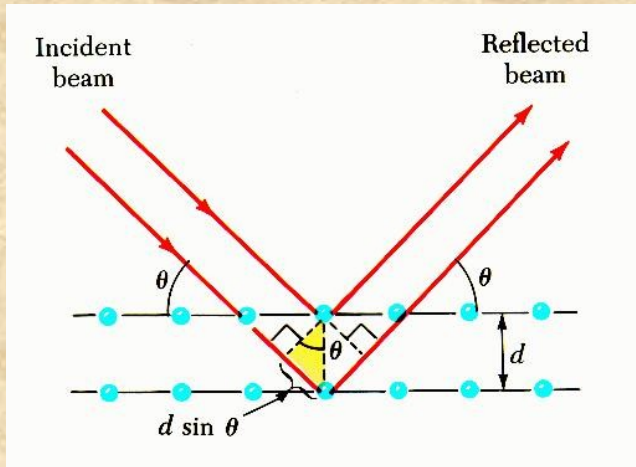


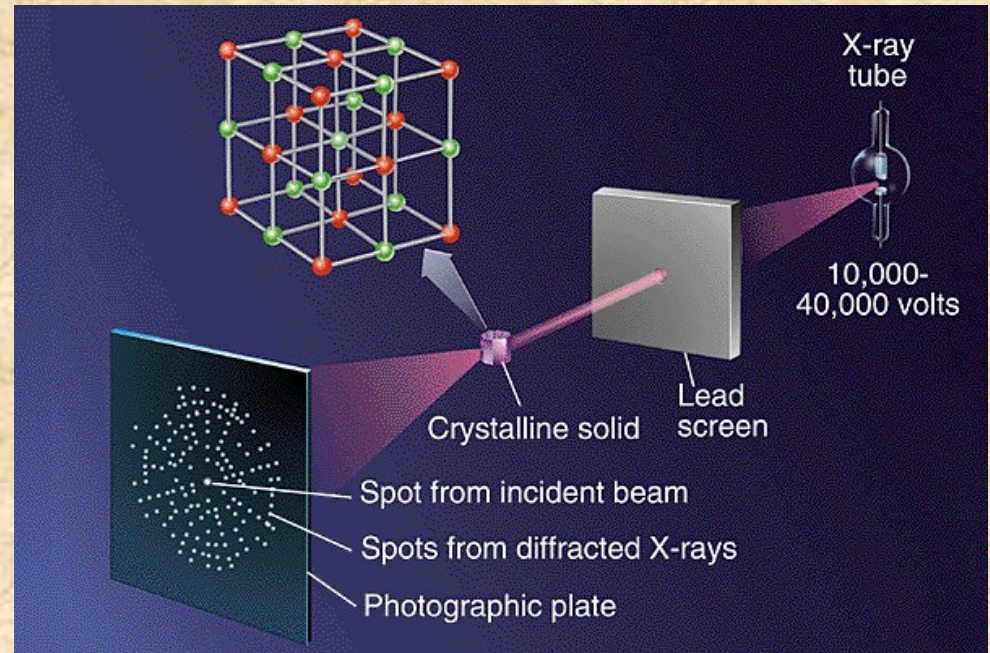
COMMENT UTILISER LES METHODES DIRECTES

Jean-Michel Leger
Stéphane Massip
EA4138-Pharmacochimie
Université de Bordeaux

RAPPELS



$$2d_{hkl} \sin \theta = k\lambda$$



$$I_{hkl} \approx |F_{hkl}|^2$$

$$F_{hkl} = v \int \int \int \rho(x,y,z) e^{i 2\pi(hx_i + ky_i + lz_i)} dx dy dz$$

$$F_{hkl} = \sum_i f_i e^{i 2\pi(hx_i + ky_i + lz_i)} = \sum_i f_i e^{i \phi} \quad f_i \text{ facteur de diffusion de l'atome } i$$

$$\rho(x,y,z) = 1/v \sum_h \sum_k \sum_l |F_{hkl}| e^{i \phi_{hkl}}$$

Les méthodes directes vont permettre de déterminer **directement** les phases ϕ

$$\rho(x,y,z)$$

Calculer les facteurs de structure pour une hypothèse structurale $\rho(x,y,z)$



F_c (calculs /positions atomiques) \longleftrightarrow F_o (mesures /intensités observées)

Facteur d'accord

$$R = \frac{\sum |F_c| - |F_o|}{\sum |F_o|}$$

3 Etapes :

Etude du groupe spatial

Résolution des phases pour une structure approchée

Affinement pour une structure finale \longleftrightarrow observations expérimentales

I - Etude préalable du mode de réseau et du groupe spatial

C'est l'étape capitale souvent source d'erreurs

Attention aux recherches automatiques des groupes spatiaux



échec dans la résolution structurale

Etude du système cristallin Ex: Monoclinique $\beta = 90,01^\circ$ (



Ambiguïté monoclinique / orthorhombique



Recherche des réflexions homologues

Monoclinique

$$I_{hkl} = I_{\bar{h}k\bar{l}} = I_{h\bar{k}l} = I_{\bar{h}kl}$$

Orthorhombique

$$I_{hkl} = I_{\bar{h}k\bar{l}} = I_{h\bar{k}l} = I_{\bar{h}kl} = I_{hk\bar{l}} = I_{\bar{h}kl} = I_{h\bar{k}\bar{l}} = I_{\bar{h}kl}$$

Calcul de la densité du cristal / nombre de motif Z par maille

$$d = \frac{M \times Z}{N \times V}$$



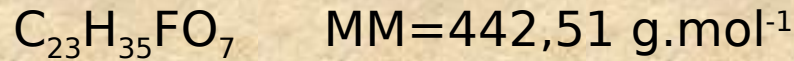
Z



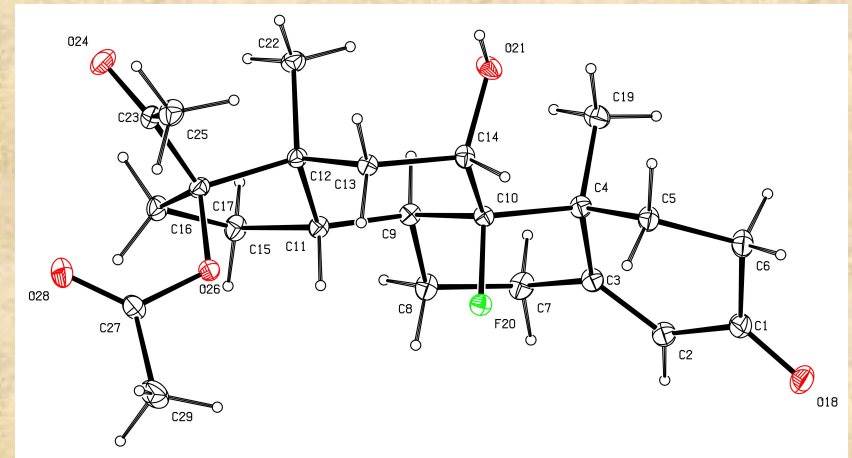
idée du groupe spatial (tables internationales)

Mode de réseau / Système	Primitif (P)	centré (I)	Bases centrées (C)	Faces centrées (F)
Cubique $a = b = c$ $\alpha = \beta = \gamma = 90^\circ$				
Rhomboédrique $a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$				
Hexagonal $a = b \neq c$ $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$				
Quadratique $a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$				
Orthorhombique $a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$				
Monoclinique $a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$				
Triclinique $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$				

Exemple de la Fluorogestone



Monoclinique
 $a = 9,743(3) \text{ \AA}$
 $b = 7,630(7) \text{ \AA}$
 $c = 15,913(3) \text{ \AA}$
 $\beta = 105,36(2)^\circ$



Volume = 1140,7 Å³

seule extinction systématique sur 0k0, k=2n (condition de présence)
 Rappel: la présence d'élément de symétrie glissant conduit à des extinctions systématiques (axes hélicoïdaux, plans de glissement)

2 groupes spatiaux possibles : P 2₁ et P 2₁/m

P 2₁ ↔ Z = 2

d = 1,288

P 2₁/m ↔ Z = 4 si molécule symétrique

1/2 molécule sur miroir ou centre

d = 2.576

Impossible car pas de symétrie dans la molécule
 pas de miroir (énantiomère 8 C*)

➡ P 2₁

Exemple de la Fluorogestone

Vérification du groupe spatial par le Patterson

↔ plan perpendiculaire à l'axe 2_1 ↔ Codes de symétrie : x, y, z et $-x, \frac{1}{2}+y, -z$

↔ Pics de Patterson en $X = 2x, Y = 0.5000$ et $Z = 2z$

↔ étude du plan de Harcker $h \ 0.5 \ l$

Instruction PATT dans name.ins de Shelx ou WinGX/solve/Shelxs-86/Patterson

	X	Y	Z	X	Y	Z		
0.0000	0.0000	0.0000	0.0000	999.	0.9235	0.1692	0.4339	52.
0.8791	0.0000	0.0316	0.0316	135.	0.4613	0.5000	0.0904	50.
0.1398	0.1634	0.1380	0.1380	118.	0.2503	0.1940	0.2040	49.
0.9048	0.5000	0.4397	0.4397	117.	0.1751	0.0000	0.2982	48.
0.1236	0.0000	0.0684	0.0684	99.	0.3069	0.5000	0.2608	47.
0.2602	0.0000	0.0477	0.0477	89.	0.1579	0.5000	0.4789	47.
0.6453	0.5000	0.1709	0.1709	86.	0.8853	0.3279	0.2692	46.
0.0034	0.1516	0.0509	0.0509	73.	0.5247	0.3654	0.2514	45.
0.8911	0.1719	0.0880	0.0880	69.	0.8764	0.3908	0.3878	42.
0.0105	0.0000	0.0793	0.0793	68.	0.0752	0.5000	0.0623	42.
0.2822	0.0422	0.2600	0.2600	68.	0.9753	0.4193	0.1185	42.
0.5658	0.5000	0.4422	0.4422	67.	0.6733	0.5000	0.1050	41.
0.1412	0.0000	0.1967	0.1967	65.	0.7850	0.1378	0.1456	41.
0.0221	0.1685	0.1660	0.1660	65.	0.5003	0.2008	0.1919	41.
0.9523	0.3423	0.4486	0.4486	62.	0.9823	0.5000	0.1736	40.
0.9055	0.0000	0.2671	0.2671	60.	0.7891	0.0000	0.2972	40.
0.0275	0.0000	0.2290	0.2290	60.	0.2132	0.3188	0.4749	39.
0.7553	0.3217	0.2949	0.2949	59.	0.3555	0.5000	0.1373	39.
0.2618	0.1632	0.1063	0.1063	58.	0.2913	0.5000	0.1813	39.
0.0000	0.3152	0.0000	0.0000	56.	0.1630	0.5000	0.1394	37.

Accumulation de pics en $Y = 0.5000$

Exemple de la Fluorogestone

Si ambiguïté : - méthodes directes qui donnent le meilleur résultat
- travailler avec symétrie plus basse (ex: triclinique P1)
↳ retrouver l'élément de symétrie manquant

Les méthodes directes = exploitation de relations entre phases
↳ densité électronique ρ positive dans la maille

★ méthodes statistiques (MULTAN, SIR, SAPI...)

Relation de Sayre, Cochran et Zachariansen (1952)

$$\Phi(H) = \Phi(H') + \Phi(H-H') \quad \text{avec } \Phi(H) = \text{phase de } F_H$$

H = ensemble des valeurs de hkl

Relation de Karle et Hauptmann (1952)

Relation de Karle et Karle (formule de la tangente, 1966)...

Dans un processus de convergence à partir d'un petit nombre de phases connues, on peut accéder à la connaissance des autres phases.

Exemple de la Fluorogestone avec SIR2004 (fichier sir.out)

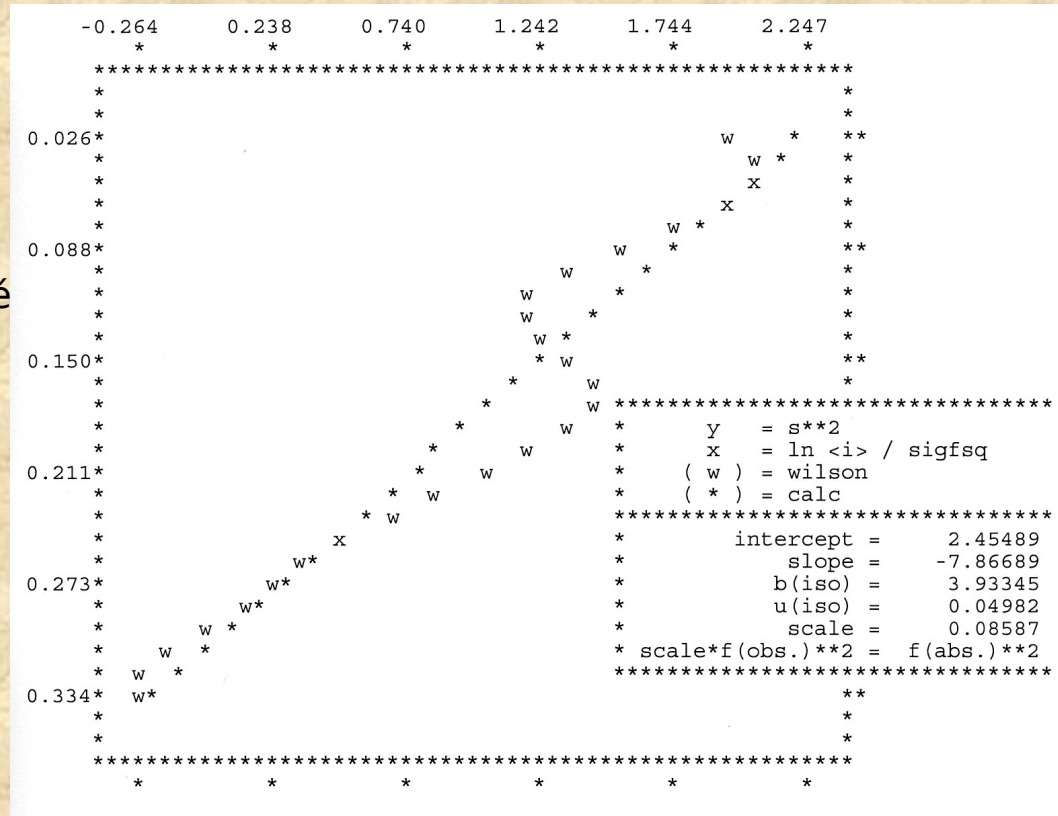
On définit un facteur de structure normalisé $E_H = \frac{|F_H|}{\sqrt{\langle |F_H|^2 \rangle}}$ qui va être utilisé dans les méthodes directes

Average	Numeric				Graphic
	all data	acentric	centric	hypercentric	a. c. h.
mod(E)	0.865	0.886	0.798	0.718	*
E**2	1.000	1.000	1.000	1.000	*
E**3	1.414	1.329	1.596	1.916	*
E**4	2.330	2.000	3.000	4.500	*
E**5	4.312	3.323	6.383	12.260	*
E**6	8.727	6.000	15.000	37.500	*
mod(E**2-1)	0.804	0.736	0.968	1.145	*
(E**2-1)**2	1.330	1.000	2.000	3.500	*
(E**2-1)**3	3.738	2.000	8.000	26.000	*
(mod(E**2-1))**3	4.205	2.415	8.691	26.903	*

mod(E**2-1) est un bon test de centricité

Extinction Group	Fig.Mer
P 1 21 1	0.491
P 1 _ 1	0.319
P 1 21/c 1	0.055
P 1 21/a 1	0.055
P 1 21/n 1	0.055
P 1 a 1	0.035
P 1 n 1	0.035
P 1 c 1	0.035
C 1 _ 1	0.006
C 1 c 1	0.006
A 1 _ 1	0.006
A 1 n 1	0.006
I 1 _ 1	0.006
I 1 a 1	0.006

Evaluation automatique du groupe spatial



Statistique de Wilson

Exemple de la Fluorogestone avec SIR2004 (fichier sir.out)

Le programme sélectionne les E les plus forts (environ 10 x Nombre d'atomes)

“boîte noire”

invariants de structure (triplets, quartets...)

↔ combinaisons de 3, 4 ...phases
indépendants du choix de l'origine

```
290 largest E-values > 1.433 to phase. ( 10 printed )
```

code	h	k	l	E	fo/sig.	code	h	k	l	E	fo/sig.
1	5	0	-6	3.103	>6	2	6	3	-9	2.948	>6
3	1	3	11	2.826	>6	4	8	1	-1	2.796	>6
5	1	0	-16	2.741	>6	6	3	0	-15	2.720	>6
7	1	0	14	2.710	>6	8	5	0	-13	2.676	>6
9	4	0	3	2.674	>6	10	8	0	-1	2.658	>6

```
242 smallest E-values for psi0 and negative quartets ( 10 printed )
```

code	h	k	l	E	fo/sig.	code	h	k	l	E	fo/sig.
2091	3	6	-9	0.000	0.00	2092	10	1	2	0.000	0.00
2093	9	0	-13	0.000	0.00	2094	4	0	8	0.000	0.00
2095	3	0	-5	0.000	0.00	2096	3	8	-5	0.000	0.00
2097	0	7	4	0.000	0.00	2098	6	6	2	0.000	0.00
2099	9	0	-14	0.000	0.00	2100	5	0	-11	0.000	0.00

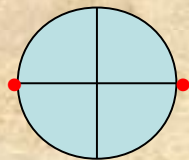
Le programme définit au plus 3 phases pour fixer l'origine de la maille
+ une phase supplémentaire en non-centrosymétrique pour fixer l'énantiomère

⇒ Grand nombre de relations entre phases

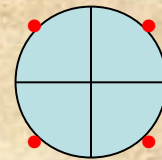
A partir des 1ères phases connues ($\Sigma 1$ avec une probabilité $> 0,9$) et les phases (pour l'origine et l'énantiomère), le processus de génération de nouvelles phases commence.

Dès qu'une phase est inconnue, le programme génère 2 à 4 valeurs possibles s
Le cadran trigonométrique.

structures centrées :



structures non-centrées :



Exemple de la Fluorogestone avec SIR2004

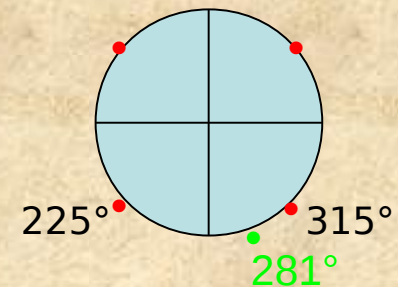
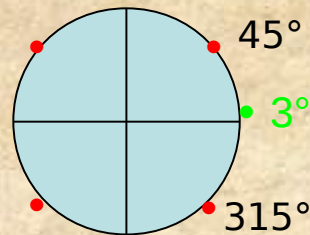
⇒ Plusieurs sets de solutions ⇒ Structure avec CFOM équivalentes

Au cours du processus, 2 phases inconnues sont utilisées.

⇒ Générations de toutes les phases (4x4 solutions).

Pour Φ_1 , 4 solutions sont utilisées, 2 convergent (45° et 315°) avec la formule de la tangente : on trouve Φ_1 à 3° .

De même pour Φ_2 à 281° .



4 solutions convergent et donnent accès à un squelette de structure acceptable

Exemple de la Fluorogestone avec SIR2004 (fichier sir.out)

Toutes les solutions sont générées.

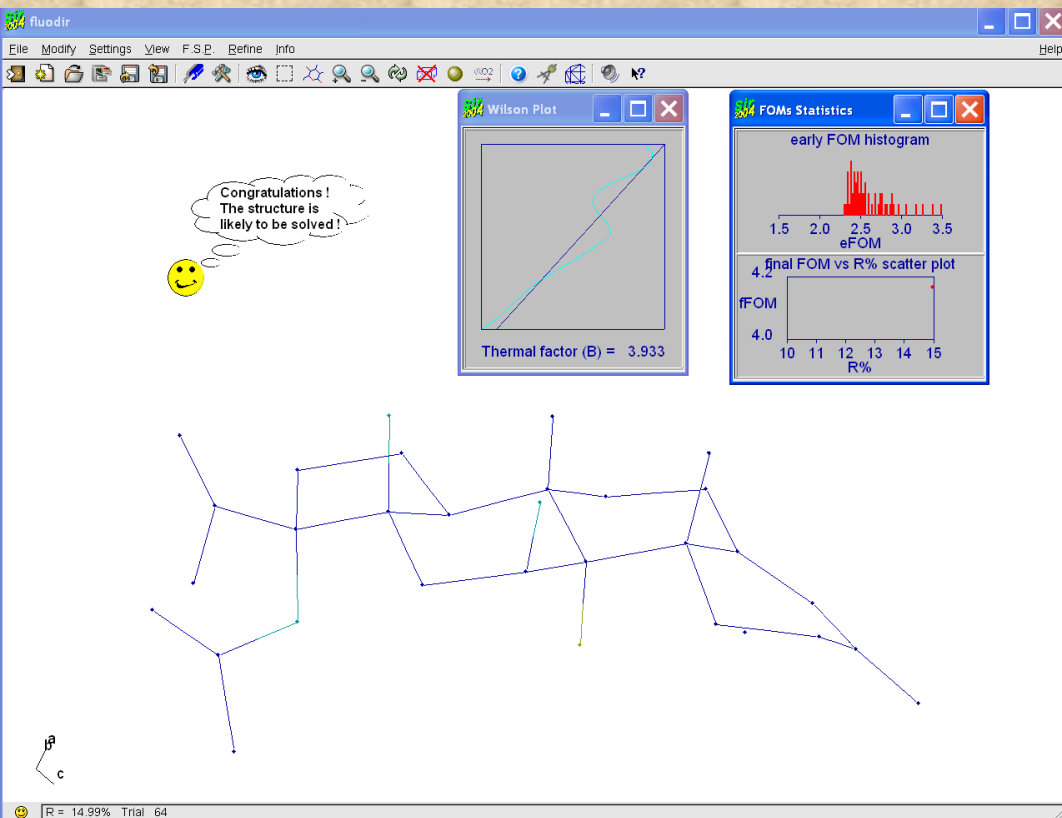
Le test eFom permet de sélectionner
La meilleure solution → eFom le plus élevé

Set	(Trial)	eFom	undetermined	Set	(Trial)	eFom	undetermined
1	(55)	3.499	54	2	(89)	3.375	61
3	(5)	3.269	68	4	(75)	3.177	74
5	(61)	3.061	71	6	(25)	2.953	75
7	(93)	2.901	81	8	(91)	2.883	69
9	(21)	2.870	74	10	(50)	2.867	71

Extending phases for Block # 1

Set	Trial	eFom	UnDet
1	55	3.499	19%
	55	End of DSR	Final FOM = 4.164

R% = 38.85 30.56 23.45 19.07 18.02 18.84 14.96 14.99 14.99 15.02



Résultat avec SIR
R ≈ 15%

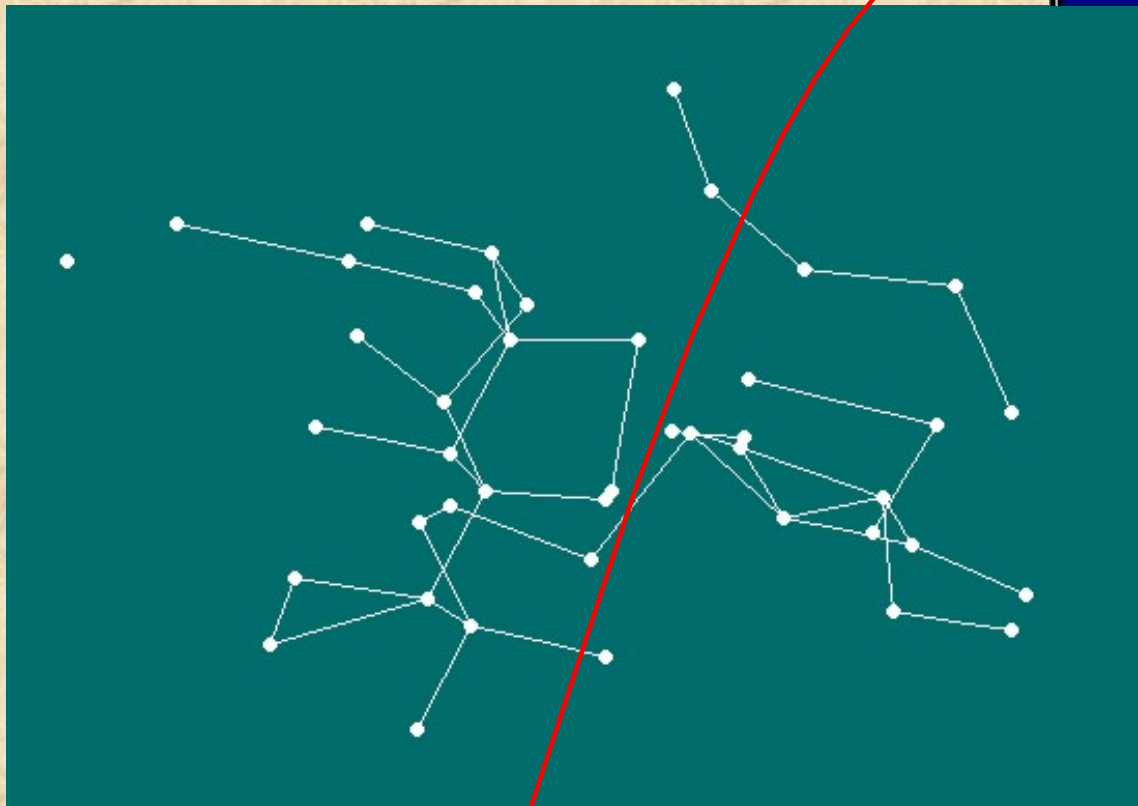


Exemple de la Fluorogestone avec SHELXS

```

Fichier  Edition  Recherche  Affichage  Options  ?
C:\shelx97\fluodir.ins
TITL FLU0
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X,Y+0.500,-Z
SFAC C H F O
UNIT 46 70 2 14
TREF 50
HKLF 4
END

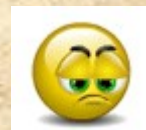
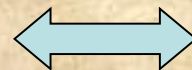
```



323 phases utilisées



UNIT



Phases mal déterminées
Nbre d'essais insuffisants

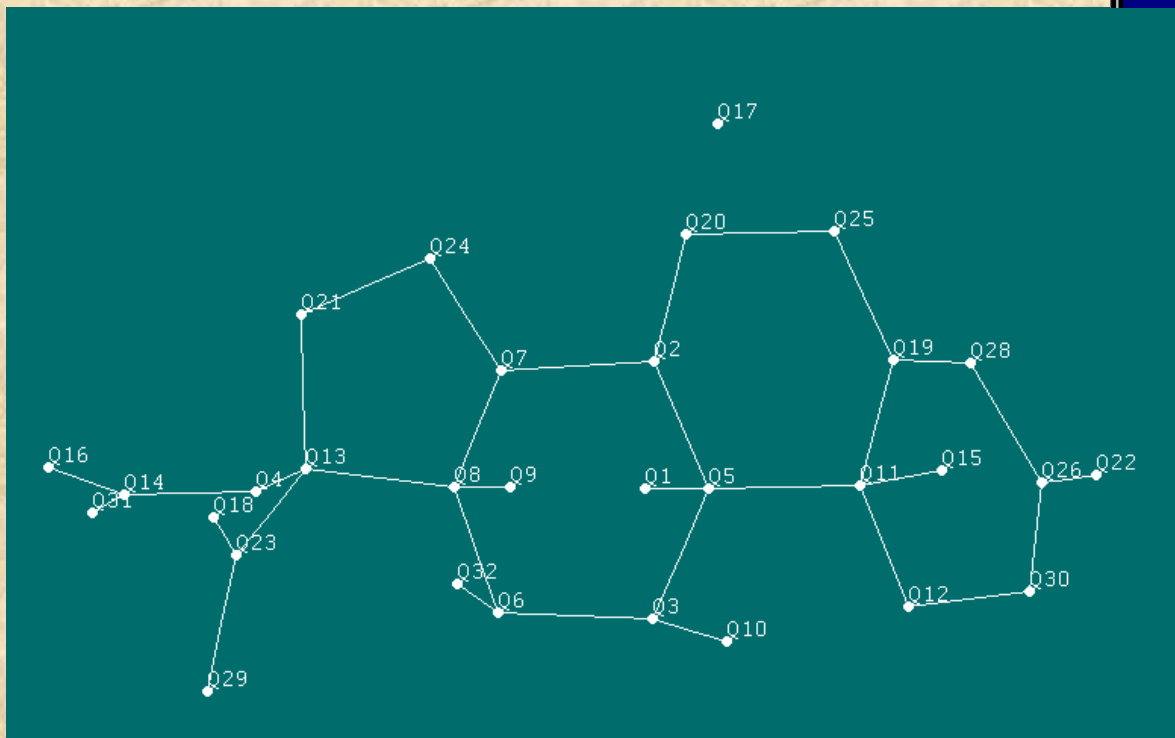
Shelxs, nbre d'essais faible, CFOM élevé (0.272)



≈ 10xNbre d'atomes

Exemple de la Fluorogestone avec SHELXS

```
Fichier Edition Recherche Affichage Options ?
C:\shelx97\fluodir.ins
TITL FLUO
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X,Y+0.500,-Z
SFAC C H F O
UNIT 46 70 2 14
TREF 290
HKLF 4
END
```



323 phases utilisées

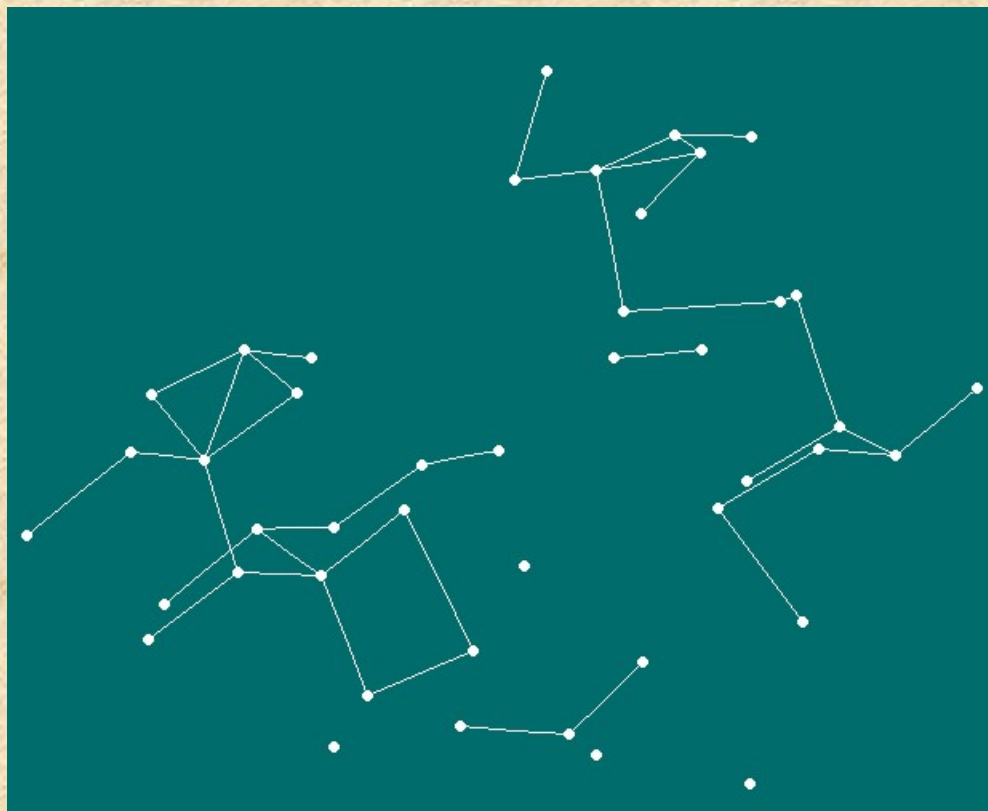


Shelxs, nbre d'essais = 290, CFOM = 0.061

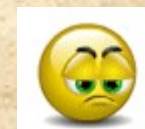
Exemple de la Fluorogestone avec SHELXS

On peut choisir le nbre de phases à déterminer

```
Fichier Edition Recherche Affichage Options ?
C:\shelx97\fluodir.ins
TITL FLUO
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X,Y+0.500,-Z
SFAC C H F O
UNIT 46 70 2 14
ESEL 2.0
TREF 290
HKLF 4
END
```



53 phases utilisées



Malgré une figure de mérite correcte

Shelxs, Nbre essais 290, Emin = 2.0, CFOM=0.030
Nbre de phases déterminées = 53

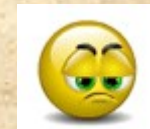
Exemple de la Fluorogestone avec SHELXS

On peut choisir le nbre de phases à déterminer (limites min et max)

```
Fichier Edition Recherche Affichage Options ?
C:\shelx97\fluodir.ins
TITL FLUO
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X,Y+0.500,-Z
SFAC C H F O
UNIT 40 70 2 14
ESEL 1.2 1.4
TREF 290
HKLF 4
END
```



238 phases utilisées



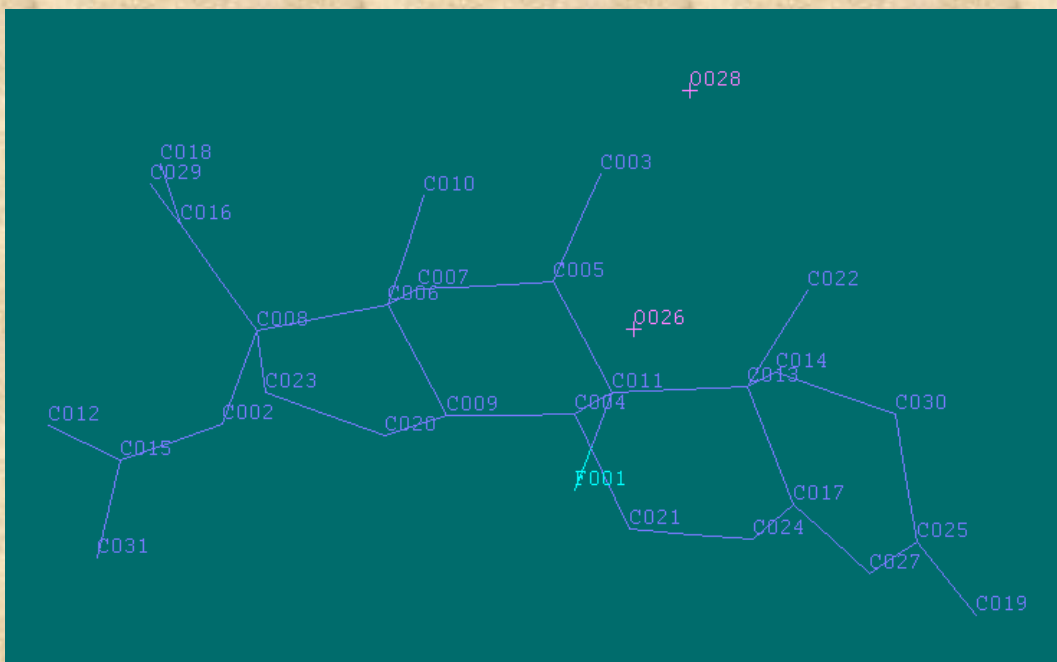
Malgré un nombre correct de phases (les bonnes?)

Shelxs, Nbre essais 290, Emin = 1.2, Emax = 1.4
CFOM=0.157, Nbre de phases déterminées = 238

Exemple de la Fluorogestone avec SHELXD

```
Fichier Edition Recherche Affichage Options ?
C:\shelx97\fluodir.ins
TITL FLUO
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X, Y+0.500, -Z
SFAC C H F O
UNIT 46 70 2 14
NTRY 10
FIND 40
PLOP 35 45
HKL 4
END
```

```
Try 5:20/t Peaks 99 83 75 72 55 54 53 47 46 45 40 38 37 35 35 34 34 33 33
R = 0.268, Min.fun. = 0.301, <cos> = 0.555, Ra = 0.194
Try 5:40 Peaks 99 75 70 69 68 67 67 66 66 64 62 59 59 56 56 55 52 50 47 46
R = 0.215, Min.fun. = 0.298, <cos> = 0.569, Ra = 0.165
Try 5, CC All/Weak 76.70 / 63.73, best 77.93 / 66.41, best final CC 88.24
```



Nbre > FIND (faux pics)
Nbre d'atomes recherchés
Nbre de solutions testées



R correct
CC = coefficient de corrélation (>65%)

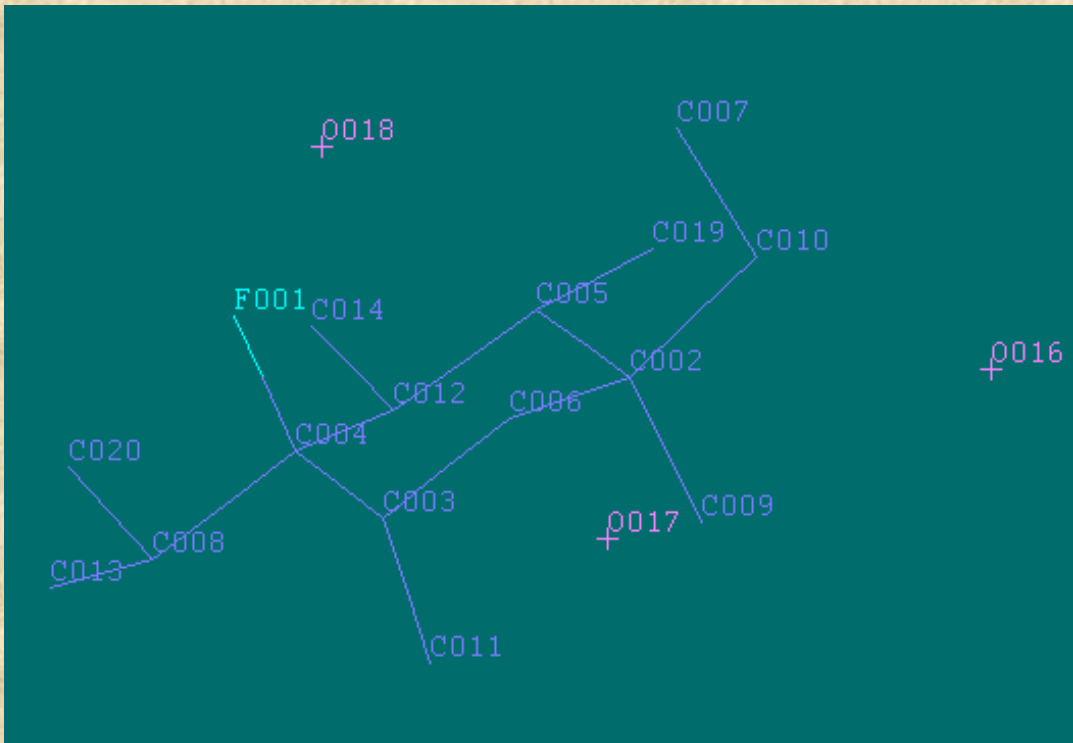
Shelxd, NTRY 10, FIND 40, R=0.21, CC=88.2%

Utilisé en général pour les moyennes et grandes molécules

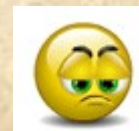
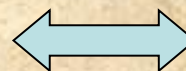
Exemple de la Fluorogestone avec SHELXD

```
Fichier Edition Recherche Affichage Options ?
C:\shelx97\fluodir.ins
TITL FLUO
CELL 1.54180 9.743 7.630 15.913 90.000 105.360 90.000
ZERR 2 0.0030 0.0070 0.0030 0.0000 0.0200 0.0000
LATT -1
SYMM -X,Y+0.500,-Z
SFAC C H F O
UNIT 46 70 2 14
NTRY 10
FIND 10
PLOP 5 15
HKLF 4
END
```

```
Try 5:20 Peaks 99 69 68 59 56 54 50 49 49 48 48 44 44 39
R = 0.377, Min.fun. = 0.430, <cos> = 0.435, Ra = 0.361
Try 5, CC All/Weak 32.68 / 15.99, best 47.92 / 32.23, best final CC 54.01
```



Importance instruction PLOP



Attention aux problèmes de Molécules indépendantes/solvants

Shelxd, NTRY 10, FIND 10, R=0.37, CC=53.7%

Exemple de la Fluorogestone avec SIR/SHELXS/SHELXD

Il existe une infinité de programmes et de logiciels « boîte noire » qui permettent d'exploiter les données de diffraction mais tous sont basés

sur des paramètres nécessaires:


Le nombre d'atomes dans le motif et/ou maille

La nature de ces atomes (lourds?)

Le mode de réseau (P, C, I ou F pour Shelx : $LATT > 0$ ou $LATT < 0$)

Les codes de symétrie (pour Shelx)


centré


non centré

Le groupe spatial (pour SIR)

Maille compatible avec les hkl (pas de $\cos > 1$)

Instruction TREF (Nbre d'essais dans Shelxs)

Instruction PLOP > Nbre d'atomes (Shelxd)

Les suites logicielles

Les diffractomètres automatiques des différents constructeurs sont livrés avec des logiciels complets allant de la collecte à la résolution de la structure

et utilisant des programmes

Exemple de logiciel :
CrystalClear et
CrystalStructure
(Rigaku-MSK)

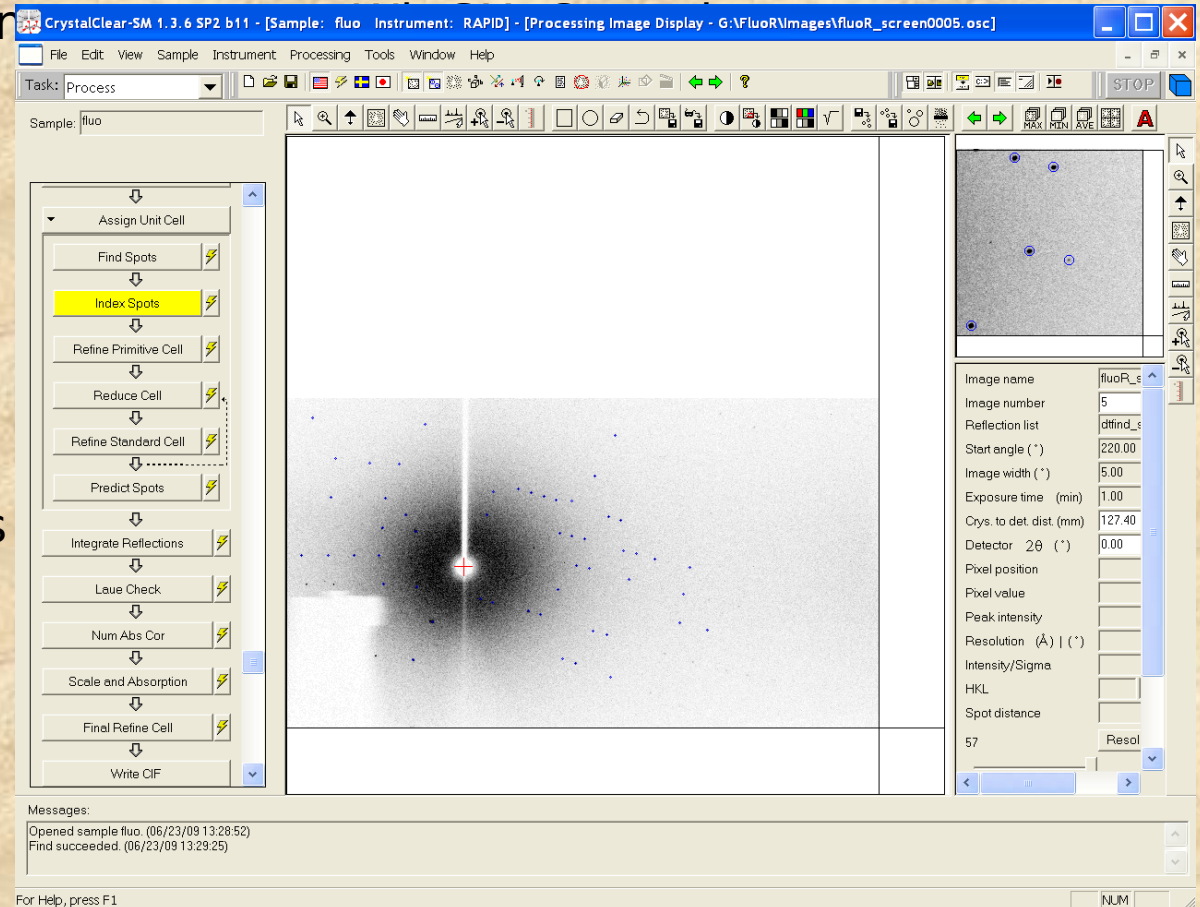
CrystalClear:
Recherche de maille
Intégration des réflexions
Créations des fichiers



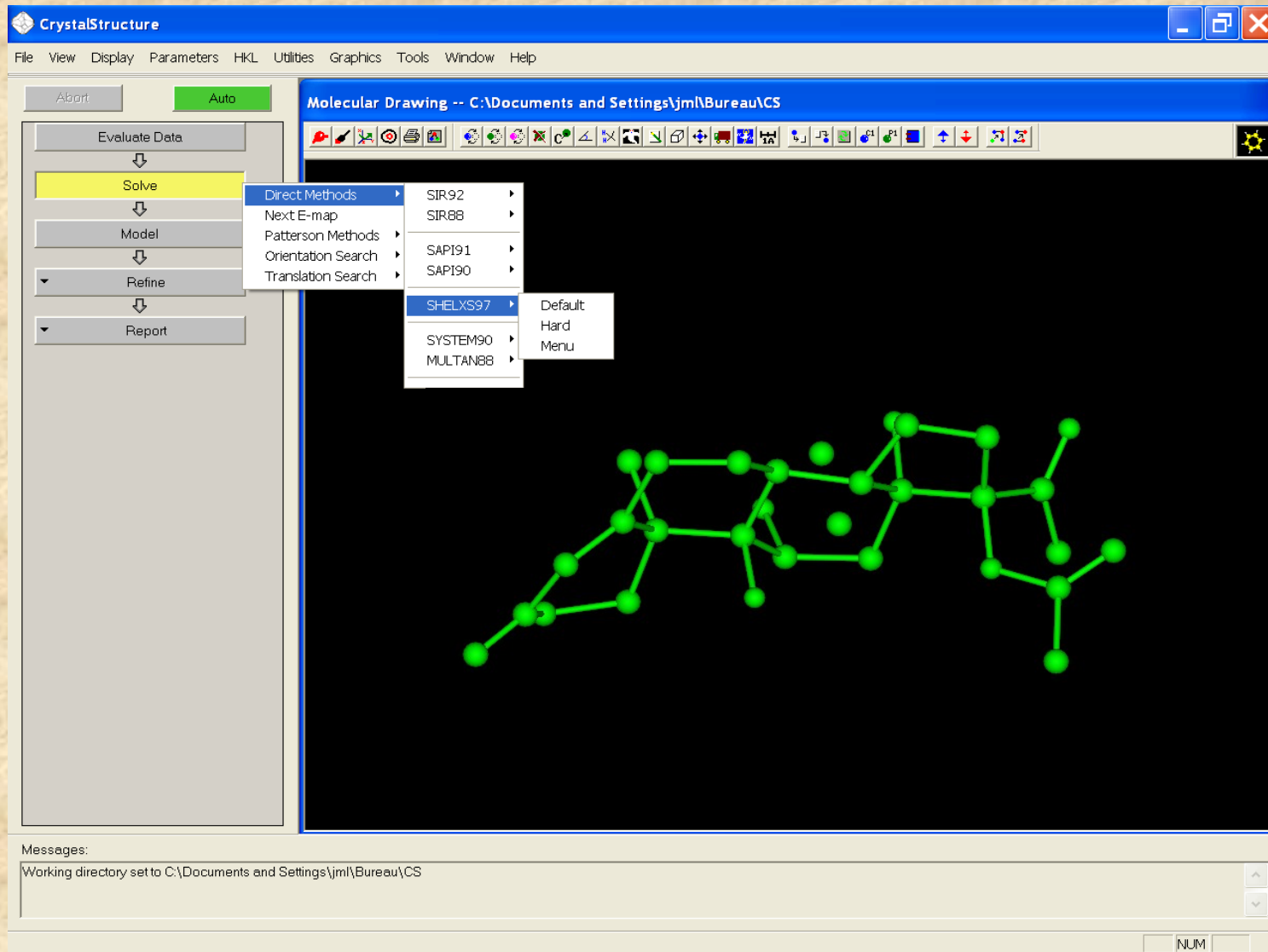
crystalclear.cif

ou texray.inp
f2plus.dat

hkl fo² au format « Rigaku » → Format shelx possible



CrystalStructure: recherche de groupe spatial, utilisation des différentes programmes de résolution de structures, affinement...



→ Format shelx .res .ins et .hkl

CONCLUSION

Dans plus de 90% des cas, les méthodes directes automatisées conduisent à l'obtention de la structure... cependant, il faut rester vigilant car chaque structure est différente.

- Bonne qualité de mesure pour une collecte complète (si possible jusqu'à 0,8Å),
- Vérifier le groupe spatial / imposer un autre groupe / lecture critique des F_H ,
- Vérifier la statistique de Wilson (centré/non centré, jouer sur le B moyen...)
- descendre en symétrie ($P-1 \rightarrow P1, P2_1/c$ ou $P2_1/m \rightarrow P2_1, Pbc_a \rightarrow P2_12_12_1...$) souvent efficace,
- En cas d'échec,
- rechercher la position du ou des éléments de symétrie,
- effectuer les translations adéquates (MOVE dans Shelxl...),
- utiliser PATSEE si on connaît la géométrie d'un fragment de la molécule (>50%) en combinant le Patterson et les méthodes directes.