Spectral Techniques for Rapid Quantification of Protein Structure in Solution

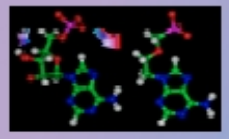
Vladimír Baumruk

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Physics

Oddělení fyziky biomolekul

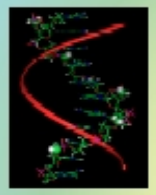
VÝZKUM FYZIKÁLNĚ–CHEMICKÝCH VLASTNOSTÍ, STRUKTURY A INTERAKCÍ KLÍČOVÝCH BIOMOLEKUL A JEJICH KOMPLEXŮ

Příklady studovaných systémů



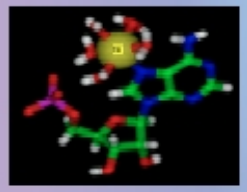
Analoga nukleotidů

Účinná virostatika podobná přirozeným nukleotidům, která blokují syntézu virových nukleových kyselin díky absenci vazebného místa pro prodlužování řetězce nukleové kyseliny.



„Antisense" oligonukleotidy

Potenciální účinná zbraň proti bakteriálním, virovým a nádorovým onemocněním. Váží se ke komplementárním úsekům nukleových kyselin a potlačují tak expresi nežádoucí genetické informace.

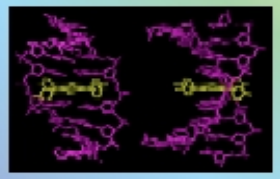


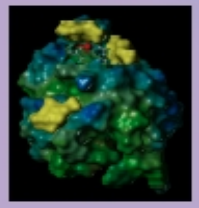
Interakce kovových iontů s nukleovými kyselinami

Uplatňují se při důležitých procesech, jako jsou rozpoznávání úseků DNA, jejich replikace, nebo přepis genetické informace do RNA. Rovněž způsobují změny jejich strukturních vlastností.

Porfyriny

Jsou součástí aktivních míst mnoha enzymů a aktivních proteinů (hemoglobin, chlorofyl). Jejich deriváty se mohou vázat i na nukleové kyseliny a zablokovat jejich funkci buď bezprostředně, nebo po světelném ozáření (tzv. fotodynamická terapie).





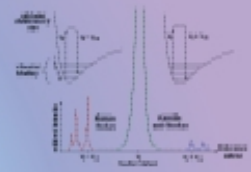
Proteiny

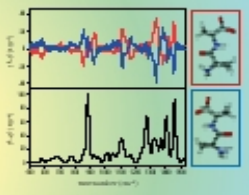
Klíčové biomolekuly odpovídající mimo jiné za enzymatickou aktivitu, transport molekul a opravy nebo replikaci DNA. Funkce proteinu úzce souvisí s jeho strukturou, a proto detailní znalost změn spojených například s vazbou malých molekul na protein vede k lepšímu pochopení jeho funkce.

Používané metody

Spektroskopie Ramanova rozptylu (RS)

Sledování změn vlnové délky u malé části záření po rozptylu na studované molekule. Tyto změny jsou způsobeny vibračním pohybem molekuly a citlivě odrážejí modifikace struktury i slabé mezimolekulové interakce. Vazbou na speciální kovové povrchy lze intenzitu RS obrovsky zvýšit, takže lze studovat i velmi malá množství molekul (metoda SERS).



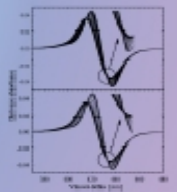


Ramanova optická aktivita (ROA)

Moderní spektroskopická metoda s vysokou citlivostí k prostorovému uspořádání (konformaci, absolutní konfiguraci) *chirálních* molekul. Je založena na měření velmi malého rozdílu v intenzitě Ramanova rozptylu excitovaného pravo- a levotočivě kruhově polarizovaným zářením.

Absorpční spektroskopie

Je založena na jevu absorpce optického záření při průchodu roztokem studované látky. Měří se v ultrafialové a viditelné (změny elektronových stavů molekul), nebo v infračervené (změny vibračních stavů) spektrální oblasti. Vedle kontroly koncentrací slouží především ke sledování konformace molekul, jejich protonačního stavu a tvorby komplexů.



Molekulárně-dynamické simulace (MDS)

Výpočetní metoda umožňující nalézt pravděpodobné strukturní uspořádání biomolekul a jejich komplexů a sledovat jejich stabilitu a dynamické chování. Lze věrohodně simulovat i vlivy okolí (např. pro molekulu ve vodném roztoku).

Nukleární magnetická rezonance (NMR)

Sledování projevů precese jaderných spinů ve vnějším stacionárním magnetickém poli po působení vysokofrekvenčních pulsů. Podává informace o chemické stavbě, konformaci, prostorovém uspořádání a dynamice studovaných molekul a jejich komplexů.

Protein structure

- primary structure
  - the linear amino acid sequence; arrangement of amino acids along a polypeptide chain
- secondary structure
  - different regions of the sequence form local regular structure like α-helices or β-strands
- tertiary structure
  - is formed by packing such structural elements into one or several compact globular units (domains)
- quaternary structure
  - several identical polypeptide chains that associate into a multimeric molecule

Why vibrational spectroscopy ?

Optical spectroscopy techniques

- Electronic Circular Dichroism (ECD, UVCD)
- Vibrational Spectroscopy
- Vibrational Optical Activity (VOA – VCD, ROA)

- ✓ **Structural information about protein can be obtained on a relatively rapid time scale**
- ✓ **The structural information is not restricted to a static picture.**
- ✓ **The size of a protein or the nature of the environment does not limit the application.**

Vibrational spectroscopy can be used for investigation of protein ***structure*** and ***function***.

Vibrational spectroscopy

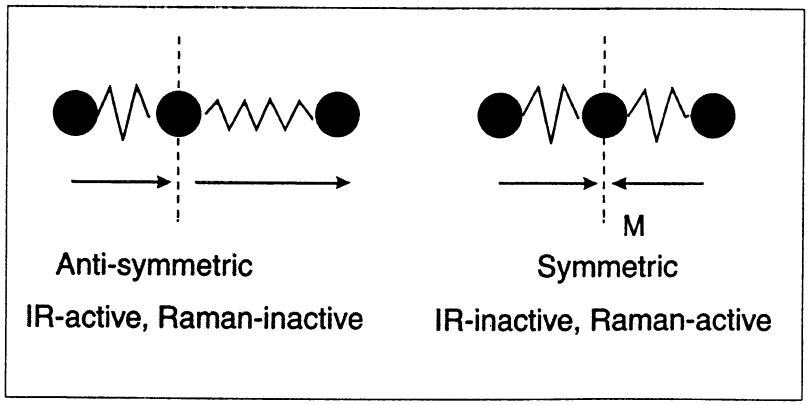
- infrared (IR) absorption
- Raman spectroscopy

CO_2 linear molecule (O=C=O) with center of symmetry

$n = 3 \Rightarrow 3n - 5 = 4$	$\Rightarrow$	2 **stretching** modes	$\Rightarrow$ symmetric	Raman active
			$\Rightarrow$ asymmetric	IR active
	$\Rightarrow$	2 **bending** modes (degenerate)		IR active

center of symmetry $\Rightarrow$ mutual exclusion

IR active vibrations are Raman inactive and vice versa



frequency of vibration

$$\nu = \frac{1}{2\pi}\sqrt{\frac{f}{\mu}} \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}$$

f **– force constant (bond strength)**

μ **– reduced mass**

Advantages

- Raman and IR spectroscopy are ***nondestructive techniques***. Ordinarily the sample may be recovered and assayed for biological activity after spectroscopic examination.
- Raman and IR methods are applicable to ***samples of*** virtually ***any morphological form***. For proteins and nucleic acids, this includes solution (aqueous and non-aqueous), suspensions, precipitates, gels, films, fibers, single crystals, and polycrystalline and amorphous solids. Data obtained from a given sample in one morphological state are generally transferable to another morphological state of the same sample. This has important practical benefits – for example, in comparing the molecular structure of a protein in the crystal with that prevailing in solution.
- ***A small sample volume*** is required for these methods (approximately ***10 μl*** is sufficient for conventional Raman spectroscopy and approximately ***20 μl*** for FT-IR spectroscopy). This represents an advantage over many other structural methods, including X-ray crystallography and NMR spectroscopy.
- Raman scattering and IR absorption processes occur on a time scale that is very short ($\approx 10^{-15}$ s) in comparison to the time scales of fluorescence ($> 10^{-9}$ s) and NMR phenomena ($\approx 10^{-6}$ s). Thus, vibrational spectroscopy is suitable for ***time-resolved studies*** of biological processes that are inaccessible by fluorescence and magnetic resonance methods.
- There exists ***a large database of IR and Raman spectra*** of proteins, nucleic acids, and their constituents (and other biomolecules as well) for which reliable band assignments, normal mode analyses, and spectra-structure correlations have been made, This facilitates interpretation of the often complex vibrational spectra obtained from proteins, nucleic acids and their complexes.

Disadvantages

- Although band resolution in vibrational spectra is superior to that of achievable in electronic spectra of condensed phases, the ***spectral resolution of vibrational spectroscopy is still inferior to that of high-field magnetic resonance spectroscopy***. Inadequate resolution can be partly overcome if chemical (isotope editing) or biological (site directed mutagenesis) modification are employed.
- Although small sample volumes are sufficient for Raman and IR analyses, relatively ***high solute concentrations*** are required (≈ 10-100 μg/μl).
- Both H_2O and D_2O are highly *problematic solvents for IR spectroscopy* (unlike Raman scattering, where H_2O and D_2O produce relatively little interference with Raman spectrum of the dissolved solute).
- Raman effect of inelastic scattering is inherently week compared to other light absorption or emission processes. Thus, considerable *sample purification and care in sample handling* is necessary.

Advantages specific to Raman spectroscopy

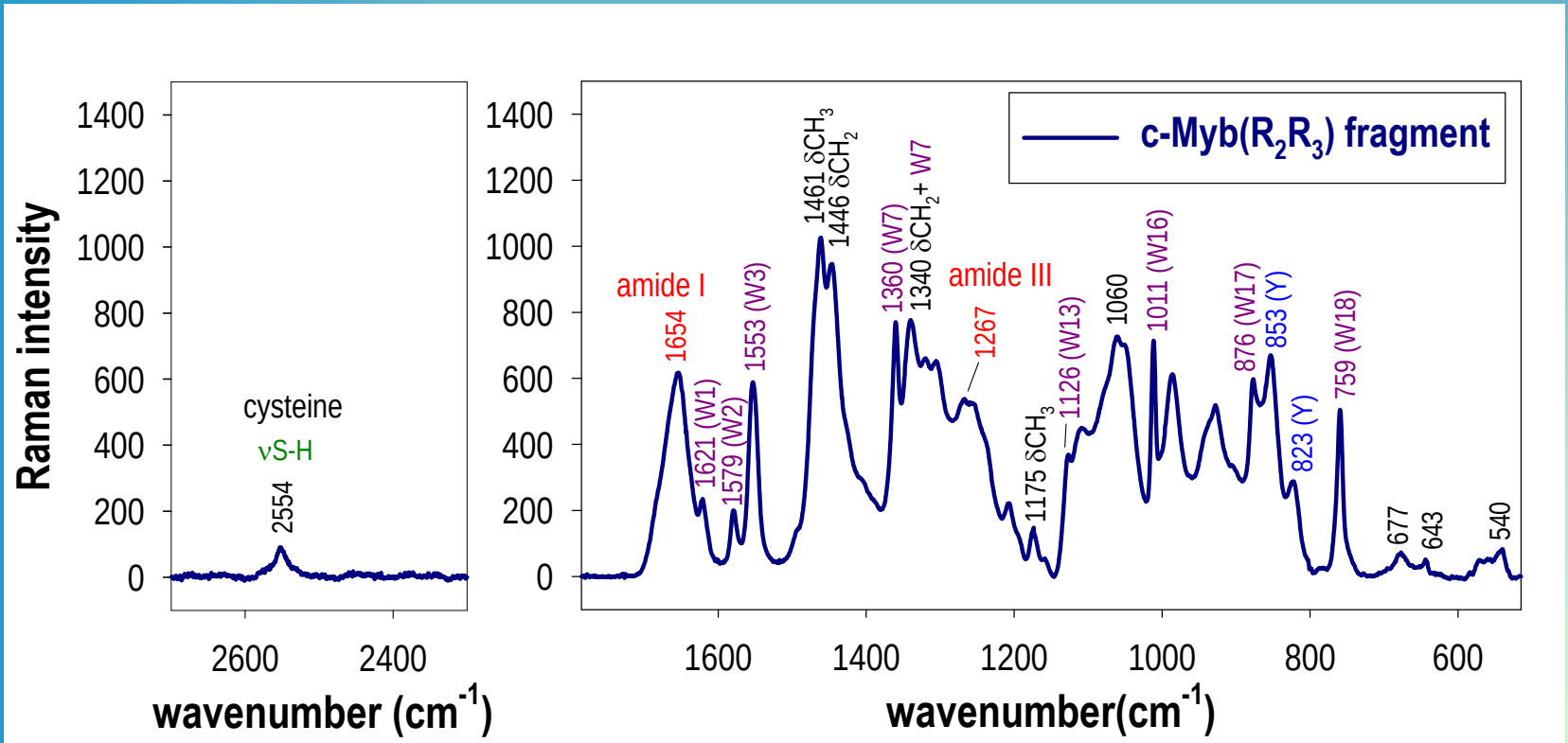
- Both H_2O and D_2O generate very week Raman signal, thus producing little interference with the spectrum of dissolved solute. This represents a significant advantage over IR spectroscopy, where both H_2O and D_2O are highly problematic solvents.

- Relatively high Raman intensities comes from molecular vibrations including large change in molecular polarizability. For proteins, Raman spectra are dominated by bands assignable to main chain peptide groups, aromatic side chains, sulfur containing side chains and side-chain carboxyls. Contribution from saturated hydrocarbon groups is intrinsically rather week, large number of such groups present in a protein result in a few relatively strong bands associated with group frequencies of methyl and methylene substituents. In case of nucleic acids, spectra are dominated by bands attributable to vibrations of purine and pyrimidine rings and backbone phosphate groups.

- Raman intensities are dramatically enhanced (by several orders of magnitude) when energy of the incident photon is in resonance with a molecular electronic transition of a chromophore. The structural information about the chromophore can be obtained from relatively dilute solution. Important is *selectivity of resonance enhnacement*.

Protein c-Myb(R_2R_3) fragment

(minimal sequence specific DNA-binding domain)



c-Myb(R_2R_3) fragment contains 6 tryptophans (3 in each repeat), 2 tyrosines (1 in each repeat), and 1 cystein (in R_2). Their signatures dominate the Raman spectrum.

Raman and resonance Raman spectroscopy

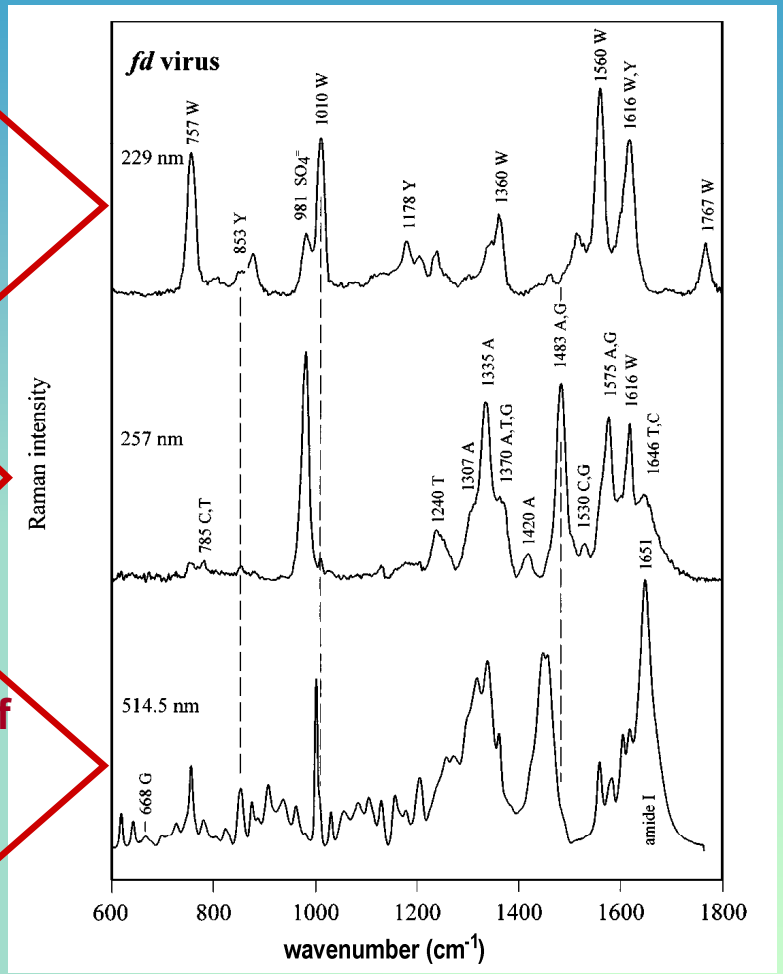
complementary nature of off-resonance (514.5 nm) and UVRR (257 and 229 nm) excitations

Selectivity !!!

The spectrum excited at 229 nm is the Raman signature of coat protein Trp and Tyr

The spectrum excited at 257 nm is the Raman signature primarily of packed ssDNA bases

Off-resonance excitation produces a rich pattern of Raman bands, primarily from viral coat subunits, which constitutes about 88% of the virion mass



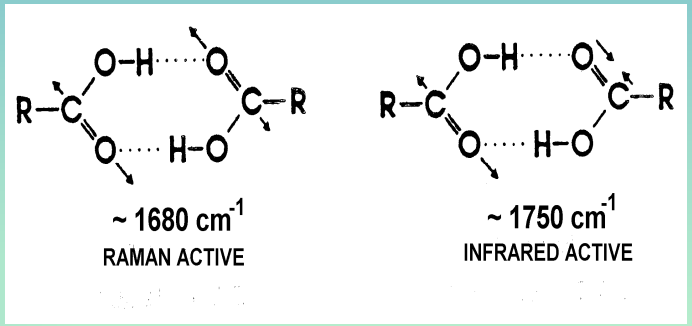
Raman spectra (600-1800 cm^{-1}) of fd virus excited at 514.5 nm (bottom, 50 μg/μL), 257 nm (middle, 0.5 μg/μL), and 229 nm (top, 0.5 μg/μL).

Vibrational coupling

Cyclic hydrogen bonding configurations are important determinants of biological structure and recognition. Examples are the Watson-Crick base pairing schemes and the interactions formed between side chains of gene regulatory proteins and their DNA recognition sites.

A simple model of cyclic hydrogen bonding is the carboxylic acid dimer. In Raman spectra of carboxylic acids, the C=O stretching mode is observed near **1680 cm^{-1}**. However, in infrared spectra no 1680 cm^{-1} band is observed. Instead an intense IR band occurs near **1750 cm^{-1}**.

- **1680 cm^{-1}** ***Raman-active mode is symmetric***, i.e. it corresponds to ***in-phase motion*** **of the two C=O oscillators**,
- ***1750 cm^{-1} IR-active mode is an antisymmetric or out-of-phase motion***,
- the two modes are shifted from a common center (1715 cm^{-1}) where uncoupled oscillators would absorb and Raman scatter.



This phenomenon, called ***intermolecular vibrational coupling***, is expected when ***two identical or nearly identical oscillators are in proximal contact and their association generates a symmetry or pseudosymmetry element*** **(typically center of inversion or rotational axis).**

In the case of proteins, the relatively large splitting observed in the IR amide I modes of antiparallel β-sheet structures (≈ 60 cm^{-1}) has been explained in terms of vibrational coupling.

Hydrogen bonding interactions

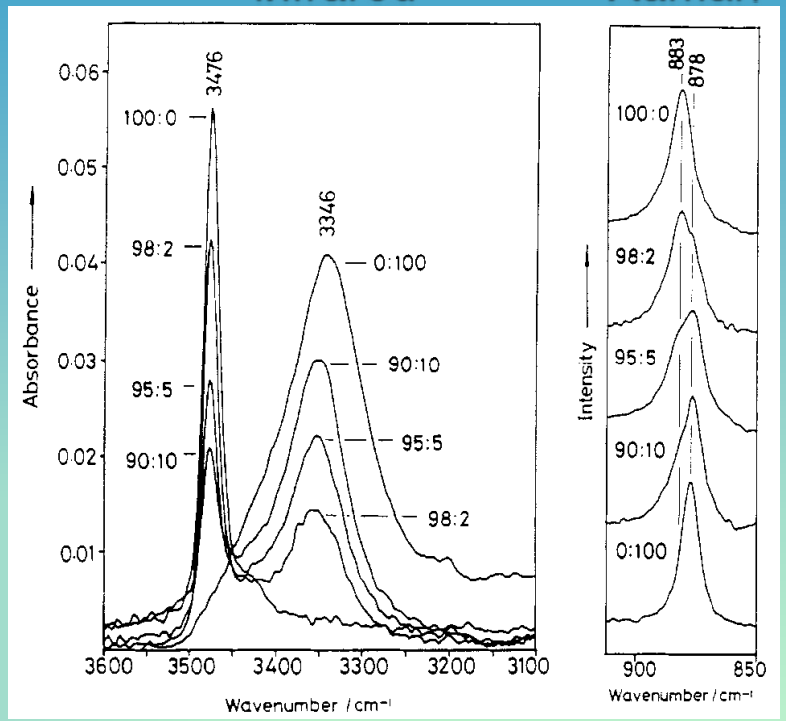
- Hydrogen bonding interaction between an appropriate donor hydrogen X-H (where X is O, N, or S) and an acceptor atom Y (usually O or N) is conveniently studied by methods of vibrational spectroscopy.
- ***The X-H stretching mode is particularly sensitive to hydrogen bonding.***

Hydrogen bonding interactions

IR and Raman spectra of skatol (3-indol acetic acid) in CS_2/dioxane mixtures

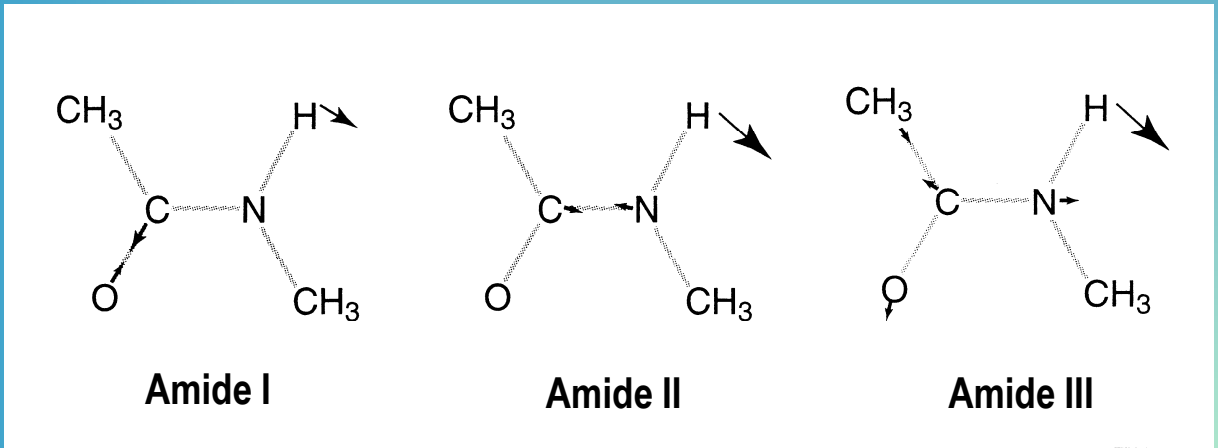
In CS_2, where ***hydrogen bond donation is disfavored***, a relatively narrow band is observed at 3476 cm^{-1}, corresponding to the imino ***N-H stretching*** vibration of ***nonhydrogen-bonded*** 3-methylindole.

As the mole fraction of dioxan is increased and ***hydrogen bonds*** of the type ***N-H···O are formed between imino N-H donors and dioxane endocyclic acceptors***, a characteristic IR band appears at 3346 cm^{-1} due to ***hydrogen-bonded imino groups***.



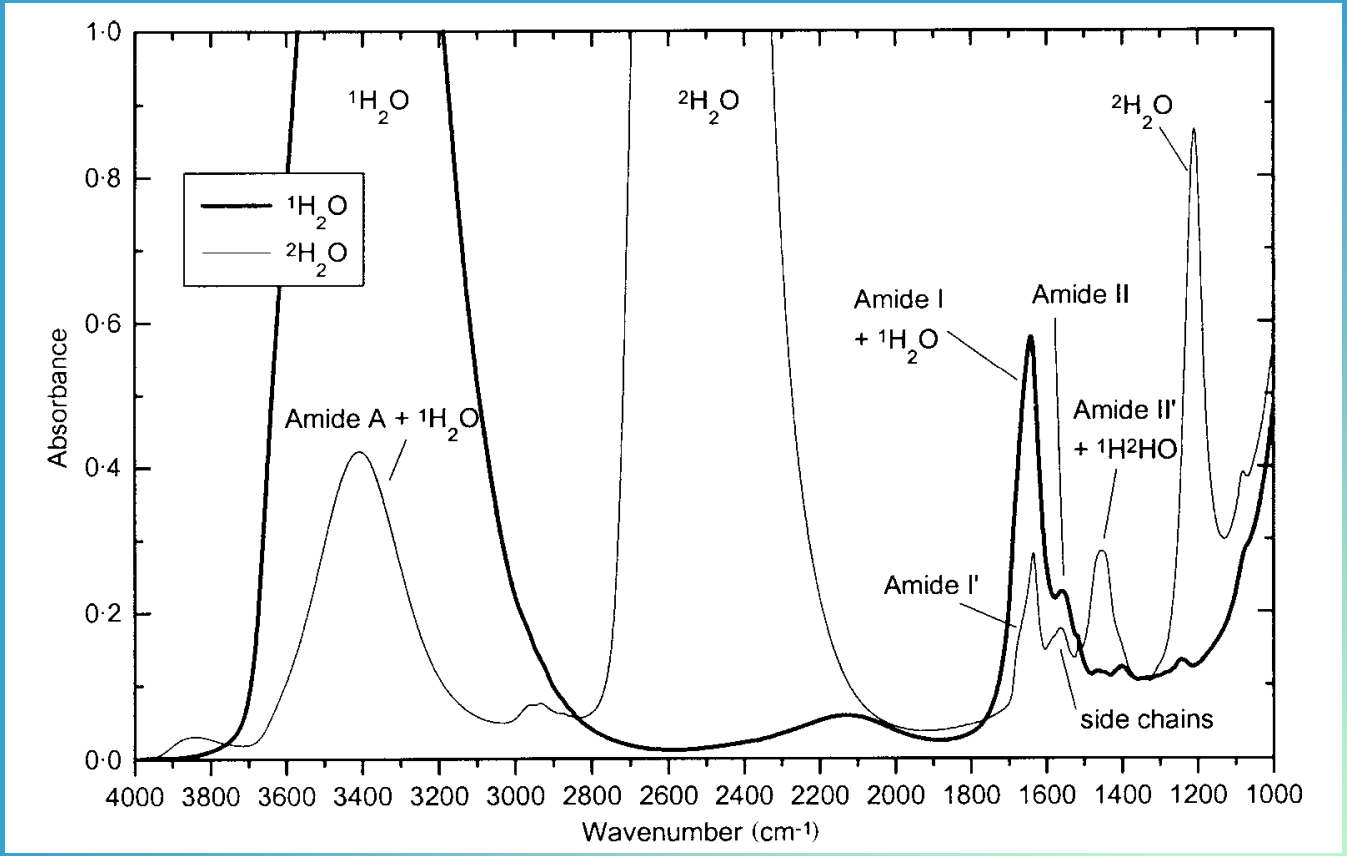
In the Raman spectrum of tryptophan, the indole ring mode near 880 cm^{-1}, which involves both ring stretching and displacement of the imino group, is *sensitive to hydrogen bonding*. *Raman spectra of 3-methylindole*, recorded at the same experimental conditions as the infrared spectra, *demonstrate that the frequency of the 880 cm^{-1} band decreases as the mole fraction of dioxane in solution increases*.

Vibrational spectroscopy and protein structure



Schematic representation of vibrational modes of amide I, amide II and amide III in N-methylacetamide, the simplest model of a peptide group in *trans* conformation. The frequencies of these modes reflect the structure of the main polypeptide chain (NH_2-$C_\alpha HR_1$-CO-NH-$C_\alpha HR_2$-CO-) irrespective of side chains (R_1, R_2, …).

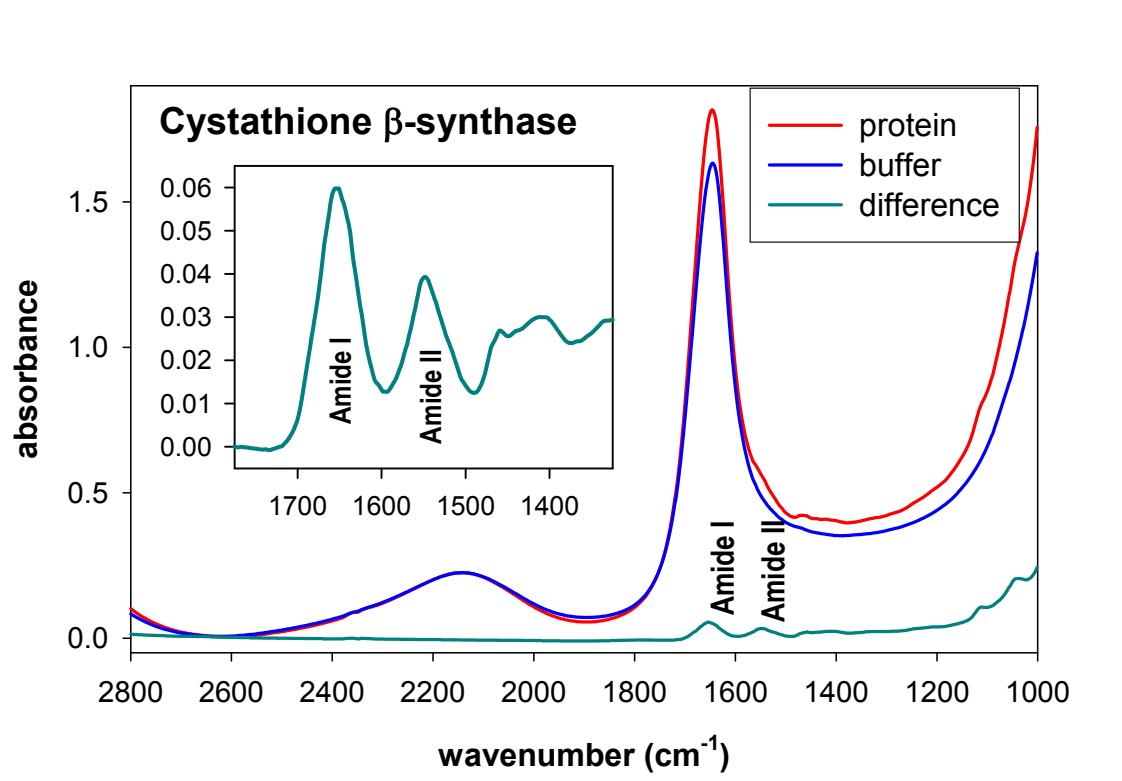
Vibrational spectroscopy and protein structure



Room temperature IR spectra of a protein in H_2O (bold line) and D_2O (thin line). Sample thickness was approximately 6 and 20 μm for H_2O and D_2O, respectively.

Vibrational spectroscopy and protein structure

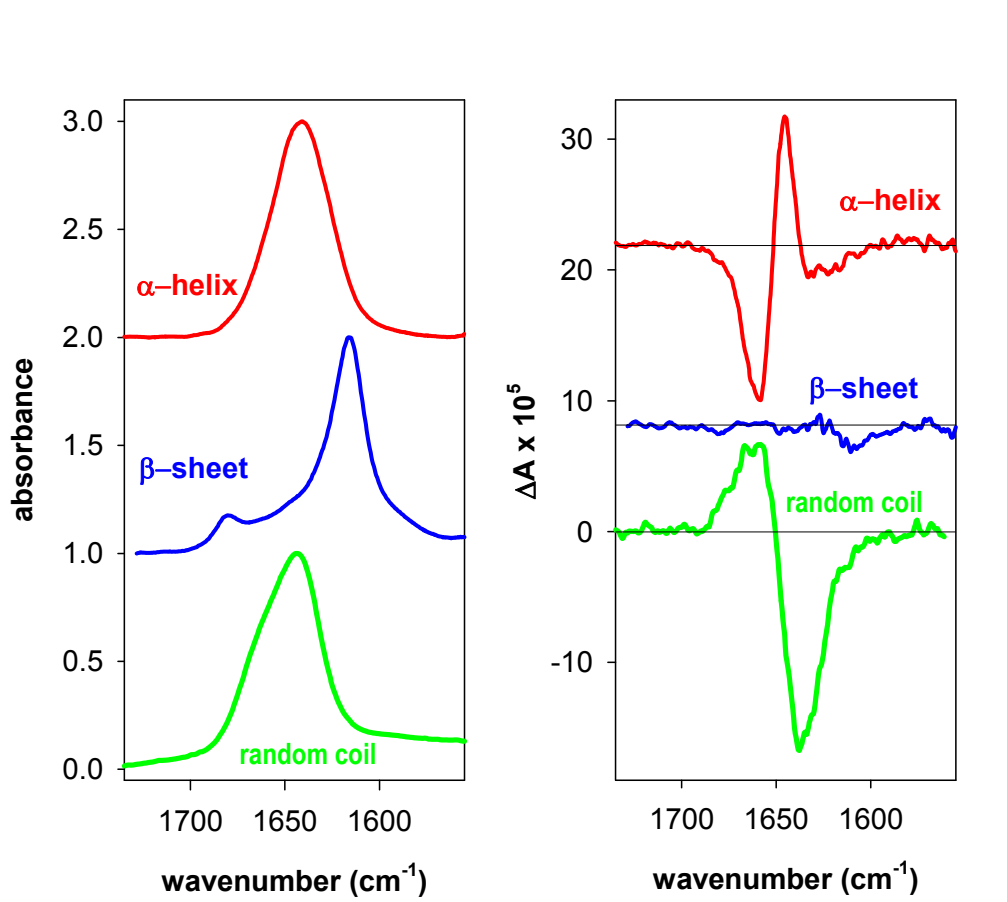
Buffer subtraction



IR spectrum of the protein cystathione ß-synthase in TrisHCl buffer pH 8.0 (red) at a protein concentration of 12.7 mg/ml placed between a pair of CaF_2 windows separated by a pathlength of 12 μm together with buffer spectrum (blue) and after subtraction of the buffer spectrum (cyan).

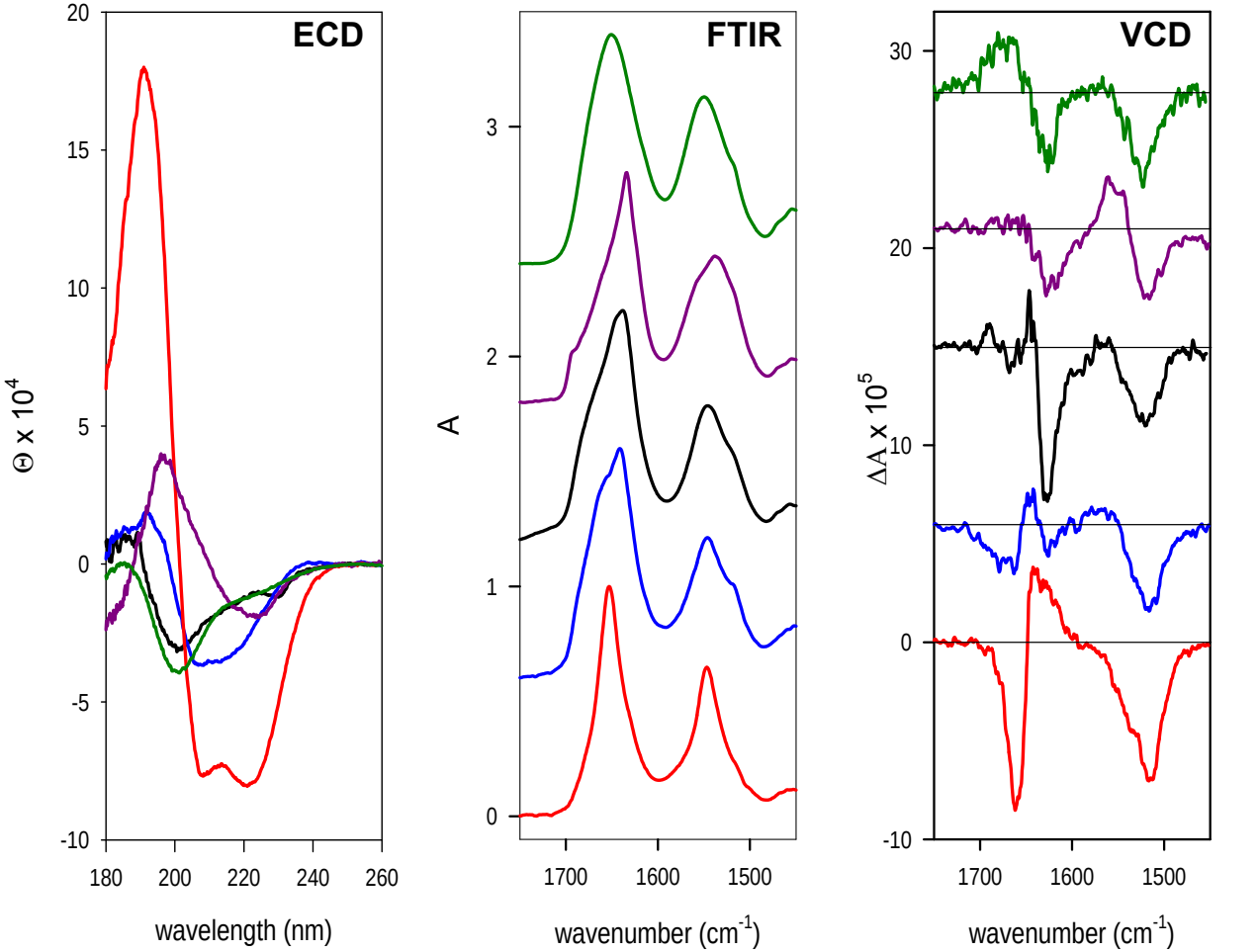
Vibrational spectroscopy and protein structure

Peptide studies



Comparison of amide I' IR absorption (left) and VCD (right) spectra of polypeptides dissolved in D_2O obtained for α-helical poly(LKKL) (top), β-sheet poly(LK) with high salt (middle), and random coil poly(K) at pH~7 (bottom), where L is L-Leu and K is L-Lys. IR and VCD spectra are normalized to A_{max} = 1 for the amide I'.

Rapid quantification of protein secondary structure



Comparison of typical ECD (left), IR (middle, amide I & II) and VCD (right, amide I & II) spectra for proteins in H_2O that have different dominant secondary structures - myoglobin (red, 77% α, 0% β), ribonuclease A (blue, 21% α, 35% β), α-chymotrypsin (black, 12% α, 32% β), concanavalin A (violet, 0% α, 40% β), α-casein (green, random coil). IR and VCD spectra are normalized to A_{max} = 1 for the amide I.

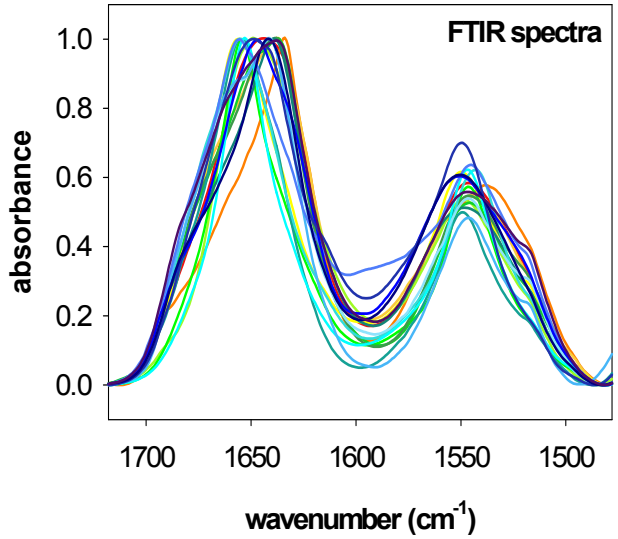
Rapid quantification of protein secondary structure

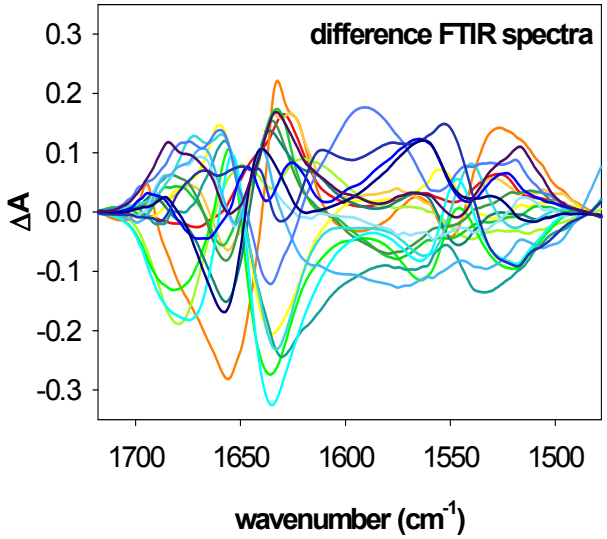
Protein reference set and secondary structure as determined by X-ray studies (Kabsch and Sander)

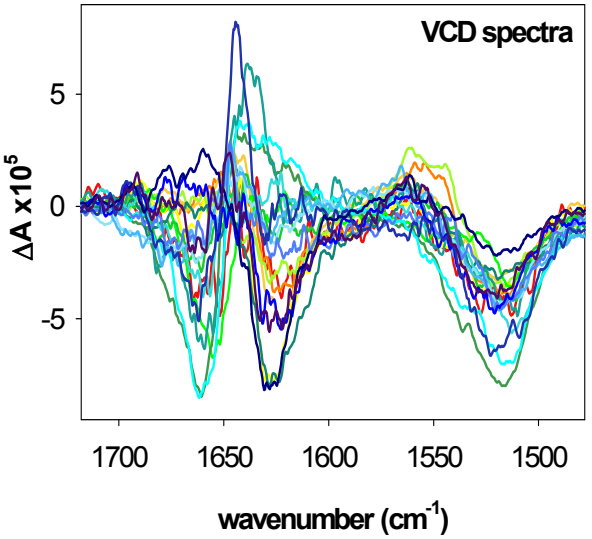
	Protein	Type	PDB code	helix	β-sheet	turn	bend	other
1	Alcohol dehydrogenase	αβ	4ADH	24.9	20.6	14.7	13.6	26.2
2	Carbonic anhydrase	ββ	1CA2	16.0	28.9	12.9	15.2	27.0
3	Chymotrypsinogen A	ββ	2CGA	14.3	32.2	14.3	12.7	26.5
4	Chymotrypsin type II	ββ	5CHA	11.8	32.1	11.4	14.4	30.4
5	Concanavalin A	ββ	3CNA	0.0	40.5	9.3	19.8	30.4
6	Cytochrome c	αα	1CYT	42.7	0.0	15.5	8.7	33.0
7	Glutathion reductase	αβ	2GRS	29.3	18.7	10.4	19.3	22.3
8	Hemoglobin	αα	1HCO	62.7	0.0	18.8	6.6	11.9
9	λ-imunoglobuline	ββ	1REI	2.8	47.7	14.0	11.2	24.3
10	Lactate dehydrogenase	αβ	4LDH	36.8	11.3	14.3	13.1	24.6
11	Lyzozyme	αβ	7LYZ	38.8	7.8	20.9	16.3	16.3
12	Myoglobin	αα	1MBN	77.1	0.0	9.8	1.9	11.1
13	Ribonuclease A	αβ	1RN3	21.0	34.7	11.3	14.5	18.6
14	Ribonuclease S	αβ	1RNS	20.8	35.2	7.2	14.4	22.4
15	Subtilisin BPN	αβ	1SBT	30.2	17.8	15.3	12.0	24.7
16	Superoxid dismutase	ββ	2SOD	2.0	38.4	14.6	20.5	24.5
17	Triosephosphate isomerase	αβ	1TIM	45.8	17.0	7.3	8.9	21.1
18	Trypsin inhibitor	αβ	3PTI	20.7	24.1	6.9	19.0	29.3
19	Trypsin	ββ	3PTN	9.9	32.3	14.8	17.9	25.1

Rapid quantification of protein secondary structure

Spectral variability of the reference set of proteins







Variability of IR (left), difference IR (middle) a VCD (right) spectra of the reference set of 19 proteins in H_2O in amide I and II region. FTIR spectra are normalized to A_{max} = 1 of amide I. Difference FTIR spectra were obtained by subtraction of the average spectrum (over the reference set) from individual spectra.

Rapid quantification of protein secondary structure

Distribution of secondary structural information as determined from X-ray crystallography studies for the 19 proteins used (in % structure).

Secondary structure	Kabsch a Sander				
	helix	β-sheet	turn	bend	other
Minimum	0	0	6.9	1.9	11.1
Maximum	77.1	47.7	20.9	20.5	33.0
Mean	26.7	23.1	12.8	13.7	23.7
Standard deviation	20.4	14.5	3.8	4.8	5.9

Rapid quantification of protein secondary structure

Determination of the secondary-structure composition of proteins

- band narrowing and curve-fitting of the amide I band

  1. component bands resolution by mathematical procedures (calculation of the second derivative of the spectrum or Fourier self-deconvolution)
  2. fitting of the component bands at the positions found
  3. assignment of the component bands to secondary structures

- calibration set of spectra from proteins with known structure to perform pattern-recognition calculations (amide I, amide I & II)

  1. data reduction to a few independent subspectra (e.g. by means of principal component method of factor analysis – PC/FA)
  2. regression fits to fractional components of secondary structure (FC)

  or

  1. neural network analysis

Determination of the secondary-structure composition of proteins

Table 3. Errors in predictions of α-helix and β-sheet fractions from optimal models for individual spectroscopic methods

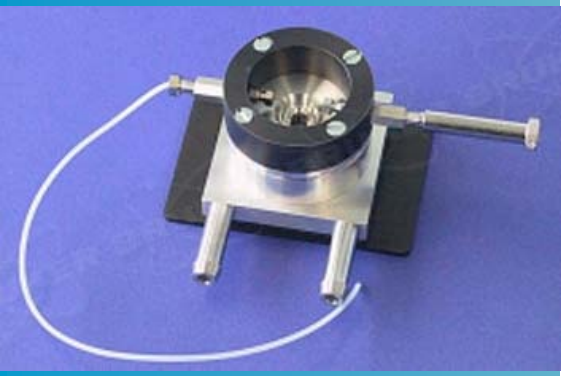
	Helix							Sheet						
Protein name	FTIR full	DF	VCD H_2O	VCD AI	VCD AII	VCD ^2H_2O	ECD	FTIR full	DF	VCD H_2O	VCD AI	VCD AII	VCD ^2H_2O	ECD
α-Chymotrypsinogen A	2.4	1.9	1.2	2.7	4.9	11.4	1.4	0.6	2.1	3.8	8.8	0.7	0.2	0.2
Alcohol dehydrogenase	7.3	3.5	9.7	14.0	0.9	1.6	2.0	8.9	12.5	7.2	7.9	1.7	3.7	4.5
α-Chymotrypsin type II	1.7	0.5	7.9	6.6	6.5	4.3	1.0	1.3	0.4	1.3	3.7	2.1	6.0	1.0
Concanavalin A	16.8	5.2	3.9	9.2	12.7	16.3	12.7	4.3	4.1	9.6	4.4	12.2	13.0	10.4
Carbonic anhydrase	5.1	2.3	13.9	11.3	13.1	2.4	8.7	6.5	5.3	11.9	13.5	7.9	1.0	10.6
Cytochrome *c*	6.6	10.8	1.8	7.3	2.7	7.3	12.2	13.8	15.7	1.9	11.9	4.1	1.9	19.2
Glutathione reductase	1.5	9.5	2.5	6.3	2.4	4.2	1.9	1.2	3.7	6.4	6.0	3.2	6.6	0.9
Hemoglobin	1.9	0.8	2.5	2.9	2.0	0.1	1.1	2.9	0.8	10.3	11.3	10.3	4.8	10.0
λ-Immunoglobulin	13.5	13.1	2.0	3.1	3.4	14.9	2.8	13.8	17.2	10.0	12.9	7.4	15.0	2.5
Lactate dehydrogenase	5.6	13.7	4.5	9.5	5.6	23.3	4.1	4.6	3.8	3.9	11.7	0.7	7.3	2.2
Lysozyme	22.4	12.1	0.2	23.3	1.0	12.1	4.5	2.4	5.0	1.8	9.2	3.3	14.1	0.7
Myoglobin	18.6	16.5	15.8	18.2	17.4	28.5	5.3	3.3	0.5	3.5	0.9	2.9	4.7	11.8
Ribonuclease A	1.6	13.4	7.6	3.4	10.1	9.6	1.8	8.1	9.3	9.6	10.6	9.3	4.1	10.1
Ribonuclease S	9.7	2.1	6.9	2.7	13.6	10.7	2.2	0.1	7.7	5.6	3.2	12.8	1.4	6.0
Subtilisin BPN'	0.7	3.8	7.6	4.7	13.7	0.8	7.4	3.3	0.7	5.9	1.8	17.4	13.0	4.3
Superoxide dismutase	25.2	6.0	12.7	15.0	3.1	2.3	17.3	16.5	10.2	6.0	3.8	8.3	3.9	5.2
Triose phosph.isomerase	1.1	4.8	0.1	1.1	1.8	9.8	1.0	10.7	2.9	4.5	1.7	6.5	2.8	10.5
Trypsin inhibitor	0.9	7.1	21.4	16.6	22.9	6.7	9.9	7.0	8.6	14.2	7.2	14.5	4.5	0.2
Trypsin	5.0	8.3	11.8	1.9	9.6	8.1	0.5	0.2	2.5	3.4	6.3	2.0	5.8	6.9

Table 4. Errors in predictions of α-helix and β-sheet fractions from optimal models for combinations of spectroscopic methods

	Helix					Sheet				
Protein name	ECD+ VCD H_2O	ECD+DF	VCD H_2O +DF	VCD ^2H_2O +DF	VCD H_2O AI+AII	ECD+ VCD H_2O	ECD+DF	ECD+ FTIR	VCD ^2H_2O+ DF	VCD H_2O AI+AII
α-Chymotrypsinogen A	1.4	0.8	1.5	3.1	5.5	1.7	2.8	1.7	7.2	6.5
Alcohol dehydrogenase	2.0	1.5	2.9	5.6	2.6	2.7	7.2	0.2	1.8	8.3
α-Chymotrypsin type II	1.0	3.5	1.6	2.2	4.2	2.8	6.6	1.2	2.5	0.2
Concanavalin A	12.7	7.3	1.8	8.9	4.6	9.9	3.9	2.3	2.2	3.9
Carbonic anhydrase	8.7	10.4	10.9	9.7	16.4	12.4	8.5	9.8	10.1	11.9
Cytochrome *c*	12.2	5.7	4.8	5.9	6.6	3.1	8.4	6.9	8.0	0.2
Glutathione reductase	1.9	5.1	0.0	2.5	3.3	4.0	6.8	4.4	2.9	0.6
Hemoglobin	1.1	3.6	1.1	4.8	3.3	4.6	10.5	3.4	5.2	10.1
λ-Immunoglobulin	2.8	8.0	1.4	7.8	0.6	4.0	9.5	16.8	17.0	1.6
Lactate dehydrogenase	4.1	0.0	4.1	13.6	2.1	6.2	3.5	5.1	5.2	4.7
Lysozyme	4.5	1.9	0.4	9.5	6.2	6.2	2.3	2.4	2.1	8.2
Myoglobin	5.3	1.2	15.8	12.7	17.3	0.8	2.2	1.9	1.5	0.0
Ribonuclease A	1.8	2.9	10.8	6.1	9.0	11.9	6.8	9.6	8.3	12.9
Ribonuclease S	2.2	3.7	6.0	7.4	11.9	9.1	3.9	11.6	8.4	4.1
Subtilisin BPN'	7.4	7.2	8.8	9.6	5.3	11.8	0.9	5.2	1.1	1.7
Superoxide dismutase	17.3	19.2	22.6	17.5	15.1	11.3	7.1	9.9	14.6	5.1
Triose phosph.isomerase	1.0	5.8	3.7	2.2	1.7	3.8	1.7	8.2	4.1	4.4
Trypsin inhibitor	9.9	3.8	19.4	17.3	18.5	7.1	2.9	17.0	15.4	13.4
Trypsin	0.5	2.8	1.4	2.5	3.4	4.0	0.5	0.6	0.3	2.5

Confocheck™ (Bruker Optics)

FT-IR system that is specially designed for protein analysis.



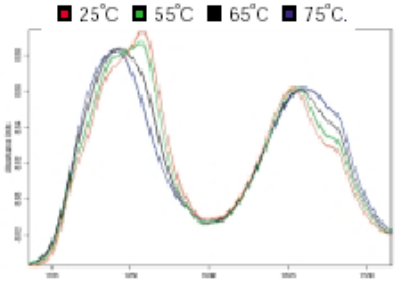
The AquaSpec™ is a dedicated flow-through transmission cell for the measurement of water solved proteins.

FT-IR Proteomics

After the successful human genome project provided almost endless information about possible sequences of proteins, Proteomics will have to provide the means to create, analyze and to understand structure and function relationships of these many proteins. Fourier Transform Infrared spectroscopy has contributed to protein structure analysis in the past but was not very attractive to many researchers, since it cannot reveal a 3D picture of the protein structure in contrast to X-ray crystallography and NMR spectroscopy. However, the restrictions of the latter methods (the need of crystals or small water-soluble proteins) provide infrared methods with a serious advantage. Using IR spectroscopy proteins can be studied in aqueous solution regardless of their molecular weight. Moreover, IR is capable to reveal conformational changes from solved proteins with very high sensitivity. It is therefore a very suitable tool for the control of protein stability or for the monitoring of induced unfolding/refolding processes.

Confocheck™ is a dedicated and compact FTIR system for fast and easy determination of protein secondary structure. Its specific configuration facilitates a fast data acquisition with a high sample throughput controlled by an user friendly software interface. Within fractions of a minute, high quality spectra from proteins in solution are acquired. Afterwards, quantitative determination of secondary structure elements (α-helix, β-strand) from the analyzed protein samples is accomplished within seconds via pattern recognition methods based on a growing protein spectra library. Confocheck™ is also designed to follow temperature induced protein conformational changes by FTIR. Therefore, an integrated cooling thermostat modulates the temperature at the internal flow-through transmission cell very accurately. Automatically, measurements are recorded for the specified temperature range with a selected increment, whereby the water vapor is subtracted from the received spectra subsequently. Representative results showing the temperature-induced unfolding/refolding from the enzyme RNase A are summarized in figures 1 and 2. In Fig 1 the protein spectra for different temperatures are displayed. It is obvious that during the heating the β-sheet was unfolded indicated by the peak at 1641 cm^{-1}. Every temperature is represented by two traces, which were obtained during the heating and subsequent cooling cycle, respectively. The similarity between both traces at any temperature reveals that the heating induced unfolding process was completely reversible and the protein was renatured during the cooling procedure afterwards. The complete protein absorption change during the temperature cycling is shown as a 2D-false color plot in Fig. 2.

Figure 1.



To introduce additional capabilities such as online dialysis, an ATR design is avaiable for protein studies - BioATRCell™ I. This accessory, which is completely compatible with the hardware and software from Confocheck™, is also very suited to study protein aggregation and assembling processes. Furthermore, protein stability can be tested on native proteins, different pH and salt conditions, after introducing mutations or for storage dependencies. In addition, protein-protein, protein-DNA, protein-substrate and protein-ligand binding can be studied using the different options of the ATR accessory. Low molecular weight compounds can be applied to the protein of investigation by dialysis. The dialysis membrane feature is essential as it avoids protein dilution during the experiment. Typically, the experimental results combine binding kinetics with detailed molecular information from the protein/ligand during interaction.

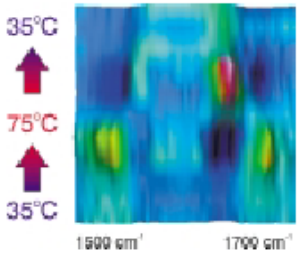


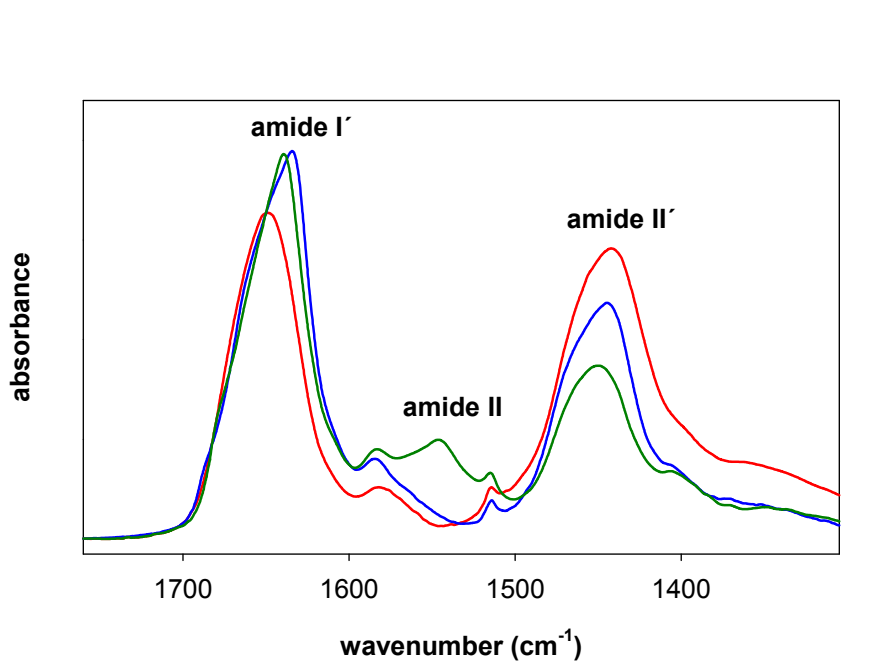
Figure 2

Further studies possible with Confocheck™ / BioATRCell™ I:

- Unrestricted secondary structure analysis and classification of proteins.
- Tertiary interactions of amino acid side chains in proteins.
- Thermal and chemical stability of proteins and DNA/RNA structures.
- Study of protein-protein and protein-DNA interactions
- Monitoring denaturation (unfolding), refolding and aggregation of proteins.
- Investigation of binding effects on protein and DNA/RNA structures.
- Light and thermally-induced kinetics studies.

Vibrational spectroscopy and protein structure

$H_2O \leftrightarrow D_2O$ exchange



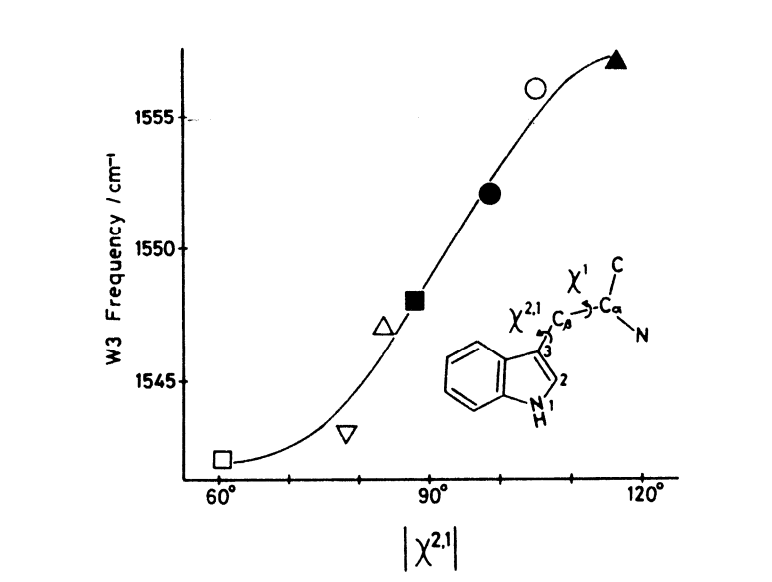
FTIR absorption spectra of ribonuclease A in D_2O in amide I , II a II region at 20 C (green - folded structure), at 70 C (red – partly unfolded structure) and after cooling to 20 C (blue – folded structure again). During following heating/cooling cycles one gets spectra identical (within experimental error) with either the red spectrum(70 C) or the blue one (20 C).

Side-Chain Conformations and Local Environments

Tryptophan

W3 mode (ca. 1550 cm^{-1}) – involving C2=C3 stretching of pyrrole nucleus

- Direct correlation between the Raman frequency of W3 and the absolute value of the torsional angle ($|\chi^{2,1}|$) of the C2-C3-Cβ-Cα network has been established by comparing Raman spectra of appropriate tryptophan derivatives with molecular conformations revealed in X-ray crystal structures.
- The correlation is applicable to both proteins solutions and solids.
- In case of proteins containing more than one Trp residue, the Raman peak position, bandwidth and band shape give information about the distribution of side-chain conformations of Trp residues.



Relationship between the frequency of tryptophan normal mode W3 and the side-chain torsion $\chi^{2,1}$ (Miura et al. *J. Raman Spectrosc.* **20** (1989) 667). Different symbols refer to different tryptophan model compounds.

Side-Chain Conformations and Local Environments

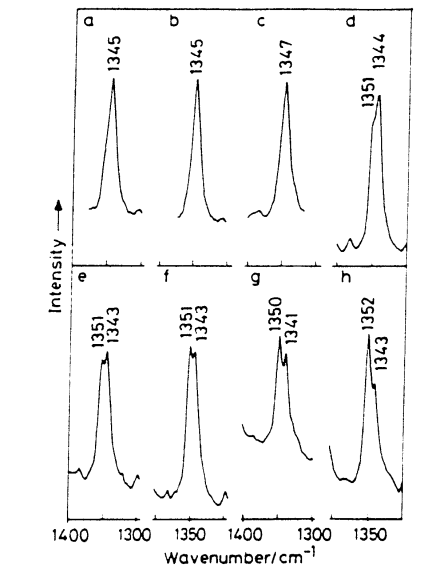
Tryptophan

Fermi doublet at ca. 1360 and 1340 cm^{-1}

- A pair of Raman bands observed near 1360 and 1340 cm^{-1} in proteins containing Trp has been assigned to Fermi resonance (involving W7 fundamental expected at 1350 cm^{-1} and one or more combination modes of out-of-plane vibrations).
- The relative intensity ratio I_{1360}/I_{1340} is a sensitive marker of the amphiphatic environment of the indole ring.
- I_{1360}/I_{1340} increases with increases hydrophobicity, exhibiting values near

  I_{1360}/I_{1340} = 0.65-0.92 in hydrophylic environment

  I_{1360}/I_{1340} = 1.23-1.32 in hydrophobic environment
- In off-resonance Raman spectra, 1360 cm^{-1} band intensity is used as the marker of hydrophobicity, because the 1340 cm^{-1} peak is usually overlapped by the C-H bending modes of aliphatic chains.



Raman spectra in the region 1300-1400 cm^{-1} of solutions of 3-methylindole in

a) methanol
b) dimethylformamide
c) dioxane
***d) o*-dichlorobenzene**
e) benzene
f) toluene
g) CS_2
***h) n*-hexane**

Side-Chain Conformations and Local Environments

Tryptophan

W6 mode (ca. 1430 cm^{-1}) – substantial contribution from indole N-H in-plane bending

- Raman frequency of the W6 mode exhibits a strong dependence upon the strength of hydrogen bonding of the NH donor
- W6 shifts from 1422 cm^{-1} (no hydrogen bonding) to 1441 cm^{-1} (strong hydrogen bonding to an appropriate acceptor.
- Easily observed in UVRR employing 266 nm excitation. In off-resonance Raman spectra usually obscured by more intense C-H bending modes of aliphatic side chains.

W17 mode (ca. 880 cm^{-1})

- deformation of the six-membered ring and the imino group displacement
- W17 frequency is sensitive to hydrogen bonding
- W17 shifts from 883 cm^{-1} (no hydrogen bonding) to 871 cm^{-1} (strong hydrogen bonding)
- W17 exhibits a distinct intensity increase in very strong hydrophobic conditions.

W18 mode (760 cm^{-1})

- indole ring-breathing vibration
- the most intense Raman band in off-resonance spectra of proteins – useful analytical indicator of tryptophan in proteins
- Additionally, its intensity is sensitive to the amphiphatic environment of indole ring – increases if the hydrophobicity of the indole environment is decreased.

Side-Chain Conformations and Local Environments

Tyrosine

Tyrosine Fermi doublet at ca. 850 and 830 cm^{-1}

- A pair of Raman bands observed near 850 and 830 cm^{-1} in proteins containing Tyr has been assigned to Fermi resonance (involving Y1 fundamental expected at 840 cm^{-1} and Y16a overtone, occurring near 420 cm^{-1}).
- The relative intensity ratio I_{850}/I_{830} is a **unambiguous marker of the hydrogen bonding state**. In a protein containing more than one Tyr I_{850}/I_{830} reflects the average of the hydrogen bonding states of all the tyrosines.
- With visible excitation, the ratio I_{850}/I_{830} ranges from 0.3 (phenolic OH is a strong hydrogen bond donor – „buried“ tyrosine) to 2.5 (strong hydrogen acceptor – „exposed“ tyrosine).

Side-Chain Conformations and Local Environments

Cysteine

S-H stretching vibration (2500 - 2600 cm^{-1})

- nearly pure group frequency
- intrinsically high Raman intensity
- it occurs in the region which is essentially devoid of interference from any other Raman bands of proteins or aqueous solvent
- the band frequency is sensitive to hydrogen bonding of the S-H donor and acceptor groups
- the band intensity is a measure of molecular concentration of thiol groups in the protein

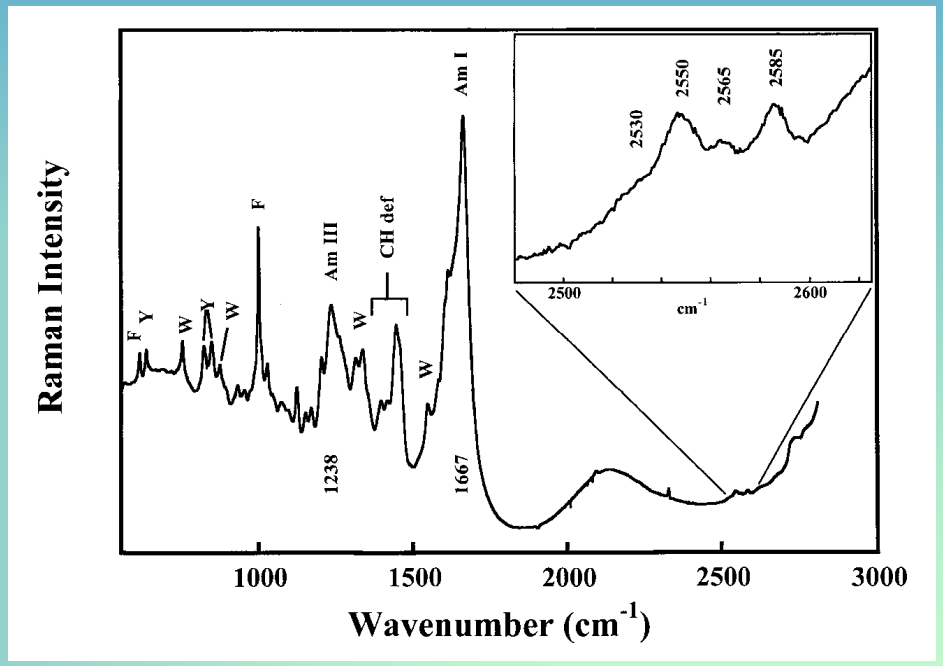
Side-Chain Conformations and Local Environments

Cysteine

S-H stretching vibration (2500 - 2600 cm^{-1})

Dependence of the Raman S-H frequency and bandwidth on hydrogen bonding

Hydrogen-bonding state of S-H group	S-H frequency (cm^{-1})	Band width (cm^{-1})	Examples
No hydrogen bond	2581–2589	12–17	Thiols in CCl_4 (dilute)
S acceptor	2590–2595	12–17	Thiols in $CHCl_3$
Weak S-H donor	2575–2580	20–25	Thiol neat liquids; thiols in thioethers
Moderate S-H donor	2560–2575	25–30	Thiols in acetone; crystal structures
Strong S-H donor	2525–2560	35–60	Thiols in dimethylacetamide; crystal structures
S-H donor and S acceptor	2565–2575	30–40	Thiols in H_2O

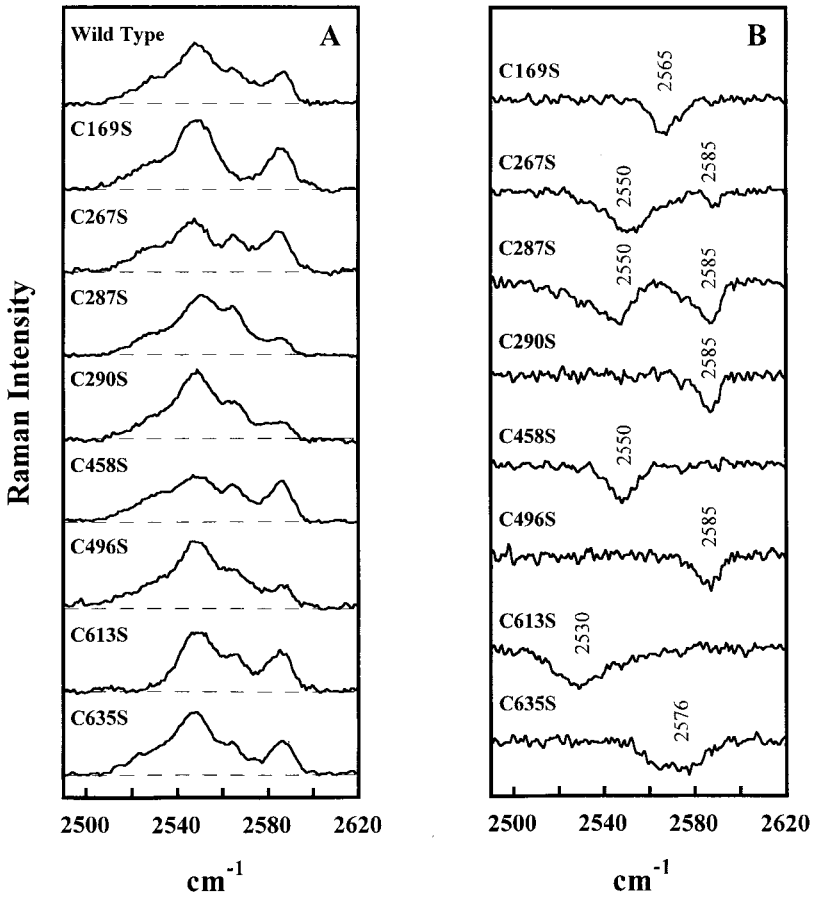


Raman spectrum (670-2800 cm^{-1}) of the P22 trimeric tailspike protein at 10°C. The inset at upper right, which shows an amplification of the spectral interval 2480-2620 cm^{-1}, exhibits the composite S-H stretching profile (bands at 2530, 2550, 2565 and 2585 cm^{-1}) of the eight cysteine residues per unit. The data are not corrected for solvent contribution. From Raso et al. *J. Mol. Biol.* 307 (2001) 899.

Side-Chain Conformations and Local Environments

Cysteine

S-H Raman signatures (2480-2630 cm^{-1}) of tailspike cysteine residues



Spectral contributions and hydrogen-bond strengths of cysteine sulfhydryl groups of the native P22 tailspike protein

Residue	Raman S-H band[a]	Hydrogen-bond strength[b]
Cys169	2565	Moderate
Cys267	2550 (90 %)	Strong
	2585 (10 %)	Very weak
Cys287	2550 (63 %)	Strong
	2585 (37 %)	Very weak
Cys290	2585	Very weak
Cys458	2550	Strong
Cys496	2585	Very weak
Cys613	2530	Very strong
Cys635	2576	Weak

[a] Raman band center (in cm^{-1} units). Numbers in parentheses are the percentages of total intensity contributed by the specified cysteine sulfhydryl at the indicated cm^{-1} value.
[b] Based upon the results reported by Li & Thomas.[11]

A. The Raman S-H profiles observed for the wild-type tailspike and for each of eight Cys → Ser mutants, as labeled.

B. Raman difference spectra computed as mutant minus wild-type, for each of the eight Cys → Ser mutants. In each trace, the S-H Raman signature of the mutated Cys site is revealed as a negative band.

Reaction-induced IR difference spectroscopy

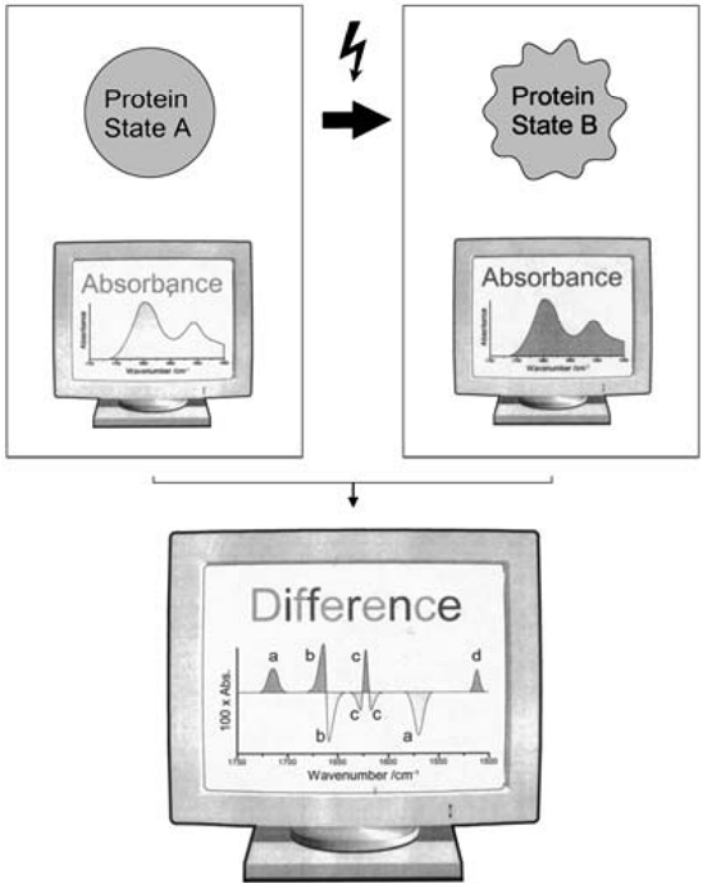


Fig. 14. Principle of reaction-induced IR difference spectroscopy. The protein is prepared in state A and the IR absorbance of this state is characterized. Then the reaction from state A to state B is triggered in the IR cuvette and the absorbance of state B measured. From the two absorbance spectra, a difference spectrum is calculated which shows the absorbance changes due to the reaction. Negative bands are characteristic of the initial state A (in light grey) and positive bands of the final state B (in dark grey). Several causes for absorbance changes are discussed in the text.

From Barth & Zscherp, Quater. Rev. Biophys. 35, 369 (2002).

Protein stability

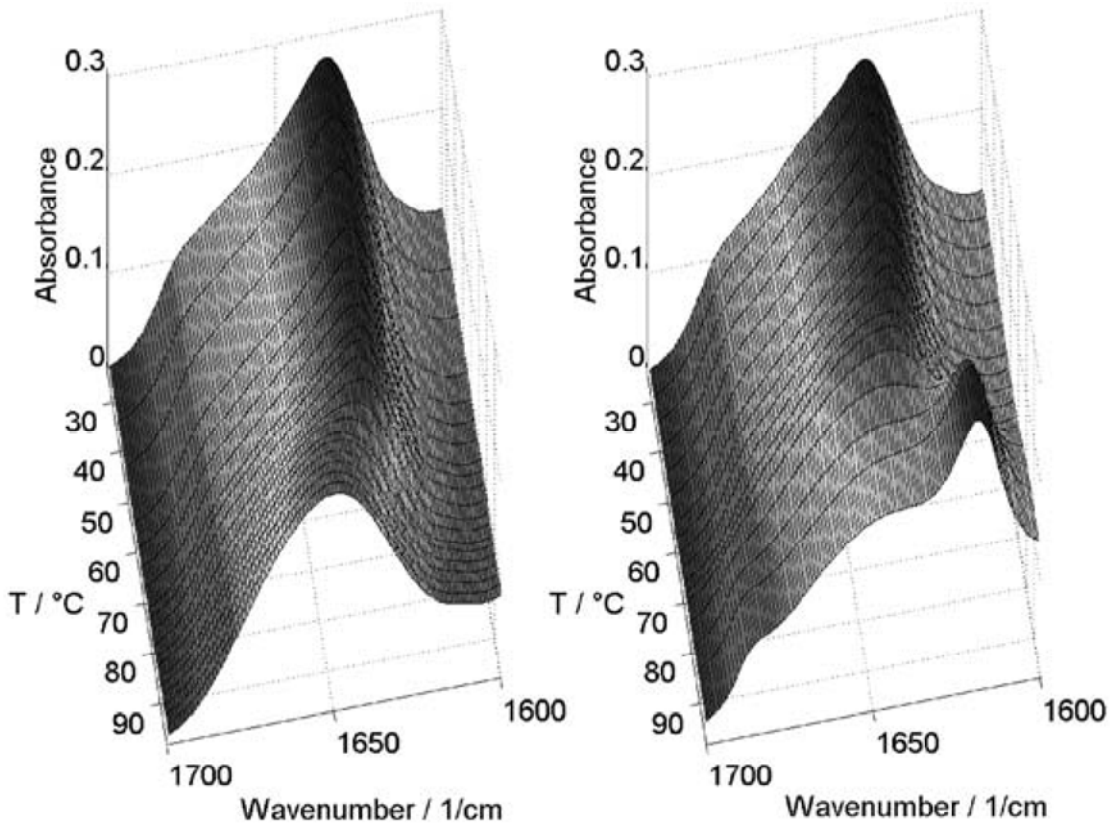


Fig. 13. Temperature-dependent IR spectra of the α-amylase inhibitor tendamistat, a small β-sheet protein. With a midpoint temperature of 82 °C, the wild-type protein unfolds and adopts an irregular structure. This leads to a broad amide I band centred at 1650 cm^{-1} (left). Mutation of the three Pro residues to Ala does not significantly alter the amide I band at room temperature (right). However, heating the Pro-free protein results in a downshift of the amide I maximum indicating aggregation of the sample. Moreover, the transition is already observed at 67 °C (C. Zscherp, H. Aygün, J. W. Engels & W. Mäntele, unpublished observations).

From Barth & Zscherp, Quater. Rev. Biophys. 35, 369 (2002).