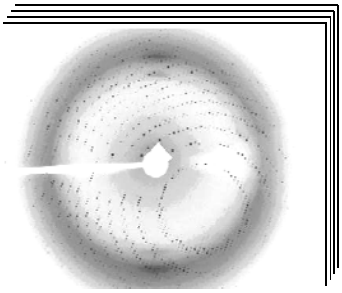
Protein structure from X-ray diffraction

Raw data: Diffraction images



Crystallize a protein with known chemical structure, say UvrB:

MSALEFGPSLKMNE...

Conformation, 3D-structure:

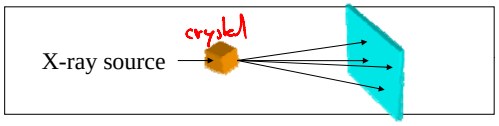
```
CRYST1  221.200   73.600   80.900  90.00  90.00  90.00 P 21 21 2     4
ATOM      1  N   ILE A   6      97.764  18.390  39.211  1.00 84.23
ATOM      2  CA  ILE A   6      97.130  18.979  37.983  1.00 84.74
ATOM      3  C   ILE A   6      96.655  17.885  37.031  1.00 84.98
ATOM      4  O   ILE A   6      97.460  17.052  36.605  1.00 85.82
ATOM      5  CB  ILE A   6      98.139  19.855  37.248  1.00 99.99
ATOM      6  CG1 ILE A   6      99.043  18.979  36.389  1.00 99.99
ATOM      7  CG2 ILE A   6      98.984  20.617  38.263  1.00 99.99
ATOM      8  CD1 ILE A   6     100.297  18.614  37.178  1.00 99.99
ATOM      9  N   ALA A   7      95.359  17.896  36.702  1.00 84.56
ATOM     10  CA  ALA A   7      94.757  16.906  35.795  1.00 83.75
ATOM     11  C   ALA A   7      95.816  16.303  34.876  1.00 83.47
.
.
.
ATOM   7604  O2*  G R   3      73.810  43.517  21.774  1.00 62.48
```

Reciprocal space

Real space

Diffraction experiment

Ingredients: X-ray beam, crystal, detector

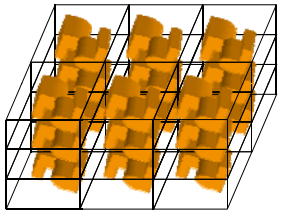


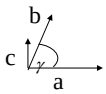
Our topics for this morning Crystal → Diffraction data

1) What are crystals, how do we obtain them?
2) What signals are detected when X-rays hit a sample?
3) What is observed when the sample is crystalline?

Three dimensional crystals

- periodic array of atoms, molecules, viruses...
- translational symmetry along three vectors a, b, c
- unit cell with edges a, b, c and angles α, β, γ is the building block for the whole crystal



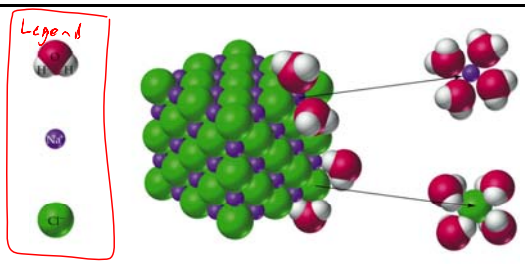


Try to crystallize some fish!

Instructions to make a 2D-crystal out of fish

- Team up with one or two other people and get some fish and overhead transparency
- Take a fish and place it on the transparency
- Place a second fish on the first fish so they superpose
- Translate the second fish in an arbitrary direction (unit repeat a)
- Repeat placing and translating until you have a row of 5 fish
- Place a new fish on the first fish and translate it in a different direction (unit repeat b)
- Complete the new row you just started
- Make another couple of rows

NaCl crystal



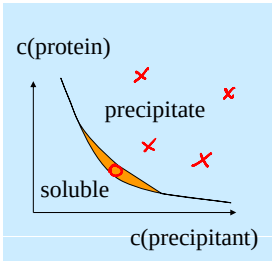
http://cwx.prenhall.com/bookbind/pubbooks/mcmurrygob/medialib/media_portfolio/text_images/FG09_01.JPG

Problem set 1: protein crystals

Differences in terms of	NaCl	protein (fish)
• size of objects	2 Å	100 Å
• shape of objects	sphere	potato
• space between objects	none	solvent
• number and nature of contacts	about 10 el	
• expected mechanical stability		
• requirement for mother liquor		
• kinetics of nucleation		
• steps required to add another object		
• kinetics of growth		

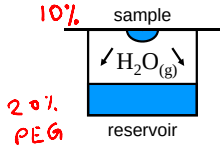
Crystallization: solubility

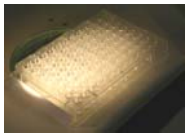
- protein solubility in water is tunable
  PEG $(NH_4)_2SO_4$
- amorphous precipitate forms easily
- protein crystals form when conditions are "just right"
  favor certain contacts

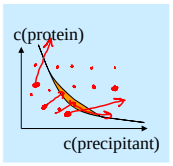


Crystallization: vapor diffusion

- mix protein solution with precipitant solution 1:1 and equilibrate against excess of the latter
- takes 1 µl of 10 mg/ml protein sol. per experiment

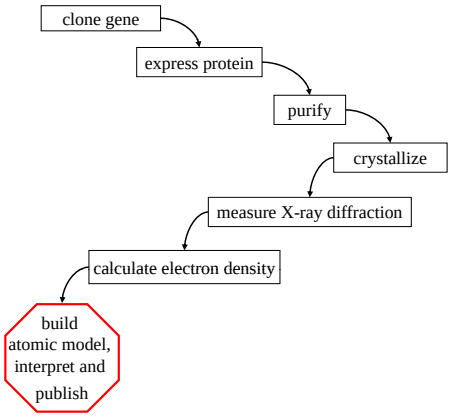






Goal: slow increase of protein and precipitant concentrations over days

Steps in solving a structure



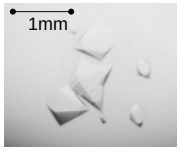
The crystallographic bottleneck

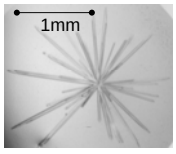
- No crystals, no structure
- "No crystals, no grant"

Problems with crystals that may prevent structure determination:
weak diffraction, twinning, very high mosaicity, too many copies in the unit cell

- Try complexes with ligand
- Try different constructs (domains, truncated forms)
- Try similar protein from different organism

Escherichia coli
7 Å resolution

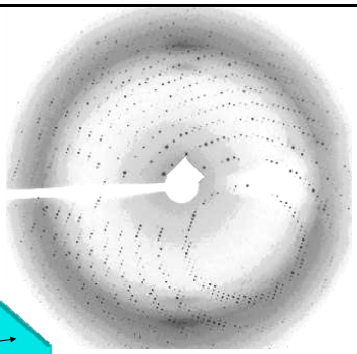


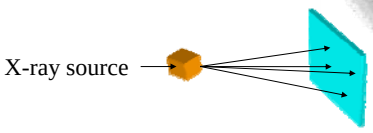


Bacillus caldotenax
2.6 Å resolution

Features of diffraction images

- Discrete spots
- Intensities are different from spot to spot
- In general, intensities decrease from the center to the edge





Structure ↔ Diffraction data

Step #1: understand how a given structure leads to observed diffraction patterns

Step #2: measure diffraction data

Step #3: solve structure based on diffraction data

Our topics for this morning

✓ 1) What are crystals, how do we obtain them?

→ 2) What signals are detected when X-rays hit a sample?

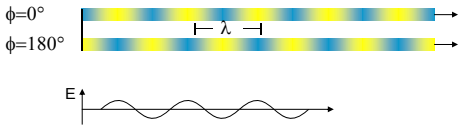
3) What is observed when the sample is crystalline? (TL)

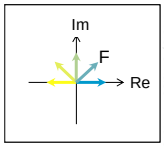
X-rays

- Electromagnetic radiation, $\lambda \approx 1$ Å
- Produced by copper anode or synchrotron
- X-ray beam described by
  - wavelength λ
  - direction

  } wave vector $\mathbf{k}$, $|\mathbf{k}| = 1/\lambda$
  - amplitude $|F|$ (how strong)
  - phase ϕ (peaks and valleys)

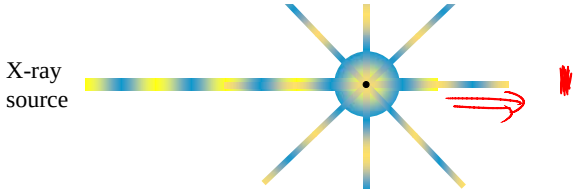
  } $F=|F| \exp(i\phi)$



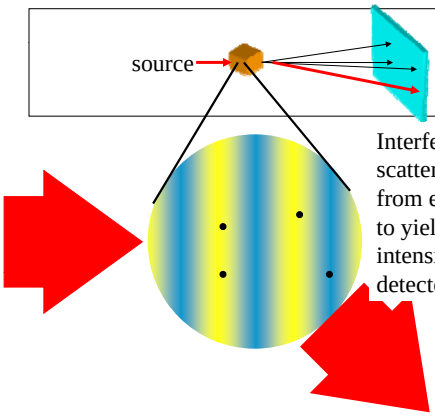


Elastic scattering

- how do X-rays interact with electrons? weak
- where do scattered X-rays go? all directions
- which wavelength? same
- what phase? unchanged

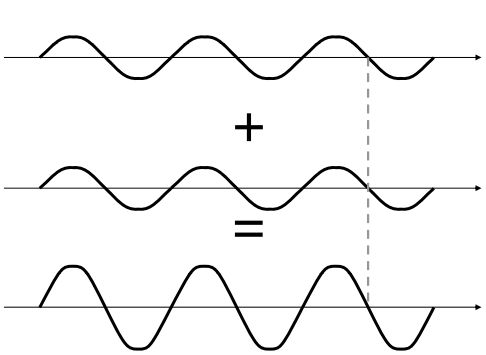


Every atom contributes to each spot



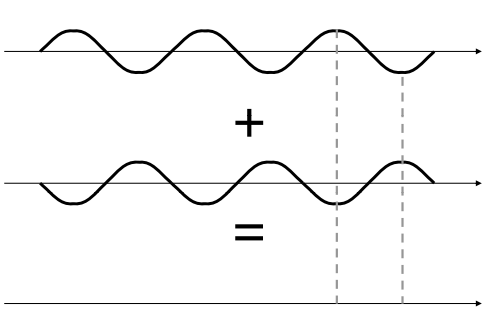
Interference explains how scattered Xrays originating from every atom combine to yield the characteristic intensities observed on the detector.

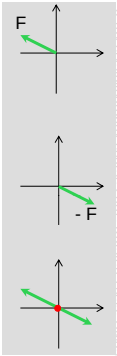
Constructive Interference: $\Delta\varphi=0$



Amplitude doubles; Intensity quadruples

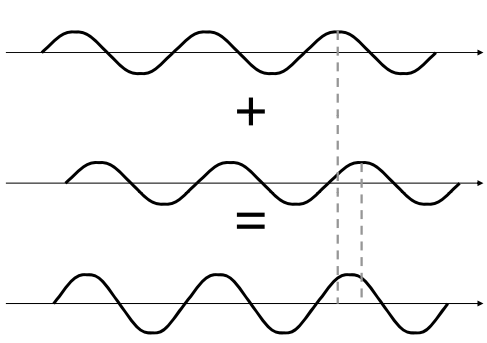
Destructive Interference: $\Delta\varphi=\pi$





Amplitude vanishes; observe zero intensity

Interference with arbitrary $\Delta\varphi$

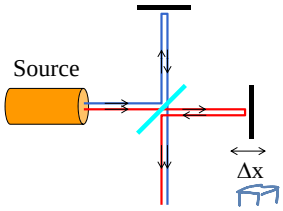


Why is the complex amplitude F a useful parameter?

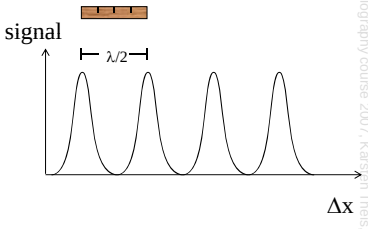
To find the resultant wave, we just add up the amplitudes F_1 and F_2

Measuring distances using light

Michelson Interferometer



Light is used as a yardstick that has a length of λ/2



To measure atomic distances through interference and thus determine the structure of molecules, our yardstick must have atomic dimensions.
This is why X-rays are required for crystallographic structure determinations

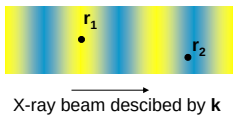
Interference

- Interference describes superposition ("addition") of X-rays with identical wavelength and direction
- Depending on the positions of the scatterers, the scattered X-rays have certain phase differences resulting in amplification or cancelation
- Interference leads to diffraction patterns on the detector that contain structural information

why is the wave vector **k** a useful parameter?

1) Two waves with the same k will interfere

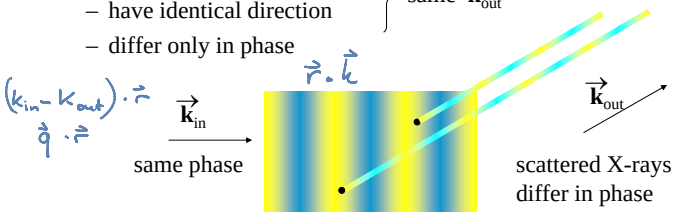
2) The scalar product **k** · **r** tells us whether a point at **r** is in a peak or valley



Atomic position determines phase

- Scattered X-rays hitting a certain location on the detector
  - originate from different atoms
  - have identical wavelength
  - have identical direction
  - differ only in phase

same $\vec{k}_{out}$



Difference in phase results from difference in distance traveled between source and detector

Crystal and diffraction data quality

UvrB crystals were grown by hanging drop vapor diffusion. Equal volumes of a solution containing 8 mg/ml UvrB in 500 mM NaCl, 20 mM Tris–HCl pH 8.2, 1 mM DTT, 0.1 mM EDTA, 0.03% dodecylmaltoside were mixed with a precipitant solution containing 14–18% PEG 6000 or PEG 20 000, 10 mM $ZnCl_2$ and 100 mM Bicine at pH 9 and equilibrated against a reservoir solution containing 20% PEG 6000, 500 mM NaCl, 100 mM Tris–HCl pH 8.5. Diffraction data of crystals, cryocooled in liquid nitrogen, were collected at beamlines X26C and X25 at the National Synchrotron Light Source in Brookhaven. The crystals belong to space group $P3_121$ with $a = b = 150.4$ Å, $c = 79.5$ Å and contain one molecule per asymmetric unit. The structure of UvrB was solved by MIR.

Table 1: Data collection and MIR statistics

	Nat-Ref	Nat-MIR
Maximum resolution (Å)	2.6	2.9
Completeness	1 (1)	0.993 (0.99)
Mean redundancy	10 (7)	7 (7)
Rsym	0.09 (0.55)	0.10 (0.60)
$\langle I/\sigma I \rangle$	29.4 (3.1)	18.0 (2.9)

Break!!!

How do we obtain
the crystal structure of a protein
from the diffraction data $|F_{hkl}|$?