Neutron Protein Crystallography

DEVA workshop, Garching

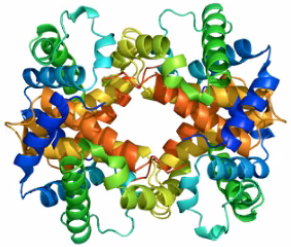
17. November 2023 | Tobias E. Schrader

Outline

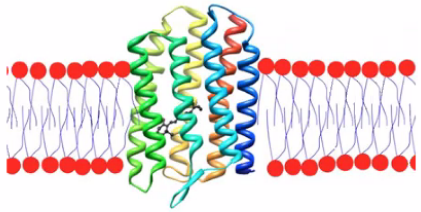
- Motivation: Why do we need protein structures at atomic resolution?
- x-ray protein crystallography
- neutron protein crystallography
- Theory of scattering from crystals
- One or two application examples: From Structure to function…

Proteine

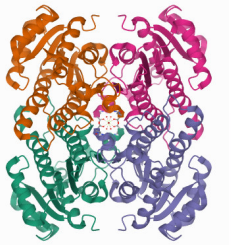
- Etwa 20 000 verschiedene Proteine arbeiten im menschlichen Körper und erfüllen die unterschiedlichsten Aufgaben:



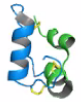
Hämoglobin,
Sauerstofftransport



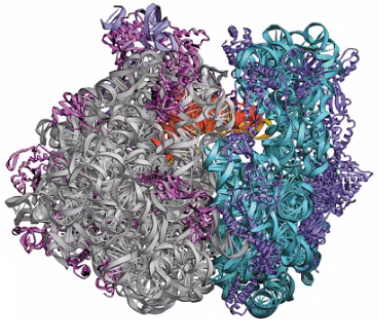
Rhodopsin,
Sehvermögen



Alkoholdehydrogenase,
Alkoholverdauung



Insulin,
Regulierung des
Zuckerhaushalts

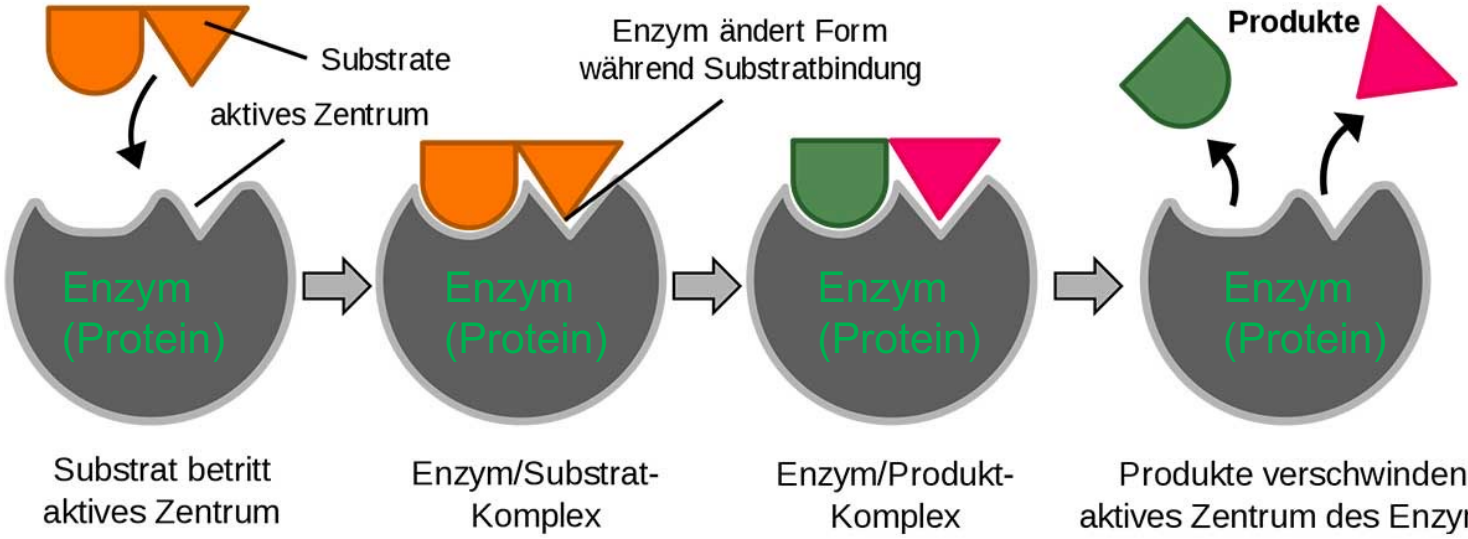


Ribosom,
Synthese von Proteinen

Einige von ihnen bieten Angriffspunkte für Medikamente...

Enzyme sind biologische Katalysatoren

- Sie beschleunigen chemische Reaktionen schon nahe Raumtemperatur

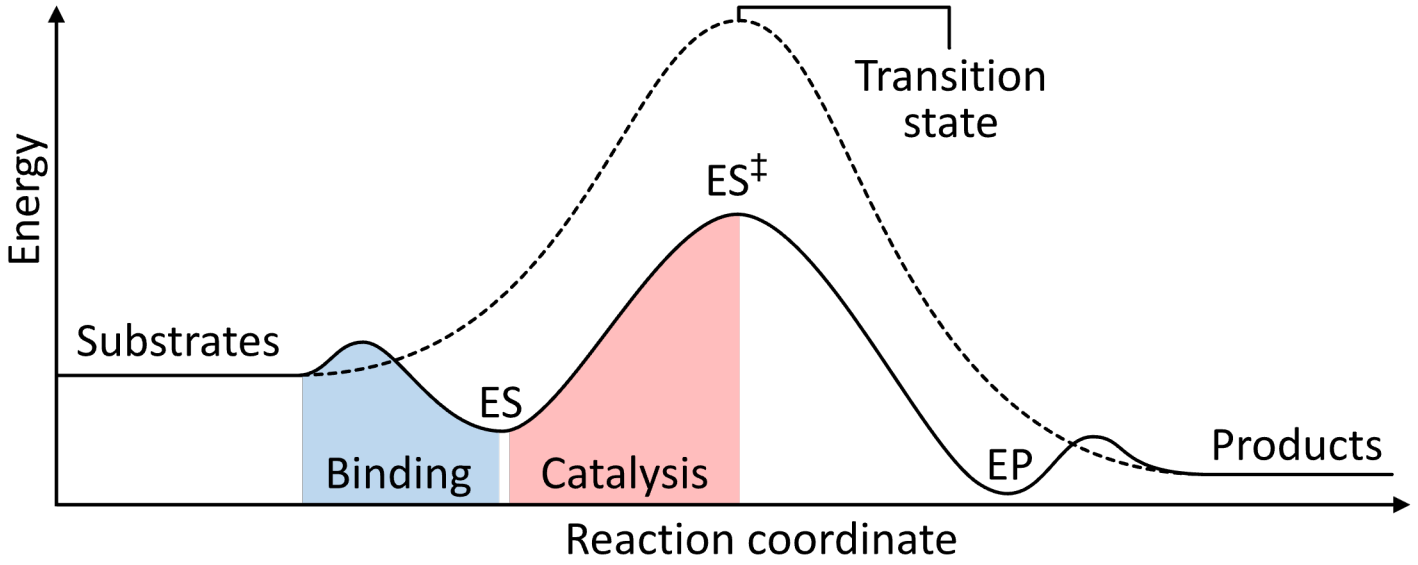


- Temperatur ca. 37°C
- Hoch substrat-spezifisch

- Im Vergleich dazu:
  Chemische Industrie

- ❖ Oft hohe Temperaturen notwendig (=nicht sehr energieeffizient)
- ❖ Mitunter viele unerwünschte Nebenprodukte, die mühsam abgetrennt werden müssen

Enzymes, a sub-class of proteins, accelerate chemical reactions



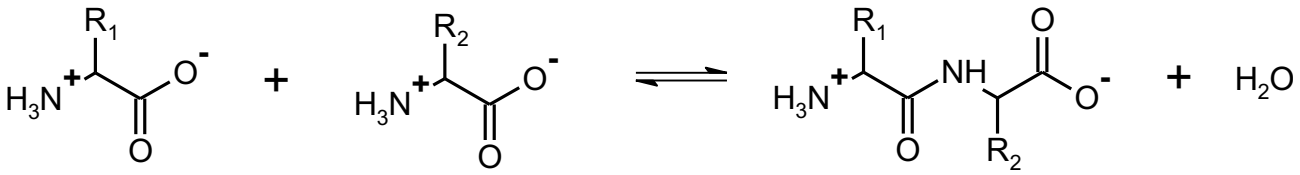
- ➡ very high substrate specificity
- ➡ rate enhancements up to 10^{18}
- ➡ enzymes can be “controlled”

Some basics on protein structures

Definition of the name "protein"

• Proteins are the most abundant class of biological macromolecules in all cellular organisms.

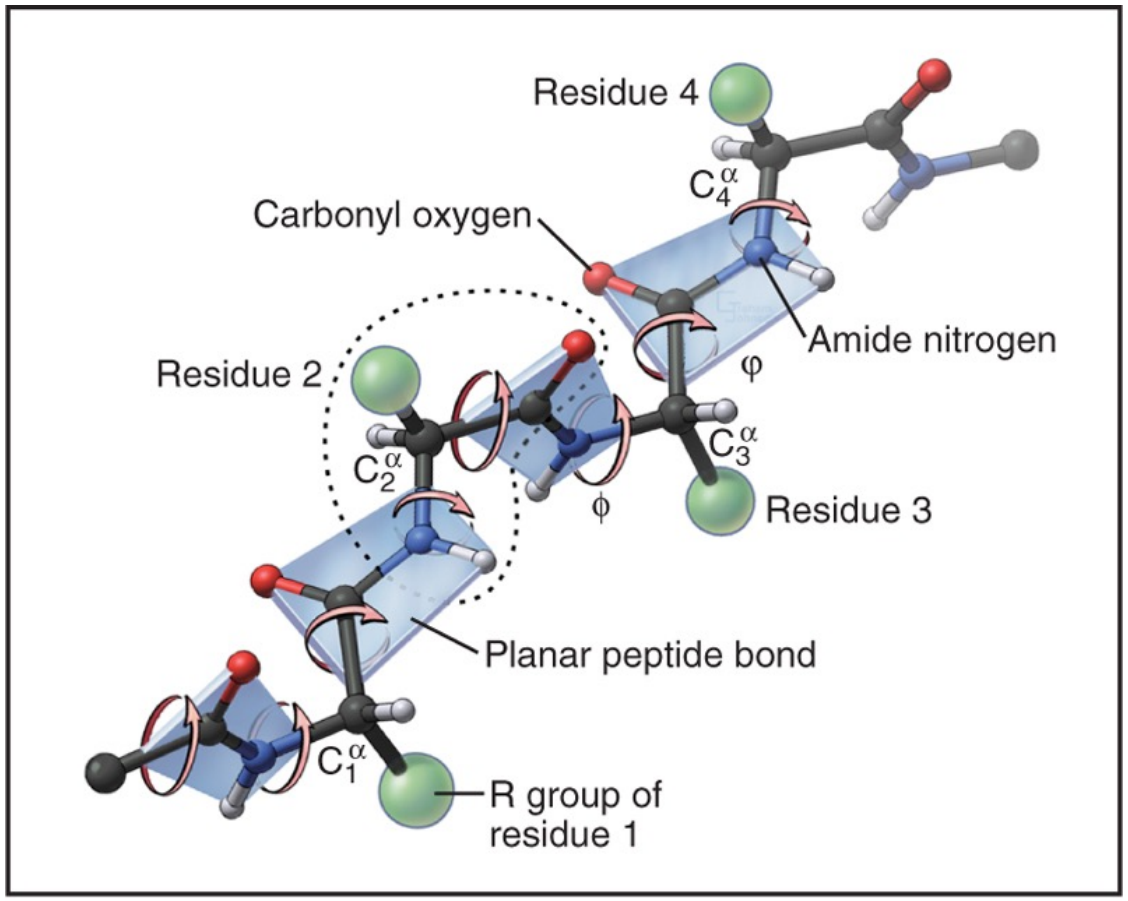
• They are polymers of α-amino acids (2-amino carboxylic acids), formed by condensation of carboxyl and amino functions of adjacent units.



• Historical nomenclature:
n < 10: oligopeptide
10 < n < 100: polypeptide
n > 100: protein

• Preferable: A protein is a polypeptide with a defined three-dimensional fold.

Definition of dihedral angles



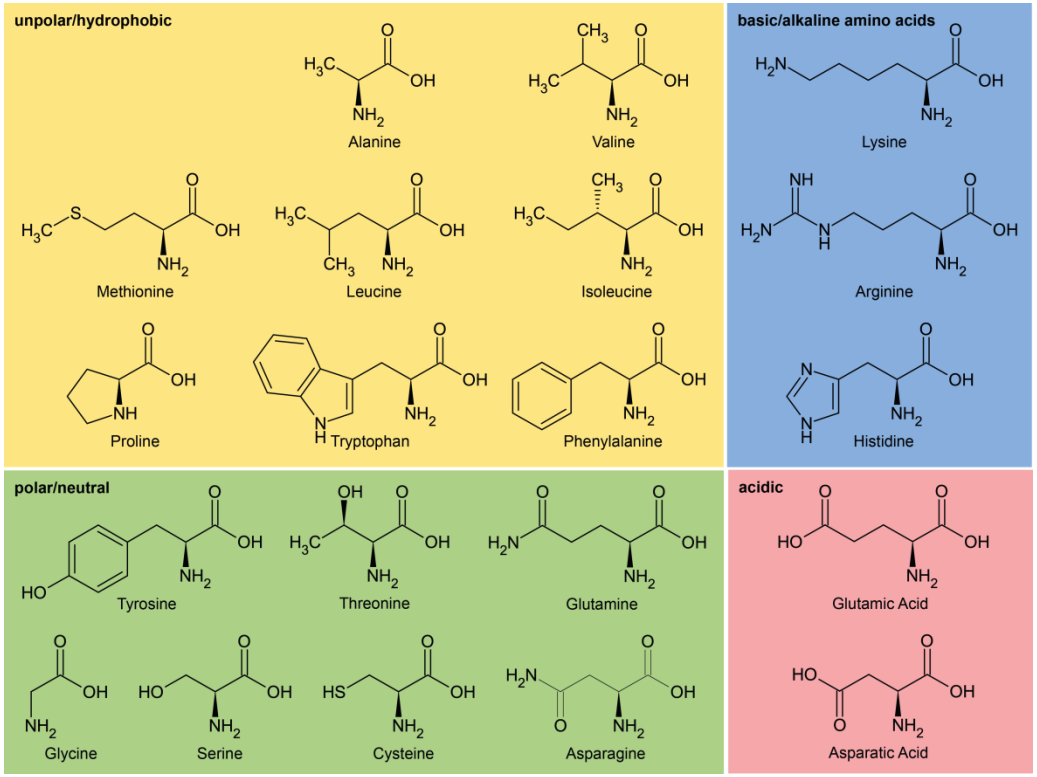
color code:
blue: nitrogen
red: oxygen
white: hydrogen
grey: carbon
green: symbol for side chains

Per amino acid the peptide chain has these two dihedral angles as degrees o freedom.

The 20 amino acids occuring in nature

- Side chains have different length and character
- α-Carbon is a chiral centre (exception: glycine)
=> D- and L-forms possible
- Cellular protein synthesis uses L-enantiomers only.

Proteins are chiral molecules!



The hierarchy of protein architecture

- Primary structure

The sequence of amino acid residues
(N to C terminus)

- Secondary structure

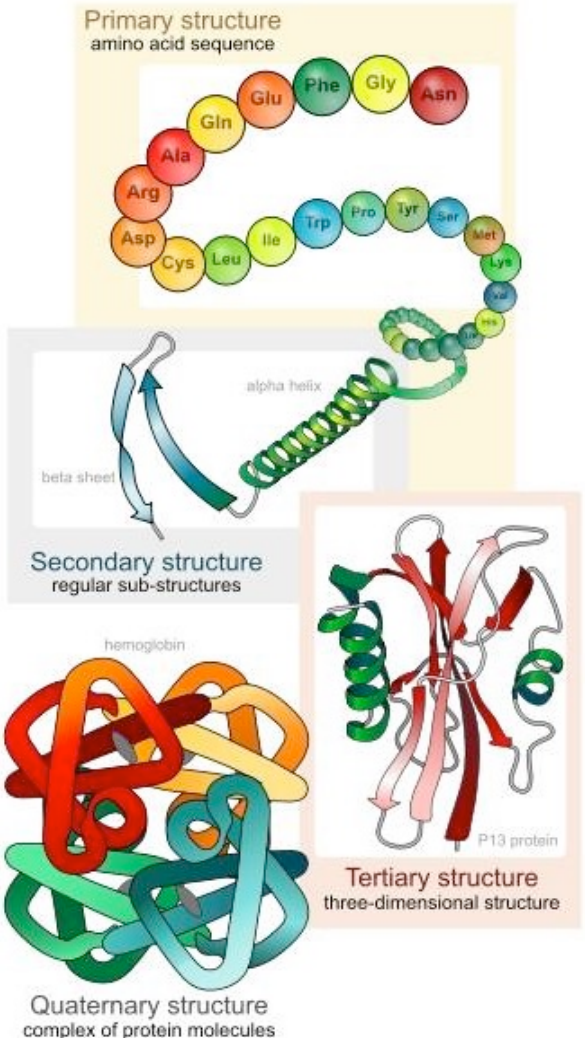
The local arrangement of main-chain atoms, often with specific hydrogen bonding pattern

- Tertiary structure

The packing of non-contiguous segments into a compact fold

- Quaternary structure

The association of individual chains into oligomeric structures



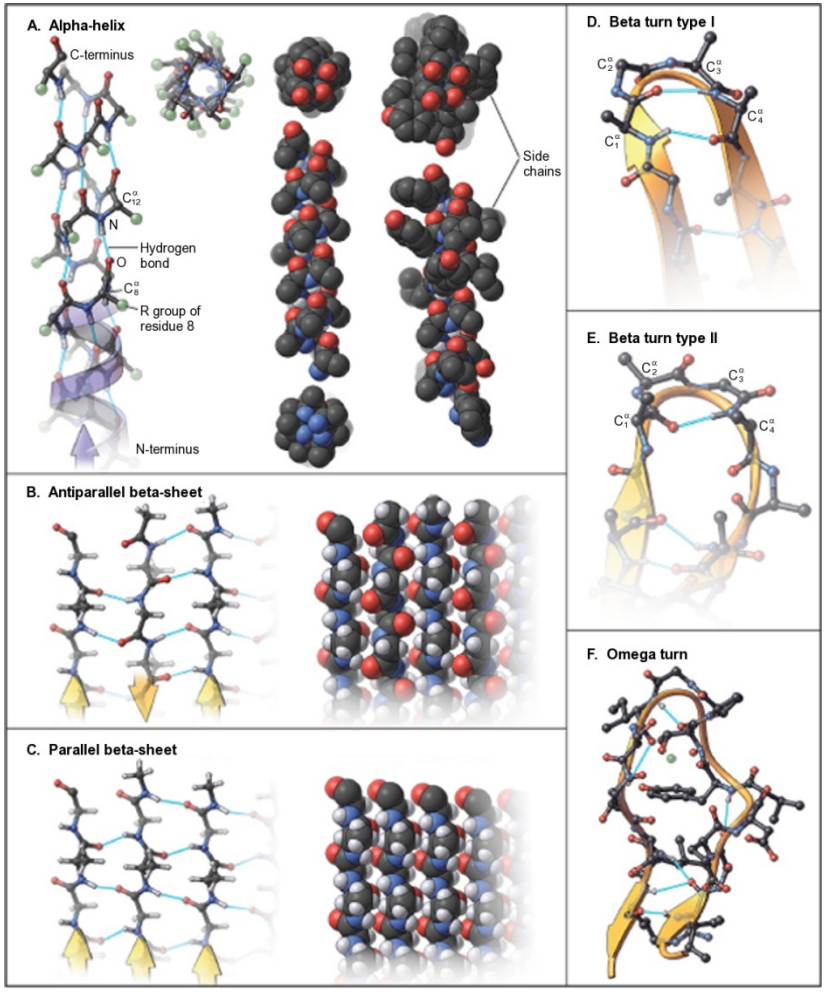
Secondary structural elements

α-helix
spiral conformation of a continuous chain

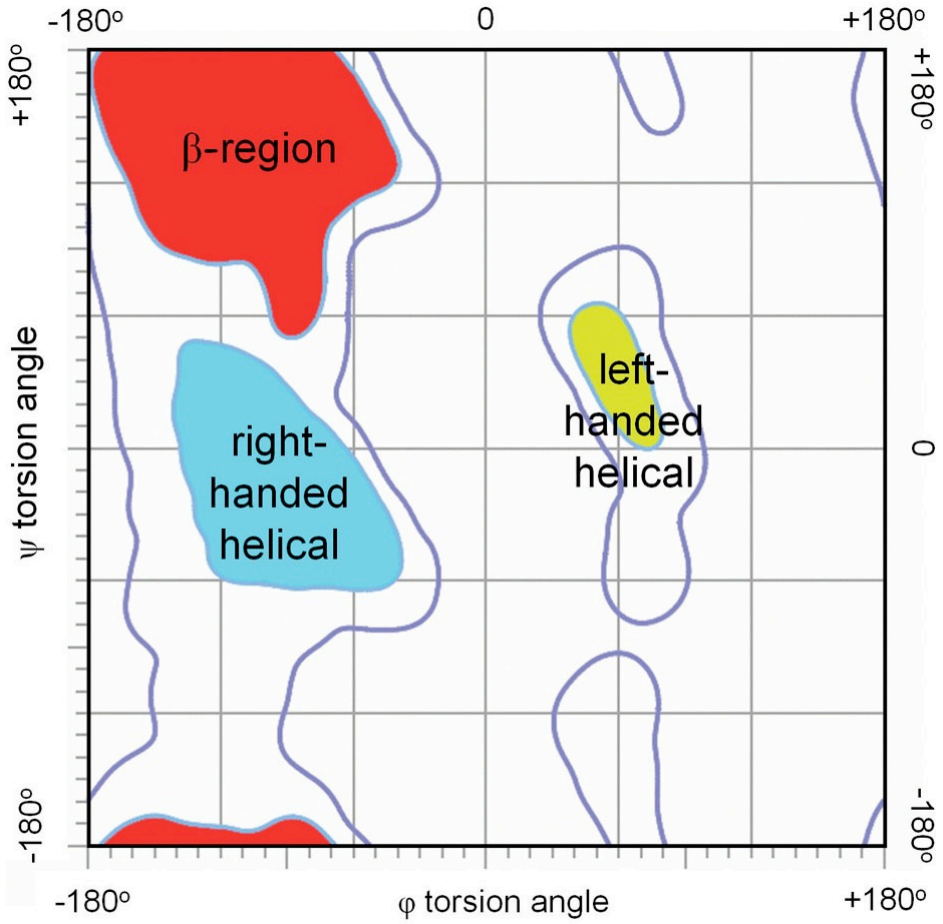
β-sheet
extended strands
aligned side-by-side

turns / loops
changing direction of chain

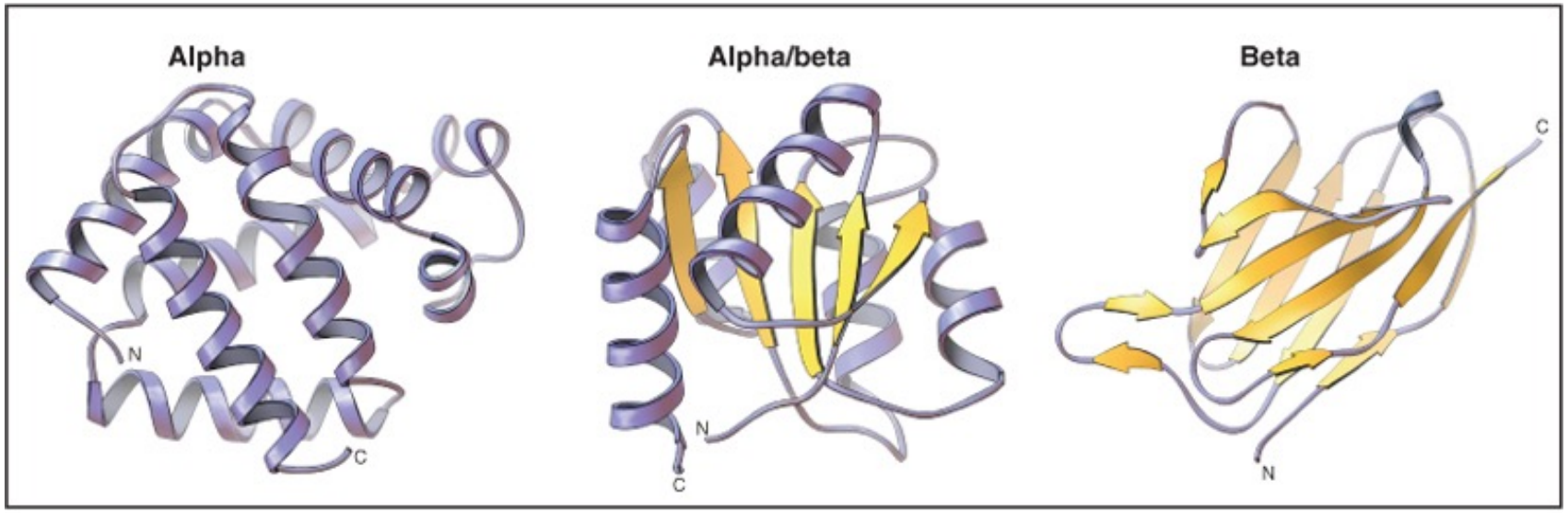
Secondary structure elements are generally stabilized by hydrogen bonds between different parts of the backbone of the protein.



Not all dihedral angles a sterically allowed or favourable: The Ramachandran plot

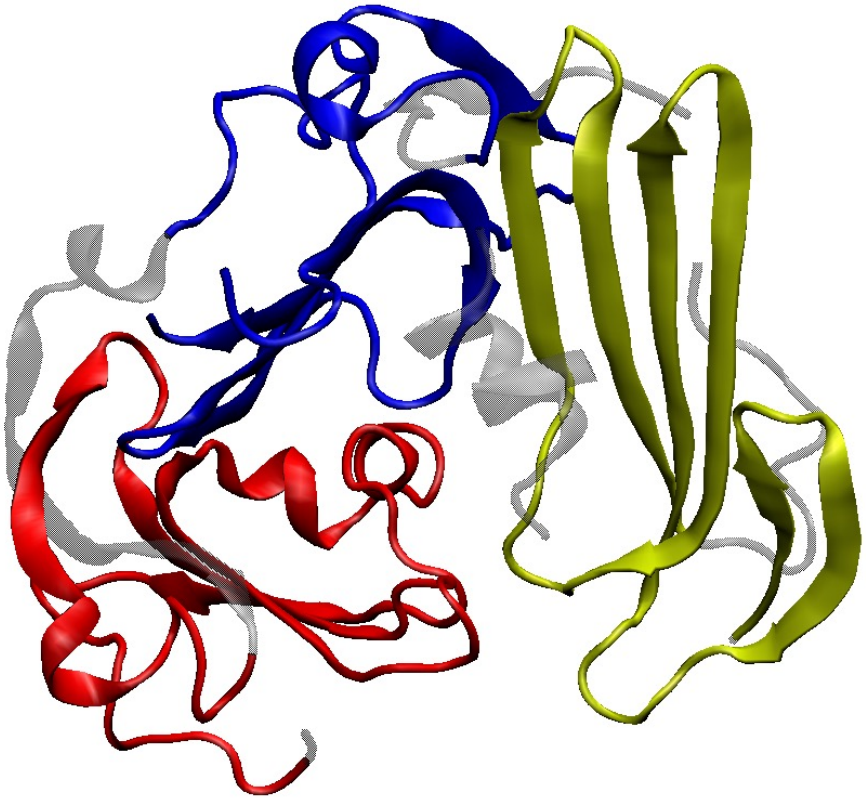


Tertiary structure: Fold classes



- Tertiary structure results from the arrangement of secondary structure elements self assembling into the native fold.
- This arrangement critically depends on side-chain interactions, the hydrophobic effect being a major driving force.

Tertiary structure: sub-dividing the protein into domains



Secondary structure plot of human urokinase plasminogen activator receptor, a complex protein (pdb code 1YWH). The domains DI (yellow), DII (blue) and DIII (red) are shown. Amino acids not belonging to any domain are depicted in grey.

What stabilizes the 3-dimensional fold: The protein folding problem

Unfolded states ⇌ Intermediate states ⇌ Native state(s)

- Polypeptide chains tend to collapse into compact structures.
- Major driving force is the **hydrophobic effect** (entropic contribution of the solvent).
- Forces stabilizing folded proteins:

Non-covalent bonds	Covalent bonds
• Hydrogen bonds	• Disulfides
• Electrostatic forces	• Isopeptide bonds
• V. d. Waals contacts	

Summary of the protein part

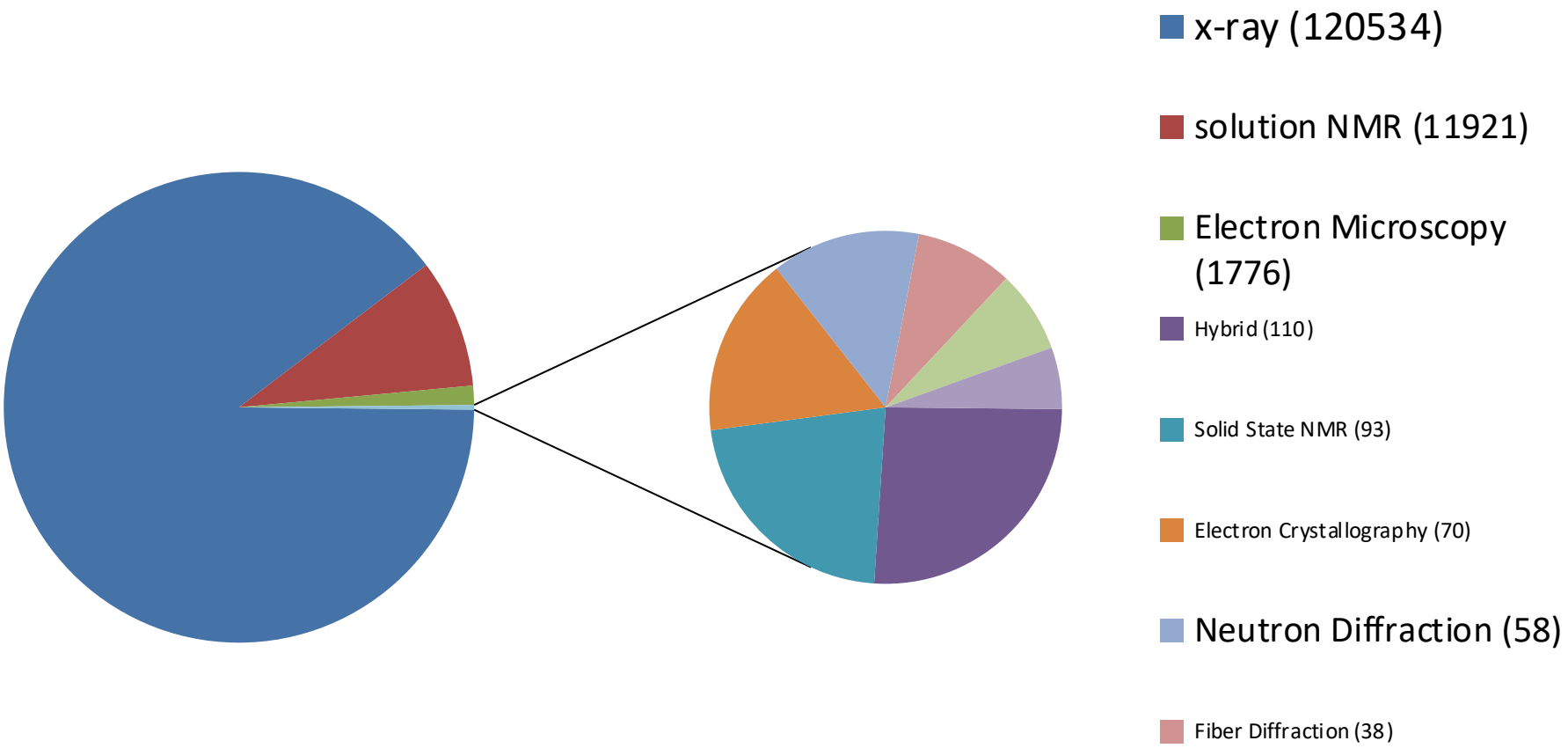
- Proteins consist of a chain of amino acids.
- There are 20 amino acids occuring in nature.
- Proteins show a special 3-D structure which is specific to their function

Why do we need to measure protein structures?
Are there no good simulations?
How do we measure protein structures?

Why do we need experimental studies on proteins?

- **MD-Simulations** suffer from non-perfect force fields: Especially the long range electrostatics is not reproduced very well. But proteins use defined and structure related electrostatics to move the acidity constants of side chains in order to make them fullfill their tasks. MD-simulations cannot model bond breaking and forming very well since the quantum chemistry nature of this process is not included in the theoretical foundation of MD.
- **Ab initio quantum chemical calculations** are still too demanding to model the complete active centre of a protein (including its substrate)

Most structures are obtained by x-ray crystallography, available neutron structures in protein data bank: ca. 100

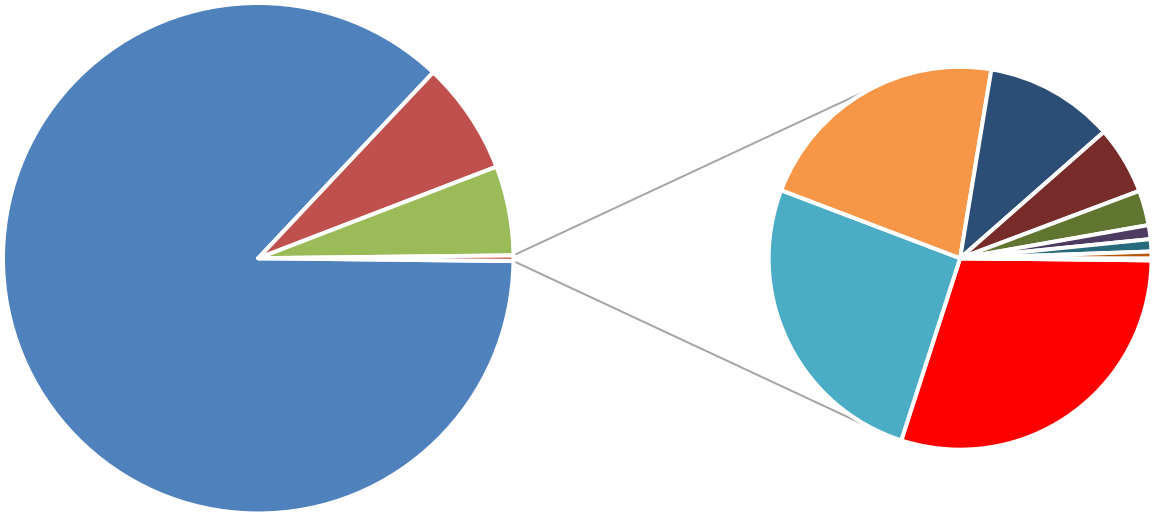


http://www.rcsb.org/

Total number of structures: 134656

Most structures are obtained by x-ray crystallography, available neutron structures in protein data bank: ca. 200

Entries in the protein data base as of 9th of May 2022



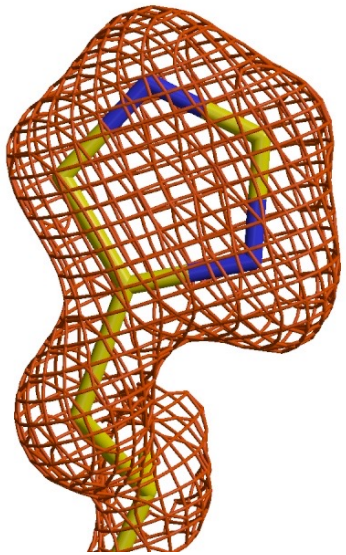
- X-RAY DIFFRACTION (165365)
- SOLUTION NMR (13598)
- ELECTRON MICROSCOPY (10777)
- NEUTRON DIFFRACTION (202)
- ELECTRON CRYSTALLOGRAPHY (175)
- SOLID-STATE NMR (148)
- SOLUTION SCATTERING (74)
- FIBER DIFFRACTION (39)
- POWDER DIFFRACTION (20)
- EPR (8)
- THEORETICAL MODEL (7)
- INFRARED SPECTROSCOPY (4)
- FLUORESCENCE TRANSFER (1)

http://www.rcsb.org/

Total number of structures: 190418

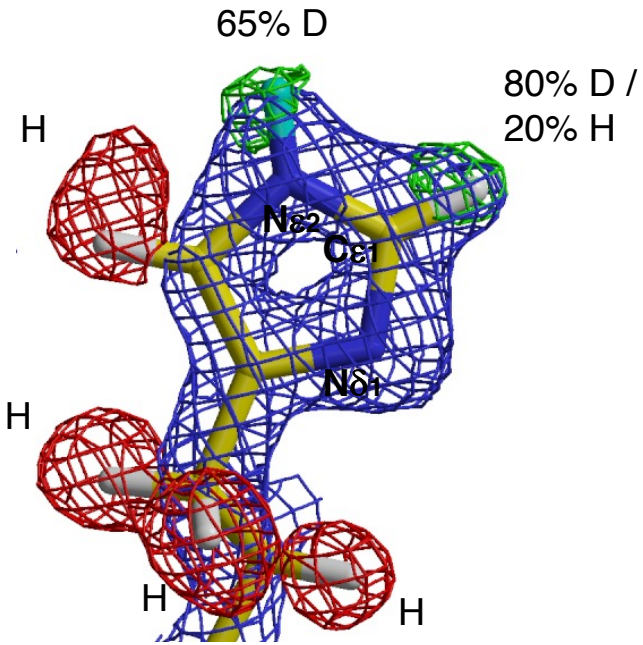
Protonation states of amino acids:

X-ray d_{min}= 1.5Å:



2Fo-Fc map; +1.5σ

neutrons d_{min}= 1.5Å:



2Fo-Fc map; +1.5σ

Fo-Fc omit-map; -3.0σ

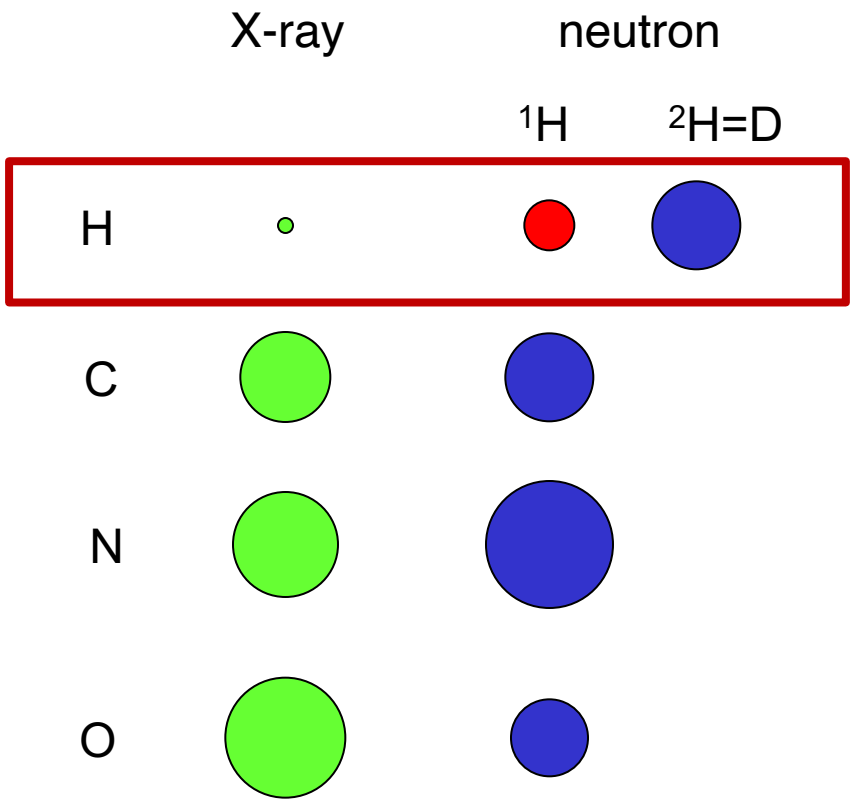
Fo-Fc omit-map; +3.0σ

Niimura N, Chatake T, Ostermann A, Kurihara K, Tanaka T. (2003) Z. Kristallogr. 218:96

Advantages of structure determination with neutrons:

Comparison of form factors (X-ray) and scattering lengths (neutrons):

Nucleus	atomic number	scattering length [10^{-12} cm]
^{1}H	1	**-0.378**
^{2}H	1	**0.667**
^{12}C	6	**0.665**
^{15}N	7	**0.921**
^{16}O	8	**0.581**



diameters correspond to:
form factor / scattering length
(scaled for C-atom)

σ_{coh} of ^{1}H is 1.8×10^{-28} m^2 but
σ_{incoh} of ^{1}H is 80.2×10^{-28} m^2
Large background from hydrogen atoms!

A crystal structure according to the protein data bank (PDB)

x,y,z coordinates (Å)

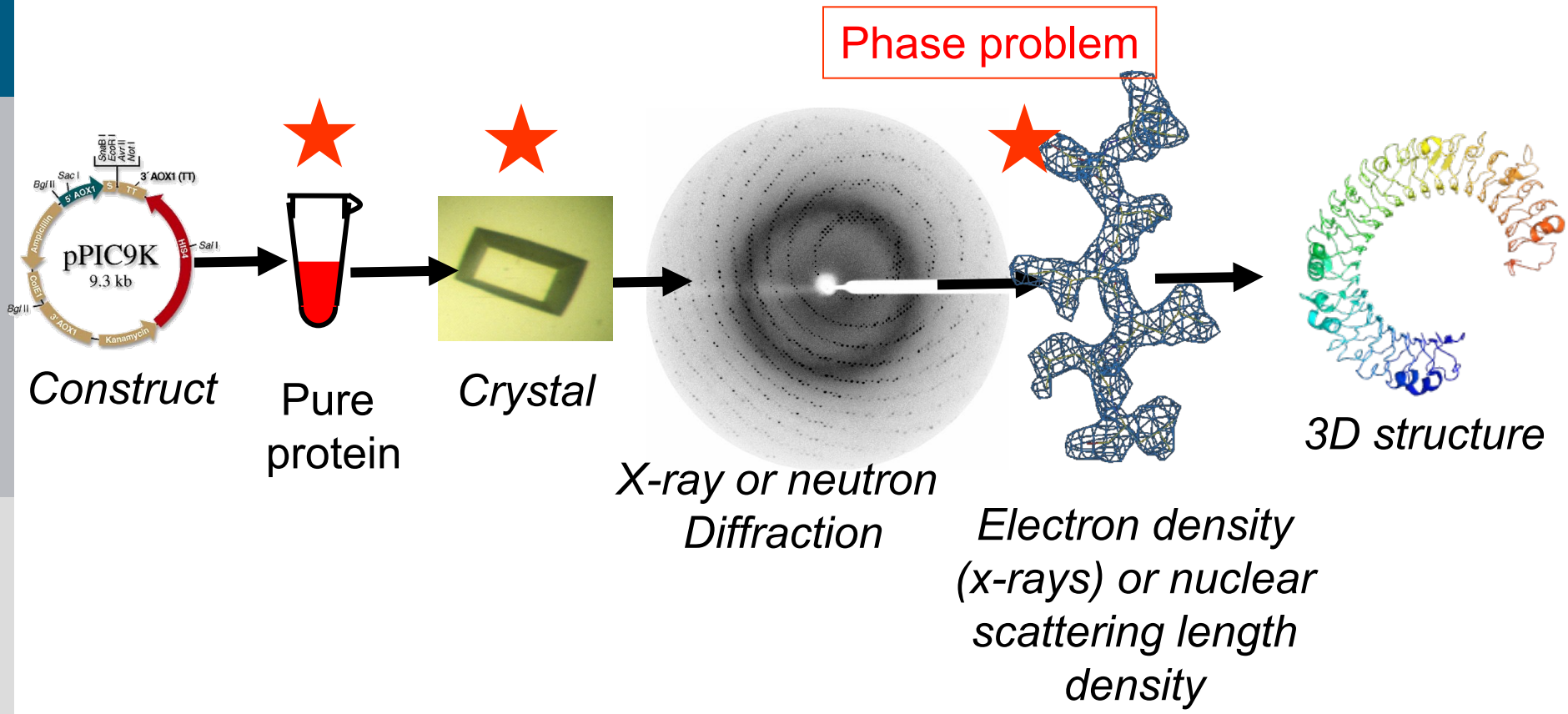
```
ATOM     25  N   ASP A 928      19.062   9.157  35.067  1.00  4.73           N
ATOM     26  CA  ASP A 928      19.770  10.123  34.232  1.00  4.58           C
ATOM     27  C   ASP A 928      19.075   9.938  32.899  1.00  4.56           C
ATOM     28  O   ASP A 928      19.074   8.824  32.351  1.00  5.39           O
ATOM     29  CB  ASP A 928      21.259   9.776  34.071  1.00  3.13           C
ATOM     30  CG  ASP A 928      22.112  10.245  35.233  1.00  5.52           C
ATOM     31  OD1 ASP A 928      21.693  11.114  36.025  1.00  5.42           O
ATOM     32  OD2 ASP A 928      23.239   9.742  35.349  1.00  7.93           O
ATOM     33  N   VAL A 929      18.417  10.985  32.405  1.00  3.68           N
ATOM     34  CA  VAL A 929      17.726  10.864  31.125  1.00  4.63           C
```

Isotropic B-factor or temperature factor is a measure of the mobility of an atom

B ($Å^2$) = $8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the mean square atomic displacement

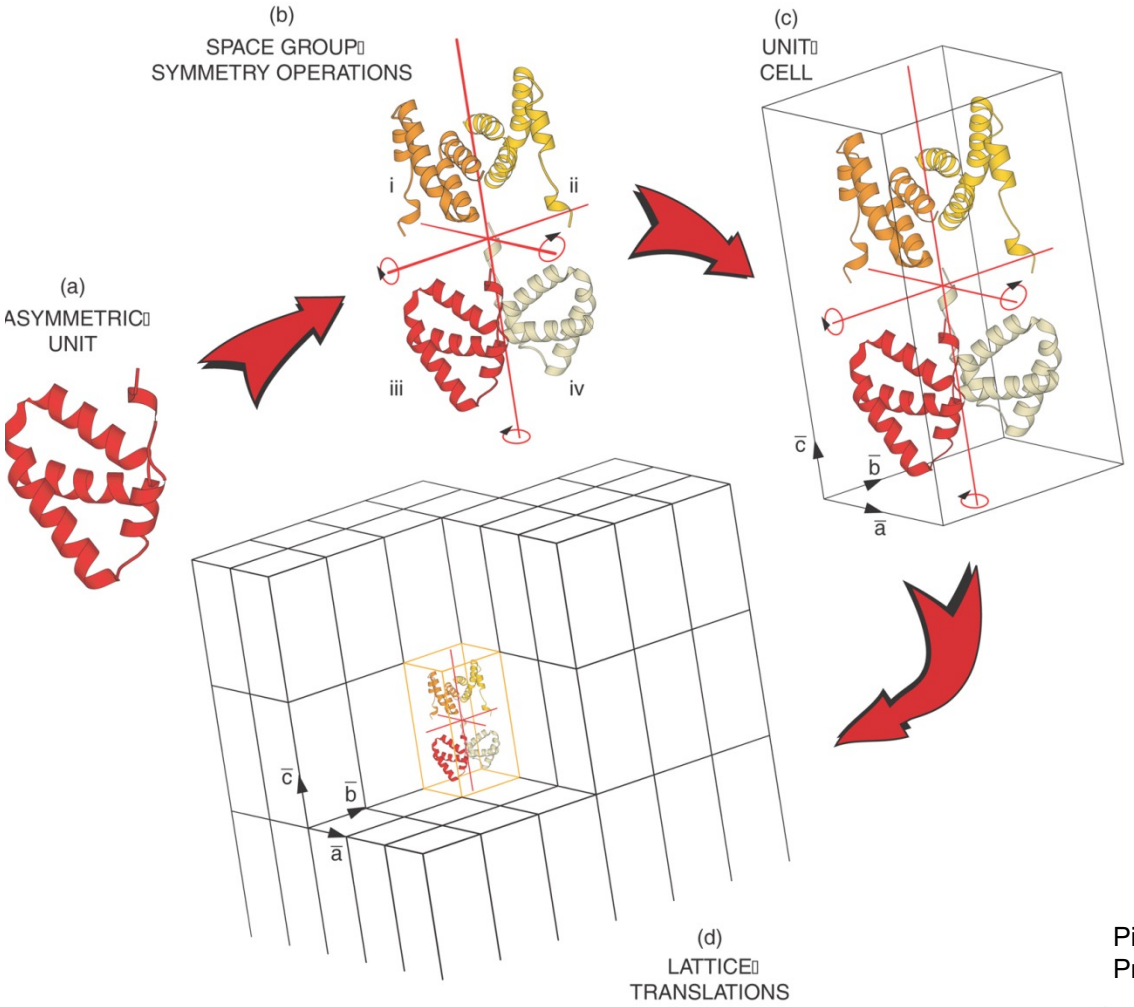
Protein crystallography in general, valid for both x-rays and neutrons as probes

Crystallography: Overiew over the process



Harma Brondijk, Crystal and Structural chemistry, Utrecht University

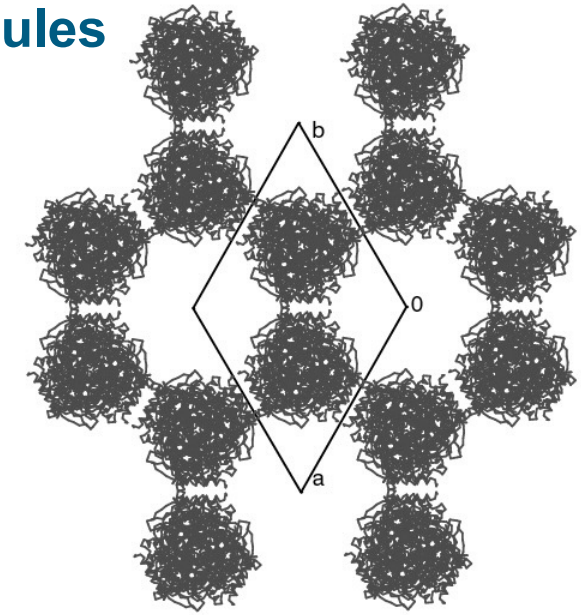
How a typical protein crystal looks like…

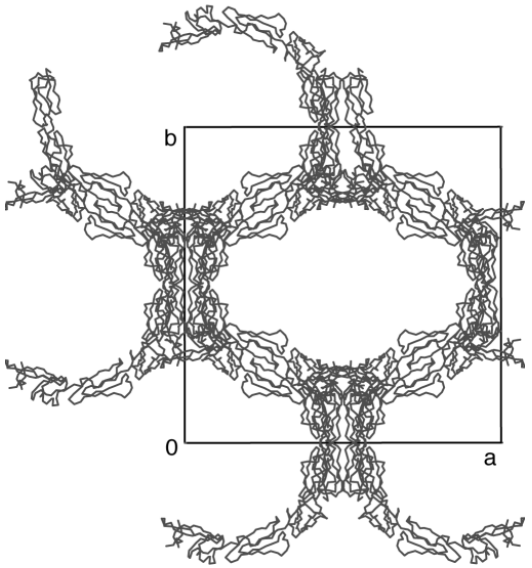


Picture taken from Lecture of Prof. Locher at ETH Zürich

fig 2.2

Protein crystals contain a lot of solvent and are held together by a limited number of weak contacts between protein molecules



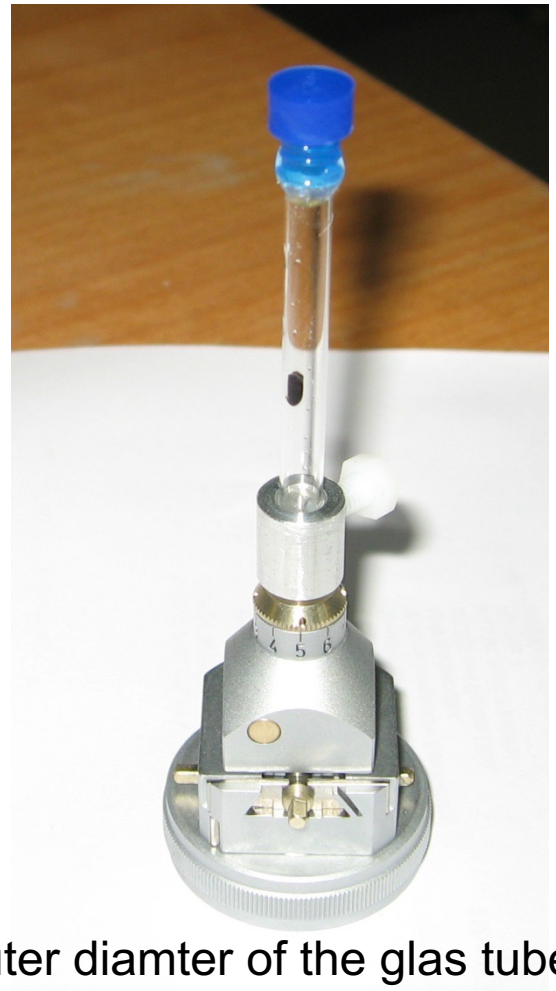


Acetylcholinesterase
~68% solvent

β2 Glycoprotein I
~90% solvent
(extremely high!)

Typical solvent content 40-60%
Solvent channels allow diffusion of compounds into crystal
Often these compounds can reach the active or binding site
Often enzymes are active in crystalline state

Size considerations of protein crystals

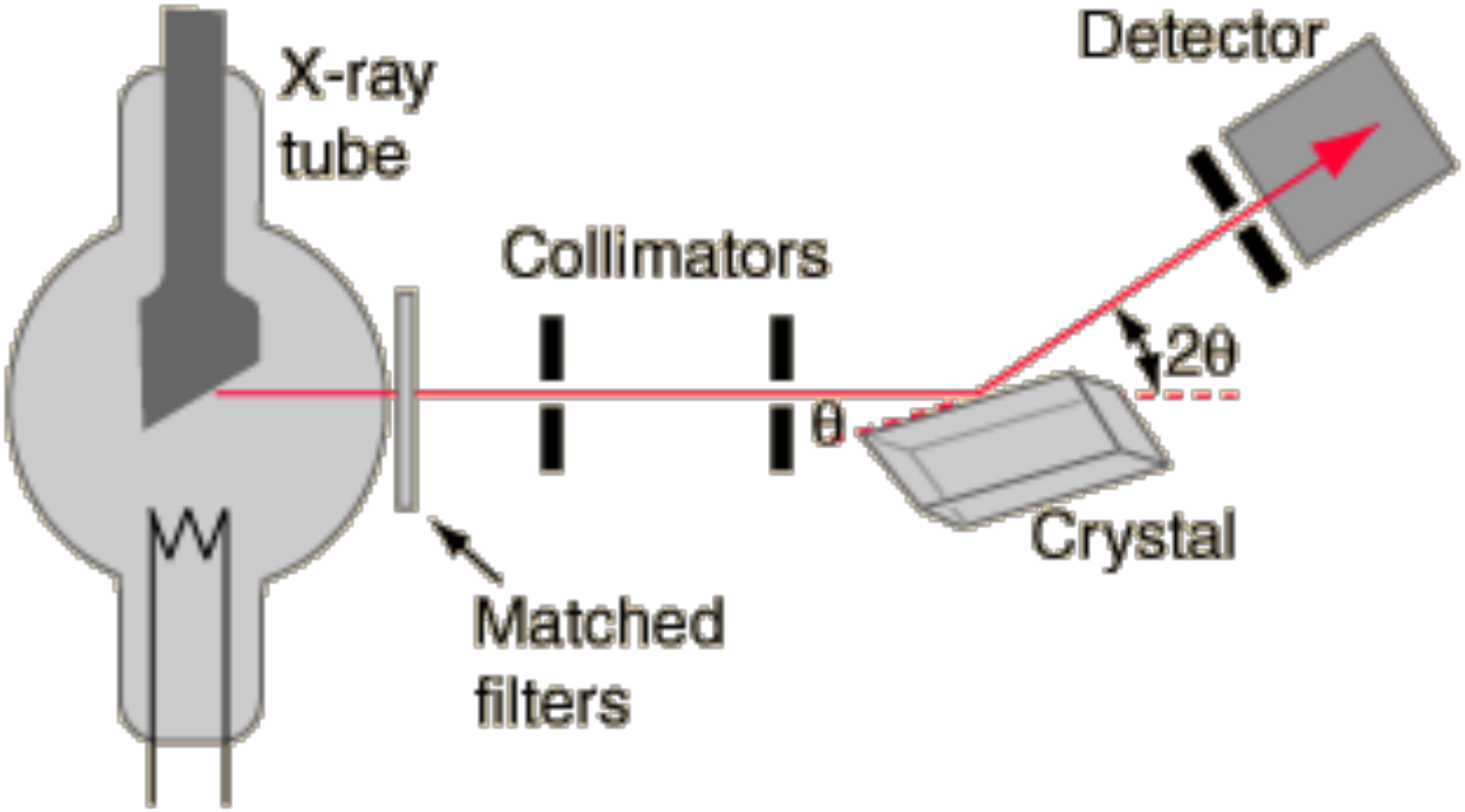


size:
x-ray-crystallography:
ca. 10 µm x 10 µm x 10 µm
typically cryoprotectants needed to facilitate measurements at low (80 K) temperatures

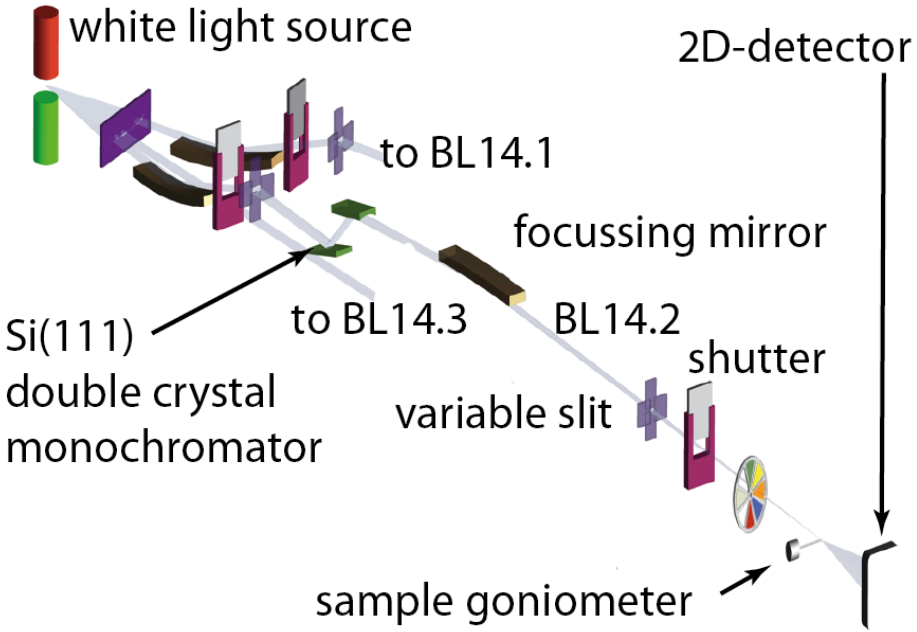
neutron protein crystalography:
The desirable size should be around 1 mm x 1 mm x 1 mm (depending on the protein/space group)

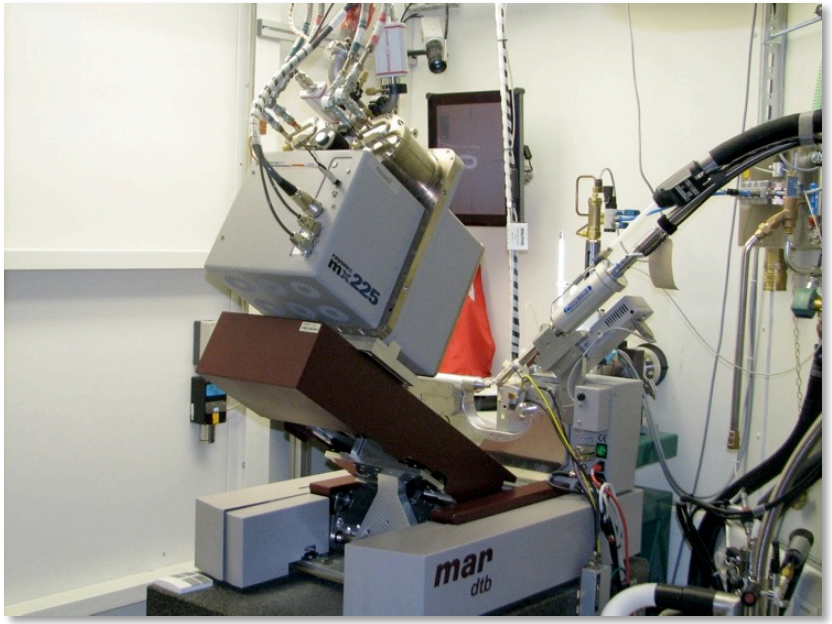
Outer diamter of the glas tube: 5 mm

Experimental set up (in case of x-rays but similar in the case of neutrons):



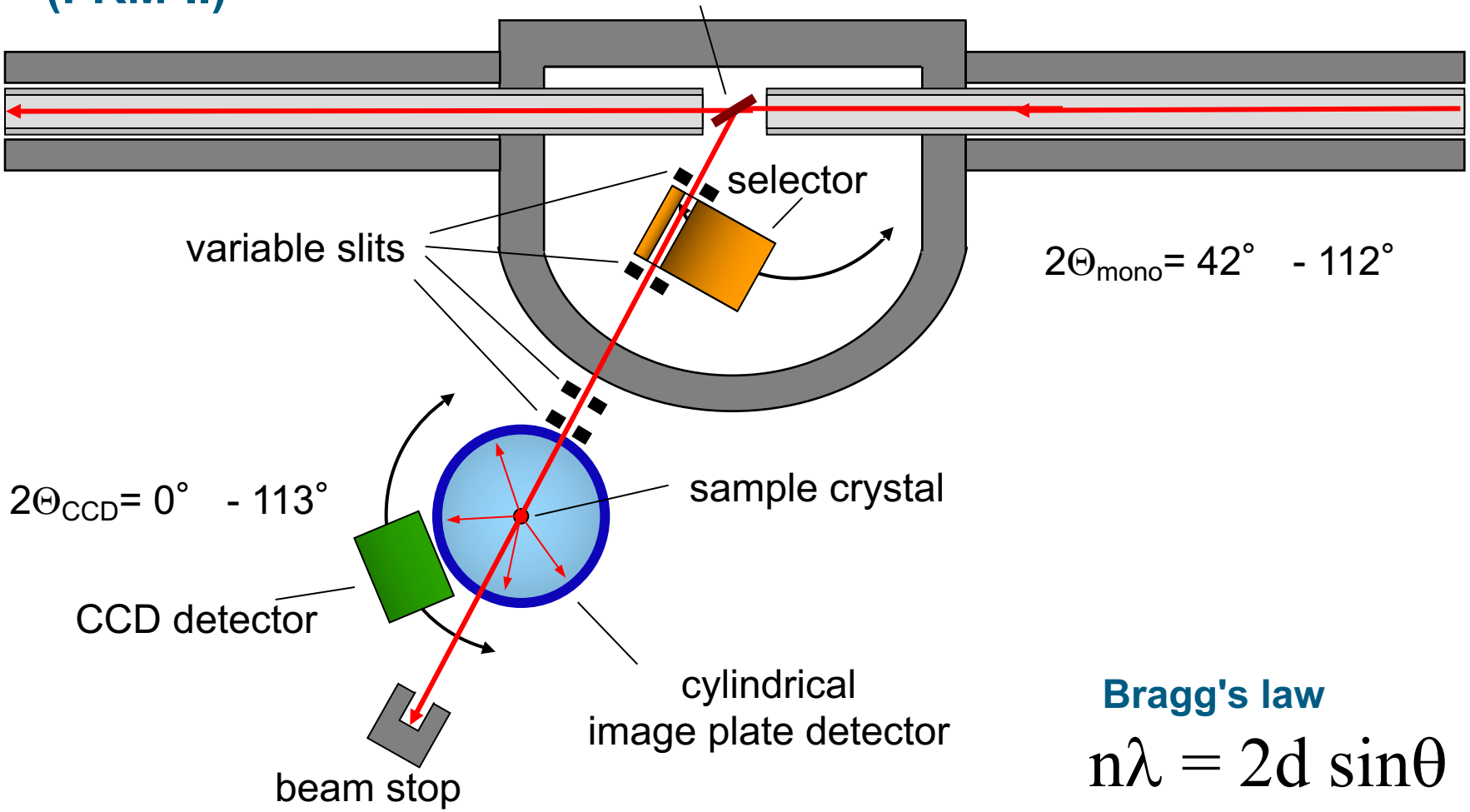
Typical x-ray protein crystallography beamline: BL 14.2 at Bessy (Berlin) run by Manfred Weiss





length scale ca. 0.5 m

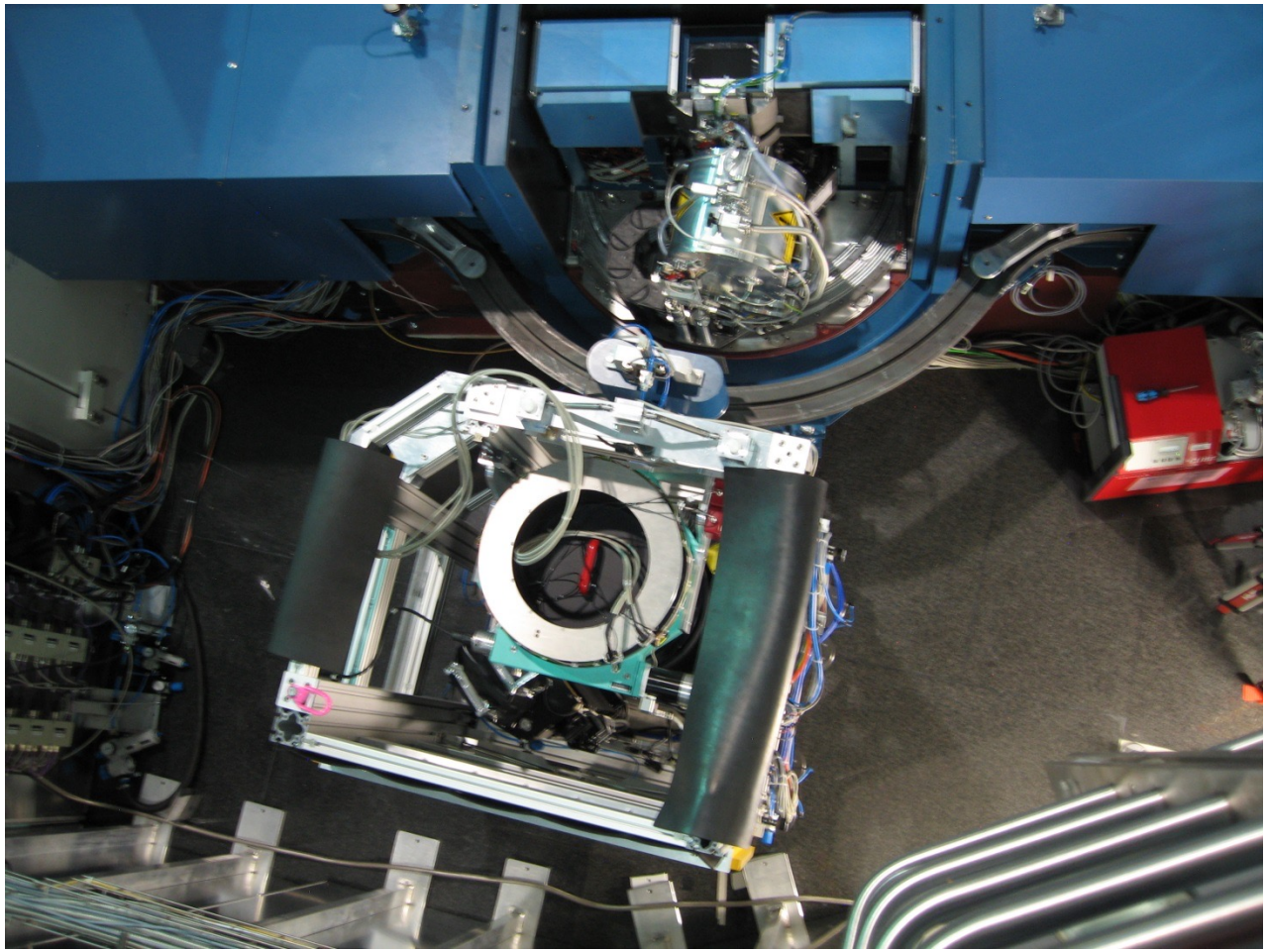
Schematic overview over BioDiff: A neutron protein diffractometer: collaboration between JCNS and TUM (FRM-II)



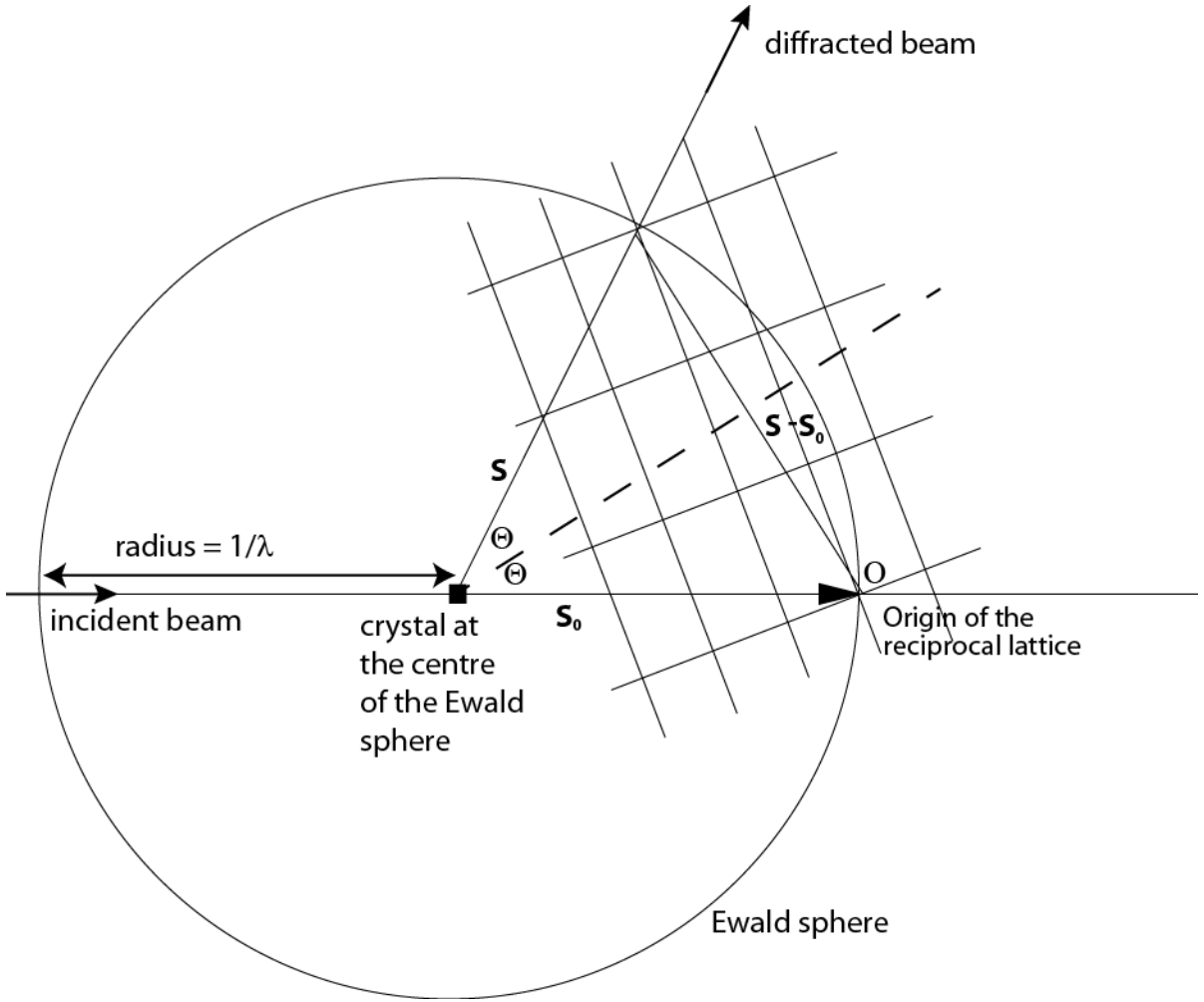
Bragg's law

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

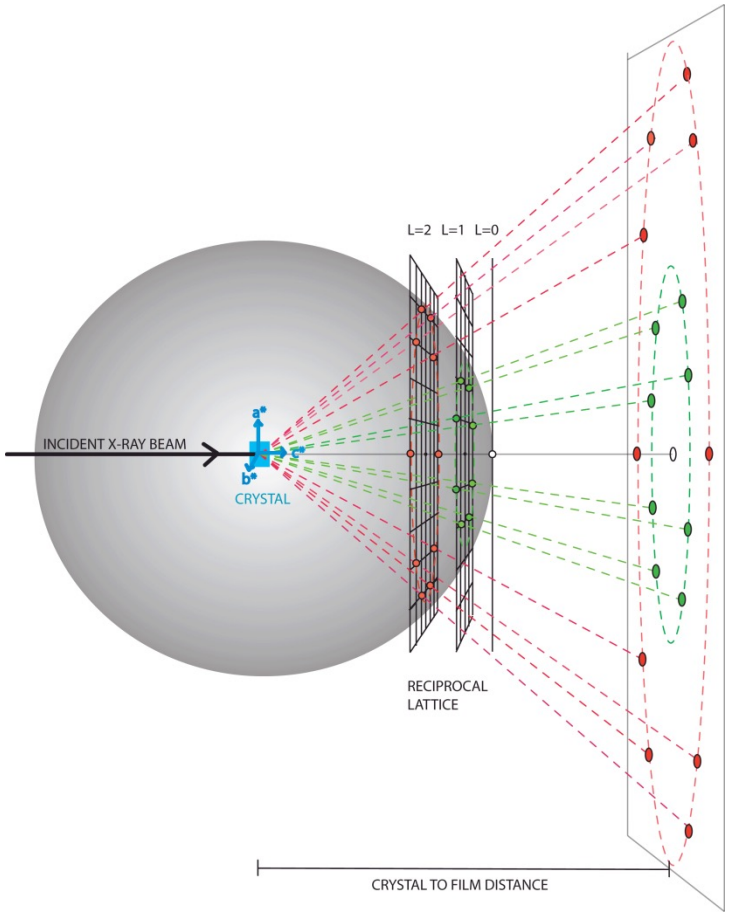
BioDiff, the corresponding view in reality:

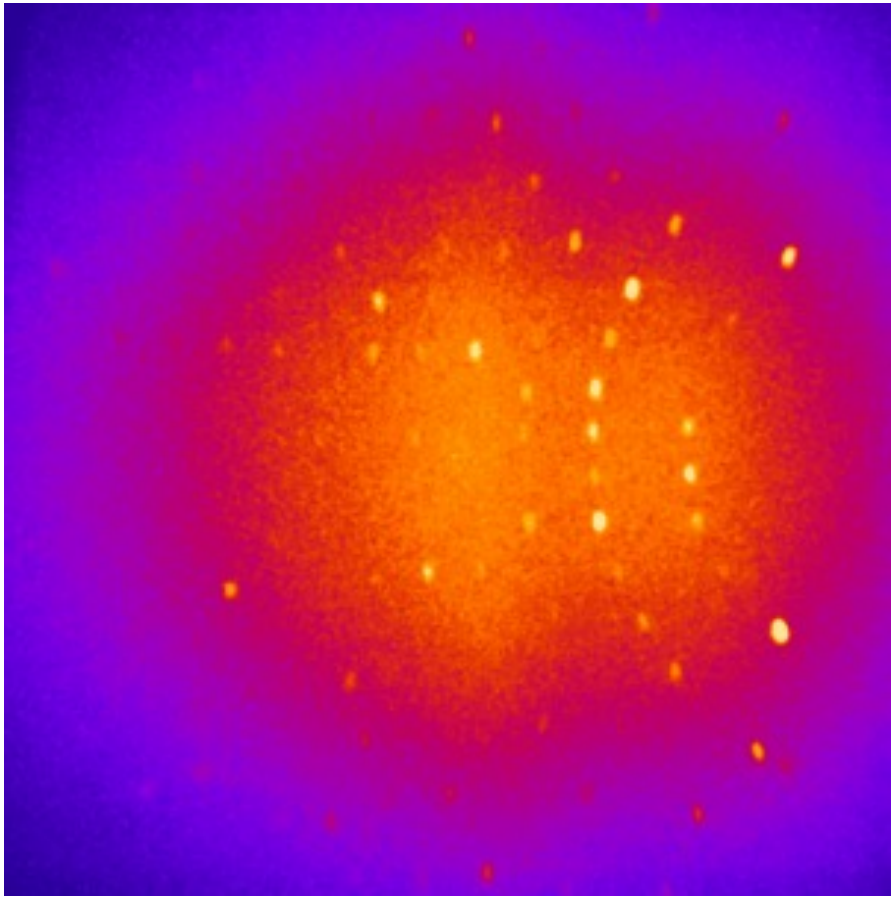


Ewald construction and Bragg‘s Law



Myoglobin protein crystal (deuterated mother liquor) full data set recorded with CCD

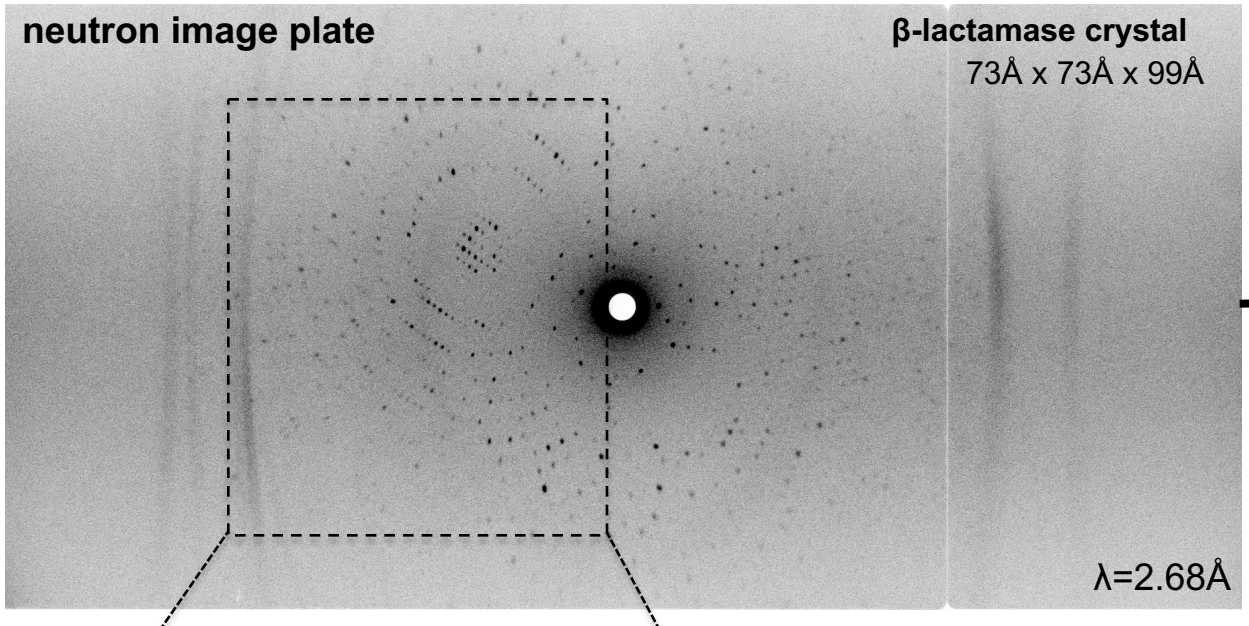


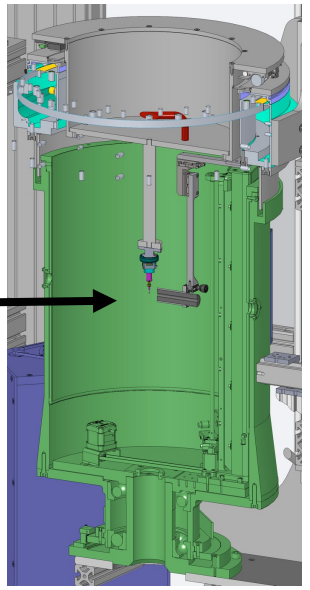


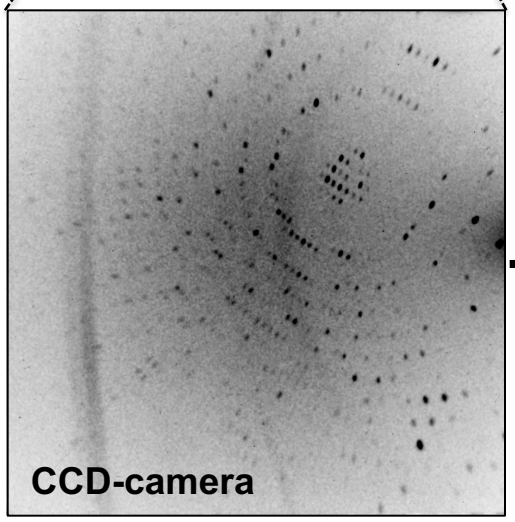
prim. beam

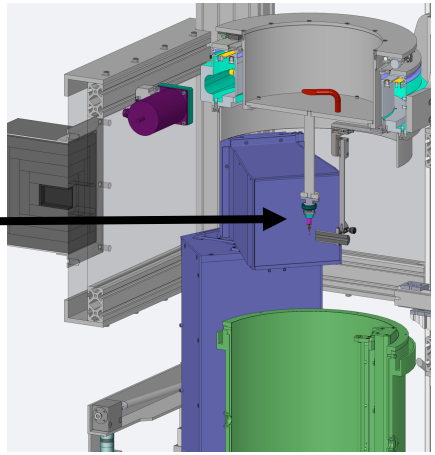
BioDiff: exposure time per frame: 20 minutes, sample: Myoglobin in deuterated mother liquor

The Two Detectors of BIODIFF









NIP-scanner

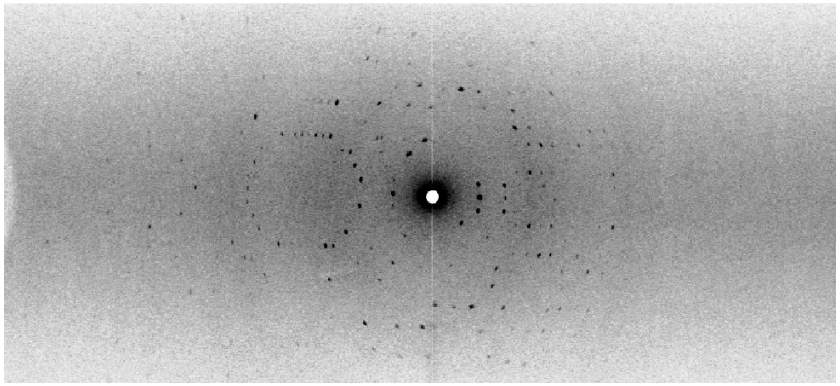
- larger solid angle
- readout time ≥ 4 min

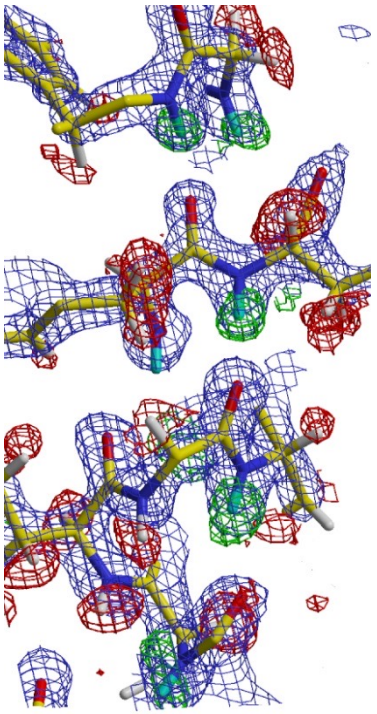
CCD-camera

- smaller solid angle
- readout time ≥ 1 sec

Theory on scattering from a crystal

3D structural analysis:





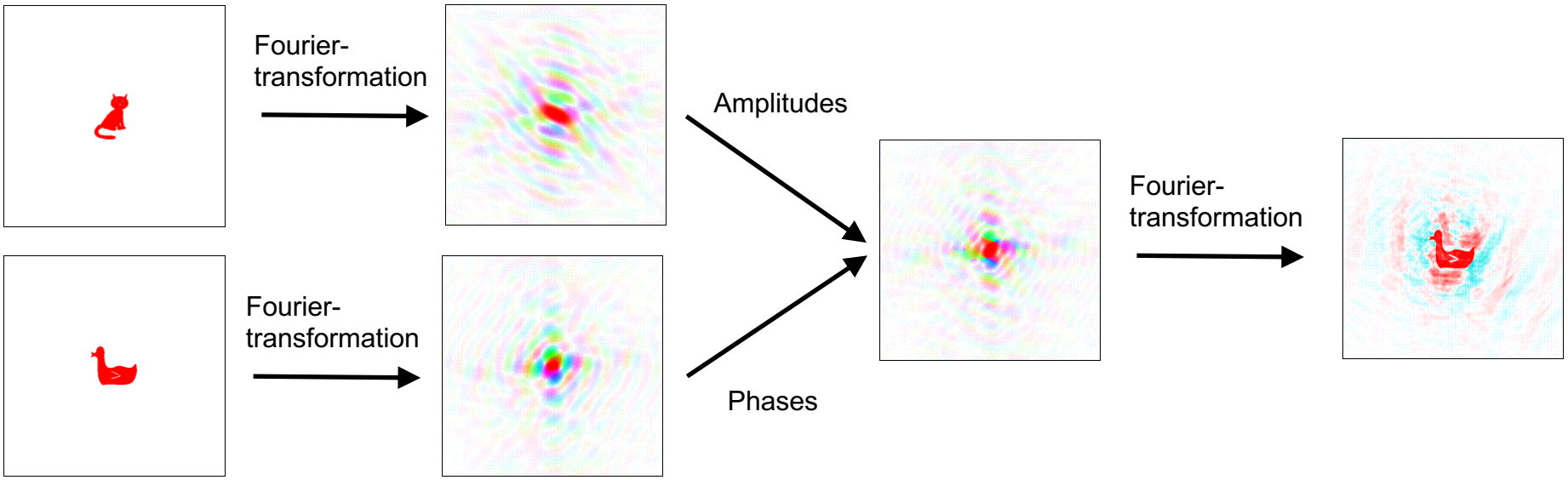
$$\rho(x,y,z) = \frac{1}{V_E} \sum_{h,k,l} F_{hkl} \cdot e^{-2\pi i(h\cdot x + k\cdot y + l\cdot z)}$$

Structure factors are complex numbers: $F_{hkl} = \|F_{hkl}\| e^{-2\pi i \alpha_{hkl}}$

with amplitudes $\|F_{hkl}\|$ and phases α_{hkl}

→ Phase Problem, because we only record intensities: $I = \|F_{hkl}\|^2$

The phases are stronger than the intensities



http://www.ysbl.york.ac.uk/~cowtan/fourier/magic.html

Solving the phase problem

• Structures of <1000 atoms can be solved by brute force statistical *Direct Methods*, given data to 1.2 Å or better. Most small molecule structures are solved this way.

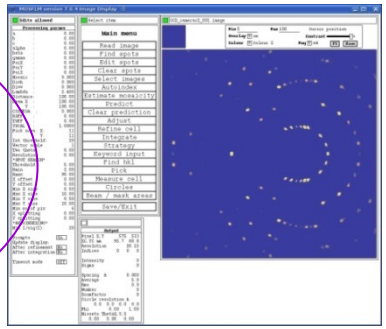
• A closely related known structure can be used as a search fragment for MR = *Molecular Replacement*.

• Heavy atoms can be introduced and the small changes in the reflection intensities exploited (SIR / MIR = Single / Multiple *Isomorphous Replacement*).

• Heavier atoms exhibit wavelength-dependent *Anomalous Scattering*, with the result that F_{hkl} and F_{-h-k-l} are not exactly equal. These small differences can be exploited in the Single Anomalous Diffraction and Multiple Anomalous diffraction (MAD) methods. For MAD, either metal atoms such as Fe present in the protein, or Se in selenomethionine (genetically modified methionine) are suitable anomalous scatterers.

Flow chart of data treatment and model building

Scans at varying crystal orientation
Scan := Series of detector images



-MOSFLM
-HKL-denzo
(comercial)

Data reduction

- determination of crystal orientation, unit cell dimensions etc.
- Calculating integral of reflection intensities

hkl-list for each scan:
h k l Intensity Intensity error

Scaling of each hkl list to match each other

-SCALA (CCP4-program package)

Unified hkl-list of measurement := complete data set

Calculation of a first map

Additional information from the solution of the phase problem

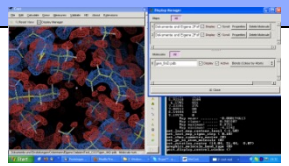
Structure refinement

-Refinement of atom coordinates displacements
- Calculation of scattering density maps (netrons) or electron density maps (x-rays)

Map-plotting

- inspection of model to fit the map)
- real space changes and refinement to the model

-nCNS
-PHENIX



-XtalView
-Coot

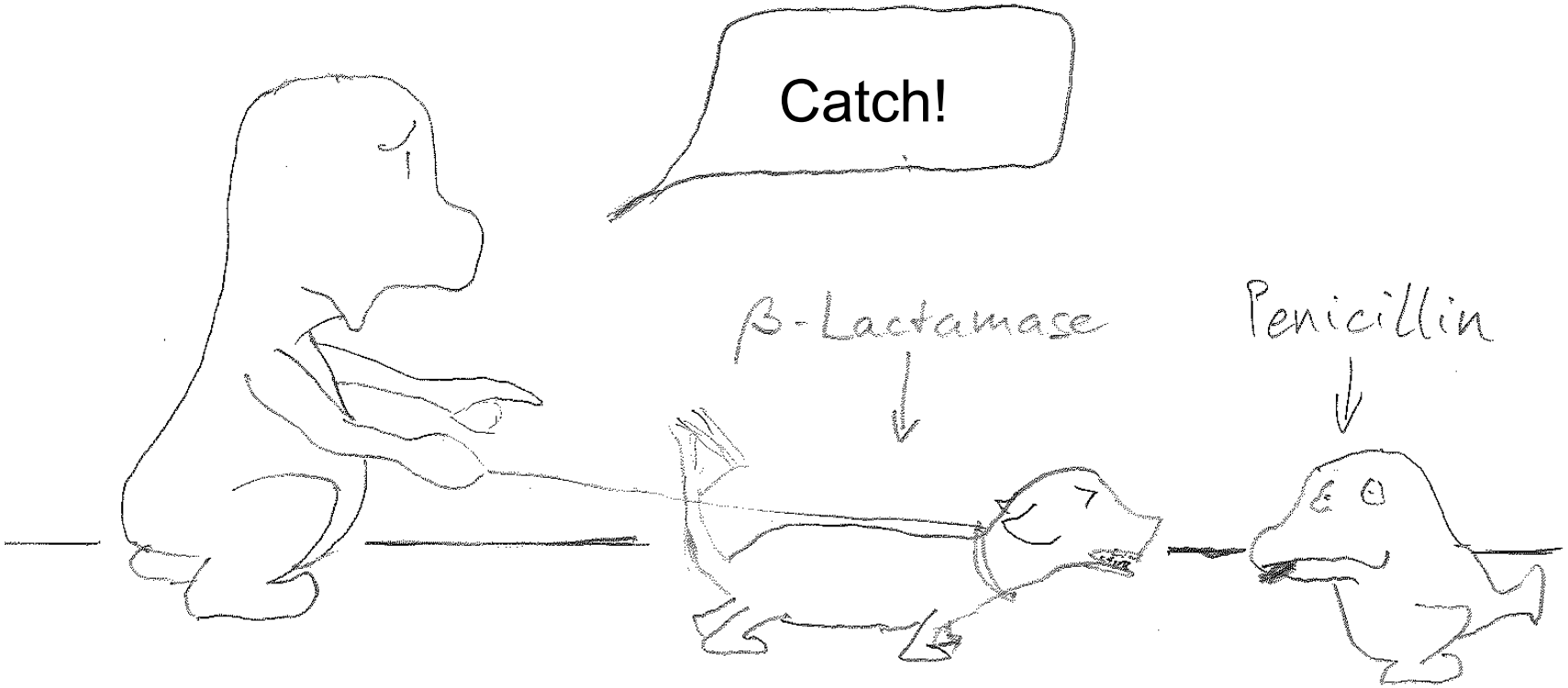
Neutron protein crystallography

Phase problem is solved by molecular replacement method using the structure obtained from the x-ray data.

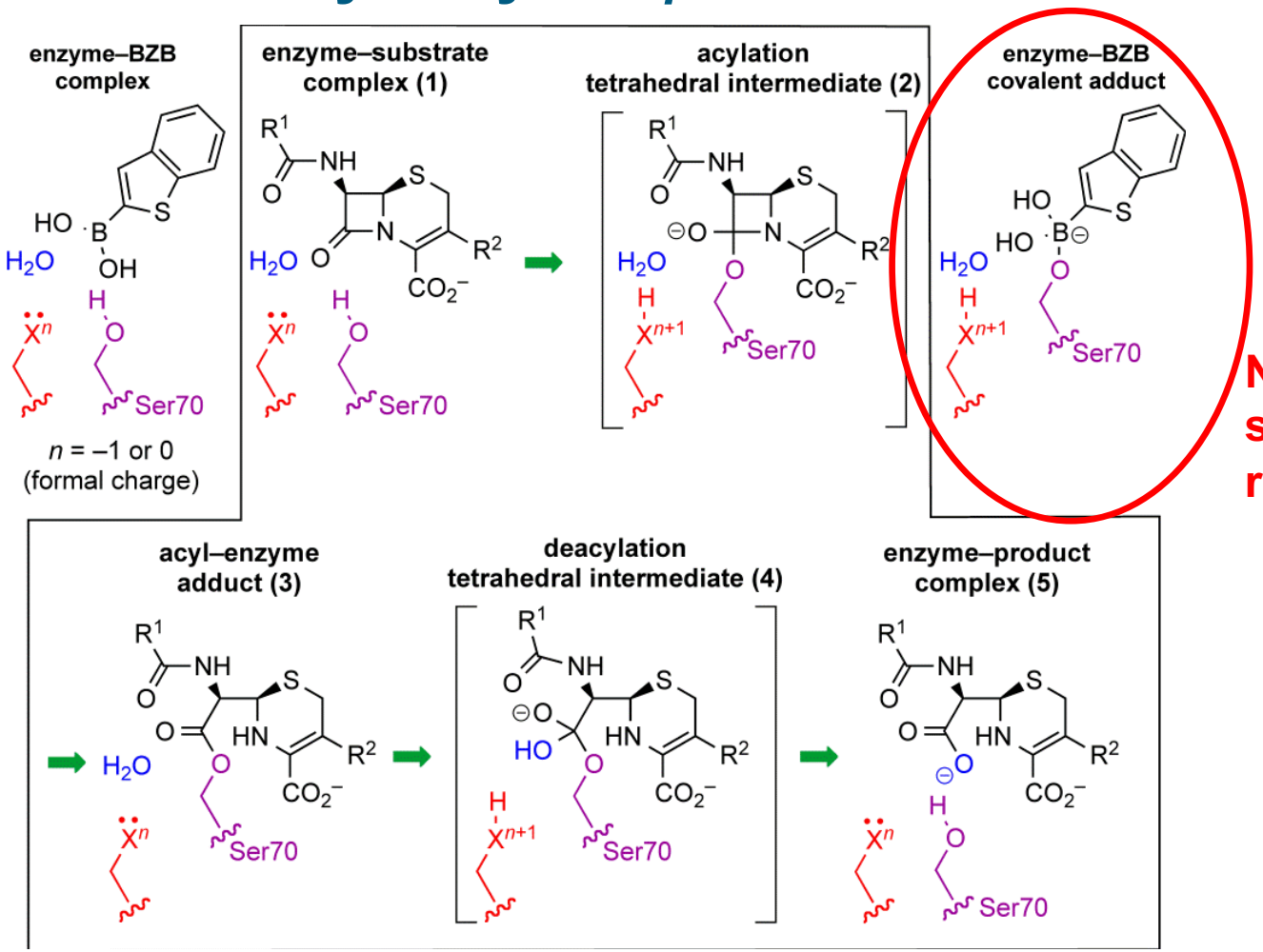
=> x-ray crystallography is a prerequisite of neutron protein crystallography.

Application Example: Protonation state of amino acid residues

The protein β-lactamase



β-lactamase: hydrolyses β-lactam antibiotics

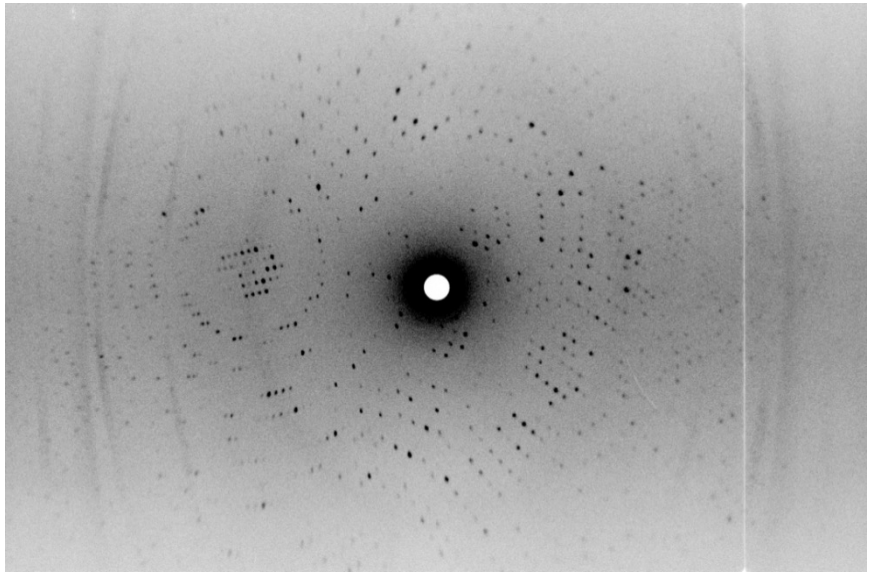


Neutron structure recorded

The catalytic cycle of a class A β-lactamase illustrated for a cephalosporin substrate (inside box) and the mode of inhibition by BZB (outside box). The general base employed is not necessarily the same for acylation and deacylation. The overall reaction pathway for β-lactam hydrolysis of a cephalosporin-like substrate by the class A β-lactamase enzymes.

Figures from: Tomanicek et al., doi: 10.1074/jbc.M112.436238

Data-set: β-lactamase with bound inhibitor



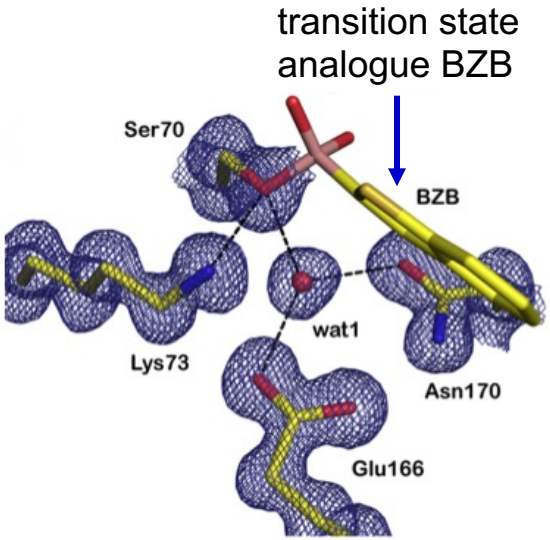
- unit cell: 73.4Å, 73.4Å, 99.1Å $P3_221$
- fully deuterated protein
- crystal size: 2.7mm^3
- Collection time: 9d

d_{min}	I/σ(I)	N_{meas}	mult.	compl. in shell %	R_{merge} %
4.31	27.8	12685	5.6	97.6	4.9
3.42	19.0	11941	5.5	98.0	8.0
2.99	10.3	10378	4.9	96.9	14.6
2.71	7.6	8757	4.3	95.5	18.7
2.52	5.9	7820	3.9	92.8	21.2
2.37	5.4	7099	3.8	89.2	21.6
2.25	5.0	6095	3.5	84.6	23.0
2.15	4.5	5906	3.4	82.9	24.7
2.07	4.1	5673	3.2	82.0	27.2
2.0	3.7	5059	2.9	81.2	27.9
overall	**7.4**	**81413**	**4.0**	**90.2**	**14.7**

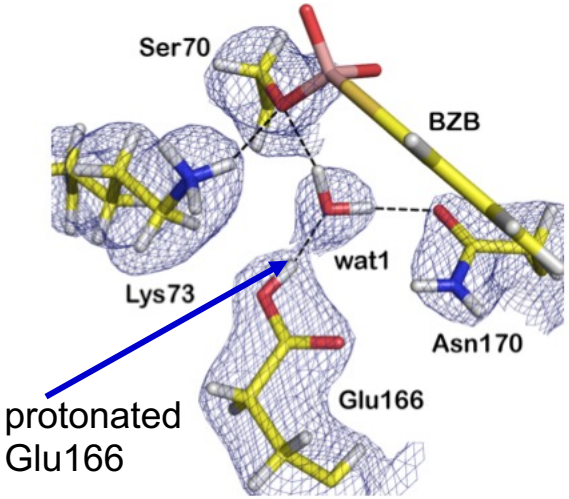
R_{pim} = 7.9% (17.9%)

Tomanicek et al., J. Biol. Chem., 288, 4715 (2013).

Catalytic Proton Network of the Toho-1 β-Lactamase



electron density map



nuclear density map from BioDiff

Glu166 acts as the general base during the catalytic action of the enzyme.

Stephen J. Tomanicek, Robert F. Standaert, Kevin L. Weiss,
Andreas Ostermann, Tobias E. Schrader, Joseph D. Ng, and Leighton Coates
J. Biol. Chem. 2013, 288:4715-4722

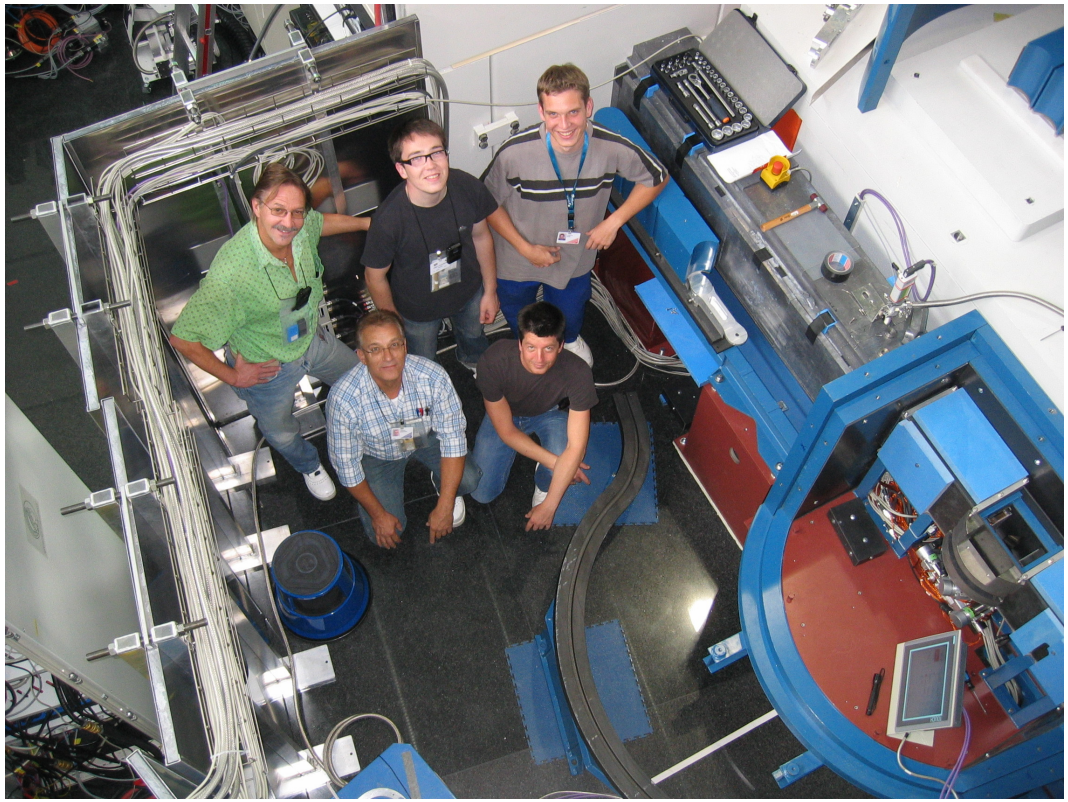
Summary

- Proteins show a special 3-D structure which is specific to their function
- x-ray crystallography: Most of the beautiful schematic pictures of proteins in textbooks of chemistry and molecular biology represent structures determined by X-ray diffraction. Advantages:
  1. only small crystals needed
  2. short measurement times enable large throuput
  3. phase problem can be solved with more and more sophisticated methods

  Disadvantages:
  1. radiation damage often observed: hydrogen abstraction, reduction of metal centres in the metalo-proteins, disulfide bond cleavage.
  2. Hydrogen positions can usually not be determined (only at high resolution)
- Neutron protein crystallography is a complementary technique as compared to x-ray crystallography. Here one can determine:
  1. protonation states of amino acid side chains (important for the function of the protein)
  2. deuterium exchange as a measure of flexibility and accessibility (discrimination between **H** / **D**)
  3. solvent structure including hydrogen atoms

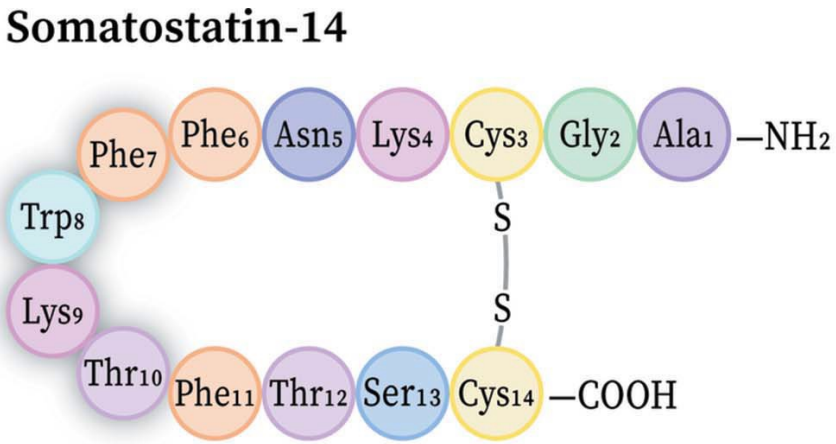
Thanks to... ... the BioDiff-Team:

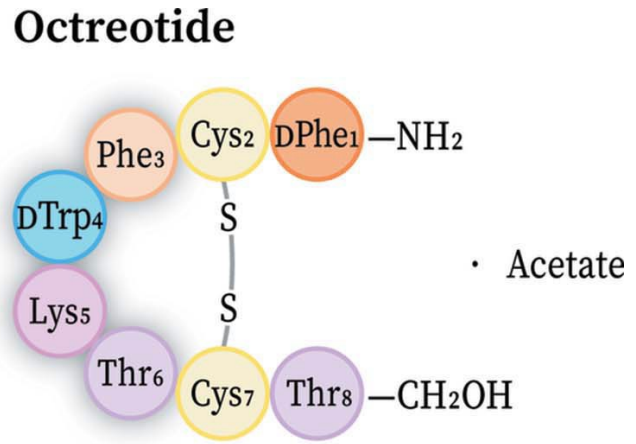
Andreas Ostermann
- Philipp Jüttner
- Reinhard Schätzler
- Bernhard Laatsch
- Frank Suxdorf
- Manfred Bednarek
- Matthias Drochner
- Harald Kleines
- Kevin Körrentz
- Karl-Heinz Mertens
- Michael Monkenbusch
- Nikolas Arendt
- Christian Felder
- Michael Wagener
- Lydia Fleischhauer-Fuss
- Vladimir Ossovyi
- Andreas Nebel
- Simon Staringer
- Harald Kusche
- Winfried Petry
- Dieter Richter



... and you for your attention!

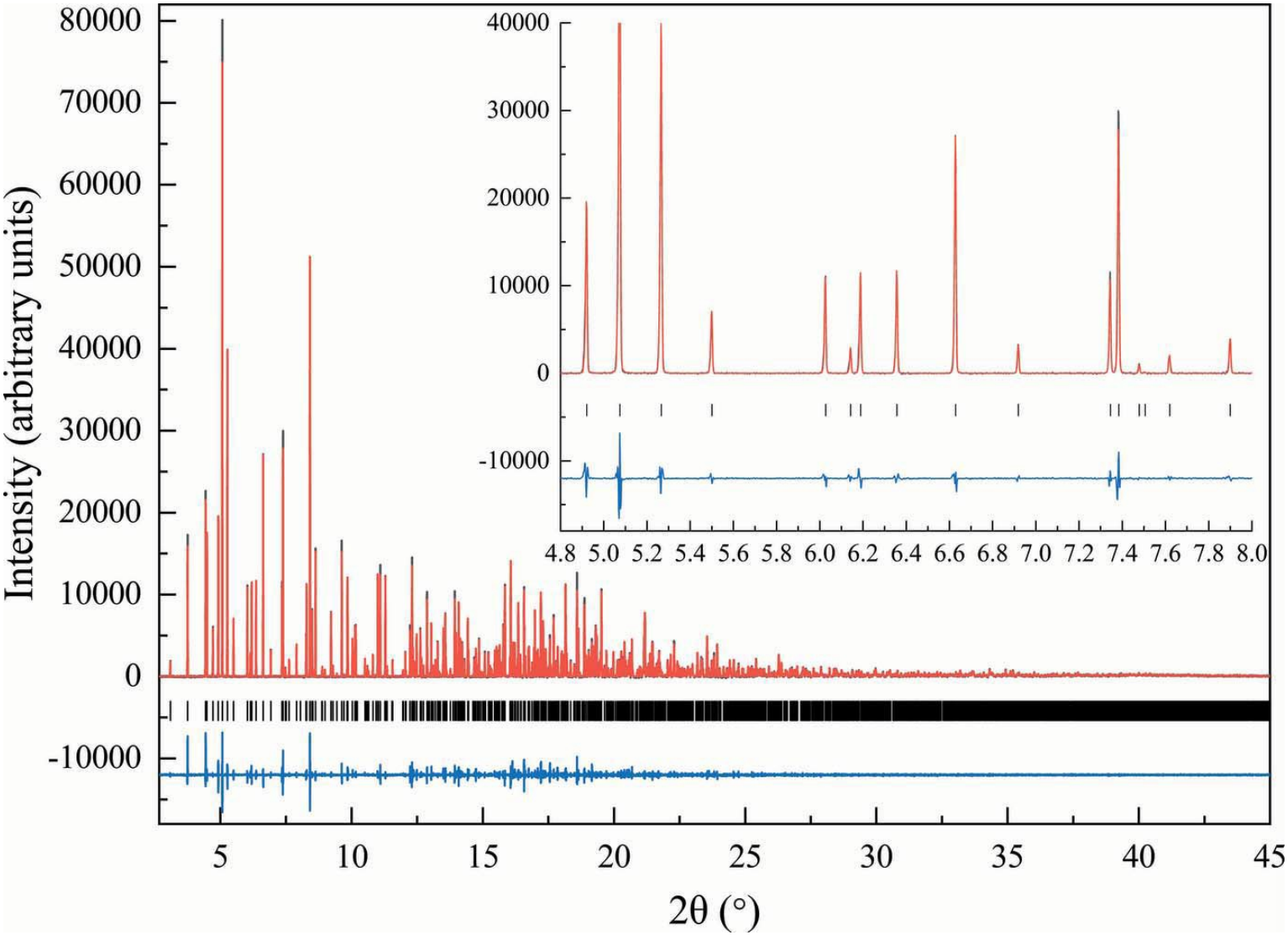
Why not do powder diffraction on protein crystals?





Unit cell size: 39 A, 30 A, 18 A

Acta Cryst. (2021). A77, 186–195


Intensity (arbitrary units)
80000
70000
60000
50000
40000
30000
20000
10000
0
-10000
5
10
15
20
25
30
35
40
45
2θ (°)
40000
30000
20000
10000
0
-10000
4.8 5.0 5.2 5.4 5.6 5.8 6.0 6.2 6.4 6.6 6.8 7.0 7.2 7.4 7.6 7.8 8.0

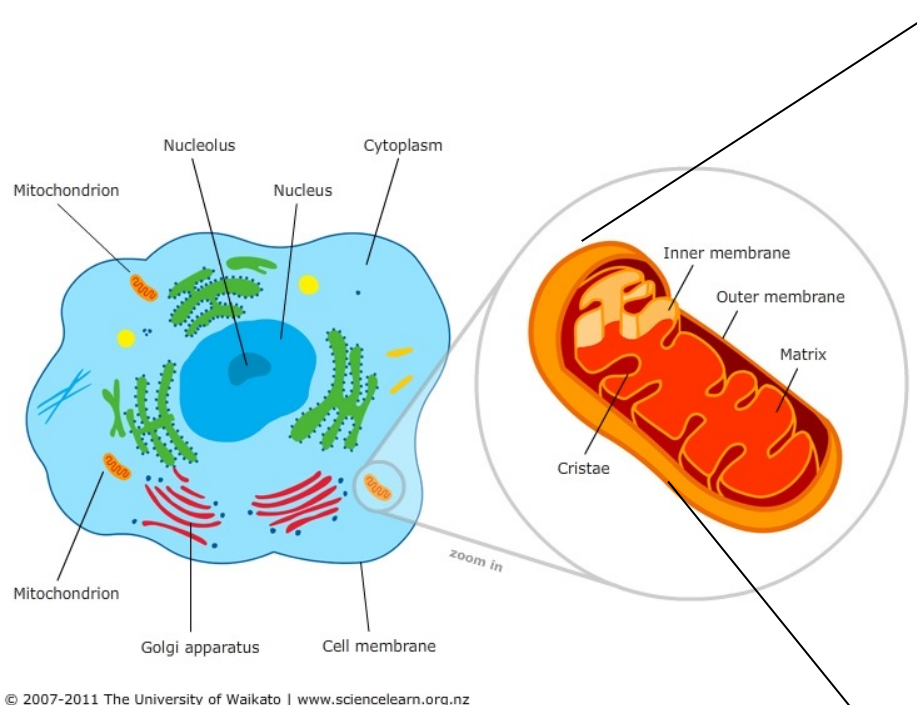
An example for a metallo-protein:

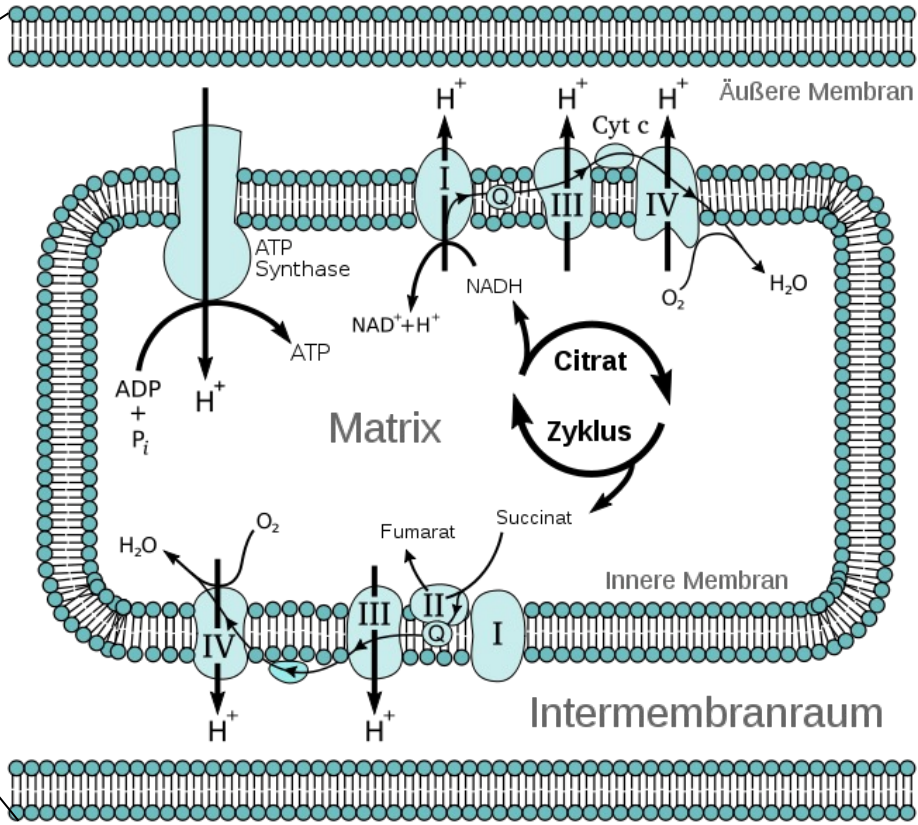
Cytochrome *c* peroxidase, or CcP is a water-soluble heme-containing enzyme of the peroxidase family that takes reducing equivalents from cytochrome *c* and reduces hydrogen peroxide to water:

$$CcP + H_2O_2 + 2\ \text{ferrocytochrome } c\ (Fe^{2+}) + 2H^+ \rightarrow CcP + 2H_2O + 2\ \text{ferricytochrome } c\ (Fe^{3+})$$

(taken from http://en.wikipedia.org/wiki/Cytochrome_c_peroxidase)

Mitochondria are the power plant of a cell (production of ATP):

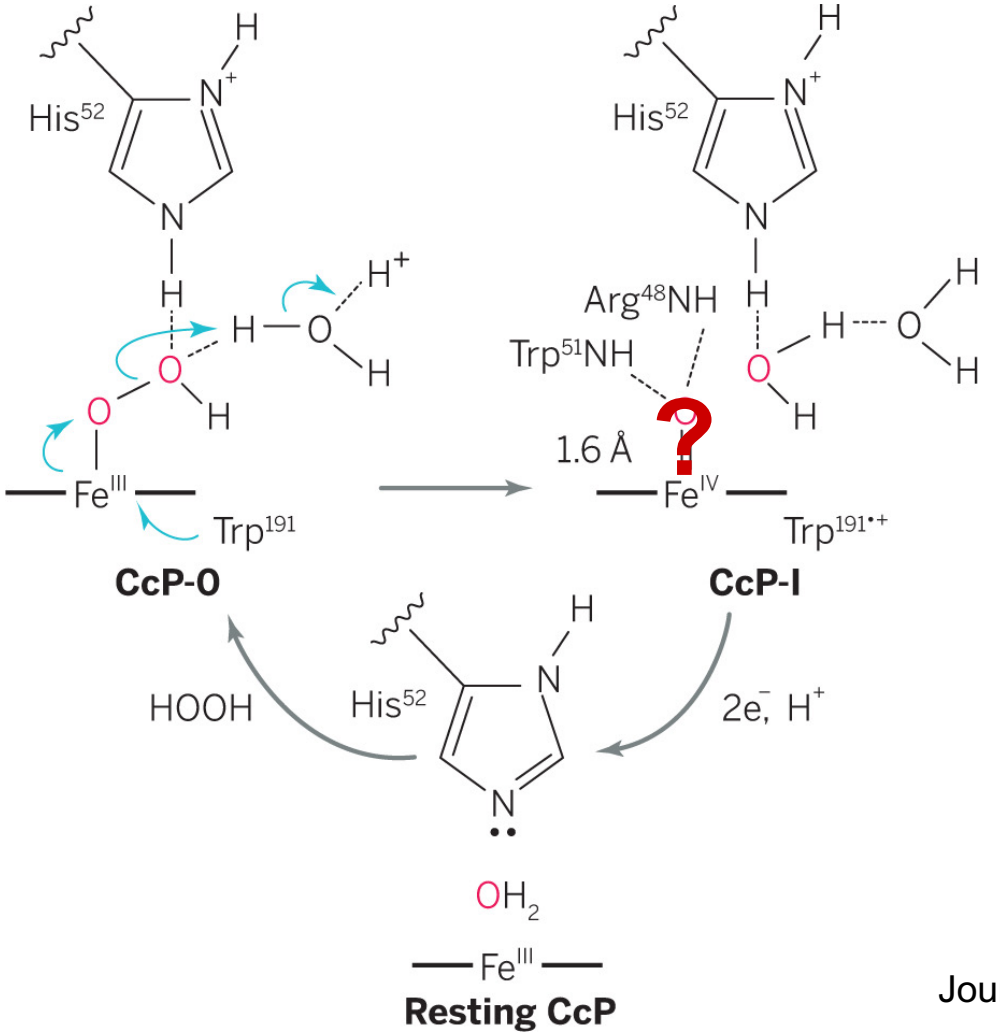




http://de.wikipedia.org/wiki/Atmungskette

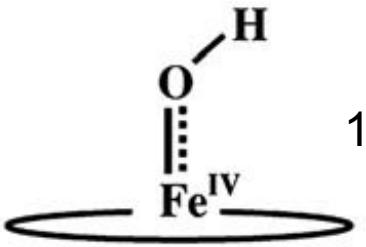
- Cytochrome C serves as an electron transporter in the respiratory chain.
- Cytochrome c Peroxidase uses two ferro-cytochrome C proteins to reduce H_2O_2 to water and two ferricytochrome C molecules

Proton-mediated mechanism.Reaction of ferric CcP with H_2O_2 first gives CcP-0, followed by O-O bond scission driven by external protonation to afford CcP-I.



J T Groves, and N C Boaz Science 2014;345:142-143

Alternative Hypothesis:

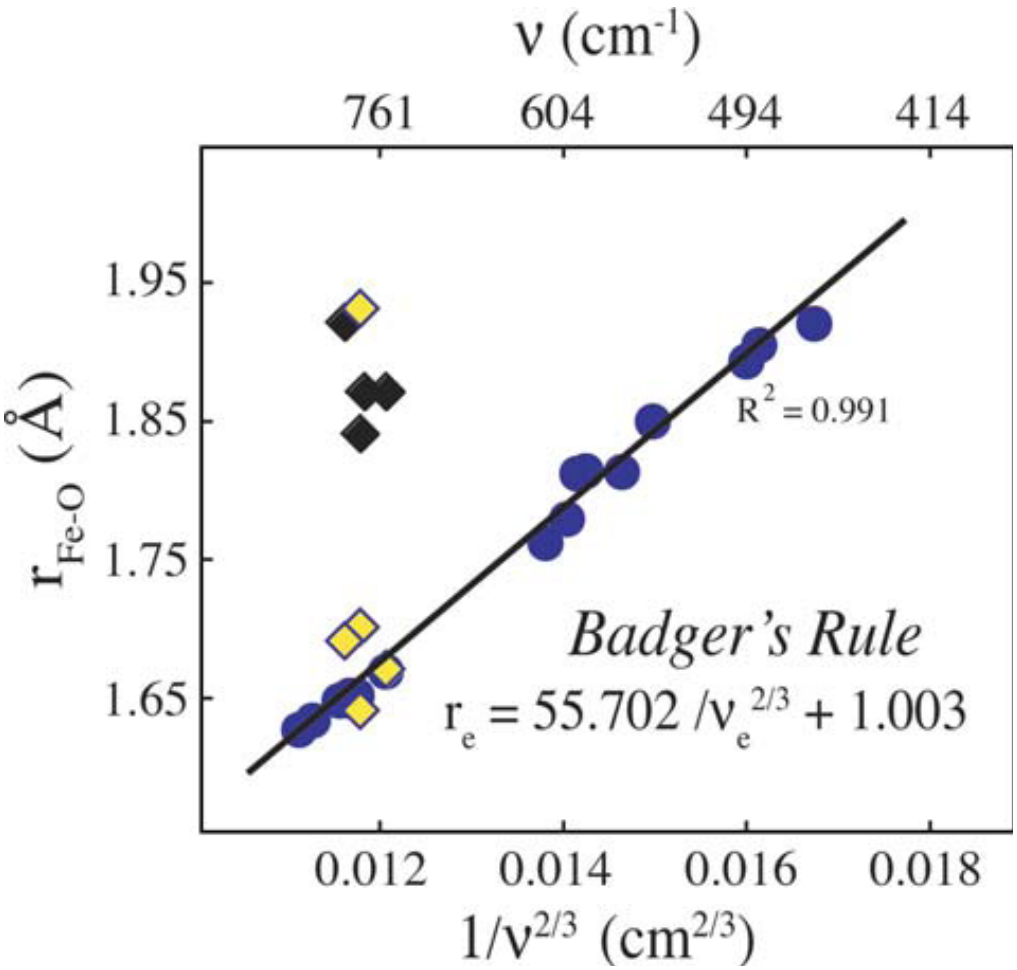


1.85 Å

Fig. 3. Compound I with an O–H bond and a bond length of Fe-O of ca. 1.85 Å.

Journal of Inorganic Biochemistry 100 (2006) 448–459

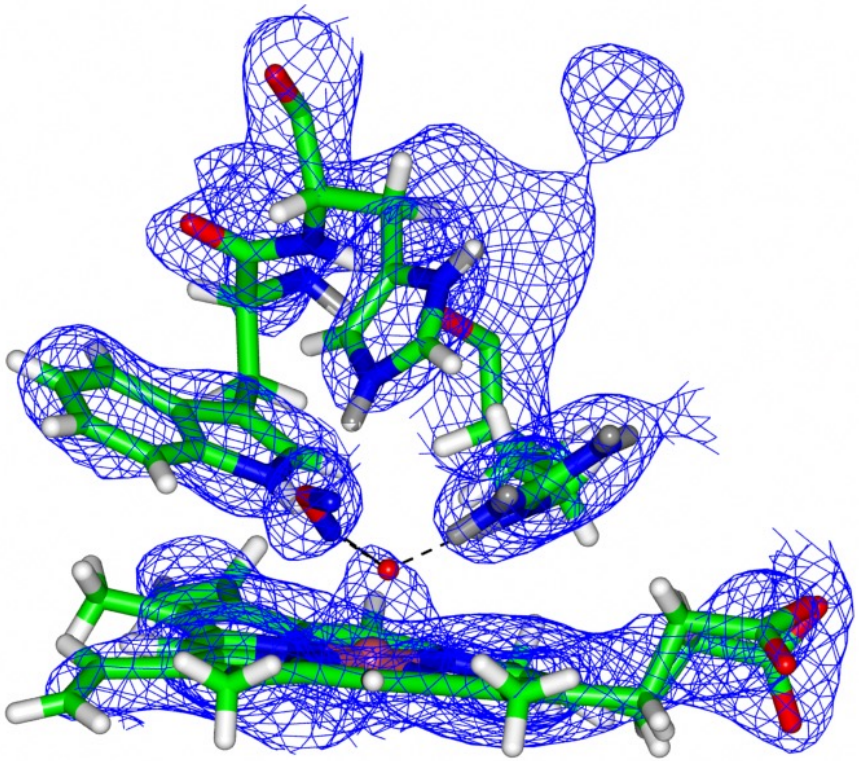
No method so far could unambigiously show the nature of the iron-oxide bond



Plot of computed stretching frequency vs Fe–O bond distance. **Yellow diamonds** are from resonance Raman or EXAFS and the solid diamonds from X-ray crystal structures. The **blue circles** are from calculations.

Figure taken from: Journal of Inorganic Biochemistry 100 (2006) 448–459

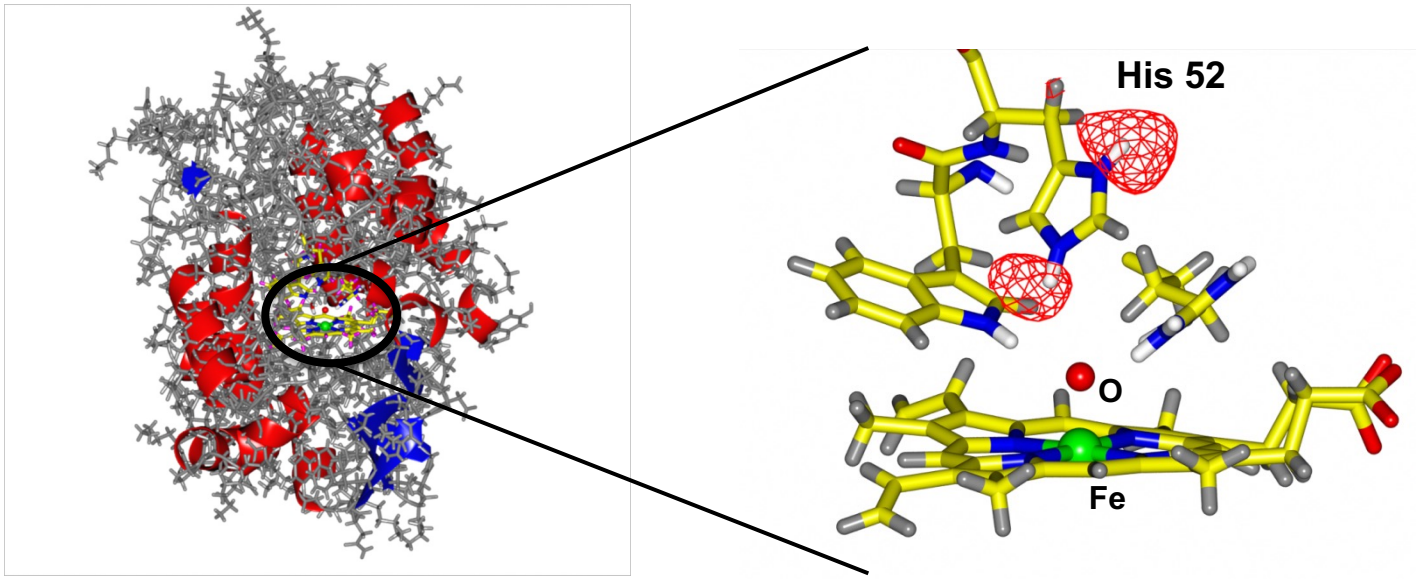
Compound I of Cytochrome c Peroxidase



Cecilia M. Casadei, Andrea Gumiero, Clive L. Metcalfe, Emma J. Murphy, Jaswir Basran, Maria Grazia Concilio, Susana C. M. Teixeira, Tobias E. Schrader, Alistair J. Fielding, Andreas Ostermann, Matthew P. Blakeley, Emma L. Raven, Peter C. E. Moody, Science 2014;345:193-197

Omit-Map for the two exchangable hydrogen atoms at His52

Cytochrome-c-Peroxidase, Compound I at 100 K:



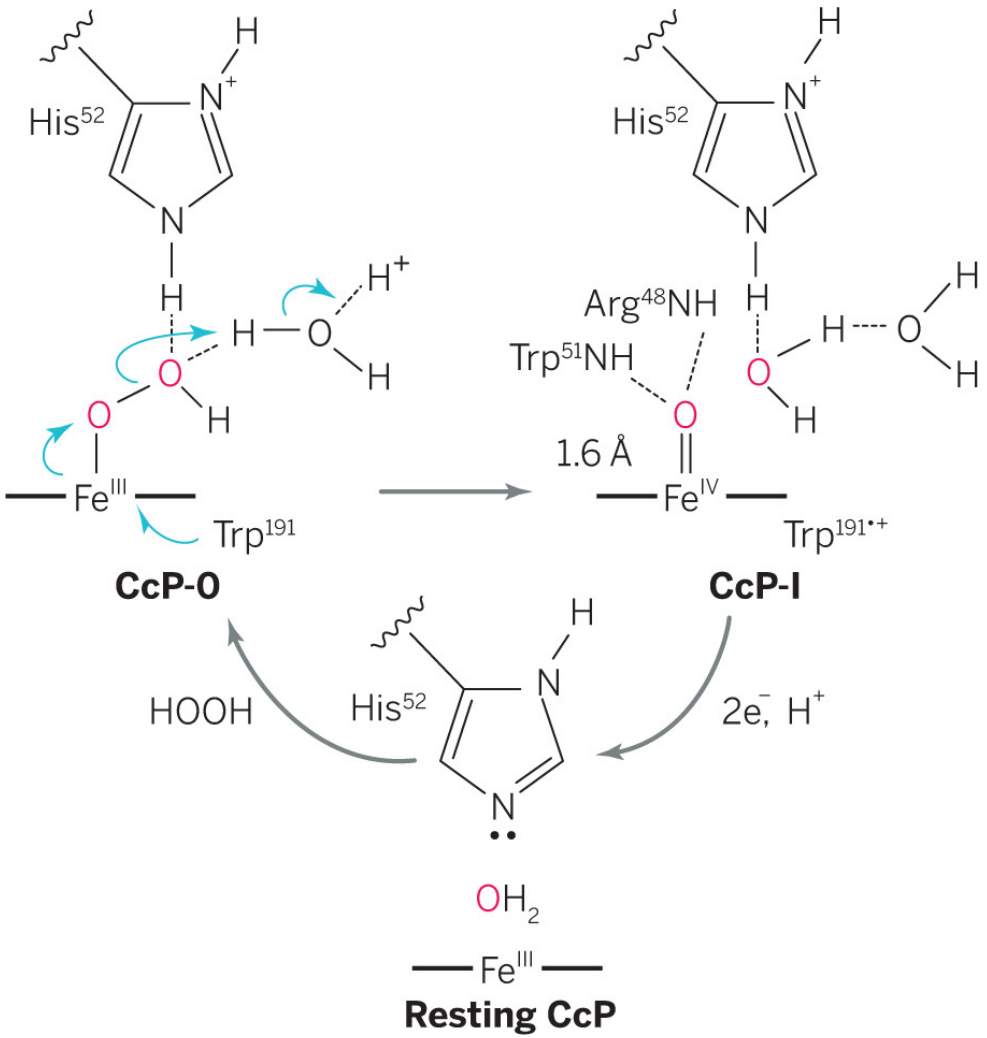
→ The oxygen atom bound to iron is not protonated.

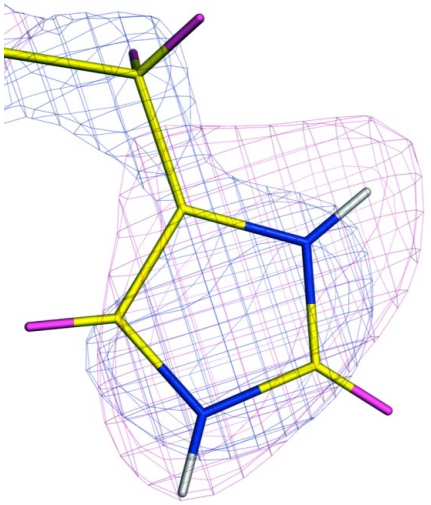
→ The amino-acid His52 is doubly protonated

↷ The reacton mechanism has to be thought over again!

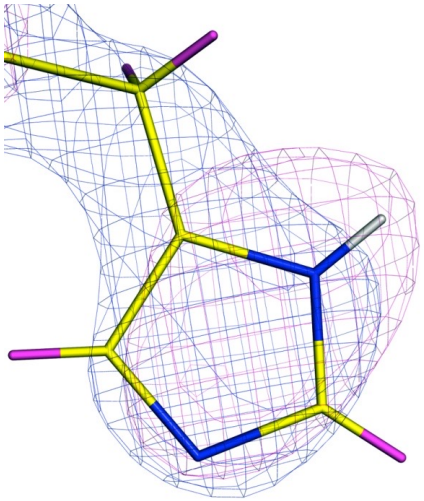
Casadei et al. Science **345**, 193 (2014)

Proton-mediated mechanism.Reaction of ferric CcP with H2O2 first gives CcP-0, followed by O-O bond scission driven by external protonation to afford CcP-I.





His 52
Compound I



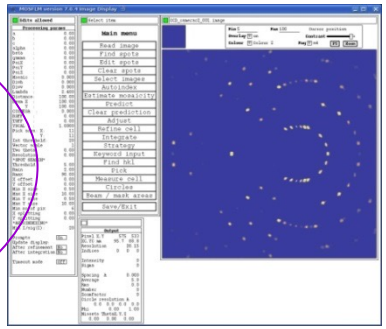
His 52 ferric
(resting)

J T Groves, and N C Boaz Science 2014;345:142-143

Tutorial: Drug development based on Fragment screening: Trypsin + Aminopyridin

Fow chart of data treatment and model building

Scans at varying crystal orientation
Scan := Series of detector images



-MOSFLM
-HKL-denzo
(comercial)

Data reduction

- determination of crystal orientation, unit cell dimensions etc.
- Calculating integral of reflection intensities

hkl-list for each scan:
h k l Intensity Intensity error

Scaling of each hkl list to match each other

-SCALA (CCP4-program package)

Unified hkl-list of measurement := complete data set

Calculation of a first map

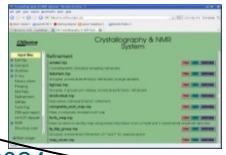
Additional information from the solution of the phase problem

Struktur refinement

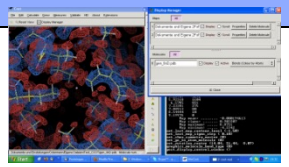
-Refinement of atom coordinates displacements
- Calculation of scattering density maps (netrons) or electron density maps (x-rays)

Map-plotting

- inspection of model to fit the map)
- real space changes and refinement to the model

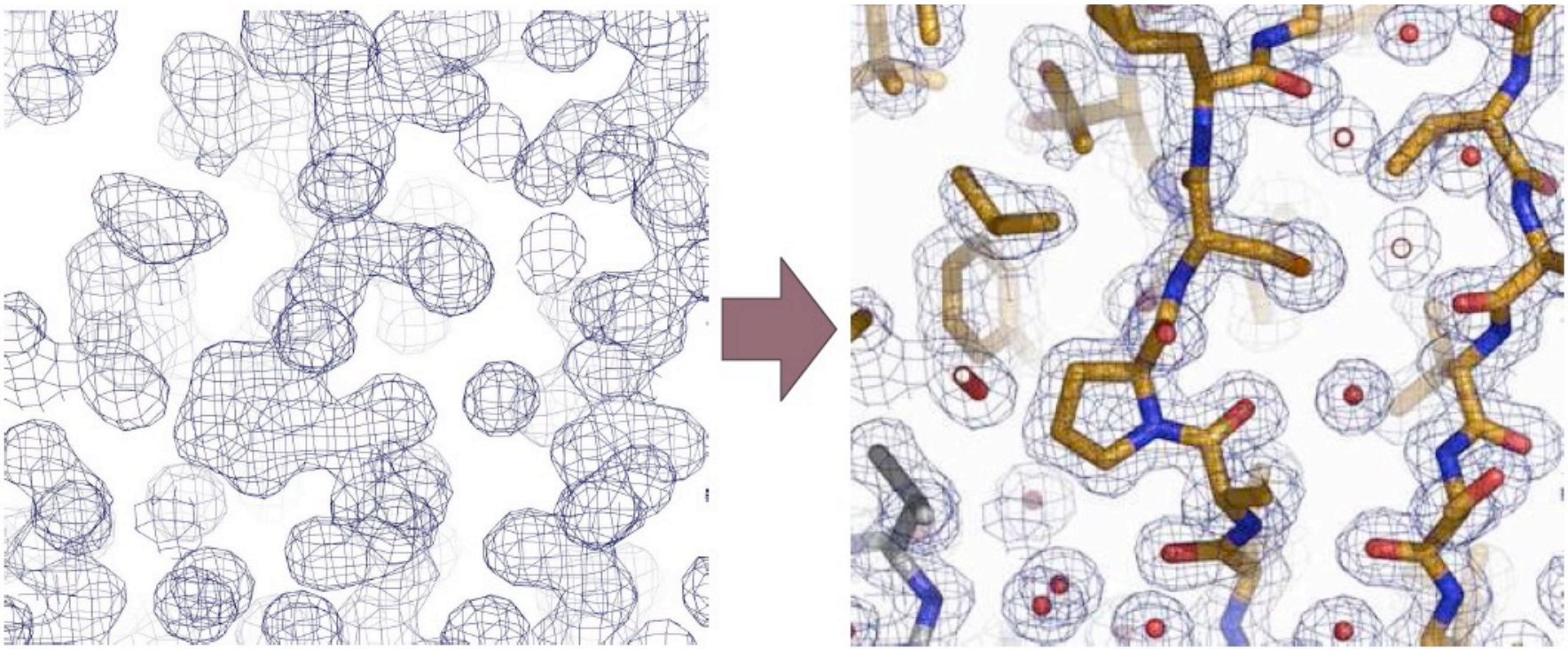


-nCNS
-PHENIX



-XtalView
-Coot

Structural Refinement: Putting the model in and applying changes in real space



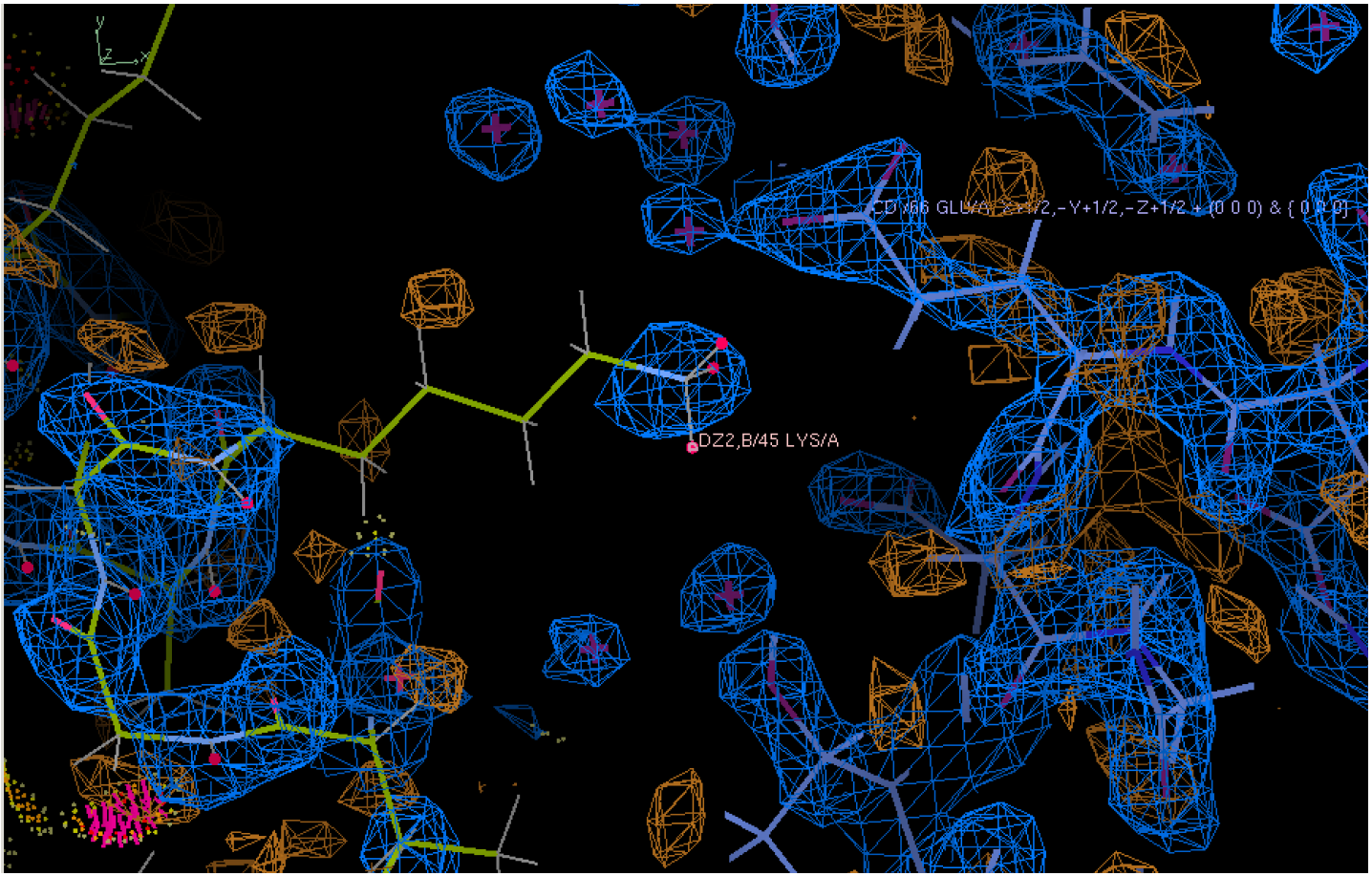
Electron density (x-ray) or nuclear density map (neutrons)

Amino acid chain is fitted into electron densities via dedicated software

from Armstrong.pdf: Qiagen

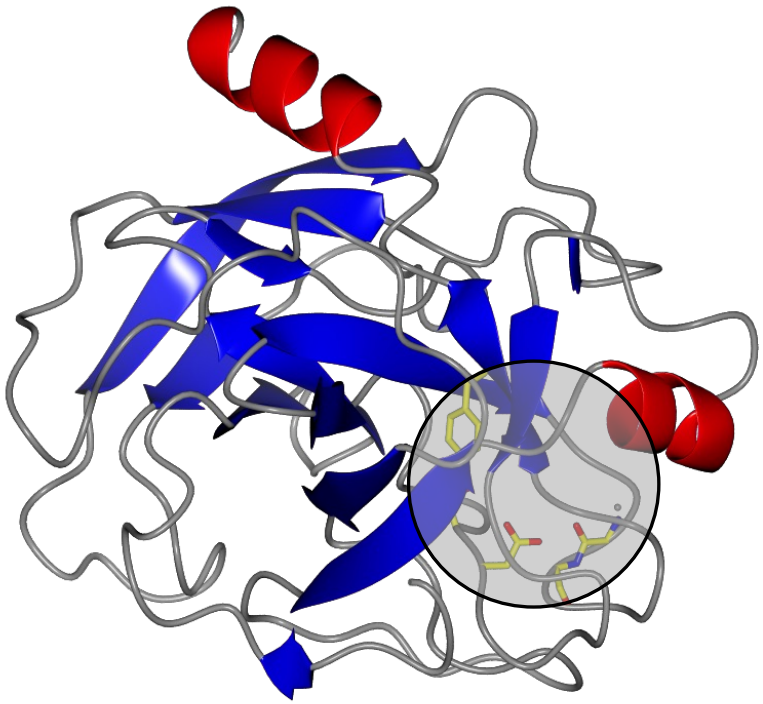
Lysine side chain on the surface of Alcoholdehydrogenase

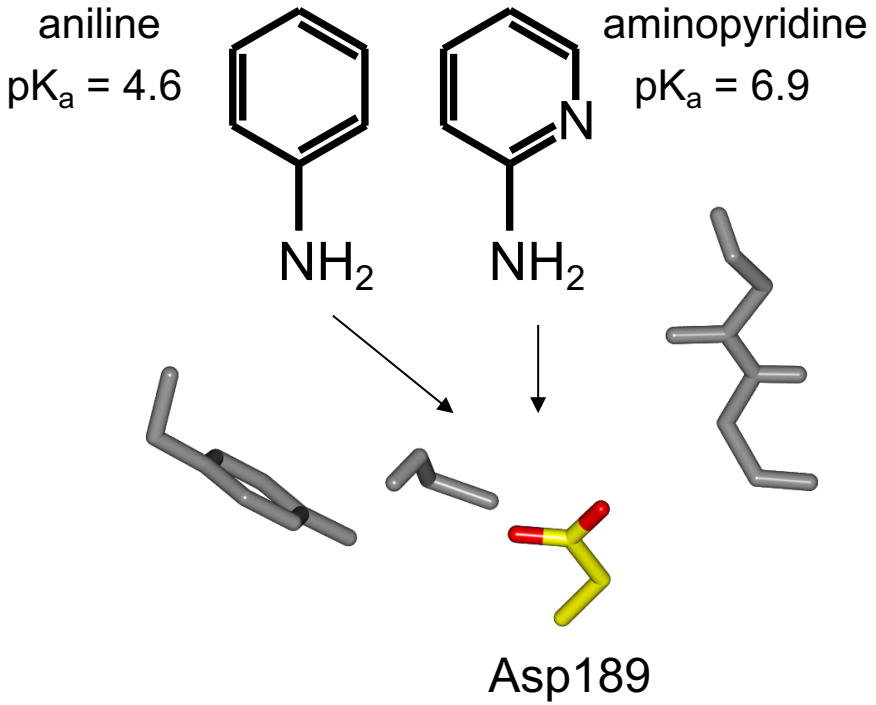
Johannes Hermann et al. Acta Cryst. (2018). F74, 754–764



Please look at the scattering (length) density not only at coordinates…

Inhibitor binding to trypsin: charges shift protonation



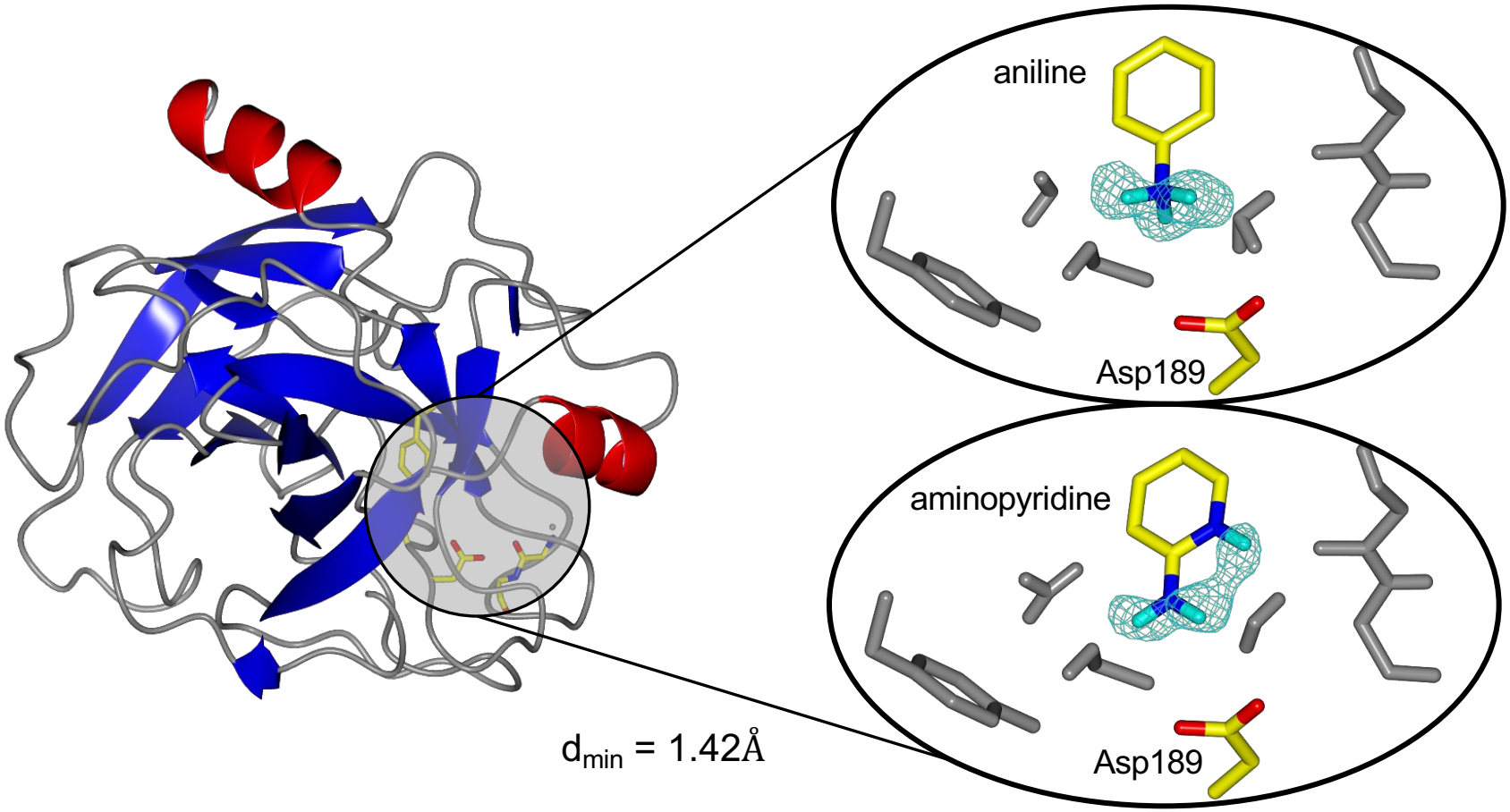


➡ question: inhibitors with less basic properties become protonated upon binding ?

Group of Prof. G. Klebe (Univ. Marburg)

Schiebel J. et al., Angewandte Chemie Int. Edition (2017), 56 (17):4887-4890

Inhibitor binding to trypsin: charges shift protonation



Group of Prof. G. Klebe (Univ. Marburg)

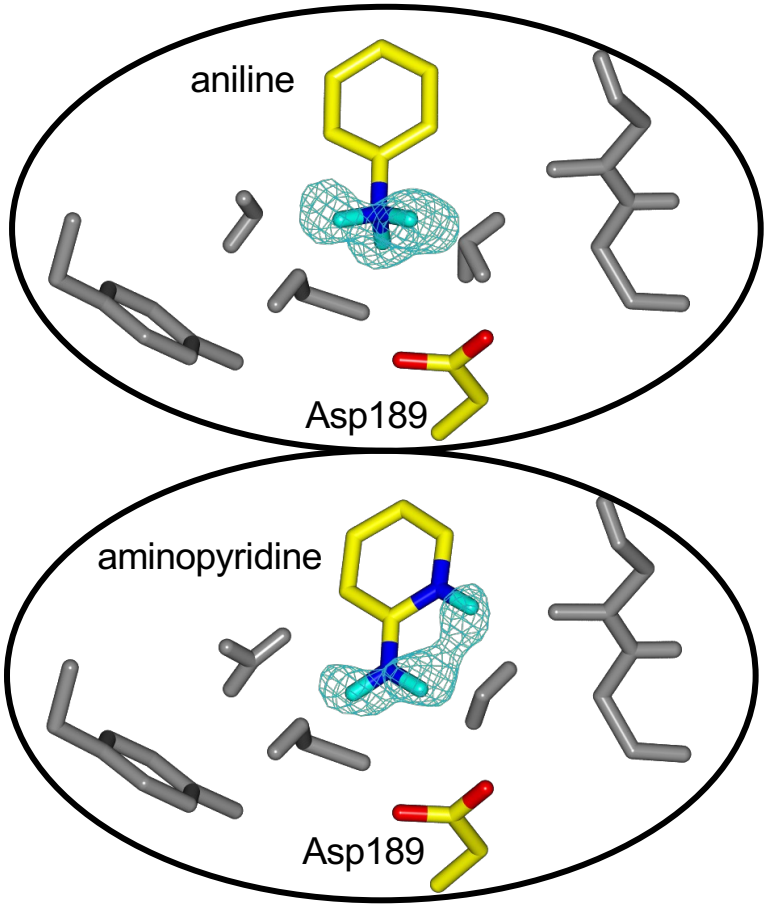
Fo-Fc omit-map; +3.5σ

Schiebel J. et al., Angewandte Chemie Int. Edition (2017), 56 (17):4887-4890

Inhibitor binding to trypsin: charges shift protonation

- Therefore, apart from charge–charge distances, tautomer stability is essential for the resulting protonation pattern

- correct prediction of such properties is key in drug development !



Group of Prof. G. Klebe (Univ. Marburg)

Fo-Fc omit-map; +3.5σ

Schiebel J. et al., Angewandte Chemie Int. Edition (2017), 56 (17):4887-4890

The R-Factors: Judging Convergence between model and data

To compare the generated electron density map and your model, you have to use the R-factor.

The R-factor is a measure of convergence between the intensities calculated from your model and the observed intensities.

$$R_{work} = \frac{\sum_{\mathbf{h}} \left| \left| F_{obs}(\mathbf{h}) \right| - \left| F_{calc}(\mathbf{h}) \right| \right|}{\sum_{\mathbf{h}} \left| F_{obs}(\mathbf{h}) \right|}$$

R:

0.6-VERY BAD

0.5 -BAD

0.4-Recoverable

0.2-Good for Protein

0.05-Good for small organic models

0-PERFECT FIT

However R_{work} can be reduced artificially by refining more parameters, so now it is usual to reserve 5 to 10% of the reflections to calculate an index R_{free} (same formula).

Table 1. Data collection statistics

Data collection	TAFI	TAFI-GEMSA	TAFI-IIYQ
Space group	P31 2 1	P31 2 1	P31 2 1
Cell dimensions			
a, *b*, *c*, Å	161.7, 161.7, 139.5	161.1, 161.1, 139.0	159.5, 159.5, 139.5
α, β, γ, °	90, 90, 120	90, 90, 120	90, 90, 120
Resolution, Å	49.51-3.10 (3.27-3.10)	49.27-3.40 (3.58-3.40)	49.03-2.80 (2.95-2.80)
R_{merge}	0.105 (0.833)	0.091 (0.630)	0.067 (0.600)
$I / \sigma I$	13.9 (2.6)	11.4 (2.2)	13.8 (2.1)
Completeness, %	100.0 (100.0)	99.6 (100.0)	100.0 (100.0)
Redundancy	8.6 (8.7)	4.9 (5.0)	4.3 (4.4)

Values in parentheses are for highest-resolution shell. A single crystal was used for each structure.

Table 2. Refinement statistics

Refinement	TAFI	TAFI-GEMSA	TAFI-IIYQ
Resolution, Å	3.1	3.4	2.8
No. reflections	36638	27424	48058
R_{work} / R_{free}	0.204 / 0.240	0.203 / 0.258	0.188 / 0.232
No. atoms			
Protein	9861	9892	9901
Ligand/ion	0 / 3	45 / 3	36 / 3
B-factors, Å²			
Protein	94.8	113.3	67.5
Ligand/ion	— / 102.9	111.9 / 117.5	72.4 / 71.3
R.m.s. deviations			
Bond lengths, Å	0.018	0.013	0.018
Bond angles, °	1.751	1.453	1.871

— Indicates not applicable.

Judging the quality of the data:

- I/σ = signal/noise >2.0
- Completeness > 90%
- R/R_{free}: difference < 0.05,
- Deviations of known geometry: weird angles

A crystal structure according to the protein data bank (PDB)

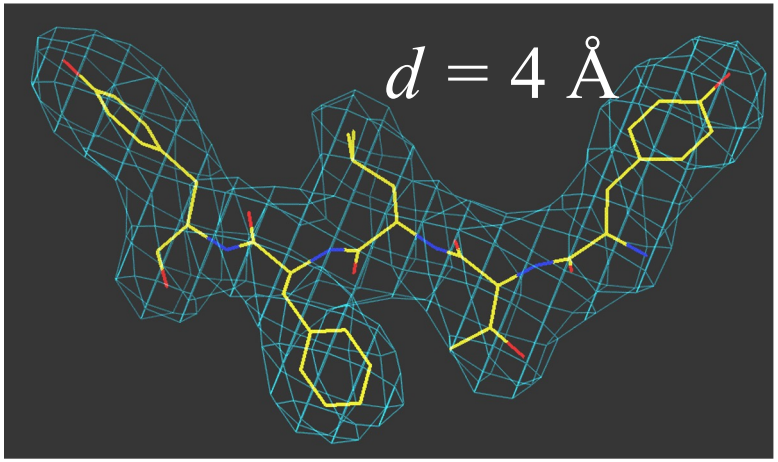
occupancy

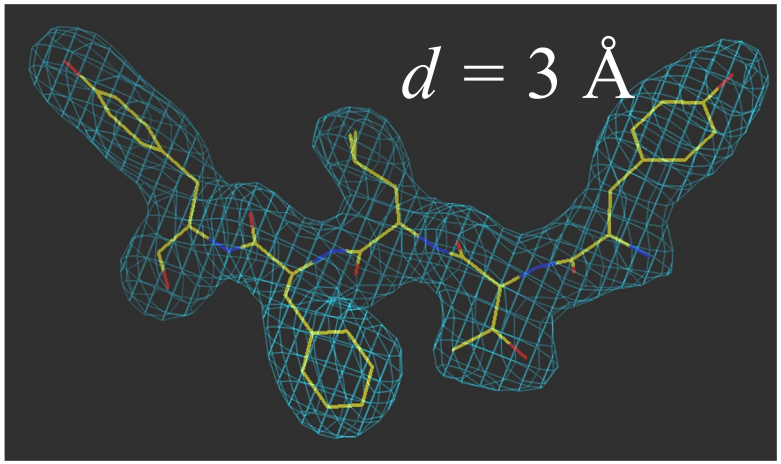
x,y,z coordinates (Å)

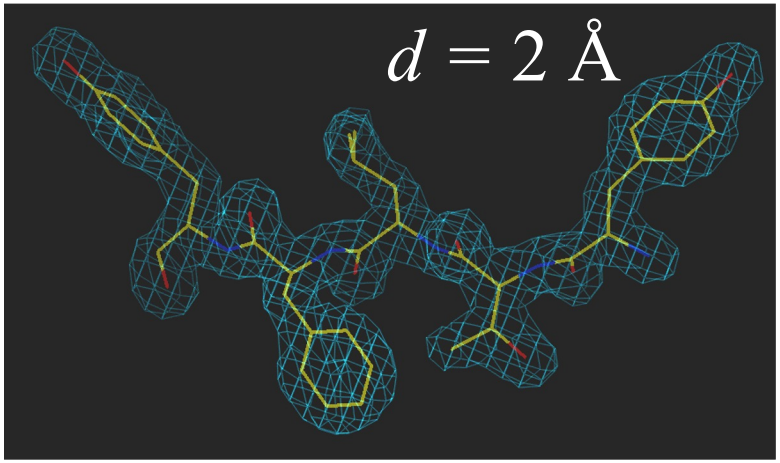
```
ATOM     25  N   ASP A 928      19.062   9.157  35.067  1.00  4.73           N
ATOM     26  CA  ASP A 928      19.770  10.123  34.232  1.00  4.58           C
ATOM     27  C   ASP A 928      19.075   9.938  32.899  1.00  4.56           C
ATOM     28  O   ASP A 928      19.074   8.824  32.351  1.00  5.39           O
ATOM     29  CB  ASP A 928      21.259   9.776  34.071  1.00  3.13           C
ATOM     30  CG  ASP A 928      22.112  10.245  35.233  1.00  5.52           C
ATOM     31  OD1 ASP A 928      21.693  11.114  36.025  1.00  5.42           O
ATOM     32  OD2 ASP A 928      23.239   9.742  35.349  1.00  7.93           O
ATOM     33  N   VAL A 929      18.417  10.985  32.405  1.00  3.68           N
ATOM     34  CA  VAL A 929      17.726  10.864  31.125  1.00  4.63           C
```

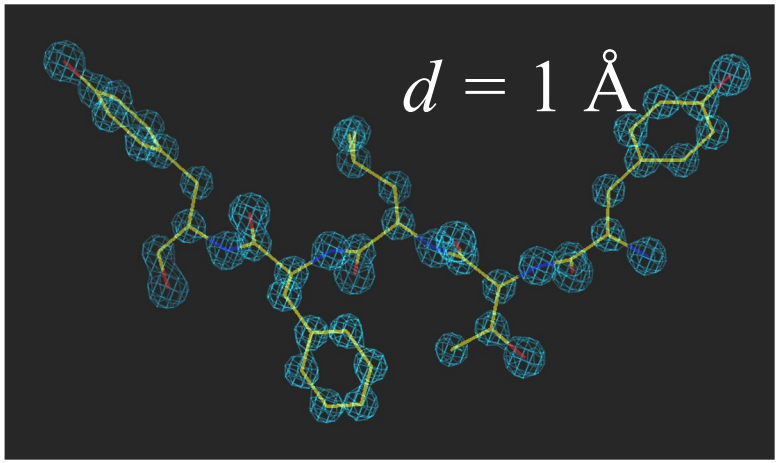
Isotropic B-factor or temperature factor is a measure of the mobility of an atom

The precision of the atomic model is mainly determined by the maximal resolution to which the crystal diffracts X-rays



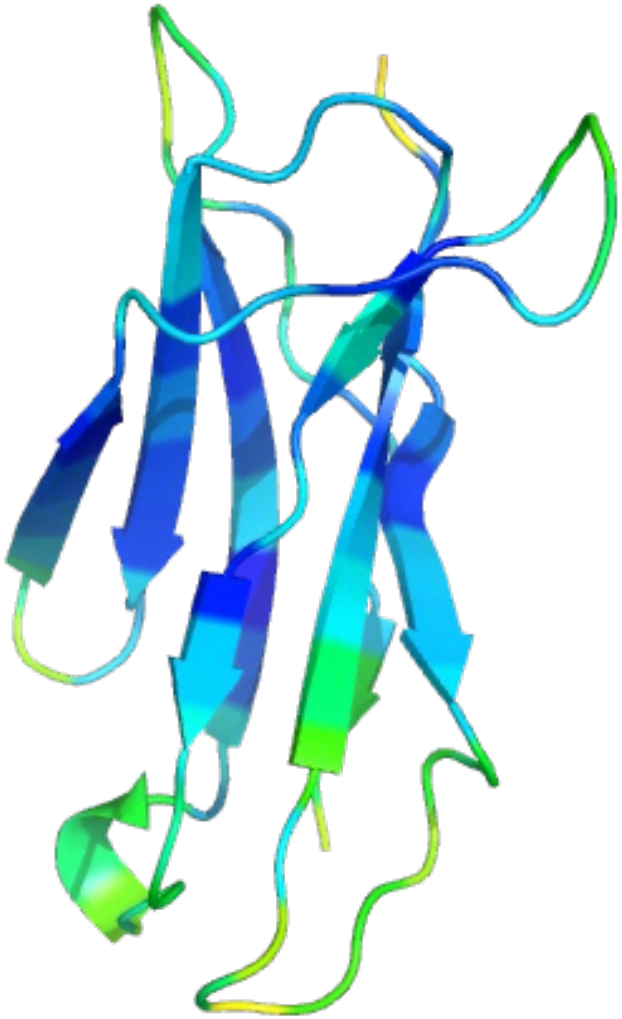






Atomic resolution

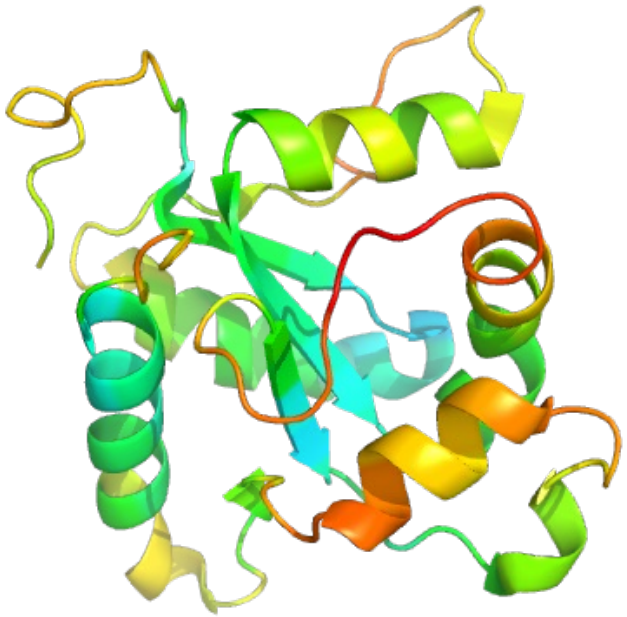
Look at B-factor distribution!



Protein coloured by B-factor:

Well defined regions have low B-factors (blue/green)

Poorly defined/more mobile regions have high B-factors (yellow/orange/red)



A protein molecule is dynamic

- The electron density is a spatial average over all molecules in the crystal and a time average over the duration of the X-ray data measurement

=> Multiple discrete conformations of a residue in different molecules are superimposed.

- A crude description of dynamics is provided in the pdb file as the isotropic B-factor
- Some dynamical aspects evident in the electron density are lost in the pdb file

Starting situation: You would like to do the neutron data refinement of Trypsin with the ligand Aminopyridin

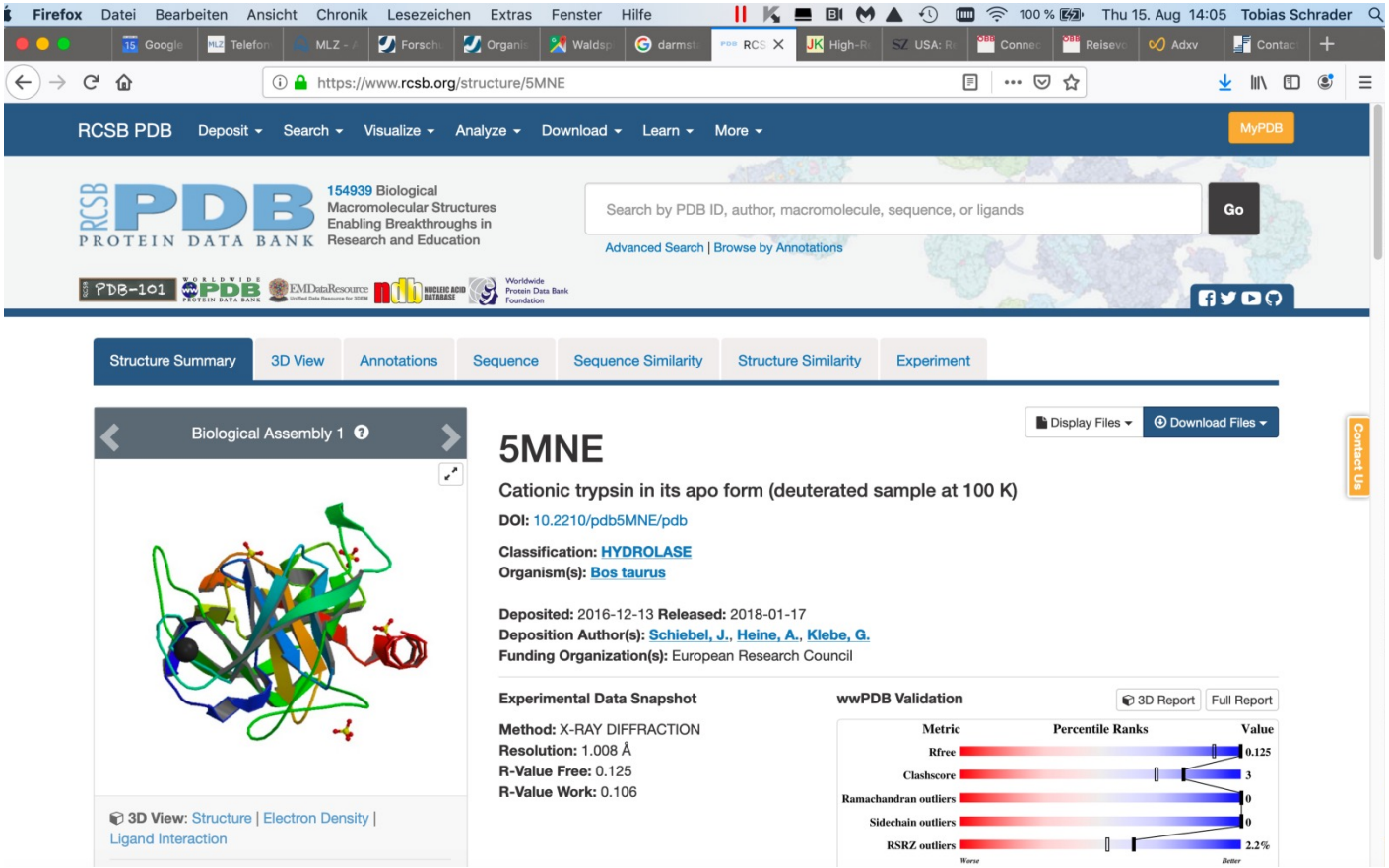
What you have:

A .sca file from the instrument responsible with the data (hkl, intensities, errors of intensities, the unit cell and the space group)

A starting model from the pdb (may be a close by x-ray model)

Starting model without the ligand

Protein data bank: www.rcsb.org



What do you need?

- The Software phenix.refine in its latest version:

https://www.phenix-online.org/download/

- The Software Coot:

https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/

- A two mouse button wheel mouse
- A good editor for text
- Some patience

A tribute to the software programmers…

Citation for the PHENIX suite:

Adams P.D., Afonine P.V., Bunkoczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., & Zwart P.H. (2010). Acta Cryst. D66, 213-221.

Validation:

Williams C.J., Hintze B.J., Headd J.J., Moriarty N.W., Chen V.B., Jain S., Prisant MG Lewis S.M., Videau L.L., Keedy D.A., Deis L.N., Arendall WB I.I.I., Verma V., Snoeyink J.S., Adams P.D., Lovell S.C., Richardson J.S., & Richardson D.C. (2018). Protein Science 27, 293-315.

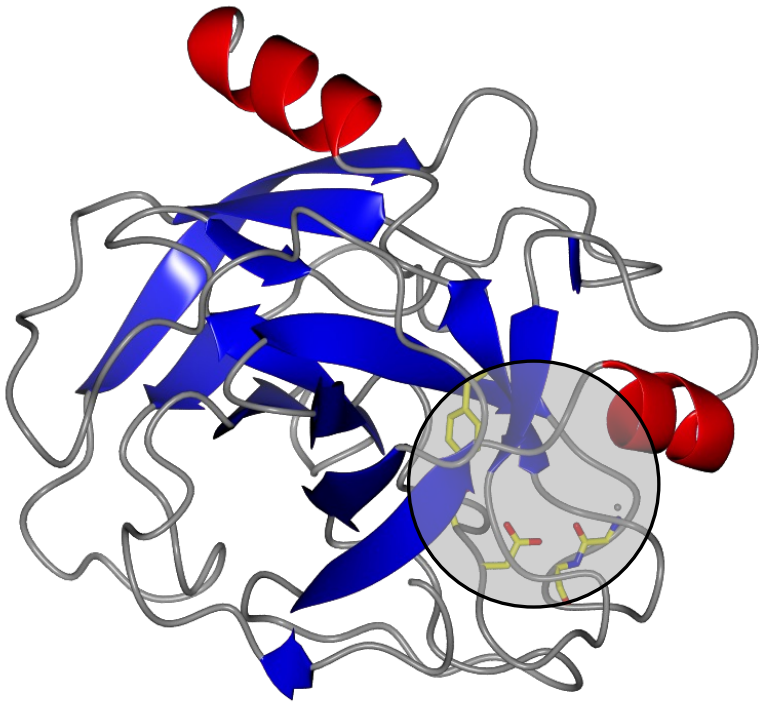
phenix.refine:

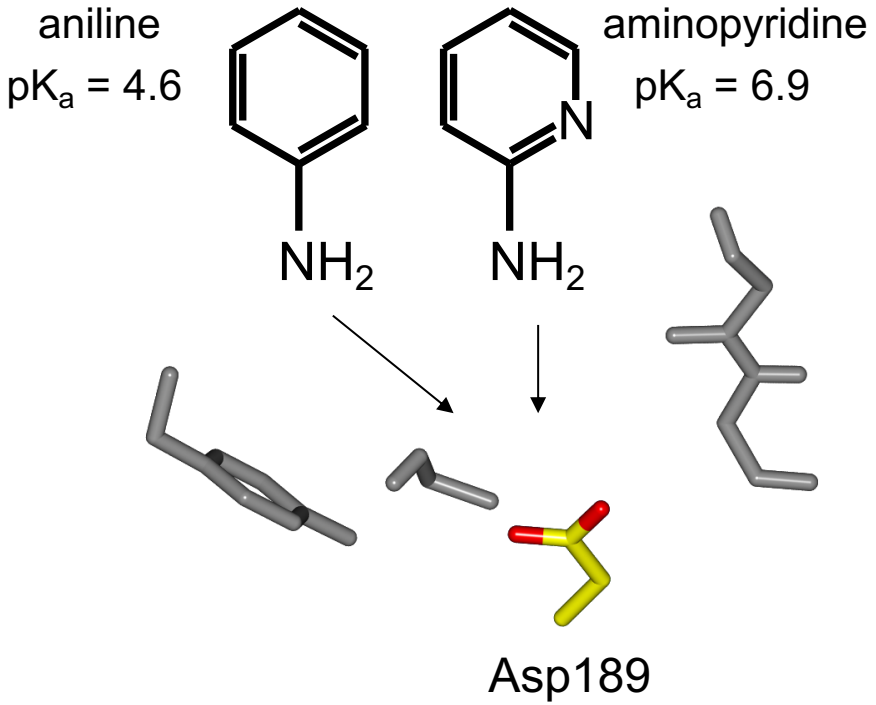
Afonine P.V., Grosse-Kunstleve R.W., Echols N., Headd J.J., Moriarty N.W., Mustyakimov M., Terwilliger T.C., Urzhumtsev A., Zwart P.H., & Adams P.D. (2012). Acta Crystallogr D Biol Crystallogr 68, 352-67.

My advice

1. Always look at the maps and difference maps produced by each refinement step
2. Disregard refinement steps when r- factors increase or the difference between r-work and r-free gets too large (more than 2.5%)
3. Always look at the output pdb model with an editor of your own choice
4. Use the pdb checking tools by phenix and later from the web-site www.rcsb.org

Tutorial: Inhibitor binding to trypsin: charges shift protonation





➡ question: inhibitors with less basic properties become protonated upon binding ?

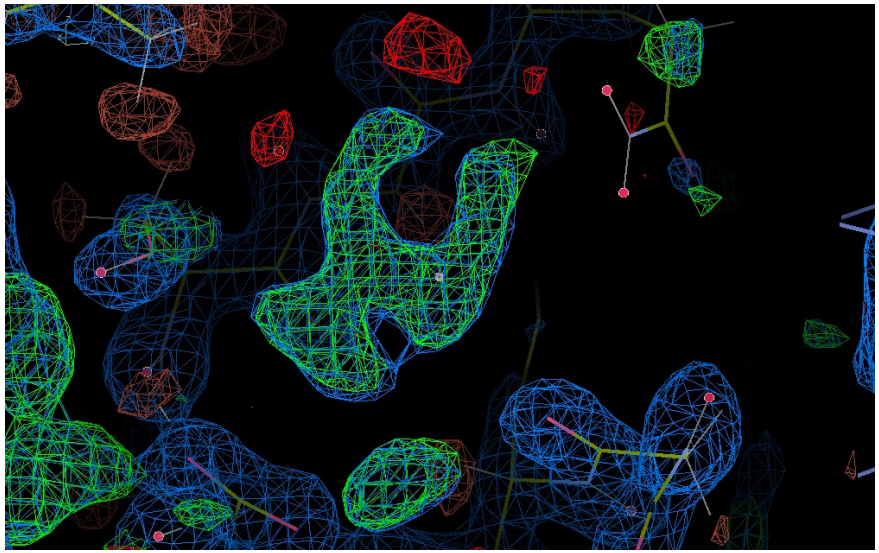
Group of Prof. G. Klebe (Univ. Marburg)

Schiebel J. et al., Angewandte Chemie Int. Edition (2017), 56 (17):4887-4890

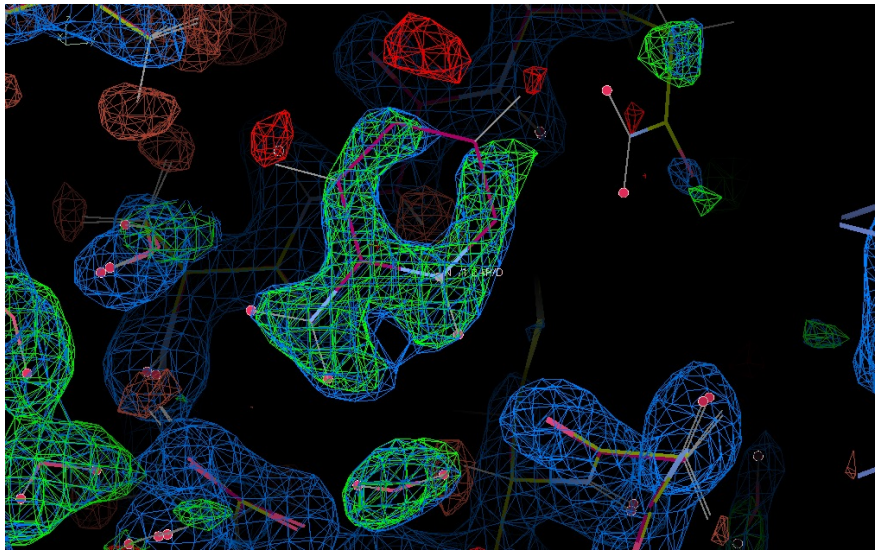
Final results

Refinement is never finished!

Result without ligand aminopyridin



Result with the published structure in



Proteins or structured macromolecules come in different shapes and sizes