

Progrès en Résolution de Structures par Diffractométrie de Poudres

Armel Le Bail

Université du Maine – CNRS UMR 6010

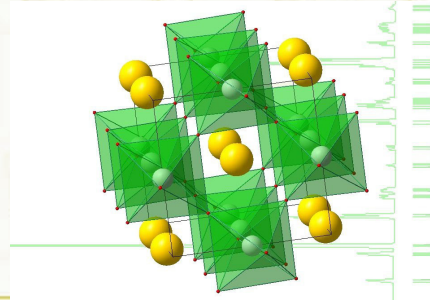
Laboratoires des Oxydes et Fluorures

alb@crystal.org

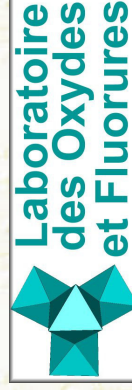
Les Rencontres LLB - SOLEIL

Diffraction de poudres

2-3 mars 2006



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE



PLAN

- A- Progrès en identification de phase**
- B- Progrès en indexation**
- C- Progrès en solution de structure**
- D- Hit parade des structures les plus complexes**
- E- Progrès en automatisation du processus complet**

Les « vraies » (?) raisons des progrès

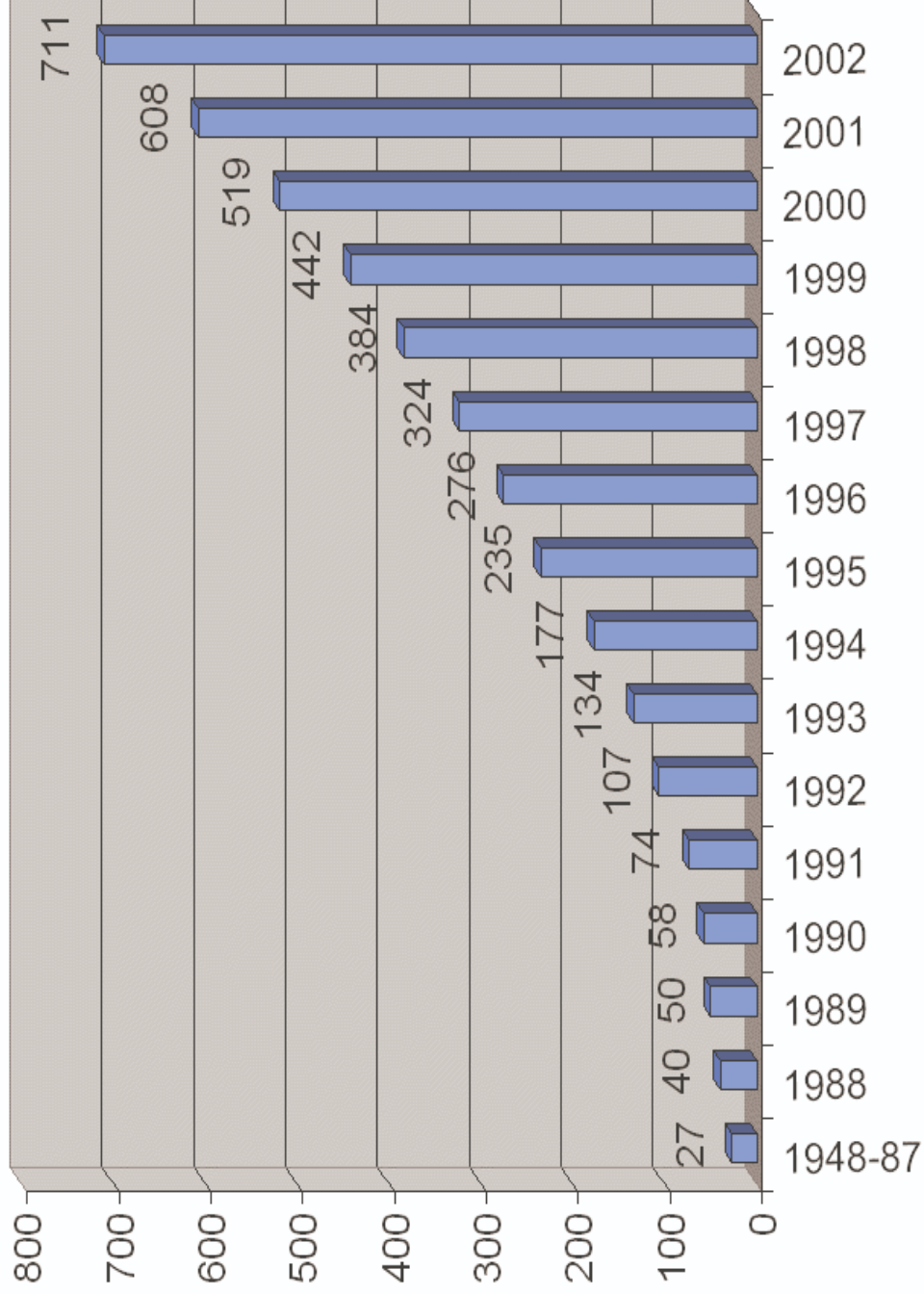
Conclusions

Progrès :

« Changement d'état graduel et continu qui consiste en une amélioration, un perfectionnement, ou un accroissement des connaissances. »

(Dictionnaire Hachette).

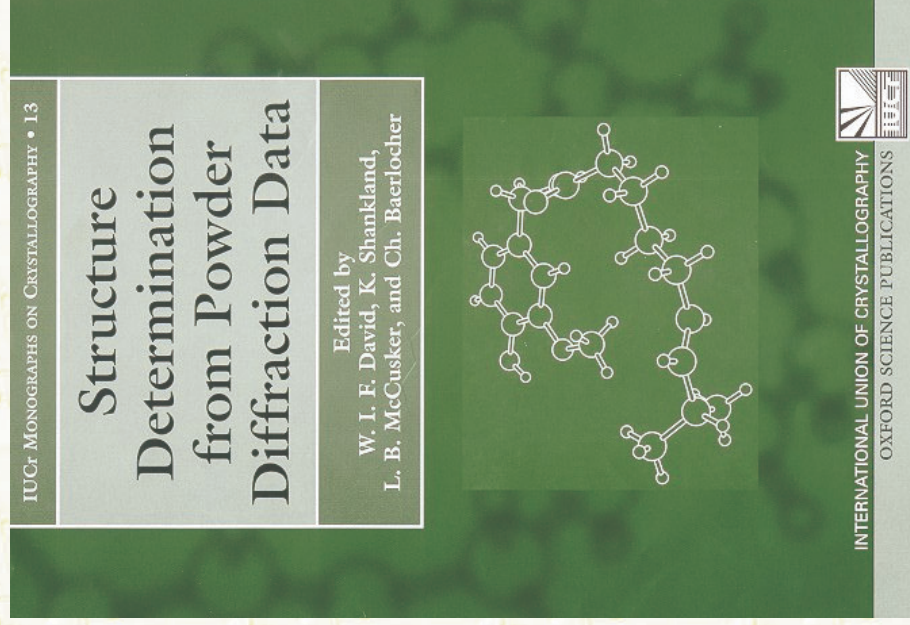
Accroissement du nombre de déterminations *ab initio* publiées



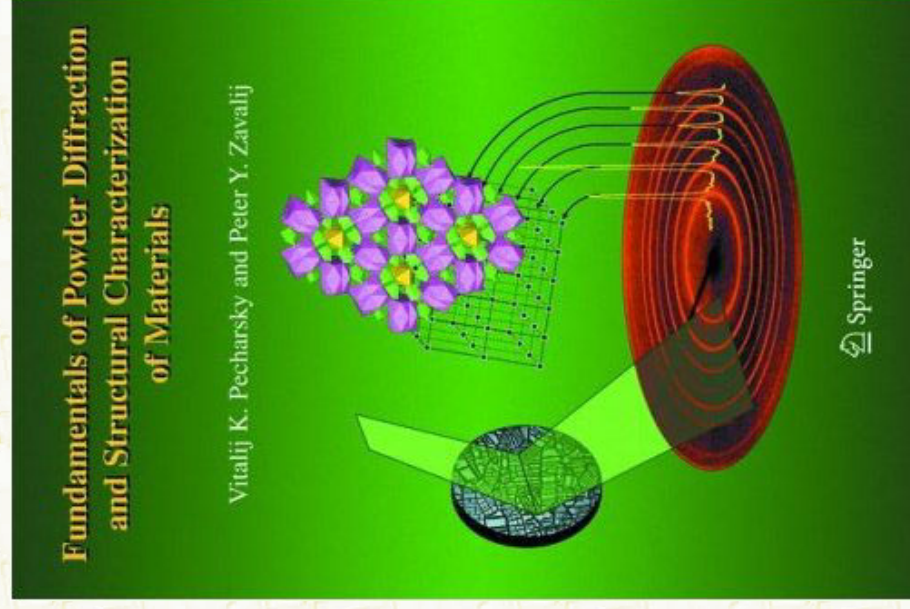
Histogramme cumulatif

Parution de livres

2002



2004



Sous-discipline reconnue : c'est un progrès...

A- Progrès en identification de phases

Par comparaison (search-match) entre diagramme théorique et diagrammes de phases connues en bases de données (JCPDS-ICDD).

2004	Entrées (total)	Inorganiques	Minéraux	Organiques	Coût (US \$)
PDF-2	157048	133370	?	25609	4920 - 875
PDF-4/Full File	157048	133370	?	25609	3785 - 720
PDF-4/Minerals	17535		17535		760 - 380
PDF-4/Organics	210000	Ajout CSD		210000	6090 - 2800
Total	339864 (il y a un peu de redondance)				
2001 (les divers PDF-4 n'existaient pas)					
PDF-2	131000				6100 - 1150
1999					
PDF-2	120000	Ajout ICSD			6100 - 1150
1995					
PDF-2	59800 (avant inclusion de données ICSD)				5750 - 1050

ICDD
Etat actuel
(pub février 2006)

Organiques :
~ CSD
286464

Inorganiques :
240050
accroissement par
inclusion du
« Pauling File »

*Identify, analyze and interpret your data with
material identification databases from the*

International Centre for Diffraction Data

PDF-4/Organics 2006

database features 286,464 organic and organometallic phases. It is a practical results-oriented product that includes drug active compounds, excipients, polymers, hydrates, acids, and pharmaceuticals.

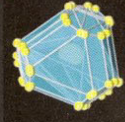


PDF-4+ 2005

database has comprehensive coverage for inorganic materials. It features 240,050 data sets, digitized patterns, molecular graphics and atomic parameters.



International Centre for Diffraction Data
Phone: 610.325.9814 Fax: 610.325.9823
marketing@icdd.com www.icdd.com www.dxcicdd.com



A-2- Progrès à venir ou en cours

Prédiction de structure cristalline
et donc prédiction de diagrammes de poudre
⇒ Bases de données
de diagrammes de poudre et
de structures hypothétiques

B- Progrès récents en indexation

B-1 – DICVOL04 remplace DICVOL91

Enfin Louër et Boultif cèdent à la tendance générale de tolérance d'impureté et de considération de décalage de zéro...;-)

B-2 – « indexing benchmarks » en 2004 :

Tout nouveau logiciel peut comparer ses performances à celles des logiciels concurrents.

Il va sans dire que ce n'est pas forcément du goût de tous les auteurs de logiciels...

**Mauvaise nouvelle
dans le domaine de l'indexation**

Robin Shirley nous a quitté en 2005

**Sa suite logicielle d'indexation CRYSFIRE
aura-t-elle un repreneur ?**

Dernière publication :

« Renewed interest in powder diffraction data indexing »,

J. Bergmann, A. Le Bail, R. Shirley and V. Zlokazov, Z. Kristallogr. 219 (2004) 783-790.

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	Note	
ITO13 standalone																						-14
DICVOL91 standalone																						-8
TREOR90 standalone																						-4
DICVOL04 standalone																						+3
McMaille standalone																						+5
First 5 above Combined																						+11
All combined																						+12
CRYSFIRE global																						+6
TAUP in CRYSFIRE																						0
KOHL in CRYSFIRE																						-5
TREOR90 in CRYSFIRE																						-8
DICVOL91 in CRYSFIRE																						-12
McMaille in CRYSFIRE																						-12
LOSH in CRYSFIRE																						-14
Mmap in CRYSFIRE																						-14
Hmap in CRYSFIRE																						-15
FJZN in CRYSFIRE																						-16
ITO12 in CRYSFIRE																						-16
LZON in CRYSFIRE																						-20
AUTOX																						
SVDindex in TOPAS																						
Xcell in Material Studio																						
	0	6	3	8	3	6	3	10	0	5	0	8	0	5	5	13	10	15	14	18		

Ne pas se limiter à un seul logiciel d'indexation...

Tendances en indexation

Approches Monte Carlo ou algorithme génétique appliquées au diagramme de poudre brut (non pas aux simples positions de pics mais aux positions + intensités) ou reconstitué.

Un logiciel précurseur (algorithme génétique) n'a jamais été disponible

J. Synchrotron Radiation **6** (1999) 87-92.

MAUD et **TOPAS (SVD-Index+)**
McMaille

Powder Diffraction **19** (2004) 249-254

C- Progrès en solution de structure

- 1- Méthodes classiques améliorées
Méthodes Directes, Patterson
- 2- Méthodes « nouvelles » dans l'espace direct
deviennent classiques
charge flipping, etc
- 3- Méthodes récentes
- 4- Approche mono ou multi diagrammes
pour se rapprocher du monocristal
- 5- Prédiction de structures cristallines
en plein boum... partielle ou totale

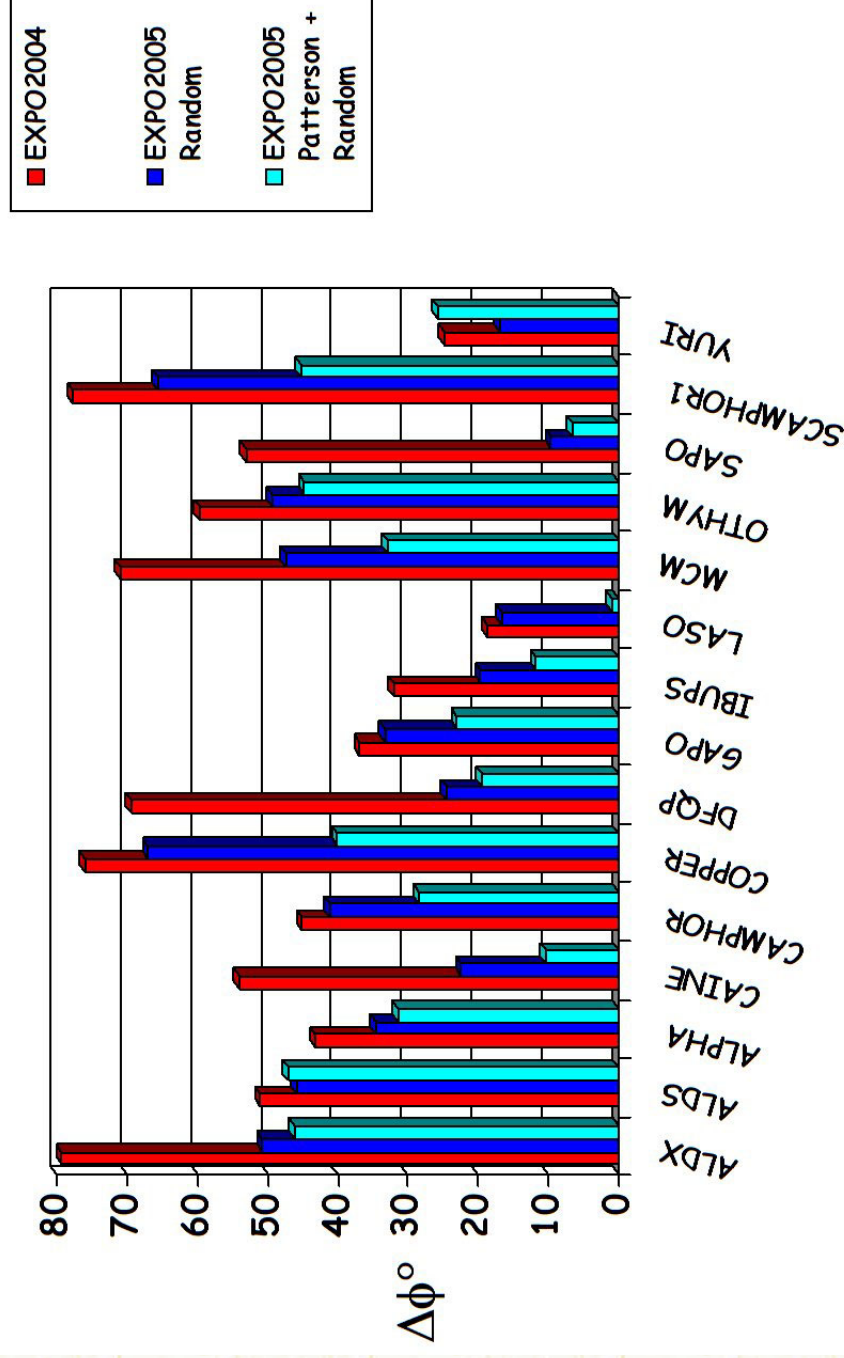
C-1 Méthodes directes / Patterson

EXPO2005, Groupe Giacobazzo

27 décompositions par Monte Carlo (partition au hasard)
20 tentatives d'attributions de phases
= 540 tentatives

Amélioration
des chances
de succès.

Réduction
des écarts de
phases.



C-2 Méthodes dans l'espace direct

Peuvent résoudre des problèmes d'autant plus complexes qu'une partie ou même tout le motif structural est connu - molécule(s), polyèdre(s).

Program	Access	GO	Data	Example	DoF
DASH	C	SA	P	Capsaicin	16
EAGER	A	GA	WP	Ph ₂ P(O)(CH ₂) ₇ P(O)Ph ₂	18
ENDEAVOUR	C	SA	I	Ag ₂ PdO ₂	45
ESPOIR	O	MC	L	Gormanite	54
FOX	O	SA	WP	Al ₂ (CH ₃ PO ₃) ₃	24
OCTOPUS	A	MC	WP	Red Fluorescein	7
POWDERSOLVE	C	MC	WP	Docetaxel	29
PSSP	O	SA	L	Malaria Pigment Beta Haematin	14
SAFE	A	SA	WP	C ₃₂ N ₃ O ₆ H ₅₃	23
SA	A	SA	WP	(CH ₂ CH ₂ O) ₆ :LiAsF ₆	79
TOPAS	C	SA	WP	Caffeine Anhydrous	93

Access : C = Commercial with academic prices, O = Open access, A = contact the authors

GO = Global Optimization : MC = Monte Carlo, SA = MC+Simulated Annealing, GA = Genetic Algorithm

Data : P = Pawley, L = Le Bail, I = Integrated intensities, WP = Whole Pattern

DoF = degrees of freedom corresponding to the example

Ref : Reference corresponding to the example (not necessarily to the program publication)

L'embaras du choix ? Open Access : **FOX, PSSP, ESPOIR**

C-3 Méthodes Nouvelles (?)

Iterated Projections, Principle of minimum charge

Acta Cryst. A **55**, 489 (1999)

Acta Cryst. A **59**, 201 (2003)

Veit Elser

two papers on

the charge-flipping algorithm:

Acta Cryst. A **60**, 134 (2004)

Acta Cryst. A **61**, 147 (2005)

Oszlányi & Sütő

Tout ce qui s'applique aux monocristaux est un jour ou l'autre testé sur les poudres, si possible... Attendons...

C-4 Résolution à partir d'un seul diagramme ou de plusieurs ?

Tendance actuelle :

Rapprocher les données de poudre des données sur monocristal
par

Exploitation maximale de tout ce qui peut réduire ou modifier le chevauchement des réflexions :
texture,
effets de dilatation thermique anisotrope,
substitution isotopique (neutrons),
dispersion anormale,
combinaison X/neutrons...

C'est de la grosse artillerie, à utiliser si nécessaire.

C-5a Prédiction Organique

Les progrès réalisés sont sensibles dans 3 publications, résultats de 'blind tests' de prédiction de structures organiques :

- Lommerse *et al.*, Acta Cryst. B56 (2000) 697-714.
- Motherwell *et al.*, Acta Cryst. B58 (2002) 647-661.
- Day *et al.*, Acta Cryst. B61 (2005) 511-527.

Les logiciels de prédiction proposent des centaines de modèles d'empilement de molécules, classés par énergie.

Pour des phases inconnues **non-indexées**, le choix final se fait par comparaison des diagrammes de poudre théoriques avec le diagramme observé.

Prédiction partielle (sans maille mais avec contenu connu)

Nouveau critère de similarité pour comparaison

« New similarity index for crystal structure determination from X-ray powder diagrams, »

D.W.M. Hofmann and L. Kuleshova,
J. Appl. Cryst. 38 (2005) 861-866.

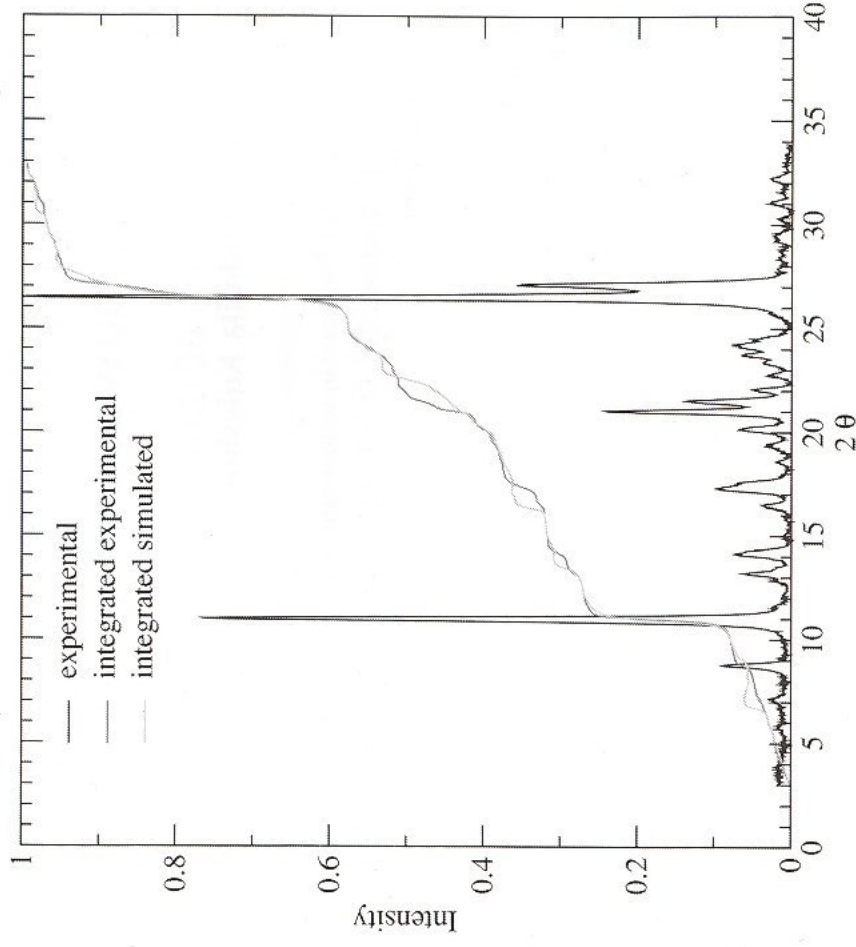


Figure 1

Comparison between experimental and integrated simulated and experimental powder diagrams for pigment PY111 for the predicted structure of rank 1, similarity index $s_{\text{int}} = 0.41\%$.

$$s_{\text{int}} = \frac{1}{n} \sum_{i_0=1}^n \left| \frac{\sum_{i=1}^{i_0} I_i^{\text{calc}}}{\sum_{j=1}^{i_0} I_j^{\text{calc}}} - \frac{\sum_{i=1}^{i_0} I_i^{\text{count}}}{\sum_{j=1}^{i_0} I_j^{\text{count}}} \right|. \quad (1)$$

C-5b Prédiction inorganique

Phases inorganiques : pas quelques centaines de modèles, produits mais des milliers ou bien plus encore dans certains cas !

>1.000.000 de zéolithes hypothétiques pour ~150 connues !

Foster et al., J. Appl. Cryst. 38 (2005) 1028-1030.

Tri pas réalisé : plusieurs modèles par zéotype.

Cette fois des bases de données s'imposent.

M.D. Foster & M.M.J. Treacy - Hypothetical Zeolites Database -
<http://www.hypotheticalzeolites.net/>

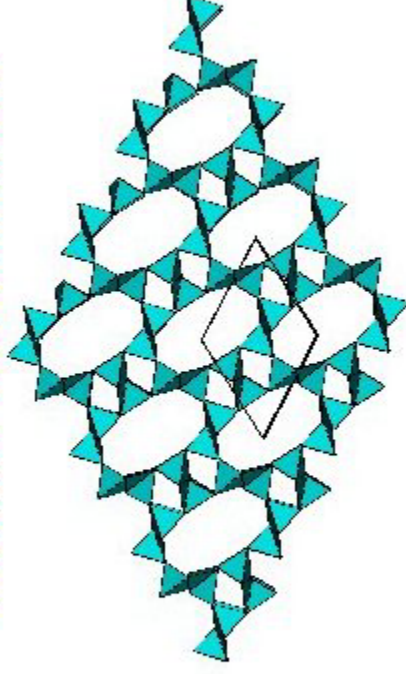
<http://www.crystallography.net/pcod/>

Predicted Crystallography Open Database

Upload data

or

Search the database



Updated February 2006: > 2600 entries in the PCOD:
 SiO_2 , B_2O_3 , AlF_3 , $\text{B}_x\text{Si}_y\text{O}_z$, $\text{B}_x\text{Re}_y\text{O}_z$, $\text{Si}_x\text{Re}_y\text{O}_z$ and $\text{Si}_x\text{Ti}_y\text{O}_z$
predicted - or enumerated - by **GRINSP** with $R_{\text{(DLS)}} < 0.02$,
or by other programs (see the "**What is New**" page).
See a **paper** published in open access at the *J. Appl. Cryst.* (2005).

All data on this site have been placed in the public domain by the contributors
Have also a look at the **COD** containing really observed crystal structures

Autre base :
PCOD

Prédictions par le logiciel

GRINSP :

J. Appl. Cryst. 38 (2005) 389-395.

Prédiction totale
de maille et de contenu

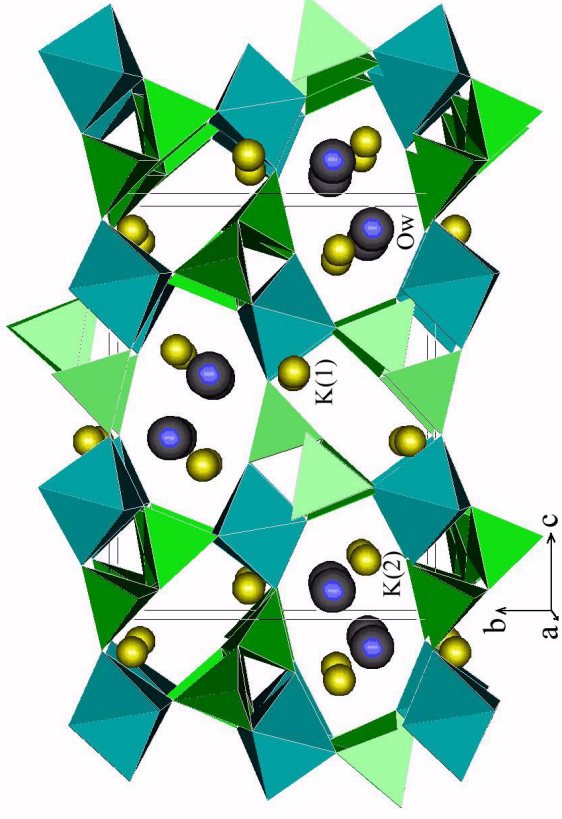
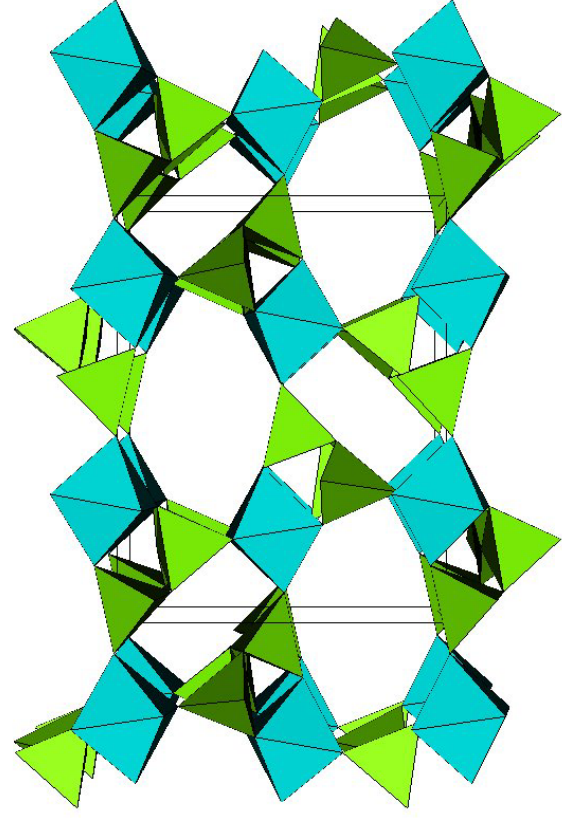
MAIS

pas général,
limitation aux réseaux
3D N-connectés

Exemple parmi les > 1000 titanosilicates dans PCOD, prédites par GRINSP

Modèle PCOD2200207 (Si_3TiO_9)²⁻ :
 $a = 7.22 \text{ \AA}$; $b = 9.97 \text{ \AA}$; $c = 12.93 \text{ \AA}$, SG $P2_12_12_1$

Connu comme $\text{K}_2\text{TiSi}_3\text{O}_9 \cdot \text{H}_2\text{O}$:
 $a = 7.1362 \text{ \AA}$; $b = 9.9084 \text{ \AA}$; $c = 12.9414 \text{ \AA}$, SG $P2_12_12_1$
(Eur. J. Solid State Inorg. Chem. 34, 1997, 381-390)



Pas si mal si on considère que K et H₂O ne sont pas pris en compte
dans la prédiction du modèle...

Les méthodes de solution (notamment dans l'espace direct) sont si efficaces qu'on se demande pourquoi prédire ? Surtout si la maille est connue....

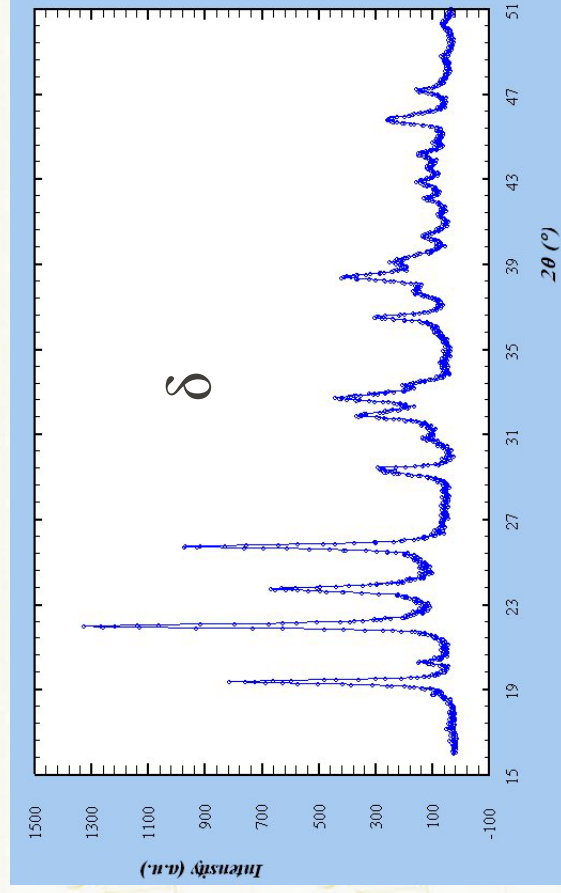
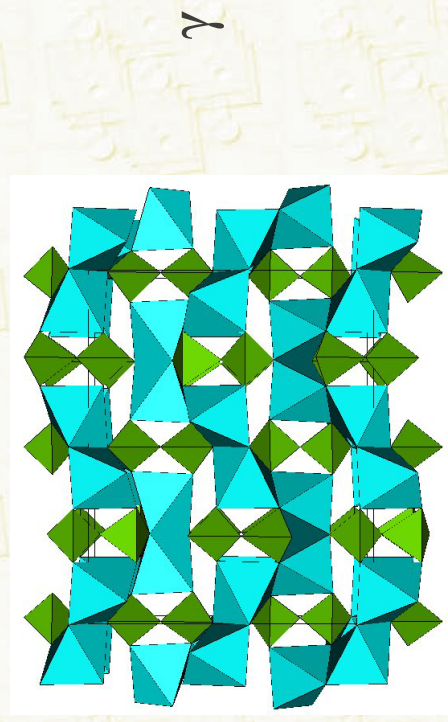
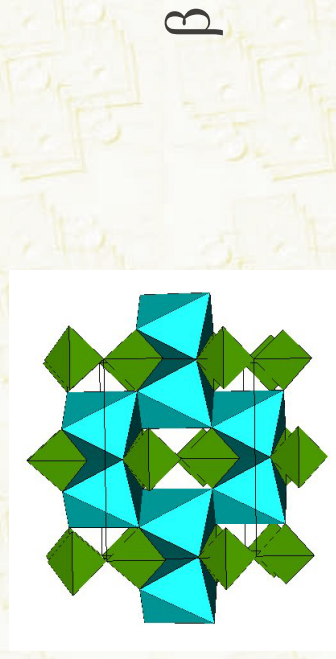
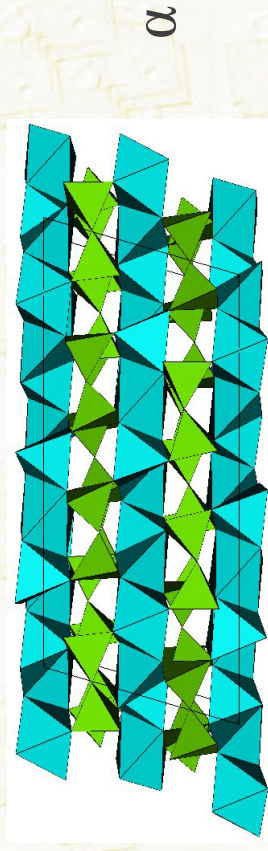
La raison est que de nombreux cas ne sont pas résolus du fait de leur complexité ou de leur mauvaise cristallinité empêchant une indexation.

Les nouvelles méthodes chassent les anciennes, avec des bases de données de structures prédites, une détermination de structure se réduira à une identification suivie d'un affinement (ce n'est pas pour demain...;-)

Cas typique dont la solution pourrait venir par prédiction

δ -Zn₂P₂O₇ Bataille *et al.*, *J. Solid State Chem.* 140 (1998) 62-70.

Indexation incertaine, profils de raies élargis par effet de taille de grains et/ou de défauts (pas mieux en synchrotron qu'en RX conventionnels...)



**Les propriétés physiques des composés
prédits peuvent elles-mêmes être prédites,
orientant les efforts de synthèse des chimistes
vers des cibles précises.**

**Ils ne vont pas aimer du tout
qu'on leur dise quoi chercher...**

D- Grands vainqueurs récents au hit parade des déterminations de structures les plus complexes :

« **Giant structure solved by
combined targeted chemistry and computational design.** »

2 structures cubiques publiées :

- 2004 - $V \sim 380.000 \text{ \AA}^3$, $a \sim 73 \text{ \AA}$, Fd-3m, 68 atomes indpdt (non-H)
- 2005 - $V \sim 702.000 \text{ \AA}^3$, $a \sim 89 \text{ \AA}$, Fd-3m, 74 atomes indpdt (non-H)

Présentées ± comme prédites – mais diagramme de poudre indexé.

Giant but not so big ?

A Hybrid Solid with Giant Pores Prepared by a Combination of Targeted Chemistry, Simulation, and Powder Diffraction

G rard F rey, Christian Serre, Caroline Mellot-Draznieks, Franck Millange, Suzy Surble, Julien Dutour, and Ir ne Margiolaki*

in the absence of single crystals. We shall prove in the following that, in the domain of MOFs, the introduction of computer simulation beside targeted chemistry of the inorganic moiety push forward these limits. It leads to a new crystalline MOF (**MIL-100**) with a giant cubic cell, the volume of which is close to $380\,000 \text{ \AA}^3$. It exhibits a hierarchy of micro- ($\varnothing \approx 6.5 \text{ \AA}$) and mesopores ($\varnothing \approx 25\text{--}30 \text{ \AA}$) and a large sorption capacity ($3100 \text{ m}^2 \text{ g}^{-1}$).

Angew. Chem. Int. Ed. 43 (2004) 2-7.

Super-tétraèdres assemblés en partage de sommets donnant des super-zéolithes (analogue MTN)

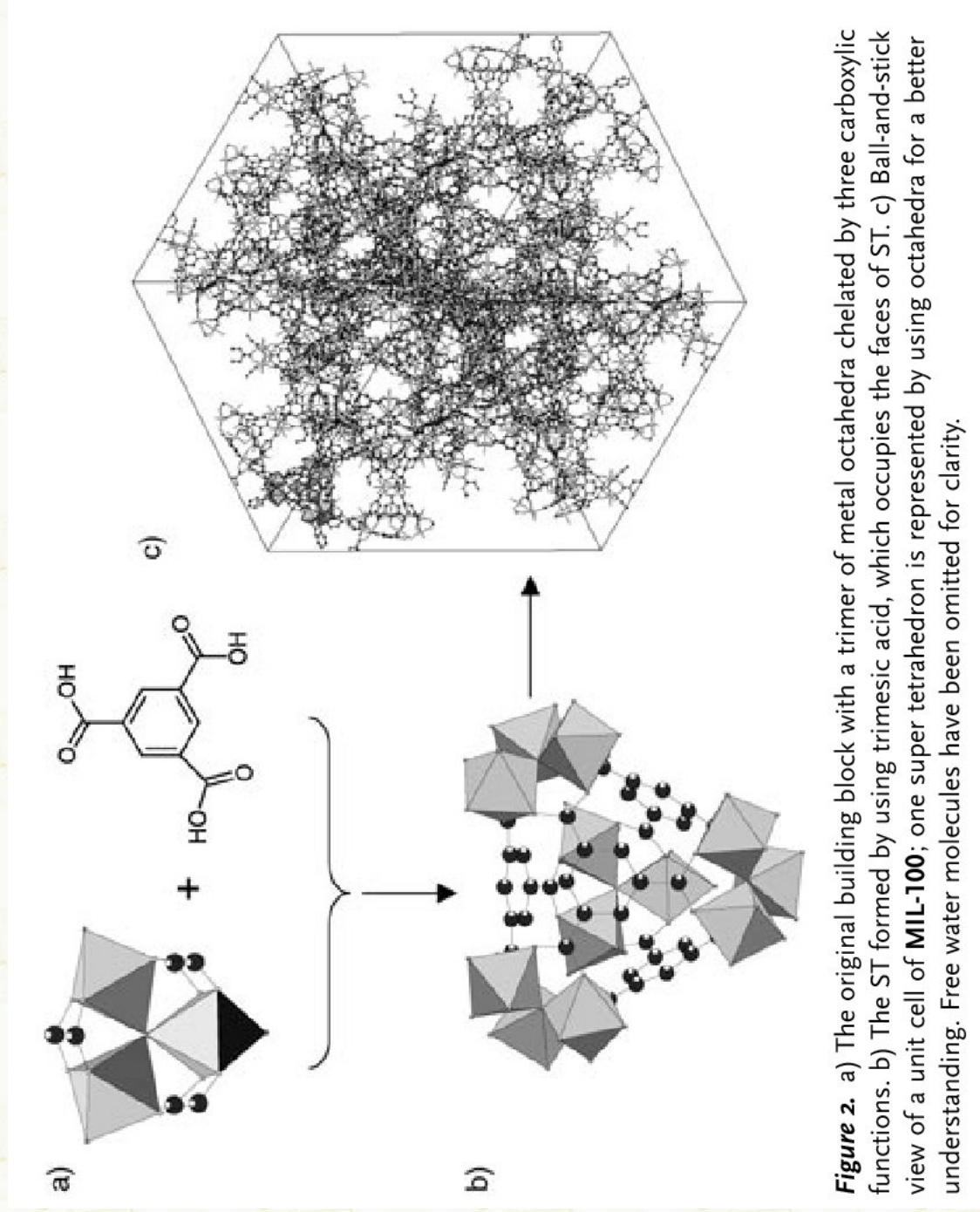


Figure 2. a) The original building block with a trimer of metal octahedra chelated by three carboxylic functions. b) The ST formed by using trimesic acid, which occupies the faces of ST. c) Ball-and-stick view of a unit cell of **MIL-100**; one super tetrahedron is represented by using octahedra for a better understanding. Free water molecules have been omitted for clarity.

A Chromium Terephthalate–Based Solid with Unusually Large Pore Volumes and Surface Area

G. Férey,^{1,2*} C. Mellot-Draznieks,³ C. Serre,¹ F. Millange,¹
J. Dutour,¹ S. Surblé,¹ I. Margiolaki⁴

We combined targeted chemistry and computational design to create a crystal structure for porous chromium terephthalate, MIL-101, with very large pore sizes and surface area. Its zeotype cubic structure has a giant cell volume ($\sim 702,000$ cubic angstroms), a hierarchy of extra-large pore sizes (~ 30 to 34 angstroms), and a Langmuir surface area for N_2 of $\sim 5900 \pm 300$ square meters per gram. Beside the usual properties of porous compounds, this solid has potential as a nanomold for monodisperse nanomaterials, as illustrated here by the incorporation of Keggin polyanions within the cages.

Science 309 (2005) 2040-2042.

Processus de prédiction identique, également analogue zéolithe MTN

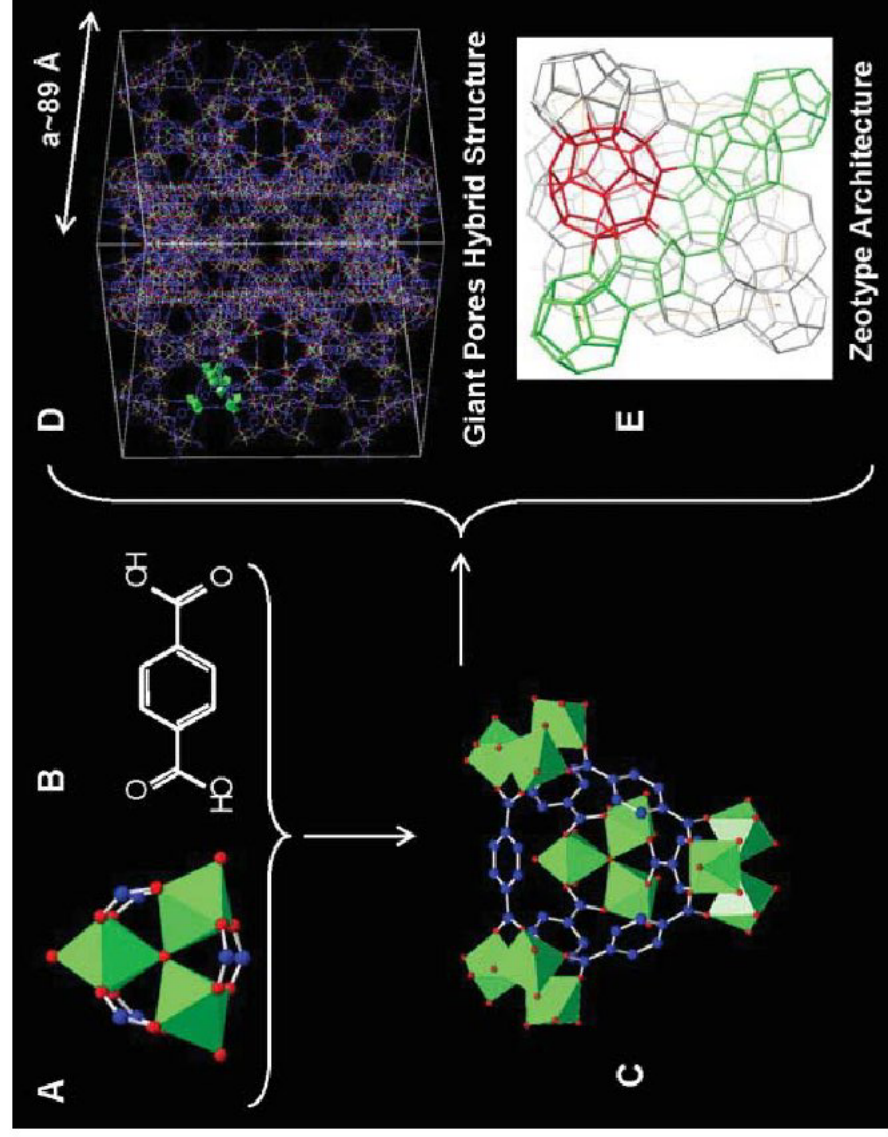


Fig. 1. (A) The computationally designed trimeric building block chelated by three carboxylic functions. The ST was constructed with (B) terephthalic acid, which lies (C) on the edges of the ST. (D) Ball-and-stick representation of one unit cell, highlighting one ST drawn in a polyhedron mode. (E) Schematic 3D representation of the MTN zeolite architecture (the vertices represent the centers of each ST) with the medium (in green, with 20 tetrahedra) and large (in red with 28 tetrahedra) cages delimited by the vertex sharing of the ST. Chromium octahedra, oxygen, fluorine and carbon atoms are in green, red, and blue, respectively.

Egaler ou surpasser ces « exploits » ?

OUI

Si la ou les molécules, la maille et le groupe d'espace sont connus, les méthodes dans l'espace direct ne nécessitent qu'environ 50 intensités par objet à localiser dans la maille par rotation et translation (6 degrés de liberté par objet), et ce, quel que soit le volume de la maille et sa symétrie.

PEUT-ETRE

Par prédiction partielle (sans maille, avec contenu).

NON

Par prédiction totale (sans maille, sans contenu).

Protéines

« The first protein crystal structure determined from high-resolution X-ray powder diffraction data, a variant of T3R3 human insulin-zinc complex produced by grinding, »

R.B. Von Dreele, P.W. Stephens, G.D. Smith and R.H. Blessing,
Acta Cryst. D56, 1549-1553 (2000).

SG : R3,

$a = 81.2780 \text{ \AA}$, $c = 73.0389 \text{ \AA}$,

$V = 417860 \text{ \AA}^3$

1630 atomes indépendants, 4893 paramètres affinés,
diagramme de poudre de 4800 points,
7981 restreintes stéréochimiques.

Facteur chance : 100%
avec maille, avec contenu.

La « complexité » (au sens de la difficulté de résolution) n'est pas proportionnelle au volume de maille

Estimation des volumes maxima à « complexité »

constante en

1996, avant

l'expansion des

méthodes de

solution dans

l'espace direct.

$$\lambda \sim 1.5 \text{ \AA}$$

$V_{\max}(\text{\AA}^3)$	General position multiplicity	Lattice	System
500	2	P	Triclinic
1000	4	P	Monoclinic
2000	8	C	"
2000	8	P	Orthorhombic
4000	16	A,B,C,I	"
8000	32	F	"
.....	etc for tetragonal, hexagonal and trigonal.....		
12000	48	P	Cubic
24000	96	I	Cubic
48000	192	F	Cubic

0.12	0.04	0.01	Minimal FWHM in $^\circ(2\theta)$
1000	3000	12000	" Fobs " extractable
70	210	840	xyz refinable parameters
23	70	280	independent atoms

La « complexité » est une combinaison de deux facteurs : largeur de raies (FWHM) et nombre d'atomes indépendants (ou de molécules).

Gros problème au final

Si :

Impossible d'affiner sans restreintes et/ou contraintes lorsque le nombre de données est à peine supérieur au nombre de paramètres de position, ou même inférieur...

Alors :

Ceci entache forcément la crédibilité des résultats.

D- Progrès en automatisation du processus complet

Suites logicielles intégrées les plus complètes :

Académique

EXPO2005

Ne contient que TREOR-N
pour indexation

N'est pas très bon en

« direct space » (essayez

FOX) ni en Rietveld

(mieux vaut passer à

FULLPROF ou **GSAS**)

En expansion

MAUD

WINPLOTR

POWDERX

...

Commerciale

TOPAS

(Bruker)

Material Studio

(Accelrys)

**Et si l'automatisme ne fonctionne pas ?...
Embauchez un cristallographe bien formé...**

Les vraies raisons des progrès

Les principales dans le désordre:

Concepteurs d'algorithmes, de bases de données ;

Synchrotrons de 3^{ème} génération ;

Côté matériel informatique : Loi de Moore, jeux vidéo (!), expansion des mémoires et disques de stockage, fréquence des processeurs x40 depuis 1995, cependant cette fréquence plafonne...

Progrès à venir ?

Les processeurs multi-cœurs prennent le relais : calcul parallèle chez Monsieur Toulemonde.

Les algorithmes de type Monte Carlo s'y prêtent parfaitement bien (indexation, solution en espace direct, prédiction).

Conclusions

Ces « progrès » sont-ils accessibles à tous ? Non.

Mises en garde :

Innovation logicielle à ne pas laisser entièrement sous contrôle de compagnies à but lucratif. L'héritage des logiciels du monde académique est encore large, mais se rétrécit devant la production commerciale.

Ne pas non plus laisser les bases de données entièrement sous contrôle de centres internationaux qui se disent à but non lucratif mais détiennent un monopole de fait sur un produit d'accès payant.

Quelques conseils supplémentaires

Choisissez une approche adaptée à votre problème
(radiation, méthodes),
celle qui vous conduira au succès.

La prédiction totale est l'approche la plus difficile.

Si vous voulez publier dans Nature ou Science,
inventez une « nouvelle » méthode,
et/ou dépassez largement 1.000.000 Å³
(visez le “hyper-giant”)
et 1500 atomes indépendants,
(pas seulement 100).

;-)

SDPD - Round Robin – Blind Test

(Structure Determination by Powder Diffractometry)

1998 <http://sdpd.univ-lemans.fr/SDPDRR/>

2002 <http://www.cristal.org/sdpdrr2/>

2006 ? **Qui veut l'organiser ?**