Solving Organic Structures from Powder Diffraction: News from the FOX

Favre-Nicolin Vincent

Université de Grenoble & CEA, DRFMC/SP2M/NRS, France

Radovan Cerný

(Université de Genève, Switzerland)

1 Presentation of FOX

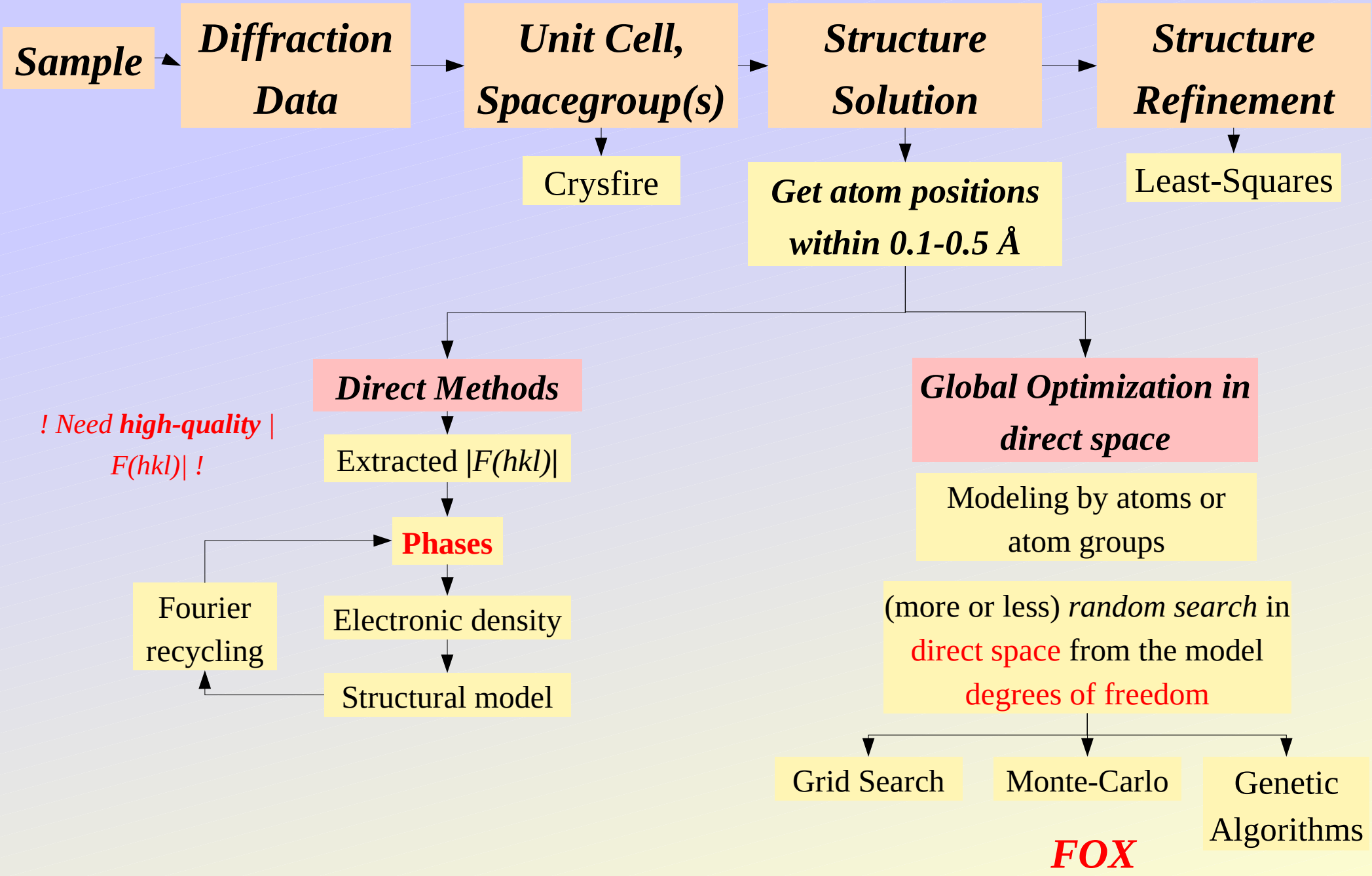
2 Improving Structure solution for organics

3 Examples of structures solved using Fox

4 New features

5 Conclusion

Structure Solution Overview



FOX features

Parametrization

- inorganic or organic materials
- description using atoms, polyhedra, molecules
- automatic, smooth correction of special positions

Data

- powder pattern (X-Ray, neutron, multi-phase)
- (pseudo-) single crystal
- joint optimization with several data sets
- use integrated profiles (no need to extract F(*hkl*))

Algorithms

- Parallel Tempering (Simulated Annealing)
- expandable to new algorithms

Other uses

- display GSAS/EXPGUI Fourier maps
- simulation of powder & single crystal diffraction

Availability

- free (http://objcryst.sourceforge.net and CCP14)
- open source (GPL)
- available for *Linux*, *MacOS X* and *windows*

Monte-Carlo Algorithm

Parametrization
-> Degrees Of Freedom

keep configuration

yes

starting sonfiguration

random change of parameters

evaluation of the new configuration:
Cost (C)

is configuration better ?
$C_n < C_{n-1}$

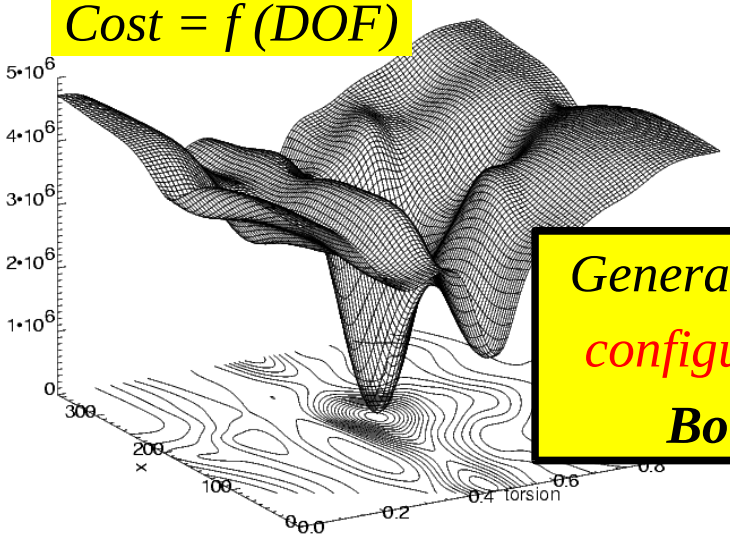
no

keep configuration with probability:

$$P = e^{\frac{-\Delta C}{T}}$$

Temperature of the algorithm

Hypersurface
Cost = f (DOF)

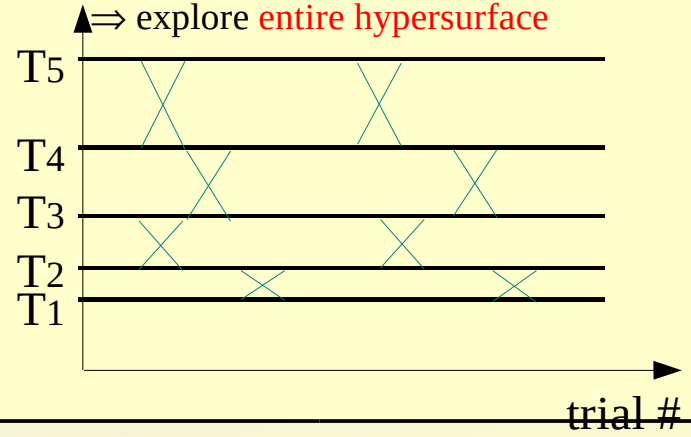


Generate a distribution of configurations following ***Boltzmann****'s law*

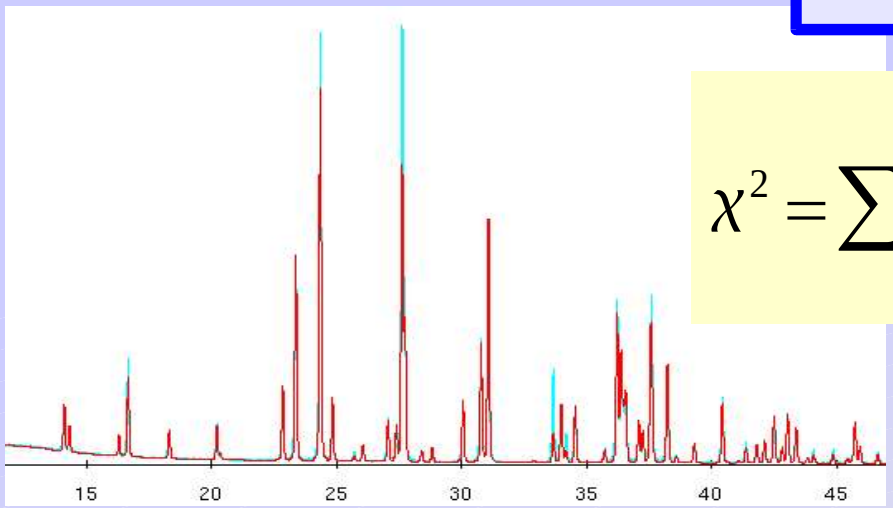
Parallel Tempering

simultaneous optimization at different temperatures

⇒ explore entire hypersurface

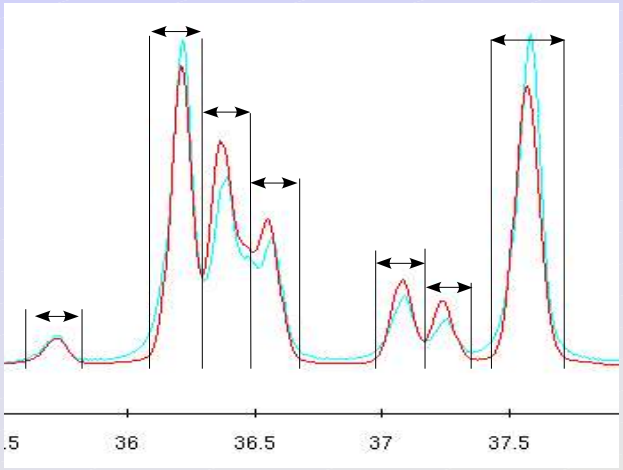


Evaluation of Trial Configurations



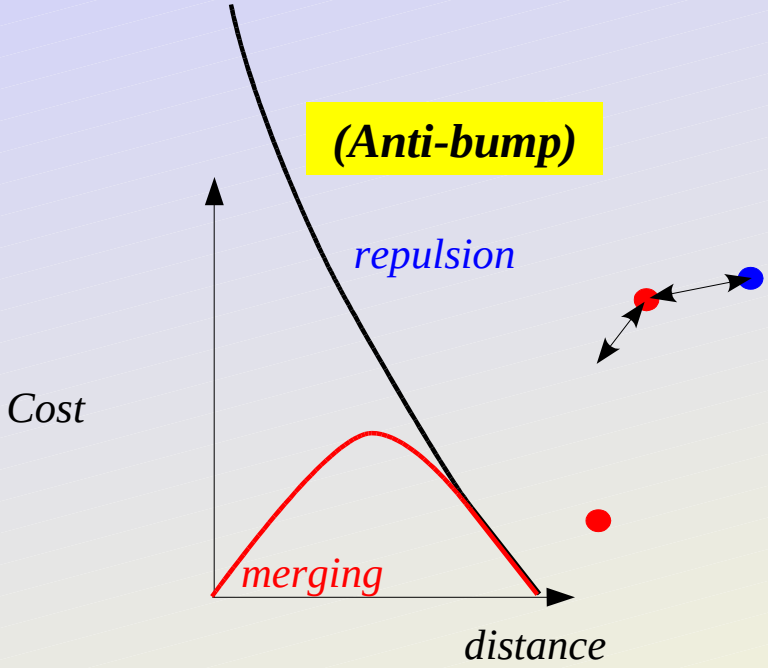
$$\chi^2 = \sum \frac{\left(y_i^{obs} - y_i^{calc}\right)}{\sigma_i^2}$$

integrated profiles



- ***No need to know the profiles***
- ***Fast***

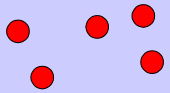
(Anti-bump)



Use any combination of criteria

$$Cost = \sum \chi_j^2$$

Crystal Structure Components

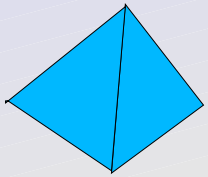


Model = n atoms ? 3N parameters !

Complexity increases exponentially with the number of parameters
⇒ use *a priori* information to reduce the number of trials required

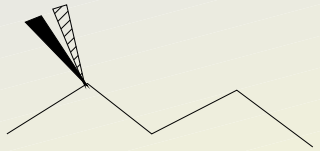
inorganic

- ***High symmetry***
- ***Special positions***
- ***Local building blocks (polyhedra)***



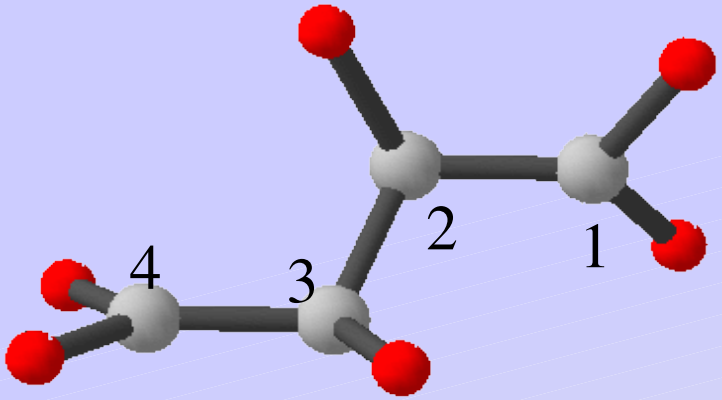
organic

- ***Low symmetry***
- ***Plane formula***
- ***Conformation information***
- ***Other: NMR***



Description from bond lengths, bond angles and dihedral angles

Parametrization of Molecular compounds



The number of trials required varies exponentially with the number of parameters

⇒ *need to use all the* **a priori** ***information*** *about the atoms coordination:*
description of all Molecules using ***bond lengths***, ***bond angles*** *and* ***dihedral angles***

```
Z-Matrix
10
C   1
C   1 1.5
C   2 1.5  1 110
C   3 1.5  2 110   1 0
O   1 1.2  2 120   3 0
O   1 1.2  2 120   5 180
O   2 1.4  1 110   3 180
O   3 1.4  2 110   4 180
O   4 1.2  3 120   2 0
O   4 1.2  3 120   9 180
```

number of atoms

free torsion angles

atom type

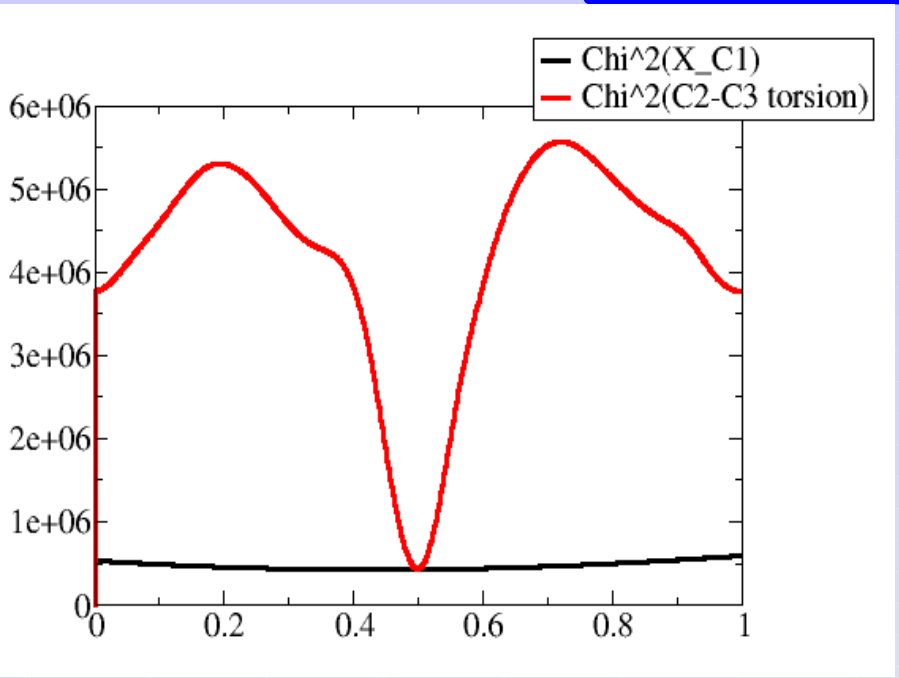
bond ***length*** with atom #

bond ***angle*** with atom #

dihedral angle with atom #

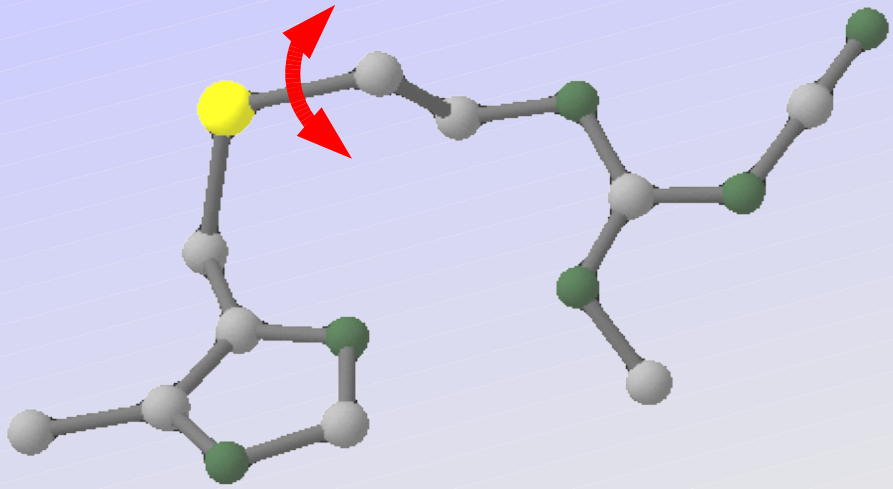
All the atomic positions are deduced from the first atom position, the z-matrix, and 3 angles for the orientation of the molecule

Pitfalls of internal coordinates (z-matrix)



*a **torsion angle** (moving many atoms) has a **much narrower minimum** than a translation parameter of an individual atom*

⇒ even if the number of degrees of freedom diminishes, the global minimum is much narrower



Atoms are deduced from previous atoms

⇒ the first atoms in the z-matrix must also be the first to be found

⇒ The convergence can depend on the order of the atoms in the z-matrix

The z-matrix approach reduces the parameter space to explore, but makes it (much) more difficult to find the solution

Flexible Approach through Restraints

idea**: keep all the coordination information, but with a **flexible approach

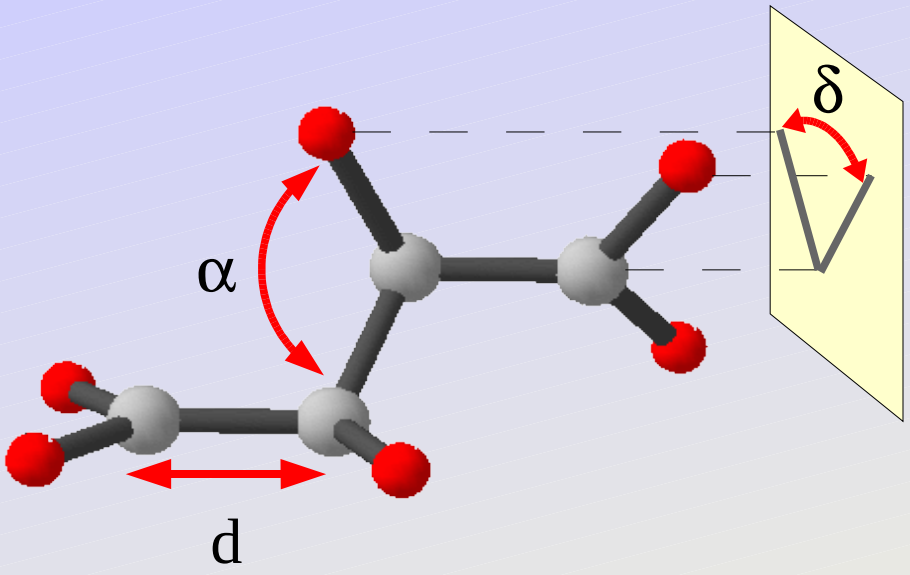
All atom positions are directly defined by their xyz coordinates

and

the coordination information is introduced by **restraints** on:

- ***bond lengths*** $\chi^2 = \frac{(d - d_0)^2}{\sigma_d^2}$
- ***bond angles*** $\chi^2 = \frac{(\alpha - \alpha_0)^2}{\sigma_\alpha^2}$
- ***dihedral angles*** $\chi^2 = \frac{(\delta - \delta_0)^2}{\sigma_\delta^2}$

The orientation of the molecule is defined by a **quaternion** (to avoid "gimbal lock" angles)

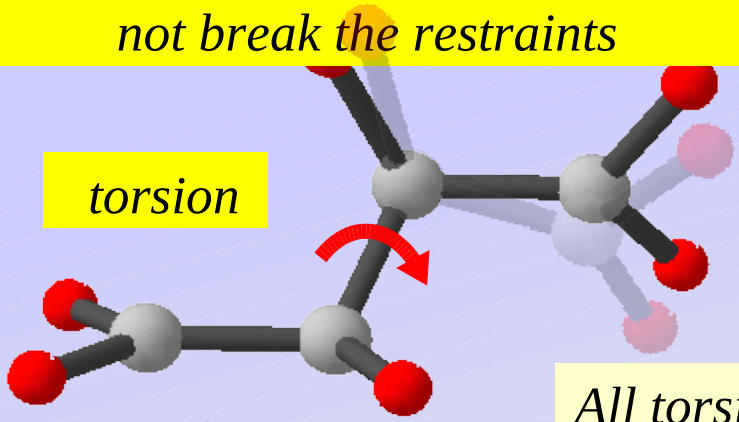


- *this modelization is **independent from the order of the atoms***
- ***any type of restraint** can be introduced*
- ***any type of movement** can be directly done (no need to compute complex torsions)*
- *any **cycle** can be defined*

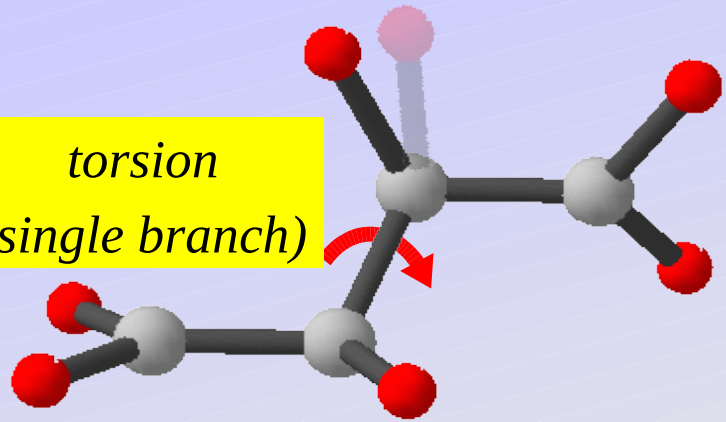
Random Moves

*With atoms defined independently, it is vital to have **intelligent moves** that do not break the restraints*

torsion

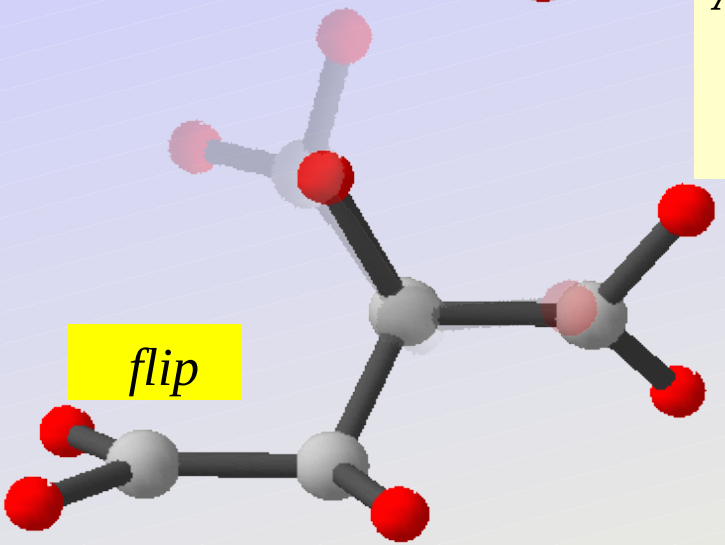


torsion (single branch)



*All torsion & flip moves that do not break restraints are **automatically identified***

flip

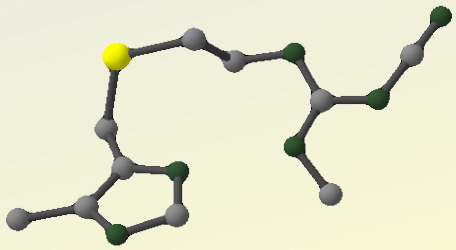


Results:

***Cimetidine**: 4.8M -> 1.6M trials*

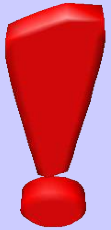
***K-tartrate** : 800 000 -> 385 000*

> 50 % faster !



After each random move, a Monte-Carlo test is made on the total internal restraint cost to see if the configuration is kept

Maximum Likelihood



WARNING :

Approximations !

In a "classical approach" :

$$\sigma^2 = y_{obs}$$

$$\chi^2 = \sum \frac{(y^i_{obs} - y^i_{calc})^2}{\sigma^2_i}$$

*assumes that the model can fit **perfectly** the observed data.*

*But there can be **errors** in the model !*

*typically **positionnal errors** during the search for a structure solution*

with a positionnal error measured by: $D(\vec{k}) = \langle \cos(2\pi \vec{k}.\Delta\vec{r}) \rangle$

*introduce a **variance** on the calculated structure factor*

$$\sigma^2_{calc} = (1 - D^2) \sum_{atoms} f^2_j$$

*Use the "**most likely**" calculated structure factor*

$$\langle F_{calc} \rangle = DF_{calc}$$

$$\sigma^2_i = \sigma^2_{calc} + \sigma^2_{obs}$$

$$\chi^2 = \sum \frac{(y^i_{obs} - \langle y^i_{calc} \rangle)^2}{\sigma^2_i}$$

Application to Global Optimization

1st application:

incomplete model

missing atoms *(H's, solvant) do not contribute to the* ***Structure Factor*** *but increase the* ***variance***

$$D(\vec{k}) = \langle \cos(2\pi \vec{k}.\Delta\vec{r}) \rangle = 0$$

$$\langle F_{calc} \rangle = DF_{calc} = 0$$

$$\sigma^2_{calc} = (1 - D^2) \sum_{atoms} f_j^2$$

Markvardsen, David & Shankland, Acta Cryst A58(2002)

2nd application:

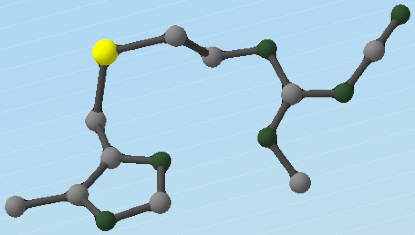
model errors

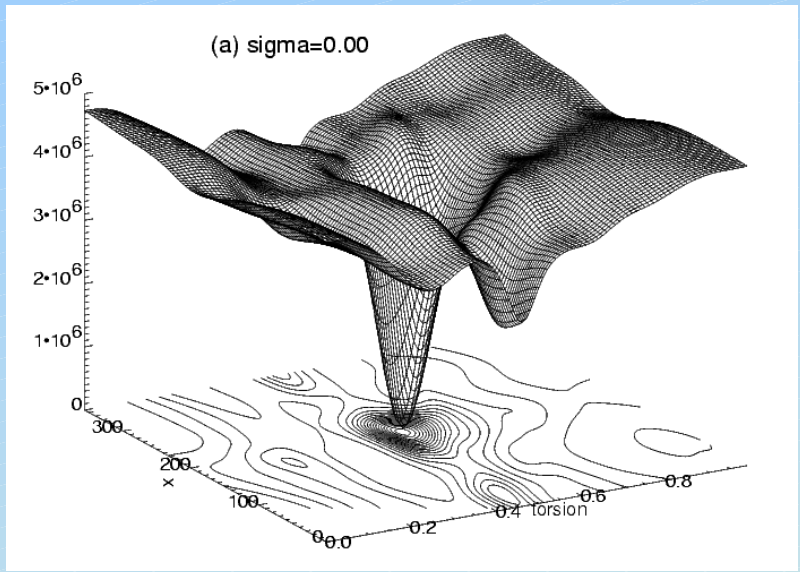
Atoms are always misplaced during a global optimisation

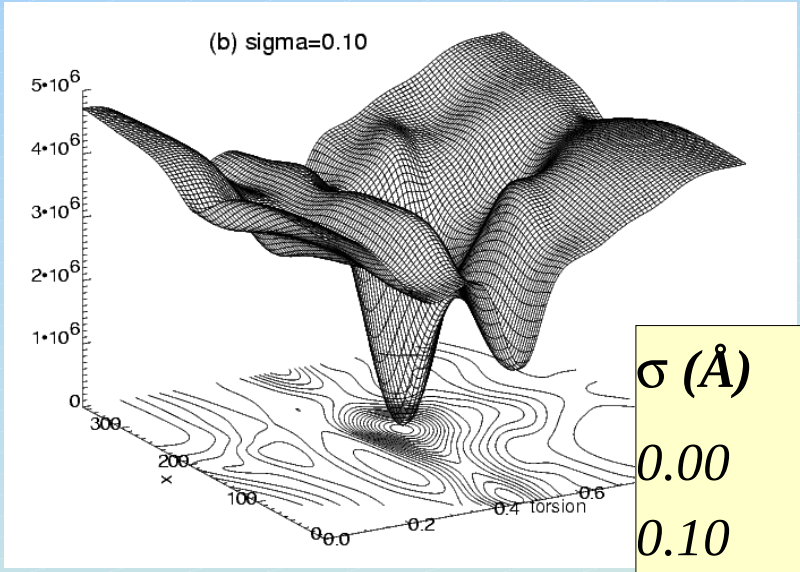
$\Rightarrow$ *taking into account* ***random positionnal errors*** *can yield a* ***better agreement between the incorrect model and the observed diffraction data***

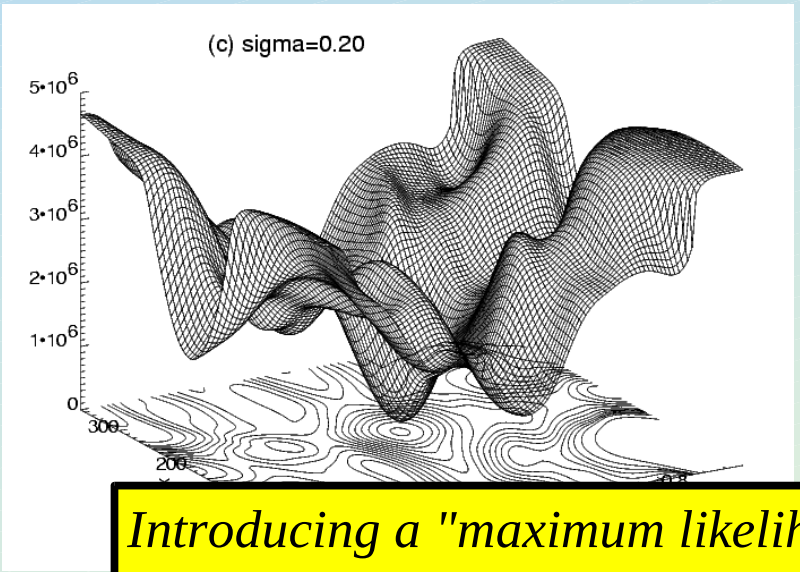
Can it help its convergence ?

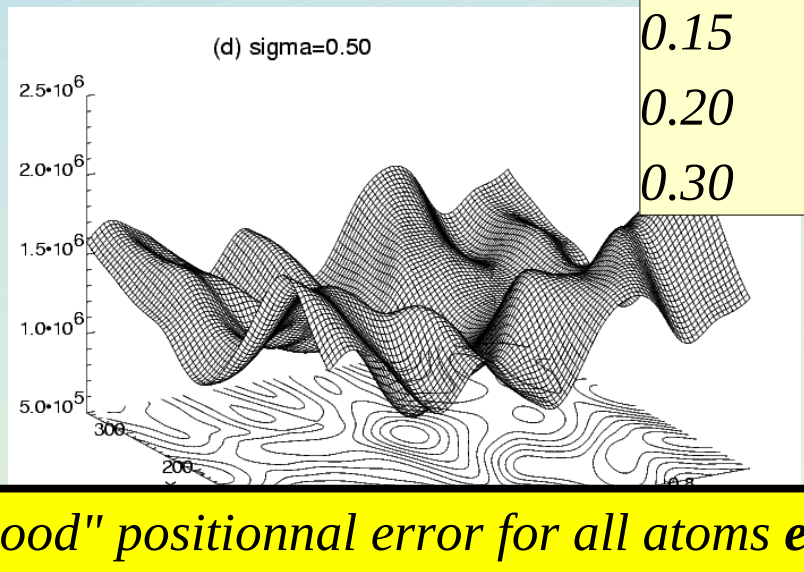
ML for Cimetidine











σ (Å)	<trials>	χ^2_{min}
0.00	1.60M	70,000
0.10	1.01M	58,000
0.15	0.90M	94,000
0.20	0.85M	156,000
0.30	0.85M	300,000

Introducing a "maximum likelihood" positionnal error for all atoms ***enlarges the hypersurface near the global minimum****.*

⇒ *This increases the* ***radius of convergence*** *of the algorithm, i.e. the probability to fall in the global minimum*

First organic (hybrid) structure: $Al_2(H_3CPO_3)_3$

RMN ^{13}C ^{27}Al ^{31}P : identification of 5 building blocks

- 3 non-equivalent H_3C-PO_3 fragments

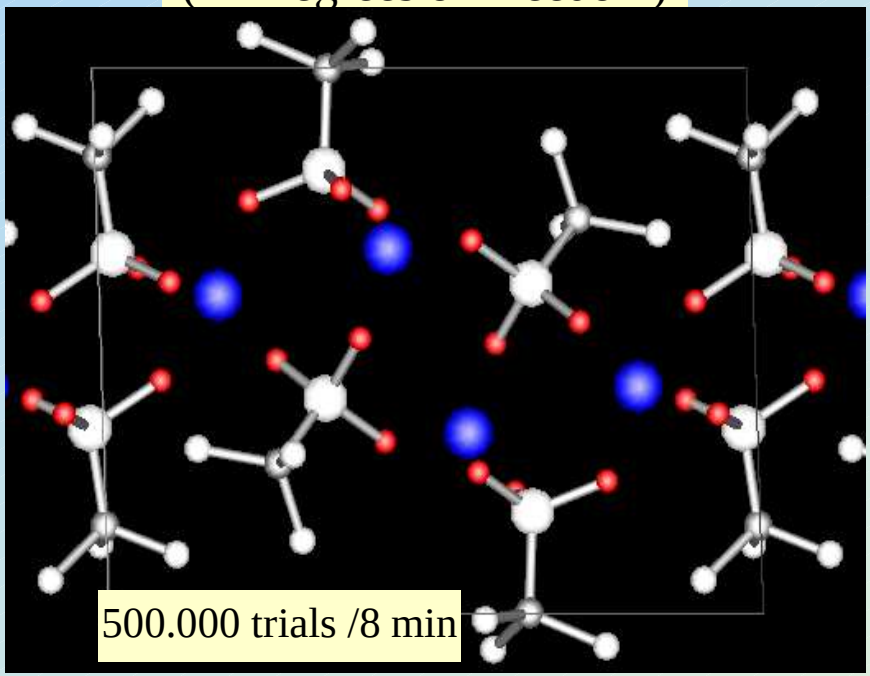
- 2 Al non-equivalent, one in tetrahedral and one in trigonal

2 possible models in Fox

including the hydrogens improves the search, by steric effect

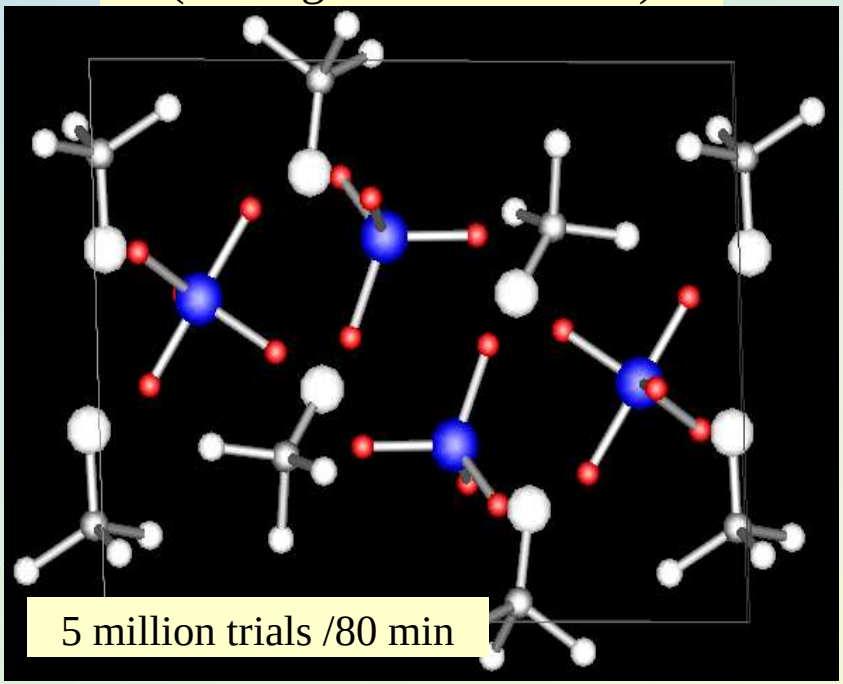
3 H_3C-PO_3 + 2 Al atoms

(24 Degrees of Freedom)



3 H_3C-P + AlO_4 + AlO_5

(27 Degrees of Freedom)



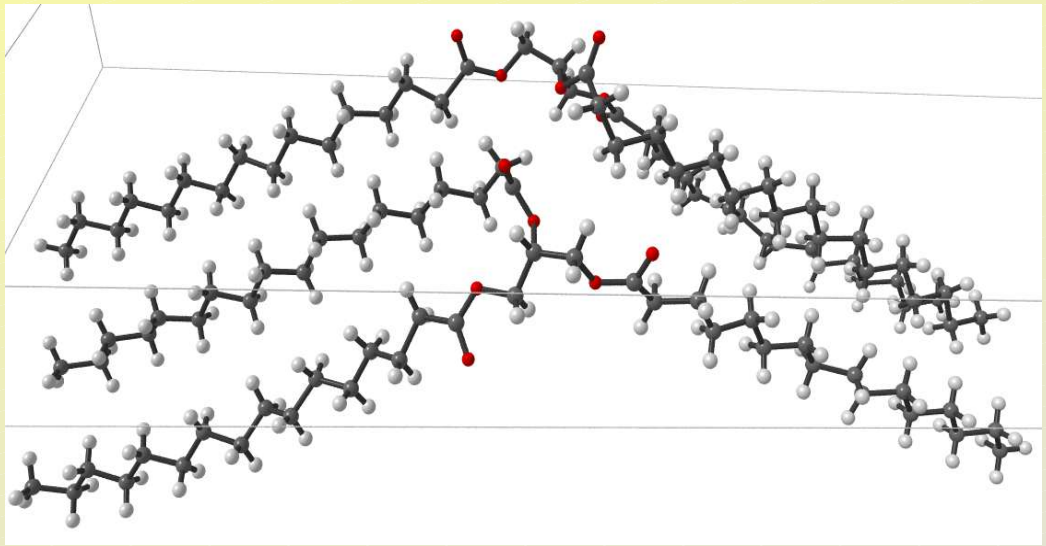
Chem. Commun. ***2002***, *808*

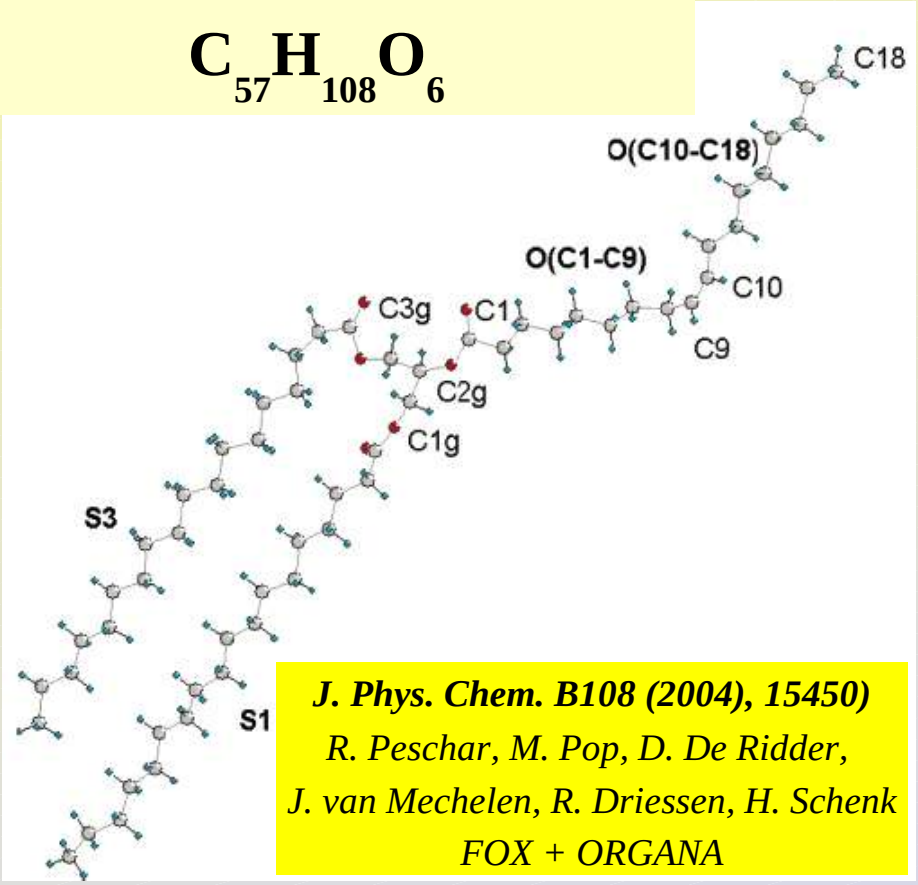
M. Edgar et al. (St Andrews, Scotland)

Triglycerides

1,3 distearoyl-2-oleoglycerol

$C_{57}H_{108}O_6$





J. Phys. Chem. B108 (2004), 15450)

R. Peschar, M. Pop, D. De Ridder, J. van Mechelen, R. Driessen, H. Schenk

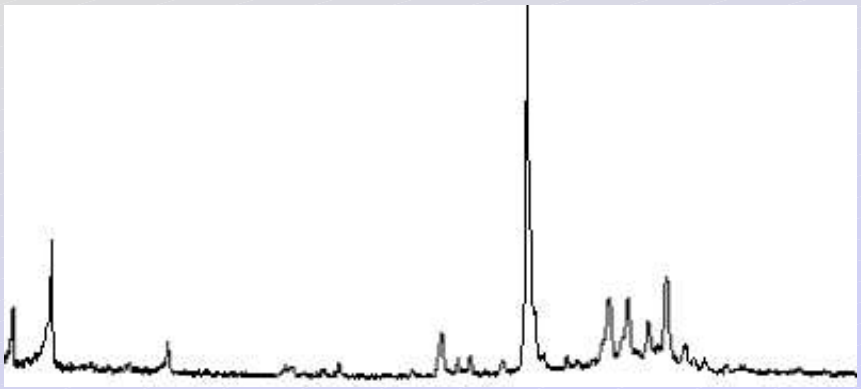
FOX + ORGANA

β' *PSP (1,3-di-n-hexadecanoyl-2-n-octadecanoyl glycerol)*

$C_{53}H_{102}O_6$

up to 56 non-H free torsion angles !

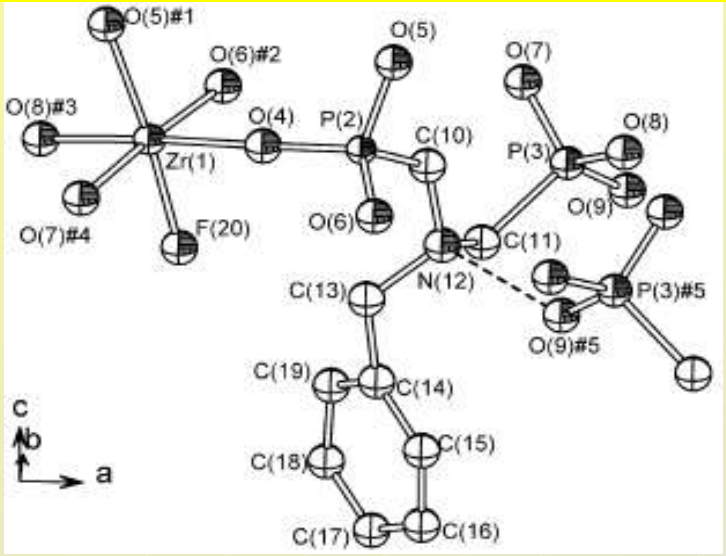
FOX > 2 months



Recent Organic Structures

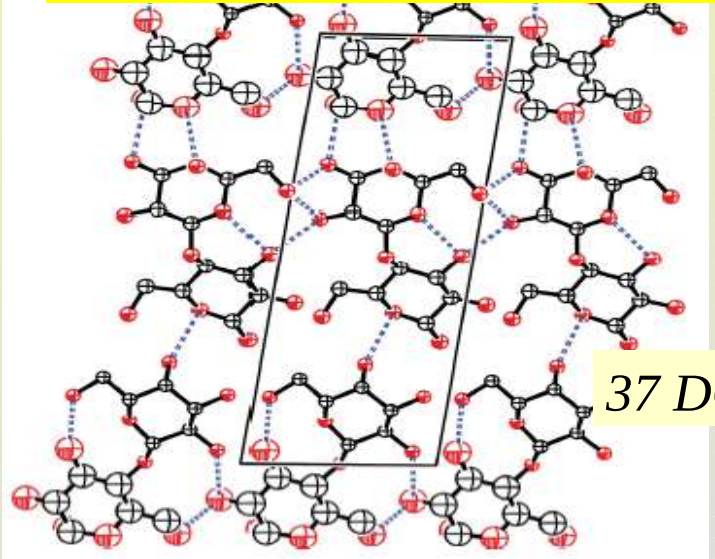
$ZrF(O_3PCH_2)NHCH_2C_6H_5$

J. Sol. St Chem. 177 (2004), 4013

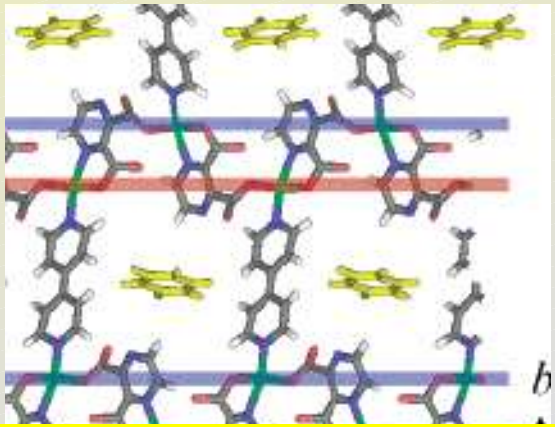


α/β *lactose* $2*C_{12}H_{22}O_{11}$

Acta Cryst. B61 (2005), 455

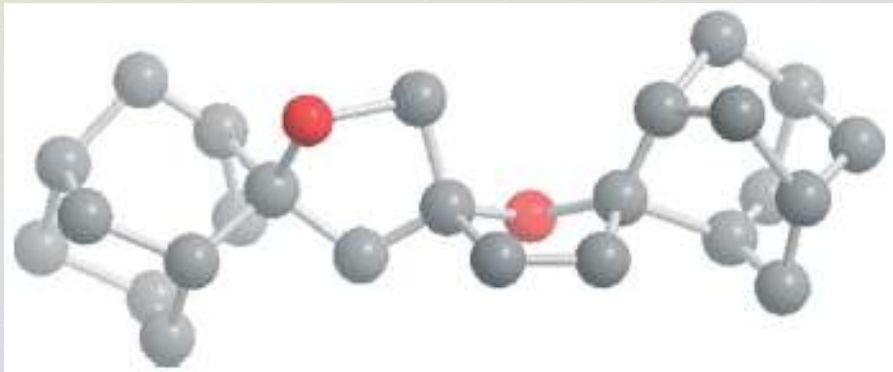


37 DOF



{Cu_2(pyrazine-2,3-dicarboxylate)$_2$ (2',4'-bipyridine)].C_6H_6}$_n$

JACS 126 (2004), 14067



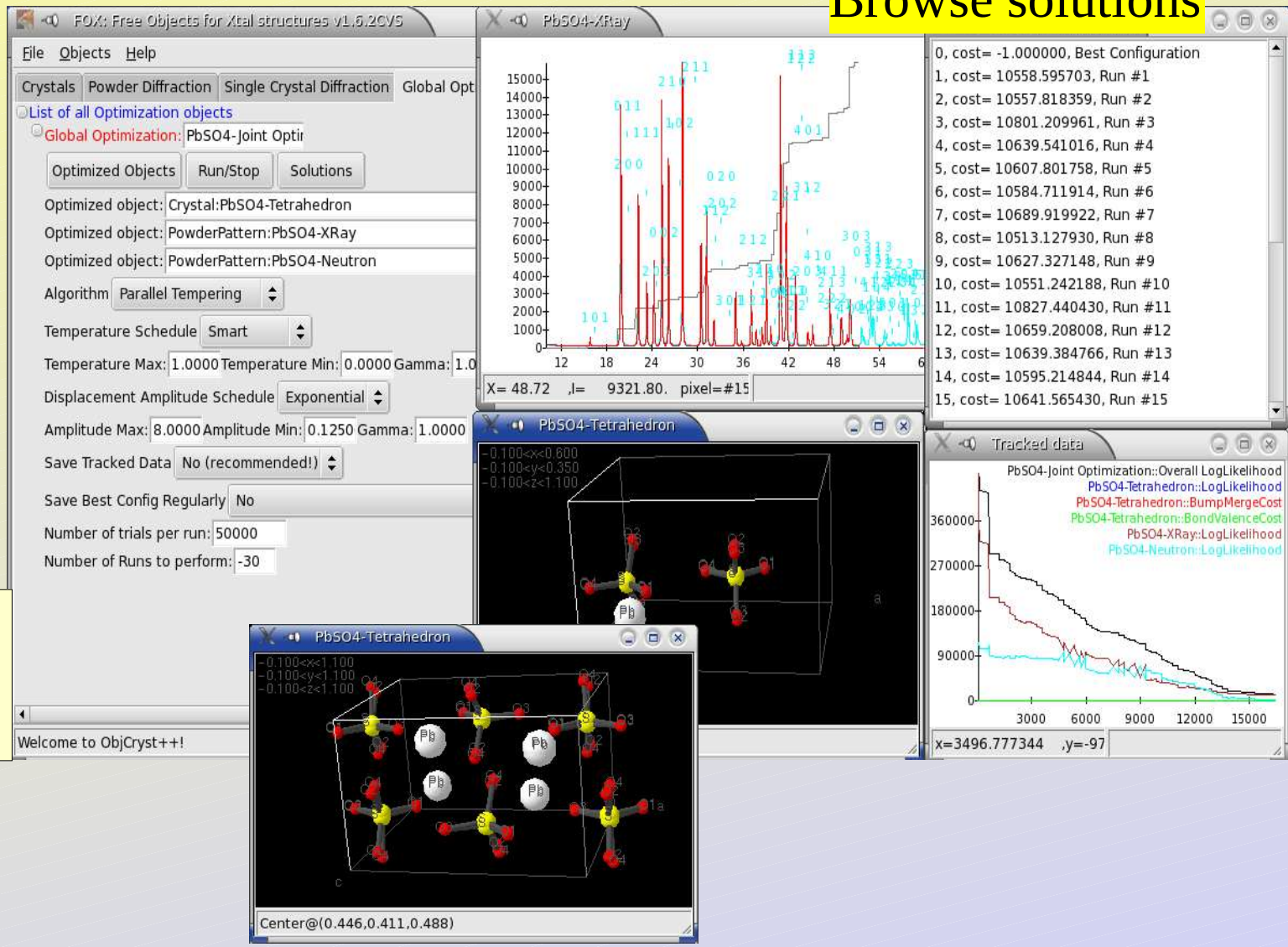
Dispiro[adamantane-2,2'-tetrahydrofuran-5',4"-tetrahydrofuran-2",1'"-adamantane]

Tetrahedron 61 (2005), 3437

New features 1
Multiple Solutions

(1)
Make multiple runs
(*Dash*-like)

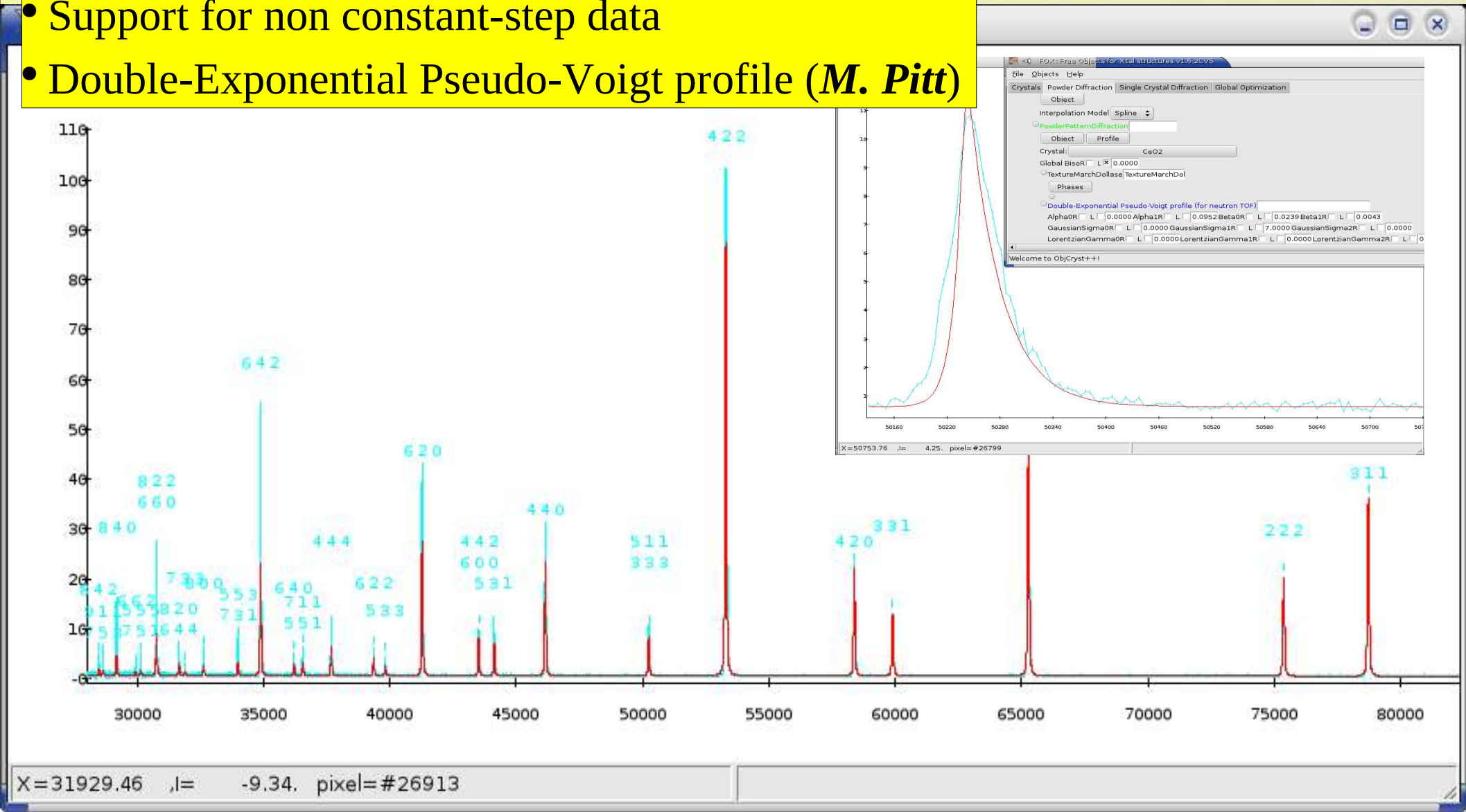
(2)
Browse solutions



Multiple solutions allow to see how "***stable***" the solution is

New features 2
Time-Of-Flight

- Support for neutron TOF data
- Support for non constant-step data
- Double-Exponential Pseudo-Voigt profile (***M. Pitt***)

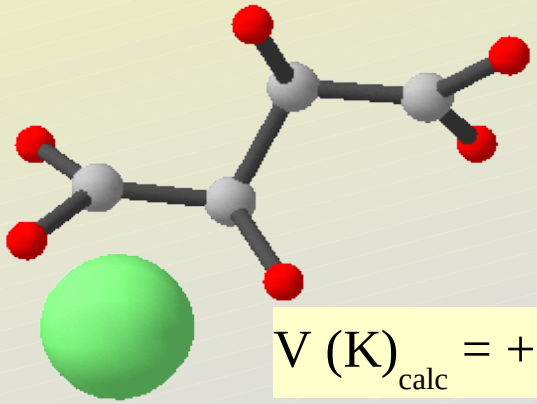


New features 3
Bond Valence

The valence of each atom depends on its neighbours and their distance :

$$V_i = \sum_{neighbours\ j} e^{\frac{(R_1 - r_{ij})}{0.37}}$$

Validation tool :



$V(K)_{calc}$ = + 0.991 OK!

Use as a ***cost function*** ?

Study with R. Cerny and T. Bataille

http://marie.epfl.ch/orlov/bondval/

Other News

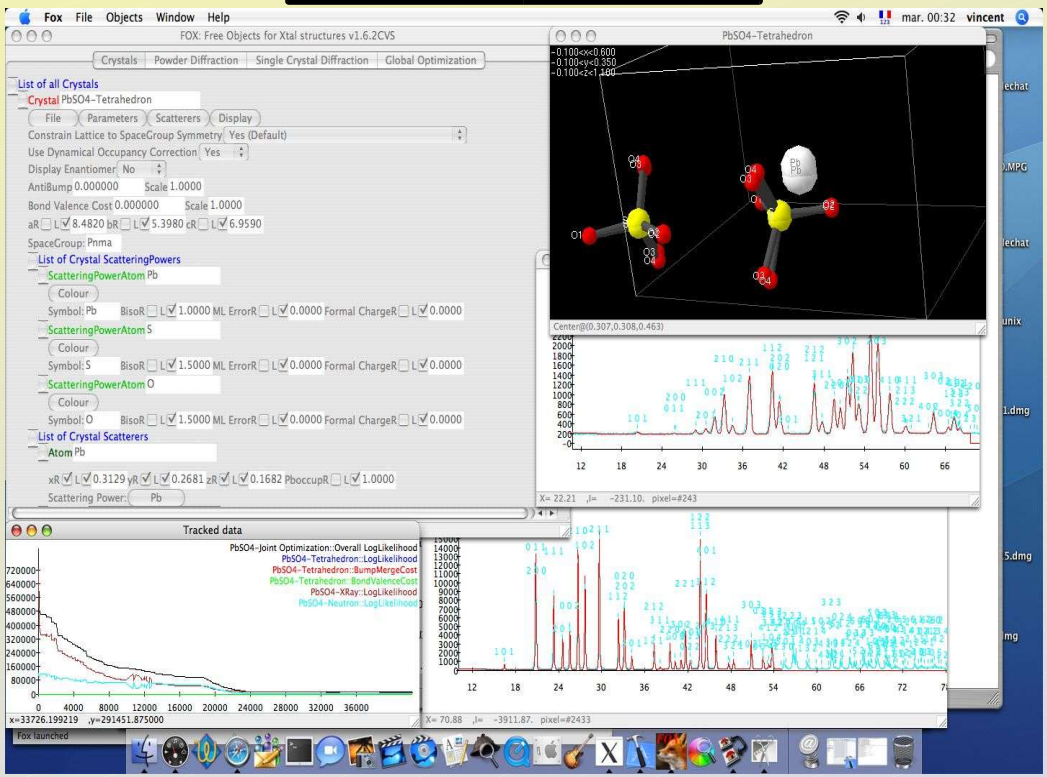
Fox wiki

http://objcryst.sourceforge.net/cgi-bin/moin.cgi/FoxWiki

The wiki will replace the current manual

- Easier to update
- *Anyone* can add information, tips...

Fox on MacOS X



Two versions:

- Aqua look (nicer, slower, few glitches)
- GTK look (faster)

Thanks to Ch. Baerlocher

Future Developments

Next version: sept/oct 2005 (read Fox mailing list for beta versions)

- ***Molecule modelling***:
  - Allow restraints to be more strict
  - Make restraints behave more like the old z-matrix
  - Better support for large Molecules
  - Allow rigid sub-groups of atoms
  - ? Inter-molecular restraints (to use NMR data – can be costly)
- ***Bond Valence support*** (validation & optimization ?)
- ***Maximum Likelihood*** (missing & disordered atoms, helping convergence ?)
- ***Input Molecules formulas directly using OpenBabel (http://openbabel.sf.net)***
  - e.g. using a SMILES: *Phenol* = "OC1=CC=CC=C1"
- ***? Energy criteria***: (difficult since can be computationnaly costly). Need testers with a good energy scheme & coefficients, the data and the will to test !
- ***? Derivative calculations*** opening the way to other algorithms (least squares, hybrid Monte-Carlo, global optimization + local minimization,...)
- ***? Genetic/Differential Evolution*** algorithm (would require altering the Molecule approach)
- ***? Extract structure factors*** [when possible], to use Fourier maps for validation/optimization

Acknowledgements

Anders Markvardsen – Maximum Likelihood help *(ugly approximations are mine),...*

Brian Toby (& Michael Polyakov) – display of GSAS/EXPGUI Fourier maps,...

Lachlan Cranswick – CCP14, suggestions...

Mark Pitt – Time of Flight

Ralf Grosse-Kunstleve - cctbx library http://cctbx.sf.net

Users of Fox

*Fox is not my research subject anymore (see **P.25.07.5** on nanostructures & anomalous diffraction),*
I continue as a hobby ⇒ help me if you want Fox to improve !

***Users advice counts** ! Send suggestions, bug reports, interesting data !*
*⇒ Get on the **Fox Mailing list** to know when beta versions are released*

July 9th 2005
EU Parliament says no to software patents

Software Fayre – FOX demonstration

Sunday 28th (10:00-12:00)

V. Favre-Nicolin & R. Cerny