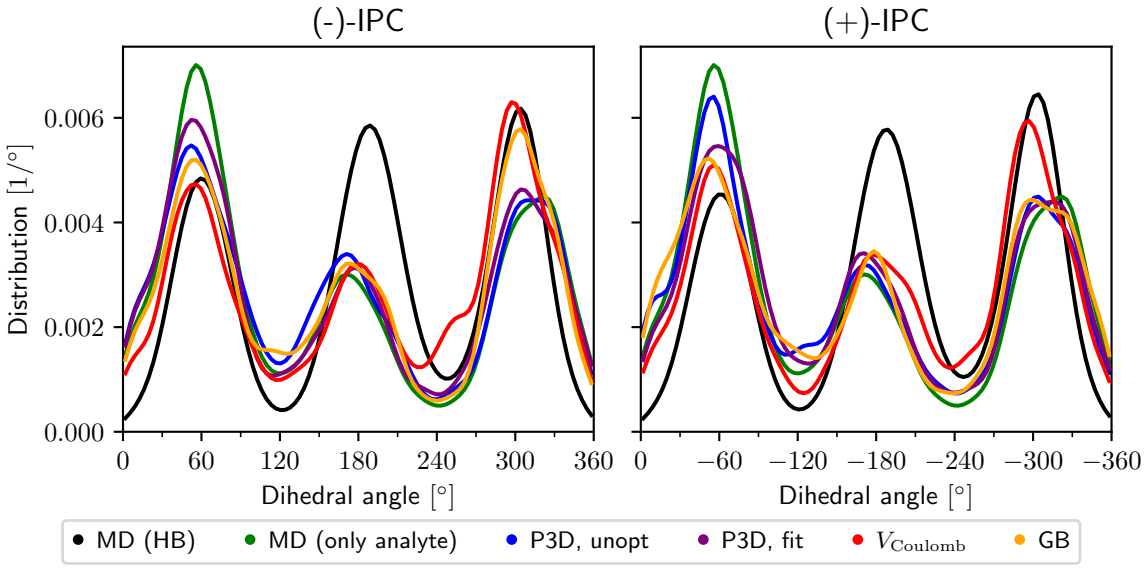


Figure 4.32.: Distribution of the C – C – O – H dihedral angle for (-)- and (+)-IPC. In black, the distribution of the poses with hydrogen bond obtained from the MD simulations ("MD (HB)") is shown, and in green, the distribution from the 1000 analyte structures obtained from MD simulations in chloroform ("MD (only analyte)"). The distributions from the MC simulations for the different models are indicated by the other colors.

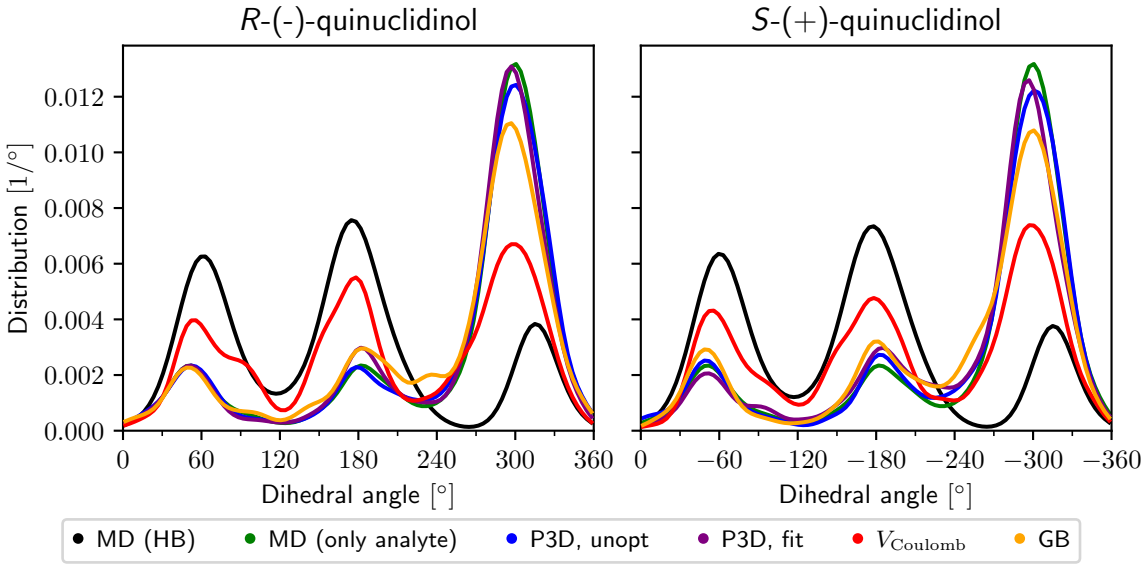


Figure 4.33.: Distribution of the C – C – O – H dihedral angle for *R*-(-)- and *S*-(+)-quinuclidinol. See Figure 4.32 for an explanation of the colors.

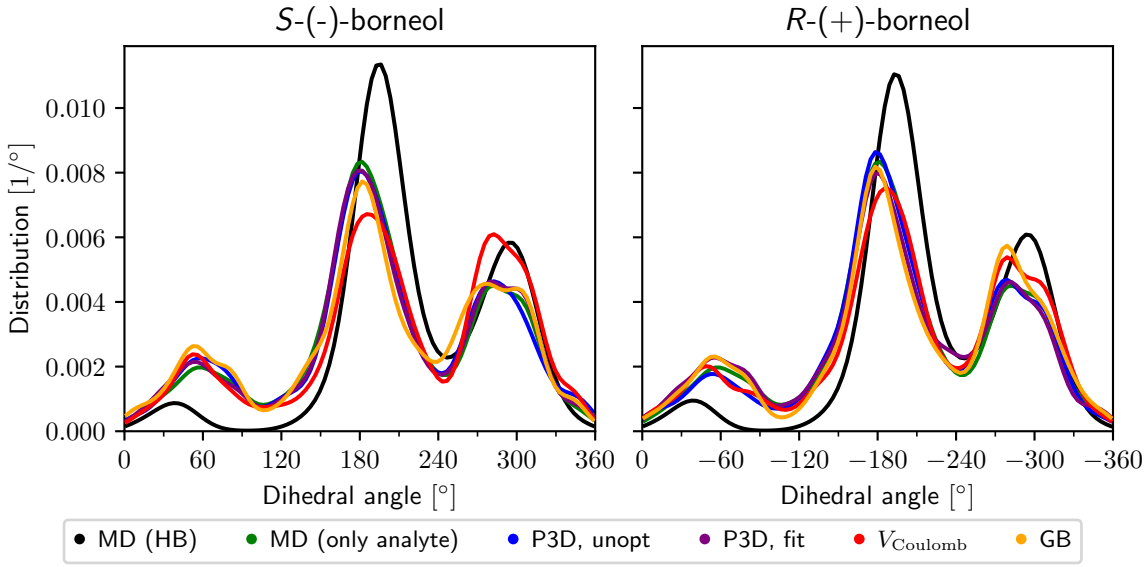


Figure 4.34.: Distribution of the C – C – O – H dihedral angle for *S*-(-)- and *R*-(+)-borneol. See Figure 4.32 for an explanation of the colors.

4.3.5. Residual dipolar couplings and alignment tensors

The RDCs calculated from the different models used in the MC calculations are compared in Figures 4.35 to 4.40 by the cosine similarity. Additionally, they are compared to the experimental RDCs and the RDCs obtained from the MD simulations (for the molecules with which I did the simulations).

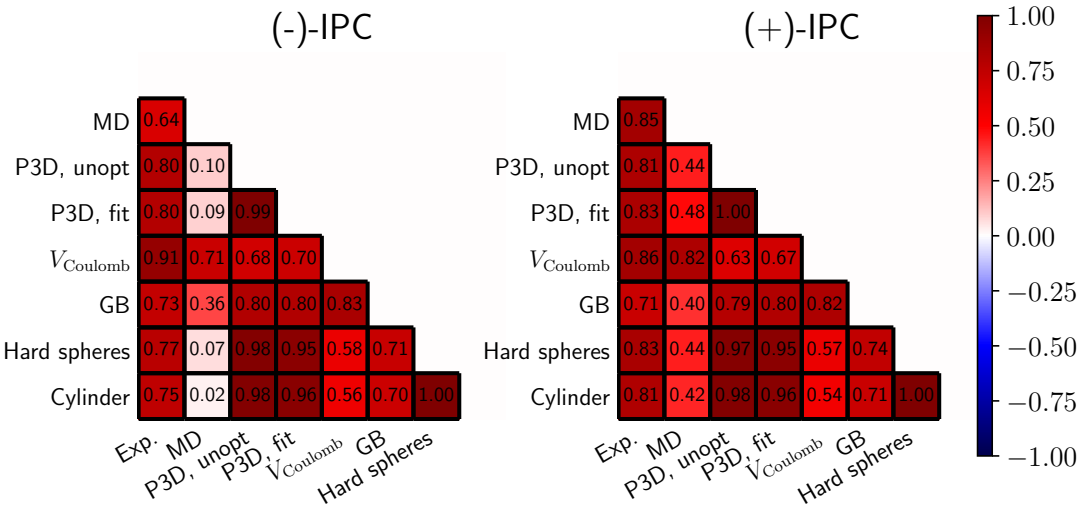


Figure 4.35.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for (-)- and (+)-IPC.

Overall, the best agreement for the molecules with polar groups (i.e. all tested molecules except for β-pinene) compared to experiment is for "V_{Coulomb}". Only for (-)-caulamidine A, the agreement is significantly lower than with the "P3D, unopt", method, but this is likely because PBLG structure for the averaging was not chosen long enough, as shown above in Table 4.4. Especially interesting are the results for quinuclidinol and borneol, where the agreement with experiment is very low in MD, but much higher for "V_{Coulomb}". A possible explanation is that the hydrogen bonds are not represented correctly in MD (which is likely considering they are only described by point charges), while the combination of V_{Coulomb} and the modeling of the atoms as hard spheres does not lead to very close H $\cdots$ O distances like in the MD simulations.

Even though camphor does not form hydrogen bonds with PBLG, the electrostatic interactions with PBLG seem essential for the alignment: As shown in Figure 4.38, the cosine similarity of the RDCs from "V_{Coulomb}" compared to experiment is 0.92 and 0.89 for the two enantiomers, second only for the MD simulations with 0.98. For "GB", where significant hydrogen bonds are still present, the agreement is 0.83 and 0.71, respectively. The other models with weaker electrostatic interactions differ significantly more. Especially the "hard spheres" (and "cylinder") model has very low agreement, which shows that the alignment of camphor is again not explained by the geometry of PBLG, but by the electrostatic interactions.

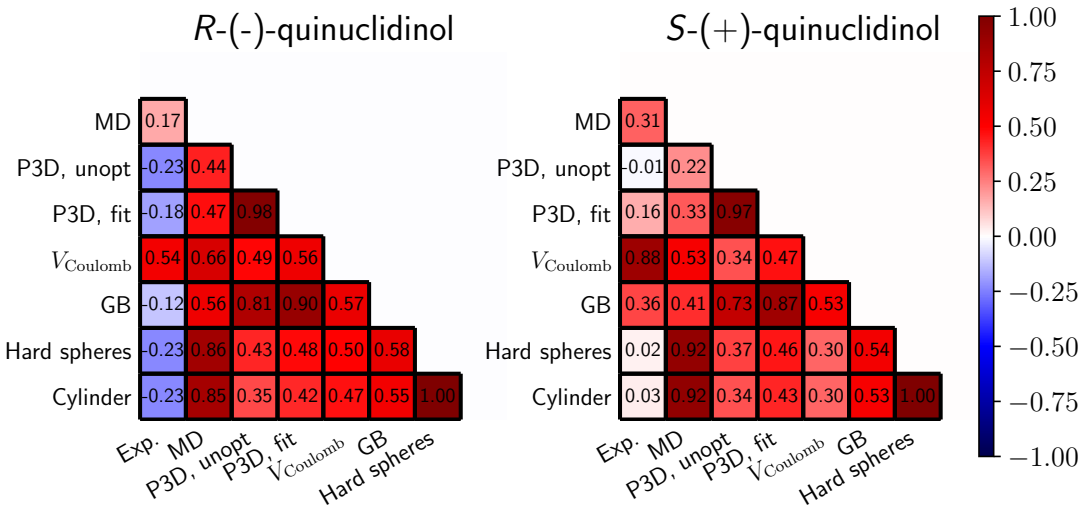


Figure 4.36.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for *R*-(-)- and *S*-(+)-quinuclidinol.

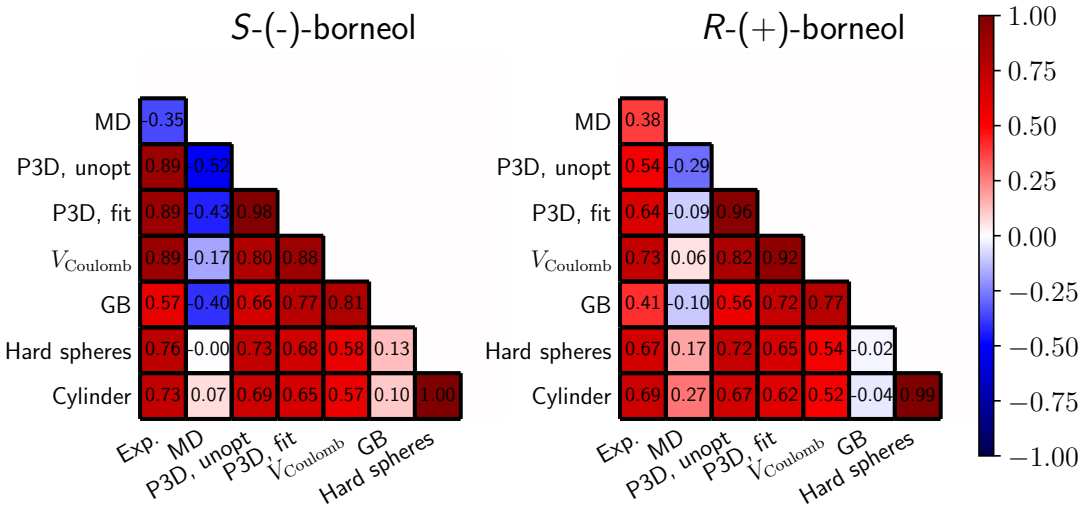


Figure 4.37.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for *S*-(-)- and *R*-(+)-borneol.

For β-pinene, the situation is different: As discussed above (see Section 4.3.1), it has a small dipole moment and consists only of carbon and hydrogen atoms, the importance of the electrostatic interactions with PBLG is much lower than for the other molecules. Here, the steric interaction with PBLG can be expected to be the main factor for the alignment. Indeed, comparing the experimental RDCs in Figure 4.39 with the steric model confirms this, with a cosine similarity of 0.94 and 0.91 for (-)- and (+)-β-pinene. Unlike for the other molecules with large dipole moment, the inclusion of electrostatic interactions strongly decreases the correlation with the experimental RDCs, although the center of mass distribution does not change significantly.

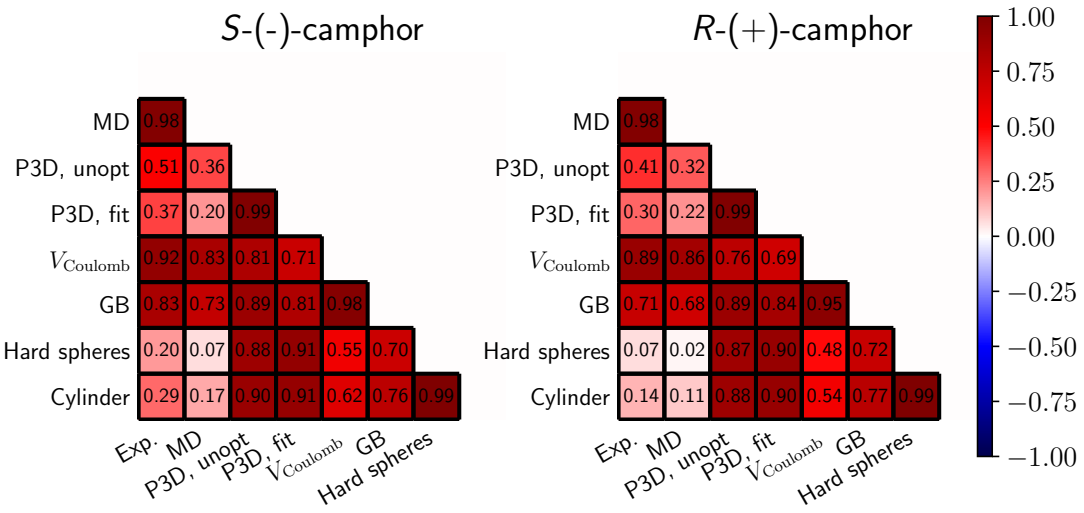


Figure 4.38.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for *S*-(-)- and *R*-(+)-camphor.

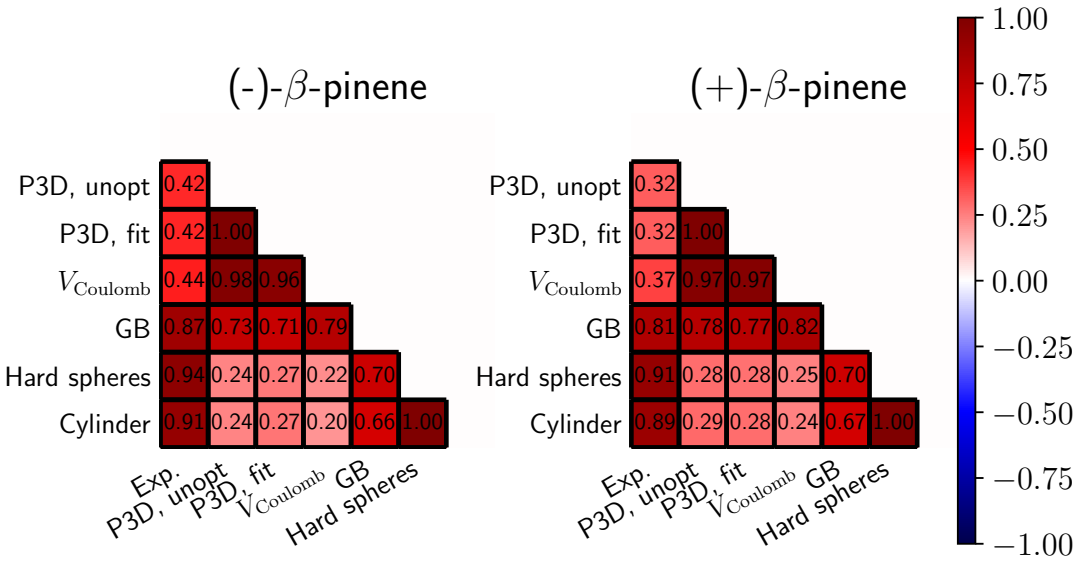


Figure 4.39.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for (-)- and (+)-β-pinene.

Also notable is the high agreement between the "hard spheres" model and the "cylinder" model: This means that the average alignment only due to the geometry of the PBLG sidechains can be described very well be replacing PBLG with a cylinder. As the alignment by a cylinder is achiral, this shows that the PBLG geometry is not sufficient for enantiodiscrimination and specific interactions (i.e. hydrogen bonds) near the chiral center of PBLG are required for the differentiation of enantiomers.

To see whether the MC simulations might be suitable for a prediction of the absolute configuration, I compare the experimental and calculated RDCs between enantiomers in

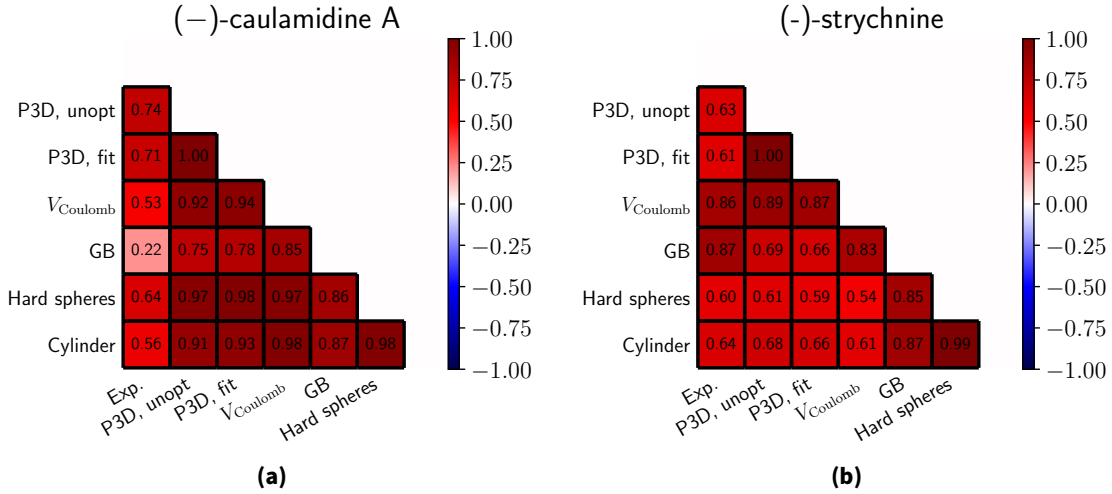


Figure 4.40.: Cosine similarities between the RDCs obtained from experiment, the MD simulations and the different models used in the MC calculations for **(a)** caulamidine A and **(b)** (-)-strychnine.

Table 4.6.: Comparison of the RDCs between the enantiomers of IPC, borneol, camphor, quinuclidinol and β-pinene for the experimental, MD and MC results. For each source of the RDCs (simulation or experiment) and each molecule, the cosine similarity between its enantiomers is shown.

Molecule	Exp.	MD	P3D, unopt	P3D, fit	V_{Coulomb}	GB	Hard spheres
IPC	0.99	0.77	0.99	1.00	0.99	0.99	1.00
Borneol	0.67	0.93	1.00	0.99	1.00	0.99	1.00
Camphor	0.99	1.00	1.00	1.00	0.99	1.00	1.00
Quinuclidinol	0.76	0.78	0.99	0.99	0.97	0.97	1.00
β-pinene	0.97		1.00	1.00	1.00	1.00	1.00

Table 4.6. Unlike for the MD simulations, where there are significant differences for the molecules with hydrogen bond donor, the cosine similarity for the MC calculations is ≥ 0.97 for all models. Thus, the differences are not large enough for a reliable differentiation. For camphor and β-pinene, the RDCs are almost identical in all tested cases as expected.

This poses the question why the differences between enantiomers are so small for the MC calculations, considering that the binding with "V_{Coulomb}" is much stronger than in the MD simulations. Therefore, I also calculated the RDCs from the MD simulations taking only the poses where atoms of the analyte and PBLG do not overlap, with the same atom radii as for the MC calculations. For the determination of the "overlap", I chose the same definition as for the hard spheres in the MC calculations, with atom radii $0.5\sigma_i$. The results are shown in Table 4.7. For camphor, this average is almost identical to the full MD average, but for the other molecules, the agreement is low. Additionally, the RDCs between enantiomers (see Table 4.7b) have a cosine similarity ≥ 0.99 just like for the MC calculations. This result is both somewhat surprising and important for explaining the enantiodiscrimination by PBLG. It means that only the alignment due to close interactions very close contact to PBLG shows significant chiral behavior, while the alignment solely due to geometry and

Table 4.7.: (a) Cosine similarity between the RDCs from MD with the poses removed where the analyte and PBLG molecules overlap (within $0.5(\sigma_i + \sigma_j)$) and the RDCs calculated from MD using the full average (S_{MD}), as well as the experimental RDCs (S_{exp}). **(b)** Cosine similarity between the RDCs of the two enantiomers calculated from the poses in the MD simulations, where the analyte and PBLG molecules do not overlap.

(a)

Molecule	S_{MD}	S_{exp}
(-)-IPC	0.21	0.86
(+)-IPC	0.68	0.94
R-(-)-quinuclidinol	0.42	-0.11
S-(+)-quinuclidinol	0.24	-0.14
S-(-)-borneol	-0.19	0.85
R-(+)-borneol	-0.01	0.67
S-(-)-camphor	0.99	0.99
R-(+)-camphor	0.99	0.98

(b)

Molecule	S
IPC	1.00
Quinuclidinol	0.99
Borneol	1.00
Camphor	1.00

the electrostatic interactions does not show a significantly difference between enantiomer. These close contacts are only made possible by the hydrogen bonds to PBLG, which, as shown in Figures 4.12 and A.5, determine most of the alignment. As mentioned above, the hydrogens attached to the O – H hydrogen bond donor are not interacting with the other atoms by the LJ potential. This means that the hydrogen bonds to PBLG are not considered as overlap between the hydrogen and oxygen atoms, except if the oxygen atoms overlap (which occurs frequently, but not for all poses with hydrogen bond).

Also shown in Table 4.7a is the cosine similarity between the RDCs calculated from MD using only the poses without overlap and to the experimental RDCs. Interestingly, the agreement with experiment is better for IPC and borneol, while it decreased for quinuclidinol. This is unlike the MC calculations with "$V_{Coulomb}$", where the agreement of the RDCs with experiment is significantly better than for the MD results. For camphor, the RDCs and thus the agreement with experiment are almost unchanged, which is not surprising considering the mostly electrostatic and steric nature of the alignment due to the lack of hydrogen bond donor.

On the other hand, the RDCs calculated from poses *with* overlap between PBLG and the analyte are compared in Table A.8. For the molecules with hydrogen bond donor, they unsurprisingly agree almost perfectly with the poses with hydrogen bond (and thus with the full MD average), because most poses with hydrogen bond have an overlap with PBLG. For camphor however, there are larger discrepancies, as these poses are not important for the overall alignment.

4.3.6. Residual dipolar couplings of hydrogen bond poses

Like for the results from the MD simulations in Figure 4.12, I also calculated the RDCs using only the poses where a hydrogen bond to PBLG is present. These RDCs of the

poses with hydrogen bond for the "P3D, unopt", "P3D, fit", "V_{Coulomb}" and "GB" models are compared among each other, to the MD results using only poses with hydrogen bonds and the experimental RDCs in Figures 4.41 to 4.43 for IPC, quinuclidinol and borneol. Notable is that the different models used in MC agree very well with each other, especially "P3D, fit", "V_{Coulomb}" and "GB". The lowest agreement is for "P3D, unopt" and "V_{Coulomb}", which is expected considering the polar binding free energies $V_{\text{Coulomb}} + \Delta\Delta G_{\text{polar}}$ for the poses with hydrogen bond in Figure 4.25 are very different between the two models.

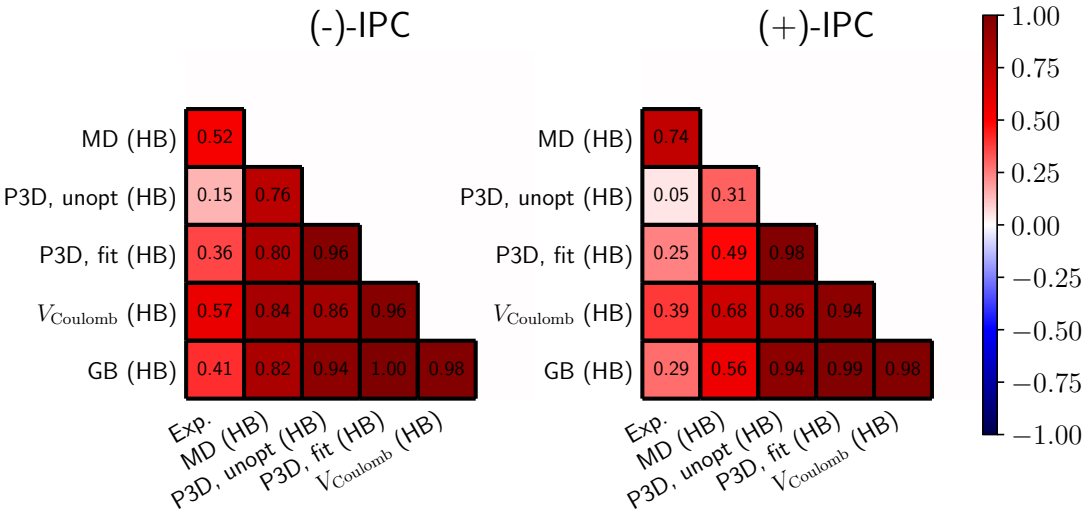


Figure 4.41.: Cosine similarities between the RDCs obtained from experiment and the poses with hydrogen bond from the MD simulations and the different models used in the MC calculations for (-)- and (+)-IPC.

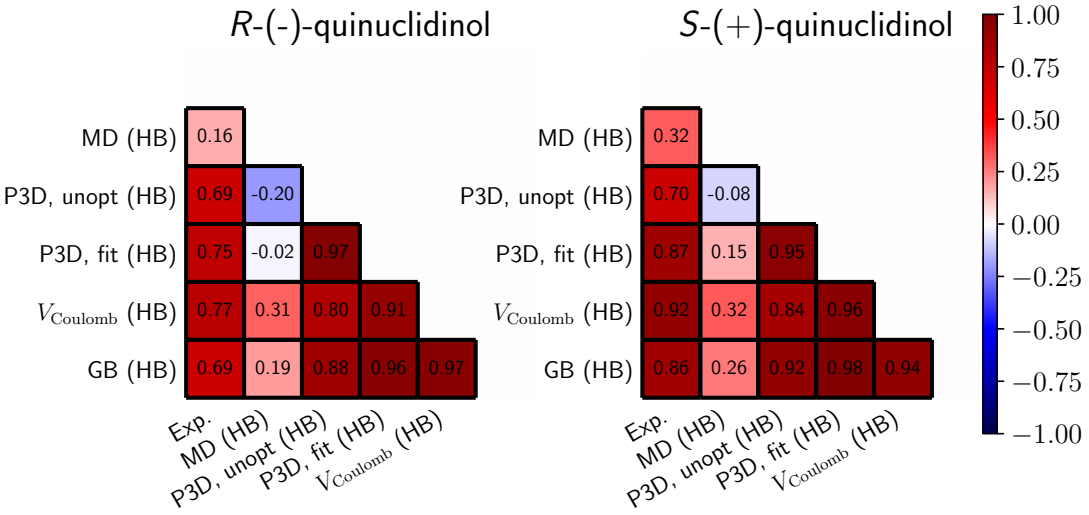


Figure 4.42.: Cosine similarities between the RDCs obtained from experiment and the poses with hydrogen bond from the MD simulations and the different models used in the MC calculations for *R*-(-)- and *S*-(+)-quinuclidinol.

Interestingly, the agreement with the hydrogen bonds from the MD simulations is low for quinuclidinol and borneol, but the agreement with experiment is much better, while it is

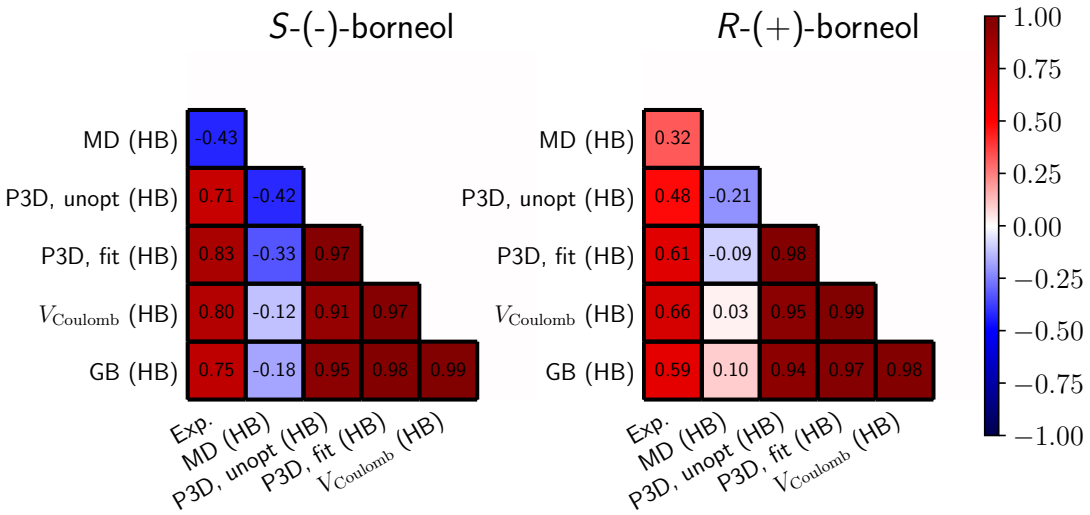


Figure 4.43.: Cosine similarities between the RDCs obtained from experiment and the poses with hydrogen bond from the MD simulations and the different models used in the MC calculations for *S*-(-)- and *R*-(+)-borneol.

Table 4.8.: Cosine similarities of the RDCs obtained from the full average and only the poses with hydrogen bond of the different MC calculations for IPC, quinuclidinol and borneol.

Molecule	P3D, unopt	P3D, fit	V_{Coulomb}	GB
(-)-IPC	-0.27	-0.05	0.80	0.44
(+)-IPC	-0.32	-0.14	0.74	0.40
R-(-)-quinuclidinol	-0.59	-0.32	0.75	0.19
S-(+)-quinuclidinol	-0.62	-0.24	0.75	0.26
S-(-)-borneol	0.63	0.80	0.98	0.90
R-(+)-borneol	0.63	0.83	0.98	0.90

low for the MD averages. This indicates that the interactions for the poses with hydrogen bonds are not represented correctly in the MD simulations. A possible reason is that with the Coulomb potential using point charges and the Lennard-Jones potential between the acceptor atom in PBLG and hydrogen atom in the analytes, the strength of the hydrogen bonds is overestimated. In the MC simulations, they are much stronger with the "V_{Coulomb}" model, their distance is however limited due to the steric exclusion by hard spheres.

In Table 4.8, the cosine similarities for the different methods are compared to the respective full average. As the hydrogen bonds are the strongest for "V_{Coulomb}" (see Figures 4.29 to 4.31), the agreement with the full average is likewise the highest. Accordingly, the agreement goes down as the strength of the hydrogen bonds decreases, with the lowest agreement for the "P3D, unopt" model. Even though the hydrogen bonds are the strongest for quinuclidinol of the three molecules (also in the MD simulations), the agreement with the full average is the lowest. A possible reason could be that the chosen definition (an

angle $\geq 150^\circ$ and a distance $\leq 3\,\text{Å}$) of whether a hydrogen bond is formed does not agree with the actual distribution in the simulation, which is considerably more narrow.

Thus, the results of the MC calculations show that the most likely reason for the low agreement of the MD results for some molecules with experiment is indeed the force field inaccuracy and not an unknown interaction in experiment. Additionally, the precise geometry of the hydrogen bond does not seem as important as the position of the hydrogen atom in the O – H hydrogen bond donors. For the distinction of enantiomers however, the hard spheres approach I used for the MC calculations does not appear suitable, as it does not result in significant differences of the RDCs between enantiomers. A more accurate implicit solvation model would require the description of the balance between the attractive dispersion interactions between analyte and PBLG, and the nonpolar solvation. For the latter, the tested model turned out not to be sufficiently accurate for calculating meaningful averages, even though I fitted the model parameters carefully to MD data. It is also still unclear whether the size of the PBLG chain is still large enough that a single PBLG structure would suffice for the integral if an exact F_{binding} were used, although this was the case for all tested methods, with the exception of "V_{Coulomb}" for two molecule, because the electrostatic interactions are strongly overestimated. Otherwise, the same approach like for multiple analyte molecules could be used, although using MD simulations might be both easier, more reliable and maybe not much more computationally expensive.

5. Summary and Outlook

In this thesis, I studied the interactions of small analyte molecules with PBLG in chloroform and used both molecular dynamics simulations and Monte Carlo calculations to predict residual dipolar couplings. For this, the MD simulations were performed with explicit chloroform solvent for both enantiomers of isopinocampheol, quinuclidinol, borneol, camphor, indolophenanthridine, ibuprofen and 1-phenylethanol. To reach sufficient convergence of the MD simulations, I used an automated workflow to run many parallel simulations, reaching a total simulation time of 100 µs or 30 µs, depending on the molecule. Analyzing the interactions between PBLG and the analyte molecules showed that there is a fundamental difference between the analyte molecules with hydrogen bond donors and those without: In the former, the alignment is mainly determined by poses where a hydrogen bond to PBLG is formed, while the latter mostly interact with the outer parts of the PBLG sidechain. For the molecules with hydrogen bond donor, there are significant differences between the RDCs both in experiment and simulation. My MD results also showed that comparing this difference in alignment ("CDDA") between the MD results and the measured RDCs results in the correct assignment for all six tested molecules. This means that given a chiral analyte molecule with hydrogen bond donor, its RDCs can be measured both with PBLG and PBDG (the other enantiomer of PBLG) and the difference can be compared with the difference from MD simulations to obtain the absolute configuration of the analyte. Additional simulations with eight analyte molecules and water molecules showed only small differences both of the RDCs directly (up to a constant scaling factor) and of the differences between enantiomers (for moderate amounts of water), which indicates that the approach is indeed relatively robust. I also confirmed the approach of modeling PBLG as a single α-helix, as there are no large difference in the RDCs if two PBLG chains are simulated.

To understand the interactions leading to the binding of the analyte molecules to PBLG, I calculated binding free energies and their contributions of various poses near PBLG using MD simulations. For the nonpolar solvation contributions, the calculations were done with two different chargeless test molecules. Additionally, I calculated total binding free energies of two test molecules with hydrogen bond donor. Comparing the contributions, the polar and nonpolar parts turned out to be of similar importance. Here, the polar solvation is mainly determined by the Coulomb potential between the test molecule and PBLG, while for the nonpolar solvation, there is a fine balance between the Lennard-Jones potential between them and the change of the nonpolar solvation free energy. Using these free energies, I fitted the parameters of both nonpolar and polar solvation models, which I adapted for the calculation of binding free energies. Here, the total Coulomb potential energy between the analytes (V_{Coulomb}) and the polar solvation turned out to be reasonably

accurate. The same can be said about the individual contributions to the nonpolar binding free energies, which however partially counteract each other. To test the feasibility of calculating RDCs with these models, I implemented all interactions in a way which is suitable for the computationally efficient application in Monte Carlo calculations.

Inspired by previous studies calculating RDCs with rigid molecules (PALES and P3D), where the poses are weighted by the Boltzmann factor calculated from electrostatic interactions, I did Monte Carlo calculations for in total 14 molecules (the first four pairs of enantiomers used in the MD simulations, as well as (-)-caulamidine A, (-)-strychnine and both enantiomers of β-pinene). Here, population annealing proved to be a suitable method for generating samples from which RDCs and other properties can be calculated, even in the presence of strong binding (free) energy minima. I performed these calculations using different models for the polar interactions between the analyte molecules and PBLG, some of which were developed specifically in this thesis for the purpose of the consideration of the polar solvation by implicit chloroform. Because I found the nonpolar solvation not to be accurate enough to be usable for the calculation of RDCs, I treated the atoms like hard spheres (as in the P3D model). For the O – H hydrogen bond molecule, calculating the average over many analyte structure was necessary, because the position of the O – H hydrogen atom strongly influences the alignment. Unexpectedly, using only "V_{Coulomb}" to describe the binding (free) energies F_{binding} showed the best agreement with experimental RDCs. The only major exception was β-pinene, where the agreement was the best when only calculating the alignment due to steric exclusions by atoms modeled as uncharged hard spheres (which gives almost identical results as the PALES ("cylinder") model), due to its nonpolar nature.

Despite the suitability of determining the absolute configuration using the results from the MD simulations for the molecules with hydrogen bond donor, there are still important limitations for the calculation of RDCs: While the RDCs calculated from the MD simulations showed very good agreement with experiment for camphor and HMIP, the agreement with experiment for the remaining molecules with O – H hydrogen bond donor is low. Although the MC calculations with "V_{Coulomb}" showed better agreement of the RDCs, this is not because a more accurate model was used. It is likely that the poses with very close contact to PBLG, which are responsible for the differences between enantiomers, are either not described accurately enough in the MD simulations or their relative importance compared to the other poses during the MD simulation is not correctly represented. The most obvious explanation is, naturally, the hydrogen bonds to PBLG: They are likely not modeled accurately because the electrostatic interactions are only described by point charges in the simulation.

Even though in the MC approach, where I modeled the atoms as hard spheres and otherwise only the electrostatic interactions, hydrogen bonds are formed and even mainly determine the alignment for "V_{Coulomb}", the difference between enantiomers is very small. This is further confirmed by considering only the poses from the MD simulations where the atoms modeled as hard spheres do not overlap between the analyte and PBLG: Like for the MD models, the difference between the enantiomers is very small. Therefore, we can conclude that the enantiodiscrimination occurs essentially only during the formation of a

hydrogen bond with PBLG, while the rest of the molecule is located between the PBLG sidechains and interacts with them very closely. Here, the actual chiral behavior does not however occur due to the hydrogen bond itself, but rather the chiral arrangement of the neighboring PBLG sidechains, with which the analyte molecule is in simultaneous close contact a substantial amount of time. Remarkably, this works well even in molecules such as HMIP, where the hydrogen bond donor is not close to the chiral center. Nonetheless, the correlation of the CDDA is clearly positive in both experiment and simulation.

To improve the MC calculation, a replacement of the hard spheres approach appears to be necessary. Instead, the dispersion interaction with both PBLG and the chloroform solvent need to be included in F_{binding}. However, the approach of splitting the nonpolar solvation into a repulsive part based on the SASA and SAV and an attractive part based on treating the solvent as a continuum proved not to be accurate enough: There were still incorrect deep minima of F_{binding} present and in the fitting procedure, the values of F_{binding} obtained from MD could only be reproduced with significant error by the implicit solvation model. The tested polar solvation model is by itself not much more accurate in terms of the correlation to the MD data. However, its contribution to F_{binding} is relatively small and together with the Coulomb interaction between the analyte molecule and PBLG, it is more accurate than the nonpolar contribution. Additionally, the effect of the polar solvation model is purely repulsive and together with hard spheres does not show deep minima. Thus, the only way forward to improve the MC calculations is the (significant) improvement of the nonpolar solvation model.

Regarding the discrepancy of the RDCs calculated from the MD simulations and the measured RDCs for the molecules with O – H donor, its origin is still unclear, although several possible reasons can be likely excluded. A possible imperfect alignment of PBLG in experiment would only lead to a scaling of the RDCs, and I demonstrated that these differences can most likely not be explained by multiple analyte molecules or water. Additionally, as the MC calculations showed better agreement by treating the atoms as hard spheres and modeling the interactions between PBLG and the analyte molecules only by the Coulomb potential, this indicates that there likely are no major unknown effects in experiment which are not accounted for in the simulations. Thus, even though the agreement of the MC calculations with experimental RDCs is currently better (ignoring the scaling) for the molecules with O – H hydrogen bond donor, future improvements of the prediction with experiment should probably focus on improving the MD simulations with explicit solvent, as they can treat the balance of the dispersion interactions more accurately and can already distinguish enantiomers in the current form. To improve the MD simulations, increasing the force field accuracy is likely essential, especially the description of hydrogen bonds to PBLG. This is further supported by the MC calculations, where the rotation of the hydrogen atom in the O – H donors proved to be very important for the alignment. For improving the accuracy, the main challenge is the computational cost, as total simulation times of many µs are necessary due to the weak alignment. This limits the applicability of QM/MM or even ML/MM approaches. The next step would therefore be simulations with more accurate classical force fields, such as applying a correction for the hydrogen bonds or using a polarizable force field. This will likely require further parametrization and validation of force field parameters.

As was previously demonstrated experimentally and what I confirmed with the MD simulations, the enantiodiscrimination using PBLG as alignment medium is sufficiently strong for analyte molecules with (accessible) hydrogen bond donor. This is not the case for other molecules however, where the alignment of enantiomers is almost identical. Thus, different chiral alignment media would be required, for which a specific interaction with the analyte close to a chiral part of the alignment medium is probably necessary.

Overall, I showed in my thesis that the prediction of RDCs up to a constant scaling factor is generally possible, even with MC calculations without taking solvent effects into account. To increase the agreement with experiment using implicit solvation models however, large improvements of the nonpolar solvation would be required. With MD simulations and explicit solvent, large discrepancies remain for molecules with O – H hydrogen bond donor, for which force field inaccuracies are the main suspected cause. The difference between enantiomers is nonetheless represented sufficiently accurately in MD to correctly determine the abolute configuration by a comparison with experimental differences for the molecules forming a hydrogen bond with PBLG.

Bibliography

[1] *alchemlyb: the simple alchemistry library, version 2.1.0*. https://github.com/alchemistry/alchemlyb. [Online; accessed 22-June-2023].

[2] Allen, M. and Tildesley, D. *Computer Simulation of Liquids*. Computer Simulation of Liquids. Clarendon Press, 1989. ISBN: 9780198556459.

[3] Almond, A. and Axelsen, J. B. "Physical Interpretation of Residual Dipolar Couplings in Neutral Aligned Media". In: *Journal of the American Chemical Society* 124.34 (Aug. 2002), pp. 9986–9987. DOI: 10.1021/ja026876i. URL: http://dx.doi.org/10.1021/ja026876i.

[4] Arnott, S. and Dover, S. D. "Refinement of bond angles of an α-helix". In: *Journal of Molecular Biology* 30.1 (Nov. 1967), pp. 209–212. DOI: 10.1016/0022-2836(67)90253-7. URL: http://dx.doi.org/10.1016/0022-2836(67)90253-7.

[5] Aroulanda, C. and Lesot, P. "Molecular enantiodiscrimination by NMR spectroscopy in chiral oriented systems: Concept, tools, and applications". In: *Chirality* 34.2 (Dec. 2021), pp. 182–244. DOI: 10.1002/chir.23386. URL: http://dx.doi.org/10.1002/chir.23386.

[6] Aroulanda, C. et al. "Investigation of the enantioselectivity of three polypeptide liquid-crystalline solvents using NMR spectroscopy." In: *Enantiomer* 6.5 (2001), pp. 281–287.

[7] Azurmendi, H. F. and Bush, C. A. "Tracking Alignment from the Moment of Inertia Tensor (TRAMITE) of Biomolecules in Neutral Dilute Liquid Crystal Solutions". In: *Journal of the American Chemical Society* 124.11 (Feb. 2002), pp. 2426–2427. DOI: 10.1021/ja017524z. URL: http://dx.doi.org/10.1021/ja017524z.

[8] Bain, A. D. and Burton, I. W. "Quadrature detection in one or more dimensions". In: *Concepts in Magnetic Resonance* 8.3 (1996), pp. 191–204. DOI: 10.1002/(sici)1099-0534(1996)8:3<191::aid-cmr3>3.0.co;2-u. URL: http://dx.doi.org/10.1002/(sici)1099-0534(1996)8:3%3C191::aid-cmr3%3E3.0.co;2-u.

[9] Bamford, C. H., Hanby, W., and Happey, F. "The structure of synthetic polypeptides I. X-ray investigation". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 205.1080 (Jan. 1951), pp. 30–47. DOI: 10.1098/rspa.1951.0015. URL: http://dx.doi.org/10.1098/rspa.1951.0015.

[10] Bax, A. "Weak alignment offers new NMR opportunities to study protein structure and dynamics". In: *Protein Science* 12.1 (Jan. 2003), pp. 1–16. DOI: 10.1110/ps.0233303. URL: http://dx.doi.org/10.1110/ps.0233303.

[11] Bayly, C. I. et al. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model". In: *The Journal of Physical Chemistry* 97.40 (Oct. 1993), pp. 10269–10280. DOI: 10.1021/j100142a004. URL: http://dx.doi.org/10.1021/j100142a004.

[12] Bennett, C. H. "Efficient estimation of free energy differences from Monte Carlo data". In: *Journal of Computational Physics* 22.2 (Oct. 1976), pp. 245–268. DOI: 10.1016/0021-9991(76)90078-4. URL: http://dx.doi.org/10.1016/0021-9991(76)90078-4.

[13] Berendsen, H. J. C. et al. "Molecular dynamics with coupling to an external bath". In: *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. DOI: 10.1063/1.448118. URL: http://dx.doi.org/10.1063/1.448118.

[14] Berger, R. et al. "Is Enantiomer Assignment Possible by NMR Spectroscopy Using Residual Dipolar Couplings from Chiral Nonracemic Alignment Media?—A Critical Assessment". In: *Angewandte Chemie International Edition* 51.33 (July 2012), pp. 8388–8391. DOI: 10.1002/anie.201107626. URL: http://dx.doi.org/10.1002/anie.201107626.

[15] Bernetti, M. and Bussi, G. "Pressure control using stochastic cell rescaling". In: *The Journal of Chemical Physics* 153.11 (Sept. 2020). DOI: 10.1063/5.0020514. URL: http://dx.doi.org/10.1063/5.0020514.

[16] Berova, N., Nakanishi, K., and Woody, R. W. *Circular dichroism: principles and applications*. John Wiley & Sons, 2000. ISBN: 9780471330035.

[17] Berthelot, D. "Sur le mélange des gaz". In: *Comptes rendus de l'Académie des Sciences* 126.3 (1898), p. 15.

[18] Beutler, T. C. et al. "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations". In: *Chemical Physics Letters* 222.6 (June 1994), pp. 529–539. DOI: 10.1016/0009-2614(94)00397-1. URL: http://dx.doi.org/10.1016/0009-2614(94)00397-1.

[19] Bijvoet, J. M., Peerdeman, A. F., and van Bommel, A. J. "Determination of the Absolute Configuration of Optically Active Compounds by Means of X-Rays". In: *Nature* 168.4268 (Aug. 1951), pp. 271–272. DOI: 10.1038/168271a0. URL: http://dx.doi.org/10.1038/168271a0.

[20] Blumofe, R. D. et al. "Cilk: an efficient multithreaded runtime system". In: *ACM SIGPLAN Notices* 30.8 (Aug. 1995), pp. 207–216. DOI: 10.1145/209937.209958. URL: http://dx.doi.org/10.1145/209937.209958.

[21] Böttcher, B. and Thiele, C. M. "Determining the Stereochemistry of Molecules from Residual Dipolar Couplings (RDCs)". In: *eMagRes*. Vol. 1. John Wiley & Sons, Ltd, Mar. 2012, pp. 169–180. DOI: 10.1002/9780470034590.emrstm1194. URL: http://dx.doi.org/10.1002/9780470034590.emrstm1194.

[22] Brunner, E. "Residual dipolar couplings in protein NMR". In: *Concepts in Magnetic Resonance* 13.4 (Jan. 2001), pp. 238–259. DOI: 10.1002/cmr.1012. URL: http://dx.doi.org/10.1002/cmr.1012.

[23] Burkardt, J. *SPHERE_LEBEDEV_RULE Quadrature Rules for the Sphere*. Accessed: 2023-01-05. URL: https://people.sc.fsu.edu/~jburkardt/datasets/sphere_lebedev_rule/sphere_lebedev_rule.html.

[24] Bussi, G., Donadio, D., and Parrinello, M. "Canonical sampling through velocity rescaling". In: *The Journal of Chemical Physics* 126.1 (Jan. 2007), p. 014101. DOI: 10.1063/1.2408420. URL: http://dx.doi.org/10.1063/1.2408420.

[25] Canet, I. et al. "Enantiomeric analysis in a polypeptide lyotropic liquid crystal by deuterium NMR". In: *Journal of the American Chemical Society* 117.24 (June 1995), pp. 6520–6526. DOI: 10.1021/ja00129a015. URL: http://dx.doi.org/10.1021/ja00129a015.

[26] Canlet, C. et al. "Deuterium NMR stereochemical analysis of threo–erythro isomers bearing remote stereogenic centres in racemic and non-racemic liquid crystalline solvents". In: *Tetrahedron: Asymmetry* 11.9 (May 2000), pp. 1911–1918. DOI: 10.1016/s0957-4166(00)00125-7. URL: http://dx.doi.org/10.1016/s0957-4166(00)00125-7.

[27] Case, D. A. et al. *Amber 2021*. University of California, San Francisco, 2021.

[28] Channell, P. J. and Scovel, C. "Symplectic integration of Hamiltonian systems". In: *Nonlinearity* 3.2 (May 1990), pp. 231–259. DOI: 10.1088/0951-7715/3/2/001. URL: http://dx.doi.org/10.1088/0951-7715/3/2/001.

[29] Chodera, J. D. "A Simple Method for Automated Equilibration Detection in Molecular Simulations". In: *Journal of Chemical Theory and Computation* 12.4 (Mar. 2016), pp. 1799–1805. DOI: 10.1021/acs.jctc.5b00784. URL: http://dx.doi.org/10.1021/acs.jctc.5b00784.

[30] Christiansen, H., Weigel, M., and Janke, W. "Accelerating Molecular Dynamics Simulations with Population Annealing". In: *Physical Review Letters* 122.6 (Feb. 2019). DOI: 10.1103/physrevlett.122.060602. URL: http://dx.doi.org/10.1103/physrevlett.122.060602.

[31] Cianciosi, S. J. et al. "Synthesis and gas-phase vibrational circular dichroism of (+)-(S,S)-cyclopropane-1,2-2H2". In: *Journal of the American Chemical Society* 111.5 (Mar. 1989), pp. 1913–1915. DOI: 10.1021/ja00187a074. URL: http://dx.doi.org/10.1021/ja00187a074.

[32] Cieplak, P. et al. "Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins". In: *Journal of Computational Chemistry* 16.11 (Nov. 1995), pp. 1357–1377. DOI: 10.1002/jcc.540161106. URL: http://dx.doi.org/10.1002/jcc.540161106.

[33] Cornell, W. D. et al. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *Journal of the American Chemical Society* 117.19 (May 1995), pp. 5179–5197. DOI: 10.1021/ja00124a002. URL: http://dx.doi.org/10.1021/ja00124a002.

[34] Cornell, W. D. et al. “Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation”. In: *Journal of the American Chemical Society* 115.21 (Oct. 1993), pp. 9620–9631. DOI: 10.1021/ja00074a030. URL: http://dx.doi.org/10.1021/ja00074a030.

[35] Courtès, L. and Wurmus, R. “Reproducible and User-Controlled Software Environments in HPC with Guix”. In: *Lecture Notes in Computer Science* (2015), pp. 579–591. DOI: 10.1007/978-3-319-27308-2_47. URL: http://dx.doi.org/10.1007/978-3-319-27308-2_47.

[36] Courtieu, J., Bayle, J., and Fung, B. “Variable angle sample spinning NMR in liquid crystals”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 26 (1994), pp. 141–169. DOI: 10.1016/0079-6565(94)80006-5. URL: http://dx.doi.org/10.1016/0079-6565(94)80006-5.

[37] Coutsias, E. A. and Wester, M. J. “RMSD and Symmetry”. In: *Journal of Computational Chemistry* 40.15 (Mar. 2019), pp. 1496–1508. DOI: 10.1002/jcc.25802. URL: http://dx.doi.org/10.1002/jcc.25802.

[38] Csizi, K.-S. and Reiher, M. “Universal QM/MM approaches for general nanoscale applications”. In: *WIREs Computational Molecular Science* 13.4 (Feb. 2023). DOI: 10.1002/wcms.1656. URL: http://dx.doi.org/10.1002/wcms.1656.

[39] Dai, N. and Jones, G. L. “Multivariate initial sequence estimators in Markov chain Monte Carlo”. In: *Journal of Multivariate Analysis* 159 (July 2017), pp. 184–199. DOI: 10.1016/j.jmva.2017.05.009. URL: http://dx.doi.org/10.1016/j.jmva.2017.05.009.

[40] Dale, J. A. and Mosher, H. S. “Nuclear magnetic resonance enantiomer regents. Configurational correlations via nuclear magnetic resonance chemical shifts of diastereomeric mandelate, O-methylmandelate, and α-methoxy-α-trifluoromethylphenylacetate (MTPA) esters”. In: *Journal of the American Chemical Society* 95.2 (Jan. 1973), pp. 512–519. DOI: 10.1021/ja00783a034. URL: http://dx.doi.org/10.1021/ja00783a034.

[41] Darden, T., York, D., and Pedersen, L. “Particle mesh Ewald: An $Nlog(N)$ method for Ewald sums in large systems”. In: *The Journal of Chemical Physics* 98.12 (June 1993), pp. 10089–10092. DOI: 10.1063/1.464397. URL: http://dx.doi.org/10.1063/1.464397.

[42] Ditchfield, R., Hehre, W. J., and Pople, J. A. “Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules”. In: *The Journal of Chemical Physics* 54.2 (Jan. 1971), pp. 724–728. DOI: 10.1063/1.1674902. URL: http://dx.doi.org/10.1063/1.1674902.

[43] Dosset, P. et al. “A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings”. In: *Journal of Biomolecular NMR* 20.3 (2001), pp. 223–231. DOI: 10.1023/a:1011206132740. URL: http://dx.doi.org/10.1023/a:1011206132740.

[44] Ehni, S. and Luy, B. "Robust INEPT and refocused INEPT transfer with compensation of a wide range of couplings, offsets, and B 1 -field inhomogeneities (COB3)". In: *Journal of Magnetic Resonance* 247 (Oct. 2014), pp. 111–117. DOI: 10.1016/j.jmr.2014.07.010. URL: http://dx.doi.org/10.1016/j.jmr.2014.07.010.

[45] Ekström, U. et al. "Arbitrary-Order Density Functional Response Theory from Automatic Differentiation". In: *Journal of Chemical Theory and Computation* 6.7 (May 2010), pp. 1971–1980. DOI: 10.1021/ct100117s. URL: http://dx.doi.org/10.1021/ct100117s.

[46] Enthart, A. et al. "The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings". In: *Journal of Magnetic Resonance* 192.2 (June 2008), pp. 314–322. DOI: 10.1016/j.jmr.2008.03.009. URL: http://dx.doi.org/10.1016/j.jmr.2008.03.009.

[47] Ewald, P. P. "Die Berechnung optischer und elektrostatischer Gitterpotentiale". In: *Annalen der Physik* 369.3 (Jan. 1921), pp. 253–287. DOI: 10.1002/andp.19213690304. URL: http://dx.doi.org/10.1002/andp.19213690304.

[48] Farahvash, A., Leontyev, I., and Stuchebrukhov, A. "Dynamic and Electronic Polarization Corrections to the Dielectric Constant of Water". In: *The Journal of Physical Chemistry A* 122.48 (Nov. 2018), pp. 9243–9250. DOI: 10.1021/acs.jpca.8b07953. URL: http://dx.doi.org/10.1021/acs.jpca.8b07953.

[49] Farjon, J. et al. "Enantiomeric excess measurements in weakly oriented chiral liquid crystal solvents through 2D 1H selective refocusing experiments". In: *Journal of Magnetic Resonance* 158.1-2 (Sept. 2002), pp. 169–172. DOI: 10.1016/s1090-7807(02)00070-8. URL: https://doi.org/10.1016%5C%2Fs1090-7807%5C%2802%5C%2900070-8.

[50] Flack, H. D. and Bernardinelli, G. "The use of X-ray crystallography to determine absolute configuration". In: *Chirality* 20.5 (2008), pp. 681–690. DOI: 10.1002/chir.20473. URL: http://dx.doi.org/10.1002/chir.20473.

[51] Floris, F. and Tomasi, J. "Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation". In: *Journal of Computational Chemistry* 10.5 (July 1989), pp. 616–627. DOI: 10.1002/jcc.540100504. URL: http://dx.doi.org/10.1002/jcc.540100504.

[52] Flyvbjerg, H. and Petersen, H. G. "Error estimates on averages of correlated data". In: *The Journal of Chemical Physics* 91.1 (July 1989), pp. 461–466. DOI: 10.1063/1.457480. URL: http://dx.doi.org/10.1063/1.457480.

[53] Fogolari, F., Brigo, A., and Molinari, H. "The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology". In: *Journal of Molecular Recognition* 15.6 (Nov. 2002), pp. 377–392. DOI: 10.1002/jmr.577. URL: http://dx.doi.org/10.1002/jmr.577.

[54] Fox, T. and Kollman, P. A. "Application of the RESP Methodology in the Parametrization of Organic Solvents". In: *The Journal of Physical Chemistry B* 102.41 (Sept. 1998), pp. 8070–8079. DOI: 10.1021/jp9717655. URL: http://dx.doi.org/10.1021/jp9717655.

[55] Frank, A. O. et al. "Direct prediction of residual dipolar couplings of small molecules in a stretched gel by stochastic molecular dynamics simulations". In: *Magnetic Resonance in Chemistry* 53.3 (Jan. 2015), pp. 213–217. DOI: 10.1002/mrc.4181. URL: http://dx.doi.org/10.1002/mrc.4181.

[56] Freudenberger, J. C. et al. "Stretched Poly(dimethylsiloxane) Gels as NMR Alignment Media for Apolar and Weakly Polar Organic Solvents: An Ideal Tool for Measuring RDCs at Low Molecular Concentrations". In: *Journal of the American Chemical Society* 126.45 (Oct. 2004), pp. 14690–14691. DOI: 10.1021/ja046155e. URL: http://dx.doi.org/10.1021/ja046155e.

[57] Fröhlich, H. "General theory of the static dielectric constant". In: *Transactions of the Faraday Society* 44.0 (1948), pp. 238–243. DOI: 10.1039/tf9484400238. URL: http://dx.doi.org/10.1039/tf9484400238.

[58] Gessert, D., Janke, W., and Weigel, M. "Resampling schemes in population annealing: Numerical and theoretical results". In: *Physical Review E* 108.6 (Dec. 2023). DOI: 10.1103/physreve.108.065309. URL: http://dx.doi.org/10.1103/physreve.108.065309.

[59] Hallwass, F. et al. "Residual Chemical Shift Anisotropy (RCSA): A Tool for the Analysis of the Configuration of Small Molecules". In: *Angewandte Chemie International Edition* 50.40 (Sept. 2011), pp. 9487–9490. DOI: 10.1002/anie.201101784. URL: http://dx.doi.org/10.1002/anie.201101784.

[60] Hansmann, S., Larem, T., and Thiele, C. M. "Enantiodifferentiating Properties of the Alignment Media PELG and PBLG – A Comparison". In: *European Journal of Organic Chemistry* 2016.7 (Feb. 2016), pp. 1324–1329. DOI: 10.1002/ejoc.201501410. URL: http://dx.doi.org/10.1002/ejoc.201501410.

[61] Hariharan, P. C. and Pople, J. A. "The influence of polarization functions on molecular orbital hydrogenation energies". In: *Theoretica Chimica Acta* 28.3 (1973), pp. 213–222. DOI: 10.1007/bf00533485. URL: http://dx.doi.org/10.1007/bf00533485.

[62] Hastings, W. K. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97. URL: http://dx.doi.org/10.1093/biomet/57.1.97.

[63] Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. "Pairwise solute descreening of solute charges from a dielectric medium". In: *Chemical Physics Letters* 246.1–2 (Nov. 1995), pp. 122–129. DOI: 10.1016/0009-2614(95)01082-k. URL: http://dx.doi.org/10.1016/0009-2614(95)01082-k.

[64] Haynes, W. M. *CRC handbook of chemistry and physics*. CRC press, 2014. ISBN: 9781482208689. DOI: 10.1201/b17118. URL: https://doi.org/10.1201/b17118.

[65] Hehre, W. J., Ditchfield, R., and Pople, J. A. "Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules". In: *The Journal of Chemical Physics* 56.5 (Mar. 1972), pp. 2257–2261. DOI: 10.1063/1.1677527. URL: http://dx.doi.org/10.1063/1.1677527.

[66] Helfrich, J. and Hentschke, R. "Molecular Dynamics Simulation of Macromolecular Interactions in Solution: Poly(.gamma.-benzyl glutamate) in Dimethylformamide and Tetrahydrofuran". In: *Macromolecules* 28.11 (May 1995), pp. 3831–3841. DOI: 10.1021/ma00115a013. URL: http://dx.doi.org/10.1021/ma00115a013.

[67] Helfrich, J., Hentschke, R., and Apel, U. M. "Molecular dynamics simulation study of poly(.gamma.-benzyl L-glutamate) in dimethylformamide". In: *Macromolecules* 27.2 (Jan. 1994), pp. 472–482. DOI: 10.1021/ma00080a022. URL: http://dx.doi.org/10.1021/ma00080a022.

[68] Hervet, H. et al. "Neutron scattering study of methyl group rotation in solid para-azoxyanisole (PAA)". In: *Journal de Physique* 37.5 (1976), pp. 587–594. DOI: 10.1051/jphys:01976003705058700. URL: http://dx.doi.org/10.1051/jphys:01976003705058700.

[69] Hess, B. et al. "LINCS: A linear constraint solver for molecular simulations". In: *J. Comput. Chem.* 18.12 (Sept. 1997), pp. 1463–1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h. URL: http://dx.doi.org/10.1002/(sici)1096-987x(199709)18:12%3C1463::aid-jcc4%3E3.0.co;2-h.

[70] Hornak, V. et al. "Comparison of multiple Amber force fields and development of improved protein backbone parameters". In: *Proteins: Structure, Function, and Bioinformatics* 65.3 (Sept. 2006), pp. 712–725. DOI: 10.1002/prot.21123. URL: http://dx.doi.org/10.1002/prot.21123.

[71] Hukushima, K. and Iba, Y. "Population Annealing and Its Application to a Spin Glass". In: *AIP Conference Proceedings*. Vol. 690. AIP, 2003, pp. 200–206. DOI: 10.1063/1.1632130. URL: http://dx.doi.org/10.1063/1.1632130.

[72] Hus, J.-C., Marion, D., and Blackledge, M. "Determination of Protein Backbone Structure Using Only Residual Dipolar Couplings". In: *Journal of the American Chemical Society* 123.7 (Jan. 2001), pp. 1541–1542. DOI: 10.1021/ja005590f. URL: http://dx.doi.org/10.1021/ja005590f.

[73] Iba, Y. "Population Monte Carlo algorithms." In: *Transactions of the Japanese Society for Artificial Intelligence* 16 (2001), pp. 279–286. DOI: 10.1527/tjsai.16.279. URL: http://dx.doi.org/10.1527/tjsai.16.279.

[74] Ibáñez de Opakua, A. and Zweckstetter, M. "Extending the applicability of P3D for structure determination of small molecules". In: *Magnetic Resonance* 2.1 (Apr. 2021), pp. 105–116. DOI: 10.5194/mr-2-105-2021. URL: http://dx.doi.org/10.5194/mr-2-105-2021.

[75] Jing, Z. et al. "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications". In: *Annual Review of Biophysics* 48.1 (May 2019), pp. 371–394. DOI: 10.1146/annurev-biophys-070317-033349. URL: http://dx.doi.org/10.1146/annurev-biophys-070317-033349.

[76] Jones, G. L. "On the Markov chain central limit theorem". In: *Probability Surveys* 1 (Jan. 2004), pp. 299–320. DOI: 10.1214/154957804100000051. URL: http://dx.doi.org/10.1214/154957804100000051.

[77] Jorgensen, W. L. et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (July 1983), pp. 926–935. DOI: 10.1063/1.445869. URL: http://dx.doi.org/10.1063/1.445869.

[78] Keeler, J. *Understanding NMR spectroscopy*. OCLC: ocm61229767. Chichester, England ; Hoboken, NJ: Wiley, 2005. 459 pp. ISBN: 978-0-470-01786-9 978-0-470-01787-6.

[79] Kirkwood, J. G. “Statistical Mechanics of Fluid Mixtures”. In: *The Journal of Chemical Physics* 3.5 (May 1935), pp. 300–313. DOI: 10.1063/1.1749657. URL: http://dx.doi.org/10.1063/1.1749657.

[80] Kirkwood, J. G. “The Dielectric Polarization of Polar Liquids”. In: *The Journal of Chemical Physics* 7.10 (Oct. 1939), pp. 911–919. DOI: 10.1063/1.1750343. URL: http://dx.doi.org/10.1063/1.1750343.

[81] Kong, L.-Y. and Wang, P. “Determination of the absolute configuration of natural products”. In: *Chinese Journal of Natural Medicines* 11.3 (Mar. 2014), pp. 193–198. DOI: 10.3724/sp.j.1009.2013.00193. URL: http://dx.doi.org/10.3724/sp.j.1009.2013.00193.

[82] Kuchel, P. W. et al. “Apparatus for rapid adjustment of the degree of alignment of NMR samples in aqueous media: Verification with residual quadrupolar splittings in 23Na and 133Cs spectra”. In: *Journal of Magnetic Resonance* 180.2 (June 2006), pp. 256–265. DOI: 10.1016/j.jmr.2006.03.002. URL: http://dx.doi.org/10.1016/j.jmr.2006.03.002.

[83] Kummerlöwe, G. and Luy, B. “Residual Dipolar Couplings for the Configurational and Conformational Analysis of Organic Molecules”. In: *Annual Reports on NMR Spectroscopy*. Ed. by Webb, G. A. Vol. 68. Academic Press, 2009, pp. 193–232. DOI: https://doi.org/10.1016/S0066-4103(09)06804-5. URL: https://www.sciencedirect.com/science/article/pii/S0066410309068045.

[84] Kummerlöwe, G. and Luy, B. “Residual dipolar couplings as a tool in determining the structure of organic molecules”. In: *Trends in Analytical Chemistry* 28.4 (Apr. 2009), pp. 483–493. DOI: 10.1016/j.trac.2008.11.016. URL: http://dx.doi.org/10.1016/j.trac.2008.11.016.

[85] Kummerlöwe, G., Schmitt, S., and Luy, B. “Cross-Fitting of Residual Dipolar Couplings”. In: *The Open Spectroscopy Journal* 4.1 (May 2010), pp. 16–27. DOI: 10.2174/1874383801004010016. URL: http://dx.doi.org/10.2174/1874383801004010016.

[86] Kummerlöwe, G. et al. “Precise Measurement of RDCs in Water and DMSO Based Gels Using a Silicone Rubber Tube for Tunable Stretching”. In: *The Open Spectroscopy Journal* 2.1 (July 2008), pp. 29–33. DOI: 10.2174/1874383800802010029. URL: http://dx.doi.org/10.2174/1874383800802010029.

[87] Lebedev, V. “Quadratures on a sphere”. In: *USSR Computational Mathematics and Mathematical Physics* 16.2 (Jan. 1976), pp. 10–24. DOI: 10.1016/0041-5553(76)90100-2. URL: http://dx.doi.org/10.1016/0041-5553(76)90100-2.

[88] Lemons, D. S. and Gythiel, A. "Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]". In: *American Journal of Physics* 65.11 (Nov. 1997), pp. 1079–1081. DOI: 10.1119/1.18725. URL: http://dx.doi.org/10.1119/1.18725.

[89] Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley & Sons, 2013. ISBN: 9780470511176.

[90] Li, G.-W. et al. "Residual Dipolar Couplings in Structure Determination of Natural Products". In: *Natural Products and Bioprospecting* 8.4 (June 2018), pp. 279–295. DOI: 10.1007/s13659-018-0174-x. URL: http://dx.doi.org/10.1007/s13659-018-0174-x.

[91] Lipsitz, R. S. and Tjandra, N. "Residual Dipolar Couplings in NMR Structure Analysis". In: *Annual Review of Biophysics and Biomolecular Structure* 33.1 (June 2004), pp. 387–413. DOI: 10.1146/annurev.biophys.33.110502.140306. URL: http://dx.doi.org/10.1146/annurev.biophys.33.110502.140306.

[92] Liu, H.-Y., Grinter, S. Z., and Zou, X. "Multiscale Generalized Born Modeling of Ligand Binding Energies for Virtual Database Screening". In: *The Journal of Physical Chemistry B* 113.35 (Aug. 2009), pp. 11793–11799. DOI: 10.1021/jp901212t. URL: http://dx.doi.org/10.1021/jp901212t.

[93] Liu, Y. et al. "A practical strategy for the accurate measurement of residual dipolar couplings in strongly aligned small molecules". In: *Journal of Magnetic Resonance* 291 (June 2018), pp. 63–72. DOI: 10.1016/j.jmr.2018.04.002. URL: http://dx.doi.org/10.1016/j.jmr.2018.04.002.

[94] Lorentz, H. A. "Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase". In: *Annalen der Physik* 248.1 (Jan. 1881), pp. 127–136. DOI: 10.1002/andp.18812480110. URL: http://dx.doi.org/10.1002/andp.18812480110.

[95] Losonczi, J. A. et al. "Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition". In: *J. Magn. Reson.* 138.2 (June 1999), pp. 334–342. DOI: 10.1006/jmre.1999.1754. URL: http://dx.doi.org/10.1006/jmre.1999.1754.

[96] Luy, B. "Disinction of enantiomers by NMR spectroscopy using chiral orienting media". In: *Journal of the Indian Institute of Science* 90.1 (2010), pp. 119–132.

[97] Luy, B., Kobzar, K., and Kessler, H. "An Easy and Scalable Method for the Partial Alignment of Organic Molecules for Measuring Residual Dipolar Couplings". In: *Angewandte Chemie International Edition* 43.9 (Feb. 2004), pp. 1092–1094. DOI: 10.1002/anie.200352860. URL: http://dx.doi.org/10.1002/anie.200352860.

[98] Luy, B. et al. "Orientational Properties of Stretched Polystyrene Gels in Organic Solvents and the Suppression of Their Residual 1H NMR Signals". In: *Journal of the American Chemical Society* 127.17 (Apr. 2005), pp. 6459–6465. DOI: 10.1021/ja043344o. URL: http://dx.doi.org/10.1021/ja043344o.

[99] Machta, J. "Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes". In: *Physical Review E* 82.2 (Aug. 2010). DOI: 10.1103/physreve.82.026704. URL: http://dx.doi.org/10.1103/physreve.82.026704.

[100] Machta, J. and Ellis, R. S. "Monte Carlo Methods for Rough Free Energy Landscapes: Population Annealing and Parallel Tempering". In: *Journal of Statistical Physics* 144.3 (June 2011), pp. 541–553. DOI: 10.1007/s10955-011-0249-0. URL: http://dx.doi.org/10.1007/s10955-011-0249-0.

[101] Maier, J. A. et al. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". In: *Journal of Chemical Theory and Computation* 11.8 (July 2015), pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255. URL: http://dx.doi.org/10.1021/acs.jctc.5b00255.

[102] Marathias, V. M. et al. "Stereochemical identification of (R)- and (S)-ibuprofen using residual dipolar couplings, NMR, and modeling". In: *Chirality* 19.9 (Nov. 2006), pp. 741–750. DOI: 10.1002/chir.20338. URL: http://dx.doi.org/10.1002/chir.20338.

[103] Markwick, P. R. L., Malliavin, T., and Nilges, M. "Structural Biology by NMR: Structure, Dynamics, and Interactions". In: *PLoS Computational Biology* 4.9 (Sept. 2008). Ed. by McEntyre, J., e1000168. DOI: 10.1371/journal.pcbi.1000168. URL: http://dx.doi.org/10.1371/journal.pcbi.1000168.

[104] Marx, A., Schmidts, V., and Thiele, C. M. "How different are diastereomorphous orientations of enantiomers in the liquid crystalline phases of PBLG and PBDG: a case study". In: *Magnetic Resonance in Chemistry* 47.9 (June 2009), pp. 734–740. DOI: 10.1002/mrc.2454. URL: http://dx.doi.org/10.1002/mrc.2454.

[105] Marx, A. and Thiele, C. "Orientational Properties of Poly-γ-benzyl-L-glutamate: Influence of Molecular Weight and Solvent on Order Parameters of the Solute". In: *Chemistry - A European Journal* 15.1 (Jan. 2009), pp. 254–260. DOI: 10.1002/chem.200801147. URL: http://dx.doi.org/10.1002/chem.200801147.

[106] Metropolis, N. et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. DOI: 10.1063/1.1699114. URL: http://dx.doi.org/10.1063/1.1699114.

[107] Milanowski, D. J. et al. "Unequivocal determination of caulamidines A and B: application and validation of new tools in the structure elucidation tool box". In: *Chemical Science* 9.2 (2018), pp. 307–314. DOI: 10.1039/c7sc01996c. URL: http://dx.doi.org/10.1039/c7sc01996c.

[108] Mölder, F. et al. "Sustainable data analysis with Snakemake". In: *F1000Research* 10 (Apr. 2021), p. 33. DOI: 10.12688/f1000research.29032.2. URL: http://dx.doi.org/10.12688/f1000research.29032.2.

[109] Nath, N., d'Auvergne, E. J., and Griesinger, C. "Long-Range Residual Dipolar Couplings: A Tool for Determining the Configuration of Small Molecules". In: *Angewandte Chemie International Edition* 54.43 (Sept. 2015), pp. 12706–12710. DOI: 10.1002/anie.201504432. URL: http://dx.doi.org/10.1002/anie.201504432.

[110] Neeman, E. M., Avilés-Moreno, J.-R., and Huet, T. R. "The quasi-unchanged gas-phase molecular structures of the atmospheric aerosol precursor β-pinene and its oxidation product nopinone". In: *Physical Chemistry Chemical Physics* 19.21 (2017), pp. 13819–13827. DOI: 10.1039/c7cp01298e. URL: http://dx.doi.org/10.1039/c7cp01298e.

[111] Neumann, T., Danilov, D., and Wenzel, W. "Multiparticle moves in acceptance rate optimized monte carlo". In: *Journal of Computational Chemistry* 36.30 (Oct. 2015), pp. 2236–2245. DOI: 10.1002/jcc.24205. URL: http://dx.doi.org/10.1002/jcc.24205.

[112] Neyman, J. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection". In: *Journal of the Royal Statistical Society* 97.4 (1934), p. 558. DOI: 10.2307/2342192. URL: http://dx.doi.org/10.2307/2342192.

[113] Nguyen, H. et al. "Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent". In: *Journal of the American Chemical Society* 136.40 (Sept. 2014), pp. 13959–13962. DOI: 10.1021/ja5032776. URL: http://dx.doi.org/10.1021/ja5032776.

[114] Nguyen, L. A., He, H., and Pham-Huy, C. "Chiral drugs: an overview". In: *International journal of biomedical science* 2.2 (2006), p. 85.

[115] Onufriev, A., Case, D. A., and Bashford, D. "Effective Born radii in the generalized Born approximation: The importance of being perfect". In: *Journal of Computational Chemistry* 23.14 (Aug. 2002), pp. 1297–1304. DOI: 10.1002/jcc.10126. URL: http://dx.doi.org/10.1002/jcc.10126.

[116] Opakua, A. I. d. et al. "Determination of Complex Small-Molecule Structures Using Molecular Alignment Simulation". In: *Angewandte Chemie International Edition* 59.15 (2020), pp. 6172–6176. DOI: https://doi.org/10.1002/anie.202000311. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202000311.

[117] Ottiger, M. and Bax, A. "How Tetrahedral Are Methyl Groups in Proteins? A Liquid Crystal NMR Study". In: *Journal of the American Chemical Society* 121.19 (May 1999), pp. 4690–4695. DOI: 10.1021/ja984484z. URL: http://dx.doi.org/10.1021/ja984484z.

[118] Ottiger, M. et al. "Measurement of Dipolar Couplings for Methylene and Methyl Sites in Weakly Oriented Macromolecules and Their Use in Structure Determination". In: *Journal of Magnetic Resonance* 134.2 (Oct. 1998), pp. 365–369. DOI: 10.1006/jmre.1998.1546. URL: http://dx.doi.org/10.1006/jmre.1998.1546.

[119] Panar, M. and Phillips, W. D. "Magnetic ordering of poly-*gamma*-benzyl L-glutamate solutions". In: *Journal of the American Chemical Society* 90.14 (July 1968), pp. 3880–3882. DOI: 10.1021/ja01016a061. URL: http://dx.doi.org/10.1021/ja01016a061.

[120] Parrinello, M. and Rahman, A. "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study". In: *Physical Review Letters* 45.14 (Oct. 1980), pp. 1196–1199. DOI: 10.1103/physrevlett.45.1196. URL: http://dx.doi.org/10.1103/physrevlett.45.1196.

[121] Parrinello, M. and Rahman, A. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied Physics* 52.12 (Dec. 1981), pp. 7182–7190. DOI: 10.1063/1.328693. URL: http://dx.doi.org/10.1063/1.328693.

[122] Patterson, D. and Schnell, M. "New studies on molecular chirality in the gas phase: enantiomer differentiation and determination of enantiomeric excess". In: *Physical Chemistry Chemical Physics* 16.23 (2014), pp. 11114–11123. DOI: 10.1039/c4cp00417e. URL: http://dx.doi.org/10.1039/c4cp00417e.

[123] Pellecchia, M. et al. "Perspectives on NMR in drug discovery: a technique comes of age". In: *Nature Reviews Drug Discovery* 7.9 (Sept. 2008), pp. 738–745. DOI: 10.1038/nrd2606. URL: http://dx.doi.org/10.1038/nrd2606.

[124] Pereda-Miranda, R. et al. "From Relative to Absolute Stereochemistry of Secondary Metabolites: Applications in Plant Chemistry". In: *Revista Brasileira de Farmacognosia* 33.1 (Dec. 2022), pp. 1–48. DOI: 10.1007/s43450-022-00333-y. URL: http://dx.doi.org/10.1007/s43450-022-00333-y.

[125] Permi, P. "A spin-state-selective experiment for measuring heteronuclear one-bond and homonuclear two-bond couplings from an HSQC-type spectrum". In: *Journal of biomolecular NMR* 22 (2002), pp. 27–35.

[126] Pierotti, R. A. "A scaled particle theory of aqueous and nonaqueous solutions". In: *Chemical Reviews* 76.6 (Dec. 1976), pp. 717–726. DOI: 10.1021/cr60304a002. URL: http://dx.doi.org/10.1021/cr60304a002.

[127] Pizzirusso, A. et al. "Alignment of Small Organic Solutes in a Nematic Solvent: The Effect of Electrostatic Interactions". In: *The Journal of Physical Chemistry B* 116.12 (Mar. 2012), pp. 3760–3771. DOI: 10.1021/jp3003799. URL: http://dx.doi.org/10.1021/jp3003799.

[128] Prestegard, J. H., Bougault, C. M., and Kishore, A. I. "Residual Dipolar Couplings in Structure Determination of Biomolecules". In: *Chemical Reviews* 104.8 (July 2004), pp. 3519–3540. DOI: 10.1021/cr030419i. URL: http://dx.doi.org/10.1021/cr030419i.

[129] Reynolds, C. A., Essex, J. W., and Richards, W. G. "Atomic charges for variable molecular conformations". In: *Journal of the American Chemical Society* 114.23 (Nov. 1992), pp. 9075–9079. DOI: 10.1021/ja00049a045. URL: http://dx.doi.org/10.1021/ja00049a045.

[130] Roux, B. and Simonson, T. "Implicit solvent models". In: *Biophysical Chemistry* 78.1–2 (Apr. 1999), pp. 1–20. DOI: 10.1016/s0301-4622(98)00226-9. URL: http://dx.doi.org/10.1016/s0301-4622(98)00226-9.

[131] Ruan, K. and Tolman, J. R. "Composite Alignment Media for the Measurement of Independent Sets of NMR Residual Dipolar Couplings". In: *Journal of the American Chemical Society* 127.43 (Oct. 2005), pp. 15032–15033. DOI: 10.1021/ja055520e. URL: http://dx.doi.org/10.1021/ja055520e.

[132] Rubinstein, R. Y. and Kroese, D. P. *Simulation and the Monte Carlo Method*. Wiley, June 2007. ISBN: 9780470230381. DOI: 10.1002/9780470230381. URL: http://dx.doi.org/10.1002/9780470230381.

[133] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (Mar. 1977), pp. 327–341. DOI: 10.1016/0021-9991(77)90098-5. URL: http://dx.doi.org/10.1016/0021-9991(77)90098-5.

[134] Sackmann, E., Meiboom, S., and Snyder, L. C. "Nuclear magnetic resonance spectra of enantiomers in optically active liquid crystals". In: *Journal of the American Chemical Society* 90.8 (Apr. 1968), pp. 2183–2184. DOI: 10.1021/ja01010a047. URL: http://dx.doi.org/10.1021/ja01010a047.

[135] Sackmann, E., Meiboom, S., and Snyder, L. C. "Relation of nematic to cholesteric mesophases". In: *Journal of the American Chemical Society* 89.23 (Nov. 1967), pp. 5981–5982. DOI: 10.1021/ja00999a062. URL: http://dx.doi.org/10.1021/ja00999a062.

[136] Sager, E. "Residual dipolar couplings: a complementary tool for stereochemistry determination of drug compounds". PhD thesis. Karlsruher Institut für Technologie (KIT), 2019. 179 pp. DOI: 10.5445/IR/1000094751.

[137] Sager, E. et al. "Hydrogen bond formation may enhance RDC-based discrimination of enantiomers". In: *Magnetic Resonance in Chemistry* 1 (May 2024). DOI: 10.1002/mrc.5448. URL: http://dx.doi.org/10.1002/mrc.5448.

[138] Samulski, E. T. and Tobolsky, A. V. "Some Unusual Properties of Poly-(*gamma*-benzyl L-glutamate) Films Cast in Strong Magnetic Fields". In: *Macromolecules* 1.6 (Nov. 1968), pp. 555–557. DOI: 10.1021/ma60006a021. URL: http://dx.doi.org/10.1021/ma60006a021.

[139] Sánchez-Pedregal, V. M., Santamaría-Fernández, R., and Navarro-Vázquez, A. "Residual Dipolar Couplings of Freely Rotating Groups in Small Molecules. Stereochemical Assignment and Side-Chain Conformation of 8-Phenylmenthol". In: *Organic Letters* 11.7 (Mar. 2009), pp. 1471–1474. DOI: 10.1021/ol8029565. URL: http://dx.doi.org/10.1021/ol8029565.

[140] Sarfati, M. et al. "Theoretical and experimental aspects of enantiomeric differentiation using natural abundance multinuclear nmr spectroscopy in chiral polypeptide liquid crystals". In: *Chemical Communications* 21 (2000), pp. 2069–2081. DOI: 10.1039/b006244h. URL: http://dx.doi.org/10.1039/b006244h.

[141] Saupe, A. and Englert, G. "High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules". In: *Physical Review Letters* 11.10 (Nov. 1963), pp. 462–464. DOI: 10.1103/physrevlett.11.462. URL: http://dx.doi.org/10.1103/physrevlett.11.462.

[142] Saupe, A. "Kernresonanzen in kristallinen Flüssigkeiten und in kristallinflüssigen Lösungen. Teil I". In: *Zeitschrift für Naturforschung A* 19.2 (Feb. 1964), pp. 161–171. DOI: 10.1515/zna-1964-0201. URL: http://dx.doi.org/10.1515/zna-1964-0201.

[143] Schäfer, A., Horn, H., and Ahlrichs, R. "Fully optimized contracted Gaussian basis sets for atoms Li to Kr". In: *The Journal of Chemical Physics* 97.4 (Aug. 1992), pp. 2571–2577. DOI: 10.1063/1.463096. URL: http://dx.doi.org/10.1063/1.463096.

[144] Schmieder, P., Kurz, M., and Kessler, H. "Determination of heteronuclear long-range couplings to heteronuclei in natural abundance by two- and three-dimensional NMR spectroscopy". In: *Journal of Biomolecular NMR* 1.4 (Nov. 1991), pp. 403–420. DOI: 10.1007/bf02192863. URL: http://dx.doi.org/10.1007/bf02192863.

[145] Schuetz, A. et al. "Stereochemistry of Sagittamide A from Residual Dipolar Coupling Enhanced NMR". In: *Journal of the American Chemical Society* 129.49 (Nov. 2007), pp. 15114–15115. DOI: 10.1021/ja075876l. URL: http://dx.doi.org/10.1021/ja075876l.

[146] Seco, J. M., Quiñoá, E., and Riguera, R. "The Assignment of Absolute Configuration by NMR". In: *Chemical Reviews* 104.1 (2004), pp. 17–118. DOI: 10.1021/cr000665j. URL: https://doi.org/10.1021/cr000665j.

[147] Shirts, M. R. and Chodera, J. D. "Statistically optimal analysis of samples from multiple equilibrium states". In: *The Journal of Chemical Physics* 129.12 (Sept. 2008). DOI: 10.1063/1.2978177. URL: http://dx.doi.org/10.1063/1.2978177.

[148] Sigalov, G., Fenley, A., and Onufriev, A. "Analytical electrostatics for biomolecules: Beyond the generalized Born approximation". In: *The Journal of Chemical Physics* 124.12 (Mar. 2006). DOI: 10.1063/1.2177251. URL: http://dx.doi.org/10.1063/1.2177251.

[149] Sinnaeve, D. et al. "Probing Long-Range Anisotropic Interactions: a General and Sign-Sensitive Strategy to Measure 1H–1H Residual Dipolar Couplings as a Key Advance for Organic Structure Determination". In: *Angewandte Chemie International Edition* 59.13 (Feb. 2020), pp. 5316–5320. DOI: 10.1002/anie.201915278. URL: http://dx.doi.org/10.1002/anie.201915278.

[150] van der Spoel, D., van Maaren, P. J., and Berendsen, H. J. C. "A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field". In: *The Journal of Chemical Physics* 108.24 (June 1998), pp. 10220–10230. DOI: 10.1063/1.476482. URL: http://dx.doi.org/10.1063/1.476482.

[151] van der Spoel, D., Zhang, J., and Zhang, H. "Quantitative predictions from molecular simulations using explicit or implicit interactions". In: *WIREs Computational Molecular Science* 12.1 (June 2021). DOI: 10.1002/wcms.1560. URL: http://dx.doi.org/10.1002/wcms.1560.

[152] Steipe, B. "A revised proof of the metric properties of optimally superimposed vector sets". In: *Acta Crystallographica Section A Foundations of Crystallography* 58.5 (Aug. 2002), pp. 506–506. DOI: 10.1107/s0108767302011637. URL: http://dx.doi.org/10.1107/s0108767302011637.

[153] Stephens, P. J. et al. "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields". In: *The Journal of Physical Chemistry* 98.45 (Nov. 1994), pp. 11623–11627. DOI: 10.1021/j100096a001. URL: http://dx.doi.org/10.1021/j100096a001.

[154] Still, W. C. et al. "Semianalytical treatment of solvation for molecular mechanics and dynamics". In: *Journal of the American Chemical Society* 112.16 (Aug. 1990), pp. 6127–6129. DOI: 10.1021/ja00172a038. URL: http://dx.doi.org/10.1021/ja00172a038.

[155] Stillinger, F. H. "Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory". In: *Journal of Solution Chemistry* 2.2–3 (1973), pp. 141–158. DOI: 10.1007/bf00651970. URL: http://dx.doi.org/10.1007/bf00651970.

[156] Sun, Q. et al. "Recent developments in the PySCF program package". In: *The Journal of Chemical Physics* 153.2 (July 2020), p. 024109. DOI: 10.1063/5.0006074. URL: https://doi.org/10.1063/5.0006074.

[157] Tan, C., Tan, Y.-H., and Luo, R. "Implicit Nonpolar Solvent Models". In: *The Journal of Physical Chemistry B* 111.42 (Oct. 2007), pp. 12263–12274. DOI: 10.1021/jp073399n. URL: http://dx.doi.org/10.1021/jp073399n.

[158] Thiele, C. M. "Scaling the Alignment of Small Organic Molecules in Substituted Polyglutamates by Variable-Angle Sample Spinning". In: *Angewandte Chemie International Edition* 44.18 (Apr. 2005), pp. 2787–2790. DOI: 10.1002/anie.200461532. URL: http://dx.doi.org/10.1002/anie.200461532.

[159] Thiele, C. M., Maliniak, A., and Stevensson, B. "Use of Local Alignment Tensors for the Determination of Relative Configurations in Organic Compounds". In: *Journal of the American Chemical Society* 131.36 (Aug. 2009), pp. 12878–12879. DOI: 10.1021/ja904536b. URL: http://dx.doi.org/10.1021/ja904536b.

[160] Thiele, C. M. et al. "Determination of the Relative Configuration of a Five-Membered Lactone from Residual Dipolar Couplings". In: *Angewandte Chemie International Edition* 45.27 (July 2006), pp. 4455–4460. DOI: 10.1002/anie.200503247. URL: http://dx.doi.org/10.1002/anie.200503247.

[161] Tjandra, N. and Bax, A. "Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium". In: *Science* 278.5340 (Nov. 1997), pp. 1111–1114. DOI: 10.1126/science.278.5340.1111. URL: http://dx.doi.org/10.1126/science.278.5340.1111.

[162] Tobolsky, A. V. and Samulski, E. T. "Solid "liquid-crystalline" films of synthetic polypeptides: a new state of matter". In: *Pure and Applied Chemistry* 23.2–3 (Jan. 1970), pp. 145–150. DOI: 10.1351/pac197023020145. URL: http://dx.doi.org/10.1351/pac197023020145.

[163] Tokdar, S. T. and Kass, R. E. "Importance sampling: a review". In: *WIREs Computational Statistics* 2.1 (Dec. 2009), pp. 54–60. DOI: 10.1002/wics.56. URL: http://dx.doi.org/10.1002/wics.56.

[164] Troche-Pesqueira, E. et al. "Computer-Assisted 3D Structure Elucidation of Natural Products using Residual Dipolar Couplings". In: *Angewandte Chemie International Edition* 56.13 (Feb. 2017), pp. 3660–3664. DOI: 10.1002/anie.201612454. URL: http://dx.doi.org/10.1002/anie.201612454.

[165] Tzvetkova, P. et al. "Configuration determination by residual dipolar couplings: accessing the full conformational space by molecular dynamics with tensorial constraints". In: *Chemical Science* 10.38 (2019), pp. 8774–8791. DOI: 10.1039/c9sc01084j. URL: http://dx.doi.org/10.1039/c9sc01084j.

[166] Vasilyev, V. "Determination of the effective dielectric constant from the accurate solution of the Poisson equation". In: *Journal of Computational Chemistry* 23.13 (July 2002), pp. 1254–1265. DOI: 10.1002/jcc.10131. URL: http://dx.doi.org/10.1002/jcc.10131.

[167] Wagoner, J. A. and Baker, N. A. "Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms". In: *Proceedings of the National Academy of Sciences* 103.22 (May 2006), pp. 8331–8336. DOI: 10.1073/pnas.0600118103. URL: http://dx.doi.org/10.1073/pnas.0600118103.

[168] Wang, L.-P. and Song, C. "Geometry optimization made simple with translation and rotation coordinates". In: *The Journal of Chemical Physics* 144.21 (June 2016). DOI: 10.1063/1.4952956. URL: http://dx.doi.org/10.1063/1.4952956.

[169] Weeks, J. D., Chandler, D., and Andersen, H. C. "Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids". In: *The Journal of Chemical Physics* 54.12 (June 1971), pp. 5237–5247. DOI: 10.1063/1.1674820. URL: http://dx.doi.org/10.1063/1.1674820.

[170] Weigel, M. et al. "Population annealing: Massively parallel simulations in statistical physics". In: *Journal of Physics: Conference Series* 921 (Nov. 2017), p. 012017. DOI: 10.1088/1742-6596/921/1/012017. URL: http://dx.doi.org/10.1088/1742-6596/921/1/012017.

[171] Weigel, M. et al. "Understanding population annealing Monte Carlo simulations". In: *Physical Review E* 103.5 (May 2021). DOI: 10.1103/physreve.103.053301. URL: http://dx.doi.org/10.1103/physreve.103.053301.

[172] Weiner, P. K. and Kollman, P. A. "AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions". In: *Journal of Computational Chemistry* 2.3 (Sept. 1981), pp. 287–303. DOI: 10.1002/jcc.540020311. URL: http://dx.doi.org/10.1002/jcc.540020311.

[173] Weiner, S. J. et al. "An all atom force field for simulations of proteins and nucleic acids". In: *Journal of Computational Chemistry* 7.2 (Apr. 1986), pp. 230–252. DOI: 10.1002/jcc.540070216. URL: http://dx.doi.org/10.1002/jcc.540070216.

[174] Wu, B. et al. "Prediction of molecular alignment of nucleic acids in aligned media". In: *Journal of Biomolecular NMR* 35.2 (May 2006), pp. 103–115. DOI: 10.1007/s10858-006-9004-2. URL: http://dx.doi.org/10.1007/s10858-006-9004-2.

[175] Zhang, J. et al. “Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents”. In: *Journal of Chemical Theory and Computation* 13.3 (Mar. 2017), pp. 1034–1043. DOI: 10.1021/acs.jctc.7b00169. URL: http://dx.doi.org/10.1021/acs.jctc.7b00169.

[176] Ziani, L. et al. “Empirical determination of the absolute configuration of small chiral molecules using natural abundance 2H NMR in chiral liquid crystals”. In: *Chemical Communications* 45 (2007), pp. 4737–4739. DOI: 10.1039/b711493a. URL: http://dx.doi.org/10.1039/b711493a.

[177] Zwanzig, R. W. “High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases”. In: *The Journal of Chemical Physics* 22.8 (Aug. 1954), pp. 1420–1426. DOI: 10.1063/1.1740409. URL: http://dx.doi.org/10.1063/1.1740409.

[178] Zweckstetter, M. “NMR: prediction of molecular alignment from structure using the PALES software”. In: *Nature Protocols* 3.4 (Mar. 2008), pp. 679–690. DOI: 10.1038/nprot.2008.36. URL: http://dx.doi.org/10.1038/nprot.2008.36.

[179] Zweckstetter, M., Hummer, G., and Bax, A. “Prediction of Charge-Induced Molecular Alignment of Biomolecules Dissolved in Dilute Liquid-Crystalline Phases”. In: *Biophysical Journal* 86.6 (June 2004), pp. 3444–3460. DOI: 10.1529/biophysj.103.035790. URL: http://dx.doi.org/10.1529/biophysj.103.035790.

List of Figures

List of Tables

A. Appendix

A.1. PBLG structure

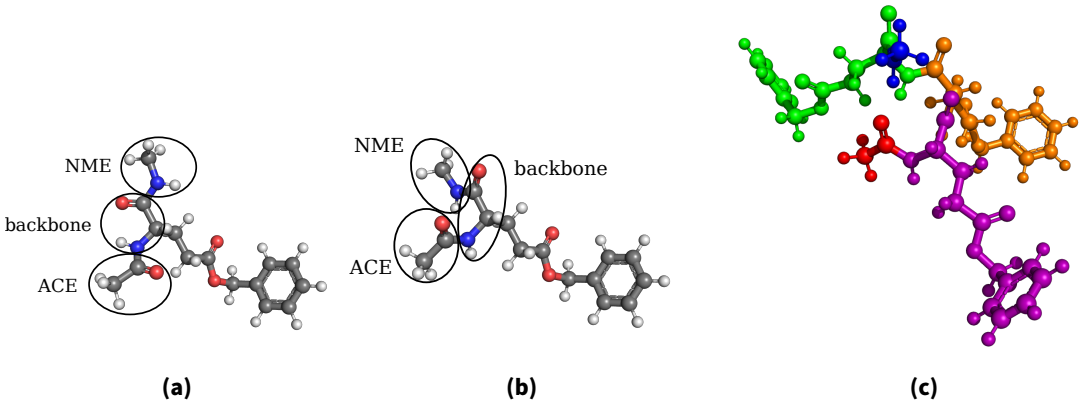


Figure A.1.: BLG monomer with $COCH_3$ (ACE) and $NHCH_3$ (NME) caps. In (a), the DFT-optimized monomer is shown, whereas in (b), the backbone bond lengths, angles and dihedrals were adjusted to experimental data of an α-helix. In (c), the molecule made of three joined BLG monomers with caps for force field parameter determination with Ambertools is shown. Here, the ACE cap is shown in red, the three monomers in purple, orange and green, and the NME cap in blue.

A.2. Scaling of the alignment

As shown in Frank et al. [55], the prediction of the RDCs in strychnine which is aligned by a polystyrene/chloroform gel is possible up to a constant factor by a molecular dynamics simulation of strychnine and a single strand of the alignment medium, which is perfectly aligned along the z-axis. In fact, this proportionality holds generally:

The dipolar couplings of a molecule are proportional (with the same constant s for all couplings) to the value obtained for the average with the isotropic part removed [165]:

$$\left\langle \cos^2(\theta_{ij}) - \frac{1}{3} \right\rangle_t = s \left\langle \cos^2(\theta_{ij}) - \frac{1}{3} \right\rangle_{t,\text{anisotropic}} . \tag{A.1}$$

Because the measurement is done over a long enough time that the molecules in question each interact with many parts of the alignment medium (i.e. different chains of a polymer) [26], the average of each coupling is very similar between all molecules. This allows for

the measurement of a single RDC value for each coupling even if the alignment medium is not perfectly aligned. Furthermore, this value is proportional to the RDC value one would measure if the alignment medium *was* perfectly aligned. This is explained in the following way:

Let $\boldsymbol{A}_z = \frac{1}{3}\,\mathrm{diag}(-1,-1,2)$ and U be the rotation from the molecule coordinate frame (or frame of an individual atom pair if the molecule is flexible) to the laboratory frame, where the alignment medium is perfectly aligned along the z-axis. Then, the alignment tensor (for the whole molecule or a single coupling) is given as:

$$\boldsymbol{A}_{\text{perfect}} = \left\langle U^\top \boldsymbol{A}_z U \right\rangle_U . \tag{A.2}$$

Here, we assume that the polymer consists of ideal chains consisting of monomers i, whose angle to the z-axis is given by the random variable $\alpha \in [0,\pi]$ (independent of U). If the monomers are rotated uniformly (and independent from α) along their longitudinal axis **or** the polymer is symmetric along its longitudinal axis (both of which are valid assumptions for PBLG), the alignment tensor is:

$$\boldsymbol{A} = \left\langle U^\top R_{x,\alpha}^\top R_{z,\gamma}^\top \boldsymbol{A}_z R_{z,\gamma} R_{x,\alpha} U \right\rangle_{\alpha,\beta,U} \tag{A.3}$$

with $R_{k,\theta}$ being the rotation matrix around the axis k by an angle θ and $\beta \sim U\big([0,2\pi)\big)$ the rotation of the monomer, which is assumed to be independent of α and U.

Because $\langle\cos^2(\beta)\rangle = \langle\sin^2(\beta)\rangle = \frac{1}{2}$, we have:

$$\left\langle R_{x,\alpha}^\top R_{z,\gamma}^\top \boldsymbol{A}_z R_{z,\gamma} R_{x,\alpha} \right\rangle_{\alpha,\beta} = \frac{1}{2}\langle 3\cos^2(\alpha) - 1\rangle_\alpha \boldsymbol{A}_z \tag{A.4}$$

and, therefore:

$$\boldsymbol{A} = \frac{1}{2}\langle 3\cos^2(\alpha) - 1\rangle_\alpha \boldsymbol{A}_{\text{perfect}}, \tag{A.5}$$

which means a scaling for all RDCs by a factor of $\frac{1}{2}\langle 3\cos^2(\alpha) - 1\rangle_\alpha$. Importantly, as the PBLG liquid crystal is strongly aligned along the z-axis [162], $\cos^2(\alpha)$ is large enough on average that this scaling is positive.

A.3. Rotating bonds in a CH_3 group

The timescale during which the rotation of CH_3 groups (present in the studied molecules) takes place, e.g. several femtoseconds in solid para-azoxyanizole [68], is much smaller than the *spectral timescale*. Therefore, special care has to be taken for such freely rotatable bonds[117]. The C – H RDCs in a CH_3 group can be correctly described by the observation that they are scaled by a factor $\frac{1}{2}\left(3\cos^2(\varphi) - 1\right)$ (where X is the non-hydrogen atom the CH_3 group is attached to) compared to the RDC of a C – H bond with the orientation of the X – C bond (see below, Equation (A.7)). Here, φ is the X – C – H bond angle, which is on average the same for all three hydrogen atoms. For $\varphi = \arccos(-1/3)$ (the bond

angle for an ideal tetrahedron), this results in a factor of $-1/3$ [139]. Therefore, when calculating the RDCs from the MD trajectory, the X – C bond can be used to determine θ_{ij} (for Equation (2.14)) and apply this scaling for calculating the average of the three C – H RDCs in a CH_3 group.

The scaling factor can derived as follows: Given the alignment tensor $\boldsymbol{A}^{\text{fixed}}$ of the imagined fixed C – H bond in a coordinate frame where it is oriented along the z-axis, the scaling is

$$\left\langle \cos^2(\theta) - \frac{1}{3} \right\rangle_{t,\text{fixed}} = \boldsymbol{A}^{\text{fixed}}_{zz}. \tag{A.6}$$

Note that $\boldsymbol{A}^{\text{fixed}}$ is not necessarily diagonal. Assuming the independence of the rotation of the CH_3 group (with (x, y) uniformly distributed in S^1) and the remaining molecular motion:

$$\begin{aligned}
\left\langle \cos^2(\theta) - \frac{1}{3} \right\rangle_{t,\text{C–H}} &= \left\langle \begin{pmatrix} x\sin(\varphi) & y\sin(\varphi) & \cos(\varphi) \end{pmatrix} \boldsymbol{A}^{\text{fixed}} \begin{pmatrix} x\sin(\varphi) \\ y\sin(\varphi) \\ \cos(\varphi) \end{pmatrix} \right\rangle_{x,y} \\
&= \frac{1}{2}\sin^2(\varphi)\left(\boldsymbol{A}^{\text{fixed}}_{xx} + \boldsymbol{A}^{\text{fixed}}_{yy}\right) + \cos^2(\varphi)\boldsymbol{A}^{\text{fixed}}_{zz} \\
&= \frac{1}{2}\left(3\cos^2(\varphi) - 1\right)\boldsymbol{A}^{\text{fixed}}_{zz}.
\end{aligned} \tag{A.7}$$

A.4. Experimental couplings

Table A.1.: $^1D_{\text{C–H}}$ couplings and experimental values (in Hz) and their measurement errors for **(a)** IPC [104] and **(b)** quinuclidinol [137]. As explained in Section 3.1.2, the CH_3-9 and CH_3-8 couplings in IPC are swapped compared to Marx, Schmidts, and Thiele [104].

	(a)	
	(-)-IPC	(+)-IPC
C1-H1	13.3(2)	14.2(2)
C2-H2	−7.4(5)	−6.2(5)
C3-H3	11.1(2)	13.4(2)
C4-H4s	10.9(2)	14.7(2)
C4-H4a	−4.0(2)	−4.3(2)
C5-H5	−5.5(2)	−5.1(2)
CH_3-9	4.0(4)	5.1(4)
CH_3-8	−2.9(4)	−4.0(4)
CH_3-10	−0.3(4)	−0.3(4)

	(b)	
	R-(-)-quinuclidinol	*S*-(+)-quinuclidinol
CH-5	22.7(3)	13.0(3)
CH_2-6’	−15.0(15)	−39.0(4)
CH_2-6”	−25.2(5)	−1.0(8)
CH_2-2’	41.3(5)	30.4(50)
CH_2-2”	−7.3(5)	−35.5(50)
CH_2-8’	−16.4(14)	−6.2(4)
CH_2-8”	32.9(40)	28.1(5)
CH-3	−4.9(3)	5.8(5)
CH_2-7’	23.1(10)	15.0(10)
CH_2-7”	10.5(20)	1.9(10)
CH_2-1’	9.1(10)	3.5(10)
CH_2-1”	−26.3(5)	−17.9(10)

Table A.2.: $^1D_{C-H}$ couplings and experimental values (in Hz) and their measurement errors for **(a)** borneol [137] and **(b)** camphor [137].

	(a)			**(b)**	
	S-(-)-borneol	*R*-(+)-borneol		*S*-(-)-camphor	*R*-(+)-camphor
CH-5	3.3(15)	29.2(4)	CH_2-6'	7.8(20)	11.0(4)
CH-4	4.9(5)	−20.5(20)	CH_2-6"	29.4(6)	26.2(7)
CH_2-6'	−30.9(20)	−11.2(10)	CH-4	2.3(30)	4.1(6)
CH_2-6"	2.4(10)	−1.6(13)	CH_2-2'	1.6(30)	5.9(20)
CH_2-2'	−21.8(30)	−9.0(24)	CH_2-2"	−41.5(30)	−52.9(30)
CH_2-2"	24.2(15)	30.4(6)	CH_3-10	−7.8(10)	−7.2(15)
CH_2-1'	19.1(10)	11.0(8)	CH_3-11	15.8(10)	15.7(3)
CH_2-1"	−25.7(50)	−22.8(10)	CH_3-8	−6.1(10)	−6.9(8)
CH_3-11	−1.6(12)	−0.2(4)			
CH_3-10	5.9(10)	1.7(4)			
CH_3-8	−1.1(5)	2.5(5)			

Table A.3.: $^1D_{C-H}$ couplings and experimental values (in Hz) and their measurement errors for **(a)** HMIP [136] and **(b)** ibuprofen [102]. As explained in Section 3.1.2, the sign of $J + D$ is swapped for the CH-5 coupling in HMIP compared to the choice in the thesis of Sager [136]. For ibuprofen, the CH-2,6, CH-3,5 and CH_3-9,13 couplings are averages due to symmetry and CH_2-7 due to unclear assignment. For HMIP, the CH_2-6 and CH_2-8 RDCs are averages over the two C – H couplings.

	(a)			**(b)**	
	R-HMIP	*S*-HMIP		*R*-ibuprofen	*S*-ibuprofen
CH-5	−44.6(150)	−7.3(45)	CH-2,6	35.2	29.0
CH_2-6	−183.0(100)	−181.2(45)	CH-3,5	38.5	23.7
CH-6a	−217.5(50)	−250.7(52)	CH_2-7	40.4(1)	37.3(1)
CH_2-8	71.0(70)	42.7(74)	CH-8	42.2	36.7
CH-9	113.4(150)	70.5(45)	CH-10	77.9	77.5
CH-12	92.7(150)	35.5(111)	CH_3-9,13	2.4(1)	2.1(1)
CH-12b	−272.4(50)	−312.7(111)	CH_3-14	9.6(1)	12.0(1)
CH-1	239.7(100)	311.6(45)			
CH-2	−251.3(150)	−188.7(45)			
CH-3	327.7(150)	400.4(45)			
CH_3	46.5(50)	60.3(37)			

Table A.4.: $^1D_{C-H}$ couplings and experimental values (in Hz) and their measurement errors for 1-phenylethanol (private communication, measured by Dr. Emine Sağer and Dr. Pavleta Tzvetkova). The CH-4,8 and CH-5,7 couplings are averages due to symmetry.

	R-1-phenylethanol	*S*-1-phenylethanol
CH_3-1	−4.7(8)	18.2(10)
CH-2	27.4(15)	22.2(5)
CH-4,8	−9.9(15)	−9.4(8)
CH-5,7	−9.9(15)	−11.1(15)
CH-6	−86.2(15)	−69.4(10)

Table A.5.: $^1D_{C-H}$ couplings and experimental values (in Hz) and their measurement errors for β-pinene [60].

	(-)-β-pinene	(+)-β-pinene
C1-H1	7.5(7)	6.6(9)
C3-H3s	3.7(6)	2.1(6)
C3-H3a	−2.3(6)	−0.8(5)
C5-H5	−6.7(1)	−7.2(4)
C7-H7s	−0.8(5)	−2.0(6)
C7-H7a	−6.6(3)	−4.5(22)
C10-H10s	5.8(11)	6.4(15)
C10-H10a	−6.1(4)	−6.6(10)
CH_3-8	2.3(4)	2.0(8)
CH_3-9	−2.8(6)	−2.2(5)

Table A.6.: $^1D_{C-H}$ couplings and experimental values (in Hz) for **(a)** (-)-caulamidine A [107] and **(b)** (-)-strychnine [93]. For both molecules, the CH_2 RDCs are averages over the two C – H couplings.

(a) (-)-caulamidine A	
CH-23	6.9
CH-24	8.1
CH-6	−1.9
CH-4	9.4
CH-1	9.1
CH-21	6.9
CH-26	−4.0
CH_2-9	0.2
CH_2-15	0.5
CH_3-29	−0.7
CH_3-17	−0.2

(b) (-)-strychnine	
CH-3	−69.8
CH-22	−1.1
CH-1	−178.9
CH-4	−182.4
CH-12	146.8
CH_2-23	92.2
CH-16	−49.6
CH-8	87.1
CH_2-20	36.3
CH_2-18	−32.6
CH-13	48.9
CH_2-17	−17.1
CH_2-11	21.2
CH-14	151.7
CH_2-15	−48.1

A.4.1. Signs of the HMIP RDCs compared to the quadrupolar coupling in deuterium

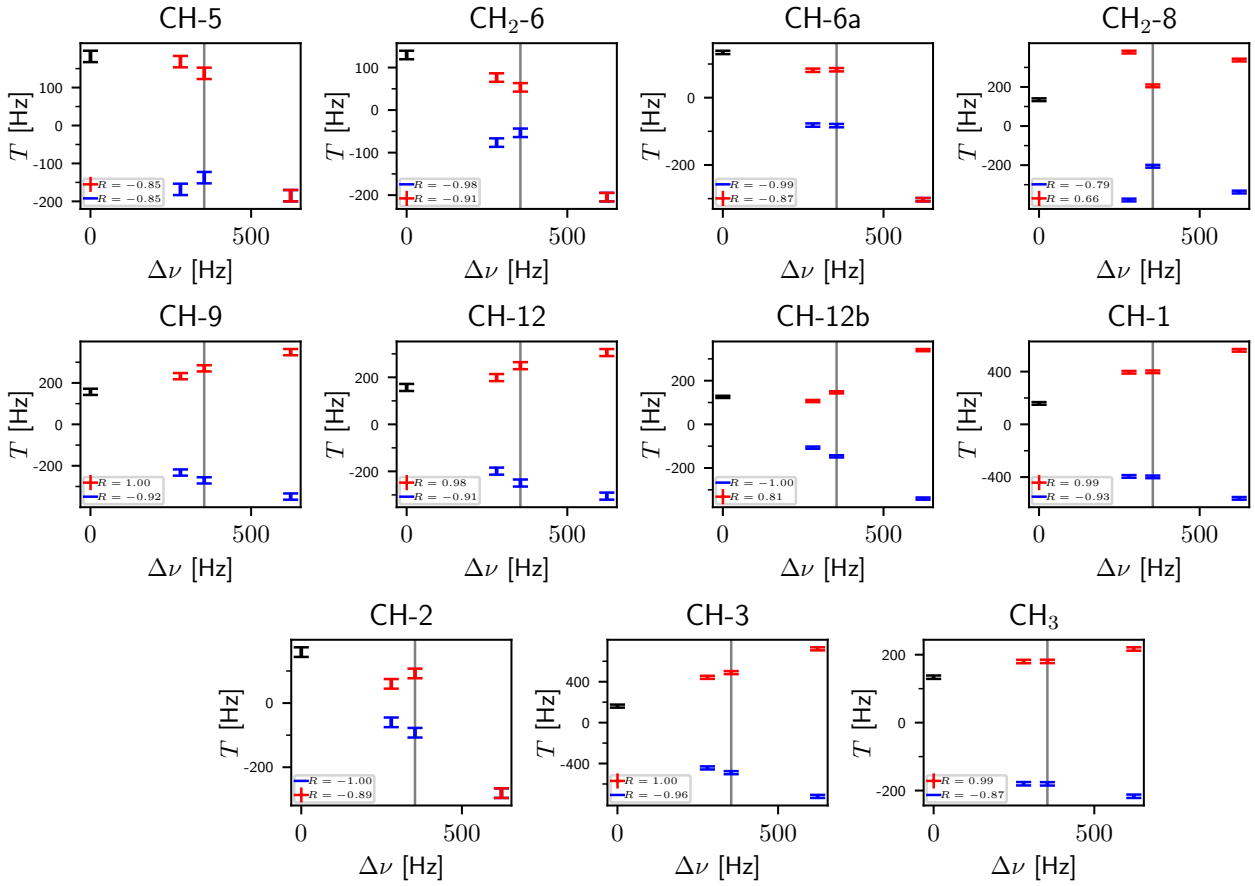


Figure A.2.: $T = J + D$ for the different measurements of D in R-HMIP with different quadrupolar couplings $\Delta\nu$ of deuterium in the deuterated chloroform reported in the thesis of Sager [136]. Shown are the two sign choices at $\Delta\nu = 354\,\mathrm{Hz}$ (negative – blue, positive – red), with the signs of the other values chosen respectively to maximize the absolute value of the correlation coefficient. The corresponding correlation coefficients are shown in the legend.

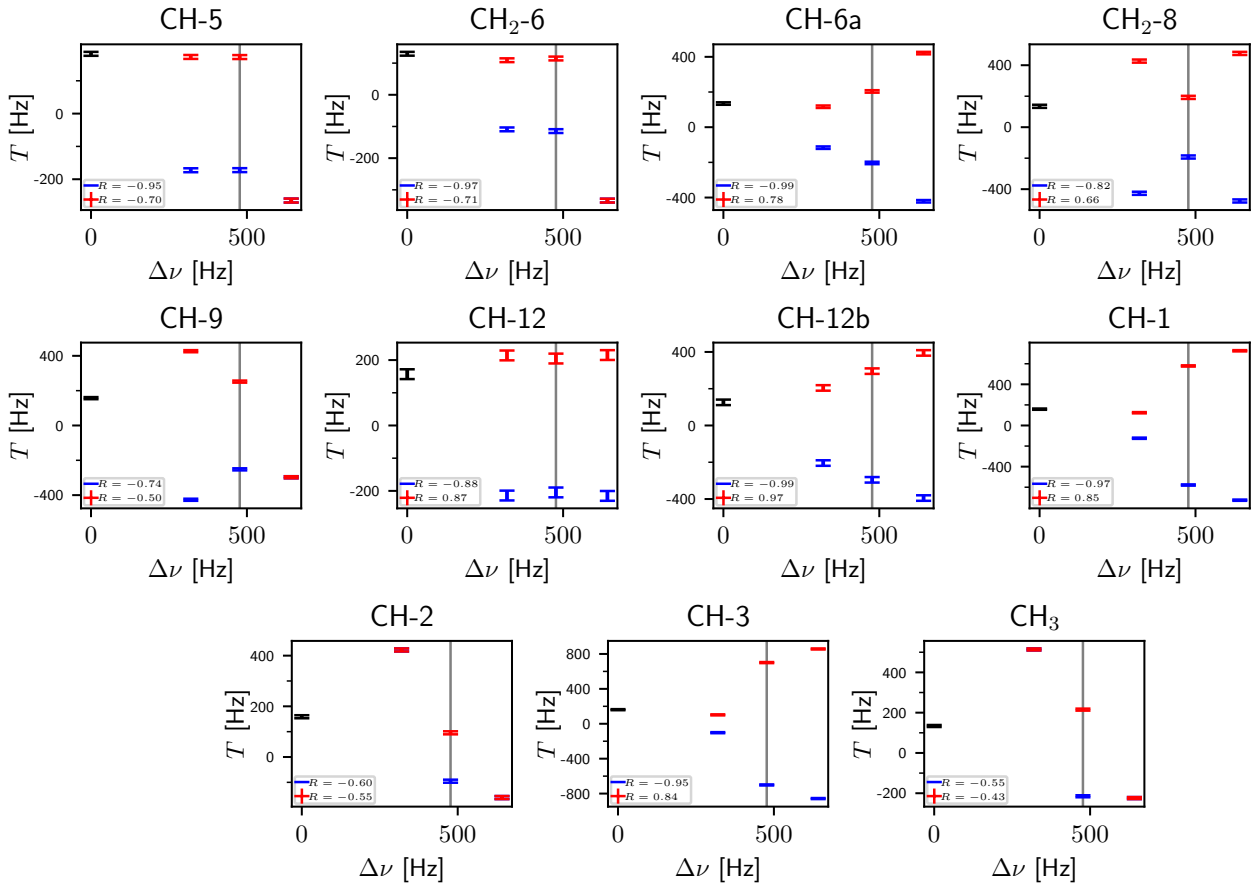


Figure A.3.: $T = J + D$ for the different measurements of D in S-HMIP with different quadrupolar couplings $\Delta\nu$ of deuterium in the deuterated chloroform reported in the thesis of Sager [136]. Shown are the two sign choices at $\Delta\nu = 477\,\mathrm{Hz}$ (negative – blue, positive – red), with the signs of the other values chosen respectively to maximize the absolute value of the correlation coefficient. The corresponding correlation coefficients are shown in the legend.

A.5. Automated workflow for the MD simulations

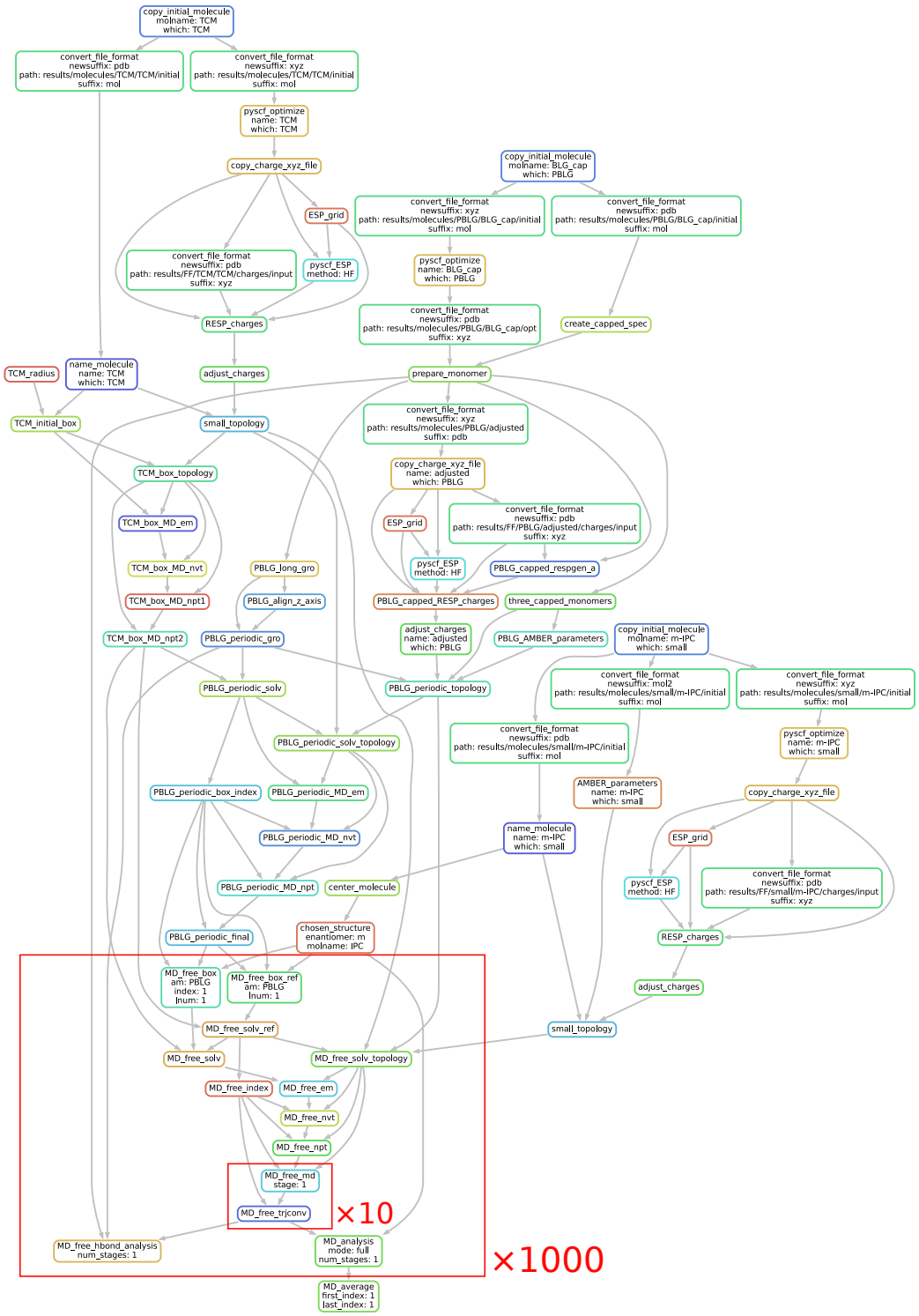


Figure A.4.: Steps in the automated workflow for a single MD chain with a simulation time of 10 ns. In the actual workflow, 1000 chains (or 300 for HMIP, ibuprofen and 1-phenylethanol) of 10 such simulations are done.

A.6. Contributions to the RDCs at different distances between the analyte and PBLG

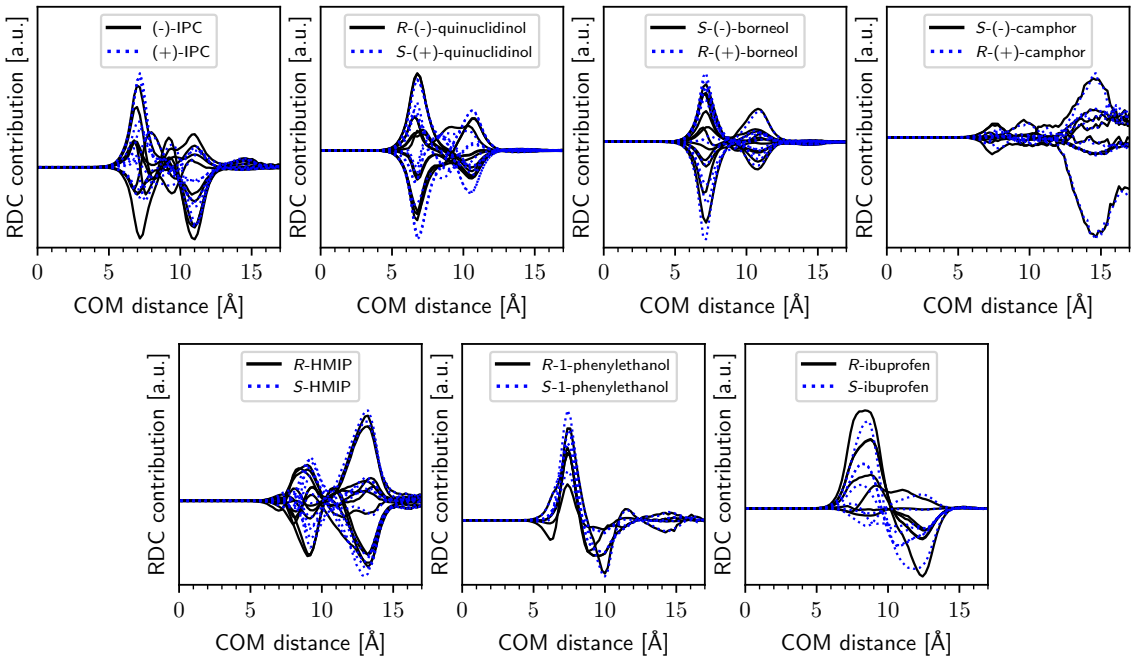


Figure A.5.: Contribution of poses at different distances between the center of mass (COM) of the analyte molecules and the backbone center of PBLG to the total RDC average.

A.7. Additional angle distributions

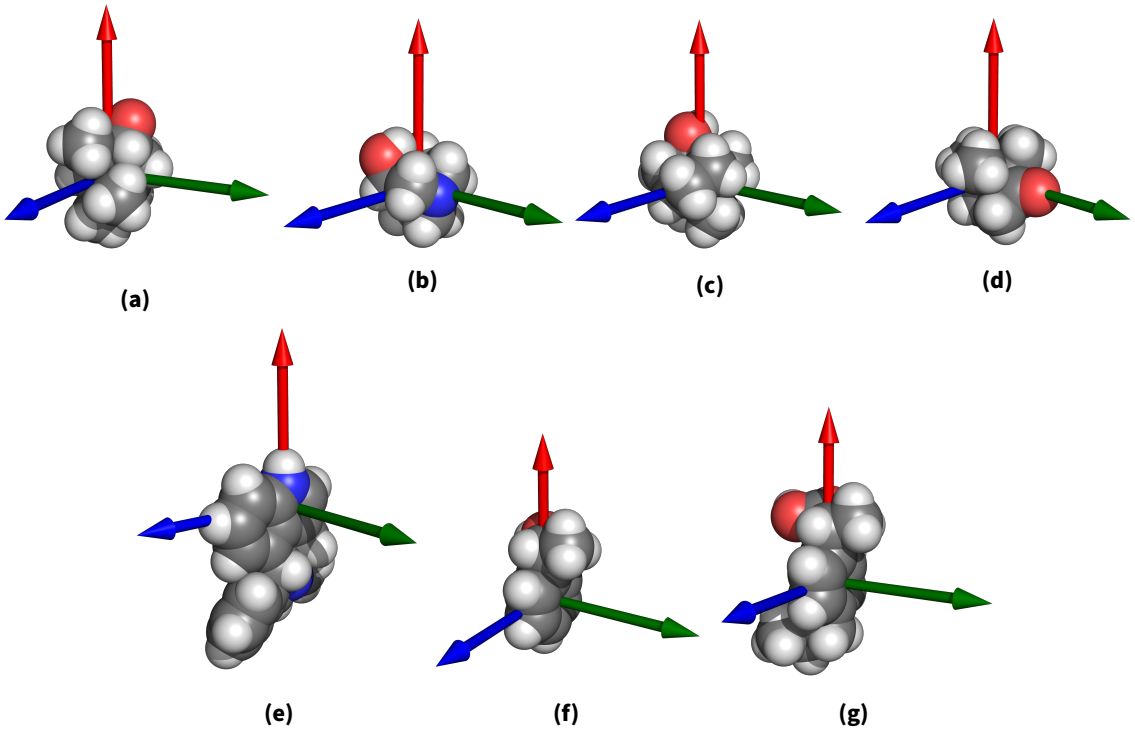


Figure A.6.: Molecular axes used for calculating the distributions in Figures 4.3 and A.8. For **(a)** (-)-IPC, **(b)** *S*-(+)-quinuclidinol, **(c)** *S*-(-)-borneol and **(d)** *S*-(-)-camphor, they are defined as the principal axes (eigenvectors of the inertia tensor) of the molecule without hydrogen atoms. For **(e)** *R*-HMIP, **(f)** *R*-1-phenylethanol and **(g)** *R*-ibuprofen, they are instead defined by the phenyl ring carbon atoms as shown in the figure.

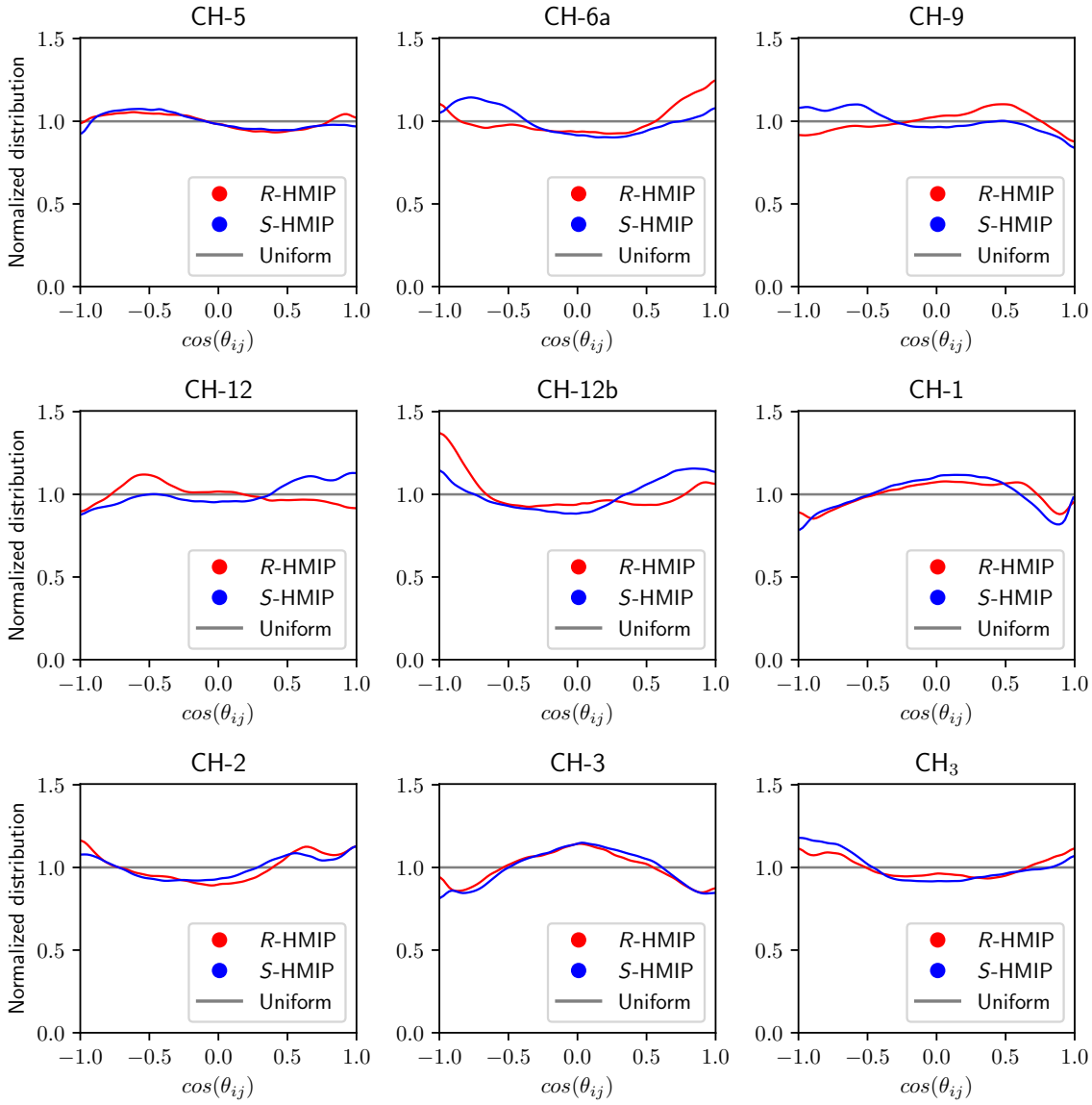


Figure A.7.: Distribution of $\cos(\theta_{ij})$ (where θ_{ij} is the angle of the C – H coupling to the z-axis) for non-averaged couplings in HMIP. For the CH_3 group, the angle of the N – C bond to the z-axis was used instead for θ_{ij}. In gray, a uniform distribution is shown, which is the distribution for isotropic movement.

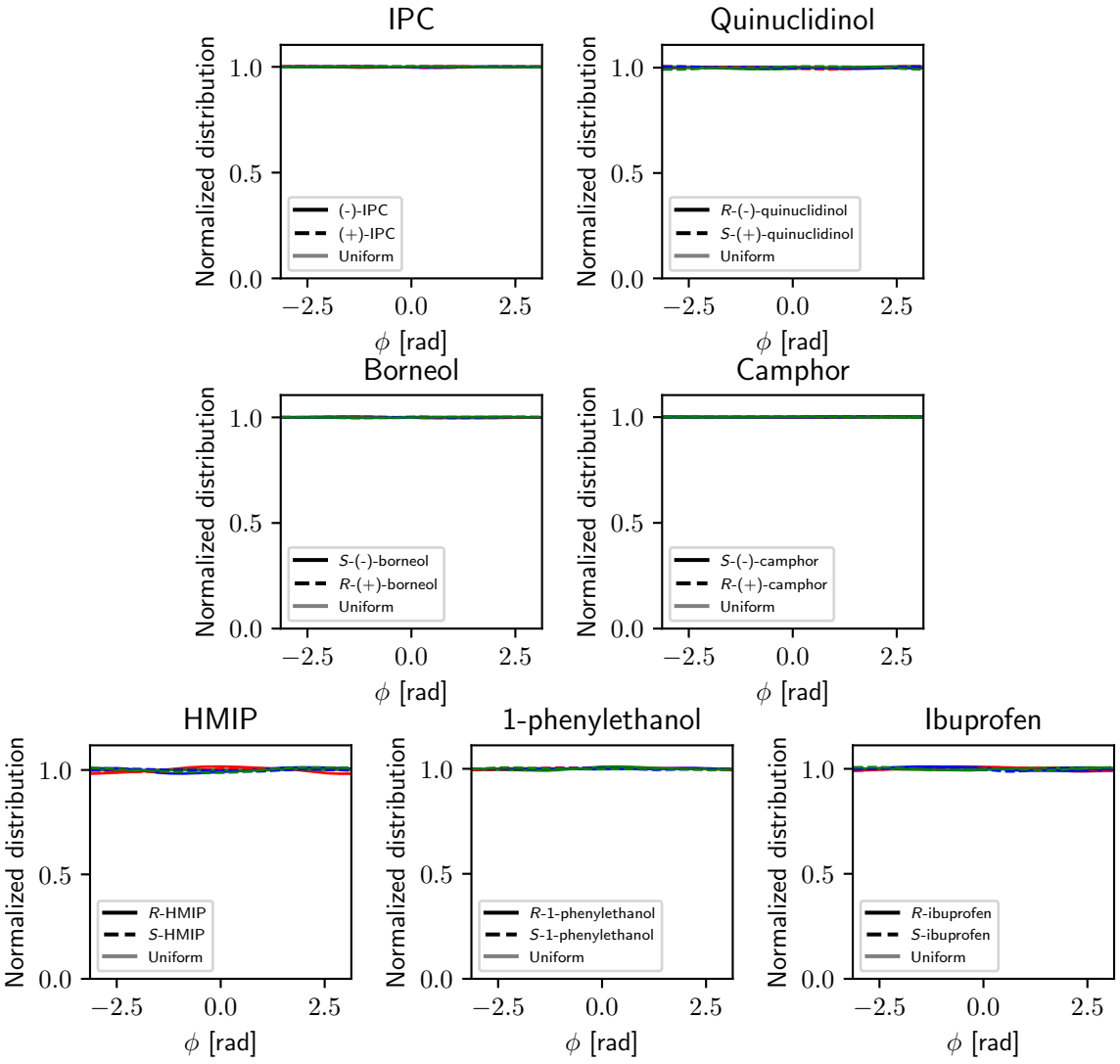


Figure A.8.: Distribution of the angle in the xy-plane ϕ for the molecule axes defined in Figure A.6 of the 7 pairs of enantiomers used in the MD simulations. Compare Figure 4.3 for the corresponding distributions of the angle θ to the z-axis. In gray, a uniform distribution is shown, which is the distribution for isotropic movement.

A.8. Comparison of additional MD averages

Table A.7.: Cosine similarity between the RDCs from MD *excluding* the poses with hydrogen bond to PBLG and the full MD average (S_{MD}), the average over the poses with hydrogen bond (S_{hbond}) and the experimental RDCs (S_{exp}).

Molecule	S_{MD}	S_{hbond}	S_{exp}
(-)-IPC	1.00	0.98	0.68
(+)-IPC	1.00	0.97	0.87
R-(-)-quinuclidinol	1.00	0.99	0.17
S-(+)-quinuclidinol	1.00	0.99	0.29
S-(-)-borneol	1.00	0.99	-0.33
R-(+)-borneol	1.00	0.99	0.40
R-HMIP	1.00	0.99	1.00
S-HMIP	1.00	1.00	0.98
R-1-phenylethanol	0.99	0.75	0.25
S-1-phenylethanol	1.00	0.92	-0.05
R-ibuprofen	1.00	1.00	0.72
S-ibuprofen	1.00	0.99	-0.54

Table A.8.: Cosine similarity between the RDCs from MD using only the poses where the analyte and PBLG molecules overlap (within $0.5(\sigma_i + \sigma_j)$) and the RDCs calculated from MD using the full average (S_{MD}), the RDCs from MD without overlap ($S_{no\ overlap}$), the RDCs from MD using only the poses with hydrogen bond to PBLG (S_{hbond}), as well as the experimental RDCs (S_{exp}).

Molecule	S_{MD}	$S_{no\ overlap}$	S_{hbond}	S_{exp}
(-)-IPC	0.98	0.01	1.00	0.49
(+)-IPC	0.97	0.48	1.00	0.70
R-(-)-quinuclidinol	1.00	0.39	1.00	0.17
S-(+)-quinuclidinol	1.00	0.21	1.00	0.31
S-(-)-borneol	0.99	-0.31	1.00	-0.45
R-(+)-borneol	1.00	-0.09	1.00	0.33
S-(-)-camphor	0.74	0.62		0.63
R-(+)-camphor	0.81	0.71		0.74

A.9. Results for 8 analyte molecules in the simulation box

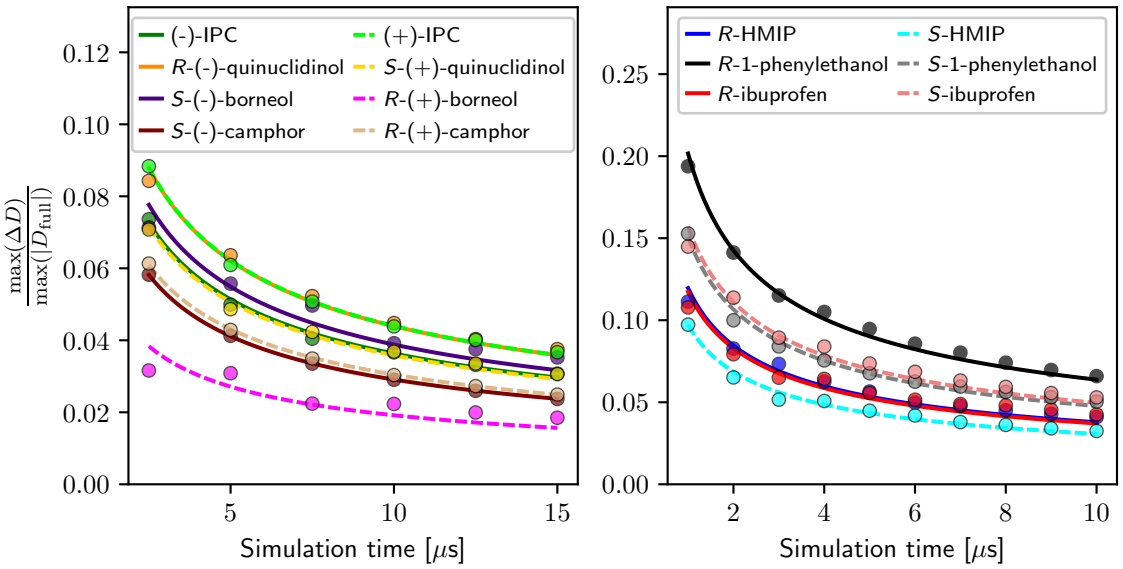


Figure A.9.: Maximal uncertainty of the RDC average relative to the maximum absolute value of the RDCs for different simulation times with 8 molecules in the simulation box. Compare with Figure 3.4.

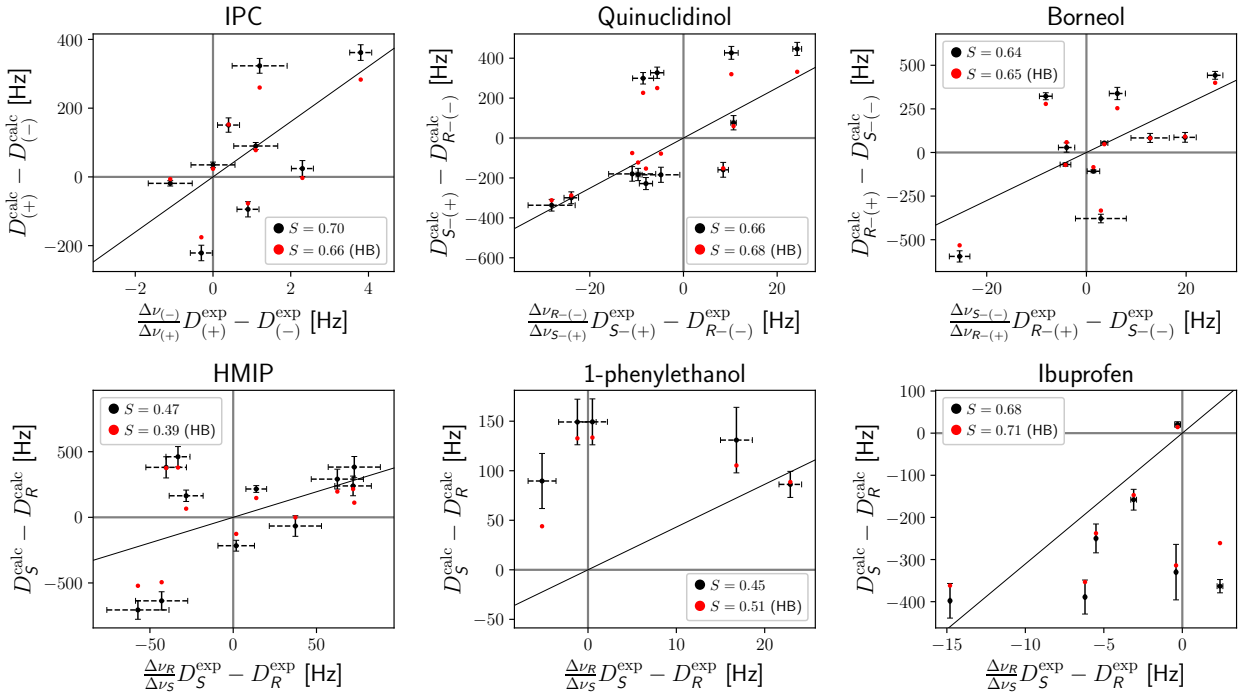


Figure A.10.: Comparison of the CDDA (difference between RDCs) obtained from the experimental RDCs and the RDCs obtained from the MD simulations with 8 analyte molecules in the simulation box for the 6 analyte molecules with hydrogen bond donor. Refer to Figure 4.13 for an explanation of the colors.

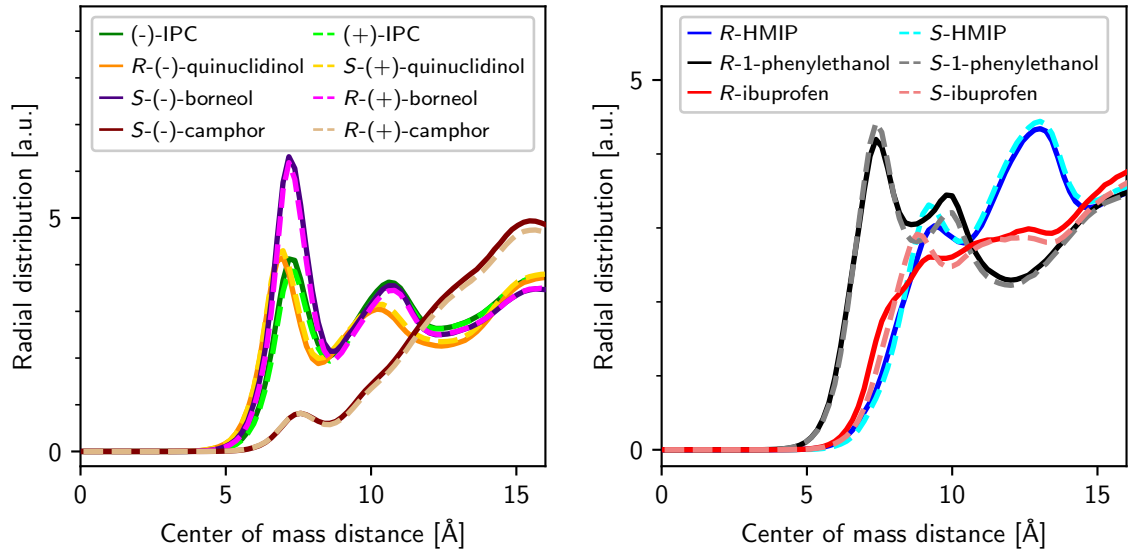


Figure A.11.: Distribution of the center of mass distance of the analyte molecules to the PBLG center in the simulations with 8 analyte molecules in the simulation box.

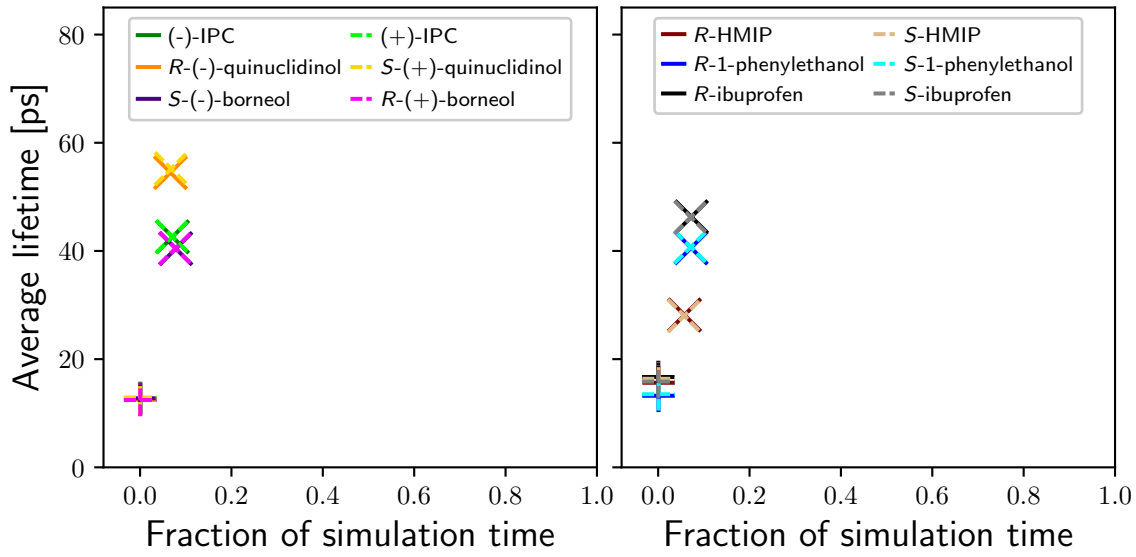


Figure A.12.: Distribution of the O – H ⋯ O hydrogen bond distances and angles in the MD simulations for the 6 pairs of enantiomers with hydrogen bond donor calculated from the simulations with 8 analyte molecules in the simulation box.

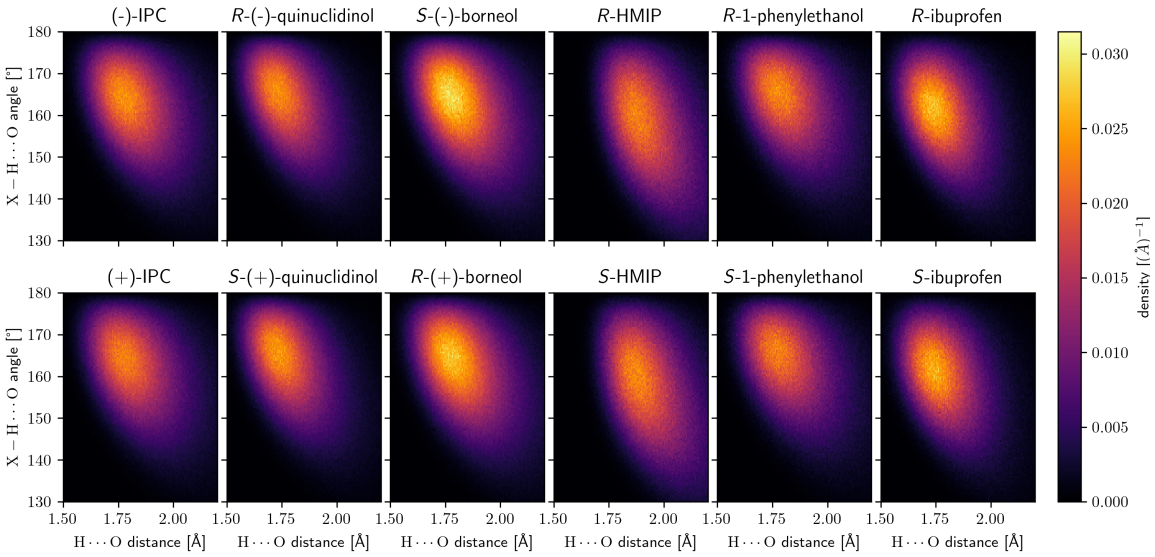


Figure A.13.: Distribution of the O – H ··· O hydrogen bond distances and angles in the MD simulations for IPC, quinuclidinol and borneol.

A.10. Comparing MD predictions

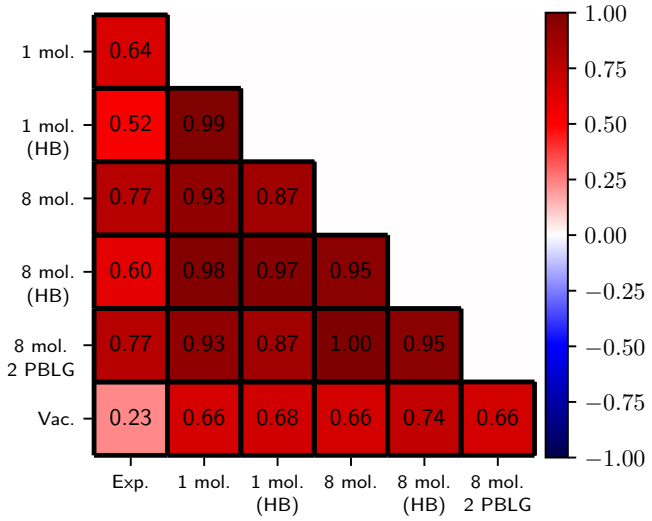


Figure A.14.: Cosine similarities for (-)-IPC between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

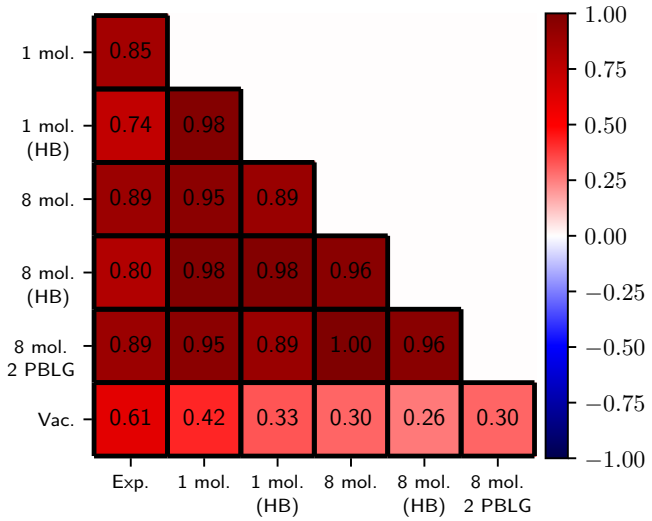


Figure A.15.: Cosine similarities for (+)-IPC between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

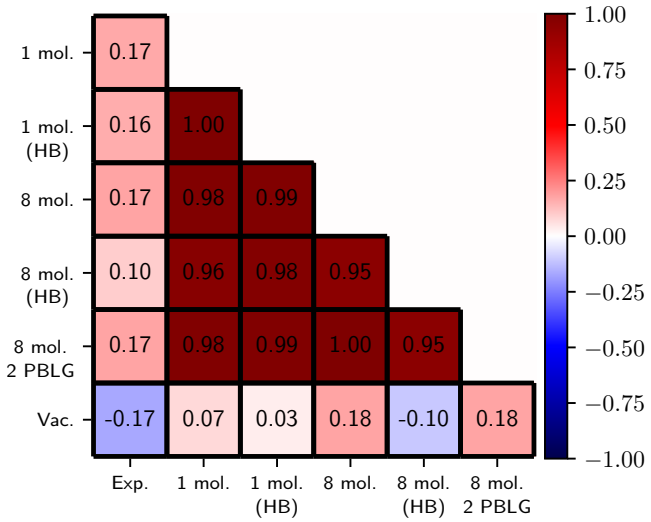


Figure A.16.: Cosine similarities for *R*-(-)-quinuclidinol between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

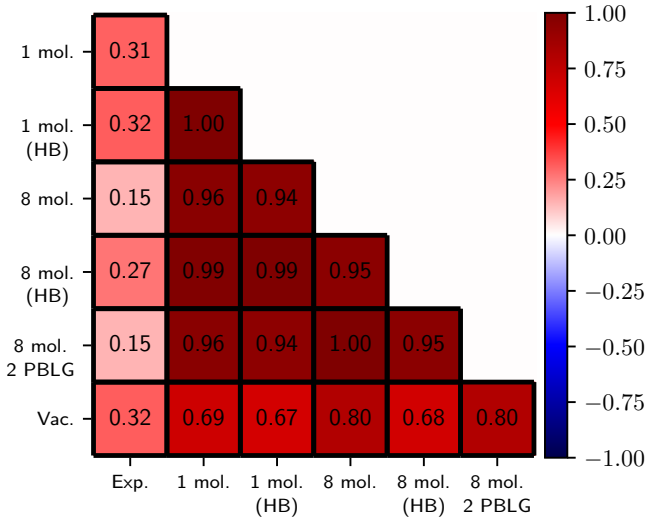


Figure A.17.: Cosine similarities for *S*-(+)-quinuclidinol between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

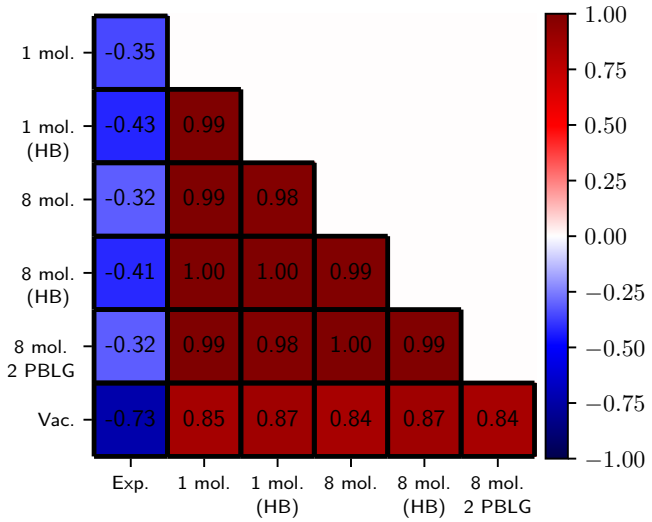


Figure A.18.: Cosine similarities for *S*-(-)-borneol between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

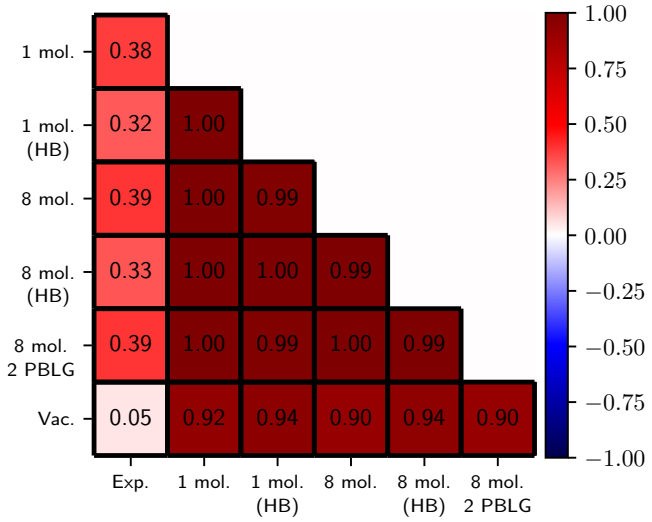


Figure A.19.: Cosine similarities for *R*-(+)-borneol between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

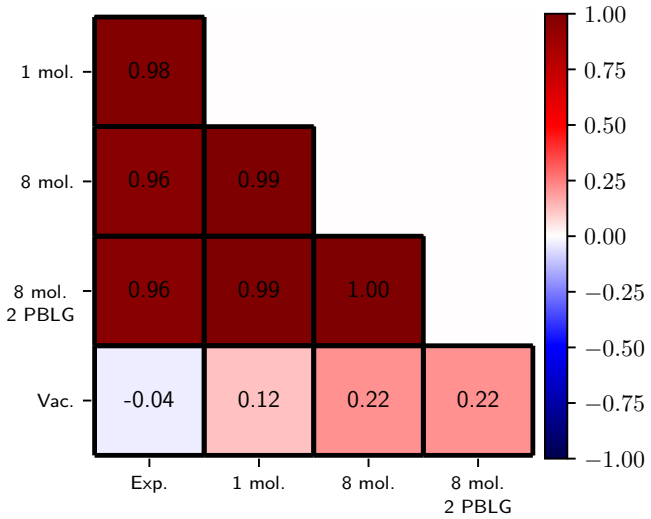


Figure A.20.: Cosine similarities for *S*-(-)-camphor between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

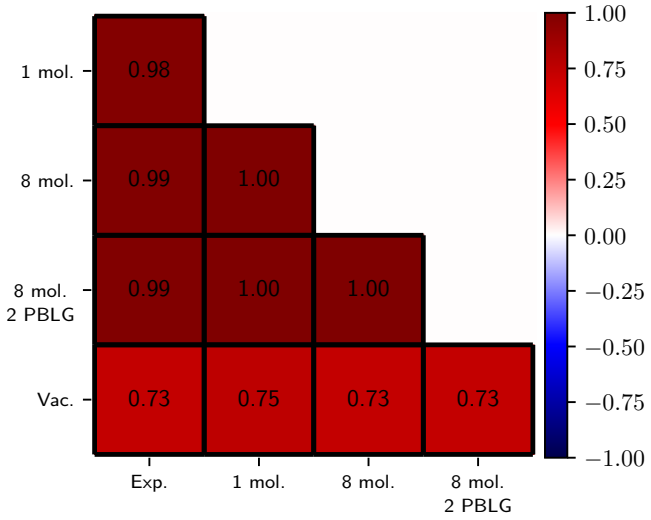


Figure A.21.: Cosine similarities for *R*-(+)-camphor between experimental RDCs, simulated RDCs with 1 analyte molecule and 1 PBLG chain, RDCs only from hydrogen-bonded poses, simulated RDCs with 8 analyte molecules, simulated RDCs with 8 analyte molecules and 2 PBLG chains, and the cylinder prediction.

A.11. Impact of water molecules on the hydrogen bond distribution

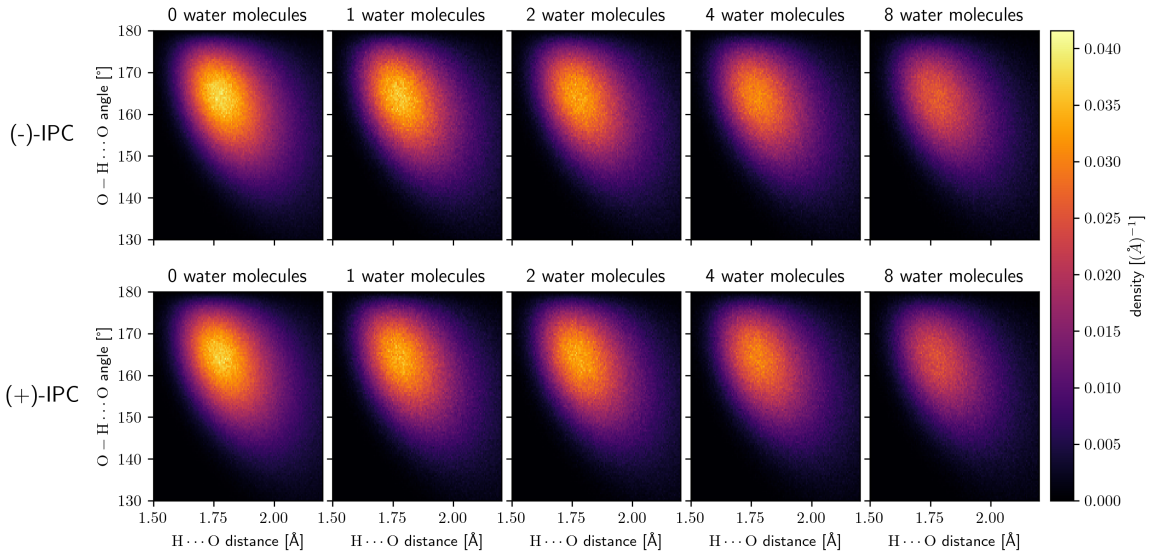


Figure A.22.: Distribution of the O – H · · · O hydrogen bond distances and angles in the MD simulations for (-)- and (+)-IPC with a single analyte molecule and different numbers of water molecules in the simulation.

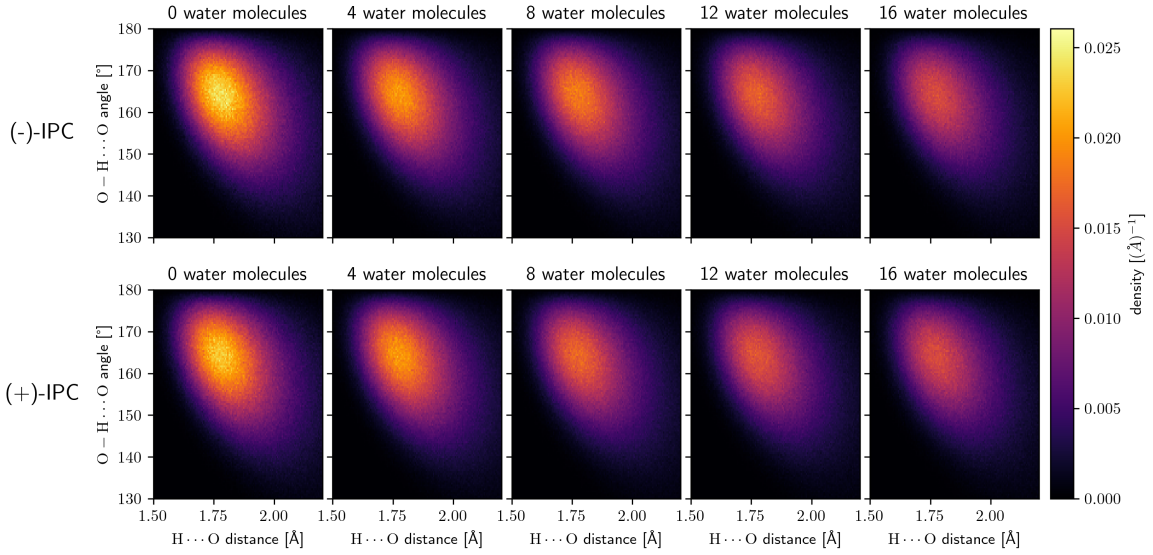


Figure A.23.: Distribution of the O – H · · · O hydrogen bond distances and angles in the MD simulations for (-)- and (+)-IPC with 8 analyte molecules and different numbers of water molecules in the simulation.

A.12. Fits for the nonpolar solvation model with the other choices for the decomposition and RDF

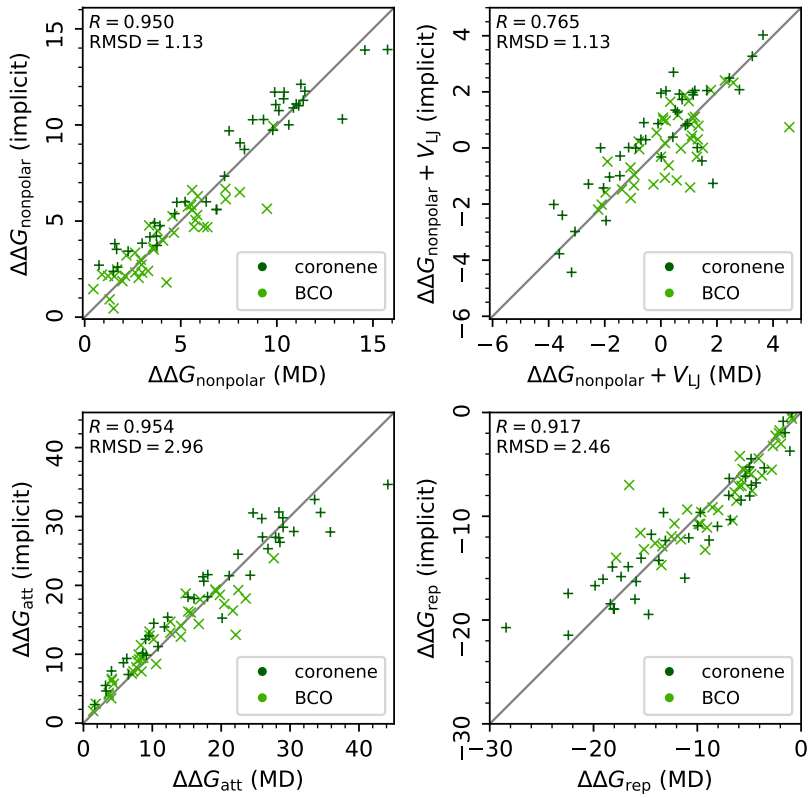


Figure A.24.: Nonpolar fit results for $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ and its contributions with the WCA decomposition and the "calc." RDF. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

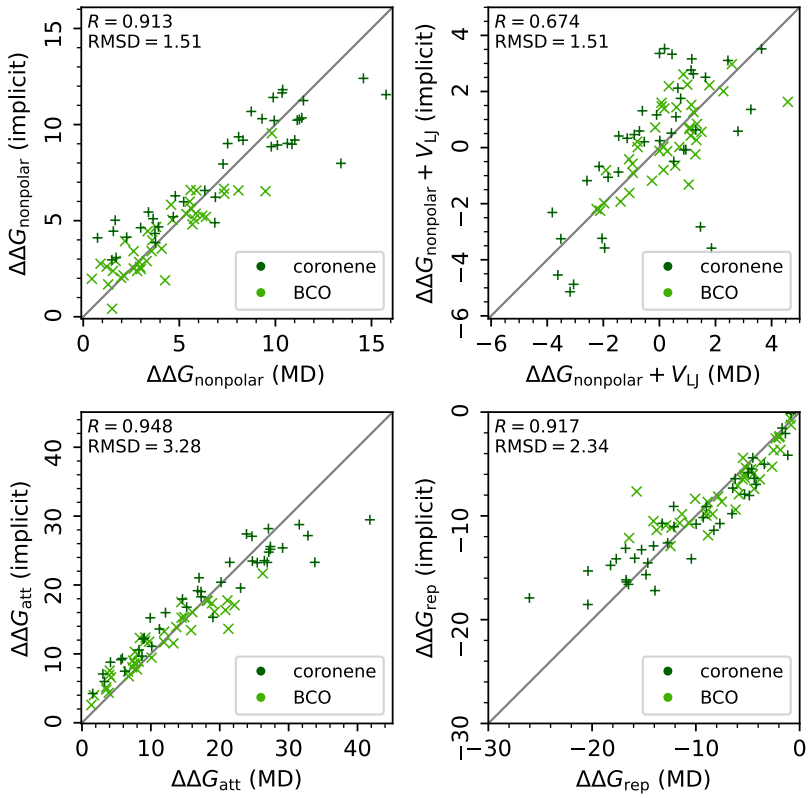


Figure A.25.: Nonpolar fit results for $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ and its contributions with the σ decomposition and the "step" RDF. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

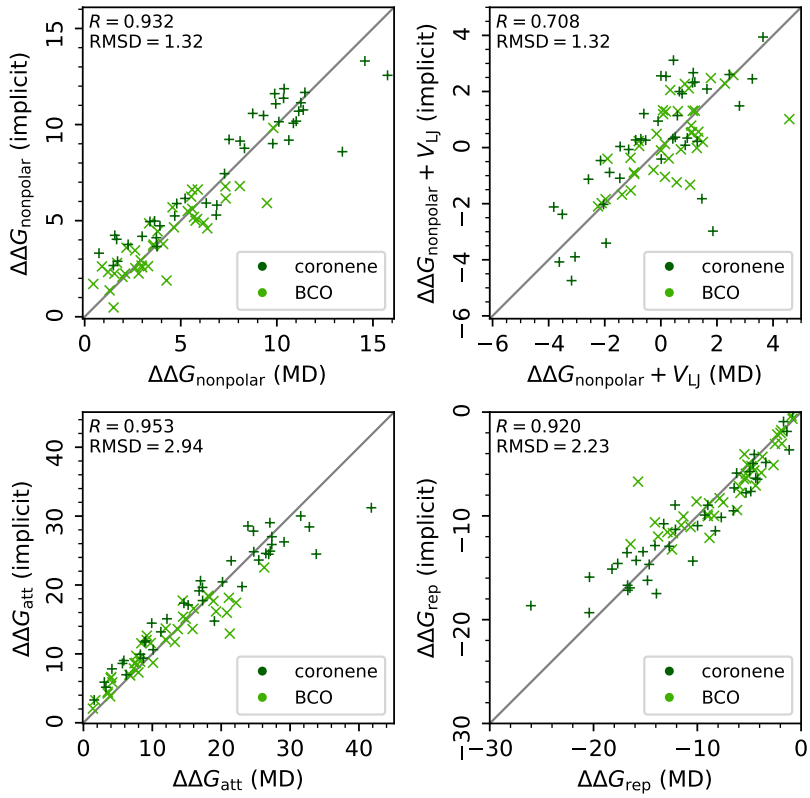


Figure A.26.: Nonpolar fit results for $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ and its contributions with the σ decomposition and the "calc." RDF. The correlation coefficient and RMSD (in kcal mol^{-1}) are given in the respective upper left corner.

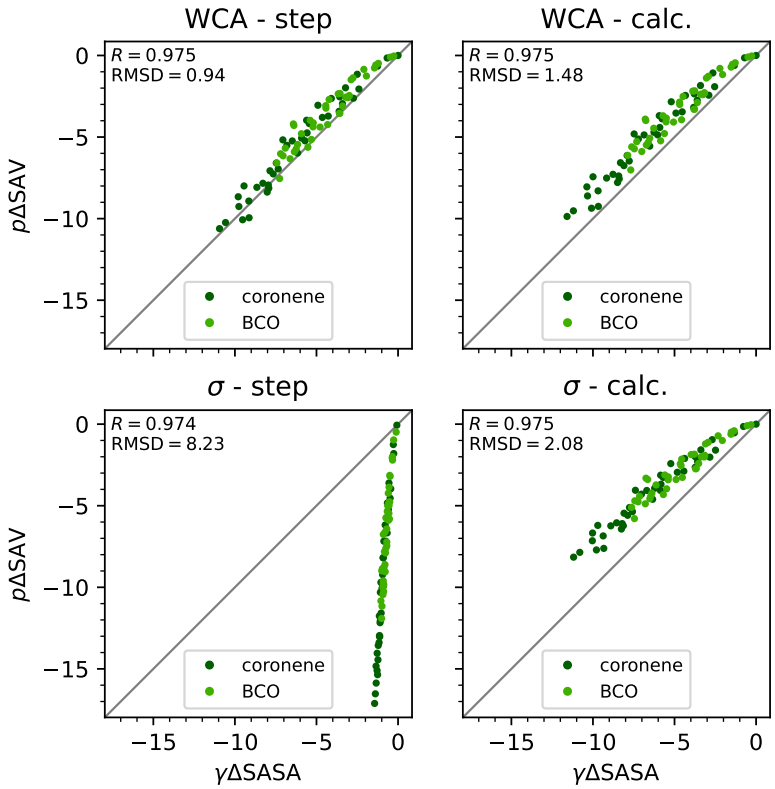


Figure A.27.: Correlation of the $\gamma\Delta$SASA and $p\Delta$SAV contributions to $\Delta\Delta G_{\text{rep}}$ for the different decompositions and RDFs.

A.13. Numerical parameters for the implicit solvation models

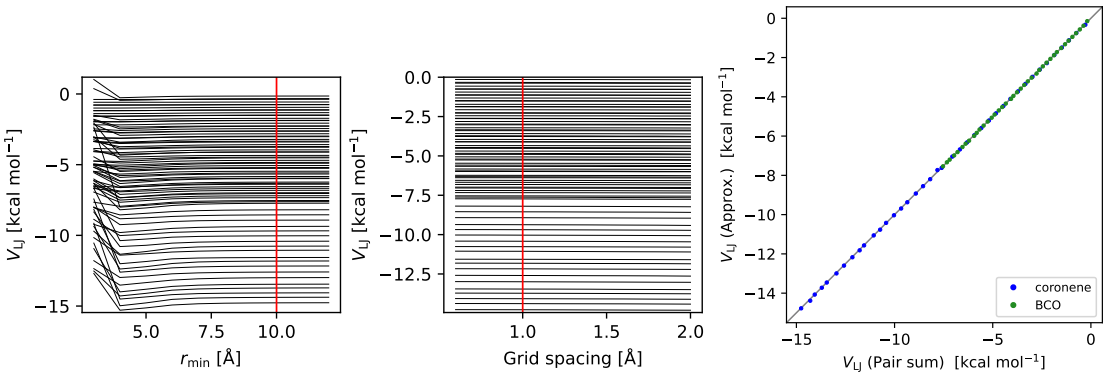


Figure A.28.: Comparison of numerical parameters for the calculation of V_{LJ}. (a) Dependence of V_{LJ} on r_{min} and the grid spacing. In red, the chosen parameters are marked. (b) Comparison of the pairwise sum with the numerical approximation.

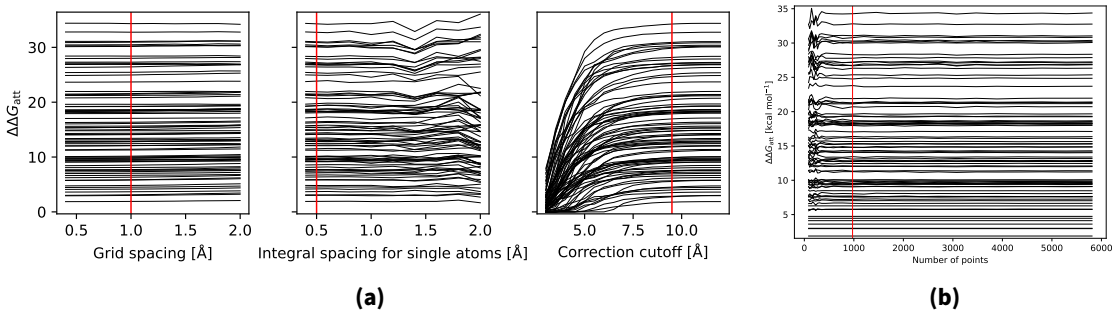


Figure A.29.: Numerical parameters for the calculation of $\Delta\Delta G_{att}$. In red, the chosen parameters are marked. (a) Dependence of $\Delta\Delta G_{att}$ on the grid spacing, the grid spacing used to calculate the integrals for single atoms around the respective other solute, and the cutoff used for the correction in Equation (3.14). (b) Dependence of $\Delta\Delta G_{att}$ on the number of Lebedev quadrature points used for each atom if it were isolated.

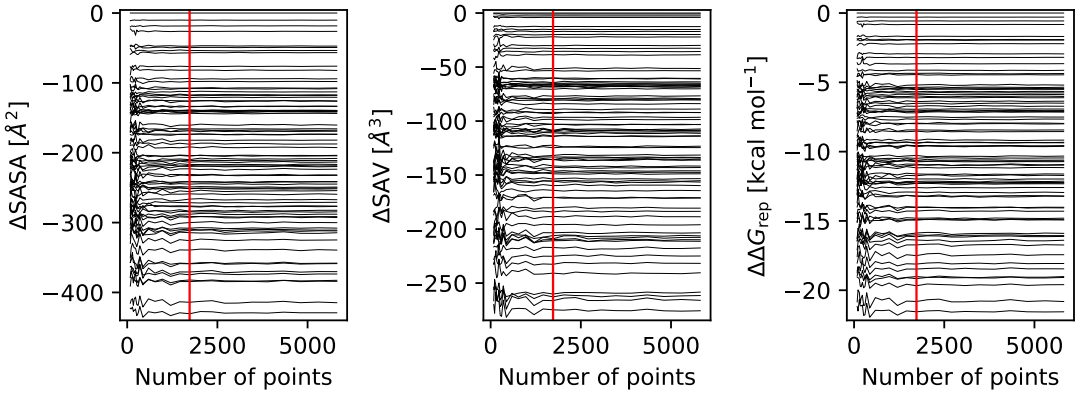


Figure A.30.: Dependence of ΔSASA, ΔSAV and $\Delta\Delta G_{\text{rep}}$ on the number of Lebedev quadrature points used for each atom if it were isolated. In red, the chosen parameters are marked.

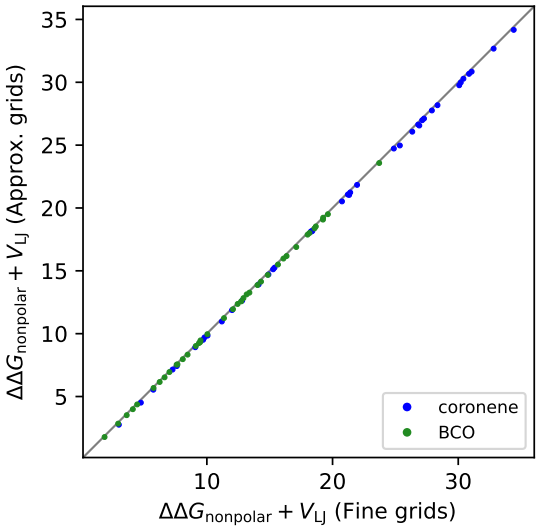


Figure A.31.

Figure A.32.: Comparison of $\Delta\Delta G_{\text{nonpolar}} + V_{\text{LJ}}$ between the chosen numerical approximation and more accurate values. For V_{LJ}, the latter means the exact pairwise sum, while for $\Delta\Delta G_{\text{rep}}$ and $\Delta\Delta G_{\text{att}}$ it means the finest grids shown in Figures A.29 and A.30.

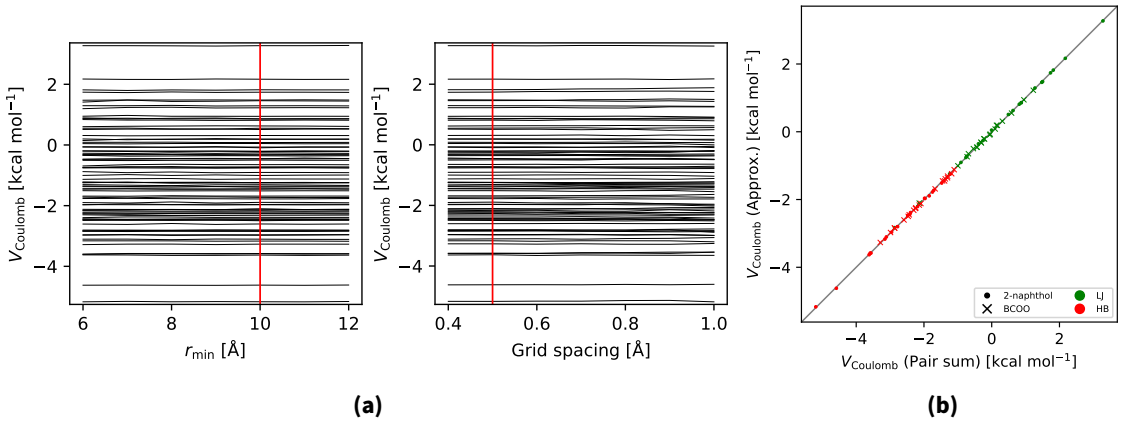


Figure A.33.: **(a)** Comparison of the numerically evaluated V_{Coulomb} for different r_{min} and grid spacings. **(b)** Comparison of V_{Coulomb} between the numerical approximation and the full pair sum.

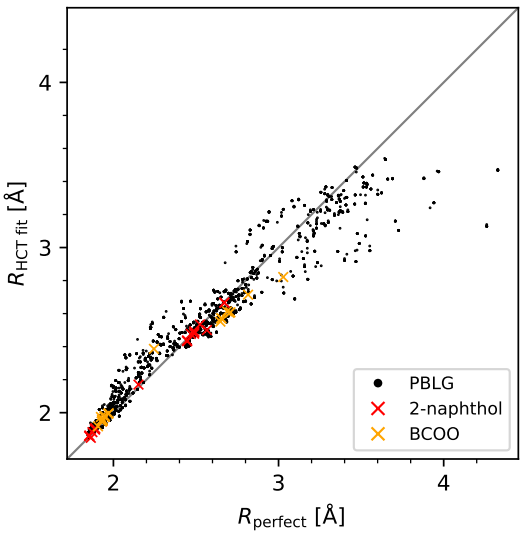


Figure A.34.: Comparison of the Born radii for the fit of the HCT scaling factors to the perfect Born radii of the isolated molecules PBLG, 2-naphthol and BCOO.

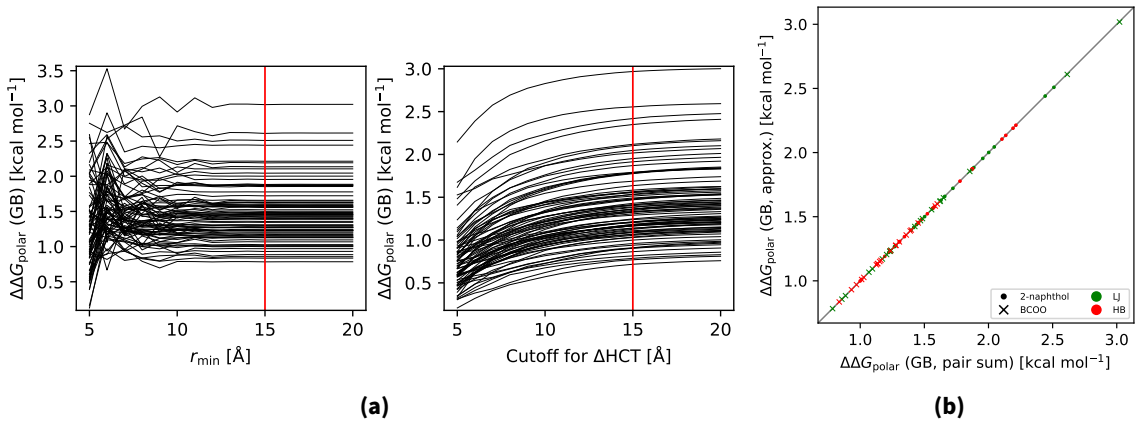


Figure A.35.: **(a)** Comparison of the numerically evaluated $\Delta\Delta G_{\text{polar}}$ from the ALPB-ΔHCT model for different r_{min} and cutoffs. **(b)** Comparison of $\Delta\Delta G_{\text{polar}}$ between the numerical approximation and the full pair sums for both the ΔHCT Born radii and the ALPB equation.

A.14. Influence of ϵ_{in} on the polar contribution to the binding free energy

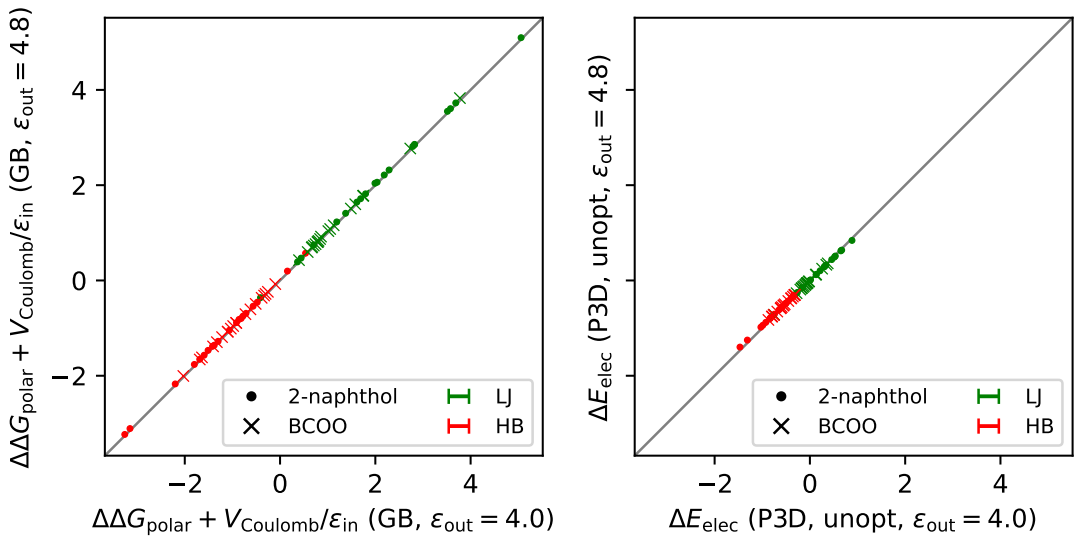


Figure A.36.: Difference of the polar binding free energy of the ALBP-ΔHCT model and the "P3D, unopt" model between $\epsilon_{out} = 4.55$ (obtained from MD) and $\epsilon_{in} = 4.8$ (from experiment). All free energy values are given in kcal mol^{-1}.

A.15. Distribution of the COM z-position in the MC calculations

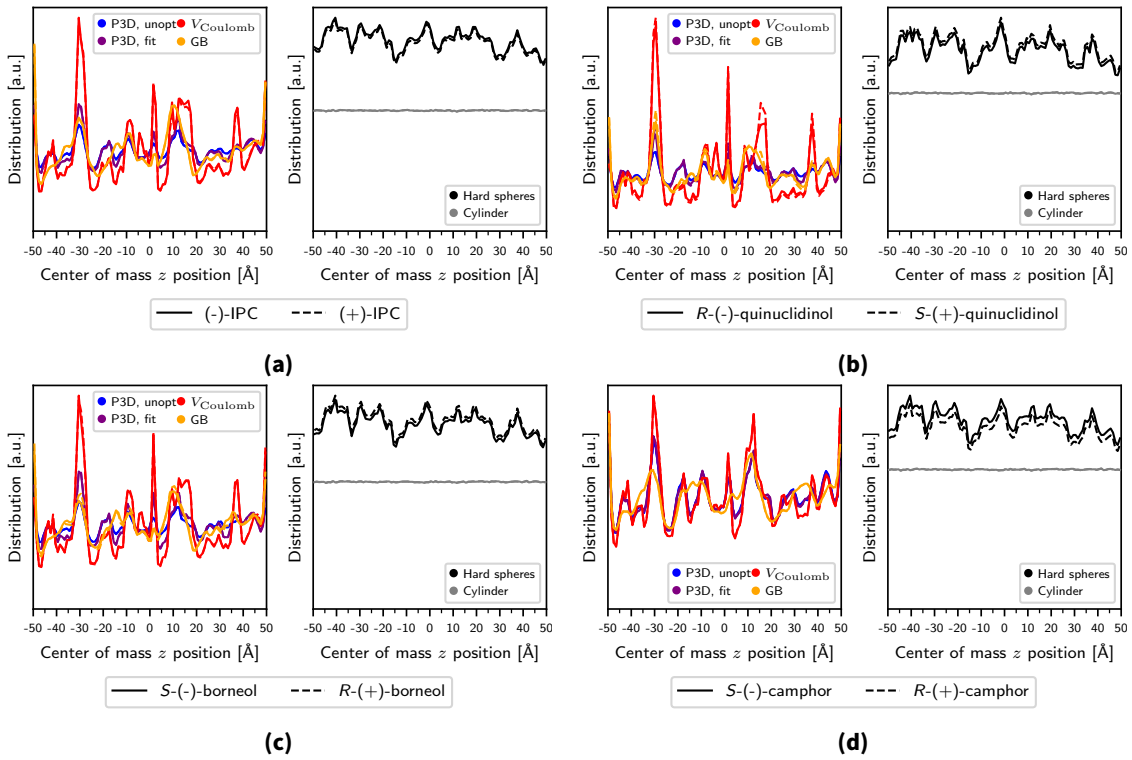


Figure A.37.: Distribution of the center of mass z position of the analyte molecules obtained from the different MC calculations for **(a)** IPC, **(b)** quinuclidinol, **(c)** borneol and **(d)** camphor. The normalization of the respective left part of the figures (in color) are the same, while the distribution of the "hard spheres" model was scaled to match the RDCs of the "cylinder" model.

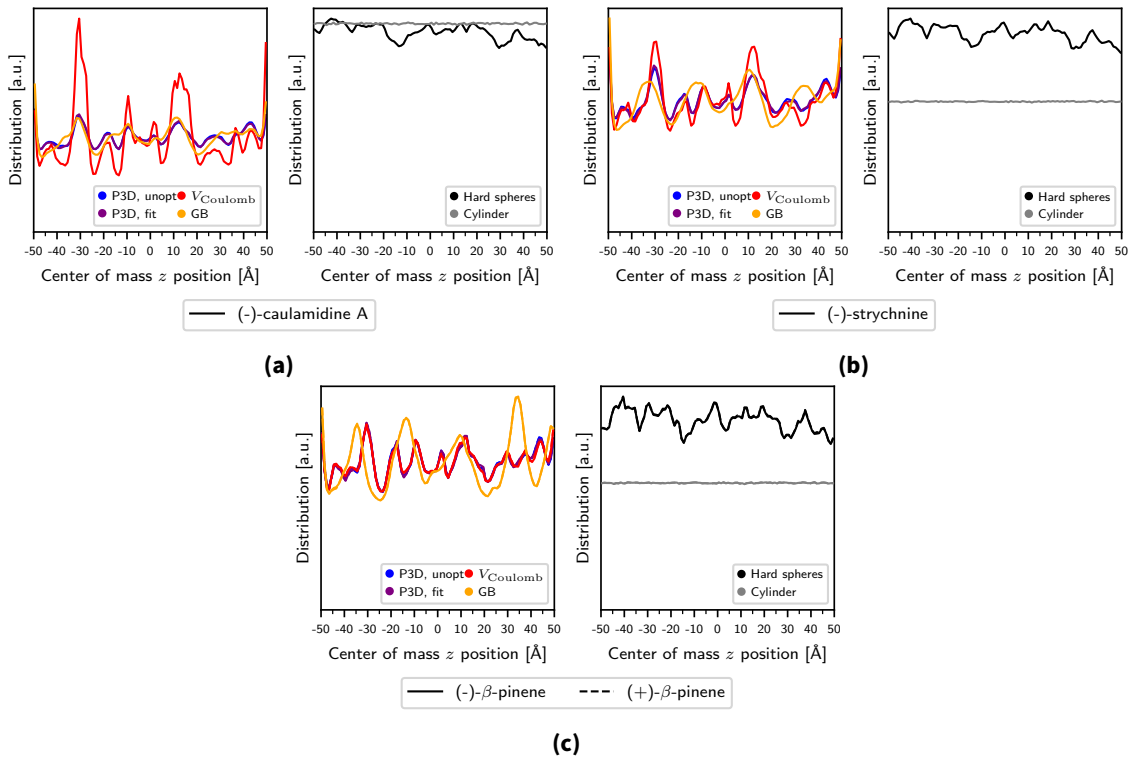


Figure A.38.: Distribution of the center of mass z position of the analyte molecules obtained from the different MC calculations for **(a)** (-)-caulamidine A, **(b)** (-)-strychnine and **(c)** (-) and (+)-β-pinene. The normalization of the respective left part of the figures (in color) are the same, while the distribution of the "hard spheres" model was scaled to match the RDCs of the "cylinder" model.

A.16. Binding (free) energy distribution for the MC calculations

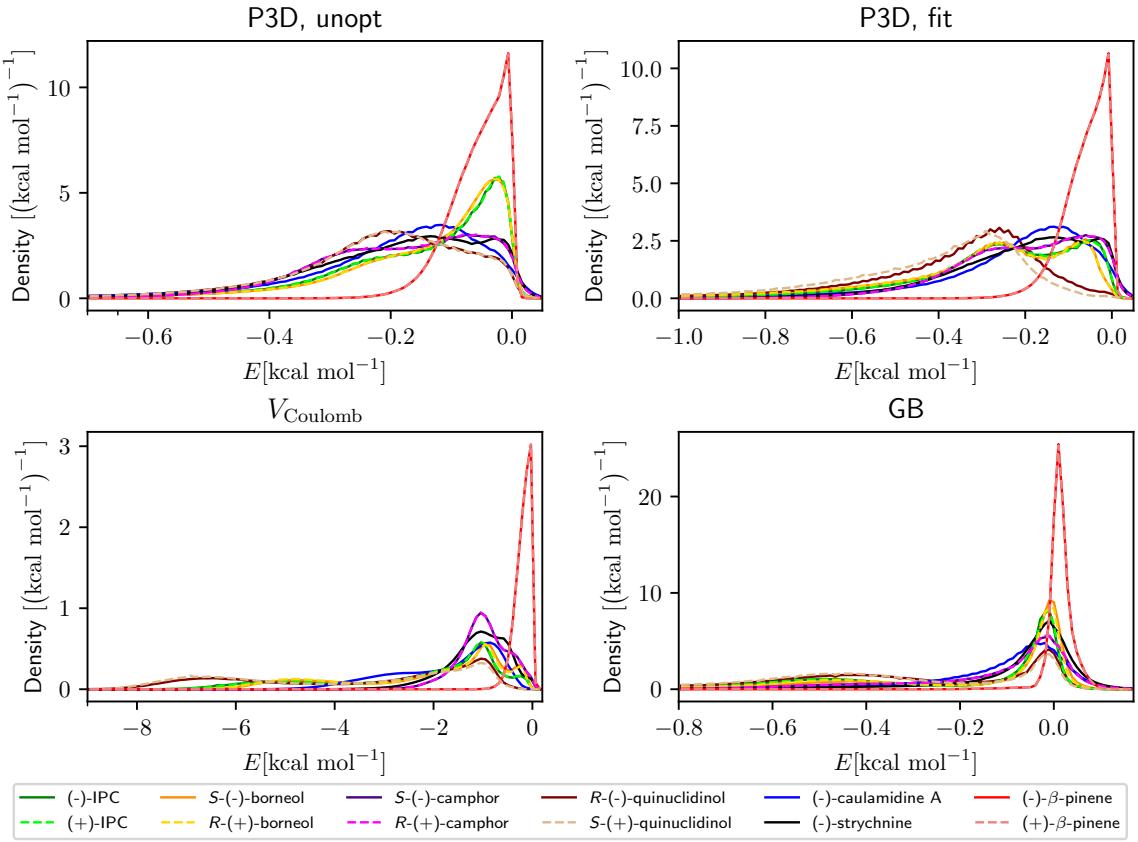


Figure A.39.: Distribution of the binding free energies to PBLG for the different models used in the MC calculations. For each model, the normalization is the same.

B. List of publications

1. Elsing, D., Luy, B., and Kozlowska, M. "Enantiomer Differentiation by Interaction-Specific Prediction of Residual Dipolar Couplings in Spherical-like Molecules". In: *Journal of Chemical Theory and Computation* 20.15 (Aug. 2024), pp. 6454–6469. DOI: 10.1021/acs.jctc.4c00441. URL: http://dx.doi.org/10.1021/acs.jctc.4c00441
2. Elsing, D., Wenzel, W., and Kozlowska, M. "An Implicit Solvation Model for Binding Free Energy Estimation in Nonaqueous Solution". In: *The Journal of Physical Chemistry B* 129.6 (Jan. 2025), pp. 1874–1889. DOI: 10.1021/acs.jpcb.4c08592. URL: http://dx.doi.org/10.1021/acs.jpcb.4c08592
3. Elsing, D.; Luy, B; Kozlowska, M. Monte Carlo prediction of residual dipolar couplings in analyte molecules aligned by poly-γ-benzyl-L-glutamate. (In preparation)
4. Elsing, D.*; Hwang, S.*; Klenin, K.*; Tzvetkova, P.*; Sager, E.; Freudenberger, J. C.; Reggelin, M.; Kondov, I.; Kozlowska, M.; Sun, H.; Luy, B et al. Absolute Configuration Determination using Chirality-Driven Difference in Alignment. (In preparation) *equal contribution, order alphabetical

List of publications unrelated to the thesis

1. Liu, M., Elsing, D., Esmaeilpour, M., Kozlowska, M., Wenzel, W., and Wöll, C. "Layer-By-Layer Assembly of Asymmetric Linkers into Non-Centrosymmetric Metal Organic Frameworks: A Thorough Theoretical Treatment". In: *Advanced Functional Materials* 34.20 (May 2023). DOI: 10.1002/adfm.202302516. URL: http://dx.doi.org/10.1002/adfm.202302516
2. Otterbach, S. A., Elsing, D., Schulz, A. D., Tappert, H., Wenzel, W., Kozlowska, M., Röhm, H., and Bräse, S. "Pseudo-Para-Substituted [2.2]Paracyclophanes for Hole Transport in Perovskite Solar Cells". In: *Advanced Functional Materials* 34.47 (Nov. 2023). DOI: 10.1002/adfm.202309226. URL: http://dx.doi.org/10.1002/adfm.202309226
3. Schulz, A. D., Otterbach, S. A., Tappert, H., Elsing, D., Wenzel, W., Kozlowska, M., Bräse, S., Colsmann, A., and Röhm, H. "Doping Strategies for Tetrasubstituted Paracyclophane Hole Transport Layers in Perovskite Solar Cells". In: *Advanced Functional Materials* 34.47 (May 2024). DOI: 10.1002/adfm.202402110. URL: http://dx.doi.org/10.1002/adfm.202402110

C. Acknowledgments

First of all, I would like to thank my supervisors Dr. Mariana Kozlowska, Prof. Dr. Burkhard Luy and Prof. Dr. Wolfgang Wenzel for the opportunity of working on an interesting project and for their continued advice and suggestions, while supporting me in the pursuit of my own research ideas. Additionally, I thank Dr. Mariana Kozlowska for her corrections of the thesis.

I also want to thank express my gratitude to all my colleagues, including Dr. Anna Mauri, Dr. Srdjan Pusara, Helmy Pacheco Hernández, Aleksandr Avdoshin, Dr. Yohanes Pramudya and Dr. Ali Deniz Özdemir, for their help and advice, both technical and non-technical.

Furthermore, I would like to thank my collaborators, especially Prof. Dr. Han Sun and Dr. Songhwan Hwang for their discussions and suggestions.

Finally, I thank my friends and family, especially my parents, brothers and my girlfriend Claire, for their large amounts of support throughout the years.