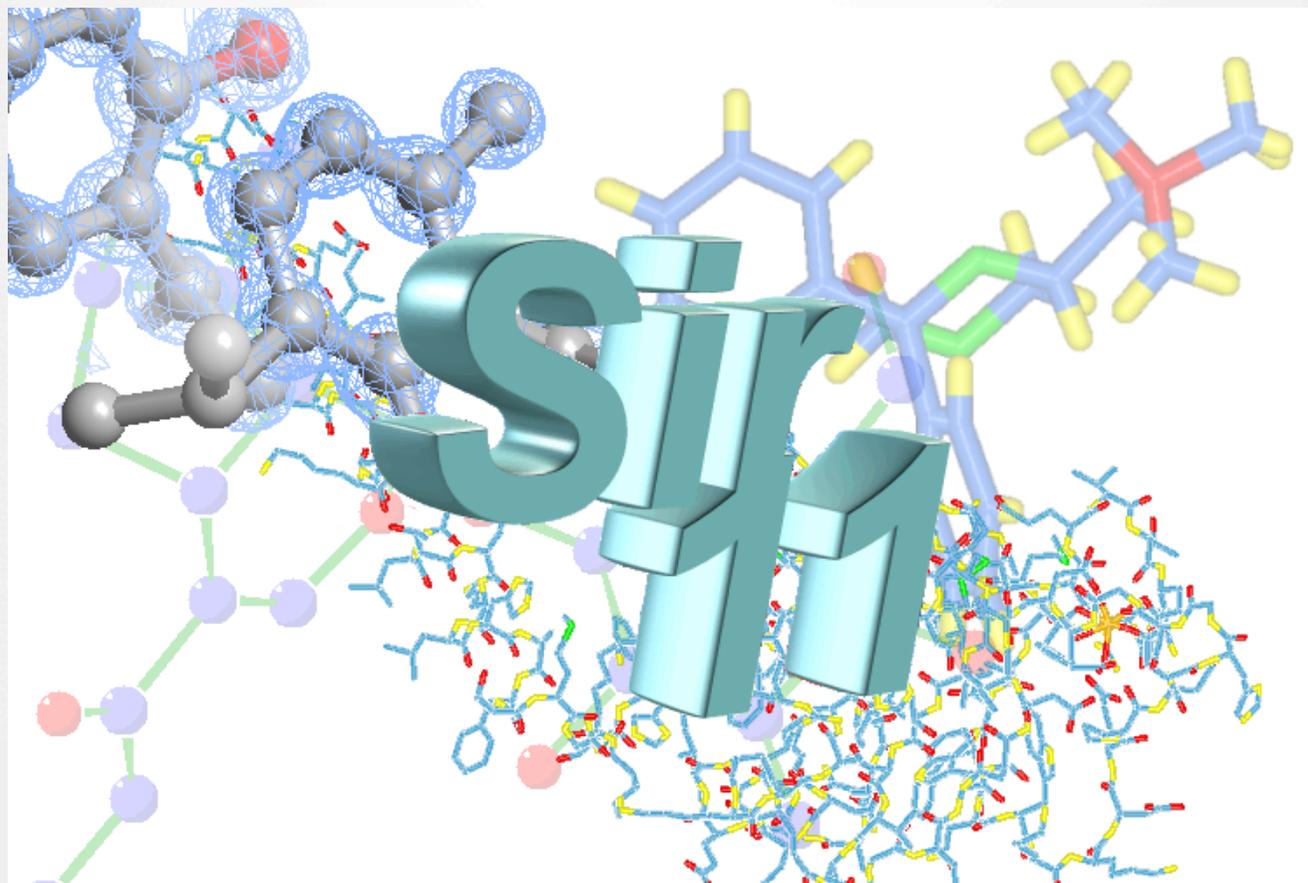




Introduction to Sir2011

IC|OI



Giovanni Luca Cascarano

CNR- Istituto di Cristallografia - Bari - Italy



The crystal structure solution

IC|OI

To solve a crystal structure, i.e. to obtain the atomic coordinates, different methods are available.

We'll focus our attention on Direct Methods, very popular and effective, able to give in a short time and in automatic way the solution to our problem: to compute the phases lost during the diffraction experiment (the so called *Phase Problem*).

Once the phases are available it is possible to apply the Fourier Transform and get the electron density map whose maxima correspond to the atomic positions.



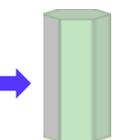
The crystal structure solution

IC|OI

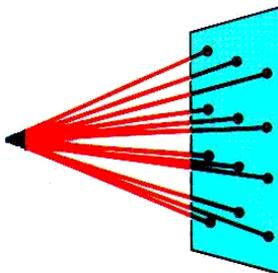
Crystallization



Beam
(X rays, electrons)



Crystal



Diffraction pattern



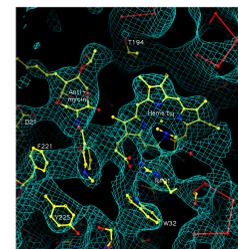
$$I = |F|^2$$
$$\varphi = ?$$



Crystallographic
Methodologies



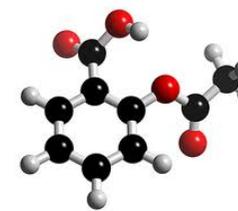
Software



Electron density map



Atomic model





The crystal structure solution

IC|OI

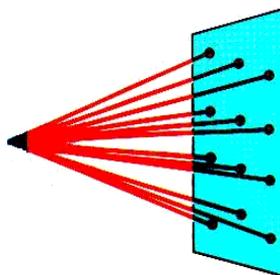
Crystallization



Beam
(X rays, electrons)



Crystal



Diffraction pattern



$I = |F|^2$
 $\varphi = ?$

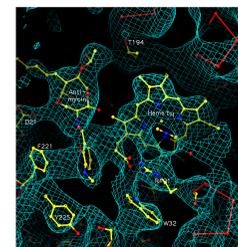
The phase problem



Crystallographic
Methodologies



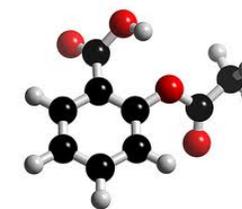
Software



Electron density map



Atomic model



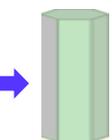


The crystal structure solution IC|OI

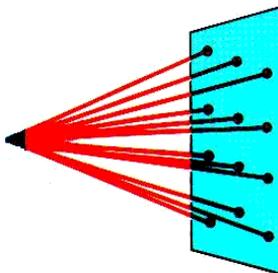
Crystallization



Beam
(X rays, electrons)



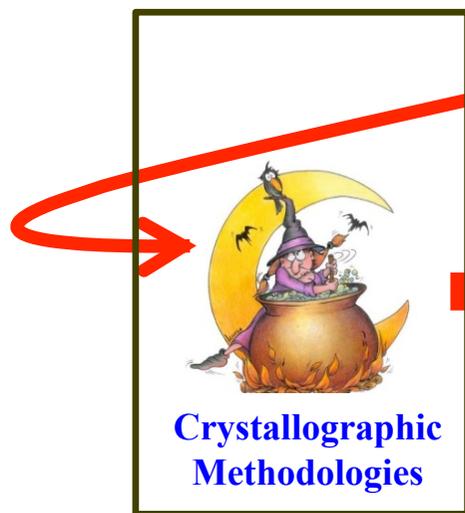
Crystal



Diffraction pattern



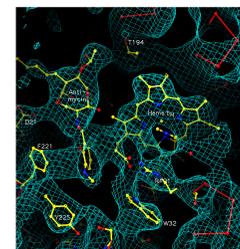
$$I = |F|^2$$
$$\varphi = ?$$



Crystallographic
Methodologies



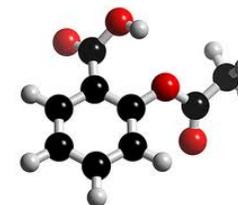
Software



Electron density map



Atomic model



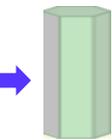


The crystal structure solution IC|OI

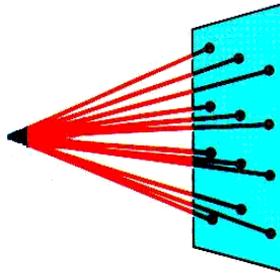
Crystallization



Beam
(X rays, electrons)



Crystal



Diffraction pattern



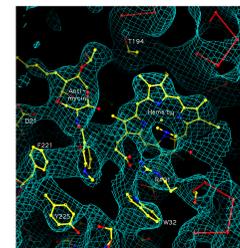
$$I = |F|^2$$
$$\varphi = ?$$



Crystallographic
Methodologies



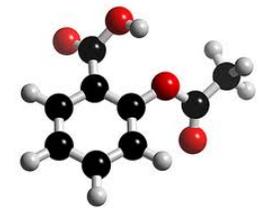
Software



Electron density map



Atomic model





The phase problem

If we define the structure factor \mathbf{F}_h with indices $\mathbf{h}=(hkl)$

$$\mathbf{F}_h = \sum_{j=1}^N f_j \exp(2\pi i \mathbf{h} \mathbf{r}_j) = |\mathbf{F}_h| \exp(i\phi_h)$$

the electron density $\rho(\mathbf{r})$ for every point \mathbf{r} in the unit cell is defined as:

$$\rho(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}} |\mathbf{F}_h| \exp(i\phi_h) \exp(-2\pi i \mathbf{h} \mathbf{r})$$

In this formula the $|\mathbf{F}_h|$ value is known from experiment, the phase of the structure factor ϕ_h has to be recovered.



Statistical analysis

IC|OI

The statistical analysis of the structure factor amplitudes gives useful information about the presence or absence of the inversion center; furthermore it allows the estimate of the overall isotropic thermal factor (B) and of the scale factor (K), necessary to transform the structure factor F_h in the normalized structure factor E_h used in Direct Methods.

To derive the theoretical probability distributions it is assumed that the atomic positions are random variables, uniformly distributed in the unit cell.

Wilson (1942) proposed a simple method to derive K and B , based on the statistical analysis of the intensities.



Statistical analysis

IC|OI

The experimental structure factors are on a relative scale. If we assume that the thermal motion is isotropic and equal for all atoms:

$$\mathbf{F}_h = \exp(-Bs^2) \prod_{j=1}^N f_j^o \exp(2\pi i \mathbf{h} \mathbf{r}_j) = \exp(-Bs^2) \mathbf{F}_h^o$$

where $s = \sin(\theta)/\lambda$.

The experimental structure factors $|\mathbf{F}_{obs}|$ are related to the calculated structure factors $|\mathbf{F}_h|$ through a scale constant K :

$$|\mathbf{F}_h|_{obs}^2 = K |\mathbf{F}_h|^2 = K |\mathbf{F}_h^o|^2 \exp(-2Bs^2)$$



Statistical analysis

IC|OI

Sir2011 divides the reflections in s^2 shells and computes the average values of the intensities in every shell; assuming that the atoms are uniformly distributed in the unit cell:

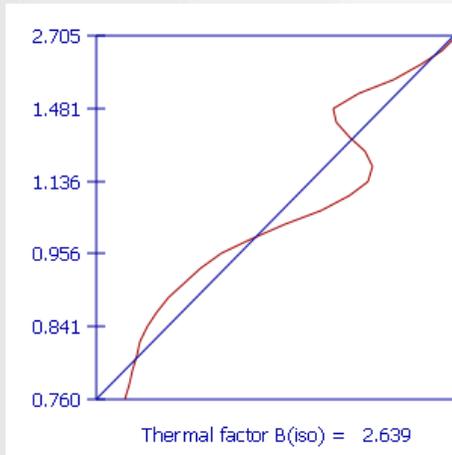
$$\langle |\mathbf{F}_h^o|^2 \rangle = \sum_{j=1}^N (f_j^o)^2 = \sum_s^o$$

from which:

$$\ln \left(\frac{\langle |\mathbf{F}_h^o|^2 \rangle}{\sum_s^o} \right) = \ln K - 2Bs^2$$



Statistical analysis



If the values of the logarithm are plotted against s^2 and the least square line is computed, its intercept on the vertical axis will give the value of $\ln K$, its slope the value of $2B$.

From the scaled intensities, corrected using the thermal factor, we obtain the normalized structure factors:

$$|\mathbf{E}_h|^2 = \frac{|\mathbf{F}_h|^2}{\epsilon \Sigma}$$

These magnitudes do not depend on resolution and on the atomic species, so they correspond to point atom structure. They are needed in the Direct Methods procedures.



Statistical analysis

IC|OI

In this example using the least square line

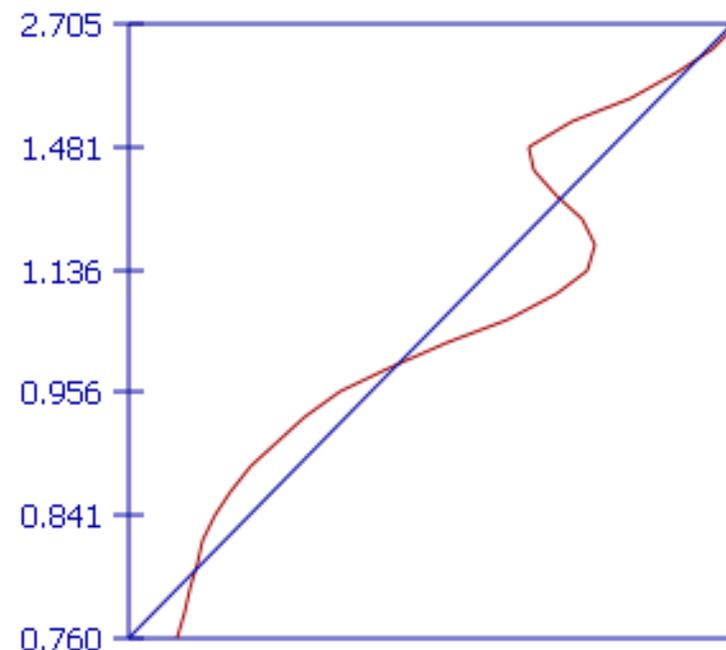
$$\ln\left(\frac{\langle |\mathbf{F}_h^o|^2 \rangle}{\Sigma_s^o}\right) = \ln K - 2Bs^2$$

we obtain the following values:

$$\text{Intercept} = \ln K = -0.08737$$

$$\text{Slope} = -2B = -5.27746$$

$$K = 1.09130 \quad B = 2.63873$$



Thermal factor B(iso) = 2.639



Structure Invariants

IC|OI

The moduli of the observed structure factors are independent from the origin but the phases depend on it.

Starting from the observed structure factor it is possible to obtain information only on linear combinations of phases independent from the origin choice called *Structure Invariants*.

The most used structure invariant is formed by three phases and is called *triplet*:

$$\Phi_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}$$

where the the sum of the indices is equal to zero.



Structure Invariants

IC|OI

Let's consider the structure factor referred to a new origin shifted by \mathbf{r}_0 :

$$\begin{aligned}\mathbf{F}'_{\mathbf{H}} &= \sum_{j=1}^N f_j \exp(2\pi i \mathbf{H} \cdot (\mathbf{r}_j - \mathbf{r}_0)) \\ &= \mathbf{F}_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}_0)\end{aligned}$$

$$= |\mathbf{F}_{\mathbf{H}}| \exp(i(\Phi_{\mathbf{H}} - 2\pi i \mathbf{H} \cdot \mathbf{r}_0))$$

The modulus does not change, the phase differs:

$$\Delta\Phi_{\mathbf{H}} = 2\pi \mathbf{H} \cdot \mathbf{r}_0$$

If we compute the structure factor respect to the new origin

$$\begin{aligned}\mathbf{F}'_{\mathbf{H}_1} \mathbf{F}'_{\mathbf{H}_2} \dots \mathbf{F}'_{\mathbf{H}_n} &= \mathbf{F}_{\mathbf{H}_1} \mathbf{F}_{\mathbf{H}_2} \dots \mathbf{F}_{\mathbf{H}_n} \exp\left[i(\Delta\Phi_{\mathbf{H}_1} \Delta\Phi_{\mathbf{H}_2} \dots \Delta\Phi_{\mathbf{H}_n})\right] \\ &= \mathbf{F}_{\mathbf{H}_1} \mathbf{F}_{\mathbf{H}_2} \dots \mathbf{F}_{\mathbf{H}_n} \exp\left[i2\pi \sum_{j=1}^n \mathbf{H}_j \cdot \mathbf{r}_0\right] = \mathbf{F}_{\mathbf{H}_1} \mathbf{F}_{\mathbf{H}_2} \dots \mathbf{F}_{\mathbf{H}_n}\end{aligned}$$



Triplet estimates: Cochran formula

IC|OI

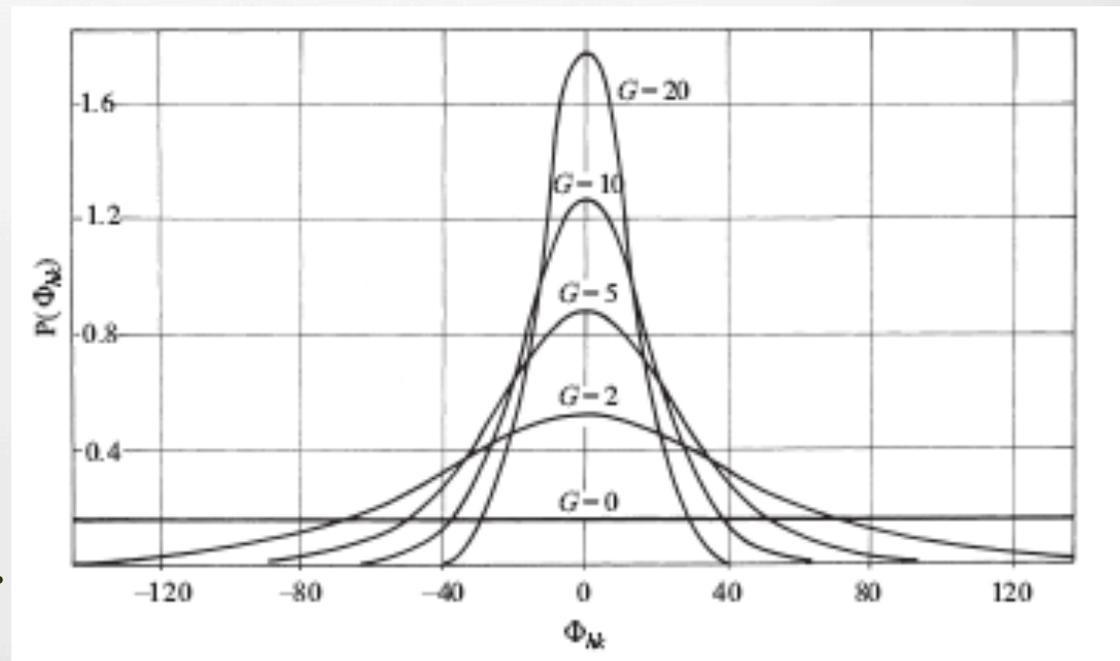
Triplets can be estimated using the classical Cochran formula. In this case the value of the triplet phase $\Phi_{h,k}$ depends only on three phases φ_h , φ_k , φ_{h+k} and its estimate on three normalized structure factors

$$\Phi_{h,k} = \varphi_h + \varphi_k + \varphi_{-h-k}$$

$$P(\Phi_{hk}) \approx [2\pi I_0]^{-1} \exp(G \cos \Phi_{hk})$$

where $G = 2 |E_h E_k E_{h+k}| / N^{1/2}$

is the concentration parameter.

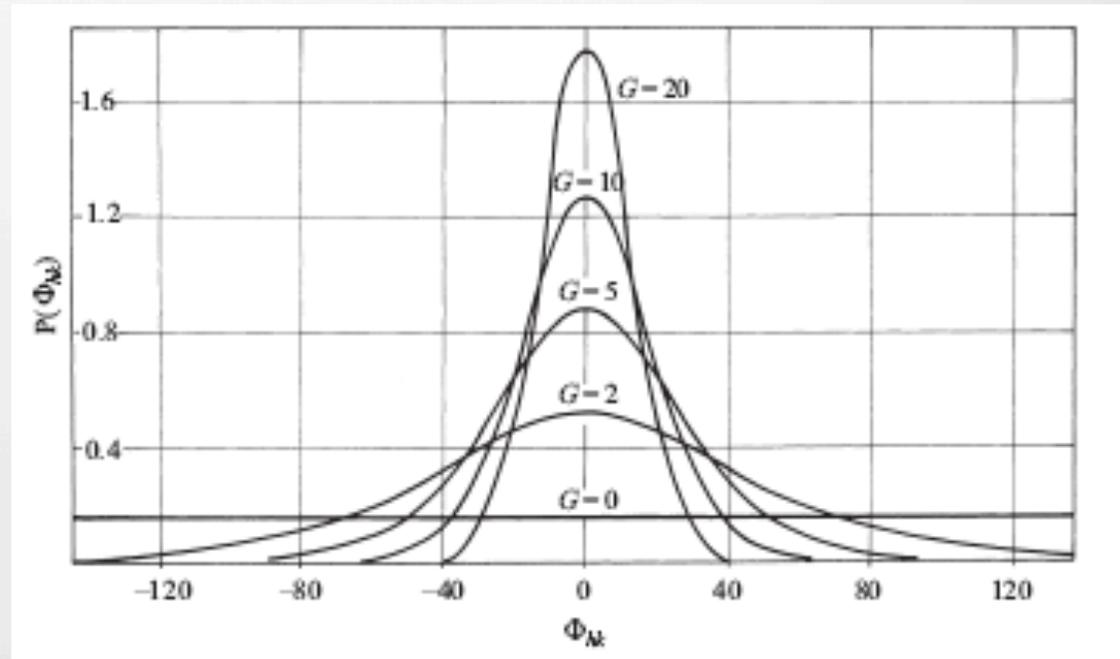




Triplet estimates: Cochran formula

IC|OI

Looking at the figure, we can deduce that larger is the value of G_{HK} , higher is the probability that the phase Φ_{HK} is close to zero. If we compute the triplet invariants using a subset of reflections with the higher values of $|\mathbf{E}|$, we can be more confident on the value of Φ_{HK} .





Triplet estimates: P10 formula

IC|OI

Triplet invariants may be replaced by a family of quintets:

$$\Phi_{h,k} = \varphi_h + \varphi_k + \varphi_{-h-k} + \varphi_l - \varphi_l$$

the values of which is perfectly equal to the basis triplet.

The vector **l** may freely vary over the set of measured reflections.

Now $\Phi_{h,k}$ depends, for each **l** vector, on 10 moduli

$$|E_h|, |E_k|, |E_{h+k}|,$$

$$|E_l|, |E_{h+l}|, |E_{h-l}|, |E_{k+l}|, |E_{k-l}|, |E_{h+k+l}|, |E_{h+k-l}|.$$



Triplet estimates: P10 formula

IC|OI

The triplet phase $\Phi_{h,k}$ may be estimated *via*

$$P(\Phi_{hk}) \approx [2\pi I_0]^{-1} \exp(G \cos \Phi_{hk})$$

but this time

$$G_{h,k} = \frac{2 |E_h E_k E_{h+k}|}{\sqrt{N}} \left(1 + \frac{1}{N} C\right)$$

C is a summation of contributions deriving from the ten E's.

If $C/N < -1$ the triplet is expected to be negative .



The tangent formula

We can write the triplet phase relationship as $\varphi_H \approx \varphi_K + \varphi_{H-K}$; the distribution of φ_H with respect to the expected value

$$\theta_H = \varphi_K + \varphi_{H-K}$$

is

$$P(\varphi_H) = K \exp(G_{HK} \cos(\varphi_H - \theta_H))$$

Let's consider all the triplets related to a reflection H

$$\varphi_H \approx \varphi_{K1} + \varphi_{H-K1}$$

$$\varphi_H \approx \varphi_{K2} + \varphi_{H-K2}$$

.....

$$\varphi_H \approx \varphi_{Kn} + \varphi_{H-Kn}$$



The tangent formula

If we assume that the phase indications are statistically independent, the joint probability of the phase φ_H will be

$$\begin{aligned} P(\varphi_h) &\approx \prod_j P_j(\varphi_h) \approx L^{-1} \prod_j \left(\exp[G_j \cos(\varphi_h - \vartheta_j)] \right) \\ &\approx L^{-1} \exp \left[\sum_j G_j \cos(\varphi_h - \vartheta_j) \right] \\ &\approx L^{-1} \exp \left(\cos \varphi_h \sum_j G_j \cos \vartheta_j + \sin \varphi_h \sum_j G_j \sin \vartheta_j \right) \end{aligned}$$



The tangent formula

If we set

$$\Sigma G_j \sin \theta_j = \alpha_H \sin \theta_j \quad \text{and} \quad \Sigma G_j \cos \theta_j = \alpha_H \cos \theta_j$$

we obtain

$$P(\varphi_H) \approx K \exp(\alpha_H \cos(\varphi_H - \theta_H)) \quad \text{where} \quad \alpha_H = (T^2 + B^2)^{1/2}$$

and

$$\tan \theta_h = \frac{\sum_j G_j \sin \theta_j}{\sum_j G_j \cos \theta_j} = \frac{T}{B}$$

This equation gives the most probable value of θ_j and is known as the *tangent formula*.

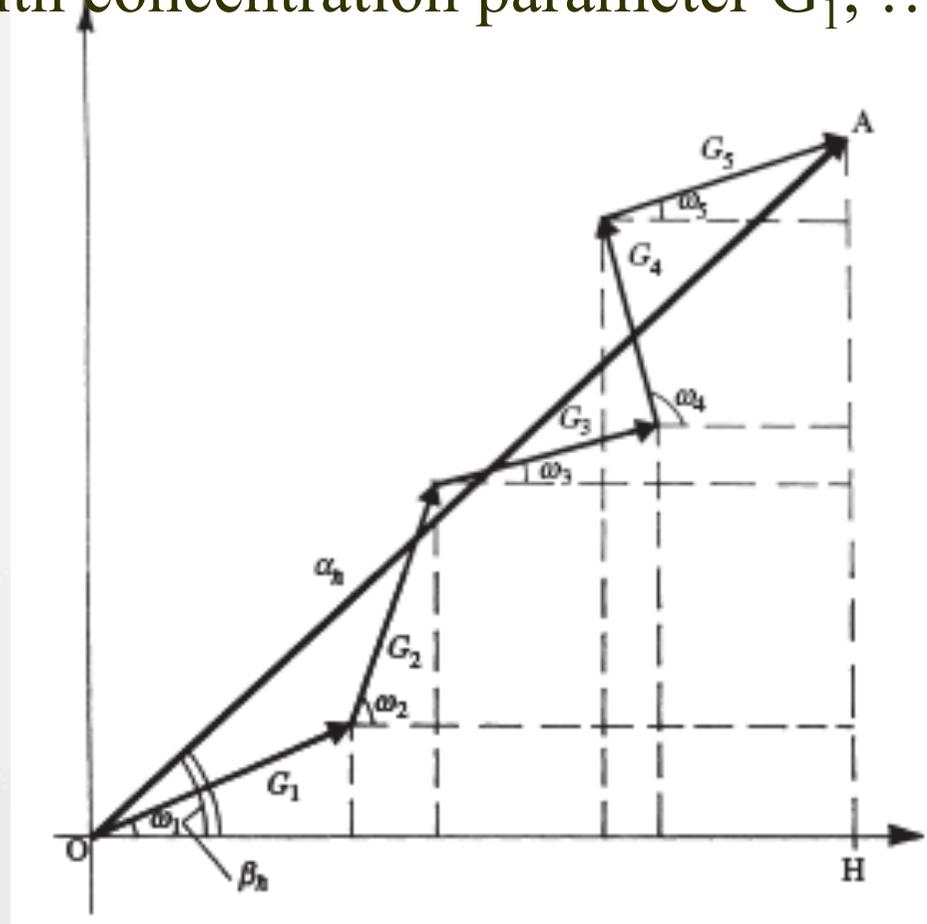
Using this formula is possible to derive the phase values of a given reflection when an estimation of n couples of phases is known.



The tangent formula

IC|OI

Vector representation, in the complex plane, of the combination of five triplets with concentration parameter G_1, \dots, G_5





The multiresolution approach

Usually random phase values are assigned to a subset of *strong reflections* (those with the higher of $|E|$); by means of the triplets and of the tangent formula almost all the strong reflections can be phased.

This procedure is repeated n times changing the values of the random phases and, for every obtained set of phases a suitable **Figure of Merit** (eFOM) is computed. Higher is the eFOM value, more likely the corresponding phase set is correct.

These trials, sorted with respect to the eFom value, will be submitted to the automatic Direct Space Refinement (DSR).



The multiresolution approach

Usually random phase values are assigned to a subset of *strong reflections* (those with the higher of $|E|$); by means of the triplets and of the tangent formula almost all the strong reflections can be

phased

This p

random

Figure

more l

These

submit

Viewing printer output file : loganin.out

Medium

+++++++ Figures of merit ++++++

| Set | (Trial) | eFom | Set | (Trial) | eFom | Set | (Trial) | eFom |
|-----|---------|-------|-----|---------|-------|-----|---------|-------|
| 1 | (13) | 4.221 | 2 | (44) | 4.217 | 3 | (52) | 4.189 |
| 4 | (6) | 3.320 | 5 | (2) | 3.319 | 6 | (56) | 3.190 |
| 7 | (45) | 3.176 | 8 | (38) | 2.619 | 9 | (43) | 2.535 |
| 10 | (92) | 2.260 | | | | | | |

+++++

the values of the

phases a suitable

the eFOM value,

value, will be

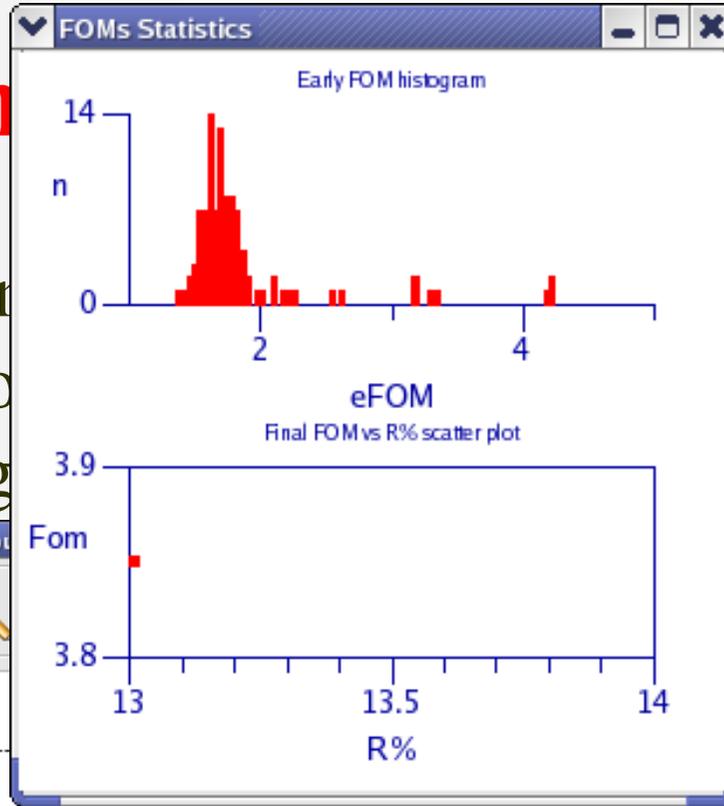
at (DSR).



The

approach

Usually random reflections (those with the highest intensity) and of the tangent plane are phased. This procedure is repeated until a suitable eFOM value is reached. These reflections are then submitted to the DSR.



igned to a subset of *strong* reflections; by means of the triplets the strong reflections can be phased. The values of the eFOM value, will be submitted (DSR).

Viewing printer output

| Set | (Trial) | eFom | Set | (Trial) | eFom | Set | (Trial) | eFom |
|-----|---------|-------|-----|---------|-------|-----|---------|-------|
| 1 | (13) | 4.221 | 2 | (44) | 4.217 | 3 | (52) | 4.189 |
| 4 | (6) | 3.320 | 5 | (2) | 3.319 | 6 | (56) | 3.190 |
| 7 | (45) | 3.176 | 8 | (38) | 2.619 | 9 | (43) | 2.535 |
| 10 | (92) | 2.260 | | | | | | |



The structural complexity

In Sir2011 the parameter related to the structural complexity governs most of the algorithms used for the automatic structure solution.

In the program there is the following classification:

- Very small* size structures: up to 6 non-H atoms in asymmetric unit
- Small* size structures: 6-80 non-H atoms in asymmetric unit
- Medium* size structures: 81-300 non-H atoms in asymmetric unit
- Large* size structures: more than 300 non-H atoms in asymmetric unit



The phase extension procedure

The number of reflections phased by the tangent formula is, in general, a small percentage of all reflections available.

To extend and refine for all suitable reflections (i.e. those for which $F_{\text{obs}} > 3\sigma(F_{\text{obs}})$) an iterative procedure is used, the DSR (Direct Space Refinement).

The procedure is constituted by n cycles

$$\rho \rightarrow \{\varphi\} \rightarrow \rho$$

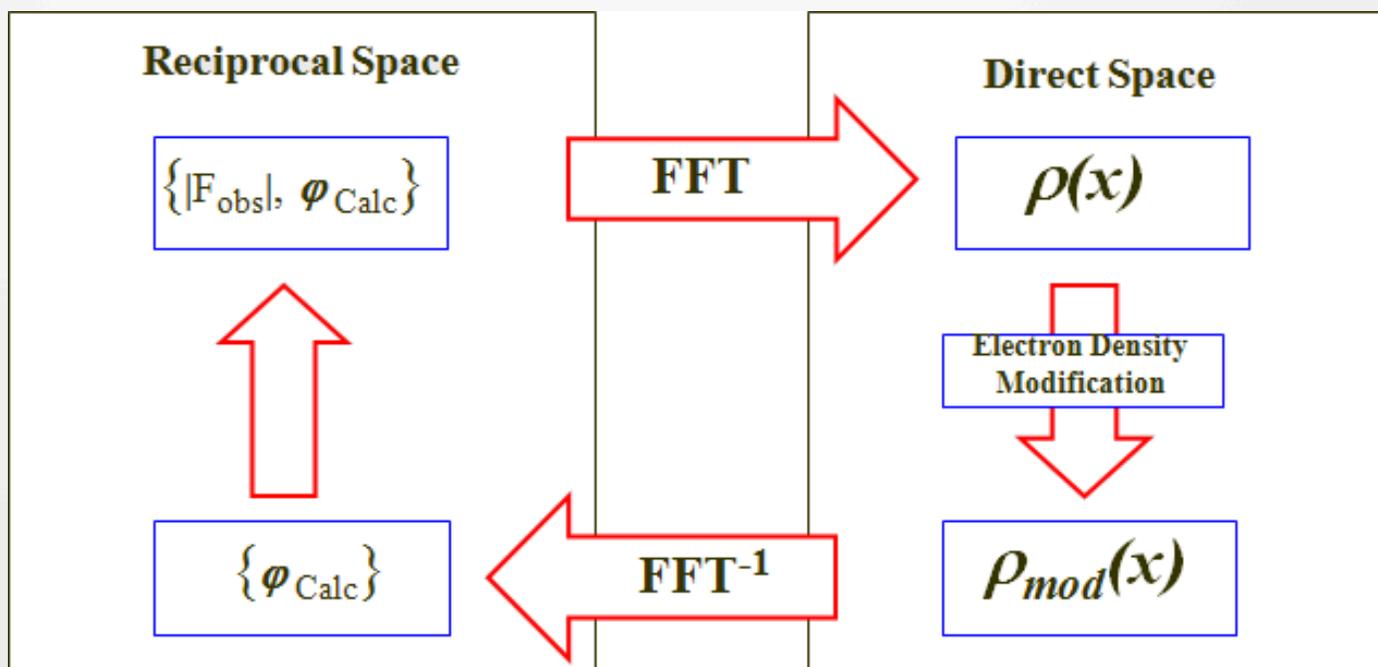
where ρ is the electron density map and $\{\varphi\}$ is the set of calculated phases. The number of cycles depends on the structural complexity.



The phase extension procedure

ICSD

The electron density maps are calculated by using a number of reflections cyclically increasing; a small fraction of ρ is used during the inversion of the electron density map, the remaining part is set to zero. To process a trial, EDM (Electron Density Modification) is applied a number of times, depending on structural complexity.





The automatic model refinement IC 51

At this stage both modulus and phase for an adequate number of reflections are available, therefore it is possible to apply the Fourier transform to the best set of phases (in terms of eFOM), extended using the DSR procedure, to obtain an electron density map.

Its peaks are labelled in terms of atomic species according to their intensity and to the chemical content of the unit cell.

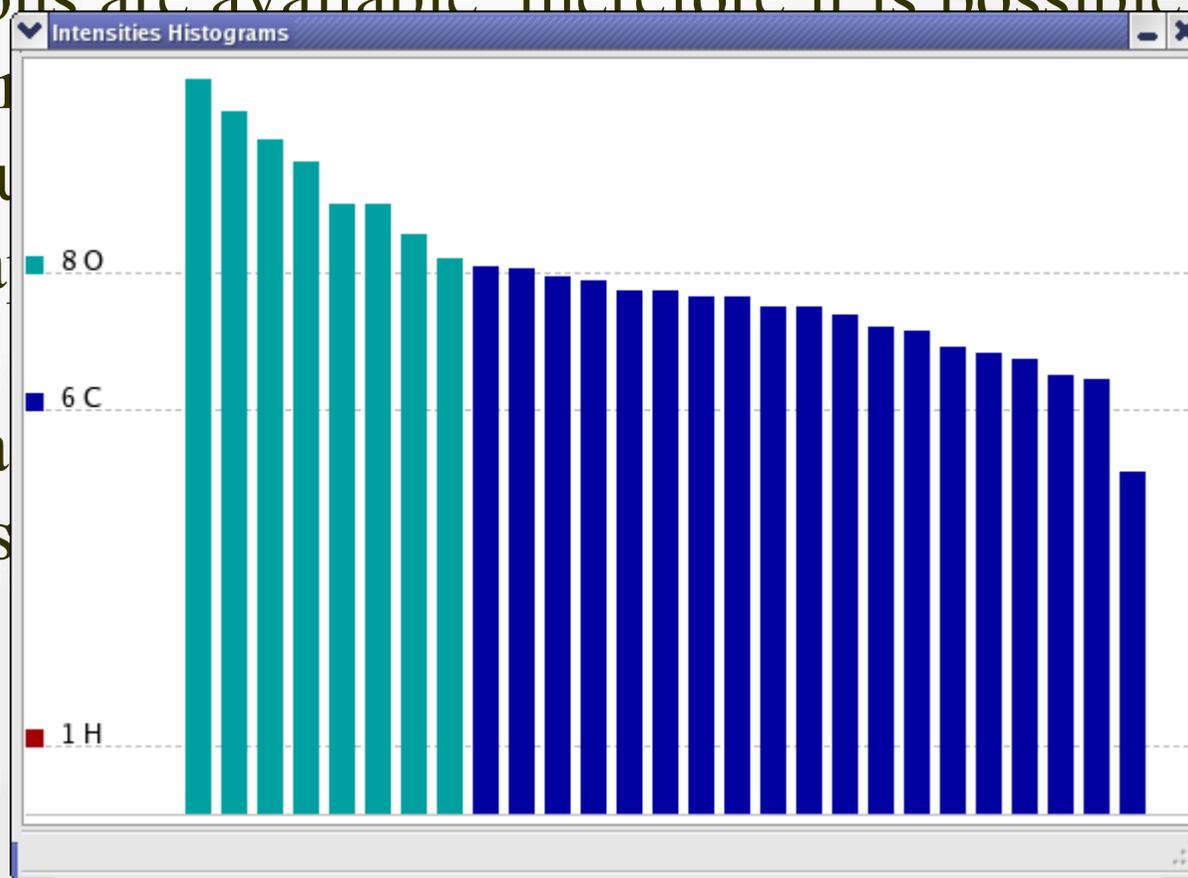


The automatic model refinement IC OI

At this stage both modulus and phase for an adequate number of reflections are available therefore it is possible to apply the

Fourier transform (of eFOM),
extended to
density map

Its peaks are
their intensities



of eFOM),
an electron

according to
it cell.



The automatic model refinement IC|OI

In the automatic model refinement there are two steps:

- six cycles of Fourier Refinement (FR)
- n cycles of Diagonal Least Squares (DLSQ)

The FR procedure modifies the isotropic thermal parameter of the atoms according to the Fourier peak height: to lower values of intensity correspond an higher value of B_{iso} .

During the DLSQ procedure the atomic coordinates and the isotropic thermal factor are refined. This procedure stops when the convergence has been attained.



The automatic model refinement ICSI

During these cycles, new values of phases and weights are computed to produce a new electron density map using $(2F_{\text{obs}} - F_{\text{calc}})$ coefficients. A new map is calculated and this iterative procedure stops when the crystallographic residual R increases, where:

$$R = \frac{\sum w(|F_o| - K|F_c|)}{\sum w(|F_o|)}$$

Fourier
Recycling

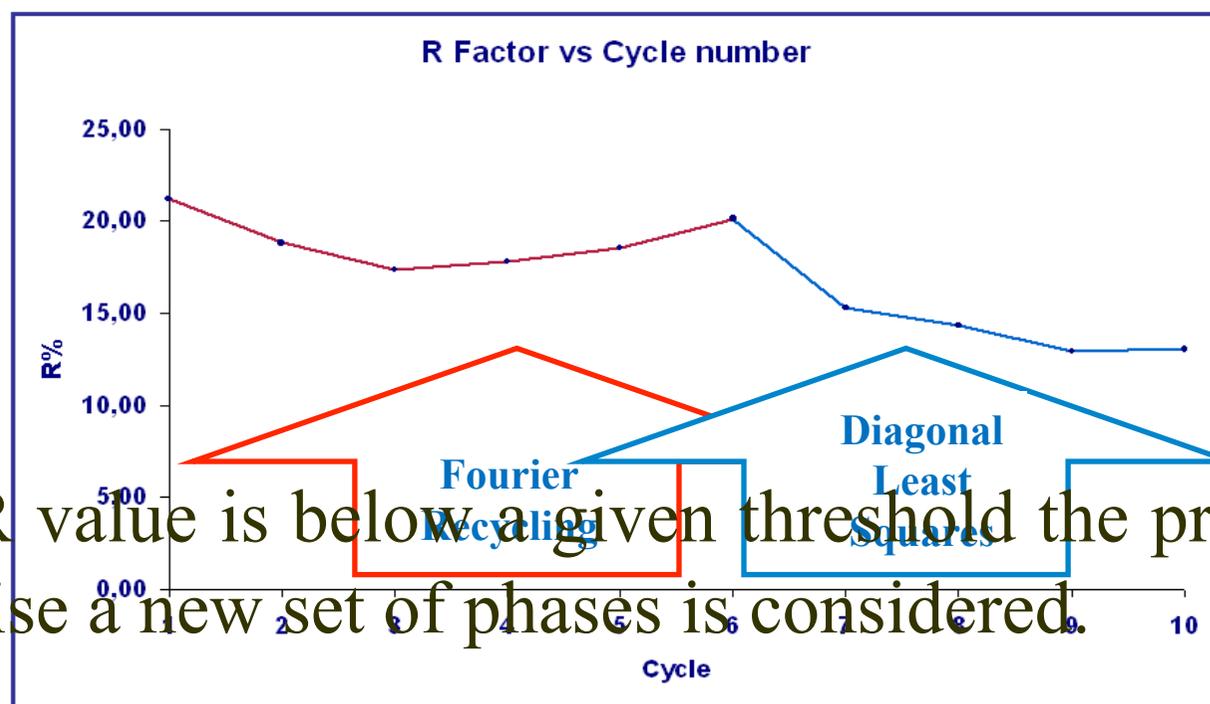
Diagonal
Least
Squares

If the R value is below a given threshold the program stops, otherwise a new set of phases is considered.



The automatic model refinement IC|OI

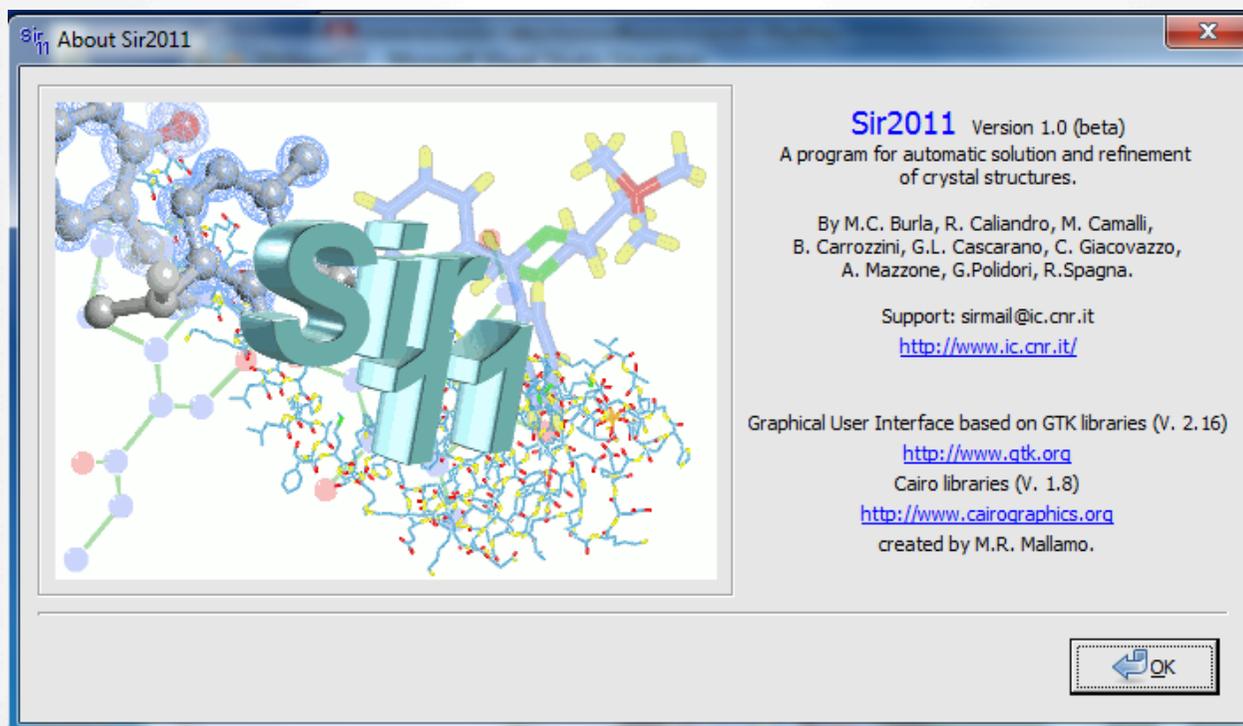
During these cycles, new values of phases and weights are computed to produce a new electron density map using $(2F_{\text{obs}} - F_{\text{calc}})$ coefficients. A new map is calculated and this iterative procedure stops when the crystallographic residual



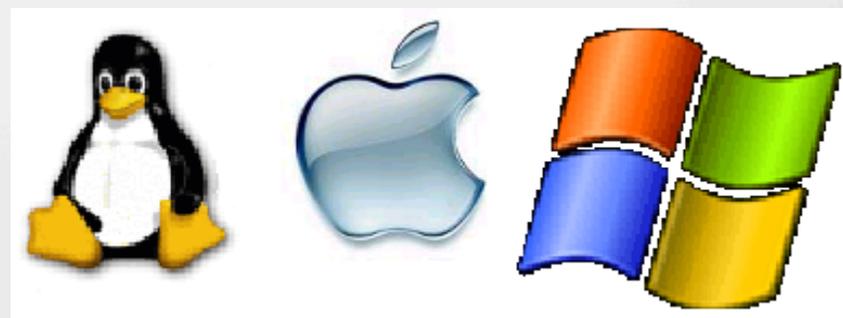
If the R value is below a given threshold the program stops, otherwise a new set of phases is considered.



The program

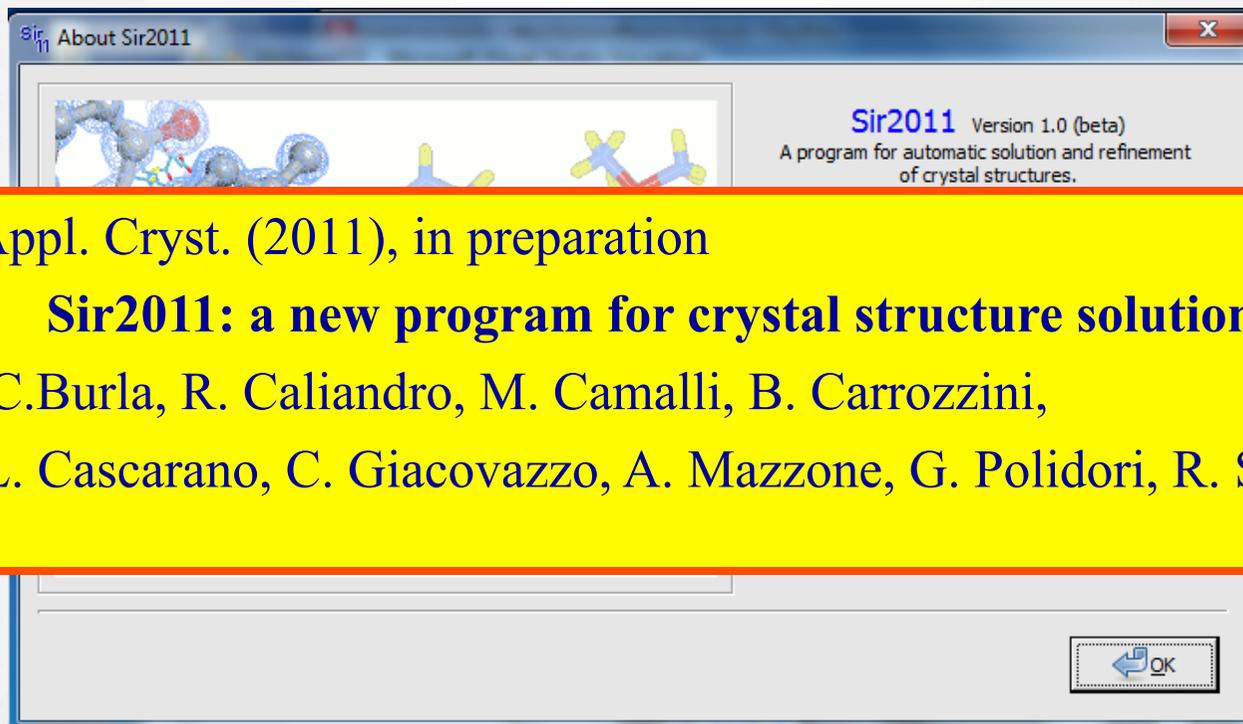


This program will be available, free of charge for academic institutions, for Linux, Mac and MS Windows systems.





The program



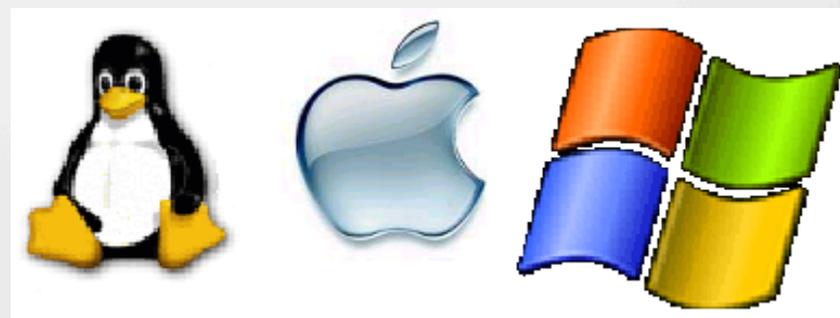
J. Appl. Cryst. (2011), in preparation

Sir2011: a new program for crystal structure solution

M.C.Burla, R. Caliendo, M. Camalli, B. Carrozzini,

G.L. Cascarano, C. Giacovazzo, A. Mazzone, G. Polidori, R. Spagna

This program will be available, free of charge for academic institutions, for Linux, Mac and MS Windows systems.





The authors

**R.Caliandro, B.Carrozzini, G.L.Cascarano,
C.Giacovazzo, A. Mazzone**
Istituto di Cristallografia del CNR – Bari

M.Camalli, R.Spagna
Istituto di Cristallografia del CNR –Rome

M.C.Burla, G.Polidori
*Dipartimento di Scienze della Terra,
University of Perugia*