FACULTY OF MATHEMATICS AND PHYSICS
Charles University

MASTER THESIS

Bc. Martin Vančura

Characterisation of polyproline I secondary structure by means of vibrational and chiroptical spectroscopy methods and quantum mechanical simulations

Institute of Physics of Charles University

Supervisor of the master thesis: RNDr. Václav Profant, Ph.D.

Study programme: Physics

Study branch: Biophysics and Chemical Physics

Prague 2020

Acknowledgments

In the first place, I would like to thank my supervisor Dr. Václav Profant (Institute of Physics of Charles University) for his guidance, teaching, patience and an excellent personal approach. I would also like to thank several people who helped us with the realization of the this thesis, namely: Prof. Petr Bouř (Institute of Organic Chemistry and Biochemistry of Academy of Sciences of the Czech Republic) for help with quantum chemistry calculations, Ing. Martin Šafařík ((Institute of Organic Chemistry and Biochemistry of Academy of Sciences of the Czech Republic) for the preparation of samples, Dr. Josef Kapitán (Palacký University Olomouc) for performing measurements on the device developed by him, Dr. Vladimír Kopecký, Jr. and Dr. Kateřina Hofbauerová for their help with experiments.

Last but not least, I would like to thank my parents for all their support. Special thanks to my mother for helping to correct grammatical errors in this thesis.

Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

We would like to thank project UNCE/SCI/010 (center for nano- and biophotonics) for financial support.

Title: Characterisation of polyproline I secondary structure by means of vibrational and chiroptical spectroscopy methods and quantum mechanical simulations

Author: Bc. Martin Vančura

Institute: Institute of Physics of Charles University

Supervisor: RNDr. Václav Profant, Ph.D., department

Abstract: Our investigation was focused on a secondary protein structure called polyproline I. This helical structure has been known for a long time, but its occurrence and significance in nature is not yet fully known. In this thesis, we use Raman spectroscopy and chiral sensitive Raman optical activity. These methods are sensitive to the structure of proteins but are more informative and sensitive to the local arrangement than the commonly used ECD and UV absorption. We were able to obtain polyproline I Raman and ROA spectra that have not yet been published. We have described important differences between the spectra of polyproline I and II and observed the process of mutarotation. The experimental part of the work is supplemented by quantum chemistry calculations of spectra using the transfer of molecular property tensor. The calculated spectra corresponded very well with the experimental spectra.

Keywords: polyproline I, mutarotation, raman optical activity, electronic circular dichroism, DFT simulations

Contents

Preface

Proteins and peptides are important compounds in living systems. They are linear polymers composed of amino acids, however, it is their higher level of organization which is vital for their biological functions. In this thesis we focus on peptides secondary structure. This structure can be described as local shape or folding of protein or peptide chain. Secondary structure is maintained by weak chemical forces between residues in local chain, interaction with solvents (water, organic, cell membrane etc.) or more distant parts of protein chain. Specific structure depends on environmental conditions and amino acid sequence. However, not only correct amino acid sequence and conditions ensure one (desired) specific conformation of protein, process of making particular molecule also plays a role. Therefore, the process of protein folding in living system is at least partially active and *in vitro* appropriate preparation and manipulation is important.

Typical secondary structure patterns are generally known and well described. These patterns represent as local periodic organization of protein chain resulting in a characteristic shape, e.g. helix. The best known examples of this secondary structure types are α-helix and β-sheet, which occur very often in natural proteins. The opposite type of structure is a random coil, which can be characterized by a pseudo-random non-periodic structure. In addition to these basic types, there is a number of other less common types of secondary structure. In this thesis we study polyproline I and II which are related to specific properties of amino acid proline. Polyproline I is right-handed helix preferred in organic solvents. In contrast, polyproline II is left-handed helix related to aqueous solution. Similar structural motives to polyproline II were described in many proteins. The occurrence and importance of polyproline I in living systems is still under investigation.

A wide variety of methods are used to study proteins. To study the structure of proteins, it is necessary to use methods that do not damage or modify their structure. Therefore, widely used methods such as chromatography or mass spectroscopy have only limited options. The most straightforward structural methods are X-ray diffraction and nuclear magnetic resonance. However, spectroscopic methods in UV, visible and infrared parts of spectra are also sensitive to (secondary and even tertiary in some cases) protein structure. A specific subgroup of spectroscopic methods are chirally sensitive (chiroptical) methods. Chirallity is property of molecule related to absence of mirror symmetry of molecule (same as right and left hand). These chiral molecules show different interaction with right- and left-handed circularly polarized radiation. The sensitivity of these methods to protein structure is related, among other things, to the fact that the different rotating helices are chiral.

One group of spectroscopic methods used for protein study is vibrational spectroscopy. Vibrational spectroscopy studies transition between vibrational states of molecule. Molecular vibrations are related mainly to movement of nuclei and are also dependent on higher level of organization of atoms in molecule. This makes vibrational spectroscopy methods useful in the study of proteins.

The main method used in this thesis is Raman scattering and its chiroptical variant called Raman optical activity (ROA). These methods use visible light as

3

incident radiation and also detected radiation scattered by sample is in visible part of spectrum. Detected energy difference between incident and scattered light lies in infrared part of spectrum, so the method belongs to vibrational spectroscopy. The use of visible light makes it possible to avoid some of the problems associated with the use of infrared radiation and also provides some kind of complementary information to infrared spectroscopy. We use also quantum chemistry calculation, which is necessary for interpretation and understanding of measured spectra. Circular dichroism is used as support method.

The main goal of this work is to obtain Raman and ROA spectra of polyproline I in different solvents, identify the characteristic spectral markers and assign them to the specific vibrations. This is challenging due to polyproline I instability in water, which is the best possible solvent for performing these measurements. Knowledge of this spectra could in principle give an option to detect the presence of this structure in proteins. We also observe mutarotation process from polyproline I to II and compare obtained spectra to expected chemical kinetics.

In the first chapter of the thesis called Introduction we provide brief overview of description of protein structure and some more detailed information about polyproline research. In Aims specific goals of this thesis are listed. In the third chapter, Theory, we summarize basics of theory necessary for description of vibrations of molecules and Raman scattering. We also provide some basics of quantum chemistry calculations. The fourth chapter Material and methods contains information about used instruments and programs (external and newly developed software). Mathematics used for description and analysis of chemical kinetics and spectral series is also provided in this chapter. Results of experiments, calculations and their discussion are in chapter Results and Discussions. Summation of obtained achievements is made in final chapter Conclusions.

1. Introduction

All known life forms and other related organic structures (such as viruses) share common principles and properties. On submicroscopic scale, all of these are made up of similar substances. The list of all types of these substances would be disproportionately long. However, most of the organic compounds can be divided into four main groups: proteins, carbohydrates, lipids, and nucleic acids. The focus of this work is a specific type of protein structure, so we will briefly introduce the importance of proteins for living systems.

Proteins and peptides are linear polymers composed of α-amino acids connected by peptide bond (a bond between carboxyl group and α-amino group). The term protein is mostly used for molecules with a chain longer than about 50 amino acids and the term peptide for shorter chains. For our purposes, however, this division is not essential and we will not make much distinction between these terms.

Peptide bond has specific properties. Formally it is usually depicted as a single bond but it is not accurate. Due to nitrogen lone electron pair of amid group, the peptide bond is stabilized by resonance and has character of partial double bond. As a result, peptide bond is practically planar.

The sequence of amino acids of protein produced by cell is encoded in DNA. The process of making protein consists of two important steps: transcription (rewriting to RNA template) and translation (making protein based on RNA template). However, the formation of a functional protein does not usually end there, it has to be folded correctly, further modified and delivered to the target destination. The complexity of these processes and their regulation are completely beyond the scope of this thesis.

Listing of functions of proteins in living systems is relatively easy – they play a role in virtually all types of processes in living systems. They have structural or building function (collagen, actin,...), transport function (haemoglobin, transferrin, membrane transporters...), signalling function (peptide hormones – e.g. insulin, receptor, ion channels...), stabilizing function (osmotic pressure, buffers), immunity function (antibodies, receptors), etc.

Enzymatic activity can be considered as a very specific function of proteins. Although proteins also need non-protein substances for many reactions and some reactions can be catalyzed by e.g. RNA, proteins are still dominant in this area. A change in just one amino acid may be sufficient to impair protein (enzymatic) function, and many inherited diseases are associated with such changes (phenylketonuria, different types of hemophilia etc.)

The importance of proteins is so crucial that they were assigned the role of genetic material before the discovery of the importance of DNA. However, the truth is that proteins also play a role in heredity, because histones play a role in epigenetic inheritance.

The size of proteins occurring in living systems varies dramatically. On the one hand, there peptides are made up of only a few amino acids. Good examples are angiotensin hormones, which play role in blood pressure and volume regulation, consisting of less than ten amino acid. On the other hand, proteins made up of thousands or tens of thousands of amino acids occur in nature. For example,

protein titin, which is part of muscle cells, is made up of approximately 30 000 amino acids.

Thus, it is clear that the study of the structure of proteins and the relations between their structural and functional properties is an important part of the research of biological systems.

1.1   Structure of proteins

Structure of proteins is described at four levels where the fourth level is optional for bigger proteins made up from more parts. Proteins may not consist of only one peptide chain. Many proteins are made from more than one chain and these chains typically form different structural parts of proteins called subunits or domains. A well-known example is haemoglobin – a heterotetramer composed of two pairs of identical units. In other proteins different protein chain can be interwined and form separate unit. This can be seen in collagen, which helix is formed from three different chains.

Primary protein structure describes sequence of amino acids used to build protein chain. As already mentioned, this order is preserved in genetic information and is modified by another reactions. The nature of the individual amino acids determines the preference of parts of the protein for the formation of higher structures and their hydrophilicity and hydrophobicity. Amino acid residues (parts outside the chain of peptide bonds) can also be modified in various ways. Small functional groups such as hydroxyls or methyls may be attached resulting in amino acids properties alteration. Amino acids (with the exception of glycine) are chiral, but in natural proteins they occur practically only in the L conformation, below we will consider only this case.

Secondary protein structure is a local spatial conformation of peptide chain. This type of structure is mainly stabilized by interaction of amid hydrogenes and carbonyl oxygen (not a case of polyproline helix) or interaction with environment (solvent). Important structural motives has their own names. The most common helical structure is the α-helix, but there are also other types – e.g. 3_{10}-, π-helix or polyproline helices. Second common structural motif is β-pleated sheet. This structure interacts with more distant parts of the chain in the same conformation (based on orientation of chains it is parallel or antiparallel). Last group of stable secondary structures are nonrepetitive coils or loops. These structures are stable in protein but it is not possible to describe them as some repetitive structure.

Tertiary protein structure is a spatial organization of secondary structure components. This structure is stabilized mainly by weak interactions among amino acids residues and also by interaction with solvent. Some proteins are also stabilized by covalent disulfide bonds between cysteine residues. Compounds different from proteins such as metal atoms or cofactors may also play a role in tertiary structure.

Quaternary structure describes organization of domains from tertiary structure. Different domains in protein may have its unique function – they can fix the protein in membrane, have specific enzymatic activity, play a role in allosteric regulation of enzymatic complex etc. Regulation of protein domains interaction by phosphorylation is an important form of regulation of metabolic pathways.[52]
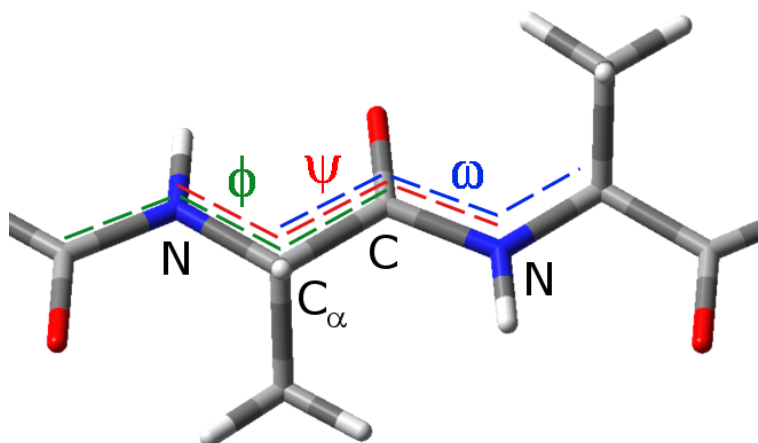
Figure 1.1: Dihedral angles in peptide chains used for specification of secondary structure.

1.2 Description of secondary structure

Secondary structure can be easily described by values of several dihedral angles of the main chain. Dihedral angle is defined by four points in space (no group of three lies on a straight line). It is defined as angle between plane determined by first three points (atoms in peptide chain) and plane defined by second, third, and fourth atom positions. Planar angles (between bonds) and bond length are well conserved in the first approximation.

Three types of dihedral angles (and three types of bonds) can be found in the peptide chain: ψ is the angle between carbonyl carbon and C_α, ϕ between atom C_α and α-amino nitrogen, and ω between α-amino nitrogen and carbonyl carbon (Fig. 1.1). In general, only the ψ and ϕ angles are used to describe the secondary structure, whose combination is characteristic of the upper structure. Omission of ω is possible because the peptide bond is almost always in the *trans* conformation. Examples of characteristic dihedral angles for different secondary structures can be found in Tab. 1.1.

Of course, only some combinations of these angles form a meaningful spatial structure. The graph showing the combination of ϕ and ψ with marked areas that are energetically and geometrically permissible is called the Ramachandran diagram. Exact shapes of suitable areas depend on calculation and used conditions. Example of this plot can be found in Fig. 1.2.

An important note is that although polyproline I and II are very close, the main difference between them lies in the value of angle ω (*cis/trans* configuration), which Ramachandran diagram does not take into account. It is also worth noting that polyproline I and II are close to the structure occupied by collagen.

1.3 Proline and its conformation

Proline is one of the basic amino acids used in synthesis of proteins. It's found as L-epimer in nature, similarly as other amino acids. However, proline is different from other amino acids because the last carbon atom in its residuum makes bond

Table 1.1: Examples of typical values of ϕ and ψ dihedral angles for different types of secondary structure.[52]

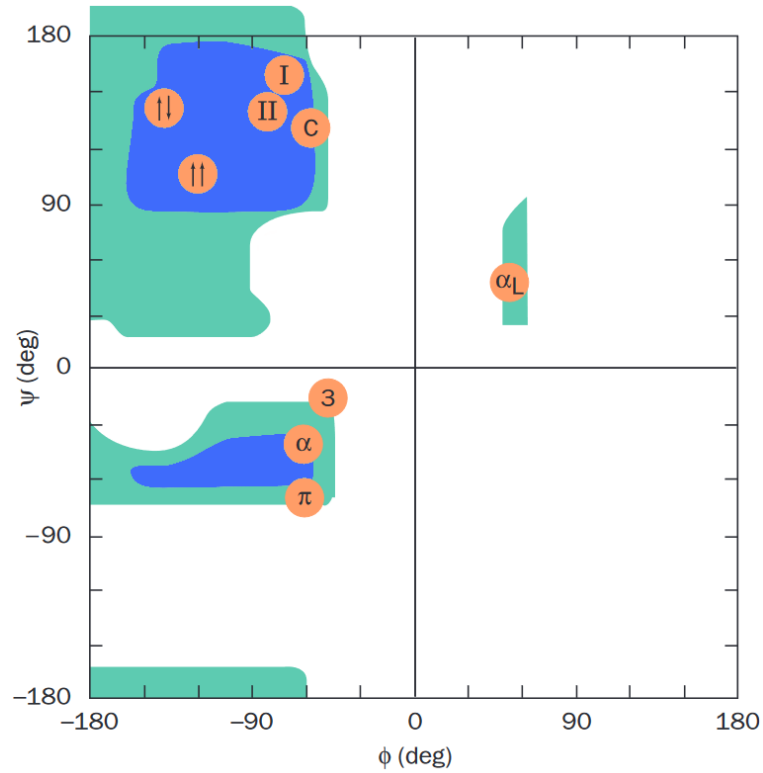| Secondary structure type	ϕ (°)	ψ (°)
Right-handed α-helix (α)	-57	-47
Parralel β pleated sheet ($\uparrow\uparrow$)	-119	113
Antiparralel β pleated sheet ($\uparrow\downarrow$)	-139	135
Right-handed 3_{10} helix (3)	-49	-26
Right-handed π-helix (π)	-57	-70
Polyproline I (I)	-75	160
Polyproline II (II)	-79	150
Collagen (C)	-51	153
Left-handed α-helix (α_L)	57	47

Figure 1.2: Ramachandran diagram. Areas that suit the conditions well are colored blue, green areas correspond to the most benevolent conditions. The legend to protein secondary structure abbreviations is given in Tab. 1.1. Image was adjusted according to [52].

with the nitrogen atom in amino group, so in fact it's imino acid. Therefore, the nitrogen atom of prolin can not provide hydrogen for hydrogen bond. However, carbonyl group can be an acceptor of hydrogen bond.[38] This additional bond also restricts possible range of phi (ϕ) angle. Furthermore, amino acids adopt mainly *trans* configuration around the peptide bond. Proline can easily adopt *cis* conformation as well (rotational barrier is around 20 kcal/mol[25]). These special properties cause proline lower occurrence in α-helix and β-sheet. It's more frequent in their endings, where prolines role is probably to disrupt the structure and end or start change conformation of pattern.

Pyrrolidin ring in proline can hold two stable conformations marked as A and B (Fig. 1.3).[43] This conformational difference from planar structure is called puckering. Simplified parameters phase angle P and puckering amplitude θ_m can be used for description of this deviation from plane. We have equations for this parameters:

$$\tan P = \frac{(\theta_2 - \theta_4) - (\theta_1 + \theta_3)}{2\theta_0(\sin \frac{\pi}{5} + \sin \frac{2\pi}{5})}, \tag{1.1}$$

$$\theta_0 = \theta_m \cos P, \tag{1.2}$$

where θ_i are torsion angles of bonds described in Fig. 1.3. For inversion of these equation it is important, that when θ is negative, 180° must be added to P.[3,54] We used $P = 170°$ as border for A ($< 170°$) and B ($> 170°$) conformation.
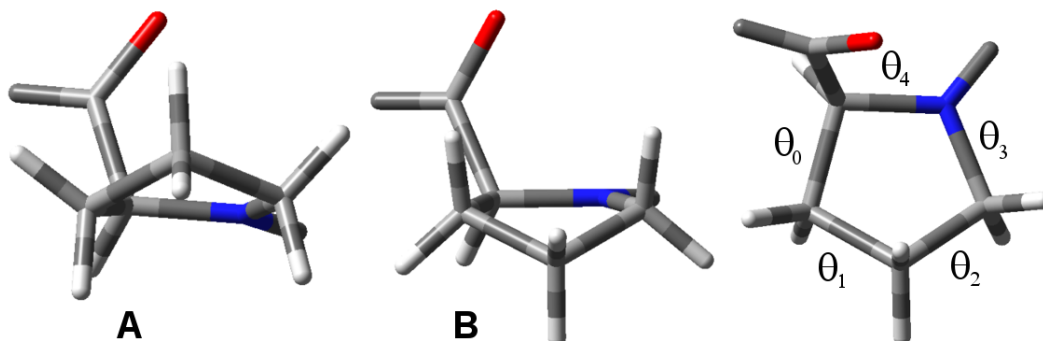
Figure 1.3: Proline A and B conformation and definitions of θ dihedral angles used for their description.

1.4 Polyproline I and II

Historically, the studies of poly-L-proline have been linked to the studies of collagen, which is a protein rich in proline. The structure of collagen was intensively investigated in the early 1950s, as evidenced by three articles in the journal Nature on this topic over three months.[17,42,44]

Proline was considered a possible model structure for the study of collagen. In X-ray diffraction experiments, the structures produced by the separate poly-L-proline chains and the structures observed in collagen were shown to partialy match.[17]

It has been found that poly-L-proline may (depending on the solvent) be present in two different conformations (secondary protein structure). Both of

these conformations are helical – right-handed polyproline I (PPI) and left-handed polyproline II (PPII). PPI has more compact structure and peptide bond occupies *cis* configuration. PPII is less compact and peptide bond occupies *trans* conformation. Visualisation of these structure is shown in Fig. 1.4[45] [50] Prior to obtaining particular chain structures, differences in optical activity were observed in the various solvents. It has also been found that these two conformations transform into one another as the solvent changes.[29] [48]

In the 1970s and 1980s, polyproline II helix was found to occur in other proteins and more interestingly not only in proline-rich regions but also in regions where there is no proline at all. It was found that the occurrence of polyproline is not only related to fibrous proteins, but it also occurs in globular proteins and is comparable to beta helix.[1,32,33] A later statistical study on Protein Data Bank[10] showed that approximately 2% of protein chains correspond to PPII.[9]

It has been also discussed that PPII may play a role in pathophysiology of amyloid-related diseases.[11] During these processes, the protein undergoes a conformational change and insoluble cell damaging aggregates are formed. α-helix was expected to be converted into a β-sheet during this process. However, it has been shown that the techniques used may not be sensitive enough to distinguish the exact conformation. Using ROA and UV Raman spectroscopy, it was found that PPII could actually play an important role in the aggregation process.[2] [4] [11]
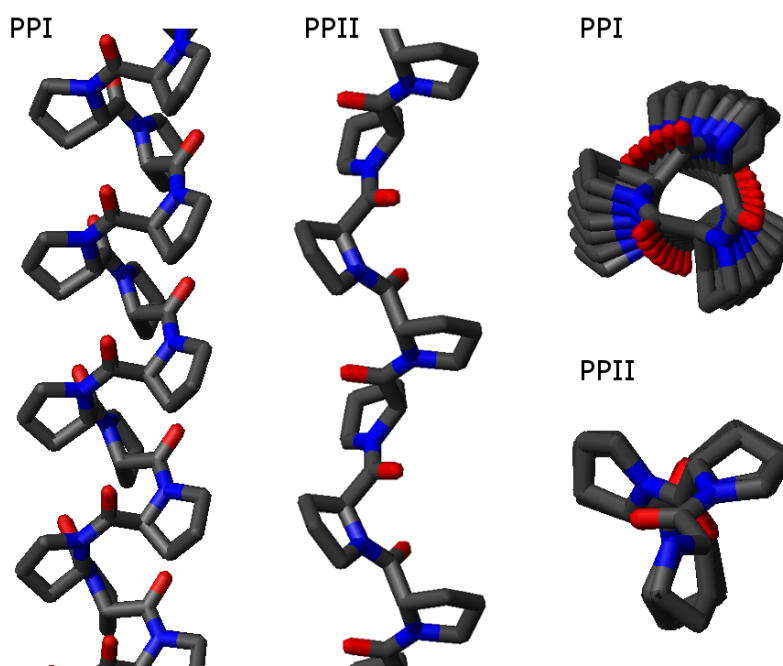

Figure 1.4: Structures of PPI and PPII, top and side view. (Public domain pictures, adjusted)

1.5 Structural and spectroscopic characterization of polyproline

First polyproline studies were based on X-ray structure analysis. The dihedral angles (ϕ, ψ, ω) determined by this method were for PPI approximately $(-75°, 160°, 0°)$ with approximately 3.3 residues per turn and for PPII $(-75°, 145°, 180°)$ with approximately 3 residues per turn.[16] [23] [50] However, the actual angles of the protein in a solution may differ from these values. For PPII, protein database analyzes were performed (structures also obtained with X-ray), showing that the maxima of the clusters in the Ramachandran plot (ϕ, ψ) corresponding to PPII are within $\pm 10°$ of these values.[2]

As mentioned before, PPI may be converted into PPII in a suitable solvent and vice versa. This transition is slow (half-life in hours or more) thanks to high energy barrier of *cis/trans*-transition. This transition was studied for different length, temperature and solvents. Using CD, it was shown for short oligopeptids (4-13 AA) that PPII stability in water and triflouroethanol decreases with shorter chain and that PPI stability in 1-propanol and methanol increases with the lenght of peptide chain. The energy difference between PPI and PPII ranged from -0.20 to 1.04 kcal/mol per residuum[23]

The study of temperature-induced transition performed on a 12 proline long peptide determined the activation energy for PPI to PPII transition in 1-propanol as $(81 \pm 4)\,\mathrm{kJ/mol^{-1}}$. It was shown that the transition is slow at temperatures close to 0° (reaction rate constant $8 \times 10^{-7}\,\mathrm{s^{-1}}$), so the samples can be stored at low temperatures for long time. The transition in 1-propanol was also determined as endothermic.[28]

The mutarotation process from PPI to PPII in water and heavy water $(\mathrm{D_2O})$ was studied with use of ECD, VCD and FTIR. Samples had molecular weight 5500-19 000 g/mol. The sample converted to PPI was dissolved in chilled solvent with concentration 10-40 mg/ml. Solution was opalescent but cleared during mutarotation process. Obtained spectra did not depend on concentration and time of preparation. Conformation were distinguishable in both VCD and IR. The authors concluded from obtained spectra that mutarotation occurs via an intermediate state with random distribution of *cis* and *trans* bonds.[18]

Intermediate states of mutarotation process were studied in more detail using Ion mobility spectroscopy–mass spectroscopy and molecular dynamics. The 13 prolines long oligopeptide was used as a sample. The study identified four different intermediate states and three possible final states (PPII or similar), the process of mutarotation is expected to start from N-terminus.[47]

ROA and Raman spectrosopy was proven to be sensitive to the length of PPII chain. ROA was also sensitive to proline puckering. The ratio A:B for polyproline peptides was determined as 6:4.[41] This is consistent with calculations showing that puckering A may be favored in short PPII oligopeptides. This article expect B puckering to be preferred in short PPI oligopeptides.[26] Other structural studies based on NMR, X-ray, ROA, and Raman experiments expect A:B ratio approximately 1:1 ratio.[27] [24] [51]

Nuclear magnetic resonance was also used to study proline-containing peptides. It was used to prove conformational stability of short proline oligopeptides and it was also demonstrated that proline *cis/trans* isomeration may be useful as a

conformational probe.[19] [55] Polyproline was also used for single-molecule Förster resonance energy transfer experiment and it was shown that longer peptides (40 AA approximately) are flexible.[46]

2. Aims

This thesis builds on previously performed unpublished experiments with polyproline I to II mutarotation process Raman scattering measurement. Our goal is to perform further measurements with better defined conditions and with better sensitivity. We also want to calculate and theoretically predict spectra. Our individual goals are:

1. Preparation of polyproline I from commercial polyproline II samples according to literature, confirmation of correct preparation by ECD, and isolation of polyproline I in solid phase.

2. Obtaining Raman spectra of polyproline I and II in solid phase using Raman microscope and identification of important spectral differences.

3. Study of kinetics of mutarotation process from polyproline I to II using ECD with temperature stabilization.

4. Measurement of Raman spectra in organic solvents and measurement of Raman and ROA spectra during the mutarotation in water. Getting Raman and ROA spectra of polyproline I in water.

5. Computational prediction of polyproline I and II spatial conformation with dependency on puckering. Preparation of geometry of longer polyproline molecules comparable to experimental samples based on our calculations.

6. Calculation of Raman and ROA spectra for prepared molecule with use of separation these larger molecules into smaller fragments, calculation in normal coordinates and transfer of molecular property tensor back to original molecule. Discussion of length and puckering dependency.

7. Comparison of our experimental and theoretical results, and their discussion.

3. Theory

Raman scattering is an inelastic two photon scattering. At first we describe general theory of molecular vibration with more details for Raman scattering and Raman optical activity. Then we give some basic description of theoretical calculations used for prediction of spectra for larger biological systems. At the end of the chapter we will also describe the fundamental principles of circular dichroism.

3.1 Vibrational spectroscopy

In this section, we briefly describe the fundamentals of the theory necessary to describe and interpret the methods of vibrational spectroscopy. The basics of theory we present in accordance with any advanced university courses focusing on quantum chemistry and quantum molecular theory as well as textbooks related to the subject [37] [49]. For a more detailed explanation and derivation readers are advised to the books.

3.1.1 Born-Oppenheimer approximation

The mass of nuclei is three order of magnitude grater than those of electrons, so they are well localized compared to electrons. Thanks to this, they can be understood to a certain extent classically. This can be mathematically formulated as a Born-Oppenheimer approximation.

Firstly, we assume a wave function dependent only on the positions of electrons $\overrightarrow{r}$ and nuclei $\overrightarrow{R}$, neglect the influence of spins and external fields. So we can write the wave function as $\Psi(\overrightarrow{r}, \overrightarrow{R})$. Secondly, we assume that the movement of electrons and nuclei can be separated and the position of nuclei can be considered constant (parametric) for the wave function of electrons.

Hamiltonian $\hat{H}$ can be divided into electron-dependent term H_E and kinetic term of nuclei T_N. The nuclear repulsion energy V_{NN} is included in electron part. From an electron perspective, the repulsive energy of nuclei for a given configuration is constant. We can write:

$$\hat{H} = T_E + T_N + V_{NN} + V_{EN} + V_{EE} = H_E + T_N. \tag{3.1}$$

This approximation is called adiabatic, so the superscript A is added to the wave function. We also assign quantum numbers e and v for vibrational and electron state of molecule. We can write Schrödinger's equation as:

$$\hat{H}\Psi_{ev}^A(\overrightarrow{r}, \overrightarrow{R}) = (H_E + T_N)\psi_e^A(\overrightarrow{r}, \overrightarrow{R})\phi_{ev}(\overrightarrow{R}) = E_{ev}^A\psi_e^A(\overrightarrow{r}, \overrightarrow{R})\phi_{ev}(\overrightarrow{R}), \tag{3.2}$$

where $\psi_e^A(\overrightarrow{r}, \overrightarrow{R})$ and $\phi_{ev}(\overrightarrow{R})$ are separated electron and nucleus parts of the wave function. Since the nuclear part does not have the position of electrons as an argument, we can solve separately the equation:

$$H_E\psi_e^A(\overrightarrow{r}), \overrightarrow{R} = E_e^A\psi_e^A(\overrightarrow{r}, \overrightarrow{R}). \tag{3.3}$$

In Born-Oppenheimer approximation we further assume that the term resulting from operation of kinetic energy of nuclei operator on the electronic wave function can be neglected. So we get the equation:

$$H_N \phi_{ev}(\overrightarrow{R}) = (T_N + E_e^A(\overrightarrow{R}))\phi_{ev} = E_{ev}^A \phi_{ev}. \qquad (3.4)$$

By solving Schrödinger's equation for electrons (3.3) and finding sets of nuclear positions $\overrightarrow{R}$ corresponding to local minima of energy surface we find stable conformation of molecules. Energy of this nuclear and electron state is referred as $E_g^A(\overrightarrow{R})$, where g marks ground electronic state. Subsequently, vibrational sublevels of this state can be found from Equation (3.4). Dependence of energy $E_g^A(\overrightarrow{R})$ on nuclei positions $\overrightarrow{R}$ can be expressed using Taylor's series as:

$$E_g^A(\overrightarrow{R}) = E_g^A(\overrightarrow{R}_0) + \frac{1}{2} \sum_{J,J'=1}^{3N} \left(\frac{\partial^2 E_g^A}{\partial \overrightarrow{R}_J \partial \overrightarrow{R}_{J'}} \right) \Delta \overrightarrow{R}_J \Delta \overrightarrow{R}_{J'} + \ldots, \qquad (3.5)$$

where $\Delta \overrightarrow{R}_J$ is displacement of the equilibrium position. The first order term in series is skipped because first derivatives are equal to zero in local minimum.

Born-Oppenheimer approximation is not able to explain the phenomena of vibrational circular dichroism (VCD), where including of dependence of electron density on movements of nuclei is required. For the description of VCD it is necessary to use an approach considering this interaction, which is called complete adiabatic approximation

3.1.2 Vibrational modes

As we discussed in the Born-Openheimer approximation, nuclei are treated in classical way. When studying the vibrational freedom of a molecule, we consider only the positions of the nuclei $\overrightarrow{R}$. After subtracting degrees of freedom of center of mass and rotation, we get $3N - 6$ for non-linear molecule and $6N - 5$ for linear molecule.

Considering vibrations as harmonic (harmonic approximation), we can neglect terms higher than second order in Equation (3.5). Then we can rewrite the kinetic and potential term in Equation (3.4) for ground vibrational state ϕ_{gv} as:

$$\left[\sum_{J=1}^{N} \frac{1}{2} M_J \dot{\overrightarrow{R}}_J^2 + \frac{1}{2} \sum_{J,J'=1}^{3N} \left(\frac{\partial^2 E_g^A}{\partial \overrightarrow{R}_J \partial \overrightarrow{R}_{J'}} \right) \Delta \overrightarrow{R}_J \Delta \overrightarrow{R}_{J'} \right] \phi_{gv}(\overrightarrow{R}) = E_{gv}^A \phi_{gv}(\overrightarrow{R}), \quad (3.6)$$

where M_J and $\dot{\overrightarrow{R}}_J$ stands for masses and velocities.

Diagonalising the matrix of second derivatives of E_g^A, we get $3N - 6$ non-zero (positive) numbers $k_{g,a}$ on the diagonal and corresponding $3N - 6$ coordinates Q_a called normal coordinates. $k_{g,a}$ has meaning of force constant of normal mode a. As a result, we separate the problem into $3N - 6$ independent systems. So for every normal mode we obtain energy $E_{gv,a}^A$ and independent equation of harmonic oscillator:

$$\left[\frac{1}{2} \dot{Q}_a^2 + \frac{1}{2} k_{g,a} Q_a^2 \right] \phi_{gv}(Q_a) = E_{gv,a}^A \phi_{gv}(Q_a). \qquad (3.7)$$

Solution of this equation is:

$$E_{gv,a}^A = (v_a + \frac{1}{2})\hbar\omega_a, \tag{3.8}$$

where $\omega_a = \sqrt{k_{g,a}}$ and v_a is index of energy level of normal mode a.

3.1.3 Infrared absorption

The transitions between the vibrational states of a molecule correspond with their energy to the infrared region of electromagnetic radiation. The most straightforward approach is to measure absorption spectra, similar to electron transitions in the visible area. This approach is called infrared spectroscopy (typically referred to as IR).

Instead of wavelength, wavenumber ν is used in absorption spectra. Typical unit used in IR is inverse centimeter cm^{-1}.

Intensity of transition depends on dipole strength of transition D_a given by formula:

$$D_a = \left| \left(\frac{\partial \overrightarrow{\mu}}{\partial Q_a} \right)_{Q_a=0} Q_a \right|^2, \tag{3.9}$$

where $\overrightarrow{\mu}$ stands for electric dipole moment of molecule. The shape of the spectral line is given by the spectral distribution function $f_a'(\nu)$.

Molar attenuation coefficient $\epsilon(\nu)$ for given molecule and wavenumber is a sum of contributions of individual vibration modes:

$$\epsilon(\nu) = \frac{8\pi^3 N\nu}{3000hc\ln(10)} \sum_a D_a f_a'(\nu), \tag{3.10}$$

where the expression before the sum corresponds to the unit conversion of dipole strength, logarithm is there for transition from natural to decadic.

Absorption $A(\nu)$ and intensity $I(\nu)$ for wavenumber ν is given by Beer-Lambert law:

$$A(\nu) = \epsilon(\nu)cl, \quad I(\nu) = I_0(\nu)10^{-A(\nu)}, \tag{3.11}$$

where c stands for molar concentration, l for length of beam in sample and I_0 beam intensity before interaction with sample.

Two main experimental approaches are used to obtain spectra. The first is the direct measurement of absorption coefficients with a similar arrangement as for the visible region. The system consists of a source, mirrors, monochromator lenses and a detector. Second approach is called Furier transform infrared spectroscopy (FTIR).

3.1.4 Raman scattering

Raman scattering is two-photon interaction between the molecule and the radiation field. Photon of incident radiation is scattered with loss or gain of some energy. This is in contrast to Rayleigh scattering, where both the excitation photon and the scattered photon have the same energy. The scattering in which the scattered radiation photon has lower energy is called Stokes, the one in which the

scattered photon energy is higher is called anti-Stockes. Schema of these scatterings is shown in Figure 3.1. Compared to the elastic (Rayleigh) scattering, the inelastic (Raman) scattering has rather low cross section (10^{-6}, only one photon per million is scatter inelastically) The relative intesity of Stokes/anti-Stokes part of Raman scattering is given by occupation of vibrational energy levels which derives from the Maxwell-Boltzmann distribution. The energy difference between incoming and outgoing photons equals the difference of two vibrational states and is independent of the wavelength of the excitation radiation.
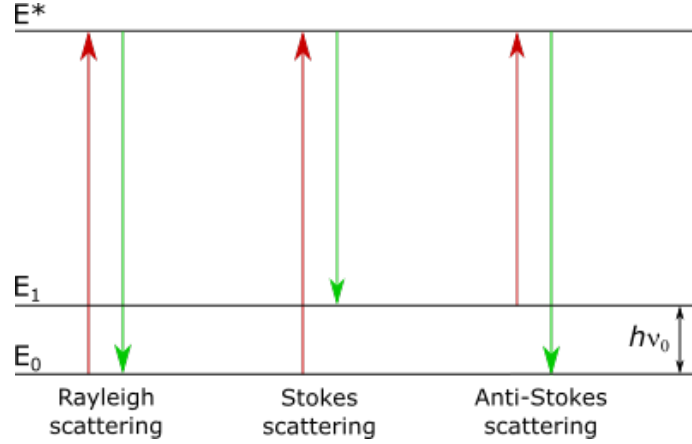

Figure 3.1: Comparison of scatterings. E_0 is energy of ground state of molecule, E_1 energy of higher vibrational state and E^* energy of virtual electron state. Excitation photon is shown as red arrow, scattered photon as green one.

At the classical level, the intensity of Raman scattering $I(\omega_0, \omega_s)$ for incident electric field $E_\beta^{(0)}(t)$ with angular frequency ω_0 is function of induced electric dipole moment $\mu_\alpha^I(t)$, which radiates scattered light with scattered angular frequency ω_s:

$$I(\omega_0, \omega_s) = \frac{1}{32\pi^2 \epsilon_0 c^3} \cdot \omega_s^4 (\mu_\alpha^I)^2 \sin^2 \theta, \tag{3.12}$$

where ϵ_0 is dielectric constant of free space, c is the speed of light and θ is direction of radiation with respect to the axis of the induced dipole. The induced electric dipole depends on the exciting electric field as:

$$\mu_\alpha^I(t) = \sum_\beta \alpha_{\alpha\beta} E_\beta^{(0)}(t), \tag{3.13}$$

where $\alpha_{\alpha\beta}$ is the polarizability response of the molecule.

Using incident electric field $E_\beta^{(0)}(t) = E_{\beta,0}^{(0)} \cos \omega_0 t$ and Taylor series we obtain for oscillation frequency ω_a of normal mode Q_a:

$$\mu_\alpha^I(t) = \alpha_{\alpha\beta,0} E_{\beta,0}^{(0)} \cos(\omega_0 t) + \frac{1}{2} \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_\alpha} \right) Q_\alpha E_{\beta,0}^{(0)} [\cos(\omega_0 - \omega_a)t + cos(\omega_0 + \omega_a)t], \tag{3.14}$$

where $\omega_s = \omega_0 - \omega_a$ represents Stokes part and $\omega_s = \omega_0 + \omega_a$ anti-Stokes part.

Important consequence of the last equation is that normal modes with:

$$\left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_\alpha} \right) = 0, \tag{3.15}$$

17

have intensity of Raman scattering equal to zero and therefore are not observable.

In the quantum description, the equation (3.12) goes to form (Einstein notation used):

$$I(\tilde{e}^d, \tilde{e}^i) = 90K \left\langle \left| \tilde{e}_\alpha^{d*} \tilde{\alpha}_{\alpha\beta} \tilde{e}_\beta^i \right| \right\rangle, \tag{3.16}$$

where tildes are used for denoting of complex quantities, $\tilde{e}_\alpha^i$ and $\tilde{e}_\beta^d$ represent unity polarization vectors of incident and scattered electric field in Cartesian coordinates (asterisk for complex conjugate), and constant K is defined as:

$$K = \frac{1}{90} \left(\frac{\omega_R^2 \mu_0 \tilde{E}^{(0)}}{4\pi R} \right)^2, \tag{3.17}$$

where ω_R is frequency of scattered electric field, $E^{(0)}$ intensity of incident electric field, μ_0 is magnetic permeability of free space, and R is detector distance from sample.

It can be shown that for liquid samples (perfect spatial averaging), the only observable quantities are invariants of tensor $\tilde{\alpha}_{\alpha\beta}$ (or their combinations) defined as:

$$\alpha^2 = \frac{1}{9} \Re \left[(\tilde{\alpha}_{\alpha\alpha})^S (\alpha_{\beta\beta})^{S*} \right], \tag{3.18}$$

$$\beta_S(\tilde{\alpha})^2 = \frac{1}{2} \Re \left[3(\tilde{\alpha}_{\alpha\beta})^S (\tilde{\alpha}_{\alpha\beta})^{S*} - (\tilde{\alpha}_{\alpha\alpha})^S (\tilde{\alpha}_{\beta\beta})^{S*} \right], \tag{3.19}$$

$$\beta_A(\tilde{\alpha})^2 = \frac{1}{2} \Re \left[3(\tilde{\alpha}_{\alpha\beta})^A (\tilde{\alpha}_{\alpha\beta})^{A*} \right], \tag{3.20}$$

where α^2 is called isotropic invariant, $\beta_S(\tilde{\alpha})^2$ the symmetric anisotropy, and $\beta_A(\tilde{\alpha})^2$ the anti-symmetric anisotropy. Symmetric and anti-symmetric combinations of tensor elements used in equations are defined as:

$$\left(\tilde{T}_{\alpha\beta} \right)^S = \frac{1}{2} \left[\left(\tilde{T}_{\alpha\beta} \right) + \left(\tilde{T}_{\beta\alpha} \right) \right], \tag{3.21}$$

$$\left(\tilde{T}_{\alpha\beta} \right)^A = \frac{1}{2} \left[\left(\tilde{T}_{\alpha\beta} \right) - \left(\tilde{T}_{\beta\alpha} \right) \right]. \tag{3.22}$$

In a far-from-resonance approximation (FFR), we assume that the energy of the lowest excited state is much higher than the energy of incident radiation. This assumption leads to the considerable simplification of theoretical description. The formula for the Raman polarizability tensor $\alpha_{\alpha\beta}$ within FFR approximation is then given by this relation:

$$\alpha_{\alpha\beta} = \frac{2}{\hbar} \sum_{e \neq g} \frac{\omega_{eg}^0}{(\omega_{eg}^0)^2 - \omega_0^2} \Re \left[\langle g | \hat{\mu}_\alpha | e \rangle \langle e | \hat{\mu}_\beta | g \rangle \right], \tag{3.23}$$

where e and g stand for excited and ground electron states, $\tilde{\mu}$ with the Greek letter subscript represents the component of the electric dipole vector, $\hbar$ is reduced Planck constant, and ω_{eg}^0 is angular frequency between excited and ground state. Tensor $\alpha_{\alpha\beta}$ is no longer complex.

Within FFR, $\beta_A(\tilde{\alpha})^2 = 0$ and $\beta_S(\tilde{\alpha})^2$ is simply labeled as $\beta(\tilde{\alpha})^2$. Definitions of these invariants are simplified to:

$$\alpha^2 = \frac{1}{9}\Re\left[\alpha_{\alpha\alpha}\alpha_{\beta\beta}\right], \tag{3.24}$$

$$\beta_S(\alpha)^2 = \frac{1}{2}\Re\left[3\alpha_{\alpha\beta}\alpha_{\alpha\beta} - \alpha_{\alpha\alpha}\alpha_{\beta\beta}\right], \tag{3.25}$$

Various experimental arrangements can be used to measure Raman scattering. The incident radiation may be non-polarized, linear or circularly polarized. The scattered radiation detection may be perpendicular to the excitation beam (right-angle scattering), against the direction of the excitation beam (backscattering), or at any other angle and also with different polarizations. However, the intensities obtained are always a linear combination of three polarizability tensor invariants.

For example, the measurement of linearly polarized radiation perpendicular to the scattering plane at rectangular scattering with excitation by non-polarized radiation $I_X^U(90°)$ is described by this combination of invariants:

$$I_X^U(90°) = 2K\left[45\alpha^2 + 7\beta_S(\tilde{\alpha})^2 + 5\beta_A(\tilde{\alpha})^2\right]. \tag{3.26}$$

3.1.5 Raman Optical Activity

Raman optical activity (ROA) is chiroptical measurement of Raman scattering. Similarly to other chiroptical methods it measures difference between interaction with right and left circular polarized light. This phenomenon is by between 3 or 5 orders of magnitude weaker then Raman scattering itself.

Four different modulation schemes are used for ROA measurements. First of them is called Incident circular polarization (ICP), where incident radiation is right- or left-handed circularly polarized and unpolarized or linearly polarized scattered light is detected. Scattered polarized light (SCP), in turn, uses unpolarized light or linearly polarized incident radiation and detects polarized right- and left-handed circularly polarized component of scattered light.

The other two methods (Dual circular polarization) use polarized light for both incidence and detection. In-phase dual circular polarization (DCP_I) detects the same circular polarization component as used for incident light (right circular incident for right circular detection, the same for left) and out-of-phase dual circular polarization detects the opposite one (left circular polarized is detected for right circular incident and vice versa). Schema of these different experimental approaches is shown in Fig. 3.2.

ICP and SCP are historically older solutions. ICP was first used by Baron in 1973 [5] [6] These first apparatuses used right-angle scattering (light perpendicular to the excitation beam was detected). Later it was shown that backscattering geometry, and SCP and DCP_I settings may be more advantageous. [7] [36]

The intensity of ROA is similar to the intensity of Raman scattering in equation (3.16) given by:

$$I(\tilde{e}^d, \tilde{e}^i) = 90K\left\langle\left|\tilde{e}_\alpha^{d*}\tilde{a}_{\alpha\beta}\tilde{e}_\beta^i\right|\right\rangle, \tag{3.27}$$

where we have replaced original $\alpha_{\alpha\beta}$ with tensor scattering tensor $\tilde{a}_{\alpha\beta}$ which in addition contains terms corresponding to the effect of the magnetic dipole $\hat{m}_\alpha$
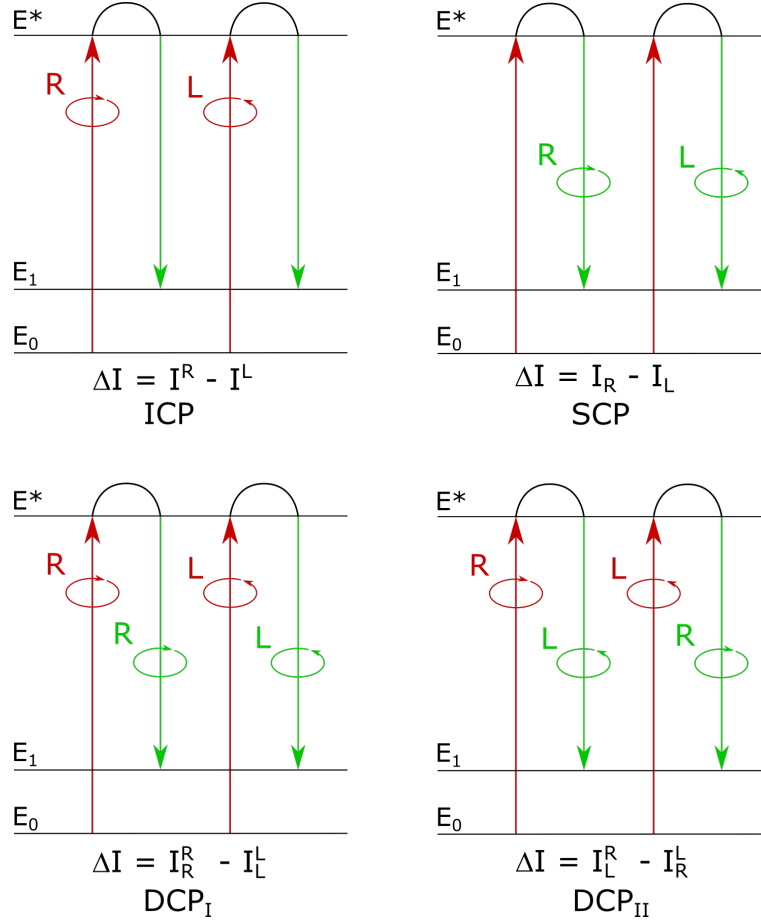
Figure 3.2: Diagrams of four ROA modulation schemes. ΔI is measured quantity. Intensities I_Y^X have their superscript for incident light polarization and subscript for scattered light polarization

and the electric quadrupole $\hat{\Theta}_{\alpha\beta}$. Definition of this tensor is:

$$\tilde{a}_{\alpha\beta} = \tilde{\alpha}_{\alpha\beta} + \frac{1}{c}\left[\epsilon_{\gamma\delta\beta}n_\delta^i\tilde{G}_{\alpha\gamma} + \epsilon_{\gamma\delta\alpha}n_\delta^d\tilde{\mathcal{G}}_{\gamma\beta} + \frac{i}{3}\left(\omega_0 n_\gamma^i\tilde{A}_{\alpha,\gamma\beta} + \omega_R n_\gamma^d\tilde{\mathcal{A}}_{\beta,\gamma\alpha}\right)\right]. \quad (3.28)$$

Einstein convention is used in last equation and $\epsilon_{\alpha\beta\gamma}$ is Levi-Civita symbol, n_δ^i and n_δ^d represents propagation vector of incident scattered light, ω_0 and ω_R are angular frequencies of incident and scattered light, and c is the speed of light. $\tilde{\alpha}_{\alpha\beta}$ is Raman polarizability tensor introduced before. The other tensors $\tilde{G}_{\alpha\gamma}$ and $\tilde{\mathcal{G}}_{\gamma\beta}$ for magnetic dipole interaction and $\tilde{A}_{\alpha,\gamma\beta}$ and $\tilde{\mathcal{A}}_{\beta,\gamma\alpha}$ for electric quadrupole interaction will be described below.

We will describe ROA again using FFR. Within this theory, we get three invariants for the scattering vector $\tilde{a}_{\alpha\beta}$ (in addition to the two that result from $\tilde{\alpha}_{\alpha\beta}$ for Raman scattering within the FFR). However, with a more general theory, it is possible to obtain a total of ten tensors for ROA (and three for Raman scattering).

Complex tensors $\tilde{T}$ can be separated to their real T and complex T' part according to relation:

$$\tilde{T} = T - iT'. \quad (3.29)$$

Within the FFR, it can be shown that tensor A and $\mathcal{A}$ have only real component, while G and $\mathcal{G}$ have only complex component. Further, the invariants of remaining

tensors pairs $\tilde{G}'_{\alpha\gamma}$, $\tilde{\mathcal{G}}'_{\gamma\beta}$ and $\tilde{A}_{\alpha,\gamma\beta}$ and $\tilde{\mathcal{A}}_{\beta,\gamma\alpha}$ become dependent on each other. So only invariants of $\tilde{G}'_{\alpha\gamma}$ and $\tilde{A}_{\alpha,\gamma\beta}$ are necessary for describing intensities within FFR description. Formulas for these tensors are given similarly as in (3.23) by these equation:

$$\tilde{G}'_{\alpha\beta} = \frac{2}{\hbar} \sum_{e \neq g} \frac{\omega^0_{eg}}{(\omega^0_{eg})^2 - \omega^2_0} \Im \left[\langle g | \hat{\mu}_\alpha | e \rangle \langle e | \hat{m}_\beta | g \rangle \right], \qquad (3.30)$$

$$\tilde{A}_{\alpha,\beta\gamma} = \frac{2}{\hbar} \sum_{e \neq g} \frac{\omega^0_{eg}}{(\omega^0_{eg})^2 - \omega^2_0} \Re \left[\langle g | \hat{\mu}_\alpha | e \rangle \langle e | \hat{\Theta}_{\beta\gamma} | g \rangle \right]. \qquad (3.31)$$

Two invariants for Raman scattering tensor within FFR α^2 and $\beta(\alpha)^2$ were already defined in previous subsection 3.1.4. Definitions of the three new ROA invariants $\alpha G'$, $\beta(G')^2$, and $\beta(A)^2$ are given here:

$$\alpha G' = \frac{1}{9} \alpha_{\alpha\alpha} G'_{\beta\beta} \qquad (3.32)$$

$$\beta(G')^2 = \frac{1}{2} \left(3\alpha_{\alpha\beta} G'_{\alpha\beta} - \alpha_{\alpha\alpha} G'_{\beta\beta} \right) \qquad (3.33)$$

$$\beta(A)^2 = \frac{1}{2} \omega_0 \alpha_{\alpha\beta} \epsilon_{\alpha\gamma\delta} A_{\gamma,\delta\beta} \qquad (3.34)$$

For example DCP$_I$-ROA backscattering intensity is proportional to this combination of invariants (FFR):

$$I^R_R - I^L_L = \frac{8K}{c} \left[12\beta(G')^2 + 4\beta(A)^2 \right]. \qquad (3.35)$$

In general, the sum intensities of Raman scattering I^S and ROA intensity I^D can be expressed by these relationships:

$$I^S = 4K \left[D_1 \alpha^2 + D_2 \beta_A(\alpha)^2 \right], \qquad (3.36)$$

$$I^D = \frac{8K}{c} \left[D_3 \beta(G')^2 + D_4 \beta(A)^2 \right], \qquad (3.37)$$

where coefficients D_{1-4} depend on geometry of experiment. Values of these coefficients for different geometries can be found in Tab. 3.1.

Table 3.1: Coefficient of invariants for different experimental arrangements.

Modulation scheme	D_1	D_2	D_3	D_3
ICP	45	7	12	4
SCP	45	7	12	4
DCP$_I$	0	6	12	4
DCP$_{II}$	45	1	0	0

The intensity of ROA signal is often described by circular intensity difference Δ_{CID}, ratio between ROA and Raman intensity. It values usually range from 10^{-3} to 10^{-4} and is defined as:

$$\Delta_{CID} = \frac{I_D}{I_S}. \qquad (3.38)$$

3.2 Calculations of Raman scattering and ROA

Raman and ROA spectra can be handled as fingerprints, e. g. for detection of the presence of a chemical in the sample, conformation of protein, etc. However, in order to understand the meaning of individual peaks or to distinguish specific enantiomers, a combination with a theoretical calculation (spectra simulation) is necessary.

The procedure of the vibrational spectra simulation consists of the following steps:

1. Finding the optimized geometry.

2. Computation of matrix of second partial derivatives of system energy with respect to nuclear coordinates. After that the matrix is diagonalized – this corresponds to the transformation into the so-called normal coordinates, the numbers on the diagonal are the corresponding frequencies.

3. Computation of (Raman scattering) intensity for each normal mode.

There are two main approaches used for these calculations (within quantum chemistry). First group is called *ab initio* (which means "from beginning"). These methods are based on approximative iterative solving of the Schrödinger equation. Their input parameters are only physical constants and no empirical parameters. These methods differ mainly in the level of electron correlation involvement. The simplest option without any involvement is Hartree–Fock self-consistent field theory. The examples of calculation methods that include correlation energy include Moeller–Plessett perturbation theory, (MP2, MP4 with different levels of approximation)[35], coupled clusters, and multireference configuration interaction. By improving the level of involvement of electron correlation, the accuracy of results can be systematically increased, but at the cost of significantly higher computational complexity.

Second used approach is called Density functional theory (DFT). This theory employs empirical functionals with empirical parameters. This means that there is no possibility of systematic improvement and the accuracy of a particular functionality depends on the type of task. Using DFT, it is possible to obtain comparable results with ab initio methods involving electron correlation with computational requirements comparable to the simple Hartree–Fock method. Typical examples of the functionals used for these calculations are the three-parameter Becke's functionals B3LYP and B3PW91[8]

In both cases the precision also depends on basis set used. By increasing the size basis set, accuracy can also be systematically increased. Finding a suitable combination of the base set and the calculation method used is to some extent empirical. For Raman and ROA calculations usage of larger basis sets including extra polarization and diffuse basis functions are important. For the calculation of Raman spectra, diffuse basis functions are important. Therefore, basis sets such as 6-31G*, DGZVP, TZVP,6-311G**, and aug-cc-pVDZ are recommended.

3.2.1 Calculation of intensities

Optimization of geometry and transformation to normal mode is a standard feature of computing software and will not be discussed in detail. However, it is

important to note that the final optimization of geometry should be carried on the same basis set as the spectra calculation. During optimization, however, it is possible to preoptimize the geometry of the system first using a smaller basis.

We will show the principle of calculation of intensities called Field-perturbation formulation. This approach to calculation is advantageous because it avoids summation across all excited states of the molecule. The equations are given only for Raman scattering. Calculation of ROA spectra use the same approach but is more complicated because more tensors are needed.

In the equation (3.14), it was shown that the intensity for a given vibration mode depends on the derivative of the polarizability tensor $\tilde{\alpha}_{\alpha\beta}$ with respect to the normal coordinate Q_a. We can rewrite this with respect to nuclear position (Cartesian) coordinates $R_{J,\gamma}$:

$$\left(\frac{\partial \left\langle \tilde{\mu}_\alpha^I \right\rangle}{\partial Q_a} \right)_{Q=0} Q_a = \left(\frac{\partial \left\langle \tilde{\alpha}_{\alpha\beta} F_\beta \right\rangle}{\partial Q_a} \right)_{Q=0} Q_a = \left(\frac{\partial \left\langle \tilde{\alpha}_{\alpha\beta} \right\rangle}{\partial R_{J,\gamma}} \right)_{R=0} S_{J,\gamma}^a Q_a F_\beta, \quad (3.39)$$

where $S_{J,\gamma}^a$ is transformation tensor between normal modes and nuclear coordinates. The partial derivative term in called Raman atomic polar tensor:

$$\tilde{P}_{\alpha\beta\gamma}^J = \left(\frac{\partial \left\langle \tilde{\alpha}_{\alpha\beta} \right\rangle}{\partial R_{J,\gamma}} \right)_{R=0}. \quad (3.40)$$

If we write an electron Hamiltonian containing the perturbation term for induced $\boldsymbol{\mu}_I$ and intrinsic $\boldsymbol{\mu}_E$ electric-dipole and force $\mathbf{F}$ (within adiabatic approximation):

$$\begin{aligned} \hat{H}_E^A(\mathbf{R}, \mathbf{F}) &= \hat{H}_{E,0}^A(\mathbf{R}) - \boldsymbol{\mu}_I \cdot \mathbf{F} - \boldsymbol{\mu}_E \cdot \mathbf{F} \\ &= \hat{H}_{E,0}^A(\mathbf{R}) - (\tilde{\boldsymbol{\alpha}} \cdot \mathbf{F}) \cdot \mathbf{F} - \boldsymbol{\mu}_E \cdot \mathbf{F} \\ &= \hat{H}_{E,0}^A(\mathbf{R}) - \tilde{\alpha}_{\beta\gamma} F_\beta F_\gamma - \mu_{E\beta} F_\beta \end{aligned} \quad (3.41)$$

and differentiate it with respect to force twice, we can write following equation for $\tilde{P}_{\alpha\beta\gamma}^J$:

$$\begin{aligned} \tilde{P}_{\alpha\beta\gamma}^J &= \left(\frac{\partial}{\partial R_{J,\alpha}} \left\langle \psi_g^A \left| \tilde{\alpha}_{\beta\gamma} \right| \Psi_g^A \right\rangle \right) \\ &= \left(\frac{\partial}{\partial R_{J,\alpha}} \left\langle \psi_g^A \left| \frac{\partial^2 \hat{H}_E^A}{\partial F_\beta \partial F_\gamma} \right| \Psi_g^A \right\rangle \right) \\ &= \left(\frac{\partial^3 \tilde{E}_g}{\partial R_\alpha \partial F_\beta \partial F_\gamma} \right)_{F=0, R=0} \\ &= \left(\frac{\partial^3 \tilde{E}_g}{\partial F_\beta \partial F_\gamma \partial R_\alpha} \right)_{F=0, R=0}, \end{aligned} \quad (3.42)$$

where the substitution of derivatives in the last equation allows the use of previously calculated derivatives according to nuclear coordinates. Derivatives can be calculated as finite difference derivatives (with use of at least five points) or analytically. The analytical calculation is less computationally demanding. The

intensities of the vibrational modes can then be calculated by substituting them into the equation (3.39).

For practical purposes, the convolution of the spectral intensities thus obtained is performed with a suitably chosen distribution function (corresponding to the spectral distribution function in the Eq. 3.10). In our case we use the Lorenz function in the form:

$$f(\nu, \gamma) = \frac{1}{\pi \gamma \left[1 + \left(\frac{\nu}{\gamma} \right)^2 \right]},$$ (3.43)

where γ is half-width at half-maximum.

3.2.2 Optimization in normal coordinates

When modeling biological systems, it may be necessary to employ a relatively inaccurate basis set or methods. The problem lies in the size of organic compounds and excessive computational demands. Therefore, optimized geometry may not match the structure we expect and know for example from X-ray. For this reason, various approaches have been developed to limit the freedom to change the geometry of the molecule.

An alternative approach is called optimization in normal modes. In this method, the energy surface gradient is calculated in Cartesian coordinates and then transformed into normal coordinates. Modes with lower frequencies can be fixed, with the result that the overall structure (e.g. dihedral angles) is changed only slightly. For this calculations we use program QGRAD developed by P. Bouř.[12] [13]

3.2.3 Transfer of molecular property tensors

Another approach dealing with the size of biological systems is to split them into smaller parts, perform calculations on those fragments and reassemble them. In our case, we work with protein, which is a linear polymer. A straightforward solution would be to divide the molecule into individual fragments, perform calculations for them and obtain the calculated spectra as their sum. However, this procedure necessarily neglects the vibration modes beyond one fragment.

Improvement can be achieved by a method called transfer of molecular property tensors. In this case, we do not compose the final spectrum from spectra of individual fragments, but we compose the tensors of the whole system from the tensors of individual fragments. From the tensors created in this way, we can obtain spectra in the manner presented earlier.

First we need to divide the molecule into individual fragments and perform optimization on them. Optimization in normal coordinates can be used to maintain geometry of fragments, which may not correspond to the basis set used, and the absence of the rest of the molecule. Then it is necessary to match the fragments to original molecule and find transformation between their coordinates system. We will not describe the exact procedure in more detail due to its technical nature.

As an example of this procedure, we can show the transformation for a matrix

of second energy derivatives with respect to coordinates (Hessian):

$$H'_{ij} = \frac{\partial^2 E}{\partial R'^i_\alpha \partial R'^j_\beta} = \boldsymbol{U}_{\alpha\delta}\boldsymbol{U}_{\beta\gamma}\frac{\partial^2 E}{\partial R^i_\delta \partial R^j_\gamma}, \tag{3.44}$$

where $\boldsymbol{U}_{\alpha\beta}$ are transformation matrix.[14]

3.2.4   Description of vibrations

Thanks to the calculation of the spectrum, the individual bands in the experimental spectrum can be analyzed in more depth. After transformation into normal coordinates, vibration movements corresponding to individual spectral bands can be visualized. These vibration movements are classified into characteristic groups and special terminology is used for them. Here we provide an overview of terms and abbreviations that we use to describe vibrational movements.

The following terms are used to describe the vibrations of the methylene group: symmetric and asymmetric stretching (ν, symmetry may be expressed by subscript); scissoring (σ), rocking (ρ), wagging (ω) and twisting (τ) bending vibrations. Greek letters in brackets are used as abbreviated notation. The graphic representation of these movements is shown in Fig. 3.3. Deformation vibration of carbon atoms in proline ring can be be divided into deformation in plane (δ) and out of plane (γ). The same vibration can occur across the chain and can be in phase (ip) or out of phase (op). In cases when vibration is related to specific atom or atoms their name is given in bracket.
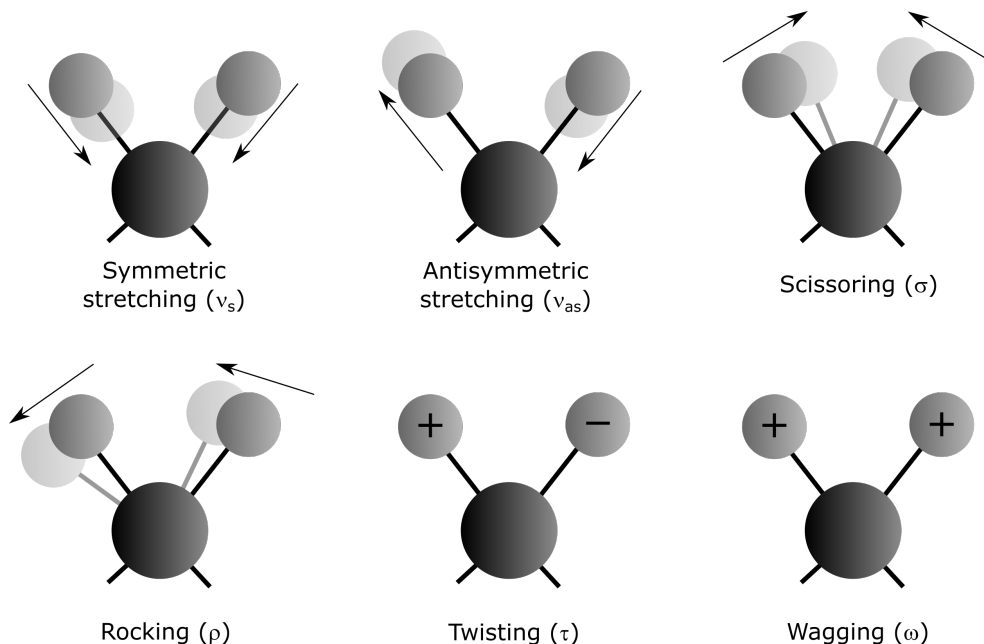
Figure 3.3: Vibration movements of methylene group. Created according to[30].

In addition, there are specific vibrations associated with the peptidic chain, which are referred to as Amide vibrations. Our case differs from most proteins in that there are no secondary amide groups in our molecules (due to the specific structure of proline). Amides within proline chain are tertiary. Nevertheless, we will use common terminology associated with the vibrations of amide bands. It

turned out that similar vibrations would occur in our structure. The only notable difference is that tertiary amides Amide II and Amide III vibrations has different character when compared to secondary ones. We describe vibrations based on vibration shapes for methylacetamide. The individual vibrations are shown in the Fig. 3.4. Amide vibrations I-IV are in the plane, vibrations V and VI are out of the plane. The exact position and intensity of amide bands strongly depends on individual molecule.

Almost every normal mode vibration is composed of several movements (elementary vibrations). We will describe the main components if they can be identified.
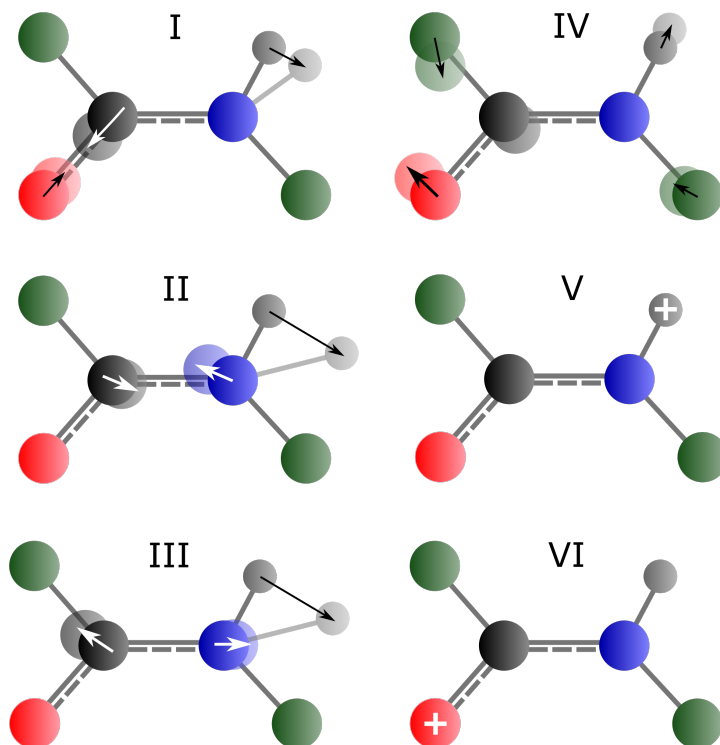
Figure 3.4: Amide vibrations of peptide bond. Dark gray is a carbon atom, red an oxygen atom, blue a nitrogen atom, light gray a hydrogen atom, green a methyl group. Created according to[39].

For a deeper insight into the problems of calculation, a more extensive description of the theory for the calculation of spectra and their interpretation can be found, for example, in[37] [49]

3.3 Electronic circular dichroism

Unlike vibrational spectroscopy, the electron circular dichroism (ECD) uses electron transitions in the molecule. It is a standard and widely used technique for protein studies. Secondary structures such as α-helix or polyprolines may have characteristic spectral signature in ECD.

Simply put, ECD is a chirally sensitive method based on UV absorption. Formally, circular dichroism $\Delta\epsilon$ is the difference between the (molar) extinction

coefficients for left-handed circularly polarized light ϵ_L and for right-handed circularly polarized light ϵ_R:

$$\Delta\epsilon = \epsilon_L - \epsilon_R. \tag{3.45}$$

The measured signal ΔA is defined as:

$$\Delta A = A_L - A_R, \tag{3.46}$$

where A_L is absorption of left-hand circularly polarized light and A_R of right-handed circularly polarized light. General expression between absorption and molar extinction coefficient is:

$$A = cl\epsilon, \tag{3.47}$$

where c is concentration and l is length of path through the sample.

In practice, however, otherwise defined quantities are commonly used for the description. One of them is called ellipticity. This description is based on fact that plane-polarized light is due to different absorption of circular components converted to elliptically polarized light. Tangents of ratio of semimajor and semiminor axes is ellipticity θ (because these angles are generally small, the tangent can be replaced by a linear function). It can be shown that the following transformation holds:

$$\theta(°) = 32.98\Delta A. \tag{3.48}$$

Molar ellipticity $[\theta]$ can be also defined, and the transformation is:

$$[\theta] = 3298\Delta\epsilon. \tag{3.49}$$

In this work, we do not introduce quantities dipole and rotational strength important for the theoretical understanding of ECD, because they are not essential in our context. We only state that the transitions of the electric dipole transition moment and magnetic dipole transition moment are important for this phenomenon. For a deeper understanding of the technique, we recommend the reader to look, for example, in book[53].

4. Materials and methods

In the first part of this chapter we describe the used samples, instrumentation and experimental processing. The second part describes the tools used to perform, process and analyze quantum chemical calculations.

4.1 Samples

For experiments we purchased three commercial poly-L-proline samples from Sigma-Aldrich company with product numbers P2129, P2254, P3886 (further referred to as *medium*, *short*, *long*) which vary in molecular weight. Molecular weights given by manufacturer and values used for calculation for preparation of the solution are shown in Table 4.1. As specified in product list M_{visc} is determined by viscosity and M_{ALLS} by multiangle or low-angle laser light scattering. We used average of these numbers M_w for calculations. The number n in the table is the approximate number of prolines that the polyproline chain contains, calculated by dividing the estimated molecular weight of the polyproline by the molecular weight of the proline. For all measurements *medium* sample is used, unless otherwise stated.

For longer storage, samples were placed in sealed tubes in the refrigerator. Samples were checked by ECD before each set of measurements and no differences were observed in the samples.

Table 4.1: Characteristics of used samples

Sample	M_{visc} (Da)	M_{ALLS} (Da)	M_w (Da)	n
2254 (short)	5800	3600	4700	40
2129 (medium)	14800	27900	21000	190
3886 (long)	32000	35400	34000	300

4.2 Preparation of PPI

In native state all commercial samples were in PPII conformation. However, we aimed on characterization and futher study of PPI conformation. Therefore, we recrystalized samples into desired form according to the method of Steinberg[48]. For each sample we prepared 5% solution of poly-L-proline in formic acic and diluted ten times in n-propanol. Then we let the samples mature for at least 72 hours at room temperature, which was enough for *trans*-to-*cis* peptide bond isomerization.

Instead of ether precipitation and drying the samples were recrystalized by volatilization which led to formation of small crystals well soluble in 1-butanol. (Volatilization was performed by Martin Šafařík, laboratory technician from Institute of Organic chemistry and Biochemistry AS CR, to whom we are very grateful.) Alternativelly, we also tried lyophilization which however led to formation of larger flake-like crystals of reduced solubility.

4.3 ECD measurement

The instrument used for ECD measurement is JASCO J-1500. The sample holder used for measurements allows thermal stabilization of the sample using a combination of water and thermoelectric temperature control. The instrument must be purged with nitrogen, which makes long measurements relatively costly. Quartz 1mm cuvettes were used for measurements.

ECD measurements were performed in both aqueous solution and organic solvents. At first we prepared samples with concentration $1\,\text{mg/ml}$ but absorption was too strong. We found concentration $0.1\,\text{mg/ml}$ optimal for measurements in organic solvents and used it for all measurements (except for one measurement where a higher concentration had to be used). However, it should be noted that the weighed quantities were close to the limit of analytic scales and the concentrations are thus rather approximate. The measurement was performed in the interval 190-300 nm with a step 0.1 nm and a scanning speed 10 nm/min.

We used 1-propanol (high purity, Made in USSR in 1970) and 1-butanol (Sigma–Aldrich, Product No. 154679) for measuring stable PPI conformation (organic solvents). The temperature was not stabilized during measurements in organic solvents. The aim of the measurements in organic solvents was to confirm the correct preparation of PPI. UV and ECD raw spectra were treated by solvent spectrum subtraction and normalized to unity at maximum of absorbance.

Water was used as solvent for measurements of mutarotation process. A concentration of $0.2\,\text{mg/ml}$ was used for these measurements. The temperature was stabilized at $35\,°C$. Due to the nature of the studied process and cost of the experimental time, only the first ten hours were continuous, followed by shorter measurements for the next three days each morning. Between measurements, the sample was stored in a thermostat at the same temperature. The measurements were performed in the interval 185-250 nm with a step 0.1 nm and a scanning speed 50 nm/min. The interval between individual measurements was 3 minutes.

Mutarotation spectra were analysed by singular value decomposition and fitted to chemical kinetics.

4.4 Solid phase Raman measurement

Solid phase Raman scattering measurements were performed both for purchased commercial samples and samples obtained by isolation of our prepared PPI from solution. Data were recorded by Raman microspectrometr *Alpha 300* made by WITec. The instrument is designed primarily for confocal Raman microscopy but also includes other techniques: atomic force microscopy, near field scanning microscopy and TrueSurface® (optical mapping of the sample surface).

The measurement itself proceeded as follows. Small sample grains were placed on a slide for measurement without the use of a cover glass. A suitable grain was then found using an optical microscope. The instrument was switched to point spectrum acquisition mode, and a setting with the maximum Raman signal was found using slight focus changes. Individual spectra were collected for approximately 1 minute. Wavelength of excitation laser light was 532 nm.

The baselines of acquired raw spectra were corrected by spline. Then the spectrum of *medium* PPI sample was normalized so that the height of the peak

at 1450 cm^{-1} was 1. All other spectra were normalized to the same integral intensity within the 250 to 1900 cm^{-1} interval to the *medium.*

4.5   Measurement of Raman and ROA spectra

Raman and ROA spectra, we measured with custom instrument, which is still under development. It is a device developed by J. Kapitán at Palacký University Olomouc in cooperation with the company Meopta. The device is based on a design by W. Hugh.[20] It allows simultaneous measurement of all ROA modulation schemes in backscattering[31]. However, it is possible to adjust it to other scattering geometries. The device also allows simultaneous measurement in the low and high frequency region. The instrument uses system of rotating half-wave plates to compensate for artefacts. The exact specifications and construction of the instrument are not made public due to ongoing development and possible patent litigation. The instrument is also equipped with a sample holder that allows thermal stabilization. The described experiments were performed by J. Kapitán to whom we are very grateful.

The aim of the performed experiments was to obtain ROA and Raman spectra of polyproline I in water, which is more suitable solvent for measuring Raman and ROA spectra than organic solvents. The second goal was to obtain Raman and ROA spectra from the mutarotation process for further processing. During the experiments, we encountered great problems with solubility and failed to prepare a concentrated sample. The result was a relatively weak ROA signal and problems with turbidity. For this reason, the samples were manipulated during the measurement (e. g. cuvette change) and various changes were made to the experimental setup. Unfortunately, this handling had a negative effect on the subsequent data processing by statistical methods.

The first measurement was performed with the sample dissolved in propanol. The initial concentration was 10 mg/ml, but the sample was poorly soluble, milky cloudy. The desired concentration was two time higher. Because PPI is stable in this solvent, the measurement could be performed under normal conditions for the required time. There were several immediately consecutive measurements with a total length of about 10 hours. The individual measurements differed slightly in the experimental settings - the time of collection of the individual spectrum and the type of recorded data. The data were used as one spectral series for processing.

The second sample (PPa) was a sample in water with an initial concentration of 10 mg/ml. The supernate after centrifugation was used for measurement due to poor sample solubility. The sample was first measured at 5 °C for about 35 hours. Most measurements were performed without setting changes and interruptions and therefore were suitable for factor analysis. From our experience gained during ECD measurements and from information in the literature[28][47], we expected that the kinetics during this temperature would be slow enough to obtain an almost pure PPI spectrum.

This first measurement was followed by another measurement with the same sample (although of course already with a decrease of PPI in the sample). The temperature was raised to 20 °C and the measurement was continued for another 60 hours. Unfortunately, during the measurement, due to a change in signal

intensity, the experimental settings were changed, which made it impossible to perform fit to reaction kinetics. This change could be due to a change in solubility caused by both a change in temperature and also an increase in the concentration of PPII with better water solubility.

The third sample (PPc) was also using water as a solvent. The sample was first measured at $5\,^{\circ}\mathrm{C}$ for about 2 hours. Then the temperature was changed to $25\,^{\circ}\mathrm{C}$. However, problems with the sample properties occurred again during the measurement. After about twenty hours of measurement, the cuvette was replaced. After this change, the measurement was carried out continuously for about 45 hours.

Data from the first hours of measurements obtained at $5\,^{\circ}\mathrm{C}$ of PPa and PPc samples were processed together by factor analysis to remove background. Continuous data obtained from long PPa sample collection at $5\,^{\circ}\mathrm{C}$ were processed separately by factor analysis and fitted to mutarotation kinetics. Data from the measurement of PPa mutarotation at $20\,^{\circ}\mathrm{C}$ were evaluated only qualitatively after baseline correction by factor analysis.

In this work we also use data measured in 2011 on a commercial instrument ChiralRAMAN-2X by Biotools inc. We use the results of two measurements. The first is the measurement in butanol which took 100 hours with intensity $150\,\mathrm{mW}$. The second is a 70-hour measurement of mutarotation in water. The sample used was short. Measurements were performed without thermal stabilization, lab temperature was approximately $25\,^{\circ}\mathrm{C}$. The concentration used was $20\,\mathrm{mg/ml}$ and the laser power was $1000\,\mathrm{mW}$.

4.6 Chemical kinetics of mutarotation

The transition from PPI to PPII is a gradual reaction with number of intermediate products as discussed in Introduction. Our measured spectra are a superposition of the spectra of all current states present at the moment in the sample. Following lines describe mathematical tools used to separate these states which are distinguishable by using the spectroscopic method.

4.6.1 Description of reaction mechanism

Kinetics of chemical reactions is described by system of differential equations. In our case, we know that the equilibrium of the reaction is significantly shifted in favor of the product. So we assume that the reaction runs in one direction only. The system of equations for the initial substance A, the final product B, and for

the general number n of intermediates X_n, is as follows:

$$[A] = -k_1 \frac{d[A]}{dt}$$

$$[X_1] = k_1 \frac{d[A]}{dt} - k_2 \frac{d[X_1]}{dt}$$

$$\ldots \tag{4.1}$$

$$[X_n] = k_n \frac{d[X_{n-1}]}{dt} - k_{n+1} \frac{d[X_n]}{dt}$$

$$[B] = k_{n+1} \frac{d[X_n]}{dt},$$

where $[\_]$ are concentrations and k_i are reaction rate coefficients. The general solution of this system is the exponential of a matrix which correlates individual concentrations and their derivatives multiplied by time. We denote the initial conditions for X_i as $X_i(0)$.

For no intermediate we have:

$$[A](t) = A(0)e^{-k_1 t}$$

$$[B](t) = A(0)(1 - e^{-k_1 t}). \tag{4.2}$$

Obviously, one of the concentrations may be neglected because it is different from the other, and we will not disclose the concentration of the product in the equations.

For one intermediate we have:

$$[A](t) = A(0)e^{-k_1 t}$$

$$[X_1](t) = -\frac{k_1(e^{-k_1 t} - e^{-k_2 t})}{k_1 - k_2}[A](0) + e^{-k_2 t}[X_1](0) \tag{4.3}$$

$$[B](t) = 1 - [A](t) - [X_1](t).$$

For two intermediates we have:

$$[A](t) = A(0)e^{-k_1 t}$$

$$[X_1](t) = -\frac{k_1(e^{-k_1 t} - e^{-k_2 t})}{k_1 - k_2}[A](0) + e^{-k_2 t}[X_1](0)$$

$$[X_2](t) = \left(\frac{-e^{-k_2 t}k_2 k_1^2 - e^{-k_1 t}k_2 k_3 k_1 + e^{-k_2 t}k_2 k_3 k_1 + e^{-k_1 t}k_2^2 k_1}{(k_1 - k_2)(k_1 - k_3)(k_2 - k_3)} + \right.$$

$$\left. + \frac{e^{-k_3 t}k_2 k_1}{(k_1 - k_3)(k_2 - k_3)} \right)[A](0) \tag{4.4}$$

$$- \frac{k_2(e^{-k_2 t} - e^{-k_3 t})}{k_2 - k_3}[X_0](0) + e^{-k_2 t}[X_1](0)$$

$$[B](t) = 1 - [A](t) - [X_1](t) - [X_2](t).$$

4.6.2 Factor analysis

Factor analysis is a statistical method that allows to reduce the dimension of the studied problem. In our case, before performing factor analysis, the number of

parameters used to describe the system in time is equal to the number of points in the spectrum. The actual number of system parameters - the concentration of individual states of the polyproline molecule - is significantly smaller. Factor analysis allows to find the main parameters that are needed to describe the whole system (statistical set) with minimal error. A more detailed description of factor analysis can be found in e.g.[34]. We are using specific method of factor analysis called singular value decomposition (SVD, but we will often refer to it as factor analysis.) Mathematically, this operation can be described as a singular value decomposition of a spectra matrix $\mathbf{Y}$ that can be written as:

$$\mathbf{Y}(\lambda) = \mathbf{V} \cdot \mathbf{W} \cdot \mathbf{S}(\lambda), \tag{4.5}$$

where $\mathbf{S}$ is matrix with normalized spectral profiles (spectral components, we will usually refer to them as subspectra). Matrix $\mathbf{W}$ is "diagonal-like"[1] and its numbers have the meaning of weights (or normalization factors) of the spectral profiles. The unitary matrix $\mathbf{V}$ combines these spectral profiles into each experimental spectrum (the elements of the matrix will usually be referred to as scores or coefficients). Thus, the meaning of its elements is the relative representation of spectral component in the individual experimental spectrum.

However, spectral components do not match the spectra of individual states present in the sample. Spectral components with the greatest weight has meaning of average spectrum – for example, it can contain common parts of the spectra of individual states. Other spectral components (second, third subspectrum, etc) match to change in experimental spectra (e. g. peak shifts) sorted by their significance (in numerical meaning). Two numbers can be used to evaluate the significance of individual spectral profiles. The first number is its singular value, which expresses its relative weight to other spectra. The second number is the residual error, which expresses the error resulting from the neglect of the rest of the spectral profiles.

To obtain the spectra of the individual substances in the sample, factor analysis must be combined with the kinetics introduced previously.

4.6.3   Fitting factor analysis with kinetics

The amount of spectral profiles that make sense to be considered can be estimated from several parameters. The first parameter is weight in matrix $\mathbf{W}$. If the corresponding weight is too small compared to previous ones, this spectrum probably doesn't carry important information about spectral change. A similar quantitative parameter is residual error. It describes the difference between original measured spectra and spectra obtained as a combination of first n spectral profiles. Like weights, when it becomes small enough, the following spectral profiles need not be considered.

The shape of the spectral profile is the last property we can evaluate. If it looks more like noise than spectral change, it should no longer be considered. All these characteristics should be consistent with each other.

When we select the appropriate number of spectral profiles, we need to find the relationship between the coefficients in the matrix $\mathbf{V}$ and the concentrations in

---

[1]It has non-zero numbers only for (some) elements w_{ii}, but it's not generally square matrix.

the kinematic model. Experimental spectra can be rewritten as the superposition of spectra of the individual components at each time of measurement t_i, which can be written as:

$$\vec{c}(t_i) \cdot \mathbf{M}(\lambda) = \mathbf{Y}(t_i, \lambda), \qquad (4.6)$$

where $\vec{c}(t_i)$ is vector of concentration described with kinetics, $\mathbf{M}(\lambda)$ is matrix with spectra of components in sample.

When we combine this equation with (4.9) in specific time t_i, we get ($\mathbf{V}$ was replaced with it's row):

$$\vec{c}(t_i) \cdot \mathbf{M}(\lambda) = \vec{V}(t_i) \cdot \mathbf{W} \cdot \mathbf{S}(\lambda). \qquad (4.7)$$

Thanks to unitarity of $\mathbf{S}$ we can write:

$$\vec{c}(t_i) \cdot \mathbf{M}(\lambda) \cdot \mathbf{S}^*(\lambda) = \vec{V}(t_i) \cdot \mathbf{W} \cdot \mathbf{S}(\lambda) \cdot \mathbf{S}^*(\lambda), \qquad (4.8)$$

$$\vec{c}(t_i) \cdot \mathbf{G} = \vec{V}(t_i) \cdot \mathbf{W}, \qquad (4.9)$$

where we have replaced $\mathbf{M}(\lambda) \cdot \mathbf{S}^*(\lambda)$ with $\mathbf{G}$. $\mathbf{G}$ is obviously time independent and its elements are generally unknown. The last equation can be used for fitting kinetics parameters such reaction rate constants and elements of $\mathbf{G}$ to obtained elements of matrix $\mathbf{V}$. After that we can get spectra of components $\mathbf{M}$ in sample with use of matrix $\mathbf{G}$, because:

$$\mathbf{G} \cdot \mathbf{S}(\lambda) = \mathbf{M}(\lambda) \cdot \mathbf{S}^*(\lambda) \cdot \mathbf{S}(\lambda) = \mathbf{M}(\lambda). \qquad (4.10)$$

4.7   Software used for spectroscopic analysis

The software supplied with the device was usually used for the initial data processing. Advanced data treatment and graph production were done in specialized software namely: GRAMS/AI by Thermo Fisher Scientific, SigmaPlot by Systat Software, Inc, and OriginPro 2020 by OriginLab Corporation. We used factor analyzes to process spectral series from mutarotation measurements. For this processing we used an inhouse program written in MATLAB. Common adjustments such as normalization, baseline subtraction, etc. were made to the spectra.

Subtraction of solvent is necessary for Raman data treatment. Raman spectra also suffer from a background caused mainly by inaccurate background subtraction and fluorescence. For individual spectra, the baseline correction was performed by interpolation with a suitable function or manually. For spectral series, baseline correction after solvent subtraction was performed on factor components. This approach allows baseline correction to be performed systematically on the entire spectral series simultaneously. The baseline correction is performed sequentially for each individual factor component and the corrected spectral series is recreated by combination of the corrected factor components instead of the original ones. Especially for higher factor components, this procedure is partially empirical and may be burdened with a certain subjective error. However trends observed in spectral series remain unaffected.

Fitting of chemical kinetics to singular values scores (coefficients of spectral components in individual measured spectra) were done in SigmaPlot. It combines some scripts with internal fitting function of SigmaPlot.

4.8 Quantum chemistry calculations

We use classical programs for quantum chemical calculations: Gaussian 09 for calculations and GaussView 6 for some input preparation analysis of vibrational movements (both by Gaussian, Inc.) The calculations were performed mainly on a computer network MetaCentrum, which allows academics to use computing power across the Czech Republic. Due to high amount of individual computation tasks it was necessary to use other external or inhouse programs. Several programs have been developed specifically for use in this thesis. A description of the main principles used is provided in the following lines.

Considerable part of this thesis is focused on the specification of particular molecular geometry for inputs of quantum-chemical software packages and likewise analysis of the molecular geometry from outputs of such programs. Differences between individual molecules are in length, sequence of A and B puckering and ϕ, ψ and ω angles. Manual building of molecular geometries and analysing of many similar well defined molecules would be too time consuming. Therefore we developed several Bash and Python scripts for input preparation, output data extraction, and analysis and used them in association with Fortran programs created by prof. Petr Bouř (programs and documentation can be found on his webpage[15]). Program for preparing proteins atoms coordinates, analysing their geometry, and performing optimization in normal modes are more complex and will be shortly described.

4.8.1 Preparing molecule geometry

We have developed a program *ProGen.py* (Protein Generator) enabling creation of almost any peptide structure based on input containing the sequences and backbone dihedral angles. As special case we use this software for generation of polyproline molecule. The program is able to make polyproline of any length with any order of puckerings and dihedral angles (including ω).

ProGen.py needs two types of input. First type are geometries of small groups of atoms in defined format (first column with atom number, the other of coordinates of this atom). It is important that first three atoms must be those in peptide backbone, first nitrogen, then C_α and C_1 (or other atoms in the same position in backbone). These small parts are building blocks – amino acids (A or B conformation of proline) or starting and ending groups.

Second type of inputs are string sequences. First with sequence of given file names and second with sequence of ϕ, ψ, ω angles both from N to C end.

This solution is in fact some adaptation of describing geometry of molecule with Z-matrix. The process starts with simple including first given block, so we have first three atoms in backbone. Now we need bond length d, angle between previous bond and new one α_n and dihedral angle of last bond θ_n. Length and angle of bonds are determined by atoms, dihedral angle is given by input sequence. We obtain new atoms position X_{n+1} as:

$$X_{n+1} = X_n + d\boldsymbol{R}_2(\theta_n)\boldsymbol{R}_1(\alpha_n)\frac{(X_{n-1} - X_n)}{|(X_{n-1} - X_n)|}, \qquad (4.11)$$

where $(X_{n-1} - X_n)$ is vector of last bond, $\boldsymbol{R}_1$ rotation matrix around axis per-

pendicular to plane defined by last two bonds and $\boldsymbol{R}_2$ rotation around last bond. This equation is just mathematical formula for calculation coordinates of next atom.

We can repeat this three times to get coordinates of backbone atoms of next block. Now we need to add the rest of atoms of the brick. If we marked coordinates of backbone atoms in input file as $\vec{a}$, $\vec{b}$ and $\vec{c}$ and coordinates in our peptide $\vec{A}$, $\vec{B}$ and $\vec{C}$, we can define base vectors $\vec{x}$, $\vec{y}$ and $\vec{z}$ and $\vec{X}$, $\vec{Y}$ and $\vec{Z}$ as:

$$\vec{b_1} = \vec{b} - \vec{a} \qquad\qquad \vec{B_1} = \vec{B} - \vec{A} \qquad (4.12)$$

$$\vec{b_2} = \vec{c} - \vec{a} \qquad\qquad \vec{B_2} = \vec{C} - \vec{A} \qquad (4.13)$$

$$\vec{x} = \frac{\vec{b_1}}{|\vec{b_1}|} \qquad\qquad \vec{X} = \frac{\vec{B_1}}{\left|\vec{B_1}\right|} \qquad (4.14)$$

$$\vec{y} = \frac{\vec{x} \times \vec{b_2}}{\left|\vec{x} \times \vec{b_2}\right|} \qquad\qquad \vec{Y} = \frac{\vec{X} \times \vec{B_2}}{\left|\vec{X} \times \vec{B_2}\right|} \qquad (4.15)$$

$$\vec{z} = \frac{\vec{x} \times \vec{y}}{|\vec{x} \times \vec{y}|} \qquad\qquad \vec{Z} = \frac{\vec{X} \times \vec{Y}}{\left|\vec{X} \times \vec{Y}\right|} \qquad (4.16)$$

It is important that we defined basis for atoms in our reference configuration and in built peptide in the same way. We can mark first atom (usually N) as origin of our coordinate system. Now we can simply transform positional vectors $\vec{r'}$ with our reference block coordinates with use of transition matrix $\boldsymbol{T}$ to our new coordinate system $\vec{r}$:

$$\vec{r} = \boldsymbol{T}\vec{r'}, \quad \boldsymbol{T} = \begin{bmatrix} \vec{a}^T \\ \vec{b}^T \\ \vec{c}^T \end{bmatrix}. \qquad (4.17)$$

At this moment we have coordinates for each atom in base which is defined in the same way in both systems. We can get position $\vec{R}$ of each atom in peptide from formula:

$$\vec{R} = \vec{R_N} + r_1\vec{A} + r_2\vec{B} + r_3\vec{C}, \qquad (4.18)$$

where r_i are components of vector $\vec{r}$ and $\vec{R_N}$ is position vector of atom marked as origin of coordinate system.

This is the end of adding one part of built peptide. Next step will be calculating position of three following atoms in peptide backbone and repeating of the same procedure. This will continue to the end of given structure in the same way.

4.8.2 Calculating of dihedral angles from obtained geometry

The second larger program *ProStat.py* allows automatic analysis of the geometry of the polyproline molecule. The program is able to automatically find and calculate dihedral angles in a backbone and analyze the puckering of proline residues.

We need to find which atoms are included in backbone to calculate dihedral angles in backbone and proline rings for calculation of puckering. One possibility is to write them manually for each type of calculated molecule, because the order of atoms in files for peptides, varying just in for example puckering, should be the same. But with higher types of peptides and scripts using different order of atoms we used more robust solution.

We made script which is able to find given sequence of atoms in molecule and identify proline rings as well. The script needs text input (e. g. "Pro Pro Pro" for proline tripeptide) and file with geometry of molecule. It returns sequence of dihedral angles and proline puckering (if there is a proline in sequence).

Algorithm is quite straightforward. At first we need to determine which atoms are connected with bond. We used simple upper constriction for distance between atoms without any dependency of atom type. Adding some dependency for CH–bond etc. would be more correct but this solution worked fine.

After building bonds we worked with molecule just like with topological graph. We applied simple depth-first search algorithm to find given sequence of atoms. For our set of molecules this is strong enough, because non uniqueness is possible just at the ends of peptide for hydrogen atoms. But this may be a problem for application to other systems and should be fixed.

With identified atoms in backbone and in proline residues we just calculated dihedral angles and puckerings of proline ring.

4.8.3   Statistical analysis of geometry descriptors

Third program *AnglesStat.py* was used to analyse obtained angles from the previous script. This analysis is mainly technical thing – reading data from output file and applying statistics after categorization. The only problem is that we are not able to just average our data.

Our data for statistics are angles. Problem is that $-180° = 180°$. If we shift this jump out of relevant part of circle, we are able to use standard mean value and standard deviation. But it would be more laborious and not so nice solution as the more correct which is directional statistics.

Each angle α_i can be also taken as complex number z_i with polar coordinates $(\alpha_i, 1)$. Other option is to thing about the angle as unit two dimensional vector with prescription $(\cos \alpha_i, \sin \alpha_i)$ (of course it is the same as real and complex part of z_i).

These complex numbers or vectors can be added and we obtain complex number or vector again. Direction of this will be obviously something like average of directions of all vectors in set. We can write:

$$C = \sum_i \cos \alpha_i, \quad S = \sum_i \sin \alpha_i, \tag{4.19}$$

where C and S has meaning of complex and real part or coordinates of sum of all vectors or complex numbers. We can easily get the mean value of angle $\bar{\alpha}$ as:

$$\bar{\alpha} = \arg(C + iS). \tag{4.20}$$

The mean value given by this definition satisfies equation $0 = \sum_i \text{arc}(\alpha_i, \bar{\alpha})$, where arc() is function returning size of smaller angle between given angles. This

mean value also minimalizes square of these values. Both statements are proven in[21]

Information about dispersion of values is provided by modulus or norm of vector $R = |(C + iS)|$. If we have n unit vectors of the same direction, we will get vector with norm n after summation. If we have set of vectors with random orientation, we will get vector with norm near to zero. It can be expressed as:

$$2\left(1 - \frac{R}{n}\right) = \frac{2}{n}\sum_i (1 - \cos(\alpha_i - \bar\alpha)) \simeq, \frac{1}{n}\sum_i (\alpha_i - \bar\alpha)^2 = s^2, \qquad (4.21)$$

where s^2 is variance and approximation is correct for small differences in angles (and is justified for our application). Proof of these equations is again possible to find in[21]. We can get standard deviation s in degrees from last equation as:

$$s = \frac{180°}{\pi}\sqrt{2 - \frac{2R}{n}}. \qquad (4.22)$$

This value varies from value calculated as sum of squares of differences but for our data and precision it's the same.

4.8.4 Optimization in normal coordinates and transfer of molecular property tensor

As described in the theory, to calculate the spectra of large molecules such as proteins, the system can be divided into smaller parts. Those optimized (e.g. in normal coordinates) and then by using the transfer of molecular property tensors, the spectrum of the whole molecule can be calculated.

This can be done with the programs created by prof. Bouř. His program *fragmentdna* divides e. g. DNA or protein molecule into fragments and creates the necessary information for the transfer of molecular properties at the same time. Program *qgrad* is able to perform optimization in normal coordinates with the possibility of restriction of low frequency modes. Finally, program *cctn* is able to perform the transfer of molecular property tensor itself.

In our case, we somewhat modified this procedure. The number of possible fragments for trimers and tetramers, differing only in puckering, is 8 and 16. Therefore, calculation of all fragments obtained by division of molecules would result in calculations of many same fragments repeatedly. To improve our capabilities, we have created a program that allows to assign the corresponding fragments to any polyproline molecule. These fragments can be calculated only once and used repeatedly for many polyproline molecules differing in puckering motif.

The non-trivial information that is needed for the transfer of molecular property tensors is mainly matrix connecting the atoms of individual fragments with the atoms of the whole molecule. To create this matrix, we used a similar approach as in the analysis of dihedral angles and puckerings. After finding the backbone atoms and determining the puckerings, a corresponding precalculated fragment is found for all positions. Then each fragment's backbone is overlapped with corresponding backbone atoms of whole molecule. The rest of the atoms is matched based on calculated distances. The implemented solution is not perfect, but it suited our conditions.

4.8.5 Parameters of performed calculations

At first we performed some calculation to compare various bases and methods to those recommended in the literature. We focused on the computational time and differences in optimized geometries. List of bases and methods can be found in Results.

The first step in calculating the spectra of PPI was to optimize the geometry. The optimization was performed on heptamers of all combinations of puckerings (128 conformers) in the 6-31G* bass set at DFT/B3PW91/cpcm(propanol) level of theory.

The optimized geometries were then statistically processed to obtain information about relationship between values of dihedral angles and proline puckering. Larger molecules of different lengths and combinations of puckerings were constructed. Shorter fragments – trimers and tetramers – were also created. Dihedral angles were selected according to the local combination of puckerings. The angles were assigned according to the Tab. 5.5 (in section 5.7.1).

Short fragments were then geometry optimized using *qgrad* and spectra calculated at B3PW91/6-31++G** level of theory. Standard geometry optimization would lead to a more significant deviation from the structure predicted with heptamers, as the interaction with the rest of the molecule would be lost. Therefore, the optimization was performed in normal coordinates with fixation of low frequency modes ($<100\,\mathrm{cm}^{-1}$).

The final choice of methods and basis sets was based primarily on recommendations in the literature[37] and previous experience with similar calculations[40][41]. An extensive comparison of basis sets and methods can be found, for example, in[22]. A detailed study of the influence of basis sets and methods on the calculated geometries and spectra is beyond the scope of this work.

According to[37], DFT methods provide a good ratio between computational complexity and fine quality results for spectroscopic applications. We performed a comparison of our chosen functional B3PW91 with the other functional B3LYP and with MP2 method (using 6-31G*). A several heptamers were used for this comparrison. Calculations using B3LYP required comparable amount of time (around 80 hours CPU time for heptamer). Calculation using MP2 method could not be performed in a reasonable time (2 weeks, 8CPU). The choice of basis set 6-31G* was given mainly by recommendations in the literature. This basis set was used because polarization functions on heavy atoms (carbon) are important for the structure of a protein chain.[37]. Again, calculation with a larger basis set would not be possible due to disproportionate computational complexity. When calculated a pentapeptide (AAAAA), the addition of a diffuse function (e.g. 6-31+G* and D95+*) caused an order of magnitude increases of computational time. When comparing optimized geometries using different basis set, the differences in dihedral angles were in units of degrees. This is a systematic error that may burden the entire calculation.

The larger basis set 6-31++G** was used to calculate the spectra of trimers and tetramers contains diffusion and polarization functions of both heavy and light atoms. The studied phenomena include changes in polarizability and therefore the distribution of electrons is important for the calculation of the spectrum itself.

Finally, the spectra of larger molecules were calculated using the transfer of

molecular property tensors. Calculations were performed for polyprolines composed only of puckerings A, only of puckering B, and random puckerings in a ratio of 3: 2. The ratio was not enforced, only the puckerings for individual places were selected with this probability. The spectra of random chains were always calculated as the average of 20 different PPI chains for which the spectra were calculated individually. Calculations were performed for all chain lengths between 10 and 100 AA with step of 5. Based on the tetramer, calculations were also performed for length 5.

We also performed calculations of prolyproline spectra in which the 3:2 ratio was maintained, but the distribution was not random. We calculated spectra for chains of length 80 and 100 with repeating blocks. For 80 the blocks were: 3A2B, 6A4B, 12A8B, 24A16B and 48A32B and for length 100: 3A2B, 6A4B, 12A8B, 30A20B and 60A40B (the number expresses the number of prolines in a given puckering). The choice of the blocks was such that it would fit on the chains as an integer multiple.

Raman spectra within individual groups (e. g. only A puckering with different length) were at first normalized to a length of individual proline molecule. Then a global maximum over the whole set of spectra at 200-1800 cm^{-1} was found and all spectra were divided by this same number. The corresponding ROA spectra were divided by the same factor to preserve the ratio between Raman and ROA intensities.

5. Results and Discussions

In this work, we combine two main mutually complementing approaches – experimental and theoretical – to the studied structure. We will first describe the experimental part of this work, because obtaining Raman and ROA spectra of PPI is the main goal of this work. Then we analyse the computational part separately, which also forms a significant part of this work. Finally, we will also compare the results of individual experiments and compare the experimental results with theoretical calculations.

5.1 Verification of correct sample preparation by ECD

We decided to use ECD to verify the preparation of samples in PPI conformation. The reason is its availability, low sample consumption and already published spectra of both PPI and PPII conformations. This allowed samples to be checked before all measurements.

We performed measurements of all three samples in two solvents: 1-butanol and 1-propanol. The obtained spectra can be found in Fig. 5.1.

Our spectra correspond to the published ones, in the Fig. 5.1 it is possible to compare with the spectrum of PPI in D_2O taken from [18] (the sample had a molecular weight of 5000-7000 Da).

ECD signal of PPI has a characteristic shape consisting of a negative band at 200 nm, a stronger positive band at 215 nm and a weak negative band at approximately 230 nm (PPII has strong negative band at 205 nm so it can be easily distinguished). We compared signals of all three available polyproline samples (*short*, *medium*, and *long*, see Material and Methods section 4.1 in Table 4.1). The signal of the *long* sample was too weak and noisy in 1-propanol. It was therefore prepared at a higher concentration of ≈1 mg/ml. To compare magnitudes of their ECD signals their absorption spectra were normalized to the same value at maximum. This normalization compensates the inaccuracies of concentrations of the samples.

Signals of all samples manifest similar intensities and shapes of ECD spectra. This shows that we successfully prepared PPI conformation from all three purchased PPII samples and that the resulting structure of all samples is similar.

Samples were always checked in the same way before measuring with other methods. No significant differences in ECD spectra were observed for several months when the solid phase samples were stored in the refrigerator.
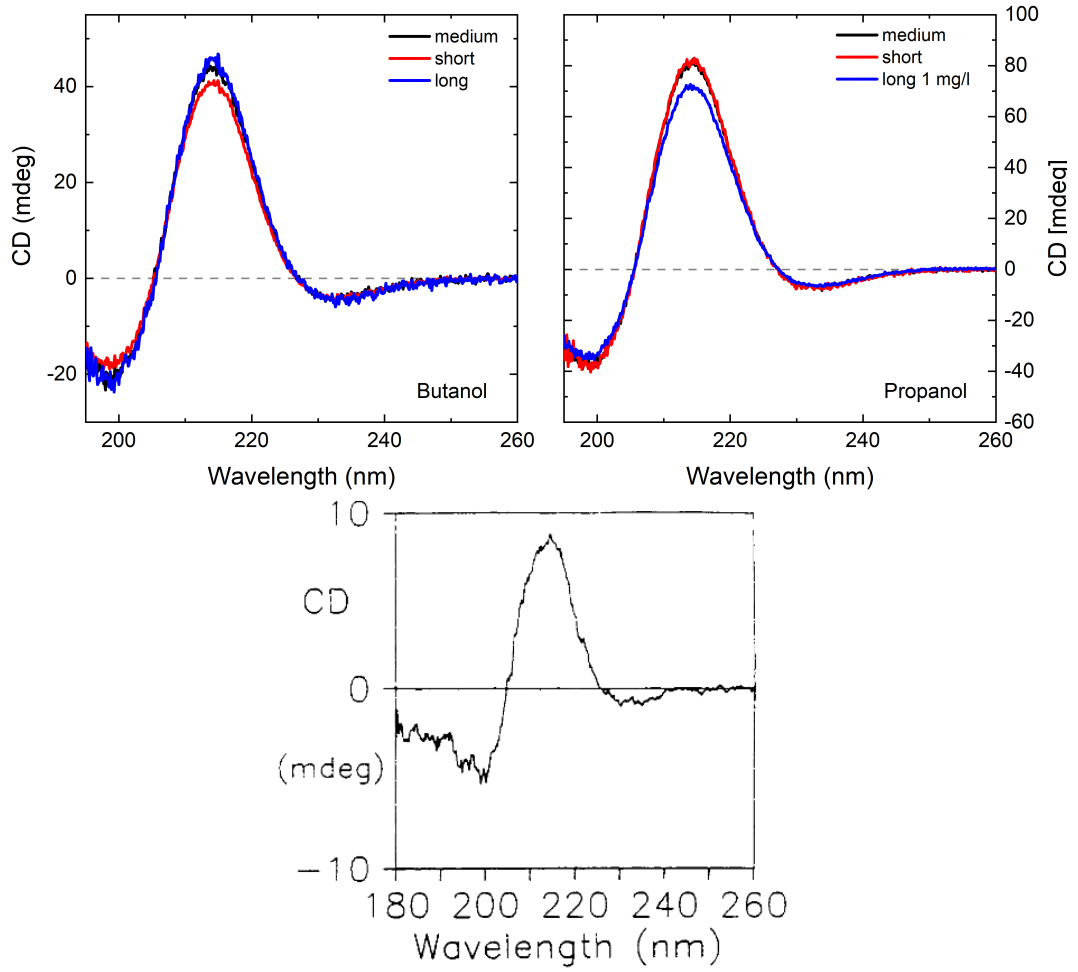
Figure 5.1: ECD spectra for all samples in 1-butanol (top, left) and 1-propanol (top, right). Sample *long* in 1-propanol was also measured at a higher concentration due to weak signal. PPI spectrum in D_2O (bottom), picture adopted from [18].

.

5.2 Solid phase Raman spectra

We used instrument *Alpha 300* from company WITec for acquiring Raman spectra of a raw samples (PPII) and prepared lyophilized samples (PPI) in solid phase.

The spectra were measured from small grains (example in Figure 5.2) on a slide in a single point Raman spectra acquisition mode. After subtracting the background, spectra were normalized to the same integral intensity.

We put small grain of sample on a slide and found it with optical microscope. Then we switched to single point Raman spectrum acquisition mode and found the best signal by small moves of sample.

After processing the spectra, we performed peak analysis with use of second derivative of spectra. We combined subjective interpretation of spectra (and their second derivative) and a tool for peak analysis in software OriginPro 2018. We identified approximately 39 peaks in spectra of each form of polyproline. There were some subtle differences between spectra of the same form and comparison of peaks with relative intensity lower then approximately 0.1 was more problematic.
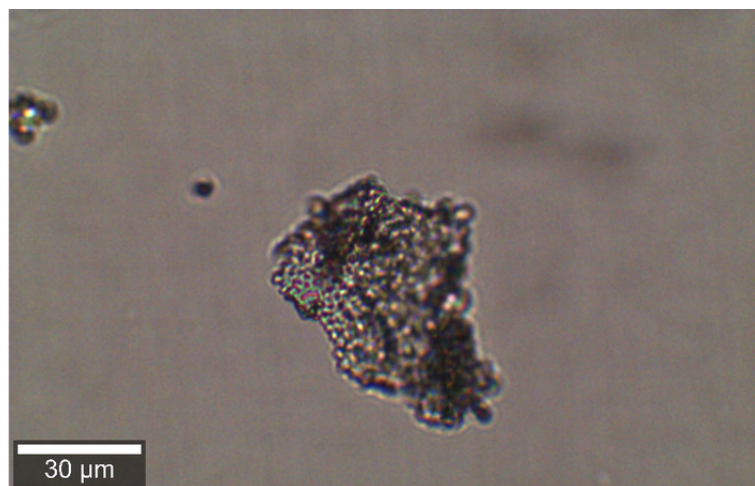
Figure 5.2: Picture of grain of *short* PPII sample in optic microscope, from Raman spectrometer *Alpha 300*.

Spectra with marked peaks are shown in Figure 5.3. Positions of bands of PPI obtained from OriginPro are in Tab. 5.1 Some peaks were impossible to identify in spectra of some samples, these are left empty. In addition, the PPI bands are assigned to the expected vibration movements in the Tab. 5.1 according to calculation of AAAA PPI tetramer. The detail description of calculation is given further (section 5.8.1). We did not deal much with PPII spectral bands, they are important for us mainly as indicators of mutarotation. Therefore, we did not make their vibrational assignment, the position of the bands is only given in the appendix A.2.

Based on the length of polypeptide chain we observed some differences in measured Raman spectra. The largest differences are in the PPII spectra between the long sample and the remaining samples, these different bands are marked with an asterisk in the Fig. 5.3. However, we do not observe similar differences in PPI samples. Therefore, we guess that those differences may be caused by impurities in the purchased *long* sample, which were removed during the PPI preparation process. Another possible option would be a slightly different crystal structure of the purchased samples. In the PPII spectra, we observe a decrease in the signal of a number of bands with increasing sample length. Namely, these are spectral bands: 6, 11, 14, 17, 21, 24, and 34. Apparently, the opposite tendency occurs only in the band 32. Differences also occur in the C-H region where the band 37 slightly shifts.
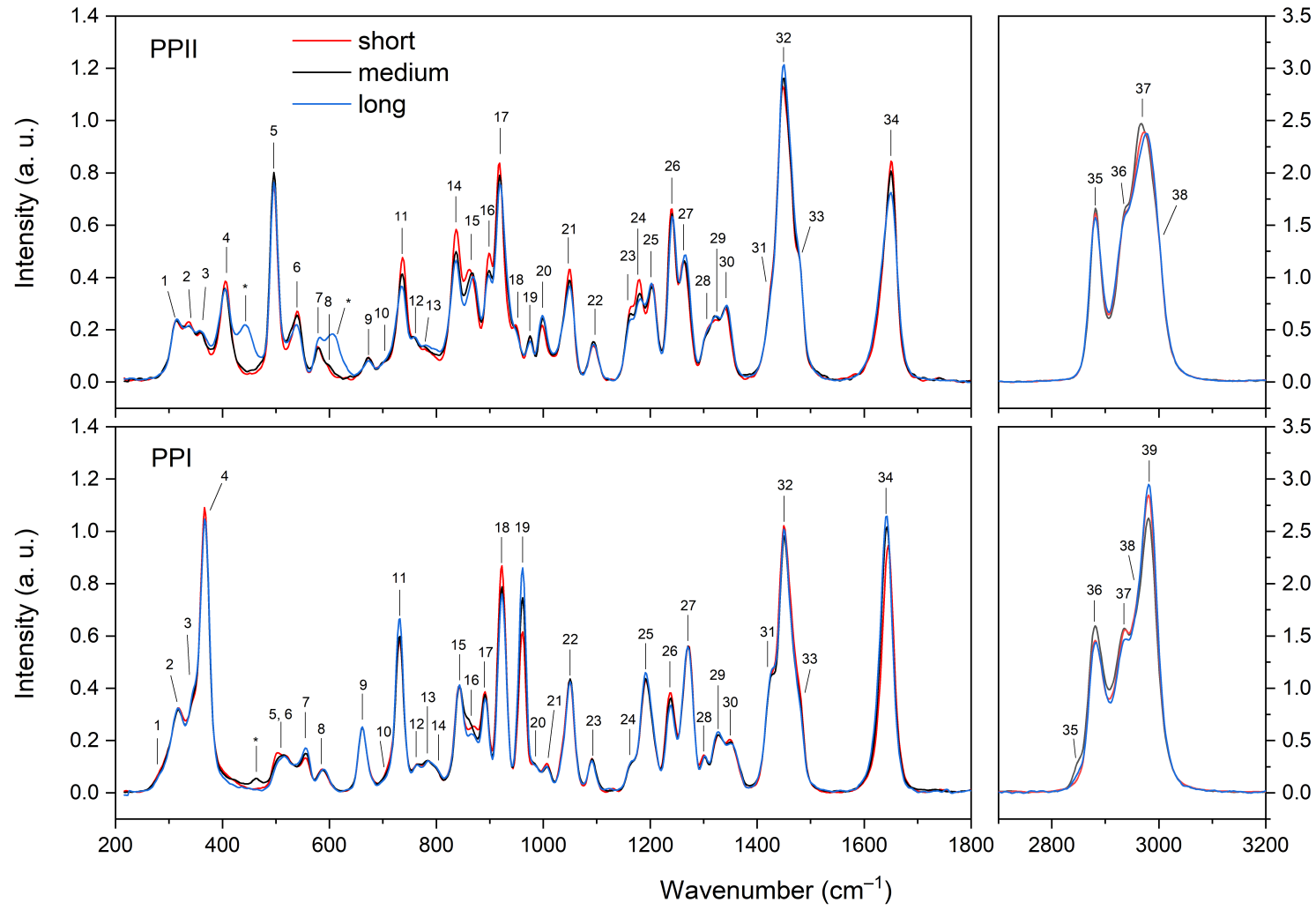
Figure 5.3: Solid phase Raman spectra of all samples in PPII and PPI conformation. The numbers indicate the bands according to Tab. 5.1 and Appendix A.2. Labelling of PPI and PPII bands does not correspond to each other.

Differences in band intensities also occur in PPI spectra. Here, too, the intensity of several bands decreases with increasing length (mainly 18, 26), but this trend is not so obvious. A markedly opposite trend occurs in band 19, where the intensity increases significantly with increasing length. This change was also observed in the calculations, but it was a relatively weak phenomenon. Differences in the shape of the spectrum also occur in the region of the band 16, where in the sample *medium* this band is weaker and shifted. Overall solid phase Raman spectra of PPII or PPI are only minimally changed by length of PP chain.

The differences between PPI and PPII spectra are highlighted in the Fig. 5.4. For PPI conformation, bands 4 ($\approx$367 cm^{-1}) and 19 (962 cm^{-1}) seem to be very characteristic. For PPII conformation, the most characteristic bands are 5 (496 cm^{-1}) and 6 (542 cm^{-1}). We suggest these bands to be used as Raman markers of PPI and PPII conformation. Also interesting is the symmetrical change in the intensity of the pair of bands 26 and 27 in both conformations. In PPI band 27 is approximately 1,4 times more intense than band 26, while in PPII the ratio of intensities of 26:27 is approximately 2:3 In the regions 800-900 and 1150-1350 cm^{-1}, there is an overall change in the shape of the spectrum. We observe significant shift at Amide I vibration (band 34): in PPII it is located at approximately 1652 cm^{-1}, while in PPI it is located at 1643 cm^{-1}. In the C-H region, there is a slight change in the intensity and shape of the spectrum.
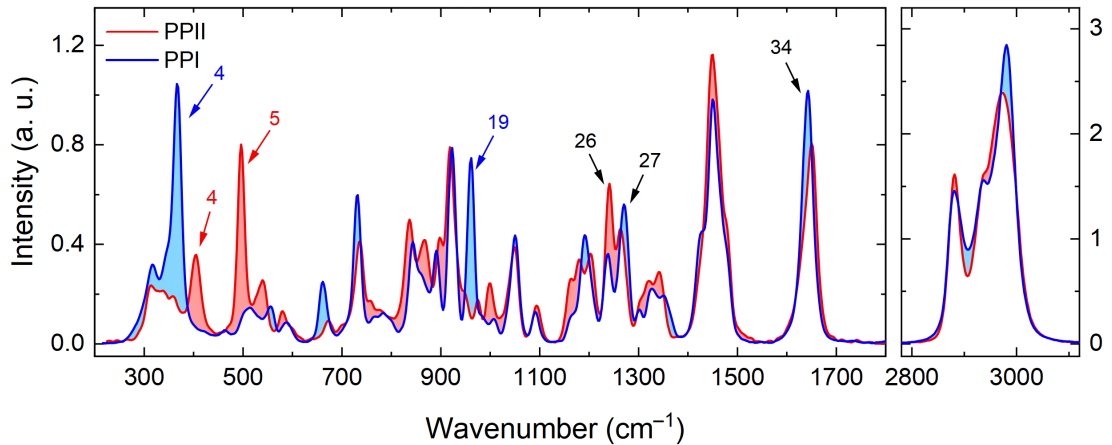
Figure 5.4: Comparison of solid phase Raman spectra of solid phase Raman spectra of medium in PPI and PPII conformation.

Table 5.1: Position of peaks in PPI Raman spectra and their assignment.

No.	*short*	*medium*	*long*	assignment
1	280 w		280 w	γ(C-C-C-C)
2	317 w	317 w	317 w	δ(ring-NCO)op
3	339 w	342 w	342 w	δ(ring-NCO)op
4	366 s	367 s	369 s	δ(ring-NCO)op
5	500 w	503w	503 w	Amid V
6	519 w	514 w	516 w	Amid V
7	557 w	557 w	557 w	δ(CCCC)
8	586 w	589 w	589 w	δ(CCCC)
9	662 w	662 w	662 w	**Amid IV**
10	703 w	700 w	698 w	**Amid VI**
11	732 m	732 m	732 m	**Amid VI**
12	764 w	766 w	764 w	δ(CH$_2$)ρ
13	787 w	784 w	787 w	δ(CH$_2$)ρ
14	798 w	800 w	798 w	δ(CH$_2$)ρ
15	841 m	841 w	841 m	δ(NCO) δ(CH)
16	869 w	864 w	867 w	δ(CH$_2$)ρ ip
17	892 w	891 w	891 w	δ(CH$_2$)ρ
18	923 s	923 s	923 s	δ(CCCC)breathing
19	962 m	962 m	962 s	ν(C$_\alpha$-N), ν(C$_\gamma$ - C$_\delta$)
20	987 w	987 w	989 w	not identified
21	1007 w	1009 w	1007 w	ν(C$_\alpha$ - C$_\beta$,) ν(C$_\gamma$ - C$_\delta$) op
22	1050 m	1050 m	1050 m	ν(C$_\alpha$ - C$_\beta$)
23	1091 w	1091 w	1091 w	δ(CH$_2$)τ, δ(CCCC) ring
24	1162 w	1162 w	1162 w	δ(CH$_2$)τ
25	1191 m	1191 m	1191 m	ν(CN), δ(CH$_2$)τ
26	1236 w	1236 w	1236 w	δ(CH$_2$)τ, δ(C$_\beta$ C$_\gamma$)
27	1273 m	1271 m	1271 m	δ(CH) δ(CH$_2$)
28	1300 w	1302 w	1302 w	δ(C$_\beta$ - C$_\gamma$), δ(C-N)
29	1325 w	1325 w	1325 w	δ(C$_\beta$ - C$_\gamma$), δ(C-N)
30	1352 w	1355 w	1357 w	δ(C$_\beta$ - C$_\gamma$), δ(C-N)
31	1423 m	1423 m	1423 m	**Amid III**, δ(CH$_2$)σ
32	1450 s	1450 s	1450 s	ν(C$_{\beta,\gamma}$H$_2$)σ op
33	1482 w	1482 w	1482 w	δ(C$_\delta$H$_2$)
34	1645 s	1643 s	1643 s	ν(C=O), **Amide I**
35	2845 w		2849 w	ν_s(CH$_2$) ip
36	2879 s	2879 s	2879 s	ν_s(CH$_2$) op
37	2933 s	2933 s	2933 s	ν(CH)
38	2956 s	2955 s	2954 s	ν_{as}(CH$_2$)
39	2983 s	2983 s	2983 s	ν_{as}(CH$_2$)

"s" – strong band, "m" – medium, "w" – weak, "vw" – weak, "sh" – shoulder. Labelling of vibrations is described in section 3.2.4.

5.3 ECD mutarotation

We introduced UV/ECD as complementary method to Raman spectroscopy for monitoring PPI to PPII mutarotation. Water was used as solvent and optimal concentration for measurement was found as 0.25 g/l ($\approx$100 times smaller than needed for Raman). We kept sample at 35 °C in thermal cell.

Due to high consumption of nitrogen it was not possible to measure continuously during the whole mutarotation process which is estimated to take approximately 90 hours. First continuous measurement lasted approximately 10 hours. Then shorter measurements were performed every morning for the three next days.

During data processing by factor analysis we found that stabilization of instrument for UV/ECD measurement was not consistent across measurements. This can be shown e. g. on weights of second factor component of UV spectra (Fig. 5.5). A detailed view of first shorter spectral series (Fig. 5.5, inset) shows that there was a process of stabilization and system wasn't at equilibrium at first. Therefore, few spectra were removed from second, third and fourth spectral series to reduce effect of this process. This change in dataset didn't change the factor dimension of problem, which was found to be 3 (Fig. 5.6). As evident from the course of singular values and residual errors, factor dimension before reduction was 4-5, while after deletion of datapoints corresponding to instrument stabilization is reduced to 3.

From the course of the singular value a and the residual error, it is evident that the change was mainly due to a reduction of factor components orders of magnitude weaker than the first factor component. This may correspond to the removal of the spectra most affected by the parasitic phenomenon.
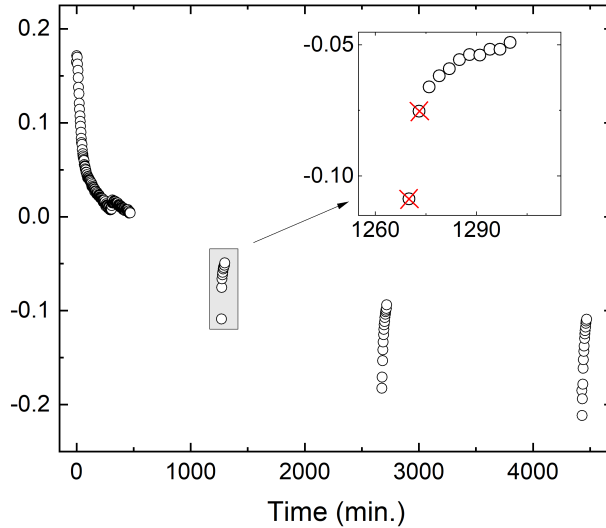
Figure 5.5: Coefficients of second factor component before dataset adjustment.

The influence of instrument stabilization was even more pronounced in ECD spectra. The coefficients of higher spectral components from different spectral series were systematically shifted to each other and badly corresponded to kinetics. Therefore, it was not possible to perform a complete fit on the CD spectra, because the data from later measurements differed significantly from the model

used (this is evident in the third subspectrum in the Fig. 5.8). Therefore, the fit was performed with fixed parameters of reaction rates $k_1 = 4.8 \cdot 10^{-4}\,\mathrm{s}^{-1}$, $k_2 = 2.3 \cdot 10^{-5}\,\mathrm{s}^{-1}$ obtained from the UV spectra (corresponding half-lives are $\tau_1 = 0,4$, $\tau_2 = 8,4$ hour). Plots showing fit of UV data can be found in the appendix A.1.
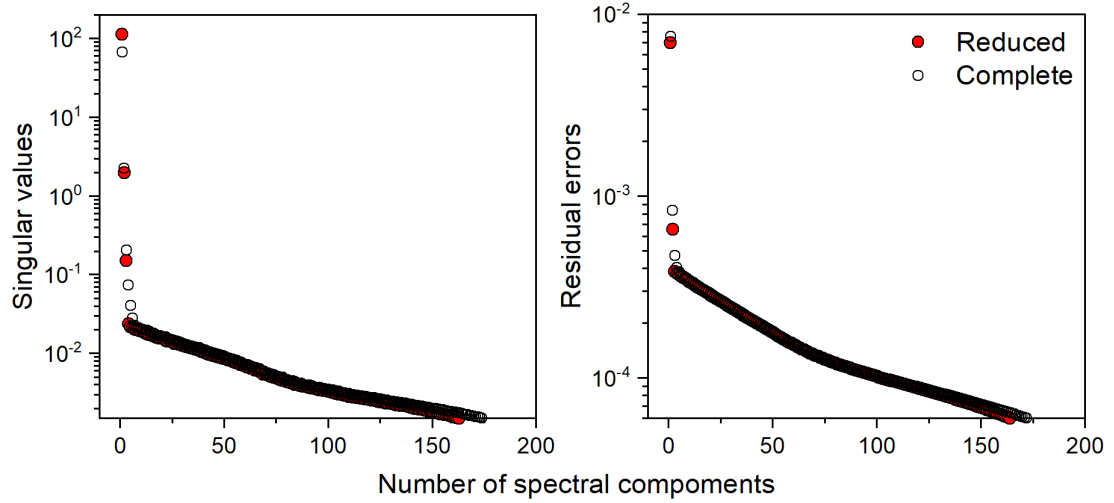
Figure 5.6: Comparison of courses of singular values and residual errors in complete and reduced dataset. The main change is taking place in higher subspectra.
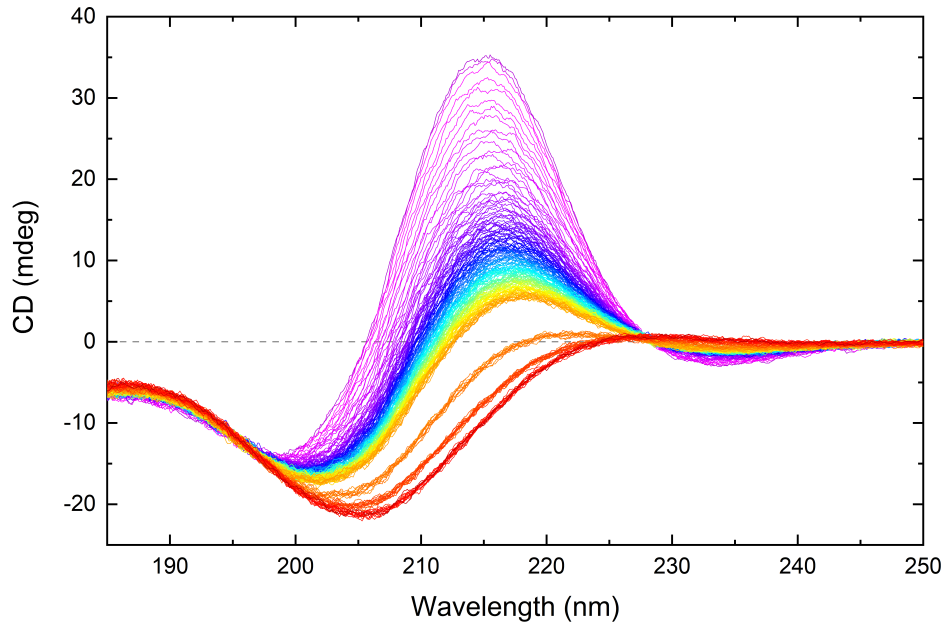
Figure 5.7: ECD spectral series measured for 4 days. Large gaps between spectra are caused by measurement interruptions. The first measured spectrum is purple, the last red.

The entire measured ECD spectral series is shown in the Fig. 5.7. The transition of the dominant positive band at 215 nm typical of PPI to the markedly negative band at 205 nm characteristic of PPII is clearly visible. The fastest transition occurs in the first hours. However, even between the third and fourth

day, the spectral shift corresponding to the conversion of PPI to PPII is still visible. The observed course corresponds to the measurements published by Dukor, et al.[18]

Fig. 5.8 shows the individual subspectra (spectral components) obtained by factor analysis and their representation (coefficients V_{ij}) in individual measured spectra. The first subspectrum has the meaning of the average spectrum, the second spectrum corresponds to the main spectral change and in a simplified manner it represents the shift between the positive band of PPI (215 nm) and the significant negative band of PPII (205 nm). The third subspectrum represents a more subtle change, its coefficients show the already mentioned problem caused by the interruption of the measurement. The fourth subspectrum then apparently represents only noise. This is evident from both the shape and the more or less random weights.

The spectra of pure components reconstructed on the basis of result of factor analysis and subsequent for of reaction kinetics mechanism (left) as well as their relative concentrations are shown in Fig. 5.9.

Spectrum of pure first component corresponds to pure PPI conformation in aqueous solution. This is well confirmed when comparison to spectrum of PPI measured in 1-butanol is made, as shown in Fig. 5.10. Spectrum of the third component appears to be similar to those of pure PPII conformation (e. g. in[18]). We were able to distinguish even the weak positive band at approximately 224 nm.

The spectrum of second factor component is referred as "metastate". However, it does not match any specific intermediate present during process of mutarotation. The process has more intermediates but we are not able to distinguish them by the method used. The spectra of all these intermediates are represented by the spectrum of metastate.

From the development of concentrations, it is evident that even after almost two days, not all the sample is still converted to the PPII state, which corresponds to the observed change between the third and fourth day of measurement. At the same time, however, the concentration of molecules in the pure PPI conformation appears to decrease rapidly. The rapid decrease in the concentration of the pure PPI conformation is consistent with the results of Shi et el. obtained on shorter peptides (length 13) using mass spectrocopy.[47]

In summary, UV absorption and ECD spectroscopy is able to distinguish 3 main components during PPI to PPII mutarotation in water. The first and last components correspond to the pure PPI and PPII conformations respectively. The second component represents an "average intermediate".

Generally, ECD is method sensitive to the conformation of larger structure elements, therefore, methods sensitive to more local arrangements may be able to distinguish more spectral components.
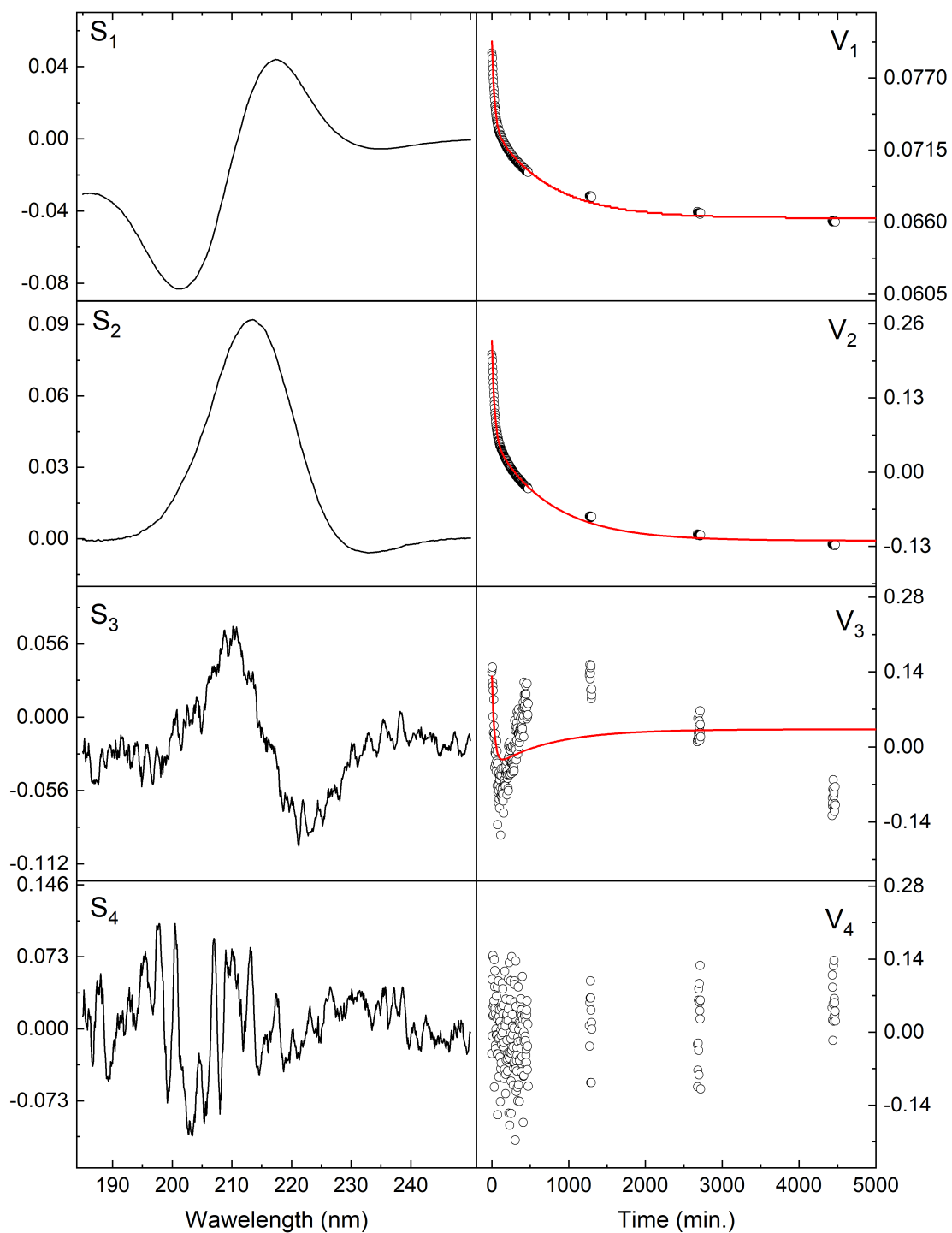
Figure 5.8: Results of SVD of ECD mutarotation measurements. The first four subspectra are shown on the left. Their corresponding coefficients along with the values from the fit are shown on the right.
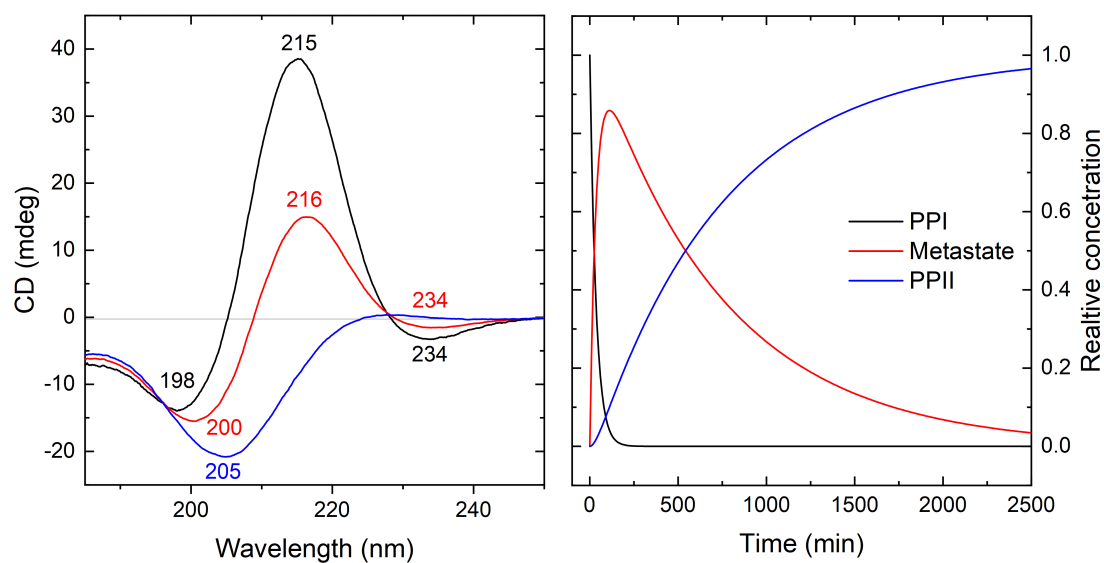
Figure 5.9: ECD spectra of pure components from three-stage fit of mutarotation reaction mechanism from factor analysis (left). The relative concentrations of those components (right).
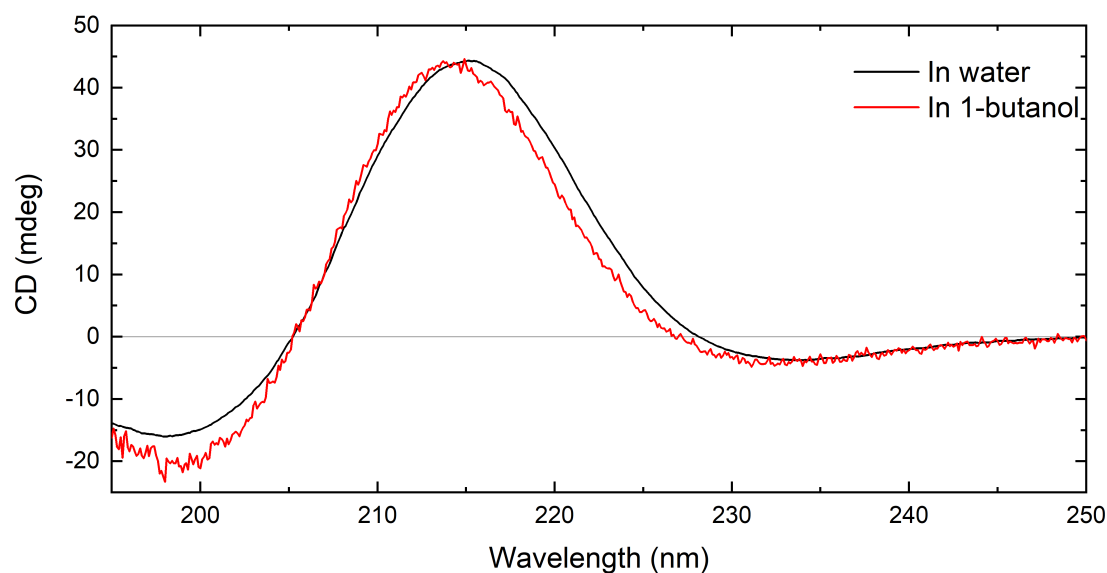
Figure 5.10: Comparison of ECD spectra of sample *medium* in PPI conformation in water and 1-butanol; solvatochromic shift is visible.

5.4 Raman and ROA spectra of PPI

In this section we show results of four experiments. Two of them represent measurement in organic solvents (both 1-propanol and 1-butanol) and other two measurement in aqueous solution (PPa, PPc) at low temperature. We used three samples dissolved in water (PPa, PPc) and one sample dissolved in propanol (PPe). As discussed in Material and methods the solubility of our sample was not satisfying. For each experiment we tried to obtain the highest concentration possible. Sample PPa was prepared in concentration 10 mg/ml, PPc in 20 mg/ml, PPe 10 mg/ml and supernate after centrifugation was used for measurement.

5.4.1 Measurements in 1-propanol and 1-butanol

Measurement in organic solvents appears to be advantageous because PPI is stable in them and no mutarotation occurs. On the other hand, compared to water, organic solvents have a much stronger Raman signal. They have a higher effective Raman cross section than water, so less beam intensity and longer data collection times are required. Another problem is caused by the fact that solvent molecules forming solvation shell interact with the polyproline chain and change both the signal of part of the solvent and the polyproline. The intense signal of the solvent, together with changes in its spectrum when interacting with the molecule, significantly complicate the subtraction of the solvent spectrum. Intense solvent bands can practically make it impossible to acquire PPI spectrum at a given location. ROA is a chiral method and solvents that are not chiral should not affect its signal (both 1-propanol and 1-butanol are not chiral). In practice, however, interference can occur in this case as well.

The spectra measured in 1-butanol and 1-propanol are shown in Fig. 5.11 and 5.12, respectively. The measurement in butanol lasted for approximately 100 hours. Solvent signal was approximately two orders of magnitude stronger than the signal of PPI itself. 1-butanol has relatively rich Raman spectrum with several narrow intense bands, which hinder correct solvent subtraction in some spectral regions (grey areas, middle panel of Fig. 5.11). Solute Raman spectrum of PPI (obtained by solvent subtraction) shares the same spectral features we found in solid phase Raman spectra: spectral markers band at $336 \, \text{cm}^{-1}$(band 4) and $963 \, \text{cm}^{-1}$ (band 19) are clearly visible. The area between $800 \, \text{cm}^{-1}$ and $1000 \, \text{cm}^{-1}$ has slightly deformed baseline due to subtraction of the solvent. Similarly, the intensity of bands 31-33 at 1451 cm^{-1} seems to be influenced by the strong band of the solvent at the same place.

Thanks to the long accumulation time, the ROA spectrum of PPI is of superior quality. To our knowledge, there is no evidence of PPI spectrum in literature. The spectrum contains over 40 well resolved bands which may be assigned to specific vibrational motives (see Tab. 5.2 in section 5.4.3). The most specific features characteristic for PPI signal appear to be the trinity of bands in low-frequency region (279, 309, and $327 \, \text{cm}^{-1}$) forming easily recognizable V-pattern. Other characteristic features seem to be two +/- couplets at $1044/1058 \, \text{cm}^{-1}$ and $1229/1273 \, \text{cm}^{-1}$ together with the overall spectral shape in 850-1100 cm^{-1} region.
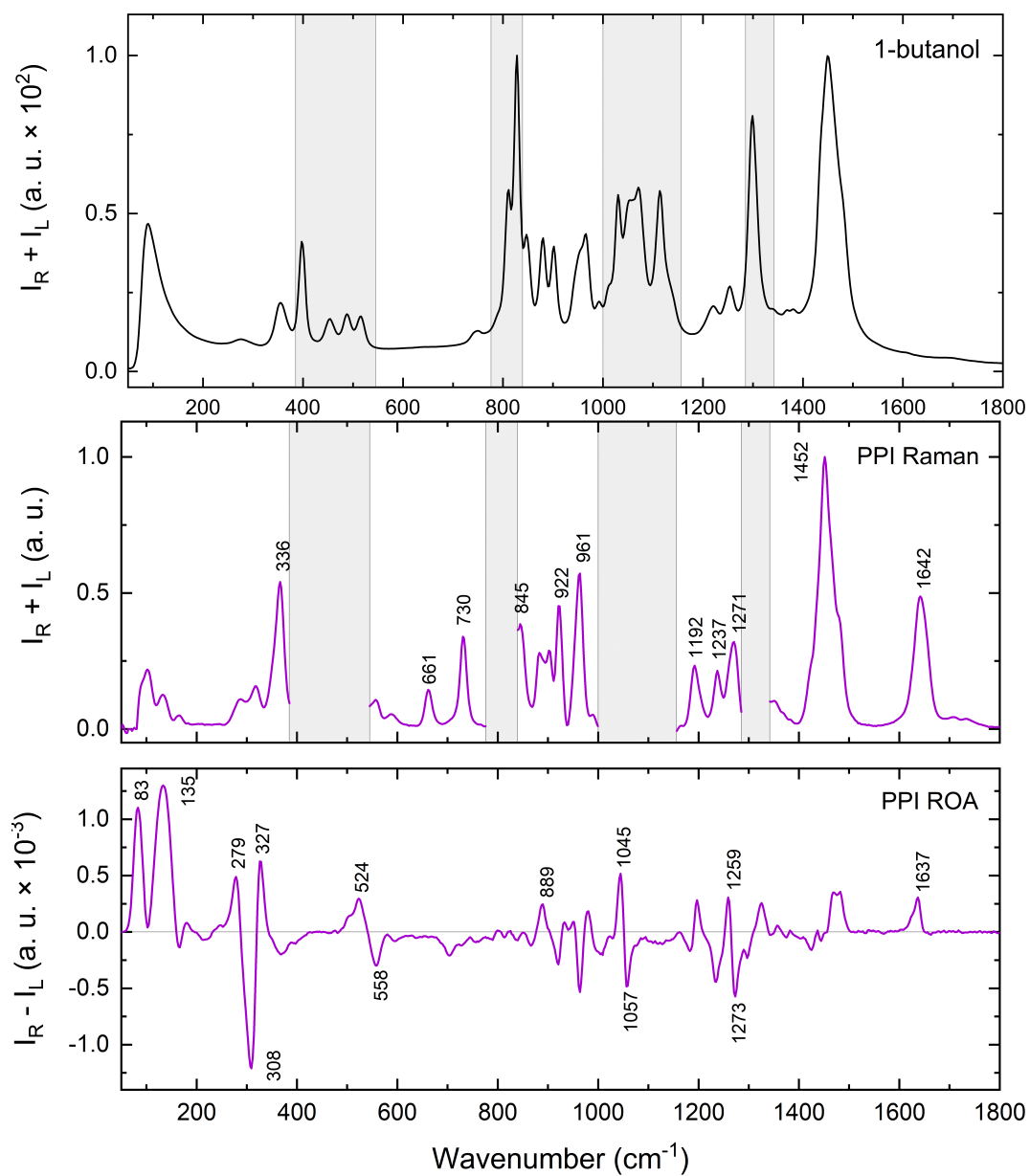
Figure 5.11: Measurement of PPI in 1-butanol (data from 2011). Spectral regions strongly interfering with the solvent signal are omitted in Raman spectrum.
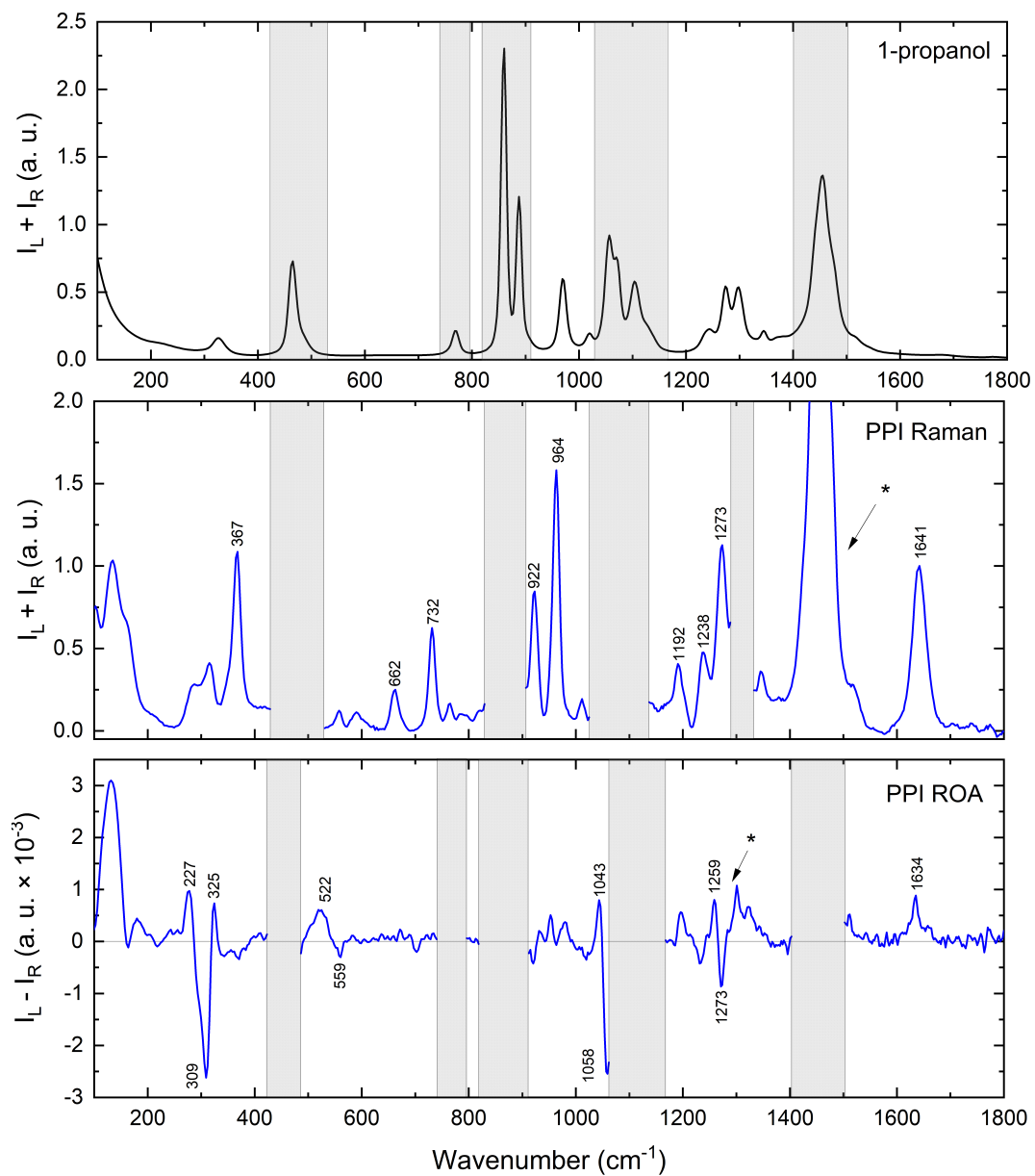
Figure 5.12: Measurement of PPI in 1-propanol (PPe sample). Spectral regions and bands strongly interfering with the solvent signal are either omitted or marked with an asterisk.

The measurement in 1-propanol took much shorter time, about 10 hours. This negatively affected the quality of obtained Raman and ROA spectra. The subtracting of the solvent from Raman data was rather challenging. Similarly to the measurement in 1-butanol, some spectral regions (strongly interfering with the solvent signal) have to be omitted in final PPI Raman spectrum (fig.) The baseline and relative intensities of several bands were also affected (e. g. bands at 922, 1273, band 32 - asterisk). Unfortunately, ROA PPI spectrum was also deformed by strong solvent signal. To acquire reasonable results, it was necessary tu subtract solvent Raman signal from PPI ROA data, however, even after that several spectral regions had to be omitted. Still, the final ROA PPI spectrum bears the same main features we obtained in those acquired in 1-butanol, namely the characteristic V-pattern.

Acquisition of the higher quality data would certainly need significantly higher accumulation time and neglect the first few hours of data acquisition, where the residual fluorescence is bleached, and system is equilibrized.

5.4.2 Raman and ROA measurements of PPI in water

During ECD measurements, we confirmed that the mutarotation kinetics is very slow at low temperatures. At $5\,°C$, the isomerization is considered to be so slow that dozens of minutes of measurements should be possible prior to any significant spectral (and system) change. Therefore, several measurements of Raman spectra and ROA were performed at temperature $5\,°C$.

At first, we analyse Raman measurements of samples PPa and PPc. Both were measured at $5\,°C$. The PPa sample was measured for 48 minutes and then removed from the instrument due to need to check another sample. Then it was returned and measurement continued for another 38 minutes. Sample PPc was measured for 164 minutes continuously. Different acquisition times of single spectra were used for each measurement.

Spectra were recorded immediately ($\approx 5\,\text{min}$) after dissolution in water, to be able to track changes induced by aqueous environment. However, the first accumulations suffered from residual florescence which slowly bleached during measurement. To prevent alterations, the solvent spectrum subtraction and baseline correction was performed on the whole set using singular value decomposition. However, the baseline correction was rather complicated due to strong fluorescence and obtained spectra may, therefore, suffer from slightly deformed baseline in lower frequency region ($100\text{-}400\,\text{cm}^{-1}$)

If we look at coefficients V_1 for first factor component S_1 from performed SVD (Fig. 5.13), it can be seen that absolute changes are small. However, SVD is sensitive to such small changes as removal and return of sample. This change does not affect obtained spectra but poses a problem for intended study and fit of reaction kinetics.

After the baseline correction, Raman spectra from all three measurements look similar (Fig. 5.14). There are some small differences related to inaccuracies mostly by baseline corrections (in the picture highlighted by red arrows), however, bands characteristic for PPI, are easily trackable (see section 5.2). We do not discuss the time dependence because this data set is made from data from separate measurements.

As a final Raman spectrum of PPI in aqueous solution we took the first factor component of the whole dataset because its practically average spectrum of all measured spectra, however with lower noise. This spectrum is in lower part in Fig. 5.14 and it can be seen that it well corresponds the whole dataset. However, it should be noted that slight change of intensity of band 19 shows, that our system is not static and isomerization from PPI and PPII runs, albeit very very slowly.
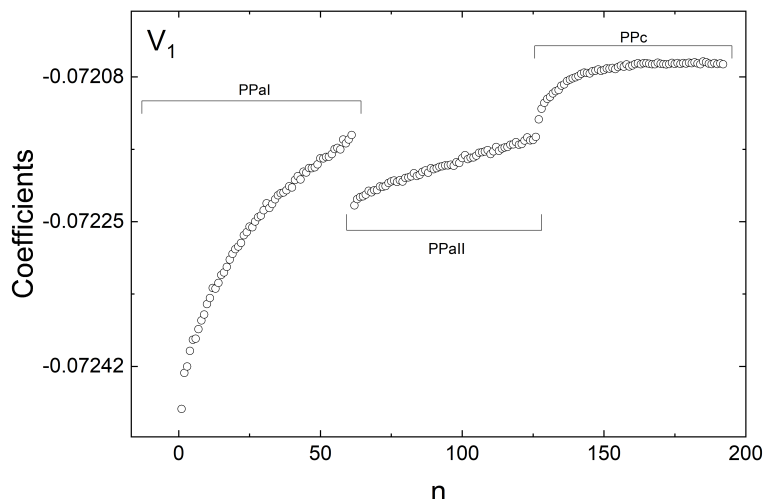
Figure 5.13: Coefficients of the first factor component. Absolute changes are small but separate measurements are distinguishable.

Similar data treatment and analysis were performed for CH-region and results are shown again in Fig. 5.14. It can be seen again that measured spectra are similar across measurements and that first factor component is similar to these spectra. When comparing to differences in CH-region observed for solid phase (Fig. 5.3), we can again say that process of mutarotation is slowly beginning.

The PPa sample was further measured for about 33.5 hours at 5 °C. Figure 5.15 shows the measured spectra after solvent subtraction and baseline correction. We observe decrease of intensities of the characteristic PPI bands (4, 11, 19) and we also see an increasing band 5 of PPII. Newly, there is an increase in intensity of band at $300 \, \mathrm{cm}^{-1}$, which can be attributed to band I of PPI. In the CH region, compared to the beginning of the isomerization, the change of the band 39 is mainly visible, while the band 37 practically does not change.

The measurement was continuous without any interruption, therefore, the SVD could be performed. The Fig. 5.16 shows the course of first 100 (of 460) singular values and residual errors. The factor dimension of the set can be estimated to 3 or 4. This is supported also by the course of coefficients V_4 which are not random (Fig. 5.18). Therefore, we performed two fits to reaction kinetics considering both 3- and 4-state reaction mechanisms.

Even in the case of 3-state process, fitting was quite difficult and failed at first when entire spectral range was considered. The fit was therefore performed on a shorter spectral interval (906 - 1022,cm^{-1}) to obtained reaction rates ($k_1 = 1.5 \cdot 10^{-4} \, \mathrm{s}^{-1}$, $k_2 = 1.3 \cdot 10^{-5} \, \mathrm{s}^{-1}$), subsequently fit was then performed on the whole spectral range 100-1800 with fixed reaction rates, resulting in the finding the approximate values of the matrix $\boldsymbol{G}$ (see section 4.6). A final fit was performed
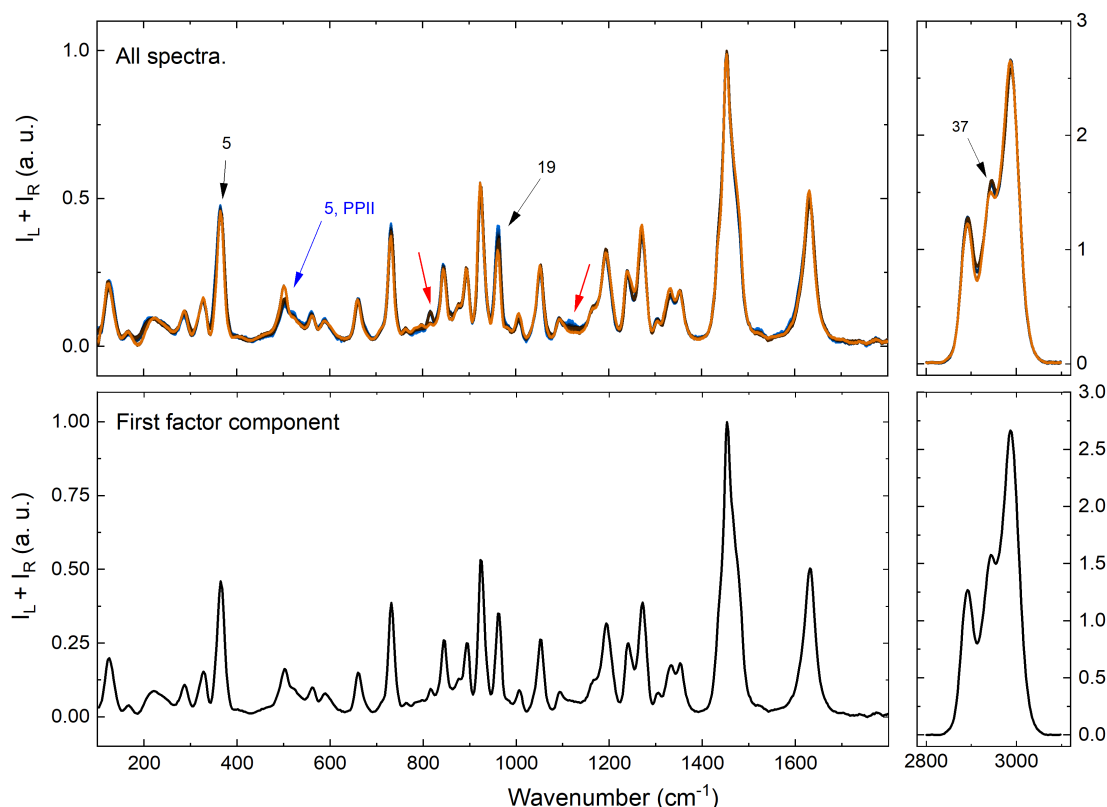
Figure 5.14: Spectra obtained from measuring PPa and PPc at 5 °C. All obtained spectra are in upper panel, first spectral component is shown in lower panel. Red arrows point at differences related to baseline correction, red picture points at characteristic bands for PPI previously identified at solid state sample (numbers notes previously used bands for PPI), blue arrows show bands characteristic for PPII.

with relaxed reaction rates and $\boldsymbol{G}$ matrix values, resulting in final reaction rates ($k_1 = 6.8 \cdot 10^{-5}\,\mathrm{s}^{-1}$, $k_1 = 1.42 \cdot 10^{-5}$), corresponding to half-lives $\tau_1 = 2.8\,\mathrm{h}$ and $\tau_2 = 13.6\,\mathrm{h}$, respectively. These fixed values of reaction rates were then used for a separate fit of CH region. When compared with the fit performed on ECD data, it is evident that the kinetics are much slower, which is of course due to the much lower temperature. The fitted values correspond very well to the courses of coefficients V_{1-3}, as shown in Fig. 5.18. Fit in the CH area, however, responded significantly worse in course of third subspectrum (not shown).

We also tried to fit to 4-state reaction mechanism. After overcoming several obstacles, we successfully reached reasonable agreement with the experiment ($k_1 = 3 \cdot 10^{-4}\,\mathrm{s}^{-1}$, $k_2 = 2 \cdot 10^{-5}\,\mathrm{s}^{-1}$, $k_3 = 5 \cdot 10^{-6}\,\mathrm{s}^{-1}$), however overall agreement between the fitted data and the 4-stage model was not better than in the case of the 3-stage model. Especially, the fit of the V_4 coefficient course was not satisfactory. Therefore, in the following text we will stick to the simpler 3-stage reaction mechanism model.

Figure 5.17 shows the obtained spectra of pure components from 3-stage fit of the reaction mechanism. The first component corresponds to the spectrum of pure PPI in H_2O. The second – "metastate" – component also resemble the PPI spectrum, however, intensities of several characteristic bands are notice-
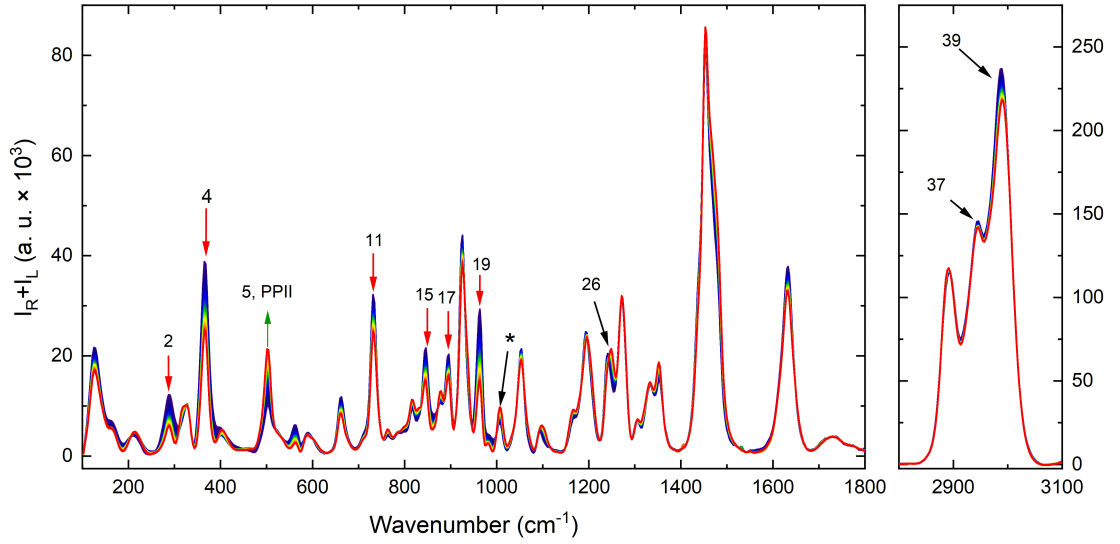
Figure 5.15: Raman spectra obtained during measurements at 5 °C, sample PPa. The numbers indicate the characteristic changing bands.
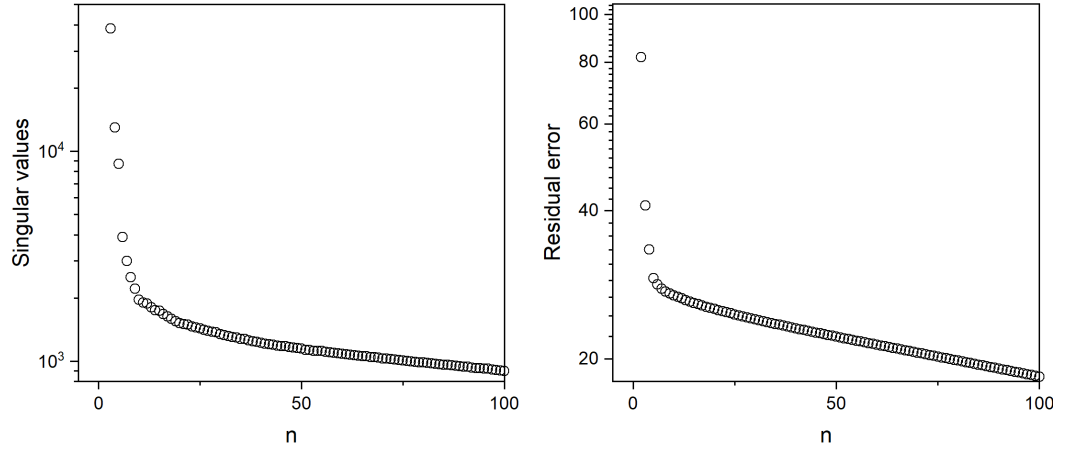
Figure 5.16: Singular values and residual errors from SVD. Raman spectra from long measurement at 5 °C, sample PPa.

ably lowered. Therefore, we estimated the PPI content to approximately 90% in metastate. The final state spectrum bears features characteristic for both PPI and PPII conformation, the PPI content could be approximately 50%.

We would like to stress out the remarkable feature of our approach utilising factor analysis (singular value decomposition). Based on the reaction mechanism we are able to reconstruct spectra of pure components which would correspond to situation prior or after the actual measurement.
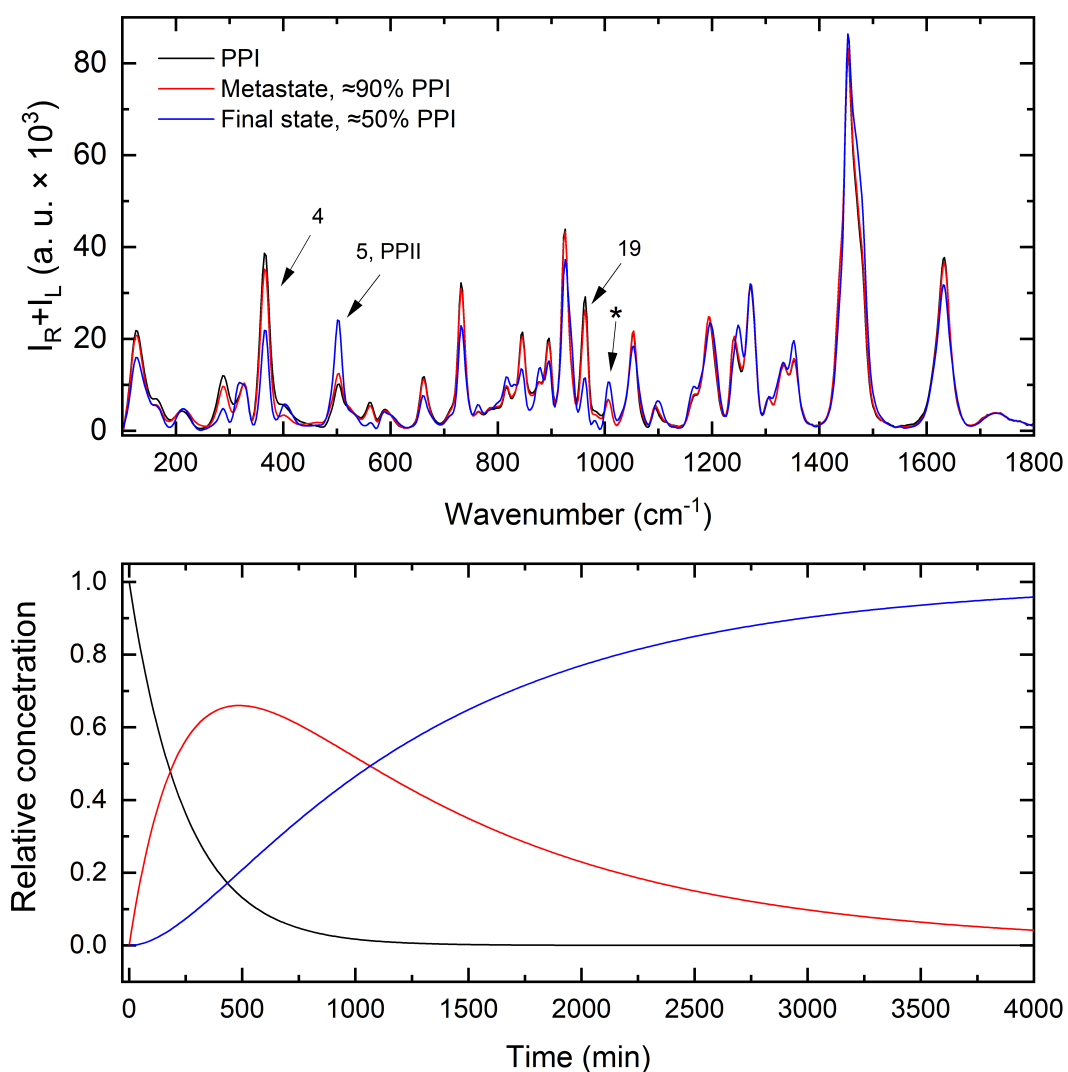
Figure 5.17: Top – Raman spectra of pure components from the 3-stage fit if 33.5 hours PPI measurement in H_2P at 5 °C. The numbers indicate the bands that indicate that the determined PPII spectrum corresponds more to the PPII than the last observed spectrum. Bottom – Related concentrations of pure components during measurement.
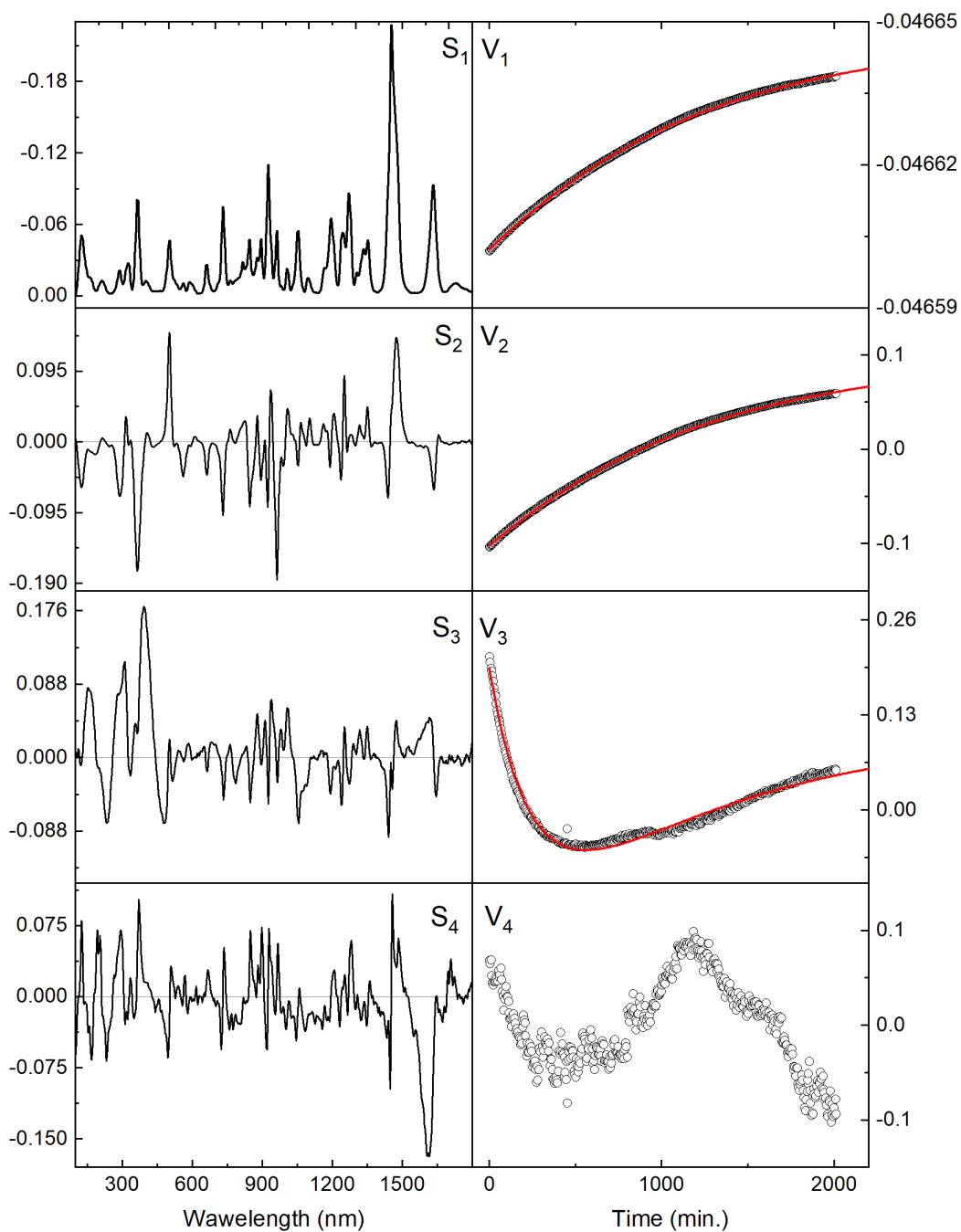
Figure 5.18: SVD of Raman spectra from 33 hours measurement at 5 °C (PPa sample). First four spectral components (left) and their coefficients (right). along with their fits according to the 3-stage reaction mechanism (red lines).

During the above described measurement (PPa 48 + 38 min, PPc 131 min, and PPa 33,5 hours) we also recorded the ROA signal. Because ROA is a significantly weaker phenomenon, longer data collection times are needed to obtain quality spectra. Unfortunately, it turned out that at the concentrations we were able to prepare, the ROA signal was relatively weak and the collection times required for a reasonably resolved spectrum were around one hour. We tried to obtain ROA spectra from two sets of data. The first spectrum was created by averaging the spectra from the beginning of the measurement of the PPa and PPc sample (first hours of measurement). The second spectrum was made by averaging a longer data collection at $5\,°C$. Nevertheless, it should be noted that a significant part of the pure PPI content has already changed to some intermediates and possibly to PPII during data collection as evidenced by above SVD analysis and fit of corresponding Raman spectra dataset.
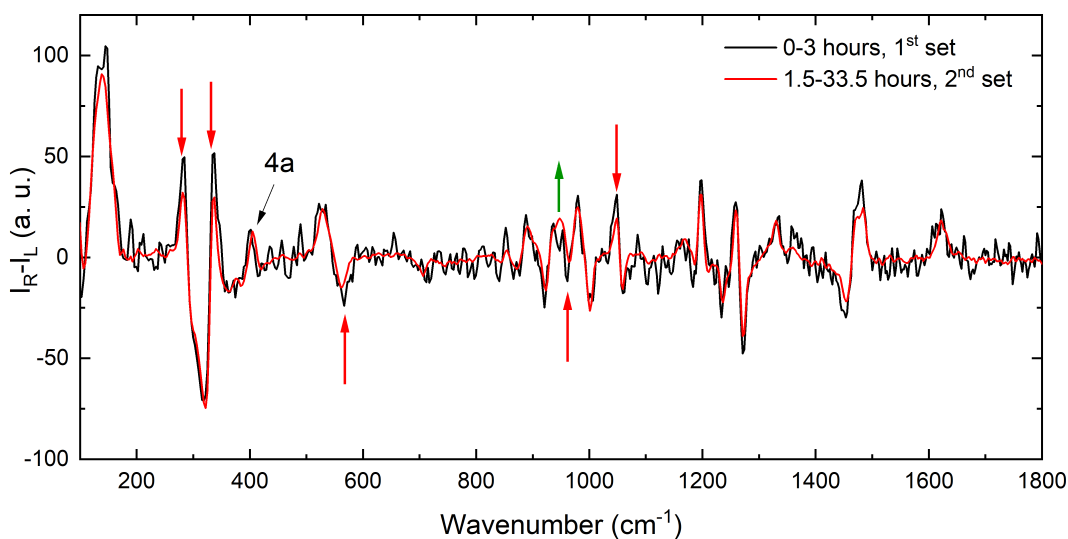
Figure 5.19: Comparison of ROA spectra. obtained from measurement in H_2O at $5\,°C$.

A comparison of those two ROA spectra can be found in Fig. 5.19. Unsurprisingly the spectrum from the first set is significantly noisier than spectrum from the second set. This is due to the significantly shorter data collection time for the first dataset. Otherwise both spectra overlap very well. However, there are several differences which cannot be attributed just to the elevated noise level (marked by arrows). These changes can be explained by the ongoing mutarotation, which is more pronounced in the spectrum with a longer data collection time. Specifically, we observe decrease of intensities of the positive bands at 297 and $327\,cm^{-1}$ forming the characteristic "V-pattern", decrease of positive band at $1045\,cm^{-1}$ and negative bands at 558 and $964\,cm^{-1}$, and slight increase of band at $948\,cm^{-1}$.

It is also worth mentioning that the highlighted band (4a) overlaps very well in both spectra. However, it does not occur in the spectra obtained in 1-propanol and 1-butanol (see Fig. 5.21). We will further see that a similar band occurs in the same place in the ROA spectrum of PPII. An explanation is therefore offered that a certain disturbance of the regularity of the PPI structure toward PPII occurs immediately after dissolution in water.
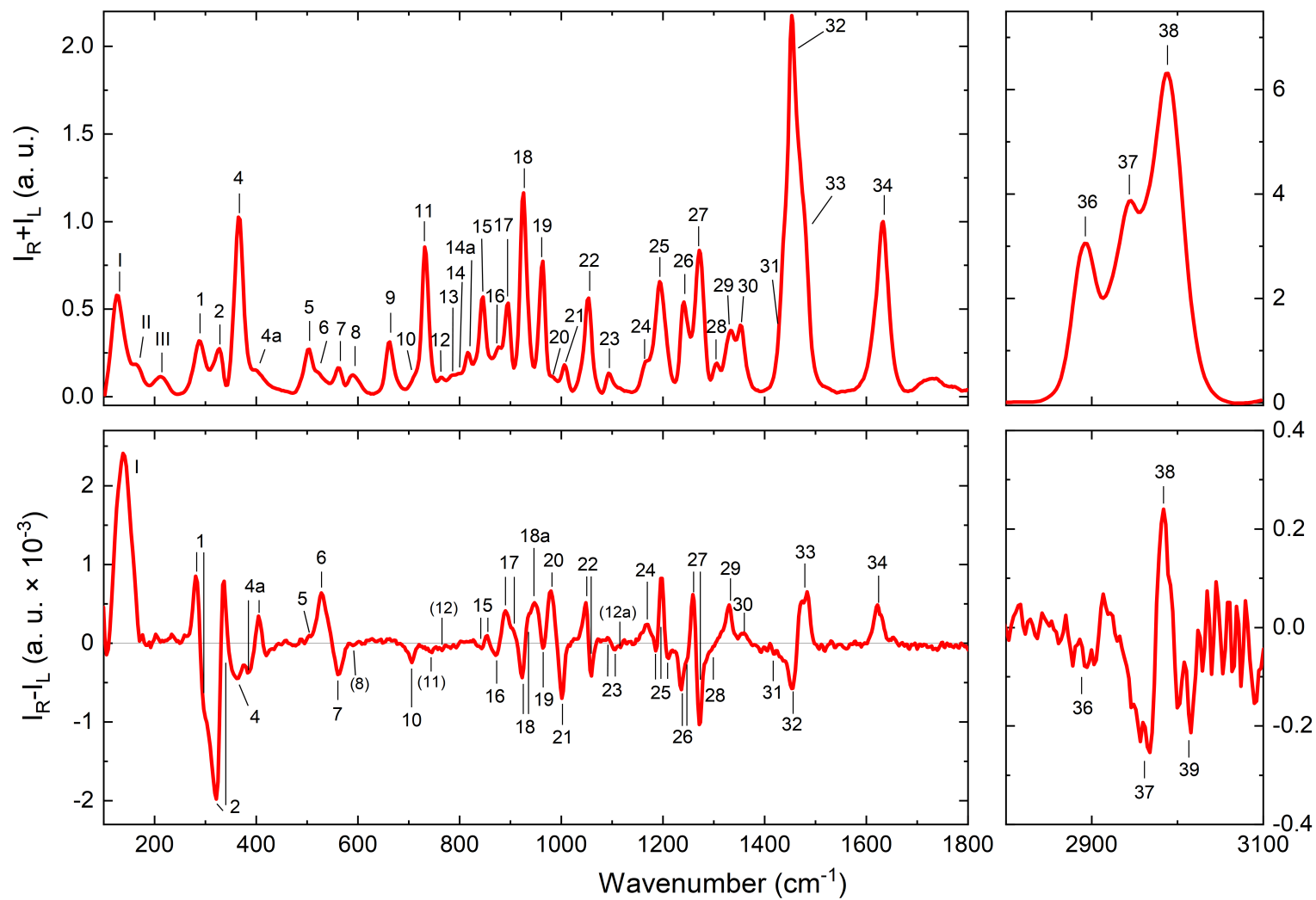
Figure 5.20: Raman and ROA spectra of PPI obtained from measurement at 5 °C in water, sample PPa. Bands labelling corresponds to the Table 5.2.

Table 5.2: Positions of bands identified in Raman and ROA spectra of PPI in solid phase, water, and 1-butanol.

No.	Solid Raman	Water Raman	Water ROA	Butanol Raman	Butanol ROA	Vibration movement
I	x	125 m	139	135 w	134 s+	$\gamma(C_\beta, C_\gamma, C_\delta)\rho$ ip
II	x	167 w	-	167 w	167 w-	$\gamma(C_\gamma, C_\delta)\rho$ op
III	x	215 w	-	-	217 w+	$\gamma(CCCCN)$
1		287 w	280m+/298s-	283 w	279 s+	$\gamma(C-C-C-C)$
2	317 w	328 w	320s-/337m+	319 m	308s-/327s+	$\delta(ring-NCO)op$
3	342 w	-	-	-	-	$\delta(ring-NCO)$ op
4	367 s	365 s	363 w-	366 s	368 m-	$\delta(ring-NCO)op$
(4a)	-	405	386 w-, 406 m+		396 sh-	(PPII)
5	503 w	501 w	506 w+	-	506 sh+	**Amid V**
6	514 w	528 w	529 m+	-	524 m+	**Amid V**
7	557 w	561 w	562 m-	559 w	558 m-	$\delta(CCCC)$
8	589 w	590 w	589 wv-	588 w	595 w-	$\delta(CCCC)$
9	662 w	663 w	-	661 w	-	**Amid IV**
10	700 w	706 w	706 w-	-	705 w-	**Amid VI**
11	732 m	731 s	744 vw-	730 m	727 w-	**Amid VI**
12	766 w	765 w	767 vw-	x	758 w-	$\delta(CH_2)\rho$
13	784 w	785 w	-	x	-	$\delta(CH_2)\rho$
14	800 w	797 w	-	x	800 w+	$\delta(CH_2)\rho$
14a		816w	-	x	823 w+	
15	841 w	846 m	841 w-, 855 w+	847 s	852 w+	$\delta(NCO)$ $\delta(CH)$
16	864 w	874 w	872 w-	-	866 w-	$\delta(CH_2)\rho$ ip
17	891 w	896 m	892 w-, 905 sh+	892	889 m+; 902 sh+	$\delta(CH_2)\rho$
18	923 s	926 s	922m-/934m+	923 s	920m-/934m+	$\delta(CCCC)$ breathing
18a	-	-	948 s+	-	951 m+	$\delta(CCCC)$ breathing
19	962 m	964 m	964 w-	961 s	964 s-	$\nu(C_\alpha-N), \nu(C_\gamma - C_\delta)$
20	987 w	988 w	979 s+	990 w	980 s	not identified
21	1009 w	1006 w	1001 s-	x	1005 m-	$\nu(C_\alpha - C_\beta), \nu(C_\gamma - C_\delta)$ op
22	1050 m	1054 m	1048m+/1060m+	x	1045s+/1057s-	$\nu(C_\alpha - C_\beta)$
23	1091 w	1093 m	1093w-/1105w-	1192	-	$\delta CH_2)\tau, \delta(CCCC)$ ring
23a	-	1116 sh	1116 vw-	-	1121 w-	$\delta(C_\delta - N)$
24	1162 w	1163 w	1168 m	1162 w	1162 w+	$\delta(CH_2)\tau$
25	1191 m	1193 m	1187w-/1198s+ /1209 w-	1192m	1183w-/1198s+ /1215 sh-	$\nu(CN)$ $\delta(CH_2)$ τ
26	1236 w	1241 m	1236m-/1247sh-	1237 m	1234m-/1247sh-	$\delta(CH_2)\tau, \delta(C_\beta C_\gamma)$
27	1271 m	1274 s	1256s+/1274s-	1271 m	1259s+/1273s-	$\delta(CH), \delta(CH_2)$
28	1302 w	1303 w	1294 sh-	x	1297 w-	$\delta(C_\beta - C_\gamma), \delta(C-N)$
29	1325 w	1330 w	1330 m+	x	1324 m+	$\delta(C_\beta - C_\gamma), \delta(C-N)$
30	1355 w	1355 w	1356 w+	1354 w	1357 w+	$\delta(C_\beta - C_\gamma), \delta(C-N)$
31	1423 w	1434 m	1428 sh-	1416 sh	1423 w-	**Amid III**, $\delta(CH_2)\sigma$
32	1450 s	1453 sh	1454 m-	1452 s	-	$\nu(C_{\beta,\gamma}H_2)\sigma$ op
33	1482 w	1479 sh	1478 m+	1485 sh	1475 m+	$\delta(C_\delta H_2)$
34	1643 s	1633 s	1623 m+	1642 s	1637 m+	$\nu(C=O)$, **Amid I**
35	-	-	x	x	x	$\nu_s(CH_2)$ ip
36	2879 s	2892 s	2890	x	x	$\nu_s(CH_2)$ op
37	2933 s	2943 s	2962	x	x	$\nu(CH)$
38	2956 s	2987 s	2985	x	x	$\nu_{as}(CH_2)$
39	2983 s	-	3007	x	x	$\nu_{as}(CH_2)$

s" – strong band, "m" – medium, "w" – weak, "vw" – weak, "sh" – shoulder.. "x" indicates that the spectrum was not evaluated in this region. "-" indicates that the band was not found. Denoting of vibrations is described in section 3.2.4.

5.4.3 Comparison of PPI spectra of different origin

In this subsection, we compare Raman and ROA spectra of PPI (and PPI-like) conformation obtained and discussed individually in previous sections. To be able to easily address observed changes, bands in all spectra were labelled (with correspondence to labelling made in section 5.2). In total we were able to identify over 40 individual spectral features (bands, shoulders, couplets, etc.) in region 100-300 cm^{-1} as shown in Fig. 5.20 on the Raman and ROA spectra of the sample PPI in water. Band positions (in water, 1-butanol, and solid phase) as well as their assignment to specific vibrational motions (according to results in section 5.8.1) are shown in Table 5.2. Positions of Raman and ROA bands in 1-propanol are not shown as they are almost identical to these in 1-butanol.

In Fig. 5.21 we compare spectra measured in 1-propanol and 1-butanol, in water, and in the solid phase. The bands in the Raman spectra correspond approximately to their position independently of solvent environment. The only exception is the position of Amide I band which is substantially shifted to lower wavenumbers (10 cm^{-1}) in water due to hydrogen bonding. However, it is possible to observe significant variations in relative intensities. In the Table 5.3, there is

a comparison of 6 and 6 selected Raman and ROA PPI bands relative intensities in all used solvents.

Raman and ROA spectra were normalized so that the Amide I band (34) has a unit height. In some cases (mostly Raman bands in 1-propanol) these intensities are deviated by problematic solvent subtraction (in table marked with asterisk), however majority of them can be directly compared. Corresponding intensities of Raman bands 5 and 26 and ROA bands 1, 6, and 27 are similar in all spectra in all used solvents. Raman band 32 is the weakest in solid phase spectrum, probably due to different character of $\delta(CH_2)\sigma$ vibrations in solid phase.

The ROA couplet 2 in the spectrum from aqueous solution has a lower intensity of both negative and positive band compared to spectra obtained in organic solvents. The ROA band 4a is visible only in the spectrum in water. Finally, ROA band 19 and couplet 22 have again lower intensities in spectrum in water. This differences may be related to some disruption of regularity in the PPI structure are discussed in more detail in section 5.6.
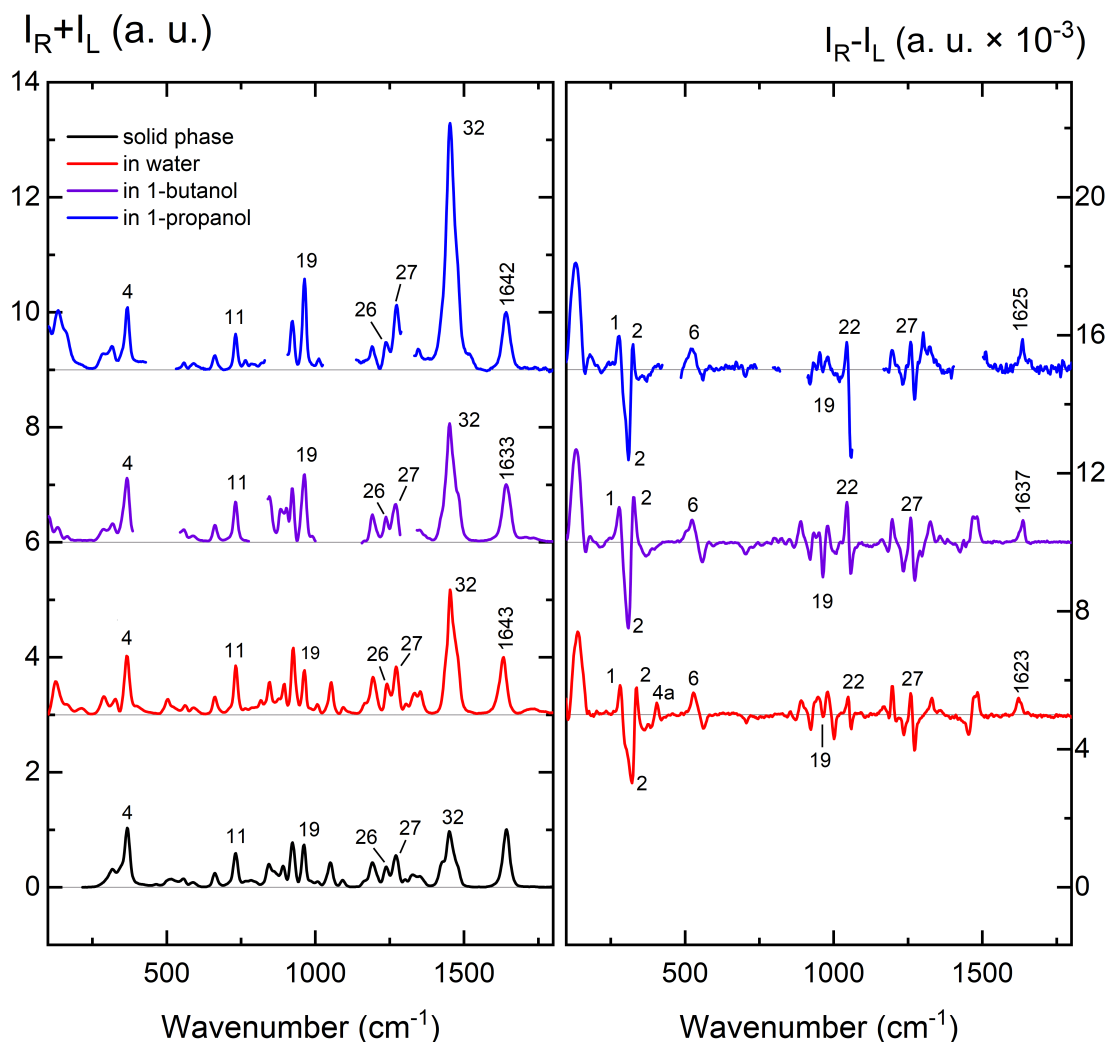
Figure 5.21: Comparison of Raman and ROA spectra obtained from measurements in solid phase (sample *medium*), water, 1-propanol, and 1-butanol.

Table 5.3: Relative intensities of several Raman and ROA bands in PPI spectra (spectra normalized on the same intensity of Amide I band in Raman spectra).

	Raman band No.					
PPI in	5	11	19	26	27	32
Solid phase	1.0	0.6	0.7	0.4	0.5	1.0
Water	1.0	0.9	0.8	0.5	0.8	2.2
1-butanol	1.1	0.7	1.2	0.4	0.7	2.1
1-propanol	1.1	0.6	1.6*	0.5	1.12*	4.3*

	ROA band No.						
PPI in	1	2	4a	6	19	22	27
Water	0.9	-2 / 0.8	0.3	0.6	-0.1	0.5/-0.4	0.6/-1.0
1-butanol	1.0	-2.49/1.3	x	0.7	-1.0	1.2/-0.8	0.7/-1.0
1-propanol	1.0	-2.6/0.7	x	0.6	-0.1*	0.8/-2.5*	0.8/-0.9

5.5 Raman and ROA study of mutarotation

In this section, we present the results of two Raman and ROA measurements at higher temperatures in order to observe the process of mutarotation.

In the former one, sample PPc was measured at 5 °C for approximately two hours and then the temperature was increased to 25 °C and data were collected for another 100 hours. Obtained spectral series after solvent subtraction and baseline correction is shown in Fig. 5.22. We see similar change of spectral bands as observed during measurement of sample PPa at 5 °C (compare Fig. 5.15) but changes are more pronounced in this measurement. Namely, we see: band 1 practically vanishes, marker bands for PPI (4, 19) descend dramatically, intensities of bands 9, 15, and 17 also visibly decrease. Characteristic bands for PPII conformation (4a, 5) get stronger, band 26 also grows. There is characteristic switch of relative intensity between bands 16 and 17. When compared to solid phase spectra (Fig. 5.3), final spectrum of this spectral series seems to still have (at least partially) some PPI component. That is in a good agreement with ECD measurement of mutarotation, where the mutarotation process had still not been completed on the third day of measurement (and the measurement took place at a higher temperature – 35 °C). SVD revealed that spectral series has a factor dimension 3, however, it was not possible to perform fit to reaction kinetics due to jumps in courses of coefficients V caused by experimental conditions during measurement (discussed in section 4.5).

The second measurement was performed already in 2011 on the *short* sample. Measurement lasted 66 hours which is less than measurement of the first sample PPc. However, when we look at Fig. 5.24, we see that band 19 characteristic for PPI vanished completely. This is worth a more detailed discussion. First, let's recall the important differences between the two experiments. Sample PPc is *medium* sample and was measured at 25 °C, second sample is *short* and measurement was performed without temperature stabilization (laboratory temperature approximately 25 °C) with relatively high laser power (1 W). So two possible explanations for this difference are offered. The first is that the rate
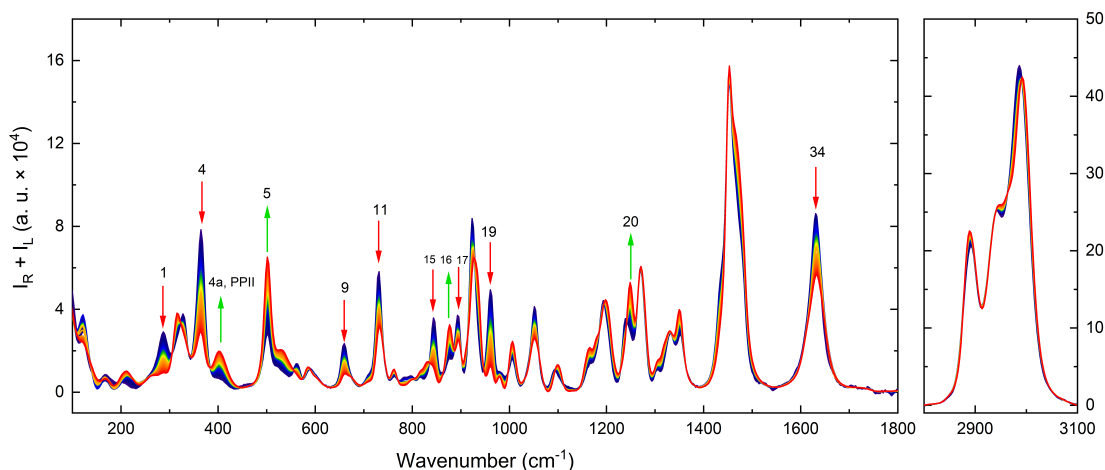
Figure 5.22: Sample PPc, kinetics at 25 °C, 100 hours of measurement. Change of bands highlighted by red and green arrows

of mutarotation may be chain length dependent. Alternatively, the combination of a relatively high laser power and a sample without thermal stabilization may result in a significantly higher effective temperature at sample (our estimate is ≈30 °C), which would cause faster mutarotation kinetics.

It this case it was possible to perform a factor analysis and fit the kinetics, which allowed quantitative comparisons of reaction rates with those from ECD monitored mutarotation. Reaction rates obtained from ECD measurements were $k_1 = 4, 8 \cdot 10^{-4}\,\mathrm{s}^{-1}$ and $k_2 = 2, 3 \cdot 10^{-5}\,\mathrm{s}^{-1}$ with corresponding half-lives $\tau_1 = 0, 4$, $\tau_2 = 8, 4$ hour, reaction rates obtained from this mutarotation measurement are $k_1 = 2, 8 \cdot 10^{-4}\,\mathrm{s}^{-1}$ and $k_2 = 2, 1 \cdot 10^{-5}\,\mathrm{s}^{-1}$ with corresponding half-lives $\tau_1 = 0, 7$ and $\tau_2 = 9.2$ hour. This means that the kinetics in the measurement with the *short* sample were slightly slower than in the ECD measurement with the *medium* sample. The results cannot be completely directly compared due to the undefined measurement temperature.

Spectra of pure components obtained by reconstruction after SVD and fitting to chemical kinetics are shown in Fig. 5.24. PPII conformation obtained from factor analysis fits to PPII Raman spectra best of all our obtained spectra during mutarotation measurements. If we look at the relative concentrations plot, we see that according to the calculated concentrations, there is almost only PPII present at the end of the measurement.

Fig. 5.25 shows subspectra and their courses of coefficients. In contrast to previous measurements, the shape of the fourth subspectrum does not look like noise and course of its coefficients is not random. Despite this, the fit for kinetics with two intermediate products (and substrate and product) did not provide better correspondence to experiment (similarly to case described in section 5.4.2). However, it is possible that with a better set-up of the experiment, Raman spectroscopy could detect a larger number of intermediate products/metastates.
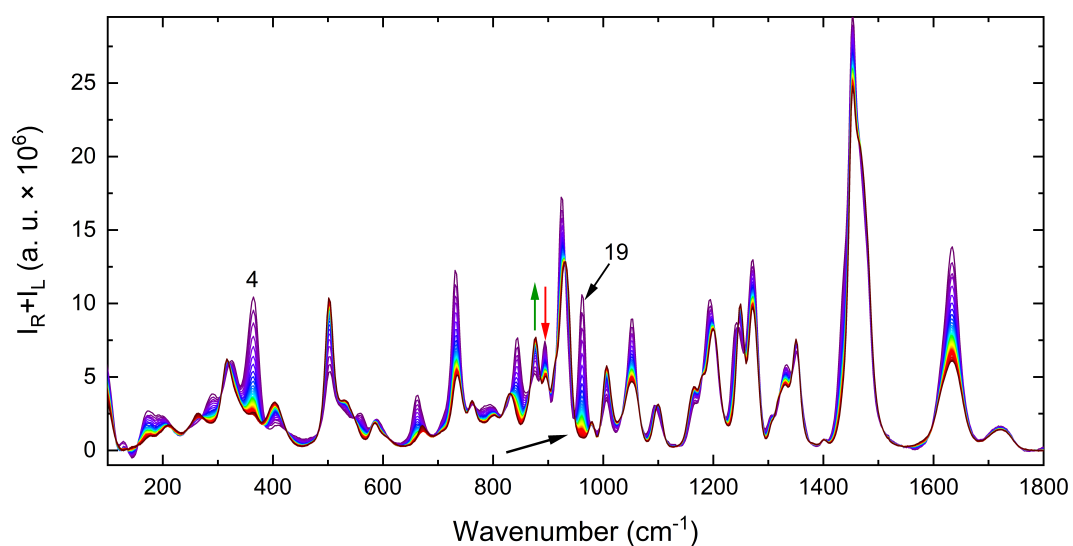
Figure 5.23: Raman spectral series from mutarotation of sample *short*. Band 19 shows that the process of mutarotation reached its end.
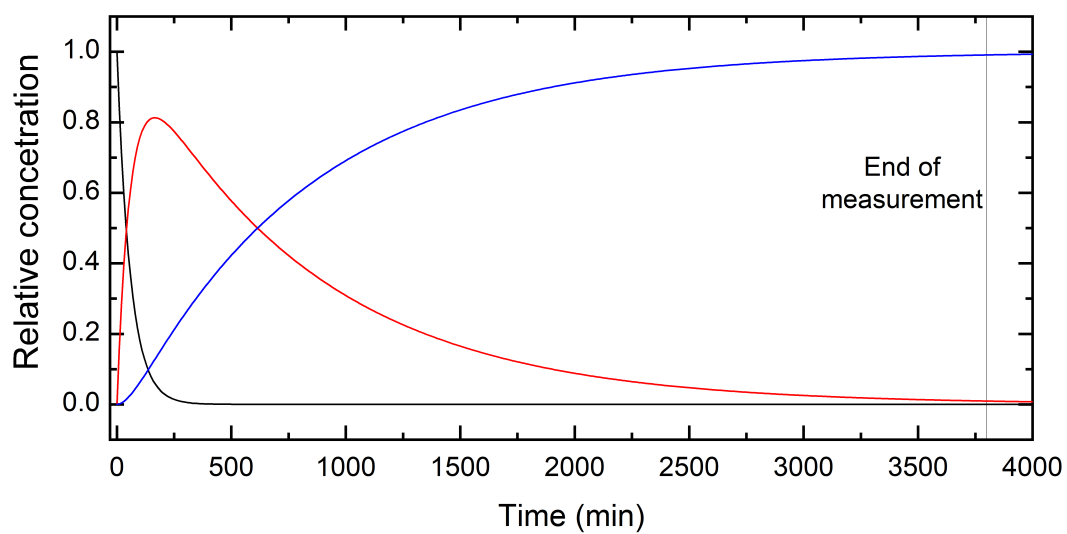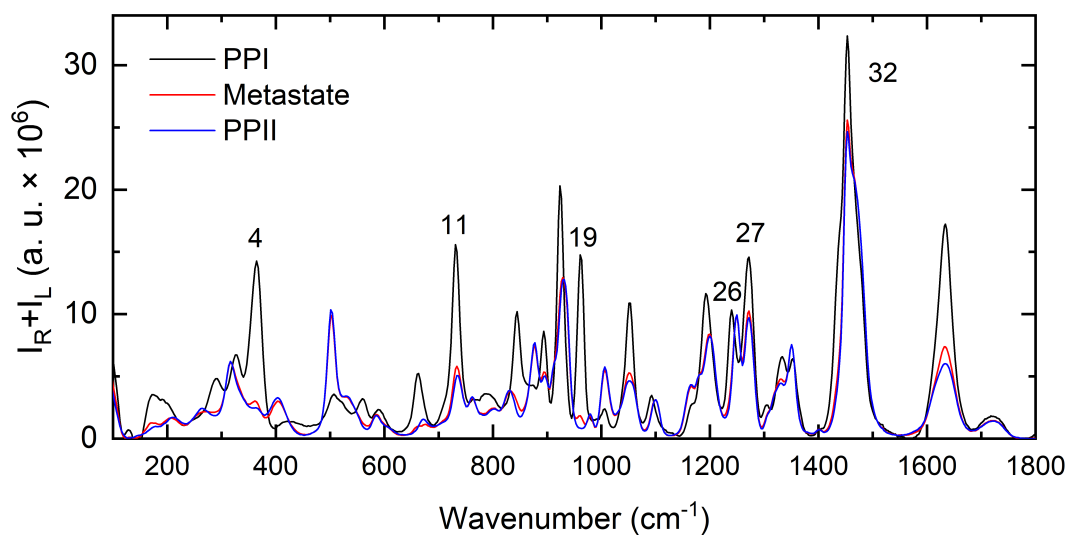
Figure 5.24: Raman spectra and their concentrations reconstructed after performing factor analysis and fit to chemical kinetics, sample *short*.
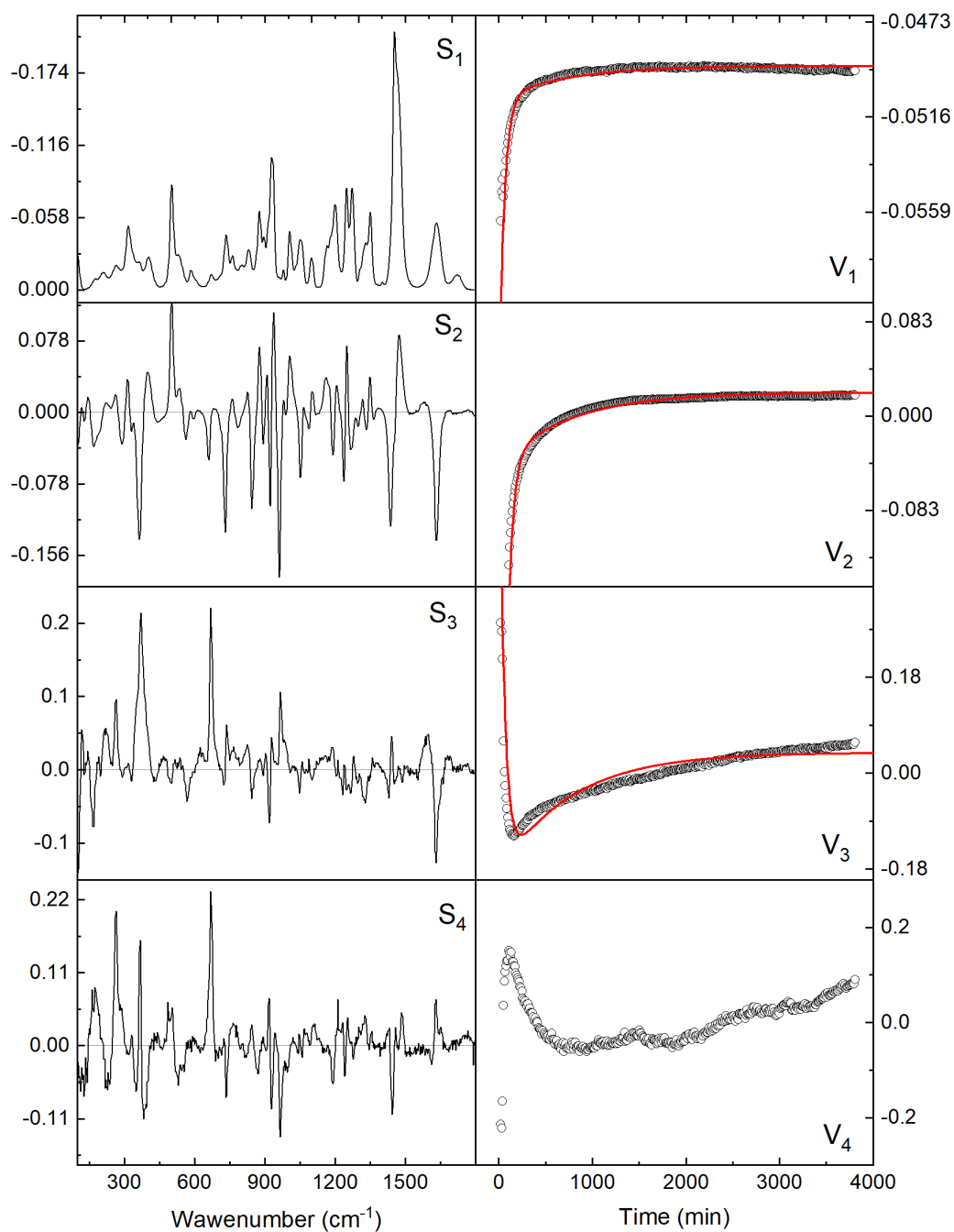
Figure 5.25: Subspectra and their coefficients for measurement of sample *short* from 2011.

5.6 Comparison of ROA spectra of PPI and PPII

ROA is a phenomenon three to four orders of magnitude weaker than its parenting Raman scattering. Therefore, the required signal acquisition times for recording of high-quality spectrum are several times longer. In our case, the time required to obtain a decent ROA spectrum (i. e. with recognizable main spectral features) was about one hour, however for well resolved ROA spectrum enabling analysis of weak spectral features and their changes it is necessary to collect data for dozens of hours (depending on the studied system).

The ROA spectra of PPI we obtained from measurements both in aqueous solution and organic solvents were already presented in section 5.4.3. We also tried to process the ROA signal from mutarotation measurements. All ROA spectral series monitoring mutarotation (PPa – 20 °C, $\approx$60 hours; PPc 25 °C, $\approx$ 70 hours and measurement of *short* sample, $\approx$64 hours) have to be averaged over larger accumulation period to lower noise (while Raman spectra had the time step about 5 minutes). However, the noise level was still so high that SVD has determined the dimension of each analysed dataset as two. In other words, due to the elevated noise level ROA sees the PPI to PPII mutarotation as two-step process only, i. e. gradual change from PPI to PPII spectral shape.

To be able to distinguish more subtle spectral changes, the kinetics would have to be really slow to accumulate enough ROA spectra with low noise. However, complete mutarotation at temperature close 0 °C would take very long time (approximately two weeks by guess)

Even so, we can make a very interesting comparison between the ROA spectra of PPI in 1-butanol (pure PPI conformation) and in water (incipient conformation) and the spectrum of PPII in water as shown in Fig. 5.26. Spectrum of PPII was adopted from publication[41]. Overall character of PPI and PPII spectra is rather different. Only few spectral bands remain unaffected by change of polyproline chain secondary structure arrangement – bands 6, 26, 27, and 32 (for their assignment see Tab. 5.2). Rest of the spectra can be divided into four regions:

(i) low-wavenumber region (200-450 cm^{-1}): – skeletal modes

(ii) middle region (800-100 cm^{-1}): $\nu(CH_2)$, $\delta(CH_2)$, $\delta(CCCC)$ breathing modes.

(iii) methylene scissoring, wagging, and twisting region (1100-1450 cm^{-1})

(iv) amide IV

(i) Both PPI and PPII have rather distinguishable spectral pattern in the low-wavenumber region. In PPII, there is a broad negative band at 325 cm^{-1}, and strong positive band at 405 cm^{-1} accompanied by weak negative band at 421 cm^{-1}.

In PPII, we detect characteristic +-+ V-shape pattern at 279, 308, and 327 cm^{-1}. Soon after initialization of mutarotation (as shown in spectrum of PPI in water) we observe slight upshift of negative 308 cm^{-1} band and lowering of its intensity, and – more specifically – the appearance of 4a PPII band.

(ii) Evidently the most specific band corresponding to the long-range regularity is 19 corresponding to $\nu(C_\alpha$-N$)$ and $\nu(C_\gamma$-C$_\delta)$ vibrations. We observe its rapid decrease (in term of absolute value) immediately after the start of the mutarotation. Another marker of PPI helix regardlessly appears to be couplet 22

69

corresponding to the $\nu(C_\alpha$-$C_\beta)$ in protein ring which looses the intensity and its positive part even switches to negative band in PPII. Increase of PPII content is accompanied by increase of intensities of bands 18, 20, and 21, while intensity of band 18a refers to the long-range regularity of PPII helix as shown in [41].

(iii) In this region isomerization from PPI to PPII is accompanied by increase of intensity of bands 24, 25, 29, and 30 mostly corresponding to $\delta(CH_2)\omega$ and $\delta(CH_2)\tau$ vibrations coupled over different proline rings.

(iv) To better compare changes in Amide I region caused by PPI to PPII transition we shifted down the position of Amide I band in 1-butanol by $10.5\,\mathrm{cm^{-1}}$ (according to the difference between Amide I bands position in Raman spectra). The result is shown in Fig. 5.27. Despite the limited quality of ROA spectra in this region (Amide I signal has quite low intensity), the differences among individual signals are sufficiently convincing to allow a meaningful analysis. Apparently, the ROA Amide I signal is not a single positive band, but strongly imbalanced couplet both for PPI and PPII. For left-handed PPII helix we observe -/+ couplet at $1603/1621\,\mathrm{cm^{-1}}$ with strongly prevailing positive components, while for right-handed PPI again there is a +/- couplet at $1626/1649\,\mathrm{cm^{-1}}$ and again with strongly prevailing positive part.

ROA signal of PPI in the water then looks like some average of the signals of regular PPI and PPII helices. This means that the ROA signal, unlike VCD (see Fig. 5.28), is sensitive to the sense of rotation of the polyproline chain.
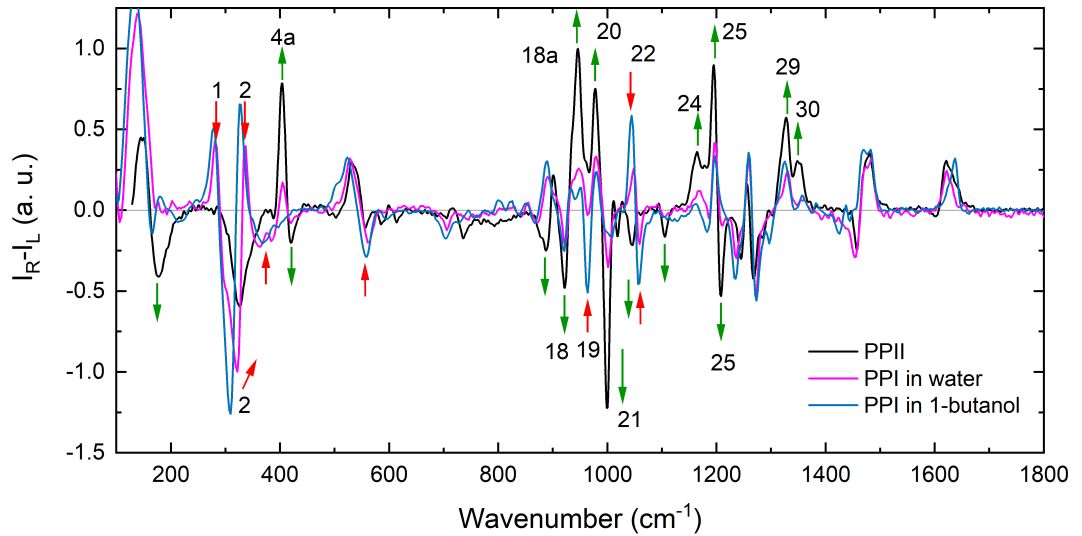
Figure 5.26: Comparison of ROA spectra of PPI and PPII. Arrows highlight changes of bands from PPI to PPII. PPI spectra from $5\,^\circ$C kinetics and PPII from measurement in aqueous solution in 2011.
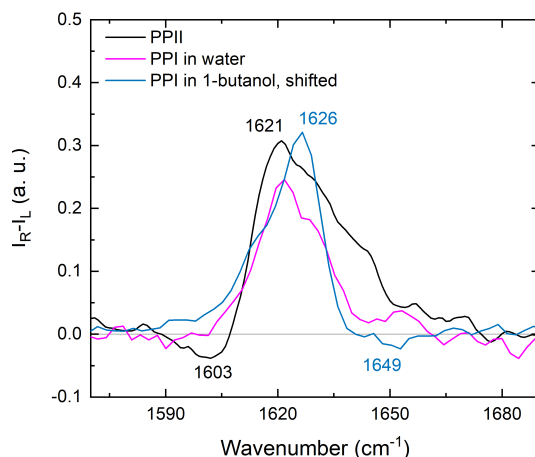
Figure 5.27: Comparison of amide I band in ROA spectra of PPI and PPII. $10.5\,\text{cm}^{-1}$ is subtracted from the scale of spectrum of sample in 1-butanol to overlay amide bands.
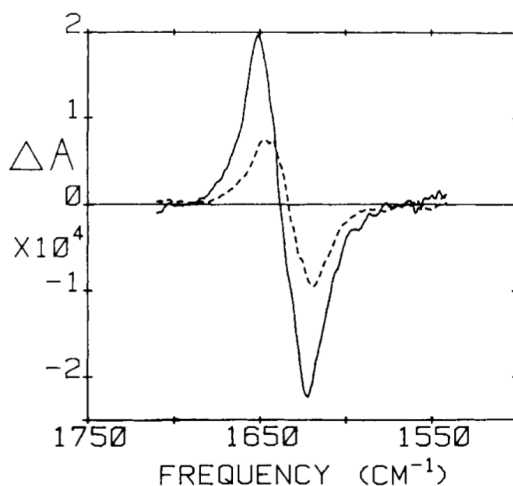
Figure 5.28: VCD spectra of PPI (- - -) and PPII (——) in 2,2,2-trifluoroethanol. Spectra are normalized to same absorbance in the Amid I band. Adopted from [18].

5.7 Structural parameters of PPI and PPII helices

To be able to interpret experimental Raman and ROA spectra in more detail, we decided to perform the corresponding spectral simulations. However, direct calculations of properties of a polyproline chain with a number of proline residua higher than six ($\approx$100 atoms) are rather costly and even impossible at the more precise level of theory. Therefore, we decided to take advantage of indirect approach based on transfer of molecular property tensors (see section 3.2.3) of smaller fragments. Those fragments are treated individually and to prevent larger deviations in geometry, their structure were optimized in normal modes with some restrictions (see section 5.7.1).

The structure of polyproline is rather simple. It can be fully described by values of three dihedral angles (ϕ, ψ, ω) specifying the conformation of peptide backbone and two pseudorotation angles (θ_m, P) characterising puckering (see

71

section 1.3). The approximate values of these angles for PPI and PPII conformations are known, however, little is known about variations in local arrangement, i. e. how specific puckering pattern affects backbone geometry, which is essential knowledge for the reasonable simulations of the longer chains.

These effects were studied (to some extent), however, only for PPII helix[41]. Therefore the first aim of our calculation was to characterize the local geometry parameters of PPI helix in dependence on puckering sequence. Fort that we studied proline heptapeptides with N-acetyl on one end and methylamide on the other.

5.7.1 Interplay of puckering and backbone geometry in proline heptamers

Heptapeptides with all combinations of puckering were used for calculating dihedral angles at B3PW91/6-31G*/cpcm(1-propanol) level of theory. This combination of DFT functional and basis set was chosen as reasonably accurate and fast.

We expected some puckering dependence so analysis (using program *Pro-Stat.py* was done for all prolines in peptide in general, with separation for puckering of single amino acid (A/B) and for sequences of puckering of three amino acids (e.g. AAB, BAB etc.) We also expected some differences at the ends of peptide so we compared results for sets where the peripheral peptides were excluded, i. e. from the central pentapeptide only.

Calculations ran successfully for 127 of 128 heptapeptides. We were unable to find energy minimum to the last one (BAABBBB). We tried to change some settings (e. g. convergence criterion), but no change which conserve comparability to calculation of other peptides helped.

Statistics for each position in heptapetide and both conformations is in Table 5.4. Number of used angles for calculation of mean value is 63 or 64 (based on which is missing because of the one not calculated heptapeptide). We can see that there are some differences for both puckerings and different positions. The largest difference between puckerings A and B is int mean value of angle ϕ (around 10°), differences for other angles are smaller and comparable to the statistical errors. First ω angle for both puckering strongly varies from the other. There is also possible to observe some differences in all angles for one or two outer prolines, which are in some cases higher than statistical error. We expect that prolines in the middle part of $(Pro)_7$ should have conformation more similar to those in longer peptide.

Results of calculation mean values for each puckering with dependency on puckering of neighbouring prolines are shown in Table 5.5. The outer one prolines were excluded from statistics. 47 or 48 angles were used for calculation of each mean value. We can see slightly better variances than before. But obvious dependence of angles ψ and ϕ is more important (highlighted in table). ϕ shows dependence on the puckering of the previous proline and ψ on the puckering of the next one. We use this results for constructing proteins for calculation of spectra.

Energy preferences of puckering has also been studied. We summed Boltzmann's coefficients calculated from energies of optimized geometries for A and B puckerings. The ratio A:B was approximately 8:2.

Table 5.4: Mean values for backbone dihedral angles at each position in $(Pro)_7$ in PPI helix for both puckerings.

A

	ω	ϕ	ψ	θ_m	P
1	-3.9 ± 0.6	-80.0 ± 0.6	164.6 ± 1.9	38.2 ± 0.1	121.7 ± 0.3
2	-7.0 ± 3.1	-79.0 ± 2.3	164.8 ± 1.4	38.4 ± 0.2	122.2 ± 2.7
3	-6.8 ± 1.6	-78.0 ± 3.0	166.6 ± 1.4	38.4 ± 0.1	120.7 ± 1.5
4	-9.9 ± 1.0	-75.4 ± 1.8	167.6 ± 1.5	38.4 ± 0.1	119.9 ± 1.1
5	-10.3 ± 1.2	-74.9 ± 2.3	167.5 ± 1.4	37.9 ± 0.1	119.5 ± 1.1
6	-10.9 ± 0.9	-74.2 ± 1.2	167.7 ± 1.3	38.0 ± 0.1	120.4 ± 0.7
7	-9.8 ± 1.1	-71.6 ± 1.0	165.6 ± 0.4	38.0 ± 0.1	116.2 ± 0.4
∅	-8.4 ± 2.8	-76.1 ± 3.3	166.3 ± 1.9	38.2± 0.2	120.1 ±2 .3

B

	ω	ϕ	ψ	θ_m	P
1	-1.7 ± 0.6	-60.0 ± 3.7	159.2 ± 1.1	38.8 ± 0.3	-84.5 ± 4.4
2	-9.4 ± 2.7	-66.9 ± 1.9	160.7 ± 2.2	38.8 ± 0.1	-94.6 ± 1.9
3	-11.4 ± 1.2	-66.8 ± 2.7	163.8 ± 1.3	38.9 ± 0.1	-96.7 ± 2.7
4	-13.5 ± 1.2	-64.4 ± 2.0	163.8 ± 1.6	38.8 ± 0.1	-95.1 ± 1.9
5	-14.4 ± 1.6	-65.9 ± 2.4	165.5 ± 1.8	38.4 ± 0.1	100.8 ± 2.9
6	-14.5 ± 0.9	-64.8 ± 1.1	165.7 ± 1.5	38.4 ± 0.1	100.3 ± 1.5
7	-12.1 ± 1.0	-61.4 ± 1.1	161.0 ± 0.6	38.3 ± 0.1	-93.8 ± 1.5
∅	-11.0 ± 4.4	-64.3 ± 3.4	162.8 ± 2.8	8.6 ± 0.3	-95.1 ± 5.6

It would be nice to find if this holds for longer oligopeptides and if consideration of longer puckering sequences will bring some benefit. Unfortunately, such calculations could be too expensive in meaning of computational time.

Heptamers in the PPII conformation were studied in the same way as these in PPI. 128 calculations for all combinations of puckering were started. Five calculations (ABAAABB, ABAABAA, ABBBBAA, BABABBA, and BBBBBBB) terminated with errors. Failure was caused by the same problem as before – convergence.

There is again dependency for A and B puckering. Average angles for all triplets are shown in Table 5.6. Compared to the PPI dependency on puckering seems to be weaker, also the variances are smaller. The only exception from this is dependency for third puckering in triplet with B in the middle which is in fact larger than in PPI ($\approx$11°, highlighted in Table 5.6). The results of PPII optimization are not directly used in this work. However, they could be used in the future to design possible mutarotation intermediates and predict their spectra.

The dependence of puckering pseudorotational angles on local puckering pattern seems to be relatively small. Differences of θ_m for both A and B puckering are within standard deviation. This holds also for angles P with the exception for puckering B in PPII. (Triplets ending with puckering A have different mean values than those which end with B and these differences are higher than standard deviations.) Therefore, the dependence of pseudorotational angles on sequence

of puckering can be neglected. It should be, however, noted that mean values of pseudorotational angles corresponding to A and B puckering differ for PPI and PPII helices.

5.7.2 Preparation of long PPI chains and their fragmentation

Long polyproline chains of arbitrary length and puckering composition in PPI conformation were prepared based on the values in Tab. 5.5 using program *ProGen.py* (section 4.8.1). Each final structure was then fragmented to either trimers or tetramers with N-acetyl group on one end and methylamide on another. Proline helices have approximately 3 residues per turn, therefore it can be expected that trimer fragments might be short to correctly describe long-range structural properties, while tetramers should be sufficient. Usage of even longer fragments would certainly be more accurate, but also more costly.

In our case, we could take advantage of the fact that the calculated fragments occured repeatedly. So we could calculate them only once and thus avoid repeated calculations. This allowed us to calculate the spectra of a number of different variants of the PPI molecules, because otherwise calculations in this range would be computationally very expensive. The geometry of trimers and tetramers was designed on the basis of previous calculations - the choice of dihedral angles was given by a combination of puckering. Larger changes in geometry were prevented by fixing low-frequency modes while optimizing in normal coordinates (cm^{-1}).

In total, there are only 8 different trimers and 16 different tetramers which may occur in any PPI chain. (For hexamers there would be 64 models.) Fragments were optimized at higher level of theory B3PW91/6-31++G**/cpcm(1-propanol) which was then used also for calculations of frequencies and intensities.

Table 5.5: Mean values for dihedral and pseudorotational angles for each triplets of puckering in PPI, colours are used for highlighting dependence on neighbouring puckering.

	ω	ϕ	ψ	θ_m	P
AAA	-9.1 ± 1.2	-74.1 ± 1.1	166.5 ± 0.6	38.2 ± 0.2	119.1 ± 1.1
AAB	-10.0 ± 1.7	-73.8 ± 1.7	168.9 ± 0.6	38.2 ± 0.2	119.5 ± 1.3
BAA	-8.1 ± 2.4	-78.0 ± 1.9	165.5 ± 0.8	38.2 ± 0.3	120.3 ± 0.7
BAB	-9.1 ± 2.1	-78.3 ± 2.2	168.0 ± 0.8	38.2 ± 0.3	121.2 ± 1.2
ABA	-12.3 ± 1.5	-63.6 ± 1.1	163.5 ± 0.9	38.6 ± 0.2	-95.2 ± 2.9
ABB	-14.1 ± 1.9	-63.6 ± 1.2	166.1 ± 1.3	38.7 ± 0.3	-97.0 ± 3.9
BBA	-12.2 ± 1.6	-67.7 ± 1.7	162.7 ± 1.1	38.7 ± 0.1	-98.2 ± 2.7
BBB	-13.7 ± 1.5	-68.0 ± 1.6	165.3 ± 1.0	38.8 ± 0.2	-99.7 ± 2.8

Table 5.6: Mean values for dihedral and pseudorotational angles for each triplets of puckering in PPII, colours are used for highlighting dependence on neighbouring puckering

	ω	ϕ	ψ	θ_m	P
AAA	176.7 ± 0.6	-72.0 ± 0.6	154.8 ± 0.9	37.5 ± 0.1	109.4 ± 0.7
AAB	176.9 ± 0.7	-72.7 ± 1.0	155.4 ± 3.5	37.3 ± 0.2	109.4 ± 1.9
BAA	177.7 ± 0.7	-72.6 ± 0.9	154.4 ± 0.8	37.5 ± 0.1	108.4 ± 0.6
BAB	177.8 ± 1.0	-73.5 ± 1.4	154.4 ± 6.0	37.3 ± 0.3	108.5 ± 2.1
ABA	172.0 ± 1.0	-60.4 ± 0.5	147.1 ± 1.7	37.9 ± 0.1	-85.9 ± 2.6
ABB	173.3 ± 0.7	-58.6 ± 0.7	135.9 ± 1.8	38.2 ± 0.1	-83.3 ± 1.8
BBA	174.1 ± 0.8	-61.7 ± 0.8	145.7 ± 1.3	38.0 ± 0.1	-77.1 ± 0.8
BBB	174.7 ± 0.6	-61.2 ± 1.0	133.4 ± 1.2	38.3 ± 0.1	$-75.\ 2\pm 0.7$

5.8 Calculation Raman and ROA spectra of PPI

5.8.1 Trimers and tetramers

The main goal of the trimer and tetramer calculations was to obtain a force field and other tensors for further calculations. However, there are two to reasons show their calculated spectra. First, spectrum of tetramer AAAA was used for assignment of experimental spectral bands to normal coordinates (vibrational movements). Second, their spectra allow us to compare the benefits of using the transfer of molecular property tensors.

Fig. 5.29 shows spectra of all individual trimers and their average spectrum. Raman spectra show only slight changes in intensities of bands and their positions (e. g. Amid IV). Differences in ROA spectra are much more significant. Characteristic regions of ROA spectra (at approximately 300, 950, $1300\,\mathrm{cm}^{-1}$) vary in intensities and shapes of spectra. Characteristic V-pattern around $300\,\mathrm{cm}^{-1}$ is missing, there is only a negative band in its place.

The number of different tetramers prevents reasonable comparison of individual spectra. Moreover, comparing the spectra of individual trimers and tetramers is not very interesting from the point of view of this thesis. Therefore, we further present only undifferentiated spectra of all trimers (Fig. 5.30) and tetramers (Fig. 5.31) together with their average spectra. Individual spectra of both trimers and tetramers vary from average spectrum, however some conserved bands can be identified.

It should be noted, that the spectra of tetramers and their average spectrum manifest more similarities to experimental PPI spectrum, i. e. V-pattern starts to emerge. The comparison of these spectra is shown in Fig. 5.32. The spectra overlap quite well, this agreement can be caused by partial fixation of the structure when optimizing in normal coordinates. The Raman spectra correspond well to each other and there are some minor differences in intensities of bands rather than significant changes of spectral shape. The same is true for ROA spectra, with the exception of region of V-pattern, bands 19 and 34, and CH region, where average spectrum of tetramers is more similar to the experimental one.

The most of significant bands identified in experimental spectra can be easily

assigned to those in calculated spectra of trimers and tetramers. (Therefore, it was possible to assign experimental spectra to vibrational moves using tetramer AAAA.) However, there are some differences in spectral shapes in some regions. Positive part of ROA couplet 2 is weaker than positive part of couplet 1 (it is missing in trimers spectrum). The spectral shape in the 850-1100 cm⁻ region is less pronounced in both ROA spectra. Some of bands in the 400-600 cm⁻ region are related to vibrations of N-acetyl and methylamide endings and do not have a corresponding counterpart in the experiment.
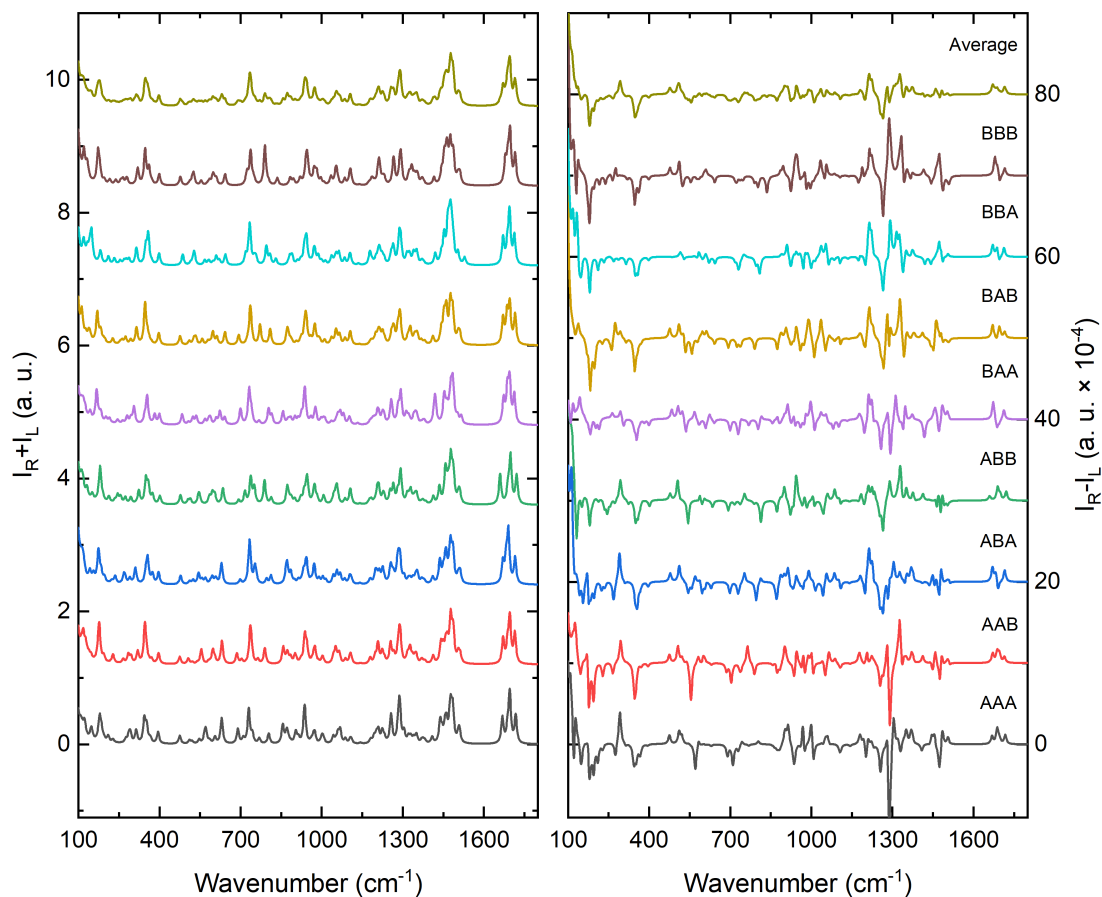
Figure 5.29: Calculated spectra (B3PW91/6-31++G**/cpcn(1-propanol) of proline trimers together with their simple average.
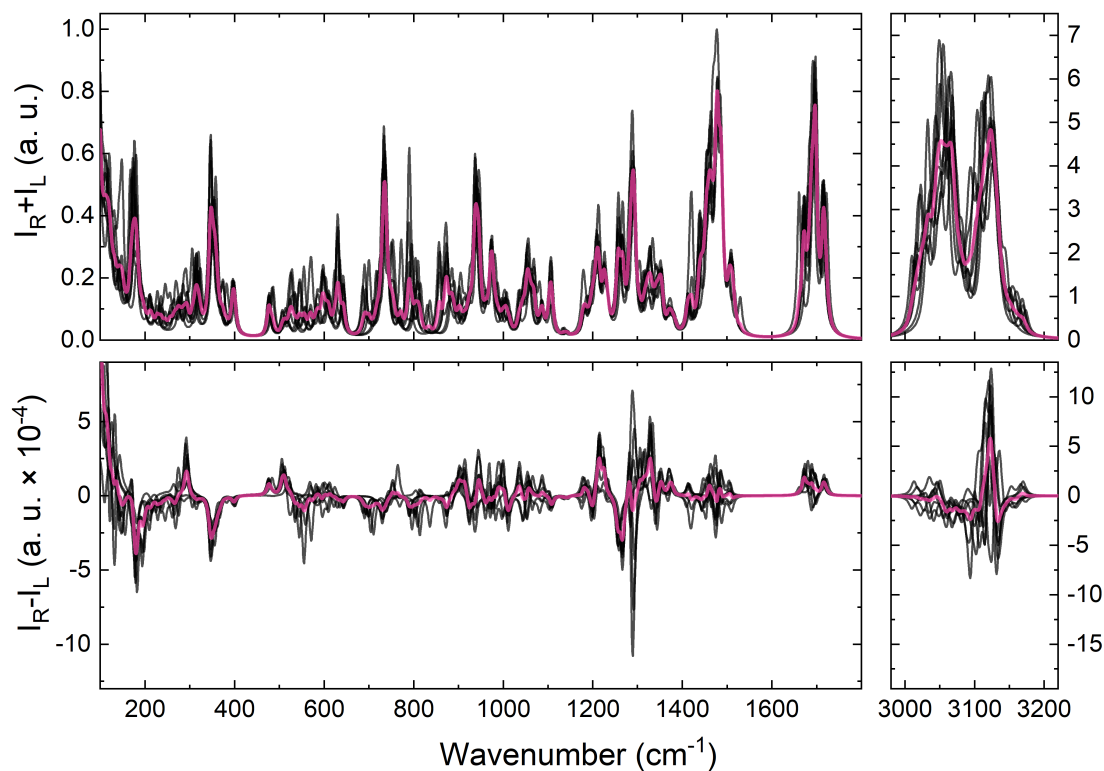
Figure 5.30: Calculated spectra of proline trimers. Spectra of individual trimers have black colour, average spectrum has pink colour.
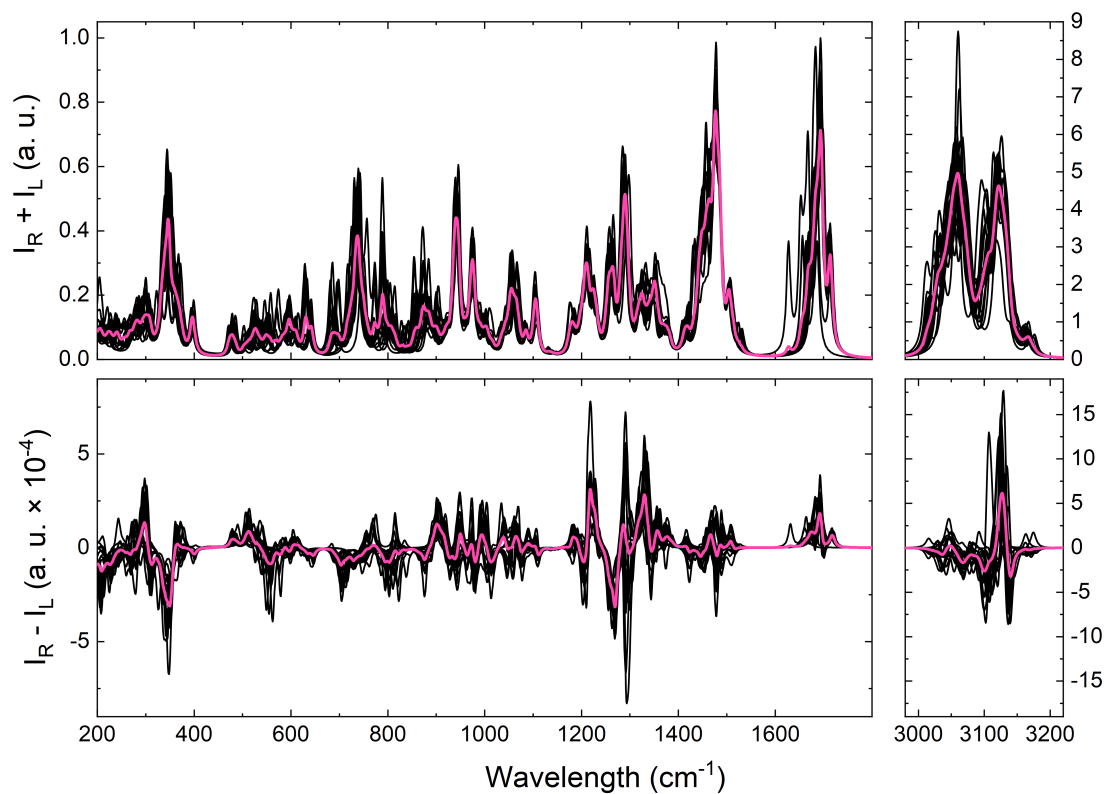
Figure 5.31: Calculated spectra of proline tetramers. Spectra of individual tetramers have black colour, average spectrum has pink colour.
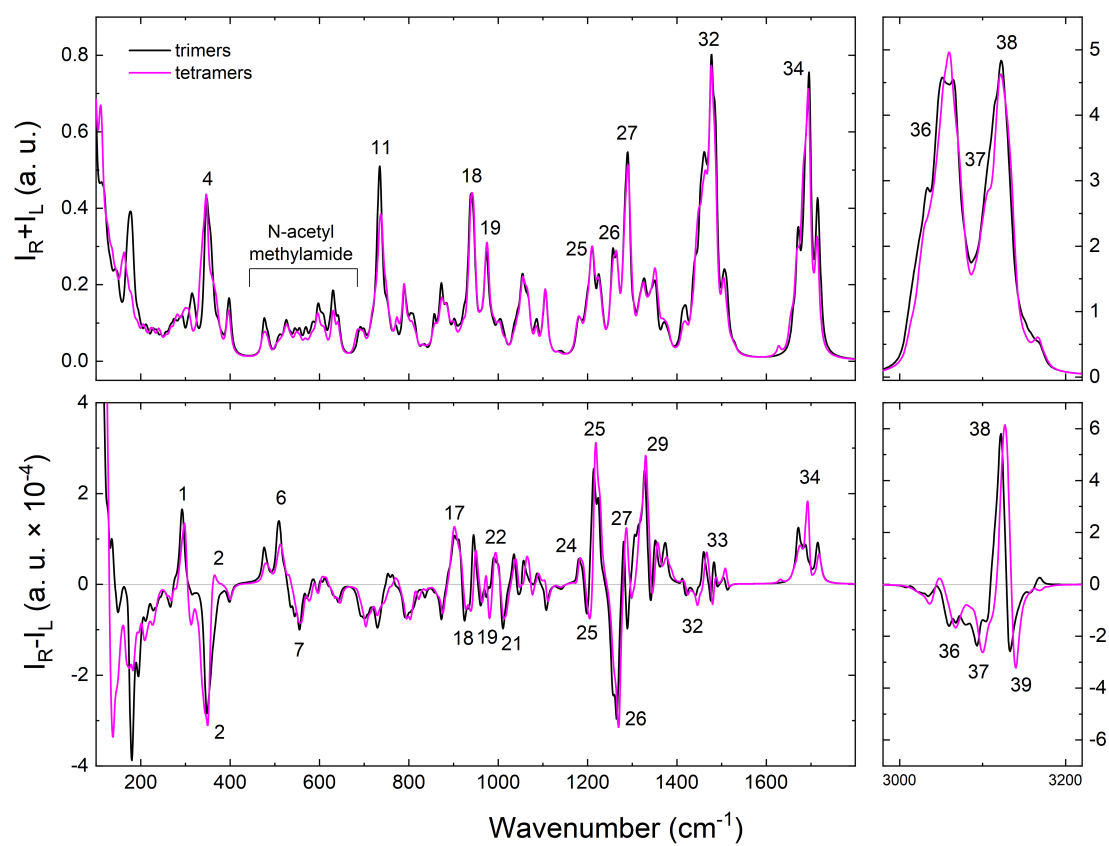
Figure 5.32: Comparison of spectra obtained as simple average from spectra of all trimers and tetramers. The most significant bands are labelled.

5.8.2   Longer PPI molecules

The main goal of calculations of spectra of longer molecules is to obtain the best correspondence between experimental and simulated Raman and ROA spectra. The longest calculated polyproline chain with a length of 100 residues is already close to the mean length of our sample *medium* (length ≈180). However, we also deal with various parameters that can affect the shape of the spectrum, above all the puckering pattern.

The calculations of the most prepared PPI molecules were performed in parallel using fragments of length 3 and 4. In general, usage of the longer fragment should provide a better prediction and agreement with the experiment. (Tetramer fragment contains one full fold of PPI helix). Quite surprisingly, the differences between spectra of long chains generated from trimers and tetramers did not turn out to be dramatic. Unless otherwise stated, the spectra shown further in this subsection are calculated using tetramers. Some additional spectra calculated using trimers can be found in the appendix A.3.
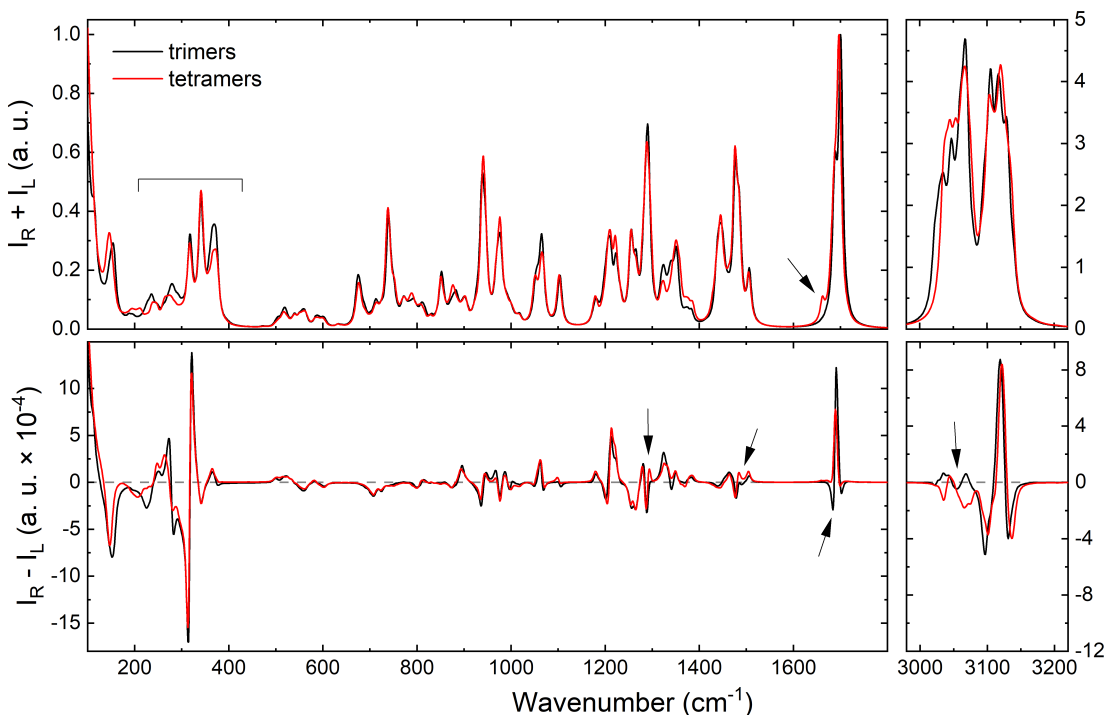
Figure 5.33: Comparison of Raman and ROA spectra of $(Pro)_{100}$ calculated from trimer and tetramer fragments. Some changes are highlighted by line and arrows (more in the text).

A comparison of Raman and ROA spectra calculated using of trimers and tetramers can be found in the Fig. 5.33. Obvious differences in Raman and ROA spectra are in the low frequency region. However, these are mainly changes in the intensity of individual bands. Just above $200 \, \text{cm}^{-1}$, an extra band appears in the Raman spectrum of tetramers and there is a slight shift of the other two bands. At $1300 \, \text{cm}^{-1}$, a positive band is formed in the ROA spectrum of tetramers. At $1400 \, \text{cm}^{-1}$, a band is shifted to positive part of spectra. The last major change in this part of the spectrum is around $1700 \, \text{cm}^{-1}$, where two negative signals in Amide I region disappear almost completely in ROA spectra of tetramers and

one weaker band is shifted in Raman spectrum. In the rest of this part of the spectrum, only some minor changes in intensities occur. In the CH region, the intensities and positions of some peaks are shifted in the Raman spectrum. The most notable difference in the ROA spectrum in the CH region is the reversed orientation of the three bands (marked by black arrow in Fig. 5.33).

Fig. 5.34 shows the length dependence of calculated spectra. Every spectrum was made as an average spectrum of 20 spectra of $(Pro)_N$ molecule with different random puckering patterns. Enlarged sections of ROA spectra are shown in bottom. The spectra for chains lengths between 5 and 100 with step 5 were calculated, however, the changes for lengths higher than $N = 20$) are relatively small, therefore some of the lengths are omitted.

Most of the bands seem to be only weakly dependent on length. The most significant changes in intensities of Raman bands occur in the 400-650 cm^{-1} region, which is dominated by vibrations related to ending groups of our molecules (N-acetyl and methylamid). Therefore, the decrease of their intensity seems to be related to the lesser role of ending groups in spectra of longer chains. Some changes are also visible in CH region of spectra, these changes may be, however, also attributed to lowering proportion of N-acetyl and methylamid groups.

In ROA spectra for $N > 20$, the most of the bands seem to be only weakly dependent on length of chain. This does not hold for V-pattern at approximately 300 cm^{-1} and Amide I band, which are slightly more sensitive to the length of PPI helix (up to chain length $\approx$35) as shown in enlarged section in Fig. 5.34. While these regions are characteristic marker of structure of PPI, this improvement in prediction is important.

Another observation can be made based on the anticipated regularity of puckering pattern. The Fig. 5.35 shows a comparison of the calculated spectra $(Pro)_{100}$ formed only from puckering A (100A), only from puckering B (100B), and then the average spectra of twenty different molecules with a random distribution of puckering in the ratio A:B=3:2 (*random*). The green line then represents the spectrum generated as a weighted sum of the spectra of 100A and 100B in a ratio of 3:2 (*average*).

The discussion of the differences between these spectra is somewhat tricky, as spectra 100A and 100B are generated using a single fragment (AAAA or BBBB), while all 16 possible fragments are involved in the production of random spectra. At first glance, spectra of 100A and 100B differ significantly. Likewise, the random spectrum differs from the average spectrum.

The Raman spectrum can be divided in two main regions. In the 900-1800 cm^{-1} region the random spectrum is remarkably similar to average spectrum, which suggests that frequency and intensity of Raman vibrational modes depends only on the number of A and B puckering in PPI chain and not on their sequence. On the contrary, in the 200-900 cm^{-1} region the sequence of A and B puckering appears to be more important, because the *average* and *random* spectrum differs significantly.
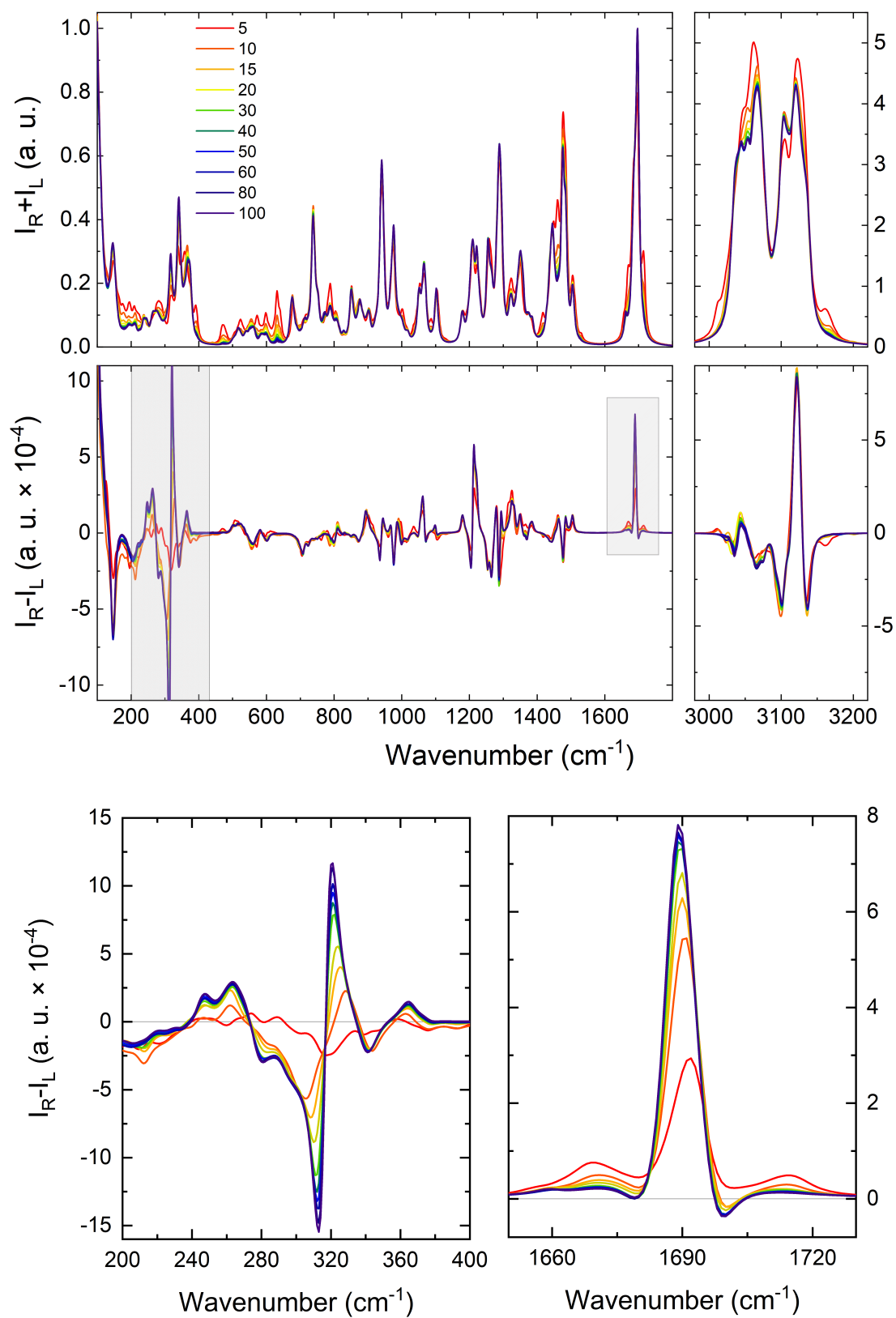
Figure 5.34: Raman and ROA spectral profiles of $(\text{Pro})_N$ in PPI conformation with random puckering pattern.
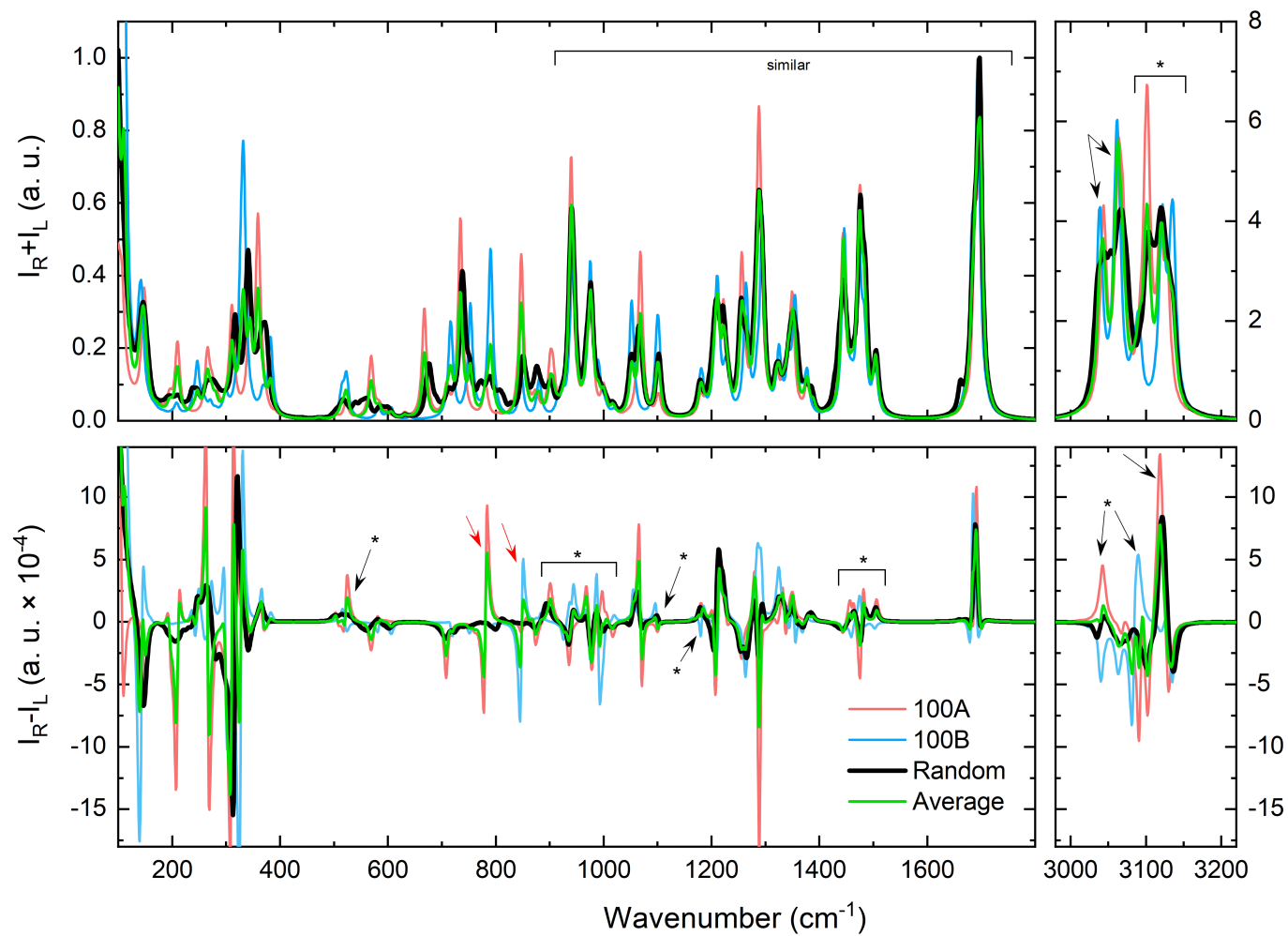
Figure 5.35: Comparison of spectra of (Pro)$_N$ polymers for 100A, 100B, *random* chain, and sum of weighted average of 100A and 100B (3:2). Some differences highlighted (arrows and asterisk)

This is even more pronounced in ROA. In the 900-1800 cm^{-1} region the *random* and *average* spectra well correspond to each other, while spectra of 100A and 100B show there are number of bands with opposite sign contributing to average spectrum (marked by asterisk). Again, in the 200-900 cm^{-1} region *random* and *average* spectrum deviates from each other which indicates that puckering pattern (and not only number of A and B residues) plays important role.

The most important differences are marked by red arrows (couplets at approximately 780 and 850 cm^{-1}), also V-pattern is much better captured in *random* spectrum. Evidentially, couplets at ≈780 and 850 cm^{-1} correspond to long-range coupling of some vibrations of adjacent proline residues with consecutive A and B puckerings respectively.

This finding allows us to estimate the maximal representation of repetitive A and B puckering sequences, because in experimental spectra there are no observable -/+ ROA couplets in the 700-900 cm^{-1} spectral region.

Fig. 5.36 provides further insight into the influence of areas formed only by puckering A (polyA) and B (polyB). In this figure we see the spectra of PPI conformation with A and B puckering composition in the ratio 3:2 in different degree of organization. The notation xAyB shows how many consecutive prolines with the corresponding puckering are in a row. This pattern is then propagated throughout the whole length of peptide. Chosen chain length was set to 80 residues, as it enables propagation of puckering patterns without creating addition residues at the end. As an extreme case of non-organization, an average from 20 random puckerings sequences of 80 residues is also included (*random*).

It should be noted, there are substantial differences in composition of prepared $(Pro)_{80}$ molecules. For 100A and 100B (discussed above), there was used only one type of tetrameric fragment. At the polyA and polyB interface, fragments AAAB, AABB, ABBB, BBBA, BBAA, BAAA start to appear. The shorter the puckering pattern, the more of these tetramer units occur. An exception is the 3A2B molecule, where fragments AAAA, BBBB, ABBB and BBBA do not occur, but a fragment ABBA is present. The rest of tetrameric units (AABA, ABAA, ABAB, BAAB, BABB, BBAB, BABA) were used only for generation of random spectrum.

Raman spectra in the 900-1800 cm^{-1} region are almost identical. Below 900 cm^{-1} there are also several band increasing in intensity with increasing length of polyA and polyB sequences (red arrows). Quite interesting are the bands where 3A2B molecule differs significantly from the trend of remaining spectra (marked as exceptional – exc.) This difference may be caused by a significantly different composition of fragments present in sequence.

However, the most important information is the change of intensity of ROA couplets at approximately 780 and 850 cm^{-1} which can be assigned to polyA and polyB sequences. In enlarged plot cutout, it is evident that the minimum length of the polyA or polyB puckering pattern necessary for creation of recognizable couplet is at least 6.
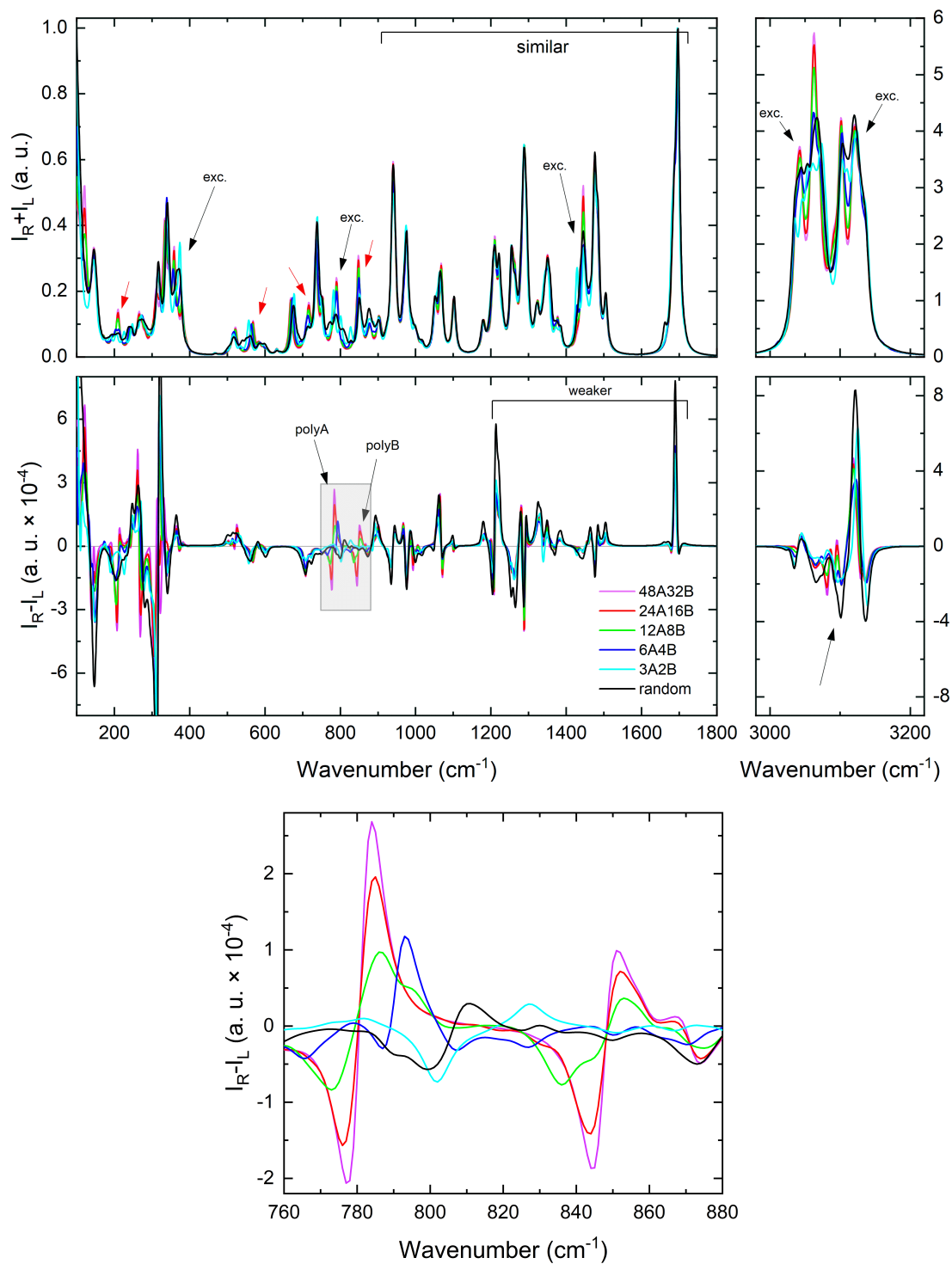
Figure 5.36: Comparison of spectra of (Pro)$_{80}$ with different lengths of polyA and polyB of repetitive sequences (top). Detail of polyA and polyB characteristic couplets (bottom).

5.9 Comparison of the experimental and calculated spectra

In this section we finally compare experimental Raman and ROA spectra with spectra predicted by quantum chemistry calculations.

Raman spectra of PPI were obtained during many different measurements. We performed measurements in solid phase, aqueous solution and organic solvents (1-butanol, 1-propanol). All measurements performed in aqueous solution varied by temperature and time of acquisition. Water is very good solvent for measurement of Raman spectra and the quality of the obtained spectra is very good, so we will use the spectrum obtained by SVD analysis of low temperature measurement in aqueous solution (section 5.4.2) for comparison with the calculated spectra.

The situation with ROA spectra is somewhat more complicated. The acquisition times were significantly longer, in addition, we found that immediately after the dissolution there is a disruption of the regularity of the PPI conformation, which is reflected in the ROA spectrum (see section 5.6). For comparison in the 100-1800 cm^{-1} region we therefore use the spectrum in 1-butanol. However, we were not successful in obtaining the ROA spectrum in 1-butanol in CH region, therefore, we use spectrum in water for comparison. (For the more detailed discussion of experimental spectra see sections 5.4.3 and 5.6)

From the calculated spectra, we selected for comparison a spectrum with a length of 20 proline with a random distribution of puckerings ((Pro)$_n$ *random*). This spectrum was chosen because the bands intensities in the ROA spectra in the region around 300 cm^{-1} (V-pattern) increase with increasing length of PPI chain. The spectrum 20 has a reasonable intensities in this region and at the same time does not otherwise differ from the spectra of longer chains. (See the section 5.8.2 for more details.)

Table 5.7: Ratios of selected bands in experimental and calculated spectra.

Bands	Calculated	Experimental
9:11	0.4	0.4
18:19	1.6	1.5
26:27	0.5	0.6
32:34	0.7	2.2

Fig. 5.37 shows comparison of experimental and calculated spectra. At first glance, the spectra correspond quite well. The positions of bands are shifted to higher frequencies in calculated spectra, this is the consequence of harmonic approximation and represents a systematic error.

Significant Raman bands (4, 9, 18, 19, 26, 27, 34) can be easily assigned. Furthermore, ratios of their intensities are also similar (see Table 5.7). The exception is pair of bands 32 and 34, where the ratio of intensities is (not exactly) reversed. Intensities and shapes in regions with weaker bands (200-300, 400-650, 760-920, 1000-1160, 1300-1400 cm^{-1}) vary more markedly, however, similar spectral pattern can still be identified. In the CH region, experimental spectrum is less symmetrical and bands are wider, which results in slightly different shape.

In the ROA spectrum, we can again identify and easily assign all significant bands. There is a characteristic V-pattern in calculated spectra, positive part of couplet 1 is slightly weaker than positive part of couplet 2. V-pattern is followed by negative band 4, which is wider in experimental spectrum. The most of the other bands are slightly weaker in calculated spectrum than in experimental (e. g. band 6). The negative part of couplet 22 is much weaker in calculated spectrum, band 33 has much lower intensity and different shape in calculated spectrum, and calculated intensity of Amide I band is much stronger than observed in experiment. However, the intensity of band 19, which is characteristic for long-order regularity of PPI chain, is well reproduced. The experimental spectrum is quite noisy in CH region, however, it is evident, that experimental and calculated spectra match very well.

Despite all the above differences, the agreement of the experimental and calculated spectra is very good. All characteristic regions have the same or at least a similar shape in the calculated spectrum. In addition, the absence of significant couplets in the region of about $800\,\mathrm{cm}^{-1}$ in the experimental spectrum and the similarity with the spectrum from random puckerings provide us with the information that long sequences with uniform A and B puckering pattern are not present in real PPI chain.) (discussed in section 5.8.2).
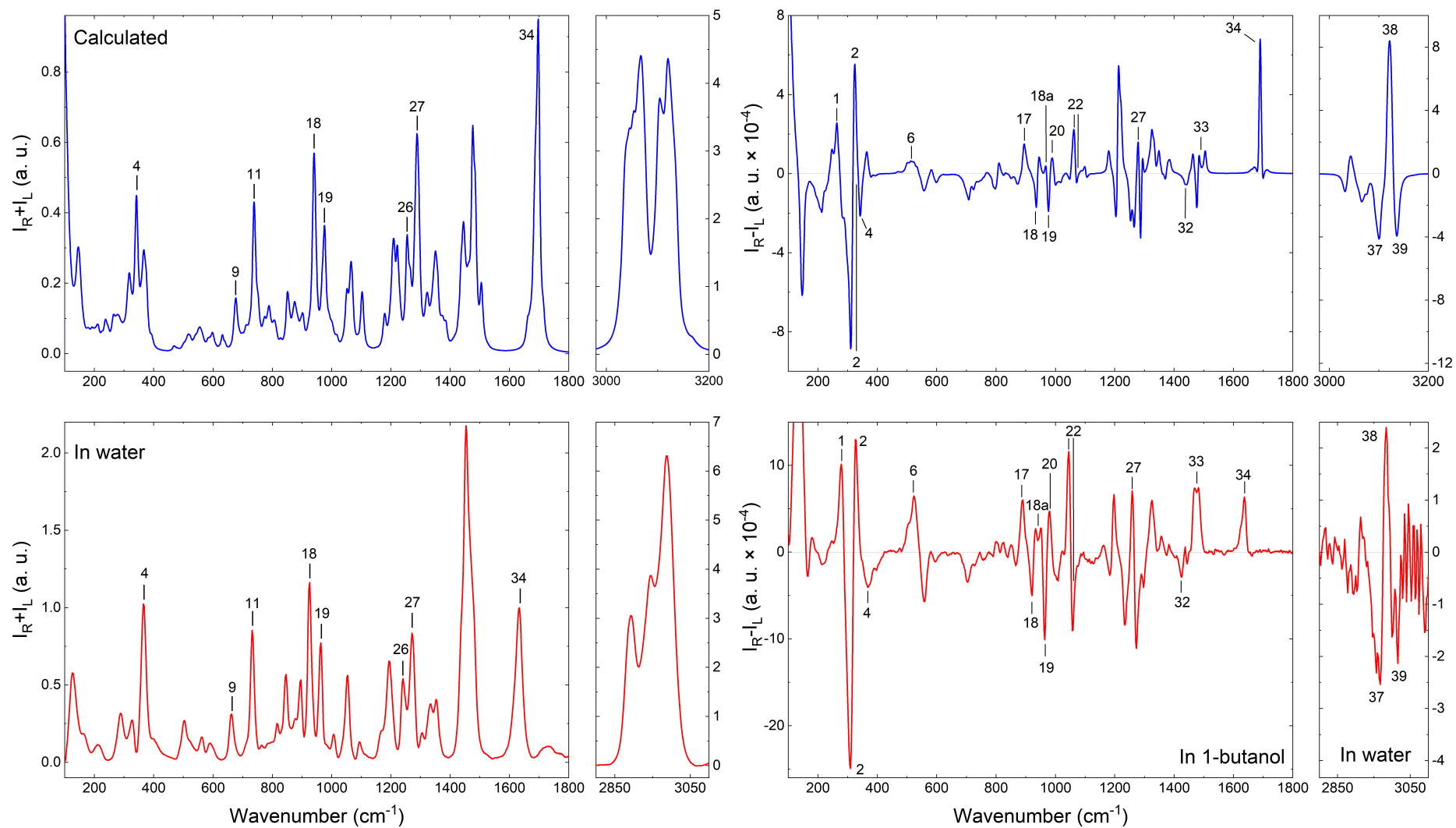
Figure 5.37: Comparison of calculated (*random* (Pro)$_{20}$ and experimental Raman (left) and ROA (right) spectra (in water and 1-butanol).

Conclusion

We focused on characterization of PPI protein secondary structure via ECD, Raman scattering, ROA, and DFT simulations. We were particularly interested in spectral manifestations of regular PPI helix, finding the characteristic spectral markers and evaluation of puckering pattern preferences of polyproline chain. We were able to satisfactory fulfil all goals specified in chapter Aims:

1. We successfully prepared a set of polyproline samples in PPI conformation differing in the mean chain length. However, there is still room for improvement as the solubility of isolated solid phase samples was far from ideal in both aliphatic alcohols and water.

2. We successfully obtained solid phase Raman spectra of both PPI and PPII conformations and identified characteristic spectral markers of each of them. We addressed the changes in spectral profiles connected to the different mean length of polyproline chain.

3. We successfully repeated measurement of PPI to PPII mutarotation process by ECD. Above the published results we were able to perform statistical analysis of recorded data, which revealed that ECD (and also UV absorption) recognizes mutarotation as a three stage process (regular PPI – intermediate – regular PPII). From the fit model of chemical kinetics, we determined corresponding conversion half-life as $\tau_1 = 0,4$, $\tau_2 = 8,4$ hour at temperature $35\,°\mathrm{C}$.

4. We acquired Raman and ROA spectra of PPI in 1-propanol and 1-butanol, while ROA spectrum from measurement in 1-butonal has exceptionally good quality. From the low temperature measurements ($5\,°\mathrm{C}$) combined with advanced analysis of recorded data set we obtained Raman spectrum of regular PPI helix in water, and also ROA spectrum from the beginning of mutarotation process with strongly prevailing PPI content. It should be noted that our results represent the very first measurement of Raman and ROA spectrum of PPI conformation.

   We studied the PPI to PPII mutarotation also via Raman and ROA measurements at various temperatures. We found that at temperatures higher than $20\,°\mathrm{C}$ the Raman scattering is able to distinguish 3 different polyproline forms (similarly to ECD) while ROA due to its inherently higher noise level only two of them.

   Yet, from the comparison of Raman and ROA PPI spectra from 1-butanol and water with spectra of PPII, we were able to identify general markers of both the conformations as well as vibrational modes related to the long-range regularity of PPI and PPII helices.

5. Based on the simulated spectra of proline tetramers with AAAA puckering we were able to assign majority of spectral features in both Raman and ROA spectra to specific vibrational motions.

We performed quantum chemical calculations for a large group of proline heptamers in PPI and PPII conformations differing in puckering and performed statistical processing of optimized geometries. We observed the dependence of dihedral angles on the local combination of puckerings. The obtained results were used for creation of long polyproline chain in PPI or PPII conformations of arbitrary puckering pattern. To facilitate the implementation of these calculations, we have created our own programs.

6. To observe changes in PPI spectra caused by different mean length of PP chain and puckering pattern, we performed quantum chemical calculations for several hundred larger molecules in PPI conformation differing in the length and distribution of puckerings. We reduced the computational cost by performing quantum chemical calculations for repeating fragments only once. We used our own program cooperating with programs for calculations of transfer of molecular property tensor (by Prof. Petr Bouř). We found that most bands in Raman and ROA spectra are sensitive to the length of PPI helix up to chain length approximately 20 residues (only few of them change in intensity in the chain length $\approx$30). Also, we showed that distribution of puckering along the PP chain is probably quite random, the occurrence of uniform sequences of more than 6 consecutive A and B puckerings is highly improbable.

7. We achieved great correspondence between experimental and simulated Raman and ROA spectra in the whole region of fundamental molecular vibrations. The key steps were usage of transfer of molecular property tensor approach and averaging of the calculated spectra over several sequences of random puckering patterns.

In the future, we would like to study the whole mutarotation at low temperature by Raman spectroscopy and ROA in combination with 2D-correlation spectroscopy analysis, which might reveal the sequence of spectral changes and give better idea about reaction mechanism.

Furthermore, we would like to extend the computational part by using the obtained information about PPI and PPII geometries and to suggest a possible structure of intermediate products based on calculations, literature and estimation. Performing spectral calculations of these intermediates and comparing with the results of using singular value decomposition and the kinetics fit could provide interesting insight into the process of mutarotation.

Bibliography

[1] A. Adzhubei, F. Eisenmenger, V. Tumanyan, M. Zinke, S. Brodzinski, and N. Esipova. Third type of secondary structure: Noncooperative mobile conformation. Protein Data Bank analysis. *Biochemical and Biophysical Research Communications*, 146(3):934–938, aug 1987.

[2] A. A. Adzhubei, M. J. Sternberg, and A. A. Makarov. Polyproline-II Helix in Proteins: Structure and Function. *Journal of Molecular Biology*, 425(12):2100–2132, jun 2013.

[3] C. Altona and M. Sundaralingam. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *Journal of the American Chemical Society*, 94(23):8205–8212, nov 1972.

[4] S. A. Asher, A. V. Mikhonin, and S. Bykov. UV Raman Demonstrates that α-Helical Polyalanine Peptides Melt to Polyproline II Conformations. *Journal of the American Chemical Society*, 126(27):8433–8440, jul 2004.

[5] L. D. Barron, M. P. Bogaard, and A. D. Buckingham. Differential Raman Scattering of Right and Left Circularly Polarized Light by Asymmetric Molecules. *Nature*, 241(5385):113–114, jan 1973.

[6] L. D. Barron, M. P. Bogaard, and A. D. Buckingham. Raman scattering of circularly polarized light by optically active molecules. *Journal of the American Chemical Society*, 95(2):603–605, jan 1973.

[7] L. D. Barron and A. D. Buckingham. Vibrational optical activity. *Chemical Physics Letters*, 492(4-6):199–213, jun 2010.

[8] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, apr 1993.

[9] R. Berisio, S. Loguercio, A. De Simone, A. Zagari, and L. Vitagliano. Polyproline Helices in Protein Structures: A Statistical Survey. *Protein & Peptide Letters*, 13(8):847–854, aug 2006.

[10] H. M. Berman. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, jan 2000.

[11] E. W. Blanch, L. A. Morozova-Roche, D. A. Cochran, A. J. Doig, L. Hecht, and L. D. Barron. Is polyproline II helix the killer conformation? a raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. *Journal of Molecular Biology*, 301(2):553–563, aug 2000.

[12] P. Bouř. Convergence Properties of the Normal Mode Optimization and Its Combination with Molecular Geometry Constraints. *Collection of Czechoslovak Chemical Communications*, 70(9):1315–1340, 2005.

[13] P. Bouř and T. A. Keiderling. Partial optimization of molecular geometry in normal coordinates and use as a tool for simulation of vibrational spectra. *The Journal of Chemical Physics*, 117(9):4126–4132, sep 2002.

[14] P. Bouř, J. Sopková, L. Bednárová, P. Maloň, and T. A. Keiderling. Transfer of molecular property tensors in cartesian coordinates: A new algorithm for simulation of vibrational spectra. *Journal of Computational Chemistry*, 18(5):646–659, apr 1997.

[15] P. Bour. Web page of prof. Petr Bouř.

[16] P. M. Cowan and S. McGavin. Structure of Poly-L-Proline. *Nature*, 176(4480):501–503, sep 1955.

[17] P. M. Cowan, S. McGavin, and A. C. North. The polypeptide chain configuration of collagen. *Nature*, 176(4492):1062–1064, 1955.

[18] R. K. Dukor and T. A. Keiderling. Mutarotation studies of poly-L-proline using FTIR, electronic and vibrational circular dichroism. *Biospectroscopy*, 2(2):83–100, dec 1998.

[19] C. Grathwohl and K. Wüthrich. The X-Pro peptide bond as an nmr probe for conformational studies of flexible linear peptides. *Biopolymers*, 15(10):2025–2041, oct 1976.

[20] W. Hug and G. Hangartner. A novel high-throughput Raman spectrometer for polarization difference measurements. *Journal of Raman Spectroscopy*, 30(9):841–852, sep 1999.

[21] S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*, volume 5 of *Series on Multivariate Analysis*. World Scientific Publishing Co. Pte. Ltd., 2001.

[22] Š. Jílek. Vibrational optical activity of 3-aminoquinuclidine. BSc. thesis, 2019.

[23] S. Kakinoki, Y. Hirano, and M. Oka. On the Stability of Polyproline-I and II Structures of Proline Oligopeptides. *Polymer Bulletin*, 53(2):109–115, jan 2005.

[24] Y. K. Kang. Cis-Trans Isomerization and Puckering of Pseudoproline Dipeptides. *The Journal of Physical Chemistry B*, 106(8):2074–2082, feb 2002.

[25] Y. K. Kang and B. J. Byun. Conformational preferences and cis-trans isomerization of azaproline residue. *Journal of Physical Chemistry B*, 111(19):5377–5385, 2007.

[26] Y. K. Kang, J. S. Jhon, and H. S. Park. Conformational Preferences of Proline Oligopeptides. *The Journal of Physical Chemistry B*, 110(35):17645–17655, sep 2006.

[27] J. Kapitán, V. Baumruk, V. Kopecký, R. Pohl, and P. Bouř. Proline Zwitterion Dynamics in Solution, Glass, and Crystalline State. *Journal of the American Chemical Society*, 128(41):13451–13462, oct 2006.

[28] M. Kuemin, J. Engel, and H. Wennemers. Temperature-induced transition between polyproline I and II helices: quantitative fitting of hysteresis effects. *Journal of Peptide Science*, 16(10):596–600, oct 2010.

[29] J. Kurtz, A. Berger, and E. Katchalski. Mutarotation of Poly-L-proline. *Nature*, 178(4541):1066–1067, nov 1956.

[30] P. Larkin. IR and Raman Spectra-Structure Correlations. In *Infrared and Raman Spectroscopy*, pages 73–115. Elsevier, 2011.

[31] H. Li and L. A. Nafie. Simultaneous acquisition of all four forms of circular polarization Raman optical activity: results for α-pinene and lysozyme. *Journal of Raman Spectroscopy*, 43(1):89–94, jan 2012.

[32] A. Makarov, N. Esipova, Y. Pankov, V. Lobachev, and B. Grishkovsky. A conformational study of β - melanocyte-stimulating hormone. *Biochemical and Biophysical Research Communications*, 67(4):1378–1383, dec 1975.

[33] A. A. Makarov, N. G. Esipova, V. M. Lobachov, B. A. Grishkovsky, and Y. A. Pankov. Secondary structure of polypeptide hormones of the anterior lobe of the pituitary gland. *Biopolymers*, 23(1):5–22, jan 1984.

[34] E. R. Malinovski. *Factor Analysis in Chemistry*. Wiley, New York, 3 edition, 2002.

[35] C. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review*, 46(7):618–622, oct 1934.

[36] L. A. Nafie. Infrared and Raman Vibrational Optical Activity: Theoretical and Experimental Aspects. *Annual Review of Physical Chemistry*, 48(1):357–386, oct 1997.

[37] L. A. Nafie. *Vibrational Optical Activity*. John Wiley & Sons, Ltd, Chichester, UK, aug 2011.

[38] National Center for Biotechnology Information. Proline. https://pubchem.ncbi.nlm.nih.gov/compound/Proline, 2019.

[39] F. S. Parker. Proteins and Polypeptides. In *Applications of Infrared Spectroscopy in Biochemistry, Biology, and Medicine*, pages 188–231. Springer US, Boston, MA, 1971.

[40] M. Pazderková, V. Profant, J. Hodačová, J. Šebestík, T. Pazderka, P. Novotná, M. Urbanová, M. Šafařík, M. Buděšínský, M. Tichý, L. Bednárová, V. Baumruk, and P. Maloň. Nonplanar Tertiary Amides in Rigid Chiral Tricyclic Dilactams. Peptide Group Distortions and Vibrational Optical Activity. *The Journal of Physical Chemistry B*, 117(33):9626–9642, aug 2013.

[41] V. Profant, V. Baumruk, X. Li, M. Šafařík, and P. Bouř. Tracking of the Polyproline Folding by Density Functional Computations and Raman Optical Activity Spectra. *The Journal of Physical Chemistry B*, 115(50):15079–15089, dec 2011.

[42] G. Ramachandran and G. Kartha. Structure of Collagen. *Nature*, 176(4482):593–595, sep 1955.

[43] G. Ramachandran, A. Lakshminarayanan, R. Balasubramanian, and G. Tegoni. Studies on the conformation of amino acids XII. Energy calculations on prolyl residue. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 221(2):165–181, nov 1970.

[44] A. Rich and F. H. C. Crick. The Structure of Collagen. *Nature*, 176(4489):915–916, nov 1955.

[45] V. Sasisekharan. Structure of poly-L-proline. II. *Acta Crystallographica*, 12(11):897–903, nov 1959.

[46] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, and W. A. Eaton. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proceedings of the National Academy of Sciences*, 102(8):2754–2759, feb 2005.

[47] L. Shi, A. E. Holliday, H. Shi, F. Zhu, M. A. Ewing, D. H. Russell, and D. E. Clemmer. Characterizing Intermediates Along the Transition from Polyproline I to Polyproline II Using Ion Mobility Spectrometry-Mass Spectrometry. *Journal of the American Chemical Society*, 136(36):12702–12711, sep 2014.

[48] I. Z. Steinberg, W. F. Harrington, A. Berger, M. Sela, and E. Katchalski. The Configurational Changes of Poly-L-proline in Solution. 82(20):5263–5279, oct 1960.

[49] P. J. Stephens, F. J. Devlin, and J. R. Cheeseman. *VCD spectroscopy for organic chemists.* CRC Press, Boca Raton, 2012.

[50] W. Traub and U. Shmueli. Structure of Poly-L-Proline I. *Nature*, 198(4886):1165–1166, jun 1963.

[51] L. Vitagliano, R. Berisio, A. Mastrangelo, L. Mazzarella, and A. Zagari. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Science*, 10(12):2627–2632, apr 2009.

[52] D. Voet and J. G. Voet. *Biochemistry.* John Wiley & Sons, Hoboken, NJ, 4 edition, 2011.

[53] R. W. Woody. Theory of Circular Dichroism of Proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, pages 25–67. Springer US, Boston, MA, 1996.

[54] D. Wu. The puckering free-energy surface of proline. *AIP Advances*, 3(3):032141, mar 2013.

[55] R. Zhang and J. S. Madalengoitia. Conformational stability of proline oligomers. *Tetrahedron Letters*, 37(35):6235–6238, 1996.

List of Figures

List of Tables

List of Abbreviations

AA   amino acid
CD   circular dichroism
DFT   density functional theory
DNA   deoxyribonucleic acid
BO   Born-Oppenheimer approximation
FTIR   Fourier transform infrared spectroscopy
FFR   Far-from-resonance approximation
IR   infrared spectroscopy
PPI   polyproline I
PPII   polyproline II
Pro   proline
ROA   Raman optical activity
RNA   ribonucleic acid
SVD   singular value decomposition
VCD   vibrational circular dichroism

A. Attachments

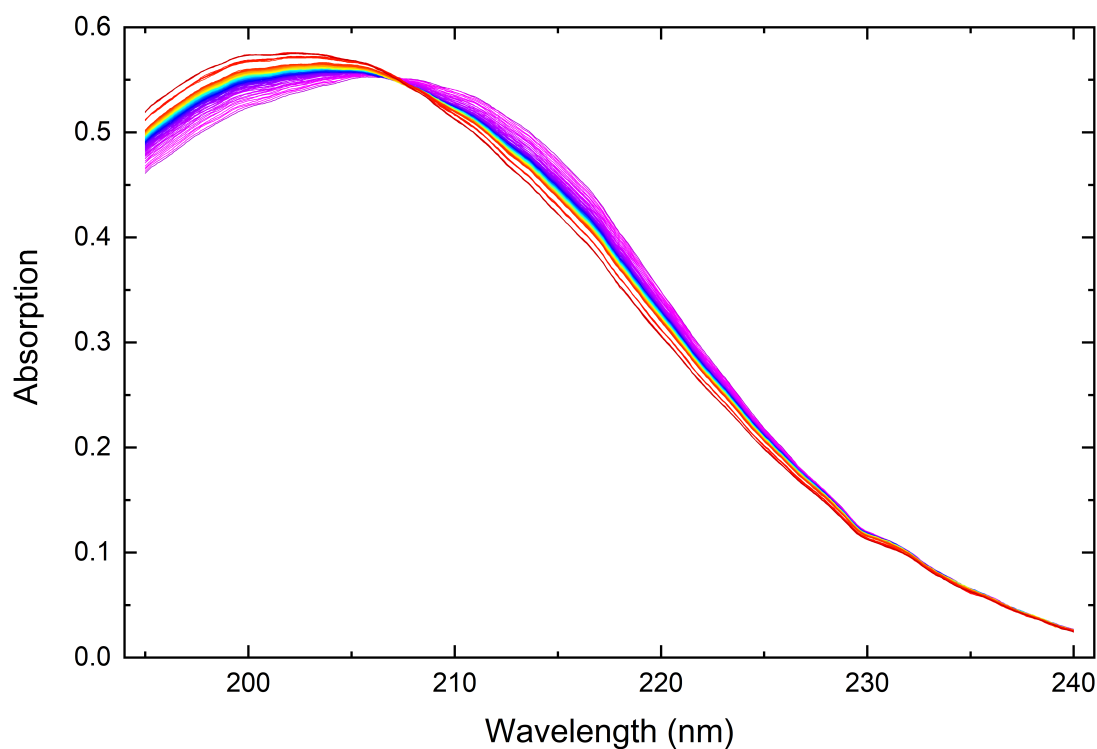## A.1 Factor analysis of mutarotation using UV absorption

Figure A.1: UV spectra obtained during mutarotation measurement. Violet first, red last. The shift of the band maximum to shorter wavelengths is noticeable.
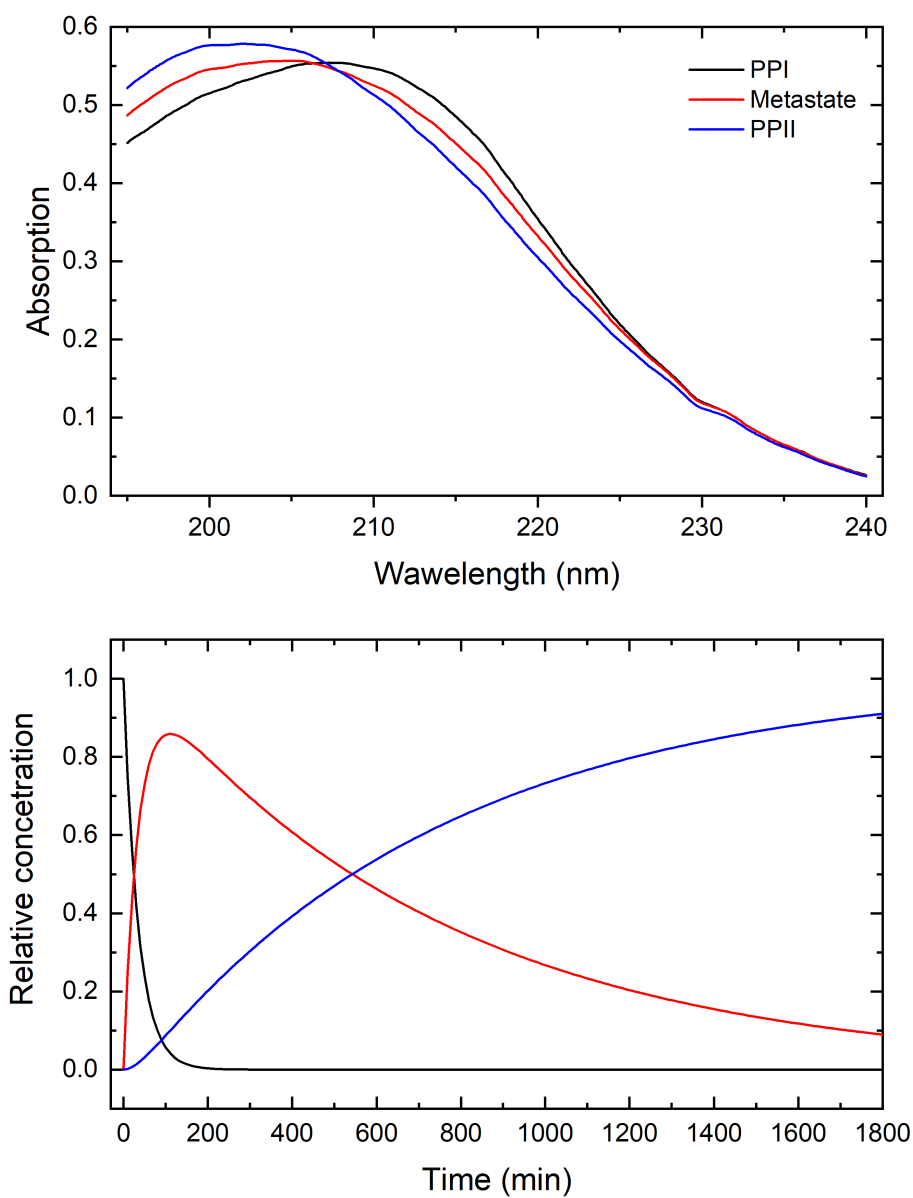
Figure A.2: UV spectra for each fitted component and their relative concentrations in time. Due to the fixed reaction rates used in the fit, the concentrations are identical to those obtained from the ECD.
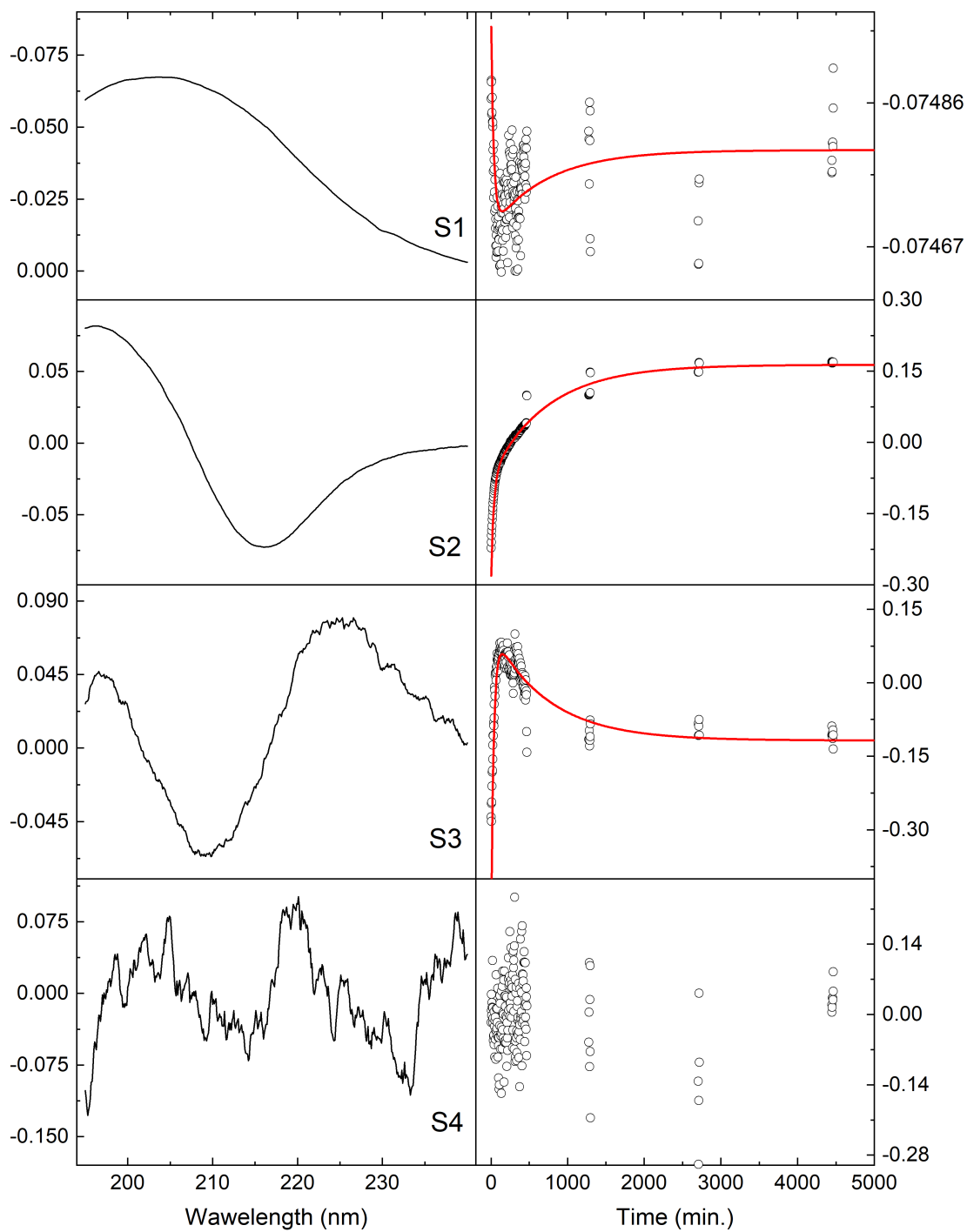
Figure A.3: First four subspectra for UV mutarotation measurement.

A.2 Solid state PPII spectral bands

Table A.1: Position of bands in PPII spectra in Fig. 5.3

No.	short	medium	long
1	312 w	312 w	312 w
2	337 w	339 w	337 w
3	362 w	359 w	362 w
4	405 w	405 w	405 w
5	496 s	496 s	496 s
6	541 w	541 w	539 w
7	577 w	580 w	580 w
8	596 w	600 w	
9	673 w	673 w	673 w
10	700 w	698 w	700 w
11	736 m	736 m	734 w
12	764 w	761 w	764 w
13	784 w	780 w	784 w
14	836 m	836 m	836 m
15	866 m	868 m	868 m
16	895 m	895 m	895 w
17	918 s	918 s	920 s
18	950 w	950 w	952 w
19	975 w	975 w	973 w
20	998 w	998 w	998 w
21	1050 m	1050 w	1050 w
22	1093 w	1093 w	1093 w
23	1159 w	1159 w	1159 w
24	1179 w	1179 w	1179 w
25	1204 w	1204 w	1204 w
26	1238 m	1241 m	1241 m
27	1263 m	1266 m	1268 m
28	1304 w	1302 w	1307 w
29	1343 w	1318 w	1318 w
30	1420 w	1343 w	1345 w
31	1447 s	1420 w	1418 w
32	1482 m	1447 s	1450 s
33	1604 w	1482 m	1479 m
34	1652 s	1652 s	1650 m
35	2881 s	2881 s	2879 s
36	2933 s	2933 s	2933 s
37	2963 s	2972 s	2956 s
38	2979 s	2981 s	2981 s
39	2999 s		

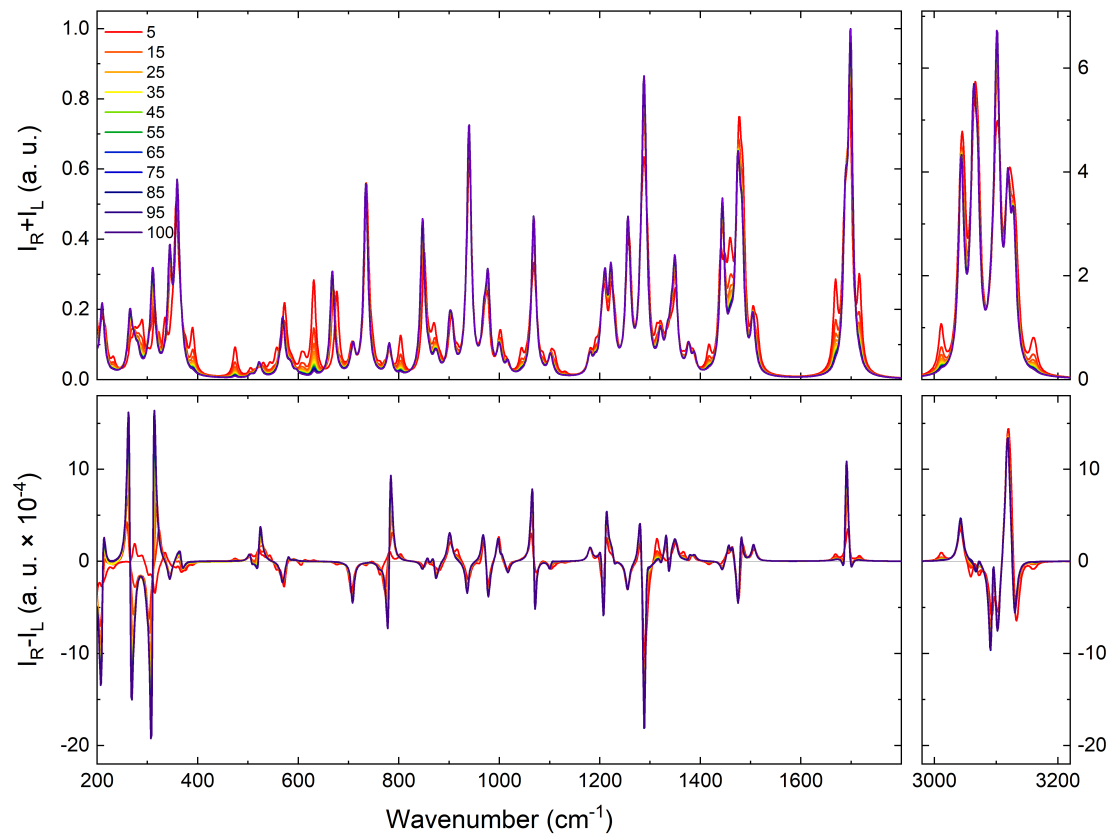# A.3 Spectra of longer PPI molecules calculated using transfer of molecular property tensors

Figure A.4: Length dependence of polyA chains of polyproline calculated using tetramers. All lengths from 5 to 100, step 5 (only some of them included in the legend)
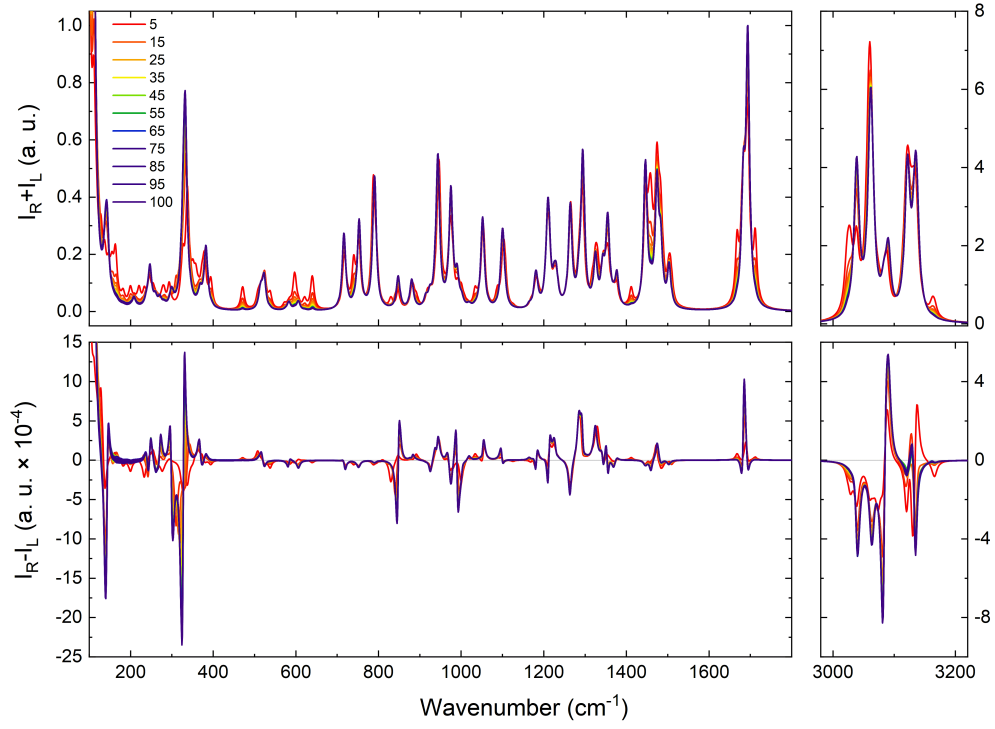
Figure A.5: Length dependence of polyB chains of polyproline calculated using tetramers. All lengths from 5 to 100, step 5 (only some of them included in the legend).
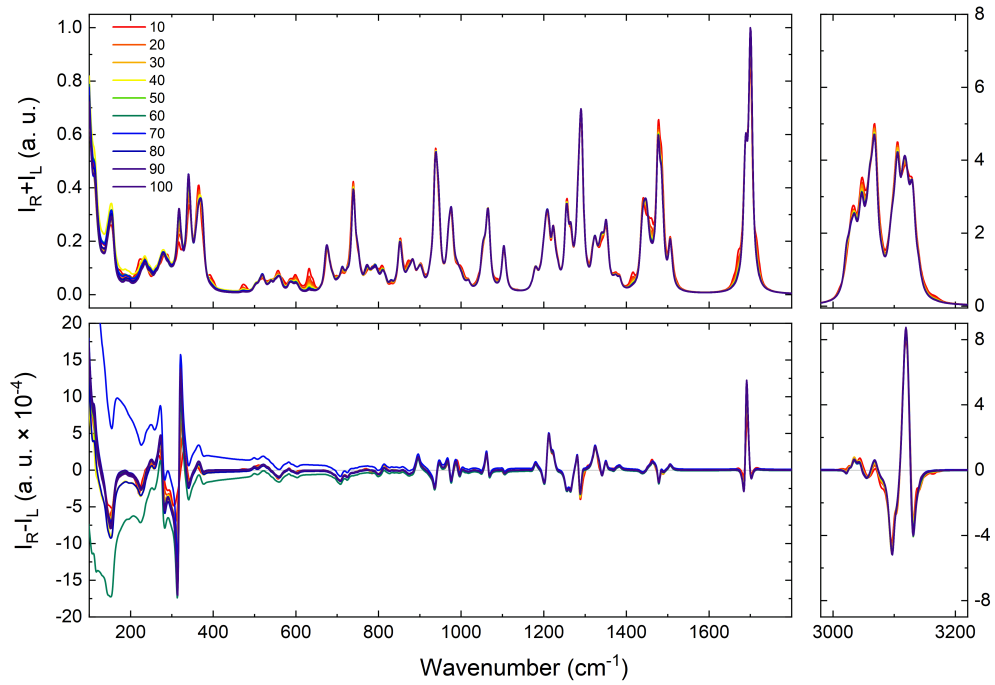

Figure A.6: Length dependence of random chains of polyproline calculated using trimers. All lengths from 10 to 100, step 5 (only some of them included in the legend). Shifts in low-frequency region are cause by strong bands with negative energy which deform the spectrum.
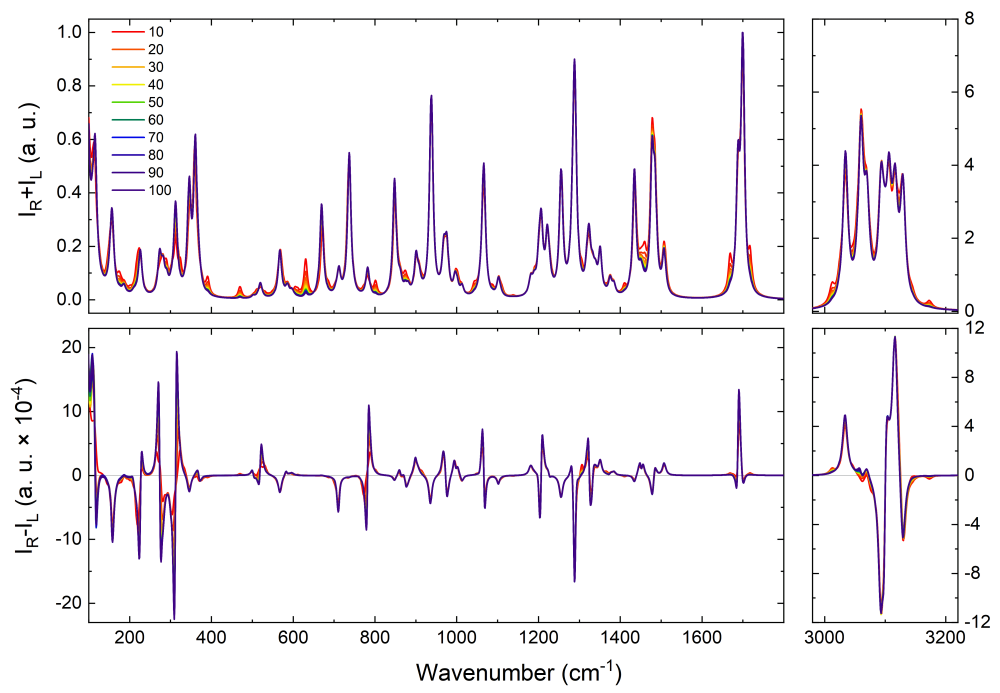
Figure A.7: Length dependence of polyA chains of polyproline calculated using trimers. All lengths from 10 to 100, step 5 (only some of them included in the legend).