Hands on McMaille, ESPOIR and GRINSP

Armel Le Bail - *Université du Maine, Laboratoire des oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France. Email : alb@cristal.org*

Abstract - Details on how installing and using the computer programs McMaille (for indexing powder diffraction patterns), ESPOIR (for solving structures in direct space) and GRINSP (for the prediction of framework inorganic compounds) are given.

Introduction

The common point between these three computer programs McMaille, ESPOIR and GRINSP (apart from the author) is that they apply the Monte Carlo method to different problems. McMaille [1] may index the powder diffraction pattern, providing the cell parameters. ESPOIR [2] may solve the structure in direct space using structure factors (either extracted from powder data or coming from single crystal measurements). GRINSP [3] may predict structures in the restricted area of the N-connected 3D nets, with N = 3 (B_2O_3), 4 (zeolites and SiO_2 dense phases: corner sharing tetrahedra), 5, 6 (corner sharing octahedra: MX_3 compounds) and binary combinations (two different polyhedra or two sizes of the same polyhedron type). GRINSP can be also used as a direct space structure solving tool for these frameworks.

'May index', 'may solve', 'may predict', the prudence is because no guarantee is provided since the process is based on the use of random numbers so that the solution may not be captured during the random walk and also because there may be no satisfying solution to the problem.

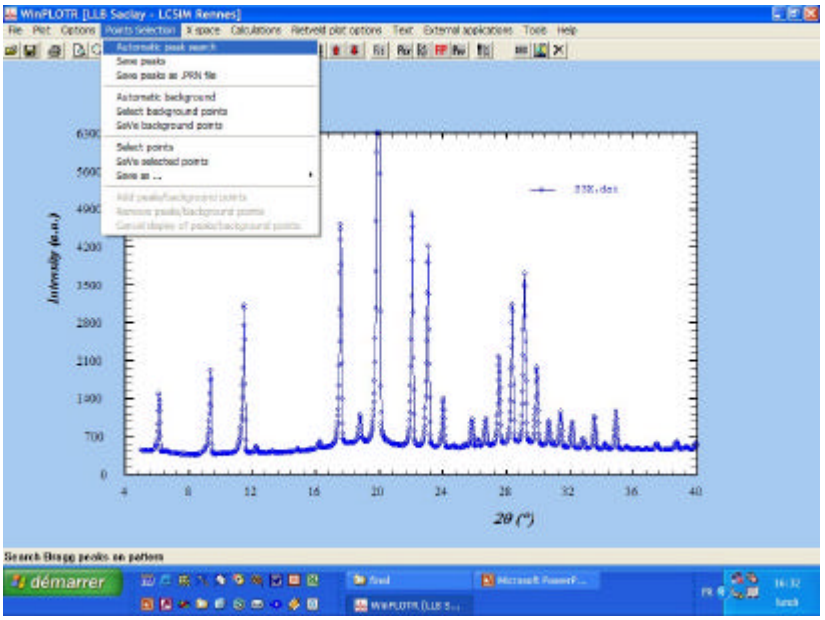
Another common point is the easy availability on the Internet [1-3], the distribution being made according to the GNU Public Licence (the source code is open and should stay open if modifications are made by users/programmers). A bad point is the austere aspect of the PC windows in which are running the programs (generally a DOS-like console application). But this is also a strength since the use of a simple text editor allows to prepare quickly a parameter file which is then combined with some experimental data file, also manageable by text edition.

Demonstrations will be provided with McMaille by using the UPPW data [4] and the indexing benchmarks [5, 6]. The ESPOIR largest problems solved to date correspond to 54 DoFs (degrees of freedom) with the mineral gormanite [7], 27 DoFs with the structure of α-$La_2W_2O_9$ [8]. In these cases, the atoms are moved independently at random in the cell till their positions fit with the intensities of a pseudo-powder pattern regenerated from the extracted "|Fobs|". ESPOIR is capable of translating and rotating a molecule as well, or several objects, up to find the best fit, this will be shown in the case of bethanecol chloride. Demonstrations with GRINSP will exhibit its abilities in the prediction of various structures: B_2O_3, zeolites (SiO_2, aluminosilicates), MX_3, boron silicates, titanium silicates, gallium phosphates, etc.

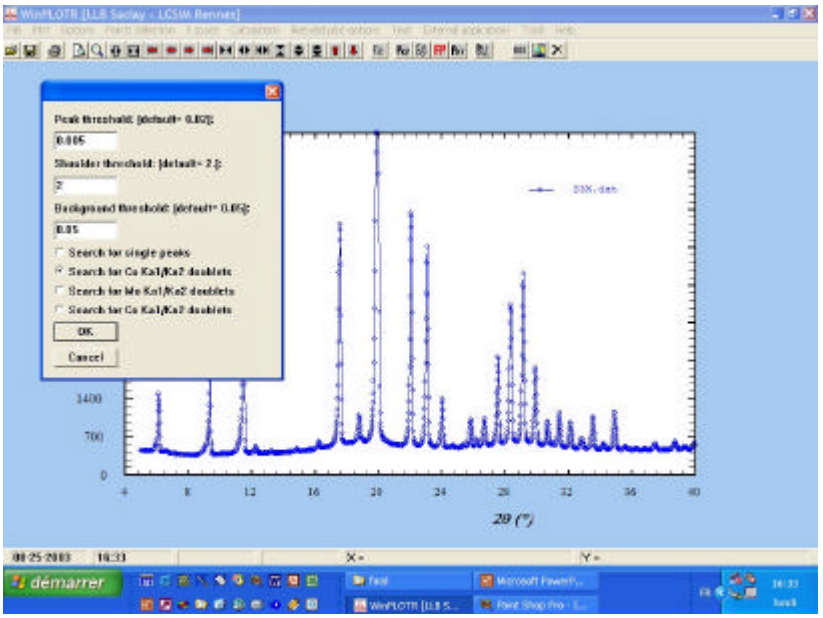
McMaille - indexing powder diffraction patterns

Installation - The package (pronounce MacMy) including the manual, examples and an executable for the MS Windows XP operating system on a Personal Computer (PC) is avaible at the URL http://www.cristal.org/McMaille/. There is no need to sign for any licence, the package is directly available and even the source code is downloadable. The installation is straightforward, after unzipping the whole package in any directory.

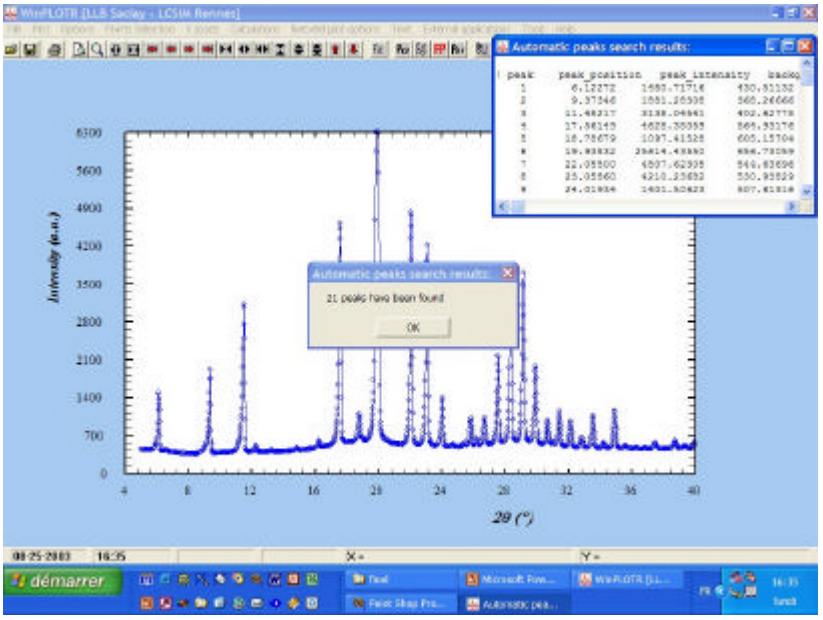
Running McMaille - Crystallographers want solutions fast... Is that possible when using McMaille ? Yes, if you consider that 5 or 15 minutes is fast. The first recommended approach when using McMaille is to use the "automatic mode". The following demonstration uses a friend software (WinPlotr, also graphical user interface for Fullprof, etc) for the extraction of the positions and intensities of the reflections from the raw powder pattern. WinPlotr has an option for the direct creation of an entry file for McMaille in automatic (or black box) mode.



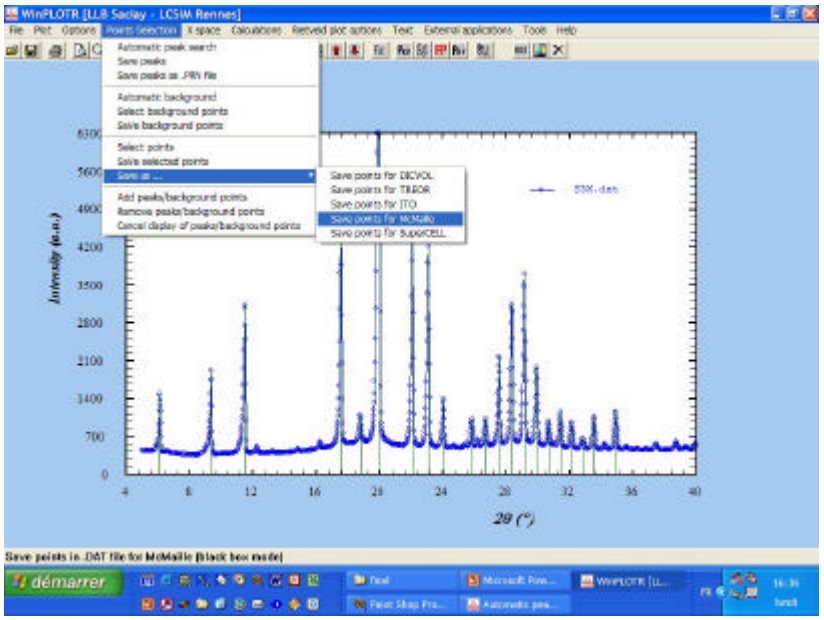
Start WinPlotr, open the powder pattern file, and select « automatic peak search »



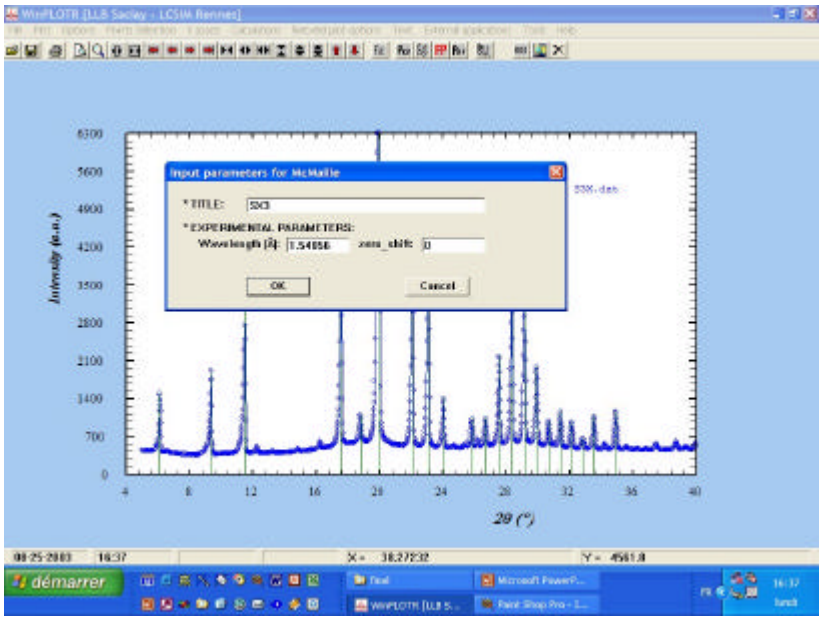
Select the appropriate conditions (Cu-Kalpha1,2 doublet, if any, etc), then, click on OK.



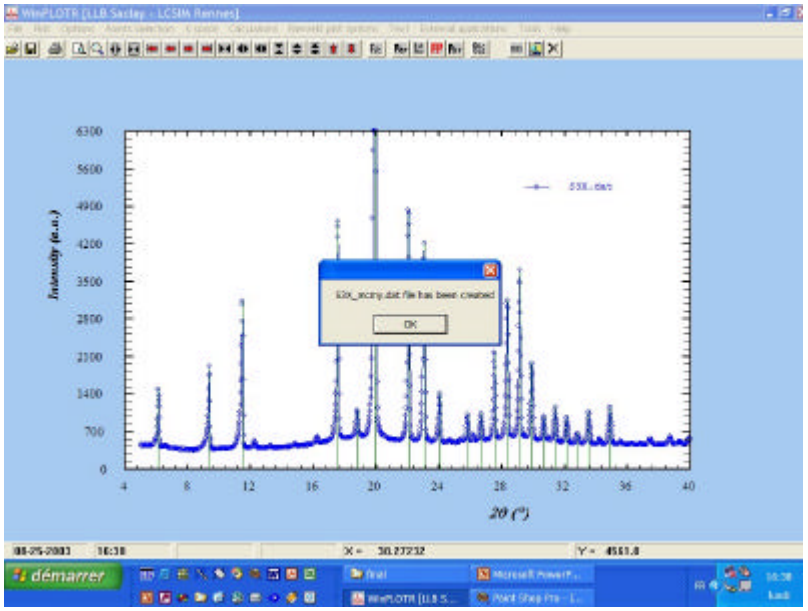
WinPlotr establishes a list of peak positions and intensities.



Save the results for the McMaille program - or if you prefer, for DICVOL, ITO or TREOR, etc.



Give the final details :
a title,
the wavelength,
a zeropoint,
that's all !



The preparation is finished, the entry file for McMaille is created.

That file looks quite simple :

```
S3X_mcmy.dat - WordPad
Fichier  Edition  Affichage  Insertion  Format  ?

SX3
! Wavelength, zeropoint and NGRID (NGRID=3: black box mode)
  1.540560  .000  3
! List of 2theta positions, intensity (min.=20)
       6.12272     1480.71716
       9.37346     1881.28308
      11.48217     3138.04541
      17.56149     4628.35059
      18.78679     1097.41528
      19.93832    25614.43550
      22.05500     4807.62305
      23.05860     4210.23682
      24.01934     1401.50623
      25.82952     1022.77277
      26.68750     1039.36450
      27.52550     2181.97827
      28.34468     3151.54053
      29.13696     3691.50562
      29.90359     1976.70300
      30.66221      977.36542
      31.40555     1166.70593
      32.12103      971.37512
      32.82693      678.58472
      33.53804     1069.94763
```

```
Raccourci vers McMaille.exe
 Entry file (no extension) ??s3x_mcmy
s3x_mcmy


McMaille version 3.04
Data file : s3x_mcmy

    WARNING - WARNING - WARNING - WARNING - WARNING
            This is the black box mode...
        Could need the whole night, if not more...
If you do not want to continue : Ctrl+C would be better
  To cancel but save, type K (capital letter) anytime


Monte Carlo search :
Cubic:          Rp      a         V      Nind
 1        81. 0.043 18.8774   6727.1  1
 1       916. 0.043 18.8774   6727.1  1
Hexagonal:      Rp      a         c        V      Nind
 1      1668. 0.038 15.4186  9.4383   1943.2  1
 1      1907. 0.038 15.4164  9.4385   1942.7  1
 3       574. 0.037 15.4160  9.4385   1942.6  1
Rhombohedral: Rp      a         c        V      Nind
Tetragonal:     Rp      a         c        V      Nind
 1       709. 0.049 13.3483  9.4547   1684.6  1
 1      3198. 0.038 13.3528  9.4382   1682.8  1
 1      4523. 0.037 13.3515  9.4385   1682.5  1
 6       328. 0.037 13.3513  9.4384   1682.5  1
 9      3247. 0.037 13.3512  9.4385   1682.4  1
Orthorhombic: Rp      a         b         c        V      Nind
 1      2012. 0.107  9.4381  3.8540 13.3564    485.8  2
 5      5845. 0.106  9.4380  3.8540 13.3563    485.8  2
 9      4905. 0.106  9.4380  3.8534 13.3571    485.8  2
 1      4558. 0.034  6.6750 13.3568  9.4379    841.5  1
 2      4861. 0.033  7.7046  9.4381 13.3563    971.2  1
```

You need then to start McMaille itself (double click on the program name McMaille.exe in its directory) and to give the entry file name (do not give the file extension .dat which is the default one) just prepared from WinPlotr :

The automatic mode in McMaille is an "expert system" which examines all symmetries in a restrained domain of volumes and cell parameters (no limit in cubic).

Symmetry	max MC events	Pmax	Vmax
cubic	V*0.5	3*dmax	(3*dmax)**3 - no limit
hex/rhomb/tetra	400000	30	4000
orthorhombic	4x1000000	20	500-1000-1500-2000
monoclinic	4x10000000	20	500-1000-1500-2000
triclinic	4x1000000000	20	250-500-750-1000

Thus, it could be necessary to explore larger ranges of volumes and cell parameters if the automatic mode failed, or if the user wishes to have a deepest look. For that, at the end of an automatic examination, McMaille creates a new entry file ready for working in manual mode. Here is such a file :

s3x_mcmy-new.dat - WordPad

Fichier Edition Affichage Insertion Format ?

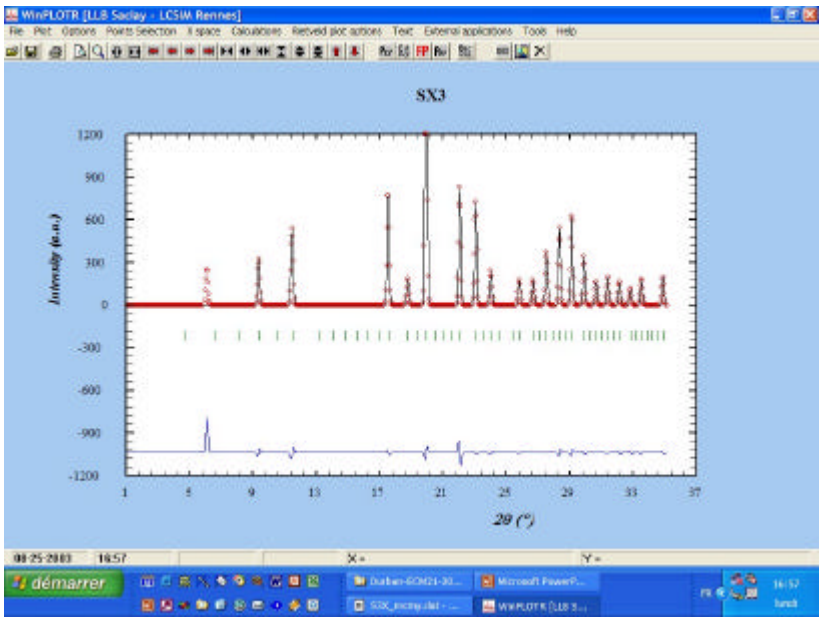
```
SX3
! Wevelength, zeropoint, Ngrid
 1.540560 0.0000 2
! Codes for symmetry
 1 0 0 0 0 0
! W, Nind
0.300  2
!Pmin, Pmax, Vmin, Vmax, Rmin, Rmax, Rmaxref
  2. 50. 8. 125000. 0.05 0.25 0.50
! Spar, Sang
 0.02 0.2
! Ntests, Nruns
 2000 1
!  2-theta    Intensity
   6.122720       1480.717
   9.373460       1881.283
   11.48217       3138.045
   17.56149       4628.351
   18.78679       1097.415
   19.93832       25614.44
   22.05500       4807.623
   23.05860       4210.237
```

Appuyez sur F1 pour obtenir de l'aide

Starting McMaille in manual mode using this file produces :

Raccourci vers McMaille.exe

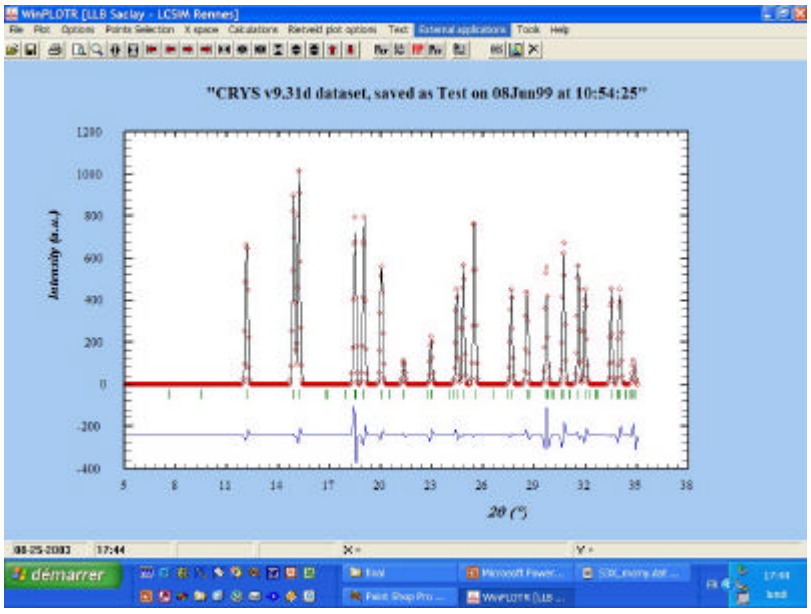
```
 Entry file (no extension) ??s3x_mcmy-new
s3x_mcmy-new


McMaille version 3.04
Data file : s3x_mcmy-new

WARNING - WARNING - WARNING - WARNING - WARNING
EXPECTED total number of tests    4401.000
EXPECTED TOTAL CPU TIME FOR A 2.4GHz:
Seconds:    0.2200500
Minutes:    3.6675001E-03
Hours:      6.1125000E-05
Days:       2.5468751E-06
Years:      6.9777397E-09
If you do not want to continue : Ctrl+C would be better


Monte Carlo search :
Cubic:          Rp      a         V      Nind
 1       455. 0.044 18.8774    6727.1   1

 YOU HAVE FOUND AN INTERESTING RESULT : Rp < Rmin !


FINAL VALUES    : (STANDARD DEVIATIONS : 2nd LINE)
   ZERO     LAMBDA       A        B        C      ALPHA     BETA     GAMMA
 -0.005     1.5406   18.8832  18.8832  18.8832   90.000   90.000   90.000
  0.003     0.0000    0.0033   0.0033   0.0033    0.000    0.000    0.000
RECIPROCAL CELL :    0.05296  0.05296  0.05296   90.000   90.000   90.000
VOLUME (A**3)   :   6733.241

  M(20) =      41.11
  F(20) =      71.74 (  0.0043,  65)
```



A .prf file is built by McMaille which contains data for a drawing by using WinPlotr.

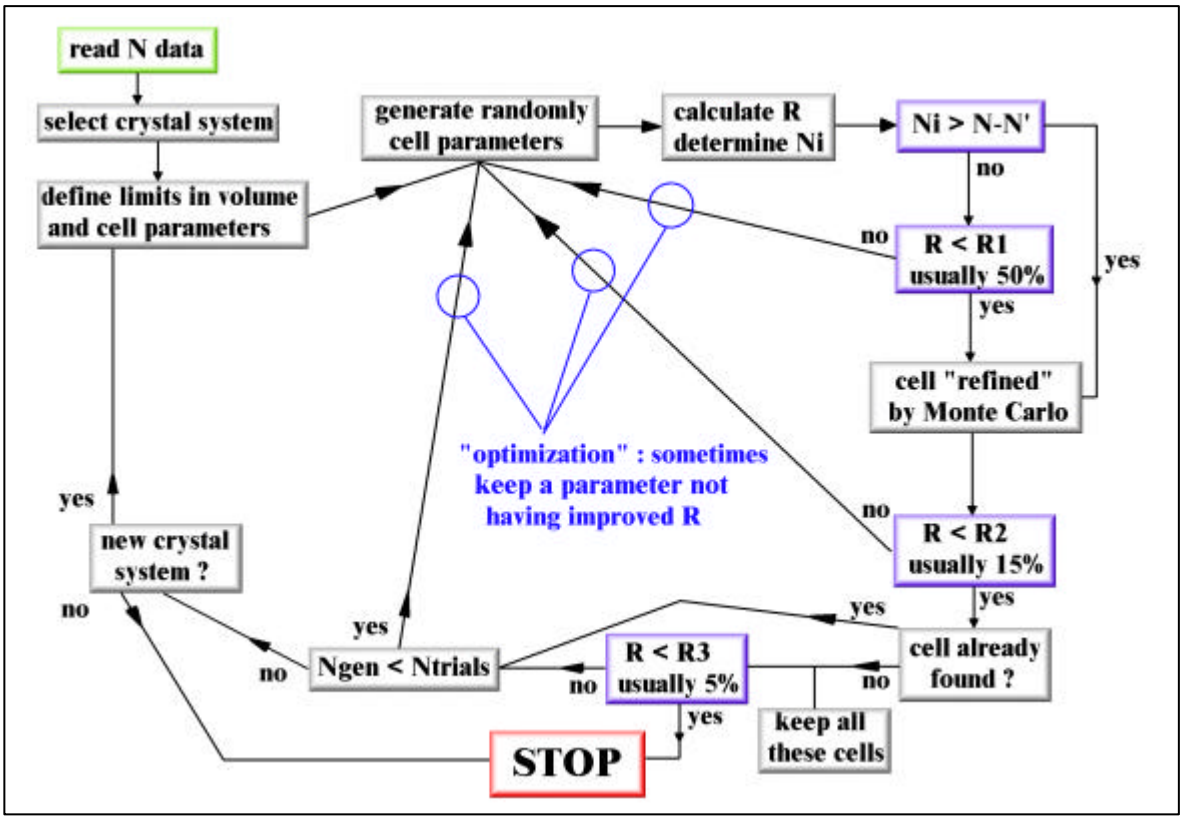
```
Raccourci vers McMaille.exe
 Entry file (no extension) ??y
y

McMaille version 3.04
Data file : y

    WARNING - WARNING - WARNING - WARNING - WARNING
            This is the black box mode...
        Could need the whole night, if not more...
If you do not want to continue : Ctrl+C would be better
  To cancel but save, type K (capital letter) anytime

Monte Carlo search :
Cubic:          Rp      a        V      Nind
Hexagonal:      Rp      a        c       V      Nind
Rhombohedral:   Rp      a        c       V      Nind
Tetragonal:     Rp      a        c       V      Nind
Orthorhombic:   Rp      a        b       c       V      Nind
Monoclinic:     Rp      a        b       c      bet     V      Nind
 2       582. 0.025  9.3801 11.6404  5.9736  96.07    648.6  0

FINAL VALUES    : (STANDARD DEVIATIONS : 2nd LINE)
   ZERO     LAMBDA       A        B        C       ALPHA     BETA     GAMMA
  0.000    1.5406    9.3817   11.6411   5.9733   90.000   96.061   90.000
  0.005    0.0000    0.0021    0.0029   0.0013    0.000    0.010    0.000
RECIPROCAL CELL :    0.10719   0.08590  0.16835   90.000   83.939   90.000
VOLUME (A**3)   :    648.718

  M(20) =      56.94
  F(20) =      94.78 (  0.0042,   50)
```

Here, it is the test example of the CRYSFIRE program (monoclinic cell) which is tested with McMaille in automatic mode. In principe the time of calculation is < 5 minutes with a > 2Ghz.PC. **The next CRYFIRE version should include McMaille**.



The corresponding .prf file.of the CRYSFIRE test:

Algorithm - Below is the McMaille chart flow where Ni is the number of indexed peaks, N is the total number of peaks, N' is the number of tolerated non-indexed peaks, R is the profile reliability, R1, R2 and R3 being the limit values defining some special behaviours of the program.:

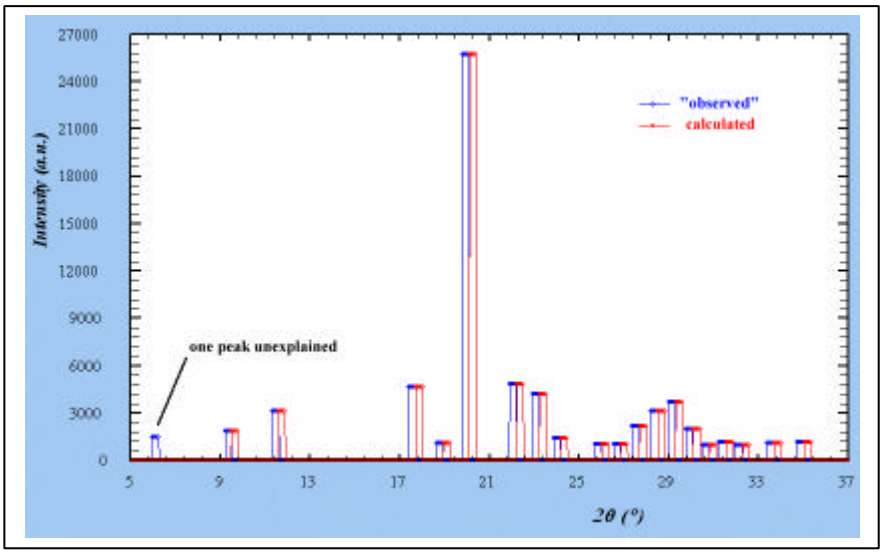


The main difficulty when using McMaille is the long calculation time. Saving time is realized with pre-established lists of Miller indices *hkl*. Between 400 and 1000 *hkl* triplets are prepared in files for each crystalline system and they are read only once at the program start, then kept in memory for subsequent use. McMaille selects the appropriate Miller indices in these lists, but does not reorder them for each test of cell parameter combinations, this would need too much time. If a calculated peak is never overlapping with an observed peak, then the corresponding Miller indices are considered as not observed.

In the preliminary versions 1 and 2 of McMaille (current version being 3), a pseudo powder pattern was regenerated from the positions and intensities of the diffraction peaks, applying an idealized peak shape (Gaussian). That choice allowed an important reduction of time calculation if compared to the use of the raw pattern (which would need to reconstitute the background, complex peak shapes, asymmetry, etc). The fit between the pseudo pattern "observed" and calculated was made by iteration of the Rietveld decomposition formula (= Le Bail method). Version 1 was working only in cubic in order to test the potentialities of the Monte Carlo approach. It was possible to realize 1000 tests per second by using a 2.4 GHz processor (corresponding to 3000 itérations of the Rietveld decomposition formula per second on a pseudo pattern including 20 peaks). Version 2 of McMaille was extended to all crystalline systems, the fastness decreasing to 300 tests per second for a triclinic cell. This was not enough for the low symmetries (monoclinic and triclinic) needing 10^8 to 10^9 tests for having chances to succeed in the indexing.

In the McMaille version 3, the idealized peak shape is even more simple : it is a column. That hyper-simplification has led to a multiplication of the speed by a factor 20, allowing to realize 20000 tests per second in cubic symmetry and 6000 for triclinic cells. The major part of the speed gain is obtained because an adjustment of the peak intensities is no longer necessary, the observed

and calculated columns have the same height and width, fixed at the beginning. The profile reliability R_P is now a function of the overlapping percentage between observed and calculated columns. The column width, in automatic mode, is calculated by the formula : 0.3 x λ / 1.54056. It can depend on the user's choice in manual mode.



One of the two cases leading to "refine" a cell is when Ni > N-N'

An observed peak is considered as indexed if some overlapping exists with a calculated peak. The second case leading to "refine" the cell parameters of a Monte Carlo proposal is when R < R1 (it is recommended to choose R1 close to 50%).

McMaille works better if the defined peak width is larger than observed (0.30°(2θ) for the Cu K-alpha-1 wavelength, 1.54056 Å, this is much more than the minimum full width at half maximum which can be as small as 0.04°(2θ)). This broadening may seem to contradict the expected accuracy, however it allows to increase the chances of success because the more the "observed" column are large and the more there is chances for them to overlap with some calculated columns. Let us take a cubic example, a column at 10°(2θ) (d = 8.838 Å) will extend between 9.85 and 10.15 °(2θ) (from d =8.972 to d = 8.707; λ = 1.54056 Å). If this position corresponds to a *200* reflection, the range of values of the cell parameter ***a*** leading in a sure way to find the solution will be [17.41 to 17.94 Å]. Any test of parameter ***a*** in this large range (> 0.5 Å) will be a winning test. Later, when refining the cell parameter, it will be the exactness of the angular position that will be important. The better will be that exactness and the smaller will be the reliability R factor (an exact overlapping of all the "observed" and calculated columns leading to R = 0), allowing to distinguish an excellent solution from several poor ones.

To say more about the "refinement" of the cell parameters if a cell proposal is retained by McMaille: if R < R1 and R1 is close to 50%, a classical refinement by a least squares process would be difficult. Thus, it is by a Monte Carlo approach that the cell parameters are "refined" for cells proposals retained if R < R1. Small cell parameters changes are tested and preserved if R improves. Such small changes are tested 200 times in cubic symmetry, and up to 5000 times in triclinic symmetry. The amplitude of the cell parameter change is determined at random but cannot be larger than |delta max| = 0.02 Å. If a very good R factor is obtained finally (at R < R3 ~ 5%) the cell parameters are really submitted to a least square refinement and the classical figures of merit (FoM) (M_{20}, F_{20}) are calculated. For that selected result, pseudo powder patterns "observed" and calculated are simulated with Gaussian peak shapes and a .prf file is created which can be displayed by WinPlotr. Other computer programs are compatible with some McMaille output files (CHEKCELL and CRYSFIRE.).

The possibility of being trapped in a false minimum is a problem occuring in most Monte Carlo process. This can be avoided by accepting time to time a change in the cell parameters which does not improve the R factor. The best probability P to accept a new cell parameter in spite of a R factor not improved was found to be close to 15%, as shown in the table below, giving the number of times the correct cell was obtained for the same number of MC events as a function of P:

P (%)	**0**	**15**	**30**	**45**	**60**	**75**	**100**
Test 1 – orthorhombic	41	45	32	27	15	6	1
Test 2 – rhombohedric	28	41	40	28	17	10	6
Test 4 – monoclinic	47	60	46	45	25	19	2
Test 6 – triclinic	36	42	36	24	18	12	12

P : a value of 15 signifies that in 15% of the tests, a change in cell parameters will be accepted even if that change does not improve R or does not improve the number of indexed reflections.
for P = 100 : a change will always be accepted
for P = 0 : a change will not be accepted if R is not improved.

McMaille is a robust indexing computer program which can possibly cope with impurity lines. The user decides of the insensibility degree by means of two adjustable parameters:

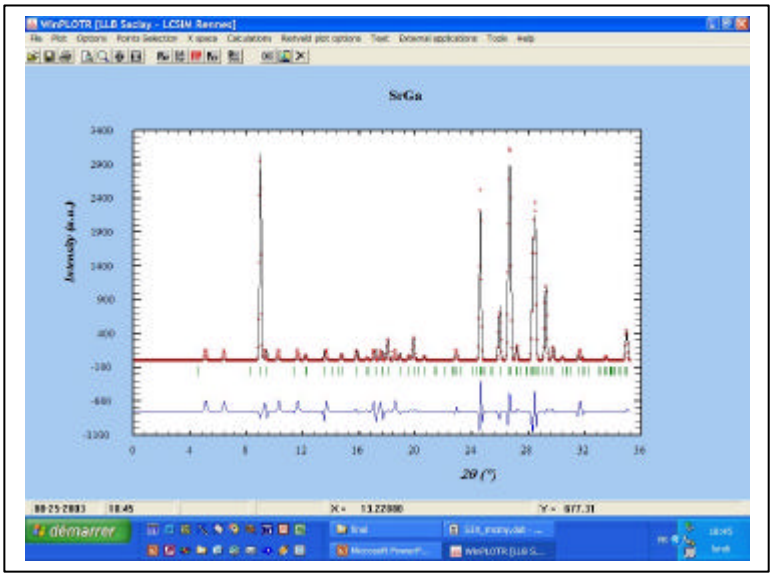
N' : the number of tolerated unindexed lines,

R2 : a R factor limit allowing to consider only solutions for which R < R2. To fix R2 at 15% signifies that cell proposals explaining at least 85% of the total intensity on the powder pattern (in the range of the selected peaks) will be preserved into the final list of the possibilities.

Thus, it is considered that an impurity phase should not concern more than 15% of the total intensity of the powder pattern. Nevertheless, the number of weak intensity peaks belonging to the impurity phase could be large, and this would force to select N' large enough (if one decides to add many weak intensity peaks in the starting list of data).

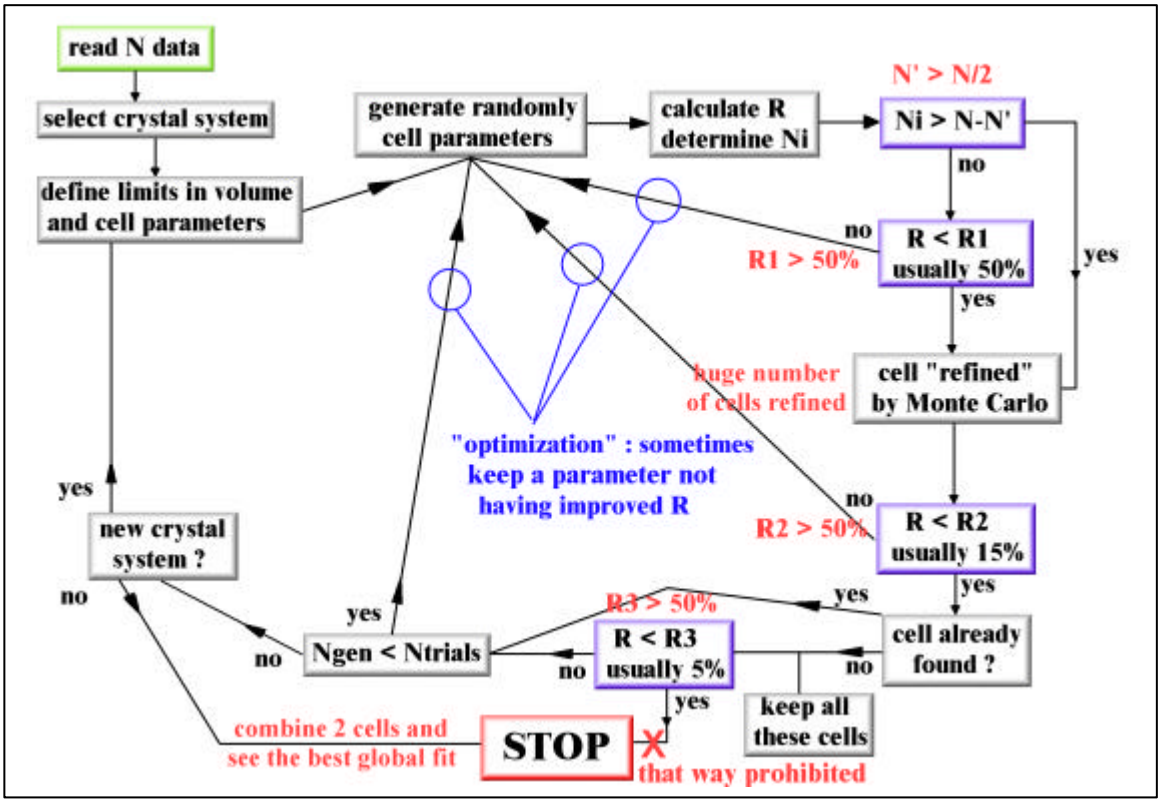
The conclusions of a systematic study about impurity peaks when using McMaille are the following. If the intensity associated to the non indexed peaks stays smaller than 15% of the total intensity, then :

- for a number of peaks smaller than 30% of the total number of peaks, McMaille generally succeeds in finding the correct cell placed at the head of the list (however the FoM decrease (R increases);
- for a number of additional peaks in the 35 to 50% range of the total number of peaks, McMaille may still propose the correct cell but not always ranked at the top of the list. It is then more difficult to conclude.



Example of an orthorhombic phase indexed in spite of 12 small peaks added to 20 peaks of the main phase (37,5% in number but less than 15% in intensité)

Beyond that problem of impurities, it can be conceivable to attempt to index multiphase powder patterns (obtaining simultaneously several cells explaining successfully the whole powder pattern). It is certainly better to try to optimize the synthesis conditions. Several synthesis made by systematically varying the chemical composition may evidence that a sample is a multiphased one (if systematic variations of the intensities of some given peaks are observed - but preferred orientation should be avoided - then peaks can be attributed to one or another phase). If really the efforts in synthesis are inconclusive and that it is believed that the samples are mixtures of several phases with unknown crystal structures, then you may try the option "multiphase" of McMaille. This option has a not negligible cost as can be seen on the flow chart :



The cost results essentially in prohibitive calculation times, because the number of cell proposals that have to be examined more carefully (by Monte Carlo adjustment if R < R1) becomes enormous given that R1 has to be fixed much larger than 50% (up to 75%). And also the R2 limit has to be large, similar to R1. Finally, R3 itself has to be larger than 50%. A series of tests using phase mixtures allowed to obtain some conclusions about the effectiveness of McMaille concerning two-phase samples. If 30 peaks are examined and that 13 to 17 peaks belong to one or to the other phase, and that 40 to 60% of the total intensity belongs to one or the other phase, then McMaille appears able to produce correct solutions in reasonable time (less than one hour) for combinations of two phases that would not be monoclinic nor triclinic for both of them. For these two low symmetries, the calculation times are really becoming too long.

```
==================================================================
Double cells with largest number of peak indexed
==================================================================

WARNING - WARNING - WARNING - WARNING - WARNING
          This is the two-phase mode
   It could be better to go back to the lab
        and try and make a pure sample

  Rp2    Vol      Vol/V1 Ind Nsol    a        b        c      alpha  beta  gamma
 0.259  1188.120  1.00  30  13  11.1880  11.1880   9.4919  90.000  90.000  90.000
 0.106   378.244  1.00  15   4  10.0276   3.4206  11.0274  90.000  90.000  90.000

 0.259  1188.120  1.00  29  13  11.1880  11.1880   9.4919  90.000  90.000  90.000
 0.100   507.781  1.00  12   1   4.5923  10.0280  11.0265  90.000  90.000  90.000

 0.259  1188.120  1.00  29  13  11.1880  11.1880   9.4919  90.000  90.000  90.000
 0.125   788.734  1.00  14   1  11.0263   7.1331  10.0282  90.000  90.000  90.000

 0.259  1188.120  1.00  29  13  11.1880  11.1880   9.4919  90.000  90.000  90.000
 0.140  1344.555  1.00  16   4  12.1596  10.0281  11.0266  90.000  90.000  90.000

 0.259  1188.120  1.00  29  13  11.1880  11.1880   9.4919  90.000  90.000  90.000
 0.110   516.988  1.00  13   2   4.6750  10.0275  11.0282  90.000  90.000  90.000
```

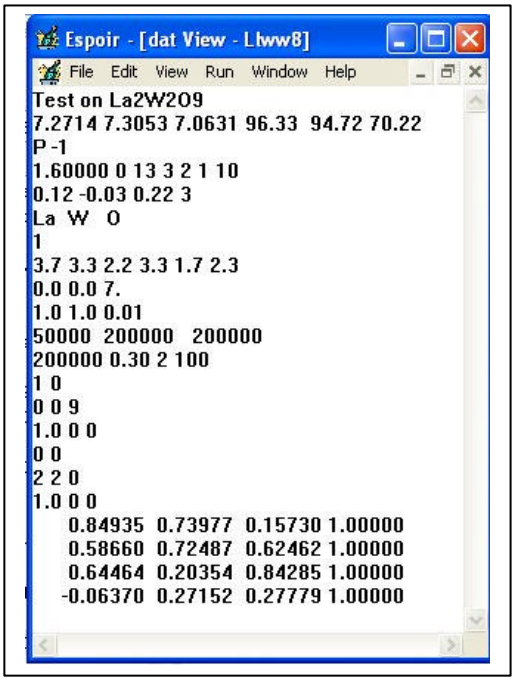
Example of indexing a mixture of 2 phases (20 minutes calculation). One phase is tetragonal, the other is orthorhombic. McMaille combines the cell proposals by two of them, and selects the best combinations indexing the largest number of peaks:

ESPOIR - Structure solution in direct space

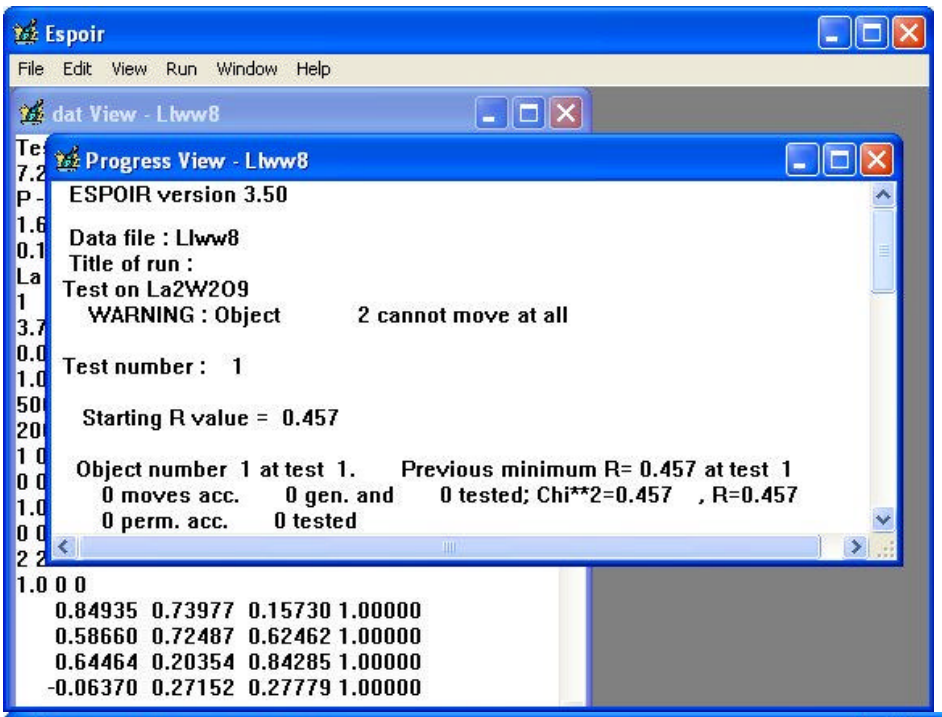
Installation - The package including the manual, examples and an executable for the MS Windows XP operating system on a Personal Computer (PC) is avaible at the URL http://www.cristal.org/sdpd/espoir/. There is no need to sign for any licence, the package is directly available and even the source code is downloadable. The installation is straightforward, after unzipping the whole package in any directory.

Running ESPOIR - At this step of a structure determination by powder diffractometry (SDPD), a cell is known (obtained from McMaille for instance), so that it is possible to extract the structure factors from the powder pattern after having selected the appropriate space group. ESPOIR is configured to accept the extracted "|Fobs|" according to the so-called Le Bail method implemented in the program Fullprof (or in GSAS, etc). Fullprof produces a .fou file containing these "|Fobs|" which will be used by ESPOIR (after removing some text from the .fou file and renaming it with .hkl extension). Other required knowledge is a good idea of the chemical composition or even more for organic and organometallic compounds : the three-dimensional molecular formula. ESPOIR, as well as many other programs working in direct space, is quite versatile and may undertake various kinds of problems characterized by partial or total chemical knowledge. Two selected examples are described below.

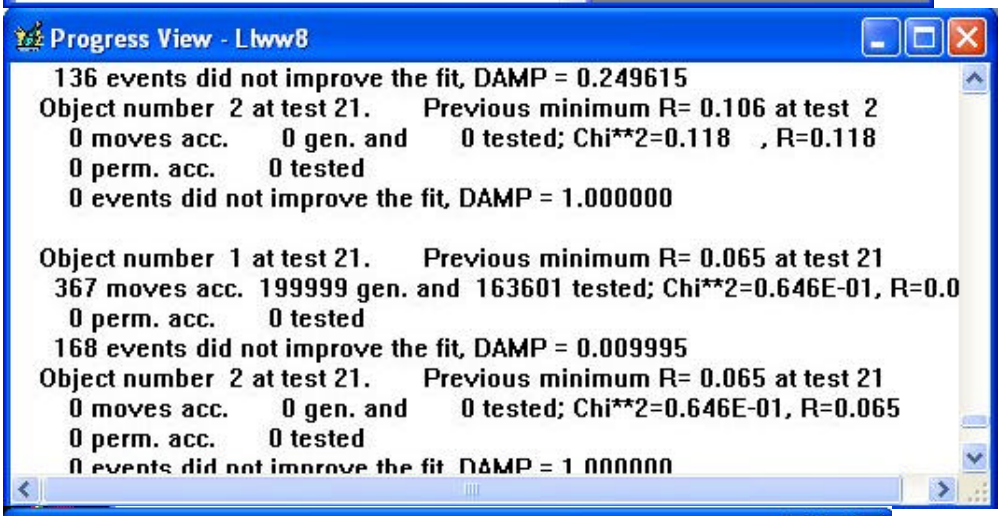
First example : $La_2W_2O_9$ - The purpose is to locate the 9 oxygen atoms from the neutron powder pattern of $La_2W_2O_9$, knowing already the positions of the La and W atoms, these atomic positions coming from powder X-ray data (located by direct methods). Why neutrons ? This is because the conventional X-ray data did not revealed accurate positions for these oxygen atoms, being quite light if compared to the tungsten and lanthanum atoms (R_B was already as small as 15% with only the W and La atoms). ESPOIR accepts the partial model built up from the W and La atoms, and does not modify it at all, allowing to introduce 9 oxygen atoms at random, moving them up to obtain the best fit between the calculated powder pattern and a pseudo "observed" pattern regenerated from the extracted "|Fobs|". Using that pseudo-observed powder pattern allows to escape from peak overlapping problems. Fifty independent tests of 200000 oxygen atom moves (10 millions Monte Carlo events) will be performed in a few minutes. The data used are the 150 first "|Fobs|" extracted from the netron powder pattern by using the Fullprof program. This is considered to be sufficient for the estimation of the 27 atomic coordinates *x,y,z* of these 9 oxygen atoms. The solution is found in 10% of the tests, obtaining R_P close to 6%.



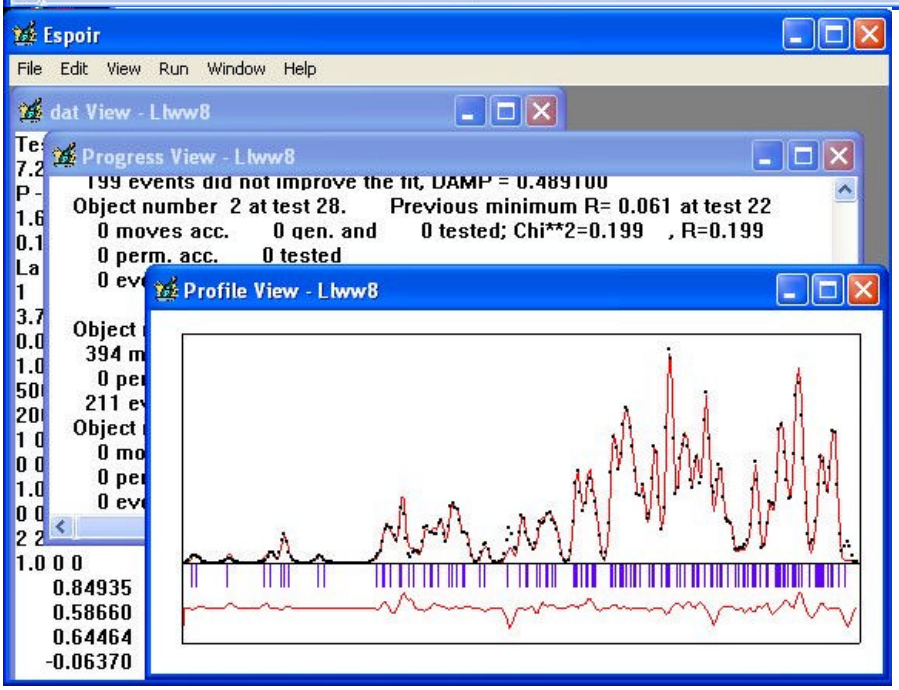
Small data file for ESPOIR containing the problem definition : Cell parameters, space group, wavelength, U,V,W, chemical content, interatomic distance restraints, maximum moves of the atoms, conditions of the Monte Carlo tests (number of events per test, number of tests), définition of the objects (here an object is stationary : 2 La + 2 W atoms, and a the other object is constituted by 9 oxygen atoms first placed at random). A second file is necessary containing the Miller indices and the "|Fobs|" extracted from the powder pattern.



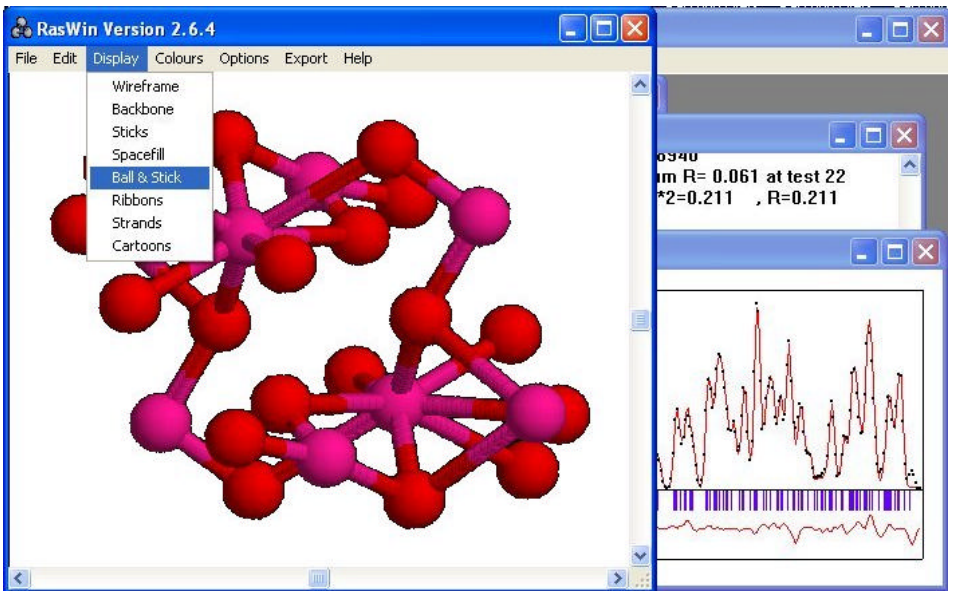
Once ESPOIR is started, reliabilities R are displayed together with various indications (numbers of moves tested, of moves accepted, of permutations made - if allowed - etc) :



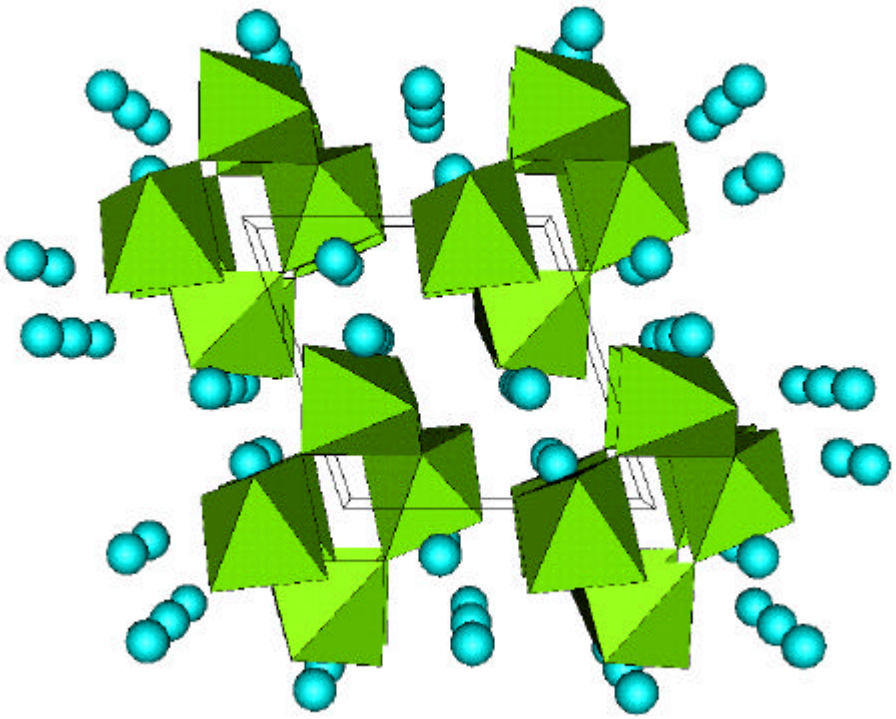
In that case, a minimum R = 0.065 is attained at the 21th test:



It is possible to display the pseudo powder pattern built from the peak positions and the extracted "|Fobs|". it is obvious here that the resolution is really bad. Peaks are overlapping a lot, and this is not really surprising since the cell is triclinic with parameters all close to 7 angstroms. The X-ray pattern had a quite better resolution.



The structure can be displayed with RASMOL, a public domain drawing program integrated in ESPOIR :



However, RASMOL is not really efficient with inorganic compounds. A projection of the structure is shown here by using STRUVIR (VRML version of STRUPLO) showing the WO_6 octahedra and WO_5 polyhedra forming the $La_2W_2O_9$ structure (The La atoms are represented by spheres) :

Second example : bethanechol chloride - In this second example, the position of the bethanechol molecule and that of a Cl atom will be established in the crystalline cell of the bethanechol chloride. When a sample is organic or organometallic and if the three-dimensional molecular formula is known, once the cell indexed and the space group established, one has only to position the molecule into the cell. The Monte Carlo method is well adapted to this kind of problem. In the case of the bethanechol chloride with ESPOIR, two objects are placed at random in the cell and will move independently. The first object is the molecule $C_7N_2O_2$ (without the 17 hydrogen atoms) describe in cartesian coordinates coming from a database. The second object is the chlorine atom. The molecule will be submitted to rotations and translations, and the Cl atom to translations, up to find a satisfying fit between the "observed" and calculated intensities.

Below is given the .dat file containing the problem description for running ESPOIR :

```
Bethanechol - molecule without H atoms + Cl random
8.880822  16.416883   7.140692  90.000000  93.829346  90.000000
P 21/N
0.6995 4 12 4 2 1 3
 0.027662   0.000000   0.000078   5
C   N   O   Cl
0
8. 8. 8. 8.
1.5 1. 0.02
20000 200000 100000
100000 0.30 2 20
2 4
7 2 2 0
1.5 0 0
0. 0. 0. 90. 90. 90.
      1.12559        1.62885        4.89741    1.00
      2.18536        3.78472        6.69582    1.00
      3.38781        2.29083        4.00240    1.00
      2.79920        2.34124        6.52033    1.00
      2.94085        0.19817        5.37523    1.00
      2.29373        4.13788        8.17289    1.00
      2.47432        5.31671        4.78811    1.00
      1.33316        5.24660        4.53231    1.00
      2.60361        1.67334        5.14159    1.00
      3.14464        4.77523        5.96655    1.00
      3.50507        6.27914        4.41895    1.00
1 0
0 0 0 1
2.5 0 0
```

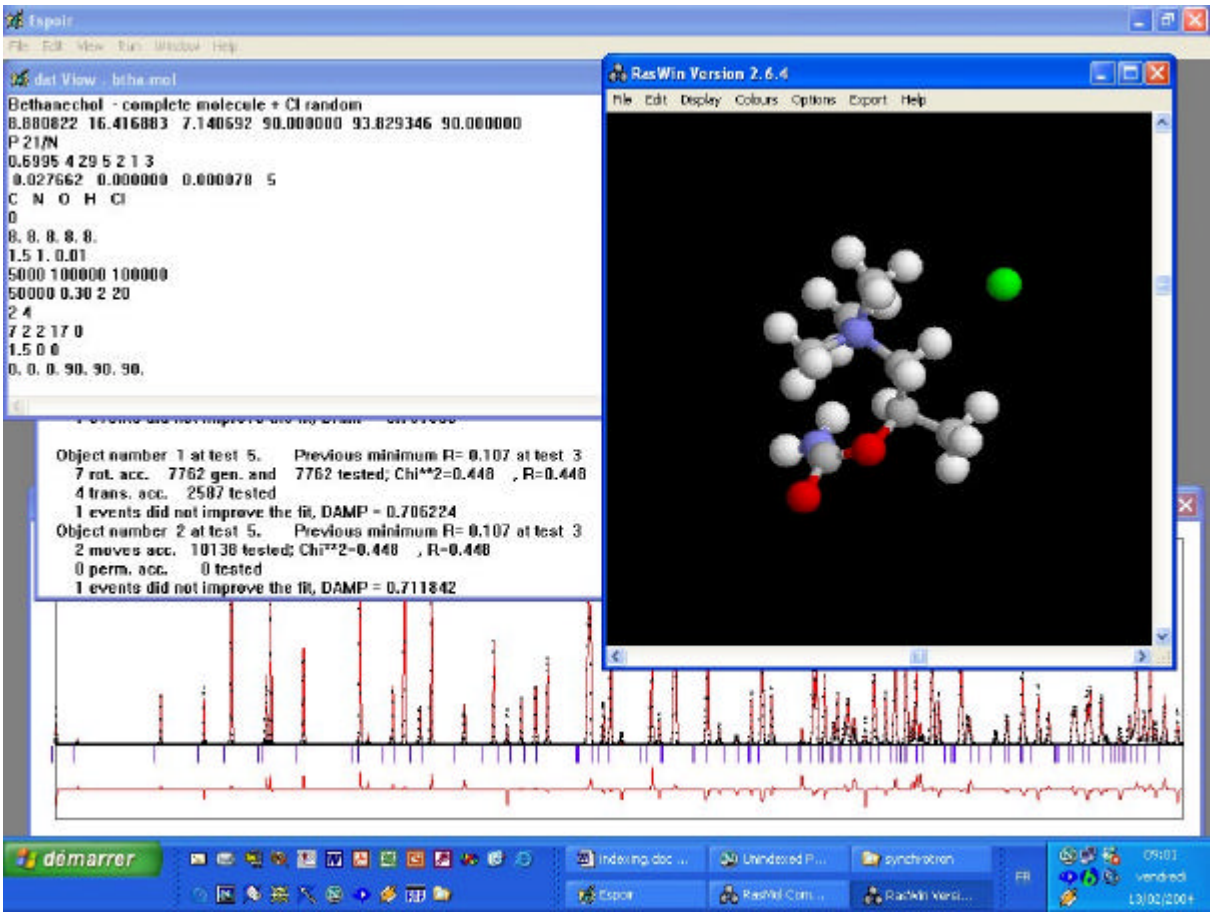
The *hkl* and "|Fobs|" extracted from the synchrotron powder pattern (by using Fullprof) are shown below. In such a case, the number of degrees of freedom is 9 (3 position and 3 orientation parameters for the molecule, and 3 position parameters for the chlorine atom) if the torsion angles are not allowed to change. So that only the first 50 "|Fobs|' will be sufficient. Using more would only lead to much longer calculation times.

```
50
   0   2   0   22.08    0.00
   1   1   0    5.72    0.04
   0   1   1   57.71    0.00
   1   2   0   53.90    0.00
  -1   0   1  187.76    0.00
  -1   1   1   86.45    0.00
   0   2   1   97.82    0.00
   1   0   1   91.13    0.00
   1   1   1  126.19    0.00
  -1   2   1  225.36    0.00
   1   3   0   15.09    0.00
   1   2   1   83.88    0.00
   2   0   0  280.27    0.00
   0   3   1   53.58    0.00
   2   1   0  265.62    0.00
   0   4   0   58.99    0.00
  -1   3   1  158.44    0.00
   2   2   0   29.81    0.00
   1   3   1   74.19    0.00
  -2   1   1   95.18    0.00
   1   4   0  135.27    0.00
   2   1   1  130.60    0.00
   0   0   2  264.85    0.00
   0   4   1  196.57    0.00
  -2   2   1   64.65    0.00
   0   1   2  130.57    0.00
```

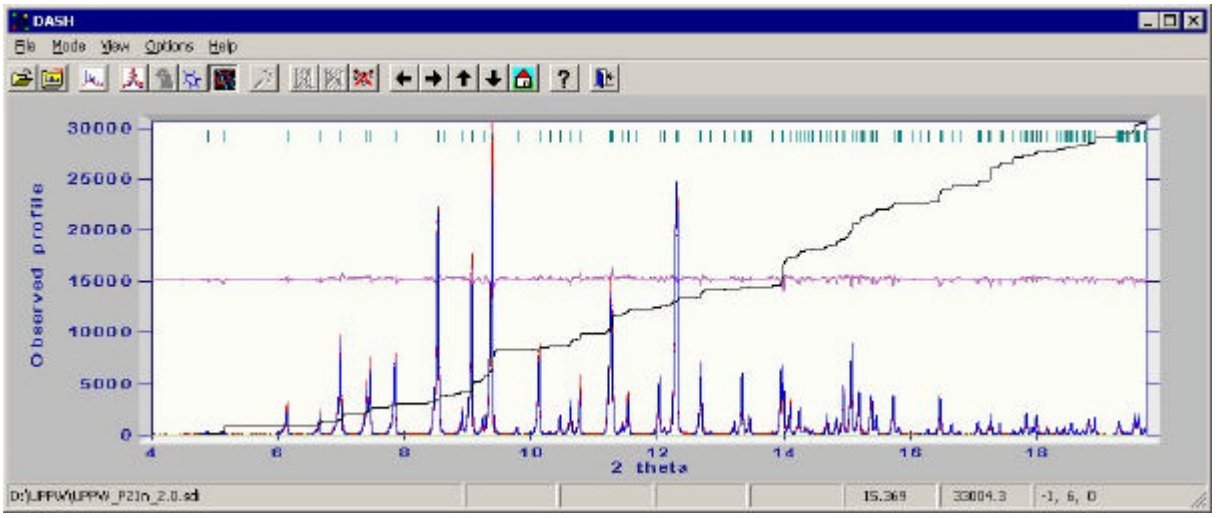
```
  -1    2    2    29.44     0.00
   1    5    0     6.48     0.24
```

Etc...

The solution is quickly obtained. It would be obtained a bit slowly if the 17 H atoms were added. Below is a screen copyshowing ESPOIR during a run including the H atoms. A reliability R value close to 10% is obtained at the third test :



An application of the DASH program to the same problem was obtained from J. van de Streek (CCDC) during the UPPW Round Robin, with an analogous success :



GRINSP - Geometrically Restrained INorganic Structure Prediction

Installation - The package including the manual, examples and an executable for the MS Windows XP operating system on a Personal Computer (PC) is avaible at the URL http://www.cristal.org/grinsp/. There is no need to sign for any licence, the package is directly available and even the source code is downloadable. The installation is straightforward, after unzipping the whole package in any directory. Note that some files like connectivity.txt, distgrinsp.txt and wyckoff.txt have to be absolutely in the same directory as grinsp.exe and the .dat file defining the run of your expected predictions.

Running GRINSP - Once it is decided to explore some chemical composition M_aX_b or $M_aM'_bX_c$ (the version 1 of GRINSP is limited to the building of N-connected 3D networks having such formulations, with N = 3, 4, 5, or 6 and mixtures N/N'), then the first thing to do is to verify if the distgrinsp.txt file distributed with the package contains the appropriate geometrical restraints : the interatomic distances for the M/X atom pairs (ideal distances for first neighbours M-X, X-X and M-M, as well as ranges of distances tolerated during the exploration: the estimated minimal and maximal second neighbour distances (M/M')-(M/M') are also requested). Such distances can be found in databases (ICSD, AMCSD, COD) by looking at known compounds belonging to the family you wish to explore, or can be found in the International Tables for Crystallography, or in books or publications about structure and bonding. Then you have to prepare a .dat file containing the description of the problem to be explored. A typical example of .dat file delivered with this version (edi.dat) is detailed below (see the GRINSP manual for full explanations):

```
EDI                                     : text for this run
P 1                                     : SG : space group
2 1 0 1                                 : nsym, npol, ncon, nmim
4                                       : ncpol (coordination for every polyhedron-type)
Si  O                                   : elements for a search in the distgrinsp.txt file
6.7   7.1   6.7   7.1   6.2  6 .6       : min and max cell parameters a, b, c
90. 90. 90. 90. 90. 90.                 : min and max angles
10. 30.                                 : min and max framework density (FD)
100000  20000  0.01                     : nruns, genmax, Rmax
6000  1                  : idls (MC cycles for distance refinement) and iref ([1] or not [0]))
4001000                  : first filename (will be 4001000.*, 4001001.*, etc) .dat, .cif, .xtl
```

Most parameters are self-explanatory for crystallographers. Some require to read the short manual.

Another example of .dat file for exploring a ternary system, note the main differences :

```
Mixte 6-3 ReO6/BO3
P 1
4 2 0 2                  npol = 2 here
6 3                      note the two coordinations defined here (octahedron and triangle)
Re  O                    one line for an ReO6 octahedron
B   O                    and one line for a BO3 triangle
3. 13. 3. 13. 3. 13.
90. 90. 90. 90. 90. 90.
10. 30.
100000 200000 0.05
6000 1
2042101
```

Below is a part of the manual :

nsym, npol, ncon, nmim : four integers (free format)

nsym = allows to define the constraints on the box cell parameters
1 : cubic
2 : tetragonal
3 : hexagonal
4 : orthorhombic
5 : monoclinic
6 : triclinic
Note that it may occur that the real symmetry will be different from the symmetry imposed here...
A way to find the true final symmetry is to apply the **PLATON** software to the CIF file. PLATON would find the symmetry elements missing in the P1 (or else) space group description.

npol = number of different types of polyhedra you wish in the predicted structure
Warning : only npol =1 or npol =2 are allowed in GRINSP version 1.00.

ncon = defines the degree of connection between polyhedra
0 : every X atom is connected to two M/M' atoms (corner-sharing)
> 0 : some X atoms are connected to only one polyhedron
< 0 : there could be edge-sharing, etc
Warning : only ncon=0 is working in GRINSP version 1.00

nmin = minimal mumber of M/M' atoms for saving a model
(if you are tired to see these small stuctures with 2, 3 or 4 M atoms only all the time)
For exploring a M/M' system, only solutions mixing M and M' will be retained (meaning that the minimum will be nmin=2, anyway)

ncpol = coordination for every of the npol different polyhedron-type
Warning ! only four values ncpol=3, 4, 5 or 6 are allowed in GRINSP version 1.00
and a maximum of 1 or 2 values can be given (because npol =1 or 2 maximum)

Elements : two elements for a given polyhedron - there may be two lines if a ternary $M_aM'_bX_c$ system is explored. **This part is formatted as 2A4 : four spaces per atom.**

GRINSP will have to find the minimal/maximal/ideal interatomic distances for this polyhedron inside of the file **distgrinsp.txt**
After the same atom codes in distgrinsp.txt is added the coordination, and then 4 lines corresponding to the prescribed interatomic distances :

Si O 4	: atom codes and coordination (here a SiO4 tetrahedron)
2.60 3.60 3.070	: minimal/maximal/ideal first distance between atom pair 1-1
1.30 1.90 1.610	: minimal/maximal/ideal first distance between atom pair 1-2
2.20 3.00 2.629	: minimal/maximal/ideal first distance between atom pair 2-2
4.40 6.00	: minimal/maximal second distance between atom pairs 1-1

You may edit distgrinsp.txt and add there your own data (anywhere) : 5 lines per kind of polyhedron as above. Care that the elements are given in 2A4 format.

Etc.

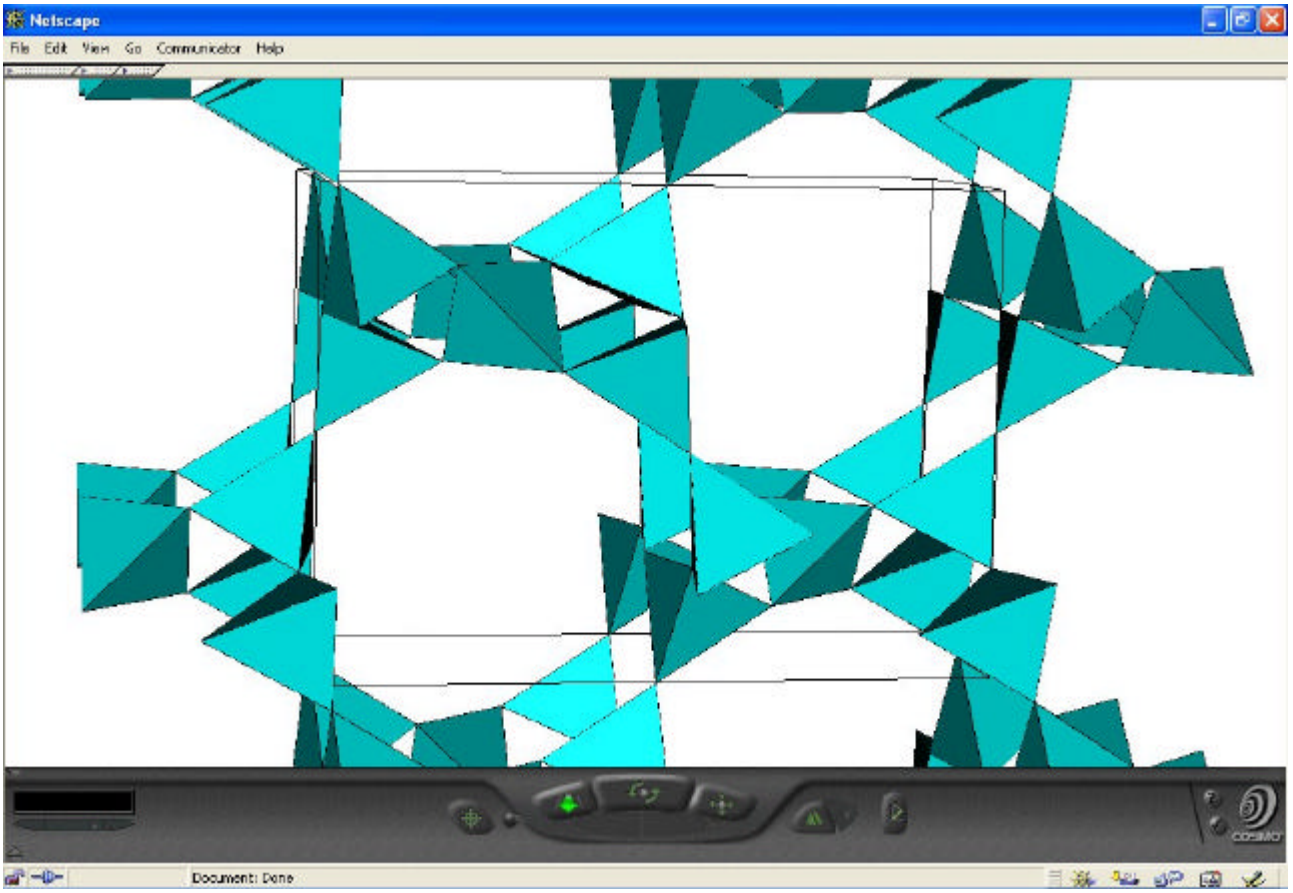
The Windows PC version will run by clicking on grinsp.exe, opening a window shown below. The name of your .dat file has to be given (test48.dat in the example below, but the .dat default extension should not be provided). The entry data parameters are displayed, as well as a summary for every successful prediction. You may stop the program execution by typing K (capital letter) anytime. The program will store and sort the results and stop at the next multiple of 50 runs (so the stop is not immediate but may need a few seconds).

```
C:\Fortran\GRINSP\Zeolites\48\GRINSP.exe

 Input file (no extension)
test48
 Test


 Space Group : P N N N

 constraint on symmetry : orthorhombic box
 Number of different types of polyhedra:            1
 Connectivity degree (0 = 3D net):            0
 Minimum number of polyhedra in the cell:            4
 Polyhedra coordinations:            4
 For polyhedra            1
 pair i-j :1-1 min/max/ideal distances : 2.570 3.570 3.070
 pair i-j :1-2 min/max/ideal distances : 1.310 1.910 1.610
 pair i-j :2-2 min/max/ideal distances : 2.229 3.029 2.629
 second pairs 1-1    min/max distances : 4.400 6.000
 Min and max a, b, c :    6.0000  16.0000   6.0000  16.0000   6.0000  16.0000
 Min and max angles  :   90.00  90.00  90.00  90.00  90.00  90.00
 Min and max framework density :   5.00 30.00
  Nrun, genmax, rdtm :       200000   300000.0      1.5000000E-02
  Monte Carlo cycles of distances refinement:        10000
  Cell refined
  Run      MCgen       a         b         c      alpha  beta   gamma  nT     R
   997     41060.  12.7539    7.1086   11.5928  90.00  90.00  90.00 20  0.0142
  1889        48.   8.9532    8.9631    8.9549  90.00  90.00  90.00 12  0.0065
  2369        14.   8.1814    7.2712    5.3331  90.00  90.00  90.00  8  0.0105
  2825         3.   8.9537    8.9612    8.9579  90.00  90.00  90.00 12  0.0066
  3189     70754.  10.8024   10.4503    7.0341  90.00  90.00  90.00 16  0.0087
  3314        11.   5.3287    8.5723    6.4562  90.00  90.00  90.00  8  0.0123
  3390     35673.   5.1767   12.3598    9.2257  90.00  90.00  90.00 10  0.0101
  3554     33578.   7.0968   10.4930   10.6955  90.00  90.00  90.00 16  0.0090
  3790    147522.   9.1001    5.1811   12.6066  90.00  90.00  90.00 10  0.0088
  4199       222.   8.9590    8.9538    8.9629  90.00  90.00  90.00 12  0.0066
  4506      1205.   9.4473    7.7709   16.2766  90.00  90.00  90.00 22  0.0146
  4771       185.  12.0784    9.2424    5.1734  90.00  90.00  90.00 10  0.0101
  4809     57567.   6.6618    6.6781   13.9134  90.00  90.00  90.00 16  0.0133
  5428    120115.   4.8837   10.6492    9.0351  90.00  90.00  90.00 12  0.0086
  5464      8104.   8.9609    8.9472    8.9501  90.00  90.00  90.00 12  0.0061
  6253      4620.  12.8591    5.1731    9.0156  90.00  90.00  90.00 10  0.0095
  6448       123.   9.0243    4.9203   10.6935  90.00  90.00  90.00 12  0.0079
```

GRINSP creates several output files :

name.imp : summarizes all the results (including FD, the framework density)
name.con : contains the coordination sequences (CS) of all predicted compounds in a run
filename.dat : atomic coordinates in STRUPLO/STRUVIR format, allowing a direct look at the structure in VRML
filename.xtl : atomic coordinates in the CERIUS Accelrys format which may be used by the zeoTsites software for coordination sequences, vertex symbols and geometry analysis.
filename.cif : atomic coordinates in the IUCr CIF format

The STRUVIR program (also distributed at the Web site: http://www.cristal.org/) produces interatomic distances calculations and prepares two VRML files (po.wrl and wi.wrl, respectively for displaying the crystal structure with polyhedra or in wireframe style). A good VRML viewer is CosmoPlayer shown on the next picture :



General prediction - If the aim is to explore the possible crystal structures showing corner-sharing in M_2X_3 (triangles) , MX_2 (tetrahedra), MX_3 (orctahedra) or mixte compounds $M_aM'_bX_c$, then it is recommended to examine first cubic cells, going from high symmetry to low symmetry. If cells are predicted in cubic, store the new coordination sequences (CS) in the connectivity.txt file, then test with tetragonal cells, store the new CS if any, test with hexagonal cells (etc : orthorhombic, monoclinic and finally triclinic completing the full exploration). New CIF files are generated only if a new CS occurs that is not already into the connectivity.txt file. However, the name.imp global file will contain all predicted results so that if you are searching for isostructural compounds, you may want to reduce the content of the connectivity.txt file to a very minimum, so that all predictions will be estimated to be new and you will obtain the CIF files. Using a negative first file number allows to create all predicted structure files.

Using GRINSP as a structure determination tool for zeolites or MX_3 compounds (or etc): If the aim is to explore a small range of cell parameters and if the symmetry is known, this is even more simple, see the EDI example above. Restrain the search to the known cell parameters (allow a small tolerance).

The final structure is still described in the P1 space group, but you will find in the .imp file the general or special positions occupied by the cations in the selected working space group before the final refinement made in the P1 space group.

GRINSP does not work by applying simulated annealing to a starting random configuration. Version 1.00 works schematically as follows, by using the Monte Carlo method :

- Manual selection of the constraints on cell parameters, of restrained interatomic distances, of the type(s) of coordinations, and of the space group. Then the Monte Carlo process starts.
- Random selection of the cell parameters inside of the predefined range.

- Random positioning of a first cation M (or M') of the future M_aX_b (or $M_aM'_bX_c$) compound on a general or special position, itself selected randomly.
- Random positioning of the next cations (random choice of M or M') in respect of the distance restraints with the previous ones, on a general or special position, itself selected randomly.
- If a model fulfills all distance criteria, place the X atoms at M-M midpoints, "refine" the atomic positions and cell parameters by a Monte Carlo process so as to improve an R factor.
- Continue to try to predict structures in that way till a certain number of independent runs are made.
- Find if the predicted structures are new or were already described (using CS - Coordination Sequences).

In the GRINSP algorithm, the number of M or M' atoms in a randomly selected cell is not predetermined, it is predicted as well. Only distances are considered (not angles - though considering a range for the second M-M distances is like restraining angles).

Currently, there are many limitations in that beta version. GRINSP 1.00 proved to be efficient for a maximum number of 64 M/M' atoms on up to 4 different general or special positions (10-15 M/M' atoms in a triclinic cell which you may force to be cubic-like or hexagonal-like or etc). It was shown to be able to predict many zeolites (ABW, ACO, AFI, BIK, EDI, GIS, LTA, SOD... but not all of them since the explorations were limited to cell parameters < 16 Å) or the compact SiO_2 phases (quartz, cristobalite, tridymite, etc), polymorphs for B_2O_3, AlF_3, hypothetical phases in binary systems B_2O_3/SiO_2, B_2O_3/ReO_3, SiO_2/ReO_3 (see the PCOD [9]). It is up to you to try GRINSP with other systems, and even the above ones have not been completely explored. One life would not be sufficient.

Further work is needed for improving the GRINSP efficiency :

- Introduction of different linkage modes than by corners (edges, faces...)
- Adding the possibility for insertion of big cations K/Sr/Ba/Cs/etc as spheres in the holes/tunnels
- Considering bond valence as an alternative to pure geometrical restraints for the model final refinements
- Increase speed by not recalculating always everything (distances)
- Increase the box size for the CS (coordination sequence) calculations (729 cells is not always enough...)
- Etc !

Good luck with McMaille, ESPOIR and GRINSP !

Think to improve them by yourself and communicate the changes (GNU Public Licence).

References

[1] A. Le Bail, *Powder Diffraction* **19** (2004) 249-254. http://www.cristal.org/McMaille/
[2] A. Le Bail, *Mater. Sci. Forum* **378-381** (2001) 65-70. http://www.cristal.org/sdpd/espoir/
[3] A. Le Bail, *IUCr Comp. Comm. Newletter* **4** (2004) 37-45. http://www.cristal.org/grinsp/
[4] UPPW - Unindexed Powder Pattern of the Week - http://sdpd.univ-lemans.fr/UPPW/
[5] Indexing Benchmarks - http://sdpd.univ-lemans.fr/UPPW/benchmarks/
[6] J. Bergmann, A. Le Bail, R. Shirley & V. Zlokazov, *Z. Kristallogr.*, in press.
[7] A. Le Bail, P.W. Stephens & F. Hubert, *Eur. J. Miner.*, **15** (2003) 719-723.
[8] Y. Laligant, A. Le Bail and F. Goutenoire, *J. Solid State Chem.* **159** (2001) 223-227.
[9] PCOD (Predicted Crystallography Open Database) : http://www.crystallography.net/pcod/