Manual for the Practical

Fourier-Transform Infrared Spectroscopy of Proteins

Advanced practical at the
Institute for Applied Physics
University of Tübingen

Assistant: F. Zhang
Room: C7 P28
Tel.: (07071) 29 75242

Tübingen
October 2018

Contents

1. Preliminary remarks

1.1. Laser safety

Attention!

During this experiment, you will use a machine containing a laser (safety class 3R) with $\sim 1\,\mathrm{mW}$ power. The laser beam can damage your eyes, if not attenuated. It is not allowed to use the apparatus in any other way than the one described in the manual, for reasons of safety.

Please note, though, that the laser (with a wavelength in the visible part of the electromagnetic spectrum) is for calibrating the distances of the mirrors and not for IR spectroscopy itself.

1.2. Educational objective and teaching goal

The aim of the experiment is to make you familiar you with a modern infrared spectroscopy technique. With the help of Fourier transform infrared spectroscopy you will examine the secondary structure of proteins. While doing so you will also get to know essential steps for the preparation of biological samples.

1.3. Prior knowledge

This section serves as a guideline for your preparation for the experiment. The points listed in the following raise no claim to completeness but rather should encourage you to work through the relevant literature by yourself. The literature information at the end of this section is to be seen in this light, too.

1.4. Guideline for the prepartion of the experiment

1. molecular vibrations
2. absorption bands of peptides
3. Fourier transform
4. Michelson–Interferometer
5. Fourier transform infrared spectroscopy

1.5. Literature

- Acquisition of basic knowledge
  1. E. Hecht, *Optik*, Oldenburg, 2001.
  2. W. Demtröder, *Molekülphysik*, Oldenburg, 2003.
  3. F. Schreiber, *Soft Matter*, lecture handouts, University of Tübingen, 2018.
- Additional information
  1. H. Günzler, H.-U. Gremlich, *IR–Spektroskopie*, Wiley-VCH, 2003.
  2. A. Barth, C. Zscherp, *What vibrations tell us about proteins*, Quart. Rev. Biophys. 35(4), 369–430 (2002)
  3. J. L. R. Arrondo, A. Muga, J. Castresana and F. M. Goñi, *Quantitative studies of the structure of proteins in solution by fourier-transform infrared spectroscopy*, Prog. Biophys. Molec. Biol. 59(1) 23–56 (1993)
  4. E. Goormaghtigh, V. Cabiaux, J. M. Ruysschaert, *Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy*, Subcell. Biochem 23 329–450 (1994)
  5. W. K. Surewicz, H. H Mantsch, D. Chapman, *Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment*, Biochem. 32(2), 389–394 (1993)

2. Infrared spectroscopy of proteins

2.1. Vibrations of molecules

The interaction of a molecule with a radiation field is directly proportional to the dipole moment of the molecule (see also books on atomic and molecular physics). Absorption or emission of radiation therefore can be observed only if the dipole moment is changed, neglecting higher order multipole transitions.

This allows to infer the structure of a molecule by observing the absorbed or emitted radiation. Since molecules composed by two identical atoms (*homonuclear diatomic* molecules e.g. N_2, O_2, Br_2) do not show any variation of the dipole moment upon vibration or rotation, no absorption or emission of radiation is normally observed for these molecules.

The number of the possible vibration modes of polyatomic molecules can be easily derived: since each single atom can move in the three directions of space (i.e.: x, y, z), a molecule consisting of N atoms possesses $3N$ degrees of freedom. However 3 of them correspond to rotations around three axis through the centre of mass and other 3 to translations of the centre of mass in one of the three directions, making the number of *vibrational* degrees of freedom Z=$3N-6$. For linear molecules there are only 2 rotational degrees of freedom: the rotation around the molecular axis has no effect since it doesn't change the coordinates of the atoms or of the centre of mass. Therefore in this case there are $Z = 3N - 5$ vibrational degrees of freedom. For each of these vibrational degrees of freedom a so-called *normal mode* can be found. All the atoms involved in a normal mode oscillate with the same frequency and constant phase. Each normal mode is assigned to a specific vibrational frequency.

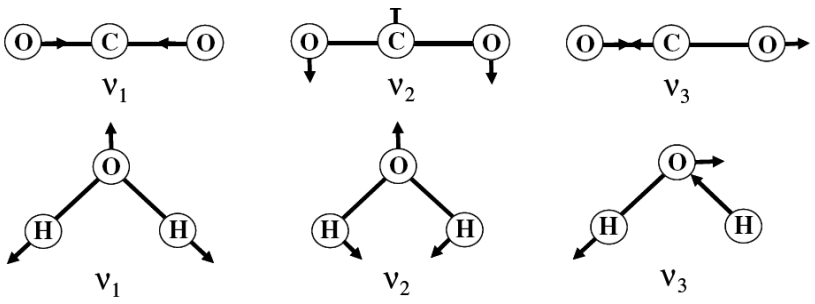

Figure 2.1.: Normal modes of the molecules CO_2 and H_2O. Upper (CO_2): symmetric stretching vibration $\nu_1 = 1337\,\mathrm{cm}^{-1}$, symmetric bending vibration $\nu_2 = 667\,\mathrm{cm}^{-1}$, antisymmetric stretching vibration $\nu_3 = 2349\mathrm{cm}^{-1}$. Lower ($H_2O$): symmetric stretching vibration $\nu_1 = 3657\,\mathrm{cm}^{-1}$, symmetric bending vibration $\nu_2 = 1595\,\mathrm{cm}^{-1}$, antisymmetric stretching vibration $\nu_3 = 3756\,\mathrm{cm}^{-1}$.

2.2. Absorption in the infrared spectral range

When a molecule absorbs light in the infrared wavelength range, it is excited to a higher energy vibrational state. Absorption can only take place, when there is a change in the dipole moment of the involved atom group during the oscillation. The intensity of an absorption band is proportional to the square of the variation of the dipole moment relative to the equilibrium. This variation is caused by *stretching vibrations* through variation of the bond lengths, and/or by *bending vibrations* and other deformations which imply variations in the bond angles. The frequency, at which a vibration absorbs, depends on the reduced mass of the atoms involved, the force constant of the bond, the binding to vicinal atoms and the coupling with other molecular vibrations. This frequency, for a diatomic molecule (composed of two different atoms) is calculated according to:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}, \tag{2.1}$$

where ν is the oscillation frequency (same as absorption frequency), k the force constant and μ the reduced mass:

$$\mu = \frac{m_1 m_2}{m_1 + m_2}, \tag{2.2}$$

Often the energy-proportional *wavenumber*, which is the inverse of the wavelength, is used instead of the frequency. For the wavenumber $\tilde{\nu}$, the unit commonly used is cm^{-1} (reciprocal centimeter).

Wavenumber / cm^{-1}	1000	2000	3000	4000	5000	6000	8000	10000
Wavelength / μm	10	5	3.33	2.5	2	1.67	1.25	1

Table 2.1.: Wavenumber and Wavelength

2.3. The structure of proteins

Proteins can be described as long chains of amino acids: for this reason there are, in principle, a lot of ways in which such a chain can be folded. A first essential constraint resides in the special properties of the peptide bond (see Figure. 2.2), which determines a planar arrangement of its CO and NH groups. It follows that a peptide chain consists of rigid, planar parts, each connected by a Carbon atom through two dihedral angles, ϕ and ψ (see Figure. 2.3). Due to boundary conditions, the ϕ and ψ angles can take only particular values. Moreover, hydrogen bonds play an important role in the formation of the structure: in particular, the hydrogen bonds formed between the NH and CO groups of the backbone. The primary structure of proteins describes the sequence of the amino acids, while the secondary structure describes the particular spatial arrangement of consecutive amino acids.

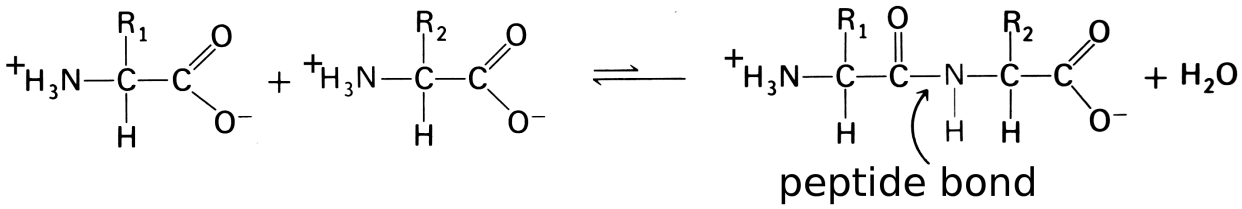

Figure 2.2.: Structure and synthesis of a peptide bond via elimination of H_2O.

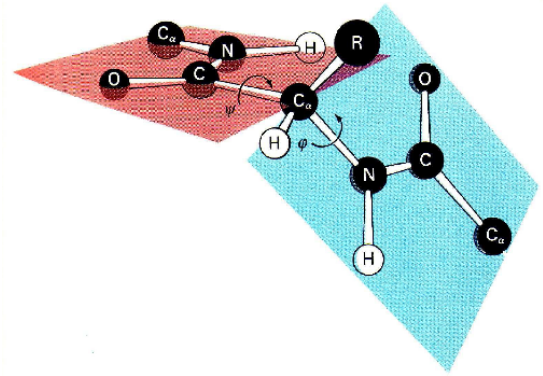

Figure 2.3.: Schematic diagram of a peptide bond.

Secondary structure

α-Helices

The α-Helix is a rod-shaped structure, in which the internal part of the rod is made by a tightly wound polypeptide chain; the side chains point outwards from this helical arrangement (see Figure. 2.4).

The α-Helix is stabilized by hydrogen bonds between the NH and CO groups of the backbone. The CO group of an amino acid forms a hydrogen bond with the NH group of the amino acid which is four residues back in the sequence. All the CO and NH groups of the backbone take part in hydrogen bonds. The distance between consecutive residues along the helix axis is 1.5 Å, and they are twisted by 100°. The diameter of the α–Helix measures 5.4 Å. Normally, right handed α-Helices are found in proteins. The amount of protein chain which is folded in an α-Helix structure varies strongly between different proteins.

β-Sheets

While in an α-Helix the hydrogen bonding network is between amino acids which belong to one strand, in the β-Sheet structure hydrogen bonds are formed between different strands (see Figure. 2.5). The β-Sheet can result in flat, bent or rod-like structures. In a β-Sheet the polypeptide chain is almost completely extended. Adjacent strands in a β-Sheet can have the same direction (parallel β–Sheets) or opposite directions (antiparallel β-Sheets).

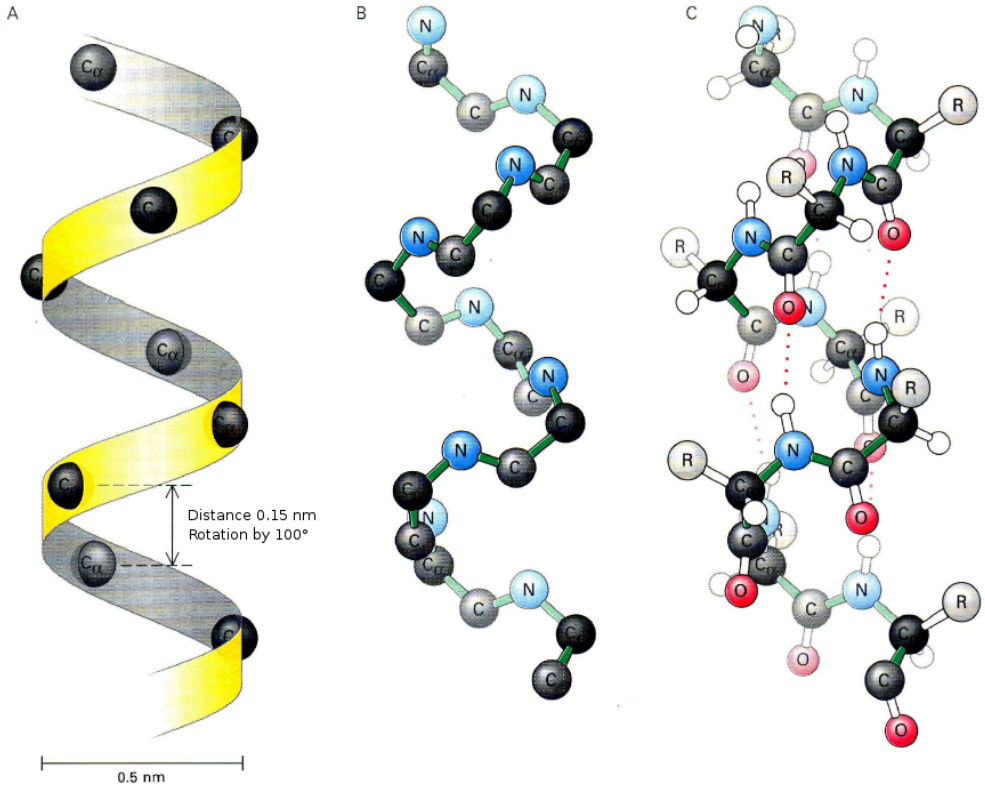

Figure 2.4.: Schematic diagram of an α-Helix.

2.4. Infrared spectroscopy of proteins

Infrared spectroscopy is becoming increasingly important for the understanding of biological molecules, in particular because it delivers information that cannot be extracted with structural elucidation methods such as NMR, Electron Microscopy and X-ray diffraction. It allows to draw conclusions on the structure and the environment of the protein backbone and of many amino acid side chains. It allows to access dynamic processes, as, for instance, the variations of bond lengths and geometries during the course of a reaction involving the protein, the changes in chemical groups (e.g. protonation) or the altered charge density in a chemical bond.

The sensitivity of modern infrared spectrometers is in principle nowadays good enough to detect, with special techniques, the contributions of individual amino acids. However, a protein absorption spectrum shows many overlapping bands of the peptide backbone (Table 2.2) and of the amino acid side chains.

Furthermore, the -C=O modes of the peptide are overlapped by a strong water band at approx. $1650\,\mathrm{cm}^{-1}$, that, while having a smaller extinction coefficient ($20\,{}^{\mathrm{l}}/_{\mathrm{mol\cdot cm}}$) in comparison with the modes of the amino acid side chains, already at a sample thickness of $10\,\mu$m leads to absorbance equal to 1 (i.e. only 10% of the incoming light are transmitted) due to the high water content of the sample. For the definition of the absorbance see Appendix A.2.

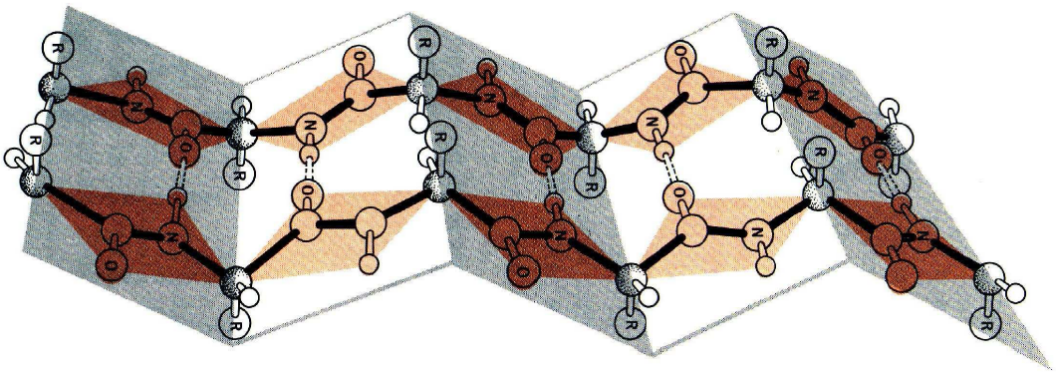

Figure 2.5.: Schematic diagram of the β-Sheet structure.

2.5. Absorption bands of the peptide group

Nine vibrational modes can be attributed to a (hypothetical) free, planar peptide group: these are denoted, in decreasing order of the wavenumber of the absorption band as Amide-A, -B and Amide-I to VII. For structural studies the Amide I is of particular importance: the position of the absorption band of the Amide-I vibration, which contributes about 80% to the C=O stretching of the peptide group, can be correlated with the secondary structure of the polypeptide. This is because the ψ and ϕ angles alter the exact geometry and therefore the position and intensity of the Amide I band (strong effect and change of the dipole interaction between the C=O vibrations). The exact analysis of the Amide I band can provide informations on the secondary structure of proteins in conjunction with mathemathical methods for the separation of overlapping absorption bands (Fourier self-deconvolution).

A hydrogen-deuterium exchange (H/D-exchange) in the peptide, as happens for example in a D_2O solution, changes the masses of the oscillating atoms, and therefore the vibrational frequencies and the position of the absorption bands in the spectrum. The band shift is dependent on the secondary structure: the weak high frequency Amide I band of β-Sheets and loops ($\approx 1670\,\mathrm{cm}^{-1}$) is often strongly shifted to lower wavenumbers by approx. $10\,\mathrm{cm}^{-1}$, while for all the other secondary structure elements the shift is only a few cm^{-1}. Tab. 2.2 shows a summary of the Amide I modes of various secondary structure elements. The Amide II band (40 . . . 60% N–H bending vibration, 18 . . . 40% C–N stretching vibration, about. 10% C–C stretching vibration), in comparison with the Amide I band, shows a much higher sensitivity towards H/D-exchange. It can shift up to $100\,\mathrm{cm}^{-1}$ to smaller wavenumbers. In this case the character of the vibration changes: it becomes mostly a C–N stretching and does not contain N–H bending vibrations anymore. This fact allows to monitor the deuteration level of proteins through the disappearance of the Amide II and the appearance of the so-called Amide II' signals at $1450\,\mathrm{cm}^{-1}$ in D_2O. A similar effect occurs for the Amide A band.

Different parts of the protein, depending on the stability of the protein structure, show different H/D-exchange rates, which can be used to assign the absorption bands to individual secondary structure elements.

Secondary structure	**Band position in H_2O / cm^{-1}**		**Band position in D_2O / cm^{-1}**	
	Mean	Range	Mean	Range
α-Helix	1654	1648 ... 1657	1652	1642 ... 1660
β-Sheet	1633	1623 ... 1641	1630	1615 ... 1638
	1684	1674 ... 1695	1679	1672 ... 1694
Loop	1672	1662 ... 1686	1671	1653 ... 1691
Disordered	1654	1642 ... 1657	1645	1639 ... 1654

Table 2.2.: Amide I modes of different secondary structure elements.
Source: after E. Goormaghtigh et al. (1994), Subcellular Biochemistry 23, 405. Note: in tables like this one different band positions are often reported for the high frequency band of parallel and antiparallel β-Sheets. This is due on the calculation for band position (for parallelβ-Sheets). Experimentally however is not possible to distinguish between parallel and antiparallel β-Sheets.

2.6. The contributions of the amino acid side chains

In the region of the Amide I band there is also a contribution to the absorption from the amino acid side chains, in particular from the ones containing a C=O group (Asp, Glu, Asn, Gln). This contribution amounts to about 10-15% of the total absorption in this spectral region. The influence of the respective protein environment (solvation, pH, dielectric constant) on the absorption properties of the side chains makes it impossible to predict the exact position of the absorption bands on the base of model spectra of amino acids in aqueous solution. Figure 2.6 gives an overview over the different amino acid side chains.

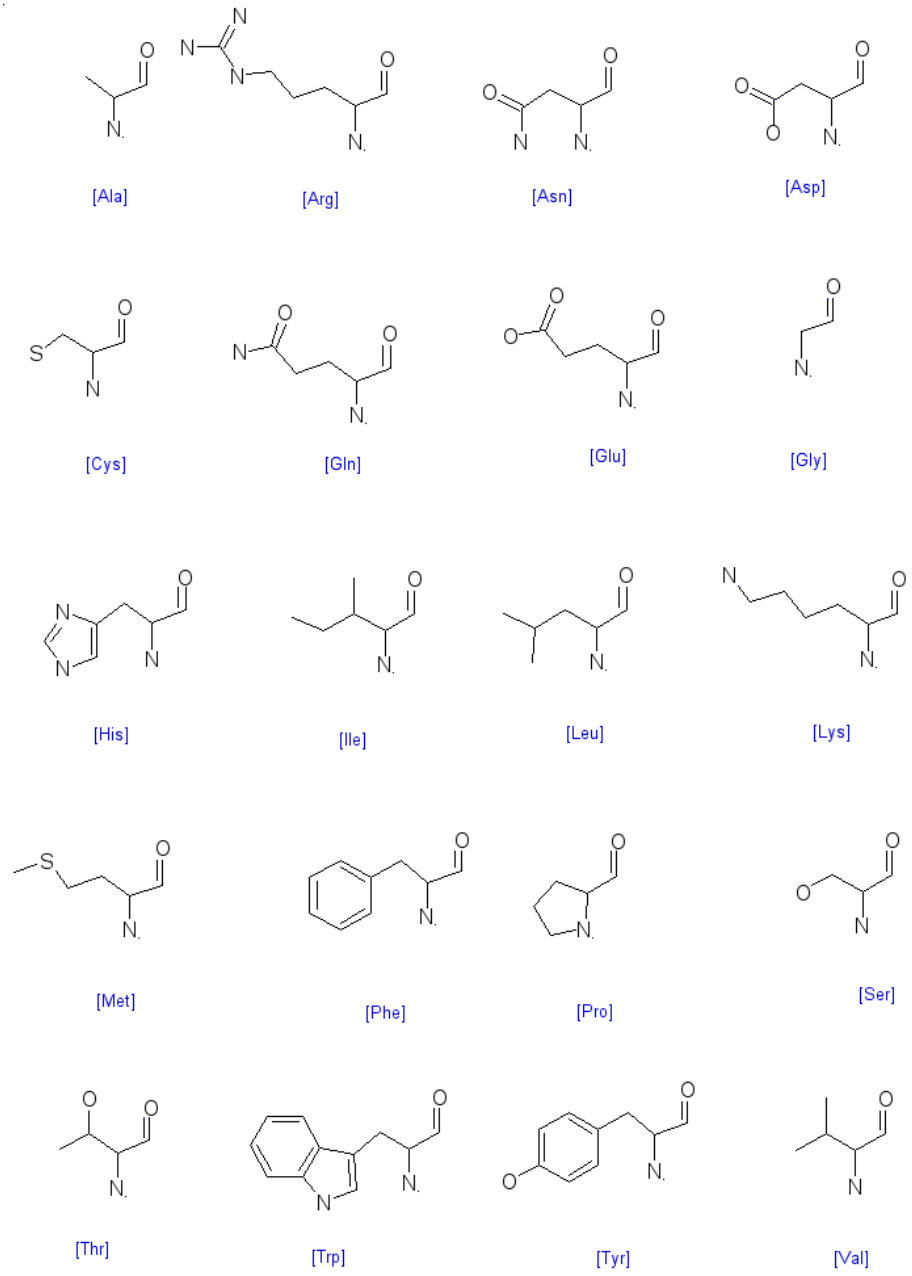

Figure 2.6.: Schematic diagram of amino acids.

3. The Fourier transform spectrometer

The modern devices to measure the infrared transmisison and the Raman signal are mostly Fourier-Transform spectrometer (FT–Spectrometer). At the heart of such an instrument there is a Michelson-Interferometer (see Figure. 3.1) instead of a monochromator (dispersive spectrometer, as, for example, spectrometers for the UV and visible spectral regions). In the former, the (polychromatic) radiation, which is to be measured, is splitted into two beams by a partially reflecting mirror. The two beams are then made to interfere.

In comparison with dispersive infrared spectroscopy, FT-infrared spectroscopy has the following advantages:

- A substantial amount of **time is saved**, since **all the frequencies are measured simultaneously**. A complete spectrum can be acquired much faster.
- Since in FT-spectroscopy the light is not attenuated by passing through the monocromator slit, the light intensity is higher, resulting in a **better signal-to-noise** ratio. This fact is also called "Fellgett's advantage" (see e.g. W. D. Perkins, *J. Chem. Ed.*, **64**, A269, 1987).
- Since the position of the mirrors is controlled by a parallel HeNe-Laser a **the wavelength can be determined with a high accuracy**.

3.1. The interferogram

With a FT-Spectrometer the spectrum of the sample is not measured directly. Instead, first a so-called interferogram is obtained, that then is converted to a spectrum by Fourier transform. The interferogram is generated by a Michelson interferometer (see Figure 3.1). The interferogram is the measured intensity in function of the position of the mirror S_2.

The light emanating from a source is splitted by a semi-transparent plate P_1 into two components (the reflecting side S is on the back), producing the light beams I and II. The transmitted light (I) goes through a second transparent plate P_2 of identical thickness and is reflected back onto itself from the precisely arranged perpendicular mirror S_1. Then part of the reflected light is directed to the detector from the reflecting side of the semi-transparent mirror S. The second beam (II) will be reflected upwards from S. Then it is again reflected on itself by the mirror S_2 and passes again through the plate P_1. At the detector, it interferes with the beam I. The path difference between the two beams can be varied by moving the mirror S_2 along the beam axis. When measuring, the sample is placed in front of the detector.

The zero position (*zero path difference*) of a Michelson interferometer, i.e. the arrangement for which the two beam paths are exactly the same, can be best checked with white light: for white light the coherence length is very short, therefore the zero path difference

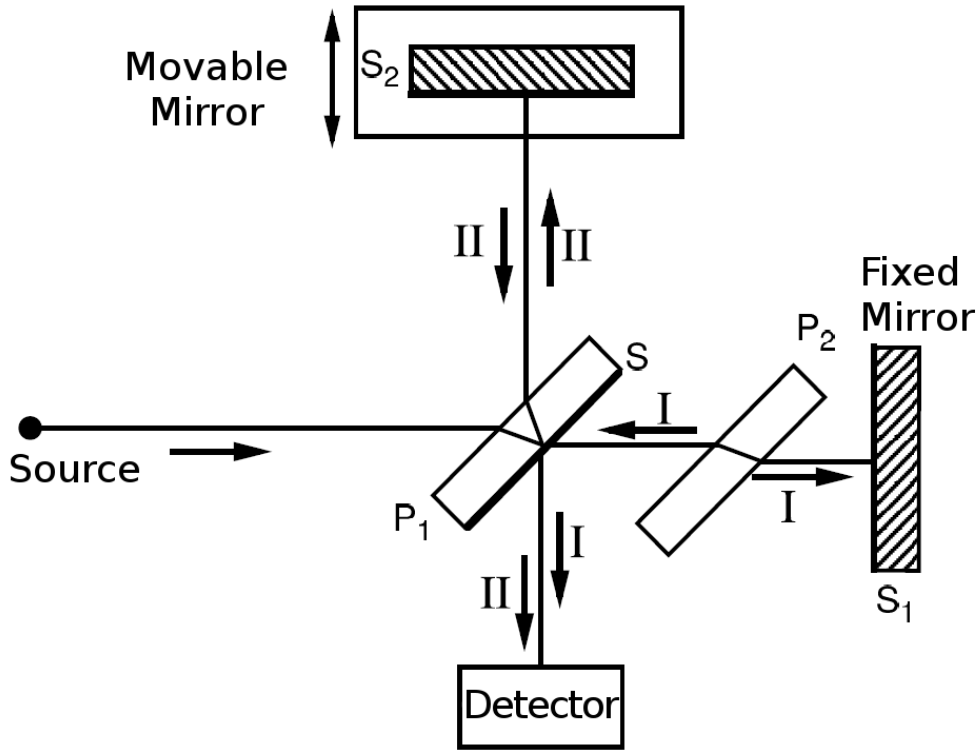

Figure 3.1.: Michelson Interferometer

can be exactly determined. The compensation plate P_2 is needed in beam path I of the interferometer in order to have exactly equal beam paths. Otherwise beam I, being reflected from the external part of the semi-transparent mirror, and beam II, reflected from the internal part, would have a phase difference of π. With the compensation plates both light beams go through the glass plate three times, experiencing also the same dispersion.

Using monochromatic light, a cosine function is obtained as interferogram. This can be exploited for the control of the mirror position, by coupling a HeNe-Laser (632 nm) to the interferometer. The infrared interferogram gets digitized at the zero-crossings of the laser interferogram. Given the wavelength of the laser, its interferogram shows zero crossings at intervals of $1/31600\,\mathrm{cm} = \lambda_{HeNe}/2$, which results in a wavenumber calibration of the spectrometer with an accuracy in the order of $0.01\,\mathrm{cm}^{-1}$.

The interferogram of a polychromatic light source is charachterized by a point of maximal interference (white light position, zero path difference), at which all the frequencies interfere constructively, because the distances from the beam splitting plate P_1 to the mobile and the fixed mirror are equal. At all the other points the condition for constructive interference is met only for a few wavelengths of the spectrum, while the others superimpose destructively (low coherence of white light).

3.2. The Fourier transform

The intensity at the detector $I(x)$ as a function of the mirror position x (the interferogram) and the intensity $S(\tilde{\nu})$ as function of the wavenumber $\tilde{\nu} = \lambda^{-1}$ of the light (the so-called

single channel spectrum) constitute a Fourier-Transform pair:

$$S(\tilde{\nu}) = \int_{-\infty}^{\infty} I(x) \exp(2\pi i \tilde{\nu} x)\, dx \tag{3.1}$$

$$I(x) = \int_{-\infty}^{\infty} S(\tilde{\nu}) \exp(-2\pi i \tilde{\nu} x)\, d\tilde{\nu} \tag{3.2}$$

Of course, in practice, a continuous and infinitely extended interferogram cannot be measured. Instead, a finite number$N$ of interferogram points with distance Δx are sampled. substituting x with $N\Delta x$, $\tilde{\nu}$ with $k\Delta\tilde{\nu}$ and the integral with a finite summation, a discrete Fourier transform is obtained:

$$S(k\Delta\tilde{\nu}) = \sum_{n=0}^{N-1} I(n\Delta x) \exp\left(\frac{2\pi i k \Delta\tilde{\nu} n \Delta x}{N}\right) \Delta x. \tag{3.3}$$

In the frequency domain the distance between the calculated points is

$$\Delta\tilde{\nu} = \frac{1}{N\Delta x}. \tag{3.4}$$

The distance $\Delta\tilde{\nu}$, defining the spectral resolution, is therefore inversely proportional to the mirror moving span and the number of sampled points. The discrete Fourier Transform introduces some problems: their corrections will be discussed in the next sections.

3.3. Apodization

Because of the finite length of the measured interferogram, artifacts can arise from the Fourier transform. The interferogram with a $x = x_{max}$ cut-off can be interpreted as the product of an infinite interferogram with a rectangular function, defined as having the value 1 for $0 \leq x \leq x_{max}$ and zero elsewhere. As a consequence secondary minima and maxima are introduced near the proper band . The amplitude of the largest secondary minimum is 22% of the amplitude of the real maximum.

This disturbance can be dealt with, by multiplying the interferogramm with a function that approaches more smoothly to the zero than the rectangular function does. These kinds of functions are called *apodization functions*. Since the apodization causes a loss of information at the edges of the interferogram, there is a deterioration of the resolution and line broadening. The apodization functions are therefore optimized to result in only a little broadening of the principal maximum while effectively suppressing side bands.

3.4. Zerofilling

Because of the finite distance of the digitized points of the interferogram and of the calculated spectrum it can happen that narrow bands, lying between the sampling points, are reproduced with smaller amplitudes. This error can be compensated for by the following trick: before the execution of the Fourier Transform, zeros are attached to the interferogram (*zerofilling*). This is equivalent to an approximation of a higher spectral resolution interferogram (i.e. simulates a longer mirror moving span): the distance between the spectral points becomes smaller, but without actually improving the resolution. In this way, the signal between the data points is interpolated.

3.5. Resolution improvement

The Amide I band of proteins consists mostly of different bands, which are characteristic for the different secondary structures. These components overlap in a large peak. Therefore some methods are applied in order to resolve the individual components of the spectrum, the spectrum is *deconvoluted*. None of these methods actually augments the spectral resolution, but they are able to emphasize the fine details of the spectrum. Three methods are presented here and the effect on the absorption bands is discussed.

3.5.1. Higher derivatives

Narrower bands are obtained, for which the 2nd (or 4th) derivative of the signal with respect to the wavenumber has a negative (or positive) sign.

Advantage: Can be performed with practically every software.

Disadvantage: Produces prominent side bands with opposite sign, making the interpretation of complicated spectra more difficult.

3.5.2. Fourier self-deconvolution

This method exploits a characteristic of the Fourier transform or the interferogram, respectively. The information on the fine structure of the spectrum is contained in the parts of the interferogram at larger mirror displacements. If this region of the interferogram is amplified, as for instance by multiplying the interferogram with an exponential function, the fine details of the spectrum are highlighted and the absorption bands become more evident.

Advantage: The bands in the adapted spectrum have the same sign as in the original spectrum.

Disadvantage: Dedicated software is needed, in which two parameters are to be choosen freely. This introduces some subjectivity in the result. It is possible to over-deconvolute and produce artifacts.

3.5.3. Strengthening of the fine structure

By subtracting from the original spectrum a smoothed version of itself, the coarse structure is lost, while the fine structure survives, making the remaining bands sharper. To avoid the production of negative side bands, the smoothed spectrum is multiplied with a factor ≤ 1 (around 0.9).

Advantages: The bands in the adapted spectrum have the same sign as in the original spectrum. Nearly every software can be used. The obtained "resolution" is comparable to the one obtained from Fourier self-deconvolution.

Disadvantage: It is possible to produce artifacts.

3.6. Carrying out an analysis of the secondary structure

The FTIR spectrum is treated with the methods explained above. A spectrum with enhanced fine details is obtained where the positions of the individual components of the spectrum are discernible. The bands at these positions are fittet and the amount of a specific type of secondary structure is calculated from the area of the fitted bands. Despite the problems mentioned below, the results from FTIR match the results form X-ray diffraction analysis quite well. Maximally, the deviations are 15%. Note, that, depending on the definition of the criterions, X-ray diffraction results may vary up to 20%. Infrared spectroscopy is especially useful to detect β-Sheet structures. Possibly occuring problems are

- The position of the bands does not allow to draw an unambiguous conclusion about the secondary structure. Particulary if there is a large amount of α-Helices, this may lead to errors. The α-Helix does not only produce a band at $1654\,\mathrm{cm}^{-1}$ but it also produces side bands at lower wavenumbers. This is also explicable theoretically. During the analysis of the secondary structure, these side chains are attributed to β-Sheet structures.

- Often, the extinction coefficients of the different secondary structure elements are assumed to be the same. This is in particular wrong for the two bands of the β-Sheet structure. The high-frequency band at $1684\,\mathrm{cm}^{-1}$ is much weaker than the one at $1633\,\mathrm{cm}^{-1}$.

- The subtraction of the water band (H_2O) in the Amide I region is not trivial and influences the relative amounts of the area of the fitted curves. Therefore, the experiments are often performed with D_2O which does not absorb that strongly in the region of the Amide I band.

- In the region of the Amide I band the absorption by side chains may may amount to $15\ldots20\%$ of the total absorption. A subtraction of this contribution is problematic because the absorption by the side chains depends on their environment in the protein.

- If the spectrum which is reconstructed from the fit is treated with the same methods as the original spectrum, the same result should be obtained. This is often not checked.

3.7. Secondary structure of typical proteins

Table 3.1 gives an overview over the secondary structure ratios in different proteins, as calculated by the X-ray structure.

Protein	**α-Helices**	**β-Sheets**	**Loops**	**Other**
Myoglobin	77 . . . 88	0	7 . . . 12	7 . . . 11
Phosphoglycerate kinase	30	23	20	27
Papain	26 . . . 29	22 . . . 29	18 . . . 23	25 . . . 30
Alcohol dehydrogenase	25 . . . 29	36 . . . 40	19	12 . . . 20
Trypsin Inhibitor	21 . . . 26	26 . . . 45	16 . . . 24	13 . . . 29
Elastase	10	40 . . . 47	26 . . . 28	16 . . . 24
Chymotrypsinogen	12	49	23	16
9–Chrymotrypsin	8 . . . 11	35 . . . 50	25 . . . 27	14 . . . 30
Trypsinogen	9	56		
Concanavalin	2 . . . 3	44 . . . 65	22 . . . 25	10 . . . 28
Bovine Serum Albumin (BSA)	50 . . . 68	16 . . . 18		
Immunoglobulin G	7.4	54	k. A.	k. A.

Table 3.1.: Secondary structure ratios in individual proteins. The data for immunoglobulin G was taken from www.rcsb.org, calculated for the entry *1HZH*. The reference which is cited there for the calculation of the secondary structure ratios is W. Kabsch and C. Sander, *Biopolymers*, 22(12), 2577, 1983.
Note: there are considerable uncertanties due to the different definitions of secondary structures by different techniques.

4. Experimental procedure

For the experiment, three samples of Poly-L-Lysine, γ-globulin, Papain, and Myoglobin in D2O will be prepared in the beginning of the experiment. The H-D exchange in the protein solution will take about one hour.

γ-globulin (molecular weight about 150 kDa, 1344 amino acids) from bovine blood consists of immunogobulins (antibodies) of the subclass immunoglobulin G (IgG). IgG has a Y-shaped structure (Fig. 4.1). The upper arms serve to bind to antigens whereas the lower part binds to body's own cells. Further information may be found in the PhD thesis of S. Da Vela, on *www.rcsb.org* (entry *1HZH*) and also e.g. in R. Rappuoli et al., *Antibodies for Infectious Diseases*, Washington DC: ASM Press, 2015.

Papain (molecular weigth about 23 kDa, more than 230 amino acids) Is a protease (protein hydrolyzing enzyme) obtained mainly from the chyle of the Papaya plant (Carica papaya). Is used as additive for cleaning products, as softener for meat, to reduce turbidity in beer, to heal wounds. Is also employed as dietary supplement to enhance protein digestion.

Myoglobin (molecular weigth about 18 kDa, 153 amino acids) extracted for example from horse cardiac muscle, is an Haem (Heme) containing protein, with high affinity for oxygen. Has an oxygen storing function in the muscle tissue of most mammals and is responsible of its red colour.

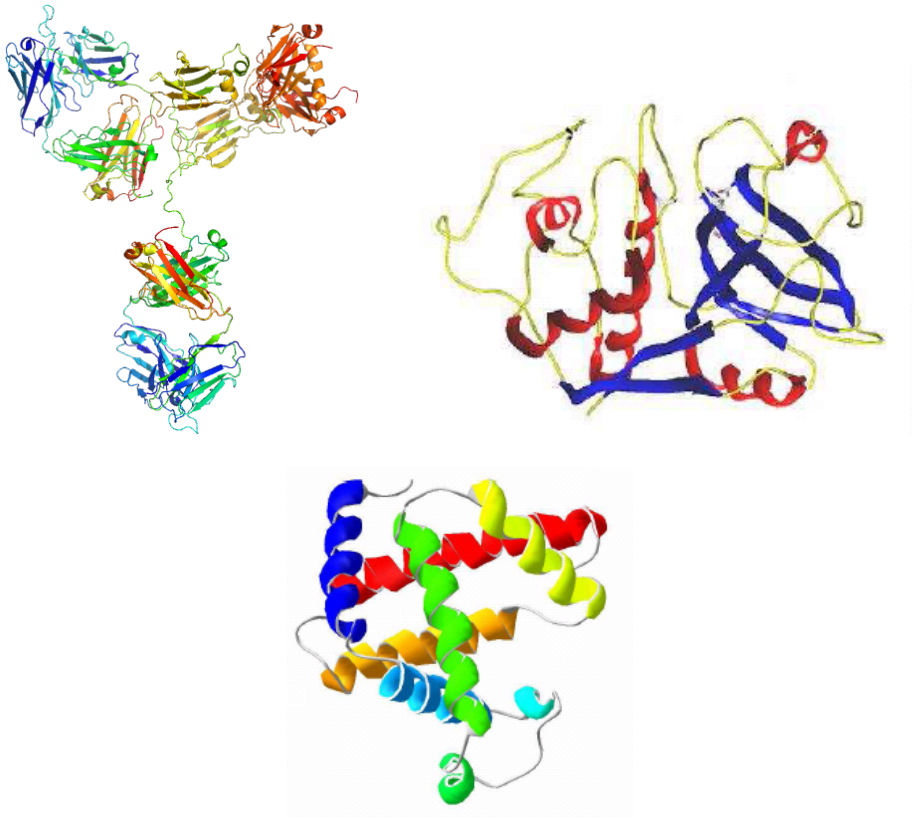

Figure 4.1.: Upper left: structure of immunoglobulin G (entry *1HZH* on www.rcsb.org), upper right: papain and bottom: myoglobin.

4.1. Preparation of the samples

For each protein about 6 mg of the protein powder are dissolved in 200 μl D_2O. For measurements with a 20 mm Transmission cell (see Figure. 4.2) and a sample thickness of about 25 μm about 30 μL of sample solution will be needed.

4.2. Operating the FTIR

The FTIR spectrometer is controlled with the OPUS software. After starting the software, set the measurement parameters in the menu *Measure/Advanced Measurement*. Choose 1000 cm^{-1} to 4000 cm^{-1} as measurement range, and 2 cm^{-1} for the wavelength resolution. Is advisable then to set the folder for the files as the local path `D:\fpraktik` . Data are saved initially in binary form, for further processing they can be saved as a table in ASCII format using *File/Save as*.

A background measurement is always performed first, and then the sample measurement.

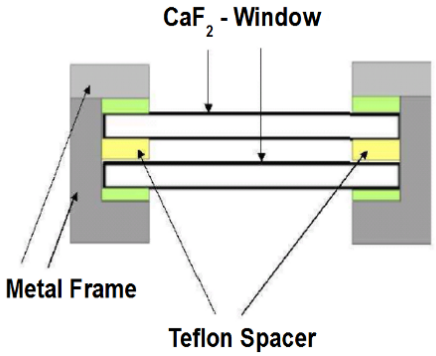

Figure 4.2.: Scheme of the transmission cell.

4.3. Preliminary measurements

- Record a spectrum with empty transmission cell and open sample chamber (so-called *empty spectrum*). Discuss the observed absorption bands with your supervisor.
- Close the sample chamber and open the dry air flow. Repeat the previous measurement after 2 and 5 minutes waiting time. Determine the necessary waiting time for spectra with the noise as low as possible.
- Record an absorption spectrum of H_2O and D_2O. Furthermore, record a background spectrum of D_2O.

4.4. The Amid-I band of proteins

Record an absorption spectrum in D_2O of the following proteins with a wavelength resolution of $2\,\mathrm{cm}^{-1}$:

- γ-globulin representing proteins with a large amount of β-sheet structures.
- myoglobin representing proteins with a large amount of helical structures.
- papain representing proteins with mixed secondary structure elements.

For all transmission measurements use the liquid cell with CaF_2–windows and variable spacers. Before each sample change clean the whole liquid cell.

5. Data Analysis

In the following, the data analysis of the FTIR measurements is described (‡ denotes optional subtasks). In order to analyze the data you need a software that is able to display and process the spectra. One option is the powerful and freely available program GNUPLOT (see appendix A.4 for more information). Of course, every other program that can handle ASCII tables as source data can be used, too. The complete analysis has to be handed in before the deadline.

Figure 5.1 shows measurements of papain, poly-L-lysin, myoglobin and the solvents H_2O and D_2O as examples.

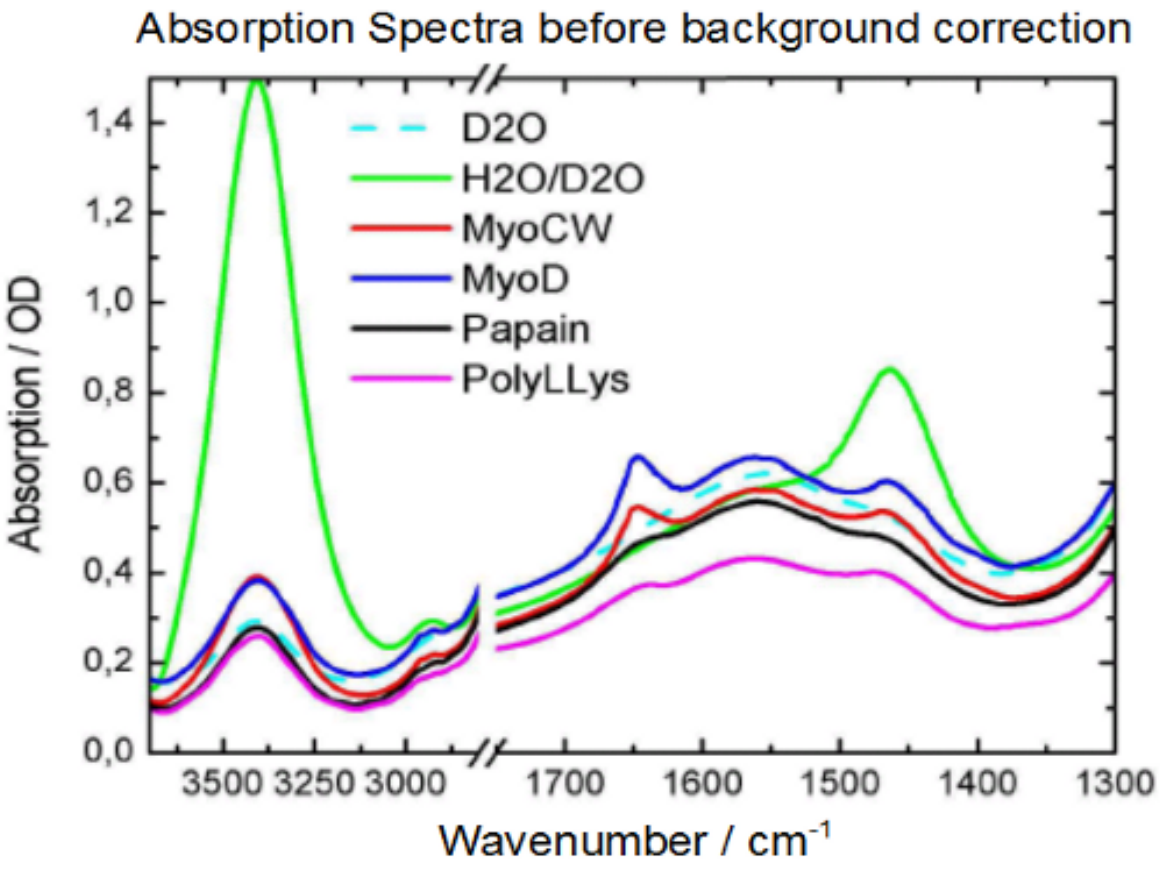

Figure 5.1.: Example spectra.

5.1. Preliminary measurements

- From the structures in the empty spectrum of the transmission cell determine the thickness of the sample compartment.
- Discuss the spectra (sample chamber open or closed with dry air flow).
- Determine the waiting time for absorption spectra with a good signal-to-noise ratio.
- Record the absorption spectra of H_2O and D_2O.

5.2. The Amid-I band of proteins

- Discuss a possible H_2O–content of the *gamma*-globulin, papain and myoglobin samples with the aid of the obtained spectra.
- Localize the typical protein absorption bands between $1400\,\mathrm{cm}^{-1}$ and $1700\,\mathrm{cm}^{-1}$ as well as the amid–A band.
- Why are the other wavelength ranges not suitable for the analysis?
- Analyze the absorption spectra of the three proteins in the region of the amid–I band. First, perform a baseline correction by subtracting a linear background from the absorption spectrum between $1600\,\mathrm{cm}^{-1}$ and $1700\,\mathrm{cm}^{-1}$. This approximately corresponds to a subtraction of the side-chains absorption.
- Identify the specific absorption bands in the mentioned spectral range according to table 5.1.
- Fit the absorption spectra using a multiple Gaussian fit (sum of five Gaussian functions). The starting values for central wavelengths and half widths (FWHM) should be chosen according to table 5.1 and kept fixed in the first fit. In a second step, leave these parameters free in order to improve the quality of the fit. Integrate the obtained single Gaussian functions for a quantitative analysis of the secondary structure. Determine the fraction of secondary structure components by comparing the obtained areas. This is done on the assumption that the extinction by the different secondary structures is approximately constant in the region of the main band. The Gaussian function which corresponds to the high frequency lateral bands of the β–sheets between $1682\,\mathrm{cm}^{-1}$ and $1689\,\mathrm{cm}^{-1}$ is neglected for the calculation of the secondary structure fractions. Otherwise the β–sheet contribution would be counted twice. Thus, only four out of the five fitted Gaussians are used to determine the specific fractions. Please note that the integral of a Gaussian can be calculated analytically.

  ‡ Integrate the absorption spectrum by summing the data values within the boundaries given in table 5.1 and determine the fraction of the respective secondary structure components using the obtained areas. Ignore the region between $1682\,\mathrm{cm}^{-1}$ and $1689\,\mathrm{cm}^{-1}$.

- Discuss your results in comparison with the values obtained by X-ray structure analysis (see table 5.1). Furthermore, discuss problems and possible sources of error of secondary structure analysis by infrared spectroscopy.

Band parameter / cm^{-1}			**secondary structure**
Range	Maximum	FWHM	
1613 . . . 1637	1632	20	β–sheets
1637 . . . 1648	1640	20	not ordered
1648 . . . 1662	1655	20	α–helix
1662 . . . 1682	1675	20	slopes
1682 . . . 1689	1685	20	β–sheets

Table 5.1.: Typical wavelength ranges of different secondary structure components.

A. Appendix

A.1. Setup and operating mode of the IFS 48

The FTIR spectrometer consists mainly of a radiation source, the interferometer unit and the detector. The IFS 48 of the company Bruker is a FTIR spectrometer with a Michelson interferometer. Figure A.1 shows the principal setup.

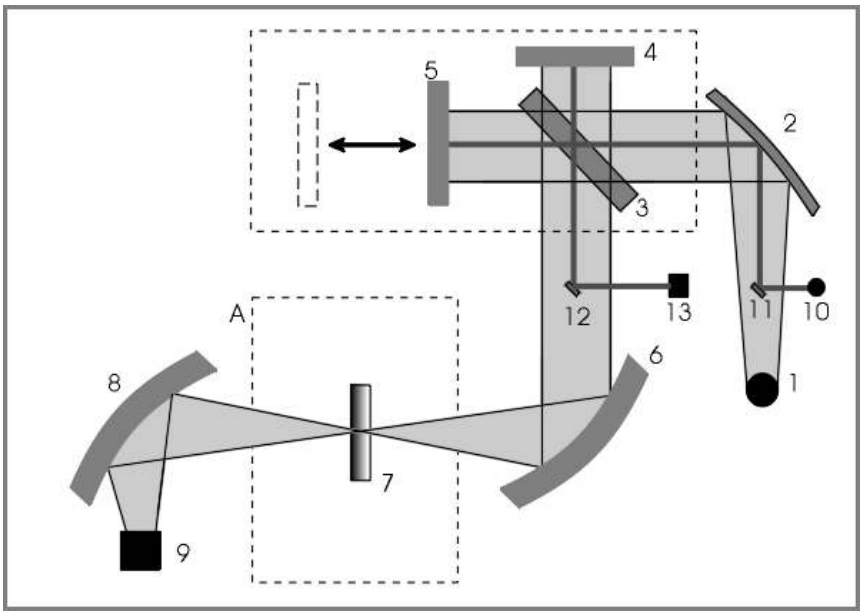

Figure A.1.: Optical path in the IFS48

The light emitted by the radiation source (1) is focused by the parabolic reflector (2) in the direction of the beam splitter (3, semi-transparent mirror). Half of the light in is reflected from (3) in the direction of the fixed mirror (4) or transmitted in the direction of the movable mirror (5) respectively. Both parts of the beam reflected from (4) or (5) are merged in the direction of the parabolic reflector (6) by (3) and interfere. (6) focuses the beam on the sample (7) in the sample chamber (A). Eventually, the elliptic mirror (8) maps the beam on the detector. In addition, a HeNe-laser (10) is directed to the parabolic reflector (2) by the mirror (11). The laser beam passes the Michelson interferometer and is directed on a separate detector (13) by the mirror (12). Interferences of the monochromatic laser are used to determine the position of the movable mirror (5). The wave number of the laser light is $15803\,\mathrm{cm}^{-1}$. This corresponds to a wavelength of 632.8 nm. The laser light is thus red.

In FTIR spectroscopy, thermal radiation sources are used which radiate a broad spectrum and are similar to a black body. The IFS 48 operates in the mean infrared and uses a

Globar. Globars are bar-shaped and consist of silicon carbide. They reach their operating temperature of 1500 K through voltage application.

The detector in the IFS 48 is a DTGS-detector (deuterated triglycinsulfate). DTGS-detectors belong to the pyroelectric detectors which consist of ferroelectric materials that polarize electrically below the Curie temperature. The infrared radiation changes the temperature. As a consequence the polarization and hence the voltage changes. Pyroelectric detectors are faster than purely thermal detectors.

Optical system	Michelson interferometer with an air-cushioned arm and a HeNe-Laser for position determination
Radiation source	Globar
Aperture	4.7 mm
Beam splitter	Ge on CsI
Detector	DTGS with CsI-window
Interferogram points per second	1×104 or 4×104
Wavenumber range	$4400 \ldots 235\,\mathrm{cm}^{-1}$
Max. resolution	$0.5\,\mathrm{cm}^{-1}$
Wavenumber precision	$0.01\,\mathrm{cm}^{-1}$
Ordinate precision	0.1% T
Beam diameter at sample position	8 mm

Table A.1.: Specifications of the IFS 48

A.2. Absorbance

The absorbance is derived from the Lambert–Beer law, which describes the absorption of radiation in a homogeneous medium:

$$I = I_0\, 10^{-\epsilon c d}, \tag{A.1}$$

I_0 is the the intensity of the incoming beam, I the intensity of the beam after passing the homogeneous medium with the extinction coefficient ϵ, the concentration c and the thickness d.

The exponent is combined to the dimensionless quantity of the absorbance (A, or optical density, OD).

A.3. Preparation of the samples

Different proteins are stored in the freezer ($-20\,^{\circ}\mathrm{C}$) in the chemistry lab. Table A.2 shows the proteins and the recommended concentrations. In order to avoid contamination of the original protein containers, small amounts have been filled into glass vials. Please use only these glass vials for the preparation of your samples and keep the original protein containers always in the freezer.

protein	concentration / μg/μl
γ-globulin	30
myoglobin	30
papain	20

Table A.2.: Concentration of the used proteins in solution.

A volume of approximately 100 μl protein solution is needed per measurement. Work carefully and precisely - the accuracy limit of the balance is about 0.1 mg. Use the balance in the chemistry lab. Place a small glass vial on the balance and press the tare button. Then fill the needed amount into the glass vial using a suitable spatula.

If the weighing is not successful, the sample volume can be doubled. Add the needed amount of D_2O using a pipette. Close and label the vial.

A.4. Gnuplot

Gnuplot is an interactive plotting program which is freely available for all common operating systems. With *Gnuplot*, ASCII data can be plotted in 2D or 3D visualization. Output formats are numerous, for example X11–terminal, ASCII–terminal, PS-, PDF-, PNG-file. Apart from the interactive control, *Gnuplot* can be also used with scripts. Table A.3 contains a list of recommended documentation. In addition, a help function is available in the *Gnuplot* prompt.

Homepage	http://www.gnuplot.info/
Download	http://www.gnuplot.info/download.html
Documentation in general	http://www.gnuplot.info/documentation.html
Detailed manual (PDF)	http://www.gnuplot.info/docs/gnuplot.pdf
Online manual (HTML)	http://www.gnuplot.info/docs/gnuplot.html
Very good short description	http://www.duke.edu/~hpgavin/gnuplot.html

Table A.3.: Download and documentation of *Gnuplot*

In the following, some plotting possibilities of *Gnuplot* are shown using an example spectrum (FTIR-Spectrum of KBr) are shown. A short excerpt of the file kbr.dat:

```
#
# KBr test measurement FTIR
# 10.03.2005
#
7500.54795    1.63501
7498.61930    1.57173
7496.64069    1.57159
7494.76198    1.56722
...
```

These data can be visualized by gnuplot with the command **plot**:

```
gnuplot> plot 'kbr.dat' using 1:2 title 'KBr test measurement FTIR'
```

This creates a graph in a default window. For further design of the plot, the following commands can be used:

```
title:                                  > set title 'KBr test measurement FTIR'
x-label:                                > set xlabel 'wavelength'
y-label:                                > set ylabel 'absorption'
x-axis range                            > set xrange [0.001:0.005]
y-axis range                            > set yrange [20:500]
autoscale                               > set autoscale
labeling                                > set label 'label' at 0.003, 260
deleting all labels                     > unset label
logarithmic scaling                     > set logscale
logarithmic scaling y-axis              > unset logscale; set logscale y
tic positions x-axis                    > set tics xtics (0.002,0.004,0.006)
automatic tics x-axis                   > unset tics; set xtics auto
```

The data can be plotted for example by the following script:

```
# Gnuplot script for plotting data file 'kbr.dat'
# This file is called kbr.gpl
set autoscale
unset log
unset label
set xtics auto
set ytics auto
set title 'KBr test measurement 10.03.2005 FTIR'
set xlabel '[cm-1]'
set ylabel 'absorption'
set key 0.01,100
set label 'Yield Point' at 0.003,260
```

Saving the plot into a postscript file can be accomplished by the following commands:

```
gnuplot> set out 'plotfile.ps'
gnuplot> set term postscript enhanced color
gnuplot> replot
```

The spectrum can be fitted by a before defined function:

```
gnuplot> l(x) = a*x+b
gnuplot> a = 1; b = 1
gnuplot> fit f1(x) 'linear.dat' using 1:2 via a, b
gnuplot> plot 'linear.dat' using 1:2 w points, l(x) w lines
```

The data can be smoothed by smoothing functions that have to be given after the plot-command:

```
gnuplot> plot 'kbr.dat' using 1:2:(0.00001) smooth acsplines
```

The value in brackets gives the weight of the points: The smaller the value, the more points are taken into account for the calculation of the splines. The smoothed data again can be saved in an ASCII-file by:

```
gnuplot> set out 'kbr_smooth.dat'
gnuplot> set term table
gnuplot> plot 'kbr.dat' using 1:2:(0.00001) smooth acsplines
```