Shape Matters

Characterization of Weak Interactions and Macrocycles by Conformational Analysis

STEFAN PEINTNER

Dissertation presented at Uppsala University to be publicly examined in A1:111a, BMC, Husargatan 3, Uppsala, Friday, 16 December 2022 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Oliver Zerbe (University of Zurich - Department of Chemistry).

Abstract
Peintner, S. 2022. Shape Matters. Characterization of Weak Interactions and Macrocycles by Conformational Analysis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2211. 78 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1643-7.

When studying conformational ensembles, it is most challenging to identify and characterize rapidly interconverting individual conformers. A precise description of the structural dynamics, however, rewards viable knowledge on conformation stabilizing forces as well as on the impact of external influences on the conformer composition. This thesis focusses on the solution NMR spectroscopic conformational analysis of flexible molecules with the aim to unveil the impact, strength and geometry of a single weak non-covalent interaction, a halogen bond. In addition, I investigated the impact of the solvent polarity on the conformational distribution of macrocyclic drugs.

Weak interactions are difficult to characterize in solution with current techniques. Therefore, I designed a peptidic β-hairpin model system that serves as platform to probe a single weak halogen bond in solution. The presented strategy benefits from the cooperativity of non-covalent forces, from preorganization, and from the entropic advantage of studying an interaction in an intramolecular setting. A weak C−I⋯O halogen bond was characterized thermodynamically and geometrically using NMR-based variable temperature, NOE, scalar coupling and RDC analyses. Time-averaged NMR parameters were deconvoluted with NAMFIS and by Singular Value Decomposition as implemented into MSpin. Characterization of such a weak interaction ($\Delta\Delta G° < 0.9$ kJ/mol) in dilute solution is remarkable.

Making use of NMR-based ensemble analysis, I further studied the impact of solvent polarity on the conformational distribution of macrocyclic drugs that do not obey Lipinski's Rule-of-5, yet experience good membrane permeability. I demonstrate experimentally that this class of compounds behaves as molecular chameleons by adjusting their conformation to shield or expose polar functionalities as an adaptation to the surrounding environment. Solution ensembles in D2O and CDCl3 mimicking the plasma/cytosol and cell membrane, respectively, were determined and thus revealed that the flexibility of studied macrocycles facilitates a major difference in size and polarity between different environments.

Overall, this thesis demonstrates the capabilities and precision of solution NMR spectroscopic conformation analysis techniques, and two possible applications of their use for solving scientific challenges of high relevance to medicinal and organic chemistry.

Stefan Peintner, Department of Chemistry - BMC, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

to my family

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Peintner, S., Erdélyi, M. (**2022**) Pushing the Limits of Characterising a Weak Halogen Bond in Solution. *Chemistry – A European Journal* 28, e202103559

II. Danelius, E., Poongavanam, V., Peintner, S., Wieske, L. H. E. Erdélyi, M., Kihlberg, J. (**2020**) Solution Conformations Explain the Chameleonic Behaviour of Macrocyclic Drugs. *Chemistry – A European Journal* 26, 23, 5231-5244

III. Umereweneza, D., Atilaw, Y., Peintner, S., Rudenko, A., Bourgard, C., Xiong, R., Muhizi, T., Gogoll, A., Erdélyi, M.; Macrocyclic Pyrrolizidine Alkaloids and Silphiperfolanol Angelate Esters from *Senecio manii. Submitted*

Reprints were made with permission from the respective publishers.

The following publications are not included as part of this thesis:

IV.   Wieske, L. H. E., <u>Peintner, S.</u>, Erdélyi, M. (**2022**) Ensemble Determination by NMR Data Deconvolution. *Nature Reviews Chemistry – in revision*

V.   Reutskaya, E., Sapegin, A., <u>Peintner, S.</u>, Erdélyi, M., Krasavin, M. (**2021**) Sulfur Oxidation Increases the Rate of HIRE-Type [1.4]Thiazepinone Ring Expansion and Influences the Conformation of a Medium-Sized Heterocyclic Scaffold. *Journal of Organic Chemistry* 86, 8, 5778-5791

VI.   Bogaerts, J., Atilaw, Y., <u>Peintner, S.</u>, Aerts, R., Kihlberg, J., Johannessen, C., Erdélyi, M. (**2021**) Employing Complementary Spectroscopies to Study the Conformations of an Epimeric Pair of Side-chain Stapled Peptides in Aqueous Solution. *RSC Advances* 11, 7, 4200-4208

VII.   Grintsevich, S., Sapegin, A., Reutskaya, E., <u>Peintner, S.</u>, Erdélyi, M., Krasavin, M. (**2020**) An Alternative Approach to the Hydrated Imidazoline Ring Expansion (HIRE) of Diarene-Fused [1.4]Oxazepines. *European Journal of Organic Chemistry* 35, 5664-5676

VIII. Yang, J., Talibov, V. O., <u>Peintner, S.</u>, Rhee, C., Poongavanam, V., Geitmann, M., Sebastiano, M. R., Simon, B., Hennig, J., Dobritzsch, D., Danielson, U. H., Kihlberg, J. (**2020**) Macrocyclic Peptides Uncover a Novel Binding Mode for Reversible Inhibitors of LSD1. *ACS Omega* 5, 8, 3979-3995

IX.   Poongavanam, V., Danelius, E., <u>Peintner, S.</u>, Alcaraz, L., Caron, G., Cummings, M. D., Wlodek, S., Erdelyi, M., Hawkins, P. C. D., Ermondi, G., Kihlberg, J. (**2018**) Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? *ACS Omega* 3, 9, 11742–11757

Contents

Abbreviations

Å	Ångström, 0.1 nm
COSY	correlated spectroscopy
DCM	dichloromethane
DFT	density functional theory
DIPEA	*N,N*-diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
eq.	equivalents
GB/SA	generalised Born and surface area
HB	hydrogen bond
HMBC	heteronuclear multiple bond correlation spectroscopy
HOMO	highest occupied molecular orbital
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single-quantum coherence spectroscopy
IUPAC	International Union of Pure and Applied Chemistry
kJ	kilojoule
LUMO	lowest unoccupied molecular orbital
MCMM	Monte-Carlo multiple minima
MeCN	acetonitrile
NAMFIS	NMR analysis of molecular flexibility in solution
NMP	N-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser enhancement spectroscopy
ppm	parts per million
R^2	coefficient of determination
RDC	residual dipolar coupling
RMSD	root-mean-square deviation
SSD	sum-of-squares differences
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofurane
TIS	triisopropylsilane
TOCSY	total correlation spectroscopy
XB	halogen bond
δ	chemical shift

1. Introduction

Chemical bonds hold together the structures necessary for life to exist with carbon as an essential component. However, these bonds are not completely rigid. They hold a certain degree of rotational freedom and flexibility. Consequently, entire molecules can adapt a variety of different conformations without the breaking of any chemical bonds. Adaptive binding as well as the triggered conversion of distinct conformations is the basis for the effective interaction of molecules. The incentive to study not just the chemical composition, but also the three-dimensional structure, awarded Derek Barton and Odd Hassel a Nobel Prize in 1969 'for their contribution to the development of the conformational analysis and its application in chemistry'.[1] Their work on simple cyclohexane and steroid *cis*/*trans*-isomers initiated the connection of conformation to internal energies. For conformationally restrained molecules or molecules with a small number of rotatable bonds, the preferred or low energy conformations are nowadays computationally accessible. However, with increasing flexibility and the dynamic formation of intermolecular non-covalent interactions, the gained complexity reduces predictability. Furthermore, as soon as molecular constraints decrease and the number of rotatable bonds rises, the energy landscape of possible conformations becomes much flatter and the calculation of precise energies becomes cumbersome.
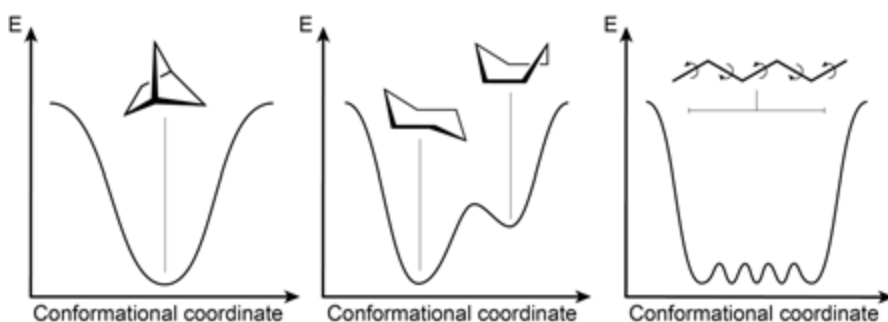

Figure 1: Energy landscapes for the conformations that three different C_6 skeletons can adapt. The number of rotatable bonds and the energy barriers between the different conformations direct the dynamics and flexibility of the molecule.

Figure 1 shows three molecules that all contain six carbon atoms, yet are vastly different in flexibility, as reflected by the dissimilar energy distributions of their adapted conformations. This example also shows that the given dynamics and flexibility of a molecule is not entirely dependent on the size of the molecule, rather the distribution of conformers is directed by their energies. The conformers are typically in constant exchange at a rate dependent on the interconversion energy barrier. Life, however, depends on functional molecules with a distinct shape and stability under physiological conditions. The coiling into specific three-dimensional shapes solely depends on non-covalent interactions. These weak bonds follow an organization determined by the sequence of the biological macromolecule and help to decrease the energy of the folded and functional conformers.[2] The resulting equilibrium between folded and unfolded conformers constitutes the fundamentals for the propagation of life. To study conformations and relate them to biological functions as well as the intentional introduction of non-covalent interactions to modify a secondary structure, aids for the development of analytical and *de novo* design tools. In this thesis, I focus on the experimental determination of the conformational properties of a variety of molecules, and the use of this knowledge for gaining improved understanding of a single weak interaction in solution (Paper I), and of the membrane permeability of large and polar macrocyclic drugs and drug candidates (Paper II). In Paper III, I utilize the restrained conformational freedom of a macrocyclic natural product to determine its relative stereochemistry, which has not yet been determined experimentally.

1.1.   Forces Determining Molecular Conformations

Molecules can assume different shapes which impacts their means of interactions and reactions with other substances. Without a way to preorganize and stabilize a reactive structure, the energy barrier would be insuperably high for any reaction to happen. Considering that structure relates to function, conformational restraints serve the purpose to shape a molecule for the task it is meant to fulfil. When designing a molecule whose conformation is important for its action, a multitude of structural elements and interactions must be balanced. Some important factors for maintaining a molecule's dynamics are: semi-rigid constraints like macrocyclisation, stereogenic elements and, for peptidic structures, turn-, sheet- or helix-inducing units. Further to this, dynamic non-covalent attractive interactions like hydrogen bonds and halogen bonds as well as salt bridges, π-π, van der Waals and hydrophobic interactions may contribute to the overall conformational stability. For the successful design of a functional structure, or for

understanding of its dynamic behavior, gaining distinct knowledge of all contributing factors is the key.

1.2.   Semi-rigid Constraints

Macrocycles

Macrocycles emerged as new modality for the inhibition of challenging targets, such as the flat interface of protein-protein interactions.[3, 4] Macrocyclisation of a known binder in an optimal case may stabilize its bioactive conformation and improve its metabolic stability, cell permeability, and in some cases, its oral bioavailability.[5] The slight reduction of flexibility upon cyclisation relates to an increased affinity as compared to the linear counterpart (Figure 2a). This is founded on the lower entropic penalty for the ligand binding to the target.[6, 7] Additionally, macrocycles have a beneficial balance between conformational restriction and flexibility that allows for a distinct adaptive behavior upon the change in polarity of the solvating environment.[8] This phenomenon is called molecular chameleonicity and will be discussed further in Paper II.

β-Turns

β-Turns were first described by C. M. Venkatachalam in 1968 and are now considered as a common structural motif in ~25% of all residues in folded peptides and proteins.[9, 10] Due to their important role in biological structures, there is a considerable interest to design β-turns and β-turn mimetics to improve stability and thus biological activity, or to enhance the bioavailability of peptidomimetic drugs. A β-turn or reverse turn, changes the direction of two secondary structural elements. It is commonly described as a four-residue sequence denoted i to $i+3$ and contains amino acids with strong turn-forming propensity (Figure 2b). The 180° rotation of the sequence backbone results in a distance of $C_\alpha i$ and $C_\alpha i+3$ of less than 7 Å, thus promoting the formation of cross-strand hydrogen bonds.[11] Different β-turns have been identified and classified depending on the Φ and Ψ dihedral angles adopted by the two central turn amino acids. Most important are the type I, I' and II, II' β-turns where I' and II' are the diastereoisomers of their corresponding original types. Due to this diastereotopic relationship, they have identical Φ and Ψ angles but with opposite sign. However, they are not energetically equivalent and I' and II' β-turns show a much higher propensity for β-sheet nucleation than either of the type I and II β-turns.[12]

Table 1: Values for the Φ and Ψ dihedral angles (depicted in Figure 2) of the β-turn residues $i+1$ and $i+2$ that classify the adopted turn.

	Ideal dihedral angles			
	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}
Type I	-60°	-30°	-90°	0°
Type I'	60°	30°	90°	0°
Type II	-60°	120°	90°	0°
Type II'	60°	-120°	-90°	0°

The different propensities may result from the conformation of these turns to be more compatible with the left-handed twist of an anti-parallel β-sheet.[13] In Paper I, a central ^{D}Pro-Gly turn unit adopting a type II' β-turn was employed to promote the formation of a stable β-hairpin secondary structure.
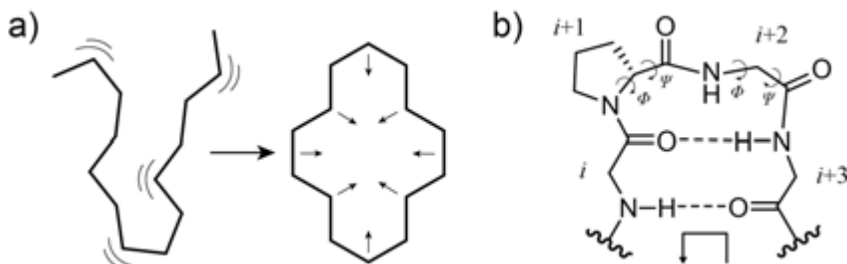

Figure 2: a) Macrocyclisation as an example of restraining and stabilizing a conformation. A Macrocycle is defined as a cyclic framework with ≥ 12 ring atoms. b) A β-turn inducing unit of a peptide strand shown as the two amino acid residue ^{D}Pro-Gly at position $i+1$ and $i+2$.

1.3. Non-covalent Interactions

Non-covalent interactions, in contrast to the previously described conformation determining forces, do not inhere a consistent connection of the atoms involved. These interactions are based on dynamic forces resulting from the formation and breaking of non-covalent interactions within a wide range of interaction strengths. The energy of attraction between both binding partners is based on the degree of orbital overlap, where the weakest interactions result from the least amount of charge effects from the involved electrons, e.g. van der Waals forces. The larger the displacement of electrons from one binding partner to the other, the more covalent the bond character and consequently the stronger the interaction between them. Non-covalent interactions are critical in the stabilization of 3D structures of large molecules, biological binding processes, chemical transformations or drug design. The utilization of a hydrogen bond network and the detailed analysis of a weak

halogen bond, both non-covalent interactions with a wide range of bond strengths, are the focus of this thesis.

Hydrogen Bonding

Hydrogen bonding, HB, is by far the most investigated non-covalent interaction.[14] It is involved in every aspect of biological function, aiding structure stability, enabling enzyme catalysis and providing the essential properties of water. It embodies a mainly electrostatic interaction between a polarized, covalently bound hydrogen atom and another electronegative acceptor atom.[15] The acceptor atom must bear a lone pair of electrons devoted to and thus attracting the hydrogen atom. A hydrogen bond generally increases in strength the more electronegative the acceptor is. The same accounts for the donor moiety carrying the hydrogen atom, increasing the dipolar moment to the electron poor hydrogen.
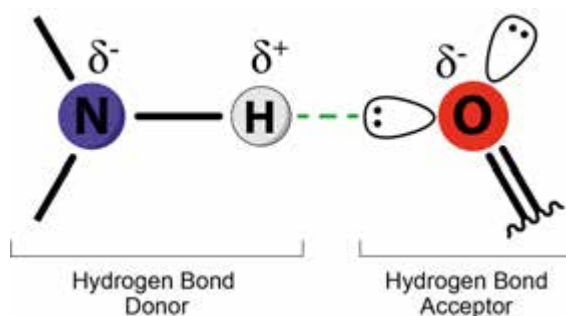

Figure 3: Example of a N−H···O hydrogen bond shown as the electrostatic interaction between two dipoles.

Thus, the hydrogen bond is considered a permanent dipole-dipole interaction with observed bond strengths ranging from 1-2 kJ/mol to >160 kJ/mol;[16] as a reference, the water-water dimer dissociation energy is about 15 kJ/mol.[17] The real power of the hydrogen bond lies in its ability to promote cluster formation in a cooperative way to determine the conformation of a molecule under physiological conditions. In the secondary structure of peptides and proteins, the backbone amide's C=O and N-H groups form intramolecular hydrogen bonds. Depending on the spacing of amino acids and orientation of the backbone strands towards each other, either α-helical or β-sheet structures are formed.[18] The cooperativity of backbone hydrogen bonds to form an isolated β-sheet structure is an integral part of this thesis and will be discussed in more detail in Paper I.

Halogen Bonding

Halogen bonding, XB, is described by a nomenclature similar to that used for the hydrogen bond.[19] Here, a halogen acts as an electrophile and is designated the role of 'halogen bond donor' (XB donor). The atom or group the halogen interacts with is a nucleophile or Lewis base and is referred to as the 'halogen bond acceptor' (XB acceptor).[20] Following the electrostatic model, this attraction is based on the fact that valence electrons on halogens are not equally dispersed around the atom's outer sphere.[21] This anisotropic distribution concentrates the electrons on a belt perpendicular to the covalent σ-bond (Figure 4).
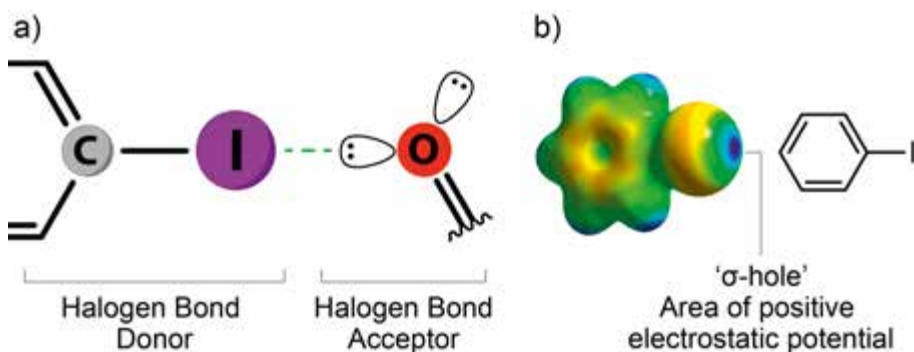

Figure 4: Example of a halogen bond depicted using the electrostatic interaction model. a) The attractive force between a non-bonding electron pair of the halogen bond acceptor (O) and the σ-hole of the halogen (I). b) The σ-hole is described as the area of positive electrostatic potential arising from the anisotropic distribution of the halogen's valence electrons.

This results in an area of positive electrostatic potential antipodal to the R-X bond, referred to as the 'σ-hole'. The more polarizable and less electronegative the halogen (I>Br>Cl>>F), the stronger the electron depletion at the σ-hole. This electrophilic area is small as compared to the size of a hydrogen atom, which makes the halogen bond more directional than a hydrogen bond. Controversial theories question whether the halogen bond is mainly driven by an electrostatic attraction of the σ-hole to the nucleophile or *via* partial electron transfer from the nucleophile into the halogen's σ*-orbital.[22-24] The charge-transfer model explains this interaction by the small energy difference of the XB acceptor's HOMO and XB donor's LUMO (the anti-bonding σ*-orbital).[25] The orbital overlap provides an increased covalent character of the halogen bond (Figure 5).
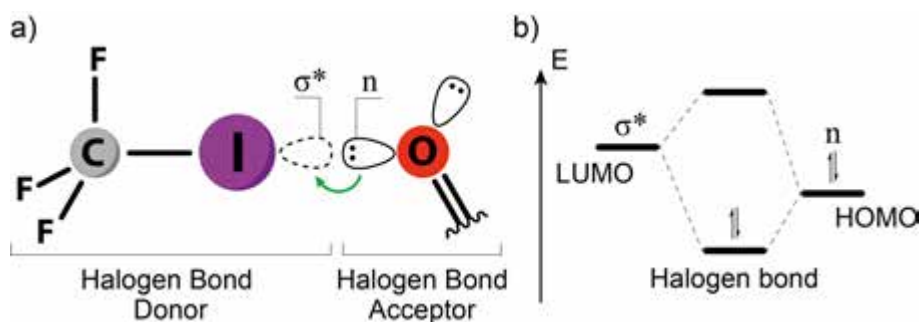
Figure 5: Example of a halogen bond depicted using the charge transfer model. a) The strong XB donor CF$_3$I overlaps its σ*-orbital with the non-bonding orbital (n) of the XB acceptor, thus forming a halogen bonded complex with covalent character. b) The small energy difference between the halogen's LUMO and the XB acceptor's HOMO promotes charge transfer.

Analysis of a large number of halogen bond complexes suggests that the major contribution to the bond strength varies. In strong complexes (stabilization energy >30 kJ/mol), the charge transfer component contributes most to the binding energy,[19] whereas weaker complexes are mainly held together by an electrostatic interactions.[26] However, a detailed description of the halogen bond − in particular guidelines for its intentional use in drug development as well as computational parametrization – is still missing. In Chapter 3, I provide the experimental characterization of a weak I···O halogen bond in solution.

1.4. Conformational Analysis by NMR

A flexible molecule in solution is present as a set of interchanging conformations. This collection of conformers is commonly referred to as the conformational ensemble. Certain conformations are lower in energy than others, and accordingly are more prevalent. Modifying the factors that stabilize certain conformers shifts the equilibrium of the ensemble or may even induce the appearance of new conformers. Conformational analysis provides information about the geometric preferences of a molecule as well as about its sensitivity to internal or external influences.

In this thesis, solution NMR conformational analysis of flexible molecules was used for the development of a method to investigate weak interactions, and to gain understanding of the impact of the environment's polarity on conformational ensembles of drug-like molecules. In Paper I, a β-hairpin peptide was designed to enable the determination of the strength and geometry of a previously undetectable weak I···O halogen bond. The cooperative folding-defolding equilibrium, i.e. the dynamic exchange between folded β-hairpin and unfolded conformations, was used to determine the

thermodynamic parameters describing a weak halogen bond. In Paper II the adaptive conformational switching between different conformers upon changing from polar to apolar environments is suggested as an approach to rationalize passive membrane permeability. The conformational analysis of macrocyclic erythronolide antibiotics in the beyond-Rule-of-5 space in D_2O and $CDCl_3$ solution revealed two distinct sets of conformational ensembles. In Paper III, conformational analysis was implemented to elucidate which diastereoisomer of the macrocyclic natural product senaetnine was isolated from *Senecio mannii*. The Configurational determination was enabled by the restrained conformational freedom created by the three stereocenters. Thus, only a very specific conformational ensemble could be adapted by the present diastereoisomer.

Since its first description in the late 1930s, NMR spectroscopy has risen to one of the most powerful methods for the structural analysis of molecules.[27-29] Within this thesis, I make use of NMR observable descriptors that provide significant information for determining the three-dimensional expansion of a molecule. In the following, a short introduction of the concept and application of the vicinal coupling constant ($^3J_{HH}$) and its relation to the dihedral angle as well as the nuclear Overhauser effect (NOE) and its relation to the interproton distance is given. The implementation of residual dipolar coupling (RDC) and its relation to C-H bond vector orientation in a magnetic field is described in Chapter 3.5.

1.5.   The Vicinal Coupling Constant ($^3J_{HH}$)

The scalar spin-spin coupling arises from indirect interactions between two spins, which is mediated by the electrons in the covalent bonds connecting the nuclei.[30] The scalar coupling describes the isotropic component of the observed coupling constant J, thus it is independent of the molecular orientation. It gives viable information about the connectivity within molecules, and also on molecular geometry. The magnitude of scalar coupling constants typically decrease with the number of bonds separating the coupling nuclei. Geometrically most important is the relationship of the coupling constant of protons three bonds apart ($^3J_{HH}$) with the dihedral angle between them. This phenomenon was first described by Martin Karplus in 1959 and his proclaimed relationship was later termed the Karplus equation.[31]

The Karplus Equation

In its original form, it describes the variation of $^3J_{HH}$ as a function of the dihedral angle:

$$^3J_{HH}(\phi) = A + B\cos\phi + C\cos 2\phi \tag{1}$$

where Φ (in radians) is the dihedral angle along a H−C−C−H bond and A, B and C are constants.[32] It allows the calculation of a coupling constant from the dihedral angle of a given conformer but the reverse is difficult. Taking an observed $^3J_{HH}$ and relate it to a dihedral angle typically results in multiple possible values (Figure 6). Therefore, the Karplus equation's primary use is to estimate the relative orientation of substituents around a single bond rather than to calculate an exact value. Karplus himself stated that "Certainly with our present knowledge, the person who attempts to estimate dihedral angles to an accuracy of one or two degrees does so at his own peril".[32] The original equation has seen continuous modifications and re-parametrizations to embrace a vast number of atom pairs in different molecular environments. Altona *et al.* developed two kinds of generalized Karplus equations based on empirical values and the inclusion of Huggins's relative electronegativities.[33, 34] Empiric parameters specifically determined for molecular classes like sugars, peptides, cyclic systems and others have further increased the accuracy of calculated coupling constants used for conformational analysis.[35]
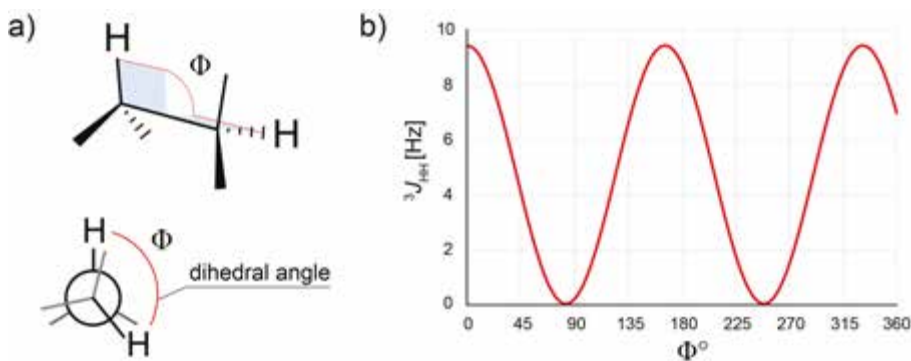

Figure 6: a) The dihedral angle (Φ) is the angle formed at the intersection of two different planes, here depicted in spatial orientation and in the Newman-projection. b) Dependence of the vicinal coupling constant ($^3J_{HH}$) on the dihedral angle. The depicted function was calculated applying equation 2.

The peptide bond is a ubiquitous part of peptides and proteins and resembles ideal properties in regards to stability and rotational restrain.[36] The dihedral angle of the H−N−C$_\alpha$−H bond (Φ) along a peptides backbone is a key factor in the determination of the geometrical propagation of a peptide or protein strand.[37] This bond has certain characteristics that promote the parametrization of a specialized Karplus equation. The O=C−N−C peptide bond can be assumed to be planar (either *cis* or *trans*) and the experimentally determined range $^3J_{NH,C\alpha H}$ for peptides is ~ 0–12 Hz.[38] Further inclusion of

electronegativity effects of all 20 amino acid side chains by Bystrov[39] led to a Karplus-type equation optimized for peptide backbones:

$$^3J_{NH,C\alpha H} = 9.4\,cos^2\phi - 1.1\cos\phi + 0.4 \qquad (2)$$

Equation 2 was utilized to calculate coupling constants from dihedral angles of theoretical conformers employed in the experimental dataset for conformational analysis of the decapeptides discussed in Paper I.

1.6.   The Nuclear Overhauser Effect (NOE)

Since Albert Overhauser proposed his ideas on dynamic nuclear polarization in 1953[40], the nuclear Overhauser effect has arguably become one of the most essential tools in NMR spectroscopy. It describes through-space interactions of nuclei via a dipole-dipole coupling. Most common is the NOE measurement via homonuclear ^{1}H-^{1}H NOESY experiments for the determination of interproton distances up to 5 Å apart.[41] When obtaining a sufficient coverage of NOEs between protons throughout the entire molecule, the fold and orientation of its skeleton can be elaborated. Like most NMR observables, it resembles a time-averaged parameter that contains both structural and dynamic information. Thus, the observed NOE is the population weighted average of the NOEs between two nuclei in all conformers present in a solution ensemble. The deconvolution of ensemble averaged NOE and $^3J_{HH}$ data will be described in Chapter 1.7. To obtain quantitative interproton distances from NOEs the origin and influences of this important phenomenon must be understood.

Spin Populations

The absolute intensity, or integral, of an NMR signal of a spin ½ nucleus such as a proton (^{1}H) is dictated by the population difference between its α and β spin states. The stronger the magnetic field of the spectrometer the larger this difference becomes. The Boltzmann equation describes the ratio between spin populations depending on the temperature (T), the gyromagnetic ratio of the spins (γ) and the magnetic field strength of the spectrometer (B_o):

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}} \qquad \text{with} \qquad \Delta E = \gamma\,B_o\left(\frac{h}{2\pi}\right) \qquad (3,4)$$

where N are the populations of spin states, k is the Boltzmann constant, ΔE is the energy difference between α and β spin states, and h is the Planck constant. In general, ΔE is very small so that the energy from thermal collisions is

sufficient to promote nuclei into a higher energy spin state (β). This is the driving force of adapting and maintaining the equilibrium between energetically different states. The ratio of the two spin states available to a proton (^1H) determines the magnitude of the macroscopic magnetization vector M and thus the signal intensity observed in an NMR spectrum.

Dipole-Dipole Coupling

The through-space coupling between two nuclear dipoles induces a distance-dependent effect of population change, which is observed as the NOE. To understand the nature of the NOE, one has to look at the relationships between spins (S and I) in an energy diagram. In the system at equilibrium, both spins reside in their mutual Boltzmann equilibriae (Figure 7a). In a steady-state NOE experiment spin S is saturated by applying a low-power irradiation pulse causing the α and β states to be equally populated (Figure 7b). Alternatively, spin S can be irradiated by a 180° pulse, resulting in an inversion of populations, which results in the transient NOE effect.
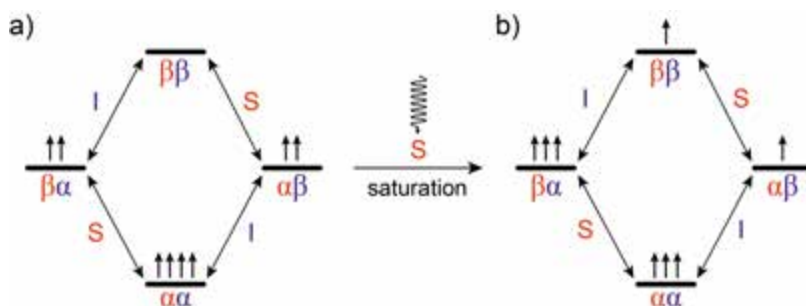

Figure 7: a) Boltzmann equilibrium of a two spin system at equilibrium. Spin S is depicted in red and spin I in blue. Lower energy states (α) are more populated relatively to the higher energy spin states (β). b) After a selective saturation pulse on spin S, its population differences are zero.

Directly after the saturation or inversion, the changed population levels will relax back to the equilibrium state via longitudinal (T_1) relaxation. The two processes, through which this can occur, are either auto-relaxation or cross-relaxation, both arise mostly from dipolar coupling. In case of cross-relaxation, the population level of spin I will be affected via two mechanisms, W_0 and W_2, the zero-quantum transition and double-quantum transition, respectively (Figure 8). The W_2 pathway involves two aligned spins to flip simultaneously giving rise to an increase in spin population difference for spin I (positive NOE). In case of W_0, two anti-parallel spins flip their orientation resulting in a reduction of spin differences for spin I (negative NOE).
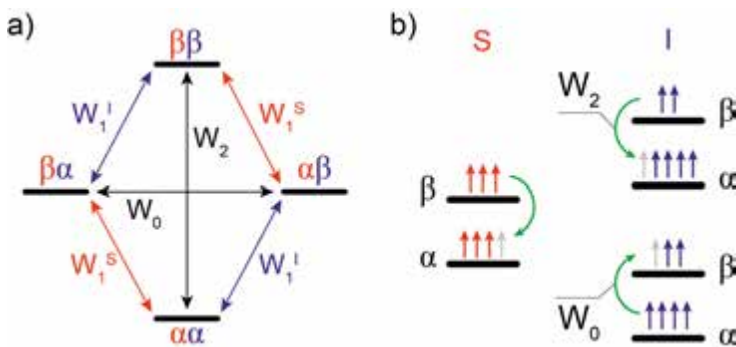
Figure 8: Origin of the NOE. a) Different pathways for population change of spin S and I. W_2 is the double-quantum transition where two aligned spins flip simultaneously and W_0, the zero-quantum transition, describes the concurrent flip of anti-parallel spins. b) During cross-relaxation, spin S regains its initial spin population. This influences the spin population of coupled spin I. W_2 increases the population difference of spin I whereas W_0 lessens it.

Whether a positive or a negative NOE is observed, is dictated by which of the two mechanisms (W_2 or W_0) is most predominant and is determined by the cross-relaxation rate (σ_{SI}).

$$\sigma_{SI} = (W_2 - W_0) \tag{5}$$

In order for spin transitions to happen, the resonance frequency has to match the energy difference of the two involved spin states. For W_2, the resonance frequency for spin transition (ω_2) is the sum of Larmor frequencies of both spins ($\omega_I + \omega_S$), whereas for W_0, ω_0 is the difference between the two ($\omega_I - \omega_S$). The molecular tumbling is decisive for which of the two mechanisms becomes more effective. Molecular motion induces an electromagnetic field via the rotation of dipolar spins within the structure. Molecules do not continuously tumble at the same rate but their size determines the maximum limit of their rotational frequencies (Figure 9). The average time required to rotate one radian ($\sim 60°$) around an axis is termed the correlation time, τ_c, which is mainly determined by the molecular weight. The τ_c for larger molecules is > 10 ns, whereas smaller molecules can range down to picoseconds. However, the correlation time is also influenced by other parameters such as temperature, solvent viscosity and spectrometer frequency (Figure 10). The relationship of correlation time and Larmor frequency on the cross relaxation pathway is described by the spectral density function.[42]

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \tag{6}$$

This indicates that slow molecular tumbling (long τ_c) generates an oscillating magnetic field with small high-frequency contributions. A short τ_c corresponds to a fast random rotation, causing a much wider frequency distribution with more high-frequency contributions. The spectral density is a measure of how much of the molecular motion is present at the correct frequency (ω_t) to cause spin transition (Figure 9).
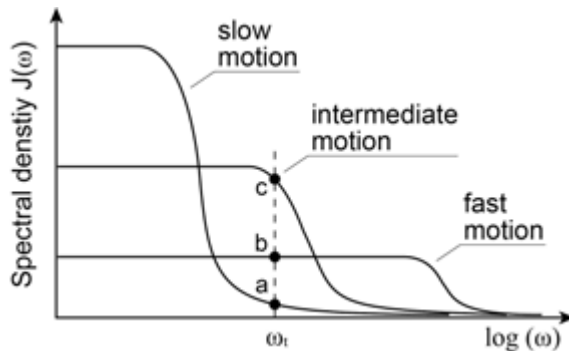

Figure 9: Spectral densities for molecules with different molecular motion. The frequency ω_t (Larmor frequency) corresponds to the energy required for spin transition of the different molecules. a) Slow motion promotes effective relaxation at low frequencies/energy differences, thus W_0 dominates for large molecules. b) Fast motion is effective for higher frequencies/energy differences, thus W_2 dominates for small molecules. c) Intermediate motion has frequencies where both relaxation pathways are equal, thus cancelling each other out.

The observed NOE can now be related to the cross-relaxation rate (σ) in dependence of the spectral densities at the resonance/transition frequencies for W_2 ($\omega_2 = 2\omega_0$) and W_0 ($\omega_0 \approx 0$).

$$\sigma_{SI} \propto \frac{1}{r_{SI}^6} \{6J(2\omega_0) - J(0)\} \tag{7}$$

When $6J(2\omega_0) > J(0)$, which requires a relatively high spectral density at $\omega = 2\omega_0$ related to a short τ_c, then W_2 dominates and a positive NOE is observed. This relates to small molecules in the extreme narrowing regime where $\omega^2\tau_c^2 << 1$. In the opposite case, when a long τ_c limits high frequency contributions, then W_0 is the more efficient cross-relaxation pathway leading to negative NOEs. A rough estimation is that molecules smaller than 600 Da will give a positive NOE and molecules larger than 1000 Da a negative NOE.[43] The crossing point is at $\omega_0\tau_c \approx 1.12$, where the effects from W_0 and W_2 cancel each other out (Figure 10).
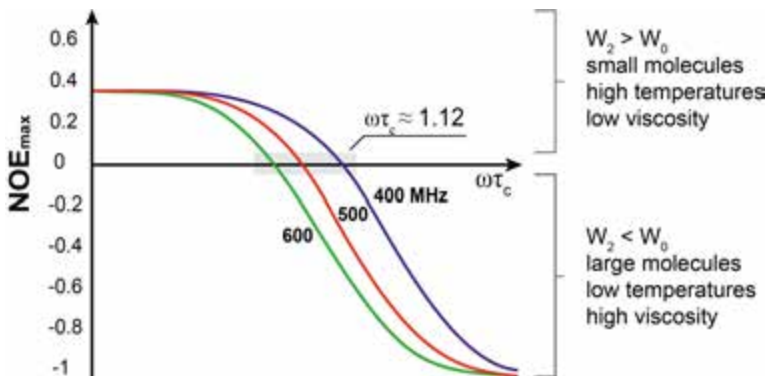
Figure 10: The maximum transient NOE enhancement given as a function of the correlation time, τ_c, of a molecule and the spectrometer frequency, ω. The path of the curves changes as the spectrometer frequency changes. At the zero-crossover section (grey box) no NOE cross-peaks can be observed in a NOESY spectrum. Further factors to whether positive, negative or no NOEs are observed are temperature and viscosity. Both influence the correlation time, with lower temperature or higher viscosity resulting in stronger negative NOEs.

From NOE to Interproton Distance

Assuming that the ratio of both mechanisms, W_0 and W_2, is the same throughout the entire molecule, then the intensity of the NOE is inversely proportional to the distance (r) between protons H_a and H_b.[30]

$$NOE_{ab} \propto \frac{1}{r_{ab}^6} \propto \sigma_{ab}^6 \tag{8}$$

The observed NOE decreases with distance (r^{-6}) that leads to a detection limit of 5-6 Å. The accuracy of NOE determination depends on the type of experiment (steady-state or transient), the resolution of the spectra, and the signal-to-noise ratio. The 2D version of a transient NOE experiment is termed NOESY. In comparison to the steady-state experiment, a NOESY experiment has the advantage that observed NOEs are symmetric. Additionally, the immediate perturbation occurring after a short inversion pulse allows for determination of the cross-relaxation rate. In the steady-state experiment spins are irradiated for a long time making it useless for the precise measurement of distances.[44]

A further increase in accuracy of r_{ab} determined from NOE intensities is obtained using either the initial-rate approximation[45] or the full-relaxation matrix method.[46] The latter is rarely used for small to medium-sized molecules, yet is popular for the analysis of large biomolecules that have rapid T_2 relaxation. The initial-rate approximation method requires determination of the NOE build-up between H_a and H_b (Figure 11). This is achieved by

recording a series of NOESY spectra varying the mixing time (t_{mix}) allowing the population change to take place. Common mixing times are set between 50 ms and 1 sec depending on the tumbling rate of the molecule under observation. If t_{mix} is too short, the NOE has little time to develop and no signal is observed; if chosen too long, other relaxation pathways will lead to a non-linear dependence of the NOE on the interatomic distance. However, in the onset of the NOE build-up, the increase is linear and solely depends on the population averaged distance between H_a and H_b. By measuring cross-peak integrals, i.e. intensities, in a series of NOESY experiments acquired with different t_{mix}, the intensity can be plotted as a function of the mixing time. The initial linear part of this curve is referred to as the initial-rate.
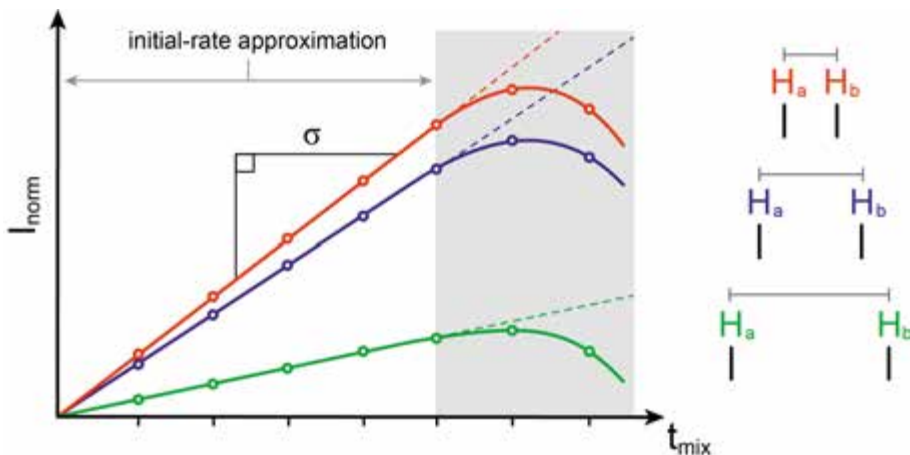

Figure 11: The NOE build-up curve is the normalized cross-peak intensity (I_{norm}) as a function of the mixing time (t_{mix}). In the initial part, the curve depends only on the dipole-dipole cross relaxation rate, σ. At longer mixing times the linearity gets lost when other relaxation effects start to disturb the NOE. The slope (σ) of the linear part is inversely proportional to the interproton distance r_{ab}^{-6}. A steeper σ (red line) corresponds to a shorter r_{ab} and a shallower σ (green line) to a longer r_{ab}.

By determining the slope of the initial build-up via linear regression analysis, random errors average out thus increasing the accuracy of the analysis. The slope (σ), reflects the NOE, and is inversely proportional to the distance r_{ab}^{-6} (Equation 8). It is recommended to normalize the intensities of diagonal- and cross-peaks by using peak amplitude normalization for improved cross-relaxation (PANIC)[47, 48] that restores the linearity of the build-up rate at longer mixing times. The normalized intensity (I_{norm}) is obtained following:

$$I_{norm} = \sqrt[2]{\frac{|cross\ peak_{ab} * cross\ peak_{ba}|}{|diagonal\ peak_a * diagonal\ peak_b|}} \qquad (9)$$

where $cross\ peak_{ab}$ and $cross\ peak_{ba}$ are the integrals of the two cross peaks on either side of the diagonal and $diagonal\ peak_a$ and $diagonal\ peak_b$ are integrals of the corresponding proton signals H_a and H_b respectively. I_{norm} is determined for all mixing times and subsequently plotted as a function of t_{mix} (Figure 11). A minimum of four normalized intensities of consecutive mixing times with $R^2 \geq 0.95$ linearity should be used to determine the build-up rate σ_{ab}. The build-up rate from a rigid pair of protons (σ_{ref}), and its known distance (r_{ref}) is taken as internal reference. Finally, the interproton distance (r_{ab}) can be derived from the buildup-rate (σ_{ab}) and the reference values using:

$$r_{ab} = r_{ref} * \left(\frac{\sigma_{ref}}{\sigma_{ab}}\right)^{-6} \tag{10}$$

Reference proton distances can either be a pair of geminal protons with 1.78 Å or a pair of *ortho* protons with 2.54 Å separation (Figure 12).[49] From the obtained interproton distances and scalar couplings, the conformation of a molecule in solution can be determined by the means of NMR data deconvolution using theoretical ensemble fitting.
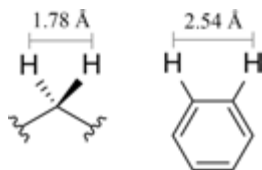

Figure 12: Reference distances (r_{ref}) from locked-in-space protons with a fixed and known interproton separation. Left: a pair of geminal protons. Right: a pair of *ortho* protons.

1.7.   Data Deconvolution

For molecules with one major conformation (due to dynamic restrain by rigidity or a network of secondary interactions), the structure can directly be determined by fulfilling all experimental restraints. Flexible and less restrained molecules however, exist as an ensemble of interchanging geometries, hence the observed signals (NOEs and $^3J_{HHS}$) reflect the average of all existing conformers. To determine the conformational ensemble of a molecule, data deconvolution driven by parallel weighing of theoretical structures is required. The term deconvolution denotes the reversal of a weighted averaging process or convolution.[50] In the following, the NMR data deconvolution method using the algorithm NAMFIS and a computationally generated theoretical ensemble is introduced.

The NAMFIS Algorithm

NAMFIS (NMR analysis of molecular flexibility in solution) was developed by D. Cicero *et al.* in the early 90s for studying the distribution of rapidly exchanging conformers of a 13-amino acid peptide. Since then, the algorithm aided the successful structure determination of numerous natural products,[51-53] peptides[54-58] and macrocycles[59-61]. NAMFIS matches the structural information from a large set of possible conformers (theoretical input ensemble) to a set of NOEs and scalar coupling constants to yield a conformational ensemble that represents the experimental data. From the theoretical input ensemble geometric information, namely the respective interproton distances (r_{calc}) and dihedral angles, are extracted. Individual energies of the conformers are not considered as these are force-field dependent. If NOE signals were observed for $-H$ to $-CH_2-$ or $-H$ to $-CH_3$, the corresponding distances in the theoretical ensemble are averaged between the two or three individual distances respectively

$$r_{calc,CH2} = \sqrt[-6]{\frac{(r_1^{-6} + r_2^{-6})}{2}} \tag{11}$$

$$r_{calc,CH3} = \sqrt[-6]{\frac{(r_1^{-6} + r_2^{-6} + r_3^{-6})}{3}} \tag{12}$$

Dihedral angles are converted to their related vicinal coupling constant ($^3J_{calc}$) as explained in Chapter 1.5. Since the experimental parameters (r_{exp} and $^3J_{exp}$) are averages, weighted over the molar fractions of the conformations present in solution, a population weighing factor (x_i) is introduced in order to match back-calculated parameters. The factor x_i must be non-negative and the sum of all values for x_i must be 1. Applying these rules, the population weighed averages are calculated following:

$$\langle r_{calc} \rangle = \sum_i x_i * r_{calc}(i) \tag{13}$$

$$\langle ^3J_{calc} \rangle = \sum_i x_i * {}^3J_{calc}(i) \tag{14}$$

The best match is obtained by a nonlinear least-squares fitting procedure that minimizes the sum-of-squares difference (SSD) between measured parameters and calculated variables:

$$SSD = \frac{1}{2}\left[\sum_{i=1}^{m}\left(\frac{\langle r_{calc}(i)\rangle - r_{exp}(i)}{r_{err}(i)}\right)^2 + \sum_{j=1}^{n}\left(\frac{\langle {}^3J_{calc}(j)\rangle - {}^3J_{exp}(j)}{{}^3J_{err}}\right)^2\right] \quad (15)$$

The error estimates r_{err} and ${}^3J_{err}$ are introduced to compensate for inaccuracies in NOE distance determination and the estimation of the Karplus equation respectively. ${}^3J_{err}$ is set constant with common values ranging from 1.5 - 3 Hz. r_{err} is staggered by distances since longer distances arise from weaker signals and thus are less reliable. Values used for r_{err} are summarized in Table 2.

Table 2: Error estimates used for experimentally determined interproton distances used in the NAMFIS analysis.[62]

NOE distance [Å]	Error estimate (r_{err})
$r_{exp} < 2.5$	0.1
$2.5 \leq r_{exp} < 3.5$	0.2
$3.5 \leq r_{exp}$	0.3

In order to be considered a feasible solution, the molar fractions and corresponding back-calculated averages ($\langle A_{calc}\rangle$) must fulfil the following constraints for any experimental parameter (A_{exp}) within the corresponding error (A_{err}):

$$A_{exp}(i) - A_{err}(i) \leq \langle A_{calc}(i)\rangle \leq A_{exp}(i) + A_{err}(i) \quad (16)$$

The final output of NAMFIS is a set of conformers with $x_i \geq 1\%$ that best fit the experimental NMR data where major fractions represent the more stable conformers. Thus, NAMFIS identifies a combination of conformers and their molar fractions, independently of their calculated energies. This gives valuable insight into the conformational preferences of a molecule in solution.

The Theoretical Input Ensemble

It is important to provide NAMFIS with a pool of conformers that covers the entire conformational space the molecule can adapt. If real conformers are missing in the input ensemble, NAMFIS is either unable to find a feasible solution within the given limits, or virtual conformers are weighted into the ensemble to mathematically reduce the deviation. Therefore a large energy window for sampled conformers is in favor for the completeness of the theoretical input ensemble.[63] Additionally the geometrical spread must be even within the given conformational space. The best method to achieve this unbiased conformational search, within an appropriate computation time, is a Monte-Carlo method.[64, 65] In this sampling approach, bonds are randomly rotated to generate new structures and are then locally minimized applying

force field parameters.[66] However, the chosen force field as well as the selected implicit solvation model can limit the sampling of real-life conformations. Thus, within this thesis, I used a set of different force fields and solvation models in combination with a large energy window (42 kJ/mol) to sample and save conformers respectively. To eliminate duplicates and ensure an even conformational distribution, a redundant conformer elimination with a RMSD cut-off at 0.5 – 2.7 Å was applied to the combined searches. Conformational searches were executed using the Monte-Carlo Multiple Minima (MCMM) method in torsional sampling mode as implemented in the MacroModel package of Schrödinger's Maestro software.[67] Final ensembles used in this thesis contained 20 – 237 unique conformers, depending on the size and flexibility of the investigated molecule.

2. Aim of Thesis

The overall goal of the work presented in this thesis is to perform ensemble determination by means of conformational analysis using solution NMR. A novel method for the thermodynamic and geometric characterization of a weak non-covalent interaction by the conformational analysis of a flexible peptide model system is demonstrated. Additionally, the impact of solvent polarity on the conformational distribution of macrocyclic drugs is investigated and correlated to their augmented cell permeability and solubility. Finally, a molecule's relative configuration is determined by comparing the conformational space specific for the present diastereoisomer with experimental data. In order to obtain constructive data and gain viable insight into the capability of conformational analysis the following projects were pursued:

- Design and synthesis of a β-hairpin forming peptide as a model system to assess the strength and geometry of a weak intramolecular halogen bond incorporated into the system.

- Thermodynamic analysis of the peptidic model system to quantify the improvement in structure stability upon introduction of the halogen bond.

- Conformational analysis of the model system by combining NOEs, ^3Js and RDCs to accurately determine the halogen bond's geometry.

- DFT guided design and synthesis of halogen bond donor groups onto amino acid building blocks to elevate the versatility of halogen bonding in biological systems.

- Ensemble analysis of a set of macrocyclic drugs in different solvents, thus revealing the adaptive behavior upon polarity change hypothesized for beyond-Rule-of-5 drugs.

- Comparison of theoretical ensembles generated for all diastereoisomers of a macrocyclic natural product with a set of interproton distances based on NOEs to determine its relative configuration.

3. A Weak Halogen Bond Characterized by Conformational Analysis (Paper I)

Since its first observation in 1814,[68] the halogen bond (XB) has evolved into a useful tool for molecular recognition in organic synthesis,[25, 69, 70] structural biology[71-73] and drug discovery.[74, 75] Just recently, in 2013, it was given an official definition by IUPAC, describing it merely as the net attractive interaction of an electrophilic region associated with a halogen and a nucleophile.[20, 76] General descriptors regarding the geometry of a halogen bond in literature are derived from crystal structures or from computational simulations.[23, 77, 78] Furthermore, weak halogen bonds of flexible and exchanging systems in the solution phase are very difficult to detect with current techniques. The molar fraction of a bimolecular complex between a halogen bond acceptor binding to a halogen bond donor in solution is typically too small to be detected. This leaves a scientific gap where the characteristics of weak halogen bonds in solution, so far, are not described. To demonstrate the difficulty in observing the formation of a weak halogen bond, I titrated 0.5 – 20 eq. of diethylether (the XB acceptor), into a solution of 20 mM iodobenzene (the XB donor) in $CDCl_3$. At this concentration, no chemical shift change was observed in ^{13}C NMR, and shifts in the ^{1}H NMR extended below 0.02 ppm at a 20-fold excess of the XB acceptor. These minor shift changes are expected for a bimolecular system that is predicted to form a weak C−I···O halogen bond. The entropic penalty and the molecular motion of the solution state are detrimental for an observable fraction of the halogen bond complex. Hence, I developed a model system that includes the XB pair as an intramolecular interaction that contributes to the stability of a secondary structure, namely a β-hairpin. This system provides an entropic advantage as the overall shape of the β-hairpin is stabilized by a series of cooperatively acting non-covalent interactions, including the XB itself. Herein I present the design, synthesis and conformational analysis of a β-hairpin as a method to probe a weak XB in solution.

3.1. Design of a β-Hairpin Model System

β-Hairpins are a ubiquitous partial structure within proteins that are defined by two hydrogen bonded antiparallel β-strands connected by a short loop.[79]

Despite its high abundance in folded proteins, they are prone to lose their organized structure when excised from the protein. Thus, when designing a sequence to fold into an isolated β-hairpin in solution, a specific peptide sequence for the β-strand and loop region is required.[80-82] Most important is the structural integrity of the loop region that initiates the proximity, backbone hydrogen bonding and antiparallel alignment between the two β-strands. Examples for β-turn templates are widely described in the literature aiding the intentional design of a β-hairpin forming peptide sequence.[80, 83-85] Upon successful design of a sequence that folds into a β-hairpin, it provides an excellent platform for conformational and thermodynamic analysis of its contributing non-covalent interactions. This chapter focusses on the development of a β-hairpin sequence that inherits the sidechain-sidechain cross-strand halogen bond of *meta*-iodo-phenylalanine with the oxygen of homo-methyl-serine.
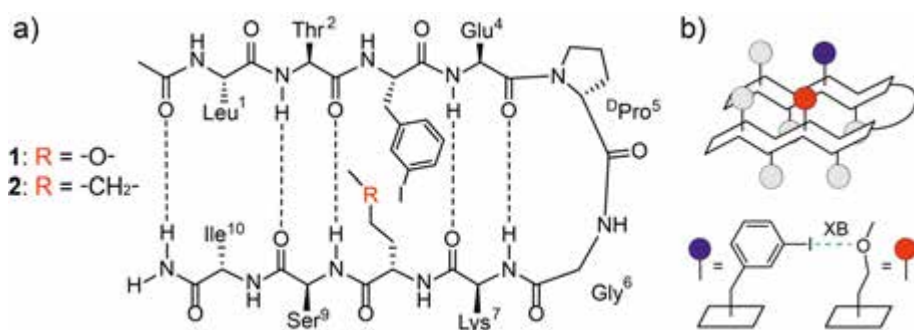

Figure 13: a) Compounds **1** and **2**, depicted as folded, antiparallel β-hairpin. b) Iterative positioning of amino acid sidechains above and below the β-hairpin plane. Residues marked in blue and red depict the XB interaction site.

The final design of the decapeptide resembles around a central ^{D}Pro^5-Gly^6 β-turn motif (Figure 13a). Gellman *et al.* demonstrated in various model peptides that this loop sequence is favorable for inducing a β-hairpin conformation in several solvents.[85-87] Adjacent to the turn, Glu^4 and Lys^7 form the first cross-strand hydrogen bond as well as a sidechain salt bridge of the -COO$^-$ and -NH$_3^+$. In peptide **1**, this pair is followed by the halogen bonding site consisting of 4-iodo-phenylalanine (Phe(I)^3) and homo-methyl-serine (Hse^8). The XB site sits at a position in the β-hairpin where no cross-strand hydrogen bonds are formed. This increases the impact of the XB formation on the fold propagation and overall stability of the hairpin. In position ^{2} and ^{9}, Thr and Ser are placed to increase polarity, solubility and inhibit peptide aggregation. The terminal groups are two β-branched amino acids, Leu^1 and Ile^{10}. Acetylation of the N-terminus introduces an additional carbonyl oxygen to form the 5th cross-strand hydrogen bond in the complete β-hairpin. The chirality of all amino acids induce an iterative orientation of sidechains above and below the

formed β-sheets plane. This orients the XB pair on one site of the plane together with only the two β-branched amino acids and thus avoiding interactions with other polar sidechains (Figure 13b). Selection of the amino acid sequence was guided by literature examples and Monte-Carlo molecular searches. The system was designed to reside at the inflection point between the fully folded and unfolded state at room temperature. Thus allowing for thermodynamic analysis of folding parameters and hence the halogen bond strength, using variable temperature NMR. For optimal quantification of the thermodynamic properties of the halogen bond, a reference, compound **2**, was essential. It resembles peptide **1** throughout but the ability to form a cross-strand XB. Thus the difference in the folding equilibrium of **1** and **2**, originates only from the intramolecular halogen bond in **1**. All other interactions contributing to the cooperative folding are identical.

3.2. Synthesis

The synthesis of compounds **1** and **2** was accomplished using solid-phase peptide synthesis (SPPS) starting with a Rink-amide resin as the solid support.[88] Syntheses were performed on a 150 µmol scale with a resin loading of 0.53 mmol/g. The first Fmoc-protected amino acid (Fmoc-Ile-OH) was anchored to the resin using a 2-fold excess of amino acid, 1.8 eq. of HCTU as the coupling agent and 4 eq. of DIPEA as a base (Scheme 1). A Kaiser-test was performed to verify sufficient coupling.[89] Consecutive amino acids were coupled using an automated peptide synthesizer following the Fmoc-protocol.[90] For compound **1**, Fmoc-Hms-OH (Fmoc-homo-methyl-serine) and for compound **2** Fmoc-Nle-OH (Fmoc-norleucine) were introduced at position [8]. After coupling of the N-terminal amino acid (Fmoc-Leu-OH) and manual deprotection using 20% piperidine in DMF, the free amine was acetylated using Ac_2O (30 eq.) and DIPEA (30 eq.) in DMF. Cleavage from the resin as well as deprotection of all side chain protecting groups was achieved by treatment with a mixture of TFA, H_2O and TIS (95:2.5:2.5, v/v). The free peptide was precipitated using cold diethylether, centrifuged and the supernatant was decanted. The remains were taken up in a mixture of H_2O:MeCN (7:3) followed by lyophilisation. Purification by prep HPLC on a C8-column provided compound **1** and **2** in 22% and 23% overall yields respectively. A detailed description of the synthesis can be found in the supporting information of Paper I.[91]
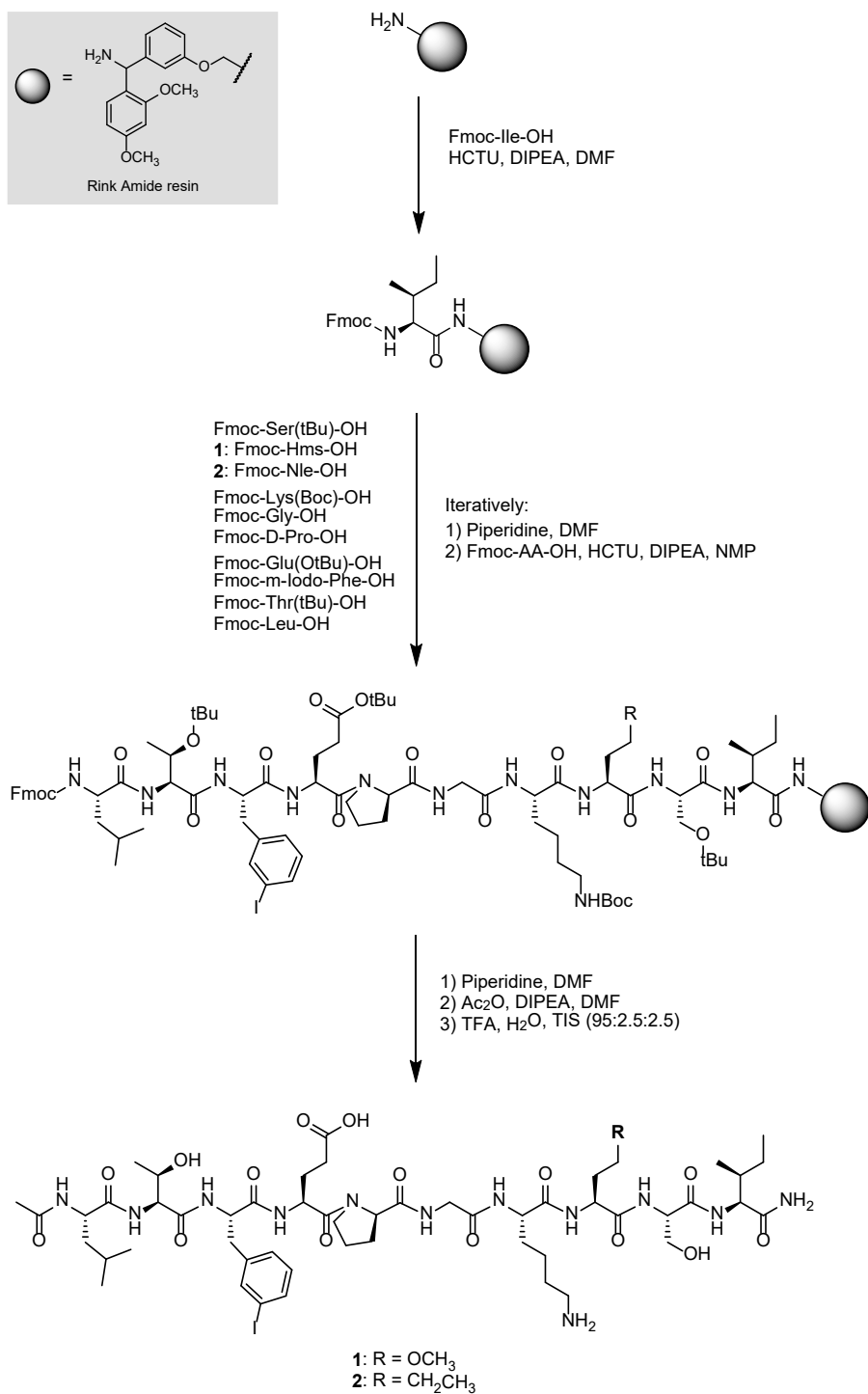
Scheme 1: Synthesis of compounds **1** and **2** on solid support. HCTU (O-(1H-6-Chlo-robenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate)

3.3. Thermodynamic Analysis

Absolute and relative stability of peptides **1** and **2** were determined following the $C_\alpha H$ chemical shift change in variable temperature NMR experiments over a 130 K temperature range. To reach detection at this extensive temperature range (218 K to 348 K), a solvent mixture of CD_2Cl_2:DMSO-d_6 (4:1) was used. ^1H shifts of **1** and **2** were assigned following the sequential resonance assignment strategy, combining NOESY and TOCSY experiments as described by Wüthrich *et al.*[92] All spectra recorded at varying temperatures were referenced to the transmitter nucleus (^1H) frequency. This avoids errors in referencing introduced by a temperature dependent reference shift (solvent residual peak or TMS). The observed changes in chemical shift for all $C_\alpha Hs$ for peptides **1** and **2** were fitted to a function describing a two-state equilibrium described by Munekata *et al.*[93]

$$\delta_{obs} = \delta_U + \left\{ \frac{\delta_F - \delta_U}{1 + exp\left[-\frac{\Delta H_m}{R} * \left(\frac{1}{T} - \frac{1}{T_m}\right)\right]} \right\} \tag{17}$$

Where δ_F and δ_U are the chemical shifts for the fully folded and unfolded structures, respectively. ΔH_m is the enthalpy change upon unfolding at the inflection point T_m. Fitting individual $C_\alpha H$ shifts provided a good estimation of their respective δ_U, however, were unreliable for the fully folded shift, δ_F. Hence, the folding curve was described by combining shifts through first taking the shift rate of individual $C_\alpha Hs$ followed by rescaling applying a z-score normalization.[94, 95]

$$\delta_{norm}^T = \frac{\left(\sqrt[2]{(\delta_{obs} - \delta_U)^2}\right) - \overline{\delta_{obs}}}{\sigma} \tag{18}$$

Where $\overline{\delta_{obs}}$ is the mean of experimental shifts and σ is the standard deviation. Since the normalized shift values reflect the folding behavior within the same range, an overall melting curve for peptide **1** and **2** could be deduced from the arithmetic means at each temperature.

$$\delta_{av}^T = \frac{1}{n} \sum_{i=1}^{n} \delta_{i,norm}^T \tag{19}$$

Where *n* is the number of amino acids $C_\alpha H$ shifts followed during temperature change. Curve fitting, using Equation 17 on the combined shift values applying a Levenberg-Marquardt least-squares protocol, provided the full two-state equilibrium melting curves for **1** and **2** (Figure 14). Additionally the change

in enthalpy and absolute melting temperature, i.e. the inflection point of the sigmoidal curve, were obtained (Table 3).
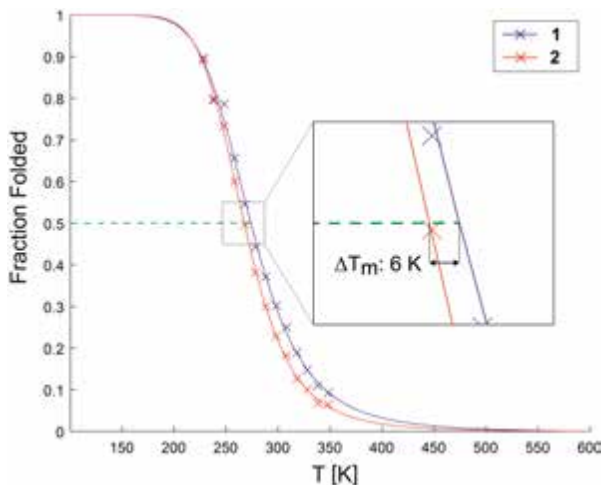

Figure 14: Overlap of the normalized melting curves of peptide **1** (blue) and the reference peptide **2** (red).

Table 3: Absolute thermodynamic constants determined from the melting curves. The melting temperature, T_m, and the change in enthalpy upon unfolding, ΔH_m, resulted from the fitting procedure. $\Delta G°$, the change of Gibbs free energy at 298 K, was calculated using Equation 20.

	T_m [K]	ΔH_m [kJ/mol]	$\Delta G°$ [kJ/mol]
1	274.0 ± 2.6	23.9 ± 3.3	2.1 ± 0.2
2	267.6 ± 1.3	26.4 ± 1.7	3.0 ± 0.1

From the thereof obtained endpoints of the melting curves, δ_F and δ_U, the Gibbs free energies at standard temperature could reliably be calculated.

$$\Delta G° = -RT * ln\left(\frac{\delta_U - \delta_{298}}{\delta_{298} - \delta_F}\right) \qquad (20)$$

Thereby the difference in energy, $\Delta\Delta G°$, between peptide **1** and **2** could be estimated to -0.9 kJ/mol. Since the possibility of forming a weak halogen bond between an aryl-iodine and an ether oxygen is the sole structural difference between the two peptides, this gain in thermodynamic stability accounts directly for the strength of the halogen bond. The herewith estimated binding energy of <0.9 kJ/mol is in line with the expected halogen bond strength. This is the first example where overall $C_\alpha Hs$ shifts were combined, revealing such

a minor change in stability. All fitting procedures and calculations of thermodynamic constants were achieved by a MatLab® script developed *ad hoc*, which I made available online.[96] However, the above approach compares the difference in thermal stability based on shifts related to absolute temperatures. The energy difference from a relative comparison of both peptides using only NMR parameters can provide a more precise value. Therefore, I investigated the two corresponding folding constants for **1** (k_F^1) and **2** (k_F^2) as the relative ratio of folding, $K_F^{1/2}$.

$$K_F^{1/2} = \frac{k_F^1}{k_F^2} = \frac{\left(\delta_U^1 - \overline{\delta_{obs}^1}\right)}{\left(\overline{\delta_{obs}^1} - \delta_F^1\right)} \bigg/ \frac{\left(\delta_U^2 - \overline{\delta_{obs}^2}\right)}{\left(\overline{\delta_{obs}^2} - \delta_F^2\right)} \tag{21}$$

When plotting $\left(\overline{\delta_{obs}^1} - \delta_F^1\right)/\left(\delta_U^2 - \overline{\delta_{obs}^2}\right)$ against $\left(\delta_U^1 - \overline{\delta_{obs}^1}\right)/\left(\overline{\delta_{obs}^2} - \delta_F^2\right)$, the ratio $K_F^{1/2}$ can be deduced from the slope of points assuming a linear relationship. The plot appears non-linear due to a difference in ΔH leading to a 'surfboard'-shape plot. However, since the shape is symmetric around the zero point and the maximum value, that is the difference of δ_{obs} at the inflection point of **2**; the slope is reliable determined by a linear fit.
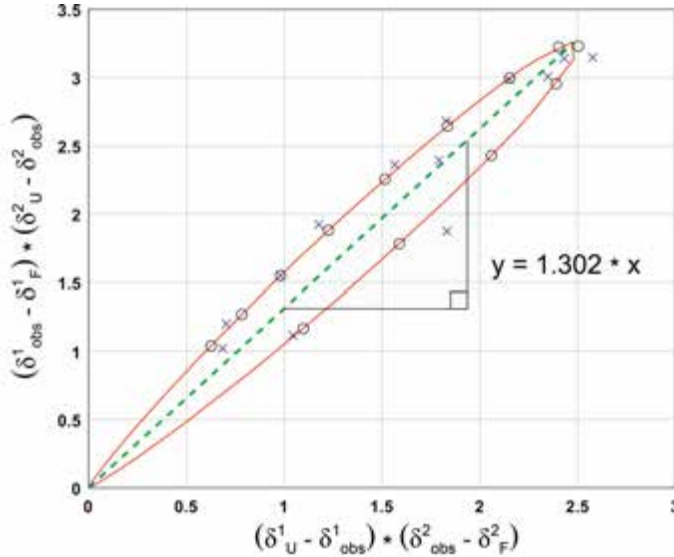

Figure 15: Relative folding ratio of **1** to **2**. The red line represents the calculated values obtained from curve fitting extended to reach the upper and lower plateau. Black circles represent fitted melting curve data points within the experimental temperature range. Blue crosses depict the $\overline{\delta_{obs}}$ values resulting from z-score rescaling and averaging experimentally determined shift changes.

The resulting relative ratio $K_F^{1/2}$ of ≈ 1.3 indicates peptide **1** to have a 30% higher folding ratio as compared to the reference **2**. The difference in stability can further be quantified as the $\Delta\Delta G$ at the melting temperature of **2**, T_m^2, and the obtained relative folding rate.

$$\Delta\Delta G = -R * T_m^2 * ln\left(K_F^{1/2}\right) \qquad (22)$$

The thus obtained $\Delta\Delta G$ = -0.6 kJ/mol is in good agreement with the value for $\Delta G°$ as obtained using the method of Munekata *et al.*[93] Since the method using the relative ratio of folding rate only relies on chemical shifts and is independent of the absolute temperature it is expected to give a more accurate result. A complete deduction of all equations used can be found in the supporting information of paper I.

The design of a linear, non-cyclic peptide inheriting a stable secondary structure at low temperatures is crucial for adequate determination of the melting curve. The herein described peptidic system provides excellent folding properties and little signal overlap to follow and analyze its thermodynamic behavior. This makes it ideal for the quantification of non-covalent interactions such as a weak I···O halogen bond. Furthermore, its sidechain diversity and a magnitude of ~45% fold at 298 K, according to the obtained melting curves, offers an excellent system to perform conformational ensemble analysis.

3.4. Secondary Structure Determination

To detect the influence of the additional weak XB on β-hairpin formation of **1** as compared to **2** in solution, conformational ensemble analysis based on NOEs and *J*-couplings with data deconvolution using NAMFIS, was performed. NMR spectra were recorded in the same solvent system as for the thermodynamic analysis (CD$_2$Cl$_2$:DMSO-d_6, 4:1) at a concentration of 3 mM. $^3J_{NH,C\alpha H}$ coupling constants were directly deduced from NH signals in the ^{1}H NMR spectra. To achieve quantitative inter-proton distances from NOE integration, a set of seven ^{1}H,^{1}H NOESY experiments were run with increasing mixing times ranging from 100 ms to 700 ms. Spectra were recorded in random order to prevent systematic errors. The recycle relaxation delay (d1) was set to 2.5 sec to attain sufficient T_1 relaxation. Interproton distances were deduced form the NOE build-up slopes as described in Chapter 1.6. Experimental data obtained for **1** and **2** showing the coverage of NOE based interproton distances and deduced *J*-coupling constants of the backbone are depicted in Figure 16.
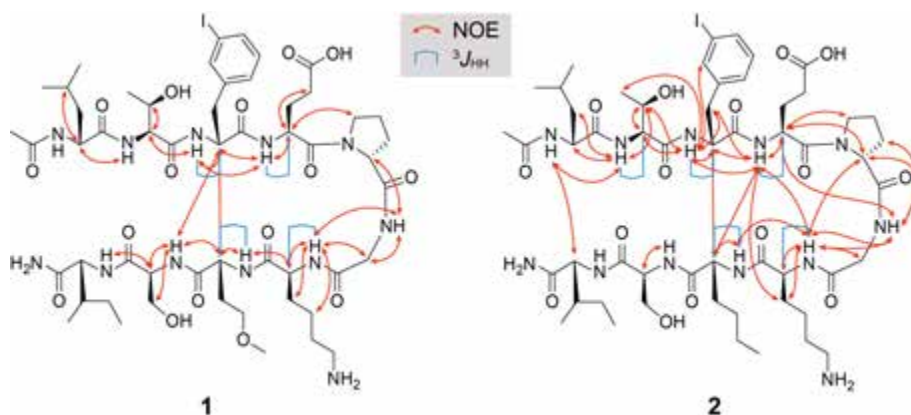
Figure 16: Depiction of interproton distances deduces from NOEs (red) and backbone vicinal coupling constants ($^3J_{HHS}$, blue) for the halogen bonding peptide **1** and the reference peptide **2** used in the NAMFIS analysis.

Theoretical ensembles for **1** and **2** were computed using the Monte-Carlo multiple minima (MCMM) method with three different force fields (OPLS3e, OPLS and AMBER*) in combination with two GB/SA solvation models (Water and Chloroform) using MacroModel as implemented in the Schrödinger Maestro software package (v. 13.0.135). In 50.000 torsional sampling steps new conformers were generated, minimized and saved within an energy window of 42 kJ/mol from the global minimum. The resulting conformers of all different conformational searches were pooled and a redundant conformer elimination with a heavy-atom cutoff at 2.7 Å was performed. Since no force field currently available in Maestro is parametrized for halogen bonds, the obtained conformers for compound **1** were enriched with geometries in which the I···O distance is at 90% of their van der Waals radii. The final ensembles contained 236 and 176 conformers for **1** and **2** respectively.

The deconvolution algorithm NAMFIS (Chapter 1.7) was implemented to reveal the set of conformers present with their according relative probabilities (molar fractions). Results are given as the best-fit solution that is the set of conformers of which the back-calculated averages deviate least from obtained experimental values. The solution set contained nine conformers for **1** and 11 for **2**. The individual conformers were analyzed by determining whether the backbone forms a β-hairpin. Threshold criteria for β-hairpin classification included type II' β-turn formation based on Φ- and Ψ- angles of turn amino acids, inter-strand distances and a Ramachandran plot analysis (for a detailed analysis see supporting information of Paper I). Thereby, four β-hairpin conformers with a combined molar fraction of 56% for peptide **1** were identified. In contrast, the solution ensemble of peptide **2** only contained three β-hairpin conformers at a sum of 39%. The 17% increase in β-hairpin

formation corroborates the increased stability found by thermodynamic analysis. Since the conformational variability of the sidechains is vast, and coverage with NOEs is sparse, reliable conclusion on conformations can only be drawn for the peptides backbone. However, results from NAMFIS demonstrate the systems optimal susceptibility for minor structural changes since the folded fraction is close to the inflection point at room temperature. The most abundant conformer (24%) in the solution ensemble of **1** exhibits a β-hairpin and is depicted in Figure 17.
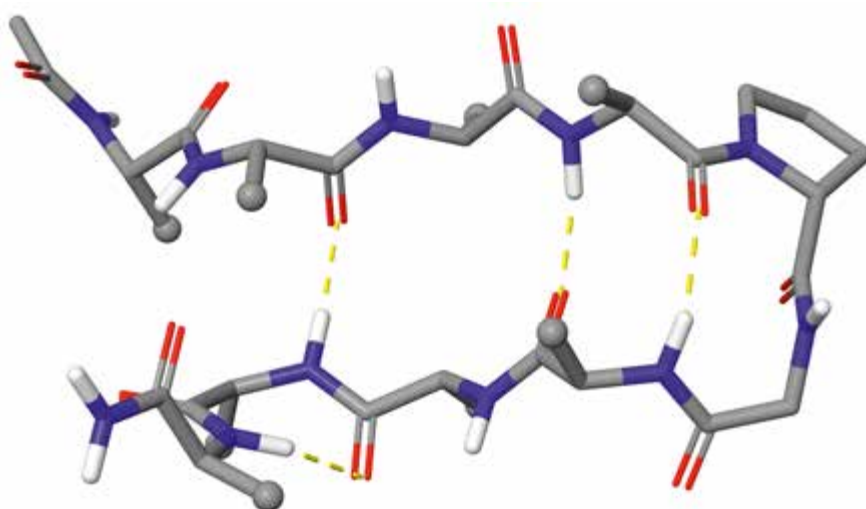
Figure 17: Conformer with highest abundance (24%) in the solution ensemble of **1** obtained from NAMFIS analysis. Sidechains are omitted for clarity.

To this point, the thermodynamic and backbone secondary structure analysis provided indirect indication for the formation of a halogen bond and thereof influencing the formation and degree of β-hairpin fold. To gain direct evidence for the correct orientation of the XB donator sidechain to the XB acceptor and an I···O interatomic distance that promotes halogen bonding, I performed RDC analysis of peptide **1**.

3.5.   Residual Dipolar Coupling (RDC) Analysis

Determining and using dipolar couplings for structure elucidation has a long history in solid-state NMR. During the last twenty years, residual dipolar coupling (RDC) has also become part of the toolbox of solution-state NMR, aided by the development of alignment media.[97-99] Commonly, a dissolved molecule faces no rotational restraints and tumbles freely in an isotropic solution. The fast molecular motion thus averages orientation-dependent

spectroscopic interactions, namely, quadrupolar coupling and dipolar coupling. Their effect is simply observed as their contribution to relaxation leading, for example, to a change in signal intensity observed as NOE. Upon introduction of a minor degree of alignment, i.e. a non-uniform restraint of a molecule's rotation, a scaled-down dipolar coupling is visible. The residual dipolar coupling is observed as a measure of Hz, whereas the real dipolar coupling is of a magnitude of kHz. However, the RDC contains the same structural information while line broadening and conformational freedom is not affected. Alignment of a molecule in solution can be obtained in two different ways: either by the use of an alignment medium or by introduction of paramagnetic entities either covalently bound to a molecule or as a binding ligand. In this thesis I used the alignment medium PBLG (poly-γ-benzyl-L-glutamate) that, above its critical concentration in solution (~130 mg/mL), forms a lyotropic liquid crystal phase. Its organization in helical columns throughout the sample induces anisotropic conditions with a global alignment of 0.1% - 1%. RDC has the major advantage to be of non-local character. This means that, in contrast to the NOE, the RDC gives orientations of bond vectors in respect to the outer magnetic field, B_0, and is independent of the relative position and orientation of other atoms in the molecule. Thus, it allows for the accurate characterization of sidechain geometries in peptide **1** that otherwise would be cumbersome using NOEs or J-couplings. In a ^{1}H, ^{13}C coupled HSQC recorded on an anisotropic sample, the observed coupling constant consists of the $^1J_{CH}$ scalar coupling and the residual dipolar coupling, $^1D_{CH}$. In an isotropic sample, the dipolar coupling is zero and thus, the RDC can be deduced as the difference to the total coupling, $^1T_{CH}$, observed under anisotropic conditions as in Equation 23.

$$^1D_{CH} = {}^1T_{CH} - {}^1J_{CH} \tag{23}$$

It is important to understand that a dipolar coupling can be also negative. A vector perfectly aligning with the outer magnetic field will have the largest positive value and a vector perpendicular to it a maximum of negative half the parallel vector's dipolar coupling. At the crossover point of 54.74°, referred to as the magic angle, the dipolar coupling is zero.[100] Thus, a zero value obtained for RDC measurements still possesses an angular information and should not be neglected.
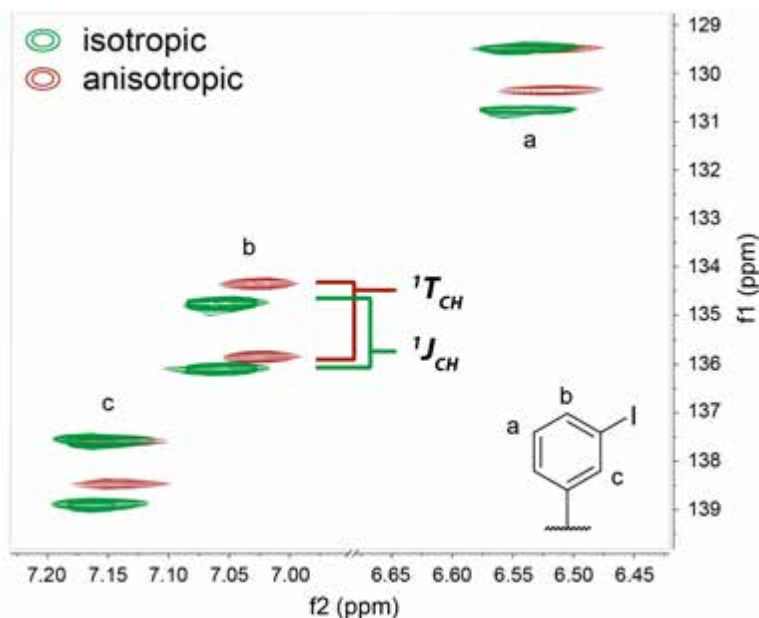
Figure 18: Overlap of the ^{1}H, ^{13}C coupled HSQC recorded under isotropic (green) and anisotropic (red) conditions of peptide **1**. Shown are the three signals of the iodo-phenylalanine residue corresponding to the letters assigned to the phenyl moiety. The coupling was measured in the indirect (f1, ^{13}C) dimension, thus increasing the signal-to-noise ratio and avoiding line broadening due to proton-proton coupling. The RDC was calculated as the difference of the total coupling ($^1T_{CH}$) in anisotropic conditions and the $^1J_{CH}$ coupling measured in isotropic conditions.

The anisotropic sample of **1** was prepared by weighing 80 mg of PBLG into an NMR tube and subsequently adding a solution of the peptide dissolved in CD_2Cl_2:DMSO-d_6 (4:1). The tube was shaken until all PBLG has dissolved followed by centrifugation of the tube. Uniform alignment was confirmed by the detection of a clear quadrupolar splitting ($\Delta\nu_Q \sim 125$ Hz) for the DMSO-d_6 signal in a ^{2}H NMR spectrum. The isotropic sample was prepared with the same concentration of **1**, but without the alignment medium. A Perfect-CLIP-^{1}H, ^{13}C-HSQC experiment[101] was run for both samples with the splitting detected in the indirect f1-dimesion. Peak asymmetry and broadening of multiple bond splitting in the proton dimension were avoided, affording an increased signal-to-noise ratio. To obtain the spatial arrangement and distance of the XB donor and acceptor site in **1**, all C-H RDCs for iodo-Phe[3] and homo-met-Ser[8] were deduced. Additional backbone C_α-H_α couplings from the turn region were determined to ensure accuracy of the alignment tensor fitting. RDCs were calculated as given in Equation 23, and the resulting values ranged from -68 to 58 Hz (Table 4). The bond vectors and their corresponding assignment are shown in Figure 19.

Table 4: RDC values measured as the difference (RDC = T – J) of the anisotropic C-H coupling constant to the isotropic sample of **1**. [a] represents the C-H vector *para* with respect to the iodine [b] represents the C-H vector *ortho* with respect to the iodine. [*] Values are from a C-H$_n$ group (n > 1); obtained RDC values were divided by n.

	C-H vector		Group	Isotropic	Anisotropic	T - J	
	Residue			$^1J_{CH}$	$^1T_{CH}$	RDC	Error
1	Phe(I)	α	CH	141.4	128.0	-13.4	4.20
2		β	CH$_2$[*]	260.2	127.5	-67.9	2.64
3		δ^a	CH	160.2	120.3	-39.9	1.61
4		δ^b	CH	165.3	109.7	-55.6	1.38
5		ε	CH	158.6	107.7	-50.9	0.94
6		ζ	CH	167.1	188.1	21.1	1.29
7	Glu	α	CH	138.9	196.4	57.5	2.37
8	Pro	α	CH	145.9	147.7	1.8	1.56
9	Gly	α	CH$_2$[*]	280.6	292.6	6.1	0.95
10	hmSer	γ	CH$_2$[*]	283.9	286.8	1.5	1.14
11		ε	CH$_3$[*]	416.6	434.7	6.0	0.23
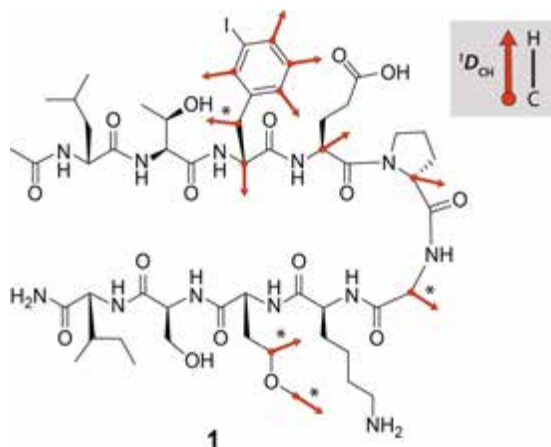
Figure 19: Depiction of C-H bond vectors (red arrows) used for RDC analysis. Vectors marked with an asterisk for CH2 and CH$_3$ groups are taken as the half- and third-sum of their originally detected dipolar coupling constants respectively.

A second theoretical ensemble was generated based on the nine backbone conformers identified by NAMFIS. Sidechain orientations were resampled, including halogen bonded ones, for all conformers individually, applying Monte-Carlo conformational sampling while retaining the peptide's backbone dihedral angles. Experimental data deconvolution resulting in a weighted set of conformers was achieved using the software MSpin. The back-calculation requires determination of the molecular alignment tensor by solving a series of linear equations in the least-square sense through the singular value deconvolution (SVD) method. For a reliable description of the alignment

tensor, a minimum of five non-parallel bond vectors must be provided. The SVD condition number obtained for peptide **1** was 4.028. Literature procedures suggest a SVD number <30 to indicate a robust fit.[102] From the back-calculated RDCs, the goodness-of-fit descriptor Q (Cornilescu's quality-factor), can be readily calculated (Equation 24). The closer to zero the Q factor, the better the fit of weighted conformers to the experimental data.

$$Q = \sqrt[2]{\frac{\sum \left(D_i^{exp} - D_j^{calc}\right)^2}{\sum D_i^{exp}}} \tag{24}$$

The obtained conformer ensemble converged remarkably well with the experimental data, as indicated by a Q factor of 0.004. From the theoretical ensemble that only contained structural information, MSpin independently weighted β-hairpin conformers at 53%. This in excellent agreement with the outcomes from both thermodynamic and NAMFIS analyses. Furthermore, and more importantly, the RDC analysis provided direct proof for the formation of a cross-strand halogen bond: 3 of the 5 conformers selected by MSpin have halogen bond geometries where one is in a β-hairpin conformation, and is present to 30%, representing 57% of the entire β-hairpin abundance. Non-hairpin conformers that display halogen bond formation are present to 45%. This observation leads to the hypothesis that the halogen bond is primarily involved in the folding process and stabilizes the β-hairpin foldamer. From the wide variety of XB donor-acceptor orientations provided in the theoretical ensemble, selected XB geometries showed a C−I···O angle of 145°, 171° and 178° (Figure 20). This bimodal distribution of the preferred XB angle of iodine is congruent with the findings of Ho *et al.*, who analysed XB geometries in a large dataset of protein crystal structures.[72]
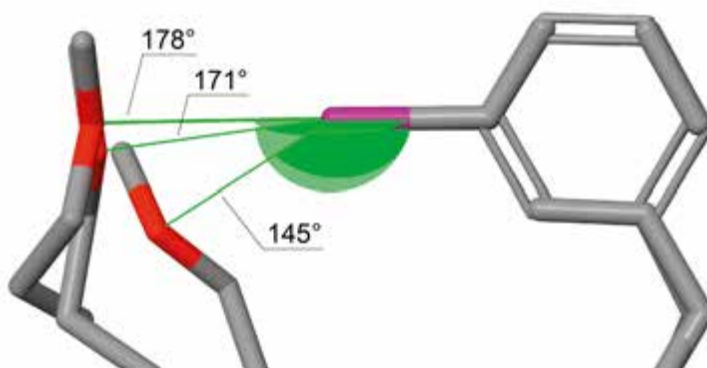

Figure 20: C−I···O bond angles of formed halogen bonds of **1** resulting from RDC analysis. The conformer with bond angle 145° is present to 30%, with 171° to 11% and 178° to 34%.

3.6. Summary

Herein I described the successful design, synthesis and evaluation of a peptide sequence inheriting a cross-strand C−I···O halogen bond. The system folds into a β-hairpin to an extent of ~45% at room temperature and is thus at its most susceptible equilibrium for internal changes affecting the stability of the secondary structure. A reference peptide that resembles the halogen bonding system, yet lacking the halogen bond acceptor, showed a significant decrease in stability. Thermodynamic analysis quantified the gain in stability as the strength of the intramolecular halogen bond to < 0.9 kJ/mol. This highlights the sensitivity of the developed model system to expose a single and weak stabilizing factor. Ensemble analysis using NOEs and ^3Js showed the molar fraction of β-hairpin formation to be 17% higher for the halogen bonding peptide as compared to the reference. By the use of an alignment medium, RDC analysis could further characterize the geometry and population of the formed halogen bond. The combined use of multiple NMR observables, as presented here for the first time, demonstrates the potential precision of conformational description of a comparably large and flexible molecule. This approach will extend the range of assessable structural diversity, raising the bar for the investigation of more complex and flexible systems. The strategy of the herein presented model system pushes the limit for the detection of weak halogen bonds towards more polar solutions, such as water.

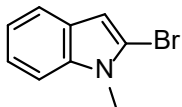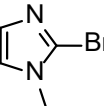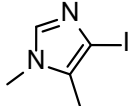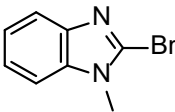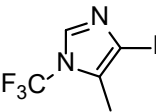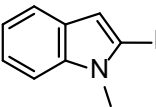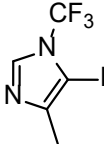# 4. Design and Synthesis of Halogen Bond Donor and Acceptor Amino Acids
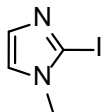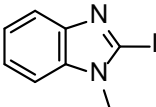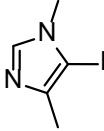
Motivated by the success of the strategy previously described in Chapter 3, I aimed to optimize the peptide model system, including the structure of the halogen bond donor and acceptor amino acids, to reach higher accuracy and wider applicability. The system's design focussed on reaching a linear, inter-strand halogen bond of two amino acid sidechains separated by an optimal distance when embedded into opposing positions of two β-sheet strands. A β-hairpin appears to have several advantages, including the chemical shift dispersion of the α-protons and amides as well as a geometry that favours the formation of cross-strand interactions that stabilize the folded structure. The cooperativity of a series of weak forces in this system are expected to compete with intermolecular interactions i.e. to the solvent. Measuring and characterizing a halogen bond in polar, protic solvents, such as water, constitutes one of the biggest challenges in the field of halogen bonding.[103, 104] A well-defined water-soluble β-hairpin combined with an optimized XB bond donor-acceptor pair should be an ideal platform to analyse a dynamic XB in water or in other polar, protic solvents. The goal was to systematically study the strength and geometry of a XB by variation of the XB acceptor moiety and by evaluating the impact of the competition by polar solvents. Herein, the computational optimisation of the sidechain length, and the development of synthetic procedures towards the XB donor and nitrogen-, oxygen- and sulphur-containing XB acceptor amino acids to be included into a water-soluble β-hairpin forming peptide is described.

4.1.   DFT Analysis of Halogen Bond Donors

A selection of aromatic systems containing one or two nitrogen atoms with either bromine or iodine at various positions acting as the XB donating group was analysed by DFT calculations (Table 5). The goal was to sample the effect of the halogens position as compared to the neighbouring heteroatoms on the magnitude of the σ-hole.[105] DFT calculations were run with the B3LYP func-tional [106-108] in combination with the aug-cc-pVTZ basis set [109, 110] for carbon, hydrogen, nitrogen and fluorine. For the halogens bromine and iodine, the

modified MDF Stuttgart/Cologne effective core potential basis set aug-cc-pVTZ-PP [111] was used. The polarizable continuum model (PCM)[112] was used to model solvation effects using the integral equation formalism variant for water. Resulting energies were revised by applying a Grimme D3 empirical dispersion correction.[113] The electrostatic potential maxima ($v_{S,max}$) at the σ-hole were determined for the XB donors **3-12** using Multiwfn (v. 3.8)[114, 115] and are listed in Table 5.

Table 5: Structures **3-12** with their associated electrostatic maximum at the σ-hole ($v_{S,max}$, in kJ/mol) at the face perpendicular to the C-X bond resulting from DFT calculations.

	Structure	$v_{S,max}$ [kJ/mol]		Structure	$v_{S,max}$ [kJ/mol]
3		106.9	**8**		74.8
4		69.6	**9**		47.6
5		122.4	**10**		115.0
6		85.6	**11**		89.3
7		112.4	**12**		138.6

The highest σ-hole maximum ($v_{S,max}$) was found for structure **12** due to the strong electron withdrawing effect of the CF$_3$-group. However, the CF$_3$ close to the iodine is likely to inflict a steric repulsion on an attracted XB acceptor. Thus, the second strongest XB donor from this series, **5**, was selected for further computational analysis. The iodo-benzimidazole group is chemically stable, synthetically available, and provides four C–H groups suitable for NOE and RDC analysis. The electrostatic potential map indicating the size and surrounding of the σ-hole in the iodine of **5** is depicted in Figure 21.
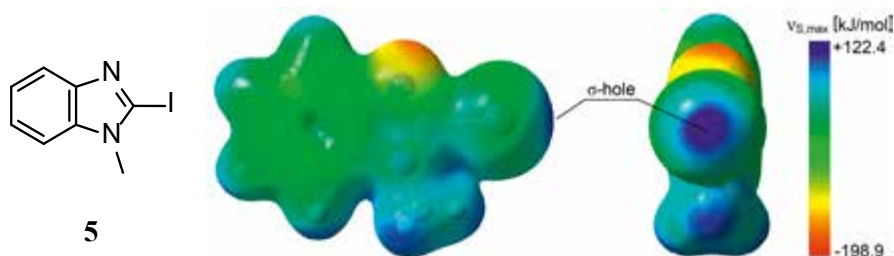
Figure 21: Electrostatic potential map calculated for **5** depicting the position and charge of the σ-hole.

4.2. Linker Sampling in a β-Sheet Context

After selection of a XB donor, the optimal side chain length from the amino acid backbone to the iodo-benzimidazole, the XB donor unit, was assessed. Simultaneously, the XB acceptors - containing a tertiary amine, an ether and a thioether - were sampled for optimal sidechain lengths. These three nucleophilic groups inherit different Lewis basicity and thus allow for the characterization of XB acceptor effects on binding strength and affinity. Such analysis has so far only been performed by computations or by X-ray diffraction methods, but not yet in solution. Besides the sidechain attachment, the heteroatoms N, O and S were methylated to minimize steric hindrance upon binding to the iodine. In an antiparallel β-sheet, the stereocenters at the C_αs cause sidechains to iteratively face above and below the β-sheet plane (Figure 22).[82] Additionally, every second pair in the backbone of an anti-parallel β-sheet forms two cross-strand hydrogen bonds between the N-H and C=O moieties. At this position, the C_αs are further apart (~5.2 Å) as compared to the position where no hydrogen bonds are formed (~4.8 Å). The XB donor and acceptor sites were studied for inclusion in both of these possible positions.
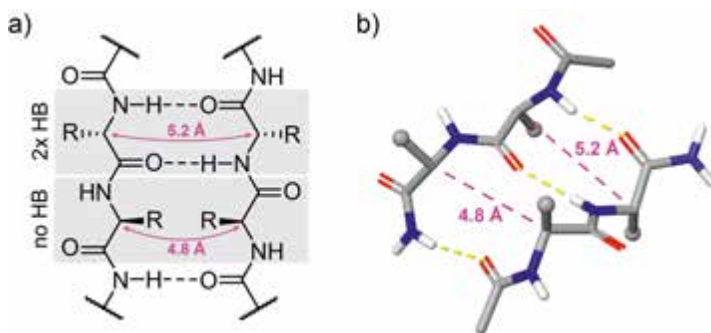
Figure 22: Depiction of the different inter-strand distances of C_α in an anti-parallel β-sheet. a) Schematic representation of a β-sheet. At the positions where two inter-strand hydrogen bonds form, the C_α groups point outwards and are further apart. In contrast, at the inwards orientation, no hydrogen bonds are formed and C_αs are closer b) 3D representation of a β-sheet showing the orientation of consecutive amino acid sidechains. Hydrogen bonds are indicated as yellow dotted lines and inter-stand distances of C_α atoms are shown as pink dotted lines. Spheres represent sidechain residues.

Conformational minimisations of C_1-C_3 and C_3-C_5 linker lengths for the XB donor and acceptor side chains, respectively, were performed. Backbone positions were frozen to maintain a β-sheet resembling geometry, mimicking the two different orientations of the sidechains. Common force fields lack parametrisation for halogen bonding, and thus the XB must be introduced by overcoming the force field biased energy barrier using manual constraints. XB formation was therefore enforced by applying a 180° C−I⋯R (R: N, O or S) bond angle constraint and an I⋯N/O distance constraint of 3.2 Å and 3.4 Å for I⋯S. The goal of this structural minimisation was the assessment whether the geometry of the inter-strand halogen bond is feasible, without comparing calculated interaction energies. Following energy minimization (Molecular Mechanics), the XB bond angle and the I⋯R distance were used to identify the optimal sidechain lengths facilitating the formation of a halogen bond. Short linkers hampered the formation of a linear C−I⋯R halogen bond, whereas long ones showed little support in aligning the XB donor and acceptor. The energy minimizations suggest that a C_2 linker on the iodo-benzimidazole XB donor and a C_3 linker on the XB acceptor in the "outwards" orientation provides optimal geometry for the formation of a linear cross-stand XB in the context of a β-sheet (Figure 23).
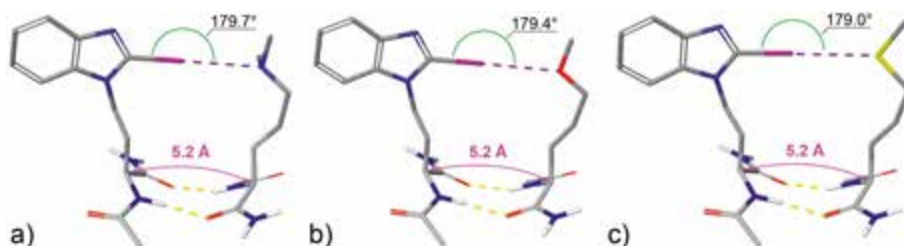
Figure 23: Optimized linker lengths of the XB donor (C_2) and the three different acceptors (C_3). XB donor with the XB acceptor group as a tertiary amine (a), an ether (b) and a thioether (c). At the selected linker lengths, the formed XB (purple dotted line) is near linear (~180°) upon a backbone inter-strand distance of 5.2 Å at the HB forming position (yellow dotted lines).

4.3. Synthesis of Halogen Bond Donor and Acceptor Amino Acids

A synthetic procedure to obtain the halogen bond donor and acceptor amino acids as Fmoc-protected building blocks suitable for SPPS was developed.
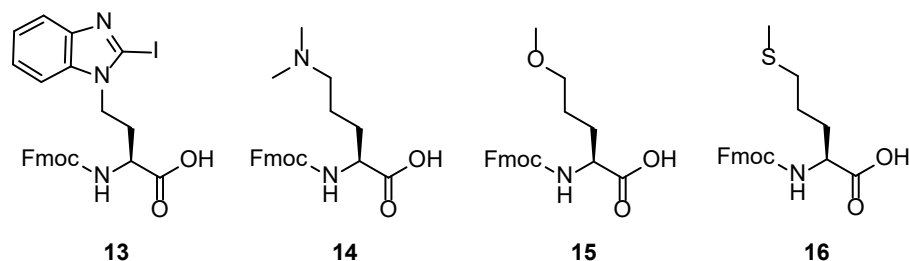

Figure 24: Fmoc protected XB donor amino acid **13** and XB acceptor amino acids **14**, **15** and **16**.
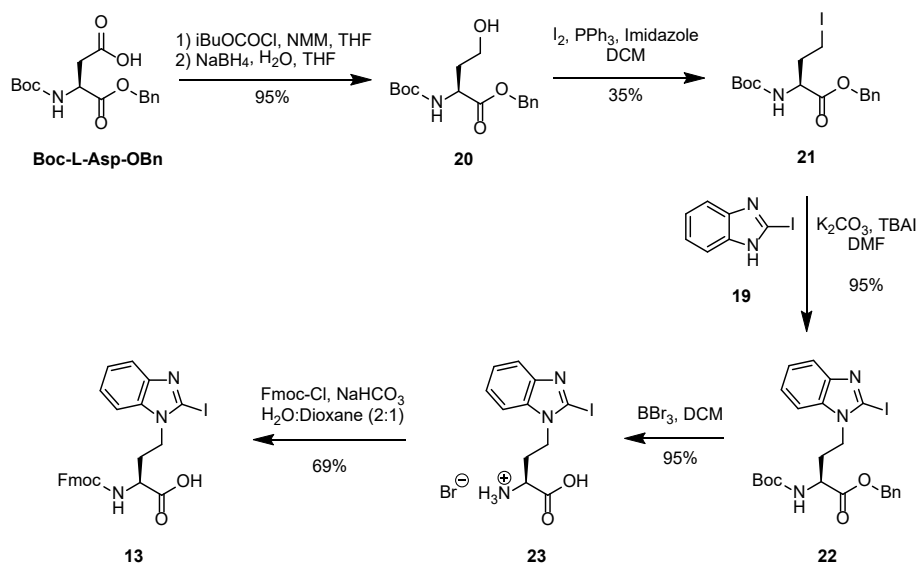
The Halogen Bond Donor Building Block

Iodo-benzimidazole **19**, which constitutes the core of the XB donor amino acid, was obtained by first protecting the NH group of 1H-benzimidazole using trityl-chloride, in the presence of triethylamine in DCM. Subsequent halogenation of **17** at the 2-position was achieved by treatment with LDA in THF at -78°C for 1h, followed by addition of I_2 providing **18** in 22% yield. Removal of the trityl group by treatment with conc. HCl in MeOH (1:4, v/v) at 70°C yielded **19** in 58%.
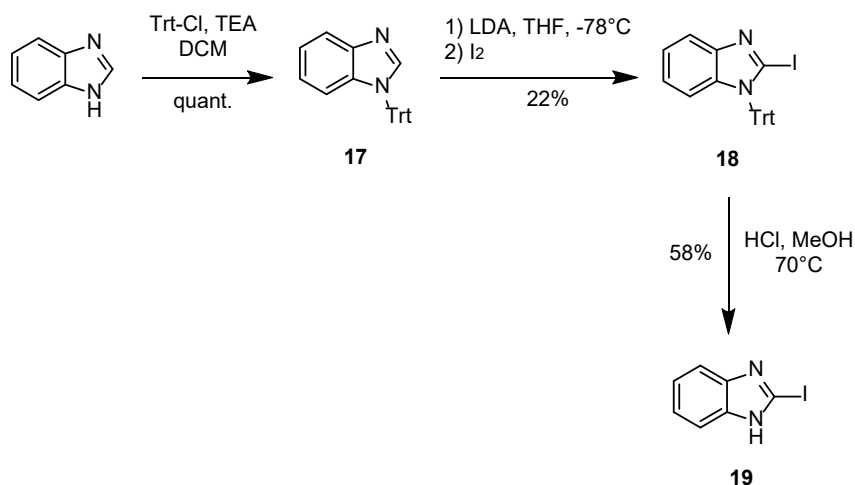
Scheme 2: Synthesis of iodo-benzimidazole **19**. LDA (Lithium diisopropylamide), Trt-Cl (Triphenylmethyl chloride).

The XB donor amino acid was obtained starting from N- and C-protected aspartate (Scheme 3). The sidechain carboxylic acid of Boc-L-Asp-OBn was reduced by first forming an asymmetric anhydride using *iso*-butyl-chloroformate, followed by reduction with $NaBH_4$. The alcohol **20** was further transformed into an alkyl iodide via an Appel-type reaction. Compound **21** was then combined with **19** in a S_N2 reaction using K_2CO_3 in DMF and TBAI as phase transfer catalyst. The alkylated iodo-benzimidazole derivative **22** was obtained in excellent yield. Simultaneous removal of the Boc- and benzyl-protecting groups was achieved using BBr_3 in dry DCM, providing compound **23** as a hydrobromide salt. In the last step, the amino group was protected using Fmoc chloride in the presence of $NaHCO_3$ in a mixture of water and 1,4-dioxane (2:1, v/v) at 0°C to provide the target compound **13**, which has a XB donor side chain, in 69% yield. Preservation of the enantiomeric purity was confirmed by chiral HPLC, using a Phenomenex Lux® 5µm Amylose-1 chiral column.

Scheme 3: Synthesis of the Fmoc-protected XB donor amino acid **22**. NMM (N-methylmorpholine), TBAI (*tert*-butyl ammonium iodide), Fmoc-Cl (9-fluorenyl-methoxycarbonyl chloride).

The Halogen Bond Acceptor Building Blocks

The XB acceptor carrying a tertiary amine group was readily synthesized starting with a selective Boc-deprotection of the terminal amine of Fmoc-L-Orn(Boc)-OH using 4 N HCl in 1,4-dioxane. Compound **24** was subjected to reductive methylation using formaldehyde and sodium cyanoborohydride. The final compound, Fmoc-protected dimethyl ornithine, **14**, was obtained in 55% yield over two steps.
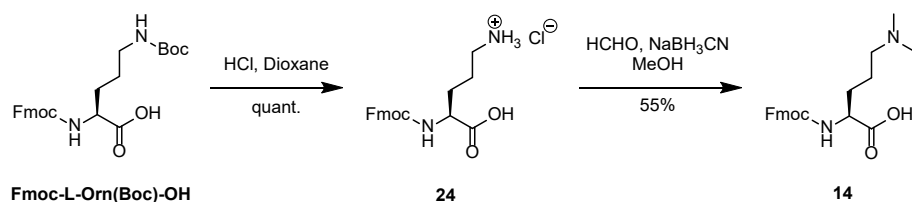
Scheme 4: Synthesis of the Fmoc-protected XB acceptor amino acid **14**.

Attempts to synthesize the O and S containing XB acceptor building blocks via a similar route to that used for generating the XB donor started from Fmoc- and tBu-protected glutamic acid. The free carboxylic acid was first protected using benzyl-bromide and Cs₂CO₃ in MeCN yielding **25** (Scheme 4). Deprotection of the tBu ester under acidic condition with water as a scavenger gave

26 in quantitative yield. Subsequent reduction of the free acid was accomplished using *iso*-butyl-chloroformate and NMM in THF forming the asymmetric anhydride, which was then converted to the alcohol **27** using NaBH$_4$. An Appel-type reaction using I$_2$, PPh$_3$ and imidazole in DCM at 0°C gave the alkyl-iodide **28** in good yield. Compound **28** was then transformed into the methyl-thioether **29** using NaSMe in DMF. After removal of the benzyl protection group, compound **16** can be obtained. Continuing from **28**, a nucleophilic substitution using a ROMe derivative forming the methyl-ether followed by benzyl deprotection is expected to yield compound **15**. The final steps from **29** to **16** and **28** to **15** have not yet been performed, however literature examples exist guiding the conversion to the final compounds. Herein I propose the developed synthetic scheme to obtain the oxygen and sulphur containing XB acceptor building blocks **15** and **16**.
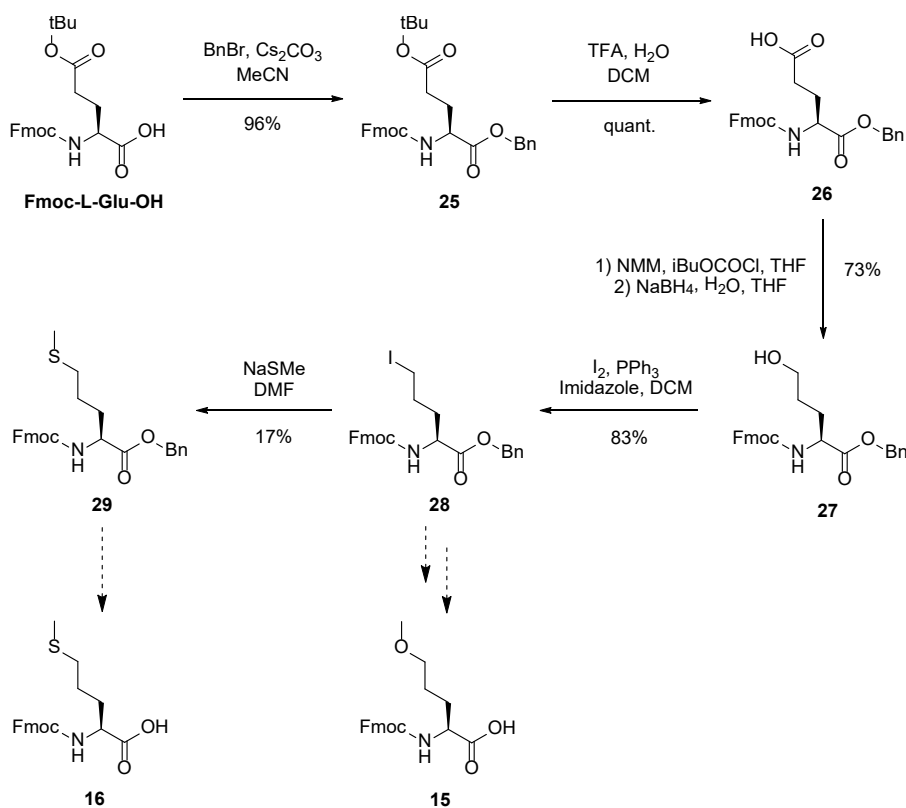
Scheme 5: Proposed synthetic route to obtain the XB acceptor amino acid building blocks **15** and **16**.

4.4. Conclusion and Outlook

I predicted the magnitude of the σ-hole (v_{max}) for a series of XB donor moieties using DFT calculations, to select one for incorporation into an amino acid sidechain. Furthermore, the amino acid sidechain length of the halogen bond donor and halogen bond acceptor amino acids, and their positions in a β-strand, were optimized by molecular mechanics energy minimization. An iodo-benzimidazole as a halogen bond donor and a tertiary amine as a halogen bond acceptor, attached to a C_2 and C_3 spacers, respectively, were identified as good interaction sites when incorporated into a peptide β-hairpin. A synthetic route for two Fmoc-protected unnatural amino acids was developed, allowing their use in solid-phase peptide synthesis.

Incorporation of these halogen bond acceptors pairwise with the halogen bond donor - and its analogues possessing bromine and chlorine - into a β-hairpin peptide consisting of 12-14 amino acids, is expected to allow systematic thermodynamic and geometric characterization of weak halogen bonds in dilute solutions. With a careful design of the amino acid sequence, these studies could allow experimental characterization of halogen bonds in polar and protic media, which has so far been a challenge. Moreover, the halogen bond donor and acceptor sites may be substituted with other interaction sites to allow the investigation of other weak non-covalent forces, such as pnictogen, chalcogen and hydrophobic interactions in dilute solutions.

5. Conformational Analysis of Macrocyclic Drugs Indicates Molecular Chameleonicity (Paper II)

At least half of all druggable targets to treat human diseases are considered unsuitable to be addressed by drugs obeying the Lipinski´s Rule-of-5 (Ro5).[116-118] To date, biologics are employed to target flat, groove-shaped or featureless binding sites. Unfortunately, those compounds have low metabolic stability, and also lack oral bioavailability and cell permeability. As to introduce small molecule properties to somewhat larger potential drugs in the beyond Rule-of-5 (bRo5) space, peptides, peptidomimetics and macrocycles are currently attractive structures.[119-122] These compounds, especially macrocycles, combine otherwise conflicting properties of solubility and cell permeability with target binding. It is hypothesized that conformational changes dynamically shield or expose polar functionalities in response to changes in the surrounding environment. A conformational change that induces a reduction in surface polarity allows a molecule to desolvate from the plasma and penetrate into the lipid bilayer and from there into the cytosol. This behaviour is termed "molecular chameleonicity" and explains their increased permeability as well as the multi-target activity.[123] However, experimental evidence so far has only been shown for the cyclic peptide cyclosporine A.[8, 124] Herein experimental proof for the solvent dependent property redistribution of the macrocyclic erythronolide antibiotics roxithromycin and telithromycin is investigated.
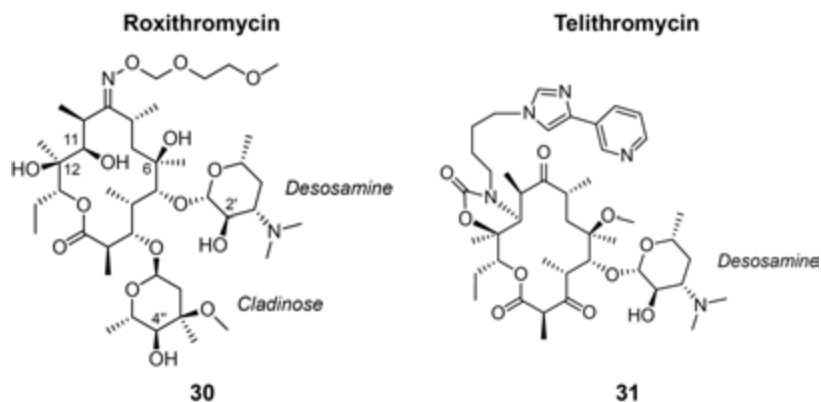
Figure 25: Structures of the two macrocyclic drugs roxithromycin and telithromycin for which conformational ensembles in D_2O and $CDCl_3$ have been determined. Names of attached sugar moieties are in italics.

5.1. Solvent Dependent Conformational Distribution

The aim of this study is to experimentally evaluate whether large flexible macrocyclic drugs adjust their conformation to the environments polarity, which may explain their simultaneous aqueous solubility and membrane permeability. Roxithromycin, **30**, and telithromycin, **31**, two flexible macrocyclic drugs have been used as model systems, and were systematically studied in environments mimicking the plasma/cytosol and the cell membrane. Chloroform is commonly used to imitate a low polar environment since its dielectric constant ($\varepsilon = 4.8$) is close to that found in a lipid bilayer ($\varepsilon = 3.0$).[125] Water, with the pH adjusted to ~7.0, adequately mimics the cytosol. Roxithromycin and telithromycin with pK_a values of 9.13 and 8.69 respectively are protonated at their desosamine moiety under these conditions.[126] Solution NMR is the method of choice to investigate solvent effects on conformation since it allows detection of even minor changes in conformational distribution at atomic resolution.

NMR samples of **30** and **31** were prepared in D_2O and $CDCl_3$. Assignment of proton shifts in both solvents were accomplished based on ^{1}H-NMR, COSY, TOCSY, HSQC, HMBC and NOESY spectra recorded at 298 K. Quantitative interproton distance were obtained from NOE build-up analysis as described in Chapter 1.6. Vicinal coupling constants of protons in the macrocycle were determined for telithromycin from the ^{1}H NMR. Experimentally determined parameters for roxithromycin contained 19 and 13 interproton distances for D_2O and $CDCl_3$, respectively. For telithromycin, 26 distances and 4 coupling constants in D_2O, and 18 distances and 2 coupling constants in $CDCl_3$ were used.

Unrestrained Monte Carlo conformational searches were run with different force fields and with the GB/SA solvation models for water and chloroform, and combined to generate an ensemble of theoretically possible conformers. Redundant conformer elimination of the combined conformer sets was used to eliminate duplicates. In addition to the computationally generated conformers, crystal structures obtained from PDB and CSD databases were added to the ensemble (Table 6). The four X-ray structures for roxithromycin and five for telithromycin completed the theoretical ensemble of 70 conformers for roxithromycin and 172 conformers for telithromycin. These theoretical conformational ensembles were used in the subsequent NAMFIS analyses.

Table 6: Crystal structures of roxithromycin and telithromycin from the Protein Data Bank (PDB) and the Cambridge Structural Database (CSD), added to the input ensembles and included in the NAMFIS analysis.

	CSD		PDB	
	entry	R-factor	entry	Resolution
Roxithromycin	EWETUV	2.71 %		
	FUXYOM	4.70 %		
	KAHWAT	1.68 %		
			1JZZ	3.8 Å
Telithromycin			1YIJ	2.6 Å
			1P9X	3.4 Å
			4V7S	3.3 Å
			4V7Z	3.1 Å
			4WF9	3.4 Å

Roxithromycin

The solution ensembles of roxithromycin, as obtained from NAMFIS analysis, contained six conformers in D_2O, whereas a single conformer was found in the $CDCl_3$ solution (**30a**). The latter conformer is identical to the X-ray crystal structure KAHWAT, which is also present to 6% probability in the solution ensemble in D_2O. In this structure, the oxime sidechain of roxithromycin is oriented above the macrocyclic core and forms two intramolecular hydrogen bonds. Additionally, the two sugar moieties are in a staggered orientation. This overall compact geometry effectively shields the polar functionalities of the molecule. The solution ensemble in water shows higher diversity in the orientation of the oxime sidechain and of the sugar moieties. In addition, the macrocycle itself adopts dissimilar conformations with RMSDs up to 1.23 Å.
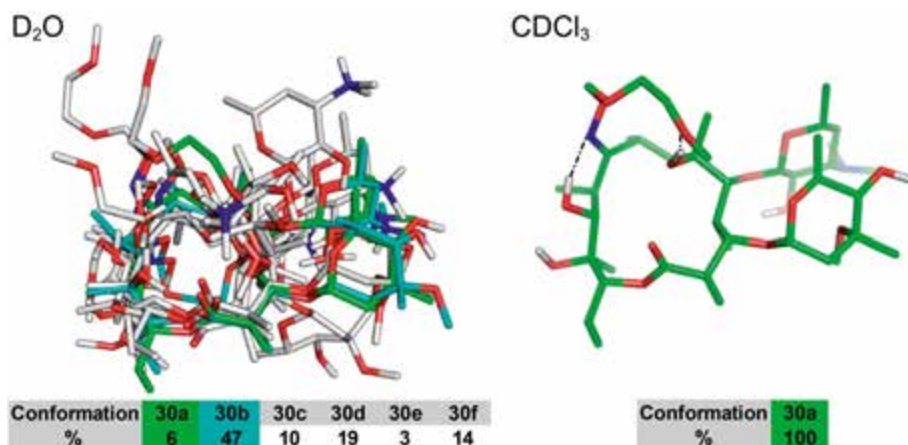
Figure 26: Solution conformers (in %) identified for roxithromycin in D_2O and $CDCl_3$. The most populated conformer in D_2O is highlighted in cyan. The single conformer found in $CDCl_3$ that is also present to 6% in the water ensemble is highlighted in green. Intramolecular hydrogen bonds of conformer **30a** from 11-OH and 6-OH are depicted as black dotted lines.

To ascertain the role of the oxime sidechain as flexible actor in the shielding of polar groups, I performed variable temperature NMR experiments of **30** in $CDCl_3$ at a temperature range of 253-298 K. The temperature coefficients ($\delta\Delta_{OH}/\Delta T$) revealed the extent of hydroxyl groups being involved in intramolecular hydrogen bonds (Table 7). The two hydroxy groups on the sugar moieties, 2'-OH and 4''-OH, as well as 12-OH situated on the macrocycle showed a < 2.4 ppb/K temperature dependent chemical shift change (see Figure 25 for numbering). This indicates the formation of strong intramolecular hydrogen bonds, most likely to neighboring oxygen atoms, forming a 5-membered ring. The two hydroxy groups, 6-OH and 11-OH, shift more pronounced upon temperature change (> 5.3 ppb/K), which indicates that the equilibrium between their intramolecular hydrogen bonded state and a state in which they interact with the surrounding solvent is shifted to the latter one.

Table 7: Temperature coefficients ($\Delta\delta_{OH}/\Delta T$ in ppb/K) of OH groups on roxithromycin in $CDCl_3$ determined by ^1H NMR at a temperature interval of 253-298 K given in absolute values.

	6-OH	11-OH	12-OH	2'-OH	4''-OH
$\Delta\delta_{OH}/\Delta T$ [ppb/K]	6.2	5.3	1.8	2.4	2.0

These findings corroborate the hypothesis that the conformation of flexible macrocycles depends on the polarity of the environment. The appearance of a conformer, which is dominant in a low-polarity environment, in the aqueous solution with a lower molar fraction, indicates that polarity change induces

58

adjustments in conformation populations by stabilizing or destabilizing certain conformers. The difference in surface polarity of roxithromycin in the solution ensembles in D_2O and $CDCl_3$, was quantified by the solvent accessible 3D polar surface, a polarity descriptor calculated for individual conformers (see Chapter 5.2).

Telithromycin

The solution ensembles of telithromycin in D_2O and $CDCl_3$ each consist of seven conformations. The *E. coli* ribosome bound conformation (PDB: 4V7S, **31g**) is present to 5% in water, while the ribosome bound forms in *T. thermophilus* (PDB: 4V7Z, **31m**) and *S. aureus* (PDB: 4WF9, **31n**) represent dominant conformers in chloroform with 52% and 16% population, respectively. The conformational distribution in $CDCl_3$ is, in contrast to roxithromycin, comparably diverse judged by the RMSD of the conformers (0.94 – 5.71 Å). In the aqueous ensemble of telithromycin, two clusters with different macrocycle geometries were identified, each comprising approximately 50% population. In chloroform however, the cores of five of the seven conformers, which comprise 94% of the ensemble, belong to the same macrocycle cluster supporting the tendency of reduced flexibility in the apolar solvent. The desosamine moiety of telithromycin in both solution ensembles appears rather rigid in relation to the macrocyclic core.
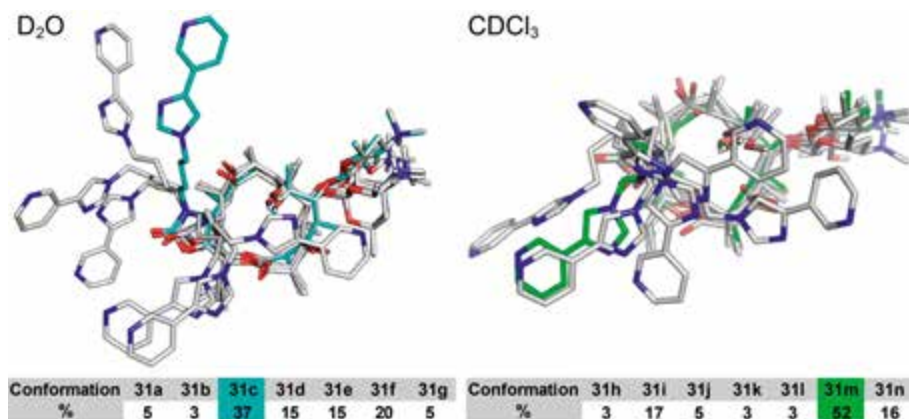

Conformation	31a	31b	31c	31d	31e	31f	31g	Conformation	31h	31i	31j	31k	31l	31m	31n
%	5	3	37	15	15	20	5	%	3	17	5	3	3	52	16

Figure 27: Solution conformers (in %) identified for telithromycin in D_2O and $CDCl_3$. The most populated conformer in D_2O is highlighted in cyan and green in $CDCl_3$.

The experimental NOE data obtained for telithromycin in both solvents provided quantitative interproton distances for the large sidechain, hence allowed the detailed description of its orientation. In chloroform, three minor conformers (**31j**, **31l** and **31n**) have the aromatic sidechain oriented above the macrocycle. The major conformer (**31m**, 52%) exhibits a linker orientation

towards the macrocycle but protrudes the aromatic rings away from the core. These conformational changes synergistically reduce the molecular radius of gyration (R_{gyr}) that reflects the general size and expansion of a compound. A discussion of the R_{gyr} and the solvent accessible 3D polar surface area for both erythronolides is presented in Chapter 5.2. Telithromycin, even if less pronounced than roxithromycin, also contains conformers that are more open and flexible in water and curled up in chloroform.

5.2.   Descriptor-based Analysis of Solution Ensembles

Knowledge about the conformational distribution and the related polar surface area of solution conformers is useful to characterise molecular chameleons. The recently popular descriptors radius of gyration[127] and solvent accessible polar surface area showed correlation to cell permeability.[128, 129] The radius of gyration (R_{gyr}) for all solution conformers was calculated as the root-mean-square distance between all atoms and the centre of mass using MOE (v2015.10). The solvent accessible 3D polar surface area (SA 3D PSA) was calculated using PyMol (v1.7.4) from the solvent accessible area defined with a probe radius of 1.4 Å (Figure 28). As expected from the direct assessment of conformers, the solution ensembles for telithromycin in comparison to roxithromycin show a much wider distribution of R_{gyr}. The major conformers of both compounds are rather similar in their R_{gyr}, which stands in contrast to the SA 3D PSA calculated for these conformers.
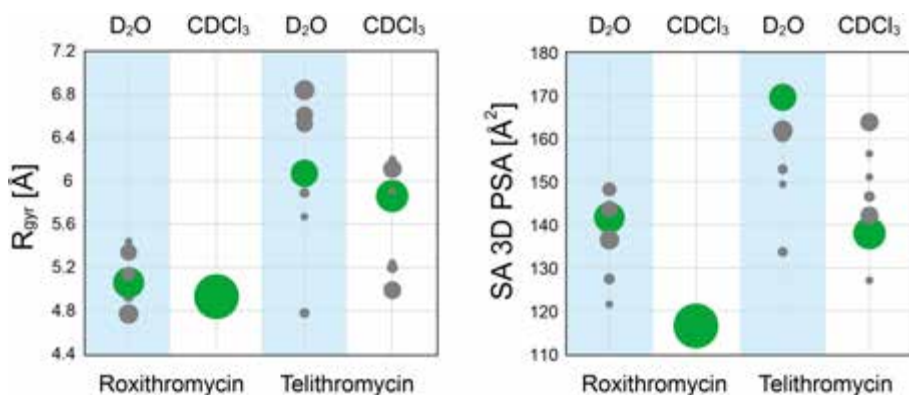

Figure 28: Calculated radius of gyration (R_{gyr}) and solvent accessible 3D polar surface area (SA 3D PSA) for the conformational ensemble of roxithromycin and telithromycin in D_2O and $CDCl_3$. Size of each circle represents the molar fraction representing the population in % of each conformer. Major conformers found in each solvent are indicated in green. The SA 3D PSA in D_2O was calculated for the protonated form whereas in $CDCl_3$ the neutral from was taken.

The spread of SA 3D PSA of telithromycin is similar in both solvents; however, the major conformer in D_2O exposes 32 Å^2 more polar area than the major one in $CDCl_3$. The same trend is observed for roxithromycin with a 25 Å^2 difference of the major conformer in D_2O and the single conformer found in $CDCl_3$. This trend seem remarkable when keeping in mind that the compounds only vary their orientation in space, but maintain the same chemical composition. This shows, that the SA 3D PSA is a more sensitive descriptor for the conformational adjustment of a bRo5 compound adapting to the polarity of the environment. Conventionally the topological polar surface area (PSA) is used to define the polarity of compounds from their 2D fragments and is a satisfactory descriptor for the drugs complying with the Lipinski Ro5.[130] For drugs in the bRo5 space, it is now suggested that the SA 3D PSA is a more viable descriptor for the chameleonicity, the compounds polarity as well as the aqueous solubility.[128]

5.3. Summary

Drugs in the bRo5 space have been hypothesized to act as molecular chameleons. Hence, their environment induced conformational adjustment allows them to display both high aqueous solubility and high cell permeability. This hypothesis has originally been based on extensive experimental and computational studies of cyclosporine A. Other drugs in the bRo5 space have investigated by inspection of their crystal structures and computationally only.

I presented experimental evidence that two non-peptidic bRo5 drugs, the erythronolide antibiotics roxithromycin and telithromycin, possess an inherent conformational flexibility that allows them to adapt to polarity changes of their environment. Solution ensembles were determined by NMR spectroscopy in aqueous and non-polar ($CDCl_3$) solutions. The solvent accessible 3D polar surface area of the conformational ensembles in water showed higher polarity as compared to that observed for chloroform. More compact conformers were found in the solution ensemble in chloroform for both compounds, indicated by their calculated radius of gyration. Synergistic structural adjustments of the macrocyclic core and the attached sidechains facilitate roxithromycin and telithromycin to attain major differences in size and polarity between different environments.

I propose that semi-flexibility is an important element to incorporate in the design of potential drugs in the bRo5 space directed towards intracellular targets. Additionally, the experimentally obtained ensembles in different solvents may guide the development of improved *in silico* sampling techniques currently struggling to reproduce the real-life conformers of macrocyclic drugs.

6. Elucidating the Relative Configuration of a Macrocyclic Natural Product (Paper III)

Senecio mannii, a species within the Asteraceae family, is a shrub common to tropical Africa, and is widely used in folk medicine.[131] Plants within the genus *Senecio* have elevated contents of pyrrolizidine alkaloids, which are heterocyclic secondary metabolites that contain esters of necic acid and a necine base.[132, 133] These alkaloids inherit strong genotoxic, hepatotoxic, tumorigenic and neurotoxic properties.[134] As common food contaminants in grain, milk, honey and eggs they exhibit a threat to public health. Phytochemical investigation of the extracts of the root and stem-bark of *Senecio manii* afforded, amongst other alkaloids, the known macrocyclic compound senaetnine, **32** (Figure 29). This compound was first isolated and described by Bohlmann[135] *et al*. in 1977, however its configuration has not been determined experimentally. Only an assumption based on the biosynthesis from the precursor necic acids was disclosed. Attempts to crystallize **32** in order to gain absolute configuration by X-ray diffraction were unsuccessful.
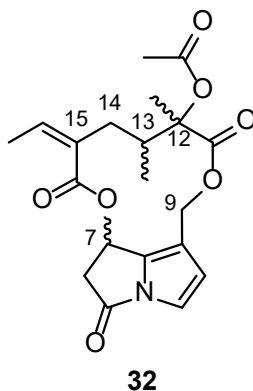

32

Figure 29: Structure of the natural compound senaetnine **32** isolated from *Senecio mannii*. Stereochemistry at carbons C-7, C-12 and C-13 is unknown.

Direct assessment of NOEs and *J*-couplings from neighboring protons within the macrocycle to the stereocenters C-12 and C-13 were inconclusive presumably due to conformational exchange around the C12-C13 bond. However, the bicyclic structure of the necin base and substituents at C-12 and

C-15 reduce the conformational freedom in large parts of the molecule. The global assessment of the macrocycle by means of ensemble analysis was expected to be applicable for the identification of the relative configuration, based on the comparison of the interproton distances describing the macrocyclic core with distances back-calculated for theoretical models of alternative stereoisomers. I used quantitative NOEs in combination with individual theoretical ensembles for the different diastereoisomers to elucidate the unknown stereochemistry of this natural product.

6.1.   Configurational Analysis of Senaetnine

NOE build-up analysis of **32** in CDCl$_3$ was performed by recording seven NOESY experiments with t$_{mix}$ set to 100-700 ms. Normalized integrals were plotted against individual mixing times to determine the build-up rate (Chapter 1.6). Only build-ups with an R^2 > 91% were selected to calculate interproton distances. The build-up rate of the geminal protons at C-14 was selected as reference for distance determination. I obtained 15 interproton distances, which covered the orientations of the rotatable bonds of the entire macrocyclic core (Figure 30).
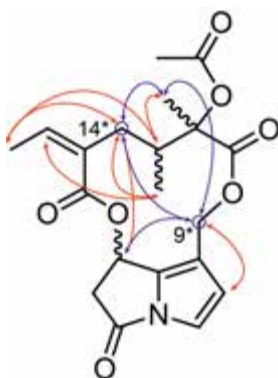

Figure 30: Interproton distances derived from NOE build-up analysis depicted as red arrows. * Diastereotopic protons on C-9 and C-14 (blue circle) involved in the analysis. For interproton distances indicated with a blue arrow, NOEs from both protons of the CH$_2$ group were observed to other protons.

Four separate theoretical input ensembles were generated for the diastereoisomers RRR, RRS, RSR and RSS respectively. The corresponding enantiomers were not included since the NOEs do not allow for differentiations between them. Monte-Carlo based conformational searches were implemented using the force fields OPLS3e, OPLS and AMBER* in two GB/SA solvation mod-

els for water and chloroform. Resulting conformers of the individual diastereoisomers were subjected to redundant conformer elimination applying a heavy-atom cut-off at 0.8 Å. This resulted in 22, 21, 21, and 20 conformers in the theoretical ensemble of RRR, RRS, RSR and RSS respectively. Despite exhaustive sampling steps within a large energy window, the obtained number of different conformers is comparably small. This reflects the limited flexibility of the system that makes it ideal for configurational analysis using NAMFIS.

Separate runs were performed using the set of 15 interproton distances against the different theoretical ensembles representing the four diastereoisomers. Experimentally determined distances (r_{exp}) were assigned staggered error estimates based on their length (Table 8).

Table 8: Error estimates used in the NAMFIS analysis for experimentally determined interproton distances (r_{exp}).

NOE distance [Å]	Error estimate
$r_{exp} < 2.5$	0.1
$2.5 \leq r_{exp} < 3.0$	0.2
$3.0 \leq r_{exp} < 3.5$	0.3
$3.5 \leq r_{exp}$	0.4

Additional permutation studies were performed in NAMFIS, changing the assignment of the diastereotopic protons on C-9 and C-14. For six of the combinations NAMFIS could not find a feasible solution due to conflicting assignment of the diastereotopic protons at C-9. The resulting combinations were ranked by their fit of back-calculated distances to experimental distances indicated by the RMSD. The diastereoisomer RRR at permutation B (Figure 31) showed a significantly lower RMSD (0.18) as compared to other combinations of chirality and assignment of H-9ab and H-14ab.
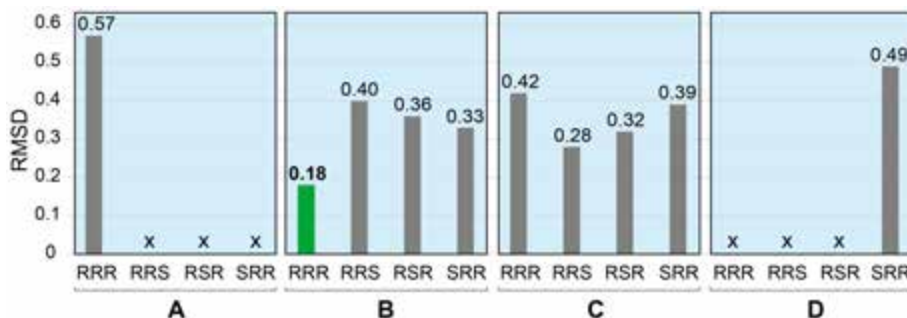

Figure 31: RMSD values obtained from all combinations run in NAMFIS. Each box (A-D) represents permutation of the assignment of protons on C-9 and C-14. Assignment for B is 9a,b (5.47, 4.69 ppm) and 14a,b (2.19, 2.33 ppm). The RRR configuration in permutation B shows the best fit to the experimental data (RMSD: 0.18). (x) No feasible solution was found for this combination.

The best-fit solution was validated by iterative removal of individual distances without a major (<10%) change in molar fractions. The 7R-12R-13R assignment of senaetnine based on the diastereoisomer differentiation of NAMFIS was in agreement with the proposed configuration by Bohlmann *et al.*[135] Besides the relative comparison of isomers, NAMFIS provided the solution ensemble for compound **32**. Out of the 22 input conformers in the theoretical ensemble, four were selected to be the present conformers of which a molar fraction of 67% is assigned to the major conformer.
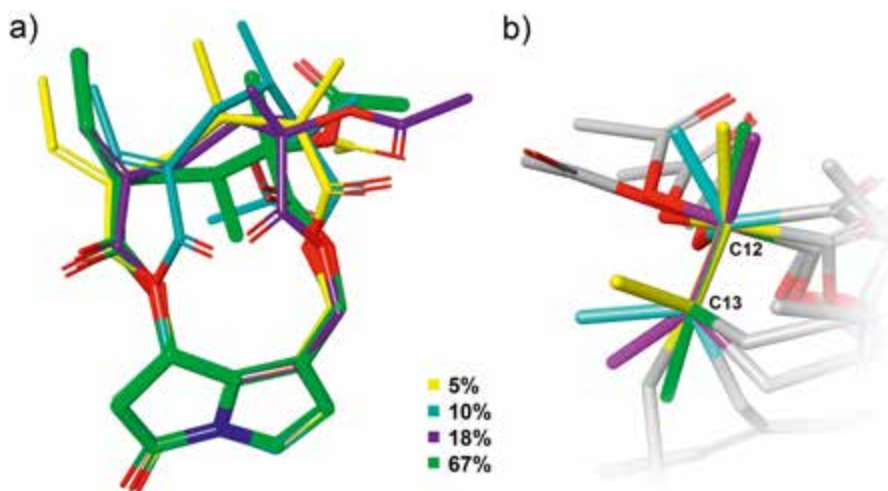
Figure 32: a) Superimposed conformers of **32** and their corresponding populations. The major conformer is colored green. b) Superposition of C-12 and C-13 show the different orientations at the two stereocenters due to rotation of bonds within the macrocycle.

An overlap of the solution geometries by superposition of C-12 and C-13 indicates the high flexibility of substituents at the two stereocenters. This corroborates the assumption that rotational freedom in this region precludes the direct deduction of the relative stereochemistry based on qualitative NOEs and *J*-couplings.

6.2.   Summary

The relative configuration of senaetnine was successfully determined using a NAMFIS based configurational ensemble analysis. Results from a combination of different input ensembles representing the four different diastereoisomer and permutation of diastereotopic assignment were examined on their fit to the experimental data. The RMSD descriptor obtained from NAMFIS revealed that the theoretical ensemble of the RRR isomer best-fit the 15 interproton distances determined by NOE build-up analysis. The

resulting RRR (or SSS) assignment of C-7, C-12 and C-13 corroborates with literature suggestions based on biosynthetic pathways yet now stands experimentally proven. This demonstrates the tactical use of NOEs within a semi-flexible molecule whose conformational distribution is determined by stereogenic centers, to allow the differentiation between possible diastereoisomers.

7. Concluding Remarks and Perspective

In this thesis, I performed conformational analysis by solution NMR spectroscopy to address three distinct fundamental questions. I deconvoluted population-averaged spectroscopic data to describe the conformational ensemble of flexible molecules, and used the gained knowledge to provide: an unparalleled understanding of a weak non-covalent interaction (Chapter 3 and 4); the impact of solvent polarity on the conformational distribution of macrocyclic drugs (Chapter 5); and elucidation of the relative stereochemistry of a natural product (Chapter 6).

In Chapter 3 (Paper I), I studied a weak halogen bond in a dilute solution. I developed a strategy that allows detection and characterization of such a weak bond by overcoming the entropic barrier through studying a weak halogen bond in an intramolecular setting. Detection of the interaction is facilitated by studying it in a cooperatively folding peptide sequence, in comparison to a reference that lacks the specific interaction site. I demonstrated that using this strategy, a C−I···O halogen bond weaker than 0.9 kJ/mol is detectable and characterizable in solution. I have determined the extent that this halogen bond stabilizes a β-hairpin conformer, the population and the geometry of this halogen bond, and have characterized the interaction thermodynamically. In Chapter 4, I furthermore designed an improved model system for studying a halogen bond in a polar and protic solvent, such as water. The halogen bond donor site was selected using DFT-based prediction of the interaction site, whereas the length of the amino acid sidechain and its positioning within a β-sheet structure was achieved by molecular mechanics minimization. As a result, I have optimized the synthesis of one halogen bond donor-acceptor pair for future incorporation into a peptide. The strategy I present is easily transferable to the investigation of any other type of weak interaction, and is hypothesized to be applicable for systematic studies of halogen bond donor and acceptor sites and of solvent effects upon these.

In Chapter 5 (Paper II) the impact of solvent polarity on the conformational distribution of macrocyclic drugs was investigated. Macrocycles are a promising class of drugs for binding to difficult-to-target interaction sites, and due to their size, they often violate the Lipinski's Rule-of-5. However, they may still show satisfactory aqueous solubility and membrane permeability for

use as drugs. I evaluated the influence of a change in the polarity of the environment on the conformational ensembles of the erythronolide antibiotics roxithromycin and telithromycin in order to experimentally prove or disprove their chameleonicity. NAMFIS-based analysis of their conformational behavior in chloroform and in water revealed that both compounds adjust their solution conformation to the solvent environment, even if to a different degree. The radius of gyration and solvent accessible 3D polar surface area calculated for the different conformers corroborated the presumption of a decreased size and surface polarity in a less polar environment. My data corroborates the hypothesis of molecular chameleonicity and provides an experimental piece of evidence to build upon in future drug design of new pharmaceuticals interacting with difficult-to-drug targets. It also provides experimental data that may be used for the improved parametrization of force fields and computational techniques in general, for the conformational prediction of macrocycles and flexible molecules.

In Chapter 6 (Paper III), the relative configuration of the natural product senaetnine isolated from *Senecio mannii* was determined. The back-calculated NMR data for the conformational ensembles of all possible diastereoisomers were compared to the experimental NMR data, so to identify its relative configuration. This was feasible due to the low flexibility of the macrocyclic core of this compound. NAMFIS analysis identified the RRR (or SSS) configuration. This result is in agreement with the literature proposal based on biosynthetic precursors. This work demonstrates the applicability of solution ensemble analysis for configuration elucidation of compounds with limited flexibility.

Herein I present a strategy for the description of a molecule's solution conformations, and for the evaluation of the interactions that determine the conformational ensemble. Using a β-hairpin model system, I could, for the first time, observe and describe a weak halogen bond in a dilute solution. Based on the strategy of utilizing a model system adapting a secondary structure and observed NOEs, *J*s and RDCs, a variety of currently undetectable interactions including pnictogen and chalcogen bonds or hydrophobic interactions will now be experimentally accessible. I also laid the foundation to design model systems that are compatible with polar solvents and for the systematic evaluation of various halogen bond acceptors. I thus expect to help the rational use of weak non-covalent interactions in drug development. My analysis of macrocyclic drugs facilitates the general perception of how flexibility influences membrane permeability. The presented experimental dataset is valuable for the parametrization of computational force fields and methods.

8. Sammanfattning på Svenska

Kemiska bindningar är krafter som håller ihop de molekyler som gör livet möjligt. Bindningarna karaktäriseras av en viss flexibilitet som möjliggör för molekylen att anpassa sig till olika miljöer. Formerna en molekyl kan anta på grund av sin flexibilitet kallas för konformationer. Det är viktigt att studera vilka konformationer en molekyl kan anta eftersom de påverkar dess aktivitet i biologiska system och dess kemiska reaktivitet med andra molekyler.

I avhandlingen beskrivs hur konformationer har studerats med hjälp av kärnmagnetisk resonansspektroskopi (NMR). Metoden gör det möjligt att studera molekyler i lösning för att skapa en bättre förståelse av hur de veckar sig till olika konformationer, hur stabila dessa är och hur externa faktorer påverkar veckningen. En flexibel molekyl antar dock ett stort antal konformationer som står i jämvikt med varandra. Jämvikten är så snabb att det enbart är möjligt att identifiera ett tidsmedelvärde med NMR. För att få en bild av de olika konformationer som står i jämvikt med varandra krävs en dekonvulotionsalgoritm som bryter ner medelvärdet till de underliggande konformationerna. Den algoritm som använts i den här avhandlingen är NAMFIS (NMR-analys av molekylär flexibilitet i lösning). NAMFIS använder två typer av experimentella data, dvs interprotonavstånd och torsionsvinklar som kan bestämmas med olika NMR-experiment. Den nukleära overhauser-effekten, vilken härrör från en dipol-dipolkoppling, gör det möjligt att bestämma det genomsnittliga avståndet mellan två protoner. Den så kallade J-kopplingen ger information om torsionsvinkeln mellan två atomer som befinner sig tre bindningar från varandra. Karplus-ekvationen gör det möjligt att beräkna torsionsvinkeln från den J-koppling som uppmätts med NMR spektroskopi.

I avhandlingen har NAMFIS-algoritmen använts för konformationsanalys på tre olika föreningar. Först designades en molekyl, dvs en dekapeptid som veckar sig till en β-hairpin i lösning. Denna sekundärstruktur är inte särskild stabil och hålls endast ihop av icke-kovalenta bindningar, vilket gör den till en idealisk model för att studera betydelsen av sådana bindningar för en β-hairpin. Mina studier fokuserade på en bestämd halogenbindning som tillsammans med andra interaktioner stabiliserar β-hairpin konformeren. Jag synteteiserade även en referenspeptid som var identisk till den förra peptiden förrutom att den inte kan forma en stabiliserande halogenbindning. Med hjälp av termodynamisk analys har jag visat att peptiden som kan stabliseras av en halogenbindning är ungefär 1 kJ/mol stabilare än referenspeptiden. Denna

skillnad är följd av halogenbindningen. I detta projekt använde jag NAMFIS för att bestämma populationen av β-hairpin konformeren med hjälp av NMR parametrar som beskriver avstånd mellan protoner samt deras dihedral vinkel. NAMFIS har selekterat konformerer från en pool av teoretiskt möjliga geometrier. Analysen visade att peptiden som stabiliseras av en halogenbindning bildar 17% mer β-hairpin strukturer jämfört med peptiden som saknar denna interaktion. För att förstå 3D-strukturen av de funktionella grupper som ingår i halogenbindningen har jag använt mig av analys av så kallade residual dipolar couplings (RDCs). β-Hairpin peptiden har blandats med en substans som orienterar den. Sådana lösningar ofta kallas alignment medium. I och med den partiella orienteringen av peptiden kunde jag mäta RDCs. Dessa gav information om orienteringen av C-H-bindningar i jämfört med NMR magnetens fält. Igenom att analysera orienteringen av ett antal C-H bindningar i närheten av halogenbindningen kunde jag visa att halogenbindningen kan ha två möjliga vinklar i denna peptid. Dessa vinklar är mycket troliga enligt literaturen. Bestämning av bindningsstyrkan och geometrin av en svag icke-kovalent interaktion, så som halogenbindning, är mycket svårt och ovanligt. Strategin jag utvecklade är därför av stort vetenskapligt värde.

Jag undersökte även hur polariteten i lösningsmedlet påverkar egenskaper hos makrocykliska läkemedel. Vissa makrocykler kan anpassa sina egenskaper till olika miljöer genom att anta olika konformationer, och kallas därför för molekylära kameleonter. Denna förmåga förmodas gynna den löslighet och membranpermebilitet som är nödvändig för användning av läkemedel. För att kunna bevisa detta experimentellt löstes de antibakteriella substanserna roxitromycin och telitromycin i vatten och klorofrom, och en NAMFIS-baserad konfomitetsanalys genomfördes. Det visade sig att båda föreningarna hade en hög flexibilitet i vattenlösning, men att de antog färre och/eller kompaktare konformationer i kloroform. I den mindre polära miljön i kloroform veckar molekylen ihop sig så att polära grupper döljs för att reduceras polariteten av molekylens yta. Studien ger idéer för hur makrocykliska läkemedel kan designas i framtiden.

Med hjälp av NAMFIS analys har jag även bestämt den relativa stereokemin av senaetnin, en substans som vi isolerade av en Östafrikansk medicinalväxt. Senaetnin är en pyrrolizidinalkaloid och isolerades från växten *Senecio mannii* och som har tre stereocentra. Faktum et att molekylen har en begränsad flexibilitet gjort de det möjligt för mig att använda bestämma stereokemin med hjälp av distanser som kunde mätas mellan olika protoner i molekylen. Jag har jämfört den experimentella datan med distanser i teoretiskt möjliga stereoisomerer och visade att teoretiska datan för den med RRR (eller SSS) stereoikemi passade bäst till datan. Det påvisade experimentellt resultatet bekräftar Bohlmanns förslag, som var den förste som isolerade senaetnin år 1977, och som härledde dess stereokemi från biosyntesen.

9. Acknowledgements

This thesis and the work I present within it would not exist without the help of many people. I would like to thank everyone who has helped and supported me in various ways throughout this journey.

First and foremost, I would like to thank *Prof. Máté Erdélyi*. Mate, you are an incredible talented, smart and caring supervisor. You have encouraged me in times where I had more data than I asked for and guided me along every step of the way. You helped me to progress, not just on a scientific but also on a personal level. I am truly grateful for your always open door, the smile and inspiration and for showing me the big picture over and over again when I dug myself in. I loved and appreciated the atmosphere you created in our group, something I truly will miss. You have been the supervisor one could only wish for.

Thanks also to my co-supervisor *Prof. Jan Kihlberg*. Jan, thank you for paving my way into becoming a PhD student here at Uppsala University. You really are a caring person and your support and calmness always motivated me to see the future in a positive way.

Big shout-out goes out to The Boys: *Scott*, *Fabio*, *Michi*, *Chris* and *Bart*. We went through so many adventures together, it's hard to remember. Whatever came our way, we found a way to celebrate it. Thanks for the amazing time here in Uppsala.

Thanks also to members of the NAMFIS subgroup, past and present: *Hermina*, *Jonathan*, *Emma*, *Yoseph*, *Yordanos*, *Christian, Bono* and *Eric*. We certainly hold the world-record in most build-ups ever made. Keep integrating!

Special thanks to *Hermina*. I really enjoyed every second discussing nerdy NMR stuff with you. Having someone like you around makes even the most difficult concepts interesting and made we want to know more.

Thanks, *Julia*, you were a super fun person to work with in the lab. However, sometimes it was not clear who was the actual supervisor. Stay the 'coole Socke' you certainly are.

References

1. Odd Hassel – Nobel Prize Outreach AB 2022. Sun. 25 Sep 2022. , https://www.nobelprize.org/prizes/chemistry/1969/hassel/facts/).
2. C. B. Anfinsen, *Biochemical Journal*, 1972, **128**, 737-749.
3. P. G. Dougherty, Z. Qian and D. Pei, *Biochemical Journal*, 2017, **474**, 1109-1125.
4. B. C. Doak, J. Zheng, D. Dobritzsch and J. Kihlberg, *Journal of Medicinal Chemistry*, 2016, **59**, 2312-2327.
5. D. Sindhikara, M. Wagner, P. Gkeka, S. Güssregen, G. Tiwari, G. Hessler, E. Yapici, Z. Li and A. Evers, *Journal of Medicinal Chemistry*, 2020, **63**, 12100-12115.
6. A. Mann, in *The Practice of Medicinal Chemistry (Third Edition)*, ed. C. G. Wermuth, Academic Press, New York, 2008, pp. 363-379.
7. J. E. DeLorbe, J. H. Clements, B. B. Whiddon and S. F. Martin, *ACS Medicinal Chemistry Letters*, 2010, **1**, 448-452.
8. C. K. Wang, J. E. Swedberg, P. J. Harvey, Q. Kaas and D. J. Craik, *The Journal of Physical Chemistry B*, 2018, **122**, 2261-2276.
9. C. M. Venkatachalam, *Biopolymers*, 1968, **6**, 1425-1436.
10. A. C. Gibbs, T. C. Bjorndahl, R. S. Hodges and D. S. Wishart, *Journal of the American Chemical Society*, 2002, **124**, 1203-1213.
11. P. N. Lewis, F. A. Momany and H. A. Scheraga, *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 1973, **303**, 211-229.
12. G. D. Rose, L. M. Glerasch and J. A. Smith, *Advances in protein chemistry*, 1985, **37**, 1-109.
13. C. Wilmot and J. Thornton, *Journal of molecular biology*, 1988, **203**, 221-232.
14. S. Scheiner, *Journal of the Indian Institute of Science*, 2020, **100**, 61-76.
15. E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure and Applied Chemistry*, 2011, **83**, 1637-1641.
16. G. R. Desiraju and T. Steiner, *The weak hydrogen bond: in structural chemistry and biology*, International Union of Crystal, 2001.
17. L. C. Ch'ng, A. K. Samanta, G. Czakó, J. M. Bowman and H. Reisler, *Journal of the American Chemical Society*, 2012, **134**, 15430-15435.
18. H. Kessler, *Angewandte Chemie International Edition in English*, 1982, **21**, 512-523.
19. G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chemical Reviews*, 2016, **116**, 2478-2601.
20. R. D. Gautam, P. S. Ho, K. Lars, C. L. Anthony, M. Roberto, M. Pierangelo, P. Peter, R. Giuseppe and R. Kari, *Pure and Applied Chemistry*, 2013, **85**, 1711-1713.
21. T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *Journal of Molecular Modeling*, 2007, **13**, 291-296.

22. S. M. Huber, E. Jimenez-Izal, J. M. Ugalde and I. Infante, *Chemical Communications*, 2012, **48**, 7708-7710.

23. E. Engelage, D. Reinhard and S. M. Huber, *Chemistry – A European Journal*, 2020, **26**, 3843-3861.

24. M. Eraković, D. Cinčić, K. Molčanov and V. Stilinović, *Angewandte Chemie International Edition*, 2019, **58**, 15702-15706.

25. A. M. S. Riel, D. A. Decato, J. Sun and O. B. Berryman, *Chemical communications (Cambridge, England)*, 2022, **58**, 1378-1381.

26. E. Margiotta, S. C. C. van der Lubbe, L. de Azevedo Santos, G. Paragi, S. Moro, F. M. Bickelhaupt and C. Fonseca Guerra, *Journal of Chemical Information and Modeling*, 2020, **60**, 1317-1328.

27. I. I. Rabi, J. R. Zacharias, S. Millman and P. Kusch, *Physical Review*, 1938, **53**, 318-318.

28. N. Bloembergen, E. M. Purcell and R. V. Pound, *Physical review*, 1948, **73**, 679.

29. F. Bloch, *Physical review*, 1946, **70**, 460.

30. T. D. Claridge, *High-resolution NMR techniques in organic chemistry*, Elsevier, 2016.

31. M. Karplus, *The Journal of Chemical Physics*, 1959, **30**, 11-15.

32. M. Karplus, *Journal of the American Chemical Society*, 1963, **85**, 2870-2871.

33. A. Navarro-Vázquez, J. C. Cobas, F. J. Sardina, J. Casanueva and E. Díez, *Journal of Chemical Information and Computer Sciences*, 2004, **44**, 1680-1685.

34. C. A. G. Haasnoot, F. A. A. M. de Leeuw and C. Altona, *Tetrahedron*, 1980, **36**, 2783-2792.

35. L. Haasnoot, Altona, *Tetrahedron*, 1980, **36**, 2783.

36. M. P. Williamson, in *Spectroscopic Methods and Analyses: NMR, Mass Spectrometry, and Metalloprotein Techniques*, eds. C. Jones, B. Mulloy and A. H. Thomas, Humana Press, Totowa, NJ, 1993, pp. 69-85.

37. A. Pardi, M. Billeter and K. Wüthrich, *Journal of Molecular Biology*, 1984, **180**, 741-751.

38. H. Kessler, C. Griesinger, J. Lautz, A. Mueller, W. F. Van Gunsteren and H. J. C. Berendsen, *Journal of the American Chemical Society*, 1988, **110**, 3393-3396.

39. V. F. Bystrov, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 1976, **10**, 41-82.

40. A. W. Overhauser, *Physical Review*, 1953, **92**, 411-415.

41. D. Neuhaus and M. Williamson, *New York, NY: VCH. xxii; ISBN: 978-0-471-24675-6*.

42. J. Keeler, *Understanding NMR spectroscopy*, John Wiley & Sons, 2010.

43. S. Macur, B. T. Farmer and L. R. Brown, *Journal of Magnetic Resonance (1969)*, 1986, **70**, 493-499.

44. S. Jurt and O. Zerbe, *Applied NMR spectroscopy for chemists and life scientists*, John Wiley & Sons, 2013.

45. C. R. Jones, C. P. Butts and J. N. Harvey, *Beilstein Journal of Organic Chemistry*, 2011, **7**, 145-150.

46. B. Vögeli, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2014, **78**, 1-46.

47. H. T. Hu and K. Krishnamurthy, *Journal of Magnetic Resonance*, 2006, **182**, 173-177.

48. S. Macur, B. T. Farmer and L. R. Brown, *Journal of Magnetic Resonance*, 1986, **70**, 493-499.

49. C. R. Jones, C. P. Butts and J. N. Harvey, *Beilstein journal of organic chemistry*, 2011, **7**, 145-150.
50. P. A. Jansson, *Deconvolution of images and spectra*, Courier Corporation, 2014.
51. P. Thepchatri, D. O. Cicero, E. Monteagudo, A. K. Ghosh, B. Cornett, E. R. Weeks and J. P. Snyder, *Journal of the American Chemical Society*, 2005, **127**, 12838-12846.
52. A. S. Jogalekar, F. H. Kriel, Q. Shi, B. Cornett, D. Cicero and J. P. Snyder, *Journal of Medicinal Chemistry*, 2010, **53**, 155-165.
53. J. P. Snyder, N. Nevins, D. O. Cicero and J. Jansen, *Journal of the American Chemical Society*, 2000, **122**, 724-725.
54. J. J. Koivisto, E. T. T. Kumpulainen and A. M. P. Koskinen, *Organic & Biomolecular Chemistry*, 2010, **8**, 2103-2116.
55. H. Andersson, H. Demaegdt, G. Vauquelin, G. Lindeberg, A. Karlén, M. Hallberg, M. Erdélyi and A. Hallberg, *Journal of Medicinal Chemistry*, 2010, **53**, 8059-8071.
56. E. Danelius, U. Brath and M. Erdélyi, *Synlett*, 2013, **24**, 2407-2410.
57. E. Danelius, H. Andersson, P. Jarvoll, K. Lood, J. Gräfenstein and M. Erdélyi, *Biochemistry*, 2017, **56**, 3265-3272.
58. H. Andersson, E. Danelius, P. Jarvoll, S. Niebling, A. J. Hughes, S. Westenhoff, U. Brath and M. Erdélyi, *ACS Omega*, 2017, **2**, 508-516.
59. F. Begnini, V. Poongavanam, Y. Atilaw, M. Erdelyi, S. Schiesser and J. Kihlberg, *ACS Medicinal Chemistry Letters*, 2021, **12**, 983-990.
60. P. Thepchatri, T. Eliseo, D. O. Cicero, D. Myles and J. P. Snyder, *Journal of the American Chemical Society*, 2007, **129**, 3127-3134.
61. V. Poongavanam, Y. Atilaw, S. Ye, L. H. E. Wieske, M. Erdelyi, G. Ermondi, G. Caron and J. Kihlberg, *Journal of Pharmaceutical Sciences*, 2021, **110**, 301-313.
62. D. O. Cicero, G. Barbato and R. Bazzo, *Journal of the American Chemical Society*, 1995, **117**, 1027-1033.
63. P. C. D. Hawkins, *Journal of Chemical Information and Modeling*, 2017, **57**, 1747-1756.
64. M. Erdélyi, V. Langer, A. Karlén and A. Gogoll, *New Journal of Chemistry*, 2002, **26**, 834-843.
65. C. Parish, R. Lombardi, K. Sinclair, E. Smith, A. Goldberg, M. Rappleye and M. Dure, *Journal of Molecular Graphics and Modelling*, 2002, **21**, 129-150.
66. T. Lewis-Atwell, P. A. Townsend and M. N. Grayson, *Tetrahedron*, 2021, **79**, 131865.
67. Schrödinger-Release, *Maestro, Schrödinger, LLC, New York, NY*, 2021.
68. J. J. Colin and H. Gaultier de Claubry, *Annales de Chimie*, 1814, **90**, 87-100.
69. R. L. Sutar and S. M. Huber, *ACS Catalysis*, 2019, **9**, 9622-9639.
70. J. Wolf, F. Huber, N. Erochok, F. Heinen, V. Guérin, C. Y. Legault, S. F. Kirsch and S. M. Huber, *Angew Chem Int Ed Engl*, 2020, **59**, 16496-16500.
71. P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 16789.
72. M. R. Scholfield, C. M. V. Zanden, M. Carter and P. S. Ho, *Protein Sci*, 2013, **22**, 139-152.
73. R. S. Czarny, A. N. Ho and P. Shing Ho, *The Chemical Record*, 2021, **21**, 1240-1251.
74. R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger and F. M. Boeckler, *Journal of Medicinal Chemistry*, 2013, **56**, 1363-1388.
75. P. S. Ho, *Journal*, 2017, **9**, 637-640.

76. P. Muller, *Pure and Applied Chemistry*, 1994, **66**, 1077-1184.
77. H. Jain, D. Sutradhar, S. Roy and G. R. Desiraju, *Angewandte Chemie International Edition*, 2021, **60**, 12841-12846.
78. C. Wang, D. Danovich, Y. Mo and S. Shaik, *Journal of chemical theory and computation*, 2014, **10**, 3726-3737.
79. W. S. Horne and T. N. Grossmann, *Nature chemistry*, 2020, **12**, 331-337.
80. M. A. Jiménez, in *Protein Design: Methods and Applications*, ed. V. Köhler, Springer New York, New York, NY, 2014, DOI: 10.1007/978-1-4939-1486-9_2, pp. 15-52.
81. E. Lenci and A. Trabocchi, *Chemical Society Reviews*, 2020, **49**, 3262-3277.
82. P. Morales and M. A. Jiménez, *Archives of Biochemistry and Biophysics*, 2019, **661**, 149-167.
83. S. H. Gellman, *Current Opinion in Chemical Biology*, 1998, **2**, 717-725.
84. J. D. Fisk and S. H. Gellman, *Journal of the American Chemical Society*, 2001, **123**, 343-344.
85. H. E. Stanger and S. H. Gellman, *Journal of the American Chemical Society*, 1998, **120**, 4236-4237.
86. T. S. Haque and S. H. Gellman, *Journal of the American Chemical Society*, 1997, **119**, 2303-2304.
87. J. F. Espinosa and S. H. Gellman, *Angewandte Chemie International Edition*, 2000, **39**, 2330-2333.
88. R. B. Merrifield, *Journal of the American Chemical Society*, 1963, **85**, 2149-2154.
89. E. Kaiser, R. L. Colescott, C. D. Bossinger and P. I. Cook, *Analytical Biochemistry*, 1970, **34**, 595-598.
90. M. Amblard, J.-A. Fehrentz, J. Martinez and G. Subra, *Molecular Biotechnology*, 2006, **33**, 239-254.
91. S. Peintner and M. Erdelyi, *Chemistry*, 2022, **28**, e202103559.
92. G. Wagner, A. Kumar and K. Wüthrich, *European Journal of Biochemistry*, 1981, **114**, 375-384.
93. N. Kobayashi, S. Honda, H. Yoshii and E. Munekata, *Biochemistry*, 2000, **39**, 6564-6571.
94. A. N. Analytical Methods Committee, *Analytical Methods*, 2016, **8**, 5553-5555.
95. S. Kappal, *London Journal of Research in Science: Natural and Formal*, 2019.
96. S. Peintner, MatLab script for peptide melting curve fitting, https://github.com/stepei/meltcurvefit).
97. R. R. Gil, *Angewandte Chemie International Edition*, 2011, **50**, 7222-7224.
98. A. Marx, B. Böttcher and C. M. Thiele, *Chemistry–A European Journal*, 2010, **16**, 1656-1663.
99. A. Marx and C. Thiele, *Chemistry – A European Journal*, 2009, **15**, 254-260.
100. G. Kummerlöwe, S. L. Grage, C. M. Thiele, I. Kuprov, A. S. Ulrich and B. Luy, *Journal of Magnetic Resonance*, 2011, **209**, 19-30.
101. L. Castañar, E. Sistaré, A. Virgili, R. T. Williamson and T. Parella, *Magnetic Resonance in Chemistry*, 2015, **53**, 115-119.
102. A. Navarro-Vázquez, *Magnetic Resonance in Chemistry*, 2012, **50**, S73-S79.
103. M. Řezanka, M. J. Langton and P. D. Beer, *Chemical Communications*, 2015, **51**, 4499-4502.
104. D. Mungalpara, S. Stegmüller and S. Kubik, *Chemical Communications*, 2017, **53**, 5095-5098.
105. R. J. Alaminsky and J. M. Seminario, *Journal of Molecular Modeling*, 2019, **25**, 160.

106. C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785-789.

107. B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chemical Physics Letters*, 1989, **157**, 200-206.

108. A. D. Becke, *Physical Review A*, 1988, **38**, 3098-3100.

109. R. A. Kendall, T. H. D. Jr. and R. J. Harrison, *The Journal of Chemical Physics*, 1992, **96**, 6796-6806.

110. D. E. Woon and T. H. D. Jr., *The Journal of Chemical Physics*, 1993, **98**, 1358-1371.

111. K. A. Peterson, B. C. Shepler, D. Figgen and H. Stoll, *The Journal of Physical Chemistry A*, 2006, **110**, 13877-13883.

112. V. Barone and M. Cossi, *The Journal of Physical Chemistry A*, 1998, **102**, 1995-2001.

113. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, 154104.

114. T. Lu and F. Chen, *Journal of computational chemistry*, 2012, **33**, 580-592.

115. J. Zhang and T. Lu, *Physical Chemistry Chemical Physics*, 2021, **23**, 20323-20328.

116. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Advanced drug delivery reviews*, 1997, **23**, 3-25.

117. A. L. Hopkins and C. R. Groom, *Nature Reviews Drug Discovery*, 2002, **1**, 727-730.

118. S. Surade and Tom L. Blundell, *Chemistry & Biology*, 2012, **19**, 42-50.

119. Bradley C. Doak, B. Over, F. Giordanetto and J. Kihlberg, *Chemistry & Biology*, 2014, **21**, 1115-1142.

120. E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, *Nature Reviews Drug Discovery*, 2008, **7**, 608-624.

121. D. A. DeGoey, H.-J. Chen, P. B. Cox and M. D. Wendt, *Journal of Medicinal Chemistry*, 2018, **61**, 2636-2651.

122. M. D. Shultz, *Journal of Medicinal Chemistry*, 2019, **62**, 1701-1714.

123. G. Caron, J. Kihlberg, G. Goetz, E. Ratkova, V. Poongavanam and G. Ermondi, *ACS Medicinal Chemistry Letters*, 2021, **12**, 13-23.

124. A. Alex, D. S. Millan, M. Perez, F. Wakenhut and G. A. Whitlock, *Medicinal Chemistry Communications*, 2011, **2**, 669-674.

125. G. Gramse, A. Dols-Perez, M. A. Edwards, L. Fumagalli and G. Gomila, *Biophysical Journal*, 2013, **104**, 1257-1262.

126. S. Şanli, N. Şanli and G. Alsancak, *Acta Chimica Slovenica*, 2010, **57**, 980-987.

127. C. R. W. Guimarães, A. M. Mathiowetz, M. Shalaeva, G. Goetz and S. Liras, *Journal of Chemical Information and Modeling*, 2012, **52**, 882-890.

128. M. Rossi Sebastiano, B. C. Doak, M. Backlund, V. Poongavanam, B. Over, G. Ermondi, G. Caron, P. Matsson and J. Kihlberg, *Journal of Medicinal Chemistry*, 2018, **61**, 4189-4202.

129. A. Whitty, M. Zhong, L. Viarengo, D. Beglov, D. R. Hall and S. Vajda, *Drug Discovery Today*, 2016, **21**, 712-717.

130. D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *Journal of Medicinal Chemistry*, 2002, **45**, 2615-2623.

131. C. Jeffrey, *Kew Bulletin*, 1986, **41**, 873-943.

132. C. Zdero, F. Bohlmann, R. M. King and L. Haegi, *Phytochemistry*, 1990, **29**, 509-511.

133. Y. Yang, L. Zhao, Y.-F. Wang, M.-L. Chang, C.-H. Huo, Y.-C. Gu, Q.-W. Shi and H. Kiyota, *Chemistry & Biodiversity*, 2011, **8**, 13-72.

134. P. P. Fu, Q. Xia, G. Lin and M. W. Chou, *Drug Metabolism Reviews*, 2004, **36**, 1-55.
135. F. Bohlmann, K.-H. Knoll, C. Zdero, P. K. Mahanta, M. Grenz, A. Suwita, D. Ehlers, N. L. Van, W.-R. Abraham and A. A. Natu, *Phytochemistry*, 1977, **16**, 965-985.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 2211

Editor: The Dean of the Faculty of Science and Technology

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022