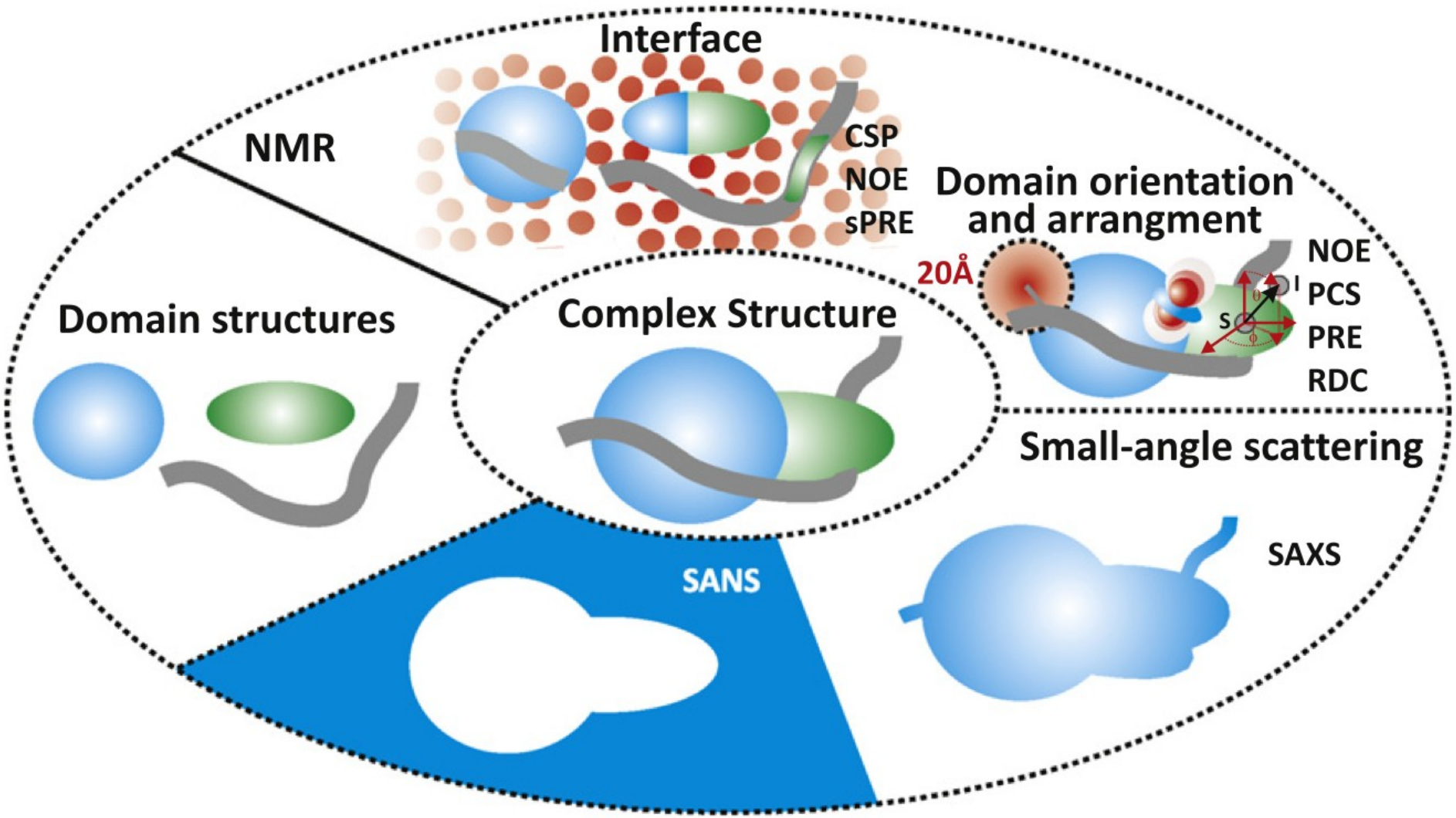
# Basics of SAXS in structural analysis (of biomolecules)

Karel Kubíček

Motivation


Interface
NMR
CSP
NOE
sPRE
Domain orientation
and arrangment
20Å
NOE
PCS
PRE
RDC
Domain structures
Complex Structure
Small-angle scattering
SANS
SAXS

Terminology

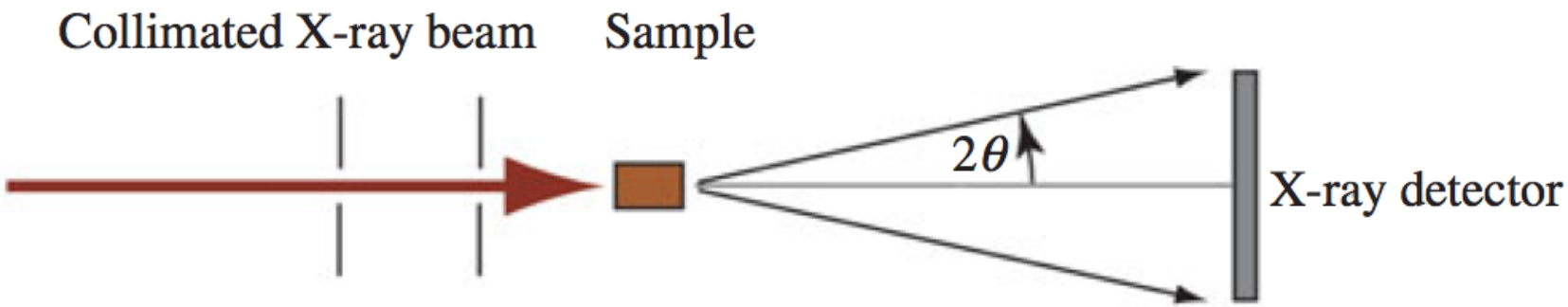
1) X-ray – diffraction / scattering
2) XS - X-ray scattering
3) SAXS/WAXS - Small/Wide Angle X-ray Scattering
4) SANS - --------”---------- Neutron ---”---

A) Otto Kratky (1902, Vienna-1995, Graz)
B) Günter Porod (1919 near Villach, 1984 Graz)
C) Dmitri I. Svergun

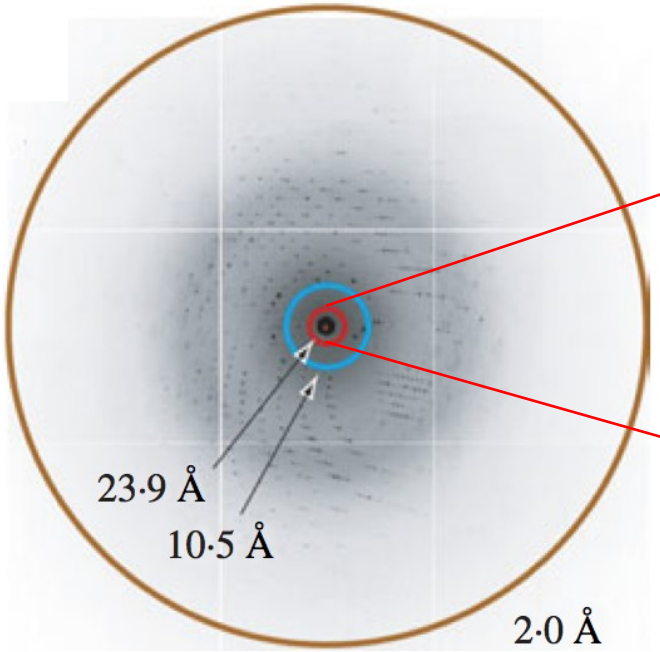
I) Scattering
II) Scattering curve
III) Guinier plot
IV) PDF (Pair-distribution function)

a) Bead model
b) Bead model - / SAXS - envelope

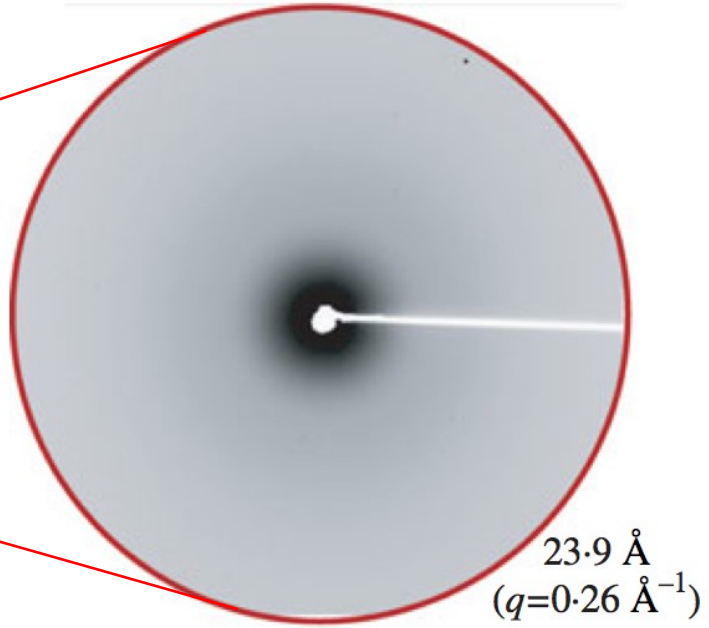
Experimental setup




X-ray

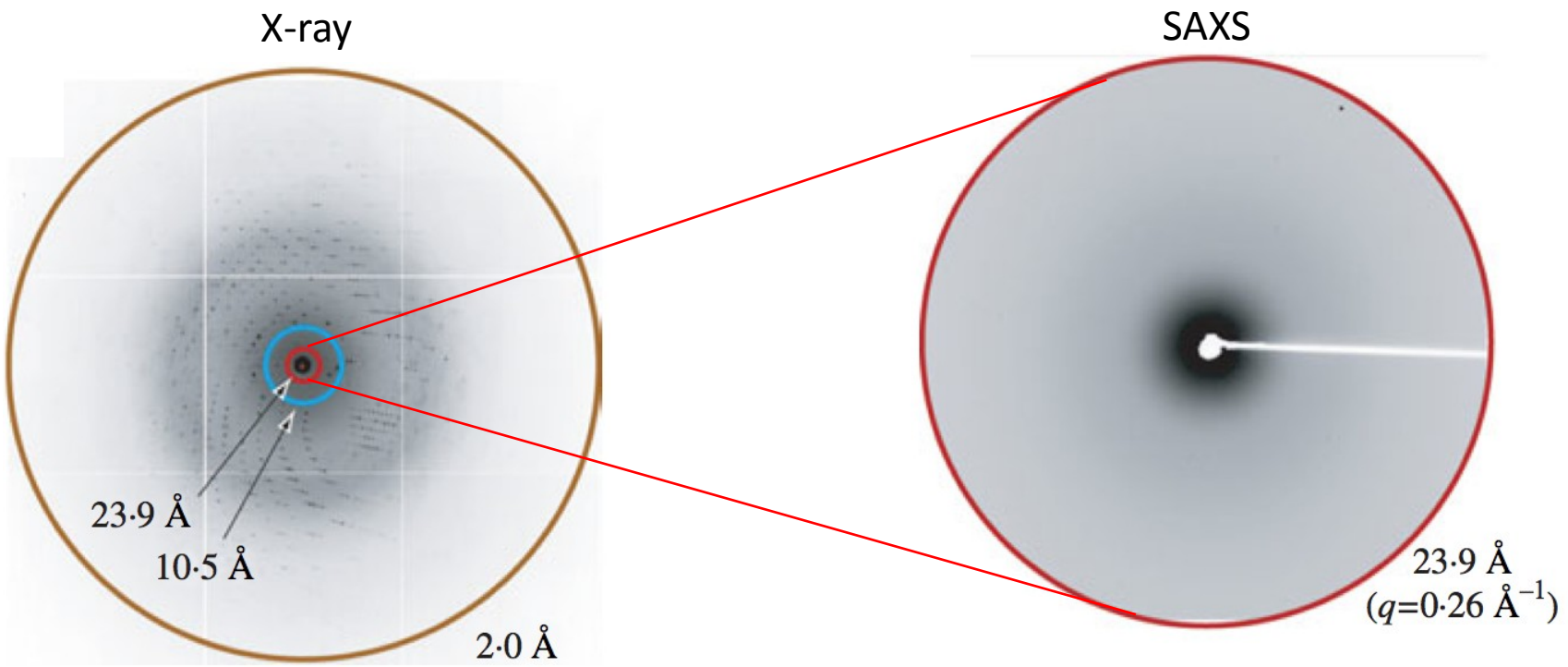



SAXS

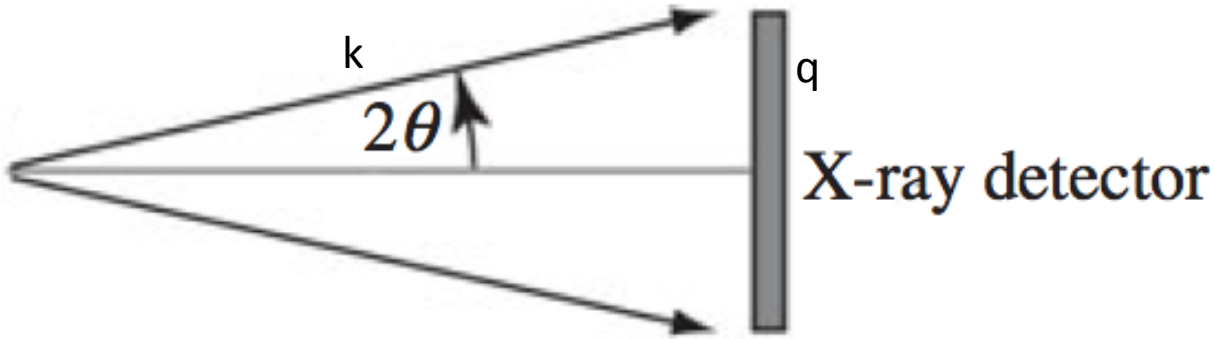


Crystalized Sample

Sample in Solution



(very tiny) bit of theory

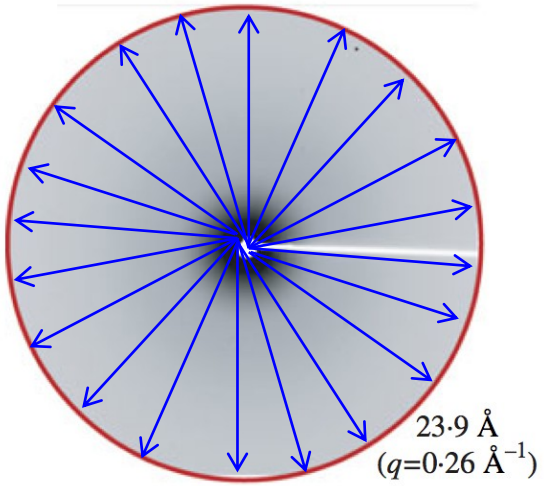



$q=2k.\sin\theta$ => $q/2/k=\sin\theta$,

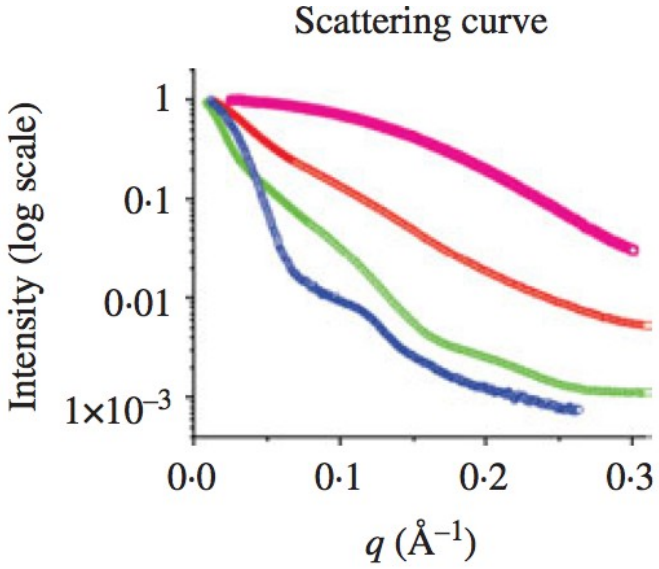
$k=2\pi/\lambda$ => **$q=4\pi/\lambda.\sin\theta$**

Often ***q*** is denoted as ***s***

1st step: scattering to scattering curve

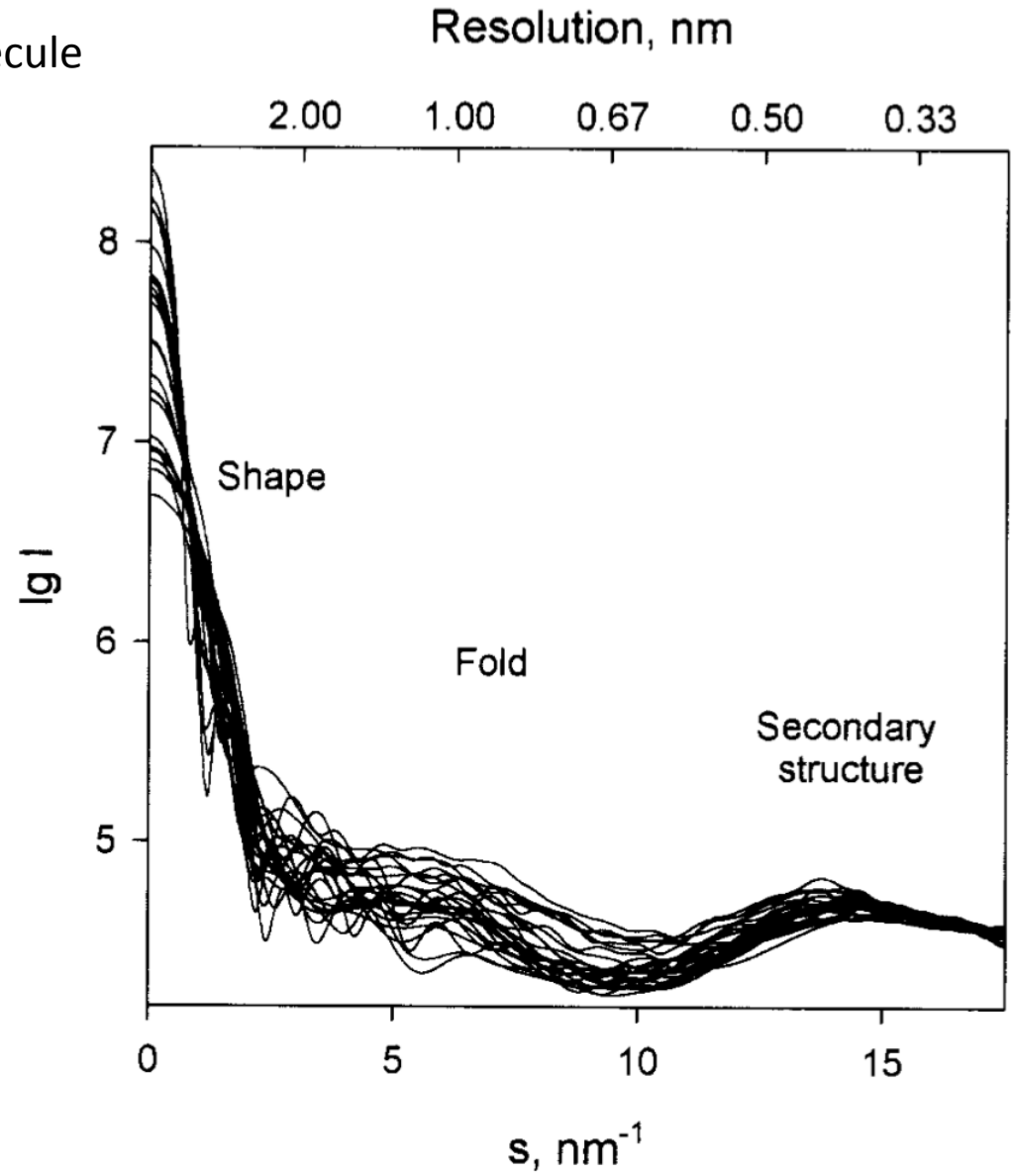



=>

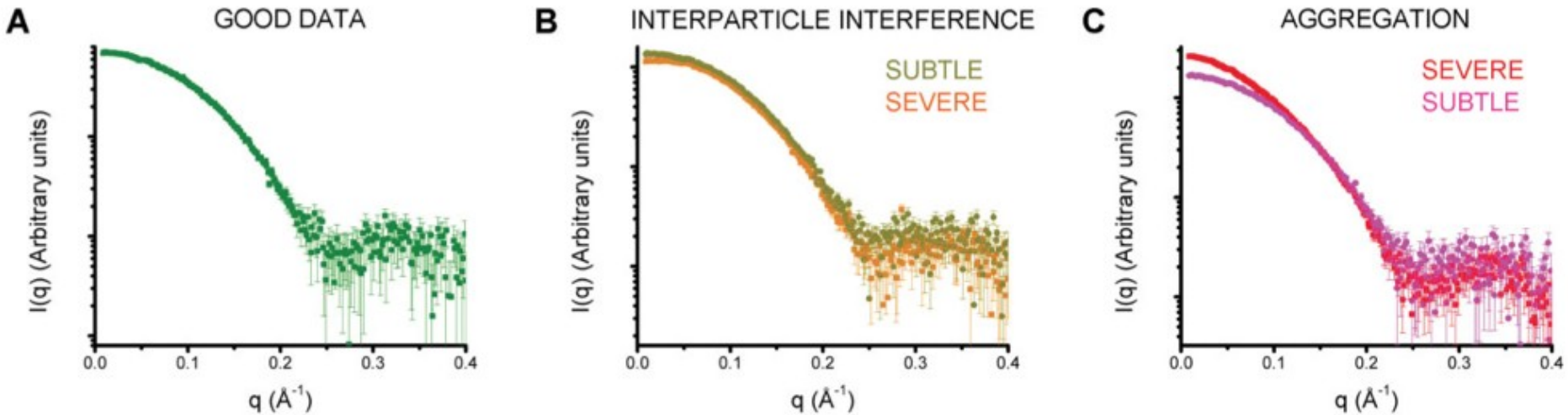


What can we learn from the scattering curve:

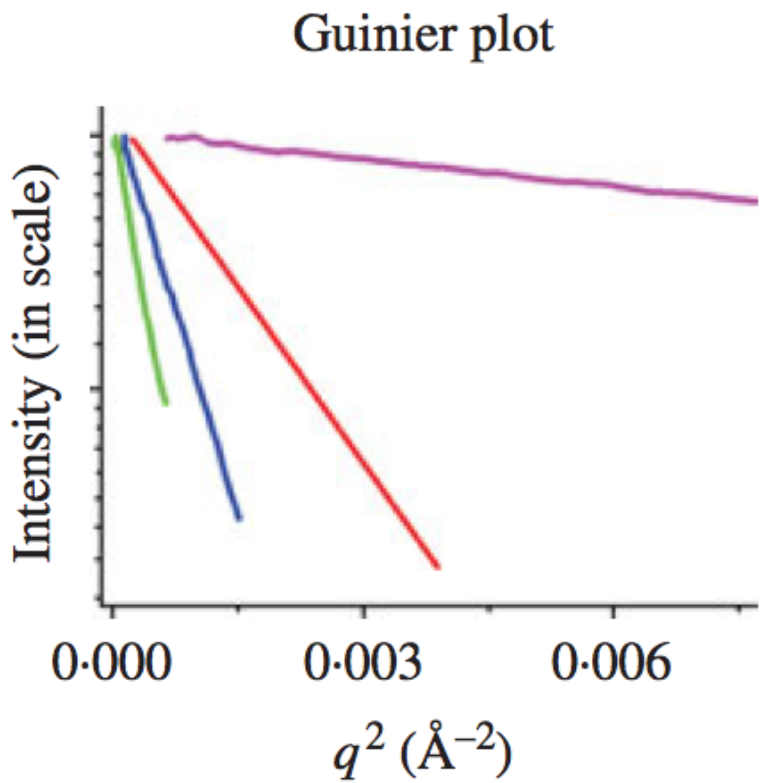
1) Shape of the studied molecule
2) Fold
3) Secondary structure



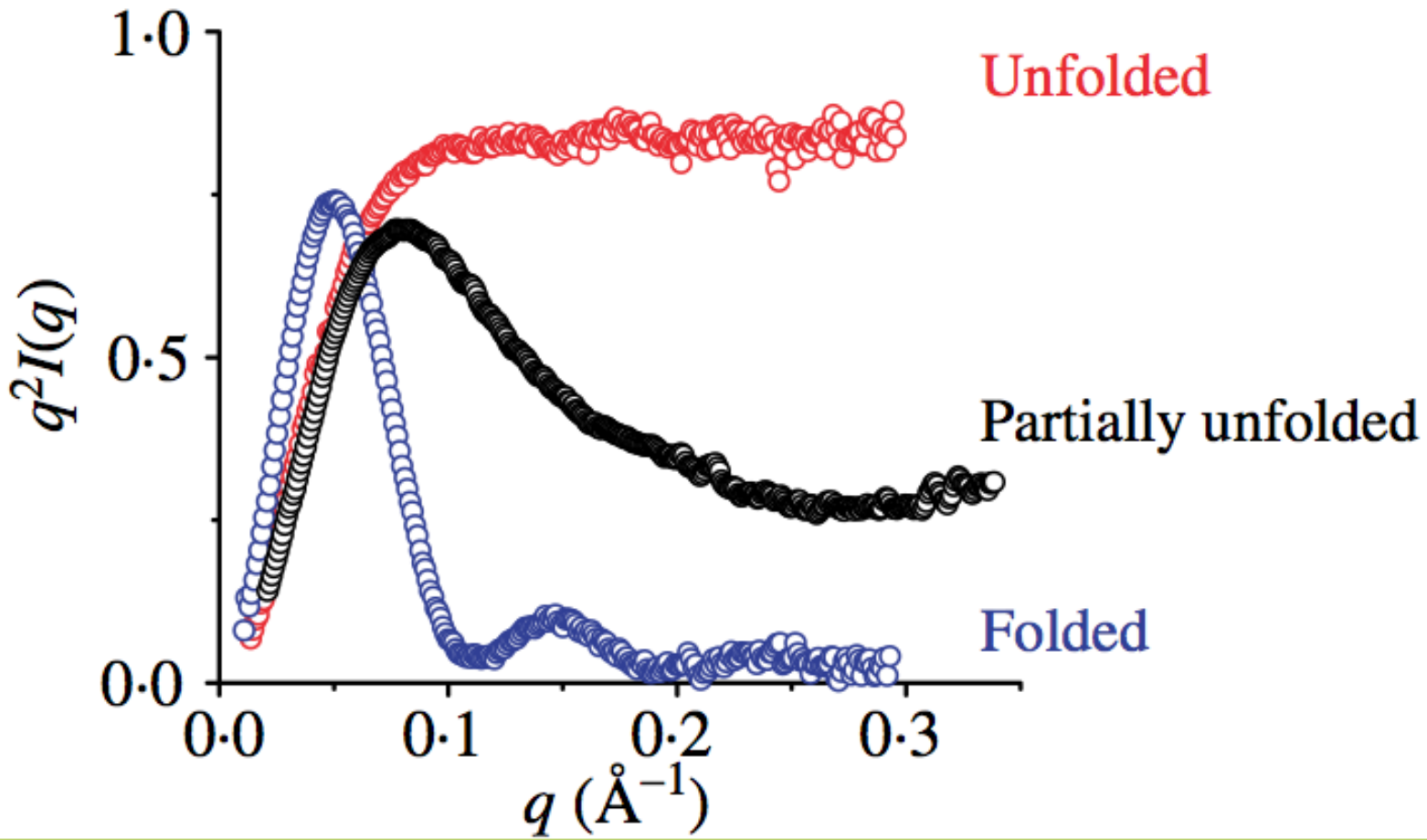
Is it really that easy?

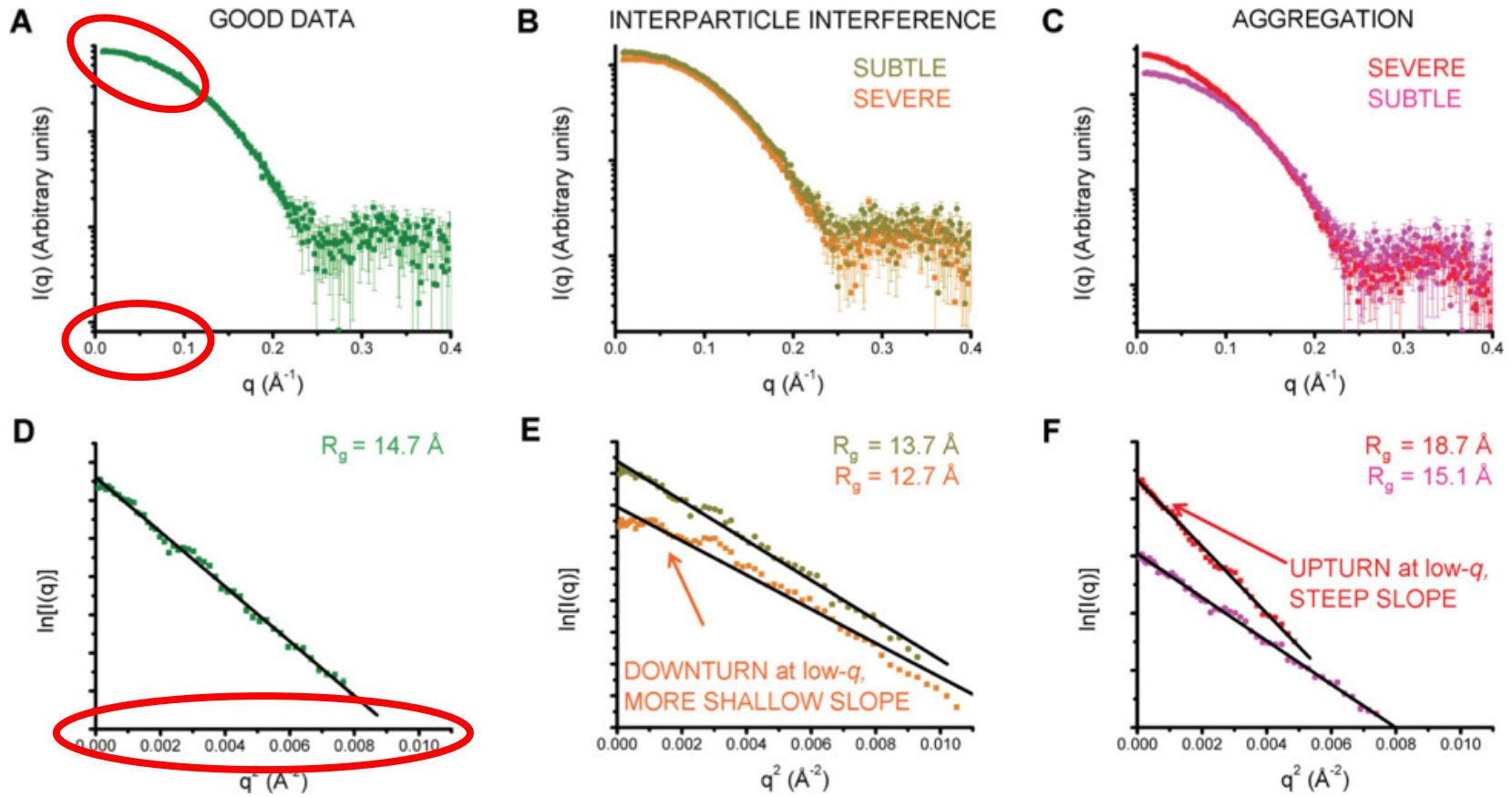


2nd step: Guinier plot & Kratky plot (from the initial region of the scattering curve)

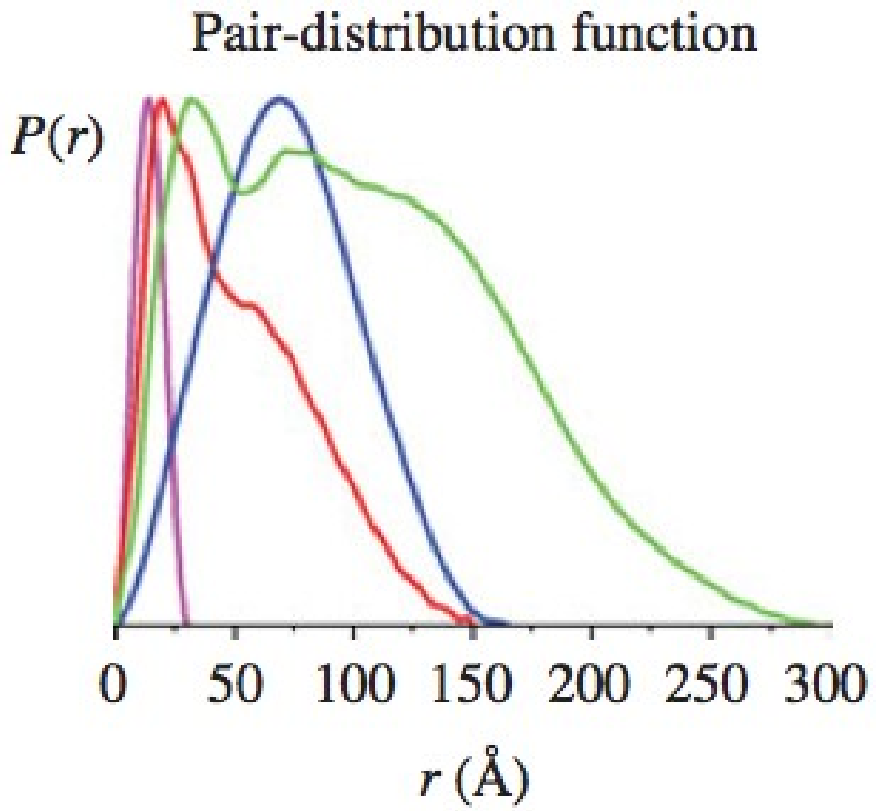


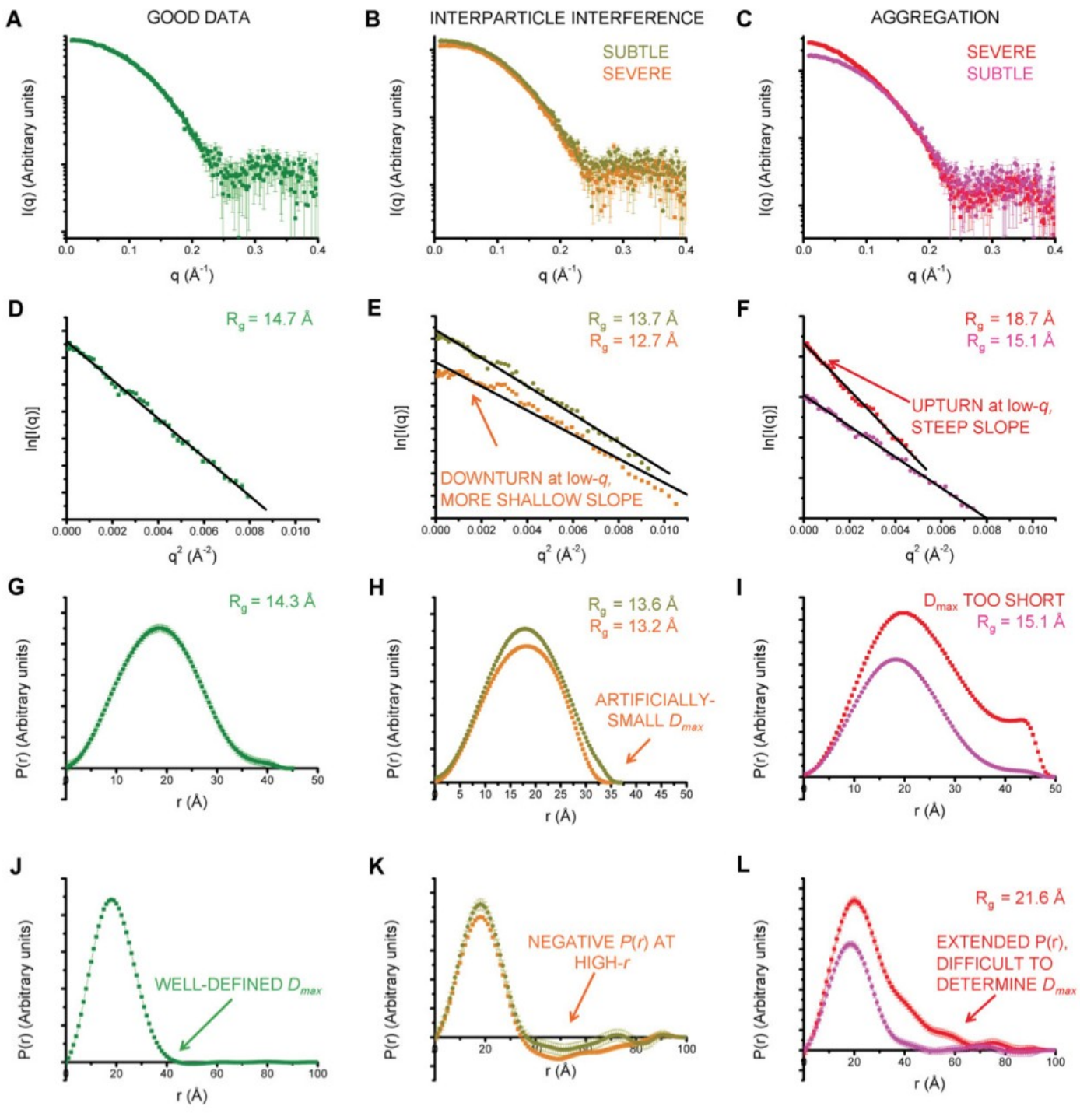
2nd step: Guinier plot & Kratky plot (from the initial region of the scattering curve)



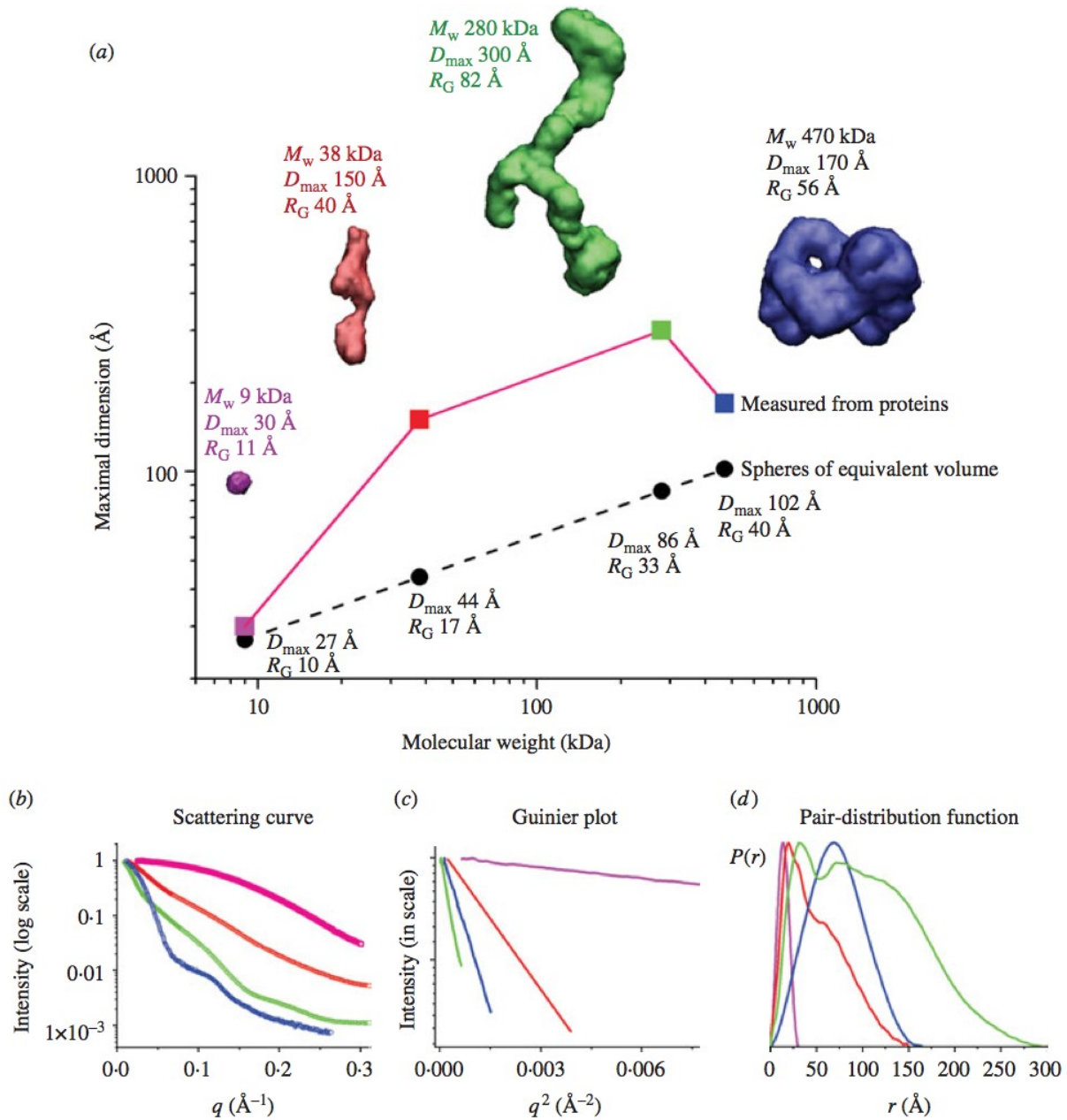

A
GOOD DATA
I(q) (Arbitrary units)
q (Å$^{-1}$)
B
INTERPARTICLE INTERFERENCE
SUBTLE
SEVERE
I(q) (Arbitrary units)
q (Å$^{-1}$)
C
AGGREGATION
SEVERE
SUBTLE
I(q) (Arbitrary units)
q (Å$^{-1}$)
D
$R_g$ = 14.7 Å
ln[I(q)]
q$^2$ (Å$^{-2}$)
E
$R_g$ = 13.7 Å
$R_g$ = 12.7 Å
DOWNTURN at low-q,
MORE SHALLOW SLOPE
ln[I(q)]
q$^2$ (Å$^{-2}$)
F
$R_g$ = 18.7 Å
$R_g$ = 15.1 Å
UPTURN at low-q,
STEEP SLOPE
ln[I(q)]
q$^2$ (Å$^{-2}$)

3rd step: PDF Pair-distribution function



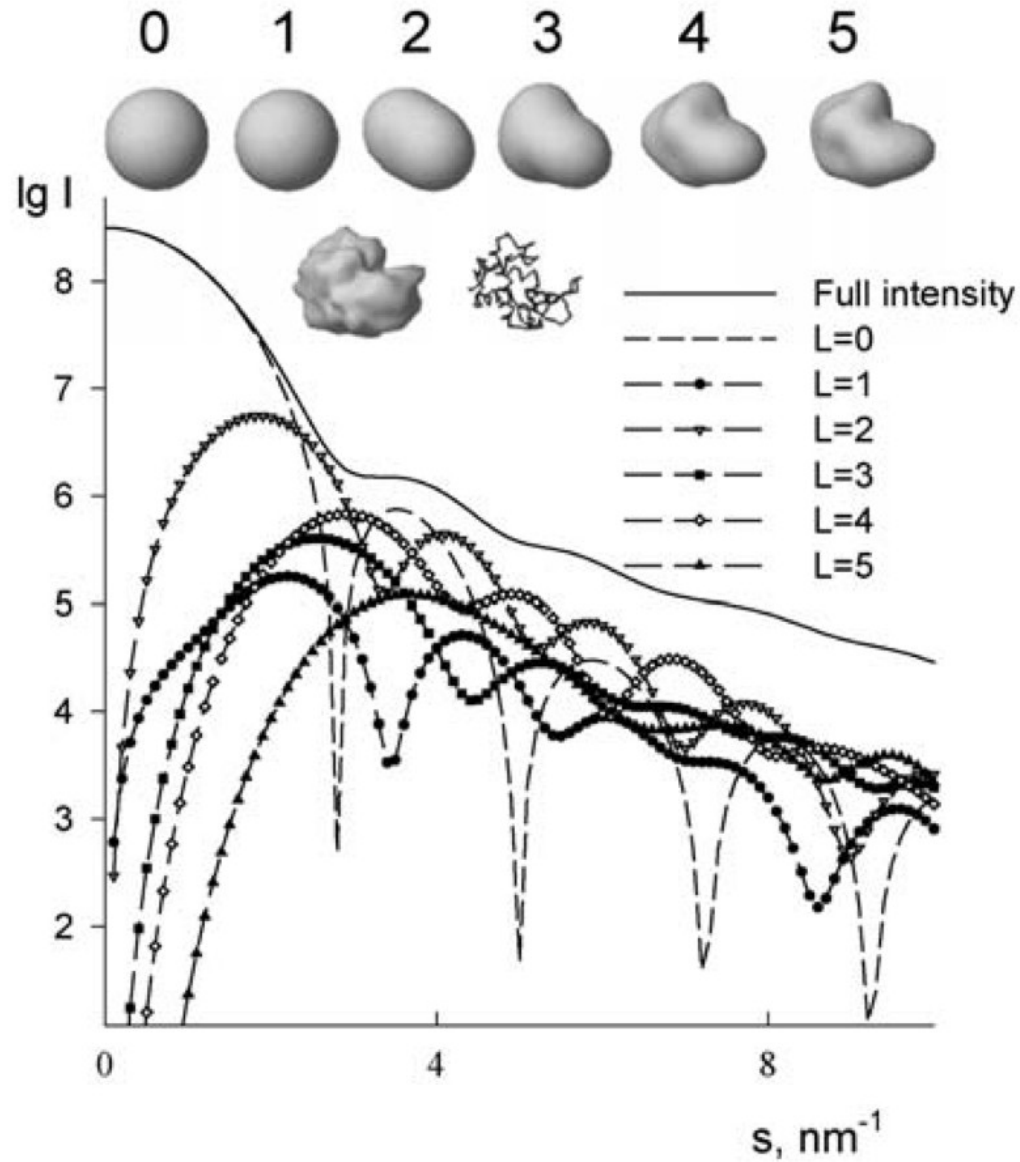

A
GOOD DATA
B
INTERPARTICLE INTERFERENCE
C
AGGREGATION
I(q) (Arbitrary units)
q (Å$^{-1}$)
SUBTLE
SEVERE
SEVERE
SUBTLE
D
E
F
ln[I(q)]
q$^2$ (Å$^{-2}$)
$R_g$ = 14.7 Å
$R_g$ = 13.7 Å
$R_g$ = 12.7 Å
$R_g$ = 18.7 Å
$R_g$ = 15.1 Å
DOWNTURN at low-$q$,
MORE SHALLOW SLOPE
UPTURN at low-$q$,
STEEP SLOPE
G
H
I
P(r) (Arbitrary units)
r (Å)
$R_g$ = 14.3 Å
$R_g$ = 13.6 Å
$R_g$ = 13.2 Å
ARTIFICIALLY-
SMALL $D_{max}$
$D_{max}$ TOO SHORT
$R_g$ = 15.1 Å
J
K
L
WELL-DEFINED $D_{max}$
NEGATIVE $P(r)$ AT
HIGH-$r$
$R_g$ = 21.6 Å
EXTENDED P(r),
DIFFICULT TO
DETERMINE $D_{max}$

All information from the scattering curve together

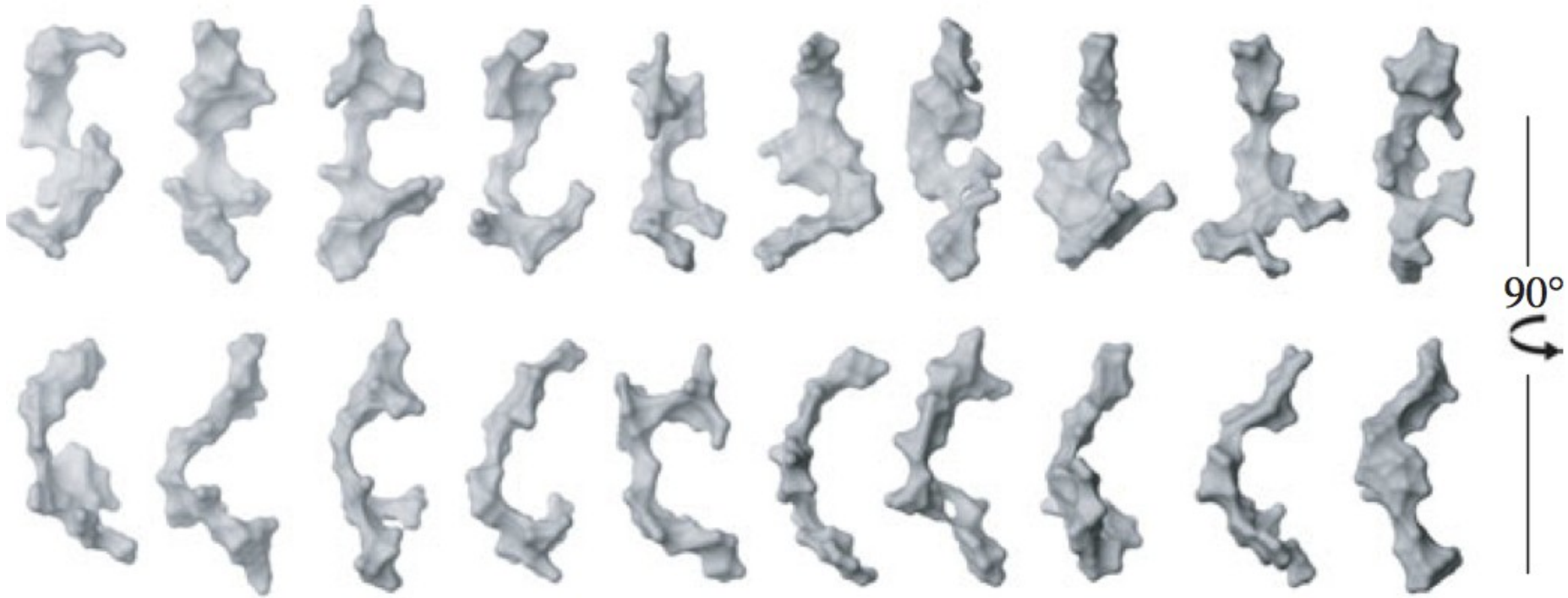


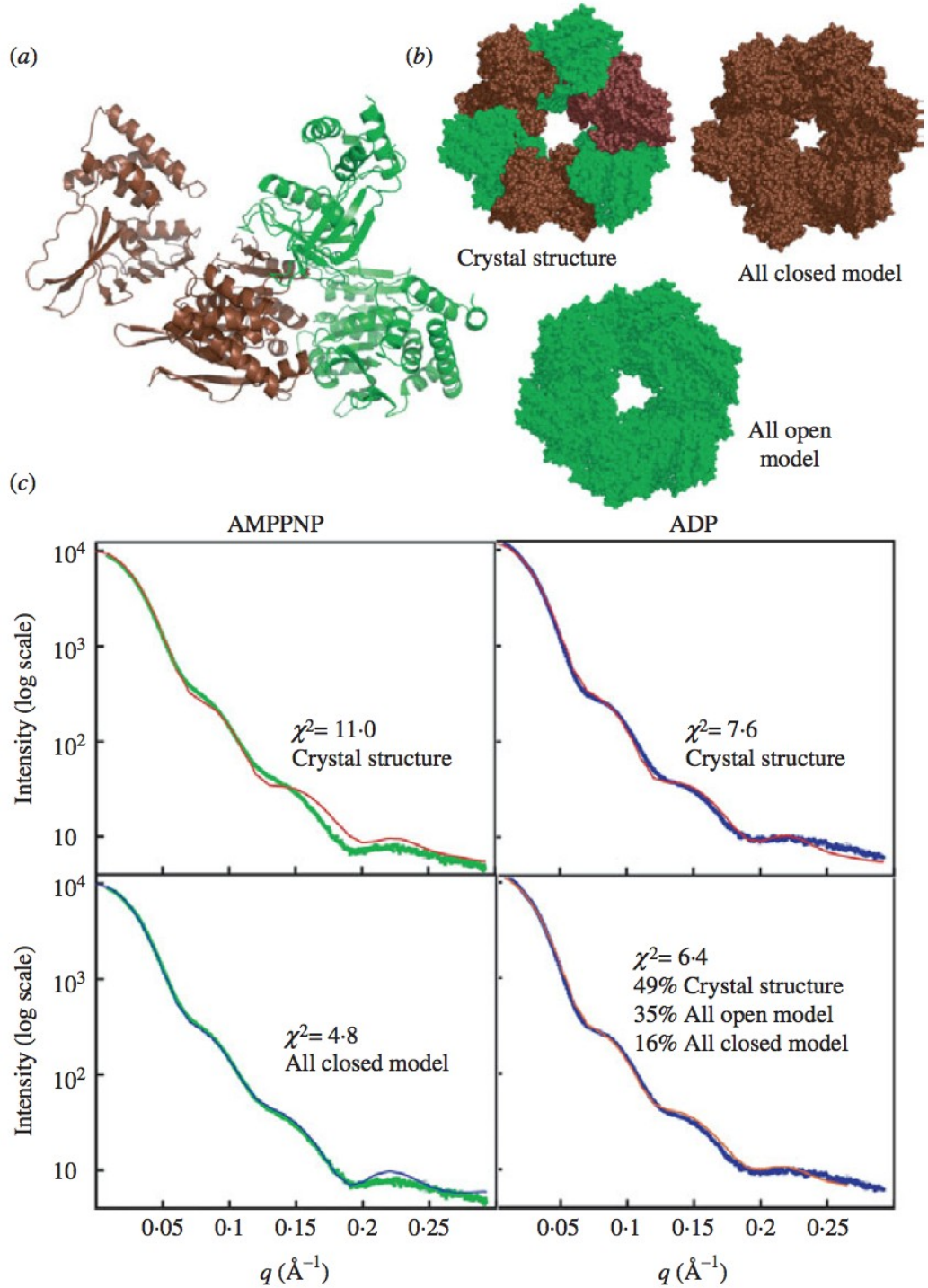
**Bead model**

Spherical harmonics problem to be solved



Solutions with similar “goodness” of fit may be obtained



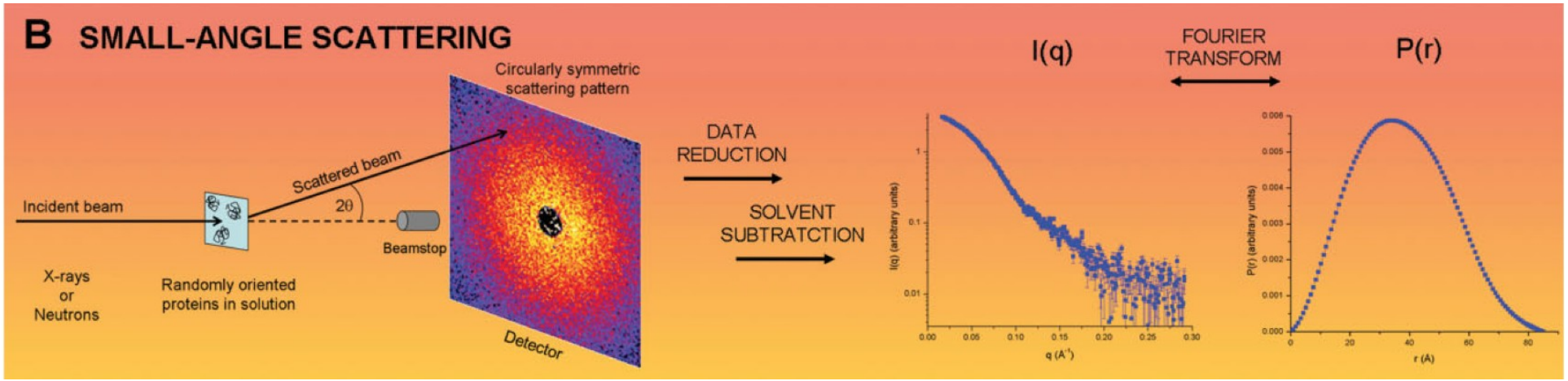

(a)
(b)
Crystal structure
All closed model
All open model
(c)
AMPPNP
ADP
Intensity (log scale)
$10^4$
$10^3$
$10^2$
10
$\chi^2$= 11·0
Crystal structure
$\chi^2$= 7·6
Crystal structure
$\chi^2$= 4·8
All closed model
$\chi^2$= 6·4
49% Crystal structure
35% All open model
16% All closed model
0·05
0·1
0·15
0·2
0·25
q (Å$^{-1}$)

# Summary

**Table 1.** *Common parameters defined by SAXS for monodisperse and homogeneous scatterers*

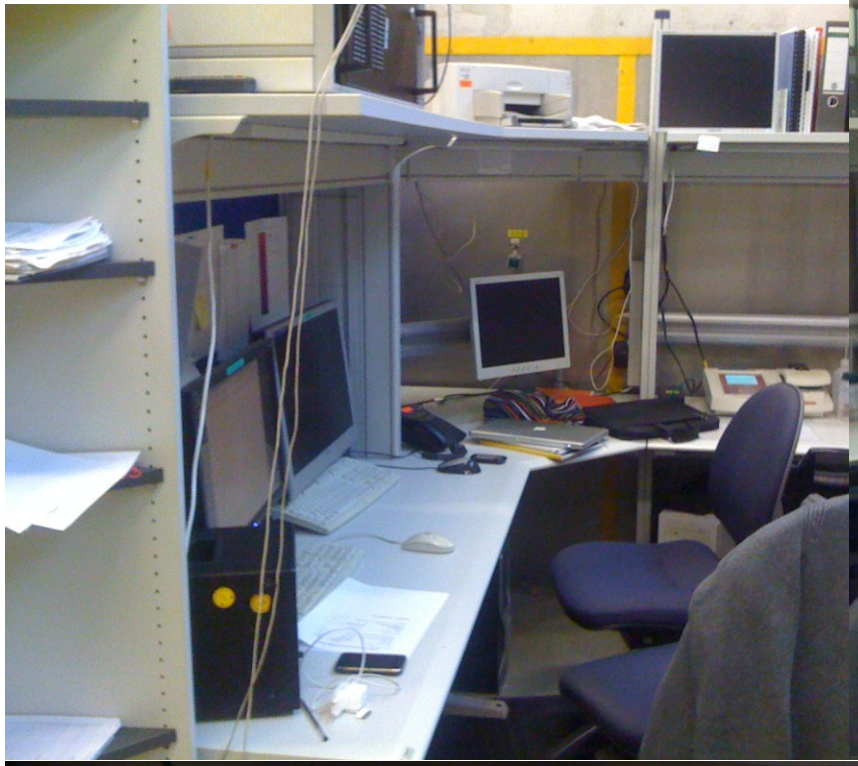
| Parameter | Formula | Range of data used and variable definitions | Comments |
|---|---|---|---|
| Radius of gyration ($R_G$): Guinier approximation | $\ln[I(q)] = \ln[I(0)] - \frac{q^2 R_G^2}{3}$ | $qR_G < 1{\cdot}3$ globular, $qR_G < 0{\cdot}8$ elongated. $I(0)$: Intensity of the scattering profile extrapolated to $q=0$ | Most common method of estimating $R_G$ Measured via the slope of the plot $\ln[I(q)]$ *vs.* $q^2$ |
| Radius of gyration ($R_G$): Debye approximation | $I(q) = \frac{2I(0)}{q^4 R_G^4}(q^2 R_G^2 - 1 + e^{-q^2 R_G^2})$ | $qR_G < 1{\cdot}4$ for elongated macromolecules | Particularly useful for elongated proteins where the Guinier approximation is valid over narrower range |
| Radius of gyration ($R_G$): defined by $P(r)$ | $R_G^2 = \int_0^{D_{max}} r^2 P(r)dr \Big/ \int_0^{D_{max}} P(r)dr$ | Entire $q$-range. $D_{max}$: Maximum dimension of particle | Good consistency check for $R_G$, $D_{max}$, and $P(r)$ |
| Pair distribution function ($P(r)$) | $P(r) = \frac{r}{2\pi^2}\int_0^{\infty} I(q) q \sin(qr)dq$ | Entire $q$-range | Indirect Fourier transform methods have been developed for calculating $P(r)$ |
| Maximum dimension ($D_{max}$) | $D_{max}$ is the value of $r$ at $P(r)=0$ for large $r$ | Requires data $q \leqslant \pi/D_{max}$ | Assignment of $D_{max}$ may be complicated by flexibility or multimerization |
| Particle volume ($V$): defined by Porod Invariant | $V = 2\pi^2 I_{exp}^2(0) \Big/ \left(\int_0^{\infty} I(q) q^2 dq\right)$ | Entire $q$-range. $I_{exp}(0)$ is the experimental intensity at $q=0$ and does not require an absolute scale | The integral portion of this equation is known as the Porod invariant. Accuracy varies for shape and size; however absolute scale and concentration information are unnecessary |
| $I(0)$: Intensity at $q=0$ which is also proportional to mass and volume | $I(0) = 4\pi\left(\int_0^{D_{max}} P(r)dr\right)$ | Entire $q$-range | Calculation of $M$ and $V$ using this version of $I(0)$ is less susceptible to aggregation and inter-particle correlations than extrapolation of low $q$ data |
| Mass ($M$) | $M = \frac{I(0)\mu^2}{N_A(1-(\rho_S/\rho_p))^2}$ | $\mu$: Average mass per number of electrons. $\rho_S$: Solvent electron density $\rho_p$: Particle electron density $N_A$: Avagadro's number | $I(0)$ must be on an absolute scale and normalized by mass/volume and not molar concentration |
| **Formulas for elongated or flexibly linked linear macromolecules** | | | |
| Radius of gyration of cross-section ($R_{XC}$) | $\ln[qI(q)] = \ln[qI(0)] - \frac{q^2 R_{XC}^2}{2}$ | Intermediate $q$ values | The slope of the linear portion of a plot of $\ln[qI(q)]$ *vs.* $q^2$ is $R_{XC}^2$; however, $R_{XC}^2$ goes to 0 as $q$ goes to 0 in regimes where scattering is dominated by $R_G$ |
| Length ($L$) | $L = (12(R_G^2 - R_{XC}^2))^{\frac{1}{2}}$ | See $R_G$ and $R_{XC}$ | The co-axial length rather than the hypotenuse ($D_{max}$) |

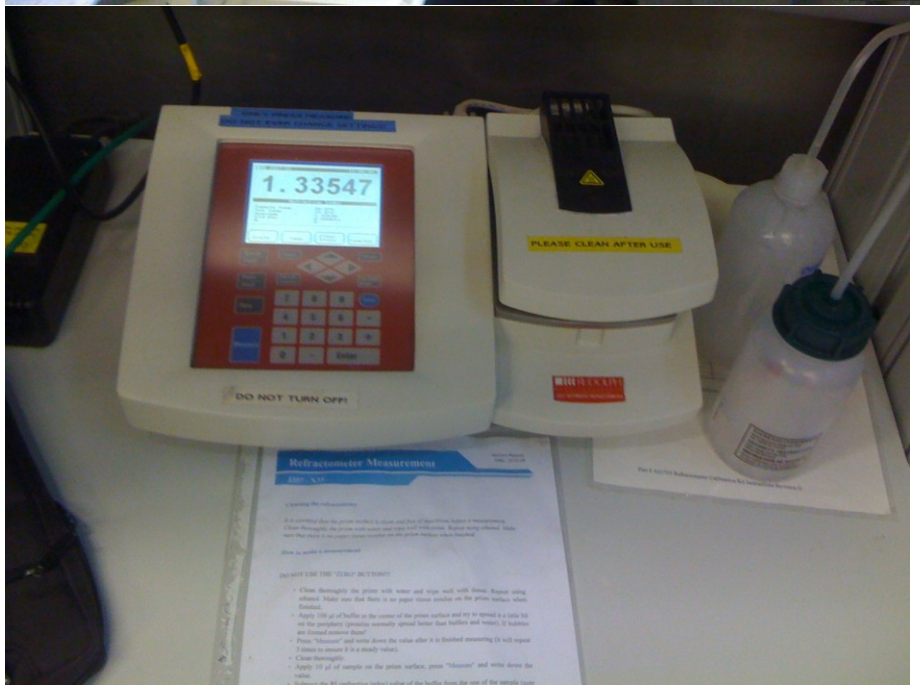
Summary (graphically)



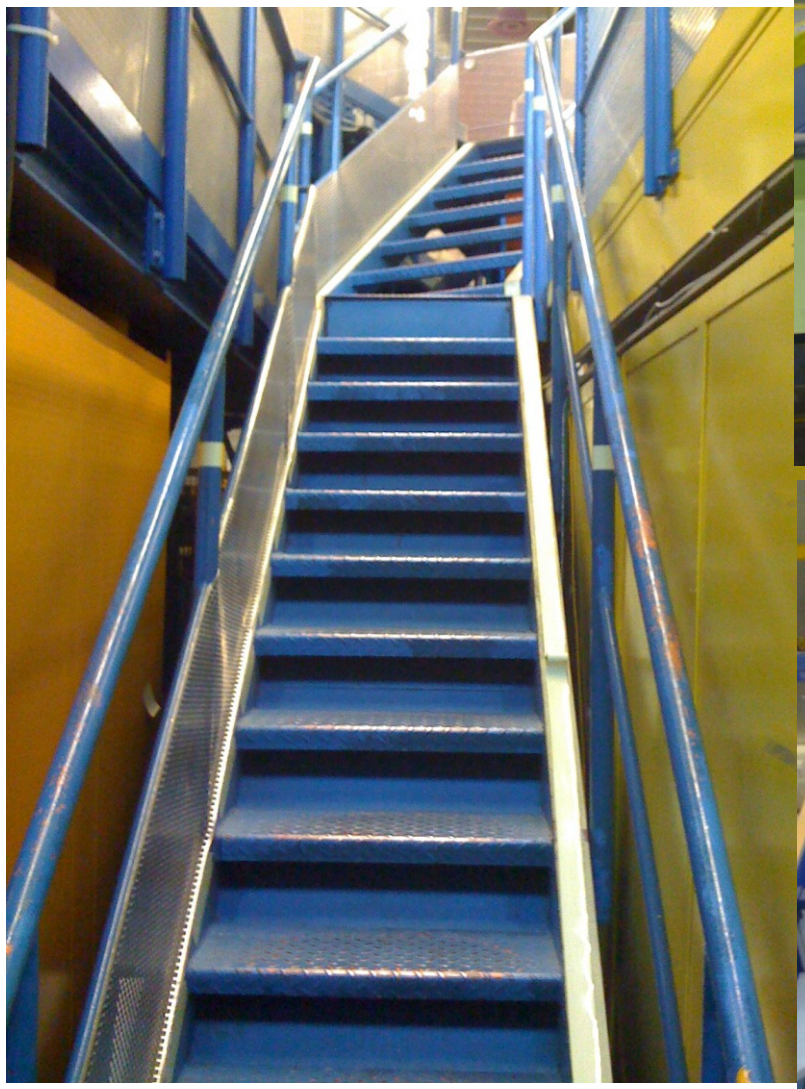
SAXS Hamburg

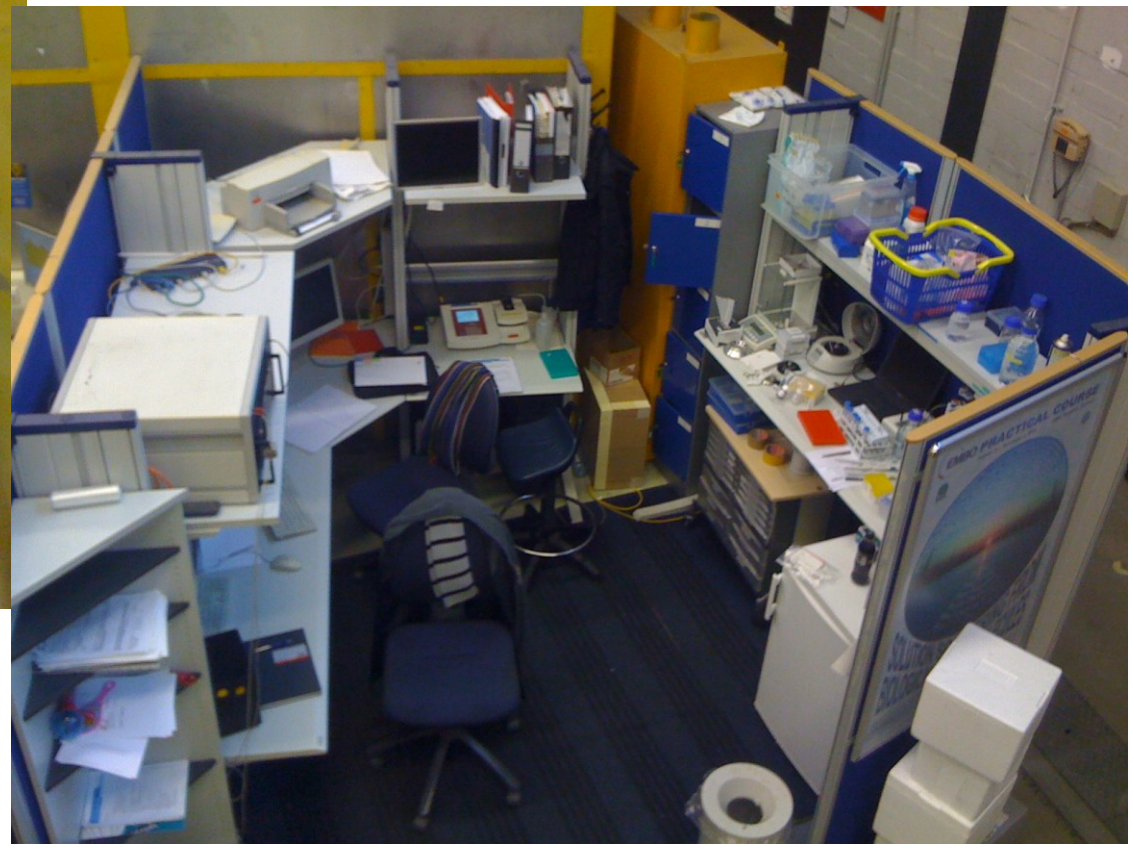

M 7-8



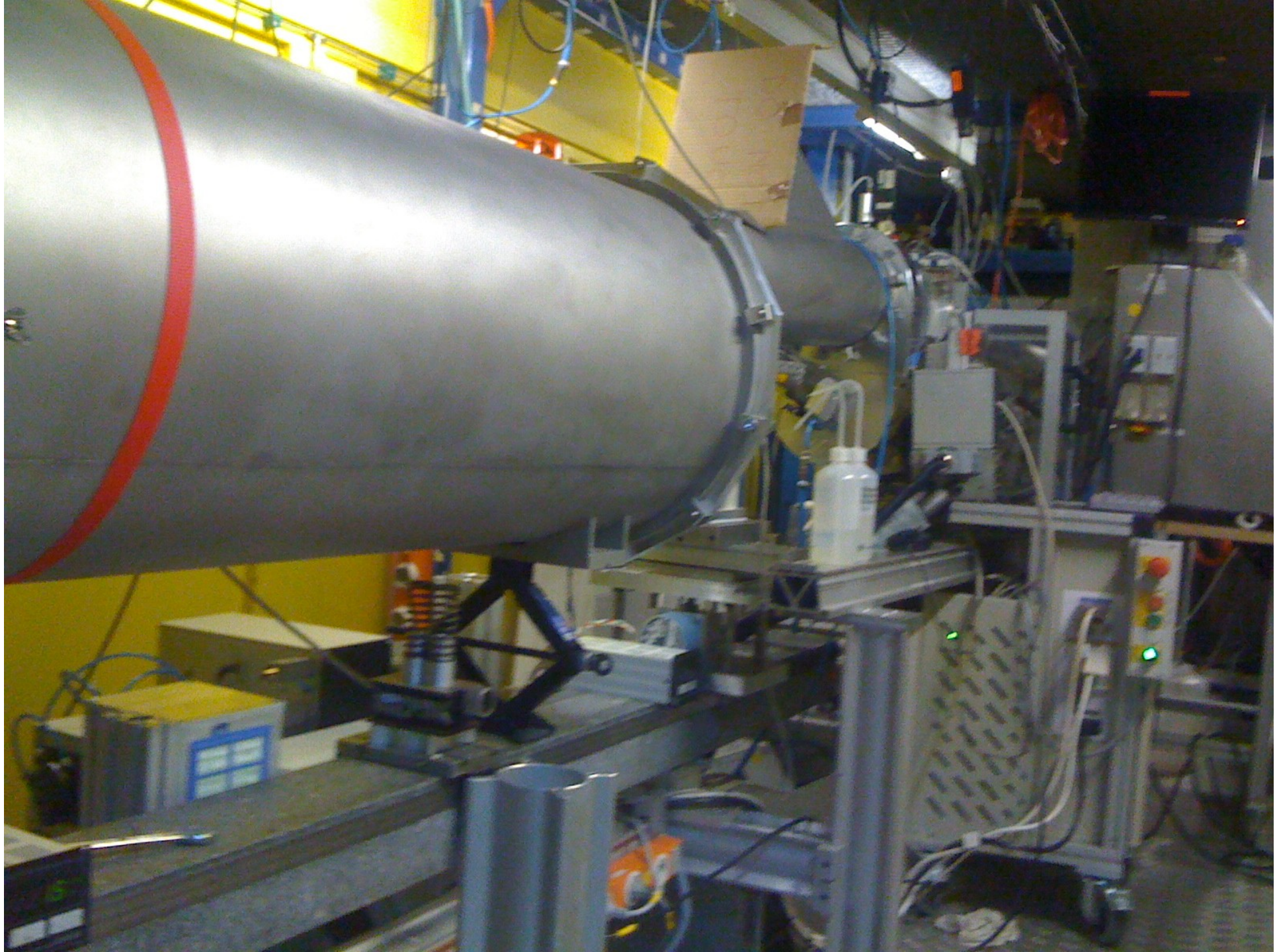

1. 33547
PLEASE CLEAN AFTER USE
DO NOT TURN OFF!
Refractometer Measurement

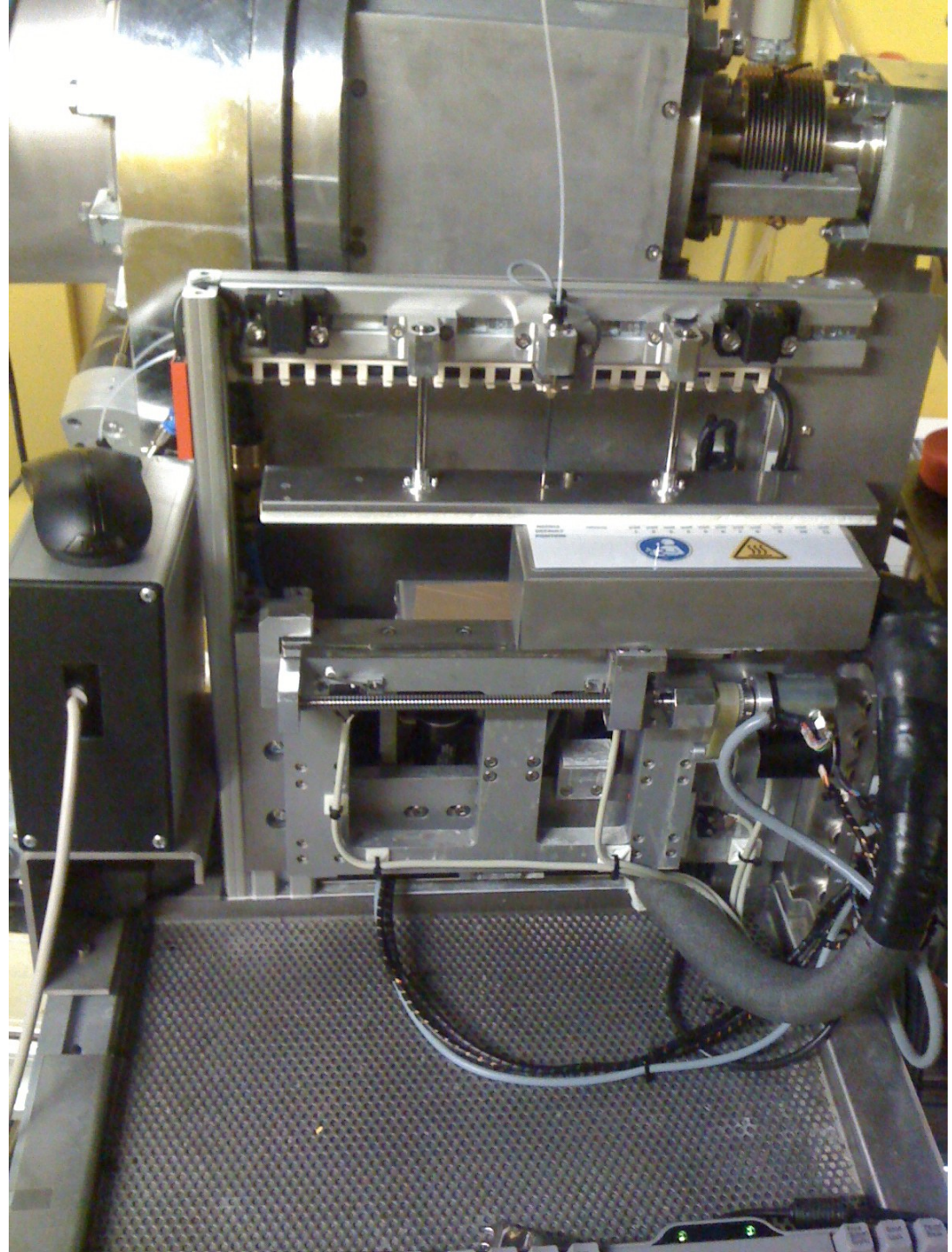

Spectrophotometer

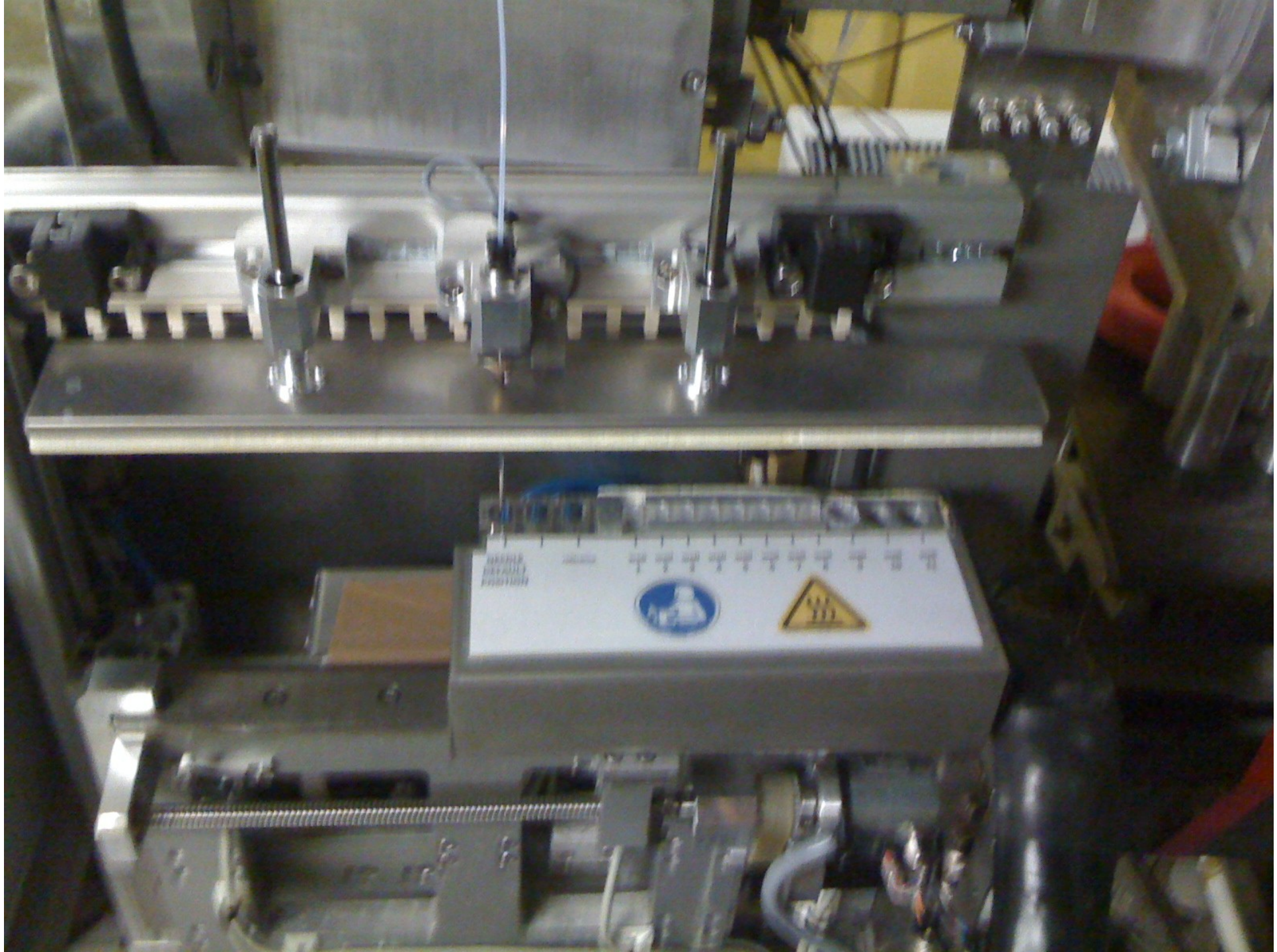




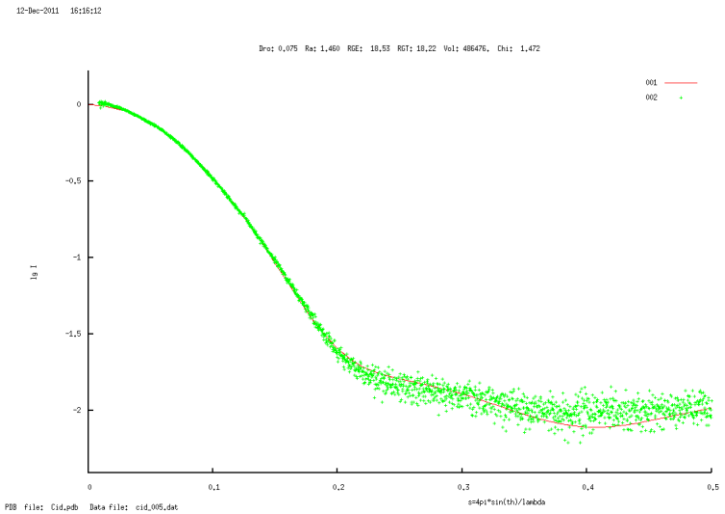

offen
TÜR-NOTAUF
NG-D1/2

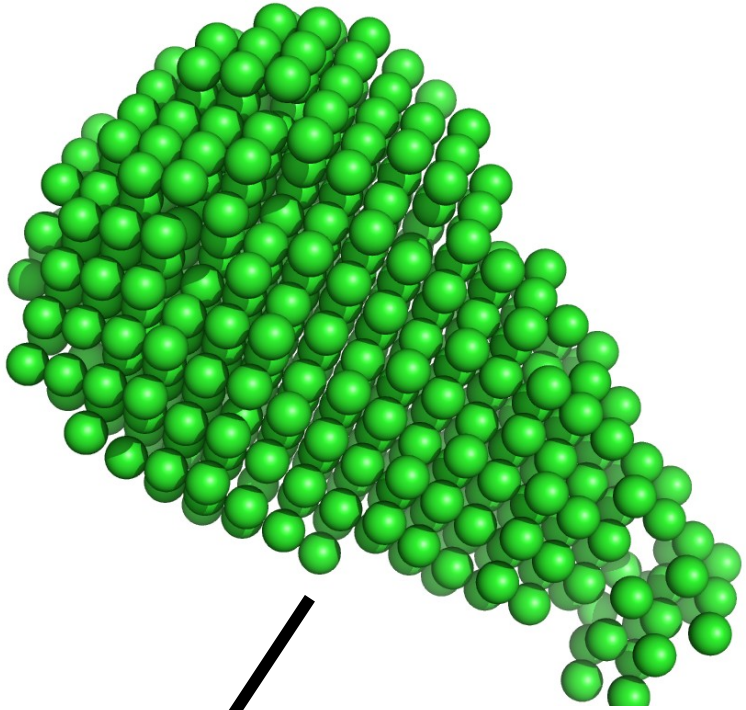


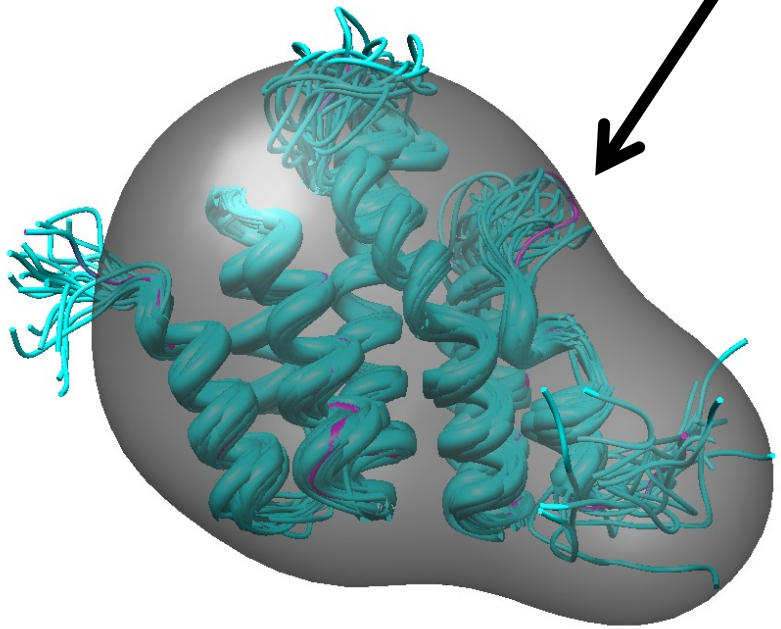




# Evaluation of the experimental data



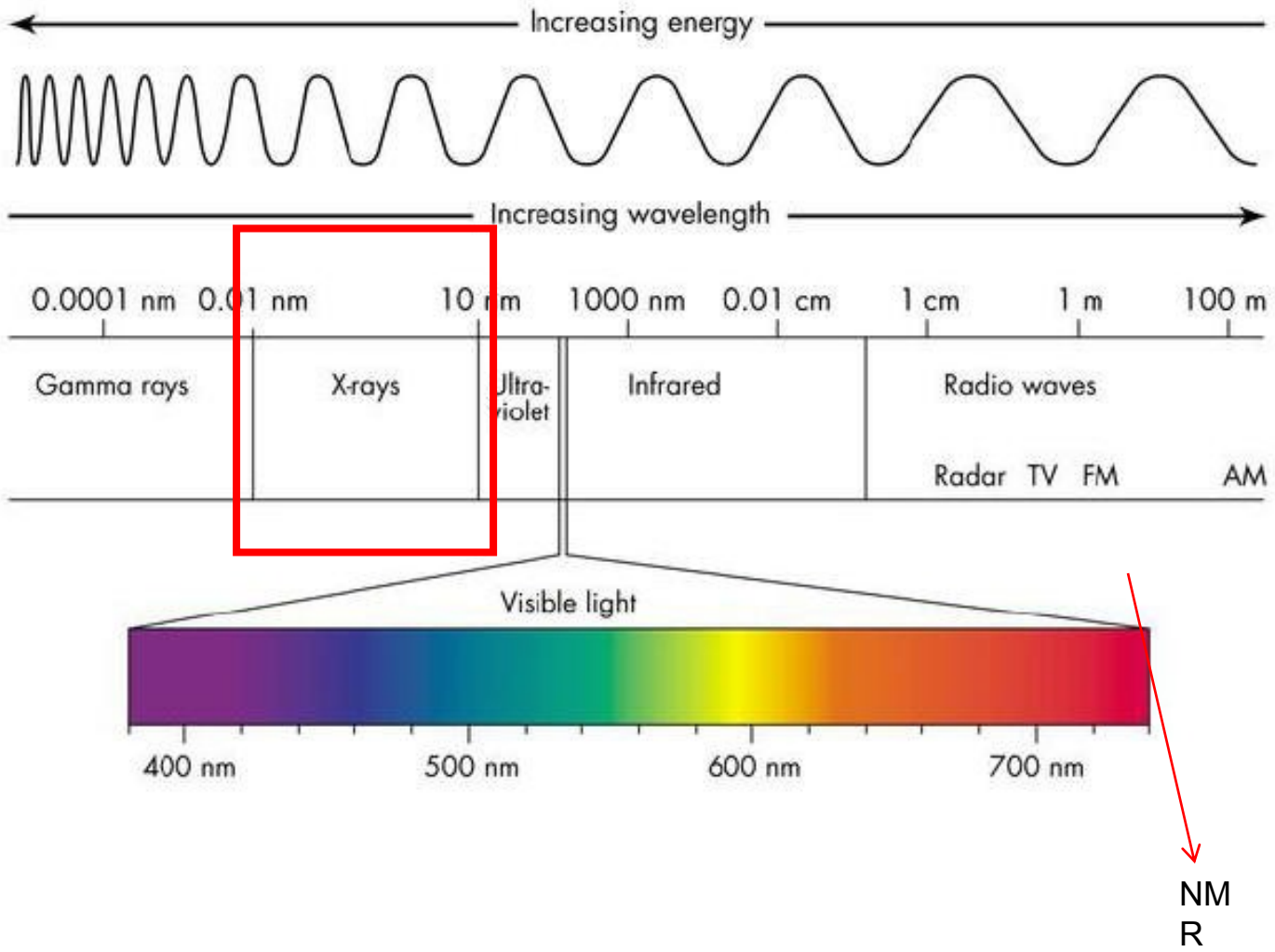


References:


1) Koch, M.H.J., Vachette, P., Svergun, D.I. *Quart Rev Biophys* **2003**
2) Jacques, D.A., Trewhella, J. *Prot Sci* **2010**
3) Svergun, D.I., Petoukhov, M.V., Koch, M.H.J. *Biophys J* **2001**
4) Wriggers, W. *Biophys Rev* **2010**
5) Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A. *Quart Rev Biophys* **2007**
6) Madl, T. Gabel, F., Sattler, M. *J Struct Biol* **2010**

# AlphaFold: a solution to a 50-year-old grand challenge in biology


https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology


Increasing energy
Increasing wavelength
0.0001 nm
0.01 nm
10 nm
1000 nm
0.01 cm
1 cm
1 m
100 m
Gamma rays
X-rays
Ultra-violet
Infrared
Radio waves
Radar
TV
FM
AM
Visible light
400 nm
500 nm
600 nm
700 nm
NM
R

# Confidence in structural features of proteins determined by X-ray crystallography

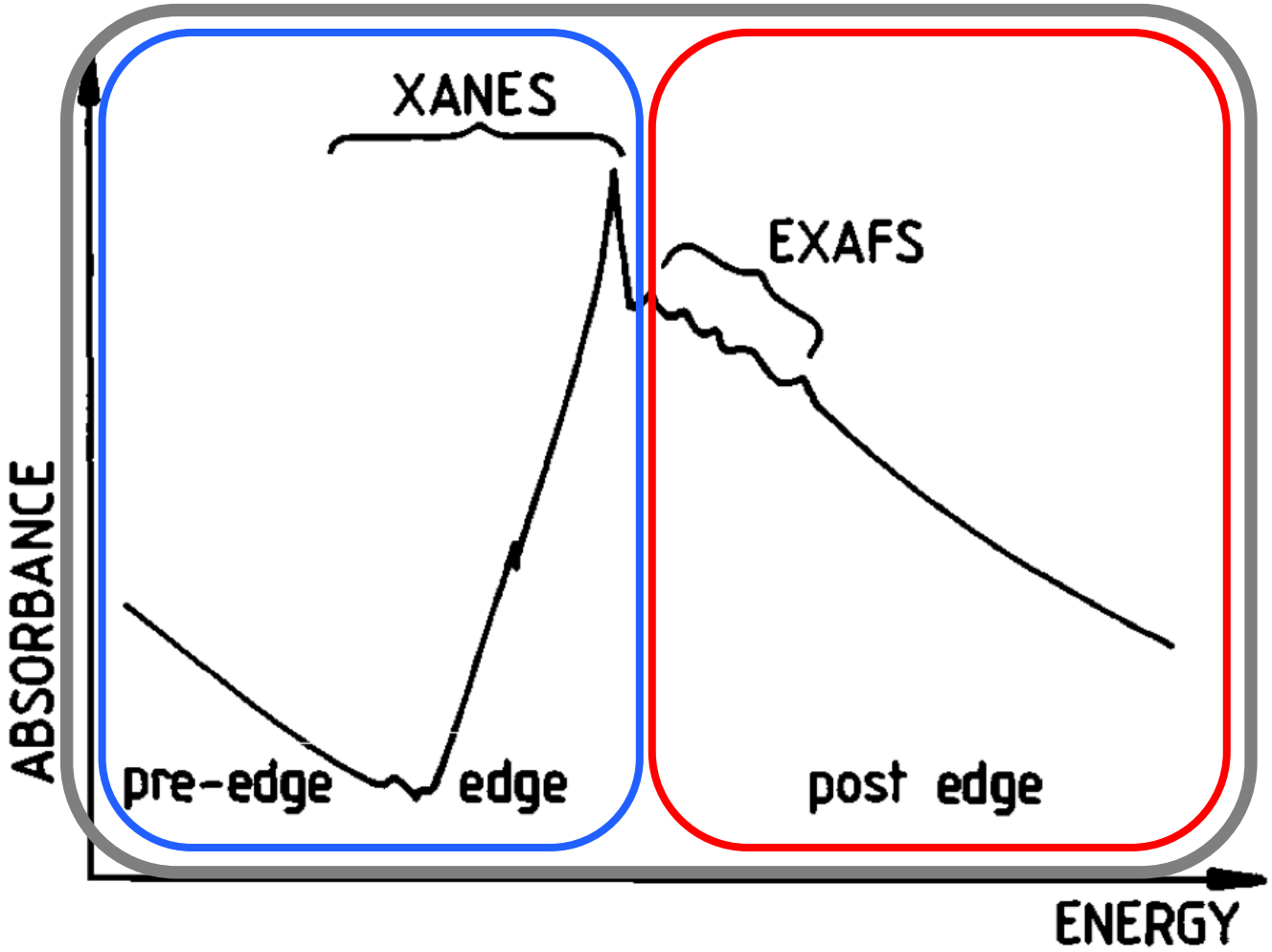
(estimates are very rough and strongly depend on the quality of the data)

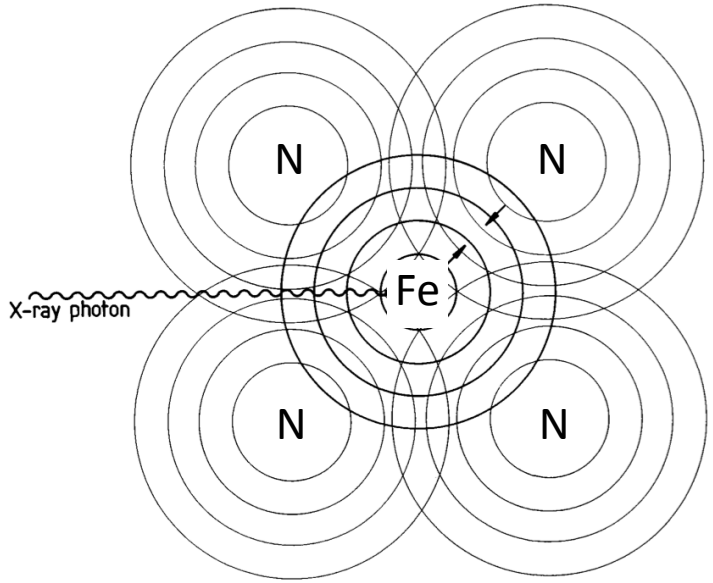
| Structural feature | | | Resolution | | |
|---|---|---|---|---|---|
| | 5 Å | 3 Å | 2.5 Å | 2 Å | 1.5 Å |
| Chain tracing | - | Fair | Good | Good | Good |
| Secondary structure | Helices fair | Fair | Good | Good | Good |
| Sidechain conformations | - | - | Fair | Good | Good |
| Orientation of peptide planes | - | - | Fair | Good | Good |
| Protein hydrogen atoms visible | - | - | - | - | Good |

XAS = X-ray Absorption Spectroscopy

XANES = X-ray Absorption Near Edge Structure

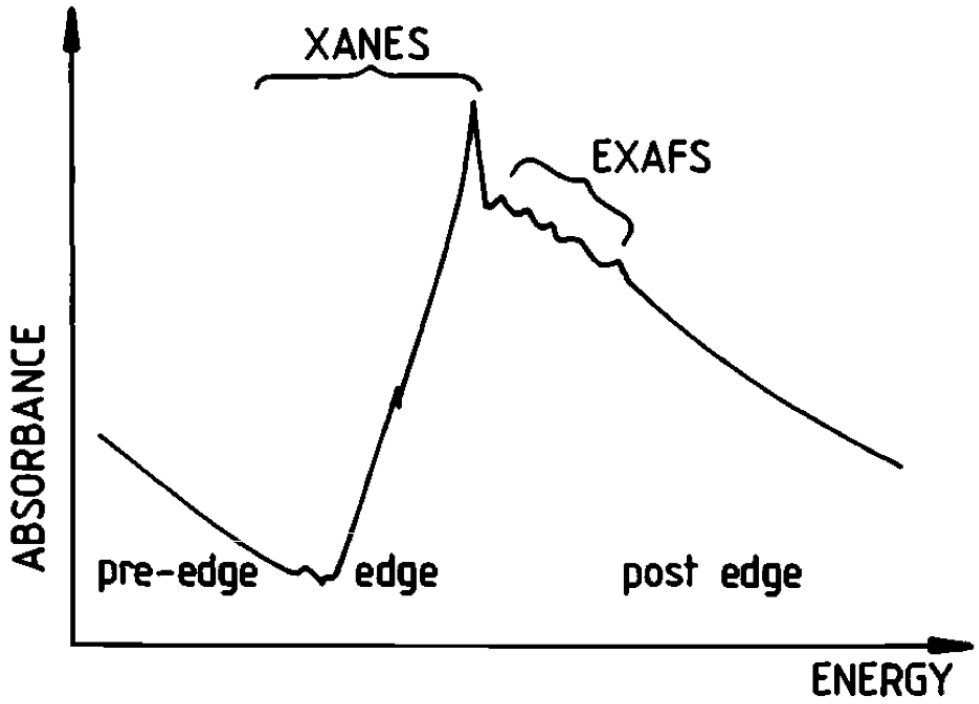
EXAFS = Extended X-Ray Absorption Fine Structure

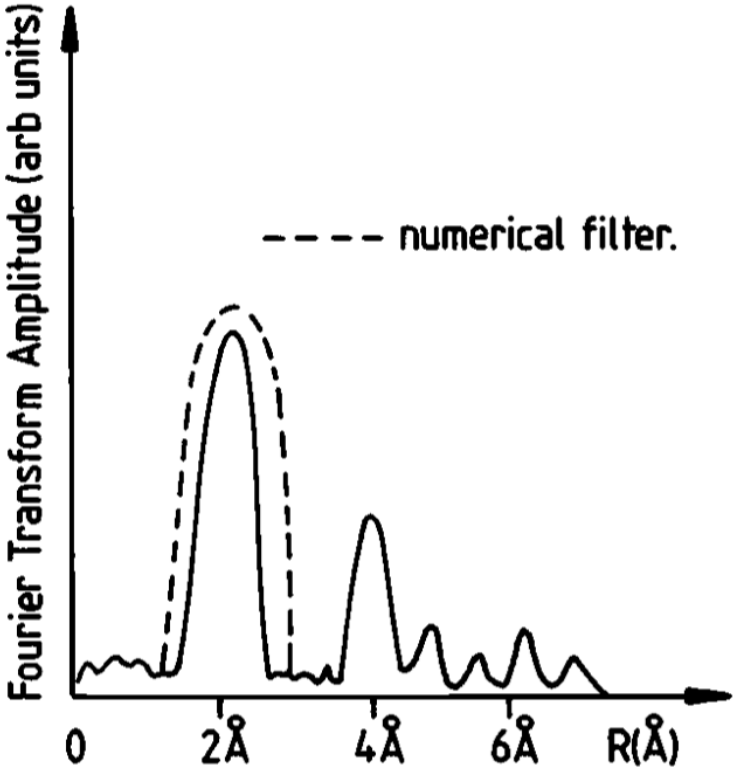




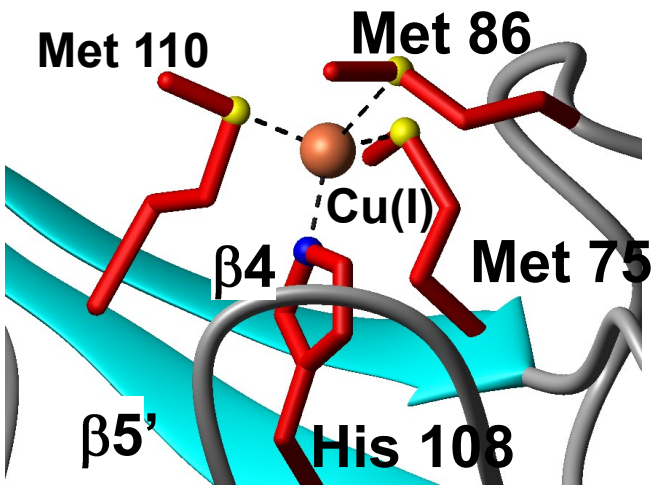

$$I = I_0 \exp(-\mu x)$$

EXAFS (Extended X-ray Absorption Fine Spectroscopy) poskytuje informace o nejbližších slupkách atomů sousedících s absorbujícím atomem



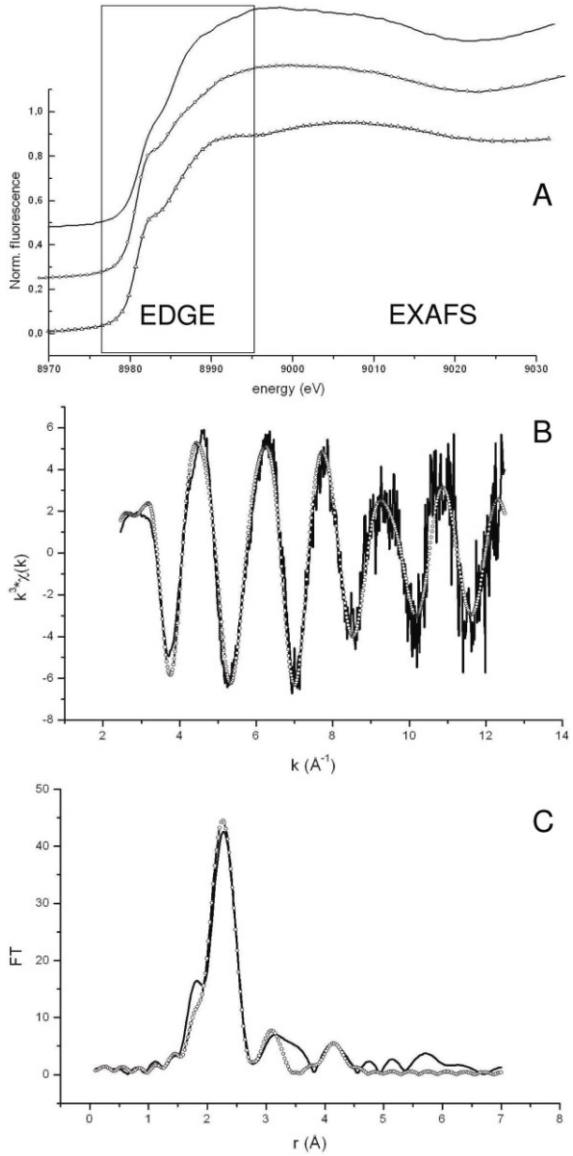
## X-Ray Absorption Spectroscopy



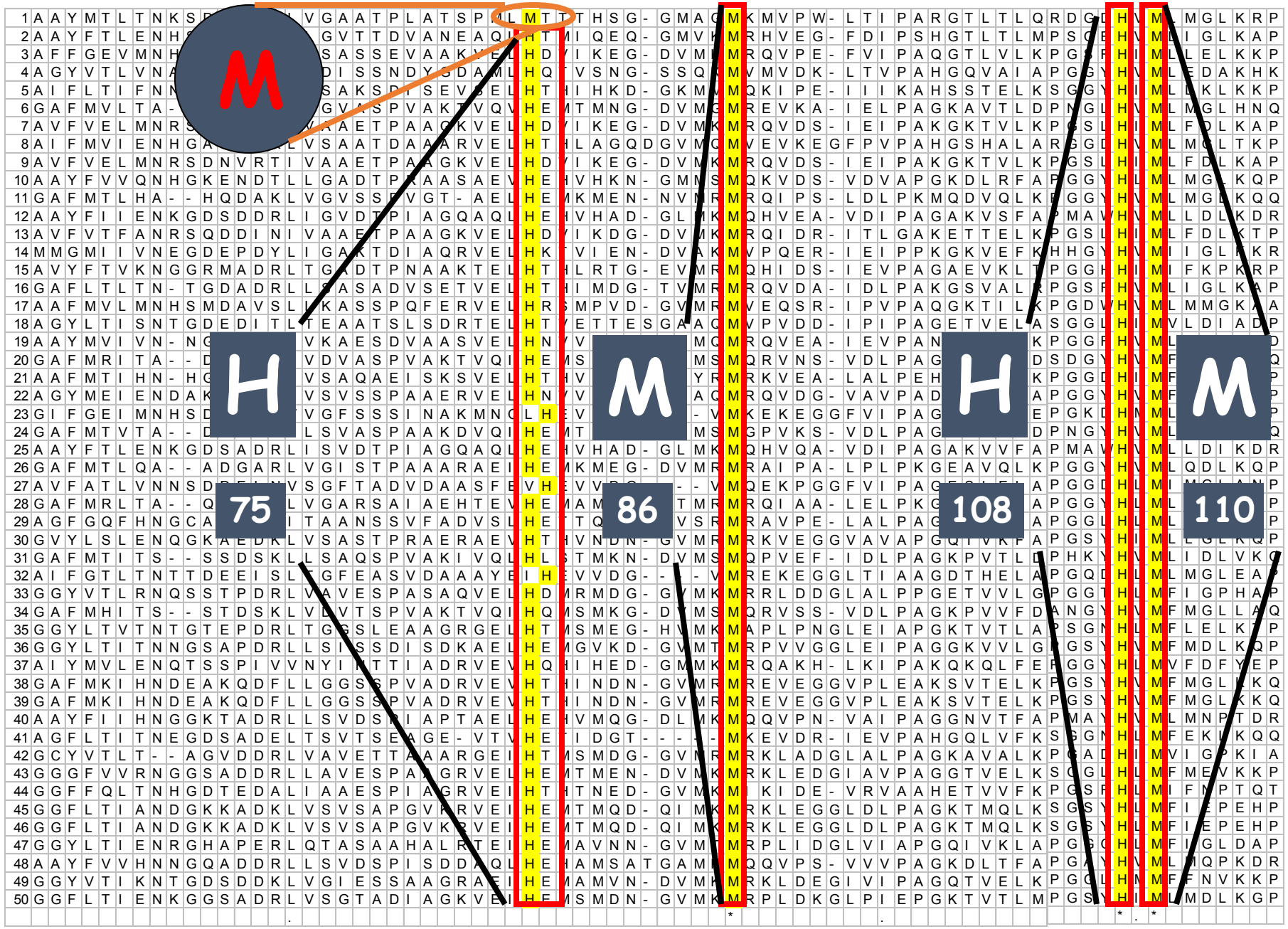

**Cu(I)DR1885 ΔE=-10.3 eV**

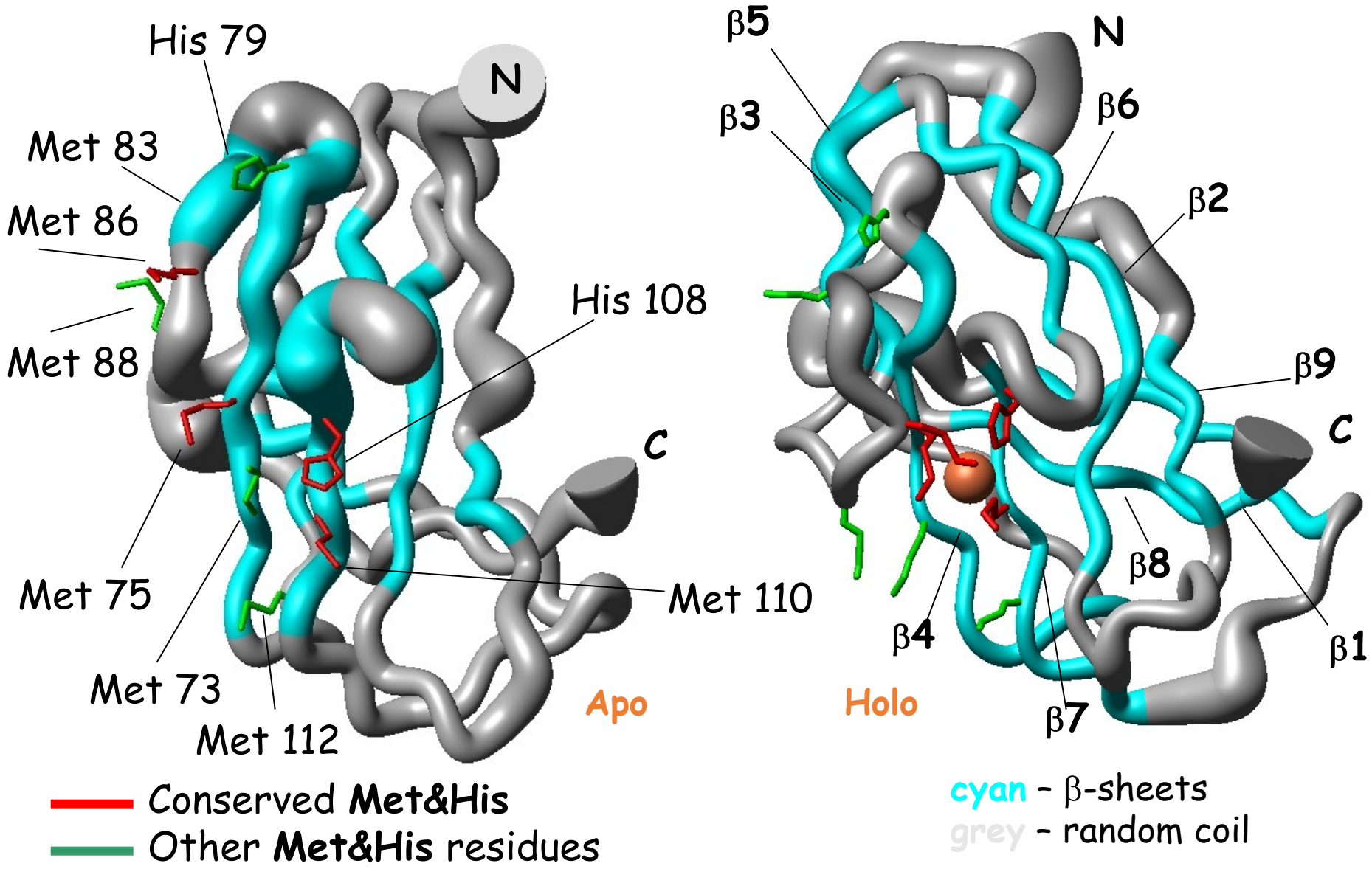
| | Ligand | r(Å) | $2\sigma^2.10^3(Å^2)$ | R-exafs | ε(fit index) |
|---|---|---|---|---|---|
| Fit1 (1shell) | 2S | 2.299 | 4(1) | 0.446 | 0.49 |
| Fit2 (1shell) | 3S | 2.301 | 9(1) | 0.403 | 0.41 |
| Fit3 (2shells) | 3S | 2.300 | 8(1) | 0.334 | 0.29 |
| | 1N§ | 1.982 | 4(1) | | |
| Fit4 (2shells) | 3S | 2.303 | 8(1) | 0.305 | 0.27 |
| | 1N* | 1.999 | 7(2) | | |

§ no MS *His, MS



# A BLAST search over all non-redundant GenBank genomes



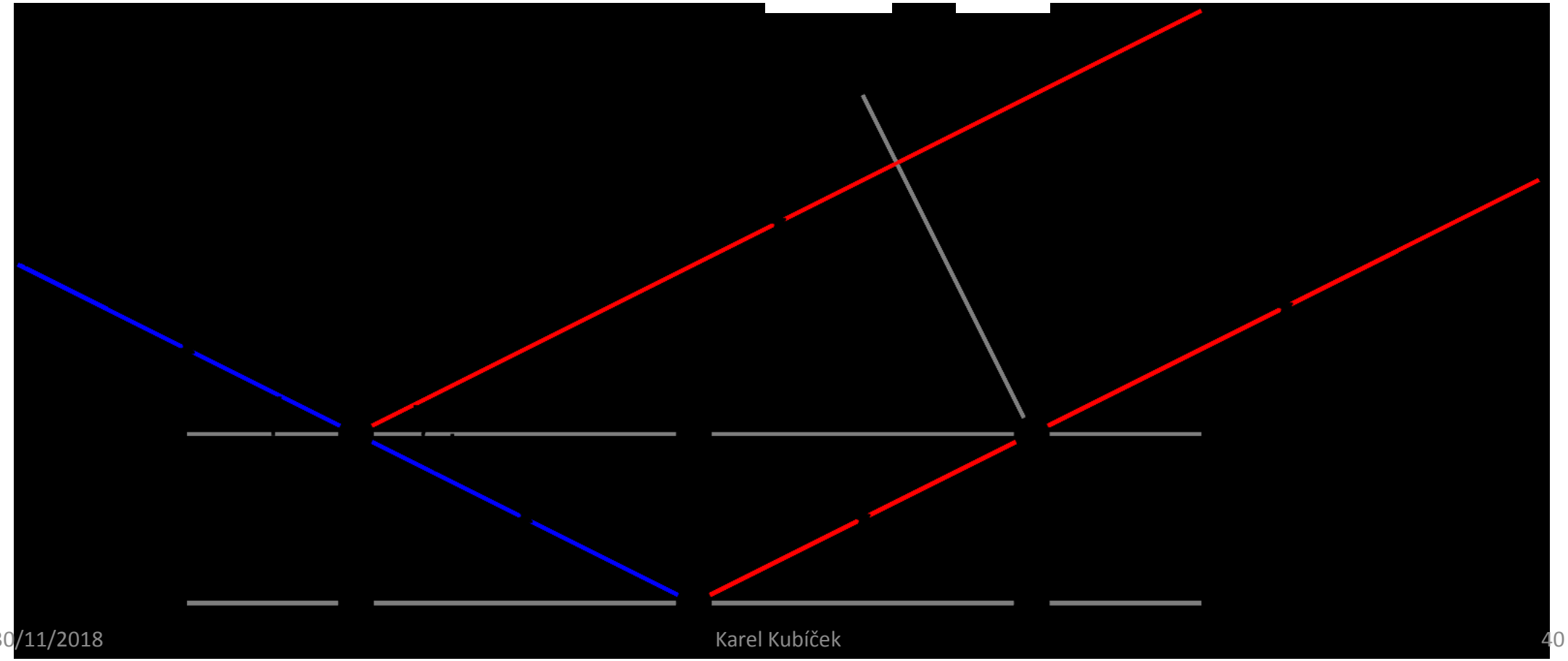

His 79
N
Met 83
Met 86
Met 88
His 108
C
Met 75
Met 110
Met 73
Met 112
Apo
β5
N
β3
β6
β2
β9
C
β8
β4
β1
β7
Holo
Conserved Met&His
Other Met&His residues
cyan – β-sheets
grey – random coil

# Rentgenstrukturní analýza

Krystalová mřížka působí na rentgenové záření jako optická mřížka na viditelné světlo. Nastávají ohybové jevy a na stínítku se objevuje difrakční obrazec. Tyto obrazce mohou být matematicky analyzovány, aby se získala informace o rozložení elektronů v molekulách tvořících krystal.

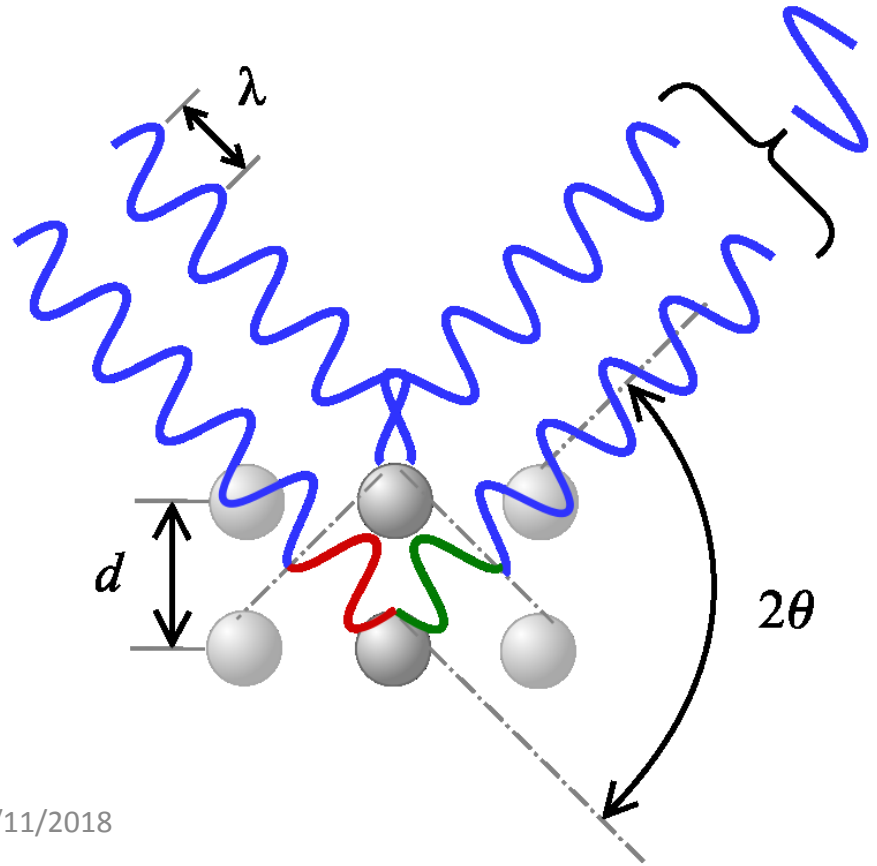
# Bragg‘s law

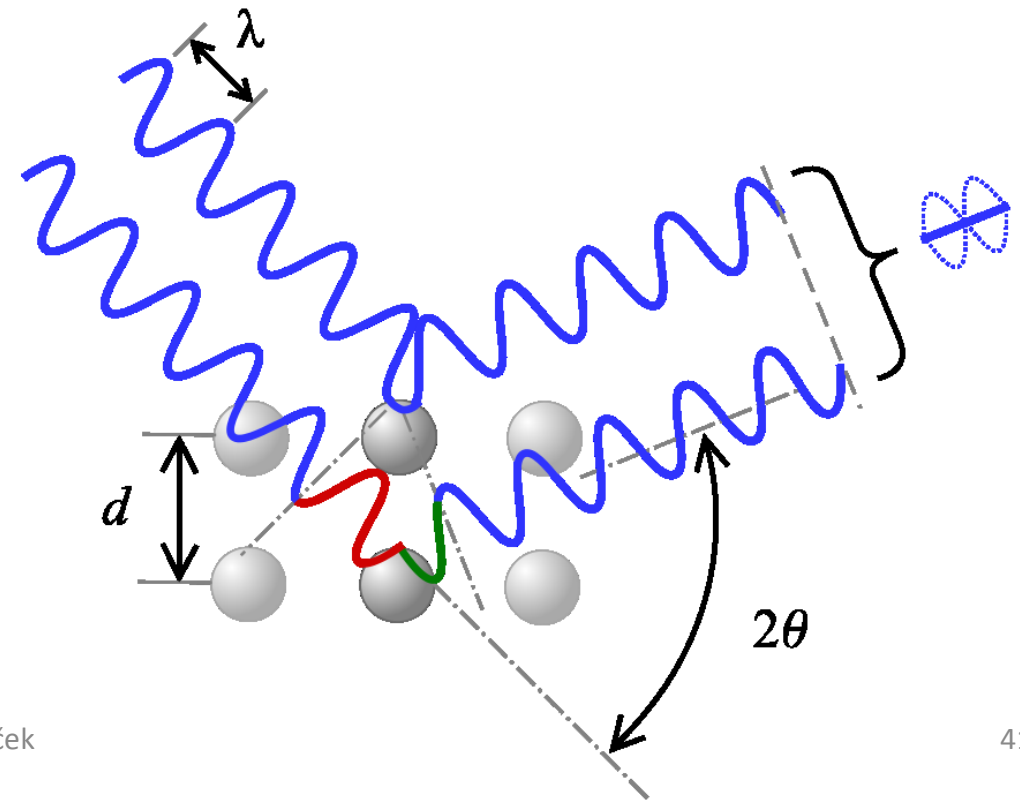
$$2d \sin\Theta = n\lambda$$



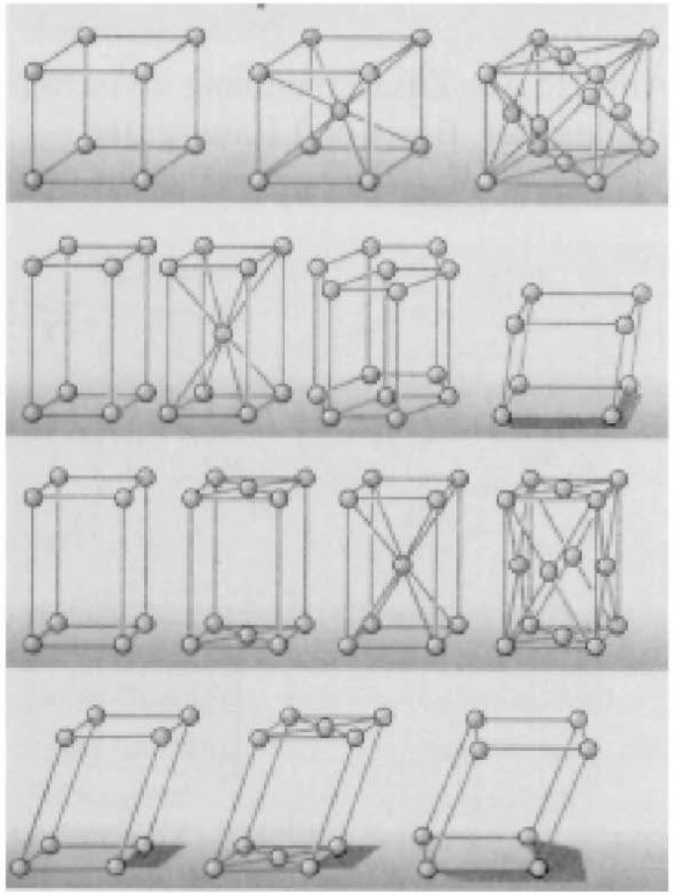
# Bragg‘s law

$$2d \sin\Theta = n\lambda$$

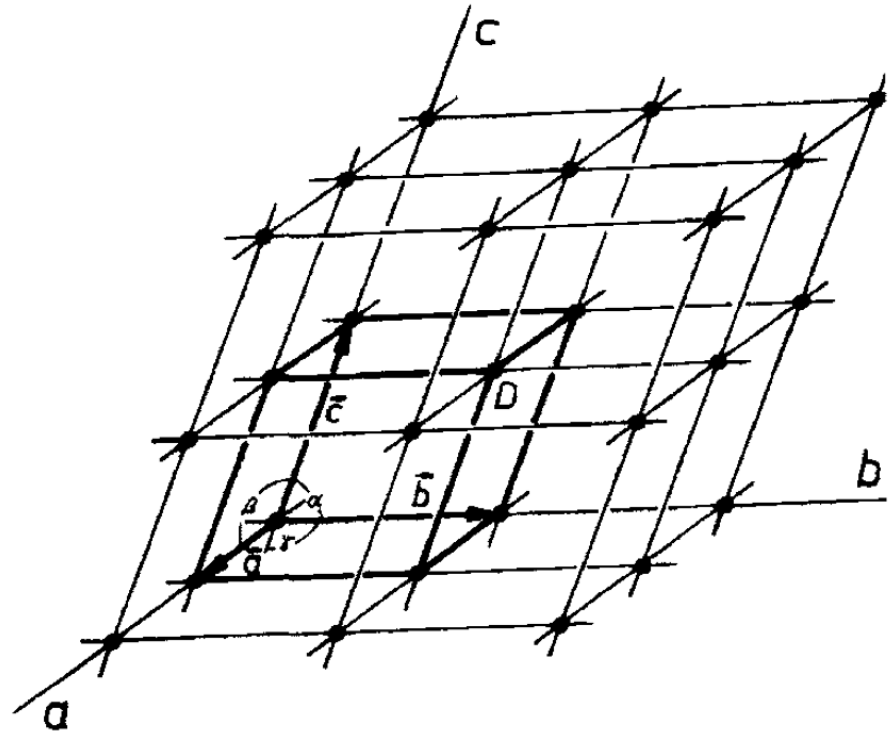


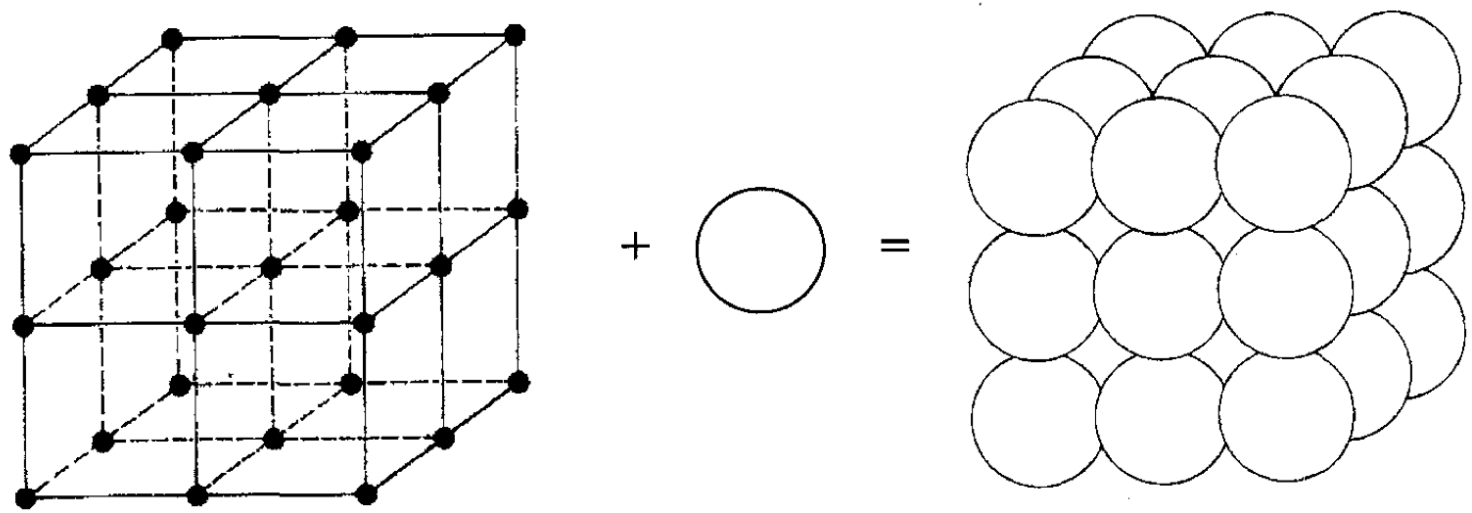



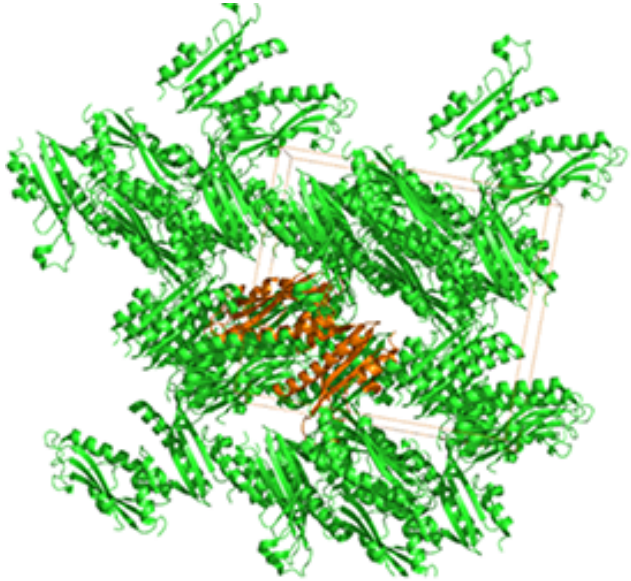
Krystalografická soustava



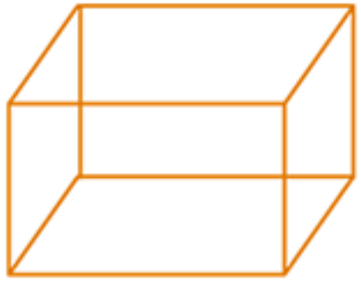
| Raumgitter | | |
|---|---|---|
| $\alpha_1, \alpha_2, \alpha_3 \neq 90^O$ : | **triklin** | Trojklonná |
| $\alpha_1 = \alpha_2 = 90^O$, $\alpha_3 \neq 90^O$ : | **monoklin** | Jednoklonná |
| $\alpha_1 = \alpha_2 = \alpha_3 = 90^O$, $a_0 \neq b_0 \neq c_0$ : | **orthorhombisch** | Kosočtverečná (ortorombická) |
| $\alpha_2 = \alpha_3 = 90^O$, $\alpha_1 = 60^O$, $b_0 = c_0$ : | **hexagonal** | Šesterečná |
| $\alpha_1 = \alpha_2 = \alpha_3 \neq 90^O$, $a_0 = b_0 = c_0$ : | **trigonal (rhomboedrisch)** | Klencová |
| $\alpha_1 = \alpha_2 = \alpha_3 = 90^O$, $a_0 = b_0 \neq c_0$ : | **tetragonal** | Čtverečná |
| $\alpha_1 = \alpha_2 = \alpha_3 = 90^O$, $a_0 = b_0 = c_0$ : | **kubisch** | Krychlová |


c
D
b
a

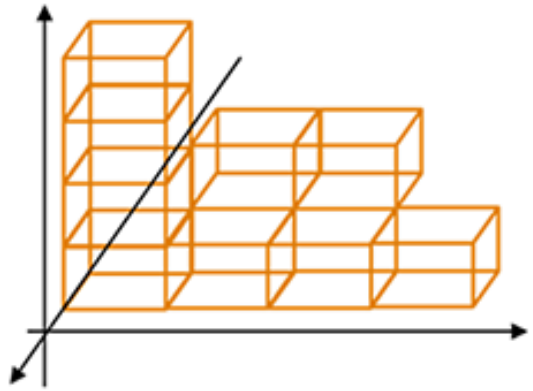




Crystal packing

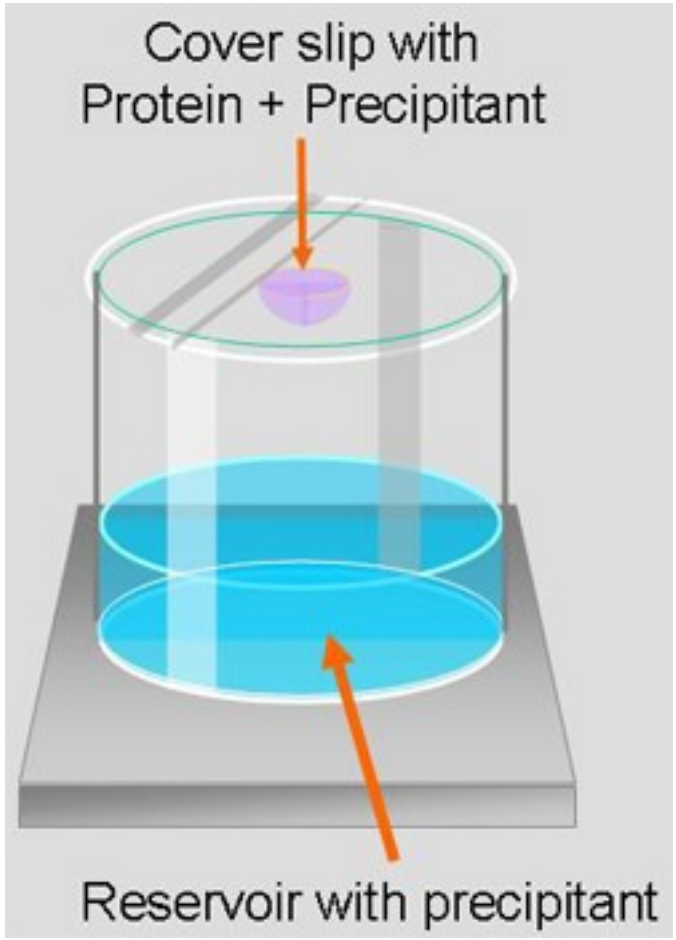


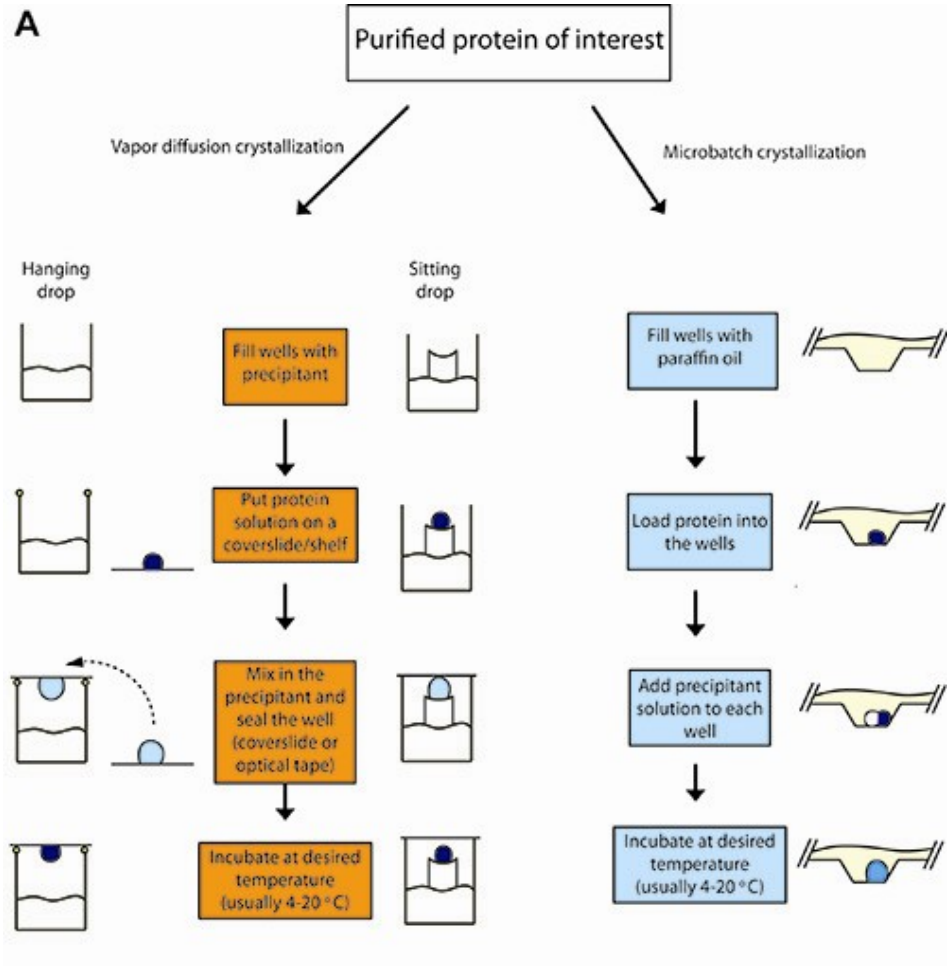
Cell

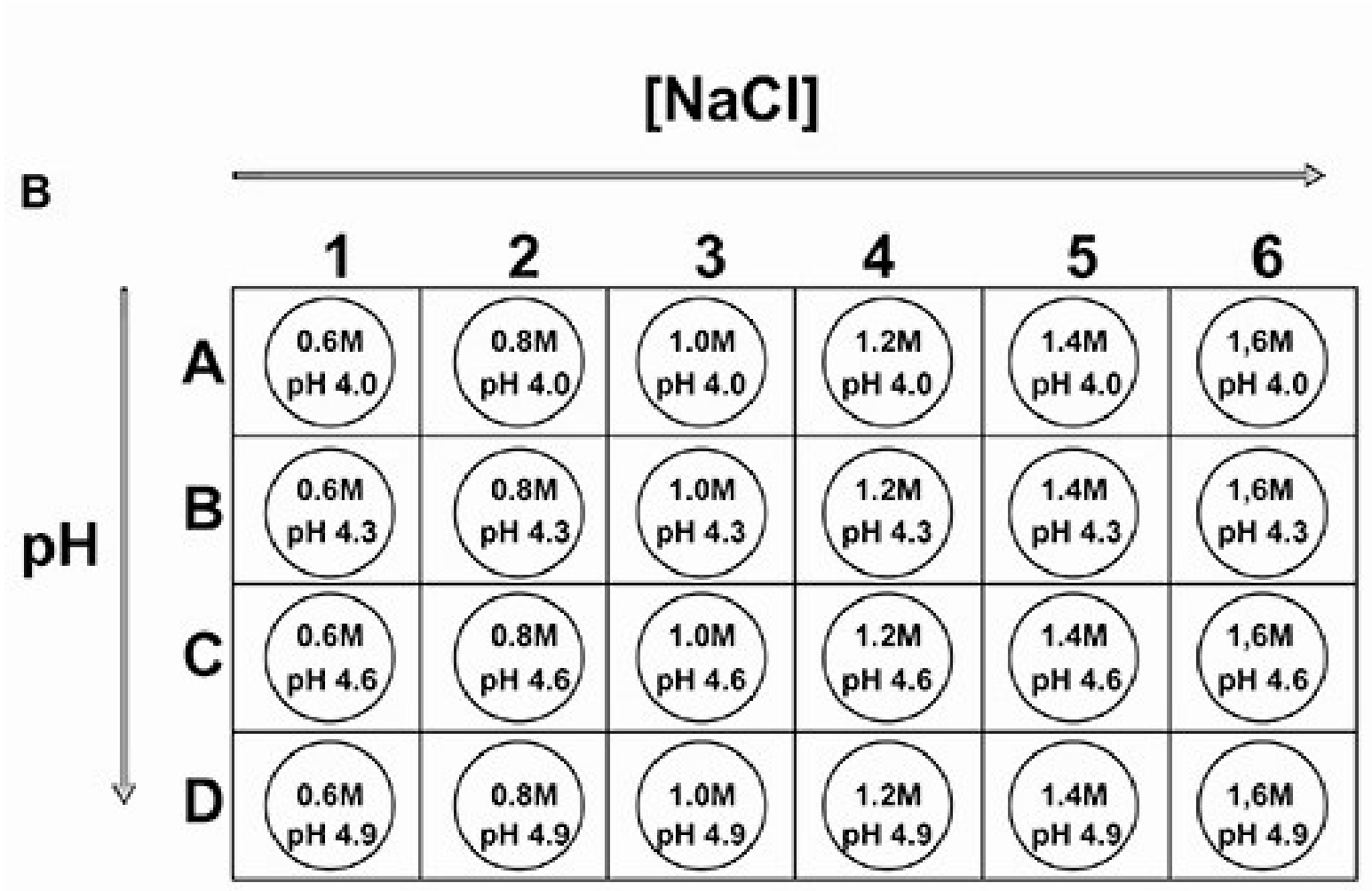


Crystal lattice

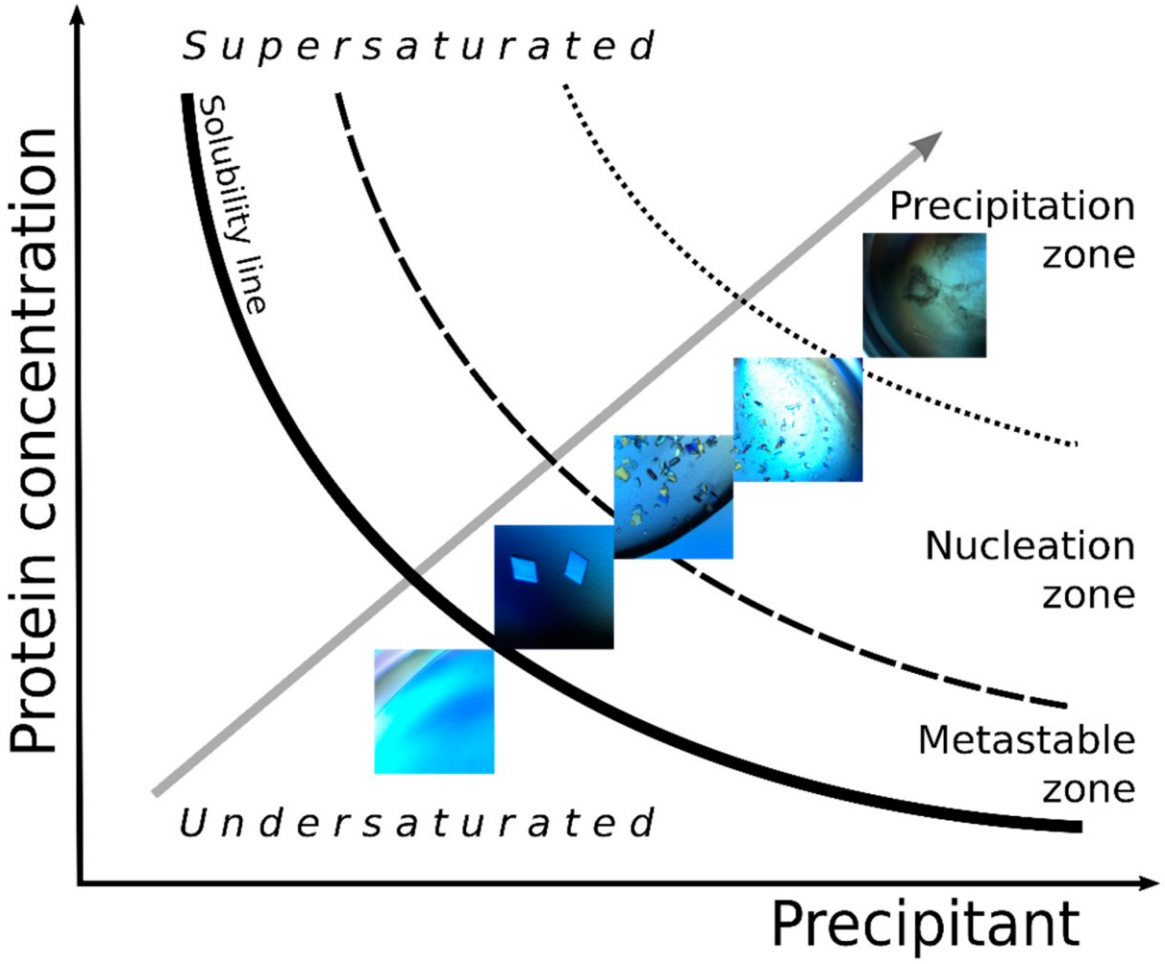
https://www.intechopen.com/books/x-ray-scattering/x-ray-diffraction-in-biology-how-can-we-see-dna-and-proteins-in-three-dimensions-

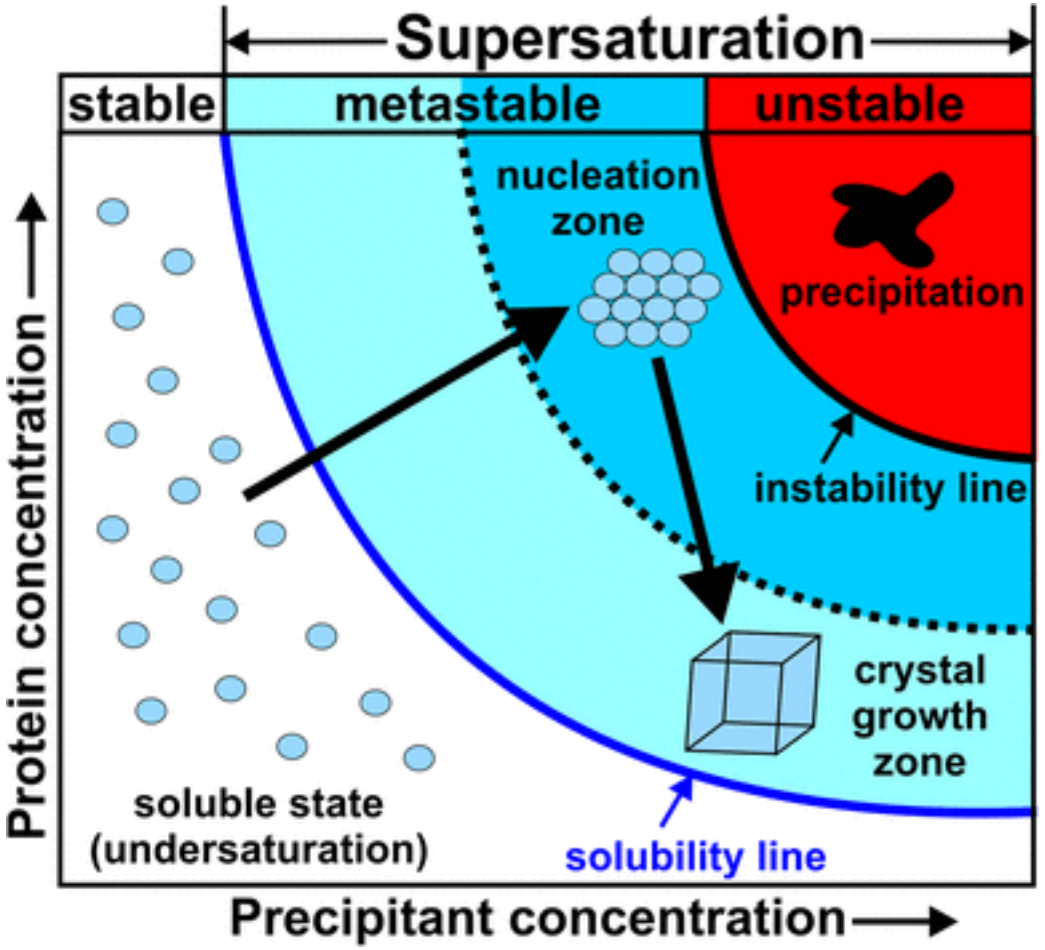

Cover slip with
Protein + Precipitant
Reservoir with precipitant


A
Purified protein of interest
Vapor diffusion crystallization
Microbatch crystallization
Hanging drop
Sitting drop
Fill wells with precipitant
Put protein solution on a coverslide/shelf
Mix in the precipitant and seal the well (coverslide or optical tape)
Incubate at desired temperature (usually 4-20 °C)
Fill wells with paraffin oil
Load protein into the wells
Add precipitant solution to each well
Incubate at desired temperature (usually 4-20 °C)

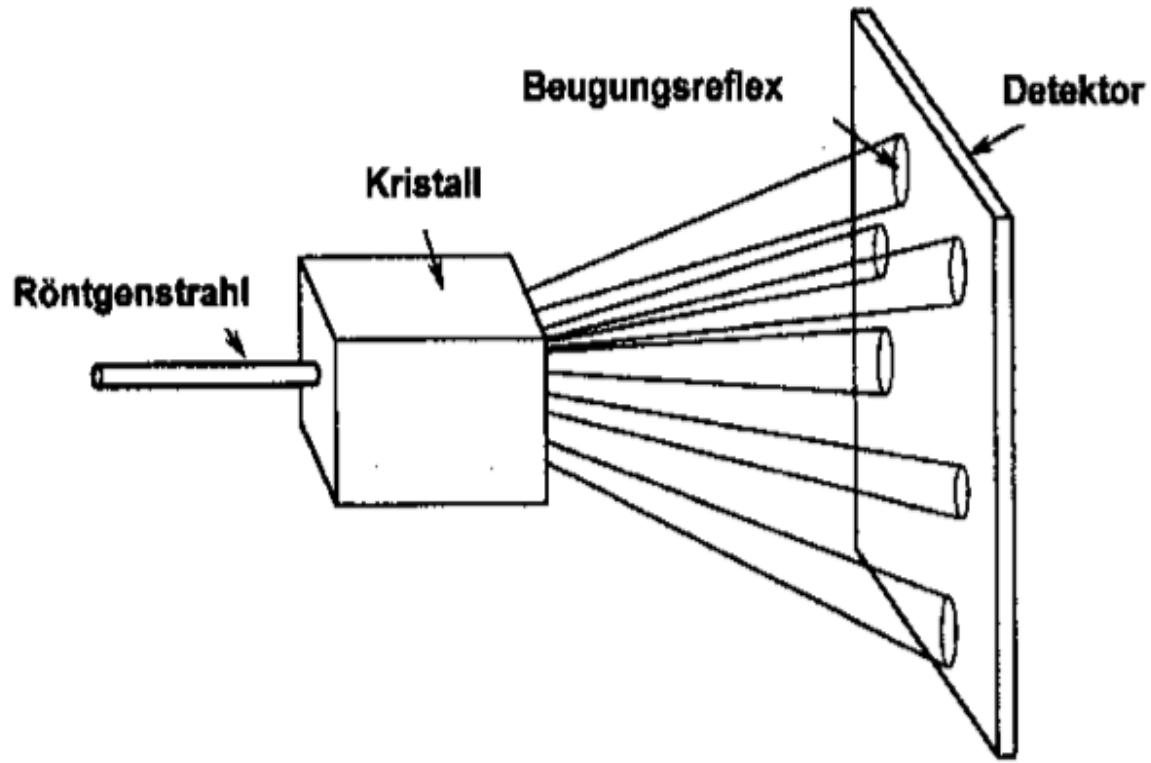

[NaCl]
B
1
2
3
4
5
6
pH
A
B
C
D
0.6M pH 4.0
0.8M pH 4.0
1.0M pH 4.0
1.2M pH 4.0
1.4M pH 4.0
1,6M pH 4.0
0.6M pH 4.3
0.8M pH 4.3
1.0M pH 4.3
1.2M pH 4.3
1.4M pH 4.3
1,6M pH 4.3
0.6M pH 4.6
0.8M pH 4.6
1.0M pH 4.6
1.2M pH 4.6
1.4M pH 4.6
1,6M pH 4.6
0.6M pH 4.9
0.8M pH 4.9
1.0M pH 4.9
1.2M pH 4.9
1.4M pH 4.9
1,6M pH 4.9

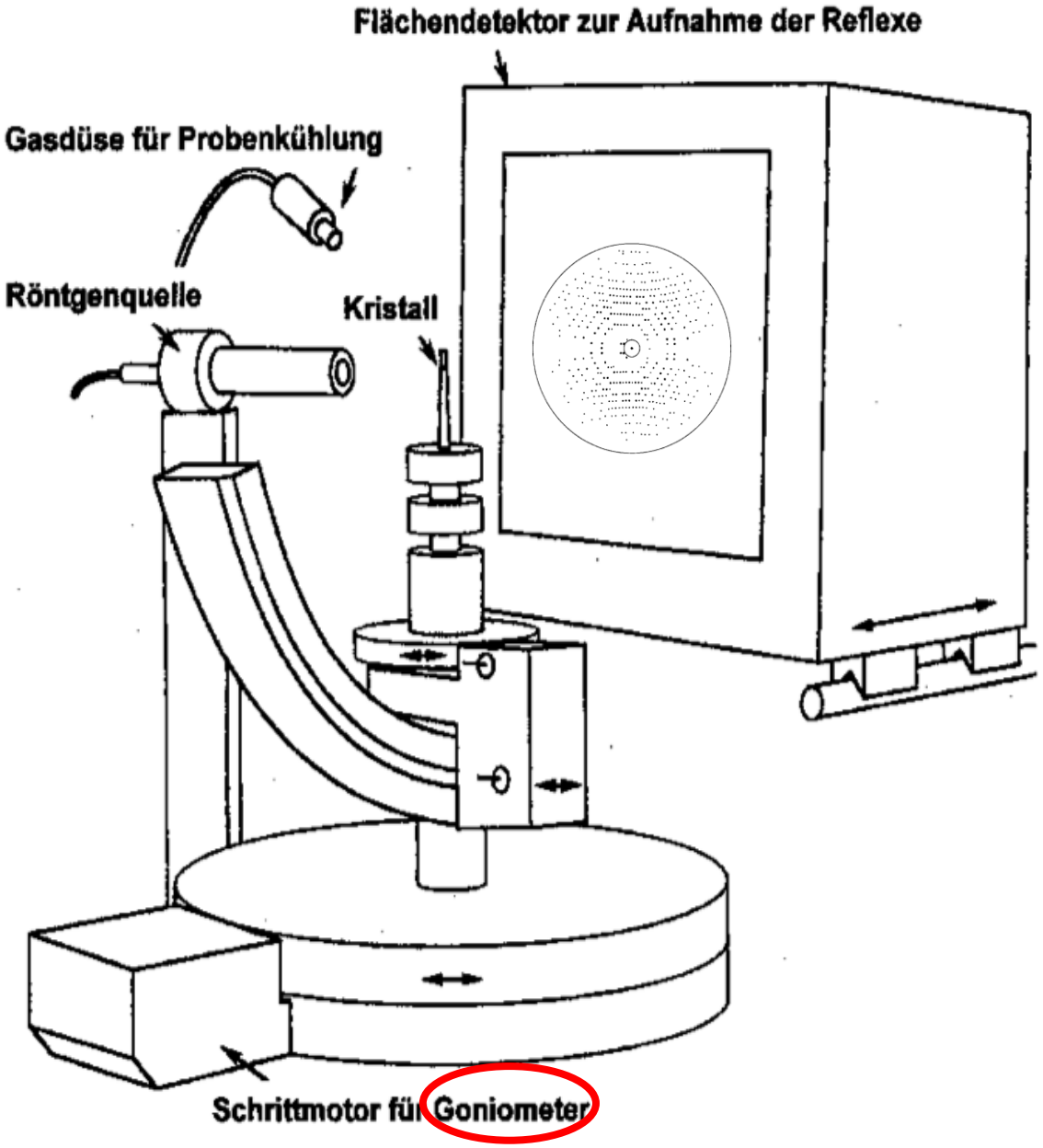
Phase diagram of solubility of a protein in solutions as a function of the concentration of the precipitant.




https://doi.org/10.3390/cryst8110434
http://skuld.bmsc.washington.edu/~merritt/bc530/local_copies/phase_methods_files/vd_xtals.jpg


Beugungsreflex
Detektor
Kristall
Röntgenstrahl

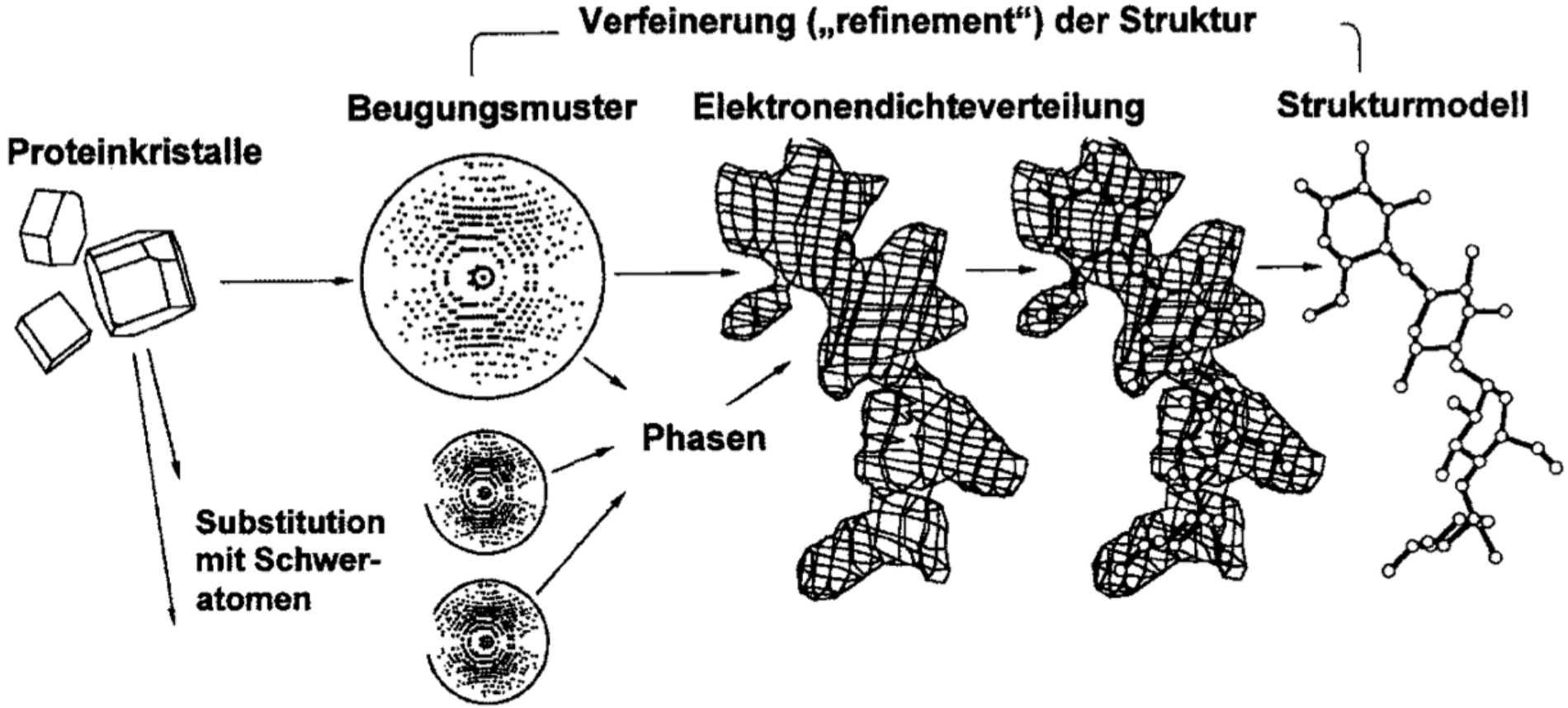


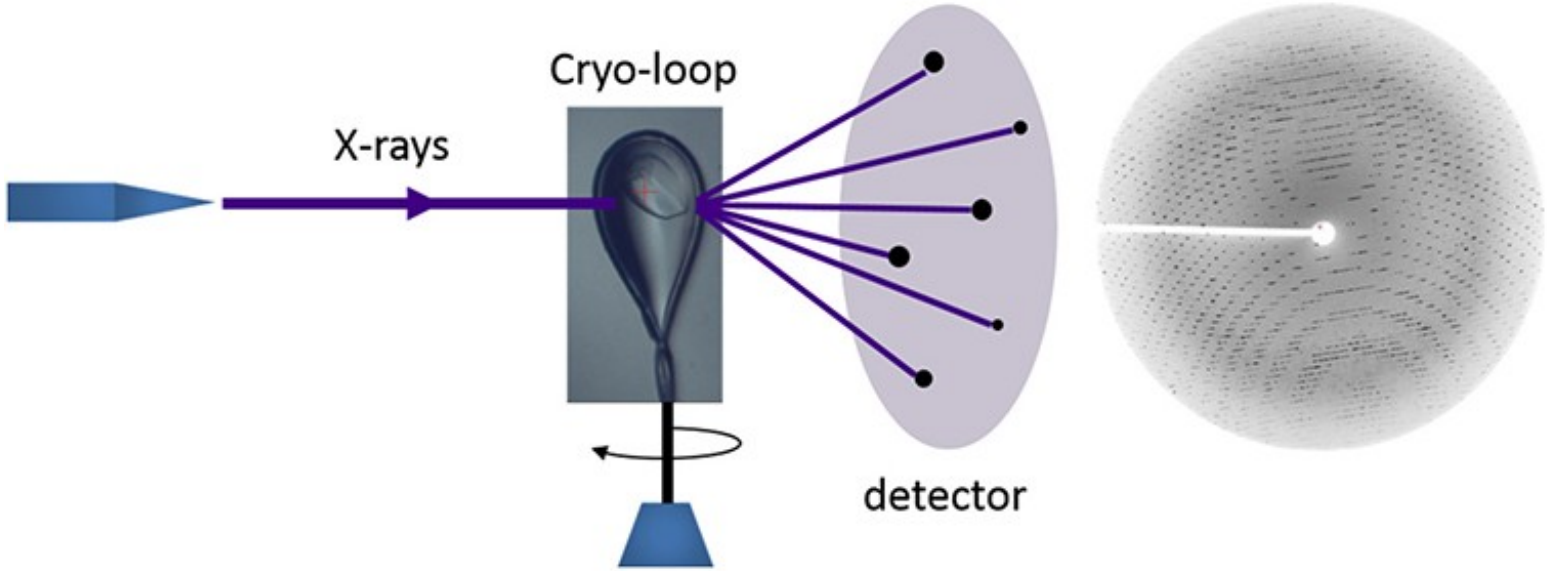

## X-Ray Goniometer

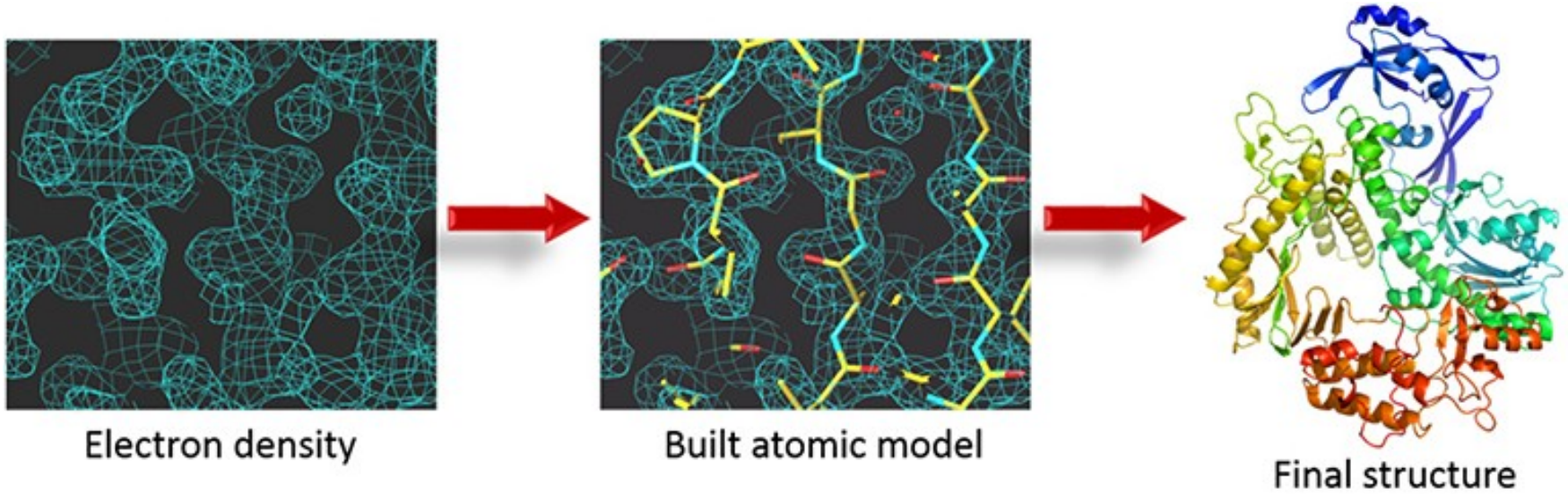
a device that **permits the simultaneous recording of the direction of X rays diffracted by a specimen** under study **and** of **the position of the specimen at the time of diffraction**.

An X-ray goniometer can be an independent device recording the diffraction pattern on photographic film; in this case it is an X-ray camera. The term “X-ray goniometer” is also applied to goniometric devices that are components of X-ray diffractometers and are used for mounting the specimen and detector in positions corresponding to the conditions necessary for the occurrence of X-ray diffraction.

https://encyclopedia2.thefreedictionary.com/X-Ray+Goniometer

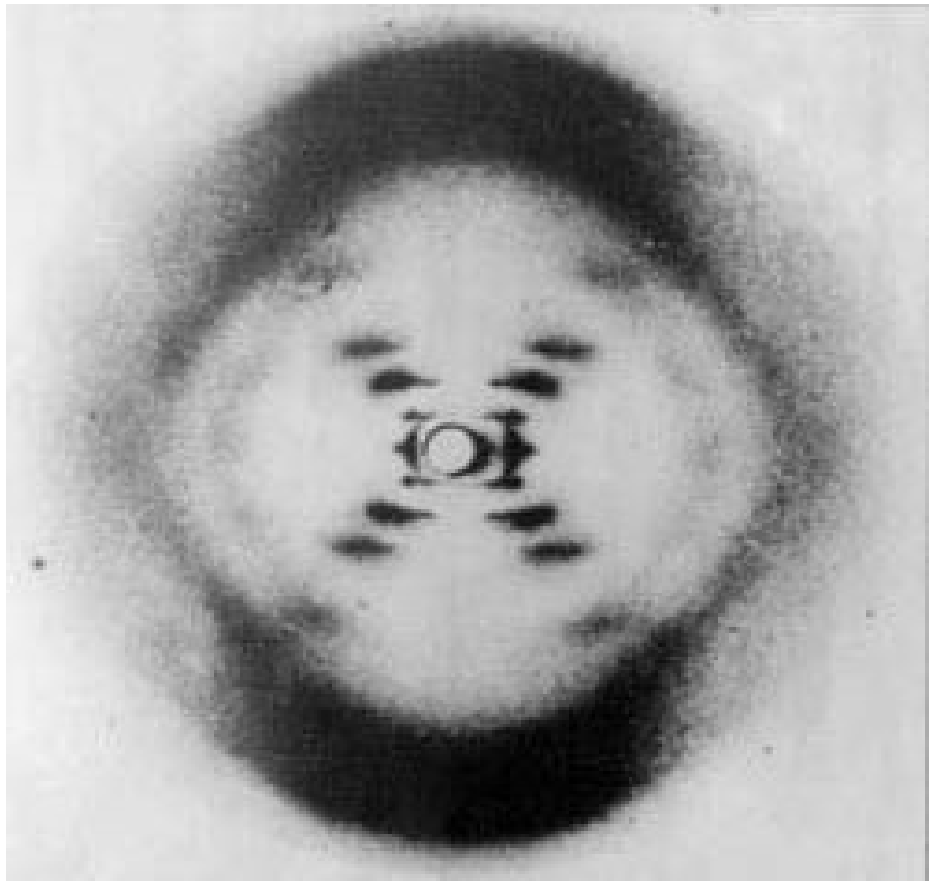

Verfeinerung („refinement") der Struktur
Beugungsmuster
Elektronendichteverteilung
Strukturmodell
Proteinkristalle
Phasen
Substitution
mit Schwer-
atomen




https://www.intechopen.com/books/x-ray-scattering/x-ray-diffraction-in-biology-how-can-we-see-dna-and-proteins-in-three-dimensions-

Krystalogram **B-DNA** získaný v **r. 1952** Rosalindou E. **Franklin**ovou, na jehož základě **Watson** a **Crick** navrhli dvoušroubovicový model struktury DNA. **C. & W.** dostali v **r.1962** společně s Mauricem Hugh Frederick Wilkinsem NC za fyziologii a medicínu „za jejich objevy týkající se molekulární struktury nukleových kyselin a jejich významu při přenosu informací v živých organizmech“



F

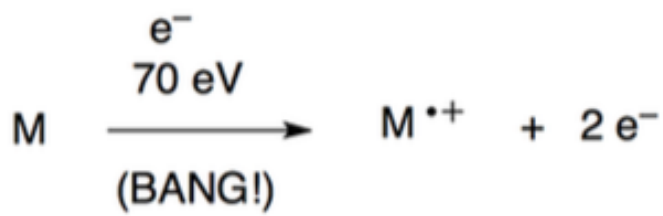
C

W

# Mass spectrometry

- Electron impact Mass spectrometry

Bombard your molecules with high energy electrons




- 70 eV = 1614 kcal/mol
  - contrast with energy from IR (1-10 kcal/mol) or NMR (0.2 cal/mol)
  - typical C-C bond = 100 kcal/mol
- Point: lots of energy in play here
  - you can eject electrons, break bonds, etc.

- don't call it spectroscopy (absorption of electromagnetic radiation)



http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

• Electron impact Mass spectrometry

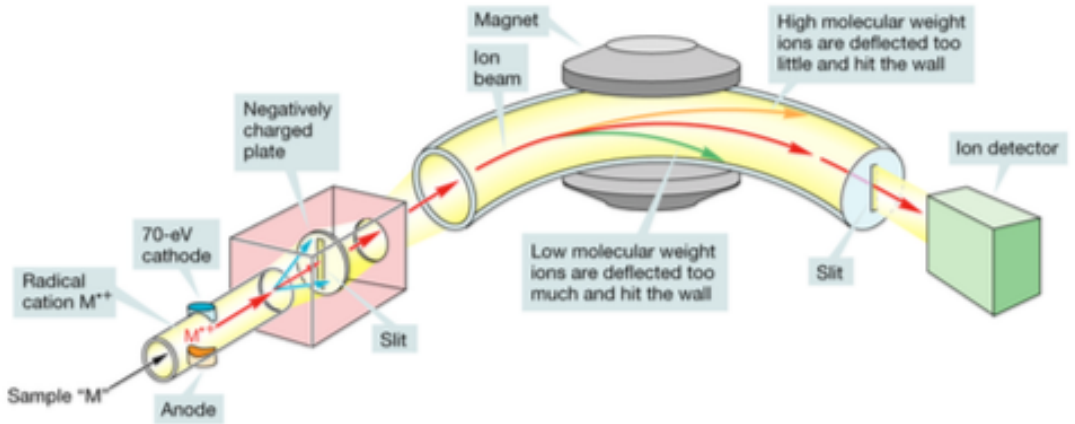



Figure 15.04

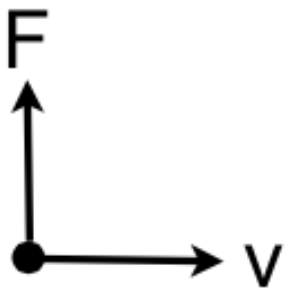

• Upon ionization, radical cations (M+) are accelerated toward a negatively charged plate with a slit. Some of the ions pass through the slit to form a beam.

• ions follow a curved path between poles of a magnet.

**uniform circular motion in the magnetic field**

**F = zvB**

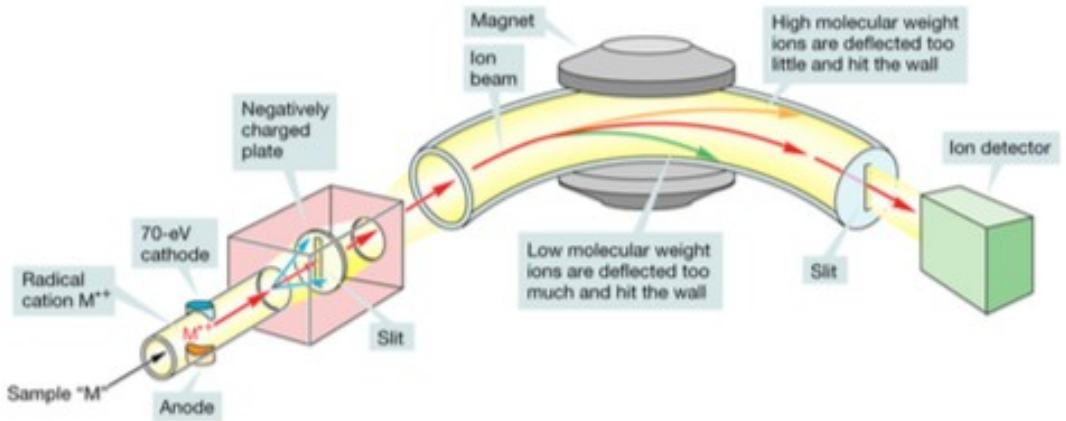
where z= point charge
v= velocity
B = Field stregnth




particles follow circular pathway as a function of time

http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

- Electron impact Mass spectrometry




$$F = \frac{mv^2}{R} = zvB$$

R = radius

**rearranged:**

$$\frac{m}{z} = \frac{RB}{v}$$

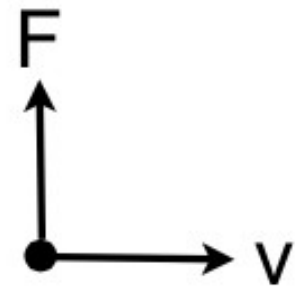
thus, we can observe different masses by holding velocity and radius constant, and varying B

usually, z = 1, so m/z = m

**uniform circular motion in the magnetic field**

**F = zvB**

where z= point charge
v= velocity
B = Field stregnth




particles follow circular pathway as a function of time

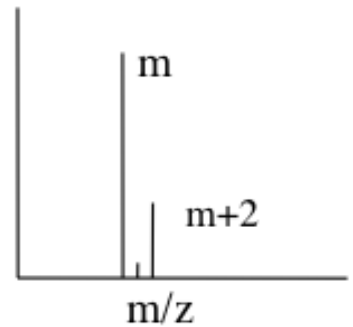
http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

# Molecular Weight vs Exact Mass

**Molecular Mass** refers the average mass of molecules made from their natural isotopic abundance:

**Exact Mass**: The mass of the most abundant isotopic form of a molecule.

**Example: $HOCH_2CH_2Cl$**




Molecular Weight:

2 x C: 2 x 12.011 : 24.022

1 x O: 1 x 15.999: 15.999

1 x Cl: 1 x 35.453: 35.453

H x 5: 5 x 1.008: 5.040

MW: 80.514

Exact Mass:

2 x C: 2 x 12.000 : 24.000

1 x O: 1 x 15.9949: 15.9949

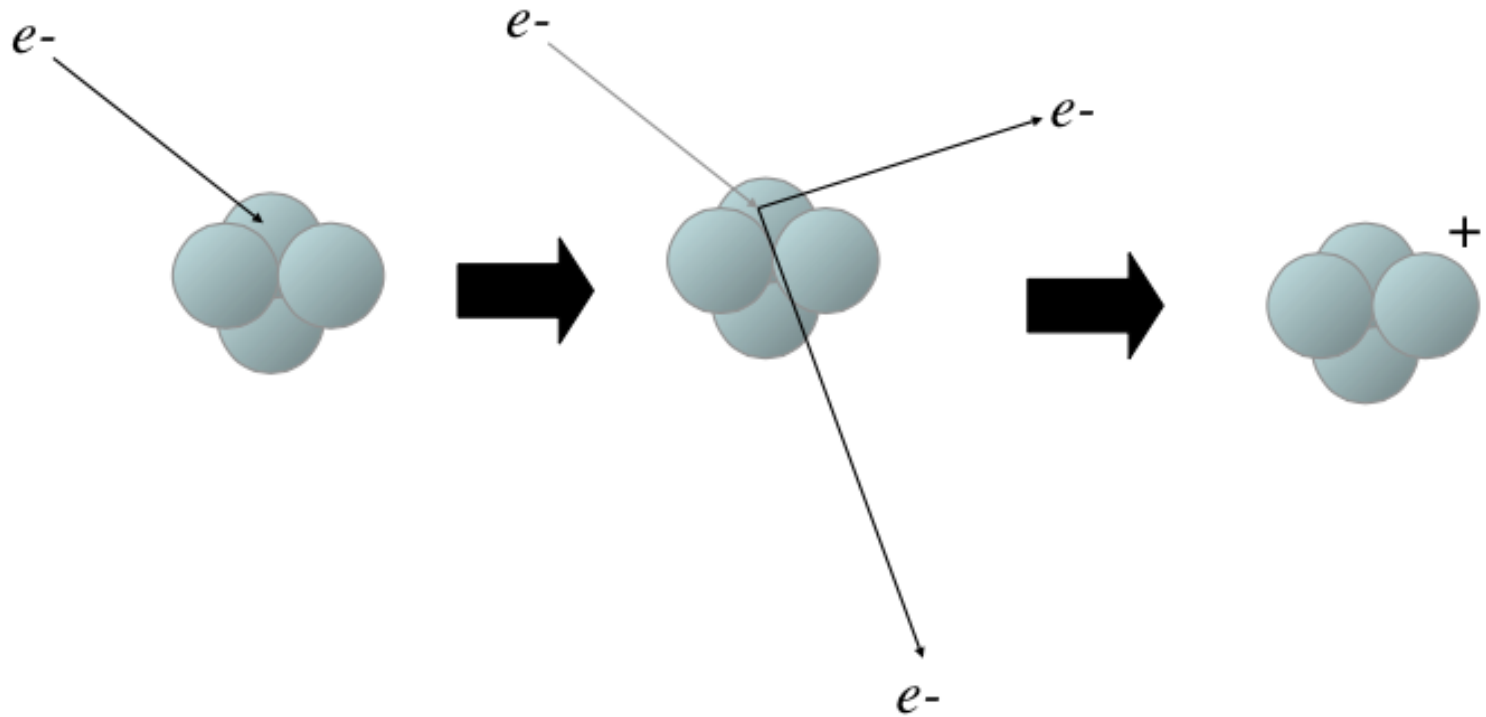
1 x Cl: 1 x 34.9689 : 34.9689

H x 5: 5 x 1.0078: 5.0390

Exact Mass: 80.003

http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

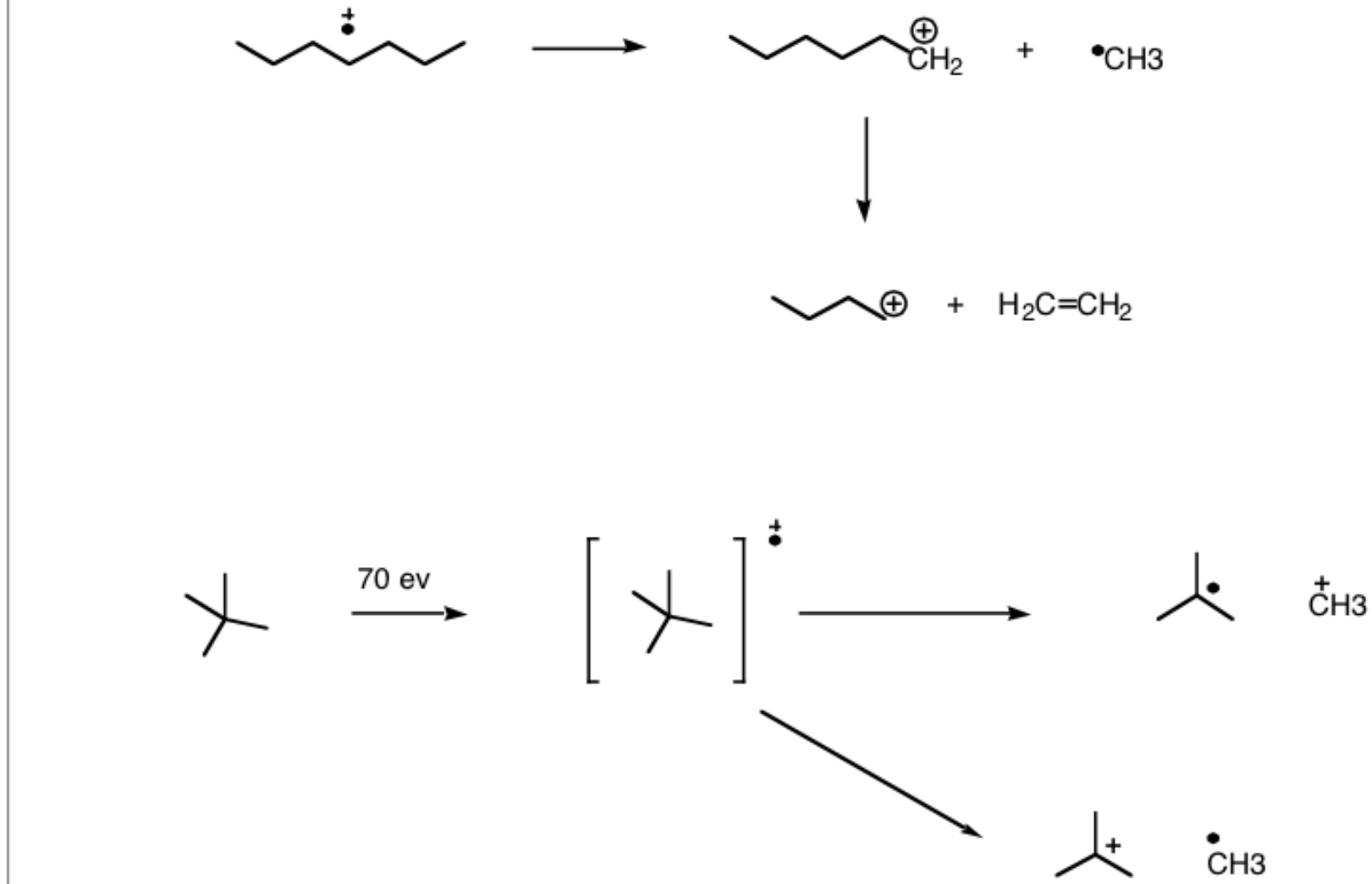
# Ionization:




- Lone pair electrons are more easily displaced than bonding electrons.
- Electrons in pi-bonds are more easily displaced than those in single bonds

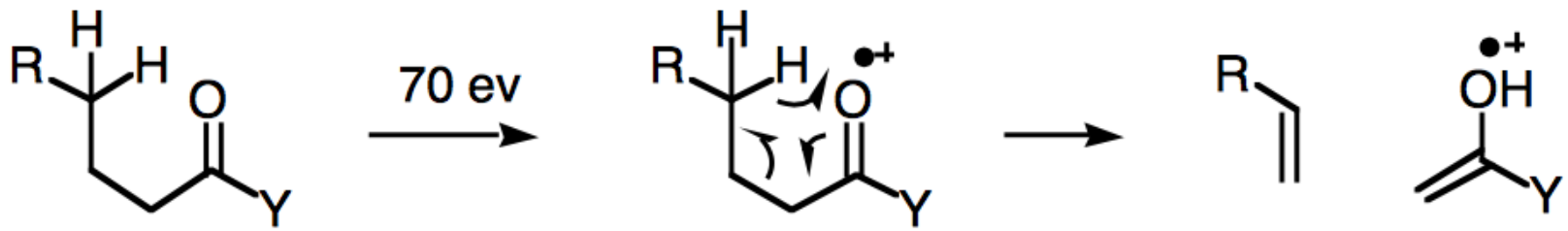
http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

# Fragmentation : Alkanes

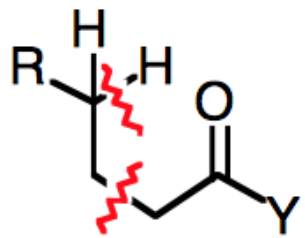



http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

# The McLafferty Rearrangment:




*must have gamma proton*




http://www.udel.edu/chem/fox/MassSpecLecture09.pdf

## Commonly Lost Fragments

m-15 $\cdot CH_3$

m-17 $\cdot OH$

m-26 $\cdot CN$

m-28 $H_2C{=}CH_2$

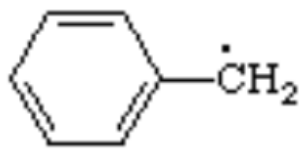
m-29 $\cdot CH_2CH_3$ $\cdot CHO$

m-31 $\cdot OCH_3$

m-35 $\cdot Cl$

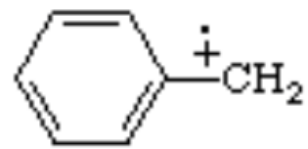
m-43 $CH_3\dot{C}{=}O$

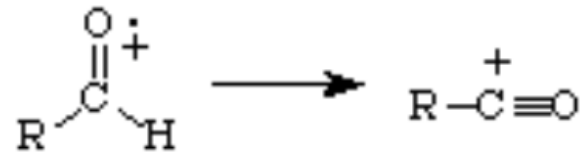
m-45 $\cdot OCH_2CH_3$

m-91 

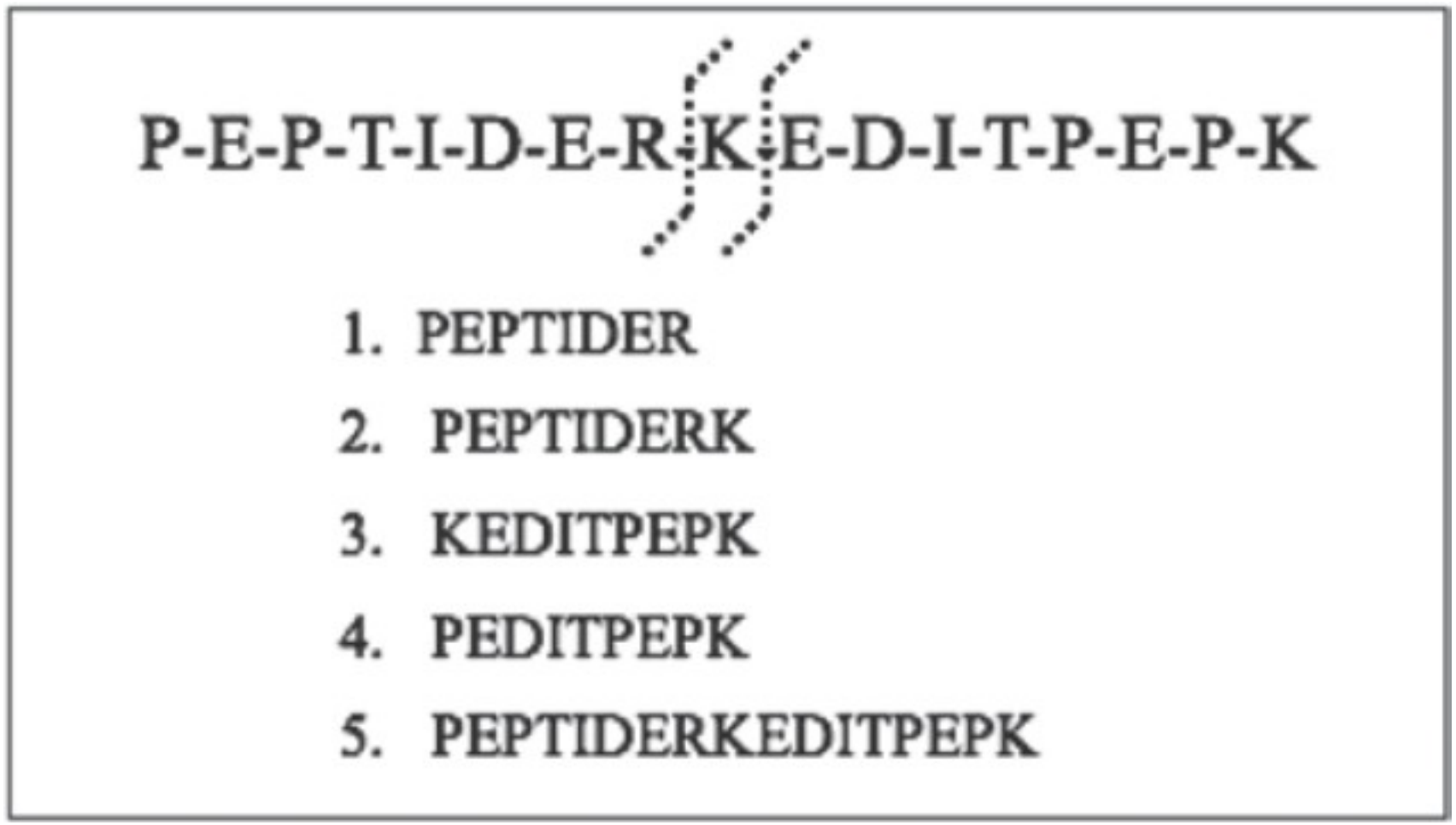
## Common Stable Ions

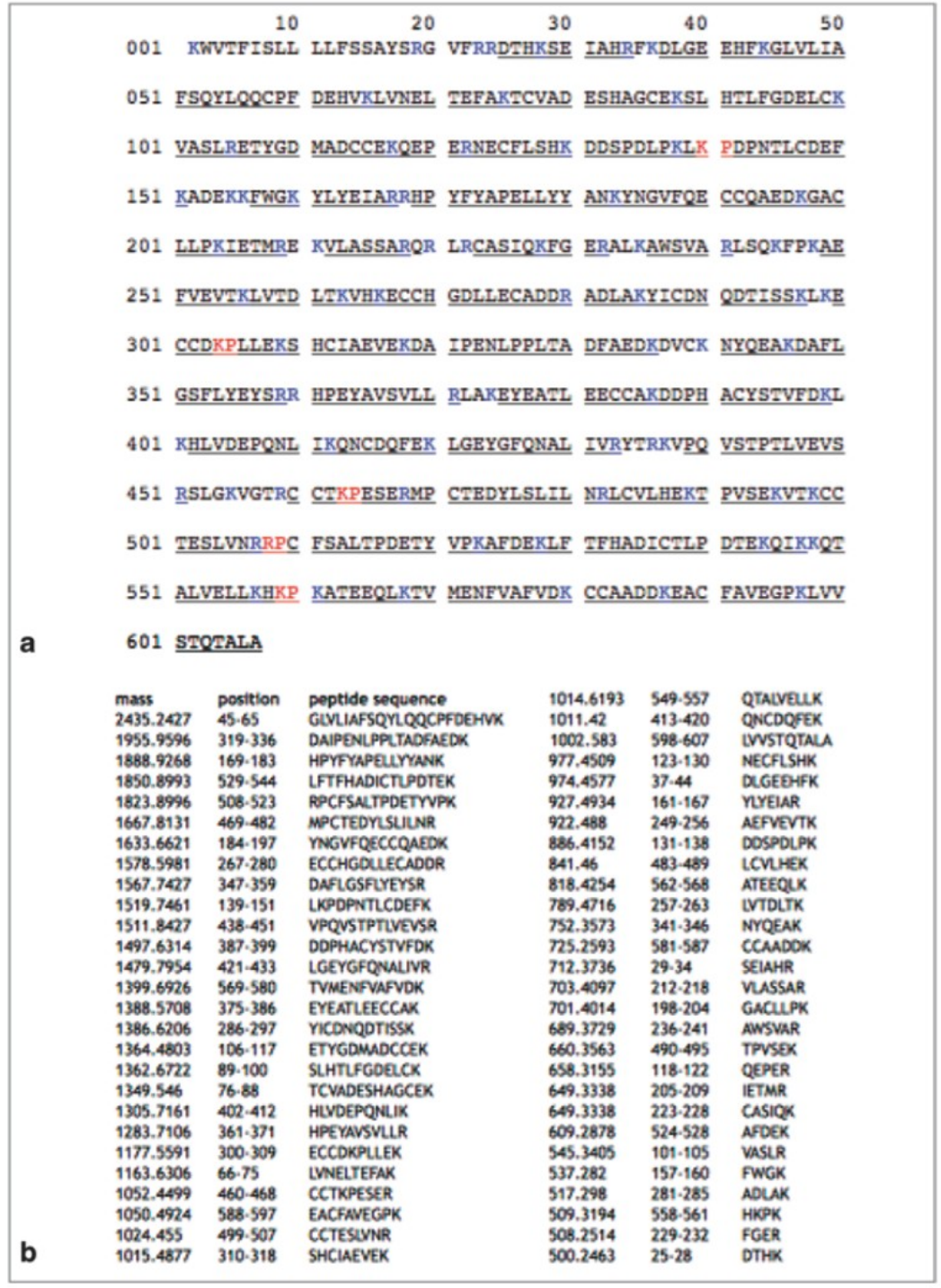
m/e = 43 $CH_3{-}C{\equiv}O^+$

m/e = 91 

m/e = m-1 

http://www.udel.edu/chem/fox/MassSpecLecture09.pdf


P-E-P-T-I-D-E-R-K-E-D-I-T-P-E-P-K
1. PEPTIDER
2. PEPTIDERK
3. KEDITPEPK
4. PEDITPEPK
5. PEPTIDERKEDITPEPK




a

| mass | position | peptide sequence | 1014.6193 | 549-557 | QTALVELLK |
|---|---|---|---|---|---|
| 2435.2427 | 45-65 | GLVLIAFSQYLQQCPFDEHVK | 1011.42 | 413-420 | QNCDQFEK |
| 1955.9596 | 319-336 | DAIPENLPPLTADFAEDK | 1002.583 | 598-607 | LVVSTQTALA |
| 1888.9268 | 169-183 | HPYFYAPELLYYANK | 977.4509 | 123-130 | NECFLSHK |
| 1850.8993 | 529-544 | LFTFHADICTLPDTEK | 974.4577 | 37-44 | DLGEEHFK |
| 1823.8996 | 508-523 | RPCFSALTPDETYVPK | 927.4934 | 161-167 | YLYEIAR |
| 1667.8131 | 469-482 | MPCTEDYLSLILNR | 922.488 | 249-256 | AEFVEVTK |
| 1633.6621 | 184-197 | YNGVFQECCQAEDK | 886.4152 | 131-138 | DDSPDLPK |
| 1578.5981 | 267-280 | ECCHGDLLECADDR | 841.46 | 483-489 | LCVLHEK |
| 1567.7427 | 347-359 | DAFLGSFLYEYSR | 818.4254 | 562-568 | ATEEQLK |
| 1519.7461 | 139-151 | LKPDPNTLCDEFK | 789.4716 | 257-263 | LVTDLTK |
| 1511.8427 | 438-451 | VPQVSTPTLVEVSR | 752.3573 | 341-346 | NYQEAK |
| 1497.6314 | 387-399 | DDPHACYSTVFDK | 725.2593 | 581-587 | CCAADDK |
| 1479.7954 | 421-433 | LGEYGFQNALIVR | 712.3736 | 29-34 | SEIAHR |
| 1399.6926 | 569-580 | TVMENFVAFVDK | 703.4097 | 212-218 | VLASSAR |
| 1388.5708 | 375-386 | EYEATLEECCAK | 701.4014 | 198-204 | GACLLPK |
| 1386.6206 | 286-297 | YICDNQDTISSK | 689.3729 | 236-241 | AWSVAR |
| 1364.4803 | 106-117 | ETYGDMADCCEK | 660.3563 | 490-495 | TPVSEK |
| 1362.6722 | 89-100 | SLHTLFGDELCK | 658.3155 | 118-122 | QEPER |
| 1349.546 | 76-88 | TCVADESHAGCEK | 649.3338 | 205-209 | IETMR |
| 1305.7161 | 402-412 | HLVDEPQNLIK | 649.3338 | 223-228 | CASIQK |
| 1283.7106 | 361-371 | HPEYAVSVLLR | 609.2878 | 524-528 | AFDEK |
| 1177.5591 | 300-309 | ECCDKPLLEK | 545.3405 | 101-105 | VASLR |
| 1163.6306 | 66-75 | LVNELTEFAK | 537.282 | 157-160 | FWGK |
| 1052.4499 | 460-468 | CCTKPESER | 517.298 | 281-285 | ADLAK |
| 1050.4924 | 588-597 | EACFAVEGPK | 509.3194 | 558-561 | HKPK |
| 1024.455 | 499-507 | CCTESLVNR | 508.2514 | 229-232 | FGER |
| 1015.4877 | 310-318 | SHCIAEVEK | 500.2463 | 25-28 | DTHK |

b

**PeptideMass**

The entered sequence is:

```
        10         20         30         40         50         60
MQQDDDFQNF VATLESFKDL KSGISGSRIK KLTTYALDHI DIESKIISLI IDYSRLCPDS

        70         80         90        100        110        120
HKLGSLYIID SIGRAYLDET RSNSNSSSNK PGTCAHAINT LGEVIQELLS DAIAKSNQDH

       130        140        150        160
KEKIRMLLDI WDRSGLFQKS YLNAIRSKCF AMDLEHHHHH
```

The selected enzyme is: Trypsin

Maximum number of missed cleavages (MC): 0

All cysteines in reduced form.

Methionines have not been oxidized.

Displaying peptides with a mass bigger than 500 Dalton.

Using monoisotopic masses of the occurring amino acid residues and giving peptide masses as $[M+H]^+$.

---

**The peptide masses from your sequence are:**

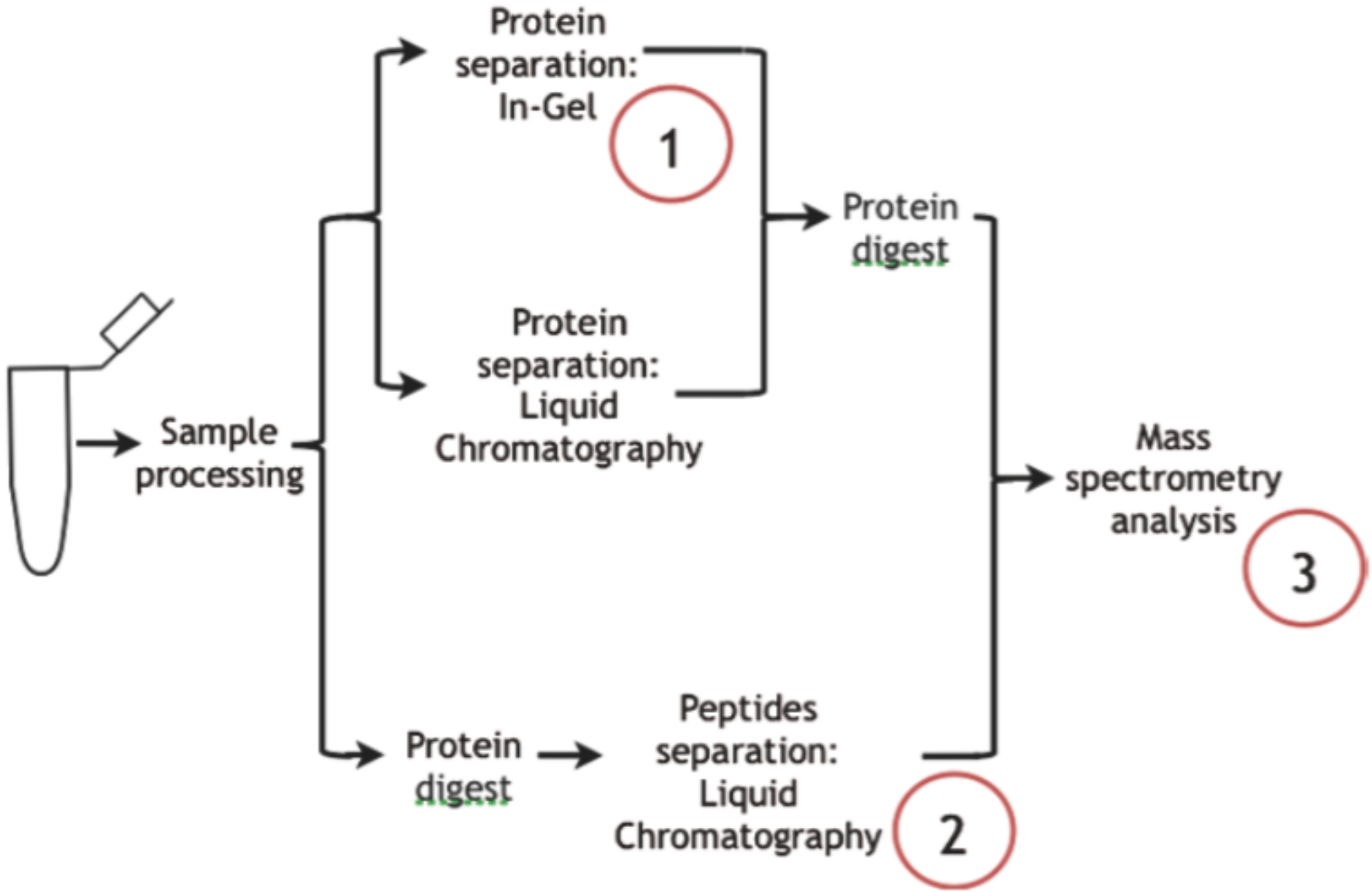
[Theoretical pI: 6.29 / Mw (average mass): 18149.49 / Mw (monoisotopic mass): 18138.14]

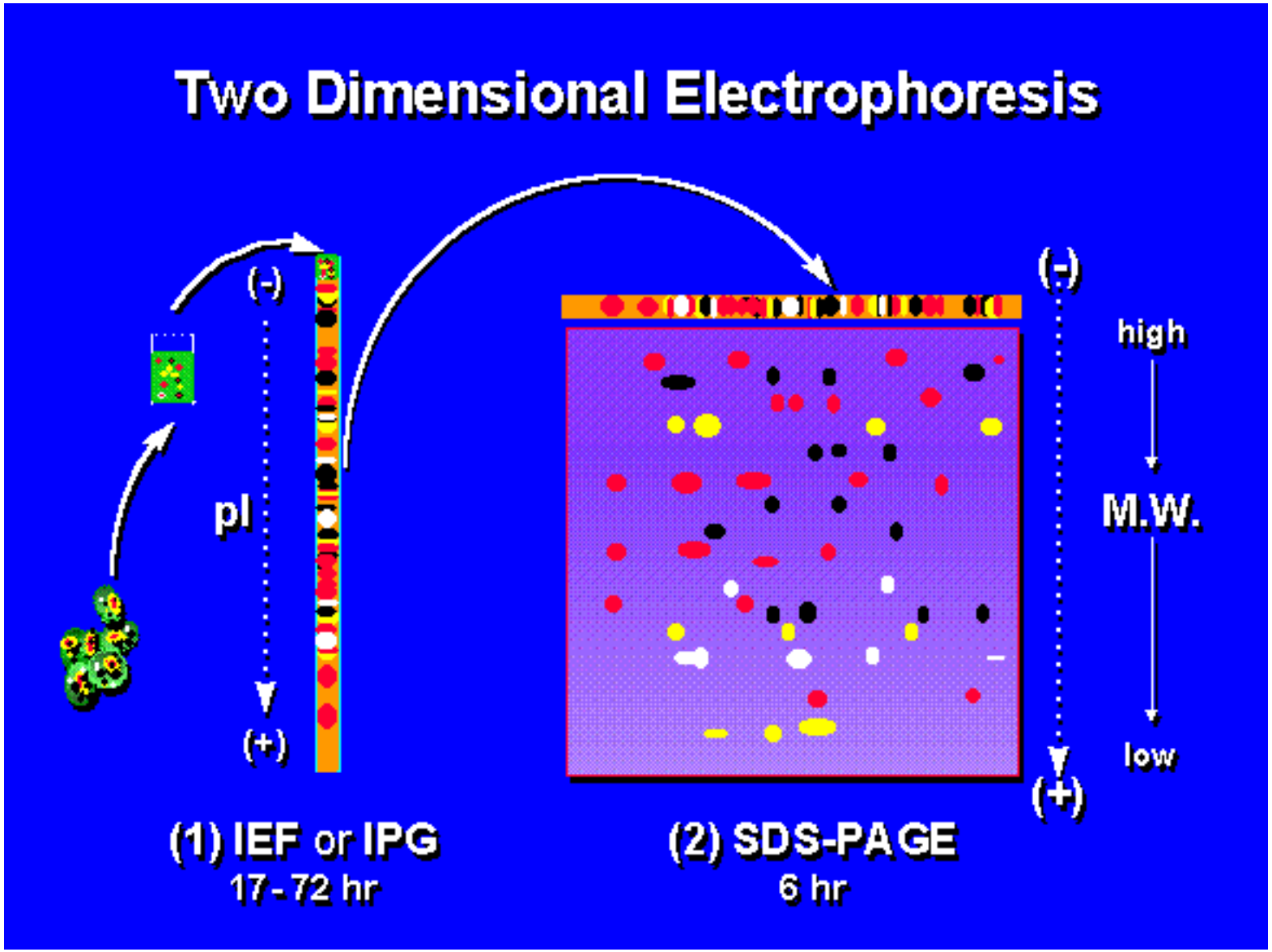
| mass | position | #MC | modifications | peptide sequence |
|---|---|---|---|---|
| 3469.7227 | 82-115 | 0 | | SNSNSSSNKPGTCAHAINTL GEVIQELLSDAIAK |
| 2162.9699 | 1-18 | 0 | | MQQDDDFQNFVATLESFK |
| 1618.8322 | 32-45 | 0 | | LTTYALDHIDIESK |
| 1513.6212 | 149-160 | 0 | | CFAMDLEHHHHH |
| 1306.7365 | 63-74 | 0 | | LGSLYIIDSIGR |
| 1192.6936 | 46-55 | 0 | | IISLIIDYSR |
| 1061.5448 | 126-133 | 0 | | MLLDIWDR |
| 867.4206 | 75-81 | 0 | | AYLDETR |
| 836.4625 | 140-146 | 0 | | SYLNAIR |
| 799.3767 | 56-62 | 0 | | LCPDSHK |
| 728.3322 | 116-121 | 0 | | SNQDHK |
| 679.3773 | 134-139 | 0 | | SGLFQK |
| 663.3420 | 22-28 | 0 | | SGISGSR |

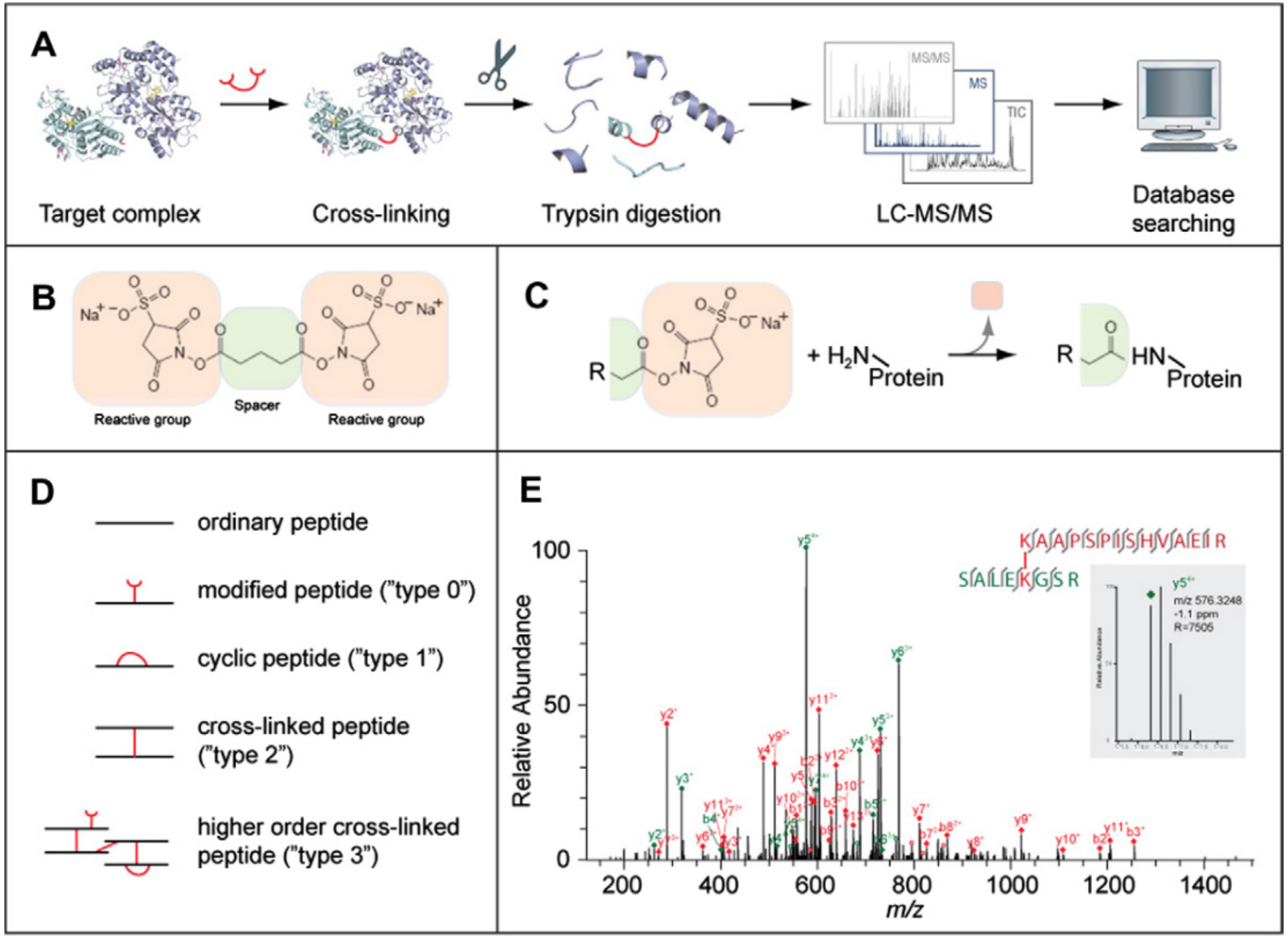
92.5% of sequence covered (you may modify the input parameters to display also peptides < 500 Da or > 100000000000 Da):

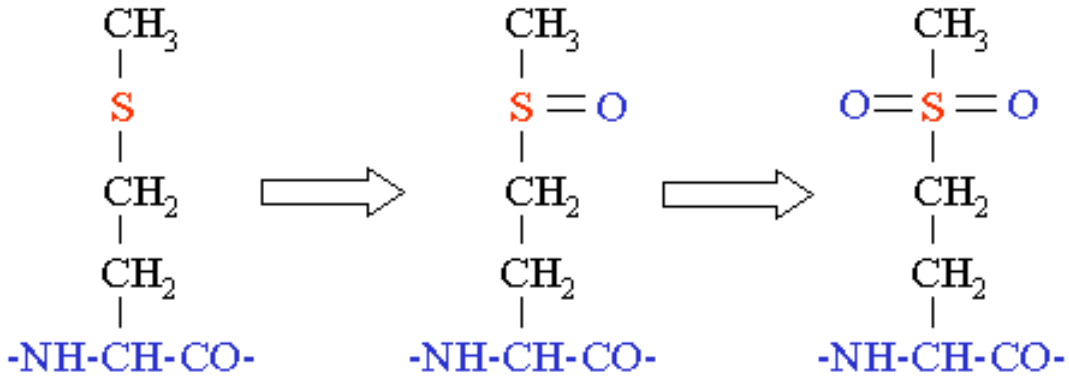
```
        10         20         30         40         50         60
MQQDDDFQNF VATLESFKdl kSGISGSRik kLTTYALDHI DIESKIISLI IDYSRLCPDS

        70         80         90        100        110        120
HKLGSLYIID SIGRAYLDET RSNSNSSSNK PGTCAHAINT LGEVIQELLS DAIAKSNQDH

       130        140        150        160
KekirMLLDI WDRSGLFQKS YLNAIRskCF AMDLEHHHHH
```

Display the list of masses in raw text format to be exported into an external application

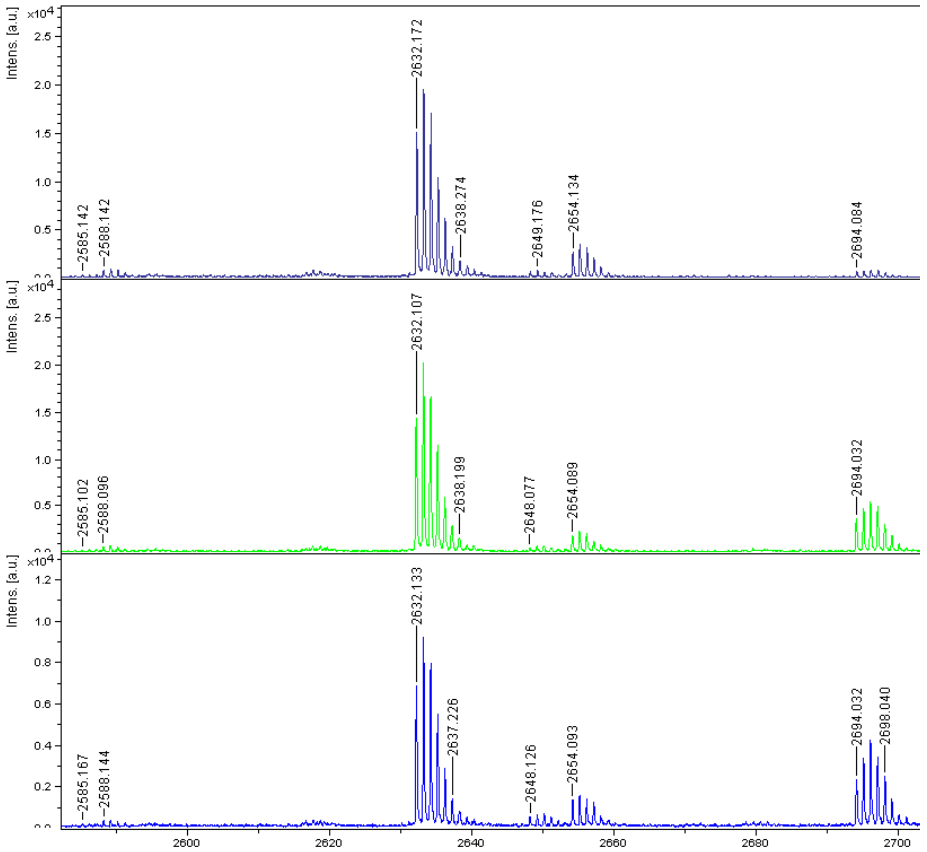

Protein separation: In-Gel
1
Protein digest
Protein separation: Liquid Chromatography
Sample processing
Mass spectrometry analysis
3
Protein digest
Peptides separation: Liquid Chromatography
2


Two Dimensional Electrophoresis
(-)
pI
(+)
(1) IEF or IPG
17 - 72 hr
(-)
high
M.W.
low
(+)
(2) SDS-PAGE
6 hr

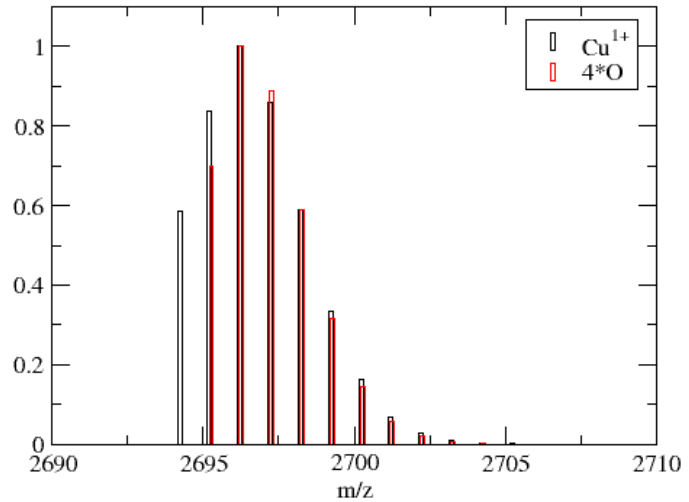

A
Target complex
Cross-linking
Trypsin digestion
MS/MS
MS
TIC
LC-MS/MS
Database searching
B
Reactive group
Spacer
Reactive group
C
+ H2N Protein
HN Protein
D
ordinary peptide
modified peptide ("type 0")
cyclic peptide ("type 1")
cross-linked peptide ("type 2")
higher order cross-linked peptide ("type 3")
E
Relative Abundance
m/z
K A A P S P I S H V A E I R
S A L E K G S R
y5 4+
m/z 576.3248
-1.1 ppm
R=7506

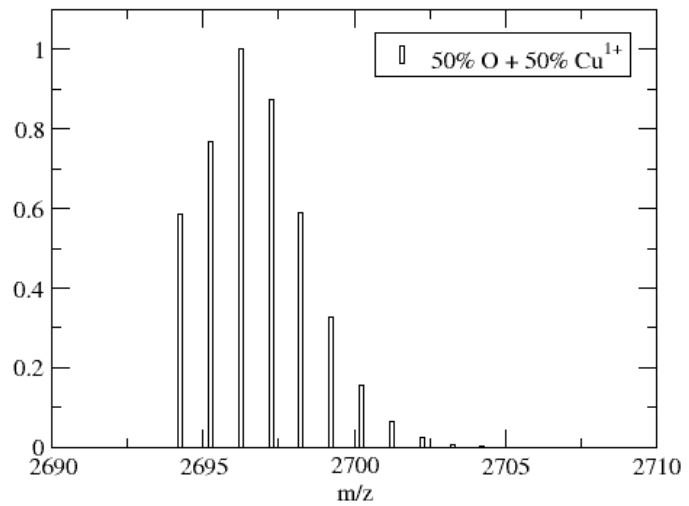



| Amino Acid | Residue Composition | Residue Monoisotopic Mass | Delta Mass |
|---|---|---|---|
| Methionine | $C_5H_9NOS$ | 131.0405 | 0 |
| Methionine Sulfoxide | $C_5H_9NO_2S$ | 147.0354 | 15.9949 |
| Methionine Sulfone | $C_5H_9NO_3S$ | 163.0303 | 31.9898 |
| Sulfur | S | 31.9721 | - |


Intens. [a.u.]
2585.142
2588.142
2632.172
2638.274
2649.176
2654.134
2694.084
2585.102
2588.096
2632.107
2638.199
2648.077
2654.089
2694.032
2585.167
2588.144
2632.133
2637.226
2648.126
2654.093
2694.032
2698.040
2600
2620
2640
2660
2680
2700

## L61-K87


Cu1+
4*O
m/z


50% O + 50% Cu1+
m/z

Mass Spectrometry

ESI - electron spray ionization – high voltage applied to a liquid jet producing highly charged droplets

MALDI - matrix-assisted laser desorption/ionization

- three step process:

1) sample mix with matrix and deposition on a metal plate,
2) laser pulse desorbs the sample with matrix,
3) analyte molecules are ionized and analyzed (TOF – time of flight MS technique)

Quadrupole mass analyzer

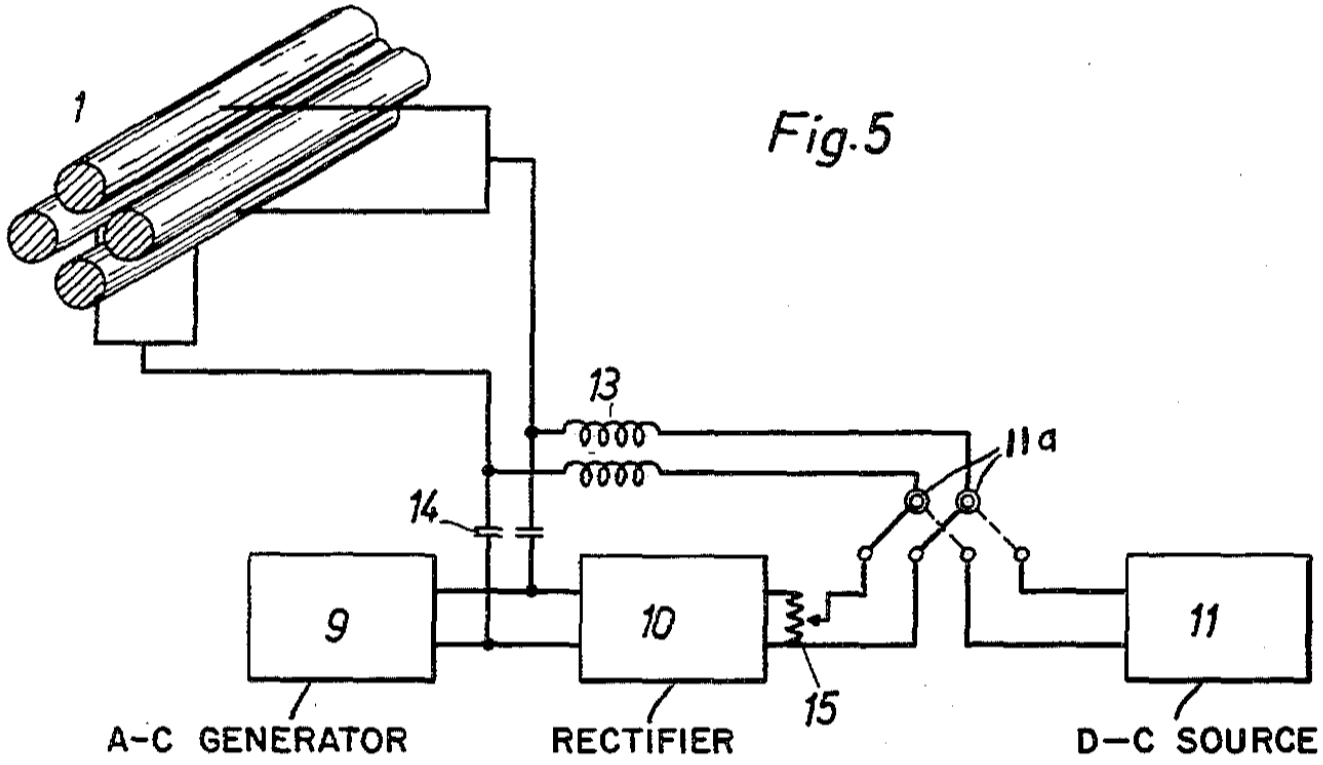
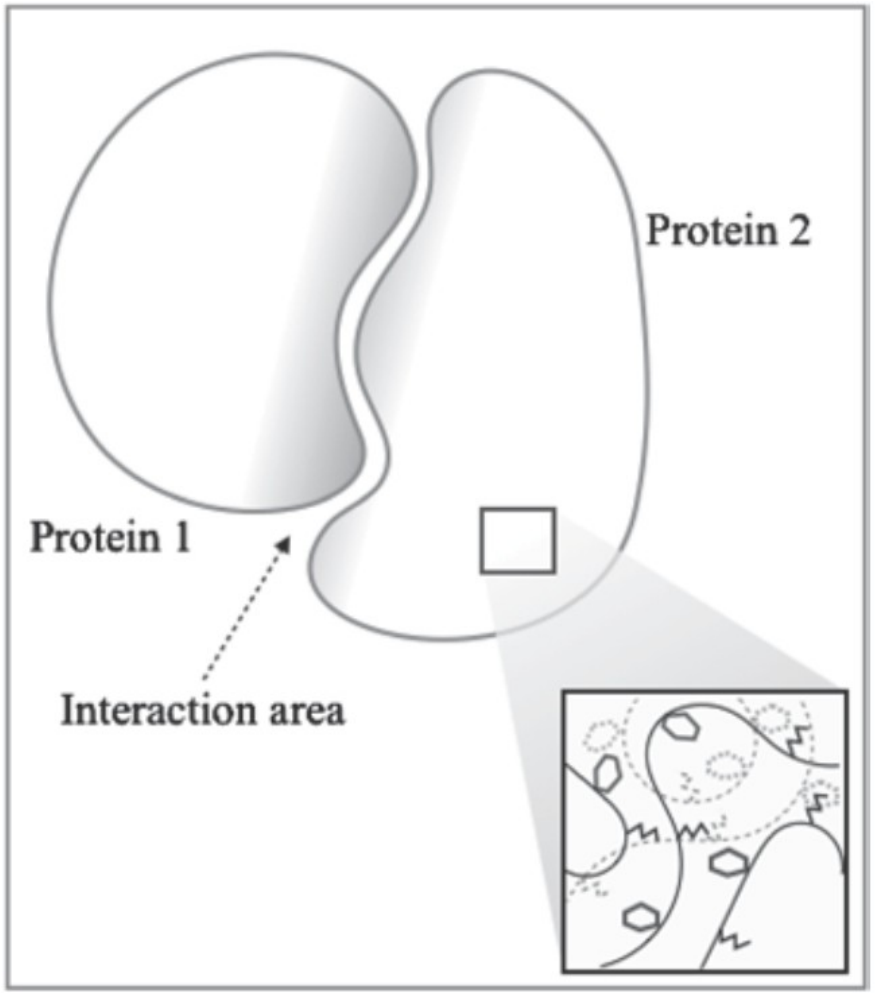



Fig.5

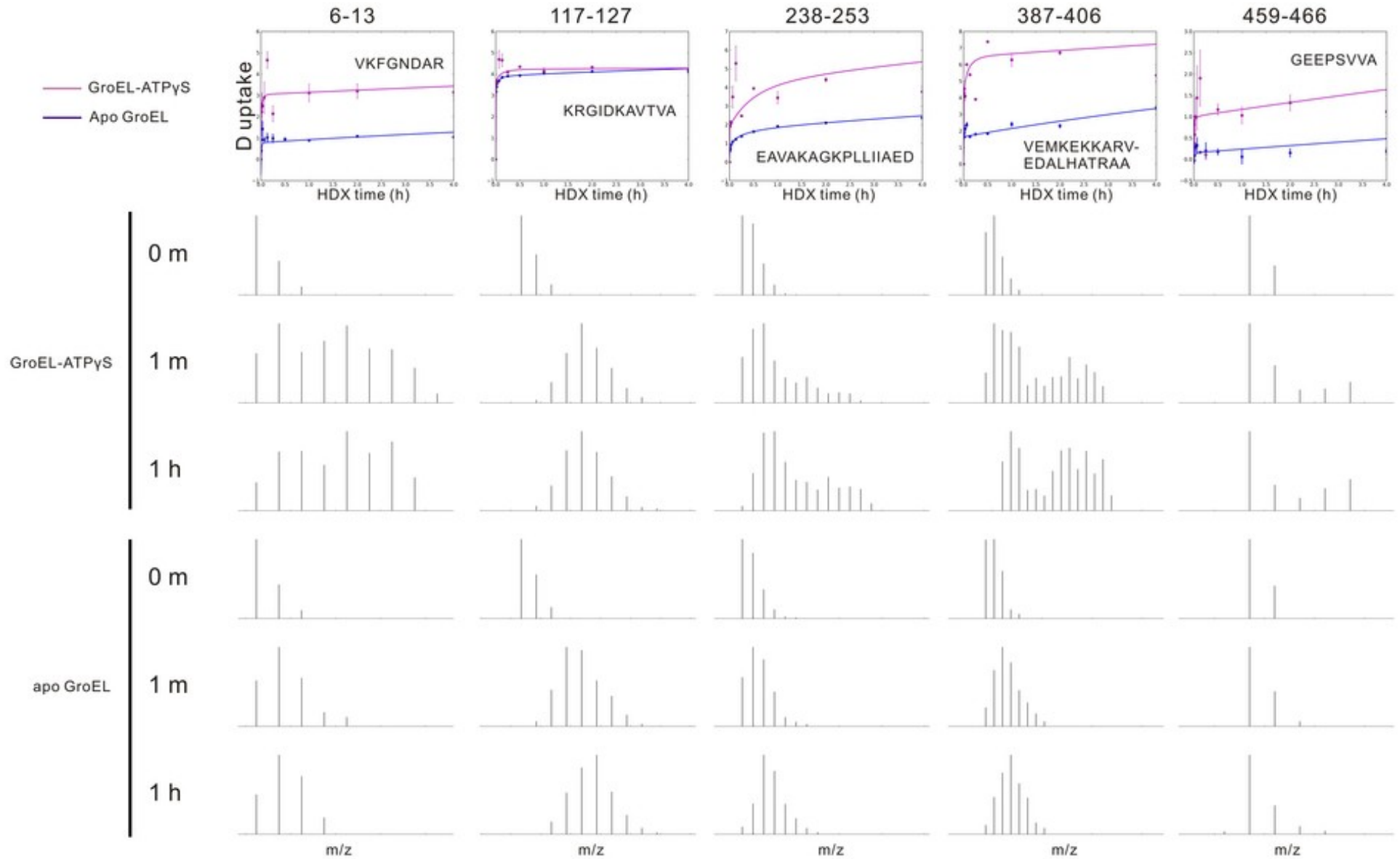
**UV MALDI Matrix List**

| Compound | Other Names | Solvent | Wavelength (nm) | Applications |
|---|---|---|---|---|
| **2,5-dihydroxy benzoic acid**[9] | DHB, Gentisic acid | acetonitrile, water, methanol, acetone, chloroform | 337, 355, 266 | peptides, nucleotides, oligonucleotides, oligosaccharides |
| **3,5-dimethoxy-4-hydroxycinnamic acid**[7][10] | sinapic acid; sinapinic acid; SA | acetonitrile, water, acetone, chloroform | 337, 355, 266 | peptides, proteins, lipids |
| **4-hydroxy-3-methoxycinnamic acid**[7][10] | ferulic acid | acetonitrile, water, propanol | 337, 355, 266 | proteins |
| **α-Cyano-4-hydroxycinnamic acid**[11] | CHCA | acetonitrile, water, ethanol, acetone | 337, 355 | peptides, lipids, nucleotides |
| **Picolinic acid**[12] | PA | Ethanol | 266 | oligonucleotides |
| **3-hydroxy picolinic acid**[13] | HPA | Ethanol | 337, 355 | oligonucleotides |

HDX MS - Hydrogen Deuterium Exchange with Mass Spec



# ARDD - average relative D-uptake difference




*Scientific Reports* **3**, Article number: 1247, 2013

## Performing Hydrogen/Deuterium Exchange with Mass Spectrometry

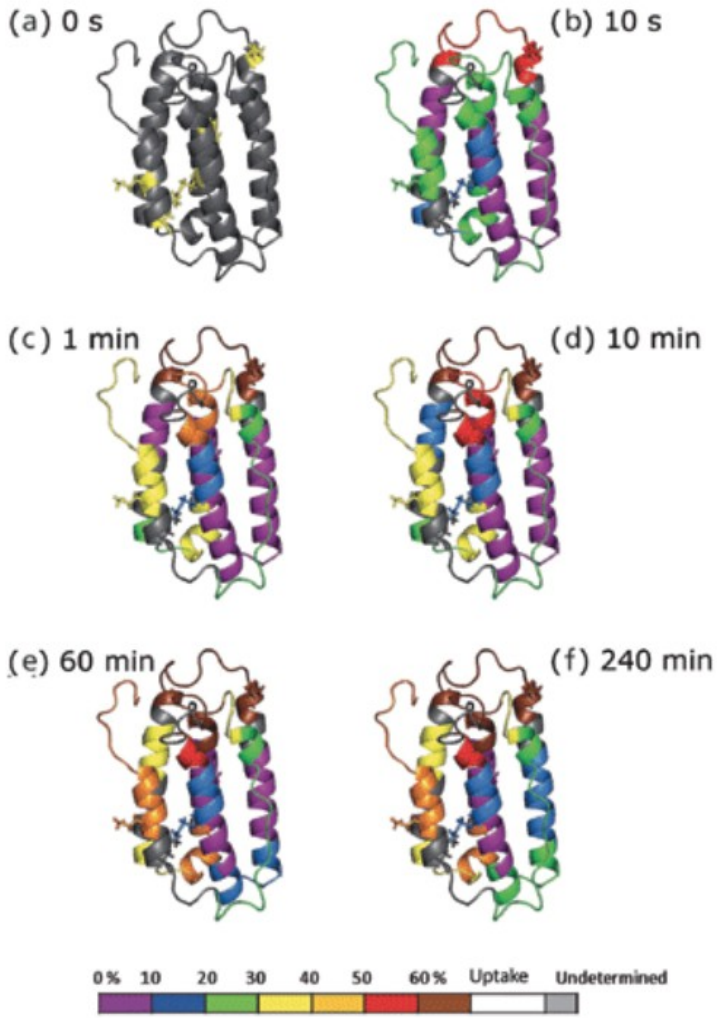


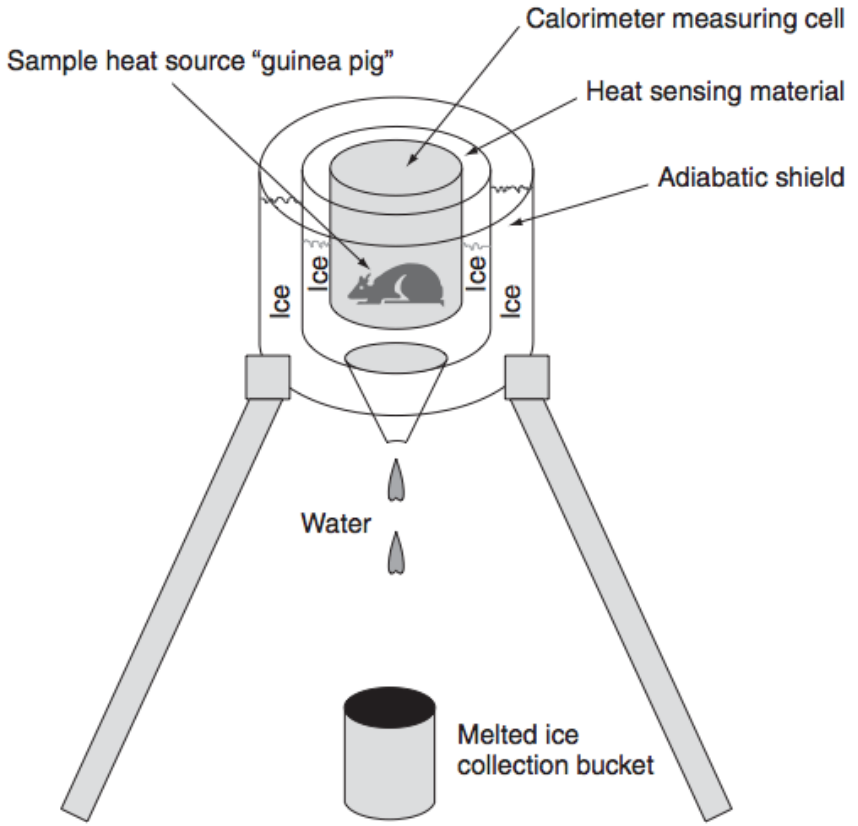

Figure 1 (HDX/MS): A depiction of the relative deuterium uptake for interferon helps one visualize and interpret the higher order protein structure related to conformational change. The uptake measurements are made at the peptide level for multiple time points across the experiment. Each uptake measurement is superimposed on the 3D structure of the protein, typically obtained from an X-ray representation.

http://www.biopharminternational.com/performing-hydrogendeuterium-exchange-mass-spectrometry

ITC – isothermal titration calorimetry

- 1760s – Black measured the heat capacity and latent heat of water
- 1780s – Lavoisier designed an ice calorimeter and used this instrument to measure the metabolic heat produced by a guinea pig confined in the measurement chamber

=> Calorimeter was one of the earliest scientific instruments & first calorimetric experiment was a biologically relevant measurement



# Why ITC?

1) ITC is a quantitative technique
2) can determine:
    I. binding affinity (**$K_a$**),
    II. enthalpy changes (**ΔH**),
    III. binding stoichiometry (**n**) of the interaction between two or more molecules in solution.

From these initial measurements:

3) Gibbs energy changes (**ΔG**)
4) entropy changes (**ΔS**)

$$\Delta G = -RT\ln K_a = \Delta H - T\Delta S$$

$$Q = V_0 \Delta H_b [M]_t K_a [L] / (1 + K_a [L])$$

$$Q = V_0 [M]_t \Sigma (n_i \Delta H_i K_{ai} [L]) / (1 + K_{ai} [L]).$$

# Thermodynamics

# Free energy change

$K_B$ – binding constant

$$K_D = 1/K_B = \frac{[L] \times [M]}{[ML]}$$

$$\Delta G = RT \ln K_D$$

$$\Delta G = \Delta H - T\Delta S$$

- $\Delta G$ is change in free energy
- $\Delta G \leq 0$ for spontaneous process
- More negative $\Delta G$, higher affinity

# Enthalpy change

· $\Delta H$ – measure of the energy content of the bonds broken and created. The dominant contribution is from hydrogen bonds.

· Negative value indicates enthalpy change favoring the binding

· Solvents play a role

$\Delta H_{observed}$ by ITC is total of :

$\Delta H_{binding}$

$\Delta H_{ionization}$

$\Delta H_{conformation}$

# Entropy change

· $\Delta S$ – positive for entropically driven reactions

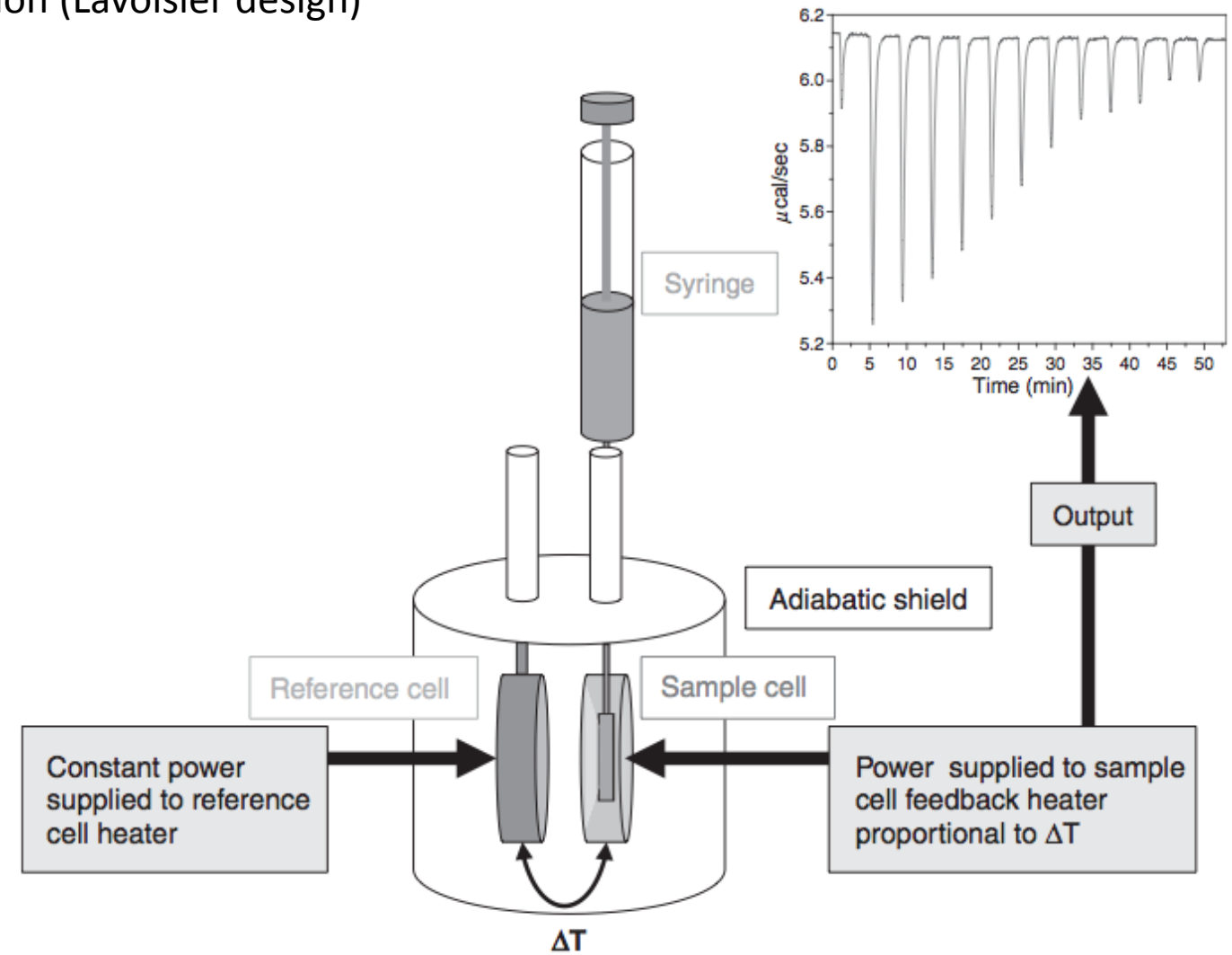
· Hydrophobic interactions

· Solvation entropy (favorable) due to release of water

· Conformational degrees of freedom (unfavorable)

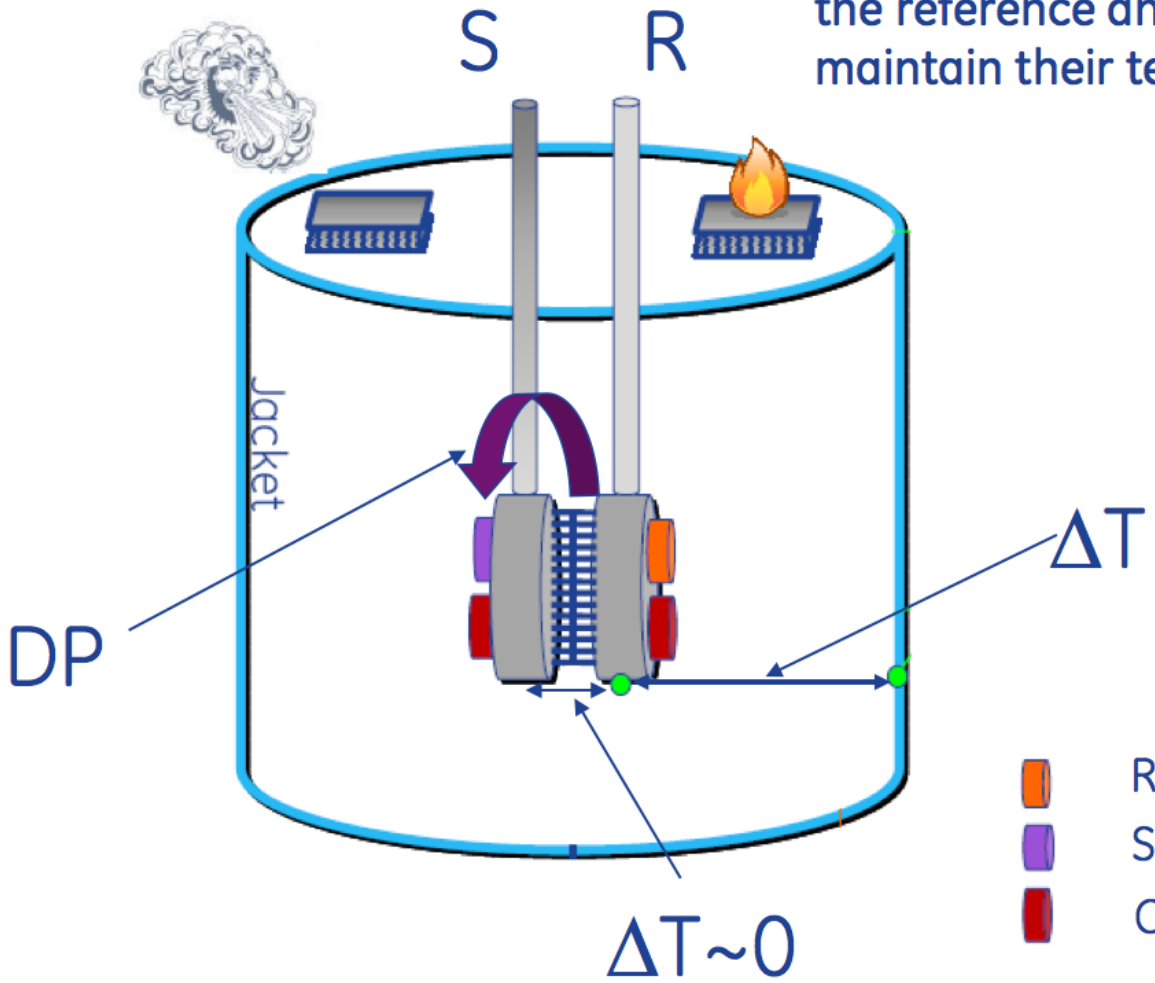
Three possibilities of calorimetric measurement:

1) temperature change (either adiabatic or isoperibol) [°C/time]
2) power compensation (often called isothermal) [µcal/time]
3) heat conduction (Lavoisier design)



# How Do ITCs Work?

The DP is a measured power differential between the reference and sample cells necessary to maintain their temperate difference at close to zero




- Reference Calibration Heater
- Sample Calibration Heater
- Cell Main Heater

# Performing an ITC experiment

Ligand in syringe

Macromolecule in sample cell

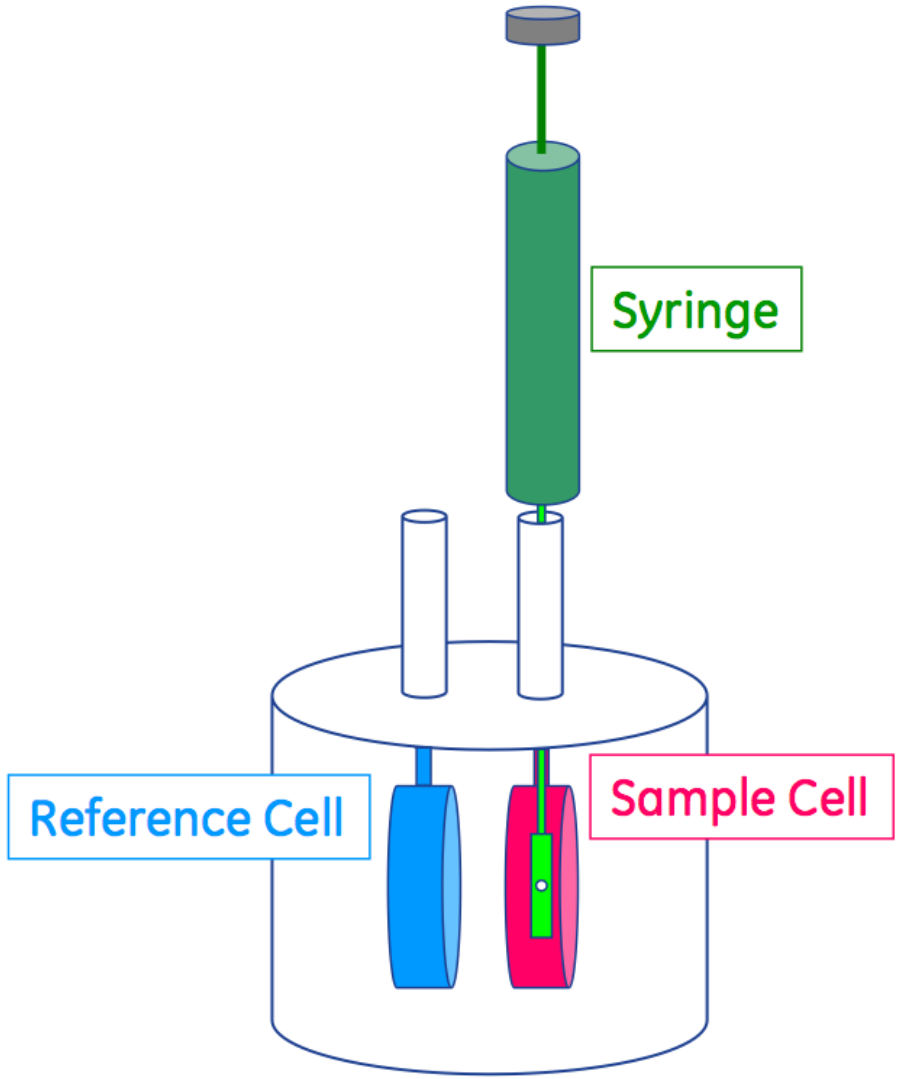
Heat of interaction is measured

Parameters measured from a single ITC experiment:

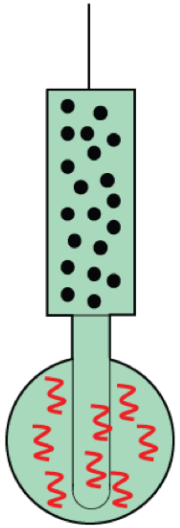
Affinity - $K_D$

Energy (Enthalpy) - $\Delta H$

Number of binding sites - n

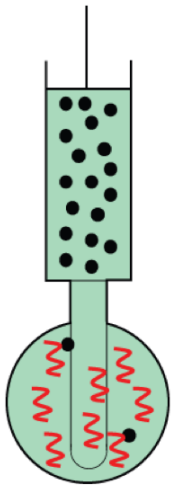


# ITC – Before titration

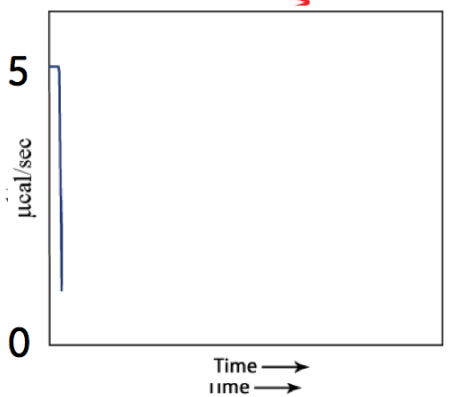


- Ligand – in syringe
- Macromolecule in ITC cell

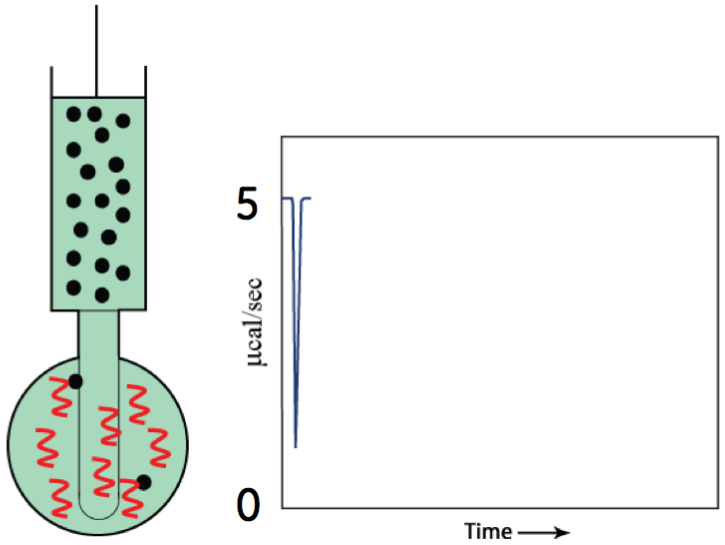
## Titration begins: First injection

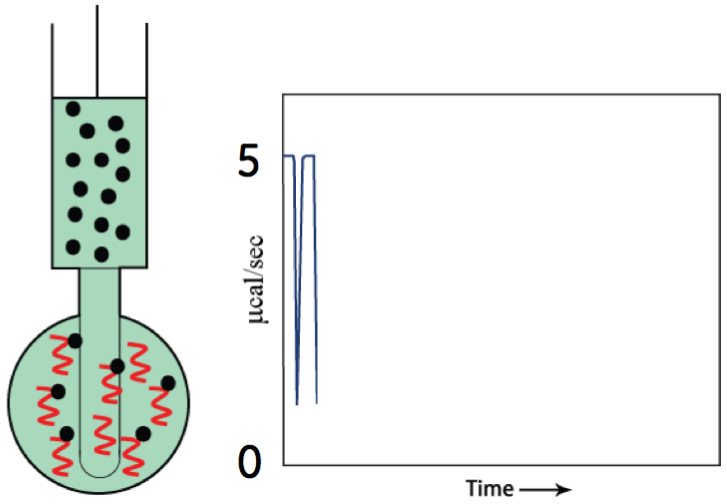


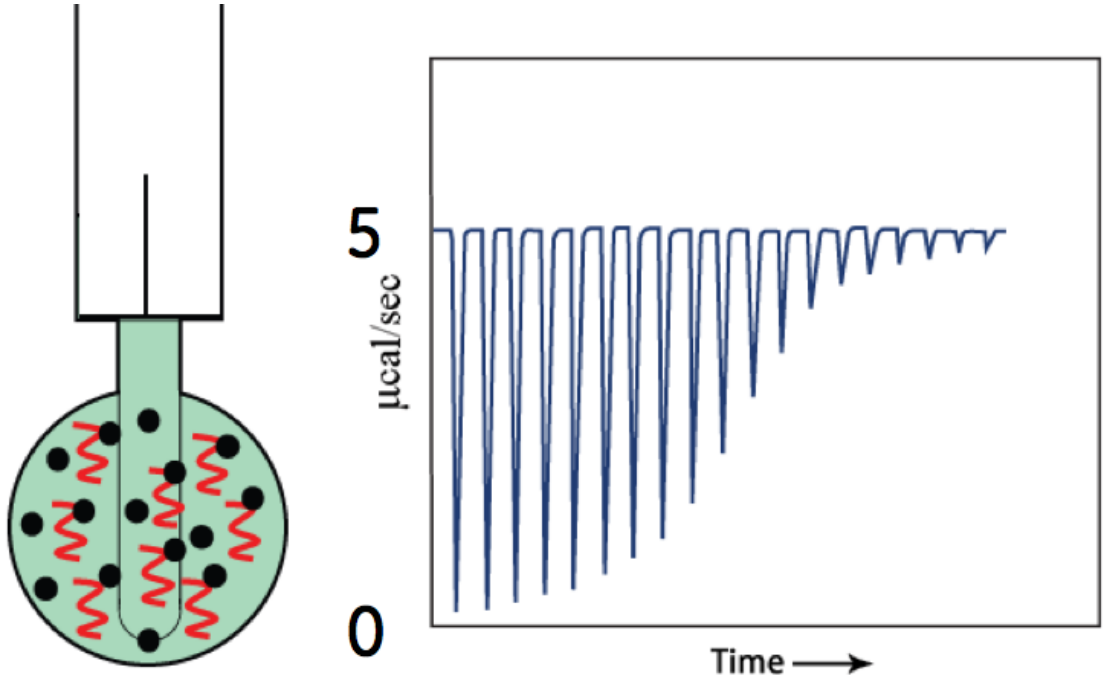
- Ligand in syringe
- Macromolecule in cell
- Macromolecule-ligand complex



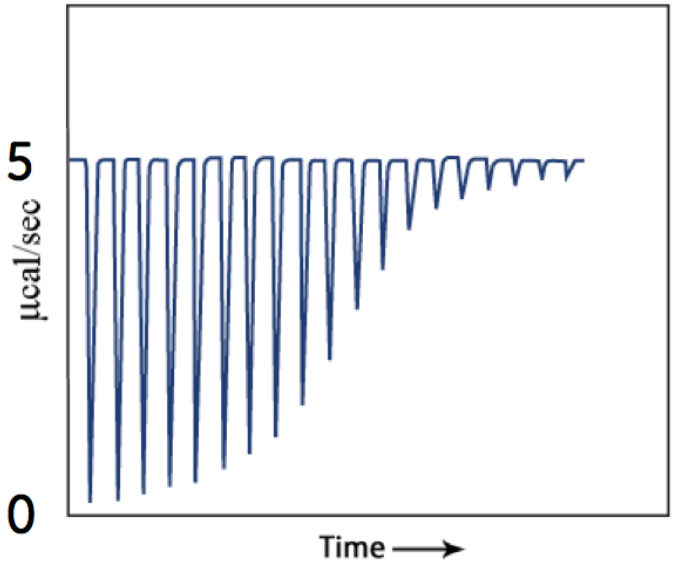

As the first injection is made, all injected ligand is bound to target macromolecule.

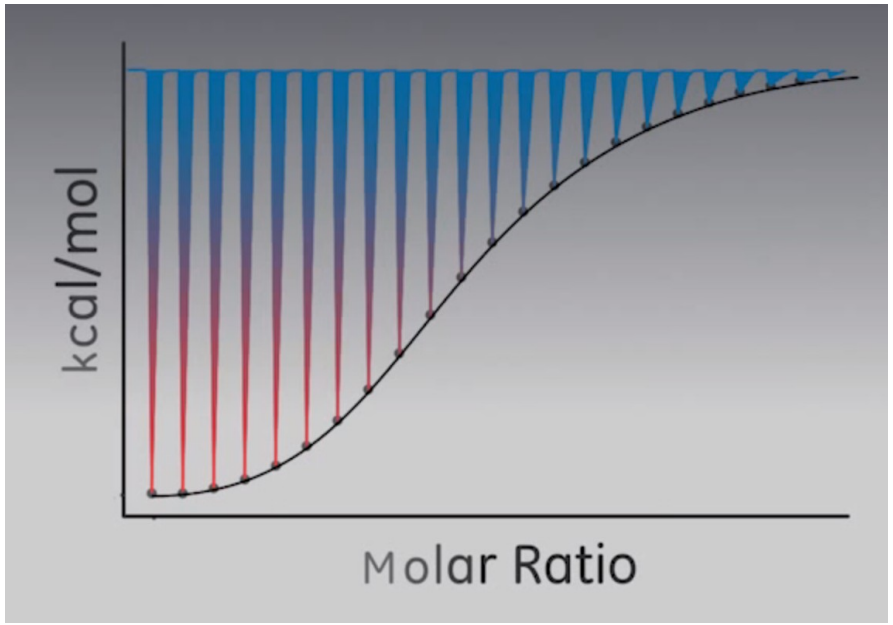


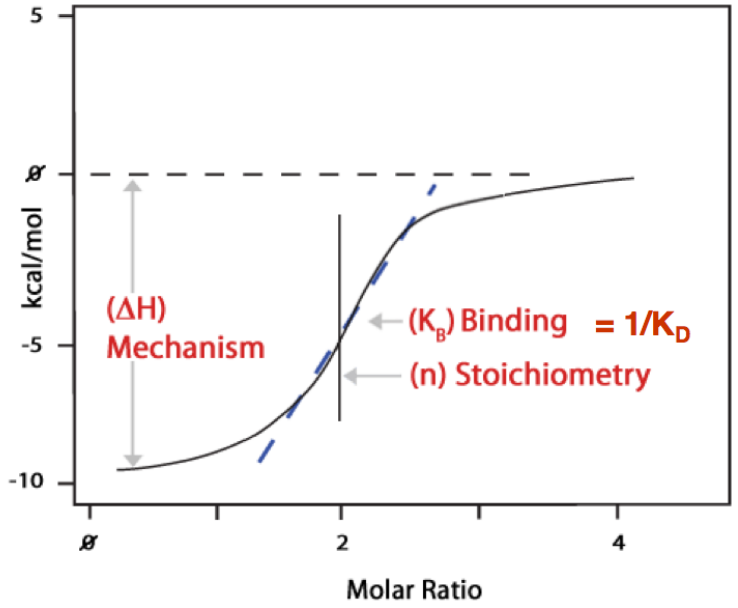




When the macromolecule is saturated with ligand, no more binding occurs, and only heat of dilution is observed.


5
µcal/sec
0
Time

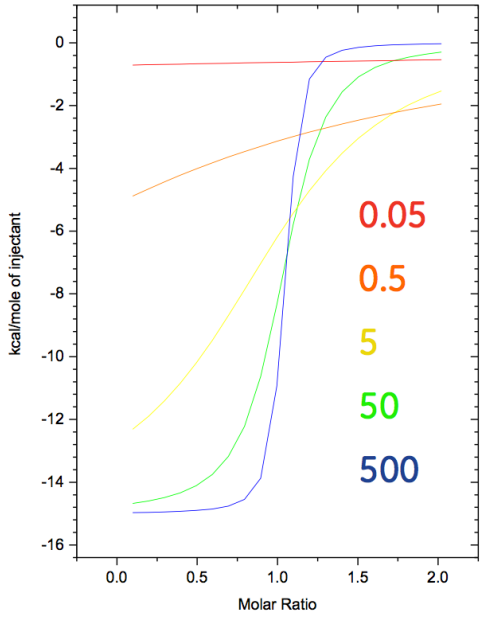

kcal/mol
Molar Ratio


5
∞
kcal/mol
-5
-10
∞
2
4
(ΔH)
Mechanism
(K_B) Binding
= 1/K_D
(n) Stoichiometry
Molar Ratio

# The experimental binding isotherm can be characterized by the unitless value *c*

$$c = K_a[\mathrm{M}]n.$$

## C Values




$$C = \frac{[M]}{K_D}$$

Example:

Kd = 100nM

[M] = 100nM, C=1
[M] = 5uM, C=50

[M]:[L] – 1:10 for n=1

## C Values in ITC

$C = \{[M]_{tot} / K_D\} * N$

C = 10-100 very good

C = 5-500 good

C = 1-5 and 500-1000 OK

C = < 1 and > 1000 not wanted