Twinning in protein crystals

Zbigniew Dauter

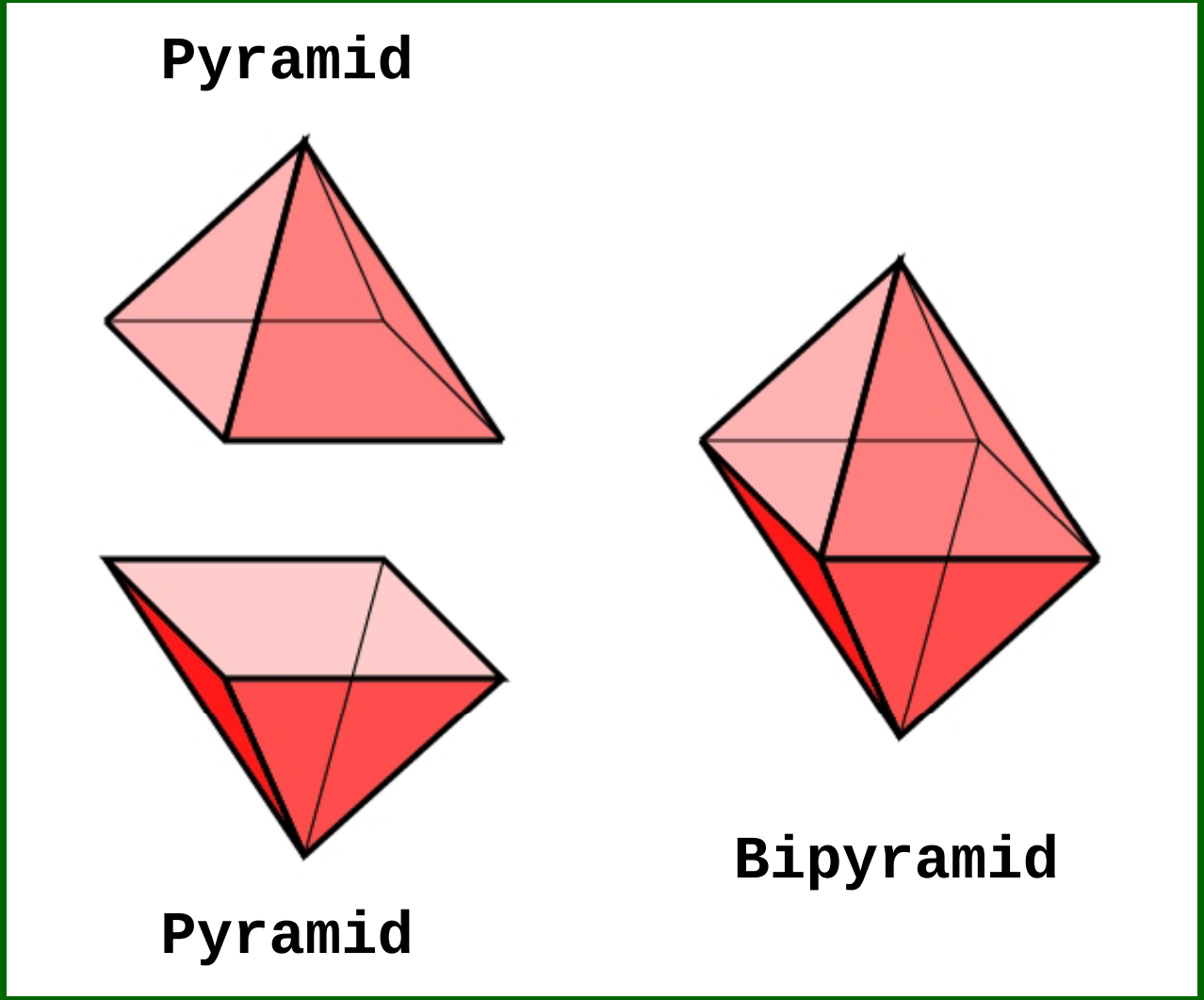
NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL

Twinning

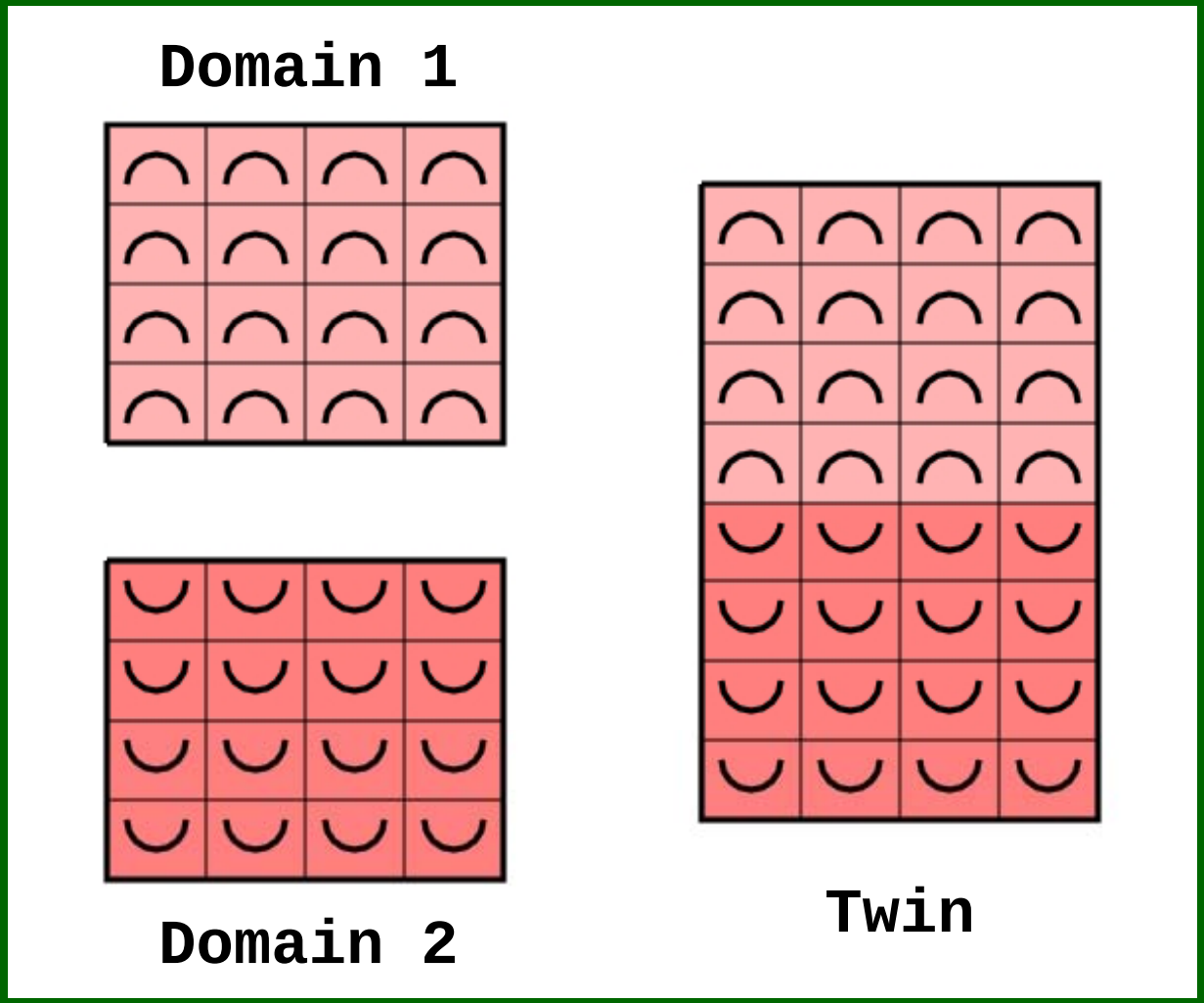
when reflections from more than one lattice (crystal) overlap

1. merohedral
   - crystal symmetry is the subgroup of the lattice symmetry
2. pseudomerohedral
   - coincidence of cell dimensions
3. non-merohedral
   - coincidence of the supercell
4. epitaxial
   - overlap in one or two dimensions
4. crystal cracking or splitting
   - this is not twinning !

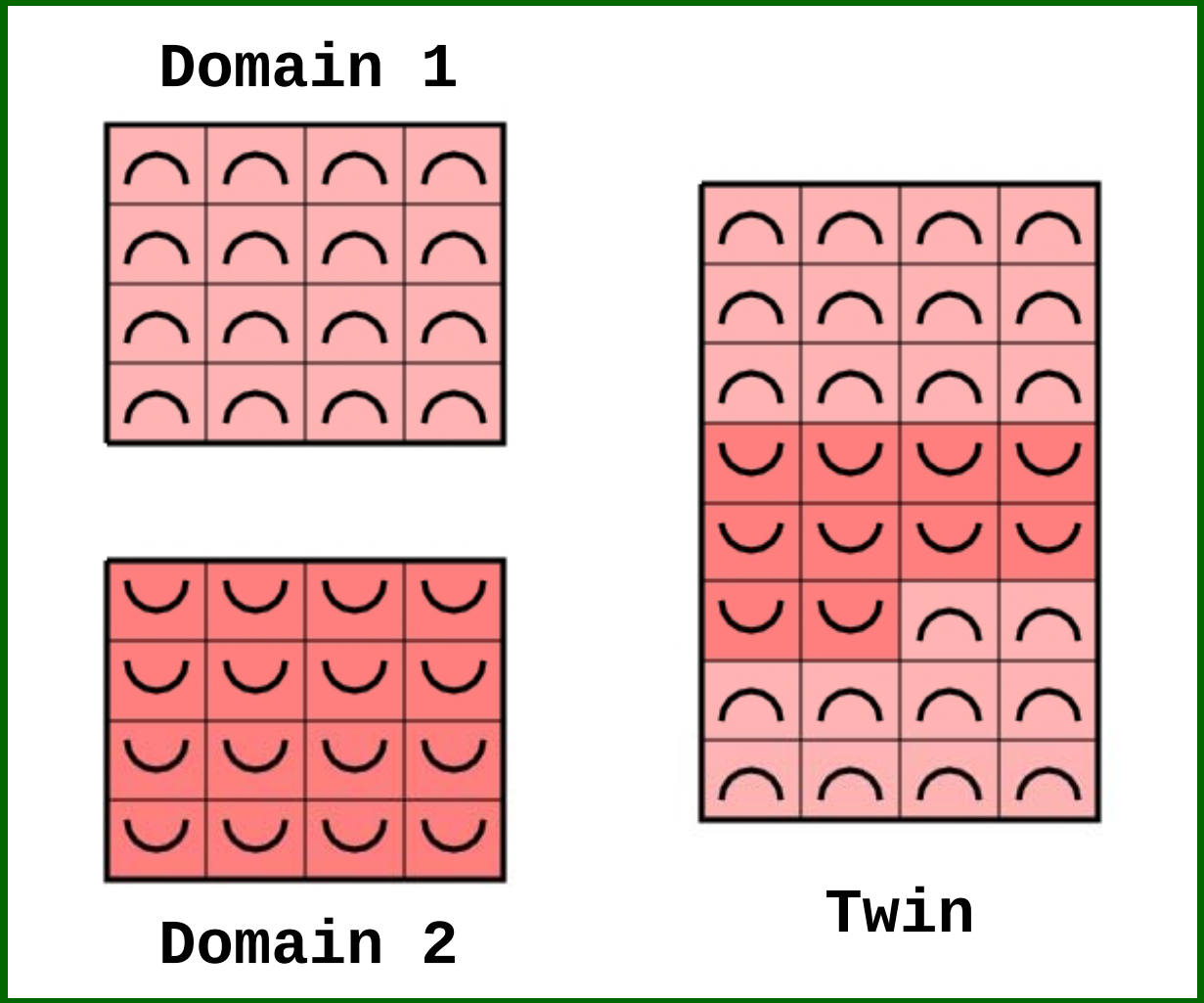
Crystal consisting of several domains



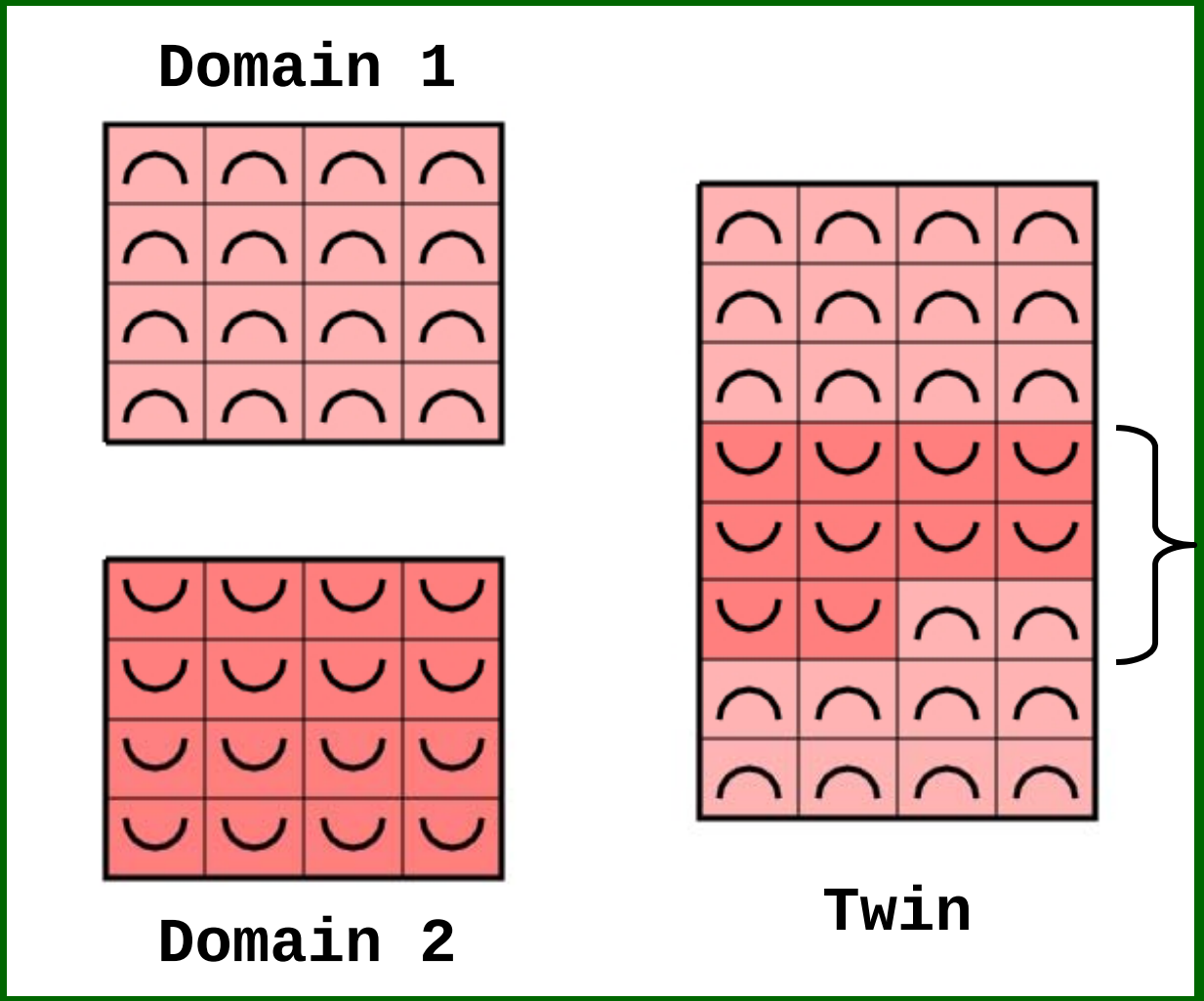
Crystal consisting of several domains



Crystal consisting of several domains



Crystal consisting of several domains

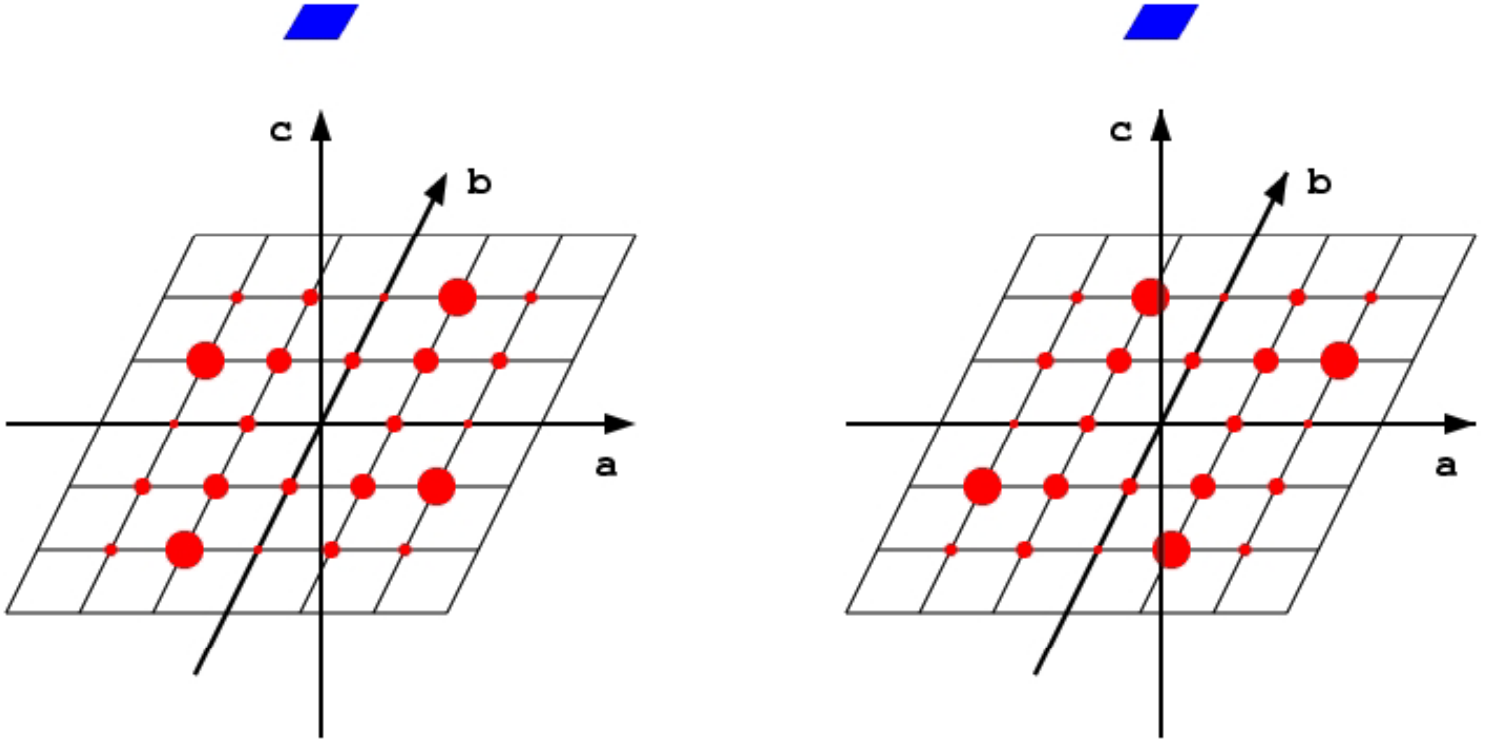


If size of domains
Is larger than
koherence length of
X-rays, addition of
intensities
= **twinning**

>> λ

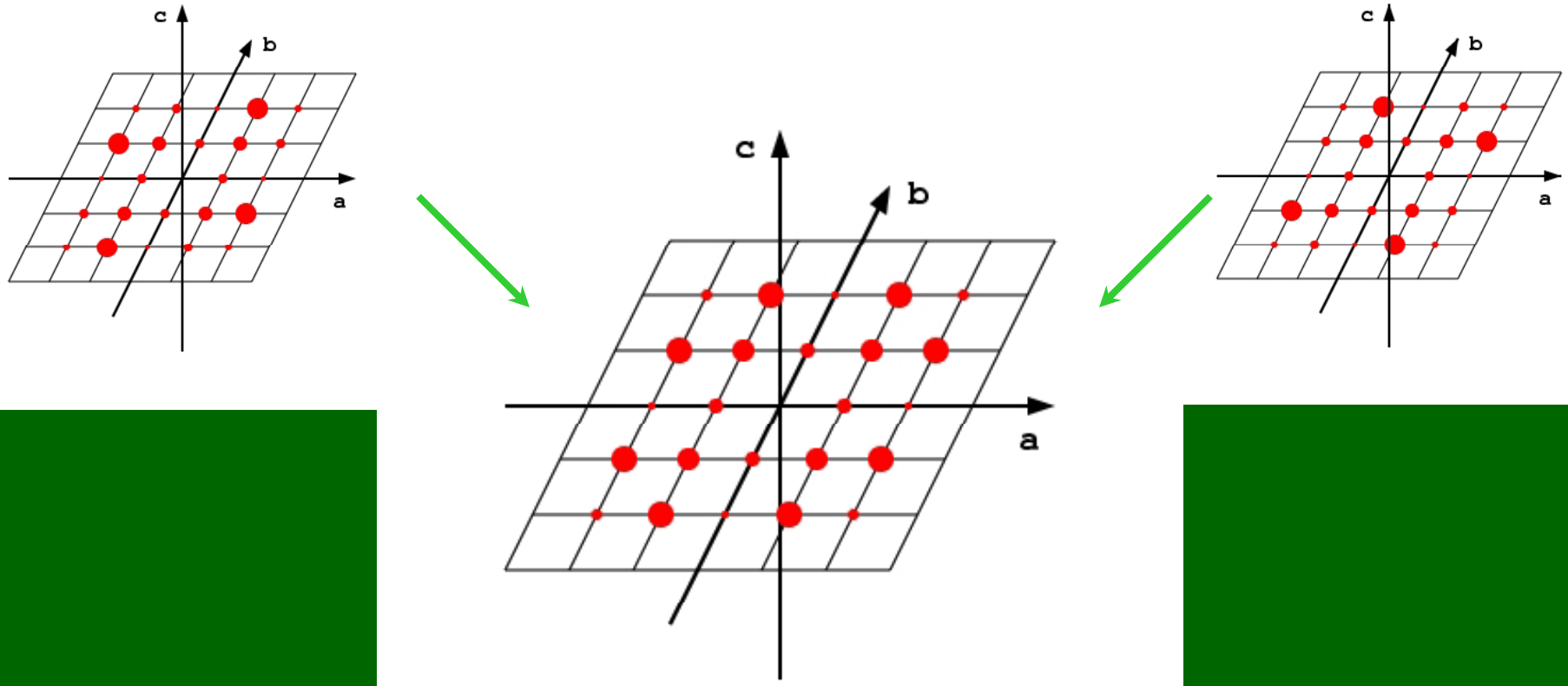
If size of domains
Is smaller than
koherence length of
X-rays, addition of
amplitudes
= **disorder**

Reciprocal lattices
symmetry 4



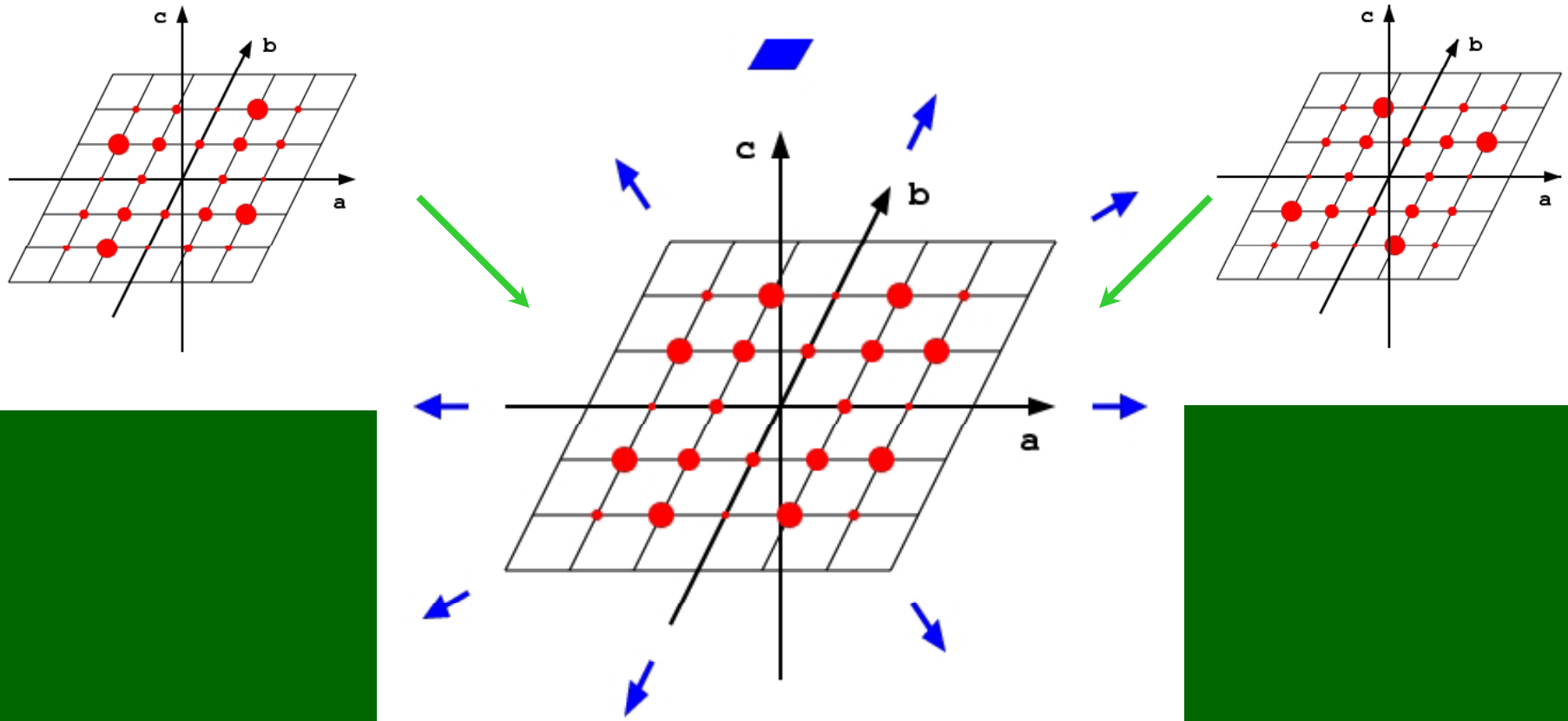
Only four-fold axis exists in both cases
but the polar axis is in opposite directions

Reciprocal lattices overlapped mimicking symmetry 422



Superposition of two lattices

Reciprocal lattices overlapped mimicking symmetry 422



Superposition of two lattices has higher symmetry

Merohedral twinning

is possible when lattice symmetry is higher than crystal symmetry

(for macromolecules,
without center of symmetry)

crystal point group		lattice symmetry
4		422 (4/mmm)
3	> 321, 312, 6 >	622 (6/mmm)
32		622 (6/mmm)
6		622 (6/mmm)
23		432 (m3m)

Twinning operators

any operation existing in lattice symmetry
but absent in crystal point group

crystal	lattice	twinning operator	
4	422	k,h,-l	2-fold || a+b
3	622	-h,-k,l	2-fold || c
		k,h,-l	2-fold || a+b
		-k,-h,-l	2-fold || a-b
321	622	-h,-k,l	2-fold || c
312	622	-h,-k,l	2-fold || c
6	622	k,h,-l	2-fold || a+b
23	432	k,h,-l	2-fold || a+b

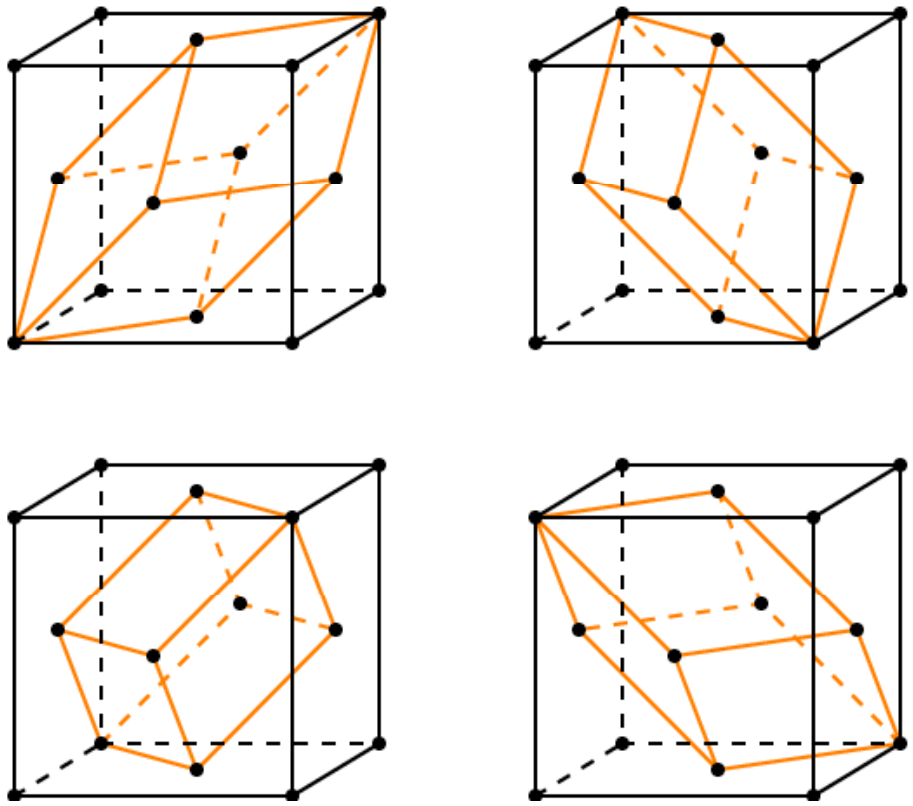
Pseudomerohedral twinning

when cell dimensions are special
and lattice has higher metric symmetry

examples:

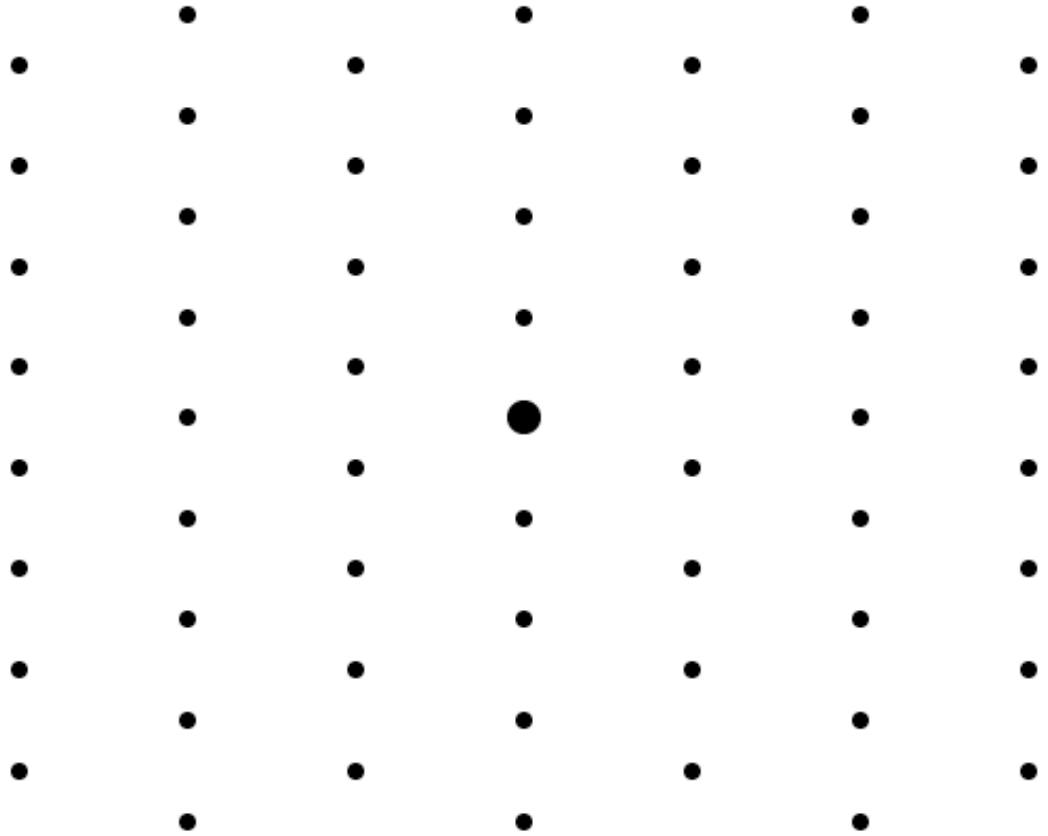
cell		lattice
monoclinic P	β~90º	-> orthorhombic P
monoclinic P	a~b	-> orthorhombic C
or	2acosβ+c = 0	-> ”
rhombohedral	α~60º	-> cubic F
rhombohedral	α~90º	-> cubic P
rhombohedral	α~109.5º	-> cubic I

Cubic F cell – Rhombohedral R with $\alpha = 60^{\circ}$

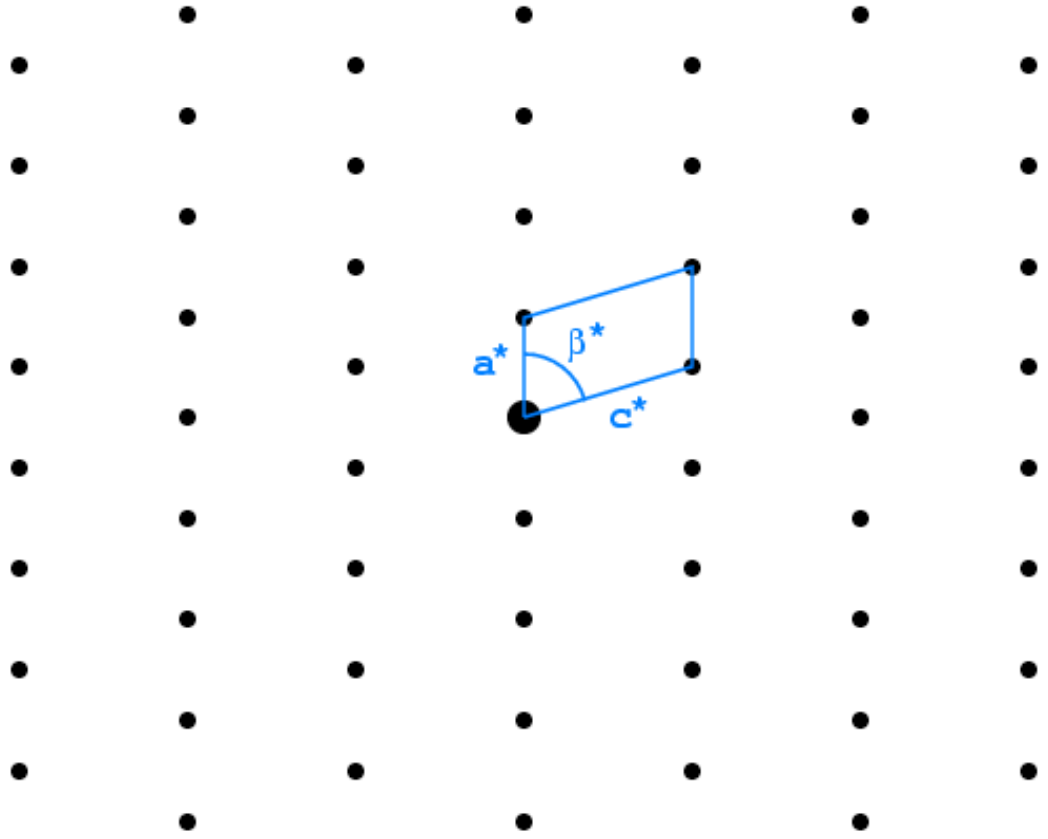


Cubic F cell can accommodate:
4 orientations of R32 crystals
8 orientations of R3 crystals

Pseudomerohedry

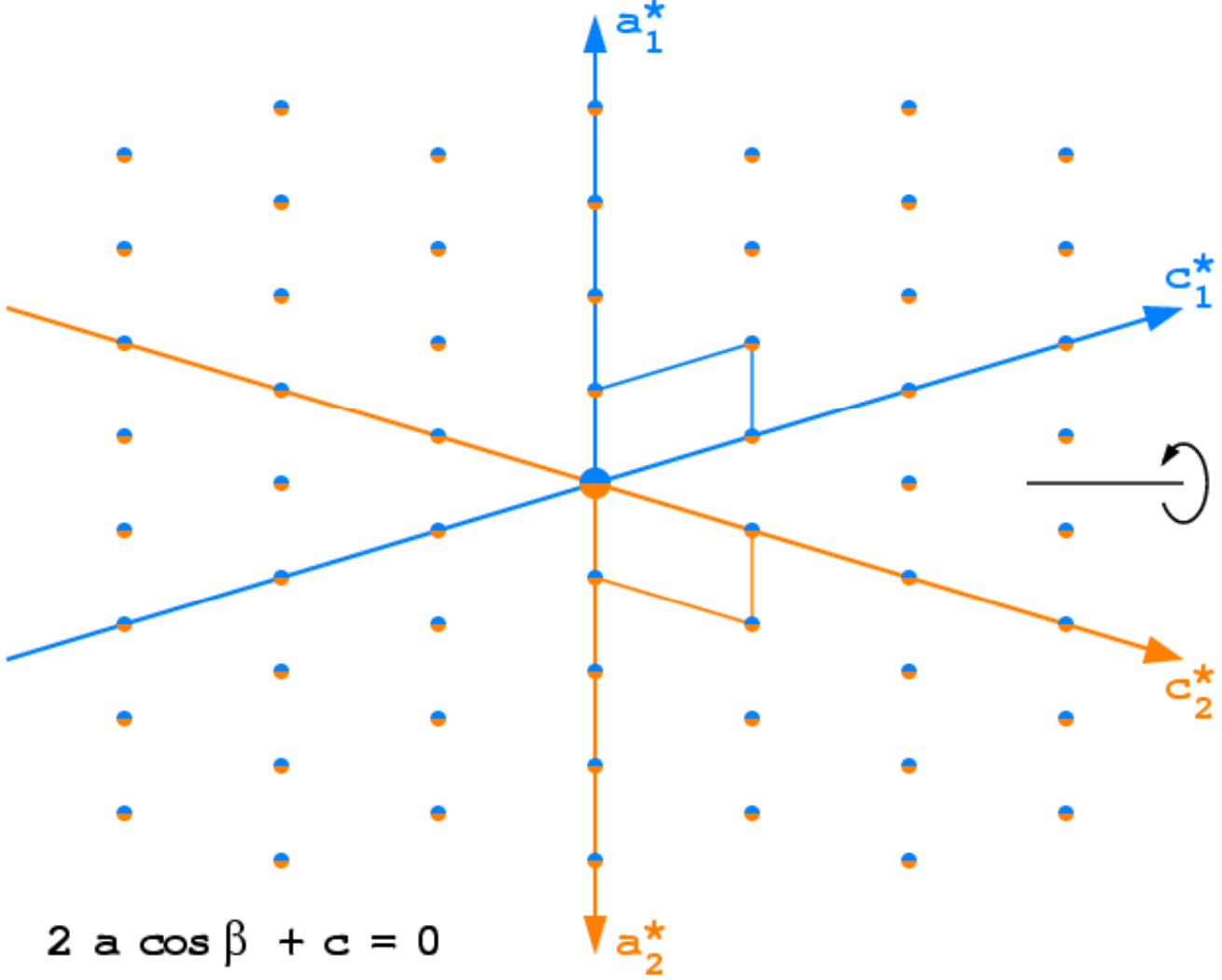


Pseudomerohedry

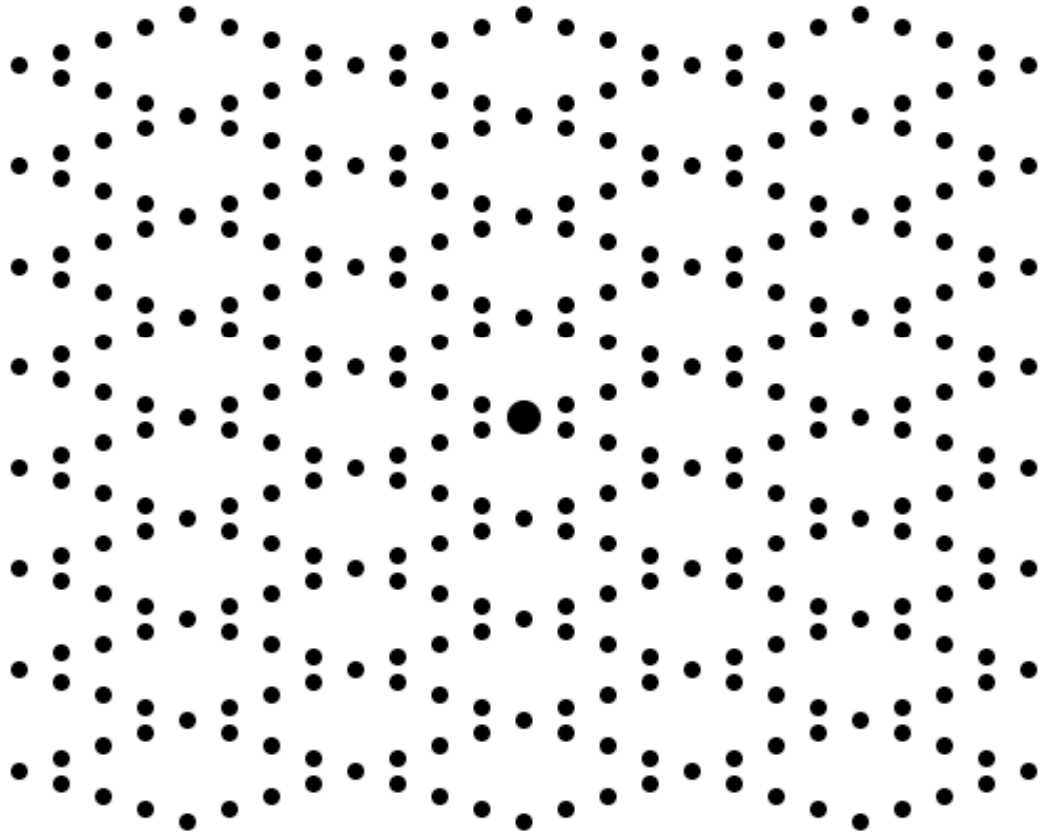


$2\ a \cos\beta + c = 0$

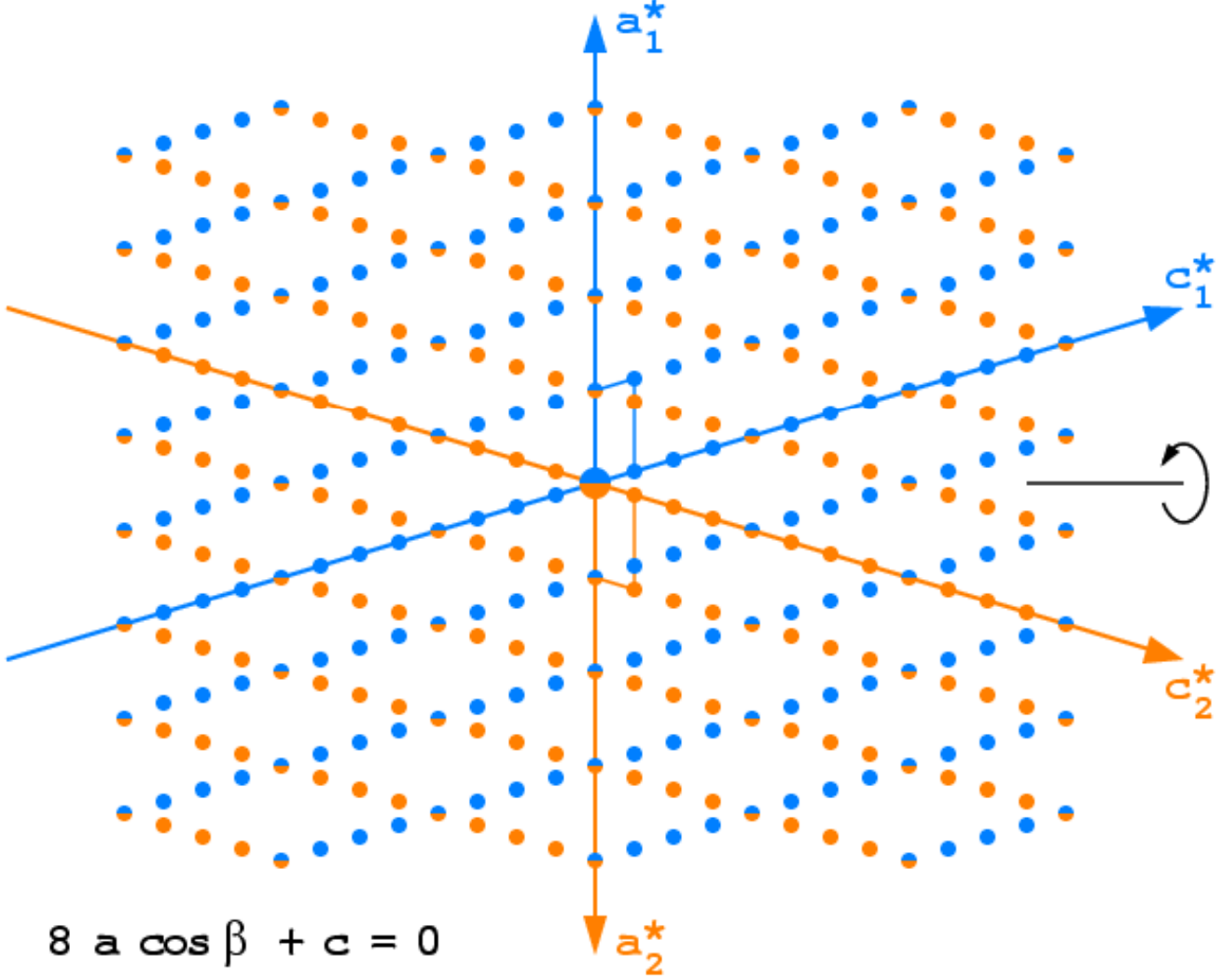
Pseudomerohedry



Non-merohedry (Reticular twinning)



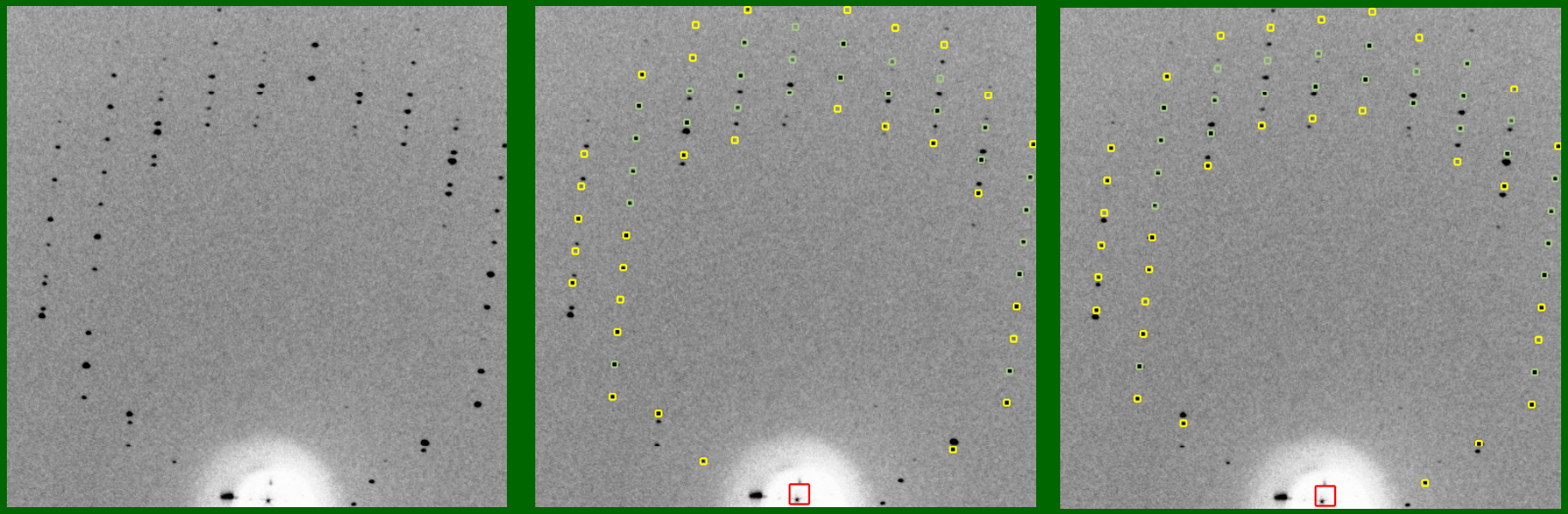
Non-merohedry (Reticular twinning)



Non-merohedry (Reticular twinning)

almo0

Reticular twinning occurs when only subset of reflections coincides.



Every fourth line of reflections coincides perfectly

Distribution of intensities differs from ordinary (Wilson) statistics

Because intensities, not amplitudes add up from different domains

There are less very weak and very strong reflections – smaller probability that two extreme intensities combine

Identifiers of twinning

based on

> statistics of all intensities
(data can be processed in low or high symmetry)

> comparison of twin-related intensities
(data have to be processed in proper. Low symmetry)

- **Wilson ratios, $\langle I^2 \rangle / \langle I \rangle^2$**
- **higher moments of E**
- **N(z) cumulative intensity test**
- **H test of Yeates**
- **negative intensity Britton test**
- **L test of Padilla & Yeates**

Wilson ratios

$\langle I^2 \rangle / \langle I \rangle^2 = 2$ for not twinned

$\langle I^2 \rangle / \langle I \rangle^2 = 1.5$ for 50 % twinned

Moments of E

$\langle |E| \rangle = 0.866$ for not twinned

$\langle |E| \rangle = 0.940$ for 50 % twinned

$\langle |E|^3 \rangle = 1.329$ for not twinned

$\langle |E|^3 \rangle = 1.175$ for 50 % twinned

capsid protein gpD from bacteriophage λ

trimer, 3 * (93 a.a. + 2 SeMet)

1.7 Å MAD data from X9B @ NSLS

monoclinic, $P2_1$

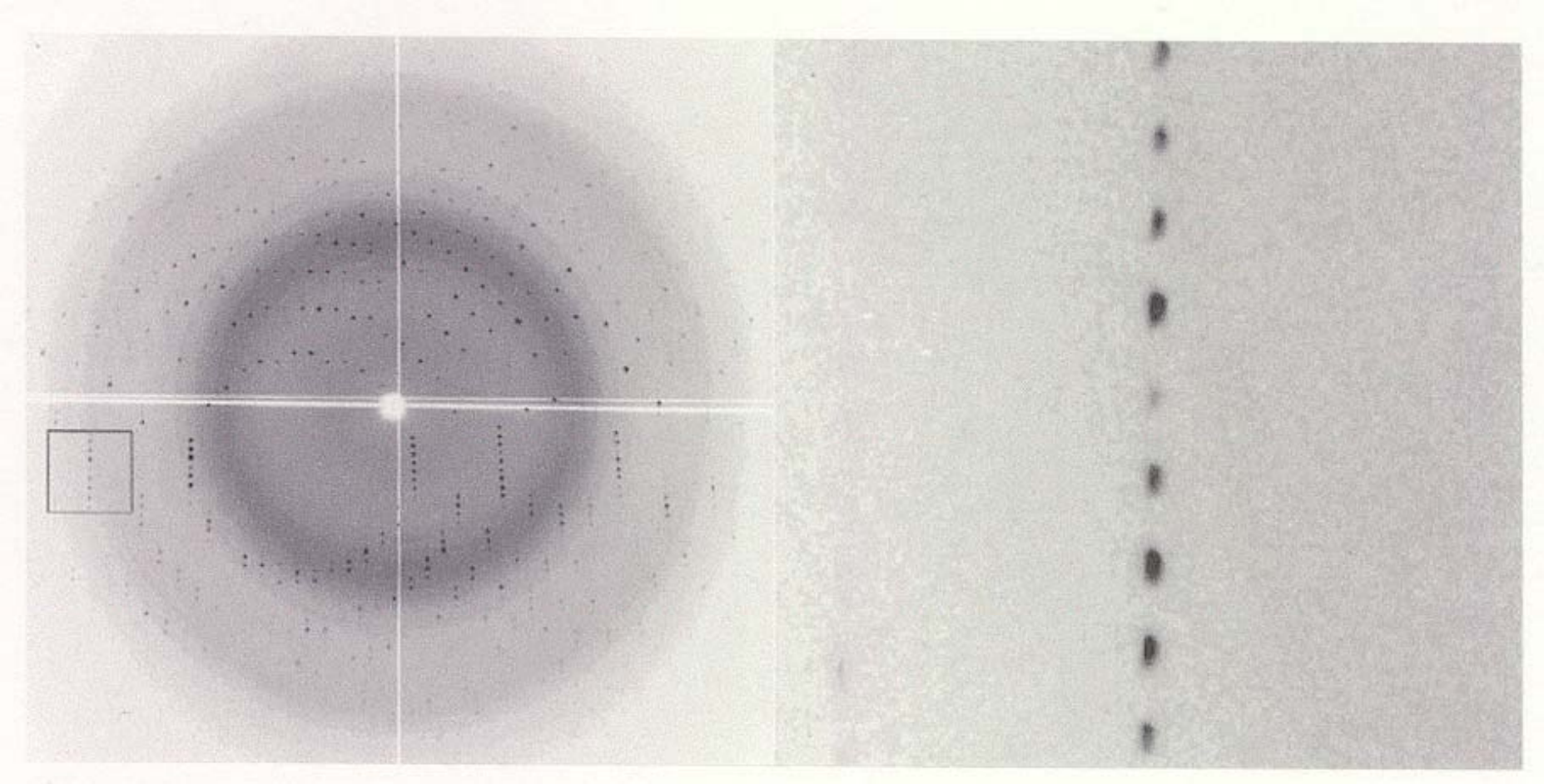
a = 45.51, b = 68.52, c = 45.52 Å

$\beta = 104.4^{o}$

45 % solvent

Refined by SHELXL with native data @ 1.3 Å, R = 13 %, α = 36 %

gpD diffraction



cumulative intensity N(z) test

(look at TRUNCATE output)

data can be processed in wrong symmetry

- number of reflections > z ,
  percent of average intensity

it is not highly probable that
both very weak or very strong
reflections will overlap after
twinning

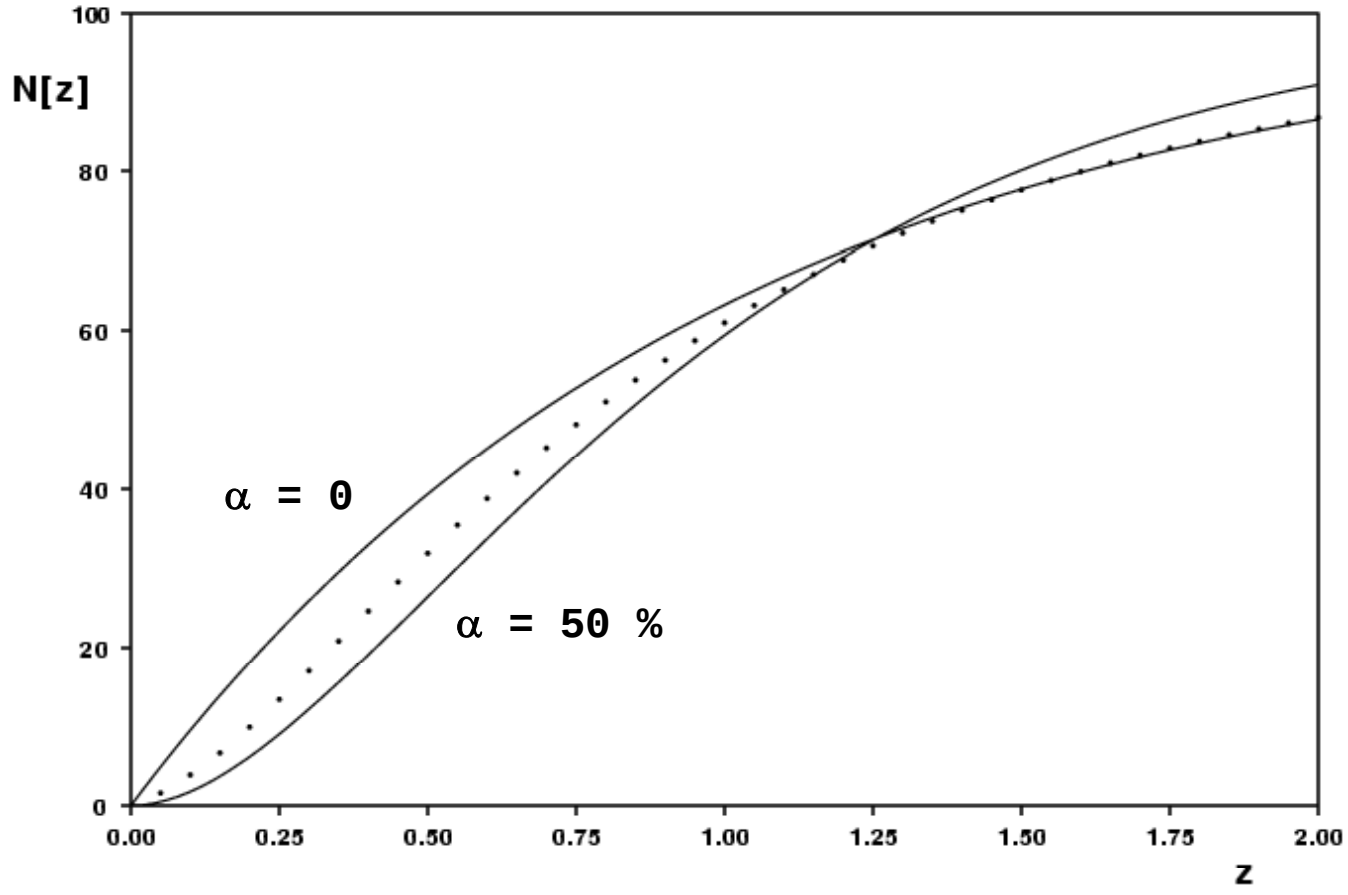
– so for twinned crystal there
  are less weak reflections

- sigmoidal N(z) curve

$\langle I^2 \rangle / \langle I \rangle^2 = 2$ for not twinned

$\langle I^2 \rangle / \langle I \rangle^2 = 1.5$ for 50 % twinned

gpD SeMet peak, cumulative intensity N(z) test



negative intensity (Britton) test

data have to be processed in low symmetry

$$I^1_{obs} = (1-\alpha)I^1_{detw} + \alpha I^2_{detw}$$

$$I^2_{obs} = \alpha I^1_{detw} + (1-a)I^2_{detw}$$

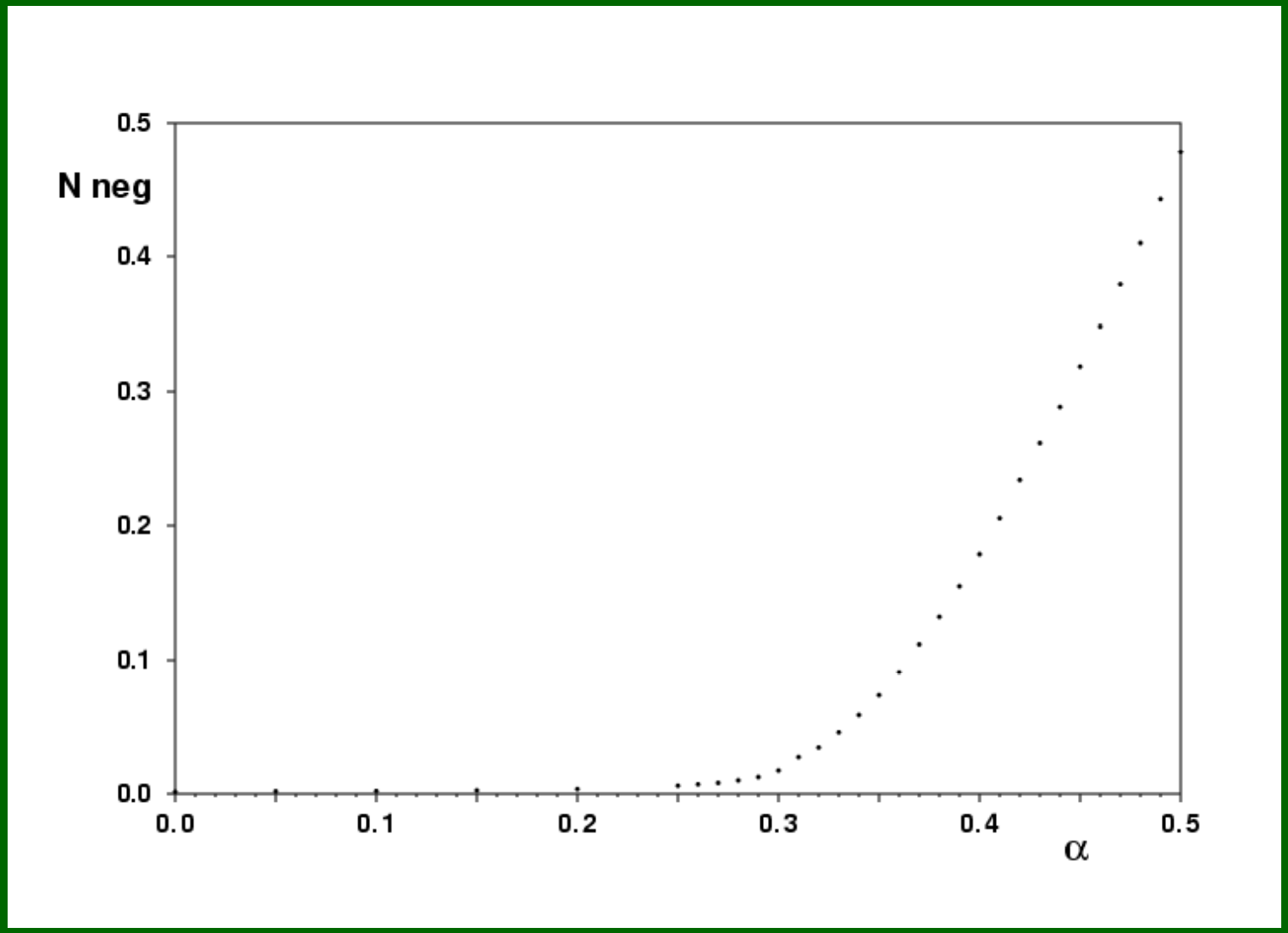
detwinning equations:

$$I^1_{detw} = [(1-\alpha)I^1_{obs} - \alpha I^2_{obs}]/(1-2\alpha)$$

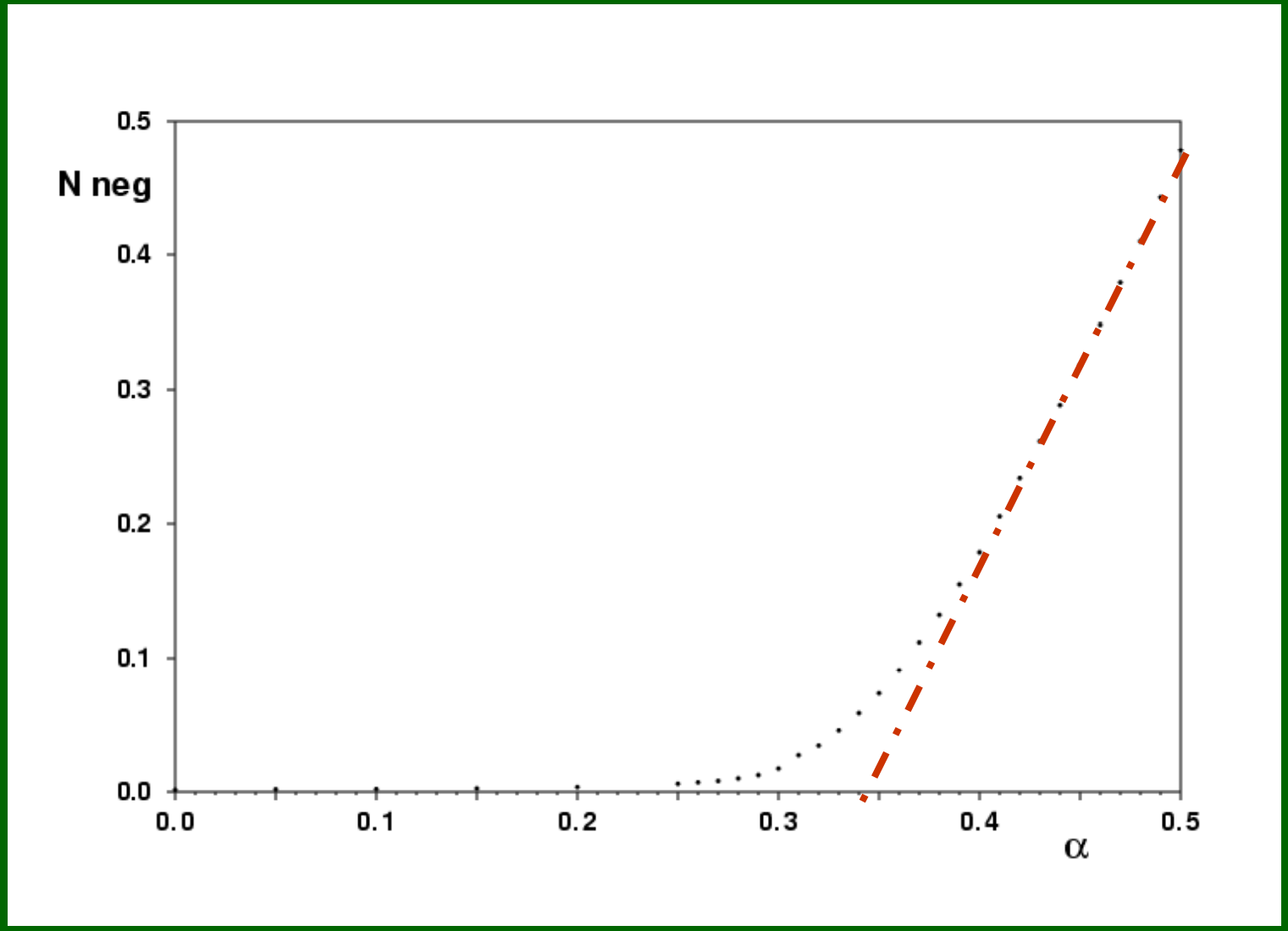
$$I^2_{detw} = [(1-\alpha)I^2_{obs} - \alpha I^1_{obs}]/(1-2\alpha)$$

if α is too high, there will be many negative estimations

gpD SeMet peak, negative intensity (Britton) test



gpD SeMet peak, negative intensity (Britton) test



Yates S(H) test

data have to be processed in low symmetry

$$H = |I^1_{obs} - I^2_{obs}| / (I^1_{obs} + I^2_{obs})$$

I^1 and I^2 are twin-related

cumulative distribution of H is linear

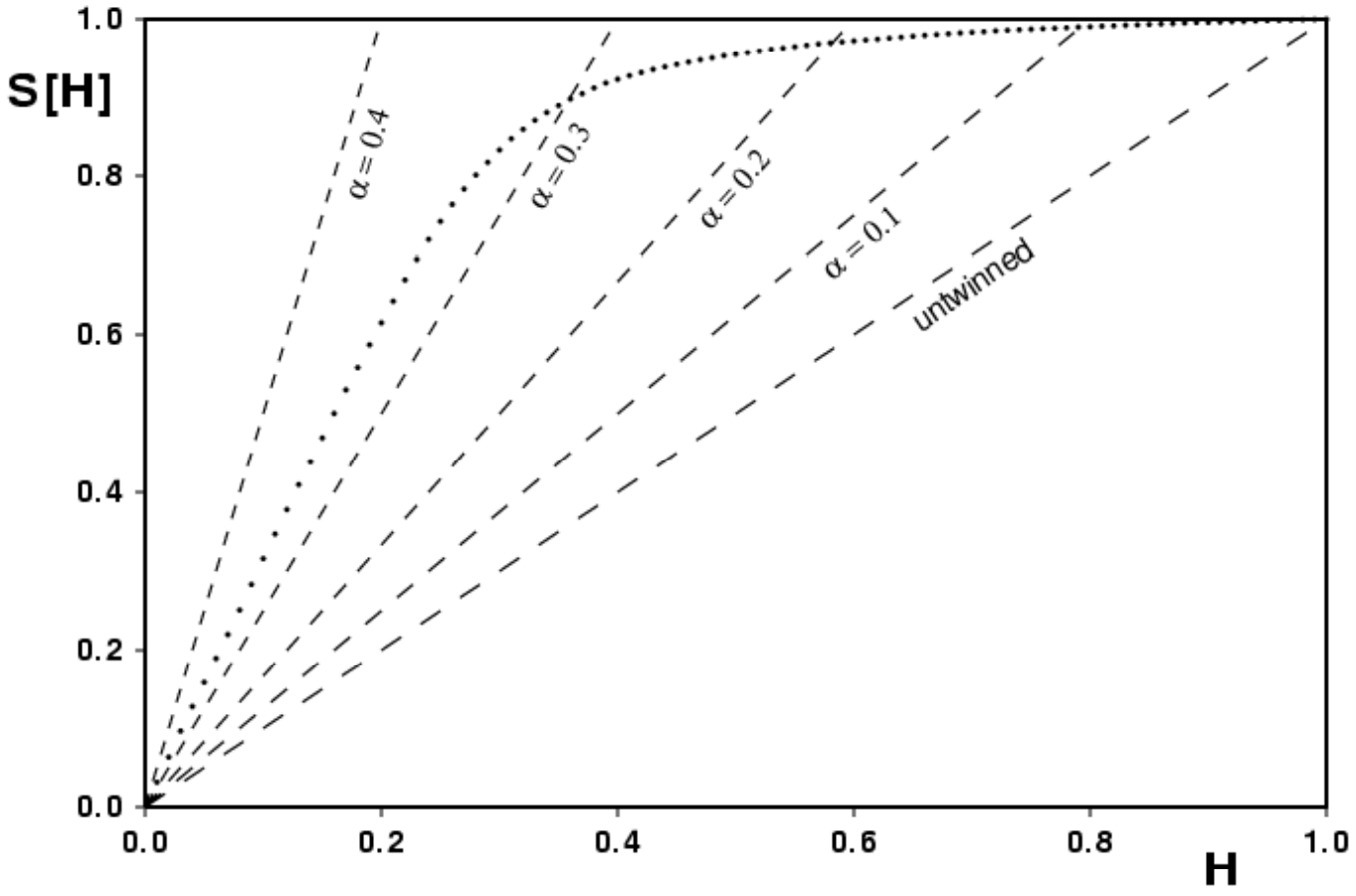
$$S(H) = H / (1 - 2\alpha)$$

$$< H > = \frac{1}{2} - \alpha$$

$$< H^2 > = (1 - 2a)^2/3$$

Weak reflections can be discarded

gpD SeMet peak, Yates S(H) test



Padilla & Yates L test

data can be processed in any symmetry

$$L = (I^1_{obs} - I^2_{obs}) / (I^1_{obs} + I^2_{obs})$$

I^1 and I^2 are close in reciprocal space
(their indices differ by a small vector)

cumulative distribution of L is linear

$$N(|L|) = |L|$$

$$< |L| > = 1/2$$
$$< L^2 > = 1/3$$

Padilla & Yates L test

data can be processed in any symmetry

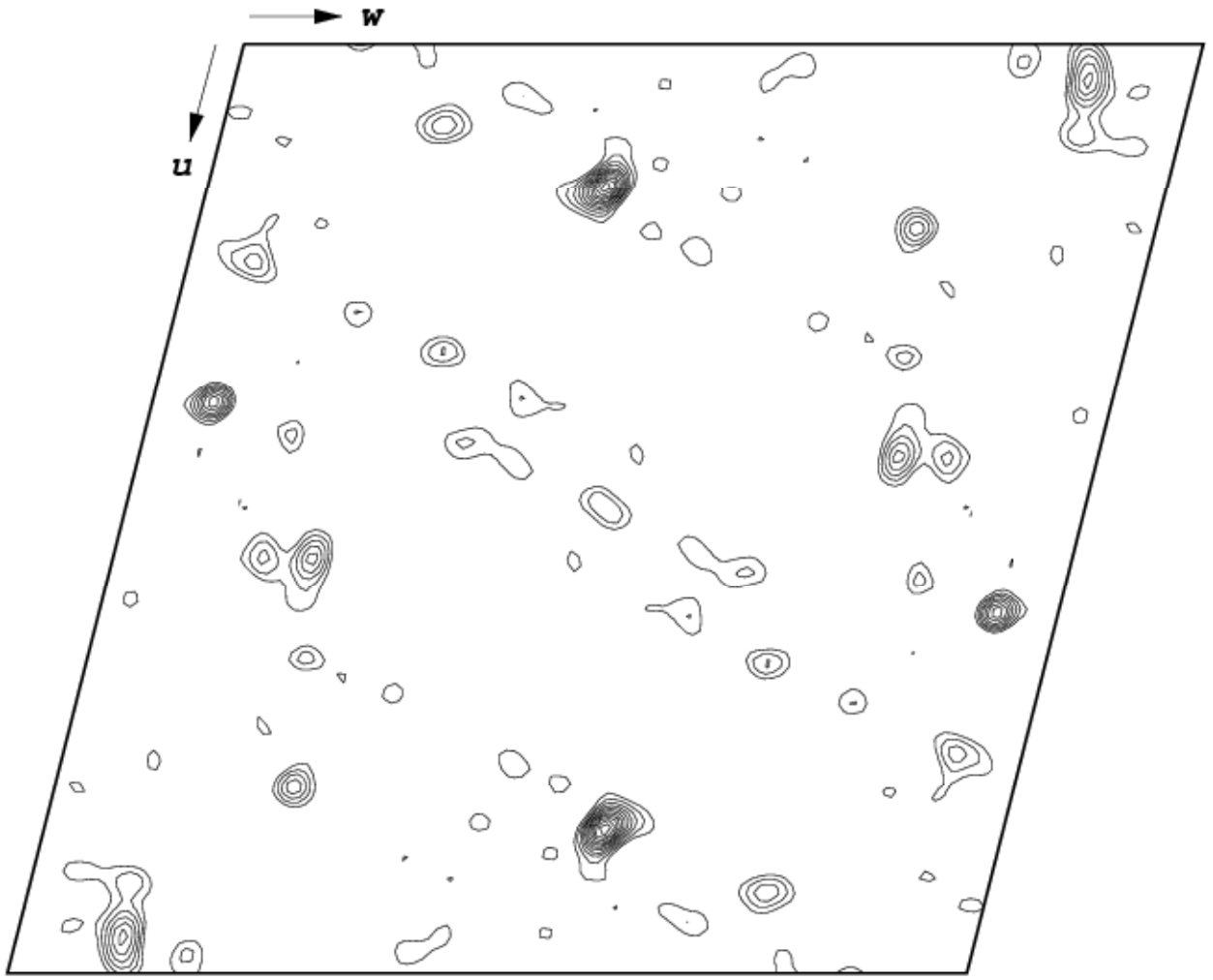
I^1 and I^2 are close in reciprocal space
their indices differ by a small vector

This statistic depends on the local differences of intensities

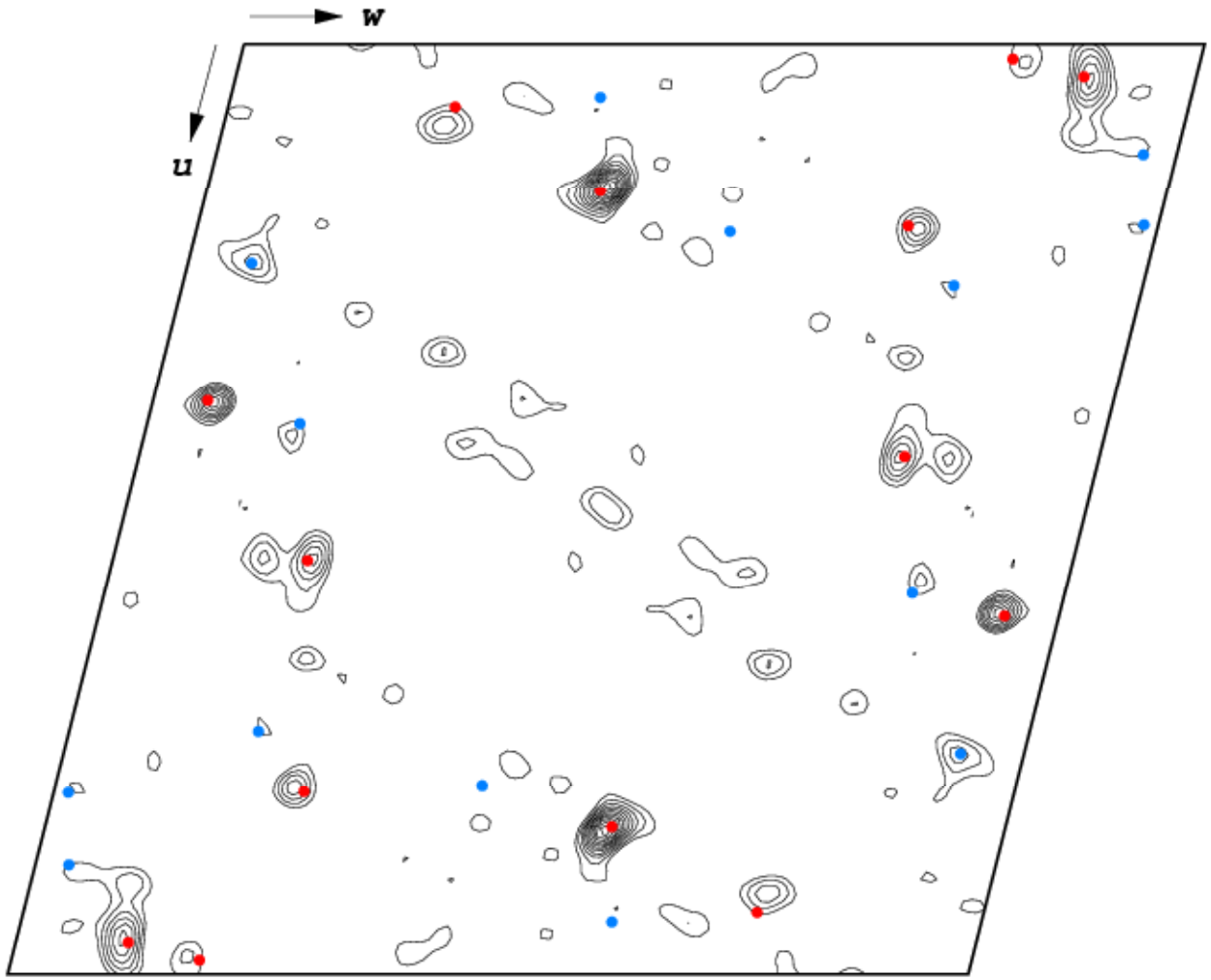
If the NCS translation is close to ½,0,0
reflections h=2n+1 are weak,
but if only reflections with the same parity of h are compared ($|h_1-h_2|$ = 2,0,0)
then the L-test in not biased

also with anisotropic diffraction

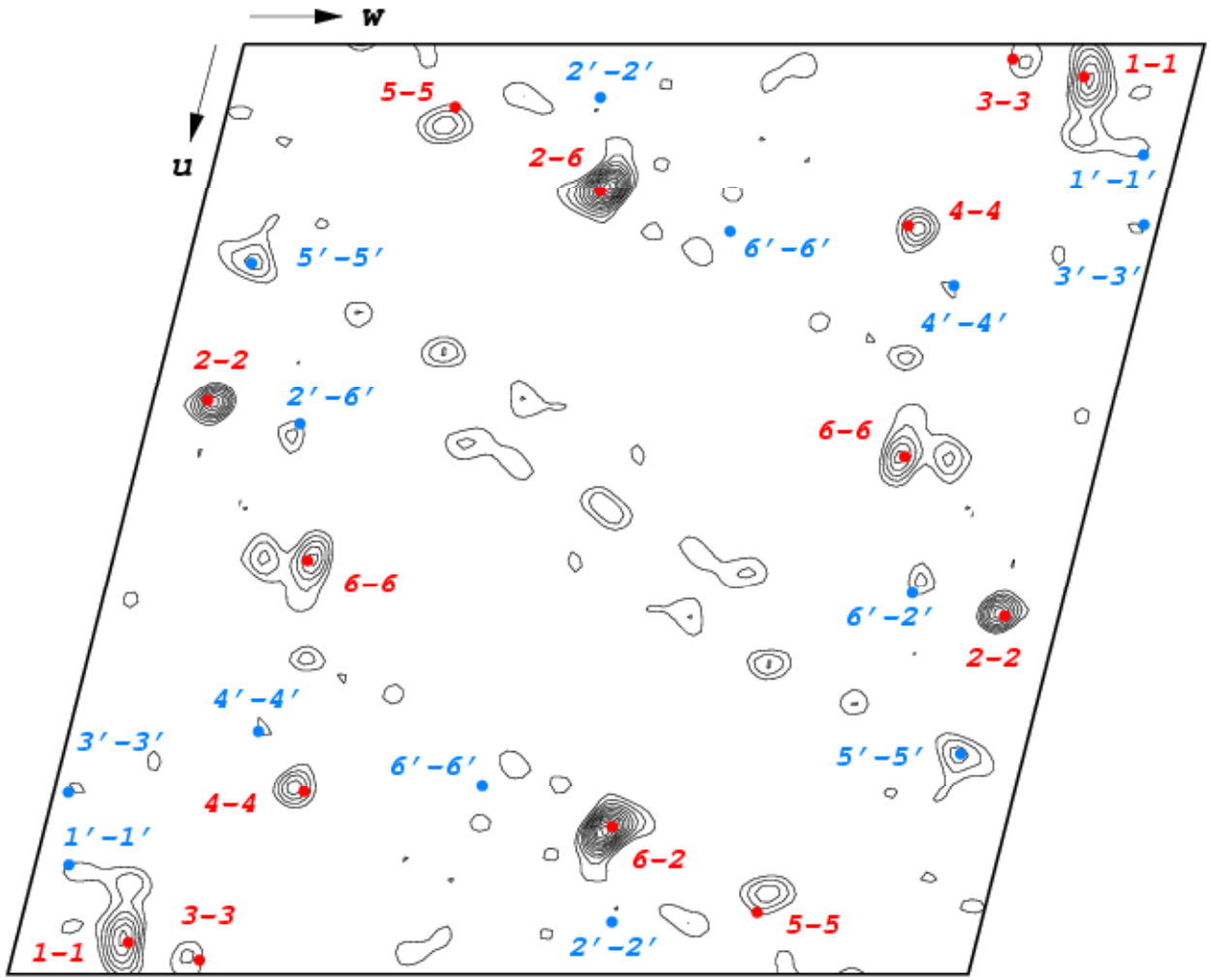
gpD anomalous Patterson



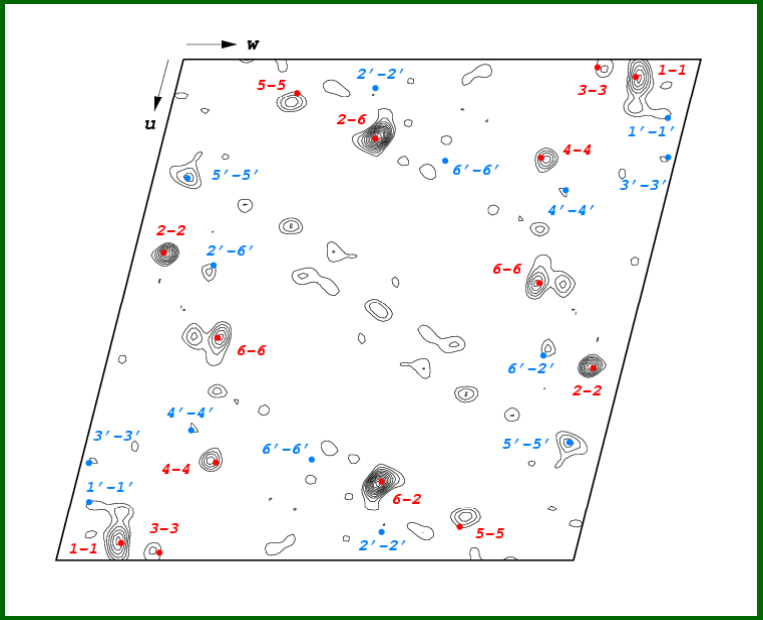
gpD anomalous Patterson



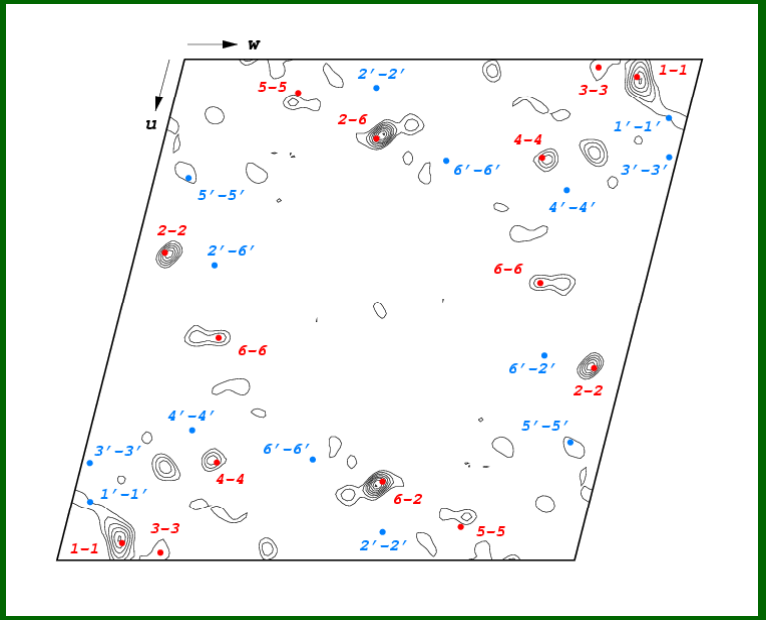
gpD anomalous Patterson



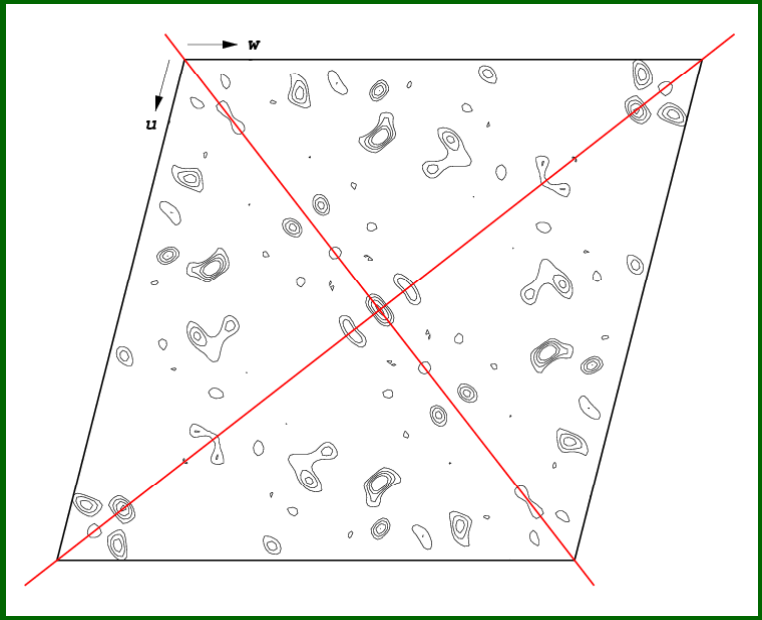
gpD anomalous Patterson



original data



detwinned data



scrambled data

Molecular replacement of twinned structures

MR is based on the overlap of Pattersons (from the known model and unknown data) therefore two solutions can be expected (but the contribution of other twinned intensities introduce noise)

MAD and SAD phasing

mixing of twinned intensities diminishes the inherently small anomalous signal (only the same Friedel mates are mixed)

Phasing results of gpD

α	orig	10%	20%	35%	50% scrambled
SHELXD results (SAD peak data)					
solutions %	98	87	100	99	97
best CC all	31.0	32.9	34.9	31.1	20.8
best PATFOM	34.9	36.5	38.4	32.3	22.8
SHELXE SAD phasing					
FOM	0.72	0.71	0.69	0.63	0.64
map CC	0.86	0.86	0.84	0.86	0.79

SHELXD peaks for peak data

(6 Se atoms expected)

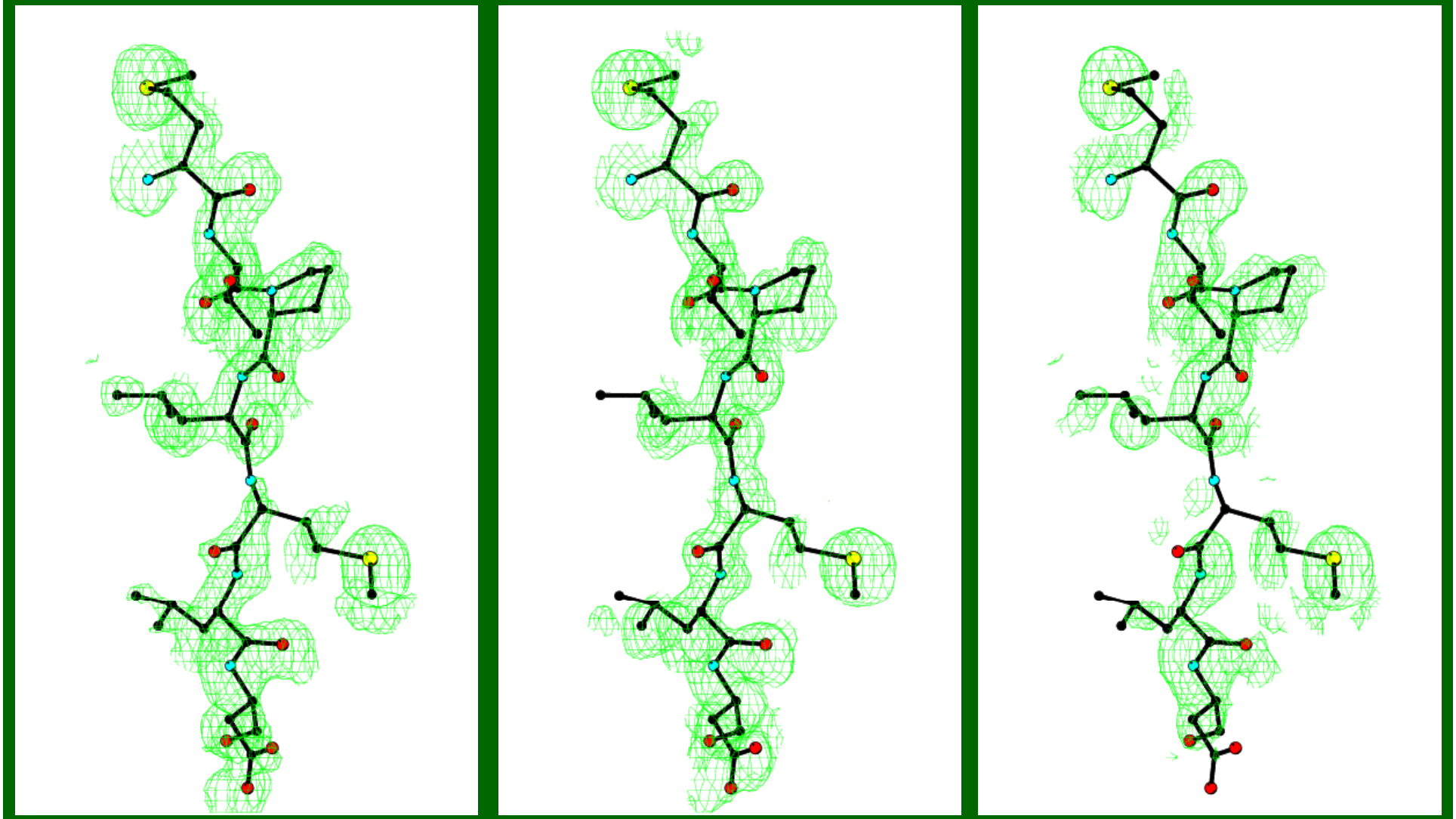
α =	orig	10%	20%	35% scrambl
peak				
1	1.00	1.00	1.00	1.00
2	0.88	0.86	0.86	0.85	0.86
3	0.82	0.83	0.83	0.81	0.79
4	0.78	0.79	0.78	0.77	0.76
5	0.70	0.72	0.71	0.69	0.70
6	0.62	0.60	0.59	0.57	0.55
7	0.14	0.11	0.13	0.16	0.20

Phasing results

original

α = 35 %

scrambled

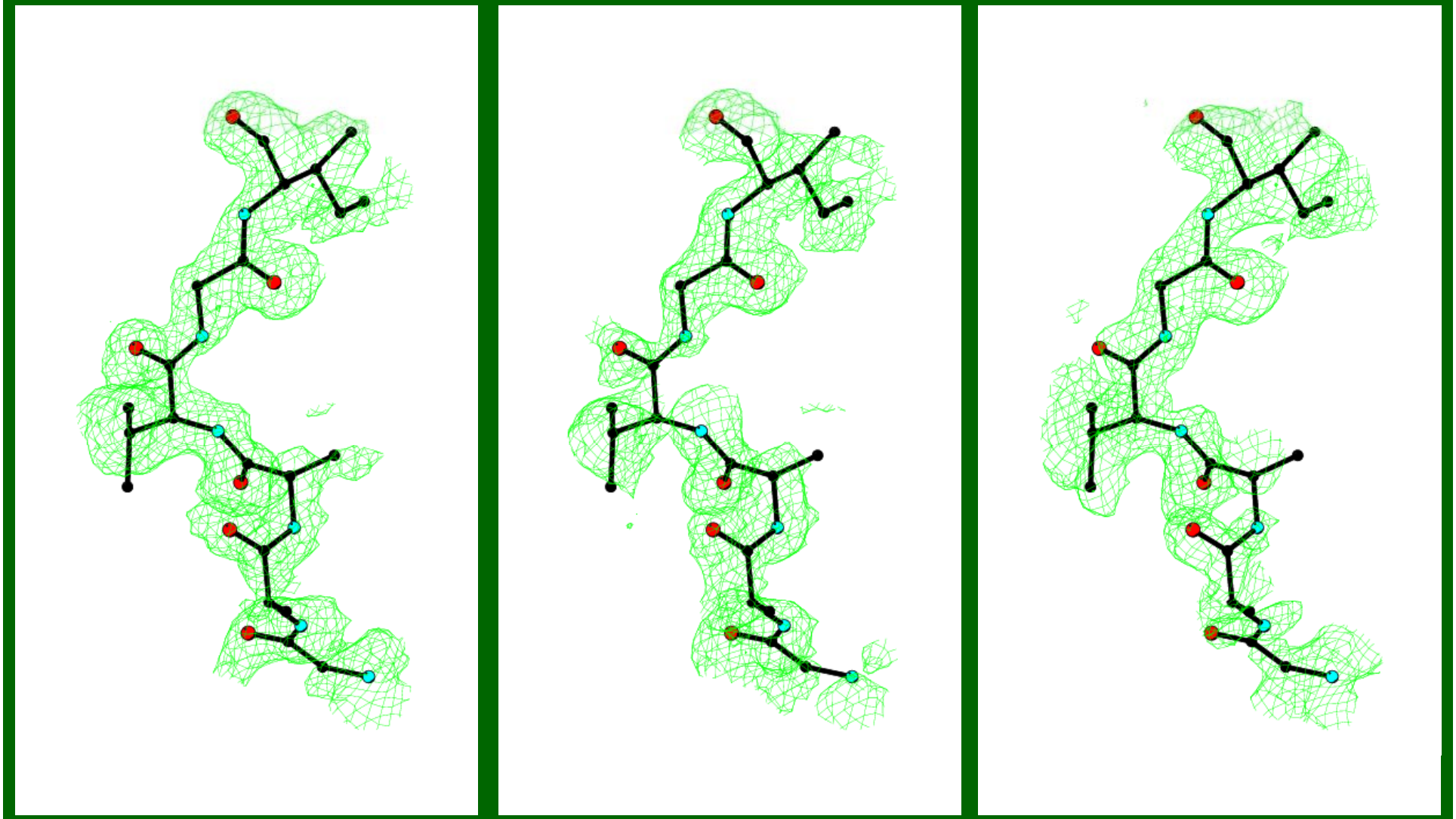


Phasing results

original

α = 35 %

scrambled



Conclusion

Do not give up too early

Try various options

If a model can be built,

It can be properly refined with SHELXL, PHENIX or CNS