

BLS1: Quality control for Biology.

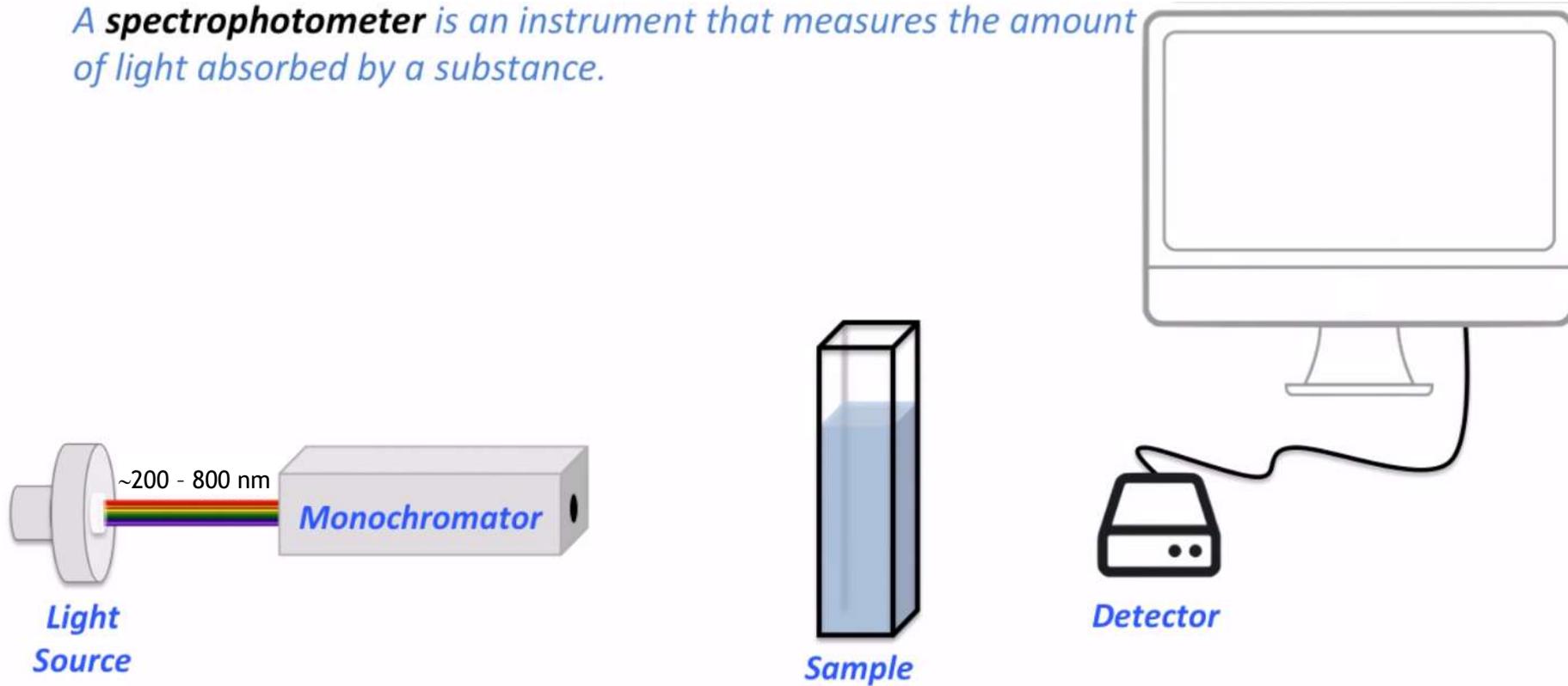
Lecture 4: UV and CD spectroscopy as qualitative and quantitative tools for proteins

Sébastien Brûlé - Institut Pasteur - sbrule@pasteur.fr



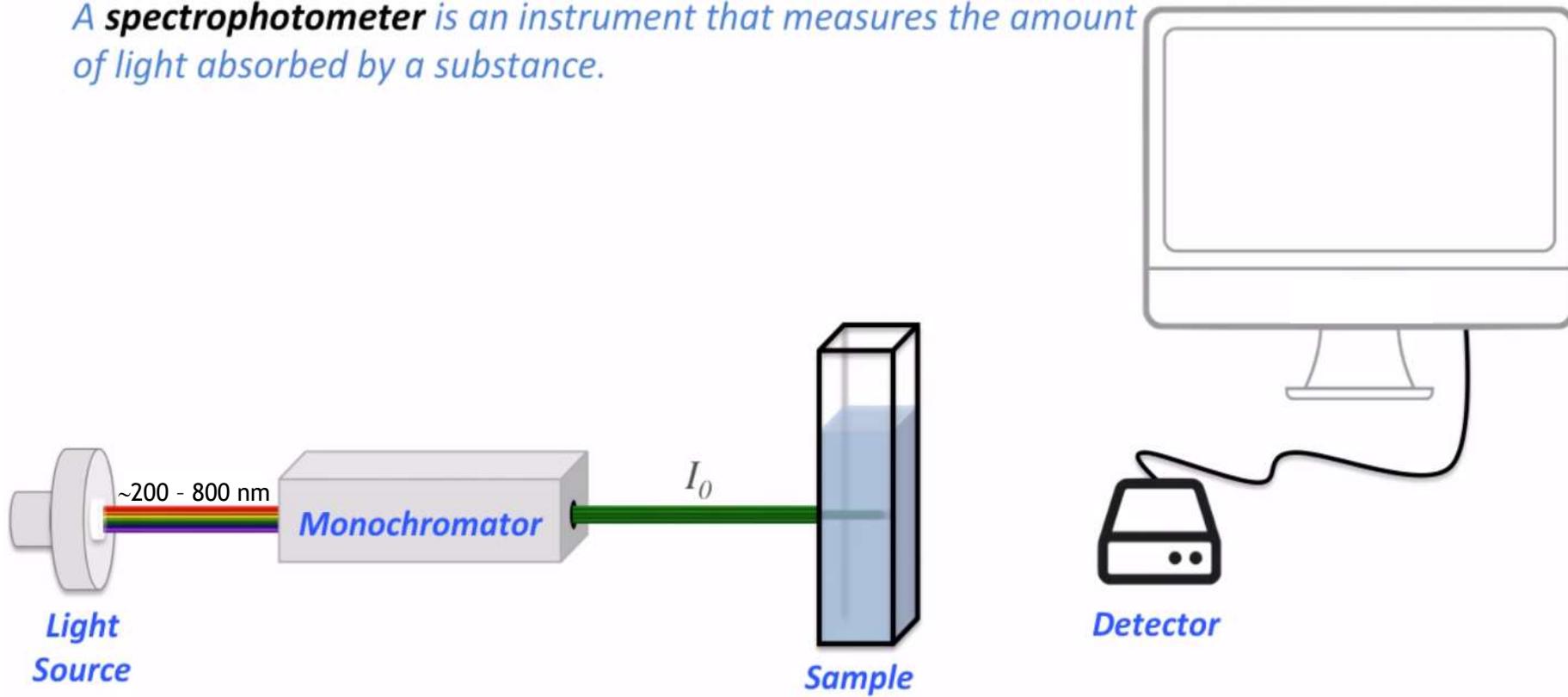
Principle of a spectrophotometer

A **spectrophotometer** is an instrument that measures the amount of light absorbed by a substance.



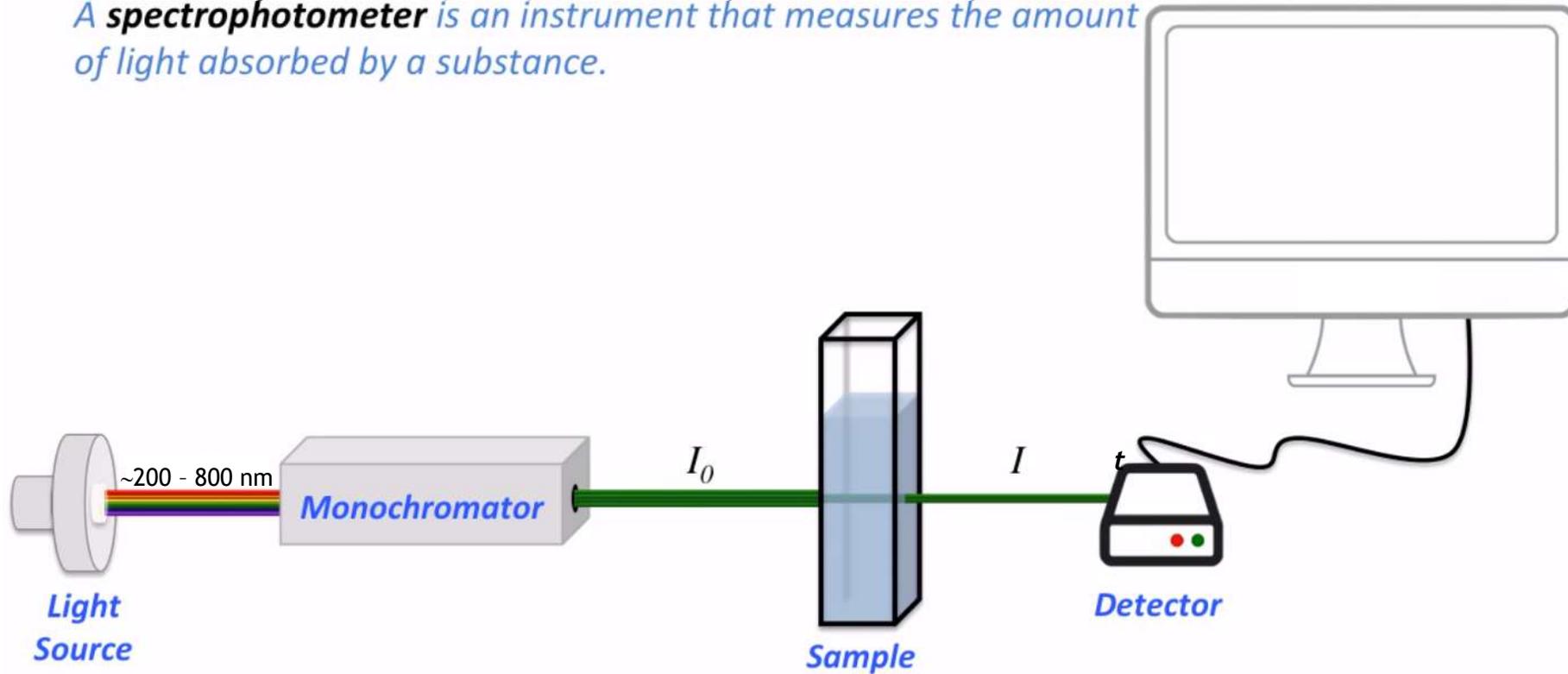
Principle of a spectrophotometer

A **spectrophotometer** is an instrument that measures the amount of light absorbed by a substance.



Principle of a spectrophotometer

A **spectrophotometer** is an instrument that measures the amount of light absorbed by a substance.

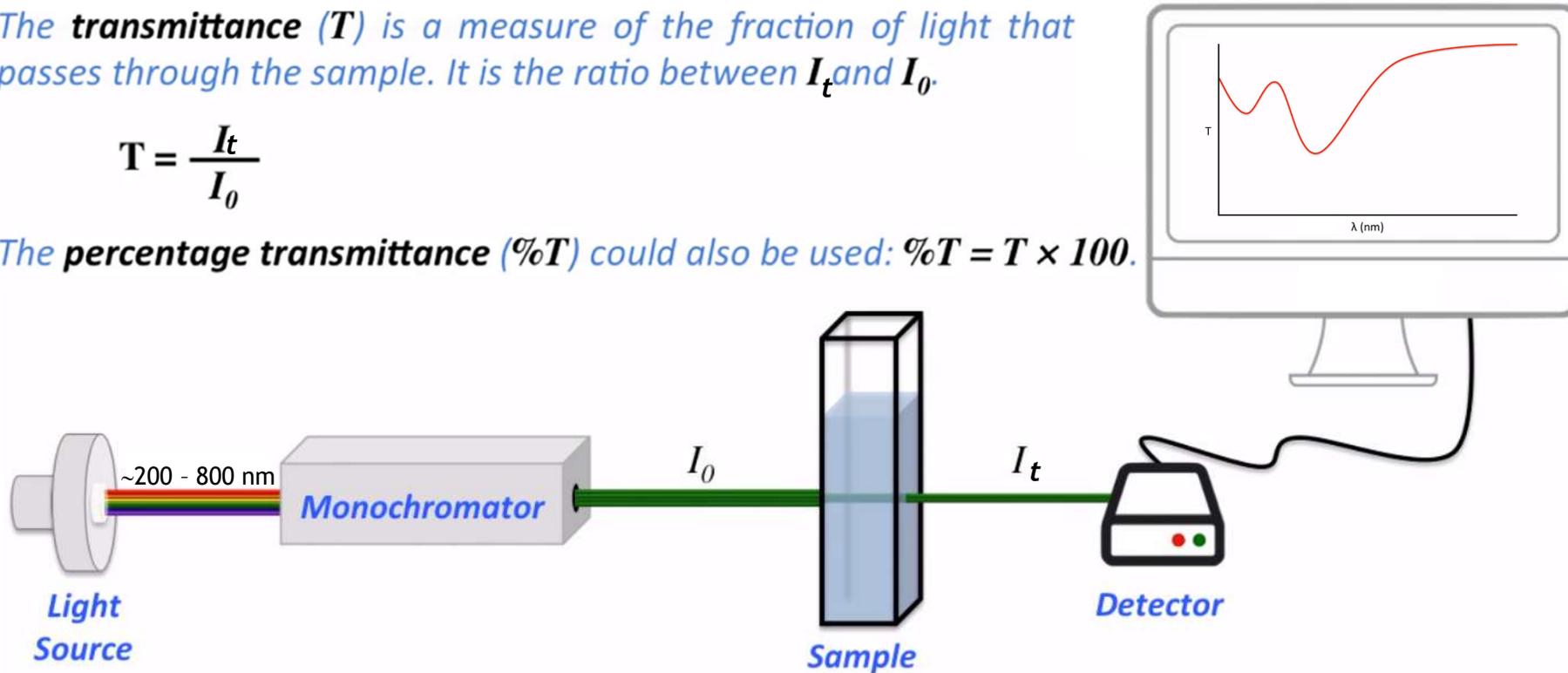


Principle of a spectrophotometer

The **transmittance** (T) is a measure of the fraction of light that passes through the sample. It is the ratio between I_t and I_0 .

$$T = \frac{I_t}{I_0}$$

The **percentage transmittance** ($\%T$) could also be used: $\%T = T \times 100$.

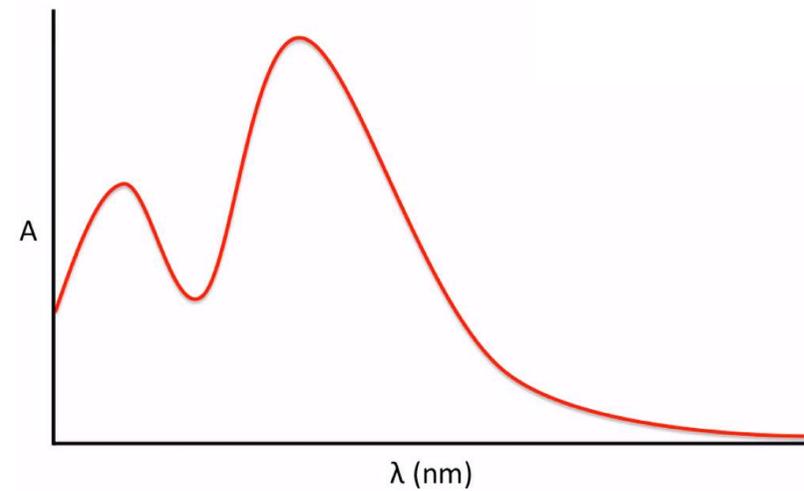
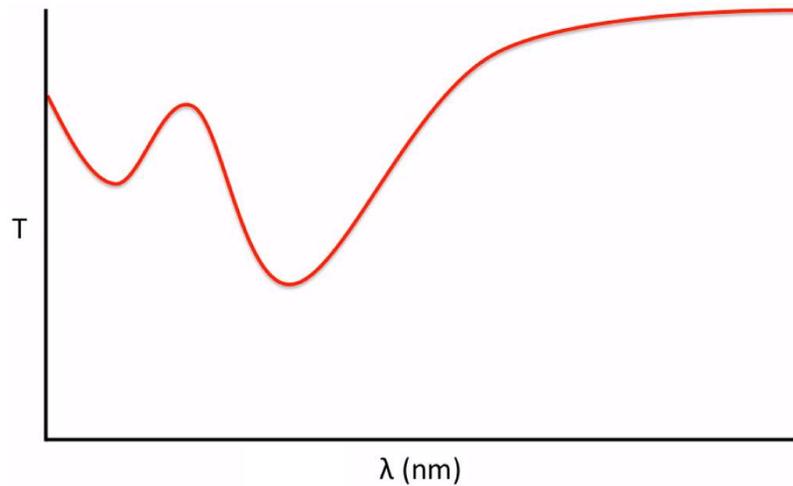


Principle of a spectrophotometer

Transmittance spectrum



Absorbance spectrum



The absorbance (A) is the amount of light absorbed by the sample. $A = -\log T = -\log \frac{I_t}{I_0}$



Beer-Lambert Law



$$C = A_{\lambda} / (\epsilon_{\lambda} \cdot l)$$

A = Absorbance at specific wavelength λ (no units).

ϵ = Molar extinction coefficient at specific wavelength ($L \cdot mol^{-1} \cdot cm^{-1}$).

l = pathlength (cm).

C = Concentration ($Mol \cdot L^{-1}$).

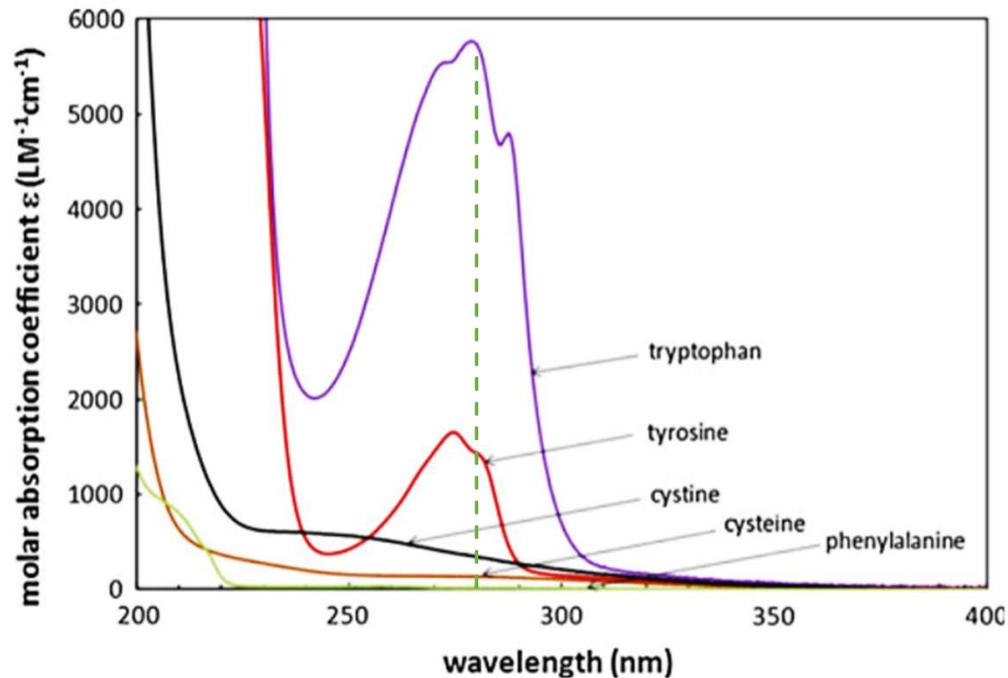
ϵ is the capacity of a molecule to attenuate (absorb) light at a given wavelength
(intrinsic property)

1 molecule => 1 epsilon!

λ is the distance from one wave cycle to the next



Where does the proteins UV contribution at 280 nm come from?



Contribution of 1 W at 280 nm in water:
3.7 times more than Y
45 times more than cystines

$$\epsilon_{280 \text{ nm}} = 5500 (\# \text{ of Trp}) + 1490 (\# \text{ of Tyr}) + 125 (\# \text{ of Cystines})$$

Pace CN et al. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 11, 2411-2423

Keith R. Millington. (2012) Diffuse reflectance spectroscopy of fibrous proteins. *Amino acids*

Tool to calculate epsilon at 280nm

ProtParam tool

ProtParam ([References](#) / [Documentation](#)) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in [Swiss-Prot](#) or [TrEMBL](#) or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) ([Disclaimer](#)).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

RESET

Compute parameters

<https://web.expasy.org/protparam/>

Tool to calculate epsilon at 280nm

BSA

Number of amino acids: 583

Molecular weight: 66432.96

Theoretical pI: 5.60

Amino acid composition: [CSV format](#)

Ala (A)	47	8.1%
Arg (R)	23	3.9%
Asn (N)	14	2.4%
Asp (D)	40	6.9%
Cys (C)	35	6.0%
Gln (Q)	20	3.4%
Glu (E)	59	10.1%
Gly (G)	16	2.7%
His (H)	17	2.9%
Ile (I)	14	2.4%
Leu (L)	61	10.5%
Lys (K)	59	10.1%
Met (M)	4	0.7%
Phe (F)	27	4.6%
Pro (P)	28	4.8%
Ser (S)	28	4.8%
Thr (T)	33	5.7%
Trp (W)	2	0.3%
Tyr (Y)	20	3.4%
Val (V)	36	6.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 99

Total number of positively charged residues (Arg + Lys): 82

Atomic composition:

Carbon	C	2934
Hydrogen	H	4615
Nitrogen	N	781
Oxygen	O	897
Sulfur	S	39

Formula: $C_{2934}H_{4615}N_{781}O_{897}S_{39}$

Total number of atoms: 9266

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient 42925
Abs 0.1% (=1 g/l) 0.646, assuming all pairs of Cys residues form cystines

Ext. coefficient 40800
Abs 0.1% (=1 g/l) 0.614, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is D (Asp).

The estimated half-life is: 1.1 hours (mammalian reticulocytes, in vitro).
3 min (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 40.11
This classifies the protein as unstable.

Aliphatic index: 76.14

Grand average of hydropathicity (GRAVY): -0.475



What's the solution for proteins w/o tryptophan?

2nd example: no W

Number of amino acids: 236

Molecular weight: 25163.13

Theoretical pI: 5.48

Amino acid composition: CSV format

Ala (A)	24	10.2%
Arg (R)	17	7.2%
Asn (N)	6	2.5%
Asp (D)	13	5.5%
Cys (C)	1	0.4%
Gln (Q)	12	5.1%
Glu (E)	12	5.1%
Gly (G)	29	12.3%
His (H)	2	0.8%
Ile (I)	10	4.2%
Leu (L)	26	11.0%
Lys (K)	5	2.1%
Met (M)	10	4.2%
Phe (F)	9	3.8%
Pro (P)	7	3.0%
Ser (S)	7	3.0%
Thr (T)	14	5.9%
Trp (W)	0	0.0%
Tyr (Y)	4	1.7%
Val (V)	28	11.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 25

Total number of positively charged residues (Arg + Lys): 22

Atomic composition:

Carbon	C	1108
Hydrogen	H	1810
Nitrogen	N	314
Oxygen	O	330
Sulfur	S	11

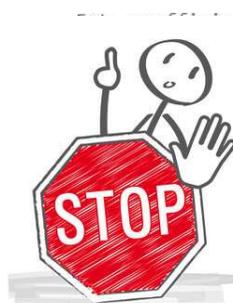
Formula: $C_{1108}H_{1810}N_{314}O_{330}S_{11}$

Total number of atoms: 3573

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.



5960
0.237, assuming all pairs of Cys residues form cystines

5960
0.237, assuming all Cys residues are reduced

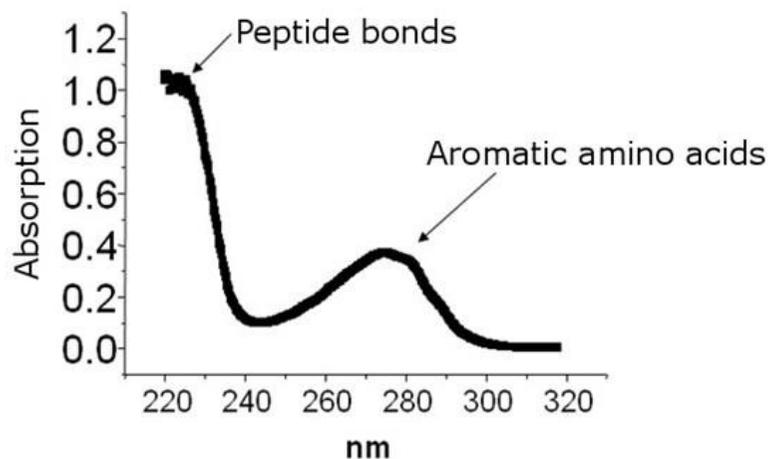
the sequence considered is Q (Gln).

life is: 0.8 hours (mammalian reticulocytes, in vitro).
10 min (yeast, in vivo).
10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 22.38
This classifies the protein as stable.

Proteins w/o W: monitoring at 205 or 214 nm



Many buffer incompatibilities!!!

Buffer and sample must be diluted 50-100 times in water to be in accepted abs range and to reduce buffer saturation.

CALCULATION OF EXTINCTION COEFFICIENTS AT 205 AND 214 NM

Sequence:

Number of S-S bonds:



Protein concentration can be determined by the absorbance at 205 or 214 nm. It is especially useful when absorbance at 280 nm cannot be used in the lack of Trp and Tyr residues. CD samples can be directly measured at these wavelengths due to the high extinction coefficients. If the spectropolarimeter is capable of converting the HT values to absorbances, then the concentrations can be determined right from the CD measurements after subtracting the baseline absorptions. Extinction coefficients at 205 and 214 nm can be calculated from the sequence.

References:

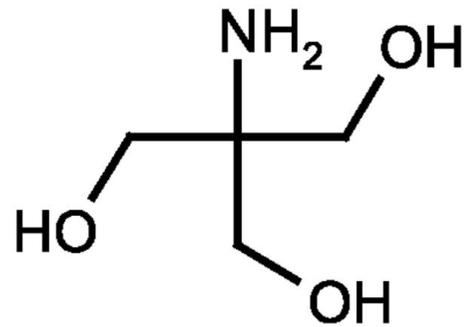
205 nm: Prot. Sci. 2013, 22, 851-858.
214 nm: J. Agric. Food Chem. 2007, 55, 5445-5451.

<https://bestsel.elte.hu/extcoeff.php>

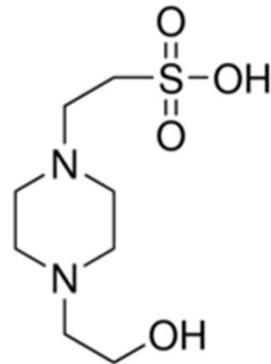
Please, type the characters above or use your password.



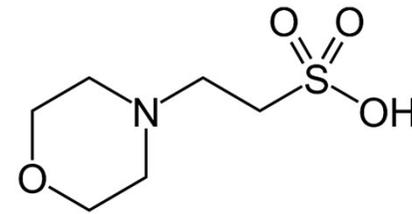
Buffer compatibility at 280 nm



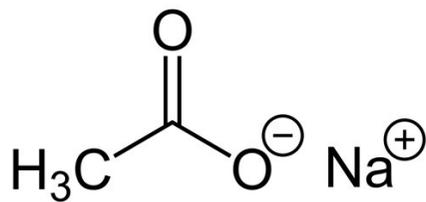
TRIS



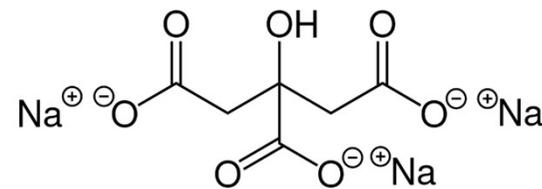
HEPES



MES



ACETATE



CITRATE

These **buffers** are compatible with the 280 nm detection

Are all molecules compatible at 280 nm?

No!

Absorbance Profile for various Reductants

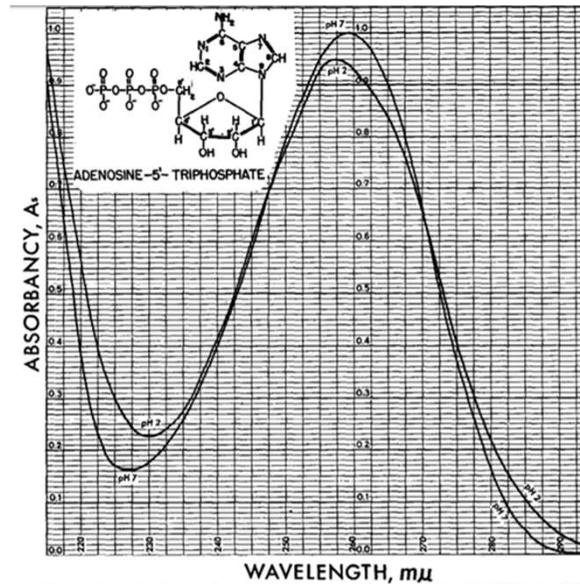
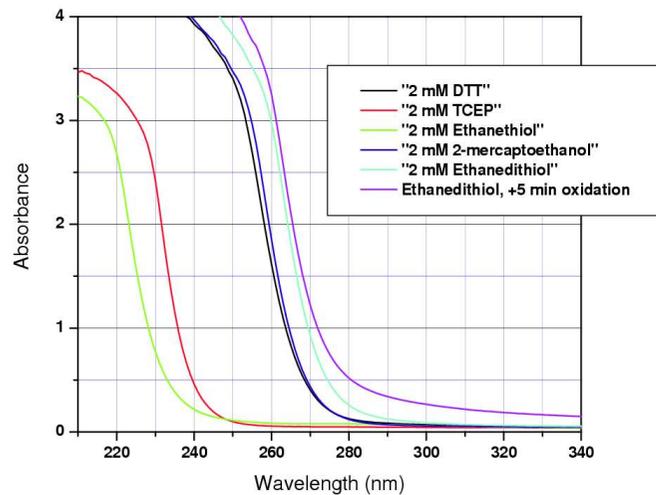


FIG. 4. Sample ultraviolet absorption curves for $6.5 \times 10^{-3} M$ ATP. At this concentration the spectrum at pH 11.3 is the same as that at pH 7 to within 0.01 unit of A_{280} .

65 μM ATP at 280 nm
= 0.15 Abs

1 mM ATP (common
concentration)
Abs= 2.3!

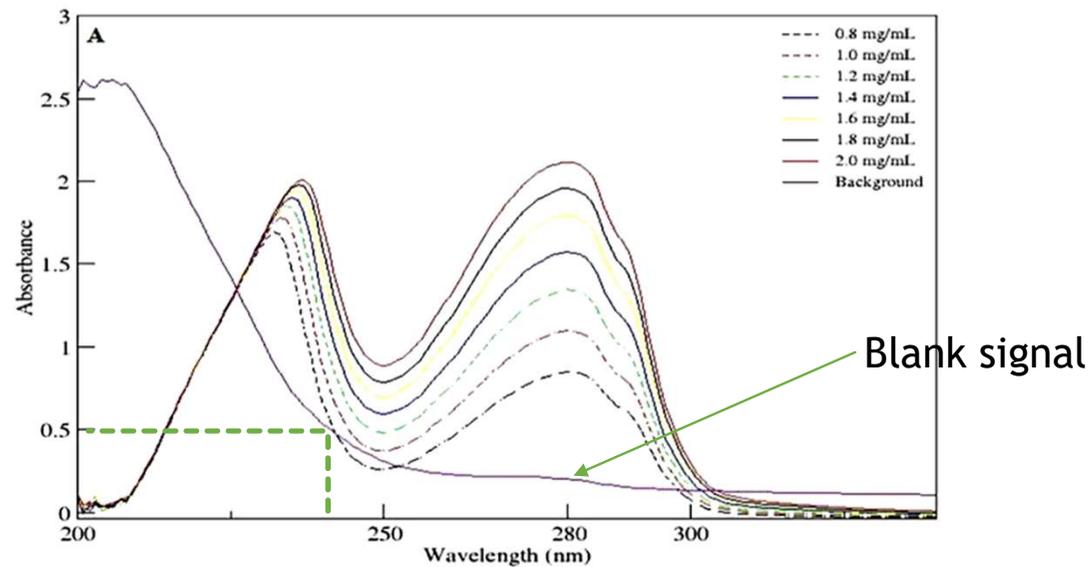
65 μM ATP at 260 nm
= 1 Abs

1 mM ATP (common
concentration)
Abs= 15.4!!!

R.M.Bock et al. (1956) Ultraviolet absorption spectra of adenosine-5'-triphosphate and related 5'-ribonucleotides. *Archives of Biochemistry and Biophysics*. Volume 62, Issue 2, June 1956, Pages 253-264

How to confirm any buffer limitations at 280 nm?

Do a blank on your buffer and visualize the transmittance signal.
or Do a blank on air and use your buffer as sample and monitor the absorbance signal.



$$A = 2 - \log \%T$$

Blank absorbance	Blank transmittance
0	100%
1	10%
2	1%
3	0.1%
4	0.01%
5	0.001%

=> Get more than 30% of transmittance signal or less than 0.5 Abs to avoid saturation

Few things to do and check before a 280 nm measurement



ϵ

Go on Protparam or Bestsel to get extinction coefficient and molecular mass.

Molecular mass must be confirmed by MS

W



Presence of Tryptophan and buffer compatibility?

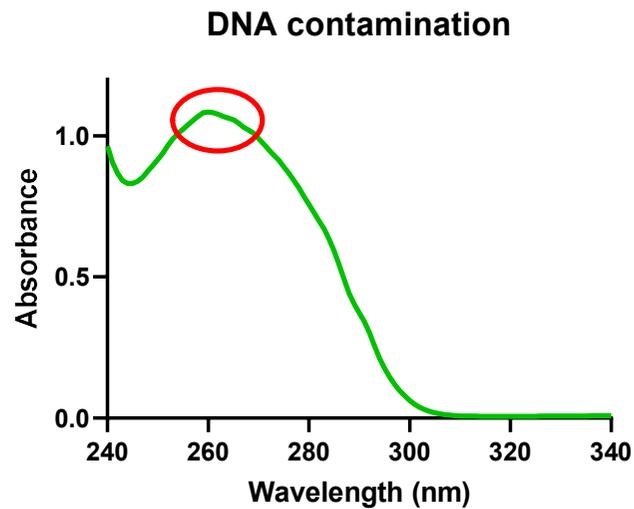




Do you know the difference between a 280 nm measurement and a UV spectrum?

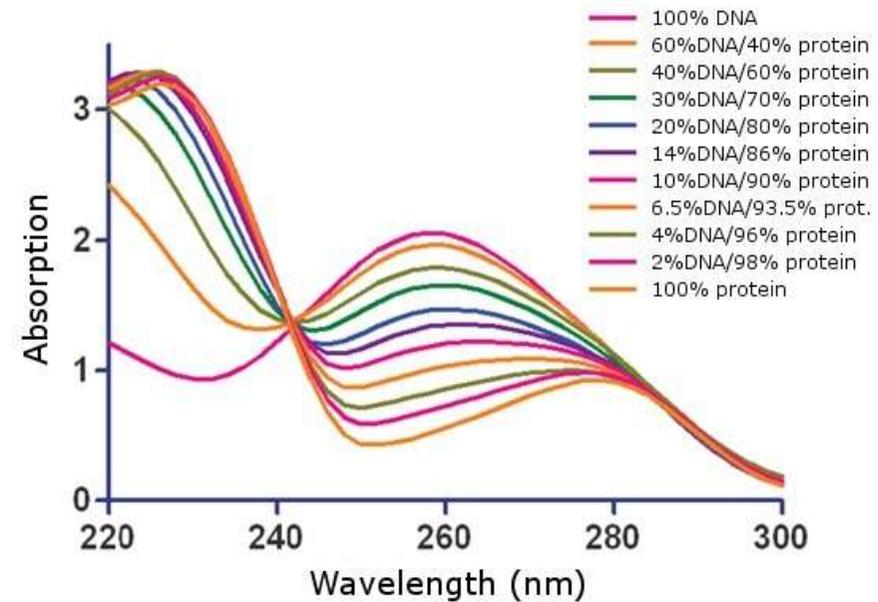


Advantages of the full UV spectrum

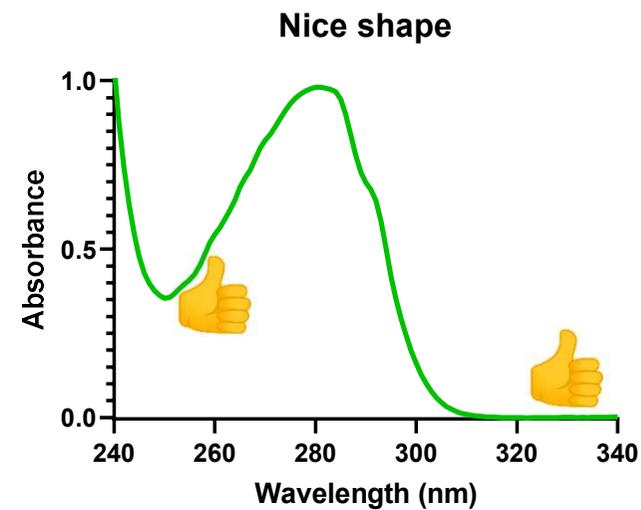
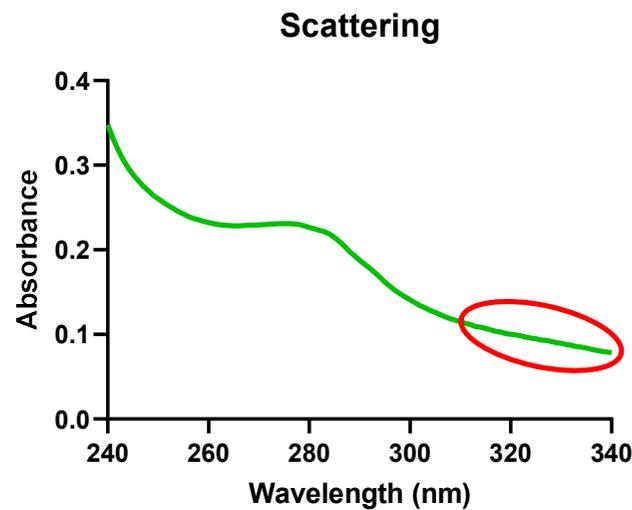


$$(\text{Abs } 260\text{nm}/\text{Abs } 280\text{nm}) < \mathbf{0.6}$$

Glasel, J.A. (1995) Validity of Nucleic Acid Purities Monitored by A260/A280 Absorbance Ratios. *Biotechniques*, 18, 62-63.



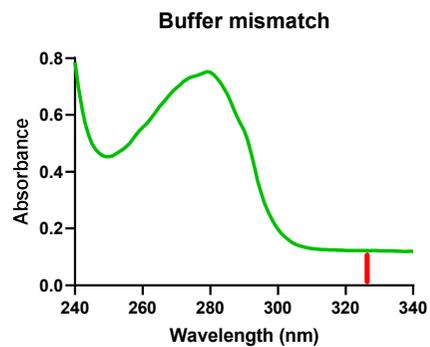
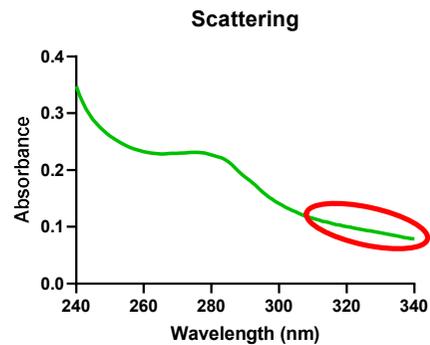
Advantages of the full UV spectrum



Aggregation Index (AI)
 $= 100 \times \text{Abs } 340 / (\text{Abs } 280 - \text{Abs } 340) < 2$

In presence of scattering, sample must be spun.

Well quantify in presence of scattering or buffer mismatch



Spectragryph

Software for optical spectroscopy
Version 1.2.15



© 2016-2020, developed by:

Dr. Friedrich Menges
Am Dummelsmoos 28
87561 Oberstdorf (GERMANY)

Email: info@effemm2.de

web site: <http://spectroscopy.ninja>

Spectragryph used to read raw data and process the data (scattering and buffer mismatch subtraction, etc.)

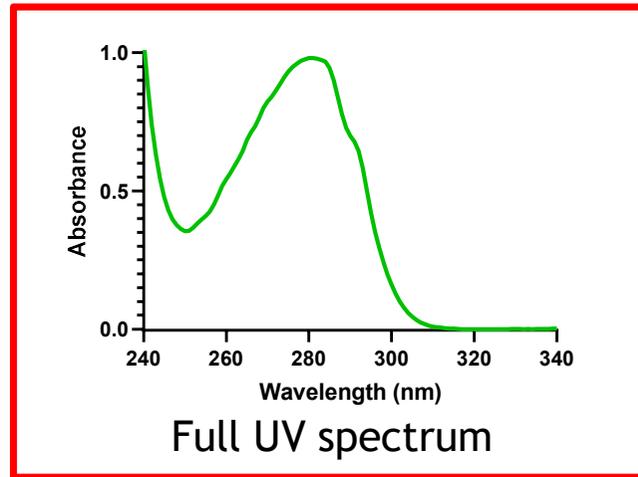
In both cases, 280 nm abs overestimated



Other approaches to quantify your protein

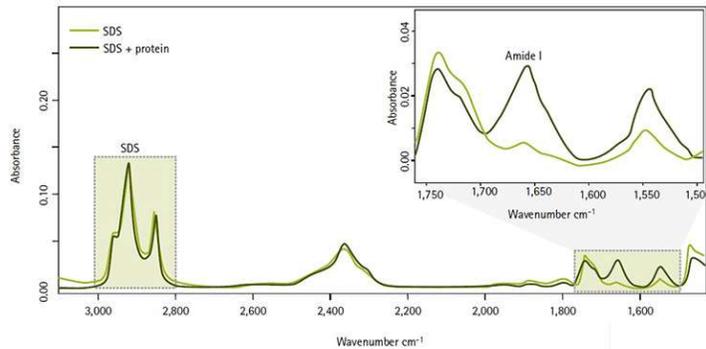


Bradford assay

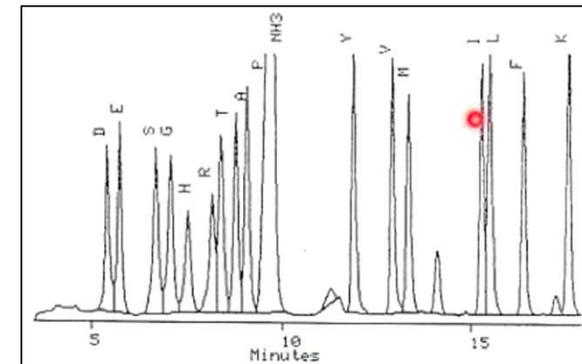


Full UV spectrum

You have to find the good technique for your system (excipient & amino acids compatibility).



Infra-red measurement



Amino acid analysis



2nd part: Introduction to Circular Dichroism

Definition:

$$\text{Circular dichroism} = \Delta A(\lambda) = A(\lambda)_{\text{LCPL}} - A(\lambda)_{\text{RCPL}}$$

Left circularly-polarized light (LCPL)
Right circularly-polarized light (RCPL)

Requirement:

CD active molecules

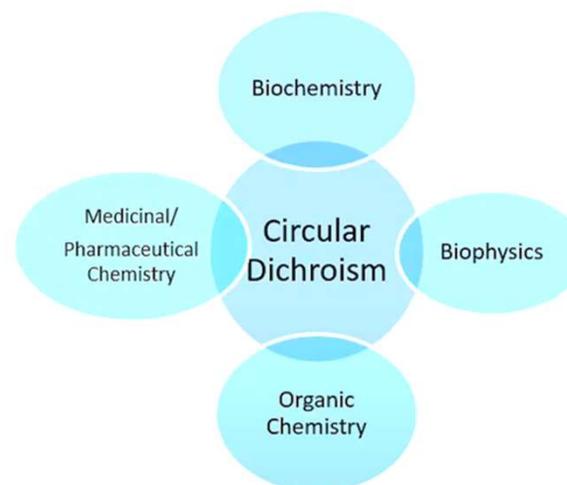
Structurally asymmetric
Exhibit absorbance

- Chiral molecules
- Non-chiral molecules attached to a chiral molecule
- Non-chiral molecule in a asymmetric environment

Applications for proteins:

Proteins

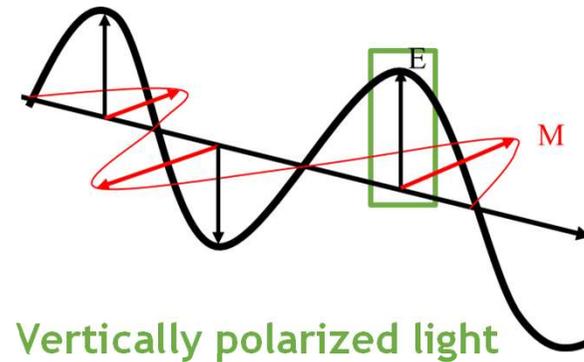
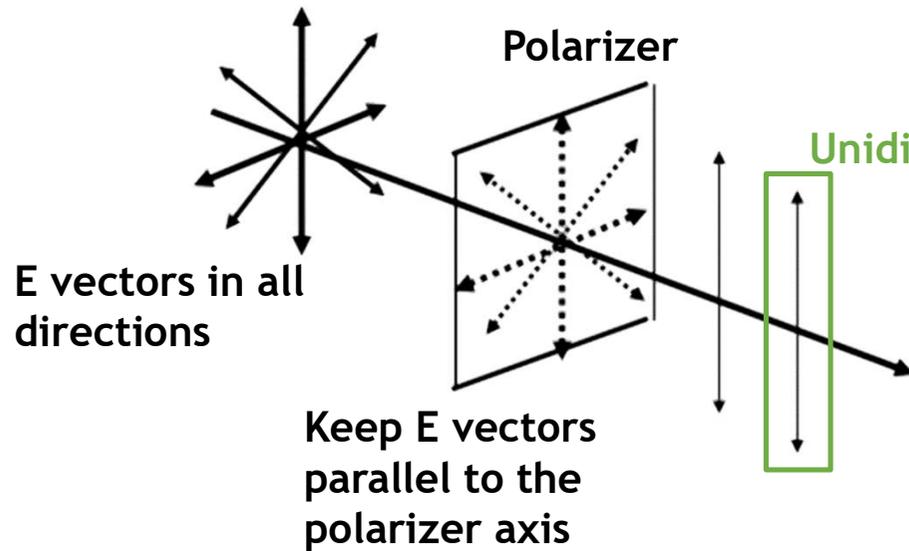
- Secondary structure analysis
- Tertiary structure fingerprint
- Detection of conformational changes
 - Structural comparison of variants (WT vs mutants)
- DNA, quadruplexes, small molecules etc.



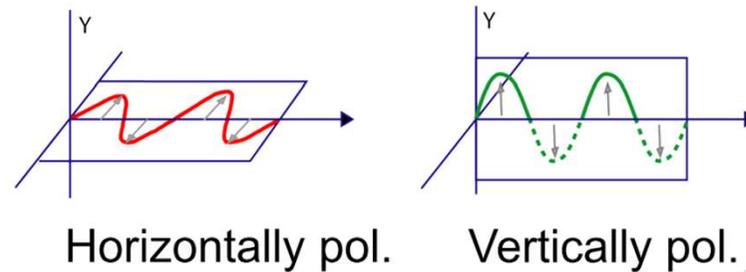
How to get circularly polarized light?

Any electronic wave consists of an **electric field** (plane of polarization) and a **magnetic field** component.

Unpolarized light

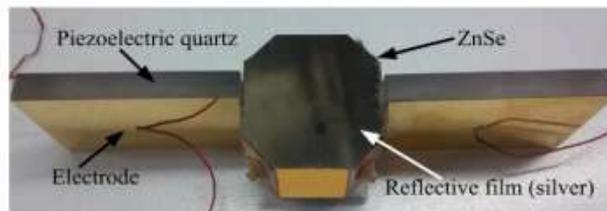


We recover both hor. and ver. polarized light...



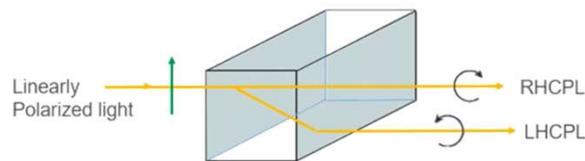
How to get circularly polarized light?

PEM converts linearly polarized light to circularly polarized light

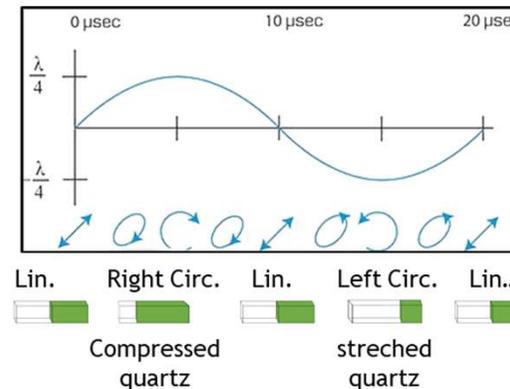


Photoelastic Modulator (PEM)

When PEM vibrates at a specific resonance frequency (50 kHz), this stress induces **birefringence** to the quartz.



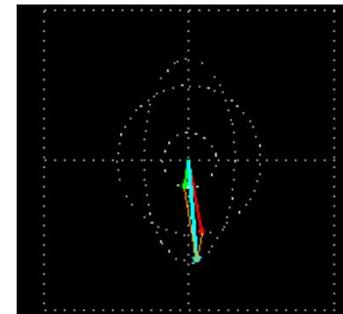
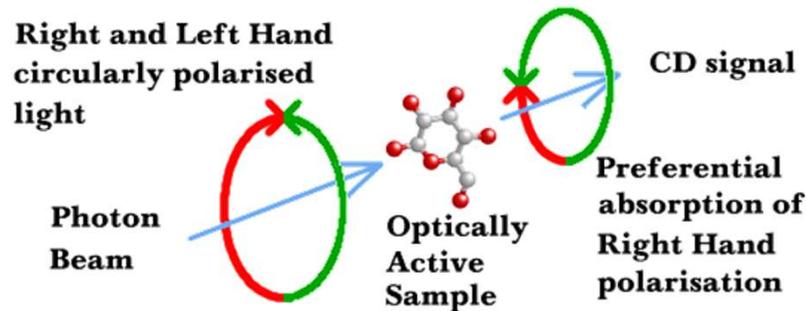
=>Under vibration, parallel and perpendicular light travels at different speed in the quartz



The modulation allows to compensate the difference in phase velocity.

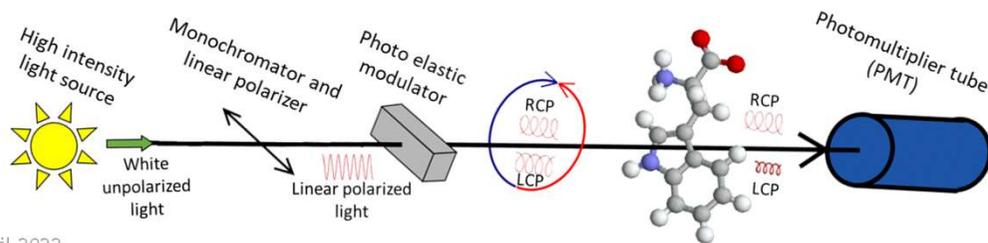
How to get circularly polarized light?

Any electronic wave consists of an electric field (plane of polarization) and a magnetic field component.



After going through a CD-active molecule → Elliptically polarized light

As the sample is optically active, there is a tiny change between RCPL and LCPL



- Right circularly-polarized component (RCPL) has been partly absorbed.
- Left circularly-polarized component (LCPL) has a larger amplitude than the right one
- Superposition of the two waves of different amplitude forms an ellipse.

Raw data in $mdeg^\circ$ of ellipticity

26

April 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004806

www.mosbri.eu

Unit used for CD measurement

From ↓ To →	Absorbance ¹	Milliabsorbance ²	Molar Extinction ³	Degrees ⁴	Millidegrees ⁵	Molar Ellipticity ⁶
(A)	A	A*1000	A*M/(C*L)	A*32.98	A*32980	A*M*3298/(L*C)
(mA)	mA/1000	mA	A*M/(C*L*1000)	mA*0.03298	mA*32.98	mA*M*3.298/(L*C)
(ε)	ε*C*L/M	ε*C*L*1000/M	ε	ε*C*L*32.98/M	ε*C*L*32980/M	ε*3298
(°)	°/32.98	°/0.03298	°*M/(C*L*32.98)	°	°*1000	°*M*100/(L*C)
(m°)	m°/32980	m°/32.98	m°*M/(C*L*32980)	m°/1000	m°	m°*M/(10*L*C)
[Θ]	[Θ]*C*L/(3298*M)	[Θ]*C*L/(3.298*M)	[Θ]/3298	[Θ]*C*L/(100*M)	[Θ]*C*L*10/M	[Θ]

¹Units are Absorbance (Abs)
²Units are milliabsorbance (mAbs)
³Units are A*L/mol*cm
⁴Units are degrees (°)
⁵Units are millidegrees (m°)
⁶Units are deg*cm²/dmol

C is concentration in g/L
M is average molecular weight (g/mol)
L is path length of cell (cm)

1. Raw data in mdeg (angle)
2. Convert mdeg in Delta epsilon or in molar ellipticity to be independent of pathlength and concentration
3. For proteins, we usually divide Delta epsilon by number of residues.

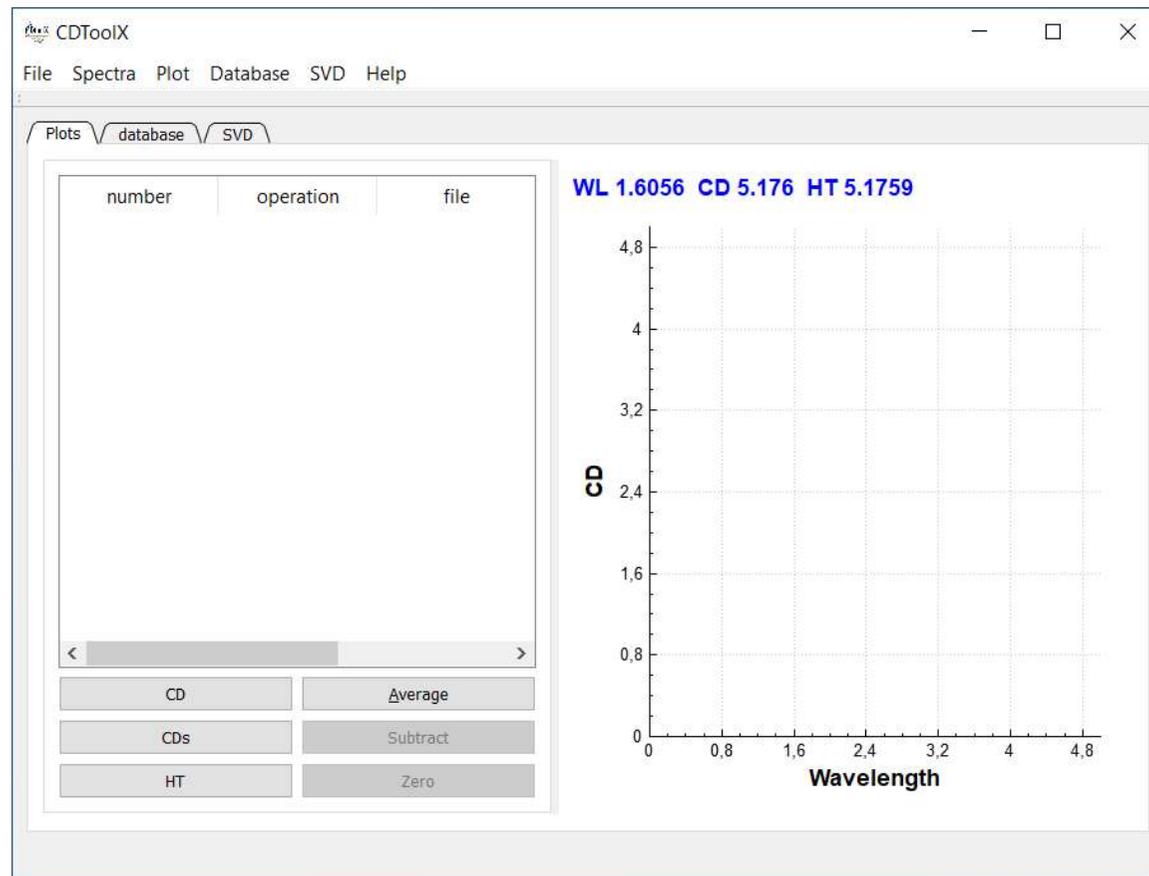
Knowing the concentration, pathlength protein sequence and purity is highly important!!!

Before CD, researchers must check their protein by UV spectroscopy and mass spectrometry

Data processing

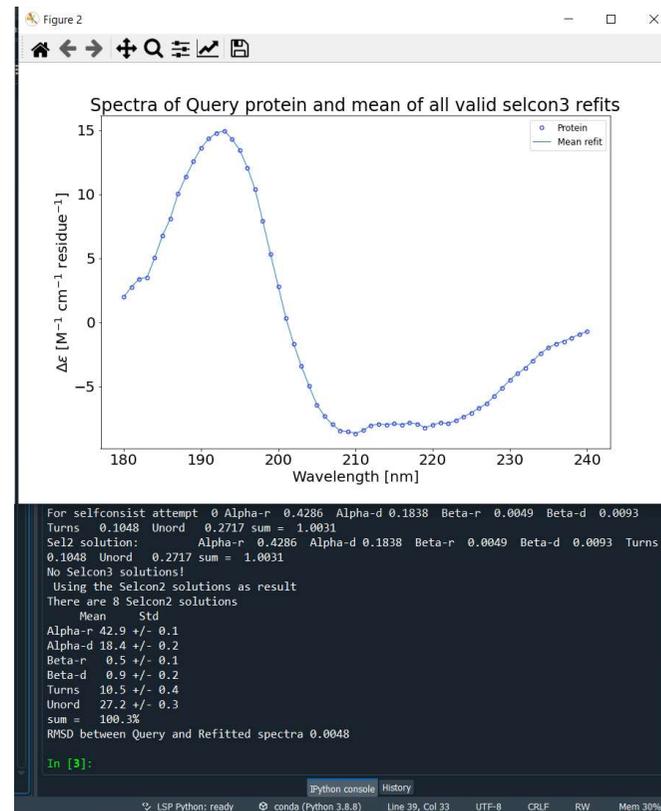
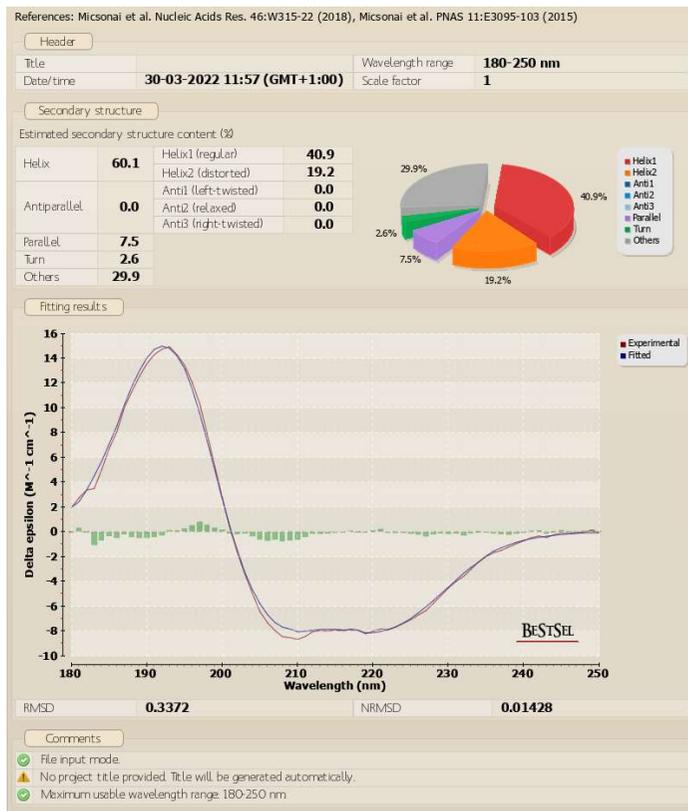
Machine softwares are usually not user friendly

Raw data are easily processed with CDToolX (sub, smoothing zero, CSA cal etc.)



Analysis softwares

Two main algorithms: BestSel and Selcon175 but other algorithms are available (See Dichroweb)



29

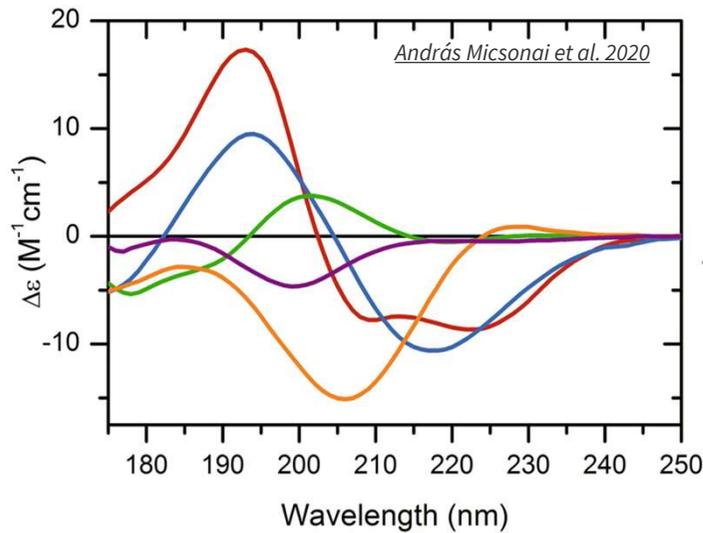
April 2022



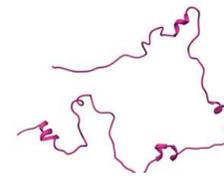
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004806

www.mosbri.eu

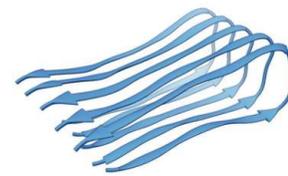
The peptide bond and secondary structure



α -helix (red)



Disordered chain (purple)



parallel β -sheet (blue)

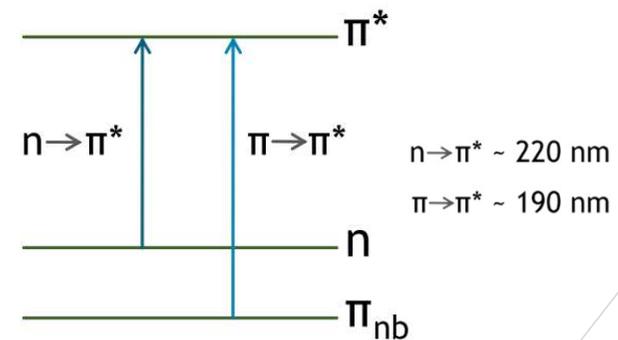
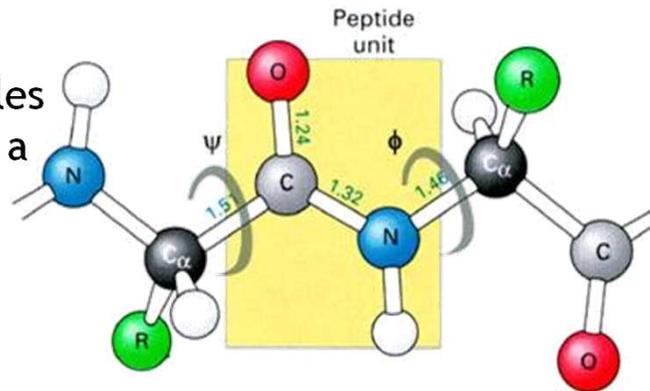


antiparallel β -sheet (green)



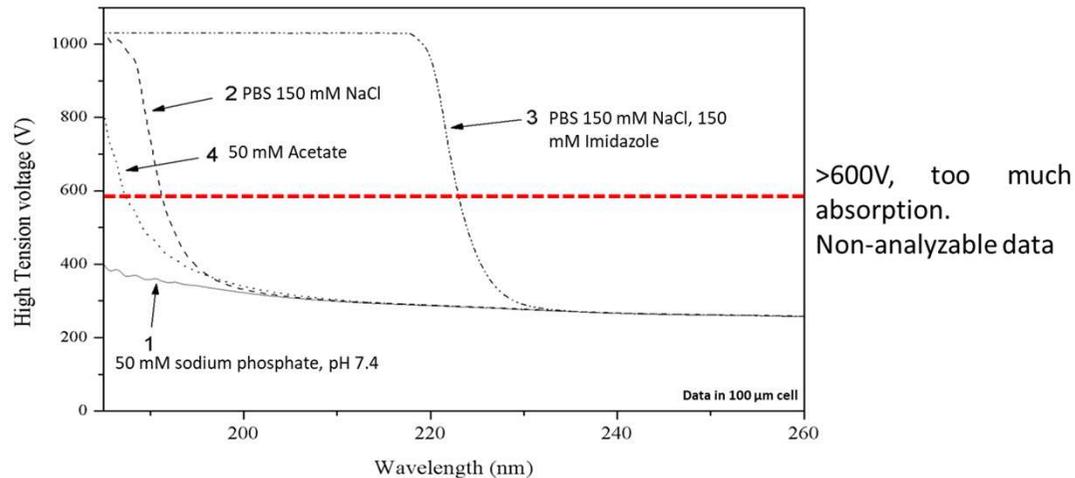
polyproline-helix (orange)

Phi (φ) and Psi (ψ) angles define the structure of a protein's backbone.



In CD, buffer limitations are more important than in UV!

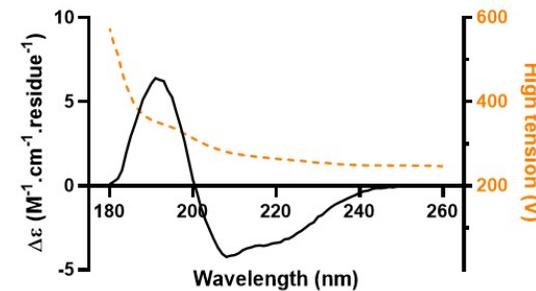
Several compounds like chloride, DTT, Imidazole, DMSO, TRIS, HEPES, etc. strongly absorb in the far-UV range



Working with high concentration (1-10 mg/ml) in low pathlength cuvettes (<100 μ m) reduces the contribution of the buffer signal in the far-UV range.

High Tension voltage (HT) is used as a gain to keep the DC of 1V in the PMT detector
HT is monitored along the spectrum to valid the CD signal just acquire

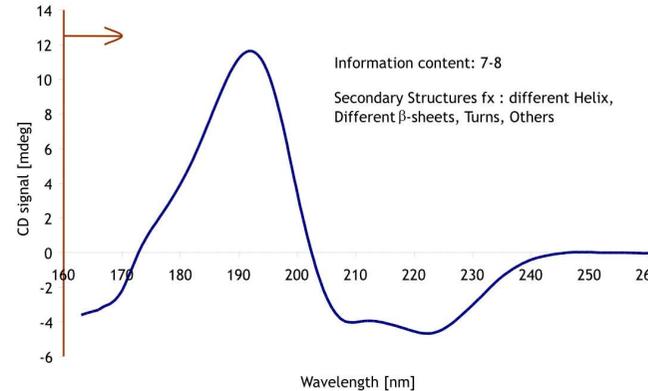
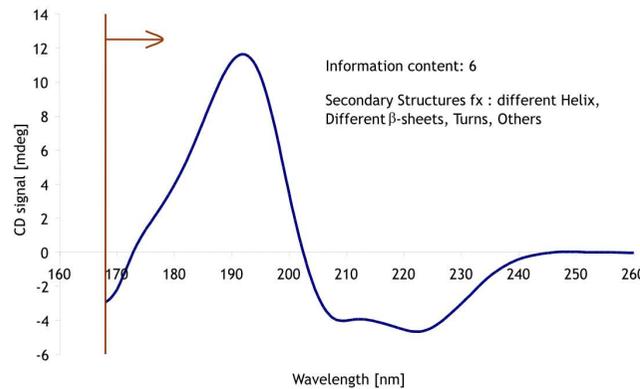
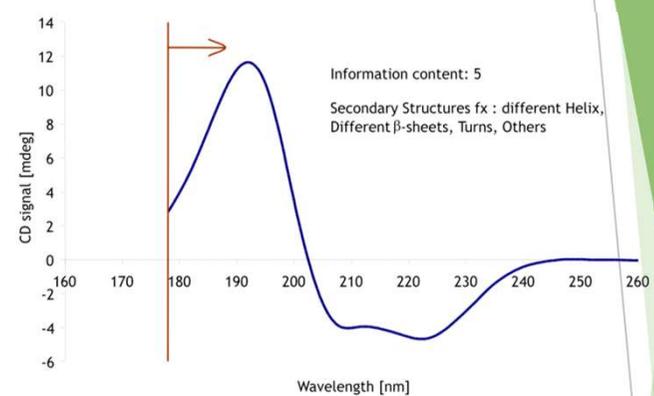
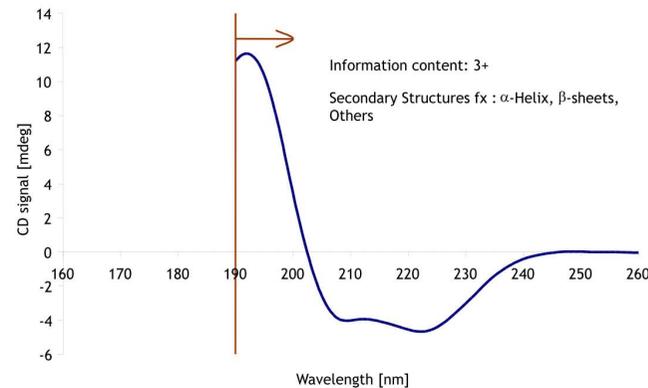
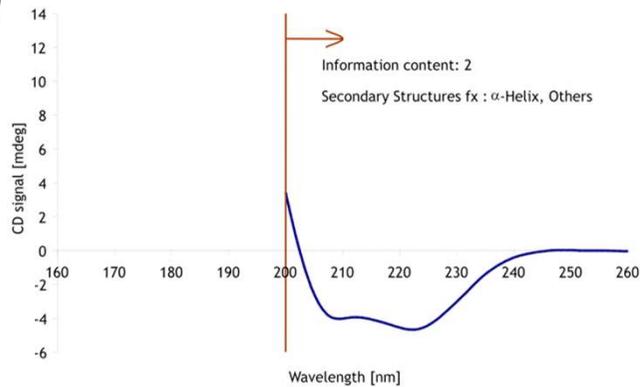
Protein in 25 mM TRIS pH 8 500 mM NaCl, 50% Glycerol



Protein at 10 mg/ml in a 10 μ m cell.
⇒ With small pathlength, it is possible to go down to 180 nm in 25mM TRIS pH8 500 mM NaCl and 50% glycerol
Against 195 nm for 150 mM NaCl in 100 μ m!

What's the good range of analysis?

Basically, more you go down in FAR-UV, better it is for the analysis

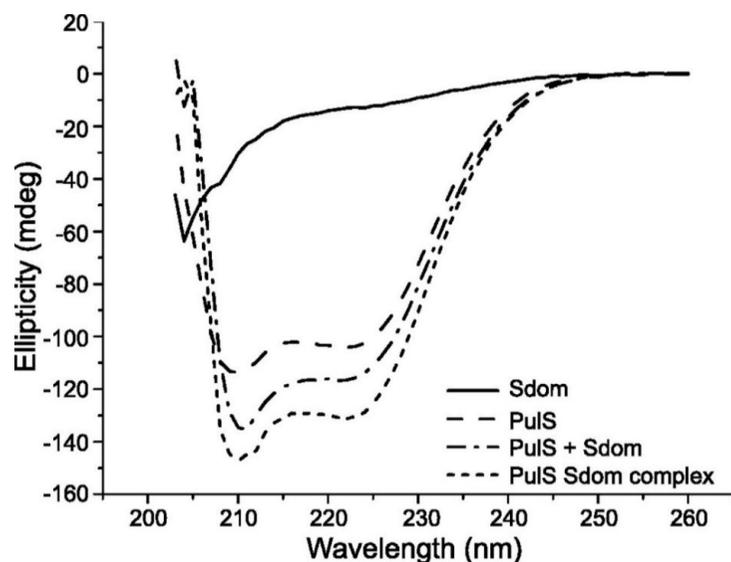


Only possible with a
synchrotron radiation circular
dichroism

Søren Vrønning Hoffmann's slide

A few results obtained at the Molecular Biophysics platform

Acquisition of secondary structure upon complex formation



Dual cuvette circular dichroism analysis shows an increase in secondary structure upon PulS-Sdom complex formation

TABLE 2

Secondary structure of PulS, Sdom, and the PulS-Sdom complex

Sample	Relative secondary structure ^a				T_m^b	ΔH^b
	α	β	Turn	Unordered		
PulS	84	0	7	8	77.3 ± 0.1	60.8 ± 1.0
Sdom	17	19	19	46	NA ^c	NA
PulS + Sdom ^d	54	5	12	30	NA	NA
Complex	63	3	11	24	80.6 ± 0.1	70.2 ± 1.0

^a Relative secondary structure was estimated from the far-UV CD spectra using the CDSSTR routine of the DICHROWEB server run on the SP175 reference dataset, see "Experimental Procedures."

^b The thermal denaturation profiles were analyzed by a nonlinear least squares fit assuming a two-state transition and were used to calculate the melting temperature (T_m) and enthalpy of unfolding (ΔH); see under "Experimental Procedures."

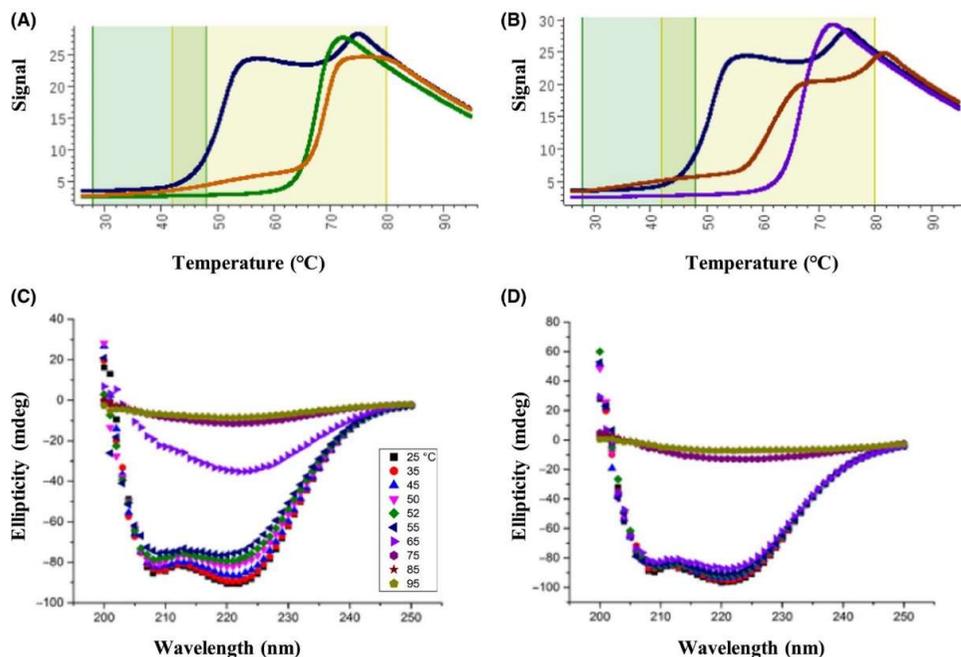
^c NA means not applicable.

^d Theoretical spectrum of an equimolar mixture of PulS + Sdom was calculated as a mean of each individual spectrum weighted according to their masses and the number of each amino acid in each protein and deconvoluted to give the predicted secondary structure.

Nickerson et al. J Biol Chem 2011

A few results obtained at the Molecular Biophysics platform

Comparison of thermal denaturation monitored by CD and TSA



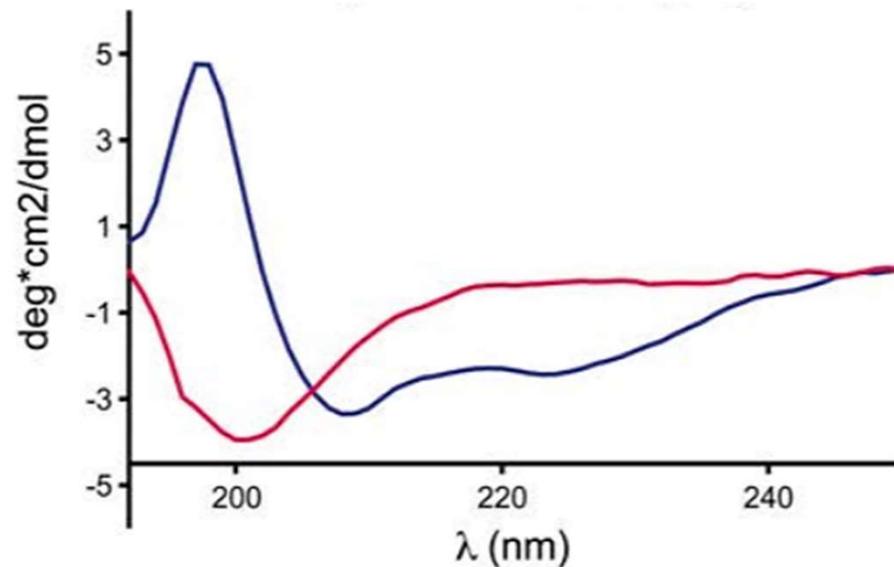
Thermal denaturation of *M. tuberculosis* UMP kinase. (A and B) TSA melting curves of *M. tuberculosis* UMP kinase. Apo enzyme (blue) in both panels. In the presence A) of GTP (1 mM, green) or UTP (1 mM, orange) B) of UMP (1 mM, brown) or Mg-ATP (7 mM, purple)

Far-UV circular dichroism spectra over a temperature range of 25–95 °C of C) the apo enzyme D) in the presence of 1 mM of Mg-ATP (D).

Walter et al. FEBS J 2022 =>Here, MgATP improves the thermal stability of UMPK both by looking at the secondary structure or tertiary environment.

A few results obtained at the Molecular Biophysics platform

Peptide stability with SUVs



SepFM peptide in solution behaves as a **random coil** and mainly folded into an **α-helix** upon interaction with SUVs, a behavior like that seen for *B. subtilis* SepF.

SepFM peptide-membrane interactions

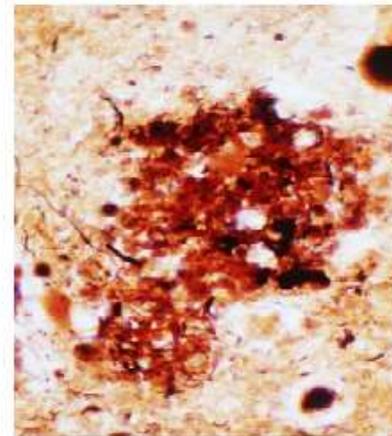
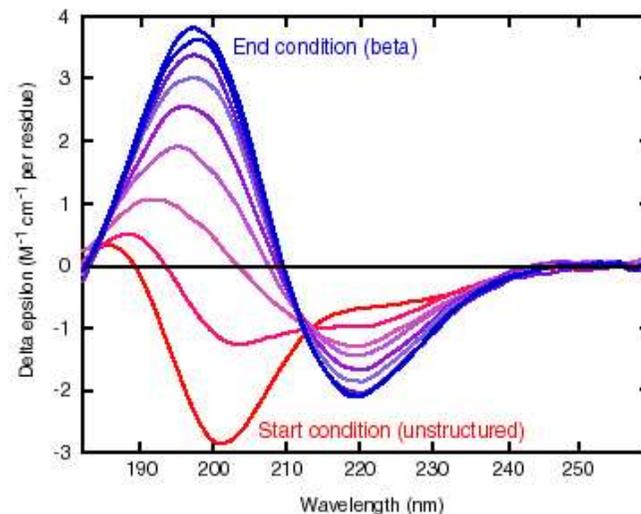
Circular dichroism spectra of SepFM in the absence (red) and presence (blue) of SUVs.

Sogues et al. *Nat. Commun.* 2020

Use of SRCD facility for amyloid diseases

A number of diseases (e.g. Alzheimer's, CJD, BSE) involve the folding of proteins and peptides into beta-sheet structures which can polymerise, forming insoluble plaques in nerve tissue (below right).

A model for the Alzheimer's peptide is LRRN, which forms spontaneously into gels with a β -sheet structure.



SRCD spectra* (left) taken during the polymerisation of LRRN peptide show that the rate of polymerisation varies with substitution of a single amino acid residue.

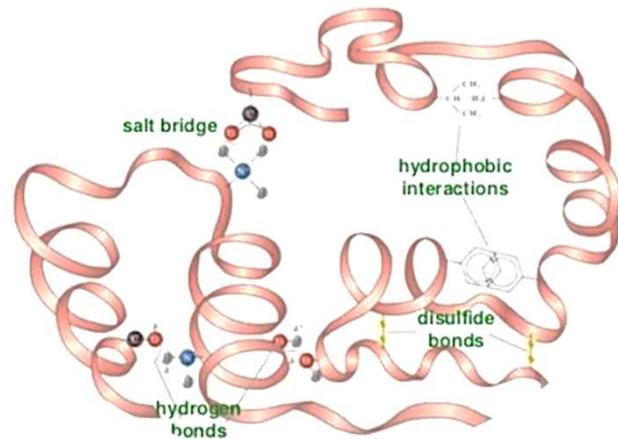
*Collaboration with N.Gay and M. Symmons, Cambridge University

The SRCD data provide important information about the processes involved in polymerisation, and may lead to the development of drugs to treat these diseases.

Alexandre Chenal' slide

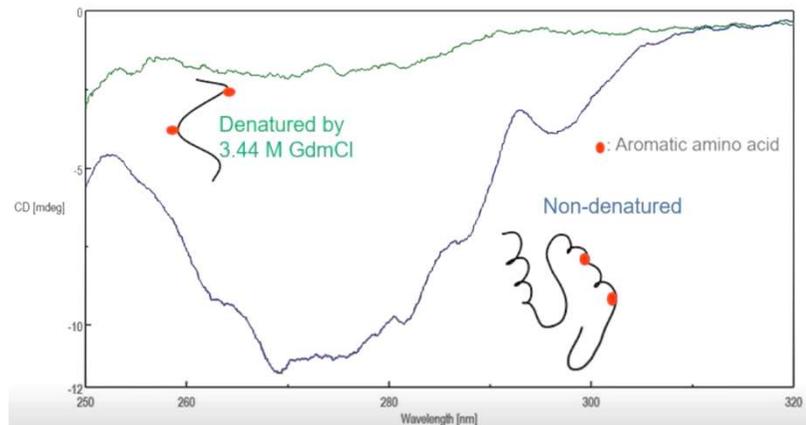
Near-UV CD spectra of protein solutions (tertiary structure)

- Mainly monitor aromatic residues environment.



Factors of influence:

- Solvent environment?
- Interactions with aromatic amino acid residues?
- Protein rigidity?



- Effect of Guanidinium chloride on the protein => loss of tertiary environment.

Why use CD?

- Simple and quick experiments
- No extensive preparation
- Measurements on solution phase
- Relatively low concentrations/amounts of sample
- Microsecond time resolution
- Any size of macromolecule



Practicalities



- CD is based on measuring a very small difference between two large signals must be done carefully
- the Abs must be reasonable max between ~0.5 and ~1.5.
- [Quartz cells](#) path lengths between 0.0001 cm and 10 cm.
- Have to be careful with buffers high UV abs
- Measure cell base line with solvent
- Then sample with same cell inserted same way around
- For accurate 2ndry structure estimation must know concentration of sample



Summary

- CD is a useful method for looking at secondary structures of proteins and peptides.
- It is an adaptation standard absorption spectroscopy in which the difference in the abs between left and right hand circularly polarized light is measured.
- CD can be measured under a wide range of conditions - e.g., good for membrane proteins.
- CD can also be used to measure tertiary structure changes.
- CD compliments other more detailed techniques such as crystallography.

