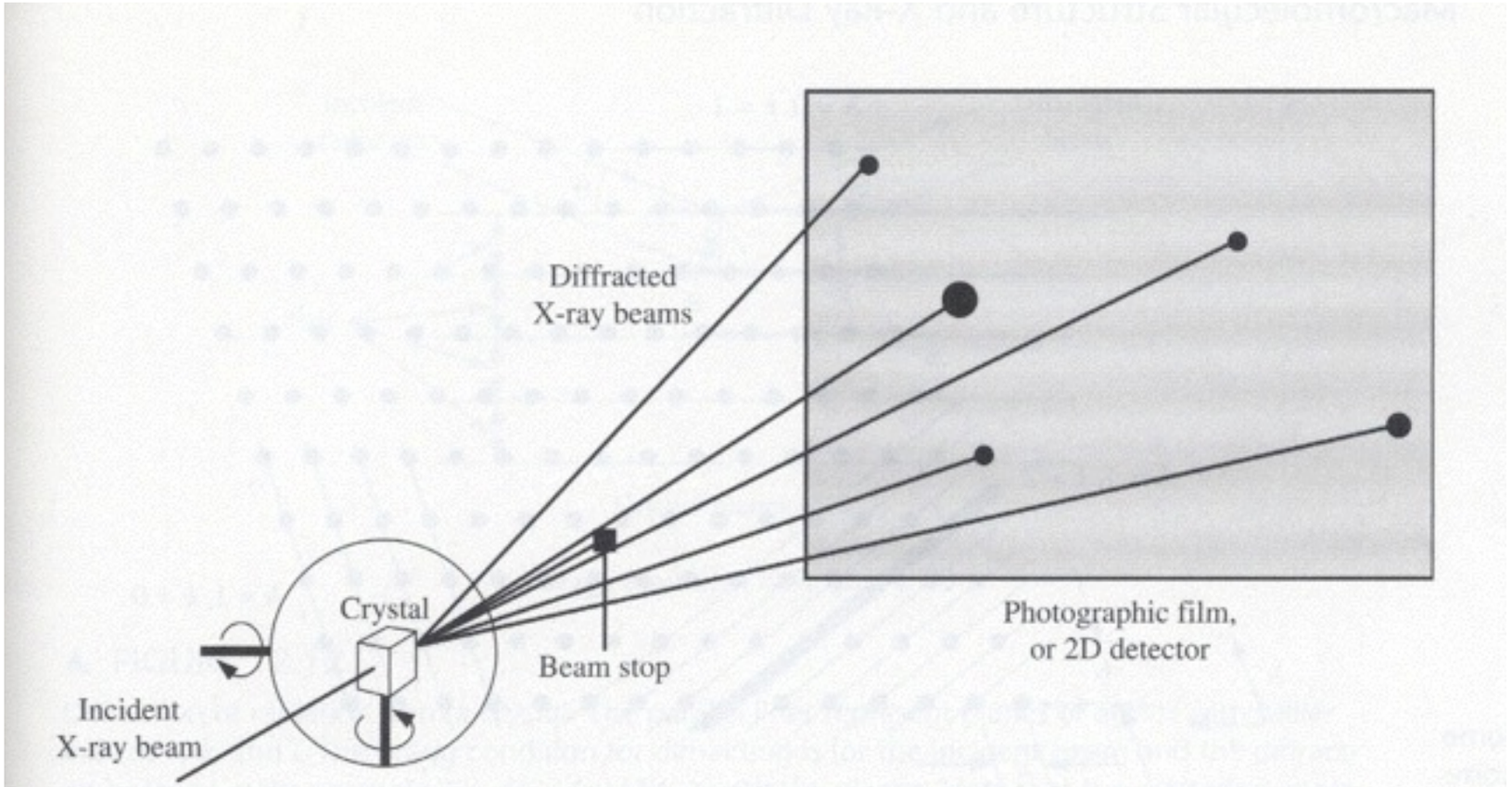
Data Collection and Resolution

Data Collection

- Measured intensity has to be corrected for background and other effects
- X-ray polarization, extinction and absorption effects and radiation damage
- Data is collected by oscillating the crystal during exposure. Usually 1° per exposure.
- How much data to collect? Depends on symmetry.
- Higher symmetry the less data to collect. Why?

Experimental Setup

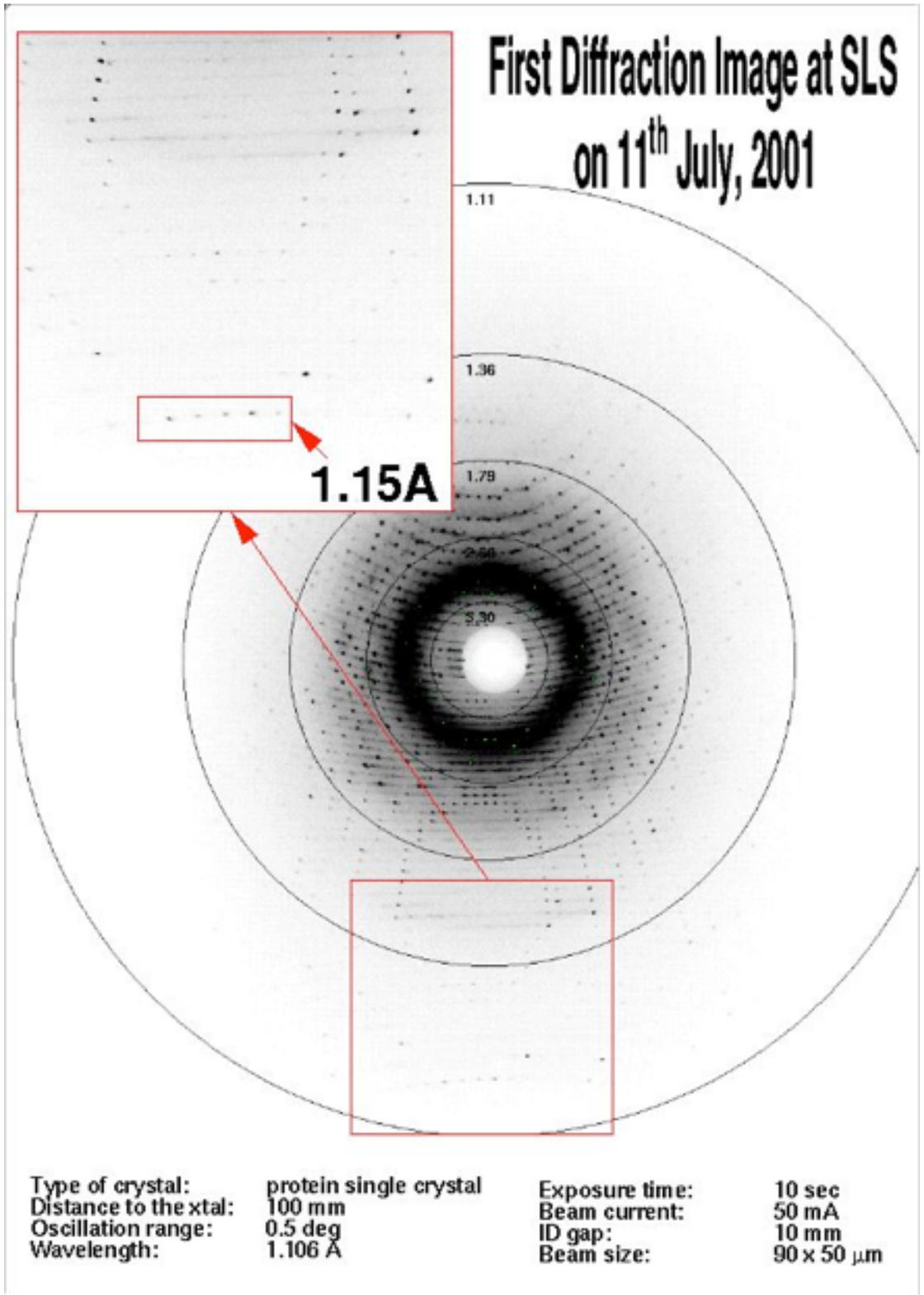


▲ FIGURE 12.10

Experimental arrangement for measuring the X-ray scattering from a crystal. A beam of X rays is incident on a crystal that is mounted so that it can be rotated around perpendicular axes. The intensities of the diffracted X-ray beams are measured with a photographic film or an equivalent imaging plate. A beam stop is used to attenuate the incident X-ray beam.

Experimental parameters are wavelength and crystal-to-detector distance.

Diffraction and Resolution

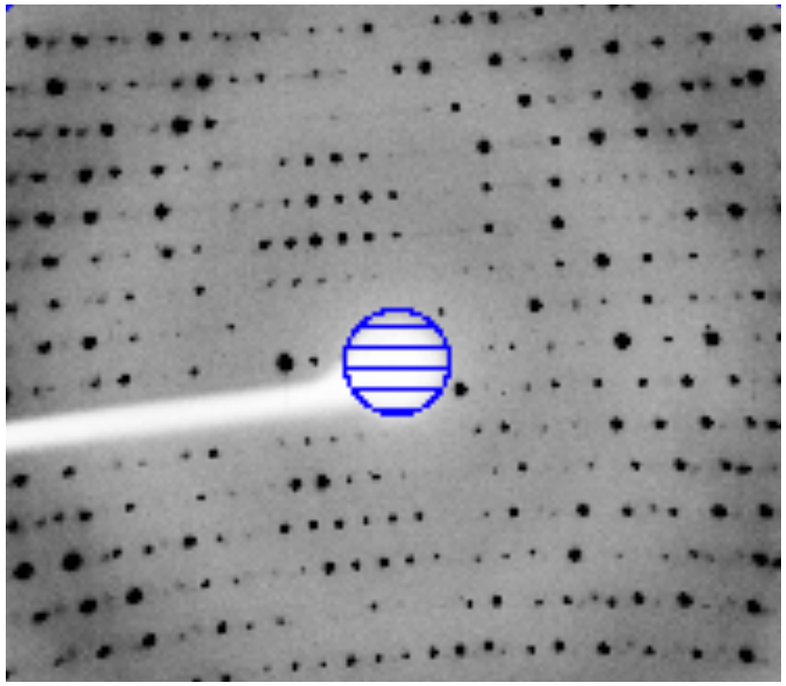


- As 2θ (distance away from the direct beam) increases, the number of reflections increases exponentially
- Every reflection contains information for the entire protein
- This allows for many more observations of the molecule and increased resolution

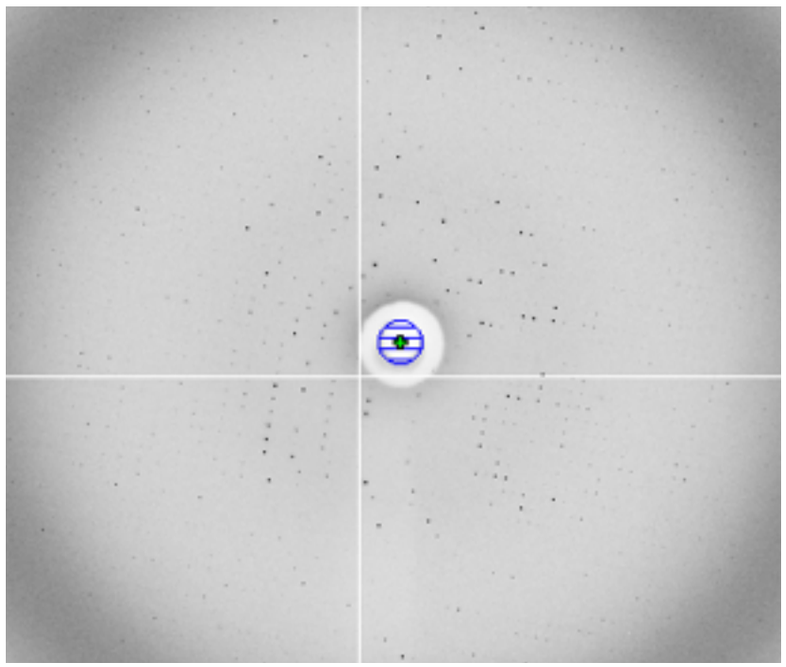
Intensity of Reflections

- For theoretical point scatter (a point that scatters X-rays in all directions) the intensity of the reflection is the same regardless of angle
- For real atoms the intensity of the reflection is a function of $(\sin\theta)/\lambda$
- Can predict the diffraction pattern if we know the lattice (space group), unit cell dimensions, distance from crystal to detector, orientation of lattice relative to X-ray beam
- BUT the intensity of the reflection depends on the composition and orientation of molecules in the asymmetric unit
- Data collection measure intensity only. Phases need to be estimated.

Reciprocal Space



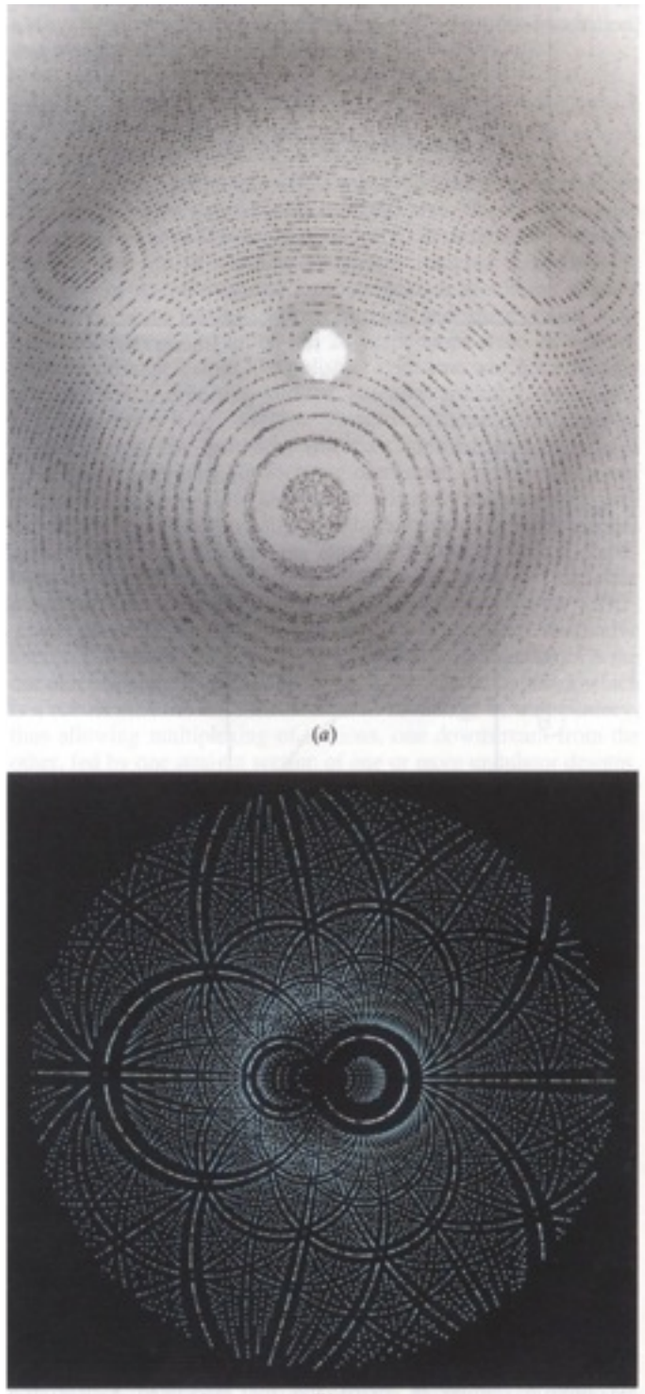
$P4_32_12$
A=78, B=78, C=36
90°,90°,90°



$P2_12_12_1$
A=111, B=173, C=308
90° 90° 90°

Larger the unit cell the closers the reflections

Monochromatic vs Polychromatic X-rays



(a)

(b)

Fig. 8.1.4.2. Single-crystal SR diffraction patterns. (*a*) Rhinovirus monochromatic oscillation photograph recorded at CHESS (Arnold *et al.* 1987; see also Rossmann & Erickson, 1983). Copyright (1987) International Union of Crystallography. (*b*) Prediction of a protein crystal Laue diffraction pattern (for an illuminating bandpass, without monochromator, $\sim 0.4 < \lambda < 2.6$ Å). The colour coding is according to the multiplicity of each spot: turquoise for singlet reflections, yellow for doublets, orange for triplets and blue for quartet or higher-multiplicity Laue spots. Reproduced with permission from Cruickshank *et al.* (1991). Copyright (1991) International Union of Crystallography.

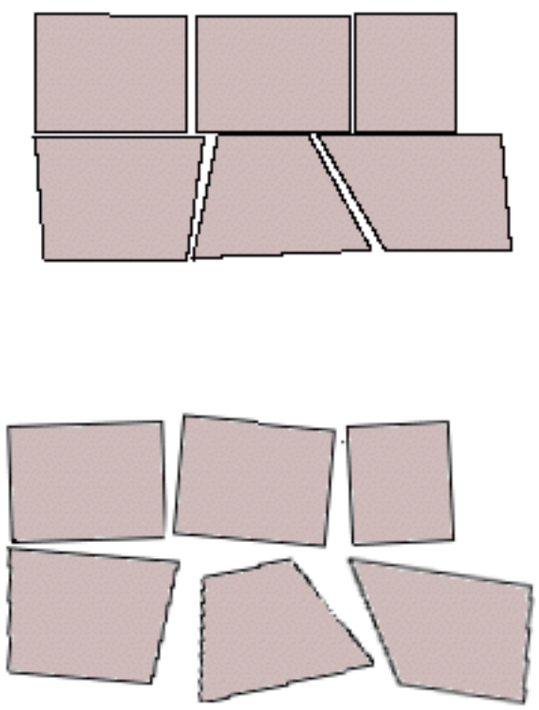
Monochromatic

Polychromatic

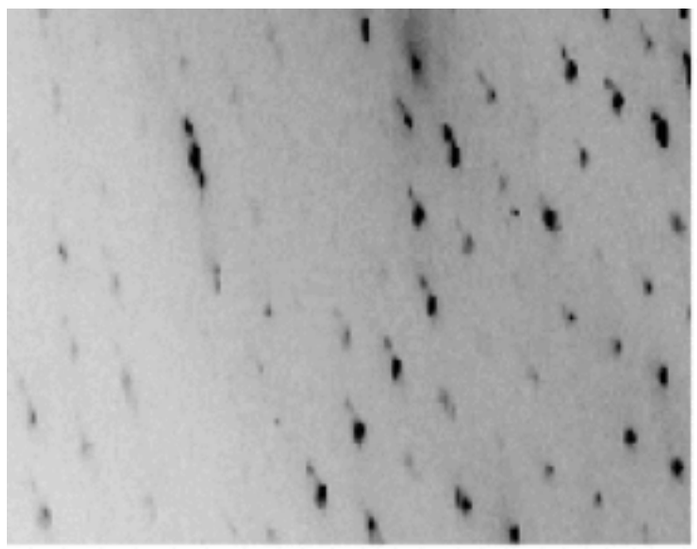
Mosaic spread

- single crystal can be seen as made up of smaller microcrystalline domains that are slightly misaligned with respect to each other.

- *mosaic spread* is an angular *measure of the misalignment*

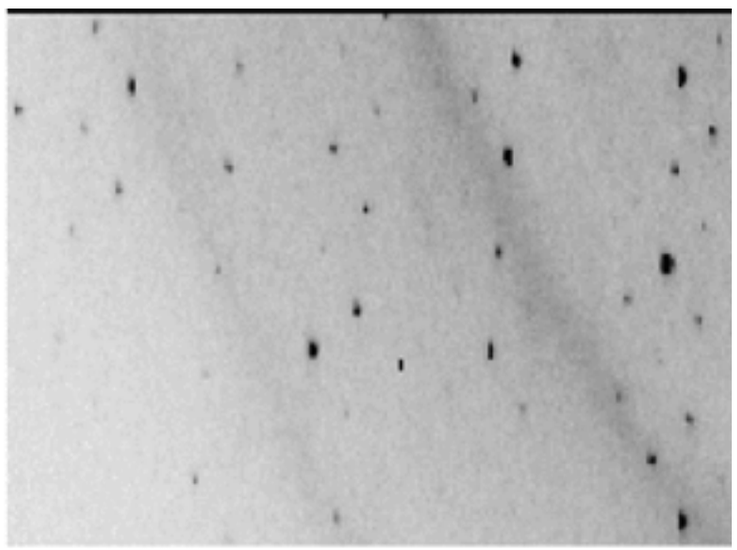


High Mosaicity



(*a*)

Low Mosaicity



(*b*)

Indexing

- Initial estimates of the unit cell dimensions are made from the positions of the reflection and from the **physical parameters of the experiment**

- Indexing programs usually work on one oscillation image at a time

- Unit cell parameters, rotation of the unit cell, mosaicity and crystal-to-detector distance.

- Measure intensity of the reflection or diffraction spot and the background.

Twinning

- Crystal growth can lead to the interdigitation of the same lattice
- Each lattice diffracts X-ray independently
- each measured reflection is due to two or more Bragg spots
- The intensity of one reflection is a mixture of the two lattices.
- merohedral twinning (from the Greek meros, part, and hedron, face), two or more lattices coincide exactly in three dimensions.

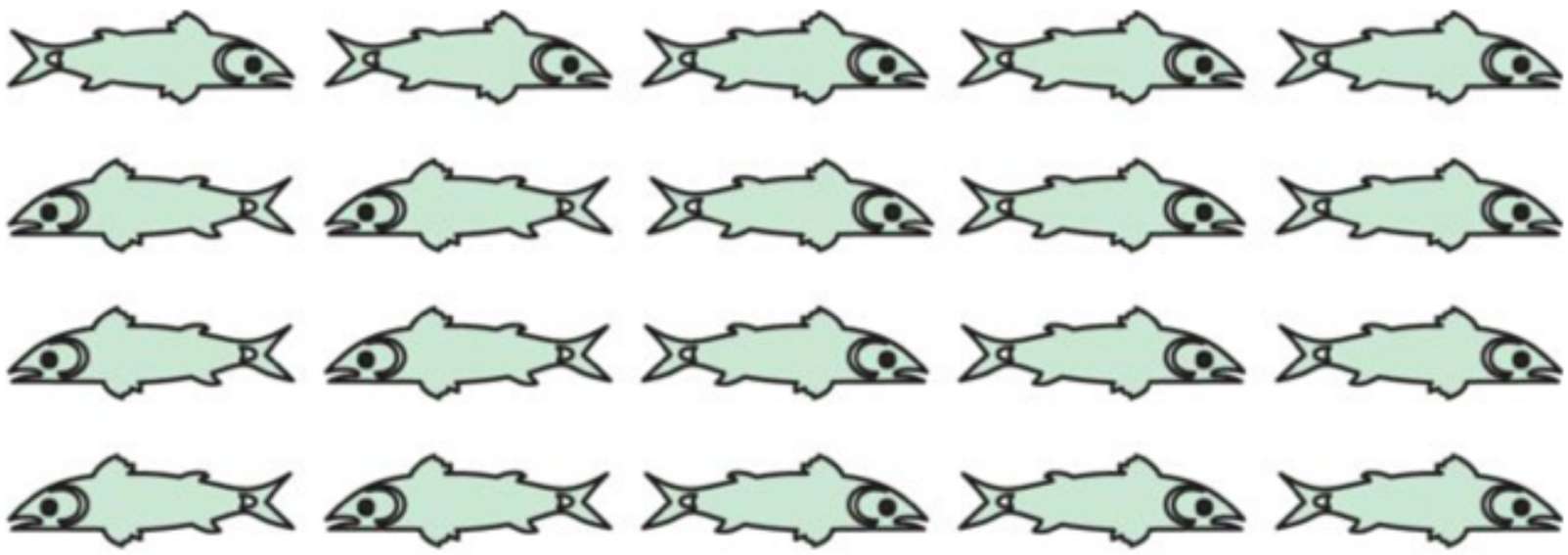
$I_{overall} = \alpha I$A + (1 − α)IB where A and B are the two separate lattices α is the twin factor α=0.5 means equal portion of both lattices.

- the presence of twinning can be identified from the average intensity, I, in each resolution range

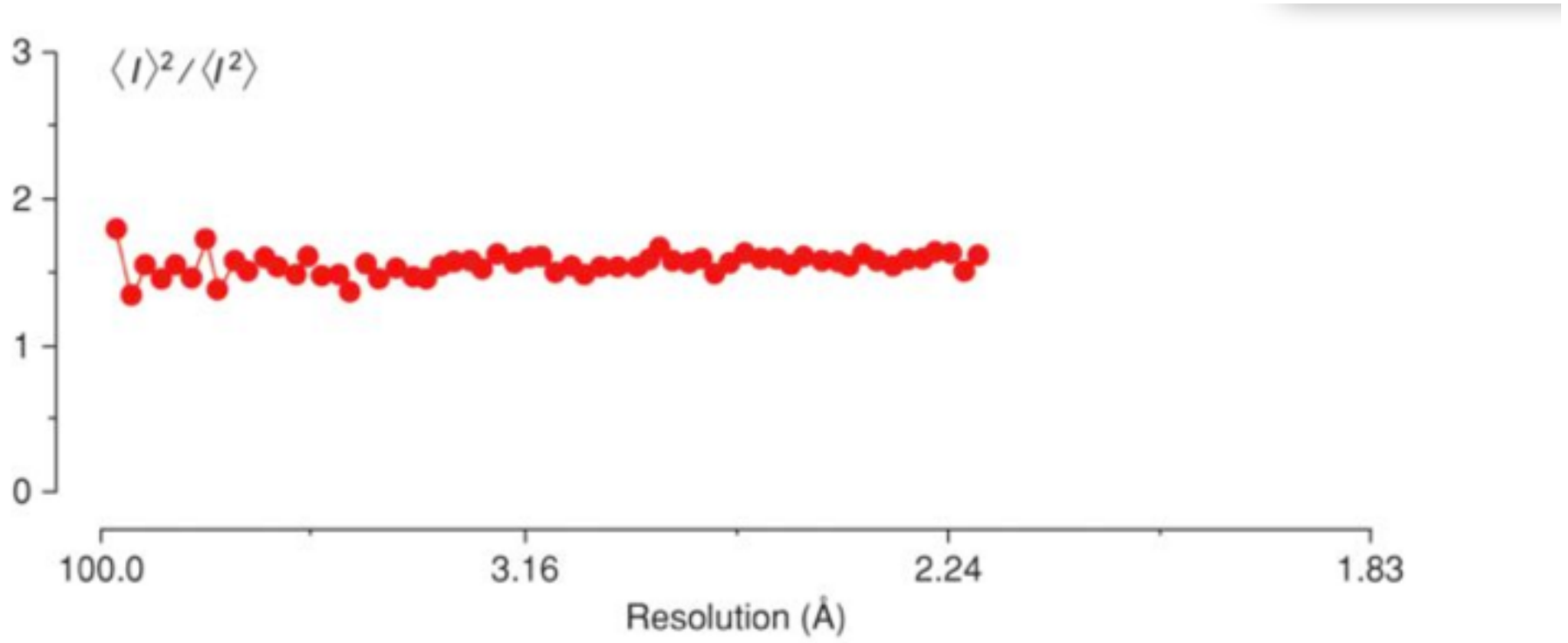
$\langle I^2 \rangle / \langle I \rangle^2 = 2$ untwinned

$\langle I^2 \rangle / \langle I \rangle^2 = 1.5$ twinned

Twinning



(a)



Redundancy and statistics

- What is the resolution limit of the data set?
- Determined by the signal to noise of the average reflection in a resolution range. (I/σ)
- Want I/σ≈2 for the highest resolution shell
- During data processing the data is broken down into resolution ranges called shells
- Redundancy is how many times a reflection is recorded
- These are critical parameters for assessing data quality

R_{merge} and $R_{p.i.m}$

- The R_{merge} parameter provides an estimate of the precision of individual measurements.

$$R = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

- R_{merge} increase as redundancy increase

- An improved multiplicity weighted R_{merge}, referred to as a Precision-Indicating Merging R factor ($R_{p.i.m.}$)

$$R_{p.i.m.} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$