

# Molecular replacement

---

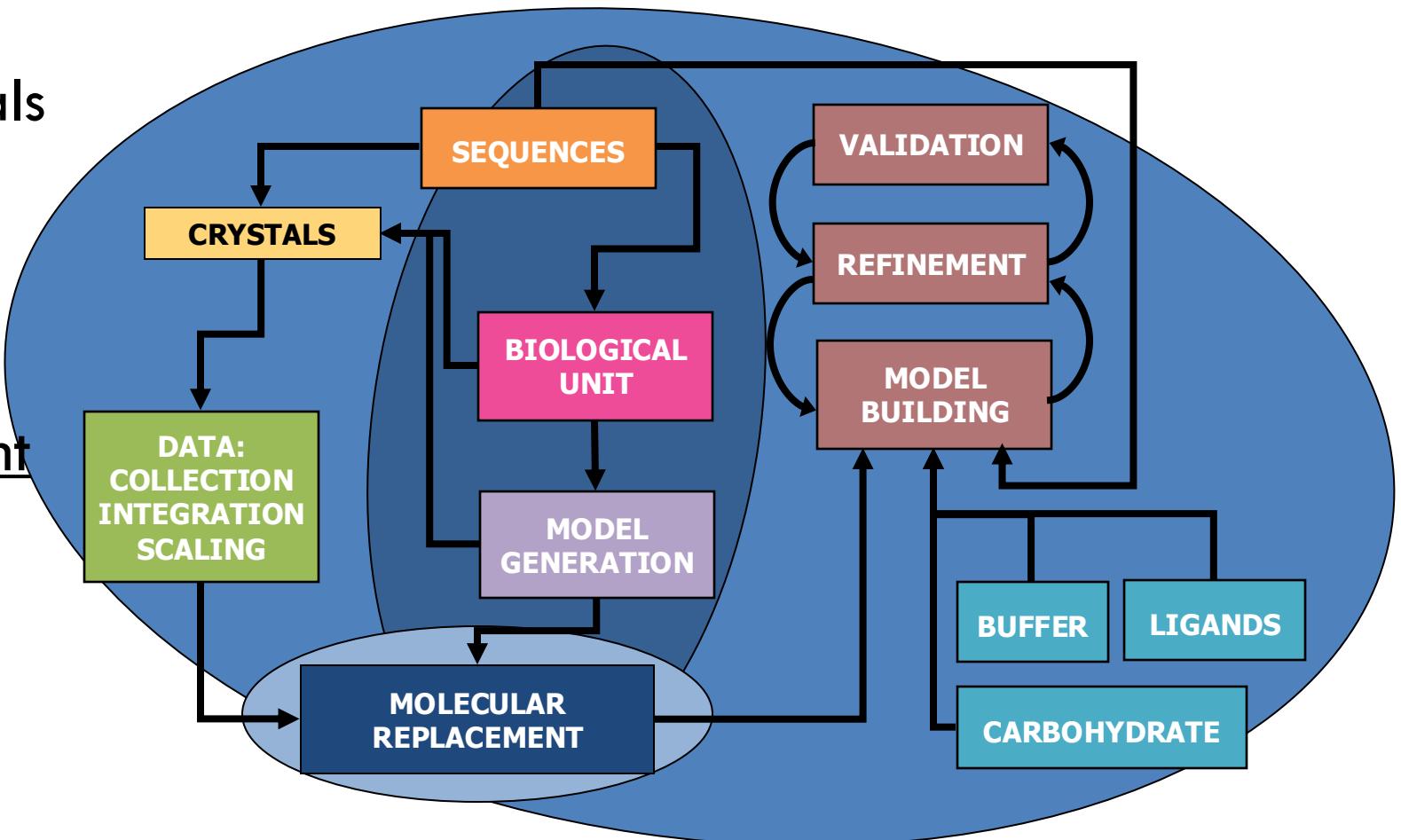
Airlie McCoy



UNIVERSITY OF  
CAMBRIDGE

# Molecular Replacement

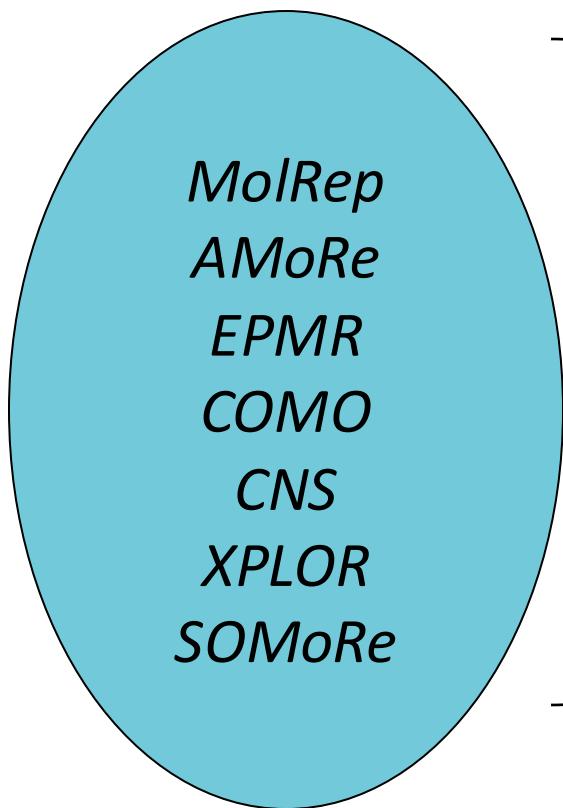
1. prepare native crystals
2. collect and prepare data
3. obtain the phases:  
molecular replacement
4. model building
5. refinement and validation



# molecular replacement terminology

- molecular replacement is a method used to solve the **phase problem** for a new crystal – the '**target**'
- it uses a structurally similar molecule – the '**model**'
  - it is usual that different 'models' are required for different parts of the target
- the model must be placed within the lattice of the new crystal to **overlie the target with low rmsd**
- molecular replacement requires a '**target function**' that determines when the model overlies the **position of the true structure**
  - match data with structure factors **calculated** from the model

# target functions for molecular replacement



Patterson based



}

Maximum Likelihood

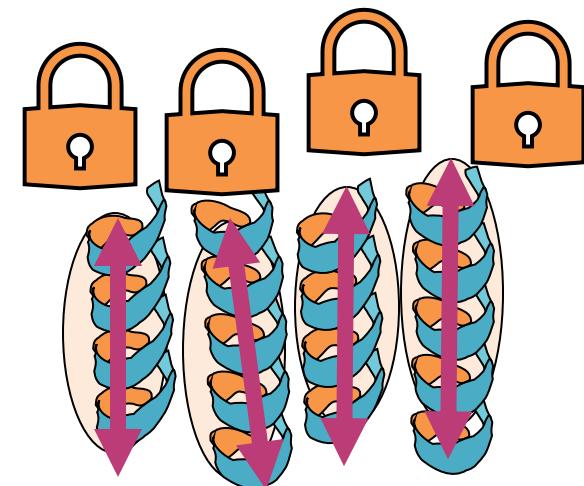
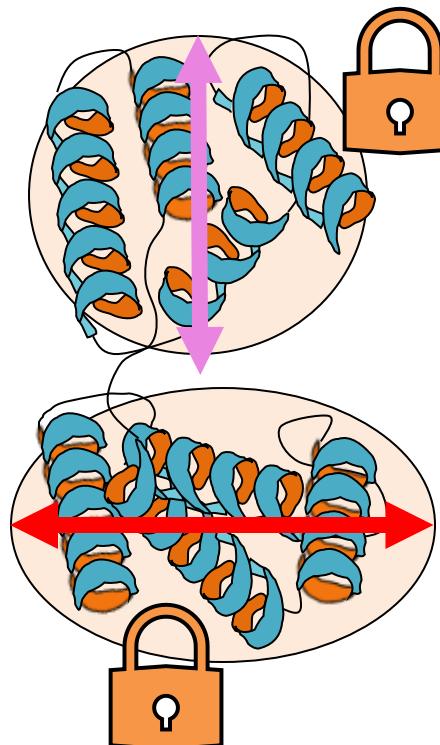
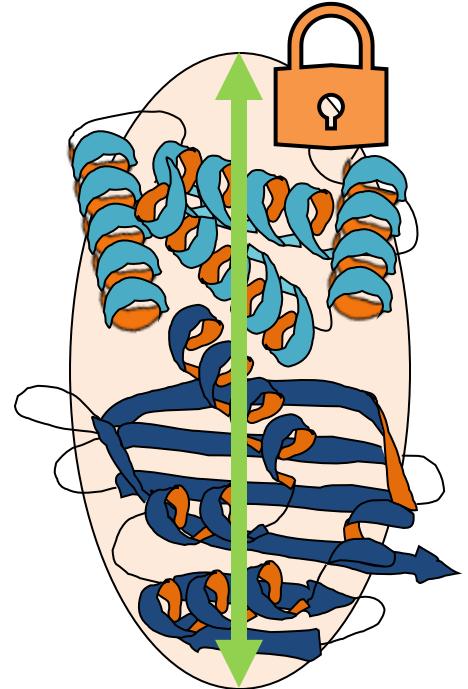
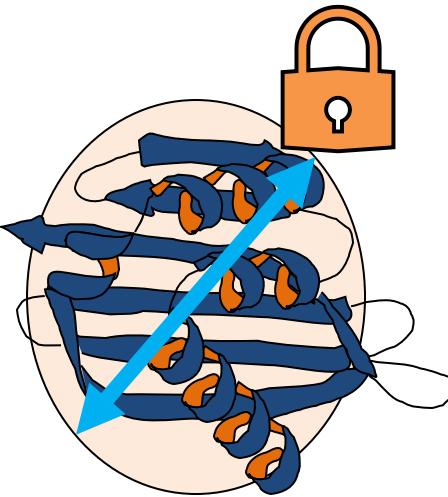
Maximum likelihood has many advantages over Patterson based methods

1. Increased signal by accounting for errors
2. Powerful statistics for predicting and judging results

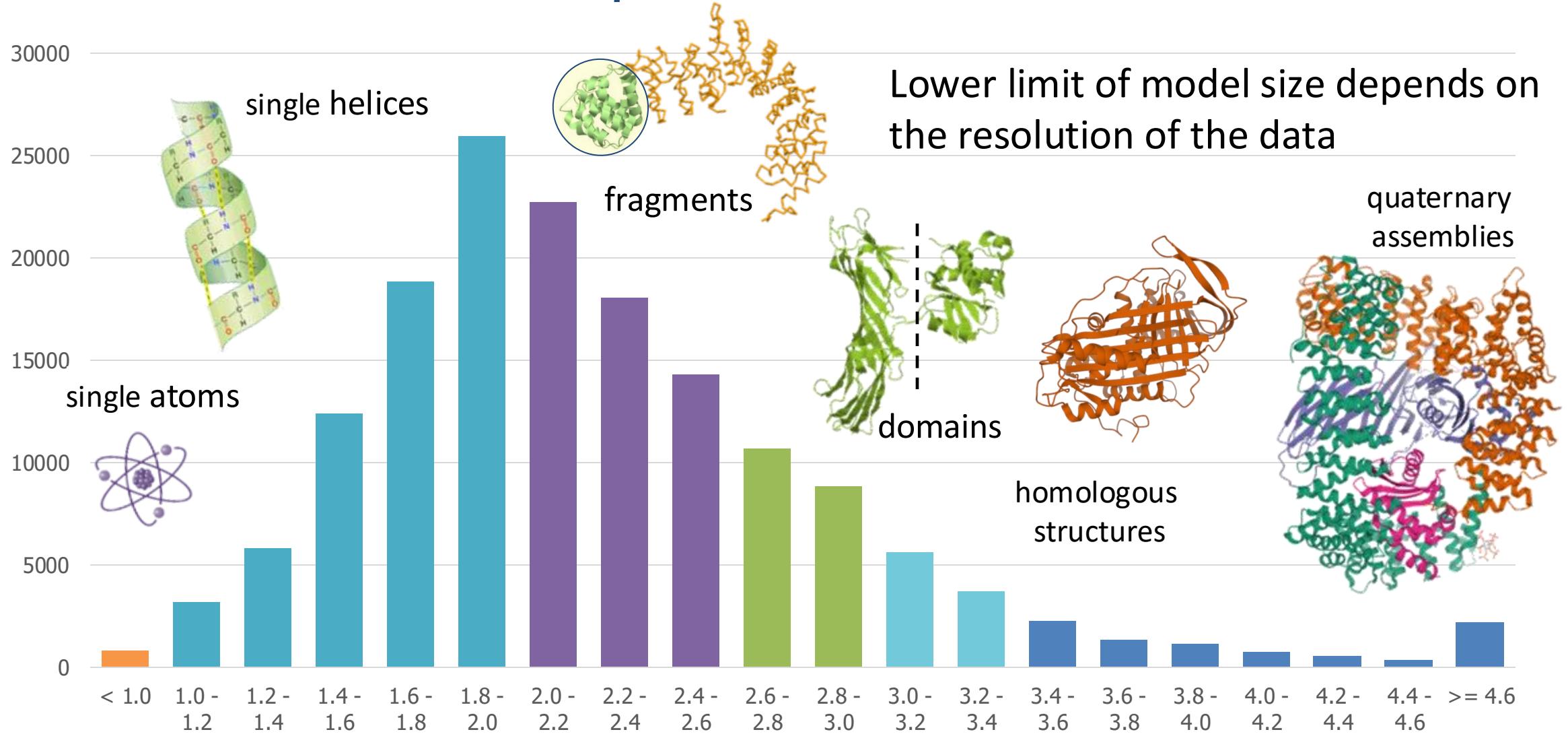
a few words about models...

## "model" – long range accuracy

- A (good) model is a (large) collection of atoms whose displacement from each other is 'locked' to be the same in the model and target
- You don't know (for certain) in advance whether the model meets this criteria... working 'blind'



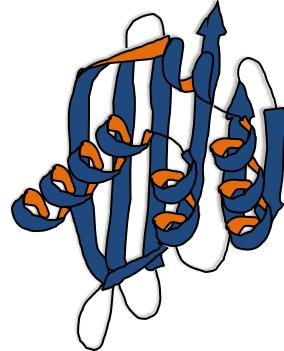
# models for molecular replacement



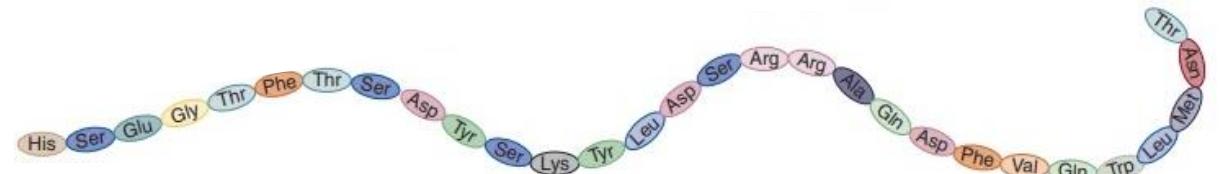
phasing by molecular replacement

# molecular replacement

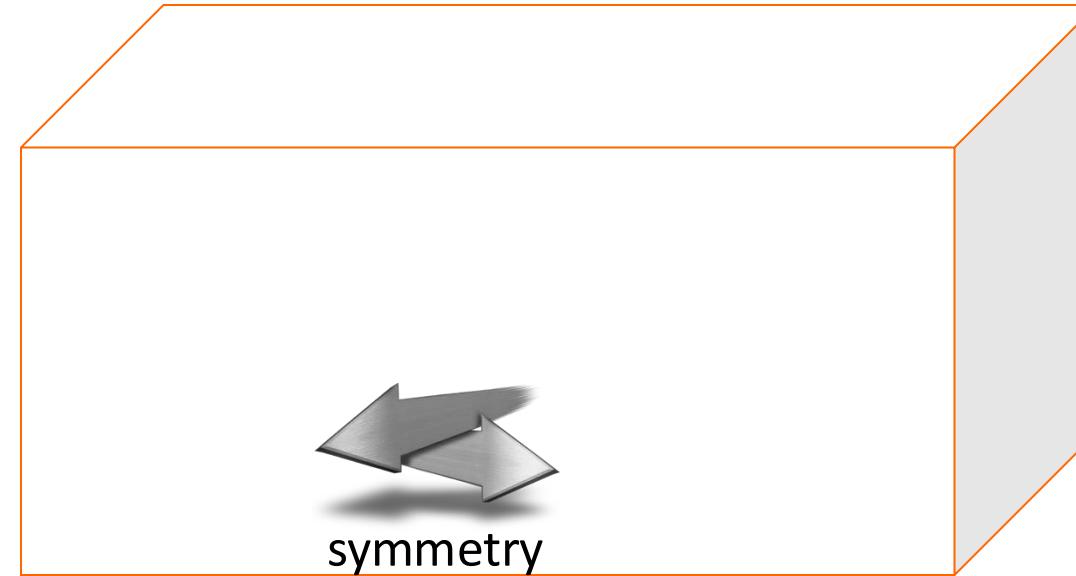
requires a model structure  
with low rmsd to the target



place the model in the  
crystal so that it has  
lowest rmsd to the  
target structure and  
use the calculated  
phases from the  
model to kickstart  
refinement

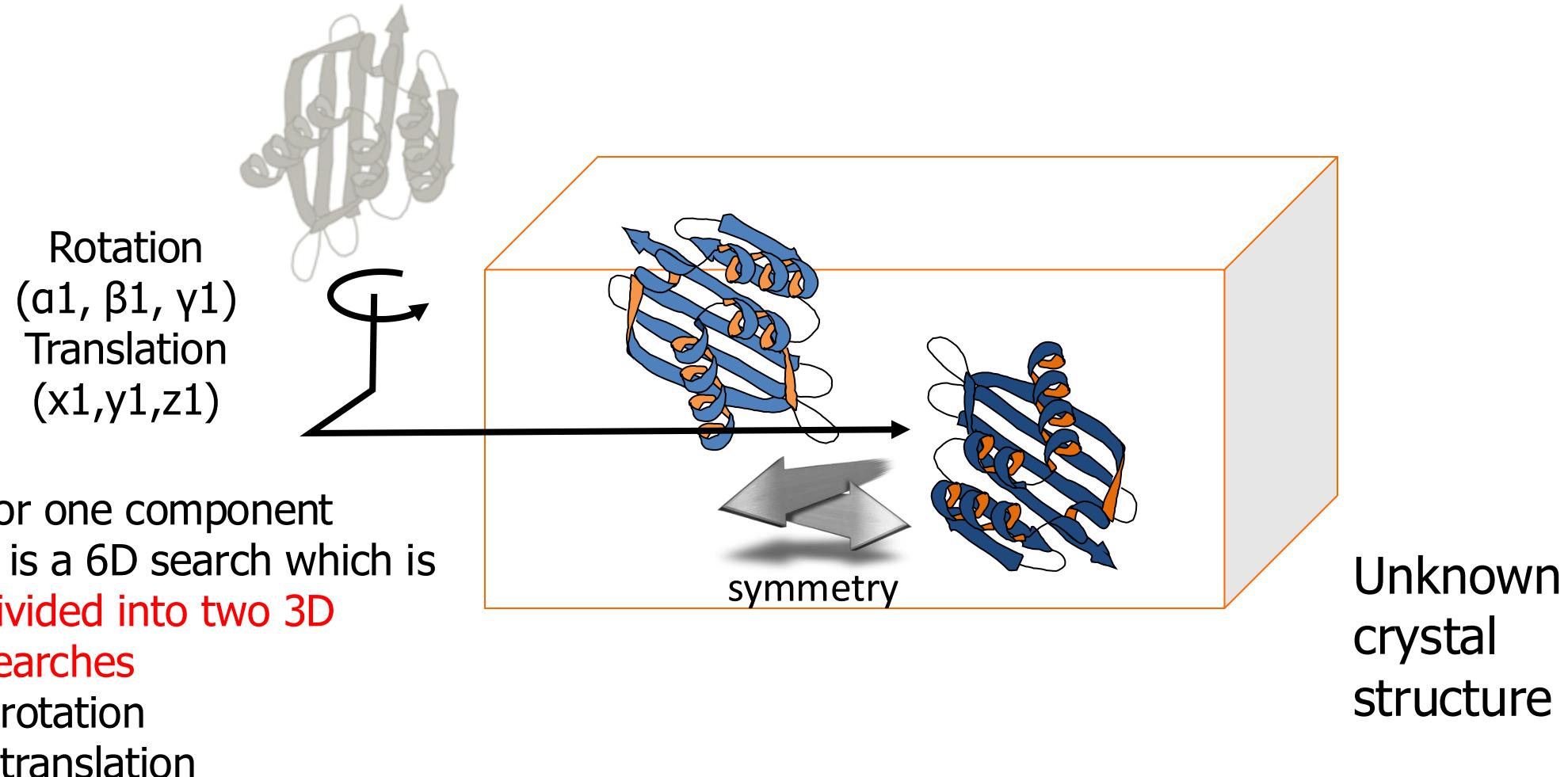


(Known) sequence/sequences  
give you a model/models



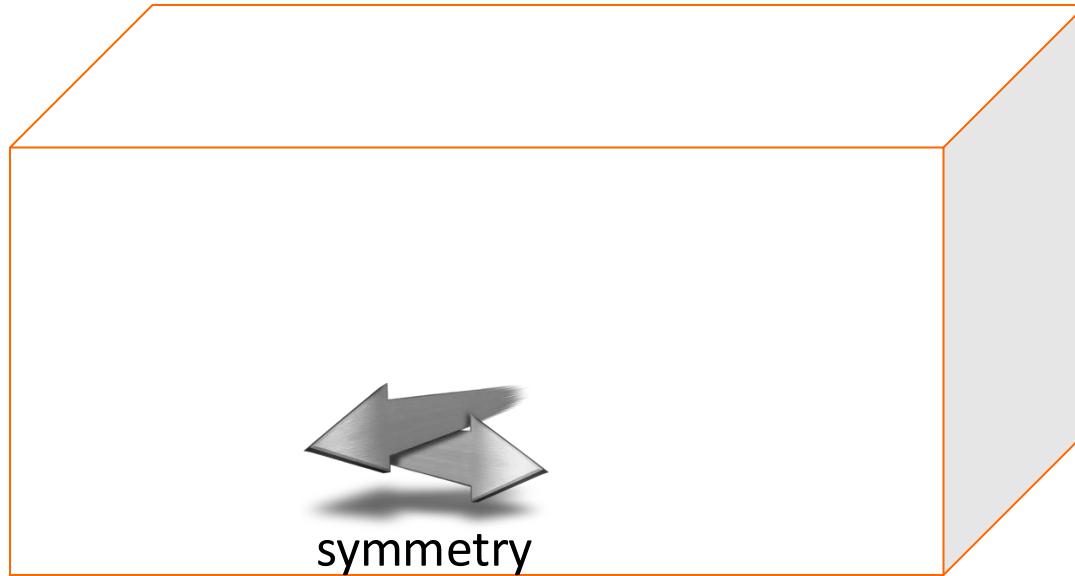
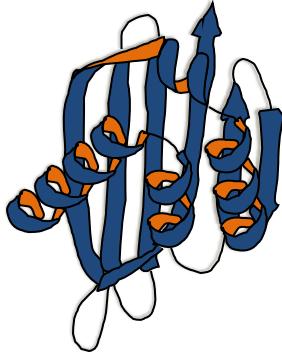
# molecular replacement – rotation and translation

requires a model structure  
with low rmsd to the target



# molecular replacement

requires a model structure  
with low rmsd to the target

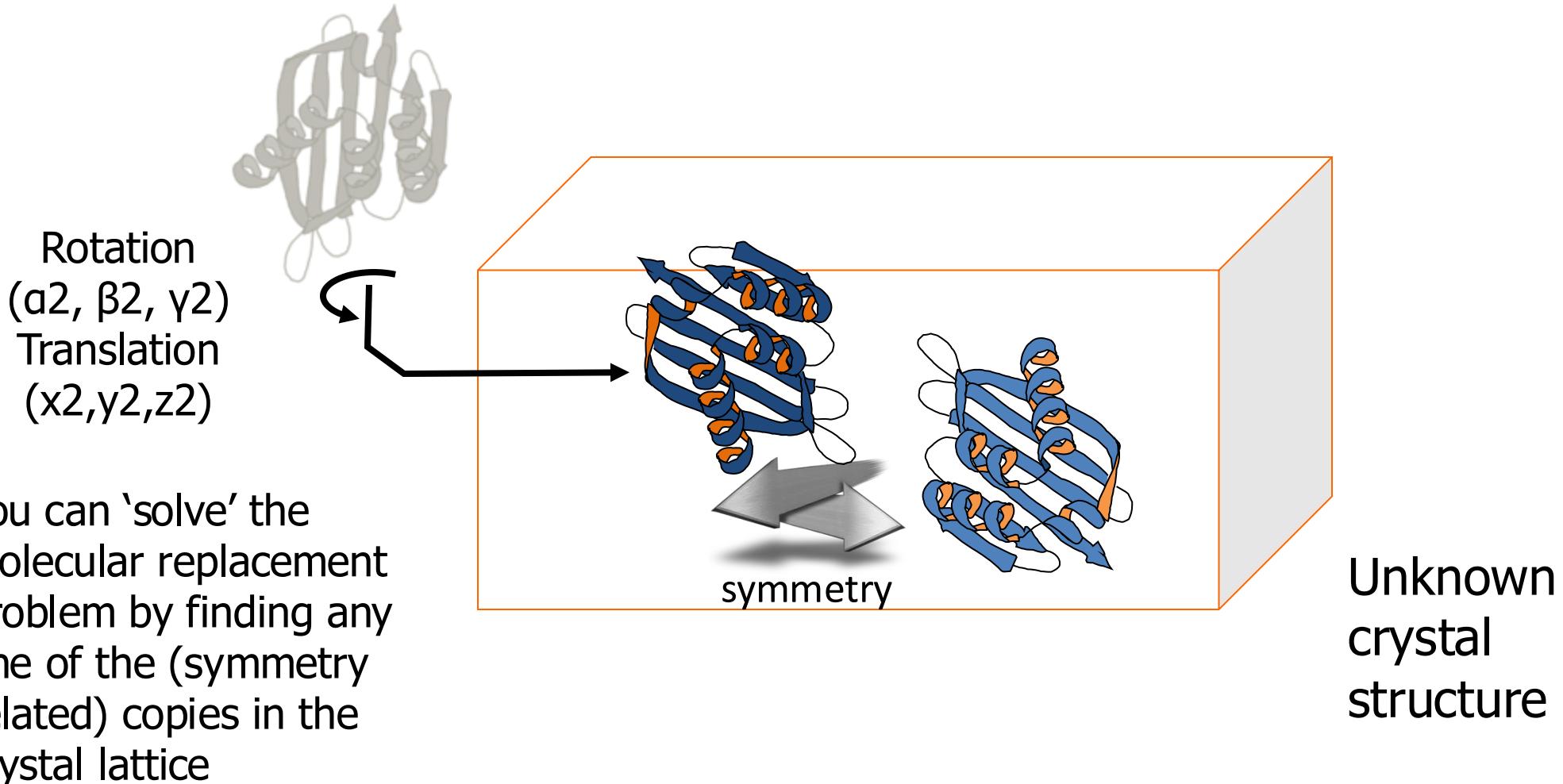


You can 'solve' the molecular replacement problem by finding any one of the (symmetry related) copies in the crystal lattice

Unknown  
crystal  
structure

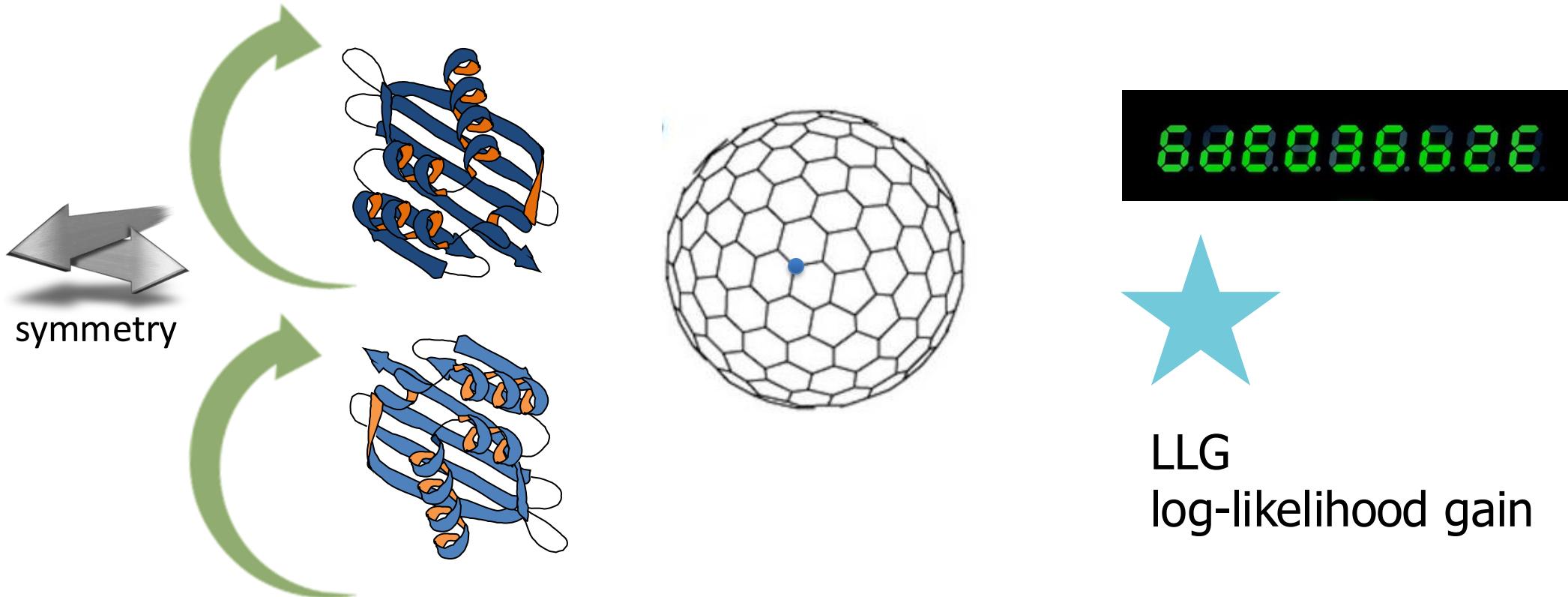
# molecular replacement – rotation and translation

requires a model structure  
with low rmsd to the target



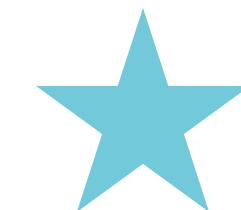
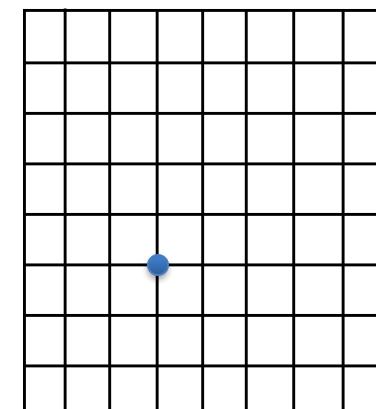
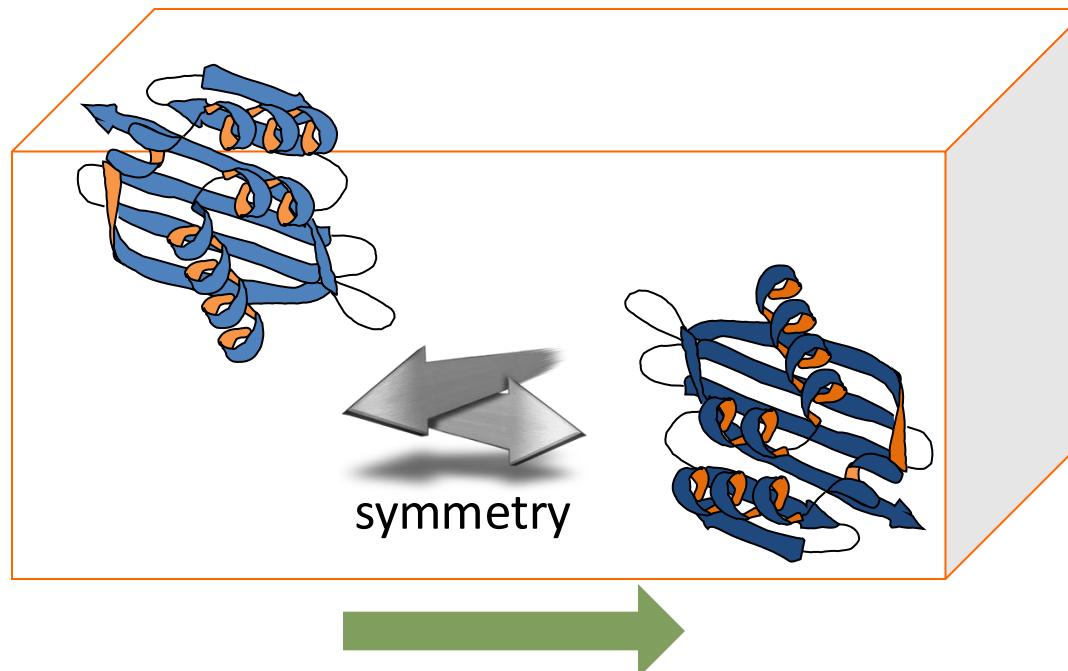
## rotation search (rotation function)

- Conceptually, orient model at in all angles on a grid
- Score each, rank and take the top or best few
- In practise, there are many speed enhancements



# translation search (translation function)

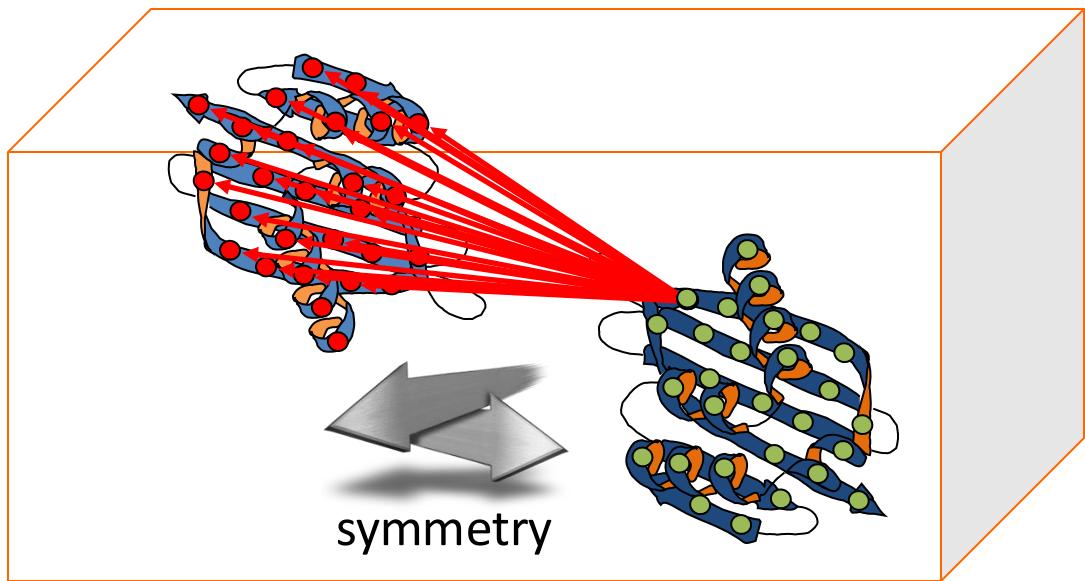
- Conceptually, place model at positions on a grid and score each
- Rank, and take the top one – or best few
- In practise, there are many speed enhancements



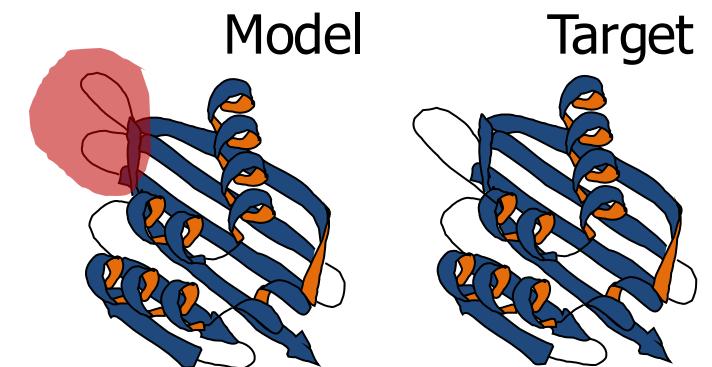
LLG  
log-likelihood gain

# packing analysis (packing function)

- C $\alpha$  clash test
- Excludes physically impossible poses, reducing search space
- No new placements



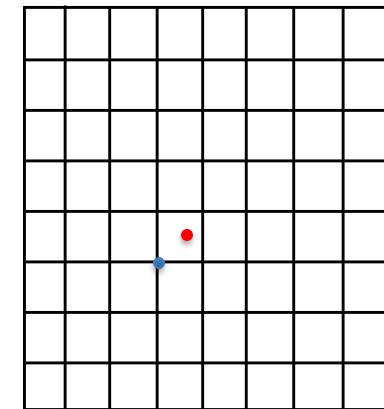
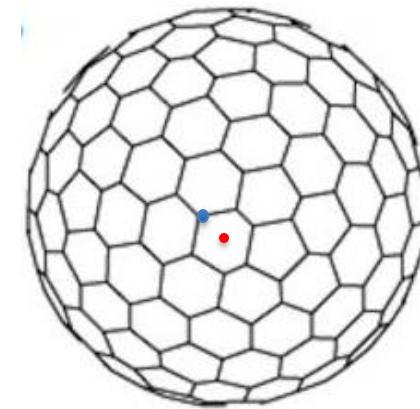
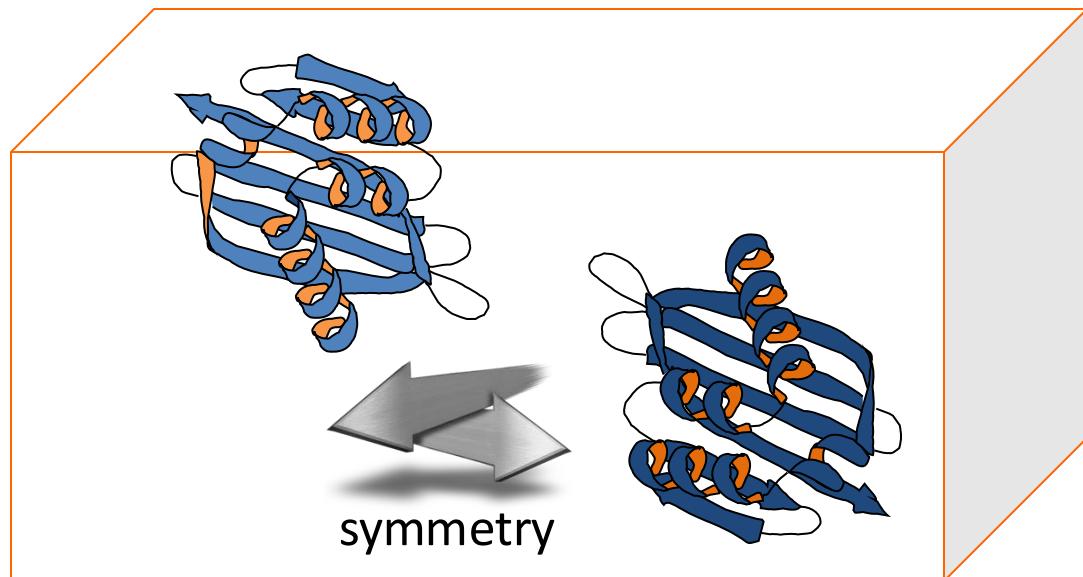
Clash with other contents  
of unit cell



Warning! Badly modelled loops can  
remove good solutions – always trim

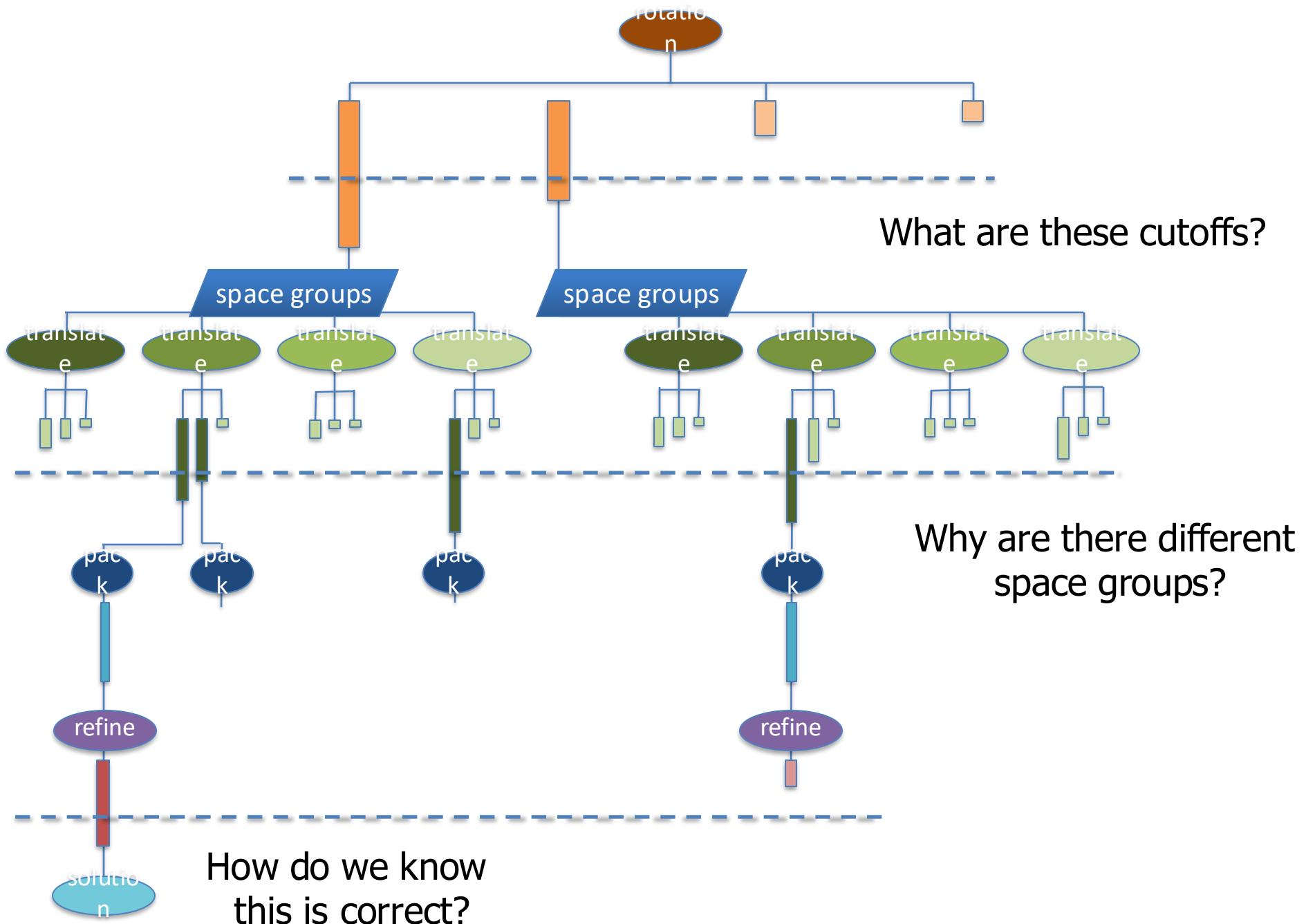
# refinement

- Optimize poses away from grid search locations



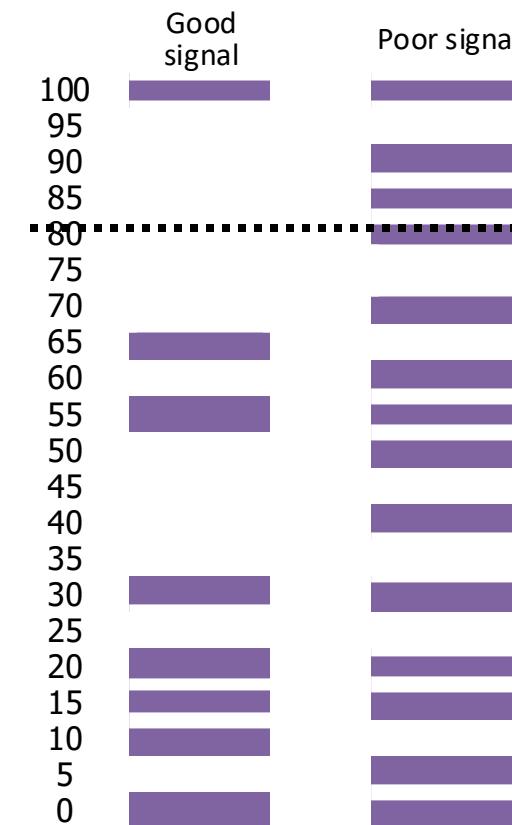
8888888888

LLG  
log-likelihood gain

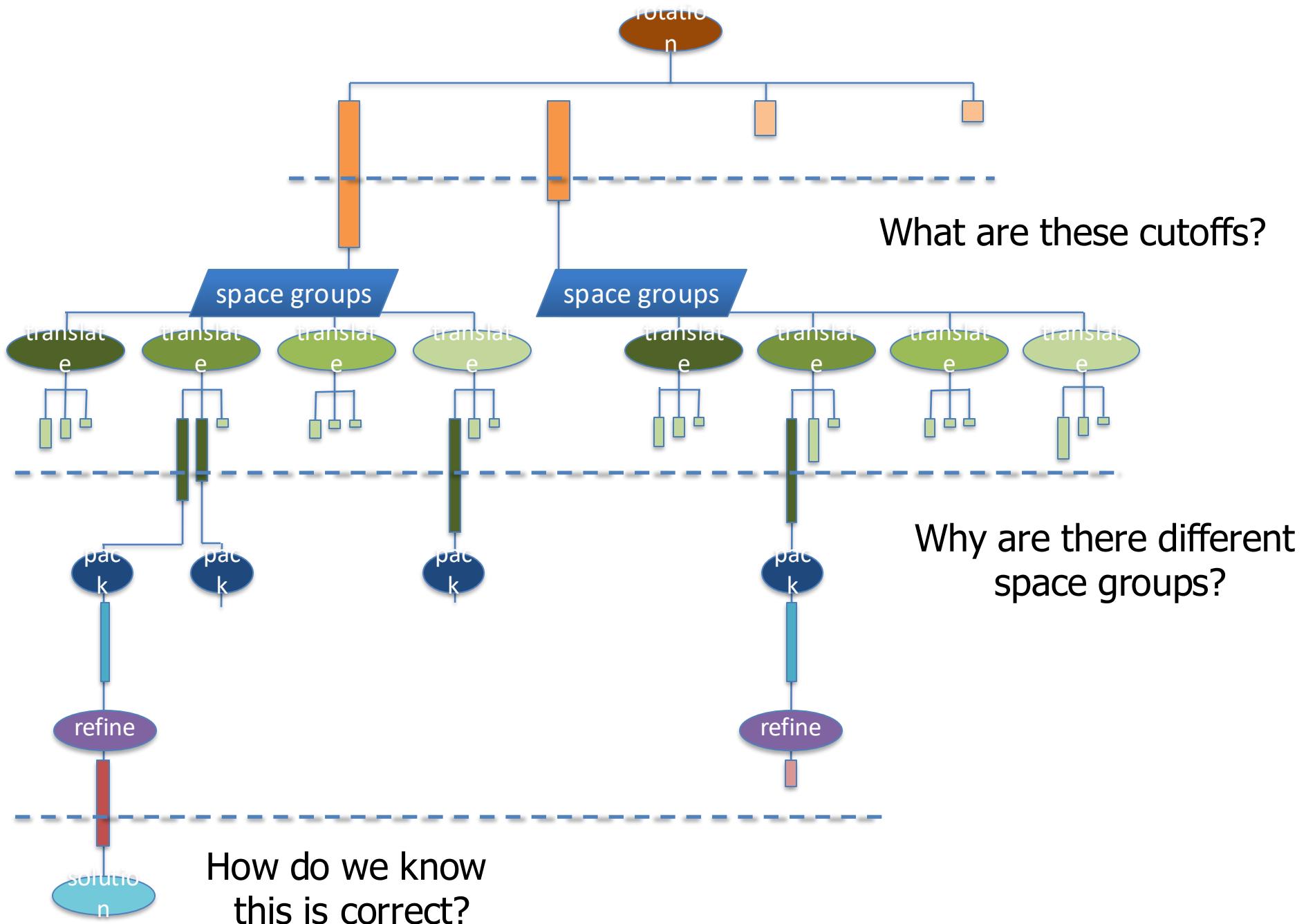


# peak selection in rotation and translation function

- Must chose a selection criteria to carry potential solutions through to the next step
- By default, solutions over 75% of the difference between the top peak and the mean are selected
  - Good signal, few potential solutions
  - Poor signal, many potential solutions
- The absolute value of the LLG is not used for decision making here\*



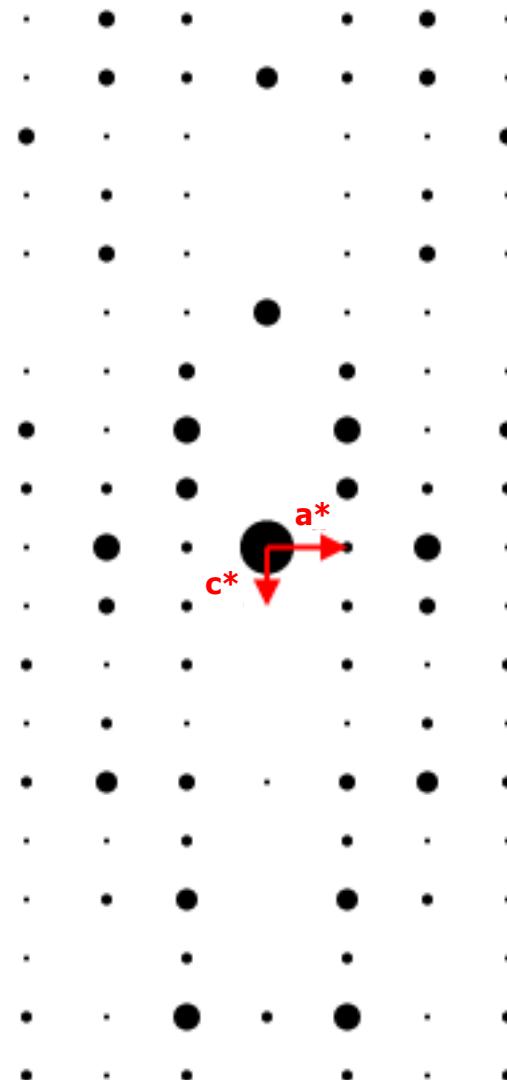
\* spoiler alert: we use percent here but the absolute value is also important!



## space group determination

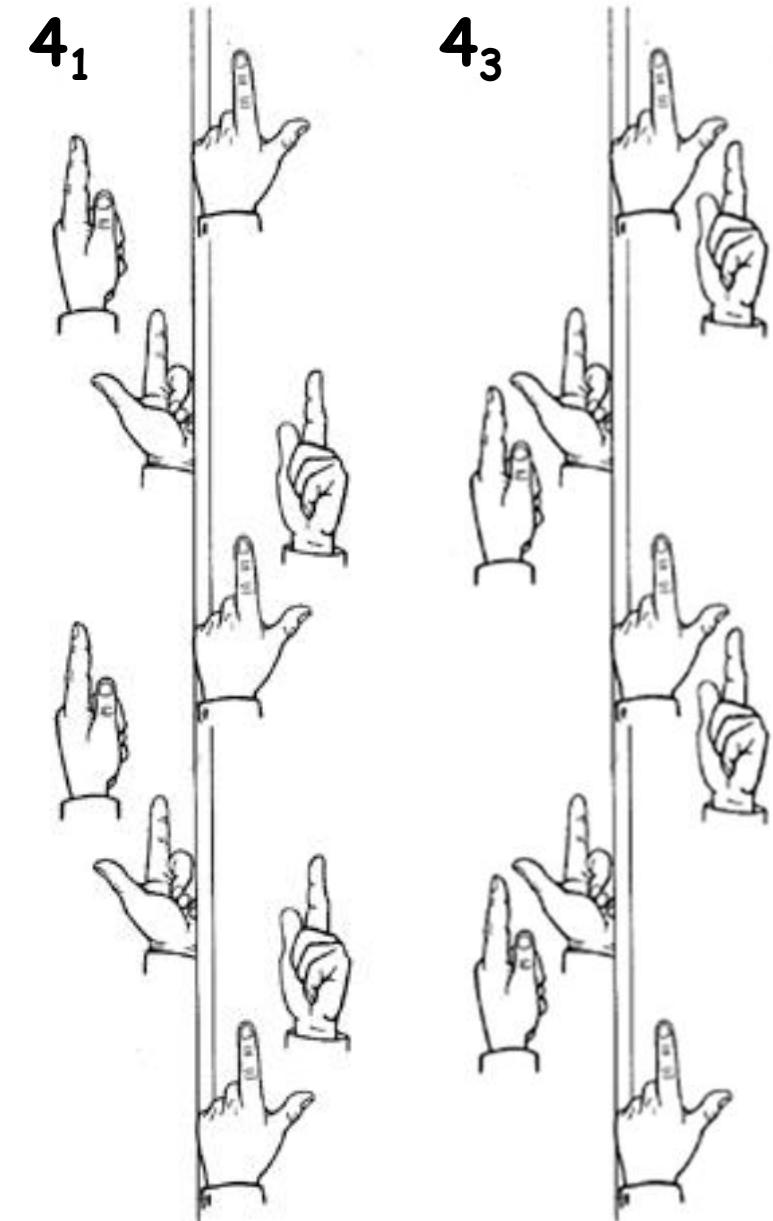
- Space groups that come in enantiomorphous pairs (e.g.  $P4_1$ ,  $P4_3$ ) cannot be distinguished at the data processing stage
- Other reasons include
  - incomplete data
  - unfortunate crystal alignment in beam
  - tiling on detector
  - short cell edge and low resolution
  - pathologies such as twinning

$4_1$  or  $4_3$



# space group determination

- Space groups that come in enantiomorphous pairs (e.g.  $P4_1$ ,  $P4_3$ ) cannot be distinguished at the data processing stage
- **The space group is only confirmed when the structure is solved**
  - or, solved ‘simultaneously’
- Phaser will run enantiomorphous pairs in the point group in the translation function
  - Or all space groups in the point group



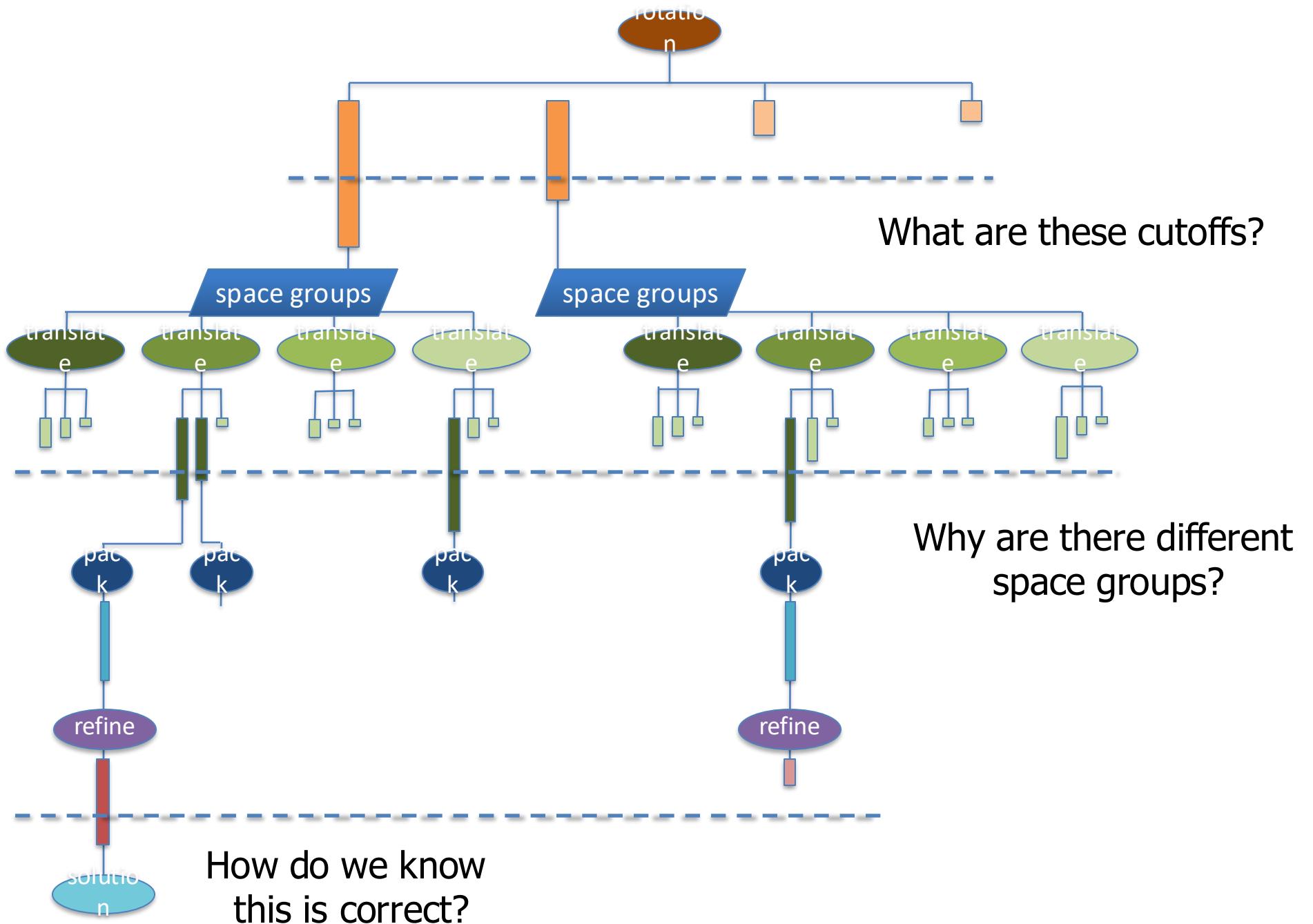


Diagram: Martin Noble

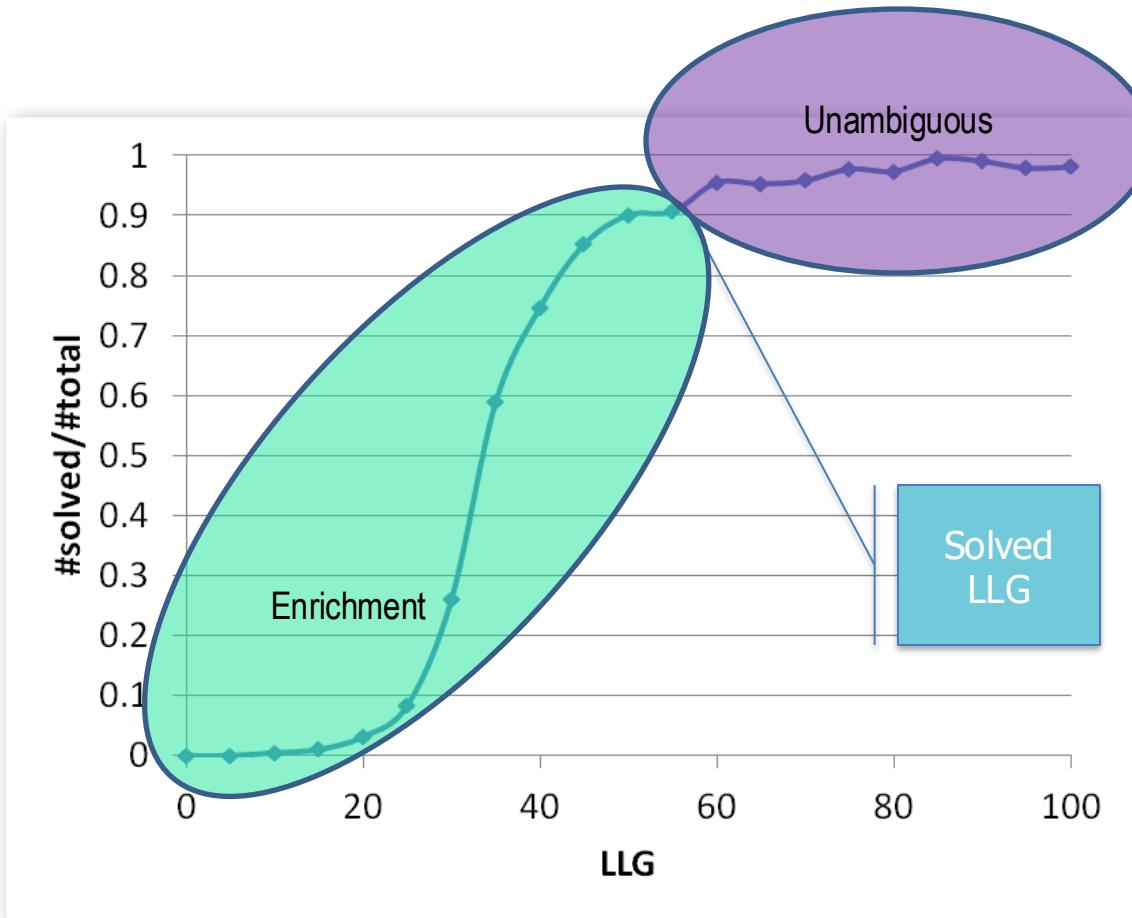
# maximum likelihood scoring in phaser

- Use probability
- Probabilities account for errors
  - Errors in the data, using  $I$  and  $\sigma I$  (intensities)
  - Errors in the model

$$LLGI = \sum_{\mathbf{h}} \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$

$E_e$  and  $D_{obs}$  are defined as in (Read & McCoy, 2016);  $E_e$  is the effective  $E$ , representing information derived from  $E_{obs}^2$ , and  $D_{obs}$  represents the reduction in correlation between observation and  $E_e$  arising from experimental error;  $E_{obs}^2 = I_{obs}/(\varepsilon \Sigma_N)$  where  $\varepsilon$  and  $\Sigma_N$  includes correction terms for anisotropy and tNCS modulations

# log-likelihood gain for solutions



Plot of LLG versus success in structure solution

Database of  
over 23000  
MR problems

R.D. Oeffner

# When is a model correctly placed?

TF Z-score	LLG score	Solved?
< 5	< 25	no
5 - 6	25 - 36	unlikely
6 - 7	36 - 49	possibly
7 - 8	49 - 64	probably
> 8	> 64	definitely

# When is a model correctly placed?

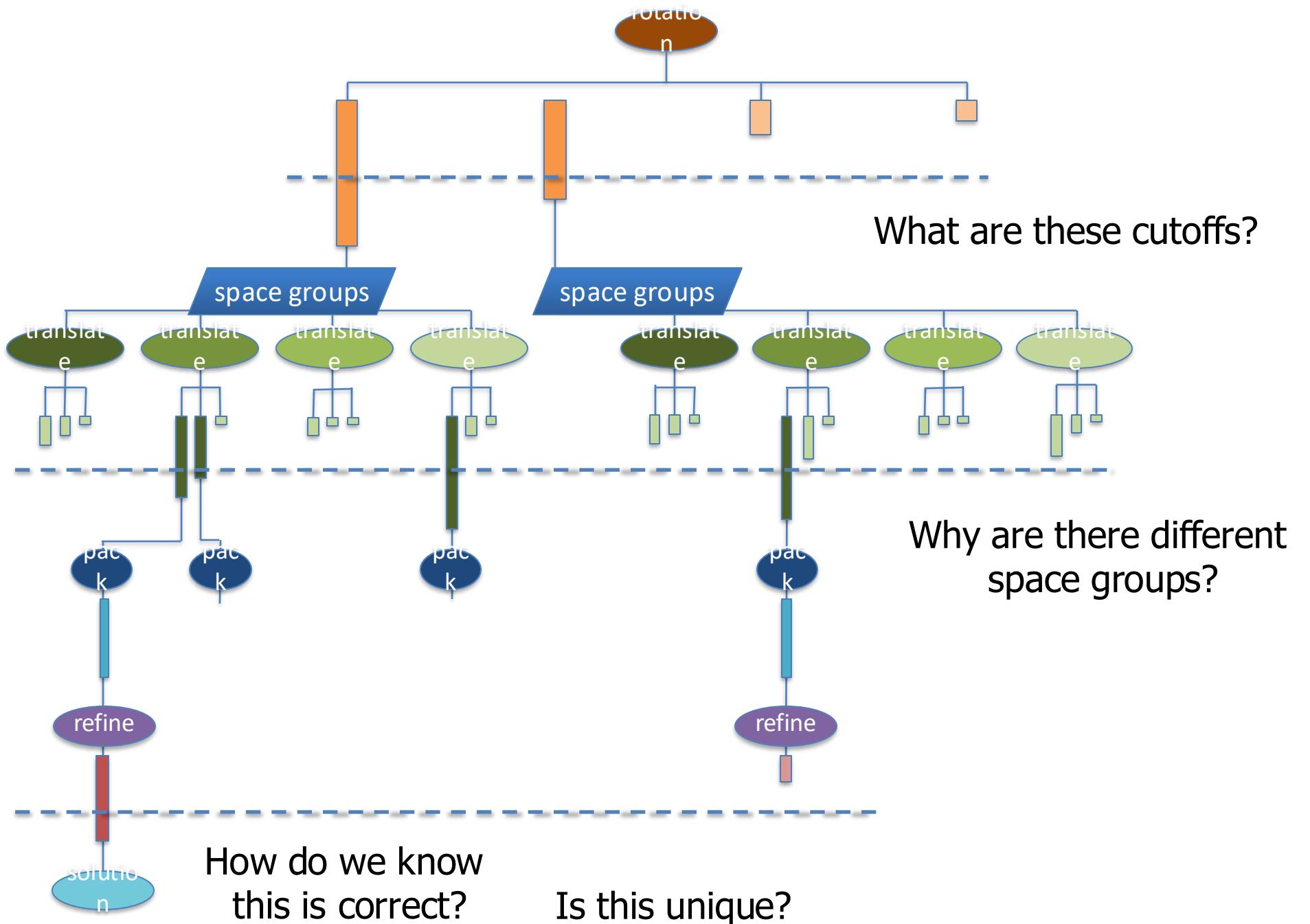
TF Z-score	LLG score	Solved?
< 5	< 25	no
5 - 6	25 - 36	unlikely
6 - 7	36 - 49	possibly
7 - 8	49 - 64	probably
> 8	> 64	definitely

If we could estimate this in advance of structure solution we could know if a structure was solvable with the data and the model.

## The ‘expected LLG’

- For an estimate of the rmsd between your model and the target...
  - And an estimate of the number of copies...
- The ‘expected LLG’ predicts how easy it will be to solve a structure with a given model
- There will be different eLLG values for different assumptions about the accuracy of the model

eLLG score	Solvable?
< 25	no
25 - 36	unlikely
36 - 49	possibly
49 - 64	probably
> 64	definitely



**alternative asymmetric units  
alternative origins**

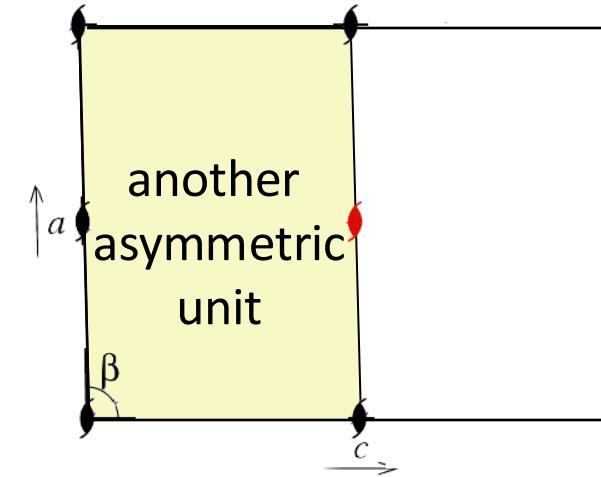
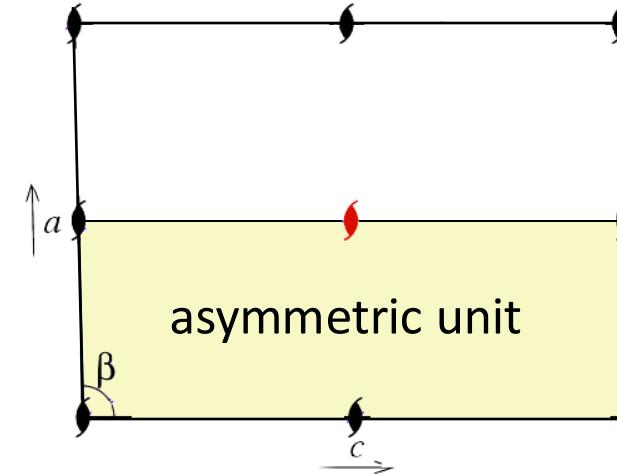
## asymmetric unit

- Even if you never solve a crystal structure you need to know about asymmetric units
- It is (one of) the fundamental concept that distinguishes crystallographic structures from cryo-em
- Crystallography gives you the structure *in the asymmetric unit*
  - *It is what is deposited!!*

# asymmetric unit

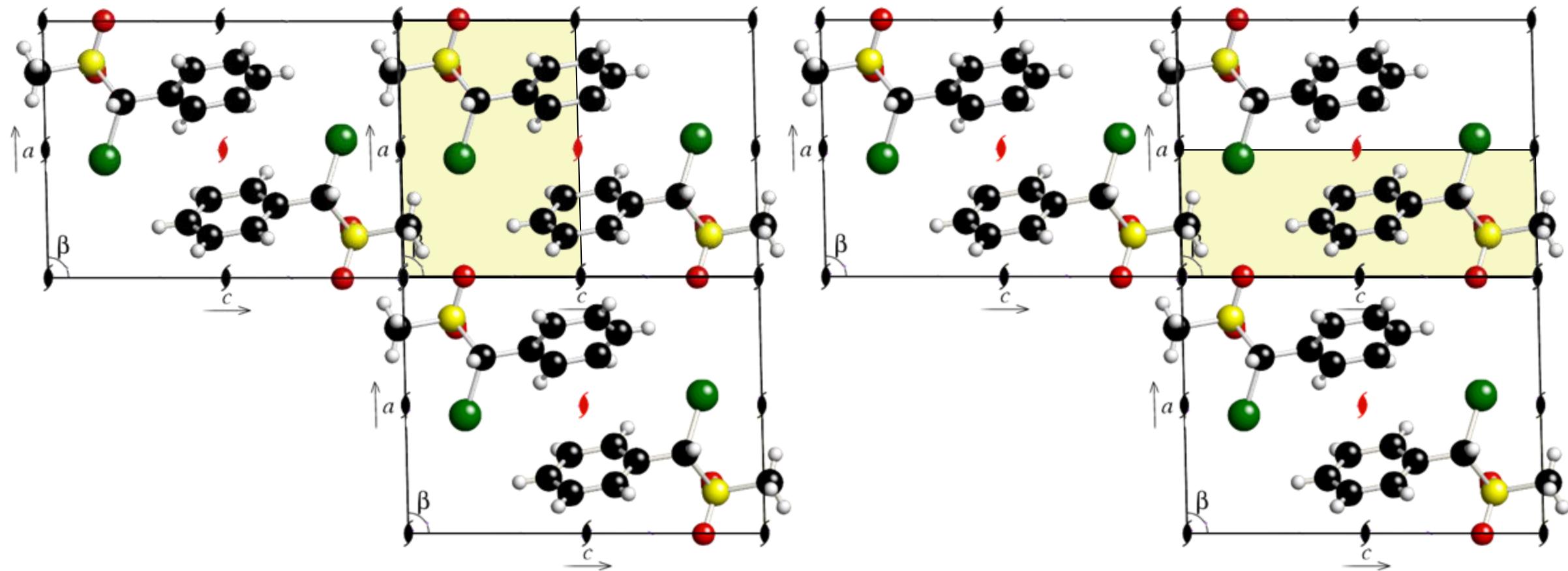
- The asymmetric unit is that part of the unit cell which can be used to generate the complete unit cell by the crystal symmetry
- The choice of asymmetric unit is not unique
  - We say '**the** asymmetric unit'
  - but mean '**an** asymmetric unit'
- The asymmetric unit is chosen by phaser according to 'closest to origin'

P2<sub>1</sub>

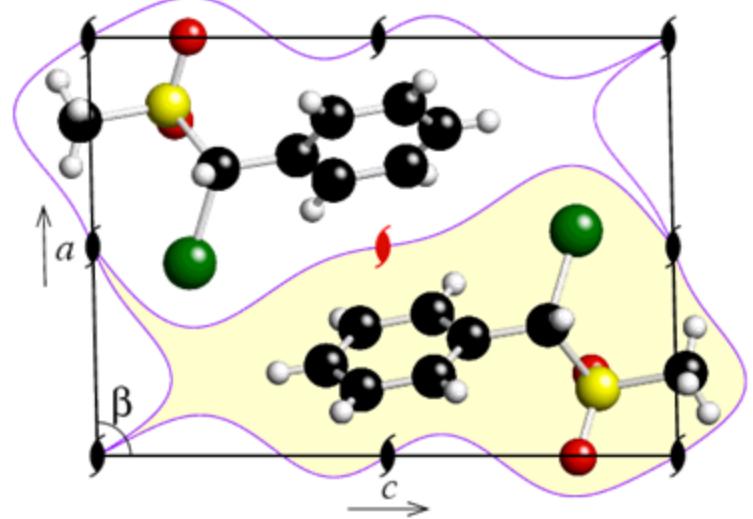


etc...

# asymmetric unit



# asymmetric unit

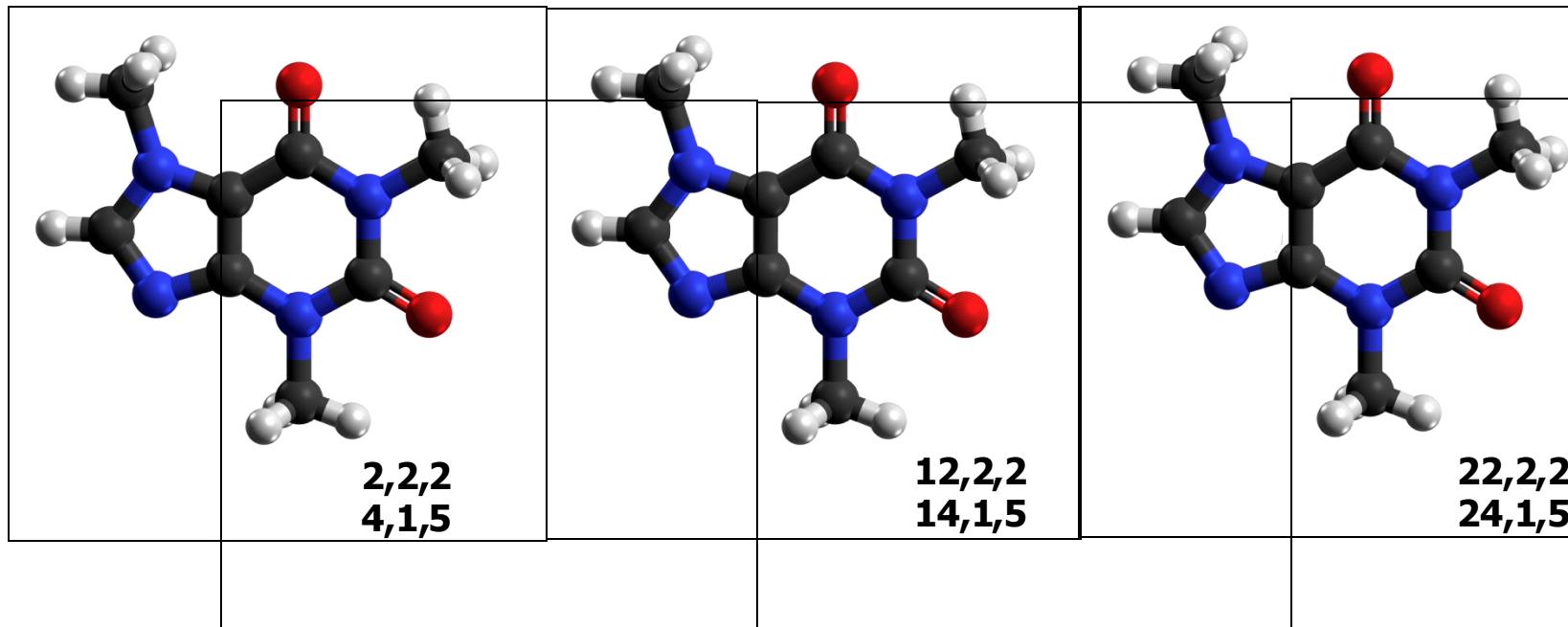


**unit cell/asymmetric unit**



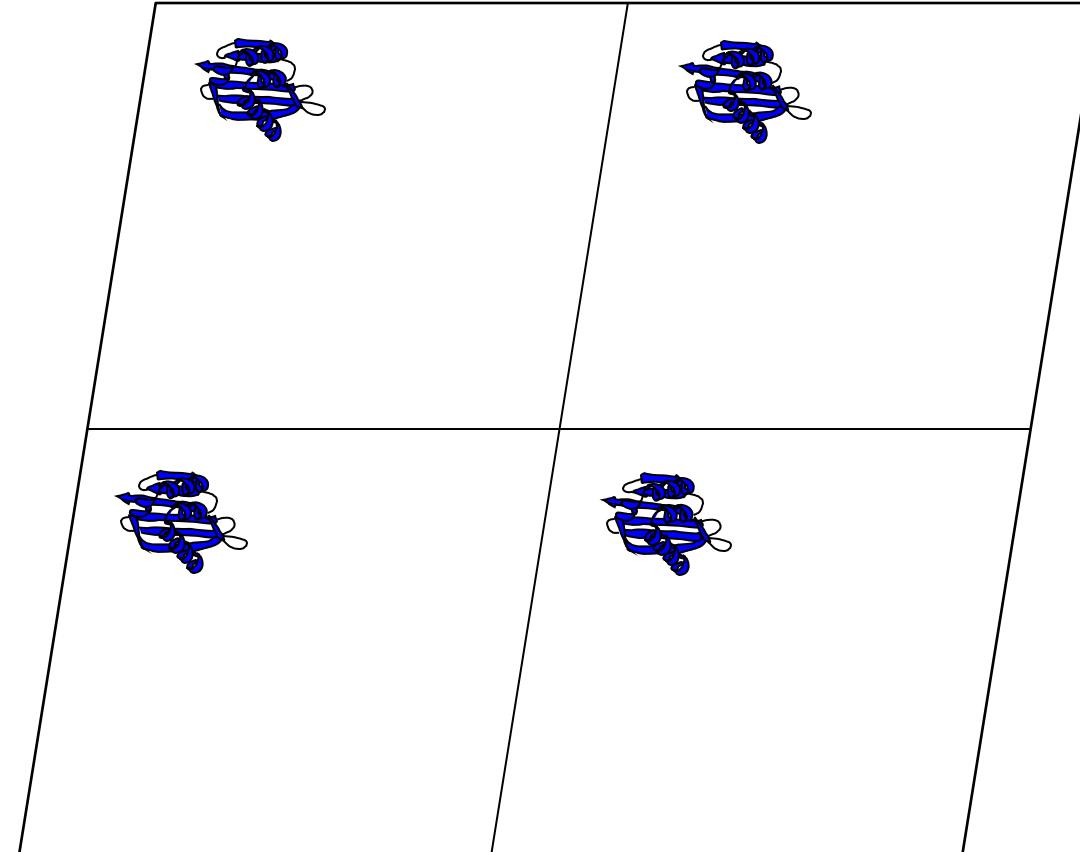
# origins

- Origins are defined by the (rotational/screw) symmetry operations
  - if there is no rotation symmetry, the origin/placement of the molecule is arbitrary



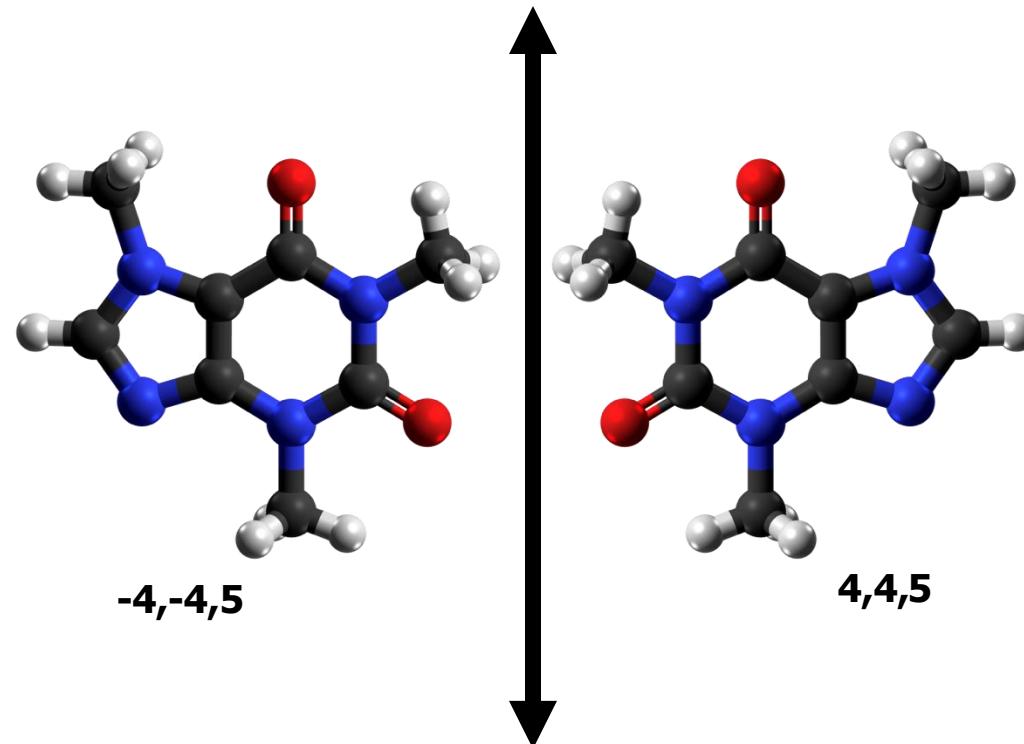
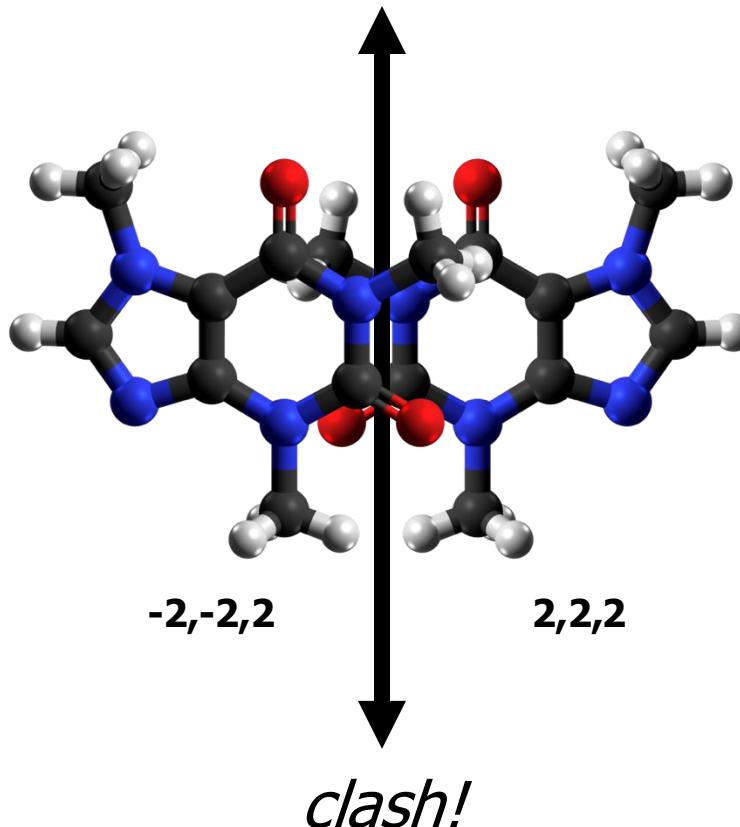
# P1

- Origin arbitrary
- No ‘translation function’ in P1



## origins

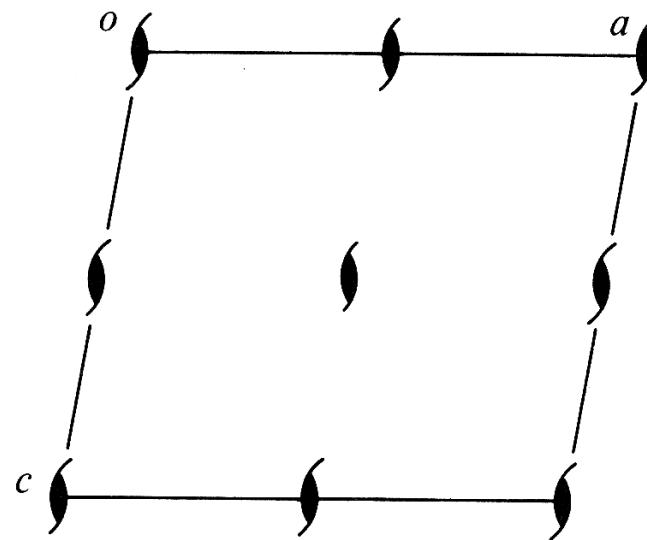
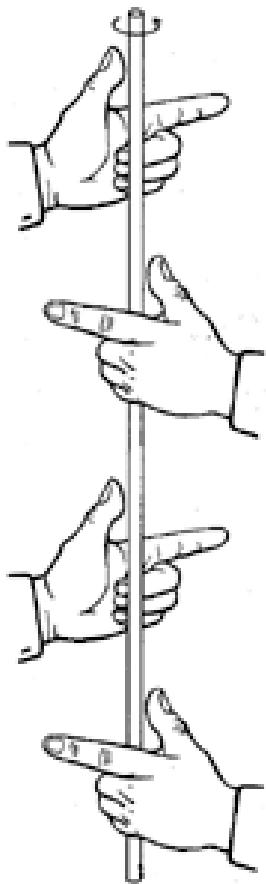
- Origins are defined by the (rotational/screw) symmetry operations, e.g.  $(x,y,z)$   $(-x,-y,z)$ 
  - translations are defined perpendicular to the rotation axis



# **P 2<sub>1</sub>**

No. 4

UNIQUE AXIS *b*

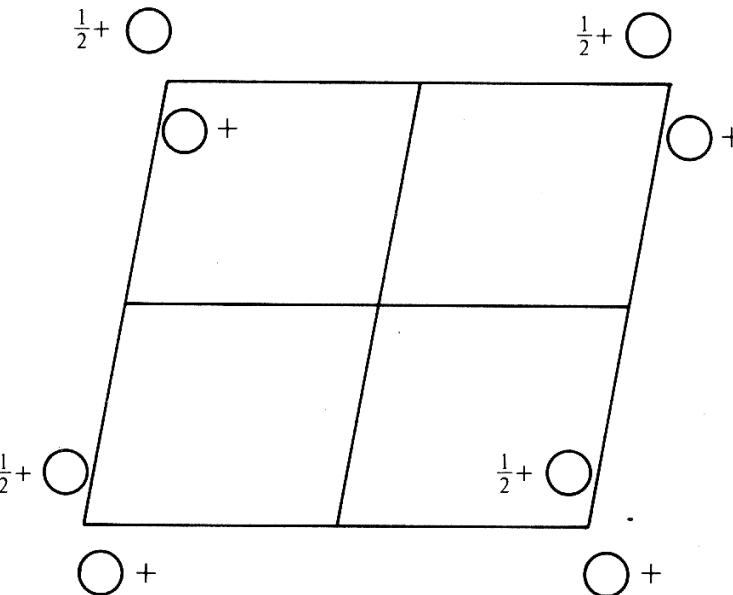


Origin on 2<sub>1</sub>

**Asymmetric unit**     $0 \leq x \leq 1; 0 \leq y \leq 1; 0 \leq z \leq \frac{1}{2}$

**Symmetry operations**

- (1) 1
- (2) 2(0,  $\frac{1}{2}$ , 0)    0,  $y$ , 0



**Positions**

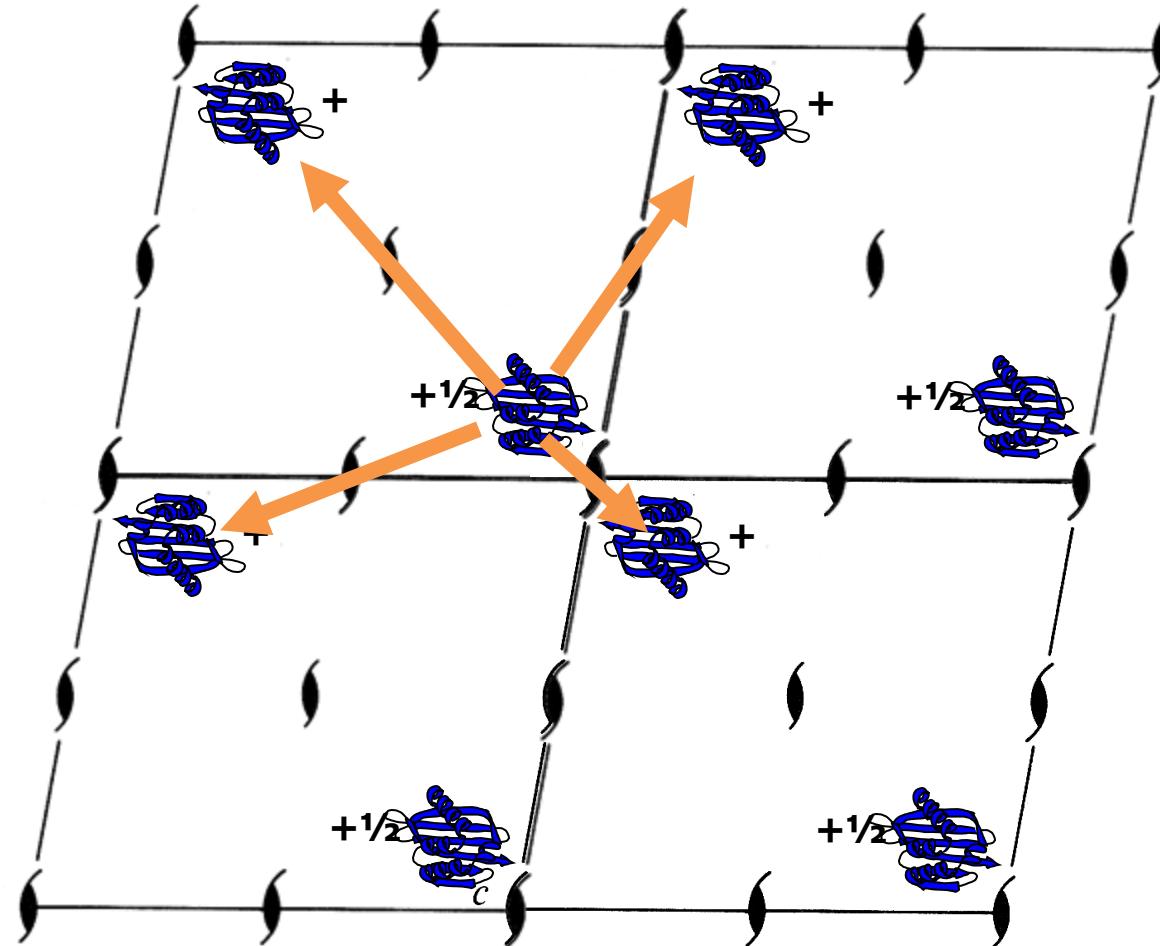
Multiplicity,  
Wyckoff letter,  
Site symmetry

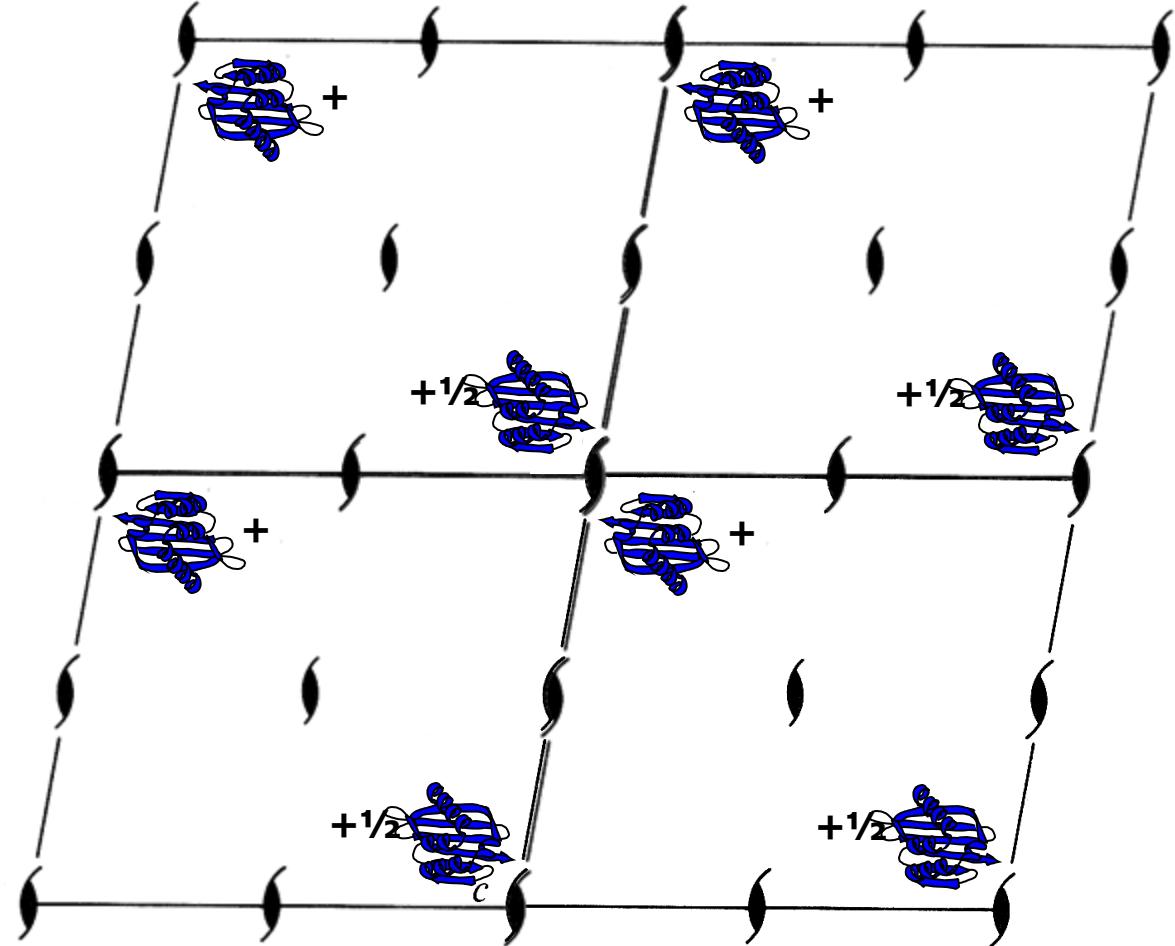
2    *a*    1    (1)  $x, y, z$     (2)  $\bar{x}, y + \frac{1}{2}, \bar{z}$

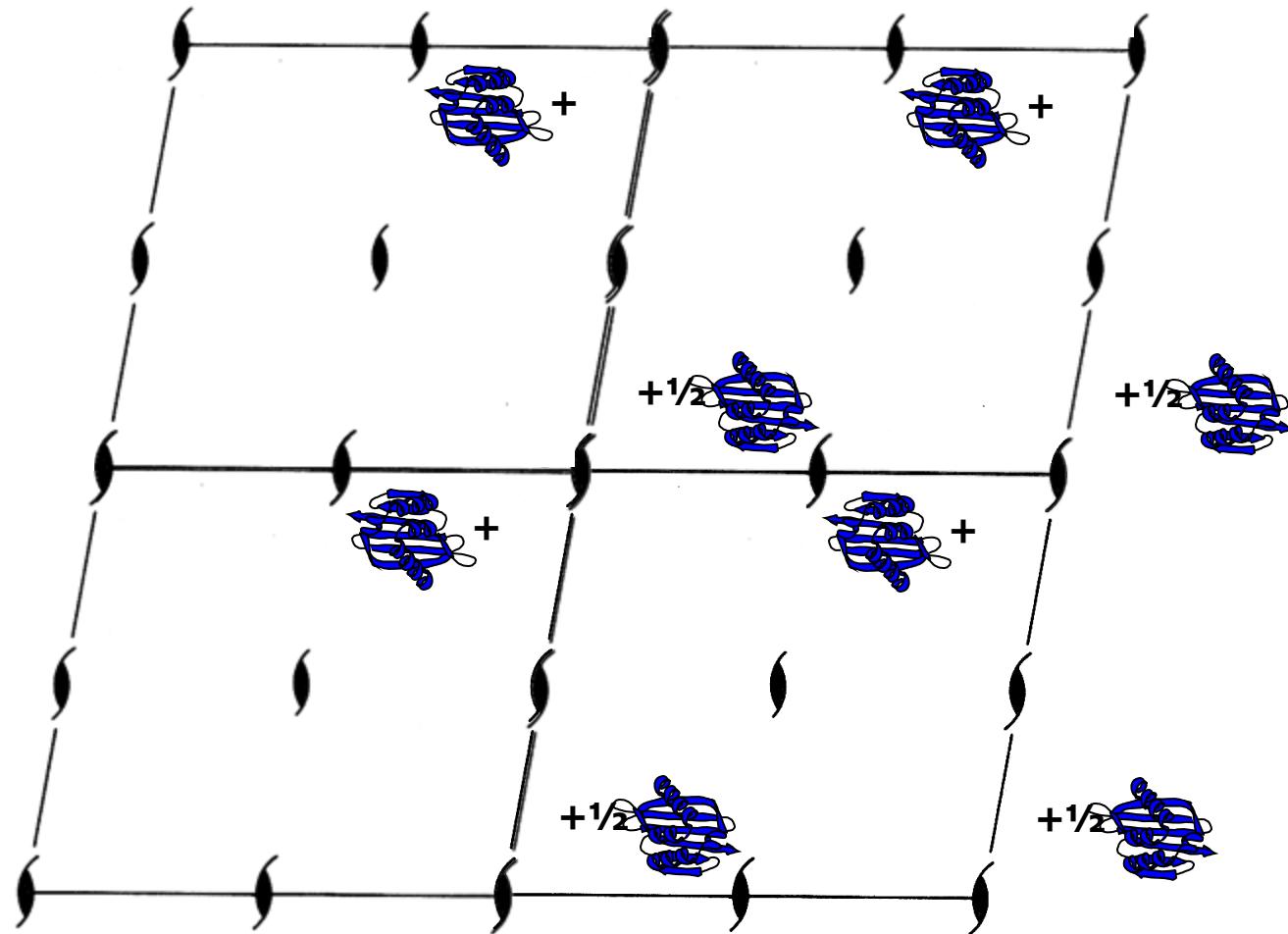
**Coordinates**

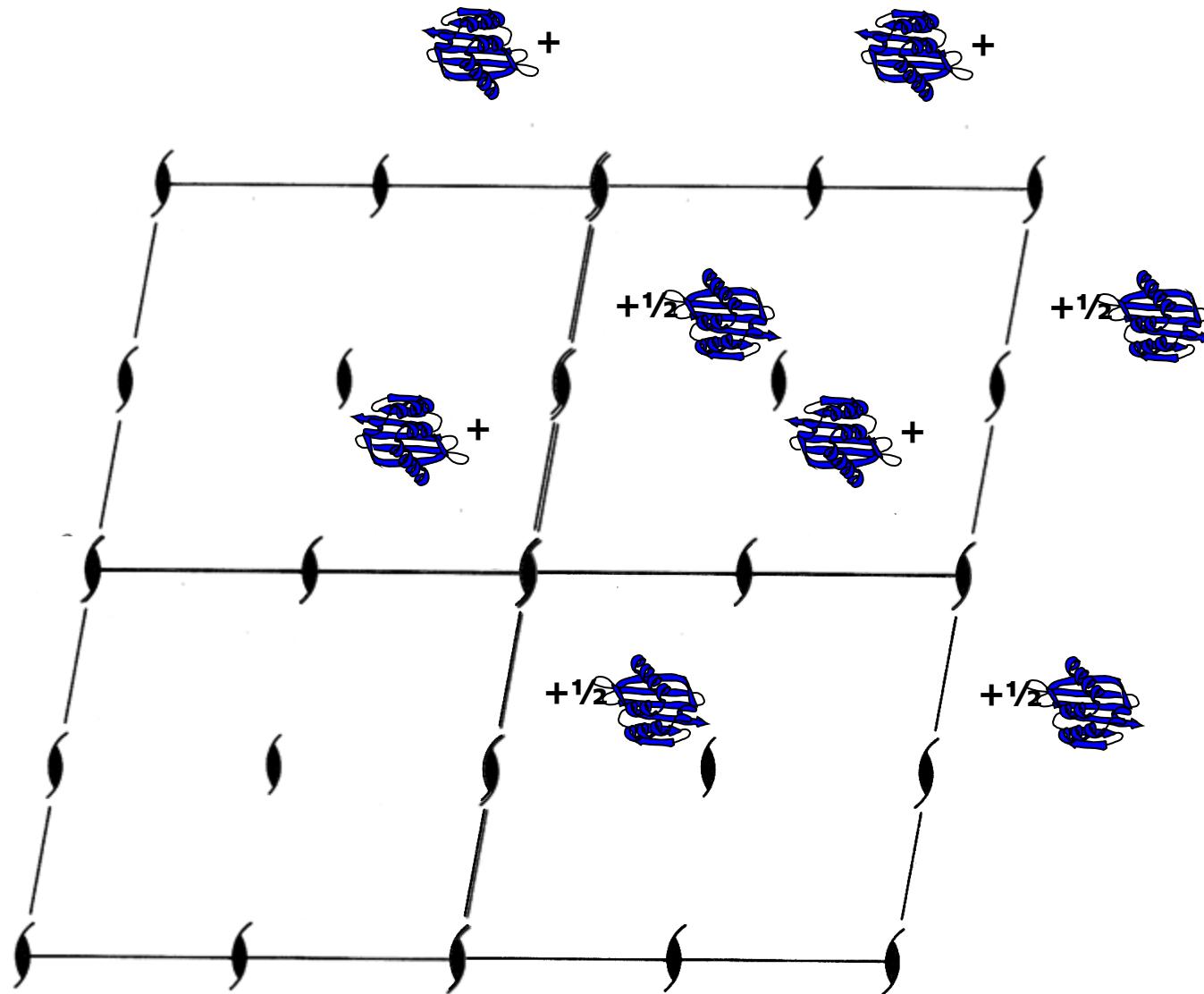
# P2<sub>1</sub>

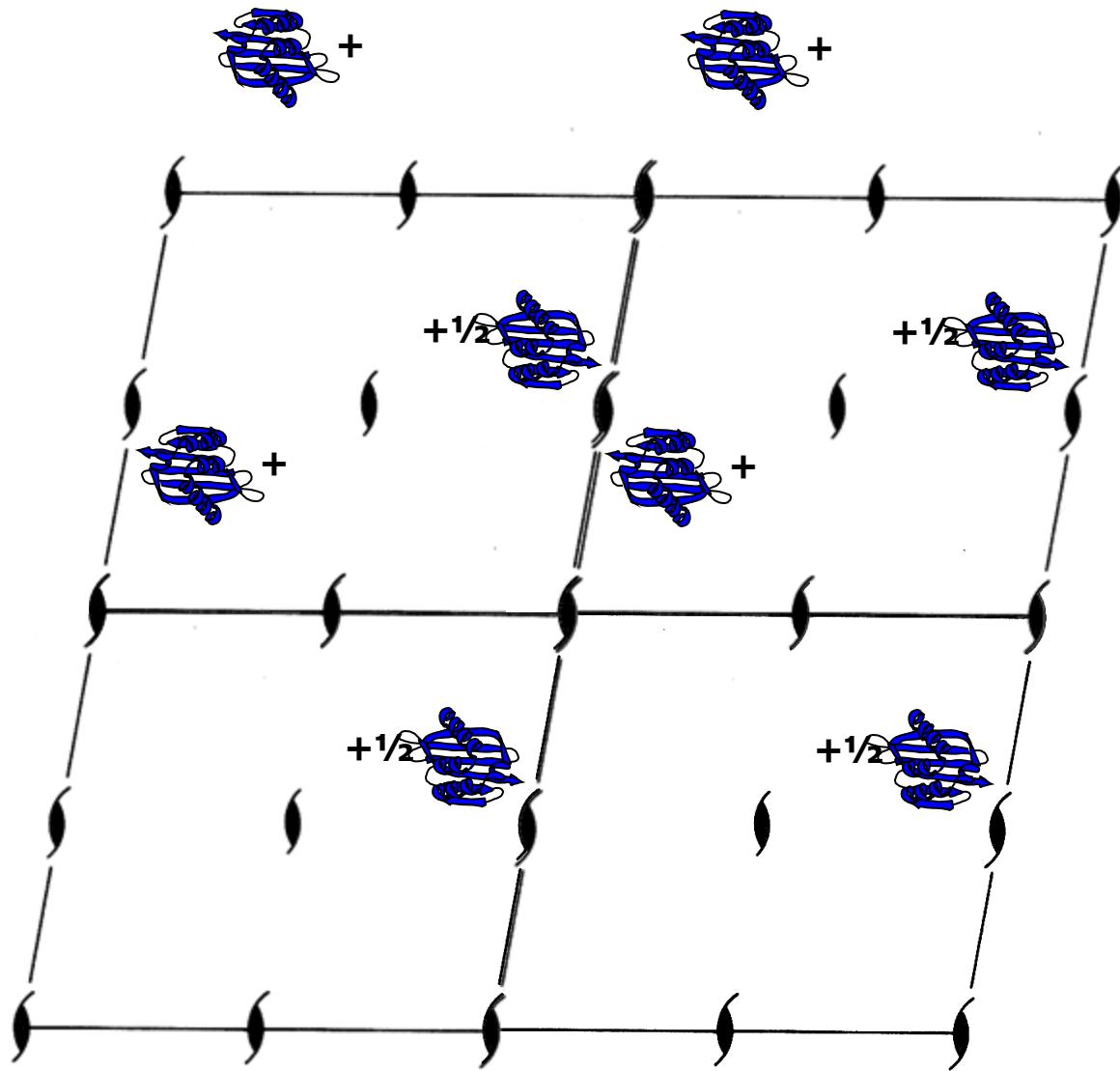
- Origin anchored at symmetry operations
- Symmetry operations  $(x,y,z)$ ,  $(-x,y+\frac{1}{2},-z)$





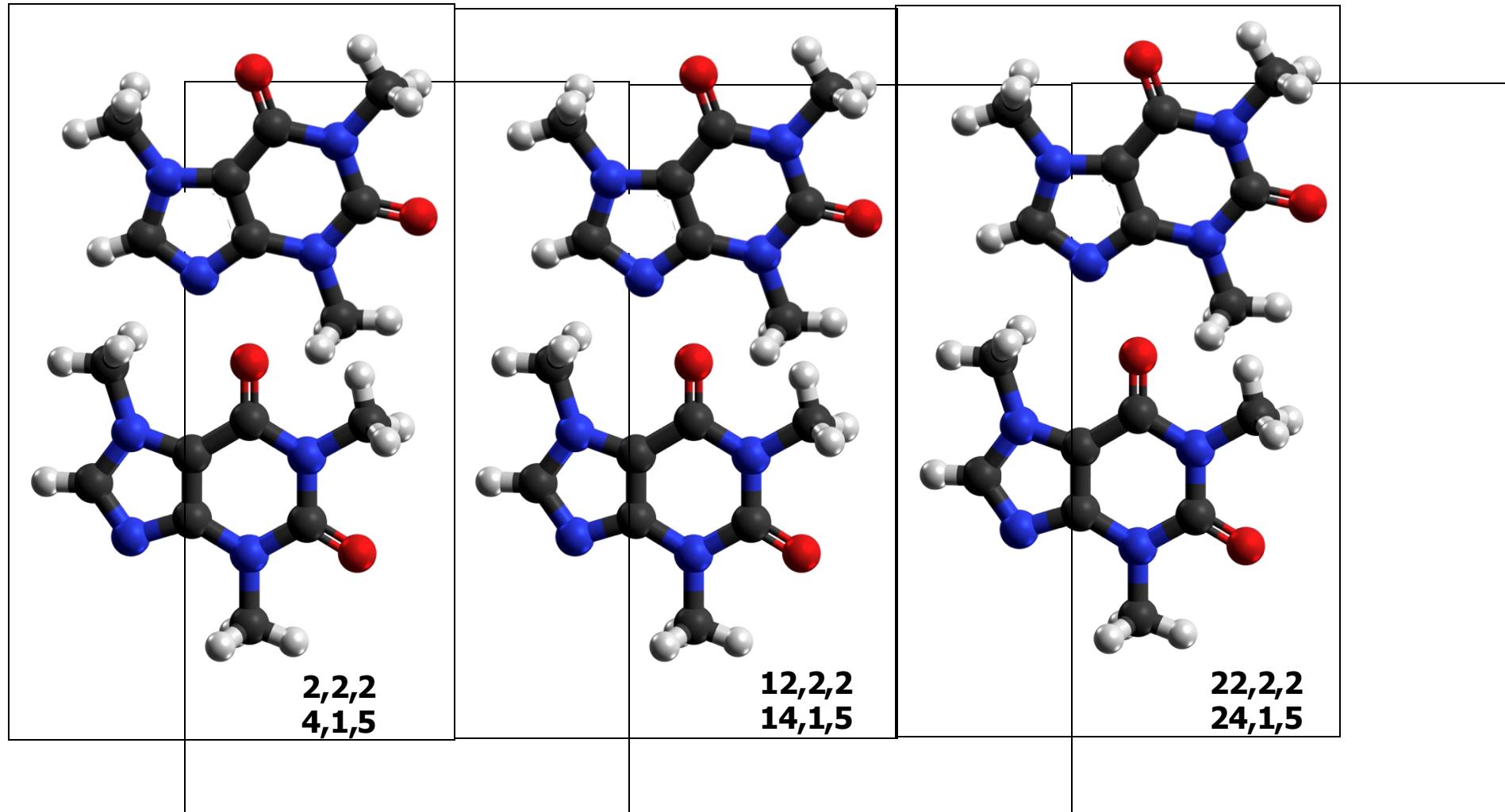






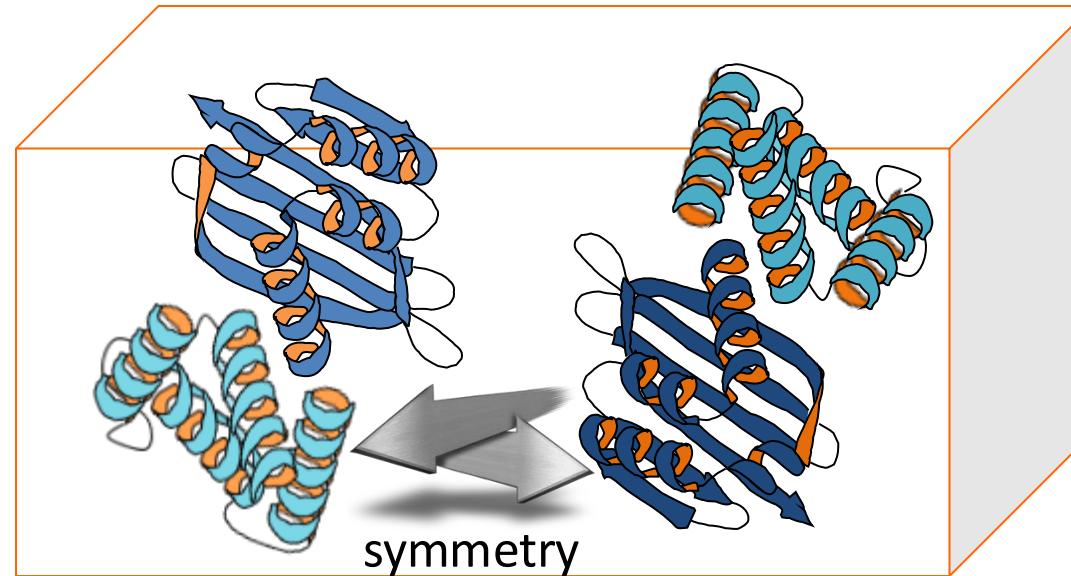
## two molecules or models...

- First molecule 'defines origin' even if space group is P1



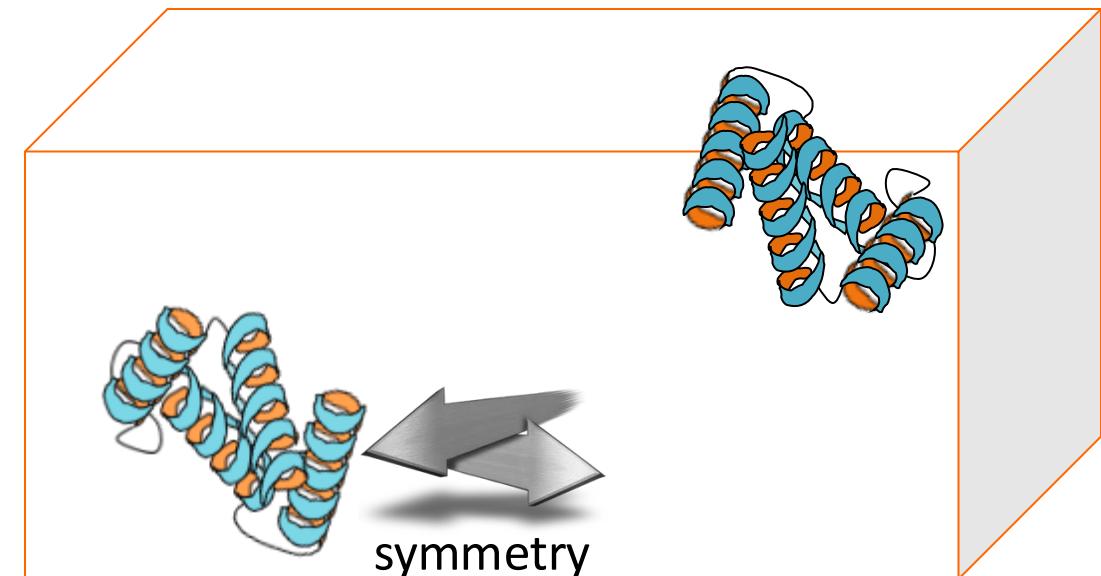
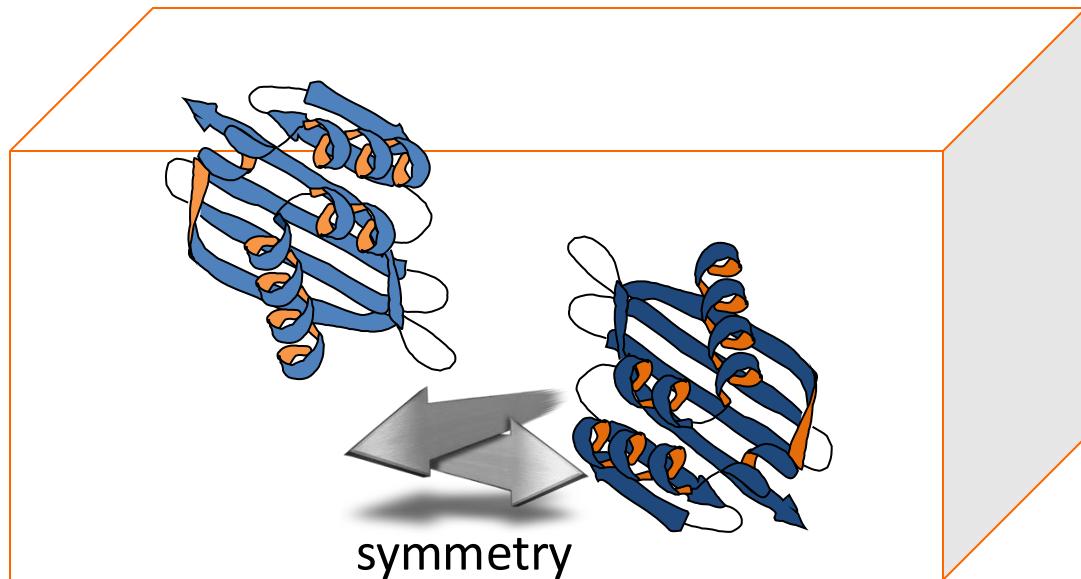
# molecular replacement

- Protein complex



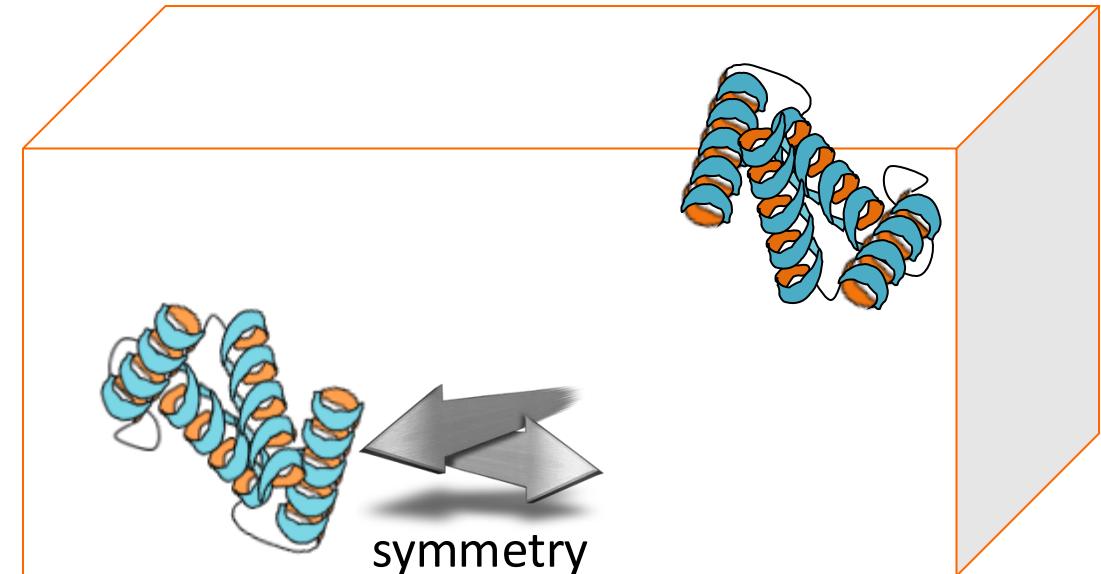
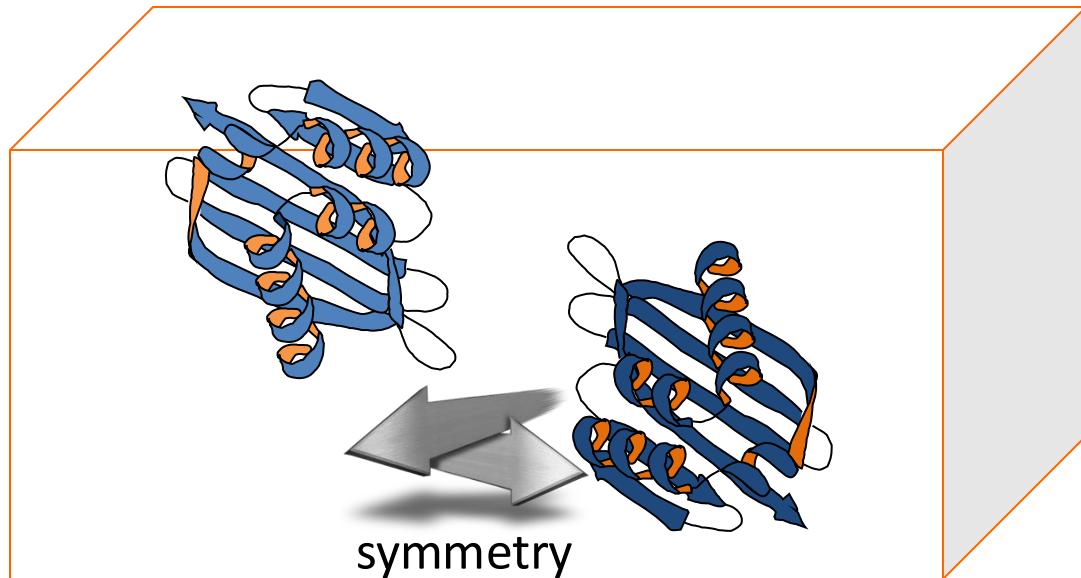
# molecular replacement

- Protein complex
- Could search for the two components separately
  - and merge solutions afterwards



# molecular replacement

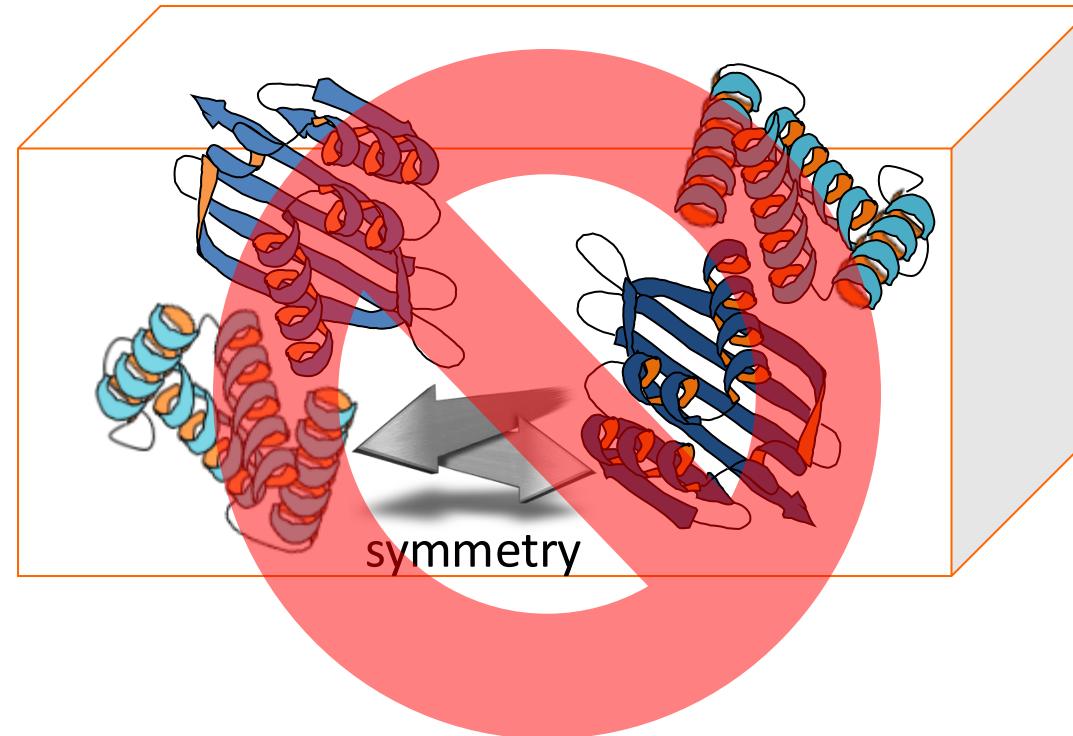
- Protein complex
- Separate solutions could be on different origins



# molecular replacement

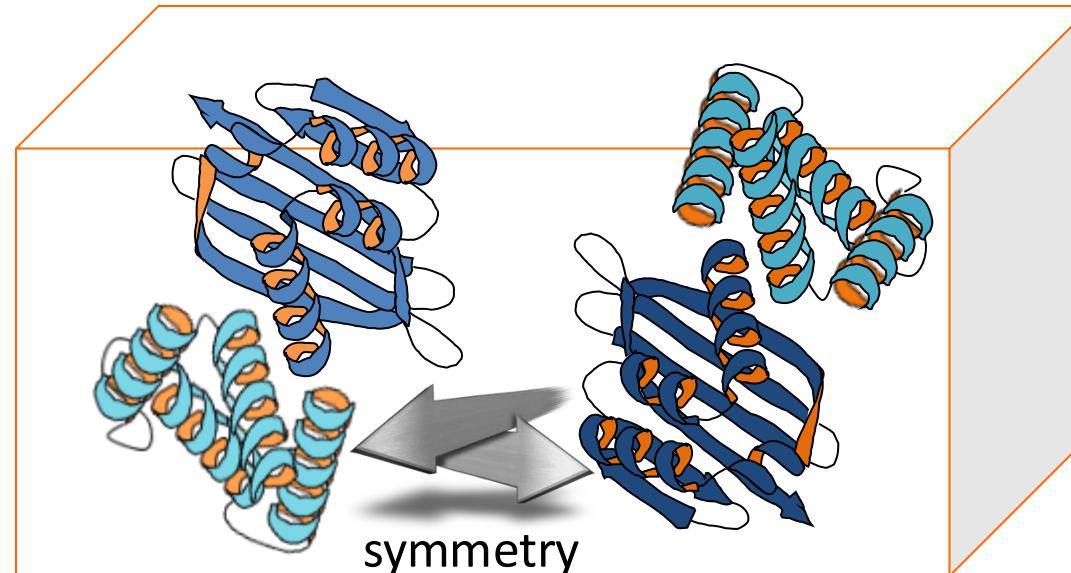
- Protein complex

- Separate solutions could be on different origins
- Signal may be low for one or more components

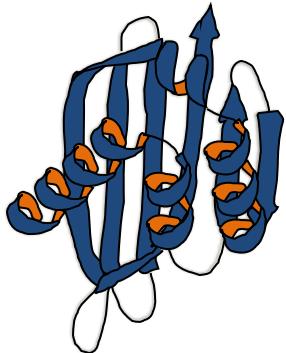


# molecular replacement

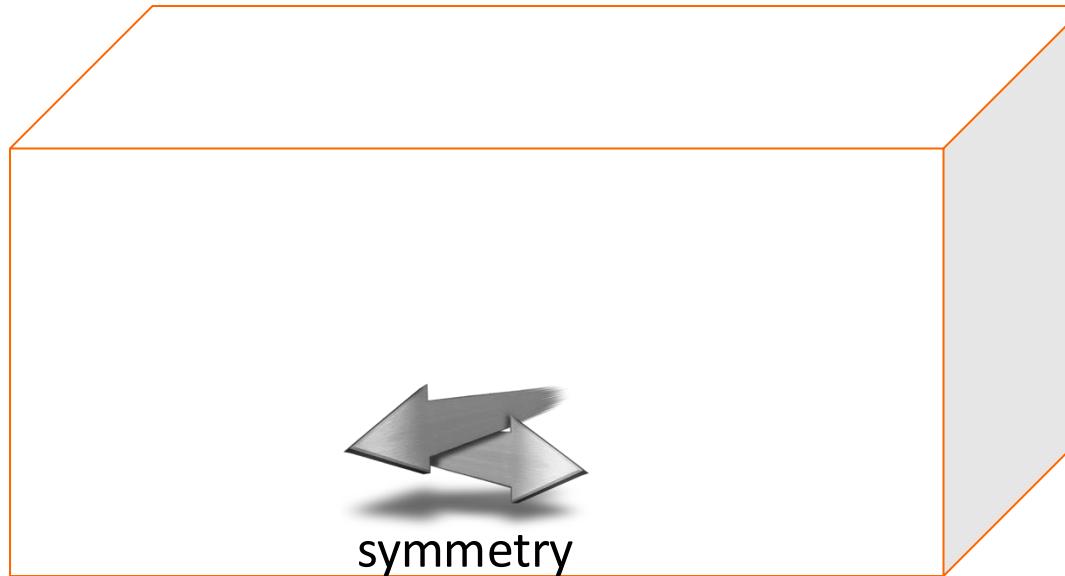
- Protein complex
- Complex built by addition



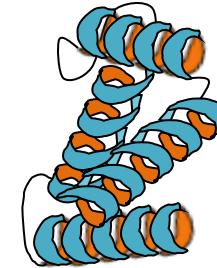
# molecular replacement



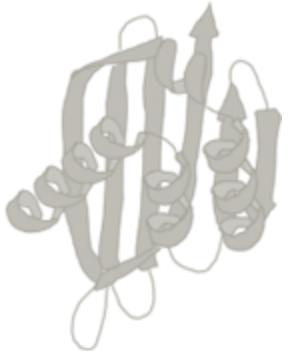
Structure built by  
addition placing  
second in the  
presence of the first



Biological  
oligomer  
placement

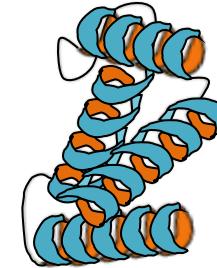
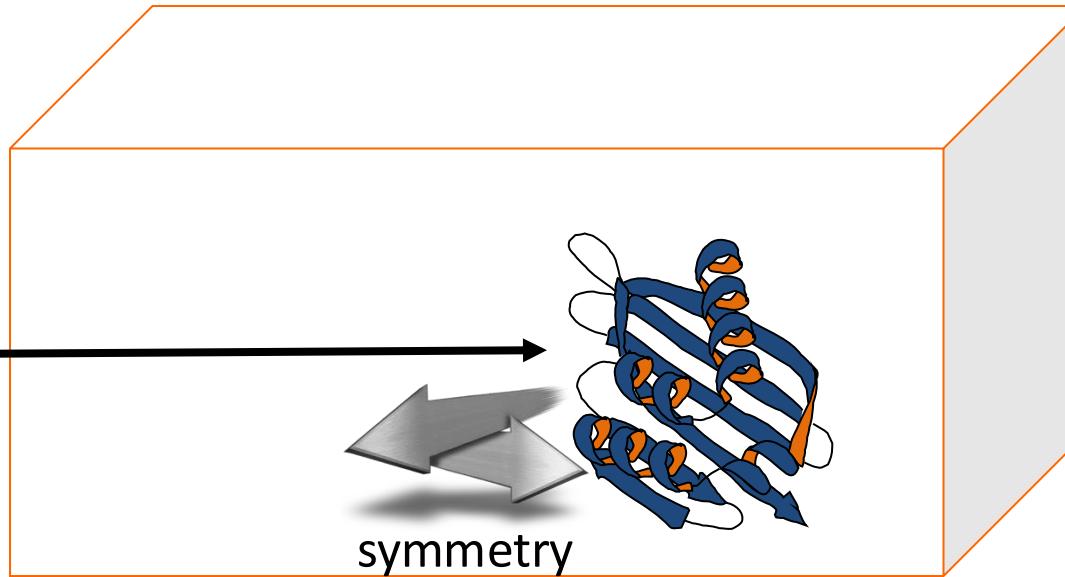
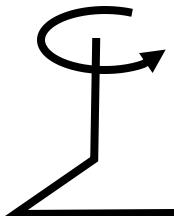


# molecular replacement



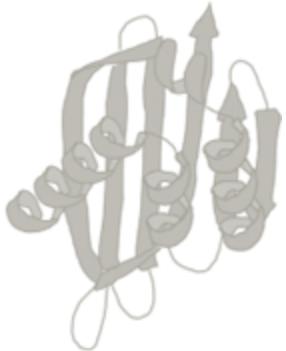
Rotation  
 $(\alpha_1, \beta_1, \gamma_1)$   
Translation  
 $(x_1, y_1, z_1)$

Structure built by  
addition placing  
second in the  
presence of the first

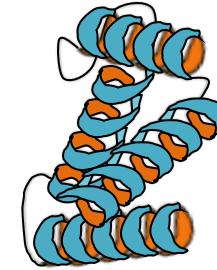
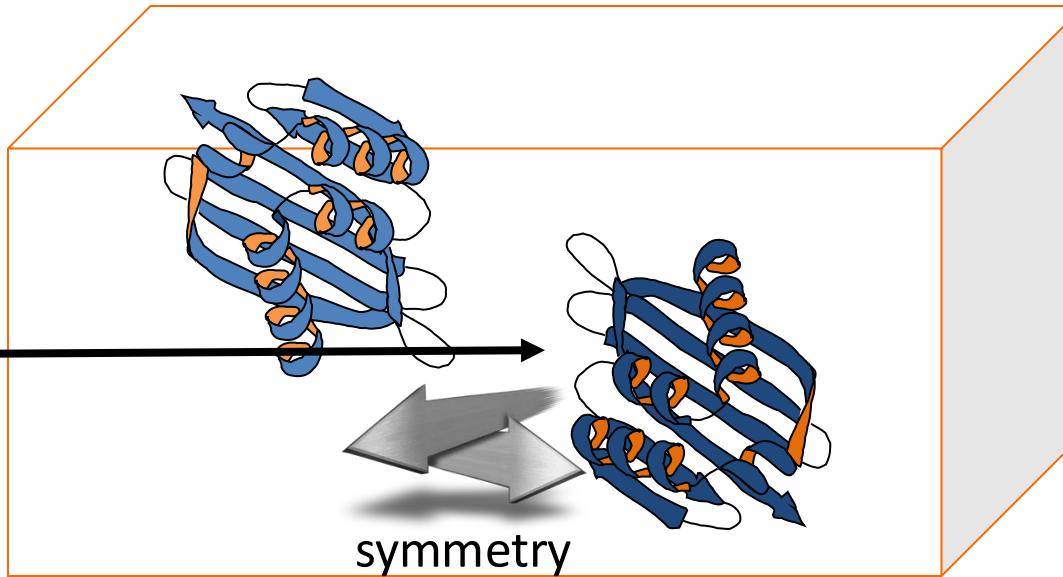
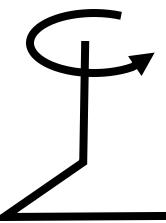


Biological  
oligomer  
placement

# molecular replacement

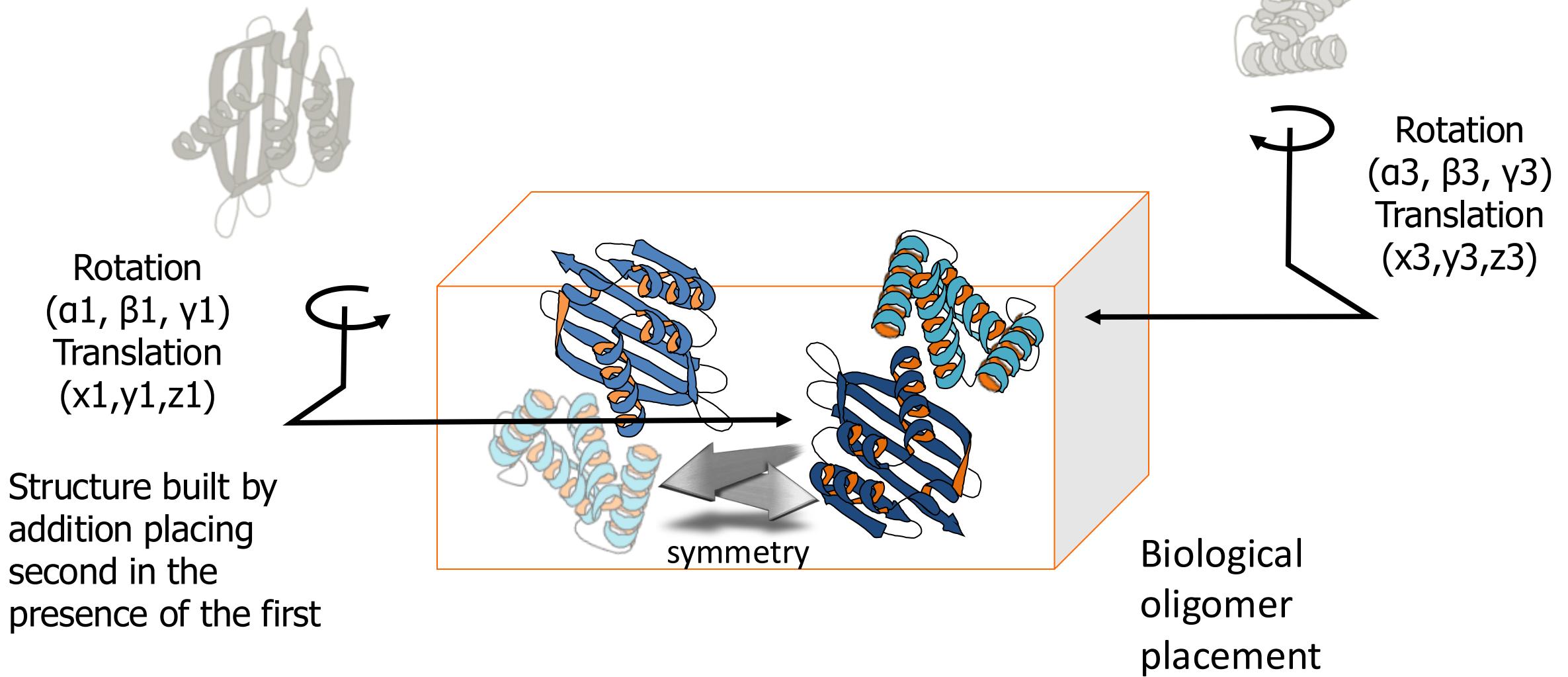


Rotation  
 $(\alpha_1, \beta_1, \gamma_1)$   
Translation  
 $(x_1, y_1, z_1)$

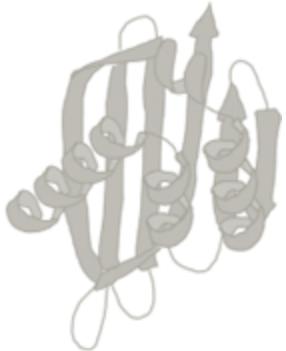


Biological oligomer placement

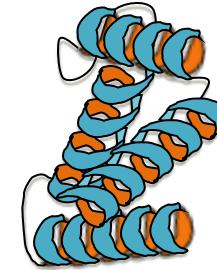
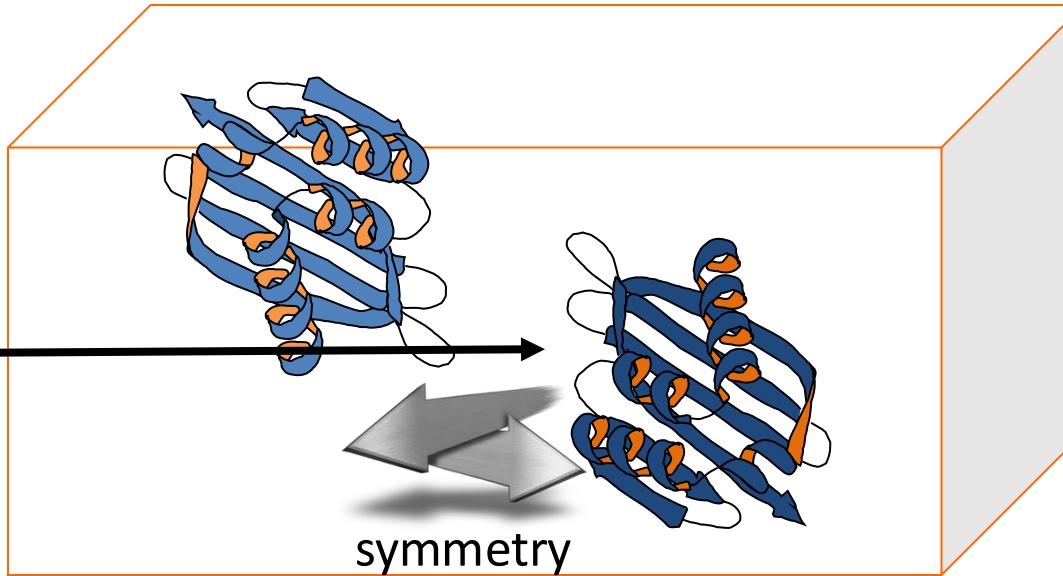
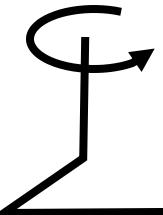
# molecular replacement



# molecular replacement

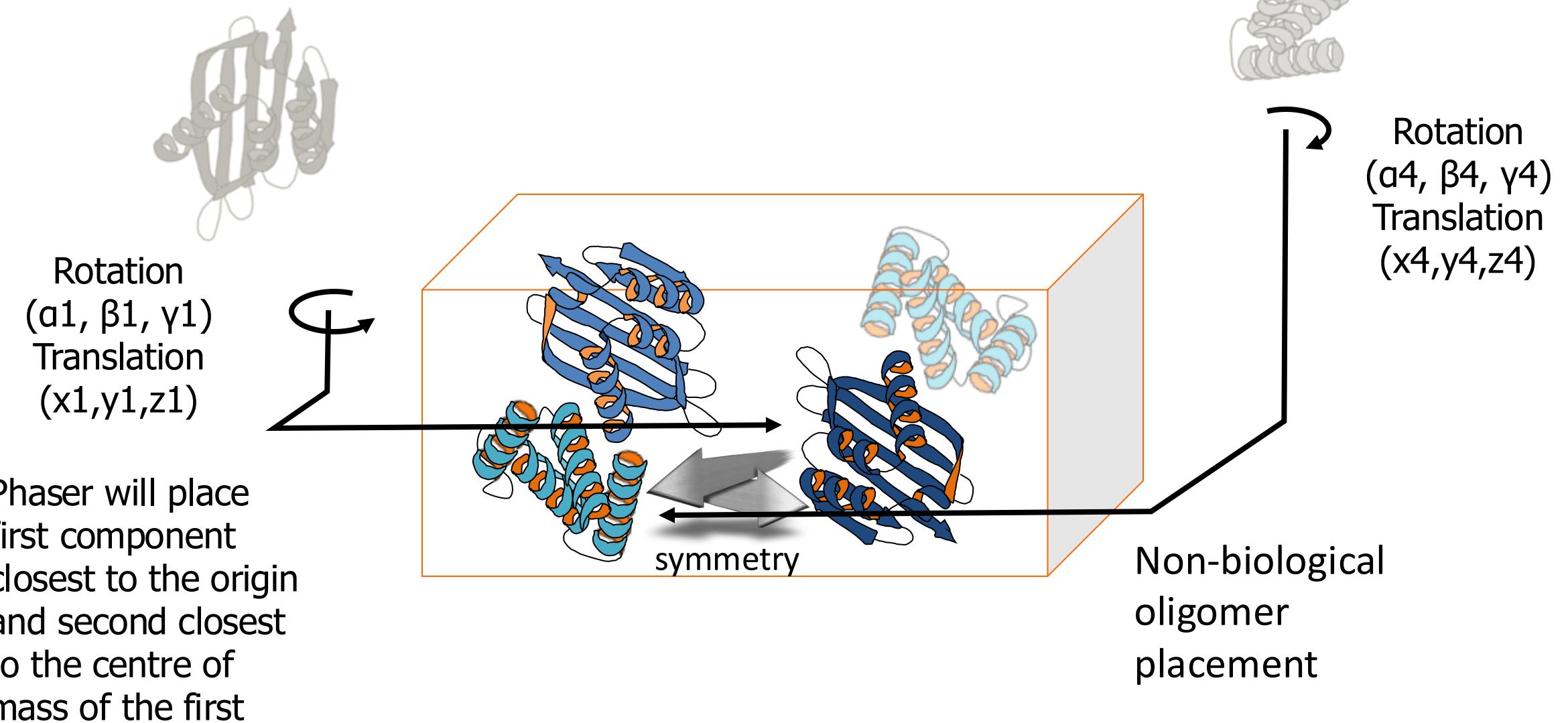


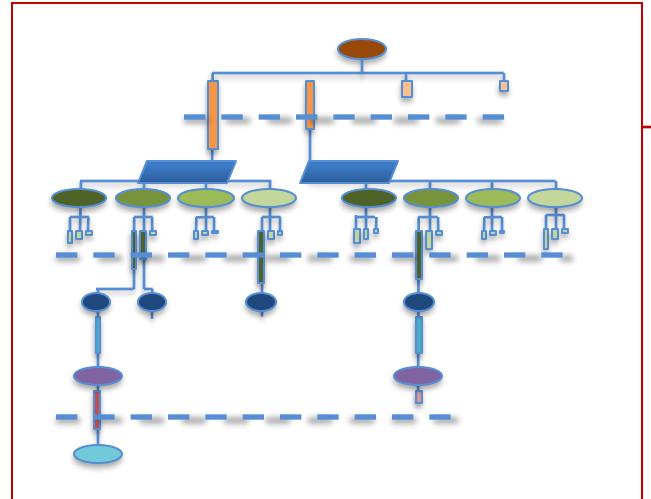
Rotation  
 $(\alpha_1, \beta_1, \gamma_1)$   
Translation  
 $(x_1, y_1, z_1)$



Biological oligomer placement

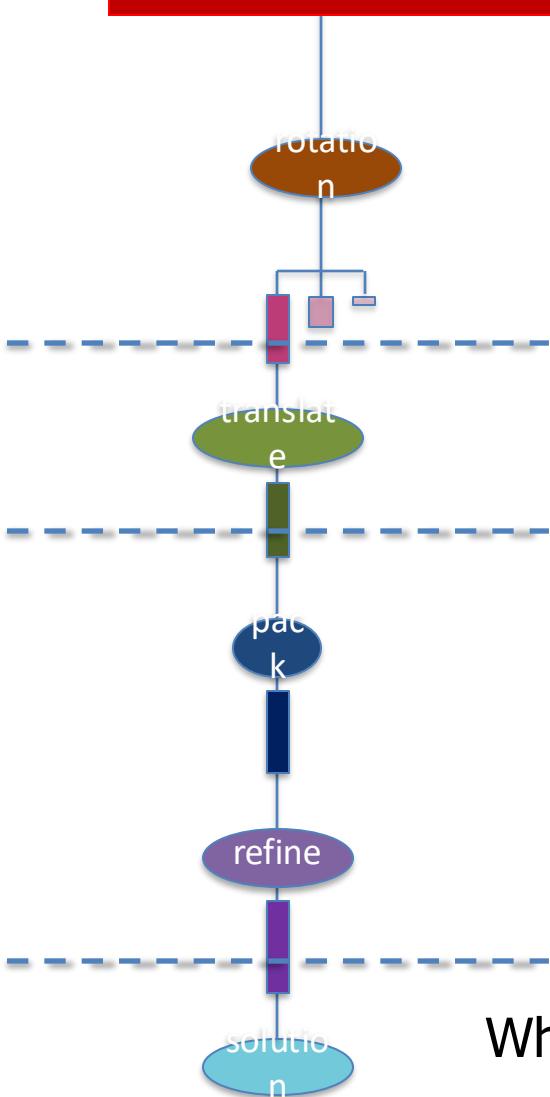
# molecular replacement





First component placed

space group



Why is placing the second so much easier than placing the first?

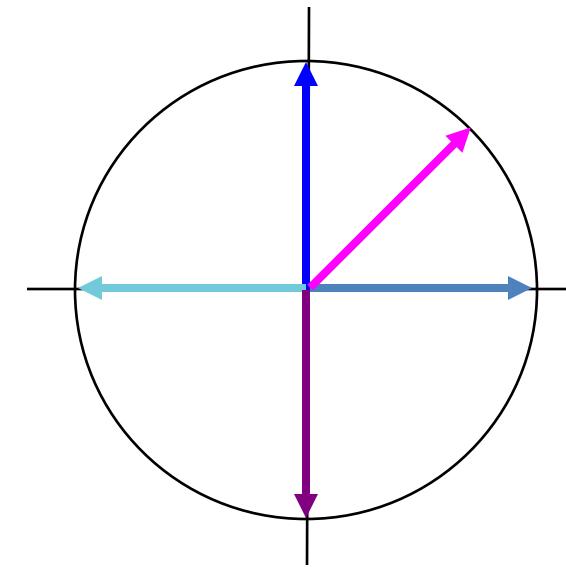
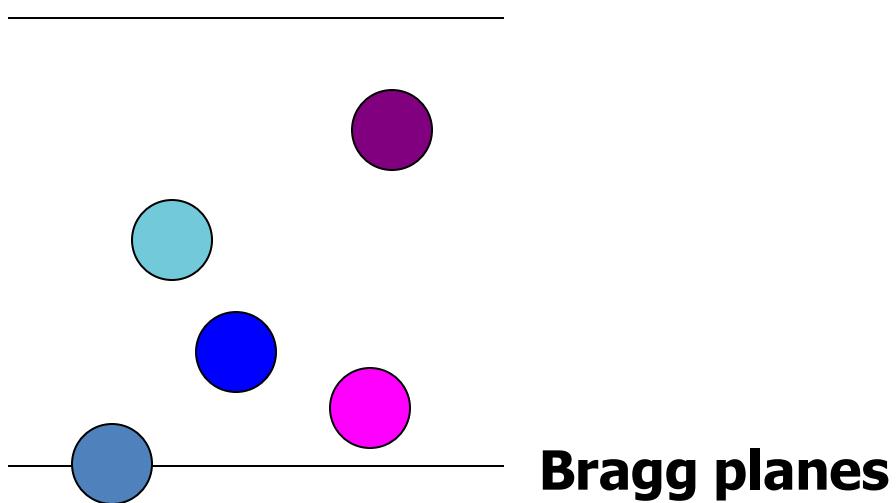
## Phaser – likelihood theory

---

$$LLGI = \sum_h \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$

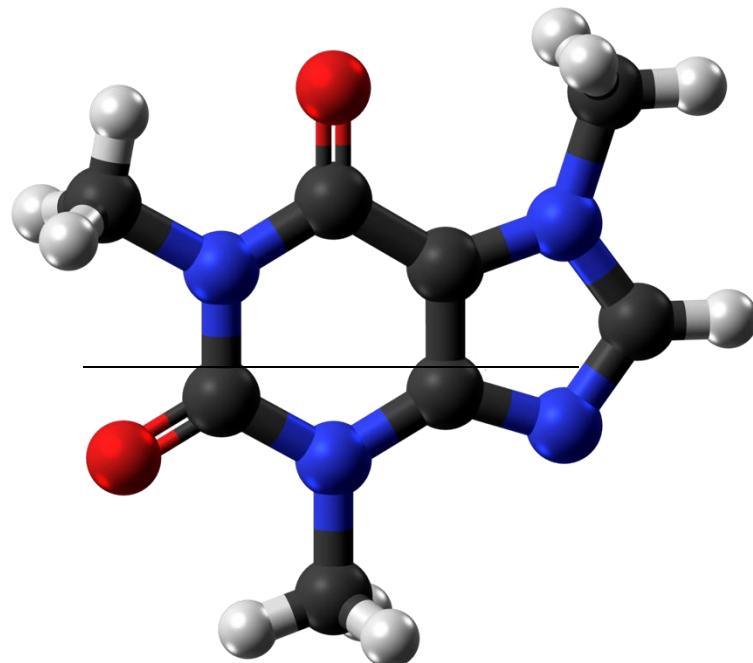
# Atom distribution

- The diffraction wave from an atom can be represented by a vector (structure factor)
- Amplitude depends on the atom-type
- Phase depends on fractional distance between Bragg planes

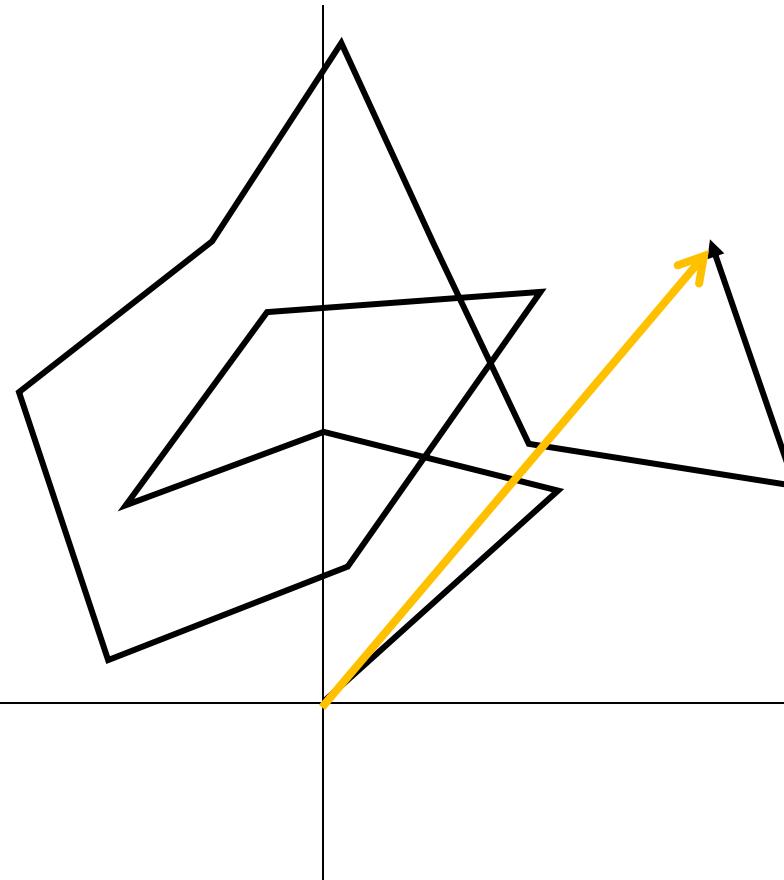


# coffee

- summing structure factors



total structure factor  
sum of atomic structure factors

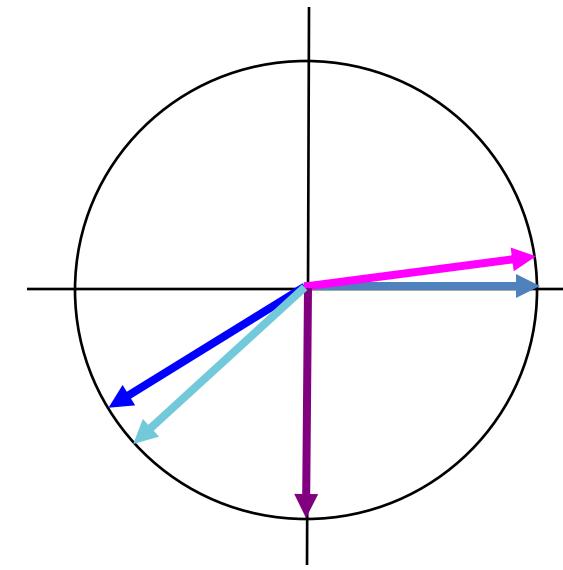
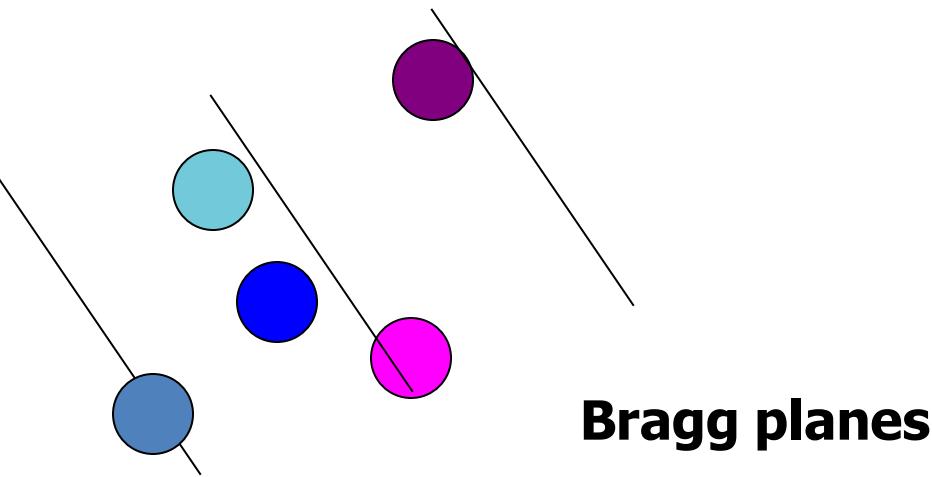


---

Bragg planes

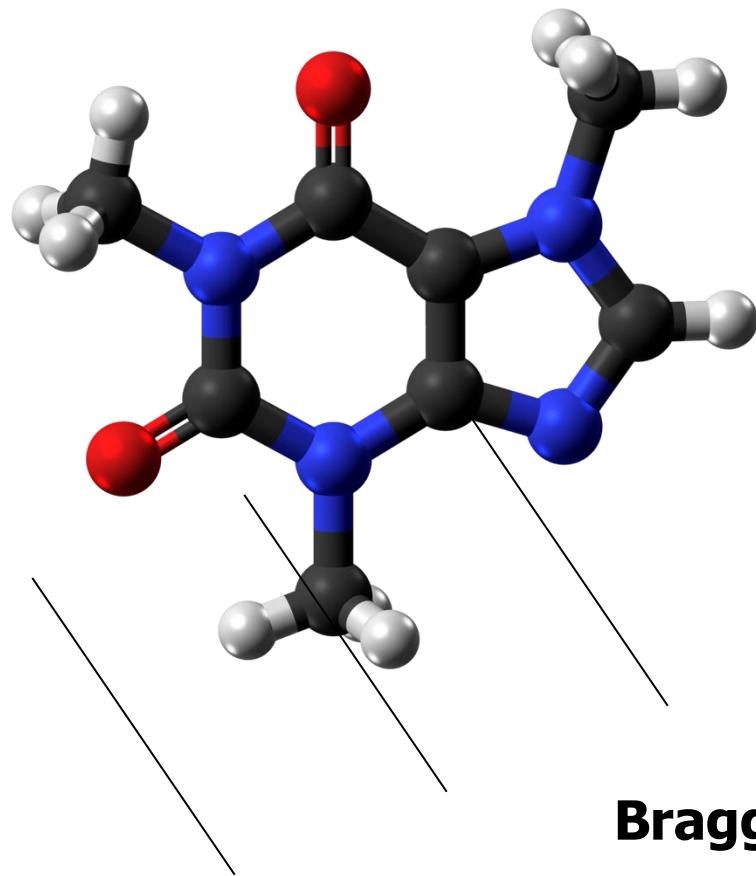
# Atom distribution

- A different Bragg plane will give different phases for the atoms
- Amplitudes of atomic structure factors will be the same (same atomtype)



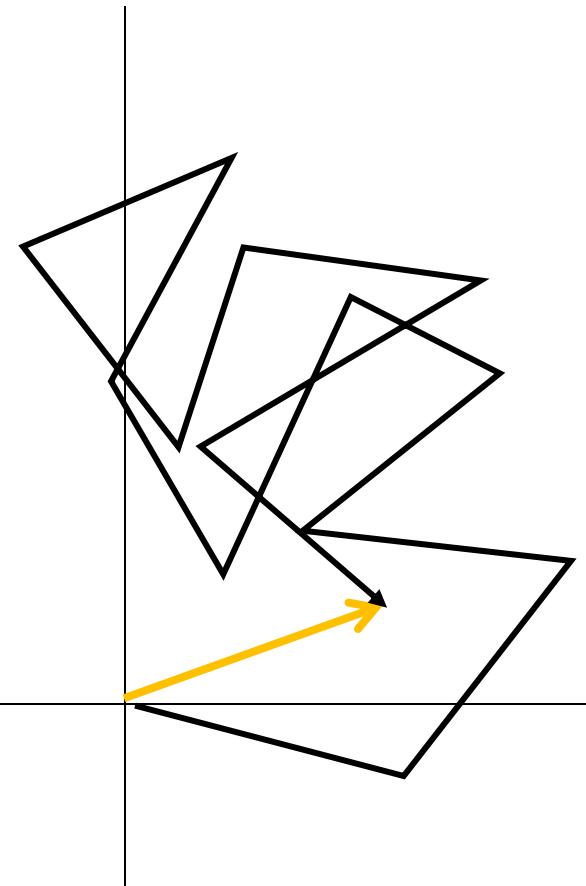
# coffee

- summing structure factors



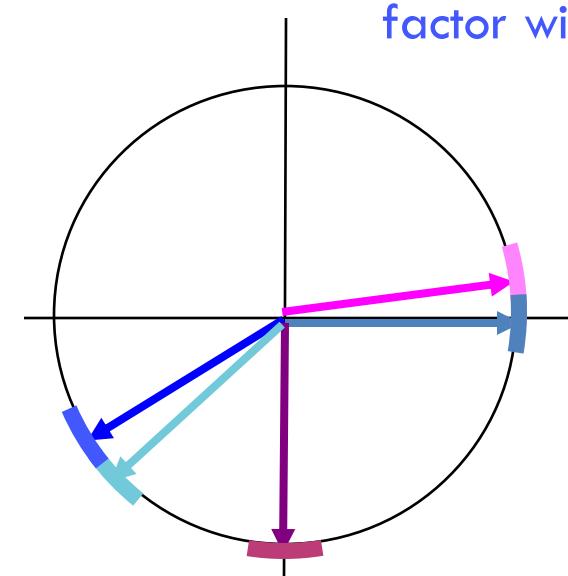
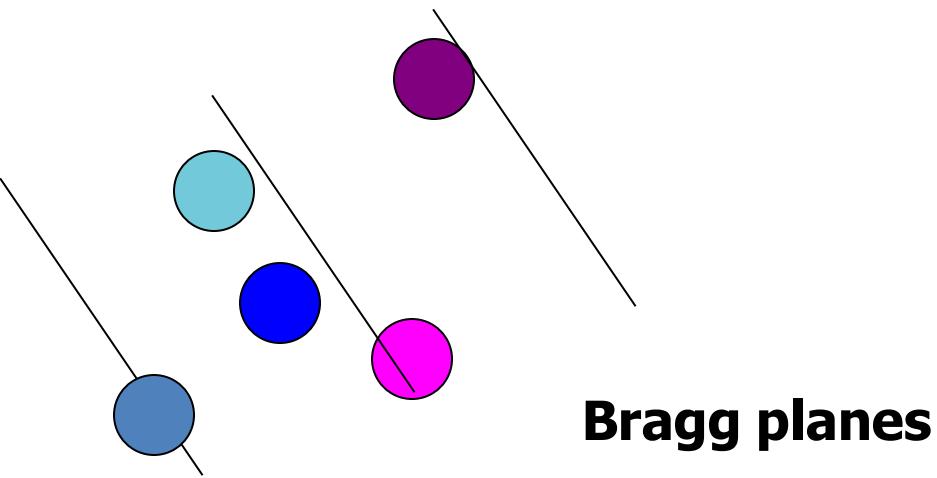
Bragg planes

total structure factor  
sum of atomic structure factors

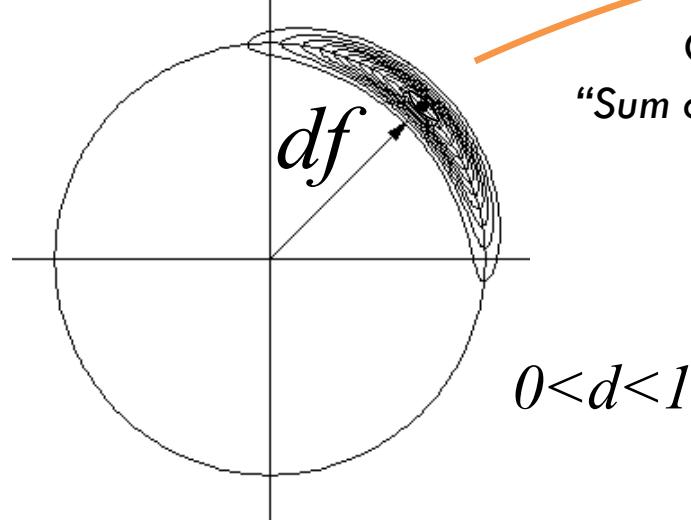


# Summing structure factors with errors

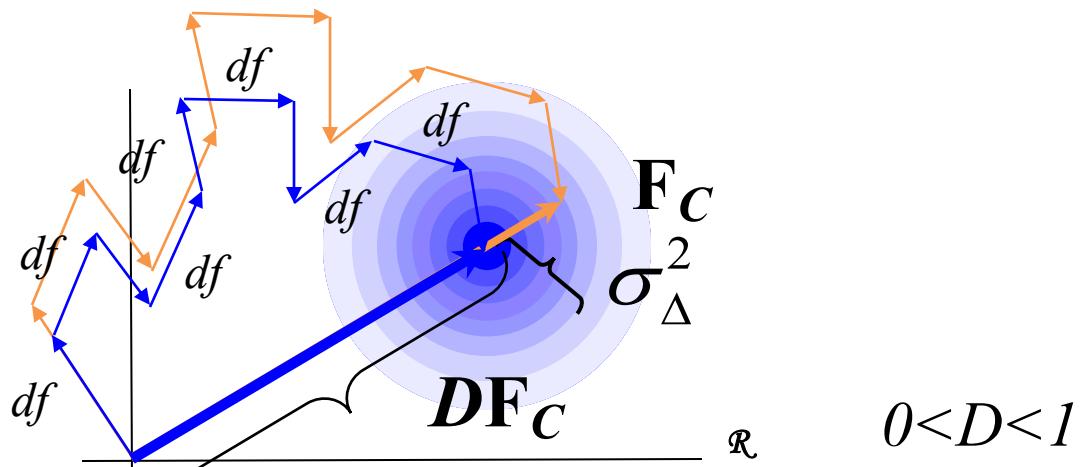
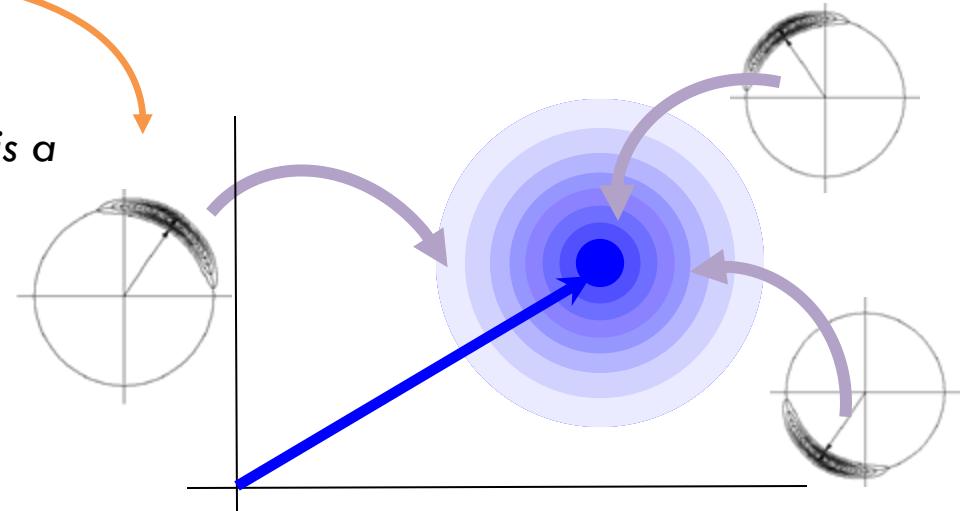
- The model has errors
- The error in the atomic position manifests as a phase error in the structure factors



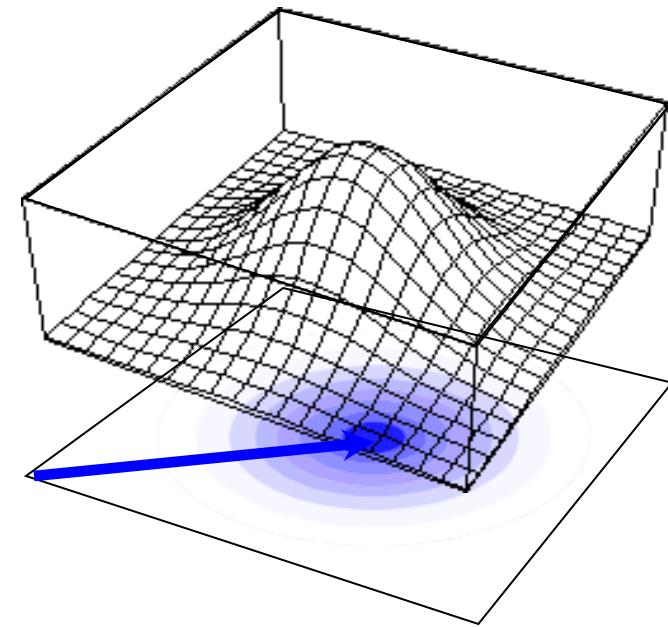
What is the probability distribution for the total structure factor with errors?



Central Limit Theorem:  
“Sum of any error distribution is a Gaussian”



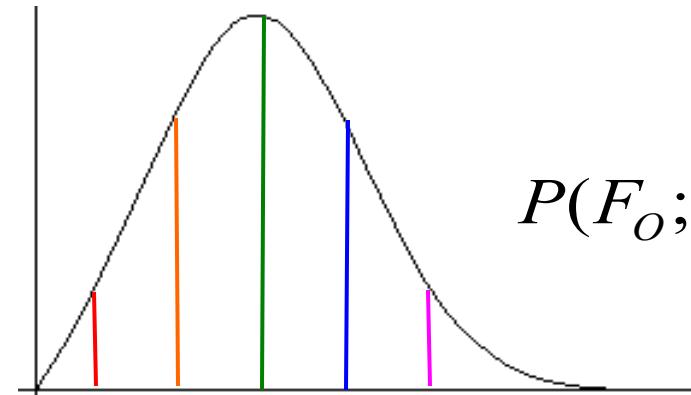
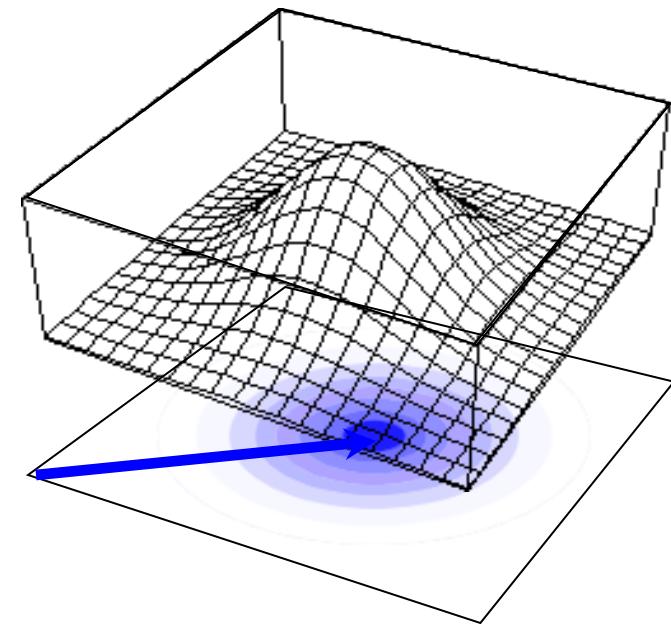
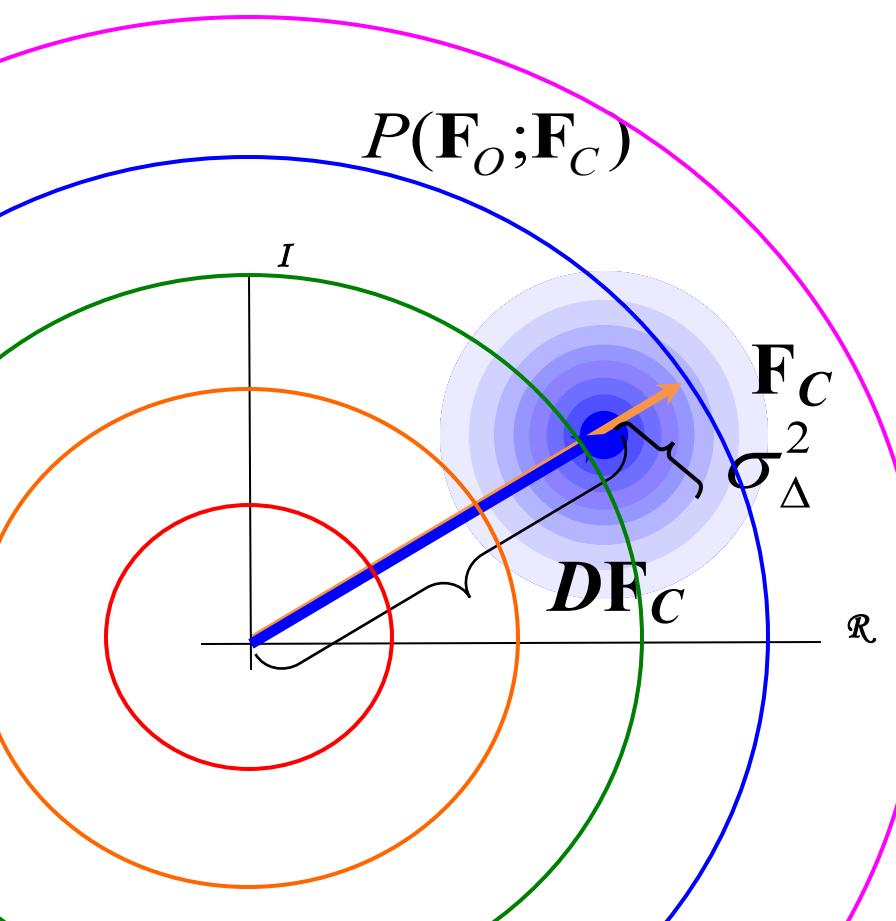
Structure Factor ( $\mathbf{F}, \mathbf{D}, \sigma_{\Delta}$ )



$$a \exp\left(-\frac{|x - b|^2}{c^2}\right)$$

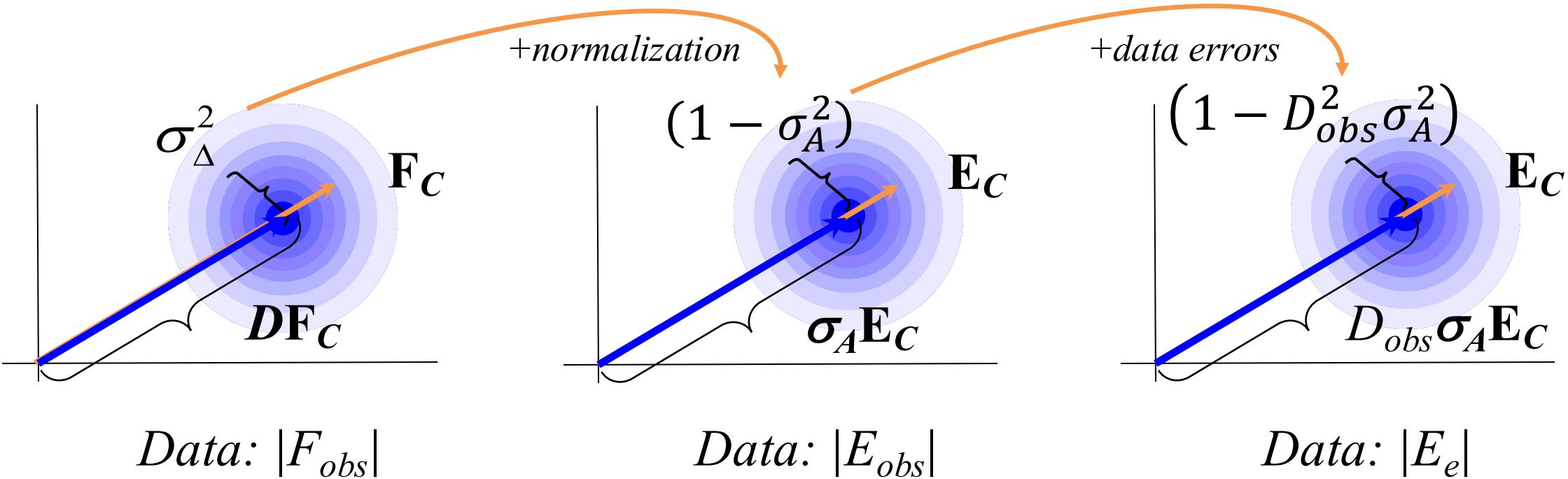
functional form for  
2D Gaussian

$$P(\mathbf{F}_o; \mathbf{F}_C) = \frac{1}{\pi\sigma_{\Delta}^2} \exp\left(-\frac{|\mathbf{F}_o - D\mathbf{F}_C|^2}{\sigma_{\Delta}^2}\right)$$



$$P(F_o; F_C) = \frac{2F_o}{\sigma_{\Delta}^2} \exp\left(-\frac{F_o^2 + D^2 F_C^2}{\sigma_{\Delta}^2}\right) I_0\left(\frac{2F_o D F_C}{\sigma_{\Delta}^2}\right)$$

## 2D Gaussians



*Advanced information:  $D_{obs}$  and  $E_e$  come from an approximation to a log-likelihood gain based on intensities (which does not have an analytical solution) cast in terms of a function that implies complex errors. The values are determined once, before phasing.*

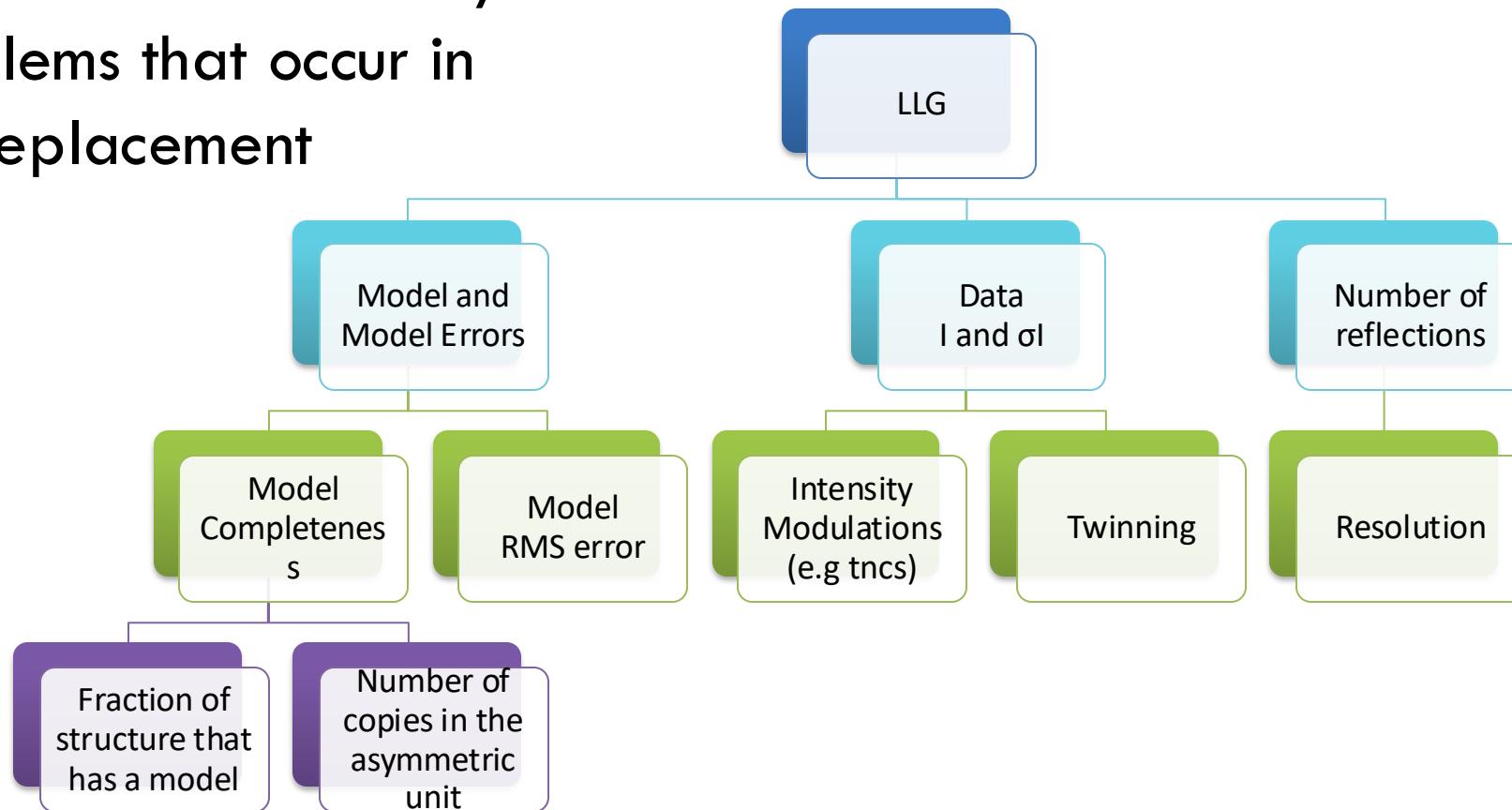
## Phaser – likelihood theory

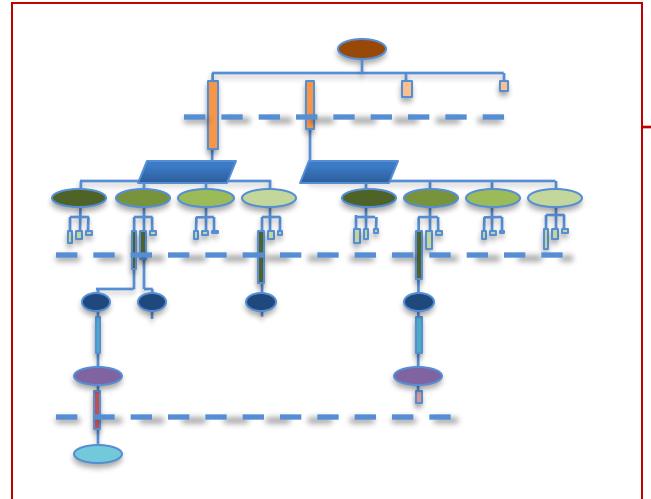
---

$$LLGI = \sum_h \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$

$$LLGI = \sum_{\mathbf{h}} \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$

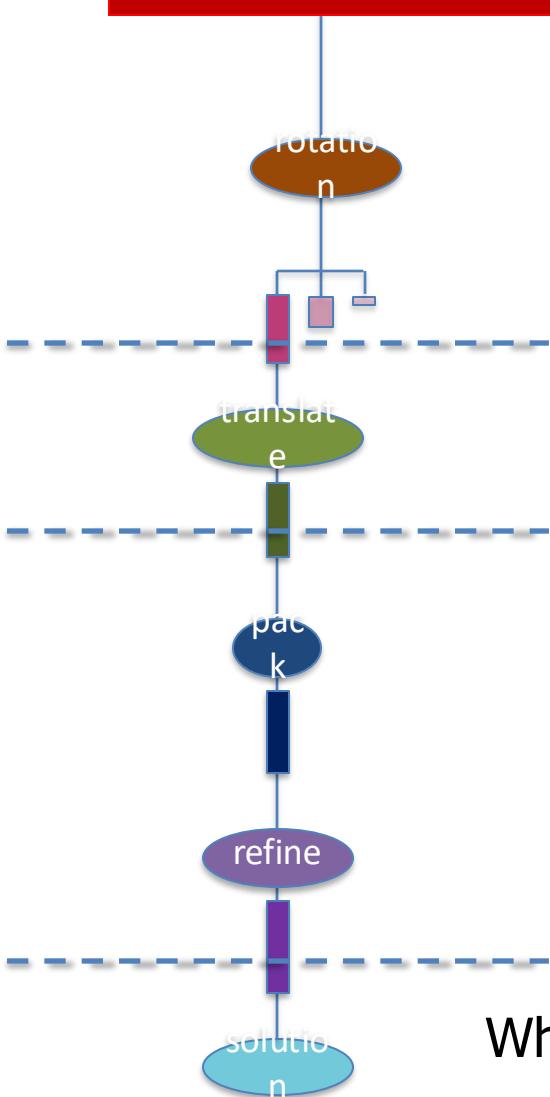
Phaser's target function  
allows us to account for many  
of the problems that occur in  
molecular replacement



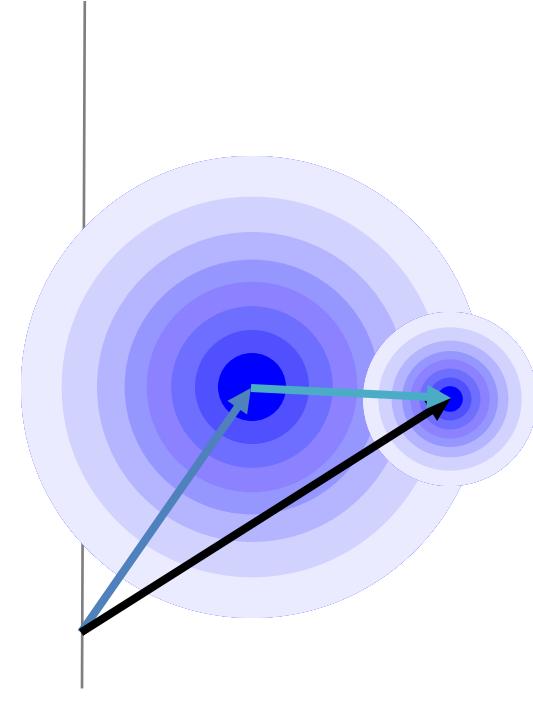
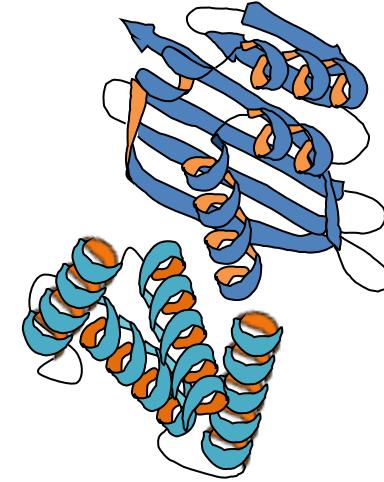
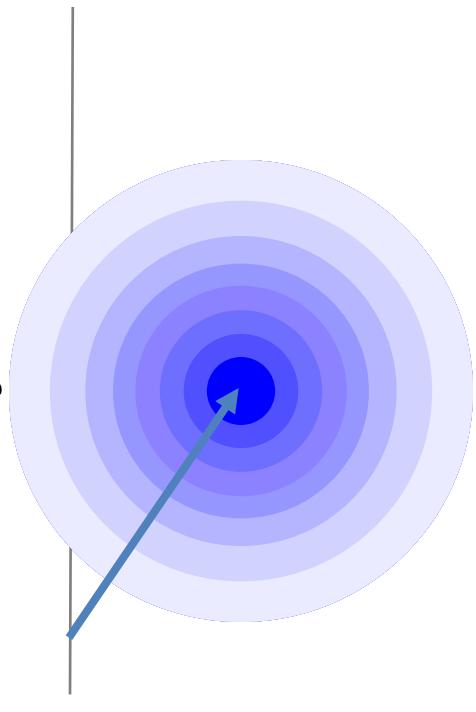
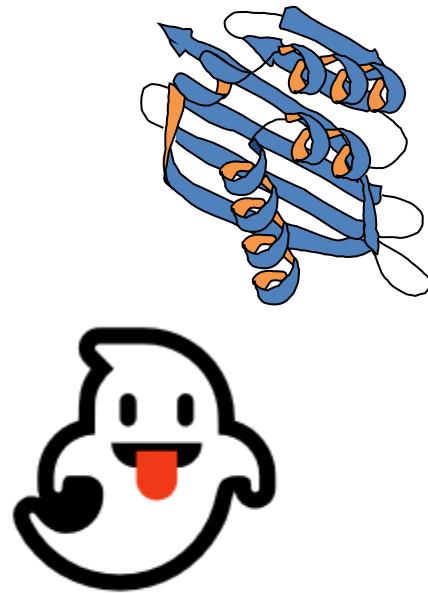


First component placed

space group



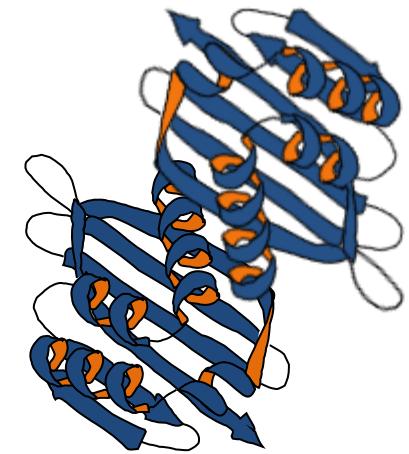
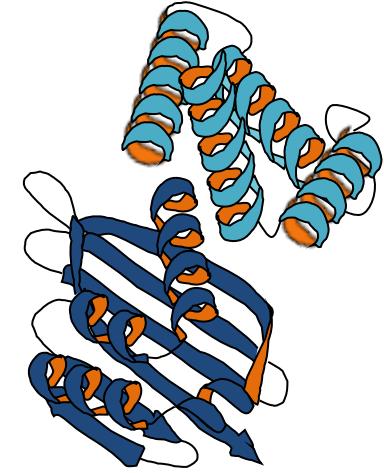
Why is placing the second so much easier than placing the first?

$\sigma_A$ 

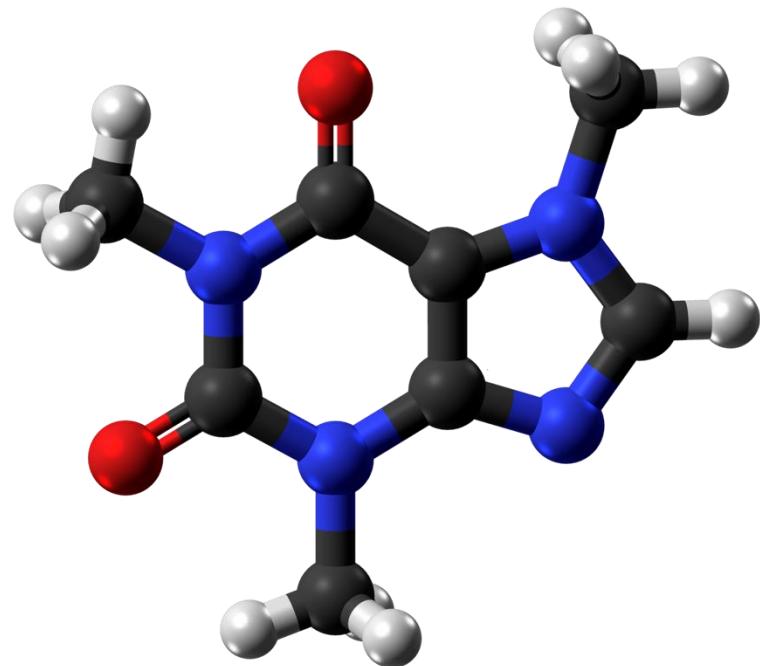
Whenever a component is placed, the  $\sigma_A$  is increased, thus increasing the signal-to-noise of the search for the next component

# asymmetric unit contents

- The problem in molecular replacement is to place all the models in the asymmetric unit
- Equivalently:
  - Multiple copies of same protein
  - Components of complexes
  - Complexes
  - Fragments of long flexible proteins
  - Fragments of badly modelled proteins
  - Non-protein components
- The more components the harder the problem

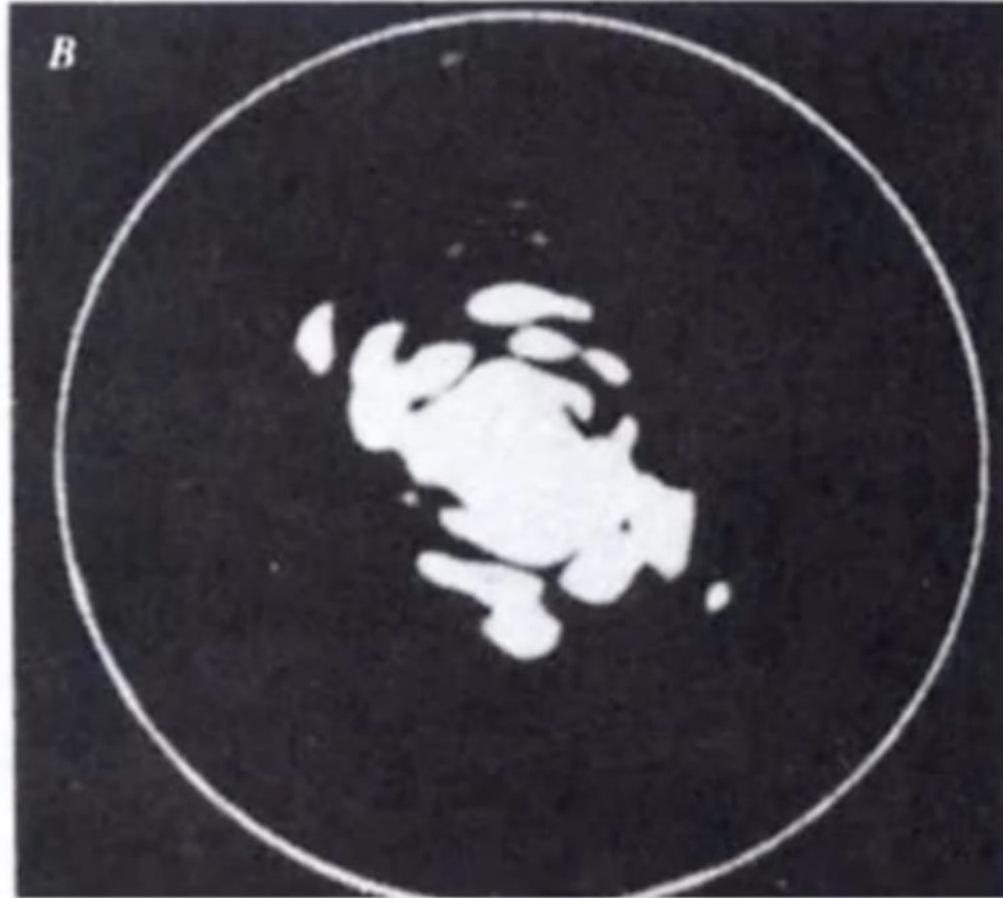
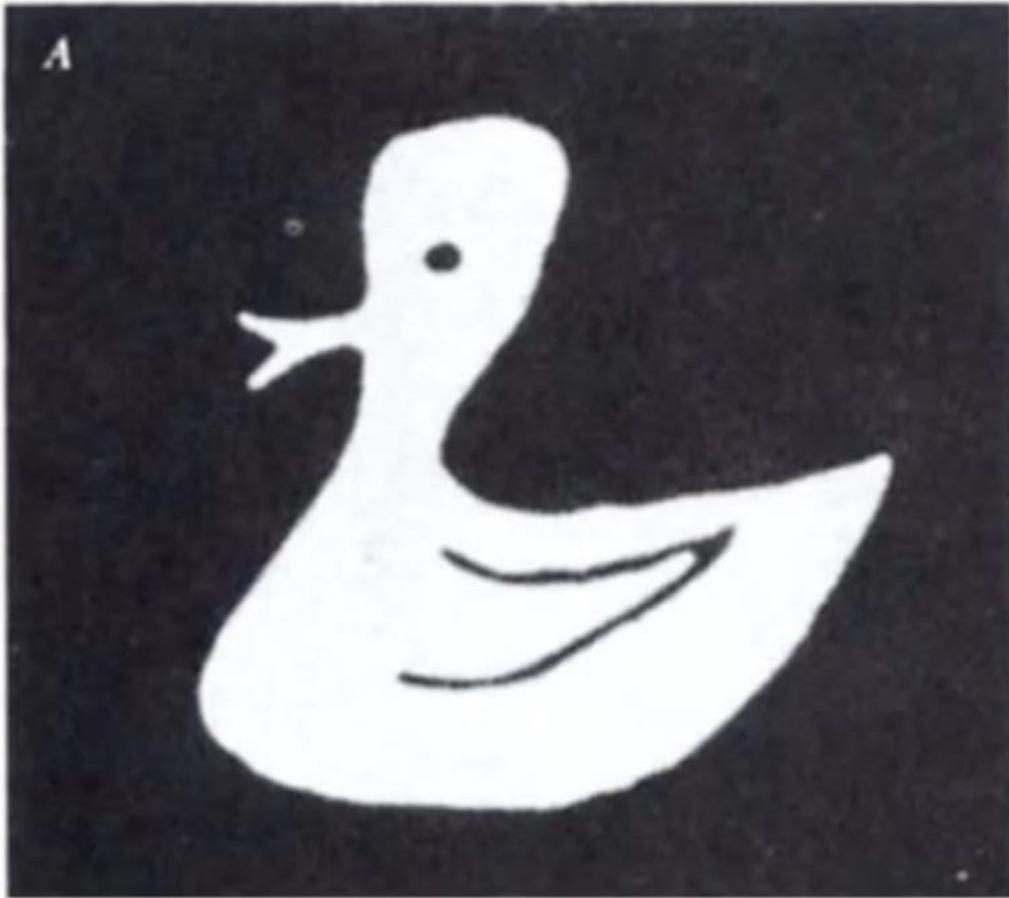


coffee



**ducks**

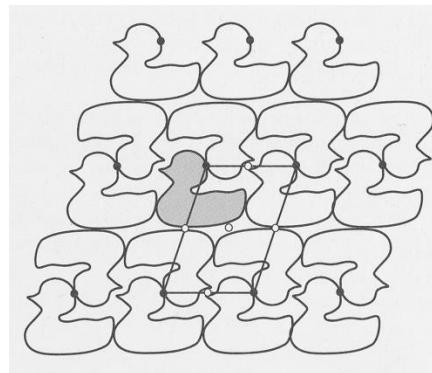
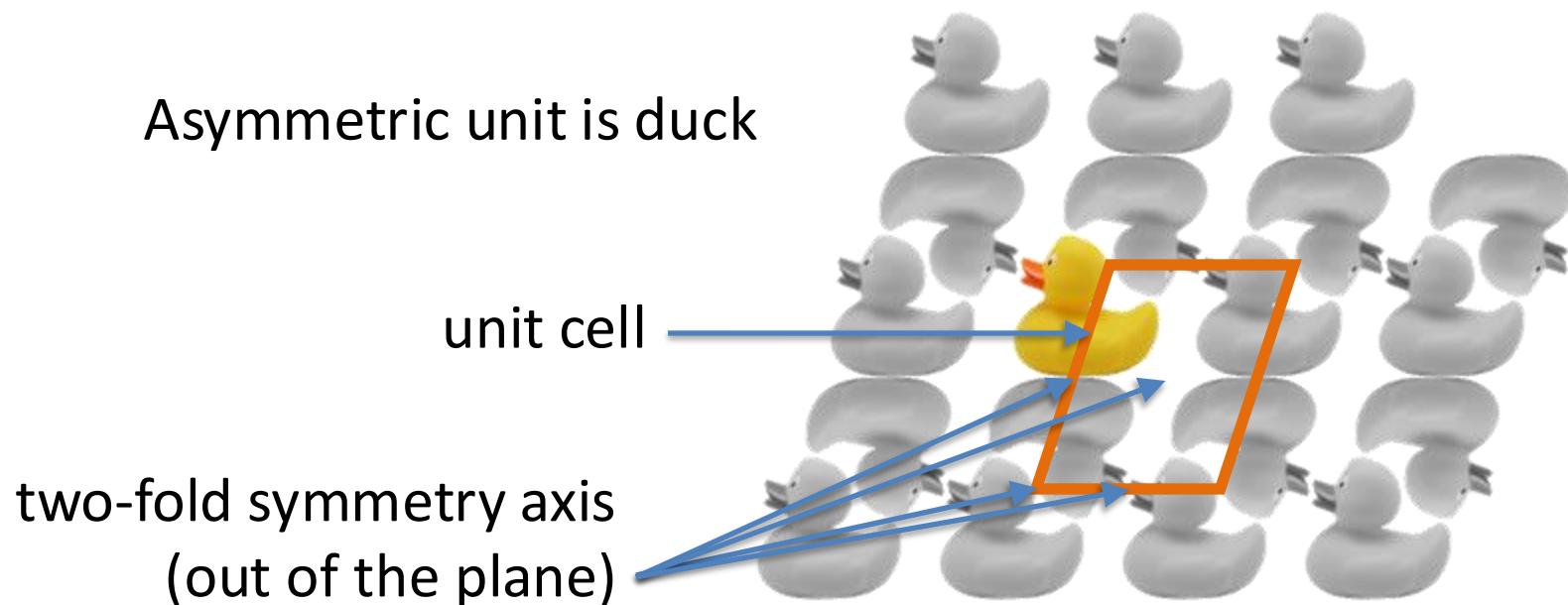
# ducks in crystallography



**Blundell and Johnson 1976**

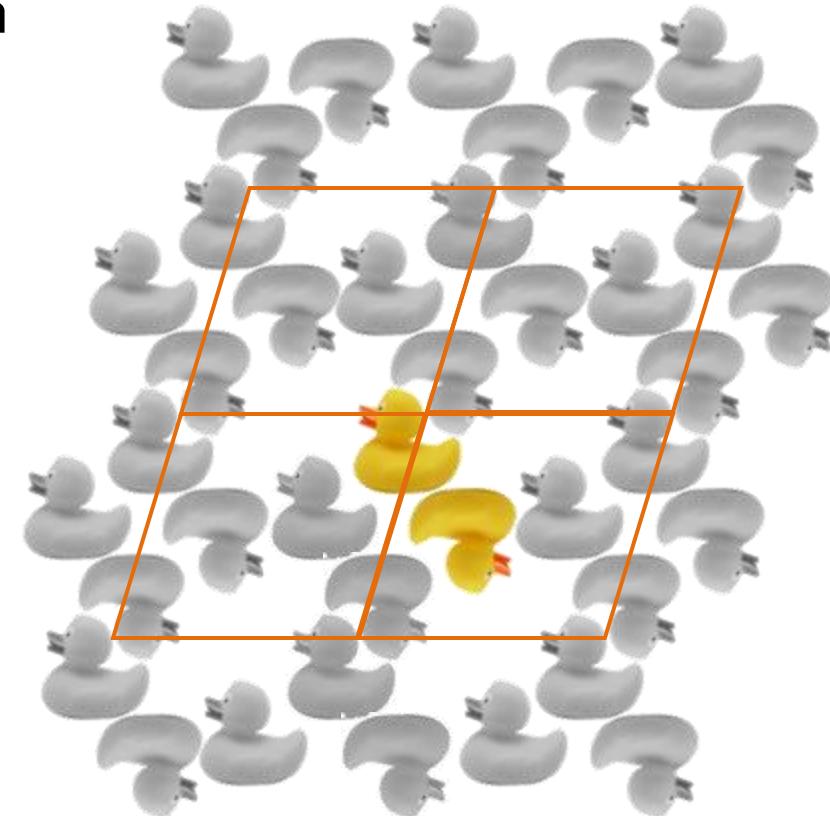
# asymmetric unit

- The asymmetric unit is the smallest unit of structure that can generate the whole crystal after application of the crystal symmetry

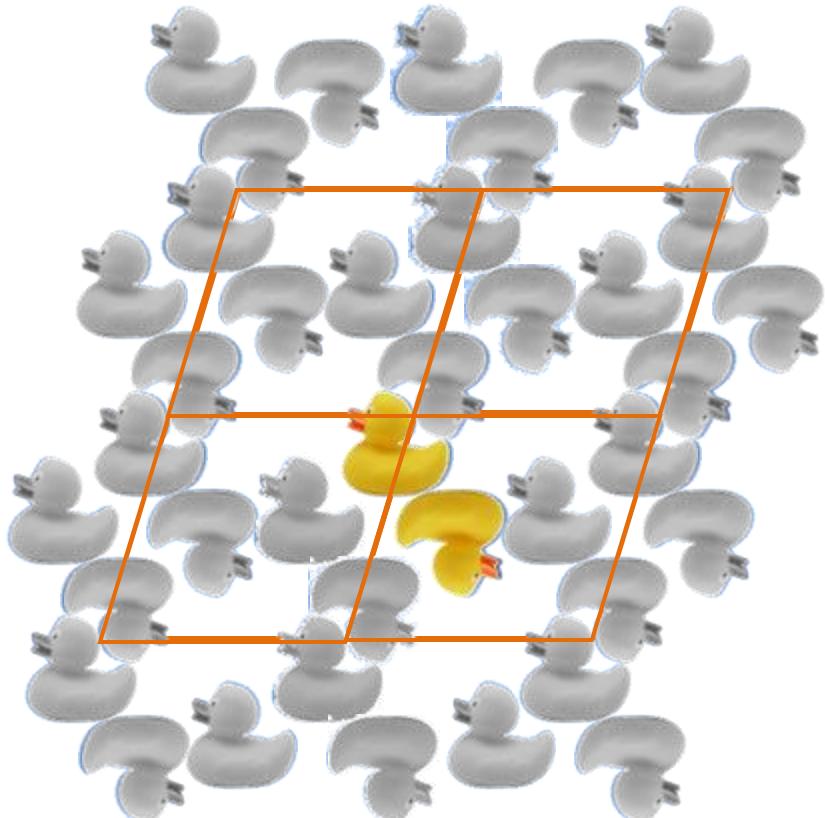


## asymmetric unit contents

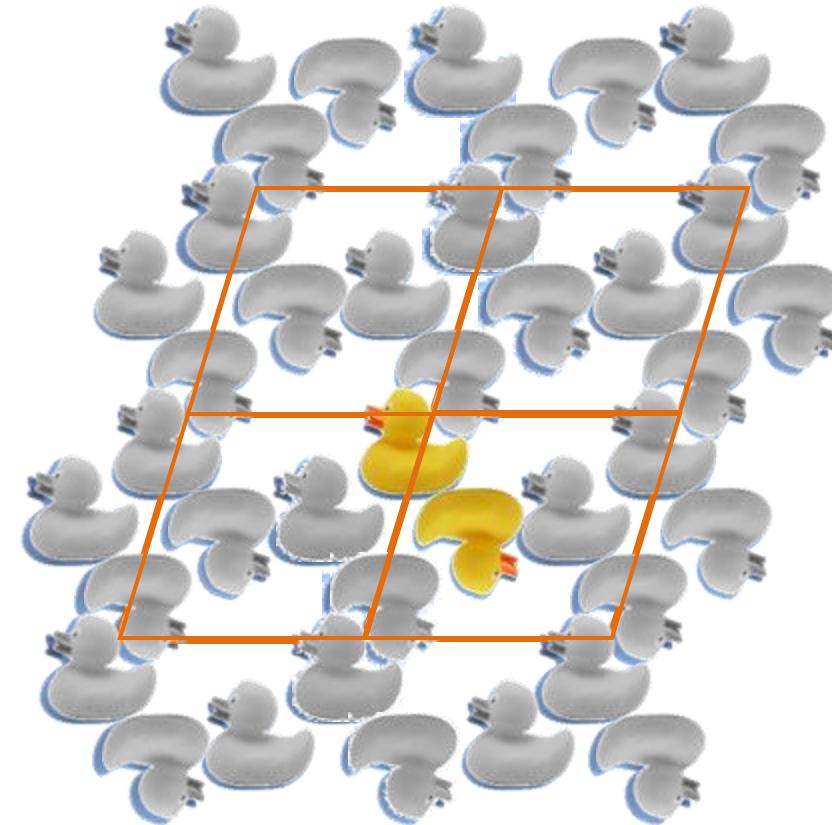
- non-crystallographic symmetry arises when there is more than one copy of a component in the asymmetric unit (asu)
- not all crystals have non-crystallographic symmetry



## Crystallographic Symmetry

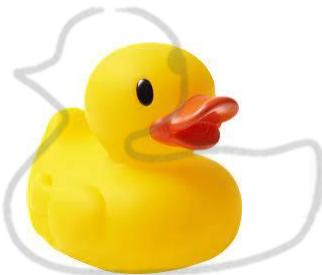


## Non-crystallographic Symmetry

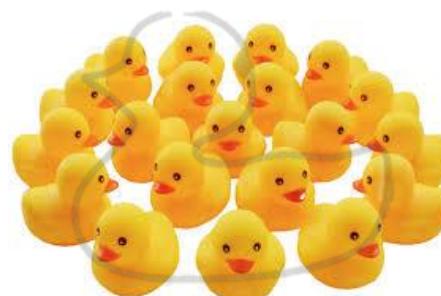


# asymmetric unit contents

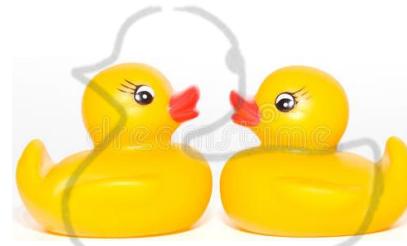
- duplication: non-crystallographic symmetry



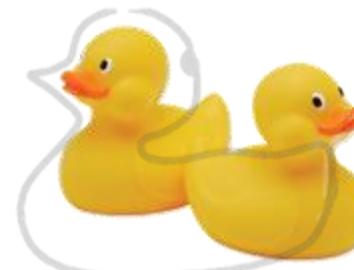
A duck



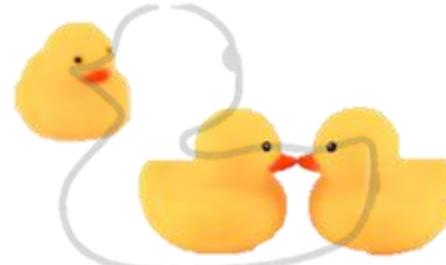
A raft of ducks  
non-  
crystallographic  
symmetry



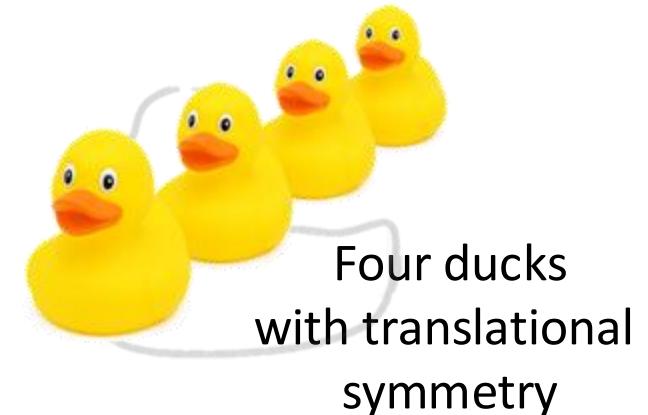
A brace of ducks  
with **point group**  
symmetry



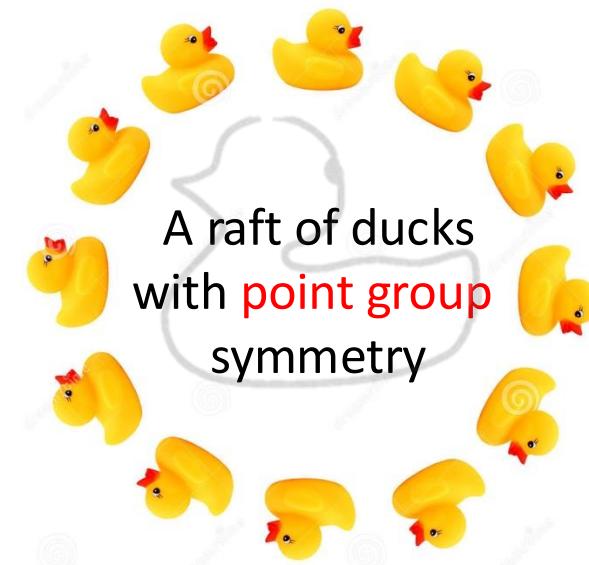
A brace of ducks  
No **point group** symmetry



A leash of ducks  
with partial **point group**  
symmetry



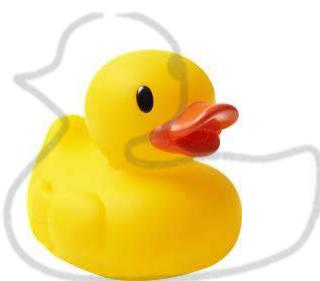
Four ducks  
with translational  
symmetry



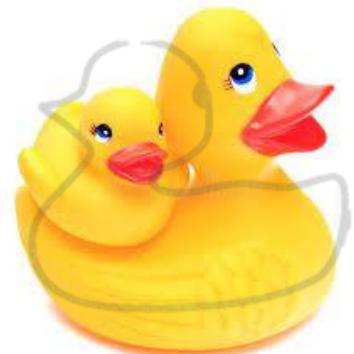
A raft of ducks  
with **point group**  
symmetry

# asymmetric unit contents

- Each ‘duck’ can itself represent a complex



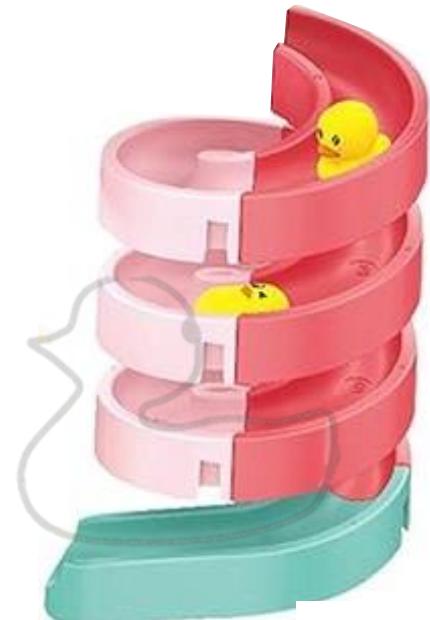
A duck



e.g. protein-ligand complex



e.g. protein complex



e.g. ribosome



e.g. (secondary) structure elements

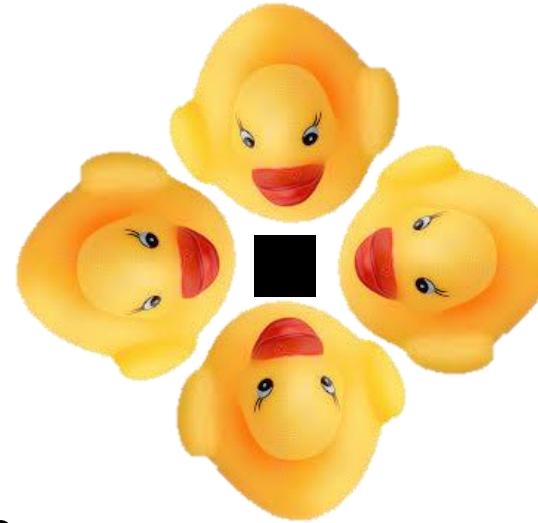


# asymmetric unit and biological assemblies



Asymmetric unit

Crystallographic 4-fold



biological assembly  
(tetramer)

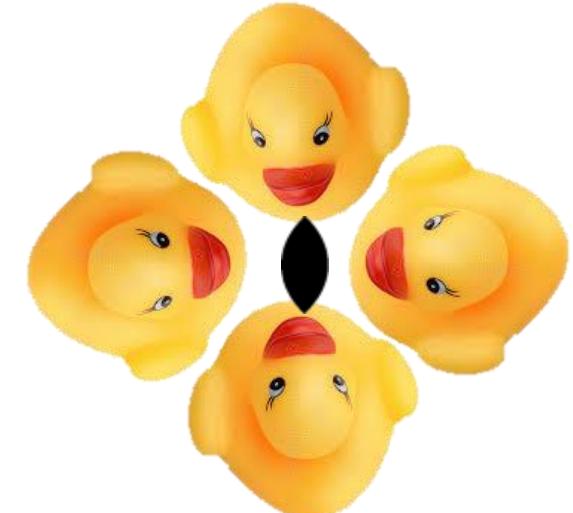
not biological assembly

*Relationship betweenasu and  
biological assembly may involve  
crystal symmetry*



Asymmetric unit

Crystallographic 2-fold



# asymmetric unit and biological assemblies

biological assembly (tetramer)



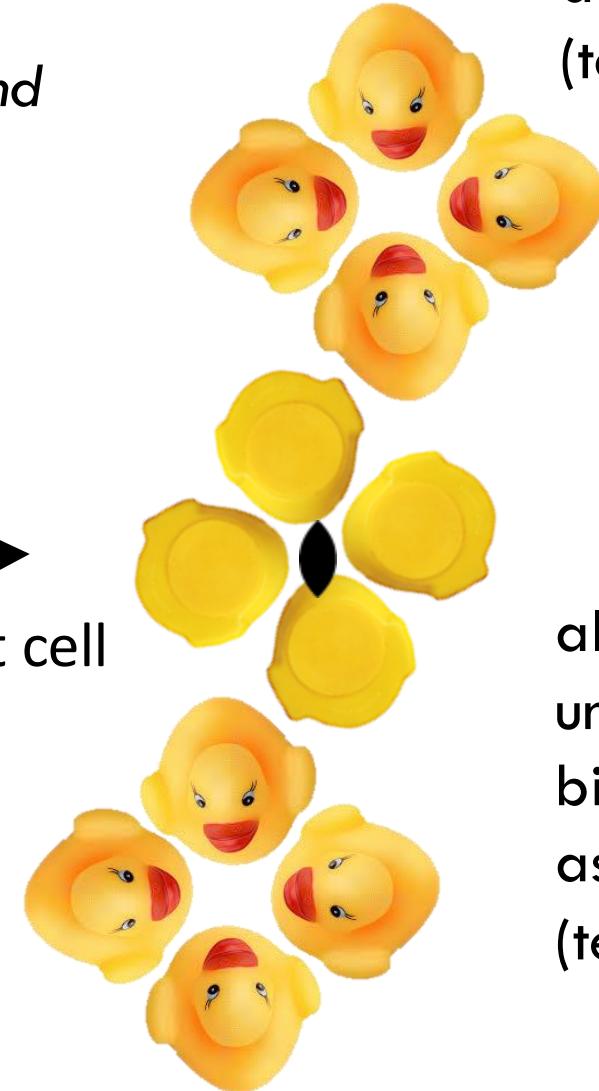
Asymmetric unit  
not biological assembly

*Relationship betweenasu and biological assembly can be complicated*

Crystallographic  
2-fold

Unit cell

biological assembly (tetramer)



biological assembly (tetramer)

# Matthew's coefficient

# Matthew's coefficient

- First calculated by Brian Matthews in 1968 (over 3500 citations)
- Most crystals are 50% protein by volume
- Can be used to estimate the contents of the asymmetric unit
- *Self Rotation Function*
- *TNCS Order*

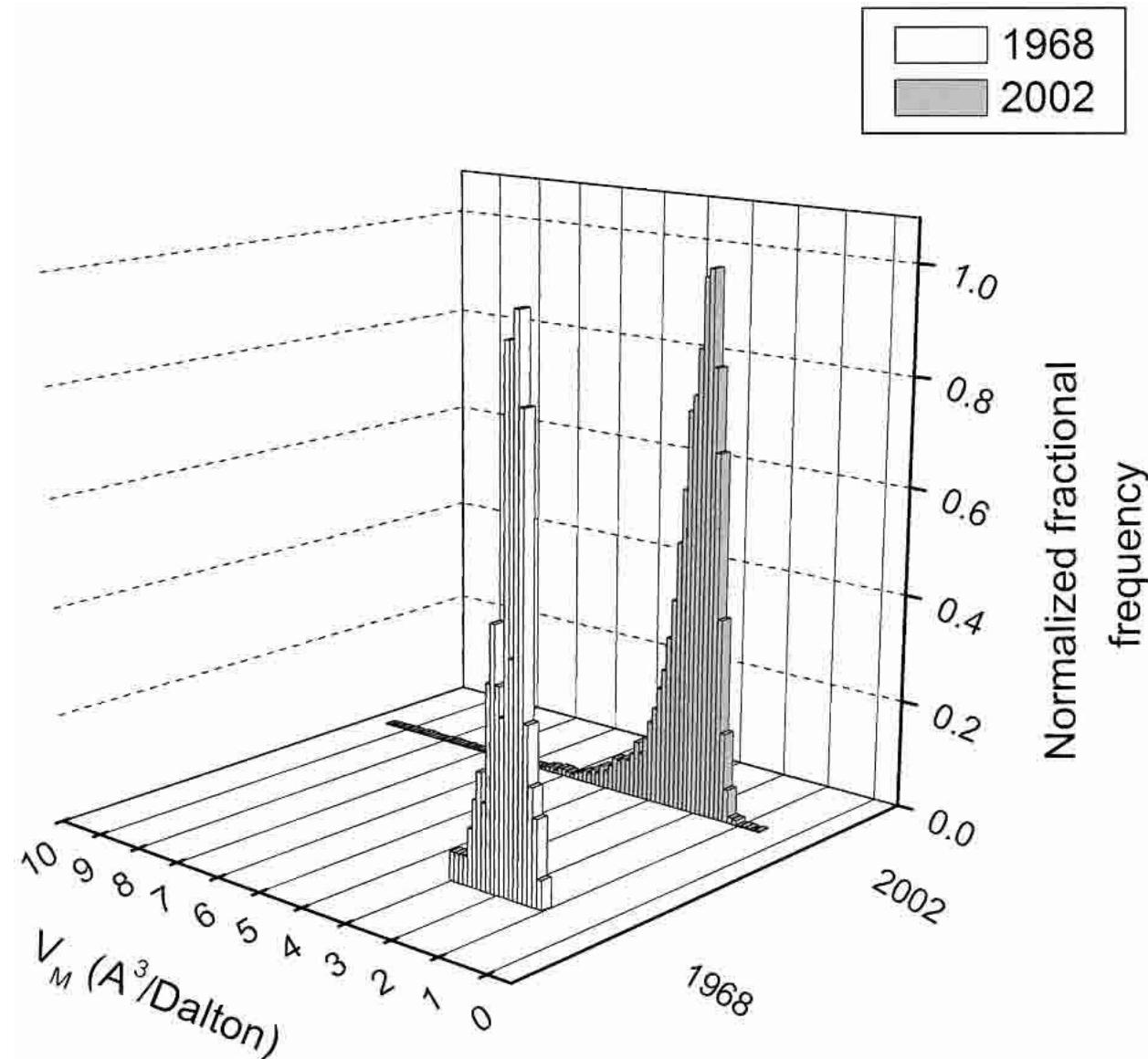
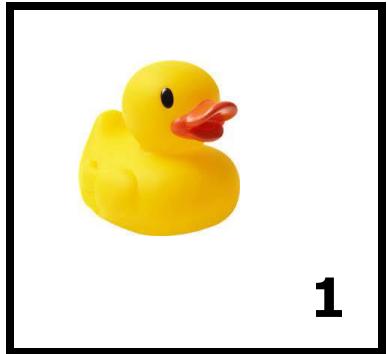


Figure 1: Kantardjieff and Rupp (2003)

# components of asymmetric unit

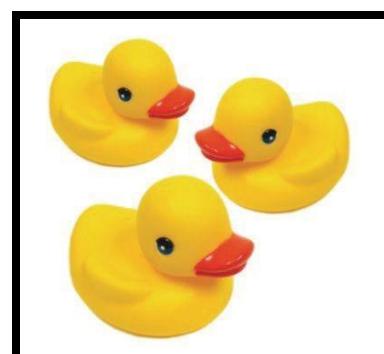
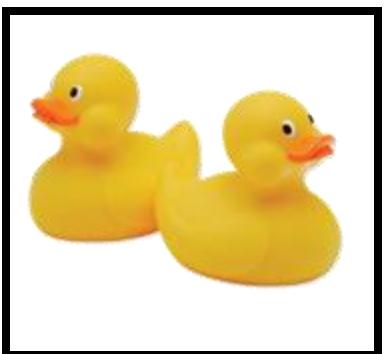
- About 50% solvent



With low numbers of possible copies, options are low

1 can only be 1

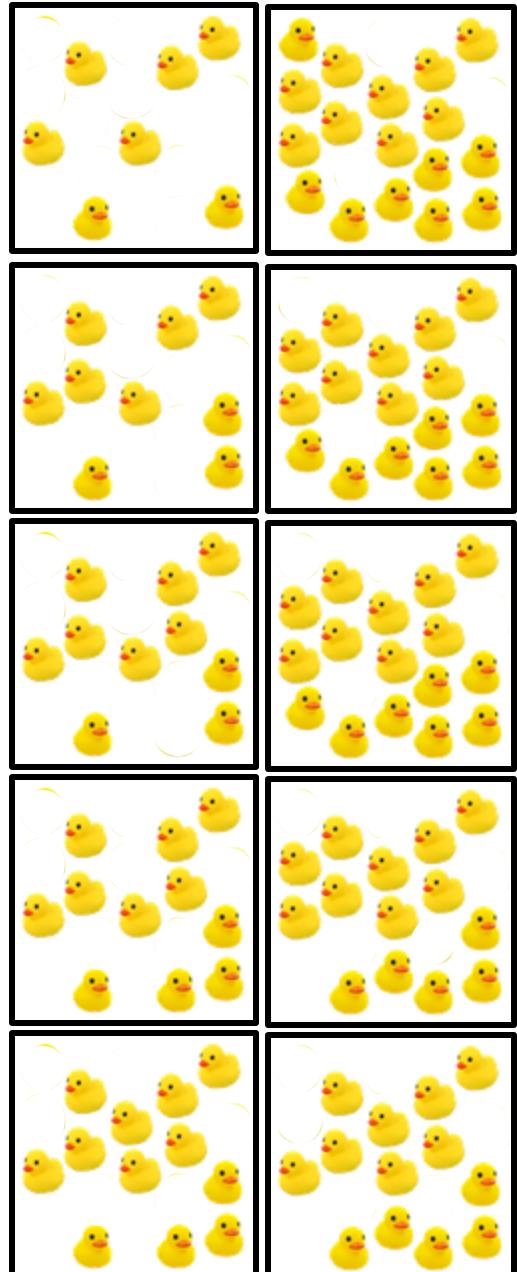
For 1, 2 or 3,  
1 is unlikely



1  
2  
3

1	7	13	19
2	8	14	20
3	9	15	21
4	10	16	22
5	11	17	23
6	12	18	24

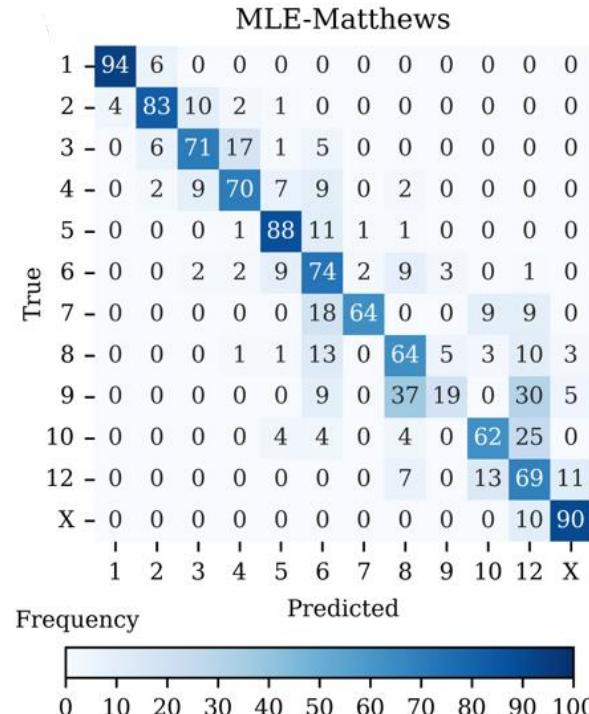
With high numbers of possible copies, options are much greater



# xtricorder

- The self rotation function and tNCS order helps predict certain asymmetric unit compositions
- The increase in overall accuracy is from 0.44 for the Matthew estimation to 0.81 for MLE-Matthews
- Gained mostly from improvements in
  - predictions for higher asymmetric unit copy numbers
  - lower symmetry space groups, including space groups  $P2_12_12_1$ ,  $P2_1$  and  $C2$

Matthews												
True	1	2	3	4	5	6	7	8	9	10	12	X
1 - 94	2	0	0	4	0	0	0	0	0	0	0	0
2 - 63	34	3	1	0	0	0	0	0	0	0	0	0
3 - 1	72	22	5	1	0	0	0	0	0	0	0	0
4 - 0	16	65	14	4	2	0	0	0	0	0	0	0
5 - 0	0	32	47	16	3	1	1	0	0	0	0	0
6 - 0	0	7	50	29	12	2	0	0	0	0	0	0
7 - 0	0	0	36	27	9	0	27	0	0	0	0	0
8 - 0	0	0	1	36	25	26	5	1	1	0	4	
9 - 0	0	0	0	28	16	33	7	0	0	2	14	
10 - 0	0	0	0	0	12	21	29	21	17	0	0	
12 - 0	0	0	0	0	0	11	24	16	22	9	18	
X - 0	0	0	0	0	0	0	0	3	20	20	57	
	1	2	3	4	5	6	7	8	9	10	12	X



search order

# search order permutations

- 3 components
- 6 search orders – search ORDER can be important

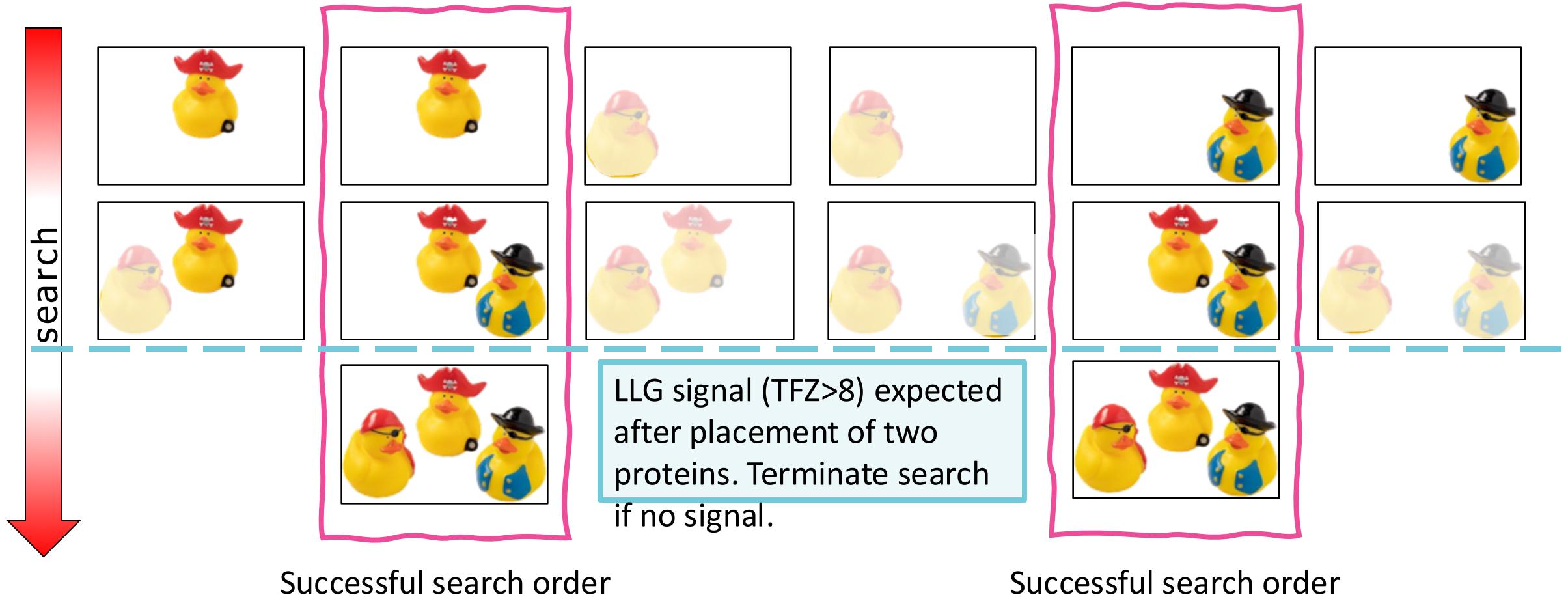
Protein/Domain present with high B-factor  
Density is poor  
Mobile in lattice and/or flexible hinge motion





# search order permutations

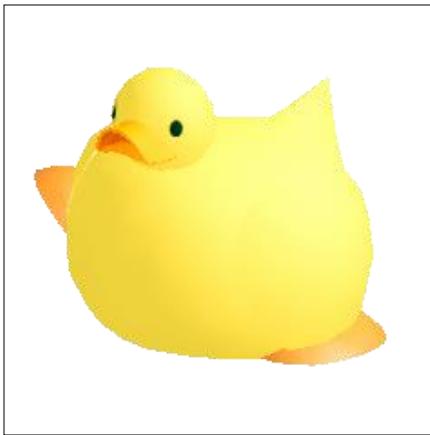
- 3 components
- 6 search orders - but only when 'bad guy' placed last is it solved



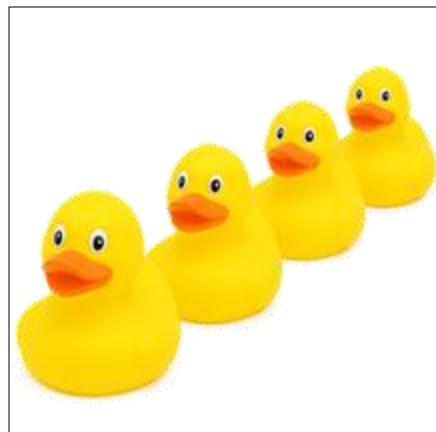
Protein present with high B-factor  
and can only be placed when other  
components are placed, which  
increases the signal  
**THE BAD GUY – BLOCKS PROGRESS**

pathologies

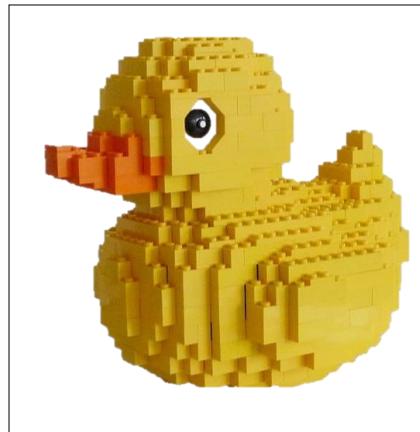
# Pathologies



Anisotropy



Translational non-crystallographic symmetry



High Mosaicity

Twinning



# Practical Pathologies

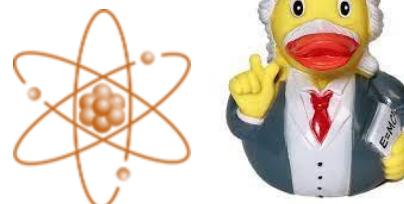
- Some pathologies can only be cured with better crystals or better crystal handling or better data collection strategies
  - High mosaicity
  - Low resolution
- Other pathologies can be cured using computational methods after data collection
  - Anisotropy
  - Twinning
  - Translational non-crystallographic symmetry



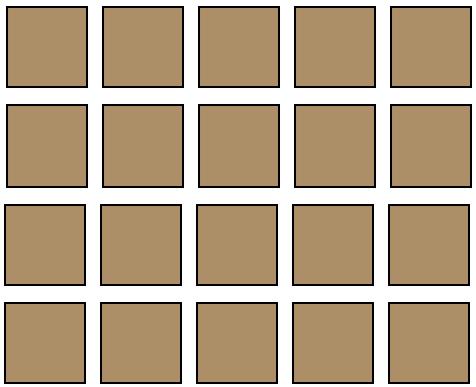
# low resolution or incomplete data

- The likelihood is summed over all reflections
- The fewer the number of reflections the lower the signal
- Collect data carefully and to as high a resolution as possible
  - Don't limit because you don't want to bother adding lots of water molecules...
  - Single atom as a model!

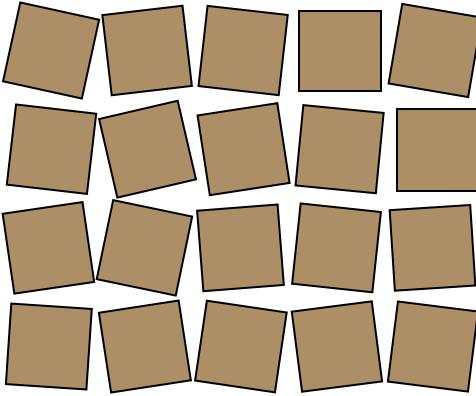
$$LLGI = \sum_{\mathbf{h}} \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$



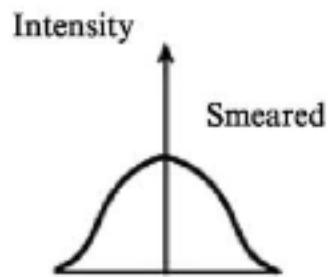
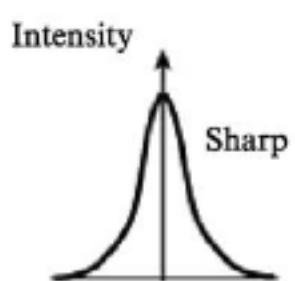
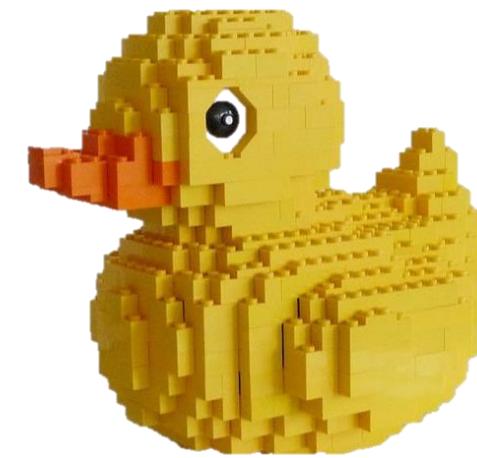
# mosaicity and low resolution



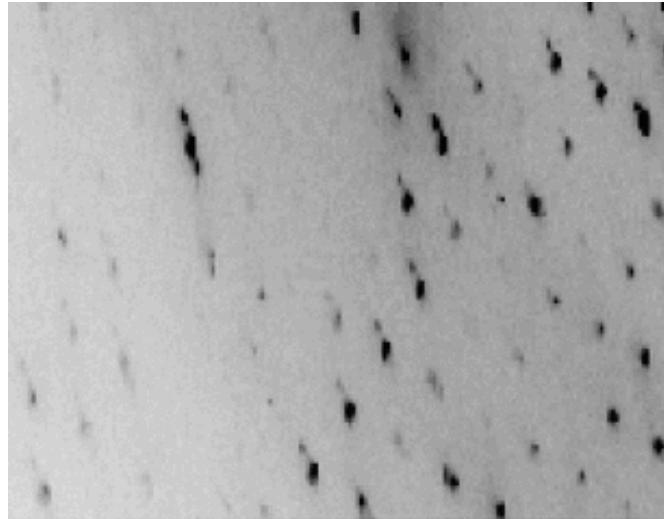
(Almost) Perfect crystals



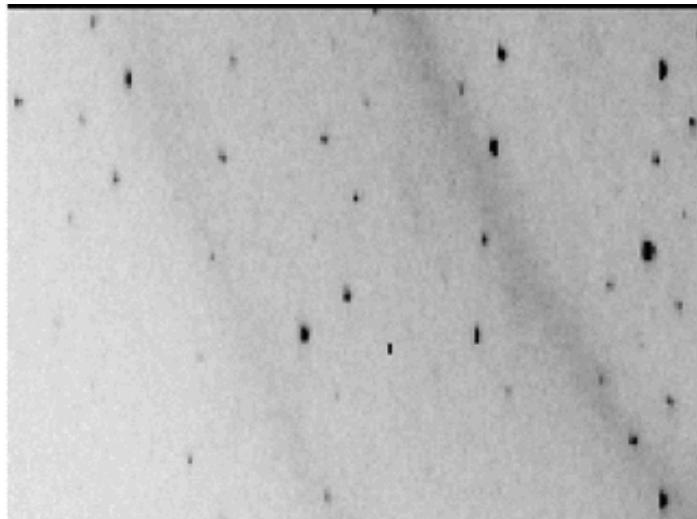
Real crystals  
Mosaic blocks  
Highly exaggerated



# mosaicity

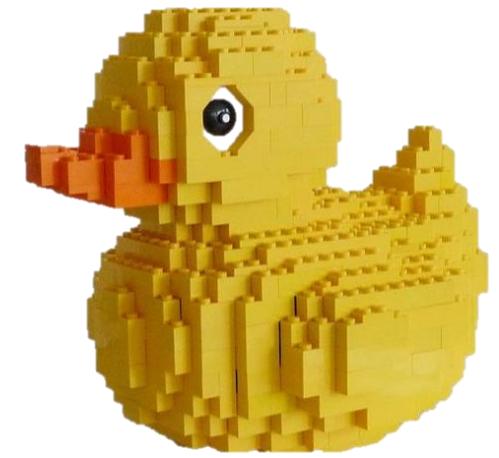


High mosaicity and split spots from a big crystal (0.5-0.8 mm in size) due to bad cryo-cooling (osmotic shock, differential expansion)



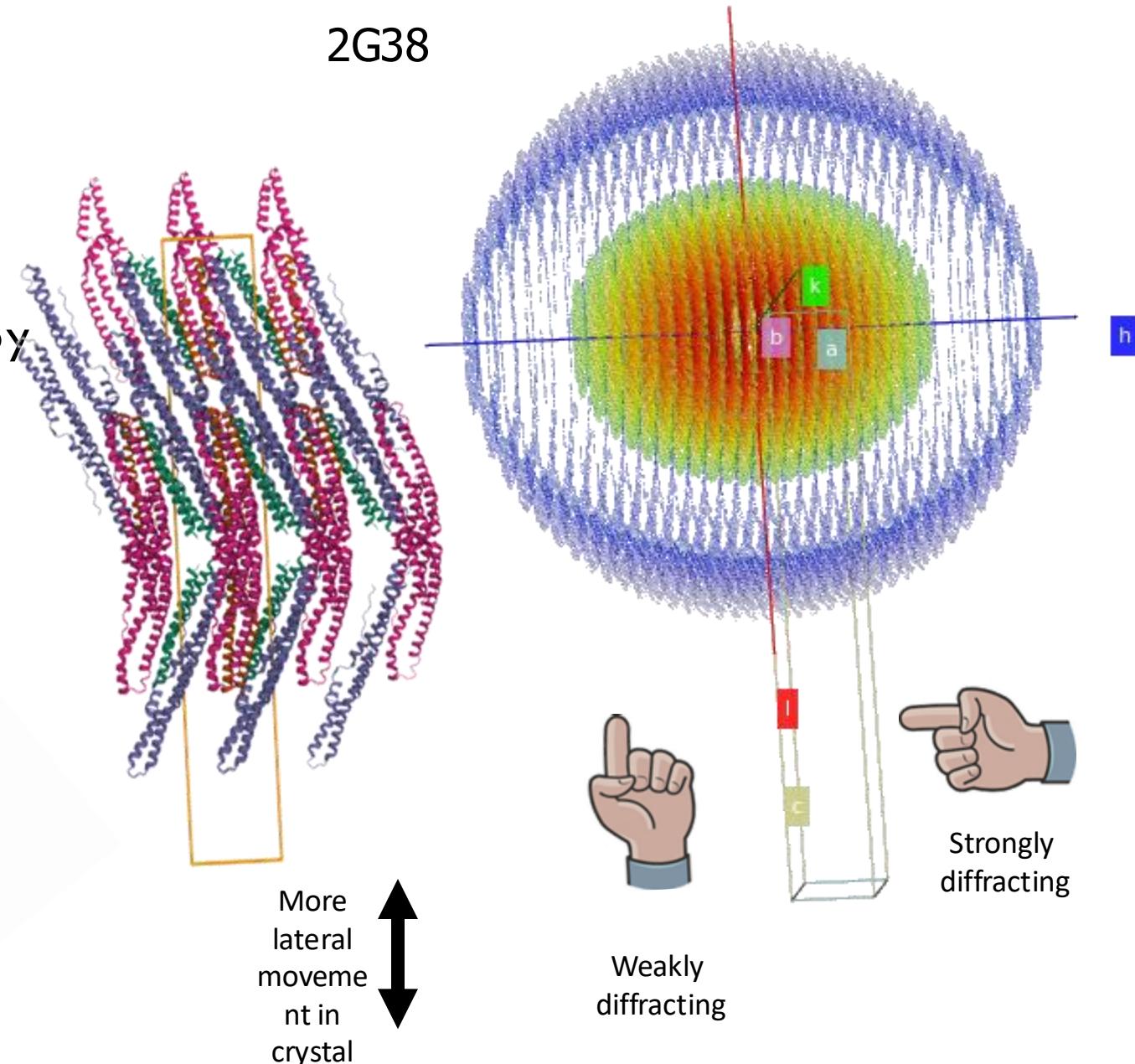
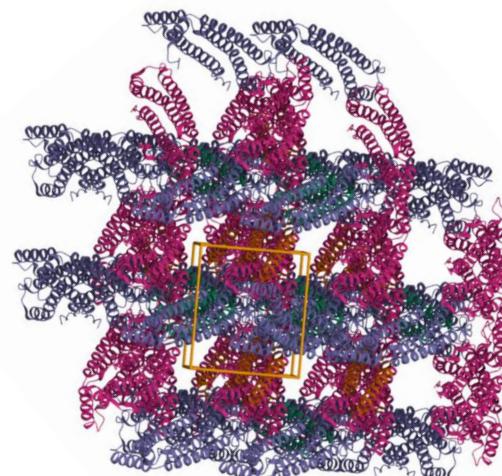
Single spots (not split) from a crystal of size 0.2 mm from which a full data set could be recorded and processed to 1.2 Å.

$$LLGI = \sum_{\mathbf{h}} \log \left( \frac{2E_e}{1 - D_{obs}^2 \sigma_A^2} \exp \left( -\frac{E_e^2 + D_{obs}^2 \sigma_A^2 E_C^2}{1 - D_{obs}^2 \sigma_A^2} \right) I_0 \left( \frac{2E_e D_{obs} \sigma_A E_C}{1 - D_{obs}^2 \sigma_A^2} \right) \right)$$



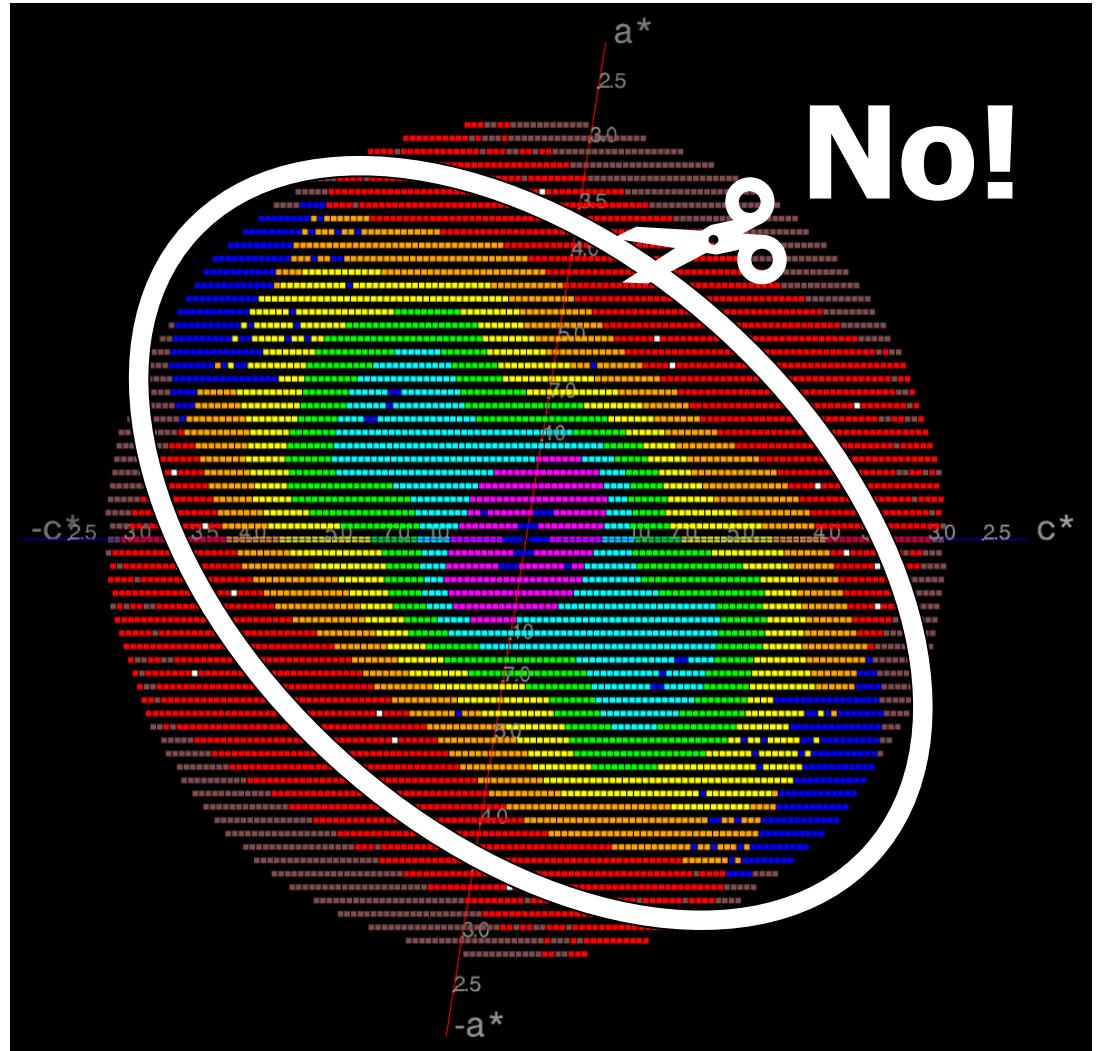
# anisotropy correction

- Many crystals diffract to different resolutions in different directions
- Correct this by making intensity distribution the same in all directions by refining parameters of the anisotropy tensor



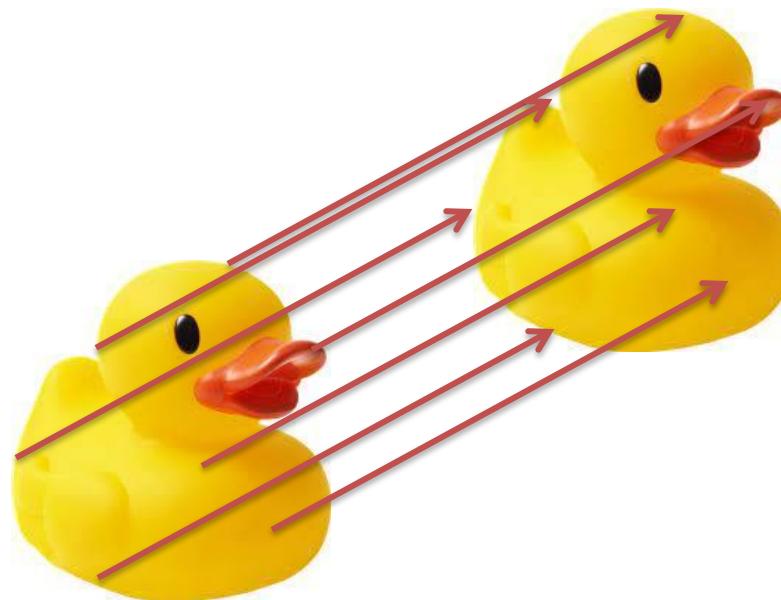
# anisotropic data in phaser

- Do not use anisotropically truncated (staraniso) data in phaser for molecular replacement

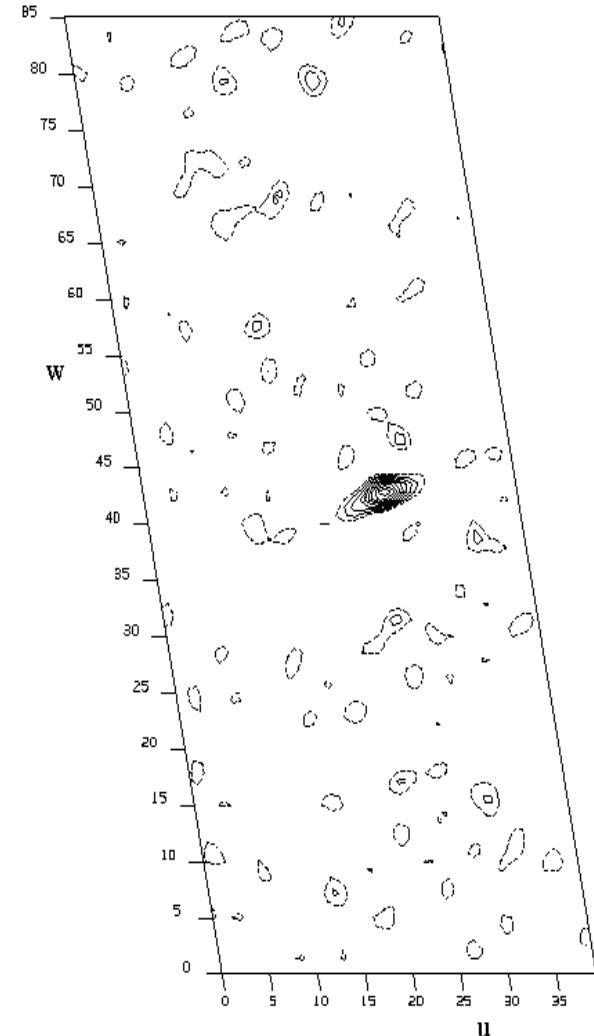


# translational non-crystallographic symmetry

- Patterson is a vector map of the crystal
- Calculated as FT of unphased intensities
- Large origin peak
- TNCS indicated by Patterson Peak
- 16% origin at 5Å

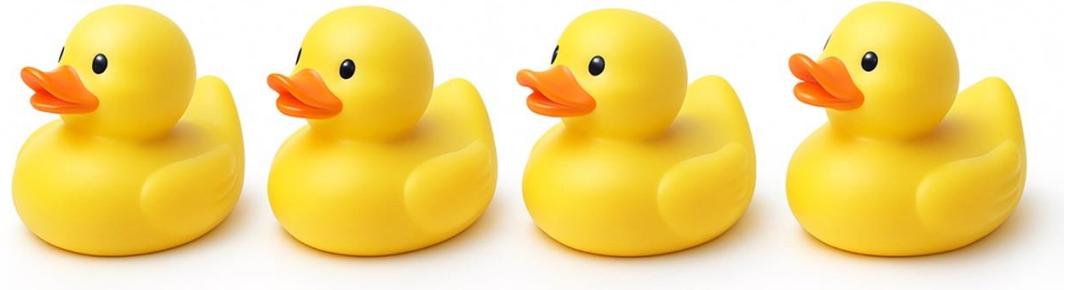


Native Patterson of mouse renin.

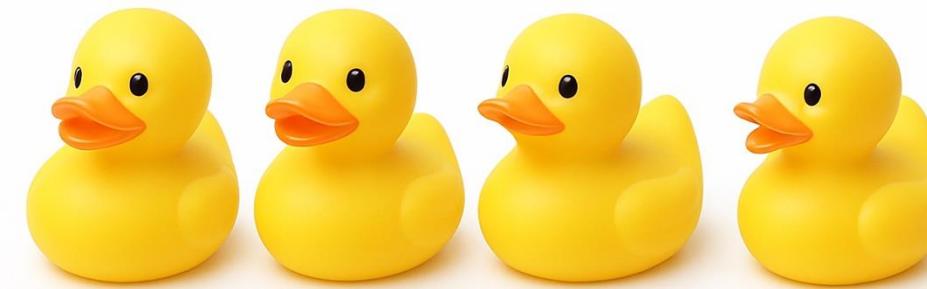


# translational non-crystallographic symmetry

- Molecules related by a vector translation
- But the translational symmetry is never exact
  - Differences from mean vector
  - Angular perturbations
- The differences from perfect translation are usually significant
- Can lead to ambiguous tncs
  - when is tncs not tncs?



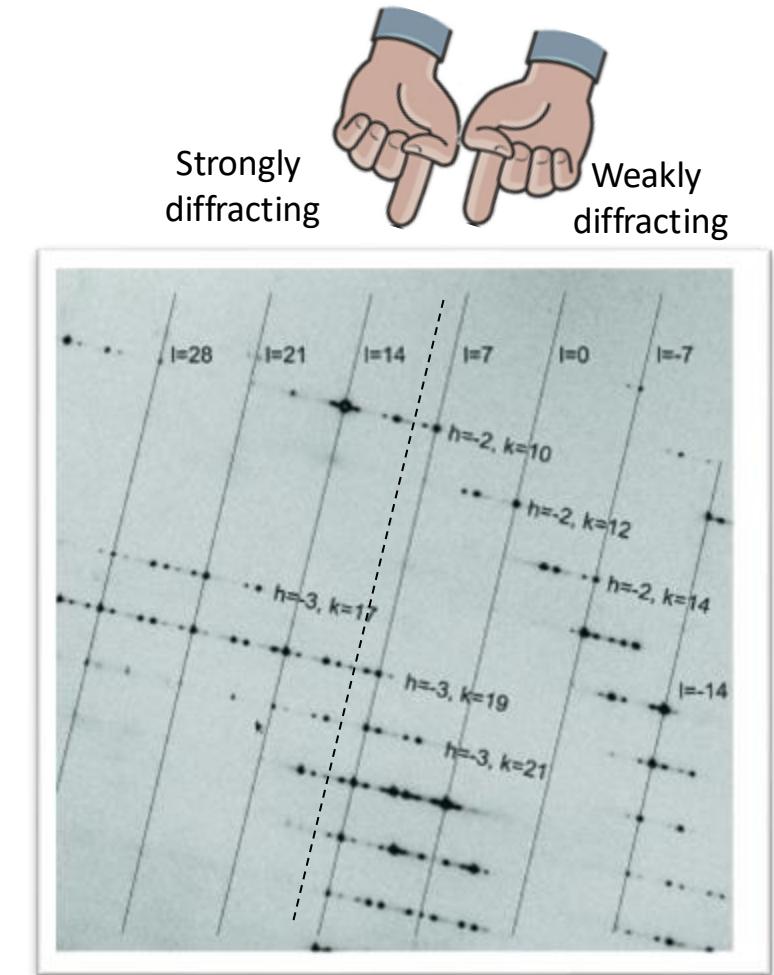
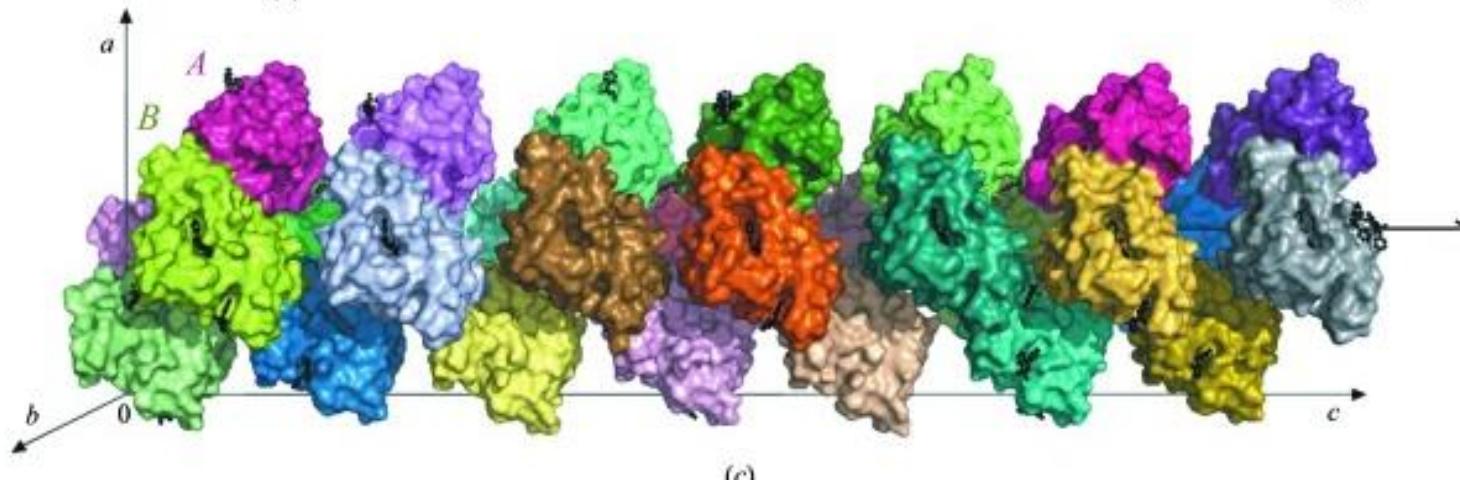
perfect translation



imperfect translation

# translational non crystallographic symmetry

- Make intensity distribution the same in all layers by refining parameters for the expected intensity factors



# translational non-crystallographic symmetry

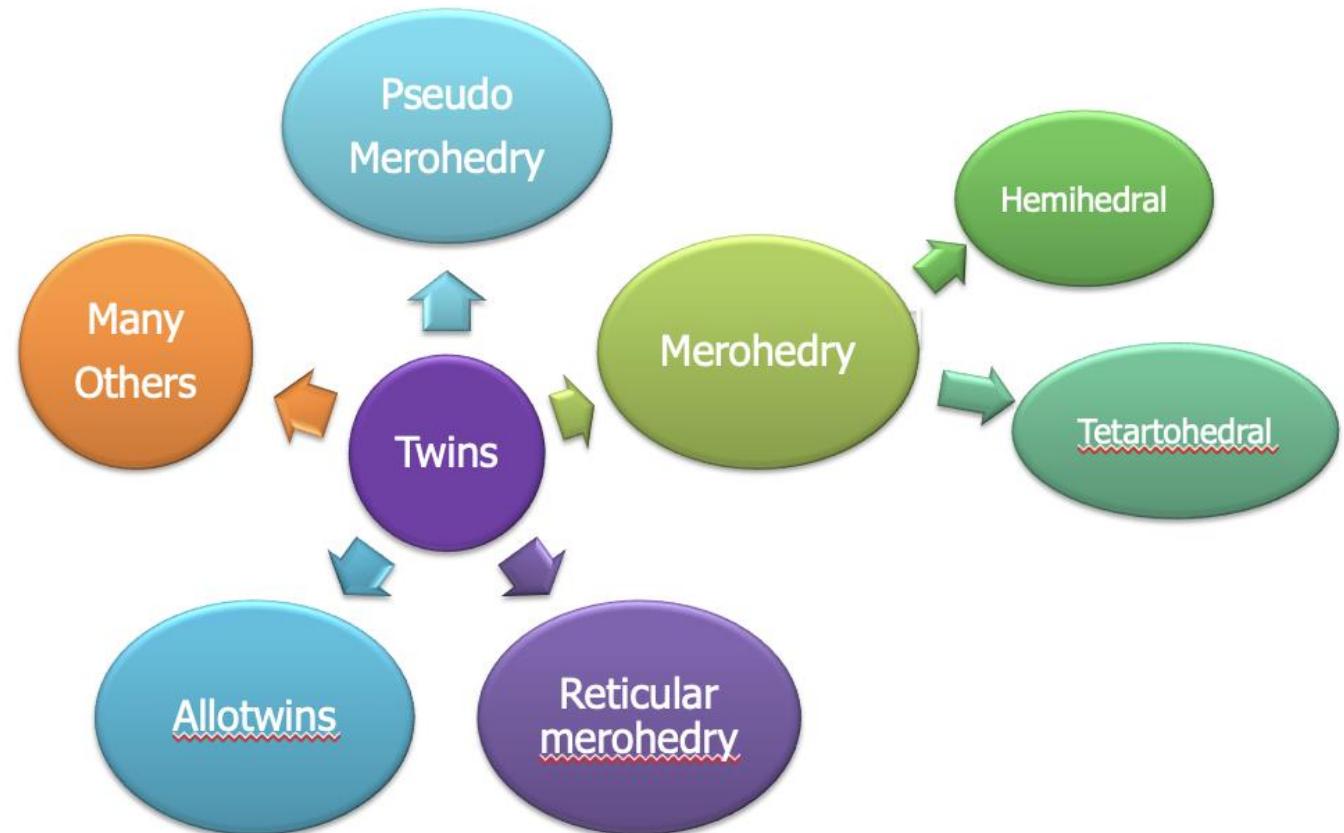
- If TNCS is not accounted for then  $\text{TFZ} > 8$  does not indicate a correct placement
  - TFZ always higher
  - TFZ very high is wrong
  - No signal, all have high TFZ

Uncorrected TF Z-score	Uncorrected LLG score	Solved?
>8	< 64	no

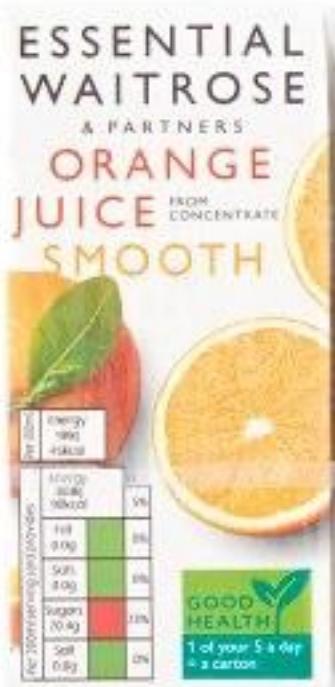
- When TNCS is ‘corrected for’ then the TFZ values are those expected of data without TNCS

# What is Twinning?

- There are many
  - many
  - many
    - many different types



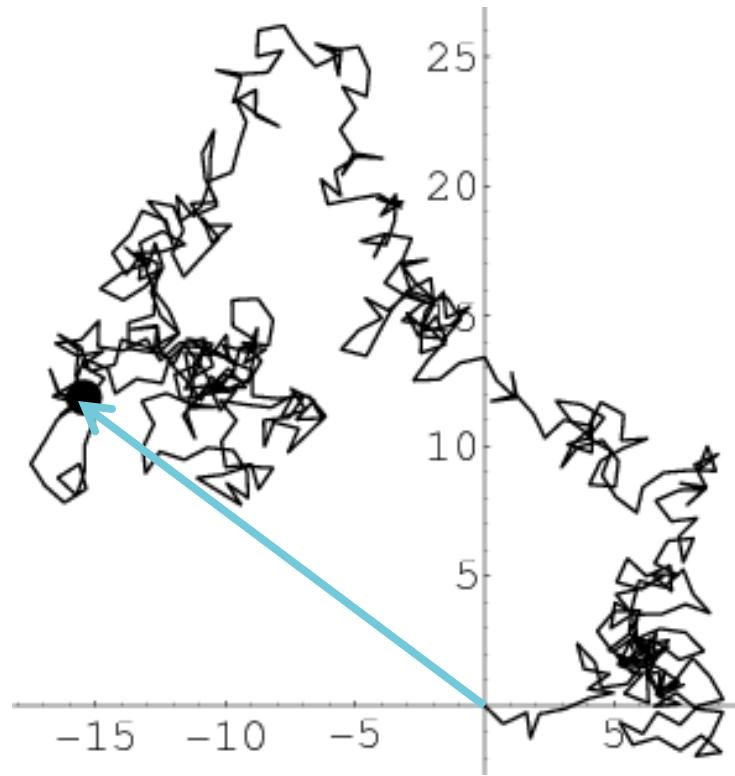
# orange juice carton



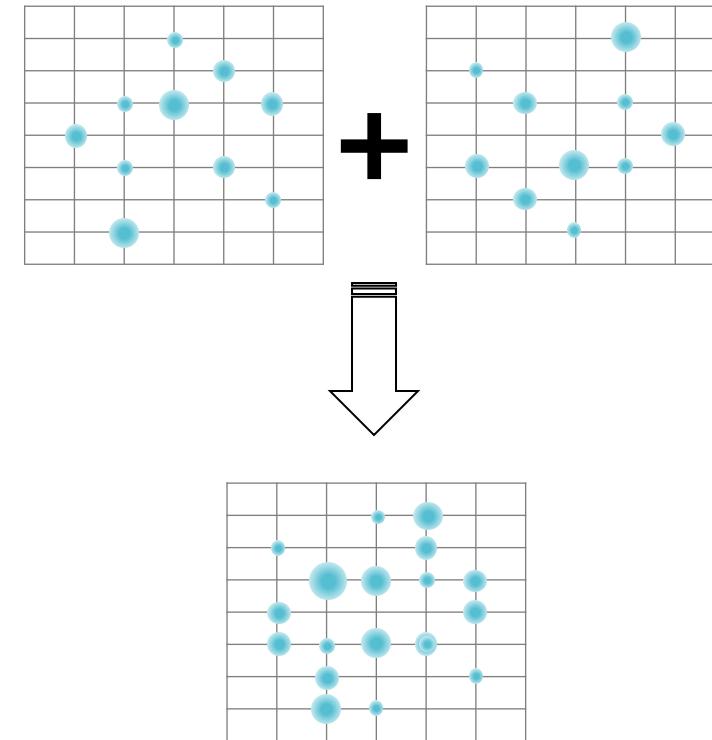


# Twinning versus Disorder

**Sum of F's**



**Sum of I's**



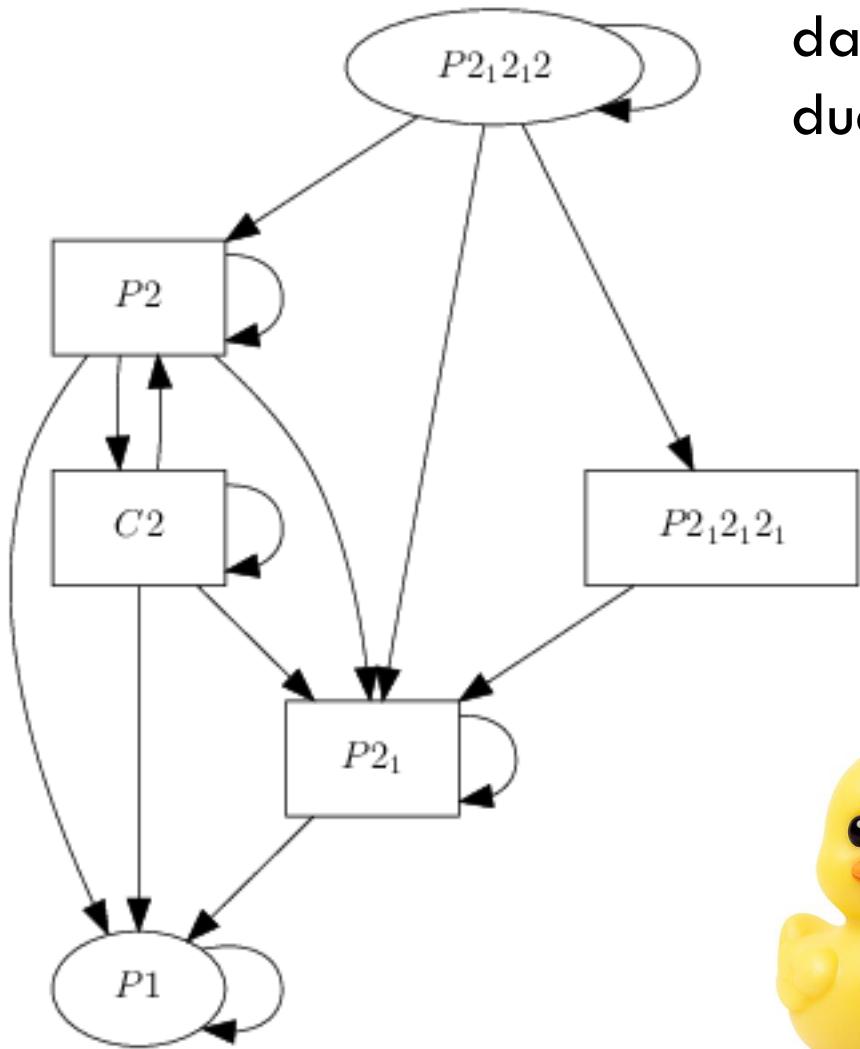
# Twinning versus Disorder

- For twinning, the size of twin domains is large compared with the cell dimensions
  - Diffracted X-rays do not interfere
  - **Sum of I's not F's**
- This is in contrast to disorder in the crystal where the differences are between neighbouring cells
  - Diffracted X-rays represent the spatially and/or temporally averaged content of the unit cells
  - **Sum of F's not I's**
- There are some nasty intermediate cases

# Twinning and Molecular Replacement

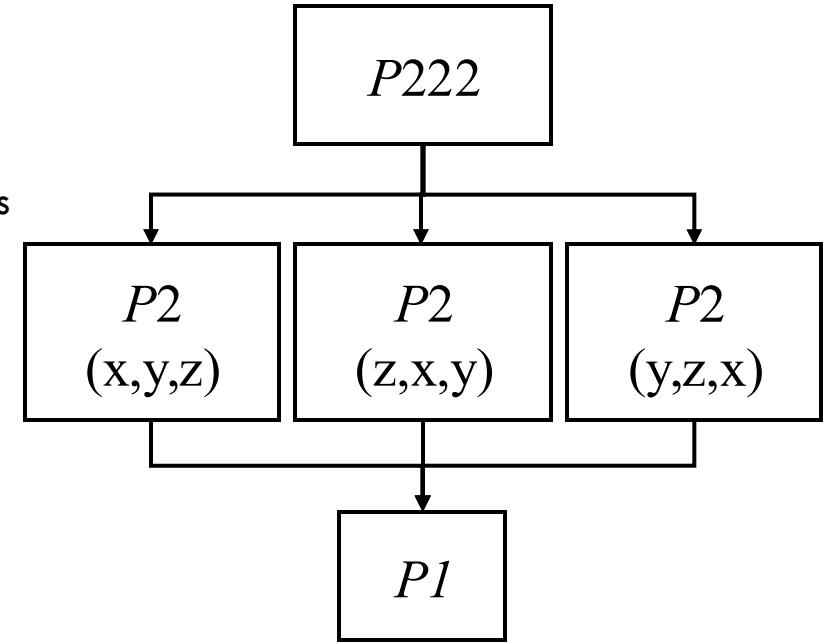
- Molecular replacement with Phaser is often successful when there is twinning without you ever knowing you have twinning
- However, sometimes you may need to try a lower symmetry space group

# subgroups



data over-merged  
due to twinning

P2 standard settings  
axis permutations



subgroups  
for molecular replacement

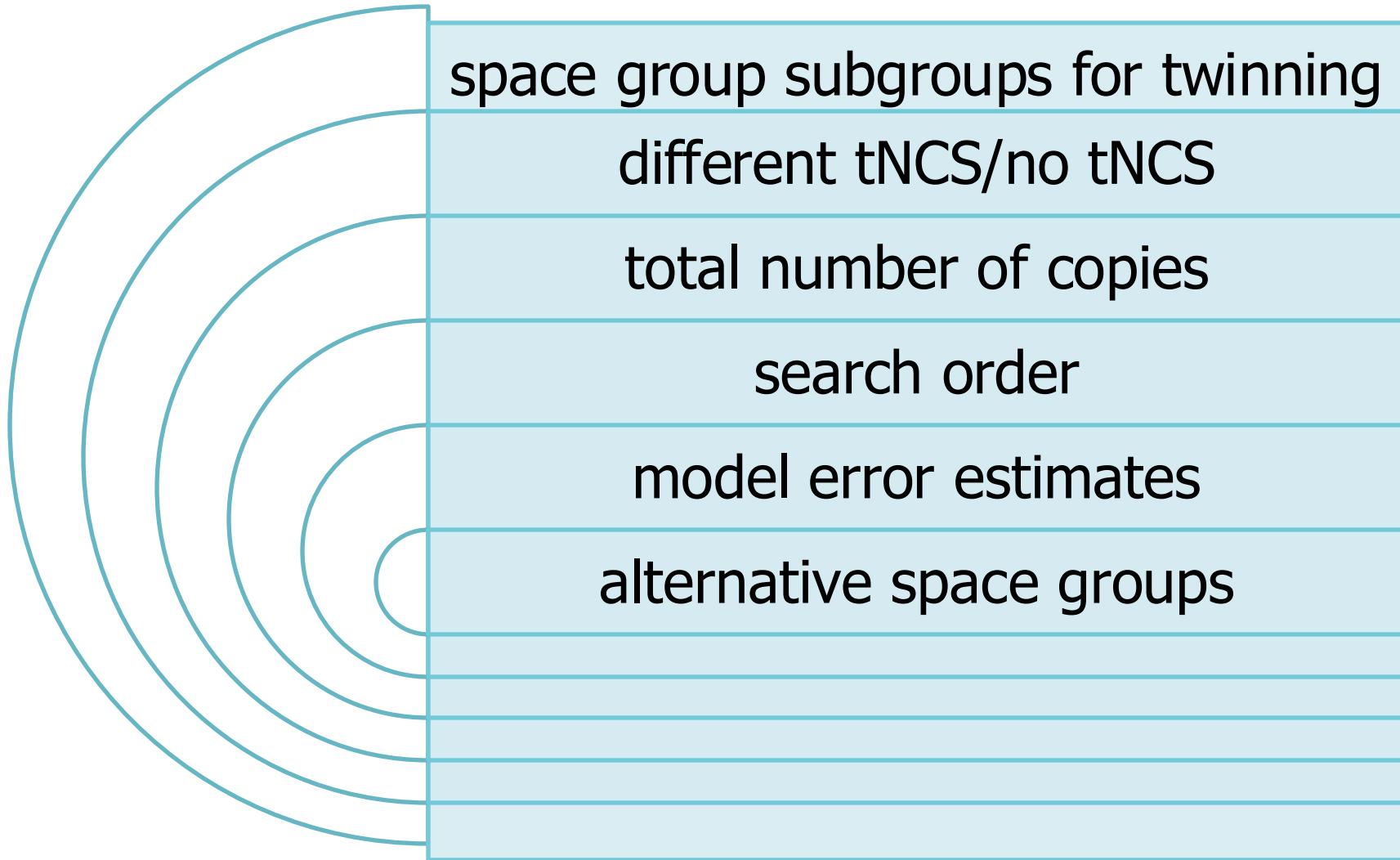
# Pathologies



Translational NCS +  
Twinning

- Translational NCS masks twinning
  - Correcting for translational NCS unmasks twinning
- Phaser gives a P-value for there being twinning in the presence of translational NCS

# PhaserTNG – new software for molecular replacement



# Phasing

- All sorts of phasing + docking
  - molecular replacement (MR)
  - single-anomalous dispersion (SAD)
  - molecular replacement with single-anomalous dispersion (MR-SAD)
  - EM docking for phased data (“EMplacement”)
- Distribution
  - CCP4
  - Phenix



UNIVERSITY OF  
CAMBRIDGE



Cambridge Institute for  
Medical Research



questions

