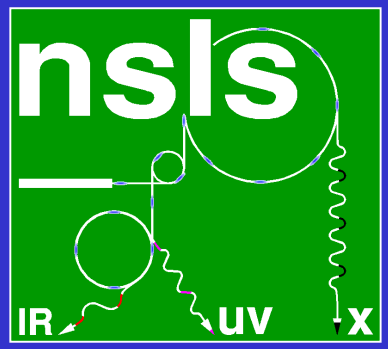
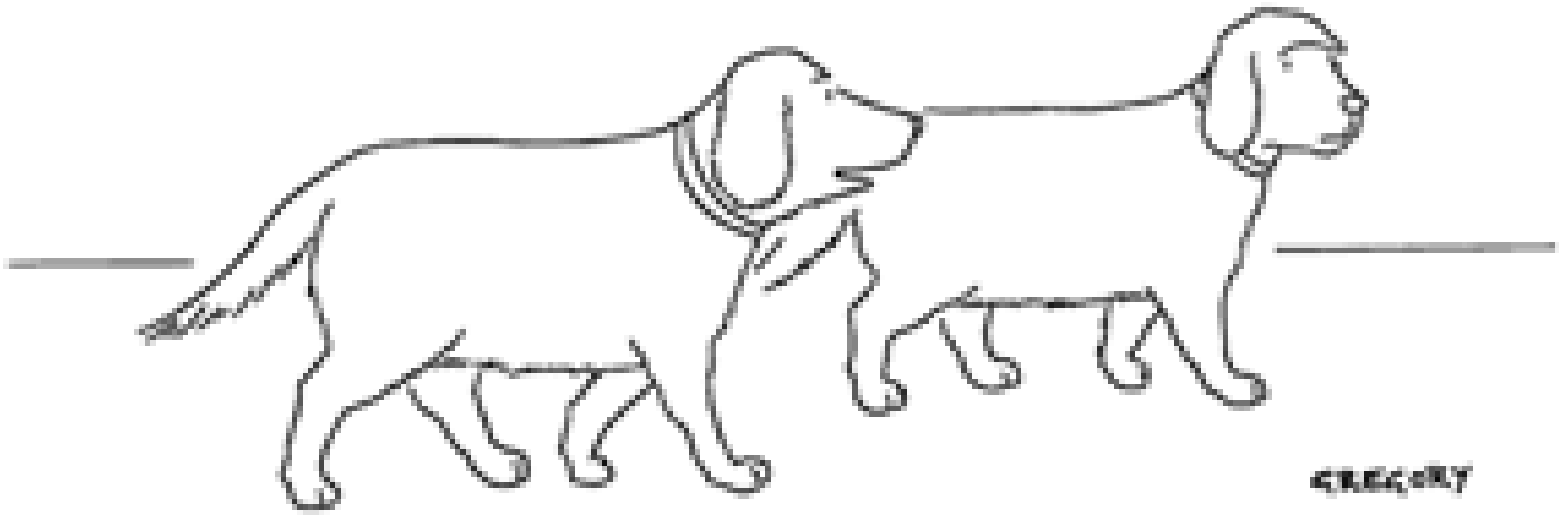
Human Intervention in the Analysis of Powder Diffraction Data (Am I the only one who has trouble solving the interesting problems?)



STONY BROOK
STATE UNIVERSITY OF NEW YORK

Lots of help from Cristian Botez, Ashfia Huq, Jae-Hyuk Her, Silvina Pagola, work here in collaboration with Raj Suryanarayanan, Cletes Nunes, Dipo Omotoso

http://powder.physics.sunysb.edu
pstephens@sunysb.edu



chemists

What do ~~cats~~ really want?

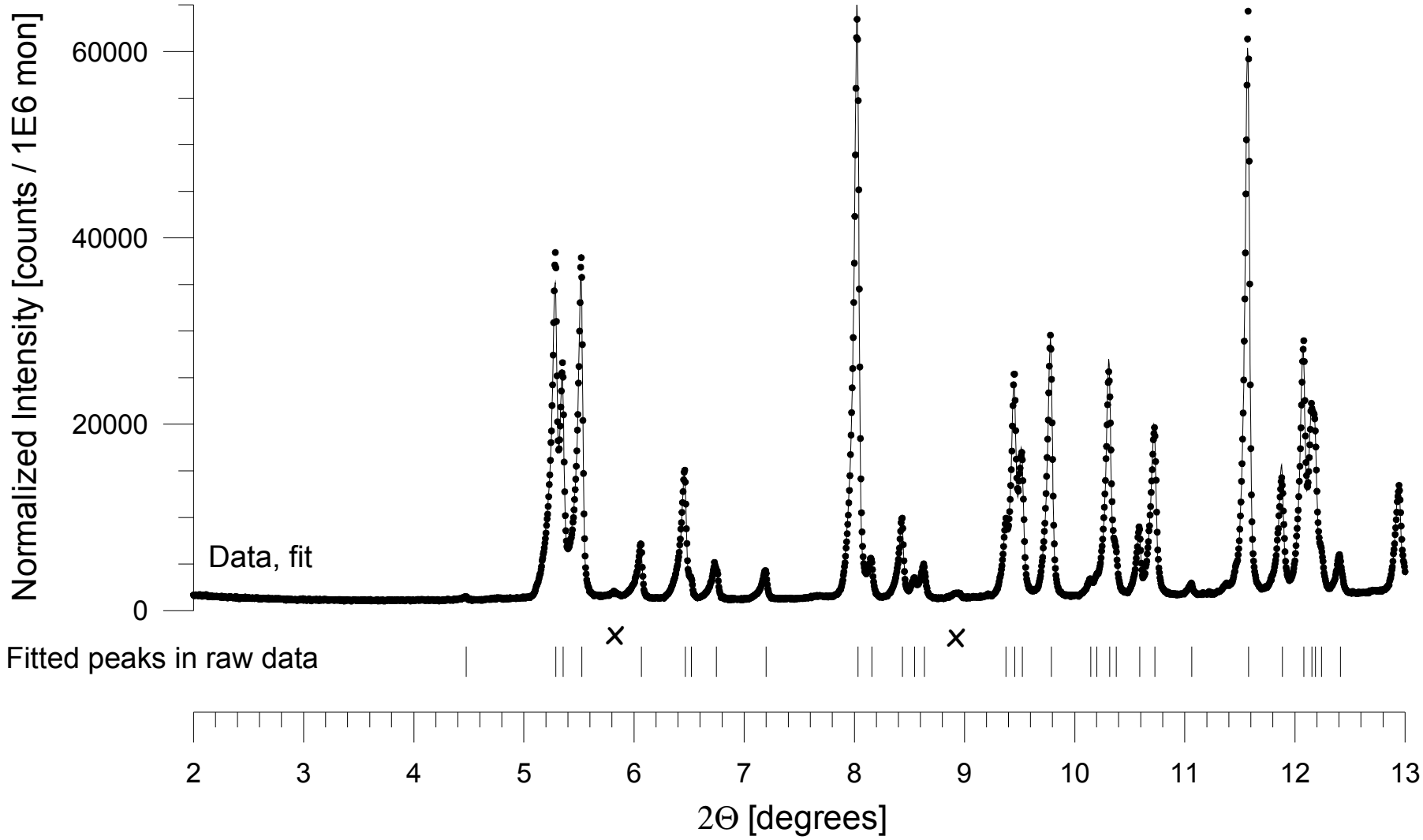
Indexing using off the shelf, public domain software (ITO)

- Good case
- Bad case
- Ugly case

Some new structures of mannitol and mannitol hydrate.

"Discovery" of a more complicated structure than was originally modeled – Ranitidine HCl (~1 yr. old – you may have heard it.)

Organic compound. 2 hours data collection (to 20°) at NSLS
Fit 32 peaks and feed their positions to ITO.



Index the peaks of a powder pattern to find the lattice

$\sin^2\theta = Ah^2 + Bk^2 + Cl^2 + Dkl + Ehl + Fhk.$

Find {A,…,F} so that (hkl) can be assigned to each peak.

☹💣☠ Inaccurate data, added impurity lines that don't belong.

ITO (J. Visser, J. Ap. Cryst. **2**, 89-95 (1969))

Zone: 2D slice (hk), so $\sin^2\theta = Ah^2 + Bk^2 + Fhk$

1. For each of the first 36 pairs of reflections, assume they define a zone and see if there is a significant number of lines in that zone.

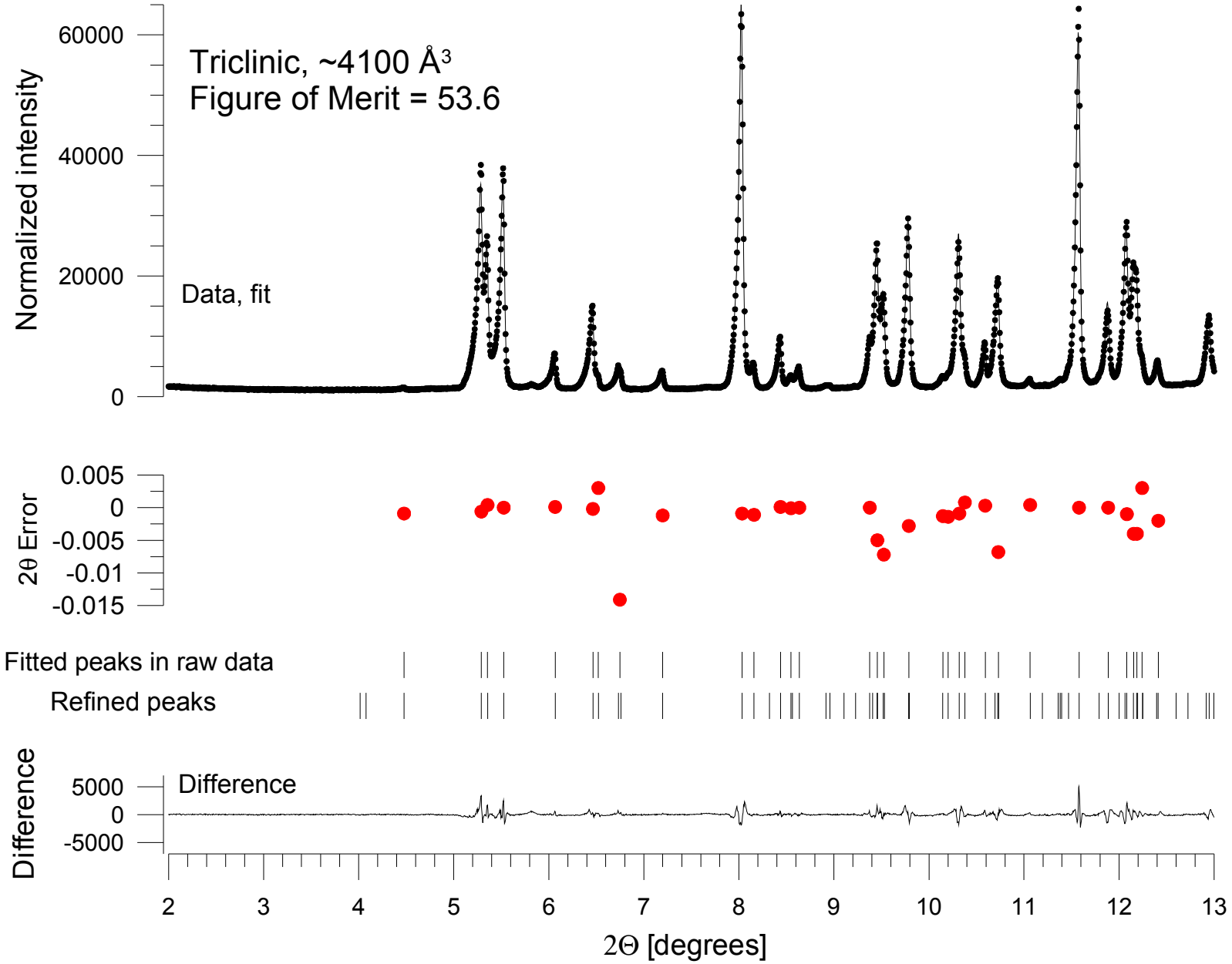
2. Refine each zone. Judge quality. Sort zones by quality.

3. For pairs of good zones, find a dihedral angle that matches some peaks not in either zone.

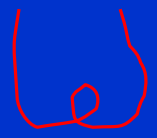
4. Judge quality of each possible solution.

5. Choose the best few solutions. Refine.

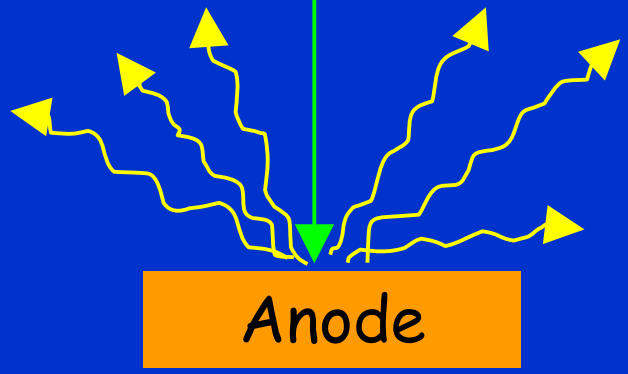
ITO was developed in an era of slow computers, lower quality data


Triclinic, ~4100 Å3
Figure of Merit = 53.6
Normalized intensity
60000
40000
20000
0
Data, fit
2θ Error
0.005
0
-0.005
-0.01
-0.015
Fitted peaks in raw data
Refined peaks
Difference
5000
0
-5000
Difference
2
3
4
5
6
7
8
9
10
11
12
13
2Θ [degrees]

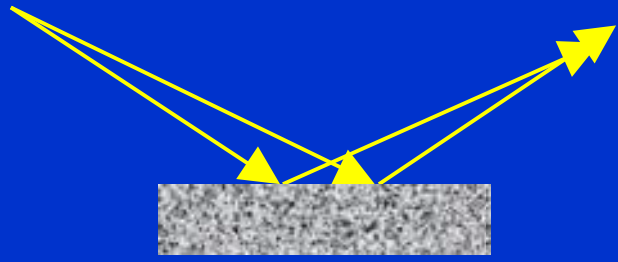
X-ray tube



Electrons,
50 keV
40 mA

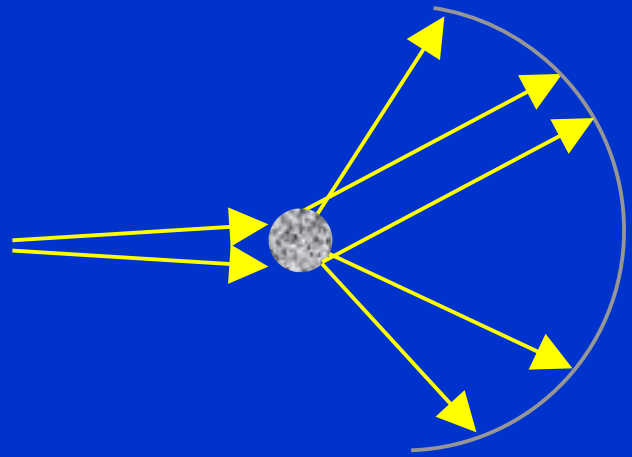


Bragg-Brentano



Focus diverging beam. Moderate resolution, sensitive to sample displacement, transparency

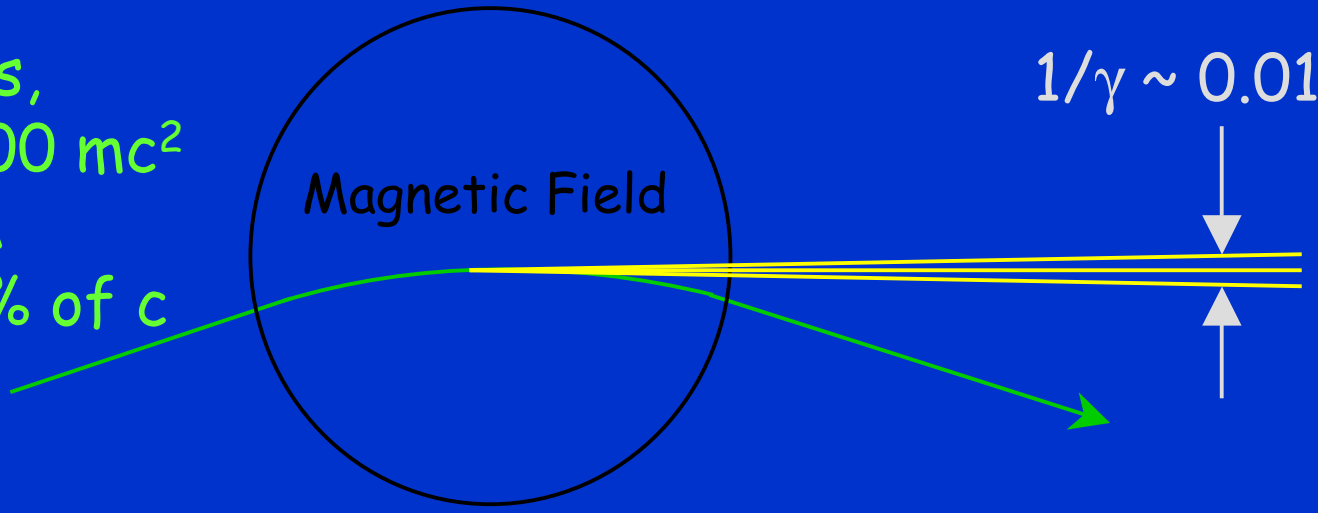
Debye-Scherrer



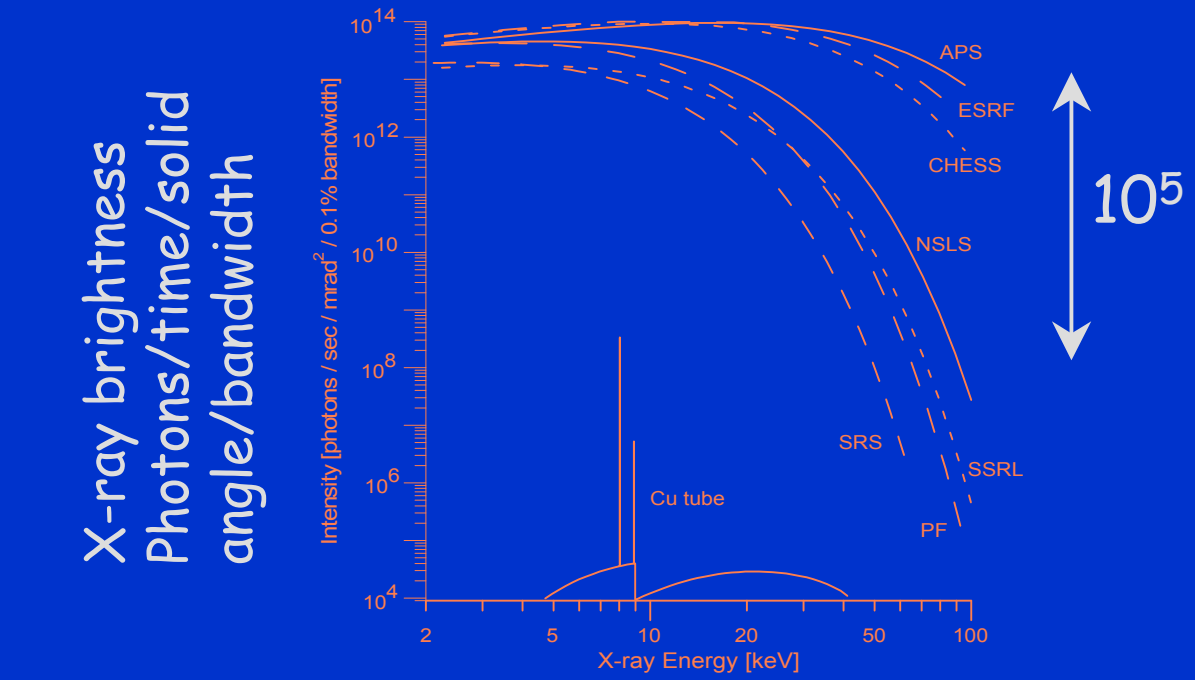
Resolution limited by divergence, parallax

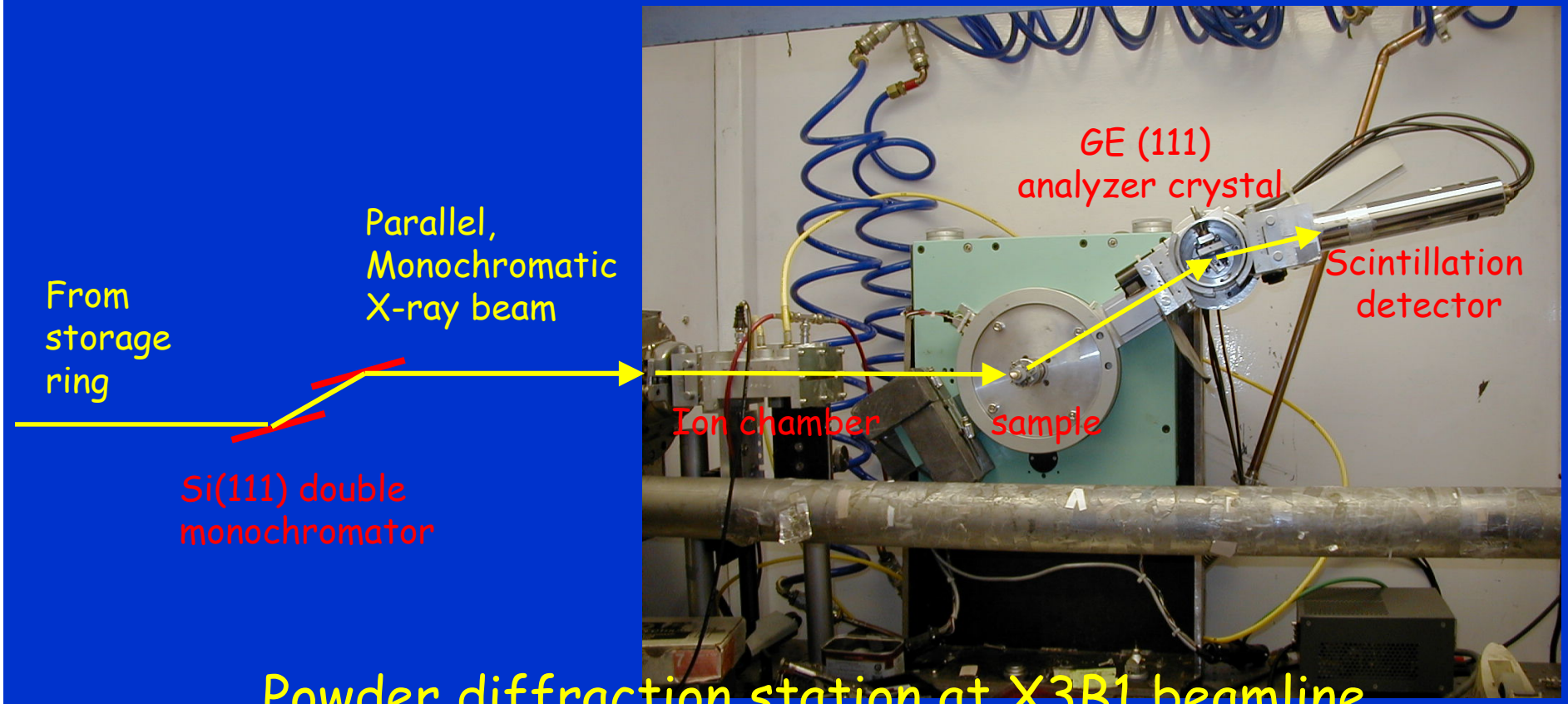
Synchrotron Radiation

Electrons,
2.8 GeV = 5500 mc^2
300 mA
99.9999983% of c



X-ray brightness
Photons/time/solid angle/bandwidth

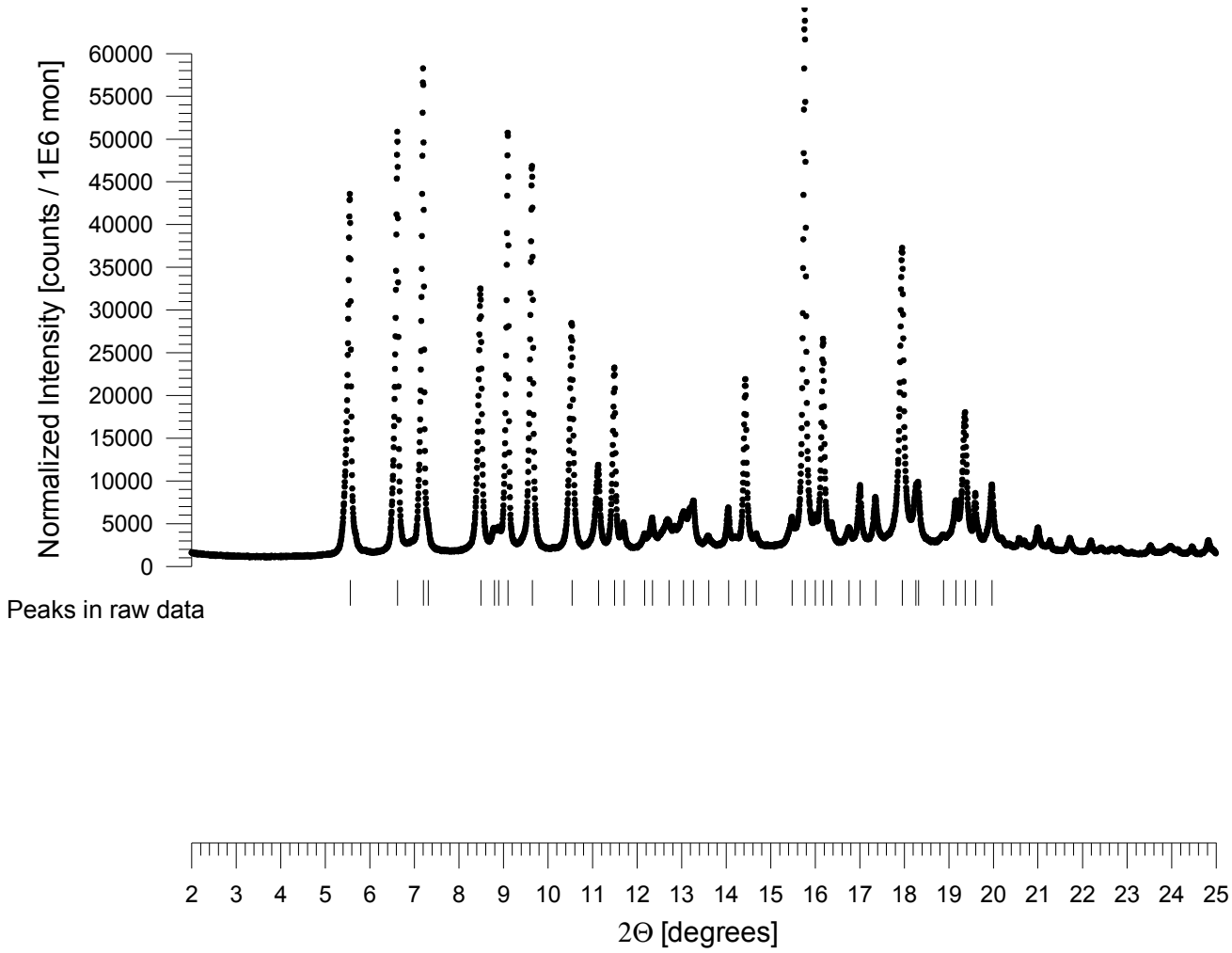




Powder diffraction station at X3B1 beamline,
National Synchrotron Light Source,
Brookhaven National Laboratory, U. S. A.

Available for general users (25%; accelerated access available),
rent ($260/hr proprietary), or collaboration.

Unknown compound #2, trouble indexing
Good looking pattern, reliable fit to raw peak positions.



Use ITO. (Other programs are widely used, and I don't want to argue about which is the best.) Particularly suited to triclinic.

TRIAL LATTICES AFTER INDEXING AND LS REFINEMENT

ONLY THE FIRST 20 LINES ARE CONSIDERED FOR INDEXING

A	B	C	D	E	F	LINES INDEXED	FIGURE OF MERIT
29.9	17.8	25.2	1.1	0.6	0.0	17.	4.9
17.9	9.5	20.2	0.0	0.0	0.0	17.	4.8
18.6	11.3	71.4	0.0	0.0	0.0	16.	8.3
18.4	11.5	71.2	0.0	0.0	0.0	16.	7.5
17.8	11.6	18.2	0.0	0.0	0.0	16.	6.0
18.7	11.1	71.0	0.0	0.0	0.0	15.	2.4
29.8	17.7	65.9	10.7	24.5	0.0	15.	5.4
17.7	9.2	20.5	0.0	0.0	0.0	15.	4.9

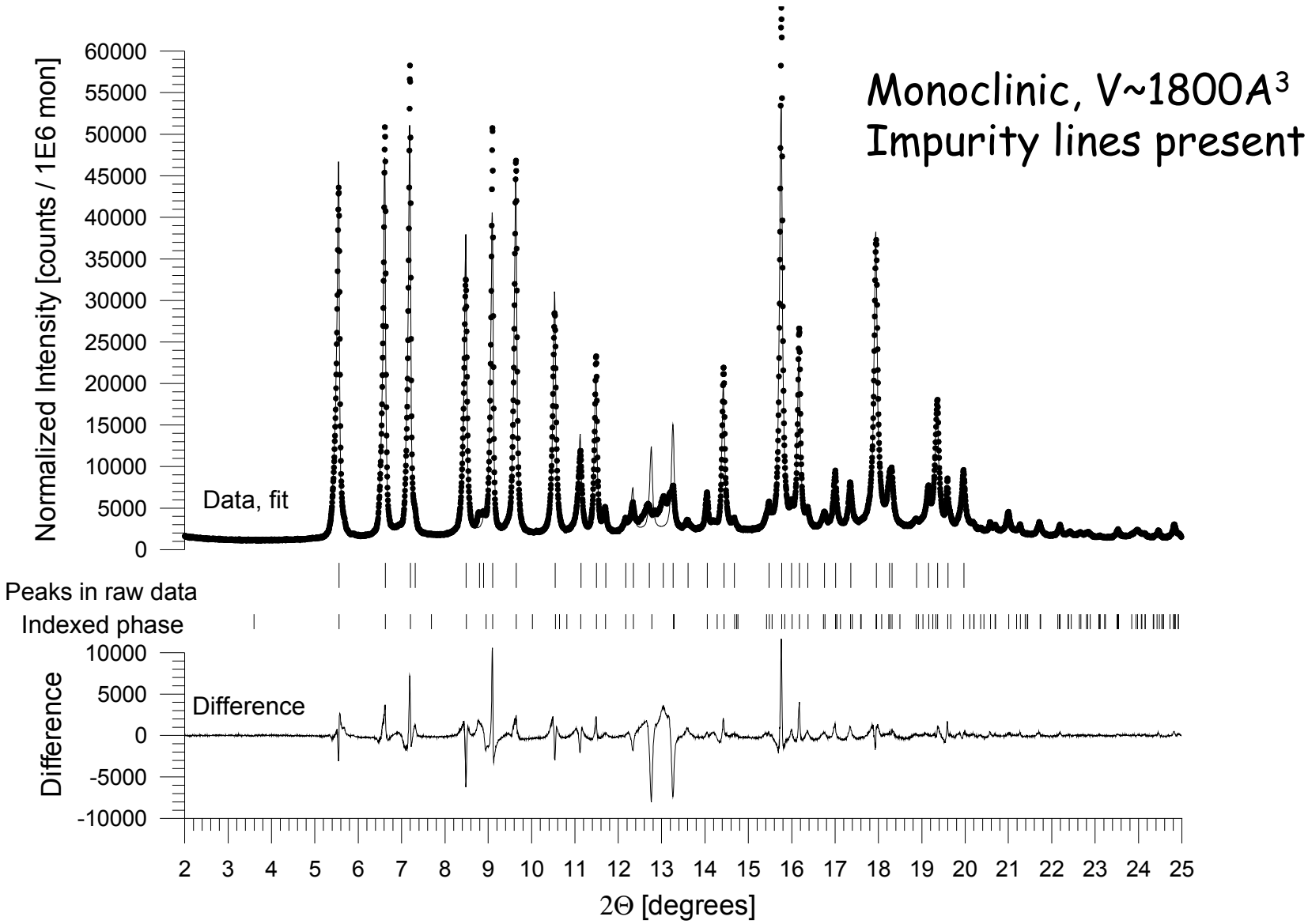
Useless.

Next pass after throwing out some of the questionable lines.

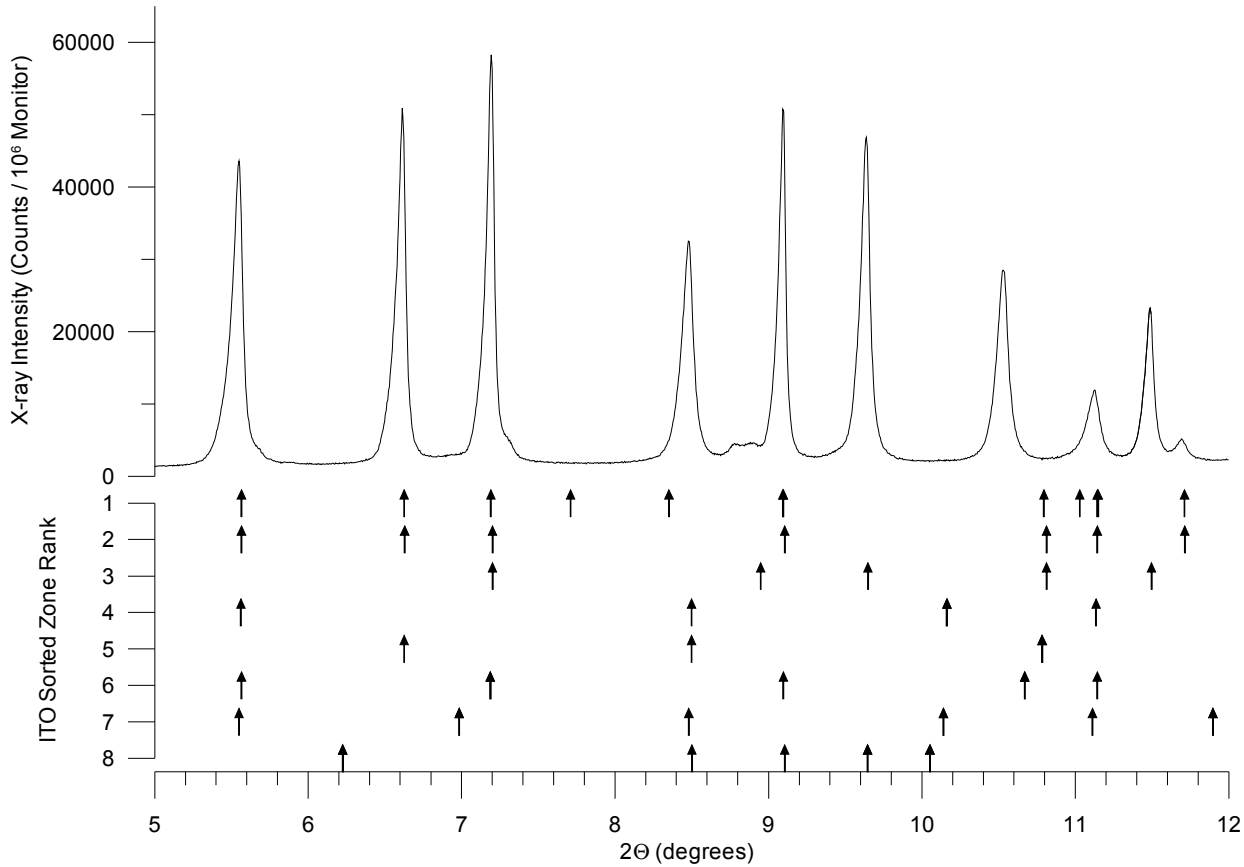
```
 TRIAL LATTICES AFTER INDEXING AND LS REFINEMENT
 ONLY THE FIRST 20 LINES ARE CONSIDERED FOR INDEXING

                                                          LINES   FIGURE
 A        B        C        D        E        F        INDEXED  OF MERIT
 74.1     39.8    115.6      0.0    23.4      0.0         16.      83.2
. . . . .

 ***************************************************************
 **  NO REALLY SATISFACTORY LATTICE HAS BEEN FOUND             ***
 **  PROBLEM IS RERUN WITH A DIFFERENT SORTING OF THE ZONES    ***
 ***************************************************************
 . . . .

 THE  4 MOST PROBABLE SOLUTIONS

                                                        LINES   FIGURE OF
 Q(A)     Q(B)     Q(C)      Q(D)      Q(E)     Q(F)   INDEXED   MERIT
 29.62    17.79    26.73      2.09      1.16     0.00      19.       6.0
 37.62    37.94    49.81     23.37     23.56     0.00      18.       7.6
 24.01    22.40    28.09     18.45     15.82   -10.51      19.       4.4
 29.41    18.73    24.90      0.74      0.83     0.00      18.       4.7
```

One actually has to dig in the middle of the file to find the plausible answer

Unknown Compound #2, indexed OK



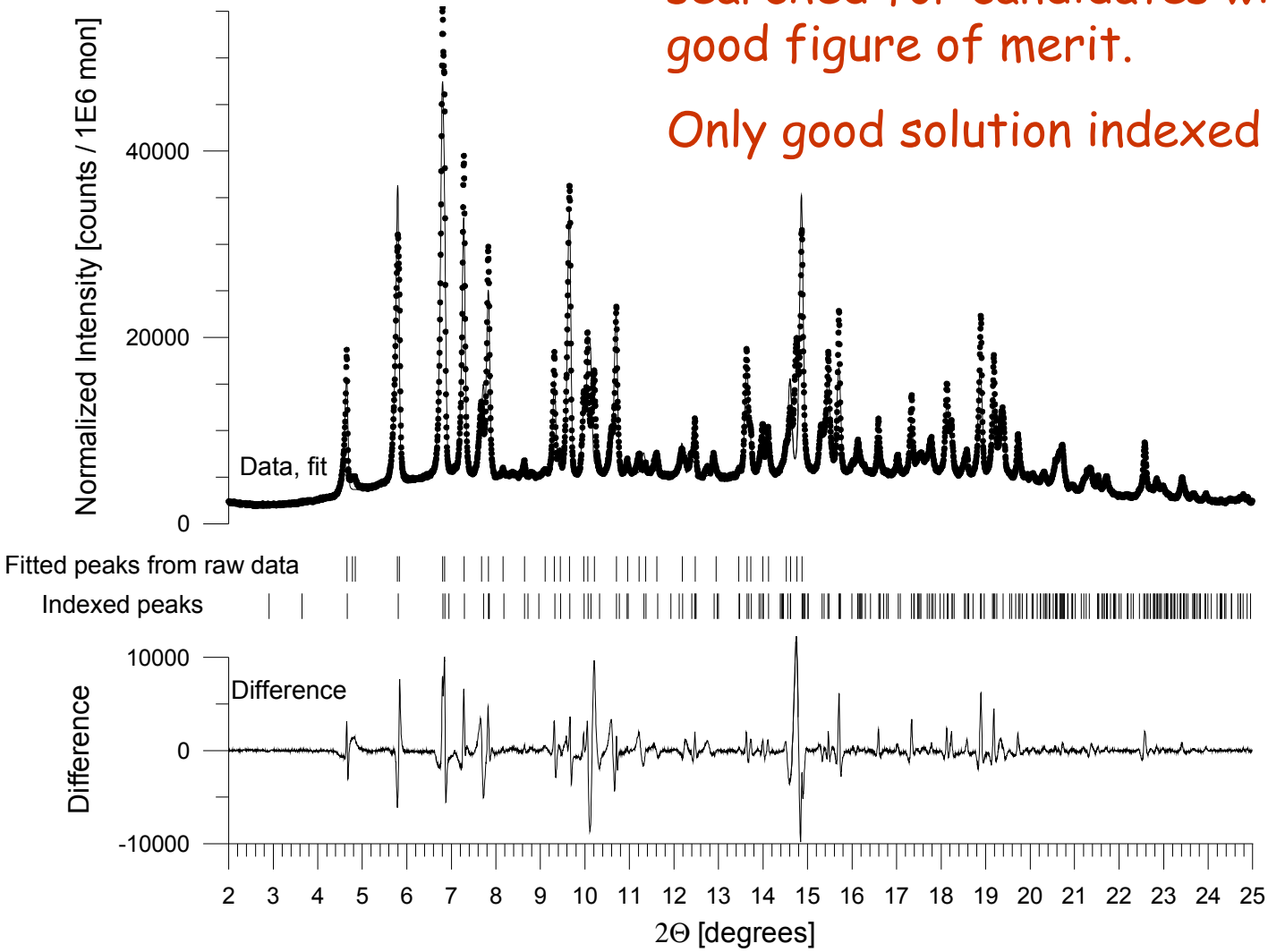
Towards a systematic approach to throwing out the observed lines that do not belong. Calculate all lines from each of ITO's candidate zones. (ITOZONES)



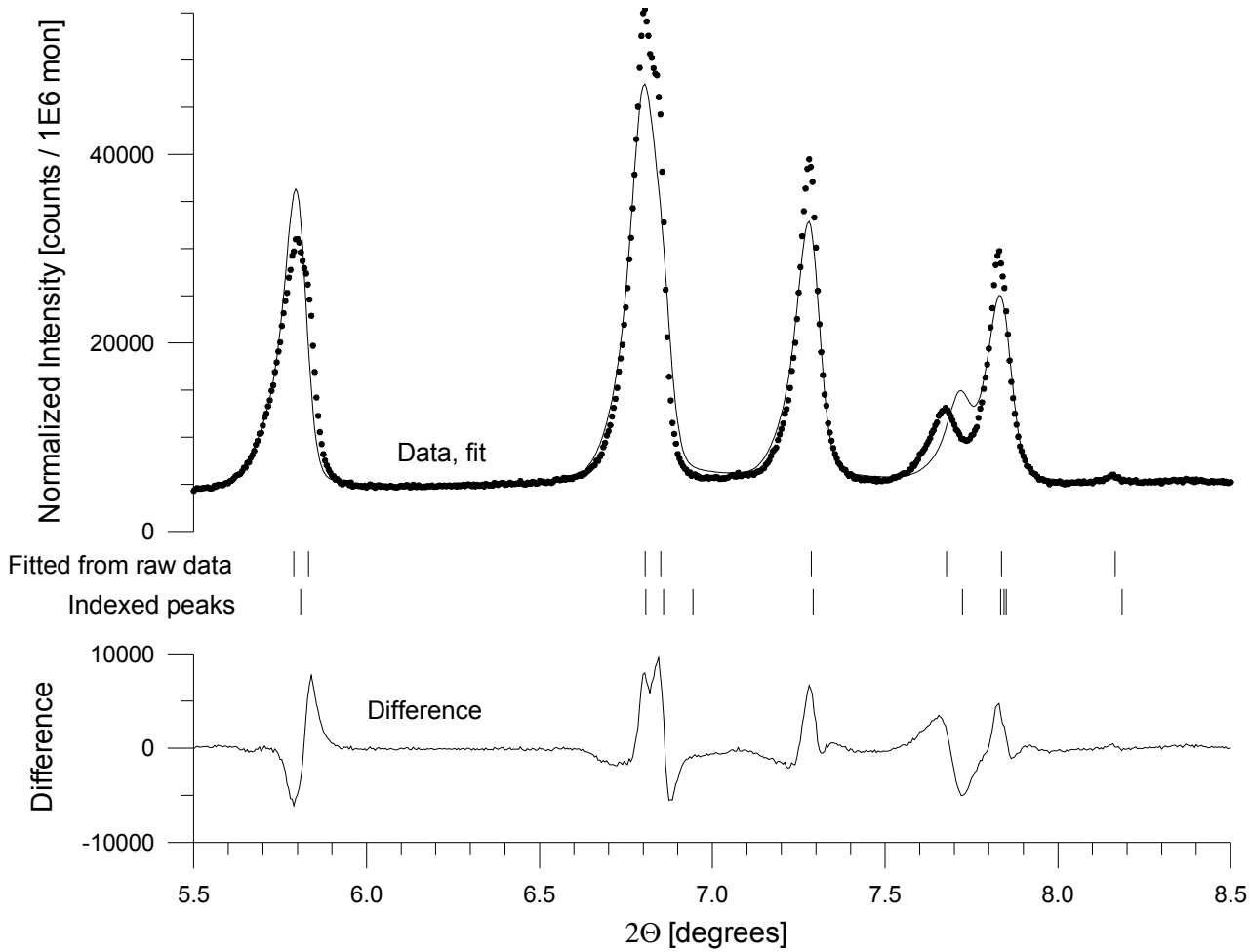
This example is even worse.

I modified ITO to randomly discard 1/4 of input lines and searched for candidates with good figure of merit.

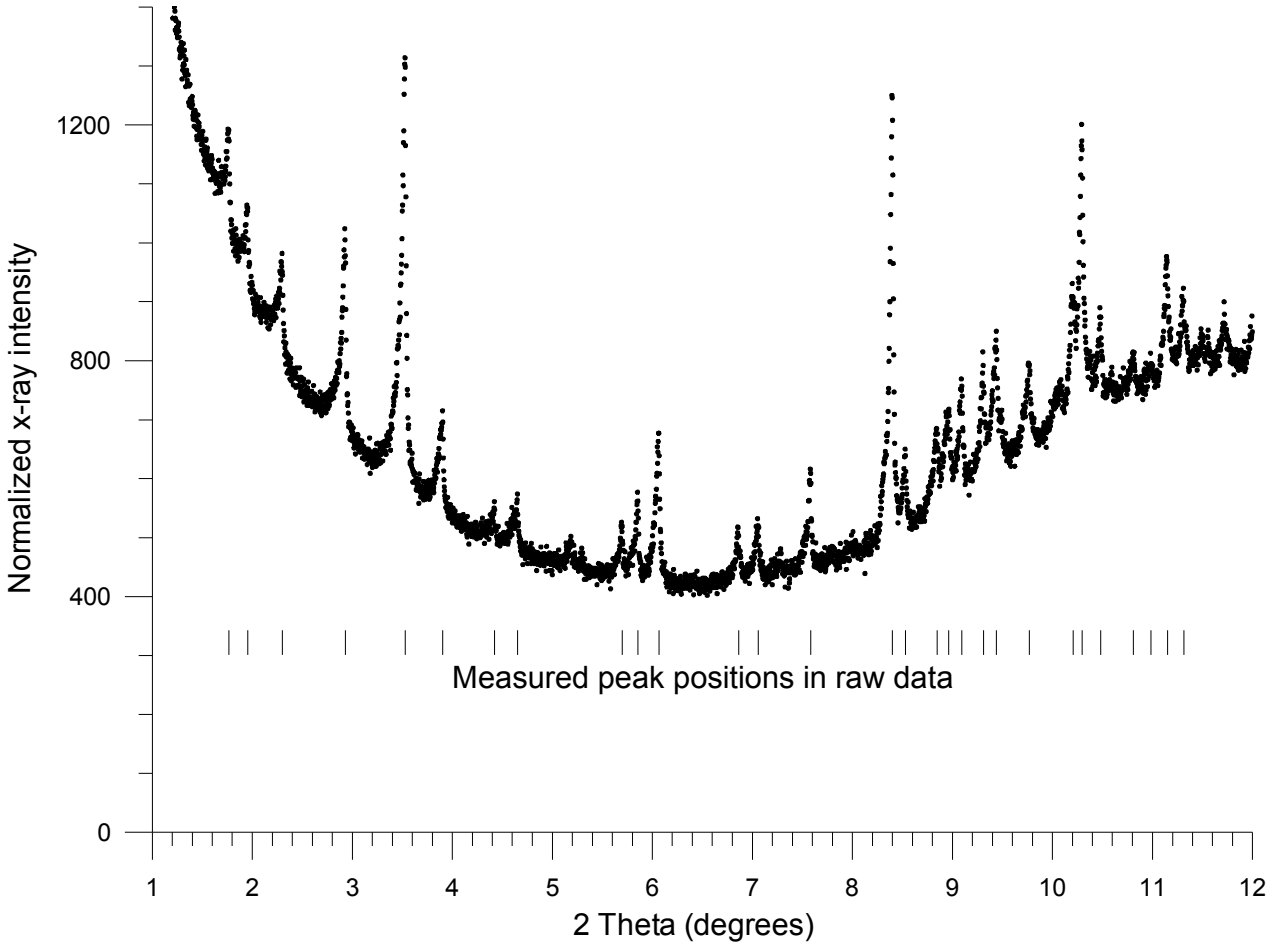
Only good solution indexed 15/20.



It is probably correct.

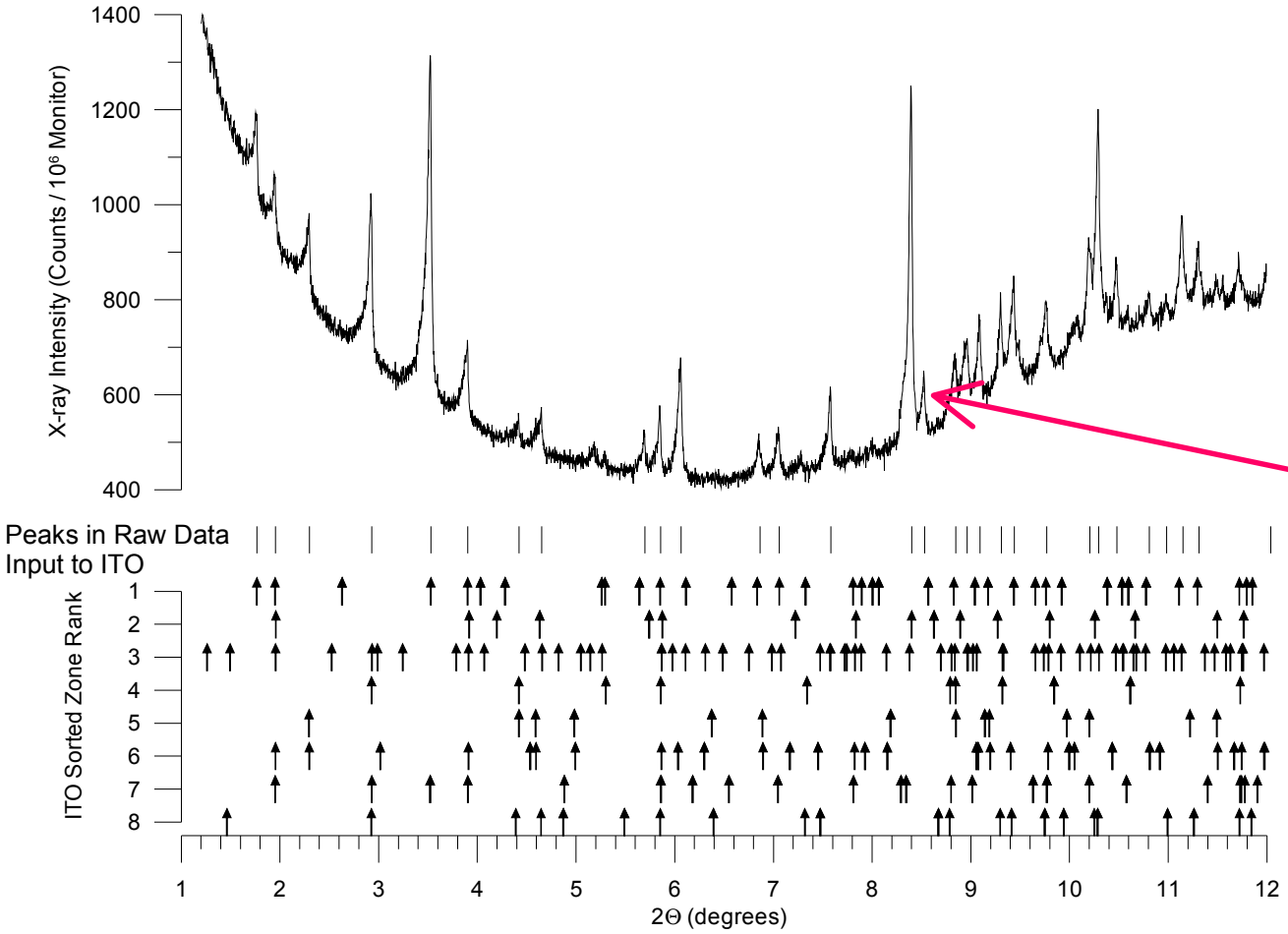


Another project. Amyloid analog from M. Balbirnie and D. Eisenberg.
(six peptides, Cd, water? Other ions?)



ITO just laughs. Data below 1°?

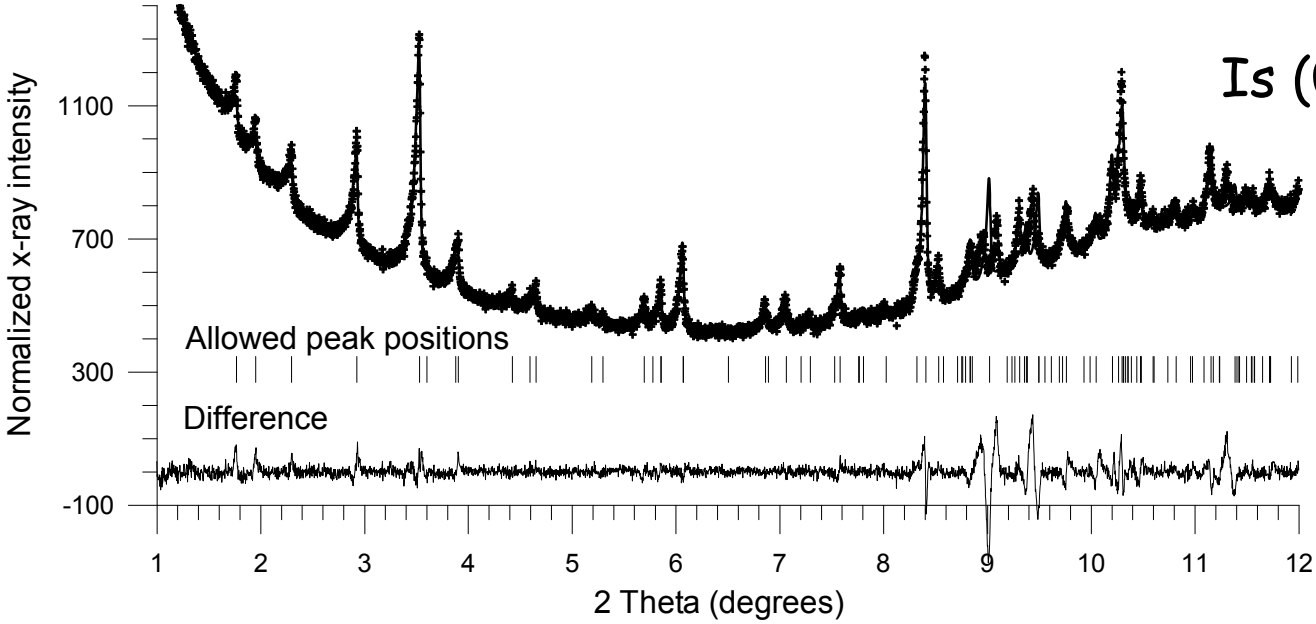
Renormalize data, divide all observed d-spacings by 5.



The first 32 allowed peaks are (h0ℓ)

D = 4.707 Å is the first observed (hkℓ)

```
**************************************************************
*     ZONE  1 INDEXES  8 LINES FROM THE  8 USED FOR ZONEFINDING     *
*     ZONE-FINDING IS RESTARTED WITH OTHER LINES.                   *
**************************************************************
```



Is (010) 4.707 Å?

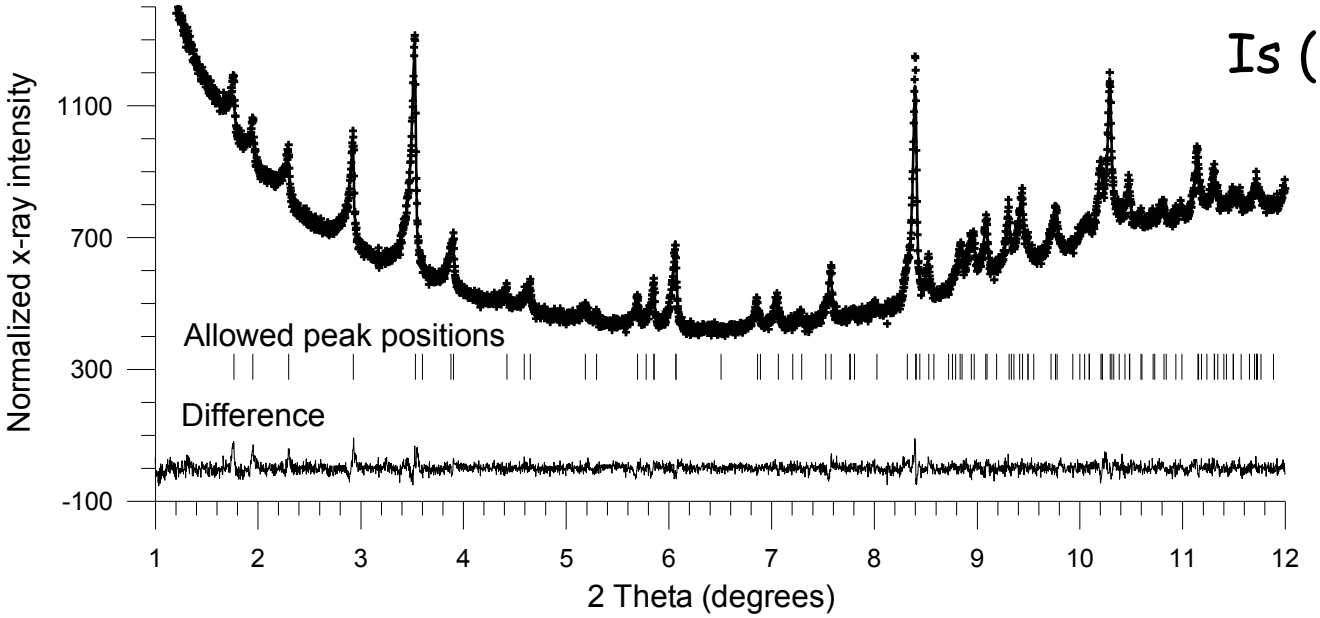
a = 23.413 Å

b = 4.707 Å

c = 21.190 Å

β= 103.86°

P 2 ?



Is (11$\bar{1}$) 4.707 Å?

a = 23.413 Å

b = 4.889 Å

c = 21.190 Å

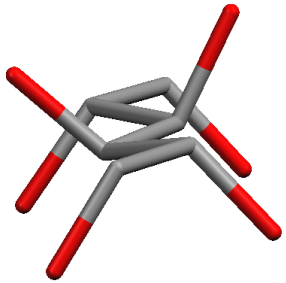
β= 103.86°

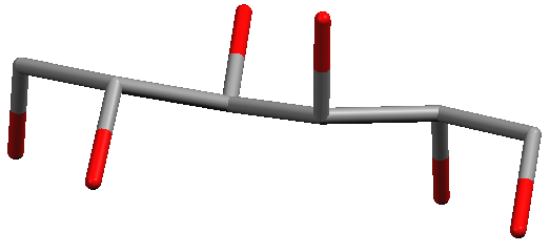
P 2_1 ?

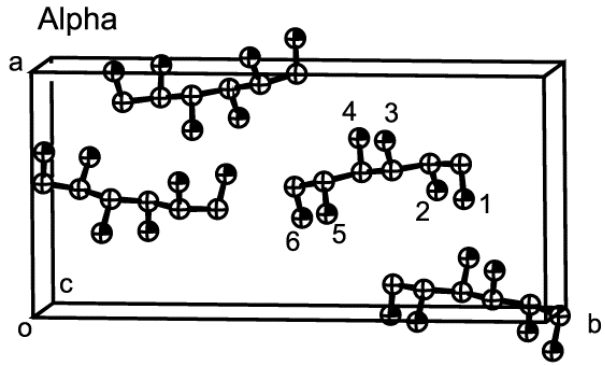
beta-sheet

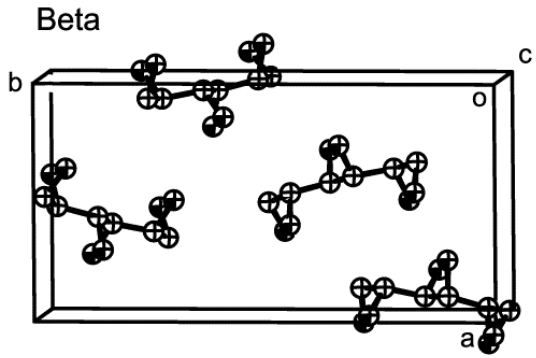
Simulated annealing solutions of organic molecules

Mannitol – commonly used excipient in pharmaceutical industry.









α and β forms were reported in 1910, structures solved in 1968.

δ reported in 1968.

δ is mechanically superior for forming tablets by compression.