UiT The Arctic University of Norway

Faculty of Science and Technology

Department of Chemistry

A computational study of cyclic peptides with vibrational circular dichroism

Karolina Di Remigio Eikås

A dissertation for the degree of Philosophiae Doctor – December 2022

Abstract

Cyclic peptides are a class of molecules that has shown antimicrobial potential. These are complex compounds to investigate with their large conformational space and multiple chiral centers. A technique that can be used to investigate both conformational preferences and absolute configuration (AC) is vibrational circular dichroism (VCD). To extract information from the experimental VCD spectra a comparison with calculated spectra is often needed and this is the focus of this thesis: the calculation of VCD spectra.

The VCD spectra are very sensitive to small structural changes, and to accurately calculate the spectra, all important conformers need to be identified. The first part of this thesis has been to establish a reliable computational protocol using meta-dynamics to sample the conformational space and *ab initio* methods to calculate the spectra for cyclic peptides.

Using our protocol, we have investigated if VCD alone can determine the AC of cyclic tetra- and hexapeptides. We show that it is possible to determine the AC of the cyclic peptides with two chiral centers while for the peptides with three and four chiral centers, VCD is at best able to reduce the number of possible ACs and further investigation with other techniques is needed.

Further, we investigated four cyclic hexapeptides with antimicrobial potential. These peptides, in contrast to the ones used for validating the protocol, consist of several amino acids with long and positively charged side chains. For these peptides, a molecular dynamics based approach provided VCD spectra in better agreement with experiment than our protocol. Reasons for this may be the lack of atomistic detail in the solvent model used during the conformational search and insufficient description of dispersion interactions during the meta-dynamics simulation.

Contents

List of Publications

This thesis is based on the following scientific publications.

I **K. D. R. Eikås**, M. T. P. Beerepoot and K. Ruud, "A Computational protocol for vibrational circular dichroism spectra of cyclic oligopeptides", *The Journal of Physical Chemistry A* **126** (2022), 5458–5471.

With belonging data set:

K. D. R. Eikås, M. T. P. Beerepoot and K. Ruud "Replication data for: A computational protocol for VCD spectra of cyclic peptides", 2022, DOI:10.18710/VITGV3, DataverseNO

My contribution: I performed all calculations and data analysis and wrote the initial draft of the computational details and results and discussion. I coordinated all revision stages, submitted to the journal, and handled the review process.

II **K. D. R. Eikås**, M. Krupová, T. Kristoffersen, M. T. P. Beerepoot and K. Ruud, "Can the absolute configuration of cyclic peptides be determined with vibrational circular dichroism?", Submitted.

My contribution: Major contribution to developing and investigating the research question. I performed all calculations and data analysis, wrote the initial draft of the paper, and coordinated all revision stages. Submitted to the journal.

III **K. D. R. Eikås**, T. Kristoffersen, M. R. Silk, J. S. M. Svendsen, K. Ruud and M. Krupová, "Conformational studies of cyclic hexapeptides with vibrational circular dichroism", Manuscript in preparation.

My contribution: Contributed to developing and investigating the to developing and investigating the research question. Together with M. Krupová, performed all calculations, data analysis and wrote an initial draft of the paper.

Acknowledgements

Finally, after years in the Arctic using countless CPU hours, my PhD journey comes to an end. Although there have been some dark hours, most of them have been full of laughter and excitement. And here is why.

Thank you, Kenneth, my main supervisor, for letting me steer my own PhD project. I am grateful that I have gotten the freedom to explore the field of chiroptical spectroscopy and that you have let me move in the direction I found more interesting in my own pace. Thank you, Maarten, my supervisor, for always having your door open. For helping me structuring my work, becoming a better writer, and developing my teaching skills. It has been great getting to know you and your family.

I am grateful that I have been allowed to do my PhD at the Hylleraas Center. The center has shown me the power of computational chemistry and the wide range of scientific problems it can be used to solve. At the center I was introduced to colleagues that turned into friends: Magnar, Marius, Lukas, Michal, and all the others. Thank you for all the scientific discussions, the jokes, the hikes, and the food.

I am also happy that I have been part of DigiBiotics. This is the team that has showed me the importance of transdisciplinarity when solving complex problems. I have learned a lot from you all. I am so grateful that I have shared my PhD journey with Tone and Laura, for all the lunches, the skiing and for letting me get to know your families. Thank you, Monika, for joining our team. I would not have made it without your help and motivation the last year. Thank you, Digital Life Norway, for the conferences, the courses, and the internship you provided me.

Finally, thank you to my family. To my parents and my brothers traveling all the way to Tromsø to spend time with me. To my husband, Roberto, for doing all the dishes and always being there for me, your support and help has been invaluable.

Abbreviations

The following abbreviations are used in the text. They are all defined in the text, but this list allows for easy reference.

CC	coupled cluster
CI	Configuration interaction
CVs	collective variables
DFT	density-functional theory
DFTB	density functional based tight-binding
ECD	electronic circular dichroism
ECP	effective core potential
FWHM	full width at half maximum
GBSA	generalized Born with surface area
GC	generic Z-matrix crossing
GGA	generalized gradient approximation
GTOs	Gaussian type orbitals
IR	infrared
KS	Kohn–Sham
LDA	local density approximation
LJ	Lennard–Jones
MD	molecular dynamics
MM	molecular mechanics
MP2	Møller-Plesset 2nd order
MP4	Møller-Plesset 4th order
MTD	meta-dynamics
NMR	nuclear magnetic resonance
PCM	polarizable continuum model
PES	potential energy surface
PWs	plane waves
QM	quantum mechanics
QM/MM	quantum mechanics / molecular mechanics
RMSD	root-mean square deviation
ROA	Raman optical activity
SAS	surface accessible surface

SCF	self-consistent field
SCRF	self-consistent reaction field
SES	surface excluded surface
STOs	Slater type orbitals
VCD	vibrational circular dichroism
xc	exchange-correlation
vdW	van der Waals
ZPVC	zero-point vibrational correction

Introduction

Spectroscopy is the study of how light interacts with matter and concerns phenomena such as absorption, emission and scattering of light by matter. In chemistry, spectroscopy is used to extract information about a molecular system such as what elements are present, the structure of the molecule and intermolecular interactions within the molecule. One widely used type of spectroscopy is Infrared (IR) spectroscopy. While other types of spectroscopies, such as nuclear magnetic resonance (NMR) and different types of X-ray spectroscopy detect individual atoms specifically, IR can identify bonds present in a molecule and is ideal for detecting functional groups in organic molecules and peptides.

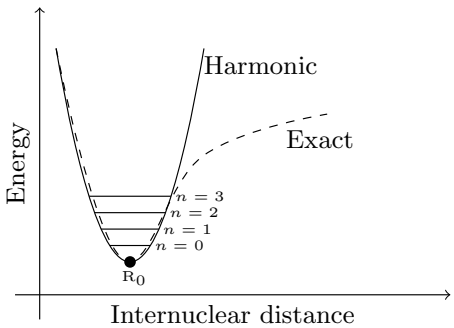


Figure 1. Illustration of the vibrational states of a molecule with a harmonic and exact potential. R_0 is the equilibrium geometry and $n = 0$ enumerates the vibrational states belonging to the electronic ground state.

Even at 0K, molecules are not rigid. Bonds and angles between atoms in the molecule oscillate and these oscillations are termed molecular vibrations[1]. A molecule can be in different vibrational states and following the laws of quantum mechanics (QM) the energy of these states is quantified

where the lowest energy state is called the ground vibrational state, shown as $n = 0$ in Figure 1. In IR spectroscopy the light shined on the sample results in vibrational excitations in the molecule, moving from $n = 0$ to $n = 1$ (or higher) in Figure 1, IR is also called vibrational spectroscopy.

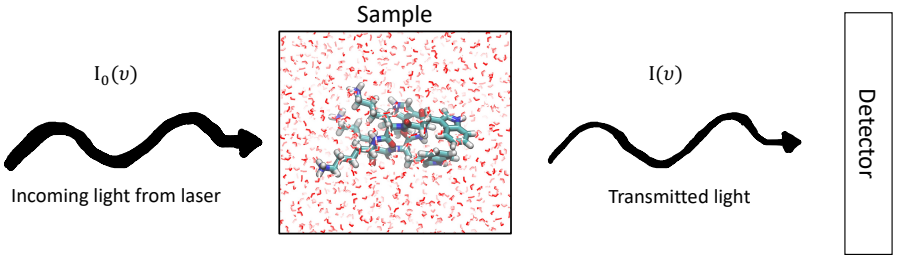


Figure 2. Illustration of absorption of light in IR spectroscopy. $I_0(\upsilon)$ and $I(\upsilon)$ are the intensity of the incoming and transmitted light, respectively, with frequency υ.

IR spectroscopy measures the absorption of light due to vibrational exitations. Light with a given intensity, I, at a given frequency or wavenumber, υ, is shined on the sample; some of the light is absorbed and induces vibrations in the sample. The transmitted light is measured as illustrated in Figure 2. The IR spectrometer measures the transmitted light over a range of different frequencies and the generated IR spectrum is the absorption as a function of the frequencies, as shown in Figure 3 where the absorption, $A(\upsilon)$, is calculated as[2]

$$A(\upsilon) = -\log_{10}\frac{I_0(\upsilon)}{I(\upsilon)}. \tag{1}$$

Molecules are either chiral or achiral. An achiral molecule is identical to its mirror image. Chiral molecules, on the other hand, are not identical to their mirror image (Figure 4). A molecule can be defined as chiral if *its mirror image cannot be superimposed on itself*[2]. A typical chiral molecule has a carbon as the chiral center surrounded by four different side groups as alanine, shown in Figure 4. The two forms of alanine have identical bond lengths and angles but they cannot be superimposed on each other. The two forms are enantiomers and the chiral center can be specified by the letter R for rectus (right) or S for sinister (left) according to the Cahn-Ingold-

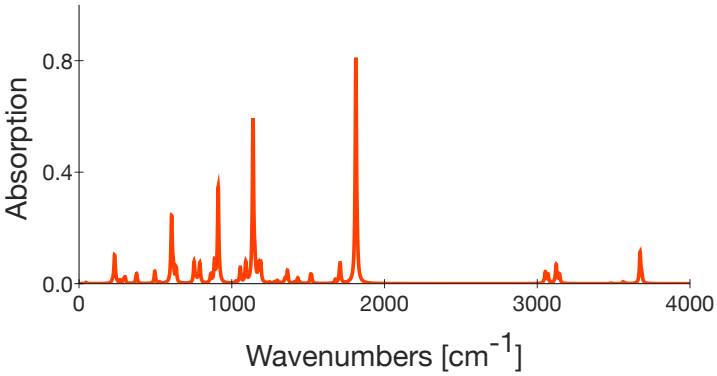


Figure 3. Calculated IR spectrum of of Alanine in vacuum. The spectrum is calculated using the computational protocol validated in **Paper I** using CREST[3–5] for the conformational search and B3LYP[6–8]/6-31+G*[9,10] for energies, geometry optimizations and spectral properties.

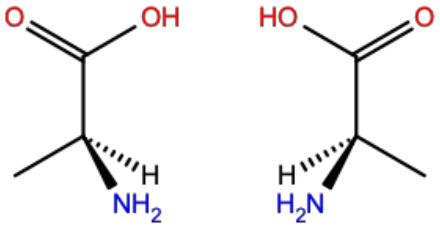


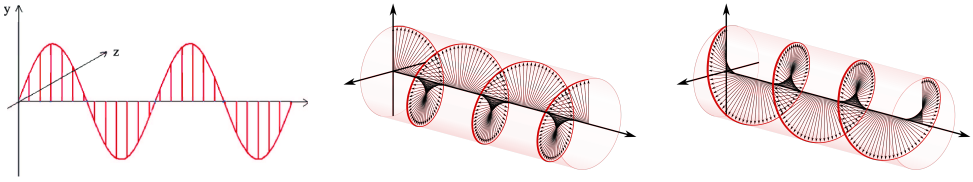
Figure 4. Both enantiomers of Alanine: the one with R configuration to the left and the one with S configuration to the right.

Prelog system[11,12]. To determine the absolute configuration is to determine the spatial arrangement of the atoms of a chiral molecular entity (or group) and its stereochemical description *e.g.* R or S.[13] There are other systems to classify the chiral center, such as the method using L and D suggested by Fischer.[14] The latter approach is commonly used for amino acids, sugars and peptides, but is less general than the R and S system that will be used throughout this thesis.

Although chiral molecules may have different functions in a biological system, they share most of their physical properties, including melting points, molecular weights and bond angles and distances. Since IR spectroscopy gives information about bond lengths and bond angles, two enantiomers with the same bond lengths, bond angles and dihedral angles have an identical IR spectrum. Hence, other types of spectroscopies have

to be used to distinguish two enantiomers.

By letting the linearly polarized light (Figure 5a) used for IR pass through an optical device, the light can be converted into circular polarized light. Circularly polarized light exists in two forms: left circularly polarized light (Figure 5b) and right circularly polarized light (Figure 5c). While linearly polarized light waves oscillates sinusoidally in a plane containing the propagation direction, the circularly polarized light waves rotate helically about the propagation direction as illustrated in Figure 5.



(a) Linearly polarized light.

(b) Left circularly polarized light as defined from the point of view of the receiver.

(c) Right circularly polarized light as defined from the point of view of the receiver.

Figure 5. Illustrations of linearly and left- and right-circularly polarized light. The illustration of linearly polarized light is from Chattopadhyay and Bharti[15] while the circularly polarized light illustrations are from Wikipedia: Circular Polarization[16].

Absorption, or scattering, of circularly polarized light can be measured and the corresponding spectroscopies are called vibrational optical activity (VOA). VOA can broadly be defined as the difference in interaction between left and right circularly polarized light with a molecular assembly undergoing vibrational transitions[2]. The chiroptical analogue of IR is named vibrational circular dichroism (VCD). Another type of VOA is Raman Optical Activity (ROA). ROA measures the scattering of circularly polarized light and is a complementary technique to VCD. In a sample consisting of randomly organized achiral molecules, the interaction with left or right circularly polarized light will be the same. If a sample consists of randomly organized chiral molecules, on the other hand, the interaction with left or right circularly polarized light will differ. Hence, VOA can be used to distinguish between enantiomers. In the papers included in this thesis, the only VOA method used is VCD. But, ROA is pointed out as a reasonable choice for further investigation of the compounds.

IR measures the absorption of linearly polarized light as

$$A = \frac{1}{2}(A_L + A_R) \tag{2}$$

while VCD measures the difference in absorption of left circularly polarized light (A_L) and right circularly polarized light (A_R) as[1]

$$\Delta A = A_L - A_R. \tag{3}$$

The resulting VCD spectrum is obtained by measuring the difference in absorption over a range of frequencies. For two enantiomers, the difference in absorbance is equal in magnitude at each light frequency but with opposite sign. Hence, the VCD spectra of two enantiomers are mirror images of each other as shown in Figure 6. The frequency range measured depends on the detector, but in general it is possible to measure VCD from approximately 600 cm^{-1} to 14 000 cm^{-1}, while the fundamental vibrations, defined in Section 3.2, are in the frequency range up to 4000 cm^{-1}.[2] In this thesis, cyclic peptides are investigated and information about the C=O stretches in the amide I region (1800-1600 cm^{-1})[17], the N-H bending and C-N stretching in the amide II region (1600-1480 cm^{-1})[17] and more complex backbone and side group vibrations in the amide III region (1350-1250 cm^{-1})[17] are considered.

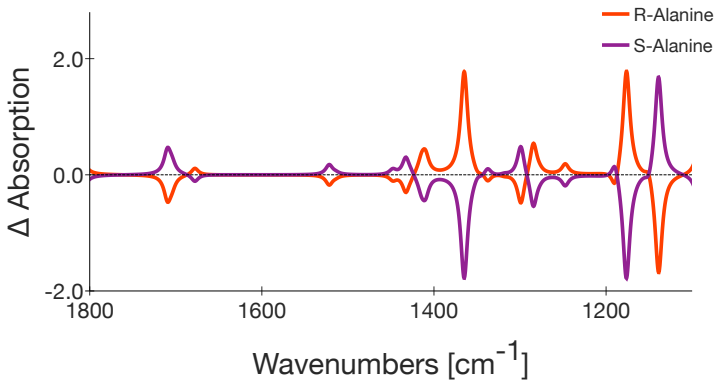


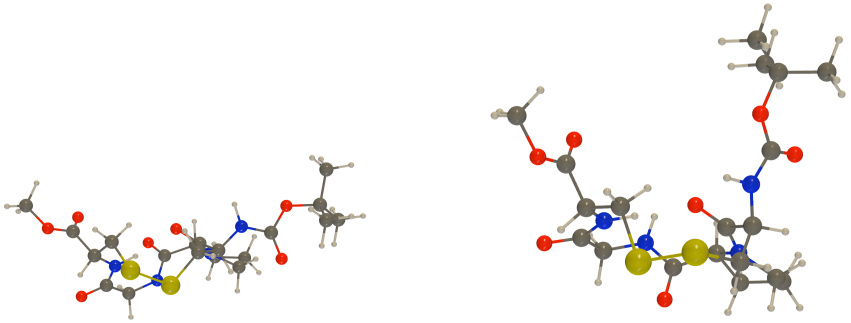
Figure 6. Calculated VCD spectra of R (red) and S (purple) Alanine in vacuum. The spectra are calculated using the computational protocol validated in **Paper I** using CREST[3–5] for the conformational search and B3LYP[6–8]/6-31+G*[9,10] for energies, geometry optimizations and spectral properties.

As seen from Figure 6 the sign of the bands is unique for the chiral element and for distinguishing the enantiomers. However, it is not possible to connect the sign of the VCD absorption *a priori* to one of the two enantiomers. To extract such information from the measured spectra a comparison with spectra simulated with electronic structure methods is important and this is the focus of this thesis: the simulation of VCD spectra.

Cyclic oligopeptides are a class of molecules that has shown potential as next-generation antibiotics and these are the class of compounds investigated in the papers included in this thesis. While alanine has one chiral atom, cyclic oligopeptides consist of several amino acids, with the possibility of multiple chiral centers. The flexibility of cyclic oligopeptides further complicates the simulation of their spectra. While IR spectroscopy measures the absorbance and the scattering intensity, VCD measures the difference in absorbance which makes VCD much more sensitive to molecular structure than its parent method.

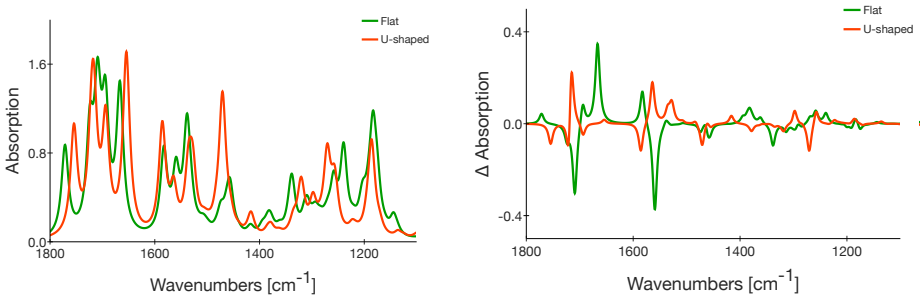
These small changes in geometry are usually referred to as conformational changes. Conformations are the three-dimensional structures a molecule can adopt. Traditionally, conformations of a molecule are defined as those arrangements of the atoms in space that can be interconverted by only rotation of single bonds.[18] Since small distortions in bond lengths and bond angles often accompany conformational changes, the traditional definition is relaxed to also include these distortions. The set of possible conformers of a molecule is called its conformational space.

In Figure 7 two different conformers of Cyclo(Boc-Cys-Pro-Gly-Cys-OMe) investigated in **Paper I** and **Paper II** are shown and in Figure 8 their corresponding IR and VCD spectra are shown. Although the differences in geometry are rather large between the two conformers, the differences in their IR spectra are small. The differences in the VCD spectra, on the other hand, are quite large. Not only the intensity of the bands differ, but also the signs. This has two consequences relevant for the work done in this thesis: first, to extract information about the chiral elements, the simulated spectra needs to be based on the relevant conformers present in the sample. Hence, the preferred conformations have to be identified. The other consequence is that if the conformers are needed to predict the spectra, also information about the conformers of the molecule can be extracted from the spectrum.



(a) Flat conformer. **(b)** U-shaped conformer

Figure 7. Two different conformers of Cyclo(Boc-Cys-Pro-Gly-Cys-OMe) where all chiral elements have the same configuration. The illustrations are made with data from **Paper I**.



(a) IR spectra **(b)** VCD spectra

Figure 8. Calculated IR and VCD spectra of the two different conformers of Cyclo(Boc-Cys-Pro-Gly-Cys-OMe) shown in Figure 7. All chiral elements have the same configuration for the two conformers. The spectra are simulated with data used in **Paper I**.

Although VCD has been widely used to investigate flexible molecules in recent years[19–24], there is a need for a more systematic approach to sample the conformational space and for robust approaches to simulate the spectra of flexible molecules with multiple chiral centres. The aim of this thesis is to establish a reliable protocol to calculate VCD spectra of flexible cyclic oligopeptides and then use the protocol in combination with experimental available data to extract information about cyclic peptides such as their absolute configuration and information about their preferred conformations. Most of the work has been focused on establishing a reliable approach to

explore the conformational space and ensure that the relevant conformers are included in the simulated spectra.

To serve as a background to the methods used in the papers, the following chapters are written with the hope that they will be useful for a student entering the field of VOA. Some basic knowledge of spectroscopy and quantum mechanics is assumed and the chapters are written to fit the knowledge level of a student starting their master studies. The order of the chapters is such that it is structured in the same way as the pipeline to simulate VCD spectra: First, approaches to sample the conformational space are presented in Chapter 1, then an introduction to electronic structure theory used to predict accurate geometries and energies is given in Chapter 2, followed by the theory needed to simulate VCD spectra using electronic structure methods, summarized in Chapter 3. In Chapter 4 a summary of the work is given together with a perspective on possible directions on future research.

Chapter 1

Conformational Search

There are different methods available to sample the conformational space and among them we find systematic searches[25–27], molecular dynamics (MD)[18] and meta-dynamics (MTD)[28]. For small molecules with only a few non-rigid dihedral angles, a systematic scan changing these variables can be a straightforward task. But for larger systems, with hundreds of dihedral angles, like the cyclic peptides investigated in the papers included in this thesis, it is a time consuming and error-prone task and it is common to simulate the system over time and analyse the changes that occur, *e.g.* with MD and MTD.

To run a MD or MTD simulation, a description of the particles and the interactions between them is needed. This can be done by molecular mechanics (MM), semi-empirical methods or quantum mechanics (QM). In MM the smallest particle is the atoms while in QM methods it is the electrons and nuclei. Hence, for a given molecular system, MM is a computationally much cheaper method than QM but also, in general, less accurate since it ignores the electrons. Semi-empirical methods are in between MM and QM, both when it comes to computational cost and, in general, accuracy. In **Paper III**, the MD simulations are performed with the computationally cheaper MM and the first section in this chapter is devoted to MM before it is shown how Newton's laws of motion can be used to simulate a molecular system over time, first with MD and then with MTD. The MTD simulations performed in the papers included in this thesis uses a tight-binding DFT method to describe the molecular system. To give the general outline of MD and MTD, the description of the molecular system

itself is not necessary and in this chapter MTD is presented (Section 1.3), while details about tight-binding DFT are given in the next chapter (Section 2.3.3.3). The MTD simulations in the papers included in this thesis are performed with the conformer-rotamer ensemble sampling tool (CREST). Details about CREST are given in the last section of this chapter.

1.1 Molecular mechanics

In molecular mechanics, the electronic motion is ignored and the energy is described by functions of the nuclear coordinates and parameters such as bond lengths, bond angles and charges. By ignoring the electrons and only taking the parameterized atoms into account, properties depending on the electron density of a molecule, such as ECD and NMR, cannot be provided. However, molecular mechanics is an acceptable compromise between computational cost and accuracy to do calculations on systems of thousands atoms and for sampling large ensembles of conformers.

The explicit expression for the energy in MM is given by a so-called force field. The force field consists of an algorithm and a collection of necessary parameters to calculate the energy. The force field algorithm is written as the sum of different energy terms and all force fields have at least three bonded energy terms; bond stretching, angle bending and dihedral angles, and two non-bonded terms; electrostatics and van der Waals interactions:

$$E = \underbrace{E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}}_{\text{Bonded}} + \underbrace{E_{\text{el}} + E_{\text{vdW}}}_{\text{Non-bonded}}. \tag{1.1}$$

Some common force fields are AMBER[29], CHARMM[30] and OPLS,[31–33] which are all developed with the aim of studying biomolecular systems. Different force fields might differ in both their description of the energy terms in Equation 1.1 and how they are parameterized. In **Paper III** the RSFF1[34] and the OPLS-AA/L[33] force fields are used and these are chosen as examples in this section. In the following, the functional description of each energy term is presented and different approaches to obtain the parameters are given.

The simplest way to model the bond stretching term and the angle bending is to use the harmonic potential and they are given as

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} k_b (R - R_0)^2 \tag{1.2}$$

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{1.3}$$

where k_b and k_θ are the force constants, R and θ are the bond lengths and angles and R_0 and θ_0 are the equilibrium bond lengths and angles. The harmonic approximation is simple and only valid for small displacements. To also be able to describe large displacements, the Morse potential can alternatively be used. However, in many force field and in both OPLS-AA/L and RSFF1, the harmonic approximation is used because it is computationally more efficient and most displacements in bond lengths and angles in molecules are small from the equilibrium geometry at room temperature.[35]

Since a rather large amount of energy is needed to create displacements in bond lengths and angles, most of the variation in structure and relative energies are due to the interactions between the dihedral angles and the non-bonded interactions.[18] The rotation of a bond can lead to rather large displacement from the equilibrium geometry and using harmonic potentials as for bond lengths and angles is not a good choice. Another consideration that has to be taken into account is the periodicity of the angle, *i.e.* that when a bond is rotated 360° the energy should return to the same value.[36] To account for this, dihedral angles are in most cases written as cosine series expansions, and one functional form is[18]

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} \sum_{n=0}^{N} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \tag{1.4}$$

where V_n is the rotational barrier with periodicity $\frac{360°}{n}$, ω is the dihedral angle, and γ is the phase factor. In both OPLS-AA/L and RSFF1, variants of Equation 1.4 are used. The number of terms included depends on the type of dihedral that is modeled. In the case of ethylene, the periodicity has to be 180° and this can be modeled by N=2 where V_0 and V_1 are zero.

The electrostatic energy between nuclei J and I is usually modeled by the Coulomb expression[36]

$$E_{\text{el}} = \sum_{I<J} \frac{q_I q_J}{\varepsilon R_{IJ}} \tag{1.5}$$

where q_I and q_J are the partial atomic charge on nuclei I and J, R_{IJ} is the distance between nuclei I and J, and ε the effective dielectric constant. The partial atomic charge is restricted to the nuclear center and gives the possibility of unequal distribution of the charge in a molecule. In both OPLS-AA/L and RSFF1 the α-carbon in an amino acid has a partial charge of 0.14 while the β-carbon has a partial charge of -0.15. The electrostatic contribution is included for all pairs of atoms separated by three or more bonds. The interaction between atoms that are closer, atoms one and two bonds apart, is already included in E_{bond} and E_{angle}. In the case where atoms are three bonds apart, the particular E_{el} term is scaled by a factor of 0.5 while interactions between atom pairs further apart are not scaled.[32]

The van der Waals interactions are usually modeled by the Lennard-Jones potential:[36]

$$E_{\text{vdW}} = 4\epsilon \sum_{I<J} \left[\left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{6} \right] \tag{1.6}$$

where σ_{IJ} is the collision parameter and ϵ_{IJ} is the depth of the minimum, written as: $\sigma_{IJ} = \sqrt{\sigma_{II}\sigma_{JJ}}$ and $\epsilon_{IJ} = \sqrt{\epsilon_{II}\epsilon_{JJ}}$. As for the electrostatic contribution, only interactions between atoms separated by three bonds or more are included and interactions involving atoms three bonds apart are scaled.

The parameters for the bond lengths, the angles and the dihedral angles are usually derived from electronic structure methods. This is also the case for the non-bonded electrostatic interactions where a common approach is to fit them to the electrostatic potential derived from the electronic wave function.[36] The van der Waals interactions on the other hand, is demanding to calculate with electronic structure methods. These parameters are therefore usually determined by fitting to experimental data. The OPLS-AA/L van der Waals parameters are derived by running a number of Monte Carlo simulations and fit them to experimental data on heat of vaporization, molecular volume and C—C radial distribution functions.[37]

The difference between OPLS/AA-L and RSFF1 are the parameters used for the dihedrals. RSFF1 is developed for proteins and the dihe-

dral parameters for the backbone and the side chains are, in contrast to OPLS/AA-L, residue-specific.[34] These residue-specific parameters are obtained by fitting each residue to data from the protein coil library[38] as described by Jiang, Zhou and Wu[34].

Not only is the type of element important in the parameterization, also the environment has to be taken into account. Both the neighbouring atoms and the type of molecule the atom is in might influence the parameteriztion. As mention above, in OPLS-AA/L and RSFF1[34], the α- and the β-carbon have different partial charge. Most force fields have a range of different atomtypes for each atom: OPLS has in total more than 800 different atomtypes parameterized, mainly different atomtypes of C, H, O and N but also the halogens, alkali metals and some other elements are available.[31–33,37] With classical approaches to calculate the energies and the forces in place, the next chapters describes how this can be used to investigate how the molecular system evolves in time.

1.2 Molecular dynamics

The basic idea behind MD simulations is to propagate a representative system over time. In molecular systems, nuclei are heavy enough to be classically described and their movement can be described by Newton's second law. Newton's second law says that the acceleration $\mathbf{a}_I$ on atom I with the mass m_I along direction $\mathbf{x}_I$ experiencing a force $\mathbf{F}_I$ can be written as[18]

$$\mathbf{a}_I = \frac{d^2\mathbf{x}_I}{dt^2} = \frac{\mathbf{F}_I}{m_I} \tag{1.7}$$

where $\mathbf{a}_I$, $\mathbf{x}_I$ and $\mathbf{F}_I$ are vectors containing x-, y- and z-components. Thus, the change in position of the atoms ($\mathbf{x}_I$) can be calculated by integrating the force $\mathbf{F}_I$ over time. In a realistic model of a system consisting of many atoms, the force acting on each atom will change whenever the atoms position is changed or when any other atom it interacts with changes position. This gives rise to the many-body problem and Equation 1.7 cannot be solved analytically.[18] A common approach is therefore to use numerical integration with a finite time step Δt. There are many algorithms for solving the equation of motion with a finite time step. The algorithm most widely used to solve the equation of motion is the Verlet algorithm where the position

$\mathbf{x}(t)$ and the acceleration $\mathbf{a}(t)$ at time t is used together with the position at the previous time step $\mathbf{x}(t-\Delta t)$ to calculate the next position $\mathbf{x}(t+\Delta t)$ as[39]

$$\mathbf{x}(t+\Delta t) = 2\mathbf{x}(t) - \mathbf{x}(t-\Delta t) + \Delta t^2 \mathbf{a}(t). \tag{1.8}$$

The Verlet algorithm is straightforward to implement and only a modest use of memory is needed.[18,39] The next position is determined by the current position. Hence, it is not a self-starting algorithm and other approaches has to be used for the very first step.[35] Another drawback is that the velocities are not explicitly given. Many MD simulations are performed with constant temperature and since the temperature is related to the kinetic energy, which again is related to the velocities, a common approach to control the temperature is to scale the velocities of each particle at each step.[18]

There are several versions of the Verlet algorithm or related algorithms such as the velocity Verlet algorithm, the leap-frog algorithm and Beeman's algorithm.[18] The leap-frog uses the following relationships[40]

$$\mathbf{x}(t+\Delta t) = \mathbf{x}(t) + \mathbf{v}(t+\frac{1}{2}\Delta t)\Delta t \tag{1.9}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t-\frac{1}{2}\Delta t) + \mathbf{a}(t)\Delta t. \tag{1.10}$$

In contrast to the Verlet algorithm, the leap-frog algorithm gives the velocities explicitly. The downside of the leap-frog algorithm is that the position and velocity is not given at the same time: the positions are calculated at t, $t+\Delta t$, $t+2\Delta t$, *etc.*, while the velocities are calculated at $t+\frac{1}{2}\Delta t$, $t+\frac{3}{2}\Delta t$, $t+\frac{5}{2}\Delta t$ *etc.* As a result, the kinetic energy contribution, which depends on the velocities, and the potential energy, which depends on the positions, cannot be determined at the same time. In **Paper III** leap-frog as implemented in the GROMACS MD software[41] is the time propagation algorithm used.

Another important aspect of a MD simulation is the time step. The time step is determined by the fastest process in the system. Rotation and vibration in a molecule typically have a frequency in the range 3-3300 cm^{-1} which correspond to a time step in the order of femtoseconds or less.[36] To sample the conformational space for cyclic peptides in **Paper III**, a total

simulation time of 1 μs is needed. Using a time step of 1 femtosecond for this simulation requires 10^9 steps. The vibrations are faster processes than the rotations and the vibrations involving hydrogen are typically the fastest ones. Since these vibrations have little impact on many properties, a common approach to save computational time is to fix bonds lengths involving hydrogen. Depending on the algorithm used to solve the equation of motion, different constraint algorithms are available, SHAKE[42] (Verlet) and LINCS[43] (leap-frog) are two example and in **Paper III** LINCS is used. Enforcing the bond length constraints typically allows for 2 or 3 times longer time steps and in **Paper III** a time step of 2 femtoseconds is used which reduses the computational effort by one half compared wit using 1 femtosecond.

1.3 Meta-dynamics

In this work, MD is used as a tool to find all relevant conformers. For the result of the MD simulation to be meaningful, the simulation time has to be long enough to visit all energetically relevant configurations. In practice this is not always possible. One frequent problem is when two relevant configurations are separated by a high energy barrier and the transition between the two states only takes place when some rare fluctuation of the system occurs. In such case, a long simulation time is needed to ensure that this rare state is reached. In other words, it is computationally demanding. An approach developed to overcome the high energy barriers without increasing the simulation time is meta-dynamics (MTD)[28]. In the papers included in this thesis, MTD as implemented in CREST[4,5] is used and serves as an example in the following.

As in MD, MTD uses Newtons equation of motion to calculate forces to change the position of the atoms for a given time step. But, in contrast to a MD simulation, a MTD simulation "remembers" the configurations visited on the PES and "fills" the PES such that the system is discouraged to revisit those configurations (Figure 1.1). This "filling" is done by adding a history-dependent bias potential (or a force) that acts on selected degrees of freedom referred to as collective variables (CVs).[44] By introducing the bias potential, the sampling of rare events to overcome high energy barriers by pushing the system away from a local minimum are accelerated. In CREST

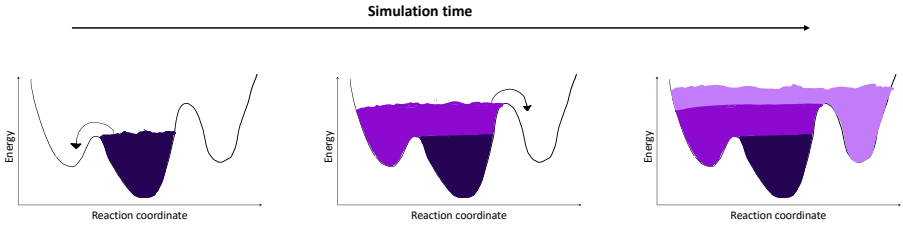


Figure 1.1. Illustration of a MTD simulation for a one-dimensional potential energy surface and how the bias potential is "filling" the potential energy surface to avoid revisiting of previous configurations. The figure is inspired by those by Grimme[4] and Barducci, Bonomi and Parrinello[44].

the bias potential is given as[4]

$$E_{bias} = \sum_{i=1}^{n} k_i e^{-\alpha \Delta_i^2} \tag{1.11}$$

where n is the number of reference structures, k_i is a positive constant discouraging the system to return to previous points, α the parameter that determines the width of the bias potential and Δ_i the CVs.

A critical point of the MTD is the choice of CV, which is in many cases not straightforward to identify. However, some guidelines exist: first, the CVs should distinguish between the initial and final state and describe all relevant intermediates. Secondly, the CVs should include all slow modes of the system, the variables defined as slow are those that cannot be sampled satisfactorily within the timescale of the simulation. Lastly, there should be a limited number of CVs.[44]

Frequently used CVs are among others bond lengths, angles and dihedral angles, dipole moment and the number of hydrogen bonds.[44] In CREST, the root-mean square deviation (RMSD) of atomic Cartesian coordinates is used and the CVs are given as[4]

$$\Delta_i = \sqrt{\frac{1}{N} \sum_{J=1}^{N} (R_j - R_j^{ref})^2} \tag{1.12}$$

where N is the number of atoms, R_J is the component of Cartesian space vector of atom J and R_J^{ref} is the corresponding atom J in the reference structure. In some cases, N does not run over all atoms in the system.

When investigating a catalyzed chemical reactions, excluding the atoms in the catalyst from the RMSD would help by only pushing the starting configuration of the substrate towards the product.[4]

In addition to running several MTD simulations, CREST runs MD simulations and compare the conformers in the MTD and MD ensembles. In the next section, more details about CREST is given.

1.4 Conformer-Rotamer Ensemble Sampling Tool

In the papers included in this thesis, the MTD simulations are performed using the conformer-rotamer ensemble sampling tool (CREST). The aim of CREST is to set up an automated routine that can be used as an exploratory and screening tool for systems containing hundreds of atoms.[4] This routine, illustrated in Figure 1.2, includes three different methods to sample conformers; MTD sampling, MD sampling and a generic Z-matrix crossing (GC) procedure.

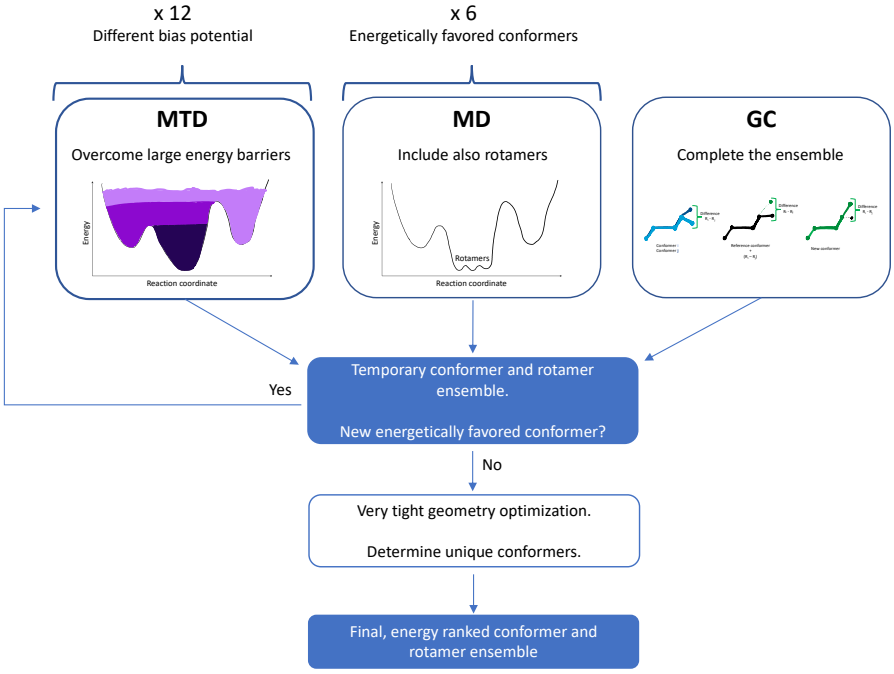


Figure 1.2. Illustration of the automated CREST routine.

In CREST, both a classical force field and a tight-binding DFT method are available for calculating the forces for the MTD and MD part. In the papers included in this thesis, the tight-binding DFT method (Section 2.4) is used for calculations performed with CREST. The CREST routine starts with 12 MTD simulations with different bias potentials before a MD simulation around the 6 MTD conformers lowest in energy is done. To help further complete the conformer ensemble, the GC procedure, illustrated in Figure 1.3, is included. This procedure is particularly useful for larger molecules with many weakly coupled rotatable functional groups, *e.g.* alkyl chains.[5] This algorithm uses internal coordinates (Z-matrix, R) and includes new structures according to

$$R_{new} = R_{ref} + (R_i - R_j) \tag{1.13}$$

where R_{new} is the energetically lowest-lying conformer and $(R_i - R_j)$ is the pairwise structural difference between conformers i and j. In this way, structural differences like a methyl rotation or flipping of a proline ring can be added to a reference structure and new rotamers and/or conformers added to the final conformer ensemble.

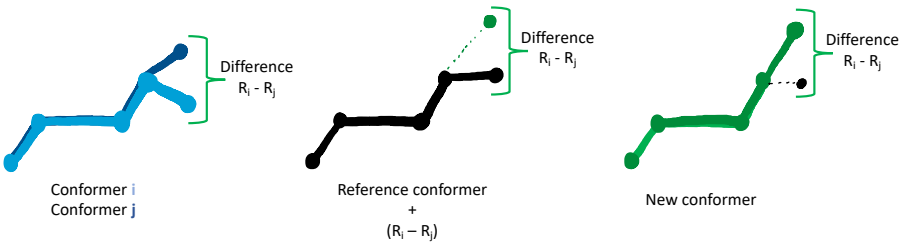


Figure 1.3. Illustration of how the generic z-matrix crossing compares two and two conformers, i and j, in the ensemble and add the difference between every two conformers to a reference conformer to create a new conformer.

The whole MTD, MD and GC procedure is done at least twice and up to five times in the case where either the MD or the GC finds a new conformer lower in energy than the one in the ensemble found with MTD. The final energy ranked ensemble serves as a starting point for more accurate calculations discussed in Chapter 2 and 3.

Chapter 2

Electronic structure theory

In the first chapter, the molecular system was described by molecular mechanics and it was shown how this simple description was useful in MD simulations to find relevant conformers. But, MM does not explicitly describe the electrons, which is needed to accurately calculate geometries, energies, IR and VCD properties. In this chapter, methods to describe molecular systems, including electrons, by solving the equation of quantum mechanics, are described.

In the first section, the non-relativistic and time-independent version of the Schrödinger equation is presented and it is shown how this equation gives rise to the potential energy surface (PES) by using the Born-Oppenheimer approximation. Thereafter the Hartree-Fock approximation is discussed before turning to density functional theory (DFT), tight-binding density functional theory (DFTB), basis sets and solvent modeling. The last section of this chapter is devoted to geometry optimizations.

2.1 The time-independent Schrödinger equation

The energy E of a system can be determined by solving the eigenvalue problem stated by the Schrödinger equation. In this thesis, the non-relativistic and time-independent Schrödinger equation is used,

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where the Hamiltonian consist of a kinetic part, $\hat{T}$, and a potential part $\hat{V}$. For a molecule, the kinetic part can be divided in a kinetic operator for the electrons, $\hat{T}_e$ and one for the nuclei, $\hat{T}_N$. The potential energy operator can be represented by a term describing the electron-electron repulsion, $\hat{V}_{ee}$, the electron-nuclei attraction, $\hat{V}_{en}$ and the nuclear-nuclear repulsion, $\hat{V}_{nn}$ and the total Hamiltonian of a molecular system can thus be written as[45]

$$\hat{H} = \hat{T} + \hat{V} = \hat{T}_N + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN}. \tag{2.2}$$

In addition to work with the time independent version of the Schrödinger equation and not including relativistic effects, the nuclei are treated as point charges and the two kinetic contributions are given as[45]

$$\hat{T}_e = -\sum_{i=1}^{n} \frac{1}{2} \nabla_i^2 \tag{2.3}$$

$$\hat{T}_N = -\sum_{I=1}^{N} \frac{1}{2M_I} \nabla_I^2 \tag{2.4}$$

with n number of electrons, N number of nucleus, M_I the mass of nuclei I, and the Laplacian operators ∇_i^2 and ∇_I^2 are the second derivatives with respect to the coordinates of the ith electron and the Ith nucleus.

The potential parts of the Hamiltonian are given as

$$\hat{V}_{ee} = \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{r_{ij}} \tag{2.5}$$

$$\hat{V}_{eN} = -\sum_{i=1}^{n} \sum_{I=1}^{N} \frac{Z_I}{r_{iI}} \tag{2.6}$$

$$\hat{V}_{NN} = \sum_{I=1}^{N} \sum_{J>I}^{N} \frac{Z_I Z_J}{R_{IJ}} \tag{2.7}$$

where r_{ij} is the distance between electrons i and j, r_{iI} the distance between electron i and nucleus I, Z_I the charge of nucleus I, and R_{IJ} the distance between nuclei I and J. Equation 2.3-2.7 is given in atomic units where the magnitude of the electron mass, the reduced Planck constant, the Coulomb constant and the electron charge are one. Atomic units are used throughout the thesis, if not specified otherwise.

An electron has a mass of 9.1×10^{-31} kg while protons and neutrons have a mass of approximately 1.7×10^{-27} kg. This means that the nuclei are at least 1800 times heavier than an electron and due to the large difference in mass, the electron has a much higher kinetic energy and moves much faster than the nuclei. The Born-Oppenheimer approximation makes use of this and states that due to the difference in mass, the motion of the electrons can be separated from the motion of the nuclei and the electrons can be considered moving in the field of the fixed nuclei.[45] Within this approximation, the wave function can be written as a product of the electronic wave function Ψ_e and the nuclear wave function Ψ_N where the nuclear wave function only depends on the nuclear coordinates and the electronic wave function depends on the electronic coordinates and only parametrically on the nuclear coordinates

$$\Psi(R, r) = \Psi_e(r; R)\Psi_N(R). \tag{2.8}$$

In a similar manner, also the Hamiltonian can be divided in a nuclear part ($\hat{T}_n$) and an electronic part

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN}. \tag{2.9}$$

In Equation 2.9 also the nuclear-nuclear repulsion is included, $\hat{V}_{NN}$, but this term is a constant for a given molecular geometry and will just add to an operator eigenvalue (E) and not affect the eigenfunctions (Ψ). The electronic Schrödinger equation can be written as

$$\hat{H}_e\Psi_e(r; R) = E_e(R)\Psi_e(r; R) \tag{2.10}$$

where the electronic energy, $E_e(R)$, when solved for a number of different geometries, creates the PES. The Born-Oppenheimer approximation is usually a good approximation but there are exceptions. For example, when the solutions are close in energy. Another example is VCD, a complete theoretical description of VCD is not possible within the Born-Oppenheimer approximation which is discussed in more details in Chapter 3.

Equation 2.10 can only be solved exact for one electron systems and in order to use the equation for more complex molecular systems other approximations and numerical methods have to be introduced. A starting

point for more sophisticated methods is the Hartree-Fock approximation which uses the variational principle to approximate a ground-state wave function.

2.2 The Hartree-Fock approximation

The essence of the Hartree-Fock method is to replace the many-electron problem with a one-electron problem where the electron-electron repulsion is treated in an average way. According to the Pauli principle, the wave function has to be antisymmetric: the wave function must change sign when the coordinates of two electrons are interchanged.[46] In Hartree-Fock, the simplest antisymmetric wave function, a Slater determinant, is used:

$$\Psi_{HF} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_n(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_n(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(\mathbf{x}_n) & \psi_2(\mathbf{x}_n) & \cdots & \psi_n(\mathbf{x}_n) \end{vmatrix} \tag{2.11}$$

where $\frac{1}{\sqrt{n!}}$ is the normalization factor for a system with n electrons, ψ_n is the one-electron wave function for electron n, including a spatial and a spin part, and $\mathbf{x}_n$ is the spatial and the spin coordinates for electron n.

The Hartree-Fock method make use of the variational principle stating that the expectation value of the Hamiltonian is an upper bound to the exact ground-state energy[45] and that the best wave function is the one that gives the lowest possible energy.[45] The one-electron wave functions ψ_i can be written as a linear combination of basis functions χ_α

$$\psi_i = \sum_{\alpha=1}^{M} c_{i\alpha}\chi_\alpha \tag{2.12}$$

where the M is the total number of basis functions and $c_{i\alpha}$ are the expansion coefficients. By applying the variational principle an optimal set of expansion coefficients is obtain by minimizing the energy.

By describing the wave function in Equation 2.10 by a Slater determinant (Equation 2.11) and minimizing the energy in line with the variational principle, the electronic Schrödinger equation (Equation 2.10) can be reduced to Hartree-Fock equations on the form

$$\hat{f}\psi_i = \epsilon_i \psi_i. \tag{2.13}$$

In the Hartree-Fock equation, $\hat{f}$ is the one-electron Fock operator given as

$$\hat{f} = -\frac{1}{2}\nabla^2 - \sum_{I=1}^{N} \frac{Z_I}{|r - R_I|} + v^{HF} \tag{2.14}$$

where v^{HF}, the Hartree-Fock potential, is the average potential electron i experiences due to the other electrons. The Hartree-Fock potential is the sum of the Coulumb operator $\hat{J}$ and the exchange operator $\hat{K}$

$$v^{HF} = \hat{J} - \hat{K} = \sum_{i=1}^{n} (\hat{J}_i - \hat{K}_i). \tag{2.15}$$

The Coulomb operator describes the electron-electron repulsion while the exchange operator does not have a simple classical interpretation, but arises due to the antisymmetry requirement of the wave function. The exchange operator is defined by its effect when operating on a one-electron wave function and for a better comparison, also the Coulomb operator is here given operating on the one-electron wave function ψ_j:

$$\hat{J}_i \psi_j(\mathbf{x}) = \left[\int \psi_i^*(\mathbf{x}') \frac{1}{|r - r'|} \psi_i(\mathbf{x}') d\mathbf{x}' \right] \psi_j(\mathbf{x}) \tag{2.16}$$

$$\hat{K}_i \psi_j(\mathbf{x}) = \left[\int \psi_i^*(\mathbf{x}') \frac{1}{|r - r'|} \psi_j(\mathbf{x}') d\mathbf{x}' \right] \psi_i(\mathbf{x}) \tag{2.17}$$

Because the Coulomb and exchange operators, and hence the Fock operators depend on all one-electron wave functions, all the ψ_j with $j \neq i$ need are needed to determine ψ_i. The equation is solved with an iterative procedure, which makes Hartree-Fock a self-consistent field (SCF) method. Starting from an initial guess, new one-electron wave functions are calculated until the one-electron functions no longer change, or more exactly; changes are smaller than a pre-established threshold.

The Hartree-Fock method is able to describe the majority of the electronic energy and gives a good estimate of the ground-state wave function. However, the approach describes the electrons moving in an average field generated by all other electrons and this description does not account for

electron correlation. There are several different methods to include electron correlation, e.g configuration interaction (CI), coupled cluster (CC) and Møller-Plesset perturbation theory of different order (MP2 and MP4). These methods are all wave-function based methods and use the Hartree-Fock wave function as a starting point and improve by including excited determinants to different extent. However, in this thesis, none of these methods are used, only Density Functional Theory (DFT), and only this method is therefore described in the next section.

2.3 Density functional theory

While wave function-based methods like CC, CI and MP2 systematically improve upon Hartree-Fock by including more and more excited determinants in the wave function, DFT replaces the wave function with the electron density. The great advantage is that the computational cost can be held at a decent level. But, DFT does not systematically improve accuracy in the same way as CI and CC. In the following, the fundamentals of DFT, the Hohenberg-Kohn theorems and Kohn-Sham theory are presented. The last section is devoted to the different DFT functionals; how to model the exchange-correlation part.

2.3.1 The Hohenberg-Kohn theorems

The idea behind DFT is to express the energy $E[\rho(r)]$ of an electronic system in terms of the density $\rho(r)$. While a wave function of a n electron system contains $3n$ variables, the density only depends on the 3 spatial coordinates independently of the number of electron. The first formal proof that the electron density determines the properties of a molecule was given by Hohenberg and Kohn in 1964.[47] They proved that there cannot exist two different external potentials which describe the same electron density. Hence, the electron density uniquely defines the external potential, the Hamiltonian, the wave function, and indeed all properties of the molecule.

Furthermore, to determine the electron density Hohenberg and Kohn formulated a variational theorem for the energy functional.[47] This theorem allows the best ground-state electron density to be determined by starting from a trial density and minimizing the energy in an analogues manner as

in Hartree-Fock.

2.3.2 Kohn-Sham density functional theory

With the Hohenberg-Kohn theorems, the energy can be expressed in terms of the density and written as the sum of the kinetic part $T[\rho(r)]$, the electron-electron interaction $W_{ee}[\rho(r)]$ and the external potential $V_{ext}[\rho(r)]$ [48]

$$E[\rho(r)] = T[\rho(r)] + W_{ee}[\rho(r)] + V_{ext}[\rho(r)] \tag{2.18}$$

where only one of them, $V_{ext}[\rho(r)]$, is known exactly. The external potential contains the nuclear-electron attraction and is in the absence of other external potentials given by

$$V_{ext}[\rho(r)] = -\sum_{I=1}^{N} \int \frac{\rho(r) Z_I}{|R_I - r|} dr \tag{2.19}$$

Although $T[\rho(r)]$ and $W_{ee}[\rho(r)]$ are exact, they are also unknown. In 1965 Kohn and Sham proposed to use a model system consisting of non-interacting electrons, thus represented exactly using a single Slater determinant. [49] The drawback is that orbitals are re-introduced, which increases the number of variables of a system with n electrons from 3 to $3n$. Despite this increase in the number of variables, the kinetic part is still easy to calculate and it gives a reasonably accurate kinetic energy, leaving just a small correction to the exact kinetic energy. For the electron-electron interaction, part of it can be described by the Coulomb repulsion as $J[\rho(r)]$ as

$$J[\rho(r)] = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr'. \tag{2.20}$$

The correction to the kinetic energy and remaining part of the electron-electron interactions is combined and denoted the exchange-correlation energy formally given as:

$$E_{xc}[\rho(r)] = (T[\rho(r)] - T_s[\rho(r)]) + (W_{ee}[\rho(r)] - J[\rho(r)]) \tag{2.21}$$

where $T_s[\rho(r)]$ is the kinetic energy of the non-interaction system. Combining Equation 2.18 and Equation 2.21 gives

$$E[\rho(r)] = T_s[\rho(r)] + J[\rho(r)] + V_{ext}[\rho(r)] + E_{xc}[\rho(r)]. \tag{2.22}$$

In this system with non-interacting electrons moving in an external potential, the Hamiltonian is separable and the one electron Kohn-Sham orbitals $\psi_i(r)$ can be determined by the Kohn-Sham equation

$$\Big(-\frac{1}{2}\nabla^2 + v_{eff}(r)\Big)\psi_i(r) = \epsilon_i \psi_i(r) \tag{2.23}$$

where the effective potential is

$$v_{eff}(r) = v_{ext}(r) + \frac{\delta J[\rho(r)]}{\delta \rho(r)} + \frac{\delta E_{xc}[\rho(r)]}{\delta \rho(r)}. \tag{2.24}$$

From the Kohn-Sham orbitals, the electron density can be calculated as

$$\rho(r) = \sum_{i=1}^{n} |\psi_i(r)|^2. \tag{2.25}$$

As for Hartree-Fock (Section 2.2), the Kohn-Sham approach is a SCF method. First, a set of initial Kohn-Sham orbitals are used as input. Secondly, a new set of Kohn-Sham orbitals are obtained by solving Equation 2.23 using an approximate exchange-correlation functional $E_{xc}[\rho(r)]$. Then the electron density can be calculated using Equation 2.25 and a new v_{eff} can be constructed. With the new v_{eff}, Equation 2.23 is again solved and this procedure is repeated until convergence. When converged, an optimal electron density can be calculated from the occupied Kohn-Sham orbitals using Equation 2.25, and the energy from Equation 2.22.

Kohn-Sham theory is formally exact: If the exact form of the exchange-correlation functional $E_{xc}[\rho(r)]$ was known, the exact electronic energy of the interacting system could be obtained. In practice, an approximate exchange-correlation functional is used, and there is a range of them to choose between and the ones used in this thesis will be presented in the next section.

2.3.3 Exchange-correlation functionals

There is no systematic way to improve the DFT calculations to an exact limit and, due to the approximate form of the exchange-correlation func-

tional, the calculated DFT energy is not an upper bound to the true energy. It is common to separate the exchange-correlation energy into a pure exchange and a pure correlation part as

$$E_{xc}[\rho(r)] = E_x[\rho(r)] + E_c[\rho(r)]. \tag{2.26}$$

In this section, the local density approximation (LDA) (Section 2.3.3.1), the generalized gradient approximation (GGA) (Section 2.3.3.2) and hybrid functionals (Section 2.3.3.3) will be discussed.

2.3.3.1 Local density approximation

The local density approximation (LDA) assumes that the density is slowly varying and is exact for a uniform electron gas with the exception for small errors due to interpolation formulas chosen for the correlation energy. The approach only depends on the electron density, $\rho(r)$, and the exchange energy of a uniform electron gas is given by the Dirac expression [50]

$$E_x^{\mathrm{LDA}}[\rho(r)] = -\frac{9}{8}\alpha\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(r)^{\frac{4}{3}} dr. \tag{2.27}$$

where α is an adjustable parameter often chosen to be $\frac{2}{3}$. [46]

The most common LDA approach is SVWN. The exchange part, S, is using Slater's X_α method [51], which is based on the Dirac expression (Equation 2.27) while the correlation part, VWN, was derived by Vosko, Wilk, and Nusair [52]. The VWN correlation functional has been interpolated as an analytical formula based on the results from quantum Monte Carlo simulations of a uniform electron gas. Another well established LDA correlation functional, which is also based on the Monte Carlo simulation energies, is the Perdew and Wang (PW) functional [53].

Despite the LDA methods simplicity, it has similar accuracy at Hartree-Fock and is widely used in the physics community for metals where the assumptions of a slowly varying density is reasonable. [36] In chemistry, on the other hand, the density is not slowly varying at the nuclei and it is not commonly used as it overestimates bond strengths significantly. [48]

2.3.3.2 Generalized gradient approximation

The assumption of a slowly varying density is not true for molecular systems. To improve upon LDA, the gradient of the electron density, $\nabla\rho(r)$, is included in the exchange-correlation functionals in addition to the electron density and the approach is called generalised gradient approximation (GGA). In practice, the dimensionless reduced density gradient [48]

$$x(r) = \frac{\nabla\rho(r)}{\rho^{\frac{4}{3}}} \tag{2.28}$$

is often used. Becke suggested in 1986 the exchange functional form [54]

$$E_x^{\mathrm{B86}}[\rho(r)] = \int \rho^{\frac{4}{3}} \left(C_x + \frac{\beta x(r)^2}{1 + \beta x(r)^2} \right) \tag{2.29}$$

which is used in the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional [55]. A couple of years later, Becke suggested the widely used B88 exchange functional as a correction to the LDA exchange energy [56]

$$E_x^{\mathrm{B88}}[\rho(r)] = \int \rho^{\frac{4}{3}} \left(C_x + \frac{\beta x(r)^2}{1 + 6\beta x(r) \times \operatorname{arcsinh}(x)} \right) dr. \tag{2.30}$$

The only parameter β in Equation 2.30 was determined by fitting the exchange energies of noble gas atoms. [36,48]

Both the PBE [55] and the Lee, Yang, and Parr (LYP) [7] correlation part are more complex equations and include four more parameters. The parameters in the LYP functional are determined by fitting to data of the Helium atom and are often used in combination with the B88 exchange functional as BLYP. The fitting of the parameter to experimental data make some classify the BLYP functional as an empirical DFT method. An example of a purely theoretical GGA functional is PBE, where the parameters of both the exchange and the correlation part are determined by satisfying theoretically exact conditions. [36,48]

While bond lengths are reasonably well predicted by GGA methods, chemical reaction barriers and NMR shielding constants are underestimated. [48] A way of improving further upon the LDA and GGA functionals is to include higher order derivatives of the density ($\nabla^2\rho(r)$) and the kinetic

energy density (Equation 2.25). These methods are referred to as *meta*-GGAs, and examples of such functionals are M06-L[57] and TPSS[58,59]. These functionals have shown improvement over the LDA and GGA functionals for many systems, but the number of parameters to fit is also significantly increased. For the GGA exchange-correlation functional PBE, there are 5 parameters to determine, where as for the M06-L functional there are 36.[48]

2.3.3.3 Hybrid exchange-correlation functionals

An approach to improve upon GGA is to include a fraction of exact exchange (*i.e.,* HF and Equation 2.17) and these types of functionals are commonly known as *hybrid* functionals. Stephens *et al.*[8] combined the BLYP functional with Becke's three parameters exact exchange to the B3LYP functional

$$E_{xc}^{\mathrm{B3LYP}} = (1-a)E_x^{\mathrm{LDA}} + aE_x^{exact} + b\Delta E_x^{\mathrm{B88}} + (1-c)E_c^{\mathrm{LDA}} + E_c^{\mathrm{LYP}}. \quad (2.31)$$

Also the PBE functional was modified to include exact exchange, this hybrid version is denoted PBE0[60,61] or PBE1PBE

$$E_{xc}^{\mathrm{PBE0}} = \frac{1}{4}E_x^{exact} + \frac{3}{4}E_x^{\mathrm{PBE}} + E_c^{\mathrm{PBE}}. \quad (2.32)$$

While B3LYP is an example of a semi-empirical hybrid functional, PBE0 is a purely theoretical functional. Hybrid functionals show in general good results for geometries, energies and vibrational frequencies.[36] Thus, the B3LYP functional is used for geometry optimizations, frequency calculations and VCD properties in all papers this thesis is based on.[62–66]

2.4 Density functional based tight-binding

The density functional based tight-binding (DFTB)[67–69] method can be described as parameterized DFT. The parametrization makes DFTB calculations roughly two to three orders of magnitude faster than DFT calculations using a GGA functional.[70] Although the accuracy of geometries and energies calculated with DFTB might not be good enough to predict spectroscopies such as VCD and NMR[71], the method can serve as a starting

point for higher level methods. In this section an introduction to general DFTB theory is given before the specific method used in this work, GFN-xTB, is presented.

2.4.1 General density functional based tight-binding theory

In the following, first the different energy terms in DFTB are presented before addressing the main sources of error: the use of a minimal basis set, the integral approximation and inherited limitations from the parent DFT method.

The density functional based tight-binding (DFTB) method is derived in the DFT framework based on a Taylor expansion (Equation 2.34) of the density $\rho(r)$ around a reference density $\rho^0(r)$ given as[72]

$$\rho(r) = \rho^0(r) + \delta\rho(r). \tag{2.33}$$

The Taylor expansion can be expressed as[73]

$$E[\rho] = E^0[\rho_0] + E^1[\rho, (\delta\rho)] + E^2[\rho, (\delta\rho)^2] + E^3[\rho, (\delta\rho)^3] + \dots \tag{2.34}$$

Depending on how many of the terms in the Taylor expansion are included, different DFTB methods appears. DFTB1 includes the zeroth and first-order term, DFTB2 the zeroth-, first-, and second-order term and so on.

The zeroth-order term is approximated by a sum of pair potentials between all atoms A and B[72]

$$E^0[\rho_0] \approx \frac{1}{2}\sum_{AB} V_{AB}^{rep}. \tag{2.35}$$

These pair potentials are determined by comparison to DFT calculations[68] or by fitting to experimental data[74]. Because this term has to be pairwise parameterized for all atom pairs, not all DFTB methods are available for all elements in the periodic table.

By defining two parameters per element, the confinement radius of the atomic orbitals and the confinement radius of the initial density, the Kohn-Sham equation (Equation 2.23) can be used to determine the atomic orbitals and the initial density. The first-order term of the energy can then be

calculated as[73]

$$E^1[\rho,(\delta\rho)] = \sum_{iAB}\sum_{\mu\in A}\sum_{\nu\in B} n_i C_{\mu i} C_{\nu i} H^0_{\mu\nu} \tag{2.36}$$

where n_i is the occupation number of Kohn-Sham orbital i, $C_{\mu i}$ and $C_{\nu i}$ are the molecular orbital coefficients and $H^0_{\mu\nu}$ is the Hamiltonian matrix element. The total energy of DFTB1 is given as the sum of E^0 and E^1 and the Kohn-Sham equation is only solved once. Hence, DFTB1 is not a SCF method in contrast to DFTB2 and DFTB3 which are SCF methods.

The second-order correction term is given as[73]

$$E^2[\rho,(\delta\rho)^2] = \frac{1}{2}\sum_{AB}\Delta q_A \Delta q_B \gamma^h_{AB}. \tag{2.37}$$

where Δq_A and Δq_B are the atomic charges obtained from a Mulliken analysis on atom A and atom B. γ^h_{AB} is an analytical function that converges to $1/R_{AB}$ for large distances. For regular interatomic distances, the γ-function is chosen such that it properly describes the atomic chemical hardness or the Hubbard parameter, which can be determined from DFT.[75]The second order term represents the long-range Coulomb interactions between point charges at different sites and includes the self-interaction contribution of the individual atoms.[69]

The third-order term is given as[73]

$$E^3[\rho,(\delta\rho)^3] = \frac{1}{3}\sum_{AB}(\Delta q_A)^2 \Delta q_B \Gamma_{AB} \tag{2.38}$$

where, as for the second-order term, Δq are the charges determined by the Mulliken analysis and Γ_{AB} is the derivative of γ_{AB} with respect to the atomic charges. Because a minimal basis set is used in DFTB, diffuse electrons are poorly described. In DFTB3 this is taken into account by including the charge derivatives of the Hubbard parameter through the Γ-function. The Mulliken analyses needed to determine the atomic charges in the second and third energy terms are found self-consistently and this makes DFTB2 and DFTB3 a SCF method.

Using a minimal basis set is important for the computational speed up but it also results in a poorer description of the electron density compared with larger basis sets. This introduces errors in noncovalent interactions,

van der Waals interactions and hydrogen bonding. Another drawback with using a minimal basis set is that some properties can be decently described while others have poor accuracy. As a result, different parameterizations are needed to accurately predict different properties.

Together with the use of a minimal a basis set, the integral approximations are the main factors giving high computational efficiency. These integral approximation makes it possible to determine the Hamiltonian matrix element in the first-order term (Equation 2.36) once and these are then read as element specific parameters in a molecular calculation.[73]

The last main short-coming in DFTB is inherited from the parent DFT method. The exact exchange functional is not know and as a result of using approximate exchange, the electron-electron repulsion is overestimated.[73] Another short-coming of DFT, is the dispersion corrections.

In some cases, these short-comings have been accounted for by adding empirical corrections or extensions to the DFTB methods,[73] *e.g.* adding dispersion corrections with the D3 or D4 model.[76–78] An example of extending the DFTB method, is to improve the description of the Mulliken charges used for the second- and third-order energy terms (Equation 2.37 and 2.38) with exchange-like one-site two-electron integrals.[73]

2.4.2 The density functional based tight-binding method GFN-xTB

The tight-binding method used in the papers in this thesis is the GFN-xTB method which is used for the geometry optimizations and energy calculations in the conformational search code CREST. There are different GFN-xTB methods: GFN0-xTB, GFN1-xTB and GFN2-xTB. They are similar and they all use only element-specific parameters and are aiming for accurate geometries, frequencies and noncovalent interactions. In the following, details for GFN2-xTB are given and some of these elements differ for the two other approaches.

One of the drawbacks of DFTB methods are the large number of parameters needed for the pair-wise potential in the zeroth-order energy term (Equation 2.35) and as a result most DFTB methods are only parameterized for smaller parts of the periodic table. The GFN2-xTB method only employs-element specific parameters and is available for elements in the

periodic table up to radon (Z=86).

The first generation of the GFN-xTB method (GFN1-xTB) uses more or less the same Hamiltonian and electrostatic energies as DFTB3, which showed good results for less polar system.[71] To improve the description of the more polar and hydrogen-bonded systems, the second generation, GFN2-xTB, includes an anisotropic term in the second-order correction term of the energy (Equation 2.37). To describe the dispersion interaction, the D4[76–78] model is included for GFN2-xTB. As for DFTB, GNF2-xTB uses a minimal basis set but polarization functions are added for the heavier elements to improve the description of hypervalent bonding arrangements.[79]

2.5 Basis sets

Both Hartree-Fock and Kohn-Sham DFT calculate self-consistently the one-electron orbitals. These one-electron orbitals can be written as a linear combination of basis functions as shown in Equation 2.12, and for any quantum chemistry calculation these basis set functions must be specified. There are different types of basis functions (or atomic orbitals[36]) available and some examples are plane waves (PWs), effective core potentials (ECPs), Slater type orbitals (STOs) and Gaussian type orbitals (GTOs). PWs are suited to describe free valence electrons in metallic materials while ECPs replace the orbitals and electrons with a potential and can be a good choice to describe core electrons of heavy metals.

The STOs (Equation 2.39) and GTOs (Equation 2.40) are atom-centered basis functions consisting of a normalization factor N, a spherical harmonics part $Y_{l,m}$, and a radial part given in polar coordinates as[36]

$$\chi^{STO}_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r} \tag{2.39}$$

$$\chi^{GTO}_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2}. \tag{2.40}$$

where n, l and m are quantum numbers, r is the radial dependence, and ζ is the orbital exponent. Although STOs give a more accurate description of an atomic orbital, the GTOs are more commonly used due to their computational efficiency. Also in this thesis all calculations are done with GTO basis sets and therefore only these type of basis sets are further considered.

If an infinite number of basis functions were used to describe the orbitals, introducing the basis set would not be an approximation. This is in practice not possible and the computational cost and the accuracy depends on the choice of basis set. What would be a good choice of basis set does not only depend on the accuracy and available computational resources, but also on what type of system is studied and what the desired properties are.

The smallest possible basis set is a minimal basis set which consists of the least number of functions needed to describe the occupied atomic orbitals.[45] For a hydrogen atom this means a single s-function and for carbon it means two s-functions ($1s$ and $2s$) and three p-functions ($2p_x$, $2p_y$ and $2p_z$).

To give a more accurate description of an atomic orbital than the minimal basis set gives, more basis functions are used. If two basis functions are used to describe one atomic orbital, the basis set is classified as a double zeta basis set and if three basis functions are used, it is called a triple zeta basis set. In the case of carbon, the double zeta basis set would have four s-functions ($1s$, $1s$', $2s$ and $2s$') and six p-functions ($2p_x$, $2p_x$', $2p_y$, $2p_y$', $2p_z$ and $2p_z$'). Increasing the number of basis functions in this way allows for a better description of the electron distribution.[36] In chemistry, the valence electrons are more important in order to describe the chemical system than the core electrons, and a common approach is therefore to use a split valence basis set. These types of basis sets use one basis function to describe the core electrons while the valence electrons are described by more basis functions.

In the zeta classification it is assumed that one basis function is sufficuent to decently describe an atomic orbital. While this is true for STOs, it is not for GTOs.[36] To improve the description, a linear combination of primitive GTOs are used to construct a contracted basis function. In the k-nlmG basis sets designed by Pople and co-workers, k gives the number of primitive GTOs used to represent one contracted GTO for the core orbitals while nlm gives both the zeta splitting and the number of primitive GTOs used for each contracted GTO in the valence; nl indicates a double split valence while nlm indicates a triple split valence. For the basis set 6-31G, the core orbitals are a contraction of six primitive GTOs and the valence is doubly split with the inner valence basis function contracted by three primitive GTOs and the outer valence with one primitive GTO.

Using s-functions to describe the s-orbitals, p-functions to describe the

p-orbitals and so on gives little flexibility in the shape of the atomic orbitals. To allow a more flexible orbital shape, functions with higher angular momentum than the angular momentum of the higher occupied orbital of that atom are added. These functions are denoted polarization functions. To polarize the s-functions on a hydrogen, a set of p-functions are added, to polarize p-functions on a carbon, a set of d-functions is added.

In cases where there are electrons far from the nucleus, *e.g.* an anion or a molecule with a dipole moment, diffuse functions are included to achieve a better description. These diffuse functions have a rather small exponent (ζ in Equation 2.40) to give a better description far from the nucleus. In the Pople style basis sets, the polarization functions are denoted with * while diffuse functions are denoted by +. When a single * or + is used, it indicates that the functions are only added for the heavy atoms. If ** or ++ is used, also polarization or diffuse functions are added on hydrogens.

In the papers included in this thesis, the double zeta Pople-style basis set with diffuse and polarizing functions on the heavy atoms, 6-31+G*, is used for geometries, frequencies and VCD properties. In **Paper I** larger and smaller Pople-style basis sets are tested for VCD spectra of smaller cyclic oligopeptides. The results show that the smallest basis set giving a reasonable VCD spectrum is the double-zeta basis set with polarization functions on the heavy atoms, 6-31G*, which is in line with the recommendations of Stephens, Devlin and Cheeseman for conformationally rigid chiral organic molecules.[1] The test shows that going to a triple zeta basis set with diffuse and polarization functions also on the hydrogens (6-311++G**) improves the frequencies slightly. The basis set test was performed for two tetrapeptides, but as the aim of this thesis is to investigate hexapeptides, 6-31+G* is a good compromise between accuracy and computational cost.

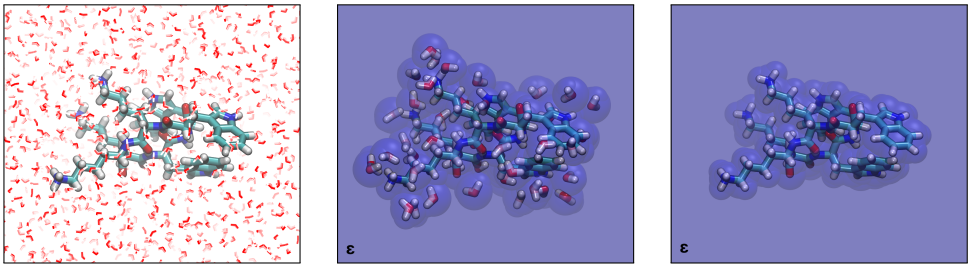
2.6 Modeling solvated systems

To accurately simulate IR and VCD spectra, the environment the molecule is in has to be taken into account. Because it is too computationally expensive to calculate both the solute and the particular solvent with QM methods, multiscale strategies are used; part of the system is modeled by QM and parts by a computationally cheaper classical description. For spectroscopies, it is common to let the classical part be described by molecular mechanics

(MM), commonly called QM/MM (Figure 2.1a), or by a continuum model (Figure 2.1c).[80] While the continuum model describes the environment as an average reaction field, the QM/MM approach has the benefit of including the actual atoms which keeps the specific solute-solvent interactions intact.

There are several implications using the QM/MM approach. One is that vibrational spectroscopies requires equilibrium geometries and in QM/MM, the geometries are typically obtained from MD or MC simulations, and they do not correspond to equilibrium geometries. There are methods to optimize these structures in a frozen environment, but the computational cost of those in addition to sampling all configurations of the solute-solvent interactions is high. Furthermore, there is no widely available software to calculate chiroptical spectroscopies. Dundas *et. al.* calculated IR and Raman spectra using a polarizable embedding QM/MM method[81] while Cappelli and co-workers have calculated VCD and ROA spectra of small chiral molecules in water with a QM/MM approach based on fluctuating charges.[82] A way of including some of the explicit solute-solvent interactions, is to add a few solvent molecules in the QM part while model the rest of the solvent as a continuum. This approach is referred to as the supramolecular approach (Figure 2.1b).

In this work, computationally cheaper continuum models are used and is the only model described below. In the papers included in this thesis, the generalized Born model with surface area contributions (GBSA) is used in the CREST code that performs the MTD simulation and the polarizable continuum model (PCM) is used for the QM calculations. These two models are used as examples in the following.



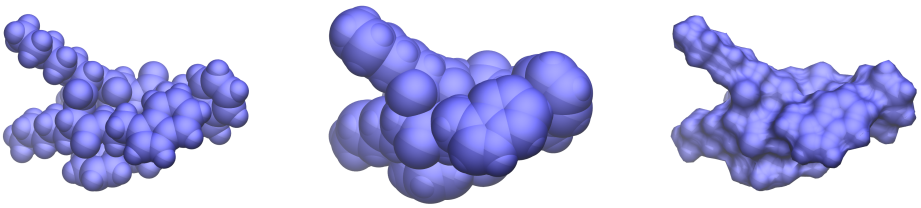
(a) QM/MM. **(b)** Supramolecular. **(c)** QM/Continuum.

Figure 2.1. Illustration of an explicit solvent model (QM/MM), a supramolecular within a continuum approach and the continuum model.

2.6.1 Continuum solvation models

The continuum model considers the solvent as a uniform dielectric medium with the solute molecule placed in a cavity within the medium as illustrated in Figure 2.1b and 2.1c. This medium is characterized by only the dielectric constant ε. This means that solvents with similar dielectric constants, *e.g.* hexanol ($\varepsilon = 12.51$) and pyridine ($\varepsilon = 12.98$), give similar results. In addition to the description of the medium itself, the cavity the solute is placed in has to be described and the cavitation energy required to create the hole in the medium.

The simplest way of describing the cavity is to use a sphere or an ellipsoid. A more realistic model is to build the cavity from overlapping van der Waals spheres of the atoms of the solute. A problem with the van der Waals surface is that small pockets where the solvent cannot reach are present. An approach to improve the surface description is to let the surface be traced out by the center of a probe sphere, which results in the solvent accessible surface (SAS) model, or an in-ward facing probe resulting in the solvent excluded surface (SES) model, see Figure 2.2. Although the SAS and SES give a more realistic picture of the cavity, they are computationally more demanding and for both GBSA and PCM, a scaled van der Waals surface is used.[36] The scaling is done to make the van der Waals surface more similar to the SAS and usually a scaling factor of about 1.2 is used.[36]



(a) A van der Waals surface.

(b) A solvent accessible surface (SAS).

(c) A solvent excluded surface (SES).

Figure 2.2. Illustration of different ways to construct the surface of the cavity.

The energy of the solute molecule placed in the medium can be written as a sum of the energy of the molecule in vacuum (E_{vac}) and the free energy

required to take the molecule from vacuum to the solvent (ΔG_{solv}) as[83]

$$E_{tot} = E_{vac} + \Delta G_{solv} \tag{2.41}$$

where the solvation free energy ΔG_{sol} can be divided into four contributions[18]

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{elec}. \tag{2.42}$$

The first contribution to the solvation energy, ΔG_{cav}, arises due to the break-up of solvent-solvent interactions to create the cavity for the solute. The second and third terms describe quantum mechanical dispersion and repulsion between the solvent electron density and the solute. The final term, ΔG_{elec}, describes the energy due to mutual polarization of solute and solvent. This electrostatic term is in practice the most relevant component of the solvation energy. The definition of ΔG_{elec} differentiates between continuum solvation models.

In GBSA the electrostatic term is given by Coulomb's law and the Born equation.[84] In this model, only the lowest-order electric moment, q, is taken into account and within this formalism the electrostatic free energy can be expressed as[84]

$$\Delta G_{elec} = -\Big(1 - \frac{1}{\varepsilon}\Big) \sum_i \sum_j \frac{q_i q_j}{f_{ij}} \tag{2.43}$$

where, in GBSA, q is the partial atomic charge, ε is the dielectric constant of the solvent and f_{ij} a function depending on the internuclear distance r_{ij} and the Born radii of the two atoms a_i and a_j:

$$f_{ij} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}}, \tag{2.44}$$

$$a_{ij} = a_i a_j \tag{2.45}$$

$$D = \frac{r_{ij}^2}{4a_{ij}^2} \tag{2.46}$$

In the case of PCM, the more computationally demanding Poisson equation is used and the electrostatic part of the free energy can be written as[36]

$$\Delta G_{elec} = \frac{1}{2} \int \rho(\mathbf{r}) \phi_{reac}(\mathbf{r}) d\mathbf{r}. \tag{2.47}$$

where $\rho(\mathbf{r})$ is the molecular charge distribution and $\phi_{reac}(\mathbf{r})$ is the electrostatic reaction potential.

The reaction potential describes the mutual polarization between solute and solvent, and it has to be solved iteratively and is commonly referred to as the self-consistent reaction field (SCRF).[18] When the standard SCF approach is used, which includes HF and DFT, the implementation and hence the computational cost of using a continuum model is similar the same calculation in vacuum.[85] With a solvent model in place, the next section describes the general procedure for geometry optimization.

2.7 Molecular geometry optimizations

This chapter has up to this point given the theory that is used to describe atoms and molecules, and their interaction with a solvent environment. In this section, the methods described above are used to find the structure with the lowest energy or a stable geometry, *i.e.* find the equilibrium structure or the optimized structure. In Figure 2.3 an illustration of the PES is shown where the equilibrium geometry is shown as the minimum. To be able to calculate the molecular properties needed to simulate the vibrational spectra in the next chapter, the structures found in the conformational search (Chapter 1) need to be equilibrium structures.

The first-order derivative of the energy with respect to nuclear displacements is called the gradient ($\mathbf{g}$). Element i of the gradient is given as

$$g_i = \frac{\partial E}{\partial R_i} \tag{2.48}$$

where R_i is a coordinate of atom i. The gradient in Cartesian coordinates is a vector with the length $3N$, where N is the number of atoms included in the optimization. The second-order derivative of the energy with respect to the nuclear displacements is called the Hessian ($\mathbf{H}$). The Hessian is a $3N \times 3N$ matrix, with elements given as

$$H_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j} \tag{2.49}$$

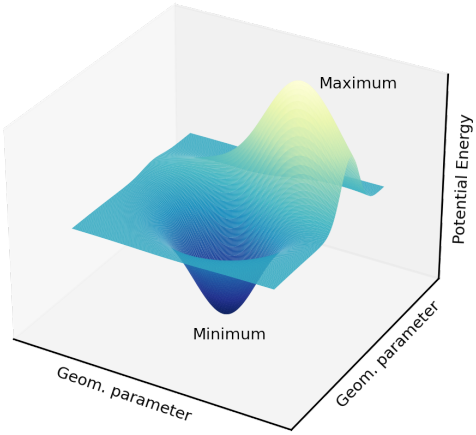


Figure 2.3. A simple illustration of the PES with only one minimum and one maximum. The minimum corresponds to the molecular geometry optimized or equilibrium structure.

where i and j are atom indices. The requirement for a structure to have reached a minimum is that $\mathbf{g}$ is zero and that all the eigenvalues of $\mathbf{H}$ are positive. Hence, to determine if a given structure is the minimum structure, both the gradient and the Hessian have to be calculated. In most cases, the gradient is calculated analytically which has a similar computational cost as calculating the energy.[86] The Hessian on the other hand, is computationally much more expensive to calculate analytically, therefore an approximated Hessian is often used for geometry optimizations. A common approach is to initially guess the Hessian and then improve it by updating it at each step of the optimization. How the Hessian is updated depends on the optimization algorithm and different approaches are shown in the following.

To find the minimum and hence the equilibrium structure, the atoms are displaced with a small step ($\Delta\mathbf{R}$) according to the forces acting on them. The forces on nucleus of atom i are given as the negative gradient

$$F_i = -g_i = -\frac{\partial E}{\partial R_i}. \tag{2.50}$$

There are different algorithms available to determine the step, among them are the steepest decent, the conjugate gradient method and the Newton-Raphson method.

The Newton-Raphson method is a nonlinear optimization method and

the starting point is the local quadratic approximation of the PES given as[86]

$$E(\mathbf{R}) = E(\mathbf{x}_0) + \mathbf{g}_0^T \Delta\mathbf{R} + \frac{1}{2}\Delta\mathbf{R}^T \mathbf{H}_0 \Delta\mathbf{R} \tag{2.51}$$

where $\mathbf{g}_0$ is the gradient and $\mathbf{H}_0$ the Hessian being evaluated at $\mathbf{R}_0$ and $\Delta\mathbf{R} = \mathbf{R} - \mathbf{R}_0$. By differentiation of Equation 2.51 with respect to the coordinates and setting to zero[86]

$$\mathbf{g}(\mathbf{R}) = \mathbf{g}_0 + \mathbf{H}_0 \Delta\mathbf{R} = 0 \tag{2.52}$$

In the local quadratic approximation to the PES, the displacement towards the minimum is given as[86]

$$\Delta\mathbf{R} = -\mathbf{H}_0^{-1}\mathbf{g}_0 \tag{2.53}$$

which is also known as the Newton or the Newton-Raphson step.[86] In the Newton-Rapson method, the Hessian is calculated at each step which is rather costly. A computationally cheaper approach to predict the step size, is to use the quasi-Newton method where an approximate Hessian is used. In the papers included in this thesis, the DFT geometry optimizations are done with the Berny algorithm as implemented in Gaussian 16,[87] which is a modified quasi-Newton method. The quasi-Newton method calculates the new Hessian ($\mathbf{H}^{new}$) as the sum of the old Hessian ($\mathbf{H}^{old}$) and a change in the Hessian ($\Delta\mathbf{H}$) .

Under the assumption that the surface is quadratic, the new Hessian must fulfill the Newton condition:

$$\Delta\mathbf{g} = \mathbf{H}^{new}\Delta\mathbf{R} \tag{2.54}$$

where $\Delta\mathbf{g} = \mathbf{g}^{new} - \mathbf{g}^{old}$ and $\Delta\mathbf{R} = \mathbf{R}^{new} - \mathbf{R}^{old}$. There are an infinite number of ways to update the Hessian that fulfill the Newton condition. One of the most commonly used updates, which is also used in the Berny algorithm, is the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[88–91] update given as[86]

$$\Delta\mathbf{H}^{BFGS} = \frac{(\Delta\mathbf{g} - \mathbf{H}^{old}\Delta\mathbf{R})(\Delta\mathbf{g} - \mathbf{H}^{old}\Delta\mathbf{R})^T}{(\Delta\mathbf{g} - \mathbf{H}^{old}\Delta\mathbf{R})^T\Delta\mathbf{R}}. \tag{2.55}$$

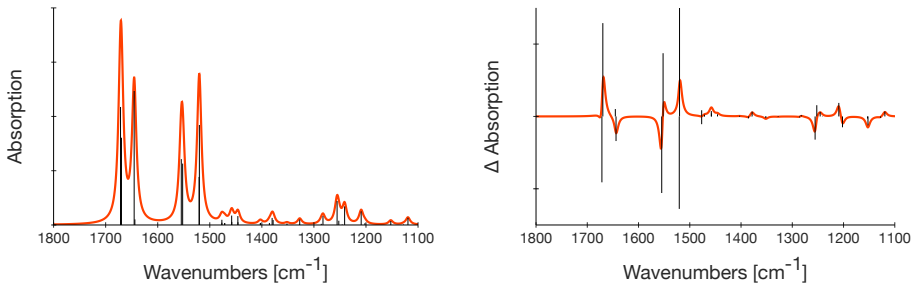
In a robust geometry optimization it is important to ensure that the step size is not too large and that the step is in the right direction towards the minimum. To ensure this, a trust radius is used. The predicted step is taken only if within the trust radius. Otherwise, a step to the trust radius in the direction of the gradient is taken. This trust radius is usually predefined and then updated during the optimization by determining how well the PES is described by the quadratic expression (Equation 2.51).[86]

Because the energy, the gradient and the Hessian are calculated with a given precision, the minimum cannot be located exactly. The gradient will never reach zero, it can only be reduced to a given value close to zero.[36] It is therefore common to let the geometry optimization go on until some predefined convergence criteria are met. These criteria are typically changes in energy, magnitude of the force and step size.

Chapter 3

Chiroptical spectroscopies

In Figure 3.1, the calculated IR and VCD spectra of Cyclo(Ala-βAla-Alaβ-Ala) is shown. To obtain the calculated spectra, the absorption (IR) or difference in absorption between right and left circularly polarized light (VCD) and the frequencies are needed. This chapter aims to describe the steps from a geometry-optimized conformer ensemble to a final, simulated spectrum that can be compared to an experimental one.



(a) IR spectra **(b)** VCD spectra

Figure 3.1. Calculated IR and VCD spectra of Cyclo(Ala-βAla-Alaβ-Ala) in trifluoroethanol (B3LYP/6-311+G*/CPCM) from **Paper I**. The sticks are the calculated dipole (IR) and rotational (VCD) strengths (Section 3.2) using a Boltzmanns' average with free energies (Section 3.3.2). The red line is a convolution with Lorentzian broadening with a 'full width at half maximum' (FWHM) of 16 cm^{-1}. The frequencies are scaled with a factor of 0.9679[92] (Section 3.3.1).

The first section shows how the frequencies are calculated while the second discusses the molecular properties needed to calculate the IR and

VCD intensities. The last section focuses on how to include the whole conformer ensemble, how to add lineshapes and scale the frequencies before analysing the similarity with experiment.

3.1 Vibrational Frequencies

In the Born-Oppenheimer approximation, the Schrödinger equation for the nuclear motion is[1]

$$[\hat{T}_N + E_e(\mathbf{R})]\Psi_N(\mathbf{R}) = E_N\Psi_N(\mathbf{R}) \tag{3.1}$$

where $E_e(\mathbf{R})$ is the PES, $\Psi_N(\mathbf{R})$ is the nuclear wave function, and E_N is the energy of the nuclear motion. The first step in solving this equation is to get an explicit expression for $E_e(\mathbf{R})$. Assuming that nuclear vibrations are only small displacements close to the equilibrium geometry $\mathbf{R_0}$ from Section 2.7, the energy at $\mathbf{R}$ can be given by a Taylor expansion to second order as[1]

$$E_e(\mathbf{R}) = E_e(\mathbf{R_0}) + \sum_{J=1}^{3N}\left(\frac{\partial E_e}{\partial \Delta\mathbf{R}_J}\right)_0 \Delta\mathbf{R}_J + \frac{1}{2}\sum_{J,I=1}^{3N}\left(\frac{\partial^2 E_e}{\partial\Delta\mathbf{R}_J\partial\Delta\mathbf{R}_I}\right)_0 \Delta\mathbf{R}_J\Delta\mathbf{R}_I \tag{3.2}$$

where the energy of the equlibrium geometry is given by $E_e(\mathbf{R_0})$ and $\Delta\mathbf{R}_J = \mathbf{R}_J - \mathbf{R}_{J,0}$ is the displacement of nucleus J from its equlibrium position. In Equation 3.2, the harmonic approximation is used: only the zeroth-, first- and second-order terms are included. By adding higher-order terms, the PES would be closer to the exact potential (illustrated in Figure 1), and will include anharmonic corrections. The first and second derivatives in Equation 3.2 have already been defined as the gradient (Equation 2.50) and the Hessian matrix (Equation 2.49), respectively, in Chapter 2.

The derivatives of E_e are evaluated at an equilibrium geometry: a minimum on the PES. Hence, the first derivative, the gradient, is zero. In addition, the constant ground state energy at the equlibrium geometry is independent of the nuclear coordinates and can therefore be set to zero and Equation 3.2 reduces to

$$E_e(\mathbf{R}) = \frac{1}{2}\sum_{J,I=1}^{3N}\left(\frac{\partial^2 E_e}{\partial\Delta\mathbf{R}_J\partial\Delta\mathbf{R}_I}\right)_0 \Delta\mathbf{R}_J\Delta\mathbf{R}_I \tag{3.3}$$

For convenience, the coordinates are next transformed: first to mass-weighted coordinates, and then to normal coordinates. In Equation 2.4, T_N is defined in terms of the mass of the nuclei. To get rid of the mass of each nucleus, we transform to mass-weighted coordinates[93]

$$\mathbf{q}_J = \sqrt{M_J}\Delta\mathbf{R}_J \tag{3.4}$$

where M_J is the mass of nucleus J and $\Delta\mathbf{R}_J$ is its Cartesian displacement. The normal coordinates can be written as a linear combination of the mass-weighted Cartesian coordinates as[2]

$$Q_a = \sum_{J=1}^{3N}\mathbf{S}_J^a\mathbf{q}_J \tag{3.5}$$

where $\mathbf{S}_J^a$ is the transformation matrix that diagonalizes the Hessian matrix (Equation 3.3). This transformation eliminates all cross-terms from different normal coordinates, Q_a and Q_b, and only includes the squares of individual normal coordinates Q_a^2. The kinetic operator in normal coordinates is given as[93]

$$\hat{T}_N = -\sum_{a=1}^{3N}\frac{1}{2}\frac{\partial^2}{\partial Q_a^2}. \tag{3.6}$$

while the PES in normal coordinates in the harmonic approximation is given as

$$E_e(\mathbf{Q}) = \frac{1}{2}\sum_{a=1}^{3N}\left(\frac{\partial^2 E_e}{\partial Q_a^2}\right)_0 Q_a^2 = \frac{1}{2}\sum_{a=1}^{3N}k_a Q_a^2 \tag{3.7}$$

where k_a is the force constant, the diagonal of the Hessian matrix in its eigenbasis.

Since a molecule is considered a rigid structure, three of the $3N$ degrees of freedom are due to translational motion and three of them, or two in the case of a linear molecule, to overall rotation of the molecule. The remaining $3N-6$ ($3N-5$ for a linear molecule) degrees of freedom correspond to

the vibrational degrees of freedom. Since the energy of the molecule does not change when the molecule translates and is only very small when the molecule rotate, the force constants of these six normal coordinates can be set to zero.

Each normal coordinate, or normal mode, is independent of all of the other normal modes. The water molecule with three atoms has $3\times3-6=3$ normal modes and these are illustrated in Figure 3.2. Each normal mode can be individually excited and has a fundamental frequency. The frequencies of these $3N-6$ motions have to be calculated to obtain the IR and VCD spectra in Figure 3.1.

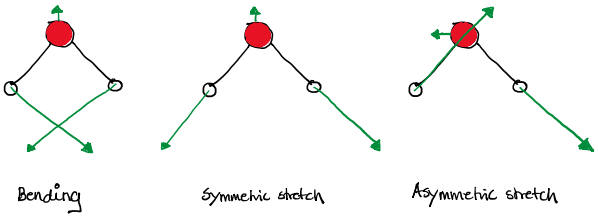


Figure 3.2. The normal modes of water. Since the normal modes are mass weighted and the center of mass is fixed[93], it is mainly the light hydrogen atoms that move in the three normal modes.

Using Equation 3.7 only including the $3N-6$ vibrational motions, and expressing $\hat{T}_N$ in normal coordinates, the nuclear Schrödinger equation is transformed into[1]

$$\left[-\frac{1}{2}\sum_{a}^{3N-6}\frac{\partial^2}{\partial Q_a^2}+\frac{1}{2}\sum_{a}^{3N-6}k_aQ_a^2\right]\Psi_N(\mathbf{Q})=E_N\Psi_N(\mathbf{Q}). \tag{3.8}$$

Equation 3.1 is now reduced to $3N-6$ identical one-dimensional harmonic oscillator Schrödinger equations, each with the eigenvalue:[1]

$$\varepsilon_a=\left(n_a+\frac{1}{2}\right)\hbar\omega_a \quad n_a=0,1,2\ldots \tag{3.9}$$

where n_a is the vibrational quantum number (illustrated in Figure 1) and ω_a is the frequency of normal mode a. Finally, the frequencies can be calculated as

$$\omega_a = \frac{1}{2\pi}\sqrt{k_a} = \frac{1}{2\pi}\sqrt{\left(\frac{\partial^2 E_e}{\partial Q_a^2}\right)_0}. \tag{3.10}$$

3.2 Calculation of IR and VCD properties

The light interacting with the molecular system in IR and VCD spectroscopy has an energy high enough to excite vibrations in the molecular system, but not to electronically excite it. Because the energy of this light has a wavelength that is much larger than the molecule, the interaction can be described as a uniform electromagnetic field:[2]

$$H' = -\sum_\alpha \hat{\mu}_\alpha \vec{F}_\alpha - \sum_\alpha \hat{m}_\alpha \vec{B}_\alpha \tag{3.11}$$

where $\hat{\mu}_\alpha$ is the electric dipole operator consisting of an electronic part ($\hat{\mu}_\alpha^{el}$) and a nuclear part ($\hat{\mu}_\alpha^{nuc}$), and $\hat{m}_\alpha$ is the magnetic dipole operator with an electronic part ($\hat{m}_\alpha^{el}$) and a nuclear part ($\hat{m}_\alpha^{nuc}$). $\vec{F}$ is the external electric field, while $\vec{B}$ is the external magnetic field. The electric and magnetic dipole operators are given in CGS units as[1]

$$\hat{\mu}_\alpha = -\sum_i e r_{i\alpha} + \sum_L Z_L e R_{L\alpha} = \hat{\mu}_\alpha^{el} + \hat{\mu}_\alpha^{nuc}, \tag{3.12}$$

$$\hat{m}_\alpha = \sum_i \frac{e}{2mc}(\mathbf{r}_i \times \mathbf{p}_i) + \sum_L \frac{Z_L e}{2M_L c}(\mathbf{R}_L \times \mathbf{P}_L) = \hat{m}_\alpha^{el} + \hat{m}_\alpha^{nuc} \tag{3.13}$$

where e and $Z_L e$, m_i and M_L, $\mathbf{r}_i$ and $\mathbf{R}_L$, $\mathbf{p}_i$ and $\mathbf{P}_L$, are the charge, the mass, the position and the momentum of electron i and nucleus L, respectively. Since the operators are given in CGS units also the rest of this sections is.

The electric dipole contribution is the dominating term in Equation 3.11 and when calculating the absorption of light as in an IR spectrum, the electric dipole approximation is used: the magnetic field is neglected and only the electric field is taken into account. In this approximation, the perturbation H' in Equation 3.11 is reduced to

$$H' = -\sum_{\alpha} \hat{\mu}_{\alpha} \vec{F}_{\alpha}. \tag{3.14}$$

Using the electric dipole approximation and the double harmonic approximation, the interaction between the light and the molecule can be described in terms of the dipole strength (D) as[1]

$$D = \langle i|\hat{\mu}_{\alpha}|f\rangle\langle f|\hat{\mu}_{\alpha}|i\rangle \tag{3.15}$$

where $\langle i|\hat{\mu}_{\alpha}|f\rangle$ is the electric dipole transition moment of the transition from vibrational state i to f.

In the electric dipole approximation, the absorption of left and right circularly polarized light is equal and as a result, the VCD intensity, defined as the difference in absorption of left and right circular polarized light (Equation 3), is zero. To predict non-zero VCD intensities, also the magnetic field has to be taken into account and the perturbation H' for VCD is given as in Equation 3.11.

To calculated the VCD intensities, in addition to the electric dipole transition moment from Equation 3.15, also the magnetic dipole transition moment, $\langle f|\hat{m}_{\alpha}|i\rangle$, is needed. The VCD intensity is calculated as the rotational strength (R) given as the imaginary part (Im) of the electric and magnetic transition dipole moment as[1]

$$R = \mathrm{Im}\left[\langle i|\hat{\mu}_{\alpha}|f\rangle\langle f|\hat{m}_{\alpha}|i\rangle\right] \tag{3.16}$$

for the transition from vibrational state i to f.

Within the harmonic approximation, only transitions between a state with adjacent energy levels are allowed, namely, $n_{excited} = n_{initial} \pm 1$. In addition, the electric dipole transition moment is only non-zero when the excited vibrational state, $n_{excited}$, is 1 for one normal mode Q_a and zero for all other normal modes. This excitation of normal mode Q_a is referred to as the fundamental vibrational excitation of the ath normal mode.[1] Thus, for a molecule with N atoms, $3N-6$ vibrational excitations are predicted to occur with frequencies equal to the frequency of the given normal mode. For these fundamental excitations, evaluation of the electric dipole transition moment using the Born-Oppenheimener wave function, transforms the dipole strength from Equation 3.15 into

$$D_a = \frac{\hbar}{2\omega_a} \frac{\partial \mu_\alpha}{\partial Q_a} \frac{\partial \mu_\alpha}{\partial Q_a}, \tag{3.17}$$

where ω_a is the frequency of normal mode a given in Equation 3.10, Q_a is the normal coordinate and $\mu_\alpha = \langle 0|\hat{\mu}_\alpha|1\rangle$ where $|0\rangle$ is the electronic and vibrational ground state and $|1\rangle$ is the electronic ground state but excited vibrational state.

The rotational strength follows the same selection rules as the dipole strength. Hence, the same fundamental vibrational transitions are allowed in VCD as in IR. Evaluation of the electric and magnetic dipole moments using the same strategy as for the dipole strength transforms the rotational strength from Equation 3.16 into

$$R_a = \frac{\hbar}{2\omega_a} \operatorname{Im} \frac{\partial \mu_\alpha}{\partial Q_a} \frac{\partial m_\alpha}{\partial Q_a} \tag{3.18}$$

where $m_\alpha = \langle 0|\hat{m}_\alpha|1\rangle$ while a, ω_a, Q_a, μ_α, $|0\rangle$ and $|1\rangle$ are the same as in Equation 3.17.

Although the normal coordinates are useful, the calculations are usually done in Cartesian displacement coordinates. By replacing the normal coordinates in Equation 3.17 with Cartesian displacement coordinates using Equation 3.4 and 3.5 the dipole strength is given as[1]

$$D_a = \frac{\hbar}{2\omega_a} \sum_{\beta=x,y,z} \sum_{J,J'}^{N} \sum_{\alpha,\alpha'=x,y,z} \left[S^a_{J\alpha} P^J_{\alpha\beta}\right] \left[S^a_{J'\alpha'} P^{J'}_{\alpha'\beta}\right] \tag{3.19}$$

where $S^a_{J\alpha}$ is the transformation matrix that diagonalizes the Hessian given in Equation 3.5 with α being the three Cartesian coordinates for the displacement $\Delta\mathbf{R}_J$ of nucleus J. $P^J_{\alpha\beta}$ is a second-rank tensor, the atomic polar tensor (APT) of nucleus J:

$$P^J_{\alpha\beta} = \left(\frac{\partial \mu_\beta}{\partial \Delta R_{J,\alpha}}\right)_0 = 2\left\langle \left(\frac{\partial \Psi^0_e(\mathbf{R})}{\partial \Delta R_{J,\alpha}}\right)_0 \middle| \hat{\mu}^{el}_\beta \middle| \Psi^0_e(\mathbf{R_0}) \right\rangle + Z_J e \delta_{\alpha\beta} \tag{3.20}$$

where μ_β is the electric dipole moment given in Equation 3.17 with $\beta = x, y, z$, $\Delta R_{J,\alpha}$ is the α-Cartesian component of the displacement from the

equilibrium geometry $\mathbf{R_0}$ of nucleus J, Ψ_e^0 is the electronic wave function of the electronic ground state, $\hat{\mu}_{\alpha,\beta}^{el}$ is the β-Cartesian component of the electronic contribution to the electric dipole moment and $\delta_{\alpha\beta}$ is the Kronecker delta equal to 1 when $\alpha = \beta$ and zero otherwise. The APT is usually divided in an electronic contribution ($E_{\alpha\beta}^J$) and a nuclear contribution ($N_{\alpha\beta}^J$):

$$E_{\alpha\beta}^J = \left\langle \left(\frac{\partial \Psi_e^0(\mathbf{R})}{\partial \Delta R_{J,\alpha}} \right)_0 \middle| \hat{\mu}_\beta^{el} \middle| \Psi_e^0(\mathbf{R_0}) \right\rangle, \tag{3.21}$$

$$N_{\alpha\beta}^J = Z_J e \delta_{\alpha\beta}. \tag{3.22}$$

Before writing the rotational strength on similar form as the dipole strength in Equation 3.19, the Born-Oppenheimer wave function has to be improved. Within the Born-Oppenheimer approximation, it is not possible to give a complete theoretical description of VCD. The reason for this failure is that the electronic contribution to the magnetic dipole transition moment is zero,[94] and that leaves us with a magnetic dipole transition moment only arising from the nuclei. To obtain also the electronic contribution to the magnetic dipole transition moment, the Born-Oppenheimer wave function is improved by mixing the Born-Oppenheimer ground-state wave function with excited-state wave functions using perturbation theory. The resulting equation for the rotational strength is then given as[95]

$$R_a = \hbar^2 \sum_{\beta=x,y,z} \sum_{J,J'}^{N} \sum_{\alpha,\alpha'=x,y,z} \left[S_{\alpha J}^a P_{\alpha\beta}^J \right] \left[S_{\alpha' J}^a M_{\alpha'\beta}^{J'} \right] \tag{3.23}$$

where $M_{\alpha'\beta}^{J'}$ is the atomic axial tensor (AAT) given as

$$M_{\alpha\beta}^J = \left\langle \left(\frac{\partial \Psi_e^0(\mathbf{R})}{\partial \Delta R_{J,\alpha}} \right)_0 \middle| \left(\frac{\partial \Psi_e^0(\mathbf{R_0}, \vec{B}_\beta)}{\partial \vec{B}_\beta} \right)_0 \right\rangle + \frac{i}{4\hbar c} \sum_\gamma Z_J e R_{J\gamma}^0 \epsilon_{\alpha\beta\gamma} \tag{3.24}$$

where Ψ_e^0 and $\Delta R_{\alpha J}$ are the same as in Equation 3.19, $\vec{B}_\beta$ is an external magnetic field interacting with the electronic part of the magnetic dipole moment, $R_{J,\gamma}^0$ is the γ-Cartesian component of the equilibrium geometry of nucleus J and $\epsilon_{\alpha\beta\gamma}$ is the alternating Levi-Civita tensor giving +1 for even permutations of the x, y and z coordinates and -1 for odd permuta-

tions. As for the APT, also the AAT is commonly devided in an electronic contribution ($I^J_{\alpha\beta}$) and a nuclear contribution ($J^J_{\alpha\beta}$):

$$I^J_{\alpha\beta} = \left\langle \left(\frac{\partial \Psi^0_e(\mathbf{R})}{\partial \Delta R_{J,\alpha}} \right)_0 \middle| \left(\frac{\partial \Psi^0_e(\mathbf{R_0}, \vec{B}_\beta)}{\partial \vec{B}_\beta} \right)_0 \right\rangle, \tag{3.25}$$

$$J^J_{\alpha\beta} = \frac{i}{4\hbar c} \sum_\gamma Z_J e R^0_{J\gamma} \epsilon_{\alpha\beta\gamma}. \tag{3.26}$$

The purpose of giving the dipole strength and rotational strength in terms of $S^a_{\alpha J}$, $P^J_{\alpha\beta}$, $M^J_{\alpha\beta}$ and ω_a is that these are the properties actually calculated. From the equilibrium geometry, the Hessian, given in Equation 2.49, is needed to calculate $S^a_{\alpha J}$ and ω_a. Both $P^J_{\alpha\beta}$ and $M^J_{\alpha\beta}$ include the calculation of derivatives of electronic ground-state wave functions with respect to the Cartesian displacement coordinates. To calculate $M^J_{\alpha\beta}$, an additional term with the derivatives of the perturbed electronic ground-state wave function with respect to the magnetic field is needed.

3.3 Visualising the calculated spectra

Plotting the normal mode frequencies and dipole or rotational strengths from the previous section results in the spectrum shown as black sticks in Figure 3.1. In the following, how the broadening of the dipole and rotational strengths is done such that they transform into the red lines in Figure 3.1 is described and a discussion on how to include the conformer ensemble is given.

3.3.1 Line broadening of the calculated spectra

An experimental spectrum contains peaks rather than spectral lines due to phenomena such as the finite lifetime of the excited state and collisional effects.[93] To get a better visual comparison with experiment, a lineshape function is added to the spectral intensity for each normal mode.[2] This lineshape can be described by a Gaussian, a Voight or a Lorentzian function depending on the nature of the experiment. In IR and VCD, it is usually assumed that a normalized Lorentzian function[2]

$$f_a(\omega) = \left[\frac{\gamma_i}{(\omega_a - \omega)^2 + \gamma_i^2}\right] \tag{3.27}$$

is a good description of the broadening where γ_i is a decay constant which is half the 'full width at half maximum' (FWHM), 2γ = FWHM, and ω_a is the frequency of normal mode a. The VCD spectrum for a given conformer n as a function of the frequency is then given as

$$R_n(\omega) = \sum_{a}^{3N-6} R_a f_a(\omega) \tag{3.28}$$

where R_a is the calculated VCD intensity given in Equation 3.23 for normal mode a. In similar manner, the line-broadened IR spectrum can be calculated.

3.3.2 Weighting spectra of conformers

After the conformational search described in Chapter 1 and optimization of those structures as described in Section 2.7, several unique conformers are identified. One way to determine the contribution of each of the unique conformers to the spectra is to use a Boltzmann average. In this case, the lower the conformer is in energy, the larger the contribution to the spectra and the weight p_n of conformer n is given as

$$p_n = \frac{e^{\frac{-\Delta G_n}{kT}}}{\sum_{n=1}^{N} e^{\frac{-\Delta G_n}{kT}}} \tag{3.29}$$

where ΔG_n is the free energy difference between the lowest free energy conformer and conformer n, k is the Boltzmann constant, T is the temperature in Kelvin and N is the total number of conformers. The Boltzmann-averaged VCD spectrum is then calculated as

$$R(\omega) = \sum_{n=1}^{N} p_n R_n(\omega) \tag{3.30}$$

where $R_n(\omega)$ is given in Equation 3.28 for conformer n, p_n in Equation 3.29 and N is the number of conformers included in the Boltzmann average.

Other types of energies can also be used in Equation 3.29, *e.g.* the

zero-point corrected electronic energy (ΔE_n^{ZPC}) or the enthalpy (ΔH_n). The relation between the free energy and the enthalpy is given by $\Delta G = \Delta H - T\Delta S$. Hence, to calculate the enthalpy the $T\Delta S$ term is not needed which makes it computationally cheaper than the free energy. For small- and medium-sized molecules the $T\Delta S$ term is assumed to be small and the differences between using the free energy and the enthalpy should be negligible. However, for the cyclic oligopeptides investigated in this thesis, there are examples of using the enthalpy and the free energy for the Boltzmann average resulting in significant different spectra. One of the contributions to the entropy is the rotational motion and in cases where the peptides have freely rotating side chains the entropy contribution may be significant with the result of different Boltzmann averages using enthalpies and free energies and hence different spectra. Therefore free energies including also the entropy are used in the Boltzmann averages in this thesis.

Another approach to take the conformer ensemble into account is to use how frequently a conformer appears in a MD or MTD simulation as a measure of how important that conformer is. This is done by geometry optimizing and calculating the spectral properties a given number of snapshots from the MD simulation and give each of them equal weight in the spectra. Although each conformer has the same weight, the assumption is that the preferred geometry will appear more frequently in the ensemble and hence give a "heavier" weight to the spectra by being included more times. In this thesis, all spectra based on the CREST conformer ensemble are Boltzmann averaged while for the spectra based on the MD simulation each of the extracted snap shots is equally weighted in the spectra. Since a significantly larger number of snap shots are needed in the latter approach, the computational cost is much higher. In **Paper III**, 200 snap shots from the MD simulation are used compared to 20 from the CREST ensemble, which makes the computational cost approximately ten times higher.

Chapter 4

Summary and outlook

In this thesis, a computational protocol to calculate VCD spectra of cyclic oligopeptides was established in **Paper I**. This protocol was then used to investigate if VCD alone can determine the AC of cyclic oligopeptides with two, three and four chiral centers in **Paper II** and the conformers of four cyclic hexapeptides in water and DMSO in **Paper III**.

The aim of **Paper I** was to test the performance of the conformational search tool CREST[4,5]. In addition, the effect of using different basis sets, adding dispersion corrections and using different energies for the Boltzmann average on the VCD spectra was tested. Systems for which experimental data was available and where the calculated spectra was based on a classical approach to find the conformers, were chosen from the literature and we showed that the overall VCD patterns are in good agreement with available experimental data. The recommended protocol consists of three steps: (1) conformational sampling with CREST and tight-binding density functional theory (xTB); (2) energy ranking based on single-point energy calculations as well as geometry optimization and VCD calculations of conformers that are within 2.5 kcal/mol of the most stable conformer using B3LYP/6-31+G*/CPCM; and (3) VCD spectra generation based on Boltzmann weighting with Gibbs free energies.

Although the overall pattern of the VCD spectra was in agreement with experiment, frequencies where in general poorly reproduced. Especially the gaps between the amide I region (1800-1600 cm^{-1}) and the amide II region (1600-1480 cm^{-1}) were poorly reproduced.

In **Paper II**, we used the computational protocol from **Paper I** to inves-

tigate the possibility to determine the AC of cyclic tetra- and hexapeptides with two, three and four chiral centers by comparing experimental VCD spectra with computed VCD spectra of all stereoisomers. We find that it is possible to unambiguously determine the peptides with two chiral centers with VCD. For the peptides with three and four chiral centers, VCD can at best limit the number of possible stereoisomers. Although disappointing, this result is not surprising. Other work has shown that combining VCD with other techniques such as ROA and NMR is a preferred strategy to correctly and reliable determine the stereochemistry of the molecules with several chiral centers.[20,96–98]

In **Paper III**, we investigated the structure of four cyclic peptides with antimicrobial potential using VCD. The experimental VCD spectra of the four peptides show similar patterns in the amide I region. This suggest, despite differences in the nature of the component amino acids and their sequence, that the four peptides have similar backbone structures. The computational protocol from **Paper I** predicted different backbone structures for the four peptides leading to computed spectra not in agreement with the experimental ones.

One of the peptides was thoroughly investigated by comparing the resulting conformers and VCD spectra from our protocol with a classical MD approach where explicit solvent molecules were present in the MD conformational sampling and then replaced by a continuum for the QM calculations. Our results show that the two computational approaches result in different conformational ensembles and different VCD spectra where the spectra simulated based on the MD ensemble is in better agreement with the experimental ones. In this work, in contrast to the systems investigated in **Paper I** and **Paper II**, the peptides have several long, flexible side chains with hydrogen-bond donor groups that can make hydrogen bonds with the backbone. While conformers in both ensembles have hydrogen bonding within the backbone, the CREST conformers also have hydrogen bonding between the side chain, which is not the case for the majority of the MD conformers.

To complete the work in **Paper III**, also spectra based on MD conformers for the remaining three peptides have to be included. If that provides calculated spectra in agreement with experiment, information about the backbone structure of the peptides can be extracted and it may be possible to obtain structural trends for the four peptides. In addition, a more

reliable investigation of the weaknesses of using CREST to sample the conformational space for these types of cyclic peptides can be determined.

In **Paper III** we find that the computational protocol established does not predict accurate spectra for the peptides with several long and charged side chains. One reason for this may be the lack of explicit solvent molecules during the conformational search. The developers of CREST recently added the possibility to run simulations using explicit solvent molecules.[99] Although a more traditional approach using snapshots from a MD simulation results in spectra in agreement with experiment, the computational cost is approximately 10 times higher than using the protocol from **Paper I**. Whether a CREST-based protocol including explicit solvation improves the results of our protocol in **Paper I**, while keeping an acceptable computational cost, is a question for further investigation.

As shown in **Paper I** and by others[96,100], adding dispersion corrections results in preferred conformers with more internal interactions. Using these energies for the Boltzmann averaged VCD spectra gives poorer agreement with experiment. The conformer ensembles predicted by CREST are based on dispersion corrected energies. The peptides investigated in **Paper III** have long and flexible side chains and CREST predicts conformers where the side chains preferentially interact with the backbone. It is possible that conformers with less internal interaction occur in the CREST simulation, but that these conformers are discarded and not included in the final ensemble due to their relatively higher dispersion corrected energies. It would be interesting to compare the CREST ensemble for these types of system with and without dispersion corrections.

Another drawback of our protocol is the poor agreement between experimental and calculated frequencies. This makes comparisons with experiment harder, particularly when the goal is to determine the stereochemistry by comparing calculated spectra of all possible stereoisomers with the experimental spectra. Although the poor agreement with experiment can, at least partially, be compensated for by applying scaling factors, there is a need for calculated frequencies of solvated systems in better agreement with experiment. Including explicit solvent molecules in the calculations might improve the frequencies and there are computational methods available: Cappelli and co-workers calculated accurate VCD spectra[82] while Dundas *et al.* calculated accurate IR spectra[81]. However, these methods

are quite computationally demanding and are at the moment feasible for smaller molecules than the ones investigated in this thesis and the method by Dundas *et al.* cannot calculate VCD properties.

The main focus of this thesis has been to on the computation of VCD spectra of cyclic peptides of sufficient quality to enable comparison with experimental data. The work done in both **Paper II** and **Paper III** would be strengthened by combining VCD with other techniques such as ROA and NMR. For ROA and the more complicated NMR techniques, calculated spectra are needed to extract information and with small adjustments, similar computational protocols used for the VCD spectra can be used also to calculate ROA and NMR spectra. With experimental expertise in place, these methods would complement the information extracted from VCD and give more insights into the structure and properties of cyclic peptides.

Bibliography

[1] P. J. Stephens, F. J. Devlin, and J. R. Cheeseman, *VCD spectroscopy for organic chemists*. CRC Press, 1st ed., 2012.

[2] L. A. Nafie, *Vibrational optical activity: principles and applications*. John Wiley & Sons Ltd., 1st ed., 2011.

[3] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, and F. Neese, "Fully automated quantum-chemistry-based computation of spin–spin-coupled nuclear magnetic resonance spectra," *Angew. Chem., Int. Ed.*, vol. 56, no. 46, pp. 14763–14769, 2017.

[4] S. Grimme, "Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations," *J. Chem. Theory Comput.*, vol. 15, no. 5, pp. 2847–2862, 2019.

[5] P. Pracht, F. Bohle, and S. Grimme, "Automated exploration of the low-energy chemical space with fast quantum chemical methods," *Phys. Chem. Chem. Phys.*, vol. 22, pp. 7169–7192, 2020.

[6] A. D. Becke, "A new mixing of Hartree Fock and local density functional theories," *J. Chem. Phys.*, vol. 98, no. 2, pp. 1372–1377, 1993.

[7] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785–789, 1988.

[8] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "*Ab Initio* calculation of vibrational absorption and circular dichroism

spectra using density functional force fields," *J. Phys. Chem.*, vol. 98, no. 45, pp. 11623–11627, 1994.

[9] A. D. McLean and G. S. Chandler, "Contracted Gaussian basis sets for molecular calculations. I. second row atoms, Z=11-18," *J. Chem. Phys.*, vol. 72, no. 10, pp. 5639–5648, 1980.

[10] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions," *J. Chem. Phys.*, vol. 72, no. 1, pp. 650–654, 1980.

[11] R. S. Cahn, "An introduction to the sequence rule: A system for the specification of absolute configuration," *J. Chem. Educ.*, vol. 41, no. 3, p. 116, 1964.

[12] R. S. Cahn, C. Ingold, and V. Prelog, "Specification of molecular chirality," *Angew. Chem., Int. Ed. Engl.*, vol. 5, no. 4, pp. 385–415, 1966.

[13] A. D. McNaught and A. Wilkinson, "IUPAC. compendium of chemical terminology," in *Gold Book*, Blackwell Scientific Publications, 2nd ed., 1997. Online version (2019-).

[14] G. P. Moss, "Basic terminology of stereochemistry (IUPAC recommendations 1996)," *Pure Appl. Chem.*, vol. 68, no. 12, pp. 2193–2222, 1996.

[15] D. Chattopadhyay and M. Bharti, "Implementation of C-PolSK modulation in FSO network and analysis of its atmospheric performance," *J. Telecommun. Inf. Technol.*, no. 4, pp. 43–49, 2019.

[16] "Wikipedia: Circular polarization," 2022. https://en.wikipedia.org/wiki/Circular_polarization (accessed: 04.04.2022).

[17] C. Merten, F. Li, K. Bravo-Rodriguez, E. Sanchez-Garcia, Y. Xu, and W. Sander, "Solvent-induced conformational changes in cyclic peptides: a vibrational circular dichroism study," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 5627–5633, 2014.

[18] A. R. Leach, *Molecular Modeling Principles and Applications*. Pearson Education Limited, 2nd ed., 2001.

[19] R. Aerts, J. Vanhove, W. Herrebout, and C. Johannessen, "Paving the way to conformationally unravel complex glycopeptide antibiotics by means of Raman optical activity," *Chem. Sci.*, vol. 12, no. 16, pp. 5952–5964, 2021.

[20] C. Merten, T. P. Golub, and N. M. Kreienborg, "Absolute configurations of synthetic molecular scaffolds from vibrational CD spectroscopy," *J. Org. Chem.*, vol. 84, no. 14, pp. 8797–8814, 2019.

[21] P. L. Polavarapu, "Determination of the absolute configurations of chiral drugs using chiroptical spectroscopy," *Molecules*, vol. 21, no. 8, p. 1056, 2016.

[22] G. Magyarfalvi, G. Tarczay, and E. Vass, "Vibrational circular dichroism," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, vol. 1, no. 3, pp. 403–425, 2011.

[23] K. H. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dračínský, and P. Bouř, "Explicit versus implicit solvent modeling of Raman optical activity spectra," *J. Phys. Chem. B*, vol. 115, no. 14, pp. 4128–4137, 2011.

[24] J. Jungwirth, J. Šebestík, M. Šafařík, J. Kapitán, and P. Bouř, "Quantitative determination of Ala-Ala conformer ratios in solution by decomposition of Raman optical activity spectra," *J. Phys. Chem. B*, vol. 121, no. 38, pp. 8956–8964, 2017.

[25] D. D. Beusen, E. Berkley Shands, S. Karasek, G. R. Marshall, and R. A. Dammkoehler, "Systematic search in conformational analysis," *J. Mol. Struct.: THEOCHEM*, vol. 370, no. 2, pp. 157–171, 1996.

[26] E. P. Jaeger, M. L. Peterson, and A. M. Treasurywala, "Conformational energy downward driver (CEDD): Characterization and calibration of the method," *J. Comput.-Aided Mol. Des.*, vol. 9, no. 1, pp. 55–64, 1995.

[27] H. Goto and E. Osawa, "Corner flapping: a simple and fast algorithm for exhaustive generation of ring conformations," *J. Am. Chem. Soc.*, vol. 111, no. 24, pp. 8950–8951, 1989.

[28] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 20, pp. 12562–12566, 2002.

[29] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[30] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[31] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *J. Am. Chem. Soc.*, vol. 110, no. 6, pp. 1657–1666, 1988.

[32] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, 1996.

[33] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," *J. Phys. Chem. B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[34] F. Jiang, C.-Y. Zhou, and Y.-D. Wu, "Residue-specific force field based on the protein coil library. RSFF1: Modification of OPLS-AA/L," *J. Phys. Chem. B*, vol. 118, no. 25, pp. 6983–6998, 2014.

[35] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models.* John Wiley & Sons Ltd, 2nd ed., 2004.

[36] F. Jensen, *Introduction to Computational Chemistry.* John Wiley & Sons, 2nd ed., 2007.

[37] G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, "Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model," *J. Phys. Chem.*, vol. 98, no. 49, pp. 13077–13082, 1994.

[38] N. C. Fitzkee, P. J. Fleming, and G. D. Rose, "The protein coil library: A structural database of nonhelix, nonstrand fragments derived from the PDB," *Proteins: Struct., Funct., Bioinf.*, vol. 58, no. 4, pp. 852–854, 2005.

[39] L. Verlet, "Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1967.

[40] R. Hockney, S. Goel, and J. Eastwood, "Quiet high-resolution computer models of a plasma," *J. Comput. Phys.*, vol. 14, no. 2, pp. 148–158, 1974.

[41] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19–25, 2015.

[42] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977.

[43] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997.

[44] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, vol. 1, no. 5, pp. 826–843, 2011.

[45] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry.* Macmilan Publishing Co., Inc., 1st ed., 1982.

[46] P. Atkins and R. Friedman, *Molecular Quantum Mechanics.* Oxford University Press Inc., 5th ed., 2011.

[47] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864–B871, 1964.

[48] D. J. Tozer, "Density functional theory," in *European Summerschool in Quantum Chemistry. Book II* (S. Reine and T. Saue, eds.), pp. 537–575, ESQC committee, 11th ed., 2019.

[49] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, pp. A1133–A1138, 1965.

[50] P. A. M. Dirac, "Note on exchange phenomena in the Thomas atom," *Math. Proc. Cambridge Philos. Soc.*, vol. 26, no. 3, p. 376–385, 1930.

[51] J. C. Slater, "A simplification of the Hartree-Fock method," *Phys. Rev.*, vol. 81, pp. 385–390, 1951.

[52] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Can. J. Phys.*, vol. 58, no. 8, pp. 1200–1211, 1980.

[53] J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," *Phys. Rev. B*, vol. 45, pp. 13244–13249, 1992.

[54] A. D. Becke, "Density functional calculations of molecular bond energies," *J. Chem. Phys.*, vol. 84, no. 8, pp. 4524–4529, 1986.

[55] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, 1996.

[56] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, 1988.

[57] Y. Zhao and D. G. Truhlar, "A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions," *J. Chem. Phys.*, vol. 125, no. 19, p. 194101, 2006.

[58] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes," *J. Chem. Phys.*, vol. 119, no. 23, pp. 12129–12137, 2003.

[59] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Erratum: "comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes"," *J. Chem. Phys.*, vol. 121, no. 22, pp. 11507–11507, 2004.

[60] M. Ernzerhof and G. E. Scuseria, "Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional," *J. Chem. Phys.*, vol. 110, no. 11, pp. 5029–5036, 1999.

[61] C. Adamo and V. Barone, "Accurate excitation energies from time-dependent density functional theory: assessing the PBE0 model for organic free radicals," *Chem. Phys. Lett.*, vol. 314, no. 1, pp. 152–157, 1999.

[62] P. J. Stephens and F. J. Devlin, "Determination of the structure of chiral molecules using *ab initio* vibrational circular dichroism spectroscopy," *Chirality*, vol. 12, no. 4, pp. 172–179, 2000.

[63] F. J. Devlin, P. J. Stephens, J. R. Cheeseman, and M. J. Frisch, "*Ab Initio* prediction of vibrational absorption and circular dichroism spectra of chiral natural products using density functional theory: α-pinene," *J. Phys. Chem. A*, vol. 101, no. 51, pp. 9912–9924, 1997.

[64] P. J. Stephens, F. J. Devlin, and J.-J. Pan, "The determination of the absolute configurations of chiral molecules using vibrational circular dichroism (VCD) spectroscopy," *Chirality*, vol. 20, no. 5, pp. 643–663, 2008.

[65] T. Kuppens, W. Langenaeker, J. P. Tollenaere, and P. Bultinck, "Determination of the stereochemistry of 3-hydroxymethyl-2,3-dihydro-[1,4]dioxino[2,3-*b*]-pyridine by vibrational circular dichroism and the effect of DFT integration grids," *J. Phys. Chem. A*, vol. 107, no. 4, pp. 542–553, 2003.

[66] M. P. Andersson and P. Uvdal, "New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple-ζ basis set 6-311+g(d,p)," *J. Phys. Chem. A*, vol. 109, no. 12, pp. 2937–2941, 2005.

[67] G. Seifert, D. Porezag, and T. Frauenheim, "Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme," *Int. J. Quantum Chem.*, vol. 58, no. 2, pp. 185–192, 1996.

[68] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, "Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon," *Phys. Rev. B*, vol. 51, pp. 12947–12957, 1995.

[69] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Phys. Rev. B*, vol. 58, pp. 7260–7268, 1998.

[70] Q. Cui and M. Elstner, "Density functional tight binding: values of semi-empirical methods in an *ab initio* era," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 14368–14377, 2014.

[71] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, "Extended tight-binding quantum chemistry methods," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, vol. 11, no. 2, p. e1493, 2021.

[72] M. Elstner and G. Seifert, "Density functional tight binding," *Philos. Trans. R. Soc., A*, vol. 372, no. 2011, p. 20120483, 2014.

[73] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, "Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications," *Chem. Rev. (Washington, DC, U. S.)*, vol. 116, no. 9, pp. 5301–5337, 2016.

[74] M. Gaus, C.-P. Chou, H. Witek, and M. Elstner, "Automatized parametrization of SCC-DFTB repulsive potentials: Application to hydrocarbons," *J. Phys. Chem. A*, vol. 113, no. 43, pp. 11866–11881, 2009.

[75] M. Gaus, Q. Cui, and M. Elstner, "Density functional tight binding: application to organic and biological molecules," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, vol. 4, no. 1, pp. 49–61, 2014.

[76] E. Caldeweyher, C. Bannwarth, and S. Grimme, "Extension of the D3 dispersion coefficient model," *J. Chem. Phys.*, vol. 147, no. 3, p. 034112, 2017.

[77] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-charge dependent london dispersion correction," *J. Chem. Phys.*, vol. 150, no. 15, p. 154122, 2019.

[78] E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme, "Extension and evaluation of the D4 london-dispersion model for periodic systems," *Phys. Chem. Chem. Phys.*, vol. 22, pp. 8499–8512, 2020.

[79] C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB - an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions," *J. Chem. Theory Comput.*, vol. 15, no. 3, pp. 1652–1671, 2019.

[80] B. Mennucci, "Modeling environment effects on spectroscopies through QM/classical models," *Phys. Chem. Chem. Phys.*, vol. 15, pp. 6583–6594, 2013.

[81] K. O. H. M. Dundas, M. T. P. Beerepoot, M. Ringholm, S. Reine, R. Bast, N. H. List, J. Kongsted, K. Ruud, and J. M. H. Olsen, "Harmonic infrared and Raman spectra in molecular environments using the polarizable embedding model," *J. Chem. Theory Comput.*, vol. 17, no. 6, pp. 3599–3617, 2021.

[82] T. Giovannini, G. Del Frate, P. Lafiosca, and C. Cappelli, "Effective computational route towards vibrational optical activity spectra of chiral molecules in aqueous solution," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 9181–9197, 2018.

[83] A. V. Onufriev, "The generalized born model : its foundation, applications, and limitations," 2010. https://people.cs.vt.edu/~onufriev/PUBLICATIONS/gbreview.pdf (accessed: 12.04.2022).

[84] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *J. Am. Chem. Soc.*, vol. 112, no. 16, pp. 6127–6129, 1990.

[85] G. Scalmani, V. Barone, K. N. Kudin, C. S. Pomelli, G. E. Scuseria, and M. J. Frisch, "Achieving linear-scaling computational cost for the polarizable continuum model of solvation," *Theor. Chem. Acc.*, vol. 111, no. 2-6, pp. 90–100, 2004.

[86] H. B. Schlegel, "Geometry optimization," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, vol. 1, no. 5, pp. 790–809, 2011.

[87] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian 16 Revision B.01," 2016. Gaussian Inc. Wallingford CT.

[88] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *J. Inst. Math. Its Appl.*, vol. 6, no. 1, pp. 76–90, 1970.

[89] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, no. 3, pp. 317–322, 1970.

[90] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comput.*, vol. 24, no. 109, pp. 23–26, 1970.

[91] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Math. Comput.*, vol. 24, no. 109, pp. 647–657, 1970.

[92] J. P. Merrick, D. Moran, and L. Radom, "An evaluation of harmonic vibrational frequency scale factors," *J. Phys. Chem. A*, vol. 111, no. 45, pp. 11683–11700, 2007.

[93] P. Norman, K. Ruud, and T. Saue, *Principles and practices of molecular properties: theory, modeling and simulations.* John Wiley & Sons Ltd, 1st ed., 2018.

[94] P. J. Stephens, "Theory of vibrational circular dichroism," *J. Phys. Chem.*, vol. 89, no. 5, pp. 748–752, 1985.

[95] P. J. Stephens, "Evaluation of theories of vibrational magnetic dipole transition moments using atomic axial tensor sum rules and paramagnetic susceptibilities," *Chem. Phys. Lett.*, vol. 180, no. 5, pp. 472–476, 1991.

[96] K. H. Hopmann, J. Šebestík, J. Novotná, W. Stensen, M. Urbanová, J. Svenson, J. S. Svendsen, P. Bouř, and K. Ruud, "Determining the absolute configuration of two marine compounds using vibrational chiroptical spectroscopy," *J. Org. Chem.*, vol. 77, no. 2, pp. 858–869, 2012.

[97] W. Dinku, J. Isaksson, F. G. Rylandsholm, P. Bouř, E. Brichtová, S. U. Choi, S.-H. Lee, Y.-S. Jung, Z. S. No, J. S. M. Svendsen, A. J. Aasen, and A. Dekebo, "Anti-proliferative activity of a novel tricyclic triterpenoid acid from *Commiphora africana* resin against four human cancer cell lines," *Appl. Biol. Chem.*, vol. 63, no. 1, pp. 1–11, 2020.

[98] J. Bogaerts, R. Aerts, T. Vermeyen, C. Johannessen, W. Herrebout, and J. M. Batista, "Tackling stereochemistry in drug molecules with vibrational optical activity," *Pharmaceuticals*, vol. 14, no. 9, p. 877, 2021.

[99] S. Spicher, C. Plett, P. Pracht, A. Hansen, and S. Grimme, "Automated molecular cluster growing for explicit solvation by efficient force field and tight binding methods," *Journal of Chemical Theory and Computation*, vol. 18, no. 5, pp. 3174–3189, 2022. PMID: 35482317.

[100] M. A. J. Koenis, O. Visser, L. Visscher, W. J. Buma, and V. P. Nicu, "Gui implementation of vcdtools, a program to analyze computed vibrational circular dichroism spectra," *J. Chem. Inf. Model.*, vol. 60, no. 1, pp. 259–267, 2020.

Paper I

A computational protocol for vibrational circular dichroism spectra of cyclic oligopeptides

Karolina Di Remigio Eikås,
Maarten T. P. Beerepoot and Kenneth Ruud

J. Phys. Chem. A **126** (2022), 5458–5471,
https://doi.org/10.1021/acs.jpca.2c02953

Replication data available at:
https://doi.org/10.18710/VITGV3

THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

A Computational Protocol for Vibrational Circular Dichroism Spectra of Cyclic Oligopeptides

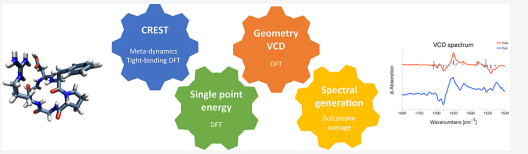
Karolina Di Remigio Eikås,* Maarten T. P. Beerepoot, and Kenneth Ruud*

Cite This: *J. Phys. Chem. A* 2022, 126, 5458−5471

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cyclic peptides are a promising class of compounds for next-generation antibiotics as they may provide new ways of limiting antibiotic resistance development. Although their cyclic structure will introduce some rigidity, their conformational space is large and they usually have multiple chiral centers that give rise to a wide range of possible stereoisomers. Chiroptical spectroscopies such as vibrational circular dichroism (VCD) are used to assign stereochemistry and discriminate enantiomers of chiral molecules, often in combination with electronic structure methods. The reliable determination of the absolute configuration of cyclic peptides will require robust computational methods than can identify all significant conformers and their relative population and reliably assign their stereochemistry from their chiroptical spectra by comparison with *ab initio* calculated spectra. We here present a computational protocol for the accurate calculation of the VCD spectra of a series of flexible cyclic oligopeptides. The protocol builds on the Conformer-Rotamer Ensemble Sampling Tool (CREST) developed by Grimme and co-workers (*Phys. Chem. Chem. Phys.* **2020**, *22*, 7169−7192 and *J. Chem. Theory. Comput.* **2019**, *15*, 2847−2862) in combination with postoptimizations using B3LYP and moderately sized basis sets. Our recommended computational protocol for the computation of VCD spectra of cyclic oligopeptides consists of three steps: (1) conformational sampling with CREST and tight-binding density functional theory (xTB); (2) energy ranking based on single-point energy calculations as well as geometry optimization and VCD calculations of conformers that are within 2.5 kcal/mol of the most stable conformer using B3LYP/6-31+G*/CPCM; and (3) VCD spectra generation based on Boltzmann weighting with Gibbs free energies. Our protocol provides a feasible basis for generating VCD spectra also for larger cyclic peptides of biological/pharmaceutical interest and can thus be used to investigate promising compounds for next-generation antibiotics.



INTRODUCTION

Chirality is a key property of many biological systems. The two enantiomers of a chiral molecule may have very different biological functions[1] with the extreme case being one enantiomer with a biological effect as a drug and the mirror image with an adverse effect when administered to a patient. This makes it important to be able to identify the stereochemistry of chiral molecules and to devise stereoselective synthetic pathways that can ensure that a particular enantiomer is synthesized. As almost all physical properties of two enantiomers are the same, the identification and separation of different enantiomers is challenging. The observation that enantiomers of chiral molecules interact differently with circularly polarized light has been used to develop a wide range of different chiroptical spectroscopies[2−4] in which the differential absorption or scattering of right- and left-circularly polarized light gives rise to a different sign for the two enantiomers. However, there is no way to *a priori* connect the sign of the differential absorption of the components of the circularly polarized light to the absolute stereochemistry of the molecule, making it important to compare the experimental spectra with spectra calculated using electronic structure methods.

Circularly polarized light can be generated in a wide frequency range. Historically, chiroptical spectroscopies in the optical and near-UV region, first measured by Cotton in 1895,[5,6] have been the most important methods for the discrimination of enantiomers. Indeed, optical rotation remains today a key quantity to determine, for instance, enantiomeric excess and to verify the stereochemistry of a chiral molecule. In order to obtain more insight into the electronic structure of chiral molecules as well as to determine the amount of α-helical content in proteins, electronic circular dichroism (CD) is commonly used.[7−9] A significant limitation of both optical rotation and CD for the determination of the chirality of small molecules is the limited number of electronic excited states available in the wavelengths accessible to experimental investigation in the UV/vis region. This is particularly critical in the case of organic molecules, which often are colorless due

Received: April 28, 2022
Revised: July 18, 2022
Published: August 5, 2022

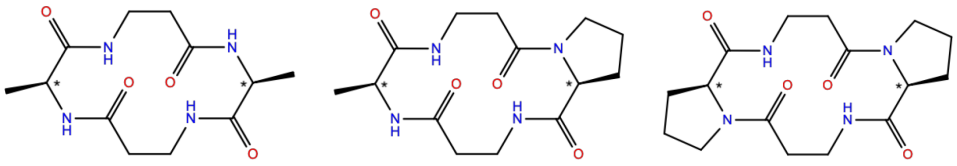


Figure 1. Structure of cyclo(**X**-*β*-Ala-**Y**-*β*-Ala) with the chiral centers indicated by an asterisk.

to the fact that there are no low-lying excited states absorbing in the visible region, leaving only a very narrow window in the near-UV region accessible to modern detectors. These challenges are further accentuated when multiple chiral centers are present in the molecule. Furthermore, the description of electronic excited states using for instance density-functional theory (DFT) is fraught with challenges.[10,11] In contrast, methods that treat electron correlation explicitly are faced with challenges arising from the size of the molecules normally involved and the lack of symmetry.[12−16]

In contrast, the infrared region is a rich source of information for structural characterization and identification of chiral molecules. Even small molecules have a large number of vibrational modes that can be probed with spectroscopies using the infared region of the electromagnetic spectrum, providing a much richer source of experimental data compared to electronic spectroscopies. Many of these vibrational modes can be associated with specific, local regions of the molecule. From a computational point of view, vibrational chiroptical spectroscopy also benefits from the fact that all relevant quantities can be calculated from a knowledge of the electronic ground state only,[17,18] which in general can be determined more accurately than electronic excited states.

Since their first observations in the early 70s,[19−21] vibrational chiroptical spectroscopies have evolved into some of the most important techniques by which the absolute stereochemistry of a molecule can be determined in a combined experimental/theoretical approach.[22,23] The two most important vibrational chiroptical techniques are vibrational circular dichroism (VCD)[24−27] in which one measures the differential absorption of left- and right-circularly polarized light, and Raman optical activity (ROA), in which one measures the differential scattering of right- and left-circularly polarized light.[2,28] VCD and ROA are the chiroptical analogues of infrared absorption spectroscopy and Raman scattering, respectively. Of these, VCD is the most common approach, to a large extent due to the availability of several different commercial instruments and easier operation. VCD has found a wide range of applications, including structural characterization of small molecules,[29−34] understanding the secondary structure of peptides and proteins[35−42] as well as for the understanding of molecular behavior and interactions in solutions,[43−51] and recently also for understanding interactions in larger molecular systems such as fibrils.[52−54] The sensitivity of the approach to even small changes in molecular structure is one of the strengths of the technique. At the same time, this sensitivity puts severe demands on the robustness of the computational model in general, and the description of molecular interactions and solvent effects in particular.

From a computational point of view, a challenge in the study of larger chiral systems is conformational flexibility due to the strong dependence of the chiroptical response on the three-dimensional structure of a molecule.[55−57] Indeed, the sign of the optical rotation as well as VCD absorption bands may change for different conformations of a molecule. The strong sensitivity on the molecular conformation also means that the calculated chiroptical response will be very sensitive to conformational sampling, both in terms of identifying all conformations present in solution, as well as the quality of the energetics used to determine the Boltzmann population of individual conformations. The complexity of this problem increases with increasing molecular size and the number of chiral centers in the molecule. The conformational sampling is thus crucial in order to reliably assign the stereochemistry of these molecules and is as such an integral part of any computation of chiroptical spectra of flexible molecules. The majority of studies in the literature use *ad hoc* schemes for conformational sampling, either based on chemical intuition or on sampling from molecular dynamics (MD) simulations.[58] In the latter case, it is important that the conformational space is sufficiently well sampled.[59] The conformational challenge gets further complicated when the molecule of interest also can form stable structures with solvent molecules, for instance, through hydrogen bonding. An early study of these effects was presented by Hopmann et al.,[59] discussing, for instance, the use of either the enthalpy or the free energy for determining Boltzmann weights as well as the challenges of comparing complexes with different number of solute−solvent bonds. Despite a large number of studies of chiroptical responses of conformationally flexible molecules in recent years,[58−63] there is a need for a more systematic approach to the sampling of conformations for flexible molecules to provide a robust and reliable approach for calculating vibrational chiroptical spectra of chiral molecules with multiple chiral centers.

At the same time, the conformational challenge is not unique to chiroptical spectroscopy. A number of strategies for sampling the conformation space of small organic molecules have been presented.[64−68] A particular promising approach was recently presented by Grimme and co-workers,[69−71] originally developed for the study of spectroscopic properties of flexible molecules with the first application being the calculation of the nuclear magnetic resonance (NMR) spectrum of the cyclic ionophore nonactin in addition to a few other (in)organic and transition-metal complexes. The approach has since been further refined and applied to the calculation of different spectroscopies, including VCD.[71−75] As such, this approach appears as a promising starting point for a robust and reliable protocol for calculating VCD spectra of conformationally flexible molecules with multiple stereocenters.

Our primary targets are cyclic oligopeptides in the so-called middle space,[76] as these systems hold promise for novel actions as antibiotics. It is important that the computational protocol is tested on well-known systems that share similarities with the compounds of interest, while at the same time being simple

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

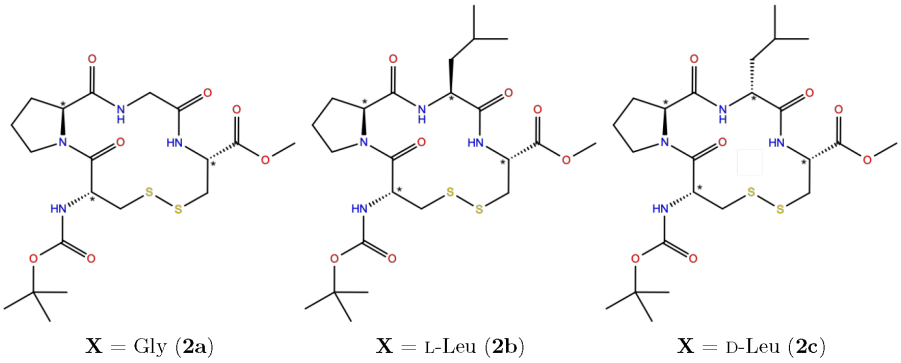


Figure 2. Structure of cyclo(Boc-Cys-Pro-**X**-Cys-OMe) with the chiral centers indicated by an asterisk.

enough to allow the protocol to be suitably optimized and its range of applicability assessed. For this purpose we have selected three classes of cyclic oligopeptides previously reported in the literature.[77−79] The first and simplest system is a series of three tetrapeptides of the form cyclo(**X**-*β*-Ala-**Y**-*β*-Ala) with **X** and **Y** both being either proline or alanine (see Figure 1). The second class is another series of three tetrapeptides of the form cyclo(Boc-Cys-Pro-**X**-Cys-OMe), where **X** can be glycine, L-leucine, or D-leucine, respectively (Figure 2). Finally, we apply the protocol to the hexapeptide cyclo(Phe-D-Pro-Gly-Arg-Gly-Asp) (Figure 3).

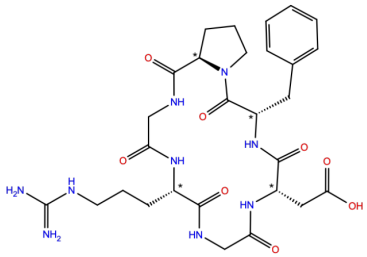


Figure 3. Structure of cyclo(Phe-D-Pro-Gly-Arg-Gly-Asp) (**3**) with the chiral centers indicated by an asterisk.

The rest of the paper is organized as follows: We first describe the details of our computational protocol, including the conformational sampling, the selection of conformers, their energy minimization and the calculation of VCD spectra. We then turn our attention to the optimization and use of the computational protocol for our reference systems, before we end by giving some concluding remarks and an outlook.

COMPUTATIONAL DETAILS

Our computational protocol for the calculation of VCD spectra of conformationally flexible molecules such as the cyclic oligopeptides that are our primary targets, consists of three steps that are summarized in Figure 4: (1) sampling of the conformational space, (2) selection of the conformers that can be expected to be important for the VCD spectrum, structural optimization of these conformers and calculation of VCD spectra, and finally (3) the generation of the Boltzmann-averaged VCD spectrum that can be compared to the experimentally observed spectrum.

Step 1: Conformational Sampling. We start by considering the selection of relevant conformers for which we used the Conformer Rotamer Ensemble Sampling Tool (CREST), version 2.8.[69−71] In our calculations on molecules **1a−c** (Figure 1), **2a−c** (Figure 2), and **3** (Figure 3), we used the iMDT-GC workflow[70,71] and the default settings with the exception of the simulation length of **1b**. The iMDT-GC workflow consists of three steps: an extensive meta-dynamic sampling (MTD) with different bias potentials, a molecular dynamics (MD) sampling around the lowest-energy conformers, and a genetic Z-matrix crossing (GC).[70,71,80] The GC compares each pair of conformers i and j and adds the structural differences $(R_i - R_j)$ to a reference structure to generate a new conformer $(R_{new} = R_{ref} + (R_i - R_j))$. The reference structure, R_{ref}, is usually the conformer lowest in energy.

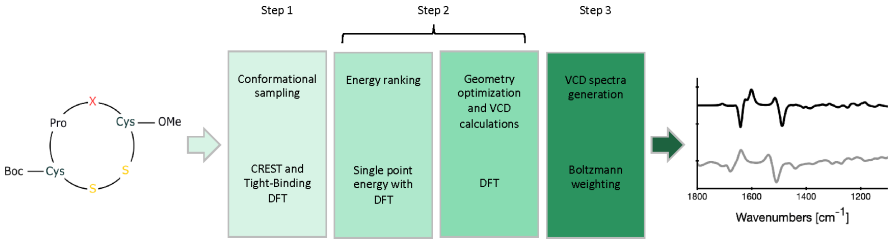


Figure 4. Overview of the protocol presented in this work.

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

The following describes the default CREST settings with the iMDT-GC workflow: On the basis of a flexibility measure for the molecule, CREST determines the length of the MTD simulation, which typically is 0.3−0.4 ps multiplied by the number of atoms in the molecule. The resulting simulation length for each molecule in this work is given in Table 1. A

Table 1. Computational Settings for the Conformer Rotamer Ensemble Sampling Tool (CREST) for Each of the Molecules Studied in This Work

	molecule	MTD time (ps)	MD time (ps)
1a	**X** = **Y** = Ala	13.0	6.5
1b	**X** = Ala, **Y** = Pro	90.0[a]	45.0[a]
1c	**X** = **Y** = Pro	11.0	5.5
2a	**X** = Gly	21.0	10.5
2b	**X** = L-Leu	28.0	14.0
2c	**X** = D-Leu	28.0	14.0
3		36.0	18.0

[a]Because of difficulties with finding conformers, the simulation time was set manually.

time step of 5.0 fs was used, coordinates were sampled every 100 fs, and a new reference structure for the bias potential was chosen every 1.0 ps.[80] The bias potential used for the MTDs combines 3 different prefactors (k), the number of atoms (N) times 0.00300, 0.0015, and 0.000075, and 4 exponents (α), 1.300, 0.780, 0.468 and 0.281, resulting in 12 different bias potentials and thus 12 MTD sequences. In addition, two MTD sequences with extreme bias potentials ($k = N \times 0.0010$, $\alpha = 0.1$ and $k = N \times 0.0005$, $\alpha = 0.8$) were performed. For each snapshot in the simulations, a loose geometry optimization with the tight-binding method GFN-xTB[81,82] was performed. If this resulted in a conformer lower in energy than the input geometry, the MTD procedure was restarted with the geometry of this conformer. This was done at least once and no more than five times. For the four conformers lowest in energy, MD simulations were performed at 400 and 500 K, respectively. Finally, geometry optimizations using GFN-xTB with tighter convergence thresholds were performed for the conformers obtained from both the MTD and MD simulations. All structures thus obtained were compared and duplicates removed. CREST identifies duplicates based on three criteria: the total energy, the root-mean-square-deviation of atomic Cartesian coordinates, and the difference in the rotational constants. The final ensemble of unique conformers was then generated by performing a geometry optimization with very tight convergence criteria, again using GFN-xTB.

For molecule **1b**, the simulation length was manually set. For **1a** and **1c**, the conformers sampled with the simulation time determined by CREST resulted in VCD spectra in agreement with experiment and a conformer ensemble including the conformer determined most important by Vass et al.[77] For **1b**, on the other hand, the simulated spectrum was not in agreement with experiment. CREST had difficulties finding the relevant conformers and the conformer determined as most important by Vass et al. was not in the ensemble. By increasing the simulation time from 12.0 ps suggested by CREST to 90.0 ps, the conformer ensemble included the relevant conformers.

In CREST 2.8, the solvent is described by a Generalized Born model where a solvent-accessible surface is used. The experimental data[77−79] for **1b−c** and **2a−c** were recorded in deuterated acetonitrile (ACN-d_3) and ACN ($\epsilon = 36$) was used in the calculation. **1a** was measured in deuterated trifluoroethanol (TFE-d_2) while **3** was measured in normal trifluoroethanol (TFE). Since TFE ($\epsilon = 27$) is not available in CREST, we chose the available solvent with the most similar dielectric constant, acetone ($\epsilon = 20$).

Step 2: Energy Ranking and VCD Calculations. Having in step 1 identified a possibly complete set of unique conformers, we next turn to the calculation of VCD spectra for the energetically relevant conformers. In selecting the relevant conformers, we perform single-point energy calculations at the DFT level of theory, following the recommendation by Grimme.[70] Single-point energies, geometry optimizations and VCD calculations were performed using DFT in Gaussian 16 (Rev. B.01).[83] Building on previous studies of VCD calculations at the DFT level of theory,[29,84−87] the B3LYP functional[26,88,89] has been used for these calculations in combination with the 6-31+G*[90,91] basis set. To evaluate the effect of the size of the basis set on the resulting VCD spectra for cyclic oligopeptides, single-point energies, geometry optimization, and VCD calculations were also performed using B3LYP with the 6-31+G, 6-31+G**, 6-31G*, 6-31++G*, and 6-311+G* basis sets for molecules **1a** and **2a**. These results can be found in the Supporting Information. In addition, Grimmes empirical dispersion correction D3 with Becke-Johnson damping[92] was tested for molecules **1a** and **2a** (B3LYP-D3). Solvent effects were included using the conductor-like polarizable continuum model (CPCM).[93,94]

Geometry optimization and VCD calculations were performed on conformers that were at most 2.5 kcal/mol higher in energy than the most stable conformer after the single-point calculations. Both larger and smaller energy gaps where tested (*vide infra*). After geometry optimization with DFT, many conformers determined to be unique by CREST end up in the same minimum. Duplicates were removed with a script[95] using the same set of criteria as used in CREST.

Step 3: VCD Spectra Generation. Boltzmann weights for all unique conformers identified after the DFT geometry optimization were used to generate VCD spectra in the DrawSpectrum program.[96] Gibbs free energies for the conformers were used to calculate the Boltzmann weights if not otherwise specified.

When comparing calculated spectra only, frequency scaling factors from Merrick, Moran, and Radom were used.[97] When comparing calculated and experimental spectra, the frequencies were scaled such that the combined IR and VCD overlap estimates (eq 1) were maximized. A more detailed discussion of scaling factors when calculated spectra are compared with experimental spectra is found in the Supporting Information (SI). The overlap estimate between two spectra a and b was calculated with DrawSpectrum as[86,98−100]

$$S_{ab} = \frac{\langle I_a | I_b \rangle}{\sqrt{\langle I_a | I_a \rangle \langle I_b | I_b \rangle}} \quad (1)$$

where I_a is the spectral intensity at a given wavenumber for spectrum a. S_{ab} ranges from −1 to 1, where 1 indicates identical spectra and −1 perfect mirror-image spectra. The frequency regions used were 1800−1500 cm^{-1} for molecules **1a−1c** and **3** and 1800−1100 cm^{-1} for molecules **2a−c**.

The spectral line shape was simulated with a Lorentzian function with a full width at half-maximum of 10 cm^{-1} for

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

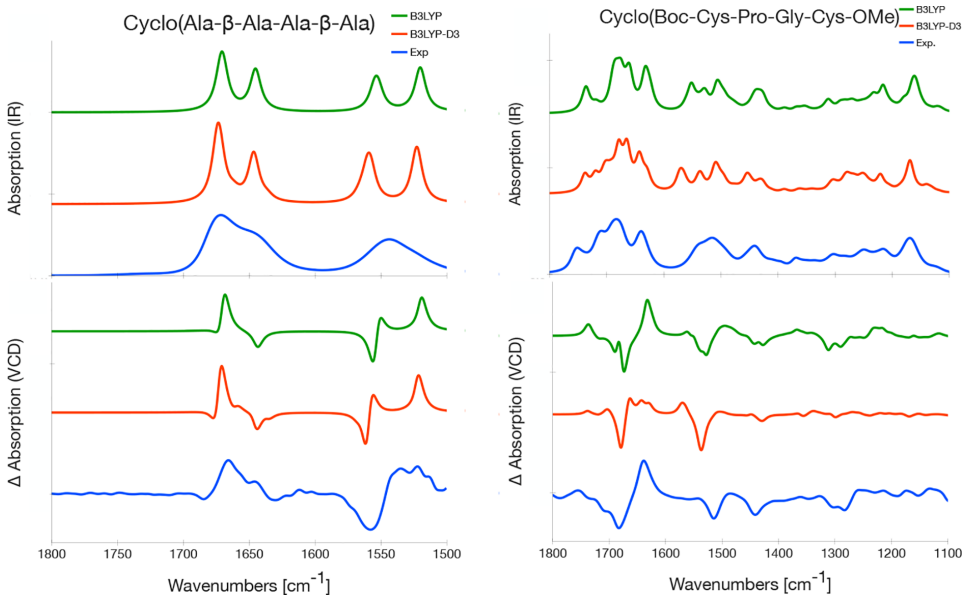


Figure 5. IR and VCD spectra of **1a** in TFE-d_2 (left) and **2a** in ACN-d_3 (right) with B3LYP/6-31+G*/CPCM and B3LYP-D3/6-31+G*/CPCM, compared with experiment. The frequencies were scaled with a factor of 0.975 for B3LYP-D3 on **2a** and 0.980 for B3LYP on **2a** as well as for all calculations on **1a**. The experimental spectrum of **1a** is measured by Vass et al.[77] while the one of **2a** by Merten et al.[78]

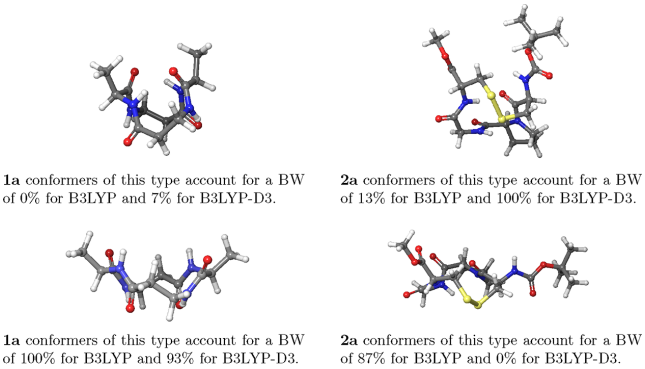


Figure 6. Computed geometries of **1a** (left) and **2a** (right) with and without dispersion interactions. The conformers can be grouped in two different types: the U-type (top row, with dispersion interactions) and flat type (bottom row, without dispersion interactions). The Boltzmann weights given in the figure are the sum of the conformers belonging to that group.

molecules **1a**–**c**[77] and **3**,[79] and 16 cm^{-1} for molecules **2a**–**c**.[78] The intensities in the calculated spectra are scaled only when comparing to experimental spectra, and in these cases the intensities are scaled such that the maximum intensity of the strongest absorption band matches the intensity of the corresponding band in the experimental spectrum.

RESULTS AND DISCUSSION

We now turn our attention to the optimization of the protocol by testing the effect of adding dispersion corrections, the selection of significant conformers, and an analysis of the relevant conformers. We then use the optimized protocol to predict IR and VCD spectra and compare to the experimental spectra of the selected molecules.

Dispersion Corrections. Dispersion corrections are automatically included in CREST calculations (step 1) through the D4 dispersion model.[101−103] We here test the effect of including dispersion corrections also in steps 2 and 3, that is, single point energies, geometries, and VCD properties calculated with DFT. The effect of adding dispersion corrections has been tested on molecules **1a** and **2a** and the results are shown in Figure 5 (IR and VCD spectra), Figure 6 (computed geometries) and Table 2 (overlap estimates S).

The IR spectra with and without dispersion corrections are similar, both qualitatively (Figure 5, top) and quantitatively

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

Table 2. Overlap Estimate S between Calculated and Experimental Spectra with B3LYP and B3LYP-D3[a]

	molecule **1a**		molecule **2a**	
	IR	VCD	IR	VCD
B3LYP	0.86	0.68	0.93	0.70
B3LYP-D3	0.86	0.63	0.93	0.46

[a]S is calculated over the frequency range 1800−1500 cm^{-1} for **1a** and 1800−1100 cm^{-1} for **2a**, using the experimental spectrum as the reference. The frequencies were scaled with a factor of 0.975 for B3LYP-D3 on **2a** and 0.980 for B3LYP on **2a** as well as for all calculations on **1a**.

(Table 2). Indeed, overlap estimates between calculated and experimental spectra are the same (0.86 for **1a** and 0.93 for **2a**) with and without dispersion corrections (Table 2). On the other hand, VCD is a much more sensitive technique. The most striking difference is the peak around 1630 cm^{-1} for **2a**, which is only reproduced by the spectrum without dispersion corrections (Figure 5). Accordingly, the overlap estimate for the calculated VCD spectrum is higher for B3LYP (0.70) than for B3LYP-D3 (0.46, Table 2). The effect of adding dispersion corrections is, however, only minor for **1a**, both qualitatively (Figure 5, left) and quantitatively (Table 2), with overlap estimates 0.68 for B3LYP and 0.63 for B3LYP-D3.

Adding dispersion corrections mainly impacts the relative ordering of the conformers. With dispersion corrections, conformers exhibiting dispersion interactions between two side groups are present: the methyl groups of both alanine residues in **1a** (Figure 6, left) and the Boc and OMe group in **2a** (Figure 6, right). For **1a**, this conformer has a low Boltzmann weight and the spectra with and without dispersion corrections are therefore similar and in agreement with experiment. On the other hand, for **2a** this conformer is the dominating conformer when B3LYP-D3 is used and hence the VCD spectrum for B3LYP-D3 differs from the one without dispersion corrections, leading to a VCD spectrum in poorer agreement with the experimental spectrum.

These results are in agreement with the results of Hopmann et al., who concluded that the Boltzmann distribution based on geometries optimized with dispersion corrections changed the VCD spectrum of a highly flexible natural compound significantly, resulting in poor agreement with experiment.[104] In Hopmann's work, enthalpies and free energies result in similar spectra and only spectra with enthalpies were shown. For molecule **2a**, using enthalpies and free energies results in similar spectra while for **1a**, including dispersion corrections and averaging over enthalpies results in a VCD spectrum in poorer agreement with experiment. Koenis et al. also observed that many key VCD bands of a rotaxane had opposite sign compared to the experimental data when using dispersion corrections. Without dispersion corrections, a different conformer of rotaxane dominated and gave a VCD spectrum in good agreement with experiment.[105] Also Merten and co-workers[106−109] and Zehnacker and co-workers[110,111] have shown in several works that including dispersion correction results in significant shifts in the conformational preferences, and that the experimental data are in better agreement with spectra calculated without dispersion corrections. On the basis of these previous observations and our results, it appears that VCD is a sensitive probe of molecular conformation and could hence be used to refine the description of dispersion interactions in quantum-chemical calculations.

Selection of Significant Conformers. To explore the dependence of the calculated VCD spectra on the number of conformers included in the spectral simulations, we consider molecules **2a** and **2c** as these display the largest number of low-energy conformers of the investigated molecules, see Table 3. The results are shown in Figure 7 and Table 4. The number

Table 3. Number of Conformers[a]

	conformational sampling	energy ranking		spectra generation	
		single point energy	geometry optimization	BW > 5%	
	CREST	ΔE < 2.5 kcal/mol	unique conformers	ΔH	ΔG
1a	7	3	1	1	1
1b	73[b]	3	2	2	2
1c	56	5	5	3	2
2a	331	55	22	10	7
2b	669	19	12	5	4
2c	778	25	15	9	7
3	725	20	20	4	3

[a]Number of conformers found by CREST, the number of conformers found within 2.5 kcal/mol of the lowest lying conformer after the single point (SP) calculations, the number of unique conformers after DFT geometry optimization and the number of conformers included in the final spectra with a Boltzmann weight (BW) higher than 5%. [b]Manually set simulation time, see Table 1.

of conformers is based on the difference in energy from the most stable conformer after single-point energy calculations, ΔE. Energy gaps of 2.0, 2.5, 3.0, and 4.0 kcal/mol result in qualitatively similar spectra with only small differences in VCD intensities for both molecules (Figure 7). Energy gaps of 2.5 and 3.0 kcal/mol result in an identical number of significant conformers and thus identical VCD spectra. High overlap estimates ($S \geq 0.99$, Table 4) confirm the high similarity of the spectra.

The number of conformers included in the geometry optimization (Table 4) and thus the main computational cost, depends strongly on the energy threshold. Increasing the energy gap from 2.0 to 2.5 kcal/mol results in one more relevant conformer for both molecules and an overlap estimate of S = 1.00 and S = 0.99 for **2a** and **2c**, respectively, using the results obtained with an energy threshold of 4.0 kcal/mol as the reference. Increasing the threshold further to 3.0 and 4.0 kcal/mol does not increase the number of relevant conformers, whereas the number of conformers in the geometry optimization increases from 55 to 68 and 97 and from 25 to 33 and 66 for **2a** and **2c**, respectively. An energy gap of 2.5 kcal/mol is also used by Grimme and co-workers in their refining approaches when selecting which conformers to optimize at the DFT level.[80]

Conformational Analysis. The conformers found by CREST and determined unique and relevant after DFT calculations are in agreement with those found with classical approaches by Vass et al. (**1a−c**),[77] Merten et al. (**2a−c**)[78] and Bour et al. (**3**).[79] **1a**, **1b**, and **1c** adopts for both chiral elements an inverse γ-turn which is in agreement with the findings of Vass et al. For molecule **2a**, we found that the dominating conformers adopts a β_{II} turn structure with small contributions from structures adopting β_I turns and classical γ turns, whereas Merten et al. assumes that there are significant contributions

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

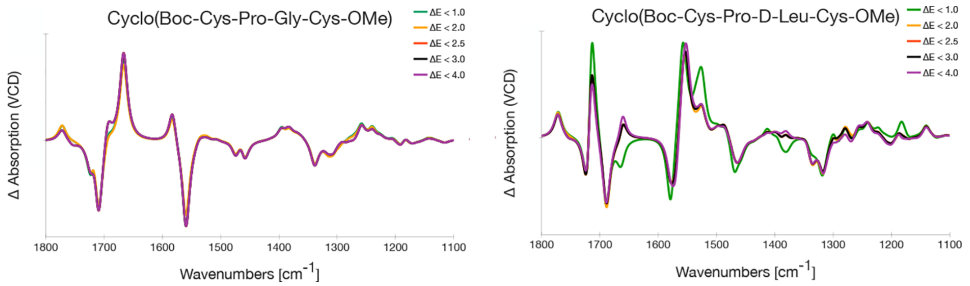


Figure 7. Calculated VCD spectra of **2a** (left) and **2c** (right) in ACN-d_3 based on a different number of conformers. The spectra are based on the conformers within a given energy gap after the single point energy calculations, ΔE in kcal/mol, indicated in the figure. The frequencies are not scaled.

Table 4. Number of Conformers of 2a and 2c that Are Geometry Optimized Based on the Energy Gap after the Single Point (SP) Energy Calculation with DFT of the 331 (2a) and 778 (2c) Conformers Found with CREST[a]

	energy gap (SP) ΔE in kcal/mol	conformers geom. opt.	unique conformers	conformers BW > 5%	lowest BW	S
2a	1.0	9	5	5	5.3%	0.99
	2.0	32	14	6	5.0%	0.99
	2.5	55	22	7	5.0%	1.00
	3.0	68	26	7	5.0%	1.00
	4.0	97	37	7	5.0%	1.00
2c	1.0	4	3	3	10.3%	0.87
	2.0	13	8	6	5.8%	0.99
	2.5	25	15	7	5.4%	0.99
	3.0	33	16	7	5.4%	0.99
	4.0	66	35	7	5.7%	1.00

[a]Only conformers with a Boltzmann weight (BW) higher than 5% are included in the spectra. The overlap estimates S are calculated with ΔE < 4.00 kcal/mol as the reference and over the frequency range shown in Figure 7: 1800−1100 cm^{-1}. The frequencies are not scaled.

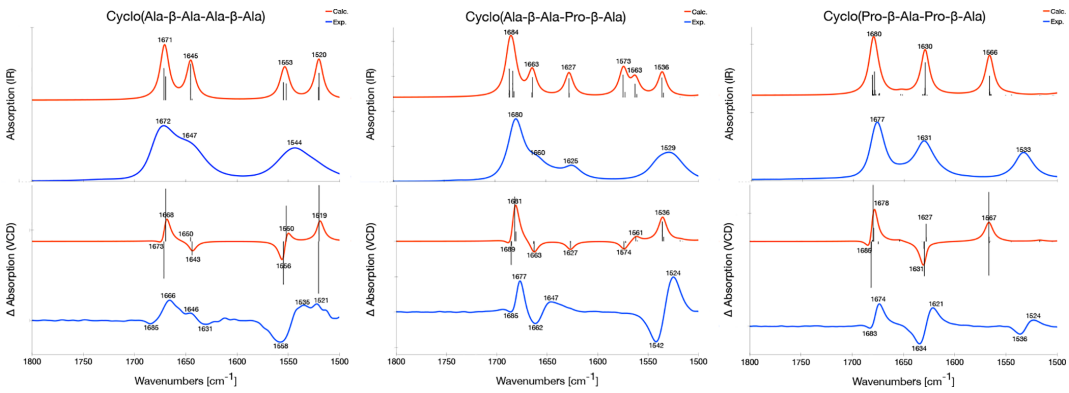


Figure 8. Experimental (blue) and calculated (B3LYP/6-31+G*/CPCM, red) IR (top) and VCD (bottom) spectra of **1a** (left), **1b** (middle), and **1c** (right). For the calculated spectra, a Lorentzian broadening with a full width at half-maximum of 10 cm^{-1} is used and the frequencies for **1a** are scaled with a factor of 0.980 and those for **1b** and **1c** with a factor of 0.990. The experimental spectra are measured by Vass et al.: **1a** in TFE-d_2 while **1b** and **1c** in ACN-d_3.[77] The sticks in the calculated VCD spectra are the rotational strengths associated with that frequency.

from both β_I and β_{II} type structures. For **2b**, we find all relevant conformers to adopt a β_I turn structure while **2c** adopts a β_{II} turn structure. These two findings are in agreement with those of Merten et al. The dominating conformers of **3** has two internal hydrogen bonds in the backbone and adopts two β_{II} turns. Also Bour et al. finds that β_{II} types structures make a major contribution to the conformer mix. The computational protocol used in this work confirms findings in the previous studies. Thus, the protocol is able to determine the important conformers of the studied compounds.

Comparison of Calculated and Experimental Spectra. We will now proceed to predict IR and VCD spectra for our selected molecules and compare to experimental spectra. For

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

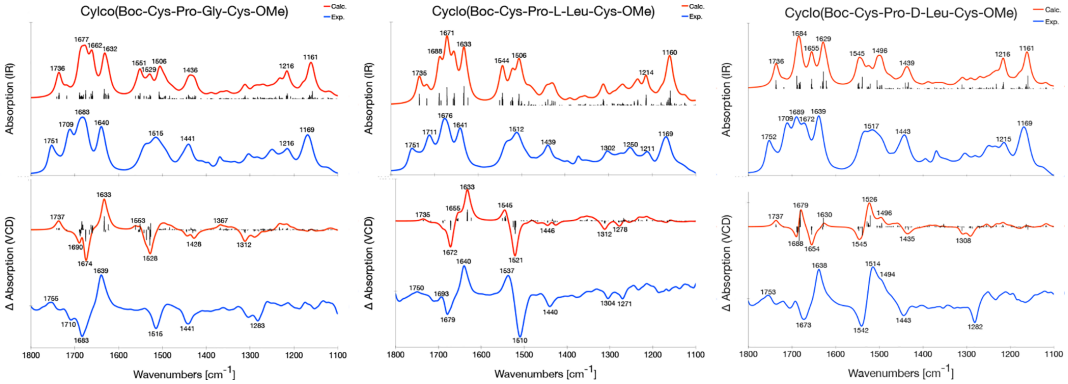


Figure 9. Experimental (blue) and calculated (B3LYP/6-31+G*/CPCM, red) IR (top) and VCD (bottom) spectra of **2a** (left), **2b** (middle) and **2c** (right). For the calculated spectra, a Lorentzian broadening with a full width at half-maximum of 16 cm^{-1} is used, and the frequencies are scaled with a factor of 0.980. The experimental spectra are measured by Merten et al. in ACN-d_3.[78] The sticks in the calculated VCD spectra are the rotational strengths associated with that frequency.

this, we use the optimized protocol: B3LYP/6-31+G*/CPCM, no dispersion corrections, and a threshold of 2.5 kcal/mol to select the most important conformers after the DFT single-point calculations. The experimental and calculated IR and VCD spectra of **1a−c**, **2a−c**, and **3** are shown in Figures 8, 9 and 10, respectively, and the overlap estimates S are given in Table 5.

In peptides, there are three important infrared regions. The amide I region from 1800 to 1600 cm^{-1} is characterized by C=O stretching modes, the amide II region from 1600 to 1480 cm^{-1} is characterized by N−H bending modes, and the amide III region from 1350 to 1250 cm^{-1} is characterized by more complex backbone vibrations in combination with side chain vibrations. For **1a−c**[77] and **3**,[79] only the amide I and II regions have been measured experimentally (1800−1500 cm^{-1}) whereas all three regions have been measured (1800−1100 cm^{-1}) for **2a−c**.[78]

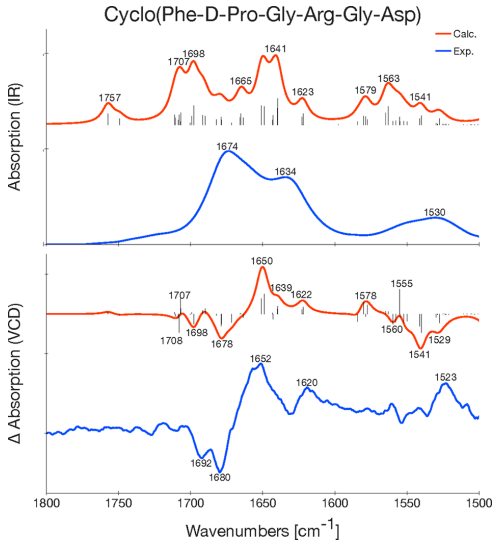


Figure 10. Experimental (blue) and calculated (B3LYP/6-31+G*/CPCM, red) IR (top) and VCD (bottom) spectra of **3**. For the calculated spectrum, a Lorentzian broadening with a full width at half-maximum of 10 cm^{-1} is used and the calculated frequencies are scaled with a factor of 0.995. The experimental spectrum is measured by Bour et al. in TFE.[79] The sticks in the calculated VCD spectra are the rotational strengths associated with that frequency.

Table 5. Overlap Estimates S between the Calculated (B3LYP/6-31+G*/CPCM) and Experimental IR and VCD Spectra and the Scaling Factor f Used for the Spectrum[a]

	S (IR)	S (VCD)	f	freq. range (cm^{-1})
1a	0.86	0.68	0.980	1800−1500
1b	0.83	0.20	0.990	
1c	0.78	0.43	0.990	
2a	0.93	0.70	0.980	1800−1100
2b	0.94	0.61	0.980	
2c	0.92	0.44	0.980	
3	0.80	0.31	0.995	1800−1500

[a]The frequency range used for S are the same as shown in the spectra (Figures 8, 9, and 10).

Molecule Class 1: Cyclo(X-β-Ala-Y-β-Ala). The shape of the experimental IR spectra of **1a−c** are well reproduced by the calculated spectra (Figure 8, top). Considering the VCD spectra of molecule **1a**, both the calculated shape and frequencies of the spectrum are in good agreement with experiment (Figure 8, bottom left panel). The VCD spectrum has a −/+/+/− sign pattern (starting from higher frequencies) in the amide I region which is reproduced well by the calculated spectra. The weak shoulder at 1646 cm^{-1} found in the experiment is in the calculated spectrum only seen when considering the individual rotational strengths (sticks in Figure 8) of the mode at 1650 cm^{-1}. The amide II region in the experimental spectrum is rather noisy due to poor transmission of the solvent,[77] but a −/+ pattern can be seen, which is reproduced in the calculated spectrum. The assignment of the different bands in the IR and VCD spectra of **1a** can be found in Table 6, together with those of **1b** and **1c**.

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

Table 6. Assigned Frequencies (in cm^{-1}) of 1a−c from Figure 8[a]

	1a		1b		1c	
	exp	calc	exp	calc	exp	calc
Amide I: C═O stretch						
freq 1	1685 (−)	1673 (−)	1685 (−)	1689 (−)	1683 (−)	1685 (−)
group	Ala, Ala (s)		Ala		Pro, Pro (s)	
freq 2	1666 (+)	1668 (+)	1677 (+)	1681 (+)	1674 (+)	1678 (+)
group	Ala, Ala (as)		Pro		Pro, Pro (as)	
freq 3	1646 (+)	1650 (+)	1662 (−)	1663 (−)	1634 (−)	1631 (−)
group	β-Ala, β-Ala (s)		β-Ala4		β-Ala, β-Ala (s)	
freq 4	1631 (−)	1643 (−)	1647 (+)[b]	1627 (−)	1621 (+)	1627 (+)
group	β-Ala, β-Ala (as)		β-Ala2		β-Ala, β-Ala (as)	
Amide II: N−H bend						
freq 5	1558 (−)	1556 (−)	1542 (−)	1574 (−)	1536 (−)	1566 (−)
group	β-Ala, β-Ala		β-Ala4		β-Ala, β-Ala	
freq 6	1532 (+)	1550 (+)	1524 (+)	1561 (+)	1524 (+)	1566 (+)
group	β-Ala, β-Ala		β-Ala2		β-Ala, β-Ala	
freq 7	–	1520 (−)	–	1536 (+)		
group	Ala, Ala		Ala		–	
freq 8	1521 (+)	1519 (+)				
group	Ala, Ala		–		–	

[a]All vibrations are assigned to the dominating functional group. Coupled vibrations are indicated as symmetric (s) or asymmetric (as). The sign of the VCD intensities is indicated with (+) or (−). The calculated frequencies for **1a** are scaled with a factor of 0.980 and those for **1b** and **1c** with a factor of 0.990 as in Figure 8. [b]Overlap of VCD bands of opposite sign gives uncertain position.[77]

Table 7. Assigned Frequencies (in cm^{-1}) of 2a−c from Figure 9[a]

	2a		2b		2c	
	exp	calc	exp	calc	exp	calc
Amide I: C═O stetch						
freq 1	1755 (+)	1737 (+)	1750 (+)	1735 (+)	1755 (+)	1737 (+)
group	Cys4		Cys4		Cys4	
freq 2	1710 (−)	1690 (−)	1711 (−)	1689 (−)	–	1688 (−)
group	Boc		Boc		Boc	
freq 3	1683 (−)	1674 (−)	1679 (−)	1672 (−)	–	1679 (+)
group	Pro		Pro		Pro	
freq 4	–	1660 (+)	–	1655 (+)	1673 (−)	1654 (−)
group	Gly		L-Leu		D-Leu	
freq 5	1639 (+)	1633 (+)	1640 (+)	1633 (+)	1638 (+)	1630 (+)
group	Cys1		Cys1		Cys1	
Amide II: N−H bend						
freq 6	1541 (+)	1553 (+)	1537 (+)	1545 (+)	1542 (−)	1545 (−)
group	Cys4		Cys4		Cys4	
freq 7	1515 (−)	1528 (−)	1510 (−)	1521 (−)	1513 (+)	1526 (+)
group	Gly		L-Leu		D-Leu	
freq 8	1487 (+)	1498 (+)	–	1483 (+)	1494 (+)[b]	1496 (+)
group	Cys1		Cys1		Cys1	

[a]All vibrations are assigned to the dominating functional group. The sign of the VCD intensities is indicated with (+) or (−). The calculated frequencies are scaled with a factor of 0.980 as in Figure 9. [b]Shoulder.

In consideration of **1b**, the signs of the peaks in the VCD experiment are partly reproduced by the calculated spectra. The amide I region in the experimental spectrum has a −/+/−/+ pattern, which is only partly reproduced by the calculated spectrum which displays a −/+/−/− pattern. The experimental IR spectrum has a band at 1625 cm^{-1} which corresponds to the band at 1627 cm^{-1} in the calculated spectrum. In the calculated VCD spectrum, this band is negative, whereas only a broad, positive band at higher frequencies is observed in experiment. In the amide II region, the experimental VCD spectrum only displays two intense bands with a −/+ pattern. In contrast, the calculated spectrum has three separate bands displaying a −/+/+ pattern and with lower intensity.

Turning to **1c**, the experiment is reproduced by the calculations when the individual rotational strengths are considered. The amide I region of the experimental VCD spectrum has a −/+/−/+ pattern which is nicely reproduced by the calculated spectrum. Because the frequencies have been calculated to have too small separation, the band in the calculated spectrum at 1631 cm^{-1} results from two overlapping modes that add up to a single negative band. However, considering the two modes separately, as shown by the sticks at 1631 and 1627 cm^{-1} (Figure 8, lower row panels), the

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

experimental pattern is reproduced by the calculations. In the amide II region of the experimental spectrum there is a −/+ pattern. In the calculated spectrum there is a positive band (1567 cm^{-1}) which is in line with the observation from the IR spectra. Again, the frequencies of the two bands have been calculated to have too small separation and the corresponding band in the calculated spectrum at 1567 cm^{-1} consists of two overlapping rotational strengths with different signs, resulting in one positive band in the calculated spectrum (Figure 8, sticks lower right panels).

Although the overall shape of the calculated VCD spectra for **1b** and **1c** is in good agreement with experiment, the overlap estimates are low. Indeed, the overlap estimates for **1b** and **1c** are 0.20 and 0.43, respectively, which is lower than that for **1a** (S = 0.68, Table 5). The overlap estimates are very sensitive to small changes in frequencies. For **1b** and **1c**, the calculated gaps between the amide I and amide II regions are much smaller than in the experiments, off by 42 and 34 cm^{-1}, respectively. The underestimation of this gap is however only 10 cm^{-1} for **1a**. This results in a low S value, particularly for **1b** and **1c**. Using a larger basis set such as 6-311++G** gives amide I−II gaps in slightly better agreement with experiment (top row Figure S1 and Figure S3). In addition, earlier work has suggested that including explicit solute−solvent interactions results in an amide I−II gap in better agreement with experiment.[79,112]

For **1b** and **1c**, we also note that some of the calculated modes are too close in frequency, which merges these vibrational modes into a single, broader band shape, and the VCD stick spectrum is needed to allow for a more detailed comparison with experiment (Figure 8, bottom middle and bottom right panels).

Molecule Class 2: Cyclo(Boc-Cys-Pro-X-Cys-OMe). The experimental IR and VCD spectra of **2a**−**c** are reproduced by the calculated spectra (Figure 9). Starting with **2a**, the amide I region of the experimental VCD spectrum (Figure 9, bottom left) has a +/−/+ sign pattern, which is reproduced by the calculations. In the amide II region, the experiment has a negative band at 1515 cm^{-1}, which corresponds to the calculated peak at 1528 cm^{-1}. Also, at lower frequencies the calculations are in good agreement with experiment. The assignment of the different bands in the IR and VCD spectra of **2a** can be found in Table 7, together with those of **2b** and **2c**.

For **2b**, both the experimental and the calculated VCD spectrum have a −/+ pattern in the amide I region and a +/− pattern in the amide II region (Figure 9, bottom middle). Also, at lower frequencies the calculated spectrum is in agreement with the experimental spectrum.

For **2c**, the sign of the bands in the experimental VCD spectrum are reproduced by the calculations when the individual sticks are considered (Figure 9, bottom right). In the experimental VCD spectrum, there are two main peaks in the amide I region: negative at 1673 cm^{-1} and positive at 1638 cm^{-1}. The negative peak is reproduced in the calculated spectrum at 1654 cm^{-1} whereas the positive peak appears as a stick at 1630 cm^{-1}. In the calculated spectrum, there is a strong positive band at 1679 cm^{-1} that is a much weaker shoulder at 1692 cm^{-1} in the experiment. The −/+ sign pattern of the amide II region is well reproduced in the calculated spectrum.

The overlap estimates of the VCD spectra for **2a**, **2b**, and **2c** are 0.70, 0.61, and 0.44, respectively. As for **1a**−**c**, the gap between the amide I and amide II region is underestimated by approximately 20 cm^{-1} with respect to the experimental spectrum. Although the shape of the amide I region for **2c** is in agreement with experiment, the intensities are not and this, combined with the too small amide I−II gap, leads to the significantly lower overlap estimate for **2c**.

Molecule 3: Cyclo(Phe-D-Pro-Gly-Arg-Gly-Asp). Molecule **3** (Figure 3) is a hexapeptide with four chiral centers and more conformational flexibility in the side chains compared to molecules **1**−**2**. The experimental IR and VCD spectra of **3** are reproduced quite well by the calculations, though there are only a few distinct bands in the experimental IR spectrum, making a detailed comparison difficult. The amide I region of the of the experimental VCD spectrum (Figure 10) has a −/−/+/+ pattern, which is is also found in the calculated spectrum. The negative peaks can be assigned to a C═O stretch on Arg and Asp, respectively, whereas the positive peaks can be assigned to a C═O stretch on Phe (strong positive peak at 1652 cm^{-1}) and different stretches and bends in the side groups of Phe and Arg (positive peak at 1620 cm^{-1}), see Table 8.

Table 8. Assigned Frequencies (in cm^{-1}) of 3 from Figure 10[a]

	3	
	exp	calc
Amide I: C═O stretch		
freq 1	–	1711 (−)
group		Pro
freq 2	–	1708 (−), 1707 (+)[b,c]
group		Gly5
freq 3	1692 (−)	1698 (−)
group		Arg
freq 4	1680 (−)	1678 (−)
group		Asp
freq 5	1652 (+)	1650 (+)
group		Phe
freq 6	–	1639 (+)[c,d]
group		Gly3
freq 7	1620 (+)	1622 (+)
group		side chains on Phe and Arg

[a]All vibrations are assigned to the dominating functional group. The sign of the VCD intensities is indicated with (+) or (−). The calculated frequencies are scaled with a factor of 0.995 as in Figure 10. [b]Two modes in combination with N−H bend on side group with different sign of the VCD intensity. [c]Different conformers give opposite sign for the VCD intensity. [d]Shoulder.

In the amide II region, there is one strong band at 1523 cm^{-1} in the experimental VCD spectrum. This peak is reproduced by a strong absorption at 1555 cm^{-1} in the calculated VCD stick spectrum. The frequency shift of this peak is one reason for the rather low overlap between calculated and experimental spectra both for IR (S = 0.80) and VCD (S = 0.31), see Table 5.

SUMMARY AND OUTLOOK

We have investigated the performance of the conformational search tool CREST developed by Grimme and co-workers,[69] originally developed for quantum chemical calculation of NMR spectra, for the calculation of IR and VCD spectra of seven conformationally flexible cyclic oligopeptides for which experimental spectra are available. Chiroptical properties of flexible molecules are particularly challenging, as different

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

conformers may have different signs of the rotatory strengths, in contrast to, for instance, IR or NMR intensities. Ensuring that all relevant conformers are identified and their relative stability is correctly predicted is therefore paramount for a robust computational protocol to reproduce experimental VCD spectra.

Our results show that CREST identifies the important conformers. Subsequent DFT energy calculations ensure sufficiently accurate results for the relative energies of the different conformers and lead to a reduction of the number of conformers for which a full geometry optimization and VCD properties calculations need to be performed. In weighing the overall contribution of different conformers to generate full IR and VCD spectra, Gibb's free energies are used. We note, however, that in many cases the enthalpy will suffice if the different conformers do not display significantly different entropic contributions. Somewhat surprisingly, but confirming earlier VCD studies,[104,105] including dispersion corrections leads to poorer agreement with experimental VCD spectra. This suggests that the sensitivity of VCD may make it an appropriate technique to use when attempting to further improve the description of dispersion corrections in DFT calculations.

The overall VCD patterns computed based on the conformers identified by the approach are in good agreement with available experimental data. However, the gap between the amide I and amide II region does not match experiment for the molecules investigated in this work. Whereas we have used continuum solvation models, explicit solvation may be required in order to correctly model the energy separation between the amide I and amide II regions.[79] Furthermore, some vibrational bands are too close in frequency in our calculations, leading to overlapping vibrational bands in the calculated VCD spectra that may hide finer patterns observed in the experimental spectra. However, these patterns can be identified by considering the computed spectrum without broadening applied (the stick spectrum). Anharmonic corrections may also also be relevant to consider in some cases.[113,114]

Despite these shortcomings, the tested protocol is able to identify the relevant conformers and the calculated VCD spectra are overall in agreement with experiments. The approach provides a feasible basis for generating VCD spectra also for larger cyclic peptides of biological/pharmaceutical interest. Whereas our protocol has been able to reproduce experimental VCD spectra of these cyclic peptides with known stereochemistry, an interesting question is to what extent experimental VCD in combination with the current computational protocol will be able to distinguish structurally similar cyclic peptides that have different stereochemistry. This will be the focus of future work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c02953.

Basis set test for cyclic oligopeptides; test of different scaling factors; comparison of using enthalpies and free energies in the Boltzmann averaged spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Karolina Di Remigio Eikås − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0003-3035-2018; Email: karolina.s.eikas@uit.no

Kenneth Ruud − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway; Norwegian Defence Research Establishment, 2027 Kjeller, Norway;* orcid.org/0000-0003-1006-8482; Email: kenneth.ruud@uit.no

Author

Maarten T. P. Beerepoot − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0003-3976-9223

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c02953

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has received support from the Research Council of Norway through a Centre of Excellence Grant (Grant 262695) and a research project grant (Grant 269425). The calculations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway (Grant NN4654K). The authors are grateful to C. Merten, E. Vass, P. Bour, T. Keiderling, and S. Grimme for helpful discussion and for kindly sharing experimental data.

■ REFERENCES

(1) Smith, S. W. Chiral Toxicology: It's the Same Thing···Only Different. *Toxicol. Sci.* **2009**, *110*, 4−30.

(2) Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press: Cambridge, 2004.

(3) Wagnière, G. H. *Linear and Nonlinear Optical Properties of Molecules*; Helvetica Chimica Acta, VCH: New York, 1993.

(4) Rizzo, A.; Coriani, S.; Ruud, K. *Computational Strategies for Spectroscopy*; John Wiley & Sons, Ltd, 2011; pp 77−135, Section: 2_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118008720.ch2.

(5) Cotton, A. Absorption inégale des rayons circulaires droit et gauche dans certain corps actifs. *C. R. H. Acad. Sci.* **1895**, *120*, 989−991.

(6) Cotton, A. Dispersion rotatoire anomale des corps absorbants. *C. R. H. Acad. Sci.* **1895**, *120*, 1044−1046.

(7) Woody, R. W. In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons: Hoboken, NJ, 2012; Vol. 2; pp 475−498.

(8) Toniolo, C.; Formaggio, F.; Woody, R. W. In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons: Hoboken, NJ, 2012; Vol. 2; pp 499−544.

(9) Kypr, J.; Kejnovská, I.; Bednářová, K.; Vorlíčková, In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons: Hoboken, NJ, 2012; Vol. 2; pp 575−586.

(10) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **2008**, *128*, 044118.
(11) Peach, M. J. G.; Sueur, C. R. L.; Ruud, K.; Guillaume, M.; Tozer, D. J. TDDFT diagnostic testing and functional assessment for triazene chromophores. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4465−4470.
(12) Polavarapu, P. L.; Chakraborty, D. K.; Ruud, K. Molecular optical rotation: an evaluation of semiempirical models. *Chem. Phys. Lett.* **2000**, *319*, 595−600.
(13) Ruud, K.; Helgaker, T. Optical rotation studied by density-functional and coupled-cluster methods. *Chem. Phys. Lett.* **2002**, *352*, 533−539.
(14) Ruud, K.; Stephens, P. J.; Devlin, F. J.; Taylor, P. R.; Cheeseman, J. R.; Frisch, M. J. Coupled-cluster calculations of optical rotation. *Chem. Phys. Lett.* **2003**, *373*, 606−614.
(15) Tam, M. C.; Russ, N. J.; Crawford, T. D. Coupled cluster calculations of optical rotatory dispersion of (S)-methyloxirane. *J. Chem. Phys.* **2004**, *121*, 3550−3557.
(16) Crawford, T. D. Ab initio calculation of molecular chiroptical properties. *Theor. Chem. Acc.* **2006**, *115*, 227−245.
(17) Ruud, K. *Comprehensive Chiroptical Spectroscopy*; John Wiley & Sons, Ltd, 2011; pp 699−727, Section: 24_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118120187.ch24.
(18) Autschbach, J. Computing chiroptical properties with first-principles theoretical methods: Background and illustrative examples. *Chirality* **2009**, *21*, E116−E152.
(19) Holzwarth, G.; Hsu, E. C.; Mosher, H. S.; Faulkner, T. R.; Moscowitz, A. Infrared circular dichroism of carbon-hydrogen and carbon-deuterium stretching modes. Observations. *J. Am. Chem. Soc.* **1974**, *96*, 251−252.
(20) Barron, L. D.; Bogaard, M. P.; Buckingham, A. D. Raman scattering of circularly polarized light by optically active molecules. *J. Am. Chem. Soc.* **1973**, *95*, 603−605.
(21) Hug, W.; Kint, S.; Bailey, G. F.; Scherer, J. R. Raman circular intensity differential spectroscopy. Spectra of (−)-α-pinene and (+)-α-phenylethylamine. *J. Am. Chem. Soc.* **1975**, *97*, 5589−5590.
(22) He, Y.; Bo, W.; Dukor, R. K.; Nafie, L. A. Determination of Absolute Configuration of Chiral Molecules Using Vibrational Optical Activity: A Review. *Appl. Spectrosc.* **2011**, *65*, 699−723.
(23) Freedman, T. B.; Cao, X.; Dukor, R. K.; Nafie, L. A. Absolute configuration determination of chiral molecules in the solution state using vibrational circular dichroism. *Chirality* **2003**, *15*, 743−758.
(24) Stephens, P. J. Theory of vibrational circular dichroism. *J. Phys. Chem.* **1985**, *89*, 748−752.
(25) Bak, K. L.; Jørgensen, P.; Helgaker, T.; Ruud, K.; Jensen, H. J. A. Gauge-origin independent multiconfigurational self-consistent-field theory for vibrational circular dichroism. *J. Chem. Phys.* **1993**, *98*, 8873−8887.
(26) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(27) Stephens, P.; Devlin, F.; Cheeseman, J. *VCD spectroscopy for organic chemists*; CRC Press, 2012; pp 1−360.
(28) Nafie, L. A. *Vibrational Optical Activity: Principles and Applications*; Wiley: Chichester, 2011.
(29) Stephens, P. J.; Devlin, F. J.; Pan, J.-J. The determination of the absolute configurations of chiral molecules using vibrational circular dichroism (VCD) spectroscopy. *Chirality* **2008**, *20*, 643−663.
(30) Polavarapu, P. L. Renaissance in chiroptical spectroscopic methods for molecular structure determination. *Chem. Rec.* **2007**, *7*, 125−136.
(31) Batista, J., Jr; Blanch, E. W.; Bolzani, V. d. S. Recent advances in the use of vibrational chiroptical spectroscopic methods for stereochemical characterization of natural products. *Nat. Prod. Rep.* **2015**, *32*, 1280−1302.
(32) Abbate, S.; Burgi, L. F.; Castiglioni, E.; Lebon, F.; Longhi, G.; Toscano, E.; Caccamese, S. Assessment of configurational and conformational properties of naringenin by vibrational circular dichroism. *Chirality* **2009**, *21*, 436−441.
(33) Batista, J. M.; Batista, A. N.; Rinaldo, D.; Vilegas, W.; Cass, Q. B.; Bolzani, V. S.; Kato, M. J.; López, S. N.; Furlan, M.; Nafie, L. A. Absolute configuration reassignment of two chromanes from Peperomia obtusifolia (Piperaceae) using VCD and DFT calculations. *Tetrahedron: Asymmetry* **2010**, *21*, 2402−2407.
(34) Chamayou, A.-C.; Lüdeke, S.; Brecht, V.; Freedman, T. B.; Nafie, L. A.; Janiak, C. Chirality and Diastereoselection of Δ/Λ-Configured Tetrahedral Zinc Complexes through Enantiopure Schiff Base Complexes: Combined Vibrational Circular Dichroism, Density Functional Theory, 1H NMR, and X-ray Structural Studies. *Inorg. Chem.* **2011**, *50*, 11363−11374.
(35) Keiderling, T. A. Structure of Condensed Phase Peptides: Insights from Vibrational Circular Dichroism and Raman Optical Activity Techniques. *Chem. Rev.* **2020**, *120*, 3381−3419.
(36) Sen, A. C.; Keiderling, T. A. Vibrational circular dichroism of polypeptides. II. Solution amide II and deuteration results. *Biopolymers* **1984**, *23*, 1519−1532.
(37) Sen, A. C.; Keiderling, T. A. Vibrational circular dichroism of polypeptides. III. Film studies of several α-helical and β-sheet polypeptides. *Biopolymers* **1984**, *23*, 1533−1545.
(38) Lal, B. B.; Nafie, L. A. Vibrational circular dichroism in amino acids and peptides. 7. Amide stretching vibrations in polypeptides. *Biopolymers* **1982**, *21*, 2161−2183.
(39) Pancoska, P.; Yasui, S. C.; Keiderling, T. A. Enhanced sensitivity to conformation in various proteins. Vibrational circular dichroism results. *Biochemistry* **1989**, *28*, 5917−5923.
(40) Pancoska, P.; Yasui, S. C.; Keiderling, T. A. Statistical analyses of the vibrational circular dichroism of selected proteins and relationship to secondary structures. *Biochemistry* **1991**, *30*, 5089−5103.
(41) Dukor, R. K.; Keiderling, T. A. Reassessment of the random coil conformation: Vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers* **1991**, *31*, 1747−1761.
(42) Gupta, V. P.; Keiderling, T. A. Vibrational CD of the amide II band in some model polypeptides and proteins. *Biopolymers* **1992**, *32*, 239−248.
(43) Polyanichko, A.; Andrushchenko, V.; Bouř, P.; Wieser, H. *Vibrational Circular Dichroism Studies of Biological Macromolecules and their Complexes*; 2012; pp 67−126.
(44) Aviles-Moreno, J. R.; Ureña Horno, E.; Partal Ureña, F.; López González, J. J. IR−Raman−VCD study of R-(+)-Pulegone: Influence of the solvent. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2011**, *79*, 767−776.
(45) Perera, A. S.; Cheramy, J.; Merten, C.; Thomas, J.; Xu, Y. IR, Raman, and Vibrational Optical Activity Spectra of Methyl Glycidate in Chloroform and Water: The Clusters-in-a-liquid Solvation Model. *ChemPhysChem* **2018**, *19*, 2234−2242.
(46) Cappelli, C. Integrated QM/polarizable MM/continuum approaches to model chiroptical properties of strongly interacting solute−solvent systems. *Int. J. Quantum Chem.* **2016**, *116*, 1532−1542.
(47) Giovannini, T.; Olszòwka, M.; Cappelli, C. Effective Fully Polarizable QM/MM Approach To Model Vibrational Circular Dichroism Spectra of Systems in Aqueous Solution. *J. Chem. Theory Comput.* **2016**, *12*, 5483−5492.
(48) Giovannini, T.; Del Frate, G.; Lafiosca, P.; Cappelli, C. Effective computational route towards vibrational optical activity spectra of chiral molecules in aqueous solution. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9181−9197.
(49) Zhang, Y.; Poopari, M. R.; Cai, X.; Savin, A.; Dezhahang, Z.; Cheramy, J.; Xu, Y. IR and Vibrational Circular Dichroism Spectroscopy of Matrine- and Artemisinin-Type Herbal Products: Stereochemical Characterization and Solvent Effects. *J. Nat. Prod.* **2016**, *79*, 1012−1023.
(50) Poopari, M. R.; Dezhahang, Z.; Xu, Y. A comparative VCD study of methyl mandelate in methanol, dimethyl sulfoxide, and

chloroform: explicit and implicit solvation models. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1655−1665.
(51) Poopari, M. R.; Dezhahang, Z.; Yang, G.; Xu, Y. Conformational Distributions of N-Acetyl-L-cysteine in Aqueous Solutions: A Combined Implicit and Explicit Solvation Treatment of VA and VCD Spectra. *ChemPhysChem* **2012**, *13*, 2310−2321.
(52) Krupová, M.; Kessler, J.; Bouř, P. Recent Trends in Chiroptical Spectroscopy: Theory and Applications of Vibrational Circular Dichroism and Raman Optical Activity. *ChemPlusChem.* **2020**, *85*, 561−575.
(53) Chi, H.; Welch, W. R. W.; Kubelka, J.; Keiderling, T. A. Insight into the Packing Pattern of β2 Fibrils: A Model Study of Glutamic Acid Rich Oligomers with 13C Isotopic Edited Vibrational Spectroscopy. *Biomacromolecules* **2013**, *14*, 3880−3891.
(54) Ma, S.; Cao, X.; Mak, M.; Sadik, A.; Walkner, C.; Freedman, T. B.; Lednev, I. K.; Dukor, R. K.; Nafie, L. A. Vibrational Circular Dichroism Shows Unusual Sensitivity to Protein Fibril Formation and Development in Solution. *J. Am. Chem. Soc.* **2007**, *129*, 12364−12365.
(55) Polavarapu, P. L.; Chakraborty, D. K.; Ruud, K. Molecular optical rotation: an evaluation of semiempirical models. *Chem. Phys. Lett.* **2000**, *319*, 595−600.
(56) Wiberg, K. B.; Wang, Y.-g.; Vaccaro, P. H.; Cheeseman, J. R.; Trucks, G.; Frisch, M. J. Optical Activity of 1-Butene, Butane, and Related Hydrocarbons. *J. Phys. Chem. A* **2004**, *108*, 32−38.
(57) Coriani, S.; Forzato, C.; Furlan, G.; Nitti, P.; Pitacco, G.; Ringholm, M.; Ruud, K. Synthesis, characterization and assignment of the absolute configuration of 4,4-dimethyl-5-oxo-tetrahydrofuran-3-carboxylic acid and its esters: a combined experimental and theoretical investigation. *Tetrahedron: Asymmetry* **2009**, *20*, 1459−1467.
(58) Merten, C.; Golub, T. P.; Kreienborg, N. M. Absolute Configurations of Synthetic Molecular Scaffolds from Vibrational CD Spectroscopy. *J. Org. Chem.* **2019**, *84*, 8797−8814.
(59) Hopmann, K. H.; Ruud, K.; Pecul, M.; Kudelski, A.; Dračínský, M.; Bouř, P. Explicit versus Implicit Solvent Modeling of Raman Optical Activity Spectra. *J. Phys. Chem. B* **2011**, *115*, 4128−4137.
(60) Aerts, R.; Vanhove, J.; Herrebout, W.; Johannessen, C. Paving the way to conformationally unravel complex glycopeptide antibiotics by means of Raman optical activity. *Chem. Sci.* **2021**, *12*, 5952−5964.
(61) Polavarapu, P. L. Determination of the Absolute Configurations of Chiral Drugs Using Chiroptical Spectroscopy. *Molecules* **2016**, *21*, 1056.
(62) Magyarfalvi, G.; Tarczay, G.; Vass, E. Vibrational circular dichroism. *Rev.: Comput. Mol. Sci.* 2011, *1*, 403−425.
(63) Jungwirth, J.; Šebestík, J.; Šafařík, M.; Kapitán, J.; Bouř, P. Quantitative Determination of Ala-Ala Conformer Ratios in Solution by Decomposition of Raman Optical Activity Spectra. *J. Phys. Chem. B* **2017**, *121*, 8956−8964.
(64) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.
(65) Miteva, M. A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622−W627.
(66) Tai, K. Conformational sampling for the impatient. *Biophys. Chem.* **2004**, *107*, 213−220.
(67) Dorfman, R. J.; Smith, K. M.; Masek, B. B.; Clark, R. D. A knowledge-based approach to generating diverse but energetically representative ensembles of ligand conformers. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 681−691.
(68) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019**, *59*, 731−742.
(69) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. Fully Automated Quantum-Chemistry-Based Computation of Spin−Spin-Coupled Nuclear Magnetic Resonance Spectra. *Angew. Chem., Int. Ed.* **2017**, *56*, 14763−14769.
(70) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.
(71) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169−7192.
(72) Bohle, F.; Seibert, J.; Grimme, S. Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules. *J. Org. Chem.* **2021**, *86*, 15522−15531.
(73) Spicher, S.; Abdullin, D.; Grimme, S.; Schiemann, O. Modeling of spin-spin distance distributions for nitroxide labeled biomacromolecules. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24282−24290.
(74) Ghidinelli, S.; Abbate, S.; Mazzeo, G.; Boiadjiev, S. E.; Lightner, D. A.; Longhi, G. Biliverdin chiral derivatives as chiroptical switches for pH and metal cation sensing. *hysical Chemistry Chemical Physics* **2021**, *23*, 20138−20151.
(75) Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **2021**, *125*, 4039−4054.
(76) Empting, M. *Cyclic Peptides: From Bioorganic Synthesis to Applications*; The Royal Society of Chemistry, 2018; pp 1−14.
(77) Vass, E.; Majer, Z.; Kőhalmy, K.; Hollósi, M. Vibrational and chiroptical spectroscopic characterization of γ-turn model cyclic tetrapeptides containing two β-Ala residues. *Chirality* **2010**, *22*, 762−771.
(78) Merten, C.; Li, F.; Bravo-Rodriguez, K.; Sanchez-Garcia, E.; Xu, Y.; Sander, W. Solvent-induced conformational changes in cyclic peptides: a vibrational circular dichroism study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5627−5633.
(79) Bouř, P.; Kim, J.; Kapitan, J.; Hammer, R.; Huang, R.; Wu, L.; Keiderling, T. Vibrational circular dichroism and IR spectral analysis as a test of theoretical conformational modeling for a cyclic hexapeptide. *Chirality* **2008**, *20*, 1104−19.
(80) Grimme et al. *Revision e19549b4: xtb doc 6.2 Documentation*. https://xtb-docs.readthedocs.io (accessed: 10.03.2020).
(81) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.
(82) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(83) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. *Gaussian 16*, Revision B.01; Gaussian Inc.: Wallingford CT, 2016.
(84) Stephens, P.; Devlin, F. Determination of the structure of chiral molecules using ab initio vibrational circular dichroism spectroscopy. *Chirality* **2000**, *12*, 172−179.
(85) Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. Ab Initio Prediction of Vibrational Absorption and Circular Dichroism Spectra of Chiral Natural Products Using Density Functional Theory: α-Pinene. *J. Phys. Chem. A* **1997**, *101*, 9912−9924.
(86) Kuppens, T.; Langenaeker, W.; Tollenaere, J. P.; Bultinck, P. Determination of the Stereochemistry of 3-Hydroxymethyl-2,3-dihydro-[1,4]dioxino[2,3-b]- pyridine by Vibrational Circular Dichroism and the Effect of DFT Integration Grids. *J. Phys. Chem. A* **2003**, *107*, 542−553.
(87) Andersson, M. P.; Uvdal, P. New Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-ζ Basis Set 6-311+G(d, p). *J. Phys. Chem. A* **2005**, *109*, 2937−2941.
(88) Becke, A. D. A new mixing of Hartree Fock and local density functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

https://doi.org/10.1021/acs.jpca.2c02953
J. Phys. Chem. A 2022, 126, 5458−5471

(89) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(90) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z = 11−18. *J. Chem. Phys.* **1980**, *72*, 5639−5648.
(91) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(92) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(93) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(94) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.
(95) Eikås, K. D. R.; Ruud, K.; Beerepoot, M. T. P. *Replication Data for: A Computational Protocol for VCD spectra of Cyclic Peptides*; 2022; DOI: 10.18710/VITGV3, DataverseNO.
(96) Liegeois, V. *DrawSpectrum, Version 1.6.100; UNamur*, 2015.
(97) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.
(98) Lamparska, E.; Liégeois, V.; Quinet, O.; Champagne, B. Theoretical Determination of the Vibrational Raman Optical Activity Signatures of Helical Polypropylene Chains. *ChemPhysChem* **2006**, *7*, 2366−2376.
(99) Liégeois, V.; Quinet, O.; Champagne, B. Vibrational Raman optical activity as a mean for revealing the helicity of oligosilanes: A quantum chemical investigation. *J. Chem. Phys.* **2005**, *122*, 214304.
(100) Zuber, G.; Hug, W. Rarefied Basis Sets for the Calculation of Optical Tensors. 1. The Importance of Gradients on Hydrogen Atoms for the Raman Scattering Tensor. *J. Phys. Chem. A* **2004**, *108*, 2108−2118.
(101) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, 034112.
(102) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.
(103) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8499−8512.
(104) Hopmann, K. H.; Šebestík, J.; Novotná, J.; Stensen, W.; Urbanová, M.; Svenson, J.; Svendsen, J. S.; Bouř, P.; Ruud, K. Determining the Absolute Configuration of Two Marine Compounds Using Vibrational Chiroptical Spectroscopy. *J. Org. Chem.* **2012**, *77*, 858−869.
(105) Koenis, M. A. J.; Visser, O.; Visscher, L.; Buma, W. J.; Nicu, V. P. GUI Implementation of VCDtools, A Program to Analyze Computed Vibrational Circular Dichroism Spectra. *J. Chem. Inf. Model.* **2020**, *60*, 259−267.
(106) Scholten, K.; Merten, C. Anion-binding of a chiral tris(2-aminoethyl)amine-based tripodal thiourea: a spectroscopic and computational study. *Phys. Chem. Chem. Phys.* **2022**, *24*, 4042−4050.
(107) Vermeyen, T.; Merten, C. Solvation and the secondary structure of a proline-containing dipeptide: insights from VCD spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 15640−15648.
(108) Bünnemann, K.; Merten, C. Solvation of a chiral carboxylic acid: effects of hydrogen bonding on the IR and VCD spectra of α-methoxyphenylacetic acid. *Phys. Chem. Chem. Phys.* **2017**, *19*, 18948−18956.
(109) Kreienborg, N. M.; Merten, C. How Do Substrates Bind to a Bifunctional Thiourea Catalyst? A Vibrational CD Study on Carboxylic Acid Binding. *Chem. - Eur. J.* **2018**, *24*, 17948−17954.
(110) Le Barbu-Debus, K.; Bowles, J.; Jähnigen, S.; Clavaguéra, C.; Calvo, F.; Vuilleumier, R.; Zehnacker, A. Assessing cluster models of solvation for the description of vibrational circular dichroism spectra: synergy between static and dynamic approaches. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26047−26068.
(111) Le Barbu-Debus, K.; Zehnacker, A. Competition between inter and intramolecular hydrogen bond evidenced by vibrational circular dichroism spectroscopy: The case of (1S,2R)-(−)-cis-1-amino-2-indanol. *Chirality* **2021**, *33*, 858−874.
(112) Kubelka, J.; Huang, R.; Keiderling, T. A. Solvent Effects on IR and VCD Spectra of Helical Peptides: DFT-Based Static Spectral Simulations with Explicit Water. *J. Phys. Chem. B* **2005**, *109*, 8231−8243.
(113) Bloino, J.; Barone, V. A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies. *J. Chem. Phys.* **2012**, *136*, 124108.
(114) Bloino, J.; Biczysko, M.; Barone, V. Anharmonic Effects on Vibrational Spectra Intensities: Infrared, Raman, Vibrational Circular Dichroism, and Raman Optical Activity. *J. Phys. Chem. A* **2015**, *119*, 11862−11874.

Paper II

Can the absolute configuration of cyclic peptides be determined with vibrational circular dichroism?

Karolina Di Remigio Eikås,
Monika Krupová, Tone Kristoffersen,
Maarten T. P. Beerepoot and Kenneth Ruud

Submitted

Can the absolute configuration of cyclic peptides be determined with vibrational circular dichroism?

Karolina Di Remigio Eikås,[*,†] Monika Krupová,[†] Tone Kristoffersen,[‡] Maarten T. P. Beerepoot,[†] and Kenneth Ruud[*,†,¶]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway*

‡*Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway*

¶*Norwegian Defence Research Establishment, P.O. Box 25, 2027 Kjeller, Norway*

E-mail: karolina.s.eikas@uit.no; kenneth.ruud@uit.no

Abstract

Cyclic peptides show a wide range of biological activities, among others as antibacterial agents. These peptides are often large and flexible with multiple chiral centers. The determination of the stereochemistry of molecules with multiple chiral centers is a challenging and important task in drug development. Chiroptical spectroscopies such as vibrational circular dichroism (VCD) can distinguish between different stereoisomers. The absolute configuration (AC) of a stereoisomer can be determined by comparing its experimental spectra to computed spectra of stereoisomers with known AC. In this way, the AC of rigid molecules with up to seven chiral centers has been assigned (Bogaerts *et al.*, *Phys. Chem. Chem. Phys.* **22**, 18014, 2020). The question arises whether this is possible with more conformationally flexible molecules such as

cyclic peptides. We here investigate to what extent the AC of cyclic peptides can be determined with VCD. More specifically, we investigate the maximum number of chiral centers a cyclic peptide can have in order to be able to unambiguously assign the AC with VCD. We present experimental and computed IR and VCD spectra for a series of eight tetrapeptides and hexapeptides with two, three and four chiral centers. We use our recently developed computational protocol with a conformational search based on sampling with meta-dynamics. We use visual inspection to compare the computed spectra of different stereoisomers with an experimental spectrum of the corresponding cyclic peptide with known AC. We find that the AC of the investigated cyclic peptides with two chiral centers can be unambiguously assigned with VCD. This is however not possible for all of the cyclic peptides with three chiral centers and for none of those with four chiral centers. At best, one can limit the number of possible stereoisomers in those cases. Our work shows that other techniques are needed to assign the AC of cyclic peptides with three or more chiral centers. Our study also constitutes a warning that the spectra of all stereoisomers should be computed before attempting to match to an experimental spectrum, to avoid an accidental erroneous match.

Introduction

Cyclic peptides show a wide range of biological activities, as antitumor drugs, immunosuppressors and antimicrobial agents, to name a few.[1–3] They are often large, flexible and can include multiple chiral centers. The determination of the stereochemistry of molecules with multiple chiral centers is a challenging and important task in drug development. Indeed, the physiological function of a medicinal drug depends strongly on its stereochemistry. Two enantiomers of a chiral compound may have different effects on the human organism; one of them can function as a life-saving drug, while the other may have no function or, in the worst case, cause severe damage to the patient.[4] Although enantiomers can have different biological effects, almost all of their physical properties are the same. Hence, common struc-

ture elucidation techniques such as mass spectroscopy, nuclear magnetic resonance (NMR) or infrared spectroscopy (IR) are often not sufficient to determine their absolute configuration (AC), even though a combination of computational methods and advanced statistical methods has shown promise in structure elucidation of small molecules with NMR.[5] One technique that has been frequently used to determine the AC of chiral molecules is X-ray crystallography. This method requires extensive experimental work and may fail in certain cases.[6] Many molecules fail to crystallize, and even if a crystal can be obtained and the AC determined, the crystal structure may deviate from the structure in solution, preventing a clear identification of structure-activity relationships.

Various methods of chiroptical spectroscopy can be used to determine the chirality of organic compounds.[7–9] Vibrational circular dichroism (VCD)[10] has recently been extensively used in the pharmaceutical industry to determine the AC of chiral drugs.[8,11,12] While IR measures the absorption of infrared light, VCD measures the difference in absorption of left- and right-circularly polarized light in the infrared energy range. Since this differential absorption varies for enantiomers, VCD, in contrast to IR, can distinguish between enantiomers and help to determine the AC of chiral compounds, since two enantiomers have equal VCD spectra with opposite sign. The situation gets more complicated, however, for molecules with more than one chiral center where one distinguishes between enantiomers and diastereomers. Enantiomers have different stereochemistry in all chiral centers, are mirror images of each other and have equal VCD spectra with opposite signs. Diastereomers have different stereochemistry in at least one but not all chiral centers and have VCD spectra that are more different than only a sign change.

Unambiguous assignment of the stereochemistry for a range of both synthetic and natural chiral compounds has been done using VCD.[8,13,14] In most cases where VCD alone has been used to determine the stereochemistry, the compounds were limited to a maximum of three chiral elements.[15] However, a recent work by Bogaerts *et al.* demonstrated that VCD could be used to unambiguously assign the AC of artemisinin and artesunate with six and seven chiral

centers, respectively.[16] This was facilitated by the fairly rigid structure of these molecules. The situation becomes more complex and challenging for conformationally flexible systems such as cyclic peptides,[17] where VCD alone may not be successful in an unambiguous AC determination.

To assign the unknown AC of a chiral compound, a measured VCD spectrum is compared with computed spectra for different stereoisomers of the compound. A computational protocol to simulate the spectra typically consists of a conformational search and quantum-chemical calculations of the VCD spectra.[8,17] Since the spectra are measured in solution, the solvent environment is often included in calculations through the use of a continuum model.[18,19] The conformational search is a key step in the computational protocol for flexible molecules, since different conformers with the same AC may provide VCD intensities with different signs in the same frequency range. When relevant conformers are missing, this could easily lead to an erroneous assignment of the AC. In a previous work,[17] we established a reliable computational protocol for the calculation of conformationally flexible molecules with known AC, using several cyclic peptides as test systems. This computational protocol was based on a meta-dynamics conformational search as implemented in CREST,[20–22] combined with energies, geometries and VCD spectra obtained using DFT. We showed that the protocol allows for the identification of relevant conformations, and that the Boltzmann-averaged spectrum agrees well with the corresponding experimental spectrum.

In this work, we investigate to what extent the AC of cyclic peptides can be determined with VCD using this computational protocol. More specifically, we investigate the maximum number of chiral centers a cyclic peptide can have in order to be able to unambiguously assign the AC with VCD.

Methods

In order to investigate to what extent the AC of cyclic peptides can be determined with VCD, we use our previously published computational protocol[17] to compute VCD spectra for all stereoisomers of a given cyclic peptide and compare them to an experimental VCD spectrum of a stereoisomer with known AC.

Investigated chiral peptides

We have selected a series of eight tetrapeptides and hexapeptides with two, three and four chiral centers (Figure 1).

With the exception of β-alanine in **1** and glycine in **2**–**6** and **8**, all amino acids have one chiral center. The AC of a chiral peptide is indicated by naming the stereochemistry of each chiral amino acid, following the order of the amino acids. For example, *SR*-cyclo(Ala-β-Ala-Pro-β-Ala) (**1**) indicates that alanine has the *S* configuration and proline has *R* wheres as *RSR*-cyclo(Val-Gly-Ser-Gly-Ala-Gly) (**5**) indicates that the chiral amino acids are *R*-valine, *S*-serine and *R*-alanine.

The experimental spectra for tetrapeptides **1**,[23] **4**[24] and **7**[24] as well as for the hexapeptide **8**[25] were taken from the literature. Cyclic hexapeptides **2**, **3**, **5** and **6** were synthesized and their VCD spectra measured as part of this study.

Synthesis of cyclic hexapeptides

Hexapeptides **2**, **3**, **5** and **6** were prepared from fluorenylmethoxycarbonyl (Fmoc)-protected amino acids by solid-phase peptide synthesis[26] using the Fmoc strategy[27] as described in detail elsewhere.[28] Using 2-chlorotrityl chloride resin and Fmoc-Gly-OH as the first amino acid, linear peptide precursors were made by an automated peptide synthesizer (Biotage Initiator+ Alstra). The linear peptides were cleaved from the resin, head-to-tail cyclized in solution, and side-chain deprotected. Crude peptides were purified by preparative reversed-

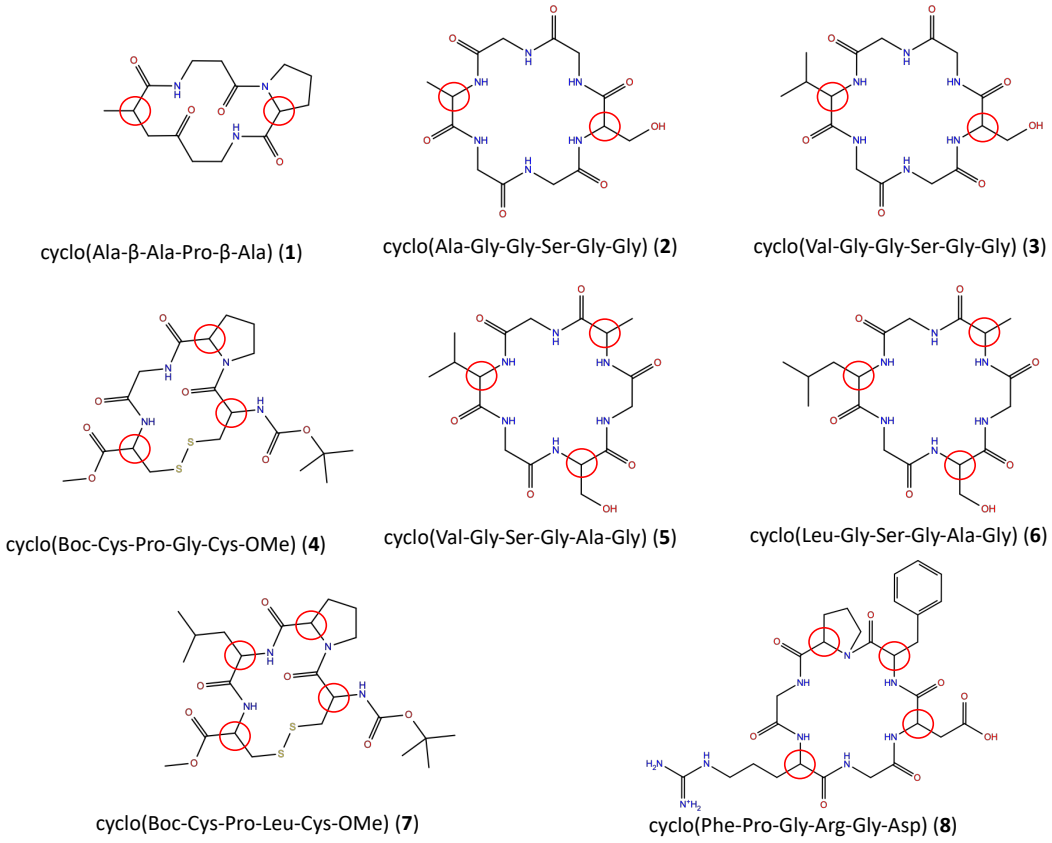


Figure 1: Structure of the investigated chiral peptides. Red circles indicate chiral centers.

phase high-performance liquid chromatography.

Measurement of IR and VCD spectra

IR absorption and VCD spectra of cyclic peptides were measured using an Invenio R FTIR spectrometer equipped with a PMA50 module for polarization measurements (Bruker) at 8 cm^{-1} resolution using a demountable BaF_2 cell and a 100 μm Mylar spacer, with photoelastic modulator module set to 1600 cm^{-1}. The peptides were dissolved in DMSO-d_6 at a concentration of 25 mg/mL, and 50 μL of this peptide solution was deposited on the bottom

cell window and covered with the top window. Six blocks of 22500 scans (16 hours of total accumulation time) were collected and averaged. Spectra of DMSO-d_6 measured at the same conditions were subtracted from the sample spectra.

Computational Details

IR and VCD spectra were calculated using our recently developed computational protocol.[17] A conformational search was performed with meta-dynamics as implemented in CREST 2.10.[21,22] The conformer ensemble was then re-ranked by a DFT single-point energy calculation (B3LYP[29,30]/6-31+G*[31,32]/CPCM[33,34]). Conformers within 2.5 kcal/mol of the lowest-lying conformer after the DFT single-point energy calculation were geometry optimized, and the resulting structures were used to calculate IR and VCD intensities. If less than 20 conformers were within the 2.5 kcal/mol range after the single-point energy calculation, we included the lowest-energy conformations above 2.5 kcal/mol up to a total of 20 conformers to ensure that the conformational space was sufficiently sampled. The single-point energy calculations, geometry optimizations and VCD calculations were performed at the DFT level (B3LYP/6-31+G*/CPCM) using Gaussian 16 (Rev. B.01).[35] For more details on the computational protocol, we refer to our previous work.[17]

Final VCD spectra were generated as Boltzmann averages based on free energies for the unique conformers by the DrawSpectrum program.[36] The calculated spectral intensities are scaled such that the peak with the highest intensity in the computed spectrum matches the peak with the highest intensity in the experimental spectrum. For the molecules measured in this work (**2**, **3**, **5** and **6**), a Lorentzian bandshape with a full width at half maximum (FWHM) of 10 cm^{-1} was applied. For the molecules measured elsewhere, the broadening was chosen to be the same as used for the calculated spectra in those works: 10 cm^{-1} for **1** measured by Vass *et al.*,[23] 16 cm^{-1} for **4** and **7** measured by Merten *et al.*[24] and 10 cm^{-1} for **8** measured by Bouř *et al.*[25]

Since the stereoisomers occur in pairs of enantiomers that have equal IR spectra and

equal but opposite-signed VCD spectra, we have reduced the computational workload by calculating the spectra for only one enantiomer in each pair. Thus, we applied the computational protocol including the conformational search for two diastereomers of the cyclic peptides with two chiral centers; for four diastereomers of the cyclic peptides with three chiral centers; and for eight diastereomers of the cyclic peptides with four chiral centers.

Comparison of the experimental and computed spectra

The most important vibrations for peptides are located in the amide I spectral region (1800-1600 cm^{-1}), associated mostly with C=O stretching modes, and the amide II spectral region (1600-1480 cm^{-1}), associated mostly with N-H bending modes. We compare experimental and computed VCD spectra for the 1800-1400 cm^{-1} region with the exception of **8**, where experimental data is available for the 1800-1500 cm^{-1} region only.

As discussed in our previous work,[17] the experimental gap between the amide I and II spectral regions is poorly reproduced in computed spectra of cyclic peptides. In general, frequencies calculated with DFT do not match the experimentally observed frequencies[37–39] and they are usually scaled to obtain a better agreement with experiment. To our knowledge, there is no established method to determine the frequency scaling factors for spectra recorded in solvent and several *ad-hoc* schemes are in use. One way is to choose the scaling factor that gives the best visual agreement with experiment. In this work, however, we compute the overlap between the experimental spectrum and spectra computed with different frequency scaling factors and choose the frequency scaling factor giving the largest overlap between experimental and computed spectra. Overlap estimates S are calculated with DrawSpectrum[36] as[40–43]

$$S = \frac{\langle I_{calc}|I_{exp}\rangle}{\sqrt{\langle I_{calc}|I_{calc}\rangle\langle I_{exp}|I_{exp}\rangle}} \tag{1}$$

where I is the spectral intensity at a given wavenumber. For IR, S can range from 0 to 1,

whereas for VCD it can range from -1 to 1. For VCD, an overlap estimate of 1 indicates identical spectra and -1 indicates mirror-image spectra.

We use overlap estimates from a combination of the IR and VCD spectra to determine the frequency scaling factor separately for each stereoisomer of each chiral peptide as

$$S = \sqrt{|S(VCD)| \times |S(IR)|}. \tag{2}$$

Overlap estimates for frequency scaling factors from 0.960 to 1.000 with increments of 0.005 were calculated and evaluated. Given the poor agreement in the experimental and computed gaps between the amide I and amide II regions, the same frequency scaling factor cannot be used for both spectral regions to calculate the overlap estimate of both regions combined. Hence, we have chosen to compute overlap estimates for the intense amide I region (1800-1600 cm^{-1}) only. The resulting frequency scaling factors are collected in Tables S1, S2 and S3.

In addition to their use in establishing the frequency scaling factors, overlap estimates are also used in this work to select the stereoisomers to show in the figures. For each enantiomeric pair, we show the enantiomer that is in best agreement with the experiment by selecting the one with the largest positive overlap estimate in the amide I region. For the cyclic peptides with four chiral centers, we show the four diastereomers that agree best with the experimental spectrum by selecting the ones with the largest positive overlap estimate.

One could in theory use overlap estimates to determine which stereoisomer matches the experimental VCD spectrum best. Indeed, Bogaerts *et al.*[16] used overlap estimates between calculated and experimental VCD and Raman optical activity (ROA) spectra to determine the AC of two rigid compounds with six and seven chiral centers. For the molecules investigated in this work, we found that the overlap estimates are not reliable enough to be used for this purpose. The overlap estimates calculated using Eq. 1 are extremely sensitive to small changes in frequencies. Indeed, a small change in the frequency scaling factor may

entirely change the AC assignment based on overlap estimates. In addition, the largest overlap estimate may correspond to a stereoisomer that shows an entirely different pattern in the VCD spectrum and which would therefore immediately be discarded by the trained eye. Bogaerts *et al.* obtained the overlap estimates using scaling factors determined by IR, VCD, Raman and ROA spectra together and these overlap estimates are likely more reliable than those calculated using scaling factor obtained from IR and VCD spectra only.

Thus, we use visual inspection to determine which computed spectrum/spectra match the experimental spectrum best. We primarily use the experimental pattern in the intense amide I region to select matching computed spectra. If this is not possible or when several computed spectra match the pattern in the amide I region, we also include the amide II region in this analysis. Another source of information is the relative intensity of the amide I and amide II regions. We do not take into account the separation between the amide I and II regions, as it is likely that explicit solvent modelling is needed to correctly reproduce this separation.[17,25]

Results

In the following, we present results for the cyclic peptides with two chiral centers (**1**-**3**), followed by those with three (**4**-**6**) and those with four chiral centers (**7**-**8**).

Cyclic peptides with two chiral centers

The experimental and computed spectra for cyclic peptides **1**, **2** and **3** are shown in Figure 2.

For tetrapeptide **1**, the experimental VCD spectrum (SS) is in fair agreement with calculated spectra for both the SS and SR stereoisomers in the frequency range 1800-1500 cm^{-1}. However, the experimental positive band at 1440 cm^{-1} is only reproduced for the SS stereoisomer (at 1470 cm^{-1}). For hexapeptide **2**, the experimental (SS) +/+/- pattern in

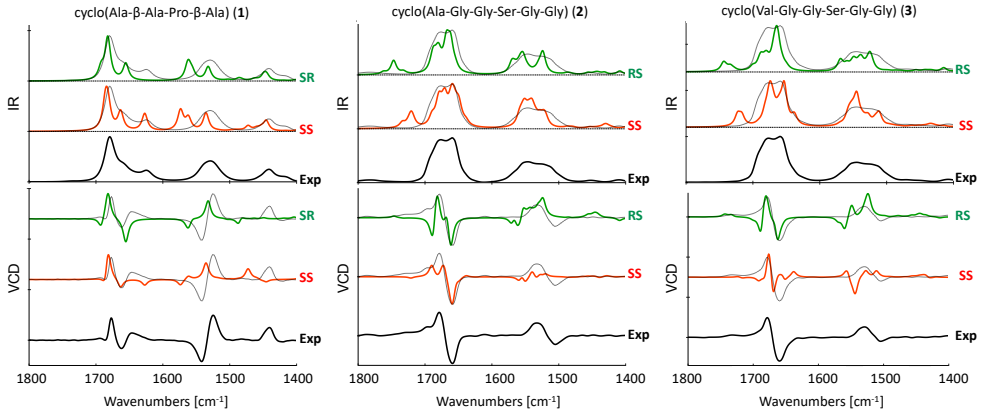


Figure 2: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **1** in ACN-d_3 [23] and **2** and **3** in DMSO-d_6. For each pair of enantiomers, the one with the largest positive overlap estimate in the amide I region is shown in the figures. The experimental spectrum is overlaid with the computed spectra as a dotted black line.

the amide I region is reproduced only in the computed VCD spectrum of the SS stereoisomer, while the RS stereoisomer shows a -/+/- pattern in this region. Neither of the calculated spectra reproduces the experimental +/- couplet in the amide II region. For hexapeptide **3**, the experimental (SS) +/- pattern in the amide I region is similarly reproduced only in the computed VCD spectrum of the SS stereoisomer, while the RS stereoisomer shows a -/+/- pattern with a prominent negative wing at 1690 cm^{-1}. Again, neither of the calculated spectra reproduces the experimental +/- couplet in the amide II region.

In summary, the AC of the three investigated cyclic peptides with two chiral centers can be determined by VCD. For tetrapeptide **1**, the calculated SS spectrum reproduces the experimental pattern in the entire investigated frequency range (1800-1400 cm^{-1}), whereas for hexapeptides **2** and **3**, only the amide I region in the experimental spectra is reliably reproduced by the calculations. This may be caused by the longer and more flexible peptide backbone in **2** and **3** compared to **1**.

Cyclic peptides with three chiral centres

The experimental and computed spectra for cyclic peptides **4**, **5** and **6** are shown in Figure 3.

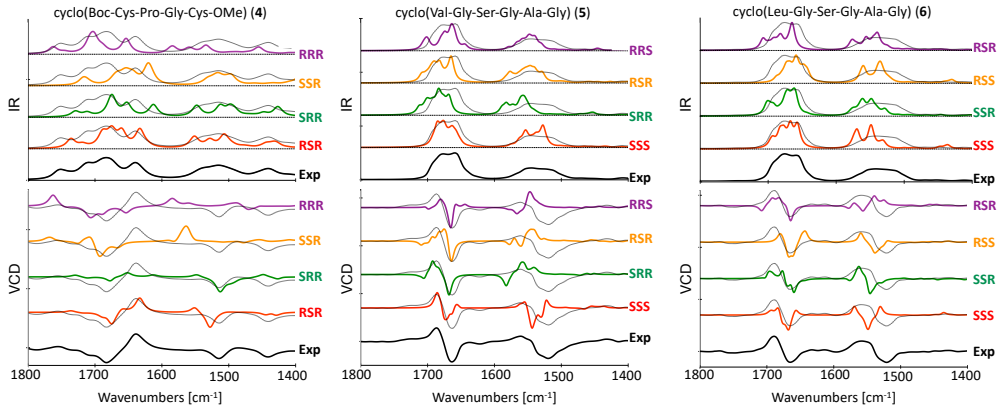


Figure 3: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **4** in ACN-d_3[24] and **5** and **6** in DMSO-d_6. For each pair of enantiomers, the one with the largest positive overlap estimate in the amide I region is shown in the figures. The experimental spectrum is overlaid with the computed spectra as a dotted black line.

For tetrapeptide **4**, the experimental (RSR) -/+ couplet in the amide I region as well as the negative band slightly above 1500 cm^{-1} are reproduced in the computed VCD spectrum of both the SRR and RSR stereoisomers. The relative intensity of the bands in the amide I and II regions is however only reproduced for the RSR stereoisomer. For hexapeptide **5**, the experimental (SSS) +/- couplet in both the amide I and II regions is reproduced only in the computed VCD spectrum for the SSS stereoisomer. For hexapeptide **6**, the experimental (SSS) +/- couplet in the amide I region is reproduced in the computed VCD spectrum of both the SSR and SSS stereoisomers. The experimental +/- couplet in the amide II region is better reproduced for the SSR stereoisomer than for the SSS stereoisomer, which shows a +/-/+ spectral pattern. Based on the amide II region, one could erroneously assign SSR as the AC, or conclude that the AC cannot reliably be assigned with VCD only.

In summary, the AC of two out of three cyclic peptides with three chiral centers can be

correctly assigned with VCD.

Cyclic peptides with four chiral centres

The experimental and computed spectra for cyclic peptides **7** and **8** are shown in Figure 4. Experimental spectra for two stereoisomers of tetrapeptide **7** are available.[24] Hence, two separate assignments can be made. We will refer to those as **7a** (RSSR) and **7b** (RSRR). Of the sixteen possible stereoisomers, only the computed spectra for the four stereoisomers with the largest positive overlap estimate in the amide I region are shown. The spectra of the four remaining diastereomers can be found in Figure S1 (**7a**), Figure S2 (**7b**) and Figure S3 (**8**).

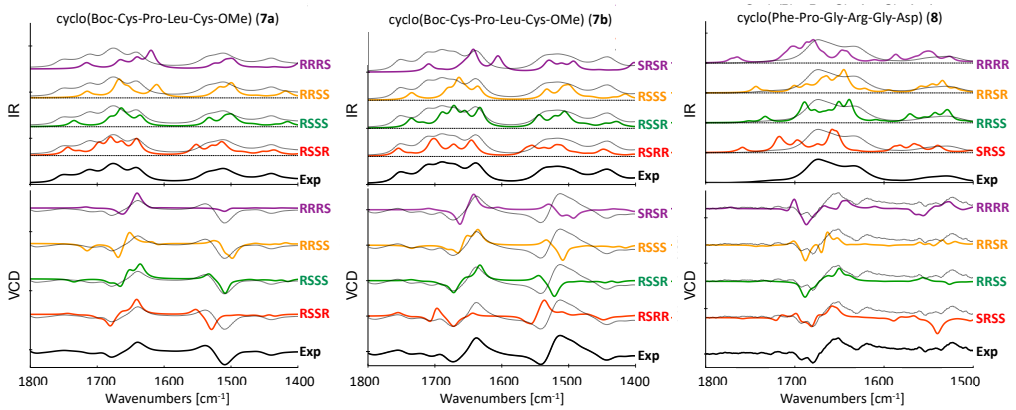


Figure 4: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **7a** in ACN-d_3,[24] **7b** in ACN-d_3[24] and **8** in TFE.[25] Computed spectra are shown for the four stereoisomers with the largest positive overlap estimate in the amide I region. The experimental spectrum is overlaid with the computed spectra as a dotted black line.

For tetrapeptide **7a**, the experimental (RSSR) -/+ couplet in the amide I region is reproduced in the computed VCD spectra of seven stereoisomers. Three of these can be eliminated due to a negative band between 1800-1700 cm^{-1} that is not observed in the experimental spectrum (RRSS in Figure 4 and SSSS and SSRS in Figure S1). For the four remaining stereoisomers, only calculated spectra for the RSSR and RSSS stereoisomers

reproduce the +/- couplet in the amide II region. For tetrapeptide **7b**, the experimental (RSRR) -/+ couplet in the amide I region is reproduced in the computed VCD spectra of four stereoisomers. The correct RSRR stereoisomer is not among those four, showing a positive band at 1700 cm^{-1} that is not observed in the experiment. The RSRR stereoisomer is, however, the only one reproducing the the experimental -/+ couplet in the amide II region. For hexapeptide **8**, the experimental (SRSS) -/-/+ spectral pattern in the amide I region is reproduced in the computed VCD spectra of the three stereoisomers SRSS, RRSS and RRSR. The experimental spectrum does not show an intense signal in the amide II region. The correct SRSS stereoisomer, however, shows a strong negative band in its computed spectrum.

In summary, the AC cannot be correctly assigned with VCD for any of the three investigated cyclic peptides with four chiral centers.

Discussion and conclusion

We have investigated to what extent VCD can be used to determine the AC of cyclic tetra- and hexapeptides with two, three and four chiral centers by comparing experimental VCD spectra with computed VCD spectra of all stereoisomers. We have shown that we can assign the AC in this way for all three cyclic peptides with two chiral centers, for two out of three cyclic peptides with three chiral centers and for none of the cyclic peptides with four chiral centers. The higher the number of chiral centers and the longer and more flexible the cyclic peptide, the more difficult it is to determine the AC. In the cases where the AC cannot be determined with VCD alone, one can at best limit the number of possible stereoisomers. At worst, an erroneous assignment is made based on VCD alone.

Although somewhat disappointing, it is not surprising that VCD alone cannot always determine the AC of molecules with three or more chiral centers. Other works have been successful in determining the AC of such molecules and have achieved this by combining

VCD with other types of spectroscopies. Hopmann *et al.* combined VCD, ROA and NMR to determine the AC of a highly flexible natural product with two chiral centers and an asymmetrically substituted double bond.[44] Dinku *et al.* combined VCD and NMR to determine the AC of a tricyclic triterpene acid with three chiral centers and three asymmetrically substituted double bonds.[45] Bogaerts *et al.* combined VCD and ROA to determine the AC of two rigid compounds with six and seven chiral centers.[16] Thus, combining VCD with other techniques such as ROA and NMR is a preferred strategy to correctly and reliably determine the AC of large and flexible molecules with multiple chiral centers.[8,14] We hypothesize that this strategy also works for cyclic peptides **6**–**8**, for which the AC could not be assigned with VCD only. Although we have shown that VCD independently can determine the correct AC for cyclic peptides **1**–**5**, a more unambiguous assignment can be made by combining several techniques.

Our study also shows that, in the absence of additional experimental data, it is important to calculate spectra for all possible stereoisomers of a flexible molecule with multiple chiral centers, to avoid an erroneous assignment based on an accidental agreement. Such an erroneous assignment becomes more probable with a higher number of chiral centers in a molecule.

Acknowledgement

This work has received support from the Research Council of Norway through a Centre of Excellence Grant (Grant No 262695) and a research project grant (Grant No 269425). The calculations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway (Grant No. NN4654K).

The authors are grateful to C. Merten, E. Vass, P. Bour, and T. Keiderling for helpful discussion and for kindly sharing experimental data.

Supporting Information Available

Supporting Information is available free of charge:

- Additional IR and VCD spectra of **7a**, **7b** and **8**.
- Used scaling factors.
- Overlap estimates.

References

(1) Jing, X.; Jin, K. A gold mine for drug discovery: Strategies to develop cyclic peptides into therapies. *Med. Res. Rev.* **2020**, *40*, 753–810.

(2) Abdalla, M. A.; McGaw, L. J. Natural Cyclic Peptides as an Attractive Modality for Therapeutics: A Mini Review. *Molecules* **2018**, *23*, 2080.

(3) Zorzi, A.; Deyle, K.; Heinis, C. Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol.* **2017**, *38*, 24–29.

(4) Smith, S. W. Chiral Toxicology: It's the Same Thing…Only Different. *Toxicol. Sci.* **2009**, *110*, 4–30.

(5) Kim, C. S.; Oh, J.; Lee, T. H. Structure elucidation of small organic molecules by contemporary computational chemistry methods. *Arch. Pharm. Res.* **2020**, *43*, 1114–1127.

(6) Devlin, F. J.; Stephens, P. J.; Besse, P. Are the absolute configurations of 2-(1-hydroxyethyl)-chromen-4-one and its 6-bromo derivative determined by X-ray crystallography correct? A vibrational circular dichroism study of their acetate derivatives. *Tetrahedron: Asymmetry* **2005**, *16*, 1557–1566.

(7) Menna, M.; Imperatore, C.; Mangoni, A.; Della Sala, G.; Taglialatela-Scafati, O. Challenges in the configuration assignment of natural products. A case-selective perspective. *Nat. Prod. Rep.* **2019**, *36*, 476–489.

(8) Bogaerts, J.; Aerts, R.; Vermeyen, T.; Johannessen, C.; Herrebout, W.; Batista, J. M. Tackling Stereochemistry in Drug Molecules with Vibrational Optical Activity. *Pharmaceuticals* **2021**, *14*, 877.

(9) Krupová, M.; Kessler, J.; Bouř, P. Recent trends in chiroptical spectroscopy: theory and applications of vibrational circular dichroism and Raman optical activity. *ChemPlusChem* **2020**, *85*, 561–575.

(10) Stephens, P.; Devlin, F.; Cheeseman, J. *VCD spectroscopy for organic chemists*; CRC Press, 2012; pp 1–360.

(11) Wesolowski, S. S.; Pivonka, D. E. A rapid alternative to X-ray crystallography for chiral determination: Case studies of vibrational circular dichroism (VCD) to advance drug discovery projects. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4019–4025.

(12) Kurouski, D. Advances of Vibrational Circular Dichroism (VCD) in bioanalytical chemistry. A review. *Anal. Chim. Acta* **2017**, *990*, 54–66.

(13) Batista Jr, J.; Blanch, E. W.; Bolzani, V. d. S. Recent advances in the use of vibrational chiroptical spectroscopic methods for stereochemical characterization of natural products. *Nat. Prod. Rep.* **2015**, *32*, 128–132.

(14) Merten, C.; Golub, T. P.; Kreienborg, N. M. Absolute Configurations of Synthetic Molecular Scaffolds from Vibrational CD Spectroscopy. *J. Org. Chem.* **2019**, *84*, 8797–8814.

(15) Polavarapu, P. L.; Santoro, E. Vibrational optical activity for structural characterization of natural products. *Nat. Prod. Rep.* **2020**, *37*, 1661–1699.

(16) Bogaerts, J.; Desmet, F.; Aerts, R.; Bultinck, P.; Herrebout, W.; Johannessen, C. A combined Raman optical activity and vibrational circular dichroism study on artemisinin-type products. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18014–18024.

(17) Eikås, K. D. R.; Beerepoot, M. T. P.; Ruud, K. A Computational Protocol for Vibrational Circular Dichroism Spectra of Cyclic Oligopeptides. *J. Phys. Chem. A* **2022**, *126*, 5458–5471.

(18) Cappelli, C.; Mennucci, B. Modeling the Solvation of Peptides. The Case of (s)-N-Acetylproline Amide in Liquid Water. *J. Phys. Chem. B* **2008**, *112*, 3441–3450.

(19) Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. Modeling solvent effects on chiroptical properties. *Chirality* **2011**, *23*, 717–729.

(20) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra. *Angew. Chem., Int. Ed.* **2017**, *56*, 14763–14769.

(21) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

(22) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(23) Vass, E.; Majer, Z.; Kőhalmy, K.; Hollósi, M. Vibrational and chiroptical spectroscopic characterization of γ-turn model cyclic tetrapeptides containing two β-Ala residues. *Chirality* **2010**, *22*, 762–771.

(24) Merten, C.; Li, F.; Bravo-Rodriguez, K.; Sanchez-Garcia, E.; Xu, Y.; Sander, W. Solvent-induced conformational changes in cyclic peptides: a vibrational circular dichroism study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5627–5633.

(25) Bouř, P.; Kim, J.; Kapitan, J.; Hammer, R.; Huang, R.; Wu, L.; Keiderling, T. Vibrational circular dichroism and IR spectral analysis as a test of theoretical conformational modeling for a cyclic hexapeptide. *Chirality* **2008**, *20*, 1104–1119.

(26) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

(27) Chan, W.; White, P. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: Oxford, 1999.

(28) Karlsen, E. A.; Stensen, W.; Juskewitz, E.; Svenson, J.; Berglin, M.; Svendsen, J. S. M. Anti-Colonization Effect of Au Surfaces with Self-Assembled Molecular Monolayers Functionalized with Antimicrobial Peptides on *S. epidermidis*. *Antibiotics* **2021**, *10*, 1516.

(29) Becke, A. D. A new mixing of Hartree Fock and local density functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(30) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(31) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11-18. *J. Chem. Phys.* **1980**, *72*, 5639–5648.

(32) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

(33) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(34) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(35) Frisch, M. J. et al. Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(36) Liegeois, V. DrawSpectrum Version 1.6.100. 2015.

(37) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

(38) Mennucci, B.; Martínez, J. M. How to Model Solvation of Peptides? Insights from a Quantum-mechanical and Molecular Dynamics Study of N-Methylacetamide. 1. Geometries, Infrared, and Ultraviolet Spectra in Water. *J. Phys. Chem. B* **2005**, *109*, 9818–9829.

(39) Cappelli, C.; Silva, C. O.; Tomasi, J. Solvent effects on vibrational modes: ab-initio calculations, scaling and solvent functions with applications to the carbonyl stretch of dialkyl ketones. *J. Mol. Struct.: THEOCHEM* **2001**, *544*, 191 – 203.

(40) Lamparska, E.; Liégeois, V.; Quinet, O.; Champagne, B. Theoretical Determination of the Vibrational Raman Optical Activity Signatures of Helical Polypropylene Chains. *ChemPhysChem* **2006**, *7*, 2366–2376.

(41) Liégeois, V.; Quinet, O.; Champagne, B. Vibrational Raman optical activity as a mean for revealing the helicity of oligosilanes: A quantum chemical investigation. *J. Chem. Phys.* **2005**, *122*, 214304.

(42) Zuber, G.; Hug, W. Rarefied Basis Sets for the Calculation of Optical Tensors. 1. The Importance of Gradients on Hydrogen Atoms for the Raman Scattering Tensor. *J. Phys. Chem. A* **2004**, *108*, 2108–2118.

(43) Kuppens, T.; Langenaeker, W.; Tollenaere, J. P.; Bultinck, P. Determination of the Stereochemistry of 3-Hydroxymethyl-2,3-dihydro-[1,4]dioxino[2,3-b]- pyridine by Vibrational Circular Dichroism and the Effect of DFT Integration Grids. *J. Phys. Chem. A* **2003**, *107*, 542–553.

(44) Hopmann, K. H.; Šebestík, J.; Novotná, J.; Stensen, W.; Urbanová, M.; Svenson, J.; Svendsen, J. S.; Bouř, P.; Ruud, K. Determining the Absolute Configuration of Two Marine Compounds Using Vibrational Chiroptical Spectroscopy. *J. Org. Chem.* **2012**, *77*, 858–869.

(45) Dinku, W.; Isaksson, J.; Rylandsholm, F. G.; Bouř, P.; Brichtová, E.; Choi, S. U.; Lee, S.-H.; Jung, Y.-S.; No, Z. S.; Svendsen, J. S. M.; Aasen, A. J.; Dekebo, A. Antiproliferative activity of a novel tricyclic triterpenoid acid from *Commiphora africana* resin against four human cancer cell lines. *Appl. Biol. Chem.* **2020**, *63*, 1–11.

Supporting Information: "Can the absolute configuration of cyclic peptides be determined with vibrational circular dichroism?"

Karolina Di Remigio Eikås,[*,†] Monika Krupová,[†] Tone Kristoffersen,[‡] Maarten T. P. Beerepoot,[†] and Kenneth Ruud[*,†,¶]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway*

‡*Department of Chemistry, UiT The Arctic University of Norway, 9037 Tromsø, Norway*

¶*Norwegian Defence Research Establishment, P.O. Box 25, 2027 Kjeller, Norway*

E-mail: karolina.s.eikas@uit.no; kenneth.ruud@uit.no

Additional IR and VCD spectra of 7a, 7b and 8

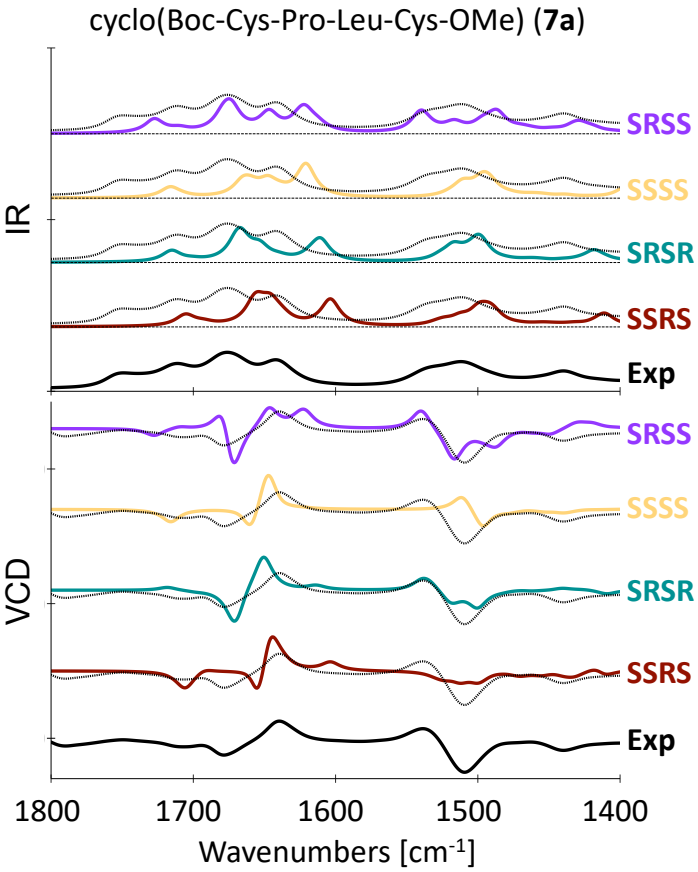


Figure S1: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **7a** in ACN-*d3*. Computed spectra are shown for the four stereoisomers with the smallest positive overlap estimate in the amide I region. The experimental spectrum is overlaid the computed spectra as a dotted black line.

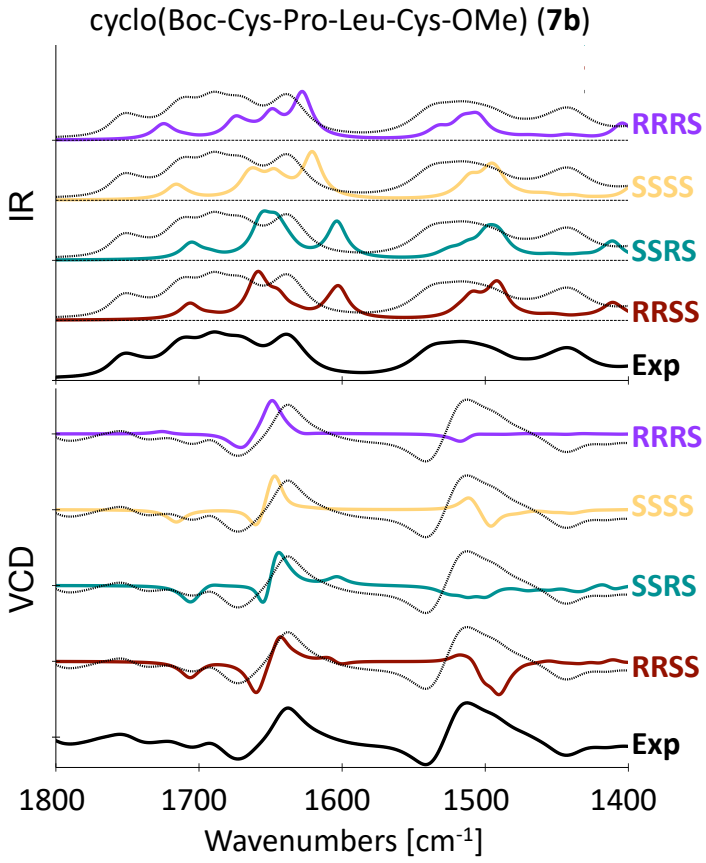


Figure S2: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **7b** in ACN-*d*3. Computed spectra are shown for the four stereoisomers with the smallest positive overlap estimate in the amide I region. The experimental spectrum is overlaid the computed spectra as a dotted black line.

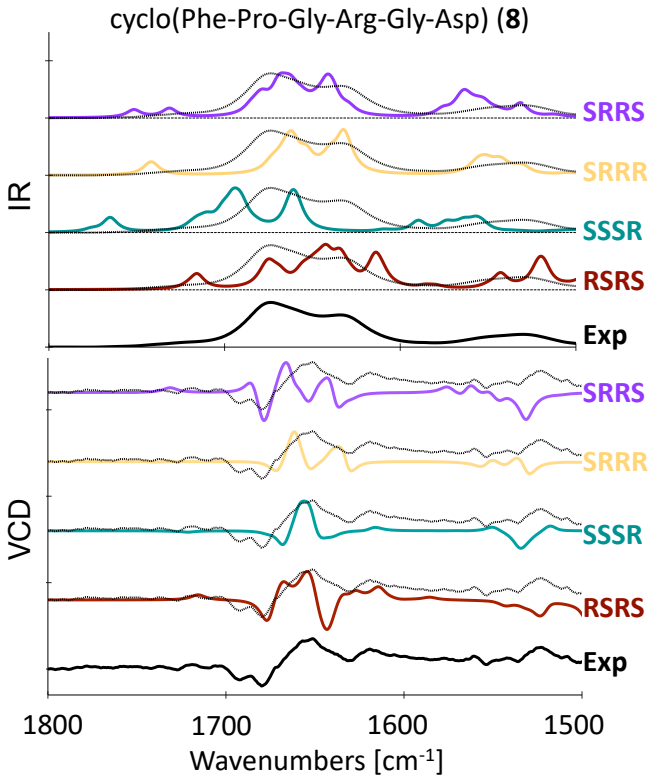


Figure S3: Experimental and computed (B3LYP/6-31+G*/CPCM) IR (top) and VCD spectra (bottom) of **8** in TFE. Computed spectra are shown for the four stereoisomers with the smallest positive overlap estimate in the amide I region. The experimental spectrum is overlaid the computed spectra as a dotted black line.

Scaling Factors

Table S1: Scaling factors used for molecules with two chiral centers: cyclo(Ala-β-Ala-Ala-β-Ala) (**1**), cyclo(Gly-Gly-Ser-Gly-Gly-Ala) (**2**) and cyclo(Gly-Gly-Ser-Gly-Gly-Val) (**3**).

	1		**2**		**3**	
	Amide I	Amide II	Amide I	Amide II	Amide I	Amide II
SS/RR	0.990	0.970	0.980	0.960	0.980	0.970
SR/RS	0.995	0.985	0.985	0.985	0.985	0.990

Table S2: Scaling factors used for molecules with three chiral centers: cyclo(Boc-Cys-Pro-Gly-Cys-OMe) (**4**), cyclo(Gly-Ala-Gly-Ser-Gly-Val) (**5**) and cyclo(Gly-Ala-Gly-Ser-Gly-Leu) (**6**).

	4		**5**		**6**	
	Amide I	Amide II	Amide I	Amide II	Amide I	Amide II
SRR/RSS	0.970	0.980	0.995	0.975	0.980	0.975
RSR/SRS	0.985	0.975	0.990	0.980	0.990	0.985
RRS/SSR	1.000	0.970	0.990	0.980	0.990	0.980
RRR/SSS	0.995	0.965	0.985	0.975	0.985	0.970

Table S3: Scaling factors used for molecules with four chiral centers: cyclo(Boc-Cys-Pro-S-Leu-Cys-OMe) (**7a**), cyclo(Boc-Cys-Pro-S-Leu-Cys-OMe) (**7b**) and Cyclo(Phe-Pro-Gly-Arg-Gly-Asp) (**8**).

	7a		**7b**		**8**	
	Amide I	Amide II	Amide I	Amide II	Amide I	Amide II
SSSS/RRRR	0.970	0.980	0.970	0.985	0.995	0.990
RSSS/SRRR	0.975	0.975	0.975	0.975	0.980	0.980
SRSS/RSRR	0.975	0.975	0.990	0.975	1.000	0.990
SSRS/RRSR	0.965	0.965	0.965	0.960	0.995	0.980
RRRS/SSSR	0.975	0.970	0.975	0.970	0.975	0.970
RRSS/SSRR	0.970	0.975	0.965	0.975	0.980	0.965
RSRS/SRSR	0.970	0.970	0.965	0.970	0.975	0.975
RSSR/SRRS	0.985	0.975	0.980	0.975	0.985	0.980

Overlap estimates

The overlap estimates are calculated over the amide I and amide II region separately. The scaling factors are determined by maximizing the combined IR and VCD overlap estimate (Equation 2) and are determined for each region separately. The scaling factors used for the amide I region are the same as those used in Figure 2-4 and S1-S3 and can be found together with those of the amide II region in Table S1-S3.

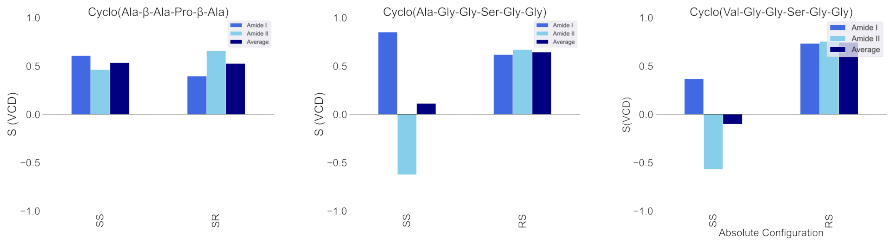


Figure S4: Overlap estimates (S(VCD)) between the calculated spectra of all diastereomers and the experimental spectra of cyclo(Ala-β-Ala-Pro-β-Ala) (**1**), cyclo(Gly-Gly-Ser-Gly-Gly-Ala) (**2**) and cyclo(Gly-Gly-Ser-Gly-Gly-Val) (**3**). The enantiomers with the positive S(VCD)-value in the amide I region are used as labels for the x-axis.

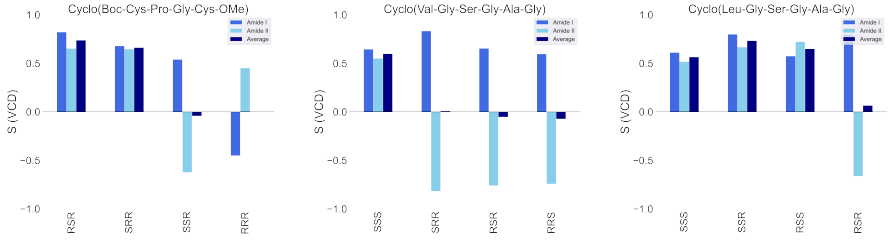


Figure S5: Overlap estimates (S(VCD)) between the calculated spectra of all diastereomers and the experimental spectra of cyclo(Boc-Cys-Pro-Gly-Cys-OMe) (**4**), cyclo(Val-Gly-Ala-Gly-Ser-Gly) (**5**) and cyclo(Leu-Gly-Ala-Gly-Ser-Gly) (**6**). The enantiomers with the positive S(VCD)-value in the amide I region are used as labels for the x-axis.

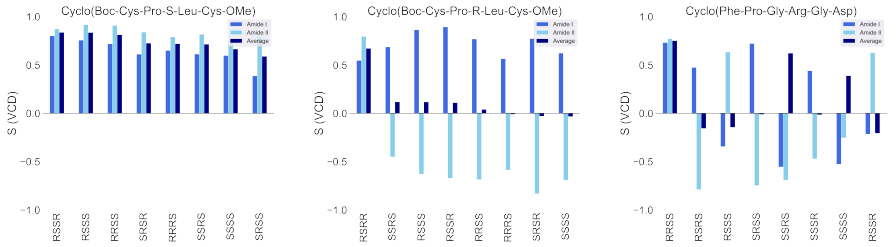


Figure S6: Overlap estimates (S(VCD)) between the calculated spectra of all diastereomers and the experimental spectrum of cyclo(Boc-Cys-Pro-S-Leu-Cys-OMe) (**7a**), cyclo(Boc-Cys-Pro-R-Leu-Cys-OMe) (**7b**) and Cyclo(Phe-Pro-Gly-Arg-Gly-Asp) (**8**). The enantiomers with the positive S(VCD)-value in the amide I region are used as labels for the x-axis.

Paper III

Conformational studies of cyclic hexapeptides with vibrational circular dichroism

Karolina Di Remigio Eikås, Tone Kristoffersen, Mitchell Ramesh Silk, John Sigurd Mjøen Svendsen, Kenneth Ruud and Monika Krupová,

Manuscript in preparation

Conformational Studies of Cyclic Hexapeptides with Vibrational Circular Dichroism

Karolina Di Remigio Eikås,[1,2] Tone Kristoffersen,[2] Mitchell Ramesh Silk,[2] John Sigurd Mjøen Svendsen,[2,3] Kenneth Ruud,[1,2,4] and Monika Krupová[1,2]

1 Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

2 Department of Chemistry, Faculty of Science and Technology, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

3 Amicoat AS, Sykehusvegen 23, 9019 Tromsø, Norway

4 Norwegian Defense Research Establishment, P.O.Box 25, 2027 Kjeller, Norway

E-mail: monika.krupova@uit.no

Abstract

Due to growing antibiotic resistance, new types of antimicrobials such as antimicrobial peptides (AMPs) are widely studied. Since their antimicrobial activity is often dependent on their structure and conformation, methods capable of providing such information quickly and easily are in high demand.

In this work, we use vibrational circular dichroism (VCD) spectroscopy and molecular modelling to study the structure and conformation of four cyclohexapeptides with antimicrobial activity. Since cyclohexapeptides generally possess significant conformational flexibility, an extensive conformational search must precede the *ab initio* calculation of VCD spectra. A typical workflow for this task is to do a statistical averaging of spectra for a large number of structures produced by molecular dynamics (MD) simulations. However, this approach is very computationally demanding, since it requires geometry optimization and VCD property calculation for a significant number of structures. Therefore, less expensive computational approaches based on Boltzmann averaging of conformers are also commonly used.

We here compare a classical MD conformational sampling approach with CREST based computational protocol based on meta-dynamics simulations and Boltzmann averaging for calculation of VCD spectra of model cyclohexapeptides in two solvents (water and DMSO). While calculations based on an MD conformational ensemble provide very good agreement with experimental spectra for both solvents, the CREST based calculations in most cases fail to reproduce the experimental spectral patterns. This is due to a strong preference of the CREST approach for conformations with internal side chain – backbone hydrogen bonding, which is caused by a combination of several factors, such as use of an implicit solvent model during the simulation, the length and nature of amino acid side chains, and/or use of dispersion-corrected energies.

Introduction

Growing antibiotic resistance is becoming one of the world's most urgent public health problems, with wide implications also for food security and technology development.[1-3] Bacterial infections are becoming increasingly difficult to treat as available antibiotics become less effective. Therefore, efforts to develop new classes of antimicrobials with reduced risk of triggering antibiotic resistance, such as antimicrobial peptides, are accelerating.[4-6]

Antimicrobial peptides (AMPs) are short and mostly positively charged peptides that can be found in a variety of organisms with direct and indirect (immunomodulatory) antimicrobial activity.[7, 8] The mode of action can vary in different AMPs but often involves peptide-membrane interactions, including association and membrane disruption.[9, 10] Although AMPs have diverse amino acid sequences, they often include hydrophobic and positively charged amino acids in their structure. The positively charged amino acids such as lysine (Lys, K) and arginine (Arg, R) are responsible for initial electrostatic interactions with the negatively charged bacterial membrane, while hydrophobic amino acids such as tryptophan (Trp, W) enable the peptide to interact with the hydrophobic parts of the phospholipid bilayer.[11-13]

In addition to the amino acid composition, the antimicrobial activity of AMPs also depends on their structure and conformation.[14, 15] Therefore, rapid and reliable experimental methods able to assess the dominating and/or antimicrobially active conformations of AMPs in various environments are much needed. The most common experimental techniques to study the conformation of peptides include X-ray crystallography[16, 17] and nuclear magnetic resonance (NMR) spectroscopy.[18, 19] Even though these techniques provide atomic-resolution structures, they have several drawbacks. For example, X-ray crystallography requires extensive experimental work in preparation of the diffracting crystal, which may in some cases even fail. In addition, the crystal structure of a molecule may deviate from its structure in the solution.[20, 21] More advanced NMR spectroscopy techniques have in many cases been successful in structural studies of small molecules in solution,[22] however, preparation of isotopically labeled samples may be complicated and expensive for bigger molecules.[23] Therefore, other techniques such as vibrational circular dichroism (VCD) that can quickly and easily provide insight into the general structure of peptides are becoming more and more important.[24-29] In VCD, the differential absorption of left- and right-handed circularly

polarized light in the infrared spectral range is measured, providing additional and often complementary information to that obtained in regular infrared absorption spectroscopy.

In the last decades, VCD has been used not only for absolute configuration determination of chiral molecules, but also for structural studies of biopolymers such as nucleic acids, proteins and peptides.[30-34] In the case of proteins and peptides, VCD has proven to be highly sensitive to backbone conformation and has enabled identification of the major structural motifs present in the biopolymers (α-helix, β-sheet, random coil, etc.).[35, 36] In peptides, also turns are common structural motifs. But, in contrast to rather rigid structural motifs such as α-helices and β-sheets, empirical comparisons of experimental VCD spectra with standardized spectral patterns for individual structural motifs cannot always reliably estimate the dominating turn. Thus, a theoretical interpretation of the spectra is needed to extract such information.[37, 38] The theoretical modeling of the spectra is straightforward and well-established for small rigid molecules; however, the situation gets more complicated with increasing size and flexibility of the molecules.

In this work we use VCD spectroscopy combined with theoretical interpretation of the experimental data to investigate the structure of four cyclohexapeptides with potential antimicrobial activity (data not published yet). All the investigated peptides contain Trp as a hydrophobic amino acid, and either Lys or Arg as a positively charged amino acid. The hydrophobic and charged amino acids either appear in alternating (cyclo(WXWXWX); X = K or R) or clustered (cyclo(WWWXXX); X = K or R) sequences.

Even though cyclization of small peptides generally results in lower flexibility, cyclohexapeptides still possess significant conformational flexibility. Therefore, *ab initio* calculation of IR and VCD spectra needs to be preceded by an extensive conformational search. In previous works, we developed a computational protocol to identify the most relevant conformations of cyclic peptides using CREST (Conformer-Rotamer Ensemble Sampling Tool).[39, 40] Compared to the cyclic peptides this protocol was validated for, the peptides investigated in this work have longer and charged side chains that can participate in intramolecular interactions. To ensure that the structures predicted by a CREST protocol are valid also for such peptides, we have compared the VCD spectra and the structures determined as relevant by the CREST protocol with a more traditional approach where the sampling of the conformational space is performed with classical molecular dynamics (MD) simulations.

There are two significant differences between the two approaches. First difference lies in how these computational methods model the solvent. While CREST uses an implicit solvent model during the simulation, explicit water molecules are present in the MD simulation allowing specific solvent-solute interactions to be explicitly accounted for, which may be important for obtaining the correct conformational ensemble, in particular for systems and solvents with strong intermolecular interactions. Second difference is related to how the relevant conformers in each approach are determined. In CREST, energies and Boltzmann averages are used to identify important conformers and to calculate the final vibrational spectra, while MD provides a statistical averaging of conformations observed during the simulation run. In this work we investigate which of these approaches leads to the best agreement with experimental spectra, and thus is best suited to aid in the structural analysis of cyclohexapeptides.

Methods

Synthesis and Vibrational Circular Dichroism of Cyclohexapeptides

Synthesis of Cyclohexapeptides. The cyclic peptides **1** – **4** (Figure 1) were synthesized using solid-phase peptide synthesis methodology[41] and fluorenylmethoxycarbonyl (Fmoc) chemistry procedures[42] as described previously. Briefly, the first amino acid was coupled to 2-Chlorotrityl chloride resin and the linear peptide precursors were prepared by an automated peptide synthesizer. The linear peptides were cleaved from the resin under mild acidic conditions and head-to-tail cyclized in solution. Next, the sidechains were deprotected and the crude cyclic peptides were purified by preparative reverse-phase HPLC. A detailed description of the synthetic procedure and characterization of the products can be found in the *Supporting Information*.

IR and VCD Spectra Acquisition. IR absorption and VCD spectra were measured with an Invenio R FTIR spectrometer with PMA50 module for polarization measurements (Bruker) at 8 cm^{-1} resolution and PEM set to 1600 cm^{-1} using a demountable BaF_2 cell and a 50 μm Mylar spacer. For spectra in water, D_2O was used as a solvent in order to avoid a strong deformation vibration band of H_2O overlapping with the amide I band of peptides. The peptides were mixed with D_2O and lyophilized overnight, and then dissolved in D_2O at a concentration of 20 mg/mL. For DMSO, non-

deuterated peptides were dissolved in DMSO-d_6 at concentration 25 mg/mL. These solutions were used for IR and VCD spectra measurements.

Usually, 50 μL of peptide solution was deposited on the bottom cell window and covered with the top window. Five blocks of 22500 scans (15 hours of total accumulation time) were collected and averaged. Spectra of D_2O / DMSO-d_6 measured at the same conditions were subtracted from the sample spectra. The values of *g*-factor were calculated as $\Delta A/A$, where ΔA is the amplitude of a VCD couplet and A is the absorption intensity of the corresponding IR band.

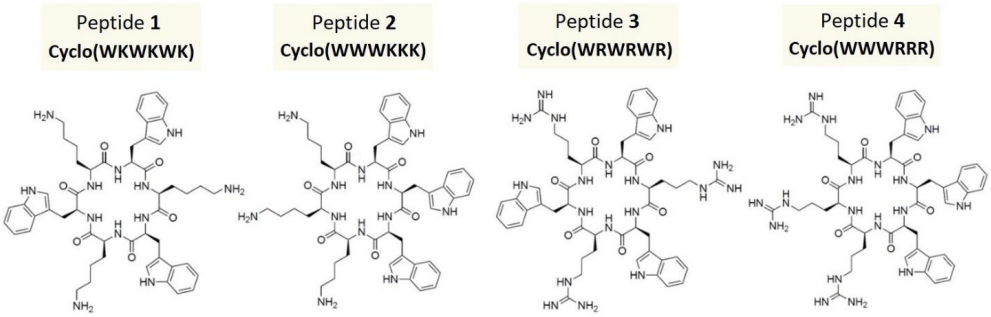


Figure 1. The investigated cyclohexapeptides **1** – **4**.

Calculation of Vibrational Circular Dichroism Spectra

Calculation of Spectra with CREST Conformational Sampling. The general workflow for conformational sampling described in our previous work[43] is based on meta-dynamics simulations as implemented in CREST 2.10. The conformer ensemble generated by CREST was then re-ranked by a DFT single-point energy calculation (B3LYP/6-31+G*[44]/CPCM[45, 46]) using Gaussian 16 (Rev. B.01)[47]. Conformers with energy within the 2.5 kcal/mol from the lowest-lying conformer after the DFT single-point energy calculations were geometry optimized, and the resulting structures were used to calculate IR and VCD intensities. If less than 20 conformers lied within 2.5 kcal/mol after the single-point energy calculation, we included the lowest-energy conformations above 2.5 kcal/mol up to a total number of 20 conformers to ensure that the conformational space was sufficiently sampled.[40]

These structures were then geometry optimized and VCD properties were calculated at the DFT level (B3LYP/6-31+G*/CPCM) using Gaussian 16 (Rev. B.01). After geometry optimization, some of the conformers ended up in the same minimum. To remove these duplicates, we compared the energies and rotational constants for the individual conformers, and only unique conformers were included in the final spectra. Final IR and VCD spectra were generated as Boltzmann averages based on free energies for the unique conformers using a Lorentzian bandshape with a full width at half maximum (FWHM) of 10 cm^{-1} with the CDSpecTech software.[48] The number of structures considered at each stage of the CREST calculation are summarized in Table 1.

Table 1. Number of conformers generated by CREST at each stage of IR and VCD spectra calculation.

	Solvent	Conformers found by CREST	Geom Opt Conformers	Unique conformers	Contributed to spectra >1 %
Peptide 1	water	326	20	19	5
	DMSO	297	34	25	3
Peptide 2	water	268	20	19	5
	DMSO	297	20	15	11
Peptide 3	water	368	20	20	5
	DMSO	230	20	20	7
Peptide 4	water	142	20	20	4
	DMSO	143	20	18	12

Calculation of Spectra with MD Conformational Sampling. The initial structure of the investigated cyclic peptides was prepared using the Maestro molecular modelling package.[49] The cyclic peptides were constructed by linking the N- and C-terminal residues of the linear peptides followed by an energy minimization. MD simulations were performed with some modifications according to the protocol of McHugh *et al.*[50] using GROMACS 2020[51-54] in water and in DMSO. For water, the RSFF1 force field[55] and TIP4P-Ew waters were used for the MD simulation. In the

case of DMSO, we used the OPLS-AA/L force field[56] and a box of pre-equilibrated DMSO molecules. Some modifications to the GROMACS library files as described by Jiang and Geng[57] were applied to generate input files suitable for cyclic peptides.

The peptides were placed in a cubic box containing pre-equilibrated solvent molecules (~2000 and ~500 molecules for the H_2O and DMSO simulations, respectively). The dimensions of the box were such that the minimum distance between solute molecule and the edges of the box was 1.0 nm. The solvated system was energy minimized with the steepest descent algorithm. The equilibration was performed in two stages with a step size of 2 fs. The first stage of equilibration consisted of a 50 ps *NVT* simulation at 300 K followed by a 50 ps *NPT* simulation at 300 K and a pressure of 1 bar. During the first step of equilibration, a position restraint with a force constant of 1000 kJ mol^{-1} nm^{-2} on the heavy atoms in the peptide was used. The second stage of equilibration was done without the position restraint on the peptide and consisted of a 100 ps *NVT* simulation at 300 K followed by a 100 ps *NPT* simulation at 300 K and 1 bar.

The production run was performed within the *NVP* ensemble at 300 K and 1 bar with a leapfrog algorithm[58] for 1 μs with a step size of 2 fs. The temperature was controlled using the V-rescale thermostat method[59] with a coupling time constant of 0.1 ps. The solute and solvent molecules were coupled to separate thermostats. The pressure was maintained by the Parrinello-Rahman barostat method with a coupling time constant of 2.0 ps and a compressibility of 4.5×10^{-5} bar^{-1}.

To constrain bonds involving hydrogens to their equilibrium value, the LINCS algorithm[60] was used. Non-bonding interactions were treated with a cut-off radius of 1.0 nm. Electrostatic interactions beyond this threshold were treated with the Particle Mesh Ewald method[61] with a Fourier spacing of 0.12 nm and an interpolation order of 4. To account for the truncation of the Lennard-Jones interactions, long-range analytic dispersion corrections to both energy and pressure were applied. The simulation trajectories were sampled every 10 ps. For each molecule, 200 snapshots were extracted from the trajectory, stripped of solvent, and the resulting structures were used for DFT calculation of IR and VCD spectra.

The geometry optimizations and VCD property calculations for conformers generated by the MD simulations were performed at the DFT level (B3LYP/6-31+G*/CPCM) using Gaussian 16 (Rev. B.01). The final spectra were generated using the CDSpecTech software[48] as a simple average of

spectra of all 200 snapshot geometries. A Lorentzian bandshape with a FWHM of 10 cm^{-1} was applied.

Clustering of the MD Conformers for Investigation of Their Backbone Structure. The geometry optimized structures from the MD simulations were clustered by the Conformer Cluster Panel script in Maestro.[49] The clustering was done based on the atomic root-mean-square deviation (RMSD) of the backbone, including hydrogens on both the α-carbons and amide groups. Based on the statistics from a test run, the Kelley index[62] was used to choose the optimal number of clusters. The clustering was then done with the number of clusters resulting in the lowest Kelley Penalty using the Average linkage method implemented in Maestro.[49] From each of the largest clusters, accounting for up to 90 % of the spectra, one structure was extracted and characterized in terms of dihedral angles, internal hydrogen bonding and turns.

Results and Discussion

We first focus on the experimental IR absorption and VCD spectra of peptides **1** – **4** in D_2O. We then evaluate the performance of the CREST-based computational protocol for these systems, followed by a comparison of the CREST and MD conformational sampling methodology for peptide **1**. Finally, we compare the two computational approaches for the calculation of vibrational spectra of peptide **1** in DMSO.

Experimental IR and VCD Spectra of Cyclohexapeptides in D_2O. Infrared absorption and VCD spectra of peptides are rich in structural information, since VCD spectral signatures can often be directly linked to the backbone conformation. Vibrational absorption spectra of proteins and peptides comprise mainly of two regions; the amide I region (1600 – 1700 cm^{-1}, mainly carbonyl stretching vibration) and the amide II region (~1550 cm^{-1} in H_2O and ~1450 cm^{-1} in D_2O, combination of N-H bending and C-N stretching vibrations). In Figure 2, we present experimental IR and VCD spectra of peptides **1** – **4** in D_2O. Heavy water was used as a solvent due to the strong absorption of H_2O in the carbonyl stretching region. This deuteration primarily affects the absorption energies in the amide II region, resulting in a significant shift to lower frequencies.

The infrared spectra of all investigated peptides in D_2O are dominated by a strong absorption band in the amide I region at 1672 cm^{-1} with a single shoulder at 1630 cm^{-1} for peptide **1** and **2**, and multiple other bands merged to a complex and wide feature with a maximum at 1610 cm^{-1} for peptide **3** and **4**. The more complex nature of the amide I region for peptides **3** and **4** originates in vibrations connected to the guanidino groups in the Arg side chains (see Figure S1 and Table S1 in *Supporting Information*). The amide II region is similar for all investigated peptides and is comprised of a complex band with a maximum at 1456 cm^{-1}.

Interestingly, all investigated peptides measured in D_2O provide similar features in their VCD spectra, primarily in the amide I region. These are characterized by a strong (+/–) couplet with a positive wing at 1652 cm^{-1} and a negative wing at 1627 cm^{-1}. The shape of the couplet is only slightly altered for individual peptides. Due to this remarkable similarity of the VCD spectra in the amide I region, we assume that the overall backbone conformation of the investigated cyclic peptides is similar and is only slightly altered by the amino acid side chains. More differences can be observed in the amide II region. For peptides **1** and **2** containing lysine, a negative band at 1423 cm^{-1} can be observed. Other features in the amide II region have too low intensity to be interpreted. For peptides **3** and **4** containing arginine, the amide II region is unusually intense and provides several spectral features. In the VCD spectra of peptide **3**, a 'W-shape' feature can be recognized with local minima at 1423 and 1387 cm^{-1}. The most prominent feature in the amide II region of the VCD spectrum of peptide **4** is a couplet with a positive wing at 1435 cm^{-1} and a negative wing at 1418 cm^{-1}, which is only slightly less intense than the amide I couplet. The *g*-factors calculated as a ratio between the absorption intensity and the amplitude of the corresponding VCD couplet for all investigated cyclohexapeptides are $\sim 5 \times 10^{-5}$ (see Table S2 in *Supporting Information*), which is a typical value for peptides.

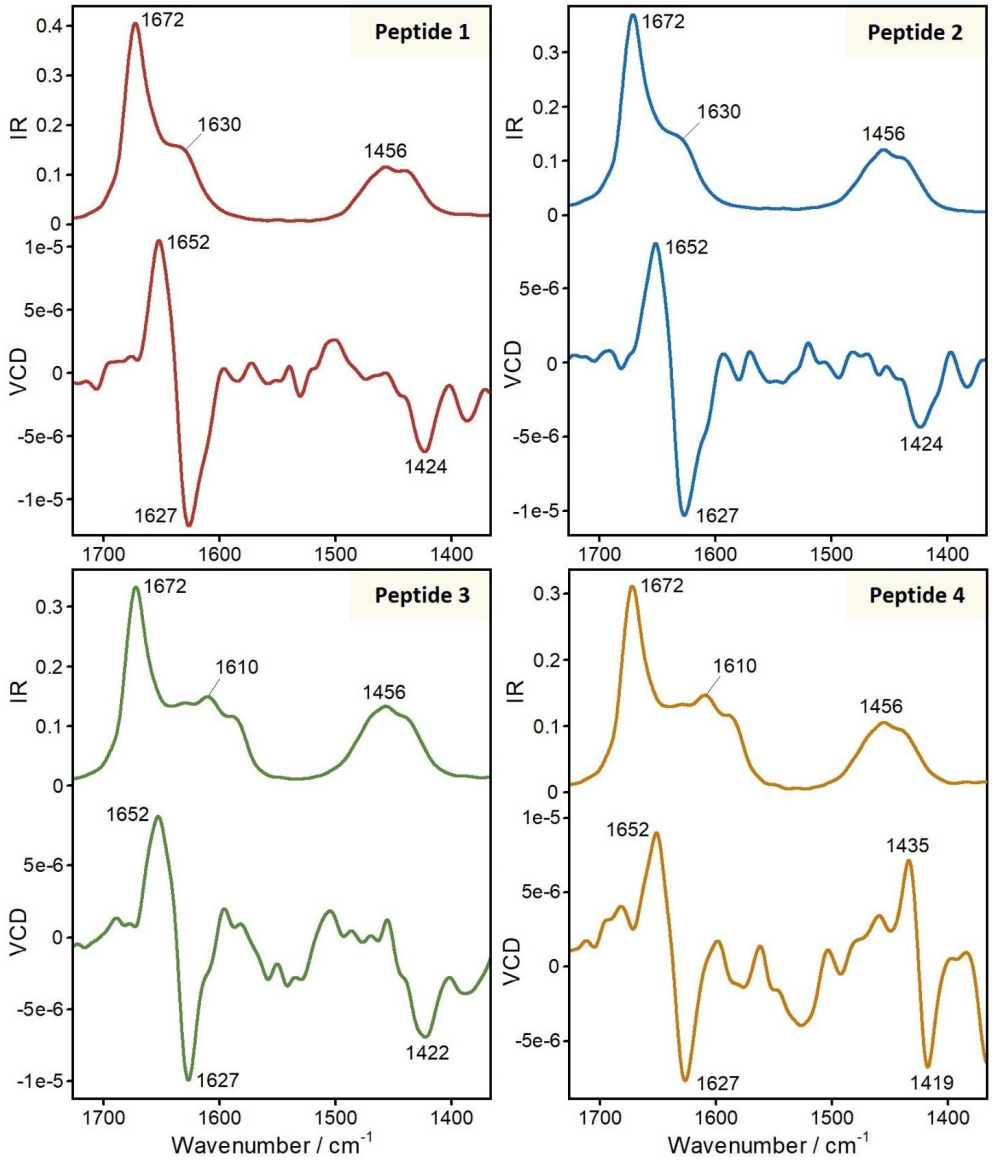


Figure 2. Experimental IR absorption and VCD spectra of investigated cyclohexapeptides **1** – **4** at 20 mg/mL in D_2O.

Calculation of Vibrational Spectra of Cyclohexapeptides Based on CREST Conformers in D_2O. To understand the structural information in the experimental spectra, we performed CREST conformational sampling and DFT spectra calculation for peptides **1** – **4** in D_2O. The results are summarized in Figure 3.

The spectra consist of the amide I region (1580 – 1720 cm^{-1}), which was in all cases simulated at significantly higher frequencies compared to experiment (shifts of 30 – 50 cm^{-1}), and the amide II region around 1450 cm^{-1}. The shift of the amide I region in the calculated spectra arises mainly from the neglect of specific solvent-solute interactions during the DFT spectra calculation using an implicit solvent model.[43] The overall agreement of the main VCD spectral signatures in the amide I region between the calculated and experimental spectra is poor. For peptide **1**, an experimental (+/–) VCD couplet centered at 1638 cm^{-1} shows a (+/–/+) signature in the calculation based on CREST conformers. For peptides **2** and **3**, more complex VCD signatures in the amide I region are not in agreement with experimental couplets. The best agreement between calculation and experiment in the amide I region was observed for peptide **4**. In this case, CREST correctly predicted a (+/–) couplet centered at 1678 cm^{-1}. The agreement of CREST signatures with experiment is much better in the amide II region. For peptides **1** – **3**, CREST was able to predict the overall spectral shapes, although the fine spectral features cannot be compared due to high signal-to-noise ratio of the experimental VCD spectra in this region. Interestingly, the CREST-based spectra for peptide **4** which showed the best agreement with experiment in the amide I region failed to correctly predict the spectral signatures in amide II region. The experimental VCD spectra of peptide **4** in the amide II region show a relatively strong (+/–) couplet centered at 1427 cm^{-1} that is not reproduced in the calculated spectra.

The backbone structures of the dominating conformers for peptides **1** – **4** predicted by CREST are illustrated in Figure S2 in the *Supporting Information*.

To investigate whether the poor agreement between the experimental and calculated VCD spectra lies in insufficient or erroneous conformational sampling, in the next section we compare the CREST conformational search with the statistical averaging of an MD conformational ensemble for peptide **1**.

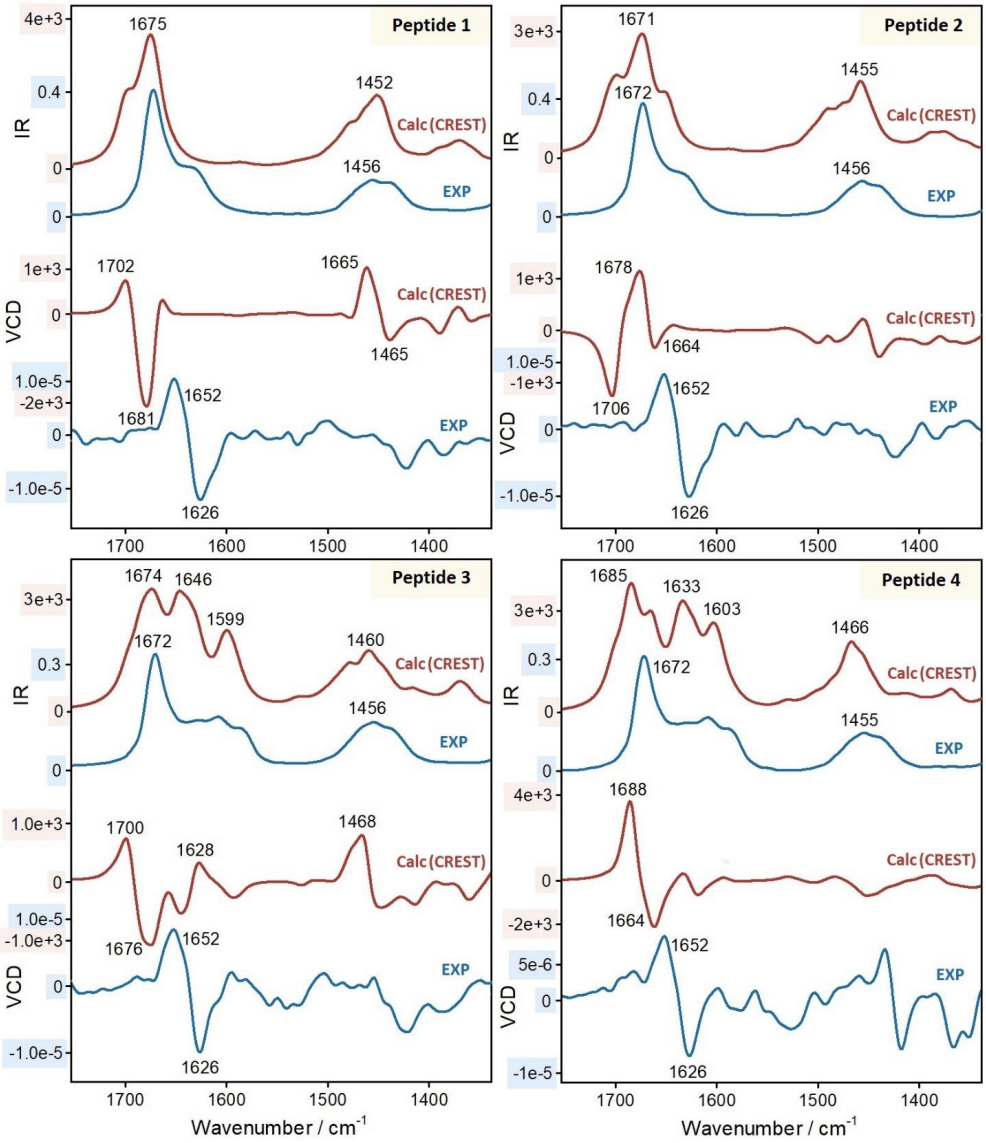


Figure 3. Experimental (D_2O, 20 mg/mL) and calculated (conformational sampling performed by CREST, spectra calculated at B3LYP/6-31+G*/CPCM level) IR and VCD spectra for cyclohexapeptides **1** – **4**.

Vibrational Spectra of Peptide 1 in D_2O based on MD and CREST Conformers. To link the experimental spectral shapes to the backbone conformation of cyclohexapeptides, we compare the experimental and calculated IR absorption and VCD spectra for peptide **1** in D_2O in Figure 4. The conformational sampling preceding the DFT calculation of the spectra was performed either by MD simulations (middle spectrum, red line) or CREST sampling (top spectrum, blue line). In the following text, we use the term 'MD based' when referring to the spectra calculated based on the MD conformational ensemble, and 'CREST based' for spectra calculated based on the CREST conformers.

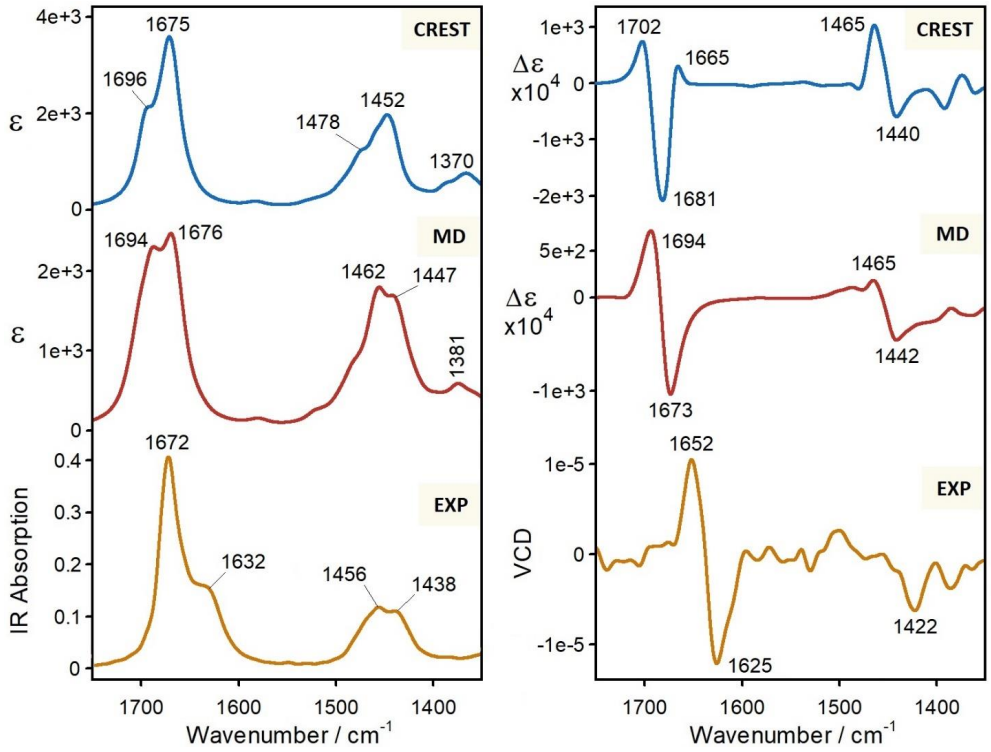


Figure 4. Comparison of experimental (orange; 20 mg/mL) and calculated IR absorption (left panel) and VCD (right panel) spectra of peptide **1** in D_2O. The conformational search for the peptide preceding DFT calculation was performed by two methods: CREST (blue) and MD (red).

Both MD and CREST based calculations correctly predicted the split of the amide I band in the IR spectrum into two components, although with opposite relative intensities. In this sense, the MD based calculation performed a bit better, predicting the relative intensities of the two peaks to be similar. In VCD, an experimental (+/–) couplet in the amide I region centered at 1638 cm^{-1} is very well reproduced in the MD based spectrum, although with a significant shift of ~45 cm^{-1} to higher frequencies due to the implicit solvent used in the DFT calculation. In the amide II region, the MD calculation provides a composite spectral pattern in VCD with a local maximum at 1465 cm^{-1} and a local minimum at 1442 cm^{-1}, which is not consistent with the experimental 'W'-like pattern. The convergence of the spectra based on the number of averaged MD structures is shown in Figure S3 (*Supporting Information*).

Analysis of the Calculated Conformations of Peptide 1 in D_2O. It is evident that using MD and CREST conformers results in different spectral patterns. To better understand the origin of these differences, we investigated the backbone structure of the conformational ensembles produced by the two approaches.

The backbone structure of the investigated peptides is determined primarily by hydrogen bonding between the carbonyl oxygens and amide hydrogens within the backbone. Various backbone conformations can be classified as different turns, characterized by several parameters. The first parameter is the number of peptide bonds between a hydrogen acceptor (C=O) and a hydrogen donor (N-H). According to this, turns in peptides can be divided into different classes: π-turn (5 peptide bonds), α-turn (4 bonds), β-turn (3 bonds), γ-turn (2 bonds) and δ-turn (1 bond). Of these, β-turns are the most usual form followed by γ-turns, while δ-turn is the least common type due to high steric hindrance. These classes of turns are further divided into different subtypes depending on the dihedral angles the backbone assumes. The more common β-turns are classified by ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2} and ψ_{i+2} dihedral angles, while the slightly less common γ-turns are classified based on the dihedral angles ϕ_{i+1} and ψ_{i+1} (Figure 5).

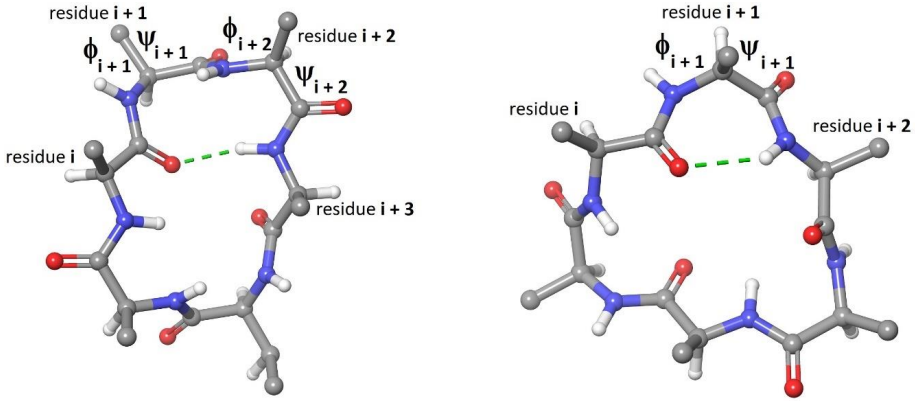


Figure 5. Example of peptide **1** backbone structure with a β- (left) and γ-turn (right). Peptide turns are further classified based on dihedral angles (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2} and ψ_{i+2} for β-turns, and ϕ_{i+1} and ψ_{i+1} for γ-turns).

The cyclohexapeptides discussed in this work contain mostly β- and γ-turns in their backbone structure. There are two types of γ-turns: a classical γ-turn (γ^{cl}) with dihedral angles approximately $\phi_{i+1} \sim 75°$ and $\psi_{i+1} \sim -65°$ and an inverse γ-turn (γ^{inv}) with dihedral angles $\phi_{i+1} \sim -75°$ and $\psi_{i+1} \sim 65°$. β-turns are in general more flexible and can assume at least 8 different conformations. Of these, five types of β-turns were identified in the investigated cyclohexapeptides and the values of the dihedral angles corresponding to these types of β-turns are summarized in Table 2.

Table 2. Ideal values for 4 types of β-turns identified in the investigated cyclohexapeptides.[63]

	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
β_I	–60 °	–30 °	–90 °	0 °
$\beta_{I'}$	60 °	30 °	90 °	0 °
β_{II}	–60 °	120 °	80 °	0 °
$\beta_{II'}$	60 °	–120 °	–80 °	0 °
β_{IV}	does not fit any other β-turn criteria			

Based on a backbone similarity analysis, extracted and geometry optimized structures of peptide **1** from the MD simulation in water were divided into 19 clusters. The 9 largest clusters account for up to 90% of the spectra, and their backbone conformations are classified by their turns in Table 3. In water, β_{II}- and $\beta_{II'}$-turns are the most prevalent in the MD structures. Both types of γ-turns are also present, but only in combination with β-turns. In the CREST ensemble, the two most populated conformers have a combination of β_{IV}- and γ^{inv}-turns in their structure. The remaining three conformers with lower Boltzmann weights have surprisingly symmetrical structures (see Figure S4 in *Supporting Information*) containing three γ^{cl}- and three γ^{inv}-turns.

Table 3. Structural analysis of conformational ensembles predicted by MD and CREST simulations of peptide **1** in water. For MD, individual clusters were obtained by backbone comparison in Maestro. Only the largest clusters that in sum contribute to 90% of the final spectra are included in the table. For CREST, all conformers that contribute with more than 1% to the final spectra are included.

Molecular Dynamics

	Turn(s)	**Weight**	**# of conf**
Cluster 1	$\beta_{II'}$ and β_{I}	23.5 %	47
Cluster 2	β_{II}	23.0 %	46
Cluster 3	β_{II} and $\beta_{II'}$	10.5 %	21
Cluster 4	β_{II} and γ^{cl}	10.5 %	21
Cluster 5	β_{II}, γ^{inv}, γ^{cl}	8.0 %	16
Cluster 6	$\beta_{II'}$	7.0 %	14
Cluster 7	$\beta_{II'}$ and γ^{inv}	5.5 %	11
Cluster 8	$\beta_{I'}$ and $\beta_{II'}$	3.0 %	6
Cluster 9	$\beta_{II'}$ and γ^{cl}	3.0 %	6

CREST

	Turn(s)	**Weight**
Conformer 1	β_{IV} and γ^{inv}	82.0 %
Conformer 2	β_{IV} and γ^{inv}	7.5 %
Conformer 3	3 γ^{cl}, 3 γ^{inv}	5.2 %
Conformer 4	3 γ^{cl}, 3 γ^{inv}	1.9 %
Conformer 5	3 γ^{cl}, 3 γ^{inv}	1.8 %

To better understand the origin of the differences in backbone structure, we compare the structure of a representative MD conformer from the largest cluster and a CREST conformer with the largest Boltzmann weight (Figure 6, **(a)** and **(c)** for MD, and **(b)** and **(d)** for CREST). A representative MD conformer has fully extended Lys side chains, allowing for hydrogen bonding with water molecules. Although the general backbone conformation differs between the MD clusters, most of the conformers (~90 %) have fully extended Lys side chains interacting with solvent molecules. In contrast to the MD structures, the dominating CREST conformer contains two Lys side chain amino groups hydrogen bonded to the peptide backbone carbonyl oxygens.

One reason for the significant difference in conformations simulated by MD and CREST may originate in the way a solvent is included. Explicit water molecules are present during the MD simulation, while implicit solvation by a continuum model is used for CREST. The use of implicit solvation during the CREST simulation results in a strong preference for conformations with hydrogen bonding between the Lys amino group hydrogens and the backbone carbonyl oxygens. This internal hydrogen bonding has a considerable influence on the backbone conformation, resulting in an altered conformational ensemble that subsequently leads to the false signatures in the CREST based VCD spectra. Another reason for the unrealistic conformational ensemble may be the use of dispersion-corrected energies when choosing relevant conformers during the CREST simulation. As shown in other works, including dispersion corrections often results in a preference for structures with more internal interactions. Therefore, although CREST may find conformers more similar to the ones predicted by MD simulations, only the ones with more internal interactions are determined relevant due to the selection based on dispersion-corrected energies.

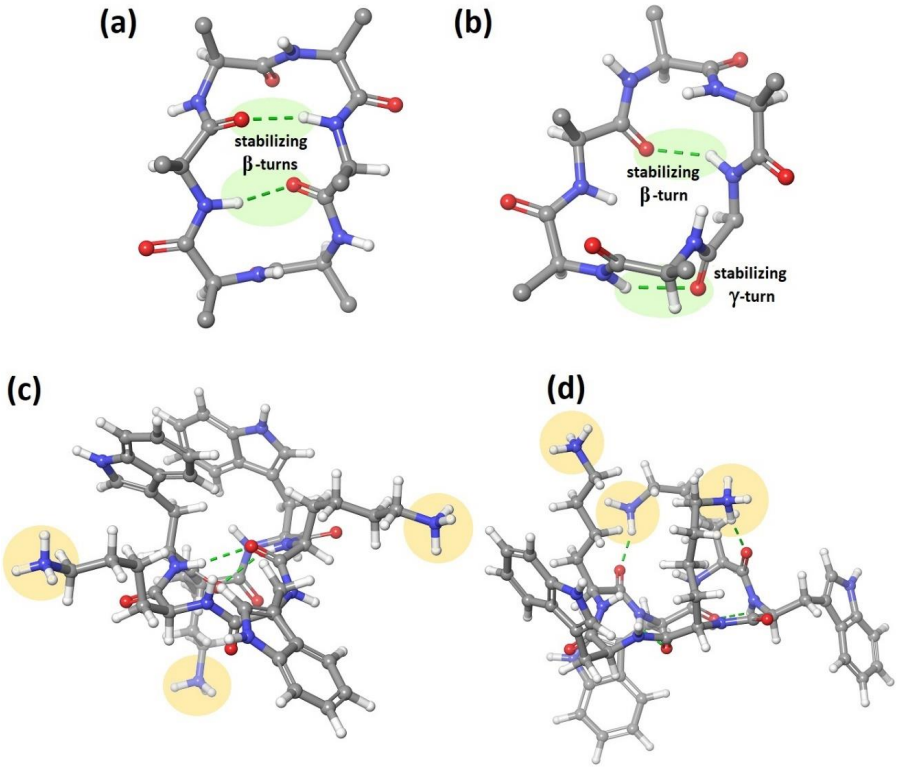


Figure 6. The dominating structures from MD (**a** and **c**) and CREST (**b** and **d**) conformational ensemble for peptide **1** in water. Hydrogen bonding is depicted by a green line. The top row shows only the backbone and hydrogen bonding within the backbone, while the bottom row shows the whole peptide including the side chains. For MD, a representative structure from the largest cluster is shown, while for CREST the conformer with the largest Boltzmann weight is shown.

Vibrational Spectra of Peptide 1 in DMSO based on MD and CREST Conformers. In our previous work, we successfully applied a protocol for CREST / DFT calculation of vibrational properties to cyclic peptides in various organic solvents. Therefore, in order to see if CREST based calculations of VCD spectra for peptides **1** – **4** work better in organic solvents than in water, we present the experimental and CREST / MD based IR absorption and VCD spectra for peptide **1** in DMSO in Figure 7. Moreover, since VCD signatures are linked to the backbone conformation,

which is strongly influenced by intra- and inter-molecular interactions, this allows us to investigate the solvent influence on peptide structure.

The experimental IR spectrum in the amide I region is dominated by a strong and relatively sharp absorption maximum at 1690 cm^{-1} accompanied by a prominent shoulder at 1653 cm^{-1}. The amide II region is broad with a maximum at 1530 cm^{-1}. In VCD, a strong (+/–) couplet in the amide I region centered at 1663 cm^{-1} can be observed. This couplet is asymmetric with a stronger positive wing at 1678 cm^{-1} and a less intense sharper negative wing at 1651 cm^{-1}. In the amide II region, a broad negative peak is located at 1506 cm^{-1}. The *g*-factor is 6×10^{-5}.

Since peptide **1** was not deuterated in DMSO, a direct spectral comparison with D_2O is not possible. However, some conclusions can still be made. Firstly, a carbonyl stretching band in DMSO is shifted to higher frequencies than in D_2O (maxima at 1672 cm^{-1} in D_2O vs. 1690 cm^{-1} in DMSO). This is due to strong non-covalent interactions between the carbonyl oxygen and water molecules, that decrease the strength of the C=O bond and subsequently cause the shift of its stretching vibration to lower frequencies. Such interactions are not present for DMSO. Another consequence of this is a slightly broader bandwidth for D_2O.

Moreover, even though we cannot directly use spectral shapes for structural comparison, a difference in backbone structure in two solvents can be expected. DMSO is a strong hydrogen bond acceptor interacting strongly with amide and amine hydrogens of the peptide backbone and polar side chains and unlike water, it lacks hydrogens which can participate in donor hydrogen bonding with carbonyls of the peptide backbone. Therefore, these varied peptide-solvent interactions in different solvents can lead to significant alterations of the peptide backbone conformation, leading to differences in the spectra.

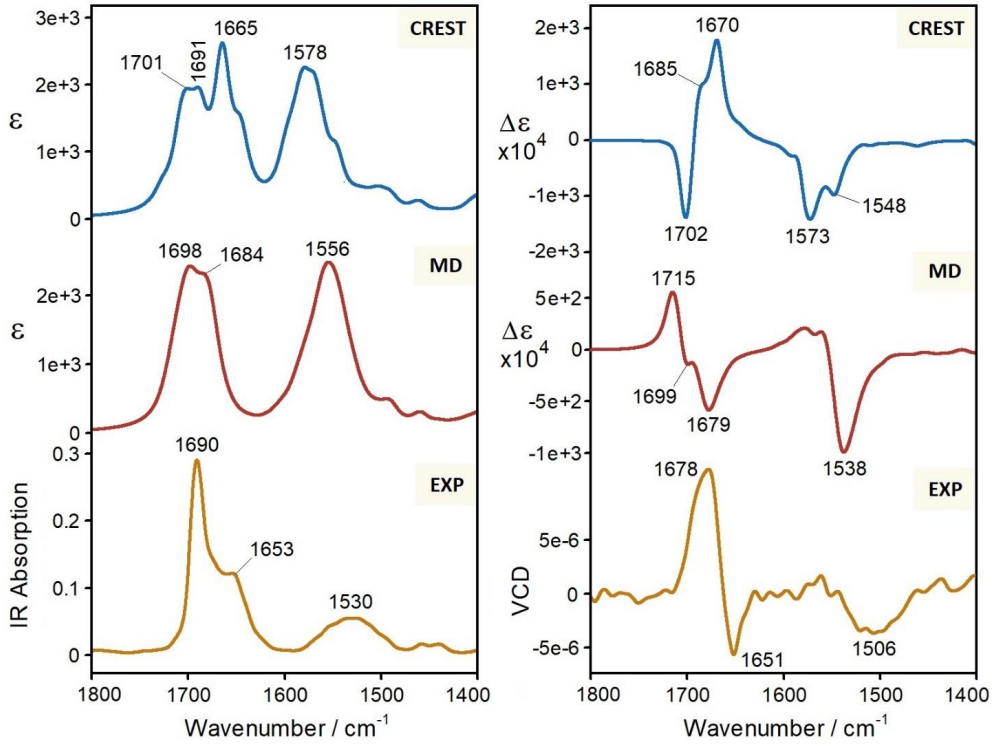


Figure 7. Comparison of experimental (orange, concentration 25 mg/mL) and calculated IR absorption (left panel) and VCD (right panel) spectra of peptide **1** in DMSO. The conformational search for cyclohexapeptide preceding the DFT calculation was performed by two methods: CREST (blue), and molecular dynamics sampling (red).

The calculated IR spectrum based on MD conformers bears some resemblance to the experiment. In the calculated spectrum, the amide I region consists of a superposition of two bands with maxima at 1698 and 1684 cm^{-1}. The separation of these two bands is smaller than in experiment (14 cm^{-1} vs. 37 cm^{-1}, respectively), and their relative intensities do not match the experiment well. In the amide II region, the calculation predicts a broad band at 1556 cm^{-1}, which is reproduced in the experiment at 1530 cm^{-1}. In VCD, the main sign of the couplet in amide I region is reproduced correctly, even though its shape is not entirely consistent with the experiment. Moreover, it is predicted at higher frequencies than in the experiment (1700 cm^{-1} vs. 1663 cm^{-1}, respectively). In

the amide II region, the calculation predicts a strong negative band located at 1538 cm^{-1}, which is in good agreement with the experiment, even though it is slightly shifted to higher frequencies. As for water, a convergence of the spectra based on the number of averaged MD structures is discussed in Figure S5 (*Supporting Information*).

For peptide **1** in DMSO, the CREST based spectra performed very poorly in predicting the correct shapes and signs of the experimental IR absorption and VCD spectra. In IR, the bandshape clearly consists of several peaks, of which the one located at 1665 cm^{-1} has the highest intensity, which is not in agreement with the experimental spectral shapes. The amide II region is correctly predicted with a broad single band, even though it is significantly shifted to higher frequencies when compared to experiment (1578 cm^{-1} vs. 1530 cm^{-1}, respectively). In the amide I region, the CREST based VCD spectra provided a strong (–/+) couplet centered at 1690 cm^{-1} having opposite signs compared to experiment. In the amide II region, two negative peaks are observed at 1573 and 1548 cm^{-1}. Even though the sign of the bands in the amide II region is predicted correctly, there is a splitting of the amide II band in the CREST based spectrum that is not observed in the experimental spectrum.

Analysis of the Calculated Conformations of Peptide 1 in DMSO. As for water, we analyze the conformational ensemble from the CREST and MD simulations in terms of their backbone conformations. The MD structures were clustered based on their backbone structure and in DMSO this resulted in 27 clusters, where the 15 largest of them account for up to 90% of the spectra (Table 4). Most of the geometries have one or two internal hydrogen bonds in the backbone and adopt conformations with β- or γ-turns. The dominating β-turn is β_I followed by β_{II}. Classical and inverted γ-turns are almost equally represented. One structural cluster with low weight (2 %) adopted a backbone conformation with a combination of β_I- and π-turns.

The three CREST conformers identified as relevant have more or less identical backbone structures with three hydrogen bonds resulting in three turns: one β_I-, one $\beta_{II'}$- and a classical γ-turn. Interestingly, all relevant conformers from CREST contain a $\beta_{II'}$-turn, which is observed only in one of the structures analyzed from the MD ensemble. As for water, both CREST and MD in DMSO produce conformational ensembles with quite different backbone structures and this directly translates to differences in their spectra.

Table 4. Structural analysis of conformational ensembles predicted by MD and CREST simulations of peptide **1** in DMSO. For MD, individual clusters were obtained by backbone comparison in Maestro. Only the largest clusters that in sum contribute to 90% of the final spectra are included in the table. For CREST, all conformers that contribute with more than 1% to the final spectra are included.

Molecular Dynamics

	Turn(s)	**Weight**	**# of conf**
Cluster 1	β_{II} and γ^{cl}	13.0 %	26
Cluster 2	β_{I}	11.0 %	22
Cluster 3	β_{I}	11.0 %	22
Cluster 4	$\beta_{I'}$	8.5 %	17
Cluster 5	γ^{cl} and γ^{cl}	8.0 %	16
Cluster 6	γ^{cl}	8.0 %	16
Cluster 7	γ^{cl} and γ^{inv}	6.0 %	12
Cluster 8	β_{I}	5.5 %	11
Cluster 9	β_{I} and $\beta_{II'}$	5.0 %	10
Cluster 10	β_{II}	3.0 %	6
Cluster 11	γ^{cl} and γ^{inv}	3.0 %	6
Cluster 12	γ^{cl} and γ^{inv}	3.0 %	6
Cluster 13	β_{I} and γ^{cl}	2.0 %	4
Cluster 14	β_{I} and π	2.0 %	4
Cluster 15	γ^{inv}	2.0 %	4

CREST

	Turn(s)	**Weight**
Conformer 1	$\beta_{II'}$, β_{I} and γ^{cl}	97.0 %
Conformer 2	$\beta_{II'}$, β_{I} and γ^{cl}	2.0 %
Conformer 3	$\beta_{II'}$, β_{I} and γ^{cl}	1.0 %

We showed that hydrogen bonding of the side chains to the backbone significantly affects CREST based conformers in water. Therefore, we compare the side chain conformation of peptide **1** in DMSO based on MD and CREST conformers in Figure 8. A representative MD conformer from the largest cluster (Figure 8; **(a)** only backbone, **(c)** backbone and amino acid side chains) and a

CREST conformer with the largest Boltzmann weight (Figure 8; **(b)** only backbone, **(d)** backbone and amino acid side chains) are shown. As for water, the Lys side chains for most of the MD conformers in DMSO (~99 %) are fully extended and do not interact with the backbone. This is not the case for the CREST conformers, where all three of the Lys side chains are hydrogen bonded to carbonyl oxygens in the backbone. As for water, this hydrogen bonding seriously affects the backbone conformation, resulting in erroneous VCD signatures.

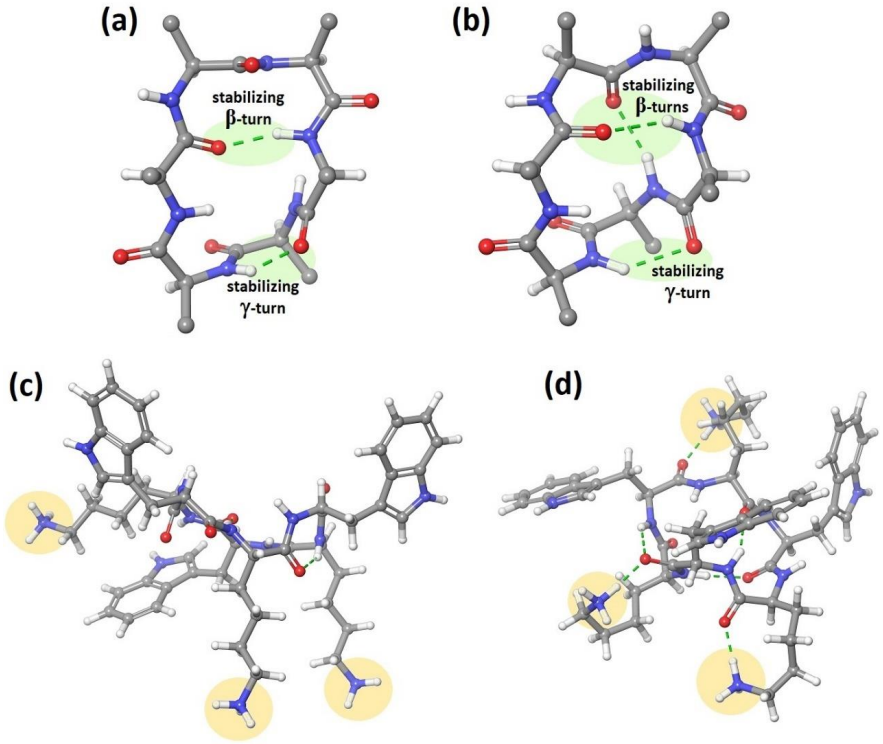


Figure 8. The dominating structures from MD (left) and CREST (right) ensemble of peptide **1** in DMSO. Hydrogen bonding is depicted by a green line. The top row shows only the backbone and hydrogen bonding within the backbone, while the bottom row shows the whole peptide including the side chains. For MD, a representative structure from the largest cluster is shown, while for CREST the conformer with the largest Boltzmann weight is shown.

Summary and concluding remarks

In this work, we used VCD spectroscopy to study the conformation of four cyclohexapeptides with antimicrobial activity in two different solvents: D_2O and DMSO. We found that all the investigated peptides in D_2O show similar VCD signatures in the amide I region, suggesting that their backbone conformation is similar and only influenced to a very minor extent by the nature of the amino acid side chains or their sequence.

To understand the structural information contained in the spectra, we used theoretical modelling. Since cyclohexapeptides are flexible molecules and a proper sampling of conformational space is needed before the calculation of spectral properties, the simulation of the spectra consisted of two steps: (i) a conformational search performed either by meta-dynamics as implemented in CREST, or by a statistical conformational sampling using molecular dynamics simulations; and (ii) subsequent DFT calculation of spectral properties.

The conformational ensemble provided by MD proved to better reproduce VCD spectral signatures for a model cyclohexapeptide both in D_2O and DMSO. The calculation of spectral properties based on the CREST conformational ensemble was not in agreement with experimental data, and in some cases even provided opposite spectral signatures. A structural analysis of the conformational ensembles generated by both methods showed that CREST favors conformations with hydrogen bonding between the amine hydrogens on the Lys side chains and carbonyls of the backbone. Such internal hydrogen bonding was only rarely present amongst the MD structures.

This erroneous internal hydrogen bonding in the CREST conformers can originate from: (i) the inclusion of solvent effects during the meta-dynamics simulation by an implicit solvent model; (ii) dispersion correction of energies favoring the internal molecular interactions; and (iii) the length and polarity of amino acid side chains. The third point is especially important for our investigated cyclohexapeptides since these contain Lys and Arg side chains which are charged (and therefore strongly interacting) and long / flexible enough to sterically interact with the backbone.

Very recently, Grimme *et al.* added the possibility of including explicit solvent molecules in the CREST simulation.[64] To what extent this would improve the predicted structures of these cyclic peptides needs to be explored in future works to improve the agreement between the calculated and experimental VCD spectra.

Acknowledgement

This work has received support from the Research Council of Norway through a Centre of Excellence Grant (Grant 262695) and a research project grant (Grant 269425). The calculations were performed on resources provided by Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway (Grant NN4654K).

Supporting Information

- Protocol for Synthesis of Cyclohexapeptides **1** – **4**
- Interpretation of Low-Frequency IR Bands in the Amide I Region of Peptides **3** and **4**
- Values of *g*-factors for VCD spectra of peptides **1** – **4** in D_2O
- Backbone structures for dominating conformers of peptides **1** – **4** calculated by CREST
- Convergence of the MD based IR and VCD spectra of peptide **1** in D_2O
- The low-energy conformer of peptide **1** in D_2O
- Convergence of the MD based IR and VCD spectra of peptide **1** in DMSO

References

1. Alanis, A. J., Resistance to antibiotics: Are we in the post-antibiotic era? *Arch. Med. Res.* **2005,** *36* (6), 697-705.

2. Frieri, M.; Kumar, K.; Boutin, A., Antibiotic resistance. *J. Infect. Public Health* **2017,** *10* (4), 369-378.

3. Gandra, S.; Barter, D. M.; Laxminarayan, R., Economic burden of antibiotic resistance: How much do we really know? *Clin. Microbiol. Infect.* **2014,** *20* (10), 973-980.

4. Chellat, M. F.; Raguz, L.; Riedl, R., Targeting antibiotic resistance. *Angew. Chem. Int. Ed.* **2016,** *55* (23), 6600-6626.

5. Selvaraj, A.; Valliammai, A.; Muthuramalingam, P.; Sethupathy, S.; Subramenium, G. A.; Ramesh, M.; Pandian, S. K., Proteomic and Systematic Functional profiling unveils citral targeting antibiotic

resistance, antioxidant defense, and biofilm-associated two-component systems of acinetobacter baumannii to encumber biofilm and virulence traits. *mSystems* **2020,** *5* (6).

6. Mor, A., Peptide-based antibiotics: A potential answer to raging antimicrobial resistance. *Drug Dev. Res.* **2000,** *50* (3-4), 440-447.

7. Laverty, G.; Gorman, S. P.; Gilmore, B. F., The potential of antimicrobial peptides as biocides. *Int. J. Mol. Sci.* **2011,** *12* (10), 6566-6596.

8. Di Somma, A.; Moretta, A.; Cane, C.; Cirillo, A.; Duilio, A., Antimicrobial and antibiofilm peptides. *Biomolecules* **2020,** *10* (4).

9. Shai, Y., Mode of action of membrane active antimicrobial peptides. *Biopolymers* **2002,** *66* (4), 236-248.

10. Kumar, P.; Kizhakkedathu, J. N.; Straus, S. K., Antimicrobial peptides: diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules* **2018,** *8* (1).

11. Malmsten, M., Interactions of antimicrobial peptides with bacterial membranes and membrane components. *Curr. Top. Med. Chem.* **2016,** *16* (1), 16-24.

12. Marcos, J. F.; Gandia, M., Antimicrobial peptides: to membranes and beyond. *Expert Opin. Drug Dis.* **2009,** *4* (6), 659-671.

13. Ciumac, D.; Gong, H. N.; Hu, X. Z.; Lu, J. R., Membrane targeting cationic antimicrobial peptides. *J. Colloid Interf. Sci.* **2019,** *537*, 163-185.

14. Hwang, P. M.; Vogel, H. J., Structure-function relationships of antimicrobial peptides. *Biochem. Cell Biol.* **1998,** *76* (2-3), 235-246.

15. Szwej, E.; O'Connor, S.; Byrne, A.; Devocelle, M.; Gallagher, W.; O'Connor, K., Enhancement of the anticancer activity of a cationic antimicrobial peptide: an investigation of mechanism and structure-function relationship. *J. Pept. Sci.* **2012,** *18*, S62-S63.

16. Fujii, K.; Young, M. T.; Harris, K. D. M., Exploiting powder X-ray diffraction for direct structure determination in structural biology: The P2X4 receptor trafficking motif YEQGL. *J. Struct. Biol.* **2011,** *174* (3), 461-467.

17. Spencer, R. K.; Nowick, J. S., A newcomers guide to peptide crystallography. *Isr. J. Chem.* **2015,** *55* (6-7), 698-710.

18. Strandberg, E.; Ulrich, A. S., NMR methods for studying membrane-active antimicrobial peptides. *Concepts Magn. Reson., Part A* **2004,** *23A* (2), 89-120.

19. Wang, G. S., Structural biology of antimicrobial peptides by NMR spectroscopy. *Curr. Org. Chem.* **2006,** *10* (5), 569-581.

20. Pusey, M. L.; Liu, Z. J.; Tempel, W.; Praissman, J.; Lin, D. W.; Wang, B. C.; Gavira, J. A.; Ng, J. D., Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol.* **2005,** *88* (3), 359-386.

21. Shi, Y. G., A Glimpse of structural biology through X-Ray crystallography. *Cell* **2014,** *159* (5), 995-1014.

22. Tormena, C. F., Conformational analysis of small molecules: NMR and quantum mechanics calculations. *Prog. Nucl. Mag. Res. Sp.* **2016,** *96*, 73-88.

23. Frueh, D. P., Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog. Nucl. Mag. Res. Spectrosc.* **2014,** *78*, 47-75.

24. Borics, A.; Murphy, R. F.; Lovas, S., Fourier transform vibrational circular dichroism as a decisive tool for conformational studies of peptides containing tyrosyl residues. *Biopolymers* **2003,** *72* (1), 21-24.

25. Keiderling, T. A.; Silva, R. A. G. D.; Yoder, G.; Dukor, R. K., Vibrational circular dichroism spectroscopy of selected oligopeptide conformations. *Bioorg. Med. Chem.* **1999,** *7* (1), 133-141.

26. Miyazawa, M.; Inouye, K.; Hayakawa, T.; Kyogoku, Y.; Sugeta, H., Vibrational circular dichroism of proline-containing oligopeptides. *Appl. Spectrosc.* **1996,** *50* (5), 644-648.

27. Petrovic, A. G.; Polavarapu, P. L.; Mahalakshmi, R.; Balaram, P., Characterization of folded conformations in a tetrapeptide containing two tryptophan residues by vibrational circular dichroism. *Chirality* **2009,** *21* (1e), E76-E85.

28. Wang, F.; Polavarapu, P. L., Conformational analysis of melittin in solution phase: Vibrational circular dichroism study. *Biopolymers* **2003,** *70* (4), 614-619.

29. Zhao, C. X.; Polavarapu, P. L.; Das, C.; Balaram, P., Vibrational circular dichroism of beta-hairpin peptides. *J. Am. Chem. Soc.* **2000,** *122* (34), 8228-8231.

30. Krupova, M.; Kessler, J.; Bour, P., Recent trends in chiroptical spectroscopy: Theory and applications of vibrational circular dichroism and Raman optical activity. *Chem. Plus. Chem.* **2020,** *85* (3), 561-575.

31. Polavarapu, P. L.; Zhao, C. X., Vibrational circular dichroism: A new spectroscopic tool for biomolecular structural determination. *Fresenius J. Anal. Chem.* **2000,** *366* (6-7), 727-734.

32. Polyanichko, A.; Wieser, H., Fourier transform infrared/vibrational circular dichroism spectroscopy as an informative tool for the investigation of large supramolecular complexes of biological macromolecules. *Biopolymers* **2005,** *78* (6), 329-339.

33. Polyanichko, A. M.; Andrushchenko, V.; Bour, P.; Wieser, H., Vibrational circular dichroism studies of biological macromolecules and their complexes. In *Circular Dichroism: Theory and Spectroscopy*, Rodgers, D. S., Ed. Nova Science Publishers, Inc.: New York, USA, 2012; pp 67-126.

34. Kurouski, D., Advances of vibrational circular dichroism (VCD) in bioanalytical chemistry. A review. *Anal. Chim. Acta.* **2017,** *990*, 54-66.

35. Keiderling, T. A., Protein and peptide secondary structure and conformational determination with vibrational circular dichroism. *Curr. Opin. Chem. Biol.* **2002,** *6* (5), 682-688.

36. Pancoska, P.; Yasui, S. C.; Keiderling, T. A., Statistical-analyses of the vibrational circular-dichroism of selected proteins and relationship to secondary structures. *Biochemistry* **1991,** *30* (20), 5089-5103.

37. Kim, J.; Kapitan, J.; Lakhani, A.; Bour, P.; Keiderling, T. A., Tight beta-turns in peptides. DFT-based study of infrared absorption and vibrational circular dichroism for various conformers including solvent effects. *Theor. Chem. Acc.* **2008,** *119* (1-3), 81-97.

38. Kessler, J.; Andrushchenko, V.; Kapitan, J.; Bour, P., Insight into vibrational circular dichroism of proteins by density functional modeling. *Phys. Chem. Chem. Phys.* **2018,** *20* (7), 4926-4935.

39. Pracht, P.; Bohle, F.; Grimme, S., Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020,** *22* (14), 7169-7192.

40. Eikås, K. D. R.; Krupova, M.; Kristoffersen, T.; Beerepoot, M. T. P.; Ruud, K., Can the absolute configuration of cyclic peptides be determined with vibrational circular dichroism? *SUBMITTED* **2022**.

41. Merrifield, R. B., Solid phase synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963,** *85*, 2149-2154.

42. Chan, W.; White, P., *Fmoc solid phase peptide synthesis: A practical approach*. Oxford University Press: Oxford, UK, 1999.

43. Eikås, K. D.; Beerepoot, M. T. P.; Ruud, K., A Computational protocol for vibrational circular dichroism spectra of cyclic oligopeptides. *J. Phys. Chem. A* **2022,** *126* (32), 5458-5471.

44. McLean, A. D.; Chandler, G. S., Contracted Gaussian-basis sets for molecular calculations .1. 2nd row atoms, Z=11-18. *J. Chem. Phys.* **1980,** *72* (10), 5639-5648.

45. Barone, V.; Cossi, M., Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998,** *102* (11), 1995-2001.

46. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V., Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003,** *24* (6), 669-681.

47. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenber, J. L.; Williams-

Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16, Revision B.01*, Gaussian, Inc.: Wallingfor CT, 2016.

48. Covington, C.; Polavarapu, P. *CDSpecTech: Computer programs for calculating similarity measures of comparison between experimental and calculated dissymetry factors and circular intensity differentials*, 2015.

49. *Schrödinger Release 2022-3: Maestro*, Schrödinger, LLC: New York, NY, 2021.

50. McHugh, S. M.; Rogers, J. R.; Yu, H. T.; Lin, Y. S., Insights into how cyclic peptides switch conformations. *J. Chem. Theory Comput.* **2016,** *12* (5), 2480-2488.

51. Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.

52. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

53. Pall, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E., Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *Lect. Notes Comput. Sc.* **2015,** *8759*, 3-27.

54. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

55. Jiang, F.; Zhou, C. Y.; Wu, Y. D., Residue-specific force field based on the protein coil library. RSFF1: modification of OPLS-AA/L. *J. Phys. Chem. B* **2014,** *118* (25), 6983-6998.

56. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001,** *105* (28), 6474-6487.

57. Jiang, F.; Geng, H., Computational methods for studying conformational behaviors of cyclic peptides. In *Cyclic peptide Design. Methods in Molecular Biology*, Goetz, G., Ed. Humana: New York, NY, 2001.

58. Hockney, R. W.; Goel, S. P.; Eastwood, J. W., Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974,** *14* (2), 148-158.

59. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1).

60. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

61. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577-8593.

62. Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J., An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **1996,** *9* (11), 1063-1065.

63. Rose, G. D.; Gierasch, L. M.; Smith, J. A., Turns in peptides and proteins. *Adv. Protein Chem.* **1985,** *37*, 1-109.

64. Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S., Automated molecular cluster growing for explicit solvation by efficient force field and tight binding methods. *J. Chem. Theory Comput.* **2022,** *18* (5), 3174-3189.

Supporting Information

Conformational Studies of Cyclic Hexapeptides with Vibrational Circular Dichroism: Theoretical and Experimental Approach

Karolina Di Remigio Eikås, Tone Kristoffersen, Mitchell Ramesh Silk, John Sigurd Mjøen Svendsen, Kenneth Ruud and Monika Krupová

CONTENTS:

Protocol for Synthesis of Investigated Cyclohexapeptides

Peptide synthesis:

Resin loading: 2-chlorotrityl chloride resin (0.15 mmol, 1.0 meq, 150 mg) was swelled in DCM (5 mL) for 30 min. The resin was drained and treated with a solution of Fmoc-AA-OH (0.3 mmol, 2 eq.) and diisopropylethylamine (1.8 mmol, 12 eq.) in DCM (5 mL). The resin mixture was left overnight under gentle agitation at room temperature. The resin mixture was drained, treated with MeOH (3 x 5 mL) to cap unreacted sites and dried with diethyl ether (3 x 5 mL).

Linear peptide synthesis: The linear peptides were prepared using an automated solid-phase peptide synthesizer (Biotage Initiator+ Alstra). The pre-loaded 2-chlorotrityl chloride resin was first swelled in DMF (20 min, 70°C). Fmoc deprotections involved treatment of the resin with 20% piperidine/DMF (4.5 mL, 3 min) once at room temp. followed by a second treatment at 70°C by microwave reactor. Amino acid couplings were performed by treating the resin with 4 eq. of Fmoc-AA-OH (0.5 M in DMF), 4 eq. of HOBt (0.5 M in DMF), 4 eq. of HBTU (0.6 M in DMF) and 8 eq. of DIEA (2M in NMP) for 5 min at 75°C by microwave reactor for Fmoc-Trp(Boc)-OH and Fmoc-Lys(Boc)-OH, or 60 min at room temp. for Fmoc-Arg(Pbf)-OH. After each Fmoc deprotection and AA coupling, the resin was washed with DMF (4 x 4.5 mL x 45 sec).

Resin cleavage: The resin-bound side-chain protected linear peptide was treated with 20% 1,1,1,3,3,3-Hexafluoro-2-propanol in DCM (2 x 5 mL x 15 min), followed by rinsing of the resin with DCM (5 mL). The filtrates were combined and concentrated under reduced pressure to yield the sidechain protected linear peptide.

Head-to-tail cyclisation and deprotection: A solution of the linear peptide (approx. 0.15 mmol) and diisopropylethylamine (0.9 mmol, 157 mL) in DMF (10 mL) was added to a solution of PyBOP (0.45 mmol, 234 mg) in DMF (100 mL) under light stir at room temp. for 1-2 hr. The mixture was concentrated by reduced pressure and treated with 95:2.5:2.5 TFA/triisopropylsilane/water (4 mL) then left to stand for 3 hr. The mixture was concentrated under N_2 gas flow followed by precipitation with ice-cold diethyl ether (15 mL). The precipitate was collected by filtration, washed with diethyl ether (15 mL), dissolved in 50% acetonitrile/water, and lyophilized to yield the crude, cyclic, sidechain deprotected peptide.

Purification: Peptides were purified by preparative reverse-phase HPLC (Waters 600 instrument with Waters 2487 Dual Absorbance detector) with a SunFire Prep. C18 OBD column (10 mm, 19 x 150 mm) using linear gradients of 0.1% TFA/water (buffer A) and 0.1% TFA/acetonitrile (buffer B) with a flow rate of 10 mL/min.

Analysis: Crude and final cyclic peptide products were analyzed by FT-MS (Thermo Scientific LTQ Orbitrap XL instrument) and by analytical reverse-phase HPLC (Waters 2795 Alliance HT system with Waters 2996 PDA Detector), using an Ascentis C18 column (3 mm, 3 x 100 mm) and solvents of 0.1% TFA/water (buffer A) and 0.1% TFA/acetonitrile (buffer B) with a linear gradient of 0-60% buffer B over 15 min and a flow rate of 0.5 mL/min.

NMR (Nuclear Magnetic Resonance) spectrometry: 1D ^{1}H NMR spectra of final cyclic peptide products were recorded at 1.6-2.0 mM in DMSO-d_6 using a Bruker Avance III 400 MHz spectrometer with cryoprobe, at 298K. NMR spectra were processed using TopSpin 4.0.9 (Bruker BioSpin GmbH).

Peptides:

1: Cyclo(WKWKWK). Purification gradient: 5-65% buffer B over 60 min, RT = 21 min (26% buffer B). Yield: 40.7 mg (41.8 %) as a white solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.82 (s, 1H), 8.13 (d, J = 6.0 Hz, 1H), 7.99 (d, J = 5.4 Hz, 1H), 7.64 (s, 3H), 7.57 (d, J = 7.8 Hz, 1H), 7.35 (d, J = 8.1 Hz, 1H), 7.16 (s, 1H), 7.08 (t, J = 7.4 Hz, 1H), 7.00 (t, J = 7.4 Hz, 1H), 4.37 (q, J = 6.7 Hz, 1H), 3.88 (q, J = 6.3 Hz, 1H), 3.27-3.12 (m, 5H), 2.55 (m, 3H), 1.57 (m, 3H), 1.36 (m, 3H), 0.96 (m, 3H). ESI-FTMS $[M + H]^{+}$ calculated: 943.5307, found: 943.5370, $[M + Na]^{+}$ calculated: 965.5126, found: 965.5176, $[M + 2H]^{2+}$ calculated: 472.2693, found: 472.2713.

2: Cyclo(WWWKKK). Purification gradient: 10-70% buffer B over 60 min, RT = 22 min (32% buffer B). Yield: 49.5 mg (39.0 %) as a white solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.83 (d, J = 1.8 Hz, 1H), 10.80 (d, J = 2.0 Hz, 1H), 10.80 (d, J = 1.9 Hz, 1H), 8.20 (d, J = 5.5 Hz, 2H), 8.06 (d, J = 7.0 Hz, 1H), 8.02 (d, J = 6.9 Hz, 1H), 7.99 (d, J = 8.1 Hz, 1H), 7.93 (d, J = 7.1 Hz, 1H), 7.72-7.64 (m, J = 6.5 Hz, 9H), 7.56 (d, J = 7.6 Hz, 1H), 7.56 (d, J = 7.4 Hz, 1H), 7.50 (d, J = 8.0 Hz, 1H), 7.35-7.33 (m, 3H), 7.10-7.01 (m, 6H), 7.0-6.95 (m, 3H), 4.30 (q, J = 6.9 Hz, 2H), 4.22 (q, J = 7.2 Hz, 1H), 4.11 (q, J = 7.0 Hz, 1H), 3.93 (q, J = 6.8 Hz, 1H), 3.86 (q, J = 7.1 Hz, 1H), 3.26 (m, 2H), 3.12-2.99 (m, 6H), 2.75 (q, J = 6.7 Hz, 2H), 2.61-2.56 (m, 2H), 1.78-1.69 (m, 4H), 1.57-1.28 (m,

12H), 1.16-1.08 (m, 3H), 0.97 (m, 1H). ESI-FTMS $[M + H]^+$ calculated: 943.5307, found: 943.5354, $[M + Na]^+$ calculated: 965.5126, found: 965.5159, $[M + 2H]^{2+}$ calculated: 472.2693, found: 472.2702.

3: Cyclo(WRWRWR). Purification gradient: 5-65% buffer B over 60 min, RT = 25 min (30% buffer B). Yield: 26.9 mg (22.6 %) as a white solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.78 (d, J = 1.7 Hz, 1H), 8.13 (d, J = 7.6 Hz, 1H), 8.10 (d, J = 7.1 Hz, 1H), 7.56 (d, J = 7.8 Hz, 1H), 7.45 (t, J = 5.5 Hz, 1H), 7.35 (d, J = 8.0 Hz, 1H), 7.16 (d, J = 2.1 Hz, 1H), 7.08 (t, J = 7.5 Hz, 1H), 7.00 (t, J = 7.5 Hz, 1H), 4.36 (q, J = 7.3 Hz, 1H), 3.94 (q, J = 7.0 Hz, 1H), 3.33-3.14 (m, 3H), 2.97 (q, J = 6.4 Hz, 2H), 1.68 (m, 1H), 1.56 (m, 1H), 1.26 (m, 2H). ESI-FTMS $[M + 2H]^{2+}$ calculated: 514.2785, found: 514.2779, $[M + 3H]^{3+}$ calculated: 343.1882, found: 343.1883.

4: Cyclo(WWWRRR). Purification gradient: 15-75% buffer B over 60 min, RT = 18 min (33% buffer B). Yield: 46.7 mg (30.0 %) as a white solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.80 (d, J = 1.9 Hz, 1H), 10.78 (d, J = 1.6 Hz, 1H), 10.74 (d, J = 1.6 Hz, 1H), 8.31 (d, J = 6.5 Hz, 1H), 8.21 (d, J = 6.5 Hz, 1H), 8.15 (d, J = 6.6 Hz, 1H), 8.07 (d, J = 6.9 Hz, 1H), 8.02 (d, J = 7.6 Hz, 1H), 7.94 (d, J = 6.8 Hz, 1H), 7.61 (t, J = 5.4 Hz, 1H), 7.56 (d, J = 7.9 Hz, 1H), 7.53-7.47 (m, 4H), 7.36-7.33 (m, 4H), 7.09-7.05 (m, 6H), 7.02-7.01 (m, 2H), 6.99-6.94 (m, 6H), 4.31-4.24 (m, 3H), 4.08 (q, J = 7.4 Hz, 1H), 4.01-3.91 (m, 2H), 3.30-3.25 (m, 2H), 3.12-3.10 (m, 9H), 2.99 (q, J = 6.6 Hz, 4H), 1.85 (m, 2H), 1.71 (m, 3H), 1.54 (m, 3H), 1.39 (m, 3H), 1.24 (m, 1H). ESI-FTMS $[M + H]^+$ calculated: 1027.5419, found: 1027.5473, $[M + 2H]^{2+}$ calculated: 514.2785, found: 514.2770, $[M + 3H]^{3+}$ calculated: 343.1882, found: 343.1872.

Interpretation of Low-Frequency IR Bands in the Amide I Region of Peptides 3 and 4

The experimental IR absorption spectra of arginine-containing peptides **3** and **4** have a complex broad band with a maximum at 1610 cm^{-1} in the amide I region that is not present in the spectra of peptides **1** and **2** containing Lys. As shown by a DFT (Density Functional Theory) calculation of the dominating CREST conformer in D_2O, this band originates in the stretching vibrations of the C-N bonds of the guanidino group on the Arg side chains (Figure S1). The frequencies of these modes are summarized in Table S1. Each of the three Arg side chains contributes with two modes: one slightly higher in frequency resulting in the bands at 1646 cm^{-1} and 1633 cm^{-1} for peptide **3** and **4**, respectively, and one slightly lower in frequency resulting in the bands at 1599 cm^{-1} and 1603 cm^{-1} for peptide **3** and **4**, respectively.

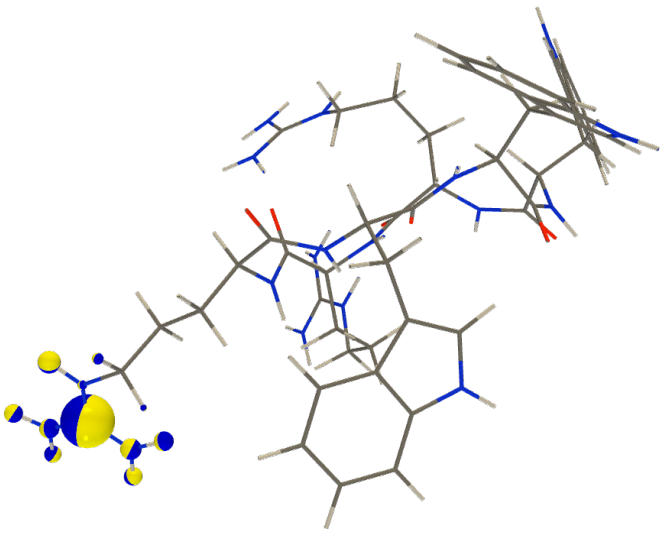


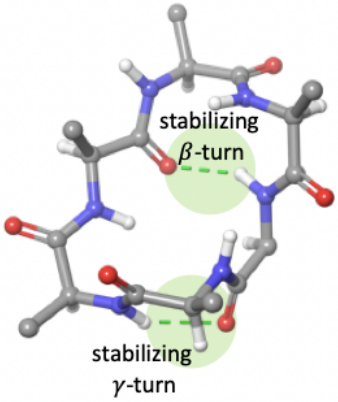
Figure S1. Illustration of C-N stretching vibrational modes in the Arg side chain resulting in the IR bands at 1610 cm^{-1} for peptide **4**. Similar C-H stretches are observed for all Arg side chains in peptide **4** and for peptide **3**.

Table S1. Calculated frequencies for the C-N stretches on the three Arg side chains of peptide **3** and **4** and their corresponding bands in experimental IR spectra.

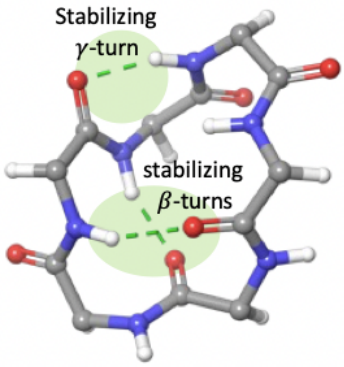
	Peptide 3		**Peptide 4**	
	Resulting IR band	Individual mode	Resulting IR band	Individual mode
Arg 1	1646 cm^{-1}	1648 cm^{-1}	1633 cm^{-1}	1637 cm^{-1}
Arg 2		1639 cm^{-1}		1632 cm^{-1}
Arg 3		1631 cm^{-1}		1622 cm^{-1}
Arg 1	1599 cm^{-1}	1598 cm^{-1}	1603 cm^{-1}	1596 cm^{-1}
Arg 2		1591 cm^{-1}		1606 cm^{-1}
Arg 3		1602 cm^{-1}		1601 cm^{-1}

Table S2. *g*-factors for experimental IR and VCD spectra of peptides **1** – **4** measured in D_2O. The values of *g*-factor were calculated as $\Delta A/A$, where ΔA is the amplitude of a VCD couplet and A is the absorption intensity of the corresponding IR band.

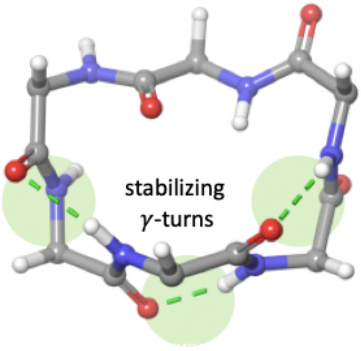
	Peptide 1	**Peptide 2**	**Peptide 3**	**Peptide 4**
***g*-factor**	5.54×10^{-5}	4.96×10^{-5}	5.52×10^{-5}	5.42×10^{-5}



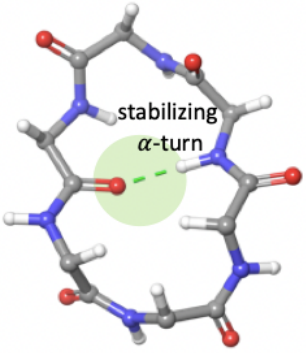
Peptide **1**
conformer with Boltzmann weight 82%



Peptide **2**
conformer with Boltzmann weight 82%



Peptide **3**
conformer with Boltzmann weight 67%



Peptide **4**
conformer with Boltzmann weight 92%

Figure S2. Backbone structures for dominating conformers of peptides **1** – **4** calculated by CREST. The dominating conformer of peptide **1** adopts a β- and a γ-turn, while the dominating conformer of peptide **2** adopts two β-turns and one γ-turn. For peptide **3**, the dominating conformer adopts three γ-turns while for peptide **4** the dominating conformer adopts an α-turn.

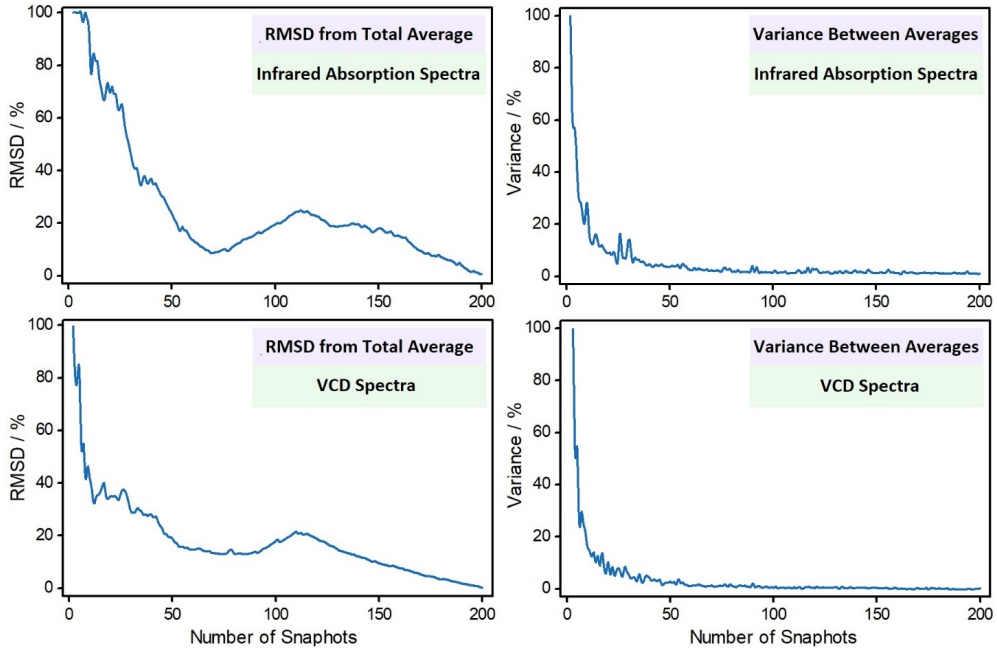


Figure S3. Convergence of the calculated IR absorption and VCD spectra on the number of averaged MD snapshots for peptide **1** in D_2O. Two ways of analysis are used: **left:** root-mean-square deviation (RMSD) of consecutive averages from a total average; **right:** variances C_n calculated as: $C_n = \frac{\int |f_n - f_{n-1}| d\omega}{\int |f_n| d\omega}$, where the cumulative spectra are $f_n = \frac{1}{n}\sum_{i=1}^{n} S_i(\omega)$; $S_i(\omega)$ is VCD intensity of the i^{th} spectrum at wavenumber ω.

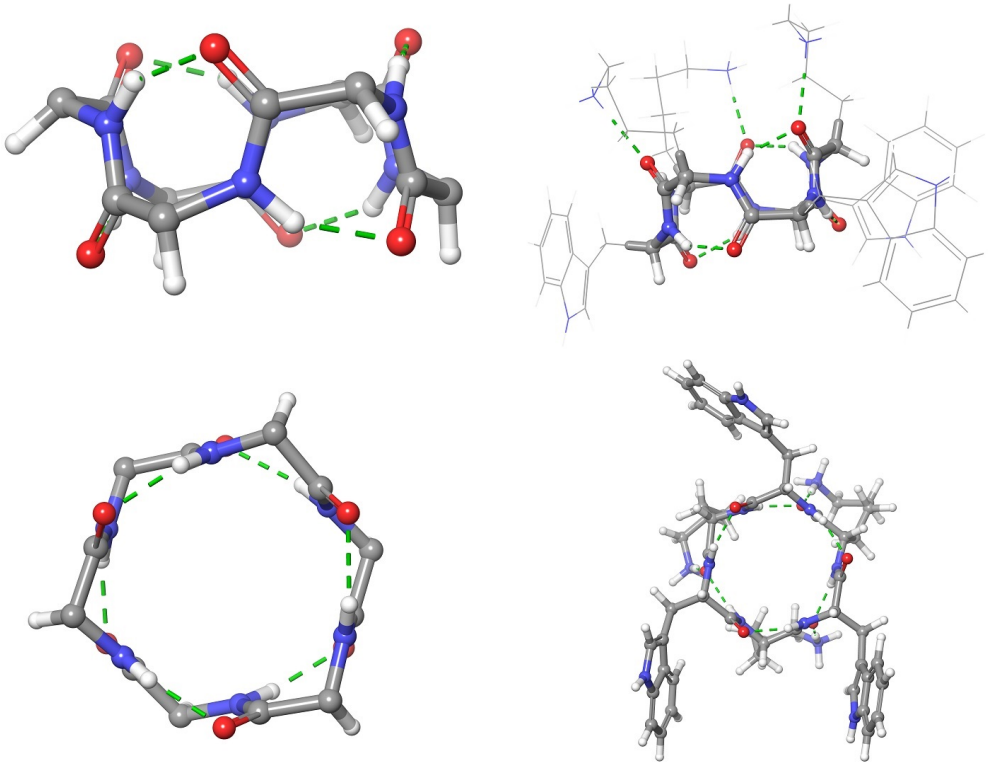


Figure S4. Structure of one of the CREST conformers for peptide **1** in water. It contains 6 turns in its backbone (3 γ^{cl} and 3 γ^{inv}), which makes it surprisingly symmetric. This particular conformer has a lower Boltzmann weight in the final spectra (5.2 %).

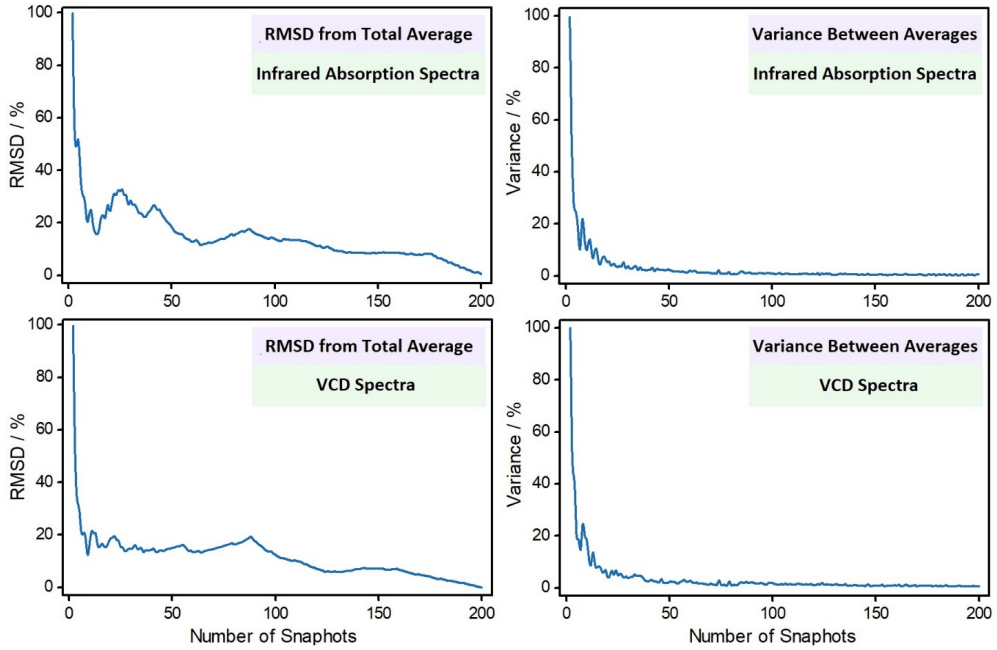


Figure S5. Convergence of the calculated IR absorption and VCD spectra on the number of averaged MD snapshots for peptide **1** in DMSO. Two ways of analysis are used: **left:** root-mean-square deviation (RMSD) of consecutive averages from a total average; **right:** variances C_n calculated as: $C_n = \frac{\int |f_n - f_{n-1}| d\omega}{\int |f_n| d\omega}$, where the cumulative spectra are $f_n = \frac{1}{n}\sum_{i=1}^{n} S_i(\omega)$; $S_i(\omega)$ is VCD intensity of the i^{th} spectrum at wavenumber ω.