

DISS. ETH NO. 28346

CONFORMATIONAL ENSEMBLES IN SOLUTION STUDIED BY NMR AND COMPUTATIONAL METHODS

A dissertation submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Thomas Hanspeter Stadelmann

MSc ETH Interdisciplinary Sciences, ETH Zurich

born on 28.06.1992

Citizen of Mauensee (LU), Switzerland

accepted on the recommendation of

Prof. Dr. Sereina Riniker

Dr. Marc-Olivier Ebert

Prof. Dr. Helma Wennemers

2022

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Marc-Olivier Ebert and Prof. Sereina Riniker for their constant support during the last years. I am grateful that you gave me the opportunity to work on these challenging and exciting projects. Sereina, you are doing an amazing job as a group leader. Thanks to you, we have such a great, collegial and motivating group spirit. Your suggestions and expertise were always extremely helpful. Oli, thank you for all your advice, your time, and your passion. Your immense knowledge about science in general is impressive. You really opened my eyes to the versatility and beauty of NMR spectroscopy.

A special thanks goes to Prof. Helma Wennemers for agreeing to be co-examiner of this thesis and for hosting some of my students.

I would also like to thank the current and former members of the Riniker and Hünenberger group and also Prof. Phillippe Hünenberger himself: Albert Hofstetter, Alžbeta Kubincová, Anna Albertini, Annick Renevey, Benjamin Ries, Candide Champion, Carmen Esposito, David Hahn, Dominik Sidler, Emília Pécora de Barros, Felix Pultar, Gerhard König, Gregory Landrum, Gregor Weiss, Jagna Witek, Jessica Braun, Lennard Bösel, Marc Lehner, Marina Pereira Oliveira, Moritz Thürlemann, Patrick Bleiziffer, Paul Katzberger, Sadra Kashef Ol Ghetta, Salomé Rieder, Shuhze Wang, and Stephanie Linker. The working atmosphere and group spirit are exceptional. Thank you for all the help, advice and chats we had in the last years during and after work. It was great to meet you all.

Further, I would like to thank Carla Rigling and Prof. Bernhard Jaun for helpful discussions as well as Claudia Hilty for the help with all the administrative tasks.

I had the pleasure to supervise excellent students and would like to thank Chantal Balmer, Dénes Tary, Hristo Bonchev, Monique Kuonen, and Stefan Feusi for their fantastic work, which contributed significantly to this thesis.

A big thank you goes also to all the excellent facilities at ETH. I would like to thank Rainer Frankenstein, René Arnold, and Stephan Burkhardt from the NMR Service of the LOC for their general help and the interesting discussions during the coffee breaks. I would also like to thank Michael Solar and Nils Trapp from the Small Molecule Crystallography Center for their advice and help to obtain great crystal structures. Another big thank you goes to Christian Marro and his apprentice from the LOC workshop for the manufacturing of the compression devices. I would also like to thank Andreas Schneider for his help during many hours trying to fix the VCD

spectrometer. Further, I would like to thank Daniel Zindel and his team for the efforts to resynthesize some of the cyclic peptides studied in this thesis.

I would also like to acknowledge my collaborators in the depsipeptides project Govindan Subramanian, Sanjay Menon, Chad Townsend, and Prof. Scott Lokey (Chapter 2), Marcel Grogg and Prof. Donald Hilvert for the synthesis of the peptides and helpful discussions (Chapter 3) as well as Prof. Erick Carreira and his lab for providing the polychlorinated-diols studied in Chapter 7.

Last but not least, I would like to thank my parents, my siblings, and my friends for their continuous support.

Summary

In this thesis, the use of spectroscopic data of small and medium-sized molecules in combination with computational approaches is investigated as a means to gain insight into the solution ensemble of the molecules and their various properties.

In Chapter 1, the importance of the correct description of the conformational ensemble is discussed. The relevant observables from nuclear magnetic resonance (NMR) spectroscopy in isotropic and anisotropic environments are presented together with the structural information that can be derived from them to describe the conformational ensemble. The basic concepts of molecular dynamics (MD) simulations and density functional theory (DFT) calculations are covered, and the importance of the choice of starting conformation in these approaches is considered.

In Chapter 2, the conformational behavior and ionophoric properties of cyclooctadepsipeptides with different anthelmintic activity are investigated with NMR spectroscopy and extensive MD simulations.

In Chapter 3, the transferability of the stabilizing effect of intramolecular hydrogen bonds between side-chain N-methylated asparagine residues, as seen for the $\beta^{6,3}$ -helix of the natural product polytheonamide B, is studied with two different model systems. Among other techniques, the characterization was carried out with distances derived from the nuclear Overhauser effect (NOE), ^{23}Na NMR spectroscopy, and MD simulations.

In Chapter 4, a new set of precise residual dipolar coupling (RDC) data for cyclosporin A is presented, recorded in a cross-linked poly(methyl methacrylate) (PMMA) gel swollen in chloroform is presented. The impact of the RDCs when used as restraints in an MD simulation of this flexible compound is investigated.

In Chapter 5, ^1H and ^{13}C chemical shifts of 35 small and rigid organic molecules are measured under standardized conditions in chloroform-d and in tetrachloromethane. The effect of directed solute-solvent interactions is evaluated, and the chemical shifts are compared to shielding constants calculated with DFT in vacuum and with an implicit solvent model.

In Chapter 6, a new approach for the interpretation of NOE data called NOE volumes affected by spin diffusion (NOVAS) is developed and used to fit experimental NOESY spectra beyond the linear

build-up regime for different organic molecules. The usefulness of NOVAS for stereospecific assignment as well as for the correct identification of relative configuration is investigated.

In Chapter 7, we aim to assign the relative configuration of eight flexible diastereomers of a trichlorinated-hexa-1-3-diol by comparison between experimental and DFT calculated ^1H and ^{13}C chemical shifts, NOESY spectra as well as infrared (IR) spectra. The chemical shieldings obtained from DFT calculations are compared to the measured chemical shifts by applying the findings from Chapter 5, while the NOESY spectra are calculated and fitted to the experimental spectra using the NOVAS approach presented in Chapter 6. For the IR spectra, we apply an improved version of the IR sequence alignment (IRSA) algorithm, developed in our laboratory.

In Chapter 8, a short conclusion and an outlook are provided regarding the topics discussed in this thesis.

Zusammenfassung

In dieser Dissertation wird die Verwendung spektroskopischer Daten kleiner und mittelgroßer Moleküle in Kombination mit rechnergestützten Methoden untersucht als ein Mittel, um Einblick in das konformationelle Ensemble der Moleküle und ihre verschiedenen Eigenschaften zu gewinnen.

In Kapitel 1 wird die Bedeutung der korrekten Beschreibung des konformationellen Ensembles diskutiert. Die wichtigsten Observablen aus der Kernspinresonanzspektroskopie (NMR) in isotropen und anisotropen Umgebungen werden vorgestellt, zusammen mit den darin enthaltenen strukturellen Informationen, die zur Beschreibung des konformationellen Ensembles verwendet werden können. Die grundlegenden Konzepte von Molekulardynamik-Simulationen (MD) und Dichtefunktionaltheorie-Berechnungen (DFT) werden behandelt und die Bedeutung der Wahl der Ausgangskonformation für diese Methoden wird betrachtet.

In Kapitel 2 werden NMR-Spektroskopie und umfangreiche MD-Simulationen verwendet für die Untersuchung des konformationellen Verhaltens und der ionophoren Eigenschaften von zyklischen Oktadepsipeptiden mit unterschiedlicher anthelmintischer Aktivität.

In Kapitel 3 wird die Übertragbarkeit des stabilisierenden Effekts von intramolekularen Wasserstoffbrückenbindungen zwischen N-methylierten Asparagin-Seitenketten, wie für die $\beta^{6,3}$ -Helix des Naturstoffs Polytheonamid B beobachtet, mit zwei verschiedenen Modellsystemen untersucht. Unter anderem wurde die Charakterisierung mit NOE-Distanzen, ^{23}Na -NMR-Spektroskopie und MD-Simulationen durchgeführt.

In Kapitel 4 wird ein neuer Satz präziser RDC Daten (Dipolare Restkopplungen) für Cyclosporin A präsentiert, die in einem vernetzten, in Chloroform gequollenen Poly(methylmethacrylat)-Gel (PMMA) aufgenommen wurden. Der Einfluss der Verwendung dieser RDCs als Restraints in einer MD-Simulation von Cyclosporin A wird untersucht.

In Kapitel 5 werden die ^1H und ^{13}C chemischen Verschiebungen von 35 kleinen, starren organischen Molekülen in Chloroform-d und in Tetrachlormethan unter standardisierten Bedingungen gemessen. Der Einfluss gerichteter Wechselwirkungen zwischen gelöster Substanz und Lösungsmittel wird untersucht, und die chemischen Verschiebungen werden mit Abschirmungskonstanten verglichen, die mit DFT im Vakuum und mit einem impliziten Lösungsmittelmodell berechnet wurden.

In Kapitel 6 wird eine neue Methode für die Interpretation von NOE-Daten namens NOVAS (NOE volumes affected by spin diffusion) entwickelt und verwendet, um experimentelle NOESY-Spektren jenseits des linearen Aufbauregimes für verschiedene organische Moleküle zu fitten. Die Nützlichkeit von NOVAS für die stereospezifische Zuordnung sowie die Identifizierung der relativen Konfiguration wird untersucht.

In Kapitel 7 versuchen wir, die relative Konfiguration von acht flexiblen Diastereomeren eines trichlorierten Hexa-1-3-diols zuzuordnen, indem wir experimentelle und DFT-berechnete ^1H und ^{13}C chemische Verschiebungen, NOESY-Spektren und Infrarotspektren (IR) vergleichen. Die aus DFT-Berechnungen gewonnenen Abschirmungskonstanten werden mit den gemessenen chemischen Verschiebungen unter Anwendung der Erkenntnisse aus Kapitel 5 verglichen, während die NOESY-Spektren mit dem in Kapitel 6 vorgestellten NOVAS-Ansatz berechnet und an die experimentellen Spektren gefittet werden. Für die IR-Spektren wenden wir eine verbesserte Version des in unserem Labor entwickelten IR-Spektrenangleichungs-Algorithmus (IRSA) an.

Kapitel 8 enthält eine kurze Schlussfolgerung und einen Ausblick in Bezug auf die in dieser Arbeit behandelten Themen.

Publications and Presentations

Article in peer-reviewed journals:

T. Stadelmann, G. Subramanian, S. Menon, C. Townsend, R. S. Lokey, M.-O. Ebert and S. Riniker. Connecting the Conformational Behavior of Cyclic Octadepsipetides With Their Ionophoric Property and Membrane Permeability. *Org. Biomol. Chem.* 2020, **18**, 7110-7126.

Related Publications:

C. Comeau, B. Ries, T. Stadelmann, J. Tremblay, S. Poulet, U. Fröhlich, J. Côté, P.-L. Boudreault, R. M. Derbali, P. Sarret, M. Grandbois, G. Leclair, S. Riniker and E. Marsault. Modulation of the Passive Permeability of Semi-Peptidic Macrocycles. N- and C-Methylations Fine-tune Conformation and Properties. *J. Med. Chem.* 2021, **64**, 5365-5383.

S. M. Linker, S. Wang, B. Ries, T. Stadelmann and S. Riniker. Passing the Barrier – How Computer Simulations Can Help to Understand and Improve the Passive Membrane Permeability of Cyclic Peptides. *CHIMIA.* 2021, **75**, 518-521.

Oral Presentations:

4th ETH – University of Tokyo Symposium: Innovations in Chemical Biology, Zürich, Switzerland, February 2020

“Using Molecular Dynamics and Nuclear Magnetic Resonance to Investigate Cyclic Peptides in Solution”

29th Intl. BIOMOS Symposium on Biomolecular Simulation, Ausserberg, Switzerland, September 2019

“Conformational changes of cyclic depsipeptide PF1022A depending on the cation Concentration”

Poster Presentations:

Fall Meeting of the Swiss Chemical Society, Online 2021

“Accessing the Conformational Space of Cyclosporin A in Chloroform Using RDCs as Tensorial Constraints”

SMASH NMR Conference, Porto, Portugal, September 2019

“Investigation of the Structure of Cyclic Depsipeptides PF1022A and Emodepside and Their Affinity to Cations”

Contents

Acknowledgements	iii
Summary	v
Zusammenfassung	vii
Publications and Presentations	ix
Contents	xi
1 Introduction	1
1.1 Conformational Ensemble in Solution.....	1
1.2 Experimental Description of the Solution Ensemble Using NMR.....	2
1.2.1 Information Obtained From Chemical Shifts.....	2
1.2.2 Information Obtained From J-Couplings.....	3
1.2.3 Information Obtained From Nuclear Overhauser Effect.....	3
1.2.4 Information Obtained From Anisotropic NMR.....	4
1.3 Molecular Dynamics Simulations.....	5
1.3.1 Classical Force Fields.....	5
1.3.2 Markov State Models.....	6
1.3.3 Incorporation of Experimental Data.....	6
1.4 Density Functional Theory.....	7
1.4.1 Structure Optimization and Frequency Calculation.....	7
1.4.2 Calculation of NMR Properties.....	8
1.5 Starting Conformations for Computational Approaches.....	9
2 Connecting the Conformational Behavior of Cyclic Octadepsipeptides with Their Ionophoric Property and Membrane Permeability	11
2.1 Introduction.....	12
2.2 Results.....	17
2.2.1 Characterization of the Conformational Behavior.....	17
2.2.2 Effect of the Presence of Monovalent Cations.....	21
2.2.3 Effect of the Presence of Monovalent Cations.....	30
2.3 Conclusion.....	31
2.4 Method Section.....	33
2.5 Appendix.....	40
3 Transferring the Stabilizing Effect of Side-Chain N-Methylations Observed in Polythienoamide B	55

3.1	Introduction	56
3.2	Results.....	60
3.2.1	Gramicidin A and its Asm/Asn Variants	60
3.2.2	Cyclic Asm/Asn Variants	62
3.3	Conclusion.....	68
3.4	Method Section.....	69
3.5	Appendix	72
4	Restrained MD-Simulations Using a Complete Set of One-Bond CH and NH Residual Dipolar Couplings for Cyclosporin A in CDCl₃	75
4.1	Introduction	76
4.2	Theory	78
4.2.1	Treatment of RDCs in Rigid Compounds.....	78
4.2.2	Treatment of RDCs in Flexible Compounds	79
4.2.3	Experimental Determination of RDCs.....	80
4.3	Results.....	82
4.4	Conclusion.....	89
4.5	Method Section.....	90
5	Comparison of Experimental and DFT Calculated Chemical Shifts Using a New Standardized Dataset Recorded in Chloroform and Tetrachloromethane.....	95
5.1	Introduction	96
5.2	Results.....	99
5.2.1	Effect of Experimental Protocol and Water Content in Chloroform	99
5.2.2	How Well Do DFT Calculations in Vacuum Reproduce Experimental Chemical Shifts in Solution?	102
5.2.3	Can the Accuracy of the Calculations Be Improved with an Implicit Solvent Model?	103
5.3	Conclusion.....	106
5.4	Method Section.....	107
5.5	Appendix	109
6	NOVAS: A Simple Protocol for Using NOESY Volumes Instead of NOE-Derived Distances for the Determination of Relative Configuration.....	133
6.1	Introduction	134
6.2	Theory	136
6.3	Results.....	139
6.3.1	Testing the NOVAS Procedure on Strychnine (1).....	139
6.3.2	Dealing with Fast Internal Methyl Group Rotation: The Example of trans-Crotonaldehyde (2).....	141

6.3.3	Diastereotopic Assignment of Methylene Groups in Morphine (3).....	145
6.3.4	Identifying the Correct Stereospecific Assignment for Androstenedione (4) out of 256 Possibilities	148
6.3.5	Differentiation between ephedrine (5) and pseudoephedrine (6)	151
6.4	Conclusion	153
6.5	Method Section	154
6.7	Appendix.....	157
7	Assignment of Relative Configuration in Linear Chlorinated Diols by Comparison of Experimental and Theoretical Spectroscopic Data.....	167
7.1	Introduction.....	169
7.2	Results	172
7.2.1	Generation of Conformational Ensembles.....	172
7.2.2	Differentiation of Diastereomers Based on Chemical Shifts.....	172
7.2.3	Differentiation of Diastereomers Based on NOVAS Approach	176
7.2.4	Combination of Chemical Shifts with the NOVAS Approach.....	178
7.2.5	Differentiation of Diastereomers Based on IR Spectra	179
7.3	Conclusion	182
7.4	Method Section	184
7.5	Appendix.....	186
8	Conclusion and Outlook	199
	References.....	203
	Curriculum Vitae	227

1 Introduction

1.1 Conformational Ensemble in Solution

Most molecules do not adopt a unique three-dimensional (3D) structure in solution at room temperature but rather exist as an ensemble of different conformations. It is essential for the understanding of a flexible system to get insight into its conformational behavior in solution and to know how the ensemble changes in different environments (i.e., polar versus apolar solvents, inside versus outside of a cell membrane).¹⁻⁵ Not only the populations of the different conformations are relevant but also the rate of interconversion between the members of the conformational ensemble.⁶ The conformational ensemble and its dynamics in a given environment are key factors to understand biological activity, molecular binding, cell permeability, and many other properties of interest.^{1,3,7,8} With computational methods, it is possible to predict such conformational ensembles for flexible systems and determine their physicochemical properties. These can be used to test hypotheses of the relative configuration, constitution, or conformational preferences by comparison with experimental data.

For small molecules, it is possible to systematically sample all accessible conformations or scan the conformational space with Monte Carlo based methods. The relative energies of the generated conformations can then be determined for instance with density functional theory (DFT) (Section 1.4), semi-empirical methods, or classical force fields (Section 1.3). Examples of such generated ensembles are presented in Chapter 6 and Chapter 7. For larger molecules, these techniques are less well suited due to the vast number of degrees of freedom and the resulting size of the conformational space. In such situations, Molecular dynamics (MD) simulations are a valuable tool to explore the conformational landscape (see Chapter 2 and Chapter 3). Nuclear magnetic resonance (NMR) spectroscopy is excellently suited for the validation of such computationally generated ensembles due to its atomic resolution.

Another approach is to directly generate a conformational ensemble that is in agreement with the experimental data. Techniques like restrained MD simulations allow the interpretation of experimental data of flexible molecules. In Chapter 4, we investigate whether residual dipolar couplings (RDCs) can be applied for restraining of an MD simulation to study the conformational ensemble in solution.

1.2 Experimental Description of the Solution Ensemble Using

NMR

NMR spectroscopy is the method of choice to determine the constitution of organic compounds.⁹ In contrast to many other spectroscopic techniques like infrared (IR), ultraviolet (UV), or circular dichroism (CD) spectroscopy, the information that can be gained from NMR spectra has atomic resolution, i.e., the signals can be assigned to the individual atoms in a molecule. Besides the connectivity information, it is also possible to get insight into the conformational and dynamic properties of the ensemble in solution.¹⁰ Due to numerous possible observables, different time scales from picoseconds to seconds can be investigated. In the following, the most important NMR observables are presented.

1.2.1 Information Obtained From Chemical Shifts

Chemical shifts are the basis for structure elucidation.^{11,12} The chemical shifts represent the type of information that is gained most easily from an NMR spectrum. In short, the chemical shift is the difference in resonance frequency of a nucleus of interest and a standard. For ^1H and ^{13}C chemical shifts this is tetramethyl silane (TMS). The chemical shift arises from local differences in the static external magnetic field B_0 between the location of the different nuclei in the sample. Depending on the electronic environment, this shielding of B_0 can be more or less pronounced. In general, electron donating groups increase the shielding (lower chemical shift), whereas electronegative groups decrease it (higher chemical shift). Therefore, the ^1H and ^{13}C chemical shifts of a methyl group connected to an sp^3 quaternary carbon are significantly lower ($\sim 0.5 - 2$ ppm, $\sim 0 - 30$ ppm)¹³ compared to the chemical shifts of an aldehyde group next to an sp^3 quaternary carbon ($\sim 9.5 - 10$ ppm, $\sim 197 - 205$ ppm).¹³ The agreement between experimentally determined chemical shifts and calculated shieldings with DFT is investigated in Chapter 5.

Chemical shifts are mainly affected by first and second neighbors surrounding the nucleus of interest. Therefore, they not only reflect local constitution but also contain information about the configuration of nearby bonds and functional groups. Thus, chemical shifts can also be used to differentiate between diastereomers or to guide the calculation of conformational ensembles.¹⁴⁻¹⁶ If they exchange slowly on the NMR time scale, even individual sets of resonances can be recorded for the respective members of the conformational ensemble. In Chapter 7, we investigate how well the small differences in ^1H and ^{13}C chemical shifts between the diastereomers of an open-chain chlorinated polyol can be predicted computationally in order to assign their relative configuration.

1.2.2 Information Obtained From J-Couplings

J-couplings, also known as scalar couplings, arise from an indirect interaction of the nuclear magnetic moments mediated by the electrons. ^3J -couplings and to a lesser extent also ^2J -couplings contain valuable information about the dihedral angles associated with the coupling nuclei. However, to make use of this information, the relation between the size of the coupling constant and a given dihedral angle needs to be calibrated, usually with a set of reference compounds. For intensely studied classes of molecules, like proteins, nucleic acids and carbohydrates, numerous parameter sets (Karplus parameters) have been developed that can be used to translate vicinal couplings constants into the dihedral angles between the coupling nuclei.¹⁷⁻¹⁹ The so-called Karplus curve shows a periodic dependence of the vicinal coupling constant on the intervening dihedral angle. Thus, given a single vicinal coupling constant, an unambiguous assignment of the respective dihedral angle is only possible when the coupling is close to a local maximum value (0° or 180°). Generally, up to four dihedral angles will yield the same coupling constant. For flexible molecules, several distinctly different distributions of local conformations can potentially yield the same scalar couplings. Despite this limitation, the combination of different J-couplings relating to the same dihedral angle can be used to differentiate between stereoisomers and for the assignment of diastereotopic protons, even in flexible molecules.^{20,21}

1.2.3 Information Obtained From Nuclear Overhauser Effect

The nuclear Overhauser effect (NOE), also known as nuclear Overhauser enhancement, is one of the most powerful observables in NMR spectroscopy.²²⁻²⁴ In contrast to J-couplings, the NOE is not limited to nuclei connected through local chemical bonds, but is a through space effect. It is based on a distance-dependent dipole-dipole interaction between the magnetic moments of two spins. As such, it contains valuable spatial information that can be used to elucidate the 3D structure of a molecule.²⁵ While this is relatively straightforward for mostly rigid molecules, more flexible compounds require the combination with advanced computational approaches to obtain information about local mobility and the conformational ensemble.²⁶⁻²⁸ The theory behind the NOE together with an application of NOE volumes affected by spin diffusion (NOVAS) to differentiate between diastereomers of organic compounds is described in Chapter 6 and Chapter 7. Exchange spectroscopy (EXSY), a technique closely related to the experimental determination of the NOE, can be employed to determine exchange rates between slowly interchanging conformations. An example of how EXSY has been used to understand the slow kinetics of cyclic octadepsipeptides is provided in Chapter 2.

1.2.4 Information Obtained From Anisotropic NMR

In addition to the NMR observables in isotropic solution briefly discussed above, there are also anisotropic interactions, which normally average out to zero in solution. These anisotropic interactions play an important role in solid-state NMR and contain additional structural information. By inducing a partial alignment in one direction, a small portion of these interactions can be observed in a system nearly behaving like an isotropic solution. This weak alignment can be achieved using stretched or compressed gels,^{29–32} liquid crystals based on helix forming peptides that self-align with the external magnetic field,^{33–35} or by tight binding of the studied molecule to a paramagnetic ion.^{36–38} The interaction most often studied for small to medium-sized molecules is the residual dipolar coupling (RDC).^{39–41} It contains information about the distance between the coupling nuclei and the angle of the internuclear vector relative to B_0 . RDCs can be obtained by measuring the isotropic J-coupling constant and the total coupling constant T (sum of J-coupling and RDC) in anisotropic solution (partial alignment). The difference between the two corresponds to the RDC. The basic theory underlying the RDC is presented in Chapter 4. Especially for flexible systems, the interpretation of the information contained in RDCs is a challenging task for which no general solution has been found yet.

A further parameter that can be determined in partially aligned samples is the residual chemical shift anisotropy (RCSA), which contains information about the orientation of the chemical shielding tensors of the magnetic nuclei.⁴² Here, the fact that chemical shifts are extremely sensitive to the environment becomes a challenge, since the isotropic shift itself will slightly change between the isotropic reference and the partially aligned sample.⁴² Due to the difficulties to extract RCSAs with high accuracy, they are less often used compared to RDCs.⁴³ For spins $>1/2$, also residual quadrupolar couplings (RQCs) can be observed in partially aligned media.^{44,45} In applied NMR spectroscopy, this effect is still of minor importance due to the low natural abundance of ^2H . However, this could change in the future with the advent of more sensitive NMR instrumentation.

1.3 Molecular Dynamics Simulations

Classical MD simulations can give insight into the conformational ensemble adopted by a molecule in solution.^{46,47} In contrast to NMR experiments, where the obtained information is averaged over many molecules and conformations, it is possible to directly observe an individual molecule and its conformational changes over time in simulations. In classical MD, atoms are described as particles that obey Newton's laws of classical mechanics. Applications of MD simulations involve folding and unfolding of peptides and proteins, interactions in host-guest systems, the role of solvent and ion concentration, and many other properties and processes.^{48–54}

1.3.1 Classical Force Fields

In order to perform an MD simulation, information about the behavior of atoms, bonds, angles, torsions and non-bonded electrostatic and van der Waals interactions are needed. They are defined in a so-called force field (FF) (Figure 1.1). The goal is to define the parameters as general as possible without losing accuracy, so that the same parameters can be used and combined to describe similar molecules. This reduces the number of parameters needed for an MD simulation.

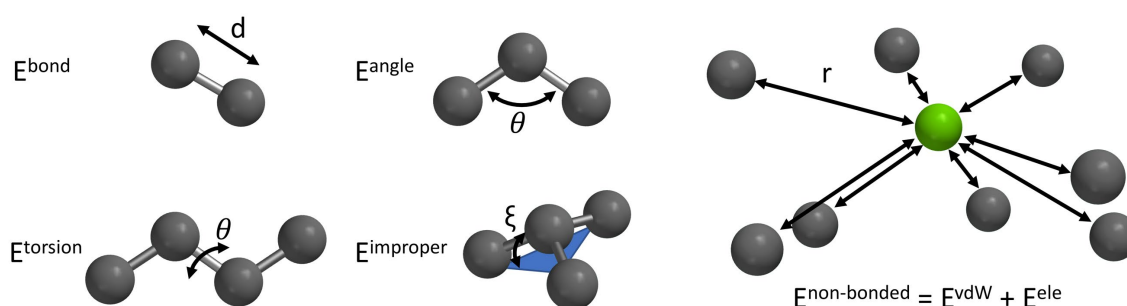


Figure 1.1: Schematic representation of the different bonded and non-bonded terms defined in a classical force field. E_{bond} describes bond stretching, E_{angle} bond angle bending, E_{torsion} the dihedral angle torsions, E_{improper} the improper dihedral angle bending, E_{vdW} the van der Waals interactions and E_{ele} the electrostatic interactions. While E_{bond} , E_{angle} , E_{torsion} and E_{improper} are bonded terms, E_{vdW} and E_{ele} are the non-bonded potential-energy contributions. Adapted from Ref. 55 with permission of the American Chemical Society.

Different FFs have been parameterized for various classes of (bio)molecules. Many design choices can be made during the parametrization procedure. In the early days of MD simulations, FFs were constructed by merging aliphatic carbons and their directly bonded hydrogen atoms into a single larger bead (united-atom approach). This was mainly done to reduce the number of particles and therefore to speed-up the calculations. Nowadays, the FFs of the GROMOS family^{56,57} still use united atoms, whereas other FF families model all hydrogens explicitly (e.g., Amber,^{58,59} CHARMM,^{60,61} and OPLS^{62,63}). Recent benchmark studies suggest that there is no clear advantage of all-atom FFs over united-atom FFs, and which FF performs best seems to be largely dependent on the properties that are to be reproduced.^{64,65} Indeed, the choice of the properties for which the FF is optimized (i.e., purely experimental properties like heat of vaporization, torsional

preferences etc., or only quantum chemical properties like van der Waals parameters or partial charges, or a mixed approach) is another important aspect during FF parametrization, together with the functional forms of the different terms. More information about the differences between the FF families and their design choices can be found in a recent review by Riniker.⁵⁵

1.3.2 Markov State Models

Markov state models (MSMs)⁶⁶⁻⁶⁹ are a powerful tool to analyze the conformational dynamics in MD simulations. This approach allows to combine several trajectories to obtain a Boltzmann distribution that describes the kinetics of the system. For this, a dimensionality reduction step is employed and the conformations are clustered into discretized microstates.⁷⁰ In a second step, these microstates are kinetically clustered into so-called metastable conformational states. Using MSMs, key conformations together with their transition probabilities can be extracted, providing information about mean first passage times. With this approach, it is possible to describe conformational changes that are slower than the length of the individual MD trajectories. In Chapter 2 and Chapter 3, MSMs were successfully applied to describe the conformational behavior of cyclic peptides.

1.3.3 Incorporation of Experimental Data

Many MD simulation packages allow the incorporation of experimental data via distances, angles and dihedrals to restrain the conformational sampling.^{26,71} Most often, data obtained from NMR experiments is employed, i.e., NOE-derived distances to restrain inter-proton distances, $^3J_{H-H}$ couplings to restrain dihedral angles, and information about stable hydrogen bonds to restrain these atom pairs in the simulation. For proteins, also chemical shifts have been used successfully as restraints for the backbone conformation.¹⁴ Recently, it was shown that RDCs can also be incorporated for conformational restraining.⁷² The use of RDCs as restraints in MD simulations is explored in Chapter 4.

1.4 Density Functional Theory

In contrast to classical MD simulations, where atoms are treated as single particles without considering the electrons explicitly, DFT models the electron density to determine the quantum-mechanical (QM) ground state of the system.⁷³ From a QM perspective, all properties of the system can be deduced from its wavefunction. Unfortunately, the Schrödinger equation cannot be solved for many-body systems and thus, approximations need to be employed. DFT uses the electron density instead of the many-body wavefunction. Often, DFT is the method of choice to study electronic and structural properties of molecules due to an optimal tradeoff between accuracy and computational efficiency.⁷⁴ The most often applied Kohn-Sham implementation of DFT depends on a functional and a basis set.⁷⁵ Since the functional form cannot be deduced from first principles, a large number of functionals and basis sets have been proposed, each with its strengths and weaknesses.^{76,77} Since the calculations are still rather expensive compared to classical MD simulations, an explicit solvation shell is usually not included in standard DFT calculations. Instead, most DFT calculations are performed in vacuum or using an implicit solvent model.

1.4.1 Structure Optimization and Frequency Calculation

Geometry optimization is a key step in DFT and computational chemistry in general.⁷⁸ Since the motion of the nuclei is much slower compared to the motion of the electrons, the energy of the system can be obtained by treating the nuclei as fixed (Born-Oppenheimer approximation).⁷⁹ The energy of a molecule can then be calculated for different relative positions of the nuclei, yielding a potential-energy surface.⁷⁸ The derivatives of the energy with respect to the coordinates of the nuclei can be used as gradients to find a configuration where the net force on each atom is reasonably close to zero (first derivative is zero, i.e., a local energetic minimum). To verify that this stationary point is indeed a minimum and not a transition state or saddle point, the second derivative needs to be computed as well. The normal modes of the Hessian of the energy with respect to the position of the nuclei give the vibrational normal modes of the molecule.⁷⁸ When all normal modes are positive, the structure corresponds to a minimum on the Born-Oppenheimer surface.

1.4.2 Calculation of NMR Properties

Numerous molecular properties that depend on the electronic structure of the system can be computed with DFT.⁸⁰ For comparison with solution state NMR, the most important properties are chemical shifts and J-couplings. To obtain the chemical shifts, it is necessary to calculate the chemical shielding tensors. In order to compute them, the second derivative of the total energy with respect to the magnetic field and the nuclear magnetic moment is needed.⁸¹ This is most often done using the gauge including/invariant atomic orbitals (GIAO) approach to avoid a dependence of the results on the choice of the origin.⁸² As an example, the DFT calculated chemical shifts or J-couplings can then be used in combination with experimental data to differentiate between stereoisomers.⁸³ The comparison between calculated shieldings and experimental chemical shifts is demonstrated in Chapter 5 and Chapter 7. To compute J-couplings, the second derivative of the energy of the molecule with respect to the magnetic moments of both involved nuclei is needed.⁸³ An example of a ¹H NMR spectrum with calculated J_{H-H} couplings is shown in Chapter 6.

1.5 Starting Conformations for Computational Approaches

Independent whether a classical MD simulation or a DFT geometry optimization is performed, the starting conformation is of general importance.^{84,85} For rigid molecules or biomolecules with stable tertiary structures like proteins, starting from a crystal structure is a valid and reasonable approach. However, when studying a newly designed molecule or a compound without a known crystal structure, this is not possible. In this case, one needs to generate a reasonable starting conformation *in silico*. In principle, an MD simulation itself can be regarded as a conformer generator. The advantage of MD simulations is that most of the generated conformers are low-energy structures (assuming a good underlying FF), but the conformational search is relatively inefficient and the simulation can get trapped in local minima. Therefore, numerous enhanced sampling methods have been developed in the past decades.^{86,87}

For very small molecules, it is feasible to sample all possible conformations systematically. Since the number of conformations grows rapidly with the number of rotatable bonds, conformer generation needs to be done in a smarter way for larger molecules. Alternative methods include for example Monte Carlo sampling techniques (applying random rotations around bonds and accepting trial moves based on a predefined criterion), or *in silico* conformer generators such as distance geometry, where a distance bounds matrix with minimum and maximum atom pair distances is created.⁸⁸ All atomic distances for each possible conformer have to lie between these bounds and are stochastically sampled during the embedding step.⁸⁸ This approach can be improved by adding information about general molecular properties (e.g., keeping aromatic rings and amides flat) and using torsional profiles fitted to a large collection of crystal structures.⁸⁹ A major advantage of distance geometry is its computational efficiency.^{89,90} However, one needs to keep in mind the applicability of the fitted torsional profiles for the compound of interest. If this is not the case, even highly populated members of the Boltzmann ensemble may be overlooked. Such a case is described in Chapter 7.

2 Connecting the Conformational Behavior of Cyclic Octadepsipeptides with Their Ionophoric Property and Membrane Permeability*

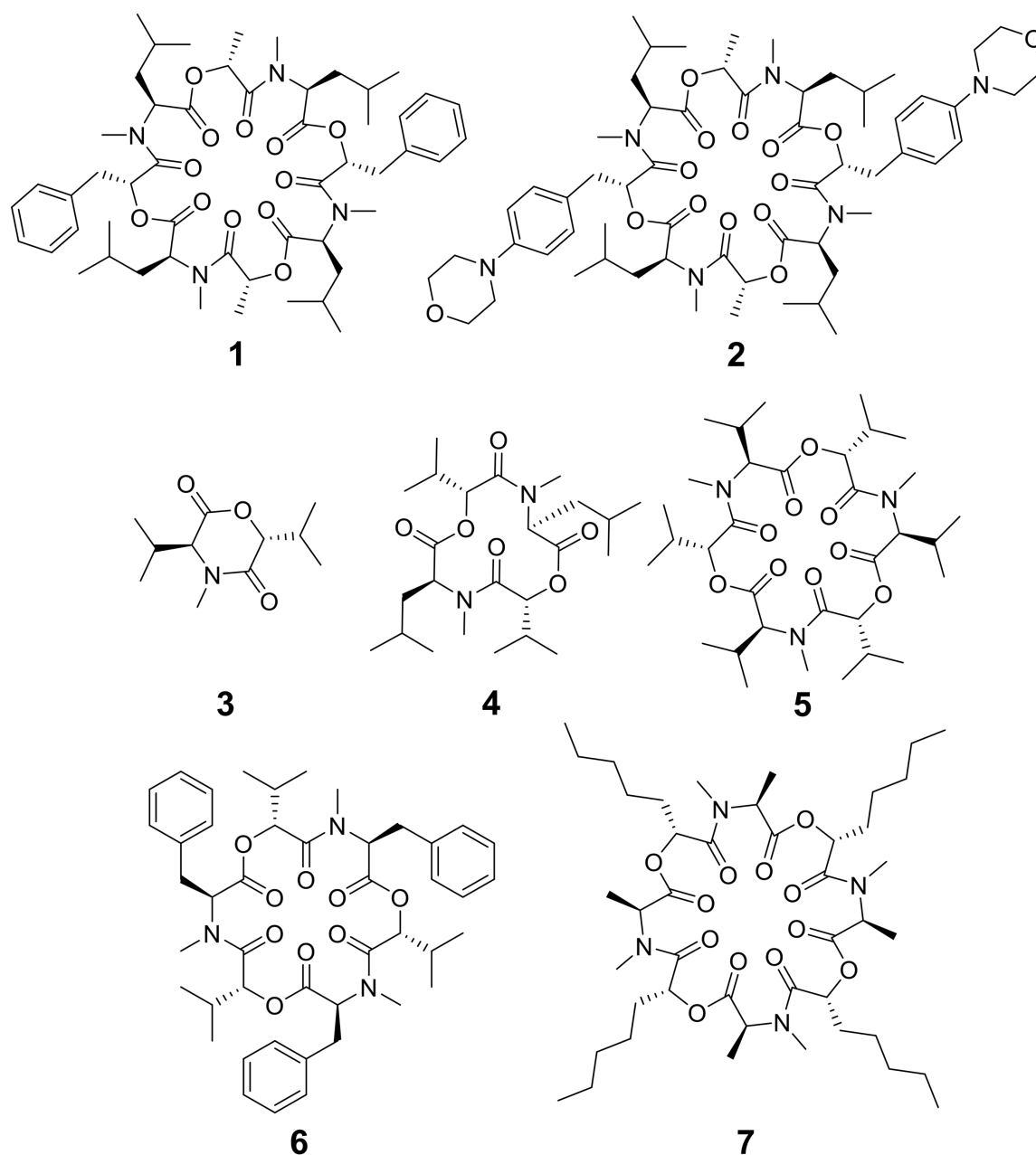
Cyclic octadepsipeptides such as PF1022A and its synthetic derivative emodepside exhibit anthelmintic activity with the latter sold as a commercial drug treatment against gastrointestinal nematodes for animal health use. The structure – permeability relationship of these cyclic depsipeptides that could ultimately provide insights into the compound bioavailability is not yet well understood. The fully N-methylated amide backbone and apolar side-chain residues do not allow for the formation of intramolecular hydrogen bonds, normally observed in the membrane-permeable conformations of cyclic peptides. Hence, any understanding gained on these depsipeptides would serve as a prototype for future design strategies. In previous nuclear magnetic resonance (NMR) studies, two macrocyclic core conformers of emodepside were detected, one with all backbone amides in *trans*-configuration (i.e., symmetric conformer) and the other with one amide in *cis*-configuration (i.e., asymmetric conformer). In addition, these depsipeptides were also reported to be ionophores with a preference of potassium over sodium. In this study, we relate the conformational behavior of PF1022A, emodepside, and closely related analogs with their ionophoric characteristic probed using NMR and molecular dynamics (MD) simulations and finally evaluated their passive membrane permeability using PAMPA. We find that the equilibrium between the two core conformers shifts more towards the symmetric conformer upon addition of monovalent cations with selectivity for potassium over sodium. Both the NMR experiments and the theoretical Markov state models based on extensive MD simulations indicate a more rigid backbone for the asymmetric conformation, whereas the symmetric conformation shows greater flexibility. The experimental results further advocate for the symmetric conformation binding the cation. The PAMPA results suggest that the investigated depsipeptides are retained in the membrane, which may be advantageous for the likely target, a membrane-bound potassium channel.

* This chapter is reproduced in part from T. Stadelmann, G. Subramanian, S. Menon, C. E. Townsend, R. S. Lokey, M.-O. Ebert and S. Riniker, *Org. Biomol. Chem.*, **18**, 7110-7126 (2020), with permission from the Royal Society of Chemistry.

2.1 Introduction

Depsipeptides are atypical peptides where one or more backbone peptide amides are replaced by ester groups. Many cyclic depsipeptides were found as secondary metabolites in nature with various applications like antibiotics, antifungal and antiviral drugs, enzyme inhibitors, ionophores, anthelmintic therapeutics etc.⁹¹⁻⁹⁷ Due to the cyclization, the flexibility of the backbone is restricted but still flexible enough to interact with potential targets.⁹⁸ This makes cyclic depsipeptides interesting lead structures for drug development.

The cyclic octadepsipeptide PF1022A (**1**) (Scheme 2.1) is a natural product, consisting of two repetitions of D-lactic acid, N-methyl L-leucine, D-phenyllactic acid and N-methyl L-leucine and has therefore a C₂ symmetry axis. PF1022A demonstrates pharmacological activity against nematodes.⁹⁹ Its synthetic derivative, emodepside (**2**) (Scheme 2.1) containing additional morpholine rings at the para position of the phenyllactic acid aromatic rings, exhibits increased anthelmintic activity¹⁰⁰ and is a commercial drug effective against a number of gastrointestinal nematodes in cats. PF1022A and emodepside belong to a subfamily of cyclic depsipeptides that have all the backbone amides methylated and possess only apolar side-chains. This means that no hydrogen bond donors are present and thus, the formation of intramolecular hydrogen bonds is not possible. Yet, the ability to adopt a conformation, which maximizes the number of intramolecular hydrogen bonds is thought to be essential for passive membrane permeation of cyclic peptides.^{3,51,101-105} Nevertheless, some members of this subfamily of cyclic depsipeptides were found to be permeable or can be easily incorporated into a lipid membrane.¹⁰⁶ Thus, to exploit their potential as therapeutics, it is important to establish a better understanding of the relationship between structure (conformational behavior) and permeability.



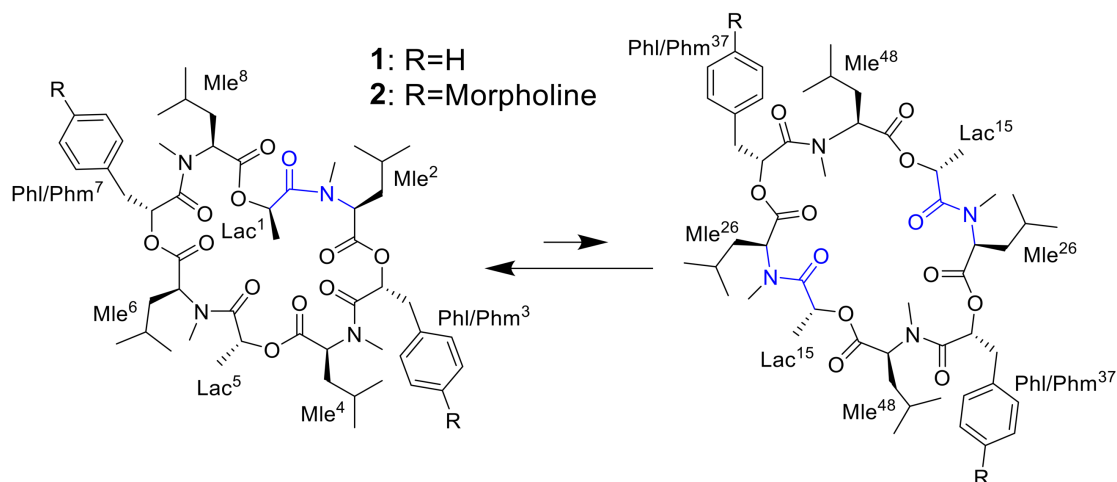
Scheme 2.1: Chemical structures of cyclic depsipeptides PF1022A (1) consisting of four L-N-methyl leucines (Mle), two D-lactic acid moieties (D-Lac) and two D-phenyllactic acid moieties (D-Phl), its synthetic derivative emodepside (2) with two additional morpholine rings in para position of the phenyllactic acid residues (D-Phm), 3,6-di-(propan-2-yl)-4-methyl-morpholine-2,5-dione (3), cyclo-(N-methyl L-leucine D-hydroxyisovaleric acid)₂ (4), enniatin B (5) consisting of three repetitions of L-N-methyl valine and D-hydroxyisovaleric acid, beauvericin (6) consisting of three repetitions of L-N-methyl phenylalanine and D-hydroxyisovaleric acid and verticilide (7) consisting of four repetitions of L-N-methyl alanine and D-2-hydroxyheptanoic acid.

Some of the known members of the subfamily of fully backbone N-methylated cyclic depsipeptides with varying core ring sizes are shown in Scheme 2.1. The smallest members consist only of one N-methylated amino acid and one hydroxy acid ($n=1$). For example, 3,6-di-(propan-2-yl)-4-methyl-morpholine-2,5-dione (3) is a natural product and was identified as a potential precursor of the cyclic hexadepsipeptide enniatin B (5).^{107,108} It showed moderate antioxidant and

antimicrobial activity.^{109,110} Structurally, both the amide and the ester bond in the six-membered ring are in *cis*-configuration.¹⁰⁸ The next larger members consist of two repetitions of an amino acid and a hydroxy acid ($n=2$). In NMR studies of cyclo-(N-methyl L-Leucine hydroxyisovaleric acid)₂ (**4**) in chloroform, a C₂-symmetric conformation was observed.^{111,112} In contrast to enniatin B (**5**) ($n=3$), it showed no activity against mycobacteria.¹¹² Enniatin B consists of three N-methyl L-valine and three D-hydroxyisovaleric acids and adopts, based on NMR studies, a C₃-symmetric conformation with all amides in *trans*-configuration in chloroform.¹¹³ It is a well-known antibacterial, anthelmintic, antifungal, herbicidal and insecticidal compound.¹¹⁴ Due to the lipophilic nature of **5**, it can be easily incorporated into lipid bilayers of cell membranes. Enniatin B was found to be ionophoric, i.e., it can carry mono and divalent cations through membranes with a selectivity for K⁺ over Na⁺.^{113,114} Further, it can form stable complexes with cations in solution. A 1:1 as well as a 2:1 sandwich (peptide : cation ratio) complex were observed.¹¹⁵ A 3:2 complex was proposed as well but with lower stability than the 1:1 and the 2:1 complexes.¹¹⁵ Enniatin B showed decent permeability ($\log P_e = -4.73$) in a passive artificial membrane permeability assay (PAMPA)¹¹⁶ and a permeability of $6.1 \cdot 10^{-4}$ cm/s in a Caco-2 permeability assay.^{106,117} Beauvericin (**6**) belongs, like **5**, to the enniatin family. It consists of three alternating N-methyl L-phenylalanine and D-hydroxyisovaleric acid residues and was observed in NMR experiments to adopt a C₃-symmetric conformation with all amides in *trans*-configuration in chloroform.^{91,118} **6** shows cytotoxic, apoptotic, anticancer, anti-inflammatory, antimicrobial, insecticidal and nematocidal activities and is able to transport cations, particularly Ca²⁺ through lipid bilayers.¹¹⁹ The passive membrane permeability of **6** was determined to be $5.8 \cdot 10^{-4}$ cm/s in a Caco-2 permeability assay,¹¹⁷ which is similar to the permeability of **5**. Verticilide (**7**) is a cyclic octadepsipeptide ($n=4$) such as PF1022A (**1**) and emodepside (**2**), and consists of four repetitions of N-methyl L-alanine and four repetitions of D-2-hydroxyheptanoic acid.¹²⁰ **7** was found to be a ryanodine-binding inhibitor and appears in NMR experiments in chloroform as two – not further studied – conformations in a ratio of 3:4.¹²⁰ Simplification of the NMR spectra of **7** was observed after the addition of a 100-fold excess of KSCN and only one conformer was detected.¹²⁰

The investigated compounds are only poorly soluble in water. Therefore, methanol and chloroform were chosen as simple mimics for a polar environment and the cell membrane, respectively. In both solvents, the NMR spectra of PF1022A (**1**) revealed two main conformations that interconvert slowly on the NMR time-scale (Scheme 2.2). The conformer ratio of **2** has been reported to be 4:1 in methanol and 3:1 in chloroform in previous studies.^{99,121} The major conformation is characterized by a single *cis*-amide bond between the D-lactic acid and the N-methyl-L-leucine residue and is thus named *asymmetric*, whereas all amide bonds are *trans* in the

minor conformation, thus named *symmetric*.^{99,121} The crystal structure of **1** apparently shows the asymmetric conformation, however, the data was not deposited with the CCDC (CCDC code MORJEI).¹²² The crystal structure of **2**, on the other hand, is the symmetric conformation (CCDC code DOMZOW).¹²³

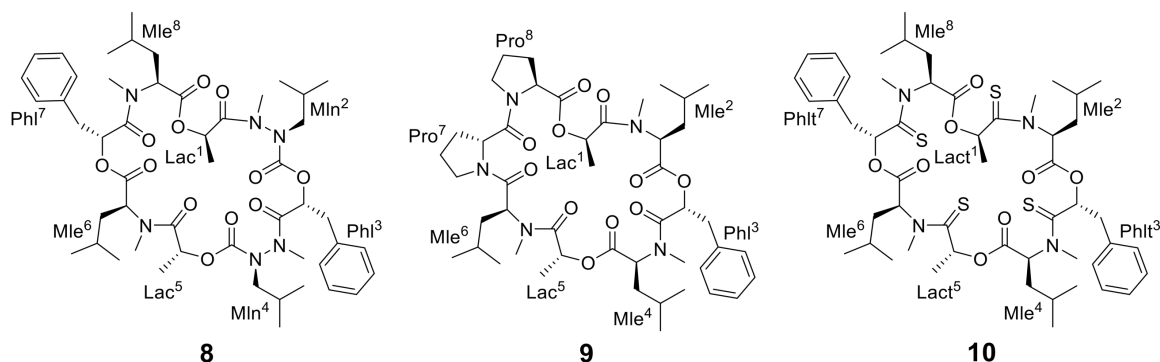


Scheme 2.2: Asymmetric (left) and symmetric (right) conformations of the two cyclic octadepsipeptides PF1022A (**1**) and emodepside (**2**) consisting of four L-N-methyl leucines (Mle), two D-lactic acid moieties (D-Lac) and two D-phenyllactic acid moieties (D-Phl) (with additional morpholine rings in para position in case of **2** (D-Phm)). In the C_2 symmetric conformation, the chemically equivalent residues share a common designation derived from their position in the asymmetric conformation.

Different side-chain and backbone modifications of **1** have been investigated in literature.^{121,124-126}

An interesting compound with regard to its conformational behavior is the bis-aza analog of PF1022A (**8**) (Scheme 2.3), where the asymmetric conformation is stabilized with a 100:7 conformer ratio in chloroform.¹²⁴ The conformation solved in the crystal structure is also asymmetric (CCDC code QOXDOW).¹²⁴ The biological activity of **8** was found to be weaker by a factor of 5 - 10 compared to **1**.¹²⁴ In another modification with a turn-inducing element consisting of two prolines (D-Pro-L-Pro) (**9**) (Scheme 2.3), it was reported that the symmetric conformation is stabilized such that only this conformer is present in solution.¹²⁵ Furthermore, the biological activity of **9** was found to be higher by a factor of 2 compared to **1**.¹²⁵ These observations led to the hypothesis that the propensity for the symmetric conformation is crucial for anthelmintic activity. However, for a third modification of **1**, in which the four peptide bonds were replaced by thiopeptide bonds (**10**) (Scheme 2.3), the activity was also increased 2.5 times compared to **1**.¹²⁶ In this case, the increased activity was attributed to a more rigid asymmetric conformation by the N-methyl-thioamides, which enhance the *cis*-amide bond between D-thiolactic acid and N-methyl-L-leucine.^{121,126} Based on the published data, no clear correlation between activity and conformational preference for the symmetric or asymmetric structure can be found, especially if it is considered that an increase or decrease of activity by a factor of 2 is mostly within the accuracy

of experiment. Additionally, no experimental membrane permeability data for **1**, **2** and **8–10** is reported in the literature.



*Scheme 2.3: Chemical structures of the bis-aza PF1022A analog (**8**) in which two C α carbons in N-methyl residues are replaced by nitrogens (Mln), of the di-proline PF1022A analog (**9**), in which residues 7 and 8 are replaced by a turn inducing D-Pro L-Pro moiety and of a tetra thioamide PF1022A analog (**10**) in which lactate and phenyllactic acid residues are replaced by their corresponding thio-analogs (Lact and Phlt).*

The mechanism of action of **1**, **2** and related compounds is not yet fully understood. Initially, their anthelmintic activity was attributed to the binding of a presynaptic latrophilin receptor.¹⁰⁰ More recently, binding to the calcium-activated potassium channel SLO-1 was proposed to be involved in the activity of **2**, possibly in combination with the latrophilin receptor.^{127–130} No crystal structure of **1** or **2** bound to one of these proteins is available. PF1022A and derivatives were reported to be ionophores with selectivity for K⁺ over Na⁺,¹³¹ similar to enniatin B. However, the ion carrier property across lipid bilayers does not appear to be related to the anthelmintic activity, because the enantiomer of PF1022A (i.e., all D- and L-residues switched) exhibited the same ionophoric ability but no anthelmintic activity.¹³¹

In this study, the interplay between the macrocyclic core conformational behavior of PF1022A, emodepside and related compounds with their ionophoric nature and their passive membrane permeability was sought out, to enhance our understanding for the rational design of such cyclic octadepsipeptides with improved profiles. For this, we characterize the conformational behavior of **1**, **2** and **8** and the effect of monovalent cations on the conformational ensembles using solution NMR measurements and extensive MD simulations in chloroform and methanol. With this data, we want to characterize how the cyclic depsipeptides interact with cations and determine a plausible coordination mode. The complexation with a cation could be an effective mechanism to bury the polar groups and thus, may be a crucial step for the incorporation of the depsipeptides into the membrane. The passive membrane permeability is assessed with PAMPA with and without the addition of potassium.

2.2 Results

2.2.1 Characterization of the Conformational Behavior

NMR MEASUREMENTS IN METHANOL AND CHLOROFORM

NMR spectra of **1**, **2** and **8** were recorded in CD₃OH and CDCl₃. The conformer ratios (Table 2.1) are in good agreement with those reported previously in the literature.^{99,121,124} A small batch-to-batch variability in the conformer ratio of **1** (ratios between 5:1 to 7:1 in methanol) was observed. The assignment of the major and minor conformer of **1** and **2** as well as of the major conformer of **8** in CDCl₃ and CD₃OH including proton, carbon and partly also nitrogen chemical shifts can be found in the Appendix.

Table 2.1: Ratios between asymmetric and symmetric conformer in CD₃OH and CDCl₃ for compounds **1**, **2** and **8**. Literature values are given in parentheses. Ratio marked with * was reported in CD₃OD.

Compound	Conformer ratio in CD ₃ OH (asymmetric : symmetric)	Conformer ratio in CDCl ₃ (asymmetric : symmetric)
PF1022A (1)	5 : 1 – 7 : 1 ^a (4 : 1 [*]) ⁹⁹	3 : 1 (3 : 1) ¹²⁶
Emodepside (2)	7 : 1	7 : 2
Bis-aza analog (8)	12 : 1	10 : 1 (100 : 7) ¹²⁴

^a The variability is likely due to residual cation content originating from synthesis, workup and purification that differs from batch to batch.

For all the investigated peptides, exchange peaks (EXSY peaks) could be detected in ROESY spectra recorded in chloroform-d. In CD₃OH, EXSY peaks could only be detected for **8** but low intensity and limited resolution did not allow further analysis. Besides the expected EXSY cross-peaks between the major asymmetric and the minor symmetric conformer for **1** and **2**, additional EXSY peaks are present, which indicate that more than the two known conformations are populated in solution. At least two additional low-intensity conformers could be identified (see Appendix). Using the volumes of the EXSY cross-peaks it is possible to calculate the site-to-site exchange rates k_1 and k_2 between the magnetic sites in the interconverting conformers (Figure 2.1). The additional two low-intensity conformers were neglected in the calculation of the exchange rates since their intensity was close to the noise level and their corresponding diagonal-peaks were partially buried under other, more intense signals. The calculated site-to-site rates are summarized in Table 2.2.

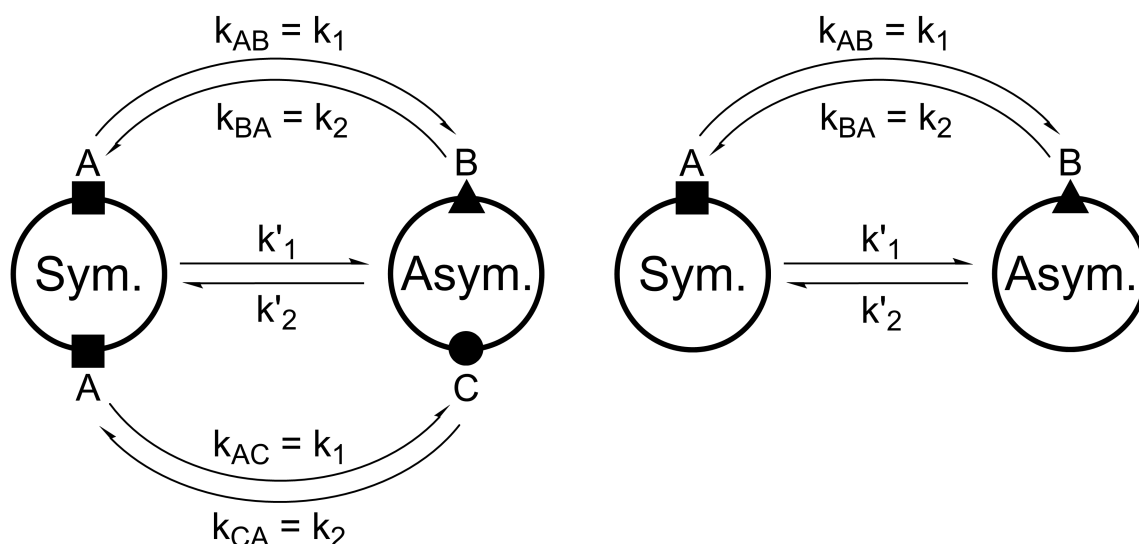


Figure 2.1: Schematic drawing of the magnetization transfer pathways used for the analysis of the EXSY data for **1**, **2** (left) and **8** (right). In the symmetric conformation of **1** and **2**, one of the two chemically equivalent amide bonds can flip into a cis-configuration to reach the asymmetric conformation (amide bond between Lac¹⁵ and Mle²⁶, see Scheme 2.2). In this process, magnetization is transferred via two different site-to-site pathways (A \leftrightarrow B and A \leftrightarrow C with $k_{AB}=k_{AC}=k_1$ and $k_{BA}=k_{CA}=k_2$), each leading to a separate set of EXSY cross-peaks. During a transition from the symmetric to the asymmetric conformation, each nucleus in a symmetric pair undergoes either pathway equally likely. In the reverse process from the asymmetric to the symmetric conformation a nucleus at site B will always follow A \leftrightarrow B whereas a nucleus at site C will always follow A \leftrightarrow C. As a consequence, the site-to-site exchange rates k_1 and k_2 for **1** and **2** differ from the mechanistic exchange rates: $k'_1=2*k_1$ and $k'_2=k_2$ where $K=k'_1/k'_2$. In **8**, the C₂ symmetry is broken by the two additional nitrogen atoms in the backbone and only a single magnetization transfer pathway has to be considered. Therefore for **8** $k_1=k'_1$ and $k_2=k'_2$.

Table 2.2: Site-to-site exchange rates between asymmetric and symmetric conformers measured in EASY-ROESY experiments with mixing time of 100 ms in CDCl₃.

Compound	k_1 [s ⁻¹]	k_2 [s ⁻¹]	k_{ex} [s ⁻¹]
PF1022A (1)	0.16	0.09	0.25
Emodepside (2)	0.12	0.06	0.18
Bis-aza analog (8)	0.17	0.02	0.19

The site-to-site exchange rates of **1**, **2** and **8** are comparable and are about twice as high compared to the exchange rate reported for cyclosporine A ($k_{ex} \approx 0.1 \text{ s}^{-1}$).¹³² This is plausible as the smaller ring size of the cyclic octadepsipeptides (24-membered ring) compared to cyclosporine A (33-membered ring) increases the ring strain. Since these results are based on a single mixing time, no direct error estimate can be given. From the comparison of the cross-peak intensities on both sides of the ROESY spectrum, errors about 20 % can be assumed.

In ¹H and ¹³C NMR spectra of the investigated cyclic octadepsipeptides, the signals for the symmetric conformer were generally found to be broader. To quantify this additional exchange broadening, presumably originating from processes on the millisecond to microsecond range, ¹³C T₂ relaxation time measurements of **1** in CDCl₃ were performed (Figure 2.2). It is clearly visible that

the symmetric conformer has shorter T_2 relaxation times for the backbone carbons compared to the asymmetric conformer. This indicates greater backbone flexibility on the μs to ms time scale for the symmetric conformer. To the best of our knowledge, this is the first time that such behavior was observed for a cyclic depsipeptide.

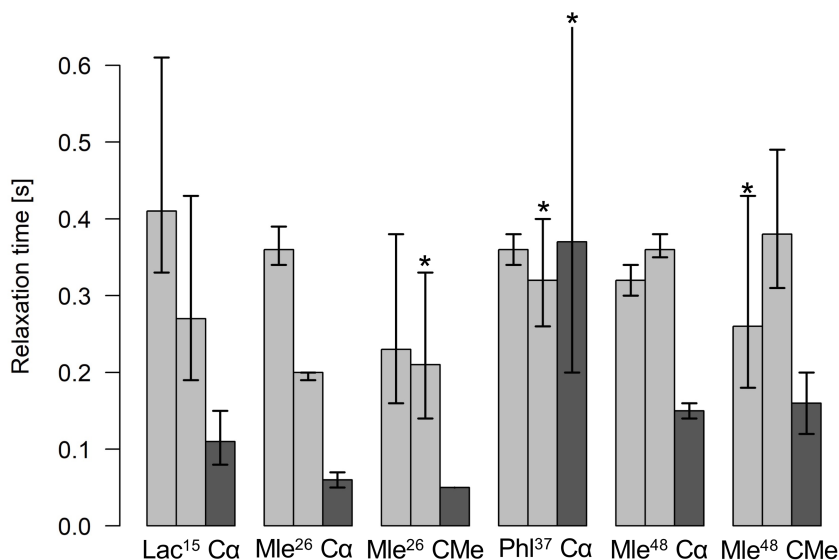


Figure 2.2: ^{13}C T_2 relaxation times measured for 20 mM PF1022A (**1**) in CDCl_3 with a series of ^{13}C -CPMG HSQC spectra with relaxation delays from 15.2 to 456 ms and with compensation of heating effects. Entries marked with * belong to partly overlapping peaks. The first two light grey bars belong to the asymmetric conformation (i.e., $\text{Lac}^1 \text{Ca}$ and $\text{Lac}^5 \text{Ca}$) whereas the third bar (dark grey) belongs to the symmetric conformation (i.e., $\text{Lac}^{15} \text{Ca}$). Error bars indicate the 95 % confidence interval of the fit.

KINETIC MODELS BASED ON MOLECULAR DYNAMICS (MD) SIMULATIONS

Extensive MD simulations of **1**, **2** and **8** were performed in methanol and chloroform using the GROMOS simulation package¹³³ and the GROMOS 54A7 united-atom force field.⁵⁷ As starting structures, the symmetric crystal structure of emodepside (**2**) (CCDC code DOMZOW) and the asymmetric crystal structure of the bis-aza analog (**8**) (CCDC code QOXDOW) were used. No significant differences in the structural ensemble could be detected between them. In general, the symmetric backbone configuration was found to be over-stabilized in the MD simulations, although the asymmetric configuration is more stable according to NMR. No *trans*-to-*cis* isomerizations were observed in the simulations, whereas five *cis*-to-*trans* isomerizations occurred. This is likely a force field issue, because *trans*-amide bonds are generally preferred in protein crystal structures. As a proof-of-principle, the partial charges in the methylated amide were slightly redistributed, which reduced the *cis*-to-*trans* isomerization rate drastically. As isomerizations are generally a rare event in finite simulations, we decided to analyze the symmetric and asymmetric conformations separately while assuming that the conformer distributions within the two sub-ensembles are correctly reproduced in the MD simulations.

Markov state models (MSMs)⁶⁶⁻⁶⁹ are a powerful tool to analyze the conformational dynamics in MD simulations. Here, we generated core-set Markov models of PF1022A (**1**) in chloroform using common nearest neighbor (CNN) based clustering^{68,134-136} and the PyEMMA package.¹³⁷ This procedure has been used successfully with other cyclic peptides.³ The MSMs were constructed separately for the asymmetric and the symmetric conformations (but with the same TICA space¹³⁸). For the asymmetric subset, only two unconnected conformational states could be identified, whereby one arose from a single simulation and was considered as noise. Therefore, the backbone with the asymmetric configuration appears to be relatively rigid. In contrast, the backbone with the symmetric configuration shows substantially more flexibility, and seven conformational states could be observed (Figure 2.3). This is in line with the NMR experiments, where shorter T_2 relaxation times were observed for the symmetric conformer, indicating higher flexibility on the μs - ms time scale.

The conformational states 3 and 5, as well as 6 and 7, are in principle the same rotated by 180° due to the C_2 symmetry of the symmetric conformation. This allows for an easy check of convergence. It can be seen in Figure 2.3 (and Table A2.8 in the Appendix) that the model is not yet fully converged. Note that the starting structure of the simulation corresponds to state 7. Conversion from state 7 to state 6 is essentially a complete reorientation of the entire backbone. Thus, very long simulations ($>10 \mu\text{s}$) would be needed to obtain the same population for state 6. Nevertheless, the results also indicate that the conformational space for the symmetric conformation is already sampled quite extensively.

Asymmetric

Symmetric

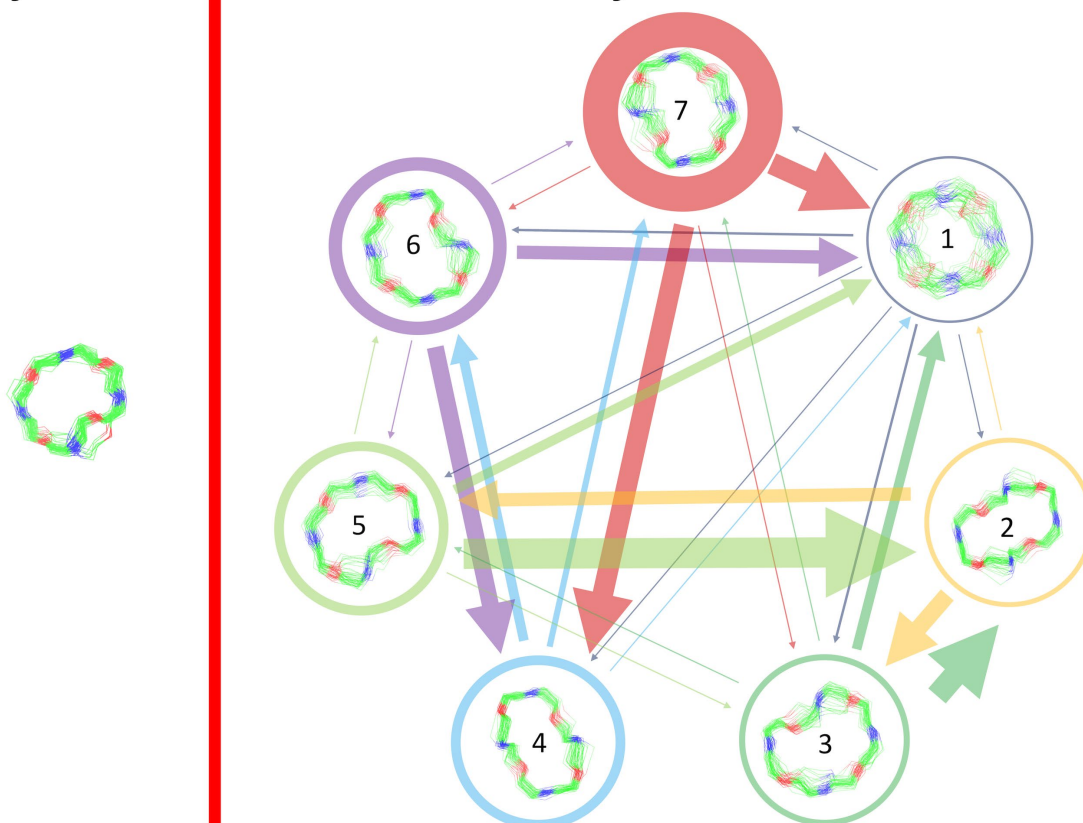


Figure 2.3: Visualization of the MSMs of the asymmetric and symmetric subsets of PF1022A (**1**) in chloroform. For each conformational state, 50 randomly picked backbone structures are shown. The thickness of the circle surrounding the state indicates the corresponding population with state 1 as the least and state 7 as the most populated conformational state. Note that state 3 and 5 as well as state 6 and 7 are chemically the same due to the C_2 symmetry of the symmetric conformation. The equilibrium populations are 7.4 % and 11.9 % for states 3 and 5, respectively, and 16.5 % and 43.2 % for states 6 and 7, respectively. A likely issue is that all simulations were started from the two available crystal structures. The arrows indicate the transition probabilities for state i going to state j within the chosen lag time (i.e., 10 ns). The arrow size corresponds to the magnitude of the probability. The subsets were analyzed separately because not enough transitions between symmetric and asymmetric conformers were observed.

2.2.2 Effect of the Presence of Monovalent Cations

BINDING AFFINITY AND CONFORMER RATIO AS A FUNCTION OF THE CATION CONCENTRATION

PF1022A (**1**) was previously reported to bind monovalent cations and act as an ionophore. However, no direct relationship between the ionophoric property and anthelmintic activity was found.¹³¹ Further, simplification of NMR spectra was observed by addition of KSCN but never described in any detail.¹²⁰ On the other hand, the connection of ion binding with the macrocycle conformational behavior as well as its importance for the membrane permeability is not yet clear. Therefore, we recorded NMR spectra in methanol of **1** and **2** in the presence of different concentrations of KSCN and NaSCN (in the case of **1** also NH_4SCN and CsSCN). In addition, the bis-aza analog **8** was titrated with KSCN. A significant change in chemical shift for the symmetric conformation was observed for **1** and **2** upon addition of the salts, with the effect being most

pronounced for Cs⁺ followed by K⁺, Na⁺ and NH₄⁺ (Figure 2.4). The cation preference is in line with the previous study.¹³¹ The change in chemical shift can be seen best for the H_α proton of PhI³⁷ in **1**, and the H_α proton of Phm³⁷ in **2** (Figure A2.6 in the Appendix). In addition, the ratio between the asymmetric and the symmetric conformer changes dramatically in favor of the symmetric conformation with increasing cation concentration (Table 2.3). Such a restriction to a single conformer was also seen for verticilide upon addition of KSCN, although no structural characterization was done in that case.¹²⁰

Table 2.3: Change in ratio between asymmetric and symmetric conformers without salt and after addition of a 40-fold excess of the salt (25-fold in case of CsSCN due to solubility issues) in CD₃OH.

Salt	PF1022A (1)	Emodepside (2)	Bis-aza analog (8)
KSCN	7 : 1 to 1 : 17	7 : 1 to 1 : 15	12 : 1 : 0.8
NaSCN	5 : 1 to 1 : 3	7 : 1 to 1 : 3	-
NH ₄ SCN	7 : 1 to 1 : 1	-	-
CsSCN	7 : 1 to 1 : 50	-	-

The changes in asymmetric : symmetric ratio upon addition of monovalent salt are comparable between **1** and **2** for KSCN and NaSCN. Consequently, the affinities of the two peptides for the cations is expected to be very similar. Therefore, for subsequent titrations only PF1022A (**1**) was used. In contrast, the bis-aza analog (**8**), that adopts the asymmetric conformer predominantly required a much higher salt concentration to observe a shift in the conformer ratio (Figure 2.5).

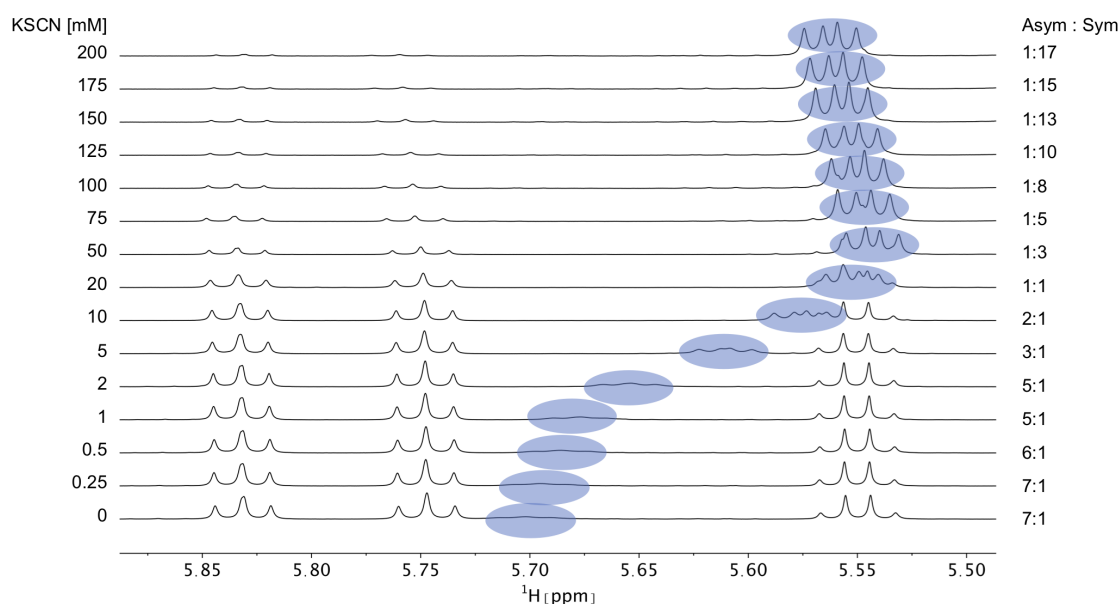


Figure 2.4: H_α region of ¹H NMR spectra of the titration of 5 mM **1** with a KSCN solution in CD₃OH. Chemical shift changes were observed for the symmetric conformation, best seen for the signal of the H_α proton in residue PhI³⁷ (blue labels). In addition, a change in the ratio between the symmetric and asymmetric conformation is observed. The asymmetric conformation shows small changes in chemical shift at high salt concentrations too. The other titration plots can be found in the Appendix.

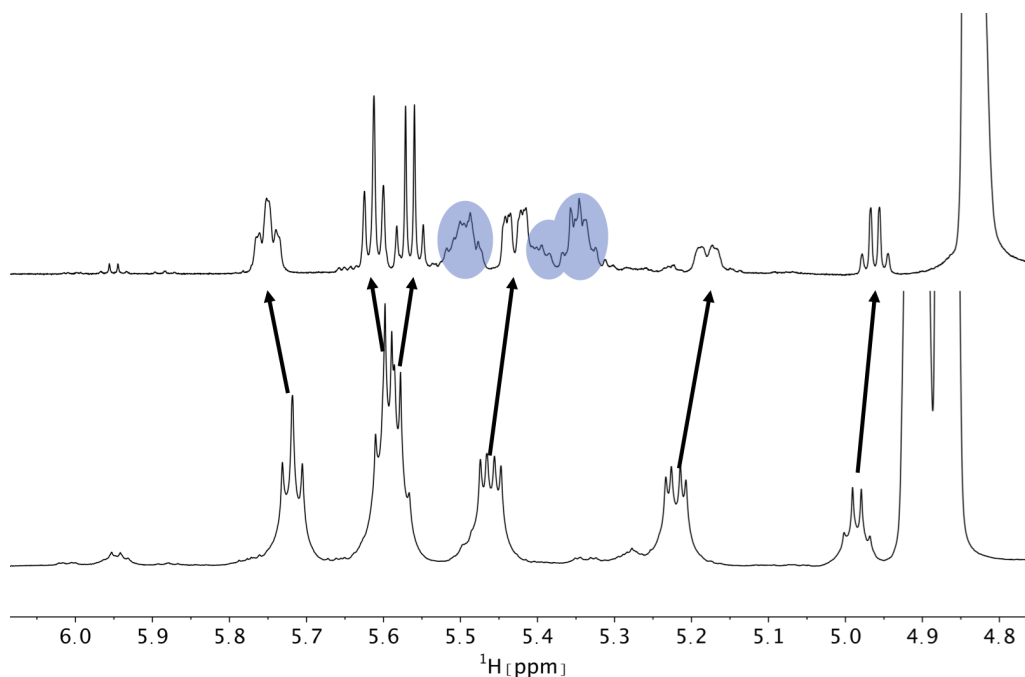


Figure 2.5: $H\alpha$ region of the 1H NMR spectra of a 5 mM solution of the bis-aza analog (**8**) without (bottom) and with 200 mM KSCN (top) in CD_3OH . Chemical shift changes were observed for the asymmetric conformation. Compared to **1** and **2**, the change in ratio between asymmetric and symmetric conformation is less pronounced and is close to 1:1 at a 40-fold excess of KSCN. Peaks of the symmetric conformation are marked in blue. The arrows indicate the movement of the asymmetric peaks upon addition of KSCN. On the right, the residual solvent peak is visible.

The titration data of PF1022A (**1**) with KSCN and CsSCN (as well as **2** with KSCN) can only be explained by a model containing at least two different ion-bound symmetric species, which are in fast exchange with the unbound symmetric conformation. In the case of a simple mixture of the free depsipeptide and only a 1:1 complex, the observed chemical shift is expected to change from the value of the free conformer towards that of the ion-bound conformer. However, we do not observe this asymptotic behavior. Instead, first the chemical shift drops with increasing salt concentration, then reaches a minimum and increases again at high concentrations. This indicates that at least a third symmetric species, which interacts with the ion, is populated. We propose a mixture of a 2:1 (peptide : cation ratio) and a 1:1 complex in solution, as was reported for enniatin B (**5**) and beauvericin (**6**).^{113,115} Such a mixture was already postulated for PF1022A (**1**) but not supported by any experimental data.¹³⁹ Normally, fitting of the equilibrium constants K_1 and K_2 is straightforward using the measured change in chemical shift in dependence of the salt concentration.¹⁴⁰ However, this system is more complicated due to the pre-equilibrium between the free asymmetric and symmetric conformers, and possibly additional species such as a 2:1 complex with one symmetric and one asymmetric conformer, or an asymmetric ion-bound conformer. We fitted our data with a model containing the free peptide in its symmetric conformation, the symmetric 1:1 complex, and the symmetric 2:1 complex. Instead of explicitly considering the pre-equilibrium, we have used the total concentration of all symmetric species

obtained from integration of the ^1H spectra. We interpret the results only qualitatively since similar fits may be achieved with different fitting parameters. Figure 2.6 clearly shows that the change in the asymmetric : symmetric ratio can be used to qualitatively measure the cation affinity of the symmetric conformer. The order of affinities with $\text{Cs}^+ > \text{K}^+ > \text{Na}^+ > \text{NH}_4^+$ is in agreement with those reported in literature,¹³¹ where alkali metals from Li^+ to Cs^+ were tested. If the change in chemical shift is plotted as a function of the salt concentration while keeping the peptide concentration constant, it can be observed that the change in chemical shift at high salt concentration is ordered by cation size. One could therefore speculate that the backbone of the depsipeptide has to adapt more extensively to accommodate smaller ions. This, in turn, leads to larger chemical shift changes in these complexes.

A consistent pattern is visible when comparing the plots on the left side and on the right side of Figure 2.6. A higher salt concentration is needed to achieve a 1:1 ratio between the asymmetric and the total symmetric species than for a 50 % change in chemical shift. The apparent lag increases with decreasing ion affinity. One can show that this behavior can already be reproduced by two coupled equilibria (ion independent conformational change and formation of the 1:1 complex). Its observation alone does not imply any cooperative phenomena or the presence of higher order complexes. Without further knowledge about the relative stabilities of the 2:1 and 1:1 complexes for each metal, a more detailed analysis is not possible at this stage.

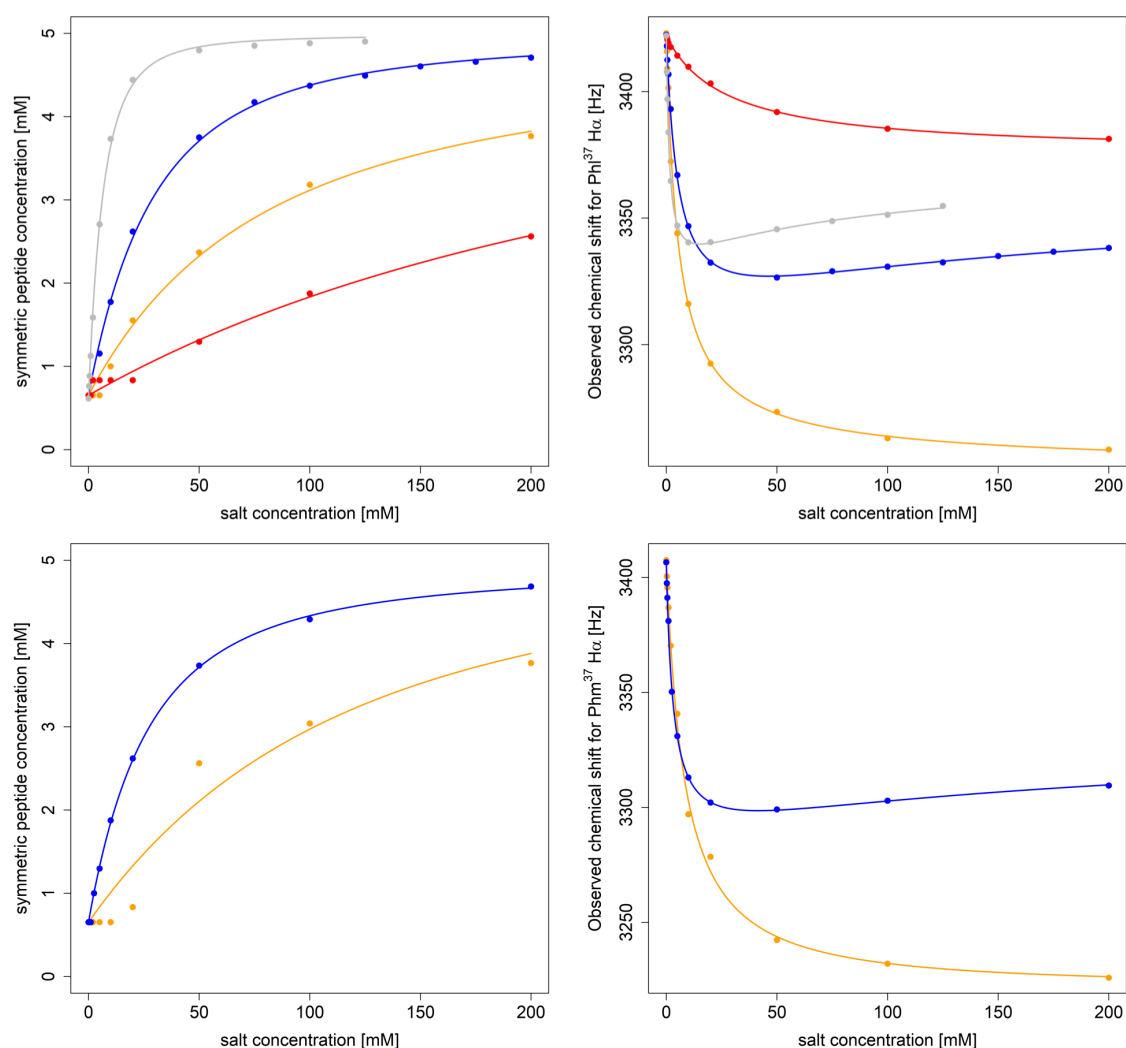


Figure 2.6: Titration of 5 mM PF1022A (**1**) (top) and 5 mM emodepside (**2**) (bottom) with different monovalent cations (CsSCN in grey, KSCN in blue, NaSCN in orange and NH₄SCN in red) in CD₃OH while the total volume was kept constant. The titration with CsSCN was only done up to 125 mM due to solubility issues. (Left): Change of the concentration of the symmetric conformation upon the addition of the corresponding salt. The data points were fitted with a damped logistic growth function (for details see Appendix). (Right): Change of the chemical shift of the PhI³⁷ H α proton as a function of the salt concentration (for details of the fit, see Appendix). The plots were generated with R.¹⁴¹

It is known that valinomycin, a cyclic dodecadepsipeptide, as well as some crown ethers can bind cations even in an apolar environment.^{142–144} This ability is an indirect evidence that the ion-bound complex may exist inside the membrane interior, i.e., that ion transport across a membrane is possible. To assess if the cyclic octadepsipeptides are also able to bind cations in an apolar solvent, KSCN was added to a solution of **1** in chloroform and sonicated for several hours. In subsequent NMR measurements, only the symmetric conformation could be detected in solution (Figure 2.7), which indicates ion binding.

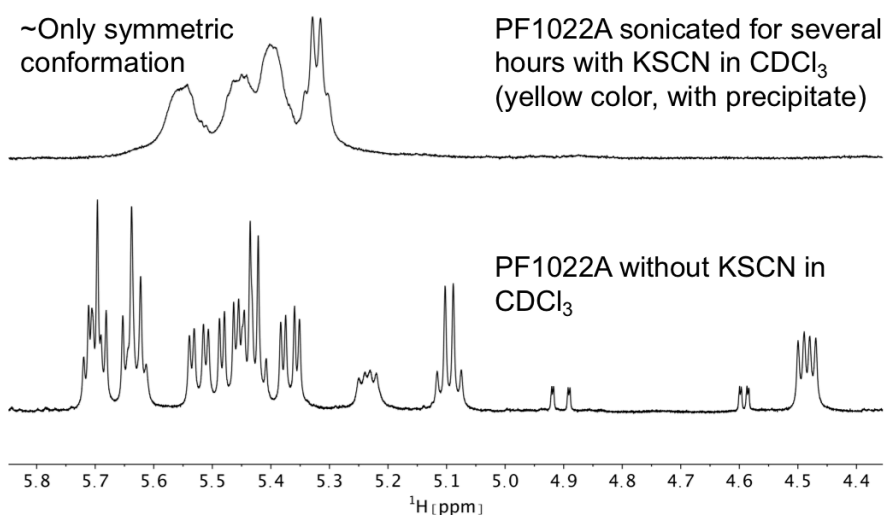


Figure 2.7: Comparison of ^1H NMR spectra of the H_α region of PF1022A (**1**) in CDCl_3 measured on a 500 MHz spectrometer. After the addition of KSCN and sonication, the symmetric conformation is present almost exclusively in solution. Note that the solution with the precipitate turned yellow.

The same effect was achieved by mixing a solution of emodepside (**2**) in chloroform with a saturated aqueous KSCN solution and letting the solution stand until phase separation has occurred (Figure 2.8). These results demonstrate that PF1022A and emodepside can carry cations from a polar phase into an apolar environment.

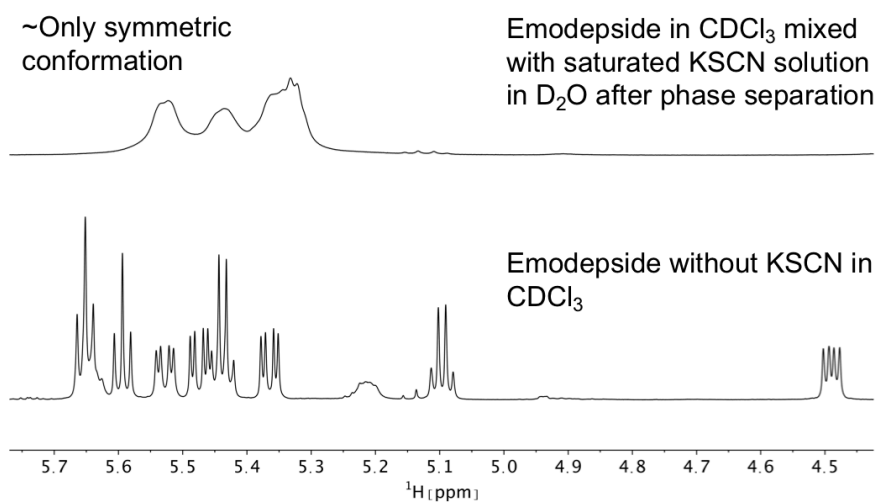


Figure 2.8: Comparison of ^1H NMR spectra of the H_α region of emodepside (**2**) in CDCl_3 measured on a 600 MHz spectrometer. After mixing with a saturated KSCN solution in D_2O , followed by sonication and phase separation, the symmetric conformation is present almost exclusively in solution.

CHARACTERIZATION OF THE ION-BOUND COMPLEX STRUCTURE

The possible structure of the depsipeptide-ion complex was first investigated *in silico*. MD simulations in presence of a K^+ ion starting from the symmetric crystal structure for **1** and **2** in methanol (10 μ s) and chloroform (1 μ s) as well as an MD simulation starting from the asymmetric crystal structure for **1** in chloroform showed that the ion binds to the peptide in the symmetric conformation independent of the starting structure, as expected from the experiment. Furthermore, a cavitand-like structure was adopted, in which the four amide oxygens and the two phenyl rings interact with the cation (Figure 2.9). In this highly symmetric conformation, the polar groups are saturated by the metal ion, whereas the side-chains of the N-methyl leucine residues shield them against the apolar environment.

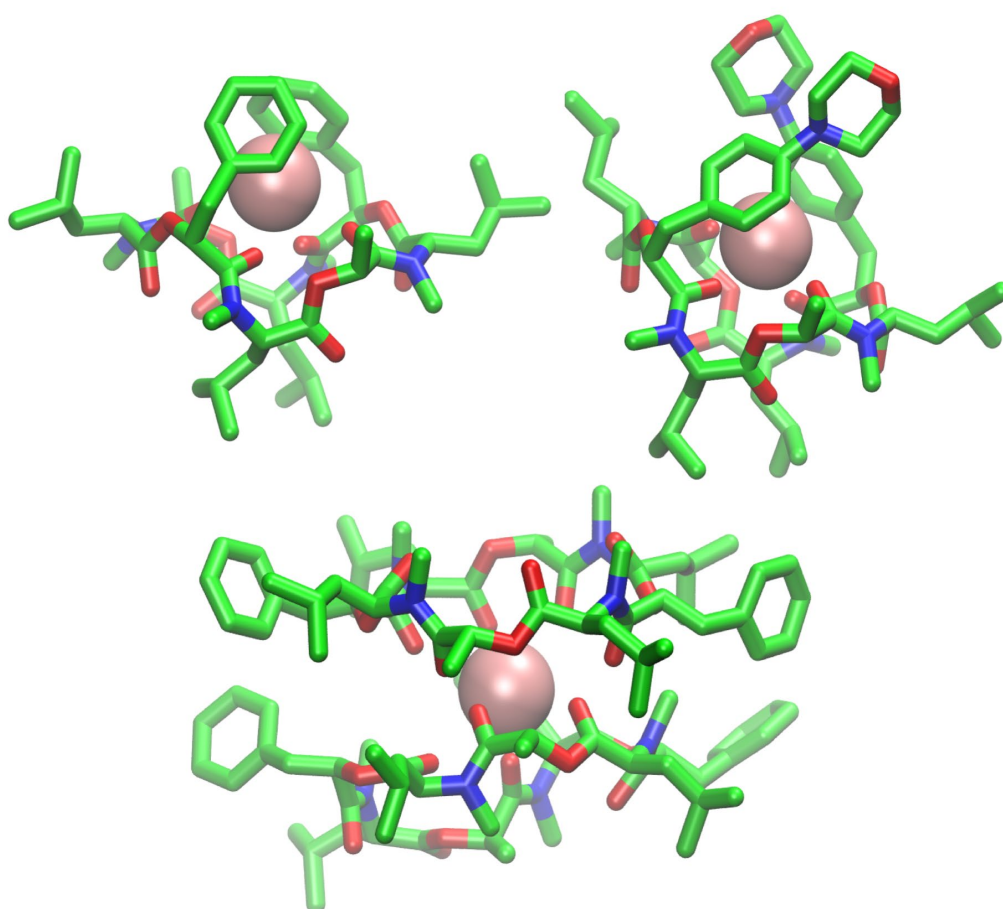


Figure 2.9: (Top): Snapshot of the 1:1 complex from the MD simulation of a single molecule of **1** (left) and **2** (right) in chloroform in presence of a single potassium ion (pink). Both depsipeptides adopt a cavity-like conformation with the cation bound in the center. The same structure could be observed for **1** in methanol after longer simulation time. (Bottom): Snapshot of the 2:1 complex from the MD simulation of two molecules of **1** in chloroform in presence of a single potassium ion. Carbons are shown in green, nitrogen atoms in blue, oxygen atoms in red and potassium ions in pink. The figures were generated with VMD.¹⁴⁵

The ion-bound conformation in the MD simulations is, however, dependent on the system setup. In simulations with two molecules of PF1022A (**1**) in chloroform (1 μ s) in presence of a single potassium ion, both a 1:1 and a 2:1 complex (Figure 2.9) could be observed over the course of the simulation, whereby the 1:1 complex did not adopt a cavitand-like structure.

To verify the cavitand-like structure of the 1:1 complex experimentally, we first aimed to crystallize PF1022A (**1**) in the presence of KSCN. Crystallization attempts with equimolar salt and peptide concentration led to separate crystals of KSCN and **1**, in which **1** is crystallized in the asymmetric conformation with one co-crystallized methanol molecule (Figure 2.10). The structure agrees well with the asymmetric crystal structure of the bis-aza analog (**8**) (CCDC code QOXDOW), justifying the use of the latter as starting structure in the MD simulations of **1**. By increasing the KSCN concentration to a 10-fold excess in methanol, an ion-bound complex of **1** could be crystallized. The crystal structure revealed a 2:3 complex (peptide : cation), with co-crystallized methanol and one water molecule (Figure 2.11). The ion-bound peptide crystallized in the symmetric conformation as observed in the NMR experiments and the MD simulations. This complex is likely not the major structure present in solution. In an MD simulation, the 2:3 complex showed very low stability.

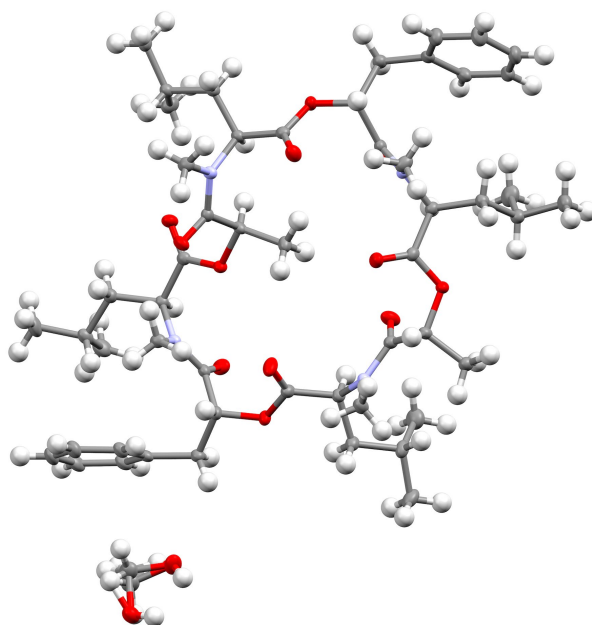


Figure 2.10: Crystal structure of PF1022A (**1**) (CCDC code: MORJEI01) crystallized in the asymmetric conformation. Carbon atoms are colored in grey, nitrogen atoms in light blue and oxygen in red. The ellipsoids represent 50 % of probability level and hydrogen atoms are shown with a radius of 0.3 Å. One methanol molecule is co-crystallized and disordered. The figure was created with Mercury.¹⁴⁶

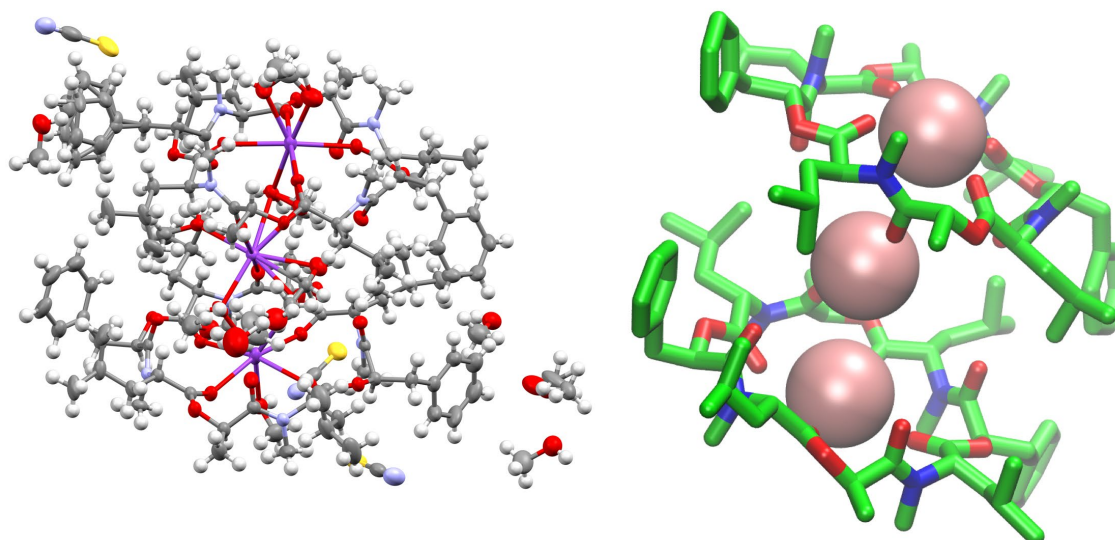


Figure 2.11: (Left): Crystal structure of a 2:3 complex of PF1022A (**1**) with KSCN (CCDC code: DUXQAR). There are three potassium ions (purple) crystalized with two molecules of the peptide. Carbon atoms are depicted in grey, nitrogen atoms in light blue, oxygen atoms in red, sulphur atoms in yellow and hydrogen atoms in white. The ellipsoids represent 50 % of probability level and hydrogen atoms are shown with a radius of 0.3 Å. One water molecule is co-crystalized as well as some methanol. The figure was generated with Mercury.¹⁴⁶ (Right): Simplified complex structure with only the non-hydrogen atoms present and without co-crystalized solvent molecules. Carbons are shown in green, nitrogen atoms in blue, oxygen atoms in red and potassium ions in pink. The figure was generated with VMD.¹⁴⁵

Since the crystallization experiments were not able to confirm the cavitand-like structure, we next turned to NMR to answer this question. The most straightforward evidence would be a through-space correlation between the two aromatic rings, which should be very close in the cavitand-like structure. However, this correlation is not experimentally accessible in these cyclic depsipeptides due to the C_2 symmetry of the symmetric conformer. One possible solution for this issue is to break the C_2 symmetry by introducing a substitution in the aromatic ring of one of the two phenyllactic acids. The PF1022A analog **11** contains an iodine substituent in para-position at one of the aromatic rings (Figure 2.12), and exhibits the same conformational behavior and ionophoric properties as **1** (experimental results summarized in the Appendix). With **11**, it should be possible to observe ROESY correlations between the two aromatic rings, if the cavitand-like structure is present in solution. However, such correlations were not observed (Figure 2.12). Therefore, the cavitand-like structure is likely an artifact of the setup in the MD simulation with a single peptide and potassium ion. This is further supported by the observation that no cavitand-like structure was adopted in the MD simulations with two peptides and a potassium ion (see discussion above).

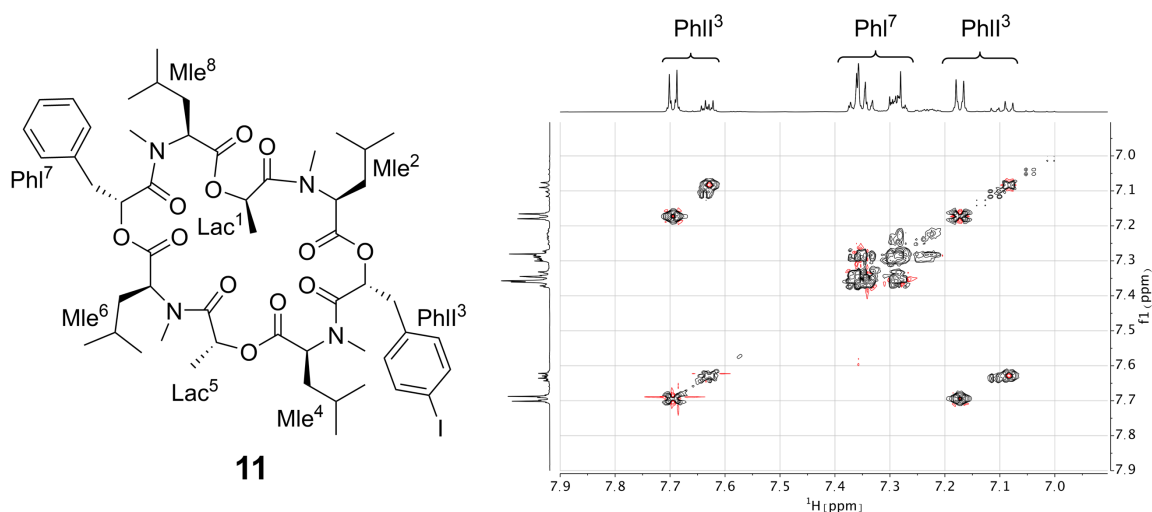


Figure 2.12: (Left): Chemical structure of the mono-iodine PF1022A analog **11**. (Right): EASY-ROESY spectrum of the aromatic region of 5 mM of **11** with 125 mM CsSCN in CD₃OH at room temperature with a mixing time of 700 ms. Only correlations within the aromatic rings were observed but no correlation between them.

2.2.3 Effect of the Presence of Monovalent Cations

Some members of the subfamily of cyclic depsipeptides with all backbone amides methylated have shown decent permeability in parallel artificial membrane permeability assays (PAMPA), e.g., for enniatin B (**5**) a log P_e value of -4.73 was determined.¹⁰⁶ For PF1022A (**1**) or emodepside (**2**), no permeability data has been reported in the literature. To assess whether the macrocyclic core conformational preference (**1** and **2** versus **8**) and the ionophoric property of the cyclic octadepsipeptides influence the passive permeability, PAMPA measurements with and without potassium salt was measured by our collaborators¹⁴⁷ using a protocol similar to that employed for enniatin B (**5**).¹⁰⁶ Surprisingly, no permeability was detected for PF1022A (**1**) and the related compounds (**2**, **8**) independent of the addition of potassium salt.¹⁴⁷ These results suggest that the investigated depsipeptides may not permeate but rather incorporate into the membrane (potentially bound to a cation in a 2:1 or 1:1 complex). Membrane incorporation would agree with the current hypothesis of the mode of action of emodepside, since SLO-1 and the latrophilin receptor are both associated with the membrane.¹²⁹ In addition, it was reported during electrophysiological studies that washout of PF1022A incorporated in membranes of CaCo-2 cells was not effective, indicating permanent incorporation into the membrane.¹³⁹ It would also not contradict the observation that emodepside is a substrate of the efflux transporter P-gp,¹⁴⁸ for which also a membrane-mediated mechanism is proposed.¹⁴⁹

2.3 Conclusion

In this work, we investigated the conformational behavior and ionophoric property of PF1022A (**1**), emodepside (**2**), and related compounds using NMR experiments and extensive MD simulations in order to establish a connection between them and potentially the membrane permeability. In support of previous literature, two major macrocyclic core conformers were detected in NMR measurements in chloroform and methanol, the major one with one amide bond in *cis*-configuration (asymmetric conformation) and the minor one with all *trans*-amide bonds (symmetric conformation). The symmetric core conformation showed a higher flexibility on the microsecond to millisecond time scale compared to the asymmetric one both in NMR (i.e., shorter T_2 relaxation times due to additional exchange contribution) and in kinetic models constructed from the MD data.

Upon addition of cations, a shift towards the symmetric conformation was observed in the NMR titration experiments, which indicates that only the symmetric conformation can bind tightly to the ions. A preference for cesium over potassium over sodium was found and in agreement with that reported previously. Furthermore, we could show that these cyclic octadepsipeptides can carry cations into an apolar solvent, like other ionophores. The titration curves indicate a mixture of both 1:1 and 2:1 (2 peptides and 1 cation) complexes. MD simulations suggest the formation of a sandwich complex, like the one observed for enniatin B (**5**). A cavitand-like structure of the 1:1 complex seen in the MD simulations could, however, not be confirmed experimentally using the mono-iodine substituted analog (**11**). Crystallization of PF1022A (**1**) with an excess of KSCN in methanol yielded a 2:3 complex (2 peptides and 3 potassium ions), where the peptides are in the symmetric conformation, confirming the findings in the NMR experiments and MD simulations.

The fact that the symmetric conformers can bind cations might still be relevant for activity, since the metal bound species may possess a higher propensity for membrane incorporation than the free peptide. This would also be in line with the location of the proposed target, SLO-1, a membrane-bound ion channel. The results of the PAMPA experiments and the ineffective wash-out of PF1022A from CaCo-2 membranes may indeed indicate that the peptides do not permeate but rather incorporate into the membrane. Extensive NMR and computational characterizations are in this case very important and provide further insight at atomic resolution beyond the scope of PAMPA. In terms of the investigated properties, no significant differences were found between **1** and **2**. The ratios between symmetric and asymmetric conformations in solutions as well as their binding affinities towards cations are similar. Thus, the difference in anthelmintic activity between **1** and **2** cannot be directly related to a difference in the conformational behavior or ionophoric

property, but likely to stem from the effect of the morpholino substitution modulating the potency at the target. The studied bis-aza analog (**8**), for which the asymmetric conformation is further stabilized, has a significantly lower affinity towards cations, which could be an indication that cation binding may be an important aspect for membrane incorporation, and potentially influence activity. Future studies with cyclic octadepsipeptides that exhibit different cation binding affinities might be able to further elucidate these connections.

2.4 Method Section

Peptide Synthesis

The methods for obtaining the depsipeptides investigated in this work have been previously reported in the literature.^{95,124,150}

NMR characterization of PF1022A (1), emodepside (2), bis-aza PF1022A analog (8) and mono-iodo analog (11) in CD₃OH and CDCl₃

20 mM solutions of **1** (12.3 mg), **2** (14.6 mg), **8** (12.4 mg) and **11** (14.0 mg) in methanol-d₃ (Armar) as well as in chloroform-d (Cambridge Isotope Laboratories) were used for the characterization by NMR. Because of solubility issues of **2** in methanol, a 6.7 mM solution was used instead (4.4 mg). A full set of spectra (¹H-NMR, ¹³C-NMR, TOCSY, double-quantum filtered COSY, multiplicity edited ¹³C-HSQC with adiabatic decoupling, ¹³C-HMBC, ¹⁵N-HMBC and EASY-ROESY¹⁵¹) was recorded for each compound except for **11** where no ¹³C-NMR spectrum was recorded. If not stated otherwise all spectra were measured at 25 °C on a Bruker Avance III HD 600 MHz spectrometer equipped with a N₂-cooled Prodigy triple resonance probe with z-gradients or on a Bruker AVANCE III 500 MHz spectrometer equipped with a BBFO broadband probe with z-gradients.

The CD₃OH signal was suppressed by presaturation or excitation sculpting.¹⁵² ¹³C-HSQC spectra were recorded with sensitivity enhancement¹⁵³ and multiplicity editing. TOCSY spectra were recorded with zero quantum filter¹⁵⁴ and 80 ms DIPSI2¹⁵⁵ isotropic mixing except for **1** in chloroform where 80 ms mlev17¹⁵⁶ mixing was used. The mixing time for the EASY-ROESY experiments was set to 100 ms if not otherwise stated. For all spectra, the time domain in both dimensions was extended to twice its size by zero filling, apodized with a cos² function, and the baseline of the resulting spectra was corrected with a polynomial of fifth order or using the Whittaker smoother algorithm.¹⁵⁷ Processing was done with Bruker TopSpin™ version 4.0 (Bruker Biospin AG) and MestReNova 12.0 (Mestrelab Research). Resonance assignment and integration of ROESY cross-peaks were performed with SPARKY 3.115.¹⁵⁸ ¹³C T₂-relaxation time measurements were done with a series of sensitivity enhanced ¹³C-CPMG-HSQC spectra¹⁵⁹ using a slightly modified version of Bruker standard pulse program hsqct2etf2gpsi with ten different evenly spaced relaxation delays between 15.2 ms and 456 ms. Heating effect compensation was used. Fitting of the exponential decays was done with Prism 8.4 (GraphPad Software).

Calculation of Exchange Rates

Following Refs. 160 and 161, site-to-site rates were determined in a straightforward way by taking the logarithm of matrix $A = M^*M_0^{-1}$ containing the volumes of cross- and diagonal-peaks of the exchanging sites (Ha of Mle², Mle⁶ and Mle²⁶) divided by their magnetic fraction M_0 (i.e., the relative intensities of the corresponding resonances in the ¹H NMR spectrum) as an approximation of $M(0)$:

$$M = e^{Lt_m} * M_0 \quad (\text{A2.1})$$

$$A = \begin{pmatrix} \frac{I_{AA}}{M_{0A}} & \frac{I_{AB}}{M_{0B}} & \frac{I_{AC}}{M_{0C}} \\ \frac{I_{BA}}{M_{0A}} & \frac{I_{BB}}{M_{0B}} & \frac{I_{BC}}{M_{0C}} \\ \frac{I_{CA}}{M_{0A}} & \frac{I_{CB}}{M_{0B}} & \frac{I_{CC}}{M_{0C}} \end{pmatrix} \quad (\text{A2.2})$$

$$\frac{1}{t_m} \ln(A) = \begin{pmatrix} k_{BA} + k_{CA} - R_1 & k_{AB} & k_{AC} \\ k_{BA} & -k_{AB} - R_1 & 0 \\ k_{BC} & 0 & -k_{AC} - R_1 \end{pmatrix} \quad (\text{A2.3})$$

L is the difference between the kinetic matrix K containing the site-to-site rate constants and the relaxation matrix R , t_m is the mixing time used in the ROESY experiment, and R_i are the auto-relaxation rates of the exchanging sites in the symmetric (A) and asymmetric (B, C) conformations.

As an example, the procedure is shown for PF1022A (1):

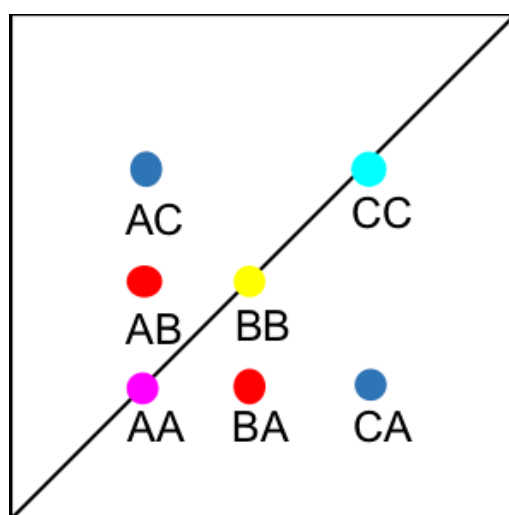


Figure 2.13: Schematic EXSY spectrum with sites A (symmetric conformation), B and C (asymmetric conformation). A exchanges with B and C but B does not exchange with C.

Peak volumes extracted from the EXSY spectrum and the magnetic fractions in a ^1H NMR spectrum are inserted in matrix *A*. Site-to-site rates are obtained by taking the logarithm of matrix *A* and dividing the result by the mixing time (0.1 s): $k_{\text{AB}} = k_{\text{AC}} = 0.09 \text{ s}^{-1}$ and $k_{\text{BA}} = k_{\text{CA}} = 0.16 \text{ s}^{-1}$ (averaged rates). Calculations were carried out in Mathematica 12.0.¹⁶²

$$A = \begin{pmatrix} \frac{77.9}{1} & \frac{1.52}{1.5} & \frac{1.55}{1.5} \\ 1.59 & 236 & 0 \\ \frac{1}{1.75} & \frac{1.5}{0} & \frac{1.5}{200} \\ \frac{1}{1} & \frac{1.5}{1.5} & \frac{1.5}{1.5} \end{pmatrix} \quad (4)$$

$$\frac{1}{0.1} \ln(A) = \begin{pmatrix} 43.5524 & 0.09 & 0.10 \\ 0.14 & 50.58 & 0.00 \\ 0.17 & 0.00 & 48.9 \end{pmatrix} \quad (5)$$

Titration With Monovalent Cations

CsSCN was prepared by dissolving Cs_2CO_3 (100 mg, 0.31 mmol, Sigma-Aldrich) and NH_4SCN (46.7 mg, 0.62 mmol, Sigma-Aldrich) in 0.5 ml water and was then crystallized at room temperature.¹⁶³ The crystals were washed with cold water and then dried in the oven at 105 °C. Aliquots of a 100 mM and a 1 M solution of KSCN (Fluka), NaSCN (Sigma-Aldrich) and NH_4SCN (Merck) and a 100 mM solution of CsSCN in methanol- d_3 were used to titrate a 5 mM solution of **1** and a 5 mM solution of **2** (only with KSCN and NaSCN) as well as a 5 mM solution of **8** (only with KSCN) and **11** (only with CsSCN). For compound **1**, ^1H spectra with solvent suppression using excitation sculpting were recorded at 0, 2.5, 5, 10, 20, 50, 100 and 200 mM KSCN. For the other titrations, 1D-NOESY spectra with presaturation (mixing time 10 ms) were recorded as it was observed that the intensities near the solvent signal were affected by the excitation sculpting. In addition, the base lines were flatter in the 1D-NOESY spectra, which was more favorable for integration of the peak intensities. Spectra were recorded at salt concentrations of 0, 0.25, 0.5, 1, 2, 5, 10, 20, 50, 100, and 200 mM, except for CsSCN, where the maximal concentration was 125 mM. For titrations of **1** with KSCN, additional data points at 75, 125, 150, and 175 mM were recorded. For titrations of **1** and **11** with CsSCN, additional data points at 75, 125 mM were recorded.

PAMPA Measurements

The measurements were carried out by Chad Townsend and Scott Lokey as described in Ref. 147.

MD Simulations

The GROMOS simulation package¹³³ was used for all simulations together with the GROMOS 54A7 united-atom force field⁵⁷ for the solvent and the peptides, and the 2016H66 force field¹⁶⁴ for the potassium ion. MD trajectories of 1 - 10 μ s length were produced under isothermal-isobaric conditions (NPT) using the leap-frog integration scheme with a time step of 2 fs.¹⁶⁵ The temperature was kept at 298 K by weak coupling to two separate temperature baths for the peptide and the solvent with a relaxation time of 0.1 ps and the pressure was kept at 1 atm by weak coupling to a pressure bath with a relaxation time of 0.5 ps and an isothermal compressibility of $4.5 \cdot 10^{-4} \text{ kJ}^{-1} \text{ mol nm}^3$.¹⁶⁶ A twin range cutoff scheme was used with cutoffs of 0.8 and 1.4 nm for the non-bonded interactions. Bond lengths were constraint with the SHAKE algorithm with a tolerance of 10^{-4} nm ¹⁶⁷ and center of mass removal was done every 1000 steps. MD simulations were performed in chloroform and methanol using dielectric permittivity coefficients taken from Ref. 168 for the dielectric continuum outside the cutoff (reaction-field method).¹⁶⁹ Simulations with K^+ ions were simulated without counter-charge because the two charged ions would aggregate in chloroform and methanol. The GROMOS++ program "ion" was used to replace the solvent molecule with the lowest electrostatic potential energy by a potassium ion.¹⁷⁰ The crystal structure of emodepside (**2**) (CCDC code DOMZOW)¹²³ was used as the symmetric starting structure, and the crystal structure of the bis-aza analog **8** as the asymmetric starting structure (CCDC code QOXDOW)¹²⁴. The peptides were minimized first in vacuum using a steepest-descent algorithm.¹⁷¹ The peptide was solvated in the corresponding solvent and the solvent was relaxed while the coordinates of the peptide were restraint with a force constant of $2.5 \cdot 10^4 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Afterwards, the system was thermalized to 298 K in five steps of 60 K and the force constant was loosened one order of magnitude in each step if not otherwise stated (Table 2.4). Initial velocities were generated using a Maxwell-Boltzmann distribution. Details of the performed simulations are summarized in Table 2.4.

Table 2.4: Details of the performed MD simulations. Thermalizations marked with * were done with a single step directly at 298 K instead of five steps. Simulations marked with # were done with modified partial charges for the methylated amides.

System	Starting structure (CCDC code)	Number of simulations	Solvent	Number of solvents	Length of thermalization per step [ps]	Length per MD simulation [μ s]
PF1022A	DOMZOW	1	CHCl ₃	329	2000*	10
PF1022A	DOMZOW	10	CHCl ₃	329	2000	1
PF1022A	DOMZOW	1	CH ₃ OH	637	20	1
PF1022A	QOXDOW	1	CHCl ₃	344	2000	10
PF1022A	QOXDOW	10	CHCl ₃	344	2000	1
PF1022A + K ⁺	DOMZOW	1	CHCl ₃	328	20	1
PF1022A + K ⁺	DOMZOW	1	CH ₃ OH	636	20	10
PF1022A + K ⁺	QOXDOW	1	CHCl ₃	343	20	1
2 PF1022A + K ⁺	DOMZOW	1	CHCl ₃	3085	2000	1
2 PF1022A + 3 K ⁺	DUXQAR	1	CHCl ₃	389	2000	1
2 PF1022A + 3 K ⁺	DUXQAR	1	CH ₃ OH	765	2000	1
Emodepside	DOMZOW	1	CHCl ₃	497	2000*	1
Emodepside	DOMZOW	1	CH ₃ OH	989	20	1
Emodepside + K ⁺	DOMZOW	1	CHCl ₃	496	20	1
Emodepside + K ⁺	DOMZOW	1	CH ₃ OH	988	20	10
Bis-aza analog	DOMZOW	11	CHCl ₃	330	2000	1
Bis-aza analog	QOXDOW	1	CHCl ₃	350	2000*	10
Bis-aza analog	QOXDOW	10	CHCl ₃	350	2000	1
Bis-aza analog	QOXDOW	1	CH ₃ OH	688	20	1
Bis-aza analog + K ⁺	QOXDOW	1	CHCl ₃	349	20	10
Bis-aza analog + K ⁺	QOXDOW	1	CH ₃ OH	687	20	1
PF1022A#	DOMZOW	1	CHCl ₃	329	2000*	10
PF1022A#	DOMZOW	10	CHCl ₃	329	2000*	1
PF1022A#	QOXDOW	1	CHCl ₃	344	2000*	10
PF1022A#	QOXDOW	10	CHCl ₃	344	2000*	1

Markov State Model (MSM) Building

MSMs were built using the PyEMMA package.¹³⁷ Ten 1 μ s and one 10 μ s MD simulations starting from the symmetric and from the asymmetric crystal structure were used to build the MSM in chloroform for PF1022A (**1**). Input features were all backbone dihedrals. Time-lagged independent component analysis (TICA)¹³⁸ was done with a lag time of 10 ns. A common nearest neighbor (CNN) density based clustering⁶⁸ with a similarity of 10 and a cutoff distance of 0.15 was applied.¹³⁶ 20 % of the input data was discarded as noise. The regions with asymmetric and symmetric conformations were not connected, since the *trans*-to-*cis* transition of the one amide bond was never sampled. The asymmetric set consisted of two non-connected subsets. One of them arose from a single simulation and was therefore discarded as noise. Implied time scales from a Bayesian MSM revealed six slow processes for the symmetric conformer. Chapman-Kolmogorov test¹³⁵ (Figure 2.14) was used to validate the model with seven conformational states. Finally, an MSM was constructed for these seven states (see main text and Figure A2.6 in the Appendix).

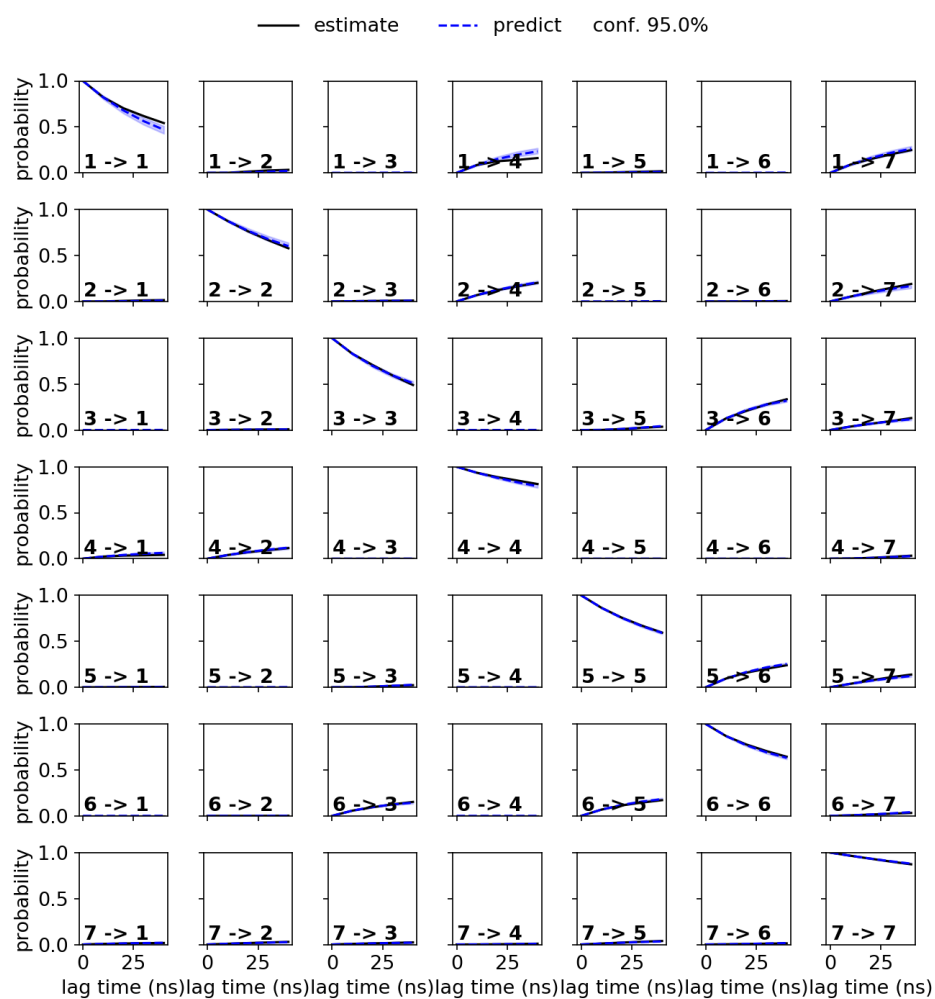


Figure 2.14: Chapman-Kolmogorov test for the symmetric conformer of **1** in chloroform with 7 states and a lag time of 10 ns.

Crystallization of 1 With KSCN in Methanol

Around 10 mg of **1** was dissolved in methanol together with an equimolar amount of KSCN (1.0 mg). The sample was put in the freezer at -28 °C. After five days, transparent crystals were obtained. Analysis was done by the small molecules crystallography center (SMOCC) at ETH Zürich. A XtaLAB Synergy, Dualflex, Pilatus 300K diffractometer was used for both measurements. The crystal was kept at 100 K during data collection. Using Olex2,¹⁷² the structure was solved with the ShelXT¹⁷³ structure solution program using Intrinsic Phasing and refined with the ShelXL¹⁷⁴ refinement package using least squares minimization. The obtained crystal structure was only the peptide without the salt in its asymmetric form. **Crystal Data** for C₅₃H₈₀N₄O₁₃ (*M* = 981.21 g/mol): monoclinic, space group P2₁ (no. 4), *a* = 14.40860(10) Å, *b* = 13.78330(10) Å, *c* = 14.46940(10) Å, *β* = 110.1570(10)°, *V* = 2697.59(4) Å³, *Z* = 2, *T* = 100.0(1) K, *μ*(Cu Kα) = 0.701 mm⁻¹, *D*_{calc} = 1.208 g/cm³, 74402 reflections measured (6.508° ≤ 2θ ≤ 159.456°), 11334 unique (*R*_{int} = 0.0423, *R*_{sigma} = 0.0233) which were used in all calculations. The final *R*₁ was 0.0280 (*I* > 2σ(*I*)) and *wR*₂ was 0.0699 (all data).

Around 10 mg of **1** was dissolved in methanol together with a tenfold excess of KSCN. The concentrated sample was put in the freezer at -28 °C. After three days transparent crystals were obtained and were given to SMOCC for analysis. The obtained crystal structure was a complex of two peptides with three ions with co-crystallized methanol molecules and one water molecule. **Crystal Data** for C₁₁₈H₁₉₈K₃N₁₁O₃₆S₃ (*M* = 2560.34 g/mol): triclinic, space group P1 (no. 1), *a* = 14.90080(10) Å, *b* = 15.83070(10) Å, *c* = 16.84690(10) Å, *α* = 111.2440(10)°, *β* = 101.4290(10)°, *γ* = 100.0950(10)°, *V* = 3495.22(5) Å³, *Z* = 1, *T* = 100.0(1) K, *μ*(CuKα) = 1.908 mm⁻¹, *D*_{calc} = 1.216 g/cm³, 95494 reflections measured (5.872° ≤ 2θ ≤ 159.716°), 28018 unique (*R*_{int} = 0.0424, *R*_{sigma} = 0.0371) which were used in all calculations. The final *R*₁ was 0.0502 (*I* > 2σ(*I*)) and *wR*₂ was 0.1453 (all data).

2.5 Appendix

NMR Assignments:

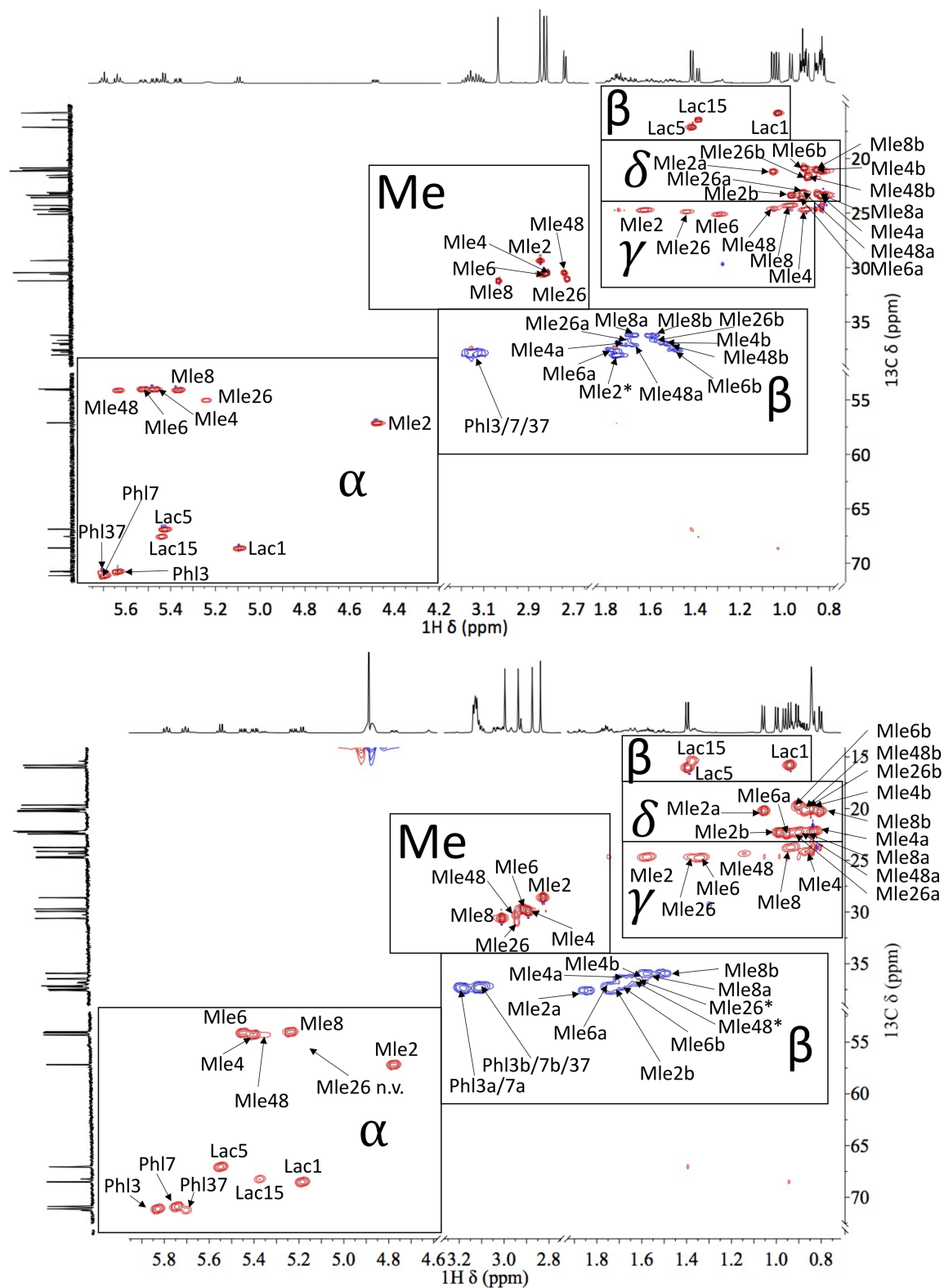


Figure A2.1: ^{13}C -HSQC spectra of PF1022A (1) in CDCl_3 (top) and in CD_3OH (bottom). Empty regions are cut out for clarity. N.v.: not visible with the employed contour levels.

Table A2.1: Assignment of the major (asymmetric) and minor (symmetric) conformation of PF1022A (**1**) in CDCl₃ referenced to the residual CHCl₃ shift in the solvent set to 7.29 ppm. * indicates non-assignable signals due to overlap.

	H α	H β	H γ	H δ	H ϵ	H ζ	HMe	
Lac ¹	5.10	1.03	-	-	-	-	-	
Mle ²	4.49	1.75	1.63	1.05, 0.97	-	-	2.85	
Phl ³	5.70	3.15	-	7.25	7.31	7.28	-	
Mle ⁴	5.47	1.73, 1.53	0.91	0.86, 0.84	-	-	2.82	
Lac ⁵	5.43	1.41	-	-	-	-	-	
Mle ⁶	5.52	1.77, 1.49	1.29	0.92, 0.91	-	-	2.83	
Phl ⁷	5.64	3.17, 3.11	-	7.26	7.30	7.27	-	
Mle ⁸	5.37	1.68, 1.60	0.98	0.83, 0.83	-	-	3.04	
Lac ¹⁵	5.44	1.39	-	-	-	-	-	
Mle ²⁶	5.23	1.72, 1.58	1.44	0.92, 0.90	-	-	2.73	
Phl ³⁷	5.70	3.16, 3.11	-	7.25-7.35*	7.25-7.35*	7.25-7.35*	-	
Mle ⁴⁸	5.63	1.68, 1.50	1.05	0.90, 0.85	-	-	2.74	
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe
Lac ¹	171.6	68.6	15.8	-	-	-	-	-
Mle ²	171.2	57.1	38.0	24.7	21.2, 23.4	-	-	29.4
Phl ³	170.0	71.2	37.6	135.1	129.5	128.6	127.2	-
Mle ⁴	169.8	54.0	37.0	24.7	21.0, 23.4	-	-	30.5
Lac ⁵	170.4	66.9	17.1	-	-	-	-	-
Mle ⁶	169.8	54.0	37.5	25.1	23.6, 20.9	-	-	30.6
Phl ⁷	170.2	70.8	38.0	135.4	129.5	128.5	127.1	-
Mle ⁸	171.0	54.1	36.2	24.3	23.4, 21.1	-	-	31.2
Lac ¹⁵	169.9	67.5	16.4	-	-	-	-	-
Mle ²⁶	170.6	55.1	36.7	24.9	23.1, 21.7	-	-	31.1
Phl ³⁷	169.4	70.8	37.8	135.4	129.7	128.5	127.0	-
Mle ⁴⁸	170.9	54.1	37.1	24.6	21.6, 23.2	-	-	30.5

Table A2.2: Assignment of the major (asymmetric) and minor (symmetric) conformation of PF1022A in CD₃OH referenced to the residual CH₃OH shift in the solvent set to 3.33 ppm. * indicates non-assignable signals due to overlap and n.d. indicates signals that could not be detected.

	H α	H β	H γ	H δ	H ϵ	H ζ	HMe	
Lac ¹	5.19	0.94	-	-	-	-	-	
Mle ²	4.78	1.85, 1.74	1.58	1.06, 0.99	-	-	2.82	
Phl ³	5.83	3.19, 3.08	-	7.24-7.35	7.24-7.35*	7.24-7.35*	-	
Mle ⁴	5.40	1.69, 1.64	0.86	0.84	-	-	2.90	
Lac ⁵	5.55	1.40	-	-	-	-	-	
Mle ⁶	5.45	1.75, 1.69	1.33	0.95, 0.90	-	-	2.92	
Phl ⁷	5.75	3.17, 3.11	-	7.24-7.35	7.24-7.35*	7.24-7.35*	-	
Mle ⁸	5.24	1.58, 1.51	0.94	0.83, 0.80	-	-	3.01	
Lac ¹⁵	5.37	1.38	-	-	-	-	-	
Mle ²⁶	5.15	1.61	1.39	0.91, 0.87	-	-	2.94	
Phl ³⁷	5.70	3.10	-	7.32	7.24-7.35*	7.24-7.35*	-	
Mle ⁴⁸	5.35	1.66	1.14	0.89, 0.87	-	-	2.95	
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe
Lac ¹	173.1	68.5	15.8	-	-	-	-	-
Mle ²	171.0	57.2	37.6	24.7	20.2, 22.3	-	-	28.6
Phl ³	171.7	71.1	37.2	134.8	129.4	128.4	126.9	-
Mle ⁴	169.3	54.3	36.4	24.1	22.3, 20.0	-	-	29.9
Lac ⁵	172.1	67.1	16.1	-	-	-	-	-
Mle ⁶	169.6	54.1	37.2	24.8	22.5, 19.7	-	-	29.7
Phl ⁷	171.7	70.9	37.5	135.1	129.3	128.3	127.0	-
Mle ⁸	170.7	54.0	35.9	23.8	22.2, 20.3	-	-	30.6
Lac ¹⁵	171.9	68.3	15.4	-	-	-	-	-
Mle ²⁶	n.d.	55.3	36.8	24.6	22.3, 20.1	-	-	31.0
Phl ³⁷	170.7	71.3	37.1	135.5	129.3	128.3	126.8	-
Mle ⁴⁸	170.2	54.3	36.9	24.3	22.2, 20.3	-	-	30.3

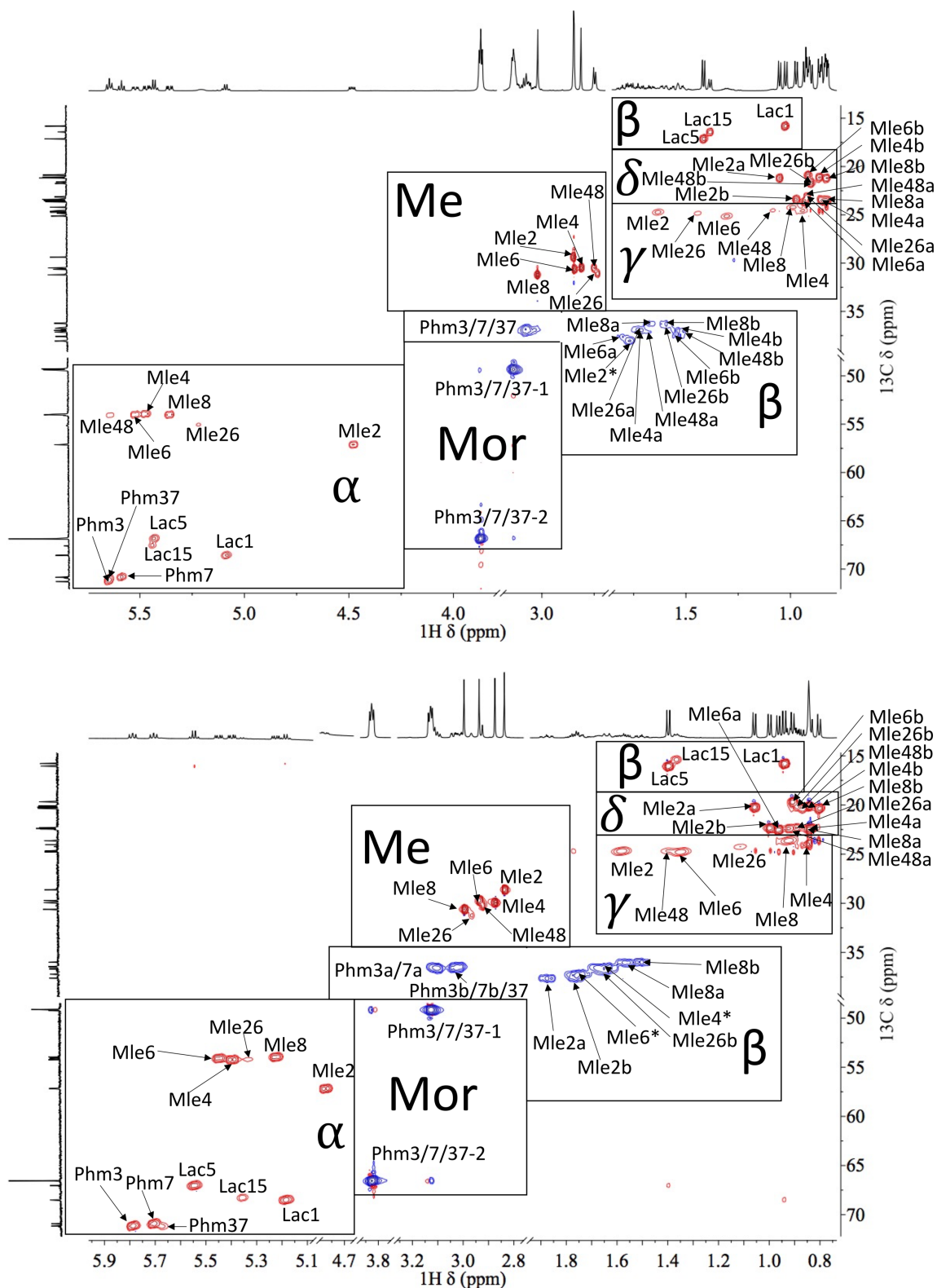


Figure A2.2: ^{13}C -HSQC spectra of emodepside (2) in CDCl_3 (top) and in CD_3OH (bottom). Empty regions are cut out for clarity.

Table A2.3: Assignment of the major (asymmetric) and minor (symmetric) conformation of emodepside (**2**) in CDCl₃ referenced to the residual CHCl₃ shift in the solvent set to 7.28 ppm. * indicates non-assignable signals due to overlap. Mor1 is the C/H next to the oxygen in the morpholine ring, and Mor2 is the C/H next to the nitrogen in the morpholine ring.

	H α	H β	H γ	H δ	H ϵ		HMe	HMor1	HMor2	
Lac ¹	5.09	1.03	-	-	-		-	-	-	
Mle ²	4.49	1.77	1.64	1.05, 0.98	-		2.85	-	-	
Phm ³	5.65	3.08	-	7.13-7.17*	6.84		-	3.12-3.16*	3.87	
Mle ⁴	5.47	1.73, 1.54	0.94	0.85	-		2.82	-	-	
Lac ⁵	5.43	1.42	-	-	-		-	-	-	
Mle ⁶	5.53	1.81, 1.54	1.31	0.92	-		2.85	-	-	
Phm ⁷	5.59	3.07	-	7.13-7.17*	6.85		-	3.12-3.16*	3.87	
Mle ⁸	5.36	1.68, 1.60	1.0	0.83	-		3.02	-	-	
Lac ¹⁵	5.44	1.39	-	-	-		-	-	-	
Mle ²⁶	5.21	1.73, 1.60	1.45	0.92, 0.90	-		2.76	-	-	
Phm ³⁷	5.64	3.04	-	7.13-7.17*	6.83		-	3.11	3.87	
Mle ⁴⁸	5.63	1.69, 1.51	1.09	0.90, 0.86	-		2.74	-	-	
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe	CMor1	CMor2
Lac ¹	171.7	68.6	15.8	-	-	-	-	-	-	-
Mle ²	171.2	57.1	38.1	24.7	21.2, 23.4	-	-	29.4	-	-
Phm ³	170.1	71.3	36.8	126.1	130.3	115.6	150.3-150.4*	-	49.3-49.4*	66.9
Mle ⁴	169.8	54.0	37.0	24.6	21.1, 23.6	-	-	30.5	-	-
Lac ⁵	170.3	66.9	17.1	-	-	-	-	-	-	-
Mle ⁶	169.8	54.0	37.6	25.1	20.9, 23.7	-	-	30.6	-	-
Phm ⁷	170.4	70.9	37.1	126.6	130.3	115.6	150.3-150.4*	-	49.3-49.4*	66.9
Mle ⁸	171.1	54.0	36.2	24.2	21.2, 23.5	-	-	31.2	-	-
Lac ¹⁵	169.9	67.6	16.4	-	-	-	-	-	-	-
Mle ²⁶	170.6	55.2	36.7	24.8	23.1, 21.6	-	-	31.2	-	-
Phm ³⁷	169.6	70.9	36.9	126.7	130.4	115.6	150.3	-	49.4	66.9
Mle ⁴⁸	171.0	54.1	37.1	24.6	21.8, 23.4	-	-	30.5	-	-

Table A2.4: Assignment of the major (asymmetric) and minor (symmetric) conformation of emodepside (2) in CD₃OH referenced to the residual CH₃OH shift in the solvent set to 3.33 ppm. * indicates non-assignable signals due to overlap and n.d. indicates signals that could not be detected. Mor1 is the C/H next to the oxygen in the morpholine ring, and Mor2 is the C/H next to the nitrogen in the morpholine ring.

	H α	H β	H γ	H δ	H ϵ		HMe	HMor1	HMor2	
Lac ¹	5.18	0.94	-	-	-		-	-	-	
Mle ²	4.78	1.88, 1.76	1.58	1.06, 1.00	-		2.84	-	-	
Phm ³	5.79	3.11, 3.02	-	7.19	6.92		-	3.12-3.14*	3.83	
Mle ⁴	5.40	1.65	0.85	0.84	-		2.87	-	-	
Lac ⁵	5.55	1.40	-	-	-		-	-	-	
Mle ⁶	5.45	1.75	1.35	0.96, 0.91	-		2.94	-	-	
Phm ⁷	5.71	3.11, 3.03	-	7.19	6.92		-	3.12-3.14*	3.83	
Mle ⁸	5.23	1.57, 1.50	0.92	0.83, 0.80	-		3.00	-	-	
Lac ¹⁵	5.36	1.37	-	-	-		-	-	-	
Mle ²⁶	5.33	1.65	1.11	0.89, 0.87	-		2.97	-	-	
Phm ³⁷	5.68	3.01	-	7.18	6.92		-	3.12	3.83	
Mle ⁴⁸	5.11	1.65	1.40	0.92, 0.89	-		2.92	-	-	
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe	CMor1	CMor2
Lac ¹	173.1	68.5	15.8	-	-	-	-	-	-	-
Mle ²	170.9	57.2	37.6	24.7	20.2, 22.3	-	-	28.6	-	-
Phm ³	171.8	71.2	36.5	125.5	130.0	115.5	150.6-150.7*	-	49.1-49.2*	66.6
Mle ⁴	169.4	54.2	36.5	24.0	22.5, 20.1	-	-	30.0	-	-
Lac ⁵	172.1	67.1	16.1	-	-	-	-	-	-	-
Mle ⁶	169.6	54.1	37.2	24.8	22.5, 19.7	-	-	29.7	-	-
Phm ⁷	171.8	70.9	36.7	125.8	130.0	115.5	150.6-150.7*	-	49.1-49.2*	66.6
Mle ⁸	170.7	54.0	36.0	23.6	22.4, 20.3	-	-	30.6	-	-
Lac ¹⁵	171.9	68.3	15.4	-	-	-	-	-	-	-
Mle ²⁶	n.d.	54.2	36.8	24.3	22.3, 20.4	-	-	31.3	-	-
Phm ³⁷	170.8	71.3	36.4	126.4	130.0	115.6	150.5	-	49.3	66.6
Mle ⁴⁸	170.4	n.d.	n.d.	24.7	22.4, 20.1	-	-	30.3	-	-

Table A2.5: Assignment of the major conformation (asymmetric) as well as assignment of H α chemical shifts for the minor conformation (symmetric) denoted with ' of the bis-aza PF1022A analog (**8**) in CDCl₃ referenced to the residual CHCl₃ shift in the solvent set to 7.28 ppm.

	H α	H β	H γ	H δ	H ϵ	H ζ	HMe	
Lac ¹	4.96	1.08	-	-	-	-	-	
MIn ²	-	3.67, 3.35	2.07	0.98, 0.86	-	-	3.13	
PhI ³	5.42	3.14	-	7.25	7.34	7.26	-	
MIn ⁴	-	3.69, 2.93	1.54	0.89, 0.89	-	-	3.16	
Lac ⁵	5.57	1.39	-	-	-	-	-	
Mle ⁶	5.56	1.72, 1.54	1.32	0.92	-	-	2.87	
PhI ⁷	5.59	3.18, 3.13	-	7.26	7.31	7.29	-	
Mle ⁸	5.38	1.67, 1.54	0.94	0.83, 0.82	-	-	2.96	
PhI ^{3'}	5.81							
Lac ^{5'}	5.28							
Mle ^{6'}	3.68							
PhI ^{7'}	5.99							
Mle ^{8'}	5.01							
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe
Lac ¹	173.4	69.6	16.0	-	-	-	-	-
MIn ²	154.7	-	58.6	27.2	20.5, 19.8	-	-	35.4
PhI ³	168.7	72.6	37.7	134.9	129.4	128.7	127.1	-
MIn ⁴	154.0	-	57.2	27.3	20.2, 19.8	-	-	38.9
Lac ⁵	171.0	67.6	17.7	-	-	-	-	-
Mle ⁶	170.2	53.7	37.8	25.2	23.6, 21.0	-	-	30.5
PhI ⁷	170.2	70.7	38.0	135.4	129.5	128.5	127.4	-
Mle ⁸	171.5	54.1	36.3	24.3	21.2, 23.3	-	-	31.1

Table A2.6: Assignment of the major conformation of the bis-aza PF1022A analog (**8**) in CD₃OH referenced to the residual CH₃OH shift in the solvent set to 3.33 ppm.

	H α	H β	H γ	H δ	H ϵ	H ζ	HMe	
Lac ¹	4.98	1.06	-	-	-	-	-	
MIn ²	-	3.69, 3.57	2.05	0.97, 0.87	-	-	3.11	
PhI ³	5.60	3.16	-	7.34	7.34	7.29	-	
MIn ⁴	-	3.64, 2.98	1.55	0.89, 0.88	-	-	3.23	
Lac ⁵	5.59	1.38	-	-	-	-	-	
Mle ⁶	5.46	1.72	1.36	0.95, 0.92	-	-	2.96	
PhI ⁷	5.72	3.19, 3.10	-	7.32	7.32	7.28	-	
Mle ⁸	5.22	1.58, 1.49	0.90	0.82, 0.81	-	-	2.98	
	C	C α	C β	C γ	C δ	C ϵ	C ζ	CMe
Lac ¹	174.2	69.3	16.1	-	-	-	-	-
MIn ²	154.7	-	58.2	26.9	19.4, 18.8	-	-	34.7
PhI ³	169.7	72.6	37.1	134.8	129.3	128.4	127.1	-
MIn ⁴	153.9	-	57.0	27.0	18.9, 19.2	-	-	38.1
Lac ⁵	172.3	67.9	16.6	-	-	-	-	-
Mle ⁶	169.9	53.9	37.4	24.9	22.4, 19.8	-	-	29.7
PhI ⁷	171.9	70.9	37.5	135.1	129.4	128.3	127.0	-
Mle ⁸	171.2	54.1	36.1	23.8	22.2, 20.3	-	-	30.5

Table A2.7: ¹⁵N chemical shifts of **1**, **2** and **8** in CDCl₃ and CD₃OH. n.d. indicates signals that could not be detected.

	Mle ² N	Mle ² N α	Mle ⁴ N	Mle ⁴ N α	Mle ⁶ N	Mle ⁸ N	Mle ²⁶ N	Mle ⁴⁸ N	Mor N
PF1022A CDCl ₃	107.4	-	110.6	-	104.8	112.6	110.2	111.4	-
PF1022A CD ₃ OH	110.5	-	114.2	-	107.2	115.6	n.d.	n.d.	-
Emodepside CDCl ₃	107.4	-	110.7	-	104.9	112.7	110.3	111.3	103.1
Emodepside CD ₃ OH	110.6	-	114.6	-	107.3	115.9	n.d.	n.d.	103.0
bis-aza PF1022A analog CDCl ₃	95.2	122.5	94.5	123.9	104.1	112.7	n.d.	n.d.	-
bis-aza PF1022A analog CD ₃ OH	96.3	122.5	96.8	123.6	106.5	115.2	n.d.	n.d.	-

Identification of Additional Conformers Based on Exchange, Exemplified for Compound 1

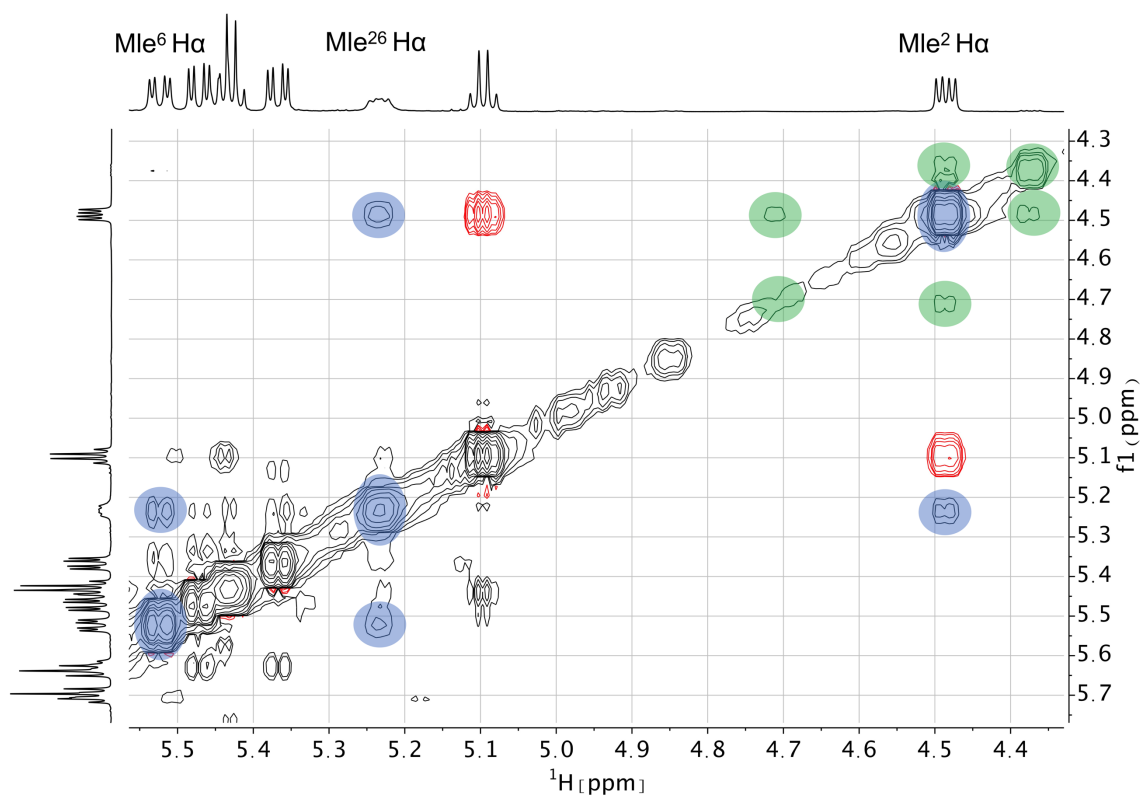


Figure A2.4: Detail of ROESY spectrum showing the exchange between Mle² Ha and Mle²⁶ Ha as well as Mle⁶ Ha and Mle²⁶ Ha of PF1022A (**1**) in chloroform (blue). ROE-peaks (red) have opposite phase relative to the diagonal whereas EXSY peaks have the same phase (black). Based on additional exchange peaks to Mle² Ha, at least two additional conformers with very low intensity could be identified.

Plots for Fitting T_2 Relaxation Times of PF1022A (1) in Chloroform

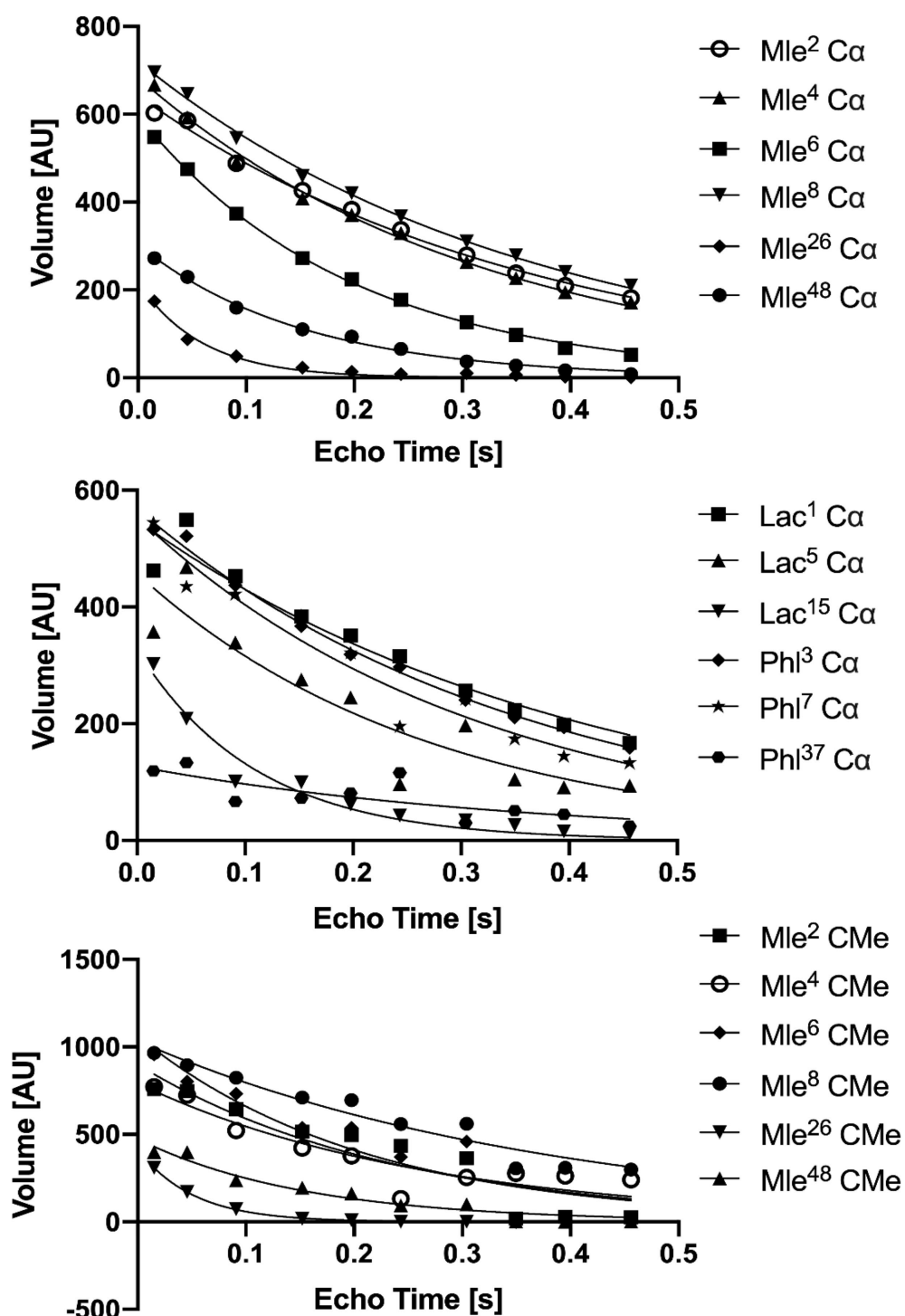


Figure A2.5: Volumes extracted from the CPMG ^{13}C -HSQC spectra plotted against the echo time to obtain the corresponding T_2 relaxation times for PF1022A (1) in chloroform. The plots were created with Prism.

Extracted Volumes and M_0 Values Used for Calculation of Site-to-Site Exchange Rates for Compounds 1 (A), 2 (B), 8 (C) and 11 (D) in Chloroform

$$A = \begin{pmatrix} \frac{77.9}{1} & \frac{1.52}{1.5} & \frac{1.55}{1.5} \\ \frac{1.59}{1} & \frac{236}{236} & \frac{0}{200} \\ \frac{1.75}{1} & \frac{0}{1.5} & \frac{200}{1.5} \end{pmatrix} \quad B = \begin{pmatrix} \frac{65.8}{1} & \frac{0.93}{1.8} & \frac{1.18}{1.8} \\ \frac{1.00}{1} & \frac{232}{232} & \frac{0}{216} \\ \frac{1.32}{1} & \frac{0}{1.8} & \frac{216}{1.8} \end{pmatrix}$$

$$C = \begin{pmatrix} \frac{12.9}{1} & \frac{0.50}{9.8} \\ \frac{0.46}{1} & \frac{387}{9.8} \end{pmatrix} \quad D = \begin{pmatrix} \frac{95}{1} & \frac{1.58}{1.5} & \frac{1.7}{1.5} \\ \frac{2.41}{1} & \frac{235}{235} & \frac{0}{218} \\ \frac{1.82}{1} & \frac{0}{1.5} & \frac{218}{1.5} \end{pmatrix}$$

MSM Building

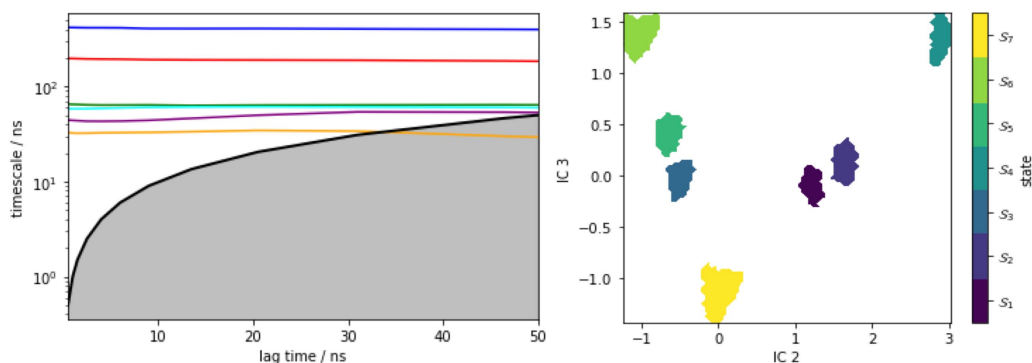


Figure A2.6: (Left): Implied time scales for the MSM of the symmetric subset of PF1022A (**1**) in chloroform. Six slow transitions were observed. (Right): The seven conformational states of the symmetric subset plotted with the second and third TICA element. The first TICA element describes mainly the dihedral undergoing cis-trans isomerization and is thus not relevant for the symmetric subset.

Table A2.8: Stationary distribution and corresponding free energies of the seven conformational states in the MSM of the symmetric subset of PF1022A (**1**) in chloroform.

State	Stationary distribution [%]	Estimated free energy [kT]
1	3.1	3.5
2	6.5	2.7
3	7.4	2.6
4	11.3	2.2
5	11.9	2.1
6	16.5	1.8
7	43.2	0.8

Titration of Emodepside with KSCN. General Procedure for the Fitting of Titration Data

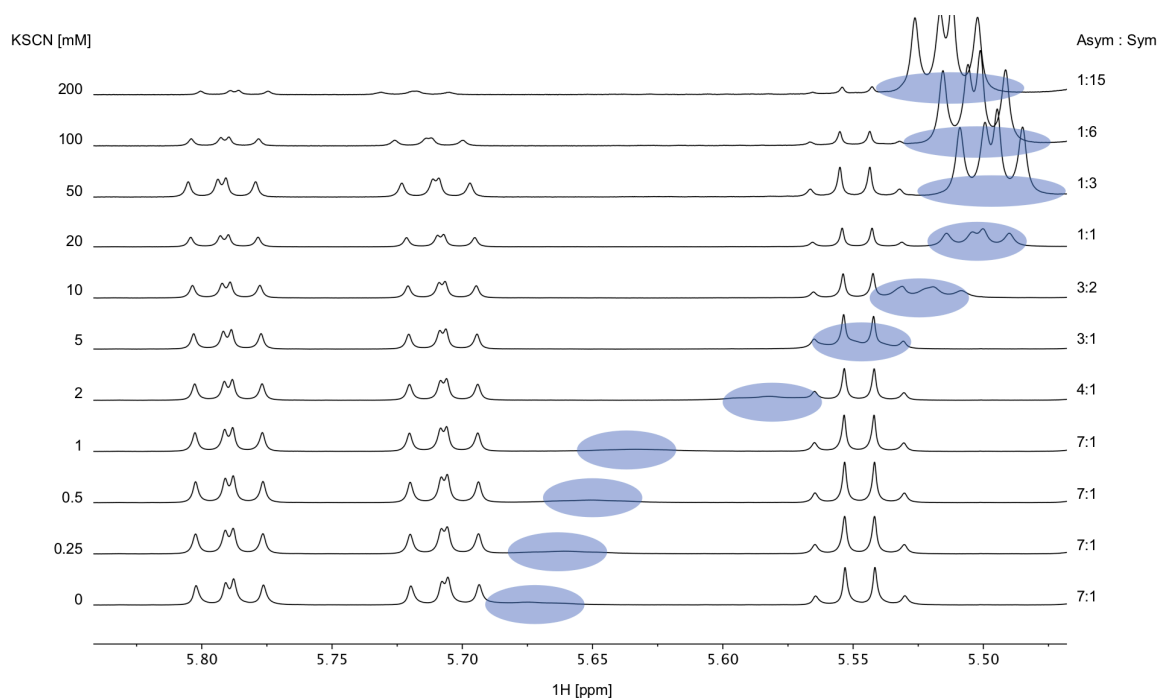


Figure A2.7: $H\alpha$ region of 1H NMR spectra of 5 mM emodepside (**2**) in CD_3OH at different KSCN concentrations. Chemical shift changes were observed for the symmetric conformation, best seen for the signal of the $H\alpha$ proton in residue Phm³⁷ (blue labels). In addition, a change in the ratio between the symmetric and asymmetric conformation is observed. Also the asymmetric conformation shows small chemical shift changes at high salt concentrations.

The concentration of the symmetric conformation in dependence of the salt concentration was fitted with the following equation (damped logistic growth function):

$$[P_{free}] = [P_{all}] - ([P_{all}] - [P_0]) * e^{\frac{-a*[S]}{[S]+b}}$$

where $[P_{free}]$ is the concentration of the free symmetric peptide, $[P_{all}]$ the total peptide concentration, $[P_0]$ the symmetric peptide concentration without salt, $[S]$ the salt concentration, and a and b are the fitting parameters.

The chemical shift change in dependence of the salt concentration was fitted using the following equations assuming a fast equilibrium between the free symmetric species, the 1:1 complex and a 2:1 complex:

$$\delta_{obs} = \delta_P - \frac{(\delta - \delta_{PS}) * [S_{tot}] * K_1 * [P_{free}] + 2 * (\delta_P - \delta_{P_2S}) * [S_{tot}] * K_1 * K_2 * [P_{free}]^2}{P_0 * (1 + K_1 * [P_{free}] + K_1 * K_2 * [P_{free}]^2)}$$

$$0 = K_1 * K_2 * [P_{free}]^3 + K_1 * (2 * K_2 * [S_{tot}] - K_2 * [P_{tot}] + 1) * [P_{free}]^2 + (K_1 * ([S_{tot}] - [P_{tot}]) + 1) * [P_{free}] - [P_{tot}]$$

with $K_1 = \frac{[PS]}{[P_{free}][S]}$ and $K_2 = \frac{[P_2S]}{[PS][S]}$

where δ_{obs} corresponds to the observed chemical shift, δ_P is the chemical shift of the symmetric free peptide, δ_{PS} is the chemical shift of the 1:1 symmetric peptide-cation complex (PS), δ_{P_2S} the shift of the 2:1 symmetric peptide-cation complex (P_2S), $[P_{tot}]$ the total concentration of the peptide (all symmetric species), directly determined from the integral of the signal of the symmetric species in 1H NMR spectrum, $[S_{tot}]$ is the total salt concentration, $[P_{free}]$ the free symmetric peptide concentration, and K_1 and K_2 are the equilibrium constants for the 1:1 and the 2:1 complexes. Fitting was done with R.

Experimental Results for mono-iodine analog (11):

Table A2.9: Ratio between asymmetric and symmetric conformer in CD_3OH and $CDCl_3$ for compounds **11**.

Compound	Conformer ration in CD_3OH (asymmetric : symmetric)	Conformer ration in $CDCl_3$ (asymmetric : symmetric)
Mono-iodo analog (11)	5 : 1	3 : 1

Table A2.10: Exchange rates between asymmetric and symmetric conformers of **11** in $CDCl_3$ determined from EASY-ROESY experiment with mixing time of 100 ms.

Compound	k_1 [s^{-1}]	k_2 [s^{-1}]	K_{ex} [s^{-1}]
Mono-iodo analog (11)	0.17	0.09	0.26

Table A2.11: Change in ratio between asymmetric and symmetric conformers of **11** without salt and after addition of a 25-fold excess of $CsSCN$ in CD_3OH .

Compound	Mono-iodo analog (11)
$CsSCN$	5 : 1 to 1 : 80

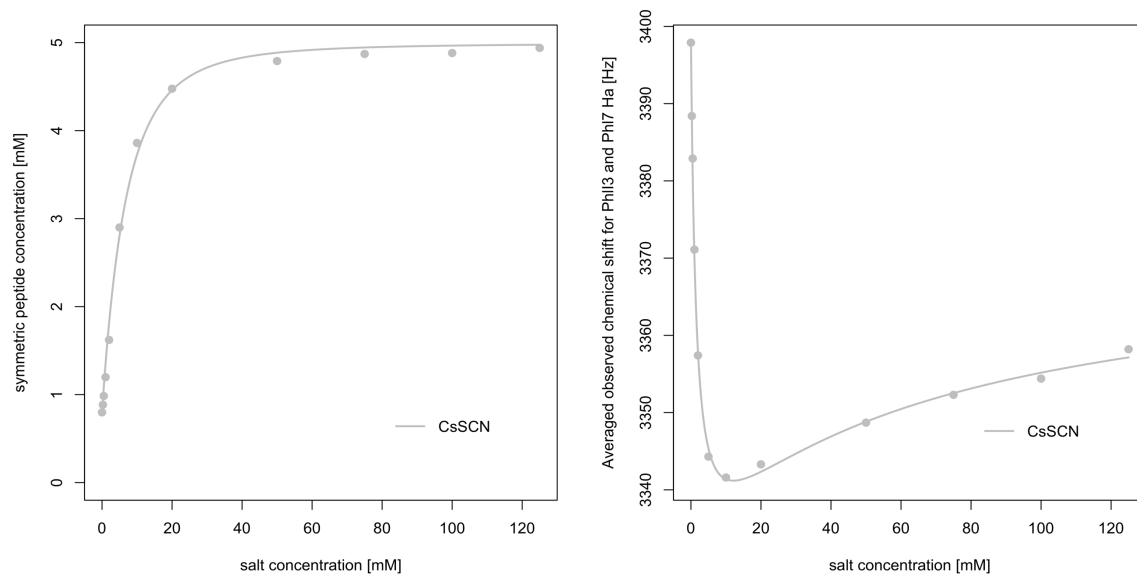


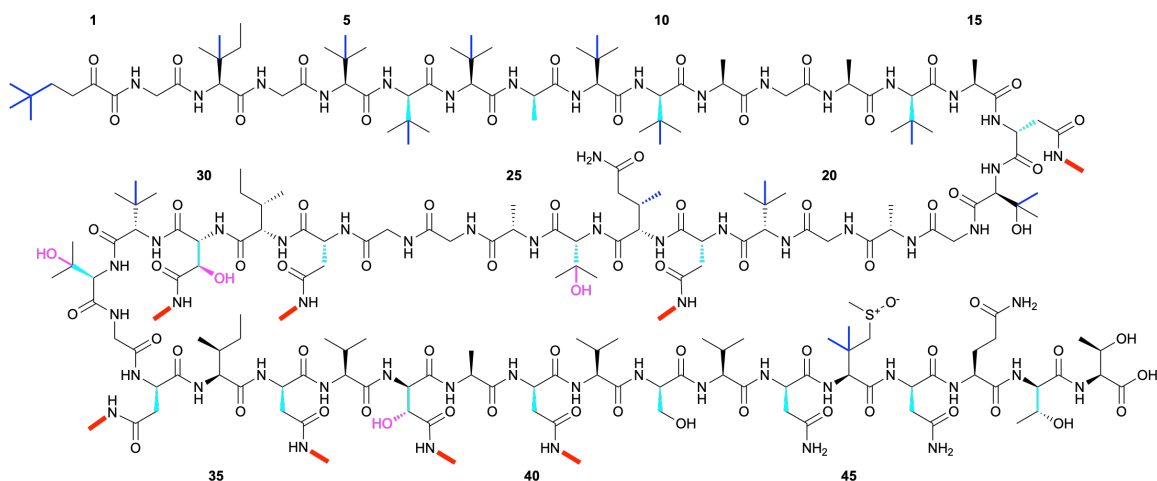
Figure A2.8: Titration of 5 mM mono-iodine analog (**11**) with CsSCN in CD₃OH while the total volume was kept constant. (Left): Change in concentration of the symmetric conformation upon the addition of CsSCN. The data points were fitted with a damped logistic growth function. (Right): Change of the chemical shift of the averaged signal of PhiI³ and PhiI⁷ Ha proton of the symmetric conformation as a function of the salt concentration.

3 Transferring the Stabilizing Effect of Side-Chain N-Methylations Observed in Polythienoamide B

The special sequence of alternating D-/L-amino acids allows the natural product polythienoamide B (pTB) to adopt a β -helix, i.e., a β -sheet wrapped into a helix. Computational studies of pTB showed that its side-chain N-methylations of asparagine residues (Asm) increase the stability of the $\beta^{6.3}$ -helical conformation in polar environments by formation of an “exoskeleton-like” network of intramolecular hydrogen bonds. The wide radius of the $\beta^{6.3}$ -helix is necessary for pTB to act as pore upon insertion into membranes, resulting in pTB’s cytotoxic function. Molecular dynamics (MD) simulations revealed that the stabilizing effect of Asm residues may be transferrable to gramicidin A (GramA), which also adopts a $\beta^{6.3}$ -helix inside a membrane but other conformations in polar environments. The Asm variant of GramA was compared to a GramA derivative with asparagine residues instead of Asm residues. For the latter, no stabilizing effect was observed in the computational study. In this work, the two GramA derivatives were studied by nuclear magnetic resonance (NMR) spectroscopy. Experimentally, the stabilizing effect could not be observed. Structure calculations based on NOE-derived distances from methanol/water samples showed no dominant secondary structure for the peptides. ^{23}Na spectra were recorded for micelle samples and also in this case no channel formation was observed. A possible reason for the discrepancy between simulation and experiment could be that the MD simulations started from the already properly folded $\beta^{6.3}$ -helix conformation and remained in this local minimum. In contrast, the $\beta^{6.3}$ -helical structure competes with other possible conformations in experiment. As an alternative test system, the Asn/Asm effect was studied with partially N-methylated cyclic octapeptides, which can potentially form (meta)stable dimers. Preliminary NMR results suggest that the dimer can indeed be stabilized through the intramolecular hydrogen bonds of the Asm side-chains. This is in agreement with Markov state models built from extensive MD simulations for the two peptides. More experiments are needed for a clear conclusion.

3.1 Introduction

Polythienoamide B (pTB), a heavily posttranslationally altered peptide (Scheme 3.1), shows high cytotoxicity in the picomolar range. Its toxicity is associated with its capability of forming a $\beta^{6.3}$ -helix.^{175,176} To form such a helix, an alternation between D- and L-amino acids is necessary. The helix of pTB is large enough to span the entire cell membrane and wide enough to allow permeation of water and ions through the membrane (Figure 3.1 right).¹⁷⁶ In case of pTB, the $\beta^{6.3}$ -helix is extraordinary stable and could be detected also in a 1:1 methanol/chloroform mixture, which is a relatively polar environment.¹⁷⁷ Extensive molecular dynamics (MD) simulations have shown that this stability is dependent on an “exoskeleton-like” network of hydrogen bonds between side-chain N-methylated D-asparagine residues (Asm) located at positions i and $i + 6$ in the sequence (Figure 3.1 left).¹⁷⁸ MD simulations without those methyl groups (i.e., normal D-asparagine residues (Asn)) resulted in unfolding of the $\beta^{6.3}$ -helix in water, showing the importance of these posttranslational modifications for the function of pTB.¹⁷⁸ Renevey and Riniker proposed that the methyl groups lead to a decreased preference of hydrogen bond formation with the solvent, which in turn results in favorable formation of intramolecular hydrogen bonds.¹⁷⁸ No experimental data are available for the Asn variant of pTB to confirm the hypothesis. However, an indirect proof of the importance of the Asm modifications is provided by the discovery that NX₅N presents a privileged motif in bacteria.¹⁷⁹



Scheme 3.1: Chemical structure of polythienoamide B (pTB). The numerous modifications are color coded. Blue: C-methylations, Cyan: epimerizations, Red: N-methylations, Pink: hydroxylations.

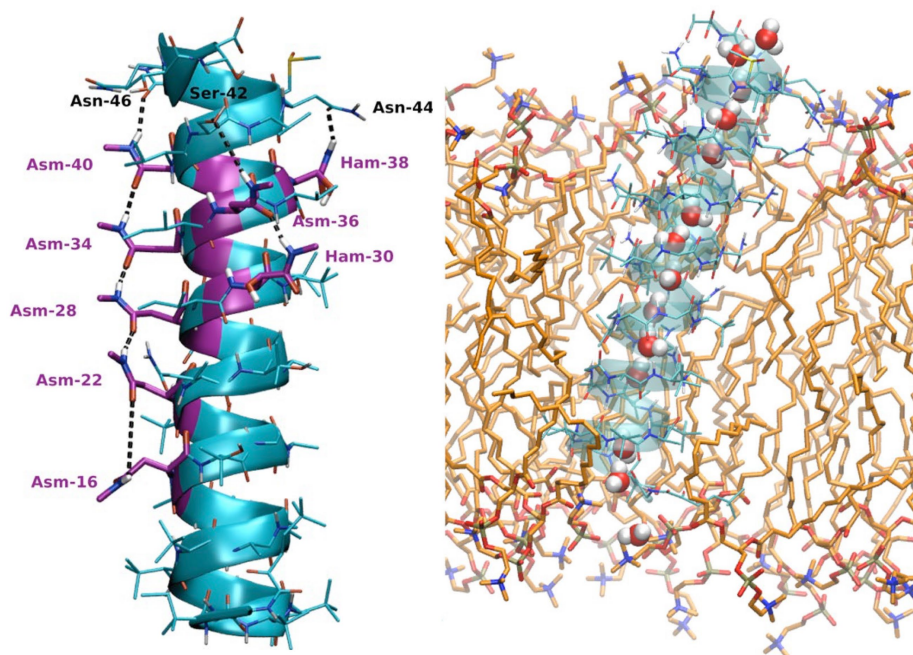
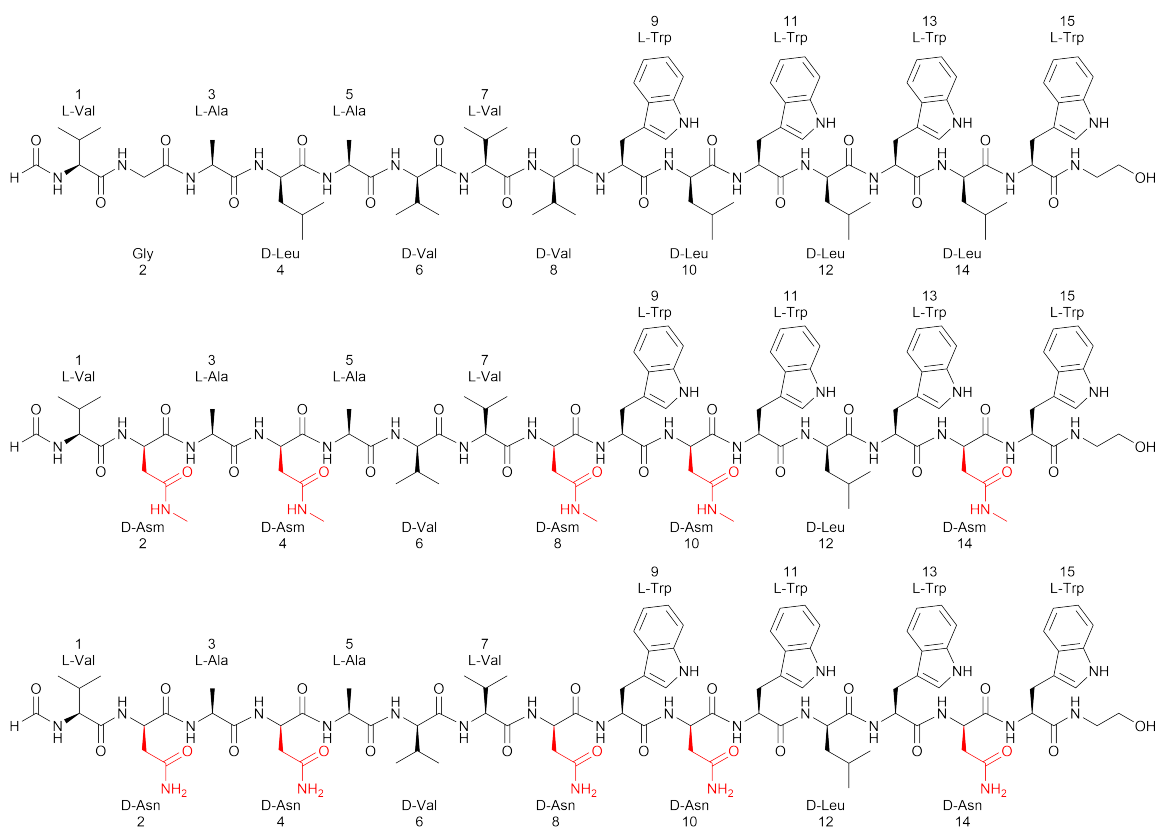


Figure 3.1: Left: Side-chain hydrogen bond network that stabilizes the $\beta^{6.3}$ -helix of pTB. Hydrogen bonds are shown with dotted black lines and side-chain N-methylated residues are highlighted in purple. Right: Snapshot of an MD simulation of pTB in a POPC membrane in water showing the pore forming capability of pTB. Adapted with permission of Springer Nature from Ref. 178.

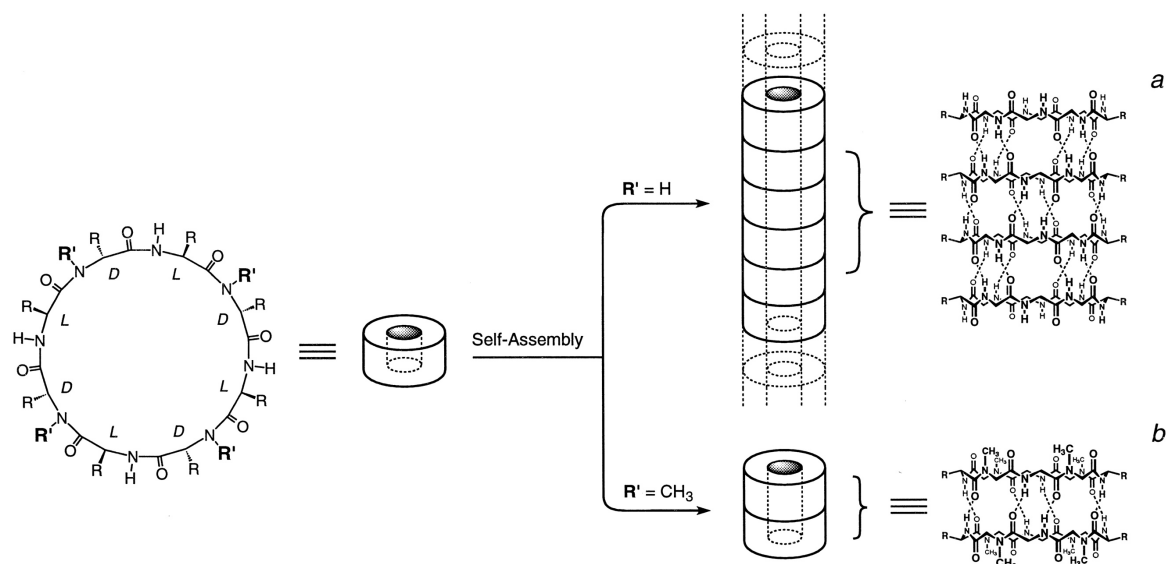
Another $\beta^{6.3}$ -helix forming peptide is the antibiotic gramicidin A (GramA) (Scheme 3.2, top).^{180,181} Wild-type GramA forms head-to-head dimers in cell membranes and micelles.¹⁸² The dimer acts also as membrane pore similar to pTB.¹⁸³ The $\beta^{6.3}$ -helix of GramA is only stable in apolar environments. In a polar medium, GramA adopts various types of conformations, strongly dependent on the experimental conditions.^{184–188} MD simulations starting from the folded $\beta^{6.3}$ -helical structure of a GramA variant, where selected residues were replaced with Asm (Scheme 3.2, middle) (GramA-Asm), suggest transferability of the stabilizing effect by the intramolecular hydrogen bonds of the Asm residues, as observed for pTB in polar environments.¹⁸⁹ The stability of the Asm variant of GramA in water as well as in a 1:1 water/methanol mixture was increased compared to the wild-type variant.¹⁸⁹ For comparison, also a variant with Asn instead of Asm (Scheme 3.2, bottom) (GramA-Asn) was simulated, which showed lower stability.¹⁸⁹



Scheme 3.2: Chemical structures of gramicidin A (GramA) (top), the Asm variant (GramA-Asm) (middle) and the Asn variant (GramA-Asn) (bottom) of GramA. Changes between GramA and the mutated variants are highlighted in red.

Along with the above-mentioned linear peptides, a second test system was devised consisting of cyclic octapeptides with the alternating D-/L-amino acid motif (cyclo[(D-aa-L-aa)₄]). Variants of these cyclic octapeptides were reported in the literature to self-assemble into nanotubes (Scheme 3.3a).¹⁹⁰ It was observed that the exchange of selected amino acids with various building blocks does not disturb the nanotube formation.^{191,192} N-methylation of the D-residues results in stable dimers with β -sheet character since the nanotube formation is prevented (Scheme 3.3b).¹⁹³ Due to the high tolerance of these dimers to changes in the amino acid sequence, these cyclic peptides could be a suitable alternative model system to investigate the stabilizing effect of Asm residues. As parent compound, cyclo[(L-Phe-D-MeN-D-Ala)₄] seems to be ideal.^{193,194} The influence of Asn/Asm can be tested by replacing one to four of the L-Phe residues with L-Asn/Asm. The modified peptides should still be able to form stable dimers and their dissociation constants can be readily determined by NMR.

If the increased stability by the presence of strategically placed Asm residues is observable experimentally, this would enrich the available tools in rational design of tertiary structures of peptides or proteins.



Scheme 3.3: Self-assembling cyclic peptides with a sequence of alternating D- and L-amino acids. **a**: No methylation of the backbone amides leads to formation of nanotubes stacked through antiparallel sheets. **b**: N-Methylation of the D-amino acids leads to formation of stable dimeric nanocylinders. Reprinted with permission of the American Chemical Society from Ref. 193.

In this study, the two variants of GramA, GramA-Asm and GramA-Asn, are investigated with nuclear magnetic resonance (NMR) spectroscopy to examine whether the increased stability of the GramA-Asm variant observed in MD simulations is also detectable in experiment. Further, extensive MD simulations of the cyclic octapeptides cyclo[L-Asm/Asn-D^{Me}N-D-Ala-L-Phe-D^{Me}N-D-Ala-L-Phe-D^{Me}N-D-Ala-L-Phe-D^{Me}N-D-Ala-] in methanol are performed to investigate whether an effect of the side-chain N-methylation is observable *in silico*. Preliminary NMR data obtained for the cyclic peptides are also discussed.

3.2 Results

3.2.1 Gramicidin A and its Asm/Asn Variants

ROESY spectra of the GramA-Asm and the GramA-Asn variants were recorded to obtain NOE-derived distances. Those were used to calculate a bundle of structures for both variants with a simulated annealing approach using the software xplor-NIH.¹⁹⁵ The conformational bundles obtained for GramA-Asm as well as for GramA-Asn do not show a dominant conformation and are far away from the desired $\beta^{6.3}$ -helix. Already the fact that for GramA-Asm only one non-sequential interresidual NOE cross-peak could be identified, is a clear indication of a poorly structured peptide (Figure 3.2, left). Addition of 30 % water seems to stabilize more compact conformations and 13 non-sequential interresidual cross-peaks were detected in the ROESY spectrum (Figure 3.2, middle). Also for the GramA-Asn variant, no dominant secondary structure is observed in methanol, yet the conformational bundle is more uniform compared to GramA-Asm (Figure 3.2, right). For wild-type GramA, also double-stranded helical conformations are observed in polar environments. A simulated annealing procedure using a single copy of the peptide cannot produce such a double-stranded helical conformation. Yet, the low number of interresidual cross-peaks do not support such a structure and it seems that both peptides are rather disordered in a polar environment.

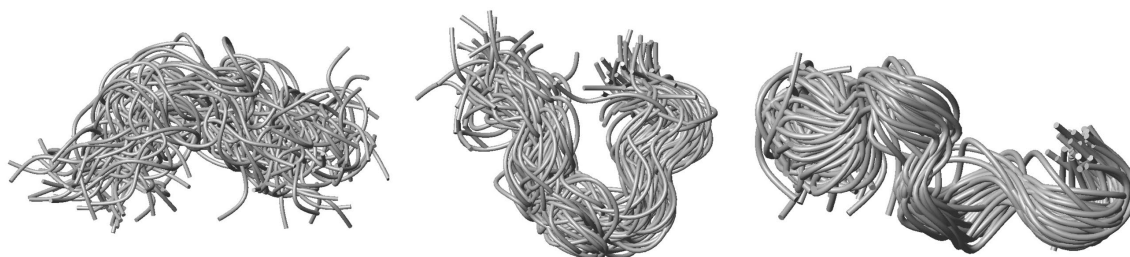


Figure 3.2: Structural bundles obtained from simulated annealing calculations using xplor-NIH¹⁹⁵. (Left): GramA-Asm in CD₃OH at 37 °C. 75 intraresidual, 24 sequential and 1 non-sequential interresidual NOE-derived distances were used for the calculation. No dominant structure was observed. Middle: GramA-Asm in 70 % CD₃OH and 30 %H₂O at 37 °C. 88 intraresidual, 36 sequential and 13 non-sequential interresidual NOE-derived distances were used together with 8 dihedral angle constraints obtained from ³J(HN-H α). In this case, it seems that the C-terminal part of the sequence adopts some kind of secondary structure but clearly not the expected $\beta^{6.3}$ -helix. (Right): GramA-Asn in CD₃OH at room temperature (RT). 33 intraresidual, 53 sequential and 28 non-sequential interresidual NOE-derived distances were used for the simulated annealing procedure. A partially helical structure is observed.

Since GramA forms the $\beta^{6.3}$ -helix conformation only in lipid bilayers, it is possible that the stabilizing effect of the Asm residues is not large enough to maintain such a conformation in polar environments. Therefore, NMR spectra were also recorded using sodium dodecyl sulfate (SDS) micelles. It has been shown for GramA that the formation of the $\beta^{6.3}$ -helix can be easily verified with a ²³Na NMR spectrum.¹⁹⁶ Due to the channel-formation, there is an exchange of Na⁺ ions between different environments (in solution and inside GramA). This leads to severe line-

broadening of the ^{23}Na signal. This could be successfully reproduced for wild-type GramA (Figure 3.3, purple). In contrast, the ^{23}Na signal in presence of GramA-Asm (Figure 3.3, blue) is nearly identical to the signal in the SDS sample without peptide (Figure 3.3, khaki) showing that no $\beta^{6.3}$ -helix is formed.

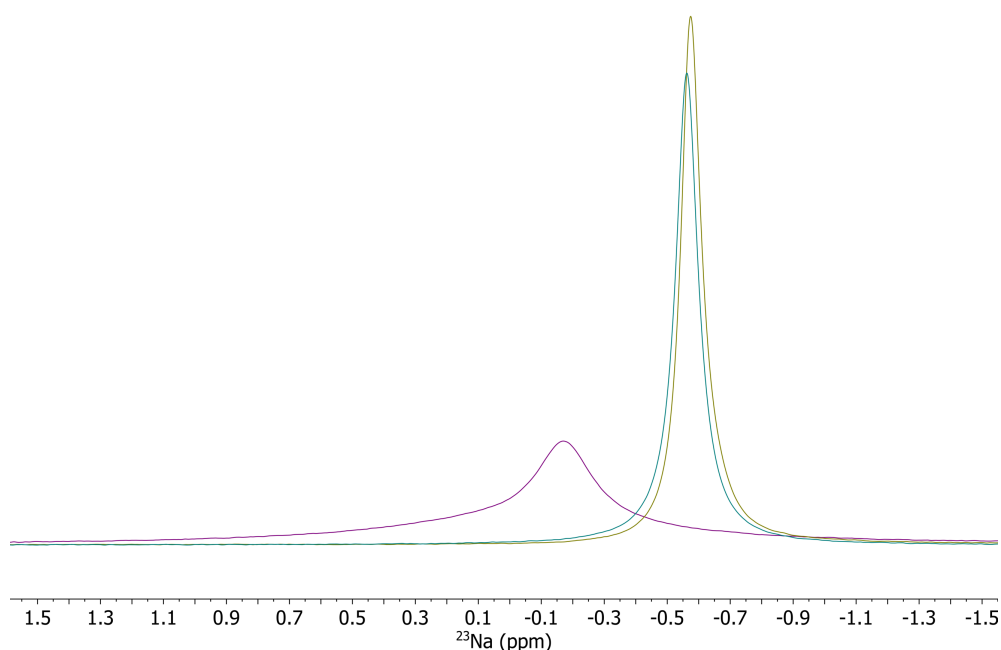
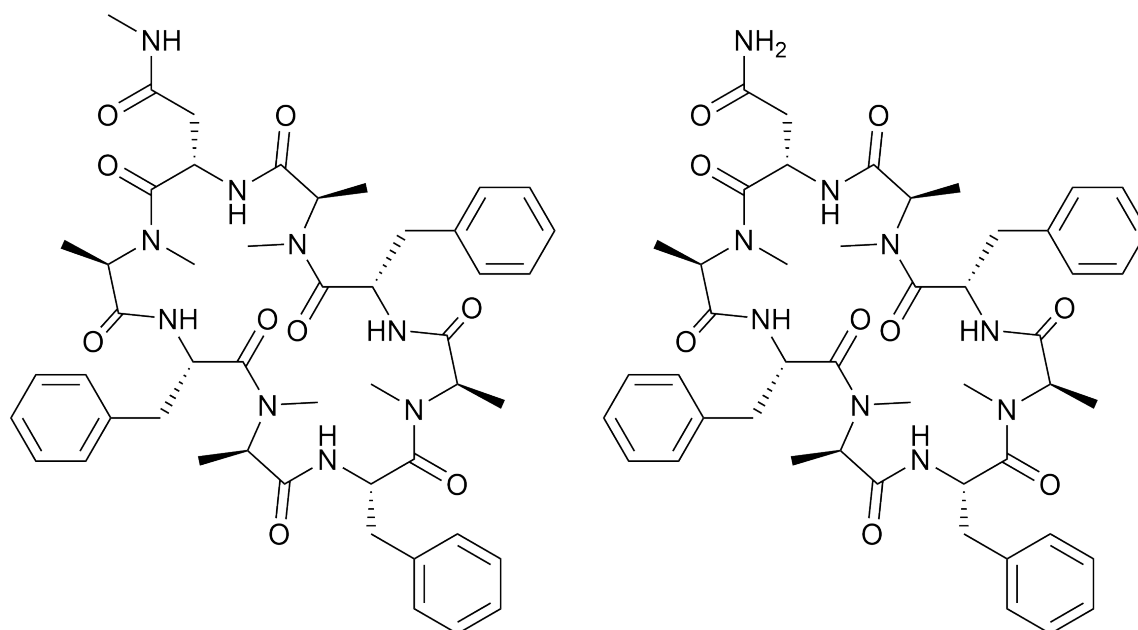


Figure 3.3: Overlay of the ^{23}Na spectra of 250 mM SDS- d_{25} in a water-trifluoroethanol (16:1 molar ratio) solution before (khaki) and after addition of 5 mM GramA (purple) or 5 mM GramA-Asm (blue). All spectra were recorded at 55 °C.

In conclusion, the increased stability of GramA-Asm compared to GramA and GramA-Asn seen in the MD simulations¹⁸⁹ could not be reproduced in experiment. Neither for GramA-Asm nor for GramA-Asn the $\beta^{6.3}$ -helical conformation was observed in methanol or in SDS micelles. A potential explanation for this discrepancy could come from the simulation set-up. The MD simulations started from the folded $\beta^{6.3}$ -helical conformation, since folding studies are still out of reach for standard MD simulations. It may indeed be possible that – once the $\beta^{6.3}$ -helical conformation is adopted – the hydrogen bond network of the Asm residues stabilizes this fold. However, the $\beta^{6.3}$ -helix competes with other conformations in solution, possibly favored by hydrogen bonds between the introduced Asm residues and the backbone. By the mutation of five out of 15 residues in the GramA sequence, the physicochemical properties of GramA-Asm are heavily altered compared to the wild-type.

3.2.2 Cyclic Asm/Asn Variants

Since the mutated variants of GramA did not show the desired structural behavior, we decided to investigate the cyclic octapeptides shown in Scheme 3.3 as an alternative test system. The simplest strategy to test the stabilizing effect of Asm residues is to exchange one L-Phe residue with L-Asm and compare it to the variant where the same residue was replaced by L-Asn (Scheme 3.3 and Scheme 3.4). It is expected that these compounds form stable dimers as the parent peptide does.



Scheme 3.4: Chemical structures of two cyclic octapeptides of the form $\text{cyclo}[L\text{-Phe-}^{\text{Me}}\text{N-D-Ala}]_4$ for which one of the L-Phe residues was replaced by L-Asm (Asm-1) (left) or L-Asn (Asn-1) (right).

Extensive MD simulations of the two peptides in methanol were performed using the GROMOS simulation package¹³³ and the GROMOS 54A7 united-atom force field.⁵⁷ As starting structure for the simulations, the dimeric crystal structure of $\text{cyclo}[L\text{-Phe-}^{\text{Me}}\text{N-D-Ala}]_4$ was used (CCDC code YAXQIX),¹⁹⁴ where a L-Phe residue was replaced by L-Asm or L-Asn in each monomer using Avogadro.¹⁹⁷

Markov state models (MSMs)⁶⁶⁻⁶⁹ are a state-of-the-art technique to analyze the exchange dynamics between different conformers. Common nearest neighbor (CNN) based clustering^{68,134-136} was applied to generate core-set Markov models of Asm-1 and Asn-1, similar to other studies conducted by our group.^{3,147} Since the introduction of the Asm/Asn residues breaks the C_4 symmetry of the parent compound, it is expected that different dimers will be observed during the MD simulations. The MSMs were constructed using inverse N-O backbone distances as input features for the TICA space.¹³⁸ For Asn-1, the constructed MSM shows five states, where state 1 (4.8 %) has the Asn residues on top of each other allowing for hydrogen bond between the

residues (Figure 3.4, top). In states 2 and 4, one of the monomers is rotated by 90°, whereas in state 3, it is rotated by 180°. State 5 corresponds to an intermediate structure, where not all hydrogen bonds between the two monomers are formed. State 6 contains mostly monomeric conformations, including also some partially interacting molecules. Assuming a stabilizing effect of Asm compared to Asn due to favored hydrogen bond formation between the two monomers via this residue, we expect state 1 to have a higher population in the MSM constructed for Asm-1. This is indeed the case, i.e., state 1 for Asm-1 has a population of 9.3 % (+4.5 % compared to Asn-1) (Figure 3.4, bottom). The same metastable states as for Asn-1 are observed. In addition, three additional intermediate states were detected (Figure 3.4). In states 7 and 8, the backbone hydrogen bonds are only partially formed and instead, one Asm side-chain forms a hydrogen bond with the backbone of the other subunit. In state 9, the aromatic ring of a Phe residue is inside the ring of the other peptide and therefore only a fraction of the backbone hydrogen bonds can be formed.

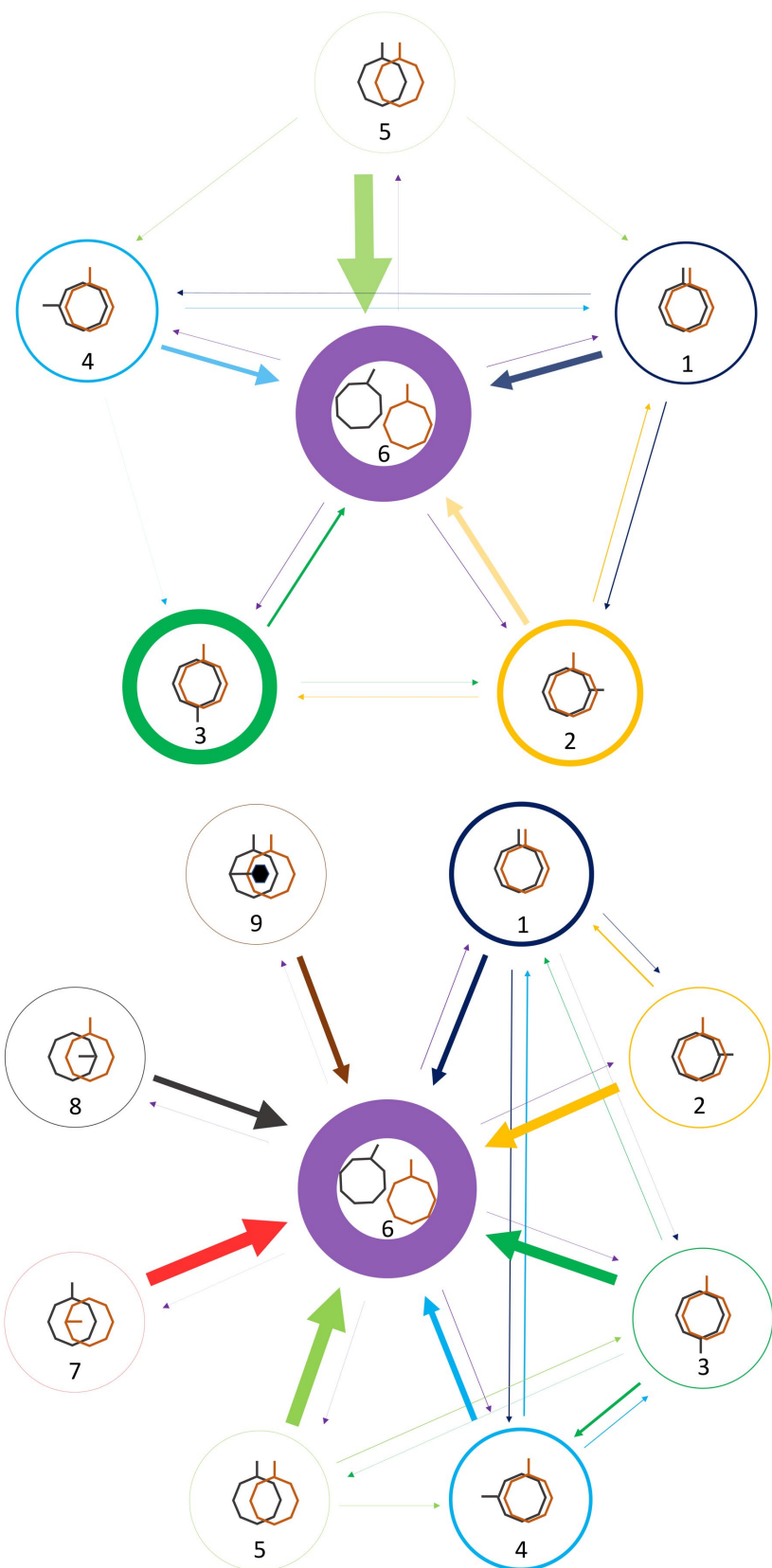


Figure 3.4: MSMs for Asn-1 (top) and Asm-1 (bottom) constructed from MD simulations in methanol. Thickness of the circles corresponds to populations of the states. Each state is represented by a scheme showing how the two dimers interact. The line at the rings indicates the position of the Asn/Asm residues. Dimers are depicted with nearly overlapping octagons, whereas partial interactions are depicted with half overlaying octagons and monomers with separate octagons, respectively.

A first batch of the two peptides was synthesized by Marcel Grogg at a 0.1 mmol scale yielding approximately 0.8 mg Asn-1 and 0.3 mg Asm-1.¹⁹⁸ Preliminary NMR results indicate that both peptides adopt at least two slowly exchanging conformations (Asn-1: 1.0:0.3, Asm-1: 1.0:0.6). More conformations might be present but they could not be identified due to the limited amount of sample. The side-chain amide protons of Asn-1 and Asm-1 can easily be spotted in the ¹⁵N-HSQC spectra as they differ significantly in ¹⁵N chemical shift from the backbone amides (Figure 3.5). Initial findings from exchange spectroscopy (EXSY) suggest that the side-chain amide of the Asm-1 variant does not form hydrogen bonds with the solvent, whereas the Asn-1 variant does (Figure 3.6). These findings need to be confirmed by additional experiments under exactly identical conditions since also differences in pH and concentration can affect amide exchange rates. More material needs to be synthesized to also determine the difference in stability between the two dimer variants.

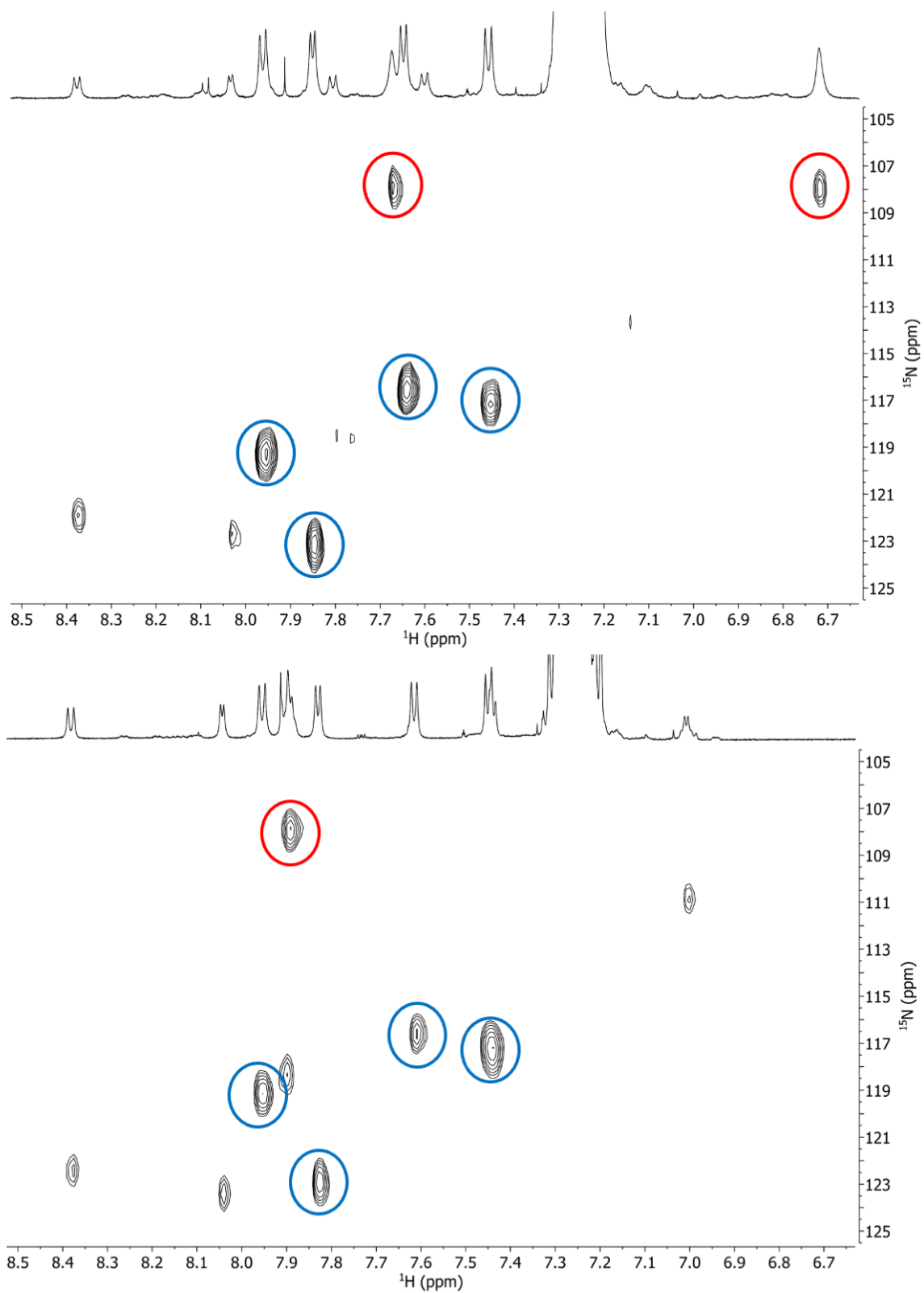


Figure 3.5: ^{15}N -HSQC spectra of Asn-1 (top) and Asm-1 (bottom). Backbone amide signals of the dominant conformation are circled in blue, whereas the side-chain amide signals are circled in red.

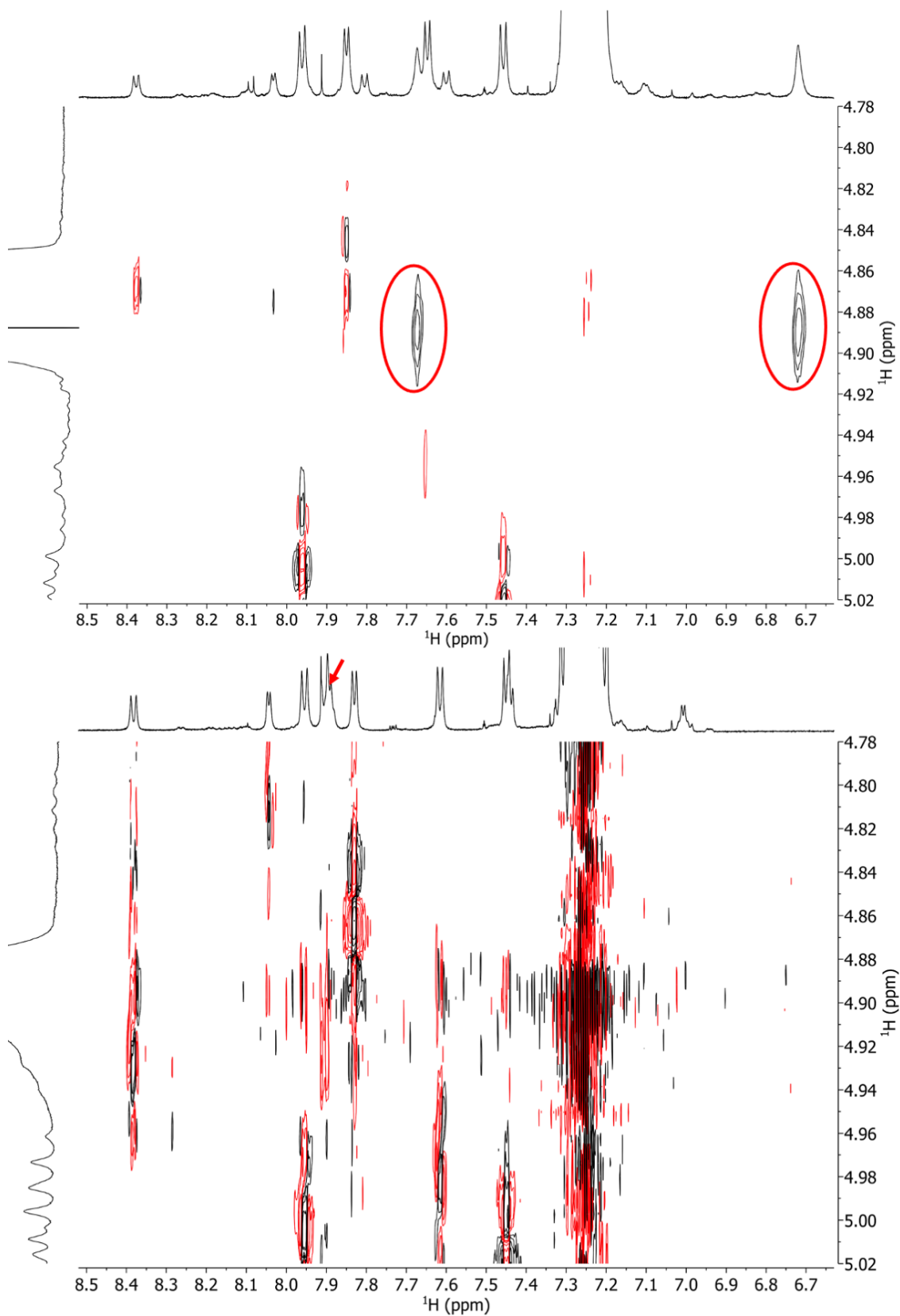


Figure 3.6: Amide region of NOESY spectra showing EXSY peaks with the solvent (red circles Asn-1, top). The red arrow in the Asm-1 spectrum (bottom) indicates the side-chain amide proton resonance. No EXSY peak with the solvent is observed.

3.3 Conclusion

In this work, we investigated whether the stabilizing effect from intramolecular hydrogen bonds of Asm residues, as seen for the $\beta^{6.3}$ -helix of the natural product pTB, can be transferred to other molecular system. The first test system was GramA, because this peptide forms a $\beta^{6.3}$ -helix similar to pTB when inside a membrane environment. Two GramA variants, GramA-Asm and GramA-Asn, were studied both with MD simulations and NMR experiments. While in simulations starting from the $\beta^{6.3}$ -helical conformation the presence of the Asm residues resulted in the expected stabilization, this finding could not be confirmed by the NMR experiments. The bundle of structures calculated based on NOE-derived distances determined in a methanol/water solution does not point to the presence of a single dominant conformation. ^{23}Na spectra revealed that GramA-Asm does not adopt a $\beta^{6.3}$ -helical conformation in SDS micelles, while the wild-type GramA is known to do that. A possible reason for the discrepancy between simulation and experiment could come from the set-up of the MD simulations, i.e., the starting structure of the properly folded $\beta^{6.3}$ -helical conformation is only a local minima and not dominant in the solution ensemble. In experiment, the $\beta^{6.3}$ -helical structure competes with other possible (meta)stable conformations that were not visited in the simulations.

As an alternative model system to assess whether the stabilizing effect of Asm in pTB can generally be employed, two cyclic octapeptides were devised (Asm-1 and Asn-1). These peptides are expected to form dimers with different stability in methanol. The four expected dimer variants were observed for both peptides in MSMs constructed from extensive MD simulations. As hypothesized, the side-chain N-methylation appears to stabilize the dimer arrangement that is able to form an additional hydrogen bond between the monomers (9.3 % for Asm-1 compared to 4.8 % for Asn-1). Preliminary NMR experiments could be performed with a small amount of the peptides. Exchange between the solvent and the side-chain amide was observed experimentally with EXSY for Asn-1, whereas no exchange could be detected for Asm-1. While these preliminary results are encouraging, more experiments are needed for a clear conclusion. Especially dissociation studies could provide more insights. Additional material necessary for these experiments is currently synthesized.

3.4 Method Section

Peptide Synthesis

The methods for the preparation of GramA-Asm, GramA-Asn as well as for the cyclic octapeptides Asm-1 and Asn-1 investigated in this work have been reported in the doctoral thesis of Marcel Grogg.¹⁹⁸

NMR Measurements

If not stated otherwise, all NMR experiments were recorded at 25 °C on a 600 MHz Bruker Avance III HD spectrometer equipped with a N₂-cooled Prodigy triple resonance probe with z-gradients. The time domain in both dimensions of the spectra were doubled by zero filling and the baseline was corrected with a third order polynomial or by applying the Whittaker smoother algorithm.¹⁵⁷ Processing was done with Bruker TopSpin™ version 4.1 (Bruker Biospin AG) or MestReNova 14.2 (Mestrelab Research). Peak assignment and integration was done using NMRFAM-SPARKY.¹⁹⁹

NMR Experiments of GramA-Asm and GramA-Asn

Each EASY-ROESY¹⁵¹ spectrum was recorded with 4096 x 512 data points using presaturation to suppress the CD₃OH signal. The spectral width was set in both dimensions to 12 ppm and the transmitter was set to the position of the CD₃OH signal. EASY-ROESY spectra of GramA-Asm in CD₃OH as well as in CD₃OH/H₂O 7:3 were recorded at 37 °C with a mixing time of 300 ms. The EASY-ROESY spectrum of GramA-Asn in CD₃OH was recorded at RT with a mixing time of 100 ms. Most of the ¹H resonances of the major conformation could be assigned by analysis of standard DQF-COSY, ¹³C-HSQC, ¹³C-HMBC, TOCSY and ¹⁵N-HSQC spectra. The ROESY cross-peak volumes were translated into distance restraints and used in a simulated annealing procedure as described in Ref. 200. The assigned ¹H chemical shifts can be found in the Appendix.

SDS micelle samples were prepared according to the procedure described by Bystrov *et al.*:¹⁹⁶ 20 µl of a 50 mM GramA or GramA-Asm solution in trifluoroethanol-d₃ (TFE-d₃) were added dropwise to 100 µl of a 500 mM SDS-d₂₅ dispersion in D₂O. 80 µl D₂O and 5 µl of a TPS solution (10 mg TPS in 2 ml D₂O) were added to obtain a final concentration of 5 mM for the peptide and a SDS concentration of 250 mM. These experimental conditions correspond to incorporation of not more than two peptide molecules per micelle.¹⁹⁶ For the sample without peptide, pure TFE-d₃ was added. ²³Na NMR experiments were recorded in a Shigemi tube at 55 °C on a 500 MHz Bruker Avance III HD spectrometer equipped with a BBFO probe with z-gradients. In total, 32768 data points were recorded with a spectral width of 60 ppm. The transmitter was set to 0 ppm. A line broadening of 2 Hz was applied.

NMR Experiments of Asm-1 and Asn-1

¹⁵N-HSQC spectra with sensitivity enhancement were recorded with spectral widths of 4 ppm in the direct and 200 ppm in the indirect dimension. Transmitters were set to 4 and 100 ppm, respectively. A total of 1024 x 256 data points were recorded. For the NOESY spectra, the spectral widths were 12 ppm in both dimensions and the transmitter was set to 4.89 ppm. Excitation sculpting was used for solvent suppression.¹⁵² The mixing time was 100 ms and a total of 4096 x 512 data points was recorded.

MD Simulations of Asm-1 and Asn-1 Dimers

MD simulations were carried out using the GROMOS simulation package¹³³ together with the GROMOS 54A7 united-atom force field.⁵⁷ For Asm residues, the same parameters were used as described in Ref. 178. For each system, 11 MD trajectories with a length of 1 μ s were produced under isothermal-isobaric conditions (NPT) using the leap-frog integration scheme with a time step of 2 fs.¹⁶⁵ The temperature was kept at 298 K by weak coupling¹⁶⁶ to two separate temperature baths for the peptide and the solvent with a relaxation time of 0.1 ps. The pressure was kept at 1 atm by weak coupling to a pressure bath with a relaxation time of 0.5 ps and an isothermal compressibility of $4.5 \cdot 10^{-4}$ kJ⁻¹ mol nm³.¹⁶⁶ The SHAKE algorithm¹⁶⁷ was used to constrain bonds with a tolerance of 10^{-4} nm. Translational motion of the center of mass was removed every 1000 steps. A twin range scheme was applied with cutoffs of 0.8 and 1.4 nm for the non-bonded interactions. Electrostatic non-bonded contributions outside the long-range cutoff were considered with the reaction-field method¹⁶⁹ and a dielectric permittivity coefficient for methanol of 27.8¹⁶⁸ was used. The dimeric crystal structure of the parent compound (CCDC code YAXQIX)¹⁹⁴ was used as template. Using Avogadro,¹⁹⁷ one Phe residue was replaced manually by an Asn/Asm residue in each monomer to obtain the dimer starting structures of Asm-1 and Asn-1. The obtained conformations were minimized in vacuum using a steepest-descent algorithm.¹⁷¹ The dimers were solvated in a box of 764 and 773 methanol molecules for Asn-1 and Asm-1, respectively. Next, the solvent was relaxed while the coordinates of the dimer were position restrained with a force constant of $2.5 \cdot 10^4$ kJ mol⁻¹ nm⁻², followed by thermalization to 298 K in five steps of 60 K, where in each step with a total length of 20 ps the force constant was loosened by one order of magnitude. For each dimer, 11 MD simulations of 1 μ s length were started with different initial velocities generated using a Maxwell-Boltzmann distribution.

Markov State Model (MSM) Building

MSMs were built separately for Asn-1 and Asm-1 using the eleven MD simulations in chloroform with lengths of 1 μ s. The Python²⁰¹ package PyEMMA¹³⁷ was used for the MSM construction in a Jupyter Notebook.²⁰² The inverse distances between all backbone oxygen and nitrogen atoms were used as input features to run a time-lagged independent component analysis (TICA)¹³⁸ with a lag time of 25 ns. A hierarchical variant of the common nearest neighbor (CNN) density based clustering⁶⁸ with a similarity of 10, a cutoff distance of 2 and a delta free energy per hierarchical layer of 0.5 was applied.¹³⁶ The MSMs were then constructed with a lag time of 100 ns. Chapman-Kolmogorov test¹³⁵ (Figure 3.7) was used to validate the models with six (Asn-1) and nine (Asm-1) conformational states. During the analysis, functionalities of the matplotlib,²⁰³ mdraj,²⁰⁴ numpy,²⁰⁵ pandas,²⁰⁶ and scipy²⁰⁷ packages were used.

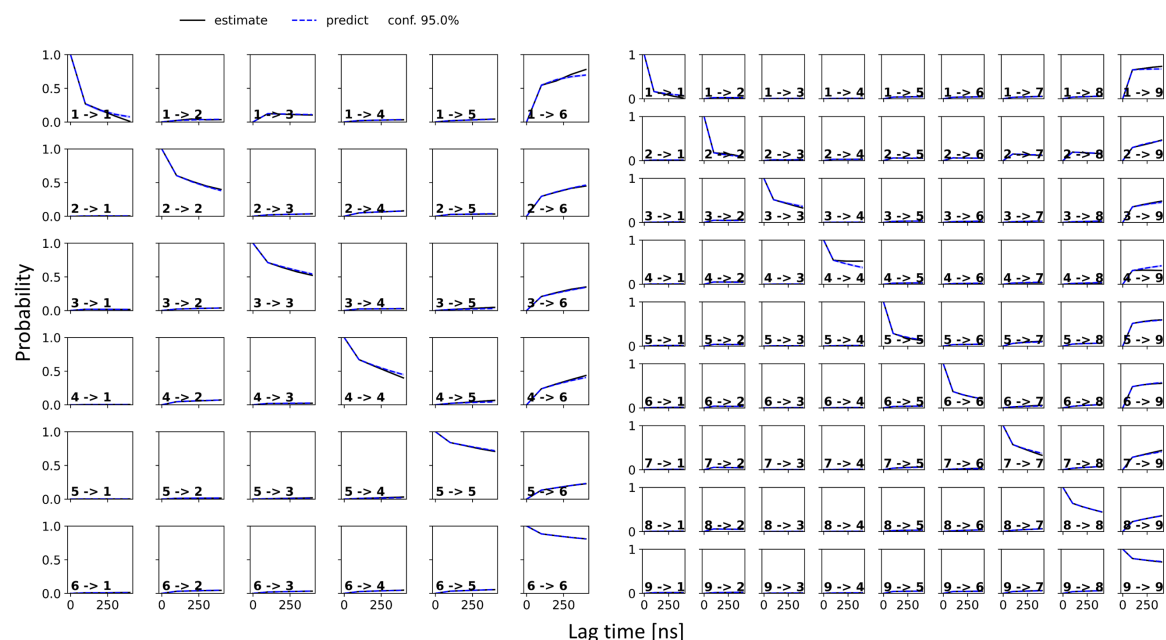


Figure 3.7: Chapman-Kolmogorov tests for Asn-1 and Asm-1 with 6 and 9 states, respectively using a lag time of 100 ns.

3.5 Appendix

Table A3.1: ¹H assignment of the major conformation of GramA-Asn in CD₃OH at RT. * indicates magnetically equivalent protons in methylene and magnetically equivalent methyl groups. † indicates aldehyde proton. Aromatic protons of Trp residues could not be assigned due to signal overlap.

	HN	H α	H β	H γ	H δ	H ϵ	HN _{sc}
Ald	n.i.†	-	-	-	-	-	-
Val ¹	8.39	4.16	2.04	0.98*-	-	-	-
D-Asn ²	8.59	4.79	2.77*	-	-	-	n.i.
Ala ³	8.18	4.22	1.34	-	-	-	-
D-Asn ⁴	8.22	4.65	2.76*	-	-	-	n.i.
Ala ⁵	7.92	4.43	1.37	-	-	-	-
D-Val ⁶	7.95	4.30	2.13	0.91*	-	-	-
Val ⁷	8.18	4.10	2.08	0.95*	-	-	-
D-Asn ⁸	8.41	4.78	2.68, 2.58	-	-	-	n.i.
Trp ⁹	8.19	4.53	3.32, 3.17	-	7.07	n.i.	10.17
D-Asn ¹⁰	8.05	4.73	2.41	-	-	-	n.i.
Trp ¹¹	8.10	4.55	3.28, 3.14	-	7.05	n.i.	10.27
D-Leu ¹²	7.96	4.14	1.23, 1.16	0.75	0.53*	-	-
Trp ¹³	8.17	4.51	3.29, 3.15	-	7.06	n.i.	10.20
D-Asn ¹⁴	7.98	4.61	2.49, 2.24	-	-	-	n.i.
Trp ¹⁵	8.00	4.57	3.37, 3.13	-	7.10	n.i.	10.15
Etam	7.78	3.24*	3.50*	-	-	-	-

Table A3.2: ¹H assignment of the major conformation of GramA-Asm in CD₃OH at 37 °C. * indicates magnetically equivalent protons in methylene and magnetically equivalent methyl groups. + indicates aldehyde proton and n.i. non identifiable protons. Aromatic protons of Trp residues could not be assigned due to signal overlap.

	HN	H α	H β	H γ	H δ	H ϵ	HN _{sc}	HMe
Ald	8.18 ⁺	-	-	-	-	-	-	-
Val ¹	8.37	4.24	2.09	0.97* ⁻	-	-	-	-
D-Asm ²	8.54	4.67	2.80, 2.70	-	-	-	7.91	n.i.
Ala ³	8.13	4.21	1.35	-	-	-	-	-
D-Asm ⁴	8.18	4.63	2.67, 2.39	-	-	-	n.i.	n.i.
Ala ⁵	7.94	4.43	1.37	-	-	-	-	-
D-Val ⁶	7.96	4.29	2.15	0.91*	-	-	-	-
Val ⁷	8.17	4.10	2.06	0.97, 0.93	-	-	-	-
D-Asm ⁸	8.41	4.79	2.67, 2.51	-	-	-	n.i.	n.i.
Trp ⁹	8.18	4.55	3.30, 3.17	-	7.08	7.32	10.23	-
D-Asm ¹⁰	7.95	4.58	n.i.	-	-	-	n.i.	n.i.
Trp ¹¹	7.98	4.57	3.35, 3.13	-	7.04	7.31	10.31	-
D-Leu ¹²	8.03	4.15	1.29, 1.21	0.79	0.56*	-	-	-
Trp ¹³	8.19	4.54	3.28, 3.12	-	7.05	7.31	10.25	-
D-Asm ¹⁴	7.96	4.61	n.i.	-	-	-	n.i.	n.i.
Trp ¹⁵	8.00	4.57	3.33, 3.12	-	7.07	7.30	10.18	-
Etam	7.82	3.24*	3.51*	-	-	-	-	-

Table A3.3: ¹H assignment of the major conformation of GramA-Asm in 70 % CD₃OH and 30 % H₂O at 37 °C. * indicates magnetically equivalent protons in methylene and magnetically equivalent methyl groups. + indicates aldehyde proton and n.i. non identifiable protons. Aromatic protons of Trp residues could not be assigned due to signal overlap.

	HN	H α	H β	H γ	H δ	H ϵ	HN _{sc}	HMe
Ald	8.17 ⁺	-	-	-	-	-	-	-
Val ¹	8.31	4.18	2.09	0.94* ⁻	-	-	-	-
D-Asm ²	7.81	4.70	2.75, 2.67	-	-	-	7.81	2.68
Ala ³	8.01	4.21	1.33	-	-	-	-	-
D-Asm ⁴	8.13	4.58	2.68, 2.60	-	-	-	7.65	2.68
Ala ⁵	7.89	4.29	1.33	-	-	-	-	-
D-Val ⁶	7.82	4.20	2.10	0.86*	-	-	-	-
Val ⁷	8.02	4.03	2.03	0.90, 0.87	-	-	-	-
D-Asm ⁸	8.36	4.75	2.59, 2.45	-	-	-	7.45	2.57
Trp ⁹	8.03	4.48	3.24, 3.14	-	7.06	7.48	10.11	-
D-Asm ¹⁰	7.87	4.59	2.7, 2.11	-	-	-	7.87	2.49
Trp ¹¹	7.96	4.53	3.23, 3.13	-	7.07	7.50	10.17	-
D-Leu ¹²	7.85	4.07	1.17, 1.11	0.91	0.59, 0.46	-	-	-
Trp ¹³	8.01	4.53	3.22, 3.07	-	7.09	7.51	10.15	-
D-Asm ¹⁴	7.89	4.49	2.33, 2.01	-	-	-	7.89	2.54
Trp ¹⁵	7.84	4.52	3.31, 3.10	-	7.07	7.55	10.05	-
Etam	7.68	3.23*	3.49*	-	-	-	-	-

4 Restrained MD-Simulations Using a Complete Set of One-Bond CH and NH Residual Dipolar Couplings for Cyclosporin A in $CDCl_3$ *

In this chapter, we present an extensive set of residual dipolar couplings (RDCs) data for cyclosporin A in chloroform. It was possible to obtain RDCs for all CH and NH bonds as well as for the homonuclear two-bonds HH in methylene groups by using a very weakly aligning compressed poly(methyl methacrylate) (PMMA) gel with a cross-linker to monomer ratio of 0.05 %. Conformational ensembles that match the RDC data were generated by molecular dynamics (MD) simulations using the recently introduced MDOC approach starting from two different crystal structures of cyclosporin A. Both ensembles could reproduce the experimental RDC data within experimental error, although they differ in the configuration of an amide bond (*cis/trans*). This indicates that even the entire set of heteronuclear CH and NH one-bond RDCs is not sufficient to unambiguously describe the conformational ensemble of cyclosporin A (at least at this alignment strength with the obtained experimental errors).

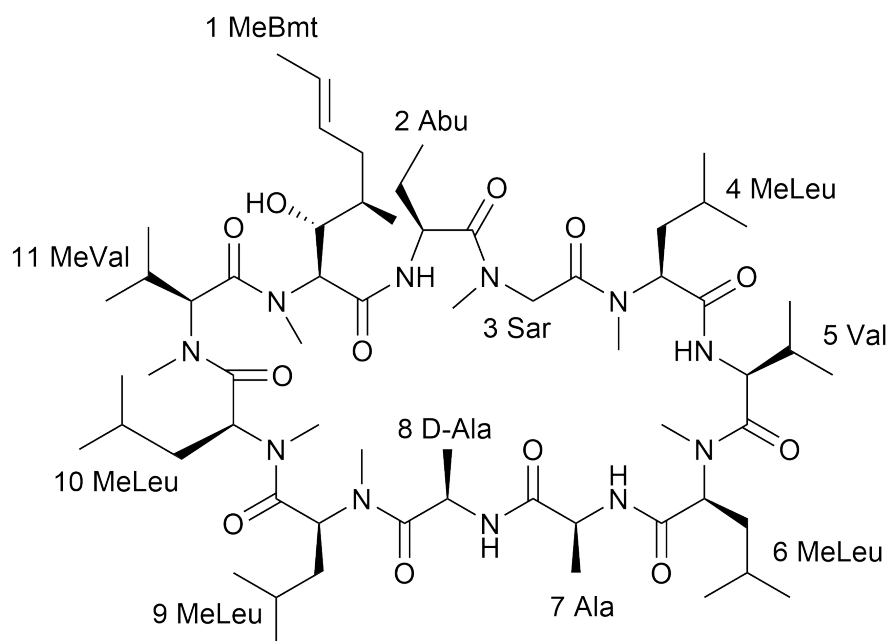
* The following semester, bachelor and master students contributed to this chapter. Stephan Feusi, Dénes Tary and Monique Kuonen: Optimization of compression device and cross-linked polymer preparation. Hristo Bonchev: Preparation of final PMMA and NMR measurements of cyclosporin A in solution and in PMMA gel.

4.1 Introduction

Residual dipolar couplings (RDCs) contain valuable structural information about the conformational ensemble of a molecule of interest. In contrast to the local information gained from chemical shifts (information about first and second sphere of atoms surrounding the nucleus), J-couplings (information about dihedral angles) and nuclear Overhauser Effect (NOE) (distance information up to 5 Å), the information gained from RDCs is of more global character.³⁹ Since the dipolar coupling is averaged to zero in solution, a partial alignment is necessary to observe RDCs. This alignment is extremely weak, thus the spectral resolution is largely similar to solution NMR.⁴⁰ RDCs have been used successfully for structure elucidation of large biomolecules with a relatively rigid core structure.^{208,209} However, making use of the information contained in the RDCs remains a major challenge for flexible systems.²¹⁰ Recently, several molecular dynamics (MD) simulation based methods were proposed to tackle this problem.^{72,211}

For the present study, we have recorded a large set of precise hetero- and homonuclear RDCs for the natural product cyclosporin A (Scheme 4.1) in a compressed poly(methyl methacrylate) (PMMA) gel in chloroform (CDCl₃). In CDCl₃, cyclosporin A shows only one dominant set of signals with about 95 % population, whereas in methanol, at least six different sets of signals can be observed.²¹² From extensive MD simulations in chloroform, it could be shown that cyclosporin A shows flexibility not only in the side-chains but also in the backbone.⁵¹

Recently, the incorporation of experimental RDCs in the COSMOS MD engine²¹³ has been implemented using tensorial constraints. We have evaluated this functionality using our newly recorded set of RDC data and the crystal structure of cyclosporin A, which resembles the major conformation in CDCl₃ (one *cis*-amide bond between residues 9 and 10),²¹⁴ as well as a more open crystal structure of cyclosporin A bound to cyclophilin (all *trans*-amide bonds).²¹⁵ Since the RDC data of the major conformation corresponds only to a subset of the conformational ensemble in CDCl₃ (about 95 %), also the generated ensemble with this approach will only cover this subfraction. If the incorporation of experimental RDCs in an MD simulation generates a reliable conformational ensemble, this method could become an extremely valuable tool to get insight into various physicochemical properties of the studied compound.



Scheme 4.1: Chemical structure of cyclosporin A with amino acids labelled in accordance to the literature.

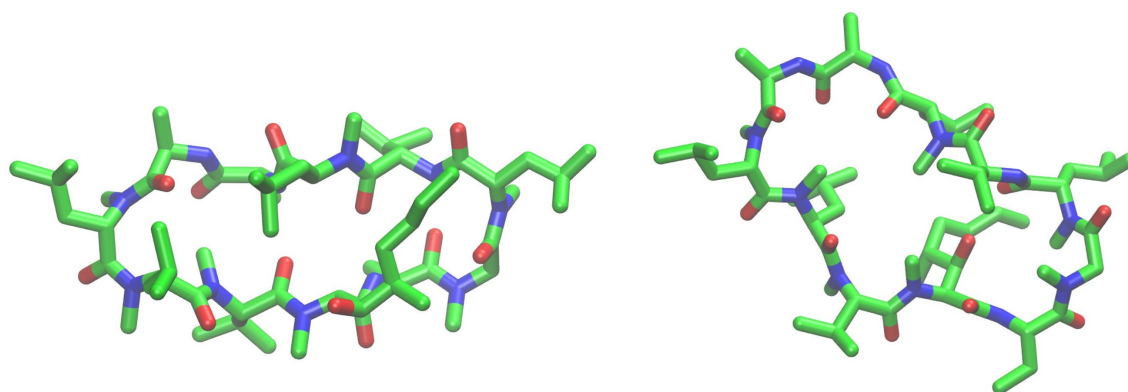


Figure 4.1: Comparison of the crystal structure similar to the solution structure in chloroform (left, CCDC code: DEKSAN) containing one cis-amide bond with the crystal structure of cyclosporin A co-crystallized with cyclophilin (right, PDB code: 2z6w) having all amide bonds in trans configuration. The crystal DEKSAN is more compact and forms four intramolecular hydrogen bonds, whereas the protein-bound crystal conformation is more open, forming hydrogen bonds with its environment. Only the heavy atoms are shown. The figure was created with VMD.¹⁴⁵

4.2 Theory

The dipolar coupling D_{IS} depends on the distance between the coupling nuclei and the angle between the external magnetic field and the internuclear vector. In the weak coupling limit, it can be expressed by:²¹⁶

$$D_{IS} = -\frac{\mu_0 \gamma_I \gamma_S \hbar}{16\pi^2} \left\langle \frac{1}{r_{IS}^3} (3 \cos^2(\theta) - 1) \right\rangle \quad (4.1)$$

where μ_0 is the magnetic permeability of the vacuum, γ_I and γ_S are the gyromagnetic ratios of the coupling nuclei I and S , and \hbar is the reduced Planck constant. r_{IS} is the internuclear distance, and θ is the angle between the internuclear vector and the external magnetic field B_0 . The angle brackets indicate the ensemble average. Note that Eq. (4.1) shows that the maximal value is twice as large as the minimal possible dipolar coupling. In an isotropic solution, all orientations are equally probable and therefore the isotropic coupling averages out to zero. In case of a weakly aligning medium, not all orientations have the same probability and therefore a residual dipolar coupling (RDC) can be observed. The size of the RDC is strongly dependent on the alignment strength of the medium.

4.2.1 Treatment of RDCs in Rigid Compounds

The simplest assumption one can make to obtain useful information from Eq. (4.1) is to consider the entire molecule as one rigid entity. For a better description of the molecule in an anisotropic alignment medium, it is beneficial to move from the laboratory frame to a frame of reference that is fixed to the molecule. In this new coordinate system, the RDC in a rigid molecule is then given by:²¹⁶

$$D_{IS} = -\frac{\mu_0 \gamma_I \gamma_S \hbar}{16\pi^2} \frac{1}{r_{IS}^3} (\vec{r}^T A \vec{r}) \quad (4.2)$$

where \vec{r} is the unit vector between the two spins I and S , and A is the alignment tensor. The latter is a symmetric and traceless 3 x 3 matrix that describes the alignment properties of the rigid molecule.

$$A = \begin{pmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{pmatrix} \quad (4.3)$$

$$A_{xx} + A_{yy} + A_{zz} = 0 \quad (4.4)$$

Five linearly independent RDCs are necessary to define the alignment tensor. With that, different possible rigid conformations can be compared to each other and the RDCs can be used for stereospecific assignment.

4.2.2 Treatment of RDCs in Flexible Compounds

For flexible molecules with more than one relevant conformation, the utilization of RDCs for structure elucidation becomes more complex. Two kinds of averaging processes have to be considered. There is still the overall tumbling motion of the compound, but now also conformational changes of the molecule have to be considered. When the time scales of the two processes are sufficiently different from each other, they can be separated.⁴¹ If the overall shape of the molecule is not affected by the conformational changes, the same alignment tensor can be used for all conformations (multi conformer single tensor approach).²¹⁷ If this is not the case, it would in principle be possible to fit a separate alignment tensor for each conformation (multi conformer multi tensor approach).²¹⁰ In practice, however, this is only rarely possible since $6n - 1$ RDCs will be needed for n conformers.

Camilloni and Vendruscolo proposed a different approach, termed “ ϑ -method”,²¹¹ to the problem of fitting an ensemble of conformations to a set of experimental RDCs. Abandoning the concept of an alignment tensor, the tensor-free ϑ -method uses directly the dependence of the dipolar couplings on the angle θ .²¹¹ To generate a conformational ensemble, the RDCs are incorporated in an MD simulation as replica-averaged structural restraints. According to Camilloni and Vendruscolo, “the generated ensemble should be the most probable one given the force field and the experimental data included, that reproduces at the same time the conformational dynamics of the system under study and the distribution of the orientations with respect to the alignment media employed to measure the RDCs”.²¹¹ The ϑ -method is subject to some controversy, as it seems to implicitly re-introduce aspects of the alignment tensor formalism, which are claimed to have been removed from the model.²¹⁸

Another approach for the treatment of RDCs measured in dynamic systems was recently presented by Tzvetkova *et al.*⁷² This method is based on tensorial constraints, which individually have to fulfill the secular dipolar interaction Hamiltonian in the laboratory frame without the assumption of an alignment tensor. For each RDC, an individual dipolar coupling tensor of the following form is constructed in its own principle axis system:⁷²

$$D_{IJ} = \begin{pmatrix} \frac{-RDC_{IJ}}{2} & 0 & 0 \\ 0 & \frac{-RDC_{IJ}}{2} & 0 \\ 0 & 0 & RDC_{IJ} \end{pmatrix} \quad (4.5)$$

A time average with exponential memory is used to introduce a new time scale for rotational reorientations and fluctuations. The difference between calculated time-averaged RDCs and experimental RDCs gives rise to a pseudo energy, which drives the system forward to a new ensemble that should be in better agreement with the experimental data. This procedure is carried out until convergence is reached. For more details, the reader is referred to two original publications, where the method is described in more detail.^{72,219}

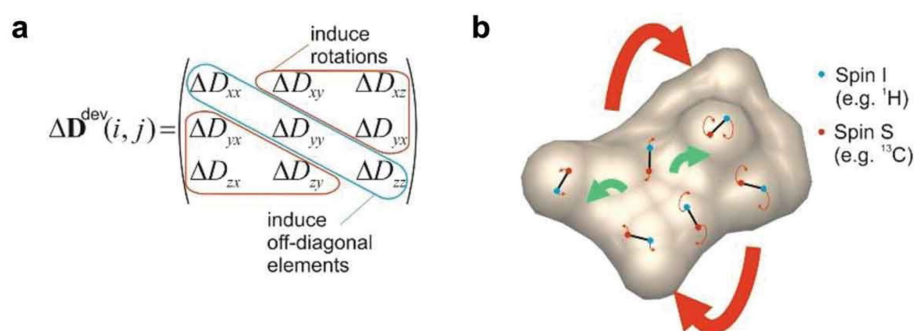


Figure 4.2: (a): Deviation matrix between experimental RDC tensor and time-averaged calculated RDC tensor. Deviations of diagonal elements induce non-zero off-diagonal elements. (b): Non-zero off-diagonal elements induce rotations of the corresponding vectors (orange arrows) that are constrained by the underlying force field. Differential rotational components lead to conformational changes (green arrows), whereas matching rotational components lead to global rotation (red arrows). Reprinted from Ref. 18 with permission from Taylor & Francis.

4.2.3 Experimental Determination of RDCs

Experimentally, an RDC is determined as the difference between a given coupling constant measured in an alignment medium ($T = J + D$) and in isotropic solution (J), respectively (Figure 4.3). Different approaches were developed to induce the necessary weak alignment. Traditionally, liquid crystals were used and are still the method of choice for large bio-molecules in aqueous solution.²²⁰ Further, it is also possible to make use of the self-aligning properties of paramagnetic ions that are tightly bound with a rigid tag to the studied molecule.²²¹ As a third method, the alignment can also be induced using an alignment gel. The anisotropy in such a gel can either be induced by stretching or compressing the swollen gel.⁴¹ The alignment strength is dependent on the cross-linker to monomer ratio in the gel and the applied strain.²²⁰ Compressed gels can be obtained either by restriction of swelling in the xy -plane directly by the NMR tube walls or in the z -direction. The first variant has some severe disadvantages since the swelling of the gel takes quite some time (for PMMA 20 – 30 days) and also diffusion of the compound of interest may take up to 3 days.²²² The second variant, originally presented by Gayathri *et al.*,²²² with compression in

z-direction swells within a few hours. Furthermore, the alignment strength is adjustable by tightening or loosening the pressure applied to the gel. Since the lines become broader with increasing alignment strength, it is essential to find optimal conditions, which give still relatively sharp signals but also RDCs of considerable size.

RDCs can in principle be measured with the entire arsenal of NMR experiments that are capable to determine ordinary J -couplings. For one-bond RDCs, the CLIP-HSQC experiment is often used to detect the coupling in the direct dimension.²²³ For crowded spectra, this can sometimes lead to overlapping signals and it is often beneficial to record the couplings in the indirect dimension. For methyl and methine protons, a ω_1 -coupled ^{13}C -HSQC spectrum with a G-BIRD^(r) element²²⁴ in t_1 together with J -scaling²²⁵ to increase the resolution in ω_1 is recommended.²²⁶ For individual CH coupling constants of diastereotopic protons in methylene groups, this experiment is not suitable as the splitting in ω_1 corresponds to the sum of the two $^1T/^1J$ -couplings. In this case, the J -HMQC-ge/se-HSQC experiment with J -scaling is the way to go.²²⁷

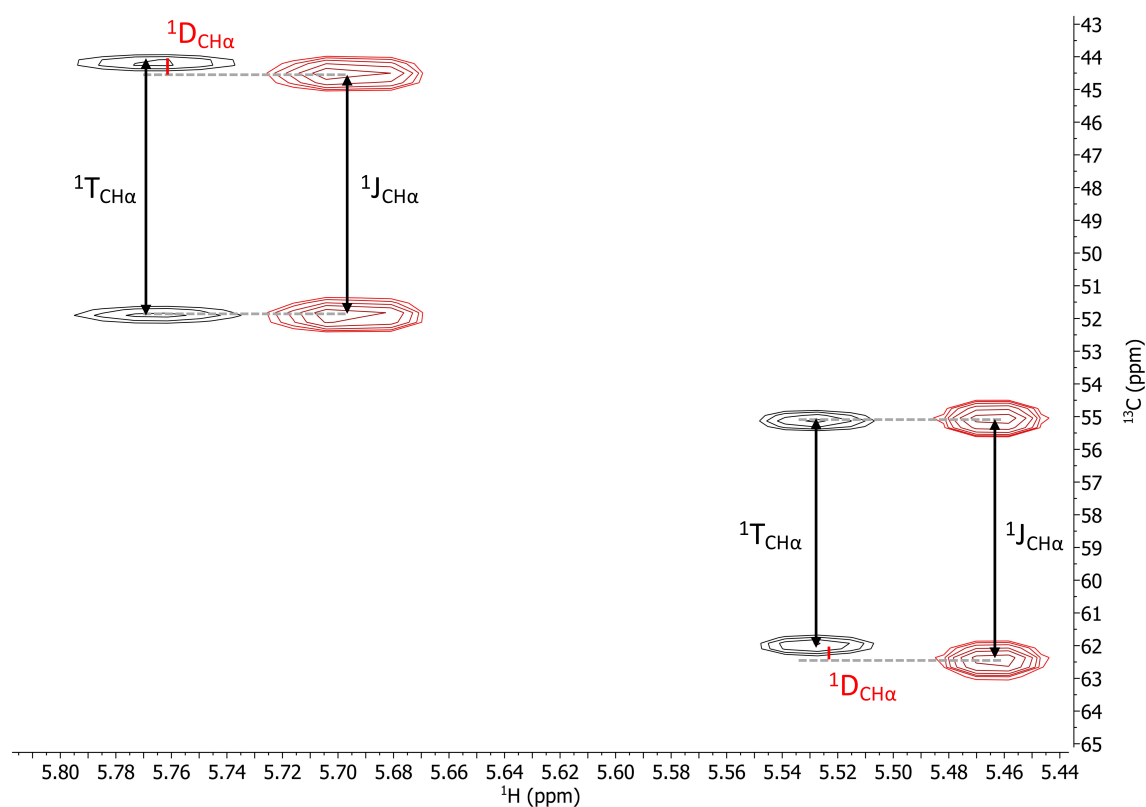


Figure 4.3: Detail view of overlaid ω_1 -coupled ^{13}C -HSQC spectra with a G-BIRD^(r) element and J -scaling of factor 8. Red: isotropic spectrum, black: spectrum in a weak alignment gel. The difference between the two measured couplings corresponds to the RDC of the one-bond CH vector. The RDC is positive for the signal on the left and negative for the signal on the right.

4.3 Results

Due to the very weak alignment in the PMMA gel with a cross-linker to monomer ratio of 0.05 %, it was possible to measure all 58 $^1D_{CH}$ RDCs for cyclosporin A. In addition, also the four $^1D_{NH}$ RDCs and the seven $^2D_{HH}$ RDCs could be measured. The RDC values with the corresponding estimated experimental errors are given in Table 4.1 and Table 4.2. The stereospecific assignment for the RDCs were taken from the literature.^{31,228} To our knowledge, this is the only complete set of 1D heteronuclear and $^2D_{HH}$ RDCs for cyclosporin A. In the RDC dataset provided by Klages *et al.*,³¹ only 35 of the 58 possible $^1D_{CH}$ RDCs could be measured, and their RDCs were in general larger. Especially RDCs of CH pairs with multiple protons in proximity were not observable in their case, probably due to long-range couplings that broadened the signals. Since we used an extremely weak alignment strength, our signals were nearly as sharp as in the isotropic solution, which allowed us to measure also these RDCs.

Table 4.1: List of one-bond CH and NH RDCs of cyclosporin A with the associated estimated error measured in 0.05 % cross-linked PMMA gel swollen in $CDCl_3$.

Residue	Coupling Nuclei	Measured RDC [Hz]	Estimated Error [Hz]
1 MeBmt	CMe-HMe*	3.0	0.1
1 MeBmt	C α -H α	-9.9	0.3
1 MeBmt	C β -H β	1.9	0.1
1 MeBmt	C γ -H γ	2.2	0.2
1 MeBmt	C δ -H δ proR	-4.8	1.3
1 MeBmt	C δ -H δ proS	0.3	0.3
1 MeBmt	C ϵ -H ϵ	1.8	0.3
1 MeBmt	C ζ -H ζ	1.3	0.1
1 MeBmt	C η -H η *	-0.5	0.1
2 Abu	C α -H α	-7.5	0.1
2 Abu	C β -H β proR	3.4	0.5
2 Abu	C β -H β proS	-2.2	1.9
2 Abu	C γ -H γ *	-2.1	0.8
3 Sar	CMe-HMe*	1.9	0.1
3 Sar	C α -H α proR	-4.8	0.6
3 Sar	C α -H α proS	-1.4	0.6
4 MeLeu	CMe-HMe*	-0.3	0.1
4 MeLeu	C α -H α	-0.1	0.1
4 MeLeu	C β -H β proR	-0.1	0.6

4 MeLeu	C β -H β proS	-4.6	0.5
4 MeLeu	C γ -H γ	-3.9	0.2
4 MeLeu	C δ -H δ proR*	-2.9	0.7
4 MeLeu	C δ -H δ proS*	0.4	0.1
5 Val	C α -H α	-4.2	0.1
5 Val	C β -H β	-4.3	0.1
5 Val	C γ -H γ proR*	-0.8	0.1
5 Val	C γ -H γ proS*	-3.3	0.2
6 MeLeu	CMe-HMe*	1.5	0.1
6 MeLeu	C α -H α	-9.5	0.2
6 MeLeu	C β -H β proR	0.0	0.5
6 MeLeu	C β -H β proS	-3.1	0.5
6 MeLeu	C γ -H γ	0.4	0.2
6 MeLeu	C δ -H δ proR*	1.5	0.1
6 MeLeu	C δ -H δ proS*	1.6	0.5
7 Ala	C α -H α	-6.0	0.3
7 Ala	C β -H β *	2.5	0.1
8 D-Ala	C α -H α	-2.2	0.2
8 D-Ala	C β -H β *	2.6	0.1
9 MeLeu	CMe-HMe*	-0.4	0.02
9 MeLeu	C α -H α	6.3	0.2
9 MeLeu	C β -H β proR	4.6	0.7
9 MeLeu	C β -H β proS	-6.4	0.6
9 MeLeu	C γ -H γ	-6.4	0.1
9 MeLeu	C δ -H δ proR*	-2.6	0.5
9 MeLeu	C δ -H δ proS*	-0.9	0.2
10 MeLeu	CMe-HMe*	2.3	0.2
10 MeLeu	C α -H α	-5.3	0.1
10 MeLeu	C β -H β proR	1.3	0.6
10 MeLeu	C β -H β proS	2.3	0.9
10 MeLeu	C γ -H γ	6.9	0.3
10 MeLeu	C δ -H δ down*	1.3	0.1
10 MeLeu	C δ -H δ up*	1.7	0.1
11 MeVal	CMe-HMe*	2.3	0.2

11 MeVal	C α -H α	-10.7	0.3
11 MeVal	C β -H β	-9.6	0.5
11 MeVal	C γ -H γ proR*	0.5	1.3
11 MeVal	C γ -H γ proS*	-4.3	0.3
2 Abu	N-HN	-3.6	0.1
5 Val	N-HN	-3.5	0.1
7 Ala	N-HN	-4.8	0.2
8 D-Ala	N-HN	1.3	0.1

Table 4.2: List of $^2D_{HH}$ RDCs of cyclosporin A with the associated estimated error measured in 0.05 % cross-linked PMMA gel swollen in $CDCl_3$.

Residue	Coupling nuclei	Measured RDC [Hz]	Estimated Error [Hz]
1 MeBmt	H δ proR- H δ proS	5.2	0.8
1 MeBmt	H δ proS- H δ proR	5.3	0.6
2 Abu	H β proR-H β proS	-1.6	0.5
2 Abu	H β proS-H β proR	-1.9	0.6
3 Sar	H α proR-H α proS	9.9	0.5
3 Sar	H α proS-H α proR	9.9	0.5
4 MeLeu	H β proR-H β proS	0.1	0.6
4 MeLeu	H β proS-H β proR	0.8	0.6
6 MeLeu	H β proR-H β proS	3.0	1.1
6 MeLeu	H β proS-H β proR	3.6	0.8
9 MeLeu	H β proR-H β proS	6.4	1.3
9 MeLeu	H β proS-H β proR	7.6	0.6
10 MeLeu	H β proR-H β proS	-1.4	0.6
10 MeLeu	H β proS-H β proR	-2.0	0.7

The $^1D_{CH}$ and $^1D_{NH}$ RDC data were used as input for molecular dynamics (MD) simulations with orientational constraints (MDOC) using the COSMOS simulation package.^{213,229} The principle how the RDC data is used for restraining and ensemble generation is shown in Figure 4.2. The $^2D_{HH}$ cannot yet be used with MDOC because the interatomic distance is strongly fluctuating. Two different crystal structures of cyclosporin A were taken as starting geometries. The first shows a conformation similar to the one observed in $CDCl_3$ solution (CCDC code: DEKSAN, one *cis*-amide bond between residues 9 and 10).²¹⁴ In the other one, cyclosporin A was co-crystallized with cyclophilin (PDB code: 2z6w, all *trans*-amide bonds).²¹⁵ The first crystal structure should be a good

starting conformation for the MDOC simulation and not much change in the overall shape is expected for the ensemble to reproduce the RDC data. For the other starting structure, the calculated ensemble will fit the RDC data initially relatively poorly and extensive structural reorganization is needed for an improved match, including a *trans*-to-*cis* isomerization of a peptide bond. Several MDOC runs were performed to determine appropriate parameters for the weight factor of the dipolar pseudo forces and the sample order parameter. The final settings were 0.0005 for the weight factor and 0.01 for the sample order parameter. A 30 ns MDOC simulation starting from the DEKSAN crystal structure fulfilled 109 of the 110 experimental RDCs within the experimental errors. No *cis*-to-*trans* isomerizations were observed (Figure 4.4, left). Starting from the alternative crystal structure (2z6w), all 110 RDC were fulfilled within experimental errors. Also here, no peptide bond isomerizations were observed (Figure 4.4, right). This means that the two obtained ensembles differ in the configuration of the peptide bond between residues 9 and 10.

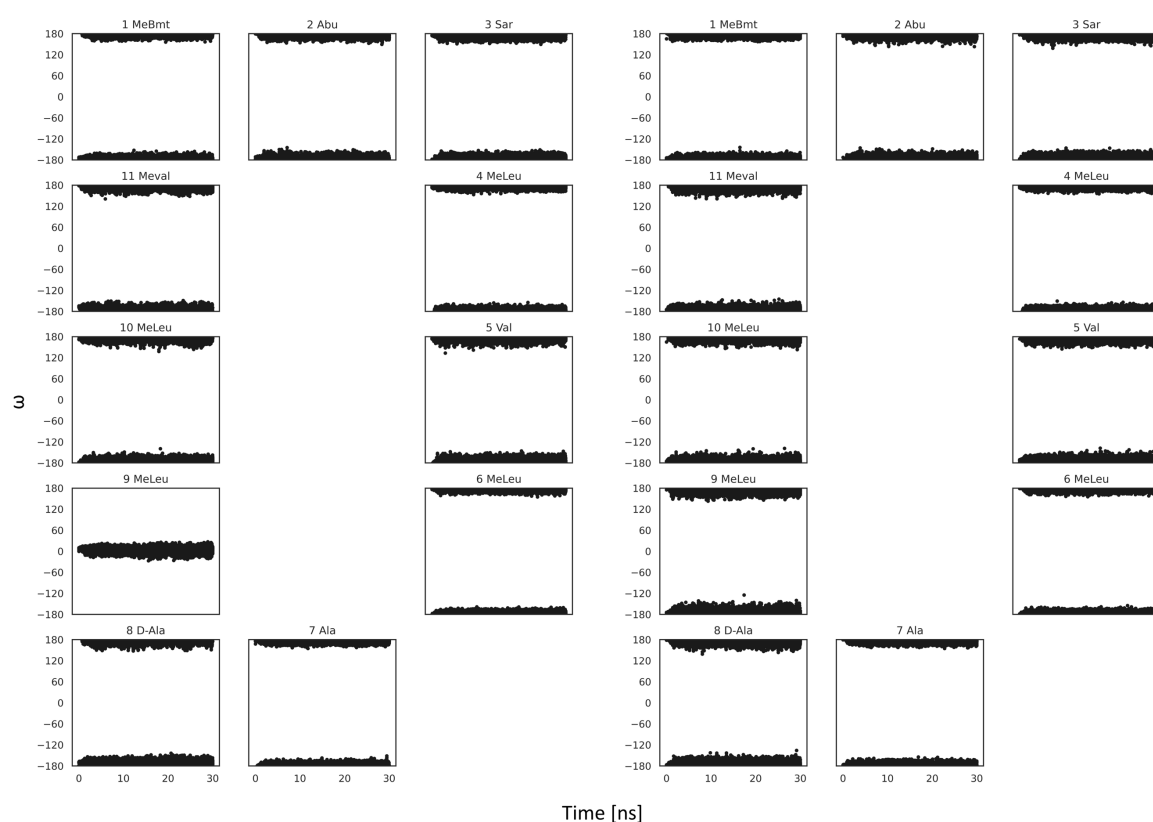


Figure 4.4: Torsional angle of the peptide bonds in cyclosporin A as a function of simulation time during the MDOC simulation starting from the crystal structures DEKSAN (left) and 2z6w (right). Note that the peptide bond between residues 9 and 10 is in the *cis*-configuration (i.e., 0°) in DEKSAN, while it is *trans* (i.e., 180°) in 2z6w. No *cis*-to-*trans* isomerization or vice versa was observed in both simulations.

Overall, cyclosporin A undergoes extensive conformational sampling in both MDOC simulations. Comparing the Ramachandran plots for the two MDOC simulations, it is notable that the distributions of the phi- and psi-dihedral angles are highly similar, except for the two residues (9 and 10) on both sides of the peptide bond whose configuration differs between the two starting structures (Figure 4.5 and Figure 4.6). Interestingly, when comparing the RMSDs of the backbone to the DEKSAN crystal structure, we find that both generated ensembles differ substantially from the NMR solution structure in chloroform (Figure 4.7). We would have expected that the backbone configurations of the simulation starting from the DEKSAN crystal structure stay relatively close to this starting structure, whereas significant changes should occur for the one starting from 2z6w including a potential *trans*-to-*cis* isomerization for the amide bond between residues 9 and 10.

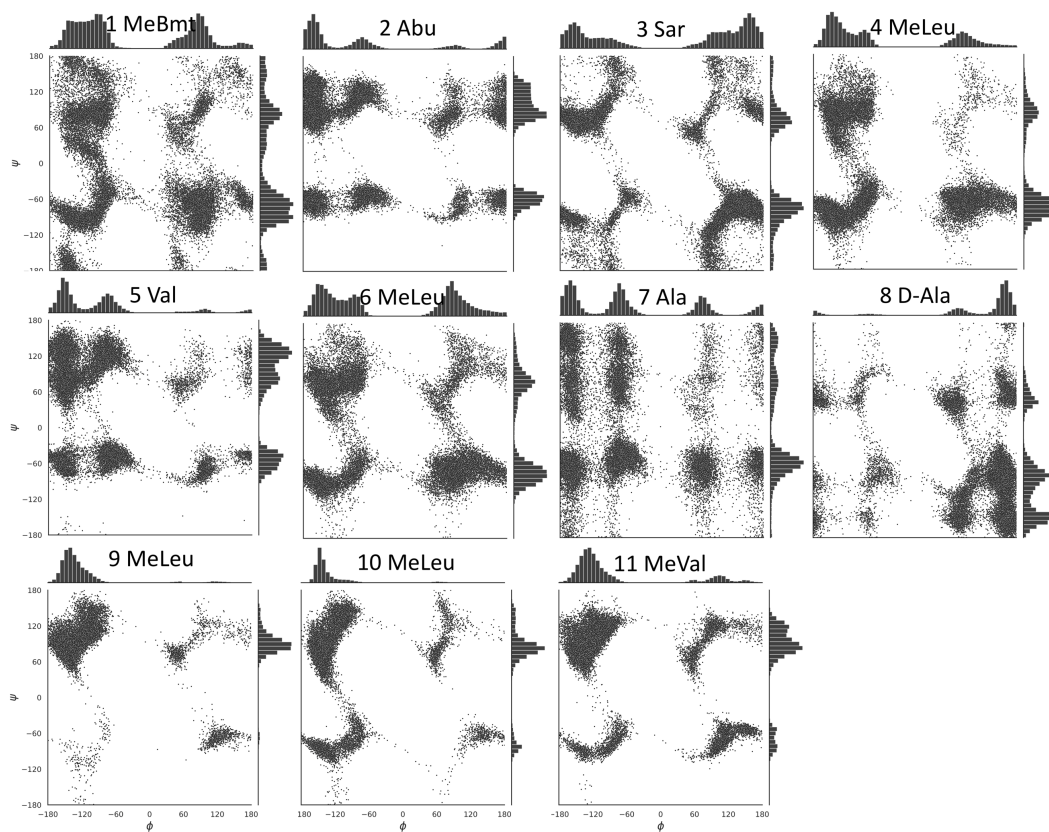


Figure 4.5: Ramachandran plot for the MDOC simulation starting from the DEKSAN crystal structure. The first 5 ns of the 30 ns simulation were discarded for equilibration.

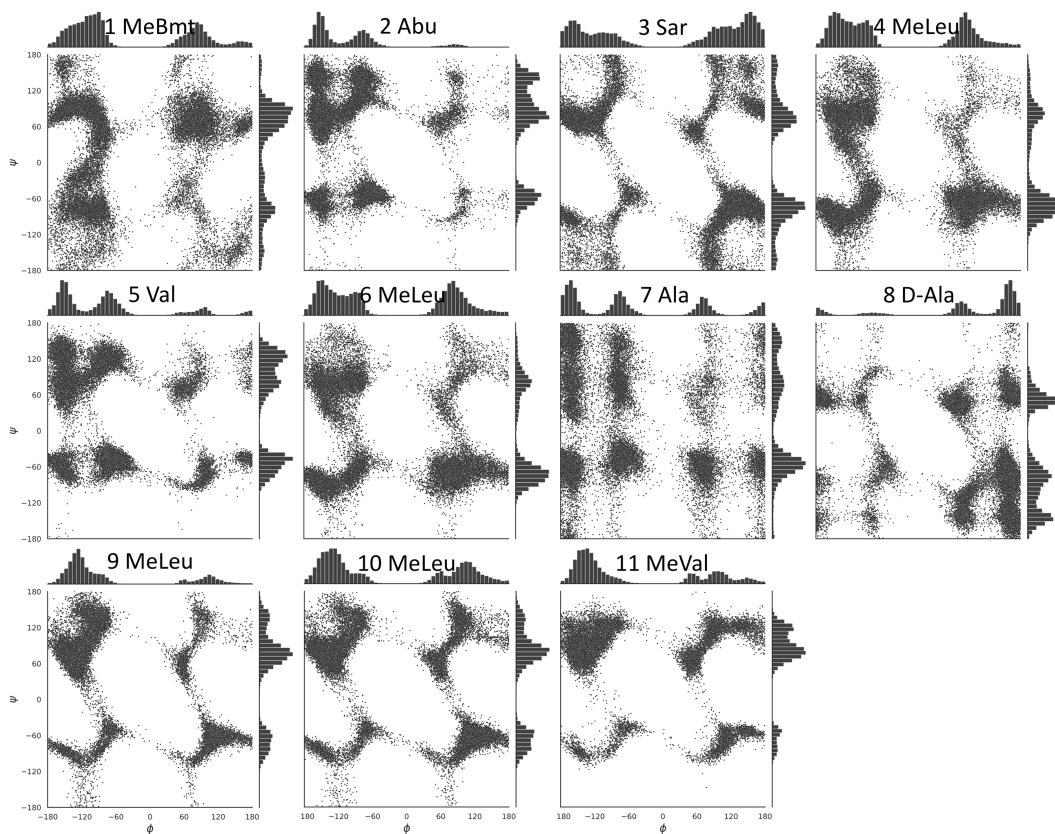


Figure 4.6: Ramachandran plot for the MDOC simulation starting from the 2z6w crystal structure. The first 5 ns of the 30 ns simulation were discarded for equilibration.

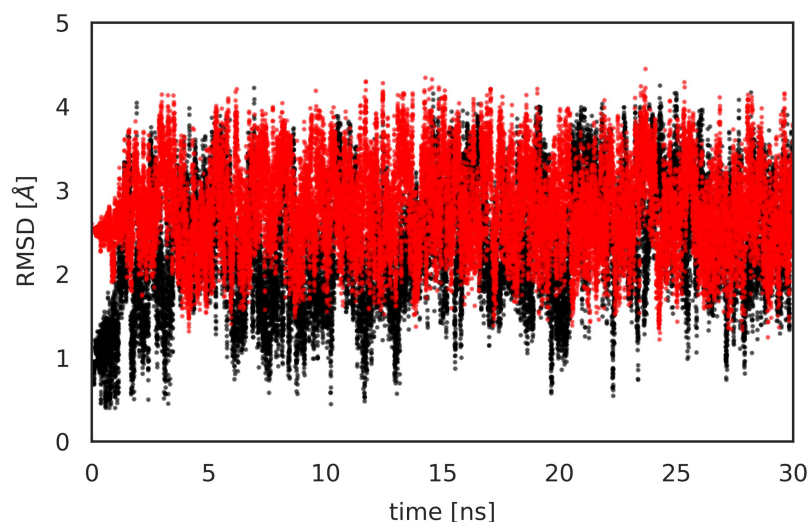


Figure 4.7: Backbone RMSD values to the DEKSAN crystal structure of the MDOC simulations starting from DEKSAN (black) and from 2z6w (red).

The fact that both ensembles fulfilled the experimental RDCs equally well is a strong indication that even when all one-bond CH and NH RDCs are available from experiment, it is not enough to unambiguously describe the conformational ensemble of cyclosporin A in solution. If we compare our Ramachandran plots to the ones from Witek *et al.*⁵¹ resulting from extensive unrestrained MD simulations in chloroform, it is evident that a much larger conformational space is visited in the MDOC simulations. This could be an indication that also unphysical states are visited due to too high pseudo forces from the RDC restraining. A possible issue could be that the side-chain RDCs account for a major part of the applied restraints. This can be considered as a dilution of the information from the backbone RDCs about the conformational space of the peptide backbone, especially, since it should be much easier to fulfill the side-chain RDCs than the backbone RDCs due to higher flexibility. However, all RDCs have the same weight in the simulation. In the current implementation, it is not possible to give them individual weights. Another aspect is certainly that our RDCs values are relatively small. On the one hand, this allows us to measure the complete set of RDCs with small absolute errors. On the other hand, the relative error becomes rather large, which also reduces the informative value of the RDC data. For further studies, it may be beneficial to compare the generated ensemble of the MDOC simulation using all available RDCs with an MDOC simulation using only the RDCs involving backbone atoms.

4.4 Conclusion

In this work, we presented a new set of precise RDC data for cyclosporin A recorded in a cross-linked PMMA gel swollen in chloroform. All one-bond CH and NH RDCs as well as all two-bond HH RDCs in methylene groups could be determined. This was made possible due to the very low cross-linker ratio of 0.05% chosen, since this gave nearly as sharp lines as the isotropic sample. MDOC simulations of 30 ns length were performed starting from two different crystal structures of cyclosporin A, one resembling the solution structure in chloroform (DEKSAN, one *cis*-amide bond) and one co-crystalized with cyclophilin (2z6w, all *trans*-amide bonds). The DEKSAN crystal structure resembles the NMR solution structure in chloroform determined from NOEs and should thus be a good starting point for the MDOC simulation. For the MDOC simulation starting from the 2z6w crystal structure, on the other hand, much more rearrangement was expected (including a *trans*-to-*cis* isomerization). Interestingly, we found that both ensembles obtained with MDOC reproduce the experimental RDCs within their experimental errors (109 of 110 in case of DEKSAN and 110 of 110 in case of 2z6w). Since no *cis*-to-*trans* isomerizations were observed during either of the MDOC simulations, the two ensembles are significantly different from each other, including the configuration of a peptide bond. Nevertheless, they reproduce the experimental RDCs equally well. This indicates that even the entire set of one-bond CH and NH RDCs is not enough information to unambiguously describe the conformational ensemble of cyclosporin A (at least at this alignment strength). Overall, it appears that the interpretation and extraction of the information obtained from RDCs is still non-trivial for flexible molecules. For future studies, the RDC dataset could be combined with J-coupling information as well as with NOE-derived distances. The combination of these datasets is possible within the MDOC approach, but additional parameters need to be optimized for the correct relative scaling of the pseudo forces.

4.5 Method Section

Experimental Details

Cross-linked poly(methyl methacrylate) (PMMA) sticks were prepared following a slightly modified version of the procedure described by Gayathri *et al.*²²² Stabilized methyl methacrylate (MMA) (Acros) and stabilized ethylene glycol dimethyl acrylate (EGDMA) (Acros) were run through a short column of basic alumina (MP Alumina B, Act I, EcoChrom) to remove the stabilizer. The polymer was prepared by mixing 10 ml MMA, 2 ml acetone-d₆ (Armar) and 3 mg of the radical starter V70 (2,2'-azobis(2,4-dimethyl-4-methoxyvaleronitrile)) (Fujifilm). Out of this solution, 10 ml were transferred into a new vessel and 7.4 μl EGDMA was added (monomer to cross-linker ratio of 0.05 %). Three freeze-pump-thaw cycles were applied to degas the solution to prevent the formation of air bubbles during polymerization. The solution was transferred into 3 mm NMR tubes (Bruker LabScape). The tubes were capped and put in an oil bath at 50 °C for four hours. Then the caps were removed and the polymer sticks were allowed to dry in the tubes for two days at RT. Afterwards the tubes were cut in half with a diamond cutter and the sticks were gently removed. The PMMA sticks were allowed to dry further for one day at RT and then cut into pieces of the desired length (between 2.5 and 4 cm). The short PMMA sticks were swollen in a 1:1 solution of acetone-d₆ and methanol-d₄ for several hours and afterwards put in a small vial with chloroform-d that was exchanged three times. Each washing step lasted at least 30 min. Finally, the sticks were allowed to dry at room temperature.

To induce the anisotropic environment, an improved variant of a commercially available gel compression device (NE-375-5, NewEra) was used. The original device was made out of Teflon, which has beneficial properties for this application since it is resistant against most solvents and does not give signals in a ¹³C-HSQC spectrum due to complete fluorination. But Teflon has only limited mechanical stability and especially the thread wears out quickly. Another issue is the mechanism by which the compression device is held on the NMR tube. For this, an NMR tube with a small collar close to the opening is provided by the vendor. But if the plunger presses tightly on the gel, the soft Teflon holding mechanism may easily slip over the collar and during such an incident also the collar may break. To tackle these issues, we decided to use a commercial screw cap NMR tube instead of the collar tube. This provides an easier mechanism for attachment and also greater stability. Second, the workshop of the Laboratory of Organic Chemistry (LOC) at ETH Zurich made an improved variant of the compression device out of polyether ether ketone (PEEK). This polymer is also very stable against most solvents and in addition shows excellent mechanical stability. The only drawback is that PEEK is not fluorinated and potentially contributes to the spectral background. Therefore, we use a small Teflon rod at the bottom of the plunger to avoid

this issue. The Teflon rod has an inner thread for easy removal from the tube. The improved device is shown together with the commercially available one in Figure 4.8.

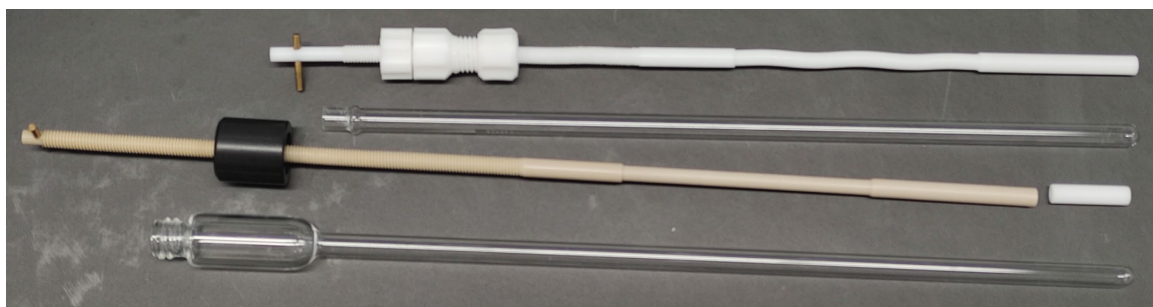


Figure 4.8: Commercially available gel compression device made out of Teflon using a collar glass tube (top) and the improved compression device made out of PEEK using a screw cap NMR tube that is more robust (bottom). After repeated use, the commercial compression device started to deform and also the threads are worn out, whereas this could not be observed for the improved variant.

For the NMR experiments in anisotropic solution, a PMMA stick of 4 cm length was put into a 5 mm screw cap NMR tube. Then 600 μl of a 50 mM cyclosporin A solution in CDCl_3 was added into the tube (the cyclosporin A was a kind gift of Prof. Seebach, ETH Zürich). The plunger of the compression device was immediately positioned at the top of the polymer stick to prevent swelling in the longitudinal direction. The stick was then allowed to swell overnight. Before measurement, the gel was compressed to a length of 3.9 cm. The isotropic sample was measured at the same concentration in a normal 5 mm NMR tube.

All spectra were recorded on a 600 MHz spectrometer equipped with a cryogenic Prodigy triple-resonance probe with z-gradients. The temperature was kept constant at 25 °C. The ^1H spectrum of the isotropic sample was recorded with a spectral width of 10 ppm. The transmitter was set to 4 ppm. 65536 data points were recorded with 32 scans. The ^1H spectrum of the compressed gel was recorded with a PROJECT- T_2 filter of 320 ms to suppress the polymer signals.²³⁰ A spectral width of 12 ppm was used and the transmitter was set to 4.5 ppm. 43268 points were recorded with 8 dummy scans followed by 8 scans. The quality of the gel was assessed by recording a ^2D spectrum and looking at the deuterium splitting of the CDCl_3 signal. A spectral width of 10 ppm was used and the transmitter was set to 5 ppm. 3686 data points were recorded with 4 scans. A deuterium splitting of 7.6 Hz was observed.

All $^1T_{\text{CH}}/^1J_{\text{CH}}$ coupling constants were measured in the indirect dimension. In case of methyl and methine protons, a ω_1 -coupled ^{13}C -HSQC spectrum with a G-BIRD^(r) element²²⁴ in t_1 to remove undesired long-range couplings was used. J -scaling²²⁵ was applied to increase the resolution in ω_1 .²²⁶ For the determination of the individual CH-coupling-constants of diastereotopic protons in methylene groups this experiment is not suitable as the splitting in ω_1 corresponds to the sum of

the two ${}^1T/{}^1J$ -couplings. In this case the J -HMQC-ge/se-HSQC experiment proposed by Fehér and Kövér was used.²²⁷ This experiment was also used to measure the two-bond homonuclear RDCs of methylene groups. The corresponding pulse sequences in the Bruker pulse program library are HSQCBIETGPJCSP.2 and HSQCBIETGPJCMQSP, respectively.

For the isotropic sample, both experiments were recorded with spectral widths of 6 ppm in the direct dimension and 150 ppm in the indirect dimension. The transmitters were set to 3 ppm and 65 ppm, respectively. For the anisotropic sample, the ω_1 -coupled ${}^{13}\text{C}$ -HSQC was recorded with spectral widths of 6 ppm in the direct dimension and 87 ppm in the indirect dimension. The transmitters were set to 3 ppm and 36 ppm, respectively. The J -HMQC-ge/se-HSQC experiment was recorded with spectral widths of 5 ppm in the direct and 65 ppm in the indirect dimension. The transmitters were set to 2.5 ppm and 35 ppm, respectively. For all four spectra, a total of 1024 x 512 points was recorded using a J -scaling factor of 8. For the two isotropic spectra as well as for the ω_1 -coupled ${}^{13}\text{C}$ -HSQC spectrum of the anisotropic sample, 8 scans per increment were accumulated whereas for the anisotropic J -HMQC-ge/se-HSQC 16 scans per increment were accumulated.

The spectra were processed with Topspin 4.1 (Bruker Biospin AG) and both time domains in all spectra were extended to twice their size by zero-filling and apodized with a \cos^2 function. Baselines were corrected using polynomials of fifth order. To obtain the RDCs, ${}^1T_{\text{CH}}$ and ${}^1J_{\text{CH}}$ couplings were determined from the ω_1 -coupled ${}^{13}\text{C}$ -HSQC and J -HMQC-ge/se-HSQC spectra. The ω_1 -trace corresponding to a given multiplet was extracted as the sum of the corresponding columns and overlaid with itself. By shifting one of the two traces such that the right half of the multiplet is directly on top of the left half of the multiplet in the other spectrum, the coupling constant can be directly read off as the offset between the two spectra. By evaluation of the offset range for which the superposition is still decent, one obtains an error range for the corresponding coupling. It is important to note that, in case of a diastereotopic CH_2 group, the ${}^1T_{\text{CH}}$ or ${}^1J_{\text{CH}}$ coupling of H proR in a J -HMQC-ge/se-HSQC spectrum is determined from the multiplet splitting at the ω_2 -position of H proS, and vice versa. The difference between a given coupling constant obtained for the anisotropic (T) and the isotropic sample (J) gives the RDC. Since a J -scaling factor of 8 was used for the J -coupling evolution in t_1 , the obtained RDC additionally needs to be divided by this number.

${}^1D_{\text{NH}}$ RDCs were determined using the corresponding ω_1 -coupled G-BIRD^(r)- ${}^{15}\text{N}$ -HSQC experiment. 512 x 128 data points were recorded for the isotropic sample with spectral widths of 1.5 and 40 ppm. The transmitters were set to 7.75 and 120 ppm, respectively. For the anisotropic sample,

512 x 256 data points were recorded with spectral widths of 1.5 and 25 ppm. The transmitters were set to 7.75 and 120 ppm, respectively.

Computational Details

The molecular dynamics with orientational constraints (MDOC)²³¹ simulations were performed using the COSMOS package²¹³ with the COSMOS-NMR force field.²²⁹ The procedure described by Tzvetkova *et al.*⁷² was followed closely. The temperature was set to 290 K with a thermostat coupling time constant of 0.02 ps, and 60'000'000 steps with a Verlet time step of 0.5 fs were carried out (total trajectory time 30 ns). The atomic charges were recalculated every four steps using bond polarization theory.²³² Translational and rotational motion was reset every 1000 steps. The π -bond torsion factor was set to 1.5. Rise time for the orientational pseudo forces were set to 2 ns. The weight factor for dipolar pseudo forces was set to 0.0005 and the sample order parameter was set to 0.01. The time constant for the property average was set to 2 ns and for the analysis the first 5 ns were discarded. For more details, the reader is referred to the publication of Tzvetkova *et al.*⁷²

5 Comparison of Experimental and DFT Calculated Chemical Shifts Using a New Standardized Dataset Recorded in Chloroform and Tetrachloromethane*

^1H and ^{13}C chemical shifts of 35 small, rigid molecules were measured under standardized conditions in chloroform-d and in tetrachloromethane. The solvent change mainly affects carbon shifts of polar functional groups. This difference cannot be adequately reproduced by DFT calculations in implicit solvent. We suspect specific solvent interactions in CDCl_3 to be the reason for this. Ignoring an incomplete representation of the solvent shell potentially obscures the direction to further improve DFT methods and hampers their fair assessment. The newly recorded datasets provide an accurate basis for the validation and calibration of DFT shift calculations, especially with respect to improved solvent models.

* Chantal Balmer contributed in the measurement of the NMR data presented in this chapter during her semester project in our group.

5.1 Introduction

Chemical shifts are the basis for the identification of organic molecules by NMR.¹² Structure elucidation and the determination of relative configuration can be assisted by comparison with chemical shieldings calculated using density functional theory (DFT). With advances in computational resources and quantum chemical software, chemical shieldings can be obtained nowadays on a routine basis.²³³ For comparison with experimental data, the chemical shieldings calculated with DFT have first to be transformed into chemical shifts. Most simply, the shielding of tetramethylsilane (TMS) can be calculated at the same level of theory and then used directly to obtain the chemical shifts:

$$\delta_{calc_i} = \sigma_{calc_{TMS}} - \sigma_{calc_i} \quad (5.1)$$

where δ_{calc_i} is the calculated chemical shift of atom i , and $\sigma_{calc_{TMS}}$ and σ_i are the calculated shieldings of TMS, and atom i , respectively. Instead of TMS, any known chemical shift determined with respect to TMS can be used as reference, provided the corresponding shielding calculation has been carried out at the same level of theory:

$$\delta_{calc_i} = \delta_{exp_{ref}} + \sigma_{calc_{ref}} - \sigma_{calc_i} \quad (5.2)$$

To compensate for shortcomings of the theoretical method in specific electronic environments, one can also employ multiple standards for different types of carbons (e.g., based on hybridization).²³⁴ In organic chemistry, the most popular way to convert calculated shieldings to chemical shifts is to use a large set of reference data and perform a linear regression between calculated shieldings and experimental shifts:

$$\delta_{calc_i} = \frac{\sigma_{calc_i} - q}{m} \quad (5.3)$$

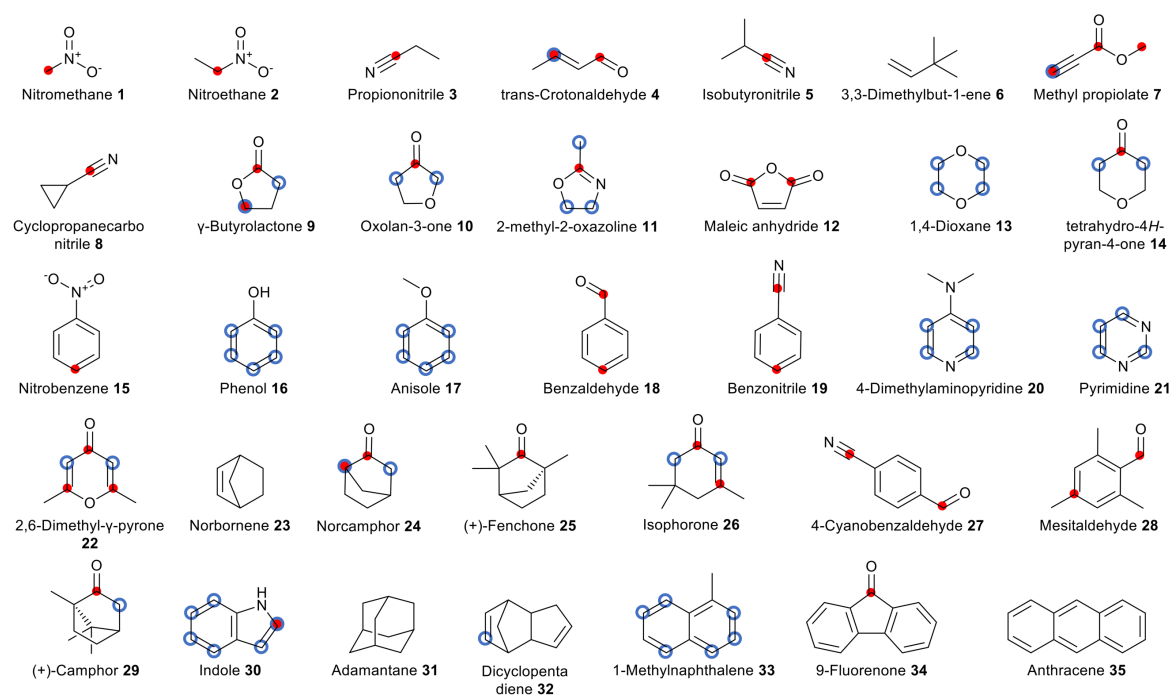
where m is the slope and q the intercept of the regression line. If the chosen level of theory could reproduce all shieldings perfectly, a slope of -1 should be obtained and q would be equivalent to σ_{TMS} (see Eq. (5.1)). This method implicitly assumes an identical relative error for all calculated shieldings, including the reference, that can be compensated by m . In general, a high-quality reference dataset of experimental chemical shifts is needed for the calibration. Chemical shifts are very sensitive to the environment and depend on the concentration and temperature.^{235,236} Also, the presence of impurities and the protonation state can influence the chemical shift.²³⁷ Therefore, control over the conditions at which the shifts were recorded is important for high-quality data. A popular reference set is the one from Tantillo *et al.*,²³⁸ which consists of experimental ^1H and ^{13}C data of 80 small and relatively rigid organic molecules. Numerous scaling

factors for different functional and basis set combinations can be found on their webpage (<http://cheshirenmr.info>). The data points were, however, collected from different sources, and experimental conditions (i.e., temperature, concentration and purity) are not indicated. In addition, the dataset contains multiple chlorinated compounds whose ^{13}C chemical shifts are affected by relativistic effects that cannot be accounted for by standard DFT. Hehre *et al.*²³⁹ used a reference set of 2000 molecules to derive an elaborate empirical correction scheme for ^{13}C chemical shifts calculated at the inexpensive $\omega\text{B97X-D/6-31G}^*$ level of theory. Also here, no special attention was paid to solvent and concentration effects or standardized experimental conditions.

Given a high-quality reference set, there are different types of errors and approximations that can impact the quality and reliability of the resulting shift prediction. One source of error is the chosen DFT method itself (functional and incomplete basis set). It is possible to go beyond DFT and perform a coupled-cluster calculation for improved accuracy, but such calculations are computationally extremely expensive and only feasible for very small systems. Numerous studies attempted to determine the best combination of functional and basis set for chemical shift calculations.^{238,240–243} Presumably, some combinations are only better than others due to fortunate error compensation. In terms of the applied level of theory, a good compromise between accuracy and computational cost are double-hybrid functionals, giving mean absolute relative errors as low as 1.9 % for the calculated shieldings compared to coupled-cluster calculations.²⁴⁴ Another potential source of error when comparing calculated shieldings with experimental chemical shifts are specific intermolecular interactions with impurities like water or the solvent itself (e.g., via hydrogen bonds), which are not trivial to account for in DFT calculations.^{245–247} In addition, also vibrational contributions to the chemical shift are usually neglected.

In this study, we focus on the solute-solvent interactions and how well these can be reproduced by DFT calculations with implicit solvent models. For this purpose, we generated a high-quality reference set of ^1H and ^{13}C chemical shifts measured in two solvents, chloroform-d (CDCl_3) and tetrachloromethane (CCl_4). Although both can be considered apolar solvents, the solute-solvent interactions are stronger in CDCl_3 because CCl_4 cannot act as a hydrogen-bond donor. This pair of solvents thus allows us to assess the impact of solute-solvent interactions on chemical shifts without solubility issues and very strong interactions. ^1H and ^{13}C chemical shifts were measured in both solvents for a set of 35 small and rigid organic molecules, consisting only of H, C, N, and O atoms (Scheme 5.1). The experimental dataset was recorded under standardized conditions, referenced to internal TMS, and all chemical shifts were reassigned to eliminate potential

incorrect assignments. To make these measurements as reproducible as possible and to reduce unwanted intermolecular interactions between different solute molecules as well as between solute and impurities, a concentration of 10 mM was chosen for all compounds. On our 600 MHz spectrometer equipped with DCH cryoprobe, this concentration still allows measuring of a ^1H and a ^{13}C spectrum within a reasonable amount of time. Only measurements were added to the dataset, for which impurity concentrations (including water) were below 20 % of the solute concentration (<2 mM) (except for **8**, where the water concentration was 20.5 %). The temperature during the measurements was kept constant at 25 °C. By comparing the experimental chemical shifts in CDCl_3 and CCl_4 , we can quantify the effect of solute-solvent interactions. The generated dataset is then used to assess the ability of standard DFT methods – with or without implicit solvent – to reproduce the experimental observations. This high-quality reference set of chemical shifts presents thus a valuable resource for the validation of DFT methods and the calibration of calculated chemical shieldings.



*Scheme 5.1: Chemical structures of the 35 molecules in the dataset used to measure ^1H and ^{13}C chemical shifts in chloroform-*d* and tetrachloromethane. Nuclei with large solvent-induced changes in chemical shift are marked. Red dots: carbons that have a difference in ^{13}C chemical shift larger 1 ppm between CDCl_3 and CCl_4 . Blue circles: hydrogens that have a difference in ^1H shift larger than 0.1 ppm between the two solvents.*

5.2 Results

After averaging magnetically equivalent nuclei, 141 ^1H and 170 ^{13}C unique chemical shifts were obtained for the 35 molecules in the dataset and unambiguously assigned in chloroform-d (CDCl_3) and tetrachloromethane (CCl_4). Exchangeable protons were not included (i.e., hydroxy proton of **16** and amine proton of **30**). Using this dataset, we first investigated the differences in the experimental data, followed by a detailed comparison with DFT calculations of NMR chemical shieldings.

5.2.1 Effect of Experimental Protocol and Water Content in Chloroform

By using a standardized protocol and carefully controlling concentration, temperature, and water content in the experiments, the chemical shifts measured in this study can be considered a highly homogeneous dataset. It presents therefore a unique opportunity to assess the effect of variations in the experimental protocol by comparing to literature values from different sources. Although the water content is naturally limited in chloroform, it may still impact the measured shifts. Figure 5.1 shows the comparison between our ^1H and ^{13}C chemical shifts in CDCl_3 and the literature values. As can be seen in Figure 5.1 and the evaluation metrics in Table 5.1, a good agreement is observed, indicating that the effects of concentration, water content, other impurities, or slight variations in temperature are relatively small for the ^1H and ^{13}C chemical shifts of the studied molecules in CDCl_3 . Along with effects of concentration, temperature and purity, also proper referencing of the spectra and potential typos might be responsible for some of the variability. If we classify our data according to their association with a functional group, the largest deviations for the ^{13}C chemical shifts were found for carbonyl carbons (RMSD of 0.43 ppm), which apparently are most affected by changes in sample or experiment conditions. The largest *individual* deviation from the values in our ^{13}C dataset likely stems from a change in protonation state (see Table 5.1).

While our shift data in CDCl_3 and the literature values (see Appendix) do not differ much on average, the spread for some of the data points is still as high or higher as the accuracy one would like to achieve in a chemical shift calculation by DFT (≤ 1 ppm for ^{13}C and ≤ 0.1 ppm for ^1H). Thus, the observed differences are still too large for the validation of computational approaches. Especially for the study of solute-solvent interactions or similar weak effects, the use of compiled literature values is not optimal since their influence on the chemical shifts is expected to be on the same order of magnitude as the spread in experimental values. The homogeneous sets of chemical shifts measured in this study will therefore serve as a valuable reference set to both validate *in silico* methods and to investigate the influence of specific solvation effects.

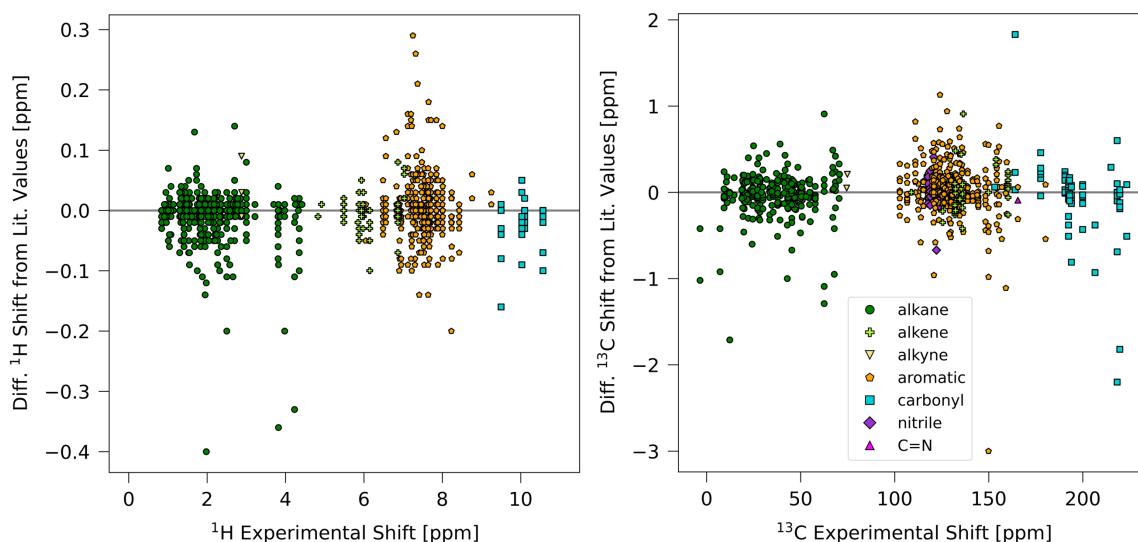


Figure 5.1: Comparison of experimental ^1H (left) and ^{13}C (right) chemical shifts measured under standardized conditions to values collected from multiple literature sources in CDCl_3 . A positive difference indicates that the literature value is larger compared to the one measured under standardized conditions. The data points are color and symbol coded with respect to the functional type of the carbon atom. Evaluation metrics are provided in Table 5.1.

Table 5.1: Root-mean-square deviation (RMSD), mean absolute deviation (MAD) and maximum absolute deviation (max. AD) when comparing the ^1H and ^{13}C chemical shifts measured under standardized conditions to values collected from multiple literature sources in CDCl_3 . (^a) methyl protons of **11**, (^b) carbons of the pyrimidine ring next to amine of **20**.

	^1H	^{13}C
RMSD [ppm]	0.05	0.26
MAD [ppm]	0.03	0.15
Max. AD [ppm]	0.40 ^a	3.00 ^b

5.2.2 Effect of Solute-Solvent Interactions in Experiment

In general, one would expect that polar groups (i.e., carbons of carbonyls, nitriles and other hydrogen bond accepting groups) experience the largest differences in chemical shifts between CCl_4 with no hydrogen-bonding capacity and CDCl_3 , which is able to form (weak) solute-solvent interactions. Figure 5.2 shows the difference between the chemical shifts in CCl_4 and CDCl_3 colored by functional group with the corresponding evaluation metrics given in Table 5.2. For the ^{13}C chemical shifts, a clear trend can be observed that correlates with the polarity of the functional groups. The largest difference is found for carbonyl carbons (cyan squares, RMSD of 3.82 ppm), followed by the nitrile carbons (violet diamonds, RMSD of 2.13 ppm), while the ^{13}C chemical shifts of the sp^3 carbons are similar in both solvents (green circles, RMSD of 0.54 ppm). An illustrative example is given for **26** in the Appendix. Interestingly, no such trend associated with the functional group was found for the ^1H shifts, although a general shift towards higher field is observed when going from CDCl_3 to CCl_4 .

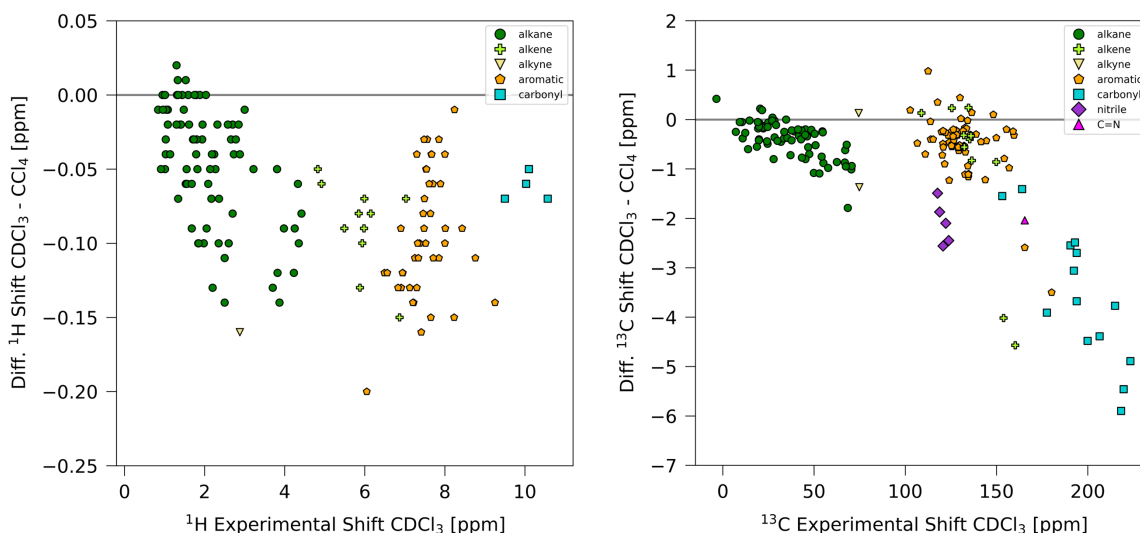


Figure 5.2: Comparison of experimental ^1H (left) and ^{13}C (right) chemical shifts measured under standardized conditions in CDCl_3 and CCl_4 . The data points are color and symbol coded with respect to the functional type of the carbon atom. The evaluation metrics are provided in Table 5.2.

Table 5.2: Root-mean-square deviation (RMSD), mean absolute deviation (MAD) and maximum absolute deviation (max. AD) when comparing the ^1H and ^{13}C chemical shifts measured in CCl_4 and CDCl_3 .

	^1H	^{13}C
RMSD [ppm]	0.08	1.40
MAD [ppm]	0.07	0.85
Max. AD [ppm]	0.20	5.90

To visualize in more detail which of nuclei experience the largest solvent-induced change in chemical shift, we selected all carbon and hydrogen chemical shifts which deviate more than 1.0 and 0.1 ppm between the two solvents, respectively (Scheme 5.1). The marked nuclei agree with positions where – according to the general concepts used in organic chemistry – one would expect the largest change in partial charge upon protonation or interaction with a hydrogen-bond donor.

Overall, the observations in our dataset confirm that there is a clear effect of the solvent-solute interactions on the chemical shifts – even for relatively apolar solvents –, which is important to take into account when comparing with computational approaches.

5.2.3 How Well Do DFT Calculations in Vacuum Reproduce Experimental Chemical Shifts in Solution?

DFT calculations to estimate NMR chemical shieldings have traditionally been performed in vacuum. To assess how large the effect of the vacuum conditions is, we compared the experimental values with DFT calculations carried out without implicit solvent using the gauche invariant atomic orbital (GIAO)⁸² approach in Orca 5.0.1^{248–250} with either the hybrid GGA functional PBE0²⁵¹ with the cc-pVTZ basis set²⁵² (called PBE0 in the following) or the 2013 version of the double-hybrid functional PBEP86²⁵³ together with the pcSseg-3 basis set²⁵⁴ (called PBEP86 in the following) on structures optimized at the BP86/def2-tzvp^{255–258} level of theory. Both methods have been shown in the past to perform well in the calculation of chemical shieldings.^{244,259} As the solute-solvent interactions are weaker in CCl₄ than in CDCl₃, we expect the vacuum condition in the calculations to be more appropriate for the former solvent. Table 5.3 gives the RMSD values when comparing calculated chemical shifts in vacuum with experimental values in CDCl₃ or CCl₄. The graphical comparisons are provided in Figure A5.5 and Figure A5.6 in the Appendix. For the ¹H shifts, only a very small change in RMSD (0.01 ppm) is observed when comparing to CDCl₃ or CCl₄ data. This might seem surprising as the two solvents gave an RMSD of 0.08 ppm for the ¹H chemical shifts in experiment (Table 5.2). However, some of the offset between the two datasets can be compensated by changing the intercept in the linear regression (q in Eq. (5.3)) without a significant increase in RMSD. A large part of the potential performance differences of the vacuum calculations can be masked by this mechanism. For the ¹³C chemical shieldings, on the other hand, a clear increase in the deviation from experiment (0.13 - 0.51 ppm) can be seen when going from CCl₄ to CDCl₃ data. The sp² carbons are thereby more affected than the sp³ carbons (as expected from the results in Figure 5.2). Again, the differences are not of the order of magnitude expected from the experimental comparison (RMSD of 1.40 ppm, Table 5.2). Here, the compensation by the regression procedure is not as efficient as for ¹H. Part of the reason for this might be that the differences in experimental ¹³C shifts between the two solvents systematically increase towards lower field. This is not the case for the ¹H shifts. The effect of the functional (PBE0 or PBEP86) is negligible in case of the ¹H shifts as noted also recently by Oliveira *et al.*²⁶⁰ For ¹³C, PBEP86 performs better than PBE0 for the values recorded in CCl₄ and worse for the values recorded in CDCl₃. The MAD and max. AD values are given in Table A5.1 and Table A5.2 in the Appendix. Also here, the real trends may be obscured by partial compensation due to the regression procedure.

Table 5.3: Root-mean-square deviation (RMSD) values in ppm of the ^1H and ^{13}C shifts calculated in vacuum with PBE0 or PBEP86 versus the experimental values in CDCl_3 or CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.11	0.10	0.09	1.37	1.47	1.05
PBEP86	CCl_4	0.11	0.11	0.09	1.27	1.27	0.91
PBE0	CDCl_3	0.12	0.11	0.10	1.50	1.69	1.12
PBEP86	CDCl_3	0.12	0.11	0.10	1.78	1.75	0.97

To assess the baseline error of the DFT method, independent of the solvent effect, we can focus the analysis only on the nuclei that showed no or a very small solvent effect in experiment (the non-marked nuclei in Scheme 5.1, 135 ^{13}C shifts and 102 ^1H shifts). Table 5.4 shows that the deviation between calculation and experiment becomes similar between the two solvents for this reduced set as expected, indicating a baseline error of PBEP86 of approximately 1 ppm for ^{13}C chemical shieldings (and 1.3 ppm with PBE0).

Table 5.4: Root-mean-square deviation (RMSD) values in ppm of the reduced or complete set of ^{13}C shieldings calculated in vacuum with PBE0 or PBEP86 versus the experimental values in CDCl_3 or CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3). The reduced set consists of the 135 ^{13}C chemical shifts not marked in Scheme 5.1.

DFT Method	Exp. Solvent	^{13}C (reduced set)			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	1.27	1.27	1.04	1.37	1.47	1.05
PBEP86	CCl_4	1.05	0.86	0.88	1.27	1.27	0.91
PBE0	CDCl_3	1.27	1.32	1.08	1.50	1.69	1.12
PBEP86	CDCl_3	1.09	1.06	0.91	1.78	1.75	0.97

5.2.4 Can the Accuracy of the Calculations Be Improved with an Implicit Solvent Model?

Next, we performed DFT calculations of the chemical shieldings with an implicit solvent (conductor-like polarizable continuum model (CPCM)²⁶¹) to explore if the agreement with the experimental data in solution can be improved. Note that the geometry optimizations were not repeated with the implicit solvent. For this, we directly compared the differences between the experimental shifts in the two solvents with the differences between the shielding values before conversion to chemical shifts (without using Eq. (5.3)), such that we avoid obscuring specific effects only present for certain functional groups by the regression procedure. Especially for ^{13}C , it is observed that CPCM cannot account for the observed experimental changes in chemical shift between CCl_4 and CDCl_3 (Figure 5.3). While the picture is less clear for protons, the largest deficits of the CPCM model for ^{13}C can again be seen for polar, hydrogen-bond accepting functional group,

especially carbonyls. By using the regression procedure in Eq. (5.3), shortcomings of the solvation model to reproduce specific interaction with the solvent will be distributed to all functional groups, also the ones not engaging in interactions with the solvent. In the worst case, selective deficiencies of the solvation model can lead to higher inaccuracies in chemical shift prediction for all atoms. After conversion to chemical shifts, the RMSDs from the experimental values in the two solvents were again investigated. While the agreement with experimental ^1H shifts generally improves slightly compared to the vacuum calculations (0.01 - 0.03 ppm), the use of an implicit solvent model only leads to an improvement for the CDCl_3 values calculated with PBEP86. Table 5.5 summarizes the performance of the two DFT methods with the corresponding implicit solvent. The MAD and max. AD values are given in Table A5.4 and Table A5.5 in the Appendix. As the limitations of the CPCM model are expected to be more severe for chloroform (lack of local hydrogen bonding capacity) than CCl_4 , it is surprising to see that the agreement with experimental data is negatively affected by CPCM for CCl_4 . Possible reasons for our findings may be the deficiencies in the CPCM implicit solvent model used and/or more favorable error cancellation in the vacuum calculations as well as the fact that the shielding calculations were performed on structures optimized in vacuum.

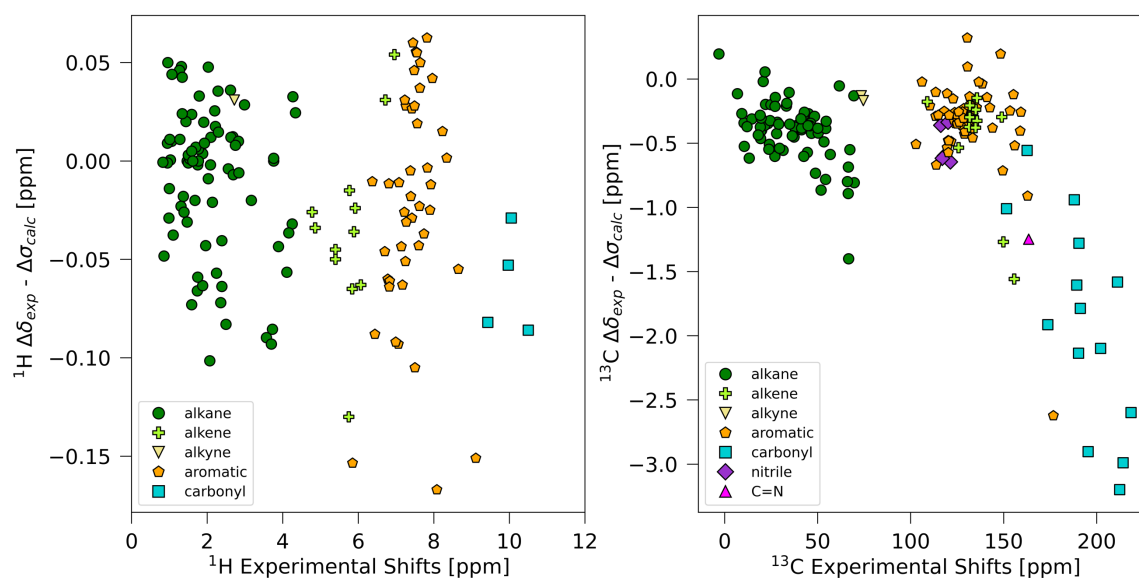


Figure 5.3: Difference between the difference in experimental shifts in CCl_4 and CDCl_3 ($\Delta\delta_{exp}$) and the difference in calculated shieldings with implicit solvent models for CHCl_3 and CCl_4 with PBEO ($\Delta\sigma_{calc}$) plotted against the experimental shifts in CCl_4 . (Left): ^1H . (Right): ^{13}C .

Table 5.5: Root-mean-square deviation (RMSD) values in ppm of the ^1H and ^{13}C shieldings calculated in the corresponding implicit solvent with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.09	0.08	0.08	1.87	1.73	1.09
PBEP86	CCl_4	0.10	0.11	0.07	1.37	1.17	0.92
PBE0	CDCl_3	0.10	0.10	0.08	1.73	1.77	1.14
PBEP86	CDCl_3	0.10	0.12	0.07	1.27	1.22	0.92

5.3 Conclusion

In this study, ^1H and ^{13}C chemical shifts of 35 small, rigid organic molecules were measured in CDCl_3 and in CCl_4 under standardized conditions. In total, 141 ^1H and 170 ^{13}C unambiguously assigned chemical shifts were obtained. This reference data is intended for the calibration and assessment of chemical shift calculations in organic chemistry, particularly with respect to the treatment of solvent-solute interactions in CDCl_3 .

Our experimental data show that specific interactions with the solute are present even in such apolar solvents. Especially the ^{13}C shifts of carbonyl and nitril groups are differentially affected by the two solvents. A direct comparison of calculated shieldings with chemical shifts recorded in CDCl_3 and CCl_4 implies that the accuracy of DFT in implicit solvent differs considerably between functional groups. The likely reason are specific interactions with the solvent (e.g., H-bonds) that cannot be adequately described by the solvent model. By converting the shieldings to chemical shifts using the common multi-standard regression method, the errors resulting from specific shortcomings of the model are redistributed over the whole shift range. This effect may have added to the ambiguous outcome found in this study regarding the usefulness of implicit CDCl_3 . For ^1H , there was no significant advantage from using an implicit solvent model. For ^{13}C , the double-hybrid functional PBE86 with implicit solvent provided the best agreement with chemical shifts recorded in CDCl_3 , while the implicit solvent decreased the performance of PBE0. The double-hybrid functional significantly increases the computation time (roughly by a factor of 15 with the chosen basis set), but the improved accuracy might be beneficial for certain applications, for example to discriminate between diastereomers. Importantly, our data also imply that the explicit treatment of solute-solvent interactions will be necessary for an even more accurate chemical shift prediction. As in most benchmark studies, we optimized the geometries in vacuum. For future studies, it would be worth comparing our findings with NMR shieldings calculated with structures that considered an implicit solvent model during geometry optimization.

5.4 Method Section

Experimental Details

The molecules included in the test set were a selection of small, rigid organic compounds containing different functional groups and consist only of H, C, N and O atoms. For each compound 10 mM solutions were prepared in chloroform-d (Apollo Scientific) and carbon tetrachloride (Sigma-Aldrich). Chloroform was previously filtered through a short column of aluminum oxide (EcoChrom, MP Alumina N, Akt. I) and was stored over molecular sieve (3 Å, Dr. Bender & Dr. Hobein AG) in the fridge. TMS was added as internal standard. NMR spectra were recorded on a Bruker AVANCE III 600 MHz spectrometer equipped with a helium-cooled CPDCH cryogenic probe with z-gradients at 25.0 °C. For each compound, a ¹H (32 scans, 16.0 ppm spectral width with 96152 fid points) and a ¹³C spectrum (512 scans, 248.5 ppm spectral width with 157890 fid points) were recorded. It was possible to shim the tetrachloromethane samples on the proton signal of TMS using the following TopSpin command:

```
topshim 1h rga lockoff o1p=0.2 selid=0.5 durmax=120
```

If peaks could not be assigned unambiguously, ¹³C-HSQC, ¹³C-HMBC, DQF-COSY, NOESY and PSYCHE²⁶² spectra were recorded as needed. Processing of the spectra was done with Bruker TopSpin™ version 4.1 (Bruker Biospin AG) and MestreNova 14.1 (Mestrelab Research). All spectra were referenced to the signal of TMS.

Computational Details

3D structures of the compounds were generated from SMILES strings using RDKit²⁶³ and a conformational search was performed. Atoms were reordered such that the hydrogen atoms follow directly to the bound heavy atom and such that magnetically equivalent groups have subsequent numbers. Hydrogens of CH₂ groups were ordered such that the proR hydrogen is first. The found structures were optimized with DFT in vacuum using Orca 5.0.1^{248–250} at the BP86/def2-tzvp^{255–257} level using the resolution of identity approximation with def2/J²⁵⁸ as auxiliary basis set and Grimme's dispersion correction D3BJ.^{264,265} Minima were verified by a frequency calculation at the same level of theory. In case of imaginary frequencies, the geometry at the most displaced point along the corresponding mode was taken as input for a new structure optimization. Only molecules with one dominant conformation were considered for the dataset. Next, NMR chemical shieldings were computed with the GIAO²⁶⁶ approach using either the hybrid GGA functional PBE0²⁶⁷ with the cc-pVTZ basis set²⁵² using cc-pVTZ/JK auxiliary basis set²⁶⁸ or the 2013 version of the dispersion corrected, spin-component scaled double-hybrid functional PBEP86²⁵³ together with the pcSseg-3 basis set²⁵⁴ with auxiliary basis sets def2/J and cc-pwCVQZ/C.^{269,270} The

resolution of identity approximation for both Coulomb and HF exchange integrals was applied for the hybrid functional (RIJK) whereas for the double-hybrid functional resolution of identity was used for the Coulomb integrals and numerical chain-of-sphere integration for the HF exchange integrals (RIJCOSX). For both, D3BJ corrections were applied. Besides the calculation in vacuum, chemical shieldings were also calculated using CPCM as an implicit solvent for chloroform and tetrachloromethane (without re-optimization of the geometry).

Analysis of the data was done with a Python script²⁰¹ in a Jupyter Notebook²⁰² and the functionalities of the matplotlib,²⁰³ nglview,²⁷¹ numpy,²⁰⁵ openbabel,²⁷² and scipy²⁰⁷ packages were used.

5.5 Appendix

Collection of Chemical Shifts From the Literature

Literature data for comparison with our measured chemical shifts in CDCl₃ were collected from different sources with at maximum ten per compound (Figure 5.1). For the literature search, publications between the period of 1980 and 2022 were considered that were measured in CDCl₃ using mainly the data found with Reaxys.^{13,273-505}

Effect of Solute-Solvent Interaction in Experiment: Example for Compounds 26 and 34

By comparison of the chemical shifts measured in CDCl₃ and CCl₄, directed solvent effects can be identified. For **26**, the ¹³C chemical shift of the carbonyl carbon, as well as the shift of the conjugated double-bound carbon (Figure A5.1, carbons 1 and 3), move towards higher field when changing the solvent from chloroform-d to tetrachloromethane, whereas the other carbons that are not involved in hydrogen bonding with the solvent, have nearly identical chemical shifts. Also, proton chemical shifts close to the functional groups involved in hydrogen bonds with chloroform (Figure A5.2, protons of carbons 2 and 6 of **26**, α to carbonyl group) change most when comparing shifts measured in CDCl₃ and in CCl₄. Also, for the ¹³C shifts of **34**, the same behavior can be observed (Figure A5.3). Having a look at the two given examples, where the order of the chemical shifts changed based on the solvent (¹H of **26** and ¹³C of **34**, Figure A5.4) and comparing them with the calculated shieldings, one can clearly see that the order is reproduced in tetrachloromethane whereas it is not when chloroform was used as an implicit solvent. It becomes evident that the hydrogen bond donor capabilities of chloroform cannot be neglected for chemical shielding calculations and the implicit solvent model cannot properly account for the directed interactions.

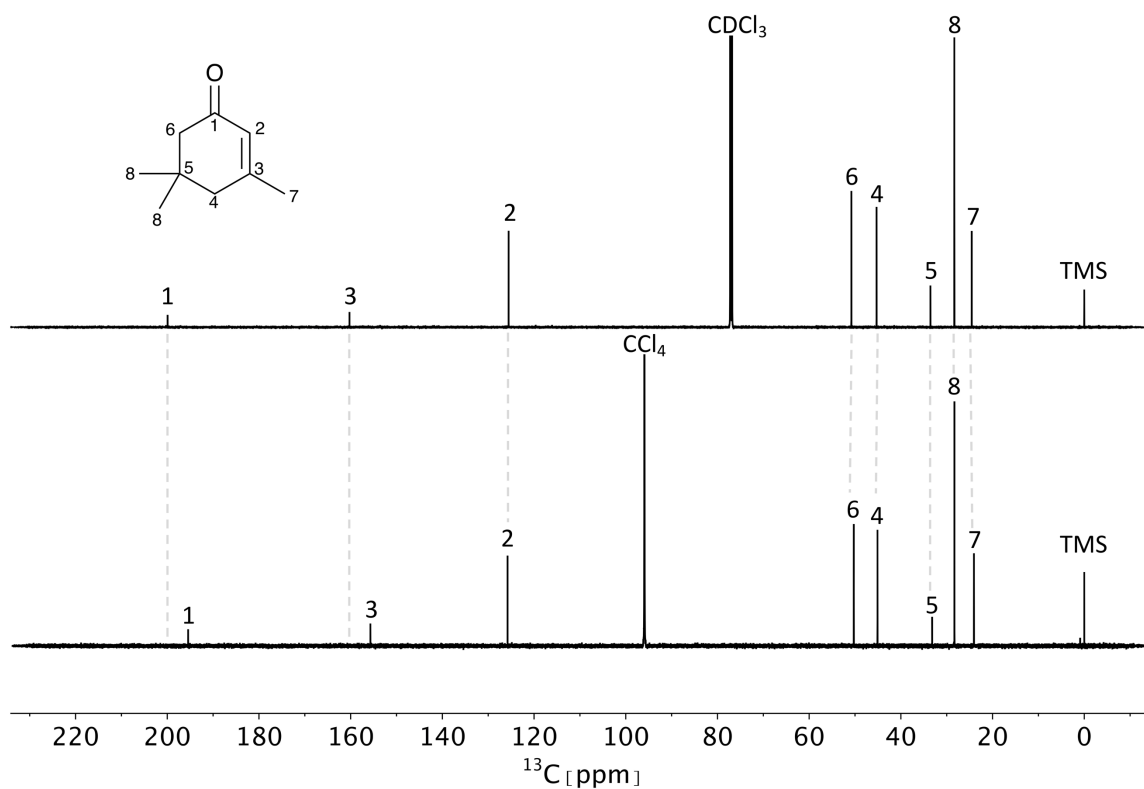


Figure A5.1: ^{13}C spectra of **26** in chloroform- d (top) and tetrachloromethane (bottom) referenced to internal TMS. Dashed grey lines are there to help visualizing the chemical shift differences between the two solvents.

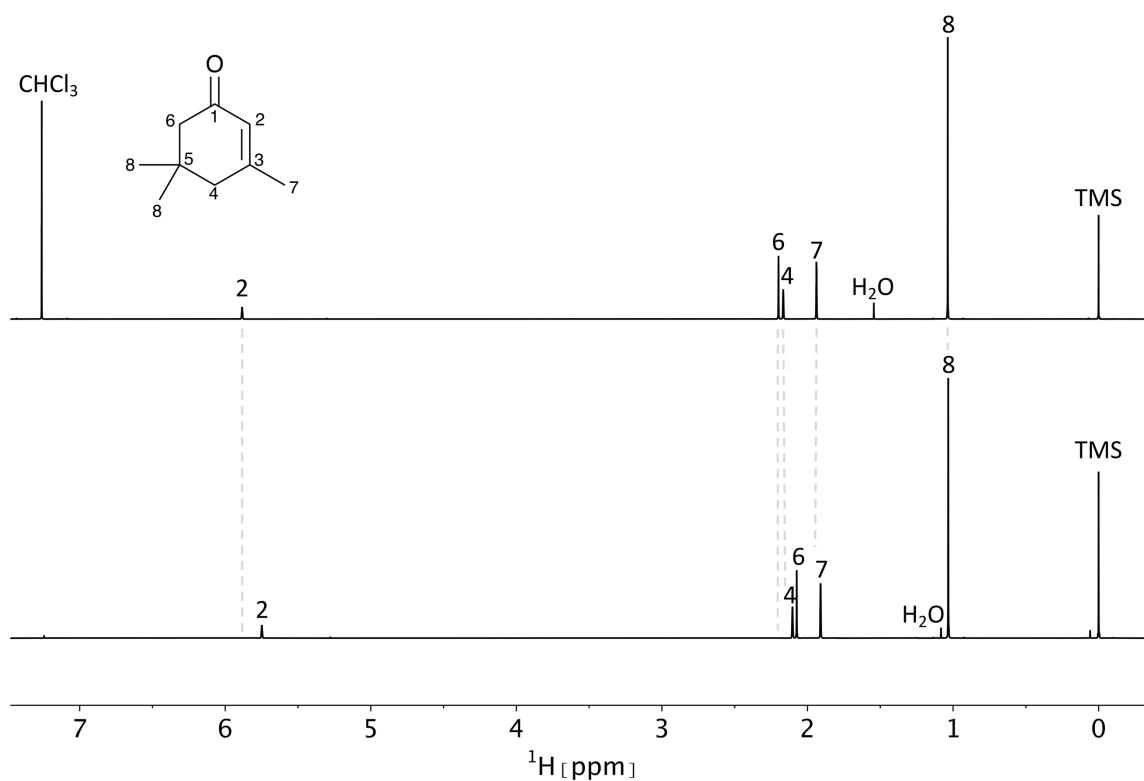


Figure A5.2: ^1H spectra of **26** in chloroform- d (top) and tetrachloromethane (bottom) referenced to internal TMS. Dashed grey lines are there to help visualizing the chemical shift differences between the two solvents.

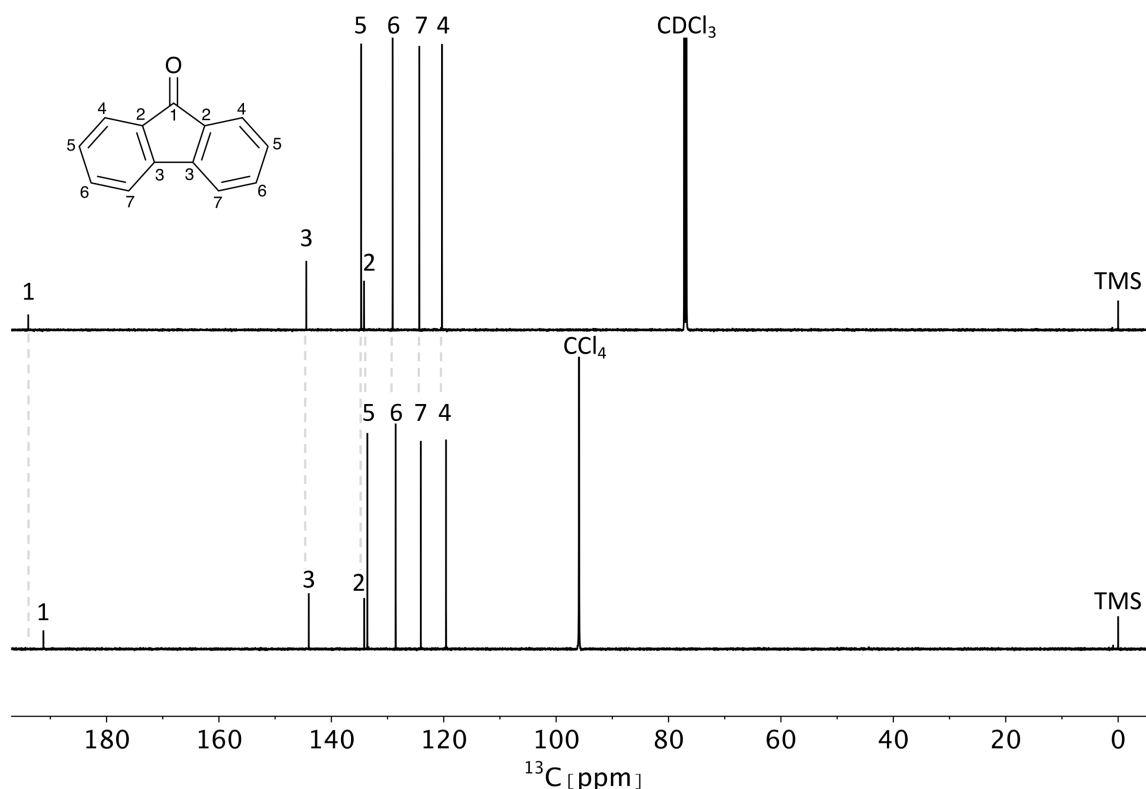


Figure A5.3: ^{13}C spectra of **34** in chloroform-*d* (top) and tetrachloromethane (bottom) referenced to internal TMS. Dashed grey lines are there to help visualizing the chemical shift differences between the two solvents. Biggest change is again observed for the carbonyl carbon 1. Shifts of carbon 5 and carbon 2 change positions.

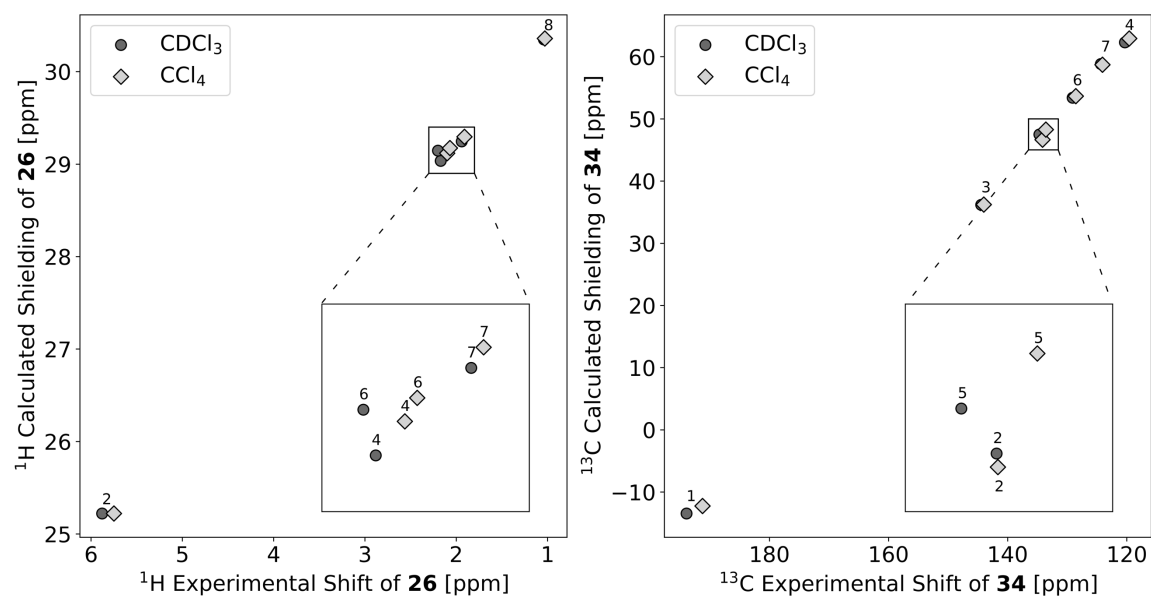


Figure A5.4: Comparison of experimental ^1H shift of **26** with the calculated shieldings using the PBE0 functional (left). The order of the shieldings using the implicit chloroform does not agree with the experiment (see numbers 4 and 6, shifts should be ordered from left bottom to top right), whereas it does when compare experimental shifts and calculated shieldings in tetrachloromethane. This is also true for the PBEP86 functional (carbons 2 and 5). The same behavior can also be observed for the comparison of experimental ^{13}C shifts of **34** with the calculated shieldings using the PBEP86 functional (right). Note that the trend was not reproduced correctly for CCl_4 when using the cheaper PBE0 functional. Numbers correspond to the ones given in Figure A5.2 and Figure A5.3, respectively.

Additional Figures and Tables:

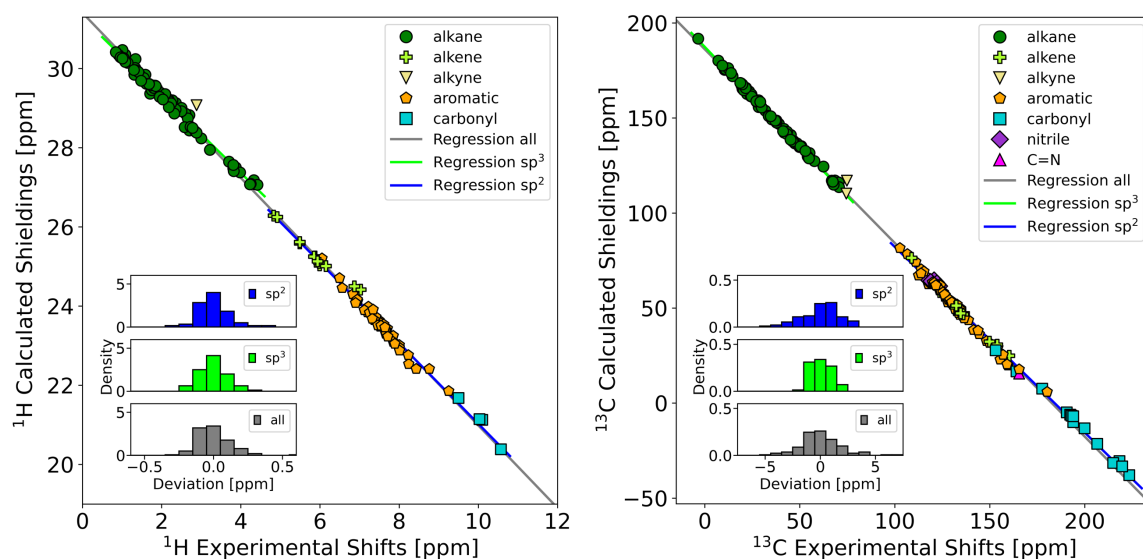


Figure A5.5: Regression of the calculated chemical shieldings (PBEP86, in vacuum) with the experimental chemical shifts measured in CDCl_3 . (Left): ^1H . (Right): ^{13}C . The histograms show the deviations after conversion of the shieldings into chemical shifts using the parameters from the regression.

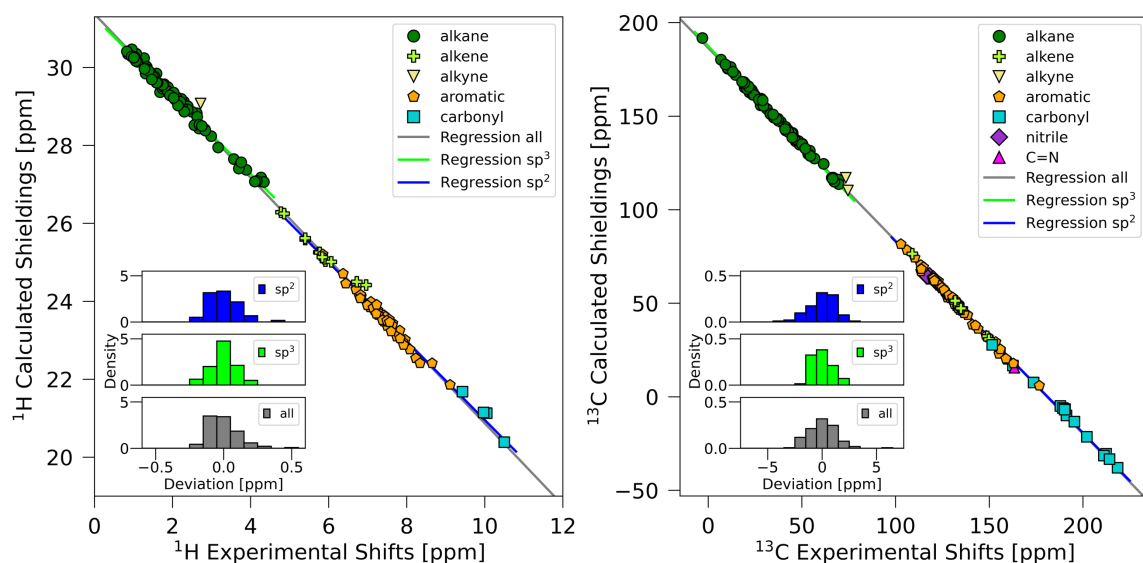


Figure A5.6: Regression of the calculated chemical shieldings (PBEP86, in vacuum) with the experimental chemical shifts measured in CCl_4 . (Left): ^1H . (Right): ^{13}C . The histograms show the deviations after conversion of the shieldings into chemical shifts using the parameters from the regression.

Table A5.1: Mean absolute deviation (MAD) values in ppm of the ^1H and ^{13}C shieldings calculated in vacuum with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.09	0.08	0.07	1.05	1.08	0.83
PBEP86	CCl_4	0.09	0.08	0.07	0.98	0.99	0.73
PBE0	CDCl_3	0.09	0.08	0.08	1.14	1.32	0.88
PBEP86	CDCl_3	0.09	0.08	0.08	1.32	1.40	0.80

Table A5.2: Maximum absolute deviation (max. AD) values in ppm of the ^1H and ^{13}C shieldings calculated in vacuum with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.32	0.34	0.23	5.10	5.04	2.56
PBEP86	CCl_4	0.53	0.38	0.23	6.10	3.52	2.27
PBE0	CDCl_3	0.43	0.31	0.26	5.17	4.65	2.82
PBEP86	CDCl_3	0.64	0.36	0.28	7.00	5.22	2.40

Table A5.3: Intercept values in ppm and slope values of the linear regression of ^1H and ^{13}C shieldings calculated in vacuum with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	31.476	31.176	31.371	185.47	187.75	185.79
		-1.062	-1.024	-1.006	-1.033	-1.048	-1.047
PBEP86	CCl_4	31.374	31.067	31.285	186.51	187.56	185.41
		-1.051	-1.011	-1.004	-1.031	-1.062	-1.023
PBE0	CDCl_3	31.508	31.316	31.368	185.14	182.17	185.67
		-1.053	-1.030	-0.983	-1.021	-1.001	-1.032
PBEP86	CDCl_3	31.405	31.209	31.281	186.17	179.88	187.43
		-1.042	-1.018	-0.980	-1.019	0.977	-1.047

Table A5.4: Mean absolute deviation (MAD) values in ppm of the ^1H and ^{13}C shieldings calculated in the corresponding implicit solvent with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.07	0.07	0.06	1.45	1.17	0.87
PBEP86	CCl_4	0.07	0.08	0.05	1.11	0.84	0.74
PBE0	CDCl_3	0.08	0.08	0.06	1.34	1.25	0.92
PBEP86	CDCl_3	0.09	0.10	0.05	1.01	0.93	0.75

Table A5.5: Maximum absolute deviation (max. AD) values in ppm of the ^1H and ^{13}C shieldings calculated in the corresponding implicit solvent with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	0.32	0.34	0.23	8.30	7.99	2.58
PBEP86	CCl_4	0.40	0.28	0.19	4.35	4.17	2.13
PBE0	CDCl_3	0.25	0.21	0.20	7.72	7.57	2.86
PBEP86	CDCl_3	0.38	0.30	0.19	4.18	3.46	2.11

Table A5.6: Intercept values in ppm and slope values of the linear regression of ^1H and ^{13}C shieldings calculated in the corresponding implicit solvent with PBE0 or PBEP86 versus the experimental values in CDCl_3 and CCl_4 . Values are given for the complete carbon set (all), for only the sp^2 carbons (sp^2) and for only sp^3 carbons (sp^3).

DFT Method	Exp. Solvent	^1H			^{13}C		
		all	sp^2	sp^3	all	sp^2	sp^3
PBE0	CCl_4	31.417	30.948	31.350	186.27	193.20	186.41
		-1.071	-1.010	-1.031	-1.047	-1.093	-1.062
PBEP86	CCl_4	31.317	30.853	31.267	187.37	191.54	188.19
		-1.061	-1.000	-1.030	-1.046	-1.074	-1.078
PBE0	CDCl_3	31.419	30.984	31.338	186.35	190.28	186.057
		-1.020	-1.011	-1.021	-1.042	-1.068	-1.054
PBEP86	CDCl_3	31.320	30.897	31.256	187.48	189.05	188.36
		-1.057	-1.002	-1.020	-1.042	-1.052	-1.072

The ^{13}C and ^1H chemical shifts are listed in the following in the same order as in the xyz files (see below). *Avgn* indicates that the next experimental chemical shift needs to be averaged over n atoms in the xyz file, whereas *ign* indicates that the next shift should be ignored (for exchangeable protons). *Mvgn_a1_a2_...ak* indicates that n atoms per experimental shift needs to be averaged and the indexes $a_1 - a_k$ indicate which atoms needs to be considered (e.g., *Mvg2_1_3_4_2* indicate that shifts need to be averaged in groups of two whereas the shifts of the first and last atom (1 and 2) as well as the second and third atom belong together).

Chemical shifts recorded in CDCl₃

Nitromethane (1)	13C 62.49 1H Avg3 4.33
Nitroethane (2)	13C 12.31 70.47 1H Avg3 1.59 Avg2 4.42
Propionitrile (3)	13C 10.48 10.92 120.69 1H Avg3 1.30 Avg2 2.36
<i>trans</i> -Crotonaldehyde (4)	13C 18.67 153.92 134.68 193.94 1H Avg3 2.03 6.87 6.15 9.50
Isobutyronitrile (5)	13C Avg2 19.95 19.84 123.77 1H Avg6 1.33 2.70
3,3-Dimethyl-1-butene (6)	13C 108.86 149.87 33.66 Avg3 29.16 1H 4.83 4.92 5.85 Avg9 1.01
Methyl propiolate (7)	13C 74.79 74.45 153.14 52.95 1H 2.88 Avg3 3.81
Cyclopropanecarbonitrile (8)	13C 122.17 -3.48 Avg2 7.12 1H 1.34 Mvg2_1_4_2_3 1.08 1.01
gamma-Butyrolactone (9)	13C 177.64 27.80 22.20 68.48 1H Avg2 2.50 Avg2 2.27 Avg2 4.35
Oxolan-3-one (10)	13C 214.99 70.62 66.85 37.07 1H Avg2 3.87 Avg2 4.25 Avg2 2.50
2-Methyl-2-oxazoline (11)	13C 13.82 165.46 54.60 67.40 1H Avg3 1.98 Avg2 3.82 Avg2 4.23
Maleic anhydride (12)	13C Avg2 164.07 Avg2 136.49 1H Avg2 7.03
1,4-Dioxane (13)	13C Avg4 67.11 1H Avg8 3.70
Tetrahydro-4H-pyran-4-one (14)	13C 206.53 Avg2 42.90 Avg2 67.85 1H Avg4 2.50 Avg4 3.98
Nitrobenzene (15)	13C 148.24 Avg2 123.52 Avg2 129.31 134.56 1H Avg2 8.24 Avg2 7.56 7.70
Phenol (16)	13C 155.44 Avg2 115.27 Avg2 129.68 120.83 1H Ign 4.61 Avg2 6.83 Avg2 7.25 6.94
Anisole (17)	13C 55.14 159.55 Avg2 113.89 Avg2 129.45 120.65 1H Avg3 3.81 Avg2 6.91 Avg2 7.30 6.95
Benzaldehyde (18)	13C 192.38 136.44 Avg2 129.76 Avg2 129.01 134.47 1H 10.03 Avg2 7.89 Avg2 7.54 7.64

Benzonitrile (19)	13C 118.85 112.50 Avg2 132.18 Avg2 129.12 132.76 1H Avg2 7.67 Avg2 7.48 7.61
4-Dimethylaminopyridine (20)	13C Avg2 39.03 154.22 Avg2 106.59 Avg2 149.88 1H Avg6 3.00 Avg2 6.49 Avg2 8.23
Pyrimidine (21)	13C 159.11 Avg2 156.94 121.58 1H 9.25 Avg2 8.76 7.34
2,6-Dimethyl-γ-pyrone (22)	13C 180.21 Avg2 113.81 Avg2 165.44 Avg2 19.76 1H Avg2 6.05 Avg6 2.24
Norbornene (23)	13C Avg2 41.74 48.51 Avg2 135.36 Avg2 24.59 1H Avg2 2.85 1.31 1.08 Avg2 5.99 Mvg2_1_3_2_4 0.95 1.61
Norcamphor (24)	13C 218.30 49.87 24.20 27.20 35.33 37.71 45.27 1H 2.60 1.53 1.82 1.80 1.44 2.66 1.55 1.73 2.05 1.84
Fenchone (25)	13C 23.37 47.40 21.71 45.35 41.67 24.96 31.85 54.16 14.63 223.41 1H Avg3 1.03 Avg3 1.03 2.14 1.53 1.79 1.71 1.79 1.39 1.56 Avg3 1.14
Isophorone (26)	13C 199.93 125.55 160.26 45.30 33.56 Avg2 28.33 50.79 24.55 1H 5.88 Avg2 2.17 Avg6 1.04 Avg2 2.20 Avg3 1.94
4-Cyanobenzaldehyde (27)	13C 190.56 138.75 Avg2 129.90 Avg2 132.92 117.66 117.70 1H 10.10 Avg2 8.00 Avg2 7.85
Mesitaldehyde (28)	13C Avg2 20.51 Avg2 141.50 Avg2 130.53 143.83 21.48 130.01 193.00 1H Avg6 2.58 Avg2 6.90 Avg3 2.32 10.57
Camphor (29)	13C 9.26 57.73 219.72 43.33 43.07 46.81 19.80 19.16 27.07 29.94 1H Avg3 0.91 2.35 1.85 2.09 Avg3 0.84 Avg3 0.96 1.95 1.34 1.68 1.41
Indole (30)	13C 127.86 135.78 124.07 102.67 110.98 122.00 119.82 120.73 1H lgn 8.14 7.22 6.56 7.41 7.20 7.12 7.65
Adamantane (31)	13C Avg4 28.34 Avg6 37.76 1H Avg4 1.88 Avg12 1.75
Dicyclopentadiene (32)	13C 54.80 132.03 132.06 34.69 41.19 46.20 132.42 136.02 45.18 50.35 1H 3.22 5.49 5.49 1.68 2.18 2.73 2.88 5.98 5.94 2.78 1.30 1.48
1-Methylnaphthalene (33)	13C 19.38 134.26 132.60 124.10 125.70 125.53 128.51 133.54 126.36 125.56 126.55 1H Avg3 2.70 8.00 7.52 7.48 7.85 7.71 7.37 7.32
9-Fluorenone (34)	13C 193.93 Avg2 134.18 Avg2 144.45 Avg2 124.35 Avg2 129.09 Avg2 134.69 Avg2 120.31 1H Avg2 7.66 Avg2 7.30 Avg2 7.49 Avg2 7.53
Anthracene (35)	13C Avg4 131.69 Avg4 128.17 Avg4 125.34 Avg2 126.22 1H Avg4 8.01 Avg4 7.46 Avg2 8.43

Chemical shifts recorded in CCl₄

Nitromethane (1)	¹³ C 61.63 1H Avg3 4.27
Nitroethane (2)	¹³ C 12.04 69.46 1H Avg3 1.59 Avg2 4.34
Propionitrile (3)	¹³ C 10.43 10.54 118.13 1H Avg3 1.32 Avg2 2.29
<i>trans</i> -Crotonaldehyde (4)	¹³ C 18.12 149.90 134.92 190.26 1H Avg3 2.03 6.72 6.07 9.43
Isobutyronitrile (5)	¹³ C Avg2 19.88 19.44 121.32 1H Avg6 1.34 2.62
3,3-Dimethyl-1-butene (6)	¹³ C 108.99 149.01 33.41 Avg3 29.05 1H 4.78 4.86 5.77 Avg9 1.01
Methyl propiolate (7)	¹³ C 73.42 74.58 151.59 51.86 1H 2.72 Avg3 3.76
Cyclopropanecarbonitrile (8)	¹³ C 120.07 -3.06 Avg2 6.87 1H 1.27 Mvg2_1_4_2_3 1.07 0.96
gamma-Butyrolactone (9)	¹³ C 173.73 27.00 22.02 66.69 1H Avg2 2.36 Avg2 2.22 Avg2 4.25
Oxolan-3-one (10)	¹³ C 211.22 69.68 65.98 36.35 1H Avg2 3.73 Avg2 4.16 Avg2 2.39
2-Methyl-2-oxazoline (11)	¹³ C 13.22 163.42 54.36 66.45 1H Avg3 1.88 Avg2 3.70 Avg2 4.11
Maleic anhydride (12)	¹³ C Avg2 162.66 Avg2 135.66 1H Avg2 6.96
1,4-Dioxane (13)	¹³ C Avg4 66.49 1H Avg8 3.57
Tetrahydro-4H-pyran-4-one (14)	¹³ C 202.14 Avg2 42.50 Avg2 67.34 1H Avg4 2.39 Avg4 3.89
Nitrobenzene (15)	¹³ C 148.34 Avg2 123.20 Avg2 128.68 133.40 1H Avg2 8.23 Avg2 7.53 7.64
Phenol (16)	¹³ C 155.24 Avg2 114.87 Avg2 129.15 120.27 1H Ign 4.27 Avg2 6.70 Avg2 7.14 6.82
Anisole (17)	¹³ C 54.26 159.23 Avg2 113.49 Avg2 128.88 120.16 1H Avg3 3.76 Avg2 6.78 Avg2 7.17 6.83
Benzaldehyde (18)	¹³ C 189.32 136.58 Avg2 129.24 Avg2 128.45 133.36 1H 9.97 Avg2 7.83 Avg2 7.49 7.56

Benzonitrile (19)	13C 116.98 113.48 Avg2 131.65 Avg2 128.58 131.65 1H Avg2 7.63 Avg2 7.45 7.55
4-Dimethylaminopyridine (20)	13C Avg2 38.79 153.43 Avg2 106.11 Avg2 149.51 1H Avg6 2.99 Avg2 6.37 Avg2 8.08
Pyrimidine (21)	13C 158.87 Avg2 155.96 120.68 1H 9.11 Avg2 8.65 7.23
2,6-Dimethyl-γ-pyrone (22)	13C 176.71 Avg2 113.77 Avg2 162.85 Avg2 19.34 1H Avg2 5.85 Avg6 2.20
Norbornene (23)	13C Avg2 41.41 48.30 Avg2 134.96 Avg2 24.44 1H Avg2 2.83 1.31 1.06 Avg2 5.92 Mvg2_1_3_2_4 0.95 1.59
Norcamphor (24)	13C 212.40 48.79 23.76 27.24 34.96 37.29 44.47 1H 2.50 1.54 1.77 1.77 1.44 2.63 1.50 1.70 1.96 1.74
Fenchone (25)	13C 23.25 46.70 21.55 45.06 41.41 24.90 31.40 53.41 14.61 218.52 1H Avg3 1.00 Avg3 0.99 2.09 1.47 1.75 1.68 1.79 1.37 1.50 Avg3 1.10
Isophorone (26)	13C 195.45 125.78 155.69 45.10 33.18 Avg2 28.31 50.27 24.06 1H 5.75 Avg2 2.10 Avg6 1.03 Avg2 2.07 Avg3 1.91
4-Cyanobenzaldehyde (27)	13C 188.01 138.45 Avg2 129.31 Avg2 132.26 118.01 116.21 1H 10.05 Avg2 7.96 Avg2 7.82
Mesitaldehyde (28)	13C Avg2 20.73 Avg2 141.05 Avg2 130.55 142.61 21.67 130.45 190.51 1H Avg6 2.56 Avg2 6.81 Avg3 2.30 10.50
Camphor (29)	13C 9.21 56.75 214.26 42.56 42.85 46.25 19.66 19.11 27.05 29.51 1H Avg3 0.86 2.25 1.75 2.03 Avg3 0.83 Avg3 0.95 1.93 1.34 1.62 1.39
Indole (30)	13C 127.64 135.47 122.84 102.86 110.28 121.74 119.57 120.50 1H Ign 7.91 7.08 6.44 7.25 7.06 6.99 7.50
Adamantane (31)	13C Avg4 27.98 Avg6 37.54 1H Avg4 1.88 Avg12 1.75
Dicyclopentadiene (32)	13C 54.54 131.48 131.74 34.69 40.99 45.86 131.86 135.69 44.87 50.06 1H 3.17 5.40 5.40 1.59 2.14 2.69 2.84 5.89 5.84 2.75 1.28 1.47
1-Methylnaphthalene (33)	13C 19.20 133.32 132.43 123.64 125.20 124.98 128.19 133.36 126.16 125.02 126.13 1H Avg3 2.68 7.90 7.42 7.39 7.74 7.60 7.27 7.22
9-Fluorenone (34)	13C 191.23 Avg2 134.16 Avg2 144.02 Avg2 124.08 Avg2 128.56 Avg2 133.59 Avg2 119.59 1H Avg2 7.62 Avg2 7.26 Avg2 7.42 Avg2 7.48
Anthracene (35)	13C Avg4 131.46 Avg4 127.86 Avg4 124.81 Avg2 125.88 1H Avg4 7.92 Avg4 7.38 Avg2 8.34

Optimized Coordinates of Molecules 1 – 35 Obtained With BP86/def2-tzvp in Vacuum

Nitromethane (1)

C	-0.62681479835640	-0.10338397388413	-0.09594534415716
H	-1.06475240570623	-0.17071521012912	0.90710471825263
H	-1.02762224265049	0.75950702999940	-0.63314190246179
H	-0.77001166026217	-1.04439927624854	-0.63342745110327
N	0.85277892047328	0.10457877546916	0.09637574377375
O	1.53986730163567	-0.90277222680733	0.26088434978263
O	1.26077088486632	1.26531388160058	0.10314688591322

Nitroethane (2)

C	-0.96696768534669	0.44305206916225	0.42801343554144
H	-0.66978856896510	0.47499099340534	1.48309292504135
H	-0.91655175367615	1.46248213595372	0.02700050034567
H	-2.00909961572277	0.09769157193784	0.37024638570311
C	-0.09510358751177	-0.50617805223681	-0.36342538055766
H	-0.11038994366408	-1.53378445482614	0.02284007796427
H	-0.35618083330943	-0.54948530126594	-1.42893508393147
N	1.38303733105249	-0.12355593890022	-0.35402271846562
O	1.72313133930475	0.87832799100153	0.27184083815888
O	2.13439331783875	-0.86432101423155	-0.98695497979996

Propionitrile (3)

C	-1.00685867612622	-0.26917392358902	0.00585853133655
H	-1.10871109372847	-0.73788179107870	0.99280360943837
H	-1.88316461082083	0.36842994838212	-0.16929767857332
H	-0.99681733805787	-1.06331233421308	-0.75128896176642
C	0.27622392115186	0.57670219501005	-0.06968060089700
H	0.36359957062600	1.06172307204197	-1.05459491256256
H	0.25246757376792	1.38466887377355	0.67831406548948
C	1.48304037049501	-0.21519880700289	0.15448526651329
N	2.43033628269259	-0.86577623332399	0.33141368102161

trans-Crotonaldehyde (4)

C	-1.75503461869207	-0.13644965430809	-0.01648719859462
H	-2.29422554402659	0.11122258577790	0.91248284205652
H	-1.94990764516937	0.65052618401412	-0.75613004486695
H	-2.19371110395776	-1.07968849012756	-0.38103999602575
C	-0.29913796390325	-0.30408687875799	0.25103146972298
H	-0.01986527584849	-1.07385403910750	0.98202451391028
C	0.69544776163276	0.40010606739442	-0.32033453957603
H	0.50130296676286	1.18415576560389	-1.05770840263235
C	2.09524580585136	0.14082840545298	0.02739903370068
H	2.23856997641628	-0.67318886212270	0.78809329267097
O	3.05321864093427	0.73393991618052	-0.44381897036572

Isobutyronitrile (5)

C	-1.09081610024817	-0.54902631910864	-0.60725977558951
H	-2.09980845963800	-0.53354263703710	-0.17610562350632
H	-1.10161756377965	0.04373772729499	-1.53200776144079
H	-0.83480370297587	-1.58639601290275	-0.86260257582979
C	1.37709150425265	-0.02226865053078	-0.19636977963813
H	2.10839853508966	0.36465110133176	0.52452446826877
H	1.64772993097083	-1.05644652057442	-0.44944719004574
H	1.43739510874756	0.58578826886683	-1.10918589568446
C	-0.05042999591797	0.01025019861822	0.38140854830145
H	-0.07258933819039	-0.60212257304712	1.29930918496698
C	-0.40862208630554	1.37390586668810	0.78416246087158
N	-0.69096083200511	2.46126555040091	1.08534493932595

3,3-Dimethylbut-1-ene (6)

C	1.87801532798471	1.20032206558087	-0.51523798527491
H	2.39688609701146	2.15669364914985	-0.59121580062901
H	2.44606291552518	0.31249914206103	-0.79748799807006
C	0.61173311736926	1.13391731903371	-0.09631188818589
H	0.09952844692424	2.06655274501715	0.17186831599941
C	-0.24196381011025	-0.10529808191836	0.06359422785347
C	-1.48721229563156	0.04969786056125	-0.83533243576005
H	-2.04394743757170	0.96589134638238	-0.58766647134789
H	-1.20058065468589	0.10297469237370	-1.89533284016335
H	-2.16706041895342	-0.80565863162688	-0.70284137163536
C	0.50855114542277	-1.38743970545845	-0.31692077576837
H	1.39555085704099	-1.53356010435832	0.31628484206945
H	-0.14591596638262	-2.26163816791868	-0.18759754200160
H	0.83581273473210	-1.36387263041545	-1.36649950103489
C	-0.69270418099616	-0.19090223017672	1.53773858792890
H	-1.36506105541205	-1.04963503844818	1.68571641785399
H	0.17293615612879	-0.31069883660434	2.20480417611486
H	-1.23216197839584	0.71884860676545	1.84073004205129

Methyl propiolate (7)

C	2.86170585055193	0.59474535093513	0.40825192625615
H	3.86083588765202	0.97010786809656	0.49542552952162
C	1.73187127243925	0.17156644142645	0.30537158821739
C	0.41834160944694	-0.42295751610995	0.19387596804153
O	-0.54071306751649	0.53332067489002	0.12382373381088
C	-1.88879066622272	0.01789115186835	0.01084425653744
H	-2.13349402001418	-0.60428006962400	0.88131791104294
H	-2.53114779924528	0.90232938468250	-0.03032410523636
H	-1.99151219066673	-0.58396035974946	-0.90130027016248
O	0.21082812357526	-1.62138392641557	0.16719746197087

Cyclopropanecarbonitrile (8)

N	2.85332611386043	-0.21089722816028	-0.22534161471361
C	1.74320841681596	-0.10297607283625	0.11017335537180
C	0.37299486904908	0.03171638069300	0.51236847117743
H	0.23228669437201	0.17102421208877	1.58484576944164
C	-0.70344570861079	-0.74776195228342	-0.23346113499028
H	-0.37052609201274	-1.38622688789919	-1.04976199226697
H	-1.51306779001485	-1.15370387640171	0.37103312793555
C	-0.61017756830057	0.73699167880322	-0.41422798541130
H	-1.35428474702151	1.37225543182025	0.06369098366953
H	-0.21325818813703	1.11588531417560	-1.35427798021377

 γ -Butyrolactone (9)

O	2.16942535470236	1.51463827727151	0.13056071743181
C	1.13426198472735	0.90628859034833	0.02730440068188
C	0.91961932801328	-0.58322095162206	-0.22670452993746
H	1.10408844971854	-0.76797008803445	-1.29719111574795
H	1.64914945007607	-1.17171078990756	0.34052365546343
C	-0.54388783623286	-0.81219223257420	0.15597860009165
H	-0.62482361021030	-1.07081500382034	1.22127963457386
H	-1.03944234001738	-1.60015612441591	-0.42459576233340
C	-1.15984106481215	0.57092072921604	-0.10034860104259
H	-1.97950301983536	0.82854675185720	0.58104741900129
H	-1.50795614842440	0.68383175763602	-1.13969419360708
O	-0.09103454770517	1.52564108404544	0.12383977542455

Oxolan-3-one (10)

O	0.55649342518305	2.26505548321312	0.38680980916114
C	0.32848710183269	1.09828912403726	0.15240202189764
C	1.34155116795781	-0.04978035831213	0.17737390157391
H	1.66505911639681	-0.21077671658887	1.22714611505822
H	2.22926944280455	0.17307302760559	-0.43028856831071
O	0.66883277867199	-1.18577726968058	-0.35519458151607
C	-0.73394233788709	-1.02832794063380	-0.04651047095007
H	-1.28899636698382	-1.67565982024770	-0.73544527917610
H	-0.93381363359855	-1.35751652945331	0.99119204996097
C	-1.01701105891593	0.46476418092388	-0.21008288516975
H	-1.24046353612780	0.72028723552310	-1.25831240027247
H	-1.82063809933372	0.86656958361344	0.41931028774327

2-Methyl-2-oxozaline (11)

C	2.15663718311260	-0.10140301552173	0.03665745019527
H	2.47968476957124	-0.60930060985291	0.95669303346582
H	2.58886354756325	-0.60462165569919	-0.83344820206060
H	2.52459385594799	0.93340052919912	0.08690348283198
C	0.67136714374403	-0.11570662474688	-0.06567577977494
N	-0.02276652158928	-0.62739478678949	-1.00896570070993
C	-1.43746842704607	-0.39407465076187	-0.66907168486305
H	-1.92021828150350	0.18555610359606	-1.47029591850657
H	-1.96490636698039	-1.35731953260693	-0.59839288640260
C	-1.40650500366645	0.36806359357739	0.68110094687305
H	-1.87844498146579	-0.17821034850049	1.50923373648674
H	-1.83069141176166	1.38013726928157	0.62814154650087
O	0.01555849407401	0.49365772882535	0.98012297596395

Maleic anhydride (12)

O	2.25180399190854	-1.07202341545776	0.00153062242008
C	1.13670450739468	-0.62653971962326	-0.00036255743090
C	-1.13870682688536	-0.62216242663804	-0.01691747890306
O	-0.00260535030570	-1.44920424653960	-0.02017587365135
C	0.67113879819006	0.78844531705918	0.01596225888820
H	1.36644343753657	1.62364378183767	0.03268138809353
C	-0.66793008068948	0.79104853147031	0.00612941850290
H	-1.36010249478928	1.62898433492929	0.01266885537958
O	-2.25554798236003	-1.06309215703781	-0.03152263329898

1,4-Dioxane (13)

O	0.83123727493900	0.64146910923991	-0.94205414005097
C	0.00247398829601	1.36991234689665	-0.03033525342119
H	0.57979359142228	1.62496107084700	0.88071969703471
H	-0.28510731008728	2.29930900354286	-0.54241100129126
C	-1.22481219388266	0.55533542173732	0.35377024830138
H	-1.84984911490983	0.37878358977326	-0.54414099600024
H	-1.82598495254671	1.08001935380733	1.11011619863233
C	-0.01451719284815	-1.42526284878955	0.02105391323473
H	0.27303302023296	-2.35466896253793	0.53312799591912
H	-0.59176034103167	-1.68032076368617	-0.89005799354922
C	1.21279533979555	-0.61066526172569	-0.36309405661047
H	1.83786577365969	-0.43423330478877	0.53482798605432
H	1.81396737222487	-1.13530901000052	-1.11947078146112
O	-0.84325425526403	-0.69684174431572	0.93275318320791

Tetrahydro-4H-pyran-4-one (14)

O	-0.11177185582409	2.38925241560384	-0.70802105894980
C	-0.03348477945725	1.32373323844633	-0.11887278800103
C	-1.24842693835919	0.51621585428084	0.30269797471077
H	-1.32636187301306	0.55156856454400	1.40231174184772
H	-2.15363641165976	0.96378291927265	-0.12930621872976
C	1.28812115974853	0.65682437650936	0.21967969152911
H	1.43326847057039	0.70562141479376	1.31193133849188
H	2.10810419233067	1.19977650353322	-0.26937659570409
C	-1.08435212460112	-0.95620135463166	-0.12422334248549
H	-1.12612240759015	-1.03083625660460	-1.22902272774667
H	-1.88875330533827	-1.57659059834255	0.29235965561573
C	1.26028099268874	-0.82661696460031	-0.20012195111060
H	2.15324288851863	-1.35285179950691	0.16207667854805
H	1.23886052507938	-0.90067328356378	-1.30551528569750
O	0.13877746690655	-1.50826002973417	0.35406188768167

Nitrobenzene (15)

O	2.74445865683561	1.40312139214155	-0.35750478024001
N	2.30422279388824	0.25297832669727	-0.27598677425740
C	0.84076100084006	0.08896934257442	-0.10074411687129
C	0.31458312841562	-1.19888030183980	-0.00659897378592
H	0.98449169425859	-2.05451746870720	-0.06342829817047
C	0.03927502289398	1.22808755949786	-0.03565716426971
H	0.50070552423236	2.21043495165014	-0.11444753988225
C	-1.06226195218816	-1.34497044739496	0.15822572290437
H	-1.49358420288808	-2.34335144632110	0.23361425722543
C	-1.33574008618060	1.06590749043235	0.12953163320893
H	-1.97991972249586	1.94379567168450	0.18280784246450
C	-1.88566634367920	-0.21643696344211	0.22623122454381
H	-2.96192058977847	-0.33700581276190	0.35510885639799
O	2.99141007584592	-0.77091129421103	-0.32940988926802

Phenol (16)

O	2.46520574722708	-0.40517145286522	0.13031188535035
H	2.93843988304569	0.44498195603022	0.16111704986696
C	1.11603964332244	-0.15232114029950	0.09722857190716
C	0.25397694724404	-1.25467339540559	0.05007198673172
H	0.67900980965558	-2.25847130619264	0.04146279841973
C	0.59361116301943	1.14716350860475	0.10912946767445
H	1.26914912324268	2.00609394880041	0.14596735883043
C	-1.12446941337521	-1.04847678929431	0.01515563768542
H	-1.79134377750212	-1.91117442408581	-0.02154659674130
C	-0.78918025840235	1.33961001292577	0.07385340678173
H	-1.18783641912889	2.35506272521292	0.08347910234942
C	-1.65572504270498	0.24553479003529	0.02672165820459
H	-2.73448040564339	0.39881156653371	-0.00074932706066

Anisole (17)

C	2.68691960331702	0.13999595834652	-0.11063501967295
H	2.68344096486603	0.51697237598338	0.92564614620555
H	3.62906909097377	-0.38412896005612	-0.30477351683199
H	2.58959406734983	0.98979960081506	-0.80662455433709
O	1.65610216090143	-0.82250711087139	-0.31760746513516
C	0.36325194451244	-0.40975818008334	-0.13413432114727
C	-0.62439710998006	-1.38422122796817	-0.35039736820156
H	-0.31241147650608	-2.38679349074143	-0.64434003670521
C	-0.00768701428547	0.88793789869890	0.24468359192055
H	0.74275666410836	1.65737724245316	0.41731117802527
C	-1.96749220268451	-1.05999566098718	-0.18826227000198
H	-2.72637573875598	-1.82509741493036	-0.35896983362851
C	-1.36363626532170	1.19705203757001	0.40348979435186
H	-1.64432607811915	2.20940020050820	0.69853532993528
C	-2.34824378107710	0.23371639787869	0.19016760211915
H	-3.40163782929883	0.48408933338406	0.31621074310409

Benzaldehyde (18)

O	2.94263597429521	1.19996172543360	-0.48513524437964
C	2.31325921547403	0.16198637846516	-0.36958263146187
H	2.82882853613303	-0.83182257260562	-0.43600783604740
C	0.85504102653235	0.07270629231235	-0.13680023871620
C	0.25363633512173	-1.18914154608347	-0.02272160343897
H	0.87298639022288	-2.08554983640121	-0.10779377118218
C	0.07186333427765	1.23419978801308	-0.02968084895190
H	0.56336941754174	2.20365224967413	-0.12216881666799
C	-1.11965379291865	-1.29493547124989	0.19682701839086
H	-1.58801284562844	-2.27580014905196	0.28568895122636
C	-1.29769250124538	1.12766435095455	0.18932282155817
H	-1.90978647067495	2.02661869155064	0.27309969606611
C	-1.89377939311800	-0.13592235577246	0.30261818292163
H	-2.96831322601321	-0.21512154523891	0.47435832068303

Benzonitrile (19)

N	3.51961945645391	0.95274746014816	-0.22345998046378
C	2.39618309209311	0.64946435935315	-0.15512784487064
C	1.01891656158161	0.27722051853738	-0.06631443673425
C	0.02406174933033	1.27091810368715	-0.01421418227689
H	0.31387400465403	2.32090192503490	-0.04259515028402
C	0.65914158848966	-1.08260558391179	-0.03097688569519
H	1.43763388053063	-1.84387177154188	-0.07193637670773
C	-1.31613138951621	0.90125223844318	0.07263095195525
H	-2.08662942518606	1.67164375139734	0.11302348103845
C	-0.68482602672322	-1.43829291718799	0.05599343573874
H	-0.96312551117540	-2.49213449703081	0.08373195382972
C	-1.67250940023984	-0.45015936338476	0.10785454977068
H	-2.72321758029252	-0.73416922354403	0.17615648469966

4-Dimethylaminopyridine (20)

C	-1.84531439626033	0.71950197920473	1.03391860582320
H	-1.60266226488247	0.37893533229128	2.05697596358828
H	-1.61353582070335	1.79390610089423	0.96801352935944
H	-2.92242634143402	0.59804227809135	0.87646430153932
C	-1.84571418392628	-1.03116248242347	-0.75859473731838
H	-2.92275616469553	-0.87724951790076	-0.63209759791962
H	-1.61520425926180	-0.93908097174788	-1.83136817983551
H	-1.60216747491177	-2.06204087355085	-0.44323520263947
N	-1.13105089226693	-0.03169260208389	0.01626658757920
C	0.24141932794724	0.01850047721867	-0.03309429849907
C	0.99103612274807	-0.79190347380210	-0.91918214400157
H	0.50610978665555	-1.50279245939913	-1.58549220434815
C	0.99150696028318	0.88538771714765	0.79739216815288
H	0.50697448440253	1.53524407362043	1.52363275698445
C	2.37744963290530	-0.68095939472968	-0.93772747468992
H	2.95199319374516	-1.30822866849467	-1.62650325855636
C	2.37788984645871	0.90618484154364	0.68644339623310
H	2.95279734426840	1.58008647934308	1.32934348000426
N	3.09505009892834	0.14813216477737	-0.16054969145609

Pyrimidine (21)

C	1.52118116530529	-0.15912926954967	-0.10713688793694
H	2.60647678468336	-0.27322603181243	-0.18349809504487
N	0.81348438665809	-1.28862713466386	0.04322875700274
C	-0.51638832913360	-1.13361511295518	0.13555654300686
H	-1.10627516888548	-2.04672030345393	0.25849699211350
C	-0.28418761559000	1.21933232818811	-0.07930724033602
H	-0.68354156158362	2.23644628446939	-0.13262792360873
C	-1.12666052491825	0.11921761130881	0.07920617106538
H	-2.20815452828203	0.23291823651957	0.15529480194089
N	1.04865439174625	1.09444139194919	-0.17440011820279

2,6-Dimethyl- γ -pyrone (22)

O	-0.04661427099177	3.26091361256264	-0.20039454133953
C	-0.02397062583758	2.02573166048813	-0.12366119556530
C	-1.23039294066410	1.19696556549629	-0.09461974565977
H	-2.19964617864795	1.69206786531735	-0.14301138830316
C	1.21189183214136	1.24425702557807	-0.05285327333156
H	2.16243636993395	1.77635788849560	-0.06842174729035
C	-1.16993889021791	-0.15201976790423	-0.01019117137378
C	1.20071779917118	-0.10615043074962	0.03016190799962
C	-2.32167508715149	-1.09823501500182	0.02766399874655
H	-2.30554622658892	-1.69107625769703	0.95447762338724
H	-3.26757760566013	-0.54830261959527	-0.02589893108323
H	-2.27096436097105	-1.80408019539463	-0.81490319197654
C	2.38622472731084	-1.00728736789481	0.10664914594879
H	2.36006496619483	-1.60417479289648	1.03060849640477
H	2.39279569801667	-1.71097312016048	-0.73920028211911
H	3.31141253655134	-0.42103709732333	0.08828149275752
O	0.02804925741076	-0.82036695332038	0.05286080279785

Norbornene (23)

C	0.22733009389728	-1.17149509215501	0.00936260834457
H	0.31773749052054	-2.25658547288321	-0.11606364836136
C	0.39205394569747	1.03960161967232	-0.41822199343095
H	0.63270732962040	1.97593760542568	-0.93458994778490
C	0.66009913112407	-0.28044145509746	-1.17881030568258
H	1.71781988225849	-0.41051722007657	-1.44450772087112
H	0.02796529552723	-0.40577592973378	-2.07104921886381
C	1.04388445952450	-0.54908439437659	1.13042582040516
H	1.36096096747469	-1.07072950787354	2.03303972270943
C	1.14225126612289	0.76752702119559	0.87561680705815
H	1.55538730211707	1.53574650568778	1.52872415750002
C	-1.12897959812691	0.86180972443035	-0.07460639966819
H	-1.42775373801686	1.48570530128791	0.77755613151880
H	-1.75023184605646	1.14943404483367	-0.93510901652341
C	-1.24263760760014	-0.66428561630921	0.21990386612966
H	-1.60498462653213	-0.87919016728499	1.23344870821492
H	-1.92099774755215	-1.15831596674297	-0.49068457069441

Norcamphor (24)

O	1.62298560450425	1.73303678685190	1.40225272749383
C	1.02951298875034	0.99837600569814	0.63864817748690
C	-0.37822115564302	1.17942059187911	0.07368597284139
H	-0.73610823827870	2.21328776532705	0.10939374778440
C	-1.25458184705401	0.15427681506602	0.85594956433443
H	-1.12743529743675	0.24807671750050	1.94247987969966
H	-2.31413751290686	0.33980106780084	0.63105918155497
C	-0.78691708758190	-1.22789359873213	0.29940192655820
H	-1.61402076567861	-1.75669875163645	-0.19421478269915
H	-0.39866328016364	-1.88858585425122	1.08691327168228
C	0.30937453654291	-0.84373673414289	-0.72143105389287
H	0.53566005131556	-1.62567762351748	-1.45651680619132
C	-0.24771717527397	0.48085633955407	-1.29729209680989
H	-1.21548885513950	0.35847193622762	-1.80444592259765
H	0.45273255376230	0.98619783591745	-1.97798644870696
C	1.53665077752522	-0.33568830304852	0.06086350117389
H	2.38998350957657	-0.12878383252071	-0.60539078148163
H	1.89289619317979	-1.00206716397332	0.85912194176951

(+)-Fenchone (25)

C	2.13684681238489	-1.10914475909141	-0.28422874202318
H	2.48656148212780	-1.69123683862987	0.57925218249778
H	1.66248896626709	-1.80406807478424	-0.98936986425493
H	3.01116480737181	-0.66258459808825	-0.78111063845628
C	1.17848247071251	0.00227329388712	0.18528790510251
C	1.94049466130799	0.94064034844076	1.12802797001573
H	2.37653385126964	0.35826558792022	1.95080416893993
H	2.75586735790959	1.44957851735348	0.59209893772468
H	1.29307832678929	1.70580677354760	1.57551219820623
C	0.45346034946934	0.67442862926121	-1.02011601082608
H	1.13725140915679	0.92102460573500	-1.84354981908450
C	-0.65418343881437	-0.35293892256100	-1.34592505841843
H	-1.36420706279911	0.01165960032108	-2.10288064640982
H	-0.28294219059159	-1.33695626383026	-1.66144147048536
C	-0.40411227622690	1.86249629794782	-0.53355843742090
H	0.16444099835961	2.59376342062734	0.05332449423795
H	-0.82136868282891	2.39869916443368	-1.39753330811857
C	-1.53026265692235	1.16571585258221	0.28447402265072
H	-1.52320855918373	1.43537216624779	1.34985787290802
H	-2.52926356871482	1.40786824828274	-0.10680805953267
C	-1.27318541600300	-0.36583960918956	0.07260206297905
C	-2.45269952913379	-1.26780461698052	0.36376393698885
H	-2.73348830460215	-1.20620859493145	1.42455222062803
H	-3.32418327613714	-0.98327852372809	-0.24396037923736
H	-2.20898946589500	-2.31801441840223	0.14960073459199
C	-0.02519316263992	-0.64372132183941	0.91565326816206
O	0.02245909736644	-1.24287296453173	1.97292145863454

Isophorone (26)

O	-0.03508167107188	3.27382019198590	0.29914510541563
C	0.09796447037172	2.05197629733294	0.33717981561615
C	1.39620047473789	1.40571902745202	0.10193052533612
H	2.24401024043403	2.07687575763782	-0.05511871454733
C	1.55160369722772	0.06323327001127	0.07124032386508
C	0.38225447221610	-0.86570786693912	0.27055657220015
H	0.52512217622040	-1.77117162950791	-0.34363080850119
H	0.40087068318325	-1.21765132812834	1.32004045914111
C	-0.99416815464530	-0.23883726410485	-0.03318121862728
C	-2.10438931022230	-1.15631529865005	0.49354305017226
H	-2.01875389003302	-1.30080112667403	1.58116082672243
H	-3.09617278598818	-0.72813266572805	0.28608437173447
H	-2.05880336559189	-2.14626524824723	0.01395466083273
C	-1.16385376315842	-0.05667429777334	-1.55157767074363
H	-2.13163809739196	0.41348535190625	-1.77873536152172
H	-0.37364483932641	0.57751679737395	-1.97651232747012
H	-1.13237256913335	-1.03011099846024	-2.06342170441039
C	-1.06597169564227	1.12922964318052	0.67150988884168
H	-1.04476283436076	0.97500944630012	1.76636665998705
H	-2.00204586180919	1.65779963922208	0.44099921481194
C	2.88564679711654	-0.57945740998336	-0.15231626672592
H	3.11703657887068	-1.28769649790783	0.66036682990006
H	2.87597826161558	-1.17105875497444	-1.08275592632275
H	3.69397998638101	0.15968496467594	-0.21453730570655

4-Cyanobenzaldehyde (27)

O	-3.33165691921501	1.14794383954928	-0.12026727290213
C	-2.67835269451153	0.12047243819101	-0.13948311794087
H	-3.17641230686136	-0.87966897985588	-0.22073128180635
C	-1.19822676505241	0.06300315788233	-0.06157079439091
C	-0.44343404100051	1.24320369619243	0.04278952116791
H	-0.96918134986837	2.19839776816991	0.06506603771749
C	-0.55336709330996	-1.18168824381182	-0.09279930690774
H	-1.14945561446872	-2.09334137807441	-0.17406234315115
C	0.94025152756633	1.18029946229198	0.11506975213940
H	1.53922841246613	2.08683879357572	0.19614339194710
C	0.83332171946212	-1.25883040827469	-0.02107089092732
H	1.34533140902890	-2.22009705267950	-0.04442118929487
C	1.58549831352192	-0.07418307410123	0.08312033804395
C	3.01090879415893	-0.14212904393466	0.15648894684634
N	4.17382060808352	-0.19716797512046	0.21145420945917

Mesitaldehyde (28)

C	-0.54153161737062	2.81047589058259	-0.00165465842191
H	-0.21039132302759	3.27767383641715	0.93698355720738
H	0.02634832000330	3.28170033716433	-0.81680236107552
H	-1.60007537769307	3.06384707971377	-0.14334029452009
C	2.28754763231463	-1.47322392288036	0.34609694991090
H	2.77536087834438	-1.22095999964878	1.29742168146190
H	2.11713152419143	-2.55761871083251	0.30460608845041
H	3.00820250738842	-1.19119634320002	-0.43345152787903
C	-0.37098310190763	1.30787673941018	0.01734283698979
C	0.98444842946880	-0.74041768730131	0.18312244748217
C	-1.50730147144000	0.51367242159436	-0.14120379126181
H	-2.47733710712799	0.99948831007561	-0.27308624691571
C	-0.18512943117920	-1.48552827702636	0.01843605677094
H	-0.11271100470854	-2.57618067357396	0.01258936516873
C	-1.43862924414959	-0.88440210579564	-0.14062009675394
C	-2.68234349841250	-1.71922285899250	-0.28149473048810
H	-3.08772043188174	-1.98015425370829	0.70952218414628
H	-3.46891939176739	-1.17995739559150	-0.82603580868240
H	-2.47569093420619	-2.66154200148214	-0.80688165884443
C	0.89320737355565	0.67935712054336	0.18428444264911
C	2.08608140908639	1.52837530718935	0.34831875108852
H	1.87995715238922	2.62498394420346	0.32469360038183
O	3.23956370812984	1.14487424313924	0.50612221313501

(+)-Camphor (29)

C	0.24287474462422	2.22895305814979	-0.33675566889983
H	-0.14876803981803	2.68257094517079	0.58511698940473
H	1.17855555500547	2.74039112092239	-0.60263357223827
H	-0.47107273633065	2.41979567609798	-1.15032169739026
C	0.48699578778780	0.74754657021135	-0.15699694625360
C	1.13667964868393	0.05726247693950	-1.36428275013315
C	0.91025477198621	-1.45087010807138	-1.17387722649692
H	0.32960841943149	-1.84261928267787	-2.02105090527832
H	1.86996092000761	-1.98756366660742	-1.16421773415754
C	0.15084079904016	-1.49544369396484	0.16458668758445
H	-0.40713927506794	-2.42523059140614	0.33909834335027
C	-0.73221173731557	-0.20657145254640	0.09849962135874
C	-1.74584833749864	-0.20221400880870	-1.05328176319560
H	-2.47307907510858	-1.01916990704087	-0.93266065713005
H	-2.31085571193109	0.74105829753514	-1.06192568625209
H	-1.28134541287610	-0.30865222497163	-2.04246352904452
C	-1.48635629188699	0.08896565569518	1.39770028818833
H	-0.83351662569020	0.14655567012388	2.27680344428090
H	-2.02674273253964	1.04361945014043	1.32014033333416
H	-2.23350564873661	-0.69665127399988	1.58839304197815
C	1.17802463140269	-1.14744333058140	1.26461618266557
H	0.77497342137907	-1.32940047790285	2.26938340391489
H	2.08835310898462	-1.75455999739639	1.16734315607248
C	1.44548366719259	0.36603637766352	1.01825635725900
H	1.19780091901977	0.97789738481683	1.89739153231123
H	2.49260811609929	0.58236919110887	0.76548530067569
O	1.72797711415513	0.59267914140014	-2.28085254590843

Indole (30)

C	0.38119393538495	0.70588923337384	0.08866553827516
C	0.23323094797444	-0.70082335375306	-0.11768173488613
N	1.50400543712975	-1.22495478934518	-0.25832438204339
H	1.71981089053872	-2.20043242847631	-0.41663529063970
C	2.43826350368318	-0.20844910423316	-0.14898704879956
H	3.49848634846739	-0.42444864188604	-0.23268152463607
C	1.78957245406254	0.98622314666021	0.06344349169903
H	2.26591341341452	1.95331727292281	0.18725123061738
C	-1.01776789169121	-1.32543166722157	-0.15219262587045
H	-1.11342110397012	-2.40095330593129	-0.31059005627740
C	-2.14140142305500	-0.52207006801005	0.02387068761064
H	-3.13251179874779	-0.97663219650882	0.00273385230909
C	-2.02093018572799	0.86889056686999	0.22916314668515
H	-2.92369103006466	1.46616555964419	0.36319138734512
C	-0.77578867328519	1.48733009004244	0.26303206549956
H	-0.69498382411353	2.56381668585199	0.42242926311158

Adamantane (31)

C	0.92556632488184	0.39899037440059	1.15304853743616
H	1.60095249129047	0.68306094146967	1.97636245283604
C	0.78391256306839	-0.73986362683728	-1.09002170397602
H	1.35787730221924	-1.26876829474970	-1.86823720080431
C	-0.65469336063021	1.26994739856852	-0.60524096818035
H	-1.11005584934628	2.17651499664152	-1.03593070946497
C	-1.12783876417951	-0.92431222860487	0.53926632075469
H	-1.92121409017332	-1.58490232818716	0.92526695008893
C	1.72403467564097	-0.34085743356350	0.06297027693535
H	2.53277373700311	0.30576212314444	-0.31693199446100
H	2.20206906585644	-1.23919566120128	0.48828976114356
C	0.14720829408536	0.52758885766467	-1.69078753611776
H	0.93203011632536	1.18718289193666	-2.09727010478968
H	-0.51480442588192	0.25713013949013	-2.53031776583820
C	0.28801682283859	1.66371978312433	0.54765392676099
H	1.07452115579506	2.34261366520630	0.17767764618787
H	-0.27300726720984	2.21200805421060	1.32286078995279
C	-1.76050897959222	0.34311152846334	-0.06557900474688
H	-2.35500897011397	0.86991363462719	0.69965515409414
H	-2.45268456441005	0.06729541494917	-0.87870188060819
C	-0.32419122733446	-1.66230843223286	-0.54789119191063
H	0.12068925079541	-2.58097174100588	-0.13024883702903
H	-0.99300663648669	-1.97212968974828	-1.36831256320990
C	-0.18443004855826	-0.52446150951751	1.68947501770322
H	-0.75342922035294	-0.00886105782624	2.48108020701046
H	0.25905260446946	-1.42525380042260	2.14589542023268

Dicyclopentadiene (32)

C	0.47210466328738	0.95494603085863	0.53357326632405
H	0.62953843400454	1.57054175674484	1.43437786698220
C	1.61954502344355	1.06270918064999	-0.42792399905448
H	1.93754852960257	2.01043397479335	-0.86593584586124
C	2.17586243696754	-0.12502963471774	-0.69687860451670
H	3.00681850089639	-0.28012284129574	-1.38768394813626
C	1.53107909548795	-1.25848806194412	0.06125895305070
H	1.15418212912413	-2.03777275345863	-0.62307496245798
H	2.25613364232205	-1.75958987175571	0.72393790506887
C	0.39029846378379	-0.58029978682954	0.86129403949678
H	0.50629996657754	-0.75854540981370	1.93885537234151
C	-1.08273460035886	-0.93891170256446	0.44837767342683
H	-1.44399472425831	-1.88667415695795	0.86531990883481
C	-1.17815424226103	-0.78947858924529	-1.06155555231413
H	-1.18178590633937	-1.61192305967756	-1.77538806661678
C	-1.10854832815015	0.52256679704602	-1.34436235560950
H	-1.03275296227497	0.97548915018025	-2.33163379546219
C	-0.97322764304945	1.26897789310958	-0.02990822706453
H	-1.22379602747532	2.33598895336589	-0.04527771767140
C	-1.80858481256785	0.35329750297117	0.89694485212895
H	-1.67482054301032	0.57880961122502	1.96614229986900
H	-2.87734509575179	0.35143601731567	0.64284593724149

1-Methylnaphthalene (33)

C	-2.20229131752364	-1.45808786592857	0.40576619303082
H	-1.89948086697823	-1.97412510381080	1.33014315480283
H	-3.26749469978238	-1.20916506946741	0.49226053330948
H	-2.09004413381956	-2.18082760025591	-0.41756149163849
C	-1.38500504172071	-0.21601431878570	0.17082298876079
C	0.03810079054451	-0.29280452065279	0.02953250128480
C	0.74458776532680	-1.52330892087219	0.10105177073529
H	0.19120446926137	-2.44732203220138	0.26866450282974
C	2.11752940672014	-1.56933895820562	-0.03769495051483
H	2.63832070753501	-2.52620006161974	0.02107756789206
C	2.85489272841775	-0.38275614460801	-0.25564210547236
H	3.93939970423843	-0.42966883356532	-0.36359238511029
C	2.20180213783321	0.82960986565494	-0.33067751534890
H	2.76318386303341	1.75126237990297	-0.49842007701721
C	0.79130586660740	0.91211743412215	-0.19236318386804
C	0.11012390745525	2.15542958206064	-0.26735511172326
H	0.69057217495433	3.06473615399658	-0.43542304715093
C	-1.25878963269998	2.20764824633804	-0.12894210524910
H	-1.78165488836811	3.16343185719394	-0.18638640020415
C	-1.99838061665994	1.02195424317831	0.08944283555080
H	-3.08347132437504	1.08399266752586	0.19710532510095

9-Fluorenone (34)

O	0.14235381233165	3.05477126007184	0.10756355551567
C	0.08681620131949	1.83486418217585	0.06442843313422
C	1.22577257024555	0.86299908070859	0.17418079097838
C	-1.13540987529764	0.98049665435299	-0.10900338653080
C	0.71838724830680	-0.45156269053156	0.07304573579939
C	-0.74970099135196	-0.37851234048635	-0.10291285153226
C	-2.46245492665635	1.35979568472521	-0.25808247560299
H	-2.73372584018578	2.41668604447868	-0.25930279774331
C	2.58113848693171	1.10891727093118	0.34658566124180
H	2.94690370481658	2.13417542170053	0.42153397027198
C	-3.42882705718290	0.35429584802655	-0.40487752517851
H	-4.47953025677878	0.62075764346871	-0.52408011801411
C	3.45170252605421	0.01205722149904	0.41985053060477
H	4.52177572666876	0.17296674962632	0.55491009706025
C	-3.05266564631569	-0.99356587082804	-0.39972195770204
H	-3.81723129245883	-1.76330625580702	-0.51507264418041
C	2.95482903846945	-1.29237303570484	0.32008386408734
H	3.64600592069529	-2.13449877170762	0.37898564210231
C	-1.70965998157658	-1.37491700306668	-0.24878501000744
H	-1.43573198380828	-2.43103130861572	-0.24690136494847
C	1.58344548312202	-1.53875333140127	0.14563085154117
H	1.21490413265132	-2.56289945361641	0.07012099910306

Anthracene (35)

C	-1.16784167980323	-0.77141076763588	-0.20901678339273
C	-1.24121539914392	0.53922120372067	0.40242538968321
C	1.18881569572516	0.76738850525300	0.20455832677834
C	1.26218983270489	-0.54325053810735	-0.40688754527032
C	-2.37683352326644	-1.50691821002172	-0.39629426022134
H	-2.31900520847248	-2.49518800432375	-0.85719163941188
C	-2.51945185990790	1.04118524599128	0.79218561695226
H	-2.57215618980655	2.02985510316391	1.25279295930435
C	2.39781853588153	1.50290288959279	0.39183863720437
H	2.33994494255190	2.49116787913751	0.85274019237800
C	2.54043065821892	-1.04520983765370	-0.79664618402551
H	2.59316735583420	-2.03387652876262	-1.25725696102451
C	-3.58738678463051	-0.98777966817984	-0.00705123192443
H	-4.50266180905821	-1.56201202425645	-0.15694681286690
C	-3.65957703645203	0.30134977756658	0.59416002636556
H	-4.62926465760216	0.69860978045780	0.89715939408386
C	3.60836279698901	0.98376262041560	0.00259630248121
H	4.52366064048157	1.55796332201490	0.15247535789986
C	3.68054785322898	-0.30537098011006	-0.59861712251219
H	4.65024251224166	-0.70261823031708	-0.90161142504540
C	-0.06070313075706	1.27034729702931	0.59127915359605
H	-0.11598337983956	2.25990100863537	1.05224000137275
C	0.08168010653825	-1.27438155532213	-0.59574384038709
H	0.13695472834396	-2.26393428828815	-1.05670355201751

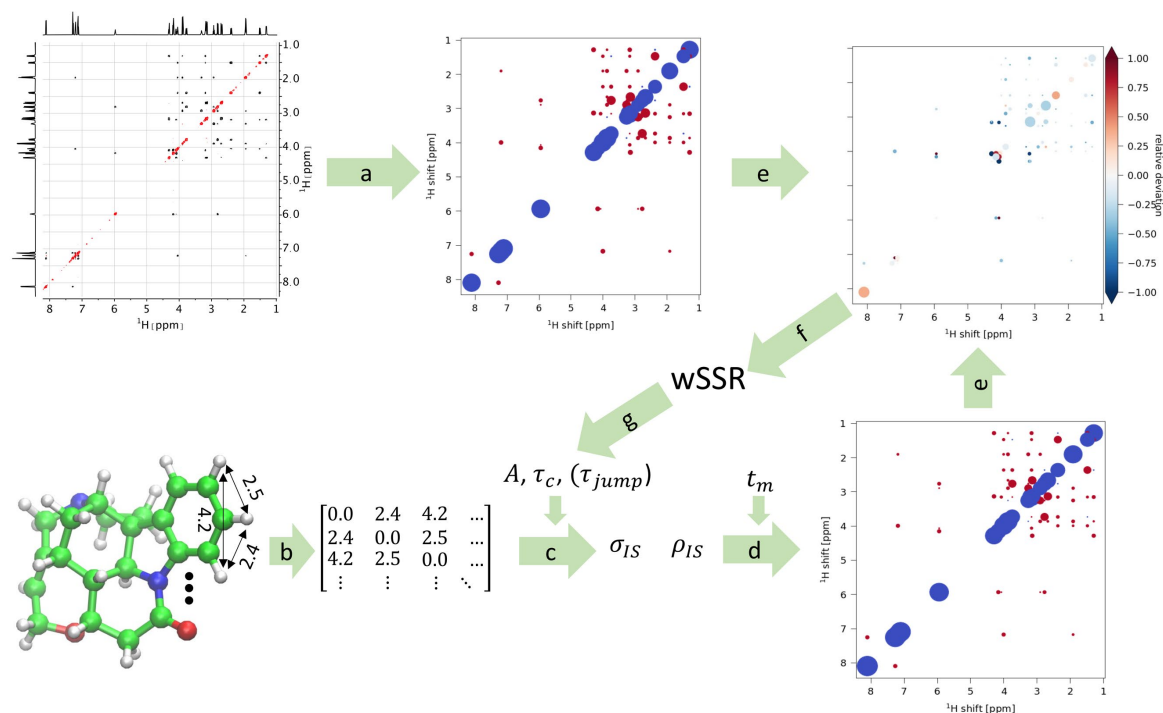
6 ***NOVAS: A Simple Protocol for Using NOESY Volumes Instead of NOE-Derived Distances for the Determination of Relative Configuration***

In this chapter, we propose a simple protocol that we call NOVAS (**NOE Volumes Affected by Spin diffusion**) for directly utilizing NOESY volumes instead of NOE-derived distances to solve structural problems in organic chemistry. Central to this approach is that the NOESY spectra are recorded at long mixing times beyond the linear build-up regime. The peaks in such spectra are close to the intensity maximum and much less affected by troublesome COSY artefacts. At long mixing times, the accuracy of NOE-derived distances deteriorates due to the influence of multi-spin effects (spin diffusion). This is in contrast to our protocol, where the rich spatial information pertinent to these effects is used in a straightforward manner and adds to the discriminating power of the experimental data. We show for different test systems that such NOESY spectra can closely be reproduced with a simple fitting procedure, and investigate how the match between experimental and calculated spectra can be used for stereospecific assignment of diastereotopic protons in methylene groups. In the NOVAS protocol, we fit experimental NOE volumes based on Boltzmann-weighted, DFT-optimized structures using a global correlation time τ_c , a scaling factor A and, if needed, a local correlation time τ_{jump} to account for the fast internal motion of methyl groups. With this protocol, no transformation of the volumes into distances is necessary and one can directly validate a computer-generated ensemble using the primary experimental data.

6.1 Introduction

The nuclear Overhauser effect (NOE), also known as nuclear Overhauser enhancement, is one of the most richest sources of structural data in nuclear magnetic resonance (NMR) spectroscopy.²²⁻²⁴ This dipole-dipole interaction can be used to obtain valuable distance information between two spins through space. The obtained information can be qualitative or quantitative.⁵⁰⁶⁻⁵⁰⁸ In a standard setup, one uses short mixing times to record a NOESY spectrum in order to stay within the initial rate approximation (two spin approximation) and to avoid spin diffusion.⁵⁰⁹ The problems of short mixing times are on one side the low intensity of the cross-peaks since the NOE has had no time to fully build-up and on the other side that NOESY spectra with short mixing times are often affected by severe artifacts (mainly of COSY-type).⁵¹⁰ To alleviate the first problem, one possibility is the so-called PANIC approach, where the initial regime of linear growth is extended to longer mixing times by dividing the cross-peak volumes of the NOESY spectrum by the diagonal-peak volumes.^{511,512} Further, it is also possible to trace the NOE build-up curve using different mixing times, resulting in so-called, exact NOEs (eNOE).^{27,513} All these approaches are used to transform the peak volumes or build-up rates into distances. The accuracy of the derived distances strongly depends on the chosen method. Most simply, this can be done by a single point calibration of the proportionality between cross-peak volume V and r^{-6} using a known volume-distance pair. The derived distances can then be used as input for structure refinement or for comparison with an ensemble resulting from molecular dynamics (MD) simulations.^{26,195,514,515} Yet, it would be more straightforward to directly use the peak volumes as a primary source of information and compare these to computational data. The background of dipolar relaxation in multi-spin systems and the calculation of NOESY spectra is known for a long time and the most important equations are summarized in the next section. For a more detailed overview, the reader is referred to Refs. 25, 160 and 516.

In this study, we show for multiple test systems that we can successfully calculate NOESY spectra based on density functional theory (DFT) optimized structures of small molecules and fit them to experimental NOESY spectra beyond the linear build-up regime (Scheme 6.1). In contrast to spectra recorded at short mixing times, cross-peak intensities are higher and often, due to spin diffusion, also indirect cross-peaks are observable, which contain additional valuable spatial information about the system under study. Since the approach presented in this Chapter is based on NOE Volumes Affected by Spin diffusion, we refer to it by the acronym NOVAS. For the fitting process, we assume that the entire molecule has one global correlation time and – except for methyl group rotations which are treated separately – internal motions are slow compared to overall tumbling.



Scheme 6.1: Schematic illustration how the NOVAS procedure works. First, a NOESY spectrum with a long mixing time (t_m) is recorded and the cross- and diagonal-peaks are integrated (a). The inter-proton distances of a DFT optimized 3D structure of the molecule of interest are calculated and stored in a matrix (b). With an initial guess for A , τ_c and, if needed, for τ_{jump} (for fast internal methyl rotation, see below) all cross- and auto-relaxation rates (σ_{IS} and ρ_{IS}) are calculated (c). With these rates the NOESY spectrum at a given mixing time is calculated (d). To quantify the relative difference between predicted and experimental spectrum (e) the weighted sum of squared residuals (WSSR) (f) is calculated. This is used as target function in the minimization process and the parameters A , τ_c (and τ_{jump}) are optimized until convergence is reached (g).

6.2 Theory

The steady-state NOE can be described by looking at the relaxation of a two-spin system with spins I and S using the well-known Solomon equations:^{516,517}

$$\frac{dI_z}{dt} = -\rho_I(I_z - I_{z,0}) - \sigma_{IS}(S_z - S_{z,0}) \quad (6.1)$$

$$\frac{dS_z}{dt} = -\rho_S(S_z - S_{z,0}) - \sigma_{IS}(I_z - I_{z,0}) \quad (6.2)$$

ρ_I is the auto-relaxation rate of spin I consisting of the longitudinal proton-proton dipolar auto-relaxation rate ρ_{IS} and an additional leakage term ρ_I^* including all other relaxation mechanisms. I_z is the magnetization in direction of the external magnetic field and $I_{z,0}$ is its equilibrium value. Analogous definitions apply for spin S . σ_{IS} is the cross-relaxation rate between spins I and S . When saturating S , S_z becomes 0 and after having reached a steady-state, the time derivative vanishes:

$$0 = -\rho_I(I_z - I_{z,0}) - \sigma_{IS}S_{z,0} \quad (6.3)$$

Solving Eq. (6.3) for I_z , one obtains:

$$I_z = I_{z,0} + \frac{\sigma_{IS}}{\rho_I} S_{z,0} \quad (6.4)$$

Rearranging and dividing by $I_{z,0}$ gives the definition of the steady-state NOE (η_{IS}) between spins I and S as the fractional enhancement of I_z with respect to its equilibrium value:

$$\frac{I_z - I_{z,0}}{I_{z,0}} = \frac{\sigma_{IS} S_{z,0}}{\rho_I I_{z,0}} = \frac{\sigma_{IS} \gamma^S}{\rho_I \gamma^I} \equiv \eta_{IS} \quad (6.5)$$

where γ^S and γ^I are the gyromagnetic ratios of spin S and I .

More generally, the time evolution of the vector ΔM of the *transient enhancements* $I_z - I_{z,0}$ of all spins in a molecule is given by:²⁵

$$\Delta M(t_m) = e^{-(\rho_D + \sigma) * t_m} \Delta M(0) \quad (6.6)$$

where t_m is called the mixing time, ρ_D is a diagonal matrix consisting of the n auto-relaxation rates ρ_I and σ is a $n \times n$ matrix consisting of all cross-relaxation rates. $\Delta M(0)$ describes the initial perturbation of the z-magnetizations. The NOESY-spectrum of such a molecule can then be described as:¹⁶⁰

$$M(t_m) = e^{-(\rho_D + \sigma) * t_m} M(0) \quad (6.7)$$

Here, $M(t_m)$ is the $n \times n$ matrix of NOE volumes at mixing time t_m . Assuming perfect pulses for all spins and complete relaxation between transients, a general scaling factor A can be introduced and $M(0)$ can be replaced by the diagonal matrix M_0 consisting of the equilibrium magnetizations.

$$M = A e^{-(\rho_D + \sigma) * t_m} M_0 \quad (6.8)$$

It is straightforward to account for multiple conformations interconverting on a time scale slower than overall tumbling by first calculating matrix M for the individual conformers and then taking the Boltzmann weighted average.

Assuming that ρ_D is dominated by proton-proton dipolar relaxation and other relaxation mechanisms (ρ_I^*) are negligible, it is possible to calculate a NOESY spectrum using the following equations:⁵¹⁶

$$\sigma_{IS} = \frac{1}{10} \left(\frac{K}{r_{IS}^3} \right)^2 (6J(2\omega_0) - J(0)) \quad (6.9)$$

$$\rho_{IS} = \frac{1}{10} \left(\frac{K}{r_{IS}^3} \right)^2 (J(0) + 3J(\omega_0) + 6J(2\omega_0)) \quad (6.10)$$

$$K = \frac{-\mu_0 \hbar \gamma^2}{4\pi} \quad (6.11)$$

$$J(\omega) = \frac{\tau_c}{1 + (\omega\tau_c)^2} \quad (6.12)$$

K is a constant with μ_0 being the permeability of the vacuum, \hbar Planck's constant, γ the gyromagnetic ratio of the ^1H nucleus, and r_{IS} the inter-nuclear distance between protons I and S . $J(\omega)$ is the reduced spectral density, ω is the Larmor frequency and τ_c is the correlation time for overall molecular tumbling. In an ensemble of interconverting conformers, τ_c can be treated as constant for simplicity.

To exemplify the difference between the two-spin/initial-rate approximation and an exact treatment of a multi-spin system, we will look at a three-spin system (I, S, X) assuming a correlation time τ_c of 10 ps and three different sets of inter-proton distances on a 600 MHz spectrometer. By plotting the volumes of the cross-peaks between spins I and S as well as X and S at varying mixing times, it can be nicely seen that, for this system, the initial linear rate approximation becomes invalid after mixing times larger than ~ 1 s (dotted lines in Figure 6.1). If the auto-relaxation rates are calculated by taking into account all three spins but cross-relaxation is treated as a two-spin process (this corresponds to PANIC^{511,512} or the eNOE^{27,513} protocol), the

time range of agreement with the exact treatment is extended. This is identical to neglecting spin diffusion. When the entire spin-system is considered, as will be the case in the NOVAS approach, one can nicely see that with increasing angle between spins S , X and I , spin diffusion becomes more and more dominant for the cross-peak between spin I and S (Figure 6.1). For small molecules, negative cross-peaks therefore contain valuable additional geometric information.

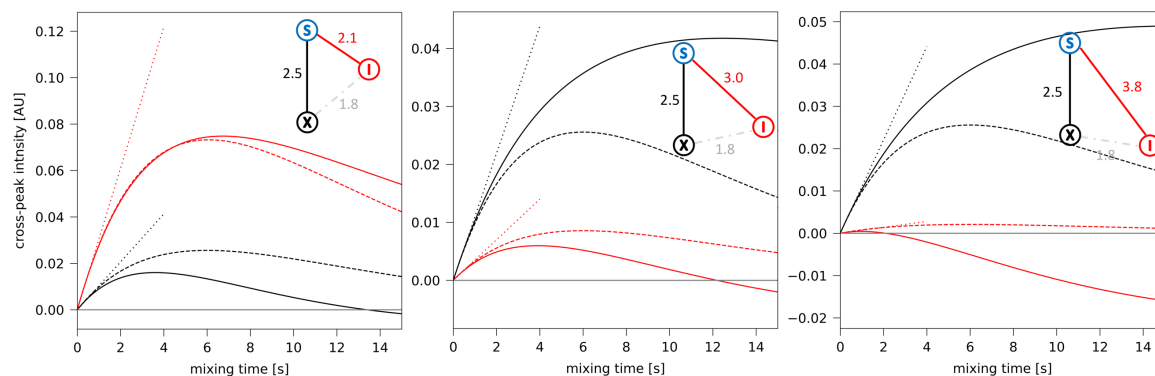


Figure 6.1: Cross-peak intensities for a three-spin system. Red lines correspond to the cross-peak between spins I and S , whereas black lines correspond to the cross-peak between spins X and S . The dotted lines are the extended initial linear slopes, the dashed lines ignore spin diffusion and the solid lines consider the three-spin system explicitly. Note that the scales for the y-axes are different for each plot.

In the simplest version of the NOVAS approach, the only two unknowns in the calculation of the NOESY spectrum are the global correlation time (τ_c) in equation (6.12) and the scaling factor (A) in equation (6.8). These two parameters are optimized using the weighted sum of squared residuals ($wSSR$) as target function:

$$wSSR = \sum \frac{(M_{exp} - M)^2}{|M_{exp}|} \quad (6.13)$$

where M_{exp} is a matrix containing the integrated volumes of the cross- and diagonal-peaks in the experimental NOESY spectrum. The $wSSR$ was chosen instead of the normally used sum of squared residuals (SSR) to compensate for the imbalance between the intense diagonal-peaks and the much weaker cross-peaks. Overlapping peaks in the experimental spectrum can still be used if the corresponding volumes are also summed up in the theoretical spectrum. This can also be done with non-assignable CH_2 protons.

6.3 Results

In the following, we applied the NOVAS approach to six compounds. First, we will present strychnine (**1**) as example of a rigid alkaloid of high complexity. As a second system, *trans*-crotonaldehyde (**2**) was selected to investigate the treatment of fast methyl group rotation in the NOVAS approach. Morphine (**3**) combines methyl group rotation with some conformational flexibility. Here, we also studied if stereospecific assignment of diastereotopic protons in methylene groups is possible using the NOVAS approach. This was also tested for the more complex molecule androstenedione (**4**) having two methyl groups. Lastly, we looked into the ability of the NOVAS approach regarding differentiation between diastereomers. For this purpose, we used the flexible diastereomers ephedrine (**5**) and pseudoephedrine (**6**).

6.3.1 Testing the NOVAS Procedure on Strychnine (**1**)

Due to its well dispersed ^1H spectrum, little overlap is observed in the NOESY spectrum of strychnine (**1**). In addition, one conformer clearly dominates the ensemble in solution. The NOESY spectrum was recorded on a 600 MHz spectrometer in CDCl_3 with a mixing time of 1 s. The experimental spectrum can be reproduced very well by our fitting protocol, including also the indirect NOE cross-peaks (Figure 6.2). A geometric situation in **1**, for which a strong indirect NOE cross-peak is observed, is shown in Figure 6.2D. The fact that the appearance of this peak is limited to a very specific relative orientation of three spins shows the potential power of the NOVAS approach. The correlation time of 30.8 ps found for the optimized spectrum is in good agreement with the proposed rotational correlation time of 25 ps derived from ^{13}C relaxation.⁵¹⁸ A relative mean absolute deviation (rMAD) of 10.8 % was obtained for the diagonal-peaks and 31.0 % for the cross-peaks. Keeping in mind that the NOE-derived distances have a $V^{-1/6}$ dependence, an error in cross-peak volume in this range translates into a distance error of about 5 % (typically 0.2 Å or smaller).

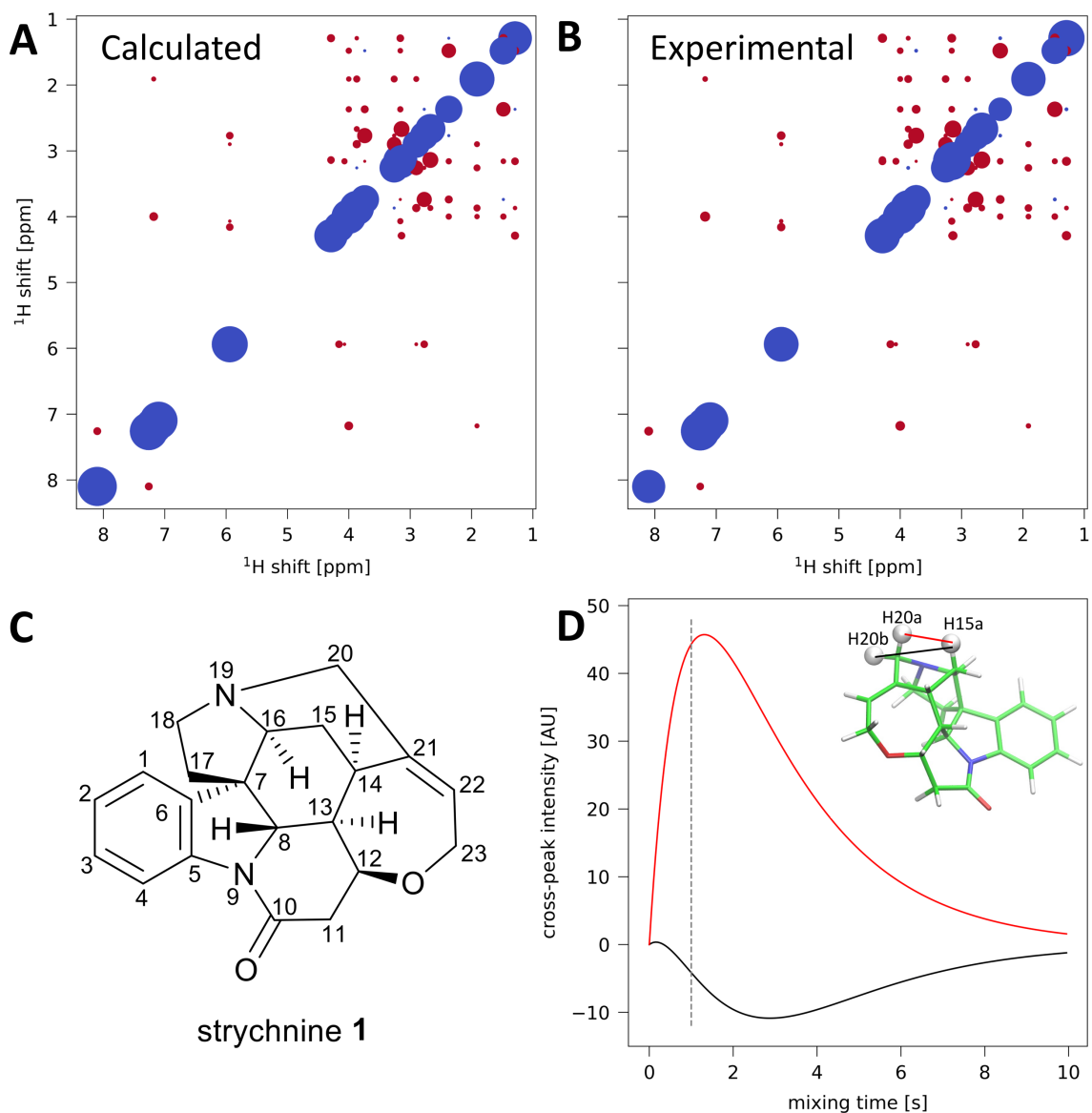


Figure 6.2: Comparison of the calculated NOESY spectrum (A) for the major conformation of strychnine (1) with the experimental NOESY volumes (B). The spectrum is reproduced very accurately including also the indirect NOE peaks (blue off-diagonal elements). Optimization of the match between experimental and calculated spectrum gave a correlation time τ_c of 30.3 ps and a scaling factor $A = 1585$. The difference between calculated and experimental NOE volumes can be seen on the left of Figure 6.3. The chemical structure of strychnine (1) is shown in C. (D) Calculated NOEs between H15a-H20a (red) and H15a-H20b (black) for different mixing times using the parameters obtained from NOVAS. The dashed grey line indicates the mixing time of 1 s that was used for the experimental NOESY spectrum. The H15a-H20a cross-peak is close to the maximum intensity and beyond the linear build-up regime. The three protons are nearly in one line and therefore a strong indirect NOE is observed due to spin diffusion.

There is a second minor conformation of strychnine reported in the literature with a population of ~2-3%.^{518,519} In our case, no improvement was achieved including the second conformer in the fitting procedure, independent of its weight. For illustration, NOESY spectra were calculated and fitted for the minor and the major conformation separately (Figure 6.3). The two structures mainly differ in the orientation of the methylene group in the seven-membered ring, whereas the rest of the structure is very similar. The obtained wSSRs of 886 for the major conformation and 1278 for

the minor conformation demonstrate that the NOVAS approach, although optimizing global parameters, is very sensitive to slight changes in the local 3D structure.

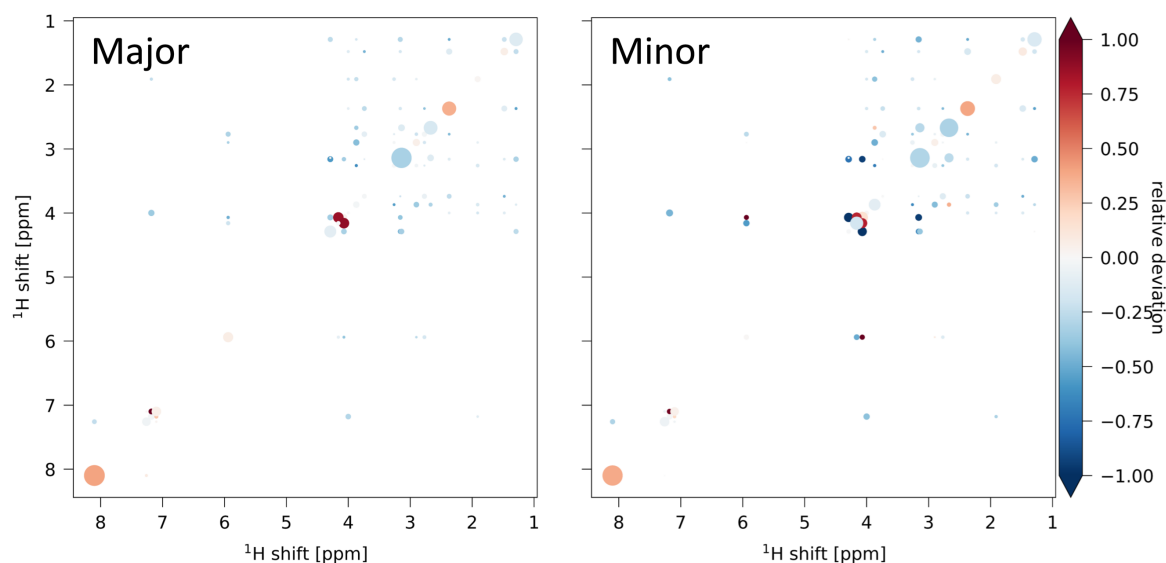


Figure 6.3: Difference between the experimental NOESY volumes of strychnine and the calculated NOESY volumes for the major (left) and the minor conformation (right). The size of the circles corresponds to the absolute difference, whereas the color denotes the relative deviation between calculated and experimental spectrum. rMAD for the diagonal elements were 10.8 % and 31.0 % for the major conformation and 12.5 % and 41.2 % for the minor conformation, respectively.

6.3.2 Dealing with Fast Internal Methyl Group Rotation: The Example of *trans*-Crotonaldehyde (2)

Since methyl group rotation is usually much faster than the overall tumbling rate, this motion needs to be considered explicitly when fitting NOESY volumes. For this purpose, we use the three-site hindered rotation model as described by James Tropp.⁵²⁰ In this model, the spectral density for an interaction involving a methyl proton is defined as follows:

$$f(n) = \left| \sum_{i=1}^3 \frac{Y_{2n}(\Phi_i^{mol})}{3r_i^3} \right|^2 \quad (6.14)$$

$$g(n) = \frac{1}{3} \sum_{i=1}^3 \left| \frac{Y_{2n}(\Phi_i^{mol})}{r_i^3} \right|^2 - f(n) \quad (6.15)$$

$$\tau_1 = \frac{1}{\frac{1}{\tau_c} + \frac{1}{\tau_{jump}}} \quad (6.16)$$

$$J^{00}(\omega) = \frac{1}{5} \left[\sum_{n=-2}^2 \frac{\tau_c f(n)}{1 + \omega^2 \tau_c^2} + \sum_{n=-2}^2 \frac{\tau_1 g(n)}{1 + \omega^2 \tau_1^2} \right] \quad (6.17)$$

r_i are the distances between the three methyl protons and another proton in the molecule and Φ_i^{mol} are the polar angles of the corresponding internuclear vectors in a common frame of

reference. Y_{2n} are second degree spherical harmonics of order -2 to 2, and τ_{jump} is the correlation time of the methyl group rotation defined as $1/(3k)$, where k is the rate constant for the jump between methyl proton positions. With that, the cross- and auto-relaxation rates between a methyl proton (X) and a proton outside the methyl group becomes:

$$\sigma_{IX} = \frac{4\pi}{10} K^2 (6J^{00}(2\omega_0) - J^{00}(0)) \quad (6.18)$$

$$\rho_{IX} = \frac{4\pi}{10} K^2 (J^{00}(0) + 3J^{00}(\omega_0) + 6J^{00}(2\omega_0)) \quad (6.19)$$

For intra-methyl relaxation, the same formula can be applied. But this time, r is constant and Φ_i^{mol} are the polar angles of the three orientations of the intra-methyl H-H-vector.

As a very simple test system to show that a proper treatment of the methyl group is essential, we fitted the experimental NOESY spectrum of *trans*-crotonaldehyde with a mixing time of 5 s. The shape of crotonaldehyde is far away from a spherical molecule and the three different correlation times along the principal axes will differ from each other. Therefore, our assumption that we can describe the system by one global rotational correlation time is not entirely appropriate anymore. Yet, we can show that, even for an anisotropically tumbling molecule we obtain a decent fit of the NOESY spectrum with rMAD of 3.7 % for the diagonal-peaks and 39.5 % for the cross-peaks, when assuming a short τ_{jump} of 0.33 ps (Figure 6.4 left). The fitted rotational correlation time is 2.7 ps with a wSSR of 412. In the literature, a ^{13}C -relaxation derived τ_c of 1.72 ± 0.11 ps is reported together with a τ_{jump} of 0.17 - 0.30 ps.⁵²¹ Optimizing also for τ_{jump} in addition to τ_c and A , gives a rMAD of 1.3 % for the diagonal-peaks and 50.5 % for the cross-peaks with a wSSR of 280, a τ_c of 3.5 ps and a τ_{jump} of 0.04 ps. Overall, the calculated τ_c values are larger but still in good agreement with the ^{13}C -derived values from the literature, especially when considering that some of our assumptions might break down for crotonaldehyde and that the literature values are reported for a sample in DMSO- d_6 at a much higher concentration (0.5 M). The obtained values for τ_{jump} imply a very fast methyl group rotation. As additional sanity check, a NOESY spectrum for *cis*-crotonaldehyde was fitted to the experimental NOESY spectrum of *trans*-crotonaldehyde. As expected, the fit is much worse (Figure A6.1 in the Appendix). The NOESY spectrum was additionally fitted without considering the methyl proton rotation, i.e., by treating the entire molecule as rigid entity. A wSSR of 858 shows that this fit is clearly worse (Figure 6.5). Also, by visual inspection it is clear that the cross-peaks are reproduced much worse without explicit treatment of methyl group mobility.

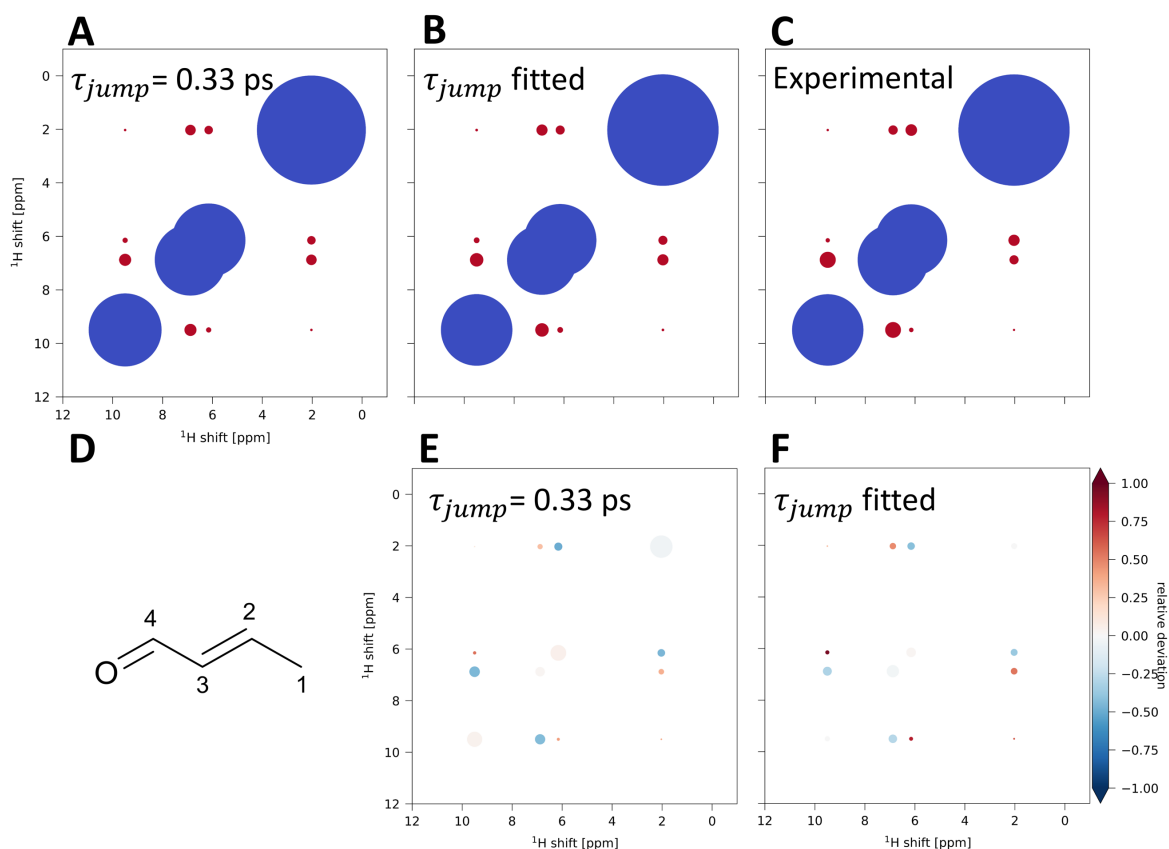


Figure 6.4: Fitting the NOESY spectrum of *trans*-crotonaldehyde with a fixed τ_{jump} of 0.33 ps (A) gave a correlation time of 2.7 ps with a wSSR of 412. B: NOESY spectrum obtained when τ_{jump} was optimized as well. This gave τ_{jump} of 0.04 ps and a correlation time of 3.5 ps with a wSSR of 280. Schematic representation of experimental NOESY spectrum of ca. 50 mM *trans*-crotonaldehyde in CDCl_3 (C). The chemical structure of *trans*-crotonaldehyde (**2**) is shown in D. The differences between the experimental and the calculated NOESY spectrum with a fixed τ_{jump} (E) and with optimized τ_{jump} (F) are also shown for easier comparison.

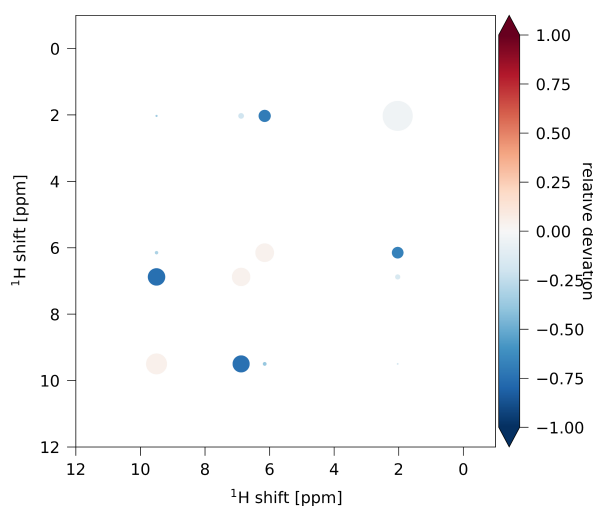


Figure 6.5: Difference between the experimental NOESY spectrum of *trans*-crotonaldehyde (**2**) and a fitted NOESY spectrum assuming no methyl group rotation. A wSSR of 858 shows that the calculated spectrum is farther away from the experimental NOESY spectrum compared to the model accounting for fast methyl group rotation (Figure 6.4).

For our sample of *trans*-crotonaldehyde in CDCl₃, we have also repeated the determination of the effective correlation time based on ¹³C relaxation. A series of inversion recovery experiments was fitted with the following equation:

$$M_{z(t)} = M_{z(0)} \left(1 - ae^{-\frac{t}{T_1}} \right) \quad (6.20)$$

where $M_{z(t)}$ is the ¹³C magnetization in z-direction at time t after the inversion pulse, T_1 is the longitudinal ¹³C relaxation time, and a is a fitting parameter accounting for imperfect pulses that should have a value between 1 and 2.

For a small molecule with $\omega\tau_c \ll 1$, the spectral density can be approximated by τ_c . By measuring the steady-state ¹H¹³C NOE (η_{obs}), one can directly calculate T_{1DD} , the dipolar contribution to T_1 (see Eqs. (6.1) – (6.5)).⁵²²

$$\frac{1}{T_1} = \frac{1}{T_{1DD}} + \frac{1}{T_{1other}} \quad (6.21)$$

$1/T_{1other}$ is the contribution to R_1 from other mechanisms than dipole-dipole relaxation.

$$T_{1DD} = \frac{2T_1}{\eta_{obs}} \quad (6.22)$$

Note that for CH₂ and CH₃ groups, T_{1DD} is then $\frac{1}{2\rho_{CH}}$ and $\frac{1}{3\rho_{CH}}$, respectively. Neglecting cross-correlation effects, τ_c can be obtained for CH and CH₂ groups from

$$\tau_c = \frac{16\pi^2 r_{C-H}^6}{N_H \gamma_H^2 \gamma_C^2 \hbar^2 \mu_0^2 T_{1DD}} \quad (6.23)$$

with r_{C-H} being the carbon-proton distance taken from the DFT-optimized structure and N_H corresponding to the number of directly bound protons. Assuming very fast internal rotation, a value for τ_c can also be derived from the dipolar T_1 of the methyl carbon:⁵²³

$$\tau_c = 3 \frac{16\pi^2 r_{C-H}^6}{\gamma_H^2 \gamma_C^2 \hbar^2 \mu_0^2 T_{1DD}} \quad (6.24)$$

The experimentally determined T_1 and heteronuclear ¹H¹³C NOE values, T_{1DD} and the resulting τ_c are listed in Table 6.1 together with the literature values for T_1 from Ref. 521. Since the values for T_1 in the literature were obtained at a lower field (100 MHz spectrometer), they should be closer to our values for T_{1DD} as there is a significant contribution from chemical shift anisotropy to ¹³C relaxation at higher magnetic fields. Our T_{1DD} values are in the same range as the literature

values.⁵²¹ Since T_{1DD} and τ_c are inversely proportional, it is clear that also the ^{13}C derived τ_c values are similar to the correlation time of 1.72 ps reported earlier.⁵²¹ The τ_c obtained from NOVAS (3.5 ps) is larger than the ^{13}C derived values would suggest. One reason for the discrepancy could be the assumption of isotropic tumbling, which is not fulfilled for molecules with a shape like crotonaldehyde. Further, the NOVAS approach assumes purely dipolar relaxation. Nevertheless, our results show that the correlation time resulting from NOVAS lies in the same range as the ^{13}C relaxation derived values. Looking at τ_c obtained from the methyl ^{13}C relaxation data, it becomes apparent that the assumptions underlying Eq. (6.24) seem to be invalid in our case. Taking the averaged τ_c from the other three carbons of 1.47 ps and applying Eq. (6.16) gives a rough estimate for τ_{jump} of 0.88 ps with an error bar of 0.35 - 2.62 ps. This would indicate that the time scales for overall tumbling and methyl group rotation are similar and therefore the correlation time for methyl rotation needs to be accounted for explicitly.

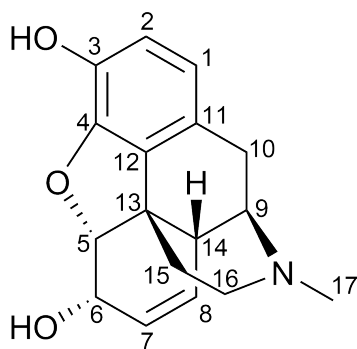
Table 6.1: Summary of the observed longitudinal relaxation times T_1 , the observed heteronuclear ^{13}C NOE, T_{1DD} (the dipolar contribution to T_1) and the corresponding correlation times τ_c for the ^{13}C nuclei of trans-crotonaldehyde with the error bar in brackets. T_{1lit} are the literature values from Ref. 521 measured in DMSO- d_6 on a 100 MHz spectrometer.

Carbon Number	T_1 [s]	η_{obs}	T_{1DD} [s]	T_{1lit} [s] ⁵²¹	τ_c [ps]
1	15.1 ± 2.8	1.42	21.2 ± 3.9	36.2 ± 3.0	6.98 [5.89 – 8.56]
2	16.2 ± 3.4	0.87	37.1 ± 7.8	22.5 ± 1.5	1.31 [1.08 – 1.65]
3	18.0 ± 2.6	1.19	30.2 ± 4.3	28.7 ± 2.9	1.57 [1.38 – 1.84]
4	15.4 ± 2.6	0.85	36.2 ± 6.2	25.7 ± 3.8	1.54 [1.31 – 1.85]

6.3.3 Diastereotopic Assignment of Methylene Groups in Morphine (3)

In addition to the examples shown above, a NOESY spectrum of morphine (Scheme 6.2) with a mixing time of 3 s was recorded in CDCl_3 with approximately 10 % CD_3OD . This molecule is more flexible than **1** and also contains a methyl group. Therefore, the NOESY spectrum of **3** is more challenging to fit with NOVAS than the previously discussed examples. With this system, it is also possible to investigate the capability of NOVAS to identify the correct stereospecific assignment of methylene protons. A conformational search was performed using the ETKDG⁸⁹ method of RDKit.²⁶³ After DFT optimization and Boltzmann-weighting, two conformers remained that significantly contribute to the ensemble, a major one contributing approximately 95 % and a minor one that contributes approximately 5 %. Since morphine has three methylene groups, there are eight possibilities for the diastereotopic assignment. Fitting these eight possible assignments to the experimental NOESY spectrum led to the eight difference spectra shown in Figure 6.6. Based on wSSR, the first assignment fits best (top left of Figure 6.6 with wSSR of 849). For this assignment the simulated coupling patterns resulting from the DFT calculated $J_{\text{H-H}}$ values are in agreement

with the experimental ^1H spectrum of **3** and were used as verification of the correct assignment (Figure 6.7). A comparison of the correlation times found by NOVAS for the eight different assignments shows that all lie in a narrow range between 43.9 and 45.1 ps (Table 6.2). This can be seen as an indication for the robustness of NOVAS.



morphine **3**

Scheme 6.2: Chemical structure of morphine (**3**). In case methylene groups, “a” is assigned to the more downfield ^1H chemical shift and “b” to the more upfield chemical shift.

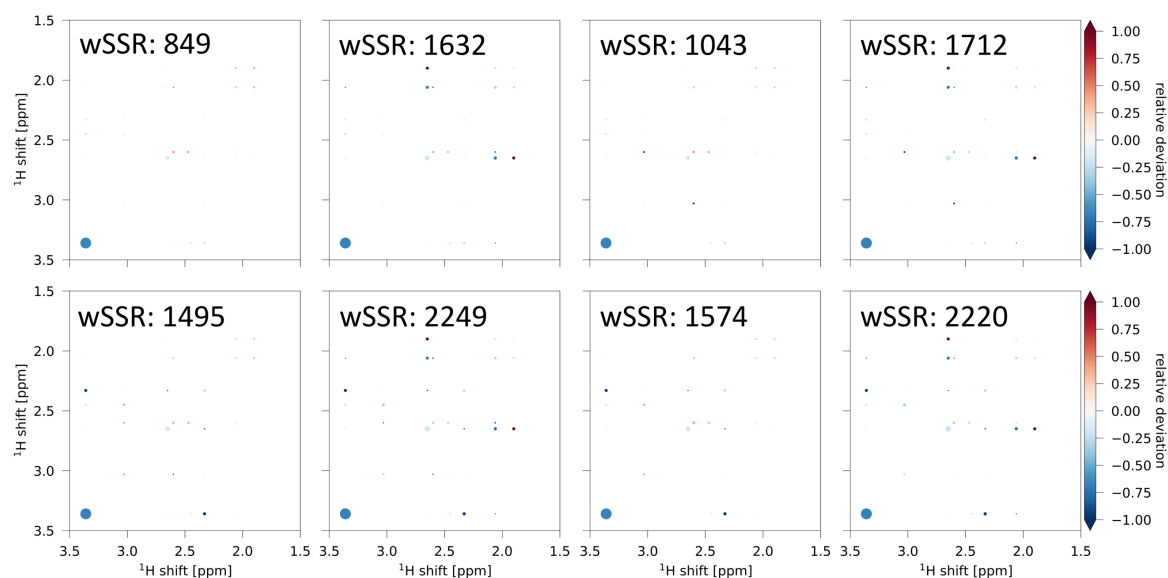


Figure 6.6: Difference between the calculated NOESY spectra for the eight possible diastereotopic assignments of morphine and the experimental spectrum from 1.5 – 3.5 ppm. The corresponding wSSR values are given in the plots. Values of τ_c , τ_{jump} as well as of the rMAD of the diagonal- and cross-peaks are given in Table 6.2.

Table 6.2: Possible stereospecific assignments of **3** with the values obtained from the NOVAS approach. The stereospecific assignment is coded in the following way: proR followed by proS, with "a" being the CH₂ proton at lower field.

Stereospecific Assignment	τ_c [ps]	τ_{jump} [ps]	rMAD diagonal-peaks [%]	rMAD cross-peaks [%]
10ab, 15ab, 16ab	44.1	2.1	19.0	21.0
10ab, 15ba, 16ab	43.9	1.7	18.8	47.1
10ab, 15ab, 16ba	44.8	2.2	18.7	33.3
10ab, 15ba, 16ba	45.1	1.7	18.2	53.0
10ba, 15ab, 16ab	44.1	1.9	21.4	44.2
10ba, 15ba, 16ab	43.9	1.4	21.1	68.9
10ba, 15ab, 16ba	44.3	1.8	21.7	46.7
10ba, 15ba, 16ba	44.4	1.3	21.0	65.5

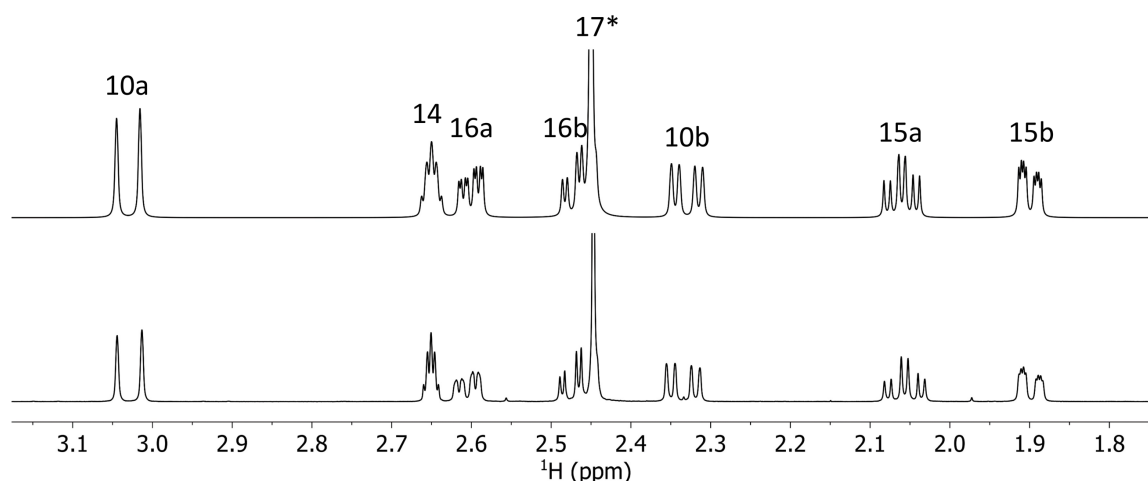


Figure 6.7: Part of the experimental ¹H spectrum (bottom) of morphine in CDCl₃ with approx. 10% CD₃OD and the simulated multiplicity patterns for the methylene protons based on DFT J-coupling calculations (top). Both conformations present in the ensemble were considered in the calculation. The agreement between simulated and experimental coupling patterns was used to verify the correct diastereotopic assignment.

The Akaike information criterion (AIC) can be used to assess whether the experimental data are statistically better represented by also including the minor conformer or if the more complex model is unjustified.⁵²⁴ In case of a least square fit, the following simplified equation can be applied:^{525–527}

$$AIC_i = n * \log SSR_i + 2k' \quad (6.25)$$

SSR_i is the sum of squared residuals of model i (in our case a model only considering a single conformation, or a model with two conformers with weights of 0.95 and 0.05), n is the number of data points (in our case the number of peak volumes that could be successfully integrated) and k' is the number of model parameters (three in case when only using the major conformation (τ_c ,

τ_{jump} and A) and four when using both (plus one for increasing complexity of the model). Since the wSSR with only one conformation is lower (844) compared to the wSSR when both structures are considered (849) and the former model has one parameter less, it is clear that the combined model also has the higher AIC score. An AIC of 653 is obtained for the model using only the major conformation, whereas an AIC of 655 is obtained for the two-conformer model (note that in the calculation of the AIC wSSR was used instead of SSR). This indicates that there is no justification to use a more complex model. The fit of the calculated NOESY for the major conformation is excellent with a rMAD of 18.8 % for the diagonal-peaks and 21.0 % for the cross-peaks. Interestingly, the diagonal-peak of H9 at 3.36 ppm relaxes significantly more slowly than anticipated from our theoretical model (Figure 6.8).

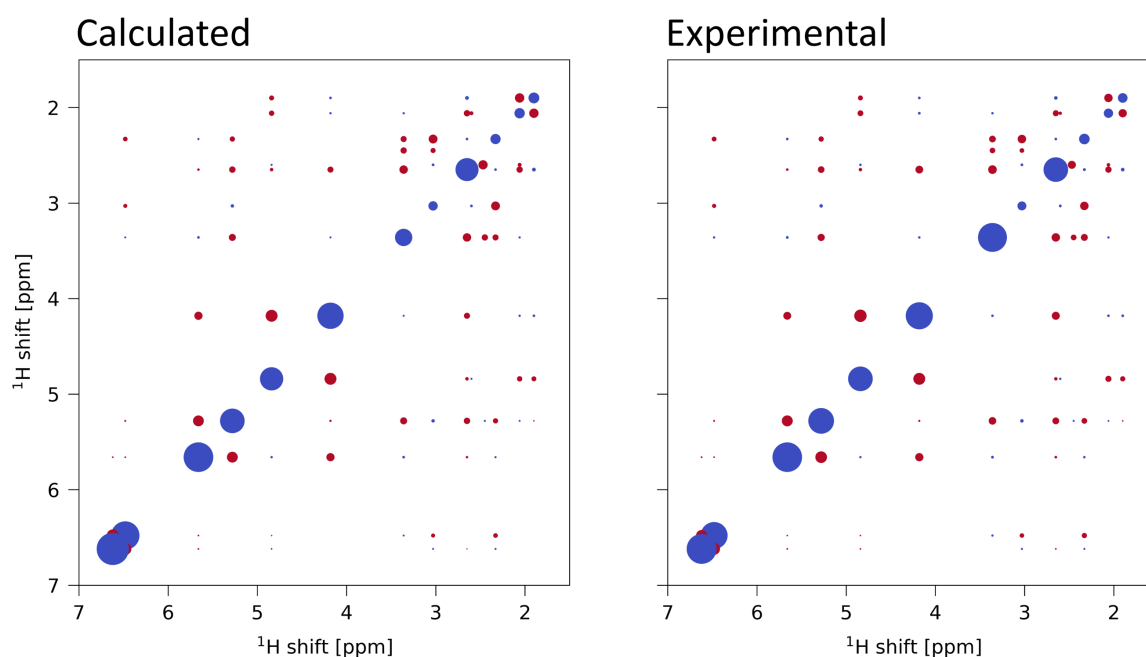


Figure 6.8: Fitted and experimental NOESY spectra of morphine. For the fit, only the major conformation was used. This gave a wSSR of 844 with a τ_c of 44.0 ps and a τ_{jump} of 2.3 ps.

6.3.4 Identifying the Correct Stereospecific Assignment for Androstenedione (**4**) out of 256 Possibilities

Next, we investigated if stereospecific assignment based on the NOVAS protocol is also possible for a more complex case. Androstenedione (**4**) has eight methylene groups, thus there are 256 possible assignments. The ^1H spectrum shows many overlapping signals, which prevented us from unambiguously assigning and integrating a large number of cross-peaks. To test if we can successfully apply the NOVAS protocol also for this difficult case, a NOESY spectrum of androstenedione with a mixing time of 3 s was recorded in chloroform. For androstenedione, six conformations were found in the conformational search. They could be summarized into two clusters with virtually identical members with a difference in energy of 0.3 kJ/mol in both sub-

ensembles. The two remaining conformations have a Boltzmann weight of 91:9. The NOVAS approach was applied for all 256 possible diastereotopic assignments. Out of these, the lowest wSSR obtained was 585 with a τ_c of 23.1 ps, a τ_{jump} of 1.7 ps, rMAD of 12.6 % for the diagonal-peaks and 28.6 % for the cross-peaks of the NOESY spectrum. The second-best fit had a wSSR of 594. The other assignments gave wSSRs >850. The two best fits differ only in the stereospecific assignment of methylene group H2. Due to signal overlap, only the cross-peaks to H1eq could be integrated for both H2 protons, with H2a having the larger cross-peak intensity by a factor of 1.6.

The AIC can also be used for the ranking of relative model probabilities (w_i), i.e., to find a weight of evidence in favor of a certain model i being the best model among the R models under consideration. In case of diastereotopic assignment, all possible models are known *a priori* and the model probabilities can readily be calculated as follows:

$$w_i = \frac{e^{-\frac{1}{2}AIC_i}}{\sum_{r=1}^R e^{-\frac{1}{2}AIC_r}} \quad (6.26)$$

In the case of androstenedione, this gives a likelihood of 66.0 % for the most probable assignment and 34.0 % for the second most probable. The stereospecific assignment of all other methylene groups apart from H2 are assigned with very high likelihood (>99.9 %) in favor of the assignment shown in Figure 6.9. The most probable stereospecific assignment given the NOESY data in CDCl₃ is in agreement with the literature.^{528,529}

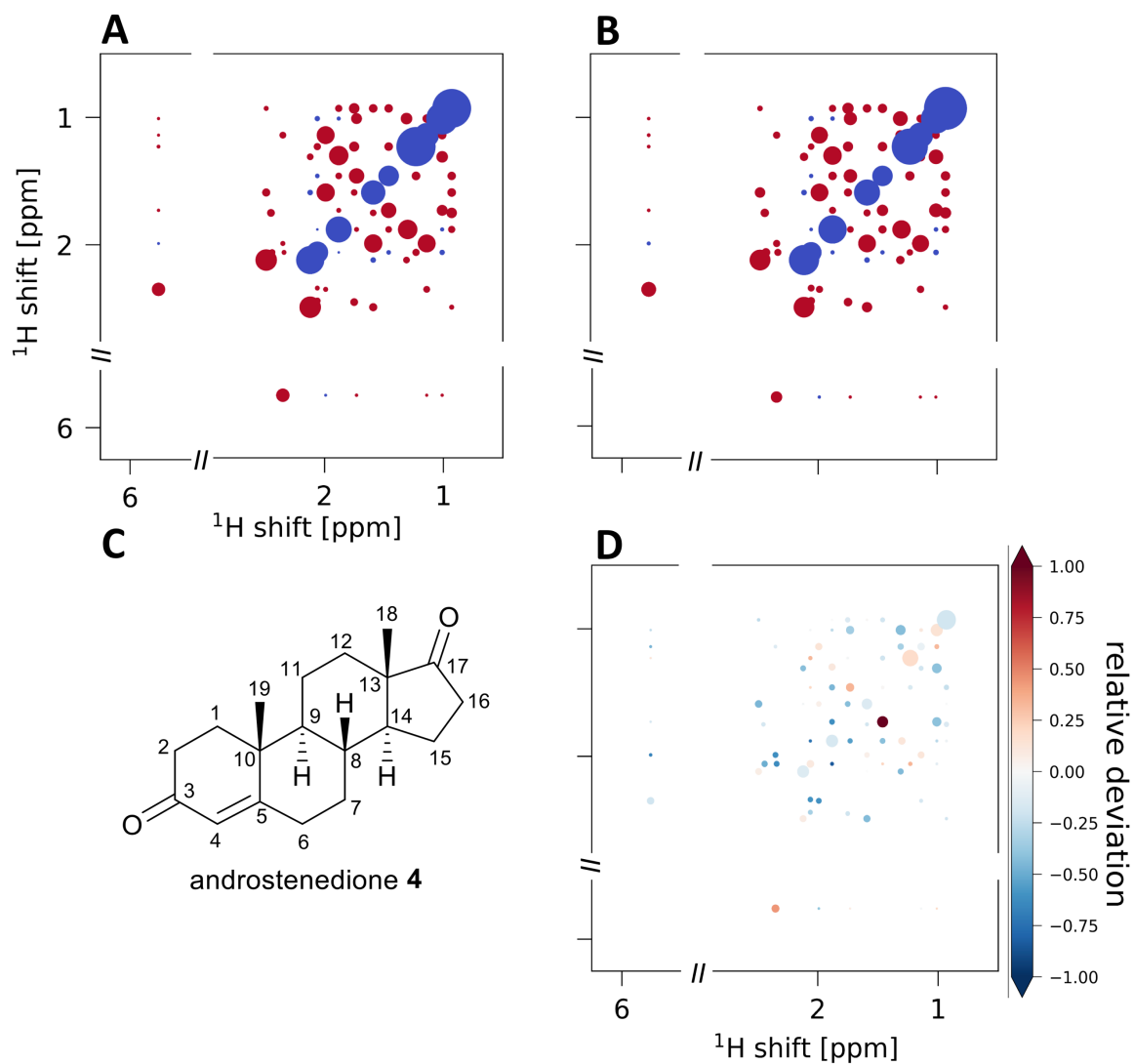


Figure 6.9: Fitted NOESY spectrum for the most probable stereospecific assignment (A) and experimental spectrum (B) of androstenedione (4) with a mixing time of 3 s recorded in CDCl_3 . The chemical structure is shown in C. The difference between experimental and calculated spectrum is shown in D. The diastereotopic assignment is as follows: (equatorial position followed by axial position with "a" being the CH_2 proton at lower field): H1ab, H2ba, H6ba, H7ab, H11ab, H12ab, H15ab and H16ab. Empty regions in the spectrum were cut out for clarity.

6.3.5 Differentiation between ephedrine (5) and pseudoephedrine (6)

Lastly, NOESY spectra of ephedrine and pseudoephedrine were recorded with a mixing time of 3 s in CDCl₃ to assess if by fitting the experimental NOESY spectra in combination with the AIC we are able to differentiate between these flexible diastereomers. This is a very challenging case since several conformations significantly contribute to the ensemble of **5** and **6**. Our fitting procedure using a fixed τ_{jump} of 0.33 ps gave wSSRs of 391.5 and 420.5 when fitting the experimental NOESY spectrum of **5** to the Boltzmann-weighted ensemble of ephedrine and pseudoephedrine, respectively. Doing the same for the experimental NOESY spectrum of pseudoephedrine gave wSSRs of 316.6 and 337.9 for ephedrine and pseudoephedrine, respectively. Both models fit better to the experimental spectrum of ephedrine. A rMAD of approx. 80 % indicates that the fitting does not work well. One possible reason for this could be that the hydroxy and the amine protons were not exchanged with deuterium before the experiment and exchange with the solvent led to magnetization transfer that cannot be modelled using NOVAS. Secondly, the aromatic signals are not well dispersed and partially overlap with the residual proton signal of CDCl₃. Because of that, there are simply not enough NOE volumes to differentiate between the two diastereomers. If the experiment is repeated, it could be worthwhile to record the spectra in a different solvent (e.g., nitrobenzene-d₅) to increase dispersion of the aromatic protons. In addition, the compounds should be first dissolved in CD₃OD to exchange the amine and hydroxy proton with deuterium.

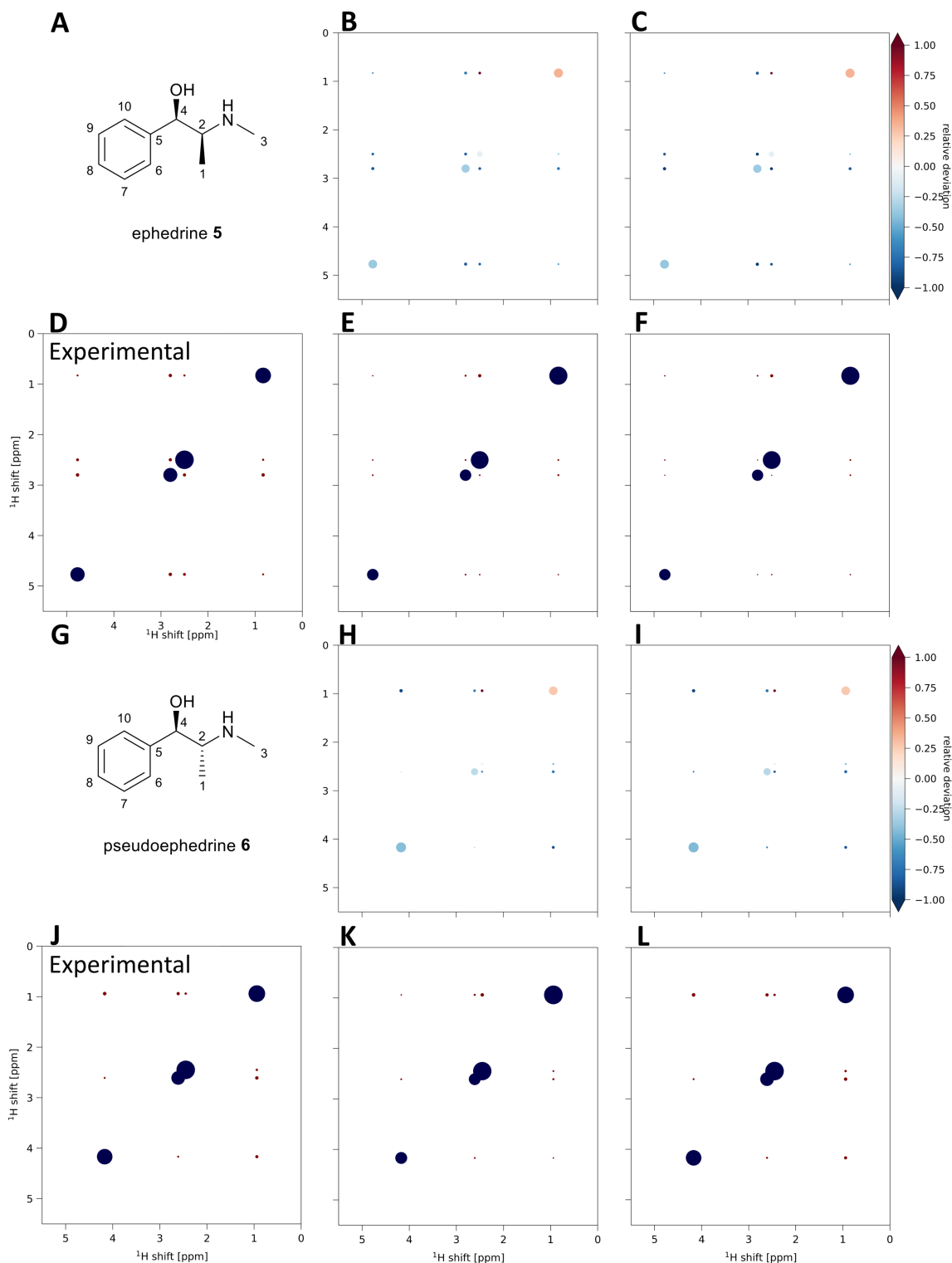


Figure 6.10: **A:** Chemical structure of ephedrine (**5**). **B:** The difference between experimental NOESY spectrum of **5** and the spectrum obtained from the calculated ensemble of **5**. **C:** The difference between experimental NOESY spectrum of **5** and the spectrum obtained from the calculated ensemble of **6**. **D:** Experimental NOESY volumes of **5**. **E:** Calculated NOESY spectrum obtained from the NOVAS approach of the ensemble of **5**. **F:** Calculated NOESY spectrum obtained from the NOVAS approach of the ensemble of **6**. **G:** Chemical structure of pseudoephedrine (**6**). **H:** The difference between experimental NOESY spectrum of **6** and the spectrum obtained from the calculated ensemble of **5**. **I:** The difference between experimental NOESY spectrum of **6** and the spectrum obtained from the calculated ensemble of **6**. **J:** Experimental NOESY volumes of **6**. **K:** Calculated NOESY spectrum obtained from the NOVAS approach of the ensemble of **5**. **L:** Calculated NOESY spectrum from the NOVAS approach of the ensemble of **6**.

6.4 Conclusion

In this study, we recorded NOESY spectra beyond the linear region of the NOE build-up curve for different test molecules. The long mixing times lead to higher intensity of the cross-peaks in the NOESY spectra as well as to the build-up of indirect NOEs, which contain valuable structural information. We showed that it is possible to fit these experimental NOESY spectra with our newly developed NOVAS protocol based on DFT-optimized structures, a global correlation time τ_c , a scaling factor A , and if needed, a local τ_{jump} to account for the fast internal motion of methyl groups. With this protocol, no transformation of the volumes into distances is necessary and one can directly compare a computer-generated ensemble to the primary experimental data, i.e., the NOE volumes. The NOVAS approach is extremely sensitive and readily allows the differentiation between different stereospecific assignments. In combination with the AIC, one obtains likelihoods for the different possible assignments as was shown for morphine and androstenedione. Not enough NOE volumes were available for the successful differentiation between the two flexible diastereomers ephedrine and pseudoephedrine. For these compounds, it would be worthwhile recording the NOESY spectra in a different solvent to potentially increase the dispersion of the aromatic protons. This would yield a higher number of usable NOE volumes and the NOVAS approach might succeed also for this extremely challenging case.

6.5 Method Section

Experimental Details

Phase sensitive gradient-enhanced NOESY^{510,530} spectra for strychnine (**1**) (Fluka), *trans*-crotonaldehyde (**2**) (Acros), morphine monohydrate (**3**) (Lipomed AG), androstenedione (**4**) (Fluka), ephedrine (**5**) (Aldrich) and pseudoephedrine (**6**) (Sigma-Aldrich) as well as the ¹³C inversion recovery experiments for *trans*-crotonaldehyde (**2**) were recorded on a Bruker Avance III HD 600 MHz spectrometer equipped with a N₂-cooled Prodigy triple resonance probe with z-gradients. ¹³C inversion recovery experiments were recorded using power gated ¹H broadband decoupling and a recycle delay of 60 s was chosen. Recovery delays and the corresponding integrals are listed in Table 6.3.

Table 6.3: Integrals for carbons 1 – 4 of **2** found in ¹³C inversion recovery experiments with variable recovery delays *t*.

Delay <i>t</i> [s]	1	2	3	4
0.5	-290.1	-631.4	-957.8	-254.2
1	-231.6	-592.7	-820.9	-188.9
3	23.8	-350.0	-650.9	-80.2
5	160.0	-278.0	-435.9	92.5
7	325.7	-41.2	-294.2	192.9
9	413.1	132.6	-125.6	290.5
11	619.4	202.1	11.1	380.0
13	650.5	390.0	120.5	537.1
15	850.9	389.0	299.8	547.4
20	955.7	620.9	472.2	676.1

The {¹H}¹³C heteronuclear NOEs for **2** were measured on a Bruker AVANCE III 600 MHz spectrometer equipped with a He-cooled DCH cryogenic probe with z-gradients using a power gated ¹H broadband decoupled ¹³C spectrum with a recycle delay of 300 s and an excitation pulse of 45 degrees. Even and odd transients were recorded in different memory locations and for every second scan the proton transmitter was moved to 1000 ppm. The observed heteronuclear NOEs are listed in Table 6.1 in the main text.

Concentrations of all samples were between 20 and 50 mM and all spectra were recorded in CDCl₃. In the case of morphine (**3**), approximately 10 % CD₃OD was added to increase solubility and to exchange the hydroxy protons with deuterium (this minimizes magnetization loss due to transfer to water). For ephedrine (**4**) and pseudoephedrine (**5**), a drop of CD₃OD was added to exchange the hydroxy protons partially with deuterium. A mixing time of 1 s was used for the NOESY

spectrum of **1**, whereas mixing times of 3 s were used for the NOESY spectra of **3**, **4**, **5** and **6**, and 5 s for **2**. The time domain in both dimensions of the NOESY spectra was extended to twice its size by zero filling. The baseline was corrected using a polynomial of third order. Processing of the spectra was done in TopSpin 4.1 (Bruker Biospin AG) and MestreNova 14.2 (Mestrelab Research). Peak assignment and partially also volume integration of the diagonal-peaks and cross-peaks of the NOESY spectra was done using NMRFAM-SPARKY.¹⁹⁹ For **1** and **4**, the integration of the NOESY peaks was done with NMRViewJ.⁵³¹

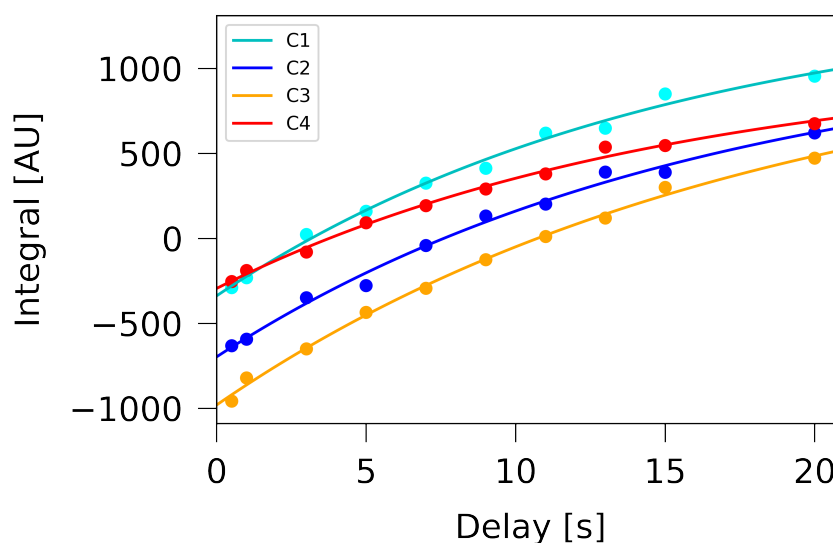


Figure 6.11: Plotted integrals of the inversion recovery experiments for *trans*-crotonaldehyde (**2**) with different delays listed in Table 6.3 to fit the corresponding ^{13}C T_1 times using Eq. (6.20).

Computational Details

3D structures were generated from SMILES strings using the ETKDG conformer generator⁸⁹ of RDKit²⁶³ with an RMSD threshold of 0.1 Å and sampling at maximum 1000 conformers with 20000 attempts. The obtained conformers were pre-optimized using the built-in version of the Merck molecular force field.⁵³² The atoms were reordered such that the attached hydrogen atoms directly follow the heavy atoms. In case of CH₂ groups, the proR proton is always first in order. The structures were then optimized with DFT in vacuum using the quantum chemical package Orca 5.0.1^{248–250} at PB86/def2-tzvp^{255–257} level of theory with resolution of identity using the def2/J auxiliary basis set²⁵⁸ and Grimme's D3BJ dispersion correction.^{264,265} Minima were verified by a subsequent frequency calculation at the same level of theory. In addition, the energy was also calculated using the conductor-like polarizable continuum model (CPCM)²⁶¹ as implicit solvent model for chloroform. The energies of the individual conformers were then computed as the sum of the Gibbs energy obtained from the vacuum calculation and the difference between the final energies of the vacuum and the implicit solvent calculation. Structures that differ less than 0.1 kJ/mol in energy were inspected for being the same minimum by a Python script from Ref.

533. When the RMSD between two conformers was below 0.05 Å, the conformers were considered to be identical and one was removed from the ensemble. If more than one conformation remained, the structures were Boltzmann weighted for further analysis.

The NOVAS protocol was carried out with a Python²⁰¹ script run in a Jupyter Notebook.²⁰² First, the proton xyz coordinates were read using the pandas package²⁰⁶ and the corresponding distance matrix was created. The ¹H chemical shifts were also read to be able to plot the calculated NOESY spectra using the matplotlib package.²⁰³ The NOE volumes were stored in a matrix containing the experimental values and zeros for cross- and diagonal-peaks that were absent or could not be integrated (e.g., because of signal overlap). Peak volumes are listed in the Appendix together with the assigned ¹H chemical shifts according to the numbering of the chemical structures in the main text. Zero entries in the NOESY matrix were then masked using the ma module of numpy.²⁰⁵ The NOESY spectrum was calculated using Eqs. (6.8) - (6.19) in the main text and was optimized based on wSSR using the minimize function from the scipy package²⁰⁷ with the Nelder-Mead algorithm.⁵³⁴ If more than one conformation had to be considered, a NOESY spectrum was calculated for each of them with the same parameters τ_c , A , and, if needed, τ_{jump} and the final spectrum was obtained by taking the average of the NOE volumes based on the Boltzmann weights.

6.7 Appendix

Table A6.1 Integrated volumes from the NOESY spectrum of **1** with a mixing time of 1 s recorded in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1	H1	-2043
H2	H1	152.2
H15 proR	H1	-1.161
H1	H15 proR	-3.241
H2	H2	-2011
H3	H2	57.90
H2	H3	42.36
H3	H3	-2270
H4	H3	104.9
H3	H4	64.71
H4	H4	-1486
H8	H8	-2471
H11 proR	H8	69.37
H8	H11 proR	91.21
H12	H8	11.80
H8	H12	10.10
H13	H8	30.21
H8	H13	31.54
H18 proR	H8	-14.23
H8	H18 proR	-20.05
H18 proS	H8	136.6
H8	H18 proS	212.9
H21	H8	6.264
H8	H21	5.702
H11 proR	H11 proR	-2471
H11 proS	H11 proR	623.4
H11 proR	H11 proS	623.6
H23 proS	H11 proR	3.430
H11 proR	H23 proS	11.59
H12	H12	-2647
H13	H12	158.3
H12	H13	156.4
H15 proS	H12	-3.813
H12	H15 proS	-1.317
H23 proR	H12	15.80
H12	H23 proR	227.1
H4	H13	5.818
H13	H13	-2665
H14	H13	144.0
H13	H14	144.5
H15 proR	H13	-19.75
H13	H15 proR	-23.83
H15 proS	H13	99.91
H13	H15 proS	174.5
H23 proR	H13	-0.901
H13	H23 proR	-0.216
H15 proR	H14	66.94
H14	H15 proR	91.25
H15 proS	H14	51.84
H14	H15 proS	73.49

H20 proS	H14	14.56
H14	H20 proS	13.01
H21	H14	7.572
H14	H21	2.862
H23 proR	H14	77.01
H14	H23 proR	101.9
H23 proS	H14	-0.201
H14	H23 proS	10.25
H15 proR	H15 proR	-1663
H15 proS	H15 proR	580.3
H15 proR	H15 proS	590.1
H16	H15 proR	88.61
H15 proR	H16	68.00
H20 proR	H15 proR	-26.69
H15 proR	H20 proR	-28.31
H20 proS	H15 proR	172.0
H15 proR	H20 proS	164.1
H15 proS	H15 proS	-1710
H16	H15 proS	91.30
H15 proS	H16	71.22
H20 proR	H15 proS	1.960
H15 proS	H20 proR	1.608
H20 proS	H15 proS	-23.73
H15 proS	H20 proS	-22.71
H1	H16	187.4
H16	H16	-2468
H18 proR	H18 proR	-2045
H18 proS	H18 proR	495.5
H18 proR	H18 proS	564.0
H20 proR	H18 proR	22.23
H18 proR	H20 proR	17.70
H18 proS	H18 proS	-1941
H20 proS	H18 proS	7.030
H18 proS	H20 proS	4.9863
H21	H18 proS	19.44
H18 proS	H21	12.68
H20 proR	H20 proR	-2006
H20 proS	H20 proR	608.9
H21	H20 proR	167.4
H20 proR	H21	96.16
H20proR	H20 proS	591.3
H20 proS	H20 proS	-2000
H21	H20 proS	-3.073
H20 proS	H21	-3.852
H21	H21	-2368
H23 proR	H21	21.52
H21	H23 proR	50.72
H23 proS	H21	87.99
H21	H23 proS	164.4
H23 proR	H23 proR	-1719
H23 proS	H23 proS	-1884

Table A6.2: ¹H chemical shifts of **1** referenced to internal TMS measured in CDCl₃ on a 600 MHz spectrometer.

Atom	¹ H chemical shift [ppm]
H1	7.18
H2	7.10
H3	7.26
H4	8.10
H8	3.87
H11 proR	2.67
H11 proS	3.10
H12	4.29
H13	1.29
H14	3.16
H15 proR	2.37
H15 proS	1.48
H16	4.00
H17*	1.91
H18 proR	3.26
H18 proS	2.90
H20 proR	2.77
H20 proS	3.74
H21	5.94
H23 proR	4.07
H23 proS	4.16

Table A6.3: Integrated volumes from the NOESY spectrum of **2** with a mixing time of 5 s recorded in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-24218
H1Me	H2	140.87
H2	H1Me	146.87
H1Me	H3	220.35
H3	H1Me	235.85
H1Me	H4	4.74
H4	H1Me	5.92
H2	H2	-9757.0
H2	H4	451.36
H4	H2	465.81
H3	H3	-10049
H3	H4	28.45
H4	H3	26.00
H4	H4	-9902.8

Table A6.4: ¹H chemical shifts of **2** referenced to internal TMS measured in CDCl₃ on a 600 MHz spectrometer.

Atom	¹ H chemical shift [ppm]
H1Me	2.03
H2	6.88
H3	6.15
H4	9.50

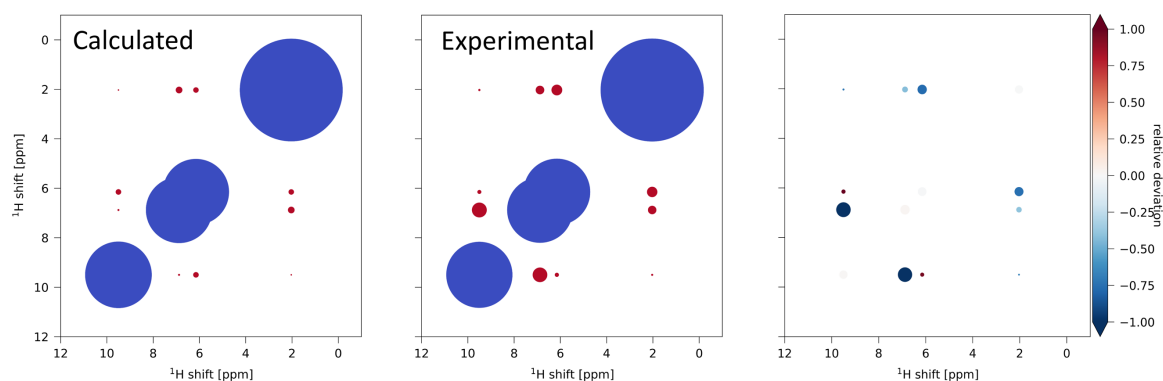


Figure A6.1: Calculated NOESY spectrum of *cis*-crotonaldehyde giving a *w*SSR = 1289 (*w*SSR = 280 for *trans*-crotonaldehyde) (left), the schematic representation of the experimental NOESY spectrum of ca. 50 mM *trans*-crotonaldehyde in CDCl₃ (middle) and the difference between experimental NOESY spectrum for *trans*-crotonaldehyde and calculated NOESY spectrum of *cis*-crotonaldehyde (right).

Table A6.5: Integrated volumes from the NOESY spectrum of **3** with a mixing time of 3 s recorded in CDCl₃ with approximately 10 % CD₃OD.

Resonance 1	Resonance 2	Volume [AU]
H1	H1	-11600
H1	H2	2010
H2	H1	1740
H1	H7	7.26
H7	H1	4.37
H1	H8	13.8
H1	H9	-28.5
H9	H1	-28.4
H1	H10 proR	164
H10 proR	H1	204
H1	H10 proS	241
H10 proS	H1	300
H2	H2	-14400
H2	H7	9.65
H7	H2	7.85
H5	H1	4.52
H5	H2	5.23
H5	H5	-9600
H5	H6	2220
H6	H5	2020
H5	H7	-26.7
H5	H14	99.4
H14	H5	93.5
H5	H15 proR	386
H15 proR	H5	437
H5	H15 proS	264
H15 proS	H5	292
H5	H16 proR	-34.1
H16 proR	H5	-30.2
H6	H6	-11700
H6	H7	893
H7	H6	801
H6	H8	19
H6	H9	-34.9
H9	H6	-39.4

H6	H14	772
H14	H6	836
H6	H15 proR	-48.7
H15 proR	H6	-41.6
H6	H15 proS	-44.5
H15 proS	H6	-48.2
H7	H7	-14100
H7	H8	1680
H8	H7	1940
H7	H9	-52.3
H9	H7	-55.8
H7	H10 proS	-40.6
H10 proS	H7	-33.9
H7	H14	39.8
H14	H7	43.6
H8	H8	-10500
H8	H9	657
H9	H8	732
H8	H10 proR	-98.5
H10 proR	H8	-103
H8	H10 proS	306
H10 proS	H8	373
H8	H14	540
H14	H8	572
H9	H9	-13400
H9	H10 proS	548
H10 proS	H9	595
H9	H14	996
H14	H9	934
H9	H15 proR	-32.6
H15 proR	H9	-27.9
H9	H17Me	353
H17Me	H9	371
H10 proR	H2	-20.4
H10 proR	H10 proR	-1080
H10 proR	H10 proS	961
H10 proS	H10 proR	947
H10 proR	H16 proR	-48.5
H16 proR	H10 proR	-52.3
H10 proR	H17Me	226
H10 proS	H2	-16.3
H10 proS	H10 proS	-1570
H10 proS	H14	-60
H14	H10 proS	-43.8
H14	H2	7.01
H14	H14	-9560
H14	H15 proR	426
H15 proR	H14	474
H14	H15 proS	-107
H15 proS	H14	-114
H15 proR	H8	-10.6
H15 proR	H15 proR	-1180
H15 proR	H15 proS	924
H15 proS	H15 proR	899
H15 proR	H16 proR	104

H16 proR	H15 proR	86
H15 proS	H8	4.91
H15 proS	H15 proS	-1280
H16 proR	H16 proR	-1050
H16 proS	H16 proR	838
H17Me	H8	-17.1

Table A6.6: ^1H chemical shifts of **3** referenced to internal TMS measured in CDCl_3 with approximately 10 % CD_3OD on a 600 MHz spectrometer.

Atom	^1H chemical shift [ppm]
H1	6.48
H2	6.62
H5	4.84
H6	4.18
H7	5.66
H8	5.28
H10 proR	3.03
H10 proS	2.33
H14	2.65
H15 proR	2.06
H15 proS	1.90
H16 proR	2.60
H16 proS	2.47
H17Me	2.45

Table A6.7: Integrated volumes from the NOESY spectrum of **4** with a mixing time of 3 s recorded in CDCl_3 .

Resonance 1	Resonance 2	Volume [AU]
H1eq	H1eq	-3306.0
H1eq	H2eq	256.89
H2eq	H1eq	279.30
H1eq	H2ax	341.30
H2ax	H1eq	444.09
H1eq	H9	-136.46
H9	H1eq	-149.49
H1eq	H11ax	-72.552
H11ax	H1eq	-71.548
H1eq	H12eq	-41.683
H12eq	H1eq	-68.611
H1eq	H19Me	220.26
H19Me	H1eq	219.60
H1ax	H4	30.018
H4	H1ax	40.517
H1ax	H9	1215.8
H9	H1ax	1343.4
H4	H4	-11157
H4	H6eq	1593.9
H6eq	H4	884.57
H4	H7eq	-58.306
H7eq	H4	-31.803
H4	H7ax	47.946
H7ax	H4	27.160
H4	H9	39.282

H9	H4	24.395
H4	H19Me	44.561
H6eq	H7eq	290.23
H7eq	H6eq	285.18
H6eq	H7ax	297.01
H7ax	H6eq	310.34
H6ax	H8	431.31
H8	H6ax	447.71
H7eq	H7ax	2116.9
H7ax	H7eq	2150.6
H7ax	H7ax	-5028.0
H7ax	H9	388.37
H9	H7ax	325.99
H7ax	H14	543.71
H14	H7ax	653.07
H8	H15ax	394.79
H15ax	H8	281.89
H8	H18Me	865.57
H18Me	H8	854.10
H8	H19Me	671.98
H9	H9	-7252.9
H9	H12eq	-84.380
H12eq	H9	-83.359
H9	H14	1546.6
H14	H9	1591.2
H11eq	H11ax	1289.9
H11ax	H11eq	863.97
H11eq	H12eq	237.25
H12eq	H11eq	278.14
H11ax	H12eq	407.53
H12eq	H11ax	438.81
H11ax	H18Me	546.34
H18Me	H11ax	562.93
H11ax	H19Me	569.64
H19Me	H11ax	550.41
H12eq	H12eq	-6439.1
H12eq	H12ax	2587.1
H12ax	H12eq	2558.2
H12eq	H18Me	291.74
H18Me	H12eq	307.24
H14	H16ax	431.31
H16ax	H14	421.97
H15eq	H15ax	2363.4
H15ax	H15eq	2362.9
H15ax	H15ax	-5397.0
H15ax	H16eq	699.40
H16eq	H15ax	726.4
H15ax	H16ax	-149.04
H16ax	H15ax	-153.02
H15ax	H18Me	865.57
H18Me	H15ax	455.21
H16eq	H16ax	3328.2
H16ax	H16eq	3311.3
H16eq	H18Me	150.05
H18Me	H16eq	140.35

H16ax	H16ax	-7367.6
H18Me	H18Me	-15355
H19Me	H19Me	-10750

Table A6.8: ^1H chemical shifts of **4** referenced to internal TMS measured on a 600 MHz spectrometer in CDCl_3 .

Atom	^1H chemical shift [ppm]
H1eq	2.06
H1ax	1.73
H2eq	2.34
H2ax	2.44
H4	5.76
H6eq	2.35
H6ax	2.45
H7eq	1.99
H7ax	1.14
H8	1.75
H9	1.01
H11eq	1.73
H11ax	1.46
H12eq	1.88
H12ax	1.30
H15eq	1.99
H15ax	1.59
H16eq	2.49
H16ax	2.12
H18Me	0.93
H19Me	1.23

Table A6.9: Integrated volumes from the NOESY spectrum of **5** with a mixing time of 3 s recorded in CDCl_3 .

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-5570
H1Me	H2	189
H2	H1Me	189
H1Me	H3Me	164
H3Me	H1Me	60.7
H1Me	H4	46.5
H4	H1Me	46.9
H2	H2	-4390
H2	H3Me	168
H3Me	H2	164
H2	H4	196
H4	H2	196
H3Me	H3Me	-7890
H3Me	H4	132
H4	H3Me	131
H4	H4	-4630

Table A6.10: ^1H chemical shifts of **5** referenced to internal TMS measured on a 600 MHz spectrometer in CDCl_3 .

Atom	^1H chemical shift [ppm]
H1Me	0.83
H2	2.80
H3Me	2.50
H4	4.77
H6-H10	7.31

Table A6.11: Integrated volumes from the NOESY spectrum of **6** with a mixing time of 3 s recorded in CDCl_3 .

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-1830
H1Me	H2	55.9
H2	H1Me	50.5
H1Me	H3Me	22.6
H3Me	H1Me	23.9
H1Me	H4	46.3
H4	H1Me	63.1
H2	H2	-1210
H2	H4	14.2
H4	H2	12.9
H3Me	H2	26.1
H3Me	H3Me	-2280
H4	H4	-1580

Table A6.12: ^1H chemical shifts of **6** referenced to internal TMS measured on a 600 MHz spectrometer in CDCl_3 .

Atom	^1H chemical shift [ppm]
H1Me	0.94
H2	2.61
H3Me	2.45
H4	4.17
H6-H10	7.32

7 Assignment of Relative Configuration in Linear Chlorinated Diols by Comparison of Experimental and Theoretical Spectroscopic Data

In organic chemistry, the assignment of relative configuration in flexible compounds is still a major challenge, if they cannot be crystallized. In this Chapter, we set out to assign the relative configuration of all eight diastereomers of a flexible trichlorinated-hexa-1-3-diol based on readily available experimental data: ^1H and ^{13}C chemical shifts, NOESY peak volumes and IR spectra. For each diastereomer, these data were compared to the properties of a conformational ensemble obtained from DFT calculations. Since it is rarely the case that all diastereomers are available experimentally, we analyzed the assignment capabilities of the different methods pretending that only experimental data from one of the eight diastereomers is at hand. Based solely on ^{13}C chemical shifts, correct assignment was obtained for six out of eight diastereomers, whereas based solely on ^1H chemical shifts, only one out of eight diastereomers was identified correctly. When ^1H and ^{13}C chemical shift data were combined, seven out of eight diastereomers were assigned correctly. Since in this study, all eight diastereomers were experimentally available, this additional information can be used in the assignment procedure. By collectively matching the ^{13}C data, an overall likelihood of 92.5 % is obtained for the correct assignment of all eight diastereomers. Collectively matching the ^1H data yields the highest likelihood for the correct assignment, although with only 10.3 %. Collectively matching the combined ^{13}C and ^1H data results in a nearly perfect differentiability between the diastereomers with a likelihood for the correct assignment of 97.6 %.

The most sensitive discrimination using a single method, with likelihoods of over 90 % for each individual stereoisomer, was achieved by comparison of NOESY spectra recorded beyond the linear build-up regime with the NOVAS approach presented in Chapter 6. Using the NOESY data of all eight diastereomers simultaneously, a likelihood for the correct assignment of over 99.9 % was achieved.

When ^1H and ^{13}C chemical shifts were combined with the NOESY data and matched individually to each compound, all eight diastereomers were assigned correctly with a likelihood over 99.6 % for each individual diastereomer. Collectively matching this extended dataset results in a perfect

differentiability between the diastereomers with an overall likelihood of over 99.99 % for the correct assignment of all eight diastereomers.

Next to assignment based on NMR data, also the assignment based on IR spectra was investigated. Five out of seven diastereomers for which experimental data could be obtained were correctly identified based on IR data alone. Combination with the NMR data is currently not possible, since the alignment score cannot readily be transferred into a likelihood that would be needed for this purpose.

7.1 Introduction

The determination of relative configuration is a crucial part in the structure elucidation of every complex organic molecule. Often, the absolute configuration of one center of chirality is known, and the relative configuration directly translates into the absolute configuration of the compound. The gold standard for identification of the absolute configuration is x-ray diffraction analysis.⁹⁶ However, it can be difficult and tedious to obtain suitable crystals and, in many cases, it is not possible at all. The relative configuration of a molecule can be obtained with numerous experimental techniques.⁵³⁵ Often, nuclear magnetic resonance (NMR) spectroscopy is used with a combined analysis of J-couplings and nuclear Overhauser Effect (NOE)-derived distances. More recently, residual dipolar couplings (RDCs) and chemical shieldings calculated with density functional theory (DFT) were also successfully applied for this task.^{536,537} Along with NMR, other spectroscopic methods can be used for diastereotopic assignment. For example, it was recently shown that the information in the fingerprint region of an experimental infrared (IR) spectrum can be used for the assignment of diastereomers by an automatic comparison to calculated spectra of the different diastereomers.^{538,539}

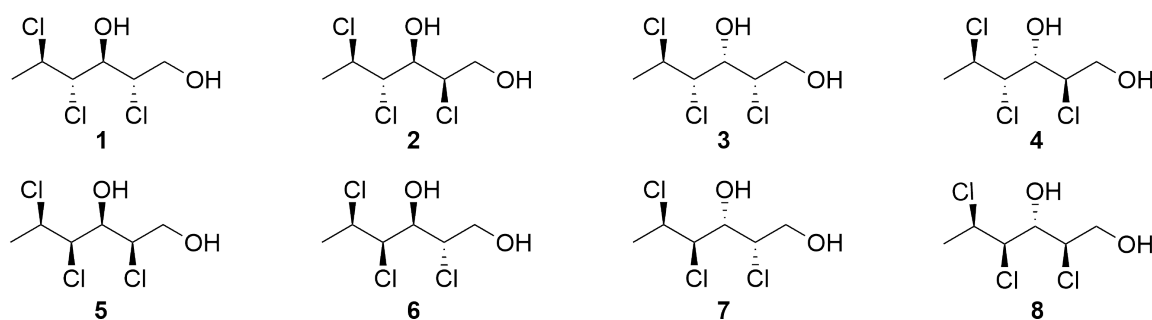
In this chapter, we aim to assign the relative configuration in a set of eight diastereomers based on the comparison between experimental data and the corresponding properties calculated by DFT. The chosen experimental data are readily available and easy to determine experimentally.

Three different methods will be evaluated to obtain the stereospecific assignment. First, ¹H and ¹³C experimental chemical shifts recorded in chloroform-d (CDCl₃) will be compared to calculated chemical shifts obtained by DFT. Here, Akaike's information criterion (AIC)^{524,525,527} is used to assess the likelihood of a given assignment. As a second method, we want to demonstrate that the identification of relative configuration is also possible based on fitting experimental NOE volumes with calculated NOESY spectra based on a set of DFT optimized conformations (NOVAS approach presented in Chapter 6.). The likelihood for a given assignment is again obtained from the AIC. In addition, combination of chemical shift data with NOE volumes is explored to improve the confidence in the assignment. As a third method, we want to obtain the relative configuration by comparing DFT calculated IR spectra with IR spectra recorded in chloroform. The improved IR sequence alignment (IRSA) algorithm is used for this purpose.^{538,540}

The calculation of chemical shieldings and IR spectra with DFT are routine operations nowadays and their background will not be discussed here (for excellent reviews on the topic see Refs. 238, 243 and 541 – 543). The necessary theory for the calculation of NOESY spectra can be found in Chapter 6. Briefly, based on the inter-proton distances in a DFT optimized molecular structure, a

set of proton-proton dipolar auto- and cross-relaxation rates is calculated. With these rates, the NOESY spectrum at a given mixing time t_m is calculated with the NOVAS approach using an overall effective correlation time τ_c and a global scaling factor A . If methyl groups are present in the molecule, an additional correlation time τ_{jump} for fast methyl rotation is introduced. By varying τ_c (τ_{jump}) and A , the difference between calculated and experimental volumes is minimized based on the weighted sum of squared residuals (wSSR), which is used as target function. If more than one conformation is significantly populated in solution, the NOESY spectrum is calculated for each conformation separately using identical values for τ_c (τ_{jump}) and A . Subsequently, the weighted averages of the peak volumes are matched with the experimental data. During this procedure, populations are fixed based on energies calculated on the BP86/def2-tzvp^{255–257} level of theory. Details are given in the methods section. Importantly, experimental mixing times are chosen long enough (3 s in this work) to also allow the build-up of indirect NOE correlations. These indirect NOEs contain valuable additional spatial information about the structure. For three protons, the indirect NOE only contributes significantly to the spectrum when two of the inter-proton vectors are relatively short and form an obtuse angle with respect to each other.⁵⁴⁴

The relative configurations of the eight diastereomers of 2,4,5-trichlorohexane-1,3-diol (Scheme 7.1, **1–8**)⁵⁴⁵ were originally assigned based on analysis of $^3J_{HH}$ and $^{2,3}J_{CH}$ couplings together with qualitative inspection of their NOESY spectra.⁵⁴⁵ The assignment was later verified by x-ray diffraction analysis.⁵⁴⁵ Analysis of proton-proton couplings is a standard task for most organic chemists, but as soon as also J_{CH} couplings have to be considered, experiments become more involved and specialized expertise and careful analysis is necessary. The range for the J_{CH} couplings is generally smaller than for the J_{HH} couplings and in case of $^2J_{CH}$ couplings not only the size but ideally also the sign of the coupling constant needs to be determined. To make use of the experimental J-couplings, they need to be first translated into dihedral angles using a Karplus curve based on reference compounds.⁵⁴⁶



Scheme 7.1: Chemical structure of the eight diastereomers of 2,4,5-trichlorohexane-1,3-diol (**1–8**) labelled in the original publication with numbers 30 – 37.⁵⁴⁵

As the relative configurations have been verified previously with x-ray data, the eight diastereomers are an ideal test case to assess and compare the capabilities of the different approaches presented in this chapter. The open-chain structure presents a particular challenge as the molecules are flexible and more than one conformation potentially contributes to the ensemble in solution. Yet, the compounds are still small enough that DFT calculations with standard functionals and basis sets are affordable.

7.2 Results

7.2.1 Generation of Conformational Ensembles

A conformational ensemble for diastereomers **1** – **8** was generated in order to compute the Boltzmann-weighted properties needed to compare with the experimental data. Conformers were generated from SMILES strings using the KDG conformer generator⁸⁹ of RDKit.²⁶³ No experimental torsion angle preferences were applied because the gauche effect between two neighboring chlorine atoms was not properly taken into account by the underlying SMARTS patterns, and thus the known crystal conformation was not always contained in the generated ensemble. The obtained 3D structures were optimized in vacuo with DFT at the BP86/def2-tzvp²⁵⁵⁻²⁵⁷ level of theory applying the resolution of identity approximation with def2/J²⁵⁸ as auxiliary basis set and Grimme's D3BJ dispersion correction.^{264,265} The energy was also calculated using the conductor-like polarizable continuum model (CPCM)²⁶¹ for chloroform to account for solvation energy. The relative free energies of the conformers were computed as the sum of the Gibbs energy obtained from the frequency calculation and the difference between the final energies obtained in vacuo and with CPCM solvation. Identical conformations were excluded and the remaining conformers were Boltzmann-weighted for further analysis. Calculated datasets based on the known relative configurations of compounds **1** – **8** were labeled **A** – **H**. For successful assignment, experimental data of compound **1** needs to match best with calculated dataset **A**, experimental data of compound **2** with calculated dataset **B**, and so on.

7.2.2 Differentiation of Diastereomers Based on Chemical Shifts

First, we aim to assign the relative configuration in **1** – **8** based on comparison between calculated and measured chemical shifts in chloroform-d (CDCl₃). To this end, ¹H and ¹³C Boltzmann-weighted chemical shieldings were calculated with DFT. The ¹H and ¹³C shieldings were then transformed into chemical shifts using slope and intercept for sp³ carbons / hydrogens bound to sp³ carbons presented in Chapter 5. Since the shieldings of chlorine bound carbons are severely affected by relativistic effects that are not accounted for in ordinary DFT calculations, a small set of ¹³C chemical shifts of 23 compounds resembling **1** – **8** was collected from the literature⁵⁴⁷⁻⁵⁵⁴ to obtain a correction factor for the shifts of the chlorine bound carbons (see Appendix). The experimental proton and carbon chemical shifts of diastereomers **1** – **6** and **8** were used directly from spectra reported by Nilewski *et al.*,⁵⁴⁵ whereas the chemical shifts for **7** were remeasured because the reported shifts were recorded in CD₂Cl₂. The ¹H and ¹³C chemical shifts can be found in Table A7.1 and Table A7.2 in the Appendix, while the averaged chemical shifts obtained from the DFT calculations are listed in Table A7.3 and Table A7.4 in the Appendix.

The Akaike information criterion (AIC) was used to obtain likelihoods of all possible assignments.⁵²⁴ In the least square case, the AIC is simply:⁵²⁵⁻⁵²⁷

$$AIC_i = n \log \sum (\delta_{calc_k} - \delta_{exp_k})^2 \quad (7.1)$$

where n is the number of experimental data points. Using Eq. (7.1), it is possible to calculate the weight of evidence (w_i) in favor of model i being the best model for the situation given the considered R models:

$$w_i = \frac{e^{-\frac{1}{2}(AIC_i - AIC_{min})}}{\sum_{r=1}^R e^{-\frac{1}{2}(AIC_r - AIC_{min})}} \quad (7.2)$$

It is seldom the case that all diastereomers are available in organic synthesis, especially when dealing with complex natural products. Therefore, let us now pretend first that only experimental data from one of the eight diastereomers is available. As can be seen directly from the likelihoods reported in Table 7.1 for the ^{13}C chemical shifts, in six of eight cases, the obtained likelihood was highest for the correct assignment and out of those, three had a likelihood >80 %. For **1**, our method suggests that the dataset **H** agrees best (81.4 %), whereas the correct assignment has a likelihood of only 13.7 %. For **5**, the highest likelihood was observed for the dataset **D** (74.6 %), whereas the correct assignment has a likelihood of 21.7 %. Looking at the ^1H chemical shifts, the obtained likelihood was highest for the correct assignment only in two of eight cases, and out of those, only one had a likelihood >80 % (Table 7.2). In this case, the experimental ^1H chemical shifts of the different diastereomers (**1 – 8**) and the calculated ^1H chemical shifts (**A – H**) agree not well enough to differentiate between diastereomers.

When using the RMSDs of the chemical shift regressions reported in Chapter 5, i.e., 1.12 ppm for ^{13}C and 0.08 ppm for ^1H (only the sp^3 carbons and the hydrogens bound to them), one can combine the AICs of ^1H and ^{13}C by dividing by their corresponding variance:

$$AIC_i = m \log \sum \frac{(^{13}\text{C}\delta_{calc_k} - ^{13}\text{C}\delta_{exp_k})^2}{\text{RMSD}_{theo\ 13\text{C}}^2} + n \log \sum \frac{(^1\text{H}\delta_{calc_k} - ^1\text{H}\delta_{exp_k})^2}{\text{RMSD}_{theo\ 1\text{H}}^2} \quad (7.3)$$

where m and n are the number of experimental data points for ^{13}C and ^1H chemical shifts, respectively.

Table 7.1: Likelihood based on the AICs obtained from the comparison of the experimental ^{13}C chemical shifts for individual compounds **1–8** to the eight calculated datasets **A–H** given in %. The highest likelihood for each experimental dataset is set in bold text.

	A	B	C	D	E	F	G	H
1	13.7	0.5	0.1	0.9	2.1	0.4	0.9	81.4
2	15.7	23.3	1.2	6.5	6.0	5.8	23.0	18.4
3	0.2	0.3	82.1	1.9	1.2	2.7	10.4	1.2
4	0.1	0.0	0.2	96.6	1.2	0.5	0.5	0.6
5	0.1	0.1	0.2	74.6	21.7	1.0	0.7	1.6
6	0.2	0.1	0.2	8.1	1.0	87.4	0.7	2.3
7	0.3	0.3	0.4	3.7	1.4	1.1	62.9	30.0
8	0.1	0.0	0.0	0.1	0.2	0.1	0.9	98.6

Table 7.2: Likelihood based on the AICs obtained from the comparison of the experimental ^1H chemical shifts for individual compounds **1–8** to the eight calculated datasets **A–H** given in %. The highest likelihood for each experimental dataset is set in bold text.

	A	B	C	D	E	F	G	H
1	10.1	6.3	4.8	0.7	68.2	0.7	2.2	7.1
2	20.8	17.6	3.1	0.4	38.8	0.4	4.9	14.1
3	0.1	0.1	94.3	1.2	3.6	0.2	0.2	0.2
4	0.2	0.1	64.6	25.8	6.3	2.6	0.1	0.2
5	0.7	0.6	68.8	3.7	23.4	1.5	0.6	0.9
6	0.5	0.4	20.5	40.2	15.4	22.5	0.2	0.4
7	6.6	4.6	9.4	0.4	21.8	0.2	22.8	34.2
8	5.8	3.5	13.7	0.5	30.8	0.2	15.5	29.9

Combining the ^1H and ^{13}C data improves the assignment confidence of the methodology with seven out of eight diastereomers being correctly assigned (Table 7.3). Out of those, four have a likelihood >80 %. For **1**, the highest likelihood is still observed for dataset **H** (66.7 %), whereas the correct assignment has a likelihood of 16.0 %.

Table 7.3: Likelihood based on the AICs obtained from the comparison of the combined experimental ^1H and ^{13}C chemical shifts for individual compounds **1 – 8** to the eight calculated datasets **A – H** given in %. The highest likelihood for each experimental dataset is set in bold text.

	A	B	C	D	E	F	G	H
1	16.0	0.4	0.0	0.1	16.5	0.0	0.2	66.7
2	24.2	30.4	0.3	0.2	17.3	0.2	8.3	19.1
3	0.0	0.0	>99.9	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.4	99.3	0.2	0.1	0.0	0.0
5	0.0	0.0	1.9	34.4	63.2	0.2	0.0	0.2
6	0.0	0.0	0.2	14.0	0.7	85.14	0.0	0.0
7	0.0	0.1	0.1	0.0	1.2	0.0	57.5	41.0
8	0.0	0.0	0.0	0.0	0.2	0.0	0.5	99.3

Since in our case, all eight diastereomers were experimentally available, we can use this additional information in the assignment procedure. We know that each experimental diastereomer **1 – 8** should correspond to exactly one calculated dataset **A – H**. For eight diastereomers, there are in total 40'320 possible assignments (=8!) to match **1 – 8** to **A – H**. For each of those possible assignments, the corresponding associated likelihoods are multiplied and divided by the sum of all obtained likelihoods. This gives the total likelihood for each of the 40'320 models.

Applying this for ^{13}C , an overall likelihood of 92.5 % is obtained for the correct assignment. Other assignments with likelihoods >0.1 % are listed in Table A7.5 in the Appendix. For the ^1H data, the likelihood for the correct assignment is the highest although only with 10.3 %. Assignments with likelihoods >1 % for ^1H are listed in Table A7.6 in the Appendix. Since the individual likelihoods for the correct assignment of the diastereomers using ^1H chemical shifts were low, this poorer agreement compared to ^{13}C can be expected. Yet, it is notable that the correct assignment had still the highest likelihood and out of the 40'320 possible assignments only a few possibilities remain with likelihoods >1 %.

Combining again the ^{13}C and ^1H data gives a likelihood for the correct assignment of 97.6 %. Assignments with likelihoods >0.1 % are listed in Table A7.7 in the Appendix. Although the ^1H chemical shifts itself are not suitable for the differentiation, they improve the confidence for the correct assignment in combination with the ^{13}C chemical shifts. The likelihood from the combined AIC results in a nearly perfect differentiability between the diastereomers in the comparison of experimental chemical shifts and calculated chemical shieldings.

To conclude, in six out of eight cases, the highest likelihood was obtained for the correct diastereomer using ^{13}C chemical shifts alone whereas this was only twice the case using ^1H chemical shifts. A confident assignment is not possible when only one diastereomer would be available. Combining the ^{13}C and ^1H data increased the differentiability of the method and seven out of eight correct assignments had the highest likelihood. A combination with other experimental data seems to be crucial when the complete experimental data for all diastereomers is not available. In the case , when all eight experimental datasets are available, the eight diastereomers were correctly assigned with high confidence using the ^{13}C chemical shifts data. In contrast, the same procedure with ^1H chemical shift did not result in a single assignment with high likelihood (although the correct assignment had the highest value of all). When combining ^1H and ^{13}C data, the likelihood of the correct assignment increased and an assignment with high confidence is possible.

7.2.3 Differentiation of Diastereomers Based on NOVAS Approach

Next, our recently developed NOVAS protocol (see Chapter 6) was applied to the assignment problem. NOESY spectra for the eight diastereomers were recorded in CDCl_3 with a mixing time of 3 s. The long mixing time leads to larger cross-peak volumes compared to the linear build-up regime and also indirect NOEs potentially contribute to the spectrum, containing additional spatial information about the system under study. A drop of D_2O was added to exchange the hydroxy protons with deuterium. This was done to minimize dipolar relaxation involving the hydroxy group. Calculated NOESY spectra were fitted to the experimental peak volumes as described in Chapter 6 by optimizing for the overall correlation time τ_c , a scaling factor A , and the correlation time of the methyl group rotation τ_{jump} with the NOVAS approach. The obtained weighted sum of squared residuals (wSSR) between the experimental NOESY spectra of diastereomers **1** – **8** and the calculated datasets (**A** – **H**) are listed in Table 7.4. The corresponding ΔAICs are listed in Table 7.5. It can be nicely seen that the lowest (best) values are obtained for the correct assignment. Correlation times between 6.5 and 11.7 ps and values for τ_{jump} between 2.7 and 3.9 ps were obtained for the correct fits between experimental and optimized spectra. In contrast to the assignment based on chemical shifts, the NOVAS procedure works perfectly when we pretend that only a single experimental NOESY spectrum of one diastereomer is available. Likelihoods of >95 % were obtained for each of the correct assignments individually, except for **8H** for which 92.6 % was obtained. The signal-to-noise ratio for the NOESY spectrum of **8** was lowest since only traces of the compound were available. Nevertheless, these results clearly show the power of the NOVAS approach and the ability to make clear and correct distinctions between diastereomers even for flexible compounds. As an illustration, the difference between calculated

and experimental NOESY spectra of **1** with the best (**A**) as well as with the second-best agreement (**F**) is shown in Figure 7.1. The corresponding plots for **2 – 8** can be found in Figure A7.2 – Figure A7.8 in the Appendix.

Making again use of the fact that NOESY data is available for all eight diastereomers, one obtains a likelihood of the correct assignment for all diastereomers of >99.9 % (**1A,2B,3C,4D,5E,6F,7G,8H**). The second-most likely assignment with a likelihood of only 0.02 % is found by exchanging the matches of **7** and **8** (**1A,2B,3C,4D,5E,6F,7H,8G**).

Table 7.4: *wSSRs* obtained from the NOVAS approach when using experimental spectra of diastereomers **1 – 8** together with the corresponding calculated datasets **A – H**. The combination yielding the lowest *wSSR* for each experimental NOESY spectrum is set in bold text.

	A	B	C	D	E	F	G	H
1	12.3	60.9	53.1	105.2	162.9	48.2	116.6	78.6
2	354.1	18.0	413.7	247.7	150.4	163.8	138.2	394.0
3	575.0	682.1	153.0	186.2	703.6	497.9	549.7	479.3
4	357.6	342.9	155.1	70.3	184.1	248.8	166.4	212.1
5	28.8	22.8	34.4	45.2	6.6	27.4	22.2	21.7
6	93.8	126.6	99.7	57.8	296.5	17.2	144.5	124.2
7	339.9	334.3	74.2	78.3	44.9	239.1	11.2	18.4
8	105.6	134.1	69.9	98.5	72.9	116.8	67.4	42.7

Table 7.5: ΔAIC values (= $AIC - AIC_{min}$) obtained from the *wSSRs* of the NOVAS approach between experimental diastereomers **1 – 8** and calculated datasets **A – H**. The lowest value for each experimental NOESY spectrum is set in bold text.

	A	B	C	D	E	F	G	H
1	0.0	40.0	36.6	53.7	64.6	34.1	56.3	46.4
2	65.5	0.0	68.9	57.6	46.6	48.5	44.8	67.8
3	45.0	50.8	0.0	6.7	51.9	40.1	43.5	38.8
4	34.2	33.3	16.6	0.0	20.2	26.5	18.1	23.2
5	23.5	19.8	26.4	30.7	0.0	22.7	19.3	19.0
6	38.9	45.8	40.4	27.8	65.4	0.0	48.3	45.4
7	71.7	71.3	39.7	40.8	29.2	64.3	0.0	10.5
8	13.6	17.2	7.4	12.5	8.0	15.1	6.8	0.0

Table 7.6: Likelihoods obtained with the NOVAS procedure for experimental NOESY spectra of diastereomers **1** – **8** with calculated datasets **A** – **H**. The highest likelihood for each experimental NOESY spectrum is set in bold text.

	A	B	C	D	E	F	G	H
1	>99.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	>99.99	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	96.55	3.45	0.00	0.00	0.00	0.00
4	0.00	0.00	0.02	99.96	0.00	0.00	0.01	0.00
5	0.00	0.00	0.00	0.00	99.98	0.00	0.01	0.01
6	0.00	0.00	0.00	0.00	0.00	>99.99	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	99.47	0.53
8	0.10	0.02	2.30	0.18	1.69	0.05	3.03	92.64

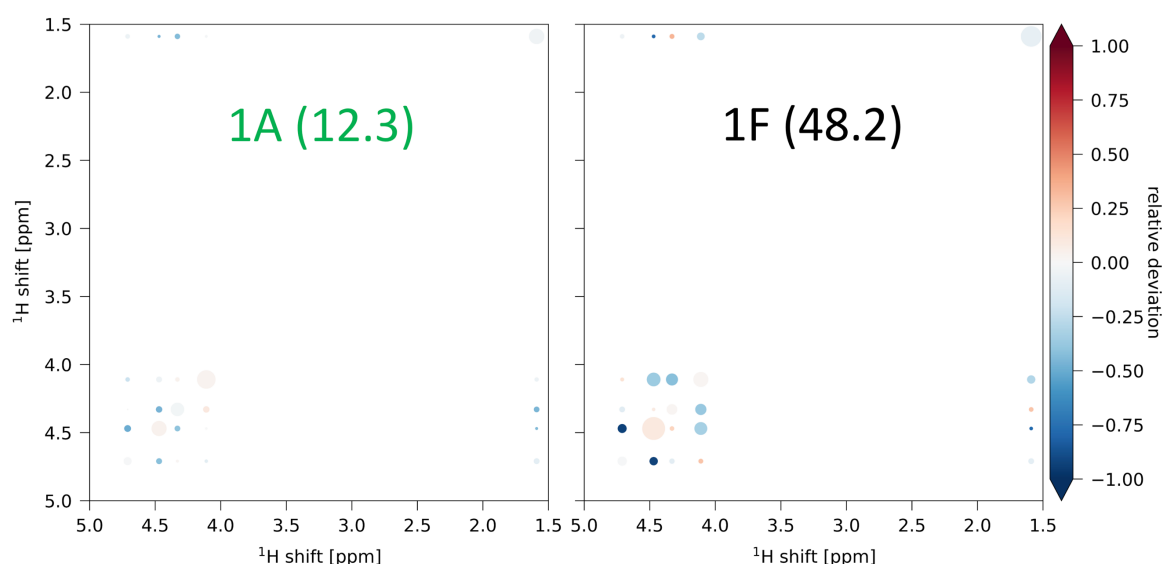


Figure 7.1: Difference between experimental NOESY spectrum of **1** and calculated NOESY spectra for the best fit **A** (green) and the second best fit **F** (black). wSSRs are given in brackets.

7.2.4 Combination of Chemical Shifts with the NOVAS Approach

As already shown for ^1H and ^{13}C chemical shifts, the use of the AIC allows to combine different experimental datasets, if an estimation for the variance is available. As a rough estimate, for the typical error in the NOVAS approach, the wSSRs obtained with the NOVAS protocol for strychnine, *trans*-crotonaldehyde, morphine and androstenedione in Chapter 6 were divided by the number of integrated NOE peaks. The average value of 8.08 gives a crude estimate of the expected RMSD in the NOVAS approach. Combination of the chemical shift data and the NOE data yield an even clearer assignment and each individual diastereomer can be identified correctly with a likelihood >99.6 % if we again pretend that only data for one diastereomer is available (Table 7.7). Making again use of the fact that chemical shift and NOESY data are available for all eight diastereomers, one obtains a likelihood of the correct assignment for all diastereomers of >99.99 %

(**1A,2B,3C,4D,5E,6F,7G,8H**). The second-most likely assignment has a likelihood of only $5.9 \times 10^{-5} \%$ (**1A,2B,3C,4D,5E,6F,7H,8G**).

The assignment confidence of the combined data is impressive. This combined approach of readily available experimental NMR data could also be a promising approach for the configurational assignment of more complicated molecules. Of course, not only chemical shifts and NOE volumes can be combined, but every other kind of experimental data that can be reproduced with computational methods can be used in principle. Among the NMR observables, incorporation of $^3J_{\text{HH}}$ couplings would be the next logical step, but also data from other spectroscopic techniques like IR or Raman would potentially add complementary valuable additional information.

Table 7.7: Likelihoods of the combination of the chemical shift data with the NOVAS procedure for experimental data of diastereomers **1 – 8** with calculated datasets **A – H**. The highest likelihood for each experimental NOESY spectrum is set in bold text.

	A	B	C	D	E	F	G	H
1	>99.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	>99.99	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	>99.99	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	>99.99	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	>99.99	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	>99.99	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	99.61	0.39
8	0.00	0.0	0.00	0.00	0.00	0.00	0.02	99.98

7.2.5 Differentiation of Diastereomers Based on IR Spectra

Lastly, we aimed to assign the relative configuration of the different chlorinated diols based on comparison of experimental and calculated IR spectra. FT-IR spectra for diastereomers **1 – 7** were recorded in chloroform. For compound **8**, not enough substance was available for a solution IR spectrum of decent quality, and thus this compound was excluded. Calculated IR spectra for **1 – 8** were obtained as Boltzmann weighted averages of the IR spectra for each significantly populated conformation. For the comparison, the improved version of the IR spectra algorithm (IRSA) from Böselt *et al.*⁵⁴⁰ was applied to the region of $900 - 1500 \text{ cm}^{-1}$, excluding the region between 1200 and 1240 cm^{-1} which shows a strong absorption band of CHCl_3 . Peak picking in both experimental and calculated spectra was done in an automated fashion. Since the experimental spectra were relatively noisy (due to the limited amount of substance or the limited solubility), the peak assignment for the experimental spectra was checked manually and corrected if needed. The agreement between an experimental and a theoretical IR spectrum is expressed in terms of an

alignment score (the more negative the better, where -1 is the best value).⁵⁴⁰ The score was computed for every combination of theoretical and experimental IR spectrum. In five out of seven diastereomers, the best alignment score was in agreement with the correct assignment (**1A**, **2B**, **5E**, **6F** and **7G**). In case of **3**, the best alignment score (-0.90) was obtained with theoretical IR spectrum **F**, whereas the correct assignment (**3C**), yielded a score of -0.83 (second best score). For diastereomer **4**, the correct assignment (**4E**) was the scored worst (-0.77) whereas a very good agreement was obtained with calculated spectrum **H** (-0.99). The alignment scores for the different assignments of a given experimental spectrum (elements within the same row) are relatively similar. This can be expected since also the experimental IR spectra are highly similar. The alignment between experimental and calculated IR spectra with best and second-best scores are shown in Figure 7.2.

The assignment of diastereomers with IRSA works when the differences in the IR spectra are large enough. In case of highly similar spectra, no clear differentiation is possible. Interestingly, IR could identify the correct configuration of **1**, whereas the identification of **1** was not possible based on chemical shift data. This is a strong indication that the information obtained from IR is at least in some cases complementary to the chemical shift information. At the current development stage, the alignment score obtained from IRSA cannot be transferred into a probability and it is therefore not possible to combine the IR data with the above presented NMR data. This will be the focus of future work.

Table 7.8: Alignment scores between experimental IR spectra for diastereomers **1** – **7** and the corresponding calculated spectra **A** – **H**. The best score is set in bold text.

	A	B	C	D	E	F	G	H
1	-0.86	-0.41	-0.68	-0.80	-0.43	-0.73	-0.75	-0.73
2	-0.51	-0.76	-0.70	-0.66	-0.65	-0.64	-0.67	-0.47
3	-0.32	-0.52	-0.83	-0.58	-0.40	-0.90	-0.43	-0.29
4	-0.91	-0.77	-0.93	-0.77	-0.91	-0.98	-0.97	-0.99
5	-0.33	-0.43	-0.38	-0.31	-0.71	-0.38	-0.39	-0.44
6	-0.72	-0.64	-0.82	-0.66	-0.77	-0.88	-0.87	-0.83
7	-0.71	-0.55	-0.77	-0.57	-0.48	-0.82	-0.88	-0.86

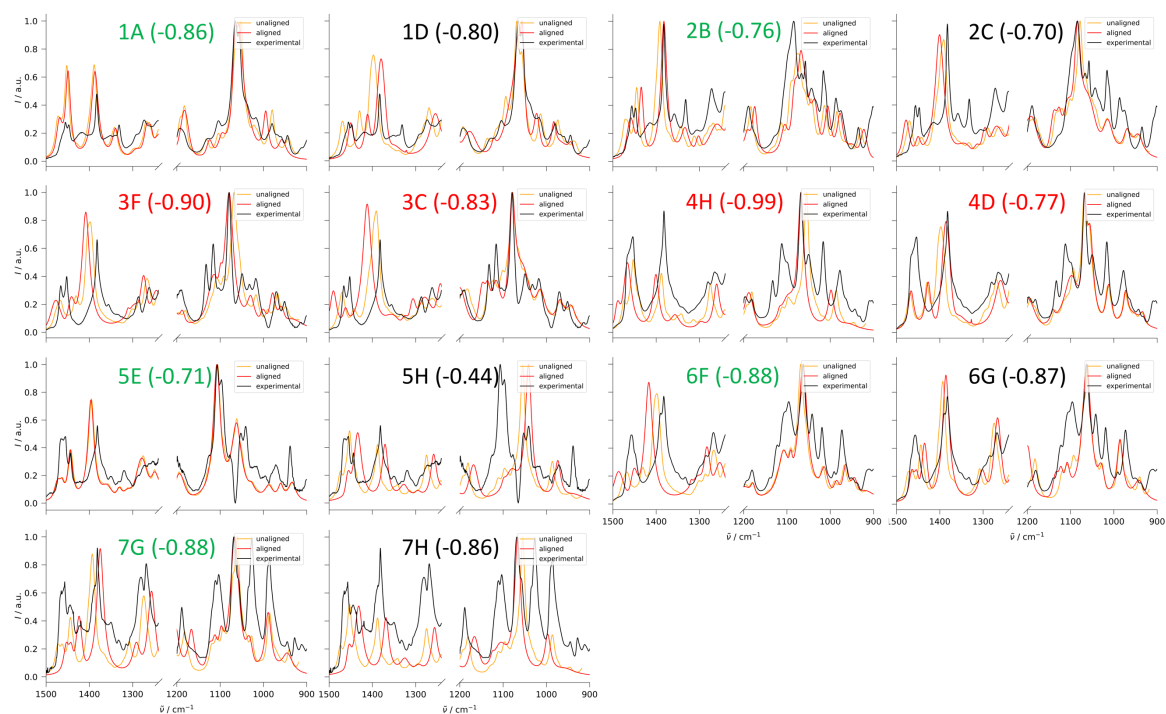


Figure 7.2: Fits for experimental IR spectra of diastereomers **1** – **7** with the corresponding theoretical spectra **A** – **H** yielding the best and second-best alignment score in case of correct assignment (green and black). For the incorrect assignments, the one yielding the best alignment score as well as the correct assignment is shown (red).

7.3 Conclusion

In this study, we aimed to assign the relative configuration of eight diastereomers **1 - 8** based on a comparison between DFT calculated and experimental ^1H and ^{13}C chemical shifts, as well as NOESY and IR spectra. Most of these data are collected routinely in an organic chemistry laboratory.

If the ^{13}C shifts were individually matched for each diastereomer, six out of eight diastereomers were identified correctly, whereas only two could be identified by an individual match of the ^1H shifts. Combination of the ^{13}C and ^1H shifts resulted in correct identification of seven out of eight diastereomers, although some of these assignments were obtained with low confidence. If the shift data for all eight diastereomers were matched collectively, assignment based on ^{13}C chemical shifts alone was possible with high confidence giving a likelihood of 92.5 % for the correct assignment. In contrast, assignment of relative configuration based on a collective math of ^1H chemical shifts alone gave a likelihood for the correct assignment of only 10.3 %, which is clearly too low for a confident assignment. Collectively matching combined ^1H and ^{13}C chemical shift data led to an increased likelihood for the correct assignment of 97.6 %.

Next, it was also investigated whether the different diastereomers can be assigned by directly matching their NOESY spectra with calculated NOESY spectra based on DFT generated conformational ensembles with the NOVAS protocol presented in Chapter 6. All eight diastereomers were correctly assigned. The power of this technique is reflected in the fact that with the NOVAS procedure correct assignment of each individual diastereomer was obtained with likelihoods over 90 %. Therefore, it could be expected that collectively matching the combined data from the NOESY spectra of all eight diastereomers would lead to an extremely high likelihood for the correct assignment of over 99.9 %. In order to answer the question how successfully this method can be applied in general, more NOESY data for challenging sets of diastereomers will be needed.

By combining the chemical shift data with the NOESY data, it was possible to assign all individual diastereomers with likelihoods over 99.6 %. When this data is matched collectively, a likelihood over 99.99 % is achieved for the correct assignment of the entire set, whereas the second-best assignment has a likelihood of only $5.9 \cdot 10^{-5}$ %. This clearly shows that combination of different experimental data in a statistically meaningful way can be a powerful strategy to assign diastereomers. It will be interesting to see how this approach will perform for even more challenging systems.

Lastly, we have also investigated if IR spectra can be used to differentiate between the diastereomers. Based on the IRSA algorithm, five out of seven diastereomers for which experimental data could be obtained were identified correctly. A reason why no correct match was found for the other two was that the fingerprint regions of the experimental IR spectra of different diastereomers were too similar. Since the compounds are flexible, it is potentially possible to observe similar bands in the fingerprint region when only one stereocenter is changed between two diastereomers. For the future, we will investigate how the alignment score can be transferred into a probability, such that it can be combined with the spectroscopic data from NMR in a statistically rigorous manner.

In conclusion, we were able to show the importance of combining different independent datasets for the correct identification of the diastereomers. Without this combination, only the NOVAS approach was able to identify all individual diastereomers correctly. For the chemical shift data, a combination with other independent data was necessary to improve confidence. Next to IR spectra, readily accessible, valuable complementary information could also come from J-couplings.

7.4 Method Section

Experimental Details

The eight diastereomers **1** - **8** were kindly provided by Prof. Erick M. Carreira (ETH Zürich). Their synthesis is described in detail in C. Nilewski *et al.*⁵⁴⁵ IR spectra were recorded in chloroform on a Spectrum Two™ FT-IR spectrometer (Perkin Elmer) using a NaCl cell with a path length of 0.2 mm. The software Spectrum 10™ (Perkin Elmer) was used for recording the spectrum and for baseline correction. NMR spectra were recorded at 25.0 °C on a Bruker AVANCE III HD 600 MHz spectrometer equipped with a N₂-cooled Prodigy triple resonance probe with z-gradients. The ¹H spectrum of **7** was recorded in CDCl₃ and referenced to internal TMS. NOESY spectra were recorded in CDCl₃ (Apollo Scientific) with a mixing time of 3 s. A recycle delay of 15 s was used to obtain symmetric NOESY spectra. A drop of D₂O (Armar) was added to the samples to promote exchange of the hydroxy protons with deuterium to minimize dipolar relaxation involving the hydroxy group. The spectral width of the NOESY spectra was 6 ppm in both dimensions, and the transmitter was set at 2.7 ppm. A total of 4096 x 256 data points were recorded. The time domain in both dimensions of the NOESY spectra was doubled by zero filling and the baseline was corrected with a third order polynomial. Processing was done with Bruker TopSpin™ version 4.1 (Bruker Biospin AG). Peak assignment and volume integration was done using NMRFAM-SPARKY.¹⁹⁹

Computational Details

The conformers for diastereomers **1** – **8** as well as for the chlorinated molecules in the literature set were generated from SMILES strings using the KDG conformer generator⁸⁹ of RDKit²⁶³ with an RMSD threshold for pruning of 0.3 Å. No experimental torsion angle preferences were applied because the gauche effect between two neighboring chlorine atoms is not properly taken into account by the underlying SMARTS patterns. Initially, the conformation found in the crystal was therefore not always contained in the generated ensemble. Between 568 and 592 structures were generated per diastereomer. The obtained 3D structures were optimized in vacuo with DFT using Orca 5.0.1²⁴⁸⁻²⁵⁰ at the BP86/def2-tzvp²⁵⁵⁻²⁵⁷ level applying the resolution of identity approximation with def2/J²⁵⁸ as auxiliary basis set and Grimme's D3BJ dispersion correction.^{264,265} For the calculation of the IR spectra and to verify that the optimized structures corresponded to a local energetic minimum, a frequency calculation was performed at the same level of theory. For structures having imaginary frequencies, the geometry at the most displaced point along the corresponding mode was taken as input for a new structure optimization. The energy was also calculated using CPCM²⁶¹ as an implicit solvent model for chloroform to account for solvation energy. NMR chemical shieldings were computed with the GIAO²⁶⁶ approach using the hybrid GGA

functional PBE0²⁶⁷ with the cc-pVTZ basis set²⁵² using cc-pVTZ/JK auxiliary basis set²⁶⁸ and D3BJ. The resolution of identity approximation for both Coulomb and HF exchange integrals was applied (RIJK). Shielding calculations were done in vacuo for ¹³C and using CPCM for chloroform for ¹H. The relative free energies of the conformers were computed as the sum of the Gibbs energy obtained from the frequency calculation and the difference between the final energies obtained in vacuo and with CPCM solvation. Structures which differ less than 0.1 kJ/mol were checked for representing the same minimum by calculating their RMSD using a Python script from Ref. 533. If the difference between structures was below 0.05 Å, the conformers were classified as identical and one was removed from the ensemble. The remaining conformers were Boltzmann-averaged for further analysis.

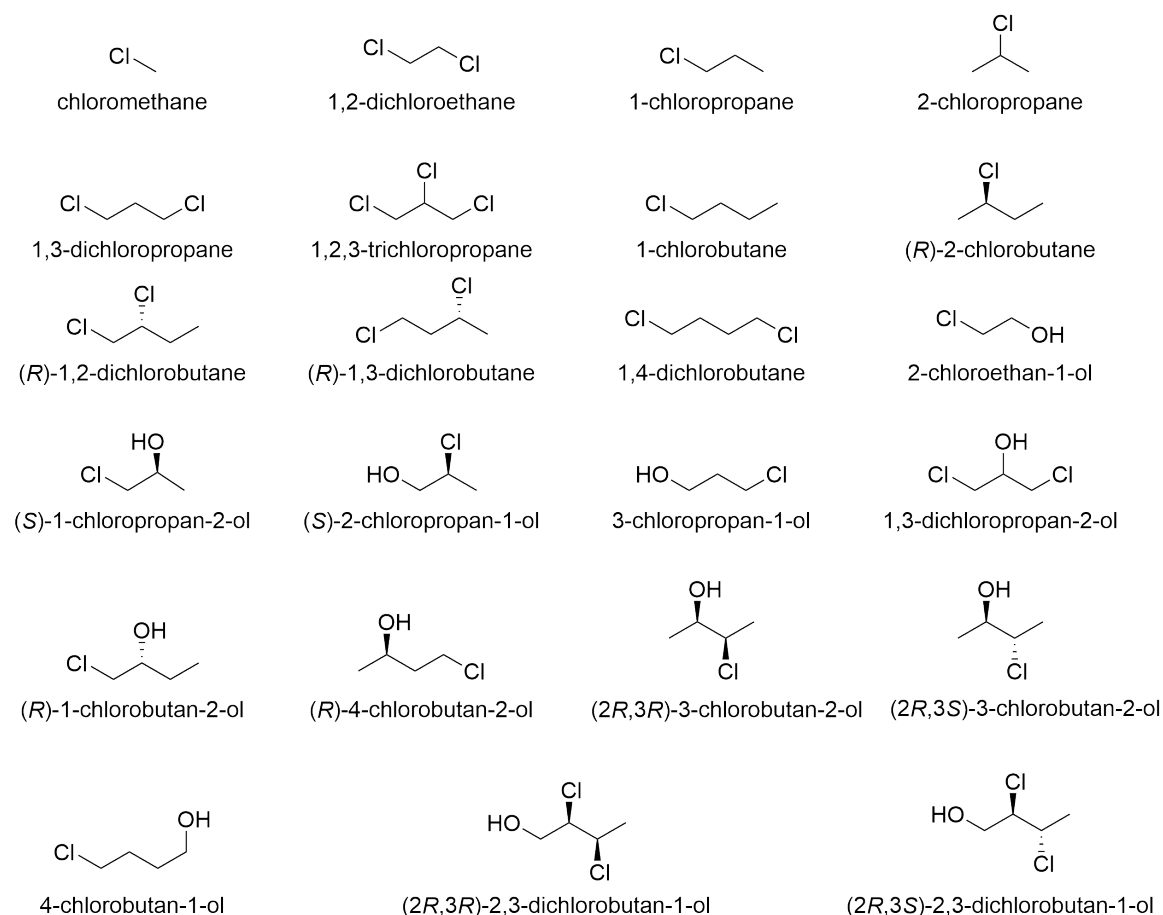
The chemical shieldings were transformed using intercept and scale from Chapter 5 with an additional correction term for carbons directly bound to chlorine (see Appendix). The NOVAS approach was applied as described in Chapter 6 with a Python²⁰¹ script run in a Jupyter Notebook.²⁰² Functionalities of the matplotlib,²⁰³ ngview,²⁷¹ numpy,²⁰⁵ openbabel,²⁷² pandas,²⁰⁶ and scipy²⁰⁷ packages were used.

IR spectra were calculated by averaging the calculated individual IR spectra of each conformer. Baseline correction of the considered region between 900 and 1500 cm⁻¹ was done with a Python script provided with the IRSA algorithm.⁵³⁸ The IR frequencies were scaled with 1.0192⁵⁵⁵ and the Lorentzian bandwidth was chosen to be 12 cm⁻¹. The region between 1200 and 1240 cm⁻¹ was ignored due to a strong absorbance band of the solvent. Peaks in the IR spectra were picked using the integrated script of the IRSA algorithm. The sequence alignment was also done using the provided scripts.⁵⁴⁰

7.5 Appendix

Determination of Correction Factor for Chlorine Bound Carbons

The compounds of the literature set contain a maximum of four carbons, bear at least one chlorine substituent, and can have an additional hydroxy substituent.



Scheme A7.1: Set of 23 organic compounds containing at least one chlorine substituent and a maximum of four carbons. For all of these compounds, ^{13}C chemical shifts in CDCl_3 are reported in the literature.⁵⁴⁷⁻⁵⁵⁴

The final chemical shifts can then be obtained from the calculated shieldings with the following equation:

$$\delta_{\text{calc}_i} = \frac{\sigma_{\text{calc}_i} - q}{m} + Cl_{\text{corr}} \quad (\text{A7.1})$$

δ_{calc_i} is the calculated chemical shift, σ_{calc_i} is the calculated shielding constant, q and m are the intercept and the slope of the regression with values of 185.67 ppm and -1.0317 for ^{13}C (calculation in vacuum) as well as 31.34 ppm and -1.0205 for ^1H (using the conductor-like polarizable continuum model (CPCM)²⁶¹ for the shielding calculation in chloroform) (both determined in Chapter 5). Cl_{corr} is the correction factor for the carbon atoms directly bound to a chlorine atom. It has a value of -5.34 ppm and was obtained by fitting a line with slope 1.0 to the

calculated ^{13}C chemical shifts of the chlorine bound atoms plotted against their experimental values from the literature.

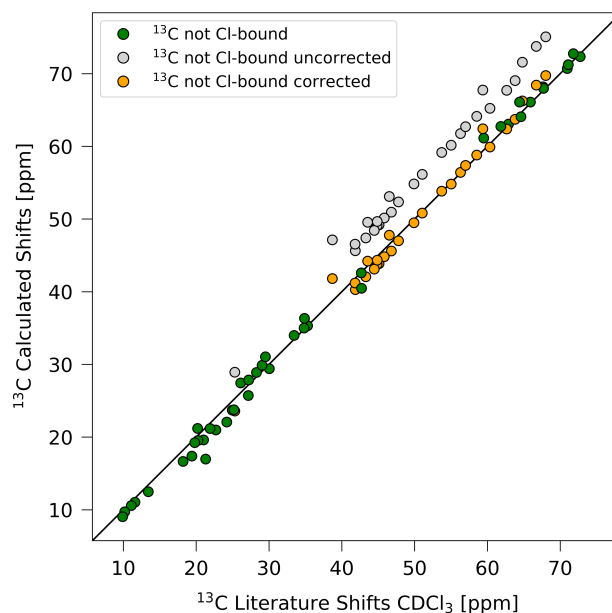


Figure A7.1: ^{13}C calculated chemical shifts versus the chemical shifts from the literature for the 23 compounds shown in Scheme A7.1. The green points are ^{13}C chemical shifts of carbons not directly connected to a chlorine atom. Calculated shifts were obtained from the shieldings with conversion parameters slope=-1.0317 and intercept=185.7 from Chapter 5. The grey points are ^{13}C chemical shifts of carbons directly bound to a chlorine atom. Calculated shifts were obtained with the same conversion parameters as for carbons not connected to a Cl atom. The calculated shifts of the orange data points are obtained by addition of a constant correction factor of -5.34 ppm to account for relativistic effects. For the corrected Cl-bound carbons, a mean absolute deviation (MAD) of 0.99 ppm, a maximum absolute deviation (max. AD) of 3.10 ppm and a root mean square deviation (RMSD) of 1.27 ppm is obtained whereas before correction values of 5.34, 8.44 and 5.48 ppm were obtained for these metrics. After the correction the overall MAD is 0.99 ppm, max. AD is 4.32 ppm and RMSD is 1.26 ppm.

Table A7.1: ^1H chemical shifts of diastereomers **1** – **8** in CDCl_3 in ppm. Shifts of **1** – **6** and **8** from Ref. 545. The shifts of the two hydroxy groups are not reported. The order of the ^1H shifts is from left to right starting at the methyl group and ending at the methylene group. * The methylene protons are averaged for easier comparison since assignment in calculated structures was not possible.

	1'	2'	3'	4'	5'	6'*
1	1.59	4.71	4.33	4.02	4.48	4.08
2	1.58	4.72	4.30	4.00	4.61	4.03
3	1.71	4.36	4.08	4.52	4.22	3.95
4	1.74	4.27	4.41	4.53	4.08	4.06
5	1.67	4.39	4.22	4.22	4.22	3.99
6	1.67	4.34	4.59	4.24	4.06	4.04
7	1.65	4.77	4.02	4.23	4.63	4.05
8	1.64	4.79	4.04	4.25	4.56	4.05

Table A7.2: ^{13}C chemical shifts of diastereomers **1** – **8** in CDCl_3 in ppm. Shifts of **1** – **6** and **8** from Ref. 545. The order of the ^{13}C shifts is from left to right starting at the methyl group and ending at the methylene group.

	1'	2'	3'	4'	5'	6'
1	19.73	56.42	67.45	75.90	62.15	63.93
2	18.76	56.33	66.78	73.71	63.05	65.86
3	22.88	56.42	66.80	71.48	67.09	63.85
4	23.45	56.29	68.04	71.51	61.75	64.55
5	23.08	56.64	69.59	73.35	62.60	64.91
6	21.87	59.63	67.27	72.27	61.81	64.18
7	22.84	56.07	65.59	73.49	62.81	65.59
8	22.91	56.76	65.44	75.80	62.78	63.41

Table A7.3: Calculated ^1H chemical shifts of diastereomers **A** – **H** in CDCl_3 in ppm. The shift of the two hydroxy groups is not reported. * The methylene protons are averaged.

	1'	2'	3'	4'	5'	6'*
A	1.55	5.11	4.52	4.08	4.82	4.13
B	1.55	5.00	4.55	4.04	4.97	4.06
C	1.64	4.59	4.35	4.60	4.33	3.90
D	1.70	4.49	4.86	4.63	4.22	4.02
E	1.63	4.76	4.52	4.37	4.48	4.00
F	1.64	4.56	5.09	4.35	4.20	4.00
G	1.61	5.09	4.22	4.36	4.99	4.06
H	1.63	5.16	4.28	4.24	4.80	4.12

Table A7.4: Calculated ^{13}C chemical shifts of diastereomers **A** – **H** in CDCl_3 in ppm.

	1'	2'	3'	4'	5'	6'
A	16.72	56.20	68.68	76.30	62.77	64.01
B	16.43	55.43	66.34	75.32	65.41	66.95
C	20.54	56.19	68.80	71.29	68.26	64.42
D	21.61	56.30	69.66	71.55	61.72	64.69
E	21.27	56.72	70.81	74.96	63.77	65.83
F	20.01	59.61	68.25	71.12	63.28	63.91
G	21.61	55.70	65.53	74.22	65.26	66.65
H	21.56	56.86	66.24	76.40	62.69	64.72

Table A7.5: Assignment probabilities >0.1 % obtained based on ¹³C shifts, if experimental data is available for **1 – 8**. Correct matches between experimental and calculated datasets (**1 – 8**) are set in bold text.

Assignment	Likelihood [%]
1A,2B,3C,4D,5E,6F,7G,8H	92.5
1A,2B,3C,4E,5D,6F,7G,8H	2.5
1B,2A, 3C,4E,5D,6F,7G,8H	2.4
1A,2G,3C,4D,5E,6F,7B,8H	0.5
1A,2B,3C,4D,5E,6F,7H,8G	0.4
1H, 2B,3C,4D,5E,6F,7G,8A	0.4
1H,2A, 3C,4D,5E,6F,7G,8B	0.1
1A,2B,3E,4D,5C,6F,7G,8H	0.1

Table A7.6: Assignment probabilities >1 % obtained based on ¹H shifts, if experimental data is available for **1 – 8**. Correct matches between experimental and calculated datasets (**1 – 8**) are set in bold text.

Assignment	Likelihood [%]
1A,2B,3C,4D,5E,6F,7G,8H	10.3
1A,2B,3C,4D,5E,6F,7H,8G	8.0
1B,2A, 3C,4D,5E,6F,7G,8H	7.5
1B,2A, 3C,4D,5E,6F,7H,8G	5.8
1E,2B,3C,4D,5A,6F,7G,8H	2.1
1E,2A, 3C,4D,5B,6F,7G,8H	2.0
1A,2B,3C,4F,5E,6D,7G,8H	1.9
1E, 2B,3C,4D,5A,6F,7H,8G	1.6
1E,2A, 3C,4D,5B,6F,7H,8G	1.6
1A,2B,3C,4F,5E,6D,7H,8G	1.5
1H,2B,3C,4D,5E,6F,7G,8A	1.4
1B,2A, 3C,4F,5E,6D,7G,8H	1.4
1A,2B,3E,4D,5C,6F,7G,8H	1.1
1B,2A, 3C,4E,5E,6D,7H,8G	1.1
1H, 2B,3C,4D,5E,6F,7A,8G	1.1
1H,2A, 3C,4D,5E,6F,7G,8B	1.0

Table A7.7: Assignment probabilities >0.1 % obtained based on combined ¹H and ¹³C shifts, if experimental data is available for **1 – 8**. Correct matches between experimental and calculated datasets (**1 – 8**) are set in bold text.

Assignment	Likelihood [%]
1A,2B,3C,4D,5E,6F,7G,8H	97.6
1B,2A, 3C,4D,5E,6F,7G,8H	1.8
1A,2B,3C,4D,5E,6F,7H,8G	0.3
1A,2B,3C,4E,5D,6F,7G,8H	0.1
1H, 2A,3C,4D,5E,6F,7G,8A	0.1

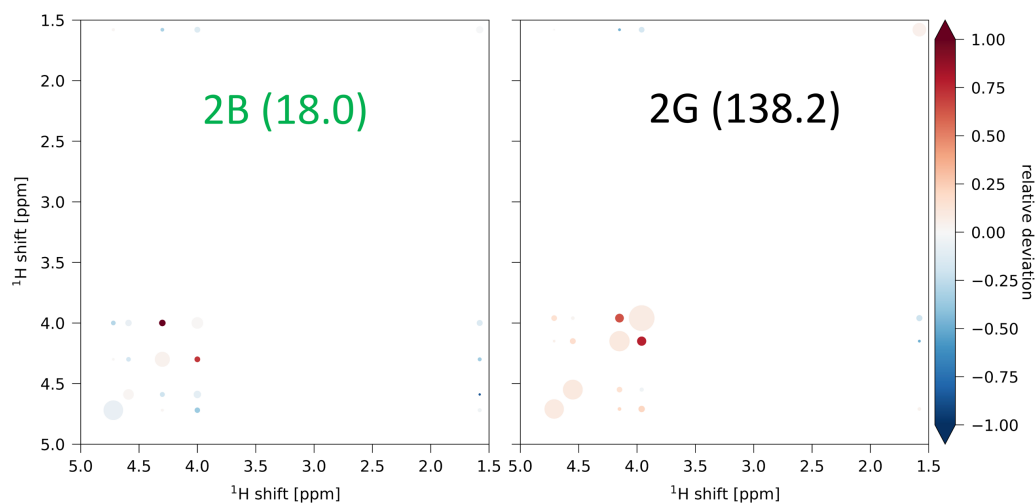


Figure A7.2: Difference between experimental NOESY spectrum of **2** and calculated NOESY spectra for the best fit **B** (green) and the second best fit **G** (black). wSSRs are given in brackets.

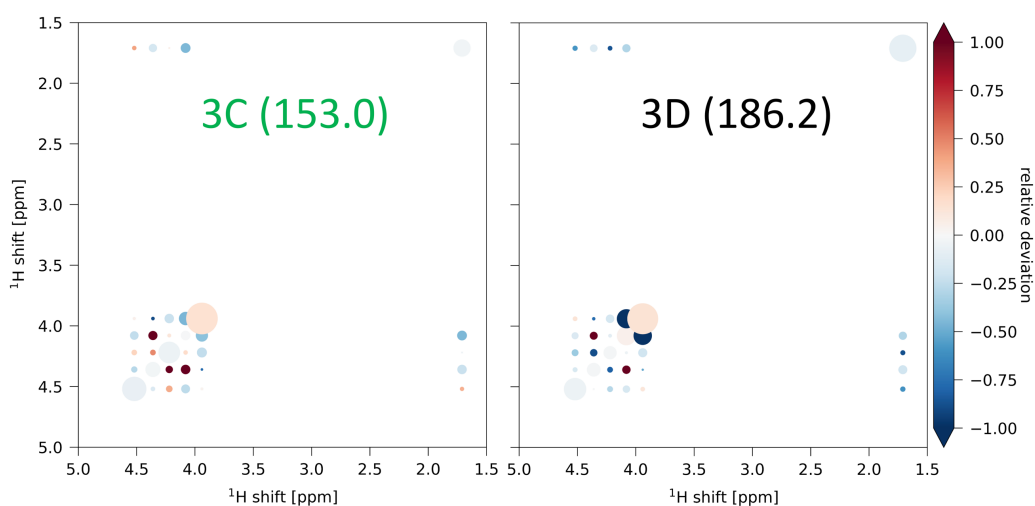


Figure A7.3: Difference between experimental NOESY spectrum of **3** and calculated NOESY spectra for the best fit **C** (green) and the second best fit **D** (black). wSSRs are given in brackets.

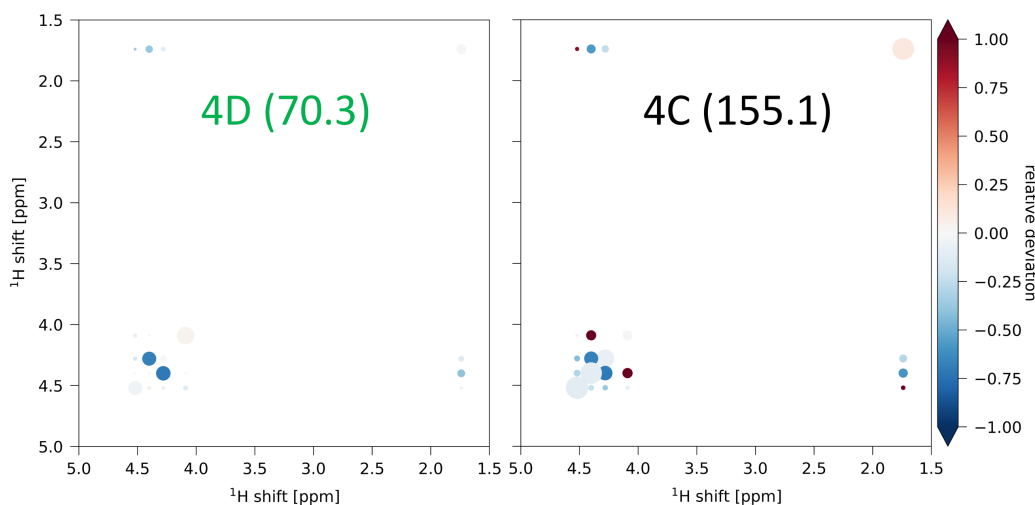


Figure A7.4: Difference between experimental NOESY spectrum of **4** and calculated NOESY spectra for the best fit **D** (green) and the second best fit **C** (black). wSSRs are given in brackets.

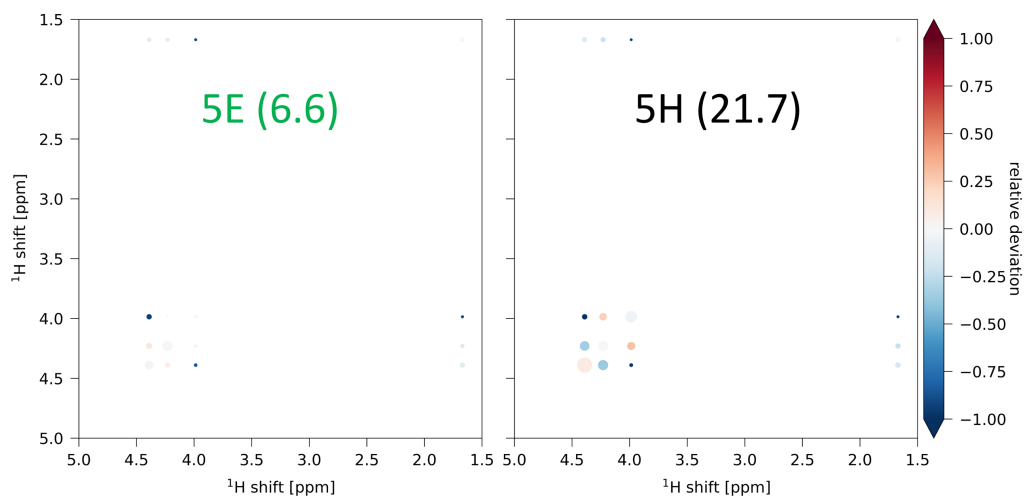


Figure A7.5: Difference between experimental NOESY spectrum of **5** and calculated NOESY spectra for the best fit **E** (green) and the second best fit **H** (black). wSSRs are given in brackets.

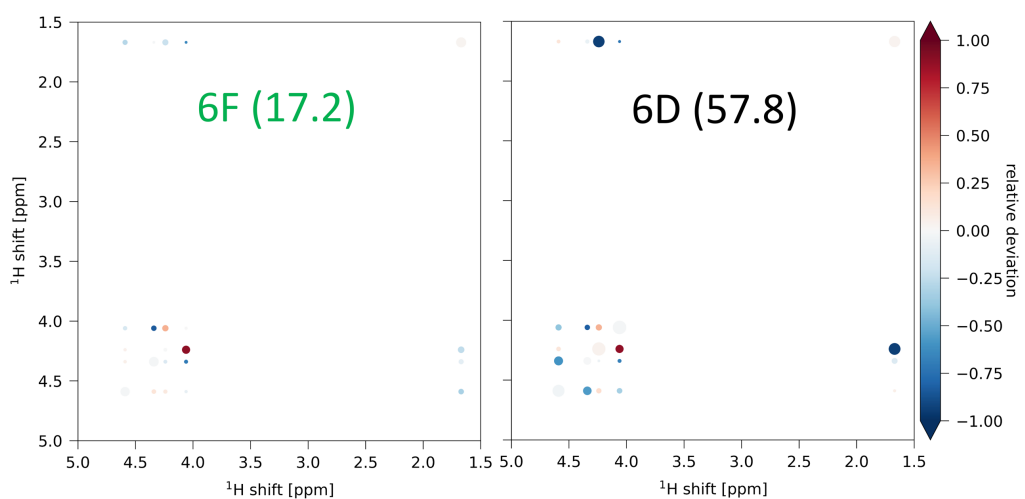


Figure A7.6: Difference between experimental NOESY spectrum of **6** and calculated NOESY spectra for the best fit **F** (green) and the second best fit **D** (black). wSSRs are given in brackets.

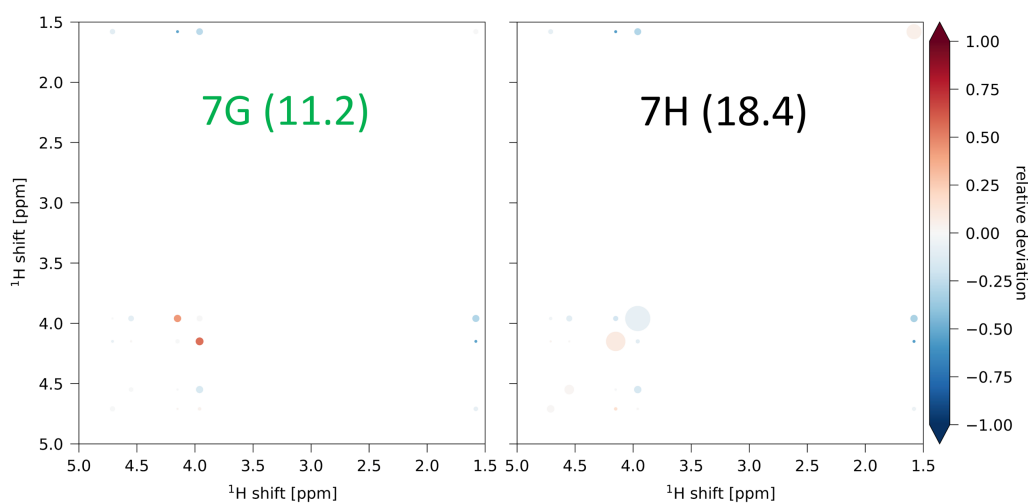


Figure A7.7: Difference between experimental NOESY spectrum of **7** and calculated NOESY spectra for the best fit **G** (green) and the second best fit **H** (black). wSSRs are given in brackets.

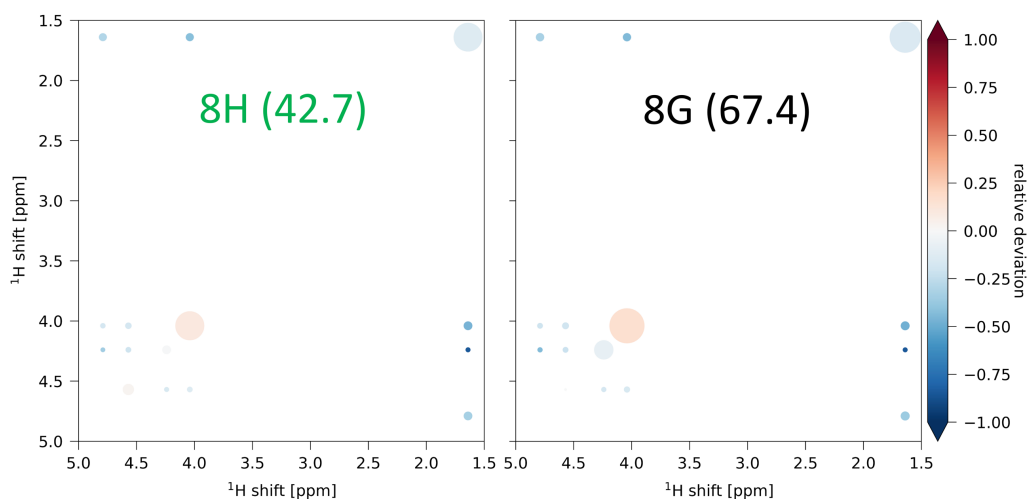


Figure A7.8: Difference between experimental NOESY spectrum of **8** and calculated NOESY spectra for the best fit **H** (green) and the second best fit **G** (black). wSSRs are given in brackets.

Table A7.8: NOESY volumes of **1** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-5490
H1Me	H2	273
H2	H1Me	261
H1Me	H3	55.3
H3	H1Me	52.9
H1Me	H4 / 6ab	207
H4 / 6ab	H1Me	198
H1Me	H5	10.5
H5	H1Me	11
H2	H2	-5730
H2	H3	207
H3	H2	203
H2	H4 / 6ab	66.34
H4 / 6ab	H2	59.2
H2	H5	78.6
H5	H2	67.9
H3	H3	-5760
H3	H4 / 6ab	333.9
H4 / 6ab	H3	313.8
H3	H5	65
H5	H3	72.8
H4 / 6ab	H4 / 6ab	-10825
H4 / 6ab	H5	488
H5	H4 / 6ab	512
H5	H5	-5520

Table A7.9: NOESY volumes of **2** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-3690
H1Me	H2	213
H2	H1Me	199
H1Me	H3	27.3
H3	H1Me	26.1
H1Me	H4 / 6ab	182
H4 / 6ab	H1Me	181
H1Me	H5	2.67
H2	H2	-4780
H2	H3	202
H3	H2	201
H2	H4 / 6ab	52.4
H4 / 6ab	H2	59.1
H3	H3	-4750
H3	H4 / 6ab	26.5
H4 / 6ab	H3	34.7
H3	H5	81.8
H5	H3	79.2
H4 / 6ab	H4 / 6ab	-7790
H4 / 6ab	H5	435
H5	H4 / 6ab	422
H5	H5	-4400

Table A7.10: NOESY volumes of **3** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-4710
H1Me	H2	206
H2	H1Me	194
H1Me	H3	112
H3	H1Me	112
H1Me	H4	20.8
H4	H1Me	20.1
H1Me	H5	12.3
H5	H1Me	11.0
H2	H2	-4370
H2	H3	14.9
H3	H2	11.1
H2	H4	56.0
H4	H2	65.6
H2	H5	30.7
H5	H2	18.8
H2	H6ab	5.40
H6ab	H2	2.58
H3	H3	-4120
H3	H4	155
H4	H3	152
H3	H5	44.8
H5	H3	47.2
H3	H6ab	213
H6ab	H3	196
H4	H4	-4630
H4	H5	62.6
H5	H4	55.7
H4	H6ab	69.0
H6ab	H4	69.6
H5	H5	-4500
H5	H6ab	207
H6ab	H5	213
H6ab	H6ab	-3990

Table A7.11: NOESY volumes of **4** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-4690
H1Me	H2	203
H2	H1Me	194
H1Me	H3	139
H3	H1Me	134
H1Me	H4	10.3
H4	H1Me	13.7
H2	H2	-4590
H2	H3	-168
H3	H2	-154
H2	H4	64.9
H4	H2	70.9
H3	H3	-4610
H3	H4	127
H4	H3	138
H3	H5 / 6ab	49.2
H5 / 6ab	H3	46.4
H4	H4	-4830
H4	H5 / 6ab	137
H5 / 6ab	H4	143
H5 / 6ab	H5 / 6ab	-8910

Table A7.12: NOESY volumes of **5** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-4870
H1Me	H2	256
H2	H1Me	250
H1Me	H3 / H4 / H5	166
H3 / H4 / H5	H1Me	164
H1Me	H6ab	-9.14
H6ab	H1Me	-8.3
H2	H2	-4610
H2	H3 / H4 / H5	457
H3 / H4 / H5	H2	474
H2	H6ab	-42.9
H6ab	H2	-18.3
H3 / H4 / H5	H3 / H4 / H5	-15100
H3 / H4 / H5	H6ab	389
H6ab	H3 / H4 / H5	372
H6ab	H6ab	-5290

Table A7.13: NOESY volumes of **6** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-3710
H1Me	H2	206
H2	H1Me	189
H1Me	H3	78.8
H3	H1Me	74.5
H1Me	H4	146
H4	H1Me	140
H5 / H6ab	H1Me	7.15
H2	H2	-4340
H2	H3	121
H3	H2	129
H4	H2	66.3
H2	H5 / H6ab	29.1
H5 / H6ab	H2	16.3
H3	H3	-4490
H3	H4	130
H4	H3	124
H3	H5 / H6ab	83.2
H5 / H6ab	H3	75.7
H4	H4	-4190
H4	H5 / H6ab	96.6
H5 / H6ab	H4	68.9
H5 / H6ab	H5 / H6ab	-8460

Table A7.14: NOESY volumes of **7** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-3240
H1Me	H2	160
H2	H1Me	166
H1Me	H3 / H6ab	125
H3 / H6ab	H1Me	121
H1Me	H4	7.48
H4	H1Me	7.46
H2	H2	-3490
H2	H3 / H6ab	143
H3 / H6ab	H2	136
H2	H4	41.5
H4	H2	36.8
H3 / H6ab	H3 / H6ab	-7110
H3 / H6ab	H4	85.2
H4	H3 / H6ab	92.3
H3 / H6ab	H5	237
H5	H3 / H6ab	221
H4	H4	-3660
H4	H5	137
H5	H4	134
H5	H5	-3560

Table A7.15: NOESY volumes of **8** recorded with a mixing time of 3 s in CDCl₃.

Resonance 1	Resonance 2	Volume [AU]
H1Me	H1Me	-1920
H1Me	H2	59.7
H2	H1Me	56.3
H1Me	H3 / H6ab	43.8
H3 / H6ab	H1Me	40.5
H1Me	H4	6.54
H2	H3 / H6ab	37.6
H2	H4	14.2
H3 / H6ab	H3 / H6ab	-2390
H3 / H6ab	H5	55.3
H5	H3 / H6ab	36.3
H4	H4	-1330
H4	H5	34.4
H5	H4	36.3
H5	H5	-1200

8 Conclusion and Outlook

In this thesis, we investigated different approaches to harness the information content of experimental data, mainly from NMR spectroscopy, through combination with computational approaches such as classical MD simulations and DFT calculations. By doing so, we could gain a better understanding of the relationship between the constitution of the conformational ensemble in solution and the associated properties.

In Chapter 2, we investigated the conformational behavior and ionophoric properties of the anthelmintic octadepsipeptides PF1022A, emodepside, and related compounds. We could show that the *symmetric* core conformation has a higher flexibility on the microsecond to millisecond time scale compared to the *asymmetric* core conformation, both in NMR and in kinetic models constructed from extensive MD data. The fact that the two approaches independently lead to the same findings simultaneously validates the MD model and aids the mechanistic interpretation of the NMR data. This exemplarily shows the power of the combination of MD and NMR. In addition, we found that the difference in anthelmintic activity between PF1022A and emodepside cannot be directly related to a difference in the conformational behavior or their capability to act as an ionophore. In general, it would be interesting to investigate the extent to which ion binding occurs in other cyclic peptides and whether there is a relationship with permeability. Further, the cavitand-like artifact structure observed in our MD simulations would be a suitable model system to assess whether MD with a polarizable force field would be able to describe the system more accurately.

In Chapter 3, we investigated whether the stabilizing effect from intramolecular side-chain hydrogen bonds of Asm residues, as seen for the $\beta^{6.3}$ -helix of the natural product pTB, can be transferred to other molecular systems. The stabilizing effect for GramA-Asm observed in MD simulations could not be confirmed by NMR experiments. ^{23}Na spectra revealed that the $\beta^{6.3}$ -helical conformation was also not adopted in SDS micelles, while the wild-type GramA is known to do that. As an alternative model system, the two cyclic octapeptides Asm-1 and Asn-1 were explored. The four expected dimer variants were observed for both peptides in MSMs constructed from extensive MD simulations. As hypothesized, the side-chain N-methylation appears to stabilize the dimer arrangement that is able to form an additional hydrogen bond between the monomers. In preliminary NMR experiments, exchange with the solvent of the side-chain amide was observed for Asn-1, whereas no exchange could be detected for Asm-1. While these preliminary results are encouraging, more experiments are needed for a clear conclusion.

Especially dissociation studies could provide more insights. Additional material necessary for these experiments is currently synthesized.

In Chapter 4, we presented a new set of precise RDC data for cyclosporin A recorded in a cross-linked PMMA gel swollen in chloroform. All one-bond CH and NH RDCs as well as the two-bond homonuclear RDCs of the methylene groups could be determined. To the best of our knowledge, this is the largest set of RDCs reported for cyclosporin A to date. Interestingly, we found that two ensembles obtained from MDOC simulations starting from different crystal structures both reproduce the experimental RDCs within their experimental errors, despite the fact that they differ in the configuration of one peptide bond. This indicates that even the entire set of one-bond CH and NH RDCs does not provide enough information to unambiguously describe the conformational ensemble of cyclosporin A. For future studies, the set of RDC data could be combined with J-coupling information as well as with NOE-derived distances to increase the restraint density.

In Chapter 5, ^1H and ^{13}C chemical shifts of 35 small and rigid organic molecules were measured under standardized conditions in CDCl_3 and in CCl_4 . The comparison of the experimental chemical shifts in CCl_4 and CDCl_3 clearly showed that distinct solvent effects are present even in such apolar environments. Especially the ^{13}C shifts of carbons in polar groups are affected. The set of chemical shifts collected in this study provides valuable reference data to validate and compare different DFT methods especially with respect to improved implicit solvent models. In future studies, it will be interesting to see whether the agreement with experimental shifts improves if the geometries of our rigid reference molecules are also optimized using an implicit solvent model.

In Chapter 6, we presented the NOVAS approach and investigated the use of information obtained from NOESY spectra recorded beyond the linear build-up regime to differentiate between stereospecific assignments of methylene protons. With NOVAS, no transformation of the volumes to distances is necessary anymore and one can directly compare a computer-generated ensemble to the primary experimental NOE data.

In Chapter 7, we aimed to assign the relative configuration of eight flexible diastereomers of a trichlorinated-hexa-1-3-diol based on a comparison between DFT calculated and experimental ^1H and ^{13}C chemical shifts, NOESY spectra, as well as IR spectra. When the calculated data was compared to each diastereomer individually, six out of eight diastereomers were correctly identified based on ^{13}C shifts, whereas only two could be correctly identified based on ^1H shifts. Combination of the two shifts resulted in correct identification of seven out of eight diastereomers. To safely identify all individual diastereomers, a combination with other

independent data (.e.g., J-couplings and IR) would be necessary. We also investigated whether the diastereomers can be assigned with the NOVAS approach presented in Chapter 6. In this case, all eight diastereomers were correctly identified with likelihoods >90 % for each individual diastereomer. A combination of chemical shifts and NOESY data gave even higher likelihoods of >99.6 % for the correct assignment of each individual diastereomer. Making use of the fact that data is available for all eight diastereomers, a likelihood of > 99.99 % was obtained for the correct assignment of the whole set of experimental data. To evaluate the full potential of NOVAS, more NOESY data for challenging sets of diastereomers will be needed. Based on the IRSA algorithm, five out of seven measurable diastereomers were identified correctly. The information obtained from IR seems to be complementary to the chemical shift data. It will be necessary to develop an approach for translating the IRSA alignment score into an assignment probability, such that it can be combined with the spectroscopic data from NMR in a statistically rigorous manner.

References

- 1 A. A. Langham, H. Khandelia, B. Schuster, A. J. Waring, R. I. Lehrer and Y. N. Kaznessis, *Peptides*, 2008, **29**, 1085–1093.
- 2 Y. Li, W. Li and Z. Xu, *Mar. Drugs*, 2021, **19**, 311.
- 3 J. Witek, S. Wang, B. Schroeder, R. Lingwood, A. Dounas, H. J. Roth, M. Fouché, M. Blatter, O. Lemke, B. Keller and S. Riniker, *J. Chem. Inf. Model.*, 2019, **59**, 294–308.
- 4 J. Egli, C. Esposito, M. Müri, S. Riniker and H. Wennemers, *J. Am. Chem. Soc.*, 2021, **143**, 5937–5942.
- 5 S. Wang, G. König, H. J. Roth, M. Fouché, S. Rodde and S. Riniker, *J. Med. Chem.*, 2021, **64**, 12761–12773.
- 6 D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.*, 2009, **5**, 789–796.
- 7 L. T. Afolabi, F. Saeed, H. Hashim and O. O. Petinrin, *PLoS One*, 2018, **13**, e0189538.
- 8 E. King, E. Aitchison, H. Li and R. Luo, *Front. Mol. Biosci.*, 2021, **8**, 712085.
- 9 D. C. Burns and W. F. Reynolds, *Optimizing NMR Methods for Structure Elucidation : Characterizing Natural Products and Other Organic Compounds*, Royal Society of Chemistry, Cambridge, 2018.
- 10 A. Y. S. Balazs, N. L. Davies, D. Longmire, M. J. Packer and E. Chiarparin, *Magn. Reson.*, 2021, **2**, 489–498.
- 11 A. C. De Dios and E. Oldfield, *Solid State Nucl. Magn. Reson.*, 1996, **6**, 101–125.
- 12 M. Elyashberg, *TrAC - Trends Anal. Chem.*, 2015, **69**, 88–97.
- 13 E. Pretsch, P. Bühlmann and M. Badertscher, *Spektroskopische Daten zur Strukturaufklärung organischer Verbindungen*, Springer, Berlin, Heidelberg, 2010.
- 14 U. Sternberg and R. Witter, *J. Biomol. NMR*, 2019, **73**, 727–741.
- 15 J. Härtner and U. M. Reinscheid, *J. Mol. Struct.*, 2008, **872**, 145–149.
- 16 P. Robustelli, K. Kohlhoff, A. Cavalli and M. Vendruscolo, *Structure*, 2010, **18**, 923–933.
- 17 M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11–15.
- 18 P. Salvador, in *Annual Reports on NMR Spectroscopy*, Elsevier Ltd., 1st edn., 2014, vol. 81, pp. 185–227.
- 19 B. Coxon, *Adv. Carbohydr. Chem. Biochem.*, 2009, **62**, 17–82.
- 20 N. Matsumori, M. Murata and K. Tachibana, *Tetrahedron*, 1995, **51**, 12229–12238.
- 21 N. Matsumori, D. Kaneno, M. Murata, H. Nakamura and K. Tachibana, *J. Org. Chem.*, 1999, **64**, 866–876.
- 22 A. W. Overhauser, *Phys. Rev.*, 1953, **92**, 411–415.
- 23 H. J. Dyson and P. E. Wright, *Annu. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 519–538.

- 24 M. R. M. Koos, K. H. G. Schulz and R. R. Gil, *Chem. Sci.*, 2020, **11**, 9930–9936.
- 25 D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, Wiley-VCH, New York, 2nd edn., 2000.
- 26 A. E. Torda, R. M. Scheek and W. F. van Gunsteren, *Chem. Phys. Lett.*, 1989, **157**, 289–294.
- 27 B. Vögeli, S. Kazemi, P. Güntert and R. Riek, *Nat. Struct. Mol. Biol.*, 2012, **19**, 1053–1058.
- 28 B. Vögeli, S. Olsson, P. Güntert and R. Riek, *Biophys. J.*, 2016, **110**, 113–126.
- 29 E. Hellemann, R. R. Teles, F. Hallwass, W. Barros, A. Navarro-Vázquez and R. R. Gil, *Chem. - Eur. J.*, 2016, **22**, 16632–16635.
- 30 R. R. Gil, C. Gayathri, N. V Tsarevsky and K. Matyjaszewski, *J. Org. Chem.*, 2008, **73**, 840–848.
- 31 J. Klages, C. Neubauer, M. Coles, H. Kessler and B. Luy, *ChemBioChem*, 2005, **6**, 1672–1678.
- 32 J. C. Freudenberger, S. Knör, K. Kobzar, D. Heckmann, T. Paululat, H. Kessler and B. Luy, *Angew. Chemie - Int. Ed.*, 2005, **44**, 423–426.
- 33 A. Marx, B. Böttcher and C. M. Thiele, *Chem. - Eur. J.*, 2010, **16**, 1656–1663.
- 34 P. Lesot, C. Aroulanda, P. Berdagué, A. Meddour, D. Merlet, J. Farjon, N. Giraud and O. Lafon, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2020, **116**, 85–154.
- 35 G. W. Li, X.-J. Wang, X. Lei, N. Liu and Z.-Q. Wu, *Macromol. Rapid Commun.*, 2022, 2100898.
- 36 G. Otting, *Annu. Rev. Biophys.*, 2010, **39**, 387–405.
- 37 J. Koehler and J. Meiler, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **59**, 360–389.
- 38 W. M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273–274**, 2–12.
- 39 G. Kummerlöwe and B. Luy, *Annu. Reports NMR Spectrosc.*, 2009, **68**, 193–232.
- 40 G. Kummerlöwe and B. Luy, *TrAC - Trends Anal. Chem.*, 2009, **28**, 483–493.
- 41 R. R. Gil, C. Griesinger, A. Navarro-Vázquez and H. Sun, in *Structure Elucidation in Organic Chemistry: The Search for the Right Tools*, Wiley-VCH Verlag GmbH & Co. KGaA, 2015, pp. 279–324.
- 42 F. Hallwass, M. Schmidt, H. Sun, A. Mazur, G. Kummerlöwe, B. Luy, A. Navarro-Vázquez, C. Griesinger and U. M. Reinscheid, *Angew. Chemie - Int. Ed.*, 2011, **50**, 9487–9490.
- 43 A. Das and N. Nath, *Magn. Reson. Chem.*, 2021, **59**, 569–576.
- 44 A. Navarro-Vázquez, P. Berdagué and P. Lesot, *ChemPhysChem*, 2017, **18**, 1252–1266.
- 45 P. Lesot, R. R. Gil, P. Berdagué and A. Navarro-Vázquez, *J. Nat. Prod.*, 2020, **83**, 3141–3148.
- 46 W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. Van Der Vegt and H. B. Yu, *Angew. Chemie - Int. Ed.*, 2006, **45**, 4064–4092.
- 47 S. A. Hollingsworth and R. O. Dror, *Neuron*, 2018, **99**, 1129–1143.
- 48 W. F. van Gunsteren, P. H. Hünenberger, H. Kovacs, A. E. Mark and C. A. Schiffer, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 1995, **348**, 49–59.

- 49 H. A. Scheraga, M. Khalili and A. Liwo, *Annu. Rev. Phys. Chem.*, 2007, **58**, 57–83.
- 50 P. S. Georgoulia and N. M. Glykos, *Arch. Biochem. Biophys.*, 2019, **664**, 76–88.
- 51 J. Witek, B. G. Keller, M. Blatter, A. Meissner, T. Wagner and S. Riniker, *J. Chem. Inf. Model.*, 2016, **56**, 1547–1562.
- 52 A. H. Mazurek, Ł. Szeleszczuk and T. Gubica, *Int. J. Mol. Sci.*, 2021, **22**, 9422.
- 53 L. Perić, C. S. Pereira, S. Perez and P. H. Hünenberger, *Mol. Simul.*, 2008, **34**, 421–446.
- 54 C. Domene, *Cent. Eur. J. Chem.*, 2007, **5**, 635–671.
- 55 S. Riniker, *J. Chem. Inf. Model.*, 2018, **58**, 565–578.
- 56 J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren and J. P. M. Postma, *Biopolymers*, 1984, **23**, 1513–1518.
- 57 N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. van Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843–856.
- 58 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 59 C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migués, J. Bickel, Y. Wang, J. Pincay, Q. Wu and C. Simmerling, *J. Chem. Theory Comput.*, 2020, **16**, 528–552.
- 60 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
- 61 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller and A. D. MacKerell, *Nat. Methods*, 2016, **14**, 71–73.
- 62 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
- 63 E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.*, 2016, **12**, 281–296.
- 64 G. C. Q. da Silva, G. M. Silva, F. W. Tavares, F. P. Fleming and B. A. C. Horta, *J. Mol. Model.*, 2020, **26**, 296.
- 65 S. Kashfolgheta, M. P. Oliveira, S. R. Rieder, B. A. C. Horta, W. E. Acree and P. H. Hünenberger, *J. Chem. Theory Comput.*, 2020, **16**, 7556–7580.
- 66 N. V. Buchete and G. Hummer, *J. Phys. Chem. B*, 2008, **112**, 6057–6069.
- 67 E. Vanden-Eijnden, M. Venturoli, G. Ciccotti and R. Elber, *J. Chem. Phys.*, 2008, **129**, 174102.
- 68 O. Lemke and B. G. Keller, *J. Chem. Phys.*, 2016, **145**, 164104.
- 69 C. Schütte, F. Noé, J. Lu, M. Sarich and E. Vanden-Eijnden, *J. Chem. Phys.*, 2011, **134**, 204105.
- 70 F. Noé and E. Rosta, *J. Chem. Phys.*, 2019, **151**, 190401.

- 71 A. E. Torda, R. M. Brunne, T. Huber, H. Kessler and W. F. van Gunsteren, *J. Biomol. NMR*, 1993, **3**, 55–66.
- 72 P. Tzvetkova, U. Sternberg, T. Gloge, A. Navarro-Vázquez and B. Luy, *Chem. Sci.*, 2019, **10**, 8774–8791.
- 73 J. Kapil, P. Shukla and A. Pathak, in *Recent Trends in Materials and Devices*, Springer, Singapore, 2020, vol. 256, pp. 211–220.
- 74 B. G. Janesko, *Chem. Soc. Rev.*, 2021, **50**, 8470–8495.
- 75 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
- 76 N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315–2372.
- 77 Y. Yuan and F. Wang, *J. Chem. Phys.*, 2021, **155**, 094104.
- 78 H. B. Schlegel, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 790–809.
- 79 M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457–484.
- 80 P. Morgante and R. Peverati, *Int. J. Quantum Chem.*, 2020, **120**, e26332.
- 81 G. Schreckenbach and T. Ziegler, *Theor. Chem. Acc.*, 1998, **99**, 71–82.
- 82 R. Ditchfield, *Mol. Phys.*, 1974, **27**, 789–807.
- 83 M. Bühl and T. van Mourik, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 634–647.
- 84 J. Shim and A. D. Mackerell, *Med. Chem. Commun.*, 2011, **2**, 356–370.
- 85 N. E. Binbay, V. Binbay, M. Aydemir, F. Durap, N. Meriç, C. Kayan and N. Arslan, *Eur. J. Sci. Technol.*, 2019, 256–269.
- 86 Y. I. Yang, Q. Shao, J. Zhang, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2019, **151**, 070902.
- 87 A. S. Kamenik, S. M. Linker and S. Riniker, *Phys. Chem. Chem. Phys.*, 2022, **24**, 1225–1236.
- 88 J. M. Blaney and J. S. Dixont, in *Reviews in Computational Chemistry*, VHC Publisher, Inc., New York, 2007, vol. 5, pp. 299–335.
- 89 S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574.
- 90 S. Wang, J. Witek, G. A. Landrum and S. Riniker, *J. Chem. Inf. Model.*, 2020, **60**, 2044–2058.
- 91 D. S. Nielsen, N. E. Shepherd, W. Xu, A. J. Lucke, M. J. Stoermer and D. P. Fairlie, *Chem. Rev.*, 2017, **117**, 8094–8128.
- 92 X. Wang, X. Gong, P. Li, D. Lai and L. Zhou, *Molecules*, 2018, **23**, 169.
- 93 G.-M. Suarez-Jimenez, A. Burgos-Hernandez and J.-M. Ezquerra-Brauer, *Mar. Drugs*, 2012, **10**, 963–986.
- 94 G. S. B. Andavan and R. Lemmens-Gruber, *Mar. Drugs*, 2010, **8**, 810–834.
- 95 S. Sivanathan and J. Scherckenbeck, *Molecules*, 2014, **19**, 12368–12420.
- 96 V. Gogineni and M. T. Hamann, *Biochim. Biophys. Acta - Gen. Subj.*, 2018, **1862**, 81–196.
- 97 X. Wang, Y. Li, X. Zhang, D. Lai and L. Zhou, *Molecules*, 2017, **22**, 2026.
- 98 E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, *Nat. Rev. Drug Discov.*, 2008, **7**, 608–624.

- 99 T. Sasaki, M. Takagi, T. Yaguchi, S. Miyadoh, T. Okada and M. Koyama, *J. Antibiot. (Tokyo)*, 1992, **45**, 692–697.
- 100 A. Harder, L. Holden-Dye, R. Walker and F. Wunderlich, *Parasitol. Res.*, 2005, **97**, S1–S10.
- 101 B. Over, P. McCarren, P. Artursson, M. Foley, F. Giordanetto, G. Grönberg, C. Hilgendorf, M. D. Lee, P. Matsson, G. Muncipinto, M. Pellisson, M. W. D. Perry, R. Svensson, J. R. Duvall and J. Kihlberg, *J. Med. Chem.*, 2014, **57**, 2746–2754.
- 102 A. Alex, D. S. Millan, M. Perez, F. Wakenhut and G. A. Whitlock, *Med. Chem. Commun.*, 2011, **2**, 669–674.
- 103 T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey and M. P. Jacobson, *J. Am. Chem. Soc.*, 2006, **128**, 14073–14080.
- 104 T. R. White, C. M. Renzelman, A. C. Rand, T. Rezai, C. M. McEwen, V. M. Gelev, R. A. Turner, R. G. Linington, S. S. F. Leung, A. S. Kalgutkar, J. N. Bauman, Y. Zhang, S. Liras, D. A. Price, A. M. Mathiowetz, M. P. Jacobson and R. S. Lokey, *Nat. Chem. Biol.*, 2011, **7**, 810–817.
- 105 T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson and R. S. Lokey, *J. Am. Chem. Soc.*, 2006, **128**, 2510–2511.
- 106 C. L. Ahlback, K. W. Lexa, A. T. Bockus, V. Chen, P. Crews, M. P. Jacobson and R. S. Lokey, *Future Med. Chem.*, 2015, **7**, 2121–2130.
- 107 M. Vukelić-Nikolić, A. Kolarević, K. Tomović, D. Yancheva, E. Cherneva, S. Najman and A. Smelcerović, *Nat. Prod. Commun.*, 2015, **10**, 1423–1426.
- 108 A. Smelcerovic, D. Yancheva, E. Cherneva, Z. Petronijevic, M. Lamshoeft and D. Herebian, *J. Mol. Struct.*, 2011, **985**, 397–402.
- 109 V. Pavlovic, A. Djordjevic, E. Cherneva, D. Yancheva and A. Smelcerovic, *Food Chem. Toxicol.*, 2012, **50**, 761–766.
- 110 V. Stankov-Jovanovic, J. C. Tabet, P. Dzodic, L. Daskalova, E. Cherneva, D. Yancheva and A. Smelcerovic, *Acta Chim. Slov.*, 2012, **59**, 939–943.
- 111 D. Lücke, T. Dalton, S. V. Ley and Z. E. Wilson, *Chem. - Eur. J.*, 2016, **22**, 4206–4217.
- 112 M. M. Shemyakin, Y. A. Ovchinnikov, V. T. Ivanov and A. A. Kiryushkin, *Tetrahedron*, 1963, **19**, 581–591.
- 113 Y. A. Ovchinnikov, V. T. Ivanov, A. V. Evstratov, I. I. Mikhaleva, V. F. Bystrov, S. L. Portnova, T. A. Balashova, E. N. Meshcheryakova and V. M. Tulchinsky, *Int. J. Pept. Protein Res.*, 1974, **6**, 465–498.
- 114 A. Prosperini, H. Berrada, M. J. Ruiz, F. Caloni, T. Coccini, L. J. Spicer, M. C. Perego and A. Lafranconi, *Front. Public Heal.*, 2017, **5**, 304.
- 115 V. T. Ivanov, A. V. Evstratov, L. V. Sumsкая, E. I. Melnik, T. S. Chumburidze, S. L. Portnova, T. A. Balashova and Y. A. Ovchinnikov, *FEBS Lett.*, 1973, **36**, 65–71.
- 116 M. Kansy, F. Senner and K. Gubernator, *J. Med. Chem.*, 1998, **41**, 1007–1010.
- 117 V. N. Tran, J. Viktorova, K. Augustynkova, N. Jelenova, S. Dobiasova, K. Rehorova, M. Fenclova, M. Stranska-Zachariasova, L. Vitek, J. Hajslova and T. Ruml, *Toxins (Basel)*, 2020, **12**, 148.
- 118 R. L. Hamill, C. E. Higgins, H. E. Boaz and M. Gorman, *Tetrahedron Lett.*, 1969, **10**, 4255–

- 4258.
- 119 Q. Wu, J. Patocka, E. Nepovimova and K. Kuca, *Front. Pharmacol.*, 2018, **9**, 1338.
- 120 K. Shiomi, R. Matsui, A. Kakei, Y. Yamaguchi, R. Masuma, H. Hatano, N. Arai, M. Isozaki, H. Tanaka, S. Kobayashi, A. Turberg and S. Omura, *J. Antibiot. (Tokyo)*., 2010, **63**, 77–82.
- 121 J. Scherkenbeck, P. Jeschke and A. Harder, *Curr. Top. Med. Chem.*, 2002, **2**, 759–777.
- 122 Y. Kodama, Y. Takeuchi and A. Suzuki, *Sci. Reports Meiji Seika Kaisha*, 1992, **31**, 1–8.
- 123 J. Baronsky, S. Bongaerts, M. Traeubel, H. C. Weiss and N. Urbanetz, *Eur. J. Pharm. Biopharm.*, 2009, **71**, 88–99.
- 124 H. Dyker, J. Scherkenbeck, D. Gondol, A. Goehrt and A. Harder, *J. Org. Chem.*, 2001, **66**, 3760–3766.
- 125 H. Dyker, A. Harder and J. Scherkenbeck, *Bioorganic Med. Chem. Lett.*, 2004, **14**, 6129–6132.
- 126 P. Jeschke, A. Harder, W. Etzel, W. Gau, G. Thielking, G. Bonse and K. Iinuma, *Pest Manag. Sci.*, 2001, **57**, 1000–1006.
- 127 A. Crisford, C. Murray, V. O'Connor, R. J. Edwards, N. Kruger, C. Welz, G. von Samson-Himmelstjerna, A. Harder, R. J. Walker and L. Holden-Dye, *Mol. Pharmacol.*, 2011, **79**, 1031–1043.
- 128 S. K. Buxton, C. Neveu, C. L. Charvet, A. P. Robertson and R. J. Martin, *Br. J. Pharmacol.*, 2011, **164**, 453–470.
- 129 R. J. Martin, S. K. Buxton, C. Neveu, C. L. Charvet and A. P. Robertson, *Exp. Parasitol.*, 2012, **132**, 40–46.
- 130 D. Kulke, G. von Samson-Himmelstjerna, S. M. Miltsch, A. J. Wolstenholme, A. R. Jex, R. B. Gasser, C. Ballesteros, T. G. Geary, J. Keiser, S. Townson, A. Harder and J. Krücken, *PLoS Negl. Trop. Dis.*, 2014, **8**, e3401.
- 131 G. Gessner, S. Meder, T. Rink and G. Boheim, *Pestic. Sci.*, 1996, **48**, 399–407.
- 132 S. Efimov, Y. Zgadzay and V. Klochkov, *Appl. Magn. Reson.*, 2014, **45**, 1225–1235.
- 133 N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger and W. F. van Gunsteren, *Comput. Phys. Commun.*, 2012, **183**, 890–903.
- 134 J. Witek, M. Mühlbauer, B. G. Keller, M. Blatter, A. Meissner, T. Wagner and S. Riniker, *ChemPhysChem*, 2017, **18**, 3309–3314.
- 135 J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.*, 2011, **134**, 174105.
- 136 R. G. Weiß, B. Ries, S. Wang and S. Riniker, *J. Chem. Phys.*, 2021, **154**, 084106.
- 137 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J. H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
- 138 Y. Naritomi and S. Fuchigami, *J. Chem. Phys.*, 2011, **134**, 065101.
- 139 R. Dornetshuber, M. R. Kamyar, P. Rawnduzi, I. Baburin, K. Kouri, E. Pilz, T. Hornbogen, R. Zocher, W. Berger and R. Lemmens-Gruber, *Biochem. Pharmacol.*, 2009, **77**, 1437–1444.

- 140 P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 1305–1323.
- 141 R Core Team, 2019.
- 142 V. F. Bystrov, Y. D. Gavrilov, V. T. Ivanov and Y. A. Ovchinnikov, *Eur. J. Biochem.*, 1977, **78**, 63–82.
- 143 C. L. Liotta and H. P. Harris, *J. Am. Chem. Soc.*, 1974, **96**, 2250–2252.
- 144 K. Neupert-Laves and M. Dobler, *Helv. Chim. Acta*, 1975, **58**, 432–442.
- 145 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 146 C. F. MacRae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Crystallogr.*, 2020, **53**, 226–235.
- 147 T. Stadelmann, G. Subramanian, S. Menon, C. E. Townsend, R. S. Lokey, M. O. Ebert and S. Riniker, *Org. Biomol. Chem.*, 2020, **18**, 7110–7126.
- 148 S. Elmshäuser, L. C. Straehle, J. Kranz, R. Krebber and J. Geyer, *J. Vet. Pharmacol. Ther.*, 2015, **38**, 74–79.
- 149 D. Xia, F. Zhou and L. Esser, *Cancer Drug Resist.*, 2019, **2**, 471–489.
- 150 Zoetis Services LLC, USA, *PCT/US2018/062749 (WO2019/108591)*, p.63.
- 151 C. M. Thiele, K. Petzold and J. Schleucher, *Chem. - Eur. J.*, 2009, **15**, 585–588.
- 152 T. L. Hwang and A. J. Shaka, *J. Magn. Reson. - Ser. A*, 1995, **112**, 275–279.
- 153 A. G. Palmer, J. Cavanagh, P. E. Wright and M. Rance, *J. Magn. Reson.*, 1991, **93**, 151–170.
- 154 M. J. Thrippleton and J. Keeler, *Angew. Chemie*, 2003, **115**, 4068–4071.
- 155 A. J. Shaka, C. J. Lee and A. Pines, *J. Magn. Reson.*, 1988, **77**, 274–293.
- 156 A. Bax and D. G. Davis, *J. Magn. Reson.*, 1985, **65**, 355–360.
- 157 P. H. C. Eilers, *Anal. Chem.*, 2003, **75**, 3631–3636.
- 158 T. D. Goddard and D. G. Kneller, *SPARKY3*, University of California, San Francisco (UCSF), 2004.
- 159 N. A. Farrow, O. Zhang, J. D. Forman-Kay and L. E. Kay, *Biochemistry*, 1995, **34**, 868–878.
- 160 R. R. Ernst, G. Bodenhausen and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1991.
- 161 C. L. Perrin and T. J. Dwyer, *Chem. Rev.*, 1990, **90**, 935–967.
- 162 Wolfram Research Inc., 2019.
- 163 R.-M. Korsi and J. Valkonen, *Thermochim. Acta*, 2003, **401**, 225–231.
- 164 B. A. C. Horta, P. T. Merz, P. F. J. Fuchs, J. Dolenc, S. Riniker and P. H. Hünenberger, *J. Chem. Theory Comput.*, 2016, **12**, 3825–3850.
- 165 R. W. Hockney, in *Methods in Computational Physics*, 1970, vol. 5, pp. 135–211.
- 166 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

- 167 J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 168 S. Riniker, A. P. E. Kunz and W. F. van Gunsteren, *J. Chem. Theory Comput.*, 2011, **7**, 1469–1475.
- 169 I. G. Tironi, R. Sperb, P. E. Smith and W. F. van Gunsteren, *J. Chem. Phys.*, 1995, **102**, 5451–5459.
- 170 A. P. Eichenberger, J. R. Allison, J. Dolenc, D. P. Geerke, B. A. C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang and W. F. van Gunsteren, *J. Chem. Theory Comput.*, 2011, **7**, 3379–3390.
- 171 W. F. van Gunsteren and M. Karplus, *J. Comput. Chem.*, 1980, **1**, 266–274.
- 172 O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Crystallogr.*, 2009, **42**, 339–341.
- 173 G. M. Sheldrick, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 2015, **71**, 3–8.
- 174 G. M. Sheldrick, *Acta Crystallogr. Sect. C Struct. Chem.*, 2015, **71**, 3–8.
- 175 T. Hamada, T. Sugawara, S. Matsunaga and N. Fusetani, *Tetrahedron Lett.*, 1994, **35**, 719–720.
- 176 T. Hamada, S. Matsunaga, G. Yano and N. Fusetani, *J. Am. Chem. Soc.*, 2005, **127**, 110–118.
- 177 T. Hamada, S. Matsunaga, M. Fujiwara, K. Fujita, H. Hirota, R. Schmucki, P. Güntert and N. Fusetani, *J. Am. Chem. Soc.*, 2010, **132**, 12941–12945.
- 178 A. Renevey and S. Riniker, *Eur. Biophys. J.*, 2017, **46**, 363–374.
- 179 A. Bhushan, P. J. Egli, E. E. Peters, M. F. Freeman and J. Piel, *Nat. Chem.*, 2019, **11**, 931–939.
- 180 R. J. Dubos, *J. Exp. Med.*, 1939, **70**, 1–10.
- 181 W. E. Herrell and D. Heilman, *J. Clin. Invest.*, 1941, **20**, 583–591.
- 182 A. S. Arseniev, I. L. Barsukov, V. F. Bystrov, A. L. Lomize and Y. A. Ovchinnikov, *FEBS Lett.*, 1985, **186**, 168–174.
- 183 S. B. Hladky and D. A. Haydon, *BBA - Biomembr.*, 1972, **274**, 294–312.
- 184 D. W. Urry, *Proc. Natl. Acad. Sci. U. S. A.*, 1971, **68**, 672–676.
- 185 W. R. Veatch, E. T. Fossel and E. R. Blout, *Biochemistry*, 1974, **13**, 5249–5256.
- 186 S. V. Sychev, N. A. Nevskaya, S. Jordanov, E. N. Shepel, A. I. Miroshnikov and V. T. Ivanov, *Bioorg. Chem.*, 1980, **9**, 121–151.
- 187 D. A. Doyle and B. A. Wallace, *Biophys. J.*, 1998, **75**, 635–640.
- 188 Y. Chen, A. Tucker and B. A. Wallace, *J. Mol. Biol.*, 1996, **264**, 757–769.
- 189 A. Renevey, ETH Zurich, 2018.
- 190 M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee and N. Khazanovich, *Nature*, 1993, **366**, 324–327.
- 191 T. D. Clark, L. K. Buehler and M. R. Ghadiri, *J. Am. Chem. Soc.*, 1998, **120**, 651–656.
- 192 W. S. Horne, C. D. Stout and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2003, **125**, 9372–9376.

- 193 T. D. Clark, J. M. Buriak, K. Kobayashi, M. P. Isler, D. E. McRee and M. Reza Ghadiri, *J. Am. Chem. Soc.*, 1998, **120**, 8949–8962.
- 194 D. T. Bong and M. R. Ghadiri, *Angew. Chemie - Int. Ed.*, 2001, **40**, 2163–2166.
- 195 C. D. Schwieters, G. A. Bermejo and G. M. Clore, *Protein Sci.*, 2018, **27**, 26–40.
- 196 V. F. Bystrov, A. S. Arseniev, I. L. Barsukov and A. L. Lomize, *Bull. Magn. Reson.*, 1986, **8**, 84–94.
- 197 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, 2012, **4**, 1–17.
- 198 M. Grogg, ETH Zurich, 2020.
- 199 W. Lee, M. Tonelli and J. L. Markley, *Bioinformatics*, 2015, **31**, 1325–1327.
- 200 C. Rigling and M. O. Ebert, *Magn. Reson. Chem.*, 2017, **55**, 655–661.
- 201 G. van Rossum, *The Python Library Reference, release 3.8.10*, Python Software Foundation, 2021.
- 202 T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing and J. Development Team, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 2016, pp. 87–90.
- 203 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.
- 204 R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L. P. Wang, T. J. Lane and V. S. Pande, *Biophys. J.*, 2015, **109**, 1528–1532.
- 205 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.
- 206 W. McKinney, in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 56–61.
- 207 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, *Nat. Methods*, 2020, **17**, 261–272.
- 208 A. Bax, *Protein Sci.*, 2003, **12**, 1–16.

- 209 J. Marchant, A. Bax and M. F. Summers, *J. Am. Chem. Soc.*, 2018, **140**, 6978–6983.
- 210 H. Sun, U. M. Reinscheid, E. L. Whitson, E. J. D’Auvergne, C. M. Ireland, A. Navarro-Vázquez and C. Griesinger, *J. Am. Chem. Soc.*, 2011, **133**, 14629–14636.
- 211 C. Camilloni and M. Vendruscolo, *J. Phys. Chem. B*, 2015, **119**, 653–661.
- 212 S. Y. Ko and C. Dalvit, *Int. J. Pept. Protein Res.*, 1992, **40**, 380–382.
- 213 U. Sternberg, F. T. Koch and P. Losso, COSMOS 6.0, 2019.
- 214 H. -R Loosli, H. Kessler, H. Oschkinat, H. -P Weber, T. J. Petcher and A. Widmer, *Helv. Chim. Acta*, 1985, **68**, 682–704.
- 215 K. Kajitani, M. Fujihashi, Y. Kobayashi, S. Shimizu, Y. Tsujimoto and K. Miki, *Proteins Struct. Funct. Genet.*, 2008, **70**, 1635–1639.
- 216 J. A. Losonczi, M. Andrec, M. W. F. Fischer and J. H. Prestegard, *J. Magn. Reson.*, 1999, **138**, 334–342.
- 217 C. M. Thiele, V. Schmidts, B. Böttcher, I. Louzao, R. Berger, A. Maliniak and B. Stevansson, *Angew. Chemie - Int. Ed.*, 2009, **48**, 6708–6712.
- 218 L. N. Wirz and J. R. Allison, *J. Phys. Chem. B*, 2015, **119**, 8223–8224.
- 219 M. E. Di Pietro, P. Tzvetkova, T. Gloge, U. Sternberg and B. Luy, *Liq. Cryst.*, 2020, **47**, 2043–2057.
- 220 C. M. Thiele, *Eur. J. Org. Chem.*, 2008, 5673–5685.
- 221 T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K. M. Fiebig and C. Griesinger, *J. Biomol. NMR*, 2004, **29**, 339–349.
- 222 C. Gayathri, N. V. Tsarevsky and R. R. Gil, *Chem. - Eur. J.*, 2010, **16**, 3622–3626.
- 223 A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler and B. Luy, *J. Magn. Reson.*, 2008, **192**, 314–322.
- 224 K. Fehér, S. Berger and K. E. Kövér, *J. Magn. Reson.*, 2003, **163**, 340–346.
- 225 J. R. Tolman and J. H. Prestegard, *J. Magn. Reson. - Ser. B*, 1996, **112**, 269–274.
- 226 C. M. Thiele and W. Bermel, *J. Magn. Reson.*, 2012, **216**, 134–143.
- 227 K. E. Kövér and K. Fehér, *J. Magn. Reson.*, 2004, **168**, 307–313.
- 228 H. Kessler, H. -R Loosli and H. Oschkinat, *Helv. Chim. Acta*, 1985, **68**, 661–681.
- 229 M. Möllhoff and U. Sternberg, *J. Mol. Model.*, 2001, **7**, 90–102.
- 230 J. A. Aguilar, M. Nilsson, G. Bodenhausen and G. A. Morris, *Chem. Commun.*, 2012, **48**, 811–813.
- 231 U. Sternberg, R. Witter and A. S. Ulrich, *J. Biomol. NMR*, 2007, **38**, 23–39.
- 232 U. Sternberg, F. T. Kock, M. Bräuer, M. Kunert and E. Anders, *J. Mol. Model.*, 2001, **7**, 54–64.
- 233 M. O. Marcarino, M. M. Zanardi, S. Cicetti and A. M. Sarotti, *Acc. Chem. Res.*, 2020, **53**, 1922–1932.
- 234 A. M. Sarotti and S. C. Pellegrinet, *J. Org. Chem.*, 2009, **74**, 7254–7260.

- 235 A. Mitra, P. J. Seaton, R. A. Assarpour and T. Williamson, *Tetrahedron*, 1998, **54**, 15489–15498.
- 236 F. C. Wermter, N. Mitschke, C. Bock and W. Dreher, *Magn. Reson. Mater. Physics, Biol. Med.*, 2017, **30**, 579–590.
- 237 S. Berger, *Tetrahedron*, 1981, **37**, 1607–1611.
- 238 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839–1862.
- 239 W. Hehre, P. Klunzinger, B. Deppmeier, A. Driessen, N. Uchida, M. Hashimoto, E. Fukushi and Y. Takata, *J. Nat. Prod.*, 2019, **82**, 2299–2306.
- 240 M. A. Iron, *J. Chem. Theory Comput.*, 2017, **13**, 5798–5819.
- 241 R. Jain, T. Bally and P. R. Rablen, *J. Org. Chem.*, 2009, **74**, 4017–4023.
- 242 P. R. Rablen, S. A. Pearlman and J. Finkbiner, *J. Phys. Chem. A*, 1999, **103**, 7357–7363.
- 243 L. B. Krivdin, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2019, **112–113**, 103–156.
- 244 G. L. Stoychev, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2018, **14**, 4756–4771.
- 245 V. A. Semenov, D. O. Samultsev and L. B. Krivdin, *Magn. Reson. Chem.*, 2014, **52**, 686–693.
- 246 E. Artikis and C. L. Brooks, *Biophys. J.*, 2019, **117**, 258–268.
- 247 P. Cmoch, P. Krzeczyński and A. Leś, *Molecules*, 2018, **23**, 161.
- 248 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78.
- 249 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, e1327.
- 250 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
- 251 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.
- 252 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- 253 S. Kozuch and J. M. L. Martin, *J. Comput. Chem.*, 2013, **34**, 2327–2344.
- 254 F. Jensen, *J. Chem. Theory Comput.*, 2015, **11**, 132–138.
- 255 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 256 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 257 J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.
- 258 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
- 259 D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville and C. Ochsenfeld, *J. Chem. Theory Comput.*, 2014, **10**, 572–578.
- 260 M. T. De Oliveira, J. M. A. Alves, A. A. C. Braga, D. J. D. Wilson and C. A. Barboza, *J. Chem. Theory Comput.*, 2021, **17**, 6876–6885.
- 261 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- 262 M. Foroozandeh, R. W. Adams, P. Kiraly, M. Nilsson and G. A. Morris, *Chem. Commun.*, 2015, **51**, 15410–15413.
- 263 G. Landrum, RDKit: Open-Source Cheminformatics Software, <http://www.rdkit.org/>,

- (accessed 8 April 2021).
- 264 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 265 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 266 K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251–8260.
- 267 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 268 *Turbomole 7.0 basis set library*.
- 269 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.
- 270 C. Hättig, *Phys. Chem. Chem. Phys.*, 2005, **7**, 59–66.
- 271 H. Nguyen, D. A. Case and A. S. Rose, *Bioinformatics*, 2018, **34**, 1241–1242.
- 272 N. M. O’Boyle, C. Morley and G. R. Hutchison, *Chem. Cent. J.*, 2008, **2**, 1–7.
- 273 Chemical Book, https://www.chemicalbook.com/SpectrumEN_1120-64-5_13CNMR.htm, (accessed 24 May 2021).
- 274 R. J. Abraham and J. Fisher, *Magn. Reson. Chem.*, 1985, **23**, 856–861.
- 275 R. I. Khusnutdinov, N. A. Shchadneva, A. R. Baiguzina, Y. Y. Lavrentieva and U. M. Dzhemilev, *Russ. Chem. Bull.*, 2002, **51**, 2074–2079.
- 276 W. Kitching, H. A. Olszowy, I. Schott, W. Adcock and D. P. Cox, *J. Organomet. Chem.*, 1986, **310**, 269–284.
- 277 R. Koçak, G. Borsato, O. De Lucchi and A. Daştan, *Helv. Chim. Acta*, 2014, **97**, 537–545.
- 278 C. Koradin, W. Dohle, A. L. Rodriguez, B. Schmid and P. Knochel, *Tetrahedron*, 2003, **59**, 1571–1587.
- 279 R. Labes, C. Battilocchio, C. Mateos, G. R. Cumming, O. De Frutos, J. A. Rincón, K. Binder and S. V. Ley, *Org. Process Res. Dev.*, 2017, **21**, 1419–1422.
- 280 V. G. Landge, R. Babu, V. Yadav, M. Subaramanian, V. Gupta and E. Balaraman, *J. Org. Chem.*, 2020, **85**, 9876–9886.
- 281 D. A. Lanfranchi, M. Vellutini, M. C. Blanc, P. Bradesi, J. Casanova and F. Tomi, *Spectrosc. Lett.*, 2010, **43**, 36–43.
- 282 T. Laurens, D. Nicole, P. Rubini, J. C. Lauer, M. Matlengiewicz and N. Henzel, *Magn. Reson. Chem.*, 1991, **29**, 1119–1129.
- 283 J. Lee, S. Wang, M. Callahan and P. Nagorny, *Org. Lett.*, 2018, **20**, 2067–2070.
- 284 M. Leyendecker, N.-C. Meyer and C. M. Thiele, *Angew. Chemie*, 2017, **129**, 11629–11632.
- 285 R. J. Abraham and M. Reid, *J. Chem. Soc. Perkin Trans. 2*, 2002, **2**, 1081–1091.
- 286 T. Li, G. B. Hammond and B. Xu, *Chem. - Eur. J.*, 2021, **27**, 9737–9741.
- 287 Y. Li, X. Liu, H. Jiang, B. Liu, Z. Chen and P. Zhou, *Angew. Chemie - Int. Ed.*, 2011, **50**, 6341–6345.
- 288 C. K. Lin and T. J. Lu, *Tetrahedron*, 2010, **66**, 9688–9693.
- 289 Z. Liu, P. Wang, Z. Yan, S. Chen, D. Yu, X. Zhao and T. Mu, *Beilstein J. Org. Chem.*, 2020, **16**,

- 645–656.
- 290 J. Luo, B. Hu, A. Sam and T. L. Liu, *Org. Lett.*, 2018, **20**, 361–364.
- 291 M. Mąkosza and M. Judka, *Chem. - Eur. J.*, 2002, **8**, 4234–4240.
- 292 A. N. Mamros, P. R. Sharrow, W. E. Weller, M. R. Luderer, J. D. Fair, K. O. Pazehoski and M. R. Luderer, *Arkivoc*, 2011, **2011**, 23–33.
- 293 S. Manna, S. Maity, S. Rana, S. Agasti and D. Maiti, *Org. Lett.*, 2012, **14**, 1736–1739.
- 294 M. Mansour, R. Giacobazzi, A. Ouali, M. Taillefer and A. Jutand, *Chem. Commun.*, 2008, 6051–6053.
- 295 G. V. Markin, Y. A. Shevelev, G. A. Domrachev, G. K. Fukin, E. V. Baranov, M. A. Lopatin, V. A. Kuropatov, A. I. Kirillov, A. S. Shavyrin and Y. A. Kurskii, *Russ. Chem. Bull.*, 2008, **57**, 1970–1974.
- 296 S. Agasti, S. Maiti, S. Maity, M. Anniyappan, M. B. Talawar and D. Maiti, *Polyhedron*, 2019, **172**, 120–124.
- 297 E. B. McLean, D. T. Mooney, D. J. Burns and A. L. Lee, *Org. Lett.*, 2022, **24**, 686–691.
- 298 C. B. McPake, C. B. Murray and G. Sandford, *ChemSusChem*, 2012, **5**, 312–319.
- 299 S. S. Meng, L. R. Lin, X. Luo, H. J. Lv, J. L. Zhao and A. S. C. Chan, *Green Chem.*, 2019, **21**, 6187–6193.
- 300 J. L. Mieloszynski, M. Schneider, D. Paquet and C. G. Andrieu, *Recl. des Trav. Chim. des Pays-Bas*, 1985, **104**, 9–15.
- 301 S. A. Miller, J. M. Bobbitt and N. E. Leadbeater, *Org. Biomol. Chem.*, 2017, **15**, 2817–2822.
- 302 T. Mori, C. Ishii and M. Kimura, *Org. Process Res. Dev.*, 2019, **23**, 1709–1717.
- 303 T. Mori, Y. Takeuchi and M. Hojo, *Tetrahedron Lett.*, 2020, **61**, 151518.
- 304 T. Moriuchi, K. Kikushima, T. Kajikawa and T. Hirao, *Tetrahedron Lett.*, 2009, **50**, 7385–7387.
- 305 Y. Motoyama, K. Kamo and H. Nagashima, *Org. Lett.*, 2009, **11**, 1345–1348.
- 306 S. R. Mudshinge, C. S. Potnis, B. Xu and G. B. Hammond, *Green Chem.*, 2020, **22**, 4161–4164.
- 307 R. A. Aitken, D. M. M. Farrell and E. H. M. Kirton, *Chem. Heterocycl. Compd.*, 2001, **37**, 1526–1531.
- 308 V. Müller, D. Ghorai, L. Capdevila, A. M. Messinis, X. Ribas and L. Ackermann, *Org. Lett.*, 2020, **22**, 7034–7040.
- 309 T. Murata, M. Hiyoshi, M. Ratanasak, J. Y. Hasegawa and T. Ema, *Chem. Commun.*, 2020, **56**, 5783–5786.
- 310 J. Nakazawa, M. Mizuki, Y. Shimazaki, F. Tani and Y. Naruta, *Org. Lett.*, 2006, **8**, 4275–4278.
- 311 G. R. Newkome, V. K. Gupta, R. W. Griffin and S. Arai, *J. Org. Chem.*, 1987, **52**, 5480–5482.
- 312 M. Niestroj and W. P. Neumann, *Chem. Ber.*, 1996, **129**, 45–51.
- 313 U. Niggli and M. Neuenschwander, *Helv. Chim. Acta*, 1990, **73**, 2199–2208.

- 314 N. F. Nikitas, D. I. Tzaras, I. Triandafillidi and C. G. Kokotos, *Green Chem.*, 2020, **22**, 471–477.
- 315 T. Niu, S. Chen, M. Hong, T. Zhang, J. Chen, X. Dong and B. Ni, *Green Chem.*, 2020, **22**, 5042–5049.
- 316 V. L. Novikov, O. P. Shestak and V. A. Denisenko, *Russ. Chem. Bull.*, 2010, **59**, 1600–1604.
- 317 R. A. Nyquist, R. Streck and G. Jeschek, *J. Mol. Struct.*, 1996, **377**, 113–128.
- 318 R. A. Aitken and N. Karodia, *Liebigs Ann.*, 1997, 779–783.
- 319 K. Ogura, M. Yamashita, M. Suzuki, S. Furukawa and G. I. Tsuchihashi, *Bull. Chem. Soc. Jpn.*, 1984, **57**, 1637–1642.
- 320 K. Ohsawa, M. Yoshida and T. Doi, *J. Org. Chem.*, 2013, **78**, 3438–3444.
- 321 U. B. Patil, S. S. Shendage and J. M. Nagarkar, *Synth.*, 2013, **45**, 3295–3299.
- 322 K. Paudel, B. Pandey, S. Xu, D. K. Taylor, D. L. Tyer, C. L. Torres, S. Gallagher, L. Kong and K. Ding, *Org. Lett.*, 2018, **20**, 4478–4481.
- 323 T. P. Petersen, A. F. Larsen, A. Ritzén and T. Ulven, *J. Org. Chem.*, 2013, **78**, 4190–4195.
- 324 R. Pilli, V. Balakrishnan, R. Chandrasekaran and R. Rasappan, *Org. Biomol. Chem.*, 2019, **17**, 1749–1753.
- 325 A. Piontek, W. Ochędzan-Siodłak, E. Bisz and M. Szostak, *Adv. Synth. Catal.*, 2019, **361**, 2329–2336.
- 326 A. I. Popov, A. J. Smetana, J. -P Kintzinger and T. T.-T. Nguyêen, *Helv. Chim. Acta*, 1980, **63**, 668–673.
- 327 S. Pradhan, V. Sharma and I. Chatterjee, *Org. Lett.*, 2021, **23**, 6148–6152.
- 328 X. Qi, C. L. Li and X. F. Wu, *Chem. - Eur. J.*, 2016, **22**, 5835–5838.
- 329 Y. Akita, Y. Itagaki, S. Takizawa and A. Ohta, *Chem. Pharm. Bull.*, 1989, **37**, 1477–1480.
- 330 W. Qiu, S. Shi, R. Li, X. Lin, L. Rao and Z. Sun, *Chinese J. Chem.*, 2021, **39**, 1255–1258.
- 331 I. Ramos-Tomillero, M. Paradís-Bas, I. De Pinho Ribeiro Moreira, J. M. Bofill, E. Nicolás and F. Albericio, *Molecules*, 2015, **20**, 5409–5422.
- 332 R. Ray, S. Chandra, D. Maiti and G. K. Lahiri, *Chem. - Eur. J.*, 2016, **22**, 8814–8822.
- 333 S. C. Richter and M. Oestreich, *Chem. - Eur. J.*, 2019, **25**, 8508–8512.
- 334 M. R. Rohman, M. Rajbangshi, B. M. Laloo, P. R. Sahu and B. Myrboh, *Tetrahedron Lett.*, 2010, **51**, 2862–2864.
- 335 B. V. Rokade, S. K. Malekar and K. R. Prabhu, *Chem. Commun.*, 2012, **48**, 5506–5508.
- 336 N. C. Rol and A. D. H. Clague, *Org. Magn. Reson.*, 1981, **16**, 187–194.
- 337 P. Saisaha, L. Buettner, M. Van Der Meer, R. Hage, B. L. Feringa, W. R. Browne and J. W. De Boer, *Adv. Synth. Catal.*, 2013, **355**, 2591–2603.
- 338 S. Saito, T. Ohwada and K. Shudo, *J. Am. Chem. Soc.*, 1995, **117**, 11081–11084.
- 339 A. Sakakura, M. Katsukawa and K. Ishihara, *Org. Lett.*, 2005, **7**, 1999–2002.

- 340 C. Altona, J. H. Ippel, A. J. A. W. Hoekzema, C. Erkelens, M. Groesbeek and L. A. Donders, *Magn. Reson. Chem.*, 1989, **27**, 564–576.
- 341 Y. Sakurai, K. Ikai, K. Hayakawa, Y. Ogiwara and N. Sakai, *Bull. Chem. Soc. Jpn.*, 2021, **94**, 1882–1893.
- 342 Y. Sawama, Y. Yabe, M. Shigetsura, T. Yamada, S. Nagata, Y. Fujiwara, T. Maegawa, Y. Monguchi and H. Sajiki, *Adv. Synth. Catal.*, 2012, **354**, 777–782.
- 343 D. Schnurpfeil, *J. für Prakt. Chemie*, 1983, **325**, 481–488.
- 344 F. Schoenebeck, J. A. Murphy, S. Z. Zhou, Y. Uenoyama, Y. Miclo and T. Tuttle, *J. Am. Chem. Soc.*, 2007, **129**, 13368–13369.
- 345 H. Schumann, M. Speis, W. P. Bosman, J. M. M. Smits and P. T. Beurskens, *J. Organomet. Chem.*, 1991, **403**, 165–182.
- 346 I. I. Schuster, *J. Org. Chem.*, 1981, **46**, 5110–5118.
- 347 I. I. Schuster, M. Parvez and A. J. Freyer, *J. Org. Chem.*, 1988, **53**, 5819–5825.
- 348 S. Seo, J. B. Taylor and M. F. Greaney, *Chem. Commun.*, 2012, **48**, 8270–8272.
- 349 A. Shahrisa, R. Tabrizi and H. R. Ahsani, *Org. Prep. Proced. Int.*, 2000, **32**, 47–55.
- 350 A. S. Shashkov, E. G. Cherepanova, L. I. Kas'yan, L. Y. Gnedenkova and M. F. Bombushkar', *Bull. Acad. Sci. USSR Div. Chem. Sci.*, 1980, **29**, 382–387.
- 351 L. Anzalone and J. A. Hirsch, *J. Org. Chem.*, 1985, **50**, 2607–2613.
- 352 S. S. Shen, V. Kartika, Y. S. Tan, R. D. Webster and K. Narasaka, *Tetrahedron Lett.*, 2012, **53**, 986–990.
- 353 S. R. Sheng, L. L. Wu and X. Huang, *J. Chem. Res. - Part S*, 2003, 258–259.
- 354 W. J. Shi and Z. J. Shi, *Chinese J. Chem.*, 2018, **36**, 183–186.
- 355 K. Shibata, K. I. Takao and A. Ogura, *J. Org. Chem.*, 2021, **86**, 10067–10087.
- 356 T. Shintou, W. Kikuchi and T. Mukaiyama, *Bull. Chem. Soc. Jpn.*, 2003, **76**, 1645–1667.
- 357 T. Shirai, K. Sugimoto, M. Iwasaki, R. Sumida, H. Fujita and Y. Yamamoto, *Synlett*, 2019, **30**, 972–976.
- 358 I. K. Sideri, E. Voutyritsa and C. G. Kokotos, *Synlett*, 2018, **29**, 1324–1328.
- 359 J. Siu, I. R. Baxendale and S. V. Ley, *Org. Biomol. Chem.*, 2004, **2**, 160–167.
- 360 A. B. Smith, R. A. Rivero, K. J. Hale and H. A. Vaccaro, *J. Am. Chem. Soc.*, 1991, **113**, 2092–2112.
- 361 G. Sun, X. Lv, Y. Zhang, M. Lei and L. Hu, *Org. Lett.*, 2017, **19**, 4235–4238.
- 362 G. Aridos and K. K. Laali, *J. Org. Chem.*, 2011, **76**, 8088–8094.
- 363 Y. Suzuki, T. Yoshino, K. Moriyama and H. Togo, *Tetrahedron*, 2011, **67**, 3809–3814.
- 364 D. W. Tan, J.-B. Xie, Q. Li, H.-X. Li, J.-C. Li, H.-Y. Li and J.-P. Lang, *Dalt. Trans.*, 2014, **43**, 14061–14071.
- 365 Y. F. Tan, Y. Chen, R. X. Li, Z. Guan and Y. H. He, *Tetrahedron Lett.*, 2022, **89**, 153593.

- 366 G. Tang, Z. Gong, W. Han and X. Sun, *Tetrahedron Lett.*, 2018, **59**, 658–662.
- 367 J. Tang, S. Zhao, Y. Wei, Z. Quan and C. Huo, *Org. Biomol. Chem.*, 2017, **15**, 1589–1592.
- 368 Y. Tang, R. I. L. Meador, C. T. Malinchak, E. E. Harrison, K. A. McCaskey, M. C. Hempel and T. W. Funk, *J. Org. Chem.*, 2020, **85**, 1823–1834.
- 369 D. Tanini, C. Dalia and A. Capperucci, *Green Chem.*, 2021, **23**, 5680–5686.
- 370 P. Thapa, S. Hazoor, B. Chouhan, T. T. Vuong and F. W. Foss, *J. Org. Chem.*, 2020, **85**, 9096–9105.
- 371 X. Tian, Y. L. Ren, F. Ren, X. Cheng, S. Zhao and J. Wang, *Synlett*, 2018, **29**, 2444–2448.
- 372 Y. Tian, X. Guo, M. Li, C. Li, X. Hu, L. Jin, N. Sun, B. Hu and Z. Shen, *Org. Lett.*, 2021, **23**, 3928–3932.
- 373 M. J. Aurell, P. Gaviña and R. Mestres, *Tetrahedron*, 1994, **50**, 2571–2582.
- 374 M. Topolski, *J. Org. Chem.*, 1995, **60**, 5588–5594.
- 375 L. A. Trimble, P. B. Reese and J. C. Vederas, *J. Am. Chem. Soc.*, 1985, **107**, 2175–2177.
- 376 D. Tsuchiya, K. Moriyama and H. Togo, *Synlett*, 2011, **18**, 2701–2704.
- 377 M. Uyanik, R. Fukatsu and K. Ishihara, *Chem. - An Asian J.*, 2010, **5**, 456–460.
- 378 E. M. M. van den Berg, F. J. H. M. Jansen, A. T. J. W. de Goede, A. U. Baldew and J. Lugtenburg, *Recl. des Trav. Chim. des Pays-Bas*, 1990, **109**, 287–297.
- 379 N. J. R. van Eikema Hommes and T. Clark, *J. Mol. Model.*, 2005, **11**, 175–185.
- 380 K. Walsh, H. F. Sneddon and C. J. Moody, *Org. Lett.*, 2014, **16**, 5224–5227.
- 381 J. C. Wan, J. M. Huang, Y. H. Jhan and J. C. Hsieh, *Org. Lett.*, 2013, **15**, 2742–2745.
- 382 A. Wang and H. Jiang, *J. Org. Chem.*, 2010, **75**, 2321–2326.
- 383 C. Wang, J. Wan, Z. Zheng and Y. Pan, *Tetrahedron*, 2007, **63**, 5071–5075.
- 384 Chemical Book, https://www.chemicalbook.com/SpectrumEN_1120-64-5_1HNMR.htm, (accessed 24 May 2021).
- 385 M. Avi, M. H. Fechter, K. Gruber, F. Belaj, P. Pöchlauser and H. Griengl, *Tetrahedron*, 2004, **60**, 10411–10418.
- 386 D. Wang, D. Kuang, F. Zhang, C. Yang and X. Zhu, *Adv. Synth. Catal.*, 2015, **357**, 714–718.
- 387 P. Wang, J. Cai, J. Yang, C. Sun, L. Li, H. Hu and M. Ji, *Tetrahedron Lett.*, 2013, **54**, 533–535.
- 388 Y. Wang, H. Zhao, Z. Hu, M. Kimura, Z. Zhiqiang, L. Peng and K. Hiratsuka, *Synthesis (Stuttg.)*, 2011, **2**, 287–291.
- 389 M. Westermayer, G. Häfelinger and C. Regelman, *Tetrahedron*, 1984, **40**, 1845–1854.
- 390 University of Florida Research Foundation, USA, PCT/US2012/068943, (WO2013/090260), p 60.
- 391 X. F. Wu, J. Schranck, H. Neumann and M. Beller, *Chem. Commun.*, 2011, **47**, 12462–12463.
- 392 A. Xia, X. Xie, X. Hu, W. Xu and Y. Liu, *J. Org. Chem.*, 2019, **84**, 13841–13857.

- 393 P. Xiao, Z. Tang, K. Wang, H. Chen, Q. Guo, Y. Chu, L. Gao and Z. Song, *J. Org. Chem.*, 2018, **83**, 1687–1700.
- 394 Y. Xiao, Y. Xu, H. S. Cheon and J. Chae, *J. Org. Chem.*, 2013, **78**, 5804–5809.
- 395 X. Xie and S. S. Stahl, *J. Am. Chem. Soc.*, 2015, **137**, 3767–3770.
- 396 N. R. Babij, E. O. McCusker, G. T. Whiteker, B. Canturk, N. Choy, L. C. Creemer, C. V. D. Amicis, N. M. Hewlett, P. L. Johnson, J. A. Knobelsdorf, F. Li, B. A. Lorsbach, B. M. Nugent, S. J. Ryan, M. R. Smith and Q. Yang, *Org. Process Res. Dev.*, 2016, **20**, 661–667.
- 397 B. Xu, E. M. Hartigan, G. Feula, Z. Huang, J. P. Lumb and B. A. Arndtsen, *Angew. Chemie - Int. Ed.*, 2016, **55**, 15802–15806.
- 398 F. Xu, Y. L. Zhong, H. Li, J. Qi, R. Desmond, Z. J. Song, J. Park, T. Wang, M. Truppo, G. R. Humphrey and R. T. Ruck, *Org. Lett.*, 2017, **19**, 5880–5883.
- 399 H. J. Xu, Y. F. Liang, Z. Y. Cai, H. X. Qi, C. Y. Yang and Y. S. Feng, *J. Org. Chem.*, 2011, **76**, 2296–2300.
- 400 R. R. Yadav, R. A. Vishwakarma and S. B. Bharate, *Tetrahedron Lett.*, 2012, **53**, 5958–5960.
- 401 Z. Yan, X.-A. Yuan, Y. Zhao, C. Zhu and J. Xie, *Angew. Chemie*, 2018, **130**, 13088–13092.
- 402 C. Yang, F. Zhang, G. J. Deng and H. Gong, *J. Org. Chem.*, 2019, **84**, 181–190.
- 403 D. Ye, Z. Liu, J. L. Sessler and C. Lei, *Chem. Commun.*, 2020, **56**, 11811–11814.
- 404 B. Yu, Y. Zhao, H. Zhang, J. Xu, L. Hao, X. Gao and Z. Liu, *Chem. Commun.*, 2014, **50**, 2330–2333.
- 405 L. Yu, H. Li, X. Zhang, J. Ye, J. Liu, Q. Xu and M. Lautens, *Org. Lett.*, 2014, **16**, 1346–1349.
- 406 S. Zahim, K. Delacroix, A. Carrier, T. Berranger, J. Bergraser, P. G. Echeverria and L. Petit, *Org. Process Res. Dev.*, 2022, **26**, 199–206.
- 407 A. Bagno, S. Gerard, J. Kevelam, E. Menna and G. Scorrano, *Chem. - Eur. J.*, 2000, **6**, 2915–2924.
- 408 W. Zhan, M. Tong, L. Ji, H. Zhang, Z. Ge, X. Wang and R. Li, *Chinese Chem. Lett.*, 2019, **30**, 973–976.
- 409 H. Zhang, S. Fu, M. Tao, Z. Liao, C. Qian and X. Chen, *Res. Chem. Intermed.*, 2016, **42**, 3325–3332.
- 410 H. Zhang, T. Guo, M. Wu, X. Huo, S. Tang, X. Wang and J. Liu, *Tetrahedron Lett.*, 2021, **67**, 152878.
- 411 K. Zhang, L. Chang, Q. An, X. Wang and Z. Zuo, *J. Am. Chem. Soc.*, 2019, **141**, 10556–10564.
- 412 L. Zhang, Z. Zha, Z. Wang and S. Fu, *Tetrahedron Lett.*, 2010, **51**, 1426–1429.
- 413 X. Zhang, J. Sun, Y. Ding and L. Yu, *Org. Lett.*, 2015, **17**, 5840–5842.
- 414 S. Zheng, C. Yu and Z. Shen, *Org. Lett.*, 2012, **14**, 3644–3647.
- 415 Y. Zheng, J. Zhang, X. Cheng, X. Xu and L. Zhang, *Angew. Chemie*, 2019, **131**, 5295–5299.
- 416 C. Zhu, Z. Zhang, W. Ding, J. Xie, Y. Chen, J. Wu, X. Chen and H. Ying, *Green Chem.*, 2014, **16**, 1131–1138.

- 417 X. Zhu, C. Liu, Y. Liu, H. Yang and H. Fu, *Chem. Commun.*, 2020, **56**, 12443–12446.
- 418 W. F. Bailey and E. A. Cioffi, *Magn. Reson. Chem.*, 1987, **25**, 181–183.
- 419 Y. Zhu, B. Zhao and Y. Shi, *Org. Lett.*, 2013, **15**, 992–995.
- 420 B. M. Zimmermann, T. T. Ngoc, D. I. Tzaras, T. Kaicharla and J. F. Teichert, *J. Am. Chem. Soc.*, 2021, **143**, 16865–16873.
- 421 M. A. Zolfigol, A. Khazaei, A. R. Moosavi-Zare, A. Zare, H. G. Kruger, Z. Asgari, V. Khakyzadeh and M. Kazem-Rostami, *J. Org. Chem.*, 2012, **77**, 3640–3645.
- 422 H. Zou, C. Hu, K. Chen, G. Xiao and X. Peng, *Synlett*, 2018, **29**, 2181–2184.
- 423 J. E. Baldwin and K. A. Black, *J. Am. Chem. Soc.*, 1984, **106**, 1029–1040.
- 424 F. Ballaschk and S. F. Kirsch, *Green Chem.*, 2019, **21**, 5896–5903.
- 425 Y. L. Ban, J. L. Dai, X. L. Jin, Q. B. Zhang and Q. Liu, *Chem. Commun.*, 2019, **55**, 9701–9704.
- 426 T. M. Barclay, I. Del Río, R. A. Gossage and S. M. Jackson, *Can. J. Chem.*, 2003, **81**, 1482–1491.
- 427 C. Berini, O. H. Winkelmann, J. Otten, D. A. Vicic and O. Navarro, *Chem. - Eur. J.*, 2010, **16**, 6857–6860.
- 428 V. Bertolini, R. Appiani, M. Pallavicini and C. Bolchi, *J. Org. Chem.*, 2021, **86**, 15712–15716.
- 429 Chemical Book, https://www.chemicalbook.com/SpectrumEN_558-37-2_1HNMR.htm, (accessed 24 May 2021).
- 430 G. Bian, H. Fan, S. Yang, H. Yue, H. Huang, H. Zong and L. Song, *J. Org. Chem.*, 2013, **78**, 9137–9142.
- 431 R. R. Biekofsky, A. B. Pomilio, R. A. Aristegui and R. H. Contreras, *J. Mol. Struct.*, 1995, **344**, 143–150.
- 432 R. T. Boeré, J. A. Eng, K. Preuss, M. Parvez, C. D. Bryan and A. W. Cordes, *Can. J. Chem.*, 1994, **72**, 1171–1180.
- 433 C. Bonfio, C. Caumes, C. D. Duffy, B. H. Patel, C. Percivalle, M. Tsanakopoulou and J. D. Sutherland, *J. Am. Chem. Soc.*, 2019, **141**, 3934–3939.
- 434 M. C. Boucenna, J. S. Davidson, A. McKee, A. L. Porte and D. C. Apperley, *J. Chem. Soc. Perkin Trans. 2*, 1995, **2**, 1381–1387.
- 435 G. Bresciani, F. Marchetti and G. Pampaloni, *New J. Chem.*, 2019, **43**, 10821–10825.
- 436 S. Budweg, K. Junge and M. Beller, *Chem. Commun.*, 2019, **55**, 14143–14146.
- 437 J. A. Campbell, G. McDougald, H. McNab, L. V. C. Rees and R. G. Tyas, *Synthesis (Stuttg.)*, 2007, **20**, 3179–3184.
- 438 B. Capon and B. Guo, *J. Am. Chem. Soc.*, 1988, **110**, 5144–5147.
- 439 P. -A Carrupt and P. Vogel, *Helv. Chim. Acta*, 1989, **72**, 1008–1028.
- 440 Chemical Book, https://www.chemicalbook.com/SpectrumEN_558-37-2_13CNMR.htm, (accessed 24 May 2021).
- 441 J. C. Chalchat, R. P. Garry, B. Lacroix and A. Michet, *Magn. Reson. Chem.*, 1985, **23**, 977–

- 984.
- 442 N. Chang and C. Chiu, *J. Chinese Chem. Soc.*, 1993, **40**, 379–384.
- 443 K. Chen, W. Chen, X. Yi, W. Chen, M. Liu and H. Wu, *Chem. Commun.*, 2019, **55**, 9287–9290.
- 444 K. Chen, C. Liu, L. Deng and G. Xu, *Steroids*, 2010, **75**, 513–516.
- 445 L. Chen, S. Wang, P. Werz, Z. Han and D. P. Gates, *Heteroat. Chem.*, 2018, **29**, e21474.
- 446 L. X. Chen, Y. H. Huang, H. Cong and Z. Tao, *Chem. Pap.*, 2018, **72**, 661–667.
- 447 C. Cheng and M. Brookhart, *Angew. Chemie - Int. Ed.*, 2012, **51**, 9422–9424.
- 448 G. Cheng and M. Luo, *European J. Org. Chem.*, 2011, 2519–2523.
- 449 Y. A. Cheng, W. Z. Yu and Y. Y. Yeung, *J. Org. Chem.*, 2016, **81**, 545–552.
- 450 C. Y. Chern, C. C. Tseng, R. H. Hsiao, F. F. Wong and Y. H. Kuo, *Heteroat. Chem.*, 2019, **2019**, 5053702.
- 451 R. J. Abraham and N. J. Ainger, *J. Chem. Soc. Perkin Trans. 2*, 1999, 441–448.
- 452 S. B. Choe and K. J. Klabunde, *J. Organomet. Chem.*, 1989, **359**, 409–418.
- 453 J. Clark and G. Hitiris, *Spectrochim. Acta Part A Mol. Spectrosc.*, 1984, **40**, 75–79.
- 454 M. D. Coburn, C. B. Storm, D. W. Moore and T. G. Archibald, *Magn. Reson. Chem.*, 1990, **28**, 16–20.
- 455 W. J. Colucci, S. J. Jungk and R. D. Gandour, *Magn. Reson. Chem.*, 1985, **23**, 335–343.
- 456 R. M. Denton, J. An, P. Lindovska and W. Lewis, *Tetrahedron*, 2012, **68**, 2899–2905.
- 457 A. Dewanji, C. Mück-Lichtenfeld and A. Studer, *Angew. Chemie - Int. Ed.*, 2016, **55**, 6749–6752.
- 458 G. Dilauro, A. Francesca Quivelli, P. Vitale, V. Capriati and F. M. Perna, *Angew. Chemie*, 2019, **131**, 1813–1816.
- 459 Y. Ding, S. Luo, L. Ma and J. An, *J. Org. Chem.*, 2019, **84**, 17.
- 460 P. G. Dumanski, C. J. Easton, S. F. Lincoln and J. S. Simpson, *Aust. J. Chem.*, 2003, **56**, 1107–1111.
- 461 R. Engelke, W. L. Earl and C. M. M. Rohlfing, *J. Chem. Phys.*, 1985, **84**, 142–146.
- 462 R. J. Abraham, B. Bardsley, M. Mobli and R. J. Smith, *Magn. Reson. Chem.*, 2005, **43**, 3–15.
- 463 H. Fang and M. Oestreich, *Angew. Chemie - Int. Ed.*, 2020, **59**, 11394–11398.
- 464 E. Fanghänel, Y. Keita, R. Radeglia and W. Schmidt, *J. für Prakt. Chemie*, 1985, **327**, 837–846.
- 465 J. C. A. Flanagan, L. M. Dornan, M. G. McLaughlin, N. G. McCreanor, M. J. Cook and M. J. Muldoon, *Green Chem.*, 2012, **14**, 1281–1283.
- 466 A. Fukuoka, T. Nagano, S. Furuta, M. Yoshizawa, M. Hirano and S. Komiya, *Bull. Chem. Soc. Jpn.*, 1998, **71**, 1409–1415.
- 467 T. Fukuyama, Y. Fujita, H. Miyoshi, I. Ryu, S. C. Kao and Y. K. Wu, *Chem. Commun.*, 2018, **54**, 5582–5585.

- 468 T. Fukuyama, Y. Hamada and I. Ryu, *Synth.*, 2021, **53**, 3404–3408.
- 469 G. F. Gauze, E. A. Basso, M. G. Campos, R. Rittner and R. J. Abraham, *J. Phys. Org. Chem.*, 2006, **19**, 376–383.
- 470 V. Gold and T. Mah, *J. Chem. Soc. Perkin Trans. 2*, 1981, 812–816.
- 471 A. Gonzalez-De-Castro, C. M. Robertson and J. Xiao, *J. Am. Chem. Soc.*, 2014, **136**, 8350–8360.
- 472 A. Gopi Krishna Reddy and G. Satyanarayana, *J. Org. Chem.*, 2016, **81**, 12212–12222.
- 473 R. J. Abraham, M. Canton and L. Griffiths, *Magn. Reson. Chem.*, 2001, **39**, 421–431.
- 474 H. Ito and T. Ito, *Chem. Lett.*, 1997, **62**, 1251–1254.
- 475 V. Goyal, J. Gahtori, A. Narani, P. Gupta, A. Bordoloi and K. Natte, *J. Org. Chem.*, 2019, **84**, 15389–15398.
- 476 G. Grossman, G. Komber, G. Kroshvits, P. A. Kirpichnikov, N. A. Mukmeneva and D. G. Pobedimskii, *Bull. Acad. Sci. USSR Div. Chem. Sci.*, 1989, **38**, 26–32.
- 477 T. Guo, Y. Gao, Z. Li, J. Liu and K. Guo, *Synlett*, 2019, **30**, 329–332.
- 478 W. Guo, L.-Q. Lu, Y. Wang, Y.-N. Wang, J.-R. Chen and W.-J. Xiao, *Angew. Chemie*, 2015, **127**, 2293–2297.
- 479 X. Guo, Z. Peng, S. Jiang and J. Shen, *Synth. Commun.*, 2011, **41**, 2044–2052.
- 480 Z. Guo, T. Pang, L. Yan, X. Wei, J. Chao and C. Xi, *Green Chem.*, 2021, **23**, 7534–7538.
- 481 S. S. R. Gupta and M. Lakshmi Kantam, *New J. Chem.*, 2021, **45**, 16179–16186.
- 482 S. Gupta, P. Sureshbabu, A. K. Singh, S. Sabiah and J. Kandasamy, *Tetrahedron Lett.*, 2017, **58**, 909–913.
- 483 M. C. Haibach, B. M. Stoltz and R. H. Grubbs, *Angew. Chemie*, 2017, **129**, 15319–15322.
- 484 R. J. Abraham, M. Canton, M. Reid and L. Griffiths, *J. Chem. Soc. Perkin Trans. 2*, 2000, 803–812.
- 485 M. Hayashi, M. Shibuya and Y. Iwabuchi, *J. Org. Chem.*, 2012, **77**, 3005–3009.
- 486 D. Heijnen, F. Tosi, C. Vila, M. C. A. Stuart, P. H. Elsinga, W. Szymanski and B. L. Feringa, *Angew. Chemie - Int. Ed.*, 2017, **56**, 3354–3359.
- 487 P. H. Helferty and P. Yates, *Org. Magn. Reson.*, 1983, **21**, 352–356.
- 488 B. Hong, K. C. C. Aganda and A. Lee, *Org. Lett.*, 2020, **22**, 4395–4399.
- 489 R. Hosseinzadeh, E. Narimani and M. Mavvaji, *Org. Prep. Proced. Int.*, 2021, **53**, 461–471.
- 490 Y. A. Ibrahim, N. A. Al-Awadi and K. Kaul, *Tetrahedron*, 2001, **57**, 7377–7381.
- 491 M. Iinuma, K. Moriyama and H. Togo, *European J. Org. Chem.*, 2014, 772–780.
- 492 B. Ilkgul, D. Gunes, O. Sirkecioglu and N. Bicak, *Tetrahedron Lett.*, 2010, **51**, 5313–5315.
- 493 S. Imai and H. Togo, *Tetrahedron*, 2016, **72**, 6948–6954.
- 494 H. Ito and T. Ito, *Chem. Lett.*, 1985, **14**, 1251–1254.

- 495 R. J. Abraham, D. J. Chadwick, P. E. Smith and F. Sancassan, *J. Chem. Soc. Perkin Trans. 2*, 1989, 1377–1384.
- 496 P. W. Jardon, E. H. Vickery, L. F. Pahler, N. Pourahmady, G. J. Mains and E. J. Eisenbraun, *J. Org. Chem.*, 1984, **49**, 2130–2135.
- 497 H. M. Jhong, Y. H. Liu, S. M. Peng and S. T. Liu, *Eur. J. Inorg. Chem.*, 2016, 5449–5455.
- 498 H. Jiang, L. Lykke, S. U. Pedersen, W. J. Xiao and K. A. Jyørgensen, *Chem. Commun.*, 2012, **48**, 7203–7205.
- 499 L. Jing, J. Wei, L. Zhou, Z. Huang, Z. Li and X. Zhou, *Chem. Commun.*, 2010, **46**, 4767–4769.
- 500 A. Joshi, R. Kumar, R. Semwal, D. Rawat and S. Adimurthy, *Green Chem.*, 2019, **21**, 962–967.
- 501 M. E. Jung and K. R. Buszek, *J. Am. Chem. Soc.*, 1988, **110**, 3965–3969.
- 502 R. G. Kalkhambkar, S. D. Bunge and K. K. Laali, *Tetrahedron Lett.*, 2011, **52**, 5184–5187.
- 503 V. Kanchupalli, D. Joseph and S. Katukojvala, *Org. Lett.*, 2015, **17**, 5878–5881.
- 504 V. P. Kashparova, V. A. Klushin, I. Y. Zhukova, I. S. Kashparov, D. V. Chernysheva, I. B. Il'chibaeva, N. V. Smirnova, E. S. Kagan and V. M. Chernyshev, *Tetrahedron Lett.*, 2017, **58**, 3517–3521.
- 505 N. Khan, A. C. Abhyankar, T. Nandi and N. Eswara Prasad, *J. Nanosci. Nanotechnol.*, 2019, **19**, 7982–7992.
- 506 C. P. Butts, C. R. Jones, E. C. Towers, J. L. Flynn, L. Appleby and N. J. Barron, *Org. Biomol. Chem.*, 2011, **9**, 177–184.
- 507 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley-VCH, New York, 1991.
- 508 A. Kumar, G. Wagner, R. R. Ernst, K and K. Wüthrich, *J. Am. Chem. Soc.*, 1981, **103**, 3654–3658.
- 509 A. Kalk and H. J. C. Berendsen, *J. Magn. Reson.*, 1976, **24**, 343–366.
- 510 R. Wagner and S. Berger, *J. Magn. Reson. - Ser. A*, 1996, **123**, 119–121.
- 511 G. Bodenhausen and R. R. Ernst, *J. Am. Chem. Soc.*, 1982, **104**, 1304–1309.
- 512 H. Hu and K. Krishnamurthy, *J. Magn. Reson.*, 2006, **182**, 173–177.
- 513 B. Vögeli, T. F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren and R. Riek, *J. Am. Chem. Soc.*, 2009, **131**, 17215–17225.
- 514 J. LAUTZ, H. KESSLER, J. M. BLANEY, R. M. SCHEEK and W. F. van Gunsteren, *Int. J. Pept. Protein Res.*, 1989, **33**, 281–288.
- 515 S. Wang, K. Krummenacher, G. A. Landrum, B. D. Sellers, P. Di Lello, S. J. Robinson, B. Martin, J. K. Holden, J. Y. K. Tom, A. C. Murthy, N. Popovych and S. Riniker, *J. Chem. Inf. Model.*, 2022, **62**, 472–485.
- 516 I. Solomon, *Phys. Rev.*, 1955, **99**, 559–565.
- 517 F. Bloch, *Phys. Rev.*, 1946, **70**, 460–474.
- 518 C. R. Jones, M. D. Greenhalgh, J. R. Bame, T. J. Simpson, R. J. Cox, J. W. Marshall and C. P. Butts, *Chem. Commun.*, 2016, **52**, 2920–2923.

- 519 A. Kolmer, L. J. Edwards, I. Kuprov and C. M. Thiele, *J. Magn. Reson.*, 2015, **261**, 101–109.
- 520 J. Tropp, *J. Chem. Phys.*, 1979, **72**, 6035–6043.
- 521 R. Rowan, J. A. McCammon and B. D. Sykes, *J. Am. Chem. Soc.*, 1974, **96**, 4773–4780.
- 522 D. D. Giannini, I. M. Armitage, H. Pearson, D. M. Grant and J. D. Roberts, *J. Am. Chem. Soc.*, 1975, **97**, 3416–3419.
- 523 B. Vögeli, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2014, **78**, 1–46.
- 524 H. Akaike, *IEEE Trans. Automat. Contr.*, 1974, **19**, 716–723.
- 525 E. J. Wagenmakers and S. Farrell, *Psychon. Bull. Rev.*, 2004, **11**, 192–196.
- 526 K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference*, Springer, New York, 2002.
- 527 K. P. Burnham and D. R. Anderson, *Sociol. Methods Res.*, 2004, **33**, 261–304.
- 528 S. Simova, H. Sengstschmid and R. Freeman, *J. Magn. Reson.*, 1997, **124**, 104–121.
- 529 A. Cotte, M. Foroozandeh and D. Jeannerat, *Chim. Int. J. Chem.*, 2012, **66**, 764–769.
- 530 J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546–4553.
- 531 B. A. Johnson, *Methods Mol. Biol.*, 2004, **278**, 313–352.
- 532 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.
- 533 J. C. Kromann, Calculate Root-mean-square deviation of Two Molecules Using Rotation, <http://github.com/charnley/rmsd>.
- 534 F. Gao and L. Han, *Comput. Optim. Appl.*, 2012, **51**, 259–277.
- 535 G. Lauro and G. Bifulco, *European J. Org. Chem.*, 2020, 3929–3941.
- 536 G. Lauro, P. Das, R. Riccio, D. S. Reddy and G. Bifulco, *J. Org. Chem.*, 2020, **85**, 3297–3306.
- 537 V. Schmidts, M. Fredersdorf, T. Lübken, A. Porzel, N. Arnold, L. Wessjohann and C. M. Thiele, *J. Nat. Prod.*, 2013, **76**, 839–844.
- 538 L. Bösel, R. Dötzer, S. Steiner, M. Stritzinger, S. Salzmann and S. Riniker, *Anal. Chem.*, 2020, **92**, 9124–9131.
- 539 F. Pultar, M. E. Hansen, S. Wolfrum, L. Bösel, R. Fróis-Martins, S. Bloch, A. G. Kravina, D. Pehlivanoglu, C. Schäffer, S. LeibundGut-Landmann, S. Riniker and E. M. Carreira, *J. Am. Chem. Soc.*, 2021, **143**, 10389–10402.
- 540 L. Bösel, R. Aerts, W. Herrebout and S. Riniker, *submitted*.
- 541 D. Porezag and M. R. Pederson, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1996, **54**, 7830–7836.
- 542 L. B. Krivdin, *Magn. Reson. Chem.*, 2019, **57**, 897–914.
- 543 L. B. Casabianca, *Magn. Reson. Chem.*, 2020, **58**, 611–624.
- 544 A. Rahman and M. I. Choudhary, *Solving problems with NMR spectroscopy*, Academic Press, 1996.
- 545 C. Nilewski, R. W. Geisser, M. O. Ebert and E. M. Carreira, *J. Am. Chem. Soc.*, 2009, **131**,

- 15866–15876.
- 546 M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870–2871.
- 547 P. D’Antuono, E. Botek, B. Champagne, M. Spassova and P. Denkova, *J. Chem. Phys.*, 2006, **125**, 24111.
- 548 C. Gropp, S. Fischer, T. Husch, N. Trapp, E. M. Carreira and F. Diederich, *J. Am. Chem. Soc.*, 2020, **142**, 4749–4755.
- 549 Chemical Book, <https://www.chemicalbook.com>, (accessed 14 December 2021).
- 550 R. E. Rosenberg and W. J. Kelly, *J. Phys. Org. Chem.*, 2015, **28**, 47–56.
- 551 R. I. Khusnutdinov, N. A. Shchadneva, R. Y. Burangulova, Z. S. Muslimov and U. M. Dzhemilev, *Russ. J. Org. Chem.*, 2006, **42**, 1615–1621.
- 552 A. Artaryan, A. Mardyukov, K. Kulbitski, I. Avigdori, G. A. Nisnevich, P. R. Schreiner and M. Gandelman, *J. Org. Chem.*, 2017, **82**, 7093–7100.
- 553 I. Mathieu-Pelta and S. A. Evans, *J. Org. Chem.*, 1992, **57**, 3409–3413.
- 554 J. J. Méndez, J. Eras, M. Balcells and R. Canela, *Synth. Commun.*, 2006, **36**, 1167–1175.
- 555 M. K. Kesharwani, B. Brauer and J. M. L. Martin, *J. Phys. Chem. A*, 2015, **119**, 1701–1714.

Curriculum Vitae

PERSONAL DATA

Name	Thomas Stadelmann
Date of Birth	June 28, 1992
Place of Birth	Sursee (LU), Switzerland
Citizen of	Mauensee (LU), Switzerland

EDUCATION

2018 – 2022	PhD with Prof. Sereina Riniker, Laboratory of Physical Chemistry and Dr. Marc-Olivier Ebert, Laboratory of Organic Chemistry at ETH Zürich, Switzerland
2015 – 2017	Master of Science in Interdisciplinary Sciences at ETH Zürich, Switzerland
2012 – 2015	Bachelor of Science in Interdisciplinary Sciences at ETH Zürich, Switzerland
2005 – 2011	Kantonsschule Sursee (LU), Switzerland

EMPLOYMENT

2017 – 2018	Research assistant (60 %) in the group of Prof. Sereina Riniker, Laboratory of Physical Chemistry with supervision of Dr. Marc-Olivier Ebert, Laboratory of Organic Chemistry at ETH Zürich, Switzerland
-------------	--