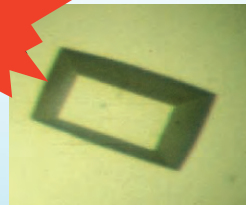


La cristallographie des macromolécules biologiques PX



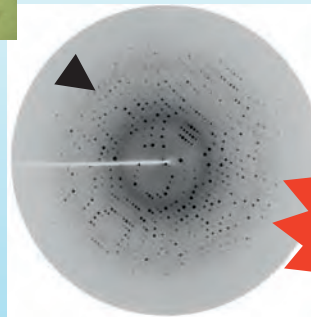
Protéine
purifiée

Cristallisation

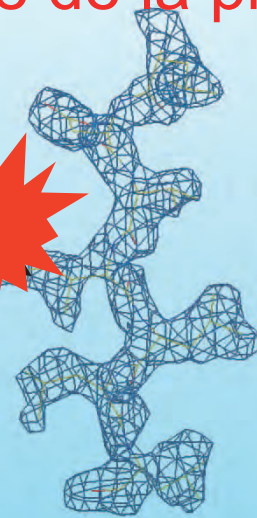


Cristal

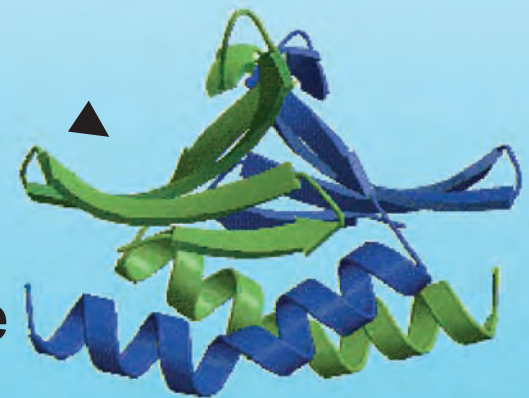
Problème de la phase



Diffraction
Rayons X



Densité
électronique



Structure 3D

Interprétation biologique ◀

La Chimie dans la Biologie

TABLEAU PÉRIODIQUE DES ÉLÉMENTS

Numéro atomique → Z

Symbole atomique → X

Masse molaire atomique (g.mol⁻¹) → M

Famille

Nom

Science.gouv.fr
http://www.science.gouv.fr

Format PDB

ATOM

ATOM	2610	CZ	ARG	A	328	60.918	8.805	63.367	1.00117.21	C
ATOM	2611	NH1	ARG	A	328	59.715	8.787	62.813	1.00117.59	N
ATOM	2612	NH2	ARG	A	328	61.968	9.144	62.636	1.00101.65	N
ATOM	2613	OXT	ARG	A	328	66.626	8.061	66.512	1.00	O
TER	2614	ARG	A	328						
HETATM	2615	CS	CS	900		74.649	19.612	41.156	0.30	CS
HETATM	2616	N6	AQP	400		51.146	29.022	31.838	0.85	N

HETATM

Zn dans 6676 structures,

Mais aussi des Terres rares,
des éléments toxiques *As*, *Hg*, etc.
des sels SO_4^{2-} (8,447), des sucres,
des polymères e.g. *MéthylPentaneDiol*

des éléments constitutifs ou adduits,
ligands, inhibiteurs, ...

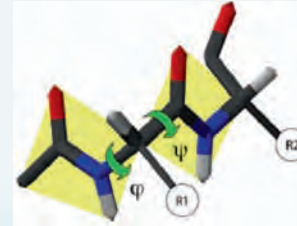
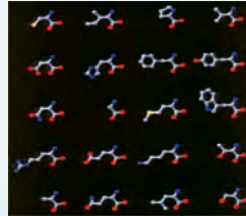
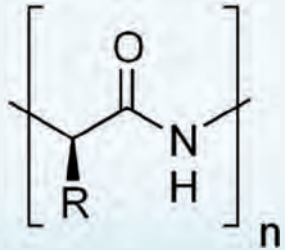
Macromolécules biologiques :

Protéines, acides nucléiques, virus, sucres, lipides, ...
des polymères d'unités a priori limitées en nombre

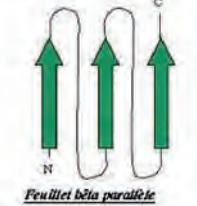
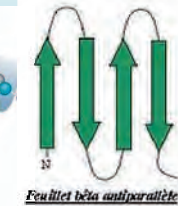
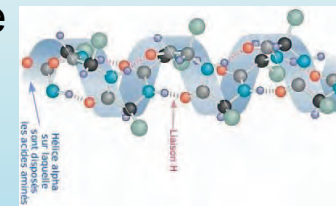
Protéine typique : env. 300 amino acides

Exemple des Protéines et de leur structuration

Chaîne composée de 20 acides aminés naturels L (séquence, **str. primaire**)



formant des éléments structuraux **secondaires** (algorithme DSSP) (hélice alpha / feuillet bêta) **prédictibles** (liaisons hydrogène) / dichroïsme circulaire



se repliant de façon **encore non prédictible** (création de motifs)



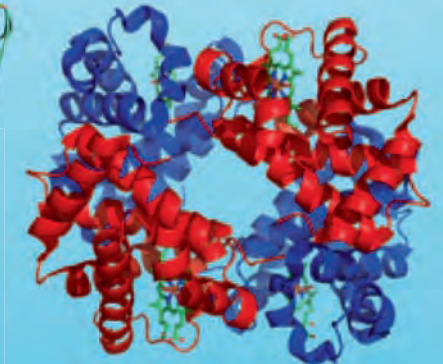
la structure tertiaire ou 3D

Ex : Myoglobine, La 1^{ère} structure résolue par cristallographie en 1958.
par Max Perutz et John Kendrew. (prix Nobel de chimie en 1962).



avec une fonction associée qui peut dépendre de l'assemblage
(**structure quaternaire**)

Ex : l'hémoglobine humaine.



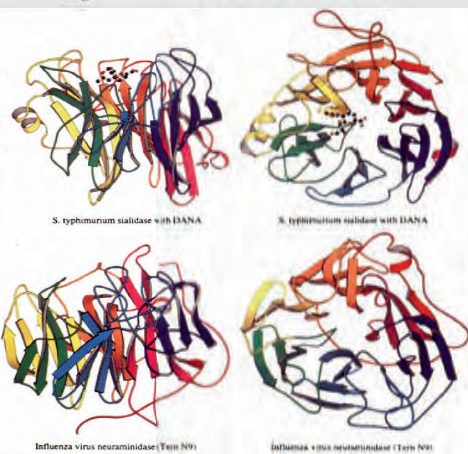
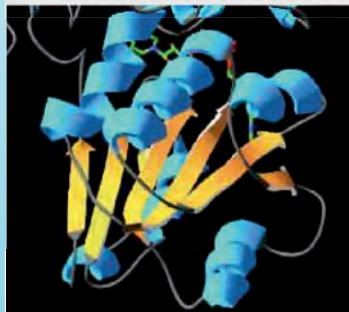
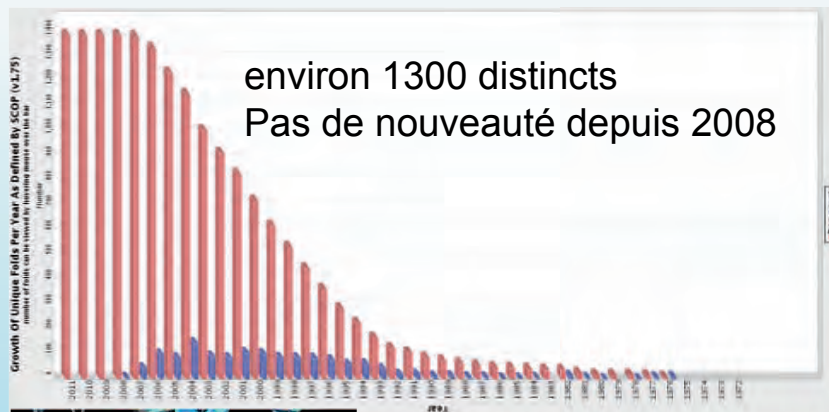
Biologie moléculaire vs chimie

Sources les plus communes :

Homo sapiens (21,059 PDB) et *E. coli*, (8,447 PDB)

Génome eucaryote environ 20000 gènes,
100000 expressions différentes ?

Classification du repliement des protéines



Exemples de repliement :

(g.) Rossman fold

Limites de la « protein engineering » :

20 aa naturels,

protéines possibles $20^N \gg$ sél. naturelle
modifs. post-translationnelles

→ *Ere de la biologie synthétique
ou synthèse chimique totale des protéines*

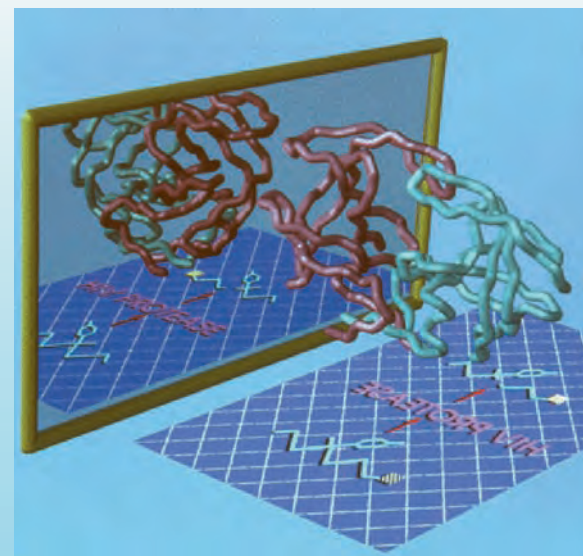
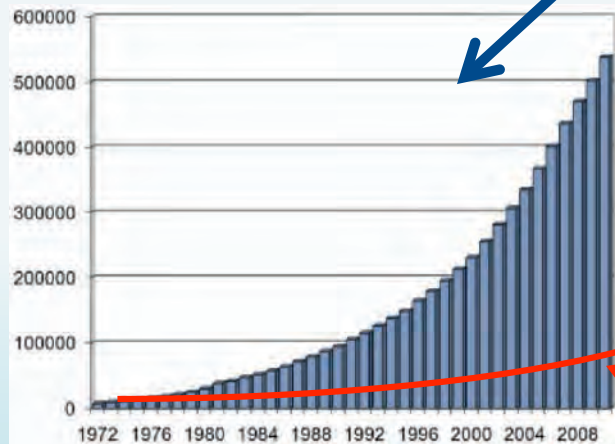


Image miroir de la protéase HIV -1

Enantiomérique D-Protéine synthétisée chimiquement
Racémisation comme aide à la cristallisation
ou pour le drug-discovery

<http://chemistry.uchicago.edu/faculty/faculty/person/member/stephen-bh-kent.html>

CSD vs PDB



Protein Data Bank

pour polypeptides et polysaccharides de plus de 4 unités

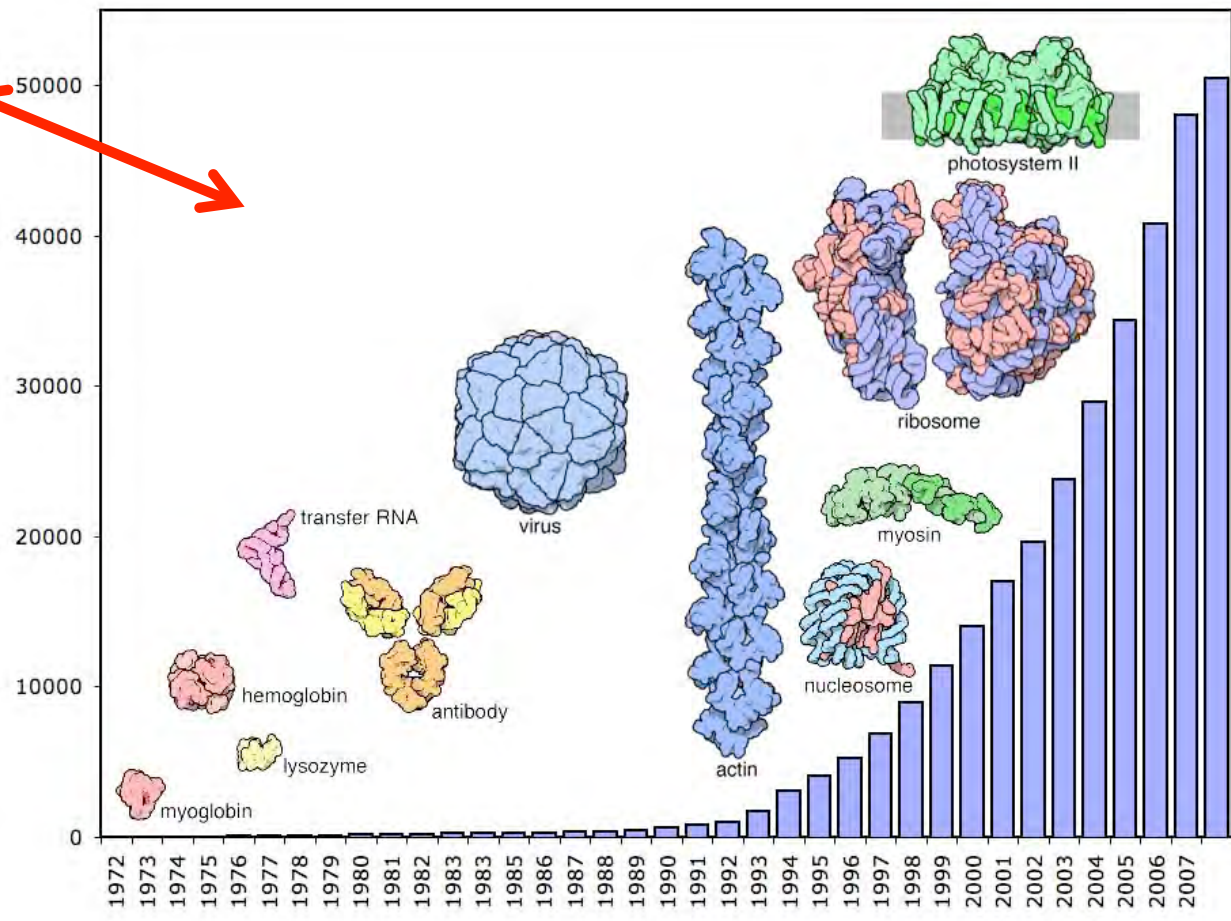
« What is the smallest protein? »

Crystal Structure of a Ten-Amino Acid Protein, Honda et al, JACS 2008

A l'instar des petites molécules il y a 20-30 ans,

La PX est entrée dans l'ère de l'automatisation

et s'attaque aux complexes de plus en plus en grands



Provenance des modèles dans la PDB (en 2011)

Any Experimental Method (75717)

Electron Crystallography (31)

Electron Microscopy (378)

Electron paramagnetic resonance (EPR) (5)

Fiber Diffraction (37)

Fluorescence Transfer (1)

Infrared Spectroscopy (4)

Neutron Diffraction (48)

Powder Diffraction (18)

Solid-State NMR (46)

Solution NMR (9028)

Solution Scattering (44)

Theoretical Model (6)

X-Ray Diffraction (66115)

12%

87%

Devinette :

À quoi reconnaît-on un cristallographe des protéines

Réponse :

À son Mac

http://sage.ucsc.edu/~wgscott/xtal/wiki/index.php/Crystallography_on_OS_X

L'essor actuel de la PX grâce à

- ✓ l'informatique (vitesse de calculs et capacité de stockage)



- ✓ la biologie moléculaire

Et conjointement aux programmes de génomiques structurales, de santé publique, des firmes pharmaceutiques, de la construction des synchrotrons, de la robotisation/miniaturisation de la cristallisation...

Beaucoup de « business » autour de la cristallogénèse et de la cristallographie (prestataires de services), développement de logiciels payants, etc.

Contrecoups avec les cas de fraude avec une douzaine de rétractations récentes

Les cristaux des protéines,

un des freins à la PX

- Mécaniquement fragiles et instables à l'air en raison d'un pourcentage élevé (20-80%) de solvant aqueux

- Très peu de contacts cristallins

- Protéines restent en général actives

- Enzyme turnover
- Ligand binding

- Taille mini. exploitable $5\mu\text{m} \times 10\mu\text{m} \times 30\mu\text{m}$

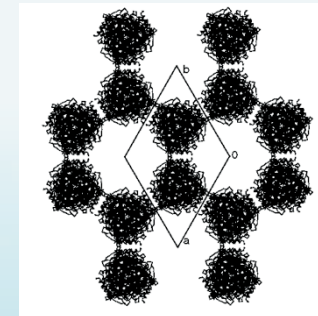
- Nanodrops des robots
- < > cristaux millimétriques (pour les neutrons)

- Grandes mailles (dim $>100\text{Å}$)

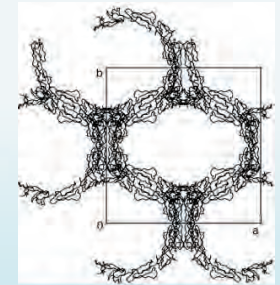
- Résolution rarement atomique

En moyenne 2Å

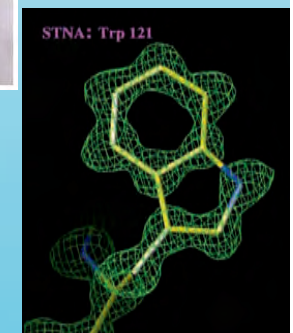
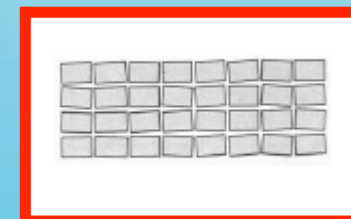
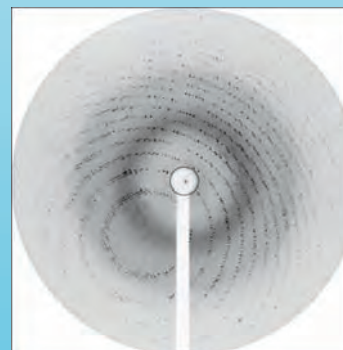
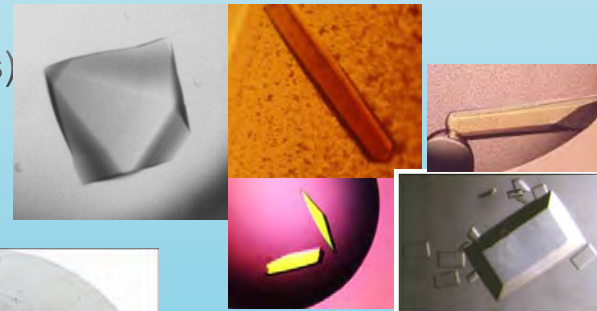
- Mosaïcité $0.2-0.5^\circ$



9 1 2
b5P- 1 2



9 1 2
b1q- 1 2



La cristallogenèse, une science à part entière

qui doit beaucoup à la chimie

(Cristech – Ateliers de formation réguliers : 2011 Cristaux biologiques 29 juin – 1 juillet 2011
Institut de Biologie Moléculaire et Cellulaire ARN – IBMC – CNRS – Strasbourg. Organisateur : Claude Sauter)

Au départ, une **solution aqueuse** avec ϵ diélectrique comme levier d'action

- Approche physico-chimique de la cristallisation

Polyélectrolytes / pH

Sels (*Couche d'hydratation, Salting-in et out, Séries d'Hofmeister*)

1. Larges protéines « nombreux domaines surfaciques hydrophobes »

- MPD, PEG *Phénomènes d'exclusion*

- Approche multifactorielle (cristallisation phénomène multiparamétrique)

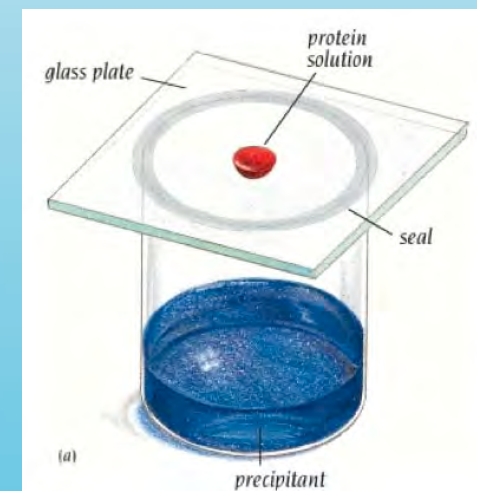
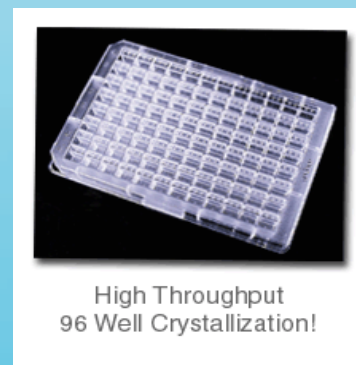
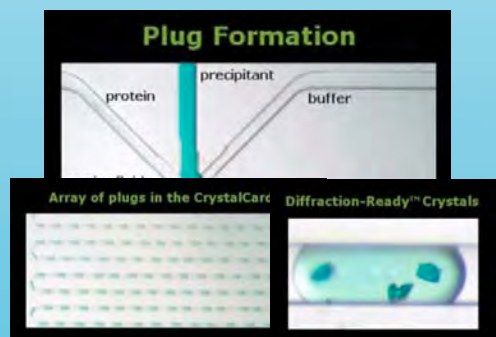
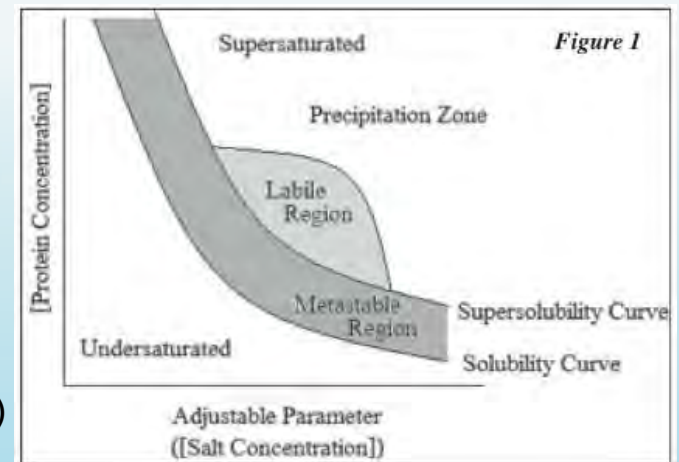
< > **Ere de la Robotisation / Miniaturisation**

- Techniques :

Diffusion de vapeur, dialyse, microbatch sous huile inerte,

Contre-diffusion en gel, capillaire,

Systèmes microfluidiques...



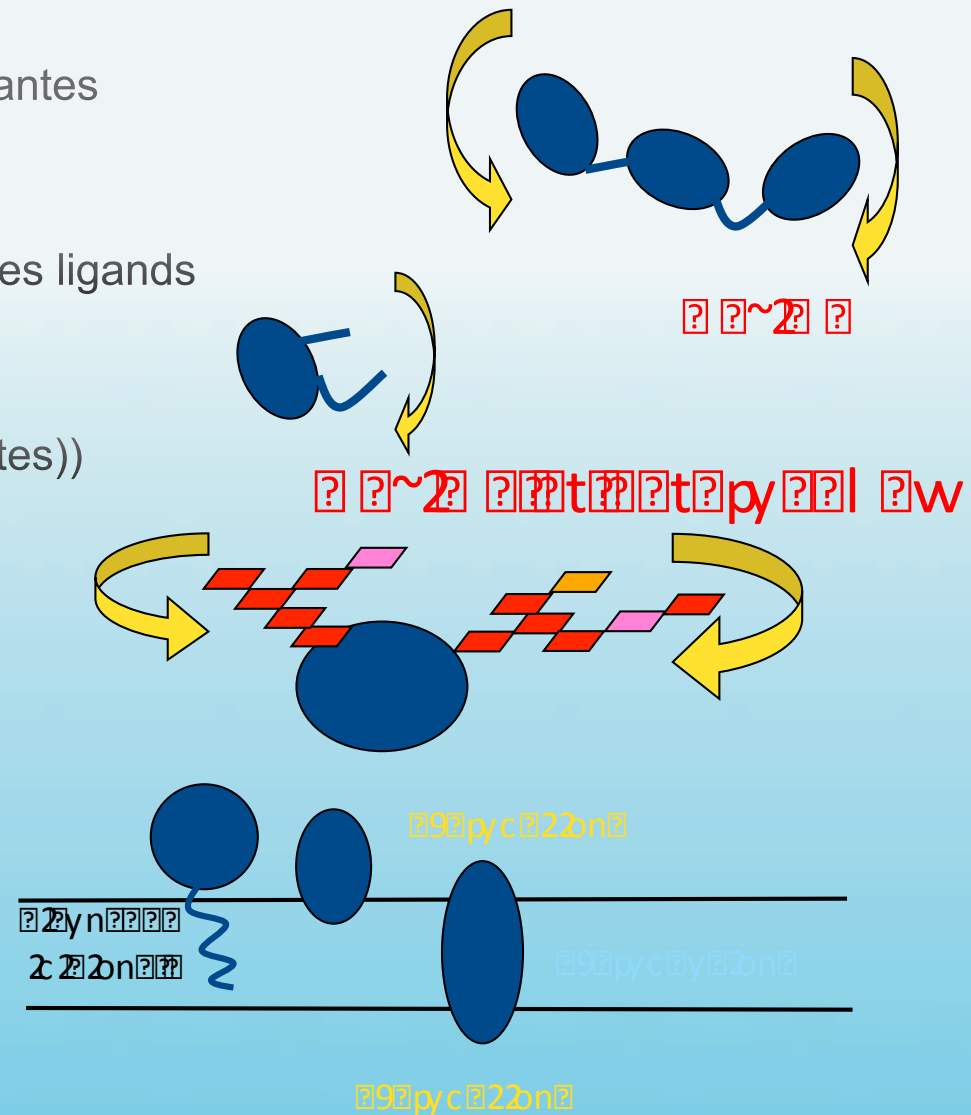
Les cas difficiles à cristalliser

- Domaines multiples / Constructions géantes
- ARN
- Co-cristallisation / solubilité et affinité des ligands
- Boucles, chaînes terminales “floppy”

(modifier processus de purification (étiquettes))

- Glycoprotéines
- Protéines Membranaires

Phases cubiques lipidiques : formant des canaux **continus** d'eau et de lipide qui facilitent la cristallisation de protéines membranaires

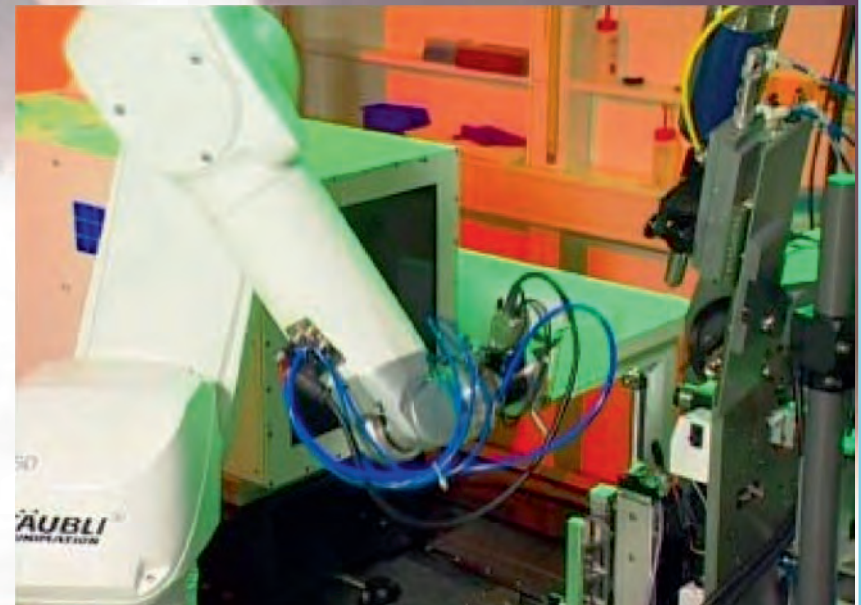


La μ -gravité pour favoriser la croissance cristalline



La DRX directement sur les boîtes de cristallisation

http://www.natx-ray.com/products/G-Rob_2D.html



(Watanabe et al., 2002; Jacquamet et al., 2004).

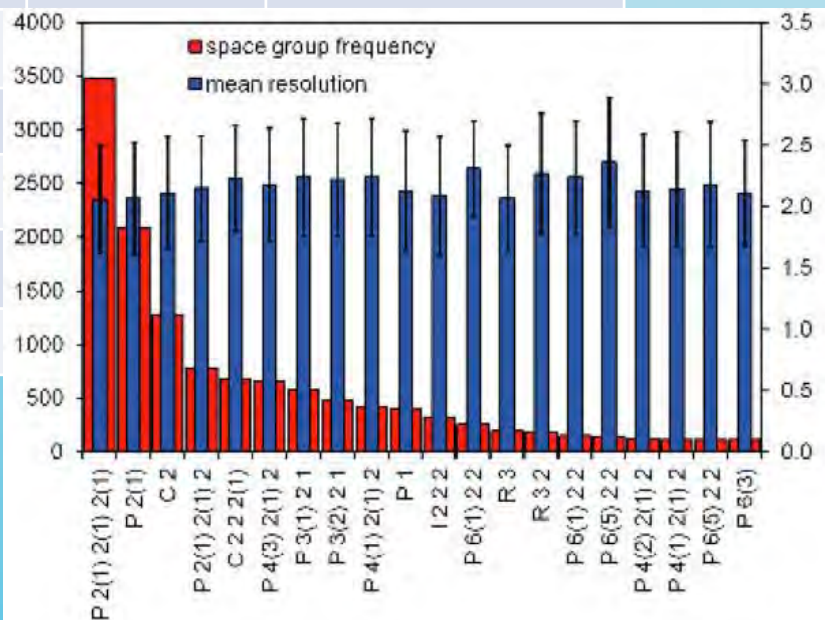
Groupes d'espace

Laue groups: $P, C, C2, C3, C4, C6, C2h, C3h, C4h, C6h, S6, T, Td, O, Oh, I, Ih$
 Point groups: $C1, C2, C3, C4, C6, C2h, C3h, C4h, C6h, S6, T, Td, O, Oh, I, Ih$
 Space groups: $P, C, C2, C3, C4, C6, C2h, C3h, C4h, C6h, S6, T, Td, O, Oh, I, Ih$

Laue group	Point group	Space group
P	$C1$	P
C	$C2$	$C2$
$C2$	$C2$	$C2$
$C3$	$C3$	$C3$
$C4$	$C4$	$C4$
$C6$	$C6$	$C6$
$C2h$	$C2h$	$C2h$
$C3h$	$C3h$	$C3h$
$C4h$	$C4h$	$C4h$
$C6h$	$C6h$	$C6h$
$S6$	$S6$	$S6$
T	T	T
Td	Td	Td
O	O	O
Oh	Oh	Oh
I	I	I
Ih	Ih	Ih

3. Point group symmetry

For chiral space groups (*i.e.* all macromolecular crystals), there is only one point group corresponding to each Laue group. It corresponds to the space group without any translations.

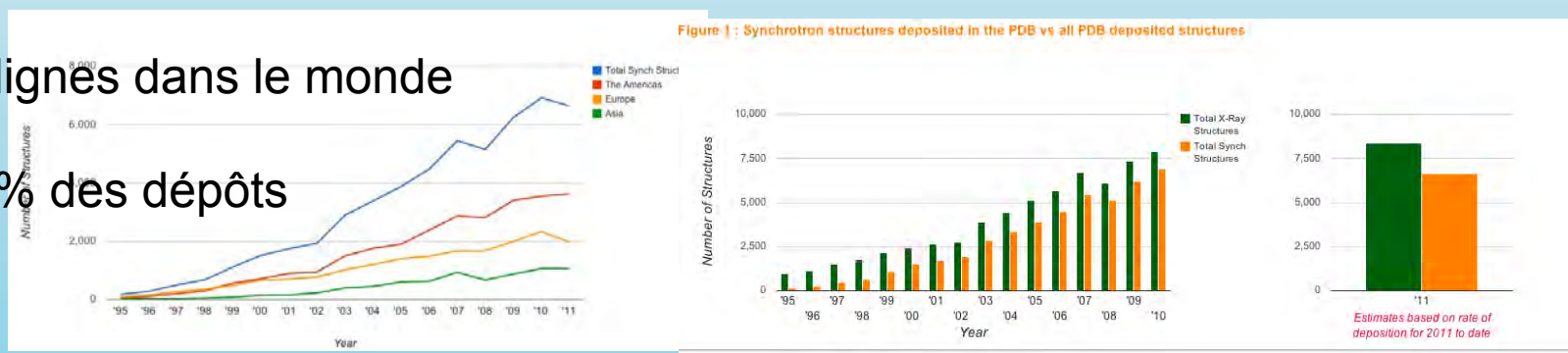


Sources et Expériences de DRX

- Anodes tournantes, Longueur d'onde Cu K α 1.54187Å, détecteur ImagePlate, Phasage au soufre possible **mais optimisé avec Cr K α 2.23Å** (Yang et al., 2003)
- Sources microfoyers (ca 70 μ m), miroirs multi-couches délivrant un faisceau comparable à *2nd-generation bending magnet synchrotron beamlines* (Arndt, Duncumb et al., 1998).
- Synchrotrons de 3^{ème} génération (flux, accordance en λ , passeur d'échantillon, FEDEX, détection CCD...) 360° en 15', qqes secondes par °,...)
Futur ? Concept du 'desktop' synchrotron ' <http://www.lynceantech.com/>

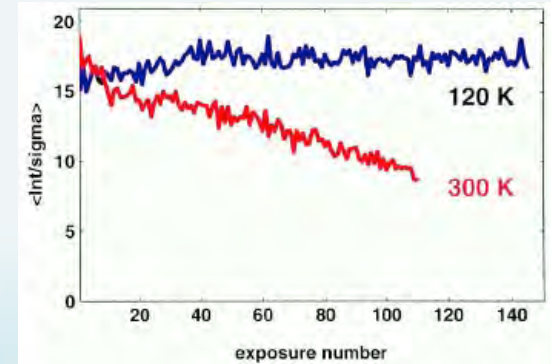
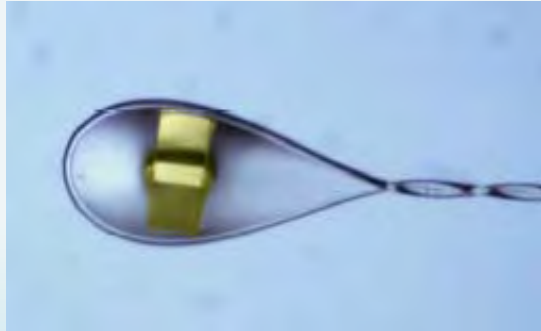
70 lignes dans le monde

75% des dépôts



- **FUTUR : coherent X-ray FEL facilities à Stanford, et à Hamburg)** pour analyser microcristaux et échantillons non cristallins in situ dans des organites (Miao et al., 1998, 2001; Huidt et al., 2003)
- Nouvelle génération de détecteurs (e.g. PILATUS) à transfert rapide d'information (mode de comptage photon unique, rotation constante du cristal, fine-slicing)

Expériences de DRX : l'apport de la cryo



Cryo à 100 K

contre les effets radiolytiques secondaires
(facteur de temps de préservation du cristal $\times 70$) $\ll 155$ K max

- (Flash/annealing) Stream cooling into gaseous cold nitrogen or Plunge cooling into liquid nitrogen 63 - 78K
- Plunge freezing into liquid propane 86-231K, N.B. SAFETY, ethane 90-184K, freon 115-243K, methylcyclopentane 130-345K, freon 14 (tetrafluoro methane) 88-145K

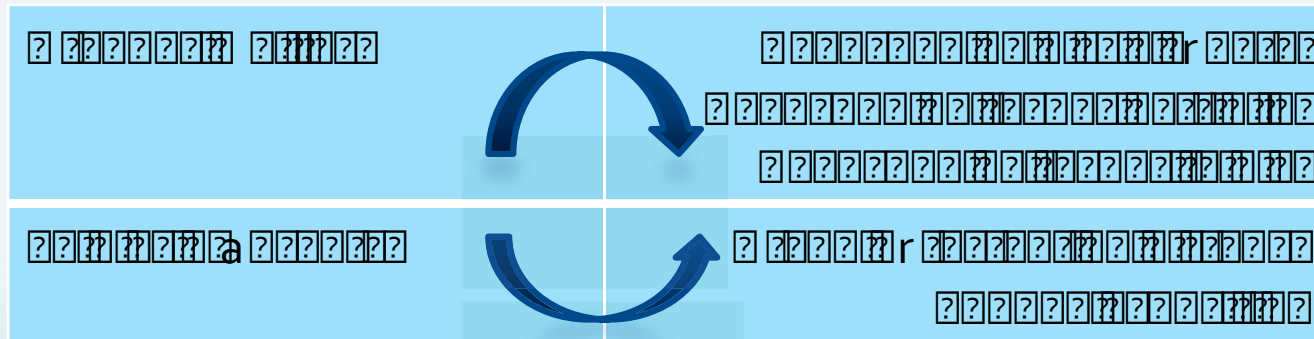
Cryoprotectants nécessaires : Ethylène glycol, glycérol, « higher alcohols », polyéthylène glycols (PEGs), ou solutions concentrées de sucrose ou de sels **mais il faut trouver les bonnes conditions !**

Bénéfices évidents :

facilité de manipulation,
conservation des cristaux (*chaîne de froid respectée*)
acquisitions multiples,
vibrations thermiques et désordre conformationnel réduits, signal sur
bruit et limites de résolution (parfois) augmentés



Autres généralités sur les cristaux



$I/\sigma(I)$ (information expérimentale) est cruciale pour résoudre les structures (notamment pour calculer les poids dans l'affinement des atomes lourds et celui du modèle)

1) Maximiser le signal et 2) Réduire l'erreur

- 1) Plus de flux, de plus gros et meilleurs cristaux, multiples cristaux, optimiser le choix de la l.o. (signal anomal)... (limites atteintes ?)
- 2) Apporter le plus grand soin à la mesure et à son traitement.

Minimiser les erreurs en les analysant, les anticipant, les réévaluant. Source d'erreurs systématiques venant des instruments : adapter les protocoles expérimentaux d'enregistrement et les programmes d'analyses

Enregistrement et traitement des images de DRX

Méthode de rotation classique (Arndt et al., 1973; Arndt & Wonacott, 1977) autour de φ axis
 $\Delta\varphi$ entre 0.2° (*fine-slicing pour améliorer le S/B des interactions faibles*) et $1-2^\circ$ (*mosaïcité + $\Delta beam$*).

$h, k, l, I(hkl), \sigma(I)$



1) Autoindexation + mosaïcité

(*>50 ou centaines de réflexions*, 2 images min. séparées de 90° si symétrie $\leq$ orthorhombique, 3 si triclin.)

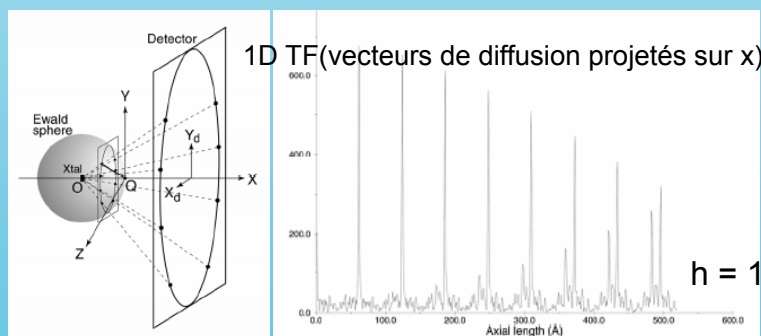
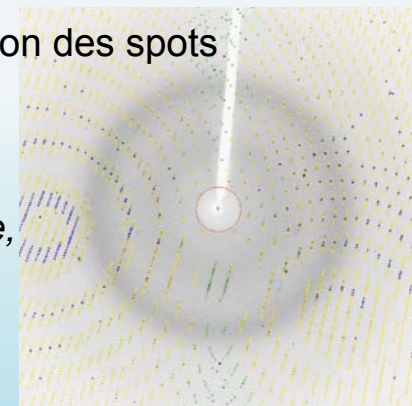
Deux méthodes :

- Approche dans l'espace réciproque, basée sur l'analyse d'un set de vecteurs inter-pics (Kabsch, 1993)
- Approche dans l'espace direct, basée sur l'interprétation de la TF du motif de diffraction (Rossmann & van Beek, 1999) **Information sur la maille, orientation / faisceau direct, détecteur, axes du gonio.**

Prédiction la meilleure possible de la position des spots sur plusieurs images

Difficultés rencontrées :

faible diffraction, macle,
 diffraction anisotrope, diffusion diffuse,
 anneaux de glace, mosaïcité élevée
 spots non résolus, « overloads »,
 zingers/rayons cosmiques.



2) **Post-refinement** : paramètres de mailles précis, idéalement qqes ° de données sur 2 segments séparés -> précision 1 pour 10000 ($\leq 2.8\text{\AA}$), orientation 0.01° (*avant intégration et fixe durant*).

3) **Intégration des images** + *affinement de la matrice d'orientation du cristal, de la mosaïcité et des paramètres de détection*

(**2D** ou **3D** selon le mode d'intégration des réflexions partielles)

Deux méthodes :

i) par simple sommation des pixels au sein d'un profil de réflexion

ii) par profile-fitting facteur d'échelle à appliquer à un profil standard à partir d'un set de réflexions bien mesurées.



Estimations $\sigma(I)$ cruciales pour multiples observations

Pour i) et ii)-partielles, statistiques de Poisson s'appliquent $N^{1/2}$ ainsi que pour les faibles et moyennes réflexions.

Fine-slicing et profile-fitting améliorent le S/B des faibles réflexions

Les logiciels les plus répandus :

MOSFLM (Leslie, 1992),
HKL2000 (Otwinowski & Minor, 1997),
XDS (Kabsch, 1988)
d*TREK (Pflugrath, 1999)
XGEN (A. J. Howard/)

Stratégies d'enregistrement

Compromis entre statistiques (assez de photons X par réflexions et multiplicité)
et « radiation damage » qui contrôle la durée d'exposition

- Rotation totale

Idéalement 180° (ou 360° en P1 pour avoir les données anomaless complètes)

Utiliser les modules « stratégie » (eg Mosflm)

*pour meilleure optimisation (eg 90° pour orthorhombique,
ou 2 x 30°) et définir le point de départ*

- Rotation/image: **pas nécessairement 1°!**

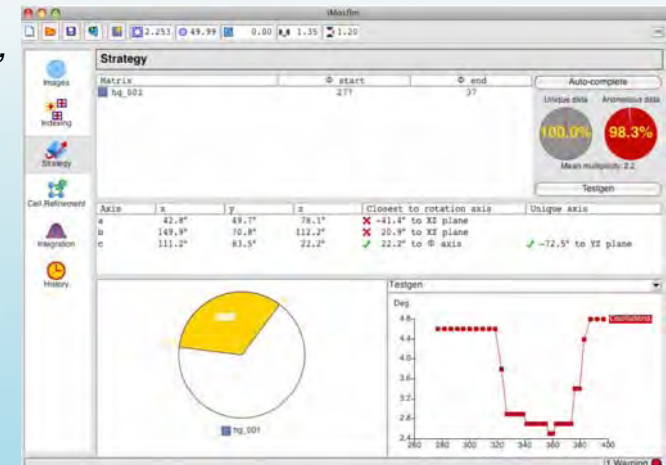
souvent entre 0.25° - 0.5°, minimise « overlap » et bruit de fond

- Temps/image

dépend souvent du temps disponible (synchrotron)

- Position du détecteur

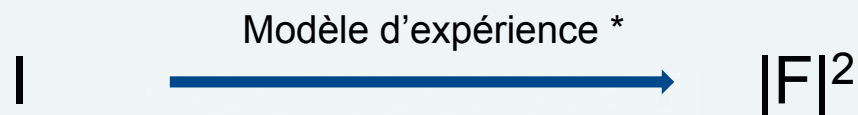
le plus loin possible pour diminuer le bruit de fond et améliorer la résolution entre pics.



2 cas à prendre en compte :

- MAD voir plus loin
- IAffinement
 - Maximiser la résolution : (2 runs long et rapide: haute et basse résolution)
 - Multiplicité haute car utile (même si moins importante que pour le AD)
 - Courte l.o. (<1Å) pour minimiser l'absorption
 - Enregistrer les réflexions symétriques à des temps et géométries différents pour une erreur absolue sur I la plus faible (e.g.rotation autour de différents axes du cristal)

Mise à l'échelle et moyennation des données

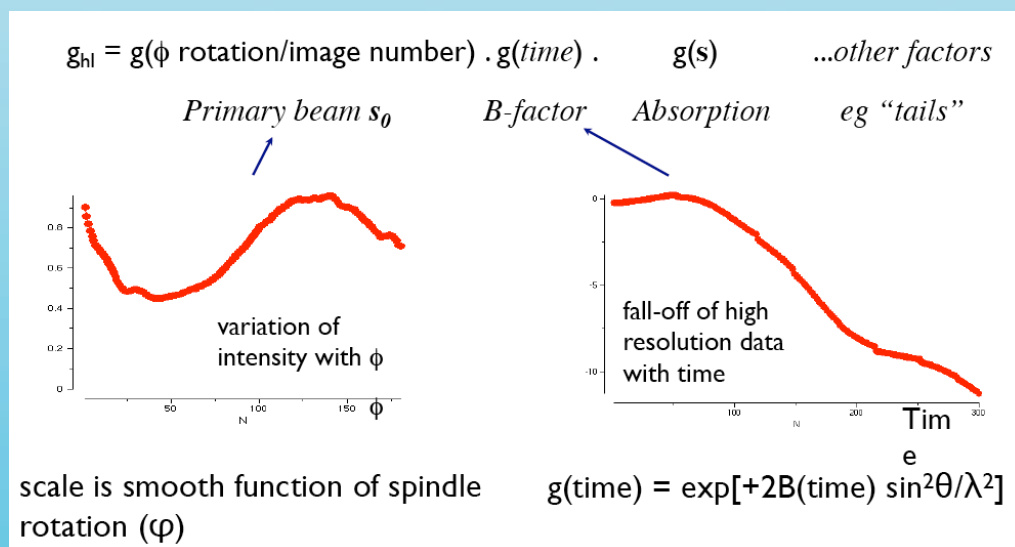


Répondre à la question le jeu de données est-il bon? (ie à quelle résolution ?, quelles images jeter ?)

*
Doit prendre en compte les facteurs liés - au faisceau incident (synchrotron) fluctuations en I, synchronisation avec l'obturateur, volume d'illumination... - au cristal et faisceau diffracté (radiation damage $\longleftrightarrow$ B-facteur) - au détecteur (calibration, détection, beamstop)

$$\Phi = \sum_{hl} w_{hl} (I_{hl} - I/k_{hl} \langle I_h \rangle)^2$$

Comparaison des réflexions équivalentes par sym., en ajustant les facteurs d'échelle pour une meilleure consistance interne des intensités.



Les critères de qualité

I. R-factors: traditional overall measures of quality

$$(a) R_{\text{merge}} (R_{\text{sym}}) = \sum |I_{hl} - \langle I_h \rangle| / \sum \langle I_h \rangle$$

This is the traditional measure of agreement, but it increases with higher multiplicity even though the merged data is better

$$(b) R_{\text{meas}} = R_{\text{r.i.m.}} = \sum \sqrt{(n/n-1)} |I_{hl} - \langle I_h \rangle| / \sum \langle I_h \rangle$$

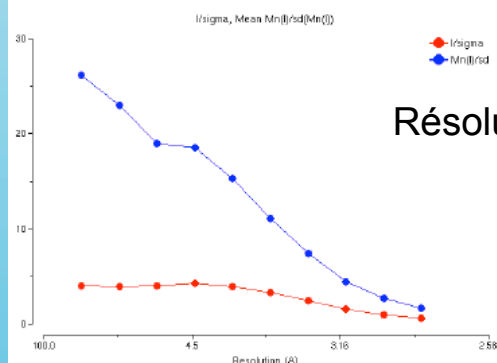
The multiplicity-weight R-factor allows for the improvement in data with higher multiplicity. This is particularly useful when comparing different possible point-groups (it is output by *POINTLESS* along with the correlation coefficient, as well as in *SCALA*)

$$(c) R_{\text{p.i.m.}} = \sum \sqrt{(1/n-1)} |I_{hl} - \langle I_h \rangle| / \sum \langle I_h \rangle$$

“Precision-indicating R-factor” gets better (smaller) with increasing multiplicity, ie it estimates the precision of the merged $\langle I \rangle$

Diederichs & Karplus, *Nature Structural Biology*, 4, 269-275 (1997)

Weiss & Hilgenfeld, *J. Appl. Cryst.* 30, 203-205 (1997)



Résolution limite des données (courbe bleue)

$$\langle I \rangle / \sigma(\langle I \rangle) \geq \sim 2 \text{ (1.5 à 1.0 si anisotropie)}$$

Générales considérations (suite)

Détection d'« outliers » (meilleure si multiplicité grande)

liés à des zones de détection (beamstop), ice spots, zingers, overlaps, twinning, mauvaise prédiction

Détection du signal anomal

Distribution des intensités et pathologies.

déviatio*n* / Wilson plot

Twinning :



The Merohedral Crystal Twinning Server
[Servers Home]

People ♦ Seminars
Lectures ♦ Webmail
Links ♦ Facilities
Software ♦ Home

UCLA

Todd Yeates and Barry Farn
UCLA Department of Chemistry and Biochemistry and
UCLA-DOE Laboratory of Structural Biology and Molecular Medicine



Reference: T. O. Yeates, *Methods Enzymol.* **276**, 344-358, 1997.

TWINNING TEST

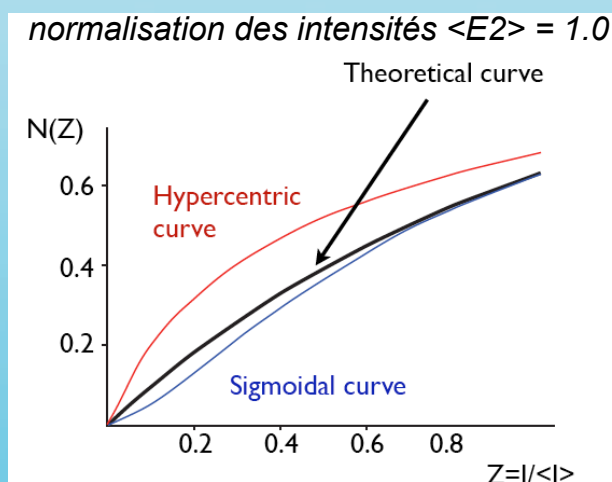
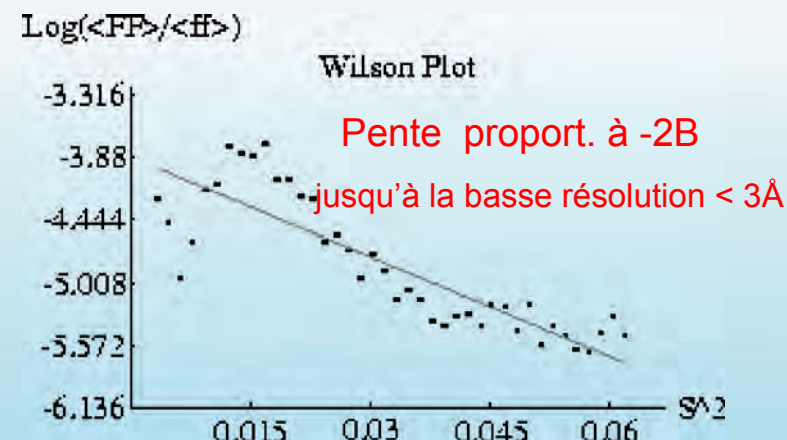
For more information on both types of twinning the please see the [Introduction to Hemiheral Twinning](#)

☐ Perfect Merohedral Twinning Test

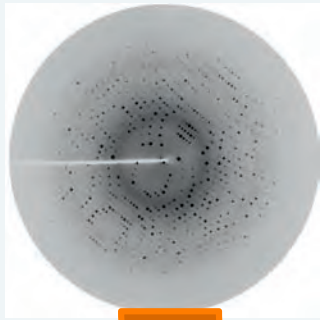
Perfect Merohedral Twinning involves a crystal with a twin fraction equal to or near 1/2. This test will attempt to determine the presence of twinning by comparing the diffraction data to that expected by Wilson statistics.

☐ Partial Merohedral Twinning Test

Partial Merohedral Twinning involves twinning with a twin fraction not equal to 1/2. This test will attempt to estimate the twin fraction alpha statistically.

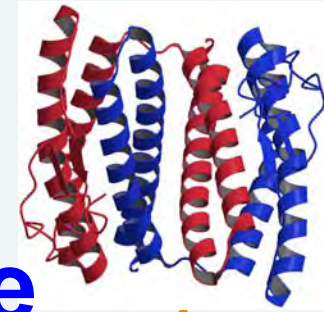


Réflexions non centrées : 18.1% des réflexions ayant une intensité $< 20\%$ à $\langle I \rangle$ à cette résolution



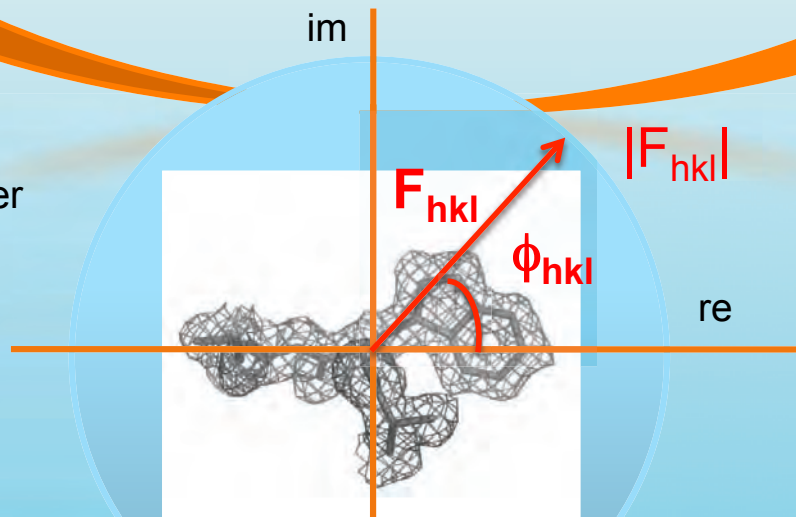
Données

La résolution structurale



Structure

Construction de Harker
ou
Diagramme d'Argand



$$\rho(x, y, z) = \frac{1}{V_{\text{maille}}} \sum_{hkl} |F_{hkl}| e^{i\phi_{hkl}} e^{-2i\pi(hx+ky+lz)}$$

Les phases cristallographiques à déterminer !

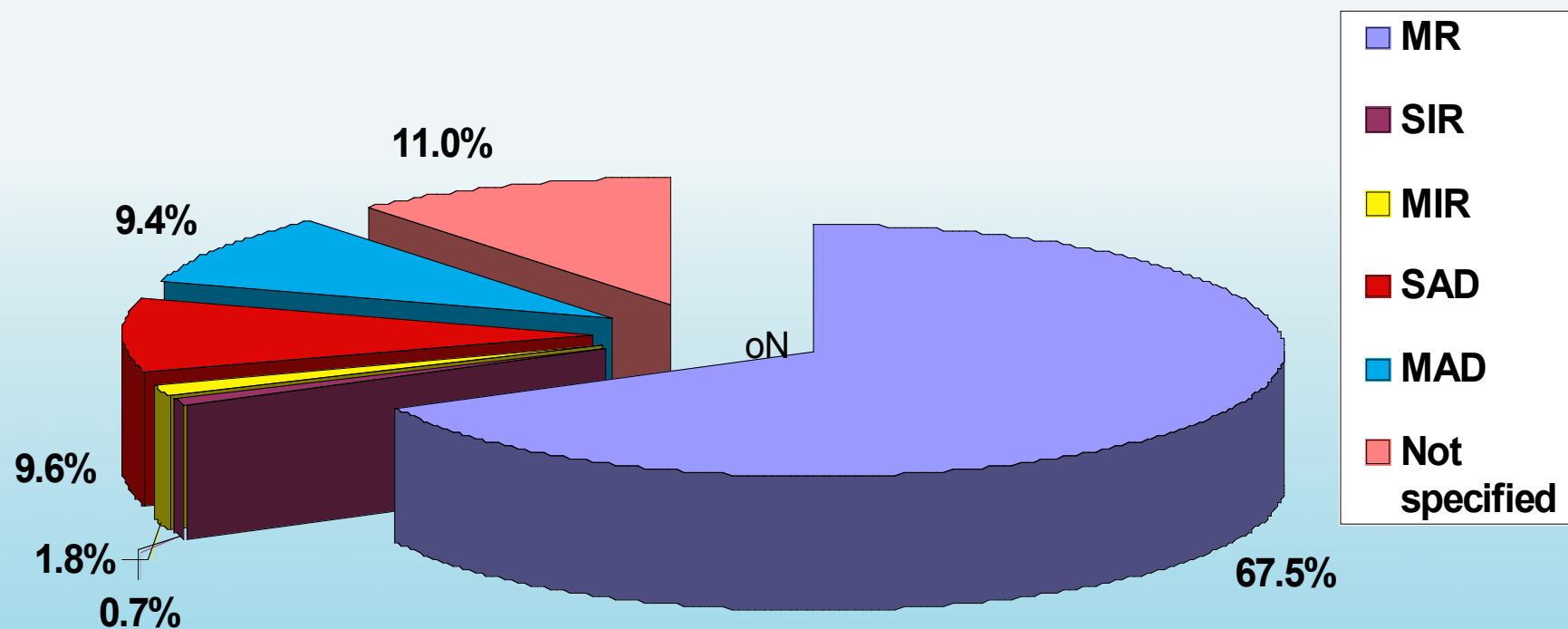
Méthodes de Phasage

Méthode	A priori – conditions
Méthodes directes	$\rho \geq 0$, atomes discrets ≤ 1000 atomes (lysozyme 124 aa 14kDa) résolution 1.2Å
Charge flipping*	Purement mathématique 6000 atomes résolution 1.0 Å
Remplacement Moléculaire	Modèle homologue
Remplacement Isomorphe	Sous-structure d'atome lourd
Diffusion anormale	Sous-structure de diffuseur anormal
Dommages induits par la radiation	
<i>Amélioration de la phase</i>	<i>Aplanissement de solvant</i> <i>Histogram matching</i> <i>Moyennation NCS</i> <i>Structure partielle</i> <i>Extension de phase</i>

* Christian Dumas and Arie van der Lee

Macromolecular structure solution by charge flipping Acta Cryst. (2008). D64, 864–873

Répartition des méthodes (PDB)



Acronymes

- MR Molecular Replacement
- SIR Single wavelength Isomorphous Replacement
- SIRAS Single wavelength Isomorphous Replacement with Anomalous Scattering
- MIR Multiple wavelength Isomorphous Replacement
- MIRAS Multiple wavelength Isomorphous Replacement with Anomalous Scattering
- SAD Single wavelength Anomalous Dispersion
- MAD Multiple wavelength Anomalous Dispersion
- RIP Radiation-damage-Induced Phasing
- RIPAS Radiation-damage-Induced Phasing with Anomalous Scattering

Le Remplacement moléculaire

Principe

Rotation et translation pour placer le modèle sur la structure du nouveau cristal

N molécules dans l'ua x 6 paramètres (α β γ , t_x t_y t_z)

i.e. **rotations** 0-360° 0-180° 0-360° (par pas de 2.5°)
+ **translations** 100 x 100 x 100 Å (par pas de 1Å)

environ 10^{12} points à tester (approche **6D**)
à 10^6 si rotation puis translation (approche **3D**)

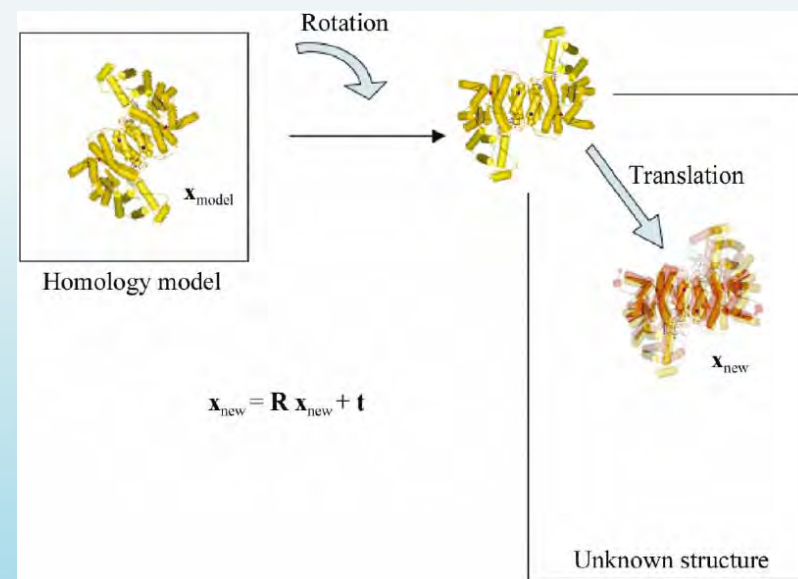
Comment accélérer le temps de calculs ?

Avantages

- Plus il y a de structures, plus les chances de trouver des modèles homologues augmentent
- Pas besoin d'obtenir des dérivés lourds ...
- Construction du modèle simplifiée car peu de tracé à réaliser...

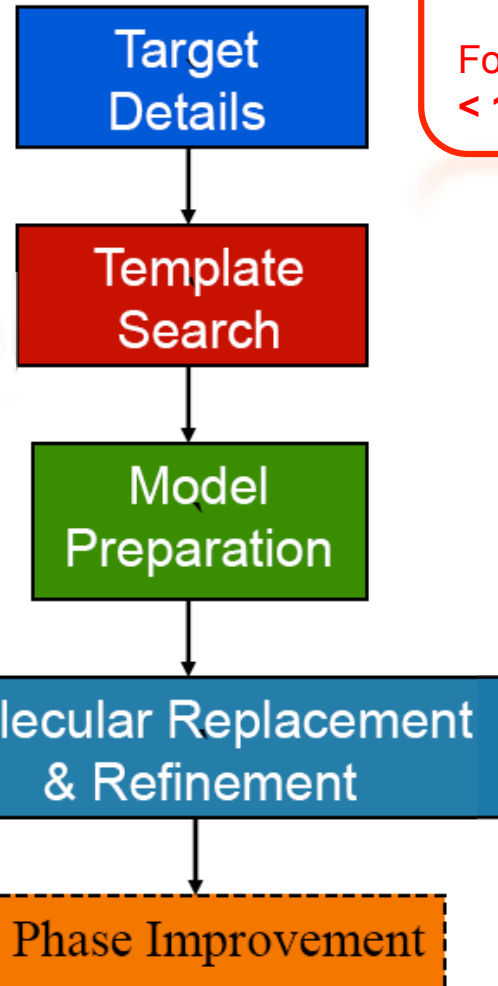
Inconvénient majeur ?

Model bias results from using an atomic model to calculate crystallographic phases. The resulting electron density map will tend to have the features present in the model even if they are not actually present in the structure (*Ramachandran & Srinivasan, 1961; Read, 1986; Bhat, 1988; Hodel et al., 1992; Adams et al., 1999; Kleywegt, 2000*).



Procédure type :

Matthews coeff.
(copies in a.s.u.)
Self RF (rotational NCS)
Native Patterson
(NCS // crystal. axe)
B factor analysis



Séquence alignement

Outils biotechnologiques type FASTA

Forte homologie de séquence >25%, min 17%
< 1.7Å r.m.s.d.!

Chainsaw

- retirer les résidus non présents dans la cible
- retirer les atomes des chaînes latérales
- retirer les régions incertaines du modèle (B factors, occupancies), les boucles flexibles

Modèle polyala

Multi-domaines

Multimères

Complexes

1 - Méthodes de Patterson

($TF|F|^2$ avec phases=0, carte de densité de vecteurs interatomiques)

Fonction de rotation (self-vectors) puis Fonction de translation enfin Rigid Body

2 - Méthodes de maximum de

vraisemblance (approche statistique dans l'espace réciproque)

$P(F_o | F_c(R,t))$ à partir de toutes les copies de la maille (self/cross vectors) ?

Critères pour valider une solution de RM (Si > 3Å de résolution , scepticisme)

Score des solution(s) RF, TRF, nettement au-dessus des autres

Empilement cristallin / groupe d'espace (dont 11 paires énantiomorphiques, e.g. P 4₁ et P 4₃).

Baisse sensible des R fact/R free (ca 50%?) après premiers cycles d'affinement

Composite OMIT Maps : les modèles tronqués réapparaissent-ils dans la carte de densité ?

Techniques pour réduire le biais de la solution RM

- **OMIT map** (Bhat, 1988); 'kicked' – (Guncar et al., 2000); Simulated-annealing – (Hodel et al., 1992; Brunger et al., 1998)

σ_A -map: $2mF_{OBS}-DF_{MODEL}$ (Read, 1986; Urzhumtsev et al., 1996)

- Model rebuilding with randomization (Zeng et al., 1997; Reddy et al., 2003)

- Prime-and-switch density modification (Terwilliger, 2004)

- Carry out the usual model building and refinement avoiding a specific model part, such as ligand

- 'ping-pong refinement' (Hunt & Deisenhofer, 2003)

- Maximiser le LOG-likelihood comme implémenté dans BUSTER (GlobalPhasing Ltd)

FUTUR : *Structures à partir de simples fragments de structures secondaires typiques,
Modèles homologues distants en utilisant ROSETTA champs de force (voir PHENIX)*

Liste nonexhaustive de logiciels la plupart « Brute force» recherches 6D – possible aujourd'hui

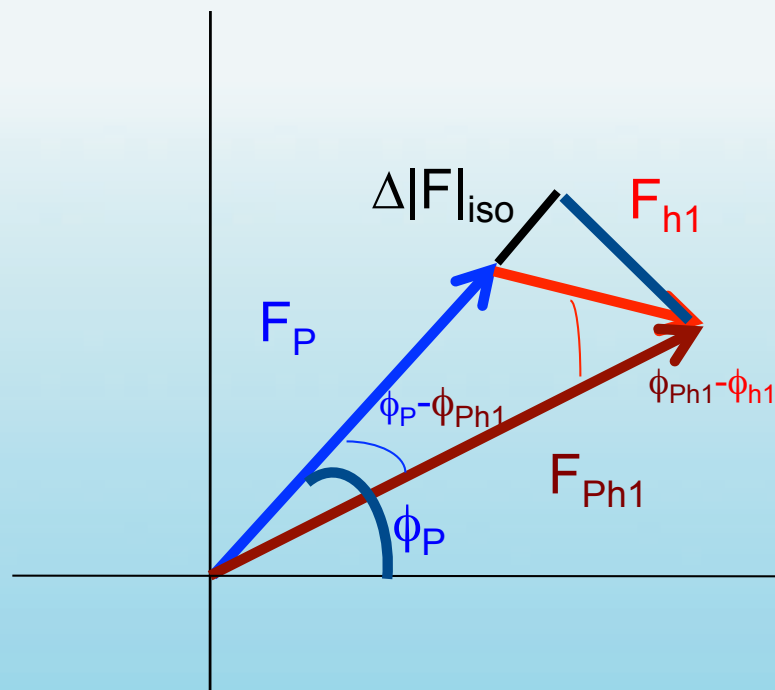
SOMoRe (low resolution coarse search followed by refinement of most likely solutions),

EPMR (evolutionary search), **BRUTE** (just plain crunching the N dimensions)

MOLREP, **PHASER** (maximum de vraisemblance, teste les groupes d'espace, utilise un "Ensemble" (i.e. plusieurs modèles superposés de type RMN), les deux sont intégrés dans **Mr BUMP**.

Phasage expérimental et relation triangulaire

Pour avoir ϕ_P , obtenir $|F_{h1}|$, ϕ_{h1}



à partir de $\mathbf{F}_{Ph1} = \mathbf{F}_P + \mathbf{F}_{h1}$

$$|F_P| \gg |F_{h1}|$$

$$\Delta|F|_{iso} = |F_{Ph1}| - |F_P| = |F_{h1}| \cos(\phi_{Ph1} - \phi_{h1})$$

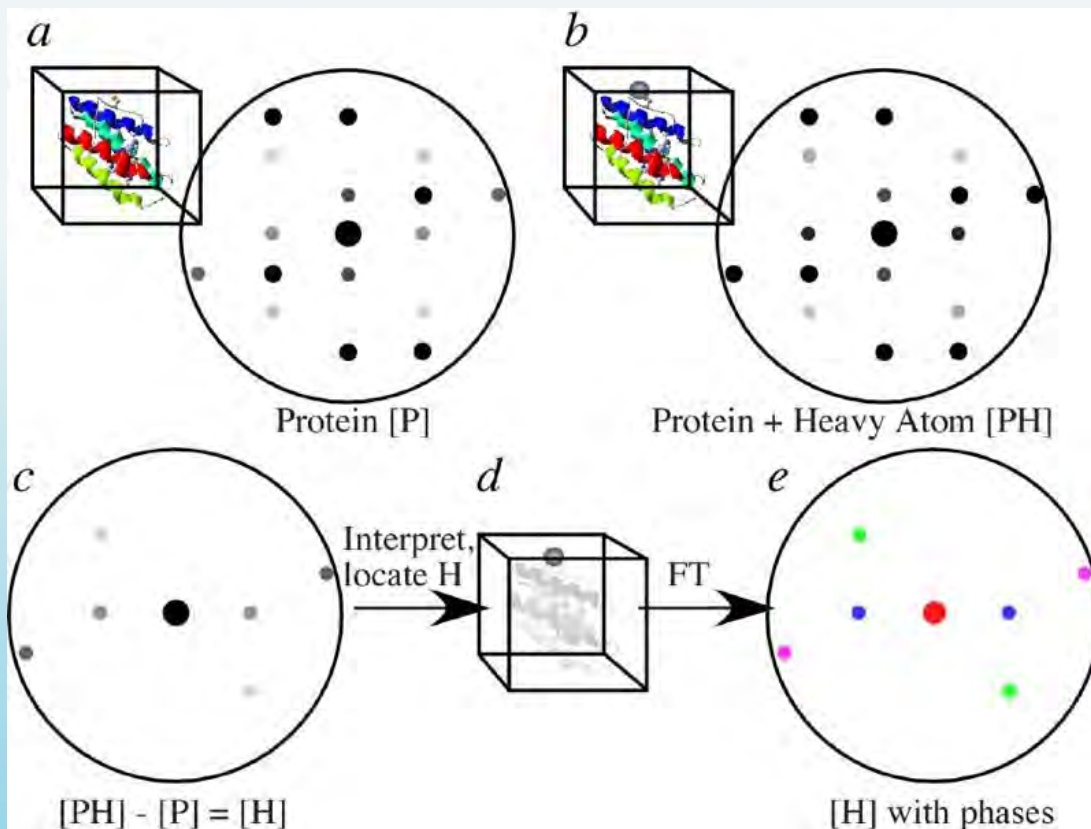
utilisé pour déterminer la sous-structure de l'atome lourd
par **méthodes directes** ou par **Patterson** :

$$\Delta|F|^2_{iso} = \frac{1}{2} |F_{h1}|^2 + \frac{1}{2} |F_{h1}|^2 \cos 2(\phi_{Ph1} - \phi_{h1})$$

(pic)

(Bruit)

Le phasage expérimental isomorphe



- a) Enregistrement sur cristal "natif"
- b) Introduction d'un atome lourd dans des sites spécifiques
 Hg^{2+} pour $-\text{CysSH}$ libre ou MET,
 Pt^{2+} pour MET,
puis enregistrement du cristal dérivé,
mesure des différences d'intensité pour chaque hkl
- c) Détermination de la position de l'atome lourd

ECUEILS :

- le manque d'isomorphisme
- le taux d'occupation

$$\cong \sqrt{2} \times \frac{f_{atomelourd}}{7} \times \sqrt{\frac{N_{atomelourd}}{N_{tot}}}$$

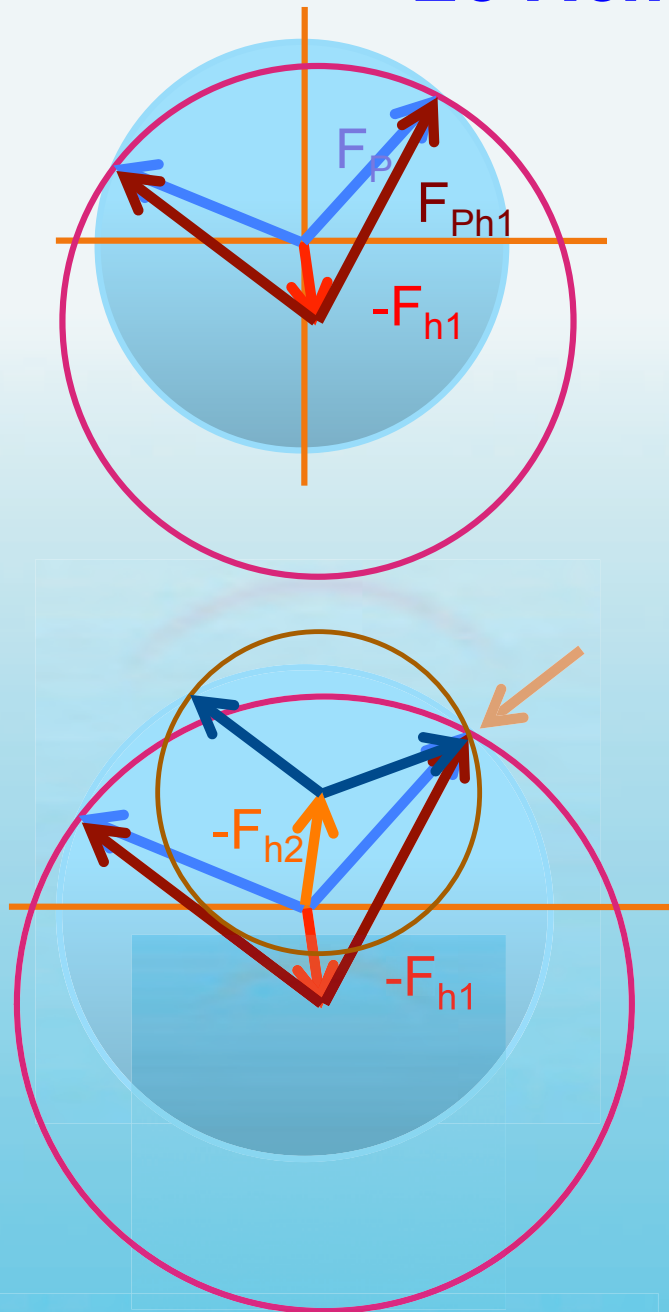
Pour les non-centrées

Ex : 1 atome Hg 80 e⁻ pour 7000 e⁻
soit 0,01

Contribution isomorphe relative

⬅ ⬅⬅
--

Le Remplacement Isomorphe?



F_p = protéine native

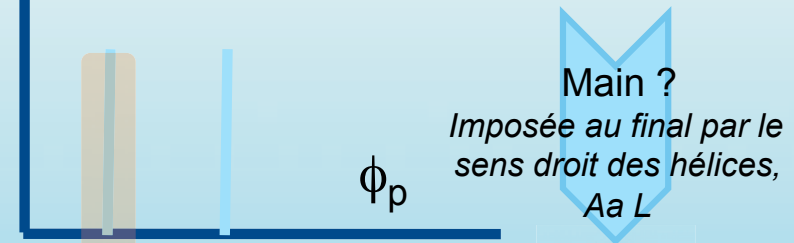
F_{h1} = atome lourd #1

F_{Ph1} = dérivé atome lourd #1

Le centre du cercle F_p est placé à l'extrémité de $-F_{h1}$

La distribution de la probabilité des phases montre que SIR résulte en une ambiguïté sur la phase (bimodale)

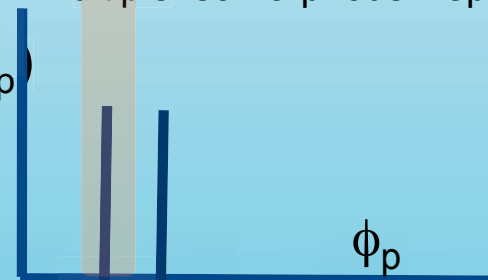
$P_{h1}(\phi_p)$



- Un second dérivé lourd :

Multiple Isomorphous Replacement alias **MIR**

$P_{h2}(\phi_p)$



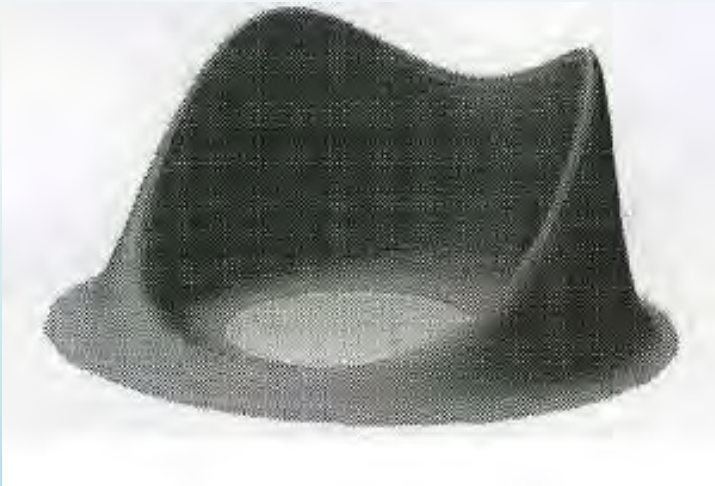
- Contribution anormale (voir plus loin)

Méthode **SIRAS**

- Aplatissement de solvant (voir plus loin)

Erreurs de fermetures

manque d'isomorphisme, erreurs sur les données, sur les positions atomiques...



$$\begin{aligned}\varepsilon &= |F_{PH(obs)}| - |F_{PH(calc)}| \\ &= |F_{PH(obs)}| - |[F_P \exp(i\alpha_P) + F_H \exp(i\alpha_H)]|\end{aligned}$$

Distribution de probabilité de phase

Maximum de la distribution : phase "la plus probable"

Centroïde de la distribution : phase "la meilleure", α_{best}

$P(\alpha) = N \exp(A \cos(\alpha) + B \sin(\alpha) + C \cos(2\alpha) + D \sin(2\alpha))$,
A, B, C, D coefficients d'Hendrickson-Lattman

Pour obtenir la distribution finale, on multiplie les probabilités.

$P(\alpha) = P_1(\alpha) \cdot P_2(\alpha) \dots P_n(\alpha)$

$i+j \quad P_i(\alpha)P_j(\alpha) A_i + A_j B_i + B_j C_i + C_j D_i + D_j$

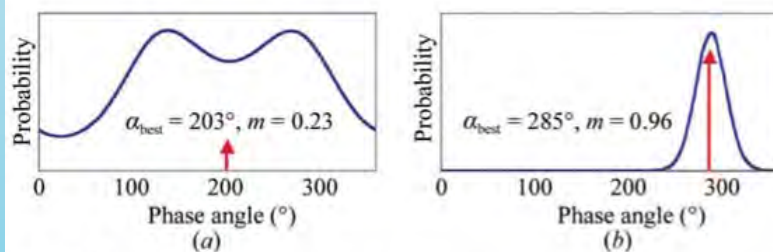


Figure 13

Phase probability for one reflection. (a) Single derivative in an SIR experiment. (b) Three derivatives. In an MIR experiment $P(\alpha_P) \propto \prod \exp(-\varepsilon_i^2/2E_i^2)$, where i is summed from 1 up to the number of derivatives.

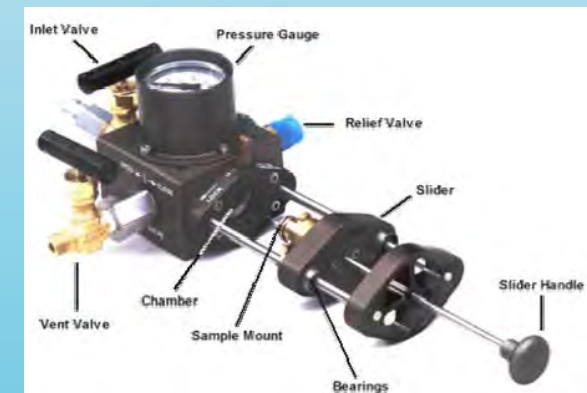
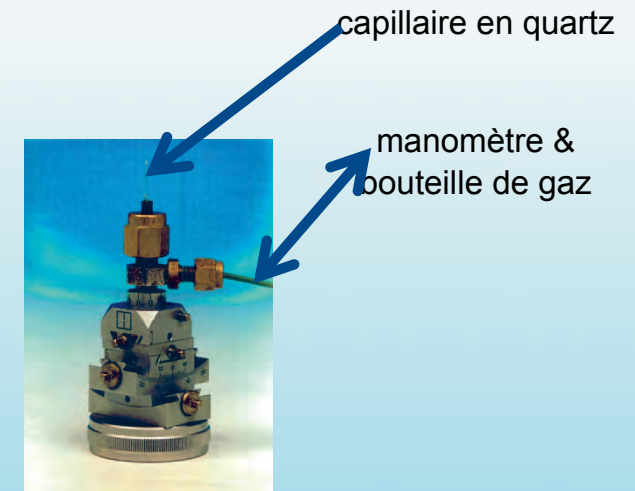
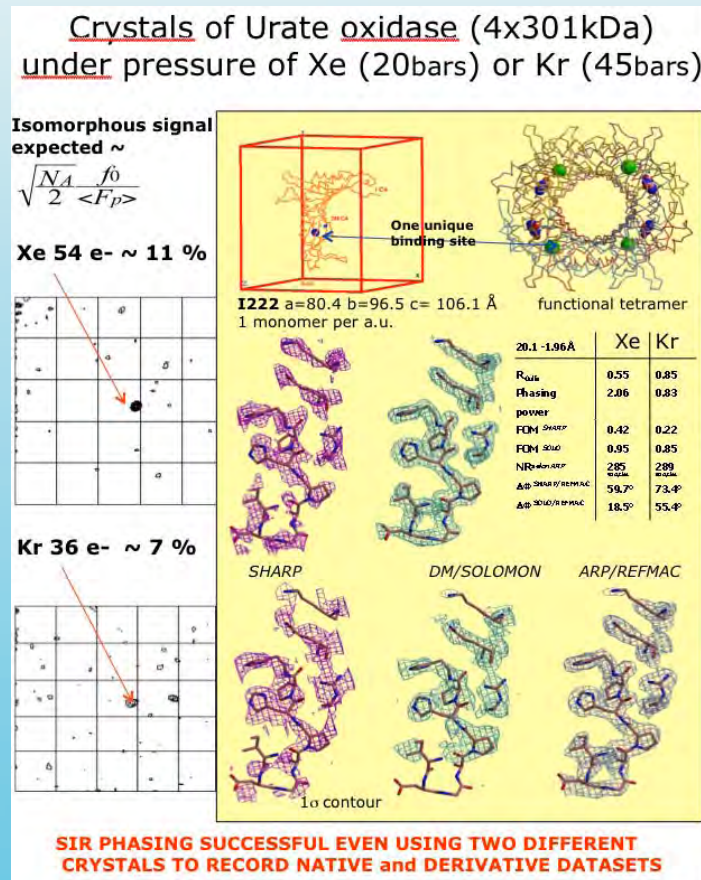
$$F_{best} = \frac{\int_0^{2\pi} P(\alpha) F_P(\alpha) d\alpha}{\int_0^{2\pi} P(\alpha) d\alpha} = |F_{best}| e^{i\alpha_{best}} = m_{hkl} |FP| e^{i\alpha_{best}} m_{hkl} = \frac{|F_{best}|}{|FP|}$$

On peut associer une pondération la « best phase » :
c'est la figure de mérite (FOM) $0 \leq m_{hkl} \leq 1$

Contribue à réduire le biais du modèle initial

Les techniques respectant l'isomorphisme

- Substitution (SeMet), co-cristallisation, trempage, adduits (caco), analogues de substrats modifiés
- Gaz noble sous pression Kr, Xe (Shiltz & Prangé)



Méthodes de Dispersion Anomale (“AD”)?

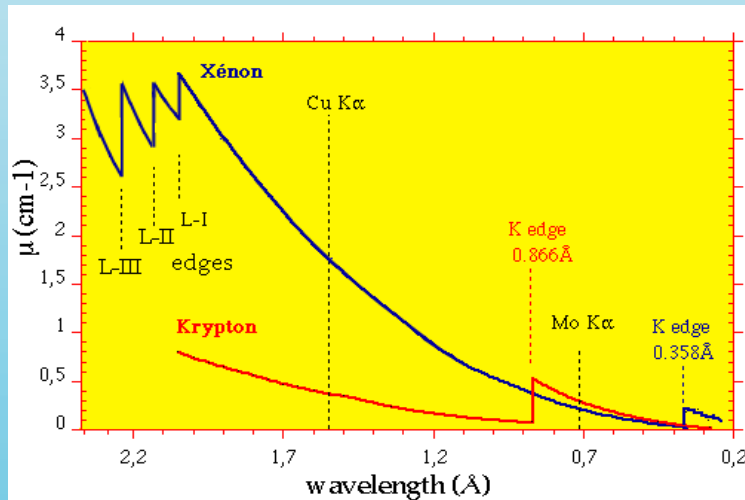
- Tout élément présente une contribution anormale intrinsèque en fonction de la longueur d'onde d'irradiation.

$$f_{\lambda} = f^{\circ} + \Delta f_{\lambda}' + i\Delta f_{\lambda}''$$

Partie constante prop. à Z,
indép. de λ et baisse avec la résolution

Corrections indépendantes de la résolution **réelle** (+/-) et
imaginaire (≥ 0 et $+\pi/2$ en avance sur $f^{\circ} + \Delta f_{\lambda}'$)

- Pour les éléments légers (H, C, N, O), les effets anomaux sont négligeables
- Pour les éléments lourds, aux seuils d'absorption, les effets AD peuvent être très importants.

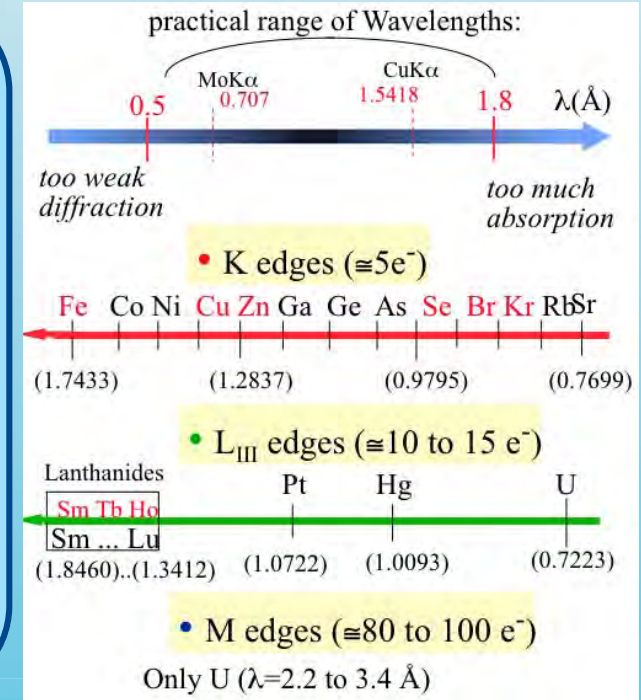


Contribution anormale

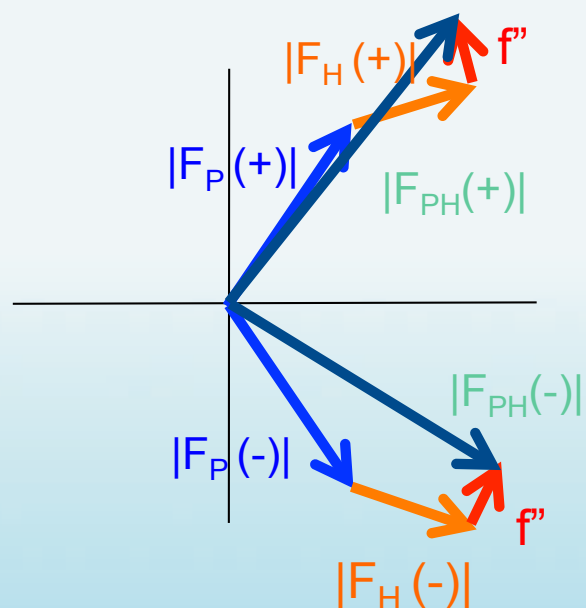
$$\frac{\Delta F^{\pm}}{F} = \frac{1}{Z_{\text{eff}}} \left(\frac{2 \sum_j q_j^2 f_j'^2}{N_P} \right)^{1/2}$$

d'un ordre de grandeur
inférieur à la contribution
Isomorphe
de l'ordre de 1 %

D'où grande précision requise



- En présence de signal anomal, la loi de Friedel n'est plus vérifiée



$$|F_P(+)| = |F_P(-)|$$

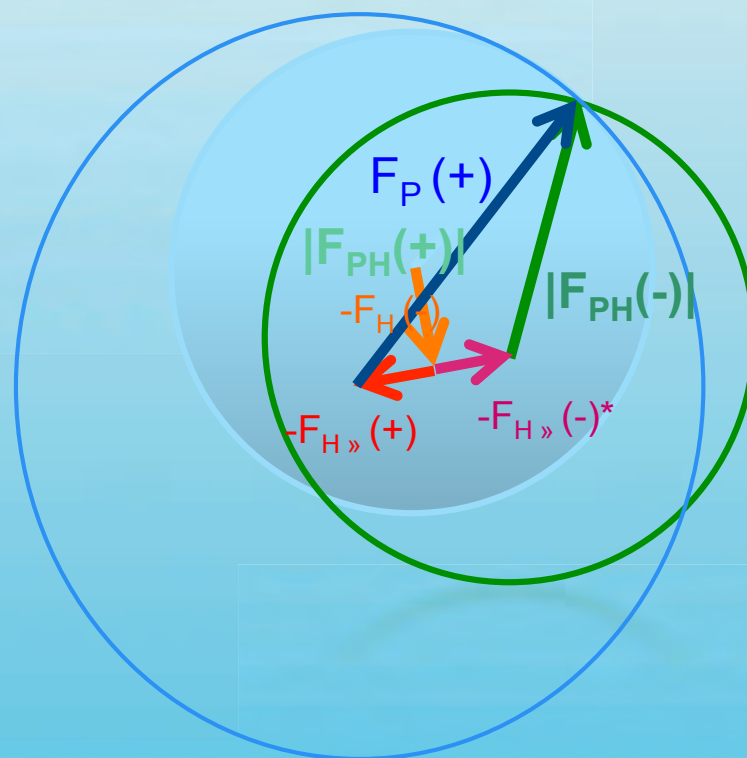
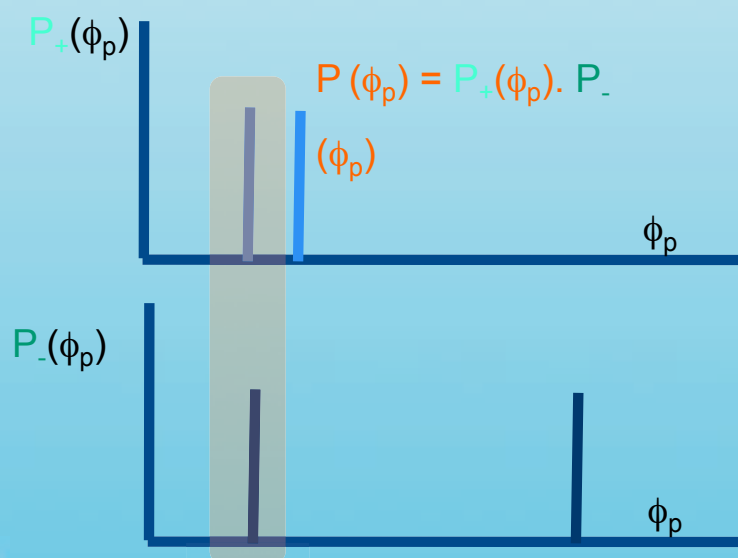
pour les acentriques

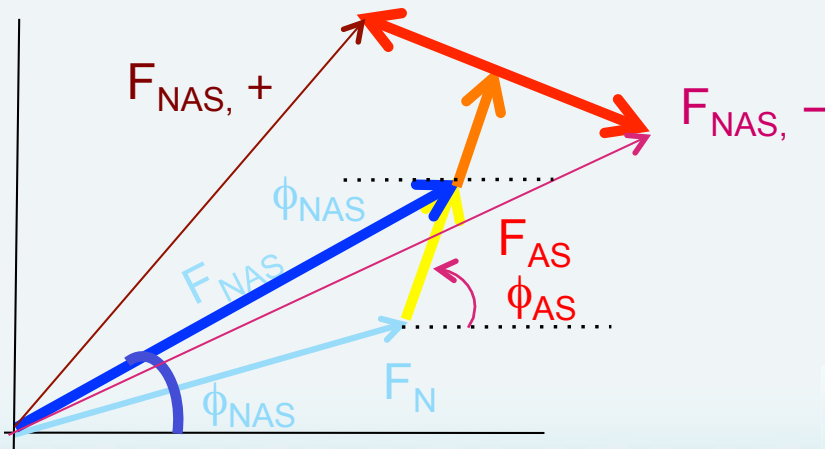
$$\phi_P(+) = -\phi_P(-)$$

$$|F_{PH}(+)| \neq |F_{PH}(-)|$$

$$\phi_{PH}(+) \neq -\phi_{PH}(-)$$

Ce qui aide à résoudre le problème de phase





Pour un type de AS, 3 inconnues :

F_{NAS}

F_{AS}

$$\phi_{NAS} - \phi_{AS} = 180 - \Delta\phi$$

MAD $\longleftrightarrow$ il s'agit de résoudre les équations de Karle/Hendrickson
une λ donnée : $|F_{NAS}(\pm h, \lambda)|^2 = |F_{NAS}(\mathbf{h})|^2 + a(\lambda) |F_{AS}(\mathbf{h})|^2$

$$+ b(\lambda) |F_{NAS}(\mathbf{h})| |F_{AS}(\mathbf{h})| \cos(\phi_{NAS} - \phi_{AS})$$

$$\pm c(\lambda) |F_{NAS}(\mathbf{h})| |F_{AS}(\mathbf{h})| \sin(\phi_{NAS} - \phi_{AS})$$

avec

$$a(\lambda) = (f'^2 + f''^2) / f^{\circ 2}$$

$$b(\lambda) = 2(f' / f^{\circ})$$

$$c(\lambda) = 2(f'' / f^{\circ})$$

$$|F_{NAS}(\pm h, \lambda)|^2 = p1 + a(\lambda) p2 + b(\lambda) p3 + c(\lambda) p4$$

$\longleftrightarrow$ Équation linéaire à 4 inconnues avec $\sin^2 + \cos^2 = 1 \longrightarrow p1 \cdot p2 - p3 \cdot p3 - p4 \cdot p4 = 0$

$$|F_{NAS}(\mathbf{h})| = \sqrt{p1}$$

$$|F_{AS}(\mathbf{h})| = \sqrt{p2}$$

$$\Delta\phi = \arctan(p4/p3)$$

Avec 2 λ , on obtient 4 observables F (+/-)

On calcule ϕ_{AS} pour la structure réelle et son énantiomère, d'où 2 ϕ_{NAS} possibles, on garde la meilleure carte de densité électronique.

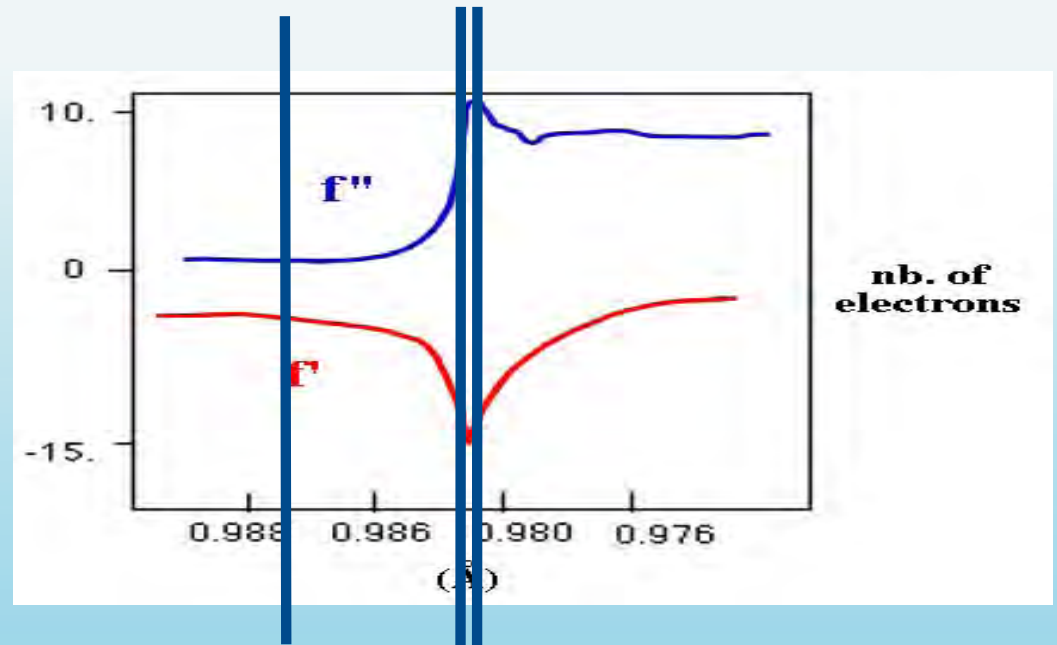
Pour 2 types de AS, 5 inconnues, 3 λ min

Enregistrement MAD, 3 λ au min

- Le signal anomal peut être optimisé par l'enregistrement à une longueur d'onde RX **at or near the absorption edge** du diffuseur anomal.
- Ceci requiert un système de DRX accordable en longueur d'onde (Synchrotron)

Avantages :

- Données à partir d'un seul cristal
isomorphisme *a priori* respecté
- Rapide (≥ 2 jeux λ pour le MAD)
- Carte de densité de suite interprétable



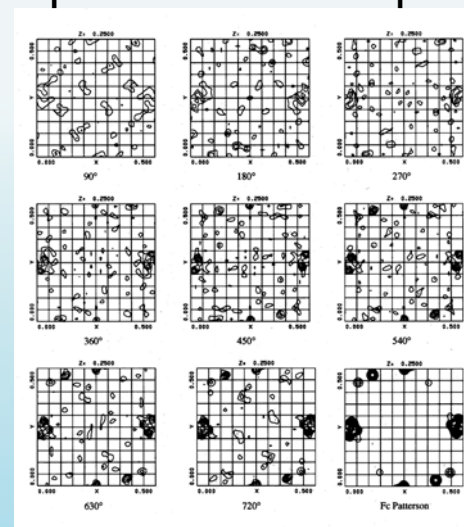
Précautions d'enregistrement

- Privilégier la redondance plutôt que la résol°
- Optimiser les paires de Bijvoet (même image ou méthode faisceau inversé)
- Vérifier radiation damage à la fin (qques images en +)

Single Wavelength Anomalous Scattering (SAS) (ou SAD))

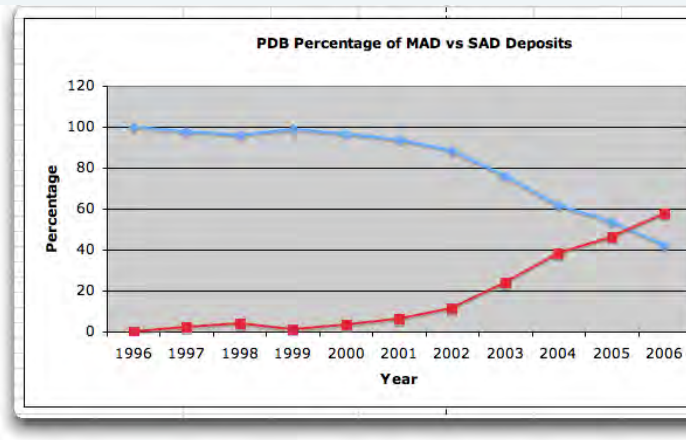
- Un seul enregistrement à une longueur d'onde qui ne coïncide pas forcément au pic d'absorption
(adaptée aux équipements de laboratoire)
- Il faut juste un très bon Signal/Bruit
(forte redondance des données)

ou des diffuseurs anomaux avec $\Delta f''$ élevés
(Fe, Cu, I, Xe, Br, lanthanides, ...)



- Combinaison avec la procédure d'aplanissement de solvant pour lever l'ambiguïté sur la phase (ou la “main”) (cf SIR, ou RIP voir plus loin)
- *Si phases initiales sont de piètres qualités, le % de solvant faible et les sites AS en positions spéciales, échec possible !*
- *Idem si la résolution est basse 3Å ou plus*

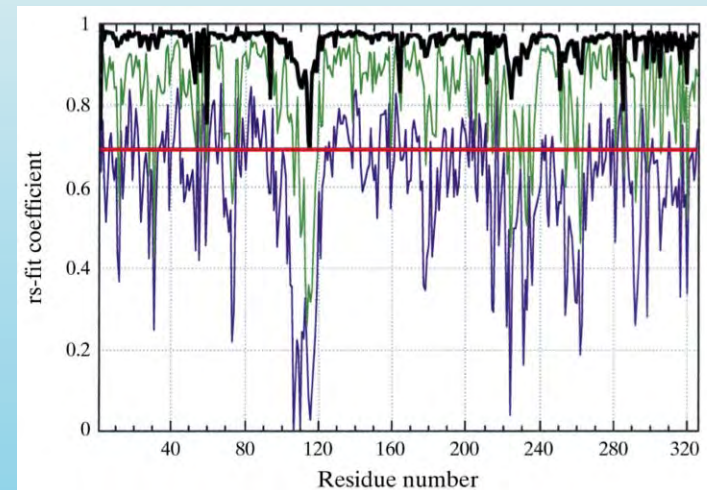
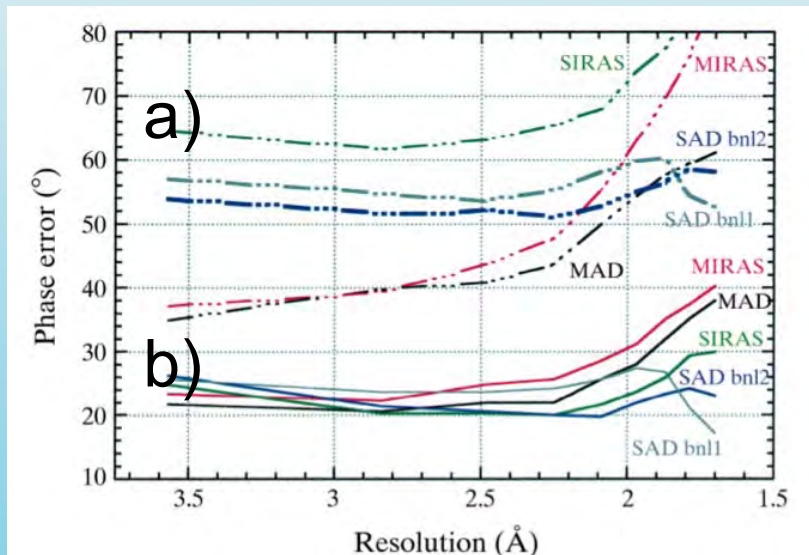
SAD vs MAD



SAD : Gain de temps indéniable / radiation
un seul jeu de données

MAD : Meilleures phases expérimentales ??

Commencer par la λ au pic (SAD) et voir ensuite



— SAD exp.
— Après DM
— Carte finale

Exemple des rscs du modèle final avec les $\neq$ cartes de densité

Etude de la sélénométh. TrpRS + (native)
Erreurs sur les phases (a) expérimentales,
(b) après aplanissement de solvant.

Table 3

Overall real-space correlation coefficients between maps and corresponding refined models at 1.7 Å (at 2.9 Å room temperature).

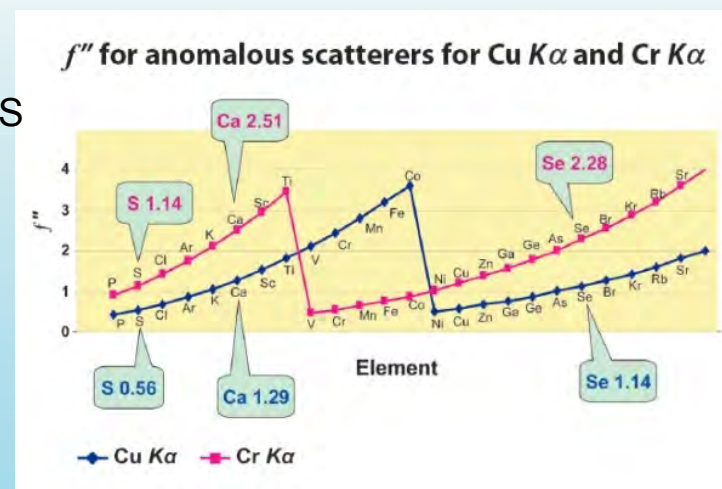
MAPS	MAD (bnl1)	SAD (bnl1)	MIRAS (bnl1-2)	SAD (bnl2, 4 crystals)	SAD (bnl2, 1 crystal)	SIRAS (bnl2)
SHARP	0.772 (0.689)	0.538 (0.442)	0.792 (0.747)	0.644 (0.540)	0.488 (0.413)	0.670 (0.605)
SOLOMON	0.907 (0.810)	0.837 (0.693)	0.895 (0.833)	0.878 (0.752)	0.904 (0.779)	0.909 (0.830)
SIGMAA	0.901 (0.770)	0.880 (0.776)	0.837 (0.702)	0.894 (0.785)	0.896 (0.801)	0.935 (0.804)
BUSTER	0.960 (0.811)	0.926 (0.753)	0.903 (0.743)	0.950 (0.821)	0.949 (0.823)	0.972 (0.820)

Des Sélénométhionines pour le phasage

- Méthionine est relativement rare : 2.4% (vs. Moyenne de 5%)
- Possible de rendre *E.coli* auxotrophe pour méthionine et d'incorporer les sélénométhionines à leur place
- L'incorporation des SeMet est stoichiométrique (occ proche de 1) et covalente, very good!
- Synchrotron beamlines conçues autour de 1Å (f'peak 0.92Å)
- Sensibles aux UV ➡ différence isomorphe utile pour UV-RIPAS

À la méthode Sulfur S-SAS

De l'intérêt des sources de laboratoires



- La plupart des protéines contiennent du soufre en quantité suffisante (pas de modification à prévoir).
- $\Delta f'' = 0.56 \text{ e}^-$ pour Cu X-rays **Haute redondance des données et bonne résolution** (ca 2Å).
- Groupe d'espace de haute symétrie augmente les chances de succès.
- “**soft X-rays**” (Cr $K\alpha$ ($\lambda = 2.2909\text{\AA}$) x2 le signal du soufre ($\Delta f'' = 1.14 \text{ e}^-$) mais problème d'absorption.
- > 20 S-SAS structures dans la Protein Data Bank.

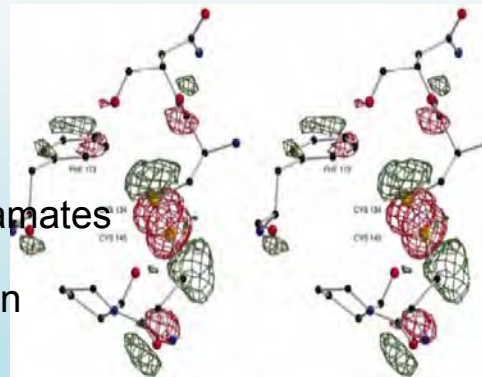
RIP pour Radiation damage Induced Phasing

ou comment détourner un problème expérimental (dégradation d'un échantillon par forte irradiation, radiolyse) en un bénéfice pour la résolution structurale

Ravelli, R. B. G., Schröder Leiros, H.-K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217-224.

Conséquence chimique de l'irradiation

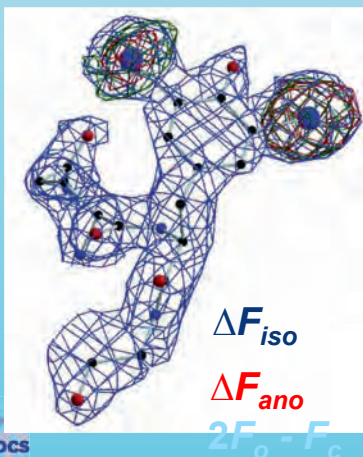
- Cassure des ponts disulfures : modifications locales de chaîne principale ou latérales.
- Décarboxylation des résidus aspartates et glutamates
- Débromination d'acides nucléiques, déiodination de tyrosine iodinée



Exemple de protocole :

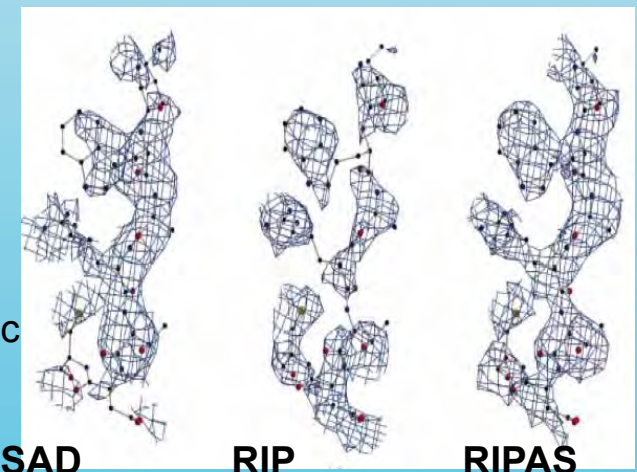
« Two highly redundant 1.4 Å data sets (ESRF beamline ID14-4) with a 24-fold attenuated beam : total absorbed dose of ca $1.3 \cdot 10^6$ Gy per data set. In between the data sets, the rotating crystal was exposed to the unattenuated beam (2 min), i.e. absorbed dose of $1 \cdot 10^7$ Gy. »

RIPAS pour Radiation damage Induced Phasing with Anomalous Scattering :
substructure solution and phasing by P H Zwart, S Banumathi, M Dauter, Z Dauter *Acta Cryst D60* (2004)



Densité électronique expérimentale (avant amélioration des phases) à partir d'un cristal de thaumatococcus trempé dans une solution iodée.

Comparaison de cartes pondérées $2f_o - F_c$
Par différentes méthodes de phasage



Méthodes directes

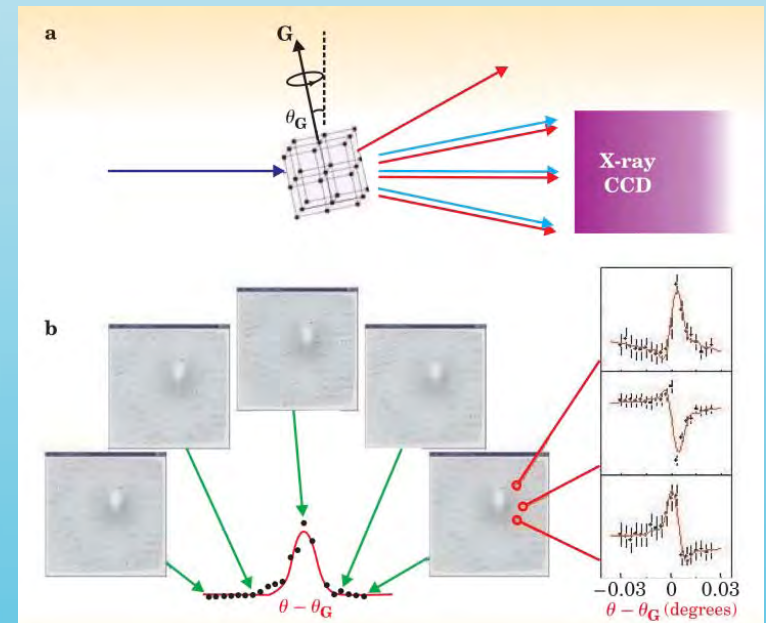
non seulement pour la localisation des atomes lourds mais aussi pour résoudre les structures à la résolution atomique. Les plus puissantes sont basées sur le principe de recyclage « dual direct- and reciprocal-space » comme dans

- SnB (Miller et al., 1994),
- SHELXD (Sheldrick & Gould, 1995)
- ACORN (Foadi et al., 2000).

Three-beam phasing, analogue à l'holographie

Géométrie « reference-beam diffraction (RBD) », Rotation du cristal autour d'un axe (θ_G) correspondant à une forte réflexion G de référence. **(a)** L'interférence de celle-ci avec celles d'un autre plan permet d'extraire l'information sur la phase. **(b)** 5 spectres à différents θ pour illustrer l'intensité de la réflexion G avec un fit Gaussien (rouge). A droite trois profils en fonction de θ pour trois réflexions. De tels profils, on peut obtenir les phases relatives de facteurs de structure ou « **phases triplet** ».

Hic : cristaux et alignements hautement parfaits



Liste non exhaustive de logiciels de phasage

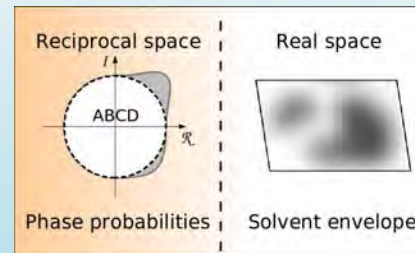
Implémentant les bases de la théorie du phasage par atomes lourds (Green et al., 1954; Blow & Crick, 1959; Rossmann, 1961; Matthews, 1966)
avec des approches notamment basées sur le **principe de maximum de ressemblance**

- MLPHARE (Otwinowski, 1991)
- **SHARP (de La Fortelle & Bricogne, 1997),**
- SOLVE (Terwilliger & Berendzen, 1999)
- PHASES (Furey & Swaminathan, 1997)
- **SIR2004 (Burla et al., 2005)**
- CRANK (Ness et al., 2004)

Méthodes d'amélioration des phases avant modèle atomique

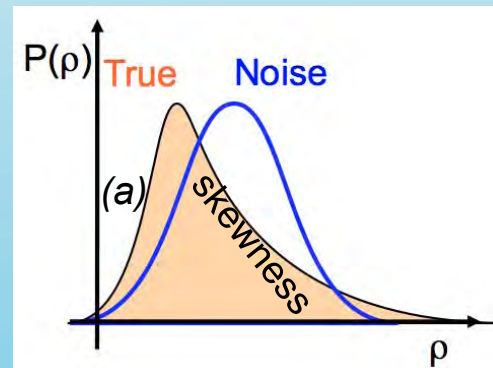
Modifications de densité électronique :
basées sur l'enveloppe de la protéine, tout ce qui n'est pas solvant est protéine et densité électronique ≥ 0

- Aplatissement de solvant

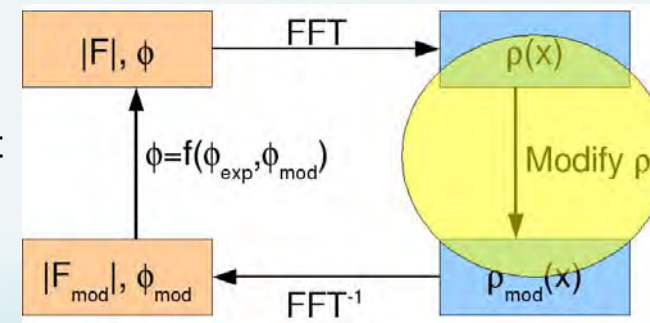


- Histogram matching

ρ indépendante de la protéine pour une résolution donnée (a) distribution représentative d'une map bien phasée, (b) bruitée (gaussienne)



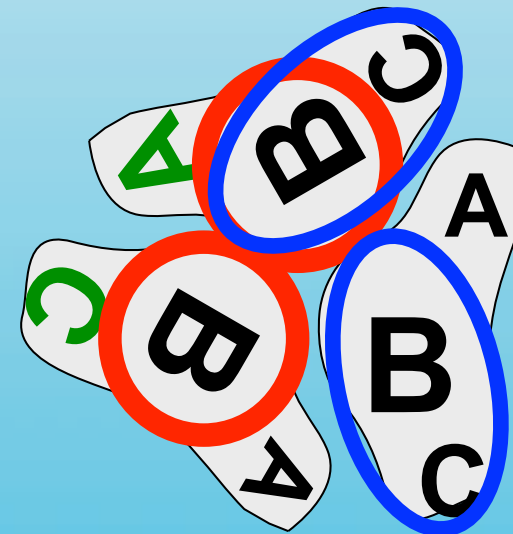
- NCS averaging (si plusieurs copies ≥ 4)
Très puissant avec DM



ISIR/ISAS program suite (Wang, 1985)

DM and PHASES, SOLOMON

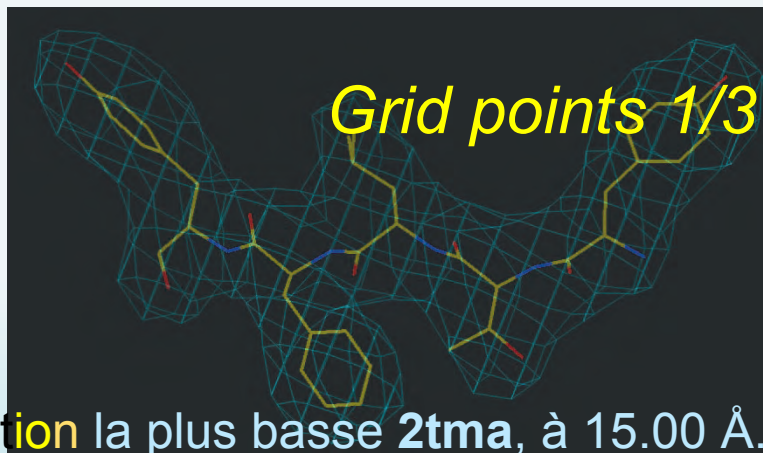
PARROT using a likelihood function which incorporates the experimental phase information as a prior.



« The best source of information about the quality of your model is your maps ! »

Dale Tronrud

Résolution carte de densité et affinement du modèle

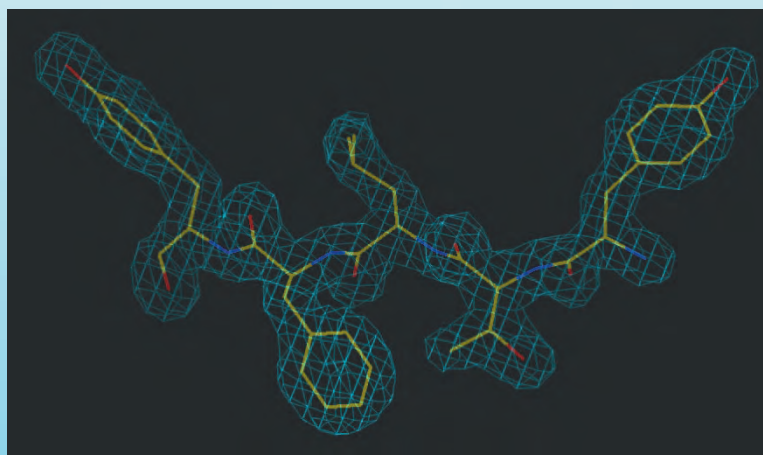
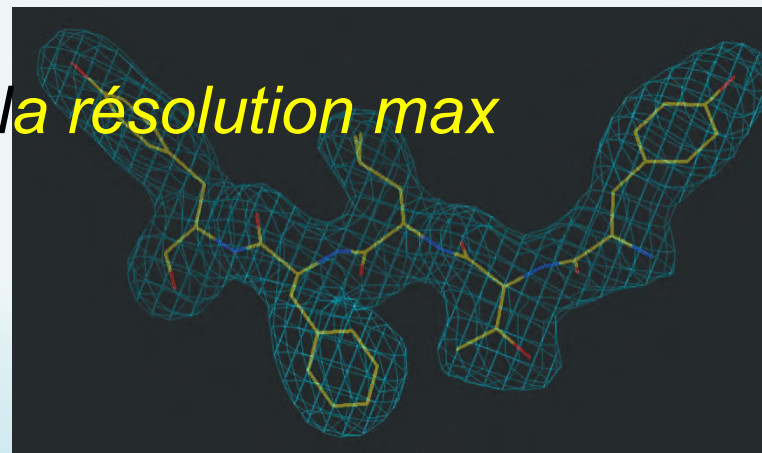


Grid points 1/3 de la résolution max

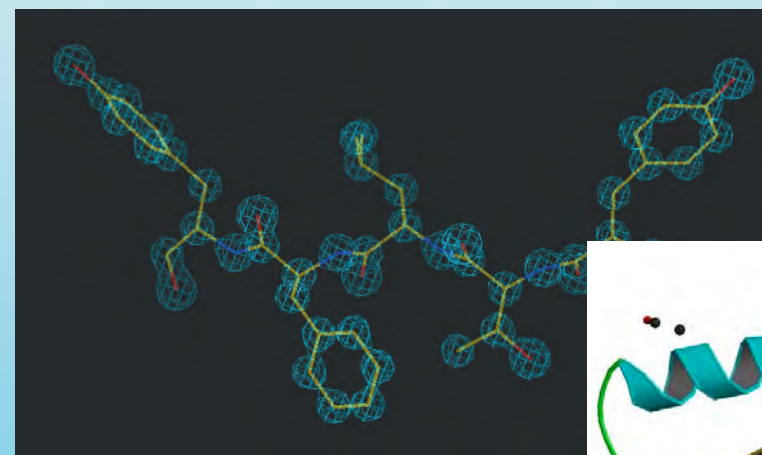
Résolution la plus basse 2tma, à 15.00 Å.

TROPOMYOSIN CRYSTAL STRUCTURE AND MUSCLE REGULATION.

CONSTRUCTION OF AN ATOMIC MODEL FOR TROPOMYOSIN AND IMPLICATIONS FOR INTERACTIONS WITH ACTIN



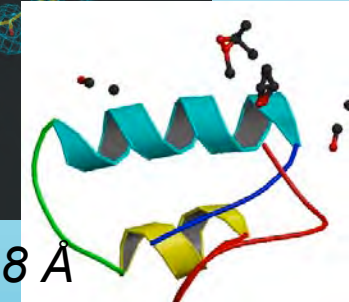
Résolution moyenne dans la PDB
2.1 Å



Résolution la plus haute 0.48 Å

Crambin (48 aa soit 5,12 kDa) 100 K (PETRA II)

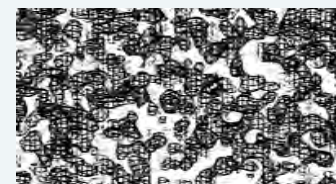
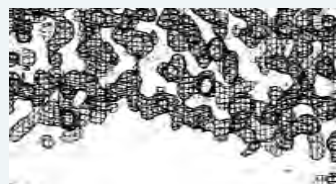
Schmidt A, et al, Acta F. 2011 (Pt, 4):424-8.



Space Group: P 1 21 1

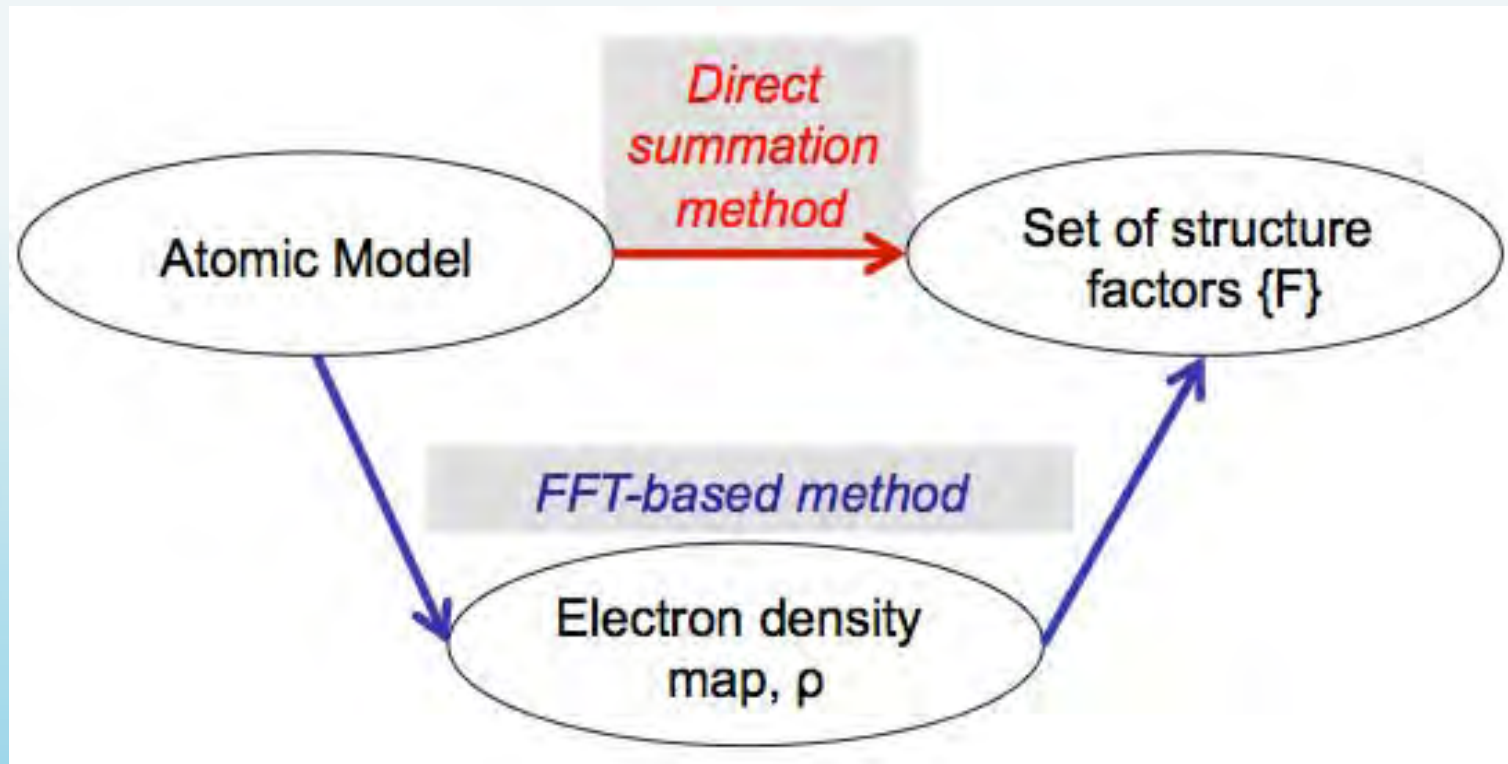
R_factor 0.127 R_Free NULL (MOPRO)

Evaluer la carte de densité électronique



<i>Base</i>	<i>Good map</i>	<i>Random map</i>
Skew of density (Podjarny, 1977)	Highly skewed (very positive at positions of atoms, zero elsewhere)	Gaussian histogram
Connectivity of regions of high density (Baker, Krukowski, & Agard, 1993)	A few connected regions can trace entire molecule	Many very short connected regions
Correlation of local rms densities (Terwilliger, 1999)	Neighboring regions in map have similar rms densities	Map has uniform rms density
R-factor in 1 st cycle of density modification (Cowtan, 1996)	Low R-factor	High R-factor

Deux façons de calculer les Facteurs de structures



La FastFourierTransform est plus rapide (ρ échantillonnée dans une sphère de $R \sim 2\text{\AA}$ autour des at.)

Temps de calcul $\sim$ calcul de densité + FFT $\sim K_{\text{grid}} \cdot (V_{\text{atom}}/V_{\text{crystal}}) + K_{\text{grid}} \cdot \ln(K_{\text{grid}})$,

where $K_{\text{grid}} = N_x N_y N_z$

Structure à haute résolution utilise la **sommation complète** 'exacte valeur pour F'

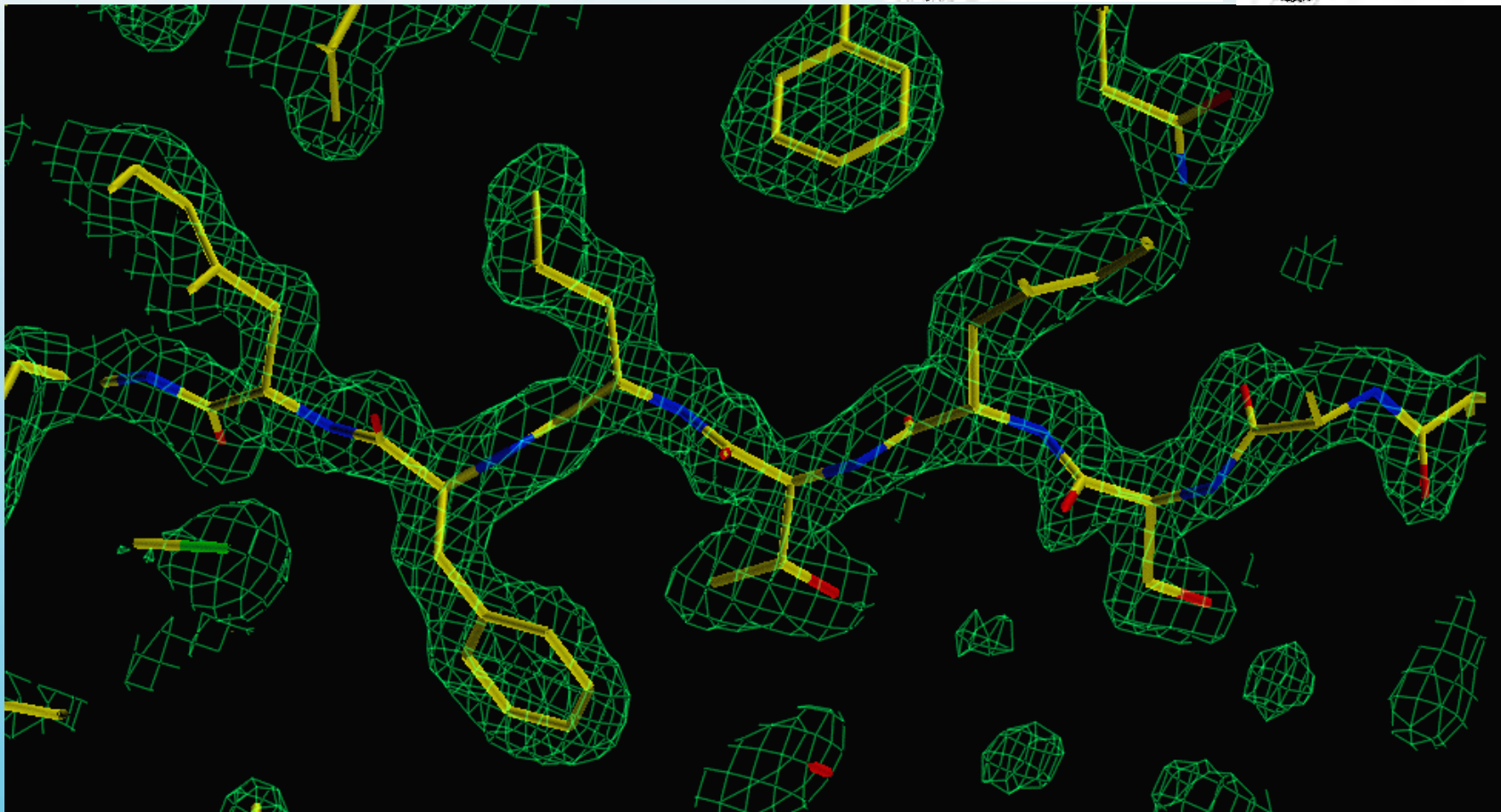
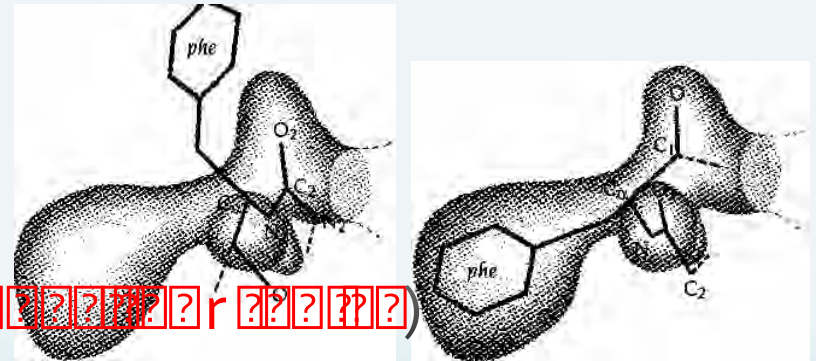
Temps de calcul $\sim$ nombre de réflexions $\cdot$ nombre d'atomes

$$\frac{4}{3} \pi \frac{V_{\text{cristal}}}{d^3}$$

Construire le modèle dans la densité électronique

- Trouver les éléments structuraux:
Hélices α , feuillets β

- Y faire entrer la séquence (???? ???? ???? ???? ???? r ???? ????)

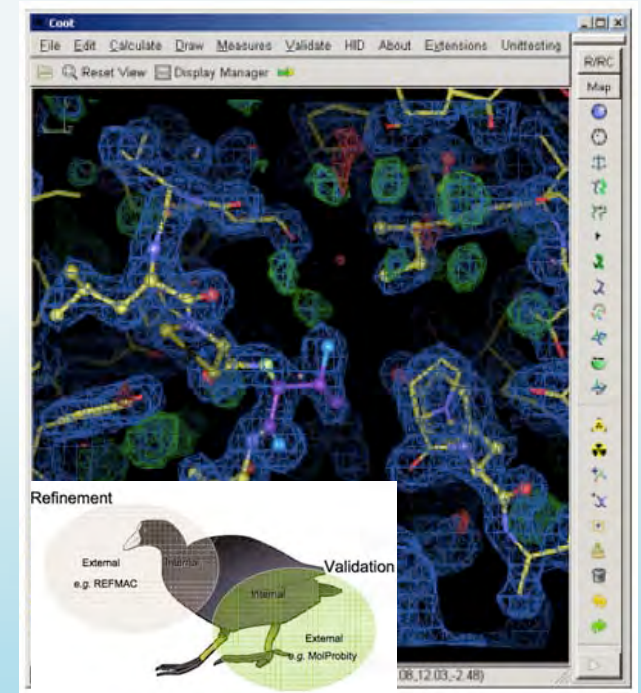


Affinement du modèle cristallographique

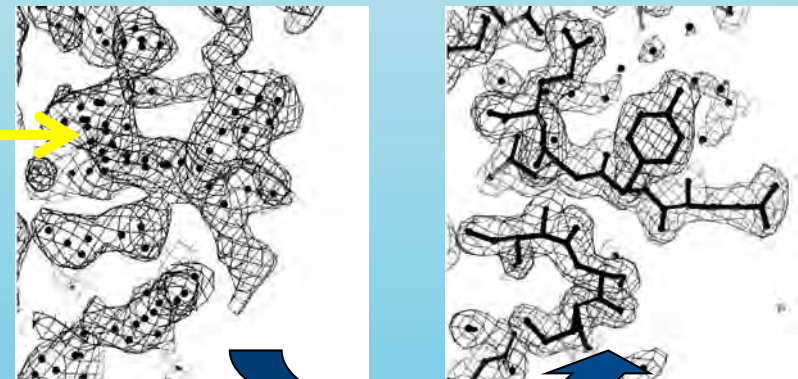
un processus itératif

- **Bad** phases
 - **poor** electron density map
 - **errors** in the protein model
- Interpretation of the electron density map
 - **improved** model (x, y, z B)
 - **improved** phases
 - **improved** map
 - even **better** model

Meilleure résolution, plus de détails sur le modèle

[illegible]

Quid de l'information (densité et fragments de modèle) manquante ?



ARP/wARP (Perrakis et al., 1999),
RESOLVE (Terwilliger, 2000)
MAID (Levitt, 2001)).

Affinement cristallographique ?

Méthodes de Gradient-descent:

?? ?B ?
PD? Pcd? ?B ?m??? ? ?e? ???e? ??? e? ??
??C? ?B ?
??B?D?PB??P???B?
?? ??B?C?t ?t ??
?? ??B?C?t ?t ?D ?C?MC? ?r???



Optimisation

$$T = T_{DONNEES}(F_{OBS}, F_{MODELE}) + wT_{RESTRAINTS}$$

d'**observables** en changeant les **paramètres d'un modèle** 

[illegible]

Engh & Huber,
quantum-chemical
calculations

Real space target
Energie empirique
Moindres carrés

$$f(p) = \sum_i^{toutes les données} [Q_{obs}(i) - Q_{calc}(i, p)]^2 / \sigma_{obs}(i)^2$$

Maximum de vraisemblance

$$f(p) = \sum_i^{toutes les données} \frac{[Q_{obs}(i) - \langle Q_{calc}(i, p) \rangle]^2}{[\sigma_{obs}(i)^2 + \sigma_{calc}(i, p)^2]}$$

d'une **fonction**

Nombre = 4 ou 9

x, y, z	3
Occ	fixée à 1
B iso	1
ou aniso	6

Simplification

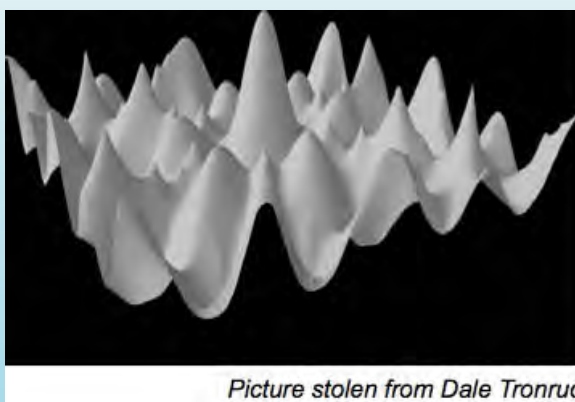
Les tous premiers modèles de structures RX n'étaient pas affinés du tout !

- 1971 Premières tentatives à partir des pics des cartes de différences de Fourier $F_o - F_c$
- 1972 Affinement par moindres carrés sur la structure de rubredoxin par le groupe de Seattle (Watenpaugh et al., 1972, 1980).
- 1978 Approche complétée par une régularisation géométrique (Agarwal, 1978; Isaacs & Agarwal, 1978).
- 1980 Restrictions géométriques incluses dans la méthode des moindres carrés dans PROLSQ (Konnert, 1976; & Hendrickson, 1980) et autres programmes (TNT; Tronrud et al 1987).
- 1987 Restrictions géométriques exprimées en termes énergétiques X-PLOR (Brünger et 1987) puis CNS (1998)
- 1997 SHELXL (Sheldrick & Schneider, 1997) issu des petites molécules et transposé aux macromolécules.
- 1997 Affinement basé sur les principes de maximum de ressemblance dans REFMAC et al., 1997) or BUSTER (Bricogne & Irwin, 1996)
- Les améliorations outils de validation Rfree (Brünger, 1992), estimation sur les incertitudes de position (Cruickshank, 1999), « merohedral twinning » (SHELXL and CNS), « TLS rigid fragment refinement » (in REFMAC5; Winn et al., 2001) et « torsion-angle restraints » (in CNS).

Le hic lié aux macromolécules?

L'affinement d'un modèle de macromolécules est une procédure non linéaire et difficile à paramétrer (avec de multiples minima locaux) et en général avec peu de données par rapport aux paramètres

ex: protéine 25kDa, à 2Å, 22000 F_{obs} Données /paramètres = 2^{iso} ou 1^{aniso},

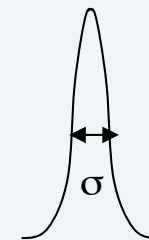


Améliorer un modèle atomique rendant compte de l'expérience et de ses erreurs se doit d'intégrer les données stéréochimiques alias « prior knowledge » $\longleftrightarrow$ chimie pour éviter le risque de sur-affinement (overfitting) et de biaiser le modèle.

But de la PX : tester une hypothèse biologique et non la confirmation ultime de cette hypothèse (paradigme bayésien) $\longrightarrow$

Les Observables :

Jauger le modèle avec $/F_{obs}/$, Phases expérimentales, dist./angles de liaisons, chiralité**... suivant des fonctions continues (pour la dérivée) accompagnées d'une indication d'incertitude (grande pour les phases par ex.)



(** exprimée à travers le signe d'une grandeur continue comme les angles dièdres impropres)

+ Restrictions stéréochimiques

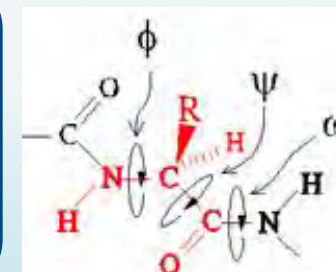
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100



Observables indép.

Paramètres = favorable

⇒ Plus aisé de réduire le nombre de Paramètres

-> B isotropes

même si atomes dans un cristal présentent une forte anisotropie ! + XYZ = paramètres atomiques

-> Occupation = 1

raisonnable pour la macromolécule / moins pour les molécules de solvant

-> affinement « corps rigide » 6 paramètres de rotation et de location (modèle préliminaire)

-> idem quand plusieurs molécules dans l'unité asymétrique reliées par NCS
(restrictions plutôt que contraintes pour permettre les ≠ entre copies)

-> modèle de liaisons/angles 5x moins de paramètres que pour un modèle atomique
(affinement compensatoire pour éviter la propagation des variations sur un paramètre donné)

Basse résolution
4Å

-> TLS B (Translation, Librations, Screw) (20 paramètres pour un groupe d'atomes ex: chaînes latérales His, Phe, Tyr
(balance à faire entre le nombre de groupes à définir et le gain en réduction de N_{paramètres})

Les fonctions

Real Space Refinement

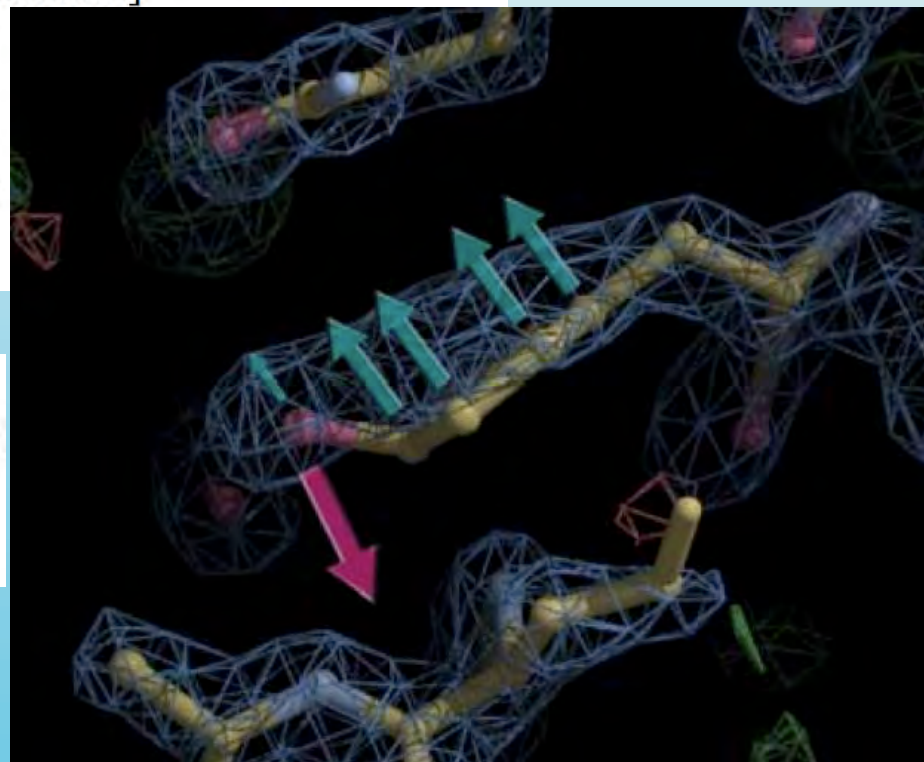
Diamond, R. (1971). *Acta Cryst. A*
27, 436-452.

- Major feature of Coot
 - Gradient minimizer (BFGS derivative)
 - Based on mmCIF standard dictionary
 - Minimizing bonds, angles, planes, non-bonded contacts, torsions, [chiral volumes, Ramachandran]
- Provides “interactive refinement”
- Different minimizer to Refmac...
 - ...means “nice & tight” geometry
 - Chi squareds

▪ Real space target

$$T_{\text{DATA}} = \sum_{\text{grid points}} (\rho_{\text{best}} - k\rho_{\text{calc}})^2$$

+ restrictions stéréochimiques
expérimental, 2mFo-DFc



Les fonctions (suite)

① **Energie empirique** ($\neq E$ liaisons ou stabilité) $\longleftrightarrow$ modèle de plus basse E, ΔE kcal.mol⁻¹

$$E = E_{chimie} + \omega_{RX} E_{RX}$$

$$E_{chimie} = \sum_{liaisons} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} k_\Phi \cos(n\Phi + d) + \sum_{chiralité, planarité} k_\omega (\omega - \omega_0)^2 + \sum_{paires d'atomes} (ar^{-12} + br^{-6} + cr^{-1})$$

$$E_{RX} = \sum_n \left[|F_{obs}(h)| - k |F_{calc}(h)| \right]^2 + restriction de phase$$

Description de E pas assez précise et théorie statistique manquante.

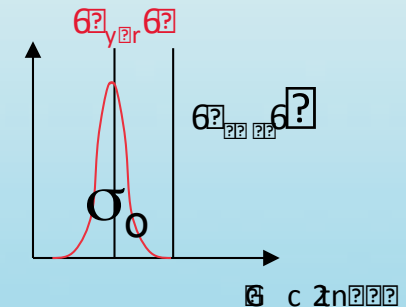
② **Moindres carrés:** $f(p) = \sum_i \frac{[Q_o(i) - Q_c(i, p)]^2}{\sigma_o(i)^2}$ à minimiser

Le meilleur jeu de paramètres donnés (modèle) sans erreur est recherché pour modéliser des observables obéissant à une distribution normale.

$\varphi_{obs} = \varphi_{calc}$ Pas d'incertitude sur la phase

- Un modèle incomplet ne peut reproduire entièrement les observables.

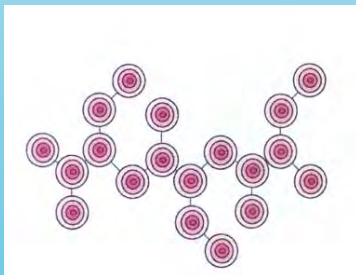
Ne prend pas en compte la non-linéarité de la relation entre les paramètres d'un modèle et les observables



③ **Maximum de Vraisemblance** ($\leq$ Bayesian Paradigm)

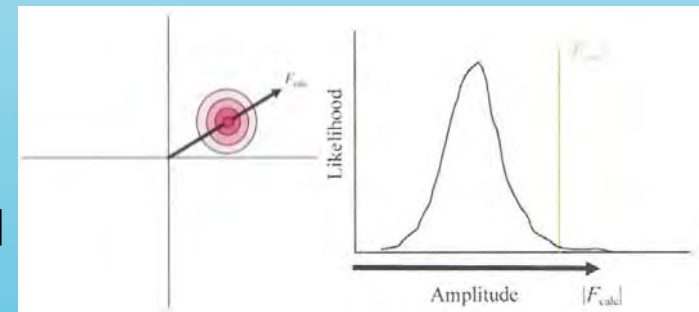
Raisonnement en termes de distribution de probabilités conditionnelles

Le meilleur modèle est celui le plus en accord avec les observables



Erreurs supposées dans le modèle

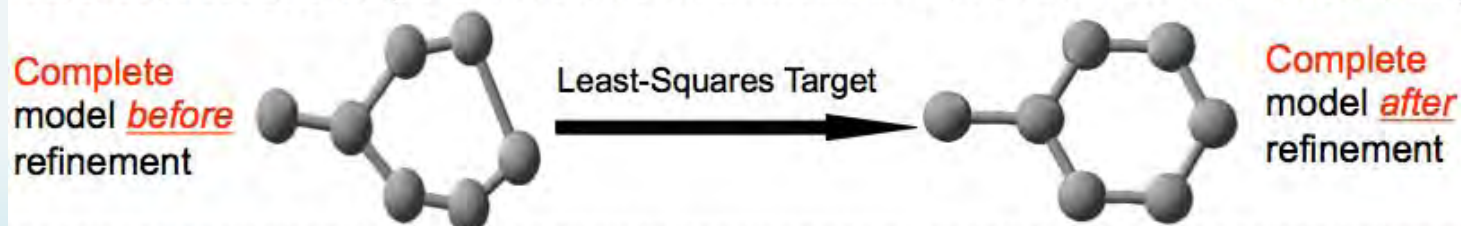
Impactent la prédiction sur $|F_{calc}|$



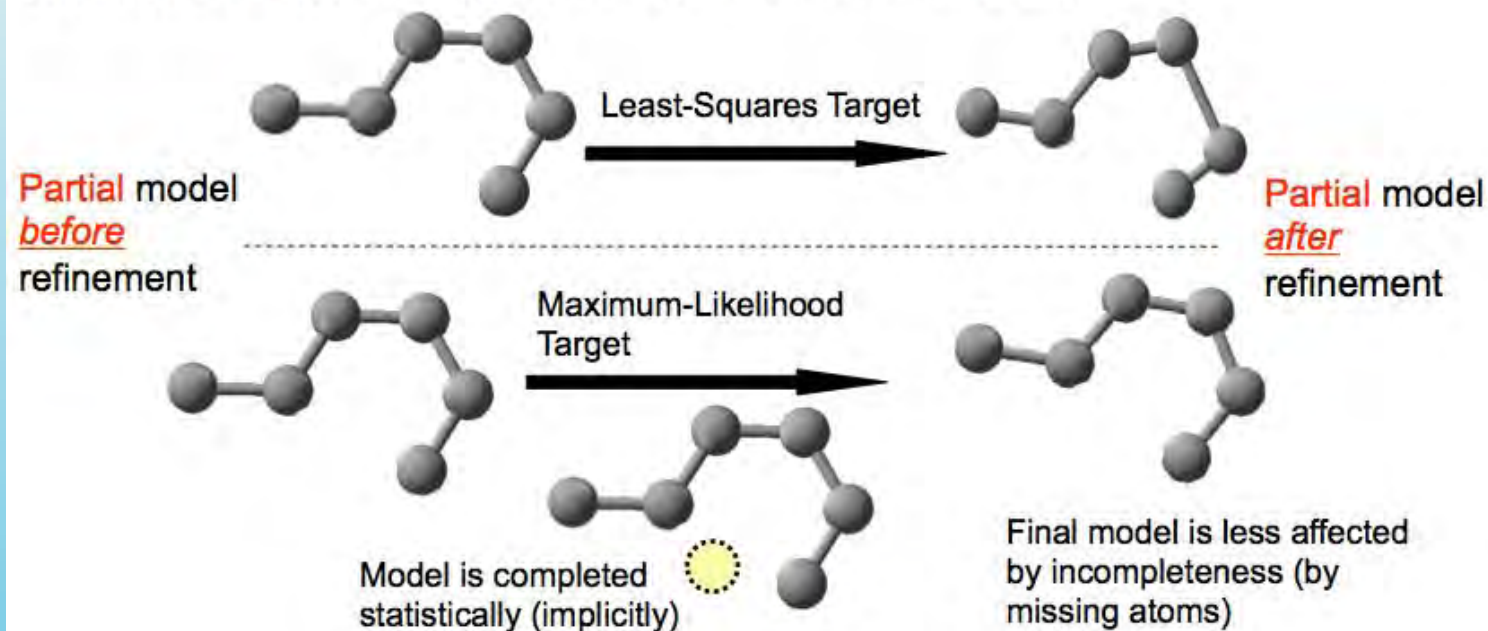
ML vs LS

Maximum de vraisemblance vs moindres carrés

- **Removable Errors** (never the case for macromolecular model, common for small molecules)



- **Irremovable Errors** (always the case for macromolecular models)



L'inférence inductive bayésienne

« On nomme **inférence bayésienne (Bayes, 1763)** la démarche logique permettant de calculer ou réviser la probabilité d'un événement. Cette démarche est régie en particulier par le théorème de Bayes, (1763) la probabilité n'est pas interprétée comme la simple traduction numérique d'un état de connaissance mais comme le degré de confiance accordé à une hypothèse. »

Raisonnement scientifique plausible qui s'applique à beaucoup de démarches rationnelles et notamment aux modèles de protéines (*Formalisme par G. Bricogne*) que l'on confronte aux données expérimentales et aux connaissances a priori $p_1, p_2, \dots$ probabilités indépendantes :

Le cadre bayésien est particulièrement adapté au traitement et à la propagation des incertitudes.

La formule de Bayes

donne la distribution de probabilité a posteriori du modèle M conditionnellement aux données observées D

$$L_M(M | D) \propto L_D(D | M) \times p_1(M) \times p_2(M) \times p_3(M) \dots$$


Fonction de vraisemblance


a priori

Buster (AutoBuster) Global Phasing Limited – TNT

Forme le modèle structural
Met à l'échelle Fobs et Fcalc
Calcule la fonction de vraisemblance

Stereochimie et NCS
Shift les paramètres du modèle

le modèle structural a trois composants

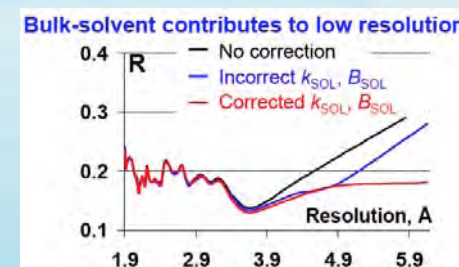
$$\rho_{\text{calc}}(x) = \rho_{\text{fragment}}(x) + \rho_{\text{missing}}(x) + \rho_{\text{solvant}}(x)$$

avec ses erreurs pour donner lieu à une distribution statistique de probabilité de

$$F_{\text{calc}} = F_{\text{fragment}}(x) + F_{\text{missing}}(x) + F_{\text{solvant}}(x)$$

Affinement du modèle partiel (fragment) par maximisation de vraisemblance tout en décrivant la partie manquante par des méthodes de maximum d'entropie directes /

Paramétrisation sophistiquée qui offre la possibilité d'affiner un modèle incomplet (10 à 45%) en présence d'un modèle basse résolution de la structure manquante!



Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Cryst.*, **D60**, 2210-2221

Les méthodes de minimisation de fonction

- La plus simple: **méthode de recherches** calcule la valeur de la fonction pour un set de paramètres p

ex: **Recuit simulé** (rayon de convergence 👉, CPU 🕒):

```

cyr22rrnr129r2n2121 t??222n22pnl 3y 2221ynt2r222tp e2n 2r3lp p 22| t??222ty22G 2| t? nt2t22n2223n22p2py 2222| t2G 2| t2cynp2n1y 2r3lp p 22| t?
nt2t??2222t nr?? 2

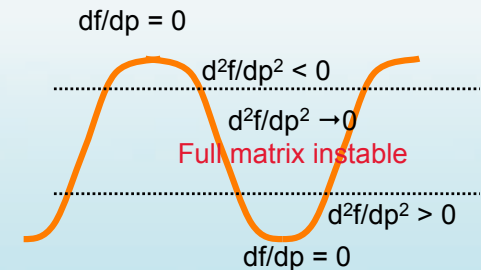
```

- Les autres requièrent en plus **les dérivées 1^{ières} et 2^{ndes}** de la fonction développée en série de Taylor

$$f(p) = f(p_0) + \left. \frac{df(p)}{dp} \right|_{p=p_0} (p - p_0) + \frac{1}{2} (p - p_0)^2 \left. \frac{d^2 f(p)}{dp^2} \right|_{p=p_0} + \dots$$

« shift vector » $s = p - p_0$
 obtenu entre modèle présent et minimum $\left. \frac{df(p)}{dp} \right|_{p=p_0+s} = 0 \Rightarrow s = - \left. \frac{d^2 f(p)}{dp^2} \right|_{p=p_0}^{-1} \left. \frac{df(p)}{dp} \right|_{p=p_0}$

Full-Matrix $\longrightarrow$ $sC_{\alpha} y_{nl} \sim \begin{bmatrix} \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \end{bmatrix} \begin{bmatrix} \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \end{bmatrix} \begin{bmatrix} \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha \end{bmatrix}$



Termes d'ordres supérieurs ignorés, s donne un minimum approché et un rayon de convergence moindre
 p_0 doit être suffisamment proche du minimum

Normal Matrix

$$\left(\begin{array}{ccc} \frac{\partial^2 f(p)}{\partial p_1^2} & \cdots & \frac{\partial^2 f(p)}{\partial p_n \partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(p)}{\partial p_1 \partial p_n} & \cdots & \frac{\partial^2 f(p)}{\partial^2 p_n} \end{array} \right)^{-1}$$

10 000 x 10 000 termes

= 100 millions

Simplificat°

Diagonale affectée par 1 paramètre

Hors-diagonale corrélation de 2 paramètres

Spare Matrix

Diagonale (incluant contributions à partir des restrict° cristallographiques et stéréochimiques)
+ **hors diagonale** pour les paramètres d'atomes connectés par restrictions géométriques.

Diagonal Matrix

Diagonale uniquement

Steepest Descent

Tous les termes diagonaux ont la même valeur

Gradient conjugué

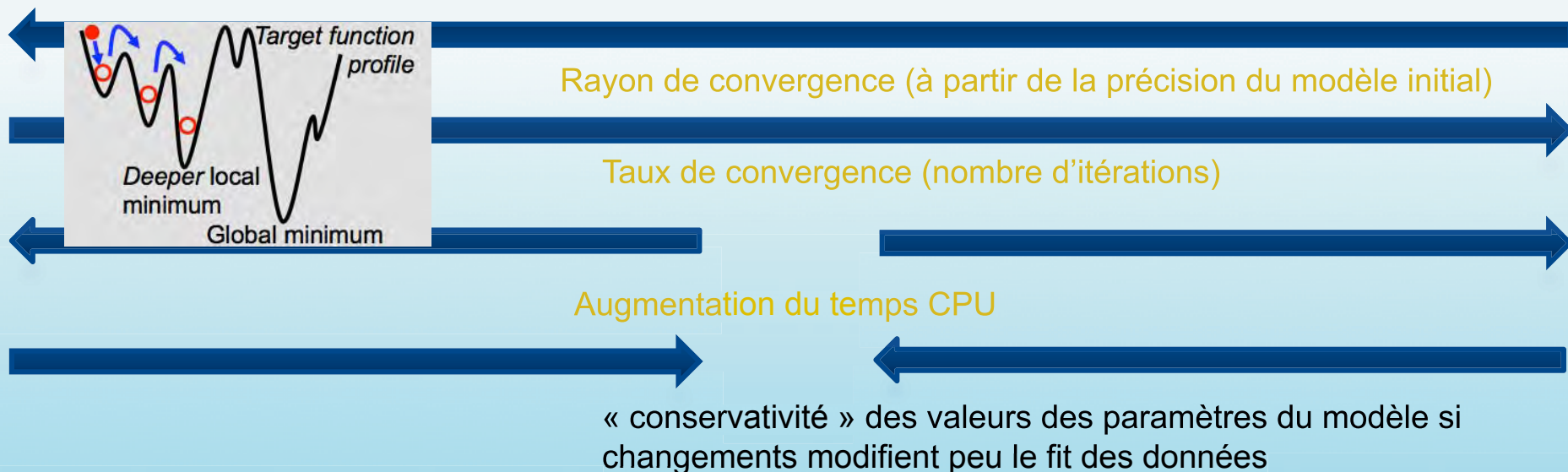
Leurs caractéristiques

Pas dérivée

1ère dérivée

2nd dérivées

Recherche recuit simulé Steepest Descent Gradient conjugué Gradient conjugué préconditionné Full -matrix



Program	Parameters	Function	Method
BUSTER/TNT	xyzb	ML, ML φ , ML?	PCG
CNS	xyzb, torsion	EE, LS, ML, ML φ	SA, CG
REFMAC	xyzb, TLS, aniso	LS, ML, ML φ	Sparse, FM
SHELXL	xyzb, aniso, free	LS	Sparse, FM
TNT	xyzb	LS	PCG
X-PLOR	xyzb, torsion	EE, LS, ML, ML φ	SA, CG

Des modèles à basse résolution...

Quelle que soit la résolution, les propriétés topologiques accessibles restent utiles à déterminer

Limite de résolution (Å)	Carte de densité	Affinement
8 - 5	Domaines visibles	En corps rigide.
5 - 4	Chaîne principale traçable / latérales parfois visibles. Régions hélicales repérables. Petites molécules visibles e.g. ADP vs ATP	Restrains sur les structures secondaires, fonction MLHL comme cible. R/Rfree et géométrie acceptable
4 - 3.5	Chaîne principale sûre, latérales plus visibles Feuillets bêta également	Conventionnel

- Privilégier le système dans sa globalité plutôt que « *Divide-and-conquer* » (fragments 3D haute résolution puis assemblage avec microscopie cryo-électronique ou SAXS..., pour déterminer l'enveloppe moléculaire.

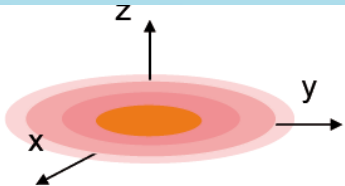
Plus que jamais à cette résolution faible, Expertise humaine + R_{free} , structures homologues, prédict° de struc.sec., posit° atomiques connues (connaissance a priori) sont essentielles

- Remplacement moléculaire et real-space fitting (régions hélicales en particulier)
- Information sur la phase de type MAD (e.g. SeMet) plutôt que SAD (DM plus aléatoire à **BR**)
 - Utiliser toutes les données mêmes faibles $I/\sigma(I) < 1$
- B-factor sharpening (pseudo mise à l'échelle de type Wilson des données de diffraction)
 - Modèle du Bulk-solvent *ad hoc* (voir CNS v1.1)
- Structure restraints, (e.g. ϕ/ψ), NCS fondamentalement utiles car le *nb param* > *nb de réflexions*
 - Multi-crystal averaging
- B facteur atomique à améliorer / technique de reconnaissance de motifs.

... aux modèles haute résolution

Key resolution limits and corresponding features

High res., Å	Size of details	Features of images	Mean Nref/at
~0.7	half of a X-X bond	deformation density	140
~0.9	X-H bond	some deformation density	75
~1.0		H atoms	60
~1.2	shortest covalent bond	individual atoms	40
~1.5	C _α -C		25

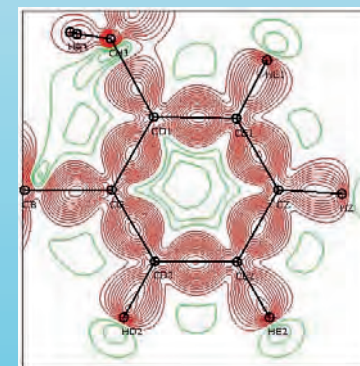


$\rho(\Delta\mathbf{r}) \sim \exp\{-\Delta\mathbf{r} \cdot \mathbf{U}^{-1}\Delta\mathbf{r}\}$

$$\mathbf{U} = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{bmatrix}$$

SHELXL pour ADP

MOPRO *déformation de densité...*



$$\Delta\rho = [P_{\text{val}} \kappa^3 \rho_{\text{val}}(\kappa r) - N_{\text{val}} \rho_{\text{val}}(r)] + \sum_l \kappa^3 R_l(\kappa r) \cdot \sum_m P_{lm} y_{lm}(\theta, \phi)$$

Ou comment relier le monde des petites aux macromolécules

86 **Suite SHELX**^{mx} 97

^{mx} pour « macromolecular crystallography ».

SHELXS résolution structurale par Patterson & Méthodes Directes

SHELX^{mx} **C/D/E** la triade pour le phasage expérimental

SHELXC préparation des données

SHELXD à l'origine pour les petites molécules par (MD), maintenant pour les solutions des sous-structures pour le phasage expérimental (sans modification de programme).

SHELXE modification de densité et auto-traçage

SHELXL^{mx} pour l'affinement haute résolution et données neutron

SHELXPRO^{mx} Collection of conversion utilities (e.g. PDB ! .ins & .res ! PDB) incluant les «standard restraints » pour les structures des peptides et acides nucléiques; création des cartes pour O et XtalView (utile pour COOT)

CIFTAB création de tableaux à partir des CIF pour publication

SHELXWAT^{mx} recherche automatique des molécules de solvant (dépassé par COOT)

Idem avec SIR 92 SIR 2008 <http://www.ic.cnr.it/>

Offres concurrentes

SHARP globalphasing.com

SOLVE <https://solve.lanl.gov>

BP3 bfsc.leidenuniv.nl/software/bp3

SnB hwi.buffalo.edu/snb,

...

DM,PARROT ysbl.york.ac.uk/~cowtan

BUCCANNEER ysbl.york.ac.uk/~cowtan/

RESOLVE <https://solve.lanl.gov>,

...

REFMAC5 ysbl.york.ac.uk/~garib/refmac/

PHENIX phenix-online.org

BUSTER globalphasing.com

TNT uoxray.uoregon.edu/tnt/

...

AVANTAGES liés à SHELXL-97 ^{mx}

(prochaine version 2012)

Philosophie conservée : programme en lignes de commandes, .ins + HKLF 4, 5(twin) ou 2(Laue)

- ✓ Enorme flexibilité pour traiter des données à partir de Laue, Neutron,... **domaines multiples**)
- ✓ Affinement contre les intensités (incluant les $I < 0$)
- ✓ Utilise calcul conventionnel des F plutôt que FFT sommation (défavorable pour dispersion ou anisotropie)
lenteur compensée par Version parallélisée (shelxl mp)
- ✓ Grand rayon de convergence

- ✓ Estimations des incertitudes standard (esds)

- ✓ Traitement général symétrie

- ✓ **Affinement anisotropique (pour $d \leq 1.5 \text{ \AA}$)**

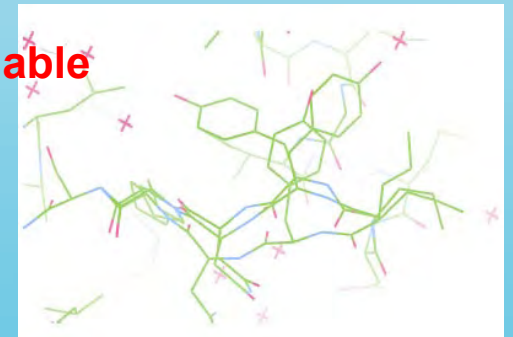
$$F(hkl) = \sum_{\substack{\text{atoms } j \\ \text{in unit cell}}} f_j(\theta_{hkl}) e^{-2\pi^2(hkl) \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix} \begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}} e^{2\pi i(hx_j + ky_j + lz_j)}$$

- ✓ Syntaxe de commandes évidente DFIX, FLAT, SADI, CHIV, SUMP, FVAR, SIMU, DELU, ISOR qui convient très bien aux protéines

- ✓ **Coordonnées fractionnées + U au lieu de B** $\longleftrightarrow$ **Free VARIable**

Affinement du taux d'occupation (études des ligands)

Traitement flexible du désordre compliqué PART + FVAR



Exemple d'affinement haute résolution avec SHELXL

Mieux vaut commencer la reconstruction ou l'affinement avec **PHENIX ou REFMAC**
Avant de finir avec SHELXL (interaction avec COOT)

Si solution MR,

En corps rigides BLOC et AFIX 6 ... AFIX 0
Restrictions sur les charnières
SHEL 8 3

L.S. (matrice complète)

Etendre ensuite la résolution puis passer à **Conjugate Gradient LeastSquares CGLS**

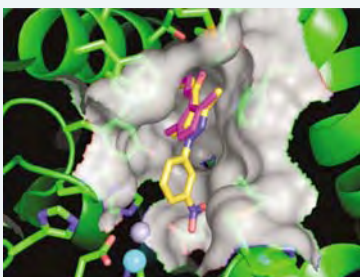
Affinement contraint et restreint dès le début pour des densités interprétables

Table 1. SHELXL refinement of rubredoxin against 0.92Å synchrotron data. 24770 reflections with $F > 4\sigma$ were used for R_1 , and 1350 for R_{free} .

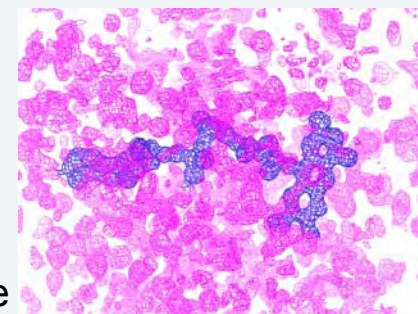
Job	Action taken	N(parm.)	N(restr.)	R_1 %	R_{free} %
1	PDB 7rxn, no water	1	1544	22.9	23.4
2	+38 waters	1734	1554	16.3	18.6
3	+14 more waters	1790	1563	15.8	18.7
4	+8 more waters	1822	1567	15.7	18.7
5	Fe,S anisotropic	1857	1597	14.8	17.7
6	+all H-atoms	1857	1600	14.0	16.8
7	all C,N,O anisotropic	4097	4888	8.8	11.3
8	+28 more waters	4358	5086	7.8	10.7
9	water $\rightarrow$ 79(1)+18(1/2)	4556	5291	7.5	10.3
10	6 disordered sidechains	4698	5530	6.9	9.7

≈ 400 at.

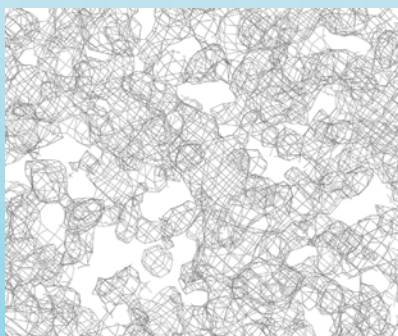
Les complexes avec les petites molécules



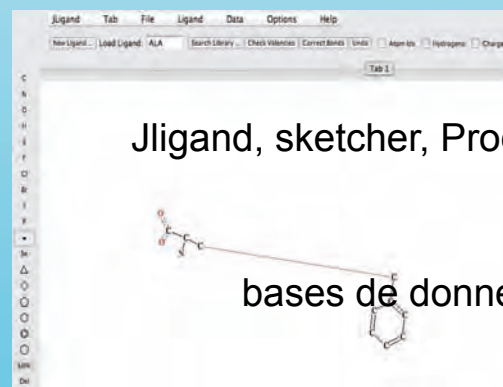
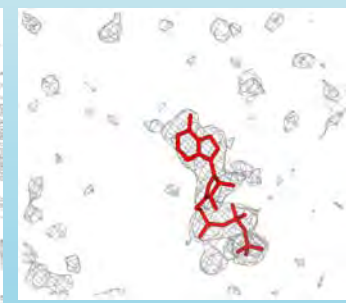
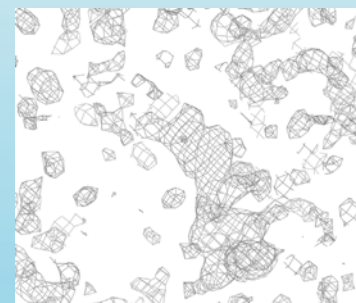
- “Structure based drug design”
- Co-cristallisation ou trempage / solubilité DMSO
- Carte de Fourier différence ($> 2\text{-}3$ sigma)
- Résolution $3\text{\AA}$ au moins (sinon “UFO”) pour étudier le mode de fixation
- Conformation peut différer des conformations provenant de la CSD,
 reconstruction, real-space fit



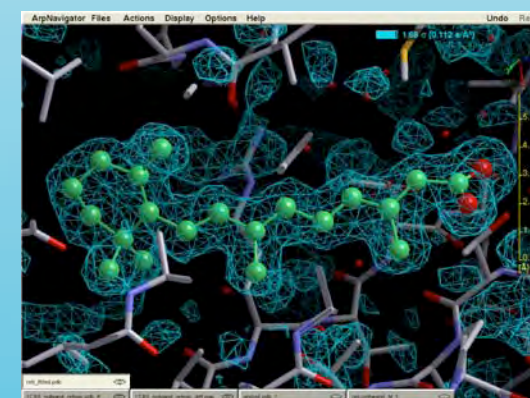
Rapide et automatique identification du site et construction du ligand identification $< 1\text{\AA}$ rmsd



- Different resolutions and data quality
- Different ligand complexity/topology
- Partial disorder of a ligand
- Different ligands at the same site?



bases de données : DrugBank,
Hic-up,...



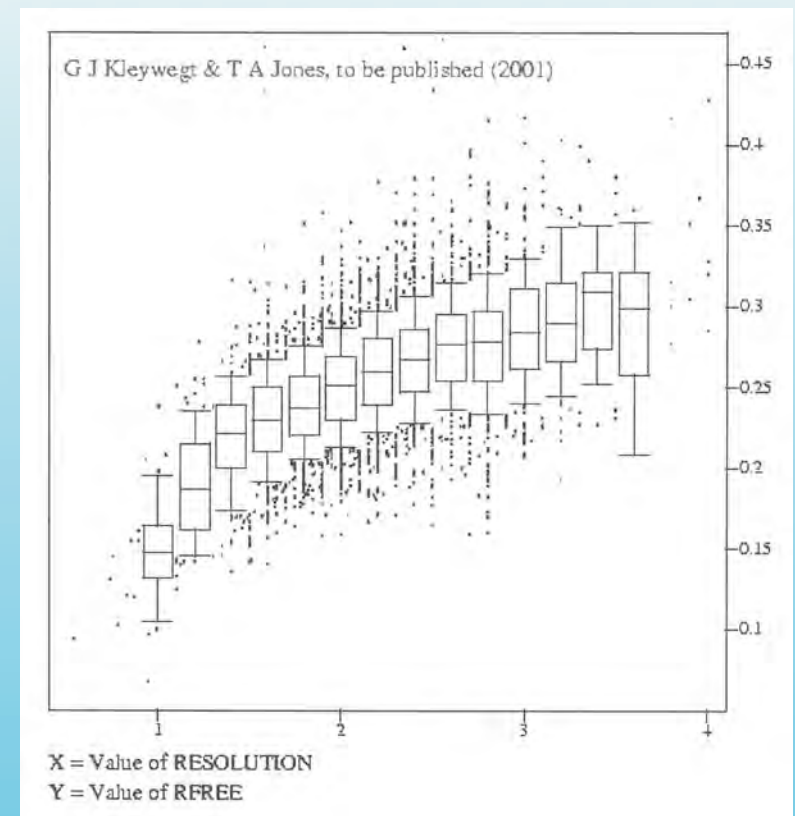
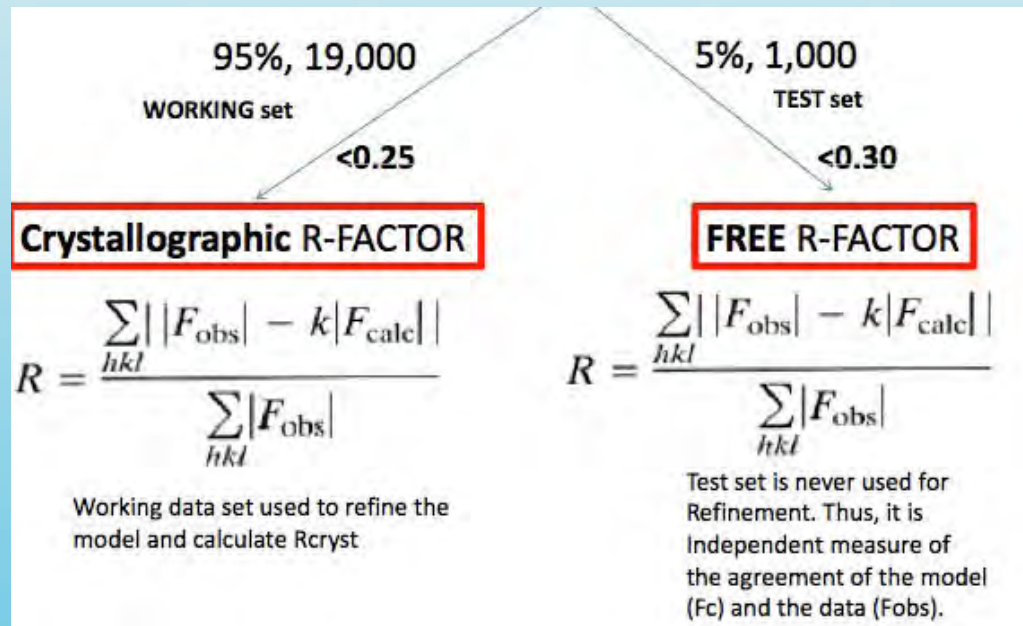
Validation des structures

Les facteurs d'accords R

(critère global de qualité du modèle, *en moyenne 20%*)

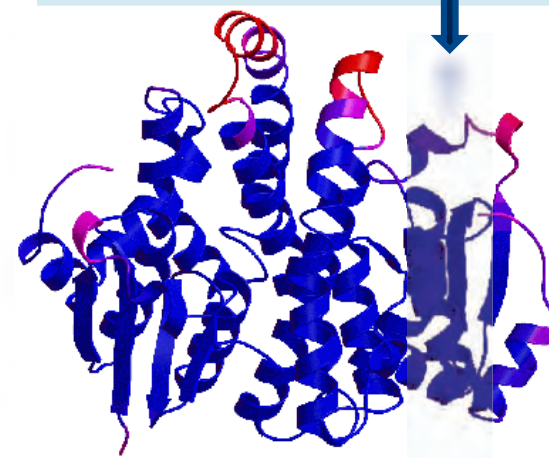
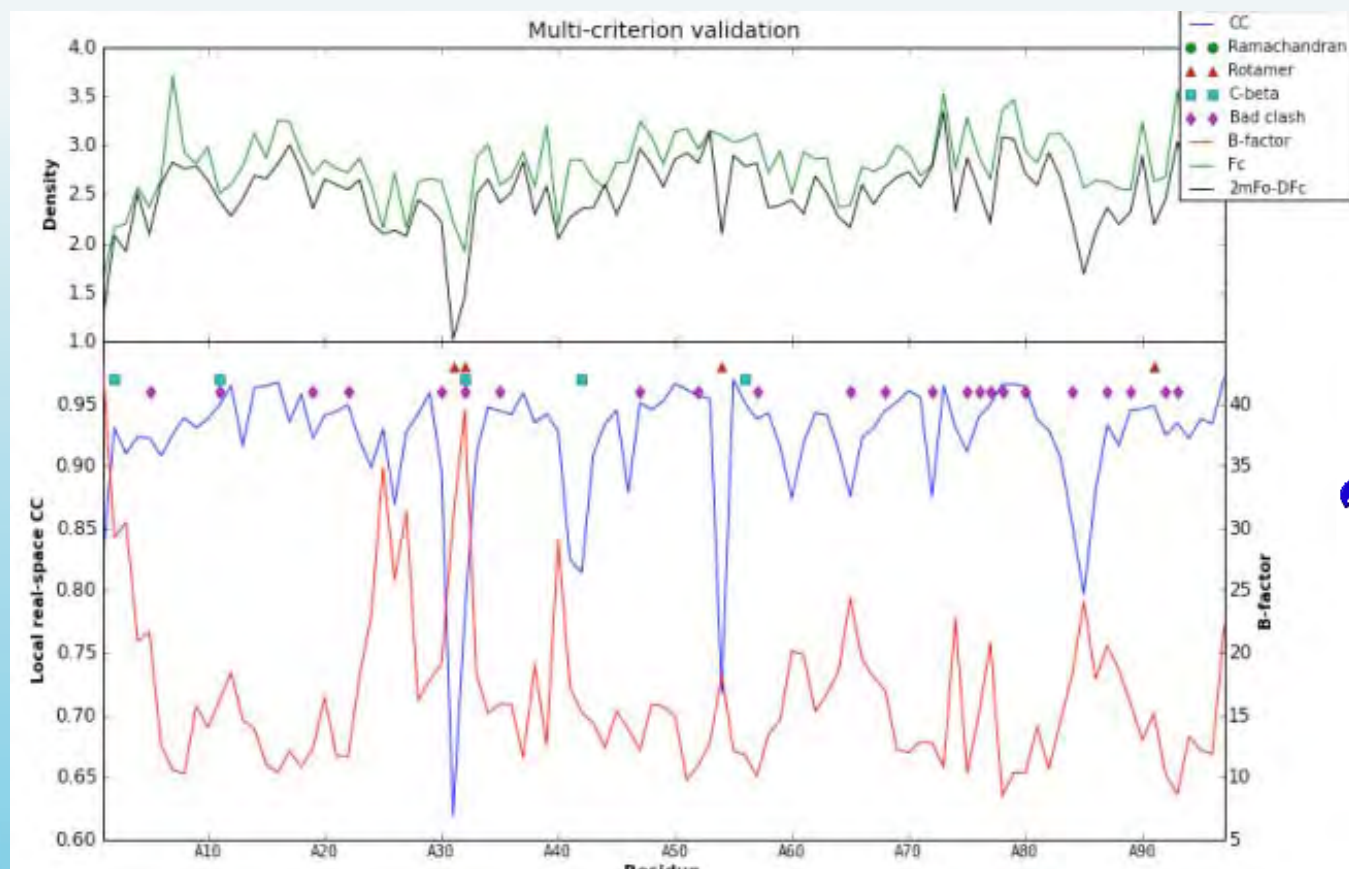
Ecart de R_{free} avec R_{fact} est significatif

- **Facteur R-free** indique le degré avec lequel le modèle prédit les données du test set (5 à 10% des données exclues du modèle et de l'affinement)
- corrélé avec la précision du modèle atomique

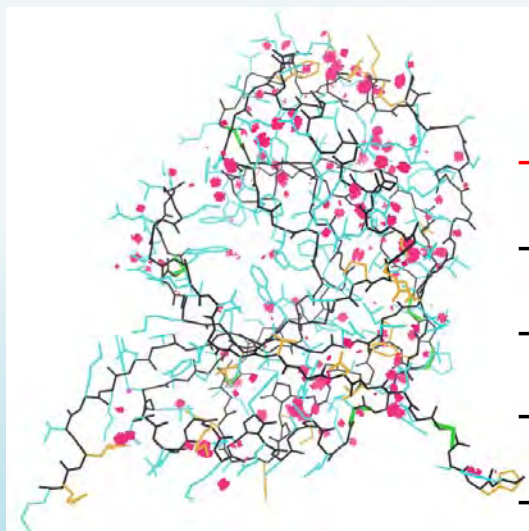


Validation des structures (suite)

Indicateurs de qualités locales du modèle tels que **Real-space Fit cc**, les **B-fac.**



Validation des structures (suite et fin)



- Ramachandran criteria

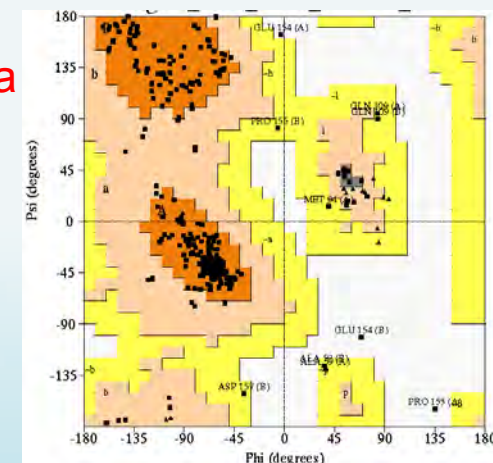
- Contacts inter-atomiques, Clash-score

- Hydrophobe dedans, hydrophile dehors

- Sites de fixations des ligands, ions, métaux

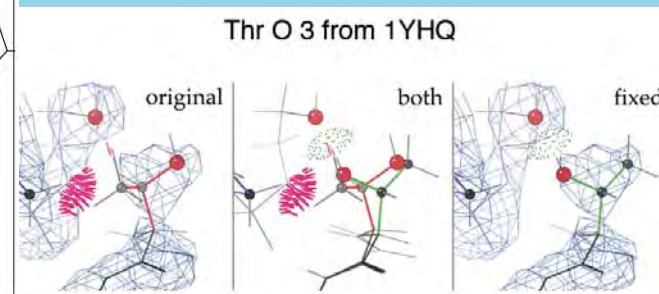
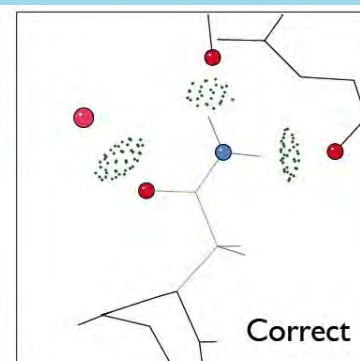
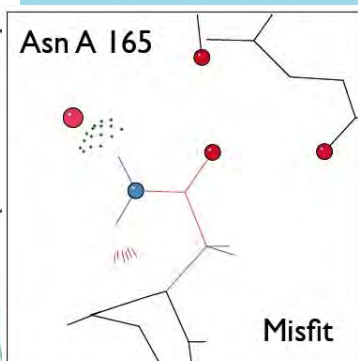
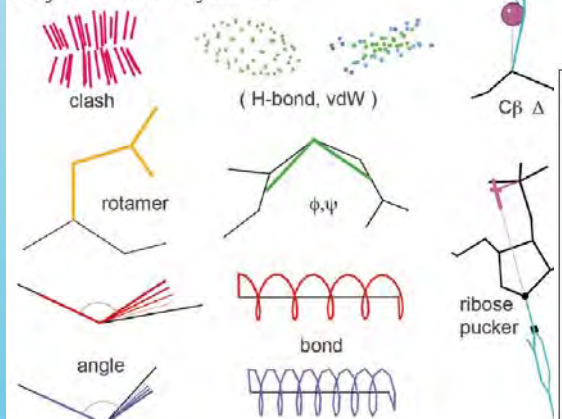
- Liaisons hydrogènes

- Rotamères des chaînes latérales

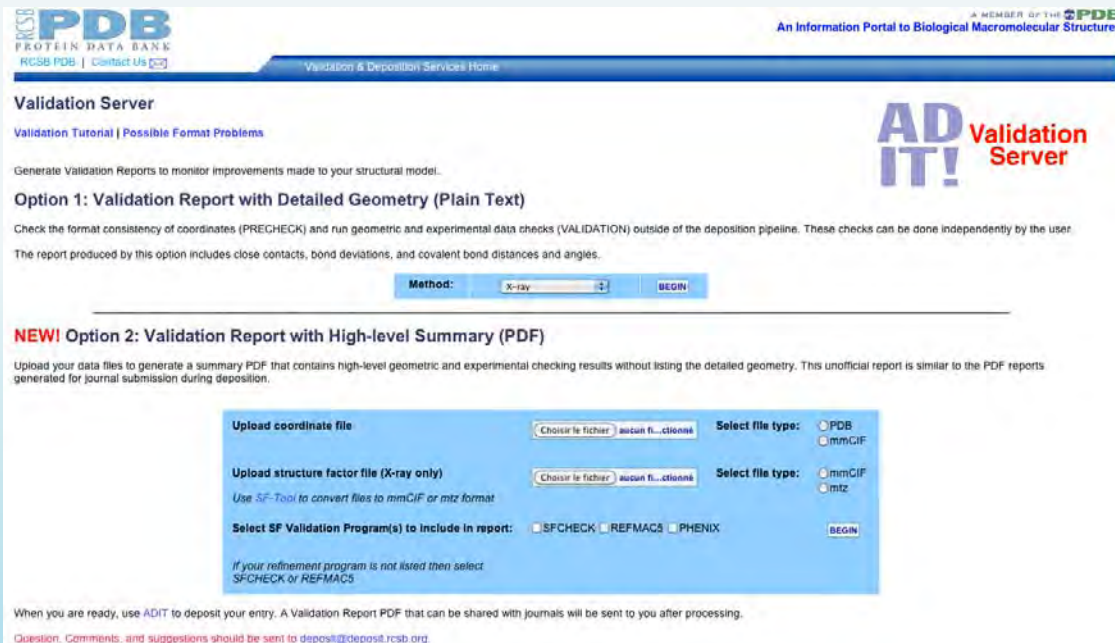


, Whatcheck, Sfccheck

Key to Outlier Symbols:



Validation et dépôt des structures



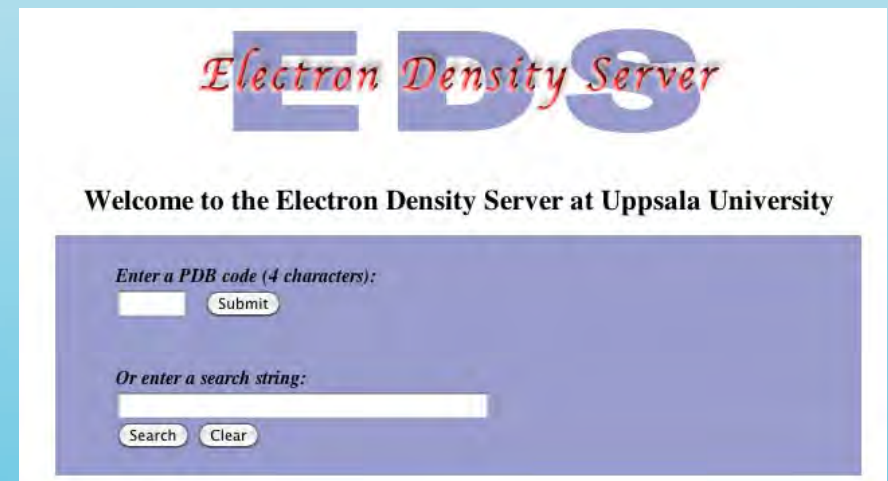
The screenshot shows the PDB Validation Server interface. At the top, it says "PDB PROTEIN DATA BANK" and "An Information Portal to Biological Macromolecular Structures". The main heading is "Validation Server". Below it, there's a "Validation Tutorial | Possible Format Problems" link. The text says: "Generate Validation Reports to monitor improvements made to your structural model." There are two options: "Option 1: Validation Report with Detailed Geometry (Plain Text)" and "Option 2: Validation Report with High-level Summary (PDF)". Option 1 includes a description of checks and a "Method:" dropdown set to "X-ray" with a "BEGIN" button. Option 2 is marked as "NEW!" and includes a description of a summary PDF. Below this is a form with fields for "Upload coordinate file", "Upload structure factor file (X-ray only)", and "Select SF Validation Program(s) to include in report:". There are also "Select file type:" radio buttons for PDB, mmCIF, and mtz. A "BEGIN" button is at the bottom right of the form. At the very bottom, it says "When you are ready, use ADIT to deposit your entry. A Validation Report PDF that can be shared with journals will be sent to you after processing." and "Question, Comments, and suggestions should be sent to deposit@deposit.rcsb.org".

2008 Dépôt des facteurs de structures observés obligatoires

A quand celui des images ?

*cf Fraud rocks protein community
Nature 462, 970 (2009)*

Serveur de visualisation des structures avec les facteurs de structure :



The screenshot shows the Electron Density Server (EDS) interface. At the top, it says "EDS Electron Density Server". Below it, it says "Welcome to the Electron Density Server at Uppsala University". There are two input sections: "Enter a PDB code (4 characters):" with a text box and a "Submit" button, and "Or enter a search string:" with a text box, a "Search" button, and a "Clear" button.

Nature. Author manuscript; available in PMC 2008 August 11.

Published in final edited form as:

Nature. 2007 August 9; 448(7154): E1–E3.

Nature 462, 970 (2009)

In a report released earlier this month, the University of Alabama at Birmingham concluded that H. M. Krishna Murthy **acted alone in fabricating and falsifying results that appeared in ten papers**^{1,2,3,4,5,6,7,8,9,10} published during the past decade. The disputed papers have been cited more than 450 times.

Crystallographic evidence for deviating C3b structure?

Bert J.C. Janssen¹, Randy J. Read², Axel T. Brunger³, and Piet Gros¹

¹*Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands* ²*Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, U.K.* ³*Howard Hughes Medical Institute, and Departments of Molecular and Cellular Physiology, Neurology and Neurological Sciences, Structural Biology, and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford CA 94305-5432, USA*

Abstract

Activation of C3 into C3b is a critical step in the complement immune response against pathogenic, immunogenic and apoptotic particles. Ajees *et al.*¹ report a crystal structure of C3b, which deviates from the one reported by Janssen *et al.*² and Wiesmann *et al.*³. Analysis of the data deposited by Ajees *et al.*¹ reveals features that are inconsistent with known physical properties of macromolecular structures and their diffraction data. Therefore, Ajees *et al.* do not provide substantial evidence for the deviating crystal structure of C3b.

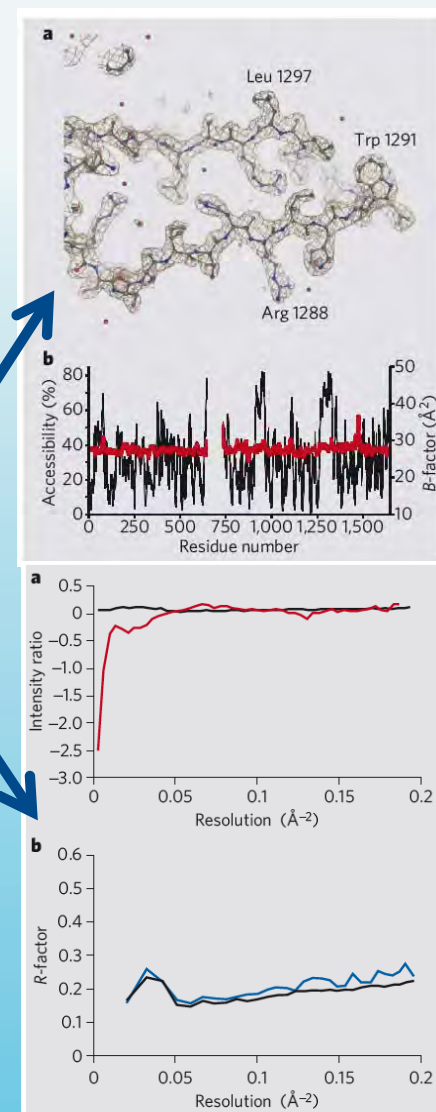


Figure 1 | Absence of bulk solvent in the deposited diffraction data of Ajees *et al.*¹. a, Plot, as a

CCP4 à l'instar du CCP14?

<http://www.ccp4.ac.uk/index.php>

1979 Collaborative Computational Project Number 4 in Protein Crystallography : *initiative UK*
Distribution « worldwide » gratuite de la suite.

Patchwork de binaires* (> 200) indépendants qui communiquent *via* des fichiers/bibliothèques au format CCP4 standard : plus grande *Efficacité / flexibilité / portabilité / simplicité* pour modifier et inclure des programmes similaires.

* C/C++ et Fortran 77 (Win, Linux, OS X)
Modèle atomique ASCII (PDB, mmCIF ou ss-structure atome lourd).
Réflexions (MTZ) et Map/mask formats binaires
Bibliothèques Clipper, MMDB, CCTBX notamment

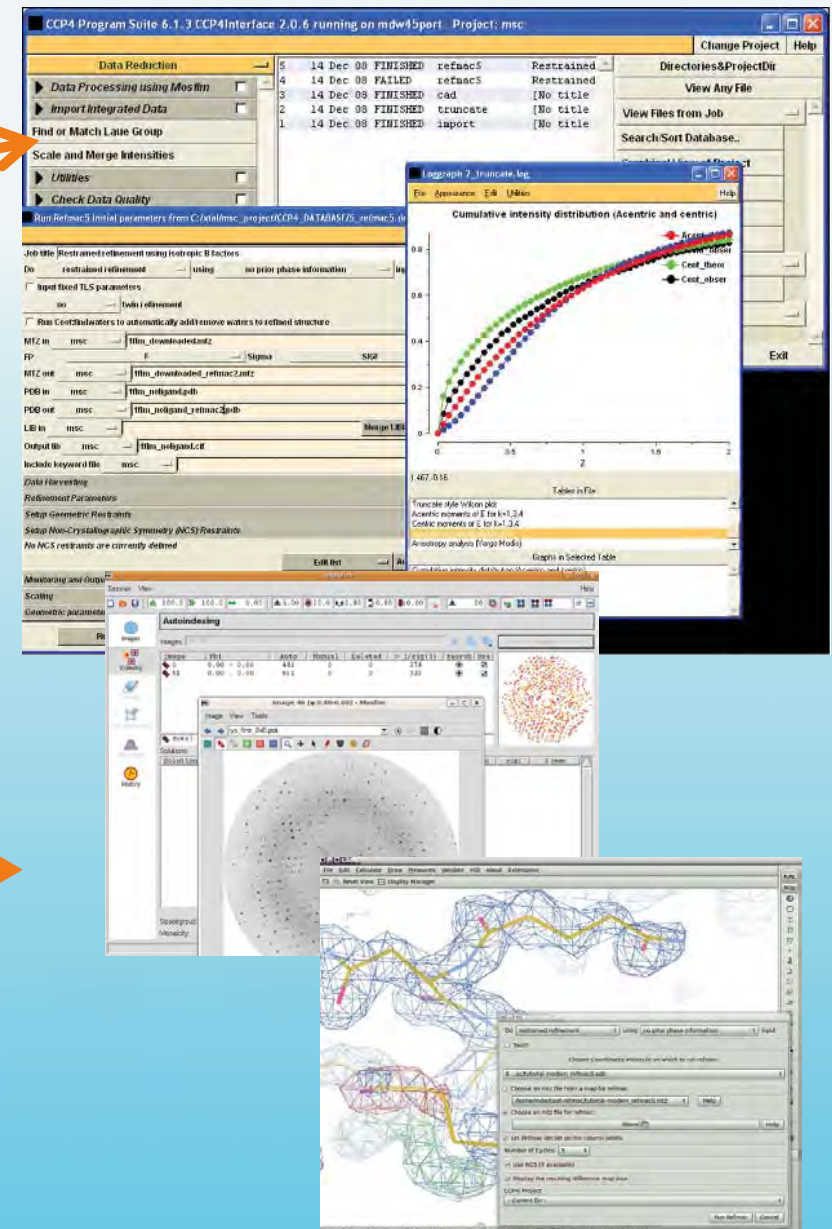
Mise à jour technique et scientifique régulière ([CCP4 6.2.0 \(15 July 2011\)](#))

GUI CCP4i

WIKI

CCP4 Bulletin board: liste de diffusion très (ré)active

Tendance à regrouper les programmes dans une suite logique en vue de l'automatisation.
ex : Mr Bump



À l'heure de l'automatisation *

Ex : (Adams et al., 2010) USA Berkeley, Los Alamos, Duke, Cambridge– UK

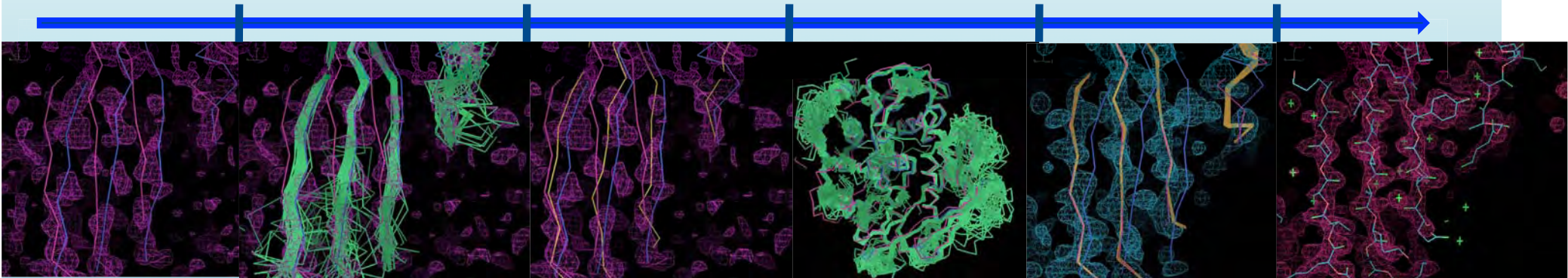
Phenix

*

- résout les cas faciles pour les non-experts
- aide les experts pour les cas difficiles,
- accélère le travail
- aide à réduire les erreurs
- Traite plus de possibilités et estime les incertitudes

De la 1^{ère} carte de densité

à celle du modèle fini



- SHARP/autoSHARP/ autoBuster (G. Bricogne - [Global Phasing Limited](#))
 - SHELX (Sheldrick, 2008) (D)
 - ARP/wARP (Langer et al., 2008)
 - Uppsala Factory (S)
- et beaucoup d'autres...

Les systèmes graphiques les plus utilisés

pour vérifier la validité du modèle et le fit dans la densité ou modifier certains fragments de la structure,

- O (Jones et al., 1991)



- XtalView (McRee, 1999) payant

- QUANTA (Accelrys Inc., San Diego; Oldfield, 2000) payant

- COOT (Emsley & Cowtan, 2004) (CCP4)

- MAIN (Turk, 2000, 2004).



+++ Vision en stéréo, écrans 3D

+++ Joystick pour bouger le curseur MOVE ATOM or MOLECULE HERE

+ Nbx menus pour faire apparaître les cartes de densité et varier leur contour
+ couplés avec calculs de dynamique, minimisation,...

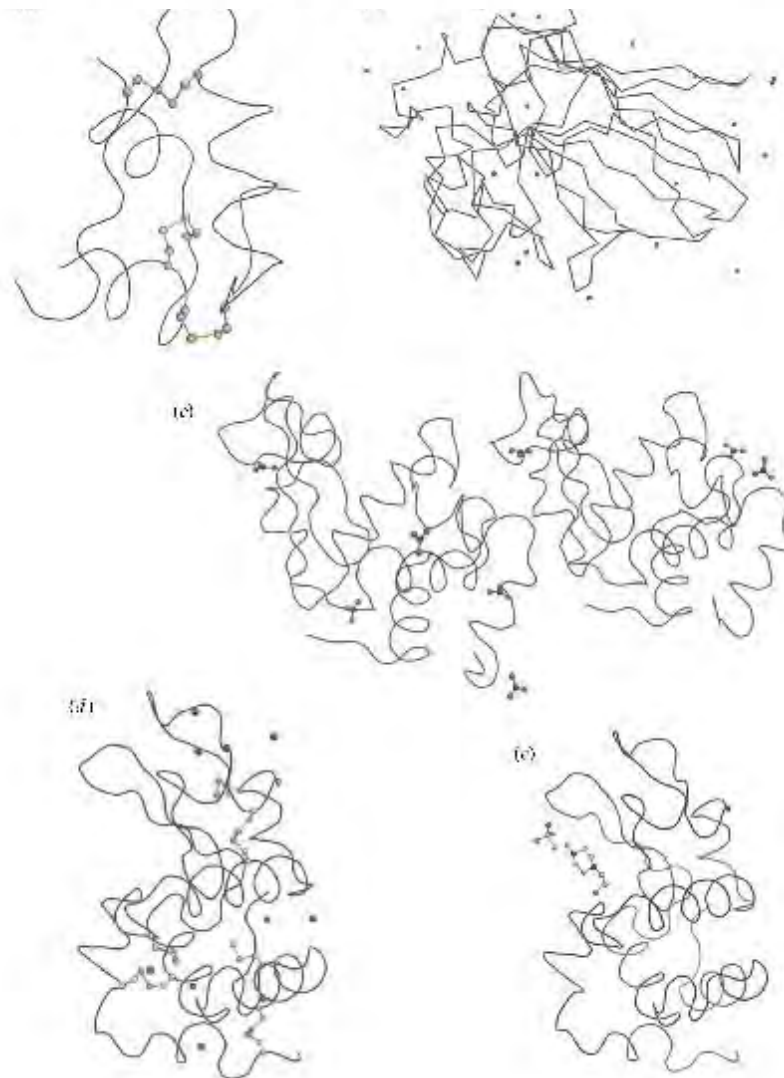
- orientés davantage structures qu'étude de l'empilement cristallin,

A tutorial for learning and teaching macromolecular crystallography

Annette Faust,^a Santosh Panjikar,^a Uwe Mueller,^b Venkataraman Parthasarathy,^a
Andrea Schmidt,^a Victor S. Lamzin^a and Manfred S. Weiss^{a*}

Received 26 May 2008
Accepted 8 October 2008

- (i) a sulfur single-wavelength anomalous diffraction (**S-SAD**) structure determination of cubic insulin,
- (ii) a bromide multiple-wavelength anomalous diffraction (**Br-MAD**) structure determination of thaumatin
- (iii) a structure determination by **molecular replacement** using monoclinic lysozyme
- (iv) the identification of bound solvent ions in lysozyme using a longer-wavelength data set
- (v) the identification of a weakly bound ligand at the active site of lysozyme.



CONCLUSION

✓ Evolution plus tardive de la PX du fait de la complexité des systèmes étudiés 100 x la taille des entités chimiques et plus emphatique

(nombreux prix Nobel : en rouge les structuralistes bio. : **1914 M. von Laue** **1915 W.H. Bragg and W.L. Bragg** **1954 L. Pauling** **1962 J.C. Kendrew, M. Perutz, F. Crick, J.D. Watson and M.H.F. Wilkins** **1964 Dorothy C. Hodgkin** **1976 W.N. Lipscomb** **1982 A. Klug** structure des complexes protéines-acides nucléiques **1985 H.A. Hauptman and J. Karle** **1988 J. Deisenhofer, R. Huber and H. Michel** centres réactifs des bactéries pourpres(R viridis) **1997 J.E Walker** **2003 P. Agre und R. MacKinnon** potassium channel **2006 R. Kornberg** RNA-Polymerase II in action **2009 V. Ramakrishnan, T.A. Steiz and Ada Yonath** (Ribosome ca 90000 atomes))

✓ mais qui doit profiter à sa parente, la cristallographie des petites molécules, comme un juste retour des choses (transfert de méthodologie, de technologie d'un côté à l'autre).

Deux orientations actuellement suivies :

- **Une d'automatisation** pour les nombreuses structures à résoudre
 - nombreux génomes, protéines encore inconnues ou non exprimées 100000 par cellule eucaryote,
 - complexes divers et variés de protéines à l'infini...) donnant lieu à des cristaux diffractant autour de 2 Å
- **Une de recherche** sur les grands complexes ou machines moléculaires dont on ne peut attendre que de petits cristaux, souvent mal formés,
 - Intégrative avec d'autres approches RMN, SAXS, EM, Neutron, AFM, simulation...
 - Cristallographie sous perturbation (haute pression, couplée à des mesures spectroscopiques...)

Avec XFEL , quel avenir pour la cristallographie traditionnelle ?