Solution of a structure

The way around the phase problem





Difference Fourier map

Structure solution

Patterson methods (vector methods)

Patterson method determine directly the position of several atoms in the unit cell. From the position of these methods, in particular if they are heavy metals with many electrons, the Fourier transformation yields an initial, relatively reasonable set of phases.

- Direct methods
- Dual space methods
- Charge flipping methods

Structure solution

Patterson methods (vector methods):

The Fourier transform of the structure factors yields the electronic density distribution:

$$TF[F(r^*)] = \rho(r)$$

We cannot determine *F*, since we lack the phases. We can only measure the intensities *I*. Thus, what do we obtain, if we Fourier transform the intensities $I = |F|^2$?

$$\frac{\text{Reminder:}}{F(hkl) = F(r^*) = \sum_i f_i e^{i2\pi(hx_i + ky_i + lz_i)} = \sum_i f_i e^{i\Phi_i(r^*)} = \sum_i f_i (\cos \Phi_i(r^*) + i\sin \Phi_i(r^*)) = A(r^*) + iB(r^*)$$
$$F^2(r^*) = A^2(r^*) + B^2(r^*) = (A(r^*) + iB(r^*)) \cdot (A(r^*) - iB(r^*)) = F(r^*)F(-r^*)$$

$$TF[F^{2}(r^{*})] = TF[F(r^{*})F(-r^{*})] = TF[F(r^{*})] * TF[F(-r^{*})] = \rho(r) * \rho(-r)$$

<u>Convolution of functions:</u> $TF[f(r)g(r)] = TF[f(r)] * TF[g(r)] \qquad f(r) * g(r) = \int f(r)g(u-r)dr = h(u)$

$$TF[F^{2}(r^{*})] = \rho(r) * \rho(-r) = \int \rho(r) \cdot \rho(u+r)dr = P(u)$$

Patterson function

$$P(u) = \int \rho(r) \cdot \rho(u+r) \, \mathrm{d}r = TF \Big[F^2(r^*) \Big]$$

 $\begin{array}{ll} \rho(r) & : \text{ maximal when } r \text{ is the position of an atom} \\ \rho(r+u) & : \text{ maximal when } r+u \text{ is the position of an atom} \\ \Rightarrow \text{ maximal when } u \text{ is an interatomic vector} \\ \Rightarrow P(u) \text{ shows thus maxima when } u \text{ is an interatomic vector} \end{array}$

The Patterson map shows maxima at the position of interatomic vectors.

We can construct it by placing in turn each atom of a structure in the origin.





Solution using Harker peaks / "Heavy atom method"

Example: $(C_5H_5)_3$ Sb in P2₁/c $P(u) = \int \rho(r) * \rho(u+r) dr = TF [F^2(r^*)]$

1. Normalisation:

The calculated values are arbitrary. For this reason, the value of peak P(0,0,0) is set to 999. P(0,0,0) is the highest maximum possible, since every atom has a distance of 0 to himself.

Since the intensity of the Patterson peaks is proportional to the product of the numbers of electrons in the two atoms forming the interatomic vector, we can calculate the theoretical intensity of each peaks, knowing that the peak 0/0/0, which involves all atoms in the unit cell, is set to 999.

| n | atom | Ζ | Z^2 | nZ² | Theoretical peak heights: | | | |
|--|---------|---------|------------|---------------|---|--|--|--|
| 4 60 | Sb C | 51 6 | 2601 36 | 10404 2160 | Sb-Sb: Sb-C: C-C: | 51*51*999/12564 = 207 51*6*999/12564 = 24 6*6*999/12564 = 3 | | |
| Peak height of a single interatomic vector | | | | 12304 | Theoretical height of P(0 | peak 0,0,0) | | |

Solution using Harker peaks / "Heavy atom method"

| Example: | | Theor | Theoretical peak heights: | | | |
|--------------------------|------------------------|----------------|---------------------------|----------------------|--|--|
| $(C_5H_5)_3$ Sb in P2 | 2 ₁ /c | Sb-Sb | o: 51*51*999/12 | 564 = 207 | | |
| | | Sb-C: | 51*6*999/125 | 64 = 24 | | |
| 2. Harker peaks | in P2 ₁ /c: | C-C: | 6*6*999/1256 | 34 = 3 | | |
| General atomic positions | x,y,z | -x,-y,-z | -x,0.5+y,0.5-z | x,0.5-y,0.5+z | | |
| x,y,z | | -2x,-2y,-2z | -2x,0.5,0.5-2z | 0,0.5-2y,0.5 | | |
| -x,-y,-z | 2x,2y,2z | | 0,0.5+2y,0.5 | 2x,0.5,0.5+2z | | |
| -x,0.5+y,0.5-z | 2x,0.5,0.5+2z | 0,0.5-2y,0.5 | | 2x,-2y,2z | | |
| x,0.5-y,0.5+z | 0,0.5+2y,0.5 | -2x,0.5,0.5-2z | -2x,2y,-2z | | | |

Harker peaks are interatomic vectors between symmetry-related atoms.

| Intensity | u | V | W | assignment | atoms | |
|-----------|-------|-------|-------|-------------|-----------|----------------|
| 999 | 0 | 0 | 0 | zero peak | | |
| 460 | 0 | 0.396 | 0.5 | Harker peak | 2 x Sb-Sb | 0,0.5±2y,0.5 |
| 452 | 0.420 | 0.5 | 0.705 | Harker peak | 2 x Sb-Sb | ±2x,0.5,0.5±2z |
| 216 | 0.421 | 0.106 | 0.205 | Harker peak | 1 x Sb-Sb | 2x,2y,2z |

Solution using Harker peaks / "Heavy atom method"

Example:

 $(C_5H_5)_3$ Sb in P2₁/c

| Intensity | u | V | w | assignment | atoms | |
|--------------------------------------|--------------------------|----------------------------|----------------------------|--|-------------------------------------|--|
| 1. 999 2. 460 3. 452 4. 216 | 0 0 0.420 0.421 | 0 0.396 0.5 0.106 | 0 0.5 0.705 0.205 | zero peak Harker peak Harker peak Harker peak | 2 x Sb-Sb 2 x Sb-Sb 1 x Sb-Sb | 0,0.5±2y,0.5 ±2x,0.5,0.5±2z 2x,2y,2z |

3. Calculation of the Sb position

| peak 2: | $0.5\pm 2y = 0.396$ | \Rightarrow | y=0.052 or y=-0.052 |
|---------|---------------------|---------------|-----------------------|
| peak 3: | 2x = 0.420 | \Rightarrow | x=0.210, z=0.103 |
| | 0.5+2z = 0.705 | \Rightarrow | or x=-0.210, z=-0.103 |
| etc. | | | |

- Can only determine the position of heavy atoms, which have remarkably higher intensities.
- Only applicable, when there are only few (1-2) heavy atoms (no heavy atom heavy atom peaks, which are not Harker peaks).

Solutions using the Patterson map Patterson map superposition (SHELXS)

One peak is chosen and the original Patterson map is shifted by this peak.

If two peaks superimpose, they are part of our searched structure. Since the Patterson function is always centrosymmetric, we obtain the structure of our molecule and its inverted image + all other symmetry equivalents.

A third superposition reduces it to only one structure and eliminates false hits. This step is not always done or necessary.



Solutions using the Patterson map

Patterson map superposition (SHELXS)

Attention: Using a second shift works only when both vectors have the same origin!

Preferably we want to use Harker peaks.



Solutions using the Patterson map

Patterson map superposition (SHELXS)

Another problem: Patterson peaks are broader than reflections. We can thus have a incorrect superposition of several peaks, generating a lot of false hits.



Patterson methods work best if there are few, very heavy atoms in the structure. The overall number of atoms is irrelevant.

Solutions using the Patterson map

Solution by "Patterson shift methods" (PATSEE, SIR, DIRDIF)

If we know a (rigid) fragment of the structure, we can look for it in the Patterson map:

- Find the correct rotation (in three dimension)
- Place each atom in the origin without changing its rotation.
- The distribution of the fragments around the origin than yields the structure



Solution des structures

Patterson methods (vector methods)

Patterson method determine directly the position of several atoms in the unit cell. From the position of these methods, in particular if they are heavy metals with many electrons, the Fourier transformation yields an initial, relatively reasonable set of phases.

Direct methods ۲

Direct methods try to establish the phases directly from the intensities of the reflections. How is this possible?

Dual space methods

Direct methods

The Patterson function is helpful if we have an unequal distribution of atoms, i. e. when a big part of the electron density is located in a small number of heavy atoms. When the electron density is more equally distributed, direct methods are more useful. They are called "direct", because we need only the intensities to determine the phases.

1948, Harker and Kasper showed that there are systematic relations between the structure factors and the symmetry elements.

Sayre equation (1953):

Starting only with two assumptions (that the electron density is never negative and that it is concentrated in well defined maxima), Sayre showed that each structure factor can be expressed by an infinite sum over all the other structure factors:

$$F_{hkl} = K \sum_{h'k'l'} F_{h'k'l'} \cdot F_{h-h',k-k',l-l'}$$

 $F_{123} = F_{100}F_{023} + F_{010}F_{113} + F_{001}F_{122} + \dots$

Triplet relationship

The Sayre equation is important, but not useful, until Hauptman and Karle showed in 1953 (Nobel price 1985) how to use it for the phase determination.

$$E_{hkl} = \sum_{h'k'l'} E_{h'k'l'} \cdot E_{h-h',k-k',l-l'}$$
 Sayre equation

When E_{hkl} is strong and if there is a pair $E_{h'k'l'}$ and $E_{h-h',k-k',l-l'}$, where both reflections are also strong, then we can ignore all the other terms to obtain **the triplet relationship** Σ_2 :

$$E_{H} \approx E_{K} \cdot E_{H-K} = |E_{K}|e^{-i\varphi_{K}} \cdot |E_{H-K}|e^{-i\varphi_{H-K}} = |E_{K}| \cdot |E_{H-K}| \cdot e^{-i(\varphi_{K}+\varphi_{H-K})}$$

And for the phases:
$$\varphi_{H} \approx \varphi_{K} + \varphi_{H-K}$$
$$H=h,k,l; K=h',k',l'$$

We use normalized reflections *E* rather than *F*. The normalisation negates the decrease of the intensities with θ which is caused by thermal motion. The electron density calculated from *E* values is thus more focalised and the Sayre equation fits better for *E* than for *F* (assumption that the electron density is located in well defined centers).

Triplet relationship Σ_2

For a centrosymmetric structure the phases can have only values of 0 or π . (This is due to the convention that an inversion center is always placed at the origin of the chosen coordinate system).



Direct methods

The determination/estimation of the initial phases in 3 steps:

- A. Find some strong reflections with known phases
- B. Estimate (somehow) the phases for a subset of strong reflections (5-10 %).
- C. Determine the rest of the phases (90-95%) using the tangent formula.



Known phases - Triplet Σ_1

A triplet Σ_1 is a special case of a triplet relationship where K=H-K:

$$E_{H} \approx E_{K} \cdot E_{K}$$
, e.g. $E_{246} = E_{123} \cdot E_{123}$

Centrosymmetric structures: If E_H is strong and E_K is strong, than E_H is always positive, independent from the phase of E_K . Σ_1 triplets are very useful, but unfortunately relatively rare.

The determination of phases is (theoretically) easier with centrosymmetric structures: each phase can have only two values: 0 or π . For non-centrosymmetric structures all values between 0 and 2π are possible.

Non-centrosymmetric (P1):

In space group P1 we can place the origin whereever we want. This means that we can **assign** arbitrary values to three phases which fix the origin in a certain position. We can thus choose three strong reflections, which help us later with the triplet relationship and we **assign the values** of these phases.



It is important to choose three reflections, which can describe the reciprocal unit cell, i. e. which are not linearly dependent.

Centrosymmetric space groups:

In centrosymmetric space groups, we are also at liberty to choose arbitrarily the value for three phases. The presence of a symmetry element, however, imposes restrictions on the phases we can choose.

The inversion center is placed by definition in the origin. Since the phases have to obey the symmetry of the lattice, only values 0 or π are permitted for all phases.

In P-1 we have 8 inversion centers and thus 8 possible origins. Changing the position of the origin (only possible to another inversion center) causes a phase change in dependence of the parity of the indices hkl.



A displacement of the origin by ½ causes an inversion for all phases which have an odd index in this direction. Phases of reflections with an even index are not affected.

Centrosymmetric space groups:

A displacement of the origin by ½ causes an inversion for all phases which have an odd index in this direction. We can thus assign arbitrary phases for three reflections to fix the origin, but we must choose reflections so that in each axis there is at least one odd index.

In the table below + indicates that the phase of the reflection remains unchanged if the origin is moved to this position, – that the phase is inversed.

| Origin | ggg | ggu | gug | guu | ugg | ugu | uug | uuu | Example for fixing | a the origin: |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|--------------------------------|---------------|
| 0,0,0 | + | + | + | + | + | + | + | + | | g the engine |
| 0,0,1⁄2 | + | — | + | — | + | — | + | - | E_{aaa} are useless fo | r fixing the |
| 0,1⁄2,0 | + | + | — | - | + | + | - | - | origin. | C |
| 0,1/2,1/2 | + | — | — | + | + | — | — | + | | |
| 1⁄2,0,0 | + | + | + | + | — | _ | _ | — | $E_{\rm ugg} = -(E_{124} = -)$ | And so we |
| 1⁄2,0,1⁄2 | + | — | + | — | — | + | - | + | | already |
| 1⁄2,1⁄2,0 | + | + | — | — | — | — | + | + | $L_{gug} = + (L_{012} = +)$ | know three |
| 1/2,1/2,1/2 | + | _ | — | + | — | + | + | _ | $E_{\rm ggu} = -(E_{221} = -)$ | phases! |

I must choose a phase, which allows to define the origin, otherwise the three phases are linearly dependent. I can thus choose between ggu, guu, ugu and uuu for the last phase, but not ggg, gug, ugg or ugu.

Non-centrosymmetic space groups with symmetry elements:

If a symmetry element is present, it has to be placed in the origin. We treat the fixation of the origin in the same way as for the centrosymmetric space groups, with the only differences that the phases are not restricted to 0 and π for all reflections, but only for a subgroup.



| Origin | g0g | g0u | u0g | u0u |
|-----------|-----|-----|-----|-----|
| 0,y,0 | + | + | + | + |
| 0,y,½ | + | _ | + | — |
| ½,y,0 | + | + | _ | — |
| 1⁄2,y,1⁄2 | + | _ | _ | + |

An axe $C_2 \parallel b$ introduces an inversion symmetry in the lattice planes *h0l* (perpendicular to the C_2 axis) and the phases of E_{h0l} are restrained to 0 and π . To fix the origin, we thus have to assign the sign of two reflections E_{h0l} and to assign an arbitrary value to a reflection *hkl* with $k \neq 0$.

For example:

$$\begin{split} E_{201} &= + \quad (\phi_{201} = 0) \\ E_{101} &= + \quad (\phi_{101} = 0) \\ \phi_{111} &= \frac{1}{2} \pi \end{split}$$



Direct methods

The determination/estimation of the initial phases in 3 steps:

- Find some strong reflections with known phases
 - Triplet relationships Σ_1
 - Origin fixation (Values free to assign)
- Estimate (somehow) the phases for a subset of strong reflections (5-10 %).
 The "somehow" will be explained later
- Determine the rest of the phases (90-95%) using the tangent formula.



Direct methods

The determination/estimation of the initial phases in 3 steps:

- Find some strong reflections with known phases
 - Triplet relationships Σ_1
 - Origin fixation (Values free to assign)
- Estimate (somehow) the phases for a subset of strong reflections (5-10 %).
- Determine the rest of the phases (90-95%) using the tangent formula.



Triplet relationship for weak phases



$$E_{hkl} = \sum_{h'k'l'} E_{h'k'l'} \cdot E_{h-h',k-k',l-l'}$$
 Sayre equation

$$E_{H} \approx E_{K} \cdot E_{H-K} \qquad \varphi_{H} \approx \varphi_{K} + \varphi_{H-K}$$

In the Hauptman and Karle approximation we can use the Σ_2 triplet (instead of using the infinite sum of all triplets) to determine the phase of $E_{\rm H}$, if all three of the reflections $E_{\rm H}$, $E_{\rm K}$ et $E_{\rm H-K}$ are strong. In that case, the probability that another triplet relation with two strong reflections exist is relatively small.

The majority of the reflections, however, are not strong and the Hauptman-Karle approximation is not valid. How can we thus determine the phases of those reflections?

Uses of multiple triplets

If the approximation is less valid for weak reflections, can we substitute accuracy by numbers and use multiple triplets to determine the same phase?

Strong reflections with estimated/determined phases:

 $\begin{array}{ccc} \mathsf{E}_{001} & \mathsf{E}_{111} \\ \mathsf{E}_{021} & \mathsf{E}_{011} \\ \mathsf{E}_{101} & \mathsf{E}_{0-11} \\ \mathsf{E}_{121} & \mathsf{E}_{1-11} \\ \mathsf{E}_{0-21} \end{array}$

Weak reflections for which we search the phases:

 $E_{122} = E_{101} \cdot E_{021} \qquad E_{112} = E_{101} \cdot E_{011} \\ E_{122} = E_{001} \cdot E_{121} \qquad E_{112} = E_{121} \cdot E_{1-12} \\ E_{122} = E_{011} \cdot E_{111} \qquad E_{112} = E_{1-11} \cdot E_{021} \\ E_{112} = E_{001} \cdot E_{101} \\ E_{102} = E_{001} \cdot E_{101} \\ E_{102} = E_{121} \cdot E_{0-21} \\ E_{102} = E_{111} \cdot E_{0-11} \\ E_{102} = E_{011} \cdot E_{1-11} \end{cases}$

- How do we average the different values obtained ?
- What is the reliability of each determination (error, standard deviation) ?

Errors and the probability function for triplets

We can trivially replace the Hauptman-Karle approximation with an exact equation, when we **introduce an error** Φ_{HK} for the determination of the phase ϕ_{H} :

$$\varphi_H = \varphi_K + \varphi_{H-K} + \Phi_{HK} \qquad \begin{array}{l} \mathrm{H} = \mathrm{h, k, l} \\ \mathrm{K} = \mathrm{h', k', l'} \end{array}$$

Less trivial: What is this error Φ_{HK} ?

The validity of the Hauptman-Karle approximation Σ_2 depends on the reflections $E_{\rm H}$, $E_{\rm K}$ and $E_{\rm H-K}$ being strong. The error $\Phi_{\rm HK}$ in the determination of the unknown phase $\varphi_{\rm H}$ depends thus on the intensities of the triplet, i. e. $E_{\rm H}$, $E_{\rm K}$ and $E_{\rm H-K}$.

The distribution of the error (probability function) is given by the Mises distribution:

$$P(\Phi_{HK}) = \frac{1}{L} e^{G_{HK} \cos \Phi_{HK}}$$

Mises distribution

$$G_{HK} = \frac{2}{\sqrt{N}} \left| E_H E_K E_{H-K} \right|$$

L: Normalisation factor *N*: Number of atoms

Errors and the probability function for triplets

Mises distribution



$$G_{HK} = \frac{2}{\sqrt{N}} \left| E_H E_K E_{H-K} \right|$$

(identical atoms)

$$G_{HK} = 2\sum_{j} Z_{j}^{3} / \left(\sum_{j} Z_{j}^{2}\right)^{3/2} \left| E_{H} E_{K} E_{H-K} \right|$$
(non-identical atoms)

L: Normalisation factor N: Number of atoms H: h,k,l; K: h',k',l'

The determination of the phases is exacter (narrow error distribution) when the intensities of the three reflection is high and the overall number of atom N is small. For this reason there is an upper size limit up to which structures can be solved with direct methods (<200 atoms per asymmetric unit).

Determination of phases using multiple triplets

If we use multiple triplets, the error distribution becomes:

$$P(\Phi_{HK}) = \prod_{j} P(\Phi_{HK_{j}}) = A \prod_{j} e^{G_{HK_{j}} \cos \Phi_{HK_{j}}} = A e^{\sum_{j} G_{HK_{j}} \cos(\varphi_{H} - \varphi_{K_{j}} - \varphi_{H-K_{j}})} = A e^{A e^{A_{H}} \cos(\varphi_{H} - \varphi_{K_{j}})} = A e^{A e^{A_{H}} \cos(\varphi_{H} - \varphi_{K_{j}})} = A e^{A e^{A_{H}} \cos(\varphi_{H} - \varphi_{K_{j}})}$$

 $\alpha_{\rm H}$ is a measure of the probability that the $\alpha_{\rm H} = 0$ determined phase is correct (equivalent to $G_{\rm HK}$ for a single triplet).



$$\left(\sum_{j} G_{HK_{j}} \cos(\varphi_{K_{j}} - \varphi_{H-K_{j}})\right)^{2} + \left(\sum_{j} G_{HK_{j}} \sin(\varphi_{K_{j}} - \varphi_{H-K_{j}})\right)^{2}$$

$$\tan \beta_{H} = \frac{\sum_{j} G_{HK_{j}} \sin(\varphi_{K_{j}} - \varphi_{H-K_{j}})}{\sum_{j} G_{HK_{j}} \cos(\varphi_{K_{j}} - \varphi_{H-K_{j}})}$$

 $\beta_{\rm H}$ is the value of the determined phase (The error Φ is zero when $\varphi_{\rm H}$ = $\beta_{\rm H}$). It is calculated from multiple triplets using the tangent formula.

Direct methods

The determination/estimation of the initial phases in 3 steps:

- Find some strong reflections with known phases
 - Triplet relationships Σ_1
 - Origin fixation (Values free to assign)
- Estimate (somehow) the phases for a subset of strong reflections (5-10 %).
- Determine the rest of the phases (90-95%) using the tangent formula.



Direct methods

Triplet relationship Σ_2

valid **only** if $E_{\rm H}$, $E_{\rm K}$ and $E_{\rm H-K}$ are strong

 $\varphi_{H} \approx \varphi_{K} + \varphi_{H-K}$ H=h,k,l; K=h',k',l'

1. step: Find all triplet relationships between strong reflections. (We will call this the "initial subset".)

- 2. step: Assign (somehow) phases to these reflections, which do not violate the triplet relationship
- We have a number r of reflections
- We have a number *n* > *r* of equations between reflections (triplet relationships)
- How do we find the set of phases for these reflections which best agrees with the triplet relationships?

| Reflections | - | | Σ_2 triplets: | Phases: | |
|-------------|-----|---|-----------------------|-----------------------------|-----------------------------|
| 0 1 0 | 2 0 | 0 | $210 = 010 \cdot 200$ | $\Phi_{010} = ?$ | Φ ₂₀₀ = ? |
| 100 | 2 1 | 0 | 121 = 010 · 111 | $\Phi_{100} = ?$ | $\Phi_{210} = ?$ |
| 1 1 1 | 2 5 | 2 | 131 = 010 · 121 | Φ ₁₁₁ = ? | Φ ₂₅₂ = ? |
| 121 | 13 | 1 | 252 = 131 · 121 | Φ ₁₂₁ = ? | $\Phi_{131} = ?$ |

Obtaining phases for the initial subset

Symbolic addition (obsolete): Unknown phases are replaced with a symbolic value.

J. Karle, I. J. Karle Acta Cryst. **1966**, 21, 848; W. H. Zachariasen Acta Cryst. **1952**, 5, 68. Used in the program SIMPLE. H. Schenk (**1985**)

Phases:

 $\Phi_{252} = \Phi_{010} \cdot \Phi_{242} = \mathbf{a} \cdot \mathbf{d}$ $\Phi_{252} = \Phi_{121} \cdot \Phi_{131} = \mathbf{a}^3 \cdot \mathbf{C}^2$ $=> \mathbf{d} = \mathbf{a}^2 \cdot \mathbf{C}^2$ $Σ_2$ triplets: $210 = 010 \cdot 200$ $121 = 010 \cdot 111$ $131 = 010 \cdot 121$ $252 = 131 \cdot 121$ $252 = 010 \cdot 242$

Have fun doing this for hundreds of reflections !

Multisolution method

First introduced in the program MULTAN

G. Germain, M. M. Woolfson Acta Cryst. 1968, B24, 91.

- Phases in the initial subset were assigned arbritrary values. These values were permutated to **cover all possible combinations!**
- Starting from this subset, all other phases are calculated with the tangent formula.
- The congruity of all obtained phases is expressed with a confidence factor. (Sum of the α values for all determined phases)
- The best solution is chosen to to calculate the initial electron density map.

Permutation of values:

Centrosymmetric: Only allowed values are 0 and π .

The phases in the initial subset are permutated through all possible combinations.

Non-centrosymetric: All values between 0 and 2π are possible.

One could show that errors in the phases of 40-50° are acceptable. Often phases are thus permutated between values of $\pm \frac{1}{2}\pi$ and $\pm \frac{3}{4}\pi$. In that way, the error is never higher than 45°.

Phase refinement

Developments in the multisolution method showed that it is possible to accept larger errors in the phases, if we increase the number of reflections in the initial subset. At the extreme of this approach, it was shown that **the refinement of arbitrary phases converges into reasonable values** when the number of reflections in the initial subset is high enough.

How to proceed:

- 1. Select an initial subset of strong reflections
- 2. Assign arbitrary values to all phases
- 3. Refine these values (several methods possible) to minimise : $\phi_H \phi_K \phi_{H-K} = \Phi$
- 4. Calculate a confidence factor for this solution
- 5. Repeat this *x* times and chose the best set for the expansion of the initial subset to all reflections with the tangent formula and for the calculation of the first electron density map (initial solution).

Problem: Free refinement using triplet relations refines in the **uranium atom solution**: all the phases φ are zero (and thus all errors Φ) and the electron density is located in one point in the origin.

How to avoid the uranium atom solution?

Negative quartets

Schenk introduced 1973 the quartet relationship:

$$E_{H} \approx E_{K} \cdot E_{L} \cdot E_{H-K-L} \Leftrightarrow \varphi_{H} + \varphi_{K} + \varphi_{L} + \varphi_{-H-K-L} = \Phi_{HKL}$$

• As for the triplet relationship, $E_{\rm H}$, $E_{\rm K}$, $E_{\rm L}$, $E_{\rm H+K+L}$ have to be all strong so that the error in determined phase $\Phi_{\rm HKL}$ is zero (small).

The quartet relationship can be described as the sum of two triplets (3 possibilities):

$$\varphi_{H} + \varphi_{K} - \varphi_{H+K} = \Phi_{11}; \quad \varphi_{H+K} + \varphi_{L} + \varphi_{-H-K-L} = \Phi_{12}; \quad \Phi_{HKL} = \Phi_{11} + \Phi_{12}$$

$$\varphi_{H} + \varphi_{L} - \varphi_{H+L} = \Phi_{21}; \quad \varphi_{H+L} + \varphi_{K} + \varphi_{-H-K-L} = \Phi_{22}; \quad \Phi_{HKL} = \Phi_{21} + \Phi_{22}$$

$$\varphi_{K} + \varphi_{L} - \varphi_{K+L} = \Phi_{31}; \quad \varphi_{K+L} + \varphi_{H} + \varphi_{-H-K-L} = \Phi_{32}; \quad \Phi_{HKL} = \Phi_{31} + \Phi_{32}$$

Attention: H,K,L are here short for H=(h,k,l); K=(h',k',l'); ...

- Thus the error Φ_{HKL} is zero, when all the errors Φ_{ij} are zero. Which means that the "cross intensities" E_{H+K} , E_{K+L} , E_{H+L} have to be strong too.
- The information gained from the quartets are weak. The possibility to find the correct phase is proportional to $1/N (1/\sqrt{N}$ for triplet relationships; *N*: number of atoms).
- Empirically it was found that $\Phi_{HKL}=\pi$ if the cross intensities E_{H+K} , E_{K+L} , E_{H+L} are all weak. We call this situation a **negative quartet**: $\phi_H + \phi_K + \phi_L + \phi_{-H-K-L} \pi = \Phi_{HKL}$
- While we do not gain any phase information from the negative quartets, adding them to the phase refinement avoids that the refinement converges in the uranium atom solution where all phases φ refine to 0 and Φ_{HKL}=0.
Steps in obtaining the initial solution

1. Normalisation: Calculation of E values

2. Triplet selection:

- Selection of reflections with high |E| (strong reflections). We need appr. 10x the number of atoms in the unit cell)
- Found all triplet relationships $E_{\rm H}$, $E_{\rm K}$, $E_{\rm H-K}$ in this subset

3. Find known phases:

- Origin fixation + enantiomorph
- Triplets Σ_1 (known phases in centrosymmetric structures)
- Additional reflections where the phases are known

4. Assign/calculate the phases for the initial subset:

- Symbolic addition
- Multisolution methods
- Phase refinement

Strong reflections

1. Calculate E values and find strong reflections



2. Find phases with known values



| Strong reflections: | | | | | | Σ_1 triplets: | | | | | | Σ_2 triplets : | | | | | | |
|---------------------|-------------|---------|---------|---------|---------|----------------------|---------|---------|---------|---------|-----------------|-----------------------|-----------------|------|----------|-----------|---------|--------------------------------|
| 0 | 1 0 + | | 1 | 1 1 | | | | 200 | = 1 | 00 | • 100 |) = | + | | 21 | 0 = | 010 |) • 200 |
| 1 0 0 - 1 2 1 | | | | | | | | | | | | | 121 = 010 · 111 | | | | | |
| 2 | 0 0 + | | 2 | 10 | | Origin fixation: 311 | | | | | | 1 = | = 200 · 111 | | | | | |
| 3 | 1 1 - | | 2 | 1 1 | | | | | | | 211 = 100 · 111 | | | | | | | |
| | | | | | | | | 3 | 811 | | | | | | | | | |
| 3. Eva | aluate each | | mbi | inat | ion | usi | ng | the | exi | stir | ng tr | iplo | et ro | elat | ion | shi | ps | Permutation of all possible |
| 01 | 0 | Ð | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | combinations (multisolution |
| 10 | 00 | Θ | - | - | - | - | _ | - | 5 | - | _ | - | | - | - | - | - | method) |
| 20 | 00 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | I |
| 31 | 1 | - | - | - | - | - | - | - | - | - | _ | - | - | - | - | - | - | |
| 12 | 21 | (+) | + | + | + | + | + | + | + | - | — | 7 | | - | - | - | - | |
| 21 | 0 | + | + | + | + | - | - | - | - | + | + | + | + | - | - | - | - | |
| 21 | 1 | (+) | + | - | - | + | + | — | - | + | + | - | - | + | + | - | — | |
| 11 | 1 | (+) | - | + | - | + | - | + | - | + | - | + | - | + | - | Ŧ | - | |
| 12 | 21=010-111 | \odot | 8 | \odot | 8 | \odot | 8 | \odot | 8 | 8 | 0 | 8 | \odot | 8 | © | 8 | | |
| 21 | 0=010-200 | \odot | \odot | \odot | \odot | 8 | 8 | 8 | 8 | \odot | \odot | (| \odot | 8 | 8 | 8 | 8 | |
| 31 | 1=200-111 | 8 | \odot | 8 | \odot | 8 | \odot | 8 | \odot | 8 | \odot | 8 | \odot | 8 | \odot | \otimes | \odot | |
| 21 | 1=100-111 | 8 | \odot | \odot | 8 | 8 | \odot | \odot | 8 | 8 | \odot | \odot | 8 | 8 | \odot | \odot | 8 | |



4. Find remaining reflections using combinations of triplets

$$E_{H} \approx E_{K} \cdot E_{H-K}$$

$$\varphi_{H} \approx \varphi_{K} + \varphi_{H-K}$$

H=h,k,l; K=h',k',l'

 $321 = 210 \cdot 111 = 411 = 211 \cdot 200 = +$ $321 = 311 \cdot 010 = -$

 $321 = 200 \cdot 121 = 411 = 311 \cdot 100 = +$

 $221 = 100 \cdot 121 = +$ $221 = 211 \cdot 010 = +$

In reality: more complex

In real cases, there is no single, analytical solution and not all triplets yield the same solution. Each calculated phase is just weighted with the probable error of its determination, which can be visualized using a vector addition, where each vector represents a phase value obtained from a triplet relationship.

$$E_{H} \approx E_{K} \cdot E_{H-K}$$

$$\tan \beta_{H} = \frac{\sum_{j} G_{HK_{j}} \sin(\varphi_{K_{j}} - \varphi_{H-K_{j}})}{\sum_{j} G_{HK_{j}} \cos(\varphi_{K_{j}} - \varphi_{H-K_{j}})}$$

The tangent formula allows us to combine multiple triplets and to calculate the resulting phase β_H and its probability α_H .



Direct methods

The determination/estimation of the initial phases in 3 steps:

- Find some strong reflections with known phases ۲
 - Triplet relationships Σ_1
 - Origin fixation (Values free to assign)
- Estimate the phases for a subset of strong reflections (5-10 %). ٠
 - Symbolic addition
 - Multisolution methods
 - Phase refinement
- Determine the rest of the phases (90-95%) using the tangent ulletformula.

$$\sum_{1} : E_{H} = E_{K} \cdot E_{K}$$
phases known
phases estimated
$$\sum_{2} : E_{H} = E_{K} \cdot E_{H-K}$$
tan
$$\beta_{H} = \frac{\sum_{j} G_{HK_{j}} \sin(\varphi_{K_{j}} - \varphi_{H-K_{j}})}{\sum_{j} G_{HK_{j}} \cos(\varphi_{K_{j}} - \varphi_{H-K_{j}})}$$
strong
Reflection intensity
Weak

The "phase annealing" in SHELX

Instead of the standard phase refinement (generate an arbitrary set of phases, refine them and repeat this process x times), SHELXS uses a different method:

- 1. The program searches *n* reflections of high *E*.
- 2. Approximately half of them are chosen (*u*, PHAN) and their phases are refined starting from random start values.
- 3. This phases are refined using "annealing" (malléabilisation): The phases were allowed to do random changes, the magnitude of which depends on a temperature factor which is slowly reduced. This allows the refinement to escape a local minimum.

How does that work ?

A simple refinement might, depending on the starting conditions, end in a false (local) minimum.



A Monte-Carlo refinement generates a large number of starting conditions, hoping that a least one refines into the global minimum.

 In an annealing procedure, the solution can move "uphill" according to a temperature factor.

A simple refinement might, depending on the starting conditions, end in a false (local) minimum.





- In an annealing procedure, the solution can move "uphill" according to a temperature factor.
- The temperature factor is slowly lowered, giving enough time for the solution to escape false minima.



A simple refinement might, depending on the starting conditions, end in a false (local) minimum.





- In an annealing procedure, the solution can move "uphill" according to a temperature factor.
- The temperature factor is slowly lowered, giving enough time for the solution to escape false minima.



A simple refinement might, depending on the starting conditions, end in a false (local) minimum.





- In an annealing procedure, the solution can move "uphill" according to a temperature factor.
- The temperature factor is slowly lowered, giving enough time for the solution to escape false minima.

A simple refinement might, depending on the starting conditions, end in a false (local) minimum.





- In an annealing procedure, the solution can move "uphill" according to a temperature factor.
- The temperature factor is slowly lowered, giving enough time for the solution to escape false minima.
- At the end the temperature factor is reduced to zero and the solution should collapse in the global minimum.

The "phase annealing" in SHELX

Instead of the standard phase refinement (generate an arbitrary set of phases, refine them and repeat this process x times), SHELXS uses a different method:

- 1. The program searches *n* reflections of high *E*.
- 2. Approximately half of them are chosen (*u*, PHAN) and their phases are refined starting from random start values.
- 3. This phases are refined using the "annealing" (malléabilisation): The phases were allowed to do random changes, the magnitude of which depends on a temperature factor which is slowly reduced. This allows the refinement to escape a local minimum.
- 4. The *u* phases obtained are combined with the remaining strong reflections *x*=*n*-*u*. The phases *x* are assigned again arbitrary values and all phases (*n*) are refined normally (without annealing), using negative quartets to avoid the uranium atom solution. At the end several confidence factors are calculated to evaluate the obtained solution.
- 5. Step 4 is repeated several times (TREF, default 250), the refinement starting always with the same phases for *u* and new arbitrary phases for *x* reflections. The best solution is retained to calculate phases for the whole data set.

209 Reflections and 2238. unique TPR for phase annealing 359 Phases refined using 8388. unique TPR 579 Reflections and 16812. unique TPR for R(alpha)

2128 Unique negative quartets found, 2128 used for phase refinement

The "phase annealing" in SHELX





Structure invariants

Several values do not depend on the placement of the origin, but only on the structure. They are called **structure invariants**:

- 1. $F_{000} = \sum Z_i$: number of electrons in the unit cell
- **2.** $F_{\rm H}F_{\rm -H} = |F|^2$: Intensity, does not contain the phases
- **3.** $F_{-H} F_{K} F_{H-K}$: triplet invariant $\phi = \phi_{-H} + \phi_{K} + \phi_{H-K}$
- **4.** $F_{-H} F_{K} F_{L} F_{H-K-L}$: quartet invariant $\varphi = \varphi_{-H} + \varphi_{K} + \varphi_{L} + \varphi_{H-K-L}$, analog quintet etc.

Structure semi-invariants do not change with the displacement of the origin, as long as it displaced to a point of identical symmetry. For example, when the origin is placed from one inversion center to another, E_{2h2k2l} does not change its phase.



126 Unique NQR employed in phase annealing

 Phase annealing cycle:
 1
 Beta =
 0.06279

 Ralpha
 0.211
 0.282
 0.609
 0.125
 0.417
 0.543
 0.509
 0.343
 0.093
 0.102
 0.280
 0.179
 0.246
 0.672
 0.097
 0.065
 0.237
 0.068
 0.223
 0.067

 Ngual
 0.194
 0.490
 0.038
 0.662
 0.187
 0.036-0.188-0.324
 0.433
 0.386-0.105-0.204-0.224
 0.205
 0.295
 0.579-0.602
 0.711-0.138
 0.624

 Mabs
 0.787
 0.729
 0.588
 0.913
 0.663
 0.608
 0.616
 0.682
 0.979
 0.924
 0.729
 0.825
 0.764
 0.565
 0.942
 1.047
 0.774
 1.046

Random values assigned for the phases not included in phase annealing

```
Ralpha Nqual Sigma-1 M(abs) CFOM Seminvariants
Try
1420309. 0.200 -0.349
                    0.065
                           0.859
                                 0.562
 810089. 0.200 -0.349
                    0.065
                           0.859
                                 0.562
1953293.0.331
              0.412
                    0.106
                          0.734
                                 2.105
1377857. 0.135 0.129
                    0.599
                           0.936
                                 1.300
 597829. 0.14 0.177
                    0.449
                          0.918
                                 1,416
                                       ----
                    0.226
                                 0.573
 891993. 0.326 2.453
                          0.71
                    0.599
                           0.936
 265661. 0.135
              0.129
                                 1.300
                                2.098
1328305. 0.318 0.384 -0.077
                           0.738
                                             -+++ +--++
                                                      ----
 350069. 0.138
              0.110
                     0.580
                           0.934
                                1.261
                                                 --+++
                                                       ++--+
              0.129
                    0.599
                           0.936
                                1.300
1750345. 0.135
                                                         -+ -++
                    0.297 0.713
                                 0.626
 363117. 0.332 -0.408
                                                   +++ -+--f
                                                            +-++- +-+--
                           0 912
                                0.783
1815585. 0.174 -0.170
                    0.355
                                                   +++ -+++- ++
                           0.752 1.631
 689317. 0.285 0.210
                    0.274
1349433. 0.289 -0.108
                    0.214
                          0.746 0.997
                                                           [...]
                                                                         Identical solutions
 CFOM Range
                Frequency
 0.000 - 0.020
                        0
                           "Number" of the solution. To
 0.020 - 0.040
                        0
                                                                     Different solutions of
 0.040 - 0.060
                        0
                           reproduce exactly this solution,
                                                                     the same "quality".
 0.060 - 0.080
                        0
                           we can repeat the structure
                                                                     Compared the
 0.080 - 0.100
                        0
                           solution using the command
 0.100 - 0.120
                        0
                                                                     different signs for the
 0.120 - 0.140
                           TREF -1420309.
                        5
                                                                     semi-invariants.
                       95
 0.140 - 0.160
 0.160 - 0.180
                        0
                                     Bad solution
 0.180 - 0.200
                        1
 [...]
256. Phase sets refined - best is code
                                                            with CFOM =
                                                 861549.
                                                                            0.1553
```

Confidence factors (structure solution)

1. R_{α} is the deviation of estimated and real α (triplet deviation). Both, correct and wrong solutions can have a small R_{α} , but only wrong solutions have a high R_{α} .

2. Confidence factor for **negative quartets**. For correct phases the value is -1, incorrect ones have positive values.

3. **Triplet congruity**. Correct phases have MABS=1.

4. CFOM = combined figure of merit: average of all confidence factors

$$\alpha_{H} = \frac{2}{\sqrt{N}} E_{H} \sum_{K} E_{K} E_{H-K}$$

Probability that the phase determined by a triplet (or a combination of triplets) is correct.

 α_{H}^{e} : α_{H} estimated (if all triplets give the same phase result)

$$\eta_{H} = \frac{1}{N} E_{H} \sum_{K,L} E_{-K} E_{-L} E_{H+K+L}$$

Probability to determine the correct phase from a quartet relationship.

 $R_{\alpha} = \frac{\sum_{h} |\alpha_{h} - \alpha_{h}^{e}|}{\sum_{h} \alpha_{h}^{e}}$

$$NQUAL = \frac{\sum_{h} \alpha_{h} \eta_{h}}{\sum_{h} |\alpha_{h}| |\eta_{h}|}$$

$$MABS = \frac{\sum_{h} \alpha_{h}}{\sum_{h} \alpha_{h}^{e}}$$

Peak list optimization

| Observed E . | GT. 1.200 | 1.300 | 1.400 | 1.500 | 1.600 | 1.700 | 1.800 | 1.900 | 2.000 | 2.100 |
|--------------|-----------|-------|-------|-------|--------|-------|--------|-------|-------|-------|
| Number | 1450 | 1243 | 1082 | 913 | 769 | 661 | 534 | 444 | 355 | 287 |
| | | | | | | | | | | |
| ESEL Emin | 1.200 | Emax | 5.000 | Del | U 0.00 |)5 1 | renorm | 0.700 | axi | is O |

Tangent expanded to 1450 out of 1450 E greater than 1.200

E-Fourier for BHG109 in P2(1)/c Heavy-atom assignments:

| | x | У | Z | s.o.f. | Height |
|-----|--------|--------|--------|--------|--------|
| TI1 | 0.7553 | 0.1017 | 0.7720 | 1.0000 | 533.3 |
| SI2 | 0.5823 | 0.2565 | 0.7807 | 1.0000 | 288.9 |

XS already does some refinement cycles using only the strongest E reflections.

Peak list optimization RE = 0.151 for 34 surviving atoms and 1450 E-values

E-Fourier for BHG109 in P2(1)/c

Peak list optimization RE = 0.147 for 34 surviving atoms and 1450 E-values

```
E-Fourier for BHG109 in P2(1)/c
Maximum = 512.38, minimum = -89.73
```

1. Bad solution, unstable refinement

• Find the rest of the structure, atom by atom, during the refinement: Delete all atoms other than the heavy atoms. Increase PLAN to 50 and refine. Use constraints such as AFIX 56 and 66 for cycles (Cp, phenyl).

Find the rest of the structure without refinement:

Delete all incorrect atoms (eventually everything other than the heavy atoms). **Put L.S. to 0** (no refinement, only calculation of the electron densitiy map) and PLAN to 50. Refine the structure and assign newly found, correct atoms. Repeat this until (nearly) all atoms are found. Than try a refinement with L.S. 4.

2. Centrosymmetric structure:

No solution or bizarre disorder with two superimposed molecules. Often caused by a slightly misplaced origin. We can solve the structure in a non-centrosymmetric space group and transform it later back into a centrosymmetric one. (Attention: For the back-transformation, we have to displace the origin.) Three possibilities to do this:

- Transformation to a non-centrosymmetric space group (change LATT x to LATT –x). Solve, translate (you have to know what to do), change back to LATT x.
- 2. Restart XPREP, chose a non-centrosymmetric space group. Solution, initial refinement and transform into a centrosymmetric space group (you have to know what to do).
- Redo the centrosymmetric solution, but introduce ESEL –1 in the .ins-file. Displace afterwards the origin to the correct position and delete additional molecules (CENT – PUSH – FUSE). (This is the fastest way, but does not always work)
- 4. Use the Patterson method, dual space refinement or charge flipping.

The transformation from a non-centrosymmetric into a centrosymmetric space group (1. & 2.) can be done by hand (XP:PUSH) or by using PLATON.

3. Change the parameters for direct methods:

a) Increase TREF to TREF 5000 (or TREF 100000).

Look at the solutions: If the program finds several, different solutions with a comparable CFOM value, try these solutions with TREF –xxxxx (xxxxxx numbre of the solution).

b) If the number of triplets (TPR) is < 20 times the number of reflections used, you can try to reduce the minimum E to accept a reflection in the initial set. Do this in steps of 0.1 (ESEL 1, ESEL 0.9, ESEL 0.8...)

209 Reflections and 2238. unique TPR for phase annealing 359 Phases refined using 8388. unique TPR 579 Reflections and 16812. unique TPR for R(alpha)

4. Still problems ?

- a) Use another space group, eventually go down to P1.
- b) Verify that all data is correct (unit cell etc.)
- c) Try a data set without absorption correction
- d) Use the Patterson function
- e) Use another program (XT, SIR, DIRDIF, etc.)
- f) Test for twinning.
- g) Crying and swearing does not solve a structure, but it helps nevertheless. ③

Structure solution

Patterson methods (vector methods)

Patterson method determine directly the position of several atoms in the unit cell. From the position of these methods, in particular if they are heavy metals with many electrons, the Fourier transformation yields an initial, relatively reasonable set of phases.

Direct methods

Direct methods try to establish the phases directly from the intensities of the reflections.

Dual space methods (Shake 'n Bake, SHELXD (XM), SIR)
 Alternate refinement in reciprocal (phase) space and real space (electron density map)

Dual space methods



Dual space methods

Differences to direct methods:

- Suitable for structures with more than 200 atoms (protein crystallography)
- Different solution approach (complementary)

Initial phases:

- From randomly distributed atoms (SnB, SHELXD (XM), SIR)
- From Patterson data (SHELXD, SIR)

Phase refinement:

- By refining the tangent formula (SnB, SHELXD (XM), SIR)
- By parameter shift (SnB, SIR)
- Using all reflections I > 3σ (SnB) or only strong ones (SHELXD)

Real space refinement:

- No real refinement. Only the strongest maxima are selected for atom positions
- SHELXD (XM) : Randomly deleting 30% of the highest maxima *improves* the solution (random omit algorithm)

SnB : http://www.hwi.buffalo.edu/snb/Tutorial/Introduction.htm SHELXD : http://shelx.uni-ac.gwdg.de/SHELX/shelx_de.pdf SIR : http://www.ic.cnr.it/sir2004manual.pdf

Solution des structures

Patterson methods (vector methods)

Patterson method determine directly the position of several atoms in the unit cell. From the position of these methods, in particular if they are heavy metals with many electrons, the Fourier transformation yields an initial, relatively reasonable set of phases.

Direct methods

Direct methods try to establish the phases directly from the intensities of the reflections.

 Dual space methods (Shake 'n Bake, SHELXD, SIR) Alternate refinement in reciprocal (phase) space and real space (electron density map)

Charge flipping methods

Charge flipping methods start with random phases and iterate by inverting the low density region of the electron density map.

Charge flipping methods

- Shiono and Woolfson, 1992: Setting areas of small electron density to 0 improves phase refinement and enables determination of phases from a random starting set.
- Charge Flipping: Introduced by Oszlnáyi & Sütő in 2004 Starting from random phases, the region of low electron density is inverted. Thus switching back between real and reciprocal space, correct phases emerge to match the observed intensities. Since regions of low electron density do not contain much structural information, inversion of this regions provides small perturbations while the overall electron density remains largely unaffected.

Charge flipping does not rely on atomicity, but only on the fact that a great part of the electron density map is actually zero.

It is not really understood why this works. \bigcirc

Charge flipping : How does it work?

Gridpoints are defined for the electron density

- We start with a random set of phases, which is used to calculate an electron density map
- For grid points with electron density lower than a certain treshhold δ , the electron density is inverted.
- Back FT yields the phases of the partially flipped density, which are used to restart the cycle.



Charge flipping



Taken from : Structure Resolution by the Charge Flipping Algorithm, Gervais Chapuis

The charge flipping algorith typically has a long initiation time, fast convergence, followed by slow optimization to the optimum solution.

Due to the two Fourier transformations required for each iteration cycle, charge flipping is relatively « computer-heavy».

Charge flipping requires only one parameter, the treshhold value δ . Its value, however, is crucial. Most programs use a dynamical treshhold, which is slowly reduced during refinement.

Online demonstration at : http://escher.epfl.ch/flip/

Combining lessons learned - SHELXT

"Intrinsic phasing" (SheIXT/XT) is a derivative of dual space refinement, but with some very useful twists.

Most importantly: SHELXT (XT) solves in P1 and derives the space group from the calculated **phases**. All other methods either rely on the observed symmetries in the reflections intensities or in the symmetries of the calculated electron density map.

George M. Sheldrick, SHELXT – Integrated space-group and crystalstructure determination, *Acta Cryst.* **2015**, *A71*, 3–8.

SHELXT – General procedure

- 1. Determination of the Laue group (program input, ins-file)
- 2. Averaging of equivalent reflections and expansion to P1.
- 3. Solving in P1 by dual space refinement
- 4. Space group determination using the **phases** and averaging of phases and electron density map.
- 5. Refinement/assignment of atoms to the electron density map

How does it work in detail?

SHELXT – Solving the structure

Assigning intial phases:

Dual space refinements typically start with random phases. XT will perform a Patterson solution and use the displacement method to get an initial atom list and thus starting phases.



The Patterson method is only used to generate the initial start phases. This can be omitted (although it is not recommended.)

For light-atom structures the Patterson solution is most likely of low quality... but not worse than random phases.

SHELXT – Dual Space Refinement



In previous dual space refinements (SHELXD/XM) it turned out to be advantageous to ignore a certain percentage of the observed maxima. This introduces a degree of randomness in the procedure and provides means to escape local minima. (This is roughly equivalent to the inversion of charge density in the charge flipping algorithm.)

SHELXT – Dual Space Refinement



Graphics adapted from Prof. Sheldrick
SHELXT – Space group determination

• After determination of the dual space refinement, the space group is determined from the obtained **phases**.

Based on work of:

- Burla, Carrozzini, Cascarano, Giacovazzo, Polidori *J. Appl. Cryst.* **2000**, *33*, 307. "Solving crystal structures in P1: an automated procedure for finding an allowed origin in the correct space group"
- Palatinus, van der Lee J. Appl. Cryst. 2008, 41, 975. "Symmetry determination following structure solution in P1"
- Only the phases of reflections with the best agreement between calculated and observed intensity are used.
- Each space group imposes a certain symmetry on the phases. Compliance to this symmetry can be expressed by a confidence factor.
 - What are symmetry conditions in reciprocal space, i. e. for reflections?
 - How to define the confidence factor?
 - How to find the correct origin?

Every symmetry operation can be expressed in the form $r' = R \cdot r + t$

- **r** is the location of the atom, r = (x, y, z)
- **r** is the symmetry-equivalent location
- **R** is symmetry matrix for rotation, inversion, or reflection at the origin
- t is the translation vector

Example: Mirror plane *m* in *P*2₁/m
Symmetry operator: *x*, ¹/₂ - *y*, *z*

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ 1/2 - y \\ 2 \\ z \end{pmatrix}$$
Example: Screw axis 3₁ in *P*3₁
Symmetry operator: -*y*, *x*-*y*, *z*+1/3
$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \\ -3 \end{pmatrix} = \begin{pmatrix} -y \\ x - y \\ z + \frac{1}{3} \end{pmatrix}$$

Please note: the translation vector *t* contains not only the translation part of screw axes or glide planes, but also the translation necessary if the symmetry element is not located at the origin. I. e. the mirror plane *m* in *P*21/m is located at y = 0.25.

For inversion symmetry, we readily know that h, k, l and -h, -k, -l are symmetry equivalent. But what is the symmetry-equivalent reflection with regard to a mirror plane at y = 0.25 or a 3_1 screw axis?

Symmetry-equivalent reflections are found using the *transpose* of the *rotational element* of the symmetry operator *R*.

$$\begin{aligned} \mathbf{reciprocal space} \\ \mathbf{r'} = \mathbf{R} \cdot \mathbf{r} + t \\ \begin{pmatrix} \mathbf{r'} \\ \mathbf{y'} \\ \mathbf{z'} \end{pmatrix} = \begin{pmatrix} \mathbf{r_{11}} & \mathbf{r_{12}} & \mathbf{r_{13}} \\ \mathbf{r_{21}} & \mathbf{r_{22}} & \mathbf{r_{23}} \\ \mathbf{r_{31}} & \mathbf{r_{32}} & \mathbf{r_{33}} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad \begin{pmatrix} \mathbf{h'} \\ \mathbf{k'} \\ \mathbf{l'} \end{pmatrix} = \begin{pmatrix} \mathbf{r_{11}} & \mathbf{r_{31}} & \mathbf{r_{31}} \\ \mathbf{r_{12}} & \mathbf{r_{22}} & \mathbf{r_{32}} \\ \mathbf{r_{13}} & \mathbf{r_{23}} & \mathbf{r_{33}} \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \mathbf{k} \\ \mathbf{l'} \end{pmatrix} \\ \end{aligned}$$

The difference between the phase ϕ of any reflection (*h*, *k*, *l*) and the phase ϕ ' of its symmetry equivalent (*h*', *k*', *l*) = $R^{T} \cdot (h, k, l)$ depends on the translation vector *t*. $\phi' = \phi - 2\pi (h \cdot t_1 + k \cdot t_2 + l \cdot t_3)$

Example: Mirror plane *m* in $P2_1/m$ Symmetry operator: *x*, $\frac{1}{2}$ -*y*, *z*

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 1/2 \\ /2 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ 1/2 - y \\ z \end{pmatrix}$$

$$\mathbf{r}' = \mathbf{R} \cdot \mathbf{r} + \mathbf{t}$$

Reciprocal space:

$$\begin{pmatrix} h' \\ k' \\ l' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h \\ k \\ l \end{pmatrix} = \begin{pmatrix} h \\ -k \\ l \end{pmatrix} \qquad \phi' = \phi - 2\pi(h \cdot 0 + k \cdot \frac{1}{2} + l \cdot 0) = \phi - \pi \cdot k$$

- Thus, the symmetry equivalent to F_{222} is F_{2-22} . The phase difference between these reflections is $\Delta \phi = 2\pi = 0$.
- The symmetry equivalent to F_{111} is F_{1-11} . The phase difference between these reflections is $\Delta \phi = \pi$.

Example: Screw axis 3_1 in $P3_1$ Symmetry operator: -y, x - y, z+1/3

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \\ /3 \end{pmatrix} = \begin{pmatrix} -y \\ x - y \\ z + \frac{1}{3} \end{pmatrix}$$

$$\mathbf{r}' = \mathbf{R} \cdot \mathbf{r} + \mathbf{t}$$

Reciprocal space:

$$\begin{pmatrix} h' \\ k' \\ l' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h \\ k \\ l \end{pmatrix} = \begin{pmatrix} k \\ -h-k \\ l \end{pmatrix} \qquad \phi' = \phi - \frac{2}{3}\pi \cdot l$$

- Thus, the symmetry equivalent to F_{333} is F_{3-63} . The phase difference between these reflections is $\Delta \phi = 2\pi = 0$.
- The symmetry equivalent to F_{321} is F_{2-51} . The phase difference between these reflections is $\Delta \phi = -2\pi/3$.
- In the presence of a 3₁ axis, the phase differences between reflections F_{hkl} and F_{k(-h-k)l} will be 0 for I=3n.

Another look at systematic absences

Systematic absences are caused by destructive interference of an atom with its symmetry equivalent for certain lattice plans. For a 3_1 axis parallel to *c* we would expect a reflection condition of I = 3n for all 00/ reflections.

The same conclusion can be reached (not surprisingly) from considering the phases. As seen on the slide before, a 3_1 axis enforces a phase difference of $\Delta \phi = 2/3 \cdot \pi \cdot I$ between reflections F(*h*, *k*, *I*) and F(*k*, -h-k, *I*).

If a reflection is *its own symmetry equivalent*, then either the phase difference must be zero or its intensity must be zero, since a reflection cannot have two phases.

If F(*h*, *k*, *l*) = F(*k*, -h-k, *l*) (only possible for h = k = 0, i. e. 00/ reflections), then only for l = 3n we have $\Delta \phi = 2/3 \cdot \pi \cdot l = n \cdot 2\pi$. For all l <> 3n, $\Delta \phi <> 0$ and destructive interference results.

On the other hand, the mirror plane in P2₁/*m* caused a phase difference of $\Delta \phi = \pi \cdot k$ between F(*h*, *k*, *l*) and F(*h*, *-k*, *l*). These reflections are identical only for k = -k = 0 and in this case the phase difference is always 0. The mirror plane does thus not introduce systematic absences.

SHELXT – Figure of merit

The concordance of the reflection data with the symmetry of a given space group is given by one single figure of merit α .

 α is the sum of the squared differences between the phases of symmetry-equivalent reflections (*q*), normalized so that $\alpha = 1$ indicates random phases.

$$q = (\phi' - \phi + 2\pi \cdot r^* \cdot t) \mod 2\pi$$
$$\alpha = \frac{3}{\pi^2} \frac{\sum_{\substack{hkl \ kl}} |F_{hkl}F_{h'k'l'}| q^2}{\sum_{\substack{hkl \ kl}} |F_{hkl}F_{h'k'l'}|}$$

SHELXT – Origin refinement/placement

The phase difference is dependent on the placement of the origin and

$$\phi' = \phi - 2\pi (h \cdot t_1 + k \cdot t_2 + l \cdot t_3)$$

is only true, if the origin is placed correctly. An incorrect placement of the origin by Δr will add an error of $2\pi \cdot (r^* - r^*) \cdot \Delta r$ to $\Delta \phi$.

For each space group, α is thus minimized by displacing the origin. In centrosymmetric space groups, this can be done analytically. For all other space groups, a (slow) grid-based search is conducted to find the origin shift which minimizes α .

XT automatically calculates the Flack-*x* parameter (Parsons' quotient) and inverts the structure if necessary.

SHELXT – Switching to the correct space group

XT will attempt further solutions for space groups below a certain α -treshhold (can be influenced using command line parameters):

- Averaging the obtained phases using the space group symmetry
- Assigning atoms to maxima
- Further dual-space refinement

Atom assignment:

- The programs attempts to use the elements provided in SFAC, but ignores the quantities in UNIT
- To bring the (meaningless) original electron count to scale, the program looks for "typical organic junk" (Sheldrick) and scales it to carbon or, alternatively, does the same with oxyanions of heavy metals (if present in SFAC).
- If this fails, the highest electron count is assigned to the heaviest element in SFAC.
- The rest of the electron density is assigned based on electron count.
- Some simple chemical rules are applied to avoid non-sense assignments. If a heavy atom is clearly present, but not specified, the program assigns either Br or I based on electron count.

SHELXT – Inspecting the results

With standard settings, XT uses the first plausible solution, which works in most cases. Sometimes, here a centrosymmetric sub-structure of the heavy atoms in a non-centrosymmetric space group, several proposals with similar α -values can be obtained (use the –a command parameter, automatically done if a heavy atom is present in SFAC). Here, the Flack-x parameter indicates the correct solution. (If the space group were indeed centrosymmetric, all solutions would have Flack-*x*=0.5).

| R1 | Rweak | Alpha | Space group | Flack_ | x File | Formula |
|-------|-------|-------|-------------|--------|--------|-----------------------|
| 0.634 | 0.338 | 0.051 | P4/mmm | | t240_a | C35 N6 Pt96 I |
| 0.203 | 0.004 | 0.069 | P4(2)/mmc | | t240_b | C42 N12 F6 C18 Pt |
| 0.637 | 0.398 | 0.070 | P4(2)/mnm | | t240_c | C22 N8 Pt46 I |
| 0.208 | 0.007 | 0.071 | P4/mnc | | t240_d | C38 N26 F14 C14 Pt |
| 0.552 | 0.137 | 0.035 | P-4m2 | 0.48 | t240_e | C59 N2 F8 C110 Pt67 |
| 0.165 | 0.004 | 0.051 | P4mm | 0.50 | t240_f | C83 N8 F25 C116 Pt2 |
| 0.168 | 0.004 | 0.053 | P422 | 0.50 | t240_g | C39 N16 F14 C15 Pt |
| 0.181 | 0.004 | 0.053 | P-42m | 0.49 | t240h | C46 N16 F44 C112 Pt I |
| 0.146 | 0.003 | 0.054 | P4(2)mc | 0.49 | t240_i | C41 N5 F5 C12 Pt |
| 0.123 | 0.003 | 0.055 | P4(2)2(1)2 | 0.09 | t240_j | C30 N6 F14 C12 Pt |
| 0.139 | 0.003 | 0.056 | P-42(1)c | 0.47 | t240_k | C26 N16 F12 Pt |
| 0.157 | 0.004 | 0.072 | P-42(1)m | 0.49 | t240_1 | C41 N14 F12 C13 Pt |
| 0.515 | 0.120 | 0.072 | P-42c | 0.46 | t240_m | C16 Pt68 I |
| 0.529 | 0.146 | 0.073 | P4 (2) 22 | 0.47 | t240_n | C20 N4 Pt58 I |
| 0.166 | 0.004 | 0.073 | P42(1)2 | 0.44 | t240_o | C48 N21 F5 C14 Pt |
| 0.142 | 0.006 | 0.074 | P4 (2) nm | 0.50 | t240_p | C42 N22 F12 C12 Pt |
| 0.192 | 0.007 | 0.075 | P4nc | 0.50 | t240_q | C41 N22 F14 C19 I |
| 0.164 | 0.007 | 0.076 | P-4n2 | 0.49 | t240_r | C42 N22 F14 C14 Pt |
| | | | | | | |

If this still does not help, we have to manually inspect all the solutions proposed (see column File).

SHELXT and OLEX

SHELXT will generate *code_x.res* and *code_x.hkl* files for all space groups successfully evaluated, with x = a, b, c... Rather unfortunately, OLEX will hide all these files in the hidden directory .olex/temp. There you will also find the SHELXT output file *code.lxt*. If a solution with ShelXT does not work out, check this directory for the output file and investigate eventually other solutions.

While ShelXT can be run without determining the space group, it still requires a *code*.ins file to read the correct Laue group. **Running XT in the wrong Laue group leads to disaster.**

In theory, SHELXT could also have been programmed to determine the Laue group, e.g. by calculating the R values or correlation coefficients when the equivalent reflections are merged. However, the Laue group has to be known to scale the data, which is an essential step for the highly focused beams now common for synchrotrons and laboratory microsources, because the effective volume of the crystal irradiated is different for different reflections and needs to be corrected for. So in practice it is best to determine the Laue group first anyway.

| S | olve 🔥 R | efine | 🛛 💙 Report | · V | | | | | |
|---|-------------------------|-----------------|---------------------------|------------|--|--|--|--|--|
| 0 | Solution Program | SheIXT | | • | | | | | |
| 0 | Solution Method | Direct Methods | | - | | | | | |
| 0 | Reflection File | NOSTRU.hkl | | • | | | | | |
| 0 | Chemical Composition | C14 H13 N1 O1 S | 1 | • | | | | | |
| 0 | Z and Z' | Atomic Vol. = 1 | 17.6 Å ³ Z = 4 | Z' = 1 | | | | | |
| 0 | Space Group Suggest SG | | P2121 | 21 🗸 | | | | | |
| 0 | Solution Settings Extra | | | | | | | | |
| 0 | TREF | | | | | | | | |
| | np nE kap | scal | ntan | wn | | | | | |
| 0 | INIT 🗖 | | | | | | | | |
| | nn nf s_plus 0.8 | s | _minus 0.2 | wr 0.2 | | | | | |
| 0 | PHAN | | | | | | | | |
| | stepscool100.9 | Boltz r | ns mtpr 40 | mnqr 10 | | | | | |
| 0 | ESEL | | | | | | | | |
| | Emin Emax | dU | renorm | axis | | | | | |
| | 1.2 5 | 0.05 | 0.7 | 0 | | | | | |
| 0 | EGEN | | | | | | | | |
| | d_min | d_max | | | | | | | |
| 0 | GRID | | | | | | | | |
| | si sa s | d di | da | dd | | | | | |
| 0 | COMMAND LINE | | | | | | | | |
| | Options | | | | | | | | |

SHELXT – Command line options

Also SHELXT does not read any instructions from a text-file (the ins-file is only required to determine the Laue group), it can be governed by command line options (which can be provided in OLEX). Possible options can be found at http://shelx.uni-ac.gwdg.de/SHELX/shelxt_keywords.php or at the beginning of the *.lxt file.

General:

-1 Laue group N (SADABS code). N=15 all hexagonal and trigonal N=16 monoclinic with a unique, N=17 monoclinic with c unique

Overrides the Laue group provided in the ins-file.

-tN use N threads, otherwise use 5 or max available, if less. -d highest resolution to be employed [-d0.8] -e fill out missing data to specified

resolution [-eX] where X is max(0.9,d-0.1) and d is the observed resolution

| olve | | Nefin | e | V Rep | ort | | | | | | | | | |
|--|---------|---------------|--|-----------|--|--|--|-------------------------------|--|-----|--|--|--|--|
| Solution Program | | ShelX | ShelXT | | | | | | | | | | | |
| Solution Method Reflection File Chemical Composition | | | Direct Methods • NOSTRU.hkl • C14 H13 N1 O1 S1 • | | | | | | | | | | | |
| | | | | | | | | 🕖 Z and Z' | | Ato | Atomic Vol. = 17.6 Å ³ Z = 4 Z' = 1 | | | |
| | | | | | | | | Space Group <u>Suggest SG</u> | | | P212121 | | | |
| Solution | Setting | s Extra | | | | | | | | | | | | |
| TREF | | | | | | | | | | | | | | |
| np 500 | nE | kapscal | | ntan | wn | | | | | | | | | |
|) INIT 🗖 | | | | | | | | | | | | | | |
| nn | nf | s_plus 0.8 | s_m 0.2 | inus 2 | 0.2 | | | | | | | | | |
| D PHAN | | | | | | | | | | | | | | |
| steps | cool | Boltz | ns | mtpr | mnqr | | | | | | | | | |
| 10 | 0.9 | | | 40 | 10 | | | | | | | | | |
| ESEL | | | | | | | | | | | | | | |
| Emin 1.2 | Emax | (| dU rei | norm | axis | | | | | | | | | |
| | | | 0.05 | | , in the second se | | | | | | | | | |
| d_min | | | d_max | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| GRID | | | | 4.5 | | | | | | | | | | |
| si | 58 | su | ai | ua | aa | | | | | | | | | |
| COMMAND | INF | | | | | | | | | | | | | |
| Options | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

SHELXT – Command line options

Phasing:

```
-q structure factors Go=E^q*Fo^(1-q) [-q0.5]
-iN NGO-(N-1)Gc map in dual space recycling [-i3]
-o switch OFF Patterson superpostion (not recommended)
-kN apply random omit every kth cycle [-k3]
-fX randomly omit fraction X of atoms [-f0.3]
-z sigma threshold for P1 peak-search [-z2.5]
-uX tangent expansion for E>X after random omit [off]
-v atomic volume threshold for P1 peak-search [-v13]
-m initial number of P1 dual space iterations [-m100]
-b spread factor for atom masks [-b3]
-jX CFOM = 0.01*CC - X*R(weak) [-j1]
-y CFOM = CHEM*CC (alternative to default -j1) [off]
```

CC is the correlation coefficient between calculated and observed intensities. R_{weak} is the R factor based on normalized intensities. Typically a space group is evaluated by calculating a combined figure of merit CFOM = $0.01 \cdot \text{CC} - x \cdot R_{\text{weak}}$ (x = 1 by default).

An alternative approach is the use of CFOM = CHEM·CC: CHEM is the fraction of all bond angles (if each peak is considered an atom) which is between 95 and 135°. (Langs, D. A. & Hauptman, H. A. (2011). Acta Cryst. A67, 396–401.). CHEM thus adresses the "chemical meaningfulness" of the solution. This is might be helpful for organic or coordination compounds, but is by default switched off, since it is not meaningfull for purely inorganic compounds.

-xX accept if CFOM > X+0.01*max(20-m, 0) where m is try number [-x0.65]

Structure solution methods

Patterson methods (vector methods)

- Works only in the presence of heavy atoms
- No upper size limitation
- Normally no complete structures in initial solution
- Today practically only used for protein phasing

Direct methods (XS)

- Works well up to 100-200 atoms / asymmetric unit
- Sometimes problems with centrosymmetric space groups

Dual space methods (XM)

- Direct methods up to 2000 atoms
- Mainly for protein crystallography, a bit more complicated setup/use

• Dual space methods (XT)

- Most likely the replacement for XS
- No assumption for space group necessary (direct determination from phases)
- Works well with low quality data
- Severe/Whole molecule disorder (e. g. unresolved twinning) can cause problems

Charge Flipping

- No assumptions for symmetry necessary, everything solved in P1
- Might be applicable for small proteins
- Data quality and completeness important, but works well with highly disordered or even modulated structures