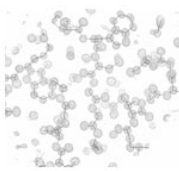
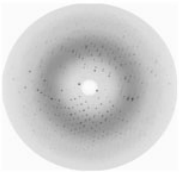
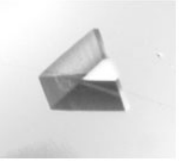
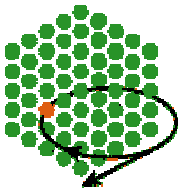


# Crystal Structure Determination of Biological Macromolecules – An Overview

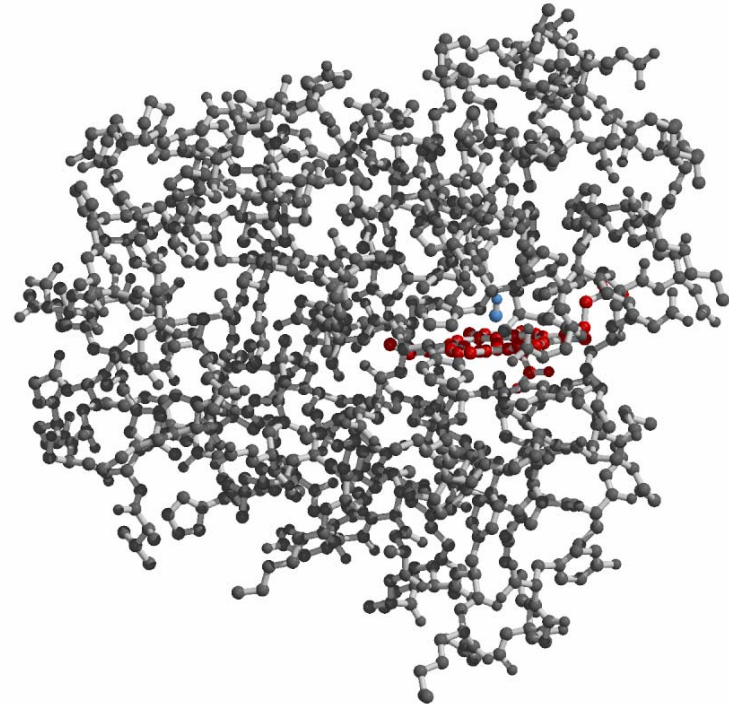
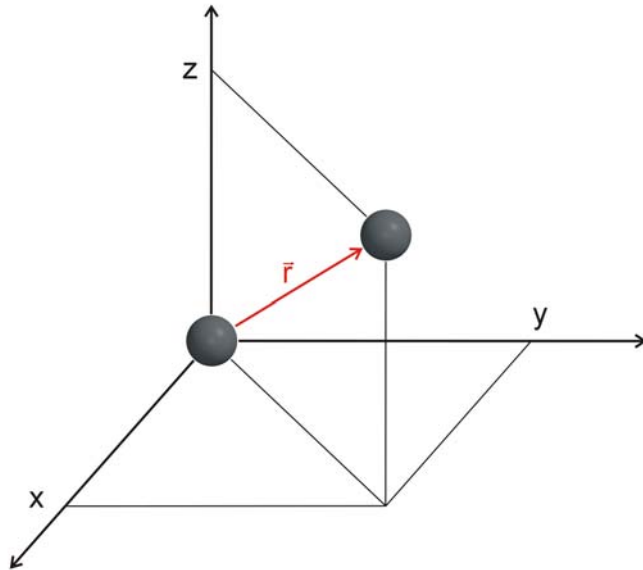
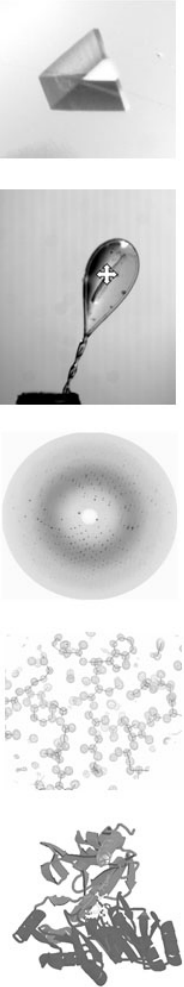


**Manfred S. Weiss**

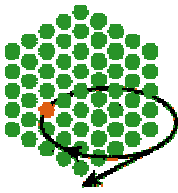
*EMBL Hamburg Outstation  
c/o DESY  
Notkestr. 85  
D-22603 Hamburg  
[msweiss@embl-hamburg.de](mailto:msweiss@embl-hamburg.de)*



# What Do We Mean By Structure?

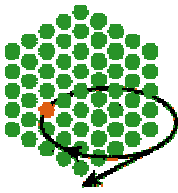


Myoglobin/CO, PDB entry 1A6G

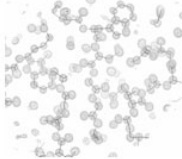
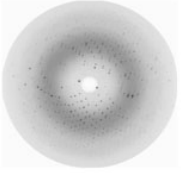
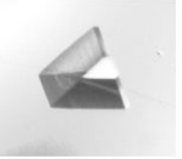


# A Coordinate File

```
HEADER          HEME PROTEIN                      25-FEB-98    1A6G
TITLE          CARBONMONOXY-MYOGLOBIN, ATOMIC RESOLUTION
EXPDTA        X-RAY DIFFRACTION
...
ATOM          1  N   VAL A   1      -4.040   15.048   13.602   1.00  37.04      N
ATOM          2  CA  VAL A   1      -3.621   15.574   14.908   1.00  25.44      C
ATOM          3  C   VAL A   1      -2.766   14.564   15.637   1.00  24.20      C
ATOM          4  O   VAL A   1      -3.155   13.399   15.760   1.00  27.65      O
ATOM          5  CB  VAL A   1      -4.863   15.869   15.757   1.00  32.48      C
ATOM          6  CG1 VAL A   1      -4.402   15.954   17.205   1.00  38.33      C
ATOM          7  CG2 VAL A   1      -5.600   17.125   15.353   1.00  45.60      C
ATOM          8  N   LEU A   2      -1.588   14.922   16.125   1.00  17.99      N
ATOM          9  CA  LEU A   2      -0.767   13.985   16.878   1.00  16.85      C
ATOM         10  C   LEU A   2      -1.326   13.777   18.289   1.00  16.94      C
ATOM         11  O   LEU A   2      -1.795   14.765   18.873   1.00  19.44      O
ATOM         12  CB  LEU A   2         0.694   14.436   17.007   1.00  15.72      C
ATOM         13  CG  LEU A   2         1.598   14.164   15.773   1.00  14.99      C
ATOM         14  CD1 LEU A   2         1.269   15.130   14.656   1.00  15.12      C
ATOM         15  CD2 LEU A   2         3.080   14.254   16.161   1.00  14.37      C
ATOM         16  N   SER A   3      -1.218   12.546   18.763   1.00  16.86      N
...
...
END
```

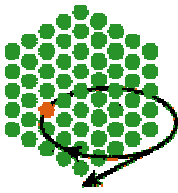


# The Protein Data Bank (PDB)

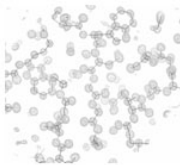
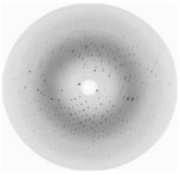
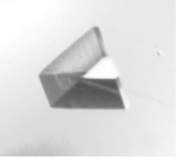


- PDB: publicly available archive of macromolecular structure data
- Link: <http://www.rcsb.org/pdb>
- Reference: H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov & P. E. Bourne (2000). The Protein Data Bank. *Nucl. Ac. Res.* **28**, 235-242.

**PDB**  
PROTEIN DATA BANK

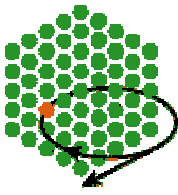


# The PDB History

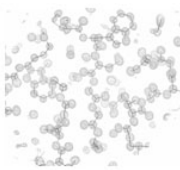
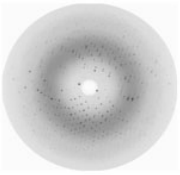
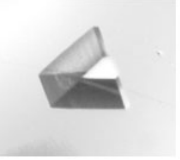


- Established at Brookhaven National Laboratory (BNL) in 1971 as one of the earliest community-driven molecular-biology data collections.
- Since October 1998 managed by the Research Collaboratory for Structural Bioinformatics (RCSB).

**P D B**  
PROTEIN DATA BANK

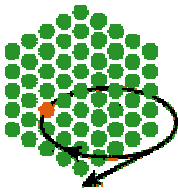


# PDB Content



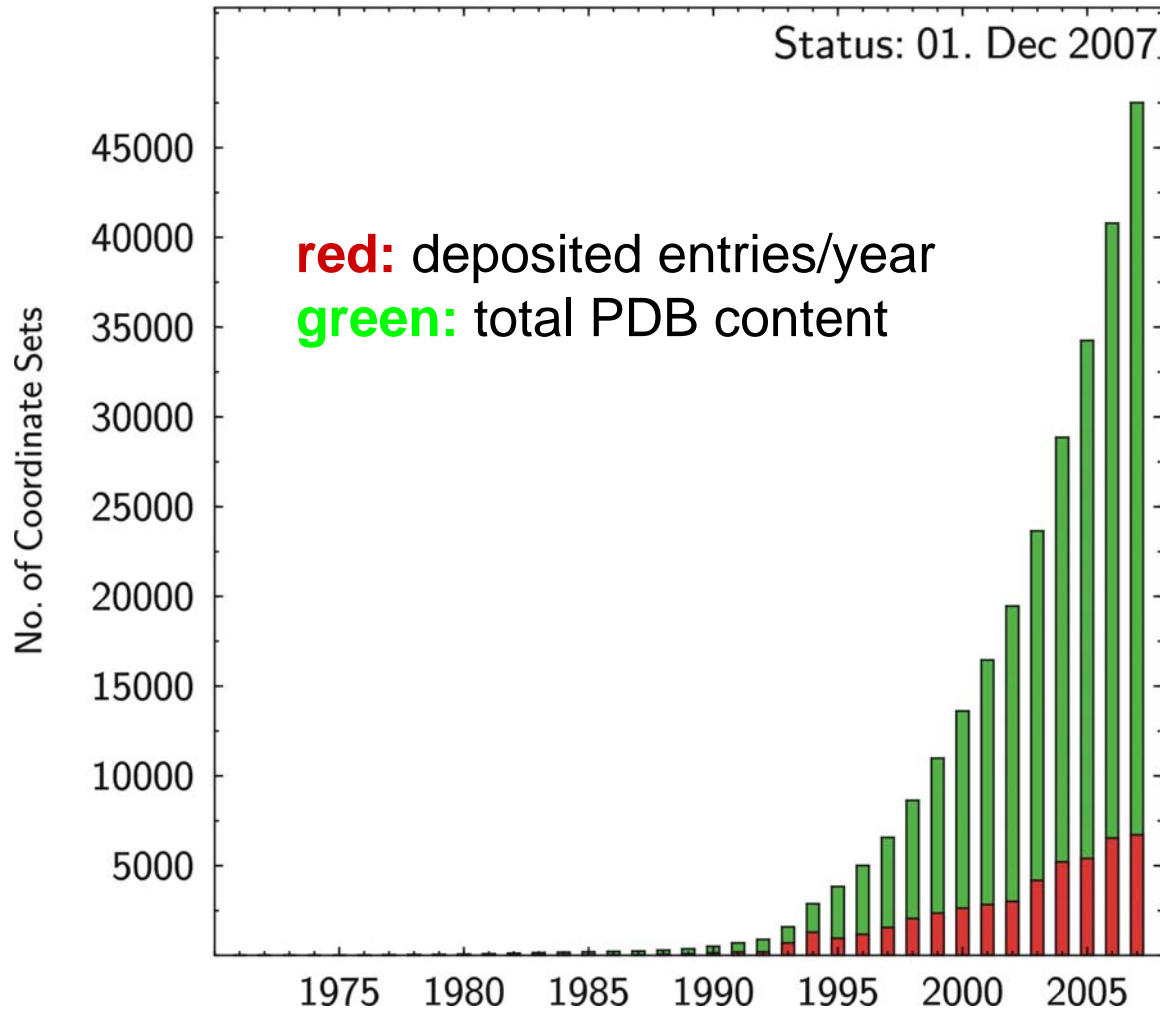
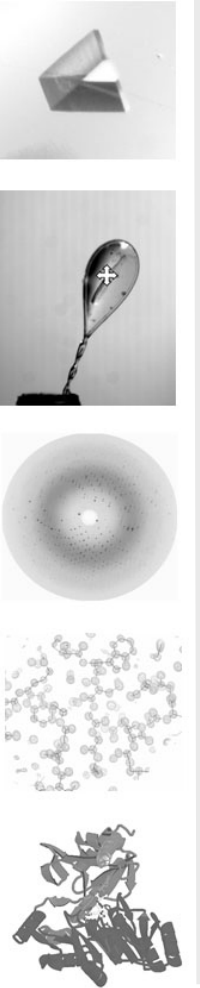
- Total contents of the PDB : 47,509 structures (as of Nov 27, 2007)
- 2006: 6,537 new structures (18 per day)

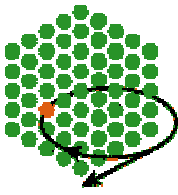
**PDB**  
PROTEIN DATA BANK



# PDB Growth Statistics

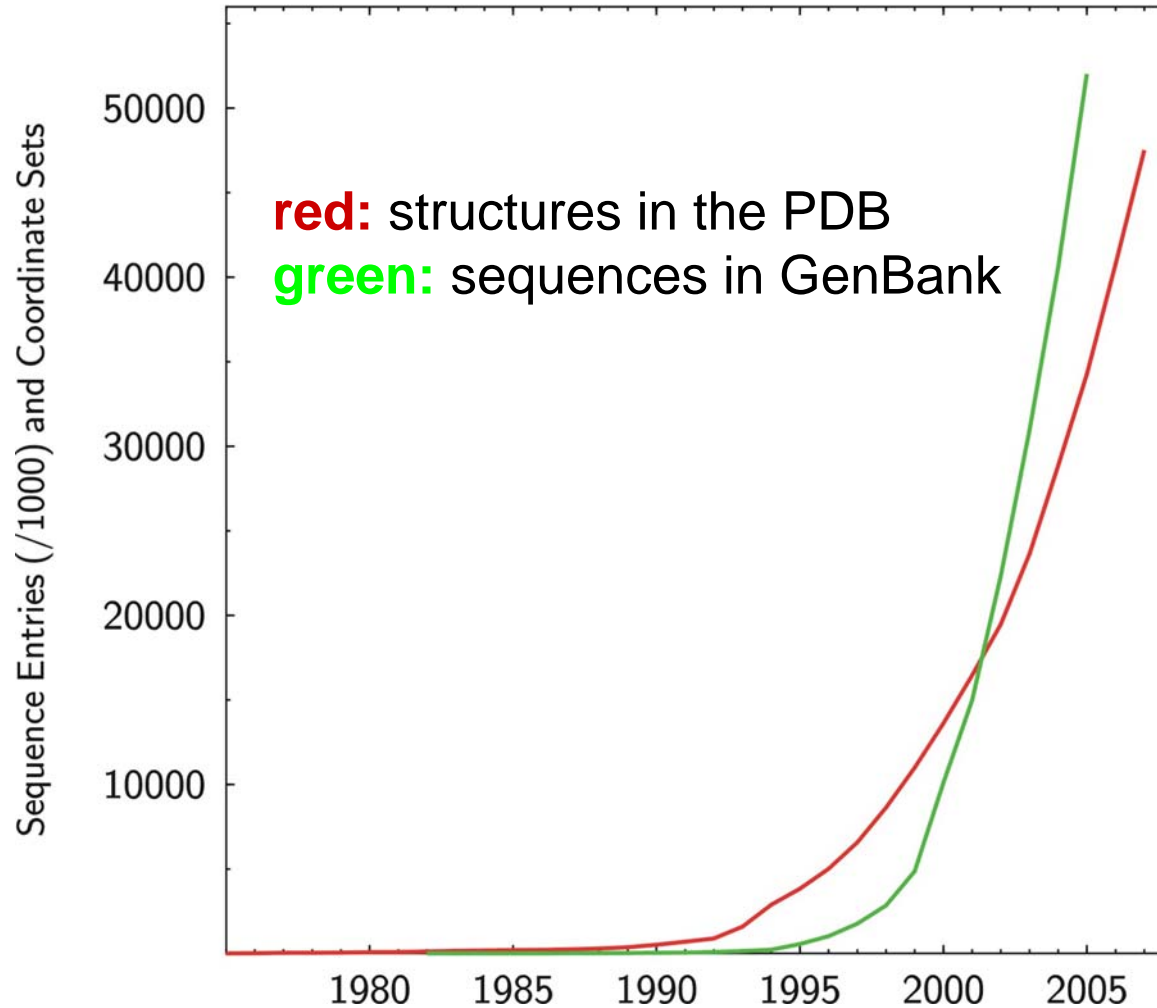
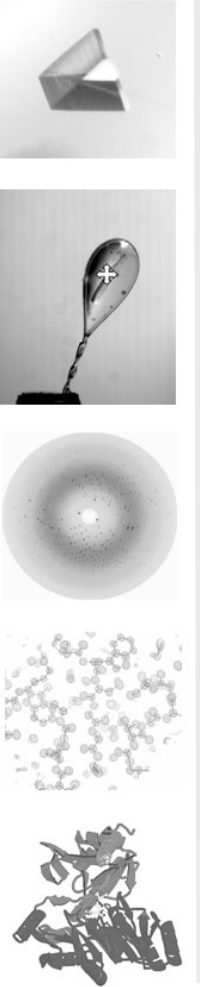
Protein Data Bank



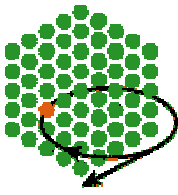


# Structures vs Sequences

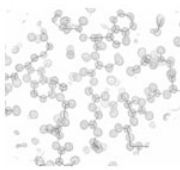
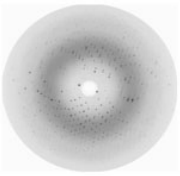
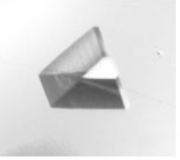
PDB vs GenBank



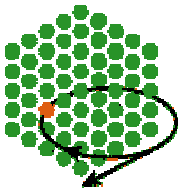




# Why Do We Want To Determine Structures?

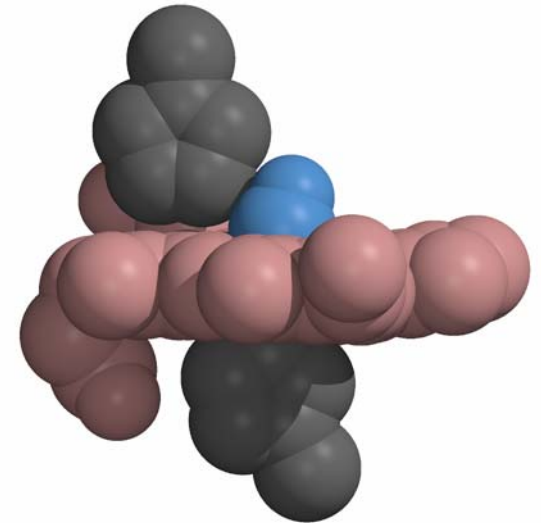
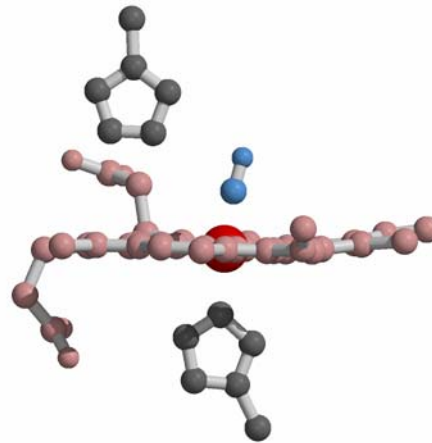
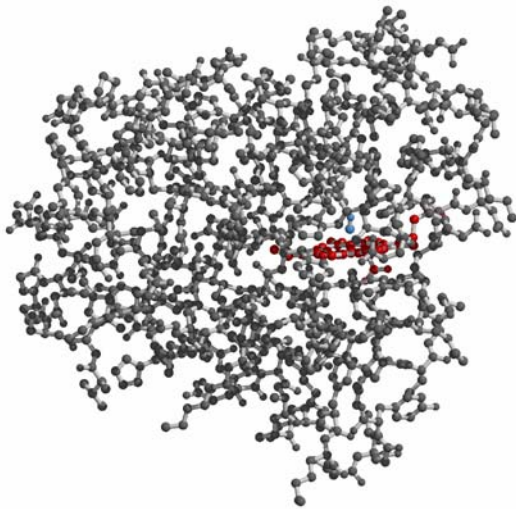


The underlying principle  
of **function** is  
**structure**.



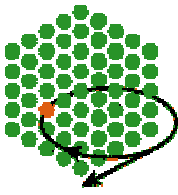
# One Example: Myoglobin

Myoglobin (oxygen carrier protein in muscles) can bind both  $O_2$  and CO at the iron ( $Fe^{2+}$ ) of its heme group.

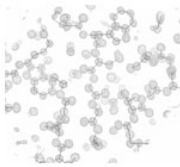
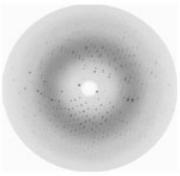
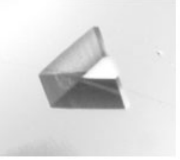


Myoglobin/CO, PDB entry 1A6G

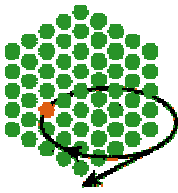
However, CO likes to bind in a linear fashion, whereas  $O_2$  prefers to bind at an angle.



# Structure Determination Methods

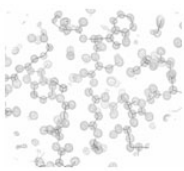
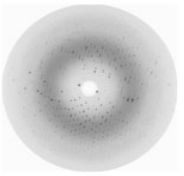
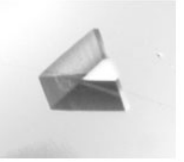


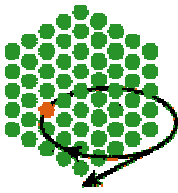
1. X-ray Diffraction
2. Nuclear Magnetic Resonance (NMR)
3. Electron Diffraction
4. Electron Microscopy
5. Neutron Diffraction
6. Molecular Modeling



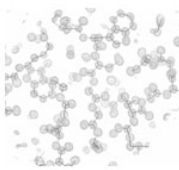
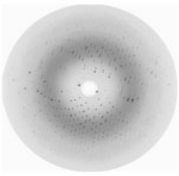
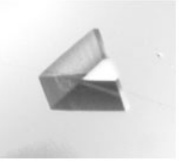
# X-ray Diffraction is ...

- . . . the most important of the techniques listed (85% of all protein structures in the PDB and 95% of all structures consisting of more than 80 amino acids have been determined by **X-ray Diffraction**).

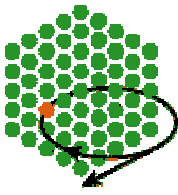




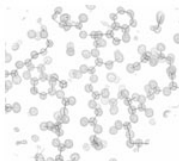
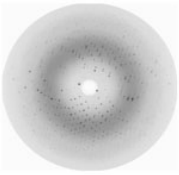
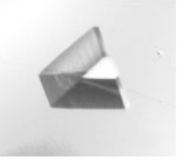
# X-ray Diffraction is ...



- . . . the most important of the techniques listed (85% of all protein structures in the PDB and 95% of all structures consisting of more than 80 amino acids have been determined by **X-ray Diffraction**).
- . . . synonymous for **Crystal Structure Analysis**.

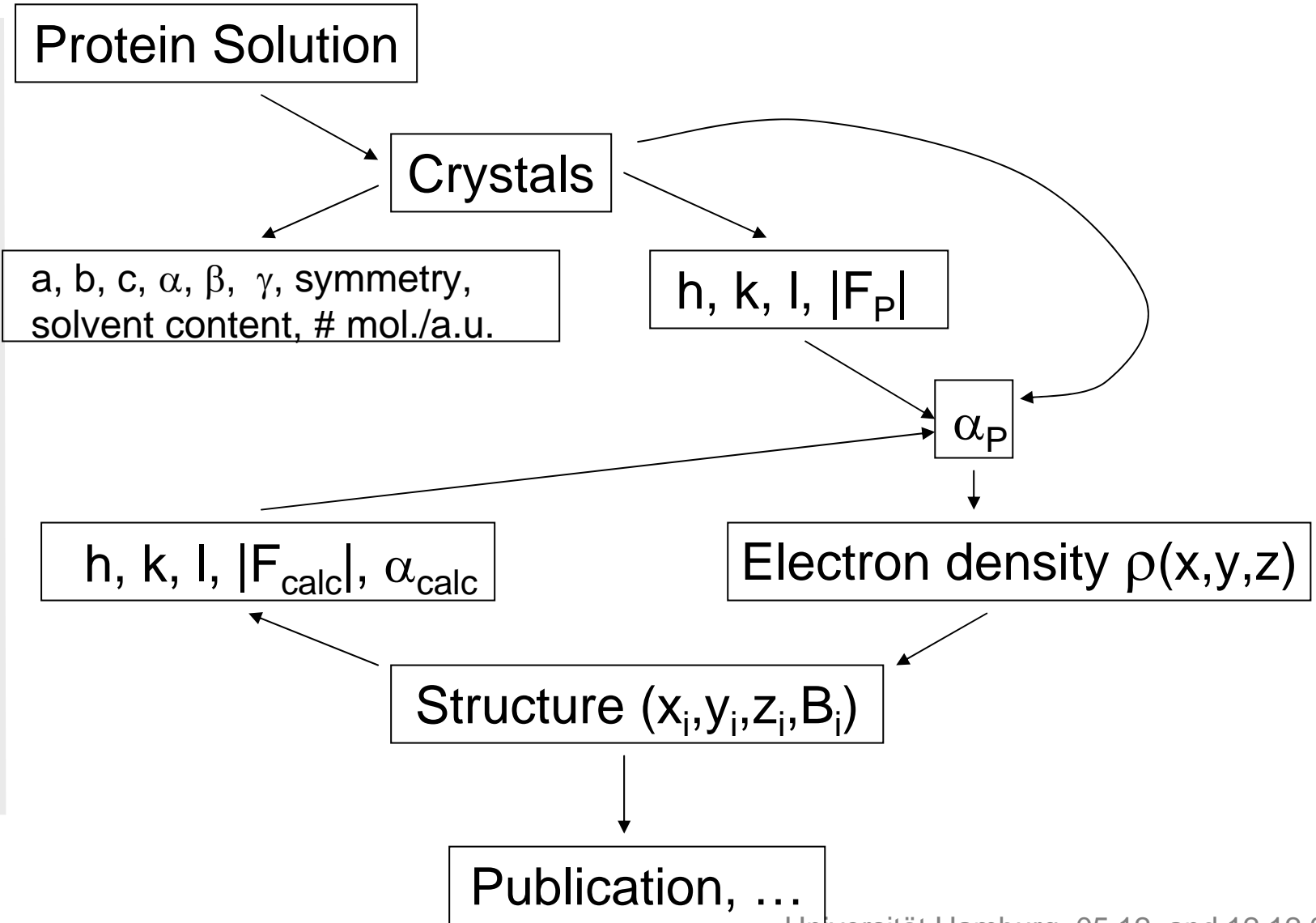
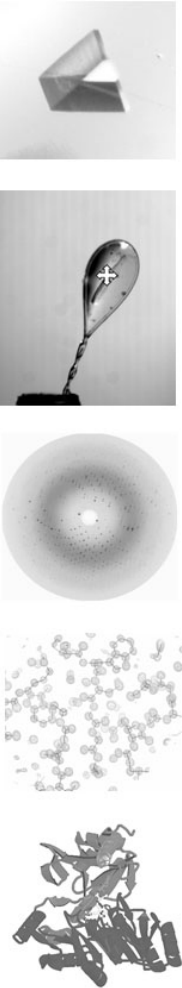
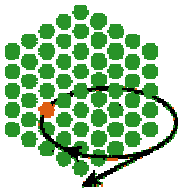


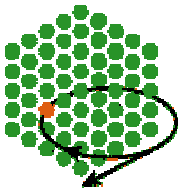
# X-ray Diffraction



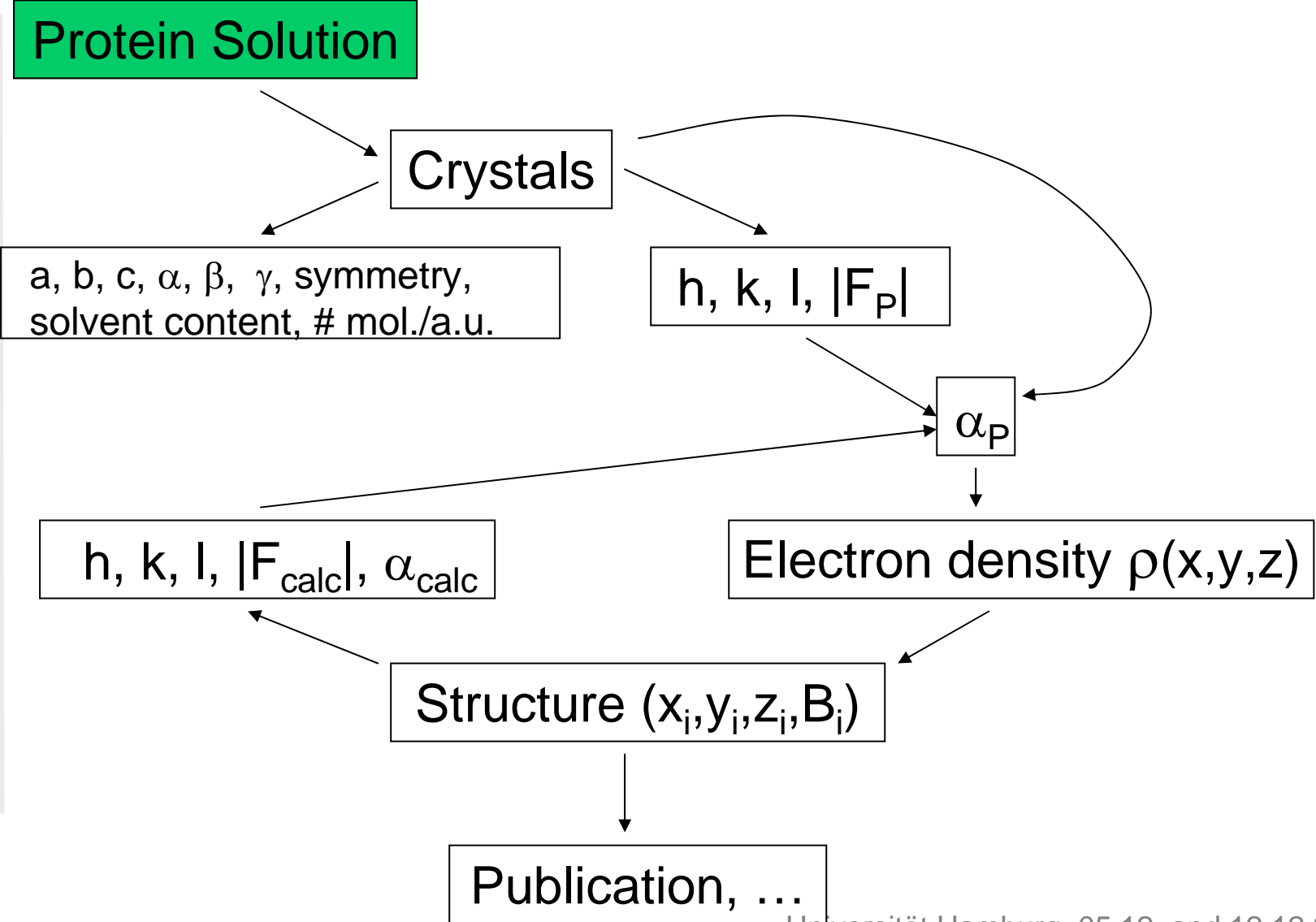
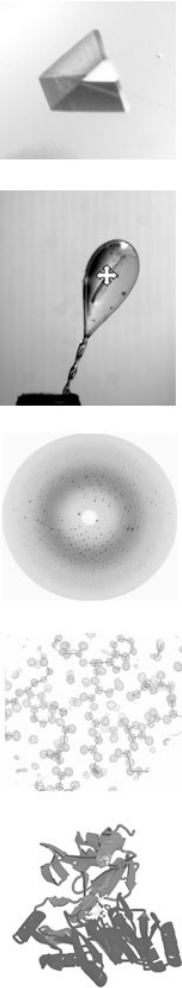
- Physical Principle: scattering of X-rays by **electrons**
- Prerequisite: three-dimensional crystals
- Resolution: atomic or sub-atomic possible
- Molecular Weights: no limit
- Application: small molecules, proteins, viruses, ribosomes

# A Flowchart of A Crystal Structure Determination



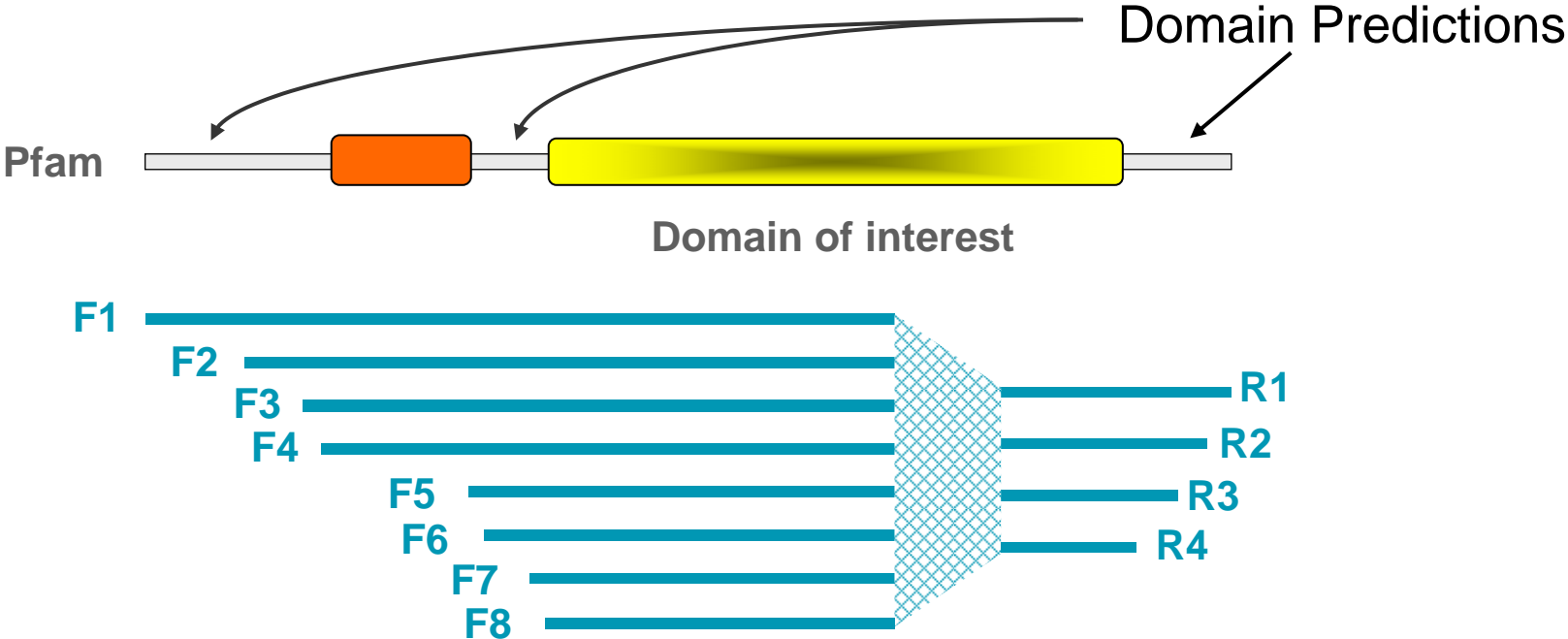
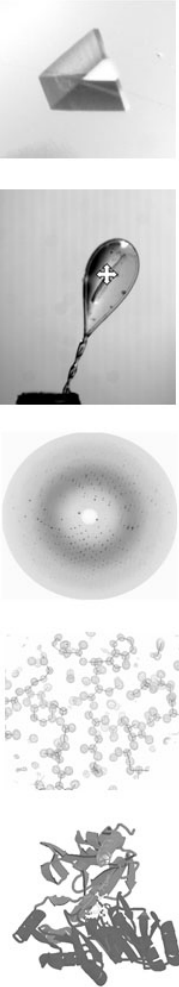
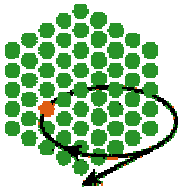


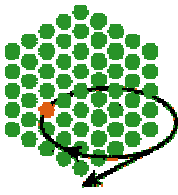
# Before We Start ...



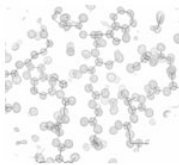
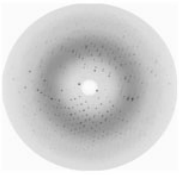
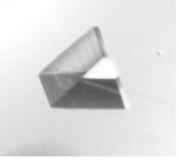


# Picking the Right System

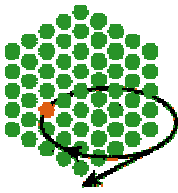




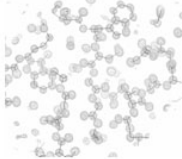
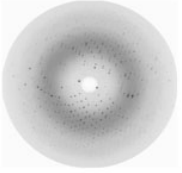
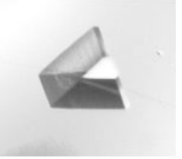
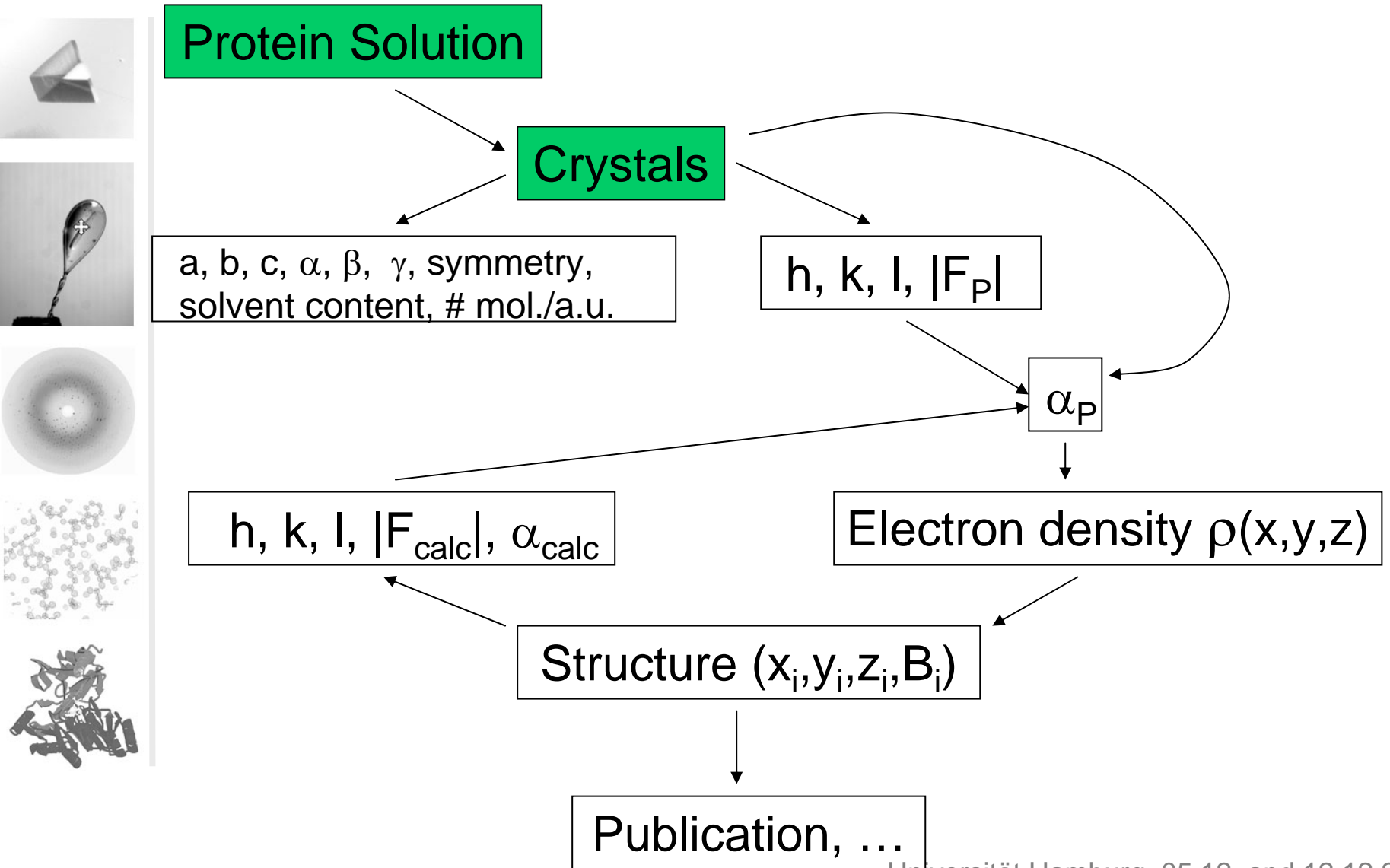
# Production of the Sample

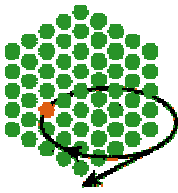


- Amplification of the Gene Using PCR
- Heterologous (Over)expression
- Purification Using Chromatographic Techniques
- Concentration of the Sample
- Biophysical Characterization
- Quality Control



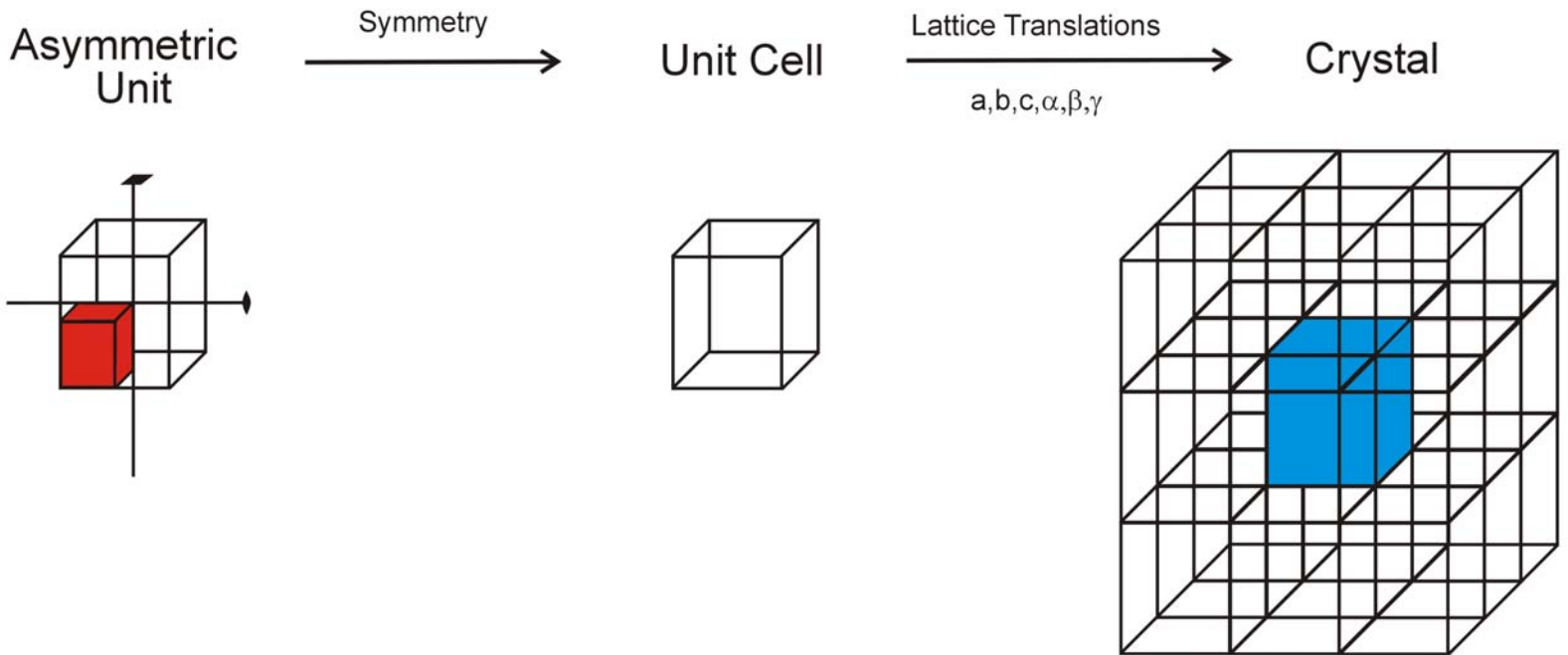
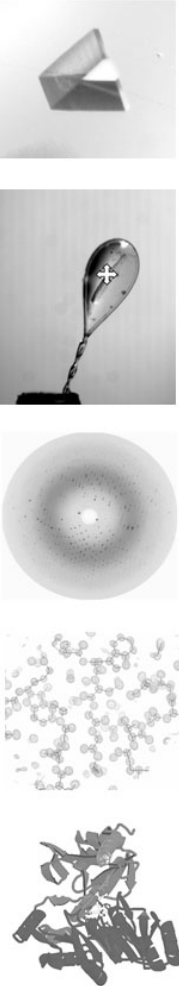
# Crystallization of a Protein

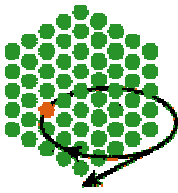




# What Is A Crystal?

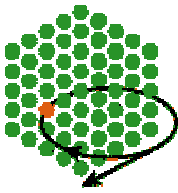
A crystal can be defined as an arrangement of building blocks which is **periodic** in **three dimensions**.





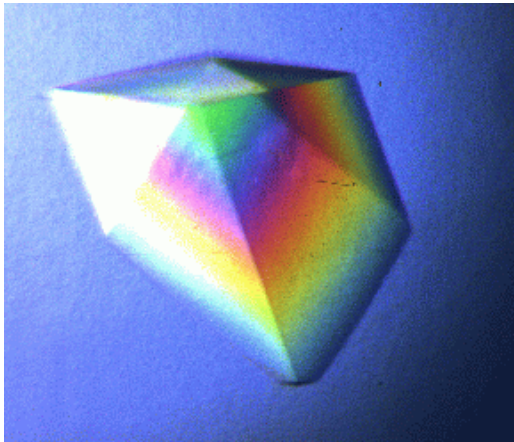
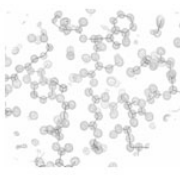
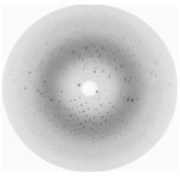
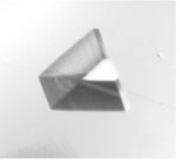
# Proteins Can Be Crystallized



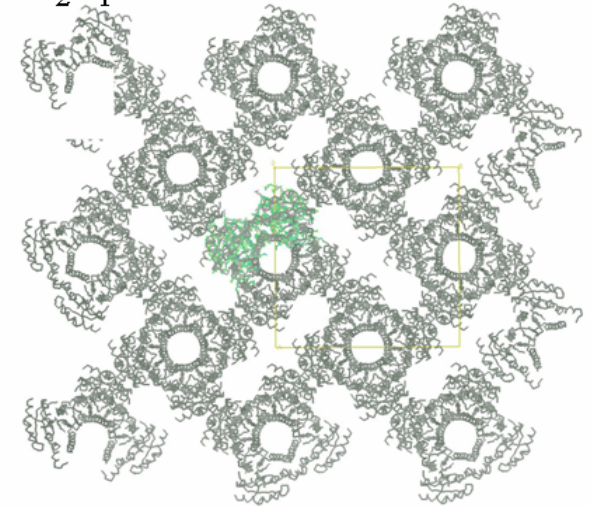


# Protein Crystals Are Different

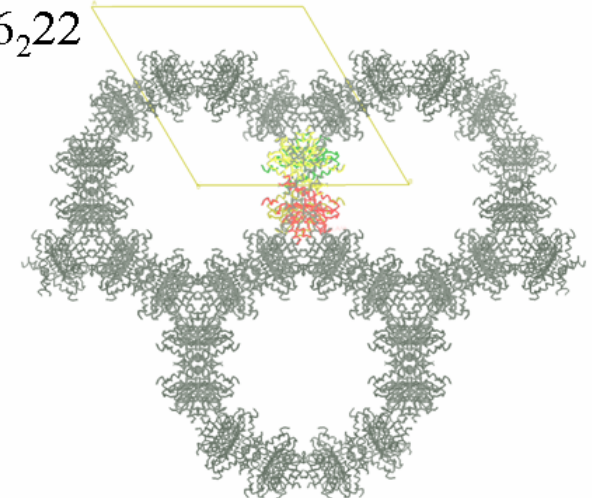
- very soft (needle test !)
- mechanically fragile
- large solvent channels (30-80%)

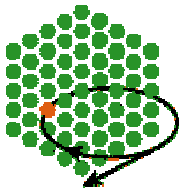


$P4_22_12$



$P6_22$

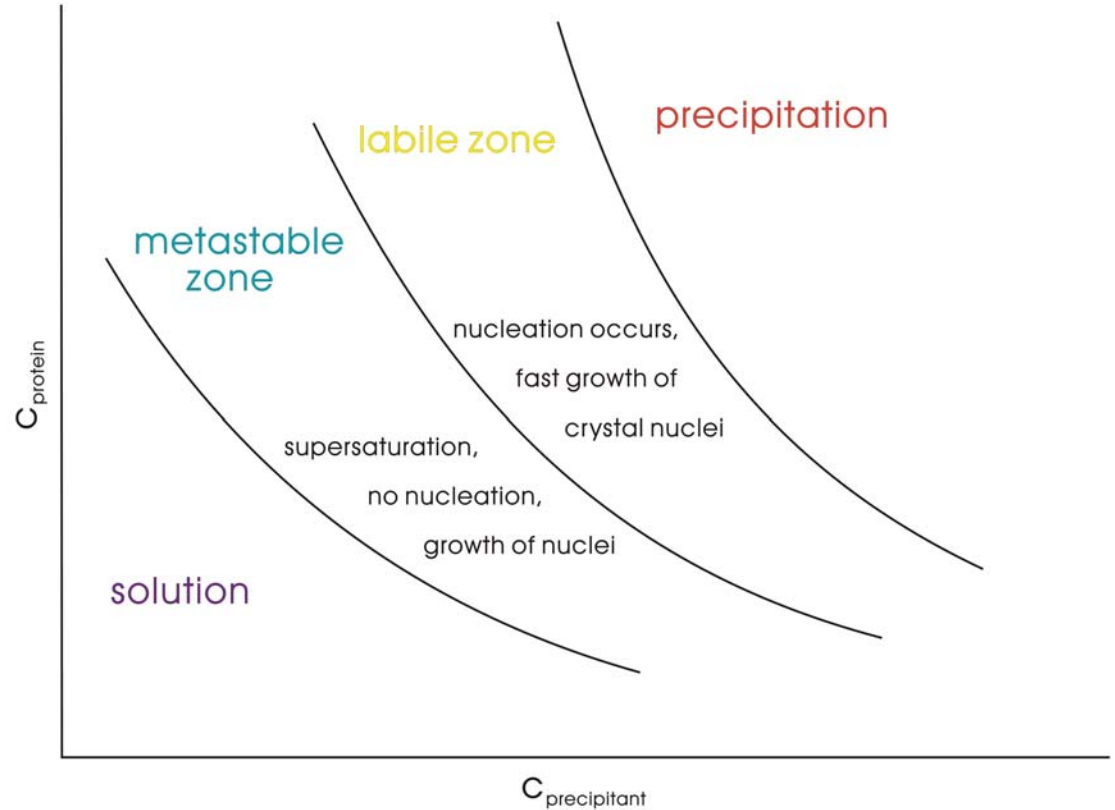
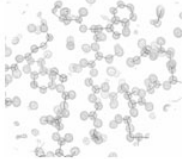
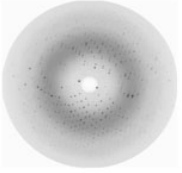
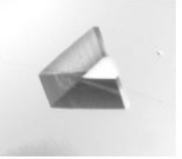




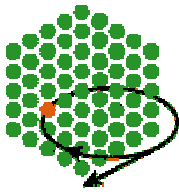
# How Can Proteins Be Crystallized?

## Two Processes

1. Nucleation
2. Growth

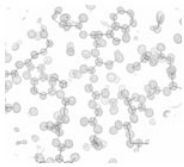
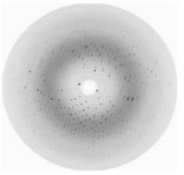
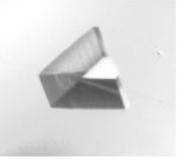
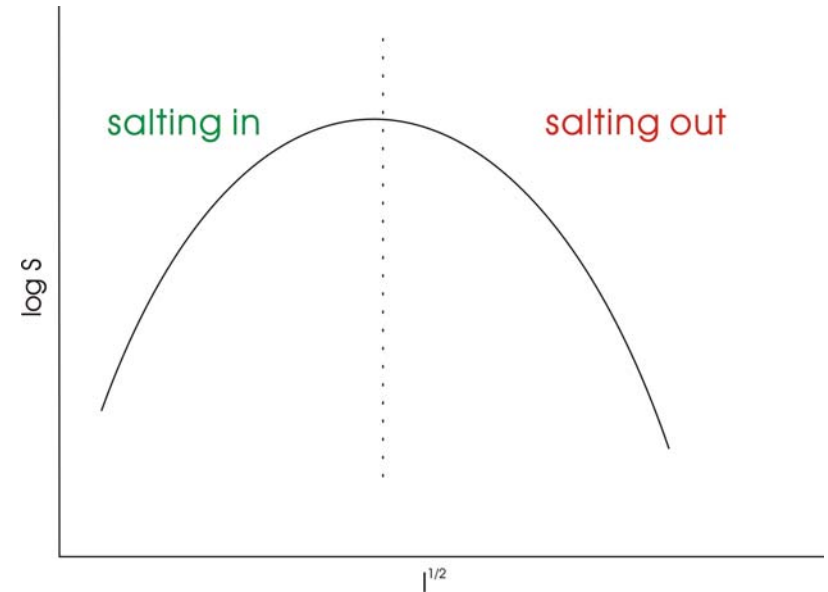


- Start off with the protein in solution ...
- Decrease solubility by some transport process (dialysis, diffusion, ...)

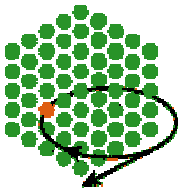


# Factors Influencing Protein Solubility

- Protein Concentration
- Ionic Strength
- pH
- Temperature
- Organic Solvents
- Additives
- ...

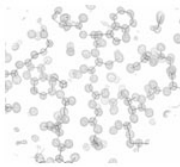
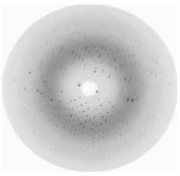
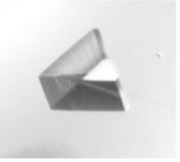




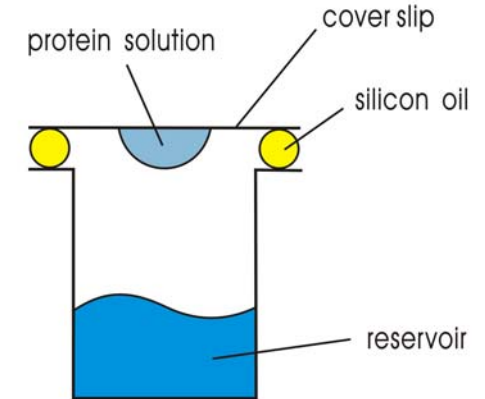


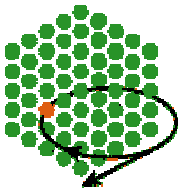
# Crystallization Methods

- Vapor Diffusion  
(hanging, sitting, sandwich-drop)
- Batch
- Dialysis
- Free Interface Diffusion
- Diffusion Across a Gel
- ...

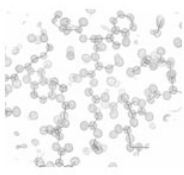
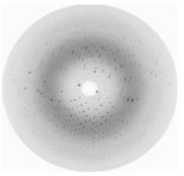
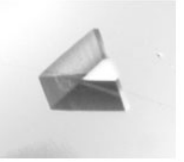


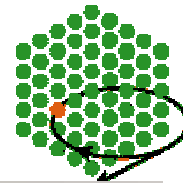
The hanging drop method



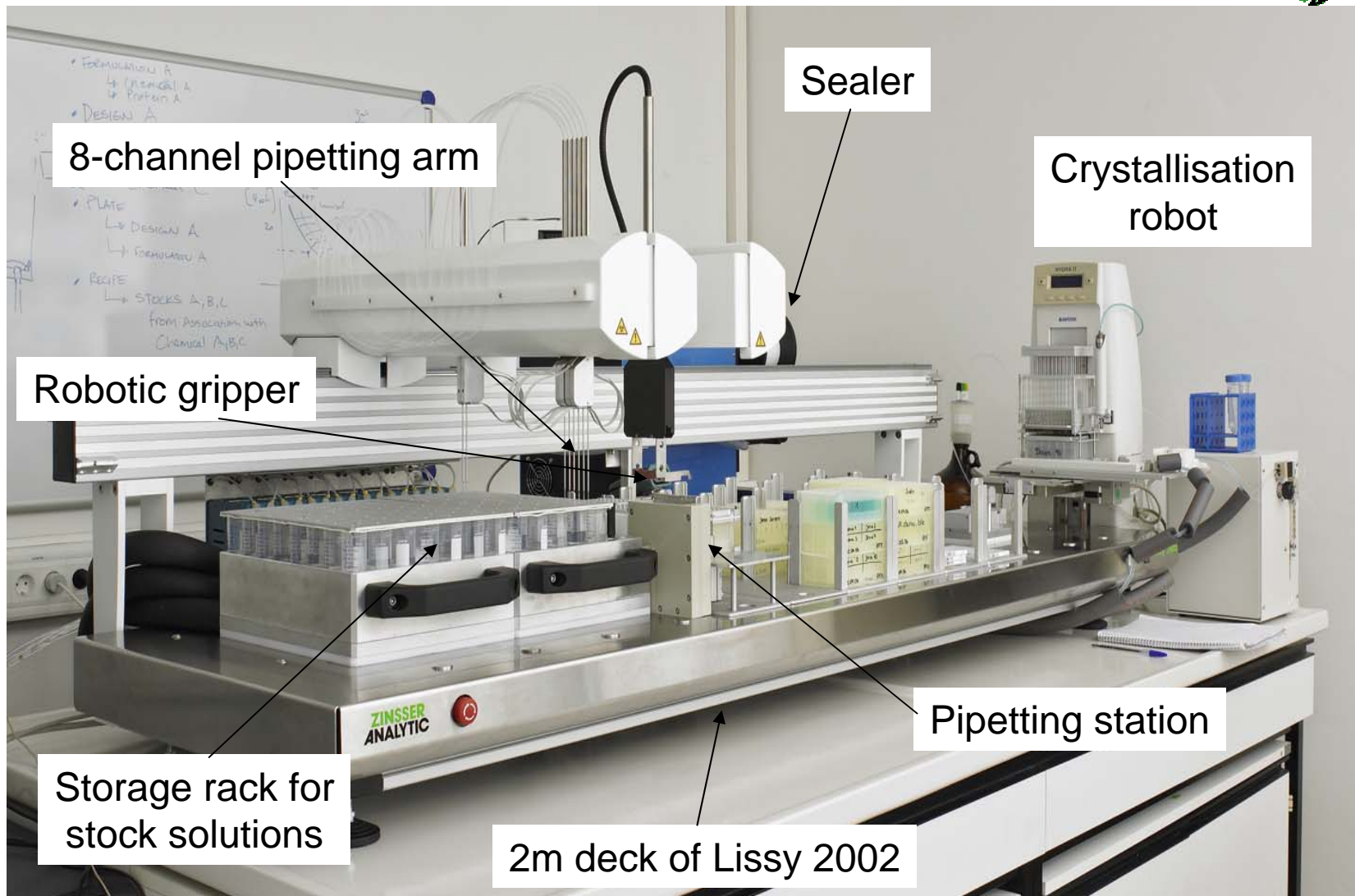
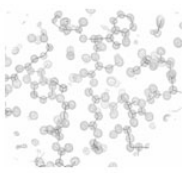
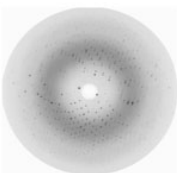


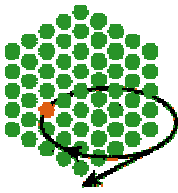
# The “Classical” Way



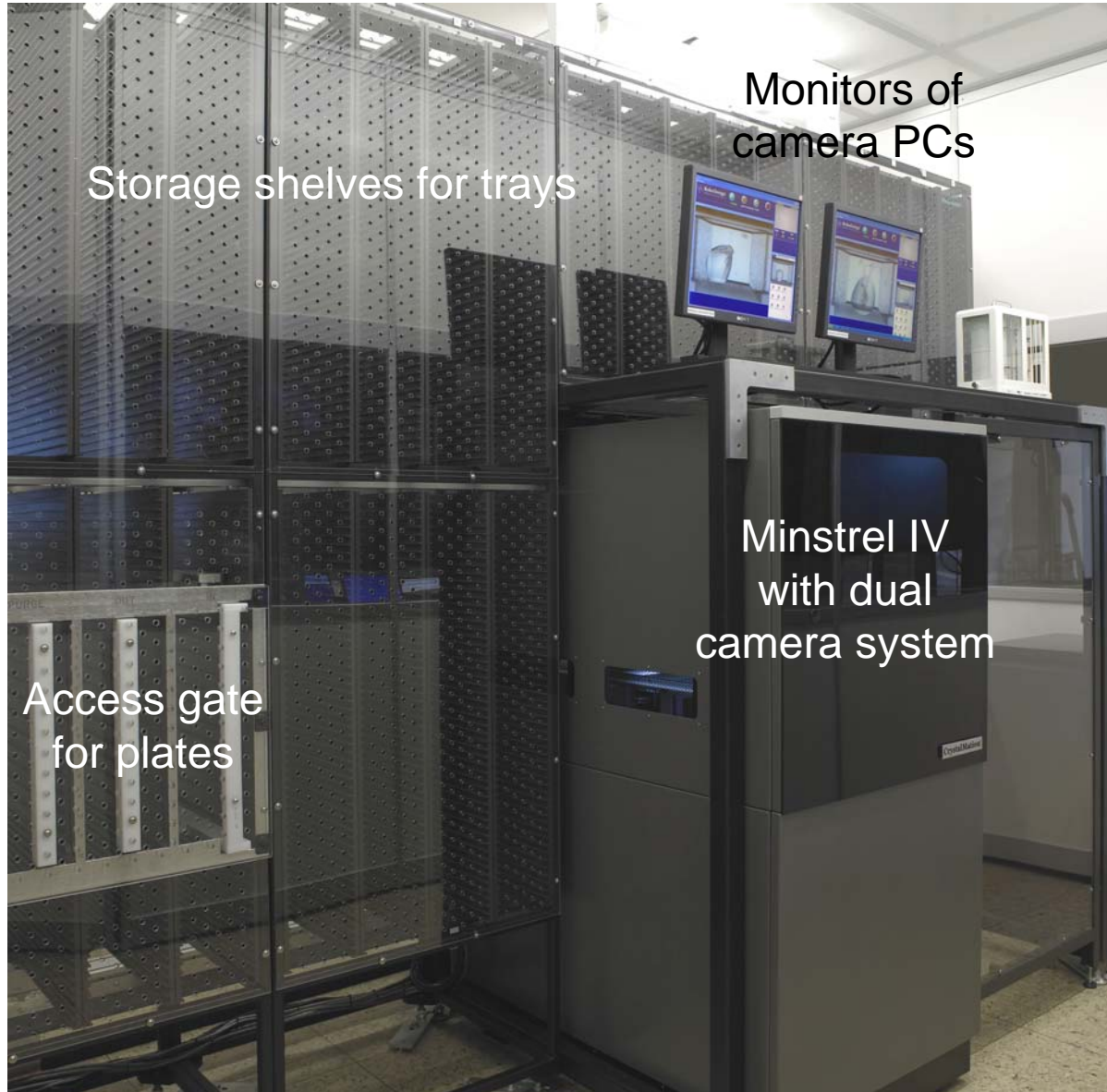
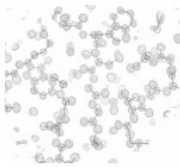
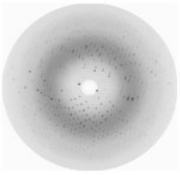
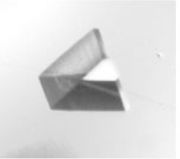


# The “Modern” HTP Way





# The “Modern” HTP Way

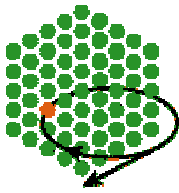


Storage shelves for trays

Monitors of camera PCs

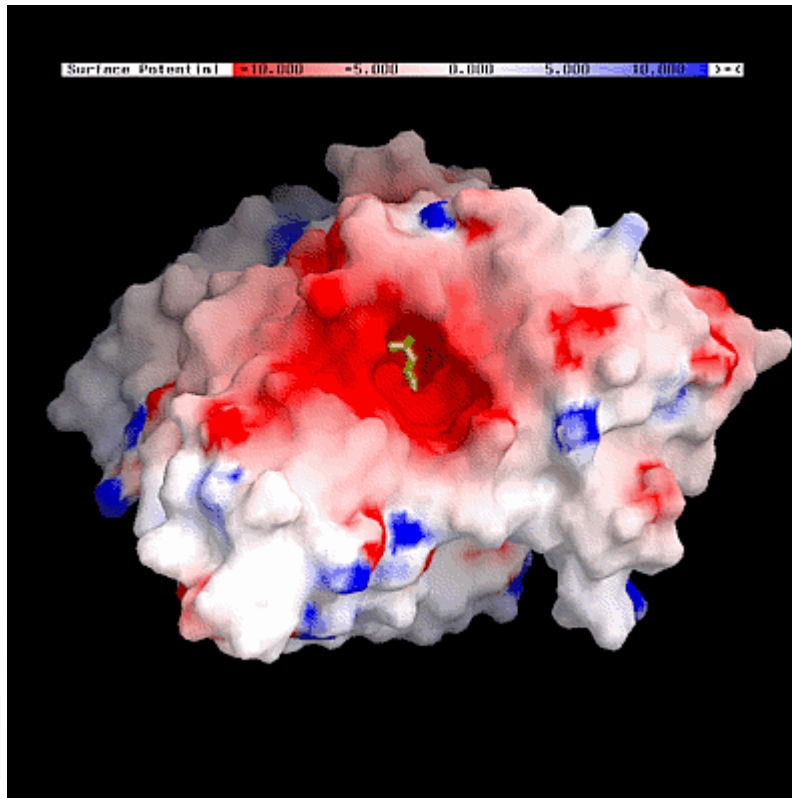
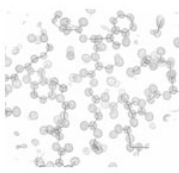
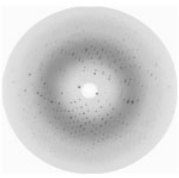
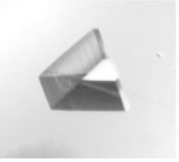
Minstrel IV  
with dual  
camera system

Access gate  
for plates

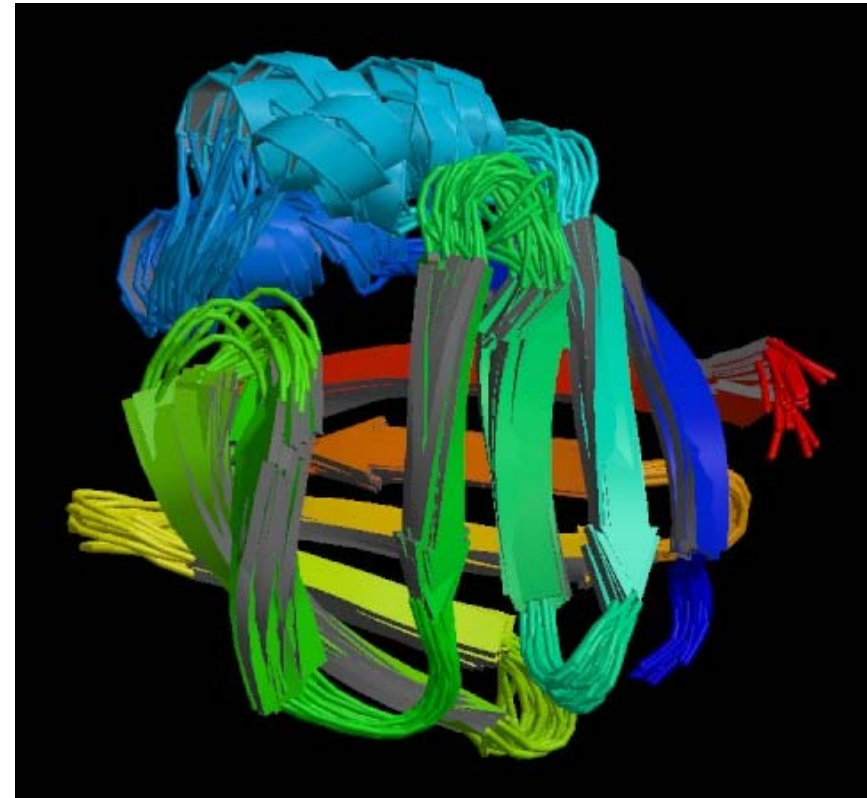


# Difficulties

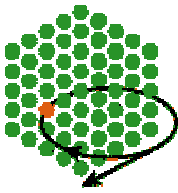
Macromolecular surfaces are irregularly shaped, carry multiple charges and are flexible.



Acetylcholine Esterase © B. Honig, Columbia University

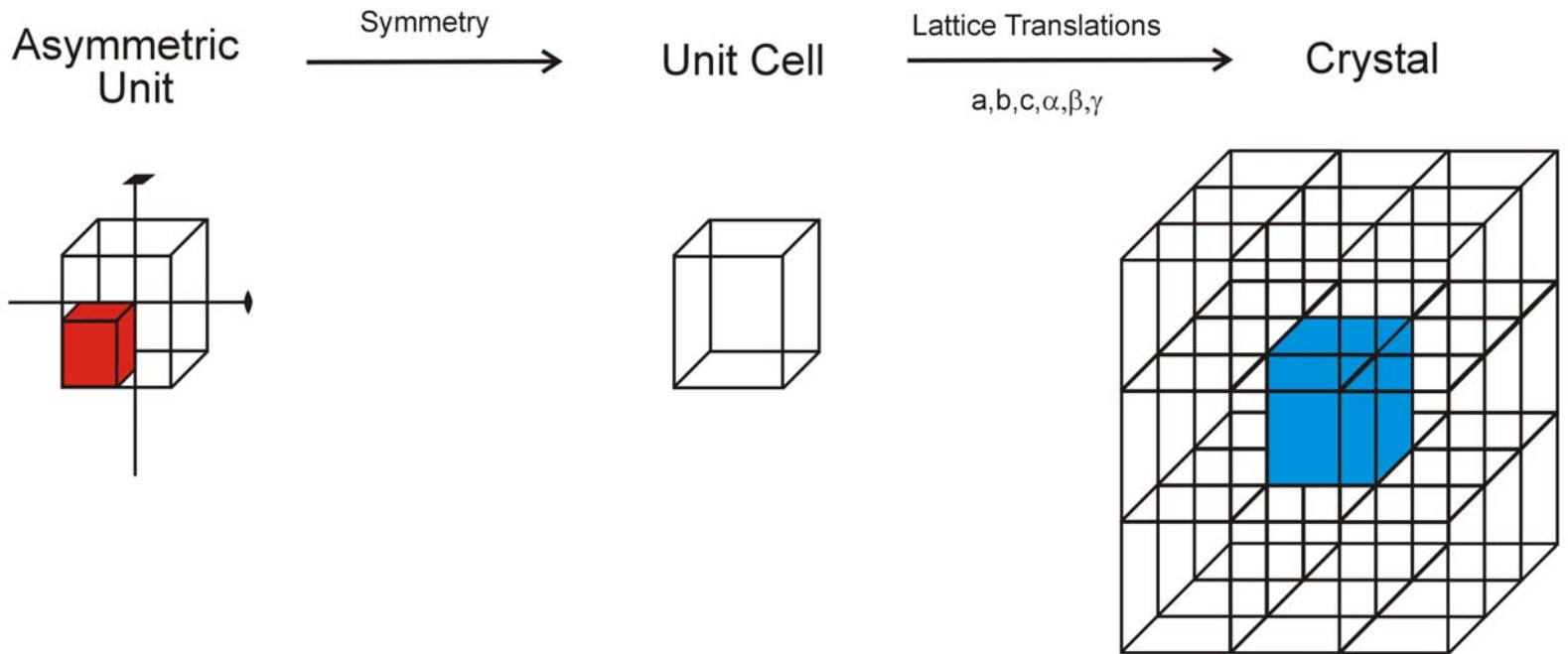
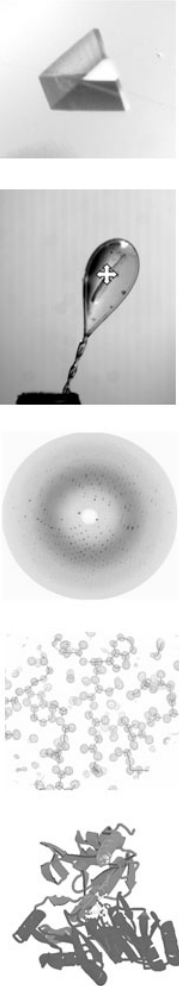


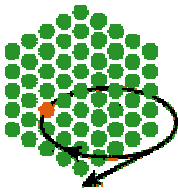
1BWY, Lassen *et al.* (1995) *Eur. J. Biochem.* **230**, 266-280.



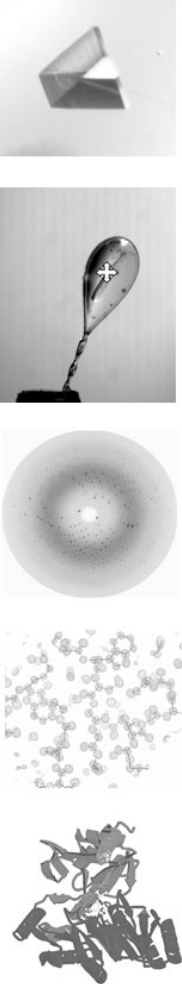
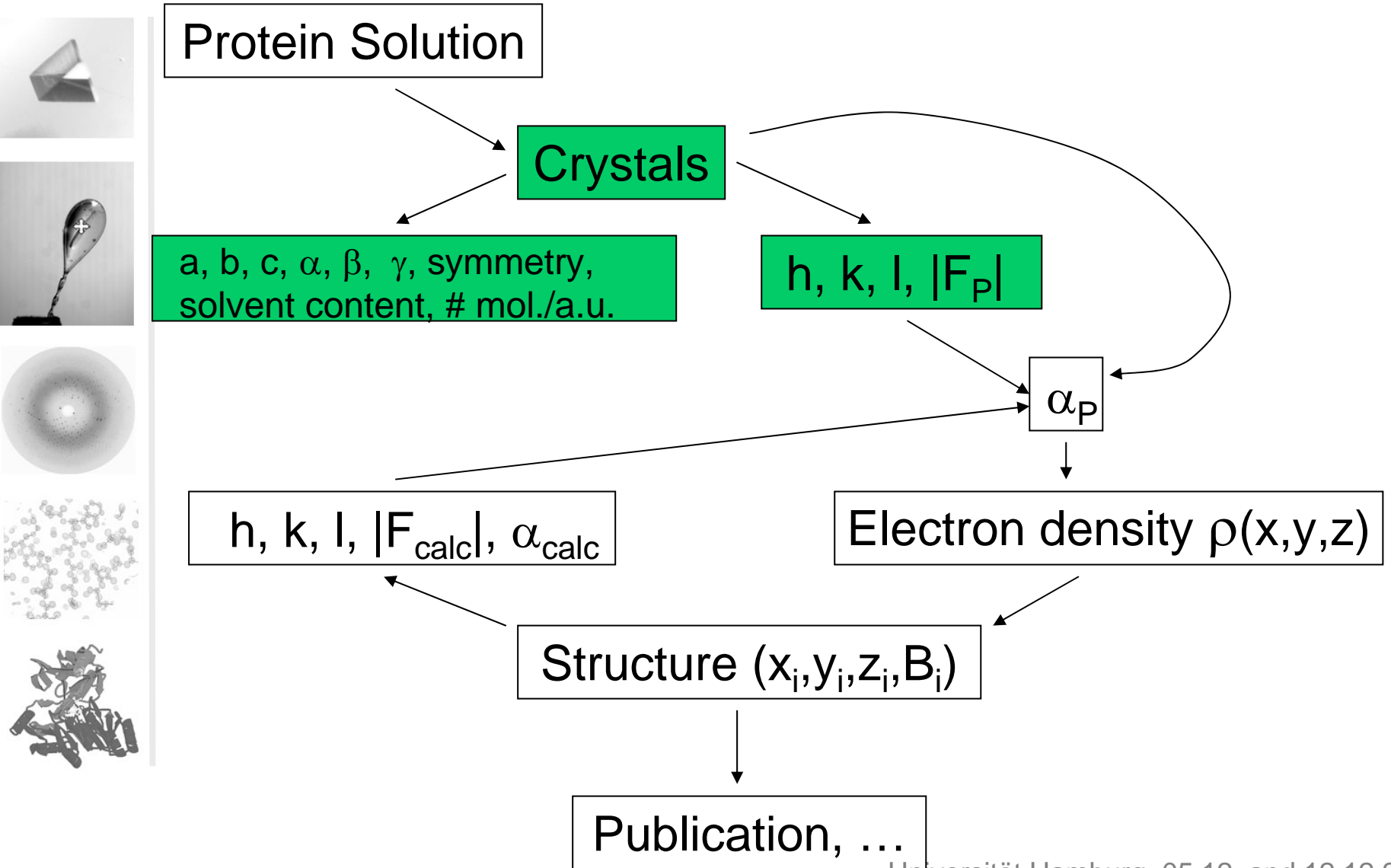
# What Is A Crystal?

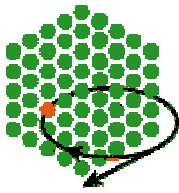
A crystal can be defined as an arrangement of building blocks which is **periodic** in **three dimensions**.



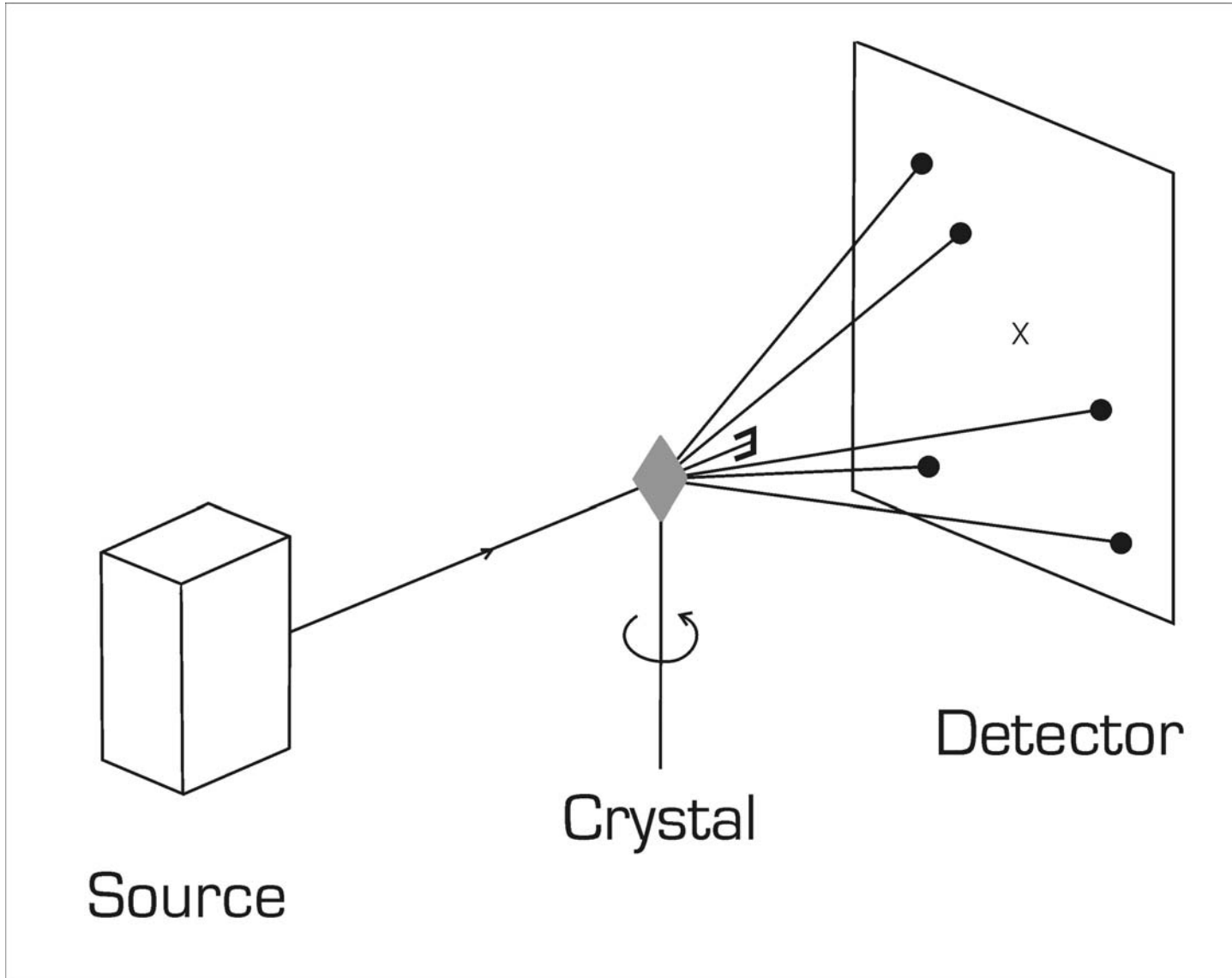
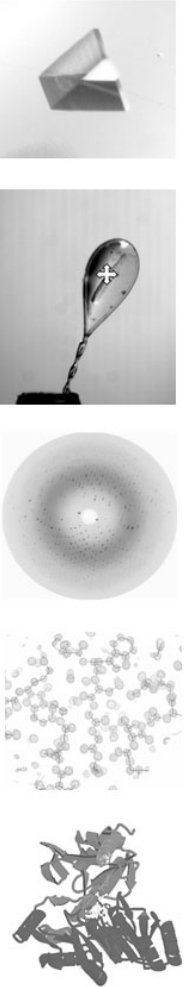


# The Diffraction Experiment



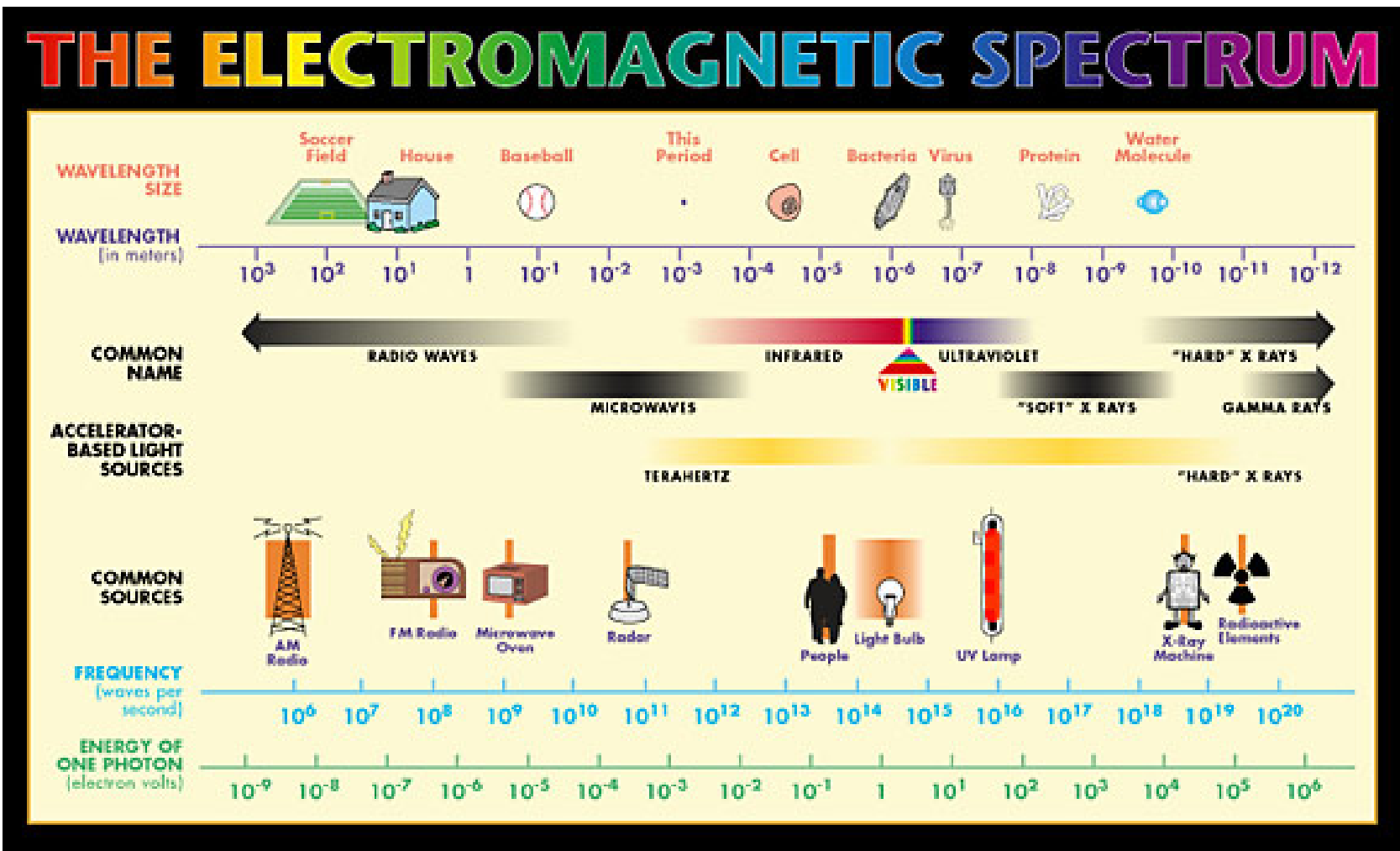
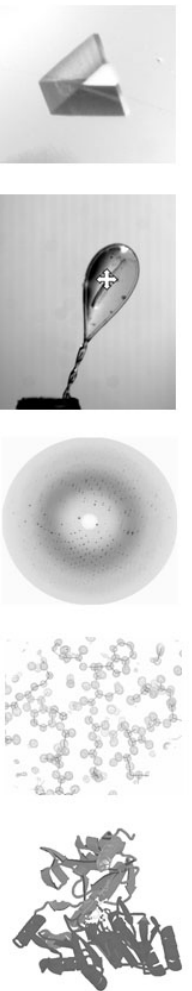
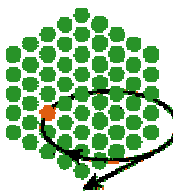


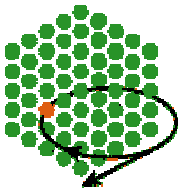
# The Diffraction Experiment



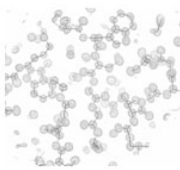
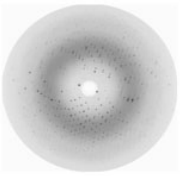
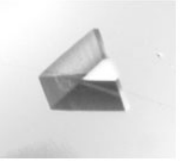


# X-Rays

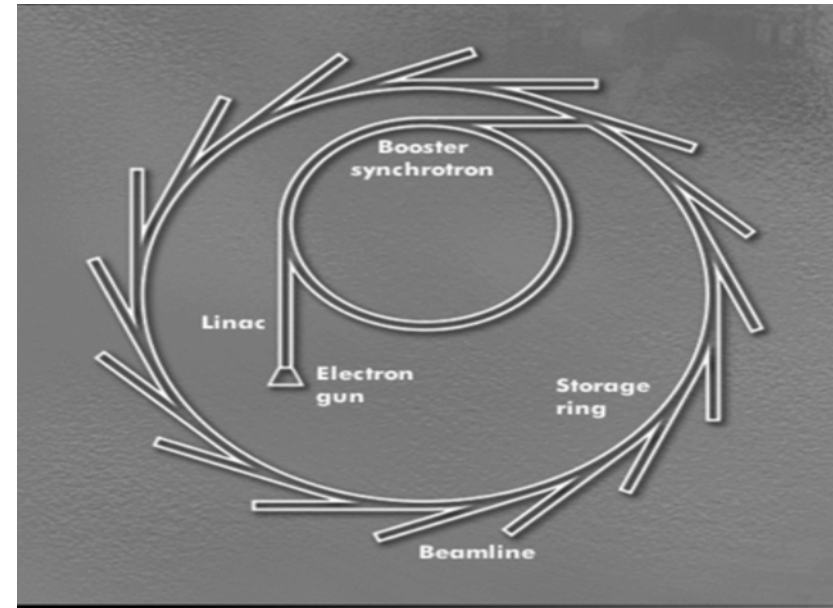




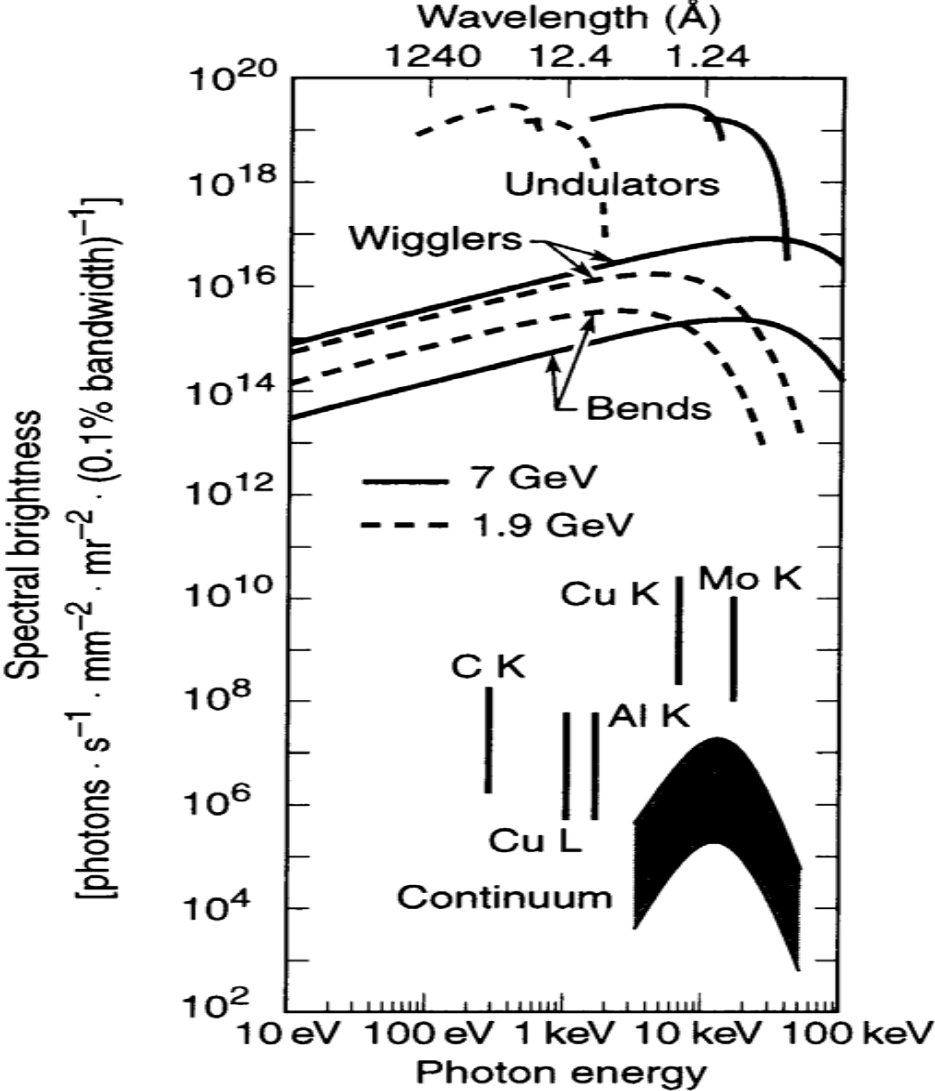
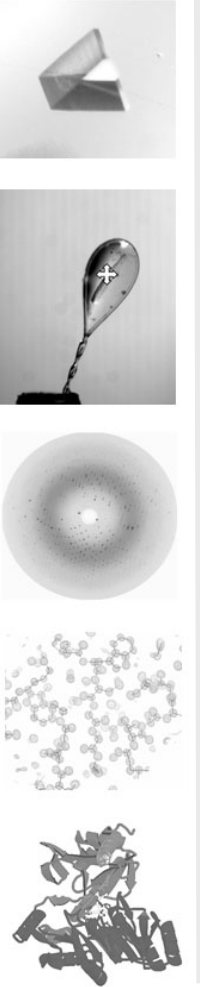
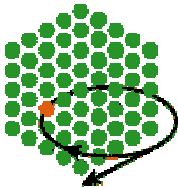
# Synchrotron Radiation

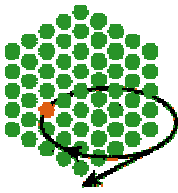


**Synchrotron Radiation** occurs when a charge moves at *relativistic speed* following a *curved trajectory*.



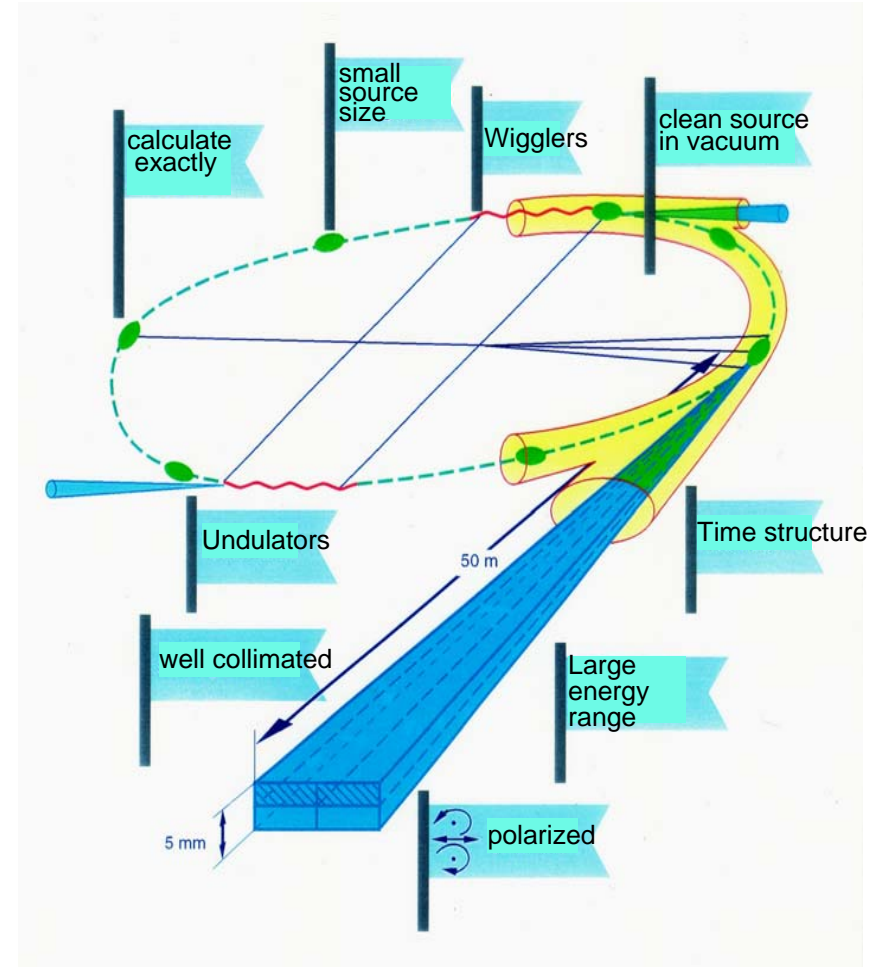
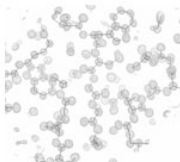
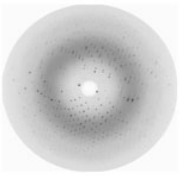
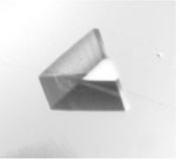
# Synchrotron Radiation

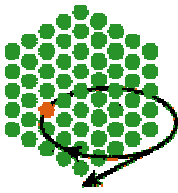




# Synchrotron Radiation

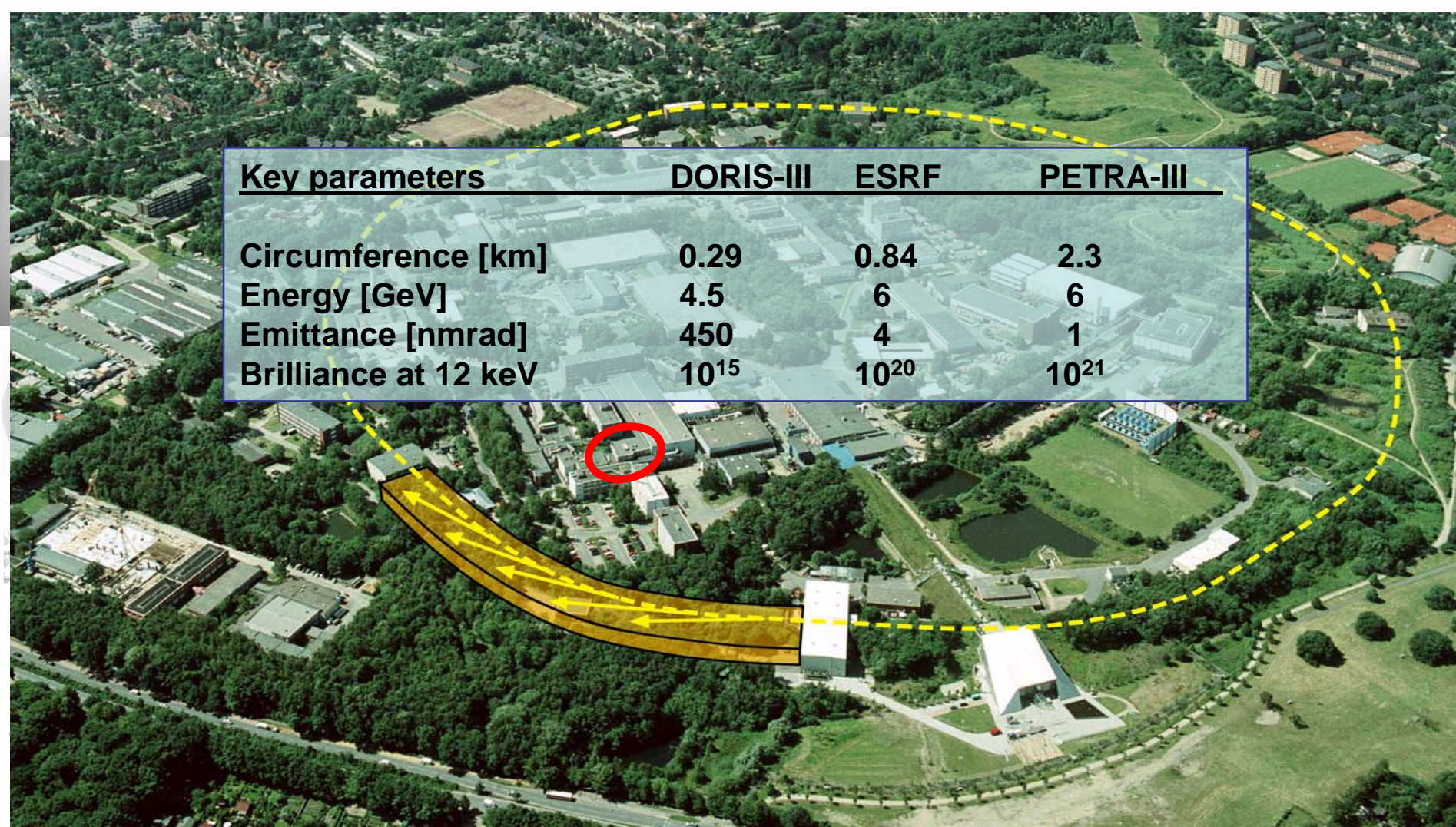
1. Continuous spectrum
2. Wavelength tunability
3. Highly collimated
4. Polarized
5. Time structure
6. High brilliance

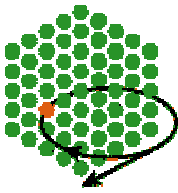




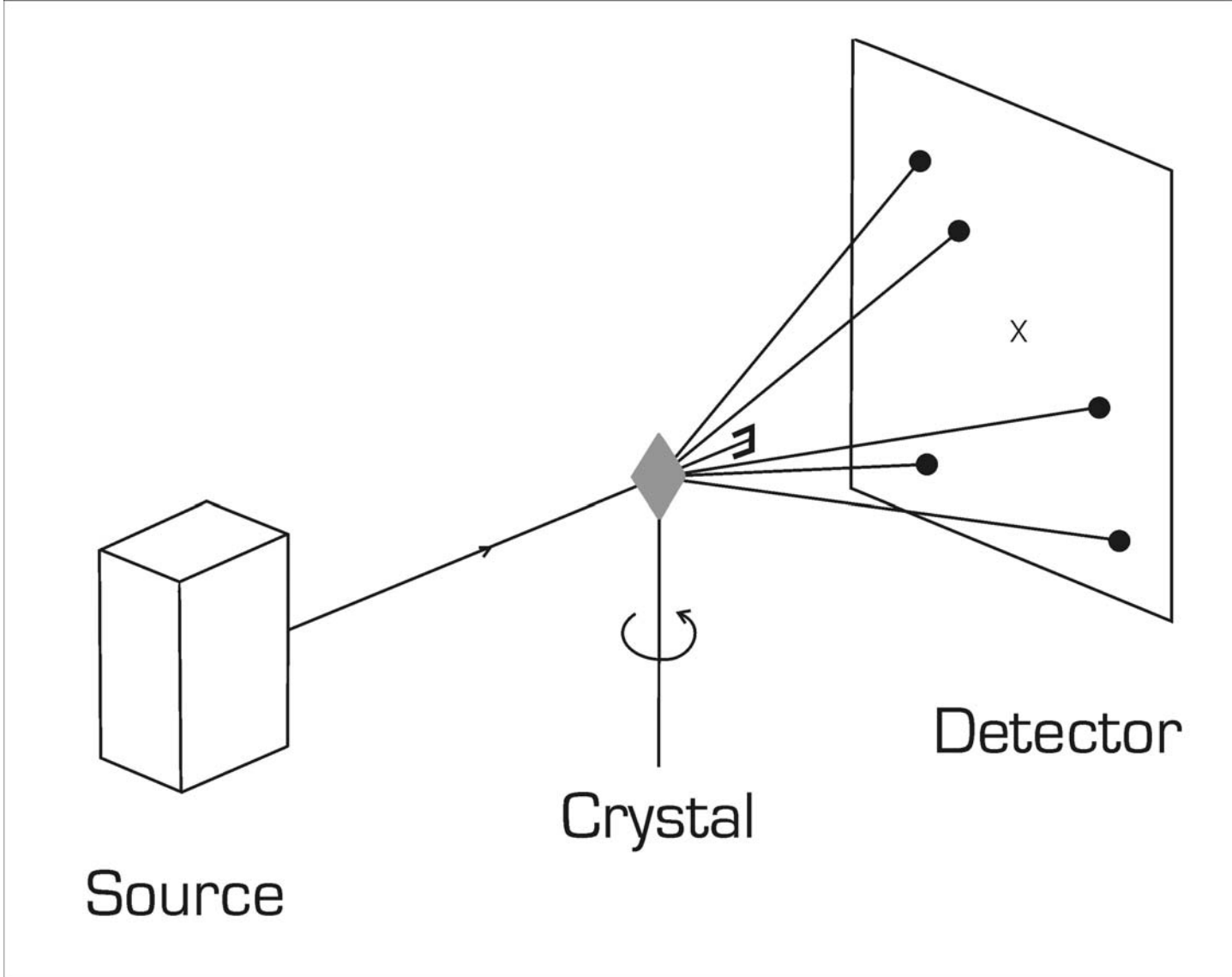
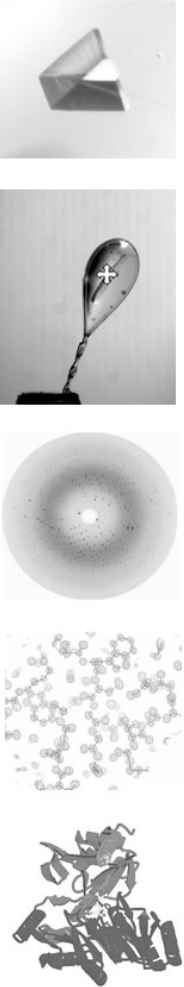
# EMBL-HH and PETRA-III

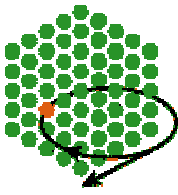
<u>Key parameters</u>	<u>DORIS-III</u>	<u>ESRF</u>	<u>PETRA-III</u>
Circumference [km]	0.29	0.84	2.3
Energy [GeV]	4.5	6	6
Emittance [nmrad]	450	4	1
Brilliance at 12 keV	$10^{15}$	$10^{20}$	$10^{21}$



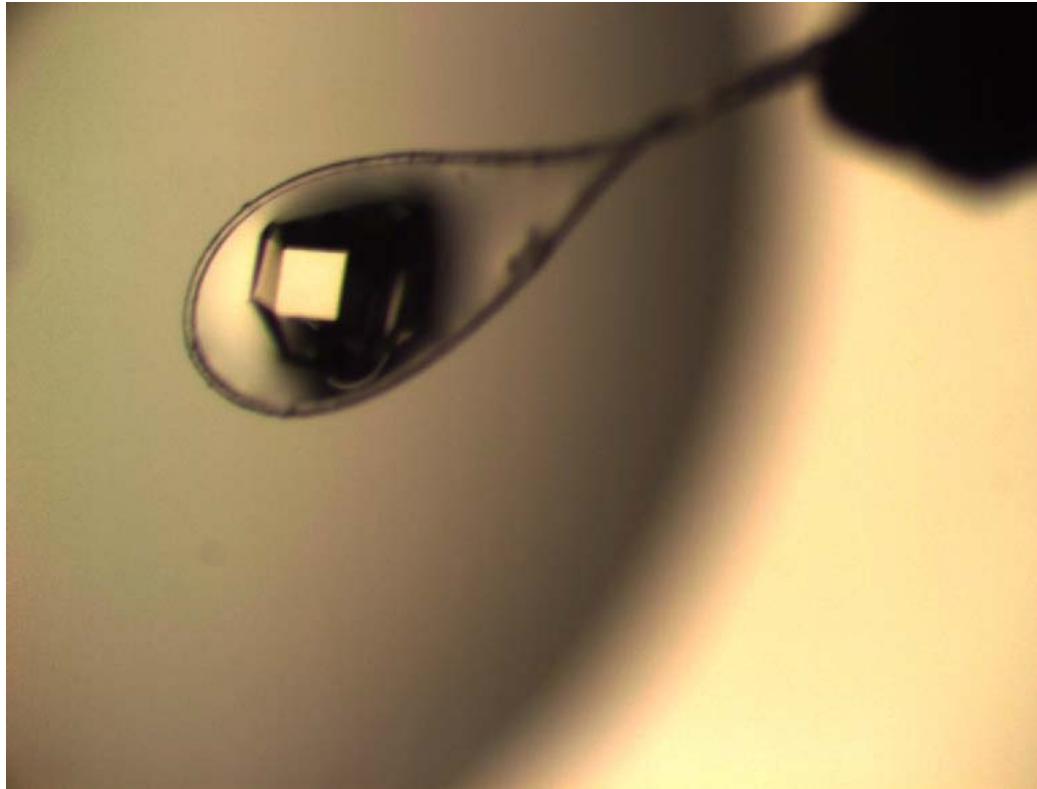
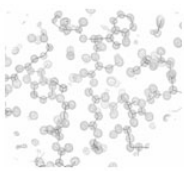
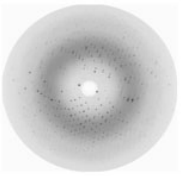
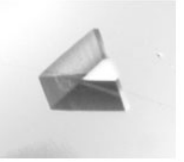


# The Diffraction Experiment



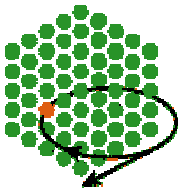


# What Does it Really Look Like?

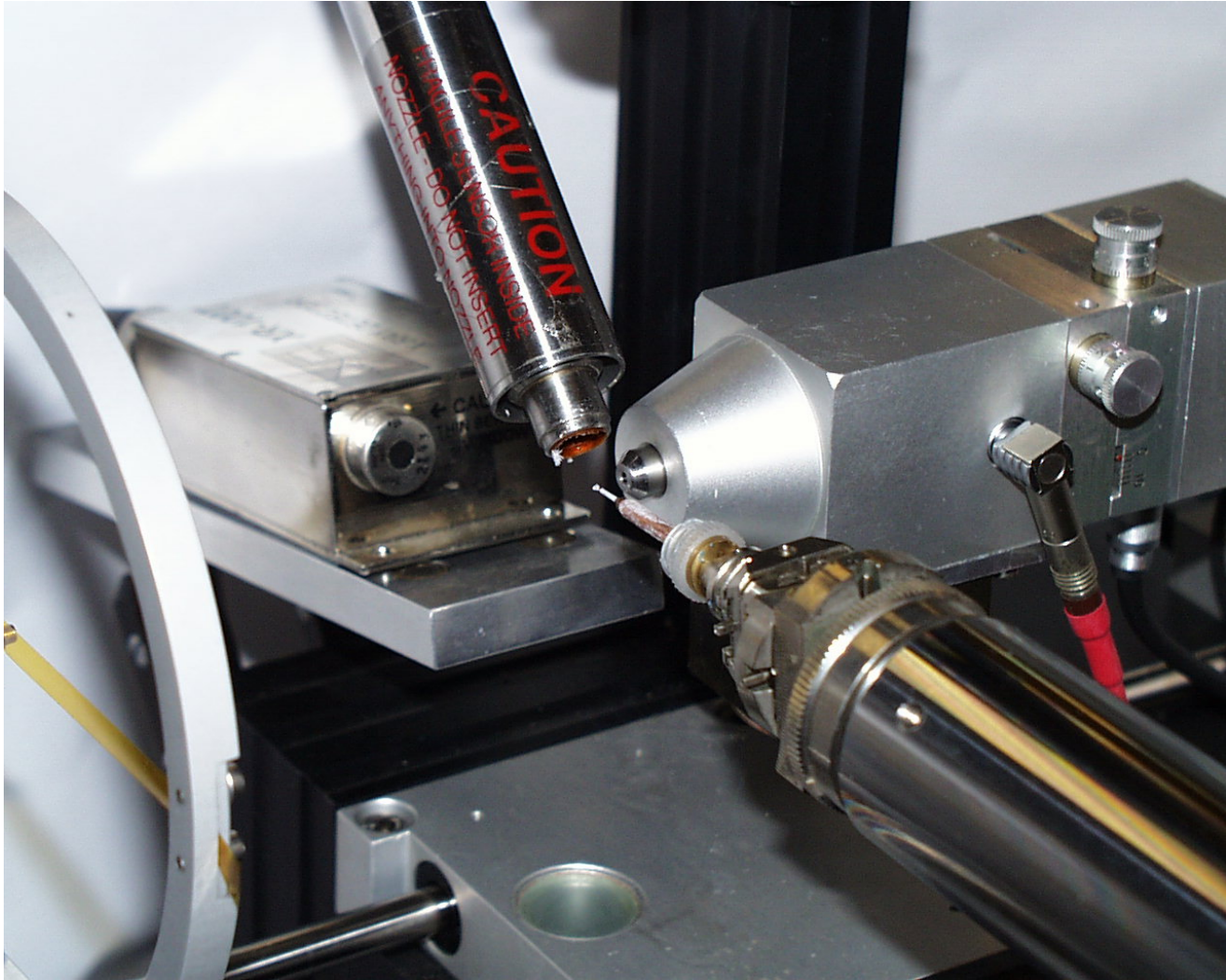
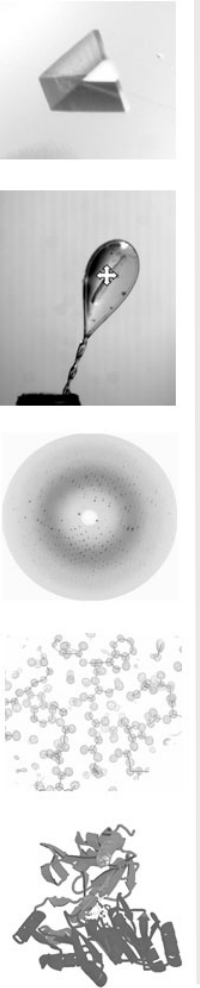


**1 mm**

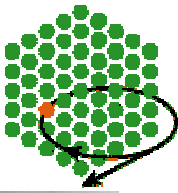
---



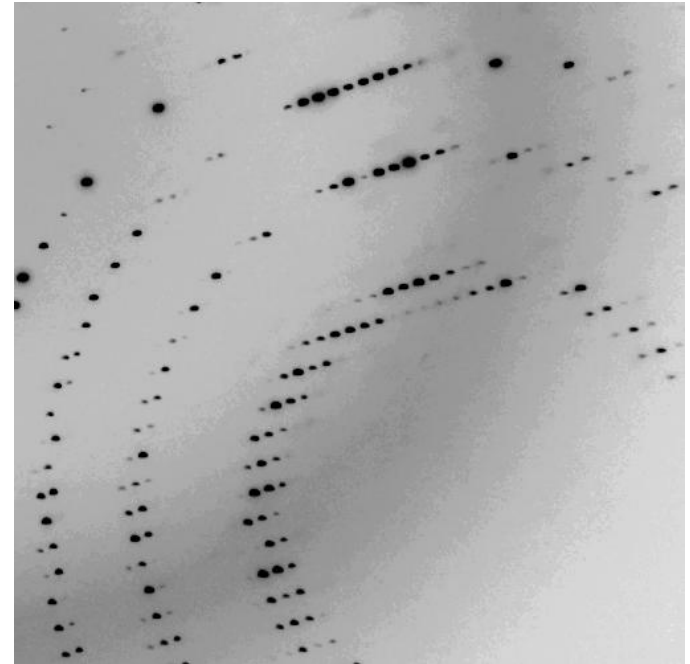
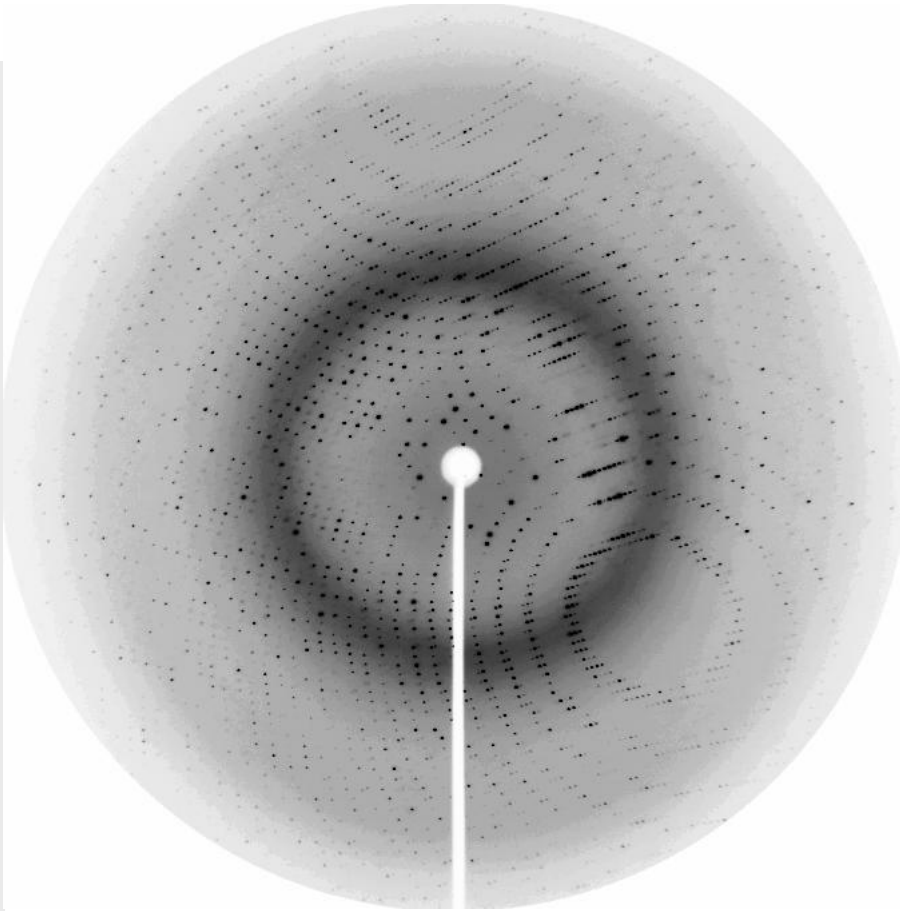
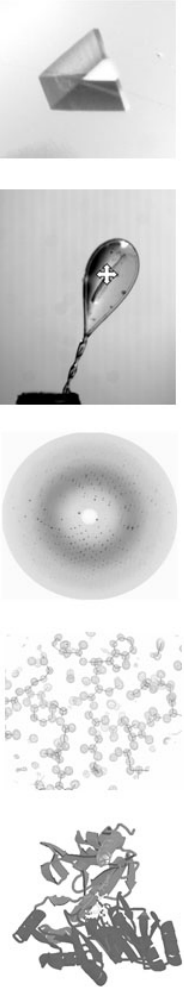
# What Does it Really Look Like?





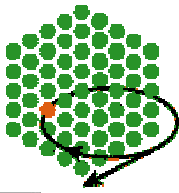


# A Typical Diffraction Pattern

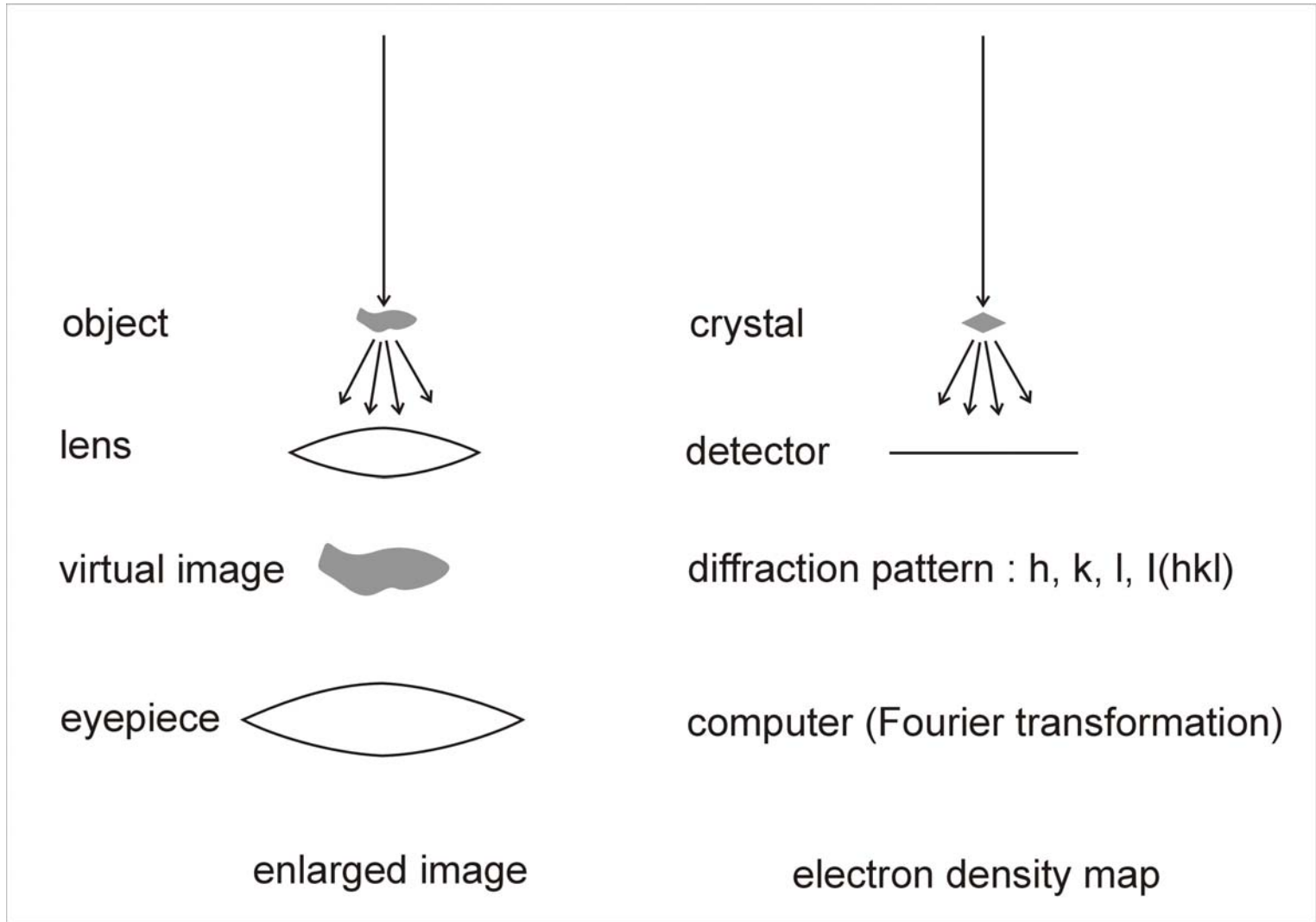
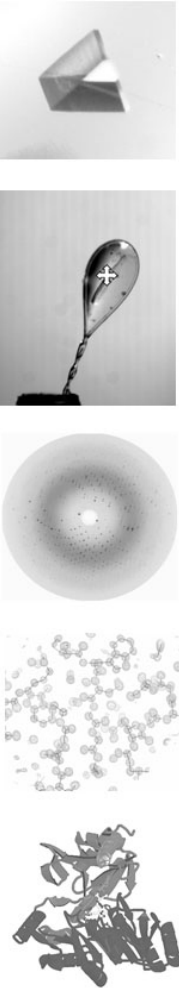


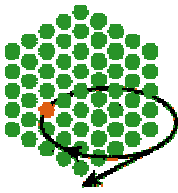
423	427	430	438	446	429	466	463	471	465	436	482	487	420	429
428	437	452	444	479	463	493	523	481	498	508	492	468	454	443
444	443	462	467	511	539	552	553	569	577	542	517	510	487	468
448	457	491	502	594	615	695	685	662	687	645	582	548	517	488
473	484	537	577	670	754	852	884	934	966	857	738	634	577	516
493	515	593	656	830	1081	1511	1932	2284	2337	1847	1128	764	622	547
502	529	617	740	1168	2605	5824	10432	14677	14750	8939	3090	1039	655	563
513	521	636	868	2304	9173	21158	39982	44400	38837	20638	8966	1286	692	564
491	533	646	957	2829	12021	26401	39395	41797	31614	15800	4664	1200	681	573
504	543	649	878	2074	6871	12827	16143	15621	11003	5464	2072	911	655	568
493	546	611	749	1202	2287	3387	3725	3386	2413	1860	991	717	583	514
487	513	578	632	783	933	1094	1141	1114	998	868	724	608	528	489
484	488	533	559	632	689	737	747	750	709	667	603	562	511	468
462	484	486	509	535	574	595	592	608	587	552	524	506	480	449
451	455	465	474	486	506	524	517	514	509	489	479	470	455	433

**h, k, l** Miller indices  
**I(h,k,l)** intensity

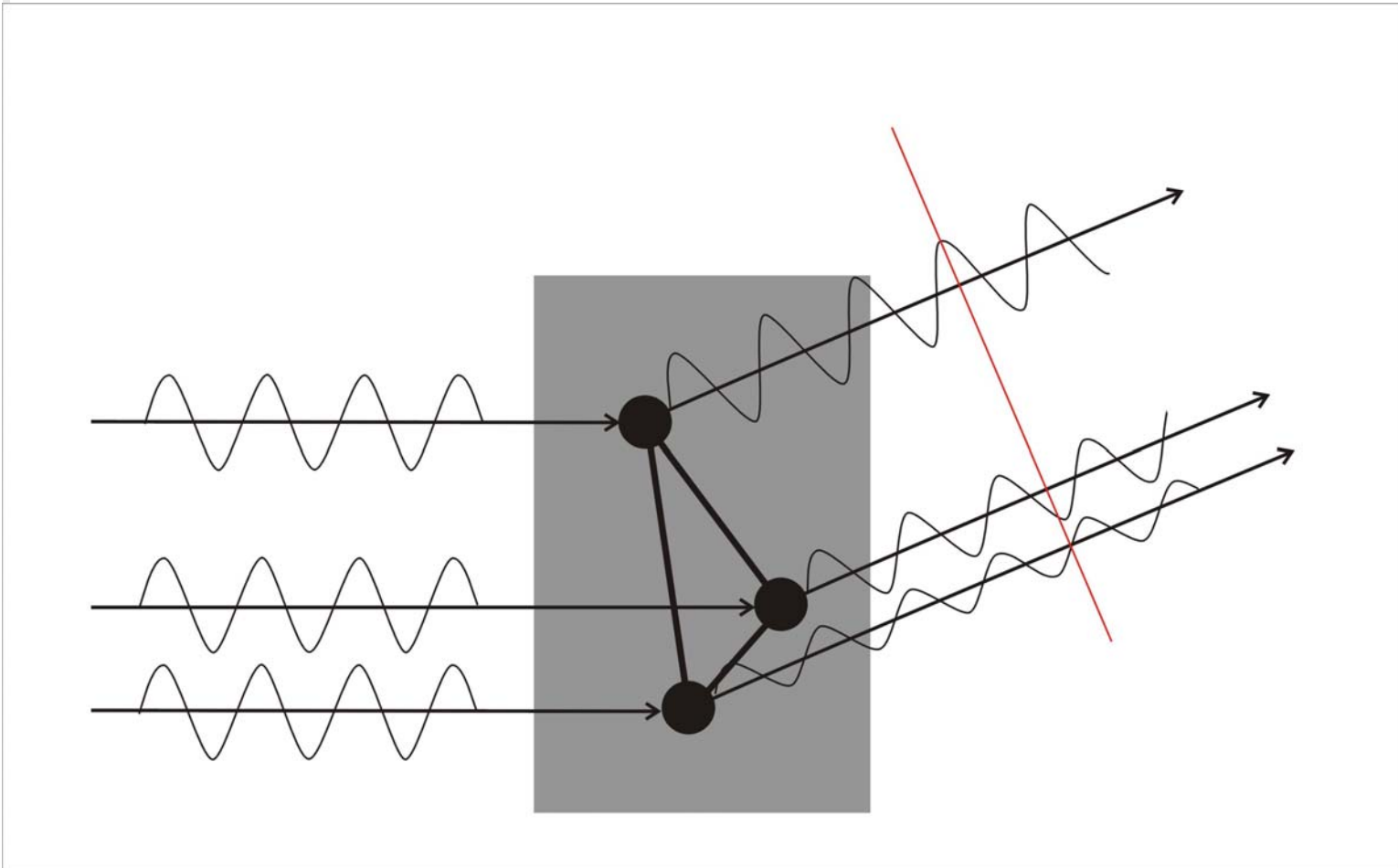
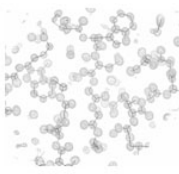
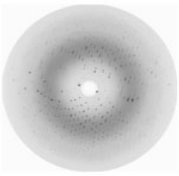
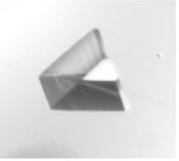


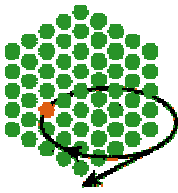
# Diffraction vs. Microscopy



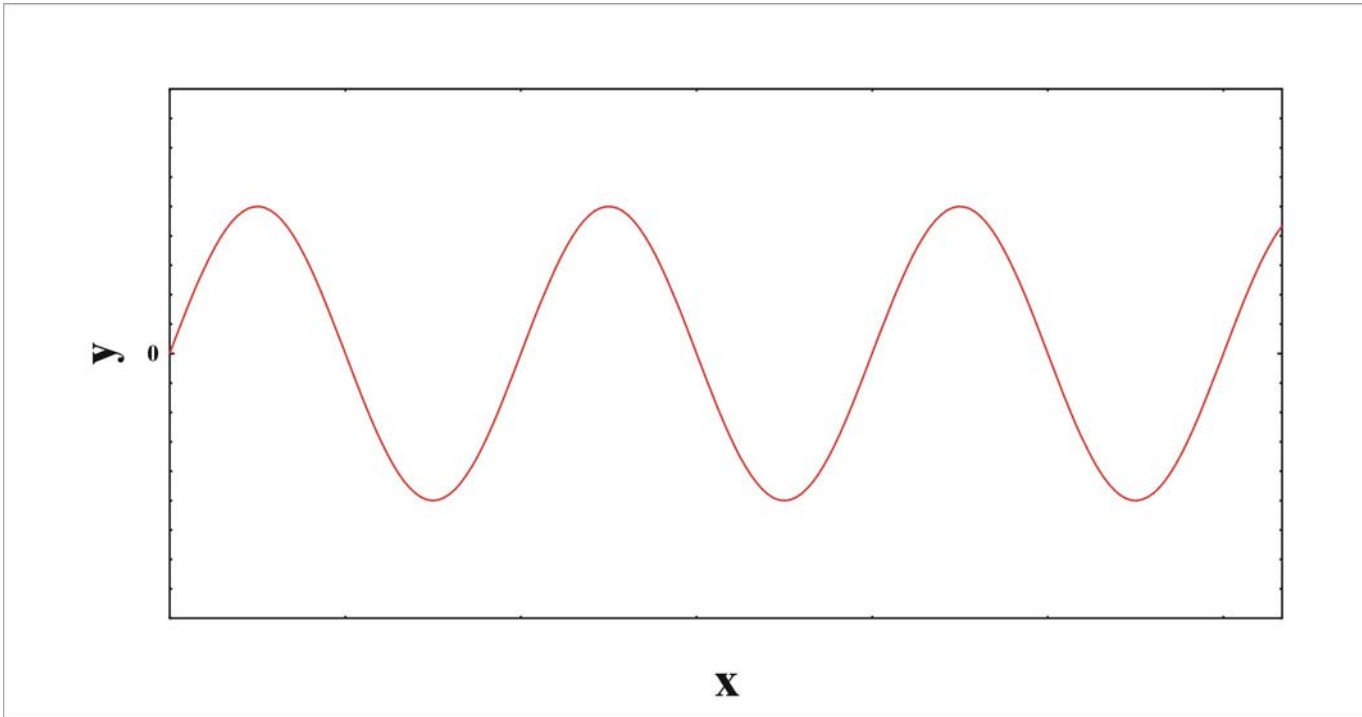
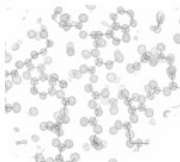
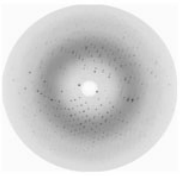
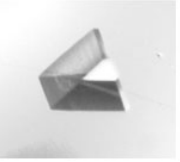


# Diffraction by a Molecule in a Crystal



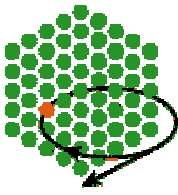


# Description of Waves - 1



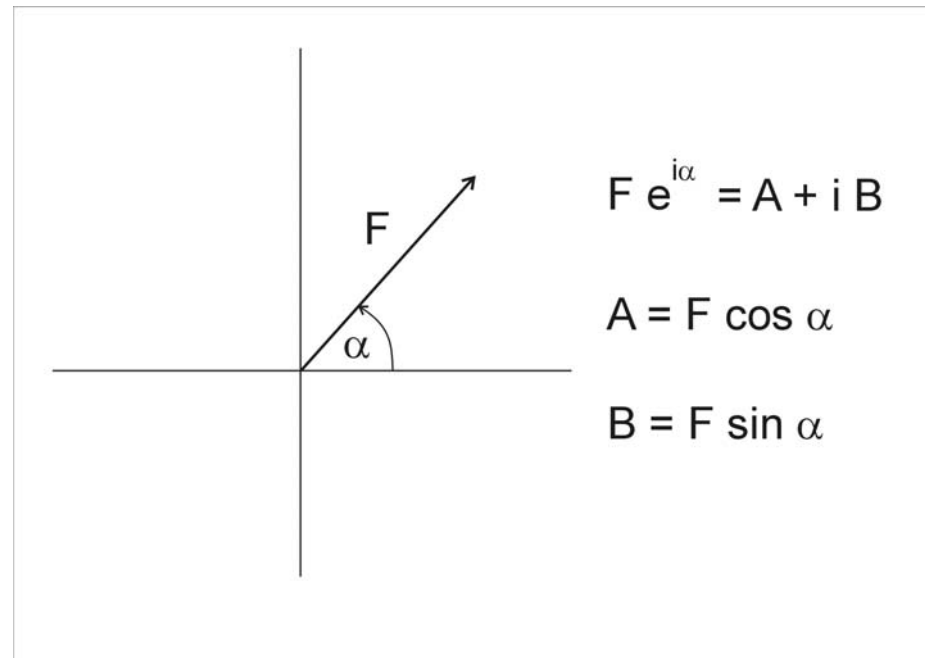
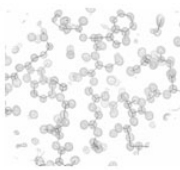
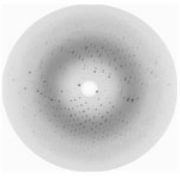
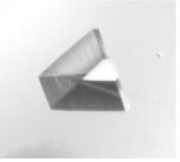
Every wave can be described by three parameters :  
amplitude  $F$ , wavelength  $\lambda$  and phase  $\alpha$

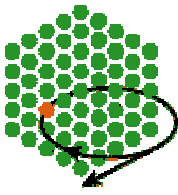
$$y(x) = F \cos (2\pi x/\lambda - \alpha)$$



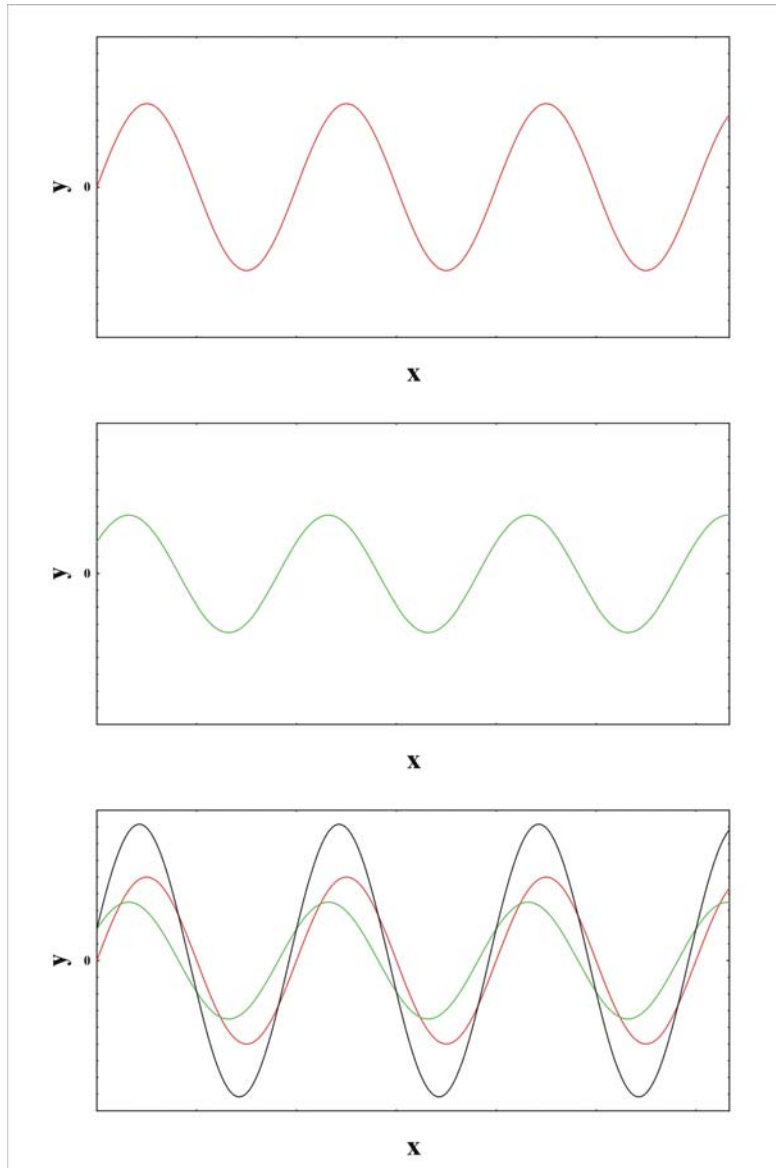
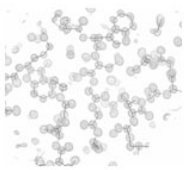
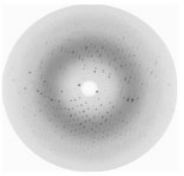
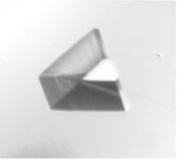
# Description of Waves - 2

Since we are only dealing with waves of the same wavelength  $\lambda$ , we only need two parameters to describe them. Complex numbers (or vectors) provide a very convenient solution to that.

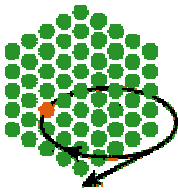




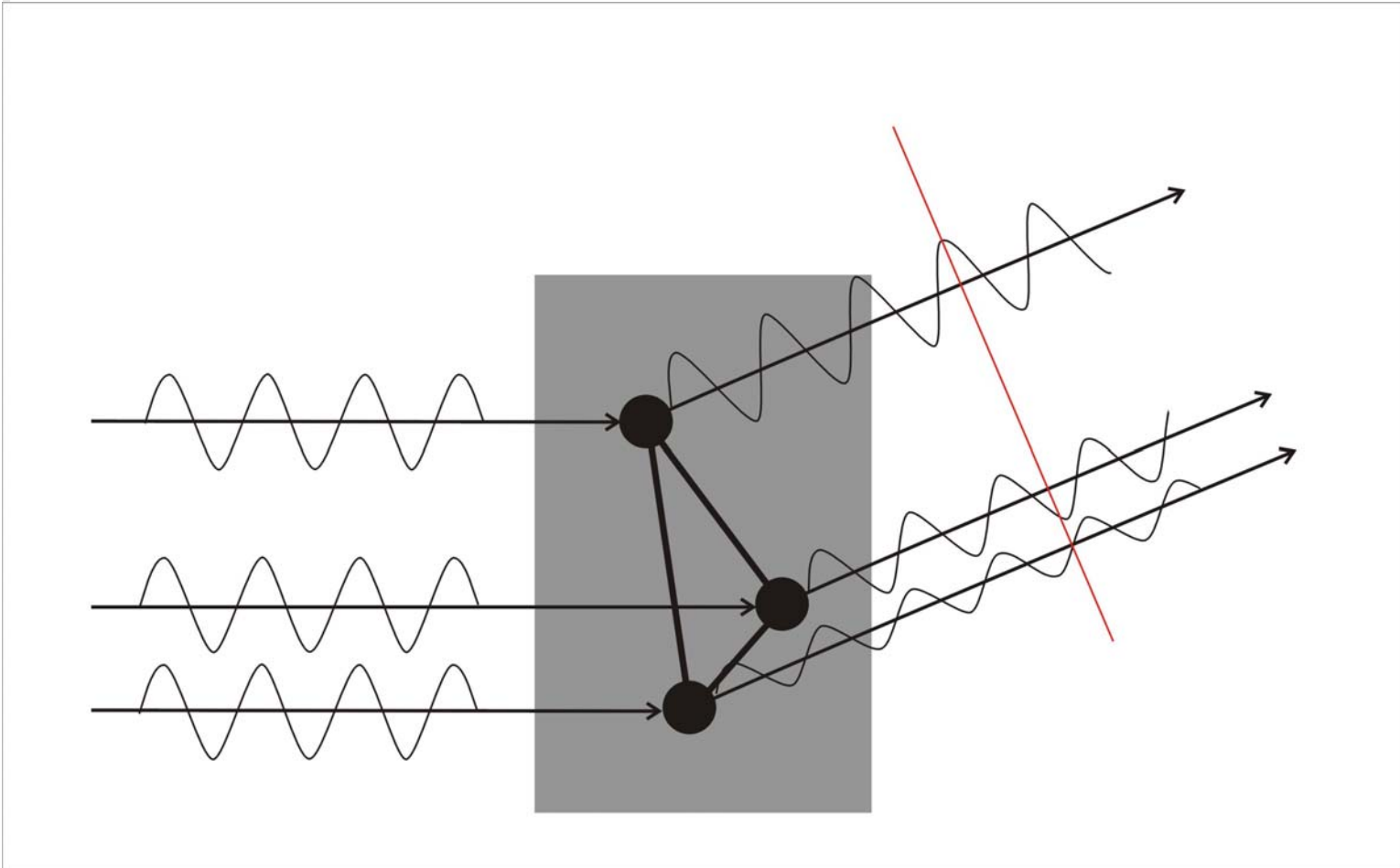
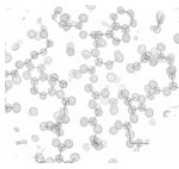
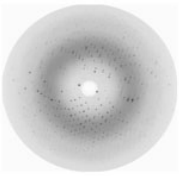
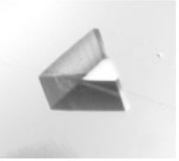
# Summation of Waves

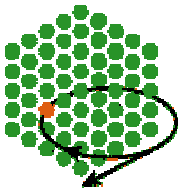


If two (or more) waves with the **same wavelength** but different **amplitudes** and **phases** are added up, the resulting wave has again the **same wavelength** but a **different amplitude** and a **different phase**.



# The Structure Factor Equation



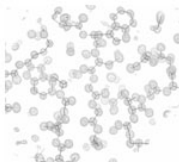
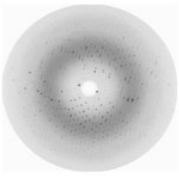
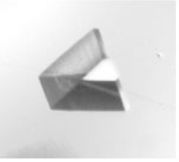


# The Structure Factor Equation

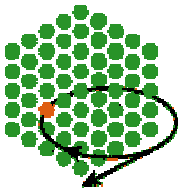
The calculation of  $F(hkl)$  from a structure  $(x_j, y_j, z_j)$  is nothing but a summation of the waves originating from each atom ( $j$ ) in the direction defined by  $(hkl)$ .

$$F(hkl) = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

It is important to note that every atom of the structure contributes to each and every reflection of the diffraction pattern.







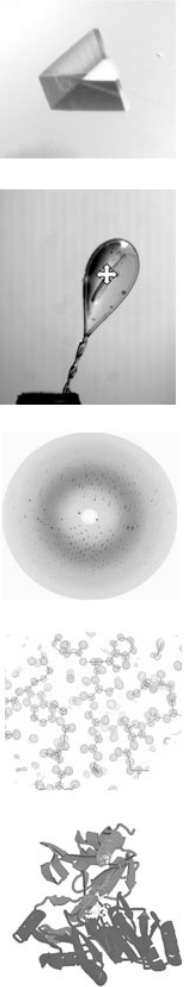
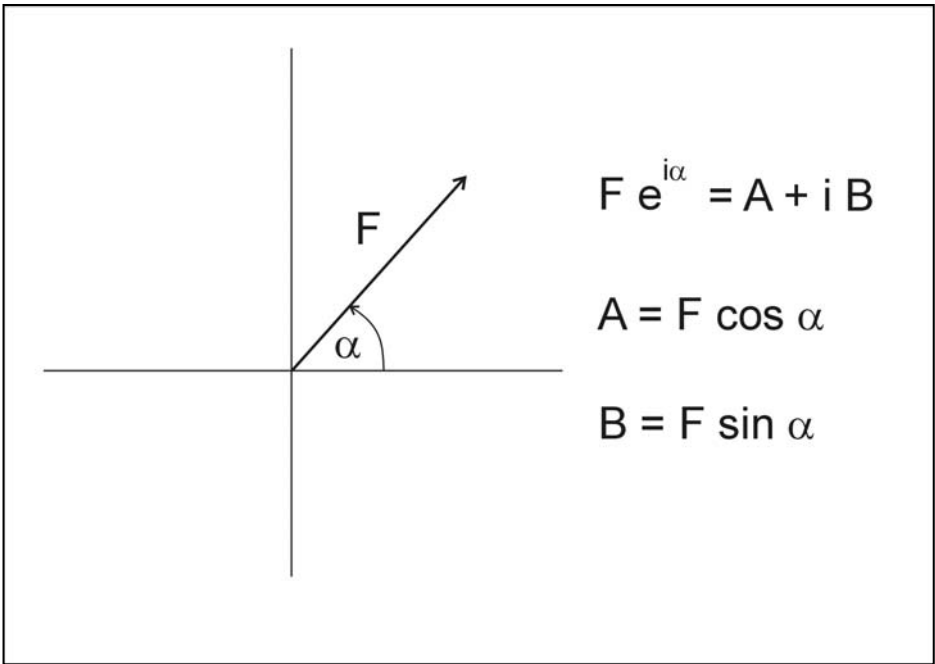
# The Structure Factor Equation

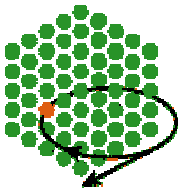
$$F(hkl) = |F(hkl)| e^{i\alpha(hkl)} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Structure factor amplitude  
 $|F(hkl)| \propto I(hkl)^{1/2}$

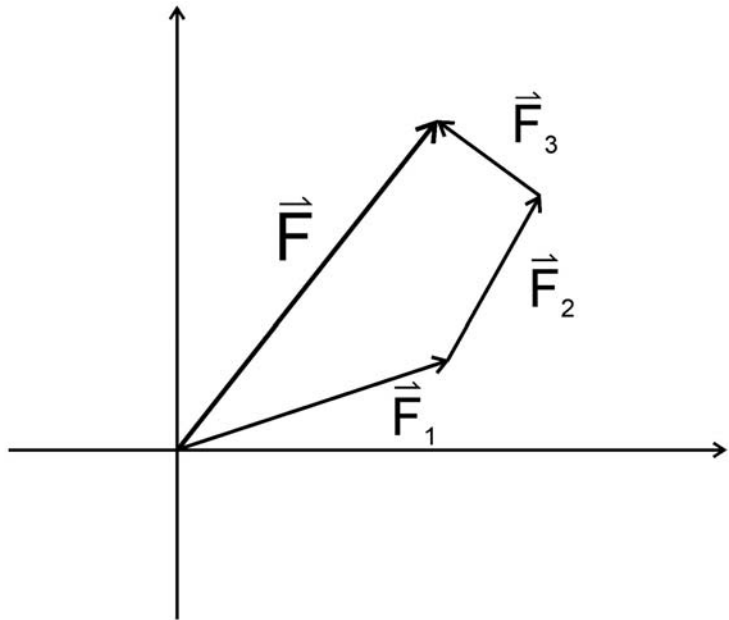
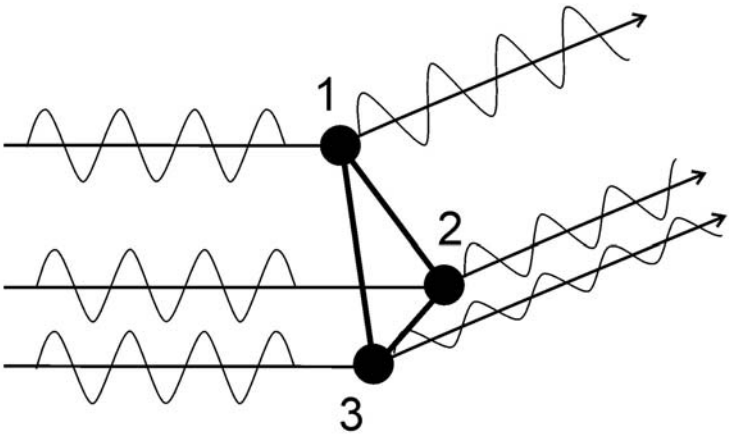
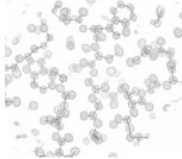
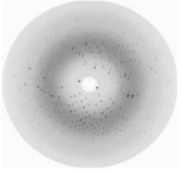
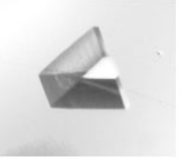
Atomic form factor  $f_j$

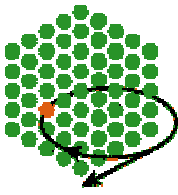
Phase  $\alpha(hkl)$





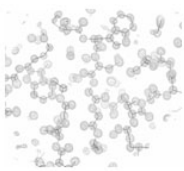
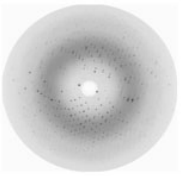
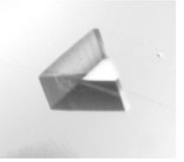
# The Structure Factor Equation



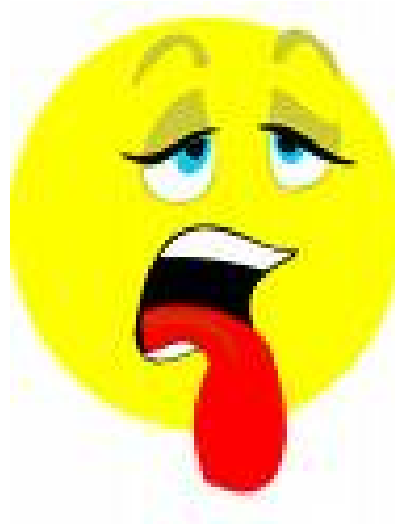
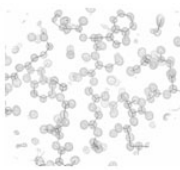
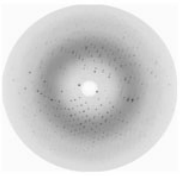
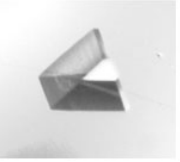
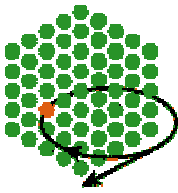


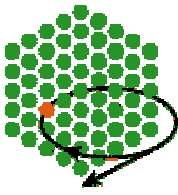
# Structure Factor Equation - Summary

- If we know the structure of the protein in a given crystal represented by  $(f_j, x_j, y_j, z_j, B_j)$ , we can calculate the diffraction pattern exactly.
- The calculated structure factor amplitudes and phases are typically referred to as  $|F_{\text{calc}}(hkl)|$  and  $\alpha_{\text{calc}}(hkl)$ .

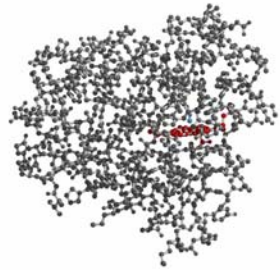
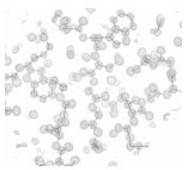
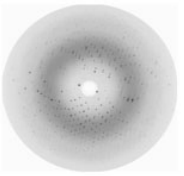
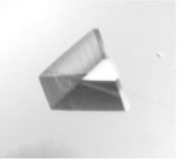


# Short Break

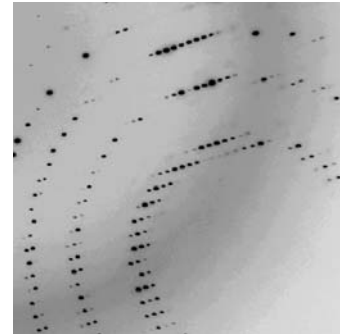
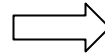




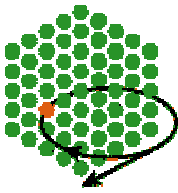
# So Far We Have Done This ...



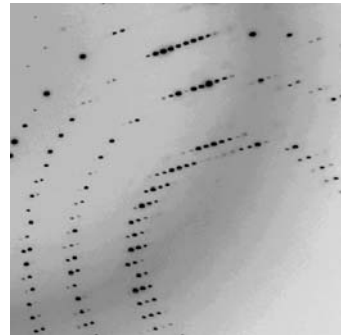
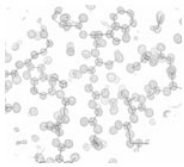
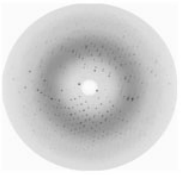
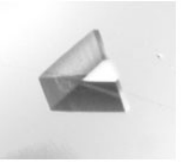
$f_j, x_j, y_j, z_j, B_j$



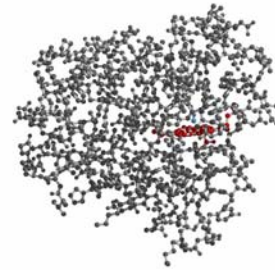
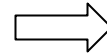
$F(hkl)$



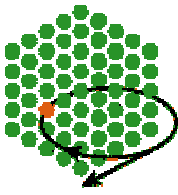
# But What We Really Want To Do Is ...



$F(hkl)$

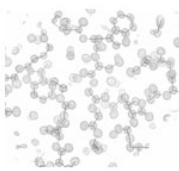
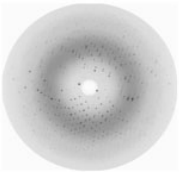
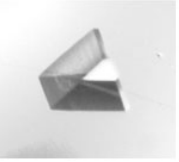


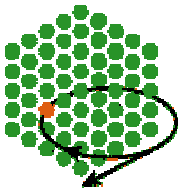
$f_j, x_j, y_j, z_j, B_j$



# The Structure Factor Equation

$$F(hkl) = |F(hkl)| e^{i\alpha(hkl)} = \sum_j f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$



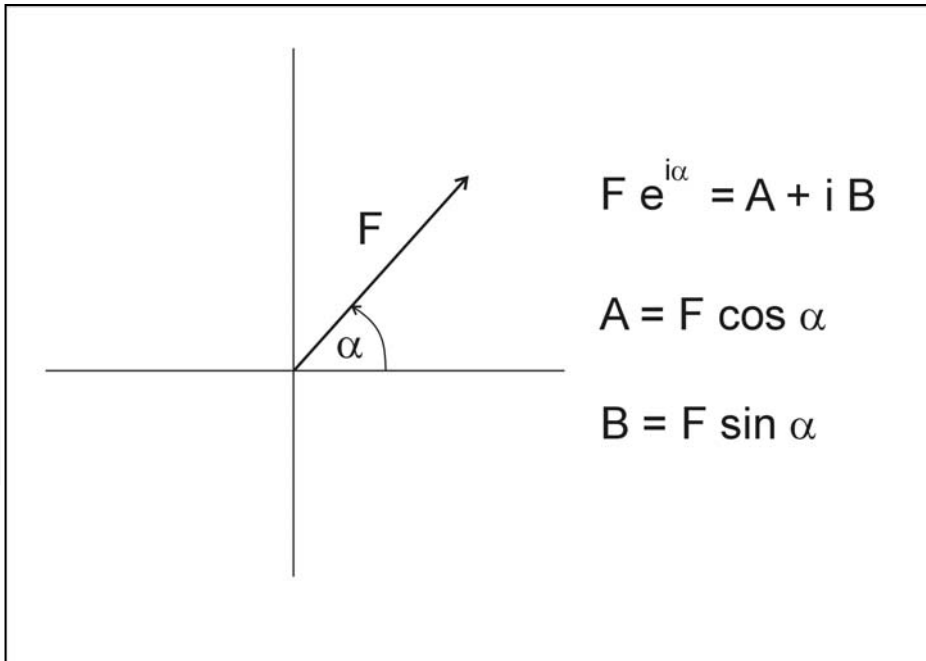


# The Electron Density Equation

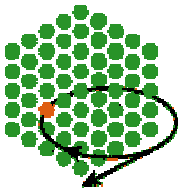
$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

Structure factor amplitude  
 $|F(hkl)| \propto I(hkl)^{1/2}$

Phase  $\alpha(hkl)$





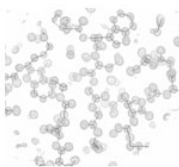
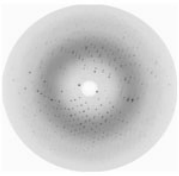
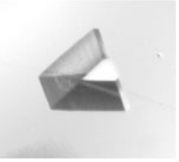


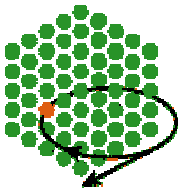
# The Electron Density Equation

The **electron density**  $\rho(x,y,z)$  is a three-dimensional function (with the unit  $e/\text{\AA}^3$ ), which describes where in the unit cell of the crystal the electrons (and therefore the atoms) are. It is basically the image of the structure we want to determine.

$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

It is important to note that every reflection (hkl) of the diffraction pattern contributes to the electron density  $\rho$  at each and every position (xyz) in the unit cell of the crystal.



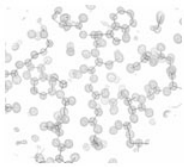
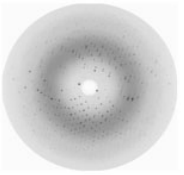
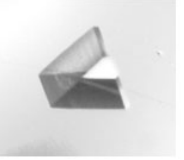


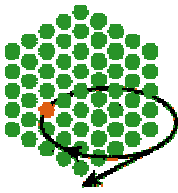
# The Phase Problem

From the diffraction pattern, we can only obtain the intensities  $I(hkl)$  of the reflections  $(hkl)$ .

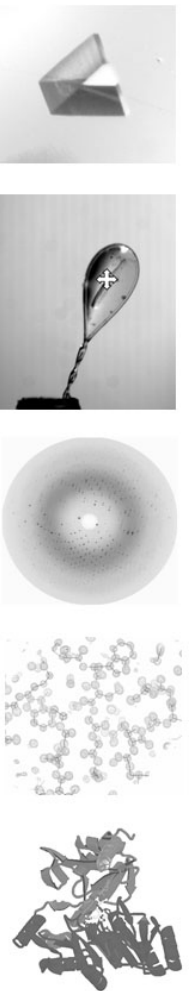
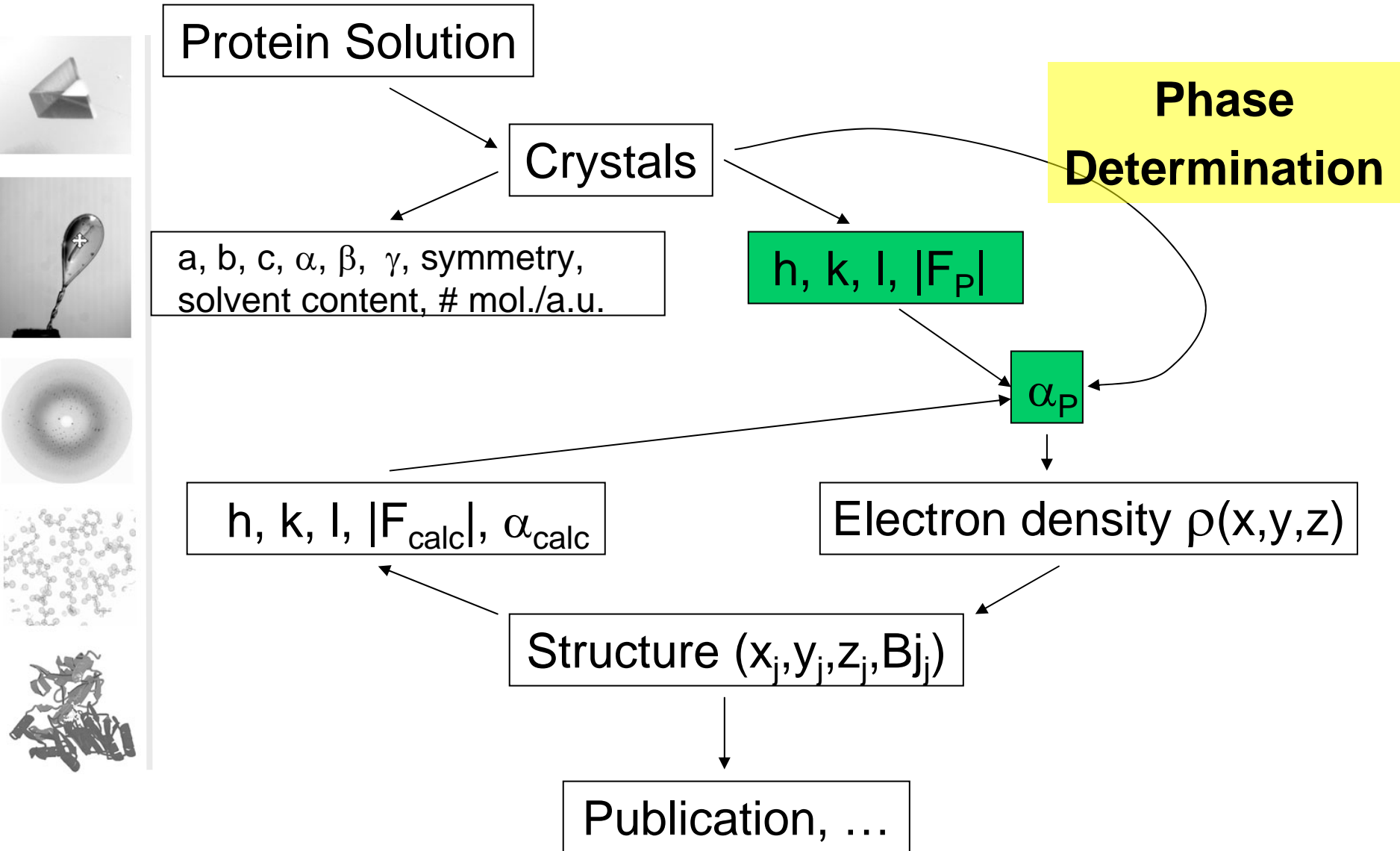
$$\begin{aligned} I(hkl) &\propto F(hkl) \cdot F^*(hkl) \\ &= |F(hkl)| e^{i\alpha(hkl)} \cdot |F(hkl)| e^{-i\alpha(hkl)} \\ &= |F(hkl)|^2 \end{aligned}$$

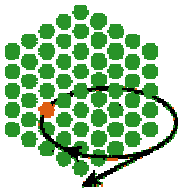
The phase  $\alpha(hkl)$  cannot be measured.





# A Flowchart of a Crystal Structure Analysis





# Phase Determination Methods

## 1. SIR, SIRAS, MIR, MIRAS

(single/multiple isomorphous replacement with anomalous scattering)

## 2. MAD

(multiple wavelength anomalous diffraction)

## 3. SAD (SAS)

(single wavelength anomalous diffraction/scattering)

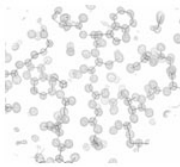
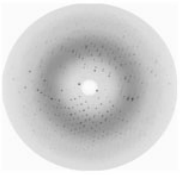
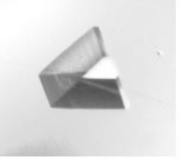
## 4. RIP, RIPAS

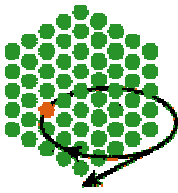
(radiation damage induced phasing with anomalous scattering)

## 5. MR

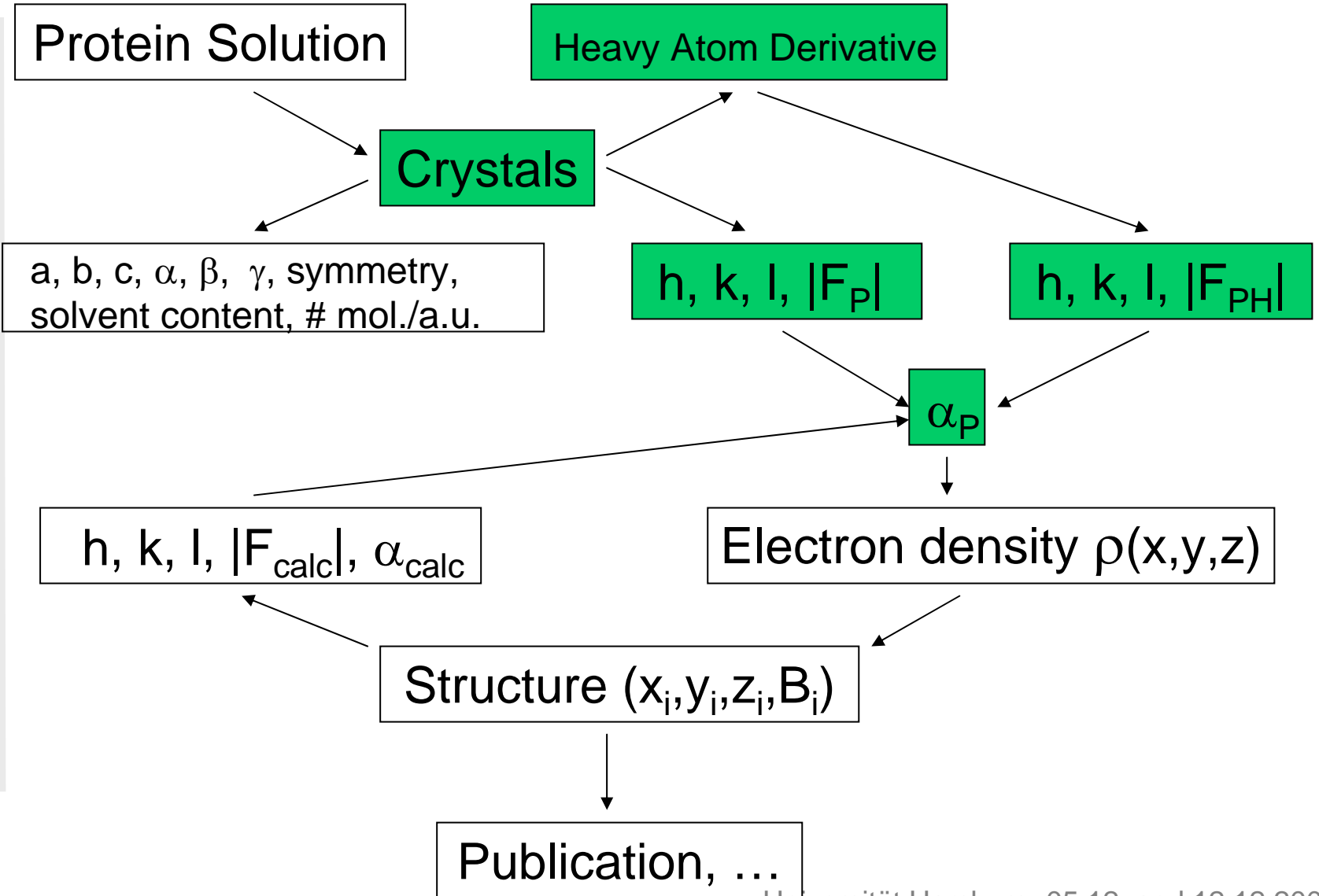
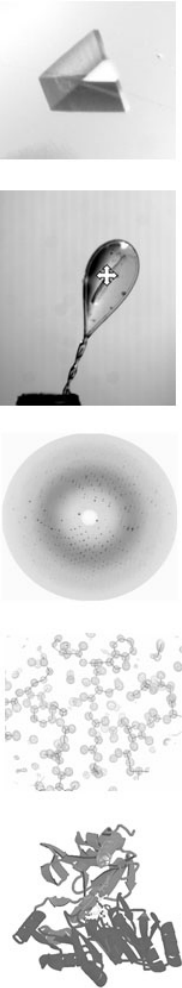
(molecular replacement)

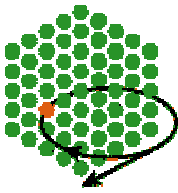
## 6. Direct Methods



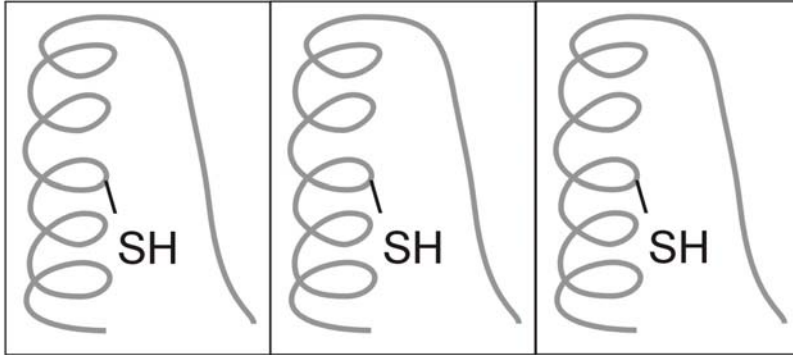
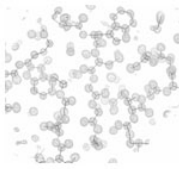
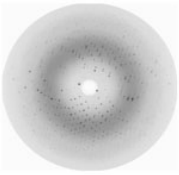
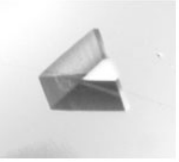


# The Process of Phase Determination



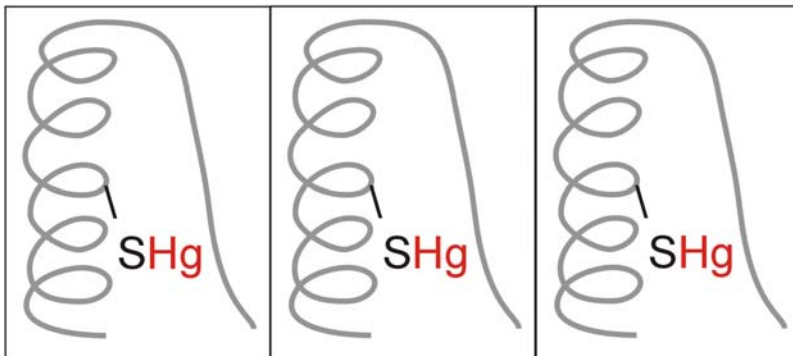


# Isomorphous Replacement

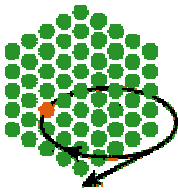


$h, k, l, |F_P(hkl)|$

$\alpha_P(hkl)$



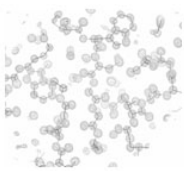
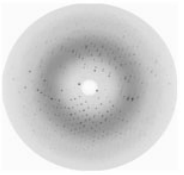
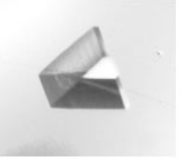
$h, k, l, |F_{PH}(hkl)|$

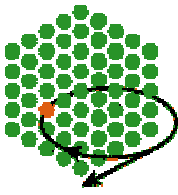


# Isomorphous Replacement

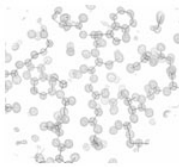
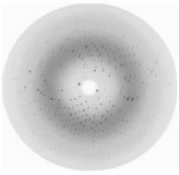
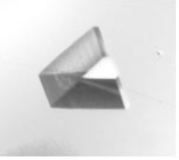
Step 1: put  $|F_P(hkl)|$  and  $|F_{PH}(hkl)|$  on the same scale

$$( R = 100 \cdot \sum_{hkl} | |F_{PH}| - |F_P| | / \sum_{hkl} |F_P| )$$





# Isomorphous Replacement



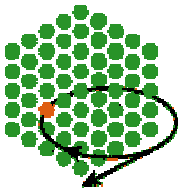
Step 1: put  $|F_P(hkl)|$  and  $|F_{PH}(hkl)|$  on the same scale

$$( R = 100 \cdot \sum_{hkl} | |F_{PH}| - |F_P| | / \sum_{hkl} |F_P| )$$

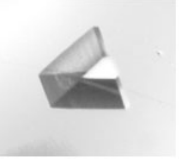
Step 2 : determine the positions of the heavy atoms (  $F_H$  )

from the differences (  $|F_{PH}(hkl)| - |F_P(hkl)|$  )





# Isomorphous Replacement



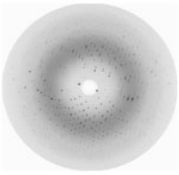
Step 1: put  $|F_P(hkl)|$  and  $|F_{PH}(hkl)|$  on the same scale

$$( R = 100 \cdot \sum_{hkl} | |F_{PH}| - |F_P| | / \sum_{hkl} |F_P| )$$

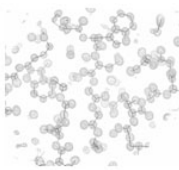


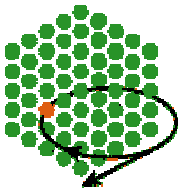
Step 2 : determine the positions of the heavy atoms (  $F_H$  )

from the differences (  $|F_{PH}(hkl)| - |F_P(hkl)|$  )

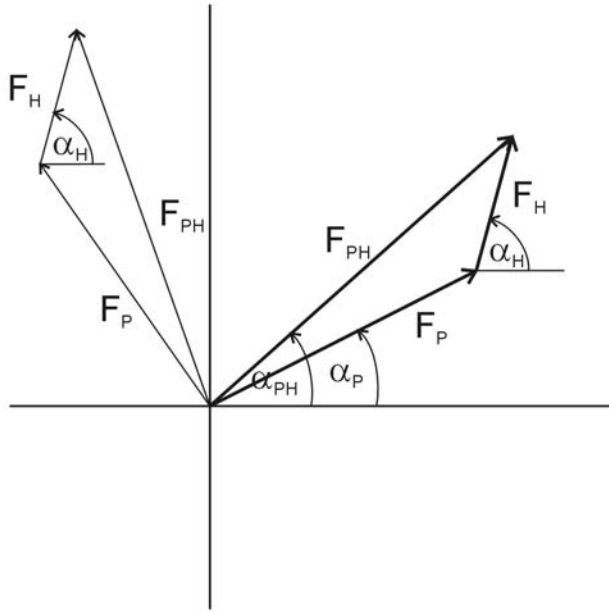
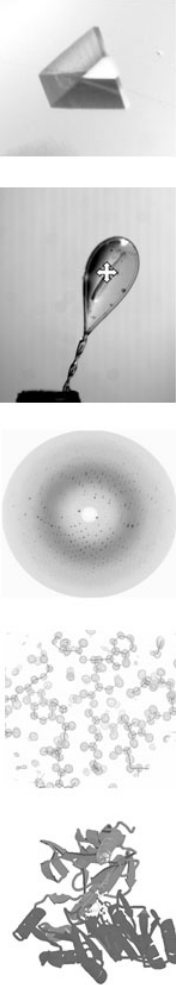


Step 3 : calculate  $\alpha_P(hkl)$  from  $|F_P(hkl)|$ ,  $|F_{PH}(hkl)|$  and  $F_H$



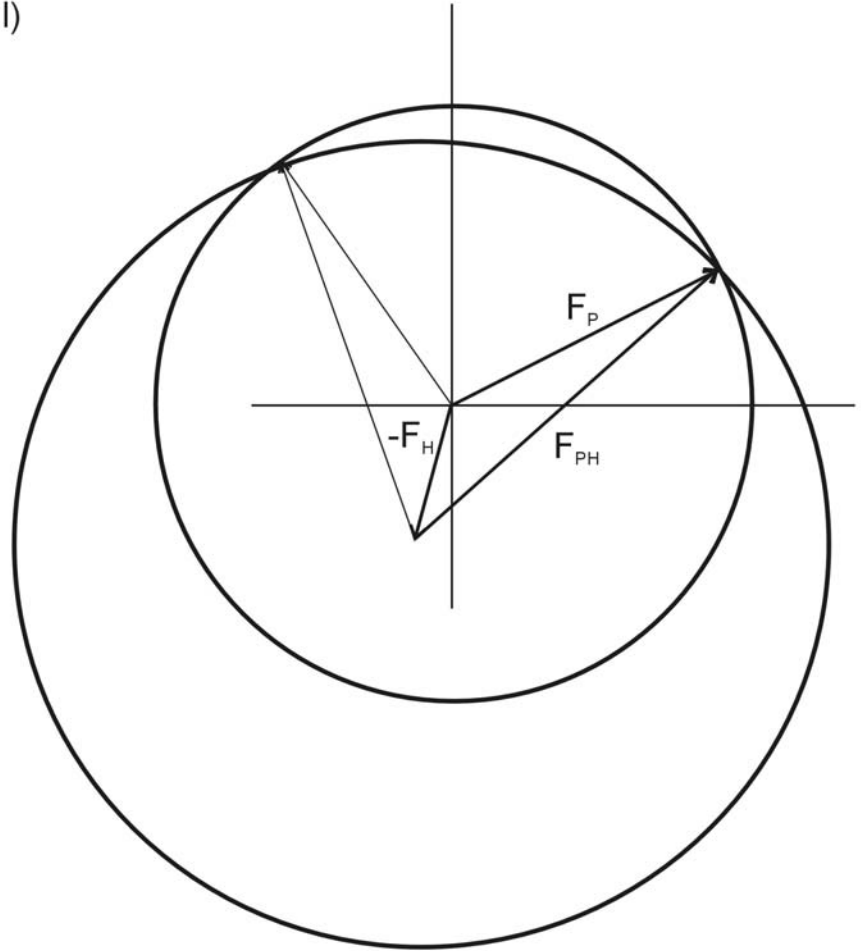


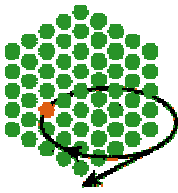
# The Harker Construction - SIR



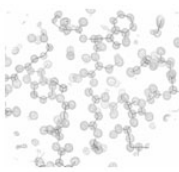
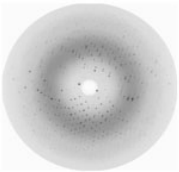
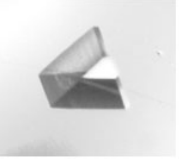
$$F_P + F_H = F_{PH}$$

$$F_P = -F_H + F_{PH}$$



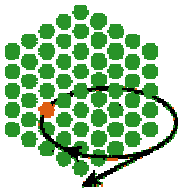


# The Harker Construction - SIR

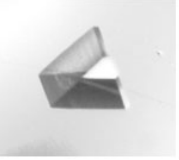


Instead of one value for  $\alpha_p(hkl)$  we obtain two possibilities  
→ *phase ambiguity*

How can this be made unambiguous ?



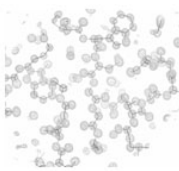
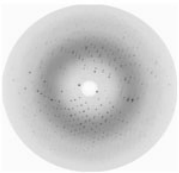
# The Harker Construction - SIR

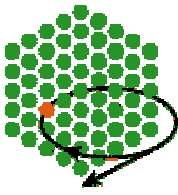


Instead of one value for  $\alpha_p(hkl)$  we obtain two possibilities  
→ *phase ambiguity*



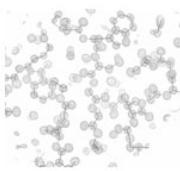
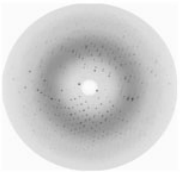
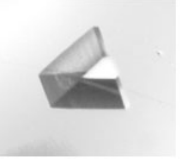
How can this be made unambiguous ?  
→ density modification

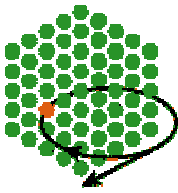




# Density Modification

Density Modification is a technique that uses additional information to improve an electron density

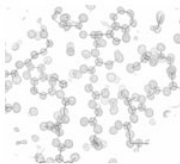
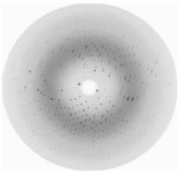
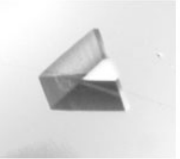




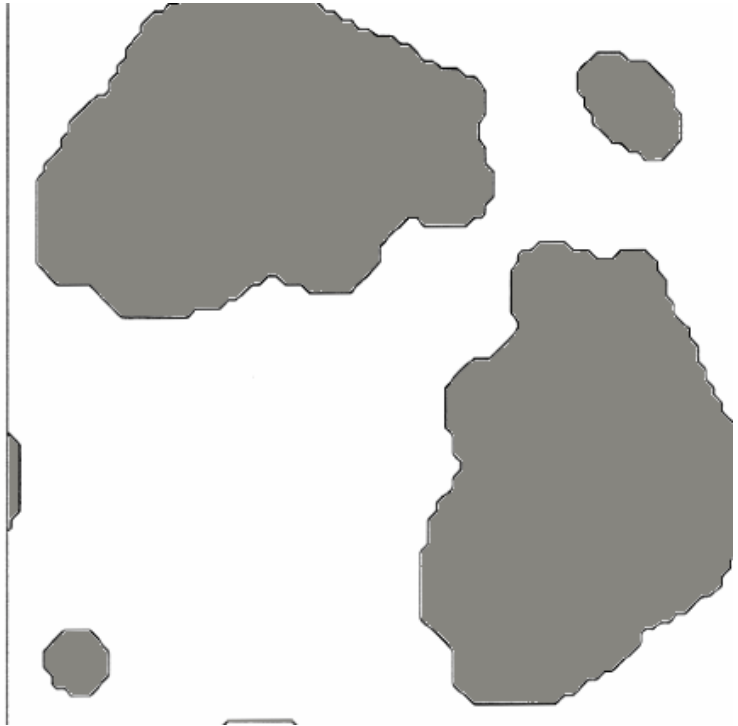
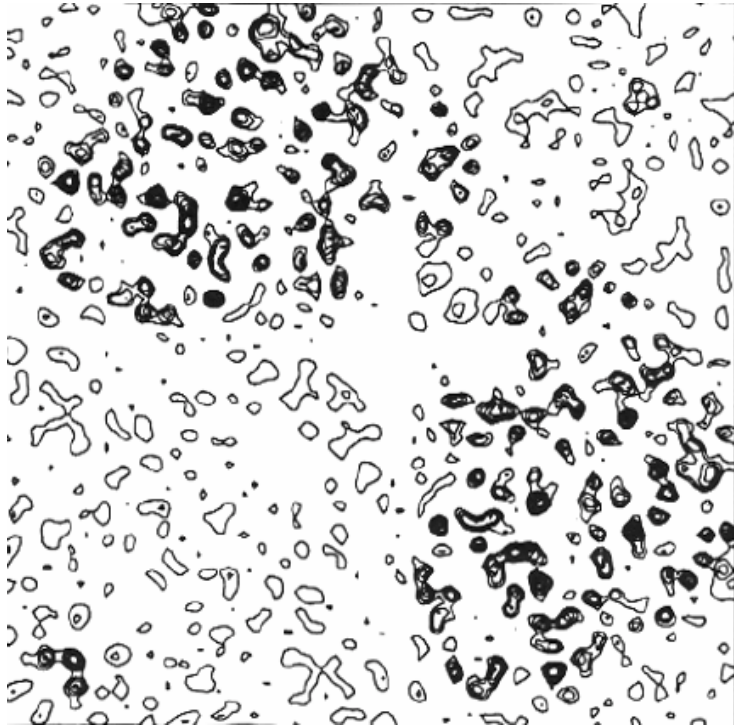
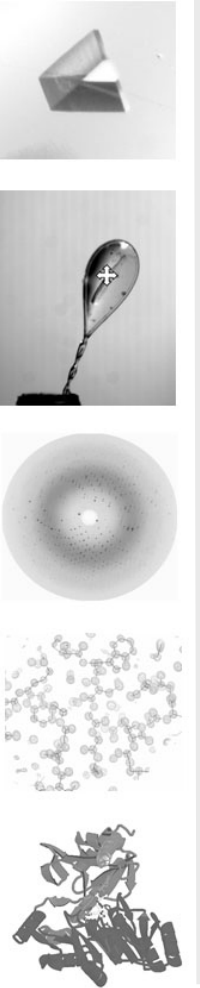
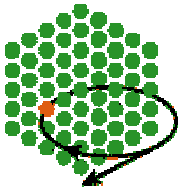
# Density Modification

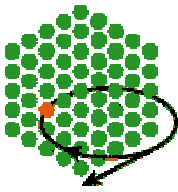
Density Modification is a technique that uses additional information to improve an electron density

- solvent flattening
- non-crystallographic symmetry averaging
- histogram matching
- map interpretation and refinement

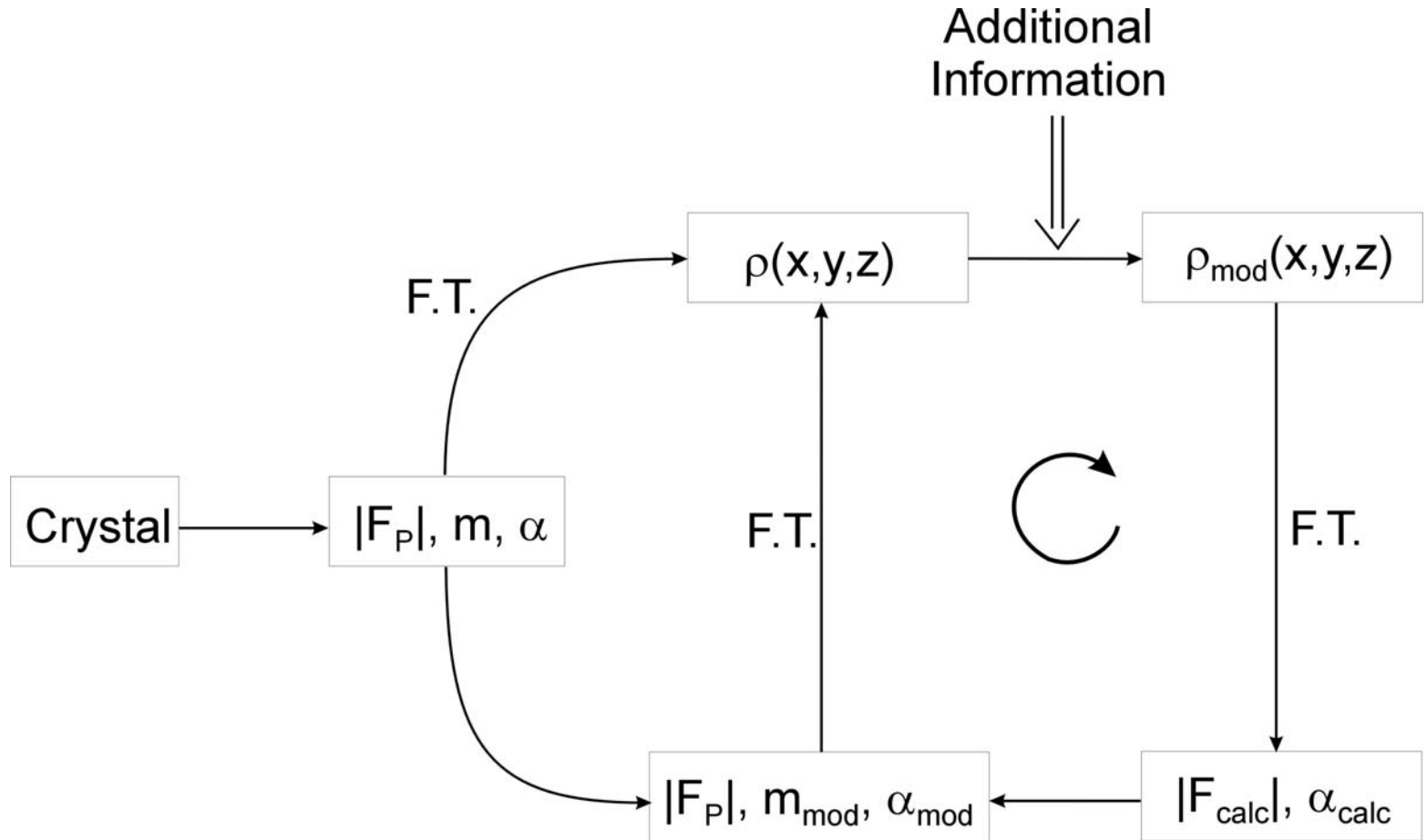
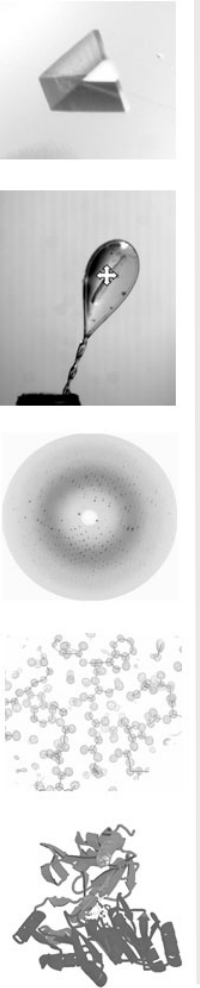


# Solvent Flattening

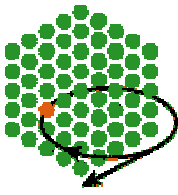




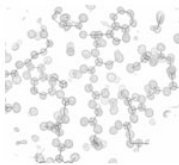
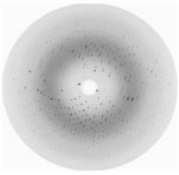
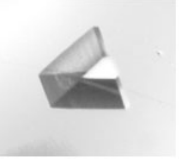
# Density Modification







# The Harker Construction - SIR

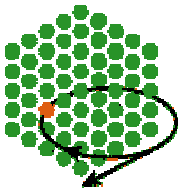


Instead of one value for  $\alpha_p(hkl)$  we obtain two possibilities  
→ *phase ambiguity*

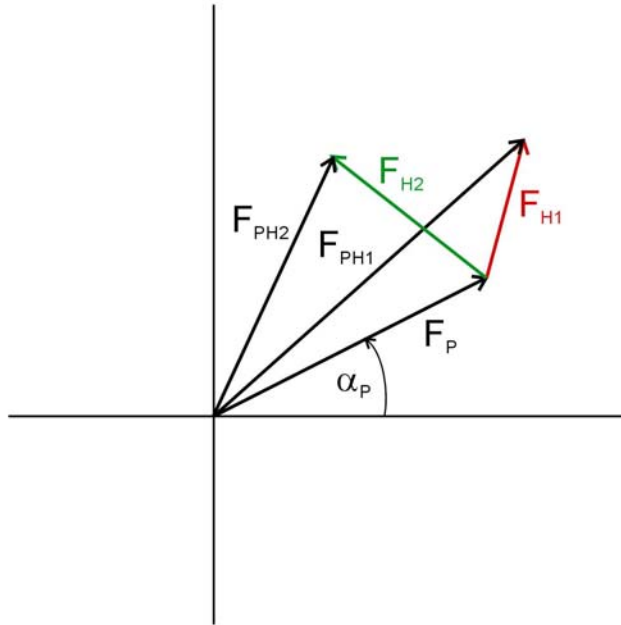
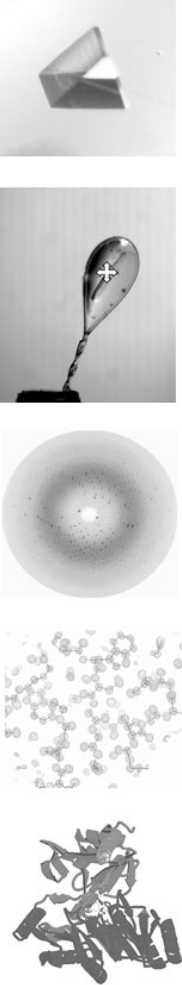
How can this be made unambiguous ?

→ density modification

→ a second heavy atom derivative (MIR)

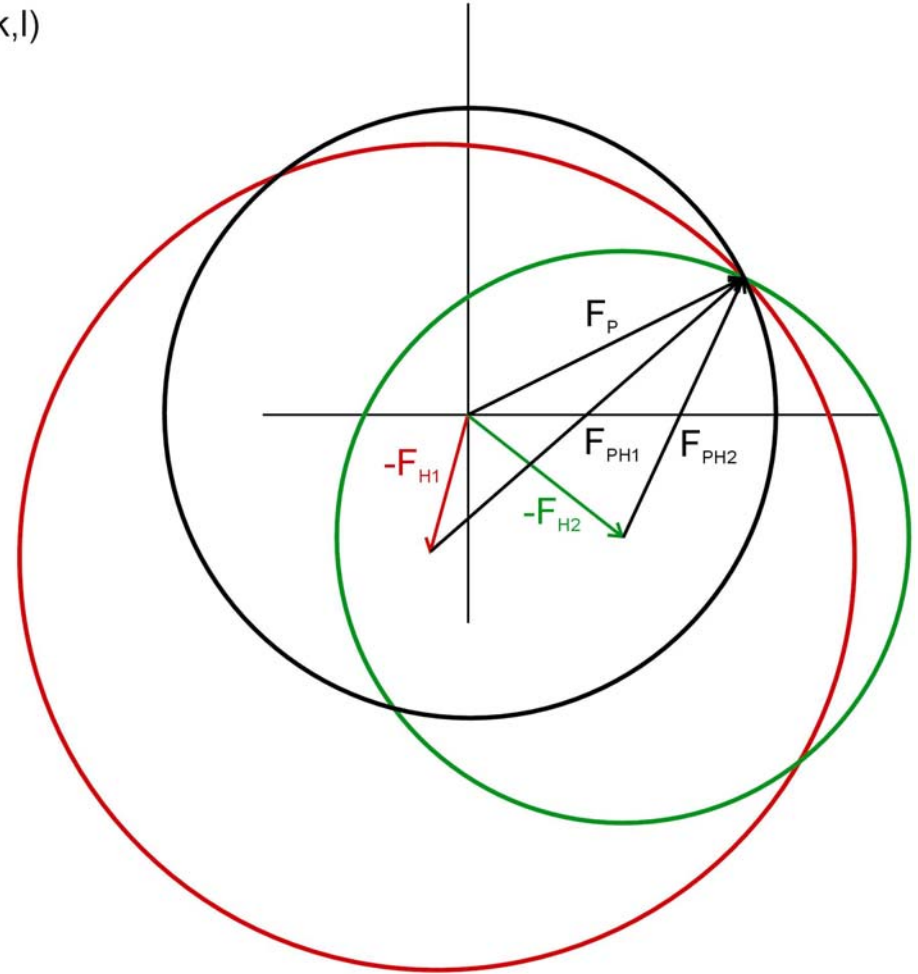


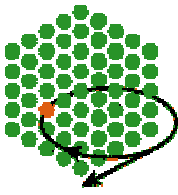
# The Harker Construction - MIR



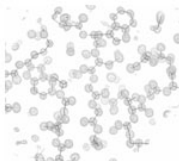
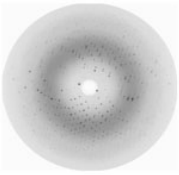
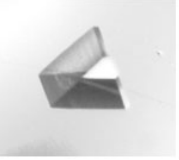
$$F_P + F_{H1} = F_{PH1}$$

$$F_P + F_{H2} = F_{PH2}$$





# The Harker Construction - SIR



Instead of one value for  $\alpha_p(hkl)$  we obtain two possibilities  
→ *phase ambiguity*

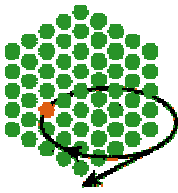
How can this be made unambiguous ?

→ density modification

→ a second heavy atom derivative (MIR)

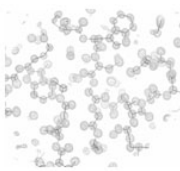
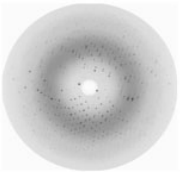
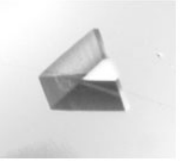
→ incorporation of anomalous scattering (SIRAS)

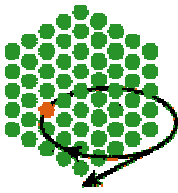
→ both ... (MIRAS)



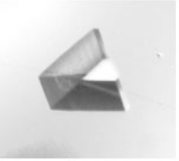
# Experiments

1. conventional heavy-atom derivatization by soaking or co-crystallization





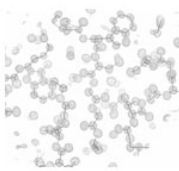
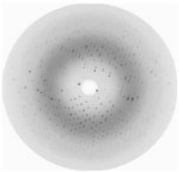
# Experiments

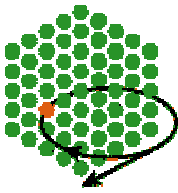


1. conventional heavy-atom derivatization by soaking or co-crystallization

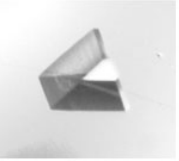


2. quick-soaking





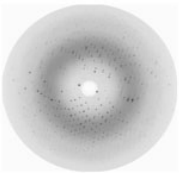
# Experiments



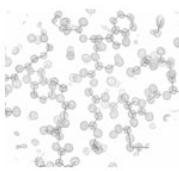
1. conventional heavy-atom derivatization by soaking or co-crystallization

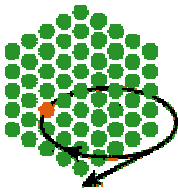


2. quick-soaking

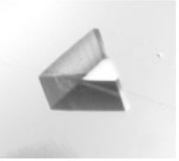


3. quick-soaking using anions (  $\text{Br}^-$ ,  $\text{I}^-$ ,  $\text{I}^-/\text{I}_2$ , ... )





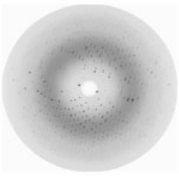
# Experiments



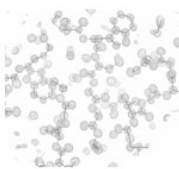
1. conventional heavy-atom derivatization by soaking or co-crystallization



2. quick-soaking

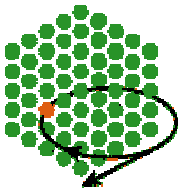


3. quick-soaking using anions (  $\text{Br}^-$ ,  $\text{I}^-$ ,  $\text{I}^-/\text{I}_2$ , ... )

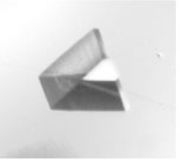


4. pressurization using noble gases ( Xe, Kr )





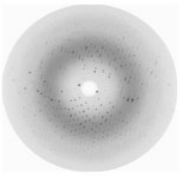
# Experiments



1. conventional heavy-atom derivatization by soaking or co-crystallization

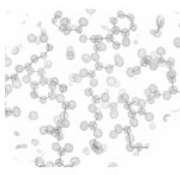


2. quick-soaking



3. quick-soaking using anions (  $\text{Br}^-$ ,  $\text{I}^-$ ,  $\text{I}^-/\text{I}_2$ , ... )

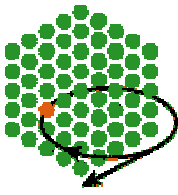
4. pressurization using noble gases ( Xe, Kr )



5. covalent modification of the protein (e.g. Met  $\rightarrow$  Se-Met)  
or of DNA (T  $\rightarrow$  Br-U)







# Phase determination methods

## 1. SIR, SIRAS, MIR, MIRAS

(single/multiple isomorphous replacement with anomalous scattering)

## 2. MAD

(multiple wavelength anomalous diffraction)

## 3. SAD (SAS)

(single wavelength anomalous diffraction/scattering)

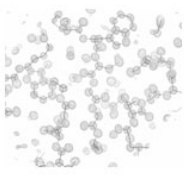
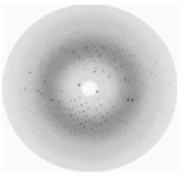
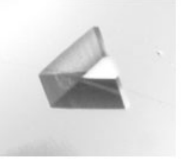
## 4. RIP, RIPAS

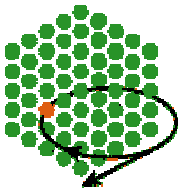
(radiation damage induced phasing with anomalous scattering)

## 5. MR

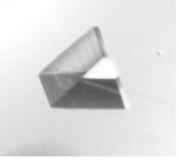
(molecular replacement)

## 6. Direct Methods





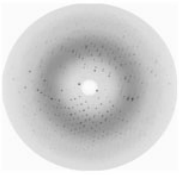
# The MR method



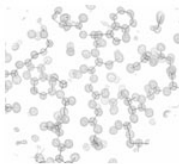
Molecular Replacement can be used to solve a structure when a homologous structure is available.

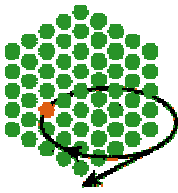


When the homologous structure can be correctly oriented and positioned (6 parameters), it can be used as a starting point for phase calculation.



Nowadays, about 2/3 of all structures are solved by Molecular Replacement.





# The Patterson Function

Introduced by and named after L. Patterson in 1934.

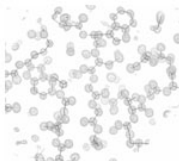
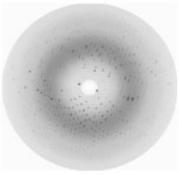
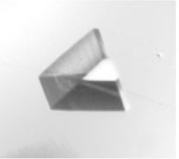
$$P(u,v,w) = 1/V \sum_{hkl} |F(hkl)|^2 \cos 2\pi(hu+kv+lw)$$

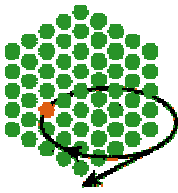
$P(u,v,w)$  can be calculated without phases.

It can be shown that  $P(u,v,w)$  is equivalent to a convolution of  $\rho(x,y,z)$  with itself.

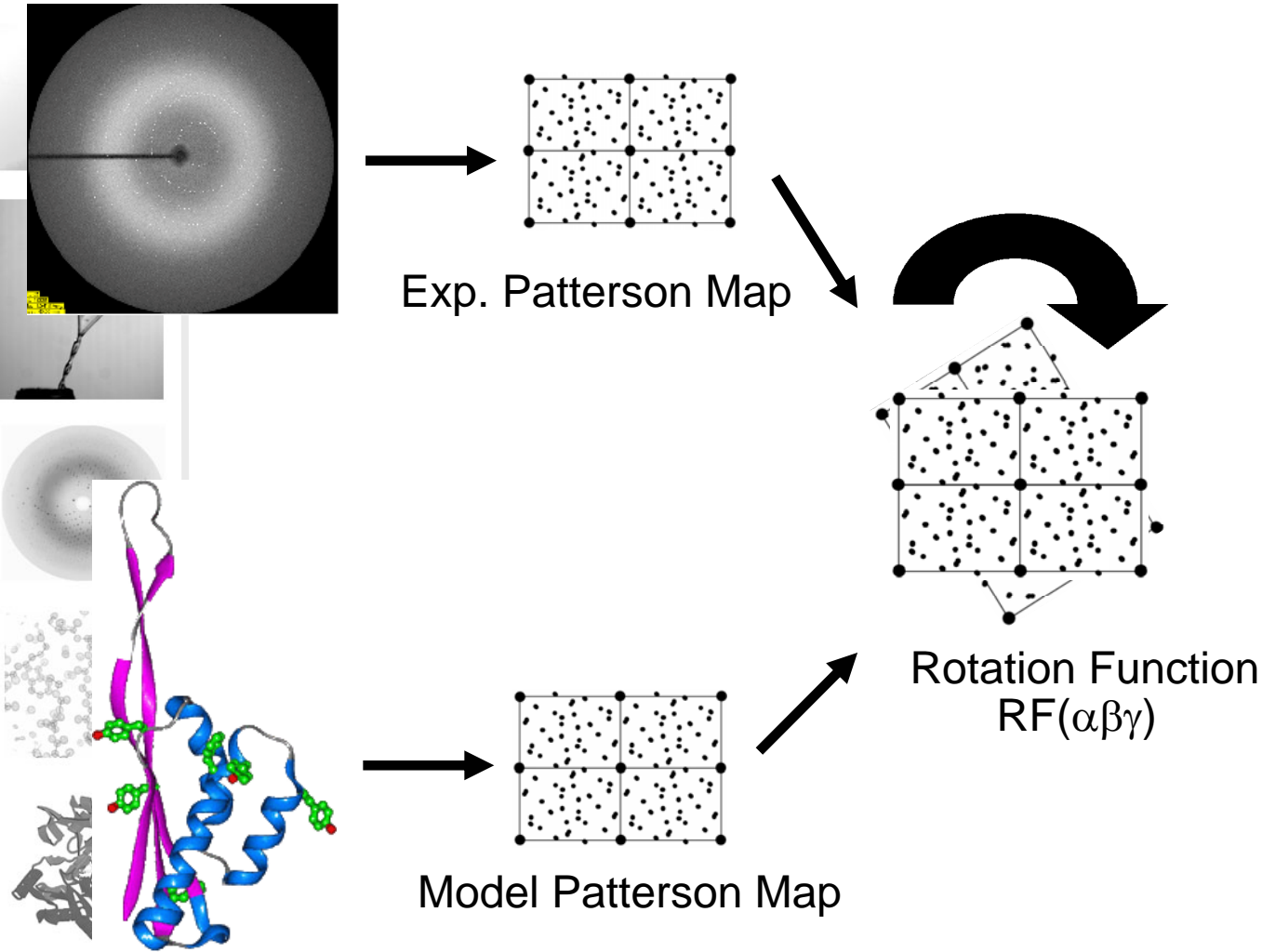
$$P(u,v,w) = 1/V \int \rho(x,y,z) \rho(x+u,y+v,z+w) dx dy dz$$

$P(u,v,w)$  can also be calculated from a known structure.

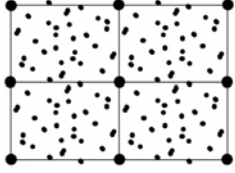
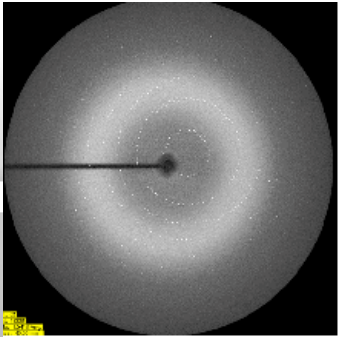
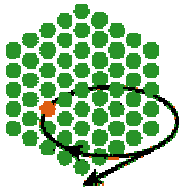




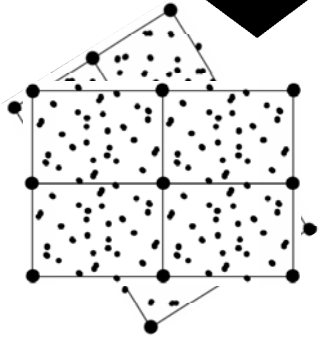
# Molecular Replacement



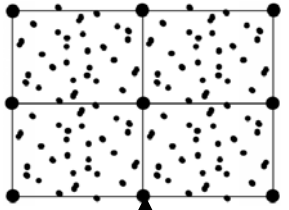
# Molecular Replacement



Exp. Patterson Map

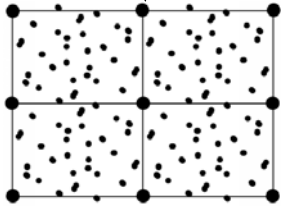


Rotation Function  $RF(\alpha\beta\gamma)$

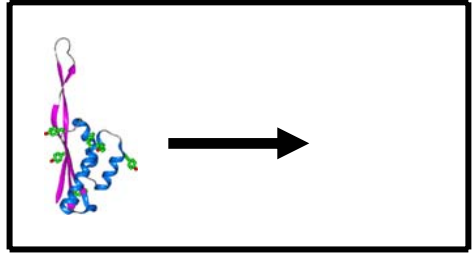


Model Patterson

$TF(xyz)$



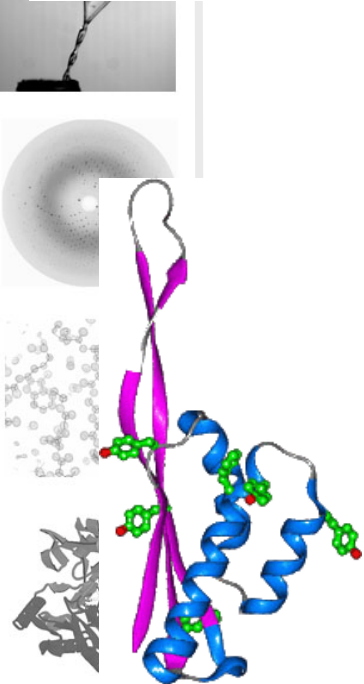
Exp. Patterson

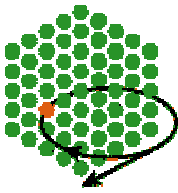


Translation search



Model Patterson Map

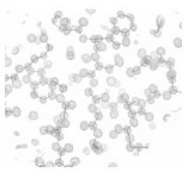
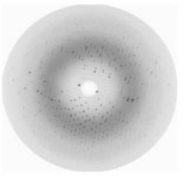
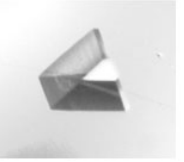


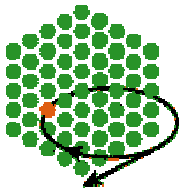


# Molecular Replacement - Difficulties

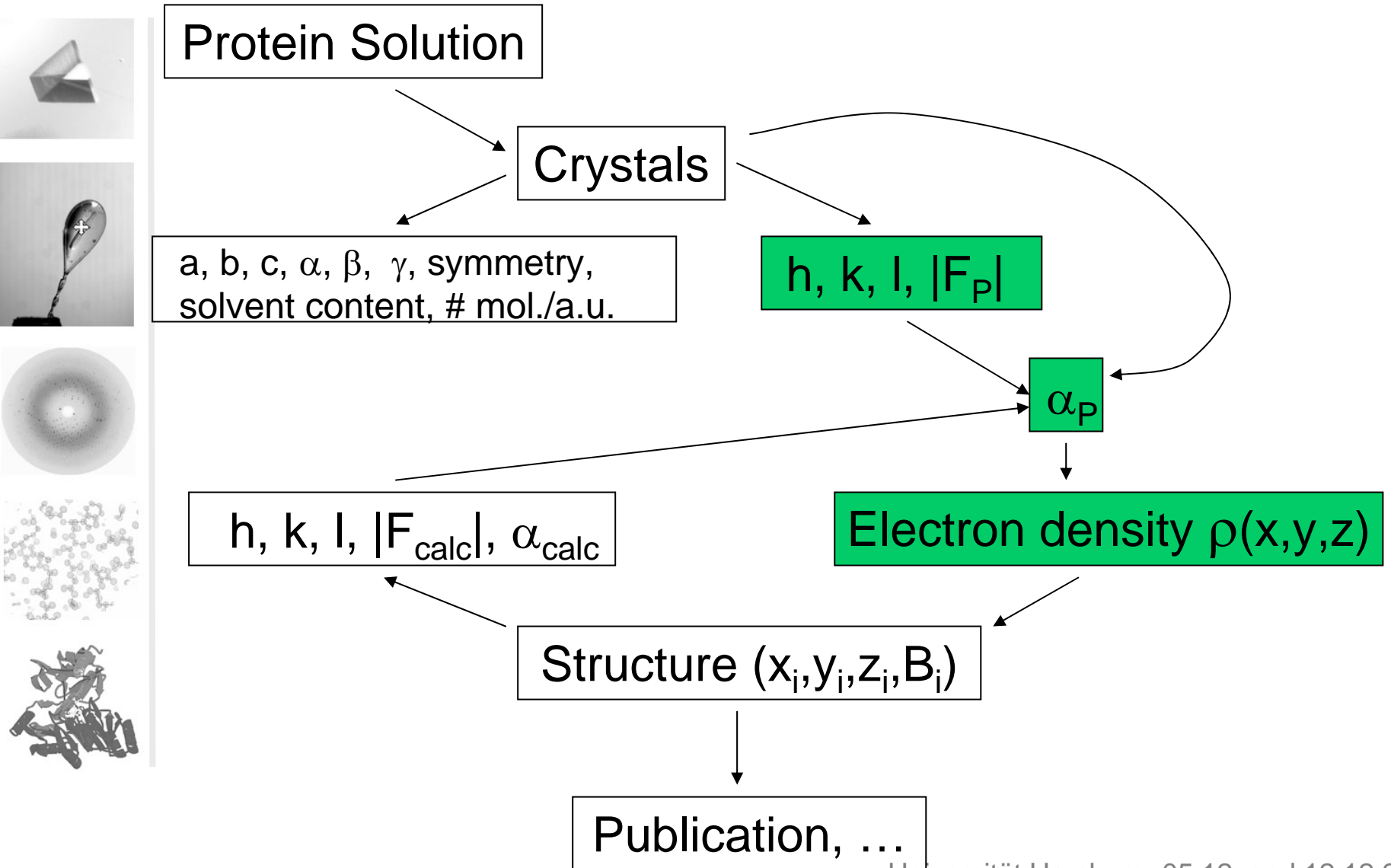
The search model is not the same as the target structure. Consequently, the respective Patterson functions will not be identical.

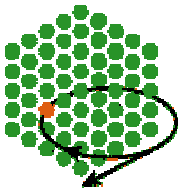
$P(u,v,w)$  from the model contains only intramolecular vectors, whereas  $P(u,v,w)$  from the diffraction data contains intra- and intermolecular vectors.





# Calculation of the Electron Density





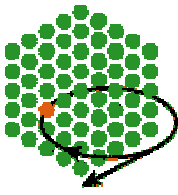
# The Electron Density Equation

Once the phases  $\alpha(hkl)$  have been determined, the three-dimensional electron density function  $\rho(x,y,z)$  can be calculated.

$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

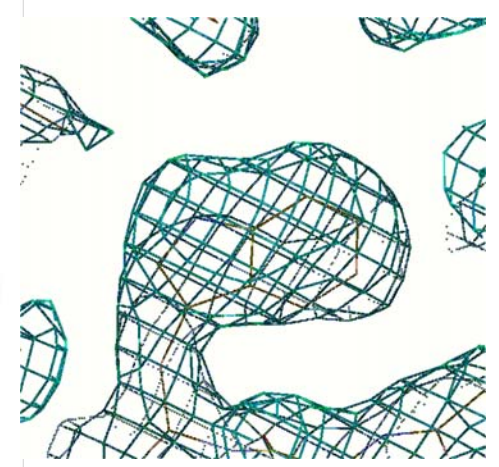
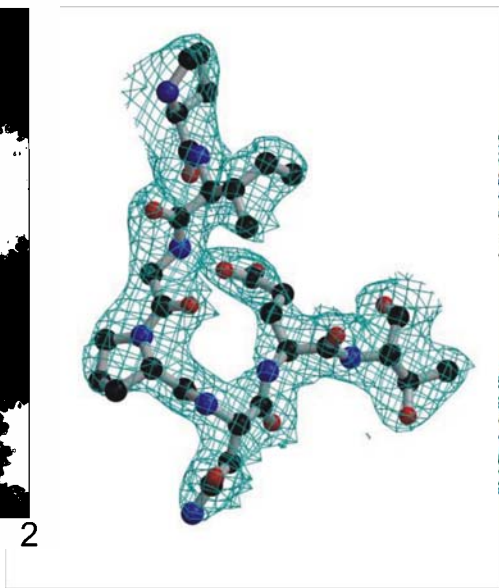
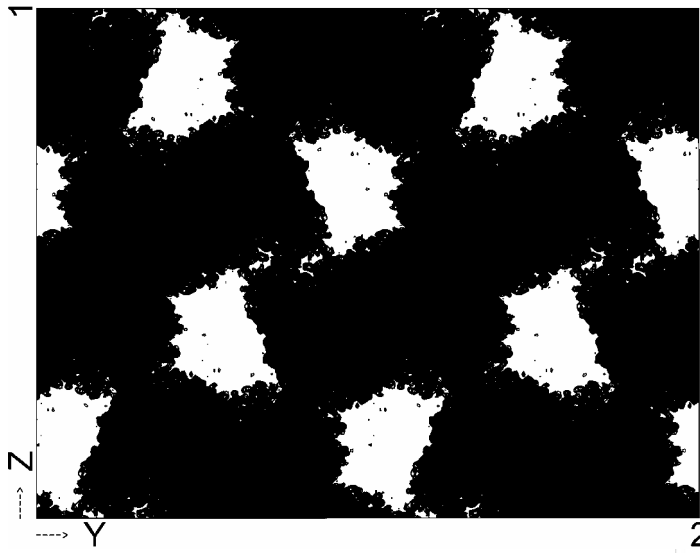
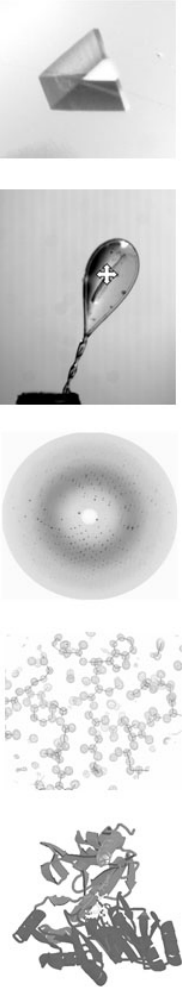
The electron density  $\rho(x,y,z)$  is a three-dimensional function (with the unit  $e/\text{\AA}^3$ ), which describes where in the unit cell of the crystal the electrons (and therefore the atoms) are. It is basically the image of the structure we want to determine.

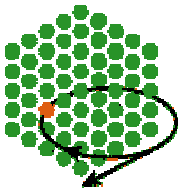




# Electron Density

All that is left to do now is to interpret the electron density in terms of the three-dimensional structure.

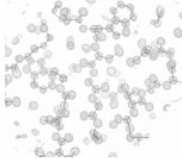
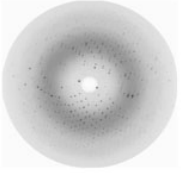
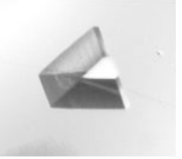


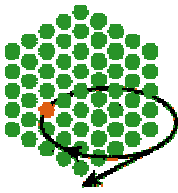


# Electron Density Maps

$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

Maps	Amplitudes	Phases	Weight(s)
$F_o$ -map	$ mF_{obs} $	$\alpha_{obs}$	f.o.m.

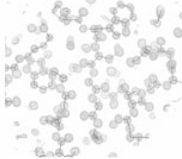
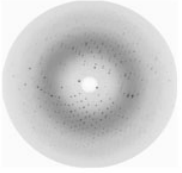
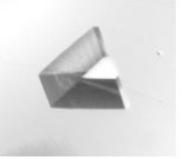


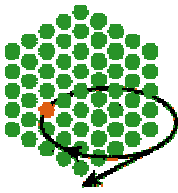


# Electron Density Maps

$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

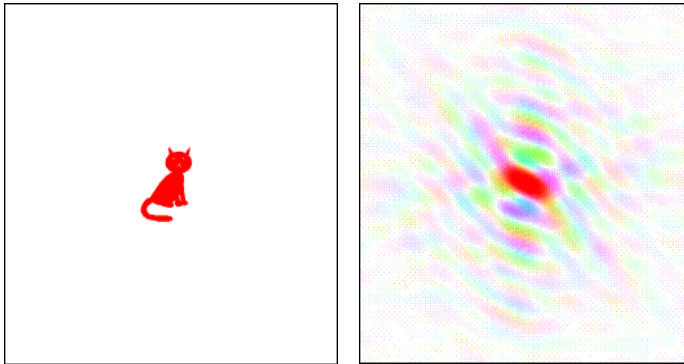
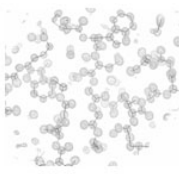
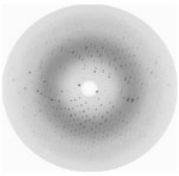
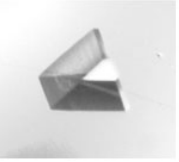
Maps	Amplitudes	Phases	Weight(s)
$F_o$ -map	$ mF_{obs} $	$\alpha_{obs}$	f.o.m.
$(F_o - F_c)$ -map	$ F_{obs} - F_{calc} $	$\alpha_{calc}$	-
$(2F_o - F_c)$ -map	$ 2F_{obs} - F_{calc} $	$\alpha_{calc}$	-
SigmaA-map	$ 2mF_{obs} - DF_{calc} $	$\alpha_{calc}$	f.o.m., D



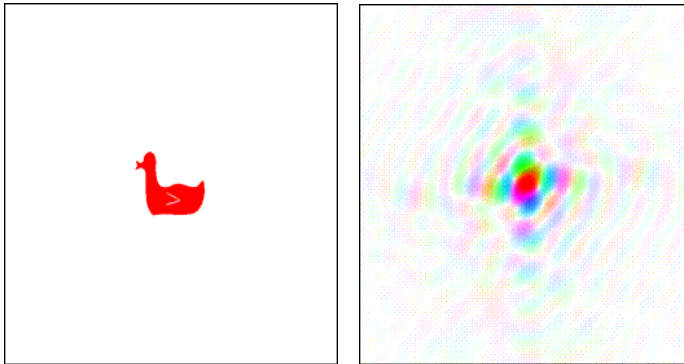
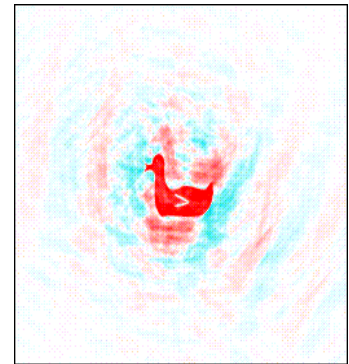


# Amplitudes vs. Phases

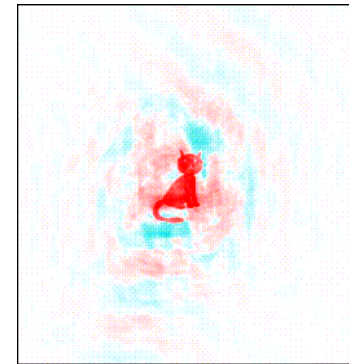
$$\rho(x,y,z) = 1/V \sum_{hkl} |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

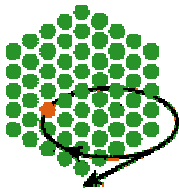


$F_{\text{Cat}}, \alpha_{\text{Duck}}$   
→



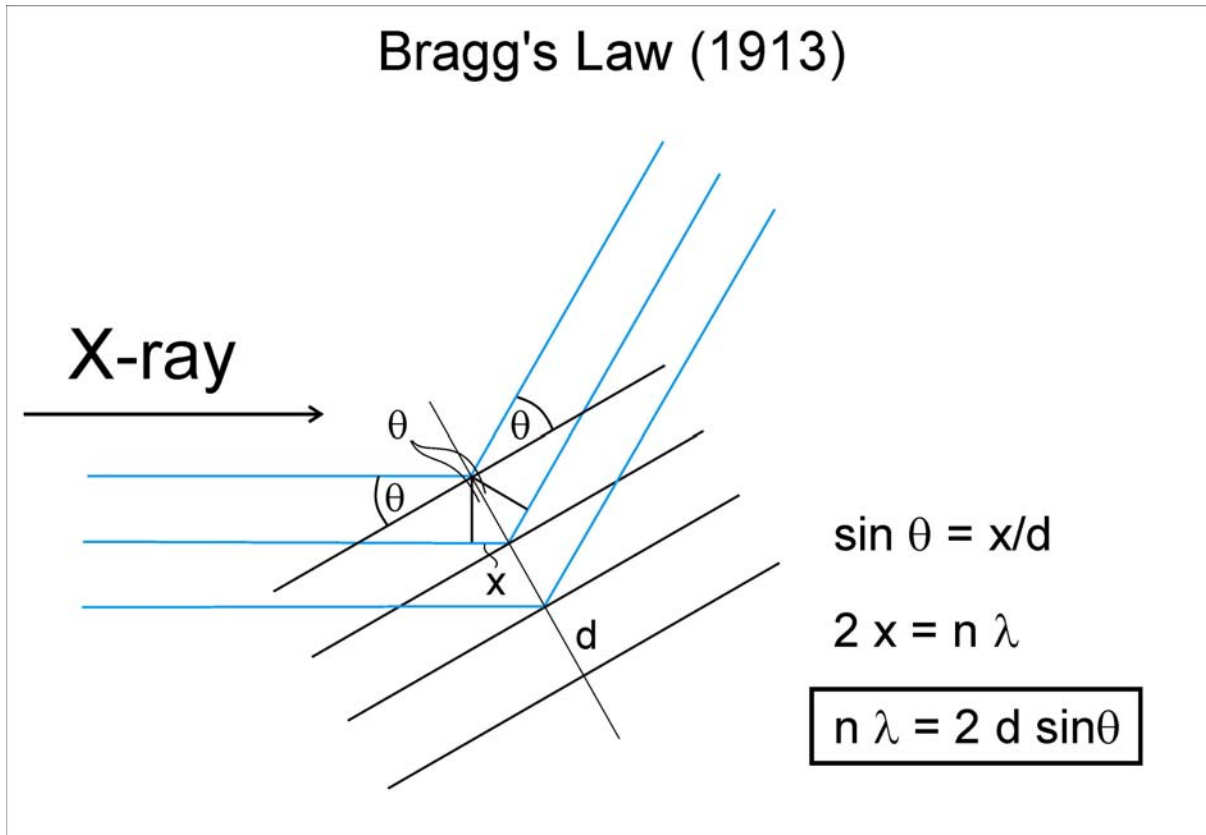
$F_{\text{Duck}}, \alpha_{\text{Cat}}$   
→

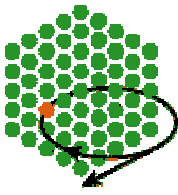




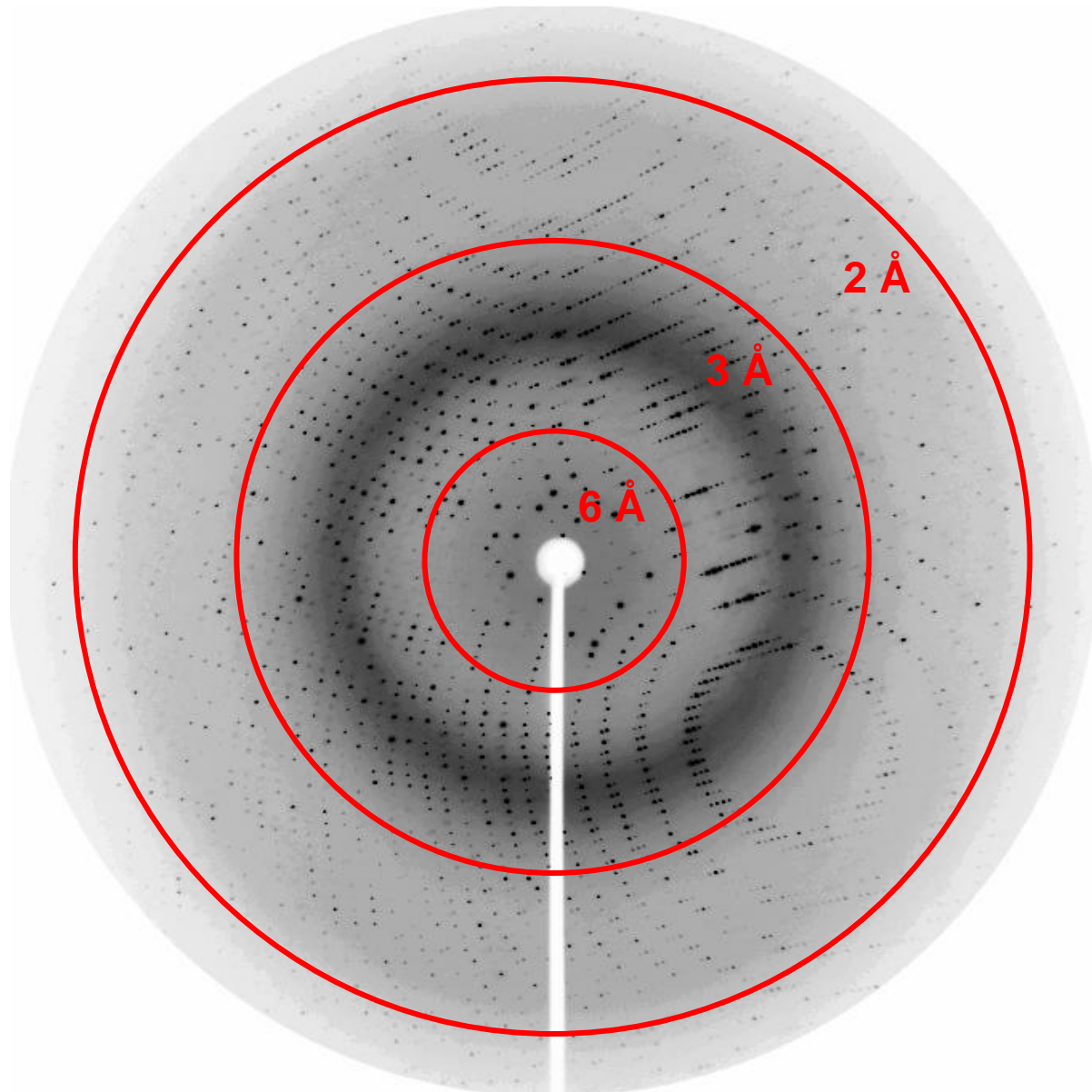
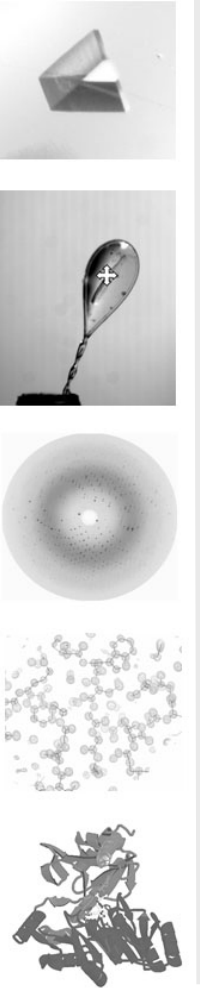
# Information Content and Resolution

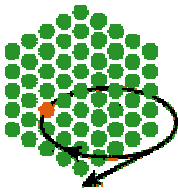
In crystallography, the resolution  $d$  is typically defined by the maximum angle  $2\theta$ , to which diffraction can be observed.



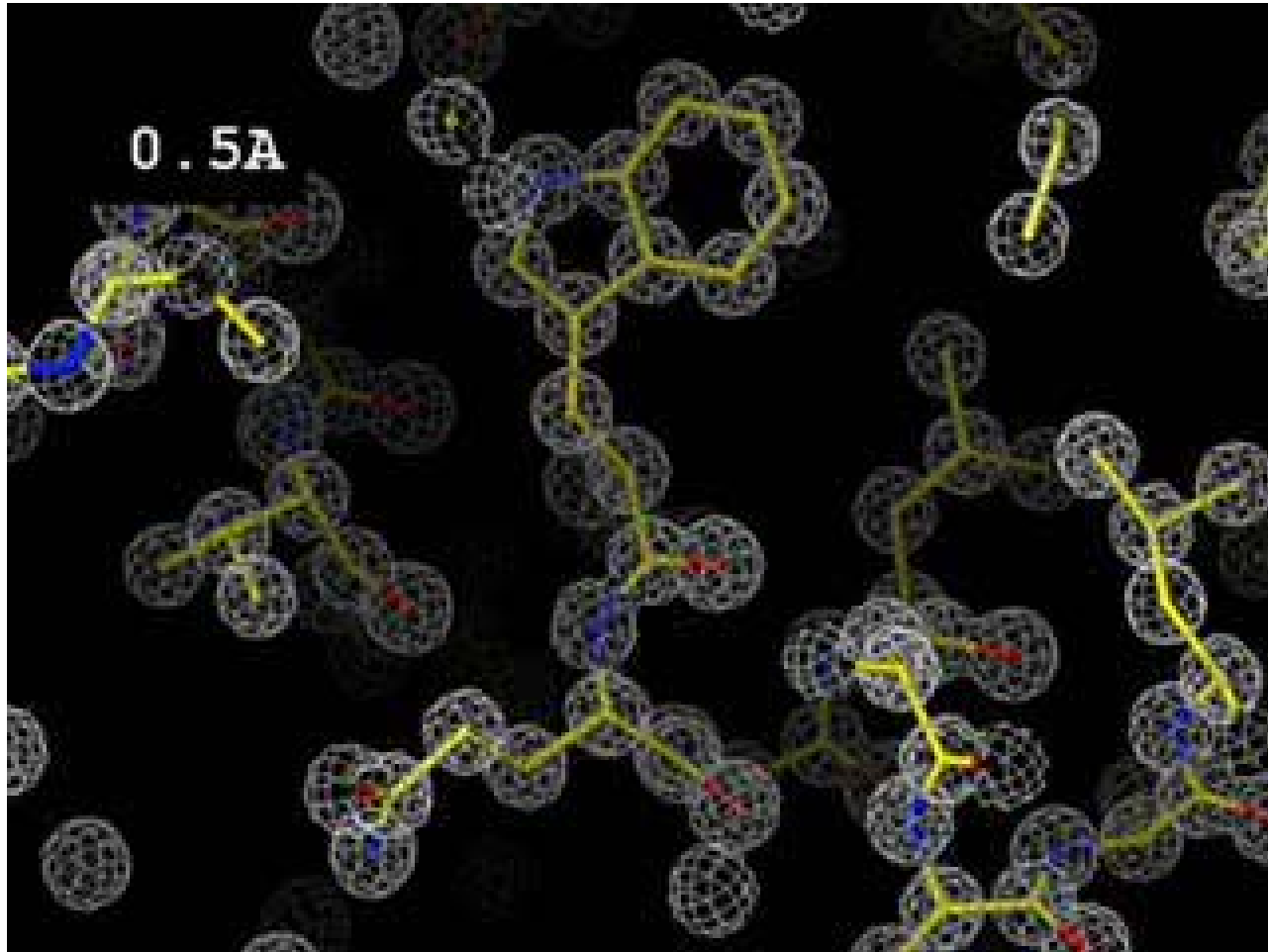
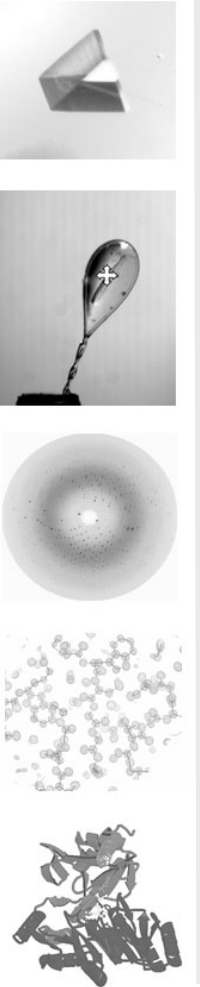


# Information Content and Resolution

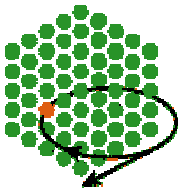




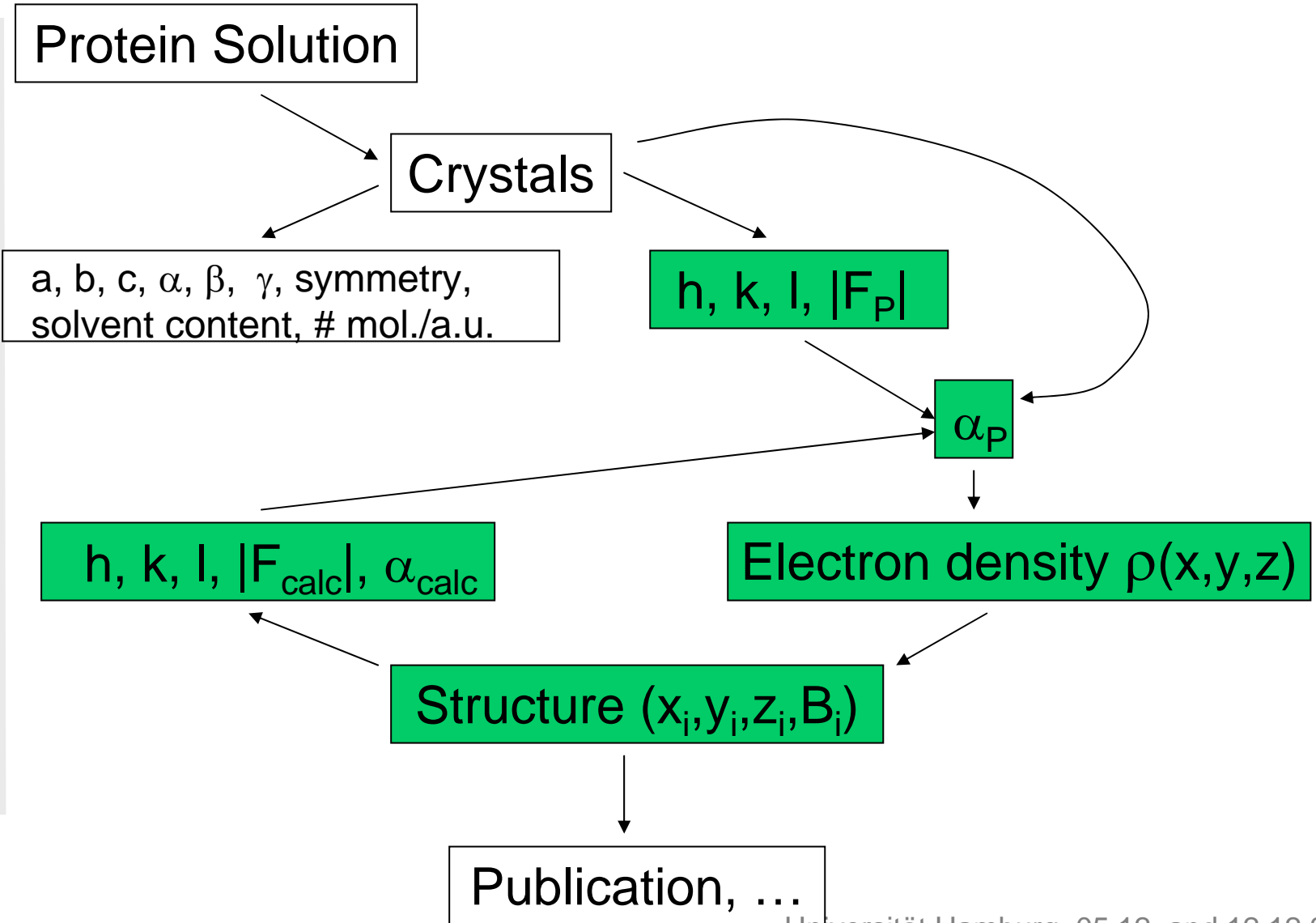
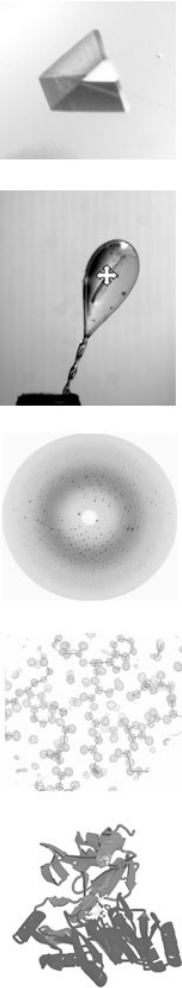
# Information Content and Resolution



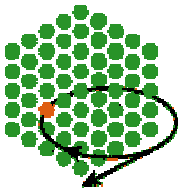
Source: <http://ucxray.berkeley.edu/~jamesh/movies/>



# Structure Refinement





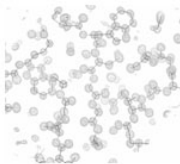
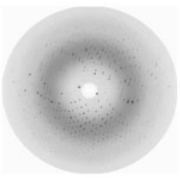
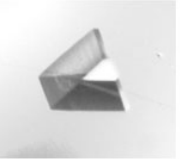


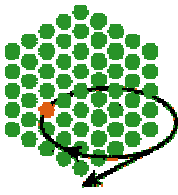
# Refinement and the R-factor

Refinement is a mathematical procedure, which iteratively improves the fit between the **experimental diffraction data** ( $F_{obs}$ ) and the theoretical diffraction data, which can be calculated from the **structural model** ( $F_{calc}$ ) at any given stage.

Mathematically, refinement is the minimization of the following function:

$$f(x_j, y_j, z_j, B_j) = \sum_{hkl} w_{hkl} (|F_{obs}(hkl)| - |F_{calc}(hkl)|)^2$$



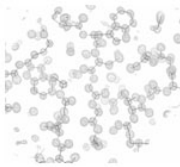
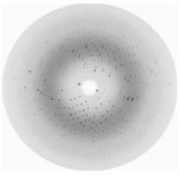
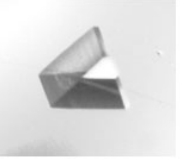


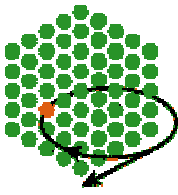
# Refinement and the R-factor

Refinement progress can be monitored using the so-called crystallographic R-factor:

$$R = 100 \cdot \frac{\sum_{hkl} | |F_{obs}| - |F_{calc}| |}{\sum_{hkl} |F_{obs}|}$$

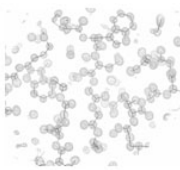
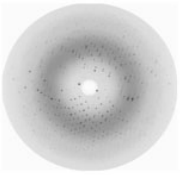
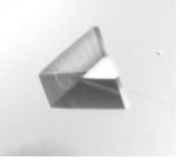
This constant feedback between the experimental data ( $F_{obs}$ ) and the model ( $F_{calc}$ ) is one of the greatest strengths of crystallography.

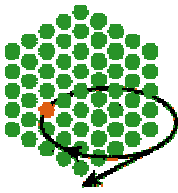




# R and $R_{\text{free}}$

However ...

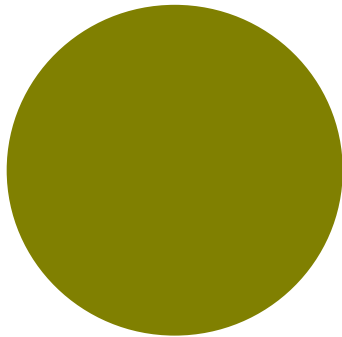




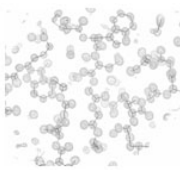
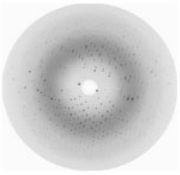
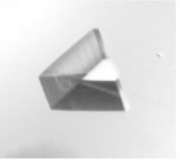
# R and $R_{\text{free}}$

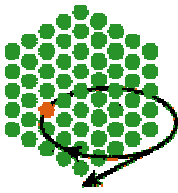
However, the target function minimized during refinement and the progress monitor have a very similar form. Anything that will minimize  $f(x_j, y_j, z_j, B_j)$ , will also lower the R-factor. Therefore, the R-factor is not a good measure for refinement progress, because it is not independent.

In 1992, the free R-factor was introduced.



all (hkl)

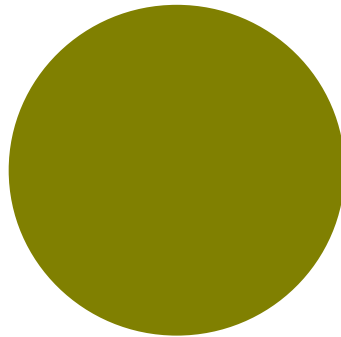




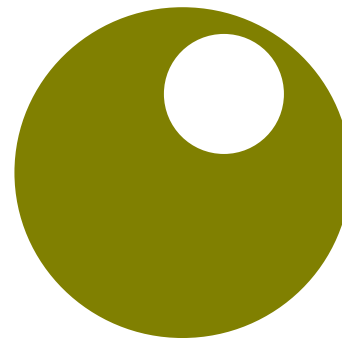
# R and $R_{\text{free}}$

However, the target function minimized during refinement and the progress monitor have a very similar form. Anything that will minimize  $f(x_j, y_j, z_j, B_j)$ , will also lower the R-factor. Therefore, the R-factor is not a good measure for refinement progress, because it is not independent.

In 1992, the free R-factor was introduced.



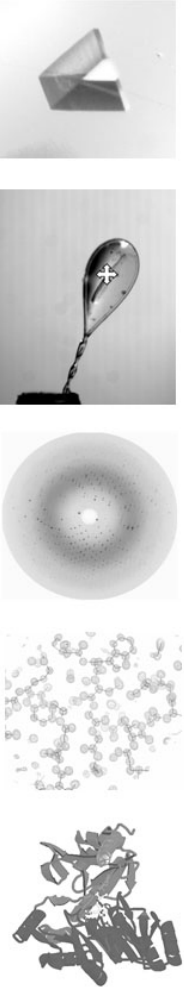
all (hkl)

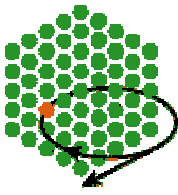


working set (hkl)



test set (hkl)





# Parameters vs. Observables

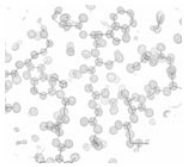
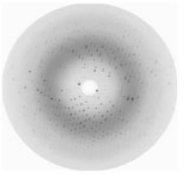
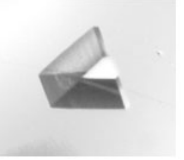
Example: A protein of 182 amino acids, 1458 atoms, 40% solvent content.

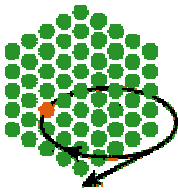
Resolution	No. of (hkl)
3.0 Å	3,500
2.5 Å	6,800
1.9 Å	13,500
1.5 Å	29,800
1.2 Å	58,800
1.0 Å	81,300

$x_j, y_j, z_j$ : 4,374 parameters

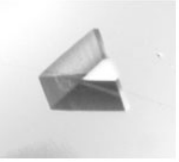
$B_{j,iso}$ : 1,458 parameters

$B_{j,aniso}$ : 8,748 parameters





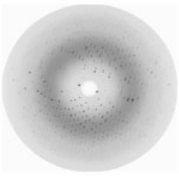
# Constraints and Restraints



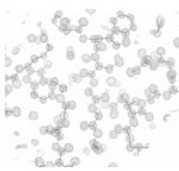
Constraints are boundary conditions, i.e. pieces of information, which reduce the number of parameters.



*E.g. all atoms of one amino acid have the same  $B_{iso}$*



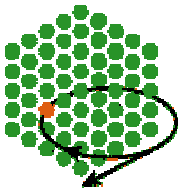
Restraints are boundary conditions, i.e. pieces of information, which increase the number of observables.



*E.g. bond lengths and bond angles should be close to the very accurate values, which can be obtained from high resolution peptide structures.*



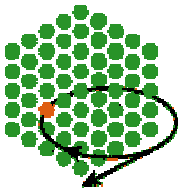
This adds the term  $g(x_j, y_j, z_j) = \sum_j w (d_{obs} - d_{ideal})^2$  to the target function for refinement.



# Information Content of a Structure

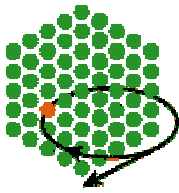
```
HEADER          HEME PROTEIN                      25-FEB-98      1A6G
TITLE          CARBONMONOXY-MYOGLOBIN, ATOMIC RESOLUTION
EXPDTA        X-RAY DIFFRACTION
...
ATOM           1  N   VAL A   1          -4.040  15.048  13.602  1.00  37.04      N
ATOM           2  CA  VAL A   1          -3.621  15.574  14.908  1.00  25.44      C
ATOM           3  C   VAL A   1          -2.766  14.564  15.637  1.00  24.20      C
ATOM           4  O   VAL A   1          -3.155  13.399  15.760  1.00  27.65      O
ATOM           5  CB  VAL A   1          -4.863  15.869  15.757  1.00  32.48      C
ATOM           6  CG1 VAL A   1          -4.402  15.954  17.205  1.00  38.33      C
ATOM           7  CG2 VAL A   1          -5.600  17.125  15.353  1.00  45.60      C
ATOM           8  N   LEU A   2          -1.588  14.922  16.125  1.00  17.99      N
ATOM           9  CA  LEU A   2          -0.767  13.985  16.878  1.00  16.85      C
ATOM          10  C   LEU A   2          -1.326  13.777  18.289  1.00  16.94      C
ATOM          11  O   LEU A   2          -1.795  14.765  18.873  1.00  19.44      O
ATOM          12  CB  LEU A   2           0.694  14.436  17.007  1.00  15.72      C
ATOM          13  CG  LEU A   2           1.598  14.164  15.773  1.00  14.99      C
ATOM          14  CD1 LEU A   2           1.269  15.130  14.656  1.00  15.12      C
ATOM          15  CD2 LEU A   2           3.080  14.254  16.161  1.00  14.37      C
ATOM          16  N   SER A   3          -1.218  12.546  18.763  1.00  16.86      N
...
...
END
```





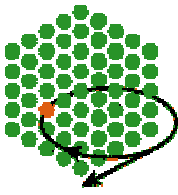
# Information Content of a Structure

```
HEADER          HEME PROTEIN                      25-FEB-98      1A6G
TITLE          CARBONMONOXY-MYOGLOBIN, ATOMIC RESOLUTION
EXPDTA        X-RAY DIFFRACTION
...
ATOM          1  N   VAL A   1      -4.040   15.048   13.602   1.00   37.04      N
ATOM          2  CA  VAL A   1      -3.621   15.574   14.908   1.00   25.44      C
ATOM          3  C   VAL A   1      -2.766   14.564   15.637   1.00   24.20      C
ATOM          4  O   VAL A   1      -3.155   13.399   15.760   1.00   27.65      O
ATOM          5  CB  VAL A   1      -4.863   15.869   15.757   1.00   32.48      C
ATOM          6  CG1 VAL A   1      -4.402   15.954   17.205   1.00   38.33      C
ATOM          7  CG2 VAL A   1      -5.600   17.125   15.353   1.00   45.60      C
ATOM          8  N   LEU A   2      -1.588   14.922   16.125   1.00   17.99      N
ATOM          9  CA  LEU A   2      -0.767   13.985   16.878   1.00   16.85      C
ATOM         10  C   LEU A   2      -1.326   13.777   18.289   1.00   16.94      C
ATOM         11  O   LEU A   2      -1.795   14.765   18.873   1.00   19.44      O
ATOM         12  CB  LEU A   2          0.694   14.436   17.007   1.00   15.72      C
ATOM         13  CG  LEU A   2          1.598   14.164   15.773   1.00   14.99      C
ATOM         14  CD1 LEU A   2          1.269   15.130   14.656   1.00   15.12      C
ATOM         15  CD2 LEU A   2          3.080   14.254   16.161   1.00   14.37      C
ATOM         16  N   SER A   3      -1.218   12.546   18.763   1.00   16.86      N
...
...
END
```



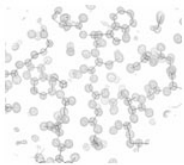
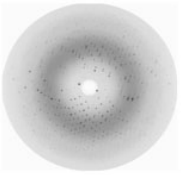
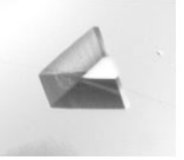
# Information Content of a Structure

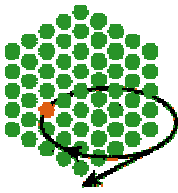
```
HEADER          HEME PROTEIN                      25-FEB-98      1A6G
TITLE          CARBONMONOXY-MYOGLOBIN, ATOMIC RESOLUTION
EXPDTA        X-RAY DIFFRACTION
...
ATOM          1  N   VAL  A   1      -4.040   15.048   13.602   1.00   37.04      N
ATOM          2  CA  VAL  A   1      -3.621   15.574   14.908   1.00   25.44      C
ATOM          3  C   VAL  A   1      -2.766   14.564   15.637   1.00   24.20      C
ATOM          4  O   VAL  A   1      -3.155   13.399   15.760   1.00   27.65      O
ATOM          5  CB  VAL  A   1      -4.863   15.869   15.757   1.00   32.48      C
ATOM          6  CG1 VAL  A   1      -4.402   15.954   17.205   1.00   38.33      C
ATOM          7  CG2 VAL  A   1      -5.600   17.125   15.353   1.00   45.60      C
ATOM          8  N   LEU  A   2      -1.588   14.922   16.125   1.00   17.99      N
ATOM          9  CA  LEU  A   2      -0.767   13.985   16.878   1.00   16.85      C
ATOM         10  C   LEU  A   2      -1.326   13.777   18.289   1.00   16.94      C
ATOM         11  O   LEU  A   2      -1.795   14.765   18.873   1.00   19.44      O
ATOM         12  CB  LEU  A   2         0.694   14.436   17.007   1.00   15.72      C
ATOM         13  CG  LEU  A   2         1.598   14.164   15.773   1.00   14.99      C
ATOM         14  CD1 LEU  A   2         1.269   15.130   14.656   1.00   15.12      C
ATOM         15  CD2 LEU  A   2         3.080   14.254   16.161   1.00   14.37      C
ATOM         16  N   SER  A   3      -1.218   12.546   18.763   1.00   16.86      N
...
...
END
```



# Information Content of a Structure

- A crystal structure constitutes a space- and time-averaged structure.
- The occupancy value reflects in how many unit cells a certain atom is in the position indicated by its coordinates.
- The atomic displacement parameter (temperature factor, B-factor) describes the smearing of the electron cloud of the atom around its equilibrium position.
- Hydrogen atoms are typically not seen in an X-ray structure, but most of their positions can be inferred.

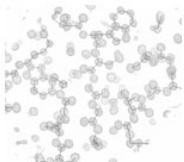
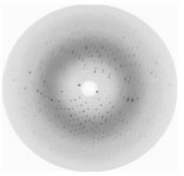
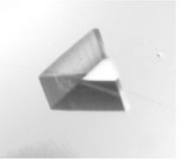


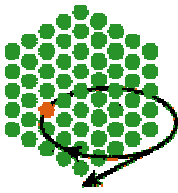


# Information Content of a Structure

- Depending on the resolution of the underlying data, and on the care which has been taken to interpret the data, there may be various uncertainties and errors in the structure.

- *coordinate errors (typically 0.2 Å for a 2.0 Å resolution structure)*
- *misinterpreted or overlooked ligands or water molecules*
- *side chain orientations (Asn-, His-, Gln-flips)*
- *incorrect side chain conformers*
- *peptide flips*
- *peptide bond conformation (cis or trans)*
- *sequence registration errors*
- *errors in the chain trace*

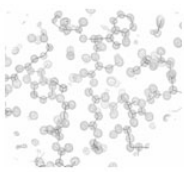
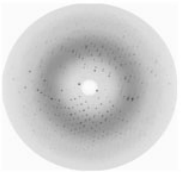
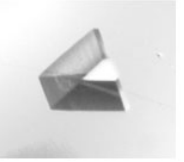


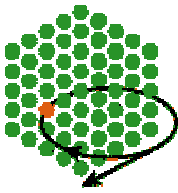


# Structure Validation

- Refinement  $R$  and  $R_{\text{free}}$

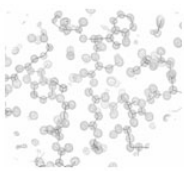
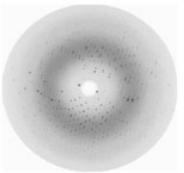
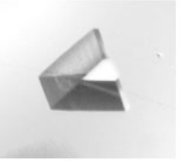
- *A well refined structure should have  $R$  and  $R_{\text{free}}$ -values of about or less than 20%.*
- *The higher the resolution, the smaller  $R$  and  $R_{\text{free}}$  should be.*
- *The higher the resolution, the smaller the gap between  $R$  and  $R_{\text{free}}$  should be.*

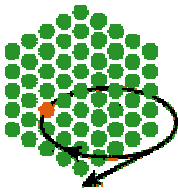




# Structure Validation

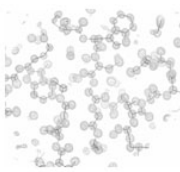
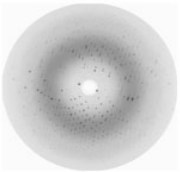
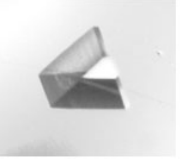
- Refinement R and  $R_{\text{free}}$
  - Fit to electron density map
- 
- *Electron density maps for all PDB Structures can be downloaded from the Electron Density Server at Uppsala University (<http://eds.bmc.uu.se/eds/>)*

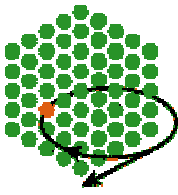




# Structure Validation

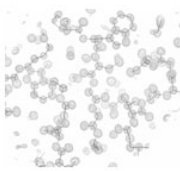
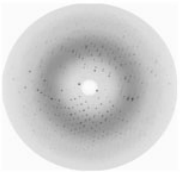
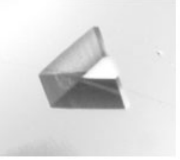
- Refinement  $R$  and  $R_{\text{free}}$
- Fit to electron density map
- Residual peaks in difference electron density map



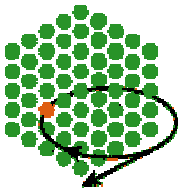


# Structure Validation

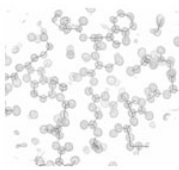
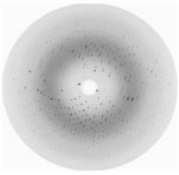
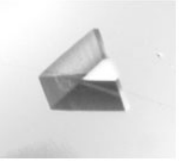
- Refinement  $R$  and  $R_{\text{free}}$
- Fit to electron density map
- Residual peaks in difference electron density map
- Omit electron density for ligands



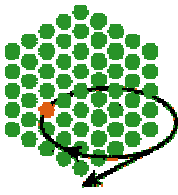




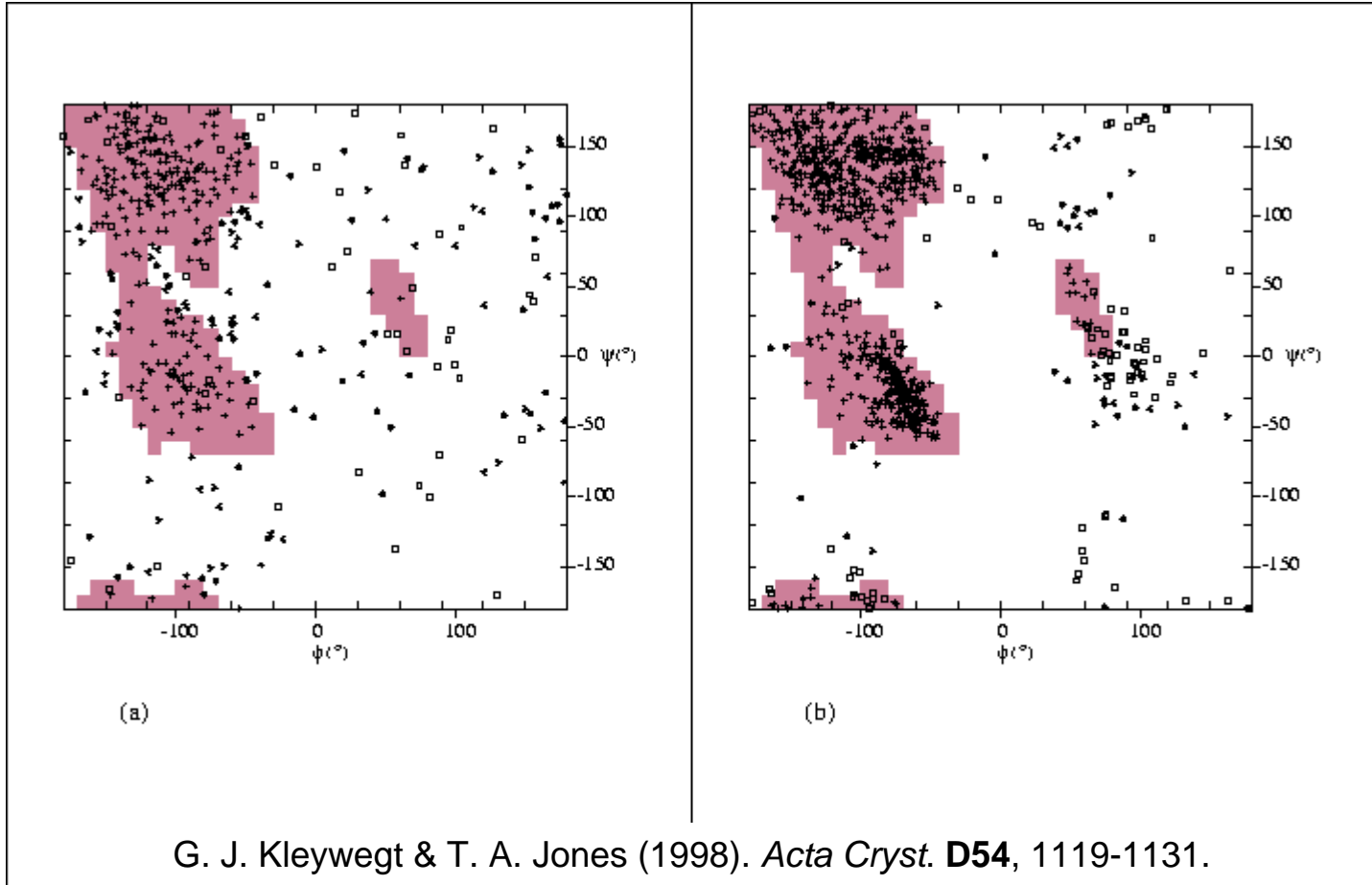
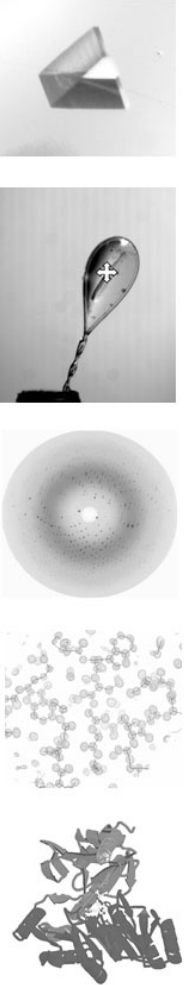
# Structure Validation

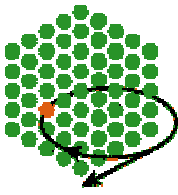


- Refinement R and  $R_{\text{free}}$
- Fit to electron density map
- Residual peaks in difference electron density map
- Omit electron density for ligands
- Ramachandran plot

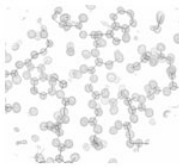
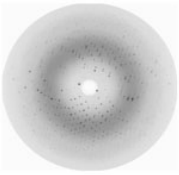
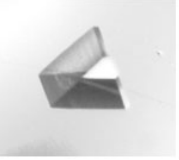


# Ramachandran Plot

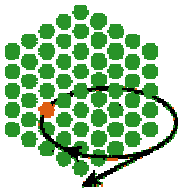




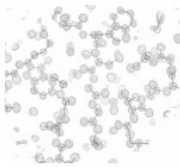
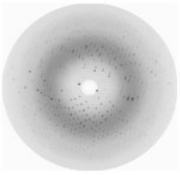
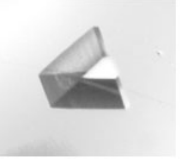
# Structure Validation



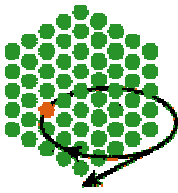
- Refinement  $R$  and  $R_{\text{free}}$
- Fit to electron density map
- Residual peaks in difference electron density map
- Omit electron density for ligands
- Ramachandran plot
- Geometric parameters
  - *A well refined structure should exhibit not more than 0.02 Å r.m.s. deviation from ideal bond lengths and not more than 2.0° r.m.s. deviation from ideal bond angles.*
  - *Needless to say, that planar groups should be planar, amino acids should have the correct chirality, etc.*



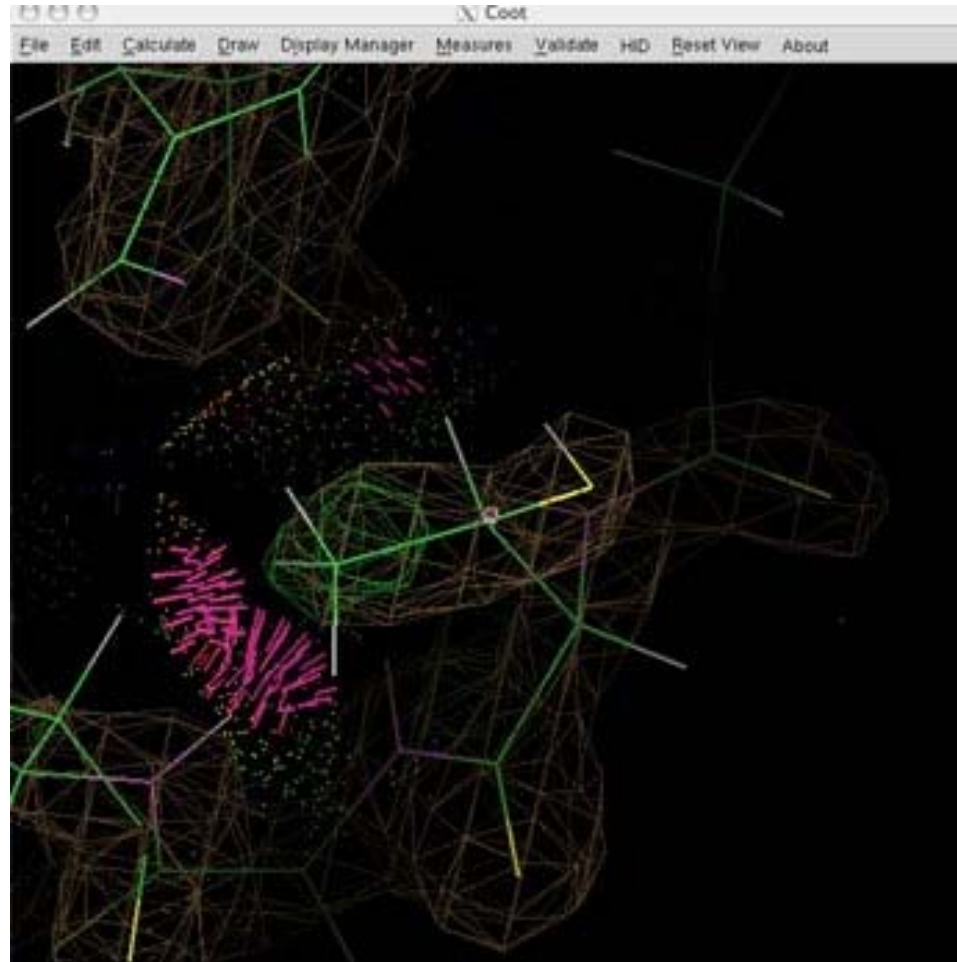
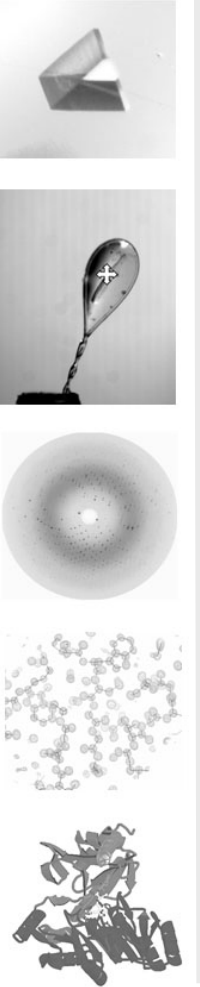
# Structure Validation



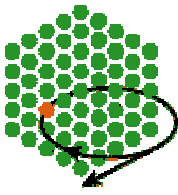
- Refinement R and  $R_{\text{free}}$
- Fit to electron density map
- Residual peaks in difference electron density map
- Omit electron density for ligands
- Ramachandran plot
- Geometric parameters
- Hydrogen atom clash analysis



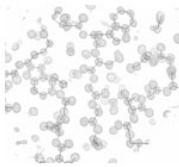
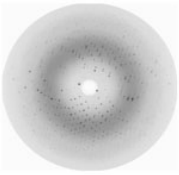
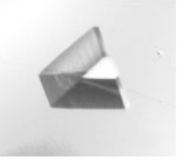
# Hydrogen Bond Clash Analysis



see: <http://molprobity.biochem.duke.edu>



# Literature

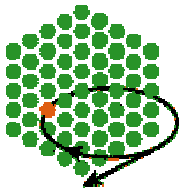


For beginners:

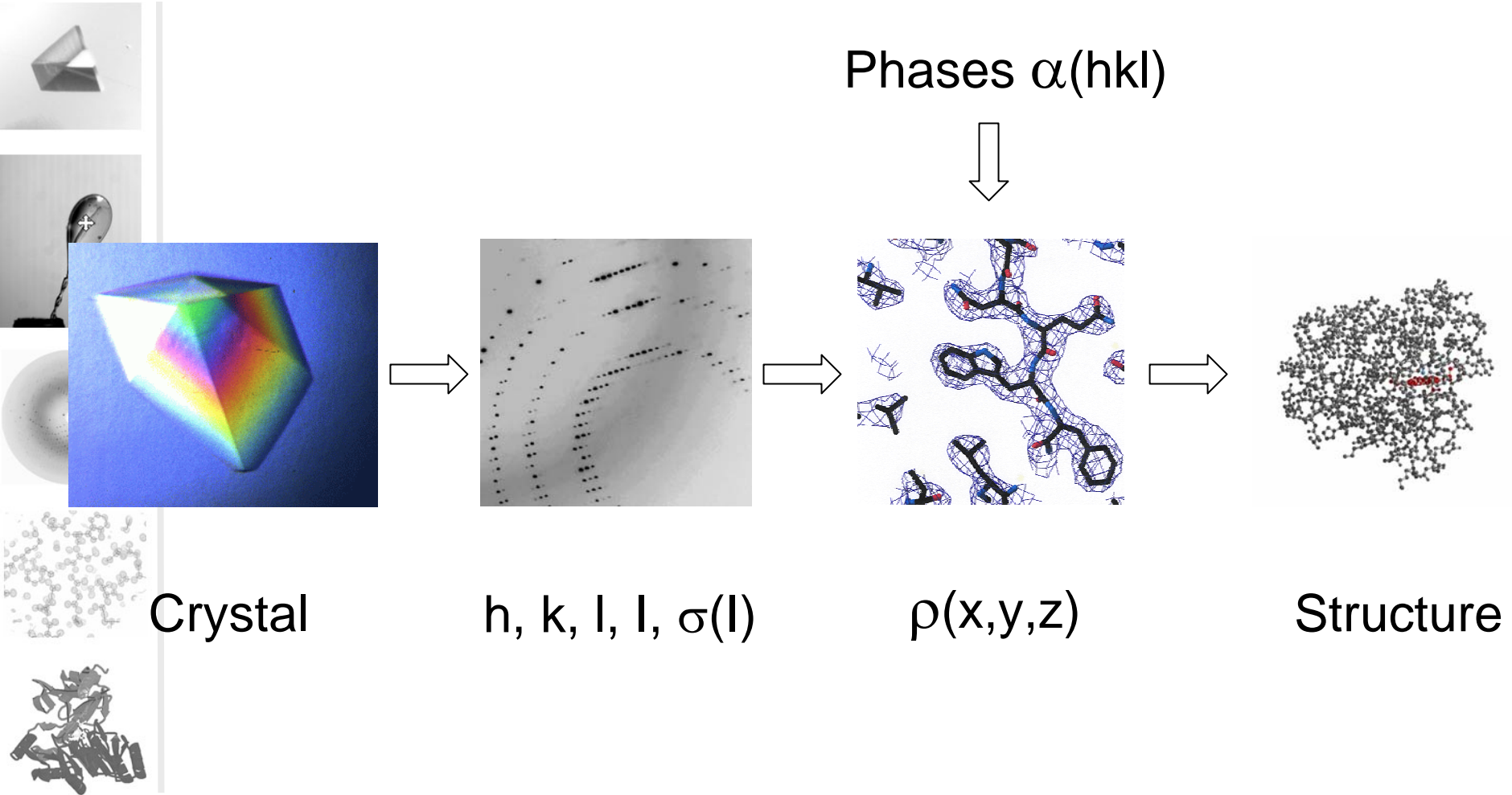
G. Rhodes. *Crystallography Made Crystal Clear*, Academic Press (1993).

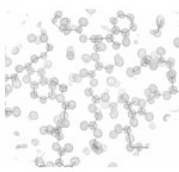
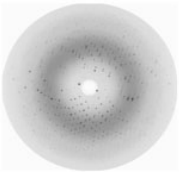
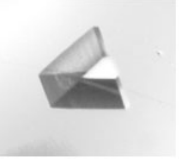
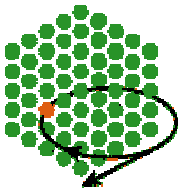
More advanced:

J. Drenth. *Principles of Protein X-ray Crystallography*, Springer Verlag (1994).



# Summary - Structure Determination





For more information, questions, etc.

[msweiss@embl-hamburg.de](mailto:msweiss@embl-hamburg.de)