EXPO&*more*
International Workshop
Organized by the Institute of Crystallography - CNR
30 September - 3 October 2019
Bari - Italy

CNR Istituto di Cristallografia
National Research Council of Italy

STRUCTURE MODEL OPTIMIZATION

Angela Altomare

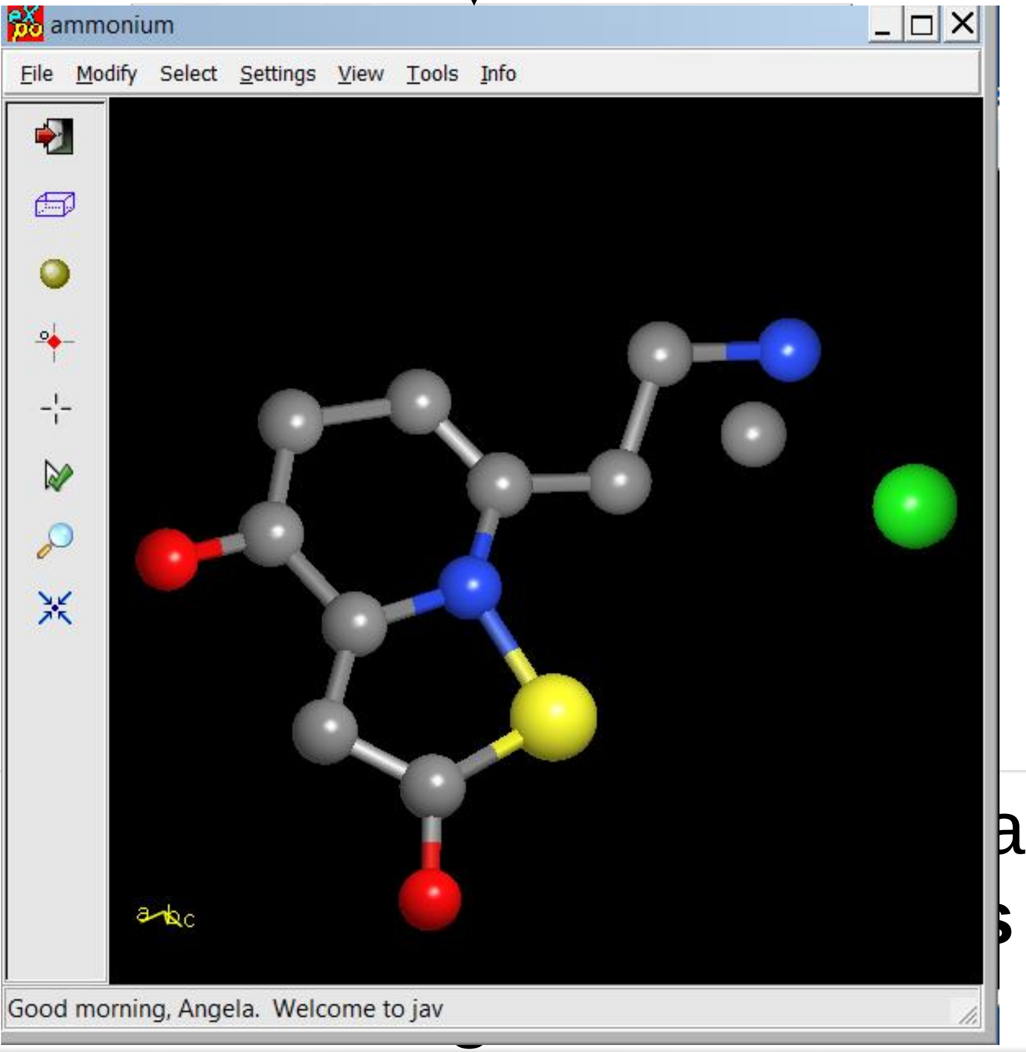
Institute of Crystallography-CNR Bari

THE CRYSTAL STRUCTURE SOLUTION PROCESS IN RECIPROCAL SPACE

- ○ Cell Parameters Determination
- ○ Space Group Identification
- ○ Full Pattern Decomposition
- ○ Structure Solution in Reciprocal Space
  (default choice in EXPO)
- ☻ ○ Structure Model Optimization
- ○ Rietveld Refinement

STRUCTURE MODEL OPTIMIZATION

PHASING PROCESS BY DIRECT METHODS



Good morning, Angela. Welcome to jav

Peak ation:

chem s and

FOURIER MAP

Observed Fourier Synthesis

$$\rho(\boldsymbol{r}) = V^{-1}\sum_{\mathbf{h}} \boxed{\left|F_{o\mathbf{h}}\right|} \exp\left(i\boxed{\varphi_{\mathbf{h}calc}}\right)\exp\left(-2\pi i\overline{\mathbf{h}}\cdot\boldsymbol{r}\right)$$

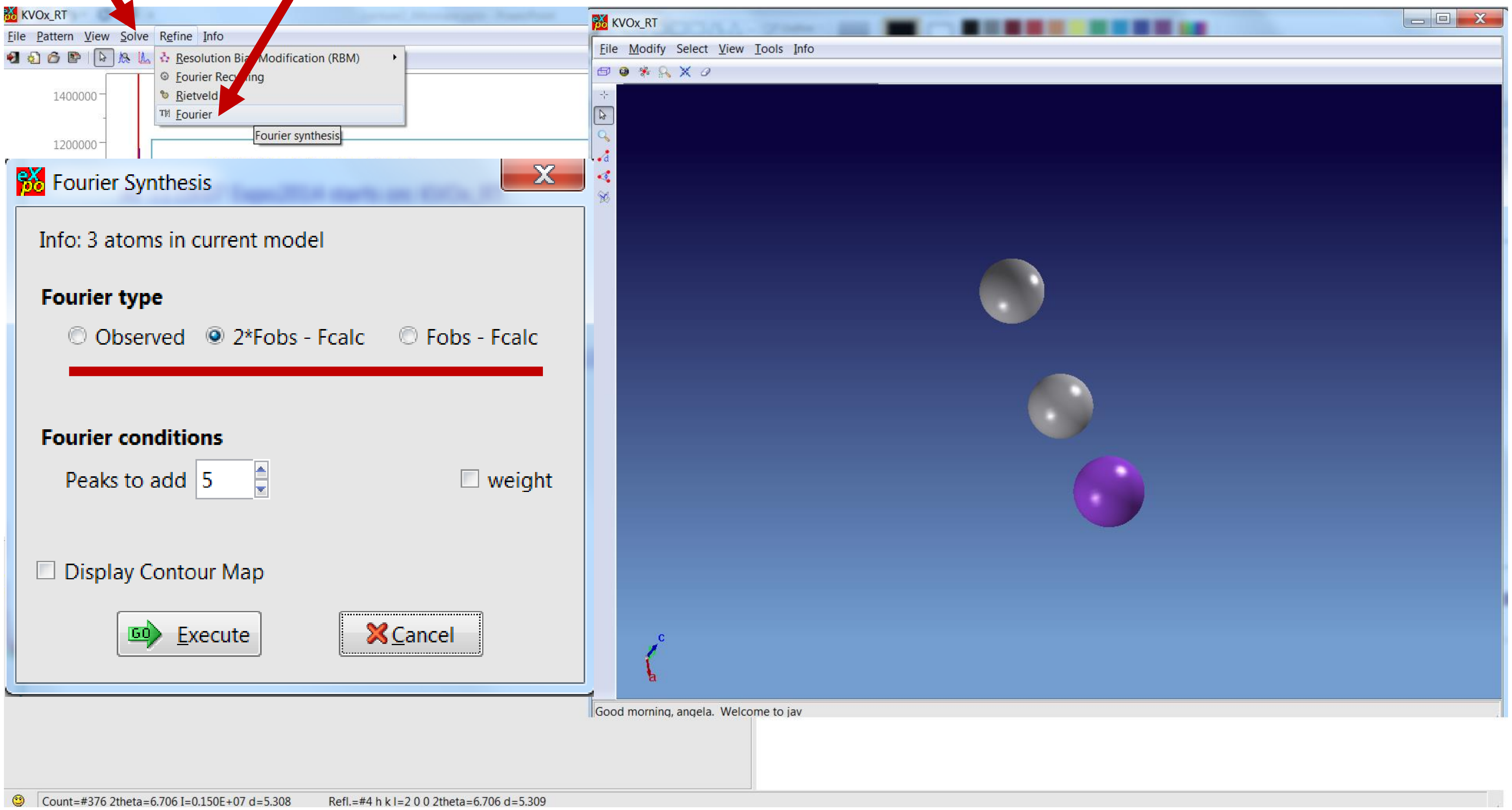
Difference Fourier Synthesis

$$\Delta\rho(\boldsymbol{r}) = V^{-1}\sum_{\mathbf{h}} \boxed{\left(\left|F_{o\mathbf{h}}\right| - \left|F_{\mathbf{h}}\right|_{calc}\right)} \exp\boxed{\left(i\varphi_{\mathbf{h}calc}\right)}\exp\left(-2\pi i\overline{\mathbf{h}}\cdot\boldsymbol{r}\right)$$

and also

$$\Delta\rho(\boldsymbol{r}) = V^{-1}\sum_{\mathbf{h}} \left(\boxed{2\left|F_{o\mathbf{h}}\right| - \left|F_{\mathbf{h}}\right|_{calc}}\right) \exp\left(\boxed{i\varphi_{\mathbf{h}calc}}\right)\exp\left(-2\pi i\overline{\mathbf{h}}\cdot\boldsymbol{r}\right)$$

Solve ->->-> Fourier



Fourier Synthesis

Info: 3 atoms in current model

Fourier type

○ Observed ● 2*Fobs - Fcalc ○ Fobs - Fcalc

Fourier conditions

Peaks to add 5 ☐ weight

☐ Display Contour Map

Execute Cancel

Good morning, angela. Welcome to jav

Count=#376 2theta=6.706 I=0.150E+07 d=5.308 Refl.=#4 h k l=2 0 0 2theta=6.706 d=5.309

STRUCTURE MODEL OPTIMIZATION

The quality of the electron density map calculated after the phasing process by Direct Methods (DM) mainly depends on

- the experimental resolution

atomic resolution is the best

- the reliability of the integrated intensities extracted from the experimental pattern (and used by DM for phasing)

peak overlap should be low (percentage of independent observations larger than 50%)

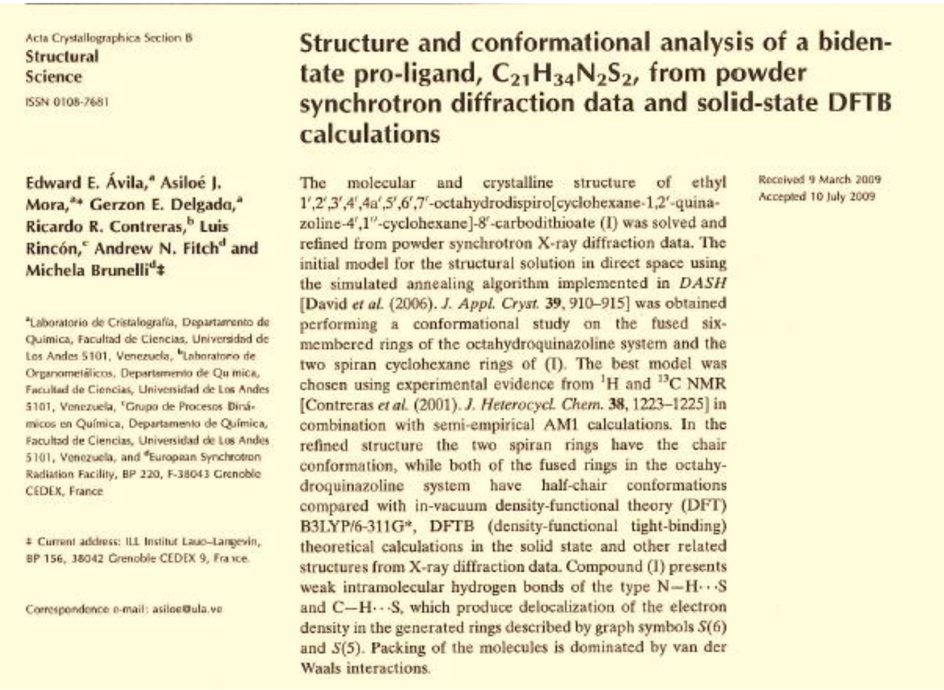
- the structure complexity

it depends on the number of non-H atoms in the asymmetric unit (30 is the current reasonable limit)

WHY DIRECT METHODS MODEL OPTIMIZATION

The structure model provided by DM is often approximate: some atoms are missed to the complete solution, additional false positions occur, some atoms are not perfectly located

It is useful to optimize the model before the final Rietveld refinement, owing to the its modest convergence rate



Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Structure and conformational analysis of a bidentate pro-ligand, $C_{21}H_{34}N_2S_2$, from powder synchrotron diffraction data and solid-state DFTB calculations

Edward E. Ávila,[a] Asiloé J. Mora,[a*] Gerzon E. Delgado,[a] Ricardo R. Contreras,[b] Luis Rincón,[c] Andrew N. Fitch[d] and Michela Brunelli[d]‡

[a]Laboratorio de Cristalografía, Departamento de Química, Facultad de Ciencias, Universidad de Los Andes 5101, Venezuela, [b]Laboratorio de Organometálicos, Departamento de Química, Facultad de Ciencias, Universidad de Los Andes 5101, Venezuela, [c]Grupo de Procesos Dinámicos en Química, Departamento de Química, Facultad de Ciencias, Universidad de Los Andes 5101, Venezuela, and [d]European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble CEDEX, France

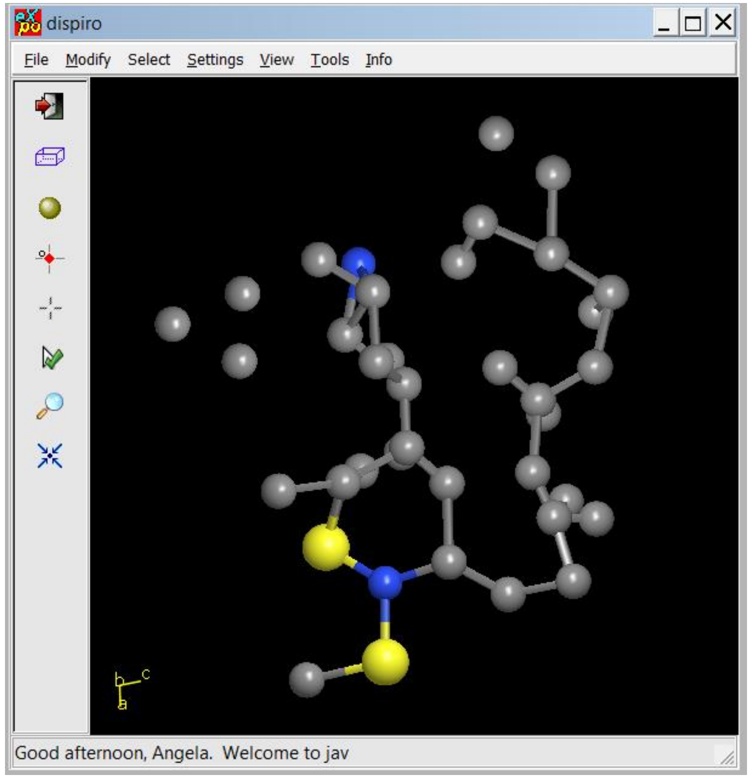
‡ Current address: ILL Institut Laue–Langevin, BP 156, 38042 Grenoble CEDEX 9, France.

Correspondence e-mail: asiloe@ula.ve

Received 9 March 2009
Accepted 10 July 2009

The molecular and crystalline structure of ethyl 1′,2′,3′,4′,4a′,5′,6′,7′-octahydrodispiro[cyclohexane-1,2′-quinazoline-4′,1″-cyclohexane]-8′-carbodithioate (I) was solved and refined from powder synchrotron X-ray diffraction data. The initial model for the structural solution in direct space using the simulated annealing algorithm implemented in *DASH* [David *et al.* (2006). *J. Appl. Cryst.* **39**, 910–915] was obtained performing a conformational study on the fused six-membered rings of the octahydroquinazoline system and the two spiran cyclohexane rings of (I). The best model was chosen using experimental evidence from ^{1}H and ^{13}C NMR [Contreras *et al.* (2001). *J. Heterocycl. Chem.* **38**, 1223–1225] in combination with semi-empirical AM1 calculations. In the refined structure the two spiran rings have the chair conformation, while both of the fused rings in the octahydroquinazoline system have half-chair conformations compared with in-vacuum density-functional theory (DFT) B3LYP/6-311G*, DFTB (density-functional tight-binding) theoretical calculations in the solid state and other related structures from X-ray diffraction data. Compound (I) presents weak intramolecular hydrogen bonds of the type N—H···S and C—H···S, which produce delocalization of the electron density in the generated rings described by graph symbols *S*(6) and *S*(5). Packing of the molecules is dominated by van der Waals interactions.

25 non-H atoms in the asymmetric unit



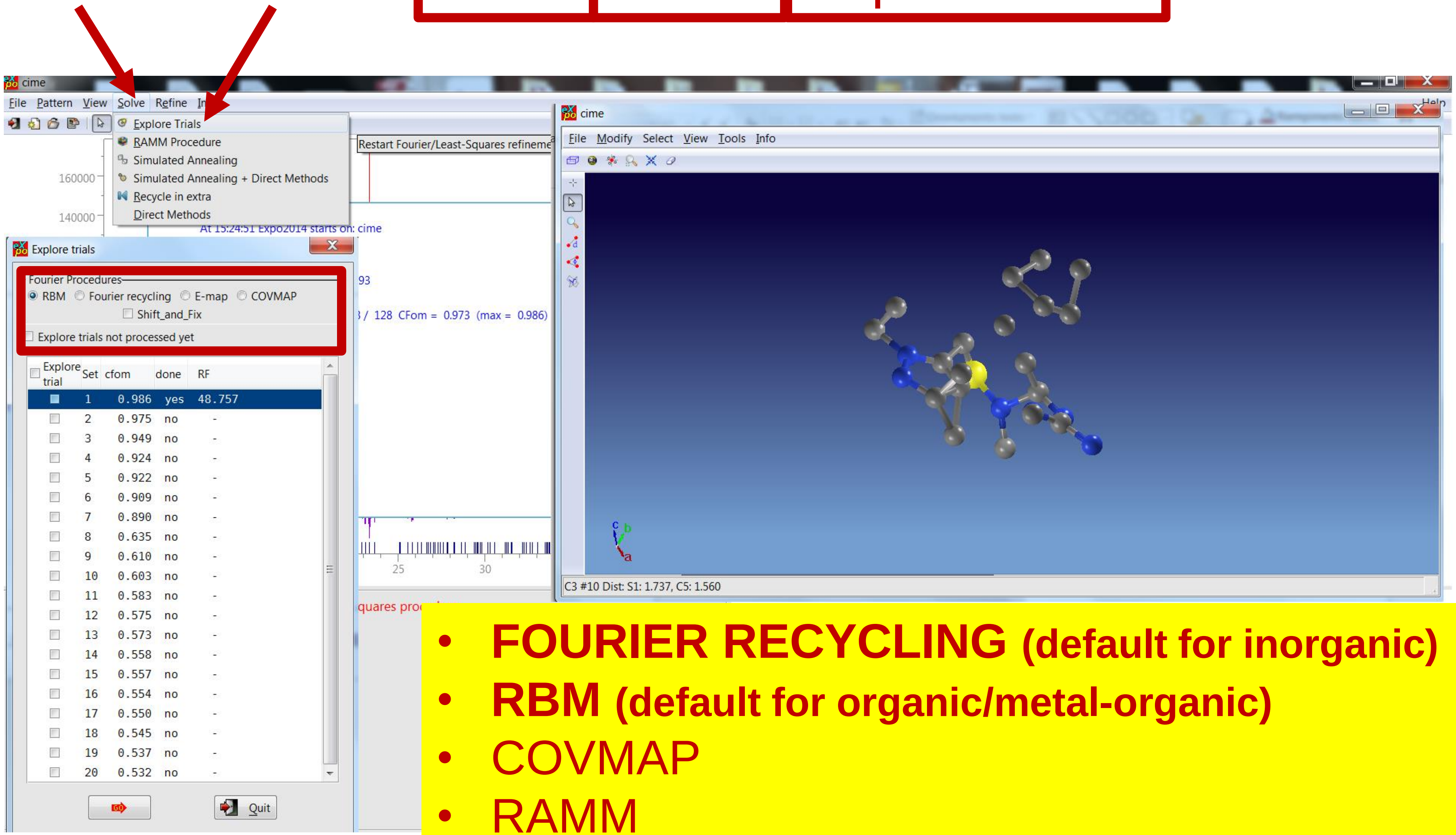
Good afternoon, Angela. Welcome to jav

The EXPO team has developed innovative theories and computing procedures devoted to improve the structure model of organic/metal-organic and inorganic structures

The model optimization requires that different methodological approaches must be adopted depending on the type of compound

- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.G.G.; Rizzi, R. *Powder diffraction: the new automatic least-squares Fourier recycling procedure in EXPO2005*. J. Appl. Cryst. **2006**, *39*, 558-562.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Kamel, G.S.; Moliterni, A.; Rizzi, R. *Minimally resolution biased electron-density maps*. Acta Cryst. **2008**, *A64*, 326–336
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *Correcting resolution bias in electron density maps of organic molecules derived by direct methods from powder data*. J. Appl. Cryst. **2008**, *41*, 592–599.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Maggi, S.; Moliterni, A.; Rizzi, R. *Correcting electron-density resolution bias in reciprocal space*. Acta Cryst. **2009**, *A65*, 183–189.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *The dual-space resolution bias correction algorithm: application to powder data*. J. Appl. Cryst. **2010**, *43*, 798–804.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *COVMAP: a new algorithm for structure model optimization in the EXPO package*. J. Appl. Cryst. **2012**, *45*, 789–797.
- Altomare, A.; Cuocci, C.; C.; Moliterni, A.; Rizzi, R. *RAMM: a new random-model-based method for solving ab initio crystal structure using the EXPO package*. J. Appl. Cryst. **2012**, *45*, 789–797.
- Altomare, A.; Cuocci, C.; Moliterni, A.; Rizzi, R., Corriero, N., Falcicchio, A. *The Shift_and_Fix procedure in EXPO: advances for solving ab initio crystal structures by powder diffraction data.* J. Appl. Cryst. **2017**, *50*, 1812–1820.

Solve ->->-> Explore Trials



Fourier Procedures
● RBM ○ Fourier recycling ○ E-map ○ COVMAP
☐ Shift_and_Fix
☐ Explore trials not processed yet

☐ Explore trial	Set	cfom	done	RF
☐	1	0.986	yes	48.757
☐	2	0.975	no	-
☐	3	0.949	no	-
☐	4	0.924	no	-
☐	5	0.922	no	-
☐	6	0.909	no	-
☐	7	0.890	no	-
☐	8	0.635	no	-
☐	9	0.610	no	-
☐	10	0.603	no	-
☐	11	0.583	no	-
☐	12	0.575	no	-
☐	13	0.573	no	-
☐	14	0.558	no	-
☐	15	0.557	no	-
☐	16	0.554	no	-
☐	17	0.550	no	-
☐	18	0.545	no	-
☐	19	0.537	no	-
☐	20	0.532	no	-

C3 #10 Dist: S1: 1.737, C5: 1.560

- **FOURIER RECYCLING (default for inorganic)**
- **RBM (default for organic/metal-organic)**
- COVMAP
- RAMM
- SHIFT_AND_FIX

LEAST SQUARES-FOURIER RECYCLING FOR INORGANIC COMPOUNDS

The model optimization default procedure for inorganic compound consists of eight blocks automatically executed; each block is constituted by weighted least squares (WLSQ) cycles combined with Fourier recycling (FR) of ($2|F_o|-|F_c|$) map calculation ($|F_o|$ and $|F_c|$ are the observed and calculated structure factor moduli, respectively)

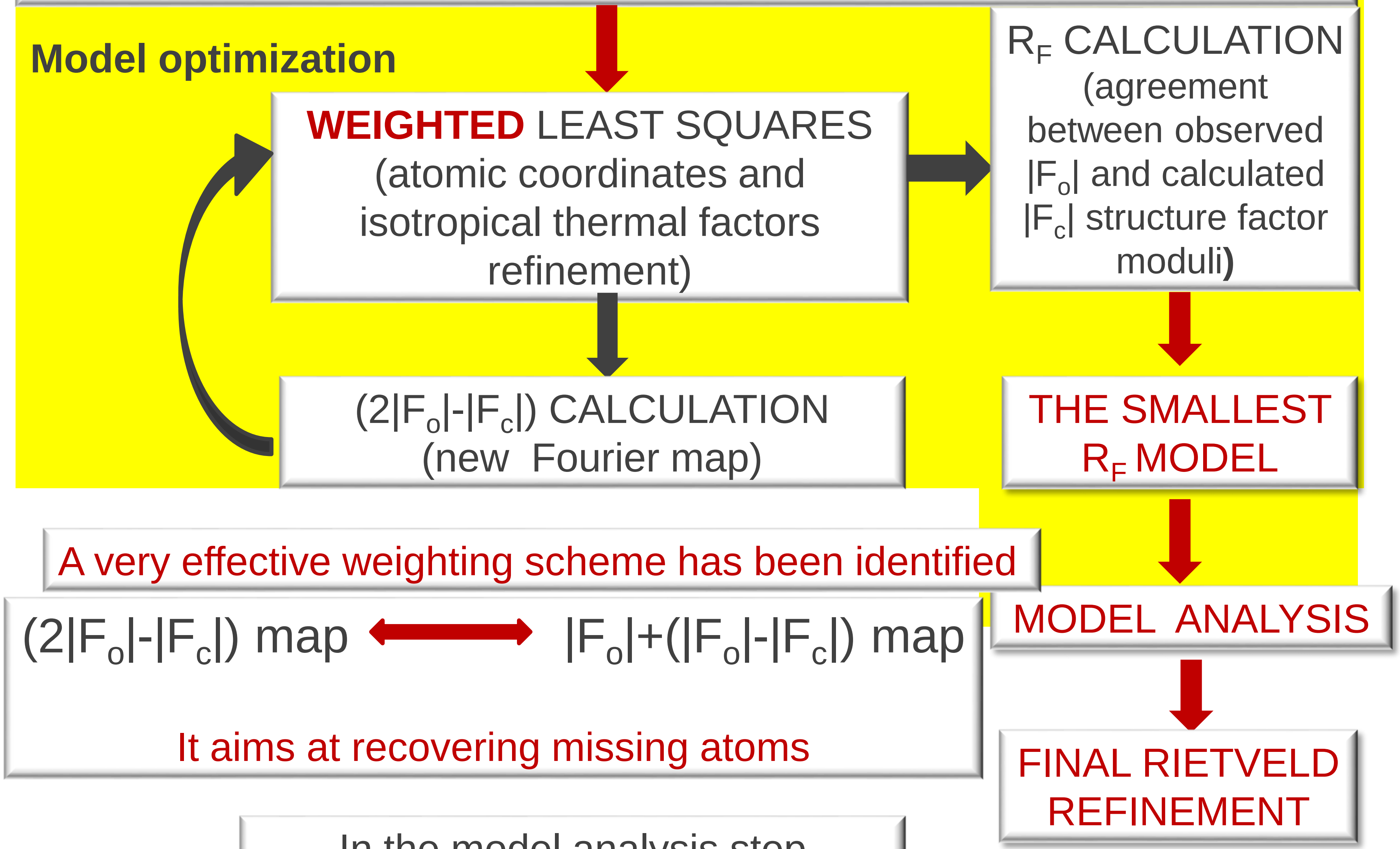
For each block, at the end of least squares, the R_F residual value is evaluated

$$R_F = \frac{\Sigma_h \left| \left|F_{oh}\right| - s \cdot \left|F_{ch}\right| \right|}{\Sigma_h \left|F_{oh}\right|}$$

At the end of the eight blocks, the structure model corresponding to the smallest R_F residual value is selected

The model is carefully analyzed: poorly connected low peak intensity atoms characterized by large refined thermal factors are discarded

LEAST SQUARES-FOURIER RECYCLING FOR INORGANIC COMPOUNDS



Model optimization

WEIGHTED LEAST SQUARES (atomic coordinates and isotropical thermal factors refinement)

($2|F_o|-|F_c|$) CALCULATION (new Fourier map)

R_F CALCULATION (agreement between observed $|F_o|$ and calculated $|F_c|$ structure factor moduli)

THE SMALLEST R_F MODEL

MODEL ANALYSIS

FINAL RIETVELD REFINEMENT

A very effective weighting scheme has been identified

($2|F_o|-|F_c|$) map ⟷ $|F_o|+(|F_o|-|F_c|)$ map

It aims at recovering missing atoms

In the model analysis step unreliable peaks are discarded.

LEAST SQUARES FOR SINGLE CRYSTAL DATA

Least squares in structure refinement by single crystal data are usually carried out for minimizing

$$R = \sum_{\mathbf{h}} w_{\mathbf{h}} \left(|F_{o\mathbf{h}}| - |F_{c\mathbf{h}}| \right)^2$$

The summation is over the total number of measured reflections

- $w_{\mathbf{h}}$ is a suitable weight associated to each reflection; it usually represents an estimate of the precision of the measured quantity
- $|F_{o\mathbf{h}}|$ is the measured (experimental) structure factor modulus
- $|F_{c\mathbf{h}}|$ is the calculated structure factor modulus

LEAST SQUARES FOR SINGLE CRYSTAL DATA

If $|F_{c\mathbf{h}}|$ is expanded by Taylor series

$$R = \sum_{\mathbf{h}} w_{\mathbf{h}} \left(\Delta |F_{\mathbf{h}}| - \sum_k \frac{\delta |F_{c\mathbf{h}}|}{\delta x_k} \Delta x_k \right)^2 ,$$

(x are the parameters to be refined)

The normal equations are obtained by setting to zero the derivatives of R with respect to the parameters to be refined

The shift Δx_k on parameters can be derived

$$x_k = x_k + \Delta x_k$$ refined parameters

LEAST SQUARES FOR SINGLE CRYSTAL DATA

$$F_{\mathbf{h}} = \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \exp[-B_j \sin^2\theta/\lambda^2]$$

f_j is the scattering factor of the *j-th* atom.

(the summation is over the number of atoms in the unit cell).

Atomic coordinates and

thermal factors are refined.

In case of single crystal, the observed structure factor moduli estimates are reliable and least squares are able to improve atom positions

The success of least squares strongly depends on the ratio ‘*number of observations (reflections)/number of parameters to be refined*’

LEAST SQUARES FOR POWDER DATA

In case of powder data, the observed structure factor moduli estimates are affected by errors (that is particularly true for overlapping reflections). In addition, the number of observations is usually small (particularly for organic structures because of their rapid scattering factor decay)

Single crystal type least squares may be unsuccessful

Special strategies can be adopted in terms of:

- quantity to be minimized
- weighting scheme

LEAST SQUARES FOR POWDER DATA

In case of powder data, for the overlapping reflections, the intensity of a cluster is more reliable than the single intensity. We developed an approach based on the use of the cluster intensities as observations in the least squares

That choice may be not profitable because of the small number of clusters (observations)

A more effective approach, available in the current version, was developed and introduced in EXPO

LEAST SQUARES FOR POWDER DATA

The main features of the EXPO approach are

- least squares minimization on reflection integrated intensities
- a robust weighting scheme capable of compensating the low accuracy of the overlapping reflection intensities
- a delayed partitioning of the observed intensity of each cluster of overlapped reflections proportionally to the calculated intensity of each reflection

LEAST SQUARES FOR POWDER DATA

In EXPO, least squares for optimizing the structure model aim at minimizing the following residual

$$R = \sum_{\mathbf{h}} w_{\mathbf{h}} (I_{o\mathbf{h}} - I_{c\mathbf{h}})^2$$

The summation is over the total number of reflections in the experimental pattern

- $w_{\mathbf{h}}$ is a suitable weight
- $I_{o\mathbf{h}}$ is the observed integrated intensity for the reflection **h**, extracted from the experimental pattern ($I_{o\mathbf{h}} \propto |F_{o\mathbf{h}}|^2$)
- $I_{c\mathbf{h}}$ is the corresponding intensity calculated from the model ($I_{c\mathbf{h}} \propto |F_{c\mathbf{h}}|^2$)

$w_{\mathbf{h}}$ is able to take into account the unavoidable errors on the integrated intensity estimates of overlapping reflections

LEAST SQUARES WEIGHTING SCHEME

$$w_{\mathbf{h}} = w_{1\mathbf{h}} \cdot w_{2\mathbf{h}}$$

$$w_{1\mathbf{h}} = w_{a\mathbf{h}} \cdot w_{b\mathbf{h}} \qquad w_{2\mathbf{h}} = 1/[1+4\cdot(m_{\mathbf{h}}\cdot LP_{\mathbf{h}}\cdot|F_{o\mathbf{h}}|)^2]$$

$$w_{a\mathbf{h}} = \exp[-0.5\cdot(\sin\theta_{\mathbf{h}}/\lambda)^2] \qquad w_{b\mathbf{h}} = \sqrt{I_{o\mathbf{h}}/I'_{o\mathbf{h}}}$$

$m_{\mathbf{h}}$ is the reflection multeplicity; $LP_{\mathbf{h}}$ is the Lorentz-polarization factor; $\theta_{\mathbf{h}}$ is the reflection angular position (Bragg angle); λ is the radiation wavelength;

$I_{o\mathbf{h}} = m_{\mathbf{h}}\cdot LP_{\mathbf{h}} \cdot|F_{o\mathbf{h}}|^2$

$$I'_{o\mathbf{h}} = I_{o\mathbf{h}} + \sum_{\mathbf{k}\neq\mathbf{h}} (1-q^2_{\mathbf{kh}})\cdot I_{o\mathbf{k}}, \quad q^2_{\mathbf{kh}} = \frac{|2\theta_{\mathbf{k}} - 2\theta_{\mathbf{h}}|}{FWHM_{\mathbf{h}}}$$

The summation is over the reflections in overlapping with **h**
$FWHM_{\mathbf{h}}$ is the full width at half maximum of the **h** reflection peak

LEAST SQUARES WEIGHTING SCHEME

$$w_{1\mathbf{h}} = w_{a\mathbf{h}} \cdot w_{b\mathbf{h}} \quad w_{a\mathbf{h}} = \exp[-0.5 \cdot (\sin\theta_{\mathbf{h}}/\lambda)^2] \quad w_{b\mathbf{h}} = \sqrt{I_{o\mathbf{h}}/I'_{o\mathbf{h}}}$$

$w_{1\mathbf{h}}$ decreases with

- increasing θ_h (it is more difficult to carefully model overlapping and background at high $2\theta_h$ angle region of the experimental pattern)
- increasing the number of reflections in overlapping (in particular, when the overlapping neighbouring intensities are large with respect to I_{oh})

$$w_{2\mathbf{h}} = 1/[1+4\cdot(m_{\mathbf{h}} \cdot LP_{\mathbf{h}} \cdot |F_{o\mathbf{h}}|)^2]$$

$w_{2\mathbf{h}}$ is the typical weight used to avoid that least squares are dominated by the largest intensities

FOURIER MAP CALCULATION

We try to improve the integrated intensity I_{oh} estimate ($I_{oh} \propto |F_{oh}|^2$) associated to each **h** overlapping reflection, before the ($2|F_o|-|F_c|$) Fourier calculation and after least squares

The observed intensity of each cluster of overlapping reflections is ‘partially’ partitioned proportionally to the calculated intensity of each reflection in the cluster

$$I^{p}_{o\mathbf{h}} = w_{1\mathbf{h}} \cdot I_{o\mathbf{h}} + K \cdot (1 - w_{1\mathbf{h}}) \cdot I_{c\mathbf{h}}$$

K is a scale factor, given by

$$K = \frac{\sum_{\mathbf{k}} (1 - w_{1\mathbf{k}}) \cdot I_{o\mathbf{k}}}{\sum_{\mathbf{k}} (1 - w_{1\mathbf{k}}) \cdot I_{c\mathbf{k}}}$$

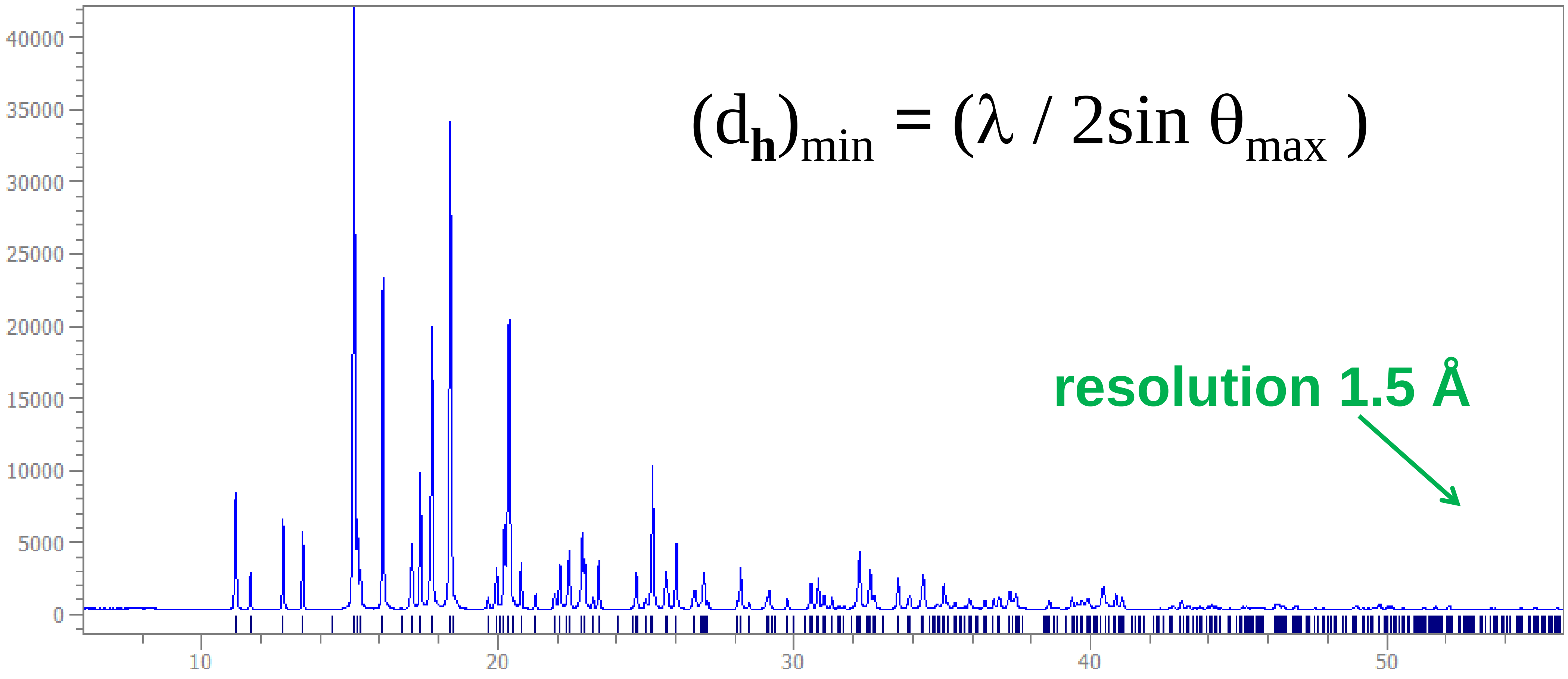
The summation goes over the reflections belonging to the cluster

K fixes the condition that the overall intensity of the cluster must not be changed after the partition

If the reflection intensity I_{oh} is well estimated (w_{1h} is ~1), $I^{p}_{o\mathbf{h}}$ slightly differs from the observed I_{oh}; on the contrary, if w_{1h} ~ 0, $I^{p}_{o\mathbf{h}}$ is proportional to the calculated intensity

MODEL OPTIMIZATION FOR ORGANIC COMPOUNDS

For organic compounds the experimental resolution *RES* ($d_{\mathbf{h}}$, the smallest interplanar distance value) is usually far away from the atomic one because of their rapid scattering factor decay



$$(d_{\mathbf{h}})_{min} = (\lambda \,/\, 2\sin\theta_{max})$$

resolution 1.5 Å

MODEL OPTIMIZATION FOR ORGANIC COMPOUNDS

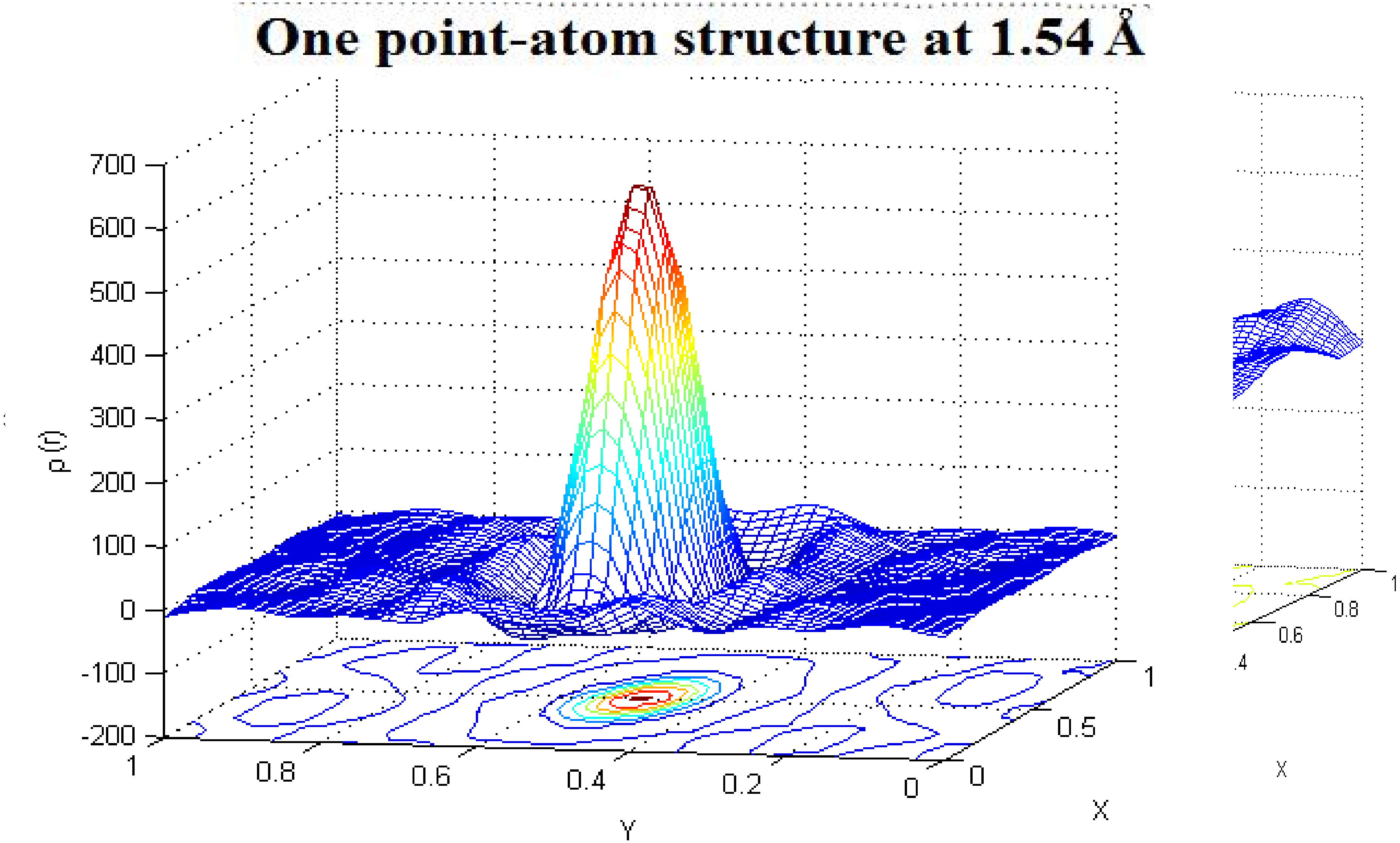
It is well known that limited experimental data resolution modifies the Fourier synthesis (even at atomic resolution)

$$\rho(\mathbf{r}) = \frac{2}{V} \sum_{h=0}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \left|F_{\boldsymbol{H}}\right| \cos\left[\left(\varphi_{\boldsymbol{H}} - 2\pi\, \overline{\boldsymbol{H}} \cdot \mathbf{r}\right)\right]$$

The map is an imperfect representation of the true electron density: the worse the experimental resolution, the less accurate the Fourier representation of the electron density

When *RES*>1 Å the resolution effects may be critical

MODEL OPTIMIZATION FOR ORGANIC COMPOUNDS



MODEL OPTIMIZATION FOR ORGANIC COMPOUNDS

- Regions of the electron density map in the unit cell become negative

- Peaks are broadened and surrounded by ripples (positive and negative with decreasing amplitude)

- Peaks are misplaced from their correct positions

- Peaks have distorted intensities

RESOLUTION BIAS MINIMIZATION PROCEDURE RBM

RBM is a very powerful procedure of EXPO providing an electron density map more informative with respect to the case in which the resolution effects have not been corrected

RBM is able to face the perennial problem of limited experimental data resolution. In case of powder, the resolution bias is mixed with errors on both structure factor moduli and phases

RBM corrects the electron density map by typical resolution bias errors: peak broadening, ripples, intensity distortion, peak shifts

RBM PROCEDURE

The method requires the generalization of the traditional concept of gaussian-like peak: each *j*-th atomic peak, in the electron density map, is replaced by a two-component function, constituted by the *main peak* and by the corresponding *ripples*

$$\rho_j(\mathbf{r}-\mathbf{r}_j) = \rho_{[\mathrm{main}]_j}(\mathbf{r}-\mathbf{r}_j) + \rho_{[\mathrm{ripples}]_j}(\mathbf{r}-\mathbf{r}_j)$$

Each component of the generalized peak is mathematically modelled in its corresponding existence domain: *e.g.*, the *j*-th main peak has its own existence domain while the ripples extend over all the rest of the unit cell

RBM PROCEDURE

RBM consists of the two following main steps

- subtracting ripples corresponding to each experimental electron density peak

$$\rho(\mathbf{r}) = \rho(\mathbf{r}) - \sum_j \rho_{[\text{ripples}]\, j}(\mathbf{r} - \mathbf{r}_j)$$

- best fitting each *j*-th main peak in its own domain by gaussian function

$$\rho(\mathbf{r}) \simeq \sum_{j=1}^{N} c_j G(\mathbf{r}; \sigma_j, \mathbf{r}_j)$$

MODEL OPTIMIZATION FOR ORGANIC COMPOUNDS

The EXPO default procedure for optimizing structure model of organic and metal-organic compounds is based on successive cycles of

$2|F|_o-|F|_c$ map calculation suitably corrected by RBM

and

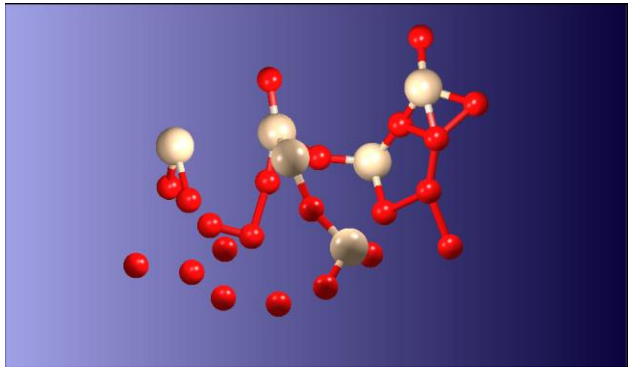
weighted least squares (restraints on distances and angles are applied)

EXPO automatically recognizes a compound as organic (or metallorganic) or inorganic and applies the designed model optimization strategy

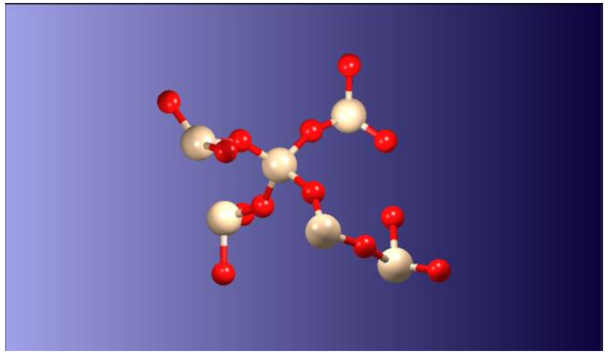
MODEL OPTIMIZATION IN EXPO

DM model

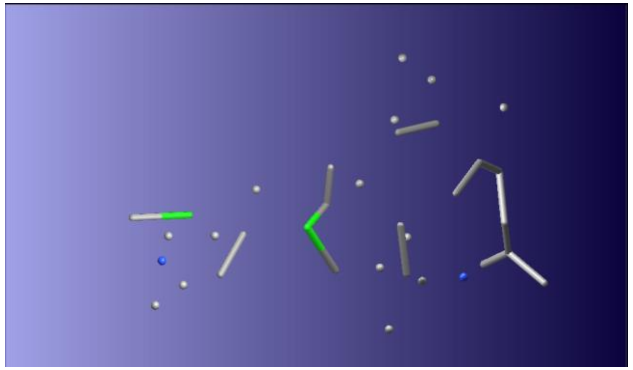
UTM-1 zeolite



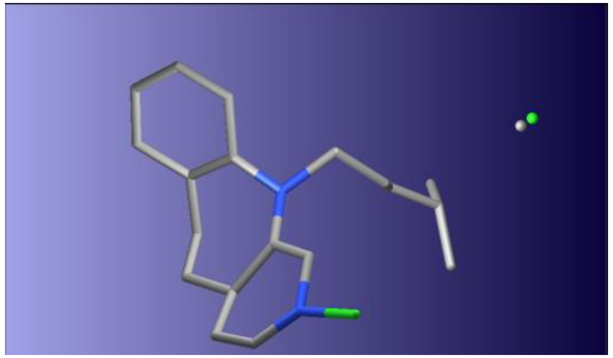
WLSQ-FR model



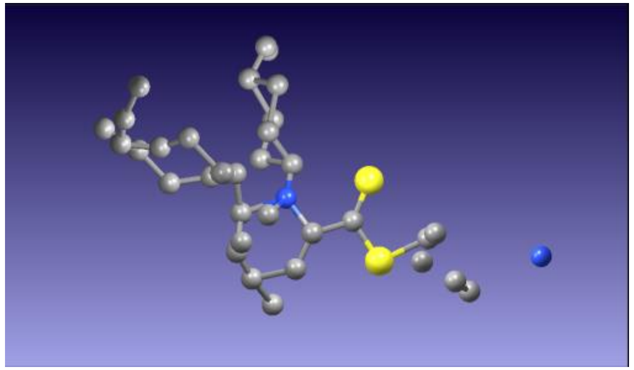
Clomipramine hydrochloride $C_{19}H_{24}ClN_2{\cdot}Cl$



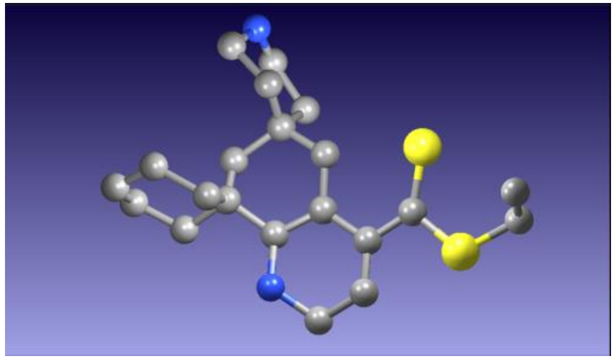
RBM model



Ethyl 1' ,2' ,3' ,4' ,4a' ,5' ,6' ,7' -octahydrodispiro[cyclohexane-1,2' -quinazoline-4' ,1'' -cyclohexane]-8' -carbodithioate



RBM model



THE RESOLUTION BIAS CORRECTION IN RECIPROCAL SPACE

The correction by RBM in the direct space can be translated into the reciprocal space

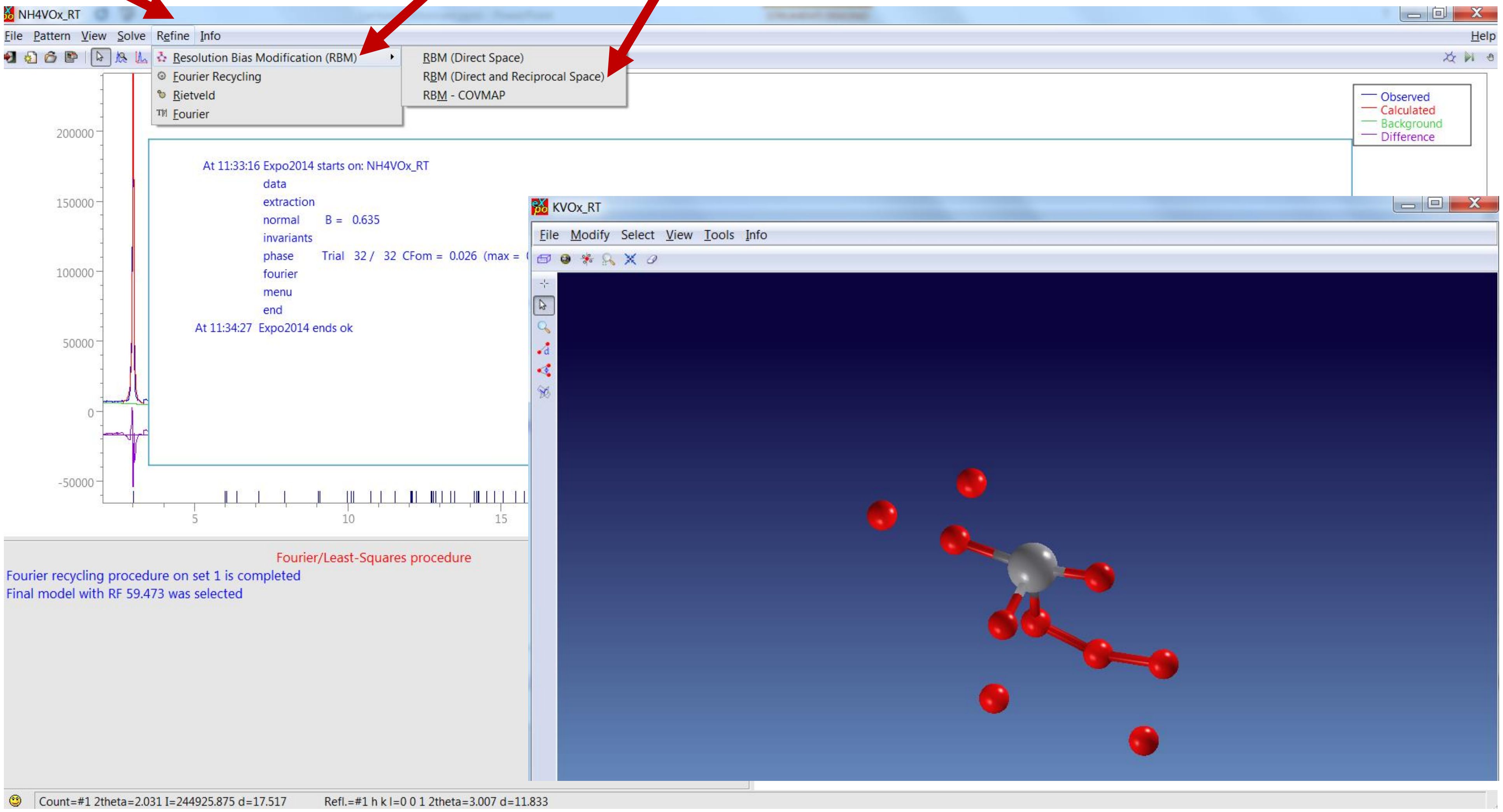
The structure factors can be modified (resolution dependent correction) in order to provide electron density map showing truncation effects smaller than the canonical case

$$\begin{tikzcd}
F_{\mathbf{h}} \arrow[r, bend left, "\text{Reciprocal Space}\ \Sigma"] & \rho(\boldsymbol{r}) \arrow[rd, bend left, "\text{Direct Space}"] & \\
F_{\mathbf{h}}^{\mathbf{mod}} \arrow[rr, bend right, "\Sigma", "\text{Reciprocal Space}"'] & & \rho'(\boldsymbol{r})
\end{tikzcd}$$

The use of the modified structure factors locates more reliable peaks

Also this non-default strategy (alternative to standard RBM in direct space), mainly based on atomic scattering factor correction, is available in EXPO for reducing the typical resolution bias errors of the electron density map

Refine ->->-> RBM ->->-> RBM (Direct and Reciprocal Space)



```
At 11:33:16 Expo2014 starts on: NH4VOx_RT
          data
          extraction
          normal      B =   0.635
          invariants
          phase       Trial   32 /   32  CFom =  0.026  (max =
          fourier
          menu
          end
At 11:34:27  Expo2014 ends ok
```

Fourier/Least-Squares procedure

Fourier recycling procedure on set 1 is completed

Final model with RF 59.473 was selected

Count=#1 2theta=2.031 I=244925.875 d=17.517   Refl.=#1 h k l=0 0 1 2theta=3.007 d=11.833

THE COVMAP METHOD
(based on the COVariance of the electron density MAP)

COVMAP is a non-default strategy in EXPO for optimizing a structure model obtained at the end of an *ab initio* solution process, in particular for recovering missing positions

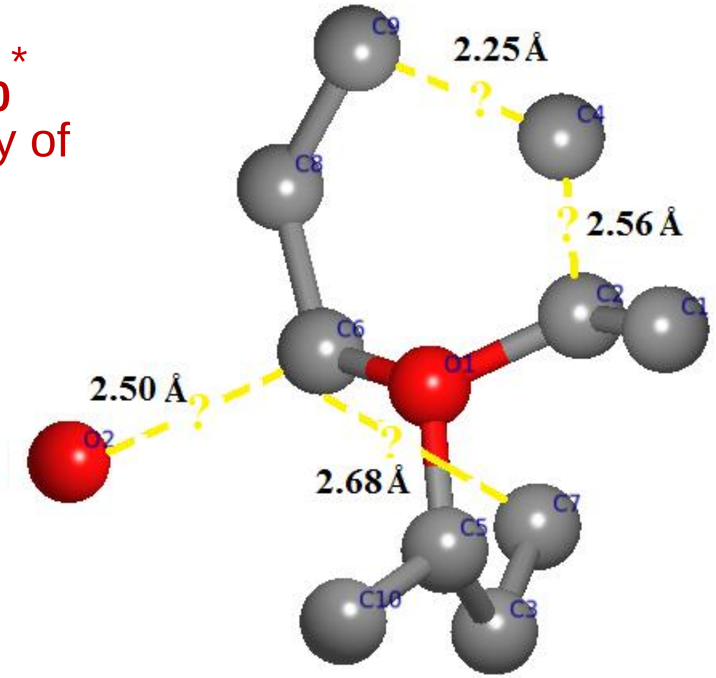
COVMAP starts from the available non-interpretable structural model containing very few well positioned and a large set of misplaced peaks

THE COVMAP METHOD

A quite common case

CAMPHOR O_2C_{10}*

*unpubl. data by courtesy of Dr. M. Brunelli



checking distances?

Only a few atoms (the yellow ones) are close to their correct positions, the others are incorrectly located or are so far from the correct positions that chemical bonds cannot be established

COVMAP is able to improve such a model

THE COVMAP METHOD

COVMAP has been inspired by a mathematical technique enabling to calculate the covariance and/or correlation between two points of an electron density map, no matter the quality of the structure model

The technique operates in the reciprocal space, by exploiting suitable Fourier series involving structure factors as components of the Fourier coefficients

COVMAP translates the covariance concept in direct space

COVARIANCE CONCEPT

the density in one point depends on the density in another point of the map if their covariance is not vanishing

THE COVMAP METHOD

Around each positioned atom (the pivot atom), at bond distance, we expect that one or more atoms are present in the correct structure (the condition is violated in case of isolated atom)

COVMAP introduces the covariance concept: the largest intensity peaks in the electron density map are considered pivots, and for each of them, the reasonable expectation that some other peaks would be present at bond distance is transformed in electron density modifications

THE COVMAP METHOD

Three kinds of information are exploited by COVMAP

1. the chemical interpretation of the structure model peaks (according to their intensities in the electron density map)

2. some basic crystal-chemical rules, essentially the bond distances expected for the pivot atoms

3. the crystallographic residual R_F corresponding to the model modified by COVMAP. It is expected to be smaller than R_F corresponding to the original improvable model

THE COVMAP METHOD

All the **pivot pairs (i, j)** of non-bonded atom positions, with distances compatible (within the electron density errors) with the presence of one atom bridging i and j atoms, are considered

$\rho_c(\mathbf{r})$ is the electron density map corresponding to the current model

$\rho_d(\mathbf{r})$ is the difference electron density map

A **new working map** $\rho_n(\mathbf{r})$ (first set to 0) is created:

If $\rho_c(\mathbf{r}) > 0$ and $\rho_d(\mathbf{r}) > 0$

$$\rho_n(\mathbf{r}) = \rho_c(\mathbf{r}) \cdot w(\mathbf{r})$$

w(r) (integer positive factor) is the number of times in which the positional vector **r** falls within the range of the expected bond distance from each position of each pivot pair
w(r) vanishes in the points where atoms are not expected to be connected with the pivot pairs

THE COVMAP BASED METHOD

A peak search on $\rho_{\mathrm{n}}(\mathbf{r})$ is carried out and new peak positions (*nd*) are added to the starting number (*ns*) of the improvable model

The model constituted by *ns*+*nd* peaks is finally retained and chemically interpreted if

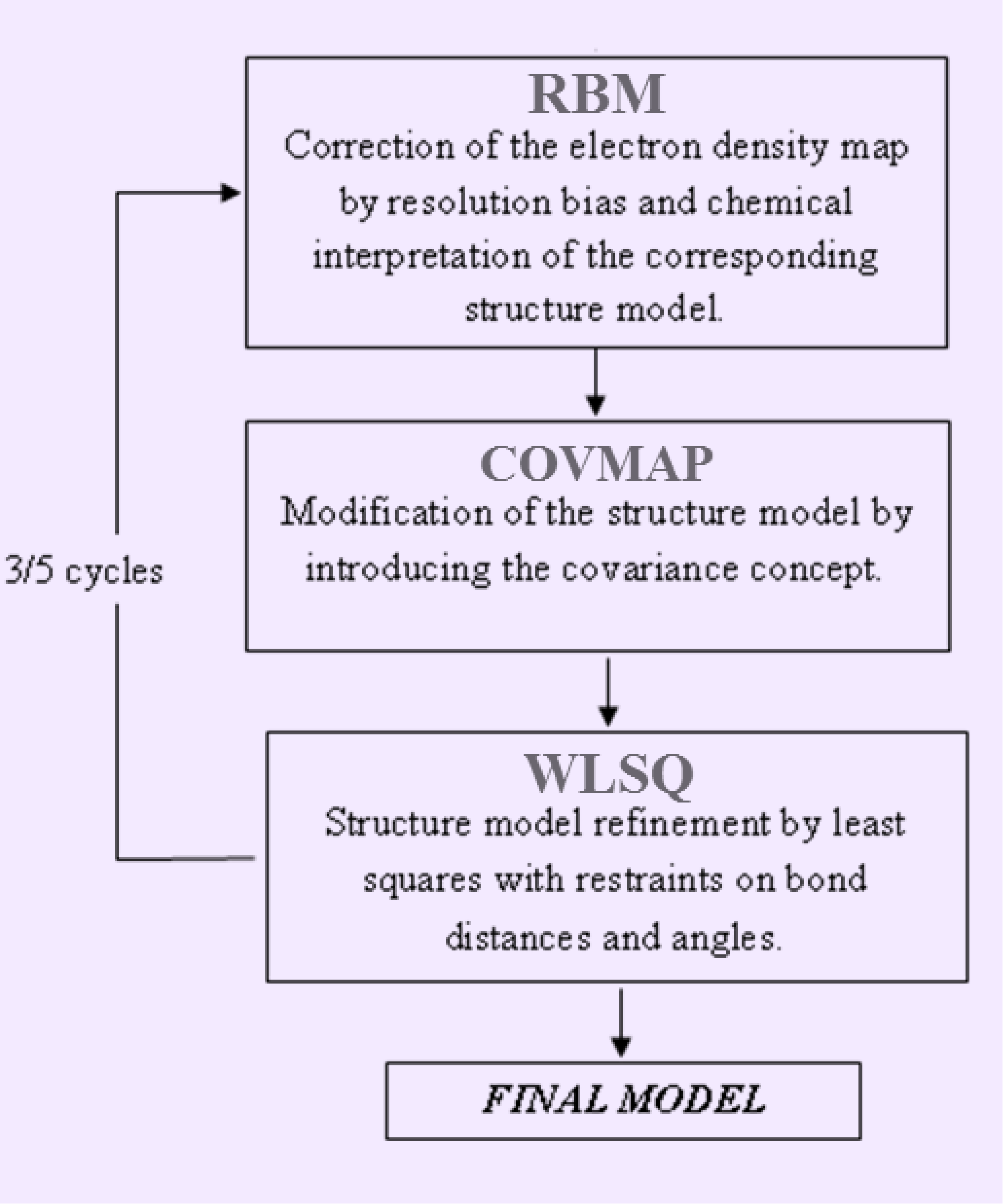
$$R_{Fns+nd}<1.2\cdot R_{Fns}$$

otherwise the work by COVMAP is cancelled.

COVMAP has been suitably combined with WLSQ (Weighted Least Squares) and RBM (Resolution Bias Minimization) procedures, which relocate, refine or reject the new peaks

This strategy highlights the benefits obtained by COVMAP

THE COVMAP BASED METHOD



RBM

Correction of the electron density map by resolution bias and chemical interpretation of the corresponding structure model.

COVMAP

Modification of the structure model by introducing the covariance concept.

WLSQ

Structure model refinement by least squares with restraints on bond distances and angles.

FINAL MODEL

THE COVMAP BASED METHOD

- **RBM** provides the structure model to COVMAP

- **COVMAP** locates new atomic positions

- **WLSQ** validate the new peaks

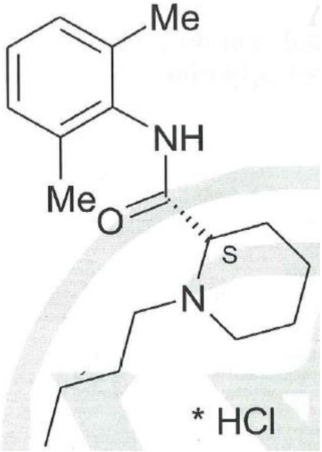
- **RBM** relocates in a better way and/or rejects the survived positions and so on cyclically

COVMAP is able to introduce supplementary structural information; RBM and WLSQ can discard false and improve good COVMAP indications

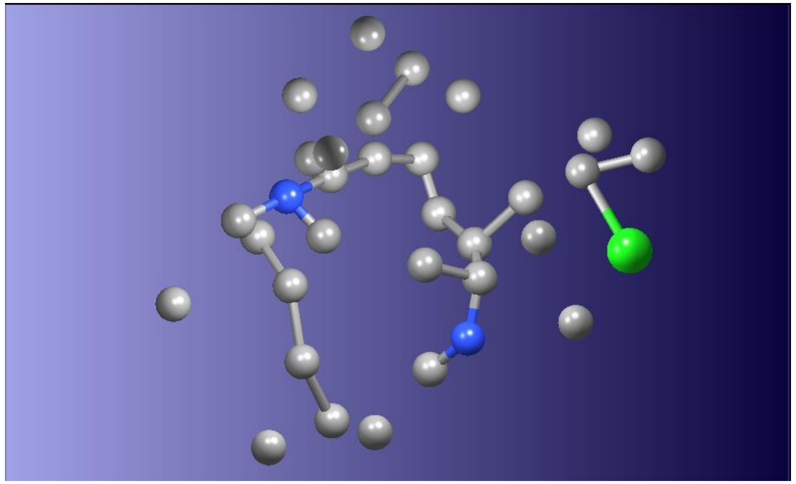
THE COVMAP BASED METHOD

S-BUPIVACAINE
HYDROCLORIDE
$C_{18}H_{28}N_2O{\cdot}HCl$
22 non-H positions in the
asymmetric unit

RES=1.46 Å

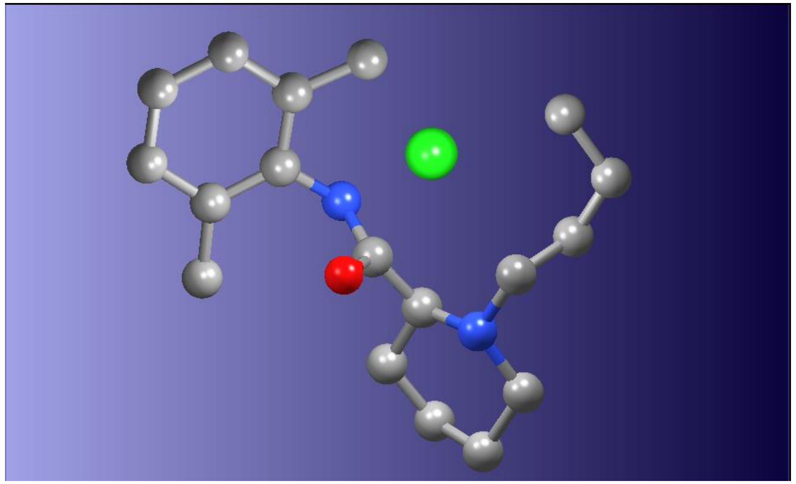


DM



ONLY 13 POSITIONS CLOSE
TO THE TRUE ONES

RBM-COVMAP-WLSQ



SOLUTION OBTAINED IN ~2 minutes

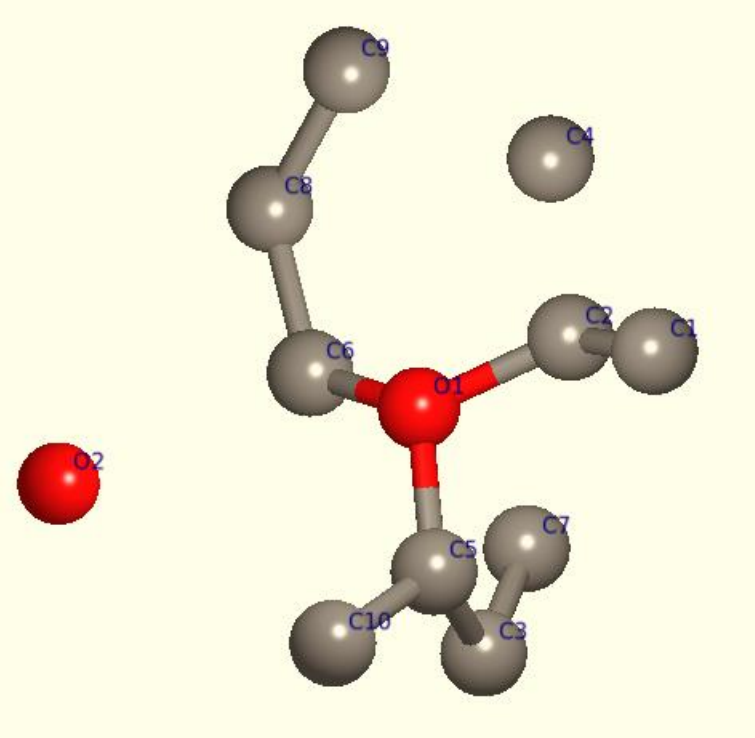
THE COVMAP BASED METHOD

CAMPHOR

O_2C_{10}

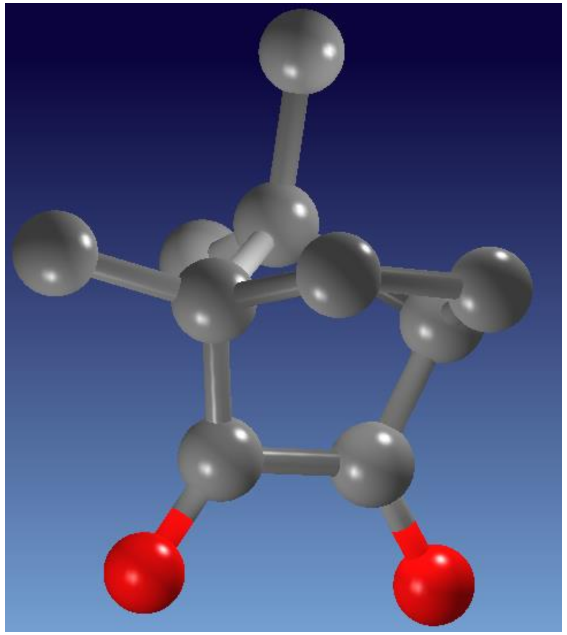
12 non-H positions in the asymmetric unit

DM

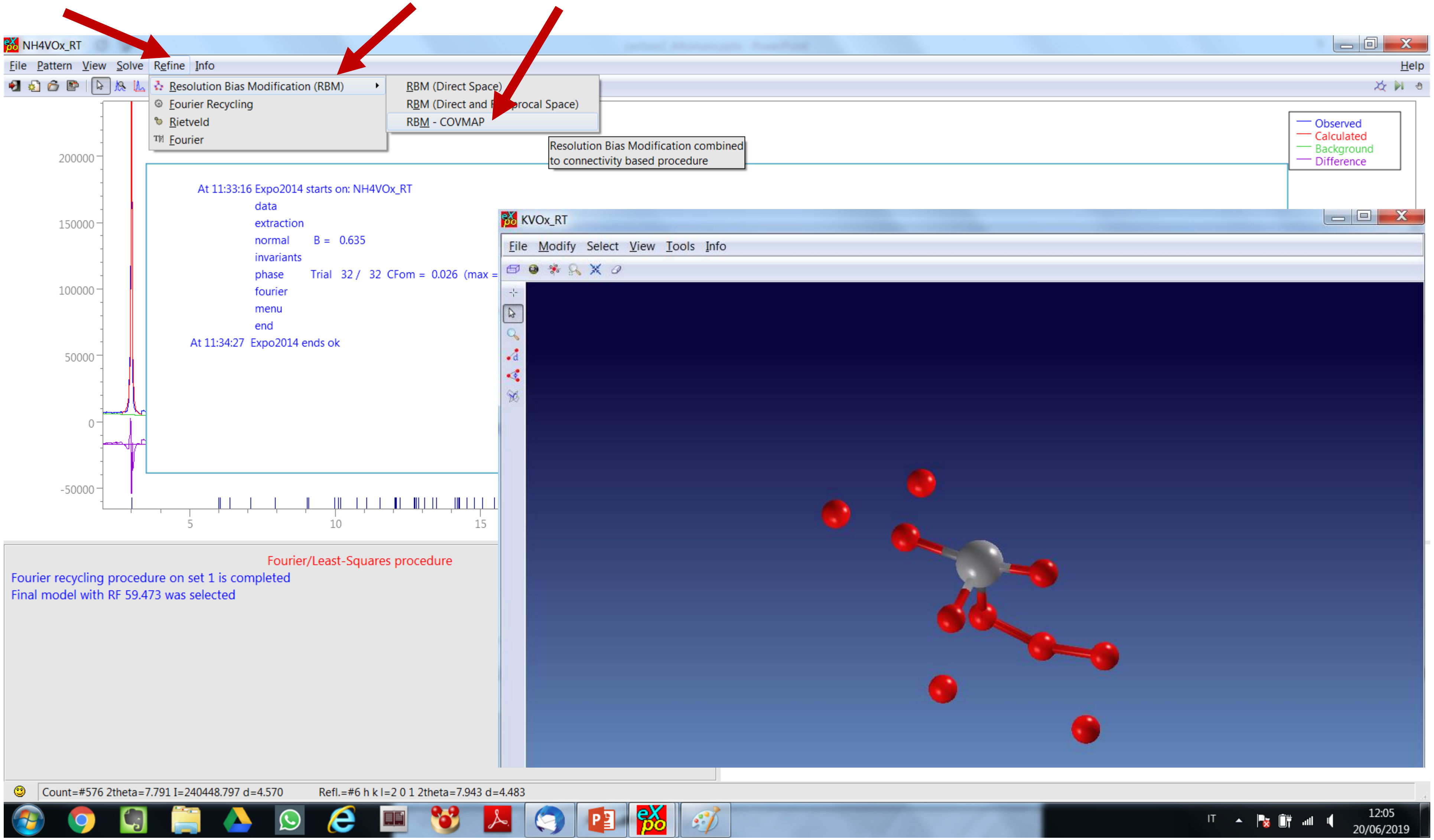


ONLY 3 POSITIONS CLOSE TO THE TRUE ONES

RBM-COVMAP-WLSQ



Refine ->->-> RBM ->->-> RBM - COVMAP



```
At 11:33:16 Expo2014 starts on: NH4VOx_RT
          data
          extraction
          normal     B =   0.635
          invariants
          phase      Trial   32 /   32  CFom =  0.026  (max =
          fourier
          menu
          end
At 11:34:27  Expo2014 ends ok
```

```
Fourier/Least-Squares procedure
Fourier recycling procedure on set 1 is completed
Final model with RF 59.473 was selected
```

```
Count=#576 2theta=7.791 I=240448.797 d=4.570      Refl.=#6 h k l=2 0 1 2theta=7.943 d=4.483
```

THE RAMM METHOD

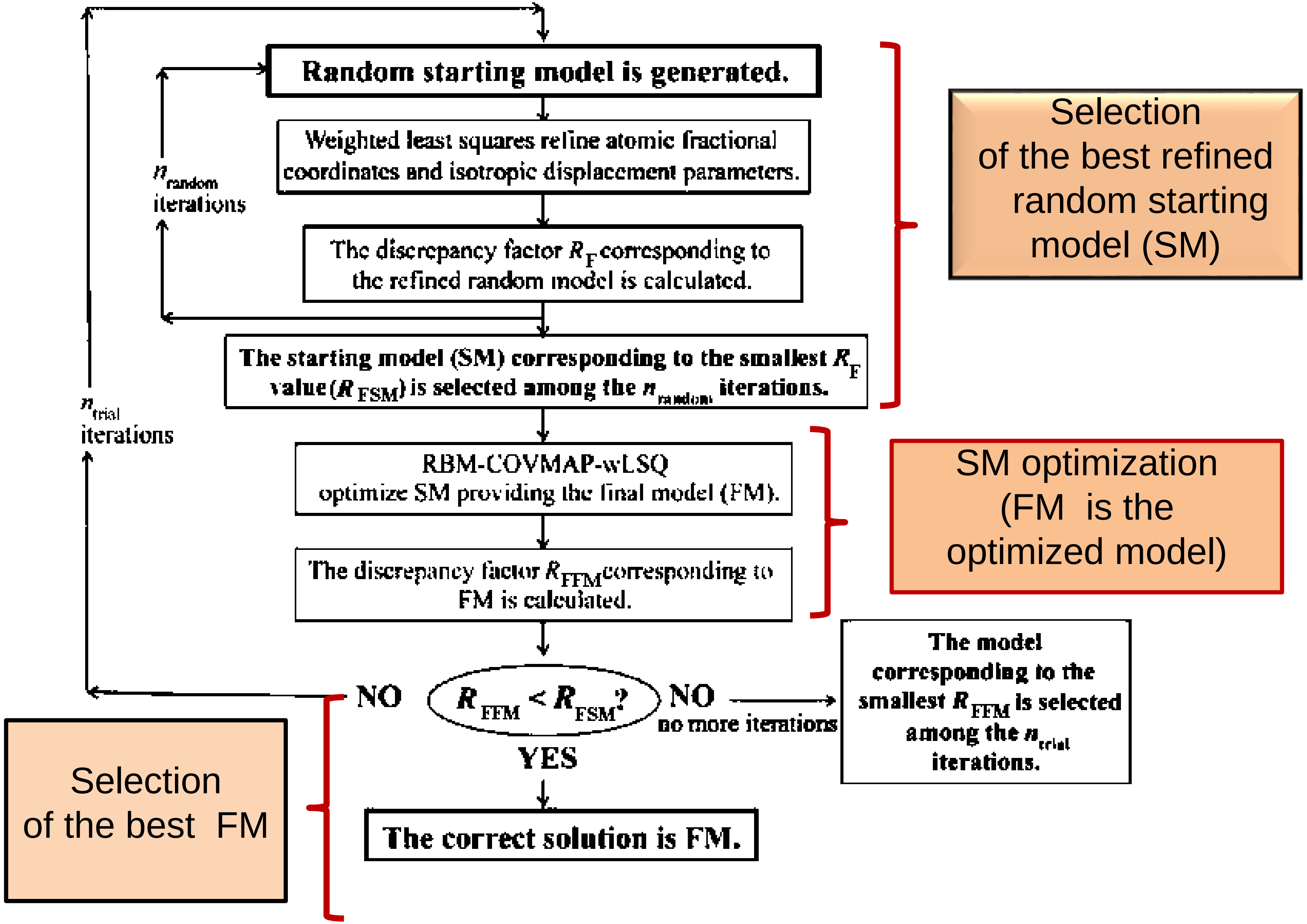
RBM-COVMAP-WLSQ is a very powerful tool for recovering the full structure from approximate structure models

If we start from a random model (not from DM model)?

RAMM: a new RAndom Model based Method for solving *ab initio* crystal structure by EXPO package

RAMM skips Direct Methods
and applies RBM-COVMAP-WLSQ to fully random starting models

THE RAMM METHOD



Random starting model is generated.

Weighted least squares refine atomic fractional coordinates and isotropic displacement parameters.

The discrepancy factor R_F corresponding to the refined random model is calculated.

n_{random} iterations

The starting model (SM) corresponding to the smallest R_F value (R_{FSM}) is selected among the n_{random} iterations.

Selection of the best refined random starting model (SM)

RBM-COVMAP-wLSQ optimize SM providing the final model (FM).

The discrepancy factor R_{FFM} corresponding to FM is calculated.

SM optimization (FM is the optimized model)

n_{trial} iterations

NO $R_{FFM} < R_{FSM}$? **NO** no more iterations

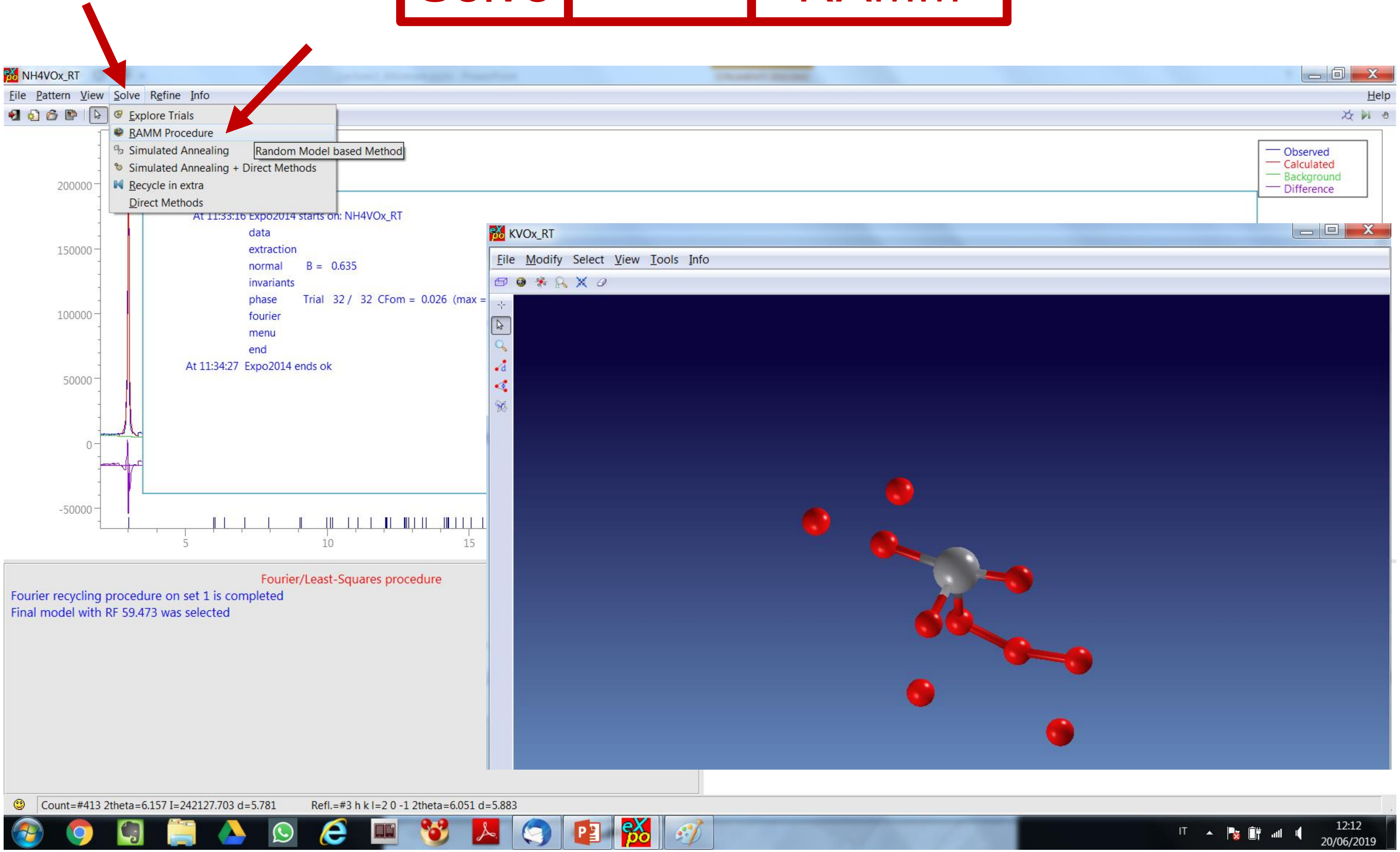
The model corresponding to the smallest R_{FFM} is selected among the n_{trial} iterations.

YES

The correct solution is FM.

Selection of the best FM

Solve ->->-> RAMM



```
At 11:33:16 Expo2014 starts on: NH4VOx_RT
            data
            extraction
            normal     B =   0.635
            invariants
            phase      Trial   32 /   32  CFom =  0.026  (max =
            fourier
            menu
            end
At 11:34:27  Expo2014 ends ok
```

```
                              Fourier/Least-Squares procedure
Fourier recycling procedure on set 1 is completed
Final model with RF 59.473 was selected
```

```
Count=#413 2theta=6.157 I=242127.703 d=5.781        Refl.=#3 h k l=2 0 -1 2theta=6.051 d=5.883
```

THE RAMM METHOD

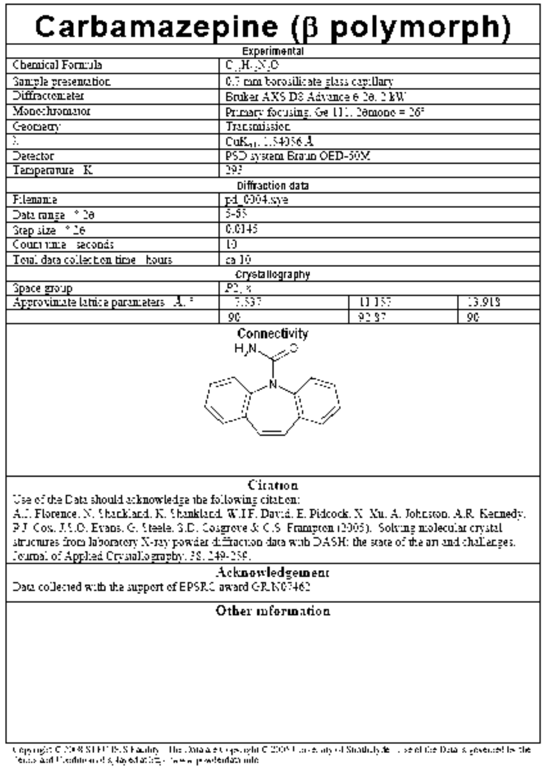
RAMM is a useful approach alternative to Direct Methods in case of their failure

RAMM is usually able not only to find the correct solution, but also to recognize it

It gives results comparable to and in some cases better than those obtained by Direct Methods

RAMM requires long execution time (it increses with the number of atoms in the asymmetric unit)

THE RAMM METHOD



Carbamazepine (β polymorph)

Experimental	
Chemical Formula	$C_{15}H_{12}N_2O$
Sample presentation	0.7 mm borosilicate glass capillary
Diffractometer	Bruker AXS D8 Advance 6 2θ, 2 kW
Monochromator	Primary focusing, Ge 111, 2θmono = 26°
Geometry	Transmission
λ	CuK$_{\alpha1}$, 1.54056 Å
Detector	PSD system Braun OED-50M
Temperature K	293

Diffraction data	
Filename	pd_0004.xye
Data range ° 2θ	5-55
Step size ° 2θ	0.0145
Count time seconds	10
Total data collection time hours	ca 10

Crystallography			
Space group	$P2_1/n$		
Approximate lattice parameters Å, °	7.537	11.157	13.918
	90	92.87	90

Connectivity

Citation

Use of the Data should acknowledge the following citation:
A.J. Florence, N. Shankland, K. Shankland, W.I.F. David, E. Pidcock, X. Xu, A. Johnston, A.R. Kennedy, P.J. Cox, J.S.O. Evans, G. Steele, S.D. Cosgrove & C.S. Frampton (2005). Solving molecular crystal structures from laboratory X-ray powder diffraction data with DASH: the state of the art and challenges. Journal of Applied Crystallography, 38, 249-259.

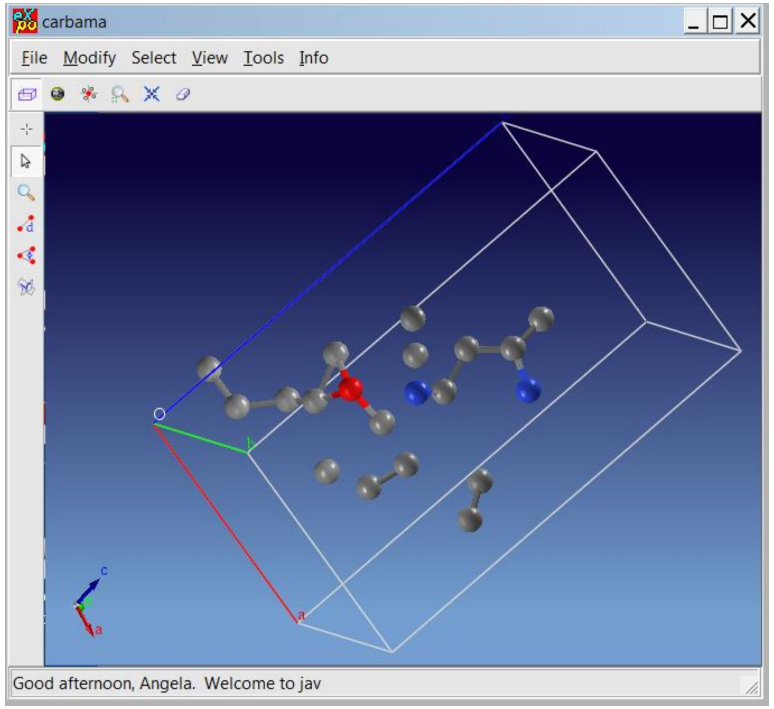
Acknowledgement

Data collected with the support of EPSRC award GR/N07462

Other information

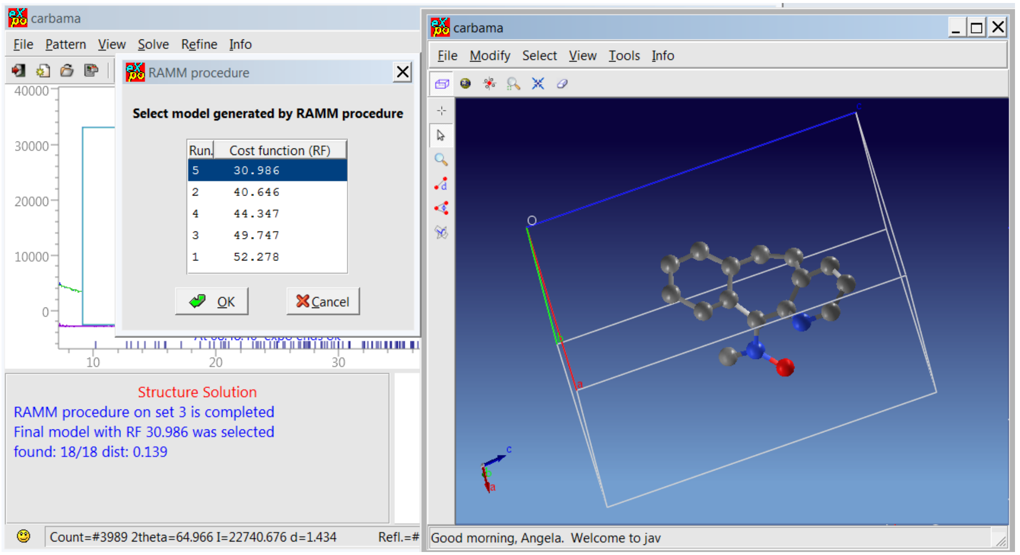
18 non-H positions in the asymmetric unit

EXPO DEFAULT



Good afternoon, Angela. Welcome to jav

RAMM



Select model generated by RAMM procedure

Run	Cost function (RF)
5	30.986
2	40.646
4	44.347
3	49.747
1	52.278

Structure Solution
RAMM procedure on set 3 is completed
Final model with RF 30.986 was selected
found: 18/18 dist: 0.139

Count=#3989 2theta=64.966 I=22740.676 d=1.434 Refl.=#

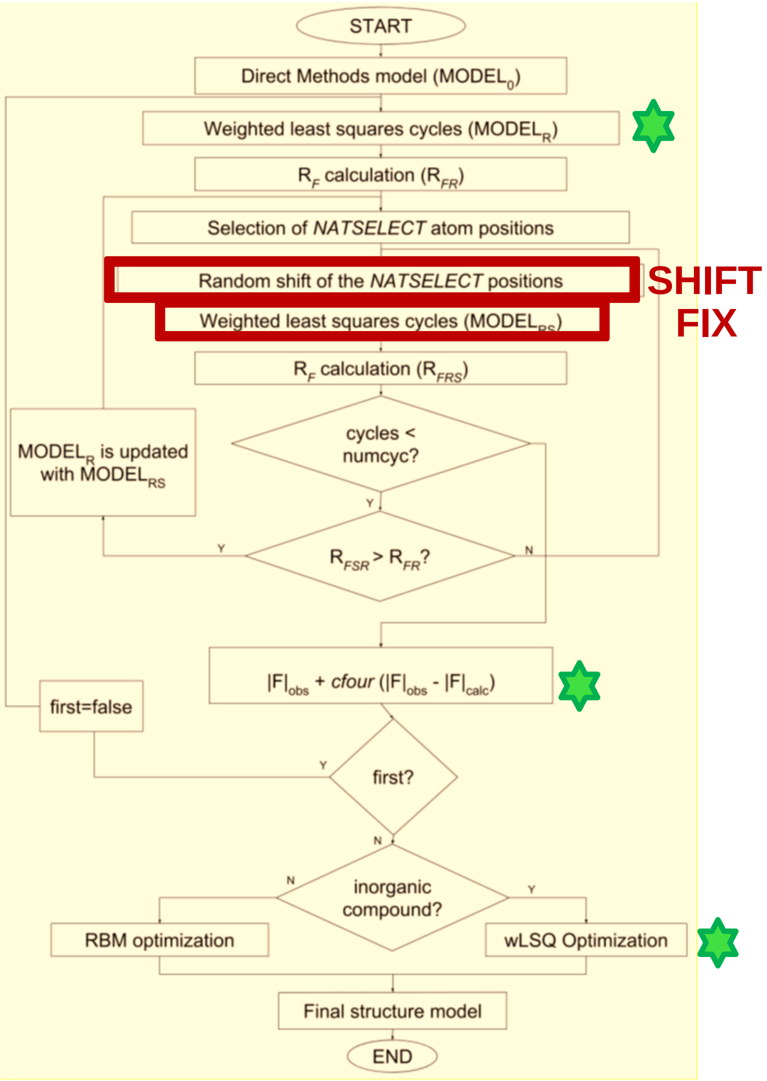
Good morning, Angela. Welcome to jav

THE SHIFT_AND_FIX METHOD

- Least squares are carried out by using experimental structure factor moduli as observations, suitably weighted

- Fourier map: difference map contribution depends on atomic content

- For the final optimization, weighted least squares are used in case of inorganic compounds (wLSQ); for organic structures, the resolution bias modification (RBM) method is applied



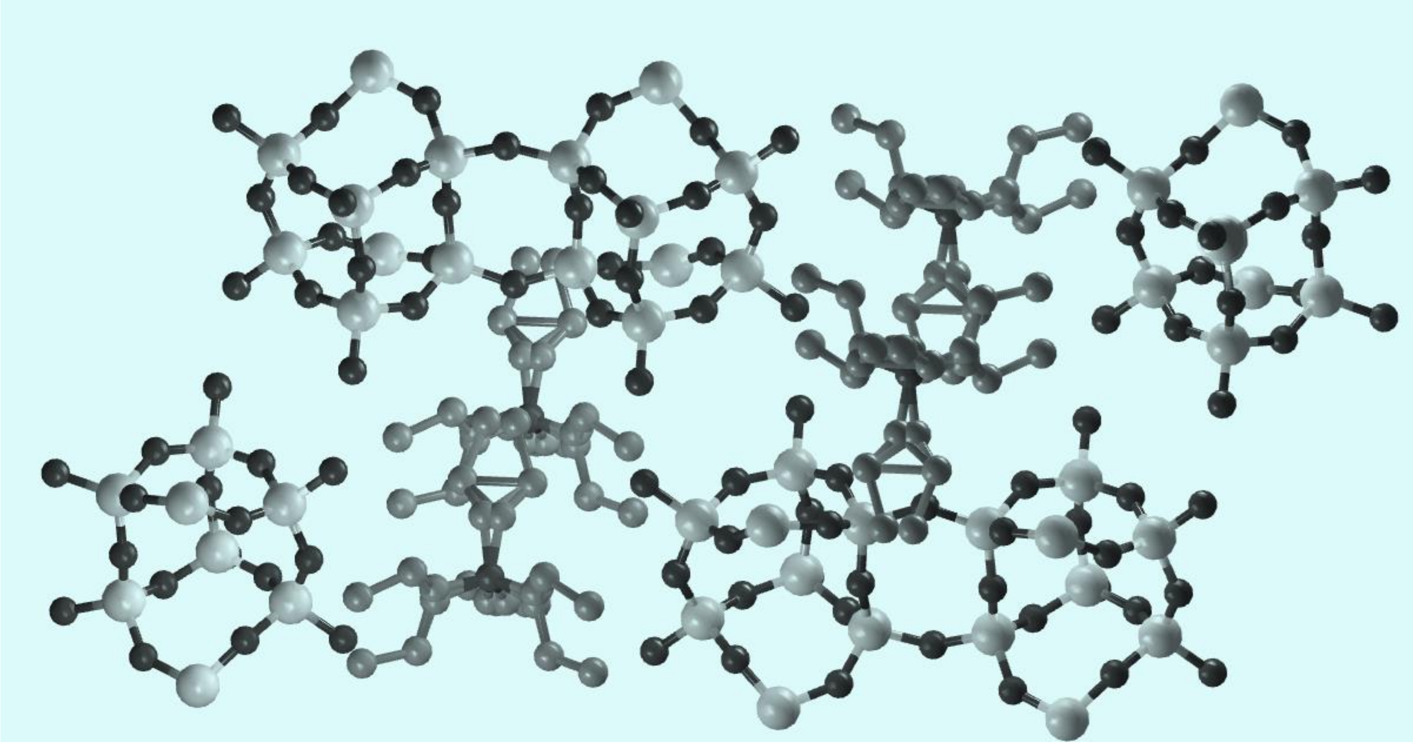
THE SHIFT_AND_FIX METHOD

Code name	Unit cell content	Number of non-H atoms in the asymmetric unit	Resolution (Å)	Space group
ALPHA	O16 N4 C20	10	1.34	$P\ 2_1\ 2_1\ 2_1$
PHENYLA	O8 C32	10	1.43	$P\ 2_1/a$
INOSITOL5A	O24 C24	12	1.20	$P\ 2_1/n$
INOSITOLD1A	O12 C12	12	1.20	$P\ 2_1$
IBUPS	O8 C52	15	1.50	$P\ 2_1/c$
TARTRATE	Ca2 O20 C8	15	1.09	$P\ \text{-}1$
DIPYR	Ni4 Cl8 N8 C40	16	1.14	$P\ 2_1/c$
LEFEBVRE	N8 C56	16	1.10	$P\ 2_1/c$
PIPERAZINE	S4 O12 N4 C48	17	1.43	$P\ 2_1/a$
DICYANO	N8 C64	18	1.09	$P\ 2_1/c$
BENZOS2	S2 O10 N2 C26	20	1.20	$P\ \text{-}1$
DITHIANON2	S4 O4 N4 C28	20	1.54	$P\ \text{-}1$
SAPO	Si32 O64 N2 C48	21	1.20	$P\ m\ m\ n$
CAPSA	O12 N4 C72	22	1.43	$P\ 2_1/c$

Conventional X-ray diffraction data

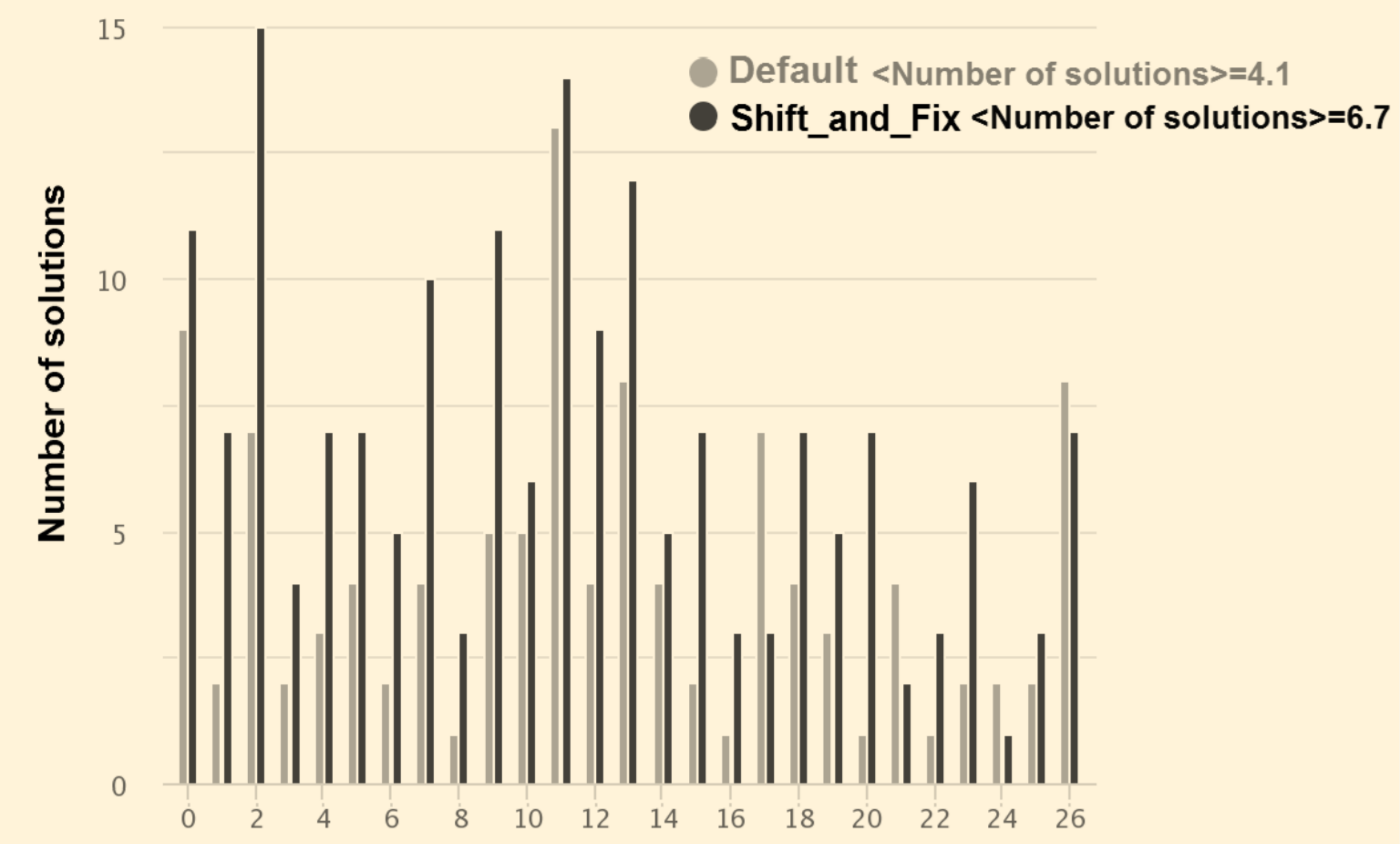
THE SHIFT_AND_FIX METHOD

Silicoaluminophosphate molecular sieve SAPO-40

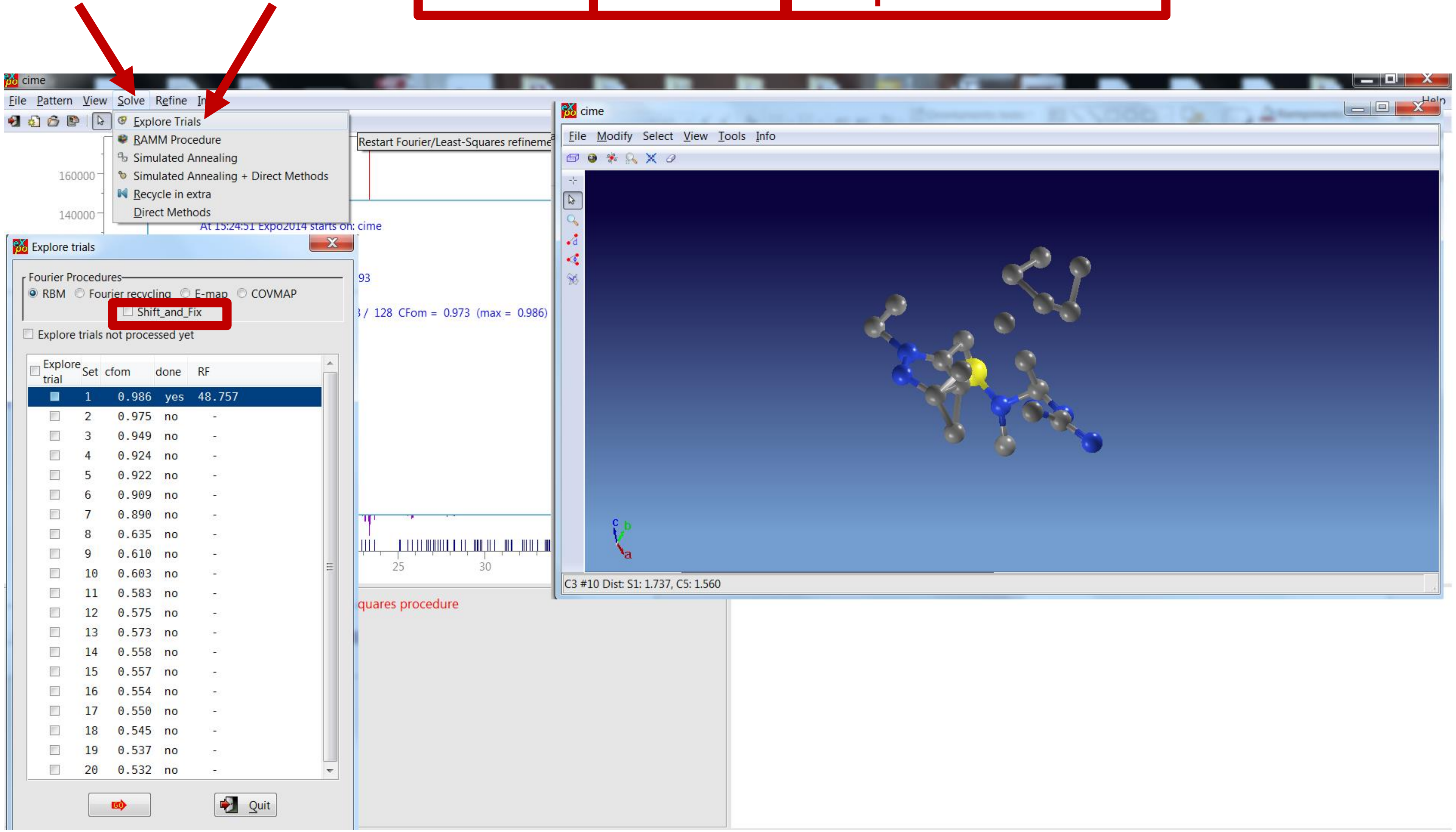


Estermann, M.A., McCusker, L.B., and Baerlocher C. (1992). *J. Appl. Cryst.*, 539-543

THE SHIFT_AND_FIX METHOD



Solve ->->-> Explore Trials



At 15:24:51 Expo2014 starts on: cime

93

3 / 128 CFom = 0.973 (max = 0.986)

quares procedure

Fourier Procedures

● RBM ○ Fourier recycling ○ E-map ○ COVMAP

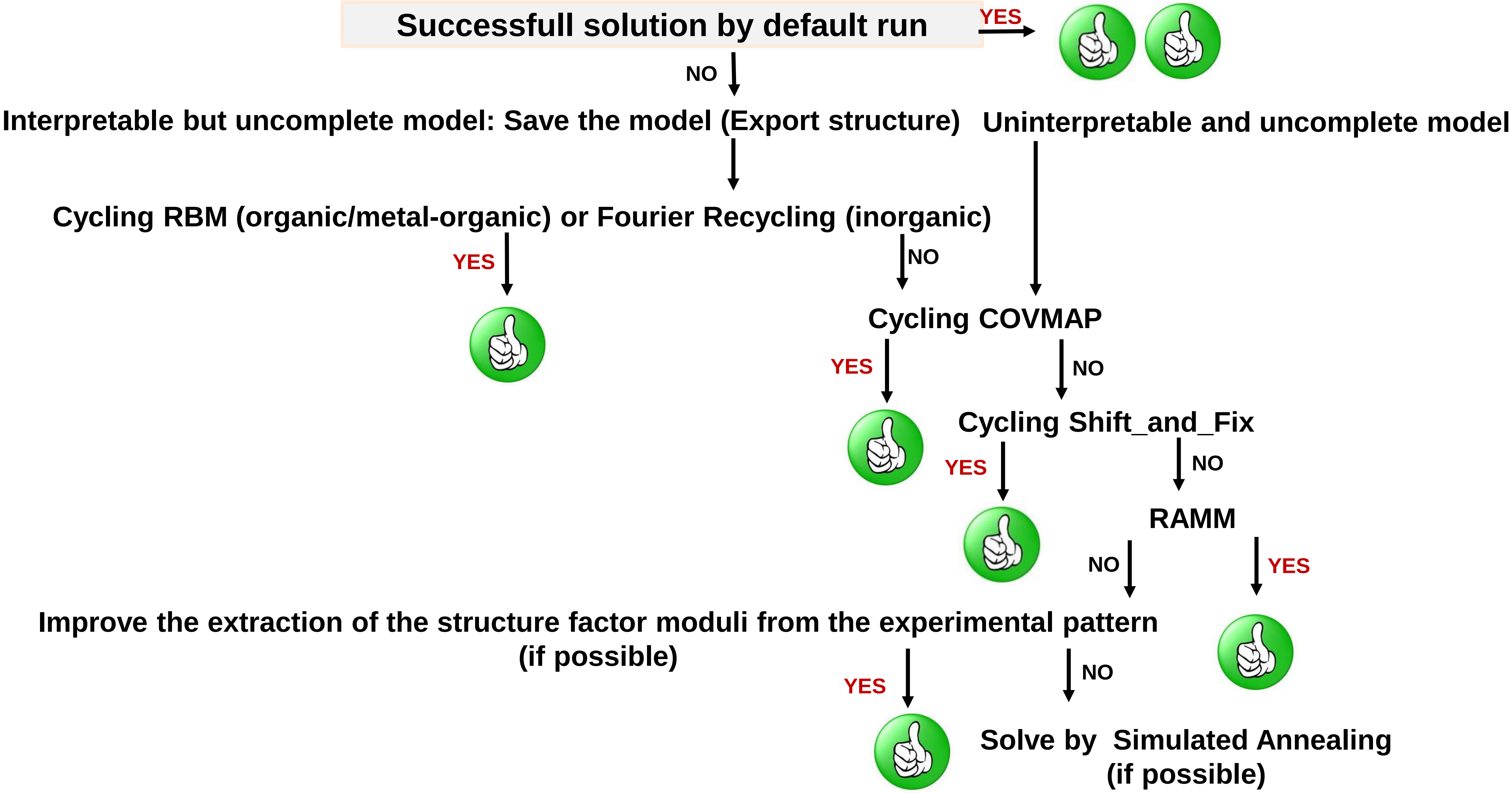
☐ Shift_and_Fix

☐ Explore trials not processed yet

☐ Explore trial	Set	cfom	done	RF
☑	1	0.986	yes	48.757
☐	2	0.975	no	-
☐	3	0.949	no	-
☐	4	0.924	no	-
☐	5	0.922	no	-
☐	6	0.909	no	-
☐	7	0.890	no	-
☐	8	0.635	no	-
☐	9	0.610	no	-
☐	10	0.603	no	-
☐	11	0.583	no	-
☐	12	0.575	no	-
☐	13	0.573	no	-
☐	14	0.558	no	-
☐	15	0.557	no	-
☐	16	0.554	no	-
☐	17	0.550	no	-
☐	18	0.545	no	-
☐	19	0.537	no	-
☐	20	0.532	no	-

C3 #10 Dist: S1: 1.737, C5: 1.560

Suggestions for structure solution in the reciprocal space



Successfull solution by default run

YES

NO

Interpretable but uncomplete model: Save the model (Export structure)

Uninterpretable and uncomplete model

Cycling RBM (organic/metal-organic) or Fourier Recycling (inorganic)

YES

NO

Cycling COVMAP

YES

NO

Cycling Shift_and_Fix

YES

NO

RAMM

NO

YES

Improve the extraction of the structure factor moduli from the experimental pattern (if possible)

YES

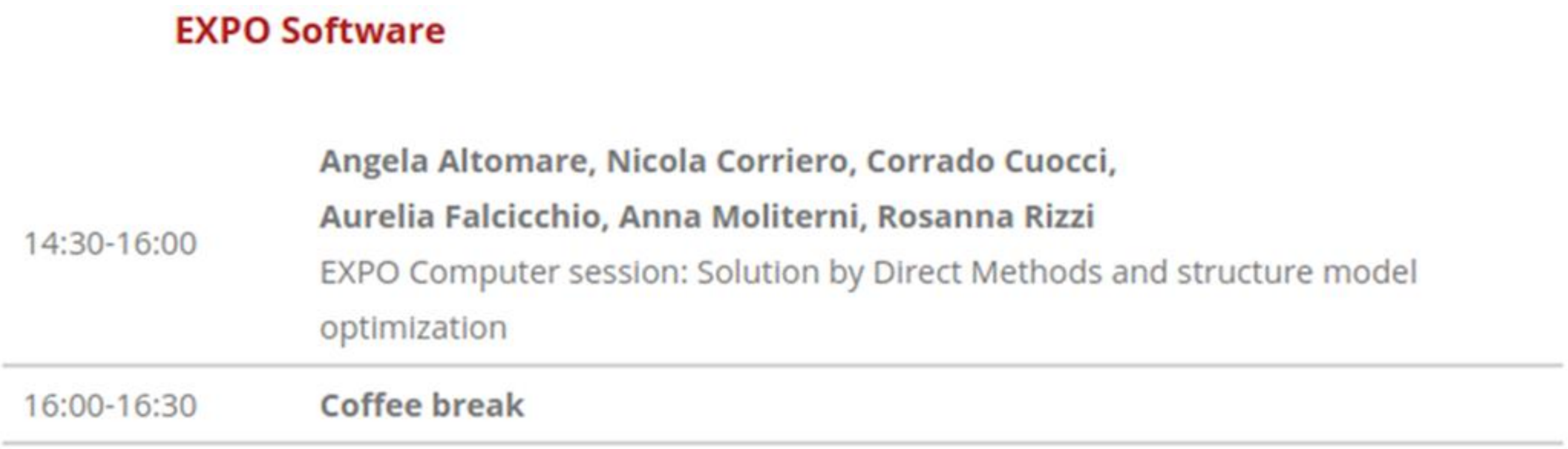
NO

Solve by Simulated Annealing (if possible)

STRUCTURE MODEL OPTIMIZATION

Scientific Program

Tuesday, 1st October 2019



EXPO Software

14:30-16:00	**Angela Altomare, Nicola Corriero, Corrado Cuocci, Aurelia Falcicchio, Anna Moliterni, Rosanna Rizzi** EXPO Computer session: Solution by Direct Methods and structure model optimization
16:00-16:30	**Coffee break**

crystals

IMPACT
FACTOR
2.061

Special Issue

Crystal Structure Characterization by Powder Diffraction

Guest Editor:

Dr. Angela Altomare

Dr. Rosanna Rizzi

Institute of Crystallography, National
Research Council-CNR, Bari, Italy

E-mail: powder@ic.cnr.it

Submission Deadline: 31 December 2019

200 CHF discounts for the paper submit by the deadline

We invite contribute of papers that, while discussing the followed computational, methodological, and/or experimental strategies, point out the essential and advanced contribution of powder diffraction in identifying the unknown crystal structure of a compound.

Keywords:
Structure solution methods
Qualitative analysis
Quantitative analysis
Structure refinement
Structure determination

MDPI
Academic Open Access Publishing
since 1996

Crystals Editorial Office
MDPI AG
St. Alban-Anlage 66
4052 Basel, Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18
www.mdpi.com
crystals@mdpi.com

Thank you for your kind attention