

# Using Anisotropic Thermal Expansion to Improve Intensity Extraction From Powder Diffraction Data

J. P. Wright, M. Brunelli,  
G. B. M. Vaughan, A. J. Mora  
and A. N. Fitch



ESRF, Grenoble, France

# Outline

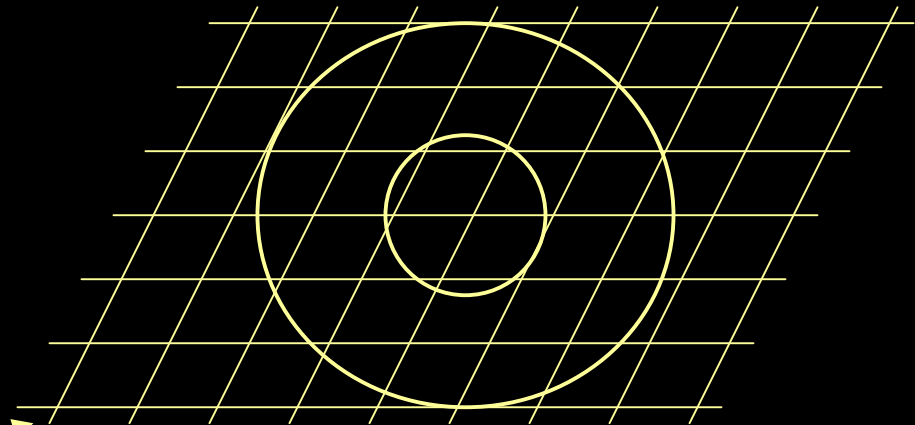
- Motivation
  - Getting better data
- First example
  - proteins
- Some data processing methods
- Application
  - small organic
- Outlook

# What's the problem?

- Peak overlap obscures structure factors
- Phase problem:
  - measure  $F(hkl)$
  - find  $f(hkl)$
- Powder diffraction gets stuck on the first step!
- Exact overlaps
  - cubic 511/333
- Inexact overlap
  - increase resolution
  - deconvolute the peaks
- Need data at "atomic resolution" for direct methods

# Overlap is 'scale invariant'

- There are more peaks overlapping at higher angles
- The units used for wavelength and cell parameters do not change the geometric problem
- The number of peaks we can measure is independent of the size of the cell



- Chemistry is not scale invariant - atoms are always the same size
- Larger structures are more difficult!!!

# Texture methods...

- Use preferred orientation to make the powder "more like" a single crystal
- Measure diffraction patterns as a function of orientation

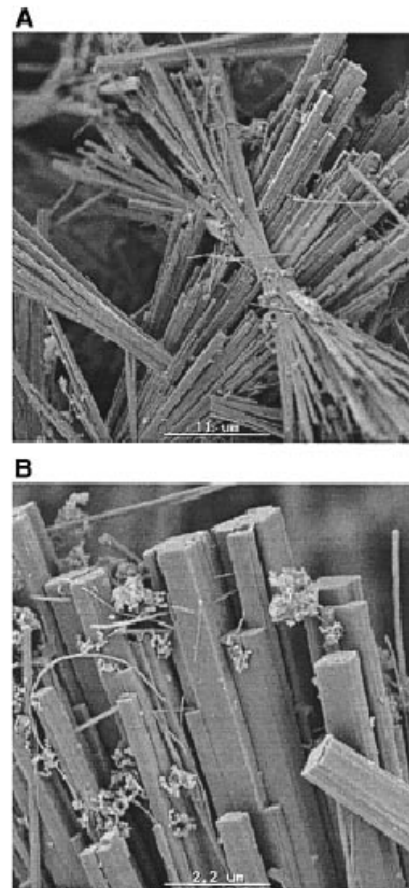


Figure 1. Scanning electron micrograph of UTD-1F (A) showing a typical bundle of fine needles and (B) a close-up of the ends of the needles.

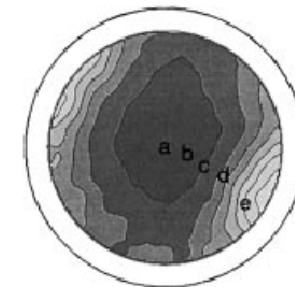


Figure 2. Pole figure for the 011 reflection. Tilt angle  $\chi$  increases along the radius ( $0^\circ$  at the center and  $90^\circ$  at the outermost circle), and the rotation angle  $\phi$  increases counterclockwise around the perimeter of the circle ( $0^\circ$  at the right). Intensity is proportional to the degree of grayness. Sample tilt and rotation angles selected for the collection of full diffraction patterns (see Figure 3) are indicated.

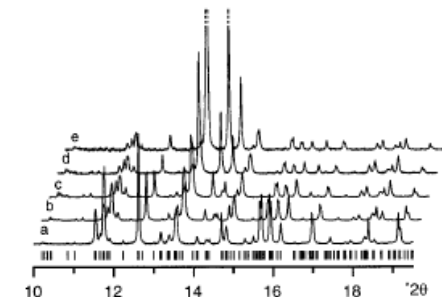


Figure 3. Small sections of the five diffraction patterns collected at different sample tilt ( $\chi$ ) and rotation ( $\phi$ ) angles. (a)  $\chi = 0^\circ$ ,  $\phi = 0^\circ$ ; (b)  $\chi = 20^\circ$ ,  $\phi = 345^\circ$ ; (c)  $\chi = 35^\circ$ ,  $\phi = 335^\circ$ ; (d)  $\chi = 50^\circ$ ,  $\phi = 335^\circ$ ; and (e)  $\chi = 70^\circ$ ,  $\phi = 325^\circ$ .

**MAS NMR Spectroscopy.** The  $^{19}\text{F}$  and  $^{29}\text{Si}$  MAS NMR spectra were recorded on a Bruker DSX 400 spectrometer.

**Powder Diffraction Data Collection for Structure Determination.** The needlelike crystallites were aligned by using shear forces in a polystyrene matrix, and the hardened sample was then polished and inserted into a circular flat-plate sample holder.

A texture goniometer was attached to the powder diffractometer on the Swiss-Norwegian Beamline (SNBL) at the European Synchrotron

Wessels T, Baerlocher C, McCusker LB, Creighton EJ,  
*J.A.C.S.* 121:6242 (1999)

Wessels T, Baerlocher C, McCusker LB  
*Science*, 284: 477 (1999)

# Anisotropic thermal expansion

- 1963
- 34 atoms in unit cell (I2/m)

*Acta Cryst.* (1963). 16, 369

## The Crystal Structure of Beta Plutonium Metal\*

BY W. H. ZACHARIASEN

*University of Chicago, Chicago, Illinois, Consultant to Los Alamos Scientific Laboratory, U.S.A.*

AND F. H. ELLINGER

*Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico, U.S.A.*

(Received 18 June 1962)

$\beta$ -Pu is monoclinic body-centered with 34 atoms per unit cell. At 190 °C

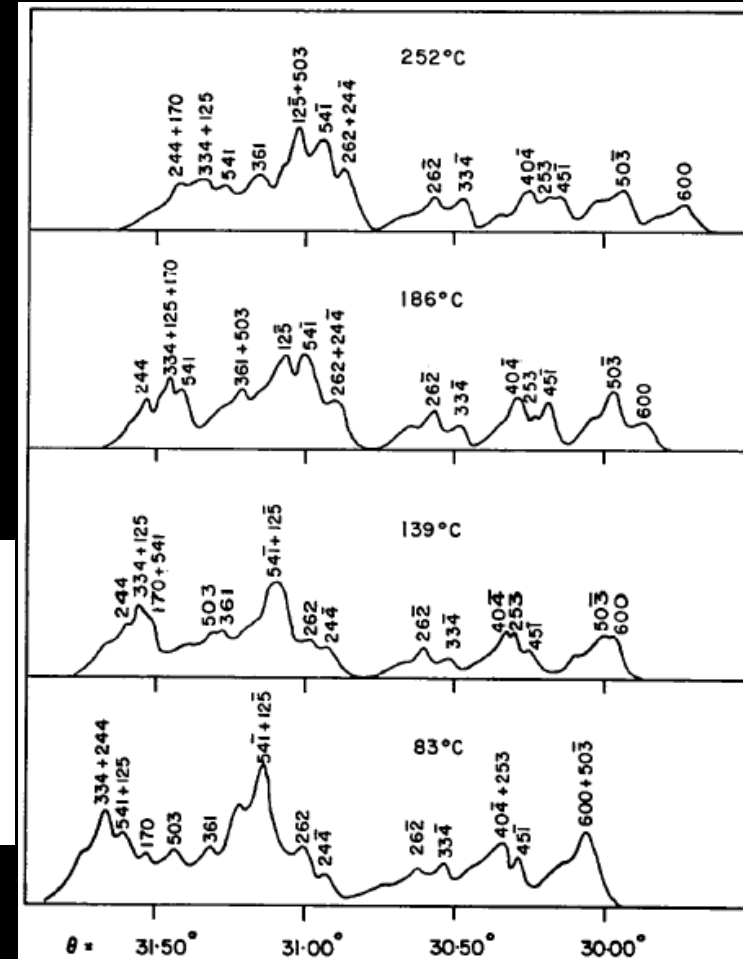


Fig. 1. Illustrates changes in the diffraction pattern due to the anisotropic thermal expansion.

# Revival/Reinvention of the method:

- J. Mat. Chem., 7 569 (1997).

---

**Routine *ab initio* structure determination of chlorothiazide by X-ray powder diffraction using optimised data collection and analysis strategies**

---

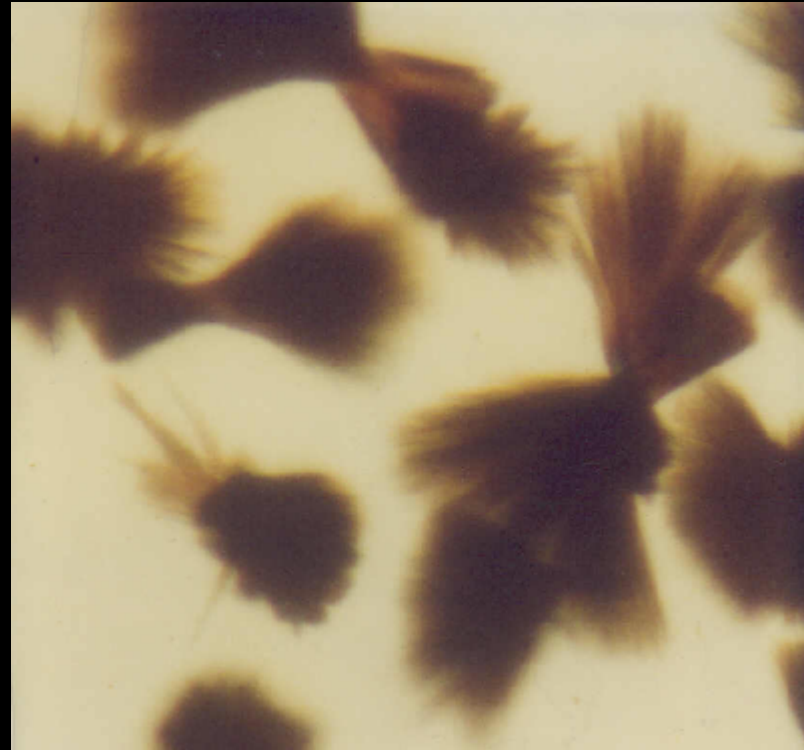
**Kenneth Shankland,\* William I. F. David and Devinderjit S. Sivia**

*ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, UK OX11 0QX*

- Used synchrotron radiation to get high resolution data
- Modern computing methods to treat data (pattern fitting!)

# Powder Methods + Proteins?

- Samples easier to produce... but,
  - Large unit cells
  - Severe peak overlap
  - Many degrees of freedom in the structure

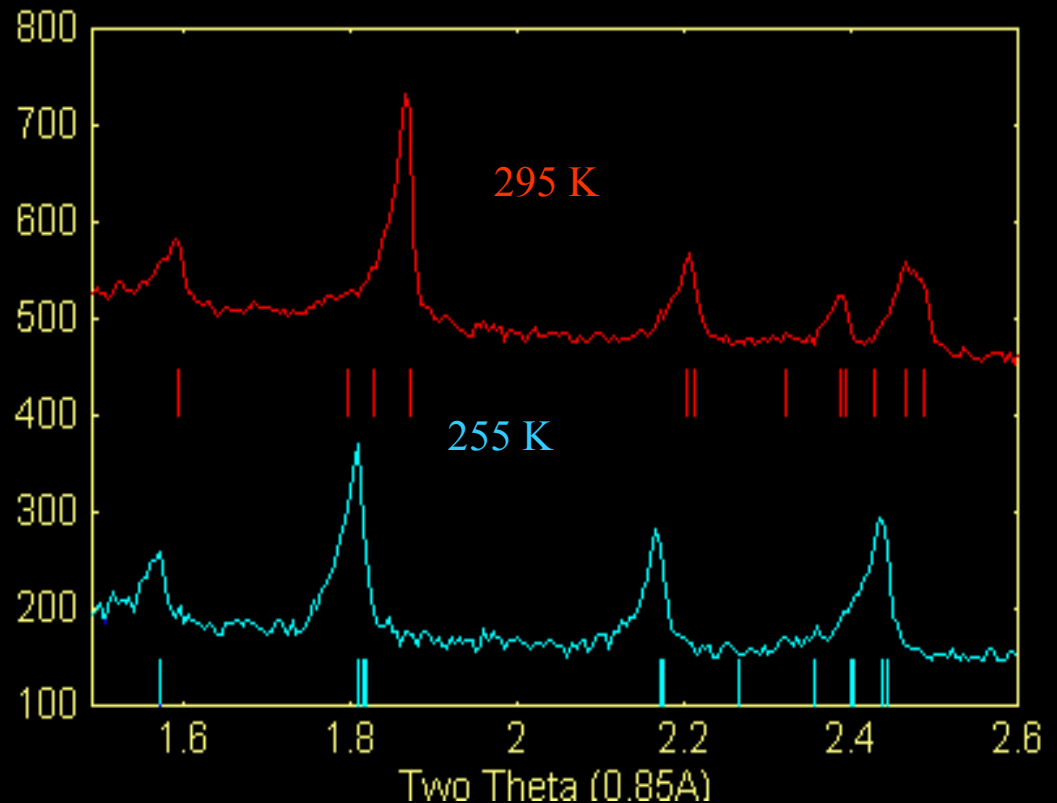


Myoglobin from horse skeletal muscle (Aldrich)  
recrystallised in 80% sat.  $(\text{NH}_4)_2\text{SO}_4$

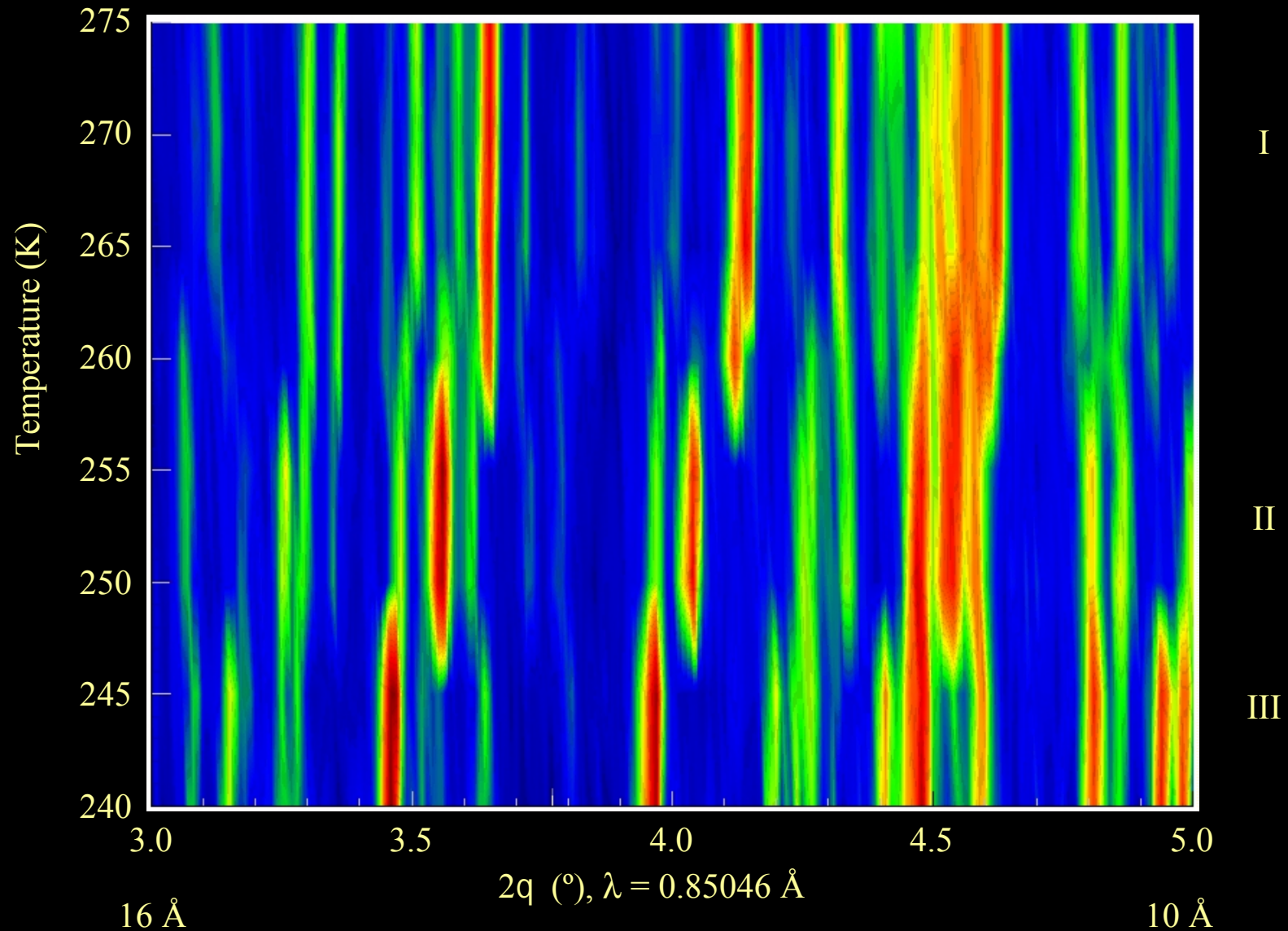


# Improving the data quality...

- Exploit anisotropic thermal expansion to alleviate peak overlap
- Lower temperatures for better data quality (B factor & radiation damage)
- Experiments carried out on BM16 (ESRF)

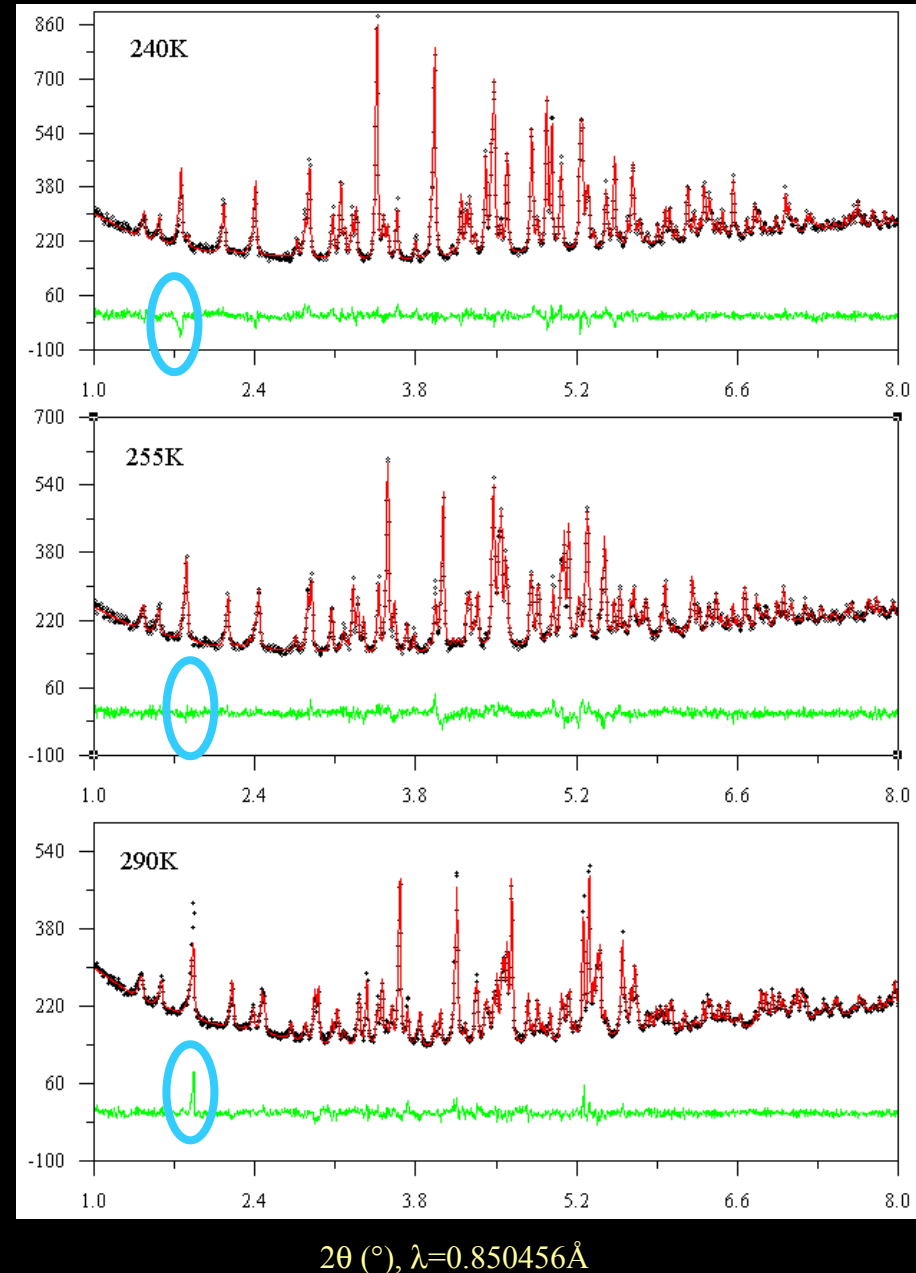


# Discrete phase transitions found...



# Combined fit for the 3 phases

- Multipattern fit with the same intensities for each
- Cell parameters and peak shape different for each pattern (indexed with DICVOL)
- Details later about pattern decompositions



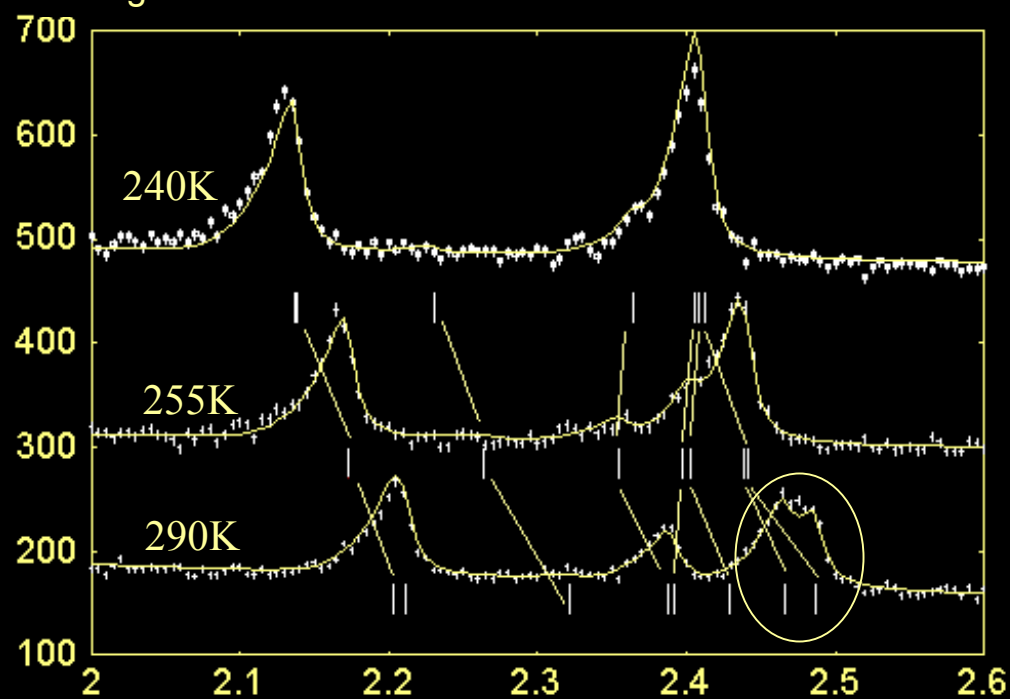
# Refinement Results

- Cell is very close to orthorhombic (a vs. [101])
- Mainly expansion along b axis
- Peaks are ~2 times instrumental width ( $\gamma_X \sim 0.006^\circ$ )
- Highlighted (-211),(111) monoclinic peaks to show distortion is resolved

(-211),(111) monoclinic =

(311),(-311) orthorhombic

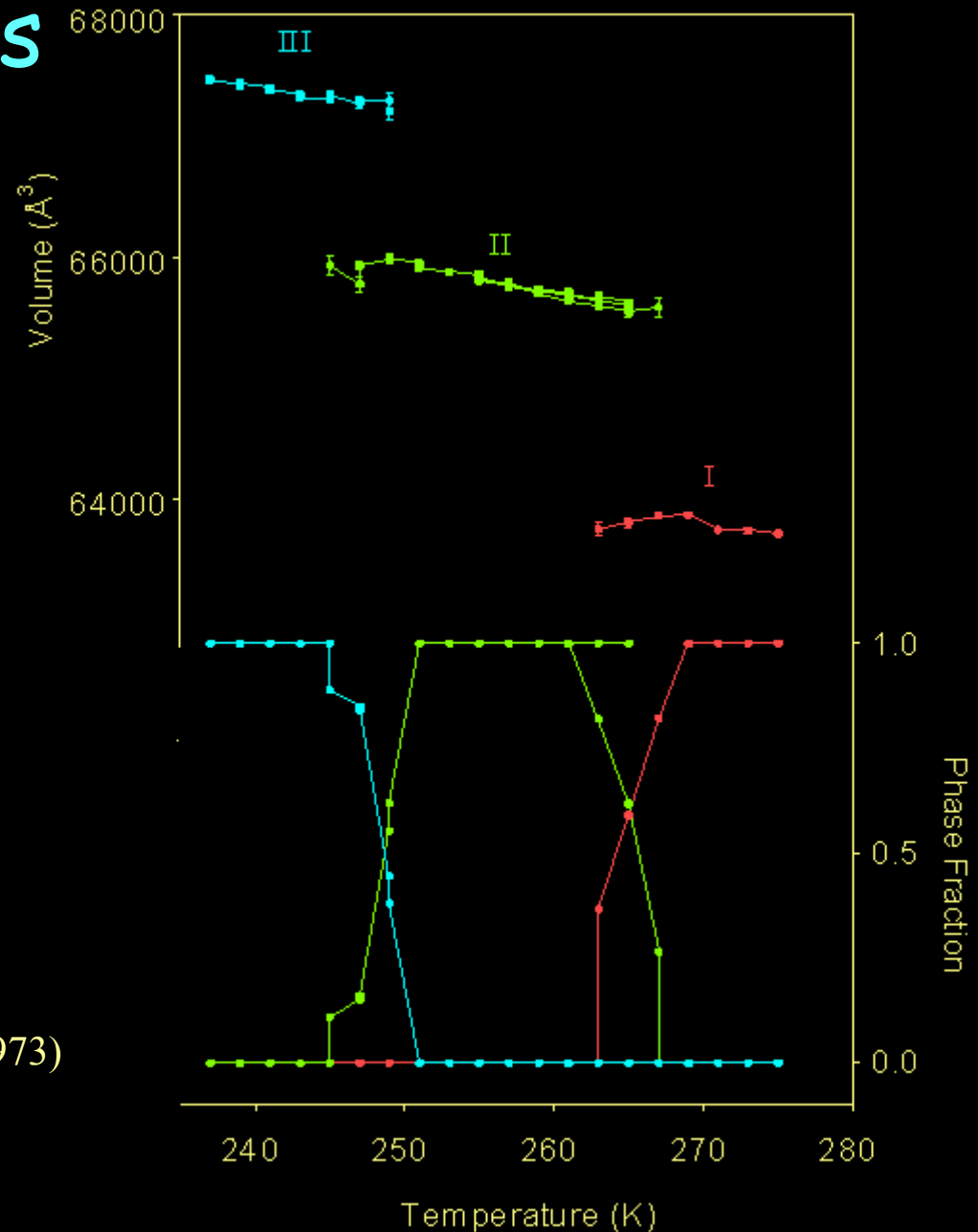
T/K	290K	255K	240K
a/Å	64.23	64.60	64.44
b/Å	28.89	29.96	30.94
c/Å	35.87	35.34	35.20
b/°	107.2	105.7	106.0
V/Å <sup>3</sup>	63592	65815	67457
[101]/Å	63.66	64.68	64.35
$\gamma_X/^\circ$	0.0149	0.0150	0.0133



# Cell volume versus temperature

- $\Delta V/V$  phase I>II is 3.5%
- $\Delta V/V$  phase II>III is 2.7%
- cf. Change in density for liquid water to hexagonal ice is ~10%
- Freezing at high pressure improves crystal quality [\*]
- Appears to show negative thermal expansion in both II and III.

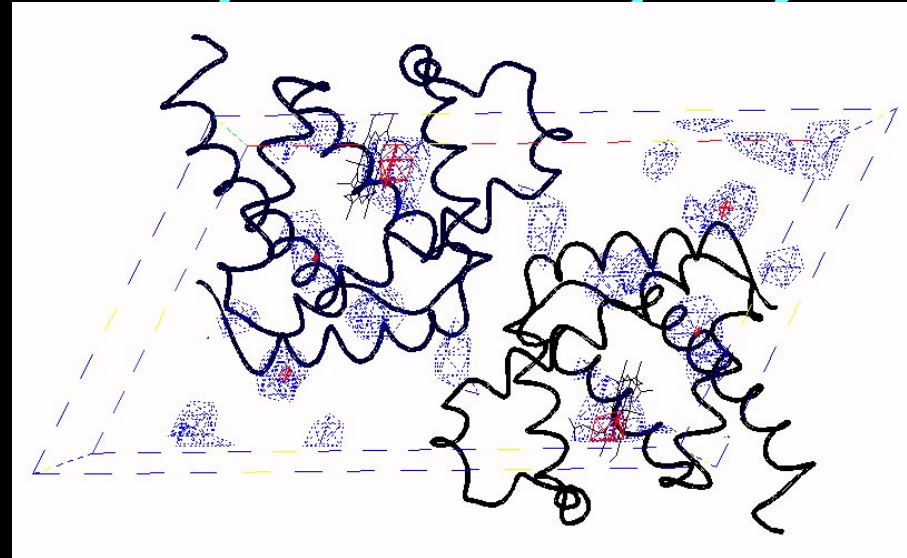
\* Thomanek *et al*, *Acta Cryst.* A29 263 (1973)



# MICE map

(C. Gilmore, University of Glasgow)

- Supplied integrated intensities to the program
- Fixed origin definition
- Eventually realised the sample is contrast matched



- Heme group can be located (!!)

# Methods for pattern decomposition

- Le Bail extraction
  - Iterative procedure. Produces one set of peak intensities which fit the powder pattern. Computationally cheap, sometimes unstable, no information about other plausible solutions.
- Pawley refinement
  - Fits the intensities as least squares variables. Soft and hard restraints improve very poor stability. Computationally expensive, but provides full information about covariance (few programs and these were limited to several hundred peaks)
- Materials studio
  - Fits structure factors,  $F = \sqrt{I}$ , to avoid negative peaks. Uses conjugate gradients for non-linear optimisation.
- Others (nowadays TOPAS at least)

(I) couldn't deal with the protein data!

# What is least squares fitting?

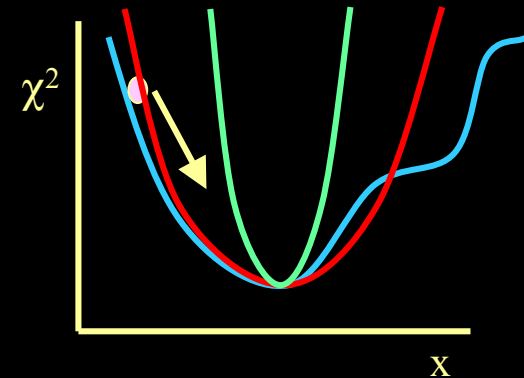
- Apologies to those who already know...
- Minimise the sum of the **squared differences** between model and data.
- Conventionally:

$$\chi^2 = \sum_{i=1}^m \frac{(y_{o,i} - y_{c,i})^2}{s_{ii}^2}$$

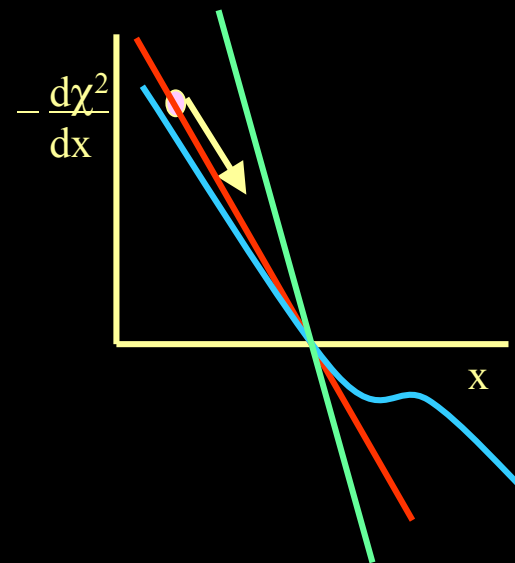
- Minimum is at  $dc^2/dx = 0$
- Linear model  $\Rightarrow$  solve for  $x$  in one step.
- Non-linear models can be approximated by a Taylor expansion and treated as linear (provided  $x$  is close to  $a$ )

$$f(x) \approx f(a) + f'(a)(x-a) + \dots$$

- For  $n$  parameters this means solving matrix equations with either an  $n \times n$  **matrix** or an  $n \times m$  **matrix**.



$\chi^2$  function is quadratic when model is linear



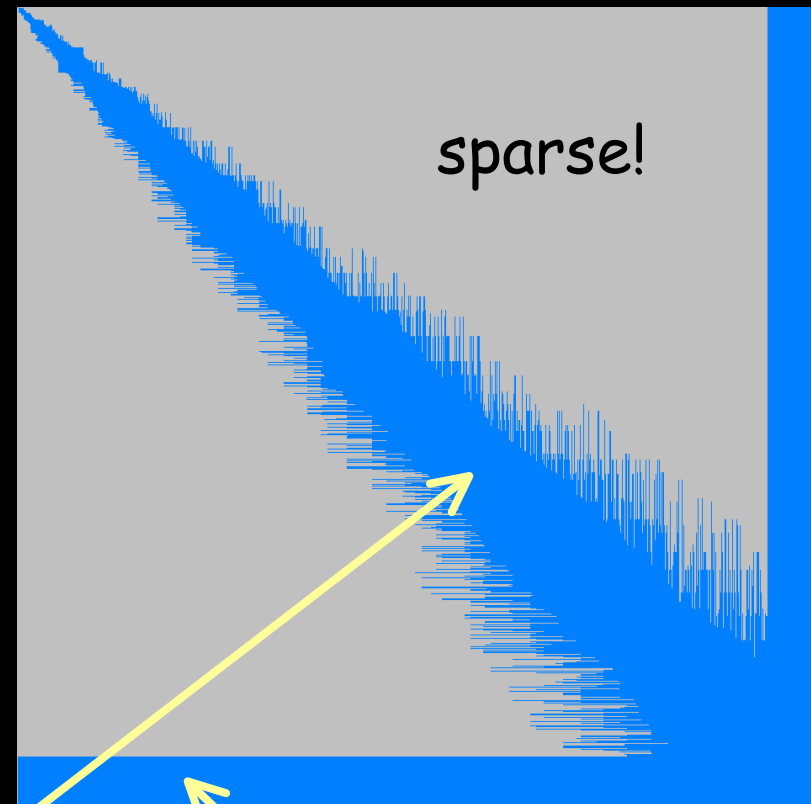
Error on fitted parameter is related to the second derivate of  $\chi^2$  (green is more precise than red)



# Intensity extractions

- Used (modified) PRODD computer program
- Does something like a Pawley fit to multiple patterns with different cell parameters
- Preserves correlation information

Least squares matrix



Intensities

cell, peakshape,  
background

# Algorithm:

x - variables, b - data

Problem:  $\mathbf{Ax} = \mathbf{b}$  ( $\mathbf{A}$  is singular)

← Singular means we can't solve for x

$(\mathbf{A} + \lambda \mathbf{I})\mathbf{x} = \mathbf{b}$  (Modify A)

← ... so we modify the problem!

$(\mathbf{L}^T \mathbf{L})\mathbf{x} = \mathbf{b}$  (Slow step)

← Some jiggery pokery to get at x

Solve:  $\mathbf{L}^T \mathbf{y} = \mathbf{b}$  and then  $\mathbf{L} \mathbf{x} = \mathbf{y}$

Truncate x to positive intensities

Use x to compute  $\mathbf{b}' = (\mathbf{b} - \mathbf{A} \mathbf{x})$

← Is x a solution to the original problem?

if  $\mathbf{b}'$  is not zero then:

    solve:  $\mathbf{L}^T \mathbf{y} = \mathbf{b}'$  and then  $\mathbf{L} \mathbf{x}' = \mathbf{y}$

← If not then solve for the residual and use that to get a better x.

    Update  $\mathbf{x} = \mathbf{x} + \mathbf{x}'$ , truncate, iterate

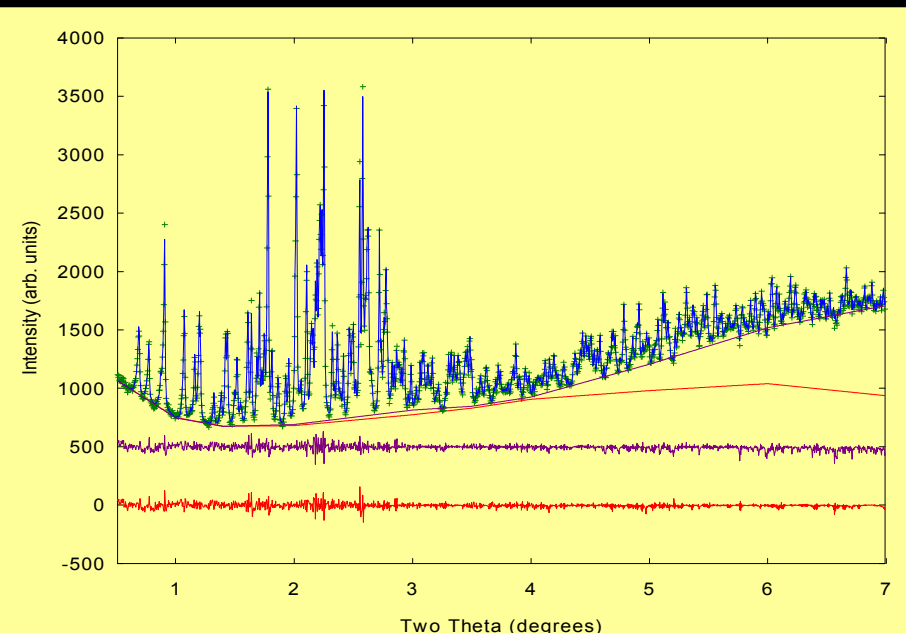
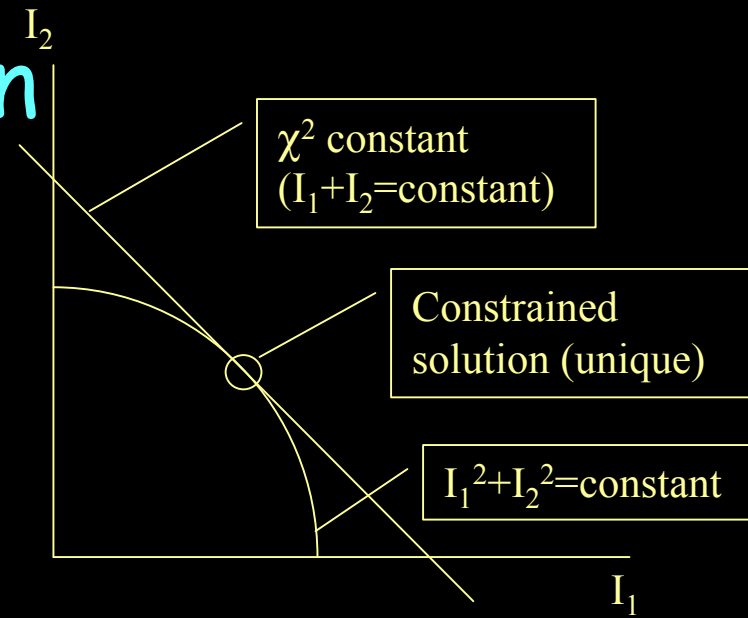
See: George and Liu, "Computer Solution of Sparse Positive Definite Systems"

# Choosing a unique solution

- Overlapped peaks can be partitioned arbitrarily
- Origin of singular matrix
- Tendency to fit noise
- Select a unique solution by minimising

$$C^2 = \sum_i \frac{(y_{o,i} - y_{c,i})^2}{S_{ii}} + \lambda \sum_j I_j^2$$

- Relative weighting remains as a tunable parameter
- Equipartitions intensity



# Why??

- These matrices are "easy" to form
- Allows the intensity information to be used without referring to the pattern(s)
- Incorporates peak overlap

$$C^2 = \sum_{ij} (I_{o,i} - I_{c,i}) A_{ij} (I_{o,j} - I_{c,j})$$

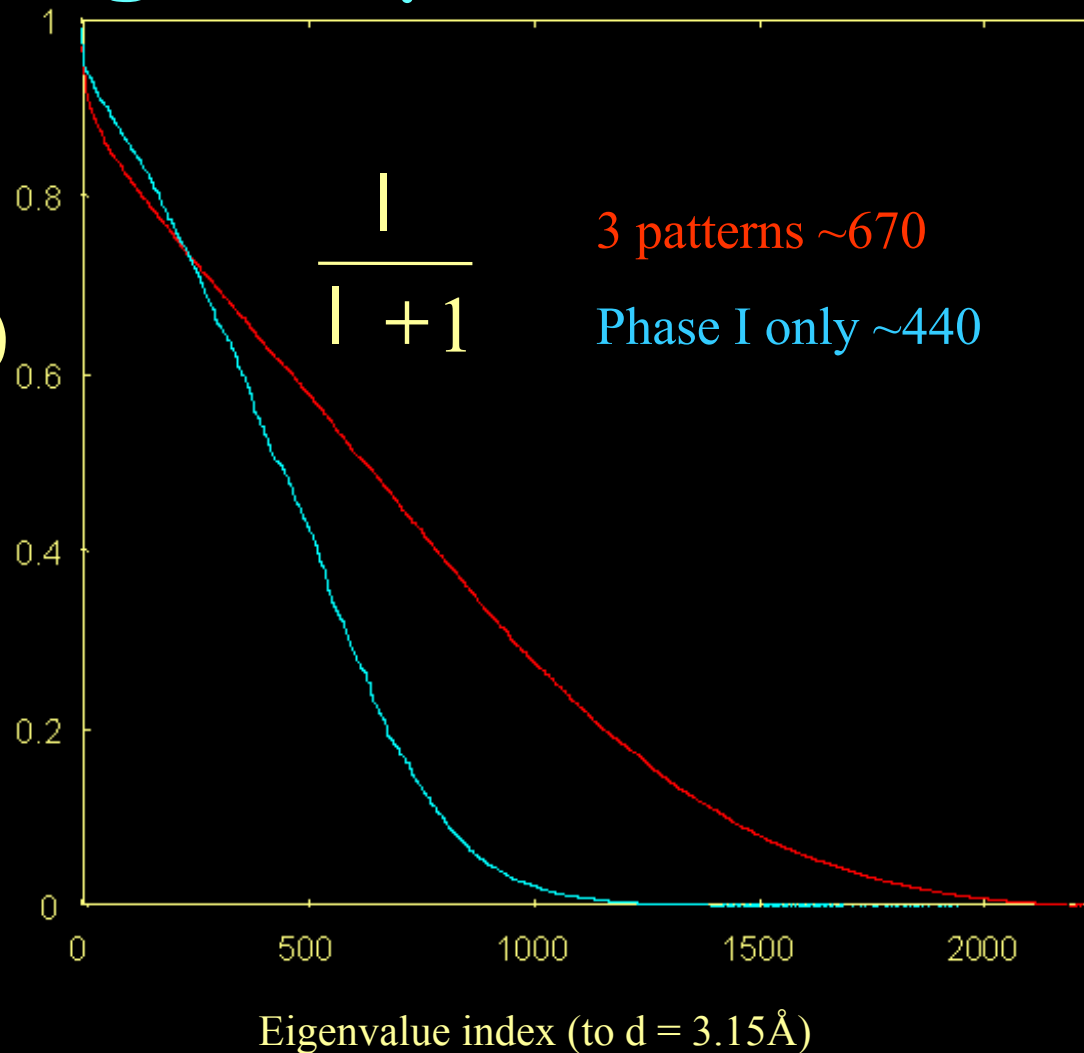
$$A_{ij} = \text{diagonal} = \frac{1}{S_{ii}^2}$$

$$C^2 = \sum_{ij} \frac{(I_{o,j} - I_{c,j})^2}{S_{ii}^2}$$

- Reduces to single crystal case when there is no peak overlap
- example
  - DASH performance

# Number of good peaks?

- Attempt to estimate the real information content in the data, with approximations (fixed background etc)
- Single crystal data would be a ~flat line with tail off for  $I \sim S^2$
- There are no "isolated" peaks
- This estimate ignores the measurement precision



# Eigenvalues/Eigenvectors

- Two exactly overlapping peaks,  $I_1$  and  $I_2$
- Eigenvectors would be
  - sum:  $I_1+I_2$
  - diff:  $I_1-I_2$
- With lots of peaks, we get uncorrelated linear combinations
- 1/eigenvalues of matrix are esds on the eigenvectors
  - 1/0 is undefined, which is the problem with inverting the matrix!
- Exact overlap
  - sum: a large number
  - difference: zero
- Inexact overlap
  - difference: small number

# Going further?

- If all of the non linear parameters are optimised the matrix faithfully represents the data
- Just a linear transformation
- Can exploit the eigenvalues and eigenvectors

$$C^2 = \sum_{ij} (I_{o,i} - I_{c,i}) A_{ij} (I_{o,j} - I_{c,j})$$

- Any further processing is independent of the raw data
- Faithfully represents errors
- "Non diagonal weight matrix"

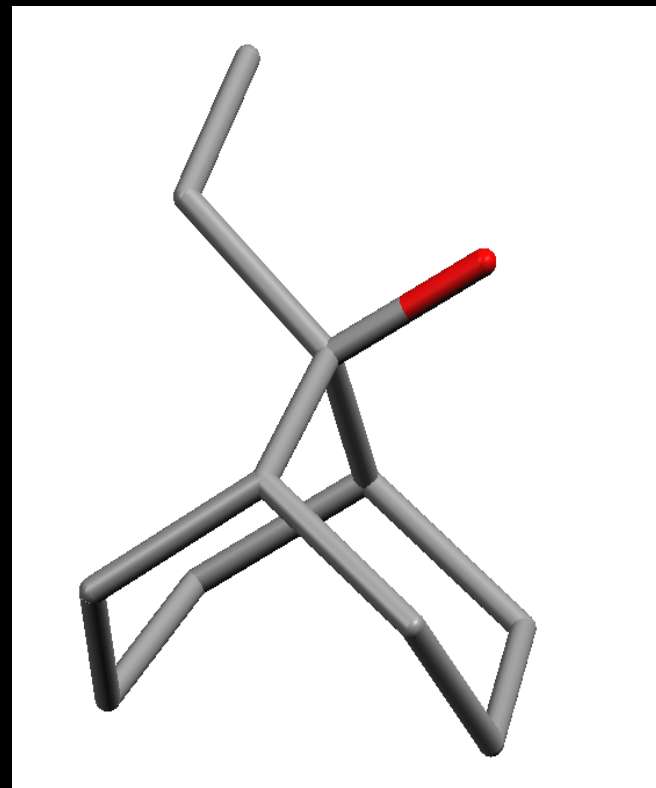
# Applications for that matrix?

- Generating other sets of intensities which fit the pattern
- Optimising some "other functions" of intensities
  - entropy
  - Wilson type statistics
  - etc!
- Refinement separated from pattern fitting
  - view as a "funny system of constraints"
  - Intensities are variables with constraints on them defined by the structure

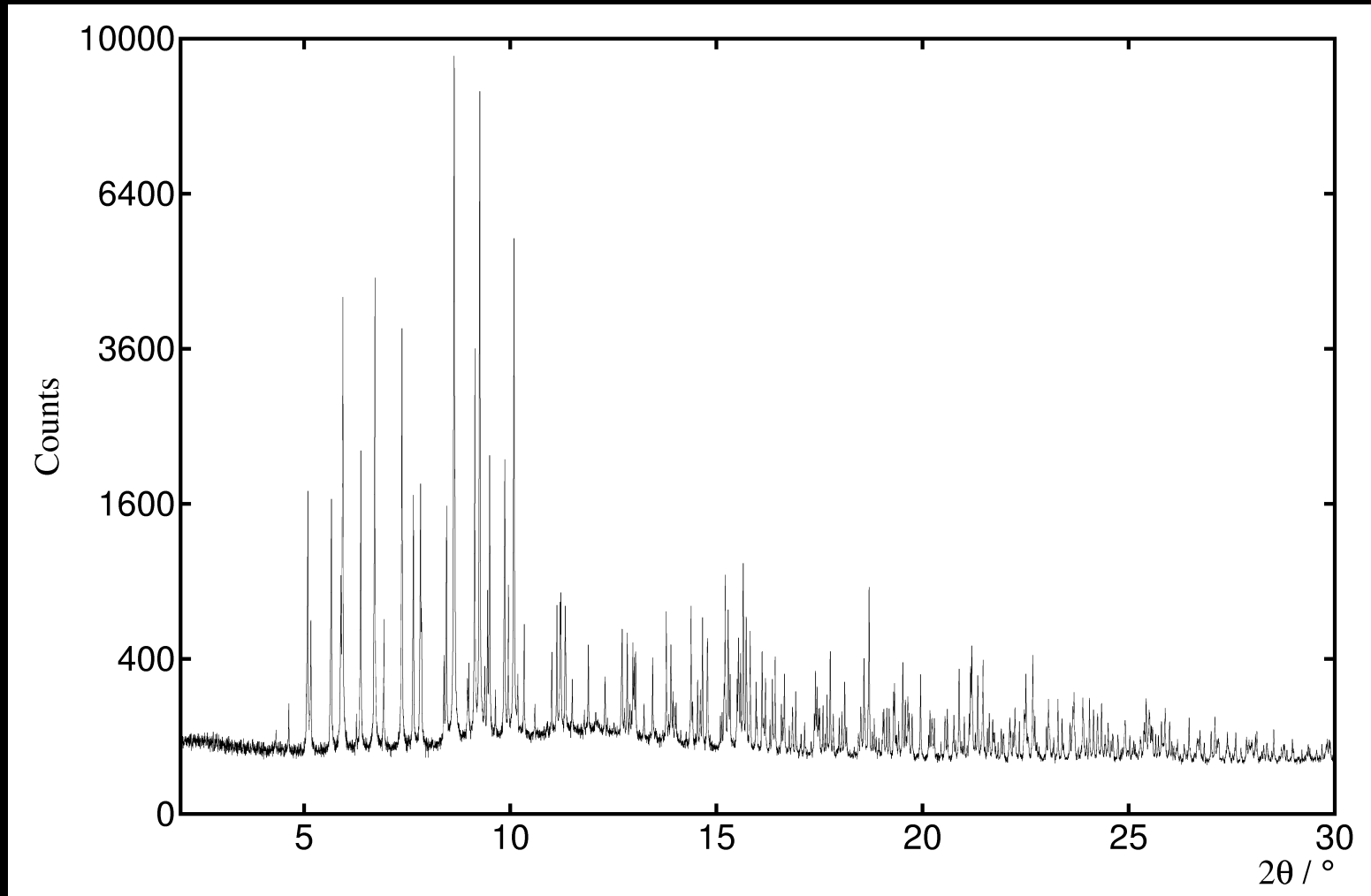


# A "small molecule" application:

- 9-ethylbicyclo[3.3.1]nona-9-ol
- $C_{11}H_{19}OH$
- Small globular organic molecule
- Part of a wider study
  - M Brunelli, PhD thesis
- Turns out to behave anomalously



# BM16 powder data



# Why is it anomalous?

- Indexed as orthorhombic from BM16:  
 $a=21.080 \text{ \AA}$ ,  $b=23.074 \text{ \AA}$ ,  $c=16.426 \text{ \AA}$   
(cell volume =  $7990 \text{ \AA}^3$ )
- Space group Pbc<sub>2</sub>a
- $Z = 32$ ;  $Z' = 4$ ; 48 atoms in asymmetric unit (384 in unit cell!)
- Large cell + No phase transitions or disordered phases

# Ideal candidate

- Diffracts beyond  $1\text{\AA}$
- Direct methods should work given good enough data
- Various attempts to solve the structure had failed
- Just need to collect sufficiently good data

# Improvements to ESRF powder diffraction beamline

- Moved the diffractometer from a bending magnet source (BM16) to undulator (ID31)
- Large increase in flux
- Multiple datasets can be collected in a reasonable time

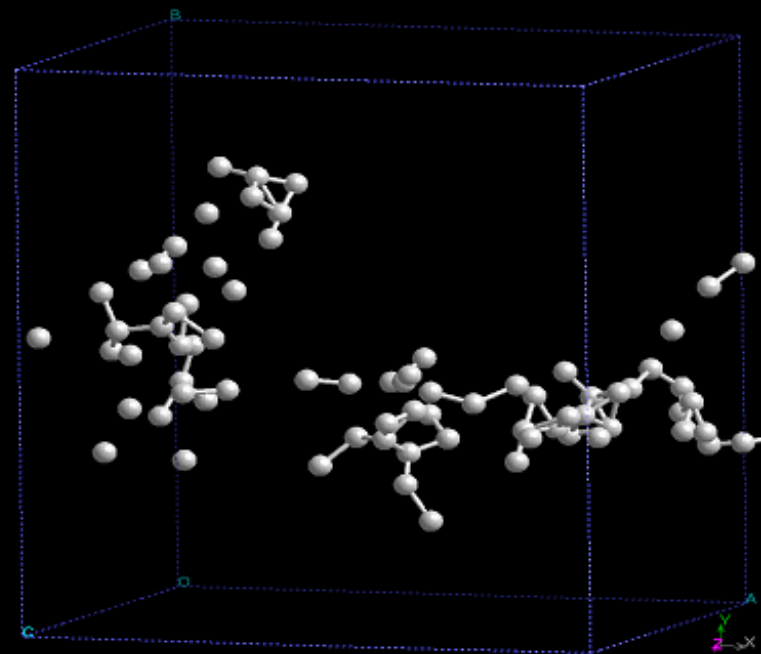


Opened to users 26 June 2002



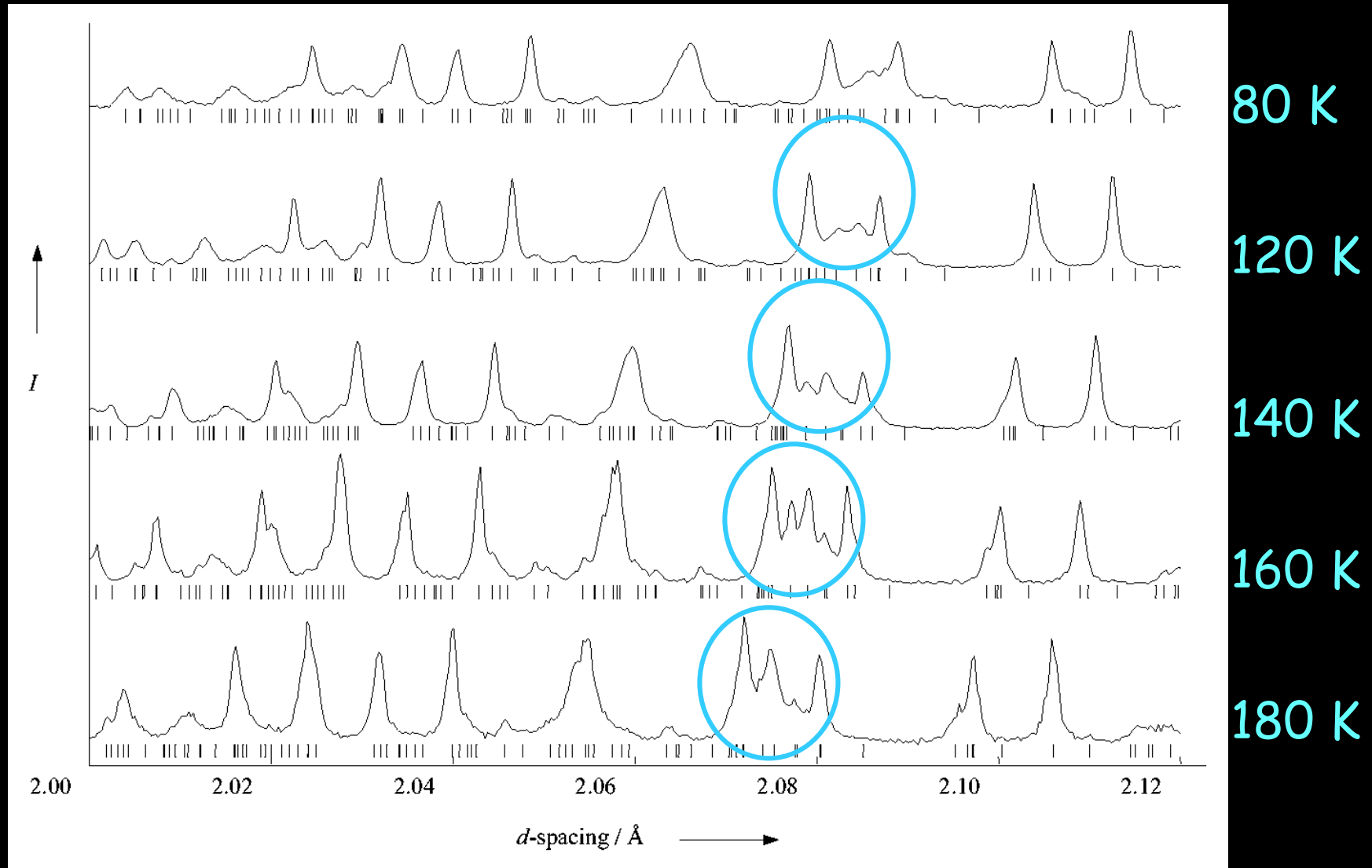
# First user experiment on ID31

- Flexible user group!
- One day of beamtime
- First attempt with 2 data sets and feeding intensities to Shelxs direct methods was unsuccessful.
- Further data sets at 140 K, 160 K and 180 K were collected.



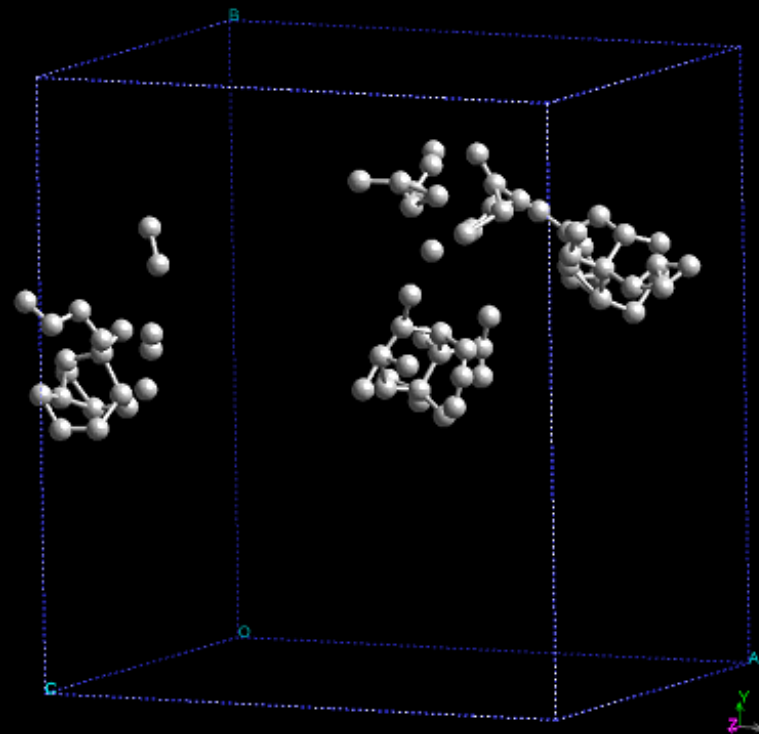
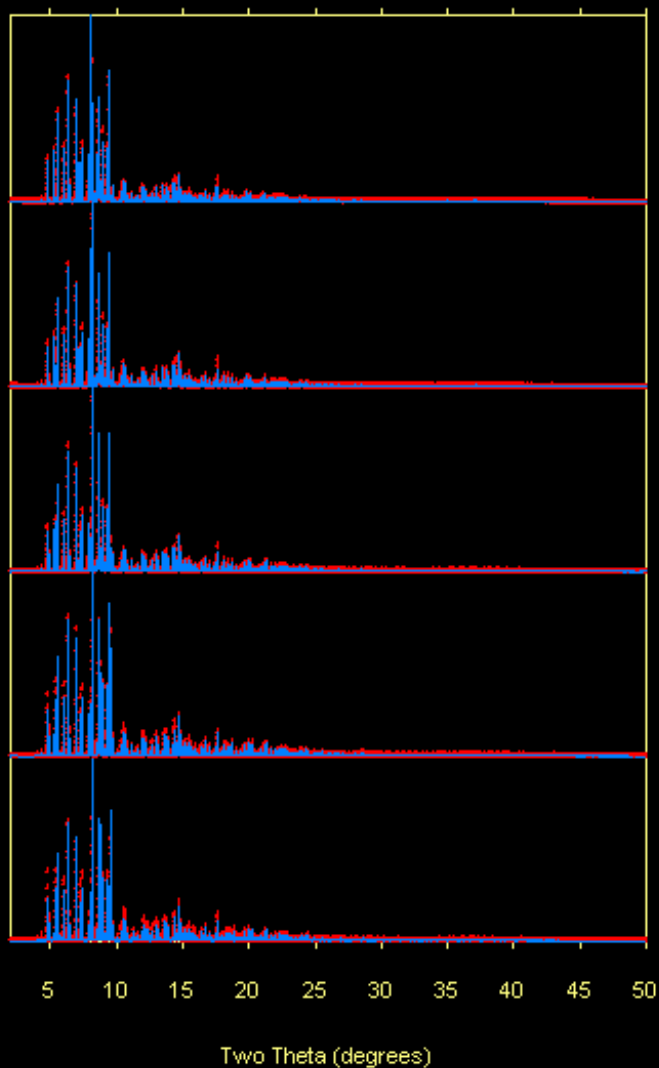
Initial structure was not interpretable with 2 datasets

# Extracts of the diffraction patterns at the different temperatures





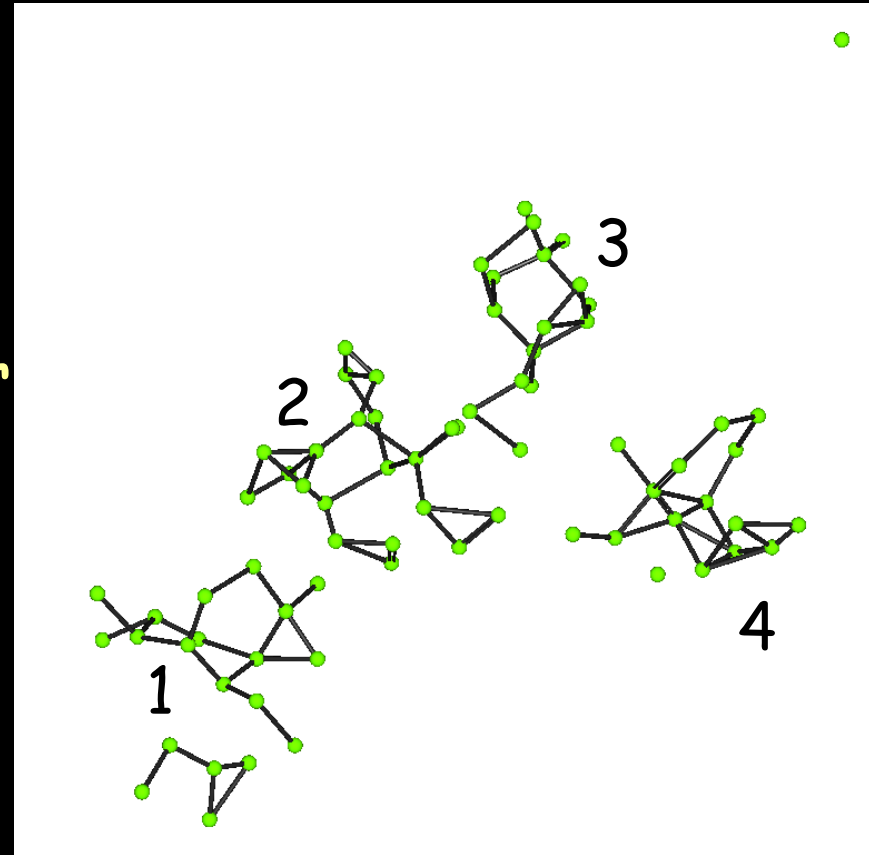
# Complications not needed?



3 molecules identified from initial Shelxs direct methods solution when all 5 patterns were used.

# Thermal factors....

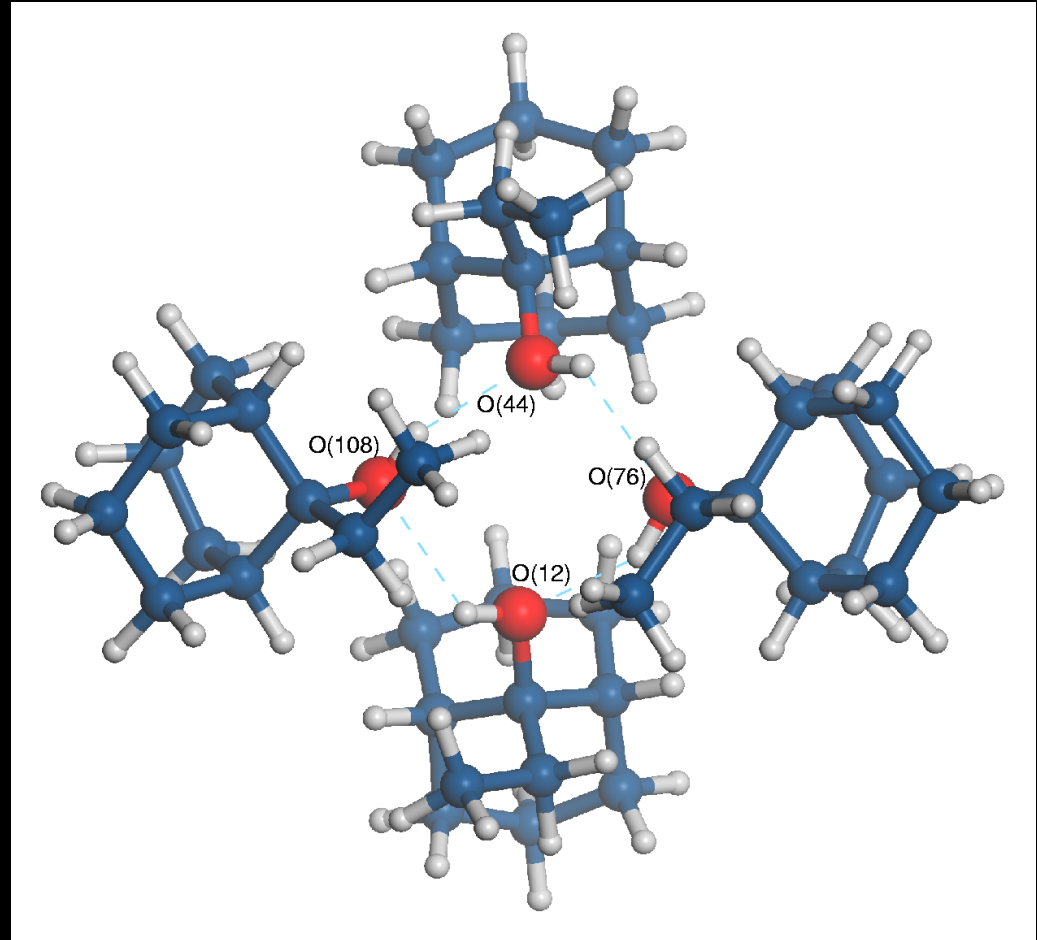
- No correction for thermal motion in the original intensity extraction
- The program was later fixed to include an overall isotropic thermal factor
- Improvement on SX R-factor was:
  - $R_1 \sim 48\% \rightarrow 41\%$



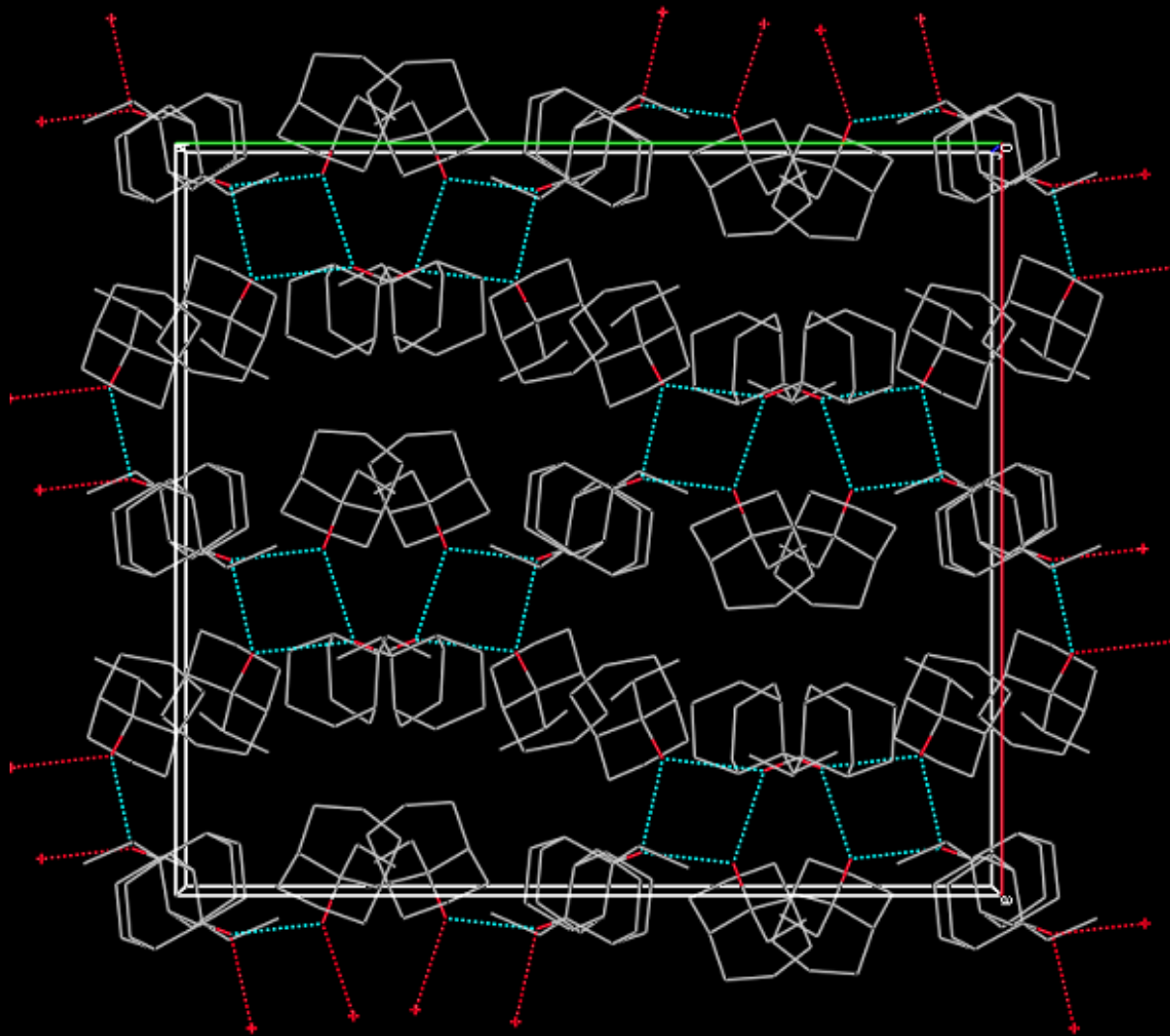
All four molecules appear on the initial E map!

# Structure completion

- Structure completed by fourier recycling against extracted intensities
  - xp (Bruker shelxtl)
- No need to repartition the overlaps!!!
  - "single crystal"  
R-factor was poor

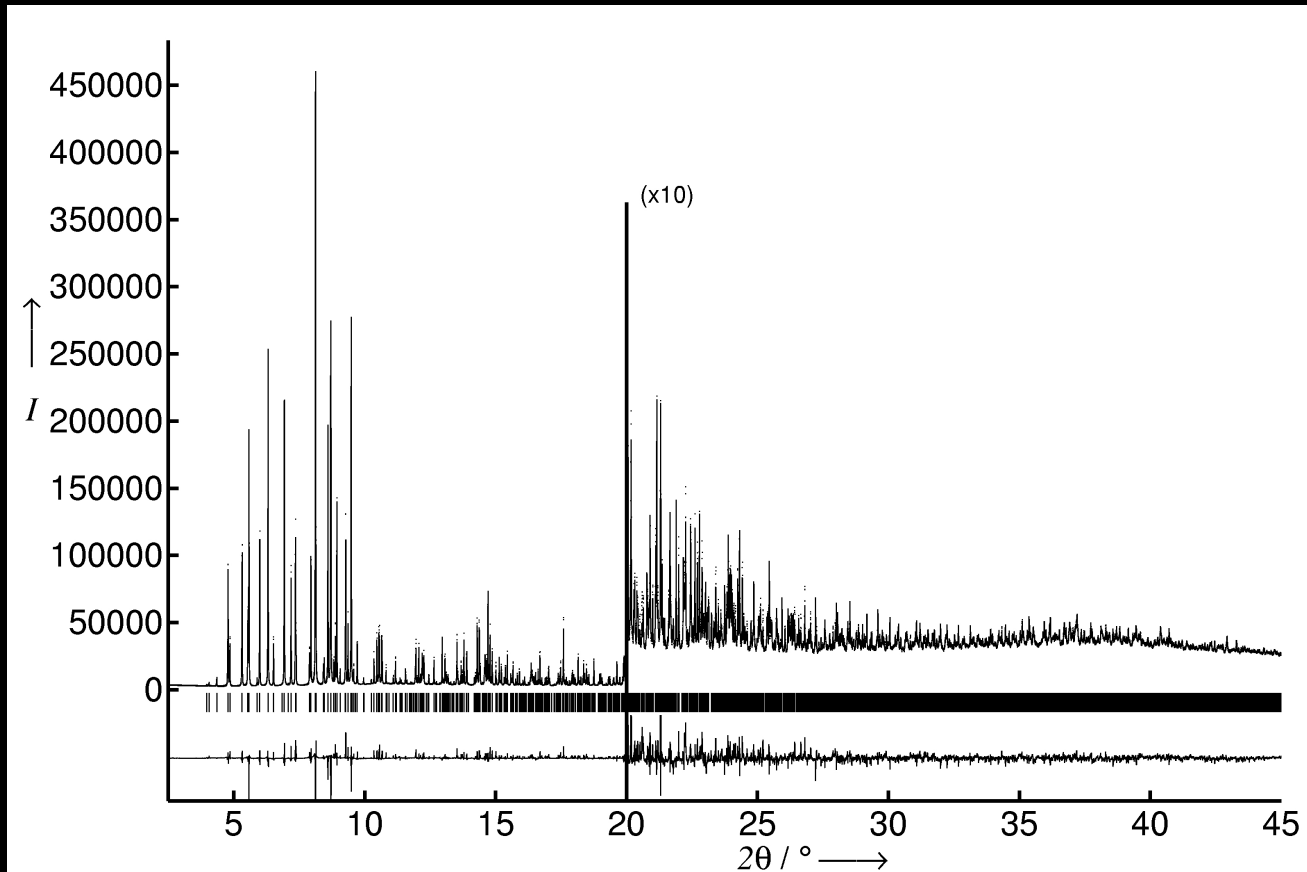


# Packing is sensible



- No bad contacts identified
- H-bonded tetramers
- Van-der-Waals interactions between tetramers

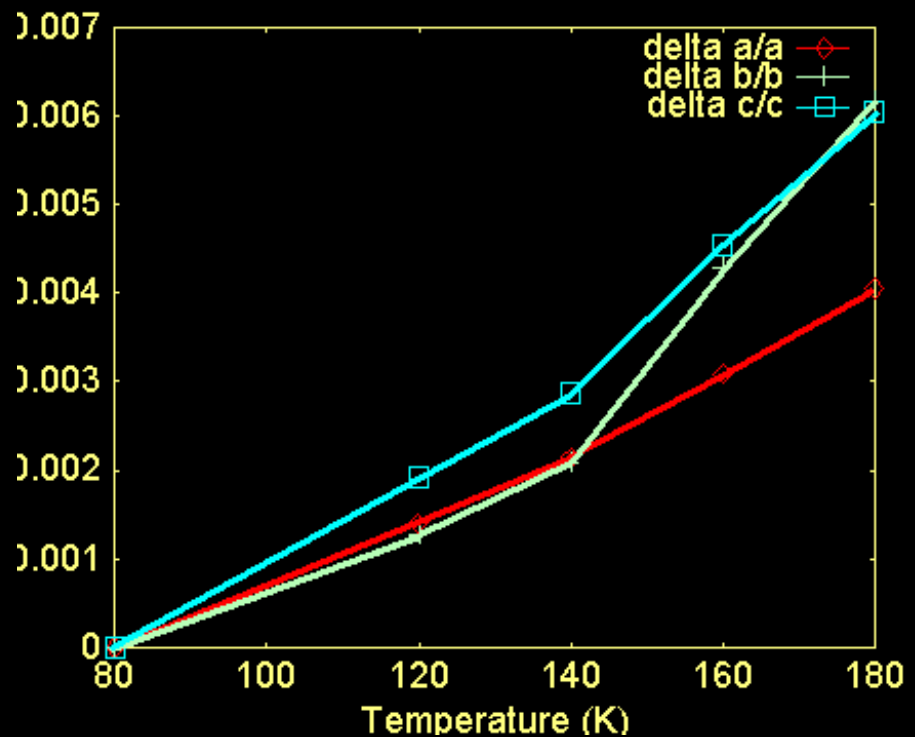
# Final Rietveld fit at 140K



- Stereochemical restraints (GSAS)
- One ethyl group apparently disordered
- $R_{wp} = 5.3\%$ ,
- $R_{F^2} = 7.5\%$
- $\chi^2 = 16.5$

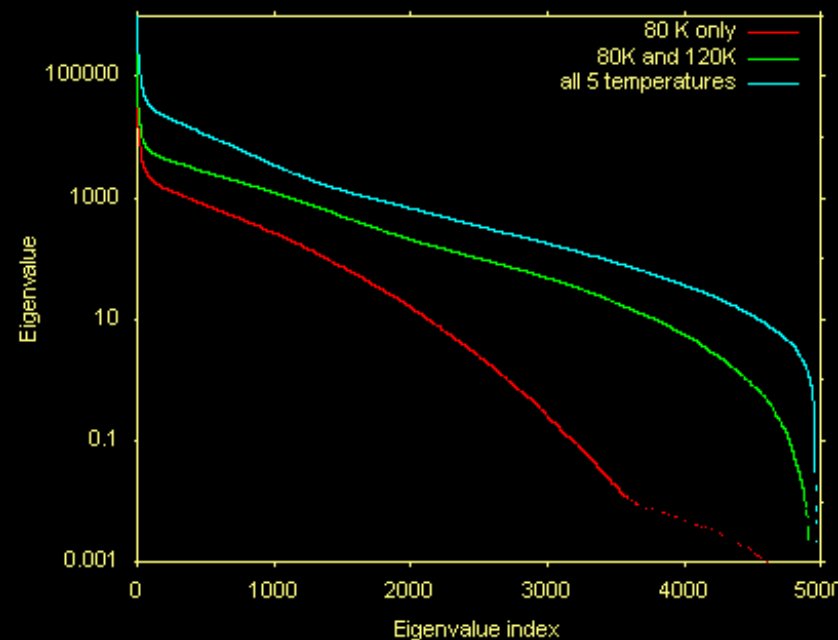
# Why success with 5 but not 2 temperatures?

- Counting for longer improves the statistics (total  $\approx 12$  hours).
- Variation of the cell parameters brought in significantly more information in later patterns
- Radiation damage might have helped by making the expansion more anisotropic?



# How many peaks were present?

- Eigenvalue spectra from the normal matrix of the Pawley fit for 80 K, 80 and 120 K and all 5 patterns.
- Difficult to assign a single number
- D.S.Sivia, J. Appl. Cryst. (2000). 33, 1295-1301



Using data to  $0.95 \text{ \AA}$   
a possible 2400 peaks exist up to  $1.2 \text{ \AA}$   
a possible 4100 peaks exist up to  $1.0 \text{ \AA}$

# Shelxs "peak counting"

- List file gives intensity statistics based on hkl file (esds are from Le Bail's algorithm when preparing hkl file)

$2\theta / ^\circ$	0	9.2	13.1	18.4	23.1	27.2	30.9	33.2	35.8	38.9	42.6	47.2	52.8	
resolution / Å	∞	5.00	3.50	2.50	2.00	1.70	1.50	1.40	1.30	1.20	1.10	1.00	0.90	
$N_{\text{observed}}$	29	63	141	214	238	239	205	284	401	536	585	228	(3163)	
$N_{\text{measured}}$	29	64	167	262	324	378	287	376	523	716	1044	892	(5062)	
$N_{\text{theory}}$	32	65	168	262	324	378	287	376	523	716	1044	1557		

- Half of the theoretically possible reflections in 1.1-1.2Å range are needed for direct methods to work (Sheldrick's rule)



# Outlook

- Suitable data for ATE can be collected on a reasonable timescale at ESRF's new powder diffraction beamline
- Samples which diffract "sufficiently well" should be tractable
  - sharp peaks
  - peaks at high angles (~atomic resolution)
  - low enough symmetry for anisotropic expansion
- Even difficult structures can be solved easily with good enough data (Single Crystals !!!)
- Simple data processing

# Outlook...

- Structure solution was "unsophisticated"
- Overlap information was extracted, but was not used
- Perhaps EXPO (Bari group) can be adapted to treat multiple datasets?
- Improved algorithms (past+future) should solve more difficult structures with improved data quality

# Acknowledgements

- M. Brunelli
- G. Vaughan
- A. Mora
- A. Fitch
- ESRF beam-line support groups
- Chris Gilmore
- Bob Von Dreele

Small molecule example:

Angew. Chem. Int. Ed (2003) 42 2029-2032