

---

# From Paramagnetic Proteins to Field Alignment of Small Molecules

Tools for Structure Determination via NMR

---

Dissertation  
for the award of the degree  
*Doctor rerum naturalium*  
of the Georg-August-Universität Göttingen

within the doctoral program  
“Physics of Biological and Complex Systems”  
of the Georg-August University School of Science (GAUSS)

submitted by  
Niels Ulrich Karschin  
from Itzehoe

Göttingen 2021



# **From Paramagnetic Proteins to Field Alignment of Small Molecules**

**Tools for Structure Determination via NMR**

Niels Ulrich Karschin

---

## **Thesis Committee**

Prof. Dr. Christian Griesinger  
NMR-based Structural Biology  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Marina Bennati  
Electron-Spin Resonance Spectroscopy  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Bert de Groot  
Computational Biomolecular Dynamics  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

---

## **Members of the Examination Board**

1st Referee:

Prof. Dr. Christian Griesinger  
NMR-based Structural Biology  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

2nd Referee:

Prof. Dr. Marina Bennati  
Electron-Spin Resonance Spectroscopy  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

---

## **Further Members of the Examination Board**

Prof. Dr. Bert de Groot  
Computational Biomolecular Dynamics  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Martin Suhm  
Vibrational Spectroscopy and Intermolecular Interactions  
Institute for Physical Chemistry, Göttingen, Germany

Prof. Dr. Ricardo Mata  
Computational Chemistry and Biochemistry  
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Dr. Stefan Glögler  
NMR Signal Enhancement  
Center for Biostructural Imaging of Neurodegeneration, Göttingen, Germany

---

**Date of oral examination:** 26 January 2022

*“Commonplace as such experiments have become in our laboratories, I have not yet lost a feeling of wonder, and of delight, that this delicate motion should reside in all the ordinary things around us, revealing itself only to him who looks for it. I remember, in the winter of our first experiments, just seven years ago, looking on snow with new eyes. There the snow lay around my doorstep – great heaps of protons quietly precessing in the earth’s magnetic field. To see the world for a moment as something rich and strange is the private reward of many a discovery.”*

– Edward M. Purcell, Nobel Lecture, 11 December 1952.



## Acknowledgement

First and foremost I would like to thank Christian Griesinger that he has accepted me into his lab and that he has shown me a general direction to walk along. I am even more thankful that he has given me the freedom to explore into topics I myself found promising and interesting, providing guidance wherever necessary. Without that kind of attitude, this dissertation would be very different from what it has become.

Many thanks to Marina Bennati and Bert de Groot for being in my thesis advisory committee. The meetings with you were always constructive and helpful, and in a pleasant atmosphere.

Thanks to Ricardo Mata, Stefan Glöggler, and Martin Suhm for being on my examination board. Ricardo and Martin, I additionally want to thank you for being excellent teachers during my earlier studies.

Pablo Trigo Mouriño has shown me how to operate an NMR spectrometer and how to deal with the perpetually unpredictable biomolecules. Without you I certainly would have had a much rougher start. I am truly grateful for that.

Many thanks to Stefan Becker, Karin Giller, Melanie Wegstroth, and Kerstin Overkamp for providing proteins and peptides, and for assistance with dealing with them. You have done exceptional work in an area where I am rather untalented.

Thanks to Petra Breiner for your support with everything regarding administration and paperwork, and for your ever positive attitude. Many thanks to Dirk Bockelmann for your constant support with any computer problems, and for the chats in the hallway.

Without the help of the institute's workshops many of my ideas could not have transformed into real-life objects. Thanks go to Matthias Kleinhans, Matthias Kulp, Frank Meyer, Christian Klaba, and Julius Bergmann. I want to thank Florian Jordan in particular, who has spent uncountable hours on my projects.

Thanks to the innumerable, unnamed authors of Wikipedia articles, python packages, Stack Overflow posts, etc. The amount of knowledge and software that is freely available owing to your commitment and effort is really astonishing.

I have worked with Phani Vemulapalli, Juan Carlos Fuentes, and Klaus Wolkenstein on small molecules. Thank you for many constructive and enriching discussions.

There are many more colleagues and members of the department that have made this a very stimulating and pleasant working environment, and I want to thank you all for that.

I have shared an office with Supriya, Sebastian, and Pablo for the best part of my Ph.D., and I have many fond memories with you guys. Leif was next door, but the same applies for him. Thank you for your friendship.

Venus joined the department only about a year ago, but he has been a true friend in a time that was not always easy for me. Thank you for that.

Many old friends from my studies have long left Göttingen, but we have stayed in contact and hopefully will for many years to come. Tjark, Lorenz, Freed, Peter, Lukas, Nils, Bene, Niklas, and Oskar, I am grateful for having you as friends.

My comrades from Schleswig (and Holstein) have given me the opportunity to get out of the science bubble on a regular basis, and we have gone through various adventures together. Thanks for always keeping me down-to-earth.

Adri, you have supported me throughout my Ph.D. time, and I could always count on you. I am truly thankful for that.

My mother, my father, and my two brothers have been there for me all my life, and I have to thank them for everything I have accomplished. You made me who I am today.

# Contents

<b>Frontmatter</b>	<b>i</b>
Acknowledgement . . . . .	v
Abbreviations . . . . .	x
List of Symbols . . . . .	xiii
List of Figures . . . . .	xviii
List of Tables . . . . .	xix
<b>1 Introduction</b>	<b>1</b>
1.1 The Quantum Revolution and the Emergence of Spin . . . . .	1
1.2 Development of Nuclear Magnetic Resonance Spectroscopy . . . . .	3
1.3 Scalar NMR Parameters for Small-Scale Structure . . . . .	6
1.4 Tensorial NMR Parameters for Large-Scale Structure . . . . .	8
1.5 Scope, Outline, and Summary . . . . .	11
<b>2 Theoretical Background</b>	<b>15</b>
2.1 Fundamentals of NMR . . . . .	15
2.1.1 Nuclear Spin and Matrix Representations . . . . .	15
2.1.2 The Secular Approximation . . . . .	18
2.2 Alignment and Mechanisms . . . . .	19
2.2.1 Description of Alignment as a Tensor . . . . .	19
2.2.2 Magnetic Field Alignment . . . . .	20
2.2.3 Electric Field Alignment . . . . .	22
2.3 Anisotropic Parameters . . . . .	25
2.3.1 Residual Dipolar Couplings . . . . .	26
2.3.2 Residual Chemical Shift Anisotropy . . . . .	28
2.3.3 Residual Quadrupolar Couplings . . . . .	29
2.4 Paramagnetism . . . . .	30
2.4.1 Magnetic Susceptibility in Paramagnetic Centers . . . . .	31
2.4.2 Pseudocontact Shifts . . . . .	32
2.5 Molecular Mechanics . . . . .	34
2.6 Density Functional Theory . . . . .	37
2.6.1 Hartree-Fock Theory . . . . .	38
2.6.2 Kohn-Sham Theory . . . . .	45
2.6.3 Basis Sets . . . . .	47
2.7 Electrostatics . . . . .	49
2.7.1 Theory of Electric Fields in Materials . . . . .	49
2.7.2 Finite Element Analysis . . . . .	51

<b>3</b>	<b>Interdomain Dynamics of the Complex Calmodulin/Munc13-1</b>	<b>55</b>
3.1	Introduction . . . . .	55
3.2	Complex Preparation and NMR Spectroscopy . . . . .	57
3.2.1	Sample Preparation . . . . .	57
3.2.2	Assignment . . . . .	58
3.2.3	Acquisition of Paramagnetic Data . . . . .	60
3.3	Data Analysis and Modeling . . . . .	63
3.3.1	Fitting of N-Terminal Data . . . . .	63
3.3.2	Fitting of C-terminal RDCs . . . . .	66
3.3.3	Sampling the Conformational Space of CaM/Munc13-1 . . . . .	69
3.3.4	Ensemble Sampling . . . . .	72
3.4	Conclusion and Outlook . . . . .	78
3.5	Materials and Methods . . . . .	82
<b>4</b>	<b>Magnetic Alignment of Natural Products</b>	<b>85</b>
4.1	Introduction . . . . .	85
4.2	Characterization of Gymnochrome G . . . . .	86
4.2.1	Constitution . . . . .	86
4.2.2	Relative Configuration by <i>J</i> -Couplings and NOE . . . . .	87
4.2.3	Relative Configuration by DP4+ . . . . .	90
4.2.4	Absolute Configuration . . . . .	91
4.3	Determination of Anisotropic Parameters . . . . .	92
4.3.1	NMR Experiments . . . . .	92
4.3.2	Chemical Shift Correction . . . . .	93
4.3.3	Field Dependence Fit and Weighting Factors . . . . .	94
4.4	Data Evaluation and Discrimination of Stereomers . . . . .	97
4.4.1	Model Generation . . . . .	97
4.4.2	Prediction of Anisotropic Data . . . . .	97
4.4.3	Comparison of Experimental and Calculated Data . . . . .	99
4.4.4	Results . . . . .	99
4.5	Addition of 1200 MHz . . . . .	101
4.6	Conclusion and Outlook . . . . .	103
4.7	Material and Methods . . . . .	103
4.7.1	Characterization of Gymnochrome G . . . . .	103
4.7.2	Anisotropic NMR . . . . .	105
<b>5</b>	<b>Design of an Electric Field Cell</b>	<b>109</b>
5.1	Introduction . . . . .	109
5.1.1	General Considerations . . . . .	110
5.2	The 1D Field Cell: Circular Tubes and Linear Electrodes . . . . .	111
5.2.1	Concept and Design . . . . .	111
5.2.2	Technical Realization . . . . .	112
5.2.3	Testing . . . . .	114
5.2.4	Conclusion and Outlook . . . . .	118

5.3	The 2D Field Cell: Shaped Tubes and Planar Electrodes . . . . .	119
5.3.1	Concept and Design . . . . .	119
5.3.2	Technical Realization . . . . .	120
5.3.3	Testing . . . . .	122
5.3.4	Conclusion and Outlook . . . . .	124
5.4	Materials and Methods . . . . .	125
<b>6</b>	<b>Extension and Improvement of the Methanol-<math>d_4</math> NMR Thermometer Calibration</b>	<b>129</b>
6.1	Introduction . . . . .	129
6.2	Results and Discussion . . . . .	130
6.3	Material and Methods . . . . .	134
	<b>References</b>	<b>136</b>
<b>A</b>	<b>Appendix</b>	<b>155</b>
A.1	Data Collection . . . . .	155
A.2	Calmodulin/Munc13-1 – Additional Data . . . . .	155
A.3	Spectral Data of Gymnochrome G . . . . .	164
	<b>Affidavit</b>	<b>175</b>

## Abbreviations and Acronyms

---

<b>AC</b>	Alternating Current
<b>AMBER</b>	Assisted Model Building with Energy Refinement
<b>APT</b>	Attached Proton Test
<b>ATR</b>	Attenuated Total Reflection
<b>BIC</b>	Bayesian Information Criterion
<b>Bis-Tris</b>	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
<b>CAD</b>	Computer-Assisted Design
<b>CaM</b>	Calmodulin
<b>CCR</b>	Cross-Correlated Relaxation
<b>CLIP-HSQC</b>	CLean-InPhase-HSQC
<b>CNC</b>	Computerized Numerical Control
<b>COSY</b>	COrrrelation SpectroscopY
<b>CS</b>	Compressed Sensing
<b>DC</b>	Direct Current
<b>DFT</b>	Density Funtional Theory
<b>DMSO</b>	DiMethyl SulfOxide
<b>ECD</b>	Electronic Circular Dichroism
<b>EDTA</b>	EthyleneDiamineTetraAcetate
<b>ESI</b>	ElectroSpray Ionization
<b>FEM</b>	Finite Element Method
<b>FID</b>	Free Induction Decay
<b>GIAO</b>	Gauge-Independent Atomic Orbitals
<b>GTO</b>	Gaussian-Type Orbital
<b>HMBC</b>	Heteronuclear Multiple-Bond Correlation
<b>HPLC</b>	High Performance Liquid Chromatography
<b>HR MS</b>	High Resolution Mass Spectrometry
<b>HSQC</b>	Heteronuclear Single-Quantum Correlation
<b>INADEQUATE</b>	Incredible Natural Abundance Double QUAntum Transfer Experiment
<b>INEPT</b>	Insensitive Nuclei Enhancement by Polarization Transfer
<b>IPAP</b>	InPhase-AntiPhase
<b>IR</b>	InfraRed

---

---

<b>IUPAC</b>	International Union of Pure and Applied Chemistry
<b>IRLS</b>	Iterative Reweighted Least Squares
<b>LCAO</b>	Linear Combination of Atomic Orbitals
<b>LSDA</b>	Local Spin Density Approximation
<b>MD</b>	Molecular Dynamics
<b>MM</b>	Molecular Modeling
<b>MMFF</b>	Merck Molecular Force Field
<b>MO</b>	Molecular Orbital
<b>MWCO</b>	Molecular Weight Cut-Off
<b>NMR</b>	Nuclear Magnetic Resonance
<b>NNLS</b>	Non-Negative Least Squares
<b>NOE</b>	Nuclear Overhauser Effect
<b>NUS</b>	Non-Uniform Sampling
<b>OPLS3</b>	Optimized Potentials for Liquid Simulations 3
<b>PCB</b>	Printed Circuit Board
<b>PCS</b>	PseudoContact Shift
<b>PDB</b>	Protein Data Bank
<b>PES</b>	PolyEtherSulfone
<b>PVD</b>	Physical Vapor Deposition
<b>PRE</b>	Paramagnetic Relaxation Enhancement
<b>RCSA</b>	Residual Chemical Shift Anisotropy
<b>RDC</b>	Residual Dipolar Coupling
<b>RF</b>	Radio Frequency
<b>RMS</b>	Root Mean Square
<b>RMSD</b>	Root Mean Square Deviation
<b>ROE</b>	Rotating frame Overhauser Effect
<b>RQC</b>	Residual Quadrupolar Couplings
<b>RSS</b>	Root Sum of Squares
<b>SCF</b>	Self-Consistent Field
<b>STO</b>	Slater-Type Orbital
<b>TMS</b>	TetraMethylSilane
<b>TOCSY</b>	TOTAL Correlation SPectroscopy
<b>TOF</b>	Time Of Flight
<b>UV/Vis</b>	UltraViolet/Visible

---

*Contents*

---

<b>VT</b>	Variable Temperature
<b>WATERGATE</b>	WATER suppression by GrAdient-Taylorred Excitation
<b>ZPC</b>	Zero-Point Corrected

---

## List of Symbols

This work follows the convention to print vector quantities in bold ( $\mathbf{r}$ ,  $\mathbf{B}_0$ ), unit vectors with a hat ( $\hat{\mathbf{r}}$ ,  $\hat{\mathbf{b}}$ ), and tensorial quantities in bold upright ( $\boldsymbol{\chi}$ ,  $\mathbf{A}$ ). It is unavoidable that sometimes different quantities share the same symbol, but the difference will be obvious from the context.

---

$\mathbf{1}$	unit matrix
$a_i$	constant parameter (triangular shape function)
$A$	area
$A$	magnetic deviation, temperature dependence
$\mathbf{A}$	alignment tensor
$\hat{A}$ , $\hat{B}$	generic operators
$b_i$	$x$ -slope (triangular shape function)
$\mathbf{B}$	magnetic field flux density
$\hat{\mathbf{b}}$	magnetic field unit vector
$c$	MO coefficients
$c$	misreferencing coordinate
$c_i$	$y$ -slope (triangular shape function)
$C$	capacitance
$d_{12}$	dipolar coupling prefactor
$D$	dipolar coupling constant
$\mathbf{D}$	electric flux density
$\mathbf{D}$	density matrix (Hartree-Fock)
$e$	elementary charge
$E$	energy
$\mathbf{E}$	electric field strength
$F$	Faraday constant
$\hat{F}$	Fock operator
$g_e$	Landé-factor
$\mathbf{G}$	two-electron integral tensor
$\hat{h}$	one-electron Hamiltonian
$\hbar$	reduced Planck constant
$\hat{\mathcal{H}}$	Hamiltonian
$\mathbf{H}$	magnetic field intensity

---

---

<b>H</b>	parameter matrix (FEM)
$I$	spin angular momentum quantum number
$\hat{\mathbf{I}}, \hat{I}_{x,y,z}$	spin operators
$i$	imaginary unit
$\mathbf{j}$	current density
$J$	scalar coupling constant
$J_{ij}$	coulomb integral
$\hat{J}_i$	coulomb operator
$k$	force constant
$k$	magnetic deviation, field dependence
$k_B$	Boltzmann constant
$K_{ij}$	exchange integral
$\hat{K}_i$	exchange operator
$l$	angular momentum quantum number
$\mathbf{l}$	angular momentum
$L$	Lagrangian
$m$	magnetic quantum number
$m/z$	mass/charge ratio
<b>M</b>	transformation matrix
$n$	natural number
$n$	refractive index
$N_i$	shape function
$\mathbf{p}$	electric dipole moment
<b>P</b>	probability tensor
$\hat{P}_\tau$	permutation operator
$q$	atomic charge
$Q$	quality factor
$Q$	nuclear quadrupole moment
<b>Q</b>	chemical shift transformation matrix
$\mathbf{r}$	distance vector
<b>R</b>	rotation matrix
$s$	estimated scatter
$S$	electron spin quantum number
<b>S</b>	overlap matrix

---

---

$t_{1,2}$	evolution time (pulse sequence)
$T$	temperature
$T$	total coupling constant
$\hat{T}$	kinetic energy operator
$U$	voltage
$V$	torsion potential Fourier coefficient
$V$	volume
$\hat{V}$	potential energy operator
$\mathbf{V}$	electric field gradient tensor
$w$	electric field energy density
$w$	weight
$Y_l^m$	spherical harmonic
$Z$	partition function
$Z$	atomic number
$\alpha$	angle between $\mathbf{B}_0$ and $\mathbf{E}$
$\alpha$	stretch anharmonicity constant
$ \alpha\rangle,  \beta\rangle$	spin functions
$\gamma$	gyromagnetic ratio
$\delta$	chemical shift
$\delta_{ij}$	Kronecker delta
$\Delta$	element area
$\epsilon_0$	vacuum permittivity
$\epsilon$	permittivity
$\epsilon$	MO energies
$\epsilon$	molar extinction coefficient
$\epsilon$	Lennard-Jones well depth
$\epsilon$	electron energy density
$\delta$	chemical shift tensor
$\theta$	polar angle
$\theta_{ijk}$	bond angle
$\lambda$	wave length
$\lambda$	Lagrange multiplier
$\Lambda$	Lagrange multiplier matrix
$\mu_0$	vacuum permeability

---

---

$\mu_B$	Bohr magneton
$\boldsymbol{\mu}$	magnetic moment
$\nu$	frequency
$\xi$	degradation coordinate
$\Pi$	product of spin orbitals
$\rho$	probability density
$\rho$	charge density
$\sigma$	standard deviation
$\tau$	permutation
$\phi$	azimuthal angle
$\phi$	molecular spin-orbital
$\varphi$	electric potential
$\chi$	angle of $\mathbf{E}$ around $\mathbf{B}_0$
$\chi$	out-of-plane angle
$\chi$	atomic orbital/basis function
$\boldsymbol{\chi}$	molecular susceptibility tensor
$\Delta\boldsymbol{\chi}$	susceptibility anisotropy tensor
$\Psi$	many-electron wave function
$\omega_0$	Larmor frequency
$\omega_Q$	first-order quadrupolar coupling
$\omega$	torsion angle
$\Omega$	solid angle
$\nabla$	gradient operator
$\nabla\cdot$	divergence operator
$\nabla\times$	curl operator

---

# List of Figures

1.1	PCS isosurfaces for Dy in an ensemble of CaM/Munc13-1. . . . .	9
2.1	Illustration of the angles describing the orientations of $\hat{\mathbf{b}}$ and $\hat{\mathbf{e}}$ . . . . .	23
2.2	Iterative process in Hartree-Fock theory. . . . .	44
2.3	$\mathbf{E}$ - and $\mathbf{D}$ -field profiles in different simple capacitor arrangements. . . . .	50
2.4	Shape functions and their parameters. . . . .	52
3.1	EF-hand $\text{Ca}^{2+}$ binding pocket and antiparallel $\beta$ -sheet in calmodulin. . . . .	55
3.2	(a) First calmodulin X-ray structure and (b) CaM/Munc13-1 NMR ensemble. . . . .	56
3.3	HNCA walk for $\text{Ca}_3\text{Yb-CaM/Munc13-1}$ . . . . .	59
3.4	Assigned resonances for the metal complexes $\text{Ca}_3\text{Ln-CaM/Munc13-1}$ . . . . .	59
3.5	Relative sizes of RDCs acquired for CaM/Munc13-1. . . . .	61
3.6	Example of an NUS sampling grid for an HNCO. . . . .	61
3.7	$\Delta\chi$ -tensor elements from the N-terminal fit. . . . .	67
3.8	Correlation plot of PCS/RDC for the N-terminal fit. . . . .	68
3.9	Plot of the $\Delta\chi$ eigenvalues of the N- and C-terminal domain. . . . .	69
3.10	RMSD as a function of ensemble size. . . . .	75
3.11	RMSD convergence for the final samplings. . . . .	76
3.12	Evolution of RMSD during iterative cross-validation. . . . .	77
3.13	Cartoon representation of the best ensembles for all $n_{\text{ens}}^{\text{init}}$ . . . . .	79
3.14	Cartoon representation of the best ensembles for all $n_{\text{ens}}^{\text{init}}$ (only members with $p > 10\%$ ). . . . .	80
3.15	Best ensemble for $n_{\text{ens}}^{\text{init}} = 20$ represented as rotation vectors in space. . . . .	81
4.1	Structure of gymnochrome G. . . . .	86
4.2	Molecular ion peak of gymnochrome G. . . . .	87
4.3	ROE contacts and Newman projections of gymnochrome G. . . . .	88
4.4	Calculated and experimental ECD spectrum of gymnochrome G. . . . .	91
4.5	Structure of strychnine. . . . .	92
4.6	RCSA correlation before and after correction. . . . .	93
4.7	Field series and weights for strychnine-RDCs. . . . .	95
4.8	Correlation of experimental and predicted anisotropic data. . . . .	100
4.9	$Q$ -factors and confidences for the different diastereomers of <b>1</b> and <b>2</b> . . . . .	100
4.10	Confidences for <b>2</b> using 1200 MHz data. . . . .	102
5.1	Electric field pattern for 1D field cell (vacuum). . . . .	113
5.2	Head assembly and PCB for the 1D field cell. . . . .	114

List of Figures

5.3	NMR spectra of strychnine in the 1D field cell. . . . .	115
5.4	FEM simulation of the 1D design. . . . .	116
5.5	Electric field histogram of the 1D cell. . . . .	117
5.6	$^2\text{D}$ spectra of MeCN at various voltages in the 1D cell. . . . .	117
5.7	Structure of disperse red 13. . . . .	117
5.8	Head assembly and coating stencil for the 2D field cell. . . . .	120
5.9	FEM simulation of the 2D design. . . . .	122
5.10	$^2\text{D}$ spectra of MeCN and nitrobenzene at various voltages in the 2D cell. . . . .	123
5.11	Electric field histogram of the 2D cell. . . . .	123
6.1	Temperature gradients as determined via NMR. . . . .	130
6.2	Magnetic field dependence of the Pt100 sensor. . . . .	132
6.3	Fit and residuals of chemical shift difference vs. corrected temperature. . . . .	133
6.4	Uncertainty of the calibration as a function of temperature. . . . .	135
A.1	$^1\text{H}$ and HSQC spectra of apo-CaM and Ca <sub>4</sub> -CaM. . . . .	156
A.2	Pulse sequence of the HNCA/HNCO/HNCO{no C $_{\alpha}$ }. . . . .	157
A.3	Pulse sequence of the HNCA{no H}. . . . .	157
A.4	Pulse sequence of the HNCO{no H}. . . . .	158
A.5	Cartoon representations of the second best ensembles by BIC. . . . .	159
A.6	Cartoon representations of the third best ensembles by BIC. . . . .	160
A.7	Cartoon representations of the fourth best ensembles by BIC. . . . .	161
A.8	Cartoon representations of the fifth best ensembles by BIC. . . . .	162
A.9	Cartoon representations of the sixth best ensembles by BIC. . . . .	163
A.10	HR mass spectrum of gymnochrome G. . . . .	165
A.11	UV/Vis and ECD spectra of gymnochrome G. . . . .	166
A.12	$^1\text{H}$ spectrum of gymnochrome G. . . . .	167
A.13	$^{13}\text{C}$ -APT spectrum of gymnochrome G. . . . .	168
A.14	DQF-COSY of gymnochrome G. . . . .	169
A.15	HSQC of gymnochrome G. . . . .	170
A.16	HMBC of gymnochrome G. . . . .	171
A.17	ROESY of gymnochrome G. . . . .	172

# List of Tables

3.1	Number of paramagnetic constraints in CaM/Munc13-1. . . . .	62
3.2	Eigenvalues of $\Delta\chi$ from the static fits to CaM/Munc13-1. . . . .	66
4.1	$J$ -couplings of gymnochrome G. . . . .	88
4.2	Comparison criteria of calculated ECD data of gymnochrome G. . . . .	90
5.1	List of relative permittivities used in FEM simulations. . . . .	126
6.1	Polynomial coefficients for the fit function in Eq. (6.4). . . . .	134
A.1	$Q$ -factors for the static fits of CaM/Munc13-1. . . . .	155
A.2	$\Delta\chi$ -tensor elements from the static fits of CaM/Munc13-1. . . . .	156
A.3	NMR-spectroscopic data of gymnochrome G. . . . .	164
A.4	$Q$ -factors and confidences $p$ for gymnochrome G. . . . .	173
A.5	$Q$ -factors and confidences $p$ for strychnine. . . . .	173



# Chapter 1

## Introduction

### 1.1 The Quantum Revolution and the Emergence of Spin

At the end of the 19th century, the theories of physics were in a quite satisfactory state. Isaac Newton's classical mechanics was well-developed into sophisticated mathematical formulations by Joseph-Louis Lagrange and William Hamilton, and it could be used to solve complex problems and successfully predict their outcome.<sup>[1-3]</sup> For example, the motion and reciprocal interaction of planets within the solar system was well understood, and the mathematician Urbain Le Verrier was able to predict both the existence and location of the then unknown planet Neptune based on orbital perturbations of Uranus.<sup>[4]</sup> Thermodynamics was the established model to deal with heat, work, and temperature. The related kinetic theory of gases was used to predict the thermodynamic behavior of gases, although many physicist regarded the occurring microscopic particles as a purely hypothetical construct. The groundbreaking theory of electromagnetism by James Maxwell was able to unify and correctly describe electrical, magnetic, and optical phenomena.<sup>[5]</sup> It was a common opinion among scientists that all main physical laws had been discovered, and that remaining inadequacies of the theory could be patched with minor modifications.

However, some of these inadequacies proved to be unfixable, and only new and revolutionary theories could account for them. In 1887, Albert Michelson and Edward Morley were unable to find evidence for a medium of electromagnetic radiation, called the luminiferous ether.<sup>[6]</sup> Albert Einstein later found a more general problem with the concept of an ether. In one of his famous thought experiments he envisioned an observer moving at the speed of light, which lead to a number of absurd results when applying the current theories of mechanics and electromagnetism. As a consequence, he went on to almost single-handedly develop the theory of relativity between 1905 and 1915, which today is considered as one of the two branches of modern physics.<sup>[7,8]</sup> It can be described in a most reductionist way as the theory of things that are either very heavy or very fast.

Another phenomenon that eluded the description via classical mechanics was the spectrum of black-body radiation. Theoretical arguments predicted that the power density should diverge for short wavelengths, which was clearly unphysical and got coined as the ultraviolet catastrophe.<sup>[9,10]</sup> Max Planck was able to solve this problem by postulating that energy can only be absorbed and emitted in discrete packets, or

quanta, and formulated what is today known as Planck's law of radiation, which was in agreement with experimental observations of thermal radiation.<sup>[11]</sup> Einstein used this idea of quantization in 1905 to explain the photoelectric effect, the observation made first by Heinrich Hertz that electrons could be emitted from condensed matter upon irradiation with light, but only if the light had a sufficiently short wavelength.<sup>[12,13]</sup> This interpretation by Einstein was incompatible with the description of light as waves, and implied that light had particle-like properties. On the other side, Ernest Rutherford demonstrated in 1911 that the positive charge of matter was concentrated into a very small fraction of the overall volume.<sup>[14]</sup> This meant that the negative electrons had to be orbiting around these nuclei, which again was incompatible with classical electrodynamics. By postulating the quantization of the electron's angular momentum, Niels Bohr was able to construct an atomic model that had the electrons orbiting the nucleus in discrete orbits with discrete energy levels.<sup>[15]</sup> Although we know today that this model fails in a number of finer details, it was a breakthrough as it successfully predicted the previously enigmatic emission lines of hydrogen gas, and the orbital radii qualitatively agreed with the atom sizes determined via the first X-ray diffraction experiments. In his Ph.D. thesis from 1924, Louis de Broglie assigned wave-like properties to particles with non-zero mass, further extending the concept of wave-particle dualism.<sup>[16]</sup> It was clear that classical physics failed on the atomic scale, and a new theory for the things that are either very light or very slow started to emerge.

In the following years, two different descriptions of this new quantum mechanics were formulated. Werner Heisenberg, Max Born, and Pascual Jordan had developed a mathematical formalism based around matrices, while Erwin Schrödinger had built upon the idea of de Broglie and developed quantum mechanics based around wave functions.<sup>[17–19]</sup> It was quickly shown that both approaches were equivalent, and it was possible to choose the description that was more convenient for the problem at hand. However, this was far from being the end of the story. Finer aspects of quantization were still being discovered, and the theory was being refined. In 1922 Otto Stern and Walther Gerlach showed experimentally that the angular momentum was not only quantized in its magnitude, but also in its direction.<sup>[20]</sup> Although they had been looking for quantization of orbital angular momentum, they actually observed the electron's unknown spin angular momentum being quantized in its orientation. As a result, they found only two instead of three discrete states. To explain this, Samuel Goudsmit and George Uhlenbeck proposed that the electron had an inherent angular momentum with quantum number  $1/2$ , although it was already clear at the time that this could not reasonably be associated with the electron literally rotating around its own axis (which would require either unreasonable sizes or rotation speeds).<sup>[21]</sup> The year before, Wolfgang Pauli had already assigned an unspecified, two-valued quantum number to the electron to explain the shell structure of the atom and the periodic system, and based on the idea of a half-integer electron spin, he went on to establish the Pauli matrices in 1927, which remains the valid description of spin phenomena today (Eq. (2.7)).<sup>[22,23]</sup> When Paul Dirac modified the Schrödinger equation to include special relativity in the following year, he found spin emerging naturally, providing a theoretical justification for it.<sup>[24]</sup>

Around the same time it was realized that atomic nuclei and its constituents also possessed spin, albeit the associated magnetic moment was about three orders of magnitude smaller. In 1933 Stern performed a similar experiment to the one from 1922 on hydrogen molecules to determine the magnetic moment of protons, but the necessary inhomogeneous magnetic field was difficult to control, so the accuracy of the measurements was low.<sup>[25-27]</sup> Isidor Rabi modified these experiments by passing atomic or molecular beams through homogeneous magnetic fields and applying an alternating magnetic field on resonance with the Larmor precession frequency of the atomic nuclei.<sup>[28,29]</sup> The frequency of such an oscillating field was much easier to measure, and the resulting accuracy in the determined nuclear magnetic moments was high. Although this may be considered the first nuclear magnetic resonance experiment, his way of detecting the resonance condition was very different from later approaches: in front and behind the section of homogeneous magnetic fields there were auxiliary sections of an inhomogeneous field that selected the molecules within the beam for their spin state. Without irradiation, the particles would pass the homogeneous section undisturbed, traverse the second inhomogeneous section as a polarized beam on a single trajectory, and were detected upon impact. On resonance irradiation in the homogeneous section would mix the spin states, and fewer particles would take the trajectory of the polarized beam. This could then be registered as a drop in intensity of the detector. For these experiments Rabi was awarded the Nobel Prize in Physics in 1944.

## 1.2 Development of Nuclear Magnetic Resonance Spectroscopy

The mayhem of the Second World War interrupted fundamental research on nuclear magnetism. Evgeny Zavoisky was most likely the first to observe proton nuclear magnetic resonance (NMR) in a condensed phase in 1941, but his magnetic fields were not very homogeneous, and he could not get reproducible results.<sup>[30]</sup> The evacuation of the Russian Academy of Sciences to his university just a month later forced him to abandon his research, and when he reassumed his experiments, he decided to focus on another type of effect. In 1944, he was the first to discover electron paramagnetic resonance.<sup>[31]</sup> On the other hand, the concentrated research effort on radio electronics and radar during the war provided much of the know-how that facilitated the development of NMR spectroscopy. It was no coincidence that Felix Bloch and Edward Purcell, who are often quoted as the fathers of NMR, were both involved in radar research during the war. In 1946, they independently detected the absorption of radio frequency (RF) radiation in condensed matter rich in protons (paraffin for Purcell, water for Bloch) within a magnetic field, setting the stage for NMR spectroscopy.<sup>[32,33]</sup> For this, they were awarded the Nobel Prize in Physics in 1952.

In further experiments the setup was improved and the resolution increased, and it became clear that the resonance frequency of a given nucleus was not independent of the chemical nature of the sample. Although this chemical shift was rather disappointing to

some purist physicists, S. Dharmatti, a chemist working in Stanford with Bloch, realized that this could be used to separately detect the chemically different hydrogen atoms in an organic compound. He recorded spectra of some simple alcohols, which featured an intensity ratio between the signals that was proportional to the number of associated protons, e.g., 1:2:3 in the case of ethanol.<sup>[34]</sup> Because their resolution was still limited by the homogeneity of the magnetic field, the group went through great lengths to improve their magnets and finally reached a homogeneity in the order of 0.5 Hz. This allowed them to observe spin-spin couplings in a number of organic compounds.<sup>[35,36]</sup> It was now clear that NMR spectra contained a plethora of valuable information for chemical analysis: the chemical shift reported on the electronic environment of the nucleus, the signal integrals gave the relative numbers of hydrogen atoms within the compound, and spin-spin coupling could be used to shed light onto the bond network of the molecule. Additionally, relaxation effects could be used to investigate dynamical processes. The brothers Russell and Sigurd Varian had founded a company to commercialize the microwave technology that they had developed during the war. They became interested in NMR spectroscopy, and they introduced the first commercial spectrometer, the Varian A-60, in 1953. It was reliable, affordable, and easy to use, and it paved the way for NMR to become a common tool in analytical chemistry.<sup>[37]</sup>

From here on there were uncountable advancements to make NMR the broad field that it is today. One of the first bigger breakthroughs was the transition from a continuous wave acquisition scheme, where radio waves at constant frequency were irradiated and the magnetic field was swept, to Fourier-transform spectroscopy.<sup>[38]</sup> In this technique, all nuclei were simultaneously excited by a broadband RF pulse, and their free induction decay (FID) was recorded. The Fourier transform was equivalent to field-sweep spectrum, but as all resonances were recorded in a single scan, the sensitivity could be improved by up to two orders of magnitude. Together with the invention of broadband decoupling, this meant that  $^{13}\text{C}$  NMR became feasible.<sup>[39]</sup> At the very beginning the FID was recorded on punch cards and the Fourier transform was calculated in a computing center overnight, but around the same time laboratory minicomputers became available, cutting down computing times and enabling the evaluation on site. Together with the application of an algorithm for fast discrete Fourier transformations, which is the origin of NMR spectroscopist's fondness for powers of two, processing times could be reduced to a matter of minutes or even seconds.<sup>[40]</sup>

As commercial Fourier-transform NMR became available in the early 1970s, this opened the door for many new experiments to be explored. Less sensitive nuclei could be investigated, and time-dependent phenomena became accessible via longitudinal and transverse relaxation times. One of the most influential new ideas was the design of pulse sequences with an incremented delay that yielded an array of time-domain data that could be Fourier-transformed into a two-dimensional spectrum. This allowed for a more direct probing of spin interactions, which manifested themselves as cross peaks in the 2D plane between the interacting nuclei. The type of interaction that was probed in this way depended on the pulse sequence design; arguably the simplest and first example was to correlate protons via their  $J$ -couplings, as it was already proposed by Jean Jeener at a summer school in 1971.<sup>[41]</sup> When Ernst et al. picked up on this idea, they demonstrated

the generality of this approach.<sup>[42]</sup> Other notable examples of interactions than can be probed this way are the longitudinal cross-relaxation rates between protons, which contain distance information,<sup>[43]</sup> and heteronuclear  $J$ -couplings, which can be achieved in various different ways.<sup>[44–46]</sup> Another benefit of these 2D techniques was that the signals were dispersed in another dimension, which reduced spectral overlap. There are no ternary interactions that would fundamentally require three-dimensional experiments, but the advantage of reducing overlap was particularly interesting for the investigation of biomacromolecules, and thanks to the development of isotopic enrichment soon higher-dimensional experiments emerged.<sup>[47–51]</sup> It must not be forgotten that these developments were only possible thanks to the rapid increase in the performance of computers, as they enabled the researchers to store and process the large amounts of data necessary for multidimensional spectroscopy. A more recent development for a similar reason was the possibility to greatly reduce the experimental time and increase the signal-to-noise ratio by undersampling the spectrum, known as non-uniform sampling (NUS). It requires orders of magnitude more computing time to process these data, and its emergence and spread is therefore linked to the availability of increasingly powerful computers.<sup>[52]</sup>

A more gradual improvement was the steady increase in magnetic field strength. Going to high fields makes the RF engineering more complex and challenging, but it improves both the sensitivity and the signal dispersion. Additionally, it reduces the effects of strong couplings, making the spectra more easily interpretable. While permanent and electromagnets were practically limited to fields of around 100 MHz, the advent of cryogenic, superconducting solenoid magnets opened the door for NMR at much higher fields.<sup>[53]</sup> Today, spectrometers operating at proton resonance frequencies of up to 1200 MHz are available.

Apart from improving existing methodology, some spectroscopists sought to broaden the applicability of NMR. NMR is somewhat unusual compared to other types of spectroscopy, such as electronic, vibrational, or rotational spectroscopy, in the sense that the highest resolution spectra can be acquired in solution and not in the gas phase. This is because the fast rotational tumbling in solution averages strong anisotropic effects such as dipolar couplings and chemical shift anisotropy, which has been a great convenience for analytical chemistry. In solids, no such averaging occurs, and spectra in this state of matter are plagued by excessive broadening. However, it was shown as early as 1958 that rapid rotation of the sample at a magic angle of  $\arctan \sqrt{2} \approx 54.74^\circ$  relative to the field, which is the root of the modulation term  $3 \cos^2 \theta - 1$ , can also be used to average out anisotropic effects. Thanks to this trick of magic angle spinning (MAS) solid-state NMR has developed into an entirely separate field, and engineers have been constantly trying to increase the spinning speed to generate even sharper lines. Today, samples can be spun at mind-blowing speeds of up to 110 kHz.

Nuclear magnetic resonance is (indirectly) well-known to the general public due to the development of magnetic resonance imaging (MRI) by Paul Lauterbur and Peter Mansfield starting from 1973.<sup>[54,55]</sup> By using carefully controlled magnetic field gradients within the sample, they were able to encode a spatial dimension into the spectrum, and together with an approach similar to multidimensional spectroscopy, this can be used to generate a three-dimensional image of the sample. They were awarded the Nobel Prize

in Physiology or Medicine for their discoveries in 2003, and their technique provides a non-invasive diagnostic tool used in every modern hospital around the world.

### 1.3 Scalar NMR Parameters for Small-Scale Structure

The field that has driven a lot of the improvements outlined in the previous section was the application of NMR in biophysics and the investigation of biomacromolecules such as proteins or DNA/RNA. The large number of individual resonances and the low concentrations of these molecules pose a serious challenge, but there is a lot of information that can be accessed with this technique. There are various methods to extract kinetic information from NMR depending on the time scale of the motion under investigation,<sup>[56–62]</sup> although the infamous supra- $\tau_c$  gap, a timescale in the micro- to nanosecond range, has eluded NMR spectroscopists so far.<sup>[63,64]</sup> Thermodynamic parameters, such as protein-ligand dissociation constants, enthalpy and entropy differences, and conformer populations can also be available.<sup>[65–67]</sup> But also in small-molecule NMR there has been some work that goes beyond the routine structure determination of new molecules, such as the identification of a minor (2%) conformation of the alkaloid strychnine,<sup>[68,69]</sup> or the characterization of exchange in organometallic complexes.<sup>[70,71]</sup>

The present dissertation will be mainly concerned with structure, and although we have made structural investigations of a protein complex and of natural products in separate chapters, here we will discuss the underlying problems and principles together to highlight similarities. The first component to structure in a molecular system is its constitution. Natural products can be very diverse in their constitution, and if they are too flexible to be crystallized, NMR is the most convenient method to determine it. The nature of functional groups can be extracted from the chemical shift. Some obvious distinctions can be made at first glance, such as the difference between an aliphatic and an aromatic proton, but also more subtle effects on the chemical shift can be used as information by comparison with traditional increment tables,<sup>[72,73]</sup> more modern prediction algorithms based on neural networks,<sup>[74,75]</sup> or more elaborate electronic structure calculations.<sup>[76]</sup> The use of scalar couplings for constitution is conceptually very simple, and it can be used to determine a molecule's bond network, although this does not mean that this is always straightforward to execute in practice. The constitution of proteins is much simpler, and mostly boils down to their amino acid sequence. This sequence can be determined by step-wise hydrolysis in a process called Edman degradation,<sup>[77,78]</sup> and more modern approaches use (tandem) mass spectrometry for this task.<sup>[79]</sup> Although this could in principle also be achieved by means of NMR, it would be a rather tedious task and require a much larger amount of protein, and scientists naturally choose the path of least resistance.

The task of structure elucidation becomes more challenging when it comes to configuration and conformation, which determine the three-dimensional arrangement of atoms in the molecule. The difference between conformation and configuration is not entirely obvious when examined more closely, and the International Union of Pure and Applied Chemistry (IUPAC) defines conformations as differing stereoisomers that can

be interconverted by rotation around formally single bonds, while configurations are all other stereoisomers.<sup>[80]</sup> In most cases, this leads to conformers being in rapid exchange with each other, while configurations cannot interconvert, although there are exceptions and edge cases such as the stereoisomerism of the aromatic system of gymnochrome G (Chapter 4). In the context of 3D structure determination this rapid interconversion is the relevant difference, although we will try to not assign too much meaning to this terminology. Both types of stereoisomerism lead to differences in the atom's spatial arrangement, which can be probed by NMR.

The longitudinal cross-relaxation rate between two nuclei and the associated nuclear Overhauser effect (NOE) provide distance information. For the most basic experimental design the resulting constraints are rather qualitative (the nuclei are within a certain distance or not), but with more effort, quantitative and accurate distances can be extracted.<sup>[68]</sup> The magnitude of the NOE scales with the inverse sixth power of the distance ( $r^{-6}$ ), and it is therefore a short-ranged effect which requires the protons to be within about 4–5 Å of each other.<sup>[81]</sup> Another parameter that is sensitive to molecular geometry is the  $J$ -coupling. This is maybe most useful for vicinal proton-proton couplings, which have a strong dependence on the associated dihedral angle described by the Karplus equation.<sup>[82]</sup> There are similar relationships for other cases such as proton-carbon couplings.<sup>[83]</sup> All these couplings also have a dependence on the chemical environment, such as the presence of electronegative substituents, and on other aspects of the geometry. Proteins have the advantage of having a very uniform chemical structure in the backbone, and parametrizations for various specific couplings found in the backbone have been developed.<sup>[84–86]</sup> Natural products on the other hand are accessible by electronic structure methods, most notably density functional theory (DFT), which can be used to predict scalar couplings for a given molecular structure. This prediction takes into account all chemical and geometrical aspects of this structure, and therefore any type of experimental coupling can be used to cross-checked against a given molecular model.<sup>[87]</sup>

Although the chemical shift is most straightforward to interpret in terms of constitution and chemical functionality, it is also sensitive to local molecular conformation. This dependence is not straightforward to describe, and the application in proteins could again profit from their chemical uniformity. For example, in different secondary structure elements the chemical shift of backbone atoms deviates systematically from the shifts observed in random coil, and the reversing of this observation lets one deduce secondary structure from chemical shifts.<sup>[88,89]</sup> There are now methods to get backbone torsion angles directly from chemical shifts, and more recent versions make use of modern computational methods such as neural networks.<sup>[90–92]</sup> The use of chemical shifts for the spatial structure of small molecules is again tied to the general availability of good *ab initio* methods, and it has seen an large amount of development in the last 20 years.<sup>[93–97]</sup>

The collection of these scalar, isotropic parameters are therefore rich in structural information and have been used for many decades for the structure elucidation of various types of molecules. One of the more impressive achievements in this regards certainly is the first protein structure in solution determined by Williamson et al. in 1985, who used NOE and  $J$ -constraints and in combination with molecular mechanics.<sup>[98]</sup> However, they

suffer from a common disadvantage, as they all only provide short-distance constraints: NOEs are useful for distances up to 4–5 Å,  $J$ -couplings mostly report on dihedral angles, and the chemical shift is only sensitive to differences in local geometry. Both for proteins and small molecules this becomes problematic if large-scale structure needs to be assessed. The typical case for natural products are molecules with two different stereochemically interesting sections, called stereoclusters, which are separated by a flexible linker, and it can be very difficult to determine the relative configuration of the two stereoclusters with respect to each other. Similarly in proteins, the relative arrangement of multiple rigid domains with respect to each other cannot be determined using these short-ranged parameters. It is therefore necessary to turn to a different class of NMR-based constraints, which will be the overarching theme of this dissertation.

## 1.4 Tensorial NMR Parameters for Large-Scale Structure

Two examples of such a different type of constraints are the effects caused by molecular alignment and by paramagnetic centers. They are closely related, and both of them are able to provide us with long-range or distance-independent structural information. Here, we will discuss them in a more prosaic and conceptual way, while a rigorous mathematical treatment can be found in Chapter 2.

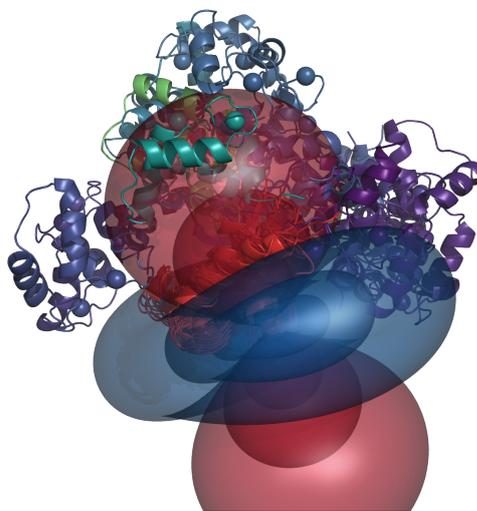
Molecular alignment describes a situation where the orientational distribution of the molecule in question with respect to the magnetic field is non-uniform. The triumph of solution-state NMR was caused by the simple fact that the large anisotropic interactions such as dipolar couplings and chemical shift anisotropy are averaged out by rapid molecular tumbling, leading to sharp lines in the spectrum. However, small deviations from isotropic tumbling can reintroduce residual anisotropic effects, which was already observed and theoretically described in the earlier days of NMR.<sup>[99,100]</sup> Notable examples for these effects are residual dipolar couplings (RDCs), which are residues of the strong through-space interaction of magnetic moments, and residual chemical shift anisotropy (RCSAs), which are differences in the motional average of the chemical shift due to alignment and chemical shift anisotropy. The application of these anisotropic effects as structural parameters, starting with RDCs, was mostly driven by the biomolecular NMR community. To make the most of such effects, the molecular alignment needs to be carefully tuned; if it is too small, the effects are difficult to measure, but if the alignment is too strong, too many anisotropic effects get reintroduced, leading to line-broadening and overly complex spectra. The development of tuneable alignment media, such as bicelles, filamentous phages, and polyacrylamide gels, opened the door for a more widespread use of anisotropic NMR.<sup>[101–103]</sup>

As the alignment is uniform within the entire (rigid) molecular system, it provides orientational constraints with respect to a global molecular frame, which therefore have no distance limitation. RDCs give angular constraints on internuclear vectors, and they can be used for protein structure refinement, although they always require some first estimate of the structure.<sup>[104]</sup> They also report on dynamics, as they get partly averaged out if the internuclear vector is in motion. This can report on motion in the microseconds

regime and faster, and by comparison with results from relaxation studies they can be used to pinpoint motion to the previously mentioned supra- $\tau_c$  window.<sup>[105]</sup> As they do not depend on local distances, they can be used to determine relative orientation on very large molecular complexes.<sup>[106–108]</sup> For the structure elucidation of small molecules RDCs are particularly useful for cases where multiple stereoclusters are relatively far apart from each other, but they can also be used as a general tool to gain additional information and certainty.<sup>[109–111]</sup> More recently, the use of RCSAs has been demonstrated to be possible, and they provide information complementary to RDCs.<sup>[112–115]</sup> To do so, it was necessary to develop ways to correct for the unavoidable isotropic chemical shift perturbations.

The second type of long-range parameters that were employed in this work are paramagnetic constraints caused by open-shell metal centers. Most of these paramagnetic metals have anisotropic  $\mathbf{g}$ -tensors and as such anisotropic susceptibilities (notable exceptions are  $\text{Mn}^{2+}$  and  $\text{Gd}^{3+}$ ), which lead to pronounced effects in NMR spectra. Firstly, the combination of a strong magnetic field and an anisotropic susceptibility leads to alignment, and the associated principles and applications we have elaborated in the previous paragraph apply similarly. A notable advantage of paramagnetic alignment is the fact that it originates from a well-defined point in the molecular system, namely the metal center. If we investigate a molecular system with two separate parts connected by a linker with a certain degree of flexibility, the part that contains the metal will experience full alignment, while the alignment in the other part will be scaled down due to the part’s relative motion. This can be exploited to investigate the dynamics of such a two-part system, which is the topic of Chapter 3. In a medium, the alignment would be induced from outside into both domains, and these outside contributions are very difficult or impossible to deconvolute from the contribution mediated via the linker.

In addition to alignment, paramagnetic centers cause a series of other effects in NMR spectra. Firstly, there is paramagnetic relaxation enhancement (PRE). It is the only effect described here that depends on the isotropic susceptibility, and it adds an additional contribution to the transverse relaxation rate. As a result, resonances of nuclei close to the metal are broadened beyond detection, but as the effect scales with the inverse sixth power of the distance to the metal, this is a rather short-ranged effect. It can be used to gather distance information in the regime where the broadening is measurable, but does not eliminate the resonance, and metals with isotropic susceptibilities are typically employed for this purpose. In the present work, PRE is undesired as it prevents us from observing certain resonances, but this is only a



**Figure 1.1.** Pseudocontact shift isosurfaces at 5, 1, and 0.2 ppm for Dysprosium, displayed on a cartoon representation of an ensemble of CaM/Munc13-1. The red and blue surfaces correspond to positive and negative PCS, respectively. For details about the ensemble, see Fig. 3.13.

minor nuisance due to the short range of about 10 Å. The second effect discussed here is the pseudocontact shift (PCS). Just as alignment it is governed by the anisotropy of the magnetic susceptibility, and the observed chemical shift perturbation is caused by an averaged dipole-dipole interaction between the electron spin and the shifted nucleus. It depends on the inverse third power of the distance to the metal, and as such differs from alignment, but it is much longer-ranged than PRE and can be observed at distances of up to 50 Å. In addition to the distance, it also depends on the orientation of the vector between the metal and the nucleus relative to the susceptibility tensor. Fig. 1.1 illustrates this orientational dependence, and it also demonstrated the far-reaching nature of PCSs. The distance-dependence of pseudocontact shifts is both a blessing and a curse: on one hand, this distance information is valuable and cannot be determined from alignment effects, but the non-linear dependence and the convolution of rotation and translation can make the interpretation much more challenging.<sup>[116]</sup>

Interestingly, the chemical shift perturbations caused by PCSs initiated the investigation of proteins by NMR spectroscopy, because strongly shifted resonances in metalloproteins could be resolved in a 1D spectrum.<sup>[117,118]</sup> To make most of these effects, they need to be accurately determined as the difference between a paramagnetic system and its diamagnetic analogue. To ensure that these differences can be exclusively attributed to paramagnetism, the diamagnetic analogue needs to be chemically as similar as possible to its paramagnetic counterpart. Maybe the best choice of such metals are the lanthanides, and they have therefore been extensively used for paramagnetic NMR.<sup>[119–123]</sup> They are all chemically very similar and can therefore be readily exchanged in a given binding pocket. The periodic table has eleven different anisotropic, paramagnetic lanthanides with varying shapes and sizes of susceptibility tensors (we have excluded Promethium from this list, which is paramagnetic, but due to its radioactivity rather inconvenient for use in NMR spectroscopy). There are four options for a diamagnetic reference (Lanthanum, Yttrium, Lutetium, and Scandium), and one can choose the reference which is chemically most similar to the paramagnetic metals that one has used (e.g., by comparison of ionic radii).

It is not always obvious how to introduce these paramagnetic centers into a system to make use of their effects. In this work, we have the convenience of a protein mutant that selectively binds lanthanides into one of its native calcium binding pockets.<sup>[124]</sup> But there has been some effort to develop tags that can be added to almost any protein to make the technique of using lanthanides more broadly applicable,<sup>[125–127]</sup> and this idea has also been extended to oligonucleotides and sugars.<sup>[128–131]</sup>

Both the effect of alignment and of paramagnetism can be conveniently described using two second-rank tensors. One tensor  $\mathbf{G}$  describes the global effect, i.e., the alignment ( $\mathbf{A}$ ) or the anisotropic susceptibility ( $\Delta\chi$ ). A second tensor  $\mathbf{M}$  depends on a local structural feature, such as the internuclear vector for RDCs, the chemical shift tensor for RCSAs, or the metal-nucleus vector for PCSs, and it is therefore unique for each individual data point. The functional form of all parameters can then be expressed in the very general form  $\text{tr}(\mathbf{GM})$  (see Eqs. (2.53), (2.59), (2.62) and (2.73)). The structural feature that goes into  $\mathbf{M}$  typically has three degrees of freedom, such as a 3D vector or the orientation of a tensor in space. Since the associated constraint (RDC/PCS/RCSA) is a scalar, it

cannot be used to unambiguously determine this vector or orientation, but only to reduce the degrees of freedom by one. For example, a PCS cannot be used to pinpoint the exact location of a nucleus. It can only be deduced that it must lay somewhere on a surface such as the one in Fig. 1.1. On its own, this is not tremendously enlightening, which is in contrast to NOEs and  $J$ -couplings, which provide much more tangible information such as distances or bond dihedrals. This is further complicated by the fact that the global tensor  $\mathbf{G}$  is typically not known *a priori* (an exception being the method developed in Chapter 4), but has to be inferred from the tensorial parameters. The strategy of using these parameters therefore has to be somewhat different.

Generally, tensorial NMR parameters are used in conjunction with an existing structural model, and the agreement of the constraints with this model is checked. Unless previously known, the five independent components of  $\mathbf{G}$  are determined by fitting them to the data and model. This is repeated for different structural models, and the agreement of the data with the various models lets us deduce which one is most likely to be correct. For natural products, the different structural options are typically the molecule's possible diastereomers, although this approach can in principle also be applied to distinguish constitutional isomerism. The necessary information within the model depends on the type of parameters that are used, but in addition to the geometry one may need molecular properties such as chemical shift tensors, electric field gradients, diamagnetic susceptibilities, and electric dipole moments. For small molecules this information can be generated by using *ab initio* methods thanks to the enormous progress in molecular modeling. In flexible cases, conformational freedom needs to be taken into account, and the populations can be either determined from calculated free energies or by taking them as additional fit parameters. Although most interest in small molecules is in determining their configuration, this is no fundamental restriction, and the exploration of the conformational space of lactose is an example for a different kind of research question.<sup>[130,131]</sup>

For proteins, structural models often exist, either from X-ray crystallography, from scalar NMR constraints, or from cryogenic electron microscopy. Molecular mechanics can be used in addition to generate different conformations derived from these structures, and additional information such as chemical shift tensors can be estimated thanks to the chemical homogeneity of proteins. The details on how the models are composed and evaluated depends on the type of question that is addressed with these tensorial parameters, and these questions are somewhat more diverse in proteins than in small molecules. In this work we investigated the interdomain dynamics of a two-domain protein complex, and the approach that we have taken will be outlined in the following section.

## 1.5 Scope, Outline, and Summary

The scope of this dissertation is to push the boundaries of what is possible with tensorial parameters in NMR. As briefly mentioned in the previous section, paramagnetic NMR is particularly well suited to study interdomain dynamics of proteins as the paramagnetic

center is localized within one domain. Therefore, the nuclei within this domain will experience the full degree of paramagnetic effects induced by the metal. The other domain, on the other hand, moves with respect to the paramagnetic center, and the measured effects will be a time average over all relative orientations. This way, it is possible to probe the conformational ensemble sampled by the two domains. This approach is not new, and the protein calmodulin and its various complexes have been a popular target for these studies for more than 20 years.<sup>[120,122,123]</sup> For the sake of simplicity, this interdomain motion is often reduced to a scalar quantity called an order parameter, and although this certainly gives some first insight into the motion, we believe that much more information can be extracted from paramagnetic constraints. Russo et al. conducted a more thorough study of the complex of Calmodulin with the IQ recognition motif from the voltage-gated calcium channel  $\text{Ca}_v1.2$  (CaM/IQ), and described the motion via an ensemble of conformations. However, as this complex restricts the motion quite strongly, which is obvious from the order parameter of around 0.9, this could not demonstrate the full potential of the technique. In Chapter 3 we report the application of this methodology to the complex of calmodulin with Munc13-1. This complex is interesting due to the unique binding motif of Munc13-1 (for a more complete introduction to the protein, see Section 3.1), and it is expected to restrict the interdomain flexibility of calmodulin much less than other complexes. To exploit the full potential of paramagnetic NMR, we have acquired four types of RDCs and four types of PCSs for six different metals, leading to a very large number of 2691 constraints for the mobile domain. From this data we have first determined an order parameter of the complex of 0.162. In contrast to the findings of Bertini et al.<sup>[122]</sup> for free calmodulin, we did not find a pronounced dependence on the metal, and we do argue why this is expected to be the norm and not the exception. We then go further to find ensembles that give us a very detailed picture of the conformational space sampled by calmodulin, and we find that the interdomain movement is a combination of both rotation and translation. Finally, we believe that this data set could be used in the future to explore more formal motional models and to investigate more strictly which characteristics of the interdomain motion can be distinguished by RDCs and PCSs and which cannot.

In Section 1.4 we have argued the case for anisotropic NMR in natural product chemistry, as it is particularly valuable in cases of separated stereoclusters. However, we believe that it is useful as a more general tool to corroborate one's findings. Natural products can be very complex molecules, and the unambiguous interpretation of scalar NMR parameters can turn out to be quite difficult. As a result, structural misassignments are more common than they should be.<sup>[132,133]</sup> It is therefore always a good idea to apply different approaches for structure elucidation and check whether they all come to the same conclusion. In particular, the combination of spectroscopic data with molecular modeling is a strategy that has seen a lot of development in the more recent past, and many techniques rely on results from DFT and molecular mechanics simulations.<sup>[134–136]</sup> Anisotropic NMR is an example of such a combination, and the field-alignment approaches that we developed in Chapters 4 and 5 benefit even more from molecular modeling. Inspired by the magnetic field alignment of paramagnetic proteins, we have shown that it is possible to use diamagnetic alignment on natural products for

the elucidation of their structure. Unlike in the case of alignment media, the degree of magnetic alignment can be reliably predicted by DFT, which greatly enhances the discriminating power of this approach. This is expected to be particularly useful for flexible molecules. In Chapter 5 we set out to design a device for the alignment via electric fields. While this shares many of the benefits with magnetic alignment, it may be even superior since the alignment could be turned on and off at the push of a button, eliminating the need for multiple magnetic fields. Additionally, it may be applicable to a larger variety of molecules. This was mostly an engineering challenge, and so far we were not able to induce alignment via electric fields. Nonetheless, we have learned a plethora of lessons in this rather unfamiliar field, and we shall continue to pursue this goal in the future.

Finally, in Chapter 6 we have made an excursus into metrology. In all our NMR experiments it was necessary to set the sample temperature as accurately as possible to avoid or reduce temperature-dependent systematic perturbations of the chemical shift. This was typically done using a standard sample with a pronounced and well-known chemical shift dependence, which is called an NMR thermometer. In our cases, we have used methanol- $d_4$  for this purpose, as was proposed and calibrated by Findeisen et al.<sup>[137]</sup> Due to technical limitations they have calibrated this NMR thermometer only in a very limited range, and as we had the capability of low-temperature NMR in our lab, we set out to extend this calibration. While we were at it, we also aimed at improving the accuracy of the calibration. We have collaborated with the German national metrology institute to get reliable calibrations of our resistance thermometer, and we also noticed and corrected a systematic deviation of these thermometers due to the strong magnetic fields. As a result, we present a revised calibration valid in the full liquid range of methanol of 175–338 K with an estimated uncertainty ( $2\sigma$ ) of around 25 mK above 240 K and up to 190 mK below 240 K.



# Chapter 2

## Theoretical Background

In this chapter we develop the theoretical framework that was necessary in the main work. It is largely based on various excellent reviews and textbooks, and we have covered some fundamentals of NMR,<sup>[81,138–140]</sup> alignment and anisotropic NMR,<sup>[104,138,141]</sup> paramagnetic NMR,<sup>[142–144]</sup> molecular modeling,<sup>[140,145]</sup> and the theory of electrostatics.<sup>[146–148]</sup> We have made an effort to unify both the notation as well as the general approach to highlight the parallels between some seemingly distinct parts, such as paramagnetic NMR and anisotropic NMR. It is refrained from explaining symbols of quantities that are considered to be common convention, such as  $\hbar$  and  $k_B$ . Instead, all symbols are found in the List of Symbols.

### 2.1 Fundamentals of NMR

#### 2.1.1 Nuclear Spin and Matrix Representations

Nuclear magnetic resonance spectroscopy is a technique that makes use of nuclear spin. This is a type of quantum mechanical angular momentum, and the first hints on its existence were a mathematical quirk found when exploring the quantum mechanics of angular momentum. Angular momentum  $\mathbf{l}$  is a vectorial quantity, and the relevant observables and operators are its three Cartesian components  $\hat{l}_x$ ,  $\hat{l}_y$ , and  $\hat{l}_z$ , and the square of its magnitude  $\hat{l}^2$ . These operators obey the following commutation relationships:

$$[\hat{l}^2, \hat{l}_x] = [\hat{l}^2, \hat{l}_y] = [\hat{l}^2, \hat{l}_z] = 0, \quad [\hat{l}_x, \hat{l}_y] = i\hbar\hat{l}_z, \quad [\hat{l}_y, \hat{l}_z] = i\hbar\hat{l}_x, \quad [\hat{l}_z, \hat{l}_x] = i\hbar\hat{l}_y. \quad (2.1)$$

They can be constructed in a specific representation (such as the position representation) and their eigenstates can be found. Since the Cartesian components of  $\hat{\mathbf{l}}$  do not commute and therefore cannot have common eigenfunctions, the conventional angular momentum eigenstates are chosen to be eigenfunctions of  $\hat{l}^2$  and  $\hat{l}_z$ . These functions are the spherical harmonics  $Y_l^m$ , characterized by the two quantum numbers  $l \in \mathbb{N}_0$  and  $m \in \{-l, -l+1, \dots, l\}$  with the following eigenrelations:

$$\hat{l}^2 Y_l^m = l(l+1)\hbar^2 Y_l^m, \quad (2.2)$$

$$\hat{l}_z Y_l^m = m\hbar Y_l^m. \quad (2.3)$$

It is also possible to derive these properties, i.e., the eigenrelations and the states as characterized by the quantum numbers, by using nothing but the commutations relationships in Eq. (2.1), and without the need to find explicit spatial functions such as  $Y_l^m$  as states. The only difference to the explicit derivation is that there is an additional set of states with half-integer  $l$  with no explicit analogue. As it turns out, the quantum mechanical spin can adopt these half-integer quantum numbers, and it is this property which is the basis of magnetic resonance.

Spin is a fundamental property of elementary particles which behaves similar to angular momentum. This means that its magnitude and with that the quantum number  $l$  is a constant for each particle. Atomic nuclei are not elementary, and they can occupy different excitation states differing in spin, but since excitation energies are many orders of magnitude larger than the thermal energy, these excited states are not found outside of radioactive processes and high-energy physics. Therefore, their spin can be considered constant for the purpose of this discussion, and the associated quantum number is commonly denoted  $I$  instead of  $l$ . As a consequence of this fixed spin, the atomic nuclei have a magnetic moment  $\boldsymbol{\mu}$  associated with them, which is proportional to the spin angular momentum  $\mathbf{I}$ :

$$\boldsymbol{\mu} = \gamma \hbar \mathbf{I}. \quad (2.4)$$

The proportionality constant  $\gamma$  is known as the gyromagnetic ratio and is also a characteristic property of each type of atomic nucleus. If the atomic nuclei are subjected to a magnetic field  $\mathbf{B}_0$ , their magnetic moment will interact with them according to the energy relation of a dipole in a field:

$$E = -\boldsymbol{\mu} \cdot \mathbf{B}_0 = -\gamma \hbar \mathbf{I} \cdot \mathbf{B}_0. \quad (2.5)$$

By letting the Cartesian axis  $z$  be along the magnetic field axis and expressing this relation in terms of quantum mechanical operators, one gets the so-called Zeeman Hamiltonian  $\hat{\mathcal{H}}_{\text{Zee}}$ , which describes the fundamental interaction in NMR. It is common to omit the ubiquitous factor of  $\hbar$ , which results in Hamiltonians expressed in angular frequency instead of energy:

$$\hat{\mathcal{H}}_{\text{Zee}} = -\gamma B_0 \hat{I}_z. \quad (2.6)$$

The energy eigenstates of this Hamiltonian are the eigenstates of  $I_z$ . As we have stated earlier, these states are characterized by the two quantum numbers  $I$  and  $m$  (or  $m_I$ ) and are denoted  $|I, m_I\rangle$ . Since  $I$  is a constant when talking about a specific nucleus, it is omitted most of time. These states are used as basis for the standard matrix representation of the spin operators, called the Zeeman eigenbasis. Using the commutation relationships it is possible to derive the effect of  $\hat{I}_x$  and  $\hat{I}_y$  on these states, but here we shall merely present the results in the form of their matrices. The form and size of these matrices obviously depends on  $I$ , and we will use the most common case  $I = 1/2$  as an example in the following illustrations. There are then two Zeeman eigenstates  $|1/2, 1/2\rangle$  and  $|1/2, -1/2\rangle$  (also known as  $|\alpha\rangle$  and  $|\beta\rangle$ ), and the matrix representation of

the Cartesian spin operators are the following:

$$\hat{I}_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{I}_y = \frac{i}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \hat{I}_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.7)$$

Again, these matrices are implicitly stated in units of  $\hbar$ . The transition that is observed in NMR is the one between the two Zeeman eigenstates. This occurs at the following angular frequency  $\omega_0$ , which is the difference between the eigenvalues of  $\hat{\mathcal{H}}_{\text{Zee}}$ :

$$\omega_0 = \left( -\gamma B_0 \frac{1}{2} \right) - \left( -\gamma B_0 \left( -\frac{1}{2} \right) \right) = -\gamma B_0. \quad (2.8)$$

This frequency is known as the nuclear Larmor frequency and depends on the magnetic field at the site of the nucleus. However, nuclei inside a molecule experience a slightly different field to  $B_0$ : the external magnetic field induces a response into the electron cloud of the molecule, leading to a local field contribution at the nucleus  $B^{\text{ind}}$  (see also Sections 2.1.2 and 2.3.2). This effect leads to a shift in transition frequency in the order of parts per million, which is a reporter on the electronic structure and therefore the chemical environment of the nucleus. This chemical shift is one of the main sources of information that can be gained from NMR spectra and it is the reason why such a fundamental interaction such as magnetic resonance is so useful for chemical analytics.

Another source of information from NMR spectra comes from the interaction of multiple nuclear spins with another. An example is the  $J$ -coupling, an interaction mediated by the electronic environment. We will use the  $J$ -coupling Hamiltonian to illustrate how to construct operators of two-spin systems:

$$\hat{\mathcal{H}}_J = 2\pi J \hat{\mathbf{I}}_1 \cdot \hat{\mathbf{I}}_2 = 2\pi J \left( \hat{I}_{1x}\hat{I}_{2x} + \hat{I}_{1y}\hat{I}_{2y} + \hat{I}_{1z}\hat{I}_{2z} \right), \quad (2.9)$$

where  $J$  is the scalar coupling constant. The index (1, 2) on the operators indicates which of the two spins the operator is associated with. When multiplying matrix representations of operators of two different spins, it is implied that the Kronecker product is to be taken, for example:

$$\hat{I}_{1x}\hat{I}_{2x} = \hat{I}_{1x} \otimes \hat{I}_{2x} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (2.10)$$

This is equivalent to constructing these matrices from the four combination states  $|\alpha\alpha\rangle$ ,  $|\alpha\beta\rangle$ ,  $|\beta\alpha\rangle$ , and  $|\beta\beta\rangle$ , and realizing that the effect of an operator on these combination states only depends on the state of the corresponding spin.

Certain phenomena are not observable for nuclei with  $I = 1/2$ . One example are residual quadrupolar couplings (Section 2.3.3), which in this work is observed on deuterium, a nucleus with  $I = 1$ . In this case there are three Zeeman states  $|1, 1\rangle$ ,  $|1, 0\rangle$ , and  $|1, -1\rangle$ ,

and the operator's matrix representations are therefore  $3 \times 3$ -matrices:

$$\hat{I}_x = \frac{\sqrt{2}}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{I}_y = \frac{i\sqrt{2}}{2} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{I}_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (2.11)$$

Using these basics, we will derive all theoretical results necessary in this work in the following sections.

## 2.1.2 The Secular Approximation

The secular approximation is one of the most fruitful approximations for Hamiltonians in NMR. As it applied on multiple occasions in the following sections, we will here present the general concept. Roughly speaking, this approximation is applicable when the Hamiltonian is the sum of a large, strong (e.g., Zeeman) and a small (e.g., chemical shift, coupling) interaction. It then allows to drop off-diagonal elements of the small contributions when expressed in the eigenbasis of the large one:

$$\hat{\mathcal{H}} = \hat{A} + \hat{B} = \begin{pmatrix} a_{11} & 0 & \dots \\ 0 & a_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \dots \\ b_{21} & b_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.12)$$

Here,  $\hat{A}$  is the large operator and  $\hat{B}$  the small one. It is valid to set  $b_{mn} \stackrel{!}{=} 0$  if  $|b_{mn}| \ll |a_{mm} - a_{nn}|$ . For example, consider a single spin under the influence of chemical shift. Its Hamiltonian consist of a Zeeman term and a chemical shift term:

$$\hat{\mathcal{H}} = -\gamma_I B_0 \hat{I}_z - \gamma_I \mathbf{B}^{\text{ind}} \cdot \hat{\mathbf{I}} = -\gamma_I B_0 \hat{I}_z - \gamma_I (B_x^{\text{ind}} \hat{I}_x + B_y^{\text{ind}} \hat{I}_y + B_z^{\text{ind}} \hat{I}_z). \quad (2.13)$$

Due to chemical shift anisotropy (see also Section 2.3.2) the induced field at the site of the nucleus  $\mathbf{B}^{\text{ind}}$  is not necessarily parallel to the main magnetic field  $\mathbf{B}_0$ , which is along  $z$  by definition. As a consequence we get contributions of  $\hat{I}_x$  and  $\hat{I}_y$  in the Hamiltonian, which contain the off-diagonal elements in the Zeeman eigenbasis. However, since  $B_{x,y,z}^{\text{ind}} \ll B_0$ , it is valid to discard them and to let

$$\hat{\mathcal{H}} \approx -\gamma_I B_0 \hat{I}_z - \gamma_I B_z^{\text{ind}} \hat{I}_z. \quad (2.14)$$

In case of a two-spin system the situation is a little more complex. For two heteronuclear spins  $I$  and  $S$  all Zeeman eigenstates are separated by large energy differences, so for small interactions all contributions to the Hamiltonian except for  $\hat{I}_z$ ,  $\hat{S}_z$ , and  $\hat{I}_z \hat{S}_z$  can be discarded. In the homonuclear case with  $I_1$  and  $I_2$  however the central states  $|\alpha\beta\rangle$  and  $|\beta\alpha\rangle$  are near-degenerate, so the corresponding off-diagonal elements may not be discarded. These two off-diagonal elements are contained in the following linear

combinations of the Cartesian operators, called flip-flop terms:

$$\hat{I}_{1x}\hat{I}_{2x} + \hat{I}_{1y}\hat{I}_{2y} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \hat{I}_{1x}\hat{I}_{2y} - \hat{I}_{1y}\hat{I}_{2x} = \frac{i}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2.15)$$

Whether these terms may be dropped or not depends on the energetic details of the spin system and typically boils down to a comparison of the chemical shift difference and the coupling strength of the two spins involved.

## 2.2 Alignment and Mechanisms

Alignment in NMR is a situation in which the orientational distribution of the analyte molecules with respect to the magnetic field is non-uniform, which means that certain molecular orientations are more probable than others. This induces anisotropic parameters which provide useful structural information, and this has been used extensively in this work. In this section we will show how the alignment of a molecule can be expressed as a tensor, and then derive the expressions for this tensor for the case of magnetic and electric field alignment. Since field alignment can be traced back to a simple dipole-field energy term, the alignment tensors can be derived from first principles using only molecular properties. This is in contrast to induced alignment, which is caused by a complex interaction between the analyte and the alignment medium, and the corresponding alignment tensors are typically difficult to impossible to predict.

### 2.2.1 Description of Alignment as a Tensor

In general, the orientational distribution of a molecule in solution is a probability density function on the surface of a sphere. If this probability density is expressed as a spherical harmonics expansion, only the five second order components are necessary to describe alignment effects in NMR due to the second-rank nature of the occurring interactions. Here, however, we will describe the alignment in terms of a second-rank tensor in Cartesian coordinates, which is an equivalent description. While in the laboratory frame the main magnetic field is static along  $z$ , this is not true in the molecular frame. From the molecule's point of view the magnetic field rapidly and randomly changes direction due to the molecule's rotational tumbling. Let this time-dependent field be expressed in terms of a magnitude and a direction:

$$\mathbf{B}_0(t) = B_0 \hat{\mathbf{b}}(t) \quad \text{with} \quad \|\hat{\mathbf{b}}(t)\| = 1. \quad (2.16)$$

Using this, we define a probability matrix  $\mathbf{P}$  that contains the time average of pairwise products of Cartesian component of  $\hat{\mathbf{b}}(t)$ :

$$\mathbf{P} = \overline{\hat{\mathbf{b}}(t)\hat{\mathbf{b}}(t)^\top} = \begin{pmatrix} \overline{\hat{b}_x^2(t)} & \overline{\hat{b}_x(t)\hat{b}_y(t)} & \overline{\hat{b}_x(t)\hat{b}_z(t)} \\ \overline{\hat{b}_y(t)\hat{b}_x(t)} & \overline{\hat{b}_y^2(t)} & \overline{\hat{b}_y(t)\hat{b}_z(t)} \\ \overline{\hat{b}_z(t)\hat{b}_x(t)} & \overline{\hat{b}_z(t)\hat{b}_y(t)} & \overline{\hat{b}_z^2(t)} \end{pmatrix}. \quad (2.17)$$

Note that because of ergodicity it makes no difference whether time or ensemble averages are used for the definition of this matrix.  $\mathbf{P}$  is symmetric and has a trace of 1 since  $\hat{\mathbf{b}}$  is a unit vector by definition, so it has five independent components. For most purposes it is more convenient to define a traceless version of this matrix:

$$\mathbf{A} = \mathbf{P} - \frac{1}{3}\mathbf{1}. \quad (2.18)$$

This matrix  $\mathbf{A}$  is called the alignment tensor and can be interpreted as the deviation of the probability density from uniformity. It will be used in the following to calculate all anisotropic parameters. In the literature it is common to diagonalize the alignment tensor and use its eigenframe as a molecular frame, and to reduce the three eigenvalues to an axial and a rhombic component. That way many equations become much more concise, elegant, and easy to work with for pen-and-paper applications. However, since the computer implementations used in this work are written for the case of arbitrary Cartesian frames, this is the form we will show throughout this chapter.

## 2.2.2 Magnetic Field Alignment

A strong magnetic field can perturb the uniform orientational distribution of a molecule. The magnetic field induces a magnetic moment in the molecule, which in return interacts with the magnetic field. If the induced magnetic moment is orientation-dependent, this interaction leads to an orientation-dependent energy term, which is the cause of alignment. Since this magnetically induced alignment is unavoidable and ubiquitous in NMR due to the ever-present strong magnetic field, it is sometimes called self-alignment. In the following we will derive the expression for the alignment tensor based on molecular properties.

The magnetic moment  $\boldsymbol{\mu}$  in a molecule is proportional to the external magnetic field and is, in general, dependent on the field's orientation. The quantity describing this proportionality, the magnetic susceptibility  $\boldsymbol{\chi}$ , is therefore a tensorial quantity:

$$\boldsymbol{\mu} = \frac{\boldsymbol{\chi}\mathbf{B}}{\mu_0}. \quad (2.19)$$

To determine the energy  $E$  needed to induce this moment one has to integrate over the magnetic field:

$$E = - \int \mathbf{B}_0 \cdot d\boldsymbol{\mu} = - \int_0^{B_0} \mathbf{B} \cdot \boldsymbol{\chi} d\mathbf{B} = - \frac{\mathbf{B}_0^\top \boldsymbol{\chi} \mathbf{B}_0}{2\mu_0} = -B_0^2 \frac{\hat{\mathbf{b}}^\top \boldsymbol{\chi} \hat{\mathbf{b}}}{2\mu_0}. \quad (2.20)$$

With this energy it is possible to calculate the time averages in  $\mathbf{A}$  (Eq. (2.18)) according to a Boltzmann distribution. The state space are all possible unit vectors  $\hat{\mathbf{b}}$ , which corresponds to the surface of a sphere. This calculation therefore involves integration over this surface, so for the solution of these integrals it is useful to express  $\hat{\mathbf{b}}$  in spherical coordinates:

$$\hat{\mathbf{b}} = \begin{pmatrix} \hat{b}_x \\ \hat{b}_y \\ \hat{b}_z \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}. \quad (2.21)$$

All integration necessary to derive the alignment tensor can be reduced to a few elementary integrals, shown below:

$$\int \hat{b}_x^2 d\Omega = \int_0^{2\pi} \int_0^\pi \hat{b}_x^2 \sin \theta d\theta d\phi = \int_0^{2\pi} \int_0^\pi \sin^3 \theta \cos^2 \phi d\theta d\phi = \frac{4\pi}{3}, \quad (2.22)$$

$$\int \hat{b}_x^4 d\Omega = \int_0^{2\pi} \int_0^\pi \hat{b}_x^4 \sin \theta d\theta d\phi = \int_0^{2\pi} \int_0^\pi \sin^5 \theta \cos^4 \phi d\theta d\phi = \frac{4\pi}{5}, \quad (2.23)$$

$$\int \hat{b}_x^2 \hat{b}_y^2 d\Omega = \int_0^{2\pi} \int_0^\pi \hat{b}_x^2 \hat{b}_y^2 \sin \theta d\theta d\phi = \int_0^{2\pi} \int_0^\pi \sin^5 \theta \sin^2 \phi \cos^2 \phi d\theta d\phi = \frac{4\pi}{15}. \quad (2.24)$$

The symmetry of these integrations allow us to exchange the Cartesian coordinates at will without changing the result, i.e.,  $\int \hat{b}_x^2 d\Omega = \int \hat{b}_y^2 d\Omega$ ,  $\int \hat{b}_x^2 \hat{b}_y^2 d\Omega = \int \hat{b}_x^2 \hat{b}_z^2 d\Omega$  etc. Additionally, integrations involving odd powers of the vector components are zero since they are symmetric integrations over odd functions, e.g.,  $\int \hat{b}_x d\Omega = \int \hat{b}_x^2 \hat{b}_y \hat{b}_z d\Omega = \dots = 0$ . This will greatly simplify the calculations that follow. The first step in Boltzmann statistics is the calculation of the partition function  $Z$ , which is the sum of (or integral over) the Boltzmann factor of all states:

$$Z = \int \exp\left(-\frac{E}{k_B T}\right) d\Omega \approx \int d\Omega = 4\pi. \quad (2.25)$$

Since in these calculations we are well within the high temperature limit  $E \ll k_B T$  the occurring exponentials can be approximated using Taylor expansions. In the case of the partition function it is appropriate to use a zeroth-order expansion, while for the probability density  $\rho$  a first-order approximation is necessary:

$$\rho(\theta, \phi) = \frac{1}{Z} \exp\left(-\frac{E}{k_B T}\right) \approx \frac{1}{4\pi} \left(1 + \frac{B_0^2}{2\mu_0 k_B T} \hat{\mathbf{b}}^\top \boldsymbol{\chi} \hat{\mathbf{b}}\right). \quad (2.26)$$

You may note that this probability density does not formally fulfill the normalization criterion due to the non-zero spherical average of the second term in Eq. (2.26). This

is a consequence of using a zeroth-order expansions of the exponential for the partition function and a first-order expansion for the Boltzmann factor. Taking this into account would be equivalent to an additional second-order term in  $E/k_B T$  for the probability density and we will therefore neglect this. Now the time averages within the alignment tensor can be calculated with this probability density. We will differentiate between on- and off-diagonal elements, starting with the latter:

$$\begin{aligned}
 A_{xy} &= \overline{\hat{b}_x \hat{b}_y} = \int \rho(\theta, \phi) \hat{b}_x \hat{b}_y d\Omega = \int \frac{1}{4\pi} \left( 1 + \frac{B_0^2}{2\mu_0 k_B T} \hat{\mathbf{b}}^\top \boldsymbol{\chi} \hat{\mathbf{b}} \right) \hat{b}_x \hat{b}_y d\Omega \\
 &= \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} \int \left( \chi_{xx} \hat{b}_x^2 + \chi_{yy} \hat{b}_y^2 + \chi_{zz} \hat{b}_z^2 + 2\chi_{xy} \hat{b}_x \hat{b}_y + 2\chi_{xz} \hat{b}_x \hat{b}_z + 2\chi_{yz} \hat{b}_y \hat{b}_z \right) \hat{b}_x \hat{b}_y d\Omega \\
 &= \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} \int 2\chi_{xy} \hat{b}_x^2 \hat{b}_y^2 d\Omega = \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} 2\chi_{xy} \frac{4\pi}{15} = \frac{B_0^2}{15\mu_0 k_B T} \chi_{xy}. \tag{2.27}
 \end{aligned}$$

We see that the off-diagonal elements of the alignment tensor are directly proportional to the corresponding elements of the susceptibility tensor. For the diagonal elements we get a similar result:

$$\begin{aligned}
 A_{xx} &= \overline{\hat{b}_x^2} - \frac{1}{3} = \frac{1}{4\pi} \int \left( 1 + \frac{B_0^2}{2\mu_0 k_B T} \hat{\mathbf{b}}^\top \boldsymbol{\chi} \hat{\mathbf{b}} \right) \left( \hat{b}_x^2 - \frac{1}{3} \right) d\Omega \\
 &= \frac{1}{4\pi} \int \left( \hat{b}_x^2 - \frac{1}{3} \right) d\Omega + \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} \int \left( \chi_{xx} \hat{b}_x^4 + \chi_{yy} \hat{b}_x^2 \hat{b}_y^2 + \chi_{zz} \hat{b}_x^2 \hat{b}_z^2 \right) d\Omega \\
 &\quad - \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} \frac{1}{3} \int \left( \chi_{xx} \hat{b}_x^2 + \chi_{yy} \hat{b}_y^2 + \chi_{zz} \hat{b}_z^2 \right) d\Omega \\
 &= \frac{1}{4\pi} \frac{B_0^2}{2\mu_0 k_B T} \left( \frac{4\pi}{5} \chi_{xx} + \frac{4\pi}{15} \chi_{yy} + \frac{4\pi}{15} \chi_{zz} - \frac{1}{3} \left( \frac{4\pi}{3} \chi_{xx} + \frac{4\pi}{3} \chi_{yy} + \frac{4\pi}{3} \chi_{zz} \right) \right) \\
 &= \frac{B_0^2}{15\mu_0 k_B T} \left( \chi_{xx} - \frac{1}{3} \text{tr}(\boldsymbol{\chi}) \right). \tag{2.28}
 \end{aligned}$$

The remaining elements of  $\mathbf{A}$  can be obtained by simply interchanging the Cartesian coordinates in the derivations in Eqs. (2.27) and (2.28) due to symmetry. The final expression for the alignment tensor is given below:

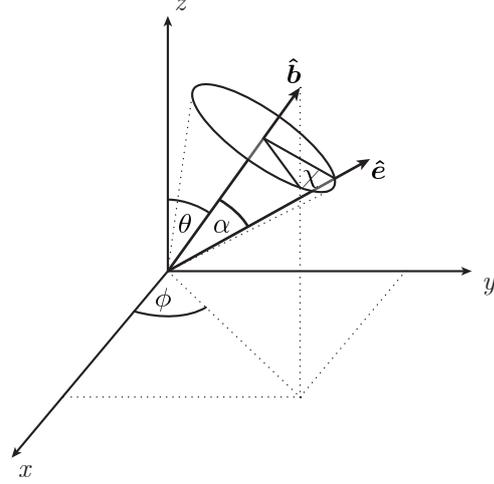
$$\mathbf{A} = \frac{B_0^2}{15\mu_0 k_B T} \left( \boldsymbol{\chi} - \frac{\text{tr}(\boldsymbol{\chi})}{3} \mathbf{1} \right) = \frac{B_0^2}{15\mu_0 k_B T} \boldsymbol{\Delta} \boldsymbol{\chi}, \tag{2.29}$$

with  $\boldsymbol{\Delta} \boldsymbol{\chi}$  being defined as the anisotropic part of the susceptibility tensor, which will play another crucial role in paramagnetic NMR (Section 2.4 and Chapter 3).

### 2.2.3 Electric Field Alignment

Molecules with permanent electric dipole moments can be aligned by external electric fields. The derivation of the corresponding alignment tensor is somewhat complicated

by the fact that this tensor describes the orientation with respect to the magnetic field, while the energy term that is responsible for the alignment is a function of the electric field. Both fields can in principle have an arbitrary, fixed angle  $\alpha$  between each other, depending on the experimental setup. While for magnetic field alignment it was sufficient to consider the magnetic field's orientation  $\hat{\mathbf{b}}$  within the molecular frame, for electric field alignment all three rotational degrees of freedom of the molecule need to be considered. In the molecular frame, the orientation of the magnetic field vector will again be described by the spherical angles  $\theta$  and  $\phi$ . Unlike before, the system is no longer invariant to rotation around the magnetic field axis, as there is an electric field vector at a fixed angle  $\alpha$  relative to the magnetic field. This rotation around  $\hat{\mathbf{b}}$  will be described by the angle  $\chi$ , and it determines the orientation of the electric field vector in the molecular frame. The zero-point of  $\chi$  is defined as the orientation where the electric field is “below” the magnetic field by an angle  $\alpha$ . These four defining angles are illustrated in Fig. 2.1. The electric field vector  $\mathbf{E}$  will again be decomposed into a magnitude  $E$  and a unit vector (i.e., orientation) component  $\hat{\mathbf{e}}$  with  $\mathbf{E} = E\hat{\mathbf{e}}$ . The expressions for the unit vectors  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{e}}$  for  $\chi = 0$  are:



**Figure 2.1.** Illustration of the angles describing the orientations of  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{e}}$ .  $\theta$  and  $\phi$  are the azimuthal and polar angles of  $\hat{\mathbf{b}}$ , respectively.  $\alpha$  is the fixed angle between  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{e}}$ .  $\chi$  describes the rotation of  $\hat{\mathbf{e}}$  around  $\hat{\mathbf{b}}$ .

$$\hat{\mathbf{b}} = \begin{pmatrix} \hat{b}_x \\ \hat{b}_y \\ \hat{b}_z \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}, \quad \hat{\mathbf{e}}(\chi = 0) = \begin{pmatrix} \sin(\theta + \alpha) \cos \phi \\ \sin(\theta + \alpha) \sin \phi \\ \cos(\theta + \alpha) \end{pmatrix}. \quad (2.30)$$

To determine the electric field orientation for an arbitrary angle  $\chi$ , we use the expression for the rotation matrix around a unit vector  $\mathbf{R}_{\hat{\mathbf{b}}}(\chi)$  (also known as Rodrigues' rotation formula):<sup>[149]</sup>

$$\hat{\mathbf{e}} = \mathbf{R}_{\hat{\mathbf{b}}}(\chi)\hat{\mathbf{e}}(\chi = 0), \quad \text{with} \quad \mathbf{R}_{\hat{\mathbf{b}}}(\chi) = (1 - \cos \chi)\hat{\mathbf{b}}\hat{\mathbf{b}}^T + \begin{pmatrix} \cos \chi & -\hat{b}_z \sin \chi & \hat{b}_y \sin \chi \\ \hat{b}_z \sin \chi & \cos \chi & -\hat{b}_x \sin \chi \\ -\hat{b}_y \sin \chi & \hat{b}_x \sin \chi & \cos \chi \end{pmatrix}. \quad (2.31)$$

Evaluating this expression is cumbersome and needs the following the trigonometric identities:

$$\sin^2 \theta + \cos^2 \theta = 1 \quad (2.32)$$

$$\sin(\theta + \alpha) = \sin \theta \cos \alpha + \cos \theta \sin \alpha \quad (2.33)$$

$$\cos(\theta + \alpha) = \cos \theta \cos \alpha - \sin \theta \sin \alpha, \quad (2.34)$$

with which it finally yields:

$$\hat{\mathbf{e}} = \begin{pmatrix} \sin \theta \cos \phi \cos \alpha + \cos \theta \cos \phi \cos \chi \sin \alpha - \sin \phi \sin \chi \sin \alpha \\ \sin \theta \sin \phi \cos \alpha + \cos \theta \sin \phi \cos \chi \sin \alpha + \cos \phi \sin \chi \sin \alpha \\ \cos \theta \cos \alpha - \sin \theta \cos \chi \sin \alpha \end{pmatrix}. \quad (2.35)$$

The next step is the determination of the probability density  $\rho$  using a Boltzmann distribution. The molecule's electric dipole moment  $\mathbf{p}$  is also separated into a magnitude  $p$  and direction  $\hat{\mathbf{p}}$ , and its energy in a field  $E_{\text{dip}}$  can then be expressed as:

$$\begin{aligned} E_{\text{dip}} &= -\mathbf{p}^T \mathbf{E} = -pE \hat{\mathbf{p}}^T \hat{\mathbf{e}} \\ &= pE \left( \hat{p}_x \sin \theta \cos \phi \cos \alpha + \hat{p}_x \cos \theta \cos \phi \cos \chi \sin \alpha - \hat{p}_x \sin \phi \sin \chi \sin \alpha \right. \\ &\quad \left. + \hat{p}_y \sin \theta \sin \phi \cos \alpha + \hat{p}_y \cos \theta \sin \phi \cos \chi \sin \alpha + \hat{p}_y \cos \phi \sin \chi \sin \alpha \right. \\ &\quad \left. + \hat{p}_z \cos \theta \cos \alpha - \hat{p}_z \sin \theta \cos \chi \sin \alpha \right) \\ &= pE \left( \hat{\mathbf{p}}^T \hat{\mathbf{b}} \cos \alpha + \sin \chi \sin \alpha (-\hat{p}_x \sin \phi + \hat{p}_y \cos \phi) \right. \\ &\quad \left. + \cos \chi \sin \alpha (\hat{p}_x \cos \theta \cos \phi + \hat{p}_y \cos \theta \sin \phi - \hat{p}_z \sin \theta) \right). \end{aligned} \quad (2.36)$$

To determine the partition function  $Z$  again a zeroth-order approximation of the Boltzmann factor is appropriate. Contrary to the derivation of magnetic alignment (Section 2.2.2) an additional integration over the third degree of freedom  $\chi$  has to be performed:

$$Z = \int_{\chi=0}^{2\pi} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \exp\left(-\frac{E_{\text{dip}}}{k_B T}\right) \sin \theta d\theta d\phi d\chi \approx \int_0^{2\pi} \int_0^{2\pi} \int_0^{\pi} \sin \theta d\theta d\phi d\chi = 8\pi^2. \quad (2.37)$$

For the probability density  $\rho$  a second-order approximation of the Boltzmann factor is necessary (first-order would yield zero alignment):

$$\rho(\theta, \phi, \chi) = \frac{1}{Z} \exp\left(-\frac{E_{\text{dip}}}{k_B T}\right) \approx \frac{1}{Z} \left( 1 - \frac{E_{\text{dip}}}{k_B T} + \frac{1}{2} \left( \frac{E_{\text{dip}}}{k_B T} \right)^2 \right). \quad (2.38)$$

Since we need the probability density to calculate averages of products of  $\hat{\mathbf{b}}$  which are not functions of  $\chi$ , we can integrate  $\rho$  over  $\chi$  beforehand to simplify the expression. We integrate from 0 to  $2\pi$  so all terms that contain odd powers of either  $\sin \chi$  or  $\cos \chi$  will vanish. When inserting the energy expression from Eq. (2.36) into the probability density

(Eq. (2.38)) we will discard these terms immediately to greatly reduce complexity:

$$\begin{aligned}
 \rho(\theta, \phi) &= \int_0^{2\pi} \rho(\theta, \phi, \chi) d\chi \\
 &= \int_0^{2\pi} \left[ \frac{1}{8\pi^2} + \frac{pE}{8\pi^2 k_B T} \hat{\mathbf{p}}^\top \hat{\mathbf{b}} \cos \alpha + \frac{1}{16\pi^2} \left( \frac{pE}{k_B T} \right)^2 \left( (\hat{\mathbf{p}}^\top \hat{\mathbf{b}} \cos \alpha)^2 \right. \right. \\
 &\quad \left. \left. + \sin^2 \chi \sin^2 \alpha (-\hat{p}_x \sin \phi + \hat{p}_y \cos \phi)^2 \right. \right. \\
 &\quad \left. \left. + \cos^2 \chi \sin^2 \alpha (\hat{p}_x \cos \theta \cos \phi + \hat{p}_y \cos \theta \sin \phi - \hat{p}_z \sin \theta)^2 \right) \right] d\chi \\
 &= \frac{1}{4\pi} + \frac{pE}{4\pi k_B T} \hat{\mathbf{p}}^\top \hat{\mathbf{b}} \cos \alpha + \frac{1}{8\pi} \left( \frac{pE}{k_B T} \right)^2 \left( (\hat{\mathbf{p}}^\top \hat{\mathbf{b}} \cos \alpha)^2 \right. \\
 &\quad \left. + \frac{1}{2} \sin^2 \alpha (-\hat{p}_x \sin \phi + \hat{p}_y \cos \phi)^2 \right. \\
 &\quad \left. + \frac{1}{2} \sin^2 \alpha (\hat{p}_x \cos \theta \cos \phi + \hat{p}_y \cos \theta \sin \phi - \hat{p}_z \sin \theta)^2 \right). \tag{2.39}
 \end{aligned}$$

After expansion of the squares this expression can again be simplified greatly by making use of the trigonometric Pythagoras:

$$\rho(\theta, \phi) = \frac{1}{4\pi} + \frac{1}{16\pi} \left( \frac{pE}{k_B T} \right)^2 \sin^2 \alpha + \frac{pE}{4\pi k_B T} \hat{\mathbf{p}}^\top \hat{\mathbf{b}} \cos \alpha + \frac{1}{16\pi} \left( \frac{pE}{k_B T} \right)^2 \hat{\mathbf{b}}^\top \hat{\mathbf{p}} \hat{\mathbf{p}}^\top \hat{\mathbf{b}} (3 \cos^2 \alpha - 1). \tag{2.40}$$

Again this probability density is not exactly normalized, as a zeroth-order expansion of the exponential for the partition function and a second-order expansion for the Boltzmann factor was used. Taking this into account is equivalent to a fourth-order term in  $E/k_B T$  and will again be neglected. Determining the alignment tensor requires the calculation of the time average of  $\hat{\mathbf{b}} \hat{\mathbf{b}}^\top - \mathbf{1}/3$ . This is simplified by realizing, just as in the case of magnetic alignment, that only second-order terms in  $\hat{\mathbf{b}}$  of the probability density will contribute to these averages, which is the last term in Eq. (2.40). By comparing this expression for  $\rho$  with the case of magnetic alignment (Eqs. (2.26) and (2.40)), the alignment tensor can be derived by analogy:

$$\mathbf{A} = \frac{1}{30} \left( \frac{pE}{k_B T} \right)^2 \left( \hat{\mathbf{p}} \hat{\mathbf{p}}^\top - \frac{1}{3} \mathbf{1} \right) (3 \cos^2 \alpha - 1). \tag{2.41}$$

We want to point out that this expression differs to the one found in the review by Brunner<sup>[104]</sup> by a factor of two, but careful examination and a literature survey led us to the conclusion that it is correct as it is stated above.

## 2.3 Anisotropic Parameters

In the previous section we have introduced alignment as a concept and the effects that cause it. Here we will describe the anisotropic parameters that were exploited in this work. They are residual effects of interactions that average to zero under isotropic

tumbling. These are residual dipolar couplings (RDCs), residual chemical shift anisotropy (RCSAs), and residual quadrupolar couplings (RQCs), which will be discussed in detail in the following.

### 2.3.1 Residual Dipolar Couplings

Dipolar couplings are a consequence of direct through-space dipole-dipole interactions between the magnetic moments of the coupling nuclei. It is roughly three orders of magnitude stronger than the  $J$ -coupling, but it has a pronounced orientational dependence and its average over a uniform distribution is zero. Therefore it is typically not observed in liquid state NMR. However, small degrees of alignment can make them reappear as an additional contribution to line splitting. The full Hamiltonian can be constructed from the expression for the classical dipole-dipole energy and is as follows:

$$\hat{\mathcal{H}}_{\text{dd,full}} = d_{12} \left( 3 \left( \hat{\mathbf{I}}_1 \cdot \hat{\mathbf{r}}_{12} \right) \left( \hat{\mathbf{I}}_2 \cdot \hat{\mathbf{r}}_{12} \right) - \hat{\mathbf{I}}_1 \cdot \hat{\mathbf{I}}_2 \right), \quad \text{with} \quad d_{12} = -\frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r_{12}^3}, \quad (2.42)$$

where  $\gamma_i$  is the gyromagnetic ratio of the nucleus  $i$ , and  $r_{12}$  and  $\hat{\mathbf{r}}_{12}$  are length and direction (as unit vector) of the internuclear vector between the coupling nuclei, respectively. Note that this Hamiltonian depends on the relative orientations of the magnetic dipoles and their connecting vector, but not on the magnetic field, since this interactions is independent of the Zeeman interaction. The well-known dependence on the angle between the internuclear vector and the magnetic field is obtained by applying the secular approximation (Section 2.1.2). The expansion of Eq. (2.42) yields

$$\hat{\mathcal{H}}_{\text{dd,full}} = d_{12} \left( 3 \sum_{p \in \{x,y,z\}} \sum_{q \in \{x,y,z\}} \hat{I}_{1p} \hat{I}_{2q} \hat{r}_p \hat{r}_q - \sum_{p \in \{x,y,z\}} \hat{I}_{1p} \hat{I}_{2p} \right), \quad (2.43)$$

where  $\hat{r}_{x,y,z}$  are the Cartesian components of  $\hat{\mathbf{r}}_{12}$ . Some contributions such as  $\hat{I}_{1x} \hat{I}_{2z}$  can be immediately discarded, others become only obvious after some rearrangement:

$$\begin{aligned} \hat{\mathcal{H}}_{\text{dd,full}} &\approx d_{12} \left( 3 \left( \hat{I}_{1z} \hat{I}_{2z} \hat{r}_z^2 + \hat{I}_{1x} \hat{I}_{2x} \hat{r}_x^2 + \hat{I}_{1y} \hat{I}_{2y} \hat{r}_y^2 + \hat{I}_{1x} \hat{I}_{2y} \hat{r}_x \hat{r}_y + \hat{I}_{1y} \hat{I}_{2x} \hat{r}_y \hat{r}_x \right) \right. \\ &\quad \left. - \left( \hat{I}_{1x} \hat{I}_{2x} + \hat{I}_{1y} \hat{I}_{2y} + \hat{I}_{1z} \hat{I}_{2z} \right) \right) \\ &= d_{12} \left( 3 \left( \hat{I}_{1z} \hat{I}_{2z} \hat{r}_z^2 + \left( \hat{I}_{1x} \hat{I}_{2x} + \hat{I}_{1y} \hat{I}_{2y} \right) \frac{\hat{r}_x^2 + \hat{r}_y^2}{2} + \left( \hat{I}_{1x} \hat{I}_{2x} - \hat{I}_{1y} \hat{I}_{2y} \right) \frac{\hat{r}_x^2 - \hat{r}_y^2}{2} \right. \right. \\ &\quad \left. \left. + \left( \hat{I}_{1x} \hat{I}_{2y} + \hat{I}_{1y} \hat{I}_{2x} \right) \hat{r}_x \hat{r}_y \right) - \left( \hat{I}_{1x} \hat{I}_{2x} + \hat{I}_{1y} \hat{I}_{2y} + \hat{I}_{1z} \hat{I}_{2z} \right) \right) \\ &\approx d_{12} \left( 3 \left( \hat{I}_{1z} \hat{I}_{2z} \hat{r}_z^2 + \left( \hat{I}_{1x} \hat{I}_{2x} + \hat{I}_{1y} \hat{I}_{2y} \right) \frac{\hat{r}_x^2 + \hat{r}_y^2}{2} - \left( \hat{I}_{1x} \hat{I}_{2x} + \hat{I}_{1y} \hat{I}_{2y} + \hat{I}_{1z} \hat{I}_{2z} \right) \right) \right). \end{aligned} \quad (2.44)$$

By making use of the fact that  $\hat{r}_x^2 + \hat{r}_y^2 = 1 - \hat{r}_z^2$  and some more elementary rearrangement, one gets:

$$\hat{\mathcal{H}}_{\text{dd,full}} \approx \hat{\mathcal{H}}_{\text{dd}} = \frac{d_{12}}{2} (3\hat{r}_z^2 - 1) (3\hat{I}_{1z}\hat{I}_{2z} - (\hat{I}_{1x}\hat{I}_{2x} + \hat{I}_{1y}\hat{I}_{2y} + \hat{I}_{1z}\hat{I}_{2z})). \quad (2.45)$$

Now an angle  $\theta$  between  $\hat{\mathbf{r}}_{12}$  and the  $z$ -axis (i.e., the magnetic field) is defined. It is  $\cos^2 \theta = \hat{r}_z^2$ . Substitution yields the well-known expression for the secular dipolar coupling in the (more general) homonuclear case:

$$\hat{\mathcal{H}}_{\text{dd}} = \frac{d_{12}}{2} (3 \cos^2 \theta - 1) (3\hat{I}_{1z}\hat{I}_{2z} - \hat{\mathbf{I}}_1 \cdot \hat{\mathbf{I}}_2). \quad (2.46)$$

The weak coupling case is satisfied if the difference in resonance of the two spins is much larger than the coupling, i.e.,  $4|\omega_1 - \omega_2| \gg |d_{12}(3 \cos^2 \theta - 1)|$ , which is obviously true for all heteronuclear cases but can also be satisfied due to sufficient chemical shift differences. In that case all contributions from  $\hat{I}_{1x}\hat{I}_{2x}$  and  $\hat{I}_{1y}\hat{I}_{2y}$  are discarded and one gets:

$$\hat{\mathcal{H}}_{\text{dd,weak}} = d_{12} (3 \cos^2 \theta - 1) \hat{I}_{1z}\hat{I}_{2z} = 2\pi D \hat{I}_{1z}\hat{I}_{2z}. \quad (2.47)$$

In the weak coupling case, which we will consider in the following, the experimentally observed splitting of the lines in Hz is in principle given by the dipolar coupling constant  $D$ . This quantity  $D$  as given in Eq. (2.47) however still contains the angle  $\theta$  as a time dependent variable. To find the experimental splitting the time average of the term  $3 \cos^2 \theta - 1$  needs to be computed, for which we will make use of the alignment tensor (Eq. (2.18)). The alignment tensor  $\mathbf{A}$  is an object in the molecular frame. Here, the internuclear vector  $\mathbf{r}_{12}$  is static, while the unit vector in the direction of the magnetic field  $\hat{\mathbf{b}}$  is time dependent. The cosine of the angle  $\theta$  is simply the scalar product of the two unit vectors:

$$\cos \theta = \hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{b}} = (\hat{b}_x \hat{r}_x + \hat{b}_y \hat{r}_y + \hat{b}_z \hat{r}_z). \quad (2.48)$$

Squaring this term yields:

$$\begin{aligned} \cos^2 \theta &= (\hat{b}_x \hat{r}_x + \hat{b}_y \hat{r}_y + \hat{b}_z \hat{r}_z)^2 \\ &= \hat{b}_x^2 \hat{r}_x^2 + \hat{b}_x \hat{b}_y \hat{r}_x \hat{r}_y + \hat{b}_x \hat{b}_z \hat{r}_x \hat{r}_z \\ &\quad + \hat{b}_y \hat{b}_x \hat{r}_y \hat{r}_x + \hat{b}_y^2 \hat{r}_y^2 + \hat{b}_y \hat{b}_z \hat{r}_y \hat{r}_z \\ &\quad + \hat{b}_z \hat{b}_x \hat{r}_z \hat{r}_x + \hat{b}_z \hat{b}_y \hat{r}_z \hat{r}_y + \hat{b}_z^2 \hat{r}_z^2 \\ &= \hat{\mathbf{r}}_{12}^T \begin{pmatrix} \hat{b}_x^2 & \hat{b}_x \hat{b}_y & \hat{b}_x \hat{b}_z \\ \hat{b}_y \hat{b}_x & \hat{b}_y^2 & \hat{b}_y \hat{b}_z \\ \hat{b}_z \hat{b}_x & \hat{b}_z \hat{b}_y & \hat{b}_z^2 \end{pmatrix} \hat{\mathbf{r}}_{12}. \end{aligned} \quad (2.49)$$

When the time average of this term is taken, the central matrix in the matrix form can be identified as the probability matrix:

$$\overline{\cos^2 \theta} = \hat{\mathbf{r}}_{12}^T \mathbf{P} \hat{\mathbf{r}}_{12}. \quad (2.50)$$

Substituting the probability matrix with the alignment tensor and multiplication with 3 yields the desired term:

$$3\hat{\mathbf{r}}_{12}^T \mathbf{A} \hat{\mathbf{r}}_{12} = 3 \left( \hat{\mathbf{r}}_{12}^T \mathbf{P} \hat{\mathbf{r}}_{12} - \frac{1}{3} \hat{\mathbf{r}}_{12}^T \mathbf{1} \hat{\mathbf{r}}_{12} \right) = \overline{3 \cos^2 \theta - 1}, \quad (2.51)$$

which can then be inserted into Eq. (2.47) to yield the time-averaged dipolar coupling Hamiltonian:

$$\overline{\hat{\mathcal{H}}_{\text{dd,weak}}} = 3d_{12} \hat{\mathbf{r}}_{12}^T \mathbf{A} \hat{\mathbf{r}}_{12} \hat{I}_{1z} \hat{I}_{2z} \quad (2.52)$$

The time-averaged dipolar coupling constant  $\overline{D}$ , which is identical to the RDC splitting  $\Delta\nu_{\text{RDC}}$ , is therefore:

$$\Delta\nu_{\text{RDC}} \equiv \overline{D} = \frac{3}{2\pi} d_{12} \hat{\mathbf{r}}_{12}^T \mathbf{A} \hat{\mathbf{r}}_{12} = -\frac{3\mu_0\gamma_1\gamma_2\hbar}{8\pi^2 r_{12}^3} \hat{\mathbf{r}}_{12}^T \mathbf{A} \hat{\mathbf{r}}_{12} = -\frac{3\mu_0\gamma_1\gamma_2\hbar}{8\pi^2 r_{12}^3} \text{tr} \left( \hat{\mathbf{r}}_{12} \hat{\mathbf{r}}_{12}^T \mathbf{A} \right) \quad (2.53)$$

This final form of the residual dipolar coupling contains, besides some constants, only the alignment tensor and the internuclear vector  $\mathbf{r}_{12}$ . In the weak coupling case the  $J$ -coupling Hamiltonian has the same form as the dipolar Hamiltonian (replace  $D$  with  $J$  in Eq. (2.47)). The observed splitting of the line, called the total coupling  $T$ , is then simply the sum of the two contributions  $T = J + D$ , which is the theoretical basis for Eq. (4.4).

### 2.3.2 Residual Chemical Shift Anisotropy

The chemical shift is a small change in the resonance frequency of a nucleus due to small magnetic fields induced into the electron cloud by the main magnetic field. This induced field is highly anisotropic, i.e., it depends on the molecule's orientation with respect to the main magnetic field. A (very rough) rule of thumb states that the typical extent of chemical shift anisotropy of a given nucleus is approximately as big as the isotropic chemical shift range of that nucleus. Since the molecules tumble rapidly in solution, the resonances observed in liquid state NMR are the averages over all orientations. Normally we assume that all orientations are equally probable, which gives us the isotropic chemical shift. If we have alignment and certain orientations are more probable than others, the orientational average will obviously change. This shift is known as residual chemical shift anisotropy (RCSA). The anisotropy of the induced magnetic field  $\mathbf{B}^{\text{ind}}$  can be described using the chemical shift (deshielding) tensor  $\boldsymbol{\delta}$ , which leads to the following chemical shift Hamiltonian:

$$\hat{\mathcal{H}}_{\text{cs}}^{\text{lab}} = -\gamma_I \mathbf{B}_{\text{ind}} \cdot \hat{\mathbf{I}} = -\gamma_I \boldsymbol{\delta} \mathbf{B}_0 \cdot \hat{\mathbf{I}}. \quad (2.54)$$

The superscript lab indicates that this Hamiltonian is expressed in the laboratory frame. Here some substantial simplifications can be made: Firstly, the main magnetic field  $\mathbf{B}_0$  is by definition along  $z$ , and secondly all contributions from  $I_x$  and  $I_y$  can be discarded using the secular approximation. After this, only scalars remain in the expression for the Hamiltonian:

$$\hat{\mathcal{H}}_{\text{cs,sec}}^{\text{lab}} = -\gamma_I \delta_{zz} B_0 \hat{I}_z. \quad (2.55)$$

While this looks concise and simple, it comes with the problem that the chemical shift tensor and its components are only constant in the molecular frame, and are time dependent in the laboratory frame due to the rapid rotational tumbling of the molecule. It is therefore necessary to calculate the time average  $\overline{\delta_{zz}}$  to determine the observed chemical shift. In simple words,  $\delta_{zz}$  is the proportionality factor for the  $z$ -component of the induced field, induced by a field along  $z$ . With this, we transition to the molecular frame, keeping in mind that the  $z$ -direction of the laboratory frame is along the direction of the magnetic field. The chemical shift tensor is now constant, while the magnetic field direction is time dependent. The induced field is given by

$$\mathbf{B}^{\text{ind}} = \boldsymbol{\delta} \mathbf{B}_0(t). \quad (2.56)$$

To find what is  $\delta_{zz}$  in the lab frame, we need to find the length of the projection of  $\mathbf{B}^{\text{ind}}$  onto  $\mathbf{B}_0$ . This projection  $B_{\parallel \mathbf{B}_0}^{\text{ind}}$  can simply be found by computing the scalar product of  $\mathbf{B}^{\text{ind}}$  with the unit vector in field direction  $\hat{\mathbf{b}}$ .  $\delta_{zz}^{\text{lab}}$  can then be identified as the proportionality factor between  $B_{\parallel \mathbf{B}_0}^{\text{ind}}$  and  $B_0$ :

$$B_{\parallel \mathbf{B}_0}^{\text{ind}} = \mathbf{B}^{\text{ind}} \cdot \hat{\mathbf{b}} = B_0 \underbrace{\boldsymbol{\delta} \hat{\mathbf{b}} \cdot \hat{\mathbf{b}}}_{\delta_{zz}^{\text{lab}}}. \quad (2.57)$$

Time averaging of  $\delta_{zz}^{\text{lab}}$  then reveals the presence of the probability matrix:

$$\overline{\delta_{zz}^{\text{lab}}} = \overline{\boldsymbol{\delta} \hat{\mathbf{b}} \cdot \hat{\mathbf{b}}} = \text{tr} \left( \overline{\boldsymbol{\delta} \hat{\mathbf{b}} \hat{\mathbf{b}}^{\text{T}}} \right) = \text{tr} (\boldsymbol{\delta} \mathbf{P}) = \delta. \quad (2.58)$$

This time average is nothing else than the experimental chemical shift  $\delta$ . To determine the residual chemical shift anisotropy  $\Delta\delta_{\text{RCSA}}$ , we need to compute the difference of the anisotropic, general case from the isotropic case where  $\mathbf{P} = 1/3 \mathbf{1}$ :

$$\Delta\delta_{\text{RCSA}} = \delta^{\text{aniso}} - \delta^{\text{iso}} = \text{tr} (\boldsymbol{\delta} \mathbf{P}) - \text{tr} \left( \boldsymbol{\delta} \frac{1}{3} \mathbf{1} \right) = \text{tr} (\boldsymbol{\delta} \mathbf{A}). \quad (2.59)$$

Note how this equation has the same functional form as the expression for the RDC (Eq. (2.53)), a fact that will become useful in data evaluation. In this section we have defined the chemical shift as the difference to the base frequency  $-\gamma_I B_0$ . This is somewhat unusual because typically the chemical shift is defined relative to some reference shift. However, since RCSA are differences in chemical shifts this distinction has no effect on the final result, so this technicality can be ignored in this case.

### 2.3.3 Residual Quadrupolar Couplings

The quadrupolar coupling is an interaction of the nuclear electric quadrupole moment with the electric field gradient at the site of the nucleus created by the electron cloud. Since the energy of a quadrupole within a field gradient depends on its orientation, the quadrupole interaction is another effect which depends on molecular orientation and

which is modulated by the molecule's rapid tumbling. Its strength depends on both the size of the quadrupole moment, which is a fixed nuclear property, and the size of the electric field gradient, which depends on chemical environment and bond geometry. Most notably all nuclei with  $I = 1/2$  have no electric quadrupole moment and therefore exhibit no such coupling. For spins with  $I \geq 1$  however it can be a very strong effect. The isotropic orientational average is zero, just like with RDCs, so in isotropic liquid state NMR it does not lead to line splitting. In many cases however it is the dominant source of relaxation, increasing line widths by orders of magnitude. Since the quadrupolar coupling is such a strong interaction, it can be a very sensitive probe for alignment, and the residual quadrupolar coupling of the solvent's deuterium nuclei is commonly used for this purpose. Since RQCs play only a minor role in this work, we will discuss them briefly.

The full form of the quadrupole Hamiltonian is the following:

$$\hat{\mathcal{H}}_{\text{Q,full}} = \frac{eQ}{2I(2I-1)\hbar} \hat{\mathbf{I}}^T \mathbf{V} \hat{\mathbf{I}}, \quad (2.60)$$

where  $Q$  is the nuclear quadrupole moment and  $\mathbf{V}$  the electric field gradient tensor. By making use of the fact that  $\mathbf{V}$  is symmetric and traceless, by applying the secular approximation, and by taking the time average, one gets the so-called first order quadrupole Hamiltonian. It contains only the field gradient component  $V_{zz}$ , similar to the case of chemical shift anisotropy (Eq. (2.55)):

$$\hat{\mathcal{H}}_{\text{Q,sec}} = \omega_{\text{Q}} \frac{1}{6} (3\hat{I}_z^2 - I(I+1)\hat{1}), \quad \text{with} \quad \omega_{\text{Q}} = \frac{3eQ\bar{V}_{zz}}{2I(2I-1)\hbar}. \quad (2.61)$$

The quantity  $\omega_{\text{Q}}$  is the (first-order) quadrupolar coupling and corresponds to the observed line splitting in angular frequency. The time averaging of  $V_{zz}$  is analogous to the time averaging of  $\delta_{zz}$  (Eq. (2.58)), and division by  $2\pi$  yields the line splitting in Hz:

$$\Delta\nu_{\text{RQC}} = \frac{\omega_{\text{Q}}}{2\pi} = \frac{3eQ}{4\pi I(2I-1)\hbar} \text{tr}(\mathbf{V}\mathbf{A}). \quad (2.62)$$

Note that since  $\mathbf{V}$  is traceless,  $\mathbf{A}$  and  $\mathbf{P}$  can be used interchangeably in this equation. We use  $\mathbf{A}$  for consistency with the corresponding equations for RDCs and RCSAs.

## 2.4 Paramagnetism

In this section the theoretical framework of paramagnetic effects in NMR is discussed. Some of these effects are a consequence of magnetic alignment, and we will refer to the previous sections wherever it is appropriate.

### 2.4.1 Magnetic Susceptibility in Paramagnetic Centers

The origin of paramagnetism is the presence of unpaired electrons in a molecular system, which in our case belong to lanthanide ions bound in a protein structure. Electrons have magnetic moments originating from their spin  $\mathbf{s}$ , and a collection of unpaired electron is unified to an overall spin  $\mathbf{S}$ :

$$\boldsymbol{\mu} = -\mu_B g_e \mathbf{S}, \quad (2.63)$$

which is the electron analogue to Eq. (2.4), and the product  $-\mu_B g_e$  can be interpreted as an electron gyromagnetic ratio. The magnetic moments of electrons are about three orders of magnitude larger than those of nuclear spins, but in the case of paired electrons they oppose and cancel each other. In unpaired systems this is not the case, and these large magnetic moments will interact with the external magnetic field of an NMR spectrometer, described by the Zeeman Hamiltonian

$$\hat{\mathcal{H}}_{Zee} = \mu_B g_e S_z B_0 \quad (2.64)$$

with the eigenstates  $|m_S\rangle$  and their energy  $E = \mu_B g_e m_S B_0$ . The states with the magnetic moment aligned with the field will be populated more, and by using a Boltzmann distribution approximated to first order one can evaluate the expectation value of the  $z$ -component of spin angular momentum  $S_z$ :

$$\begin{aligned} \langle S_z \rangle &= \frac{1}{2S+1} \sum_{m_S} \langle m_S | S_z | m_S \rangle \left( 1 - \frac{\mu_B g_e m_S B_0}{k_B T} \right) \\ &= -\frac{\mu_B g_e B_0}{3k_B T} S(S+1). \end{aligned} \quad (2.65)$$

Inserting this expression into Eq. (2.63) yields the induced magnetic moment:

$$\mu_z = \frac{\mu_B^2 g_e^2 B_0}{3k_B T} S(S+1), \quad (2.66)$$

The quantity relating the induced magnetic moment with the magnetic field is known as the susceptibility  $\chi$  (compare Eq. (2.19)), and we can identify:

$$\chi = \frac{\mu_B^2 g_e^2 \mu_0}{3k_B T} S(S+1) \quad (2.67)$$

which is also known as the Curie law.

So far we have treated the electron independently of its environment, and as a result we have found the susceptibility to be isotropic. In asymmetric chemical environments however, contributions from orbital angular momentum lead to an anisotropy of the magnetic susceptibility, which is then described as a tensor  $\boldsymbol{\chi}$ :

$$\boldsymbol{\mu} = \frac{\boldsymbol{\chi} \mathbf{B}}{\mu_0}. \quad (2.68)$$

In complex systems with multiple electrons, zero-field splitting, and nearly-degenerate excited states it is difficult to *ab initio* predict the susceptibility tensor both in its magnitude and orientation (one prominent approach being the Van Vleck equation<sup>[150]</sup>), but it is the  $\chi$  anisotropy that governs most of the paramagnetic effects in NMR and it can therefore readily be determined experimentally. It is also common to define a traceless tensor  $\Delta\chi$  that contains only the anisotropic part:

$$\Delta\chi = \chi - \frac{\text{tr}(\chi)}{3}\mathbf{1}. \quad (2.69)$$

This traceless, symmetric tensor lies at the heart of the expressions for the paramagnetic effects observed in this work. Residual dipolar couplings (RDCs) are alignment effects, and as such do not differ from diamagnetic alignment. The derivation of the alignment mechanism can be found in Section 2.2.2, and the derivation for the RDCs in Section 2.3.1. Pseudocontact shifts on the other hand are an effect unique to paramagnetic NMR, and they are described in the following section.

## 2.4.2 Pseudocontact Shifts

The interaction of a nuclear spin with the electron magnetic moment leads to a chemical shift perturbation of that spin. As with the interaction between two nuclear spins, there are two possible mechanisms: the contact shifts are caused by the delocalization of the unpaired electron throughout the bond network of the protein. This through-bond mechanism, which is the analogue of  $J$ -coupling, needs a non-vanishing electronic density at the site of the nucleus. For lanthanide-protein complexes, this mechanism tends to be of minor importance. First, the ionic character of the metal with unpaired electrons in a well-shielded  $f$ -orbital leads to low delocalization of the electron into the ligand amino acids, and second, the resonances of spins close to the metal are broadened beyond detection and cannot be observed. Pseudocontact shifts, in contrast, are chemical shift perturbations that arise due to the through-space dipole-dipole interaction between the electron spin and the atomic nuclei. Unlike in nuclear dipole-dipole coupling this interaction is not averaged to zero by molecular tumbling if the electron magnetic moment varies with the magnetic field orientation, or in other words, if its susceptibility tensor is anisotropic. The Hamiltonian for a through-space interaction of two dipoles  $\mu_1$  and  $\mu_2$  is given below (compare also Eq. (2.42)):

$$\hat{\mathcal{H}}_{\text{dd,full}} = -\frac{\mu_0}{4\pi r_{12}^3} \left( 3 \left( \hat{\boldsymbol{\mu}}_1^T \hat{\mathbf{r}}_{12} \right) \left( \hat{\boldsymbol{\mu}}_2^T \hat{\mathbf{r}}_{12} \right) - \hat{\boldsymbol{\mu}}_1^T \hat{\boldsymbol{\mu}}_2 \right). \quad (2.70)$$

Let us consider this equation in the molecular frame, where the magnetic field direction may be arbitrary ( $\mathbf{B}_0 = B_0 \hat{\mathbf{b}}$ ). For one of the dipoles we insert the expression for the nuclear dipole moment  $\hat{\boldsymbol{\mu}} = \gamma_I \hat{\mathbf{b}} \hat{I}_b$ . The secular approximation allows us to eliminate all components of  $\hat{\mathbf{I}}$  that are not parallel to  $\hat{\mathbf{b}}$ , so only  $\hat{I}_b$  remains (which is identical to  $\hat{I}_z$  in the lab frame). For the other dipole we insert the expression of the electron dipole moment (Eq. (2.68)). As we are interested in the effect on the nuclear spin states and the

electron spin states are much more short-lived, it is appropriate to insert the expectation value for the electron magnetic moment instead of an operator form. The connecting vector  $\mathbf{r}_{12}$  can be approximated with the lanthanide-nucleus vector  $\mathbf{r}_{\text{Ln}}$ . Although the electron is not exactly localized at the metal center, this approximation is reasonable at longer distances, especially since lanthanide metal centers have low delocalization as mention before. With this, Eq. (2.70) becomes:

$$\begin{aligned}
\hat{\mathcal{H}}_{\text{dd,PCS}} &= -\frac{\mu_0}{4\pi r_{\text{Ln}}^3} \left( 3 \left( \gamma_I \hat{I}_b \hat{\mathbf{b}}^\top \hat{\mathbf{r}}_{\text{Ln}} \right) \left( \frac{B_0}{\mu_0} (\boldsymbol{\chi} \hat{\mathbf{b}})^\top \hat{\mathbf{r}}_{\text{Ln}} \right) - \gamma_I \hat{I}_b \hat{\mathbf{b}}^\top \frac{B_0}{\mu_0} (\boldsymbol{\chi} \hat{\mathbf{b}}) \right) \\
&= -\frac{\gamma_I B_0 \hat{I}_b}{4\pi r_{\text{Ln}}^3} \left( 3 \hat{\mathbf{b}}^\top \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\chi} \hat{\mathbf{b}} - \hat{\mathbf{b}}^\top \boldsymbol{\chi} \hat{\mathbf{b}} \right) \\
&= -\frac{\gamma_I B_0 \hat{I}_b}{4\pi r_{\text{Ln}}^3} \hat{\mathbf{b}}^\top \left( 3 \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\chi} - \boldsymbol{\chi} \right) \hat{\mathbf{b}} \\
&= -\frac{\gamma_I B_0 \hat{I}_b}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{\mathbf{b}} \hat{\mathbf{b}}^\top \left( 3 \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\chi} - \boldsymbol{\chi} \right) \right). \tag{2.71}
\end{aligned}$$

Now we take the time average and convert back into the laboratory frame to get the PCS Hamiltonian  $\hat{\mathcal{H}}_{\text{PCS}}$ . That way we find the probability tensor  $\mathbf{P} = \overline{\hat{\mathbf{b}} \hat{\mathbf{b}}^\top}$  as we have seen in the derivation of RDC, RCSA, and RQC before. In this case however, the matrix right of it is not traceless, and therefore the isotropic part of  $\mathbf{P}$  is by far dominating. We may therefore ignore alignment and set  $\mathbf{P} = \mathbf{1}/3$ :

$$\begin{aligned}
\hat{\mathcal{H}}_{\text{PCS}} &= -\frac{\gamma_I B_0 \hat{I}_z}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \overline{\hat{\mathbf{b}} \hat{\mathbf{b}}^\top} \left( 3 \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\chi} - \boldsymbol{\chi} \right) \right) \\
&= -\frac{\gamma_I B_0 \hat{I}_z}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\chi} - \frac{1}{3} \boldsymbol{\chi} \right) \\
&= -\frac{\gamma_I B_0 \hat{I}_z}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\Delta} \boldsymbol{\chi} \right). \tag{2.72}
\end{aligned}$$

By comparing this to the chemical shift Hamiltonian (e.g., Eq. (2.55)), we can identify the contribution of the PCS to the chemical shift:

$$\Delta\delta_{\text{PCS}} = -\frac{1}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^\top \boldsymbol{\Delta} \boldsymbol{\chi} \right) \tag{2.73}$$

The alignment contribution to the PCS should not be confused with the RCSA (Eq. (2.59)), which is an entirely different interaction. It scales approximately with the isotropic part of the  $\boldsymbol{\chi}$ -tensor, and accounts for less than 2% of the overall chemical shift perturbation.

## 2.5 Molecular Mechanics

Molecular mechanics, also referred to as force-field methods, are a way to model molecules and molecular interactions using atoms as the basic building blocks. Unlike in electronic structure methods, electrons and nuclei are not treated explicitly. Instead, the potential hypersurface of the nuclei is expressed as a parameterized function of the nuclear coordinates. This essentially corresponds to the simple view of molecules being spherical atoms connected by springs. To do so, in addition to the atomic coordinates the binding network of the molecule has to be specified. Then, molecular mechanics builds upon the basic idea of organic chemistry that certain building blocks, or functional groups, behave similarly in various molecular environments. Atoms of the same atomic number are further distinguished into atom types; for example, carbon atoms could be separated into alkylic, vinylic, carbonyl, and acetylenic carbons. Based on these atom types and the binding network, the energy  $E_{\text{MM}}$  of the system is expressed as a sum of a handful of contributions:

$$E_{\text{MM}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{oop}} + E_{\text{cross}} + E_{\text{vdW}} + E_{\text{el}}. \quad (2.74)$$

These energy terms are all simple parametric functions of the nuclear coordinates, and therefore forces can be calculated by simple differentiation with respect to these coordinates. In the following we will elaborate on these terms using the MMFF<sup>[151]</sup> force field as an example, which we have used extensively in this work.

The term  $E_{\text{str}}$  describes the bond stretching interaction between two bound atoms  $i$  and  $j$ , and therefore the force or energy needed to change the length  $r_{ij}$  of their bond. In general, this term should have a minimum at some equilibrium length  $r_{0,ij}$ , should diverge at zero distance, and converge to some dissociation energy for infinite distance. However, most force fields are not designed to correctly describe dissociation (or nuclear fusion, for that matter), and a polynomial expansion of the energy around the equilibrium distance is employed. Since the absolute energy is meaningless and the first derivative at the equilibrium vanishes, the first non-zero term is quadratic in the distance deviation  $\Delta r_{ij} = r_{ij} - r_{0,ij}$ , and by neglecting all other terms, one gets a harmonic potential:

$$E_{\text{str},ij}^{\text{harm}} = \frac{1}{2}k_{ij}\Delta r_{ij}^2. \quad (2.75)$$

Here,  $k_{ij}$  is the force constant of this bond, and it depends on both the atoms types of  $i$  and  $j$ . This harmonic potential is symmetric around the minimum and is a reasonable approximation at small displacements. Higher order terms would improve the agreement, but also increase the number of necessary parameters, and parametrization is a significant problem in molecular mechanics. One way to include a cubic and quartic term without additional parameters is to make their coefficients a constant fraction of the harmonic force constant, as it is done in the MMFF force field:

$$E_{\text{str},ij}^{\text{MMFF}} = \frac{1}{2}k_{ij}\Delta r_{ij}^2\left(1 + \alpha\Delta r_{ij} + \frac{7}{12}\alpha^2\Delta r_{ij}^2\right). \quad (2.76)$$

The additional parameter  $\alpha = 2 \text{ \AA}^{-1}$  is the same for all atom pairs, so it does not add to the parametrization complexity. This particular form is obtained by requiring that the higher-order derivatives of  $E_{\text{str},ij}^{\text{MMFF}}$  match those of the Morse potential.

The energy required to change an angle between two bonds is described by  $E_{\text{bend}}$ , and it is treated very similarly to stretching. Since this bending is rather stiff, the expression for  $E_{\text{bend}}$  needs to be only accurate for relatively small displacements from equilibrium.  $E_{\text{bend}}$  is a function of  $\Delta\theta_{ijk} = \theta_{ijk} - \theta_{0,ijk}$ , where  $\theta_{ijk}$  is the angle between three atoms  $i$ ,  $j$ , and  $k$  bound in series, and  $\theta_{0,ijk}$  is their equilibrium angle. The simplest function is again a harmonic potential in  $\Delta\theta_{ijk}$ , whereas MMFF employs a cubic form using a similar approach as for the stretching term:

$$E_{\text{bend},ijk}^{\text{harm}} = \frac{1}{2}k_{ijk}\Delta\theta_{ijk}^2, \quad (2.77)$$

$$E_{\text{bend},ijk}^{\text{MMFF}} = \frac{1}{2}k_{ijk}\Delta\theta_{ijk}^2(1 + a\Delta\theta_{ijk}). \quad (2.78)$$

The bending force constant  $k_{ijk}$  is again dependent on the atom types of  $i$ ,  $j$ , and  $k$ , whereas the cubic-bend constant  $a$  is the same for all bending potentials, and it is  $a = 0.4 \text{ rad}^{-1}$  in MMFF.

$E_{\text{tors},ijkl}$  is the energy associated with the torsional angle  $\omega$  of four atoms  $i$ ,  $j$ ,  $k$ , and  $l$  bound in series. Since torsion is much less stiff than bending, the potential function needs to give reasonable results for all possible angles between 0 and  $2\pi$ , and it needs to describe the periodicity of  $\omega$ . Therefore, a Fourier series is typically used for this purpose, and MMFF uses terms up to order 3:

$$E_{\text{tors},ijkl}^{\text{MMFF}} = \frac{1}{2}(V_{1,ijkl}(1 + \cos \omega) + V_{2,ijkl}(1 + \cos 2\omega) + V_{3,ijkl}(1 + \cos 3\omega)). \quad (2.79)$$

The third-order term describes a potential with three-fold symmetry, such as the torsion of ethane. The first-order term allows for distinction between *syn*- and *anti*-forms such as in butane. The second-order term has two-fold symmetry, which occurs for example in the rotation around a peptide bond. The parameters  $V$  are again dependent on the four atom types involved.

In cases where four atoms  $i$ ,  $j$ ,  $k$ , and  $l$  are bound in a trigonal-planar fashion with  $j$  in the center, it is often necessary to include a term to enforce this planarity. In principle this is already achieved by the three in-plane angle potentials, but since significant out-of-plane bending can be done without perturbing the three internal angles very much, the resulting potential curve is much too shallow. Therefore, an additional four-atom term is included in these cases. This can be done with a dihedral constraint (since the four atoms are not bound in series, this is called an improper dihedral), but better results can be yielded with a distinct out-of-plane term  $E_{\text{oop}}$ . In the case of MMFF this is a simple harmonic of the angle  $\chi_{ijk,l}$  between the plane  $ijk$  and the bond  $jl$ :

$$E_{\text{oop}}^{\text{MMFF}} = \frac{1}{2}k_{ijk,l}^{\text{oop}}\chi_{ijk,l}^2. \quad (2.80)$$

In addition to planar atom arrangements, this term can also be assigned to trigonal-pyramidal geometries to improve modeling of the inversion barrier.

The term  $E_{\text{cross}}$  describes the coupling between the terms described above, and it does not appear in all force fields. MMFF includes only one cross-term in the form of a stretch-bend interaction between three bound atoms  $i$ ,  $j$ , and  $k$ :

$$E_{\text{sb},ijk}^{\text{MMFF}} = (k_{ijk}^{\text{sb}} \Delta r_{ij} + k_{kji}^{\text{sb}} \Delta r_{kj}) \Delta \theta_{ijk}. \quad (2.81)$$

This term can account for the effect that equilibrium bond lengths may depend on the bond angles. In the water molecule, for example, compressing the bond angles to lower values results in larger bond lengths, and the opposite is observed for a widened bond angle.

All previous interactions described the energy profiles in bonded situations, and the remaining two terms are concerned with non-bonded interactions. The van-der-Waals term describes the subtle interaction of the atom's electron clouds. It is attractive at intermediate distances due to induced dipole-dipole interactions, and it becomes strongly repulsive at short distances due to Coulomb and exchange interaction of the electron clouds. This term is responsible for the finite size of the atoms and prevents them to come too close to each other. Simple theoretical arguments can justify an  $r^{-6}$ -dependence of the attractive part, and so a very well-known model for the van-der-Waals interaction is the Lennard-Jones potential:

$$E_{\text{vdW},ij}^{\text{LJ}} = \varepsilon_{ij} \left( \left( \frac{r_{0,ij}}{r} \right)^{12} - 2 \left( \frac{r_{0,ij}}{r} \right)^6 \right), \quad (2.82)$$

where  $r_0$  is the minimum energy distance and  $\varepsilon$  is the well depth. There is no physical justification for the repulsive part to depend on  $r^{-12}$ , but it was chosen for its computational convenience. There are a number of other models for these van-der-Waals forces and MMFF employs a buffered 14-7 potential.<sup>[152]</sup> This is a modification of the Lennard-Jones potential based on experimental data on noble-gas atoms:

$$E_{\text{vdW},ij}^{\text{MMFF}} = \varepsilon_{ij} \left( \frac{1.07r_{0,ij}}{r + 0.07r_{0,ij}} \right)^7 \left( \frac{1.12r_{0,ij}^7}{r^7 + 0.12r_{0,ij}^7} - 2 \right). \quad (2.83)$$

The parameters  $\varepsilon_{ij}$  and  $r_{0,ij}$  again depend on the atom types of the two atoms  $i$  and  $j$ . As the interaction of bonded atoms is already described with the bonded terms, the van-der-Waals energy is only applied to atoms which are three or more bonds apart.

Finally,  $E_{\text{el}}$  describes the electrostatic interaction, which is governed by the Coulomb law:

$$E_{\text{el},ij}^{\text{MMFF}} = \frac{q_i q_j}{\epsilon(r_{ij} + \delta)}. \quad (2.84)$$

The dielectric constant  $\epsilon$  is typically set to its vacuum value  $4\pi\epsilon_0$ , but deviations from this can be used to model solvent and other polarization effects. The ‘‘electrostatic buffering’’ constant  $\delta$ , which is set to a small value of 0.05 Å, does not appear in the correct form of Coulomb's law. It is rather a pragmatic modification that has little impact at typical

distances but prevents the attractive Coulomb interaction of oppositely charged atoms to overcome the van-der-Waals repulsion at very short distances, which would lead to the coalescence of these atoms. The partial atomic charges  $q_i$  are calculated as a sum of contributions from polarized bonds.

Due to the simple functional form of the nuclear potential energy surface in molecular mechanics, it allows for the treatment of very large molecular systems. Protein complexes of several thousand atoms can be modeled on a simple desktop computer on a reasonable timescale, and exploring 100 000 points in the conformational space of a medium-sized natural product such as gymnochrome G can be done in a matter of hours. This is a great advantage of these methods when compared to electronic structure methods. However, the quality of the results is tightly bound to the parametrization, and whereas they perform excellently in chemically uniform and well-studied systems such as proteins, they may give poorer results for more diverse molecules such as natural products, and fail completely if there are no parameters available for given structural elements. Also, the type of information is limited, and molecular properties beyond the energy and the geometry are typically not accessible. It may then be necessary to turn to electronic structure calculations.

## 2.6 Density Functional Theory

Electronic structure methods describe the movement of electrons in the potential of the atomic nuclei to gain insight on molecular properties. The first group of such methods do this by searching for the electronic wave function  $\Psi$  as a solution of the time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi. \quad (2.85)$$

$\Psi$  is a highly complex function, and for a system with  $N$  electrons it depends on  $4N$  variables, three spatial and on spin coordinate for each electron. Walter Kohn and Pierre Hohenberg proved in 1964 that it is in principle not necessary to know the wave function, as all molecular properties and observables such as the energy can be expressed as functionals of the electron density  $\rho$ , which is the basis of density functional theory. The electron density is only a function of three spatial variables, and as such much simpler than the wave function, and the ground state density can be found by minimizing the energy.

Unfortunately, the energy functional, that is relation between an electron density and the energy, is not known. This is in contrast to wave-function methods, where the wave function is unknown, but its connection with the system's energy, i.e., the Hamiltonian, is a relatively simple expression. Therefore, only approximate energy functionals can be employed, and as a result pure orbital-free density functional methods are not accurate enough to be of general use.

A much better approach to this problem is a hybrid method proposed by Kohn and Lu Jeu Sham, which combines wave function and density functional concepts. It can be seen as a density-functional exchange correction to the Hartree-Fock self-consistent field method, and it is both conceptually and computationally very similar to it. We shall

therefore first describe Hartree-Fock theory, and then elaborate on the modifications from Kohn-Sham theory.

### 2.6.1 Hartree-Fock Theory

Hartree-Fock theory is a procedure to find an approximate wave function as solution to the time-independent Schrödinger equation. First, we will separate the motion of electrons and nuclei, as they move on different time scales due to their vastly different masses. In the description of the electron motion, the nuclei are therefore not treated explicitly, and only their positions appear as parameters for the Coulomb potential in the Hamiltonian. This approach is known as Born-Oppenheimer approximation. The most complex part of the remaining Hamiltonian, and the terms that lead to the Schrödinger equation to be unsolvable, is the electron-electron interaction. In Hartree-Fock theory each electron is described in the average field of all the other electrons, and we will see in the following how this leads to an iterative procedure to find a wave function. However, this treatment neglects electron correlation, which is the fundamental flaw of Hartree-Fock.

The wave function  $\Psi$  for an  $N$ -electron system will be constructed from  $N$  one-electron wave functions, or spin-orbitals. These orbitals  $\phi$  are functions of three spatial and one spin coordinate of a single electron, and they are chosen to be orthonormal:

$$\phi_i(x_j, y_j, z_j, \sigma_j) =: \phi_i(j), \quad \langle \phi_i | \phi_j \rangle = \delta_{ij}. \quad (2.86)$$

The total wave function  $\Psi$ , which is a function of the  $4N$  coordinates of the  $N$  electrons, has to obey the Pauli exclusion principle, which means that the sign of  $\Psi$  has to change when interchanging two arbitrary electrons. This condition can be fulfilled by constructing it as a Slater determinant of molecular orbitals (MOs):

$$\Psi_{\text{SD}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix}. \quad (2.87)$$

The trial wave function in Hartree-Fock consists of a single such Slater determinant, which is equivalent to neglecting electron-electron correlation.

As the total wave function is constructed from orbitals  $\phi$ , these orbitals have to be found and optimized to get the optimal Slater determinant. This is done by employing the variational principle: the energy of a trial ground state wave function will always be greater or equal to the energy of the true solution; an optimal solution can therefore be found by minimizing the trial function's energy. To find the energy, we need the system's

electron Hamiltonian  $\hat{\mathcal{H}}_e$ :

$$\hat{\mathcal{H}}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}, \quad (2.88)$$

$$\hat{T}_e = - \sum_i^{N_{\text{elec}}} \frac{1}{2} \nabla_i^2, \quad (2.89)$$

$$\hat{V}_{ne} = - \sum_i^{N_{\text{elec}}} \sum_A^{N_{\text{nuc}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}_i\|}, \quad (2.90)$$

$$\hat{V}_{ee} = \sum_i^{N_{\text{elec}}} \sum_{j>i}^{N_{\text{elec}}} \frac{1}{\|\mathbf{r}_j - \mathbf{r}_i\|}, \quad (2.91)$$

$$\hat{V}_{nn} = \sum_A^{N_{\text{nuc}}} \sum_{B>A}^{N_{\text{nuc}}} \frac{Z_A Z_B}{\|\mathbf{R}_A - \mathbf{R}_B\|}. \quad (2.92)$$

This Hamiltonian consist out of the four contributions of electron kinetic energy  $\hat{T}_e$ , nuclear-electron interaction  $\hat{V}_{ne}$ , electron-electron interaction  $\hat{V}_{ee}$ , and nuclear interaction  $\hat{V}_{nn}$ .  $\mathbf{r}$ ,  $\mathbf{R}$ , and  $Z$  are electron coordinates, nuclear coordinates, and atomic numbers, respectively, and the Hamiltonian is expressed in atomic units (i.e.,  $\hbar = e = a_0 = m_e = 1$ ). As the constituents of the Slater determinant are one-electron functions, it is helpful to rearrange the terms of  $\hat{\mathcal{H}}_e$  by electron index, distinguishing between two-electron contributions  $r_{ij}^{-1}$ , one-electron contributions  $\hat{h}_i$ , and zero-electron contributions  $\hat{V}_{nn}$ :

$$\hat{\mathcal{H}}_e = \sum_i^{N_{\text{elec}}} \hat{h}_i + \sum_i^{N_{\text{elec}}} \sum_{j>i}^{N_{\text{elec}}} r_{ij}^{-1} + \hat{V}_{nn}, \quad (2.93)$$

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_A^{N_{\text{nuc}}} \frac{Z_A}{\|\mathbf{R}_A - \mathbf{r}_i\|}, \quad (2.94)$$

$$r_{ij}^{-1} = \frac{1}{\|\mathbf{r}_j - \mathbf{r}_i\|}. \quad (2.95)$$

We can use these parts to calculate the energy of a Slater determinant. The zero-electron contribution is trivial: as  $\hat{V}_{nn}$  does not depend on any electron coordinates, its expectation value is independent of the wave function, and yields a constant  $V_{nn}$ :

$$\langle \Psi_{\text{SD}} | \hat{V}_{nn} | \Psi_{\text{SD}} \rangle = V_{nn}. \quad (2.96)$$

Before continuing, let us examine more closely the structure of a Slater determinant: it is a sum of products of orbitals. Each product contains each orbital as a factor, and the sum goes over all  $N!$  permutations of the electron coordinates:

$$\Psi_{\text{SD}} = \frac{1}{\sqrt{N!}} \sum_{\tau} \text{sgn}(\tau) \hat{P}_{\tau} \Pi, \quad (2.97)$$

$$\Pi = \prod_i^{N_{\text{elec}}} \phi_i(i). \quad (2.98)$$

Here,  $\Pi$  is the product of the diagonal elements of  $\Psi_{\text{SD}}$ .  $\tau$  corresponds to a permutation, and the operator  $\hat{P}_\tau$  permutes the electron indices (not orbital indices!) of the function it operates on, according to  $\tau$ . If we now calculate the expectation value of  $\hat{h}_i$ , we create all combinations of addends of  $\Psi_{\text{SD}}$ . However, only the diagonal terms contribute, as  $\hat{h}_i$  is only a one-electron operator, and in a cross term with different permutations we will therefore always create an overlap of two different orbitals:

$$\begin{aligned} \langle \Psi_{\text{SD}} | \hat{h}_i | \Psi_{\text{SD}} \rangle &= \frac{1}{N!} \sum_{\tau} \sum_{\tau'} \langle \hat{P}_\tau \Pi | \hat{h}_i | \hat{P}_{\tau'} \Pi \rangle \\ &= \frac{1}{N!} \sum_{\tau} \langle \hat{P}_\tau \Pi | \hat{h}_i | \hat{P}_\tau \Pi \rangle \\ &= \frac{1}{N!} \sum_j (N-1)! \langle \phi_j | \hat{h}_j | \phi_j \rangle = \frac{1}{N} \sum_j h_j. \end{aligned} \quad (2.99)$$

In the last step we have made use of the fact that for each permutation, a different orbital may be a function of the  $i$ th electron's coordinates. More specifically, for each orbital  $j$  there are  $(N-1)!$  permutations for which this is the case. So the expectation value of  $\hat{h}_i$  is the sum over all diagonal elements of  $\hat{h}_i$  with the orbitals  $\phi_j$ . As such, it is independent of the coordinate  $i$ , which we have tried to clarify by switching to the index  $j$ , which is the expected result since all electrons are indistinguishable. For the same reason, we will drop the index on the operator  $\hat{h}$  in the following.

With a similar argument we can evaluate the expectation value of  $r_{ij}^{-1}$ . We again get a non-zero result if we calculate the overlap of same permutations. As an example, for the identity permutation  $\hat{P}_1$  we get

$$\langle \hat{P}_1 \Pi | r_{ij}^{-1} | \hat{P}_1 \Pi \rangle = \langle \Pi | r_{ij}^{-1} | \Pi \rangle = \langle \phi_i(i) \phi_j(j) | r_{ij}^{-1} | \phi_i(i) \phi_j(j) \rangle =: J_{ij}. \quad (2.100)$$

This sort of term  $J_{ij}$  is known as a Coulomb integral. However, as  $r_{ij}^{-1}$  is a two-electron operator, there are cross terms with different permutations that produce non-zero results, in the case that those permutations differ only by two coordinates. Consider, for example, the unity permutation  $\hat{P}_1$  and the permutation that switches the coordinates of  $i$  and  $j$ ,  $\hat{P}_{ij}$ :

$$\langle \hat{P}_{ij} \Pi | r_{ij}^{-1} | \hat{P}_{ij} \Pi \rangle = \langle \phi_i(i) \phi_j(j) | r_{ij}^{-1} | \phi_i(j) \phi_j(i) \rangle =: K_{ij}. \quad (2.101)$$

These terms are known as exchange integrals, and as they are a consequence of the Pauli principle, they have no classical analogue. If one now collects all these terms together, the full expectation value is

$$\langle \Psi_{\text{SD}} | r_{ij}^{-1} | \Psi_{\text{SD}} \rangle = \frac{1}{N(N-1)} \sum_i \sum_{j>i} (J_{ij} - K_{ij}). \quad (2.102)$$

The negative sign of  $K_{ij}$  arises from the fact that it arises from a term with odd-parity permutation ( $\text{sgn}(\tau) = -1$ , Eq. (2.97)). By summing over all  $N$  different  $\hat{h}_i$  and  $N(N-1)$

different  $r_{ij}^{-1}$  (compare Eq. (2.93)), the total energy  $E$  evaluates to:

$$E = \langle \Psi_{\text{SD}} | \hat{\mathcal{H}}_e | \Psi_{\text{SD}} \rangle = \sum_i^N h_i + \frac{1}{2} \sum_i^N \sum_j^N (J_{ij} - K_{ij}) + V_{\text{nn}}. \quad (2.103)$$

The factor of 1/2 allows the nested sums to both run over all electrons, as the diagonal terms cancel each other exactly ( $J_{ii} - K_{ii} = 0$ ).

For the variational procedure, it is helpful to express this in terms of Coulomb and exchange operators  $\hat{J}_j$  and  $\hat{K}_j$ :

$$E = \sum_i^N \langle \phi_i | \hat{h} | \phi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N (\langle \phi_i | \hat{J}_j | \phi_i \rangle - \langle \phi_i | \hat{K}_j | \phi_i \rangle) + V_{\text{nn}}, \quad (2.104)$$

$$\hat{J}_j | \phi_i(2) \rangle = \langle \phi_j(1) | r_{12}^{-1} | \phi_j(1) \rangle | \phi_i(2) \rangle, \quad (2.105)$$

$$\hat{K}_j | \phi_i(2) \rangle = \langle \phi_j(1) | r_{12}^{-1} | \phi_i(1) \rangle | \phi_j(2) \rangle. \quad (2.106)$$

The minimization of  $E$  has to be performed under the condition that the orbitals remain orthonormal, which can be expressed as the constraint  $\langle \phi_i | \phi_j \rangle - \delta_{ij} = 0$ . Such constraints can be included minimizing a Lagrangian  $L$ , which includes the constraints using Lagrange multipliers  $\lambda$ :

$$L = E - \sum_i^N \sum_j^N \lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}). \quad (2.107)$$

A constrained minimization corresponds to finding a stationary point in  $L$  instead of  $E$ :

$$\delta L = \delta E - \sum_i^N \sum_j^N \lambda_{ij} (\langle \delta \phi_i | \phi_j \rangle + \langle \phi_i | \delta \phi_j \rangle) \stackrel{!}{=} 0. \quad (2.108)$$

The variation of the energy  $\delta E$  can be expressed as:

$$\begin{aligned} \delta E &= \sum_i^N (\langle \delta \phi_i | \hat{h} | \phi_i \rangle + \langle \phi_i | \hat{h} | \delta \phi_i \rangle) \\ &+ \frac{1}{2} \sum_i^N \sum_j^N (\langle \delta \phi_i | \hat{J}_j - \hat{K}_j | \phi_i \rangle + \langle \phi_i | \hat{J}_j - \hat{K}_j | \delta \phi_i \rangle \\ &\quad + \langle \delta \phi_j | \hat{J}_i - \hat{K}_i | \phi_j \rangle + \langle \phi_j | \hat{J}_i - \hat{K}_i | \delta \phi_j \rangle) \\ &= \sum_i^N (\langle \delta \phi_i | \hat{h} | \phi_i \rangle + \langle \phi_i | \hat{h} | \delta \phi_i \rangle) + \sum_i^N \sum_j^N (\langle \delta \phi_i | \hat{J}_j - \hat{K}_j | \phi_i \rangle + \langle \phi_i | \hat{J}_j - \hat{K}_j | \delta \phi_i \rangle) \\ &= \sum_i^N (\langle \delta \phi_i | \hat{F} | \phi_i \rangle + \langle \phi_i | \hat{F} | \delta \phi_i \rangle) \quad \text{with} \quad \hat{F} = \hat{h} + \sum_j^N (\hat{J}_j - \hat{K}_j). \end{aligned} \quad (2.109)$$

In the final step, we have introduced a Fock operator  $\hat{F}$ , which is an effective one-electron energy operator. Since it was found in the variation of the energy, it is not directly associated with the total energy. In particular, the sum of Fock operators does not yield the Hamiltonian. With these Fock operators, the variation of the Lagrangian simplifies to:

$$\delta L = \sum_i^N \left( \langle \delta\phi_i | \hat{F} | \phi_i \rangle + \langle \phi_i | \hat{F} | \delta\phi_i \rangle \right) - \sum_i^N \sum_j^N \lambda_{ij} \left( \langle \delta\phi_i | \phi_j \rangle + \langle \phi_i | \delta\phi_j \rangle \right). \quad (2.110)$$

Using the Hermitian property of  $\hat{F}$ , this can be rewritten as:

$$\delta L = \sum_i^N \langle \delta\phi_i | \hat{F} | \phi_i \rangle - \sum_i^N \sum_j^N \lambda_{ij} \langle \delta\phi_i | \phi_j \rangle + \sum_i^N \langle \delta\phi_i | \hat{F} | \phi_i \rangle^* - \sum_i^N \sum_j^N \lambda_{ij} \langle \delta\phi_i | \phi_j \rangle^*. \quad (2.111)$$

As both the variation in  $\langle \delta\phi |$  and  $\langle \delta\phi |^*$  should lead to a stationary point of the Lagrangian, both the first two and the last two terms of Eq. (2.111) should vanish. The same must therefore be true for the difference of the first two terms and the complex conjugate of the last two terms:

$$\sum_i^N \sum_j^N (\lambda_{ij} - \lambda_{ij}^*) \langle \delta\phi_i | \phi_j \rangle = 0. \quad (2.112)$$

This tells us that the Lagrange multipliers can be interpreted as the elements of a Hermitian matrix  $\mathbf{\Lambda}$  ( $\lambda_{ij} = \lambda_{ij}^*$ ). The final so-called Fock equations can be deduced from Eq. (2.111):

$$\hat{F}\phi_i = \sum_j^N \lambda_{ij} \phi_j. \quad (2.113)$$

By inserting this into Eq. (2.111) it can be seen that a solution to Eq. (2.113) will satisfy the condition in Eq. (2.111). We can now make use of the Hermiticity of  $\mathbf{\Lambda}$  by expressing Eq. (2.113) in its eigenframe, which leads to a pseudo-eigenvalue equation:

$$\hat{F}\phi'_i = \varepsilon_i \phi'_i. \quad (2.114)$$

The quantities  $\varepsilon_i$  are the eigenvalues of  $\mathbf{\Lambda}$  and can be interpreted as orbital energies. The reason why Eq. (2.114) is only a pseudo-eigenvalue equation is that the operators  $\hat{J}_j$  and  $\hat{K}_j$  within  $\hat{F}$  depend themselves on the molecular orbitals. It is therefore typically necessary to find a solution iteratively, where a first guess of orbitals is used to calculate  $\hat{F}$ . The solution of Eq. (2.114) is then used to calculate a new  $\hat{F}$ , etc. A set of converged solutions for such a process are called Self-Consistent Field (SCF) orbitals.

So far we have not specified how the variation of the orbitals  $\phi$  is conducted in detail. In almost all applications, the molecular orbitals are expanded in terms of a set

of basis functions  $\chi$ . Often the functional form of these basis functions is similar to atomic orbitals, and therefore this approach is known as Linear Combination of Atomic Orbitals (LCAO). A notable difference of common basis functions to the exact solutions of the hydrogen atom is that instead of having an exponential radial decay, a Gaussian function is typically employed. Although exponential functions provide better results, they are computationally much more difficult, and the gain in computational speed with Gaussian functions more than compensates for the slightly poorer representation. Gaussian functions are therefore almost exclusively used for basis sets, and they are the namesake of the computational chemistry program employed in this work.<sup>[153]</sup>

Let us assume that we have a set of  $M$  basis functions to expand our molecular orbitals. They can then be expressed as linear combinations with coefficients  $c$ :

$$\phi_i = \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha}. \quad (2.115)$$

The choice of a (finite) basis set lets us not only expand our MOs, but it enables us to express all our equations in terms of matrix and vector algebra, which should be familiar from the quantum mechanical description of NMR spectroscopy. Eq. (2.115) can be thought of as representing the molecular orbital  $\phi$  as a vector  $\mathbf{c}$ . If we insert Eq. (2.115) into Eq. (2.114), we get:

$$\hat{F} \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha} = \varepsilon_i \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha}. \quad (2.116)$$

The matrix-vector form of this equation can be generated by multiplying from the left with a specific basis function and integrating. If this is done with all basis functions  $\chi_{\beta}$  and the resulting equations are collected, this can be expressed as:

$$\mathbf{F}\mathbf{c} = \mathbf{S}\mathbf{c}\boldsymbol{\varepsilon}, \quad (2.117)$$

$$\text{with } F_{\alpha\beta} = \langle \chi_{\alpha} | \hat{F} | \chi_{\beta} \rangle \quad \text{and} \quad S_{\alpha\beta} = \langle \chi_{\alpha} | \chi_{\beta} \rangle. \quad (2.118)$$

These are known as the Roothaan-Hall equations, which have the form of a generalized eigenvalue problem of matrices  $\mathbf{F}$  and  $\mathbf{S}$ .  $\boldsymbol{\varepsilon}$  is a diagonal matrix of the eigenvalues, which correspond to MO energies. In systems with an orthonormal basis set the so-called overlap matrix  $\mathbf{S}$  would be a unity matrix, and Eq. (2.117) would reduce to a regular eigenvalue problem of  $\mathbf{F}$ . For molecular systems however, this is never the case, as atomic orbitals at different nuclear centers will not be orthogonal (and depending on basis set design, neither are those at the same nuclear center).

The solution of this equation will yield  $M$  MOs, and the  $N$  with lowest energy will make up our Slater determinant. Unfortunately, it is not quite as straightforward. As mentioned earlier, the Fock operator, and therefore the Fock matrix, depends itself on the

occupied MOs. Let us more closely look at the expression for the Fock matrix elements:

$$\begin{aligned}
 \langle \chi_\alpha | \hat{F} | \chi_\beta \rangle &= \langle \chi_\alpha | \hat{h} | \chi_\beta \rangle + \sum_j^{N(\text{occ.MO})} \langle \chi_\alpha | \hat{J}_j - \hat{K}_j | \chi_\beta \rangle \\
 &= \langle \chi_\alpha | \hat{h} | \chi_\beta \rangle + \sum_j^{N(\text{occ.MO})} \left( \langle \chi_\alpha \phi_j | r_{12}^{-1} | \chi_\beta \phi_j \rangle - \langle \chi_\alpha \phi_j | r_{12}^{-1} | \phi_j \chi_\beta \rangle \right) \\
 &= \langle \chi_\alpha | \hat{h} | \chi_\beta \rangle + \sum_j^{N(\text{occ.MO})} \sum_\gamma^M \sum_\delta^M c_{\gamma j} c_{\delta j} \left( \langle \chi_\alpha \chi_\gamma | r_{12}^{-1} | \chi_\beta \chi_\delta \rangle - \langle \chi_\alpha \chi_\gamma | r_{12}^{-1} | \chi_\delta \chi_\beta \rangle \right) \\
 &= \langle \chi_\alpha | \hat{h} | \chi_\beta \rangle + \sum_\gamma^M \sum_\delta^M D_{\gamma\delta} \left( \langle \chi_\alpha \chi_\gamma | r_{12}^{-1} | \chi_\beta \chi_\delta \rangle - \langle \chi_\alpha \chi_\gamma | r_{12}^{-1} | \chi_\delta \chi_\beta \rangle \right) \quad (2.119)
 \end{aligned}$$

$$\text{with } D_{\gamma\delta} = \sum_j^{N(\text{occ.MO})} c_{\gamma j} c_{\delta j} \quad (2.120)$$

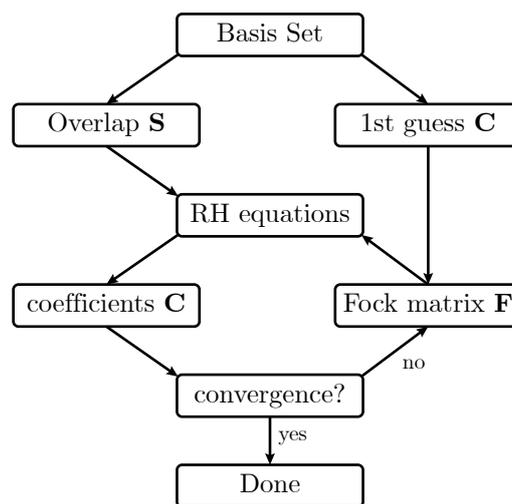
$$\text{and } \langle \chi_\alpha \chi_\gamma | r_{12}^{-1} | \chi_\beta \chi_\delta \rangle = \langle \chi_\alpha(1) \chi_\gamma(2) | r_{12}^{-1} | \chi_\beta(1) \chi_\delta(2) \rangle. \quad (2.121)$$

The four-function integrals have been expressed in a short-hand notation where the order indicates the variable (Eq. (2.121)). The Coulomb and exchange operator depend on the electron distribution, so for their calculation it was necessary to sum over the  $N$  occupied MOs. This could be written more concisely with the use of a density matrix  $\mathbf{D}$ , which is defined analogously to the density matrix known from spin systems: instead of averaging the products of basis coefficients of all spins in a systems, here we average over the occupied molecular orbitals (Eq. (2.120)). Eq. (2.119) can be written even more concisely by using the elements of a rank-4 tensor  $\mathbf{G}$  containing all possible four-function integrals of the basis functions:

$$F_{\alpha\beta} = h_{\alpha\beta} + \sum_\gamma^M \sum_\delta^M G_{\alpha\beta\gamma\delta} D_{\gamma\delta}. \quad (2.122)$$

The iterative process in which a SCF solution is found is represented in Fig. 2.2. With a basis set and an initial guess for the MO coefficients, a Fock matrix and the corresponding Roothaan-Hall equations are constructed. The resulting solution coefficients are compared with the previous ones, and if they are the same (within a certain criterion), the iteration stops. Otherwise, a new Fock matrix is generated to start a new cycle.

The matrices  $\mathbf{h}$  and  $\mathbf{S}$  only depend on the basis set and do not change during the iteration, and they can be calculated in advance. In principle, the same is true for the



**Figure 2.2.** Flowchart of the iterative process to find self-consistent field solutions of the Hartree-Fock problem.

tensor  $\mathbf{G}$ . Since it has  $M^4$  elements, this becomes quickly unfeasible with an increasing basis set size. For example, the calculations performed in Chapter 4 all had around 1200 basis functions, and  $\mathbf{G}$  would have around  $2 \cdot 10^{12}$  elements. No computer can store such an amount of data, and it is therefore necessary to calculate its elements transiently during the calculation of the Fock matrix. Thanks to many clever approximation techniques which can identify terms that do not contribute to  $\mathbf{F}$ , Hartree-Fock and DFT calculations often scale considerably better than the theoretical  $M^4$ , and large molecular systems become accessible with this method.

### 2.6.2 Kohn-Sham Theory

Kohn-Sham theory is a variant of DFT in which the concept of orbitals is reintroduced. This comes at the price of increasing the complexity from 3 to  $3N$  variables, but it provides a better approximation to the kinetic energy, which is the main difficulty of orbital-free DFT. If the system is described as a Slater determinant, the kinetic energy can be calculated as follows:

$$T_S = \sum_i^N \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \right\rangle. \quad (2.123)$$

As this kinetic energy is calculated from a single Slater determinant (as indicated by its subscript), it is not exact, and the missing contribution is considered as a kinetic correlation energy. However, it is a substantial improvement over available purely density-functional forms. The electron density can be expressed in terms of the orbitals as

$$\rho_S = \sum_i^N |\phi_i|^2, \quad (2.124)$$

which is again an approximation for the same reason as above. With this density, the electron-nuclear as well as the Coulomb interaction can be calculated using the (exact) functionals:

$$E_{\text{ne}}[\rho] = - \sum_A^{N_{\text{nuc}}} \int \frac{Z_A \rho(\mathbf{r})}{|\mathbf{R}_A - \mathbf{r}|} d\mathbf{r}, \quad (2.125)$$

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (2.126)$$

Using these expressions, the final form of the DFT energy functional can be written as

$$E[\rho] = T_S[\rho] + E_{\text{ne}}[\rho] + J[\rho] + E_{\text{xc}}[\rho]. \quad (2.127)$$

The last term in this equation is called the exchange-correlation energy functional, which is the key in the formulation of an accurate orbital-based DFT theory. It is conceptually a correction to Hartree-Fock theory which has the potential to be computationally very simple and fast to apply. This is in contrast to more rigorous treatments of electron

correlation, such as configuration interaction or perturbation theory, which substantially increase the computational cost of the method. Unfortunately, it is not possible (so far) to derive an exact expression for  $E_{xc}$ , and it is also not obvious how to systematically improve it to get better results. Therefore, a plethora of different exchange-correlation functionals have been proposed in the past decades, which all have slightly different performances for different systems and properties.

Often,  $E_{xc}$  is divided into separate exchange and correlation contributions, and the functionals are expressed in terms of energy densities  $\varepsilon$ :

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] = \int \rho(\mathbf{r})\varepsilon_x(\rho)d\mathbf{r} + \int \rho(\mathbf{r})\varepsilon_c(\rho)d\mathbf{r}. \quad (2.128)$$

Since the correlation of electrons of same spin and opposite spin is different, it is necessary to distinguish between the separate densities  $\rho_\alpha$  and  $\rho_\beta$  for all but closed-shell cases. The simplest approach for these functionals is the local spin density approximation (LSDA), which uses functional expressions for the uniform electron gas. For the exchange energy, this functional has a rather simple form:

$$E_x^{\text{LSDA}}[\rho] = -\frac{3}{2^{5/3}} \left(\frac{3}{\pi}\right)^{1/3} \int (\rho_\alpha^{4/3}(\mathbf{r}) + \rho_\beta^{4/3}(\mathbf{r}))d\mathbf{r}. \quad (2.129)$$

For the correlation energy in the LSDA, there are only analytical expressions for high- and low-density limits. The general case can be calculated accurately with quantum Monte-Carlo methods, and there are parametric functions for the corresponding energy density  $\varepsilon_c^{\text{LSDA}}(\rho_\alpha, \rho_\beta)$  which are considered to be accurate fits. One example, called VWN,<sup>[154]</sup> has a total of eight free parameters, but as it is a very complex function we shall not show it here.

An improvement over these local density methods can be achieved by including the gradient of the density. Again, there is no obvious, analytical way how to do this, so there have been various proposals. A popular option for the exchange energy density, as proposed by Axel Becke in 1988, is the following:<sup>[155]</sup>

$$\varepsilon_x^{\text{B88}} = \varepsilon_x^{\text{LSDA}} - \beta\rho^{1/3} \frac{x^2}{1 + 6\beta x \text{arsinh } x}, \quad \text{with} \quad x = \frac{|\nabla\rho|}{\rho^{4/3}}. \quad (2.130)$$

Here,  $\beta$  is a parameter which is fitted to known data for noble gas atoms. This functional is a considerable improvement for the exchange energy, reducing the error by almost two orders of magnitude. For the correlation energy, there are also a wide variety of proposed functionals, one of them being the LYP functional.<sup>[156]</sup> Similar to the LSDA correlation it has a very complex functional form which will not be shown here. It contains four free parameters which were obtained by fitting it to data for the helium atom.

Another way of improving the performance of exchange-correlation functionals is by including exchange as defined in Hartree-Fock theory ( $E_x^{\text{HF}}$ ). These form are called hybrid functionals, and one example is the B3LYP functional which was used extensively

in this work. Its energy expression is:

$$E_{\text{xc}}^{\text{B3LYP}} = (1 - a - b)E_{\text{x}}^{\text{LSDA}} + aE_{\text{x}}^{\text{HF}} + bE_{\text{x}}^{\text{B88}} + (1 - c)E_{\text{c}}^{\text{LSDA}} + cE_{\text{c}}^{\text{LYP}}. \quad (2.131)$$

The parameters  $a$ ,  $b$ , and  $c$  are determined by fitting to experimental data, and take the values  $a = 0.20$ ,  $b = 0.72$ , and  $c = 0.81$ .

Once an exchange functional has been chosen, the Kohn-Sham problem is very similar to Hartree-Fock. It can be expressed as an eigenproblem of an effective one-electron operator  $\hat{h}_{\text{KS}}$ , analogous to the Fock operator:

$$\hat{h}_{\text{KS}} = -\frac{1}{2}\nabla^2 + \hat{V}_{\text{ne}} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \hat{V}_{\text{xc}}, \quad (2.132)$$

$$\text{with } \hat{V}_{\text{xc}} = \varepsilon_{\text{xc}}(\mathbf{r}) + \int \rho(\mathbf{r}') \frac{\delta\varepsilon(\mathbf{r}')}{\delta\rho(\mathbf{r})} d\mathbf{r}'. \quad (2.133)$$

Using this Kohn-Sham operator, the iterative solution of its pseudo-eigenproblem using an expansion of basis function is in complete analogy to the Hartree-Fock procedure.

### 2.6.3 Basis Sets

The choice of basis set is an important decision for any *ab initio* method, and it can significantly influence the accuracy and computational efficiency of the calculation. Expanding an orbital in a complete set of functions is not an approximation, but as such complete sets are infinite in size, any applicable, finite basis set will lead to truncation errors. Since *ab initio* computations all formally scale with at least the fourth power of the basis set size (compare Eq. (2.119)), keeping the number of functions as small as possible is of utmost importance for practical applicability. In addition, they should be easy to handle in the types of calculation that occur in a given method. The LCAO approach for constructing molecular orbitals is based around the idea of using functions that resemble the solutions of the hydrogen atom. Slater-Type Orbitals (STOs) are the type of basis functions that follow this idea most closely, and they have the general form

$$\chi^{\text{STO}}(r, \theta, \phi) = NY_l^m(\theta, \phi)r^{n-1}e^{-\zeta r}. \quad (2.134)$$

Here,  $N$  is a normalization constant, and the integers  $n$ ,  $l$ , and  $m$  are similar in meaning as the quantum numbers of hydrogen wave functions. The only difference to them is that STOs do not have radial nodes, but these can emerge by forming suitable linear combinations of STOs with different  $n$ . Although STOs are very similar to atomic orbitals and therefore give excellent per-function accuracy, they are computationally very unfavorable. In particular, there is no analytical expression for three- and four-center integrals, and their numerical calculation is very slow. As such, they are mostly limited in use to atomic or diatomic cases or for methods where three- and four-center integrals do not occur (such as semi-empirical calculations).

The other option for basis functions are Gaussian-Type Orbitals (GTOs), which in Cartesian coordinates have the following functional form:

$$\chi^{\text{GTO}}(x, y, z) = N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2}. \quad (2.135)$$

The fact that there is a square dependence of the distance in the exponent leads to two disadvantages compared to STOs: first, GTOs have zero slope at the site of the nucleus, unlike atomic wave functions, which have a “cusp” at zero distance. The description near the nucleus is therefore poorer. Second, the function decays too rapidly with distance, and properties depending on the tail of the wave functions will be described less accurately. However, this can be remedied by including a larger number of basis functions, or by using linear combinations of GTOs with fixed coefficients (contracted GTOs), and the computational benefit from using GTOs by far compensates for the necessary larger number of functions. The sum of the Cartesian components in a GTO corresponds to the angular momentum quantum number, i.e.,  $l_x + l_y + l_z = l$ . There is no explicit occurrence of  $n$  in GTOs, and this distinction is achieved by a different choice of  $\zeta$ . Sets of GTOs with same  $l$  are called  $s$ -,  $p$ -,  $d$ -functions, etc., again in analogy to the hydrogen atom.

A minimal basis set contains a single set of functions up to the highest  $n$  of the neutral atom. Hydrogen and helium would only get a single  $s$  function, lithium through neon would be described with two  $s$ - and a set of  $p$ -functions, etc. This is also known as a single-zeta basis set. In a double-zeta basis set all functions are doubled, so hydrogen would be modeled with two  $s$ -functions with different exponent  $\zeta$ . A triple-zeta set is generated by tripling the amount of functions. Since often the relevant parts of molecular binding happens in the valence orbitals, this multiplication is sometimes only applied to these functions, yielding a so-called split valence basis. Functions of higher angular momentum than those in a minimum basis are called polarization functions. For example,  $p$ -functions can be added to hydrogen to better model the non-spherical distribution of its electrons due to a chemical bond.

An example for a family of basis sets are the Pople-style functions (6-31G etc.). They were designed as a general-purpose basis and have been used extensively in the past. The resulting large amount of reference data is beneficial to the users, but it also hinders the transition to more modern and computationally efficient basis sets. General-purpose basis sets have the intrinsic disadvantage of having to make trade-offs to perform reasonably well in various types of applications. However, different methods and calculations have different requirements for basis sets, for example, the basis functions that are ideal for electron correlation methods differ from those designed for independent-particle methods (HF, DFT). The calculation of certain molecular properties may also benefit from adapted basis sets, and it is rather obvious that the calculation of NMR properties requires a good description at the site of the nucleus, which may be less important for chemical binding. Wherever possible, we have therefore chosen to use Jensen-type basis sets (pcseg- $n$ , pcSseg- $n$ ) in this work. They have been specifically designed to be used for density functional methods, and modified sets optimized for the calculation of NMR shielding constants or  $J$ -couplings are available.

## 2.7 Electrostatics

### 2.7.1 Theory of Electric Fields in Materials

To assess the behavior of the electric field within a cell of arbitrary geometry and composed of various materials, a basic understanding of the theory of electromagnetism is vital. In this section we will introduce these necessary basics.

Classical electromagnetism is governed by Maxwell's equations, which are presented here in their macroscopic form:<sup>[5]</sup>

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}, \quad (2.136)$$

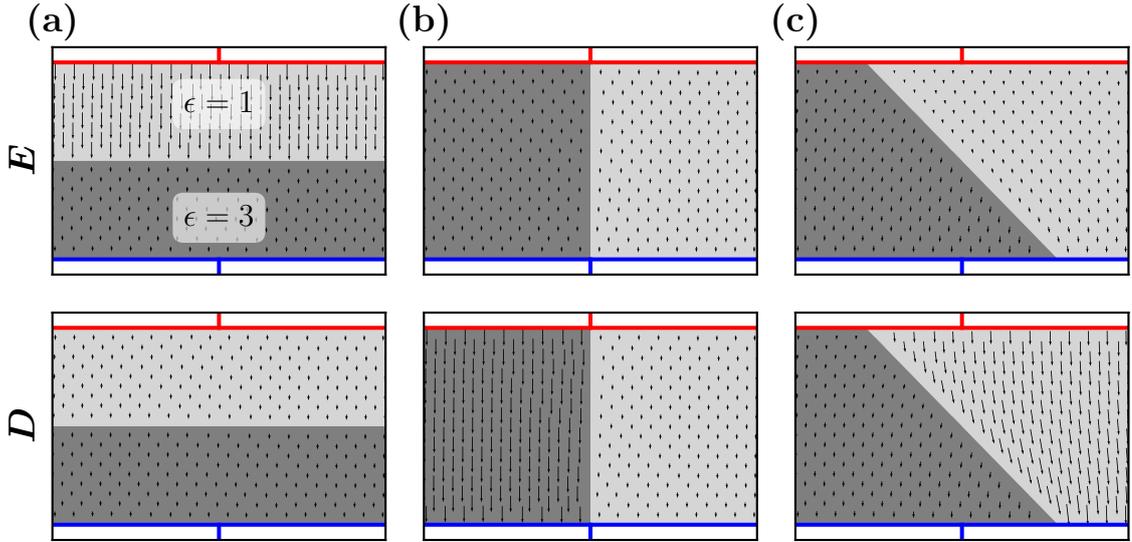
$$\nabla \times \mathbf{H} = \mathbf{j} + \dot{\mathbf{D}}, \quad (2.137)$$

$$\nabla \cdot \mathbf{D} = \rho, \quad (2.138)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (2.139)$$

$\mathbf{H}$  and  $\mathbf{B}$  are the magnetic field intensity and the magnetic flux density that may be familiar to a reader coming from magnetic resonance. The description of the electric field knows a similar distinction between the electric field strength  $\mathbf{E}$  and the electric flux density  $\mathbf{D}$ , also known as the electric displacement field. Furthermore,  $\mathbf{j}$  and  $\rho$  denote the current and charge density, respectively. The distinction between  $\mathbf{H}$  and  $\mathbf{B}$  on one hand and  $\mathbf{D}$  and  $\mathbf{E}$  on the other hand is the feature of this macroscopic form of electromagnetism, where the influence of microscopic magnetic and electric dipoles in matter are taken into account as a functional dependence between the two fields. For our system we will consider the static case, which is also applicable in the case for AC fields with low frequencies whose wave length is much larger than the system size. In this case all time derivatives vanish and equations for the magnetic and electric fields can be separated, and we need to consider only the latter of the two.

Every electric point charge can be imagined to emit an electric flux equal to its charge (both have the same unit) uniformly in all directions, and the electric flux density  $\mathbf{D}$  is simply this flux per surface area. If the  $\mathbf{D}$  is integrated over a closed surface, the result will be the total charge inside the enclosed space. This is formulated in Eq. (2.138). The electric field strength  $\mathbf{E}$  however describes the effect of this field on charges, namely the coulomb force. In the static case it has no curl ( $\nabla \times \mathbf{E} = \mathbf{0}$ ), and can therefore be described as the gradient of a potential:  $\mathbf{E} = \nabla\phi$ . In vacuum both are connected by a fundamental constant, the vacuum permittivity  $\epsilon_0$ :  $\mathbf{D} = \epsilon_0\mathbf{E}$ . In the most common case of linear materials, the permittivity is scaled by a constant factor  $\epsilon_r$ :  $\mathbf{D} = \epsilon_0\epsilon_r\mathbf{E}$ . As a consequence, both  $\mathbf{D}$  and  $\mathbf{E}$  will change at the interface of two materials with different permittivities. The laws governing this change can be explored using simple model systems. Let us consider two parallel, conducting and charged surfaces with constant distance and infinite extent. Since they are conducting, they have the same potential  $\phi$  everywhere. Between these surfaces there are two layers of different materials, and the material interface is parallel to the charged surfaces (Fig. 2.3a). As the materials are electrically neutral, no electric flux is generated within them. The electric flux that passes through the first layer necessarily needs to pass through the second layer as well;



**Figure 2.3.** Electric field strength  $E$  (top) and electric flux density  $D$  (bottom) in different model capacitor geometries and two materials with permittivities 1 (light gray) and 3 (dark gray). The material interfaces are either perpendicular (a), parallel (b), or slanted (c) with respect to the charged plates.

therefore  $D$  must be equal in both layers. We can conclude from this model that the normal component of  $D$  must be continuous at an interface. The electric field  $E$  must therefore be inversely proportional to the material's permittivity. In the second model the interface between the two materials is not parallel to the surface, but perpendicular, so in different parts of the model the surfaces are separated by different materials (Fig. 2.3b). However, as the potential difference as well as the distances are the same, the electric field strength  $E$  must therefore be the same in both cases. From this we can conclude that the tangential component of  $E$  must be continuous at an interface. The resulting difference in flux density results from a difference in charge density on the conducting surfaces in the two different regions. From this we can derive the general situation at an interface. Let  $E_{it}$  and  $E_{in}$  be the tangential and normal components, respectively, of the field strength in a material  $i$  at an interface. The angle  $\alpha_i$  shall be the angle of the field with the interface normal, so that it is  $\tan \alpha_i = E_{it}/E_{in}$ . At an interface between two materials 1 and 2 we have  $E_{1t} = E_{2t}$  and  $E_{1n}/\epsilon_1 = E_{2n}/\epsilon_2$ . From this we can construct the electric field's law of refraction, named after the analogy to the refraction of light:

$$\frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\epsilon_2}{\epsilon_1}. \quad (2.140)$$

This change of the field angle at an interface in the general case is depicted in Fig. 2.3, where there is a slanted interface between the two materials in the capacitor model.

### 2.7.2 Finite Element Analysis

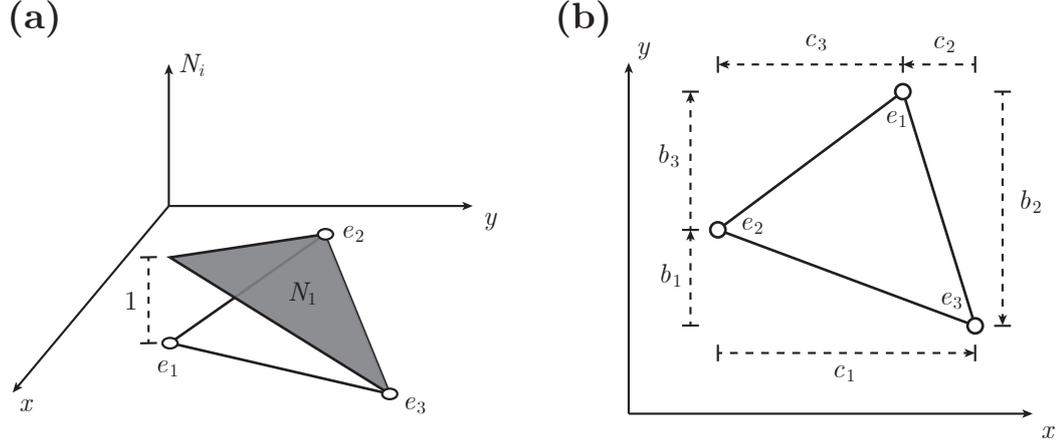
For the field profiles in complex cell geometries there are typically no analytic solutions. It is therefore necessary to use a numerical, approximate algorithm to calculate these fields. In this work we made use of finite element analysis, which is a very general numerical technique capable of solving many problems in physics and engineering such as mechanical stress and strain, heat conduction, fluid dynamics, electromagnetism, and many others. In this method the domain of the problem, i.e., a surface or a volume, is discretized into a collection of basic geometrical objects such as triangles, squares, or tetrahedrons. These objects of finite surface or volume are the eponymous elements. The physical quantities in question are modeled within each element using very simple, often linear, shape functions. One advantage of this method and our rationale in using it is the possibility of varying the size of the elements depending on the location. In regions where strong gradients are expected, e.g., close to the electrodes, the element mesh can be made very fine, while in regions of less interest and where only small gradients are expected the meshing can be done more coarsely. A detailed discussion of this technique is beyond the scope of this work, but we will explain the basic concepts of the algorithm on the example of electrostatics in two dimensions.

The problem is set by a geometry of regions with different permittivity, representing the different materials, as well as a collection of boundary conditions that can be either electric potentials or electric field strengths that are fixed in certain parts of the domain. The goal of the analysis is to find all desired electrical parameters in the unconstrained regions of the domain. The first step is to find a condition that needs to be satisfied in such a stationary solution, and express it in a suitable way. While there are different pathways to find these solutions, here we will present one based on the variational principle which may be familiar to readers from theoretical chemistry. It is based on the fact that the system will adopt a state of minimum energy. The energy is calculated by integrating the electric field's energy density  $w$  over the entire area  $A$ :

$$w = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} = \frac{1}{2} \epsilon \mathbf{E}^2 = \frac{1}{2} \epsilon (\nabla \varphi)^2, \quad (2.141)$$

$$E = \int_A \frac{1}{2} \epsilon (\nabla \varphi)^2 dA. \quad (2.142)$$

This expression needs to be in a form of a single independent variable, such as  $\varphi$  in the case of the Eq. (2.142). This form is called irreducible, meaning that it cannot be reduced to fewer independent variables. It is sufficient to find a solution for the electric potential  $\varphi$ , all other parameters such as the field can be derived from it. The domain is discretized into a triangular mesh which consists of two basic objects: the nodes of this mesh are points on the surface, and the potential values at these nodes  $\varphi_i$  will be the parameters that determine the solution. The triangular areas in between are the (finite) elements, and we will assume that the material properties, such as the permittivity, are constant within each element. Based on the potential values of each element's three nodes, we need to find a linear function that interpolates the potential within the area of the element. This is done by using so-called shape functions, which are



**Figure 2.4.** (a) Illustration of the shape function  $N_1$  within its element. It evaluates to 1 at node 1 and to 0 at nodes 2 and 3. (b) Geometrical interpretation of the linear parameters  $b_i$  and  $c_i$ .

conceptually similar (but significantly simpler) to basis functions in electronic structure theory. Within each element there are three shape functions  $N_i$ , one for each node  $i$ . These shape functions will be constructed as linear functions in two dimensions, with parameters  $a_i$ ,  $b_i$ , and  $c_i$  and the element's area  $\Delta$ :

$$N_i(x, y) = \frac{1}{2\Delta} (a_i + b_i x + c_i y) \quad (2.143)$$

$$\text{with } \nabla N_i(x, y) = \frac{1}{2\Delta} \begin{pmatrix} b_i \\ c_i \end{pmatrix}. \quad (2.144)$$

The condition that determines these parameters is that each shape function, e.g.,  $N_1$ , should have a value of 1 at its own node  $((x_1, y_1))$  and a value of 0 at the other two nodes  $((x_2, y_2)$  and  $(x_3, y_3))$ , which is illustrated in Fig. 2.4a. This results in the following system of linear equations (expressed in matrix form):

$$\frac{1}{2\Delta} \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \quad (2.145)$$

The solution to this system and the ones for the two other nodes is the following:

$$a_1 = x_2 y_3 - x_3 y_2, \quad b_1 = y_2 - y_3, \quad c_1 = x_3 - x_2, \quad (2.146)$$

$$a_2 = x_3 y_1 - x_1 y_3, \quad b_2 = y_3 - y_1, \quad c_2 = x_1 - x_3, \quad (2.147)$$

$$a_3 = x_1 y_2 - x_2 y_1, \quad b_3 = y_1 - y_2, \quad c_3 = x_2 - x_1, \quad (2.148)$$

and the area can be expressed as  $\Delta = (x_1 b_1 + x_2 b_2 + x_3 b_3)/2$ . A geometrical interpretation of the linear parameters  $b_i$  and  $c_i$  is shown in Fig. 2.4b. The interpolation of the potential

within the element  $\varphi_e$  is now executed as a linear combination of  $N_i$  with the node's potential as linear parameters:

$$\varphi_e = \sum_{i=1}^3 \varphi_i N_i, \quad (2.149)$$

$$\nabla \varphi_e = \sum_{i=1}^3 \varphi_i \nabla N_i. \quad (2.150)$$

Due to the construction of the shape functions this model function  $\varphi_e$  has the node's potential value at its position ( $\varphi_e(x_i, y_i) = \varphi_i$ ) and interpolates linearly along the edge, which ensures that this model is continuous over element boundaries. It is limited by having a discontinuous first derivative as well as a vanishing second derivative, so no second derivatives may appear in the governing equations. This model function for the potential may now be inserted into the equation for the total energy (Eq. (2.142)), and the integration over the whole domain is split up into an integration over element areas  $A_e$  and summation over elements  $e$ . As the potential within an element is modeled with a simple linear function (Eq. (2.150)), this integration is rather straightforward:

$$\begin{aligned} E &= \frac{1}{2} \sum_e \epsilon_e \int_{A_e} (\nabla \varphi_e)^2 dA \\ &= \frac{1}{2} \sum_e \epsilon_e \int_{A_e} \sum_{i=1}^3 \sum_{j=1}^3 \varphi_i^e \varphi_j^e \nabla N_i^e \nabla N_j^e dA \\ &= \frac{1}{2} \sum_e \epsilon_e \sum_{i=1}^3 \sum_{j=1}^3 \varphi_i^e \varphi_j^e \int_{A_e} \frac{1}{4\Delta_e^2} (b_i^e b_j^e + c_i^e c_j^e) dA \\ &= \frac{1}{2} \sum_e \epsilon_e \sum_{i=1}^3 \sum_{j=1}^3 \varphi_i^e \varphi_j^e \frac{1}{4\Delta_e} (b_i^e b_j^e + c_i^e c_j^e). \end{aligned} \quad (2.151)$$

This final expression for the energy is now only a function of the element's permittivities  $\epsilon_e$ , the node positions, and the node potentials. The first two are given by the model geometry, as well as a number of node potential which are the boundary conditions. It is therefore necessary to find the minimum of  $E$  as a function of all  $\varphi_i^e$ , which is done by finding the simultaneous roots of its derivatives:

$$\frac{\partial E}{\partial \varphi_i^e} = \sum_{\{e|i \in e\}} \epsilon_e \sum_{j=1}^3 \varphi_j^e \frac{1}{4\Delta_e} (b_i^e b_j^e + c_i^e c_j^e) \quad (2.152)$$

Note that since  $i$  and  $j$  are local indexes we only sum over elements  $e$  that contain  $i$  as a node. This is now a linear equation with independent variables  $\varphi_j^e$ , which are the potentials of the nodes that share an element with the node  $e, i$ , including itself ( $\varphi_i^e$ ). The double indexing  $e, i$  is ambiguous as the same node may belong to various elements, so it is best to switch to a global indexing of all nodes (denoted  $k$  or  $l$ ). The derivative

with respect to a node potential in a global indexing  $\varphi_k$  has the following form.

$$\frac{\partial E}{\partial \varphi_k} = \sum_l \varphi_l \left( \sum_{\{e|k,l \in e\}} \epsilon_e \frac{1}{4\Delta_e} (b_k b_l + c_k c_l) \right) \stackrel{!}{=} 0. \quad (2.153)$$

The inner summation is over all elements that contain both nodes  $k$  and  $l$ , which, if  $k \neq l$ , are those two that share the edge between  $k$  and  $l$ . For the diagonal elements  $k = l$  the summation is over all elements that contain this node, typically between four and eight in triangular meshes. This system of linear equations may be expressed as a simple matrix equation with the parameter matrix  $\mathbf{H}$  and the solution vector  $\boldsymbol{\varphi}$ , which is a vector of all node potentials  $\varphi_k$ . It has dimensions of  $n \times n$ , where  $n$  is the total number of nodes in the mesh:

$$\mathbf{H}\boldsymbol{\varphi} = \mathbf{0}, \quad \text{with} \quad H_{kl} = \sum_{\{e|k,l \in e\}} \epsilon_e \frac{1}{4\Delta_e} (b_k b_l + c_k c_l). \quad (2.154)$$

For unconstrained systems the trivial solution to Eq. (2.154) is  $\boldsymbol{\varphi} = \mathbf{0}$ , which corresponds to a vanishing potential and field everywhere and is a rather uninspiring result. If some elements of  $\boldsymbol{\varphi}$  are however fixed to a certain value (e.g., at the edges of electrodes), the solution has to be modified. Let  $\bar{\boldsymbol{\varphi}}$  be the vector of potential boundary conditions where all unconstrained potentials are zero. A single index  $u$  ( $\mathbf{H}_u, \boldsymbol{\varphi}_u$ ) indicates that the constrained rows have been removed from the vector or matrix, and a double index  $u$  ( $\mathbf{H}_{uu}$ ) indicates that both the constrained rows and columns have been removed. The constrained problem can then be formulated as follows:

$$\mathbf{H}_{uu}\boldsymbol{\varphi}_u = -\mathbf{H}_u\bar{\boldsymbol{\varphi}}. \quad (2.155)$$

It is a linear equation of the unconstrained potentials  $\boldsymbol{\varphi}_u$  and may be solved using standard algorithms from linear algebra.

This section aimed at illustrating the general concept of finite element analysis. Aspects that were not covered include, for example, boundary conditions on the field instead of the potential, which require a modification of the energy term in Eq. (2.142), higher dimensions or different forms of elements (tetragons, tetrahedrons, trigonal prisms etc.), higher order shape functions, time dependent problems, and many more.

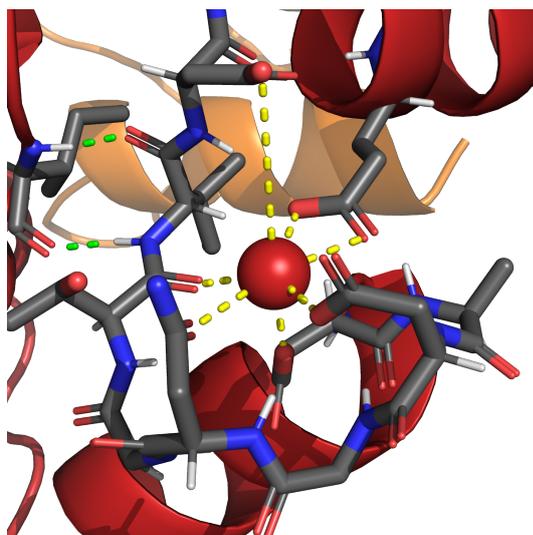
## Chapter 3

# Interdomain Dynamics of the Complex Calmodulin/Munc13-1

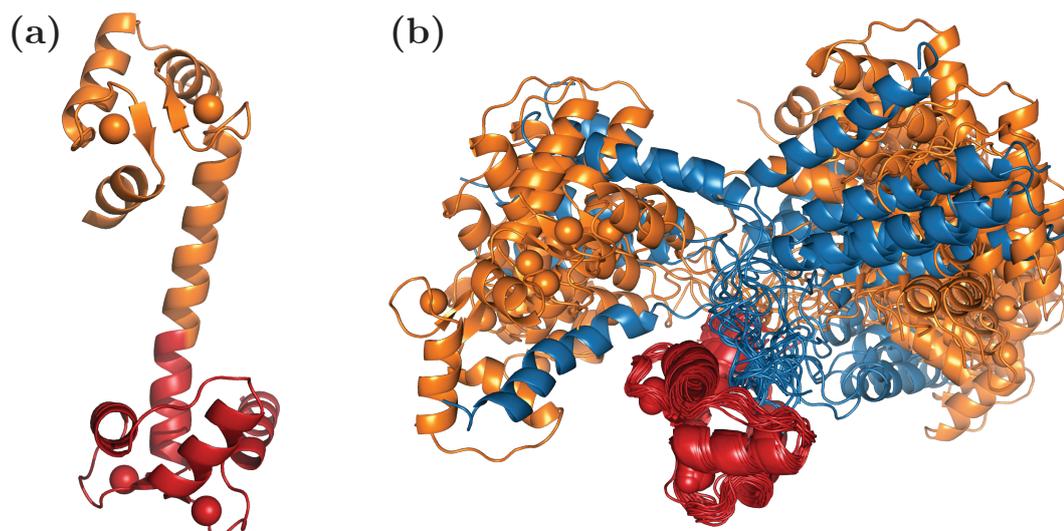
In this Chapter we will describe how the interdomain dynamics of the complex calmodulin/Munc13-1 was investigated by means of paramagnetic NMR. The expression and purification of N60D-calmodulin was performed by Melanie Wegstroth, Karin Giller, and Stefan Becker. Peptide synthesis of the Munc13-1 fragment was done by Kerstin Overkamp.

### 3.1 Introduction

Calmodulin (CaM) is a 148-amino acid regulatory protein that can be found in all eukaryotic cells, and it belongs to the family of EF-hand proteins.<sup>[158]</sup> The EF-hand motif is a helix-loop-helix structural element that forms a pocket for  $\text{Ca}^{2+}$  with a pentagonal-bipyramidal binding geometry. In calmodulin, the loops of two such motifs are held together by a short stretch of antiparallel  $\beta$ -sheet, forming a rigid, globular domain. The binding pocket as well as the  $\beta$ -sheet are shown in Fig. 3.1. Two of such domains, containing approximately 70 residues each, are connected by a linker. The binding pockets give the protein the ability to sense intracellular  $\text{Ca}^{2+}$  fluctuations and to activate downstream signaling pathways upon binding to it, as explained in the following. The two domains of calmodulin undergo a structural rearrangement when binding  $\text{Ca}^{2+}$ , which exposes a hydrophobic cleft. With this cleft calmodulin is able to bind to different proteins, acting either as an inhibitor or activator, depending on the specific target. The apparent dissoci-



**Figure 3.1.** Illustration of the EF-hand  $\text{Ca}^{2+}$ -binding pocket in calmodulin, with the pentagonal bipyramidal coordination shown in yellow. The hydrogen bonds of the antiparallel  $\beta$ -sheet that bind both EF-hand motifs to a globular domain are shown in green in the top left of the image. The geometry is taken from 2BE6/B.<sup>[157]</sup>



**Figure 3.2.** (a) Early crystal structure of calmodulin (PDB code 3CLN<sup>[168]</sup>), showing the overall dumbbell shape and the continuous  $\alpha$ -helix connecting the two domains. (b) NMR ensemble of CaM/Munc13-1 (PDB code 2KDU<sup>[169]</sup>), aligned onto the N-terminal domain. Munc13-1 is shown in blue.

ation constant of the N- and C-terminal domain are approximately  $10\ \mu\text{M}$  and  $1\ \mu\text{M}$ , respectively,<sup>[159,160]</sup> so changes in  $\text{Ca}^{2+}$  concentration within these limits are interpreted as signals and induce conformational rearrangements in calmodulin, leading to target activation or inhibition upon binding.<sup>[161–167]</sup>

The rigidity of the linker that connects calmodulin's two domains has been the subject of various studies. The earliest X-ray structure of calmodulin implied that the protein has a rather rigid dumbbell shape with the linker forming a continuous  $\alpha$ -helix connecting the two domains (Fig. 3.2a),<sup>[168,170]</sup> and small-angle X-ray scattering was used to get first insights into the domain's rearrangement upon calcium binding.<sup>[171,172]</sup> NMR studies later revealed that the linker region of  $\text{Ca}_4$ -CaM was highly dynamic and allowed for relative motion of the two domains with respect to each other,<sup>[173,174]</sup> an X-ray study found disorder in the linker residues,<sup>[175]</sup> and a structure of the protein crystallized in a closed conformation was published.<sup>[176]</sup> This led to the proposal that calmodulin may sample a wide, continuous range of relative domain orientations in solution, which was supported by studies of molecular dynamics.<sup>[177,178]</sup>

Bertini et al. paved the way for paramagnetic studies of calmodulin by screening a variety of mutants in the binding sites. They discovered that an N60D mutant allowed them to selectively load the second binding site (residues 56–67) with a lanthanide,<sup>[124]</sup> and proceeded to study the conformational space sampled by calmodulin using pseudocontact shifts and residual dipolar couplings as observables.<sup>[122]</sup> Their findings confirm the wide range of possible domain orientations that were proposed earlier, and they composed three- to four-membered ensembles to fit their data. In these ensembles, the C-terminal domain spans a region of space approximately in the form of a cone with a semiangle

of 40° relative to the N-terminal domain. They also determined an order parameter to quantify the flexibility, defined here as the ratio of alignment (RDCs) between the C- and N-terminal domain, for which they found a value between 0.05 and 0.15.

When binding to target peptides and proteins, calmodulin typically binds to a short section of amphiphilic  $\alpha$ -helix within the target with both its domains, which forces it into a more rigid state not dissimilar to the closed crystal structure mentioned earlier.<sup>[179,180]</sup> There are many examples of this in the literature,<sup>[181–186]</sup> one of them being the complex with the IQ recognition motif from the voltage-gated calcium channel Ca<sub>v</sub>1.2 (CaM/IQ).<sup>[157]</sup> Russo et al. applied paramagnetic NMR to study the conformational ensemble of this complex in solution, and indeed found significantly reduced interdomain dynamics with an order parameter of around 0.9.<sup>[123]</sup>

Munc13-1 is a large multi-domain protein involved in synaptic vesicle fusion and neurotransmitter release,<sup>[187–189]</sup> and its activity is highly dependent on calcium concentration due to its regulation by calmodulin.<sup>[190]</sup> Unlike many other targets of calmodulin, Munc13-1 features a unique 1-5-8-26 recognition motif consisting of two independent sections binding to the N- and C-terminal domain.<sup>[169,190]</sup> It is therefore reasonable to assume that through their interaction it restricts the conformational freedom of calmodulin far less than any of the other targets. In the work of Rodríguez-Castañeda et al. an NMR structure of the complex was determined (Fig. 3.2b), but since no long-range constraints were used for this ensemble, it does not properly describe the interdomain dynamics present in the complex. In this chapter we will explore the conformational space sampled by calmodulin using paramagnetic NMR, and compile an ensemble that is in agreement with the acquired constraints. From a methodological point of view we will maximize the amount of gathered paramagnetic data, and we will show innovative ways of reducing the structural information of conformations to its essentials, of sampling this conformational space, and of finding an appropriate ensemble size which is neither too big nor too small.

## 3.2 Complex Preparation and NMR Spectroscopy

### 3.2.1 Sample Preparation

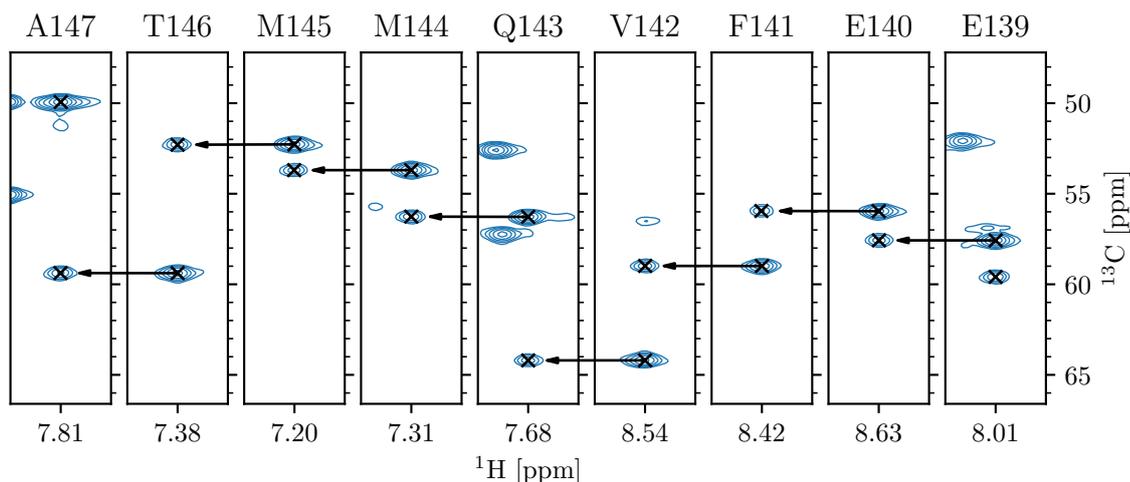
The preparation of uniformly <sup>15</sup>N, <sup>13</sup>C-labeled N60D-calmodulin followed largely established procedures.<sup>[191,192]</sup> It is however important to ensure that any remaining ethylenediaminetetraacetate (EDTA) from the protein purification procedure is eliminated before preparing the complex. EDTA is a strong chelating agent and is able to remove calcium ions from the binding pockets of calmodulin, but since only calcium-bound calmodulin (Ca<sub>4</sub>-CaM) binds strongly to Munc13-1,<sup>[169]</sup> full calcium saturation needs to be ensured before proceeding to prepare the complex. This is best done by means of NMR: apo-CaM and Ca<sub>4</sub>-CaM are easily distinguishable by their <sup>15</sup>N-HSQC spectrum, and EDTA can be identified in the proton spectrum (see Fig. A.1). To remove EDTA, an additional dialysis step against an EDTA-free buffer was added to the existing purification protocol, and afterwards CaCl<sub>2</sub> was added until no further changes in the HSQC spectrum could be

observed. Munc13-1 was prepared by solid-phase synthesis and lyophilized. As it is not soluble by itself in the required concentrations of  $> 1$  mM, the complex was prepared by adding a solution of CaM to lyophilized Munc13-1. This has the problem that it is very difficult to control the amount of added peptide, as the quantities are too small to be reliably controlled gravimetrically. As an additional complication, there is evidence that CaM can weakly bind to a second equivalent of Munc13-1, so care has to be taken to only add a single equivalent of Munc13-1.<sup>[193]</sup> The process of complex formation can again be easily monitored by NMR spectroscopy. The binding of the first equivalent of Munc13-1 occurs in the slow exchange regime, so in the case of incomplete saturation, two sets of peaks of CaM and CaM/Munc13-1 can be observed. The binding of the second equivalent of Munc13-1 happens on the intermediate exchange regime, so in excess of Munc13-1 certain affected peaks will shift and eventually broaden beyond detection. To prevent dissolving an excess of Munc13-1, which is very difficult to reverse, we simply saturated only half of the protein solution by adding an excess of solid Munc13-1, removed the remaining peptide, and combined it again with the protein stock. This generated a solution with too little peptide, as could be easily determined by NMR. By repeating this process multiple times, saturating increasingly smaller fractions of the total volume, and acquiring an NMR spectrum after each step, the stoichiometry of the complex could be controlled very precisely. The resulting solution of 1.85 mL had a concentration of approximately 0.9 mM.

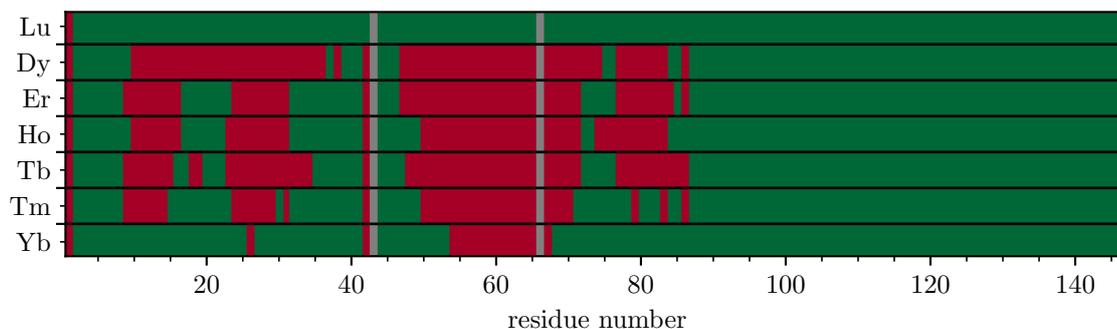
To maximize the NMR signal in all further steps of the project, this was concentrated further to a final volume of 1.12 mL and a concentration of 1.43 mM as determined via NMR.<sup>[194]</sup> All lanthanide-substituted samples were prepared from this stock solution to ensure that they have exactly the same complex conditions. The mutation in N60D-CaM achieves about a 10-fold increase in the affinity of lanthanides (Ln) to the second metal binding site, so upon addition of lanthanide the calcium ion in this binding site gets substituted preferentially.<sup>[124]</sup> This was done by titrating a 30 mM solution of  $\text{LnCl}_3$  to the complex solution while monitoring via NMR. The lanthanide binding occurs on the slow exchange regime, so two sets of peaks can be observed in the HSQC spectrum during the titration. Care must be taken not to add excess lanthanide as it will substitute calcium in the other binding pockets as well as soon as the second site is saturated. This way six paramagnetic samples with the lanthanides dysprosium (Dy), erbium (Er), holmium (Ho), terbium (Tb), thulium (Tm), and ytterbium (Yb) were prepared, as well as a diamagnetic reference sample with lutetium (Lu).

### 3.2.2 Assignment

The paramagnetic lanthanide ions induce strong chemical shift perturbations in the vicinity of the binding site, and a number of resonances in the immediate surrounding of the lanthanide are broadened beyond detection by paramagnetic relaxation enhancement (PRE). Even for the diamagnetic reference with lutetium some chemical shift changes relative to  $\text{Ca}_4\text{-CaM/Munc13-1}$  were observed due to the different charge and ionic radius of  $\text{Lu}^{3+}$  and  $\text{Ca}^{2+}$ . Due to these shift perturbations specific to each lanthanide, separate resonance assignment had to be done for the different samples. This was done following



**Figure 3.3.** Illustration of an HNCA-based backbone walk on the example of ten residues of  $\text{Ca}_3\text{Yb-CaM/Munc13-1}$ . Each slice also differs in  $^{15}\text{N}$  chemical shift, which is not depicted here.



**Figure 3.4.** Available (green) and missing (red) amide resonance assignment for the different metal complexes  $\text{Ca}_3\text{Ln-CaM/Munc13-1}$ . The gray lines indicate the position of the two proline residues.

established methods using triple-resonance experiments, and we acquired HNCA and HNCOC spectra of each sample. An HNCA is a three-dimensional NMR spectrum that correlates the amide proton and the amide nitrogen resonance with the  $\text{C}_\alpha$  resonance of the same ( $i$ ) and of the preceding ( $i - 1$ ) residue. The spectrum will therefore feature two peaks for each amino acid in the protein. Since the intraresidual  $J$ -coupling of the amide nitrogen to its own  $\text{C}_\alpha^i$  is slightly larger than the interresidual coupling to the preceding  $\text{C}_\alpha^{i-1}$ , the two corresponding peaks can typically be distinguished by their signal intensity. Therefore, the peaks from two neighboring amino acids will differ in their HN-resonance, but share a peak in the carbon dimension, and the intensity of these two peaks will indicate which amino acid comes first in the sequence. If this is done for each pair of neighboring amino acids, the whole sequence can be assigned. This so-called backbone walk is illustrated in Fig. 3.3. Due to the finite resolution of NMR spectra a *de novo* assignment using only an HNCA is typically prevented by different  $\text{C}_\alpha$  having

very similar resonance frequencies, and more sophisticated experiments are necessary. However, since an assignment of Ca<sub>4</sub>-CaM/Munc13-1 was available in literature<sup>[192]</sup> and the chemical shift changes of Ca<sub>3</sub>Lu-CaM/Munc13-1 were quite small due to the lack of PCSs, the HNCA was sufficient to assign the entire protein backbone. The HNCO spectrum, which correlates the amide proton and nitrogen resonances with the preceding ( $i - 1$ ) carbonyl C' resonance, does not provide sequential information, but could be used to cross-check the proposed assignment with the data from the literature. Based on the assignment of Ca<sub>3</sub>Lu-CaM/Munc13-1 we assigned the paramagnetic derivatives. While the differences in chemical shift (i.e., the PCSs) could be substantial close to the metal center, we could make use of the fact that the PCS varies slowly in space. That means that resonances of atoms that are close in space (e.g., belong to the same residue) will have similar PCSs. Although there are exceptions to this rule particularly in the spatial regions where the PCS has a zero-crossing, this trend greatly simplified the resonance assignment of the paramagnetic complexes. Since some parts of the sequence close to the metal were broadened beyond detection, the assignment of all paramagnetic complexes was incomplete. Fig. 3.4 illustrates which residues could be observed and assigned in which complex, and the complete assignment can be found in the data collection (Appendix A.1).

### 3.2.3 Acquisition of Paramagnetic Data

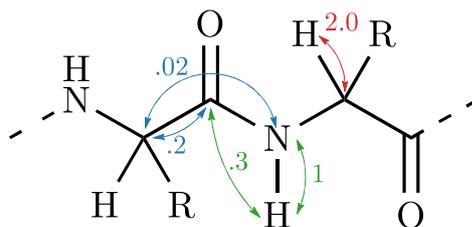
It was our goal to acquire as many paramagnetic constraints as possible, within reasonable effort, to maximize the amount of available data. PCSs are determined by simply taking the chemical shift difference between the paramagnetic sample and the diamagnetic reference (Lu). Instead of limiting ourselves to amide H and N resonances, as it was done in previous work,<sup>[123]</sup> we determined PCSs for all resonances that were available from the assignment process, namely amide-H, amide-N, C<sub>α</sub>, and C'. RDCs are determined in a similar way. In the spectrum of the diamagnetic reference only  $J$ -couplings occur, while in the paramagnetic samples a combined coupling  $T = J + D$  is observed. The RDCs ( $D$ ) are then determined as the difference between the two. The simplest RDCs to measure are the amide-HN-RDCs. They can be extracted from an uncoupled HSQC, or, more commonly, from an IPAP-HSQC where the two doublet components are separated into two spectra, but both approaches suffer from peak overlap in the crowded central region of the HN-plane. As most RDCs are relatively small effects in the order of single-digit Hz, their measurement requires very accurate peak positions, and any overlap essentially disqualifies a peak from being used for an RDC.

To minimize this issue we determined all RDCs from 3D NMR spectra. This has the additional benefit that multiple RDCs can potentially be determined from one experiment. For example, we acquired an HNCO{no H} spectrum without proton decoupling during  $t_1$  and  $t_2$ . As a consequence, each resonance is split into a doublet. Since there is no proton decoupling in neither of the indirect evolution periods, these two peaks are separated in the <sup>15</sup>N dimension by  $T_{\text{HN}}$  and in the <sup>13</sup>C dimension by  $T_{\text{HC}'}$ . The second experiment we acquired was an HNCO{no C<sub>α</sub>} without C<sub>α</sub> decoupling. This gives  $T_{\text{C}'\text{C}_\alpha}$  in the carbon dimension and  $T_{\text{NC}_\alpha}$  in the nitrogen dimension. Since  $D_{\text{NC}_\alpha}$  is very small due to the

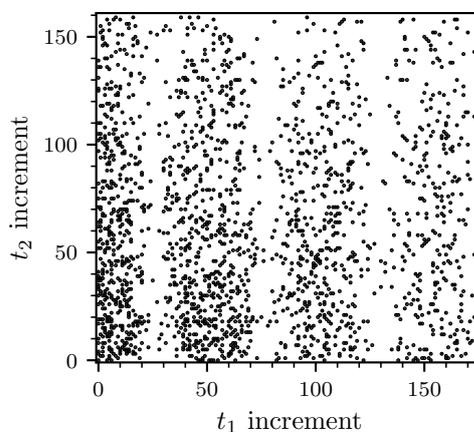
relatively large distance and the low gyromagnetic ratios of the nuclei involved, we did not use them as RDCs. As a third experiment we acquired an HNCA{no H} without proton decoupling in the  $t_1$  evolution time, which yielded  $T_{C_\alpha H_\alpha}$  in the carbon dimension. We did not remove proton decoupling during  $t_2$  as this would introduce the unwanted  $T_{NH}$  coupling in the nitrogen dimension. RDCs in paramagnetic samples are caused by magnetic alignment, so they scale with the square of the magnetic field. Therefore we sought to measure them at as high fields as possible. The HNCOs without H-decoupling (for  $T_{HN}$  and  $T_{HC'}$ ) were acquired at 950 MHz, the other two sets of experiments at 900 MHz. In Fig. 3.5 an overview of the acquired RDCs and their relative sizes is given.

All HNCA and HNCO experiments mentioned above were acquired in a way to maximize accuracy in peak position. This was done by adjusting the acquisition time in all dimensions such that the signal has fully decayed, which corresponds to acquisition times of about 45–60 ms for H, N, and  $C'$ . The lifetime of  $C_\alpha$ -magnetization is much shorter due to the dipolar coupling to the directly attached protons as well as the 35 Hz- $J$ -coupling to  $C_\beta$ , so about 12 ms of acquisition time were sufficient for this nucleus. Most modern triple resonance sequences combine the magnetization transfer of  $C_\alpha/C'$  to N with the chemical shift labeling of N into a single block of constant time,<sup>[195]</sup> which reduces losses due to relaxation but limits the acquisition time to the transfer time of  $< (2J_{CN})^{-1}$ . To be free of this limitation, all our triple resonance experiments used semi-constant time blocks.<sup>[196,197]</sup> These are a clever modification of the constant time block which removes the constraint on the acquisition time  $t_{aq}$ . For the case of  $t_{aq} < (2J_{CN})^{-1}$  it is identical to the constant time block. If the acquisition time  $t_{aq}$  is greater than  $(2J_{CN})^{-1}$ , the block length gets slowly incremented from  $(2J_{CN})^{-1}$  at the beginning to  $t_{aq}$  at the end, and the  $180^\circ$ -pulses within are shifted in a way that the evolution of the chemical shift increases from 0 to  $t_{aq}$  and the evolution of the coupling remains constant at  $(2J_{CN})^{-1}$ . This is best understood by inspecting a pulse sequence such as the one in Fig. A.2.

By increasing the acquisition times in the indirect dimensions, the number of points to be sampled inevitably increases as well, and the number of transients for each sample



**Figure 3.5.** RDCs and their size relative to  $D_{NH}$  acquired for CaM/Munc13-1. The coloring groups the couplings acquired in a single experiment. Due to the small size  $D_{NC_\alpha}$  were not used as RDC constraints.



**Figure 3.6.** NUS sampling grid for an HNCO without  $C_\alpha$ -decoupling. The density of points decreases exponentially with larger  $t_{1,2}$  times. Additionally, the point density follows a 52.5 Hz-modulation in  $t_1$  to match the  $T_{C'C_\alpha}$ -coupling.

**Table 3.1.** Number of paramagnetic constraints in CaM/Munc13-1, separated by lanthanide, by type (PCS/RDC), and by the protein domain.

		Dy	Er	Ho	Tb	Tm	Yb	sum	
N-terminus	PCS	57	103	117	90	129	199	695	1026
	RDC	13	42	58	18	55	145	331	
C-terminus	PCS	237	233	241	230	238	244	1423	2691
	RDC	213	200	213	209	209	224	1268	

should not be reduced below a certain number (typically 8) due to the phase cycle. In this case one runs into the so-called sampling limit, which describes the case where the minimum experimental time is not restricted by the signal-to-noise ratio, but by the number of samples that have to be acquired. This limit is about 3 and 7 days for HNCA- and HNC0-based experiments, respectively, which is prohibitive given that multiple experiments need to be acquired for multiple samples. To cope with this, we employed non-uniform sampling (NUS), which is a technique in which only a fraction of the samples are acquired in a non-regular pattern.<sup>[52]</sup> By acquiring more samples in regions of the grid where more signal is expected, i.e., at shorter evolution times and at the maxima of cosine-modulations, this approach can even improve the signal-to-noise ratio achieved in a given amount of time (see Fig. 3.6). The drawback of NUS is that before Fourier transform, the missing samples need to be reconstructed using complex and computationally expensive algorithms. Nonetheless, as the necessary computation time is typically orders of magnitude less expensive than high-field NMR time, it is a valuable tool to acquire high resolution 3D NMR spectra.

The amount of paramagnetic constraints that could be extracted from this collection of spectra is summarized in Table 3.1. It is evident that far fewer data could be collected in the N-terminal domain than in the C-terminal domain. This is expected as the paramagnetic lanthanide is located within the N-terminal domain, and many nearby resonances are broadened beyond detection (compare also Fig. 3.4). Metals with larger susceptibilities (e.g., Dy) provide fewer constraints than lanthanides with a smaller susceptibility (e.g., Yb) for the same reason. This does not make them less valuable, however, since the effects are also larger and therefore less affected by noise, providing better constraints. It is also worth noting that the aim of this work is to analyze the protein's interdomain motion, and the paramagnetic effects that report on this motion are the ones located in C-terminal domain (i.e., in a different domain than the lanthanide).

## 3.3 Data Analysis and Modeling

### 3.3.1 Fitting of N-Terminal Data

The first step in data analysis is the fitting of the N-terminal data to determine the susceptibility tensors  $\Delta\chi$ . As the N-terminal domain has little to no motion this is also well-suited to try different fitting approaches and to assess the data quality. The link between the paramagnetic data, the structure, and the susceptibility tensor is found in Eqs. (2.29), (2.53) and (2.73) and is repeated for the reader below:

$$\Delta\delta_{\text{PCS}} = \frac{1}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{\mathbf{r}}_{\text{Ln}} \hat{\mathbf{r}}_{\text{Ln}}^{\text{T}} \Delta\chi \right), \quad (3.1)$$

$$\Delta\nu_{\text{RDC}} = -\frac{3\mu_0\gamma_1\gamma_2\hbar}{8\pi^2 r_{12}^3} \text{tr} \left( \hat{\mathbf{r}}_{12} \hat{\mathbf{r}}_{12}^{\text{T}} \mathbf{A} \right) = -\frac{B_0^2\gamma_1\gamma_2\hbar}{40\pi^2 k_{\text{B}} T r_{12}^3} \text{tr} \left( \hat{\mathbf{r}}_{12} \hat{\mathbf{r}}_{12}^{\text{T}} \Delta\chi \right). \quad (3.2)$$

With the geometrical parameters  $\mathbf{r}_{\text{Ln}}$  and  $\mathbf{r}_{12}$  from a structural model and the paramagnetic data set, these can be expressed as a system of linear equations with the five independent components of  $\Delta\chi$  as unknowns in an equivalent manner as shown in Eqs. (4.11) and (4.12). These five components are  $\Delta\chi_{\text{ax}} = 2\Delta\chi_{zz} - \Delta\chi_{xx} - \Delta\chi_y$ ,  $\Delta\chi_{\text{rh}} = \Delta\chi_{xx} - \Delta\chi_y$ ,  $\Delta\chi_{xy}$ ,  $\Delta\chi_{xz}$ , and  $\Delta\chi_{yz}$ . The least-squares solution for these components can be found easily and deterministically, and the agreement of the data with the structural model is assessed by a  $Q$ -Factor (see also Eq. (4.13)):

$$Q = \sqrt{\frac{\sum_i (x_{\text{exp},i} - x_{\text{calc},i})^2}{\sum_i x_{\text{exp},i}^2}}. \quad (3.3)$$

Here  $x$  refers to any type of data point (RDC/PCS) and the index indicates whether it is experimental or back-calculated using the fitted susceptibility tensor. We will use subscripts on the  $Q$ -factor to indicate when only a certain type of data was considered ( $Q_{\text{PCS}}$ ,  $Q_{\text{RDC}}$ ), and superscripts to differentiate between the two domains ( $Q^{\text{N}}$ ,  $Q^{\text{C}}$ ). In this analysis there are still some finer aspects to consider, namely the choice of the structural model, the simultaneous evaluation of PCSs and RDCs, the optimization of the lanthanide position, and the inclusion of RCSAs. We will discuss each of these aspects in the following. Whenever we report an improvement in  $Q$ -factor due to any of these points, this was calculated with the other three aspects already optimized.

An accurate structural model is crucial for the successful fitting of data. While there is an NMR structure of the complex CaM/Munc13-1, deposited in the protein data bank (PDB)<sup>[198]</sup> with code 2KDU,<sup>[169]</sup> it provides a subpar model for the N-terminal domain of calmodulin, which is reflected in the comparatively high  $Q^{\text{N}}$ -factor when fitting the data. We fitted our data to the 20 different members of the 2KDU-ensemble individually, which yielded a set of  $Q^{\text{N}}$ -factors with a minimum of 0.1361, a mean of 0.1758, and an RMSD of 0.0214. There is no crystal structure of CaM/Munc13-1, but if one assumes that the structure of the individual domains is not significantly affected by the type of binding partner, one can also use the structure of a different complex. We chose to

use the structure of the complex CaM/IQ with PDB code 2BE6,<sup>[157]</sup> which had also been studied by paramagnetic NMR.<sup>[123]</sup> This X-ray ensemble contains three different structures (A, B, and C), and any of them yielded much lower  $Q^N$ -factors than the NMR structures (0.0615, 0.0461, and 0.0690, respectively). We also had the idea that the domain in solution could be better modeled by a combination of the three X-ray structures with equal contributions, but this fit yielded a  $Q^N$  of 0.0557, rejecting this hypothesis. We therefore chose to use the N-terminal domain structure of 2BE6/B in all further analyses. Another point concerning the structural model is the fact that small-scale local motion such as bond libration reduce the observed RDC and therefore leads to an underestimation of the tensor size when fitting to a rigid model. A simple way to cope with this is to use larger effective bond lengths. For the most affected RDCs, which are  $D_{\text{NH}}$  and  $D_{\text{C}_\alpha\text{H}_\alpha}$ , we only took the corresponding bond orientation from the structure and fixed the length to 1.041 Å and 1.117 Å, respectively, as described by Ottiger et. al.<sup>[199]</sup>

Eq. (3.3) hides the problem of simultaneously evaluating RDCs and PCSs, which inherently come in different units. There are different ways to overcome this problem, for example by converting both into frequency units (similar to how it was done in Chapter 4 for gymnochrome G), or by scaling them by the maximum observed experimental value.<sup>[123]</sup> However, we believe it is best to scale different types of data with the associated statistical error (as for strychnine in Chapter 4) to minimize the influence of this scatter, which is a very general way of scaling any type of data for simultaneous evaluation. To do so, we devised an iterative approach to determine the relative scaling factor of RDCs and PCSs. First, we fitted RDCs and PCSs separately to the structural model, and estimated the associated error by the RMSD of experimental against back-calculated data. We then used these RMSDs as scaling factors to simultaneously fit PCSs and RDCs to the model. From this simultaneous fit, we again determined the RMSDs for PCSs and RDCs, and used them as scaling factors for a new simultaneous fit. This cycle was repeated until the RMSDs did not change any further, which was typically achieved after 3–5 iterations. The RMSDs used for scaling converged to 44 ppb for PCSs and 4.6 Hz for RDCs, yielding a relative scaling factor of 105 Hz ppm<sup>-1</sup>. The scaled data was then also used for the calculation of  $Q$ -factors (Eq. (3.3)).

The X-ray structures used for the fit were acquired for a purely calcium-loaded form of the complex (in our case, Ca<sub>4</sub>-CaM/IQ), while the paramagnetic data was acquired from a complex where one calcium atom was replaced by a lanthanide ion. For the fit on PCSs, the vector between the lanthanide and the nucleus in question is the relevant geometrical parameter, and in the simplest case this is determined by simply assuming that the lanthanide takes the exact same position as the calcium ion in the structural model. However, this may not be entirely accurate as the binding pocket can adopt a slightly different geometry when binding an ion with different charge and ionic radius. An inaccurate lanthanide position most strongly affects the PCSs from nuclei close to it, and results in a poorer fit for these data points. We therefore chose to optimize the lanthanide position with the criterion of minimizing the  $Q_{\text{PCS}}^N$ -factor. This resulted in an improvement in  $Q_{\text{PCS}}^N$  from 0.0514 to 0.0375 with a change in lanthanide position of 0.60 Å.

Although the chemical shift perturbation caused by the paramagnetic lanthanide is dominated by the PCS, there is a small contribution from alignment in the form of an RCSA as well. Both the PCS and the RCSA depend (apart from the geometry) only on the susceptibility tensor  $\Delta\chi$ , and therefore the inclusion of RCSA comes essentially “for free”, in the sense that it does not add additional fit parameters (unlike, e.g., optimizing the lanthanide position). The expression for the chemical shift perturbation is adapted to be

$$\Delta\delta = \Delta\delta_{\text{PCS}} + \Delta\delta_{\text{RCSA}} = \frac{1}{4\pi r_{\text{Ln}}^3} \text{tr} \left( \hat{r}_{\text{Ln}} \hat{r}_{\text{Ln}}^T \Delta\chi \right) + \frac{B_0^2}{15\mu_0 k_B T} \text{tr} (\delta \Delta\chi). \quad (3.4)$$

If this expression is expanded into the aforementioned system of linear equations, the inclusion of RCSAs therefore simply corresponds to a small additional contribution to the linear coefficient matrix. To do this in practice, the chemical shift tensor  $\delta$  is necessary. We chose to include RCSAs only for carbonyl carbon and amide nitrogen nuclei, as these exhibit the largest chemical shift anisotropy. The CSA eigenvalues and the orientation of the eigenvectors within the peptide plane, which we took as local reference frame for each nucleus, were taken from Loth et al.<sup>[200]</sup> This yielded an improvement in  $Q_{\text{PCS}}^{\text{N}}$ -factor from 0.0467 to 0.0383. The root mean square (RMS) of the back-calculated PCS contribution is 3.515 ppm, while it is 0.058 ppm for the RCSAs. This effect is expected to be even more relevant in the case of the C-terminal domain, since the RCSA does not scale with distance like the PCS. Indeed, if one takes the absolute best ensemble determined in Section 3.3.4 and predicts the contributions to the chemical shift perturbation, one finds that the RCSAs contribute a remarkable 20% to the overall chemical shift perturbation in the C-terminal domain. We will continue referring to the chemical shift perturbations as PCSs, but from here on this will always imply that the RCSAs have been included as well.

Taking into account all these previous points, we fitted the N-terminal paramagnetic data to yield the susceptibility tensors  $\Delta\chi$  of the six lanthanide ions. The full tensor has an orientation in space and is therefore dependent on the coordinate system. All structures of CaM/Munc13-1 used in this work were aligned with the backbone of the N-terminal domain to the crystal structure 2BE6/B, which served as a reference frame in this way. The uncertainty of these tensor elements were determined via 1000 steps of bootstrap resampling of the data.<sup>[201]</sup> Briefly, in a bootstrapping analysis points from the original data set are drawn with replacement to yield new, synthetic data sets with the same number of points as the original one. They are then subjected to the same fitting as the original data set. This way, a statistic of the resulting values (such as  $\Delta\chi$ ) can be done. The uncertainty in the tensor components is calculated as the RMSD of all bootstrap resamples. The (orientation-independent) eigenvalues are found in Table 3.2, and the full five tensor elements are plotted in Fig. 3.7 and tabulated in Table A.2. Fig. 3.7 shows also the tensor if only one type of data (PCS/RDC) is used for fitting. This demonstrates that both types of data do lead to the same result within the margin of error. The tensors determined from RDCs are associated with a higher uncertainty, first because there are fewer data points available (most notably for Dy and Tb), and second

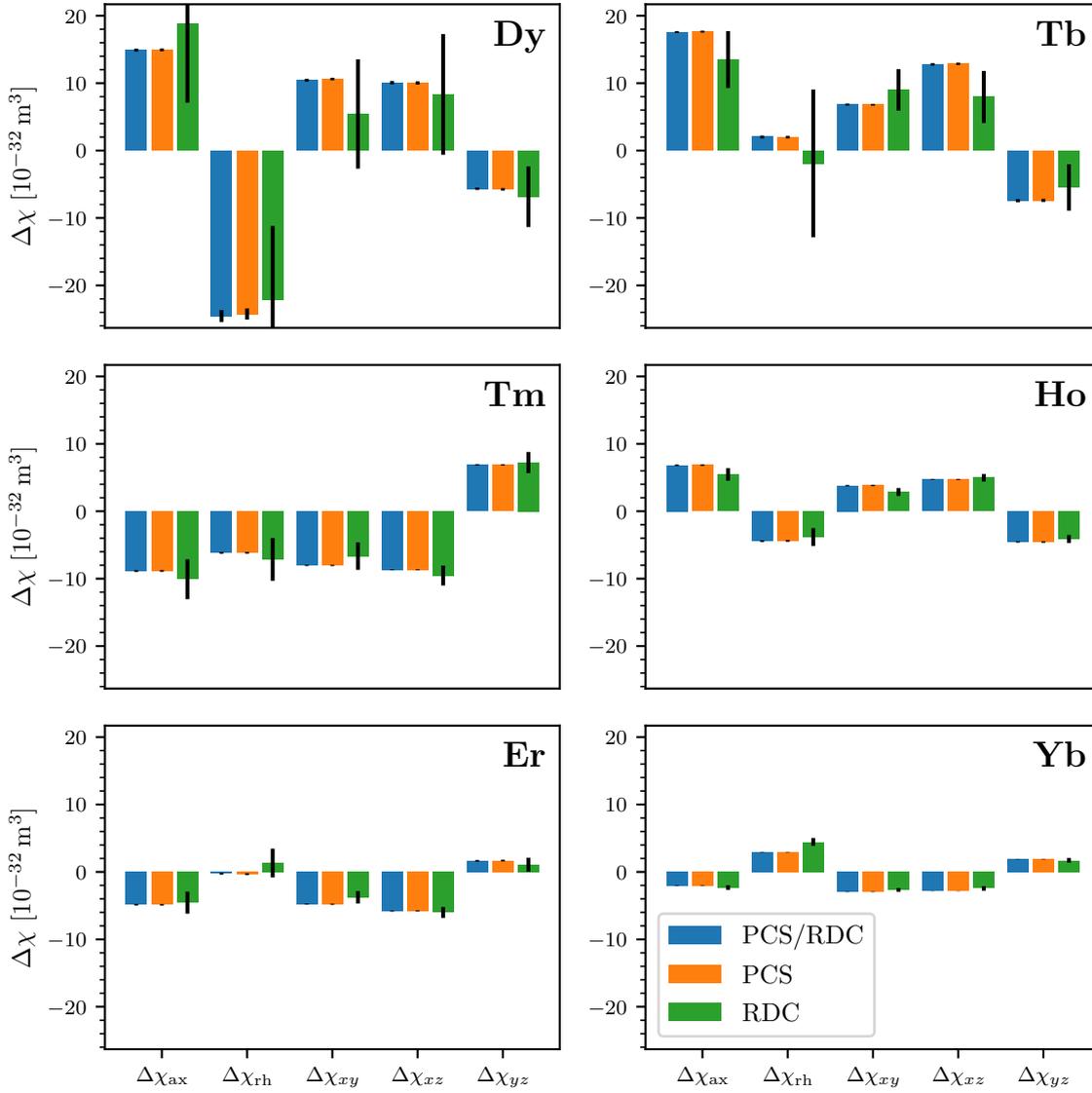
**Table 3.2.** Eigenvalues of  $\Delta\chi$  in  $10^{-32} \text{ m}^3$ , obtained from fitting both RDCs and PCSs to the N-terminal domain, and by fitting RDCs to the C-terminal domain, for all six lanthanides. The ordering follows the convention  $|\Delta\tilde{\chi}_{xx}| < |\Delta\tilde{\chi}_{yy}| < |\Delta\tilde{\chi}_{zz}|$ . The uncertainty is given in parentheses.

	domain	$\Delta\tilde{\chi}_{xx}$	$\Delta\tilde{\chi}_{yy}$	$\Delta\tilde{\chi}_{zz}$
Dy	N	8.34(30)	18.36(23)	-26.69(18)
	C	0.17(27)	3.76(29)	-3.93(30)
Er	N	-1.05(13)	-8.20(12)	9.25(8)
	C	-0.27(16)	-1.21(17)	1.48(18)
Ho	N	0.74(11)	9.46(13)	-10.21(8)
	C	-0.19(22)	-1.86(22)	2.05(23)
Tb	N	-2.55(21)	-20.94(14)	23.49(19)
	C	0.22(26)	3.76(32)	-3.98(31)
Tm	N	-3.31(11)	-14.08(14)	17.39(14)
	C	0.17(21)	2.67(26)	-2.84(27)
Yb	N	-2.27(5)	-3.55(5)	5.82(4)
	C	0.00(16)	0.65(14)	-0.65(16)

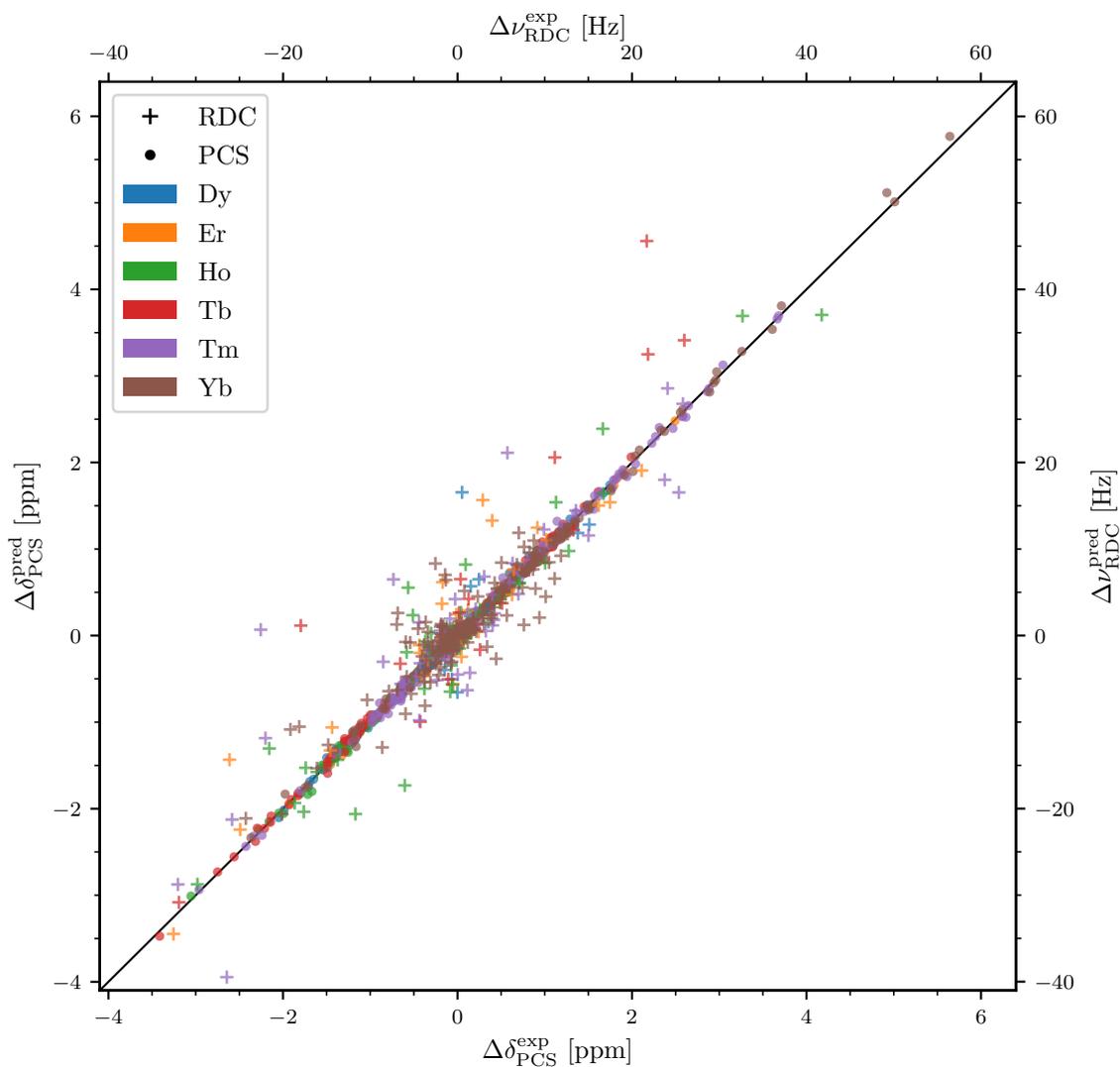
because they are associated with a higher relative scatter. The  $Q_{\text{PCS}}^{\text{N}}$ -factor is 0.038 and the  $Q_{\text{RDC}}^{\text{N}}$ -factor is 0.392, which tells us that the RDCs have ten times higher relative scatter than the PCSs. The overall  $Q^{\text{N}}$ -factor is 0.046. Across the various metals the  $Q^{\text{N}}$ -factor is relatively consistent, although there are larger variations for the RDC-only fits. All  $Q^{\text{N}}$ -factors are tabulated in Table A.1. A correlation plot of the experimental and back-calculated data is shown in Fig. 3.8.

### 3.3.2 Fitting of C-terminal RDCs

While it is necessary to have a motional model for the interdomain dynamics to fit PCSs due to their complex, non-linear dependence on distance to the lanthanide center, this is not true for RDCs. They are an alignment effect, and, under the assumption that the C-terminal domain is in itself rigid, one can fit an alignment tensor to the C-terminal RDCs and a domain structure. This alignment tensor can also be expressed as an effective  $\Delta\chi$ -tensor, so that it is comparable to the results from the N-terminal domain. Due to the interdomain motion, this C-terminal tensor will be reduced compared to the N-terminal tensor, and this reduction can be used to quantify the motion as a scalar, similar to an order parameter. As a structural model we have used the C-terminal domain of 2BE6/C; although for this analysis the three models of 2BE6 showed very little difference, we later found that 2BE6/C is a significantly better model for the C-terminal domain when including PCSs (see Section 3.3.4), so we chose to use this structure here as well. As the RDCs are associated with a relatively high statistical error, and the size of the RDCs is much

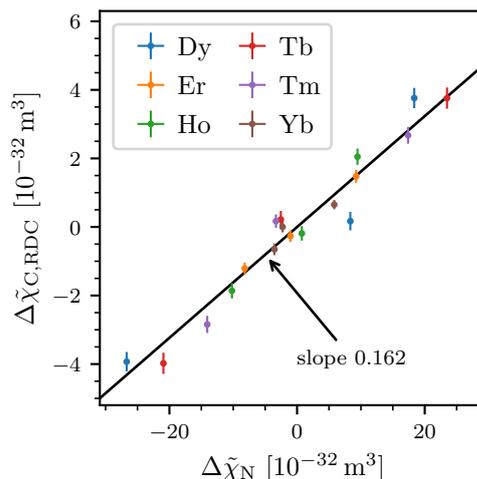


**Figure 3.7.** Plot of the five independent tensor elements for the six different lanthanides. The tensor elements obtained from fitting only PCSs or RDCs are displayed as well. Both are in good agreement with each other, except for Dy and Tb, where only very few RDCs are available. This leads to a high uncertainty in the fitted tensor elements. The reference coordinate system was taken from the crystal structure 2BE6/B. This data is also tabulated in Table A.2.



**Figure 3.8.** Correlation plot of experimental and back-calculated PCSs and RDCs for the N-terminal fit of CaM/Munc13-1. For better visibility, they are scaled as  $1 \text{ ppm} \hat{=} 10 \text{ Hz}$ . This is not to be confused with the scaling used in the data analysis.

smaller than in the N-terminal domain, these fits produced a very high  $Q_{\text{RDC}}^{\text{C}}$ -factor of overall 0.63, ranging between 0.52 and 0.93 for the different metals (Table A.1). But since we had acquired a large number of RDCs in the C-terminal domain, the resulting uncertainty in the effective  $\Delta\chi$ -tensor was still reasonable. The tensor eigenvalues are reported in Table 3.2, while the full form can be found in Table A.2. The size reduction of the tensor between the N- and C-terminal domain can best be seen by plotting their eigenvalues against each other. Note that for this correlation the eigenvalues have to be ordered by size and not by absolute values. Fig. 3.9 illustrates that there is a consistent linear scaling between the eigenvalues of the two domain's tensors with a scaling factor of 0.162 as determined via a least-squares fit. It is not immediately obvious why all metals have a similar scaling, and this is also in contrast to previous findings from Bertini et al., who had found 0.15 for thulium and 0.05 for terbium in free calmodulin.<sup>[122]</sup> Here, one needs to consider the fact that the anisotropy of the susceptibility tensor is not independent for all metals, as it all originates from asymmetry of the same binding environment. As a result the spatial orientation of the  $\Delta\chi$ -tensor is similar for all metals, and the mean angle between the eigenframes of our tensors is only around  $16^\circ$ . The tensors Bertini et al. have determined are very similar to the ones that we have found, and the angular difference between the eigenframes of Tm and Tb is as low as  $9^\circ$ . To further investigate this issue, we have performed a computer experiment by generating ensembles of three random rotations (the size that Bertini et al. have proposed in their work) and computed the averaged tensors of both Tm and Tb under these three rotations with equal population. We then evaluated individual order parameters for both metals. Only in about one in 15 000 of these random rotational ensembles we have found the order parameter for Tm to be more than three times bigger than the one for Tb. This lets us conclude that is the norm and not the exception to find order parameters that are approximately equal for all metals, and it raises some doubt about the data of Bertini et al.<sup>[122]</sup>



**Figure 3.9.** Plot of the eigenvalues of the tensor determined from the N-terminal fit  $\Delta\chi_{\text{N}}$  against the effective tensor eigenvalues from the RDC-only fit to the C-terminal domain  $\Delta\chi_{\text{C,RDC}}$ . There is a clear linear dependence with a scaling factor of 0.162. Error bars for  $\Delta\chi_{\text{N}}$  are too small to be seen.

### 3.3.3 Sampling the Conformational Space of CaM/Munc13-1

A common approach to a motional model of a complex such as Cam/Munc13-1 is to represent it as an ensemble of discrete conformations. The ensemble 2KDU from Rodríguez-Castañeda et al.<sup>[169]</sup> was generated without long-distance constraints such as PCSs and RDCs, so it was not surprising that it does not describe the interdomain motion very well. If one tries to fit the C-terminal paramagnetic data to this ensemble,

it agrees only very poorly.<sup>[193]</sup> A simple way to assess the mobility of a given model is to calculate its order parameter, just as we have done it in the previous section. This evaluates to be 0.48 for the ensemble 2KDU, which is about a factor of three larger than the experimental result, so this ensemble is clearly more rigid than the real complex.<sup>[193]</sup> To find an ensemble that fits to the experimental data, it was necessary to generate a new pool of conformations from which an ensemble could be compiled. We generated this pool by doing a conformational search with the structures from 2KDU as starting points.

It was our goal to generate a set of conformations that spanned a motional range as large as possible, and which only excluded conformations that were sterically impossible, to then later choose a subset of this pool as the final model. We therefore did not worry about modeling the protein as realistically as possible or to calculate accurate energies. Instead, we modeled the protein in vacuum using the OPLS3 force field.<sup>[202]</sup> We chose this force field since other popular options for proteins, such as AMBER,<sup>[203]</sup> did not include parameters for  $\text{Ca}^{2+}$ . The conformational search was very similar to the ones performed in Chapter 4, consisting of a random sampling of certain bond torsion angles. We chose the two backbone dihedrals ( $\phi$ ,  $\psi$ ) of the residues 76–81 in calmodulin’s linker region to be sampled in the search, which corresponds to the region where Tjandra et al. have found increased mobility.<sup>[174]</sup> Sampling a larger stretch of the linker lead to distortions of the adjacent  $\alpha$ -helices, so we limited the search to this relatively small range of residues. Each sampling step changed a random selection of up to six of the aforementioned torsion angles by up to  $10^\circ$ , and then performed a minimization of the resulting structure. The convergence criterion was set to a relatively large  $1 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$  as the minimization was the computational bottleneck of the search. The main additional complication compared to a standard conformational search was the presence of the peptide linker. The change in torsion angles was done such that the N-terminal domain remained immobile, while the C-terminal domain moved through space. We needed to ensure that the binding of Munc13-1 to its pocket in the C-terminal domain was not undone by this step, which was achieved by adding a number of artificial distance constraints. A key residue in this binding is Munc’s tryptophan 489, which is located inside of a hydrophobic pocket within Calmodulin. We added constraints to 18 of calmodulin’s backbone atoms that were less than  $8 \text{ \AA}$  away from the central carbon atom  $\text{C}_{\delta 2}$  of W489. These constraints were executed as flat-bottom potential wells centered around the mean distance of the corresponding atoms in the 20 structures of 2KDU, a half-width of the flat section of  $1 \text{ \AA}$ , and a force constant in the outer part of  $200 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ . These constraints ensured that after altering the linker torsion angles, W489 was pulled back into its binding pocket in the ensuing minimization step, and distributing this stress onto 18 different atoms prevented a distortion of the binding pocket. Another measure to preserve the original domain structure was constraining the two hydrogen bonds in each domain’s section of antiparallel  $\beta$ -sheet. As an energy cutoff we used  $3000 \text{ kJ mol}^{-1}$ , which is twice the difference between the highest and lowest energy structure in the 2KDU ensemble. We ran this conformational sampling with 9900 steps, starting from each of the 20 structures from 2KDU, and with approximately two thirds of the steps producing a compatible

geometry, we generated a total of 122 700 conformations of CaM/Munc13-1. These served as a basic pool of conformations in all further steps.

Such a large number of protein structures can become a challenge to handle computationally, and the total size of structure files was 26 GB. It was therefore desirable to reduce this data to the essential pieces of information and to discard all others. While the most obvious approach would be to eliminate the positions of atoms not involved in any paramagnetic constraints, which would already cut down the necessary space in memory significantly, we decided to go even further. As explained in Section 3.3.1, the structure of the individual domains is represented rather poorly by the NMR ensemble 2KDU, and X-ray structures such as 2BE6 agree much better to the available paramagnetic constraints. Therefore we again chose to use the C-terminal domain from one of the three structures of 2BE6. As a consequence, the only relevant piece of information contained within a conformation of CaM/Munc13-1, as generated by the conformational search, is the relative orientation of the two domains, which can be expressed as three translational and three rotational degrees of freedom. This is obviously an enormous reduction in complexity and facilitates data handling significantly.

In practice, this was implemented using homogeneous coordinates, which are a convenient way to simultaneously express translation and rotation. For regular three-dimensional coordinates translation is expressed as vector addition and rotation as matrix multiplication. Consider a point  $P$  with position vector  $\mathbf{p}$ . Translation and rotations are computed as follows:

$$\mathbf{p}_{\text{trans}} = \mathbf{p} + \mathbf{q}, \quad \mathbf{p}_{\text{rot}} = \mathbf{R}\mathbf{p}. \quad (3.5)$$

Where  $\mathbf{q}$  is the translation vector and  $\mathbf{R}$  a rotation matrix (i.e.,  $\mathbf{R}^T\mathbf{R} = \mathbf{1}$  and  $|\mathbf{R}| = 1$ ). However, it is not obvious how to pack these two operations into one. Homogeneous coordinates provide a solution to this problem. They are generated by adding an additional coordinate  $w$ , and for any point  $(x, y, z)$  the tuple  $(xw, yw, zw, w)$  is a set of homogeneous coordinates of this point. Any choice of  $w$  refers to the same point, so it is typically set to 1. In these coordinates, both translation and rotation can be expressed as matrix multiplication, as is shown below:

$$\mathbf{p}_{\text{trans}}^{\text{homo}} = \begin{pmatrix} 1 & 0 & 0 & q_x \\ 0 & 1 & 0 & q_y \\ 0 & 0 & 1 & q_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \\ 1 \end{pmatrix} = \begin{pmatrix} q_x + p_x \\ q_y + p_y \\ q_z + p_z \\ 1 \end{pmatrix}, \quad (3.6)$$

$$\mathbf{p}_{\text{rot}}^{\text{homo}} = \begin{pmatrix} & & 0 \\ & \mathbf{R} & 0 \\ & & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \\ 1 \end{pmatrix} = \begin{pmatrix} \\ \mathbf{R}\mathbf{p} \\ \\ 1 \end{pmatrix}. \quad (3.7)$$

With this it is possible to compute a matrix representing two or more concatenated translations and rotations by simply taking the matrix product of the individual transformation matrices. Using this formalism we expressed each conformation as a  $4 \times 4$ -matrix,

which described the transformation of the C-terminal domain of some reference structure (in our case, 2BE6/B) to the desired location. The transformation matrix was found as the transformation that minimizes the distance RMSD between their backbone atoms (N, C $_{\alpha}$ , C') of residues 84–145, so both structures were represented as the collection of these atom positions. First, both the reference and the conformation were translated such that the mean of all atom positions would lay in the origin, with the translation matrices  $\mathbf{T}_{\text{ref}}$  and  $\mathbf{T}_{\text{conf}}$ . From there it is a pure rotation  $\mathbf{R}$  that transforms the reference into the conformation, which can be found using the Kabsch algorithm.<sup>[204]</sup> The total transformation matrix  $\mathbf{M}$  is then found by concatenating these three operations:  $\mathbf{M} = \mathbf{T}_{\text{conf}}^{-1} \mathbf{R} \mathbf{T}_{\text{ref}}$ . With these matrices and the knowledge about a single reference structure, any interdomain orientation could be reconstructed. As an additional feature of this formalism, it is possible to construct translation-invariant vectors by setting the coordinate  $w$  to zero (i.e.,  $(x, y, z, 0)^{\text{T}}$ ). This is useful if one has already computed internuclear vectors for RDCs within the C-terminal domain, which are only affected by domain rotation, but not by domain translation.

### 3.3.4 Ensemble Sampling

Now that we had found a set of domain arrangements that were sterically possible, it was necessary to find a subset that is in agreement with the experimental data. As the susceptibility tensors were already very accurately determined in the N-terminal fits, we chose an approach that used these tensors instead of fitting them again. The first step was to predict the paramagnetic constraints in the C-terminal domain using the  $\Delta\chi$ -tensors from the N-terminal fits, the rigid C-terminal domain structure 2BE6/C, and its arrangement relative to the N-terminal domain (and with that the  $\Delta\chi$ -tensor) from the transformation matrices, considering also the contribution from RCSAs as explained in Section 3.3.1. This yielded an array of  $122\,700 \times 2691$  data points, and the problem of finding a matching ensemble was now equivalent of finding a linear combination (with certain constraints) of rows of this array that reproduce the 2691 experimental data points. The coefficients of such a linear combination can then be interpreted as populations of the corresponding conformation. Although this is in principle no different from the many other systems of linear equations that we have solved in this work, the sheer size of the coefficient matrix of  $122\,700 \times 2691$  makes it impossible to find a solution by applying the usual algorithms.

We therefore chose a random-sampling approach to find approximate solutions to this problem. We randomly chose a small subset of up to 100 ( $n_{\text{ens}}$ ) rows, or conformations, and checked its agreement with the data. To do this, there are two ways to determine the individual populations of the chosen conformations. Arguably the simplest way is to populate them all equally by the inverse of the number of conformations. Alternatively one can fit for the populations, under the constraint that they are positive and sum up to one. Populating them equally has the advantage that evaluating the agreement is very fast, as one has to simply calculate the mean across rows and compare it to the experimental values. This allows for sampling many more ensembles in the same amount of computation time than when one needs to fit for each population. Larger

populations can in principle be described by choosing the same conformation multiple times. However, this approach comes with a number of drawbacks.

First, there is no clear hierarchy in these models, and it is a mistake to equate the number of ensemble members to a formal number of fit parameters. In hierarchical model systems, adding a degree of freedom (i.e., a fit parameter) can never lead to a decrease in agreement, as higher level models include lower level models. In the case of fitted populations, a model with a smaller ensemble size can simply be reproduced by setting some populations to zero. For the case of equally populated ensembles however, an increase in ensemble size can very well lead to an overall decrease in agreement and slower convergence, as too many members lead to a model that is too mobile, and due to the random nature of the sampling it is rather unlikely to pick the same conformation multiple times. Second, if one designs an iterative sampling strategy (which will be described in detail further down), it is not obvious how to systematically improve an existing ensemble, while for ensembles with fitted populations one simply needs to replace the members with low populations. Third, finding ensembles with good agreement with the data is much more unlikely, as one needs to find just the right combination, and therefore the improvement via random sampling is very slow. It turns out that this effect by far overcompensates for the faster evaluation of the ensemble in terms of computation time. Fourth, we find that setting all population to an equal value is a very artificial choice with no justification from reality. It is basic knowledge from chemistry that conformations of molecules of all sizes are populated to different extents based on their free energy, and we believe it is therefore the more realistic choice.

As briefly mentioned above, this fitting procedure is different from the linear systems that we have encountered before as they need to be solved under the constraint that the sum of populations  $p_i$  equals to one and that there may be no negative populations. The first constraint can be solved by adding an additional, synthetic data point with a very high value (in our case  $10^6$ ) that is the same for both the experimental as well as for all predicted data sets. The system of linear equations then has the following matrix form:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n_{\text{ens}}} \\ x_{21} & x_{22} & \cdots & x_{2n_{\text{ens}}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_{\text{data}}1} & x_{n_{\text{data}}2} & \cdots & x_{n_{\text{data}}n_{\text{ens}}} \\ 10^6 & 10^6 & \cdots & 10^6 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_{\text{ens}}} \end{pmatrix} = \begin{pmatrix} x_{\text{exp},1} \\ x_{\text{exp},1} \\ \vdots \\ x_{\text{exp},n_{\text{data}}} \\ 10^6 \end{pmatrix}. \quad (3.8)$$

It is easy to see that the equation for the last row in Eq. (3.8) is equivalent to the constraint of  $\sum p_i = 1$ , and due to the high value small deviations from this lead to a steep increase in the minimization criterion (i.e., RMSD). The second constraint leads to the non-negative least squares (NNLS) problem, which can be solved with deterministic algorithms,<sup>[205]</sup> and fast implementations are available.

As mentioned before, we devised an iterative sampling approach to find increasingly better ensembles, vaguely inspired by the concept of evolution with the survival of the fittest and mutation. First, we drew  $n_{\text{sample}}$  different ensembles with size  $n_{\text{ens}}$ , fitted their populations, and calculated their agreement with the data as the RMSD. We then

took the 10% of these ensembles that fitted best, and duplicated them ten times. All conformations that had a fitted population of zero were replaced randomly with new conformations. Apart from that, the original set remained unchanged. For the first copy, we replaced an additional 10% of ensemble members randomly, and for the  $i$ th copy, we replaced  $10i$ % of ensemble members randomly. This way we generated a new set of ensembles with same size  $n_{\text{sample}}$  with varying degrees of variation (mutation) compared to the most successful samples from the previous step. Since for one copy we only replaced unpopulated members, we ensured that each generation would never be worse than the previous one, and the agreement as measured by the best member of the sample would therefore decrease monotonically with each step.

As before with the N-terminal domain, we were still faced with the issue of choosing a structural model and the simultaneous evaluation of PCSs and RDCs. To evaluate PCSs and RDCs simultaneously, we scaled them again with their (estimated) standard deviation. For the RDCs this was simple, as we could use the RMSD from the RDC-only C-terminal fit as an estimate for their scatter. Since there was a significant dependence on the type of RDC, we chose to scale them independently, and their RMSDs were 1.63, 3.42, 2.20, and 0.78 Hz for  $D_{\text{NH}}$ ,  $D_{\text{C}\alpha\text{H}\alpha}$ ,  $D_{\text{HC}'}$ , and  $D_{\text{C}'\text{C}\alpha}$ , respectively. For the N-terminal domain this distinction between RDC types was not feasible due to the very low number of RDCs for some metals. We estimated the scatter of the PCSs by a short sampling procedure using exclusively PCSs, using 100 steps, a sample size of 1000, and an ensemble size of 50 (taking about 10 minutes). After such a sampling the RMSD had already converged up to a few percent to the asymptotical value, and we took the resulting RMSD of 9.8 ppb as a reasonable approximation for the PCS scatter. This type of scaling with the estimated standard deviation leads to the effect that the combined, scaled RMSD for a reasonable model will approach unity, and this is what we will report as RMSD in the following. We proceeded in a similar way for the three candidates from 2BE6 for the domain structure, and to our surprise we found that the C-terminal structure of 2BE6/B provided the worst agreement with our experimental data, yielding an about 30% larger RMSD than the other two structures. Between them the difference was less striking, but 2BE6/C yielded about 5% smaller RMSDs than 2BE6/A, so we chose it as our model for the C-terminal domain structure. It is rather surprising to find the PCSs to be the discriminating factor between different domain structures, and it is conventional wisdom that they are less sensitive to small structural variations than RDCs. It is possible that this is simply due to the lower relative uncertainty that is associated with the PCSs. However, as the domain structure was not the scope of this work, we did not investigate this in more detail.

To find an appropriate ensemble size, we did again short test samples, and employed the Bayesian Information Criterion (BIC)<sup>[206]</sup> as a preliminary measure for model selection. Under the assumption of independently and normally distributed errors, it can be formulated as:<sup>[207]</sup>

$$\text{BIC} = n_{\text{ens}} \ln(n_{\text{data}}) - 2n_{\text{data}} \ln(\sigma_{\text{data}}), \quad (3.9)$$

where  $\sigma_{\text{data}}$  is the data's standard deviation, which can be estimated by the RMSD. When comparing two models, the one with the lower BIC is preferable. This allows us

to compute a factor by which the RMSD needs to decrease for each additional ensemble member:

$$(n_{\text{ens}} + 1) \ln(n_{\text{data}}) - 2n_{\text{data}} \ln(\sigma_{\text{data}}^{n+1}) < n_{\text{ens}} \ln(n_{\text{data}}) - 2n_{\text{data}} \ln(\sigma_{\text{data}}^n) \quad (3.10)$$

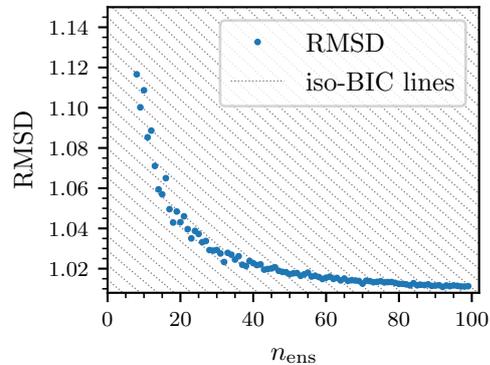
$$\ln(n_{\text{data}}) < 2n_{\text{data}} \left( \ln(\sigma_{\text{data}}^n) - \ln(\sigma_{\text{data}}^{n+1}) \right) \quad (3.11)$$

$$\exp\left(\frac{\ln(n_{\text{data}})}{2n_{\text{data}}}\right) \approx 1.0015 < \frac{\sigma_{\text{data}}^n}{\sigma_{\text{data}}^{n+1}} \quad (3.12)$$

This tells us that for each additional ensemble element the achieved RMSD should drop by at least 0.15%. While we later used cross-validation to find an optimal ensemble size for the final model, this simple comparison via the BIC helped us get an idea of the size of ensemble needed. We again ran a short test sampling with 100 steps and 1000 samples with all ensemble sizes between 1 and 100 and plotted the achieved RMSD against the ensemble size  $n_{\text{ens}}$  (Fig. 3.10). As expected each additional ensemble member brought less of an improvement, so that a certain point we should go below the BIC threshold of 1.0015. This can easily be seen by plotting lines of constant BIC, corresponding to the relation  $\text{RMSD} \propto 1.0015^{-n_{\text{ens}}}$ . Inspection of Fig. 3.10 reveals that the models with lowest BIC correspond to ensemble sizes somewhere between 15 and 30.

Based on this, we chose the ensemble sizes for the final samplings. As it is always possible to reduce the ensemble size afterwards, we chose the previously determined size of 15–30 as a lower bound and ran longer samplings with ensemble sizes of 15, 20, 50, and 100. As the evaluation time scales quadratically with  $n_{\text{ens}}$ , we had to run fewer steps and samples for the larger ensembles, and the final samplings were performed with 15 000, 10 000, 4000, and 2000 steps and samples for sizes of 15, 20, 50, and 100, respectively. As this is a probabilistic procedure, we repeated each sampling 50 times to evaluate the distribution of results. In Fig. 3.11 we have shown the best sample’s RMSD as well as the BIC as a function of steps performed.

The final ensembles with sizes of 15 and 20 perform similarly well. However, even when taking into account the ensemble size with the BIC, the 20-membered ensembles still performed slightly better. On the other hand, the ensembles with sizes of 50 and 100 showed clear signs of overfitting. First, in agreement with the preliminary results, the RMSD did not decrease by the necessary  $1.0015^{-\Delta n_{\text{ens}}}$  going from 20 to 50 to 100 structures. Secondly, the fitted populations for some structures in the larger ensembles were very low, and particularly for  $n_{\text{ens}} = 100$  there was always 10–20% of structures that had a fitted populations of



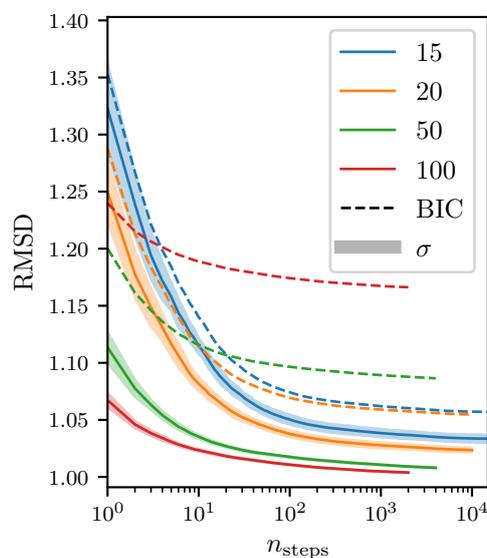
**Figure 3.10.** RMSD as a function of ensemble size for a short test sampling. The dotted lines show lines of constant BIC. The best models (smallest BIC) are found in the bottom left region of the plot, corresponding to an ensemble size between 15–30.

exactly 0. We therefore devised a cross-validation algorithm to iteratively reject all unnecessary ensemble members.

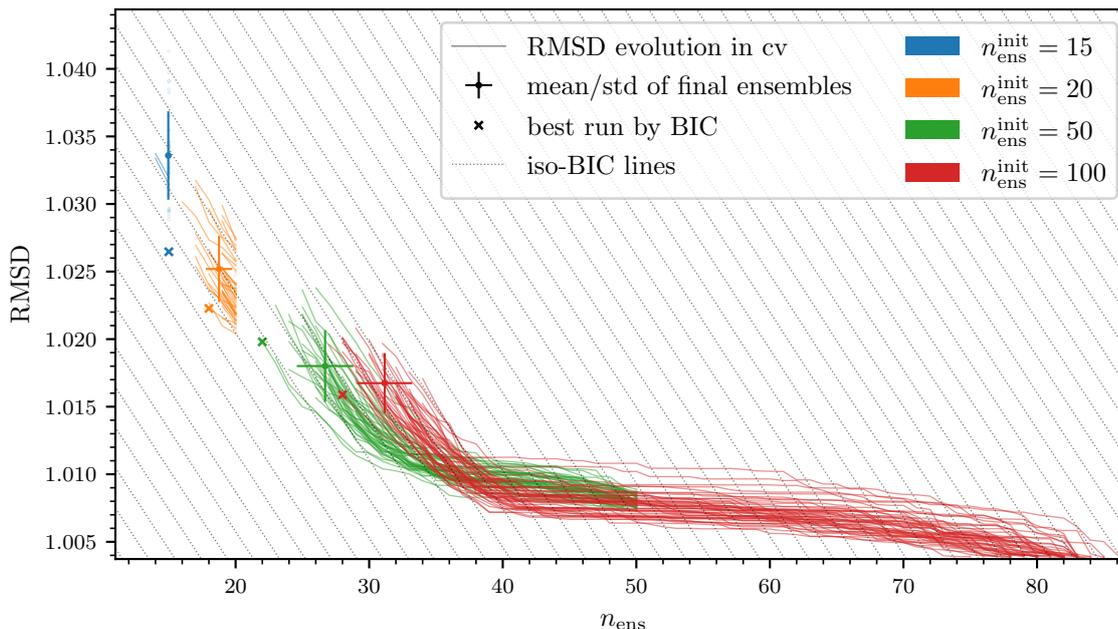
Cross-validation is a very simple and elementary technique to detect overfitting, and it involves splitting the available data into two parts, fitting the model to the first part (the training set) and then evaluate the agreement with the second part (the test set). In the case of overfitting the agreement of a more complex model is only higher with the training set, and not with the test set.<sup>[208]</sup> This iterative cross-validation was done in the following way: We split our data randomly in half in 100 different ways into training and test sets of equal size. Taking an ensemble, we fitted populations against the 100 training sets and evaluated the RMSD against the 100 corresponding test sets. We then compiled  $n_{\text{ens}}$  sub-ensembles by removing one member each, and repeated cross-validation process for these sub-ensembles. If all members of the original ensemble were necessary, one would expect all of the sub-ensembles to perform poorer against the test sets. We therefore computed the ratio of the sub-ensemble RMSDs and the full-ensemble RMSD and checked for values smaller than the BIC-threshold of 1.0015.

If we found any, we removed the member with the lowest ratio from the ensemble. We then went back to the start and repeated the process. If no ratio below 1.0015 was found, this indicated that every member accounted for an improvement of at least this much and was considered necessary, so no overfitting was present any more. Although it is somewhat redundant to combine cross-validation with the BIC, and it would in principle be sufficient to look for RMSD ratios smaller than unity, we wanted to apply a somewhat stricter measure to ensure that each ensemble member would contribute to an improvement instead of just stagnation, and so the BIC threshold that we had computed earlier was a convenient and somewhat justified quantity for such a tightened criterion. Also, there is some statistical uncertainty to these ratios as this is again a probabilistic algorithm, but by taking 100 independent splits of the data, we could both reduce and estimate this uncertainty to about an order of magnitude smaller than the threshold of 1.0015.

Several key aspects and results of this iterative cross-validation are presented in Fig. 3.12. Firstly, it shows the evolution of the RMSD upon successive removal of structures for each of the 50 runs. This again demonstrates very clearly that both 50 and 100 are too large for an appropriate model. In both cases the removal of ensemble members only reduces the agreement with the data very slowly, until a critical ensemble



**Figure 3.11.** Convergence of RMSD for ensemble sizes of 15, 20, 50, and 100, run with 15 000, 10 000, 4000, and 2000 steps and samples, respectively. The shaded region shows the estimated standard deviation calculated as the RMSD of 50 individual runs. The dashed line is calculated as  $\text{RMSD} \cdot 1.0015^{n_{\text{ens}}}$ , which corresponds to the BIC.



**Figure 3.12.** Evolution of RMSD during iterative cross-validation of the 50 different runs, starting at  $n_{\text{ens}}^{\text{init}}$  of 15, 20, 50, and 100. The dots with error bars show the mean and standard deviation of both RMSD and ensemble size for the final 50 ensembles resulting from the cross-validation procedure, and the cross marks the best ensemble by BIC. Lines of equal BIC are dotted in gray.

size of around 35 is reached. There, the RMSD lines curve upwards, until reaching a critical slope that halts the iterative cross-validation. This trend is not observed for starting sizes of 15 or 20, where removal of single ensemble members sharply increase the RMSD from the beginning. Although the cross-validation was able to remove redundant structures as far as possible, the fact that they originate from a sampling with a large ensemble size still remains noticeable. Fig. 3.12 also shows the average and the standard deviation of both the ensemble size and achieved RMSD for the final, cross-validated structures originating from different samplings. The final ensembles from the  $n_{\text{ens}}^{\text{init}} = 100$  sampling have an average size of 31.2, almost five more members than the ones starting at a size of 50, which got reduced to an average size of 26.7. The slightly better RMSD of the former does not justify this larger size when considering the BIC. This is again best seen by inspecting the points relative to the lines of constant BIC (iso-BIC) depicted in Fig. 3.12. The trend remains the same when going to the samples that started with 20 members. They were only reduced to an average size of 18.8, indicating that they were already much closer to the ideal size. In terms of BIC these models are again better than the counterparts originating from larger ensembles. Only when going to a starting size of 15 the reduced ensemble sizes has a sufficient effect on the RMSD so that they on average perform poorer in terms of BIC than the ones starting from 20. This discussion was based on the mean of the 50 runs, and a similar trend is observed in the best ensembles, marked as crosses in Fig. 3.12. The exception to this is the absolute best

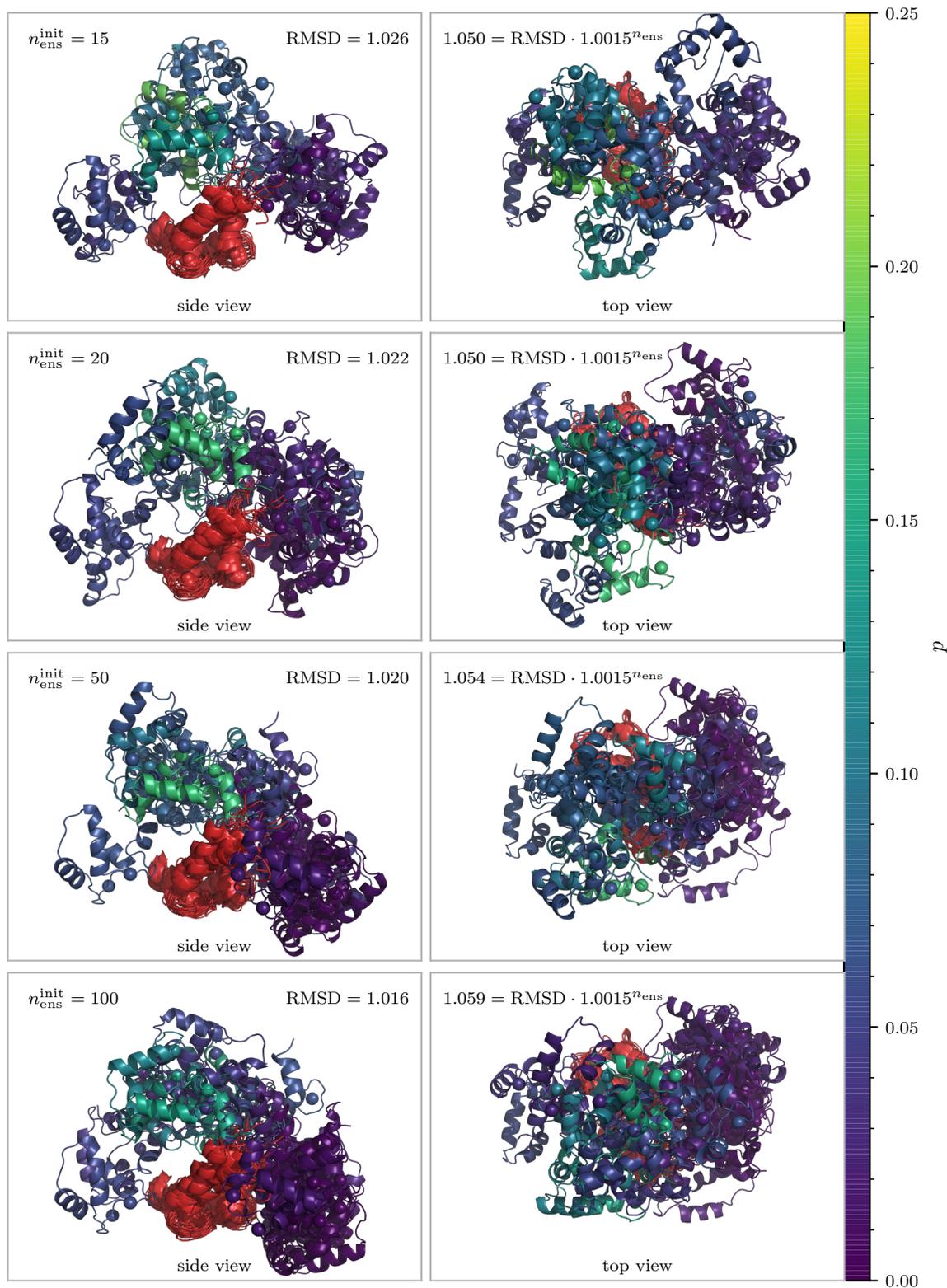
ensemble, which originated from a sampling with 15 ensemble members; the difference to the best result starting from 20 members is small though (less than 0.1%), and this can be attributed to the probabilistic nature of the sampling. The  $Q^C$ -factor for this absolute best ensemble evaluates to 0.290.

When comparing the different ensembles found by this procedure, they all populate similar conformations. Fig. 3.13 shows cartoon representations of the best ensembles by BIC for the four different starting ensemble sizes (15, 20, 50, and 100), aligned on the N-terminal domain and shown from two different angles (called side and top view). It becomes immediately obvious that the C-terminal domain of CaM/Munc13-1 samples an ample region of space around the N-terminal domain. Seen from the side, this roughly corresponds to an arc of around 180 degrees, with the highest population in the center and minor conformations on the sides. When this arc is seen from the top, the shapes are somewhat more distinct, forming a belt with uniform width for the larger ensembles ( $n_{\text{ens}}^{\text{init}} = 50, 100$ ), while it is less regular for the smaller ensembles ( $n_{\text{ens}}^{\text{init}} = 15, 20$ ). Fig. 3.13 shows only ensemble members with a population above 10%. This highlights that in all ensembles the most strongly populated conformations are also very similar.

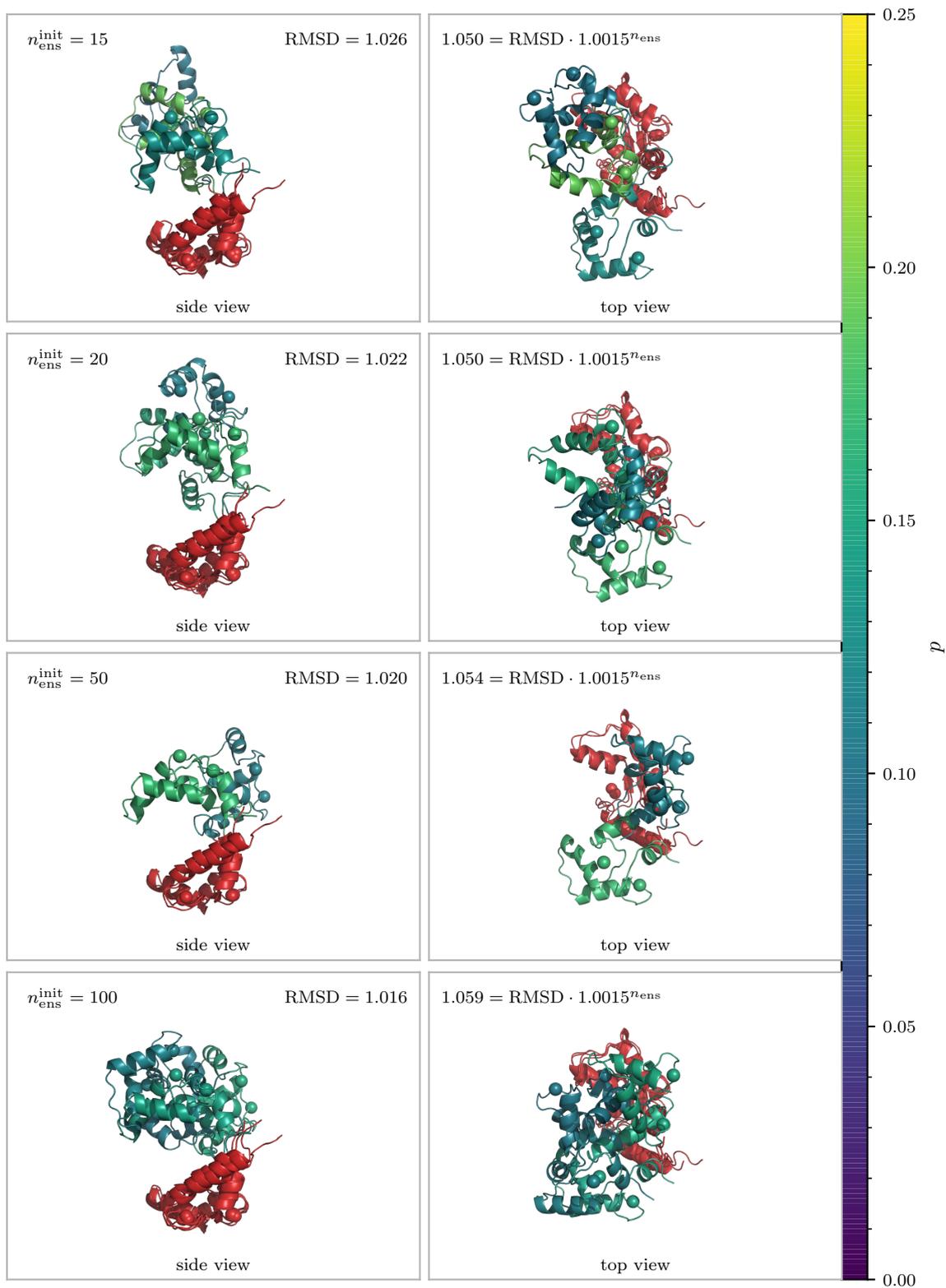
In all cases, this motion cannot be reduced to rotation along one or two degrees of freedom. In addition to the rotation of one domain around the other, the ensembles show varying degrees of torsion (rotation around a vector connecting the two domains), and none of the ensemble rotations can be reduced to a single pivot. We have shown the latter by trying to shift the origin so that the translational components of an ensemble's transformation matrices would vanish. However, this proved to be impossible, and the remaining translational components had a mean length of around 13–17 Å, which was quite consistent across the various ensembles. We have tried to capture these findings in Fig. 3.15 on the example of the absolute best ensemble (Fig. 3.13, first row). It shows a cartoon of 2BE6/C, which is the starting point for all rotation matrices. The optimal origin of rotation is depicted as a white sphere, which is expectedly located in the linker region. We have then represented the ensemble members as arrows. The distance between the origin of rotation and the base of these arrows is the remaining translational component, whereas the arrow's length and orientation represent the angle and axes of rotation, respectively. The population is color-coded as before. It is visible from the arrow's wide distribution how much translation contributes to the domain reorientation. There seems to be no discernible pattern or correlation between translation and rotation, and this image clearly illustrates the wide range of motion that the domains are spanning.

### 3.4 Conclusion and Outlook

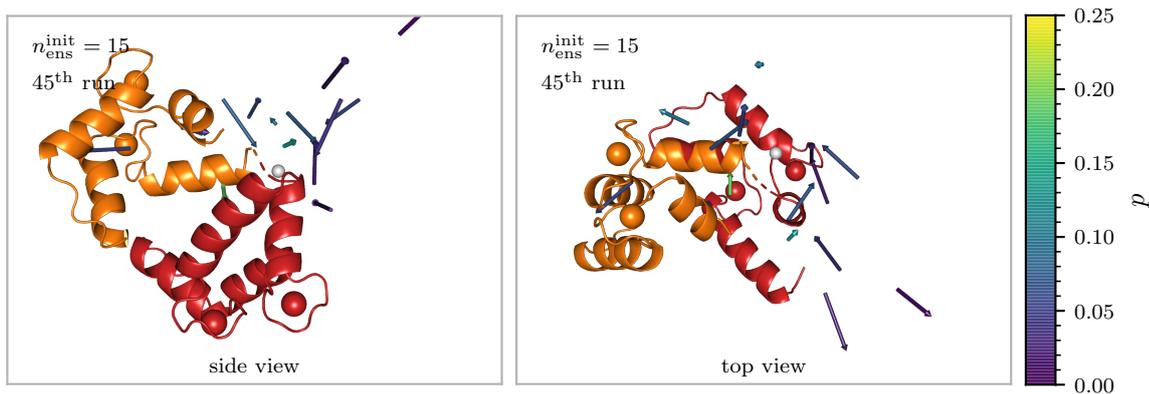
In this chapter we have shown how the interdomain motion of the complex CaM/Munc13-1 could be investigated and modeled using paramagnetic NMR. To reduce the impact of unavoidable experimental scatter, we have acquired as many paramagnetic constraints in the backbone region as possible, and used exclusively triple-resonance experiments to reduce the amount of data rejected because of overlap. This way we have been able to acquire a sizable 1026 constraints in the N-terminus and 2691 constraints in the



**Figure 3.13.** Best ensembles by BIC for the different starting ensemble sizes, viewed from two different angles. The C-terminal is color-coded by population  $p$ .



**Figure 3.14.** Best ensembles by BIC for the different starting ensemble sizes, viewed from two different angles. Only members with a population higher than 10 % are shown. The C-terminal is color-coded by population  $p$ .



**Figure 3.15.** Cartoon of 2BE6/C, and transformations of the absolute best ensemble from  $n_{\text{ens}}^{\text{init}} = 15$  represented as arrows. The optimal origin of rotation is depicted as a white sphere. The distance between the origin of rotation and the base of these arrows is the remaining translational component, whereas the arrow's length and orientation represent the angle and axes of rotation, respectively.

C-terminal domain. The large amount of data and the careful inclusion of minor effects such as the lanthanide positioning and RCSAs enabled us to determine the lanthanide's susceptibility tensor with a very low relative uncertainty between 1–3%, which laid the foundation for the subsequent data analysis. Following the simplistic approach from Bertini et al.,<sup>[122]</sup> we determined an order parameter of 0.162 by comparing the relative degrees of alignment in both domains. Unlike their results for free calmodulin, we could not find any pronounced dependence on the metal. Our order parameter is very close to the value of 0.15 that Bertini et al. had found for free CaM in the case of thulium. This supports the idea that the interdomain motion of calmodulin is only marginally restricted by its binding to Munc13-1. We have also argued that it is expected to find similar order parameters for all metals, since their tensors all share a similar eigenframe (not eigenvalues!). This result should therefore not be interpreted as isotropic motion, which is supported by the fact that the ensembles that we have found to describe our data do not exhibit a high degree of symmetry.

To find these ensembles, we have sampled the conformational space of CaM/Munc13-1 extensively using molecular mechanics, including a number of synthetic force contributions to keep the complex together during the sampling. We then devised a way to reduce the information content of each conformation to the relative interdomain arrangement by borrowing a tool from projective geometry, the homogeneous coordinates. From these arrangements and the  $\Delta\chi$  tensors we predicted the paramagnetic data for each conformation, which yielded the basic data matrix for the subsequent sampling.

This sampling was executed by randomly choosing subsets of conformations and then iteratively improving them by randomly modifying the ones with the best agreement, a process vaguely inspired by evolution. We have then used a combination of cross-validation and the Bayesian information criterion (BIC) to eliminate redundant conformations and to compare models with different ensemble sizes against each other. This way, we have found the necessary ensemble size to be between 15 and 20. Various

ensembles that perform similarly have been found by repeating this non-deterministic process multiple times, and they all span a similar region in space. By examining them we have found that they describe an interdomain motion that comprises both translation in the order of 15 Å and rotation about all three spatial directions. This further confirms our result that the interdomain motion is hardly restricted upon binding to Munc13-1, which is in stark contrast to the case of CaM/IQ which was investigated in a similar manner.<sup>[123]</sup>

As we have gathered a large amount of paramagnetic constraints of a highly dynamic system, this could be used as test case for some more fundamental questions about the types of motion that are detectable and distinguishable this way. In the most general way the interdomain motion can be thought of as a probability density in the six motional degrees of freedom, which is in contrast to the description as an ensemble, which is merely a collection of points in these six dimensions. D’Auvergne et al. have already formulated the extensive theory of frame ordering that unifies the description of rotational ordering of rigid body frames, and they found that the averaging of an alignment tensor can be reduced to a rank-4 tensor with 15 independent components, which is therefore the maximum amount of information that can be gathered from RDCs.<sup>[116]</sup> PCSs however also encode the distance of the two domains as an inverse third power, so they are both more rich in information and much more difficult to model. D’Auvergne et al. have proposed a variety of parameterized, continuous motional models such as the isotropic cone rotation or the free rotor. As none of these models incorporate translation, based on our findings they should not be able to model the motion of CaM/Munc13-1. Also, they do not form clear model hierarchies in the sense that they do not allow incremental increase in model complexity, unlike the ensemble approach, where the ensemble size is such an hierarchical increment. It would be quite interesting whether it is possible to find a suitable set of 6D-functions in which the probability density can be expanded, and whether this is able to outperform an ensemble-based approach.

## 3.5 Materials and Methods

**3.5a Sample Preparation** Uniformly  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled N60D-Calmodulin was expressed from *E. coli* following published procedures,<sup>[191,192]</sup> with the addition of a dialysis step against buffer A (20 mM Bis-Tris, 150 mM KCl, 150  $\mu\text{M}$   $\text{CaCl}_2$ , pH 6.8) to yield a solution of 1.8 mL at  $\approx 0.9$  mM protein concentration. 50  $\mu\text{L}$   $\text{D}_2\text{O}$  ( $\approx 3\%$ ) were added for field locking purposes. Ca-loading was checked via NMR and adjusted by addition of 1 equiv.  $\text{CaCl}_2$  (9.1  $\mu\text{L}$  200 mM  $\text{CaCl}_2$ ). The complex was prepared by repeatedly adding an aliquot of 500–1000  $\mu\text{L}$  to lyophilized Munc13-1, agitate for 30 min, adjust the pH, and reunite the protein solutions, until full 1:1-CaM/Munc13-1-saturation was observed via NMR. The solution was then concentrated to 1120  $\mu\text{L}$  at 1.43 mM with a Vivaspin 20 with PES membrane and 5 kDa molecular weight cut-off (MWCO). The lanthanide samples were prepared by taking aliquots of 130  $\mu\text{L}$  and titrating 30 mM  $\text{LnCl}_3$  in buffer A until 1:1 Ln loading was observed via NMR, with  $\text{Ln} \in \{\text{Lu}, \text{Dy}, \text{Er}, \text{Ho}, \text{Tb}, \text{Tm}, \text{Yb}\}$ . The samples were prepared in 3 mm NMR tubes and stored at 4 °C.

**3.5b NMR Spectroscopy** All NMR experiments were acquired on Bruker Avance III HD spectrometers operating at proton frequencies between 600–950 MHz using inverse cryogenically cooled (QCI/TCl) probes. The sample temperature was set to 298 K and checked using 99.8 % MeOD.<sup>[137]</sup> Triple-resonance experiments were acquired using the following parameters: the carrier offsets were set to 4.7, 116.85, 174.8, and 52.65 ppm and the spectral widths to 14, 28.1, 12, and 28 ppm for H, N, C', and C<sub>α</sub>, respectively. The number of dummy scans was 512, the number of scans per increment 8, and the relaxation delay 1 s. For the uncoupled spectra, the number of acquired real points were 1024, 256, 256, and 128 for H, N, C', and C<sub>α</sub>, respectively. For the coupled spectra (for RDCs), the number of real points were 1024, 320, 352, and 192 for H, N, C', and C<sub>α</sub>, respectively. The NUS sampling schedule was generated using exponential weighting with effective  $T_2$  times of 50, 50, and 20 ms for N, C', and C<sub>α</sub>, respectively. For the coupled spectra a cosine modulation of 52.5 Hz for  $T_{C'C_α}$  and 143.5 Hz for  $T_{C_αH_α}$  was taken into account for the schedule generation.

<sup>15</sup>N-HSQC spectra were acquired using an in-house sequence with 3-9-19 water suppression,<sup>[209]</sup> gradient filters, and <sup>13</sup>C decoupling during  $t_1$ . Decoupled triple resonance experiments were acquired using standard sequences `hncogpwg3d_sct` and `hncagpwg3d_sct`, featuring watergate water suppression,<sup>[210]</sup> gradient filters, and semi-constant time <sup>15</sup>N evolution. The coupled experiments were acquired using slight modifications of these sequences. Diagrams of all 3D pulse sequences are found in Figs. A.2 to A.4, and all sequences are found in the data collection (Appendix A.1).

All spectra were processed with NMRpipe<sup>[211]</sup> using zero-filling to twice the number of points and cosine-squared apodization. NUS reconstruction was done using MddNMR<sup>[212]</sup> using the CS-IRLS<sup>[213]</sup> algorithm. A sample of processing scripts can be found in the data collection (Appendix A.1). Peak picking and assignment was done in CCPNMR AnalysisAssign.<sup>[214]</sup>

**3.5c Conformational Search** The conformational search was done in MacroModel<sup>[215]</sup> using the 20 structures of 2KDU<sup>[169]</sup> as starting points. The search was done as a random sampling of the two backbone angles of residues 76–81. The minimization convergence criterion was 1 kJ mol<sup>-1</sup> Å<sup>-1</sup> and the energy cutoff was 3000 kJ mol<sup>-1</sup>. 18 backbone atoms within 8 Å of W489-C<sub>δ2</sub> were constrained using flat-bottom potential wells with 1 Å half-width and a force constant of 200 kJ mol<sup>-1</sup> Å<sup>-2</sup>. The four hydrogen bonds in the antiparallel β-sheets were constrained using MacroModel's FXHB operation code. A sample script for the search can be found in the data collection (Appendix A.1).

**3.5d Data Analysis** All computations and data analysis not mentioned elsewhere were performed using a series of self-made python scripts, which make extensive use of the SciPy ecosystem.<sup>[216–218]</sup> The paramagnetic data, the transformation matrices for generating the conformers, and the final, cross-validated ensembles can be found in the data collection (Appendix A.1).



## Chapter 4

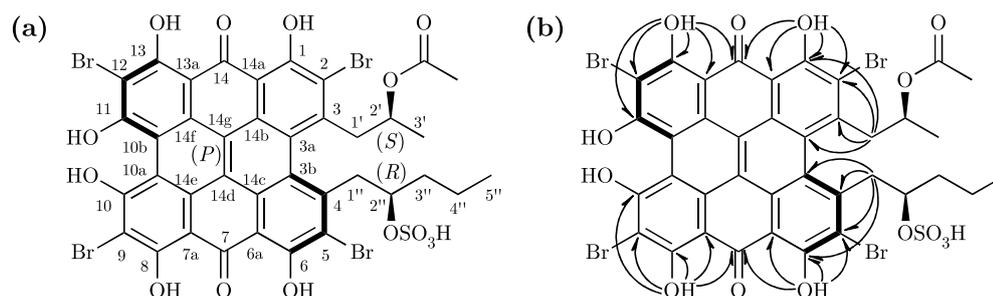
# Magnetic Alignment of Natural Products

In this chapter the structure determination of natural products using anisotropic parameters from magnetic alignment is described. It is based largely on work that was published in a journal article in 2020, and contains identical text passages and figures.<sup>[219]</sup> The isolation of gymnochrome G and the determination of its constitution as reported in this chapter was performed by the co-author Klaus Wolkenstein.

### 4.1 Introduction

As introduced in Chapter 1, anisotropic NMR can provide valuable information for the elucidation of the structure of unknown compounds such as natural products, and to acquire them, these unknown molecules have to be aligned. In the established protocols alignment is achieved by introducing the analyte into an anisotropic medium, such as liquid crystals, filamentous phages, or deformed polymer gels.<sup>[220–222]</sup> However, this process comes with certain drawbacks, especially if only small amounts of analyte are available (as it is common for natural products). First, any transfer of the analyte suffers from inevitable sample losses; second, recovery of the analyte from anisotropic media can be difficult; third, it prohibits the acquisition of any new experiments under normal, isotropic conditions for the time being; and fourth, the alignment tensor needs to be fitted and can yet not be sufficiently accurately predicted.

Therefore, we set out to obtain alignment without an alignment medium. A simpler way to achieve anisotropic conditions is by exploiting the alignment that is induced by the external magnetic field. This alignment occurs if the analyte molecule has an anisotropic magnetic susceptibility, and it scales with the square of the external magnetic field  $B_0$ .<sup>[223]</sup> It is most commonly used in the investigation of metal-binding proteins, where paramagnetic metal centers (e.g., transition metals or lanthanides) with substantially anisotropic susceptibilities lead to large degrees of alignment, and the isotropic reference experiment can be acquired on diamagnetic counterparts (for example, Zn, La, or Lu).<sup>[106,122,123,142,224,225]</sup> For diamagnetic cases, aromatic systems are the most relevant sources for anisotropic susceptibilities. RDCs and RCSAs can be determined by performing field-dependent measurements, but the effects are typically much smaller than for paramagnetic metal centers. RDCs have already been determined this way,



**Figure 4.1.** (a) Constitution, configuration, and atom numbering of gymnochrome G (**1**). (b) Relevant HMBC correlations for the assignment of the aromatic carbon resonances.

both for biomacromolecules such as proteins and oligonucleotides as well as for small molecules with large aromatic systems such as porphyrins.<sup>[106,226–230]</sup> However, to our knowledge, the measurement of RCSAs and the use of either RDCs or RCSAs for small molecule structure (conformation and configuration) determination through magnetic alignment have not been reported so far.

In this work, we show the feasibility of this approach on two molecular examples. First, gymnochrome G (**1**, Fig. 4.1) is a previously unknown marine natural product isolated from the deep-sea crinoid *Hypalocrinus naresianus*, featuring a large, proton-deficient aromatic system and two side chains with one stereocenter each. Its aromatic system with eight annealed rings leads to an unusually large degree of alignment and makes this an ideal target for this approach. We corroborate the configuration using established NMR-based techniques as ( $2'S, 2''R$ ) (see Section 4.2). Second, to show that this method is more universally applicable, we apply it to strychnine (**2**, Fig. 4.5), a well-studied alkaloid with only a single aromatic ring. Despite its relatively low degree of magnetic alignment, we were able to identify the correct configuration from the set of its diastereomers with high confidence.

## 4.2 Characterization of Gymnochrome G

Gymnochrome G is a natural product that was isolated from a specimen of *Hypalocrinus naresianus* along a variety of other compounds.<sup>[231,232]</sup> Before it could be used as a model compound for testing the approach of magnetic alignment, its structure had to be determined independently. This was achieved using various spectroscopic methods, including NMR, UV/Vis, ECD, and MS.

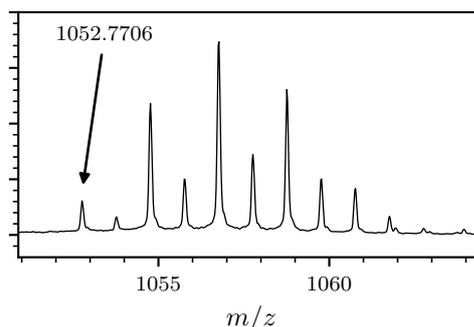
### 4.2.1 Constitution

Gymnochrome G was isolated from successive extracts of lyophilized animal material as a violet solid in a quantity of around 2.7 mg or 2.6  $\mu\text{mol}$ . The first characterization was done using UV/Vis, ECD, and HR ESI MS (Figs. A.10 and A.11). Brominated

phenanthroperylene quinones have previously been reported as occurring in crinoids, and the UV/Vis and ECD spectrum recorded for gymnochrome G was very similar to the reported spectra.<sup>[233–235]</sup> This was the first hint that the substance belonged to the family

$m/z = 1052$  and  $m/z = 1061$  with a monoisotopic mass of  $m/z = 1052.7706$  (Fig. 4.2), which was assigned to the deprotonated molecule ion  $[M-H]^-$ . The isotope pattern with a characteristic intensity ratio of 1:4:6:4:1 indicated the presence of four bromine atoms. The overlaid  $^{13}\text{C}$  isotope pattern enabled us to estimate the number of carbon atoms to be around 35–40. A fragment peak with  $\Delta m/z = -97.9678$ , corresponding to the mass of  $\text{H}_2\text{SO}_4$ , indicates the presence of a sulfate group. With these constraints the molecular formula could be determined to be  $\text{C}_{38}\text{H}_{26}\text{Br}_4\text{O}_{14}\text{S}$ .

For the structural characterization a number of basic NMR spectra were acquired, including  $^1\text{H}$  1D,  $^{13}\text{C}$  APT, COSY,  $^1\text{H},^{13}\text{C}$ -HSQC, and  $^1\text{H},^{13}\text{C}$ -HMBC spectra (Figs. A.12 to A.16). The carbon spectrum shows 38 peaks which confirmed the molecular formula determined previously, and which are compatible with the proposed phenanthroperylene quinone base structure. There are two peaks in the region around 185 ppm, corresponding to the quinone-carbonyl carbons, seven peaks in the region of 160–170 ppm which contain the six phenolic carbon atoms, and 20 signals in the aromatic region between 100–145 ppm. The remaining nine signals in the aliphatic region between 15–80 ppm belong to the two side chains, whose constitution could be straightforwardly determined using proton-based experiments to be propyl-2-acetate and pentyl-2-sulfate. The full structure including numbering is shown in Fig. 4.1. While the resonance assignment was again straightforward in the case of the side chains, it proved to be challenging in the case of the aromatic core. While in principle this could have been achieved using an INADEQUATE experiment,<sup>[236]</sup> which is based on one-bond carbon-carbon correlations, this was utterly unfeasible due to the very low sensitivity of this experiment. Large parts of the aromatic system could be assigned with an HMBC optimized for a small coupling constant of 3 Hz (Fig. A.16), making use of the correlations of the hydroxy protons in positions 1, 6, 8, and 13. These correlations are illustrated on the structure in Fig. 4.1b. However, the central carbons 10a–b and 14b–g showed no correlation and therefore remain unassigned. A list of proton and carbon chemical shift can be found in Table A.3.



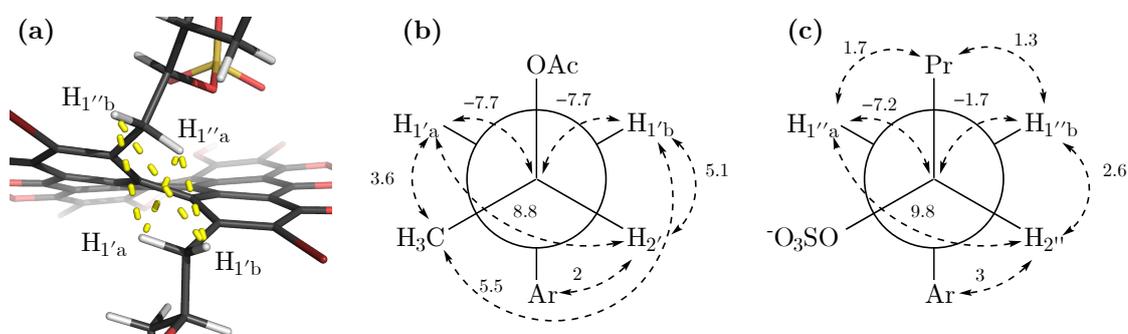
**Figure 4.2.** Molecular ion peak in the mass spectrum of gymnochrome G. The monoisotopic peak is annotated with its exact mass.

#### 4.2.2 Relative Configuration by *J*-Couplings and NOE

Gymnochrome G does not only exhibit chirality in its side chains with the chiral carbon centers. Its aromatic system is non-planar due to steric clashes of the side chains as well as the so called bay-OH groups at positions 10 and 11. The aromatic rings in the

**Table 4.1.**  $J$ -couplings of gymnochrome G relevant for the configuration determination, sorted by type of experiment and side chain.

	coupling	$J$ in Hz	coupling	$J$ in Hz
$J$ -HMBC	${}^3J_{C3a-H1'a}$	3	${}^3J_{C3b-H1''a}$	4
	${}^3J_{C3a-H1'b}$	5	${}^3J_{C3b-H1''b}$	5
	${}^3J_{C2-H1'a}$	8	${}^3J_{C5-H1''a}$	8
	${}^3J_{C2-H1'b}$	4	${}^3J_{C5-H1''b}$	4
	${}^3J_{C3-H2'}$	2	${}^3J_{C4-H2''}$	3
1D ${}^1H$	${}^3J_{H2'-H1'a}$	8.8	${}^3J_{H2''-H1''a}$	9.8
	${}^3J_{H2'-H1'b}$	5.1	${}^3J_{H2''-H1''b}$	2.6
HECADE	${}^3J_{C3'-H1'a}$	3.6	${}^3J_{C3''-H1''a}$	1.7
	${}^3J_{C3'-H1'b}$	5.5	${}^3J_{C3''-H1''b}$	1.3
	${}^2J_{C2'-H1'a}$	-7.7	${}^2J_{C2''-H1''a}$	-7.2
	${}^2J_{C2'-H1'b}$	-7.7	${}^2J_{C2''-H1''b}$	-1.7

**Figure 4.3.** (a) Visible intersidechain ROE contacts in gymnochrome G. (b,c) Newman projections along  $C2'-C1'$  (b) and  $C2''-C1''$  (c) with the experimental  $J$ -couplings in Hz used to determine the major conformer and the configuration of gymnochrome G.

corners are therefore bent out of plane, and the ring system exhibits axial chirality. This is very useful for the determination of the absolute configuration of the chiral centers at the side chains, since the aromatic system can be used as chiral reference. A comparison with the literature suggest that that the configuration should be propeller-( $P$ ),<sup>[237]</sup> which we will assume for now. This was later confirmed by comparison of experimental and predicted ECD (see Section 4.2.4). In the following we will show how to determine the configuration of the molecule's chiral carbon atoms relative to the aromatic system using distance and dihedral constraints.

A Rotating frame Overhauser Enhancement Spectroscopy (ROESY) spectrum was acquired to get distance information.<sup>[238]</sup> This was preferable over a conventional NOESY since the molecular weight of the compound and thus the rotational correlation time was such that the NOE effect was close to its zero-crossing. Information about dihedral angles

was extracted from proton-proton  ${}^3J_{\text{HH}}$  and long-range proton-carbon  ${}^nJ_{\text{CH}}$ -couplings. The former were simply extracted from the 1D proton spectrum. For the latter two different experiments were acquired: the HECADÉ is an experiment with an HSQC block followed by a TOCSY transfer and proton acquisition.<sup>[239]</sup> It is set up in such a way that it yields ECOSY-type crosspeaks with a splitting of  ${}^1J_{\text{CH}}$  in the indirect dimension and  ${}^nJ_{\text{CH}}$  in the direct dimension. This enables the measurement of very small  ${}^nJ_{\text{CH}}$ -couplings due to the large splitting in the indirect dimension, and the determination of their absolute sign as  ${}^1J_{\text{CH}}$  is always positive. The disadvantage of this type of experiment is that it can only measure couplings to non-quaternary carbon atoms due to the HSQC block. To get these couplings, a  $J$ -HMBC was acquired as well.<sup>[240]</sup> This experiment is different to a regular HMBC in that it features a splitting in the indirect dimension corresponding to  ${}^nJ_{\text{CH}}$ . However, since the smaller couplings are in the same order as the line width, the accuracy of these couplings is lower due to the overlap of the doublet components. All couplings are reported in Table 4.1. To account for the lower accuracy of couplings from  $J$ -HMBC they were rounded to integer Hz, while all other couplings were rounded to a tenth of a Hz.

These data were then used to determine the configuration and major conformation of the side chains. First, the only ROE contacts between the side chains came from the protons at positions 1' and 1''. This suggests that the side chains point away from each other to reduce steric interactions. This finding was confirmed in the conformational search performed for the DP4+ analysis (see Section 4.2.3), where all conformations within the energy window of 21 kJ mol<sup>-1</sup> adopted this arrangement. The side chain orientation and the key ROE contacts are illustrated in Fig. 4.3a. Now that the conformation about the rotatable bonds 3–1' and 4–1'' was known, the diastereotopic protons at 1' and 1'' could be assigned using their  ${}^3J_{\text{CH}}$ -couplings to the carbon atoms 2, 3a, 3b, and 5. These couplings follow a Karplus-type dependence on dihedral angles,<sup>[82,83]</sup> so for antiperiplanar arrangements (dihedral around 180°) larger couplings are expected than for synperiplanar arrangements (dihedral around 0°). The protons H1'a and H1'b are indexed in order of chemical shift. H1'a has a larger coupling to C2 than to C3a, while H1'b shows the opposite trend. Therefore H1'a must be facing inward (pro-*R*), while H1'b is facing outward (pro-*S*). The same argument applies to H1''a (pro-*R*) and H1''b (pro-*S*). Now that this assignment is established, the dihedral angles along the bonds 1'–2' and 1''–2'' can be analyzed, which is best illustrated on a Newman projection (Fig. 4.3b,c). H2' has a large coupling to H1'a and a small one to H1'b, so it is arranged antiperiplanar to the former and synclinal to the latter. C3' has the opposite trend in couplings, so it is arranged synclinal to H1'a and antiperiplanar to H1'b. The small coupling of H2' and C3 is in agreement with this arrangement. Finally, the  ${}^2J$ -couplings between H1'a,b and C2' confirm that the acetoxy group must be located at the only remaining position.  ${}^2J_{\text{CH}}$  are sensitive to the dihedral between the proton and an electronegative substituent (e.g., AcO) on the carbon atom. If the dihedral is close to 0°, the coupling is large and negative. If it is close to 180°, the coupling is small and positive. Both H1'a and H1'b have large, negative couplings to C2', so the acetoxy group must be arranged synclinal to both. This yields the configuration at the center C2' to be *S*. Using the same logic the configuration of C2'' can be determined to be *R*. Here the trends in the couplings are even more

**Table 4.2.** Comparison criteria of calculated ECD data of gymnochrome G.

Functional	PBE0	B3LYP	CAM-B3LYP
Scaling factor	0.25	0.26	0.42
Bandwidth $\sigma$ (eV)	0.17	0.17	0.30
UV shift (nm)	9	-1	43
Similarity factor $s$	0.80	0.81	0.69
Similarity factor $s$ (enantiomer)	0.04	0.07	0.05

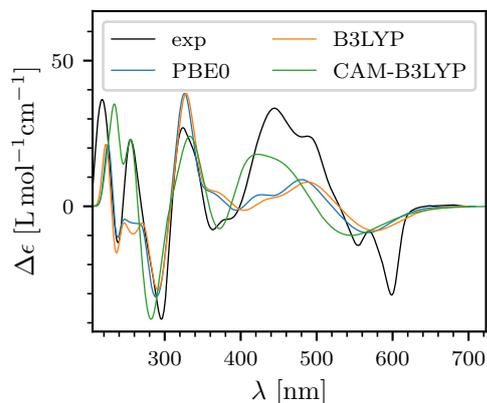
pronounced, which is an indication that there is even less rotational freedom around  $1''-2''$  than there is around  $1'-2'$ . This trend is also confirmed by the conformational search, where all conformations within  $21 \text{ kJ mol}^{-1}$  adopted this same rotamer around  $1''-2''$ .

### 4.2.3 Relative Configuration by DP4+

After determining the side chain configuration using ROE contacts and  $J$ -couplings, we decided to corroborate this result with DP4+ as an independent method. DP4 and DP4+ are probabilistic methods based on the comparison of experimental and predicted chemical shifts.<sup>[96,97]</sup> To predict chemical shifts, first a conformational search needed to be performed. The resulting geometries were optimized and chemical shieldings were calculated using density functional theory (DFT). The conformation's individual shieldings were averaged using a Boltzmann distribution, typically with zero-point corrected (ZPC) free energies from a frequency calculation. Conversion of shieldings into chemical shifts is typically done using some sort of empirical relationship. Grimblat et al.<sup>[97]</sup> performed this process on 72 different organic compounds and evaluated the statistical properties of the difference between calculated and experimental chemical shifts. With a set of experimental chemical shifts and predicted shifts for some structural candidates, a relative probability can be assigned to each of these candidates using the deviation of predicted and experimental shifts and the knowledge about their statistical distribution. We predicted shieldings for all four possible side chain configurations using the methods and levels of theory as required (for details see Section 4.7), and compared them to the experimental shifts using a spreadsheet provided by the authors. To have a result independent from the one using  $J$ -couplings and ROE, the diastereotopic methylene protons at  $1'$  and  $1''$  were assigned to the better fitting value for each configuration. We also only used shifts from the side chain as well as aromatic ipso- and ortho-carbon atoms, as the aromatic system is the same for all configurations and a full assignment was not available. The resulting DP4+ probabilities were 0.00 % for  $(2'R, 2''R)$ ,  $(2'R, 2''S)$ , and  $(2'S, 2''S)$ , and 100.00 % for  $(2'S, 2''R)$ , which confirmed our previous result with high certainty.

#### 4.2.4 Absolute Configuration

The absolute configuration of the aromatic system was confirmed by comparing the experimental electronic circular dichroism (ECD) spectrum with predictions from time-domain DFT, largely following the recommendations by Pescitelli et al.<sup>[241]</sup> One important recommendation is to use different types of functionals (with and without long-range corrections, and with different exchange fractions), so the three functionals PBE0,<sup>[242]</sup> B3LYP,<sup>[154,156,243,244]</sup> and CAM-B3LYP<sup>[245]</sup> were used. For these calculations the structures of the various possible conformations were necessary, and the same structures as in the calculation of anisotropic parameters were used (see Section 4.7). To reduce computational demand, only structures less than  $11.4 \text{ kJ mol}^{-1}$  above the ground state were used, which corresponds to about 1 % population at 298 K. To get transition energies, an



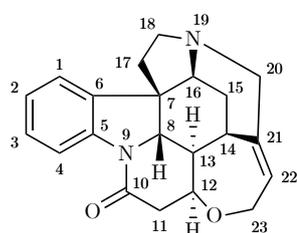
**Figure 4.4.** Comparison of the experimental ECD spectrum (black) and the predicted ECD spectra for the different functionals PBE0 (blue), B3LYP (orange), and CAM-B3LYP (green).

excited state calculation using 100 states was performed. The first calculation with the PBE0 functional was done on all four side chain configurations, which revealed that the side chains have a negligible effect on the predicted ECD spectrum. For the other two functionals the calculations were therefore only performed on the correct diastereomer ( $2'S, 2''R$ ). These calculations generate a list of transition energies and intensities. To generate a simulated spectrum and compare it to the experiment, SpecDis 1.71 was used.<sup>[246]</sup> This software performed the task of Boltzmann averaging the individual conformers using the zero-point corrected free energies, and also the calculation of the UV correction, the bandwidth  $\sigma$ , and the scaling factor. The UV correction is an empirical shift along the wavelength axis which accounts for systematic errors in the transition energies. The bandwidth is simply the line width of the individual transitions, and using a Gaussian line shape model the list of transitions is converted into a continuous spectrum. The scaling factor accounts for the fact that absolute transition intensities are generally inaccurate in ECD calculations. The predicted spectrum is generated with these three parameters and compared to the experiment, and SpecDis generates a similarity factor  $s \in [0, 1]$  for both enantiomers as a quantitative way to distinguish between the two. Note that the predicted spectra of the enantiomers are generated by simply changing the sign of  $\Delta\epsilon$ . The spectra without long-range corrections (PBE0 and B3LYP) show a slightly better agreement, but in all three cases both visual comparison (Fig. 4.4) as well as the similarity factors (Table 4.2) clearly confirm the configuration of the aromatic system to be (*P*).

### 4.3 Determination of Anisotropic Parameters

To increase the confidence in the result obtained from anisotropic NMR parameters, it is desirable to acquire as many of such constraints as possible. Magnetic field alignment is limited by the very small magnitude of the alignment, so only the strongest anisotropic parameters could be measured, which were  $^{13}\text{C}$  RCSAs, one-bond carbon-proton RDCs ( $^1D_{\text{CH}}$ ), and geminal proton-proton RDCs ( $^2D_{\text{HH}}$ ). For gymnochrome G, measurement of some vicinal RDCs ( $^3D_{\text{HH}}$ ) was also possible. Proton RCSAs are expected to be similar in size in Hz to carbon RCSAs, but bigger line widths and homonuclear couplings decrease the measuring accuracy, so it was not possible to acquire useful  $^1\text{H}$  RCSAs. Since the size of RDCs scales with the third inverse power of the internuclear distance, most long-range RDCs were also unfeasible to acquire.

#### 4.3.1 NMR Experiments



**Figure 4.5.** Constitution, configuration, and atom numbering of the alkaloid strychnine (**2**).

RCSAs manifest themselves as small changes in chemical shifts, so these have to be accurately measured. To do so, we acquired attached proton test (APT) spectra, which are a type of 1D  $^{13}\text{C}$  experiment using a  $J$ -modulated spin echo to include multiplicity information in the sign of the signal. Regular pulse-acquire sequences on  $^{13}\text{C}$  suffer from baseline distortions caused by pulse ring-down, especially in the cryogenically cooled, inverse-geometry probes used in our facility, and in experiments with a large number of scans.<sup>[247]</sup> In the APT sequence acquisition does not occur immediately after a pulse which prevents these baseline distortions. This is the main reason for the choice of this experiment to measure carbon chemical shifts. For RDCs, all nuclear pairs of interest also have a substantial scalar  $J$ -coupling, and the RDC is observed as a small change in the overall coupling. For gymnochrome G the proton-proton couplings were extracted simply from 1D  $^1\text{H}$  spectra, and carbon-proton couplings were not acquired. For strychnine we measured couplings using a CLIP-HSQC (clean-inphase-HSQC)<sup>[248]</sup> with perfect-echo INEPT periods<sup>[249]</sup> to suppress phase and amplitude modulations from homonuclear proton couplings. While proton-proton couplings could have been also been extracted from the 1D proton spectrum, the use of an HSQC for this purpose has the advantage of having almost no signal overlap due to the additional dimension. Only a few peaks showed overlap with long-range correlations of neighboring protons and had to be discarded. It was also important to bear in mind that couplings (and especially RDCs) are a signed quantity. As all RDCs were much smaller than the corresponding  $J$ -coupling, the sign of the total coupling  $T$  follows from the sign of  $J$ . All  $^2J_{\text{HH}}$ -couplings were assumed to be negative, all others positive.

In methods involving induced alignment anisotropic parameters are determined as the difference between values in isotropic and anisotropic, aligned conditions. In the case of magnetic field alignment there is always some degree of anisotropy as there is always

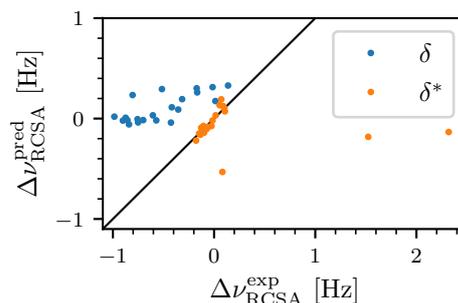
a non-zero magnetic field. The necessary approach is therefore to measure at different magnetic fields and then fit for the expected square-field dependence of the anisotropic parameters. We acquired the above-mentioned experiments at fields of 400 MHz (except for APT of **1**), 600 MHz, 700 MHz, 800 MHz, 900 MHz, and 950 MHz (except for APT of **2**).

### 4.3.2 Chemical Shift Correction

Since the desired anisotropic effects are very small, any source of error has to be carefully excluded. The chemical shift is very sensitive to the experimental conditions, so the perturbation of  $^{13}\text{C}$  chemical shifts could not be associated to alignment alone. As the spectra at different fields were acquired at different spectrometers on different days, we identified the following systematic errors for  $^{13}\text{C}$  chemical shifts: first, differences in sample temperature due to imperfections in temperature calibration and different decoupling powers on the order of 0.2 K; second, slow degradation of the sample over the course of time; and third, the referencing. These contributions could be eliminated using basic linear algebra. We postulate that all these errors can be described as linear perturbations as a function of temperature  $T$ , some degradation coordinate  $\xi$  (which is not necessarily proportional to time), and a constant  $c$  quantifying the amount of misreferencing:

$$\boldsymbol{\delta} = \boldsymbol{\delta}_{\text{corr}} + T\boldsymbol{\delta}_T + \xi\boldsymbol{\delta}_{\text{degrad}} + c\boldsymbol{\delta}_{\text{ref}}, \quad (4.1)$$

where the five vectors represent the measured chemical shift, the corrected chemical shift, and the chemical shift contributions from temperature, degradation, and misreferencing, respectively. Each assigned  $^{13}\text{C}$  shift corresponds to one component of these vectors, making them 30-dimensional for **1** and 21-dimensional for **2**. For temperature, we quantified the relative effect by acquiring spectra at the same field, but at different temperatures (297 K, 298 K, and 299 K for **1**, 298 K and 301 K for **2**). For each resonance, a linear fit of the form  $\delta = \delta_0 + T\delta_T$  was performed, and the resulting  $\delta_T$  were arrayed to the vector  $\boldsymbol{\delta}_T$ . For degradation, APT spectra at different time points, but equal fields were acquired. We acquired three pairs of spectra for **1** and two pairs of spectra for **2** with time differences between one and three months and calculated  $\boldsymbol{\delta}_{\text{degrad}}$  as the chemical shift difference of each resonance between these pairs. The fact that the resulting  $\boldsymbol{\delta}_{\text{degrad}}$  for different spectral pairs were close to parallel confirms that the assumption of linearity in Eq. (4.1) is a reasonable approximation for degradation. For practical application we recommend acquiring the data as quickly as possible, and to repeat the measurement of



**Figure 4.6.** Correlation plot of experimental and DFT-calculated RCSAs for the correct diastereomer of strychnine. While there is overall very poor agreement of the uncorrected data, in the corrected data all deviation gets concentrated into three defined points, which are then discarded.

the first field point at the end of the measurement series. This way the degradation effect during the measurement series is captured, requiring only one additional measurement and no extra waiting time. To give the reader an idea of the magnitude of these effects, the root mean square of the temperature dependence of the carbon chemical shifts was 3.3 ppb K<sup>-1</sup> for **1** and 5.9 ppb K<sup>-1</sup> for **2**, while for the degradation it was 3.7 ppb for **1** and 0.4 ppb for **2**. The correction of misreferencing was only necessary for **2**, supposedly due to small temperature and degradation effects on the reference shift (TMS), and the smaller size of the desired RCSAs.  $\delta_{\text{ref}}$  is trivial to find as every peak is shifted by the same amount and therefore all elements of  $\delta_{\text{ref}}$  are equal (e.g., 1). Now, instead of attempting to correct the measured chemical shifts by determining the exact values of  $T$ ,  $\xi$ , and  $c$  for each spectrum, another approach proved to be much more effective. We performed an orthonormal transformation of the chemical shift vectors, using the temperature, degradation, and referencing vectors as the first basis vectors. This can be done by arraying these perturbation vectors into an  $m \times n$  matrix  $\mathbf{A}$ , where  $m$  is the number of resonances and  $n$  is the number of perturbation vectors, and a full  $\mathbf{QR}$  decomposition of this matrix is performed. Note that only the direction and not the magnitude of the perturbation vectors is relevant, as they are normalized during the process of  $\mathbf{QR}$  decomposition. The orthogonal matrix  $\mathbf{Q}$  can now be used to perform the above-mentioned orthonormal transformation of chemical shift vectors  $\delta$  by matrix multiplication  $\delta^* = \mathbf{Q}^T \delta$ . Since the perturbation vectors are the first basis vectors of the new system, all following base vectors are orthogonal to these vectors and are therefore not affected by them. If Eq. (4.1) is expressed with chemical shift vectors  $\delta^*$  in the new basis, it changes as follows:

$$\delta^* = \delta_{\text{corr}}^* + T \begin{pmatrix} \delta_{T,1}^* \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix} + \xi \begin{pmatrix} \delta_{\text{degrad},1}^* \\ \delta_{\text{degrad},2}^* \\ 0 \\ 0 \\ \vdots \end{pmatrix} + c \begin{pmatrix} \delta_{\text{ref},1}^* \\ \delta_{\text{ref},2}^* \\ \delta_{\text{ref},3}^* \\ 0 \\ \vdots \end{pmatrix}. \quad (4.2)$$

It is easy to see that only the first three components are affected by temperature, degradation, and misreferencing, which have to be discarded. The remaining components are now free from these systematic errors and could be used for structure determination. The success of this approach becomes obvious when inspecting a correlation plot of experimental and predicted RCSAs with and without basis transformation, which is shown in Fig. 4.6. Without the transformation, there is little correlation between the data. After the transformation, all the systematic error is concentrated into three points, and the other ones show very good agreement.

### 4.3.3 Field Dependence Fit and Weighting Factors

The RDCs and RCSAs are the square-field dependent components of the total coupling and the chemical shift, respectively, and they are found by a fitting procedure as outlined in the following. To determine RCSAs, the corrected chemical shifts were expressed as

### 4.3 Determination of Anisotropic Parameters

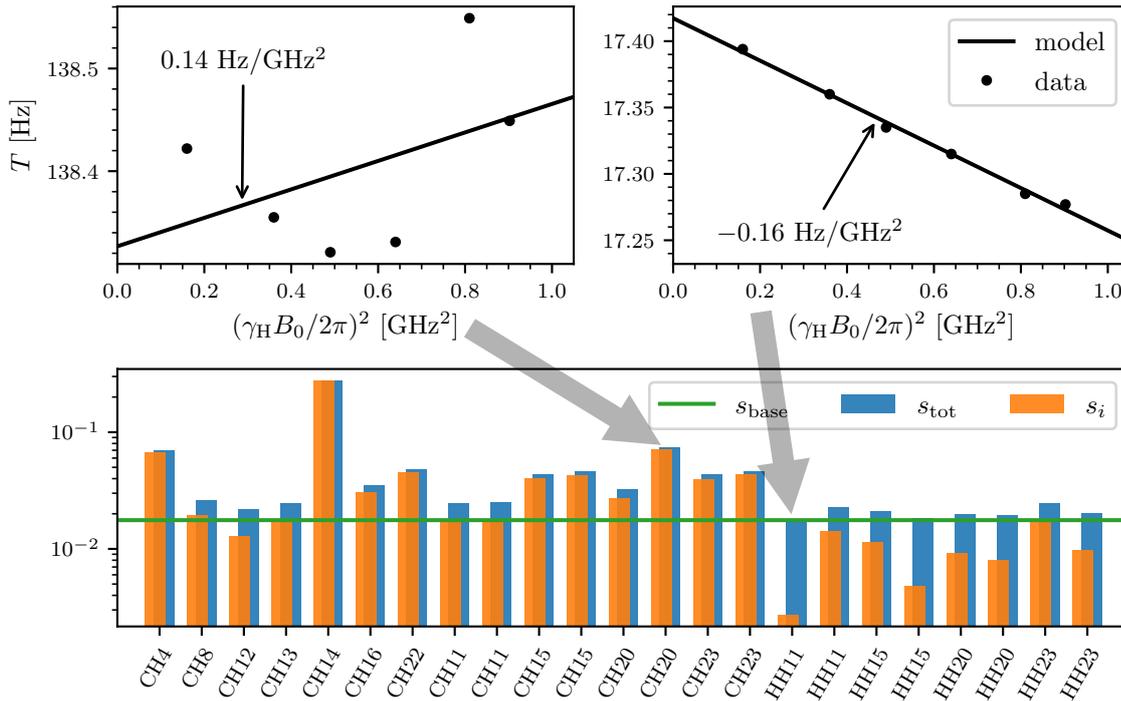
peak positions  $\nu$  in Hz with  $\nu = \delta\nu_{\text{ref}}$ , which obey the following relation:

$$\nu = \delta_{\text{iso}}\nu_{\text{ref}} + \delta_{\text{aniso}}^B\nu_{\text{ref}}B_0^2 = \delta_{\text{iso}}\gamma B_0 + \delta_{\text{aniso}}\gamma B_0^3. \quad (4.3)$$

In this context, the unknown parameters  $\delta_{\text{iso}}$  and  $\delta_{\text{aniso}}^B$  correspond to the isotropic chemical shift in ppm and to the desired RCSA in ppm T<sup>-2</sup>, respectively. The superscript  $B$  illustrates that  $\delta_{\text{aniso}}^B$  dimensionally contains the inverse square of the field. As both parameters appear in linear form in Eq. (4.3), the least-squares solution to a given set of  $(\nu, B_0)$  data can be determined easily and deterministically. Similarly, the measured, total coupling  $T$  obeys the following relation as a function of the field:

$$T = J + D^B B_0^2. \quad (4.4)$$

Here,  $J$  and  $D^B$  correspond to the  $J$ -coupling in Hz and to the RDC in Hz T<sup>-2</sup>, respectively, and can again be determined as least-squares solution to a given set of  $(T, B_0)$ . As mentioned above,  $\delta_{\text{aniso}}^B$  and  $D^B$  have units of ppm T<sup>-2</sup> and Hz T<sup>-2</sup>, respectively, but for all subsequent data interpretation they were converted into frequency units by evaluating the anisotropic components of Eqs. (4.3) and (4.4) at  $B_0 = 23.49$  T (1 GHz proton frequency). While it is a somewhat arbitrary choice, it enables the simultaneous



**Figure 4.7.** Two examples of the field dependence of the total coupling  $T$  in strychnine. While the  ${}^2T_{\text{HH}10}$ -coupling shows excellent agreement with the model (top right), the  ${}^1T_{\text{CH}20}$  is affected by large deviations. The devised weighting method using the inverse of  $s_{\text{tot}}$  (bottom, blue bars) ensures that the  ${}^2D_{\text{HH}11}$ -RDC does not get overvalued, but it appropriately devalues the  ${}^1D_{\text{CH}20}$ -RDC.

interpretation of RDC and RCSA. For gymnochrome G, we give them both the same weight since at this field the RDCs are in size between aliphatic and aromatic  $^{13}\text{C}$  RCSAs, while for strychnine, a more sophisticated weighting procedure was devised (see below). For strychnine,  $^1D_{\text{CH}}$ -couplings from the two protons of a methylene group were evaluated as their mean value to remove the need for diastereotopic assignment. A given geminal  $^2T_{\text{HH}}$ -coupling was determined individually from each proton peak and averaged.

While for the chemical shifts the main error contribution was corrected as outlined above, the issues with couplings were less predictable and less easy to mitigate. The biggest problem were peak distortions due to overlap with long-range correlation as well as strong coupling effects. While some of the more obvious outliers could have been sorted out manually, this introduces unwanted human bias. Instead, we devised a weighting strategy for each point based on the agreement with Eqs. (4.3) and (4.4). Although this was originally developed with the RDCs in mind, we applied it to both RDCs and RCSAs, since the determination of weighting factors from uncertainties for both types of parameters is a general way to make simultaneously evaluation possible. In the following, we illustrate the procedure on the example of RDCs.

All data points are affected by random scatter, and the uncertainty in the determined anisotropic component is proportional to this scatter. In principle, the scatter of each field-series of couplings can be quantified by calculating the RMSD  $s_i$  of the fit to Eq. (4.4):

$$s_i = \sqrt{\frac{1}{n_j} \sum_j (T_{\text{meas}}^{ij} - T_{\text{model}}^{ij})^2}, \quad (4.5)$$

where the indices  $i$  and  $j$  refer to the coupling (atom pair) and the field, respectively, and  $n_j$  is the number of fields. Since there are only six data points (fields) along  $j$ , there is a significant uncertainty associated with this. If this RMSD is used as a weighting factor to the corresponding RDC, data points that by chance have a very low RMSD will be overvalued (Fig. 4.7, top right). However, since some few couplings are affected by stronger peak distortion (Fig. 4.7, top left), there was the need to devalue them relative to the more accurate RDCs, i.e., determine individual weighting factors. We made the following assumption: All couplings are equally affected by the same (random) base scatter. Some few couplings are then additionally affected by individual deviations due to peak distortions etc. If there were no individual deviations, the base scatter could be estimated by taking the RMS of the RMSD of all field-series fits. Some of these field-series are affected by additional deviations which lead to outliers in their corresponding RMSD, which greatly affects this RMS of RMSD. To cope with that, we estimate the base scatter  $s_{\text{base}}$  by calculating the root median square instead:

$$s_{\text{base}} = \sqrt{\text{median}(s_i^2)}. \quad (4.6)$$

Now, we estimate the individual error as the RMSD of the field-series fit and calculate the total error  $s_{\text{tot}}$  as RMS of base and individual error. We use the inverse of the error

as weight  $w_i$ :

$$w_i = \frac{1}{s_{\text{tot}}} = \frac{1}{\sqrt{s_i^2 + s_{\text{base}}^2}}. \quad (4.7)$$

This enables a smooth transition between small and large errors, prevents the overvaluation of data points, and enables to give smaller weights to data points with large error. In Fig. 4.7 the different error components  $s_i$ ,  $s_{\text{base}}$ , and  $s_{\text{tot}}$  are illustrated for the RDCs of strychnine as a logarithmic bar plot.

## 4.4 Data Evaluation and Discrimination of Stereomers

Generally, anisotropic NMR data do not provide individually interpretable structural constraints, but they are used in their entirety and validated against different structural models. These models are generated for each possible configuration using molecular modeling.

### 4.4.1 Model Generation

In the case of **1**, the aromatic system was fixed to its known configuration ( $P$ ) previously determined via ECD (Section 4.2). Therefore, the two side-chain stereocenters  $2'$  and  $2''$  produce four possible diastereomers. We conducted an MM-based conformational search of the side-chain torsions to generate a conformational ensemble for each diastereomer, containing between 13 and 33 conformers. For **2**, we assumed no conformational flexibility. The multiple fused and bridged cycles of **2** allow only a subset of the theoretically possible 32 diastereomers, and we were able to build 22 configurations. Although some of these have strongly distorted carbon binding geometries and could have been discarded as they are unlikely to be stable, we purposefully left them in the analysis as this further demonstrates the discriminating power of the method. We cross-checked the resulting geometries with the structures of Bifulco et al.,<sup>[250]</sup> who conducted a thorough study of the possible diastereomers of strychnine. All conformations for both **1** and **2** were geometry-optimized and chemical shift tensors as well as magnetic susceptibility tensors were predicted via DFT. For technical details about these calculations see Section 4.7. These molecular models are finally used to predict a set of anisotropic data which is compared to the experiment.

### 4.4.2 Prediction of Anisotropic Data

The RDCs and RCSAs can be calculated from the expressions found in Eqs. (2.53) and (2.59) and are repeated for the reader below:

$$\Delta\nu_{\text{RDC}} = -\frac{3\mu_0\gamma_1\gamma_2\hbar}{8\pi^2r_{12}^5}\text{tr}\left(\mathbf{r}_{12}\mathbf{r}_{12}^{\text{T}}\mathbf{A}\right), \quad (4.8)$$

$$\Delta\nu_{\text{RCSA}} = \nu_{\text{ref}}\text{tr}\left(\delta\mathbf{A}\right). \quad (4.9)$$

In the case of magnetic alignment the alignment tensor  $\mathbf{A}$  can also be calculated from molecular properties (derivation in Section 2.2.2):

$$\mathbf{A} = \frac{B_0^2}{15\mu_0 k_B T} \left( \boldsymbol{\chi} - \frac{\text{tr}(\boldsymbol{\chi})}{3} \mathbf{1} \right). \quad (4.10)$$

Since all quantities are either constants ( $\gamma, \mu_0, \hbar, k_B$ ), known experimental parameters ( $T, B_0$ ), or molecular properties that can be calculated via DFT ( $\mathbf{r}_{12}, \boldsymbol{\chi}, \boldsymbol{\delta}$ ), theoretical RDC and RCSA can be calculated by simply inserting all constants and parameters into Eqs. (4.8) to (4.10). In the multiconformer case of  $\mathbf{1}$ , Boltzmann averaging is done as a final step with the anisotropic parameters.

In the more widespread case of measuring anisotropic NMR with induced alignment, the prediction of the alignment tensor is typically not possible and it has to be determined using a fitting procedure. To assess whether the possibility to predict the alignment is an advantage of our method, we also determined the alignment tensor via a fitting procedure to compare both approaches. To do so, Eqs. (4.8) and (4.9) have to be brought into a form where the linearity of the components of  $\mathbf{A}$  is apparent. It is helpful to realize that both equations have the exact same form:

$$\Delta\nu = \text{tr}(\mathbf{MA}). \quad (4.11)$$

For  $\Delta\nu_{\text{RDC}}$  it is  $\mathbf{M} = -(3\mu_0\gamma_1\gamma_2\hbar)/(8\pi^2r_{12}^5)\mathbf{r}_{12}\mathbf{r}_{12}^T$ , while for  $\Delta\nu_{\text{RCSA}}$  it is obviously  $\mathbf{M} = \boldsymbol{\delta}$ . By using the fact that  $\mathbf{A}$  and  $\mathbf{M}$  are real symmetric this can be expanded and rearranged into

$$\begin{aligned} \Delta\nu = & \frac{1}{3}(A_{xx} + A_{yy} + A_{zz})(M_{xx} + M_{yy} + M_{zz}) \\ & + \frac{1}{6}(2A_{zz} - A_{xx} - A_{yy})(2M_{zz} - M_{xx} - M_{yy}) \\ & + \frac{1}{2}(A_{xx} - A_{yy})(M_{xx} - M_{yy}) + 2A_{xy}M_{xy} + 2A_{xz}M_{xz} + 2A_{yz}M_{yz}. \end{aligned} \quad (4.12)$$

The reason that the products of diagonal elements are rearranged in this way is that now the first summand can be discarded since  $\mathbf{A}$  is traceless, i.e.,  $A_{xx} + A_{yy} + A_{zz} = 0$ . What is left is a linear equation of five independent components of  $\mathbf{A}$ . This becomes an (ideally overdetermined) system of linear equations if all data points  $\Delta\nu$  (RDC/RCSA) with their individual  $\mathbf{M}$  are taken together, and the least-squares solution of this system yields the fitted alignment tensor. This fitted alignment tensor can then in return be used to back-calculate the anisotropic data. For  $\mathbf{1}$ , there is the additional complication of multiple conformers. While in principle each conformer has its individual alignment tensor, this would increase the number of free variables beyond feasibility, but it is also not necessary. This is because the main source of alignment (i.e., the aromatic system) has no conformational freedom. Therefore, we can ignore these differences and apply a single tensor approximation. To do so, we use a common frame defined by the aromatic system for all conformers. In practice this is done by defining the vectors

$\mathbf{v}_1 = \overrightarrow{\text{C5C12}}$ ,  $\mathbf{v}_2 = \overrightarrow{\text{C2C9}}$ ,  $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$ , orthonormalize them, and use them as base vectors for a common Cartesian coordinate system. In this common frame, we calculate the components of the matrices  $\mathbf{M}_k$  for each conformer  $k$  and apply the Boltzmann averaging to them. Note that it would be a mistake to average over coordinate vectors as they contribute quadratically to the RDC coefficient matrix. The averaged  $\mathbf{M}$  is then used in the alignment tensor fit (Eq. (4.12)) to determine  $\mathbf{A}$ . For **2**, the weighting has to be included in this tensor fitting. This is achieved simply by multiplying both sides of Eq. (4.12) with  $w_i$  for each data point  $i$ .

### 4.4.3 Comparison of Experimental and Calculated Data

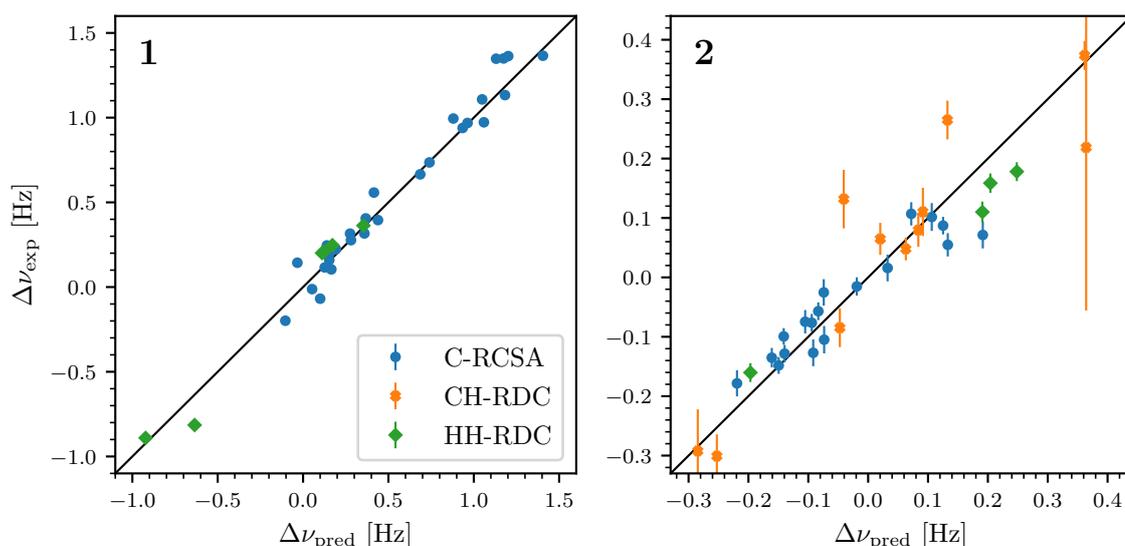
Both ways of determining the alignment tensor (predicted or fitted) yield a set of calculated data for each configuration, which is compared with the experimental data to identify the best match. This comparison is done in the same way for both back-calculated and DFT-predicted data sets. As primary criterion for the agreement of the calculated and experimental data we use the  $Q$ -factor,<sup>[251]</sup> which is the RMSD between the two, scaled by the RMS of the experimental data:

$$Q = \sqrt{\frac{\sum_i w_i^2 (\Delta\nu_{\text{exp},i} - \Delta\nu_{\text{calc},i})^2}{\sum_i w_i^2 \Delta\nu_{\text{exp},i}^2}}. \quad (4.13)$$

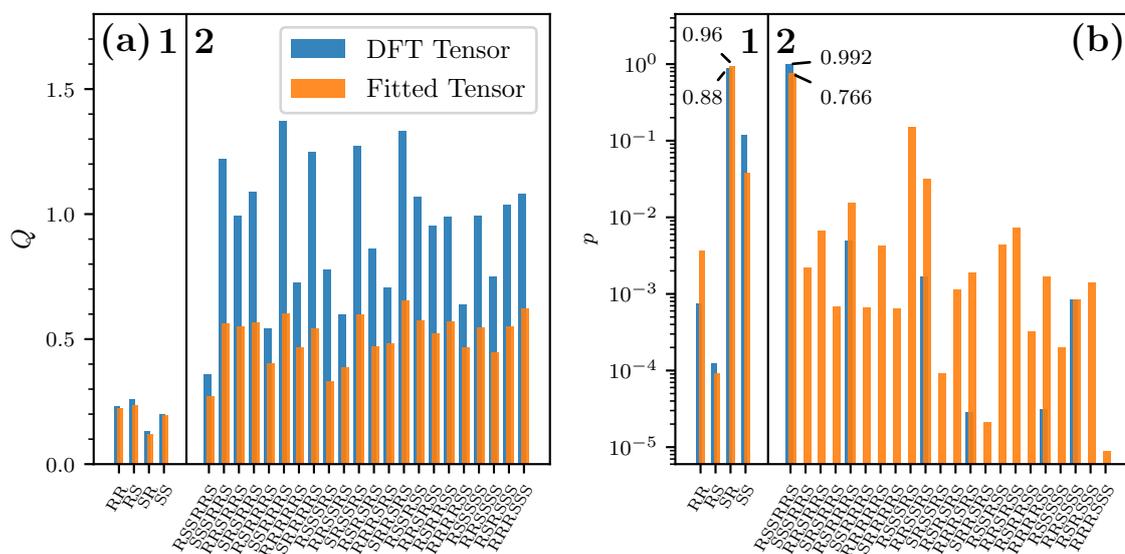
The configuration whose calculated data gives the lowest  $Q$ -factor is assumed to be the correct configuration. However, this approach gives no quantitative estimate for the confidence of the result since differences in  $Q$ -factors cannot be directly interpreted as a measure for the confidence. We use a bootstrapping test to examine the statistical properties of the data and get an insight on the confidence of our result.<sup>[201]</sup> Bootstrapping has the advantage that no assumptions about source, size, and distribution of error have to be made. In this procedure, resampled data sets of the same original size are generated by drawing randomly with replacement from the experimental data. These resampled sets are then each subjected to the same analysis as the original data set. The DFT-predicted data has to be simply re-matched to the sampled experimental data points. However, the procedure of fitting the alignment tensor and back-calculating the data has to be repeated for each resample. By repeating this resampling many times (in our case  $10^6$ ) one gets a distribution of any quantity that results from the data evaluation. As mentioned before, we assume the correct configuration to be the one with the lowest  $Q$ -factor, so we did this for each resample. The fraction of resamples that yields a certain configuration is then interpreted as confidence  $p$  in this result. Both  $Q$ -factors and confidences are illustrated in Fig. 4.9 as well as tabulated in Tables A.4 and A.5.

### 4.4.4 Results

First, it is noteworthy that the predicted anisotropic data are in excellent agreement with the experimental data, especially considering that the predicted data are purely obtained by DFT and contain no empirical or fitted contribution. A correlation of



**Figure 4.8.** Correlation plot of the experimental and DFT-predicted anisotropic data for the correct diastereomers of **1** (left) and **2** (right). The error bars in the plot for **2** correspond to the combined error  $s_{\text{tot}}$ . The values are calculated for a field of 23.49 T (1 GHz proton frequency).

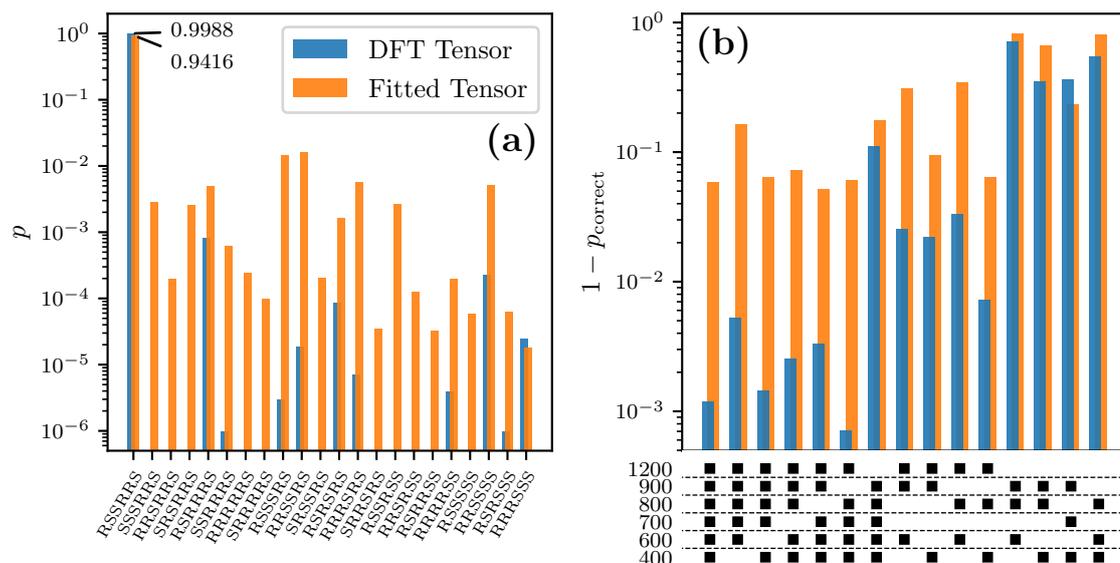


**Figure 4.9.** (a)  $Q$ -factors for the different diastereomers of **1** and **2**. The values obtained from predicting the alignment tensor with DFT are shown in blue, the ones obtained from the fitted alignment tensor are shown in orange. In all cases, the true configuration is correctly identified by the lowest  $Q$ -factor. (b) Confidences  $p$  from bootstrap resampling for the two different approaches of data evaluation, plotted on a logarithmic scale.

experimental and predicted anisotropic data that illustrates this is shown in Fig. 4.8. Figure 4.9 shows the  $Q$ -factors for the possible configurations of **1** and **2** for both methods of obtaining the alignment tensor. In all cases, the true solution ( $(2'S, 2''R)$  for **1**,  $(7R, 8S, 12S, 13R, 14R, 16S)$  for **2**) is correctly identified by the lowest  $Q$ -factor. Generally, the  $Q$ -factor is lower for fitted compared to predicted alignment tensors, as more fitting parameters always lead to smaller residuals. Yet, more important than the absolute values of  $Q$  are the power of the two approaches to discriminate configurations. In the case of **2**, the use of predicted alignment tensors clearly increases the degree of discrimination, as the  $Q$ -factors of the incorrect configurations increase much more than the one of the correct diastereomer: the ratio between the second-smallest and the smallest  $Q$ -factor is only 1.22 for the fitted tensor approach, while it improves to 1.57 by using the predicted tensor. By using the results from the bootstrapping analysis, the increase in confidence of the result for **2** by using predicted alignment becomes obvious, as the confidence increases from around 77% to 99.2% (see Fig. 4.9). This trend has a simple explanation: when the alignment tensor is fitted, it gives the system some leeway when fitting the data to an incorrect structure, thus reducing the  $Q$ -factor and therefore improving the agreement for this incorrect model. For **1** however, the confidence goes from 96% (fitted) to 88% (predicted), so there seems to be no advantage in using the predicted tensor. We hypothesize that this is caused by the structural similarity of the possible configurations, which all share the large aromatic system that is the source of alignment. The many data points available in this system with comparatively large anisotropic effects already characterize the alignment tensor very well, while there are relatively few data points and marginal contribution to alignment in the relevant side-chain regions. The data from the aromatic ring therefore dominates the fitting procedure and the differences between the four diastereomers are small. Also, there is more uncertainty in the structural model of **1** due to the conformational flexibility of the side chains, that is, the regions containing the stereocenters and the conformational averaging using calculated free energies. It is plausible that this uncertainty masks the expected improvement from the predicted alignment tensor compared to the fitted tensor. If the energy is calculated in a different way, for example, without thermal corrections or with a different basis set, the confidences change in the order of 10%.

## 4.5 Addition of 1200 MHz

While it was possible to use anisotropic NMR data from magnetic field alignment to distinguish between diastereomers, the small size of the effects remained the main challenge of this approach. As the degree of magnetic alignment is proportional to the square of the magnetic field, this methodology should greatly profit from the use of even higher fields in NMR. As a 1200 MHz NMR spectrometer became available in our laboratory shortly after the publication of the previous results,<sup>[219]</sup> it was natural to test how the inclusion of this ultra-high field instrument could potentially improve the discrimination of this method. We therefore repeated the measurement of  $^{13}\text{C}$ -RCSAs and RDCs on the example of strychnine (**2**) using instruments at frequencies of 400, 600, 700, 800, 900,



**Figure 4.10.** (a) Confidences  $p$  from bootstrap resampling for the two different approaches of data evaluation for **2** using 1200 MHz data. (b) Sum of confidences for the incorrect diastereomers ( $1 - p_{\text{correct}}$ ) for different combinations of field data used.

and 1200 MHz. The inclusion of the 1200 MHz instrument did not come without its own trade-offs and issues. First, at this field so far only 3 mm probeheads are available, forcing us to prepare our sample in a 3 mm NMR tube. This led to a reduced signal-to-noise ratio of our experiments and as such a reduced accuracy of our shifts and couplings, most notably at the 400 MHz instrument, which was not equipped with a cryogenically cooled probehead. Second, the experiments on the 1200 MHz instrument were performed only weeks after the charging of the superconducting magnet, and the resulting shim instabilities led to increased line widths. The data evaluation was performed exactly as with the previous data with fields up to 950 MHz so that both results remain comparable. And indeed the confidences in the result improved significantly from 99.2 % to 99.88 % using the predicted alignment tensor (see Fig. 4.10a). That corresponds to a reduction in the remaining uncertainty ( $1 - p_{\text{correct}}$ ) of almost an order of magnitude. We also investigated the effect of using only a subset of fields on the final result to assess how many different fields are necessary to get to a certain result, which is illustrated in Fig. 4.10b. First we repeated the data evaluation after removing the data from a single field. Here it became obvious that the data point from 1200 MHz is by far the most important one, and without it the uncertainty goes up to 10 %. By removing any other data point the uncertainty stays below 0.5 %. We also tested different combinations of three field points to see with how few data the approach remains feasible. This expectedly led to an increase in the uncertainty, presumably because the weighting approach relies on having a sufficient number of points to assess the agreement with the expected square-field dependence of RDCs and RCSAs (see Section 4.3). Here the data point at 1200 MHz again proved to be crucial, as all the sets without it reached uncertainties

between 35–70%, which effectively means that the correct diastereomer could not be determined. The three-point data sets that contained the point from 1200 MHz however all had uncertainties between 0.7–3.3%, which is an expected increase compared to using more data, but still identifies the correct diastereomer with reasonable certainty.

## 4.6 Conclusion and Outlook

We have demonstrated how both RDCs and  $^{13}\text{C}$ -RCSAs caused by magnetically induced alignment can be acquired, including a way to efficiently remove systematic errors from chemical shifts which is crucial for the measurement of the very small RCSAs, and an approach for weighting the individual data points. These data are accurate enough to successfully distinguish between all different diastereomers of our sample molecules. Second, we have shown that anisotropic parameters can be fully and reliably predicted using DFT, and how this can significantly improve the discriminating power in structure determination. For multiconformer cases, this approach eliminates the need for a common conformer frame, so fewer additional assumptions have to be made. The feasibility of our approach does not directly depend on the degree of alignment, but more on the differences in anisotropic parameters between the diastereomers. As these can be accurately predicted, the feasibility for any given molecule can be evaluated with inexpensive DFT calculations before investing expensive NMR time. The molecule strychnine is certainly near the limit of what we assume to be feasible at this point; we have calculated expected anisotropic parameters for various non-aromatic compounds (e.g., sugars) and they are typically about an order of magnitude smaller than for an aromatic compound such as strychnine. But also the geometrical differences of a set of diastereomers obviously affect how well they can be discriminated. Lastly, the availability of high magnetic fields plays an important role. Since the magnitude of anisotropic effects scales with the square of the field, while the measuring accuracy is (in first approximation) constant, having very high fields such as 1200 MHz greatly enhances the feasibility of our approach. Since even strychnine with only a single aromatic ring can be subjected to the magnetic alignment method, we conclude that it is an elegant alternative to alignment media for the determination of the relative configuration of flexible natural products.

## 4.7 Material and Methods

### 4.7.1 Characterization of Gymnochrome G

**4.7.1a Spectroscopic Experimental Procedures** UV spectra were recorded on a Jasco V-630 UV-visible spectrophotometer, ECD spectra were recorded on a Jasco J-810 spectropolarimeter, and IR spectra were measured on a Jasco 4100 FT-IR spectrometer equipped with a Pike Gladi ATR (attenuated total reflection) accessory. 1D and 2D NMR spectra were recorded in  $\text{MeOH-}d_3$  at 298 K on Bruker Avance III HD and Avance Neo spectrometers at 800 and 900 MHz equipped with TCI CryoProbes. Chemical shifts were referenced using residual solvent peaks ( $\text{MeOH}$ :  $\delta_{\text{H}} = 3.31$  ppm,  $\delta_{\text{C}} = 49.0$  ppm).

Standard Bruker pulse sequences (zgpr, jmod, hsqcetgpsp.2, hmbcetgpl2nd, hmbcetgpjcl2nd, hsqcdietgpjcdnsisp) were used.  $^1\text{H}$ ,  $^{13}\text{C}$ -HMBC experiments were optimized for  $^nJ_{\text{CH}} = 8\text{ Hz}$ , as well as  $^nJ_{\text{CH}} = 3\text{ Hz}$  for long-range correlations. Long range CH-couplings to quaternary carbons were measured using  $J$ -HMBC experiments,<sup>[240]</sup> all other signed CH-couplings were determined using a HSQC-HECADE-type sequence.<sup>[239]</sup> High-resolution MS spectra were obtained using a Bruker micrOTOF mass spectrometer with electrospray ionization in the negative-ion mode. HPLC was performed on an Agilent 1200 Series system using a Phenomenex Gemini C18 column ( $250 \times 10\text{ mm i.d.}$ ,  $5\ \mu\text{m}$ ).

**4.7.1b Animal Material, Extraction, and Isolation** Two specimen of *H. naresianus* were collected from Shima Spur, Kumano-nada Sea, Japan from depths of 763–852 m. Voucher samples were deposited in the collection of the Systematische Zoologie am Museum für Naturkunde Berlin (ZMB Ech 7415 and ZMB Ech 7416).<sup>[231]</sup> Lyophilized *H. naresianus* material (9.0 and 10.1 g) was successively extracted with MeOH/ $\text{CH}_2\text{Cl}_2$  (1:1), MeOH/ $\text{H}_2\text{O}$  (9:1), MeOH/ $\text{H}_2\text{O}$  (1:1), and distilled water. The methanol-soluble pigments of the MeOH/ $\text{CH}_2\text{Cl}_2$  extract were combined with the other extracts and subjected to semipreparative HPLC using a linear gradient of MeCN/20 mM aqueous ammonium acetate (45:55) to 85% MeCN. Fractions were purified and concentrated using solid phase extraction (Bondesil C18,  $40\ \mu\text{m}$ ). Pigments were washed with water and eluted with MeOH/ $\text{H}_2\text{O}$  (9:1) followed by evaporation of the eluates to dryness to give compound **1** in addition to several anthraquinone and biaryl pigments as well as further phenanthroperylene quinones reported elsewhere.<sup>[231,232]</sup>

**4.7.1c Molecular Data** violet solid (2.7 mg); UV (MeOH)  $\lambda/\text{nm}$  ( $\log \epsilon/\text{L mol}^{-1}\text{ cm}^{-1}$ ) 220 (4.64), 235 (4.65), 257 (4.59), 300 (4.46), 331 (4.37), 404 (3.86), 496 (4.07), 553 (4.21), 597 (4.50); ECD (MeOH)  $\lambda/\text{nm}$  ( $\Delta\epsilon/\text{L mol}^{-1}\text{ cm}^{-1}$ ) 218 (+36.65), 238 (−12.43), 255 (+22.91), 296 (−38.76), 324 (+26.98), 363 (−8.03), 445 (+33.67), 488 (+24.08), 555 (−13.44), 599 (−30.47); IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$  1570, 1451, 1238, 1128;  $^1\text{H}$  NMR (MeOH- $d_3$ , 800 MHz), see Table A.3;  $^{13}\text{C}$  NMR (MeOH- $d_3$ , 200 MHz), see Table A.3; HRESIMS  $m/z$  1052.7706  $[\text{M}-\text{H}]^-$  (calcd for  $\text{C}_{38}\text{H}_{25}\text{Br}_4\text{O}_{14}\text{S}$ , 1052.7704), 954.7986  $[\text{M}-\text{H}_3\text{SO}_4]^-$  (calcd for  $\text{C}_{38}\text{H}_{23}\text{Br}_4\text{O}_{10}$ , 954.8030), 525.8836  $[\text{M}-2\text{H}]_2^-$  (calcd for  $0.5 \times \text{C}_{38}\text{H}_{24}\text{Br}_4\text{O}_{14}\text{S}$ , 525.8816).

**4.7.1d Computational Methods** The starting geometries of gymnochrome G were built in Maestro 11.4.<sup>[252]</sup> For all configurations, the quinoid carbonyl acted as H-bond acceptor for the 1-, 6-, 8-, and 13-OH, the 11-OH group acted as H-bond acceptor for 10-OH, and 12-Br acted as H-bond acceptor for 11-OH. In all following steps this OH conformation was kept fixed to greatly reduce sampling complexity. The sulfate group was built in its deprotonated form. The aromatic ring was built in propeller conformation (*P*). All four combinations of configuration at the two stereogenic centers were generated. A conformational search was performed with MacroModel 11.8<sup>[215]</sup> using the MMFF force field<sup>[151]</sup> in vacuum. The method was a Monte-Carlo torsional sampling<sup>[253]</sup> of all non-terminal rotatable side chain bond with 100 000 steps, a mini-

mization convergence of  $1.0 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$ , and an energy threshold of  $21 \text{ kJ mol}^{-1}$ . The resulting conformers were subjected to a finer minimization with  $0.001 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$  convergence, and structures with a maximum atom deviation below  $0.5 \text{ \AA}$  were discarded as duplicates. For the DP4+ analysis, all conformers from the ensemble were geometry optimized at the B3LYP<sup>[154,156,243,244]</sup>/6-31G<sup>[254]</sup> level of theory, including the calculation of vibrational frequencies, and all conformers below  $8.4 \text{ kJ mol}^{-1}$  of the minimum zero-point corrected (ZPC) free energy were discarded. The shielding constants were calculated at the mPW1PW91<sup>[255,256]</sup>/6-31+G(d,p) level of theory using GIAO<sup>[257]</sup> and an implicit solvent model.<sup>[258]</sup> The resulting shieldings were then averaged assuming a Boltzmann distribution with the ZPC free energies from the geometry optimization. For the ECD calculation, all conformers were geometry optimized at the B3LYP/pcseg-1<sup>[259]</sup> level of theory, including vibrational frequencies to get ZPC free energies. All conformers with a free energy of less than  $11.4 \text{ kJ mol}^{-1}$  above the ground state were used for an excited state calculation with 100 states using time-domain DFT using three different functionals PBE0<sup>[242]</sup> (all configurations), B3LYP,<sup>[154,156,243,244]</sup> and CAM-B3LYP<sup>[245]</sup> (only (2'S, 2''R)), and the pcseg-1 basis set. All DFT calculations were performed with Gaussian09,<sup>[153]</sup> and Jensen basis sets were retrieved from the Basis Set Exchange.<sup>[260]</sup>

## 4.7.2 Anisotropic NMR

**4.7.2a Sample Preparation** For **1**, 2.0 mg of dry material from purification was dissolved in ca.  $150 \mu\text{L}$  MeOH- $d_3$  and transferred into a 3 mm NMR tube. For **2**, a 200 mM solution of strychnine (Sigma-Aldrich) in  $\text{CDCl}_3$  with ca. 1% v/v TMS was washed twice 1:1 with  $250 \text{ g L}^{-1} \text{ K}_2\text{CO}_3(\text{aq})$  to remove traces of chloroform degradation products. The solution was then dried over  $\text{K}_2\text{CO}_3(\text{s})$ , transferred into a 5 mm NMR tube, and degassed using five freeze-pump-thaw cycles. Finally, the tube was flame-sealed under vacuum. For the measurements at 1200 MHz, a 200 mM solution of strychnine in  $\text{CDCl}_3$  was washed twice 1:1 with  $250 \text{ g L}^{-1} \text{ K}_2\text{CO}_3(\text{aq})$ , dried over  $\text{K}_2\text{CO}_3(\text{s})$ , transferred into a 3 mm NMR tube, and degassed using five freeze-pump-thaw cycles. The tube was flame-sealed under argon atmosphere.

**4.7.2b NMR Parameters** 1D and 2D NMR spectra were recorded on Bruker Avance III HD or Bruker Avance Neo spectrometers at fields between 9.4–28.2 T (proton frequencies of 400, 600, 700, 800, 900, 950, and 1200 MHz) using inverse cryogenically cooled (QCI/TCI) probes (600–1200 MHz) and a room temperature inverse (QXI) probe (400 MHz). The sample temperature was set to 298 K. For **1**, the temperature control was cross-checked and fine-tuned using a 99.8% MeOD- $d_4$  reference sample (Sigma-Aldrich).<sup>[137]</sup> For **2**, the chemical shift of the residual water peak (2.397 ppm @298 K,  $-23 \text{ ppb K}^{-1}$  temperature shift) served as internal temperature standard. Proton 1D spectra were acquired using simple pulse-acquire sequences. For **1**, low power presaturation was added for solvent suppression. Spin-echo based solvent suppression (WATERGATE etc.) is less suitable as it introduces phase errors in the multiplet structure due to homonuclear coupling evolution. Carbon 1D spectra were acquired using an APT

sequence (Bruker standard sequence jmod) with 4–5 kHz proton decoupling. For **2**, HSQC spectra were acquired using a perfect-clip-hsqc sequence.<sup>[249]</sup> For data processing, zero-filling to 1024k ( $2^{20}$ ) points (1D spectra) and 64k ( $2^{16}$ ) points (HSQC) was applied in the direct dimension. For the low-sensitivity carbon-based experiments (APT/HSQC), 0.3 Hz line broadening was applied before Fourier transformation. Peak picking and multiplet analysis were done in MestreNova 11. For **1**,  $J_{\text{HH}}$  couplings were taken from the proton 1D, and peaks showing overlap were discarded. For **2**, all couplings were extracted from the HSQC. The slices in the proton dimension containing the relevant peaks were extracted for this purpose. Any HSQC peaks showing obvious overlap with long-range correlations of neighboring peaks were discarded. This was the case for the atom positions 1, 2, and 3. The unprocessed chemical shifts and couplings used in the data analysis can be found in the data collection (Appendix A.1).

**4.7.2c Computational Methods** For **1**, the conformer geometries from the DP4+/ECD calculations at the MMFF level of theory were used as starting points (Section 4.7.2). For **2**, starting geometries for 22 possible diastereomers were built in Maestro 11.4. All geometries were optimized at the B3LYP/pcseg-1 level of theory, including vibrational frequencies for **1** to get ZPC free energies. We calculated NMR shieldings and magnetic susceptibility tensor at the B3LYP/pcSseg-1<sup>[261]</sup> (**1**) and B3LYP/pcSseg-2<sup>[261]</sup> (**2**) level of theory using GIAO and an implicit solvent model. All DFT calculations were performed with Gaussian09,<sup>[153]</sup> and Jensen basis sets were retrieved from the Basis Set Exchange.<sup>[260]</sup> The logfiles from the NMR calculations containing all relevant data for the analysis can be found in the data collection (Appendix A.1).

During the first attempts to predict magnetic susceptibilities, we encountered severe basis set convergence problems, which was most prominent for large Pople-type basis sets. We investigated this by calculating shieldings and susceptibilities for strychnine using 57 different basis sets, from STO-3G to 6-311++G(3df,3pd). For relatively small basis sets (e.g., 6-31G, 6-31G(d)) the agreement of the calculated alignment (i.e., the susceptibility anisotropy) with the experimental data was good, but with larger basis sets it became increasingly worse. This was most extreme for basis sets with diffuse functions (i.e., functions with small exponents decaying slowly with distance), which points to numerical artifacts due to near-degeneracy of some basis set functions. It is noteworthy that these numerical artifacts had a significant effect only on the susceptibilities, but not on geometries or nuclear shielding tensors. These numerical artifacts could be avoided by using a finer than default integration grid using the Gaussian “UltraFine” keyword. As an additional measure, we decided to switch to a more balanced family of basis sets with a clear hierarchy. We tested Jensen’s polarization-consistent segmented basis sets as well as Karlsruhe (def2) basis sets<sup>[262]</sup> on strychnine, which yielded very similar results. We finally chose to do all subsequent calculations with Jensen-type basis sets as they were specifically designed for DFT calculations and provided a subfamily optimized for nuclear properties (pcSseg-n).

**4.7.2d Data Analysis** All computations and data analysis were performed using a series of in-house python scripts, which make extensive use of the SciPy ecosystem.<sup>[216–218]</sup> Intermediate results, such as carbon resonance assignment, chemical shift error vectors  $\delta_T$  and  $\delta_{\text{degrad}}$ , the transformation matrix  $\mathbf{Q}$ , and the fitted RDCs and RCSAs with the fit residuals and the weighting factors are reported in the data collection as tab-separated text files with extensive comments (Appendix A.1).



# Chapter 5

## Design of an Electric Field Cell

This chapter describes the design of a an electric field cell for anisotropic NMR. All electrotechnical work, in particular the design and construction of an AC high voltage source, was done by Florian Jordan. Matthias Kulp and Matthias Kleinhans performed the coatings using physical vapor deposition, and Julius Bergmann designed and machined the stencils for coating.

### 5.1 Introduction

While the method of magnetic field alignment as presented in the previous chapter was an elegant and physically simple way to acquire anisotropic NMR data, it suffers from a series of drawbacks, which are mostly connected to the very small magnitude of the effects. First, very high magnetic fields are necessary to observe appreciable alignment, which may not be available in routine analytics. Second, a number of different fields are necessary, which increases the experimental effort and raises the need for error correction. Third, only molecules with sufficiently large anisotropic susceptibility can be examined, which is mostly the case for aromatic molecules. Using electric fields as aligning agents has the potential of having the advantages of magnetic field alignment, such as predictability via DFT and the lack of aligning media, while avoiding some of its disadvantages. Magnetic alignment is based on the interaction between the magnetic field and the molecule's susceptibility, whereas electric field alignment is caused by the interaction of the molecule's electric dipole moment with an external electric field. This type of alignment is therefore also governed by simple physical interactions and molecular properties, and can potentially be predicted just as easily as magnetic alignment using molecular modeling such as DFT. On the other hand, as the electric field could be applied independently of the magnetic field, the isotropic and anisotropic experiments could be acquired in a single session by simply turning the electric field off or on. This may greatly reduce the necessity for error correction. In magnetic alignment, the chemical shift errors were mostly caused by temperature, degradation, and referencing instabilities which are attributable to using different spectrometers on different days. Similarly, problems with RDCs were often caused by the different strong coupling regimes present at different magnetic fields. Another advantage of using electric fields is that the degree of alignment can potentially be tuned by simply increasing or decreasing the electric field strength.

Also, it is more universally applicable since most small molecules such as natural products have a nonzero dipole moment.

There are also challenges and drawbacks of this approach. One fundamental drawback shared with magnetic alignment is that there is only one possible shape of the alignment tensor, which is governed by the molecular properties. Here, induced alignment has the advantage of offering the possibility of acquiring data in different media with different alignment tensors and as such gaining independent constraints. As with magnetic alignment, very strong fields are necessary to achieve appreciable alignment. A quick back-of-the-envelope calculation and the study of previous work<sup>[263–266]</sup> reveals that the fields need to be in the order of 10–100 kV cm<sup>-1</sup>, so care needs to be taken to avoid arcing and electric breakdown in the apparatus. Indeed, the success or failure of electric alignment hinges on the design of the electric field cell. Fundamentally, such cells have been designed and constructed and this effect has been experimentally verified before. Plantega et al. used alignment by DC electric fields to measure the orientation of dipole moments in substituted benzenes<sup>[263]</sup> as well as the degree of alignment in binary mixtures of liquids with different polarities.<sup>[264]</sup> Peshkovsky et al. used pulsed AC electric fields at 28.5 kHz to study the alignment of nitrobenzene, observing both RQCs<sup>[265]</sup> and RDCs.<sup>[266]</sup> In both cases the field cells were too large to fit into a regular 5 mm NMR probe, and the experiments were conducted in wide-bore magnets with large-diameter probes. We believe however that electric field alignment can only become widespread and useful in routine structure determination if the field cells are applicable in a standard liquid state NMR spectrometer, so this was the main goal in the design of our cell.

### 5.1.1 General Considerations

As briefly introduced in Section 5.1, there are certain requirements that a field cell for electric alignment should fulfill. Firstly, it should be applicable in a standard spectrometer. Since most routine instruments are fitted with 5 mm probeheads, the cell has to be limited to a 5 mm NMR tube as outer boundary. Also, any electrical connections need to be attached to the top as this is the only access point to a regular probe.

To better estimate the necessary field strengths it is instructive to calculate the expected anisotropic parameters of an exemplary sample. The alignment tensor for electric alignment is shown in Eq. (2.41). However, the effective dipole moment  $p$  and electric field  $E$  differ from the gas phase dipole moment  $p_0$  and the theoretical electric field  $E_0$  as computed by the fraction of voltage and electrode distance. This is described by Onsager’s theory,<sup>[267]</sup> which comes to the following result for the general case of polar liquids:

$$p = \frac{(2\epsilon_r + 1)(n^2 + 1)}{3(2\epsilon_r + n^2)}p_0 \quad \text{and} \quad E = \frac{3\epsilon_r}{2\epsilon_r + 1}E_0, \quad (5.1)$$

where  $\epsilon_r$  and  $n$  are the solvent’s dielectric constant and the refractive index, respectively. An example could be a <sup>1</sup>D<sub>CH</sub>-RDC of strychnine dissolved in chloroform. We use the parameters of this sample ( $\epsilon_r = 4.81$ ,  $n = 1.45$ ,  $p_0 = 5.3$  D), a temperature of 298 K, a perpendicular arrangement of  $\mathbf{E}$  and  $\mathbf{B}$  ( $\alpha = \pi/2$ ), a bond length of 1.1 Å, and a parallel

arrangement of the dipole moment and the bond vector (which yields the orientational maximum of the RDC) to calculate the RDC  $\bar{D}_{\max}$  as a function of the field strength:

$$\bar{D}_{\max} = \frac{2}{3} \frac{3\mu_0\gamma_H\gamma_C\hbar}{8\pi^2r_{\text{CH}}^2} \frac{1}{30} \left( \frac{\epsilon_r(n^2 + 1) p_0 E_0}{(2\epsilon_r + n^2) k_B T} \right)^2 (3 \cos^2 \alpha - 1) \approx -3 \cdot 10^{-4} \text{ Hz} \left( E_0 / \text{kV cm}^{-1} \right)^2 \quad (5.2)$$

If one aims for these maximum RDCs being no smaller than 0.3 Hz, which roughly corresponds to what was achieved by magnetic alignment, the electric field should be no smaller than about 30 kV cm<sup>-1</sup>. Such field strengths pose severe engineering challenges: a potential apparatus needs to be designed in a way that prevents arching between the cell electrodes, but also any arching to the the NMR coil needs to be prevented with certainty, as such an event may be catastrophic to the NMR hardware.

Apart from the need to reach these high field strengths, the electric field also needs to be sufficiently homogeneous within the active volume of the NMR sample. If this is not the case, different parts of the sample exhibit different RDCs or RCSAs, which leads to additional line broadening. Formally, the resulting line shape is the convolution of the isotropic line shape and the distribution function of the RDCs or RCSAs (which are proportional to  $E^2$ ). Here we will not formulate a formal requirement for the homogeneity, but it is a property to be considered and examined for a potential cell design.

Finally, the components and electrodes of the field cell should also not compromise the NMR experiments. Firstly, for polymeric materials and fluids the NMR background signal has to be taken into account. Secondly, the conducting materials of the electrodes have to be arranged in a way that they do not act as a Faraday cage, shielding the sample volume from the RF radiation that is applied from outside.

## 5.2 The 1D Field Cell: Circular Tubes and Linear Electrodes

### 5.2.1 Concept and Design

Based on the abovementioned considerations two major design choices were made for the first cell design. First, we considered various arrangements of electrodes. The textbook example of creating homogeneous electric fields is the plate capacitor with plane-shaped electrodes, where the distance of the planes is much smaller than their dimensions. Arranging the plates above and below the active volume violates this geometric condition, which leads to field inhomogeneity and poses the additional challenge of connecting the bottom electrode from above. Arranging the plates sideways is hardly compatible with the round cross-section of the NMR tube and also creates a large conducting surface between the NMR coil and the sample, which potentially attenuates or blocks the RF radiation necessary for the NMR experiment. We therefore chose to use a number of line-shaped electrodes parallel to the NMR tube axis; this design will be referred to as the 1D field cell.

Second, the cell should consist of a 3 mm tube containing the sample within a 5 mm tube, where the electrodes and electrical connections are positioned in the ring-shaped gap between the tubes. This ensures that both the sample as well as the NMR probe are separated from the high-voltage components by an insulating layer of borosilicate glass, which has a high dielectric strength of around  $300 \text{ kV cm}^{-1}$ .<sup>[268]</sup>

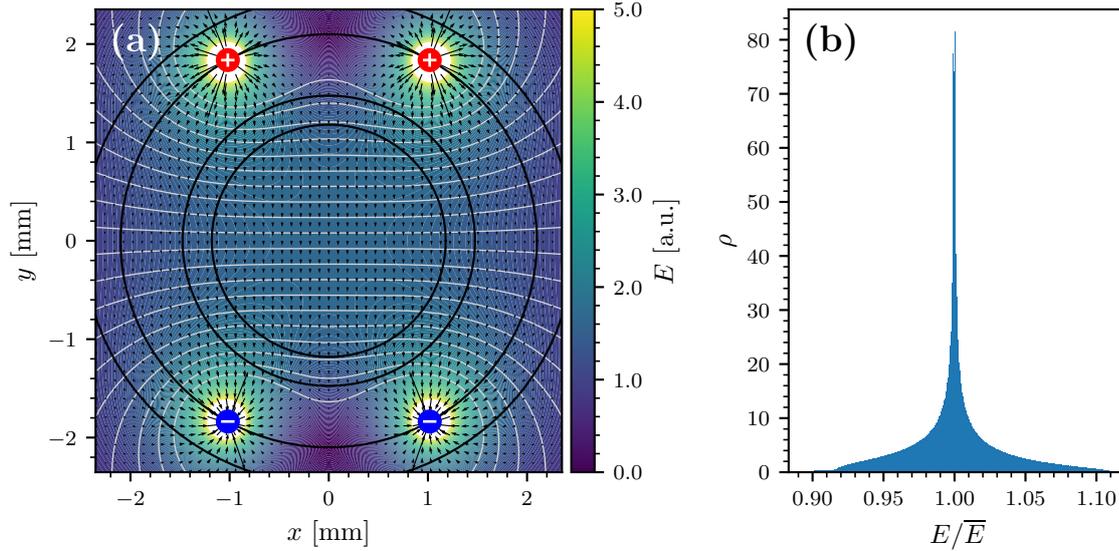
To determine the basic arrangement of the electrodes, we experimented with different options using a very simple mathematical model of line-shaped electrodes in vacuum. Also, the region around the active volume is translationally symmetric along the tube axis, so we limited the field modeling to a 2D cross section. The electric field of a charged line (infinitely long and thin) in vacuum is given by:

$$\mathbf{E} = \frac{\lambda}{2\pi\epsilon_0 r} \hat{\mathbf{e}}_r, \quad (5.3)$$

where  $\lambda$  is the linear charge density,  $r$  and  $\hat{\mathbf{e}}_r$  the distance and direction to the electrode, respectively, and  $\epsilon_0$  the vacuum permittivity. With these 1D electrodes field inhomogeneities are unavoidable, and they are smallest within the active volume when the electrodes are as far away as possible, so we positioned them on the inner wall of the outer tube. We chose to use four electrodes (two positive and two negative ones) since fewer (i.e., two) would lead to a very inhomogeneous field, similar to a dipole field, while more would require different same-polarity electrodes to be connected to different magnitudes of electric potential. Symmetry consideration require these four electrodes to be arranged in a rectangular pattern (see Fig. 5.1a), and there is only one degree of freedom for this arrangement: the ratio of side lengths of this describing rectangle, or, as we express it, the angle  $\alpha$  between its diagonals. In this description an angle  $\alpha$  of  $0^\circ$  corresponds to the case where the same-polarity electrodes coincide, which corresponds to the (suboptimal) dipole case. At an angle  $\alpha$  of  $180^\circ$  electrodes of opposite polarity coincide, which is obviously also not a viable arrangement. The optimum must therefore lie somewhere in between. The criterion for the optimization was the field homogeneity, or more precisely the RMSD of the electric field magnitude within the active volume. We found the minimum of the RMSD at an angle of  $\alpha = 58^\circ$ . The field distribution for this case is depicted in Fig. 5.1a, and a histogram of relative field strengths within the active volume is shown in Fig. 5.1b. It shows a narrow and symmetric peak at the center which is exactly desirable for our application.

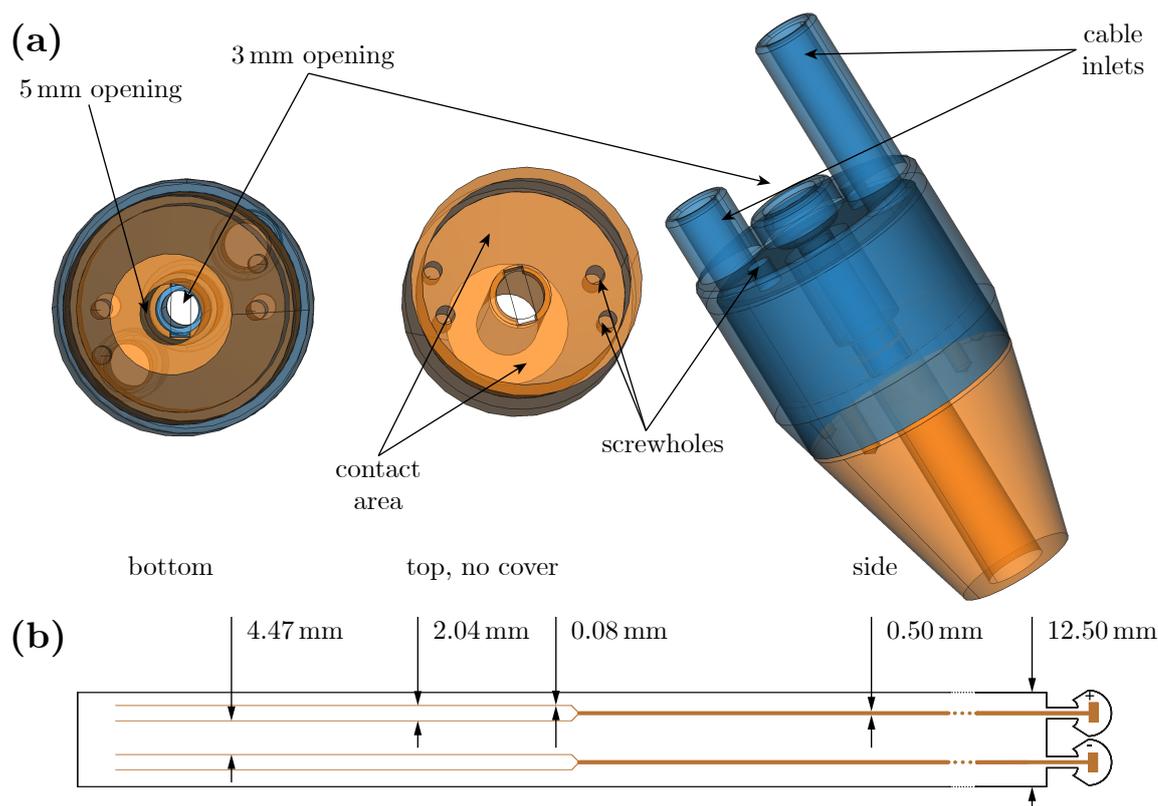
## 5.2.2 Technical Realization

While the most natural choice for constructing line electrodes would be to use thin wires, it is not trivial to ensure that they are perfectly straight and properly positioned in the desired rectangular arrangement, and to safely connect them to a high-voltage cable. We therefore used flexible printed circuit boards (PCB) as framework for the electrodes. Specifically, we designed a PCB with a width corresponding to the inner circumference of a 5 mm tube that could be rolled and inserted into such a tube. The electrodes were implemented as two pairs of two conducting paths on the PCB. The minimum



**Figure 5.1.** (a) Electric field distribution for four electrodes at their optimal positions: the separation in polar angle of the electrodes with same polarity is  $58^\circ$ . The field is calculated for line-shaped electrodes in vacuum. The color and arrows indicate the field, equipotential lines are drawn in gray, and inner and outer circumferences of the two NMR tubes are drawn in black. (b) Histogram of relative field strengths within the active volume (inner black circle).

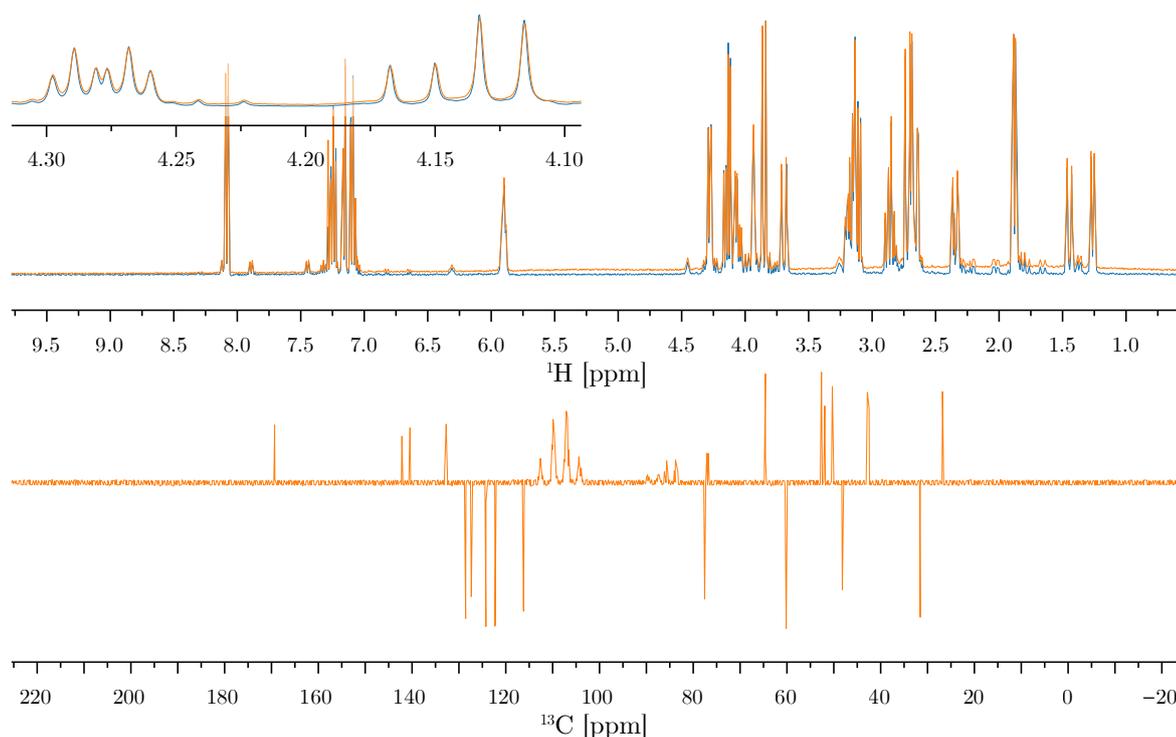
cross section of these electrodes were given by the minimum feature size on the PCB ( $80\ \mu\text{m}$ ) and the thickness of the PCB's copper layer ( $18\ \mu\text{m}$ ). Above the active volume the same-polarity electrodes were united into a larger conducting path and ended in a soldering pad on a foldable flap. A layout of the PCB is depicted in Fig. 5.2b. To hold the various components together a 3D-printed plastic head was designed. It consisted of two pieces: the bottom part had a 5 mm opening from the bottom that would hold the outer tube. The PCB's top flaps would go further through this opening and fold horizontally onto a flat surface, where the connecting cables could then be soldered on. All solder joints were covered in epoxy for electrical insulation. The head's lid contained three holes for the two cables and for the inner tube, which could potentially be chosen to be longer than the outer tube. The computer-assisted design (CAD) model of the head is depicted in Fig. 5.2a. To ensure that both tubes are concentric at the bottom, a  $2 \times 1\ \text{mm}$  O-ring was placed at the bottom of the outer tube before inserting the inner tube. Having air in the gap between the two tubes is detrimental both because of its low dielectric strength and its low permittivity, which leads to high local field strengths. Therefore we filled this gap with different insulating fluids (See below). The high voltage was supplied by two commercial DC voltage supplies, one with positive polarity and one with negative polarity to ensure symmetric potential of the two electrode pairs around ground.



**Figure 5.2.** (a) CAD model of the 1D field cell's two-pieced head. The 5 mm tube is inserted from the bottom and is fixed by friction, the inner 3 mm tube from the top. The pads of the flex-PCB are folded onto the contact area and soldered to the cable coming from the top. (b) Layout of the flex-PCB (to scale). The dotted region indicates a cut-out of 50 mm.

### 5.2.3 Testing

We first tested the setup with various choices of insulating fluids for electrical stability outside the NMR instrument. Since the insulating fluid is located between the electrodes, its permittivity  $\epsilon_r$  has a large effect on the field inside the sample volume; for maximum field in the sample the insulation's permittivity should be much larger than the sample's permittivity. Also, a fluid with minimal NMR background should be chosen, so the first choice were deuterated solvents. We tested dichloromethane with  $\epsilon_r \approx 8.9$  as well as chloroform with  $\epsilon_r \approx 4.8$  as readily available solvents. Unfortunately, the flexible PCB were not resistant against these solvents and its glued layers started to deteriorate and curl. Additionally, electrolytic gas formation was observed upon application of high voltage, which may potentially be catastrophic in combination with arcing, so these solvents were quickly rejected as potential insulators. Another option for insulating fluids without (proton) NMR background that are chemically very inert are perfluorinated hydrocarbons. We chose to use perfluorodecalin mostly for its high boiling point (low volatility) and its availability. While it has a low permittivity ( $\epsilon_r \approx 1.9$ ), it ensured electrical stability of our setup up to a voltage of 10 kV which could not be achieved

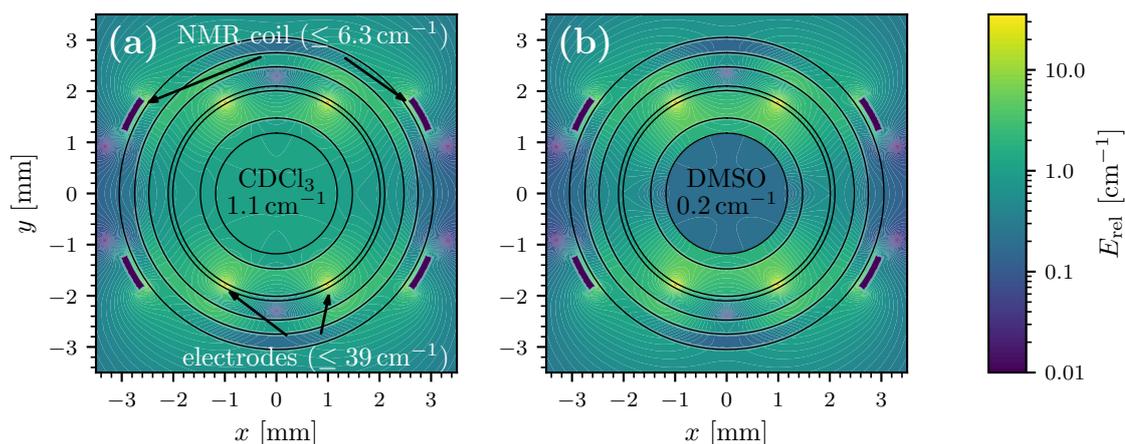


**Figure 5.3.**  $^1\text{H}$  and  $^{13}\text{C}$  spectra of 200 mM strychnine in  $\text{CDCl}_3$  acquired within the 1D field cell (orange) as compared to the reference spectrum (blue, only  $^1\text{H}$ ). In  $^1\text{H}$  the polymer background from the flex-PCB is visible as a low-intensity hump between  $-1$  and  $9$  ppm. Neither the signal intensity nor the line shape are affected. In  $^{13}\text{C}$  the background from perfluorodecalin is limited to the region between  $80$  and  $115$  ppm.

with the other tested options. The electrical stability in the head region was challenging to achieve; only after several iterations of adding epoxy and rearranging the cables we could safely exclude arcing even at the highest voltage.

To see if and how the field cell components would affect the NMR experiments, we acquired  $^1\text{H}$  1D spectra of a test sample (200 mM strychnine in  $\text{CDCl}_3$ ) with and without the field cell (Fig. 5.3). In terms of line shape and intensity there was no measurable difference, and the sample within the field cell generally behaved no different from a regular NMR sample. Only the background signal from the flex-PCB polymer made it possible to distinguish between the two spectra, which appeared as a very broad hump between  $-1$  and  $9$  ppm. In a  $^{13}\text{C}$  spectrum the signal from perfluorodecalin was clearly visible as a number of signals between  $80$  and  $115$  ppm; in many natural products this is an empty region of the spectrum between the aliphatic and the aromatic region, so this should also only be a minor nuisance.

While the first field calculations in vacuum gave a very rough idea of the field distribution given the electrode geometry, it ignored the electric properties of the materials within and around the cell. To get a more accurate and realistic picture, we used finite element calculations to simulate the field distribution in the cell. As before, we exploited the translational symmetry along  $z$  around the active volume and restricted the simulation



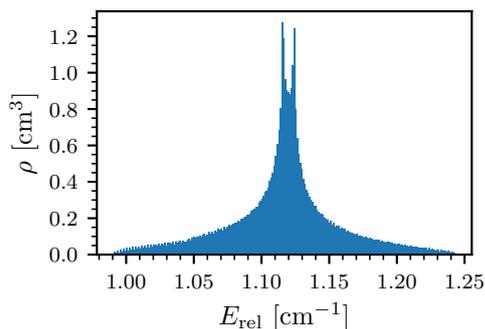
**Figure 5.4.** Finite element simulation of the field distribution of the 1D design. The concentric circles mark the borders between the different components, which are, from inside to outside: sample, glass tube, insulating fluid, flex-PCB, glass tube, air, probe glass, air. The electrodes and the NMR coil are annotated. The coloring indicates the magnitude of the relative electric field  $E_{\text{rel}} = E/U$  in a logarithmic scale. The simulation was done for (a) CDCl<sub>3</sub> with  $\epsilon_r \approx 4.8$  as well as (b) DMSO with  $\epsilon_r \approx 46.7$ .

to two dimensions. A disc in the center of the geometrical model represented the sample volume, which is surrounded by several concentric rings. These represent the inner 3 mm tube, followed by the insulating fluid and the polyimide foil of the flex-PCB. Within the PCB the  $80 \times 18 \mu\text{m}$  conducting electrodes were embedded. It is surrounded first by the outer 5 mm tube and then by air. Since it is relevant to simulate the field distribution of the sample cell within an NMR probe, and the conducting parts of the NMR coil may have a significant effect on the electric field, we also included the first layer of a room temperature NMR probe in the geometrical model. The sample in the probe is surrounded by a sapphire glass wall with an i.d. of 5.5 mm and an o.d. of 6.1 mm, and the Helmholtz coil right behind it. While we did not know the exact coil dimensions as they are corporate secrets of the manufacturer, we could measure approximate dimensions on a disassembled probe. The coil is made from wire with a rectangular cross section measuring approximately  $0.8 \times 0.2\text{ mm}$ , and in a horizontal cross section through the active volume these wires are positioned in a very similar, rectangular arrangement as the electrodes. In principle the field cell can be freely rotated against the probe part, and we chose an arrangement so that they are as far apart as possible. This geometry is best understood by inspecting Fig. 5.4 which shows the borders between the different materials as well as the predicted field distribution.

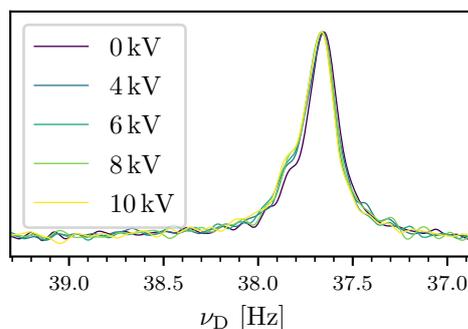
For the simulation, we applied equal and opposite potentials  $\pm U/2$  to the electrodes and set the potential of the NMR coil to 0 (ground). We then generally report the field as a relative value  $E_{\text{rel}} = E/U$  so that it becomes independent of the applied voltage  $U$ . Also, it is sufficient to consider the magnitude of the electric field. For alignment, only the electric field magnitude and its angle  $\alpha$  to the magnetic field  $B_0$  is relevant, and since  $B_0$  is along  $z$ , the angle  $\alpha$  is independent of the orientation of  $\mathbf{E}$  in the  $xy$ -plane. As we

assume translational symmetry along  $z$ ,  $\mathbf{E}$  cannot have a  $z$ -component, and therefore it is always  $\alpha = \pi/2$ .

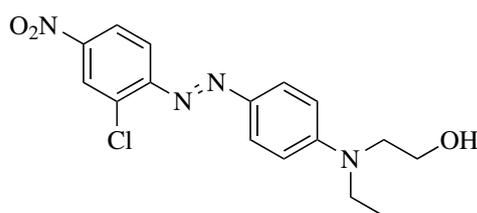
The result of these FEM simulations for two different sample solvents are depicted in Fig. 5.4. Similar to the vacuum calculations, these simulations confirm that the field within the sample region is reasonably homogeneous, as can be seen in Fig. 5.5. However, the flaws of this cell geometry also becomes immediately obvious. First, let us consider the case for a low-permittivity solvent such as  $\text{CDCl}_3$  (Fig. 5.4a). The relative field strength within the sample of  $1.1 \text{ cm}^{-1}$  is less than half the value one would expect by taking the inverse of the electrode distance. This inverse of the electrode distance would be the valid approach for a plate capacitor with a single material between the plates, and marks an upper limit for the relative field. While our geometry is fundamentally different from a plate capacitor which already limits the achievable fields, it is further reduced by the very low-permittivity insulating fluid perfluorodecalin and the interaction with the NMR coil, which is at ground potential. This effect is even more pronounced when the sample solvent is a high-permittivity material such as DMSO with  $\epsilon_r \approx 46.7$ . In that case, the effective field is more than five times smaller as compared to  $\text{CDCl}_3$  as a solvent (Fig. 5.4b). Even with  $\text{CDCl}_3$ , the maximum voltage possible with the available power supplies of 10 kV would yield a field of  $\approx 11 \text{ kV cm}^{-1}$ , which may be sufficient to see first effects in model systems, but certainly not for broad applicability in natural products. An additional fundamental flaw of the line-electrode geometry is the occurrence of very high local fields at the electrode surface. In these simulations, this field has a value of  $\approx 40 \text{ cm}^{-1}$  and is therefore two orders of magnitude larger than the field within the sample. This facilitates electrical breakdown at the electrodes and therefore reduces the maximum applicable voltage. Similarly, the field around the NMR coil is also very high ( $\approx 10 \text{ cm}^{-1}$ ), which may lead to arcing during pulsing and potentially damaging the probe.



**Figure 5.5.** Histogram of the electric field within the sample region for  $\text{CDCl}_3$  as a solvent. It is normalized such that integration yields the cell area.



**Figure 5.6.** Deuterium signal of MeCN with applied voltages between 0 and 10 kV.



**Figure 5.7.** Structure of disperse red 13.

Finally, we tested whether we could observe anisotropic effects using this 1D field cell under the most promising conditions. Firstly, we used benzene as a solvent which has a very low permittivity of  $\approx 2.3$ , yielding a relative field strength in the FEM simulation as high as  $1.6 \text{ cm}^{-1}$ . Secondly, we tried to observe deuterium residual quadrupolar couplings (RQC) of acetonitrile. The deuterium signal of MeCN is remarkably narrow with a line width between 0.1–0.2 Hz due to its low molecular weight, and the comparatively large electric dipole moment of 4.8 D (predicted via DFT) should lead to strong electric alignment. Taking all parameters into account, we predicted the RQC to be around 0.1 Hz, which would be well observable given the line width. In a sample of 2% v/v of MeCN- $d_3$  in  $\text{C}_6\text{D}_6$ , we acquired deuterium spectra at voltages of 0, 4, 6, 8, and 10 kV, which are shown in Fig. 5.6. Unfortunately, no RQC splitting could be observed, not even a slight broadening at higher voltages. We also attempted to observe RDCs in disperse red 13, which is an azo dye with a (predicted) dipole moment of around 11 D. Here we predicted an ortho- $^3D_{\text{HH}}$  coupling of 0.06 Hz at a voltage of 10 kV, which is already at the lower limit of detectability. Again, we were not able to detect any measurable difference between the spectra with and without applied voltage.

In a second set of experiments we observed clear signs of arching at the NMR coil on deuterium at voltages of only 6 kV. As this can potentially damage the probe, we decided to abort the experiments and fundamentally review the cell design.

## 5.2.4 Conclusion and Outlook

Apart from the obvious failure in generating alignment, the first cell showed a number of other flaws and drawbacks. Due to the relatively large distance of the electrodes, high potentials are necessary to generate the required fields. This is worsened by the fact that between the electrodes there are other components of the cell (tube, insulating fluid) which, due to their relatively low permittivity, reduce the field within the sample region even further. This also leads to a great dependence of the field on the sample properties: when going from  $\text{CDCl}_3$  to DMSO as solvent, the field within it is reduced more than five-fold (Fig. 5.4). Generally, the theoretically achieved field strengths with this design were barely enough to yield significant alignment on model systems, and would certainly be too low for general applicability. The high necessary voltages are unfavorable insofar as they require more powerful voltage sources, and more care in ensuring electrical stability and insulating the design to the outside. As we have seen the insulation to the outside does not only concern the prevention of electrical breakdown across the tube, but also the shielding of the NMR coil from high field strengths. Such a shielding could in principle be achieved by adding a conductive layer around the outside of the outer tube. However, FEM simulations of such a modification show that this would reduce the field within the sample again by more than a factor of two, while simultaneously increasing the local field around the electrodes.

A possible explanation for the missing alignment may be the use of a DC voltage. In the resulting static field the migration of ions in the solvent can occur, which creates a field with opposite polarity. Given a sufficient concentration of ions, this continues until the fields cancel completely. A quick calculation can give us an order of magnitude for

the necessary ion concentration. As a model system we assume a plate capacitor with electrode distance  $d = 4$  mm, surface area  $A$ , and a dielectric with  $\epsilon_r = 5$ . The charge  $Q$  loaded onto the capacitor is given by

$$Q = CU = \epsilon_0\epsilon_r \frac{A}{d}U \quad (5.4)$$

To cancel the field, the total charge of the dissolved ions must be at least as big as this capacitor charge  $C$ . We get the ion amount by dividing by the Faraday constant  $F$ , and the concentration  $c$  by dividing again by the Volume  $V = Ad$ :

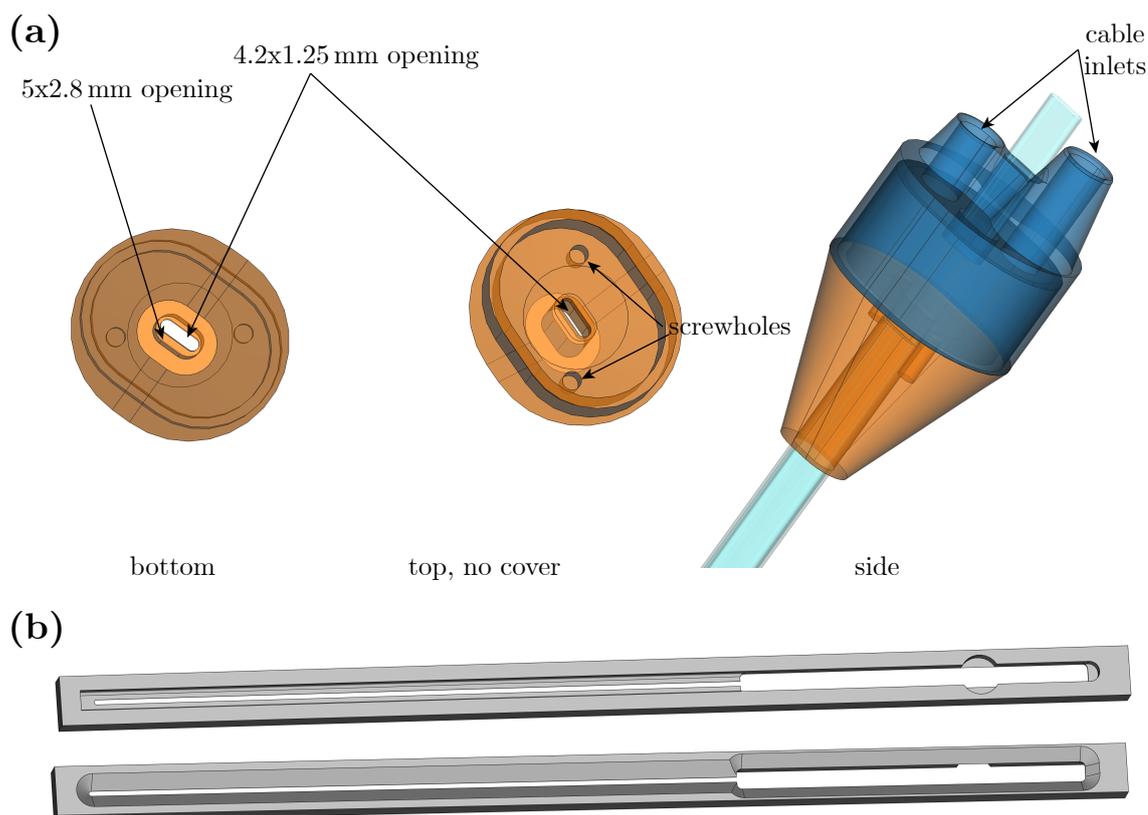
$$c = \frac{n}{V} = \frac{Q}{FAd} = \epsilon_0\epsilon_r \frac{U}{d^2F} \quad (5.5)$$

A potential of 10 kV and the aforementioned capacitor parameters yield a required ion concentration of only 0.3 nM. Since even very non-polar solvents are never completely free of ions, it is quite plausible that this mechanism leads to a complete cancellation of the field within the solution in our case.

## 5.3 The 2D Field Cell: Shaped Tubes and Planar Electrodes

### 5.3.1 Concept and Design

Many of the obvious issues of the 1D field cell were a consequence of the line shape of the electrodes. We therefore chose to revisit the possibility of building a field cell which is more close to the geometry of a plate capacitor, with plane-shaped, 2D electrodes. One of the issues of such a design is that the sample region is largely encompassed by conducting surfaces, which potentially act as a Faraday cage and shield the sample from the RF irradiation. However, in the work of Peshkovsky et al. this problem was solved by using electrodes from the semiconducting material indium tin oxide deposited onto a glass surface.<sup>[265]</sup> The combination of a low layer thickness and a low specific conductance of the material yielded a sufficiently low surface conductivity such that the high frequency RF irradiation is not significantly damped. Their field cell did not meet our requirement of being compatible with a standard probe, but it showed that the basic geometry of a plate capacitor can be made compatible with an NMR experiment, so we chose to follow this basic design for our second design, hereinafter called the 2D field cell. As we suspected the use of a DC electric field as one possible cause for the failure of the first experiments, we chose to employ an AC field with a frequency in the order of 10 kHz as it was done in most of the previous work done on electric field alignment. Off-the-shelf high voltage AC power supplies for scientific purposes offer a wide range of functions, but are prohibitively expensive. For our purpose of alignment a very simple device with a voltage range up to 10 kV and a frequency range in the single-digit kHz range should be sufficient. This type of technology is, in principle, cheaply available as



**Figure 5.8.** (a) CAD model of the 2D field cell's two-pieced head. Both tubes are inserted from the bottom. The ring-shaped PCB (not shown) is mounted onto the flat surface shown in the center view and fixed using screws, so that the spring contacts touch the inner tube on the large sides upon insertion. (b) CAD model of the aluminum stencil used to coat the shaped tubes.

it appears in various consumer electronics, such as cathode-ray tube monitors or plasma lamps. We therefore chose to build a suitable voltage supply ourselves in a price range of around 1000 €.

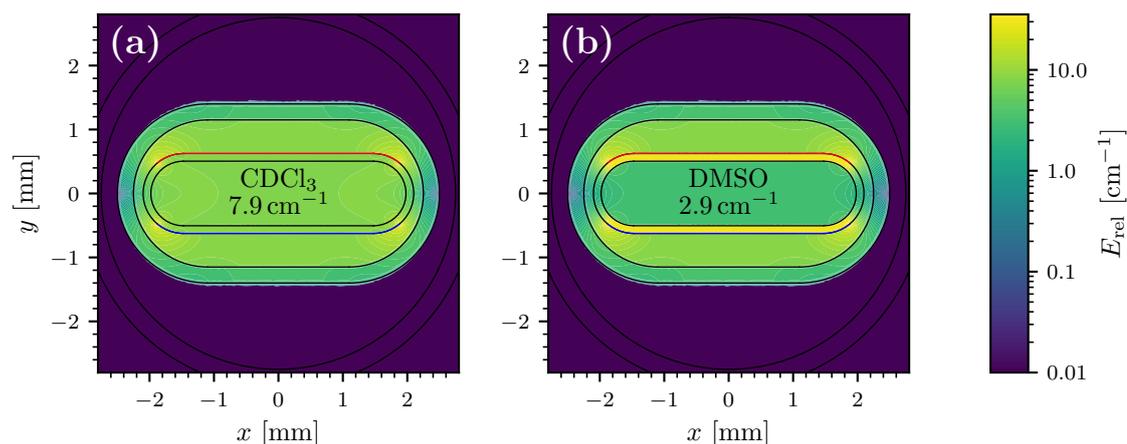
### 5.3.2 Technical Realization

To construct a cell that is compatible with a standard NMR probe and still approximately corresponds to the geometry of a plate capacitor, we chose to use shaped tubes with a near-rectangular cross section. Such tubes with dimensions  $5 \times 2.8$  mm are already available for NMR for an entirely different purpose: their shape avoids the region of the probe where the electric component of the RF pulses is most prominent, and has therefore advantages for lossy samples with high salt concentrations.<sup>[269]</sup> Thanks to this, appropriate sample holders are already available. We chose to keep the tube-in-tube design as this again ensures electrical insulation to the outside. The outer tube had the common dimensions of  $4.95 \times 2.8$  mm to ensure compatibility with existing equipment. For the inner tube we chose tubes with dimensions  $4.2 \times 1.25$  mm. The wide side of

4.2 mm is just small enough to fit inside the outer tube, which has a wall thickness of 0.25 mm. This maximizes sample volume and reduces possible lateral motion of the inner tube. Having a thin side of only 1.25 mm, which will correspond to the electrode distance, increases the possible electric field and reduces field inhomogeneities on the sides. A thin wall of only 0.12 mm maximizes sample volume and, as the electrodes will be applied on the outside, reduces the influence of the glass on the field strength within the sample.

We applied the electrodes onto the glass surface as a thin film using thermal physical vapor deposition (PVD). In this technique, the film material is loaded into a metal crucible, called “boat”, within a high vacuum chamber below the glass tube. The boat is then heated by passing an electric current through it, evaporating the material. The material then deposits onto all surfaces within the chamber, including the glass tube. We tested different semiconducting materials such as indium tin oxide and silicon, but since these materials are sensitive to the oxidation conditions within the chamber which could not be controlled with the PVD chamber that was available to us, we were not able to produce conducting surfaces. Silver and gold produced films that were too conductive and significantly hindered the NMR experiments. Also, they were very sensitive to mechanical abrasion (touching, scratching etc.). Finally, we discovered that titanium was a suitable material: it could easily be deposited with our PVD chamber and produced conductive and physically robust films. Due to limited capabilities of the truly ancient PVD chamber we could not control the film thickness very well and therefore the resulting (approximate) surface resistivities fluctuated in the range between 1.5–4  $\Omega$ . However, in no case those layers would measurably influence the NMR experiments (see Section 5.3.3). The cross-section of the inner tubes is approximately that of a rectangle with two semicircles on the short sides. To maximize field homogeneity, we aimed at coating the full width of the flat surface as well as short section of the curved part as well. This was done on a length of 70 mm from the bottom to encompass the whole sample region. Above that, the titanium surface had merely the purpose of a conducting layer, so we reduced the width to 1 mm to increase the distance from opposite electrodes as well as to reduce the capacitance of the resulting cell. To achieve this precise non-planar shape, we designed and produced a CNC-milled aluminum stencil in which the tube was placed during coating. It is depicted in Fig. 5.8b. Again, the space between the tubes was filled with perfluorodecalin to a height of approximately 100 mm so that the wide part of the electrodes was safely covered.

We designed a slightly modified, 3D-printed head to hold all parts together (see Fig. 5.8a). The outer tube was 25 mm shorter than the inner tube and reached 10 mm into this head. Within the head there was a ring-shaped PCB onto which the connecting cables were soldered. The ring’s hole in the center was sized so that the inner tube could be guided through. On the side of this hole spring contacts ensured the electrical contact to the electrode surfaces. With this design it was possible to ensure reliable contact to the titanium surface (onto which it was not possible to solder) with the added benefit of being reversible, i.e., the inner tube could be easily removed from the setup without the need for tools or soldering. After observing probe arching in first tests similar to before (see Section 5.2.3), we added a titanium coating all the way around the outer tube from



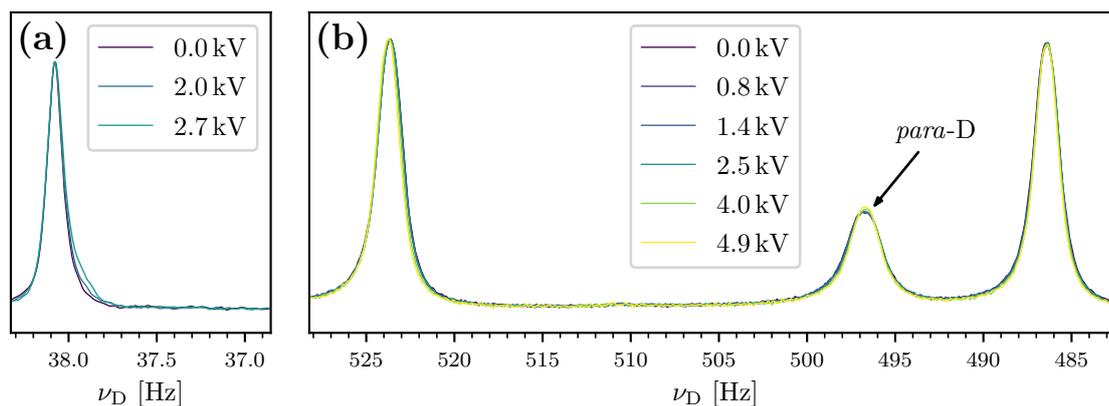
**Figure 5.9.** Finite element simulation of the field distribution of the 2D design. The black lines mark the borders between the different components. The electrodes are marked in red and blue. The coloring indicates the absolute value of the relative electric field  $E_{\text{rel}} = E/U$  in a logarithmic scale. The simulation was done for (a)  $\text{CDCl}_3$  with  $\epsilon_r \approx 4.8$  as well as (b)  $\text{DMSO}$  with  $\epsilon_r \approx 46.7$ .

the bottom to a height of approximately 100 mm, which shields the outside from the high electric fields present within the cell.

Our high voltage AC supply consisted of a two-stage amplification. The basic waveform, a sine wave of the desired frequency, was provided by a signal generator at a low voltage of around 1 V. The mid-range DC voltage of  $\pm 60$  V was provided by industrial power supplies. Both of these were then fed into a home-built inverter which outputted AC voltage with a maximum of 60 V (peak) or 42 V (RMS) with the appropriate frequency. The voltage level could be tuned by adjusting the voltage of the input wave from the signal generator. Here, it was important to ensure that both the DC input as well as the inverter itself could handle several amperes of current that were necessary for the next amplification step. This next step consisted of a simple high voltage transformer with 2400 turns on the output side. We wound 8 turns on the input side by hand, which yields a theoretical voltage increase by a factor of 300. The output side was directly connected to the head of the field cell with a cable, and it was also equipped with a high voltage probe to directly measure the occurring voltages and waveforms.

### 5.3.3 Testing

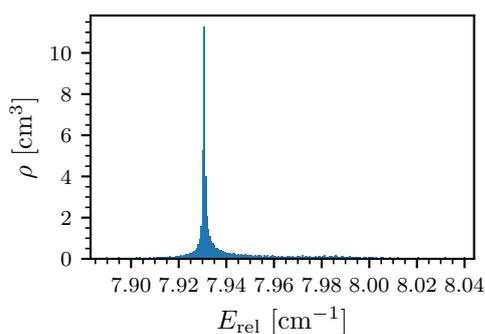
As with the first design we tested the electrical performance of the cell outside the magnet to assure that the setup is electrically stable and poses no harm to the NMR equipment. Unlike with the first design, we had no problems with electrical breakdown in the region of the 3D-printed head or the electrodes, and the cell was stable at the maximum possible voltage of around 4.9 kV RMS. To maximize the achievable voltage, we exploited the fact that the field cell (a capacitance) and the transformer's secondary winding (an inductance) together make a resonant circuit. If the frequency of the AC current in the primary winding matched this resonance, which was between 4–10 kHz



**Figure 5.10.** (a) Deuterium signal of 5% v/v MeCN- $d_3$  in benzene at voltages between 0 and 2.7 kV RMS and 2.3 kHz. (b) Deuterium signal of neat nitrobenzene- $d_5$  at voltages between 0 and 4.9 kV RMS and 4 kHz. Mind the different frequency scale due to the very different line widths of MeCN and nitrobenzene.

depending on the solvent, we were able to approximately double the voltage at the cell. We also tested the effect of the setup without any voltage on the NMR experiment. Here the reference was two shaped tubes stacked into each other, but without any titanium coating or perfluorodecalin between the tubes. Although there were now two layers of titanium in the final field cell, neither the signal intensity nor the shim was significantly affected. Since we were working on a room temperature probe which is not designed to be used with shaped tubes, it was not possible to insert the field cell with a defined orientation relative to the probe's Helmholtz coil, and indeed we did observe a dependence of the tuning and matching of the probe circuit on this rotational positioning of the cell. However, this did not affect the signal intensity or the shim quality.

The simulation of the field distribution was done in a very similar manner as for the 1D cell. It was limited to two dimensions, and the probe part was kept identical. The shaped tubes were modeled as rectangles with semicircles attached to the short sides. The outer side of the inner tube as well as an additional  $40^\circ$  along the semicircle was defined to be the electrode surface. The outer side of the outer tube was set to ground potential, which models the outer shield. The features of the probe such as the outer glass tube and the coil cross sections were left in the simulation, although they obviously did not experience any field as they were completely cut off from the electrodes by the shield. The resulting field distributions for both  $\text{CDCl}_3$  and DMSO as sample



**Figure 5.11.** Histogram of the electric field within the sample region for  $\text{CDCl}_3$  as a solvent. It is normalized such that integration yields the cell area.

solvents are depicted in Fig. 5.9. For  $\text{CDCl}_3$  the relative field strength has increased by an order of magnitude compared to the 1D cell. In the case of the high-permittivity solvent DMSO the difference is even more pronounced, since the 2D design is less sensitive to the sample properties, which is additional advantage. The shielding achieves a complete elimination of the electric field outside in the probe region, and the field within the sample is not reduced at all by this shield. There are no significant field spikes except for the expected ones at the electrode edges, which well-insulated by perfluorodecalin. The homogeneity of this design is excellent, as becomes evident by inspecting the field histogram in Fig. 5.11. In summary these FEM simulations showed that the new design is, in theory, an immense improvement over the 1D cell.

We again attempted to observe RQCs on deuterated solvents to see whether we could induce alignment with our field cell. The concentration of  $\text{MeCN-}d_3$  in benzene was increased to 5% v/v for higher signal-to-noise. These experiments were done without exploiting the resonance of the cell-transformer circuit, so the maximum voltage was 2.7 kV RMS at a frequency of 2.3 kHz. FEM simulations using benzene as a solvent predict a relative field of  $8.9 \text{ cm}^{-1}$ , which would together yield a field of  $24.0 \text{ kV cm}^{-1}$  and an RQC in MeCN of 0.27 Hz. Given a  $^2\text{D}$  line width of 0.12 Hz, such a splitting should be resolved almost to the baseline. Unfortunately, no splitting or even line broadening could be observed, as is seen in Fig. 5.10a. The small broadening at the base of the peak at higher electric field fields could be attributed to slow deterioration of the shim, as this broadening remained even after switching off the field again. To ensure that there are no unknown sample effects that prevent alignment, we tested a sample where electric alignment was observed before by Peshkovsky et al.,<sup>[265]</sup> namely pure nitrobenzene- $d_5$ , and applied voltages up to 4.9 kV RMS using the cell's resonance at 4 kHz. With the FEM-predicted relative field strength of  $3.5 \text{ cm}^{-1}$  for nitrobenzene this yields a predicted RQC of the *para*-deuteron of 1.6 Hz. Interpolation of the previously observed RQCs suggests that an even higher value of around 3.8 Hz should be observed. These larger experimental alignments were hypothesized to come from clustering effects in pure solvents.<sup>[104,265]</sup> At a line width of 2.0 Hz for the *para*-deuteron a splitting or broadening should be visible in any case. However, no such observation was made upon application of the voltage, which is depicted in Fig. 5.10b.

### 5.3.4 Conclusion and Outlook

Even with the new design of the field cell it was not possible to observe electrically induced alignment, and the reason for this failure is not obvious to us. We used AC fields at lower frequencies of 2–4 kHz than in previously reported cases of electrical alignment with 28.5 kHz,<sup>[265,266]</sup> but as ion migration is expected to occur on the timescale of seconds,<sup>[146]</sup> this should not lead to a cancellation of the electric field. One significant difference to previous cells is that in our design there is a layer of glass between the sample and the electrodes. While in principle this has the advantage of eliminating issues with electrode degradation and electrolytic reactions, it reduces the aligning field if the sample permittivity is much larger than the permittivity of the glass. Although we expected the electric field to penetrate the glass unhindered, it is not unthinkable that this is a cause

for the missing alignment. Another difference to the work of Peshkovsky et al.<sup>[265,266]</sup> is that they used pulsed electric fields during an indirect evolution period instead of having a continuous field. This was done to further reduce issues with electrode and sample degradation, as they had an immediate interface between the sample and the electrode. As this is relatively simple to implement, it would be worthwhile to do so to see whether it makes a difference in the result. An independent method of determining the presence of a field within the cell would also be desirable. A possible way would be to make use of the exceptionally large Kerr effect of nitrobenzene.<sup>[270,271]</sup> Upon the application of an electric field, nitrobenzene becomes birefringent and is able to change the polarization of incoming light. Using two polarization filters in an orthogonal arrangement on both sides of the cell, this rotation can be detected. This would correspond to using the field cell as part of a Kerr shutter. As the reaction times of Kerr shutters are known to be on the nanosecond time scale, this should also work for the AC fields employed in this work. Should it be possible to demonstrate alignment using electrical alignment using AC fields, it would also be desirable to redesign the high voltage source. Particularly the input for the high voltage transformer could be designed from off-the-shelf audio equipment, where high power amplifiers and wave generators (e.g., a smartphone) are cheaply available without the need for custom design of electronic components.

Although the present work was not able to achieve alignment via the use of electric fields, we believe that it is worth to persist on this work. It is an effect that has been proven before to be measurable, and the advantages and potential benefits for the field of anisotropic NMR are too significant to give up on.

## 5.4 Materials and Methods

**5.4a Cell Construction** Glass tubes were made from Borosilicate 3.3 glass and all procured from Hilgenberg GmbH. Circular tubes were of “economy” quality. The nominal 5 mm tube had dimensions of 4.95(5) mm o.d., 0.38 mm wall thickness, and 178 mm length, the nominal 3 mm tube had dimensions of 2.95(3) mm o.d., 0.295 mm wall thickness, and 203 mm length. The outer rectangular tubes had dimensions of 4.95(5) × 2.80(5) mm outer cross section, 0.25(3) mm wall thickness, and 178.0(5) mm length, the inner rectangular tubes had 4.2(2) × 1.25(5) mm outer cross section, 0.12(5) mm wall thickness, and 203.0(5) mm length. All dimensions and uncertainties are as reported by the manufacturer. The 1D cell’s head was printed on an in-house fused deposition printer from polylactate filament, the 2D cell’s head was printed using stereolithography with the resin WaterShed XC 11122 by Proto Labs Germany GmbH. The flexible PCBs for the 1D cell electrodes were ordered from LeitOn GmbH. They consisted of a sandwich of 25 μm polyimide, 20 μm epoxy glue, 18 μm copper, 13 μm epoxy glue, and 13 μm polyimide. Perfluorodecalin was procured from Sigma-Aldrich. Titanium coating was done using a 1961 Balzers PVD chamber using resistive heating. A quantity of 0.45 g of titanium granulate was weighed into a tungsten boat and fully evaporated. For the 1D cell, a DC high voltage was provided by two Heinzinger HNC 20.000 - 5 pos./neg. sources (one positive, one negative). For the 2D cell, a home-built AC voltage source was employed.

The frequency was determined by a sine wave at an amplitude of 0–1 V peak-to-peak supplied by a signal generator. This input was amplified by a home-built driver using a  $\pm 60$  V, 4.3 A DC input as a power source. The schematics of this driver can be found in the data collection (Appendix A.1). Its output was then fed into a high-voltage transformer with six primary turns and 2400 secondary turns, which was procured from Voltagezone Electronics e.U.<sup>[272]</sup>

**5.4b FEM** The meshes were generated with Gmsh 4.7<sup>[277]</sup> using its python API. The mesh for the 1D cell consisted of eight concentric circles with diameters 2.36, 2.95, 4.022, 4.2, 4.95, 5.5, 6.1, and 36.95 mm. The appropriate materials were, from inside to outside: sample, glass, perfluorodecalin, polyimide, glass, air, glass, air. Both the electrodes and the NMR coil were modeled as four annulus sectors. The electrodes sectors had a width of 80  $\mu\text{m}$  and thickness of 18  $\mu\text{m}$ , and their radial position matched the layering of the flex-PCB (Par. 5.4a). The polar positions were at angles of  $-29$ ,  $29$ ,  $151$ , and  $-151^\circ$ . The coil sectors had dimensions of  $0.8 \times 0.2$  mm and were positioned at polar angles of  $-61$ ,  $61$ ,  $119$ , and  $-119^\circ$ . The radial positioning was such that they were located in the outermost air region with a gap of 0.1 mm to the glass ring. The mesh element sizes were 10  $\mu\text{m}$  in the sample region, 5  $\mu\text{m}$  in the electrode region, 50  $\mu\text{m}$  in the rest of the cell, and it increased to 500 mm in the large air region around.

The mesh for the 2D cell shared the outer glass ring and NMR coil with the 1D cell. The outermost diameter was reduced to 8 mm. The shaped tubes were modeled as rectangles with semicircles attached to the short sides, and the dimensions were chosen to match the real-world counterpart (Par. 5.4a). The mesh size within the inner tube and up to 1 mm around it was set to 20  $\mu\text{m}$ , falling off to 100  $\mu\text{m}$  in the outer regions of the model. The python scripts for generating these meshes can be found in the data collection (Appendix A.1).

The finite-element analysis was run in Elmer 8.4 using the electrostatics solver.<sup>[278]</sup> The permittivities were set to the values as shown in Table 5.1. The potential of the electrodes were set to  $\pm 0.05$  V yielding a potential difference of 0.01 V; this was done so that the simulation would calculate convenient numbers in the SI-unit  $\text{V m}^{-1}$ , and all field values were then divided by the potential difference. The potential of the NMR coil was set to 0 (ground). For the 1D cell, the outermost circle at a diameter of 36.95 mm corresponds to the metal casing of the probe, and it was set to ground. For the 2D cell, the outermost circle at 8 mm diameter was fit with an infinity boundary condition, and the outer boundary of the outer tube was set to ground (shielding). The solver input files for all simulations mentioned in this work can be found in the data collection (Appendix A.1). Visualization and analysis was done using Paraview 5.8.<sup>[279]</sup>

**Table 5.1.** List of relative permittivities used in FEM simulations.

material	$\epsilon_r$
air	1
polyimide <sup>[273]</sup>	3.1
glass <sup>[274]</sup>	4.6
chloroform <sup>[275]</sup>	4.81
DMSO <sup>[275]</sup>	46.7
perfluorodecalin <sup>[276]</sup>	1.86
benzene <sup>[275]</sup>	2.27
dichloromethane <sup>[275]</sup>	8.93
nitrobenzene <sup>[275]</sup>	34.82

**5.4c NMR Spectroscopy** Deuterated solvents were acquired from Deutero GmbH, disperse red 13 was acquired from Sigma-Aldrich. All substances were used without purification. NMR spectra were acquired on a Bruker Avance III HD spectrometer at a field of 9.4 T (400 MHz) using room temperature probes (QXI/TBO). The sample temperature was set to 298 K. Carbon spectra were acquired using an APT sequence (Bruker standard sequence jmod). Proton and deuterium spectra were acquired using simple pulse-acquire sequences. For proton spectra under alignment conditions (disperse red 13) an acquisition time of 4.0 s was employed, for deuterium we used acquisition times of 4.7 s (nitrobenzene) and 19.1 s (MeCN).

**5.4d Molecular Modeling** For disperse red 13, a conformational search was performed by sampling torsions of freely rotatable bonds, using an energy cut-off of  $21 \text{ kJ mol}^{-1}$  and a minimization convergence of  $1.0 \text{ kJ mol}^{-1} \text{ \AA}^{-1}$ . This was done in MacroModel 11.8<sup>[215]</sup> using the MMFF force field<sup>[151]</sup> in vacuum. We performed a total of 10 000 sampling steps. For MeCN and nitrobenzene, we assumed a single conformation. All structures were geometry-optimized at the B3LYP<sup>[154,156,243,244]</sup>/pcseg-1<sup>[259]</sup> level of theory. We calculated NMR shieldings and electric field gradients at the B3LYP/pcSseg-1<sup>[261]</sup> level of theory using GIAO and an implicit solvent model. All DFT calculations were performed with Gaussian09,<sup>[153]</sup> and Jensen basis sets were retrieved from the Basis Set Exchange.<sup>[260]</sup> The Gaussian log files containing all relevant information can be found in the data collection (Appendix A.1).



# Chapter 6

## Extension and Improvement of the Methanol- $d_4$ NMR Thermometer Calibration

This chapter is a reformatted version of a journal article published in 2021, and is identical to it in text and content.<sup>[280]</sup> The calibration of the Pt100 thermometer against the ITS-90 was performed by the co-authors Stephan Krenek and Dieter Heyer.

### 6.1 Introduction

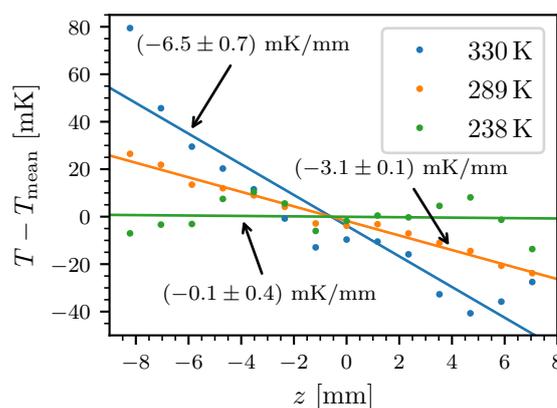
The sample temperature is an important parameter in many NMR experiments. For example, when studying kinetics or thermodynamics temperature is inherently relevant.<sup>[65,281–283]</sup> In other cases, such when accurate chemical shifts need to be determined, temperature variation between experiments can lead to undesirable systematic errors.<sup>[219]</sup> The spectrometer's variable temperature unit has a temperature sensor which is typically located directly below the sample tube, and depending on the flow and input temperature of the tempering gas there are significant systematic differences between the sensor and sample temperature of 5 K and more.<sup>[284]</sup> To correct for these differences, it is necessary to experimentally determine the temperature inside the NMR sample. A very convenient way of doing this is a sample with a strongly temperature dependent chemical shift, where this dependence is known and established. The measured chemical shift then directly relates to the temperature of the sample. Two very common examples for this are methanol and glycol, where the chemical shift difference between hydroxy and aliphatic protons are a very good sensor for temperature.<sup>[285,286]</sup> Both these examples were established more than half a century ago when sensitivity of NMR spectrometers was much lower than today, and therefore fully protonated substances were used. Nowadays however these very high sensitivity samples cause problems due to radiation damping, which makes accurate determination of shifts challenging to impossible. To solve this, Findeisen et al. proposed the use of methanol- $d_4$  with 99.8 % deuteration, which is cheaply available as NMR solvent, and its residual proton signals are readily detectable in a modern NMR spectrometer with sufficient intensity.<sup>[137]</sup> Due to these advantages it has become a popular choice as NMR thermometer, and there are ready-to-use sealed samples of 99.8 % methanol- $d_4$  from various manufacturers. Since the degree of deuteration

has an influence on the exact shift-temperature-relationship, the authors determined an individual calibration curve for this sample. Unfortunately, due to limitations of their hardware, they could only determine this calibration in the temperature range of 282–330 K. In this work, we aim to extend this calibration to the full liquid range of methanol and decrease its uncertainty using temperature sensors with higher accuracy. An additional complication to the calibration of NMR thermometers is the magnetic field influence on the commonly used Pt100 resistance thermometers. Due to the effect of magnetoresistance the measured temperature is subject to an offset depending on the magnetic field, which is significant at the fields present in NMR spectrometers. This effect is known in the literature but was ignored in the work of Findeisen et al.<sup>[287,288]</sup> By acquiring data at different fields we eliminate this effect by extrapolating to zero field.

## 6.2 Results and Discussion

In preliminary experiments we made observations that determined our decisions on the details of the calibration procedure. Firstly, the reproducibility and stability of temperature, which was checked by repeatedly inserting the Pt100 thermometer, depended on the instrument. More specifically, our spectrometers  $\geq 600$  MHz, equipped with cryoprobes, showed excellent stability and reproducibility smaller than the digital resolution of our thermometer of 10 mK. Our 400 MHz instrument, which is equipped with a room temperature probe, showed similar stability but much poorer reproducibility. The temperature after removing and reinserting of the thermometer would vary up to 100 mK. Also, the equilibration time needed after a change in temperature setting until there was a stable temperature at the Pt100 sensor was about an order of magnitude larger on the 400 MHz instrument (1 h vs. 5 min). Tests on a 300 MHz instrument with a room temperature probe showed the same behavior, so we hypothesize that this is a consequence of the probe design. Secondly, the deviation between measured temperature due to magnetoresistance at 400 MHz and 800 MHz is in the order of 300 mK, which is well above the accuracy that we were aiming for. It was therefore necessary to correct for this effect. Therefore, we conducted the main measurement series on the smallest field available with a cryoprobe, namely 600 MHz. Since our cryoprobes are not capable of going below 233 K, we conducted the low temperature part of the calibration at 400 MHz, which is therefore associated with a higher random uncertainty.

Another issue that needed to be addressed is the effect of possible temper-

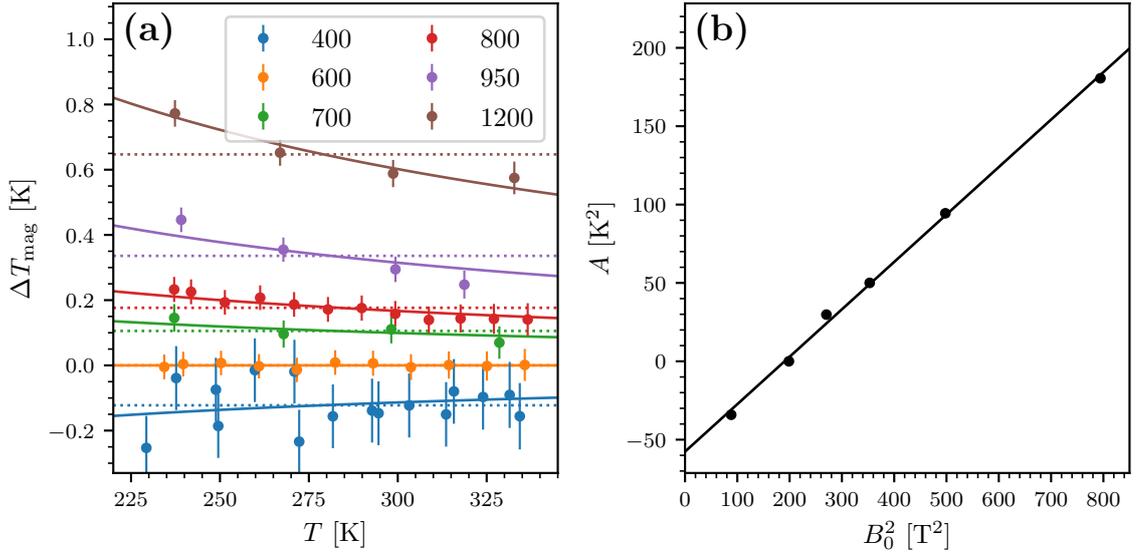


**Figure 6.1.** Temperature differences from the mean as a function of the vertical displacement  $z$  as determined via NMR.

ature gradients and convection, which may lead to systematic errors not captured by a statistical error analysis. Temperature gradients and convection are a phenomenon that is present in an NMR tube under most circumstances, and depends on sample viscosity, temperature settings, and tube diameter, among others.<sup>[289–291]</sup> For our experiments vertical temperature gradients are most problematic, as they could lead to systematic errors if the gradients are large compared to the achieved accuracy in sensor positioning. We therefore estimated these gradients from both experimental approaches. First, we acquired slice-selective spectra of the methanol standard sample at temperatures of 330, 289, and 238 K (roughly corresponding to the temperature range of our cryoprobes) and calculated the temperatures from the observed chemical shift differences. By fitting a linear equation to these data, we estimated the temperature- $z$ -gradients  $\partial_z T$  to be  $-6.5(7)$ ,  $-3.0(1)$ , and  $-0.1(4)$  mK mm<sup>-1</sup> at 330, 289, and 238 K, respectively (the positive  $z$ -direction is upward). These data is depicted in Fig. 6.1. On the other hand, we measured the temperature using the Pt100-thermometer while deliberately displacing the sensor by  $\pm 3$  mm in the vertical direction. While there was no measurable difference at 289 K and 238 K, there was a barely measurable difference of 10–20 mK for this displacement at 330 K. This is in agreement with the data from the NMR measurements. Since both ways of measuring the temperature (NMR and Pt100) are averaging over a similar volume, and the Pt100 sensor can be centered in the NMR-active volume to an accuracy of approximately  $\pm 1$  mm, we concluded that temperature gradients in the sample do not lead to a significant systematic deviation. A statistical uncertainty contribution resulting from these measurements is taken into account accordingly as a small contribution in the total uncertainty budget (see Par. 6.3d).

In the first step for the main measurement series, we acquired 11 data pairs of temperature ( $T$ ) as measured by the Pt100 sensor and chemical shift difference ( $\Delta\delta$ ) between the methyl and the hydroxyl peak in the range of 234–336 K in steps of approximately 10 K. Since the purpose of the calibration is to find the temperature as a function of shift difference, we chose  $\Delta\delta$  to be the independent variable for our model (and not vice versa). We chose a polynomial as the model function. However, a second degree polynomial as used by Findeisen et al. proved to be insufficient to describe the data, and the residuals showed a very clear trend of a third degree component. A standard approach to determine the necessary model complexity and to avoid overfitting is cross-validation:<sup>[208]</sup> We randomly and repeatedly (10 000 times) chose eight data points ( $\approx 70\%$ ) to perform the fit, and then checked the root mean square deviation (RMSD) of the model with the three remaining data points. This RMSD had a minimum for a fourth degree polynomial, which is then the appropriate model to choose. With this model, the RMSD of the fit with all data points was as low as 6.3 mK (similar to Fig. 6.3b above 230 K). This indicates that our data is associated with very little statistical scatter.

It is known that the magnitude of magnetoresistance of a given sensor depends on the geometry of its winding.<sup>[287,292]</sup> A correction must therefore be adjusted to the specific sensor model. Therefore, we determined this correction ourselves by measuring additional data at different magnetic fields between 9.4 T and 28.2 T (400–1200 MHz proton resonance frequency), with at least 4 points per field between 235 K and 335 K. The field dependent deviation cannot be measured in absolute terms since this would

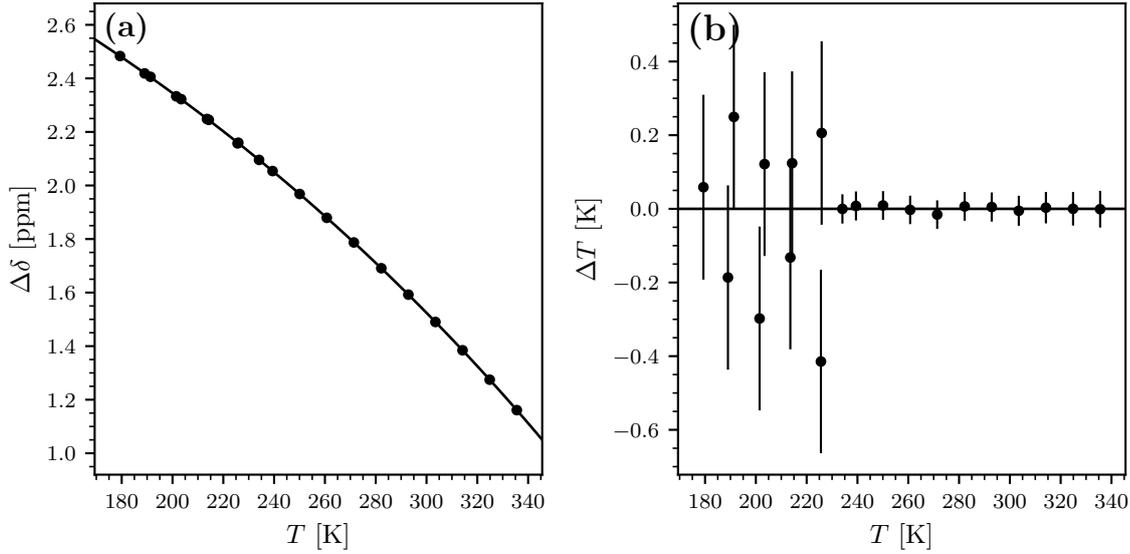


**Figure 6.2.** (a) Deviation of Pt100 temperatures as a function of field and temperature, relative to the uncorrected calibration at 14.1 T (600 MHz), and the proposed models: temperature-independent deviation (dotted line) and deviation inversely proportional to temperature (solid line). The error bars show  $2\sigma$  for each data point. (b) Deviation constant  $A$  as a function of the square magnetic field and its linear fit function.

require data at zero field which we cannot acquire. The deviation is therefore determined relative to a reference field. Since it is also very challenging to impossible to acquire data at different fields at the exact same true temperature (i.e., the same  $\Delta\delta$ ), we simply used the uncorrected model that we had determined at 14.1 T (600 MHz proton resonance frequency) before as a reference. This model gives the theoretical Pt100-temperature at 14.1 T as a function of  $\Delta\delta$ . Comparing this with the measured Pt100-temperature at the given field gives the field-dependent temperature deviation relative to 14.1 T. This deviation is known to be proportional to the square of the magnetic field.<sup>[287,288]</sup> An appropriate model for the temperature dependence of the magnetic field error is less obvious, and we explored two possibilities. The simplest solution is to ignore it and to assume that the magnetic field deviation is independent of temperature. However, after looking into the data from Brandt et al.<sup>[287]</sup> we realized the deviation in our temperature range can be more accurately described by

$$\Delta T_{\text{mag}} = \frac{A}{T}, \quad (6.1)$$

where  $\Delta T_{\text{mag}}$  is the temperature deviation,  $T$  the (absolute) temperature, and  $A$  a free parameter. This model is equally simple as the assumption of a constant deviation, as it only contains a single parameter, but it improves the agreement (RMSD) with our data by a factor of 2.4 and was therefore our model of choice. The field-dependent deviation as well as the two models can be seen in Fig. 6.2a. The parameter  $A$  then exhibits the



**Figure 6.3.** (a) Fit of chemical shift difference versus corrected temperature. (b) Residuals of the fit as a function of temperature. The data  $<230$  K were acquired at 9.4 T (400 MHz) with a  $\text{LN}_2$  heat exchanger and show a much larger scatter. The error bars show  $2\sigma$  for each data point.

aforementioned square-dependence on the magnetic field:

$$A = kB_0^2 + A_{\text{ref}}, \quad (6.2)$$

where  $k$  is the field-dependence parameter specific to the sensor and  $A_{\text{ref}}$  is an offset that is a consequence of using a non-zero field (14.1 T) as a reference for the temperature deviation. The least-squares fit yielded  $k = 0.3026(66) \text{ K}^2 \text{ T}^{-2}$  and  $A_{\text{ref}} = -57.6(29) \text{ K}^2$ . Fig. 6.2b shows  $A$  as a function of the square field as well as the fitted Eq. (6.2).

With this data it is possible to correct for the systematic effects introduced by the magnetic field. The measured temperatures  $T_{\text{Pt100}}$  were corrected as follows:

$$T_{\text{corr}} = T_{\text{Pt100}} - \frac{kB_0^2}{T_{\text{Pt100}}}. \quad (6.3)$$

For the final calibration curve the 11 data points between 235 K and 335 K at 14.1 T were combined with 9 data points between 180 K and 230 K at 9.4 T, both corrected for the magnetic field shift. The data from 9.4 T was acquired in two different sessions on two different probes (TBO and QXI) using a  $\text{LN}_2$  heat exchanger to further cool down the variable temperature (VT) gas. Indeed, it seemed as the data from 9.4 T was also affected by systematic effects. These could partly be associated with the order of data acquisition (i.e., from high to low temperatures or vice versa) and possible incomplete thermal equilibration after temperature changes. However, this cannot explain all effects. We included both datasets in the fitting assuming that the true value lies somewhere in between.

This combined data set was then used again as basis for a fourth-degree polynomial  $T$  as a function of  $\Delta\delta$  in ppm, which is the main result of this work :

$$T = \sum_{i=0}^4 a_i \Delta\delta^i. \quad (6.4)$$

Table 6.1 contains the polynomial coefficients of the fit function, and Fig. 6.3a shows the graph of this function including the data points. In Fig. 6.3b the deviation of the individual data points from the fitted curve is depicted.

We used a Monte-Carlo approach to estimate the uncertainty of this calibration curve (see Fig. 6.4).<sup>[293]</sup> Each data point is associated with an uncertainty from different contributions (see Par. 6.3d), and we generated 100 000 datasets where we added an additional, synthetic deviation to each point. Here it was important to distinguish between the different contributions to the uncertainty, because the contributions from the temperature regulation  $\sigma_{\text{reg}}$ , the thermometer  $\sigma_{\text{therm}}$ , and the sensor positioning  $\sigma_z$  affect each data point independently, whereas the contribution from the magnetic field correction  $\sigma_k$  affects each data point the same way. Each synthetic data point  $T_{i,j}^{\text{syn}}$ , where  $i$  is the index for the corresponding experimental data point  $T_i^{\text{meas}}$  and  $j$  the index for the synthetic data set, was calculated as follows:

**Table 6.1.** Polynomial coefficients for the fit function in Eq. (6.4).

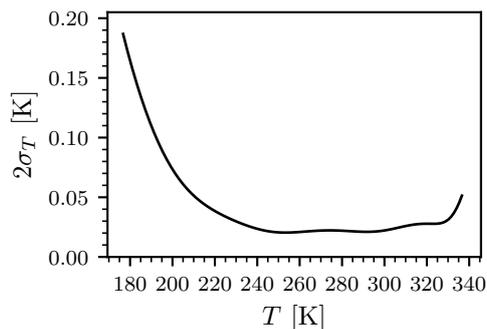
i	$a_i/\text{K ppm}^{-i}$
0	416.4745
1	-39.5133
2	-36.0620
3	11.4869
4	-2.4340

$$T_{i,j}^{\text{syn}} = T_i^{\text{meas}} + x_{i,j} \sqrt{\sigma_{\text{reg}}^2 + \sigma_{\text{therm}}^2 + (\sigma_z \partial_z T)^2} + x_i \frac{\sigma_k B_0^2}{T_i^{\text{meas}}}, \quad (6.5)$$

where  $x_{i,j}$  and  $x_i$  are standard normally distributed ( $\sigma = 1$ ) pseudorandom numbers. For each synthetic data set  $j$  we determined the fourth-degree polynomial fit function. From this set of fit functions, we estimated the standard deviation of temperature  $\sigma_T$  as a function of temperature, which we interpret as uncertainty of our calibration. In Fig. 6.4 we have plotted  $2\sigma_T$ , which represents a confidence interval of 95%. Roughly speaking, this interval is  $\pm 25$  mK in the range between 330 K and 240 K and then increases to  $\pm 190$  mK at 180 K.

## 6.3 Material and Methods

**6.3a NMR Spectroscopy** We used commercial NMR standard reference samples from Bruker Biospin. They are water-free 99.8% methanol- $d_4$  flame-sealed under inert atmosphere. NMR experiments were conducted on Bruker Avance III HD spectrometers (400, 600, and 950 MHz) and Avance Neo spectrometers (700, 800, and 1200 MHz) equipped with a TBO or QXI probe (400 MHz) or Q/TCI cryoprobes (all other fields). The probe at 1200 MHz was a 3 mm probe. For temperatures below 230 K we used a liquid nitrogen heat exchanger to further cool down the VT gas flow (only 400 MHz). All spectra were acquired using a simple pulse-acquire sequence with a single scan. The acquisition time was



**Figure 6.4.** Uncertainty ( $2\sigma_T$ ) of the calibration as a function of temperature as determined by a Monte Carlo analysis.

5.85 s and the spectral width 14 ppm. Before Fourier transform exponential line broadening between 0.3 Hz and 3 Hz (depending on achieved line shape/shim) and a zero filling to 1 048 576 points was applied, yielding a spectrum with digital resolution of 0.013 ppb. Peak picking was done in MestreNova 11. Imaging experiments were performed using slice-selective excitation with a  $z$ -gradient of  $\approx 0.5 \text{ T m}^{-1}$  and an EBURP2<sup>[294]</sup> pulse with a length of 250  $\mu\text{s}$  ( $\approx 15 \text{ kHz}$  bandwidth). 14 slices were acquired with pulse offsets in steps of 25 kHz (corresponds to  $\approx 1.2 \text{ mm}$ ).

**6.3b Pt100 Resistance Thermometer** We used a Pt100 sensor acquired from Klasmeier GmbH with model number P100/2515. It is a wire-wound ceramic sensor with a length of 25 mm and a diameter of 1.5 mm in the accuracy class 1/10 DIN B. We connected a low temperature compatible PTFE cable in 4-wire configuration. Resistance/Temperature measurement was done using a Greisinger GMH 3750 thermometer. The sensor was inserted into an NMR tube and fixed in place to be exactly ( $\pm 1 \text{ mm}$ ) in the center of the NMR coil (active volume). It was submersed in ethanol to a filling height of 40 mm to mimic the conditions of the NMR sample as closely as possible. The whole sensor-thermometer measurement chain was calibrated traceable to the International System of Units. For this purpose, the display values of the thermometer were compared with the measurements of a standard platinum resistance thermometer (SPRT) that was directly calibrated at the fixed points of the International Temperature Scale (ITS-90). The comparison was carried out in stirred alcohol and water baths with high homogeneity and stability, so that uncertainties of a few millikelvin can be achieved.<sup>[295]</sup> From the deviations measured several times at intervals of approximately 20 K, a correction function was determined for the display values. Its total uncertainty results from the uncertainties of the calibration and the confidence band of the fit. It varies with temperature from about 19 mK to 27 mK and is used as standard deviation  $\sigma_{\text{therm}}$  in the data analysis.

**6.3c Measurement Procedure** To measure both spectra and temperature at different temperatures, we alternated between the NMR sample and the Pt100 sensor. The temperature setting of the spectrometer was changed while the Pt100 sensor was inserted. That way we could record a temperature value both before and after acquisition of the NMR spectrum at all given temperatures. When using the  $\text{LN}_2$  heat exchanger, the measured temperature would fluctuate, so we recorded an upper and lower limit of this fluctuation. These values were used to estimate the temperature stability of the spectrometer's temperature regulation. Before recording the temperature or the spectrum we waited until stability within the measurement accuracy was reached (approx. 5 min).

At 400 MHz the data was associated with much larger fluctuations, so data was acquired in two sessions on two different probes (TBO and QXI).

**6.3d Fitting and Data Analysis** All analysis and data evaluation were performed using Python3 relying heavily on the SciPy ecosystem.<sup>[216–218]</sup> The fits to Eqs. (6.1) and (6.2) were least-squares solutions with no weighting. The uncertainty of the fit parameters of Eq. (6.2) was calculated as the root of the estimated variance reported by the fit routine `numpy.polyfit`.<sup>[296]</sup> To estimate the standard deviation  $\sigma_{\text{reg}}$  of the temperature regulation in the spectrometer, we calculated the root mean square deviation (RMSD) from the temperatures recorded before and after the acquisition of the spectrum. We assumed this standard deviation of the temperature control is independent of temperature within one measurement series, so we calculated a common  $\sigma_{\text{reg}}$  for all points by taking the root mean square (RMS) of the individual RMSDs. The standard deviation of the temperature measurement  $\sigma_{\text{therm}}$  was estimated during the sensor calibration as a function of temperature. The contribution from the magnetic field correction  $\sigma_{\text{corr}}$  was estimated by using a linear propagation of uncertainty for Eq. (6.3) using the standard deviation of  $k$ . Note that this is a minor contribution, adding less than 5% to the total uncertainty. The contribution from temperature gradients was calculated by estimating the uncertainty of the sensor position to be  $\sigma_z = 1$  mm, and to assume a linear model for the temperature dependence of the temperature gradient in a cryoprobe:  $\partial_z T = -6.9 \cdot 10^{-5} \text{ mm}^{-1} T + 16.5 \text{ mK mm}^{-1}$ , obtained as least-squares fit to the three gradient data points determined from slice-selective NMR experiments. The contribution from gradients to the data at 400 MHz was neglected as this is a minor contribution and these data are already associated with a much higher uncertainty. The combined standard deviation of each temperature point was then calculated as the root sum of squares (RSS) of these contributions. The polynomial fits for the function  $T(\Delta\delta)$  were determined as the least-squares solutions using the inverse of the uncertainties as weighting. The uncertainty of the fit itself was calculated with a Monte-Carlo approach, as described above.

# References

- [1] I. Newton, *Philosophiae Naturalis Principia Mathematica*, Royal Society, London, **1687**.
- [2] J.-L. Lagrange, *Mécanique Analytique*, 1re éd., Chez La Veuve Desaint, Paris, **1788**.
- [3] W. R. Hamilton, “On a General Method in Dynamics; by Which the Study of the Motions of All Free Systems of Attracting or Repelling Points is Reduced to the Search and Differentiation of One Central Relation, or Characteristic Function”, *Phil. Trans.* **1834**, *124*, 247–308, DOI 10.1098/rstl.1834.0017.
- [4] J. F. Encke, “Schreiben des Herrn Professors Encke an den Herausgeber.”, *Astron. Nachr.* **1846**, *580*, 49–52, DOI 10.1002/asna.18470250402.
- [5] J. C. Maxwell, “On Physical Lines of Force”, *Philos. Mag.* **1861**, *24*, 12–24, DOI 10.1080/14786431003659180.
- [6] A. A. Michelson, E. W. Morley, “On the Relative Motion of the Earth and the Luminiferous Ether”, *Am. J. Sci.* **1887**, *34*, 333–345, DOI 10.2475/ajs.s3-34.203.333.
- [7] A. Einstein, “Zur Elektrodynamik bewegter Körper.”, *Ann. Phys. (Berlin)* **1905**, *17*, 891–921, DOI 10.1002/andp.19053221004.
- [8] A. Einstein, “Die Feldgleichungen der Gravitation.”, *Ber. Akad. Wiss. Berlin* **1915**, 844–847, DOI 10.1002/3527608958.ch5.
- [9] L. Rayleigh, “Remarks upon the Law of Complete Radiation”, *Philos. Mag.* **1900**, *49*, 539–540, DOI 10.1080/14786440009463878.
- [10] J. H. Jeans, “On the Partition of Energy between Matter and Aether.”, *Philos. Mag.* **1905**, *10*, 91–98, DOI 10.1080/14786440509463348.
- [11] M. Planck, “Über das Gesetz der Energieverteilung im Normalspectrum”, *Ann. Phys. (Berlin)* **1901**, *309*, 553–563, DOI 10.1002/andp.19013090310.
- [12] H. Hertz, “Über den Einfluss des ultravioletten Lichtes auf die electriche Entladung”, *Ann. Phys. (Berlin)* **1887**, *267*, 983–100, DOI 10.1002/andp.18872670827.
- [13] A. Einstein, “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”, *Ann. Phys. (Berlin)* **1905**, *322*, 132–148, DOI 10.1002/andp.19053220607.
- [14] E. Rutherford, “The Scattering of  $\alpha$  and  $\beta$  Particles by Matter and the Structure of the Atom”, *Philos. Mag.* **1911**, *21*, 669–688, DOI 10.1080/14786440508637080.
- [15] N. Bohr, “On the Constitution of Atoms and Molecules”, *Philos. Mag.* **1913**, *26*, 1–25, DOI 10.1080/14786441308634955.
- [16] L. de Broglie, *Recherches sur la Théorie des Quanta*, Doctoral Thesis, University of Paris, France, **1924**.
- [17] A. Einstein, “Quantisierung als Eigenwertproblem”, *Ann. Phys. (Berlin)* **1924**, *384*, 361–377, DOI 10.1002/andp.19263840404.
- [18] M. Born, P. Jordan, “Zur Quantenmechanik”, *Z. Phys.* **1925**, *34*, 858–888, DOI 10.1007/bf01328531.

## References

- [19] M. Born, W. Heisenberg, P. Jordan, “Zur Quantenmechanik II.”, *Z. Phys.* **1926**, *35*, 557–615, DOI 10.1007/bf01379806.
- [20] W. Gerlach, O. Stern, “Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld”, *Z. Phys.* **1922**, *9*, 349–352, DOI 10.1007/bf01326983.
- [21] G. E. Uhlenbeck, S. Goudsmit, “Spinning Electrons and the Structure of Spectra”, *Nature* **1926**, *117*, 264–265, DOI 10.1038/117264a0.
- [22] W. Pauli, “Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren”, *Z. Phys.* **1925**, *31*, 765–783, DOI 10.1007/bf02980631.
- [23] W. Pauli, “Zur Quantenmechanik des magnetischen Elektrons”, *Z. Phys.* **1927**, *43*, 601–623, DOI 10.1007/bf01397326.
- [24] P. A. M. Dirac, “The Quantum Theory of the Electron”, *Proc. Math. Phys. Eng. Sci.* **1928**, *117*, 610–624, DOI 10.1098/rspa.1928.0023.
- [25] R. Frisch, O. Stern, “Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. I”, *Z. Phys.* **1933**, *85*, 4–16, DOI 10.1007/bf01330773.
- [26] I. Estermann, O. Stern, “Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. II”, *Z. Phys.* **1933**, *85*, 17–24, DOI 10.1007/bf01330774.
- [27] I. Estermann, O. Stern, “Über die magnetische Ablenkung von isotopen Wasserstoffmolekülen und das magnetische Moment des ‚Deutons‘”, *Z. Phys.* **1933**, *86*, 132–134, DOI 10.1007/bf01340190.
- [28] I. I. Rabi, J. R. Zacharias, S. Millman, P. Kusch, “A New Method of Measuring Nuclear Magnetic Moment”, *Phys. Rev.* **1938**, *53*, 318–318, DOI 10.1103/PhysRev.53.318.
- [29] J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey, J. R. Zacharias, “The Magnetic Moments of the Proton and the Deuteron – the Radiofrequency Spectrum of H<sub>2</sub> In Various Magnetic Fields”, *Phys. Rev.* **1939**, *56*, 728–743, DOI 10.1103/PhysRev.56.728.
- [30] G. R. Eaton, S. S. Eaton, K. Salikhov, *Foundations of Modern EPR*, World Scientific, Singapur, **1998**.
- [31] E. Zavoisky, *Paramagnetic Absorption in Perpendicular and Parallel Fields for Salts, Solutions and Metals*, Doctoral Thesis, Kazan State University, USSR, **1944**.
- [32] F. Bloch, “Nuclear Induction”, *Phys. Rev.* **1946**, *70*, 460–474, DOI 10.1103/physrev.70.460.
- [33] E. M. Purcell, H. C. Torrey, R. V. Pound, “Resonance Absorption by Nuclear Magnetic Moments in a Solid”, *Phys. Rev.* **1946**, *69*, 37–38, DOI 10.1103/physrev.69.37.
- [34] J. T. Arnold, S. S. Dharmatti, M. E. Packard, “Chemical Effects on Nuclear Induction Signals from Organic Compounds”, *J. Chem. Phys.* **1951**, *19*, 507–507, DOI 10.1063/1.1748264.
- [35] J. T. Arnold, “Magnetic Resonances of Protons in Ethyl Alcohol”, *Phys. Rev.* **1956**, *102*, 136–150, DOI 10.1103/PhysRev.102.136.
- [36] W. A. Anderson, “Nuclear Magnetic Resonance Spectra of Some Hydrocarbons”, *Phys. Rev.* **1956**, *102*, 151–167, DOI 10.1103/PhysRev.102.151.
- [37] R. Freeman, J. Robert in *NMR at Very High Field. NMR (Basic Principles and Progress), Vol. 25*, (Ed.: J. Robert), Springer, Berlin, Heidelberg, **1991**, DOI 10.1007/978-3-642-48814-6\_1.
- [38] R. R. Ernst, W. A. Anderson, “Application of Fourier Transform Spectroscopy to Magnetic Resonance”, *Rev. Sci. Instrum.* **1966**, *37*, 93–102, DOI 10.1063/1.1719961.
- [39] R. R. Ernst, “Nuclear Magnetic Double Resonance with an Incoherent Radio-Frequency Field”, *J. Chem. Phys.* **1966**, *45*, 3845–3861, DOI 10.1063/1.1727409.

- [40] J. W. Cooley, J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series", *Math. Comp.* **1965**, *19*, 297–301, DOI 10.1090/s0025-5718-1965-0178586-1.
- [41] R. R. Ernst, *Nobelpreisträger aus Winterthur – Autobiografie*, Hier und Jetzt, Verlag für Kultur und Geschichte GmbH, Baden, Schweiz, **2020**.
- [42] W. P. Aue, E. Bartholdi, R. R. Ernst, "Two-Dimensional Spectroscopy. Application to Nuclear Magnetic Resonance", *J. Chem. Phys.* **1976**, *64*, 2229–2246, DOI 10.1063/1.432450.
- [43] A. Kumar, R. R. Ernst, K. Wüthrich, "A Two-Dimensional Nuclear Overhauser Enhancement (2D NOE) Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules", *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6, DOI 10.1016/0006-291x(80)90695-6.
- [44] G. Bodenhausen, R. Freeman, "Correlation of Chemical Shifts of Protons and  $^{13}\text{C}$ ", *J. Am. Chem. Soc.* **1978**, *100*, 320–321, DOI 10.1021/ja00469a073.
- [45] G. Bodenhausen, D. J. Ruben, "Natural Abundance  $^{15}\text{N}$  NMR By Enhanced Heteronuclear Spectroscopy", *Chem. Phys. Lett.* **1980**, *69*, 185–189, DOI 10.1016/0009-2614(80)80041-8.
- [46] A. Bax, R. H. Griffey, B. L. Hawkins, "Correlation of Proton and  $^{15}\text{N}$  Chemical Shifts by Multiple Quantum NMR", *J. Magn. Reson.* **1983**, *55*, 301–315, DOI 10.1016/0022-2364(83)90241-X.
- [47] C. Griesinger, O. W. Sørensen, R. R. Ernst, "A Practical Approach to Three-Dimensional NMR Spectroscopy", *J. Magn. Reson. (1969)* **1987**, *73*, 574–579, DOI 10.1016/0022-2364(87)90027-8.
- [48] H. Oschkinat, C. Griesinger, P. J. Kraulis, O. W. Sørensen, R. R. Ernst, A. M. Gronenborn, G. M. Clore, "Three-Dimensional NMR Spectroscopy of a Protein in Solution", *Nature* **1988**, *332*, 374–376, DOI 10.1038/332374a0.
- [49] S. W. Fesik, E. R. P. Zuiderweg, "Heteronuclear Three-Dimensional NMR-Spectroscopy - a Strategy for the Simplification of Homonuclear Two-Dimensional NMR-Spectra", *J. Magn. Reson.* **1988**, *78*, 588–593, DOI 10.1016/0022-2364(88)90144-8.
- [50] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, "Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins", *J. Magn. Reson.* **1990**, *89*, 496–514, DOI 10.1016/0022-2364(90)90333-5.
- [51] L. E. Kay, G. M. Clore, A. Bax, A. M. Gronenborn, "4-Dimensional Heteronuclear Triple-Resonance NMR Spectroscopy of Interleukin-1-Beta in Solution", *Science* **1990**, *249*, 411–414, DOI 10.1126/science.2377896.
- [52] K. Kazimierczuk, V. Orekhov, "Non-Uniform Sampling: Post-Fourier Era of NMR Data Collection and Processing", *Magn. Reson. Chem.* **2015**, *53*, 921–926, DOI 10.1002/mrc.4284.
- [53] F. A. Nelson, H. E. Weaver, "Nuclear Magnetic Resonance Spectroscopy in Superconducting Magnetic Fields", *Science* **1964**, *146*, 223–232, DOI 10.1126/science.146.3641.223.
- [54] P. C. Lauterbur, "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance", *Nature* **1973**, *242*, 190–191, DOI 10.1038/242190a0.
- [55] P. Mansfield, P. K. Grannell, "NMR 'Diffraction' in Solids", *J. Phys. C Solid State Phys.* **1973**, *6*, L422–L426, DOI 10.1088/0022-3719/6/22/007.
- [56] L. E. Kay, D. A. Torchia, A. Bax, "Backbone Dynamics of Proteins as Studied by  $^{15}\text{N}$  Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease", *Biochemistry* **1989**, *28*, 8972–8979, DOI 10.1021/bi00449a003.
- [57] C. Deverell, R. E. Morgan, J. H. Strange, "Studies of Chemical Exchange by Nuclear Magnetic Relaxation in the Rotating Frame", *Mol. Phys.* **1970**, *18*, 553–559, DOI 10.1080/0026897700100611.

## References

- [58] D. G. Davis, M. E. Perlman, R. E. London, "Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes via the  $T_{1\rho}$  and  $T_2$  (CPMG) Methods", *J. Magn. Reson. B* **1994**, *104*, 266–275, DOI 10.1006/jmrb.1994.1084.
- [59] N. L. Fawzi, J. F. Ying, R. Ghirlando, D. A. Torchia, G. M. Clore, "Atomic-Resolution Dynamics on the Surface of Amyloid- $\beta$  Protofibrils Probed by Solution NMR", *Nature* **2011**, *480*, 268–272, DOI 10.1038/nature10577.
- [60] P. Vallurupalli, G. Bouvignies, L. E. Kay, "Studying 'Invisible' Excited Protein States in Slow Exchange with a Major State Conformation", *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161, DOI 10.1021/ja3001419.
- [61] N. A. J. Van Nuland, V. Forge, J. Balbach, C. M. Dobson, "Real-Time NMR Studies of Protein Folding", *Acc. Chem. Res.* **1998**, *31*, 773–780, DOI 10.1021/ar9700791.
- [62] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, "Investigation of Exchange Processes by Two-Dimensional NMR-Spectroscopy", *J. Chem. Phys.* **1979**, *71*, 4546–4553, DOI 10.1063/1.438208.
- [63] A. G. Palmer, "Chemical Exchange in Biomacromolecules: Past, Present, and Future", *J. Magn. Reson.* **2014**, *241*, 3–17, DOI 10.1016/j.jmr.2014.01.008.
- [64] J. Reddy, S. Pratihari, D. Ban, S. Frischkorn, S. Becker, C. Griesinger, D. Lee, "Simultaneous Determination of Fast and Slow Dynamics in Molecules Using Extreme CPMG Relaxation Dispersion Experiments", *J. Biomol. NMR* **2018**, *70*, 1–9, DOI 10.1007/s10858-017-0155-0.
- [65] F. A. A. Mulder, A. Mittermaier, B. Hon, F. W. Dahlquist, L. E. Kay, "Studying Excited States of Proteins by NMR Spectroscopy", *Nat. Struct. Biol.* **2001**, *8*, 932–935, DOI 10.1038/nsb1101-932.
- [66] L. Fielding, "NMR Methods for the Determination of Protein-Ligand Dissociation Constants", *Curr. Top. Med. Chem.* **2003**, *3*, 39–53, DOI 10.2174/1568026033392705.
- [67] M. P. Williamson, "Using Chemical Shift Perturbation to Characterise Ligand Binding", *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16, DOI 10.1016/j.pnmrs.2013.02.001.
- [68] C. P. Butts, C. R. Jones, J. N. Harvey, "High Precision NOEs as a Probe for Low Level Conformers – a Second Conformation of Strychnine", *Chem. Commun.* **2011**, *47*, 1193–1195, DOI 10.1039/c0cc04114a.
- [69] A. Kolmer, L. J. Edwards, I. Kuprov, C. M. Thiele, "Conformational Analysis of Small Organic Molecules Using NOE and RDC Data: A Discussion of Strychnine and  $\alpha$ -Methylene- $\gamma$ -Butyrolactone", *J. Magn. Reson.* **2015**, *261*, 101–109, DOI 10.1016/j.jmr.2015.10.007.
- [70] R. D. Thomas, M. T. Clarke, R. M. Jensen, T. C. Young, "Fluxional Exchange of Tert-Butyllithium Tetramers from Temperature-Dependent  $^{13}\text{C}$ - $^6\text{Li}$  Coupling", *Organometallics* **1986**, *5*, 1851–1857, DOI 10.1021/om00140a016.
- [71] R. Lichtenberger, S. O. Baumann, M. Bendova, M. Puchberger, U. Schubert, "Modification of Aluminium Alkoxides With Dialkylmalonates", *Monatsh. Chem.* **2010**, *141*, 717–727, DOI 10.1007/s00706-010-0317-1.
- [72] E. Pretsch, P. Bühlmann, C. Affolter, *Structure Determination of Organic Compounds - Tables of Spectral Data*, 4th ed., Springer Science & Business Media, Berlin Heidelberg, **2013**.
- [73] S. Bienz, L. Bigler, T. Fox, H. Meier, *Spektroskopische Methoden in der Organischen Chemie*, 9th ed., Georg Thieme Verlag, Stuttgart, **2016**.
- [74] J. Aires-de-Sousa, M. C. Hemmer, J. Gasteiger, "Prediction of  $^1\text{H}$  NMR Chemical Shifts Using Neural Networks", *Anal. Chem.* **2002**, *74*, 80–90, DOI 10.1021/ac010737m.
- [75] D. Banfi, L. Patiny, "www.nmrd.org: Resurrecting and Processing NMR Spectra Online", *Chimia* **2008**, *62*, 280–281, DOI 10.2533/chimia.2008.280.

- [76] M. W. Lodewyk, M. R. Siebert, D. J. Tantillo, "Computational Prediction of  $^1\text{H}$  and  $^{13}\text{C}$  Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry", *Chem. Rev.* **2012**, *112*, 1839–1862, DOI 10.1021/cr200106v.
- [77] P. Edman, "A Method for the Determination of the Amino Acid Sequence in Peptides", *Arch. Biochem.* **1949**, *22*, 475–476.
- [78] P. Edman, G. Begg, "A Protein Sequenator", *Eur. J. Biochem.* **1967**, *1*, 80–91, DOI 10.1111/j.1432-1033.1967.tb00047.x.
- [79] K. F. Medzihradsky, "Peptide Sequence Analysis", *Methods Enzymol.* **2005**, *402*, 209–244, DOI 10.1016/S0076-6879(05)02007-0.
- [80] G. P. Moss, "Basic Terminology of Stereochemistry", *Pure Appl. Chem.* **1996**, *68*, 2193–2222, DOI 10.1351/pac199668122193.
- [81] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, N. J. Skelton, M. Rance, *Protein NMR Spectroscopy*, Elsevier LTD, Oxford, **2007**, 996 Pages.
- [82] M. Karplus, "Vicinal Proton Coupling in Nuclear Magnetic Resonance", *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871, DOI 10.1021/ja00901a059.
- [83] N. Matsumori, D. Kaneno, M. Murata, H. Nakamura, K. Tachibana, "Stereochemical Determination of Acyclic Structures Based on Carbon-Proton Spin-Coupling Constants. A Method of Configuration Analysis for Natural Products", *J. Org. Chem.* **1999**, *64*, 866–876, DOI 10.1021/jo981810k.
- [84] V. Bystrov, "Spin-Spin Coupling and the Conformational States of Peptide Systems", *Prog. Nucl. Magn. Reson.* **1976**, *10*, 41–82, DOI 10.1016/0079-6565(76)80001-5.
- [85] A. Pardi, M. Billeter, K. Wüthrich, "Calibration of the Angular Dependence of the Amide Proton- $\text{C}^\alpha$  Proton Coupling Constants,  $^3j_{\text{HN}\alpha}$ , in a Globular Protein - Use of  $^3j_{\text{HN}\alpha}$  for Identification of Helical Secondary Structure", *J. Mol. Biol.* **1984**, *180*, 741–751, DOI 10.1016/0022-2836(84)90035-4.
- [86] J. S. Hu, A. Bax, "Determination of  $\phi$  and  $\chi_1$  angles in proteins from  $^{13}\text{C}$ - $^{13}\text{C}$  three-bond  $J$  couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond?", *J. Am. Chem. Soc.* **1997**, *119*, 6360–6368, DOI 10.1021/ja970067v.
- [87] M. R. M. Koos, A. Navarro-Vazquez, C. Anklin, R. R. Gil, "Computer-Assisted 3D Structure Elucidation (CASE-3D): The Structural Value of  $^2J_{\text{CH}}$  in Addition to  $^3J_{\text{CH}}$  Coupling Constants", *Angew. Chem. Int. Ed.* **2020**, *59*, 3938–3941, DOI 10.1002/anie.201915103.
- [88] D. Wishart, B. Sykes, F. Richards, "Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure", *J. Mol. Biol.* **1991**, *222*, 311–333, DOI 10.1016/0022-2836(91)90214-q.
- [89] D. S. Wishart, B. D. Sykes, F. M. Richards, "The Chemical Shift Index – a Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy", *Biochemistry* **1992**, *31*, 1647–1651, DOI 10.1021/bi00121a010.
- [90] L. Szilagyi, "Chemical Shifts in Proteins Come of Age", *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *27*, 325–443, DOI 10.1016/0079-6565(95)01011-2.
- [91] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, "TALOS+ : A Hybrid Method for Predicting Protein Backbone Torsion Angles from NMR Chemical Shifts", *J. Biomol. NMR* **2009**, *44*, 213–223, DOI 10.1007/s10858-009-9333-z.
- [92] B. Han, Y. F. Liu, S. W. Ginzinger, D. S. Wishart, "SHIFTX2: Significantly Improved Protein Chemical Shift Prediction", *J. Biomol. NMR* **2011**, *50*, 43–57, DOI 10.1007/s10858-011-9478-4.

## References

- [93] G. Barone, L. Gomez-Paloma, D. Duca, A. Silvestri, R. Riccio, G. Bifulco, "Structure Validation of Natural Products by Quantum-Mechanical GIAO Calculations of  $^{13}\text{C}$  NMR Chemical Shifts", *Chem. Eur. J.* **2002**, *8*, 3233–3239, DOI 10.1002/1521-3765(20020715)8:14<3233::Aid-Chem3233>3.0.Co;2-0.
- [94] G. Barone, D. Duca, A. Silvestri, L. Gomez-Paloma, R. Riccio, G. Bifulco, "Determination of the Relative Stereochemistry of Flexible Organic Compounds by Ab Initio Methods: Conformational Analysis and Boltzmann-Averaged GIAO  $^{13}\text{C}$  NMR Chemical Shifts", *Chem. Eur. J.* **2002**, *8*, 3240–3245, DOI 10.1002/1521-3765(20020715)8:14<3240::Aid-Chem3240>3.0.Co;2-G.
- [95] R. Jain, T. Bally, P. R. Rablen, "Calculating Accurate Proton Chemical Shifts of Organic Molecules with Density Functional Methods and Modest Basis Sets", *J. Org. Chem.* **2009**, *74*, 4017–4023, DOI 10.1021/jo900482q.
- [96] S. G. Smith, J. M. Goodman, "Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability", *J. Am. Chem. Soc.* **2010**, *132*, 12946–12959, DOI 10.1021/ja105035r.
- [97] N. Grimblat, M. M. Zanardi, A. M. Sarotti, "Beyond DP4: an Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts", *J. Org. Chem.* **2015**, *80*, 12526–12534, DOI 10.1021/acs.joc.5b02396.
- [98] M. P. Williamson, T. F. Havel, K. Wüthrich, "Solution Conformation of Proteinase Inhibitor IIa From Bull Seminal Plasma by  $^1\text{H}$  Nuclear Magnetic Resonance and Distance Geometry", *J. Mol. Biol.* **1985**, *182*, 295–315, DOI 10.1016/0022-2836(85)90347-x.
- [99] A. Saupe, G. Englert, "High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules", *Phys. Rev. Lett.* **1963**, *11*, 462–464, DOI 10.1103/PhysRevLett.11.462.
- [100] A. Saupe, "The Average Orientation of Solute Molecules in Nematic Liquid Crystals by Proton Magnetic Resonance Measurements and Orientation Dependent Intermolecular Forces", *Mol. Cryst.* **1966**, *1*, 527–540, DOI 10.1080/15421406608083291.
- [101] N. Tjandra, "Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium", *Science* **1997**, *278*, 1111–1114, DOI 10.1126/science.278.5340.1111.
- [102] M. R. Hansen, L. Mueller, A. Pardi, "Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions", *Nat. Struct. Biol.* **1998**, *5*, 1065–1074, DOI 10.1038/4176.
- [103] R. Tycko, F. J. Blanco, Y. Ishii, "Alignment of Biopolymers in Strained Gels: A New Way to Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR", *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341, DOI 10.1021/ja002133q.
- [104] E. Brunner, "Residual Dipolar Couplings in Protein NMR", *Concepts Magn. Reson.* **2001**, *13*, 238–259, DOI 10.1002/cm.1012.
- [105] O. F. Lange, N. A. Lakomek, C. Fares, G. F. Schroder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger, B. L. de Groot, "Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution", *Science* **2008**, *320*, 1471–1475, DOI 10.1126/science.1157092.
- [106] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, "Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins: Information for Structure Determination in Solution.", *Proc. Natl. Acad. Sci.* **1995**, *92*, 9279–9283, DOI 10.1073/pnas.92.20.9279.
- [107] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, "NMR Evidence for Slow Collective Motions in Cyanometmyoglobin", *Nat. Struct. Biol.* **1997**, *4*, 292–297, DOI 10.1038/nsb0497-292.

- [108] C. D. Schwieters, J. Y. Suh, A. Grishaev, R. Ghirlando, Y. Takayama, G. M. Clore, "Solution Structure of the 128 kDa Enzyme I Dimer from *Escherichia coli* and its 146 kDa Complex with HPr Using Residual Dipolar Couplings and Small- and Wide-Angle X-ray Scattering", *J. Am. Chem. Soc.* **2010**, *132*, 13026–13045, DOI 10.1021/ja105485b.
- [109] D. Menche, "New Methods for Stereochemical Determination of Complex Polyketides: Configurational Assignment of Novel Metabolites from *Myxobacteria*", *Nat. Prod. Rep.* **2008**, *25*, 905–918, DOI 10.1039/b707989n.
- [110] J. L. Yan, A. D. Kline, H. P. Mo, M. J. Shapiro, E. R. Zartler, "A Novel Method for the Determination of Stereochemistry in Six-Membered Chairlike Rings Using Residual Dipolar Couplings", *J. Org. Chem.* **2003**, *68*, 1786–1795, DOI 10.1021/jo020670i.
- [111] G. Kummerlöwe, B. Luy, "Residual Dipolar Couplings for the Configurational and Conformational Analysis of Organic Molecules", *Annu. Rep. NMR Spectrosc.* **2009**, *68*, 193–232, DOI 10.1016/S0066-4103(09)06804-5.
- [112] F. Hallwass, M. Schmidt, H. Sun, A. Mazur, G. Kummerlowe, B. Luy, A. Navarro-Vazquez, C. Griesinger, U. M. Reinscheid, "Residual Chemical Shift Anisotropy (RCSA): A Tool for the Analysis of the Configuration of Small Molecules", *Angew. Chem. Int. Ed.* **2011**, *50*, 9487–9490, DOI 10.1002/anie.201101784.
- [113] G. Kummerlöwe, S. L. Grage, C. M. Thiele, I. Kuprov, A. S. Ulrich, B. Luy, "Variable Angle NMR Spectroscopy and Its Application to the Measurement of Residual Chemical Shift Anisotropy", *J. Magn. Reson.* **2011**, *209*, 19–30, DOI 10.1016/j.jmr.2010.11.019.
- [114] N. Nath, M. Schmidt, R. R. Gil, R. T. Williamson, G. E. Martin, A. Navarro-Vazquez, C. Griesinger, Y. Z. Liu, "Determination of Relative Configuration from Residual Chemical Shift Anisotropy", *J. Am. Chem. Soc.* **2016**, *138*, 9548–9556, DOI 10.1021/jacs.6b04082.
- [115] N. Nath, J. C. Fuentes-Monteverde, D. Pech-Puch, J. Rodriguez, C. Jimenez, M. Noll, A. Kreiter, M. Reggelin, A. Navarro-Vazquez, C. Griesinger, "Relative Configuration of Micrograms of Natural Compounds Using Proton Residual Chemical Shift Anisotropy", *Nat. Commun.* **2020**, *11*, 4372, DOI 10.1038/s41467-020-18093-5.
- [116] E. J. d’Auvergne, C. Griesinger, "The Theory of Frame Ordering: Observing Motions in Calmodulin Complexes", *Q. Rev. Biophys.* **2019**, *52*, 1–13, DOI 10.1017/S0033583519000015.
- [117] H. Senn, K. Wüthrich, "Amino-Acid-Sequence, Hem-Iron Coordination Geometry and Functional Properties of Mitochondrial and Bacterial C-Type Cytochromes", *Q. Rev. Biophys.* **1985**, *18*, 111–134, DOI 10.1017/S0033583500005151.
- [118] L. M. Weiner, "Magnetic Resonance Study of the Structure and Functions of Cytochrome-P450", *Crit. Rev. Biochem.* **1986**, *20*, 139–200, DOI 10.3109/10409238609083734.
- [119] L. Lee, B. D. Sykes, "Use of Lanthanide-Induced Nuclear Magnetic-Resonance Shifts for Determination of Protein Structure in Solution - EF Calcium Binding Site of Carp Parvalbumin", *Biochemistry* **1983**, *22*, 4366–4373, DOI 10.1021/bi00288a004.
- [120] R. R. Biekofsky, F. W. Muskett, J. M. Schmidt, S. R. Martin, J. P. Browne, P. M. Bayley, J. Feeney, "NMR Approaches for Monitoring Domain Orientations in Calcium-Binding Proteins in Solution Using Partial Replacement of  $\text{Ca}^{2+}$  by  $\text{Tb}^{3+}$ ", *FEBS Lett.* **1999**, *460*, 519–526, DOI 10.1016/S0014-5793(99)01410-6.
- [121] M. Allegrozzi, I. Bertini, M. B. L. Janik, Y. M. Lee, G. H. Lin, C. Luchinat, "Lanthanide-Induced Pseudocontact Shifts for Solution Structure Refinements of Macromolecules in Shells up to 40 Å from the Metal Ion", *J. Am. Chem. Soc.* **2000**, *122*, 4154–4161, DOI 10.1021/ja993691b.
- [122] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, M. A. Zoroddu, "Experimentally Exploring the Conformational Space Sampled by Domain Reorientation in Calmodulin", *Proc. Natl. Acad. Sci.* **2004**, *101*, 6841–6846, DOI 10.1073/pnas.0308641101.

## References

- [123] L. Russo, M. Maestre-Martínez, S. Wolff, S. Becker, C. Griesinger, “Interdomain Dynamics Explored by Paramagnetic NMR”, *J. Am. Chem. Soc.* **2013**, *135*, 17111–17120, DOI 10.1021/ja408143f.
- [124] I. Bertini, I. Gelis, N. Katsaros, C. Luchinat, A. Provenzani, “Tuning the Affinity for Lanthanides of Calcium Binding Proteins”, *Biochemistry* **2003**, *42*, 8011–8021, DOI 10.1021/bi034494z.
- [125] C. Ma, S. J. Opella, “Lanthanide Ions Bind Specifically to an Added ‘EF-Hand’ and Orient a Membrane Protein in Micelles for Solution NMR Spectroscopy”, *J. Magn. Reson.* **2000**, *146*, 381–384, DOI 10.1006/jmre.2000.2172.
- [126] M. Prudêncio, J. Rohovec, J. A. Peters, E. Tocheva, M. J. Boulanger, M. E. P. Murphy, H. J. Hupkes, W. Kisters, A. Impagliazzo, M. Ubbink, “A Caged Lanthanide Complex as a Paramagnetic Shift Agent for Protein NMR”, *Chem. Eur. J.* **2004**, *10*, 3252–3260, DOI 10.1002/chem.200306019.
- [127] A. Leonov, B. Voigt, F. Rodríguez-Castañeda, P. Sakhaii, C. Griesinger, “Convenient Synthesis of Multifunctional EDTA-Based Chiral Metal Chelates Substituted with an S-Mesylycysteine”, *Chem. Eur. J.* **2005**, *11*, 3342–3348, DOI 10.1002/chem.200400907.
- [128] Z. Y. Wu, M. D. Lee, T. J. Carruthers, M. Szabo, M. L. Dennis, J. D. Swarbrick, B. Graham, G. Otting, “New Lanthanide Tag for the Generation of Pseudocontact Shifts in DNA by Site-Specific Ligation to a Phosphorothioate Group”, *Bioconj. Chem.* **2017**, *28*, 1741–1748, DOI 10.1021/acs.bioconjchem.7b00202.
- [129] S. Täubert, Y. H. Zhang, M. Maestre-Martínez, F. Siepel, E. Woltjen, A. Leonov, C. Griesinger, “Lanthanide Tagging of Oligonucleotides to Nucleobase for Paramagnetic NMR”, *ChemBioChem* **2020**, *21*, 3333–3337, DOI 10.1002/cbic.202000417.
- [130] S. Yamamoto, T. Yamaguchi, M. Erdelyi, C. Griesinger, K. Kato, “Paramagnetic Lanthanide Tagging for NMR Conformational Analyses of N-Linked Oligosaccharides”, *Chem. Eur. J.* **2011**, *17*, 9280–9282, DOI 10.1002/chem.201100856.
- [131] M. Erdelyi, E. d’Auvergne, A. Navarro-Vazquez, A. Leonov, C. Griesinger, “Dynamics of the Glycosidic Bond: Conformational Space of Lactose”, *Chem. Eur. J.* **2011**, *17*, 9368–9376, DOI 10.1002/chem.201100854.
- [132] K. C. Nicolaou, S. A. Snyder, “Chasing Molecules That Were Never There: Misassigned Natural Products and the Role of Chemical Synthesis in Modern Structure Elucidation”, *Angew. Chem. Int. Ed.* **2005**, *44*, 1012–1044, DOI 10.1002/anie.200460864.
- [133] T. L. Suyama, W. H. Gerwick, K. L. McPhail, “Survey of Marine Natural Product Structure Revisions: A Synergy of Spectroscopy and Chemical Synthesis”, *Bioorg. Med. Chem.* **2011**, *19*, 6675–6701, DOI 10.1016/j.bmc.2011.06.011.
- [134] Y. Liu, A. Navarro-Vázquez, R. R. Gil, C. Griesinger, G. E. Martin, R. T. Williamson, “Application of Anisotropic NMR Parameters to the Confirmation of Molecular Structure”, *Nat. Protoc.* **2018**, *14*, 217–247, DOI 10.1038/s41596-018-0091-9.
- [135] G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, “Determination of Relative Configuration in Organic Compounds by NMR Spectroscopy and Computational Methods”, *Chem. Rev.* **2007**, *107*, 3744–3779, DOI 10.1021/cr030733c.
- [136] T. Bruhn, A. Schaumlöffel, Y. Hemberger, G. Bringmann, “SpecDis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra”, *Chirality* **2013**, *25*, 243–249, DOI 10.1002/chir.22138.
- [137] M. Findeisen, T. Brand, S. Berger, “A  $^1\text{H}$ -NMR Thermometer Suitable for Cryoprobes”, *Magn. Reson. Chem.* **2007**, *45*, 175–178, DOI 10.1002/mrc.1941.
- [138] M. H. Levitt, *Spin Dynamics – Basics of Nuclear Magnetic Resonance*, 2nd ed., John Wiley & Sons, New York, **2008**.

- [139] J. Keeler, *Understanding NMR Spectroscopy*, 2nd ed., John Wiley & Sons, New York, **2011**.
- [140] P. W. Atkins, R. S. Friedman, *Molecular Quantum Mechanics*, 5th ed., Oxford University Press, New York, London, **2011**.
- [141] F. Kramer, M. Deshmukh, H. Kessler, S. Glaser, “Residual Dipolar Coupling Constants: An Elementary Derivation of Key Equations”, *Concepts Magn. Reson.* **2004**, *21A*, 10–21, DOI 10.1002/cmr.a.20003.
- [142] G. Otting, “Protein NMR Using Paramagnetic Ions”, *Annu. Rev. Biophys.* **2010**, *39*, 387–405, DOI 10.1146/annurev.biophys.093008.131321.
- [143] I. Bertini, C. Luchinat, G. Parigi, “Magnetic Susceptibility in Paramagnetic NMR”, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249–273, DOI 10.1016/s0079-6565(02)00002-x.
- [144] A. A. Bothner-By in *eMagRes*, John Wiley & Sons, Ltd, **2007**, DOI 10.1002/9780470034590.emrstm0290.
- [145] F. Jensen, *Introduction to Computational Chemistry*, 3rd ed., John Wiley & Sons, New York, **2017**.
- [146] A. KÜchler, *Hochspannungstechnik – Grundlagen – Technologie – Anwendungen*, 4th ed., Springer Science & Business Media, Berlin Heidelberg, **2009**.
- [147] M. G. Larson, F. Bengzon, *The Finite Element Method: Theory, Implementation, and Applications*, Springer Science & Business Media, Berlin Heidelberg, **2013**.
- [148] O. C. Zienkiewicz, R. L. Taylor, J. Zhu, *The Finite Element Method: Its Basis and Fundamentals*, 7th ed., Butterworth-Heinemann, Oxford, **2013**.
- [149] J. Hanson, “Rotations in Three, Four, and Five Dimensions”, *arXiv e-prints* **2011**, arXiv:1103.5263.
- [150] H. Lueken, *Magnetochemie*, Vieweg+Teubner Verlag, **1999**, DOI 10.1007/978-3-322-80118-0.
- [151] T. A. Halgren, “Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94”, *J. Comput. Chem.* **1996**, *17*, 490–519, DOI 10.1002/(sici)1096-987x(199604)17:5/6<490::aid-jcc1>3.0.co;2-p.
- [152] T. A. Halgren, “The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters”, *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843, DOI 10.1021/ja00046a032.
- [153] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision C.01, Wallingford CT, **2010**.
- [154] S. H. Vosko, L. Wilk, M. Nusair, “Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis”, *Can. J. Phys.* **1980**, *58*, 1200–1211, DOI 10.1139/p80-159.
- [155] A. D. Becke, “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior”, *Phys. Rev. A* **1988**, *38*, 3098–3100, DOI 10.1103/physreva.38.3098.

## References

- [156] C. Lee, W. Yang, R. G. Parr, “Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density”, *Phys. Rev. B* **1988**, *37*, 785–789, DOI 10.1103/physrevb.37.785.
- [157] F. Van Petegem, F. C. Chatelain, D. L. Minor, “Insights into Voltage-Gated Calcium Channel Regulation from the Structure of the Ca<sub>v</sub>1.2 IQ Domain-Ca<sup>2+</sup>/Calmodulin Complex”, *Nat. Struct. Mol. Biol.* **2005**, *12*, 1108–1115, DOI 10.1038/nsmb1027.
- [158] R. H. Kretsinger, C. E. Nockolds, “Carp Muscle Calcium-binding Protein: II. Structure Determination and General Description”, *J. Biol. Chem* **1972**, *248*, 3313–3326, DOI 10.1016/S0021-9258(19)44043-X.
- [159] S. Linse, A. Helmersson, S. Forsen, “Calcium Binding to Calmodulin and Its Globular Domains”, *J. Biol. Chem.* **1991**, *266*, 8050–8054, DOI 10.1016/S0021-9258(18)92938-8.
- [160] W. S. VanScyoc, B. R. Sorensen, E. Rusinova, W. R. Laws, J. A. Ross, M. A. Shea, “Calcium Binding to Calmodulin Mutants Monitored by Domain-Specific Intrinsic Phenylalanine and Tyrosine Fluorescence”, *Biophys. J.* **2002**, *83*, 2767–2780, DOI 10.1016/s0006-3495(02)75286-7.
- [161] B. E. Finn, J. Evenas, T. Drakenberg, J. P. Waltho, E. Thulin, S. Forsen, “Calcium-Induced Structural Changes and Domain Autonomy in Calmodulin”, *Nat. Struct. Biol.* **1995**, *2*, 777–783, DOI 10.1038/nsb0995-777.
- [162] A. Crivici, M. Ikura, “Molecular and Structural Basis of Target Recognition by Calmodulin”, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 85–116, DOI 10.1146/annurev.bb.24.060195.000505.
- [163] O. B. Peersen, T. S. Madsen, J. J. Falke, “Intermolecular Tuning of Calmodulin by Target Peptides and Proteins: Differential Effects on Ca<sup>2+</sup> Binding and Implications for Kinase Activation”, *Protein Sci.* **1997**, *6*, 794–807, DOI 10.1002/pro.5560060406.
- [164] C. Sola, S. Barron, J. M. Tusell, J. Serratos, “The Ca<sup>2+</sup>/Calmodulin Signaling System in the Neural Response to Excitability. Involvement of Neuronal and Glial Cells”, *Prog. Neurobiol.* **1999**, *58*, 207–232, DOI 10.1016/S0301-0082(98)00082-3.
- [165] D. Chin, A. R. Means, “Calmodulin: A Prototypical Calcium Sensor”, *Trends Cell. Biol.* **2000**, *10*, 322–328, DOI 10.1016/S0962-8924(00)01800-6.
- [166] E. Carafoli, “Calcium Signaling: A Tale for All Seasons”, *Proc. Natl. Acad. Sci.* **2002**, *99*, 1115–1122, DOI 10.1073/pnas.032427999.
- [167] H. Tidow, P. Nissen, “Structural Diversity of Calmodulin Binding to Its Target Sites”, *FEBS J.* **2013**, *280*, 5551–5565, DOI 10.1111/febs.12296.
- [168] Y. Babu, C. E. Bugg, W. J. Cook, “Structure of Calmodulin Refined at 2.2 Å Resolution”, *J. Mol. Biol.* **1988**, *204*, 191–204, DOI 10.1016/0022-2836(88)90608-0.
- [169] F. Rodríguez-Castañeda, M. Maestre-Martínez, N. Coudevylle, K. Dimova, H. Junge, N. Lipstein, D. Lee, S. Becker, N. Brose, O. Jahn, et al., “Modular Architecture of Munc13/Calmodulin Complexes: Dual Regulation by Ca<sup>2+</sup> and Possible Function in Short-Term Synaptic Plasticity”, *EMBO J.* **2009**, *29*, 680–691, DOI 10.1038/emboj.2009.373.
- [170] Y. S. Babu, J. S. Sack, T. J. Greenhough, C. E. Bugg, A. R. Means, W. J. Cook, “Three-Dimensional Structure of Calmodulin”, *Nature* **1985**, *315*, 37–40, DOI 10.1038/315037a0.
- [171] B. A. Seaton, J. F. Head, D. M. Engelman, F. M. Richards, “Calcium-Induced Increase in the Radius of Gyration and Maximum Dimension of Calmodulin Measured by Small-Angle X-Ray-Scattering”, *Biochemistry* **1985**, *24*, 6740–6743, DOI 10.1021/bi00345a002.
- [172] N. Matsushima, Y. Izumi, T. Matsuo, H. Yoshino, T. Ueki, Y. Miyake, “Binding of Both Ca<sup>2+</sup> and Mastoparan to Calmodulin Induces a Large Change in the Tertiary Structure”, *J. Biochem.* **1989**, *105*, 883–887, DOI 10.1093/oxfordjournals.jbchem.a122773.

- [173] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, A. Bax, “Backbone Dynamics of Calmodulin Studied by  $^{15}\text{N}$  Relaxation Using Inverse Detected 2-Dimensional NMR-Spectroscopy - the Central Helix Is Flexible”, *Biochemistry* **1992**, *31*, 5269–5278, DOI 10.1021/bi00138a005.
- [174] J. L. Baber, A. Szabo, N. Tjandra, “Analysis of Slow Interdomain Motion of Macromolecules Using NMR Relaxation Data”, *J. Am. Chem. Soc.* **2001**, *123*, 3953–3959, DOI 10.1021/ja0041876.
- [175] M. A. Wilson, A. T. Brunger, “The 1.0 Å Crystal Structure of  $\text{Ca}^{2+}$ -Bound Calmodulin: An Analysis of Disorder and Implications for Functionally Relevant Plasticity”, *J. Mol. Biol.* **2000**, *301*, 1237–1256, DOI 10.1006/jmbi.2000.4029.
- [176] J. L. Fallon, F. A. Quioco, “A Closed Compact Structure of Native  $\text{Ca}^{2+}$ -Calmodulin”, *Structure* **2003**, *11*, 1303–1307, DOI 10.1016/j.str.2003.09.004.
- [177] W. Wriggers, E. Mehler, F. Pitici, H. Weinstein, K. Schulten, “Structure and Dynamics of Calmodulin in Solution”, *Biophys. J.* **1998**, *74*, 1622–1639, DOI 10.1016/S0006-3495(98)77876-2.
- [178] N. P. Barton, C. S. Verma, L. S. A. Caves, “Inherent Flexibility of Calmodulin Domains: A Normal-Mode Analysis Study”, *J. Phys. Chem. B* **2002**, *106*, 11036–11040, DOI 10.1021/jp026692q.
- [179] P. M. Bayley, W. A. Findlay, S. R. Martin, “Target Recognition by Calmodulin: Dissecting the Kinetics and Affinity of Interaction Using Short Peptide Sequences”, *Protein Sci.* **1996**, *5*, 1215–1228, DOI 10.1002/pro.5560050701.
- [180] M. Ikura, “Calcium Binding and Conformational Response in EF-Hand Proteins”, *Trends Biochem. Sci.* **1996**, *21*, 14–17, DOI 10.1016/0968-0004(96)80879-6.
- [181] W. E. Meador, A. R. Means, F. A. Quioco, “Target Enzyme Recognition by Calmodulin - 2.4 Å Structure of a Calmodulin-Peptide Complex”, *Science* **1992**, *257*, 1251–1255, DOI 10.1126/science.1519061.
- [182] M. Ikura, G. M. Clore, A. M. Gronenborn, G. Zhu, C. B. Klee, A. Bax, “Solution Structure of a Calmodulin-Target Peptide Complex by Multidimensional NMR”, *Science* **1992**, *256*, 632–638, DOI 10.1126/science.1585175.
- [183] M. Ikura, G. Barbato, C. B. Klee, A. Bax, “Solution Structure of Calmodulin and Its Complex with a Myosin Light Chain Kinase Fragment”, *Cell Calcium* **1992**, *13*, 391–400, DOI 10.1016/0143-4160(92)90052-T.
- [184] B. Elshorst, M. Hennig, H. Forsterling, A. Diener, M. Maurer, P. Schulte, H. Schwalbe, C. Griesinger, J. Krebs, H. Schmid, T. Vorherr, E. Carafoli, “NMR Solution Structure of a Complex of Calmodulin with a Binding Peptide of the  $\text{Ca}^{2+}$  Pump”, *Biochemistry* **1999**, *38*, 12320–12332, DOI 10.1021/bi9908235.
- [185] M. A. Schumacher, A. F. Rivard, H. P. Bachinger, J. P. Adelman, “Structure of the Gating Domain of a  $\text{Ca}^{2+}$ -Activated  $\text{K}^{+}$  Channel Complexed with  $\text{Ca}^{2+}$ /Calmodulin”, *Nature* **2001**, *410*, 1120–1124, DOI 10.1038/35074145.
- [186] M. Osawa, H. Tokumitsu, M. B. Swindells, H. Kurihara, M. Orita, T. Shibamura, T. Furuya, M. Ikura, “A Novel Target Recognition Revealed by Calmodulin in Complex with  $\text{Ca}^{2+}$ -Calmodulin-Dependent Kinase”, *Nat. Struct. Biol.* **1999**, *6*, 819–824, DOI 10.1038/12271.
- [187] I. Augustin, C. Rosenmund, T. C. Sudhof, N. Brose, “Munc13-1 Is Essential for Fusion Competence of Glutamatergic Synaptic Vesicles”, *Nature* **1999**, *400*, 457–461, DOI 10.1038/22768.
- [188] C. Rosenmund, A. Sigler, I. Augustin, K. Reim, N. Brose, J. S. Rhee, “Differential Control of Vesicle Priming and Short-Term Plasticity by Munc13 Isoforms”, *Neuron* **2002**, *33*, 411–424, DOI 10.1016/S0896-6273(02)00568-8.

## References

- [189] F. Varoquaux, A. Sigler, J. S. Rhee, N. Brose, C. Enk, K. Reim, C. Rosenmund, “Total Arrest of Spontaneous and Evoked Synaptic Transmission but Normal Synaptogenesis in the Absence of Munc13-Mediated Vesicle Priming”, *Proc. Natl. Acad. Sci.* **2002**, *99*, 9037–9042, DOI 10.1073/pnas.122623799.
- [190] H. J. Junge, J. S. Rhee, O. Jahn, F. Varoquaux, J. Spiess, M. N. Waxham, C. Rosenmund, N. Brose, “Calmodulin and Munc13 form a  $\text{Ca}^{2+}$  sensor/effector complex that controls short-term synaptic plasticity”, *Cell* **2004**, *118*, 389–401, DOI 10.1016/j.cell.2004.06.029.
- [191] D. Guerini, J. Krebs, E. Carafoli, “Stimulation of the Purified Erythrocyte  $\text{Ca}^{2+}$ -ATPase by Tryptic Fragments of Calmodulin”, *J. Biol. Chem.* **1984**, *259*, 5172–5177, DOI 10.1016/S0021-9258(17)42530-0.
- [192] F. Rodríguez-Castañeda, N. Coudeville, S. Becker, N. Brose, T. Carlomagno, C. Griesinger, “ $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  Resonance Assignments of the Calmodulin/Munc13-1 Peptide Complex”, *Biomol. NMR Assign.* **2009**, *4*, 45–48, DOI 10.1007/s12104-009-9204-2.
- [193] N. Karschin, *Paramagnetic NMR in Proteins*, Master Thesis, University of Göttingen, Germany, **2017**.
- [194] G. Wider, L. Dreier, “Measuring Protein Concentrations by NMR Spectroscopy”, *J. Am. Chem. Soc.* **2006**, *128*, 2571–2576, DOI 10.1021/ja055336t.
- [195] J. Santoro, G. C. King, “A Constant-Time 2D Overbroaden Experiment for Inverse Correlation of Isotopically Enriched Species”, *J. Magn. Reson.* **1992**, *97*, 202–207, DOI 10.1016/0022-2364(92)90250-B.
- [196] S. Grzesiek, A. Bax, “Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly  $^{13}\text{C}/^{15}\text{N}$ -Enriched Proteins”, *J. Biomol. NMR* **1993**, *3*, 185–204, DOI 10.1007/BF00178261.
- [197] T. M. Logan, E. T. Olejniczak, R. X. Xu, S. W. Fesik, “A General Method for Assigning NMR Spectra of Denatured Proteins Using 3D HC(CO)NH-TOCSY Triple Resonance Experiments”, *J. Biomol. NMR* **1993**, *3*, 225–231, DOI 10.1007/BF00178264.
- [198] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, “The Protein Data Bank”, *Nucleic Acids Res.* **2000**, *28*, 235–242, DOI 10.1093/nar/28.1.235.
- [199] M. Ottiger, A. Bax, “Determination of Relative  $\text{N}-\text{H}^{\text{N}}$ ,  $\text{N}-\text{C}'$ ,  $\text{C}^{\alpha}-\text{C}'$ , and  $\text{C}^{\alpha}-\text{H}^{\alpha}$  Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase”, *J. Am. Chem. Soc.* **1998**, *120*, 12334–12341, DOI 10.1021/ja9826791.
- [200] K. Loth, P. Pelupessy, G. Bodenhausen, “Chemical Shift Anisotropy Tensors of Carbonyl, Nitrogen, and Amide Proton Nuclei in Proteins through Cross-Correlated Relaxation in NMR Spectroscopy”, *J. Am. Chem. Soc.* **2005**, *127*, 6062–6068, DOI 10.1021/ja042863o.
- [201] B. Efron, “Bootstrap Methods: Another Look at the Jackknife”, *Ann. Statist.* **1979**, *7*, 1–26, DOI 10.1214/aos/1176344552.
- [202] E. Harder, W. Damm, J. Maple, C. J. Wu, M. Reboul, J. Y. Xiang, L. L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, “OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins”, *J. Chem. Theory Comput.* **2016**, *12*, 281–296, DOI 10.1021/acs.jctc.5b00864.
- [203] 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, P. A. Kollman, “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules”, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197, DOI 10.1021/ja00124a002.
- [204] W. Kabsch, “A Solution for the Best Rotation to Relate Two Sets of Vectors”, *Acta Cryst.* **1976**, *32*, 922–923, DOI 10.1107/s0567739476001873.

- [205] C. L. Lawson, R. J. Hanson, *Solving Least Squares Problems*, Society for Industrial and Applied Mathematics, **1987**, DOI 10.1137/1.9781611971217.
- [206] G. Schwarz, “Estimating the Dimension of a Model”, *Ann. Statist.* **1978**, *6*, 461–464, DOI 10.1214/aos/1176344136.
- [207] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, Amsterdam, Boston, **1981**.
- [208] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions”, *J. R. Stat. Soc. B Stat. Methodol.* **1974**, *36*, 111–147, DOI 10.1111/j.2517-6161.1974.tb00994.x.
- [209] V. Sklenar, M. Piotta, R. Leppik, V. Saudek, “Gradient-Tailored Water Suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC Experiments Optimized to Retain Full Sensitivity”, *J. Magn. Reson. A* **1993**, *102*, 241–245, DOI 10.1006/jmra.1993.1098.
- [210] M. Piotta, V. Saudek, V. Sklenář, “Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions”, *J. Biomol. NMR* **1992**, *2*, 661–665, DOI 10.1007/bf02192855.
- [211] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, “NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes”, *J. Biomol. NMR* **1995**, *6*, DOI 10.1007/bf00197809.
- [212] V. Y. Orekhov, M. Mayzel, V. Jaravine, K. Kazimierczuk, *MDDNMR*, version 2.6, **2005**.
- [213] K. Kazimierczuk, V. Y. Orekhov, “Accelerated NMR Spectroscopy by Using Compressed Sensing”, *Angew. Chem. Int. Ed.* **2011**, *50*, 5556–5559, DOI 10.1002/anie.201100370.
- [214] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, “CcpNMR AnalysisAssign: A Flexible Platform for Integrated NMR Analysis”, *J. Biomol. NMR* **2016**, *66*, 111–124, DOI 10.1007/s10858-016-0060-y.
- [215] Schrödinger Release 2017-4: Macromodel, New York, **2019**.
- [216] G. van Rossum, F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA, **2009**.
- [217] 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, I. 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. N. H. Ribeiro, F. Pedregosa, P. van Mulbregt, S. 1. 0. Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”, *Nat. Methods* **2020**, *17*, 261–272, DOI 10.1038/s41592-019-0686-2.
- [218] J. D. Hunter, “Matplotlib: A 2D Graphics Environment”, *Comput. Sci. Eng.* **2007**, *9*, 90–95, DOI 10.1109/Mcse.2007.55.
- [219] N. Karschin, K. Wolkenstein, C. Griesinger, “Magnetically Induced Alignment of Natural Products for Stereochemical Structure Determination via NMR”, *Angew. Chem. Int. Ed.* **2020**, *59*, 15860–15864, DOI 10.1002/anie.202004881.
- [220] J. H. Prestegard, A. I. Kishore, “Partial Alignment of Biomolecules: An Aid to NMR Characterization”, *Curr. Opin. Chem. Biol.* **2001**, *5*, 584–590, DOI 10.1016/s1367-5931(00)00247-7.
- [221] A. Marx, C. Thiele, “Orientational Properties of Poly- $\gamma$ -benzyl-L-glutamate: Influence of Molecular Weight and Solvent on Order Parameters of the Solute”, *Chem. Eur. J.* **2009**, *15*, 254–260, DOI 10.1002/chem.200801147.
- [222] B. Luy, K. Kobzar, H. Kessler, “An Easy and Scalable Method for the Partial Alignment of Organic Molecules for Measuring Residual Dipolar Couplings”, *Angew. Chem. Int. Ed.* **2004**, *43*, 1092–1094, DOI 10.1002/anie.200352860.

## References

- [223] A. A. Bothner-By in *Encyclopedia of Nuclear Magnetic Resonance*, (Eds.: D. M. Grant, R. K. Harris), Wiley, Chichester, **1995**, pp. 2932–2938.
- [224] F. Peters, M. Maestre-Martínez, A. Leonov, L. Kovačič, S. Becker, R. Boelens, C. Griesinger, “Cys-Ph-TAHA: A Lanthanide Binding Tag for RDC And PCS Enhanced Protein NMR”, *J. Biomol. NMR* **2011**, *51*, 329–337, DOI 10.1007/s10858-011-9560-y.
- [225] X.-C. Su, K. McAndrew, T. Huber, G. Otting, “Lanthanide-Binding Peptides for NMR Measurements of Residual Dipolar Couplings and Paramagnetic Effects from Multiple Angles”, *J. Am. Chem. Soc.* **2008**, *130*, 1681–1687, DOI 10.1021/ja0765641.
- [226] N. Tjandra, S. Grzesiek, A. Bax, “Magnetic Field Dependence of Nitrogen-Proton  $J$ -Splittings in  $^{15}\text{N}$ -Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling”, *J. Am. Chem. Soc.* **1996**, *118*, 6264–6272, DOI 10.1021/ja960106n.
- [227] H. C. Kung, K. Y. Wang, I. Goljer, P. H. Bolton, “Magnetic Alignment of Duplex and Quadruplex DNAs”, *J. Magn. Reson. B* **1995**, *109*, 323–325, DOI 10.1006/jmrb.1995.9987.
- [228] M. A. Lisicki, P. K. Mishra, A. A. Bothner-By, J. S. Lindsey, “Solution Conformation of a Porphyrin-Quinone Cage Molecule Determined by Dipolar Magnetic Field Effects in Ultra-High-Field NMR”, *J. Org. Chem.* **1988**, *92*, 3400–3403, DOI 10.1021/j100323a018.
- [229] C. Gayathri, A. Bothner-By, P. V. Zijl, C. Maclean, “Dipolar Magnetic Field Effects in NMR Spectra of Liquids”, *Chem. Phys. Lett.* **1982**, *87*, 192–196, DOI 10.1016/0009-2614(82)83585-9.
- [230] F. A. L. Anet, “Magnetic Alignment Effects in the 500 MHz Proton NMR Spectrum of *o*-Dichlorobenzene in Acetone- $d_6$ ”, *J. Am. Chem. Soc.* **1986**, *108*, 1354–1355, DOI 10.1021/ja00266a069.
- [231] K. Wolkenstein, J. C. Fuentes-Monteverde, N. Nath, T. Oji, C. Griesinger, “Hyalocrinins, Taurine-Conjugated Anthraquinone and Biaryl Pigments from the Deep Sea Crinoid *Hyalocrinus naresianus*”, *J. Nat. Prod.* **2018**, *82*, 163–167, DOI 10.1021/acs.jnatprod.8b00803.
- [232] S. P. B. Vemulapalli, J. C. Fuentes-Monteverde, N. Karschin, T. Oji, C. Griesinger, K. Wolkenstein, “Structure and Absolute Configuration of Phenanthro-*perylene* Quinone Pigments from the Deep-Sea Crinoid *Hyalocrinus naresianus*”, *Mar. Drugs* **2021**, *19*, 445, DOI 10.3390/md19080445.
- [233] F. D. Riccardis, M. Iorizzi, L. Minale, R. Riccio, B. R. de Forges, C. Debitus, “The Gymnochromes: Novel Marine Brominated Phenanthroperylenequinone Pigments from the Stalked Crinoid *Gymnocrinus Richeri*”, *J. Org. Chem.* **1991**, *56*, 6781–6787, DOI 10.1021/jo00024a016.
- [234] H. V. K. Wangun, A. Wood, C. Fiorilla, J. K. Reed, P. J. McCarthy, A. E. Wright, “Gymnochromes E and F, Cytotoxic Phenanthroperylenequinones from a Deep-Water Crinoid, *Holopus rangii*”, *J. Nat. Prod.* **2010**, *73*, 712–715, DOI 10.1021/np900526y.
- [235] C. A. Mulrooney, E. M. O’Brien, B. J. Morgan, M. C. Kozłowski, “Perylenequinones: Isolation, Synthesis, and Biological Activity”, *Eur. J. Org. Chem.* **2012**, *2012*, 3887–3904, DOI 10.1002/ejoc.201200184.
- [236] A. Bax, R. Freeman, S. P. Kempell, “Natural Abundance  $^{13}\text{C}$ - $^{13}\text{C}$  Coupling Observed via Double-Quantum Coherence”, *J. Am. Chem. Soc.* **1980**, *102*, 4849–4851, DOI 10.1021/ja00534a056.
- [237] R. Altmann, C. Etlstorfer, H. Falk, “Chiroptical Properties and Absolute Configurations of the Hypericin Chromophore Propeller Enantiomers”, *Monatsh. Chem. Chem. Mon.* **1997**, *128*, 785–793, DOI 10.1007/bf00807089.
- [238] A. Bax, D. G. Davis, “Practical Aspects of Two-Dimensional Transverse NOE Spectroscopy”, *J. Magn. Reson. (1969)* **1985**, *63*, 207–213, DOI 10.1016/0022-2364(85)90171-4.

- [239] W. Koźmiński, D. Nanz, “Sensitivity Improvement and New Acquisition Scheme of Heteronuclear Active-Coupling-Pattern-Tilting Spectroscopy”, *J. Magn. Reson.* **2000**, *142*, 294–299, DOI 10.1006/jmre.1999.1939.
- [240] A. Meissner, O. W. Sörensen, “Measurement of  $J_{\text{HH}}$  and Long-Range  $J_{\text{XH}}$  Coupling Constants in Small Molecules. Broadband XLOC and  $J$ -HMBC”, *Magn. Reson. Chem.* **2001**, *39*, 49–52, DOI 10.1002/1097-458x(200101)39:1<49::aid-mrc798>3.0.co;2-s.
- [241] G. Pescitelli, T. Bruhn, “Good Computational Practice in the Assignment of Absolute Configurations by TDDFT Calculations of ECD Spectra”, *Chirality* **2016**, *28*, 466–474, DOI 10.1002/chir.22600.
- [242] C. Adamo, V. Barone, “Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model”, *J. Chem. Phys.* **1999**, *110*, 6158–6170, DOI 10.1063/1.478522.
- [243] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”, *J. Phys. Chem.* **1994**, *98*, 11623–11627, DOI 10.1021/j100096a001.
- [244] A. D. Becke, “Density-Functional Thermochemistry. III. The Role of Exact Exchange”, *J. Chem. Phys.* **1993**, *98*, 5648–5652, DOI 10.1063/1.464913.
- [245] T. Yanai, D. P. Tew, N. C. Handy, “A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP)”, *Chem. Phys. Lett.* **2004**, *393*, 51–57, DOI 10.1016/j.cpllett.2004.06.011.
- [246] T. Bruhn, A. Schaumlöffel, Y. Hemberger, G. Pescitelli, *SpecDis*, version 1.17.1, Berlin, **2017**.
- [247] Y. Xia, S. Moran, E. P. Nikonowicz, X. Gao, “Z-Restored Spin-Echo  $^{13}\text{C}$  1D Spectrum of Straight Baseline Free of Hump, Dip and Roll”, *Magn. Reson. Chem.* **2008**, *46*, 432–435, DOI 10.1002/mrc.2195.
- [248] A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler, B. Luy, “The CLIP/CLAP-HSQC: Pure Absorptive Spectra for the Measurement of One-Bond Couplings”, *J. Magn. Reson.* **2008**, *192*, 314–322, DOI 10.1016/j.jmr.2008.03.009.
- [249] L. Castañar, E. Sistaré, A. Virgili, R. T. Williamson, T. Parella, “Suppression of Phase and Amplitude  $J_{\text{HH}}$  Modulations in HSQC Experiments”, *Magn. Reson. Chem.* **2014**, *53*, 115–119, DOI 10.1002/mrc.4149.
- [250] G. Bifulco, R. Riccio, G. E. Martin, A. V. Buevich, R. T. Williamson, “Quantum Chemical Calculations of  $^1J_{\text{CC}}$  Coupling Constants for the Stereochemical Determination of Organic Compounds”, *Org. Lett.* **2013**, *15*, 654–657, DOI 10.1021/o13034803.
- [251] G. Cornilescu, A. Bax, “Measurement of Proton, Nitrogen, and Carbonyl Chemical Shielding Anisotropies in a Protein Dissolved in a Dilute Liquid Crystalline Phase”, *J. Am. Chem. Soc.* **2000**, *122*, 10143–10154, DOI 10.1021/ja0016194.
- [252] Schrödinger Release 2017-4: Maestro, New York, **2019**.
- [253] G. Chang, W. C. Guida, W. C. Still, “An Internal-Coordinate Monte Carlo Method for Searching Conformational Space”, *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386, DOI 10.1021/ja00194a035.
- [254] W. J. Hehre, R. Ditchfield, J. A. Pople, “Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules”, *J. Chem. Phys.* **1972**, *56*, 2257–2261, DOI 10.1063/1.1677527.
- [255] C. Adamo, V. Barone, “Exchange Functionals with Improved Long-Range Behavior and Adiabatic Connection Methods without Adjustable Parameters: The mPW and mPW1PW Models”, *J. Chem. Phys.* **1998**, *108*, 664–675, DOI 10.1063/1.475428.

## References

- [256] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, “Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation”, *Phys. Rev. B* **1992**, *46*, 6671–6687, DOI 10.1103/physrevb.46.6671.
- [257] R. Ditchfield, “Self-Consistent Perturbation Theory of Diamagnetism”, *Mol. Phys.* **1974**, *27*, 789–807, DOI 10.1080/00268977400100711.
- [258] J. Tomasi, B. Mennucci, R. Cammi, “Quantum Mechanical Continuum Solvation Models”, *Chem. Rev.* **2005**, *105*, 2999–3094, DOI 10.1021/cr9904009.
- [259] F. Jensen, “Unifying General and Segmented Contracted Basis Sets. Segmented Polarization Consistent Basis Sets”, *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085, DOI 10.1021/ct401026a.
- [260] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, T. L. Windus, “New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community”, *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820, DOI 10.1021/acs.jcim.9b00725.
- [261] F. Jensen, “Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding”, *J. Chem. Theory Comput.* **2015**, *11*, 132–138, DOI 10.1021/ct5009526.
- [262] F. Weigend, R. Ahlrichs, “Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy”, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305, DOI 10.1039/b508541a.
- [263] T. Plantenga, B. Ruessink, C. MacLean, “<sup>13</sup>C NMR of Molecules Partially Aligned by an Electric Field: A New Method for Determining the Orientation of the Dipole Moment”, *Chem. Phys.* **1980**, *48*, 359–368, DOI 10.1016/0301-0104(80)80066-8.
- [264] T. Plantenga, P. V. Zijl, C. Maclean, “Studies of Quadrupolar and Dipolar Electric Field Effects in the NMR Spectra of Binary Mixtures of Liquids”, *Chem. Phys.* **1982**, *66*, 1–9, DOI 10.1016/0301-0104(82)88001-4.
- [265] A. Peshkovsky, A. E. McDermott, “NMR Spectroscopy in the Presence of Strong AC Electric Fields: Degree of Alignment of Polar Molecules”, *J. Phys. Chem. A* **1999**, *103*, 8604–8611, DOI 10.1021/jp991993u.
- [266] A. Peshkovsky, A. E. McDermott, “Dipolar Interactions in Molecules Aligned by Strong AC Electric Fields”, *J. Magn. Reson.* **2000**, *147*, 104–109, DOI 10.1006/jmre.2000.2167.
- [267] L. Onsager, “Electric Moments of Molecules in Liquids”, *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493, DOI 10.1021/ja01299a050.
- [268] Schott Borofloat 33 – Technical Brochure, <https://web.archive.org/web/20170409202239/http://www.vdg-ev.org/technik/kataloge/schott/borofloat.pdf> (visited on 2021-10-18).
- [269] L. C. Robosky, M. D. Reily, D. Avizonis, “Improving NMR Sensitivity by Use of Salt-Tolerant Cryogenically Cooled Probes”, *Anal. Bioanal. Chem.* **2007**, *387*, 529–532, DOI 10.1007/s00216-006-0982-4.
- [270] J. Kerr, “XL. A New Relation between Electricity and Light: Dielectrified Media Birefringent”, *Philos. Mag.* **1875**, *50*, 337–348, DOI 10.1080/14786447508641302.
- [271] P. P. Ho, R. R. Alfano, “Optical Kerr Effect in Liquids”, *Phys. Rev. A* **1979**, *20*, 2170–2187, DOI 10.1103/PhysRevA.20.2170.
- [272] Voltagezone Electronics e.U. – online shop, [https://highvoltageshop.com/epages/b73088c0-9f9a-4230-9ffc-4fd5c619abc4.sf/de\\_DE/?ObjectPath=/Shops/b73088c0-9f9a-4230-9ffc-4fd5c619abc4/Products/TRANSHF\\_15KVAC](https://highvoltageshop.com/epages/b73088c0-9f9a-4230-9ffc-4fd5c619abc4.sf/de_DE/?ObjectPath=/Shops/b73088c0-9f9a-4230-9ffc-4fd5c619abc4/Products/TRANSHF_15KVAC) (visited on 2021-10-18).

- [273] DuPont Kapton Summary of Properties – Technical Brochure, <https://www.dupont.com/content/dam/dupont/amer/us/en/products/ei-transformation/documents/EI-10142-Kapton-Summary-of-Properties.pdf> (visited on 2021-10-18).
- [274] Technisches Datenblatt - Borosilicatglas 3.3, [https://www.hilgenberg-gmbh.de/uploads/tx\\_ttproducts/datasheet/Technisches\\_Datenblatt\\_0500\\_d27bbc.pdf](https://www.hilgenberg-gmbh.de/uploads/tx_ttproducts/datasheet/Technisches_Datenblatt_0500_d27bbc.pdf) (visited on 2021-10-18).
- [275] Dalton Research Group home page, University of Washington, [https://depts.washington.edu/eoopic/linkfiles/dielectric\\_chart%5B1%5D.pdf](https://depts.washington.edu/eoopic/linkfiles/dielectric_chart%5B1%5D.pdf) (visited on 2021-10-18).
- [276] <http://www.stenutz.eu/chem/solv6.php?name=perfluorodecalin> (visited on 2021-10-18).
- [277] C. Geuzaine, J. F. Remacle, “Gmsh: A 3-D Finite Element Mesh Generator with Built-In Pre- And Post-Processing Facilities”, *Int. J. Numer. Meth. Eng.* **2009**, *79*, 1309–1331, DOI 10.1002/nme.2579.
- [278] *Elmer FEM – Open Dource Multiphysics Simulation Software*, version 8.4, **2018**.
- [279] J. Ahrens, B. Geveci, C. Law in *Visualization Handbook*, (Eds.: C. D. Hansen, C. R. Johnson), Academic Press, Amsterdam, Boston, **2005**.
- [280] N. Karschin, S. Krenek, D. Heyer, C. Griesinger, “Extension and Improvement of the Methanol- $d_4$  NMR Thermometer Calibration”, *Magn. Reson. Chem.* **2022**, *60*, 203–209, DOI 10.1002/mrc.5216.
- [281] K. K. Frederick, M. S. Marlow, K. G. Valentine, A. J. Wand, “Conformational Entropy in Molecular Recognition by Proteins”, *Nature* **2007**, *448*, 325–329, DOI 10.1038/nature05959.
- [282] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, D. Kern, “Linkage between Dynamics and Catalysis in a Thermophilic-Mesophilic Enzyme Pair”, *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–949, DOI 10.1038/nsmb821.
- [283] N. J. Baxter, M. P. Williamson, “Temperature Dependence of  $^1\text{H}$  Chemical Shifts in Proteins”, *J. Biomol. NMR* **1997**, *9*, 359–369, DOI 10.1023/A:1018334207887.
- [284] Variable Temperature Control for NMR Probes User Manual, version 002, Bruker Biospin AG, Fällanden, Switzerland, **2012**.
- [285] A. L. Van Geet, “Calibration of the Methanol and Glycol Nuclear Magnetic Resonance Thermometers with a Static Thermistor Probe”, *Anal. Chem.* **1968**, *40*, 2227–2229, DOI 10.1021/ac50158a064.
- [286] O. Yamamoto, Yanagisa.M, “A Method for Calibration of Nuclear Magnetic Resonance Standard Samples for Measuring Temperature”, *Anal. Chem.* **1970**, *42*, 1463–1465, DOI 10.1021/ac60294a019.
- [287] B. L. Brandt, L. G. Rubin, H. H. Sample, “Low-Temperature Thermometry in High Magnetic Fields. 6. Industrial-Grade Pt Resistors above 66 K – Rh-Fe and Au-Mn Resistors above 40 K”, *Rev. Sci. Instrum.* **1988**, *59*, 642–645, DOI 10.1063/1.1139849.
- [288] K. Nara, H. Kato, M. Okaji, “Magnetoresistance of a Highly Stable Industrial-Grade Platinum Resistance Thermometer between 20 K and 240 K”, *Cryogenics* **1991**, *31*, 16–20, DOI 10.1016/0011-2275(91)90185-Y.
- [289] T. M. Barbosa, R. Rittner, C. F. Tormena, G. A. Morris, M. Nilsson, “Convection in Liquid-State NMR: Expect the Unexpected”, *RSC Adv.* **2016**, *6*, 95173–95176, DOI 10.1039/c6ra23427e.
- [290] N. Hedin, I. Furó, “Temperature Imaging by  $^1\text{H}$  NMR and Suppression of Convection in NMR Probes”, *J. Magn. Reson.* **1998**, *131*, 126–130, DOI 10.1006/jmre.1997.1352.
- [291] I. Swan, M. Reid, P. W. Howe, M. A. Connell, M. Nilsson, M. A. Moore, G. A. Morris, “Sample Convection in Liquid-State NMR: Why It Is Always with Us, and What We Can Do about It”, *J. Magn. Reson.* **2015**, *252*, 120–129, DOI 10.1016/j.jmr.2014.12.006.

## References

- [292] K. Nara, H. Kato, M. Okaji, “Design of Platinum Resistance Thermometer with Small Magnetic-Field Correction”, *Cryogenics* **1994**, *34*, 1007–1010, DOI 10.1016/0011-2275(94)90094-9.
- [293] Evaluation of Measurement Data - Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement” - Propagation of Distributions Using a Monte Carlo Method. JCGM 101:2008, Joint Committee for Guides in Metrology.
- [294] H. Geen, R. Freeman, “Band-Selective Radiofrequency Pulses”, *J. Magn. Reson. (1969)* **1991**, *93*, 93–141, DOI 10.1016/0022-2364(91)90034-Q.
- [295] S. Rudtsch, C. von Rohden, “Calibration and Self-Validation of Thermistors for High-Precision Temperature Measurements”, *Measurement* **2015**, *76*, 1–6, DOI 10.1016/j.measurement.2015.07.028.
- [296] 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 Rio, M. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, “Array Programming with NumPy”, *Nature* **2020**, *585*, 357–362, DOI 10.1038/s41586-020-2649-2.
- [297] L. Emsley, G. Bodenhausen, “Gaussian Pulse Cascades: New Analytical Functions for Rectangular Selective Inversion and In-Phase Excitation in NMR”, *Chem. Phys. Lett.* **1990**, *165*, 469–476, DOI 10.1016/0009-2614(90)87025-m.
- [298] L. Emsley, G. Bodenhausen, “Optimization of Shaped Selective Pulses for NMR Using a Quaternion Description of Their Overall Propagators”, *J. Magn. Reson.* **1992**, *97*, 135–148, DOI 10.1016/0022-2364(92)90242-y.

# A Appendix

## A.1 Data Collection

Most numerical and other digital data are too bulky to be printed, and are more conveniently provided as compressed archive of data files. Please click [here](#) to open the archive, or access it via the list of attachments of your pdf viewer.

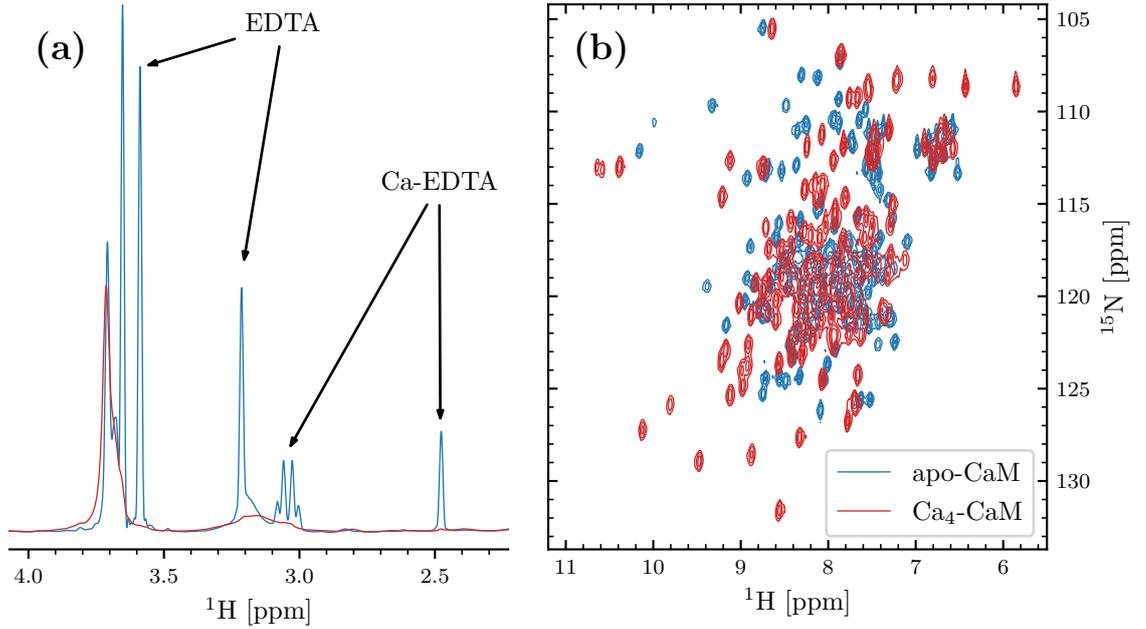
## A.2 Calmodulin/Munc13-1 – Additional Data

**Table A.1.** *Q*-factors for the fits to static structural models of CaM/Munc13-1.

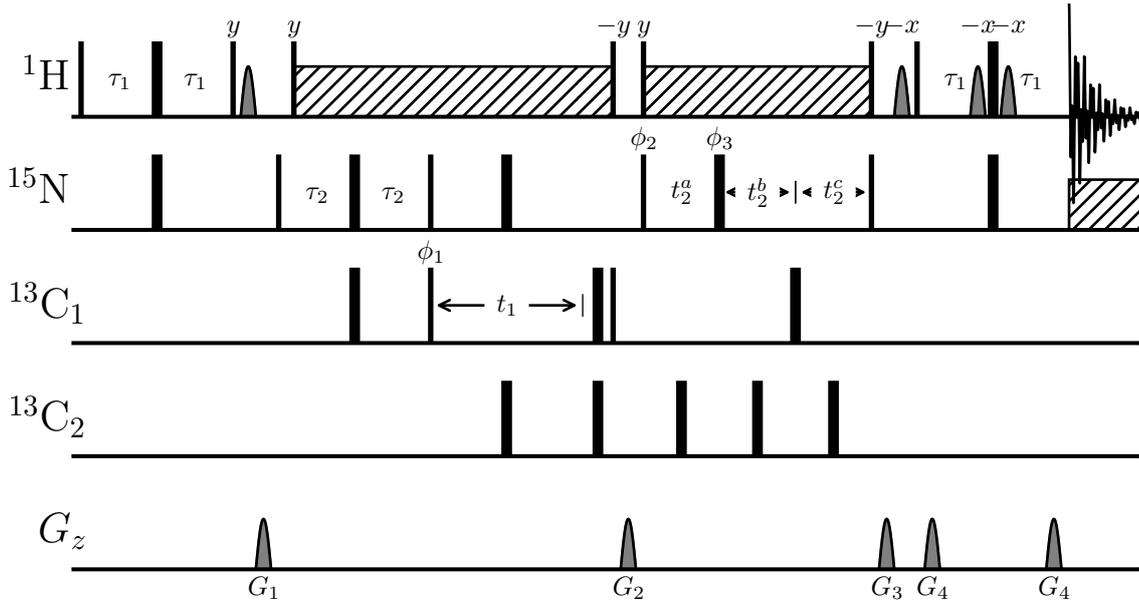
	N-terminal		C-terminal	
	PCS/RDC	PCS	RDC	RDC
Dy	0.036	0.027	0.125	0.516
Er	0.052	0.042	0.353	0.805
Ho	0.049	0.042	0.307	0.749
Tb	0.048	0.039	0.220	0.555
Tm	0.047	0.036	0.381	0.613
Yb	0.043	0.038	0.453	0.933
all	0.046	0.038	0.392	0.625

**Table A.2.** The five independent components of  $\Delta\chi$ , given in  $10^{-32} \text{ m}^3$ , obtained from fitting both RDCs and PCSs to the N-terminal domain, and from fitting RDCs to the C-terminal domain, for the six lanthanides. The uncertainty from bootstrap analysis is given in parentheses.

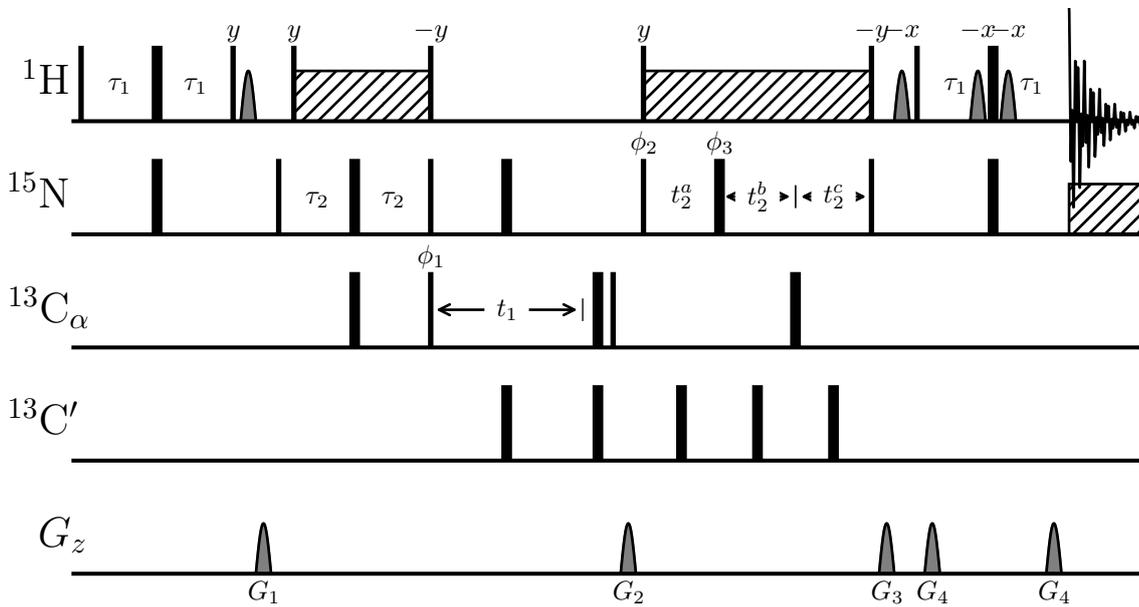
		$\Delta\chi_{ax}$	$\Delta\chi_{rh}$	$\Delta\chi_{xy}$	$\Delta\chi_{xz}$	$\Delta\chi_{yz}$
Dy	N	14.90(24)	-24.54(95)	10.43(23)	10.08(26)	-5.66(19)
	C	-0.14(28)	-1.74(42)	-1.95(24)	3.16(29)	-0.49(24)
Er	N	-4.87(11)	-0.27(13)	-4.76(8)	-5.81(8)	1.64(12)
	C	0.15(16)	1.46(38)	0.05(14)	-1.00(17)	0.55(17)
Ho	N	6.83(11)	-4.43(13)	3.80(9)	4.71(7)	-4.55(10)
	C	-0.07(23)	-1.42(40)	-0.69(17)	1.46(24)	-0.86(17)
Tb	N	17.58(15)	2.02(22)	6.83(14)	12.79(18)	-7.43(24)
	C	0.24(30)	-2.80(53)	-0.09(23)	2.91(32)	-2.13(23)
Tm	N	-8.86(14)	-6.17(14)	-8.03(9)	-8.68(6)	6.89(10)
	C	0.73(25)	3.27(49)	-0.21(22)	-1.29(17)	1.69(18)
Yb	N	-2.04(5)	2.94(4)	-2.96(4)	-2.86(2)	1.91(3)
	C	0.09(17)	0.60(28)	0.27(14)	-0.47(14)	0.20(16)



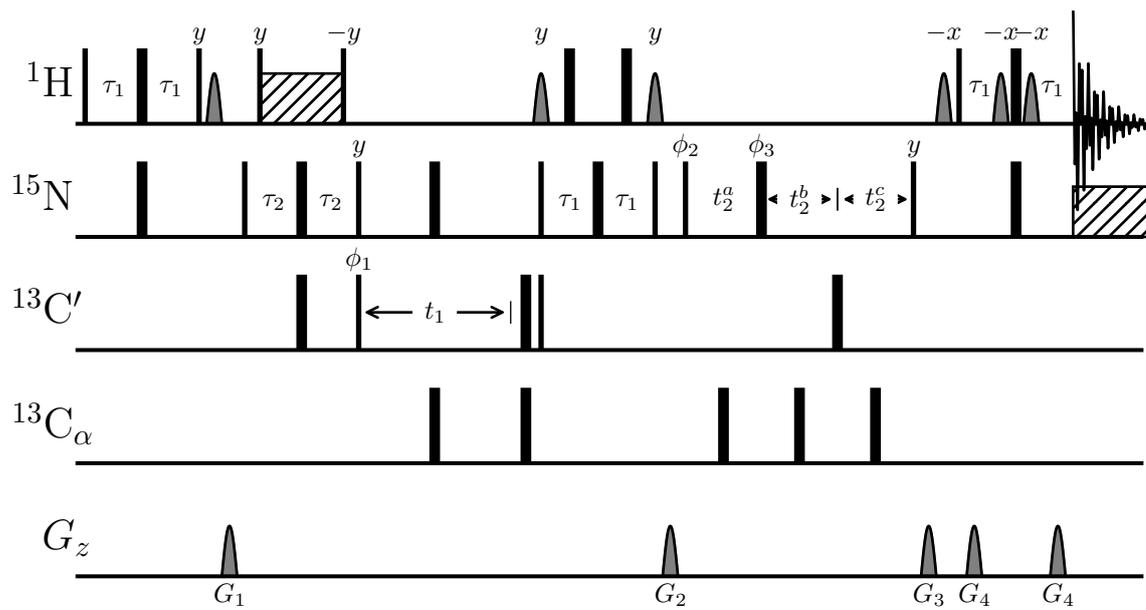
**Figure A.1.**  $^1\text{H}$  (a) and  $^{15}\text{N}$ -HSQC (b) spectra of apo-CaM and Ca<sub>4</sub>-CaM. Both the Ca-loading and the presence of EDTA can be clearly discerned via NMR.



**Figure A.2.** HNCA/HNCO pulse sequence. Cycled phases are  $\phi_1 = (x, -x)$ ,  $\phi_2 = (x)_2(-x)_2$ ,  $\phi_3 = (x)_4(-x)_4$ , and  $\phi_{\text{rec}} = (x, -x, -x, x)_2$ . The INEPT delays were set to  $\tau_1 = 2.3$  ms and  $\tau_2 = 12.4$  ms. The semi-constant time delays were  $t_2^a = \tau_2 - t_2\tau_2/t_{2,\text{max}}$ ,  $t_2^b = t_2(1/2 - \tau_2/t_{2,\text{max}})$ , and  $t_2^c = \tau_2 + t_2/2$ . For selective  $C_\alpha$  and  $C'$  excitation Gauss-cascade type shaped pulses are used,<sup>[297,298]</sup> and water suppression is achieved via a watergate block in the reverse-INEPT period.<sup>[210]</sup> For the HNCA, it is  $C_1 = C_\alpha$  and  $C_2 = C'$ , and for the HNCO, it is  $C_1 = C'$  and  $C_2 = C_\alpha$ . The HNCO{no  $C_\alpha$ } was acquired by omitting all pulses on  $C_\alpha$ .



**Figure A.3.** HNCA{no H} pulse sequence. All annotated parameters were identical to the regular HNCA (Fig. A.2).



**Figure A.4.** HNC{no H} pulse sequence. All annotated parameters were identical to the regular HNC (Fig. A.2).

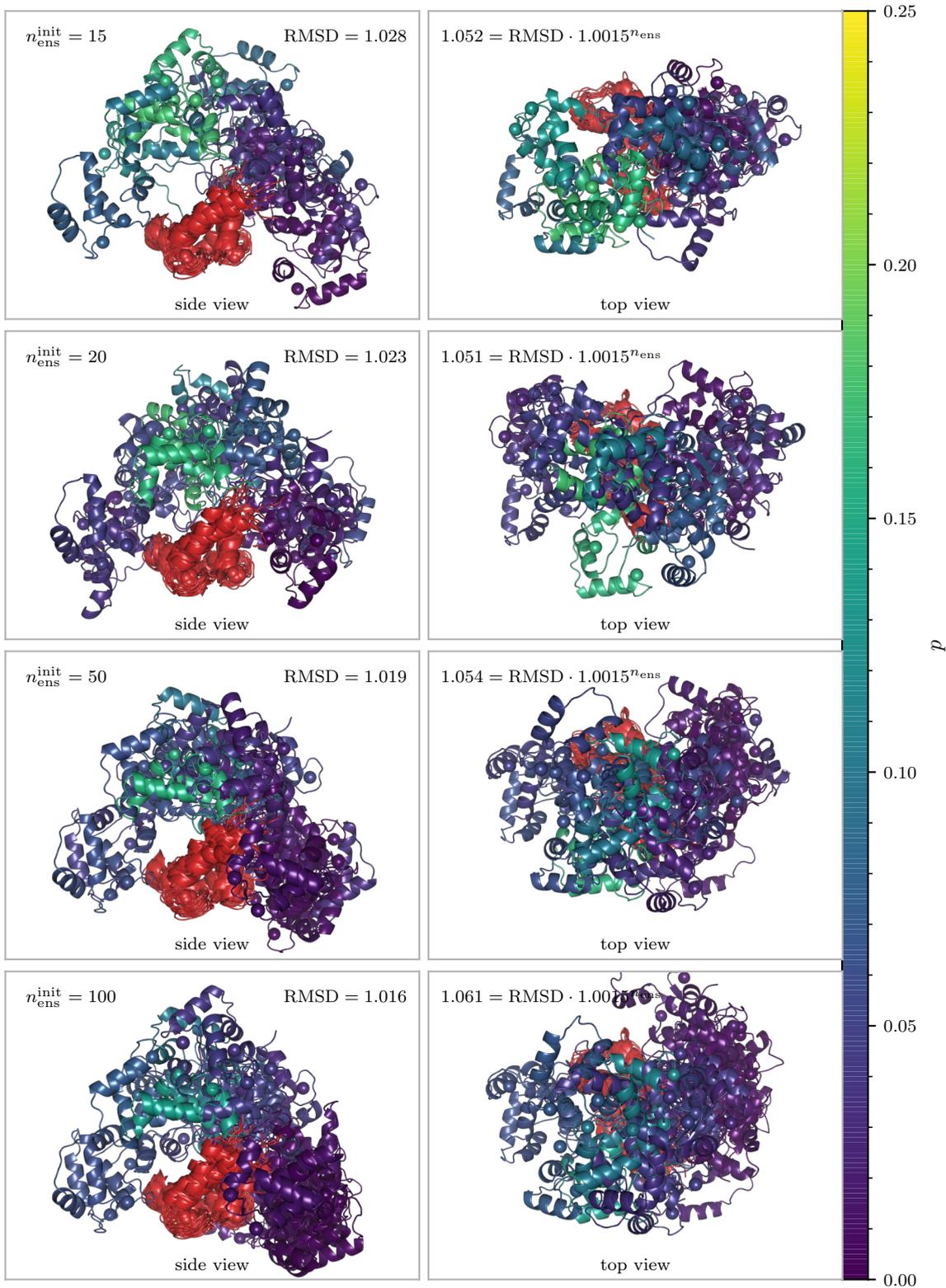


Figure A.5. Cartoon representations of the second best ensembles by BIC, for each starting size.

A Appendix

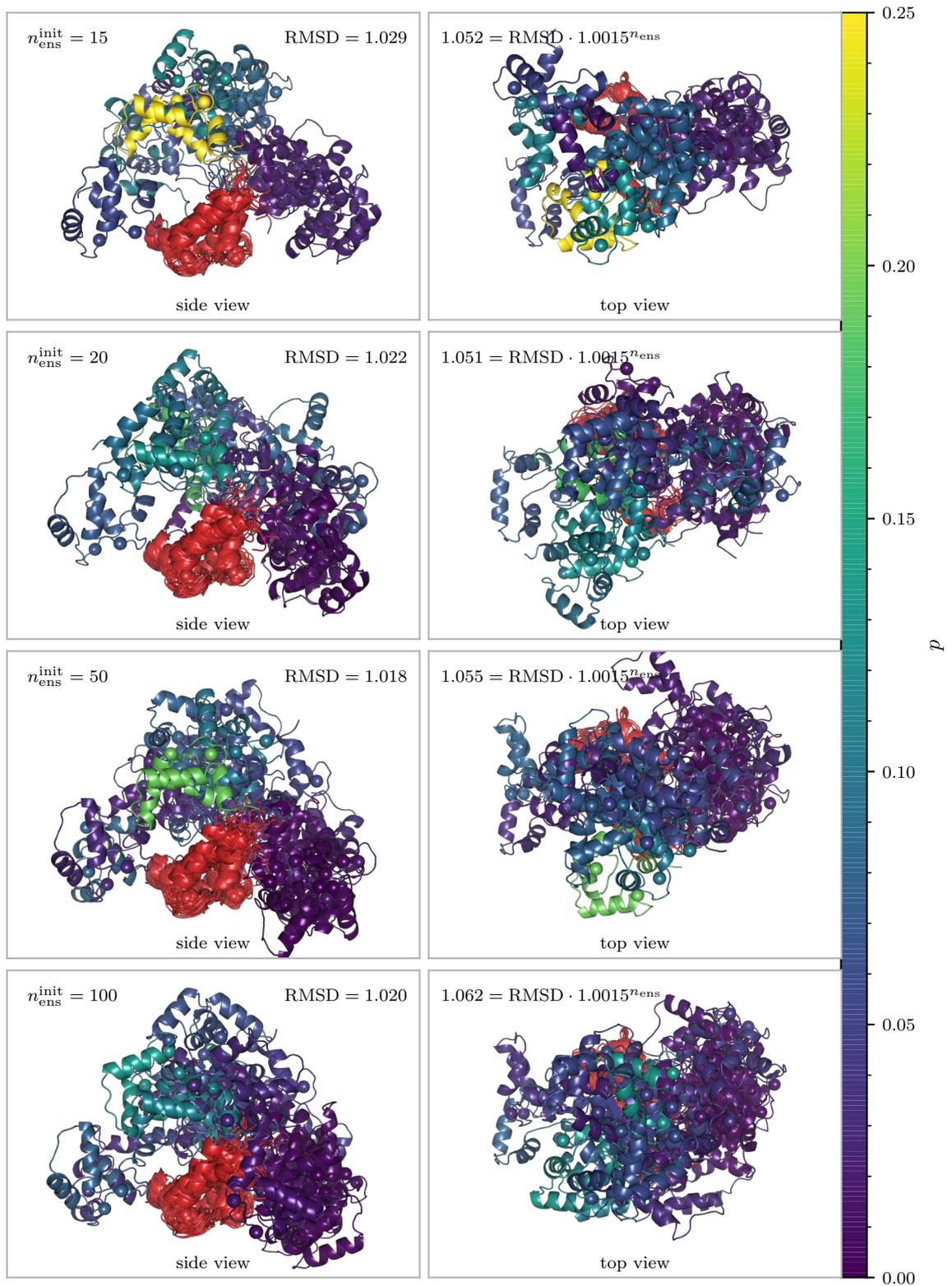


Figure A.6. Cartoon representations of the third best ensembles by BIC, for each starting size.

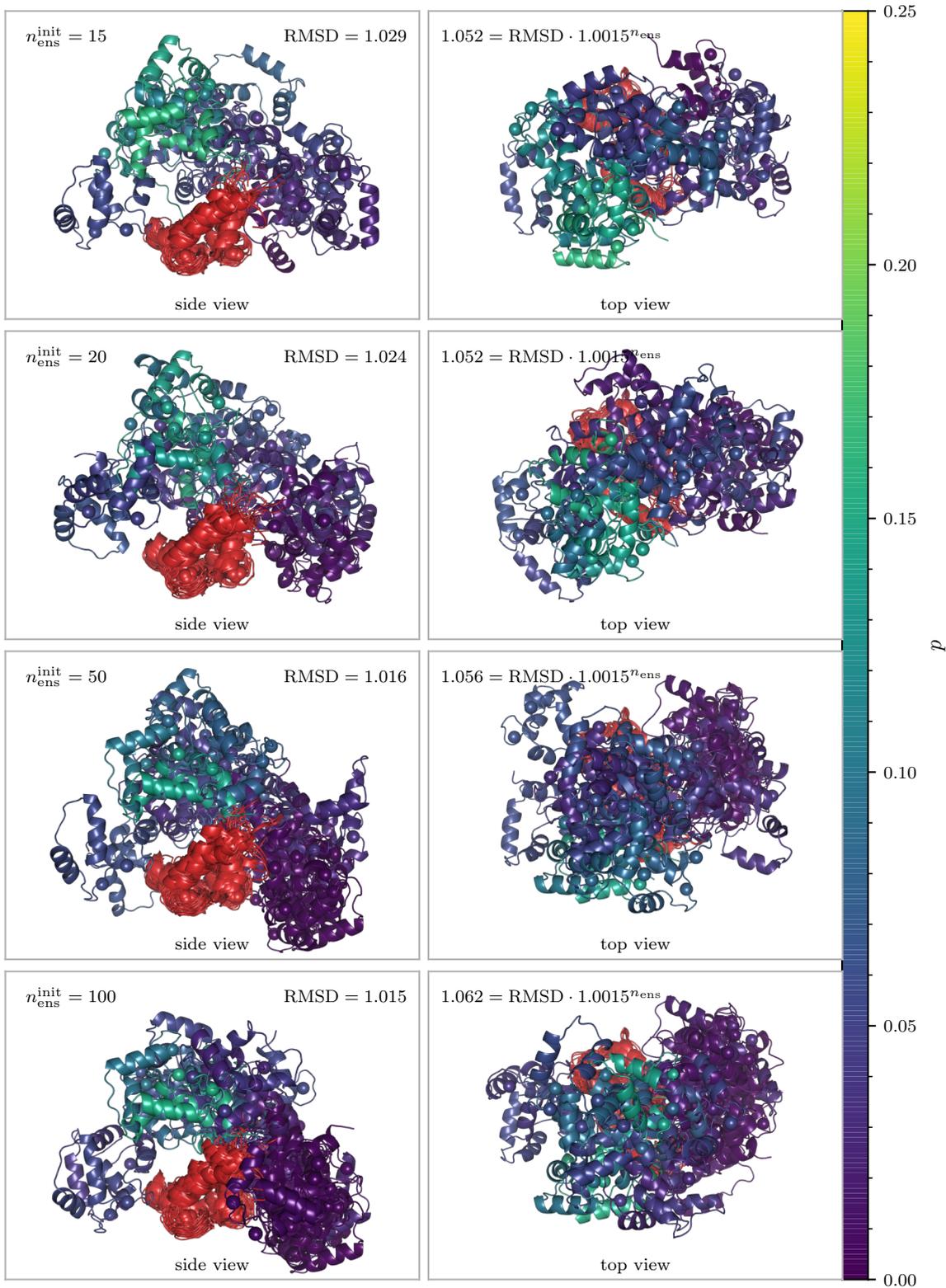


Figure A.7. Cartoon representations of the fourth best ensembles by BIC, for each starting size.

A Appendix

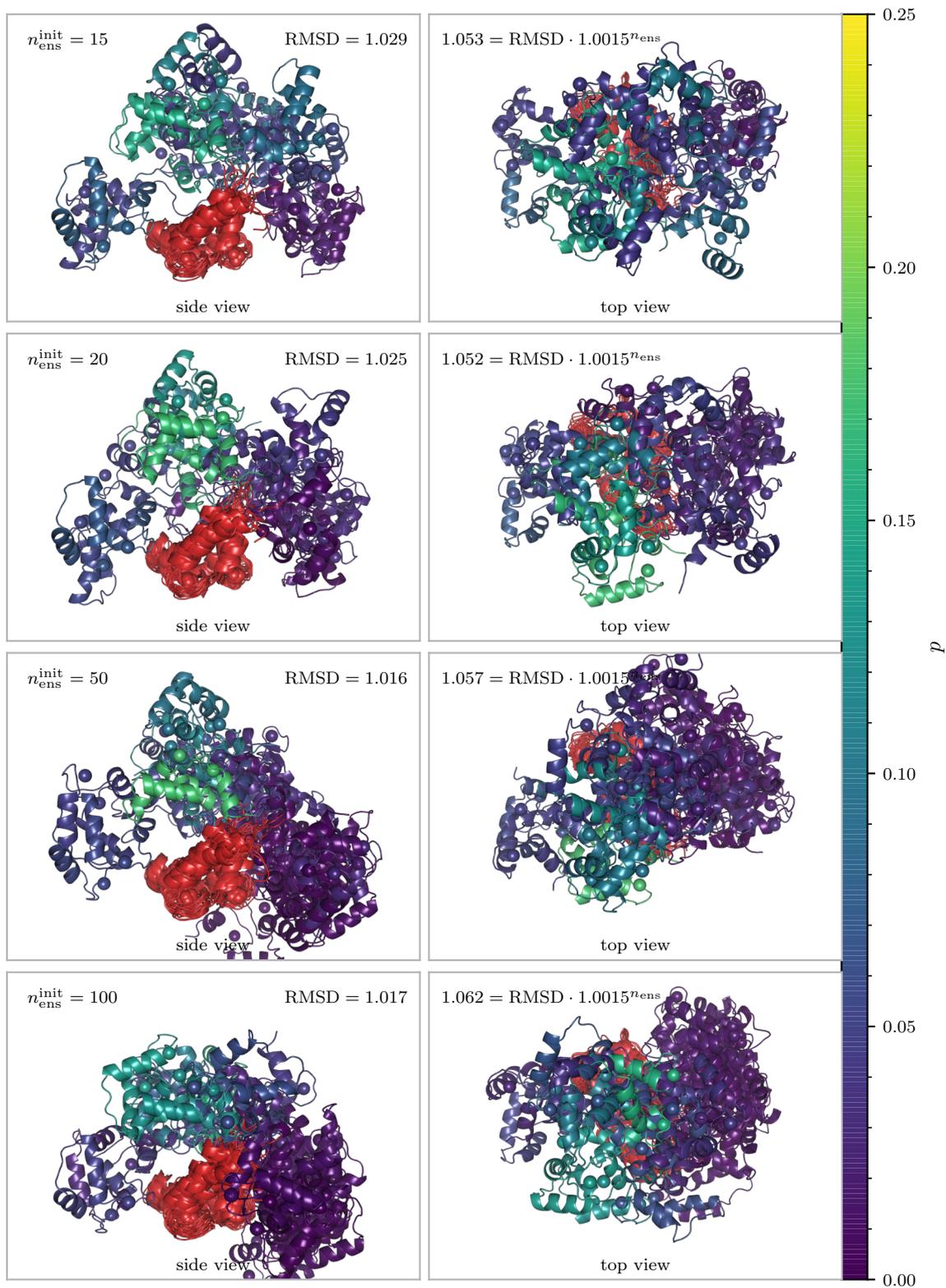


Figure A.8. Cartoon representations of the fifth best ensembles by BIC, for each starting size.

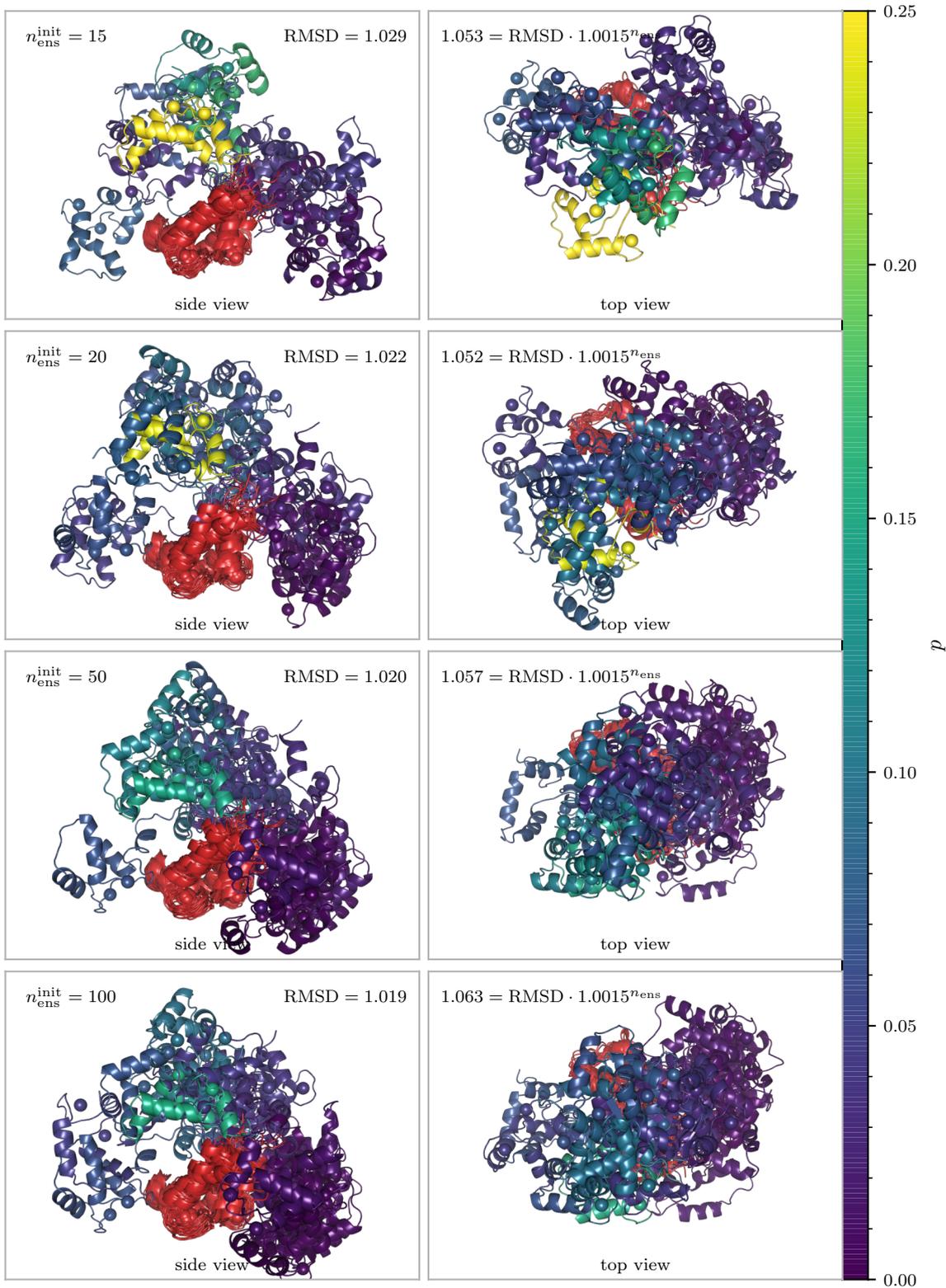
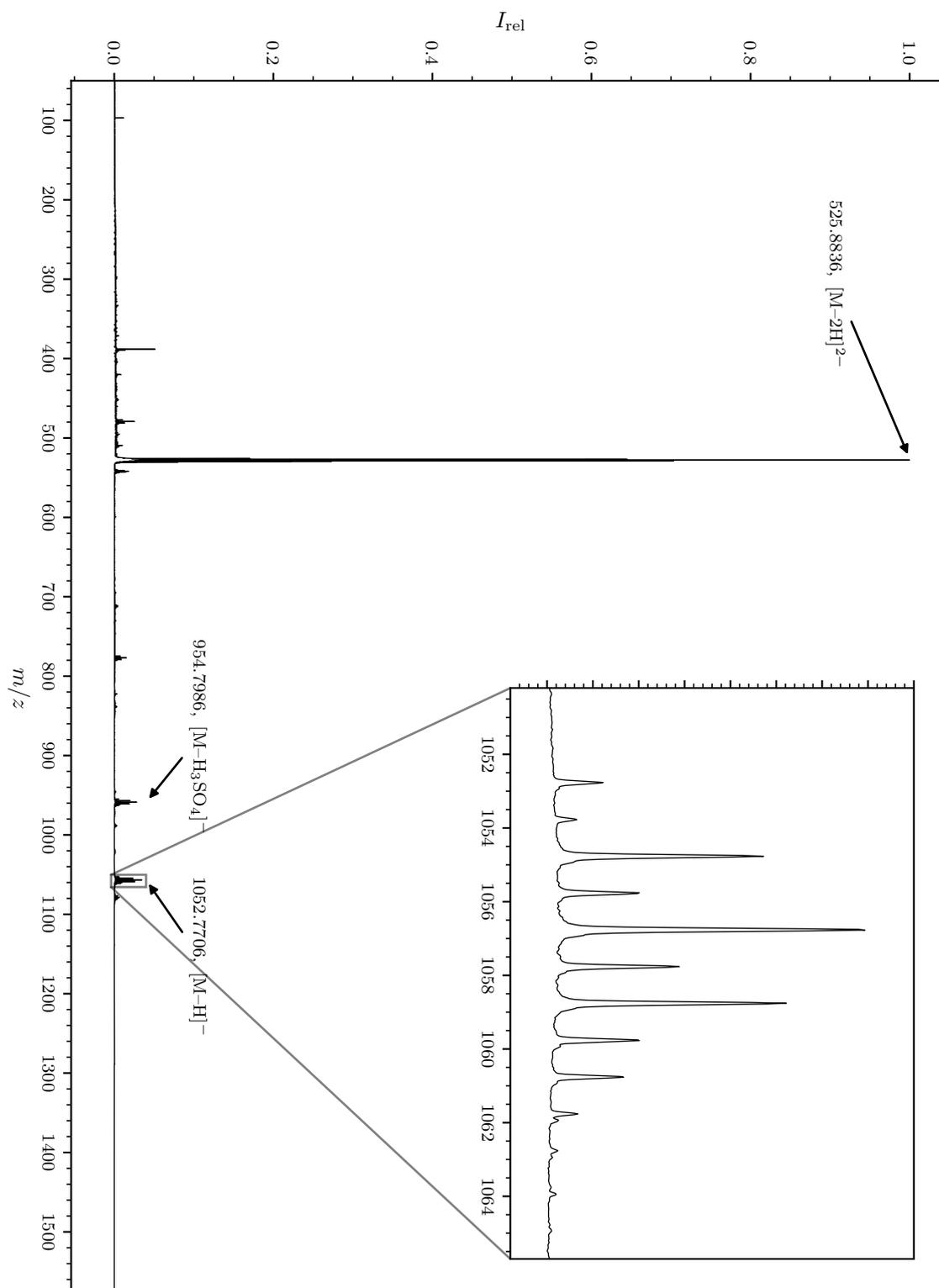


Figure A.9. Cartoon representations of the sixth best ensembles by BIC, for each starting size.

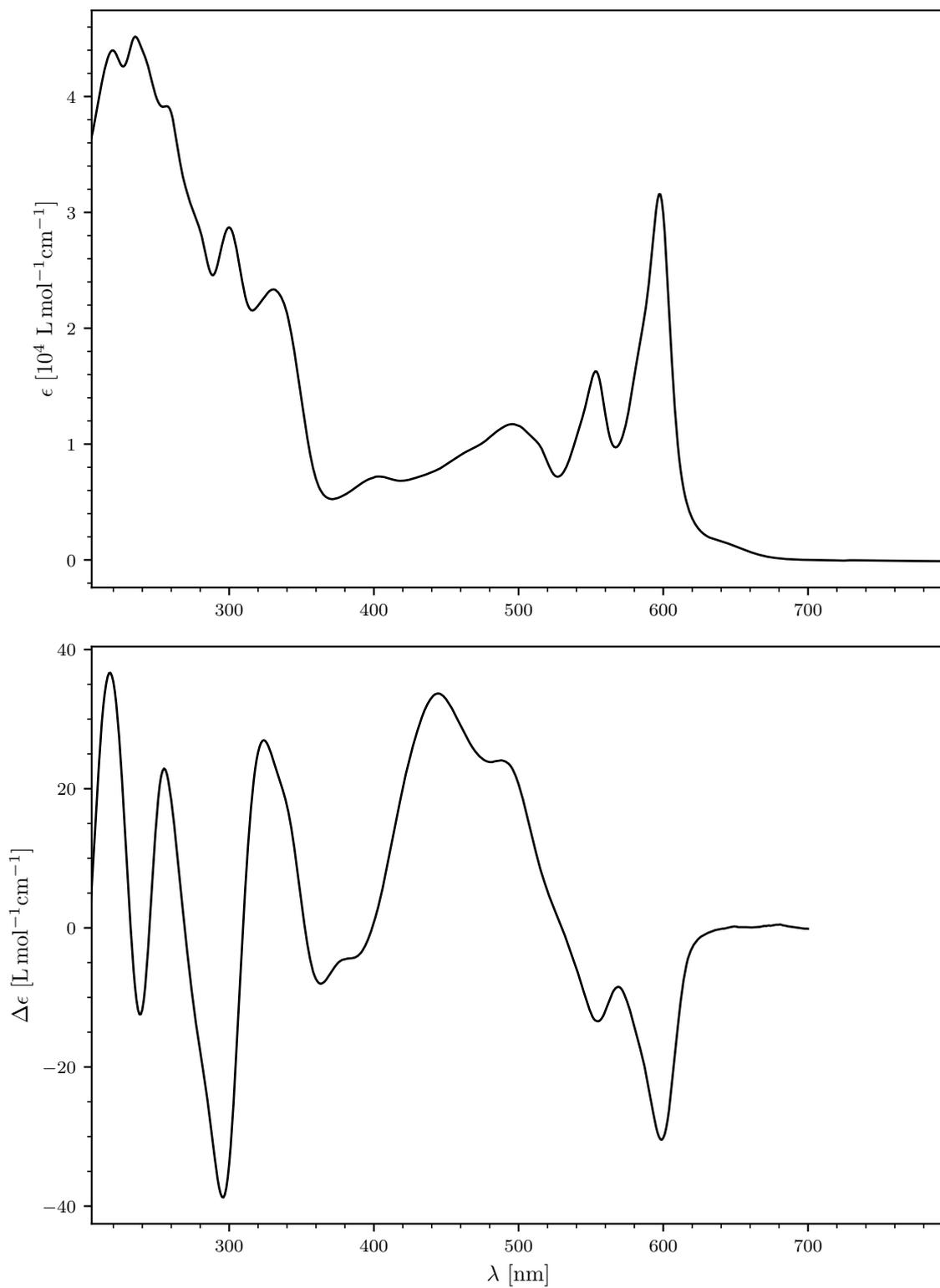
### A.3 Spectral Data of Gymnochrome G

**Table A.3.** NMR-spectroscopic data of gymnochrome G. The unassigned carbon resonances are located in the center of the aromatic core and could not be assigned due to a lack of proton correlations. They belong to the atom numbers 10a, 10b, and 14b–g.

index	$\delta_C$	index	$\delta_C$	$\delta_H$ ( $J$ in Hz)	HMBC
7	185.83	14a	110.85		
14	185.70	6a	110.84		
11	170.31	13a	104.25		
10	170.09	7a	104.16		
13	166.43	9	103.80		
8	166.37	12	103.77		
6	160.92			4.09 dd (13.5; 5.1)	
1	160.70	1'	44.51	3.82 dd (13.4; 8.8)	2, 3, 3a, 2'
4	145.00	2'	71.15	4.79 m	OCOCH <sub>3</sub>
3	142.53	2'- OCOCH <sub>3</sub>	171.53		
	127.65	2'- OCOCH <sub>3</sub>	20.83	1.81 s	OCOCH <sub>3</sub>
	127.55	3'	18.59	0.15 d (6.2)	1', 2'
	126.60	1''	44.06	3.92 dd (14.3; 2.6)	
	126.55			3.86 dd (14.2; 9.6)	3b, 4, 5, 2''
3b	123.79	2''	79.95	4.27 m	
3a	123.47	3''	38.72	1.48 m, 1.64 m	2'', 5''
	122.93	4''	18.87	1.42 m, 1.47 m	5''
	122.92	5''	14.42	0.87 t (7.3)	3'', 4''
	119.46	1-OH		14.94 s	1, 2, 14a, 14
	119.38	6-OH		14.92 s	5, 6, 6a, 7
5	116.82	8-OH		15.47 s	7a, 8, 9, 10, 7
2	116.52	13-OH		15.44 s	11, 12, 13, 13a, 14



**Figure A.10.** HR mass spectrum of gymnochrome G. The main peaks are annotated with the species and its monoisotopic mass.



**Figure A.11.** UV/Vis spectrum (top) and ECD spectrum (bottom) of gymnochrome G in methanol.

### A.3 Spectral Data of Gymnochrome G

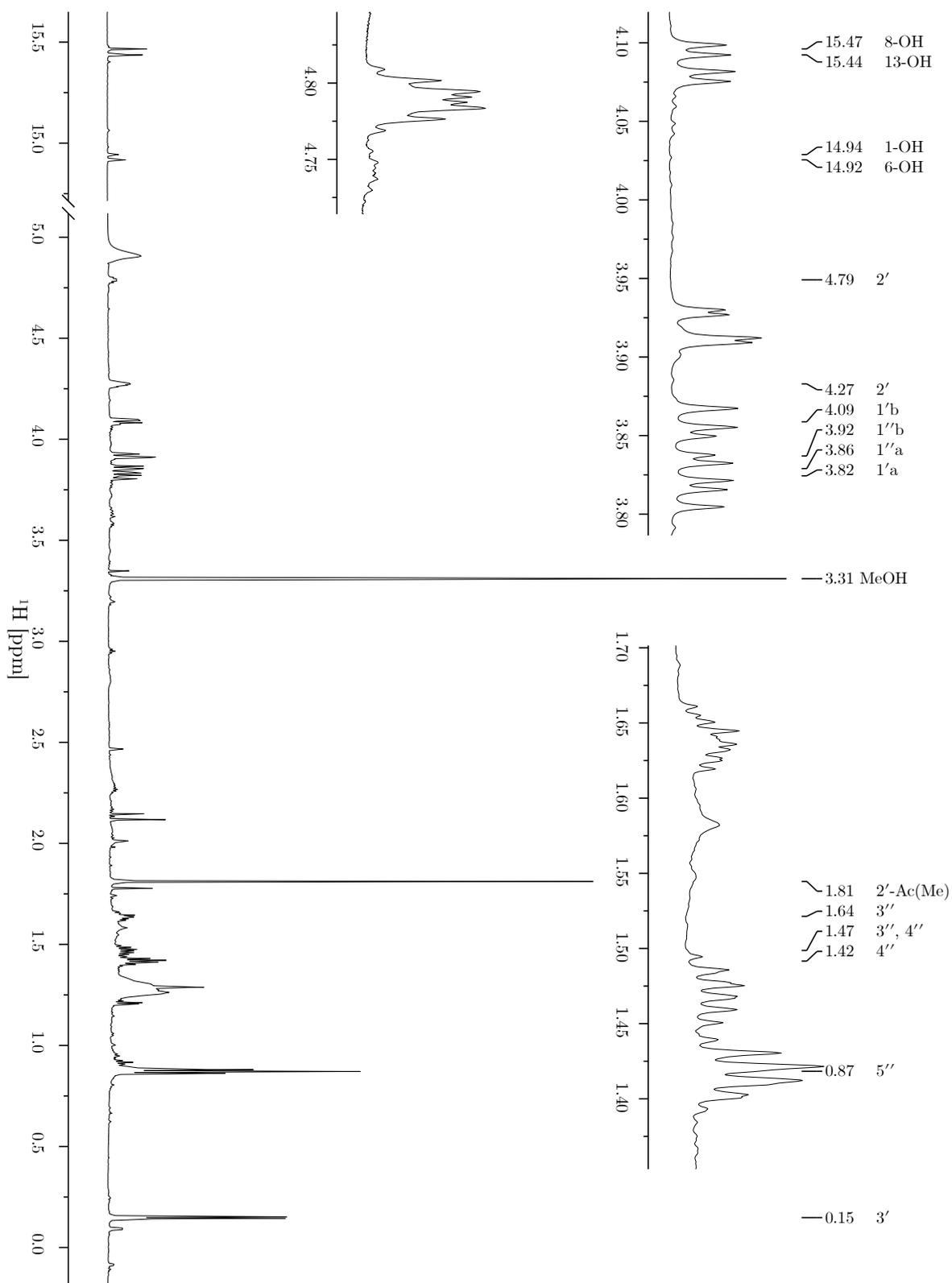


Figure A.12. 800 MHz  $^1\text{H}$  spectrum of gymnochrome G.

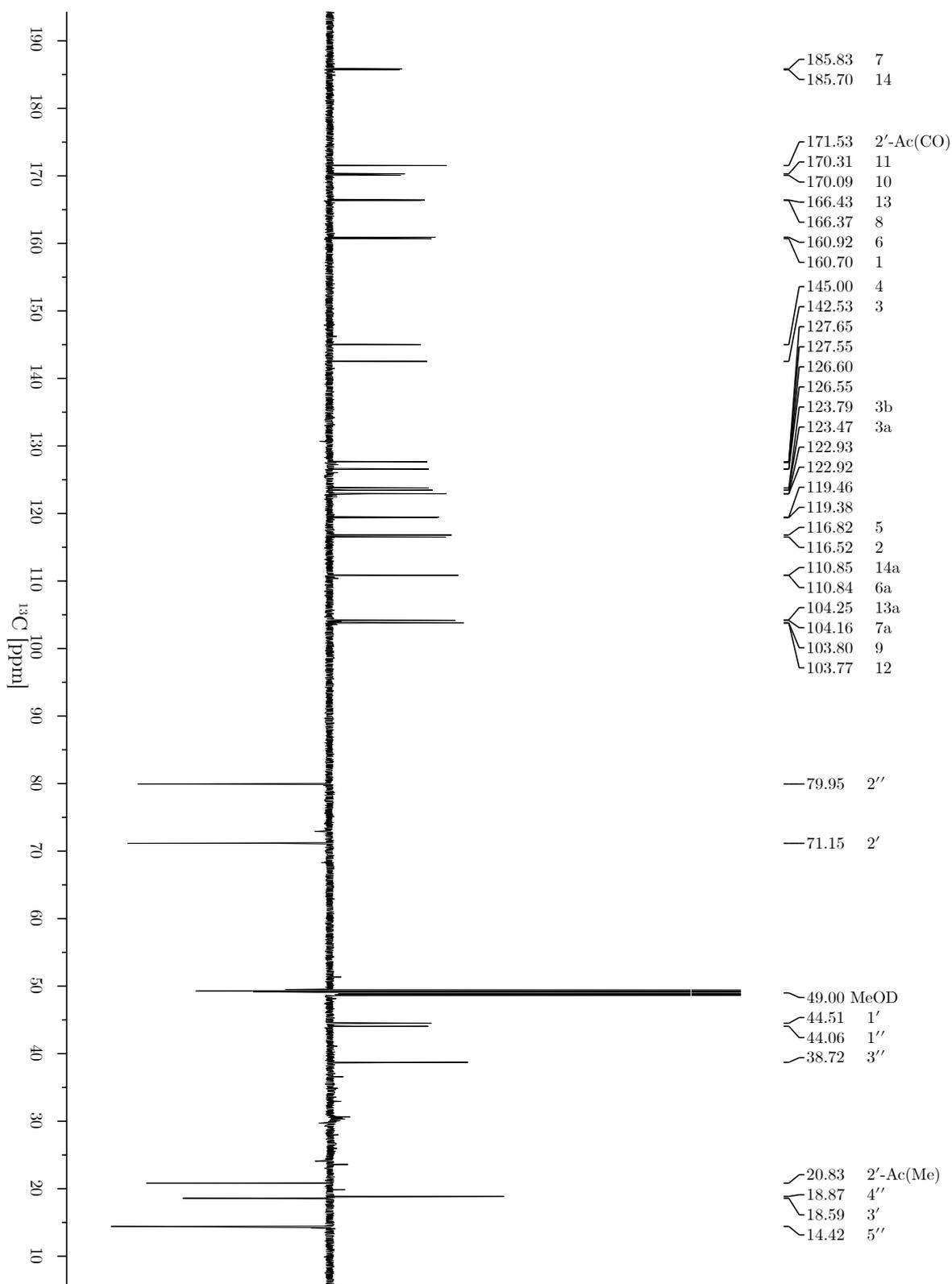


Figure A.13. 200 MHz  $^{13}\text{C}$ -APT spectrum of gymnochrome G.

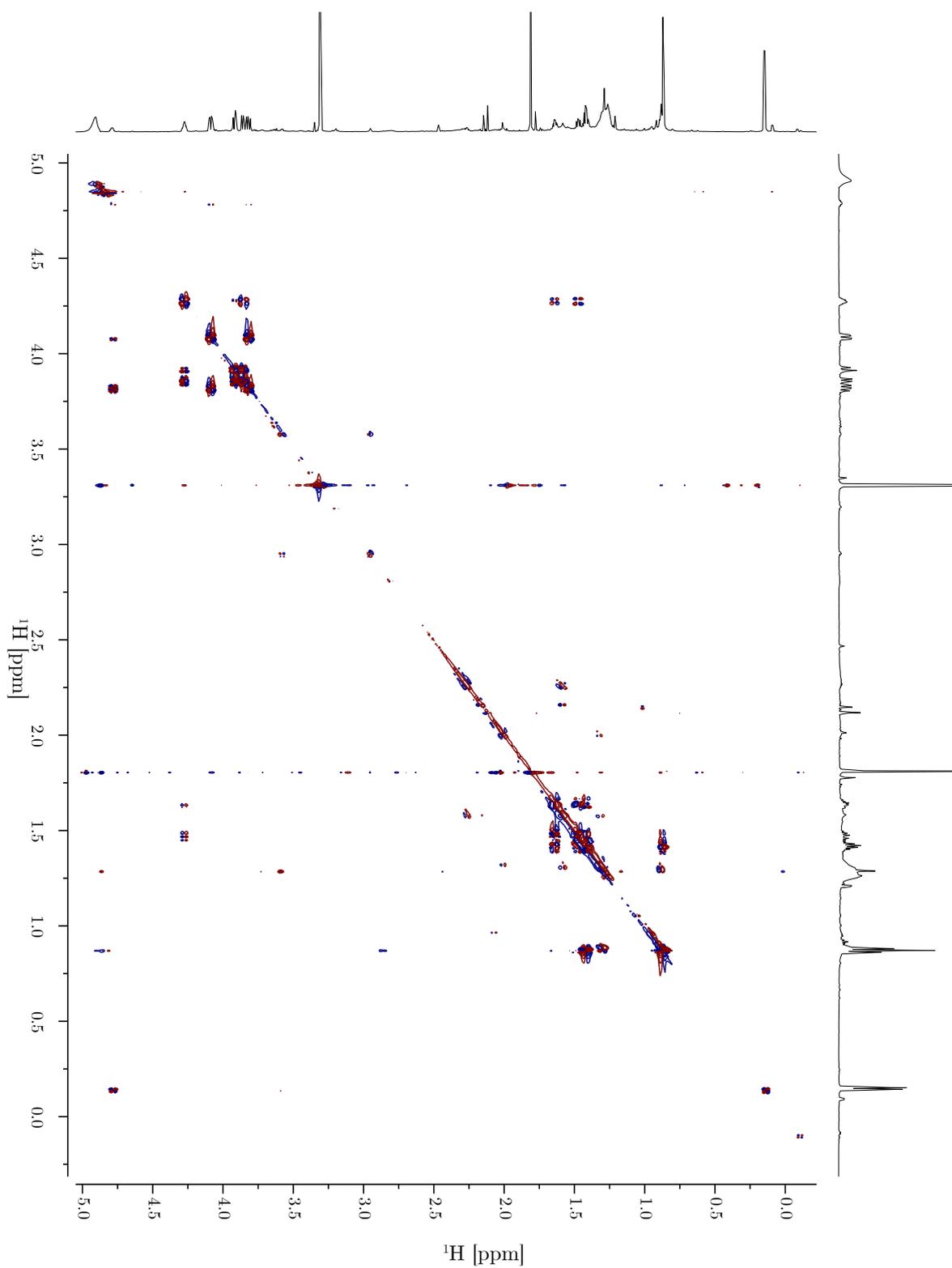


Figure A.14. 800 MHz DQF-COSY of gymnochrome G.

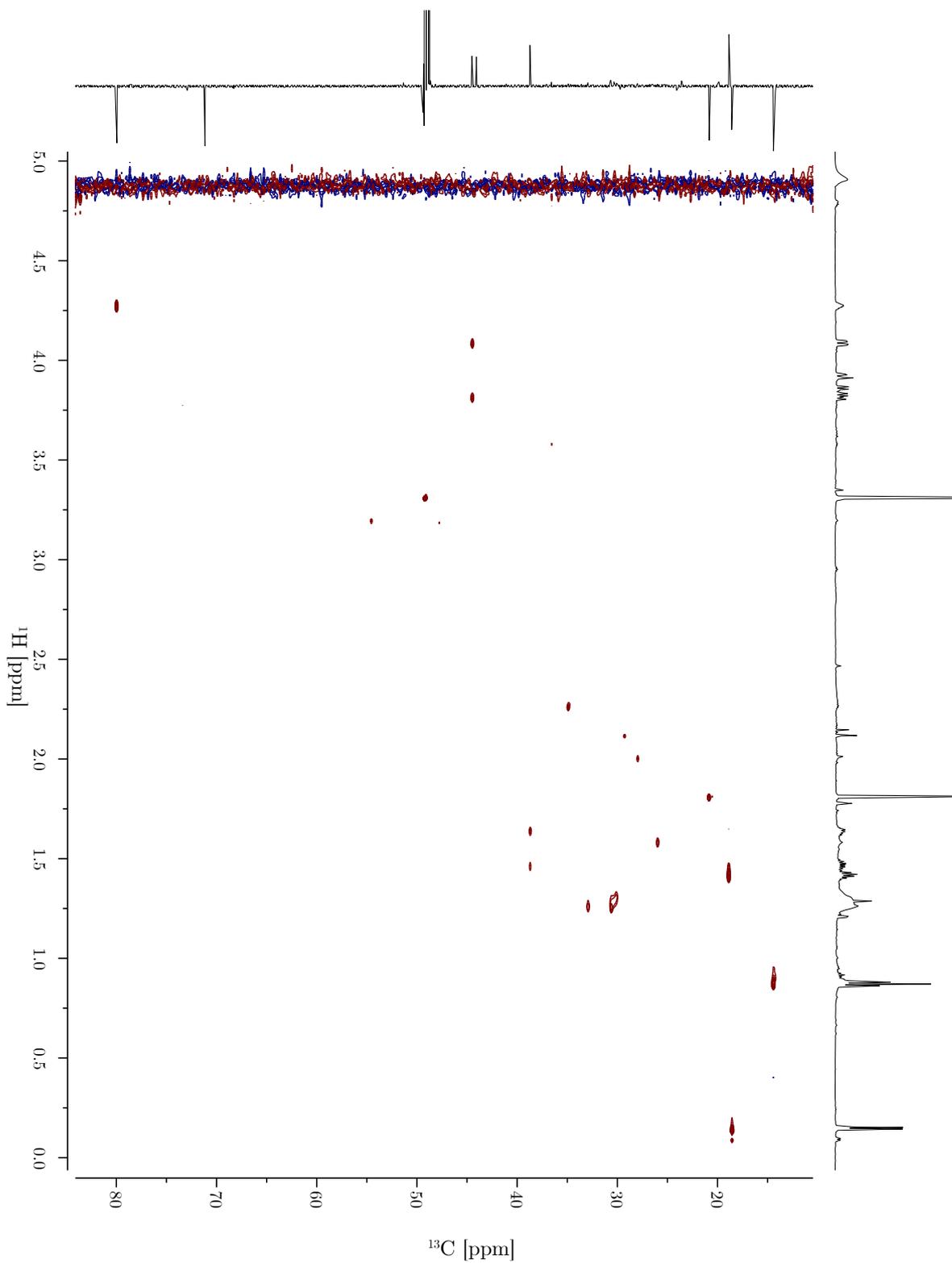
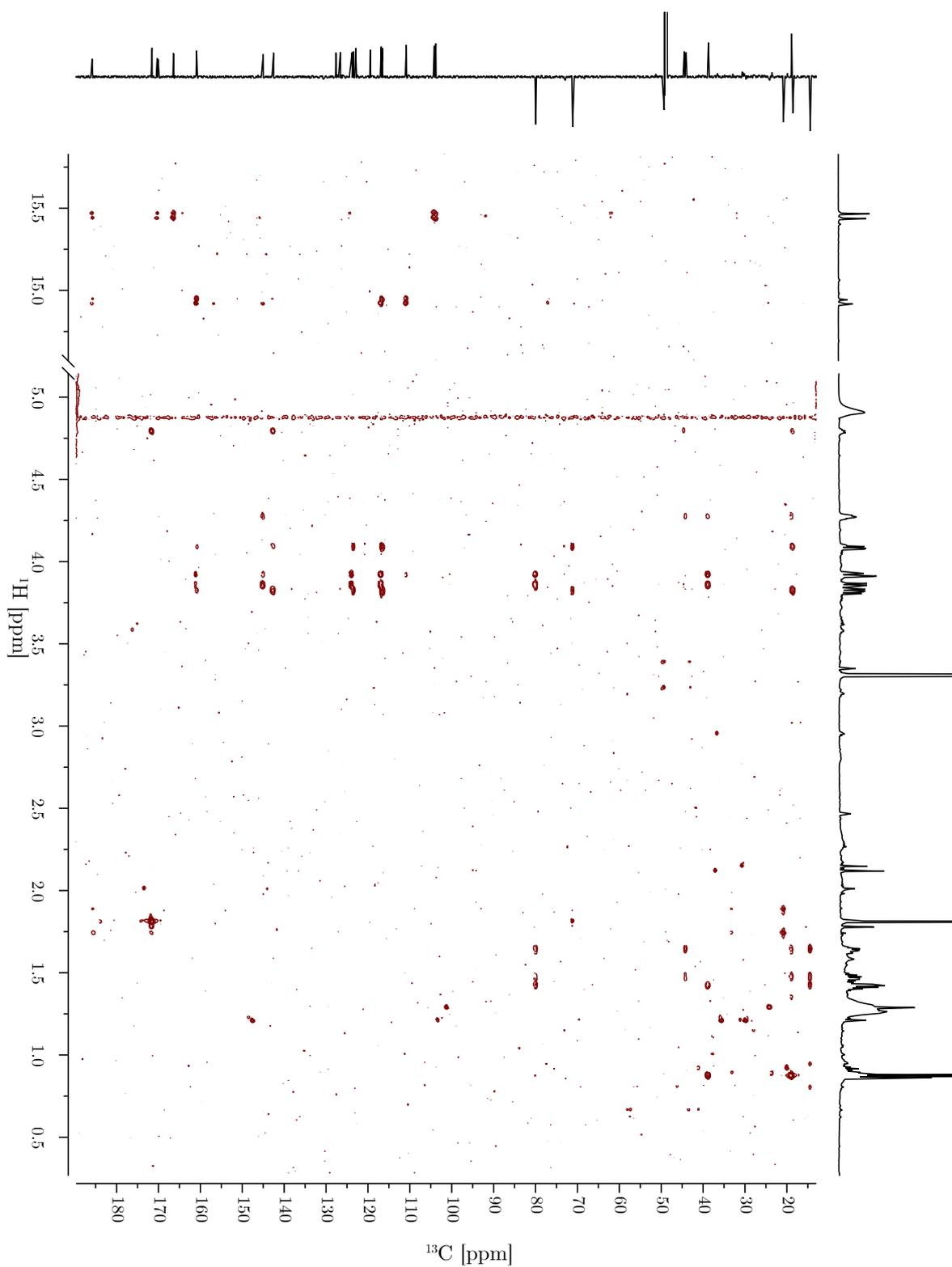


Figure A.15. 900 MHz  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC of gymnochrome G.



**Figure A.16.** 800 MHz  $^1\text{H}$ ,  $^{13}\text{C}$ -HMBC of gymnochrome G, optimized for a small coupling of 3 Hz.

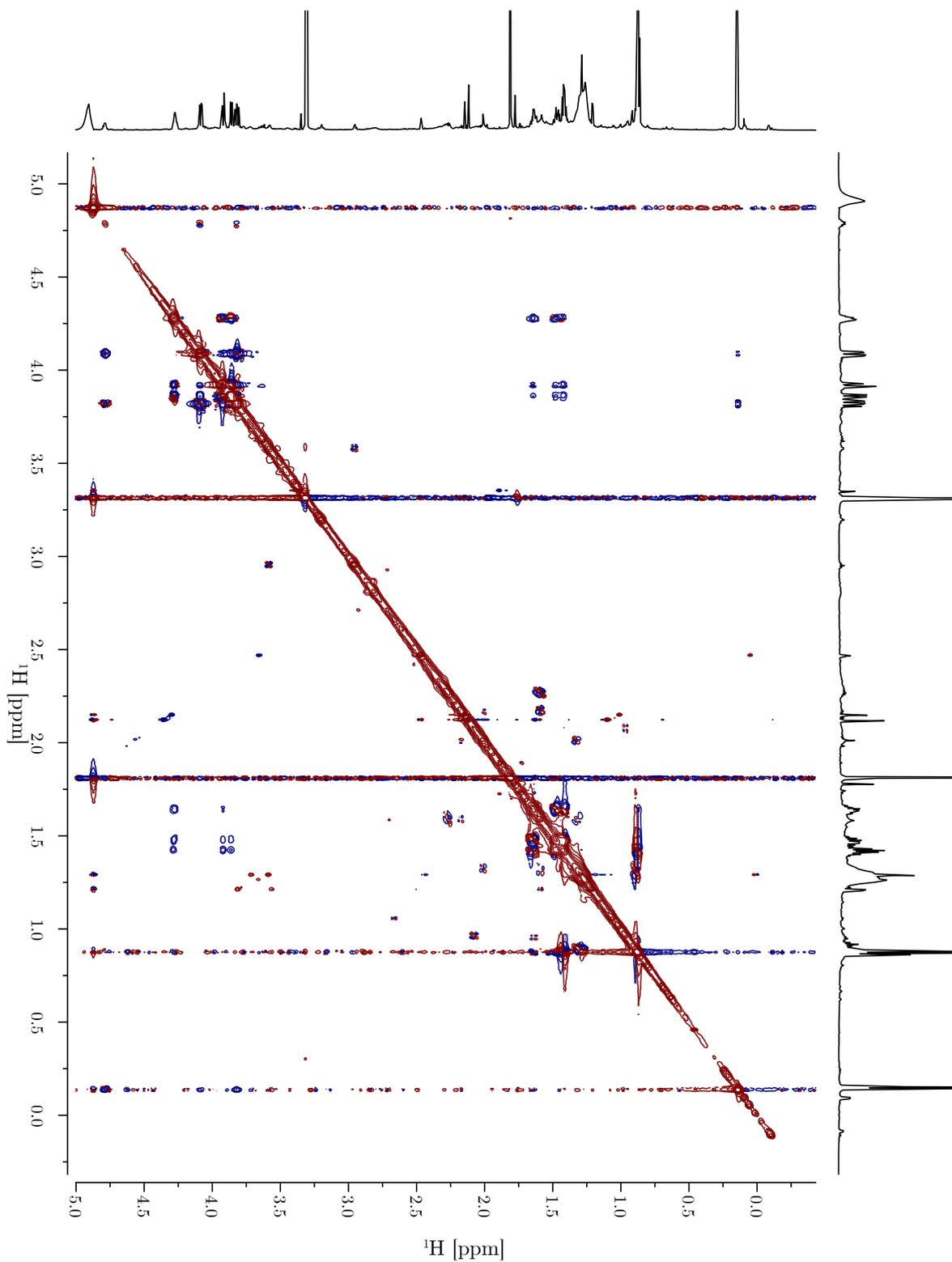


Figure A.17. 800 MHz ROESY of gymnochrome G.

**Table A.4.**  $Q$ -factors and confidences  $p$  for gymnochrome G (**1**) for both approaches.

Configuration	$Q$ (DFT)	$Q$ (fitted)	$p$ (DFT)	$p$ (fitted)
<i>RR</i>	0.229	0.223	$7.46 \cdot 10^{-4}$	$3.72 \cdot 10^{-3}$
<i>RS</i>	0.257	0.235	$1.23 \cdot 10^{-4}$	$9.20 \cdot 10^{-5}$
<i>SR</i>	0.132	0.118	$8.80 \cdot 10^{-1}$	$9.57 \cdot 10^{-1}$
<i>SS</i>	0.201	0.195	$1.19 \cdot 10^{-1}$	$3.88 \cdot 10^{-2}$

**Table A.5.**  $Q$ -factors and confidences  $p$  for strychnine (**2**) for both approaches. A confidence of 0 is equivalent to it being smaller than  $10^{-6}$ .

Configuration	$Q$ (DFT)	$Q$ (fitted)	$p$ (DFT)	$p$ (fitted)
<i>RSSRRS</i>	0.326	0.273	$9.92 \cdot 10^{-1}$	$7.66 \cdot 10^{-1}$
<i>SSSRRS</i>	1.200	0.563	0.00	$2.19 \cdot 10^{-3}$
<i>RRSRRS</i>	0.969	0.552	0.00	$6.75 \cdot 10^{-3}$
<i>SRSRRS</i>	1.073	0.569	0.00	$6.97 \cdot 10^{-4}$
<i>RSRRRS</i>	0.512	0.407	$5.00 \cdot 10^{-3}$	$1.55 \cdot 10^{-2}$
<i>SSRRRS</i>	1.354	0.602	0.00	$6.73 \cdot 10^{-4}$
<i>RRRRRS</i>	0.705	0.470	0.00	$4.30 \cdot 10^{-3}$
<i>SRRRRS</i>	1.239	0.545	0.00	$6.47 \cdot 10^{-4}$
<i>RSSSRS</i>	0.751	0.334	0.00	$1.52 \cdot 10^{-1}$
<i>RRSSRS</i>	0.573	0.392	$1.68 \cdot 10^{-3}$	$3.15 \cdot 10^{-2}$
<i>SRSSRS</i>	1.263	0.602	0.00	$9.10 \cdot 10^{-5}$
<i>RSRSRS</i>	0.822	0.475	0.00	$1.13 \cdot 10^{-3}$
<i>RRRSRS</i>	0.684	0.485	$2.90 \cdot 10^{-5}$	$1.93 \cdot 10^{-3}$
<i>SRRSRS</i>	1.323	0.658	0.00	$2.10 \cdot 10^{-5}$
<i>RSSRSS</i>	1.057	0.577	0.00	$4.37 \cdot 10^{-3}$
<i>RRSRSS</i>	0.931	0.525	0.00	$7.30 \cdot 10^{-3}$
<i>RSRRSS</i>	0.976	0.568	0.00	$3.21 \cdot 10^{-4}$
<i>RRRRSS</i>	0.620	0.470	$3.10 \cdot 10^{-5}$	$1.68 \cdot 10^{-3}$
<i>RSSSSS</i>	0.986	0.548	0.00	$2.03 \cdot 10^{-4}$
<i>RRSSSS</i>	0.713	0.452	$8.57 \cdot 10^{-4}$	$8.44 \cdot 10^{-4}$
<i>RSRSSS</i>	1.020	0.550	0.00	$1.43 \cdot 10^{-3}$
<i>RRRSSS</i>	1.044	0.626	0.00	$9.00 \cdot 10^{-6}$



## **Affidavit**

I hereby declare that I have developed and written this dissertation “From Paramagnetic Proteins to Field Alignment of Small Molecules – Tools for Structure Determination via NMR” without any external unauthorized help, and that I have not used sources or means without declaration in the text.

Göttingen, 25 November 2021