

Macromolecule Structure determination by X-ray crystallography. A primer.

Pr. Jean Cavarelli

Structural biology of epigenetic targets



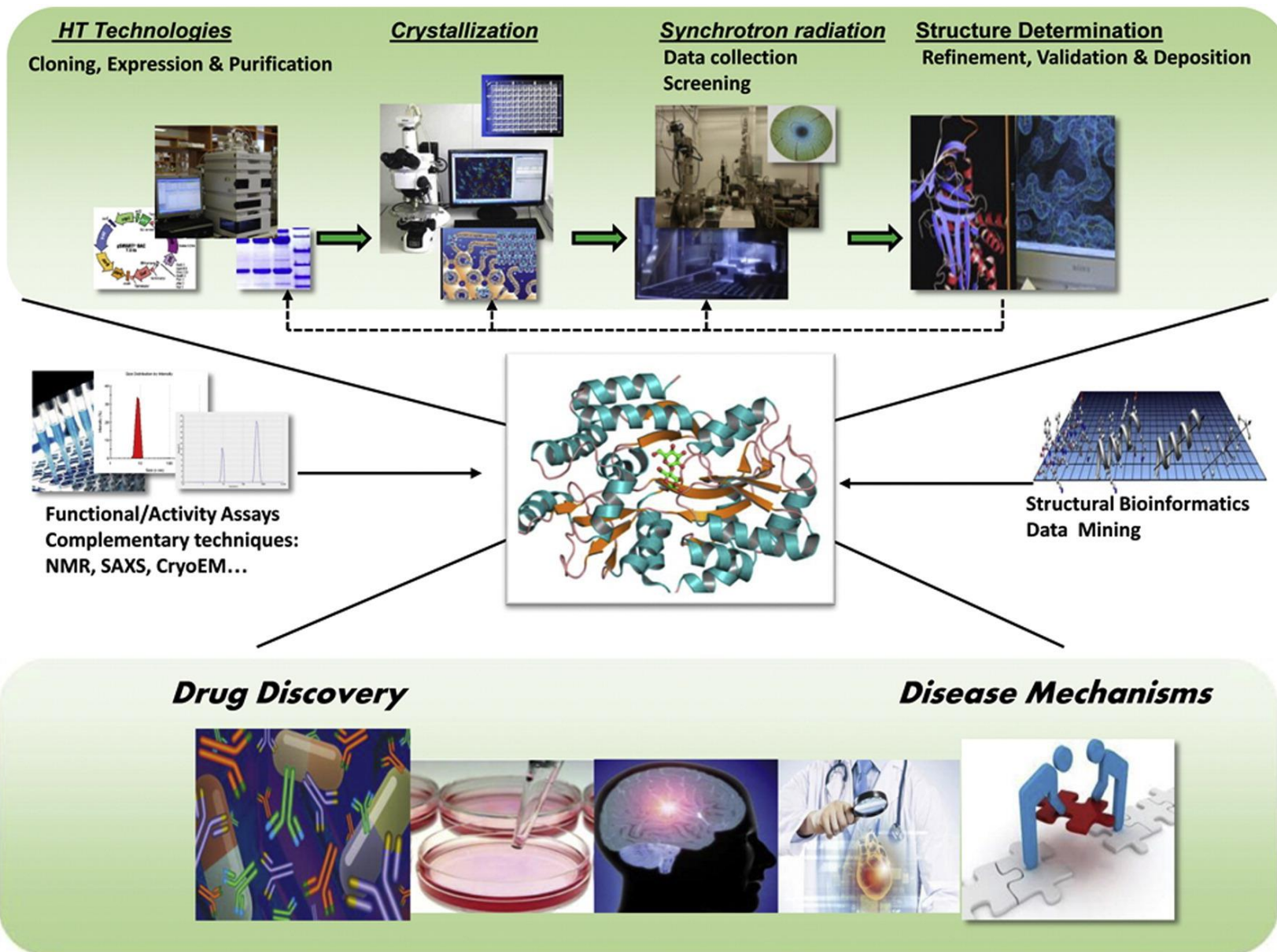
jean.cavarelli@unistra.fr

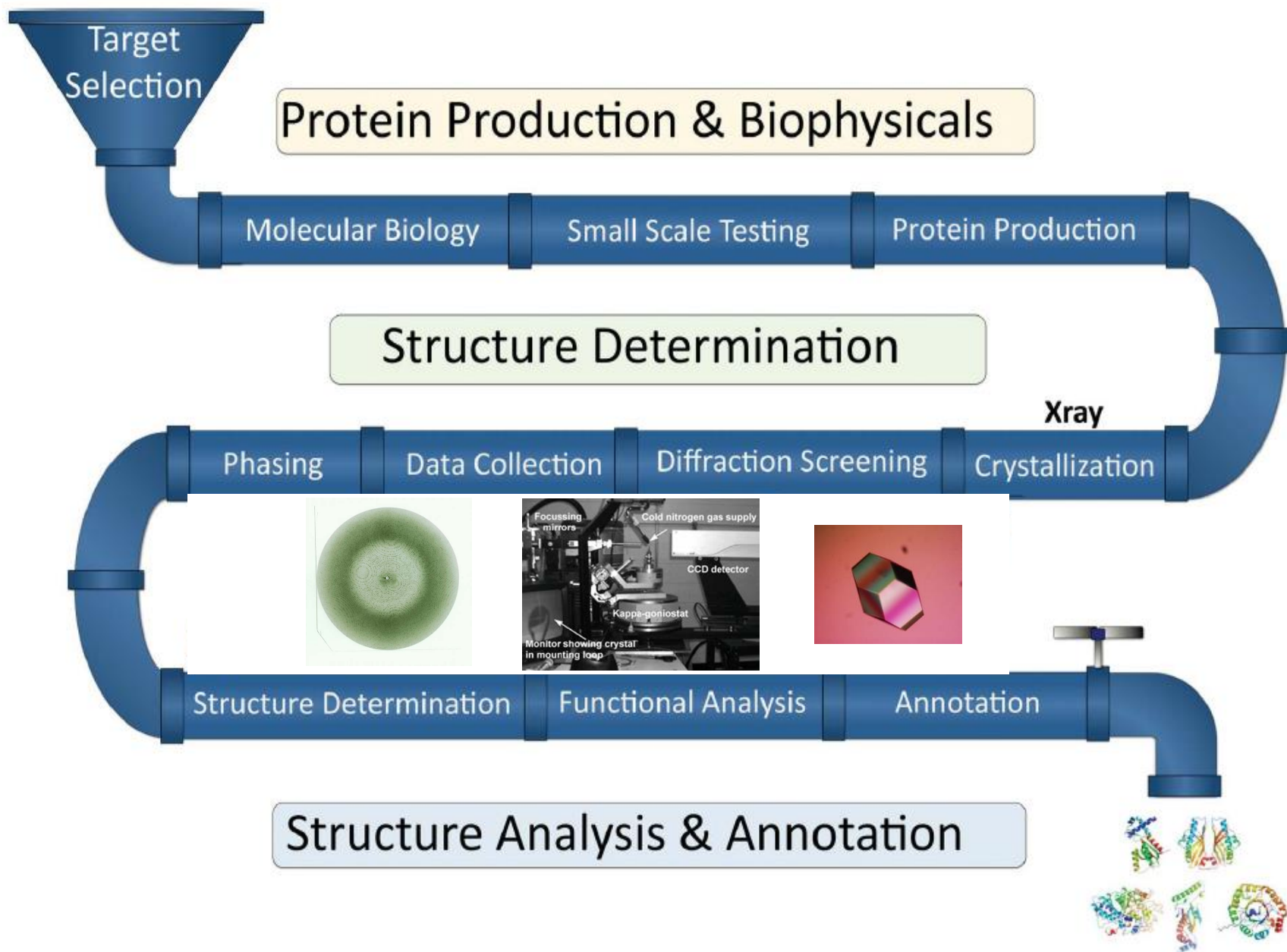
jean.cavarelli@igbmc.fr



Inserm

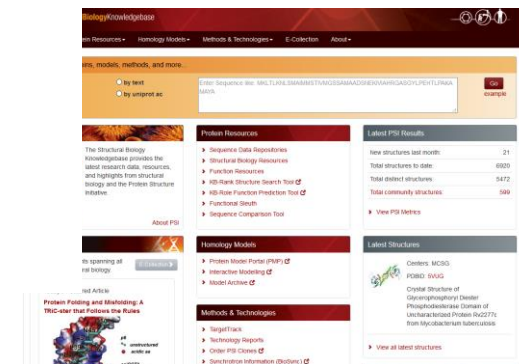
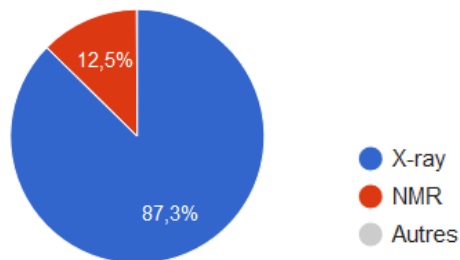
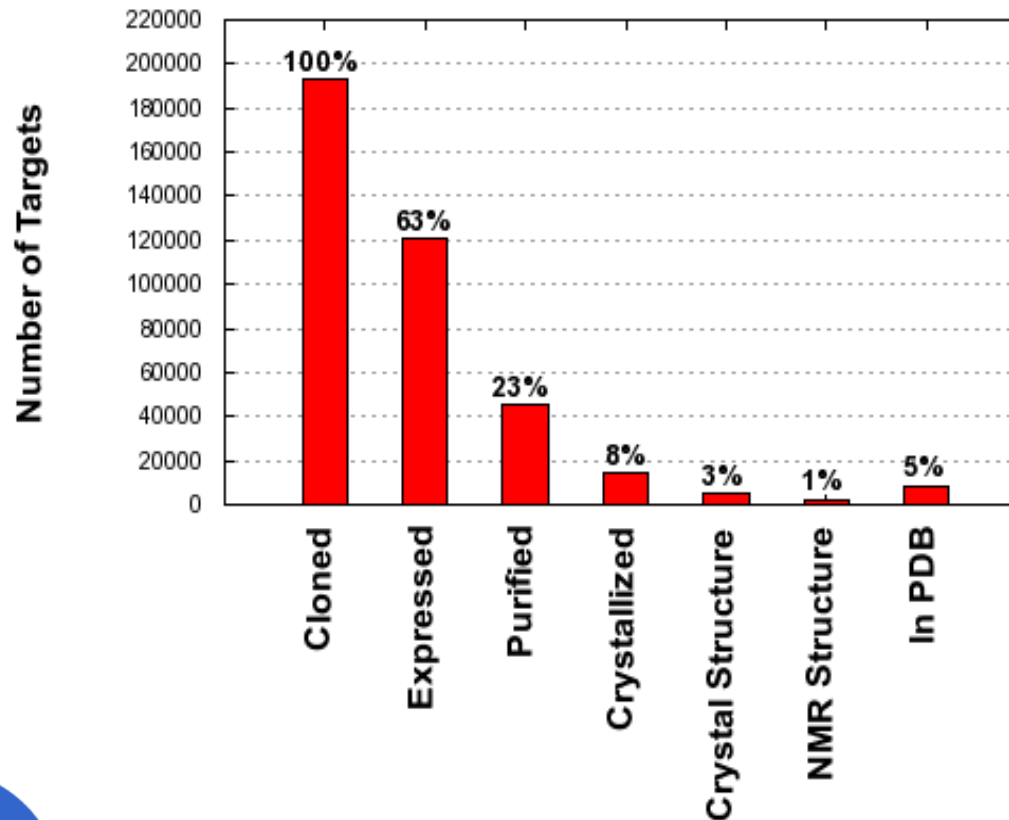
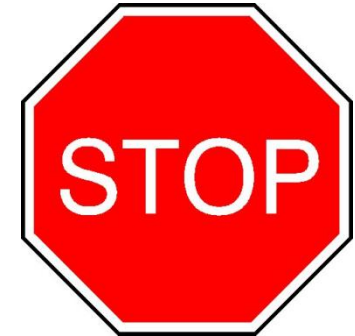






DATA from Structural genomics centers

<http://targetdb.pdb.org/>



Interactions rayonnements - matière : Notions de base

1. Modification du rayonnement sous l'action de la matière :

Diminution de l'intensité : absorption du rayonnement

Changement du module du vecteur d'onde : une partie des photons ou des particules ont perdu de l'énergie

Changement du vecteur d'onde en direction : diffusion

2. Modification de la matière sous l'action du rayonnement :

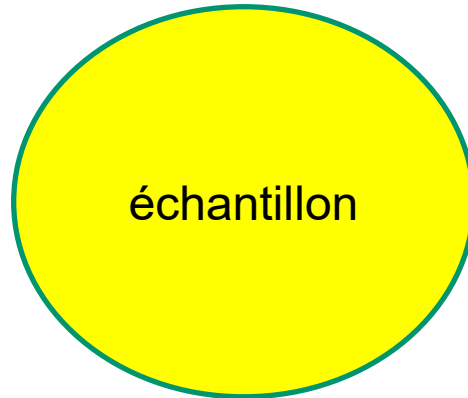
L'énergie perdue par le rayonnement est intégralement transférée aux atomes de la matière sous différentes formes

énergie potentielle, énergie cinétique, énergie de vibration thermique

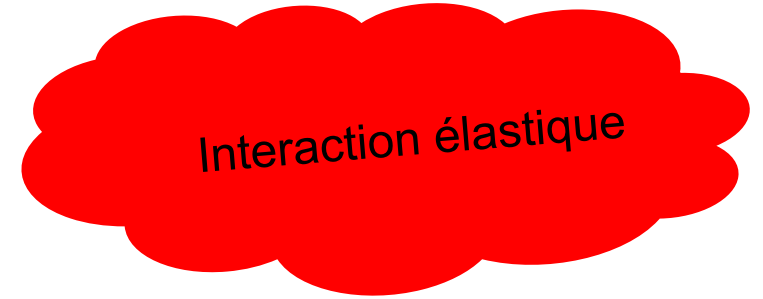
une partie de cette énergie pourra être réémise sous forme de rayonnements ou de chaleur

Interactions rayonnements - matière

Photons/RX
Neutrons
Electrons



absorption

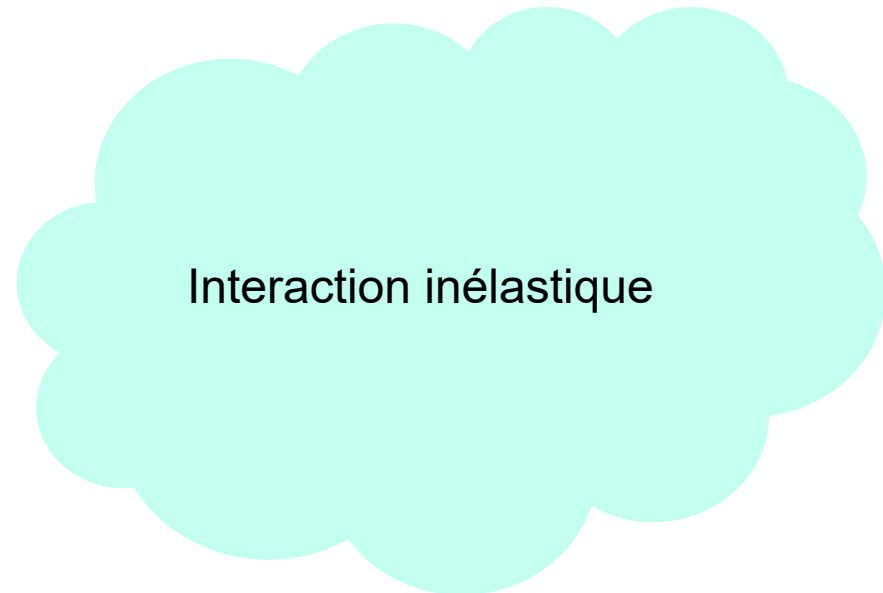


Interaction élastique

Cohérente



Incohérente



Interaction inélastique

Méthodes spectroscopiques

Interaction RX avec un cristal "typique " de protéine
(valeurs numériques pour $\lambda=1\text{\AA}$, 12.4 keV)

La majeure partie du rayonnement n'est pas « diffusée » (98%)
(pas de changement du vecteur d'onde en direction)

Pour la partie diffusée (2%), 3 possibilités :

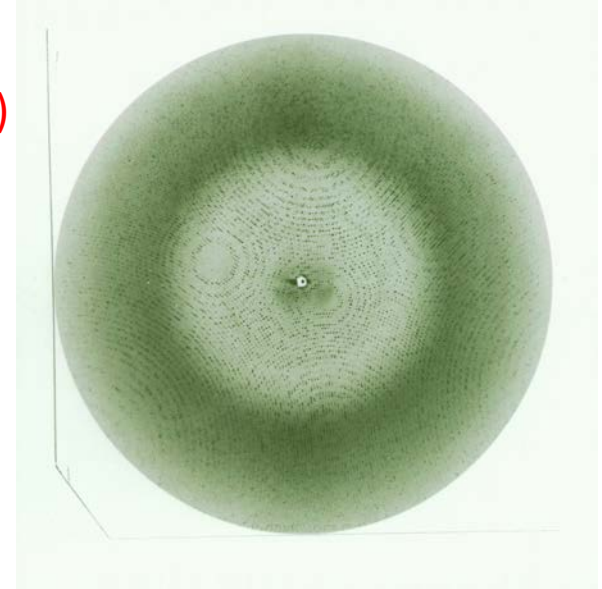
Interaction par effet photoélectrique (84%)
éjections d'électrons des atomes

Diffusion " Thompson " ("Rayleigh") (8%)

interaction cohérente et élastique, donnant lieu au phénomène d'interférences constructives observé (diffraction)

Diffusion "Compton" (8%)

Interaction inélastique et incohérente



Diffusion élastique des Rayons X

Les Rayons X interagissent avec particules chargées :
électrons, protons

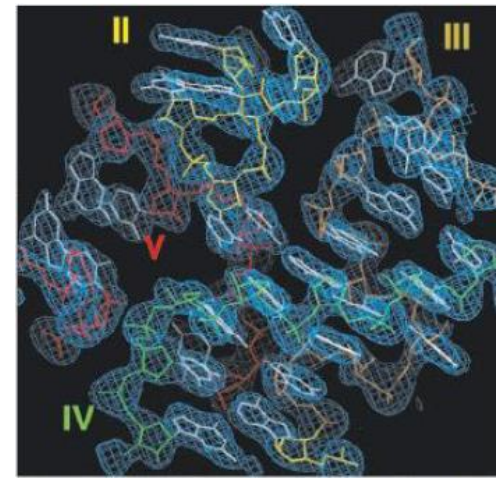
Section efficace de diffusion élastique électrons/protons

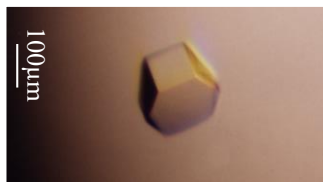
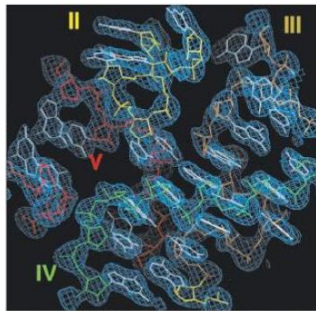
$$\frac{\sigma_e}{\sigma_p} \div \left(\frac{M_p}{M_e} \right)^2 \cong \left(\frac{1.6 \cdot 10^{-27}}{9.1 \cdot 10^{-31}} \right)^2 \cong 3.4 \cdot 10^6$$

La participation des protons est négligeable pour la diffusion élastique des RX

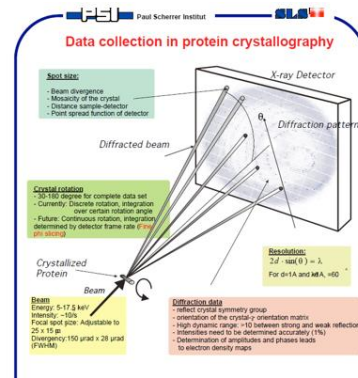
Pour les RX, la matière est vue comme une répartition d'électrons

La diffusion élastique des rayons X apporte des informations sur la densité électronique de l'échantillon.

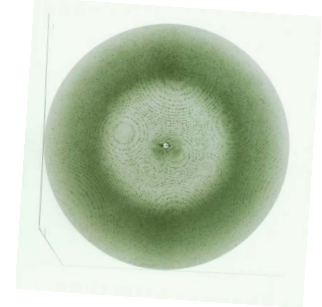




Espace direct



Espace réciproque



Une structure cristalline

- Un groupe d'espace, une maille cristalline
- N atomes dans une maille
- Pour chaque atome j ($j = 1, N$), les "paramètres" :
 - l'interaction avec le rayonnement
 - facteur de diffusion f_j
 - les coordonnées de position
 - (x_j, y_j, z_j) , (coordonnées relatives)
 - le désordre statique et dynamique
 - q_j, w_j

$\rho(x, y, z)$

Un ensemble de m ondes "diffusées"

- Les rayons X interagissent avec les particules chargées : électrons, protons
- On utilise l'interaction élastique (cohérente) des rayons X
 - la participation des protons est négligeable (par rapport aux électrons)
 - pour les rayons X, la matière est vue comme une distribution d'électrons
- m lié à la limite de diffraction du cristal (d_{min}),
- chaque onde est caractérisée par 3 paramètres :
 - une direction de diffusion,
 - un module,
 - une phase (une différence de phase par rapport à une onde de référence).

$m F_{hkl}, hkl \in D^*$

$$\rho(x, y, z) \xrightleftharpoons[\hat{\mathcal{F}}]{\hat{\mathcal{F}}^{-1}} F_{hkl} \quad \left\{ \begin{array}{l} \rho(x, y, z) = \hat{\mathcal{F}}[F_{hkl}] \\ F_{hkl} = \hat{\mathcal{F}}^{-1}[\rho(x, y, z)] \end{array} \right.$$

Le résultat de processus cristallographique

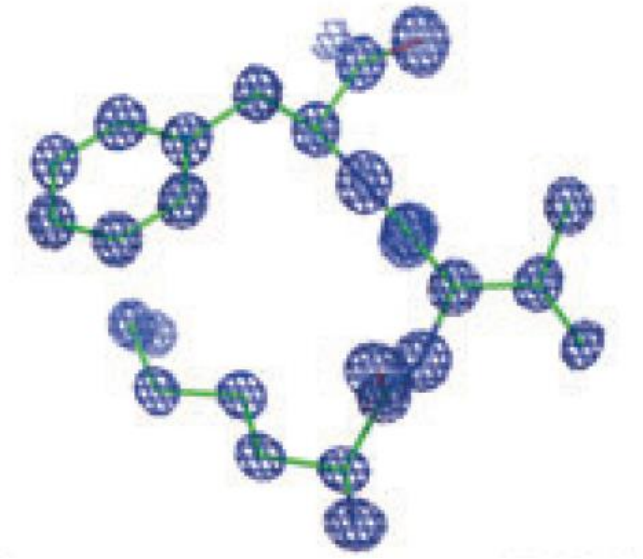
La référence et le juge de paix

La fonction densité électronique

Fonction densité électronique théorique

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{h=+\infty} \sum_{k=-\infty}^{k=+\infty} \sum_{l=-\infty}^{l=+\infty} |F_{hkl}| e^{i\phi_{hkl}} e^{-2i\pi(hx+ky+lz)}$$

- fonction réelle (**en absence de diffusion anormale**) et positive (si tous les termes sont inclus)
- fonction qui présente des maxima dans les régions occupées par les atomes
- fonction qui est donc en général interprétée en termes de positions atomiques



Espace direct $\Leftrightarrow$ Espace réciproque

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} F_{hkl} e^{-2i\pi(hx+ky+lz)} = \left(\frac{1}{V}\right) \sum_{(h,k,l) \in \mathbb{Z}^3} |F_{hkl}| e^{i\phi_{hkl}} e^{-2i\pi(hx+ky+lz)}$$

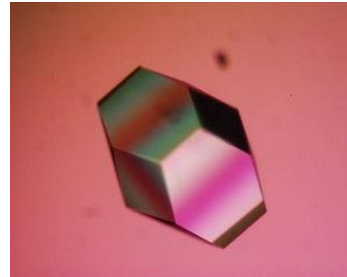
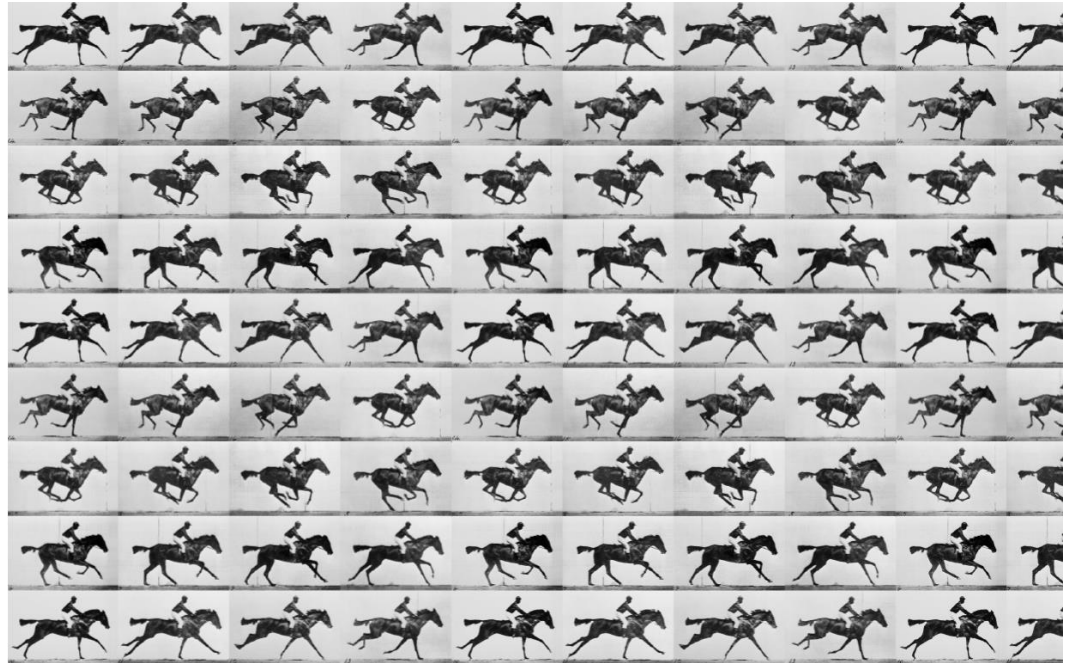
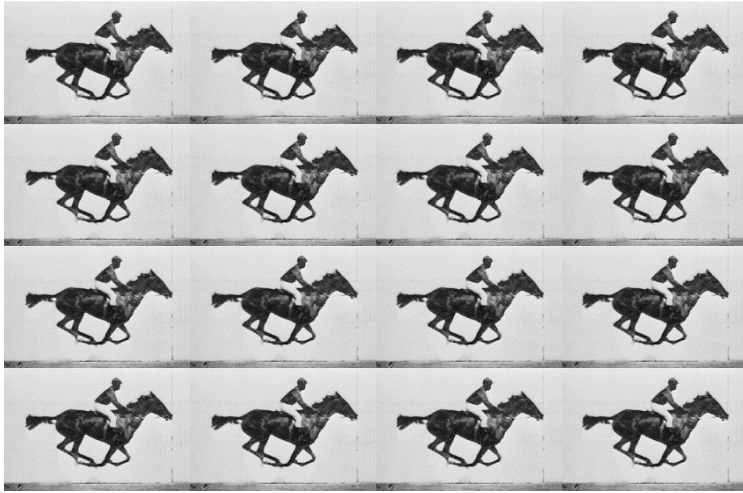
$$F_{hkl} = \sum_{j=1}^N q_j f_j^{hkl} w_j^{hkl} e^{2i\pi(hx_j+ky_j+lz_j)}$$

Structure	positions atomiques	facteur de diffusion	Facteur de Structure	module FS	Phase FS	Densité électronique
A	$\vec{r}_j$	f_j	F_H	$ F_H $	α_H	$\rho(r)$
B	$-\vec{r}_j$	f_j	$\overline{F_H}$	$ F_H $	$-\alpha_H$	$\rho_B(r) = \rho(-r)$
C	$\vec{r}_j$	$-f_j$	$-F_H$	$ F_H $	$\pi + \alpha_H$	$\rho_C(r) = -\rho(r)$
D	$-\vec{r}_j$	$-f_j$	$-\overline{F_H}$	$ F_H $	$\pi - \alpha_H$	$\rho_D(r) = -\rho(-r)$

structure cristallographique = moyennes dans le temps et dans l'espace :

ensemble des structures présentes dans le cristal pendant le temps de l'expérience

le modèle construit en termes de paramètres atomiques se doit de refléter au mieux les positions et les mouvements des atomes dans le cristal



$\sim 10^{13}$ molecules

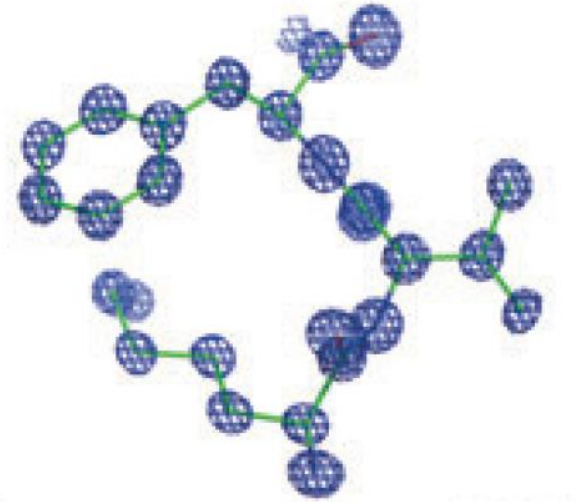
Paramètres de déplacements atomiques

$$F_{hkl} = \sum_{j=1}^N q_j f_j^{hkl} w_j^{hkl} e^{2i\pi(hx_j + ky_j + lz_j)}$$

- Modéliser aux mieux les mouvements et le "désordre" des atomes dans la structure
- Les paramètres de déplacements atomiques

Désordre statique et dynamique

- Facteur d'occupation : q_i ($j=1,n$)
- Facteur de Debye Waller: w_i , ($j=1,n$)
 - Modèle simple isotrope, facteur d'agitation thermique
 - Modèle plus sophistiqué, anisotrope







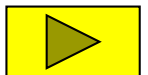
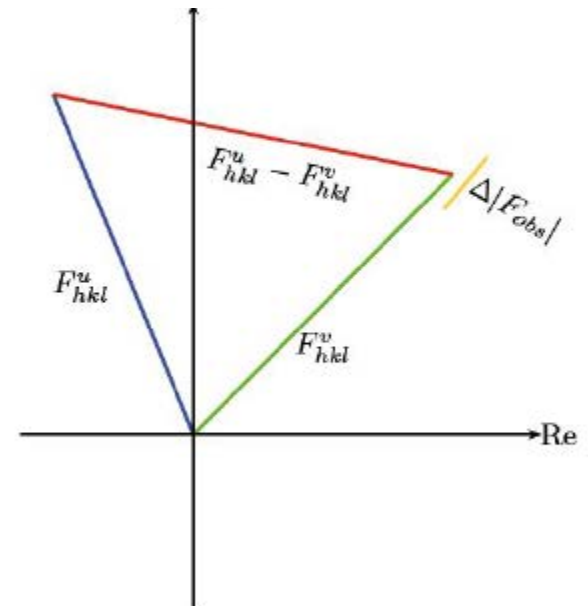
Fonction densité électronique. En théorie :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} |F_{hkl}^{vrai}| e^{i\phi_{hkl}^{vrai}} e^{-2i\pi(hx+ky+lz)}$$

L'information de phases est essentielle

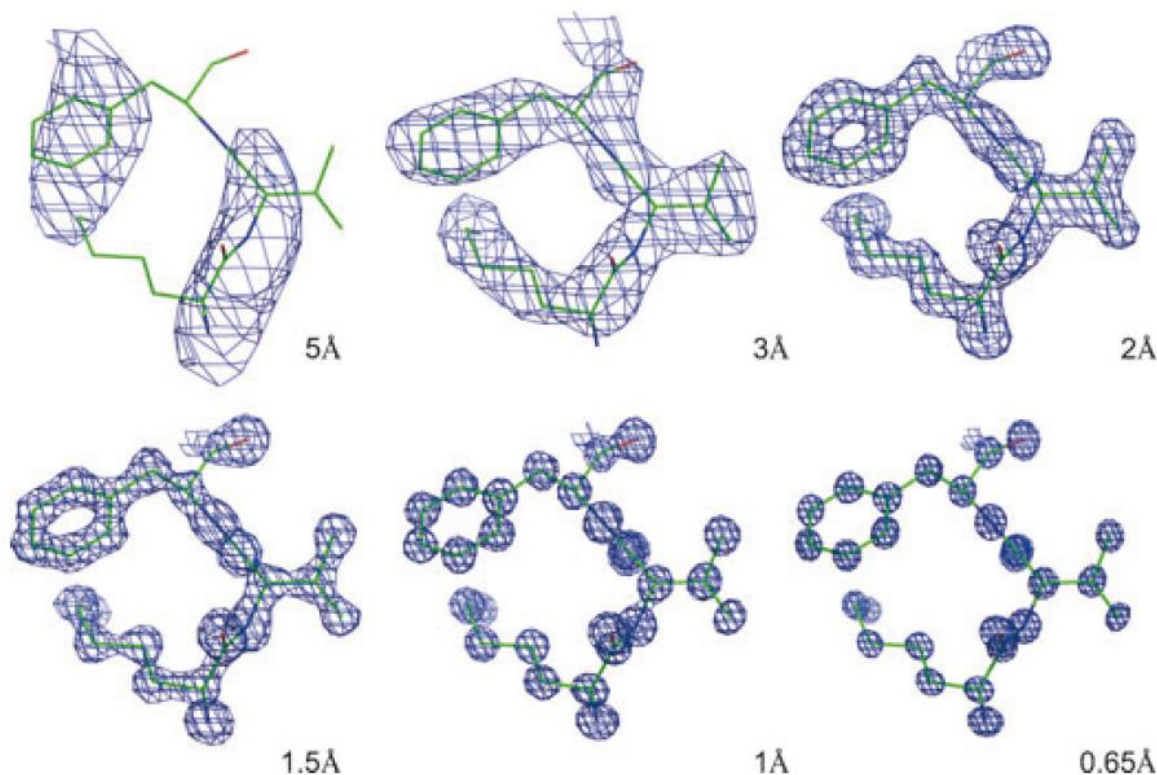
Théoreme de Parseval

$$\langle (\Delta\rho)^2 \rangle = \left(\frac{1}{V^2}\right) \sum_{h \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} |F_{hkl}^v - F_{hkl}^u|^2$$



Fonction densité électronique. La réalité :

$$\rho(x, y, z)^{obs} = \left(\frac{1}{V}\right) \sum_{hkl \in D^*} |F_{hkl}^{obs}| e^{i\phi_{hkl}^{calc}} e^{-2i\pi(hx+ky+lz)}$$



Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures

Alexander Wlodawer¹, Wlodek Minor^{2,3}, Zbigniew Dauter⁴ and Mariusz Jaskolski^{5,6}

FEBS Journal (2007) Journal compilation © 2007 FEBS. No claim to original US government works

Fonction densité électronique. La réalité :

$$\rho(x, y, z)^{obs} = \left(\frac{1}{V}\right) \sum_{hkl \in D^*} |F_{hkl}^{obs}| e^{i\phi_{hkl}^{calc}} e^{-2i\pi(hx+ky+lz)}$$

$\rho(x, y, z)^{obs} \neq \rho(x, y, z)^{vrai}$. 3 sources d'erreurs principales :

- les erreurs sur la valeur des coefficients utilisés

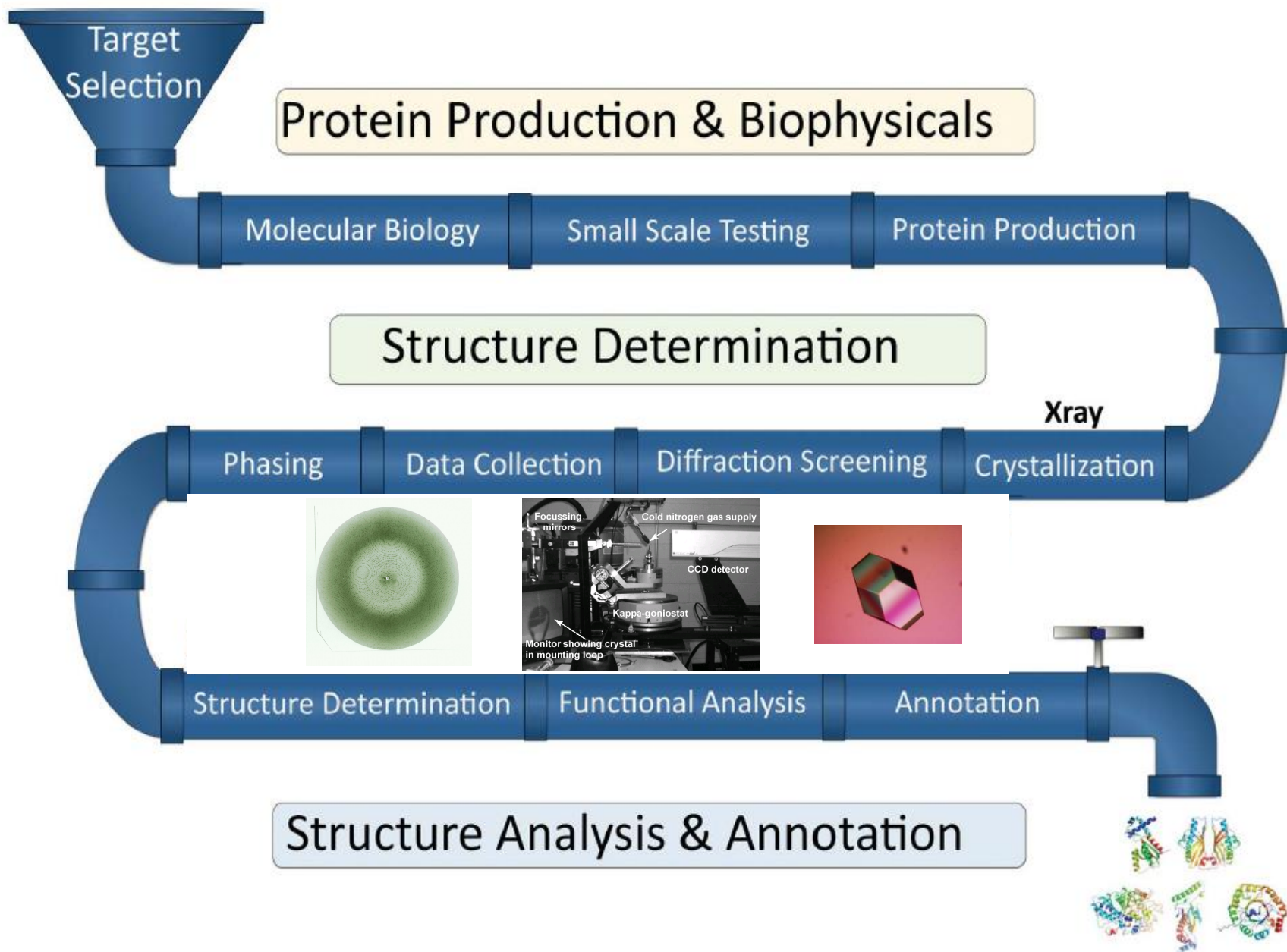
$$|F_{hkl}^{obs}| \neq |F_{hkl}^{vrai}| \quad \phi_{hkl}^{calc} \neq \phi_{hkl}^{vrai}$$

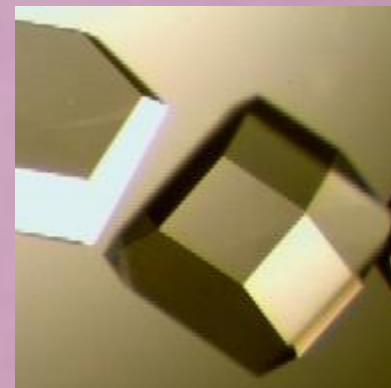
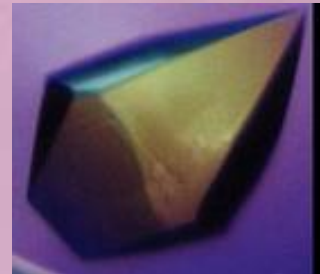
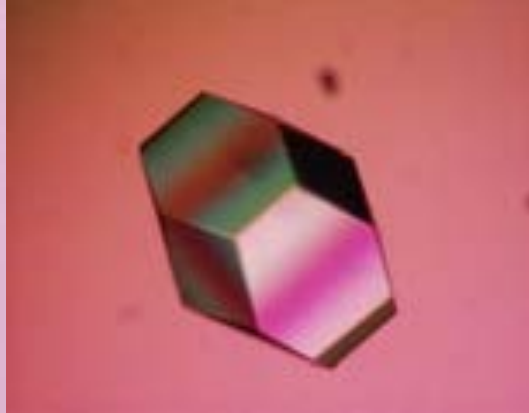
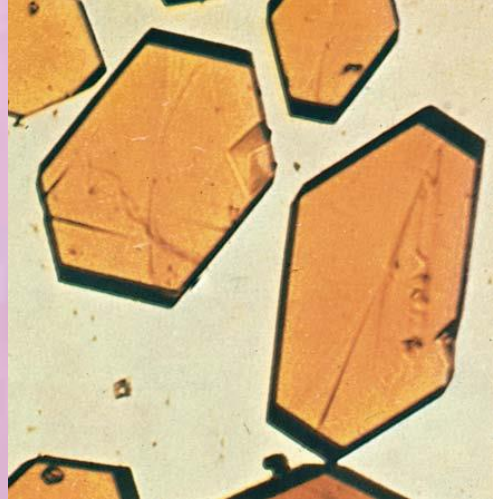
- les erreurs de troncatures

$$D^* \neq \mathbb{Z}^3$$

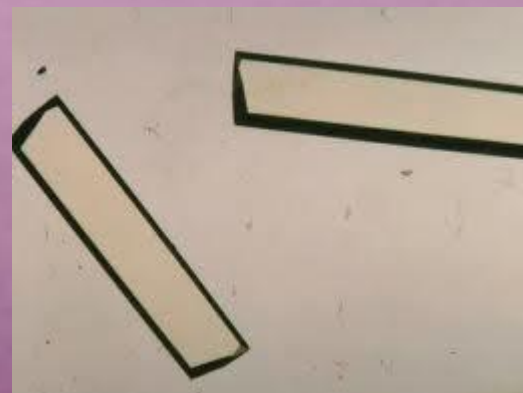
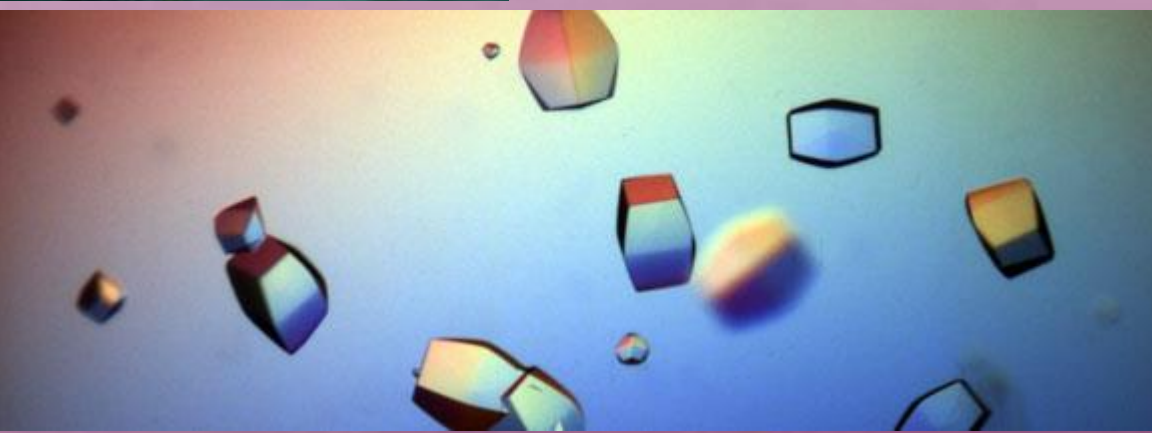
Conséquences des erreurs :

- perte d'interprétabilité de la fonction en terme de positions atomiques
- distance minimale pour que 2 "détails" soient vus séparément, en théorie $\approx 0,7d_{min}$ pratiquement $\approx d_{min}$
- des maxima "élargis" avec des "hauteurs" réduites
- des faux maxima, des minima négatifs
- des superpositions d'extrema (maxima, minima)

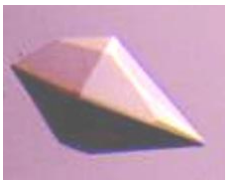
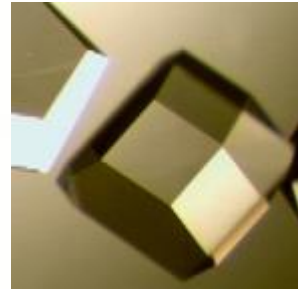
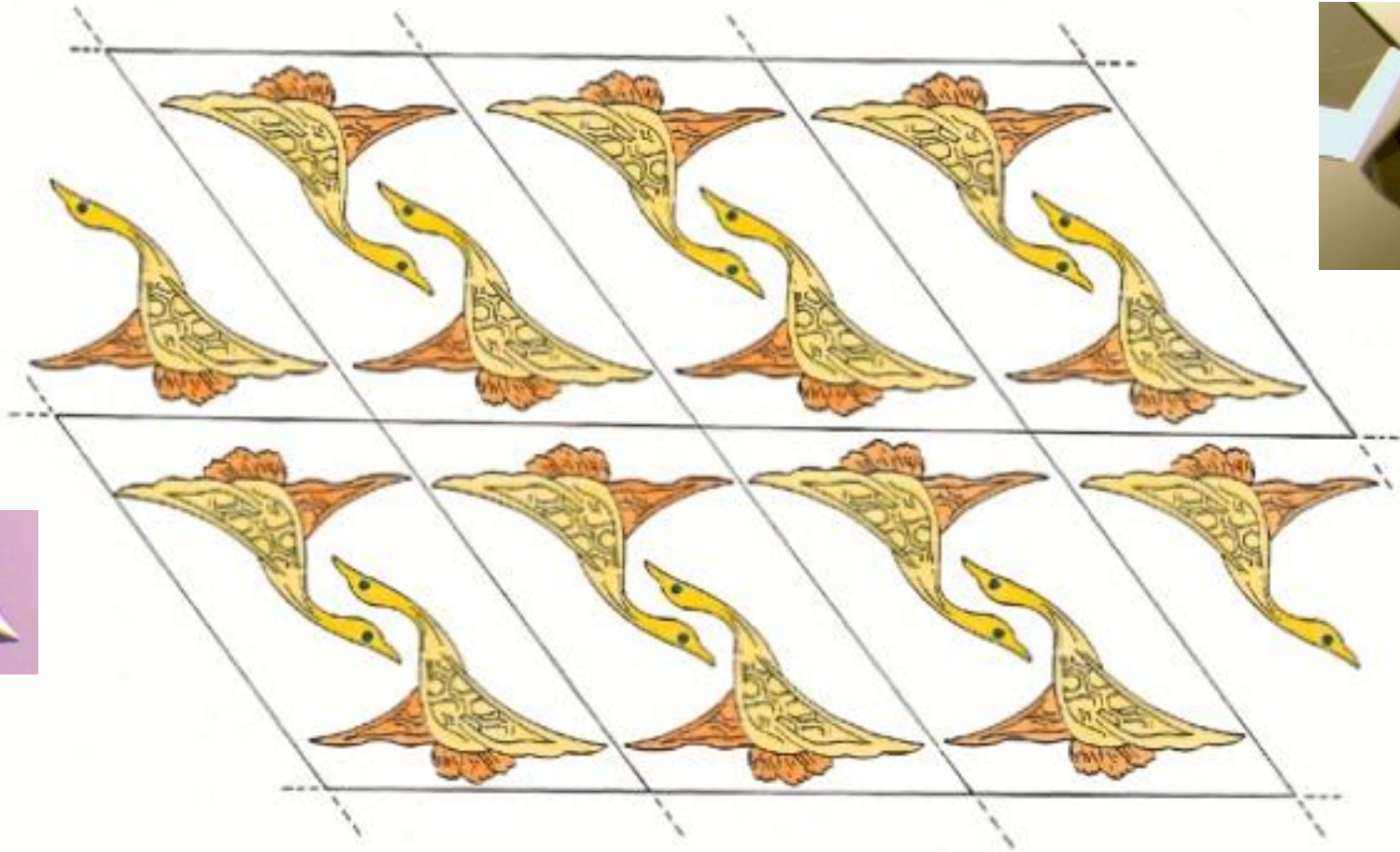




$\sim 10^{13}$ molecules



Un cristal : Arrangement périodique ordonné



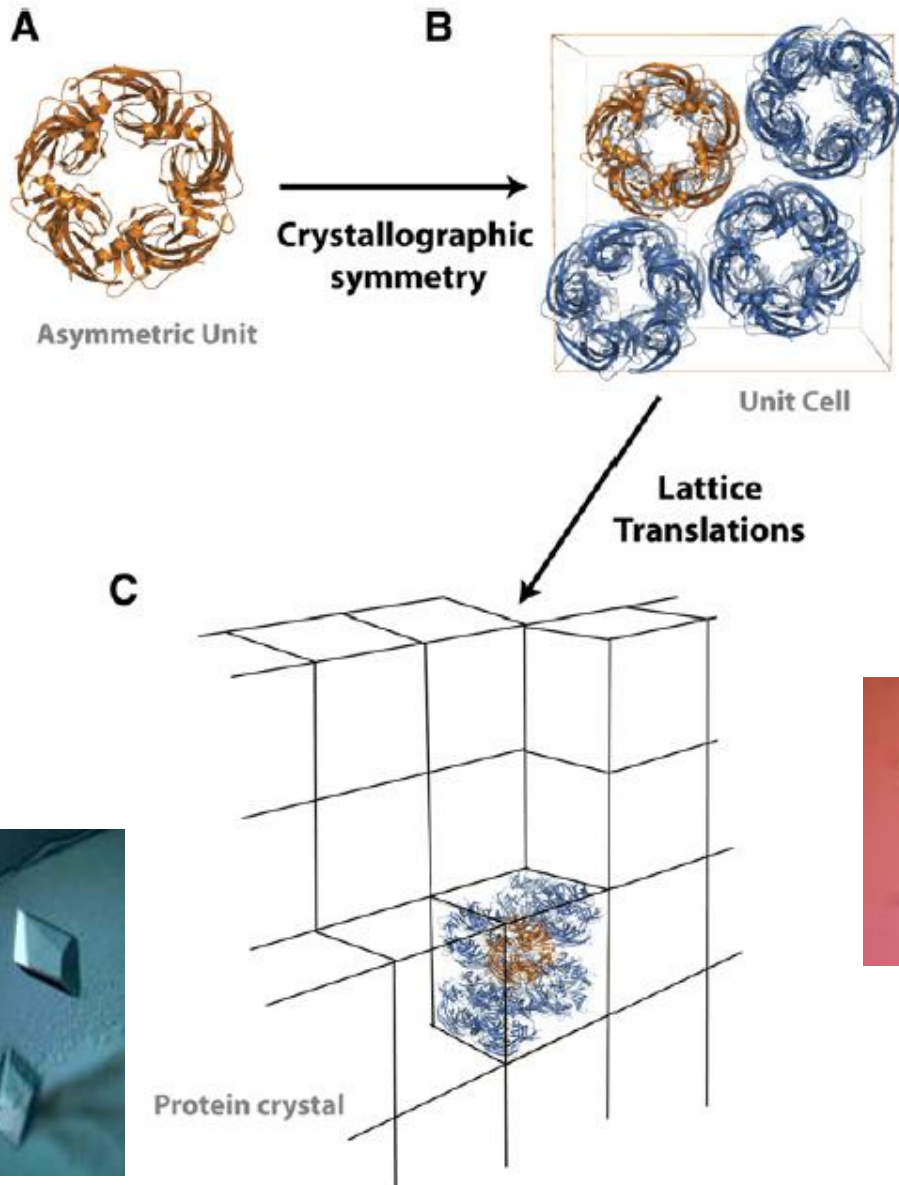
Reproduction ULP Strasbourg. Autorisation CFC-Paris

©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

230 Groupes d'espaces

Maille cristalline caractérisée par l'un des 7 systèmes cristallins

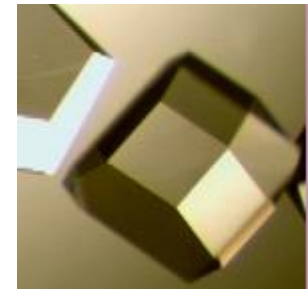
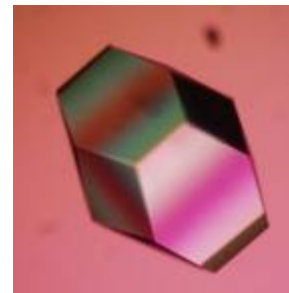




Un cristal :
Arrangement
périodique ordonné

7 systèmes cristallins

230 groupes d'espaces



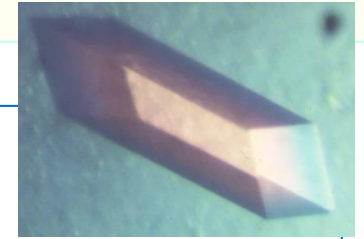
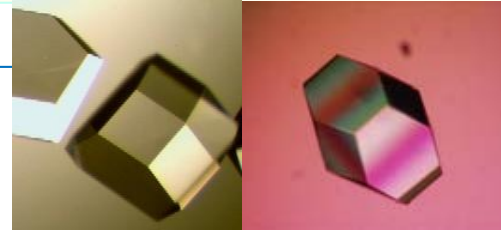
Daniel L. Minor, Jr.
Neuron 54, May 24, 2007 ©2007 Elsevier Inc.

Reproduction ULP Strasbourg. Autorisation CFC-Paris

Symétrie des cristaux

Symétrie macroscopique

- gouverne les formes extérieures des cristaux
- correspond à la **symétrie ponctuelle** ou **symétrie d'orientation**
- Ne rend pas compte de la structure atomique



Travaux de Haüy (1784)

Classification en 7 systèmes : Mohs 1820

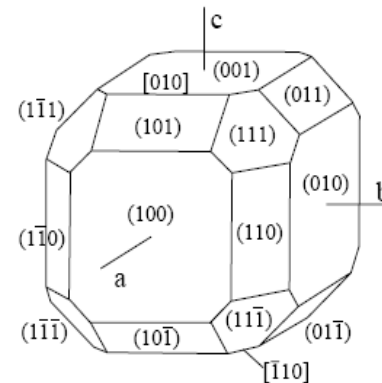
Les 32 groupes ponctuels : Hessel, 1830

Indexation des faces : Miller, 1839



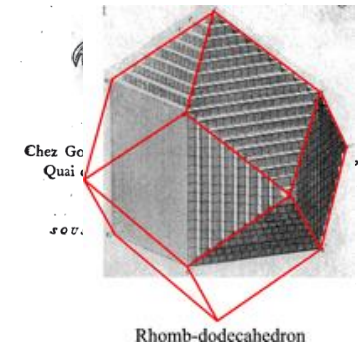
Symétrie microscopique

- Vision atomique, **Symétrie spatiale**
- Bravais : **14 réseaux**, 1848
- **Groupes spatiaux. 230 groupes**
 - Sohncke, 1879, Schoenflies-Fedorov 1891; Mauguin-Hermann, 1933



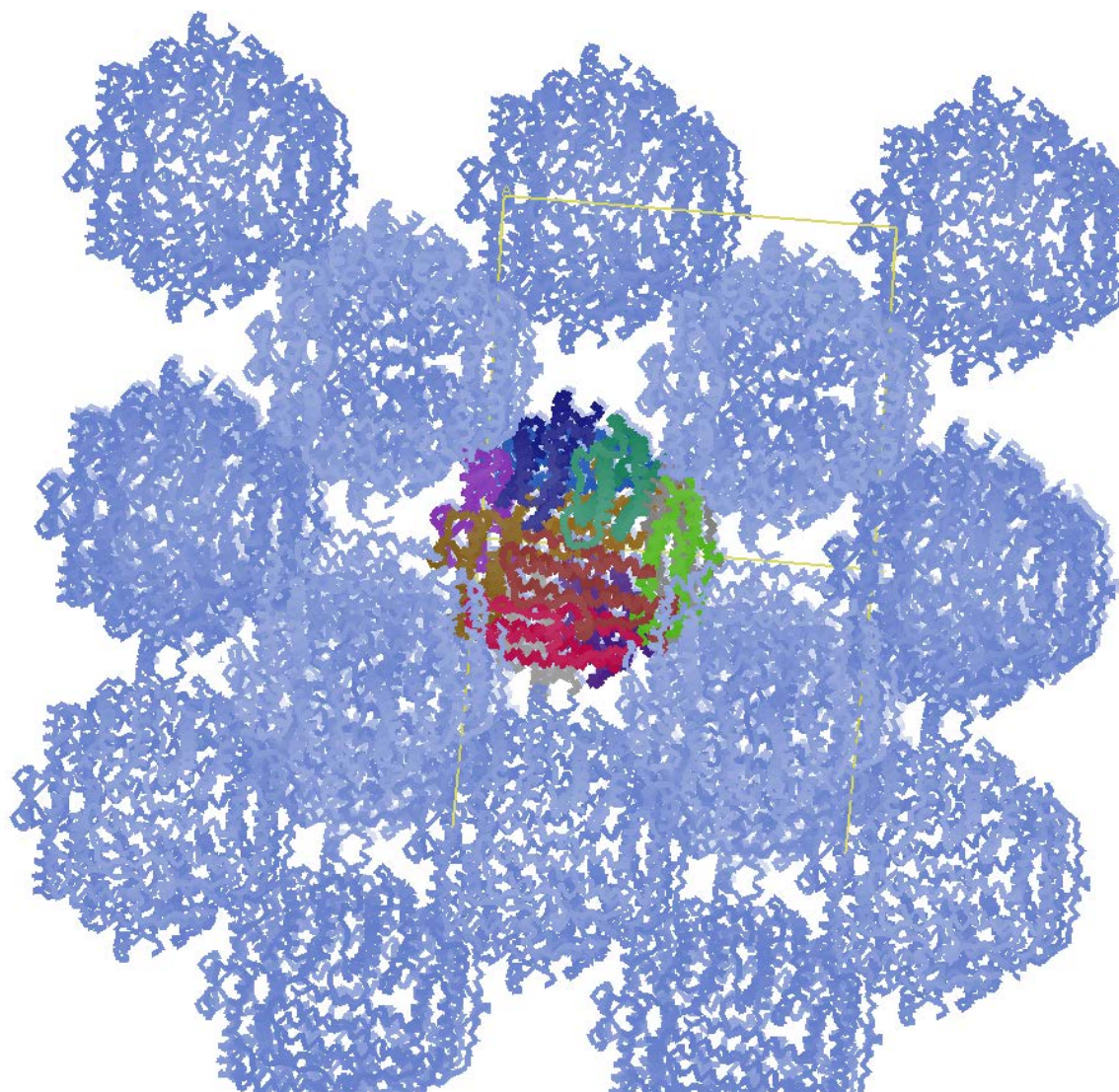
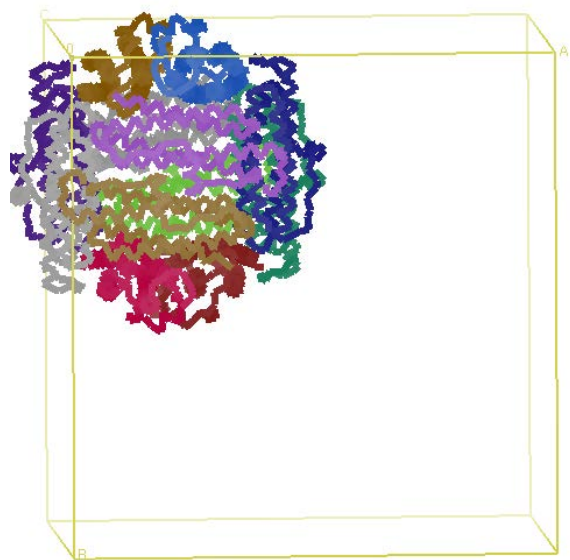
*ESSAI
D'UNE THÉORIE
SUR LA STRUCTURE
DES CRYSTAUX,
APPLIQUÉE
À PLUSIEURS GENRES DE SUBSTANCES
CRYSTALLISÉES;*

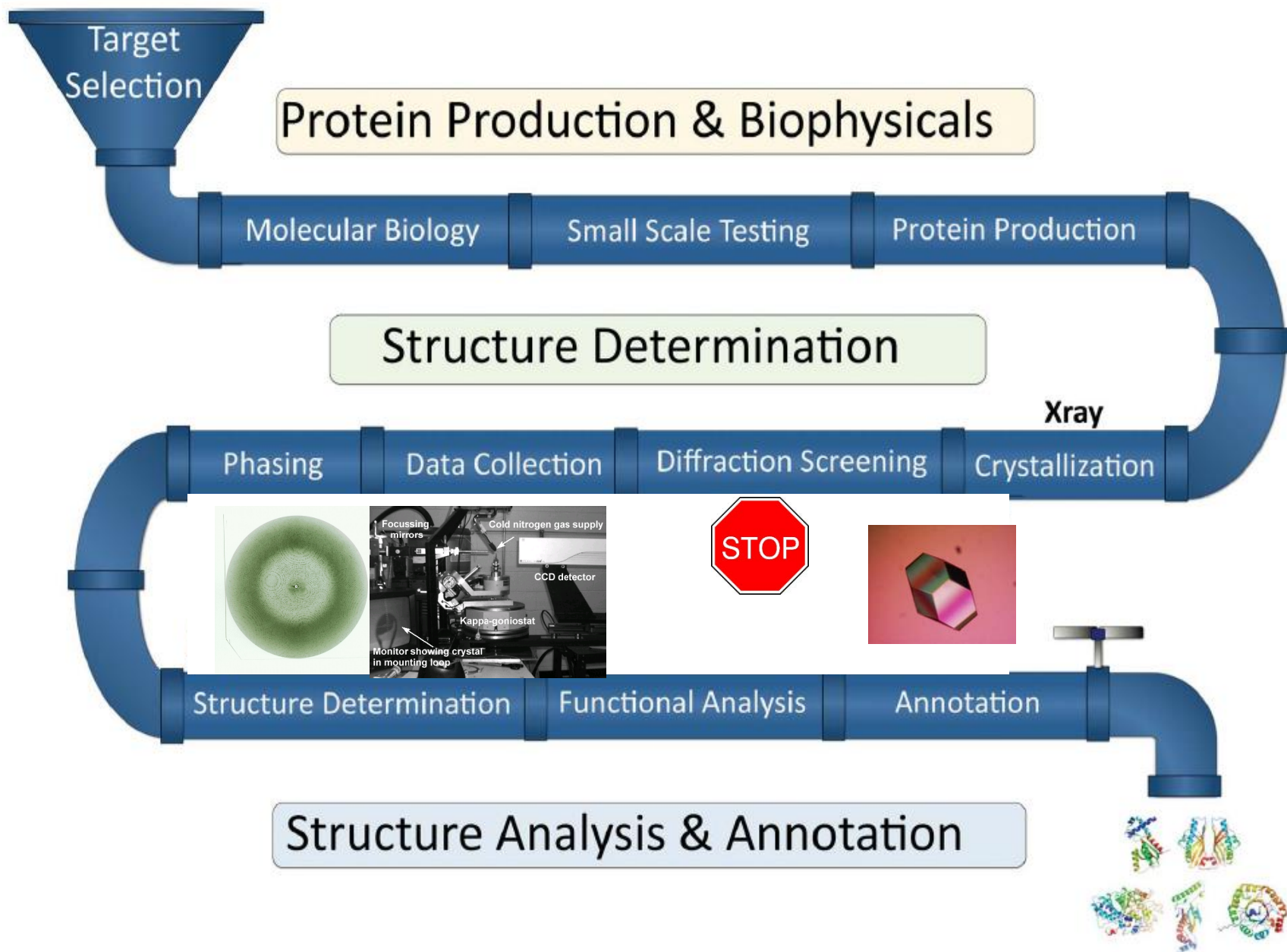
*Par M. l'Abbé HAÛY, de l'Académie Royale des
Sciences, Professeur d'Humanités dans l'Université de
Paris.*



Rhomb-dodecahedron

Système	Symétrie	Groupes ponctuels	Contraintes imposées	Réseaux de Bravais
Triclinique	Identité ou inversion	$1, \bar{1}$	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	P
Monoclinique	1 sym d'ordre 2	$2, m, 2/m$	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$	P, C
Orthorhombique	3 syms d'ordre 2 $\perp$	$222, mm2, mmm$	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	P, C, I, F
Quadratique	axe d'ordre 4	$4, \bar{4}, 4/m$ $422, 4mm, \bar{4}2m$ $4/mmm$	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	P, I
Hexagonal	axe d'ordre 6	$6, \bar{6}, 6/m$ $622, 6mm, \bar{6}2m$ $6/mmm$	$a = b \neq c$ $\alpha = \beta = 90^\circ \neq \gamma = 120^\circ$	P
Trigonal / Rhomboédrique	axe d'ordre 3		$a = b \neq c$ $\alpha = \beta = 90^\circ \neq \gamma = 120^\circ$ axe 3 $\parallel \vec{c}$	P
		$3, \bar{3}$ $32, 3m, \bar{3}m$	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$ axe 3 $\parallel \vec{a} + \vec{b} + \vec{c}$	R
Cubique	4 axes d'ordre 3	$23, m\bar{3}$ $432, \bar{4}3m, m\bar{3}m$	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	P, I, F



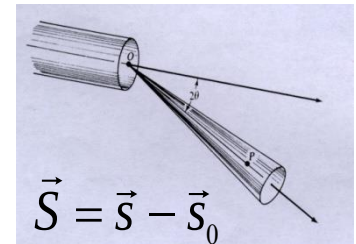


Diffusion élastique cohérente des RX par un atome

chaque atome est caractérisé par une certaine densité électronique $\rho(r)$

En prenant comme unité l'amplitude diffusé par un électron en O,

$$F(\vec{S}) = \oint_V \rho(r) \exp(2i\pi \vec{r} \cdot \vec{S}) dr^3$$



Facteur de diffusion (f_j) d'un atome j

caractérise l'interaction de l'atome j avec les rayons X
(par rapport à un électron libre)

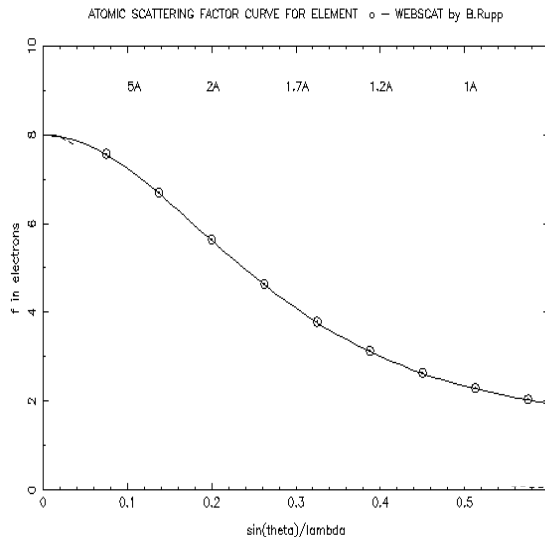
Diffusion élastique cohérente des RX par un atome

facteur de diffusion (f_j) d'un atome j

Diffusion non anormale

- l'énergie des photons X est assez différente de l'énergie de liaison d'un électron dans l'atome, les électrons diffusent comme s'ils étaient libres
- facteur de diffusion (f_j) d'un atome j est un nombre réel

$$f_j(S) = \int 4\pi r^2 \rho(r) \frac{\sin(2\pi r S)}{2\pi r S} dr$$



pour $\frac{\sin \theta}{\lambda} = 0, \quad f_i^0 \cong Z_i$

Approximation polynomiale (Cromer, Waber 1968)

$$f^0\left(\frac{\sin \theta}{\lambda}\right) = \sum_{i=1}^{i=4} a_i e^{-b_i \left(\frac{\sin \theta}{\lambda}\right)^2} + c$$

diffusion anormale :

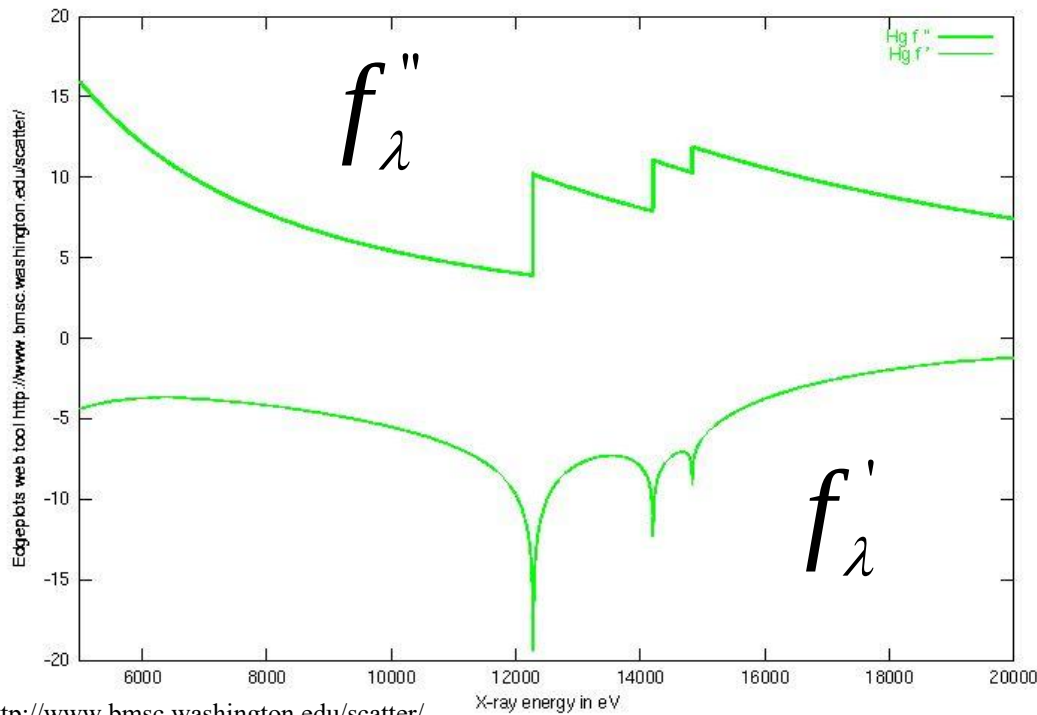
- l'énergie des photons X est très proche de l'énergie de liaison d'un électron dans l'atome,
- effets de résonance entre le photon incident et l'électron importants,

→

- changement de phase et d'amplitude de l'onde diffusée dépendant de la nature de l'atome
 - l'onde diffusée par l'atome n'est plus en phase avec l'onde de référence diffusée par un électron à l'origine.
- facteur de diffusion (f_j) d'un atome j est un nombre complexe

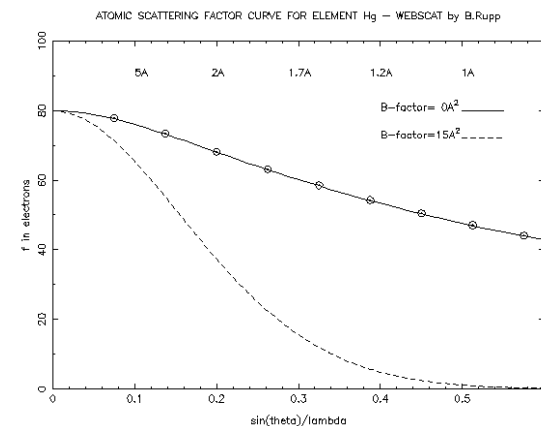
$$f_j(\lambda) = f_j^0 + f_j'(\lambda) + i f_j''(\lambda)$$

$$f(\lambda) = f_0 + f'(\lambda) + i f''(\lambda)$$



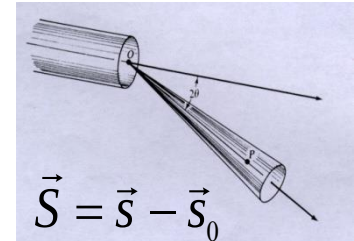
<http://www.bmsc.washington.edu/scatter/>

$$E_{(\text{keV})} \cong \frac{12,3985}{\lambda_{(\text{\AA})}}$$



Diffusion par n atomes

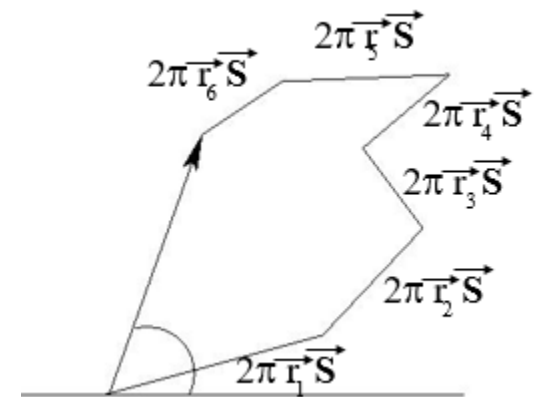
- N atomes caractérisés par :
 - facteur de diffusion f_j ($j=1,n$)
 - position $\vec{r}_j$ ($j=1,n$)



- En prenant comme unité l'amplitude diffusé par un électron en O,

l'onde diffusée par n atomes dans une direction quelconque caractérisée par l'angle 2θ est :

$$F(\vec{S}) = \sum_{j=1}^{j=n} f_j \exp(2i\pi \vec{r}_j \cdot \vec{S})$$



Diffraction par un réseau tridimensionnel d'atomes identiques

- Un réseau caractérisé par les vecteurs $(\vec{a}, \vec{b}, \vec{c})$
- Pour (n_a, n_b, n_c) grands,
 - Interférences constructives si et seulement si

$$\vec{a} \cdot \vec{S} = h, \quad h \in \mathbb{Z}$$

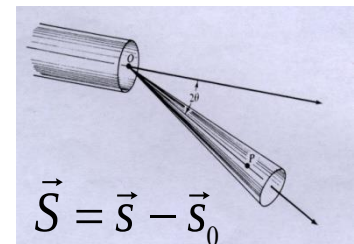
$$\vec{b} \cdot \vec{S} = k, \quad k \in \mathbb{Z}$$

$$\vec{c} \cdot \vec{S} = l, \quad l \in \mathbb{Z}$$

Equations de Laue



Nobel Prize in Physics 1914



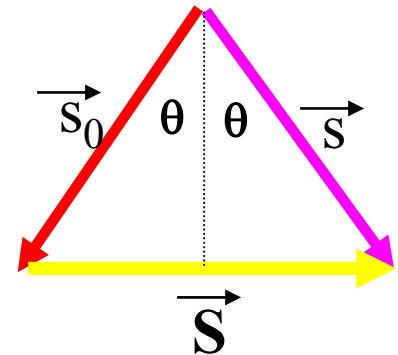
Diffraction par un réseau tridimensionnel d'atomes identiques

$$\vec{a} \cdot \vec{S} = h, \quad h \in \mathbb{Z}$$

$$\vec{b} \cdot \vec{S} = k, \quad k \in \mathbb{Z}$$

$$\vec{c} \cdot \vec{S} = l, \quad l \in \mathbb{Z}$$

Equations de Laue



- Les équations de Laue sont vérifiées si et seulement si

- $\vec{S}$ est un vecteur du réseau réciproque

$$\exists (h, k, l) \in \mathbb{Z}^3 / \vec{S} = h \vec{a}^* + k \vec{b}^* + l \vec{c}^*$$

- On a alors **diffraction dans la direction** $\vec{S} = \vec{S} + \vec{S}_0$



Max von Laue

1879 - 1960

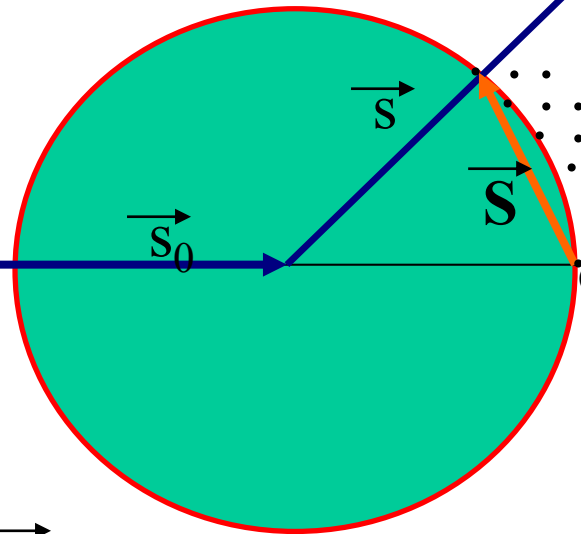
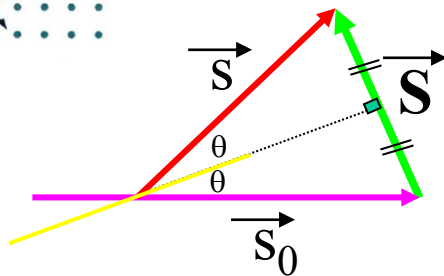
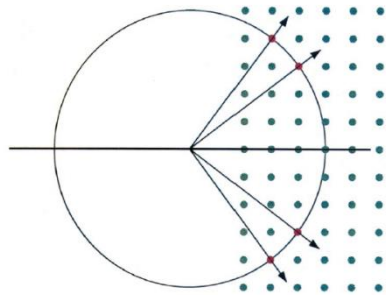
Nobel Prize in Physics 1914

Construction d'Ewald. Sphère d'Ewald

Interprétation géométrique des équations de Laue

$$\vec{S} = \vec{s} - \vec{s}_0 \quad |\vec{s}_0| = |\vec{s}| = \frac{1}{\lambda} \quad \vec{S} = h \vec{a}^* + k \vec{b}^* + l \vec{c}^*$$

$$|\vec{S}| = 2 \frac{\sin \theta}{\lambda}$$



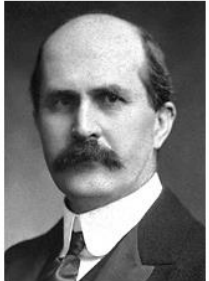
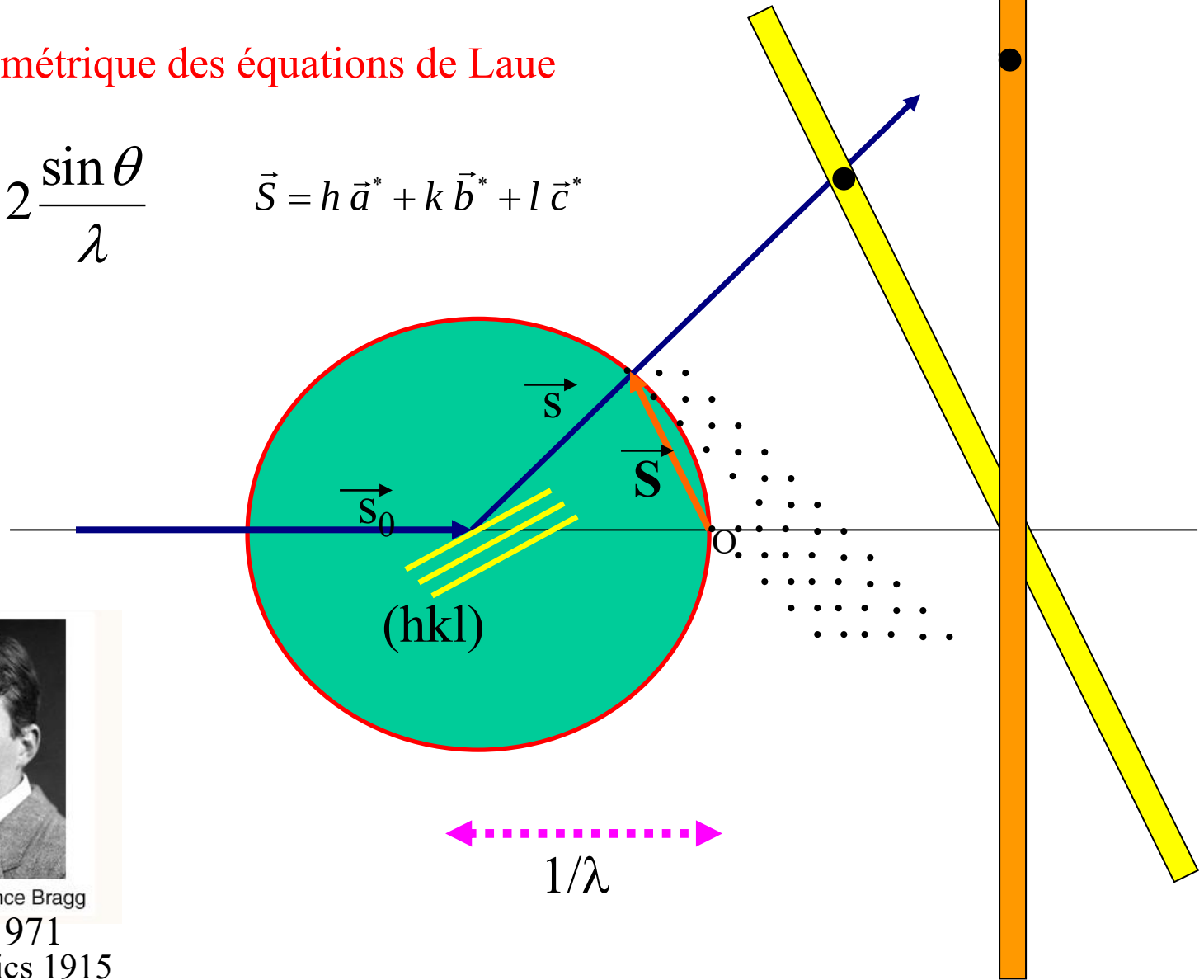
Paul Peter Ewald
1888 - 1985

détecteur

Interprétation géométrique des équations de Laue

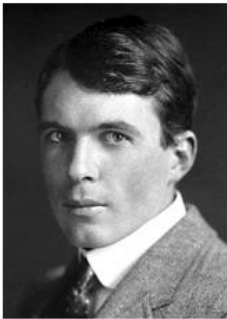
$$|\vec{S}| = \frac{1}{d_{hkl}} = 2 \frac{\sin \theta}{\lambda}$$

$$\vec{S} = h \vec{a}^* + k \vec{b}^* + l \vec{c}^*$$



William Henry Bragg

1862 - 1942



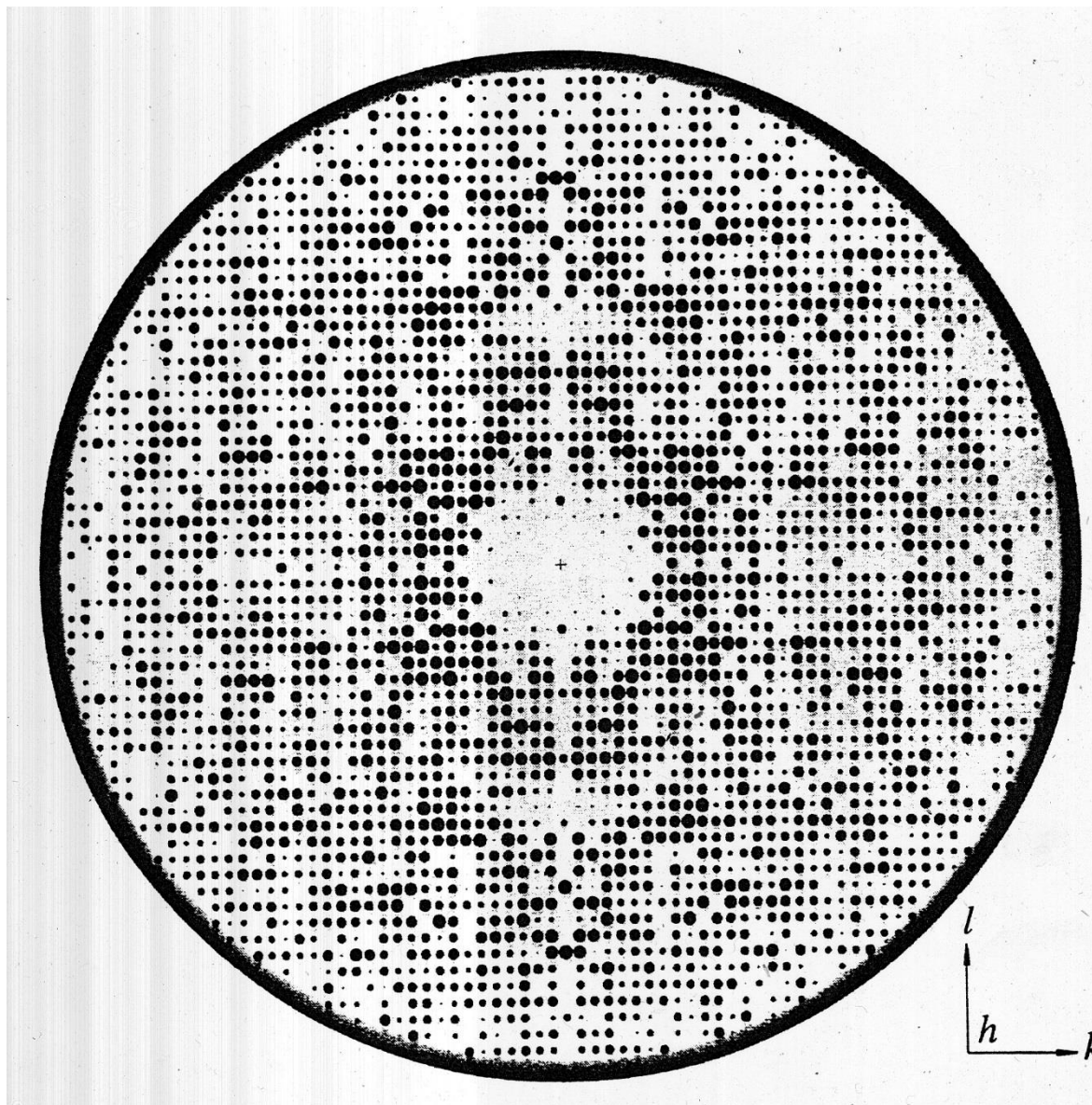
William Lawrence Bragg

1890 - 1971

Nobel Prize in Physics 1915

Loi de Bragg

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



N_{ov}

$$\lambda = 1.22 \text{ \AA}$$

$$a = 411 \text{ \AA} \quad b = 410 \text{ \AA}$$

$$c = 420 \text{ \AA}$$

$$\alpha = 59^\circ \quad \beta = 58,9^\circ$$

$$\gamma = 64^\circ$$

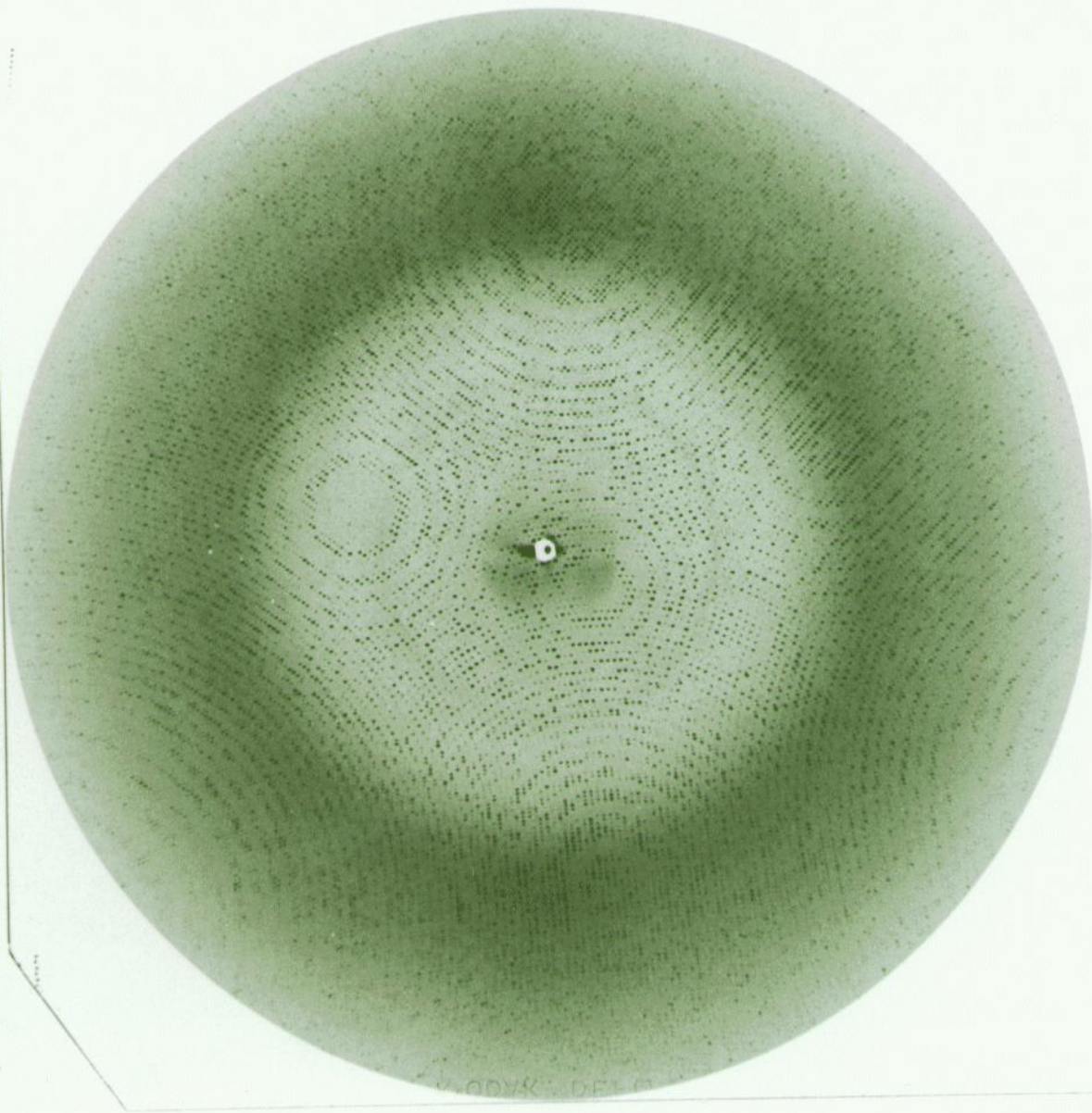
$$\Delta\phi = 0.3^\circ$$

$$DF = 120 \text{ mm.}$$

15245 réflexions

$$2.7 \text{ \AA}$$

Brookhaven (USA),
1988.



Facteurs de Structure d'un cristal

$$\sum_{j=1}^{j=n} f_j \exp(2i\pi \vec{r}_j \cdot \vec{S})$$

n_1 mailles selon $\vec{a}$, n_2 mailles selon $\vec{b}$, n_3 mailles selon $\vec{c}$

$$G(\vec{S}) = F(\vec{S}) L(\vec{S})$$

- $F(\vec{S})$ **facteur de structure (pour une maille)**
 - Dépend du contenu atomique de la maille, rend compte du déphasage entre atomes de la maille.
 - Indépendant de la taille du cristal.
- $L(\vec{S})$ **facteur de forme**
 - Amplitude diffusé par le cristal, en prenant comme unité celle diffusée par une maille. Dépend des dimensions ou de la forme du cristal, rend compte du déphasage entre les mailles.
 - Indépendant du contenu atomique de la maille.

$$L(\vec{S}) = \sum_{u=0}^{u=n_1-1} \sum_{v=0}^{v=n_2-1} \sum_{w=0}^{w=n_3-1} \exp(2i\pi(u\vec{a}.\vec{S} + v\vec{b}.\vec{S} + w\vec{c}.\vec{S}))$$

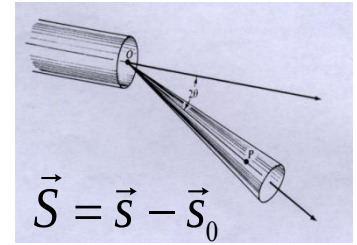
$$L(\vec{S}) = \sum_{u=0}^{u=n_1-1} \exp(2i\pi u\vec{a}.\vec{S}) \sum_{v=0}^{v=n_2-1} \exp(2i\pi v\vec{b}.\vec{S}) \sum_{w=0}^{w=n_3-1} \exp(2i\pi w\vec{c}.\vec{S})$$

$$|L(\vec{S})|^2 = \left(\frac{\sin(n_1\pi\vec{a}.\vec{S})}{\sin(\pi\vec{a}.\vec{S})} \right)^2 \left(\frac{\sin(n_2\pi\vec{b}.\vec{S})}{\sin(\pi\vec{b}.\vec{S})} \right)^2 \left(\frac{\sin(n_3\pi\vec{c}.\vec{S})}{\sin(\pi\vec{c}.\vec{S})} \right)^2$$

Pour un cristal de dimension infinie

$$L(\vec{S}) \neq 0 \quad \text{si et seulement si} \quad \vec{S} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$$

$$G(\vec{S}) \perp n_1^2 n_2^2 n_3^2 F(\vec{S})$$



Pour un cristal de dimension "infinie"

$$G(\vec{S}) \propto F(\vec{S}) \neq 0$$

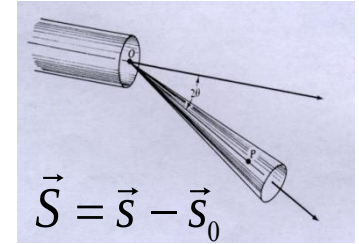
$$\text{si et seulement si} \quad \vec{S} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$$

$\{F(\vec{S})\}$ facteurs de structure

- dépend du contenu atomique de la maille
- rend compte du déphasage entre atomes de la maille
- Indépendant de la taille du cristal

Facteur de structure

- N atomes dans la maille
- $(f_j, \vec{r}_j), j = 1, N)$



$$F(\vec{S}) = \sum_{j=1}^{j=N} f_j \exp(2\pi i \vec{S} \cdot \vec{r}_j)$$

$$\vec{r}_j = x_j \vec{a} + y_j \vec{b} + z_j \vec{c} \quad 0 \leq x_j, y_j, z_j < 1$$

$$\vec{S} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$$

$$F(\vec{S}) = \sum_{j=1}^{j=N} f_j \exp(2\pi i (hx_j + ky_j + lz_j)) = F_{hkl}$$

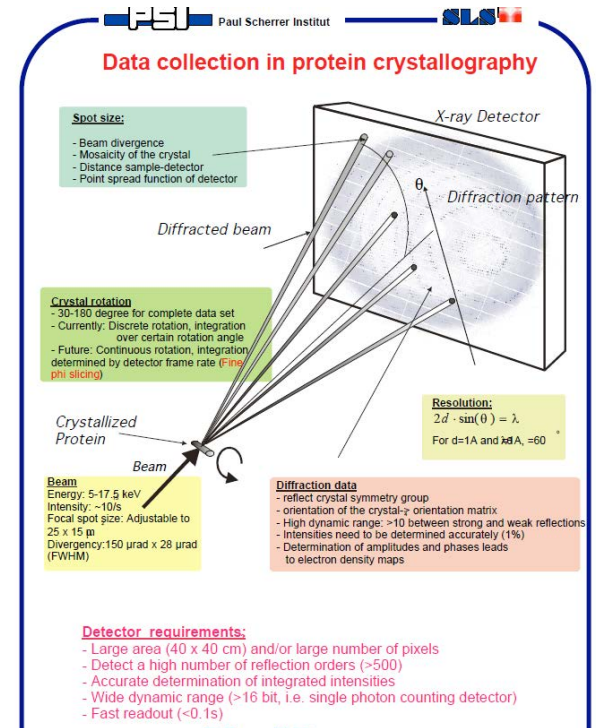
Limite de diffraction du cristal

$$|\vec{S}| = |\vec{S}_{\max}|$$

Attention, on utilise dans la pratique

$$d_{\min} = \frac{1}{|\vec{S}_{\max}|}$$

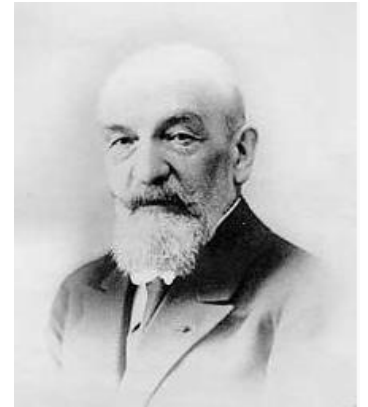
La limite de diffraction n'est pas un critère fixé par l'utilisateur, mais dans les conditions les plus optimales, elle reflète la **qualité intrinsèque (ordre) d'un cristal**



Relations entre facteurs de structures

En absence de diffuseur anomal

Loi de Friedel (Georges Friedel)



1865-1933

$$F_{hkl} = \sum_{j=1}^{j=N} f_j \exp(2\pi i(hx_j + ky_j + lz_j))$$

$$F_{\bar{h}\bar{k}\bar{l}} = \bar{F}_{hkl}$$

$$\left| F_{\bar{h}\bar{k}\bar{l}} \right| = \left| \bar{F}_{hkl} \right| \quad \text{et} \quad \phi_{\bar{h}\bar{k}\bar{l}} = -\phi_{hkl}$$

Pour un groupe d'espace donné,

- Espace direct : Opérateurs de symétrie :

$$C_s : (R_s, T_s), (s=1, n)$$

$$\vec{r}_{js} = R_s \vec{r}_j + T_s$$

- Espace réciproque : Contraintes correspondantes sur les facteurs de structure

$$F_{R_s^t H} = F_H \exp(-2i\pi H^t T_s)$$

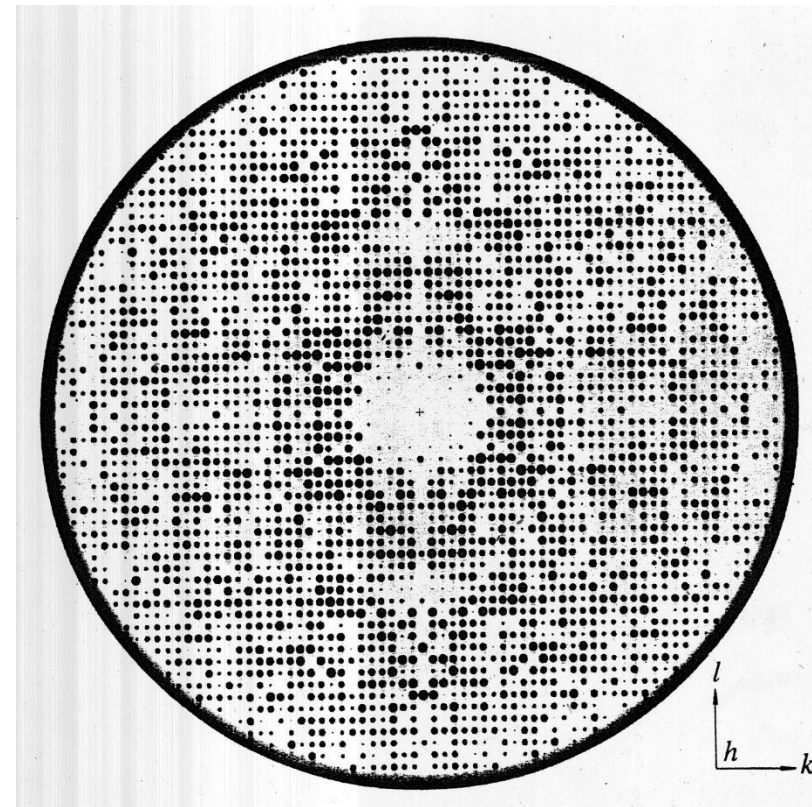
$$\left| F_{R_s^t H} \right| = \left| F_H \right| \quad \text{et} \quad \phi_{R_s^t H} \equiv \phi_H - 2\pi H^t T_s \quad [2\pi]$$

$$F_{R_S^t H} = F_H \exp(-2i\pi H^t T_S)$$

$$\left| F_{R_S^t H} \right| = \left| F_H \right| \quad \phi_{R_S^t H} = \phi_H - 2\pi H^t T_S$$

Réflexions équivalentes

Extinctions systématiques



Exemple. Groupe $P6_1$

- Positions équivalentes

x, y, z

$-y, x-y, z+1/3$

$y-x, -x, z+2/3$

$-x, -y, z+1/2$

$y, y-x, z+5/6$

$x-y, x, z+1/6$

- Réflexions équivalentes en absence de diffuseur anomal

hkl

$-h, -k, -l$

$k, (-h - k), l$

$-k, (h + k), -l$

$(-h - k), h, l$

$(h + k), -h, -l$

$-h, -k, l$

$h, k, -l$

$-k, (h + k), l$

$k, (-h - k), -l$

$(h + k), -h, l$

$(-h - k), h, -l$

Contraintes sur les facteurs de structure : Extinctions systématiques

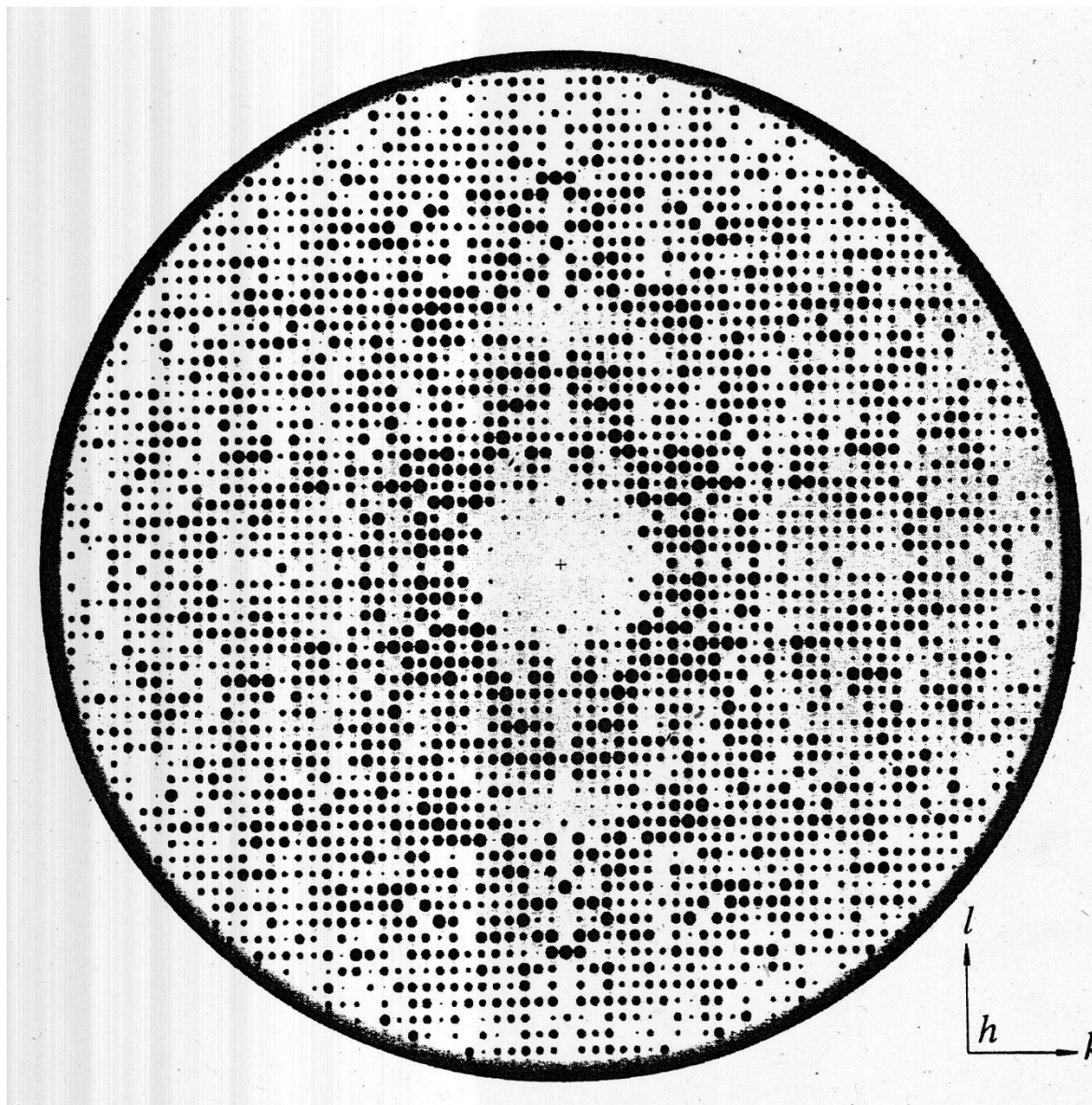
$$F_{R_S^t H} = F_H \exp(-2i\pi H^t T_S)$$

$$\left| F_{R_S^t H} \right| = \left| F_H \right| \quad \phi_{R_S^t H} = \phi_H - 2\pi H^t T_S$$

$$\text{si } H = R_S^t H \quad \text{et} \quad H^t T_S \text{ non entier} \Rightarrow \left| F_H \right| = 0$$

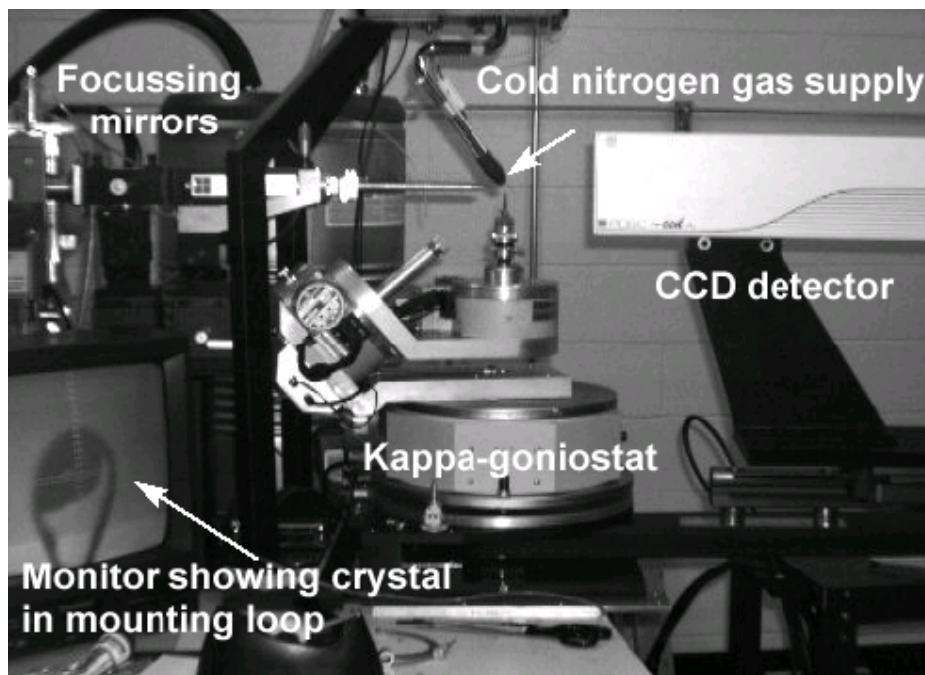
- **axes hélicoïdaux :**
 - extinctions à périodicité unidimensionnelle : situés sur une rangée particulière du réseau réciproque
- **plans de symétrie translatatoires :**
 - extinctions à périodicité bidimensionnelle, situés sur un plan particulier du réseau réciproque
- **modes de réseau :**
 - extinctions à périodicité tridimensionnelle.

Réflexions affectées	Conditions d'existence	Éléments de symétrie
h00	$h=2n$	$2_1 // a, 4_2 // a$
0k0	$k=2n$	$2_1 // b, 4_2 // b$
00l	$l=2n$	$2_1 \text{ ou } 4_2 \text{ ou } 6_3 // c$
	$l=3n$	$3_1 \text{ ou } 3_2 \text{ ou } 6_2 \text{ ou } 6_4 // c$
	$l=4n$	$4_1 \text{ ou } 4_3 // c$
	$l=6n$	$6_1 \text{ ou } 6_5 // c$
hkl	$h+k=2n$	Mode C
	$h+l=2n$	Mode B
	$k+l=2n$	Mode A
	$h+k+l=2n$	Mode I
	$h,k,l, \text{ tous pairs ou impairs}$	Mode F



Collecte des données de diffraction

1. cristal
2. source de Rayons X (cours Jean-Luc Ferrer)
3. détecteur
4. Méthode de collecte des données
5. Traitements des données



www.esrf.fr

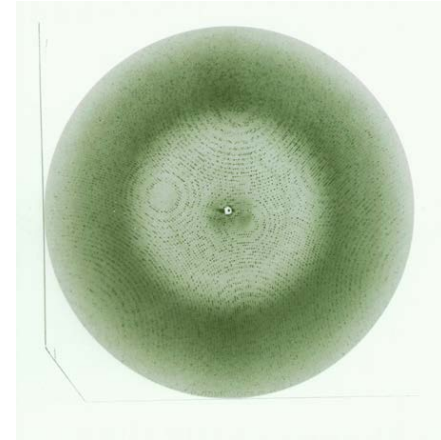
Reproduction ULP Strasbourg. Autorisation CFC-Paris

Collecte des données de diffraction

$$I_{hkl}(\lambda) = K \frac{V_c}{V^2} I_0(\lambda) C(\lambda, r) \lambda^3 |F_{hkl}(\lambda)|^2$$

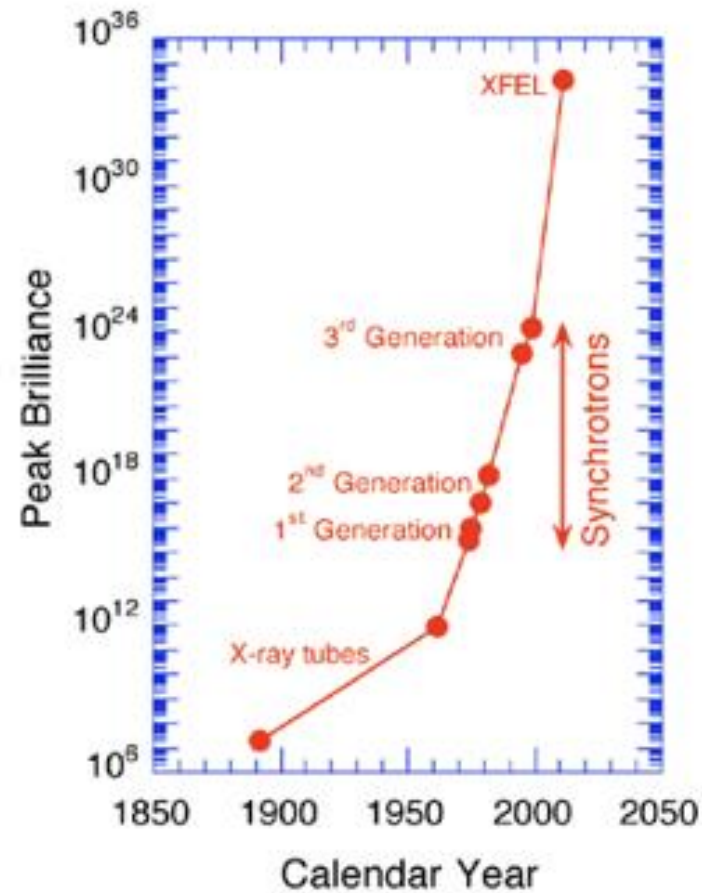
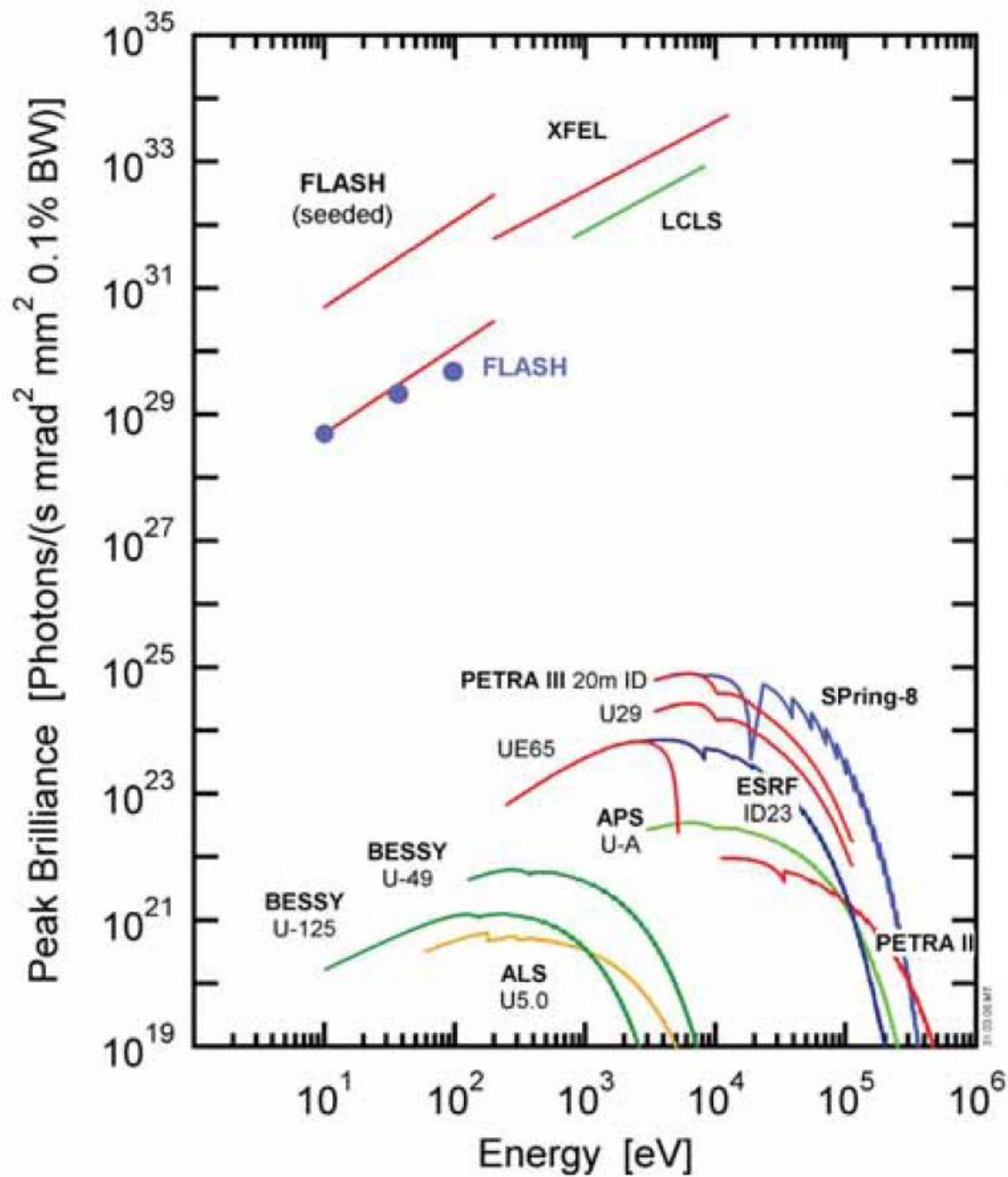
Spécificité des macromolécules biologiques

- Intensités faibles
- Réseaux réciproques fortement peuplés
- Dégradation du cristal dans faisceau de rayons X



Des progrès **technologiques** et **méthodologiques** à toutes les étapes du processus

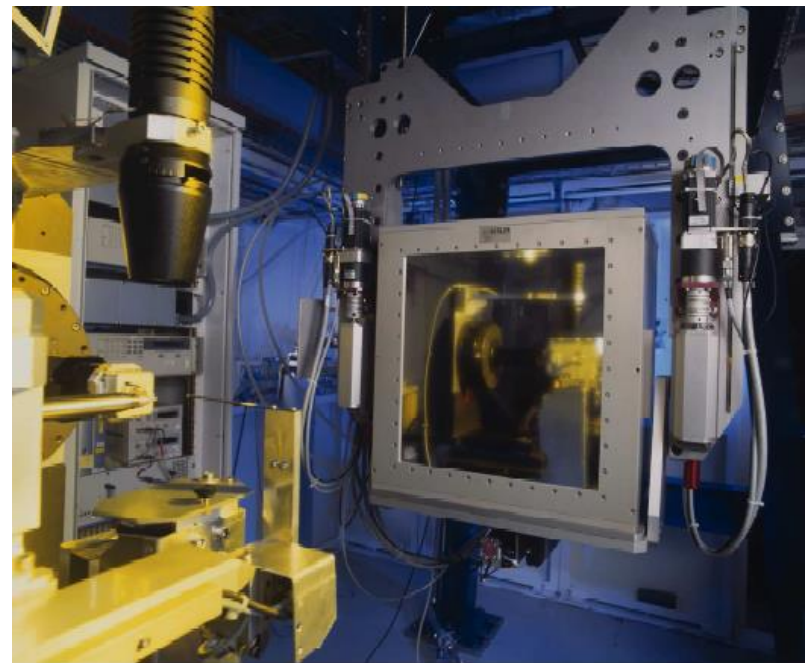
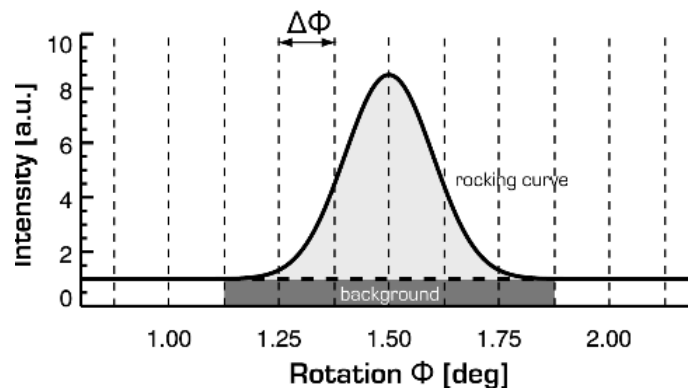
- Source de Rayons X
- Détecteurs
- Méthodes de Collecte et de traitement des données
- Préparation des cristaux



PILATUS La révolution de ces dernières années

Pixel ApparATUS for the Swiss Light Source Paul Scherrer Institut (PSI), CH-Villigen

- Qualités
 - Temps de lecture : quelques ms
 - Pas de bruit de fond
 - Grande dynamique
 - Collecte en continu possible
 - 0.02° - 0.2° , exposition 100ms
 - Lecture 3ms sans arrêt



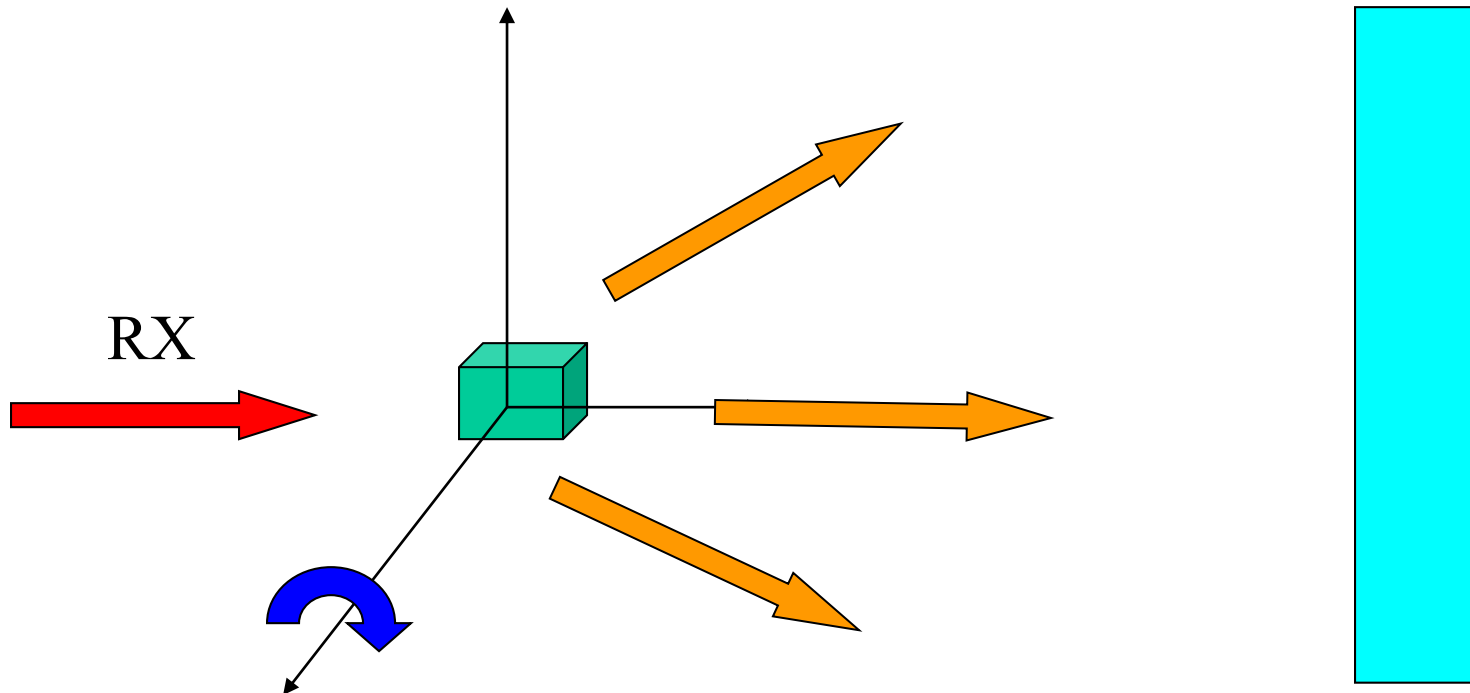


Technical specifications

EIGER X	1M	4M	9M	16M
Number of detector modules	1 × 2	2 × 4	3 × 6	4 × 8
Sensitive area, width × height [mm ²]	77.2 × 79.9	155.2 × 162.5	233.2 × 245.2	311.2 × 327.8
Pixel size [μm ²]	75 × 75			
Total number of pixels	1030 × 1065 = 1,096,950	2070 × 2167 = 4,485,690	3110 × 3269 = 10,166,590	4150 × 4371 = 18,139,650
Gap width, hor. / ver. [pixel]	– / 37	10 / 37	10 / 37	10 / 37
Inactive area [%]	3.5	5.6	6.3	6.6
Defective pixels [%]	< 0.03			
Maximum frame rate* [Hz]	3000	750	238	133
Readout time	continuous readout, 3 μs dead time, duty cycle > 99 %			
Point-spread function [pixel]	1			
Sensor thickness [μm]	450			
Threshold energy [keV]	2.7 - 18			
Maximum count rate [phts/s/mm ²]	5 · 10 ⁸			
Counter bit depth [bit]	12			
Image bit depth [bit]	16 or 32			
Data format	HDF5 / NeXus			
Dimensions (WHD) [mm ³]	114 × 133 × 240	235 × 237 × 372	340 × 370 × 500	400 × 430 × 500
Weight [kg]	3.9	15	41	55
Power consumption [W]	75	300	750	1200
* Theoretical limit at maximum bandwidth				

Méthode d'oscillation

- Donner au cristal un **mouvement de rotation de faible** amplitude autour d'un axe
 - Amplitude de la rotation ($\Delta\phi$) limitée par la superposition de réflexions différentes en un même endroit du détecteur
- L'espace complet de diffraction est enregistré en une succession de N "clichés", correspondant chacun à une oscillation $\Delta\phi$



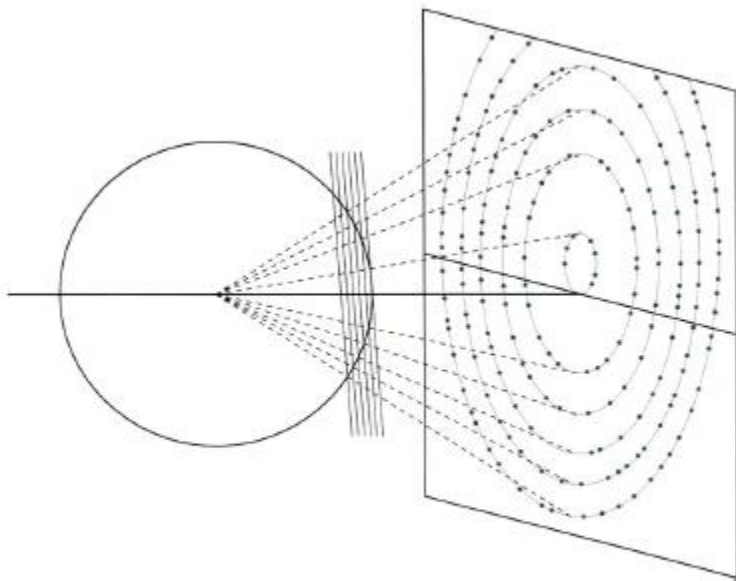


Figure 3
A still exposure with a stationary crystal contains only a small number of reflections arranged in a set of narrow ellipses.

© Zauter, (1997) ,Methods in Enzymology, Vol 276
Reproduction ULP Strasbourg. Autorisation CFC-Paris

© Zauter, (1999) , Acta Cryst., D55, 1703-1717
Reproduction ULP Strasbourg. Autorisation CFC-Paris

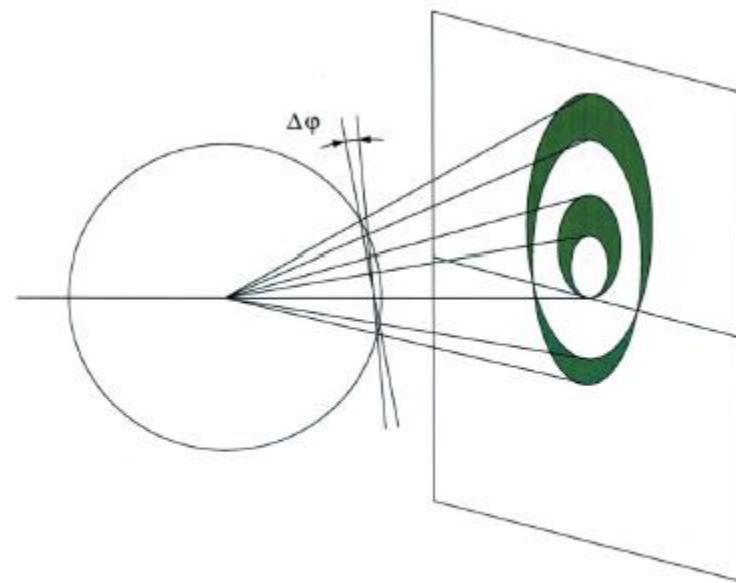
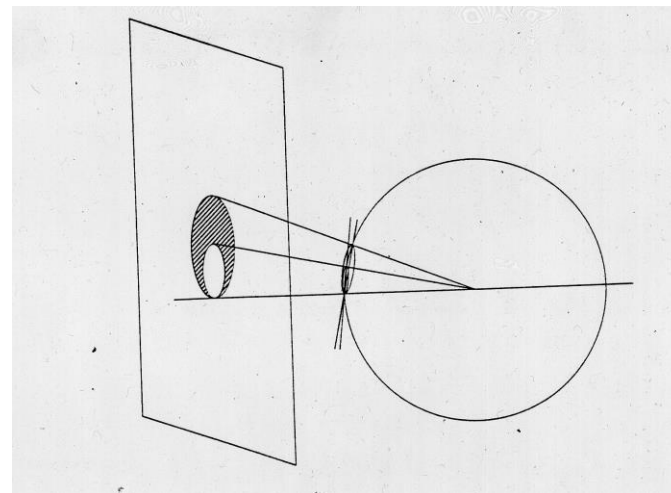
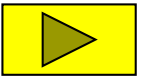
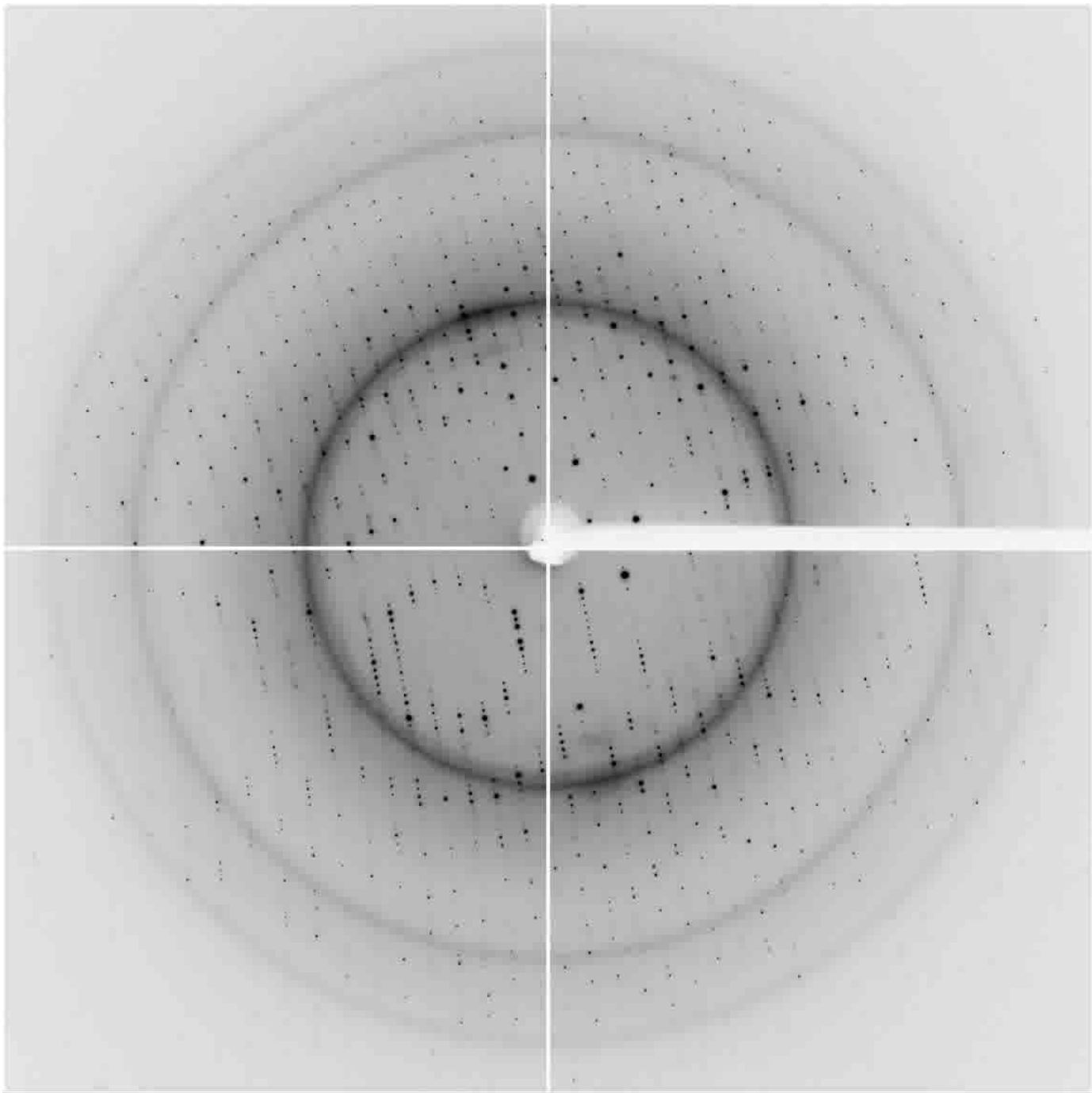


Figure 4
When the crystal is rotated, reflections from the same plane in the reciprocal lattice form a lune, limited by two ellipses corresponding to the start and end positions.

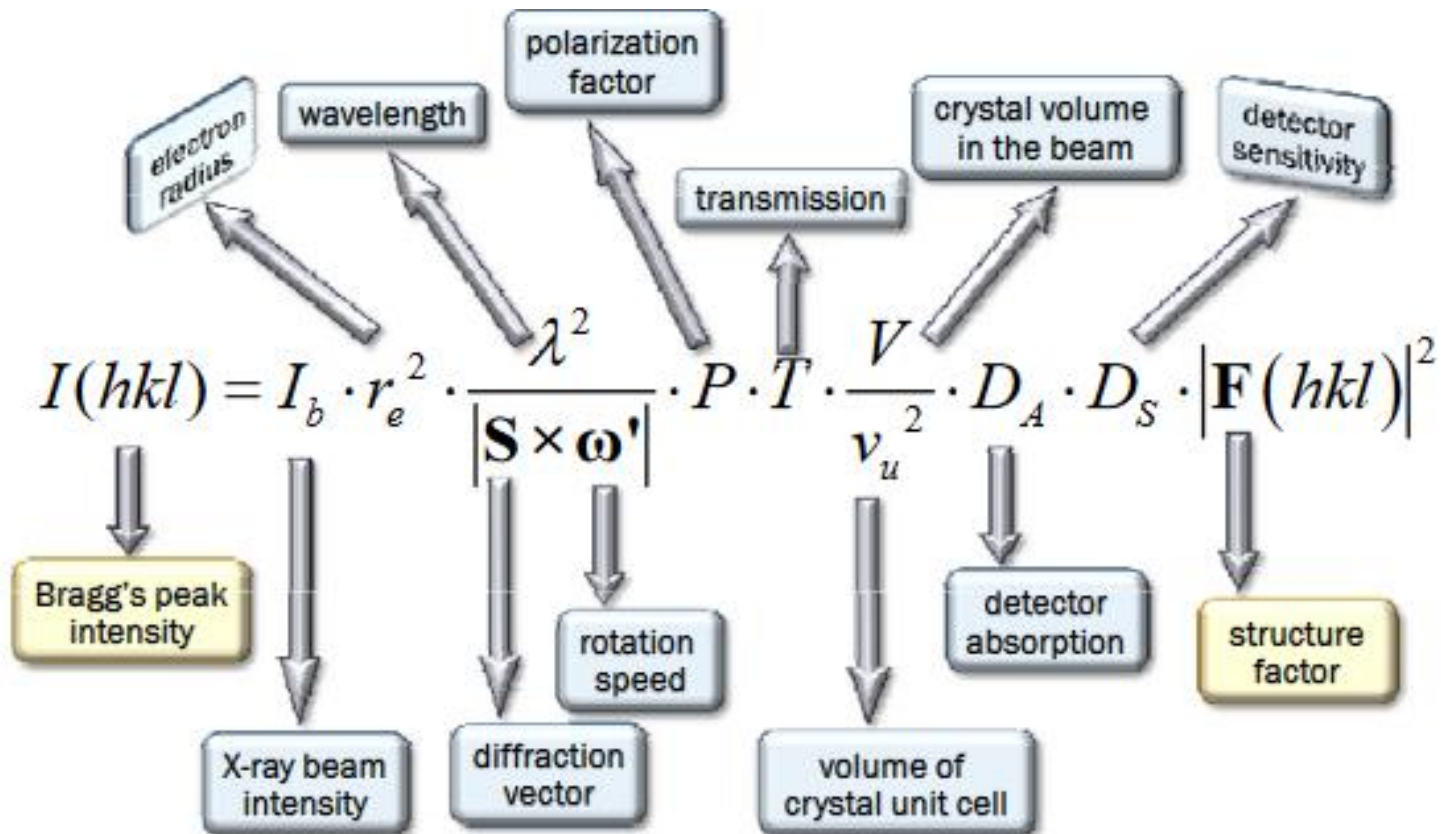


Méthode d'oscillation

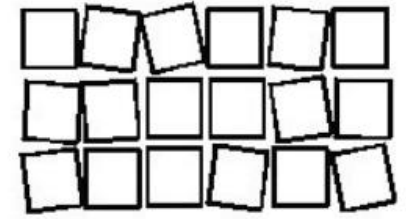
Enregistrement du signal pendant toute la durée du passage du volume à travers la sphère d'Ewald

$$I_{hkl} = \oint \text{énergie reçue}$$

Pour un cristal parfaitement mosaïque

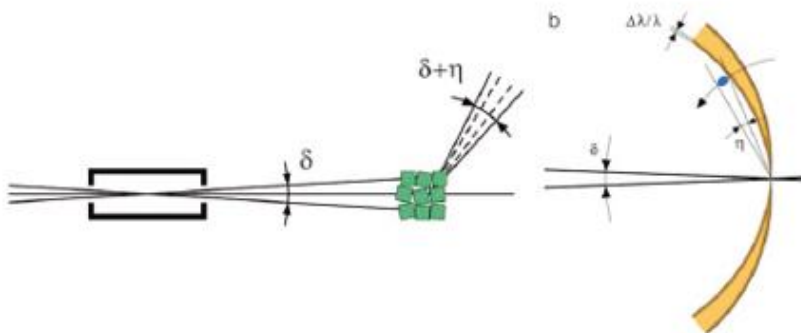


Méthode d'oscillation

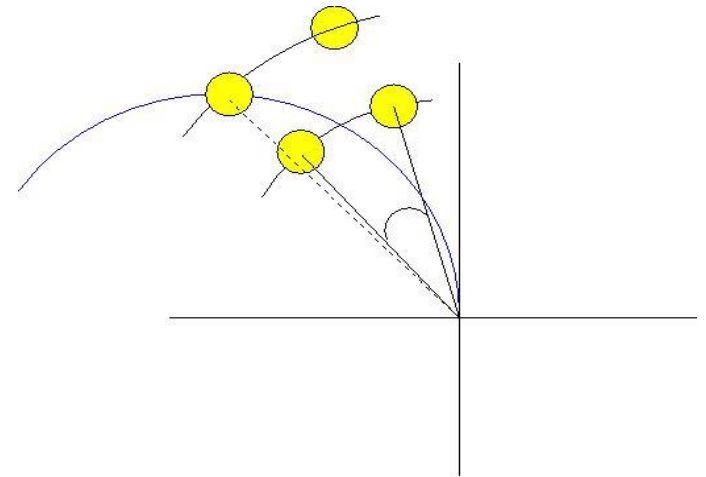


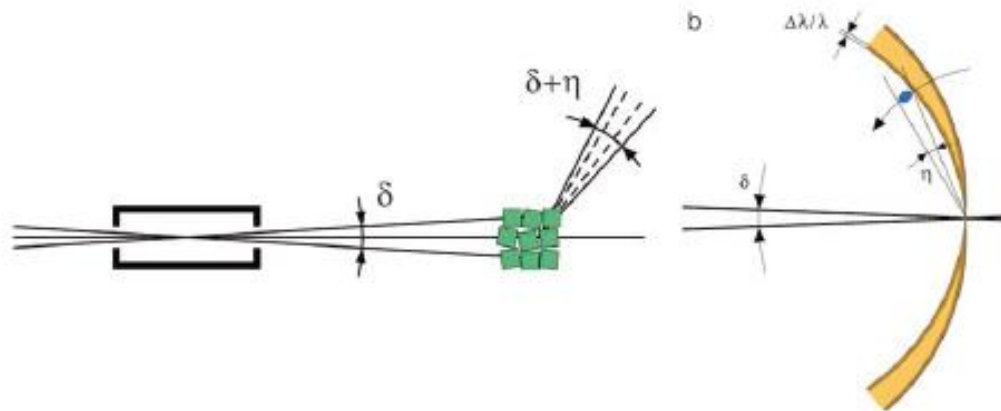
- Les réflexions sont en position de diffraction sur un certain domaine angulaire qui dépend :
 - **mosaicité du cristal**
 - organisation du cristal en petits domaines cristallins désorientés les uns par rapport aux autres
 - **caractéristiques du faisceau de rayons X incidents**
 - divergence, dispersion en longueur d'onde, optique
- Modèle simple. Volume sphérique, de rayon ε

$$\varepsilon \approx (1/2) \Delta d^* \cos\theta$$

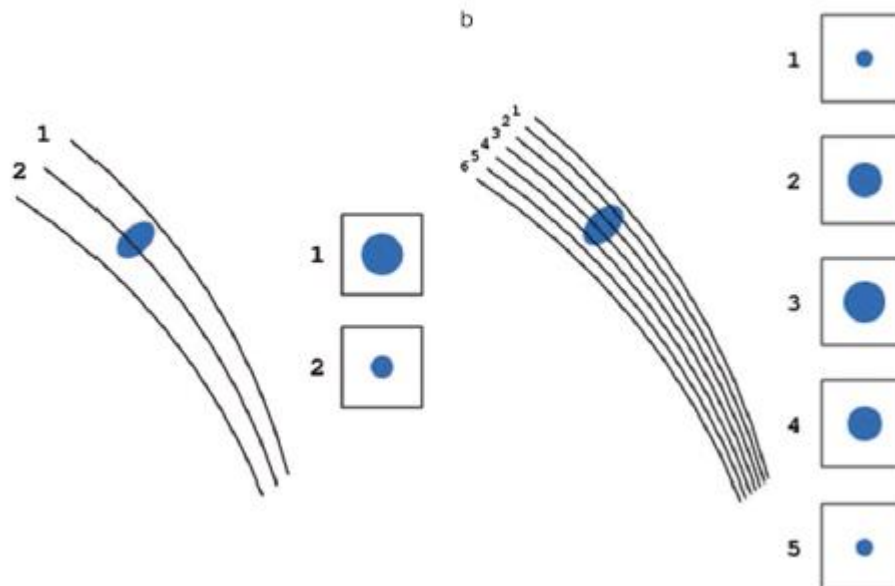


A schematic representation of the beam divergence (δ), crystal mosaicity (η) and beam wavelength bandpass ($\Delta\lambda/\lambda$) in the direct (a) and reciprocal (b) space



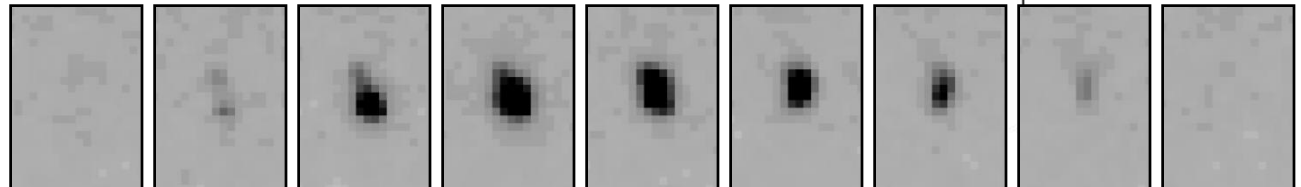
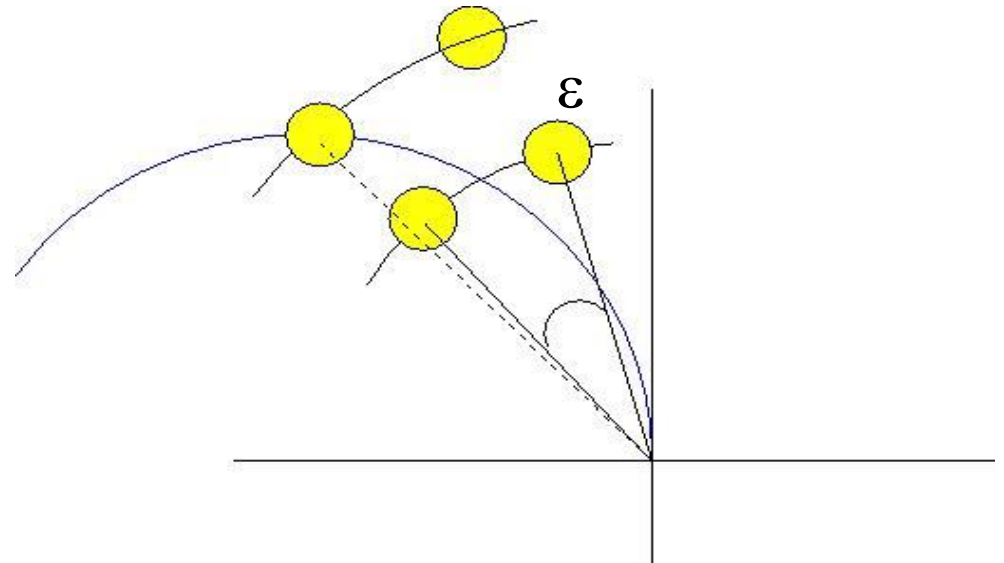
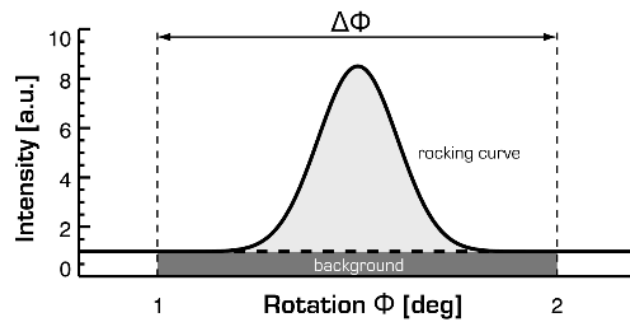


A schematic representation of the beam divergence (δ), crystal mosaicity (η) and beam wavelength bandpass ($\Delta\lambda/\lambda$) in the direct (a) and reciprocal (b) space

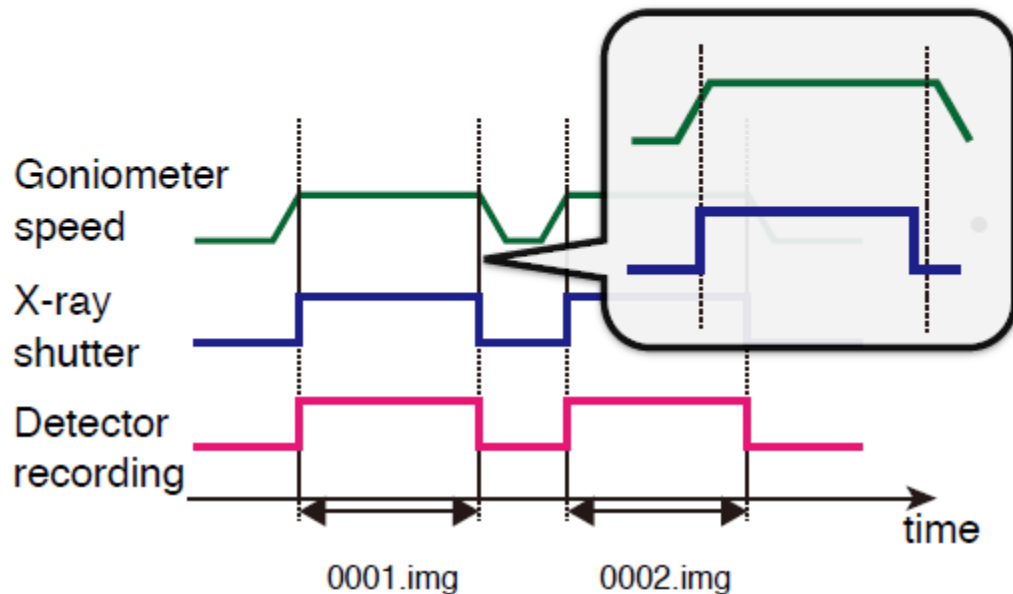


2 types de réflexions sur le cliché de diffraction :

- Réflexions entières
- Réflexions partielles

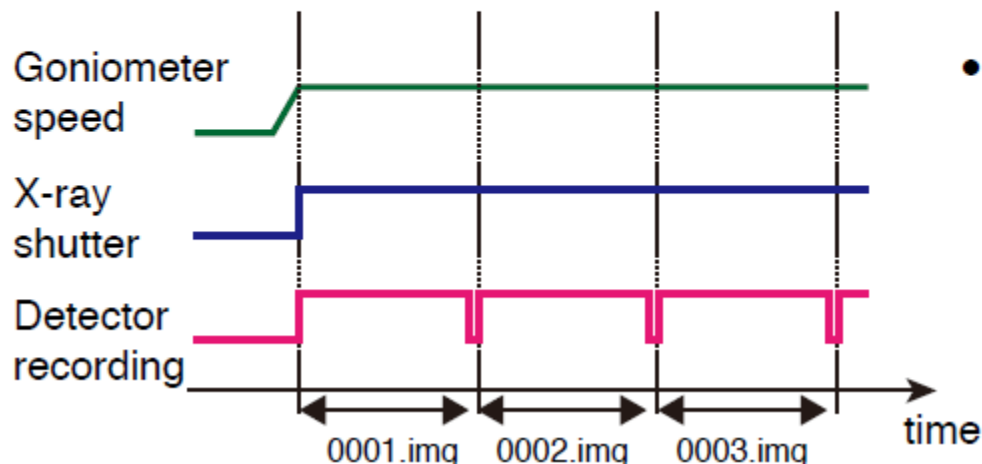


What is 'shutterless'?



With-shutter data collection

- Slow.
- Goniometer/X-ray shutter synchronization error and shutter jitter may exist.



Shutterless data collection

- Fast.
- No mechanical error.
- Requires detector with ignorable readout time.

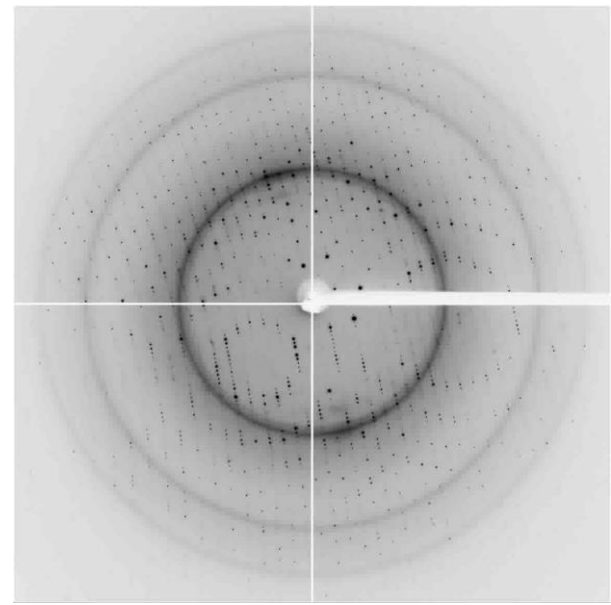
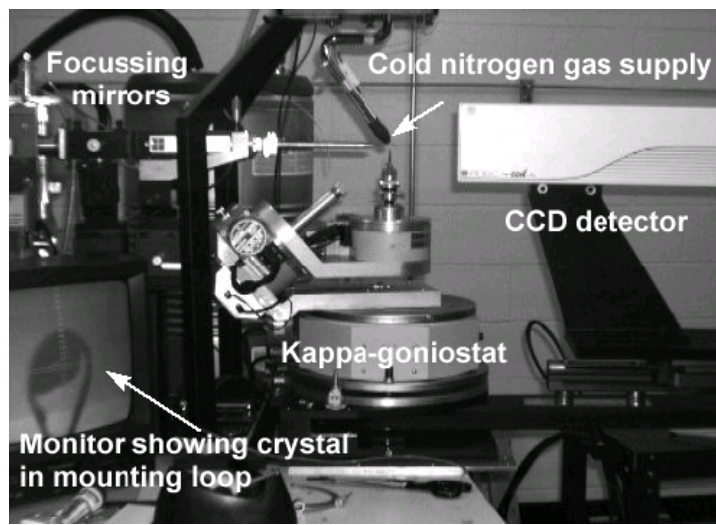
La collecte des données

- **Un moment crucial :**
 - Obtenir un espace complet de diffraction sur toute la gamme de résolution
 - Des mesures précises et de qualité
- **Des choix et des compromis, des décisions à prendre rapidement :**
 - Paramètres géométriques de l'enregistrement
 - Domaine à enregistrer en priorité mais obtenir aussi la redondance maximale
 - Angle d'oscillation
 - Temps d'exposition
 - Un cristal à durée de vie limitée

[Optimization of data collection taking radiation damage into account.](#)

Bourenkov GP, Popov AN.

Acta Crystallogr D Biol Crystallogr. 2010 Apr;66(Pt 4):409-19. doi: 10.1107/S0907444909054961.



Data collection, especially at synchrotron sites, is the time where you may solve your structure in a few minutes or destroy months (or years) of work and waste a lot of money in a few seconds. **Be prepared !**

Crystallographic programs are now very easy to use, require few parameters but are based on complicated algorithms.

Try to understand what is going on and where crucial decisions need to be done

It may be fast to build 90% of a structure but it will take a longer time to finish the last 10%.

Quelques logiciels :

- Denzo - HKL2000/HKL3000
 - <http://www.hkl-xray.com/>



- Mosflm (CCP4)
 - <http://www.mrc-lmb.cam.ac.uk/harry/mosflm>
- XDS
 - <http://xds.mpimf-heidelberg.mpg.de/>



Le traitement des données est de plus en plus intégré à la collecte.
Résultats peu de temps (qq mn) après la collecte.

Pipelines

Xia2 (CCP4)

```
xia2 pipeline=3d /here/are/my/images
```

autoPROC (GlobalPhasing Ltd)

```
process -I /where/ever/images -d sub.dir > out.put
```

http://jean.cavarelli.free.fr/igbmc/labo_services/xray_progs.html

Traitements des données de diffraction

1. Inspection visuelle d'une image de diffraction

2. Indexation initiale

3. Choix des paramètres optimaux d'enregistrement

4. Collecte d'un espace complet de diffraction

5. Affinement des paramètres

$$\sum ((P,Q)^{hkl}_{\text{calc}} - (P,Q)^{hkl}_{\text{obs}}) = f(\text{param}(i), i=1,n)$$

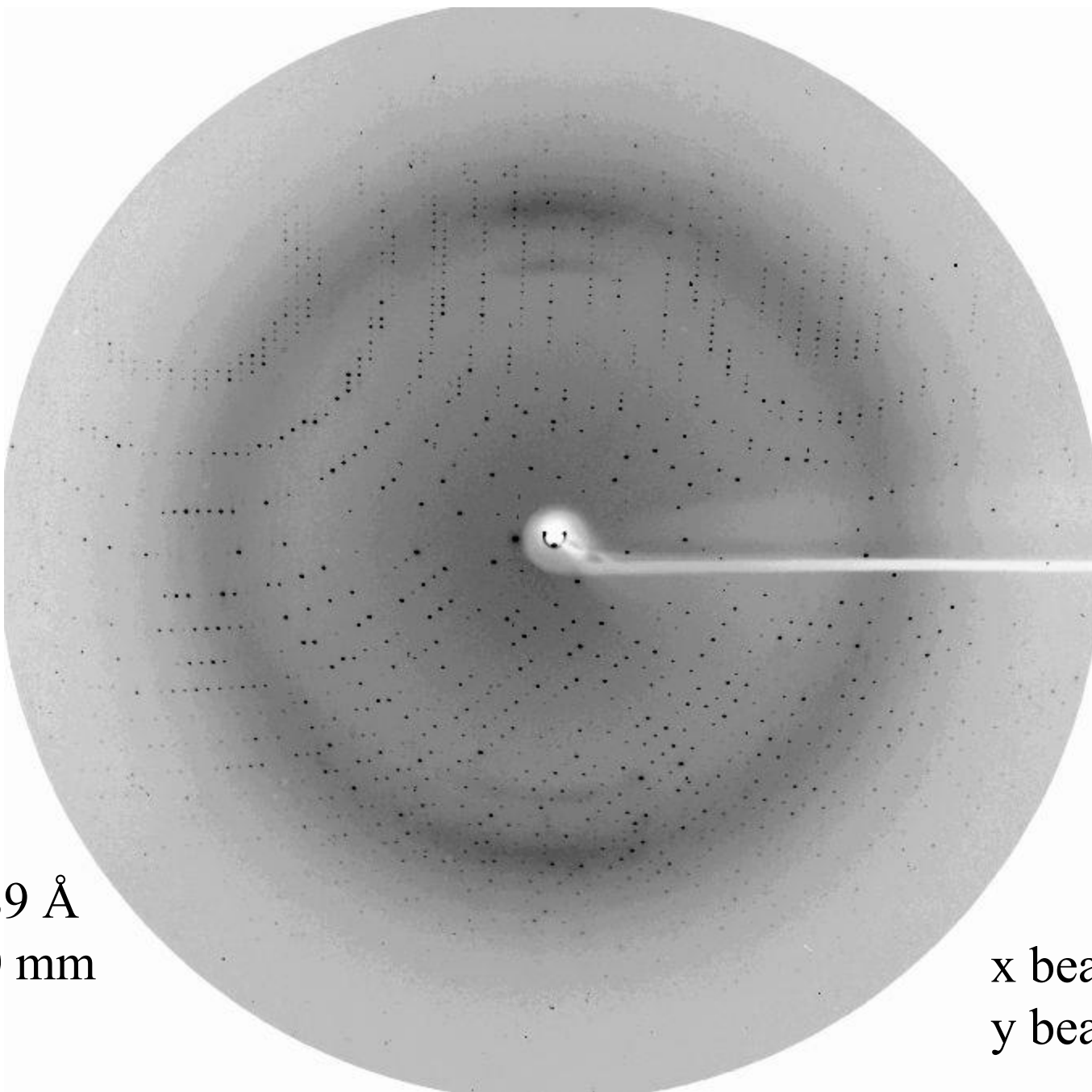
6. mesures des intensités par cliché

7. Mises à l'échelle

8. Réduction des données, statistiques, erreurs

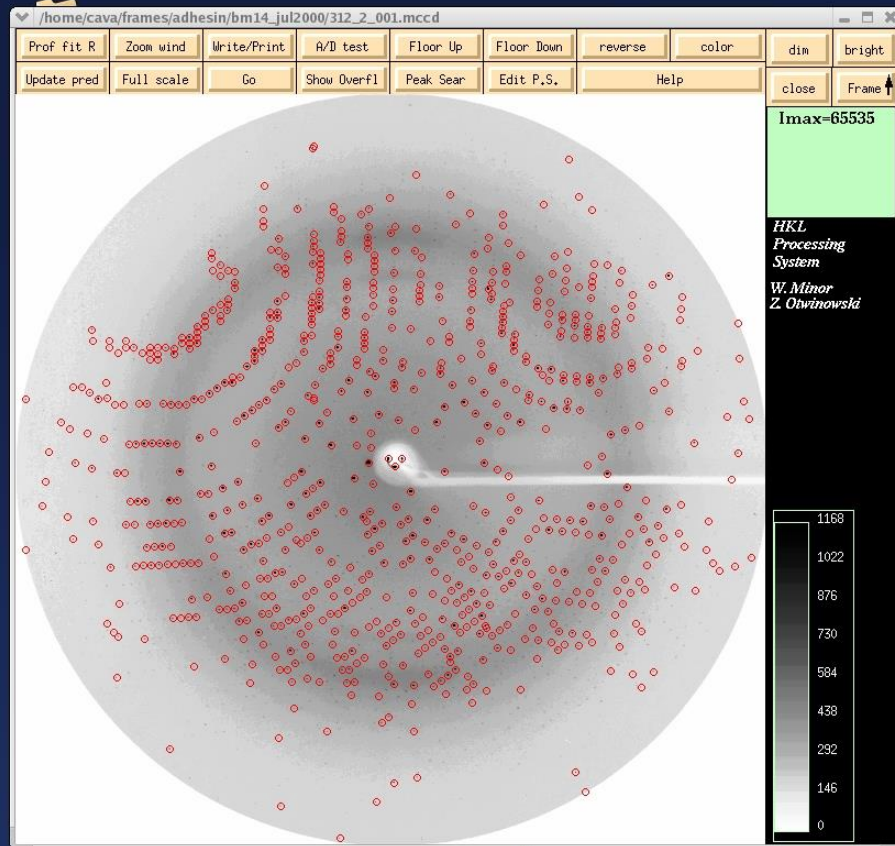
- Réflexions équivalentes
- Extinctions

9. Résultats



$\lambda = 0.9789 \text{ \AA}$
DF = 140 mm
 $\Delta\phi = 0.7^\circ$

x beam 64.1
y beam 66.9



avarelli at I.G.B.M.C. - collaborative license Academic license

Configuration Crystal Information Report Help

Strategy/Simulation Integration Scaling Macros Credits Copyrights

Refinement Information

Space Group: P1

Resolution: 50.00 - edge

Positional: X- χ^2 : Y- χ^2 :

Partiality: χ^2 :

X Beam: Y Beam:

a: b: c:
 α : β : γ :

Crystal Rotation X:
Crystal Rotation Y:
Crystal Rotation Z:
Detector Rotation X:
Detector Rotation Y:
Detector Rotation Z:
Crossfire X:
Crossfire Y:
Crossfire XY:
Distance:
Mosaicity:

Controls

Refinement Sigma Cutoff 5.0

3D Window 5

Peak Search

Display Change Display to frame 1

Index on set 1

Refine for 5 cycles

Bravais Lattice Check Mosaicity

Abort Refinement Reference Zone Alignment Zone

Integrate Sets Integration Setup

Set Beam Position Set Blind Region

Reject Criteria



Integration Box

Profile Fitting Radius 10.0

Box Size 24

Spot Size 0.35

☐ Elongation Limit 0.7

More Options



Terminal

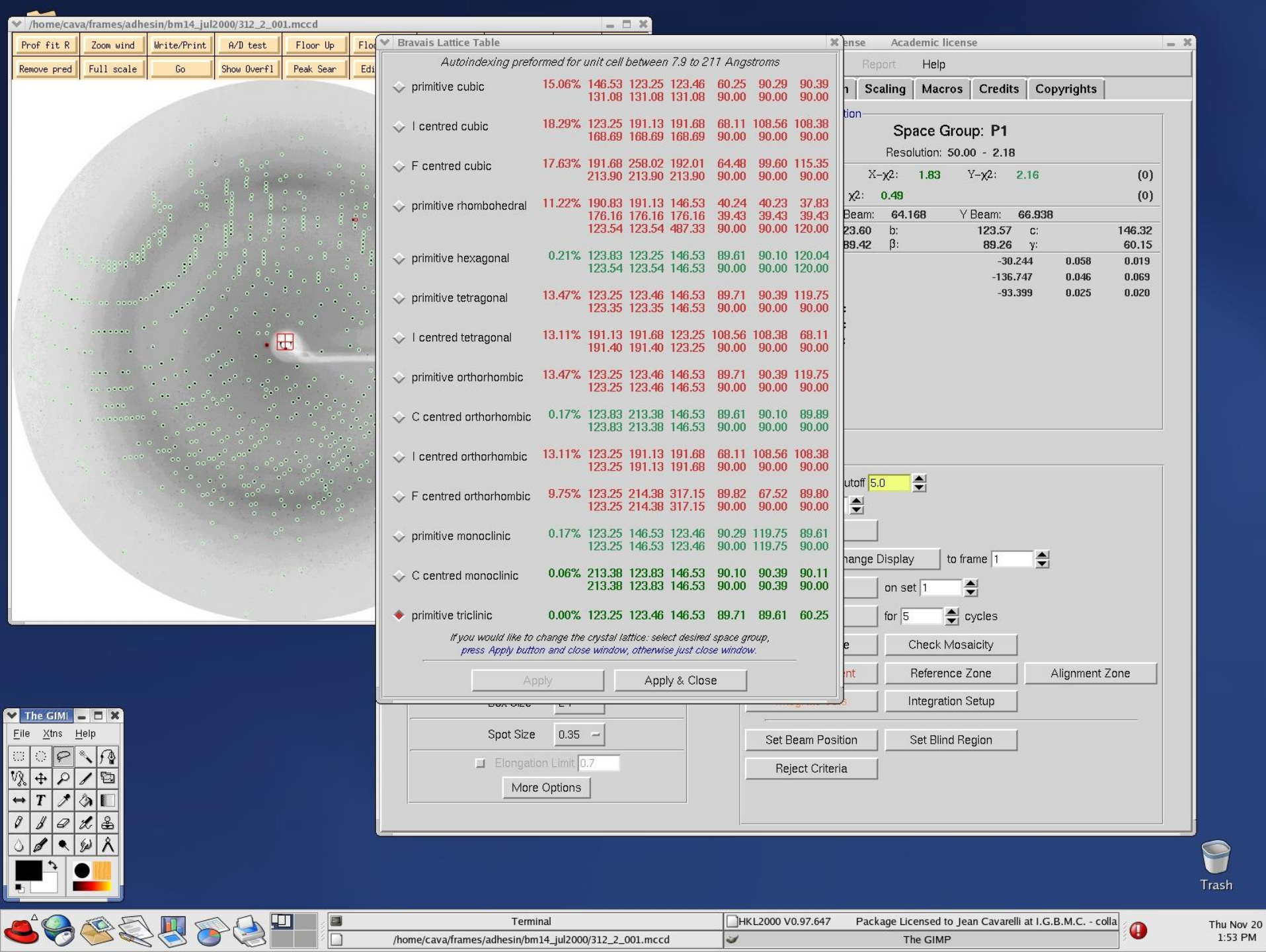
/home/cava/frames/adhesin/bm14_jul2000/312_2_001.mccd

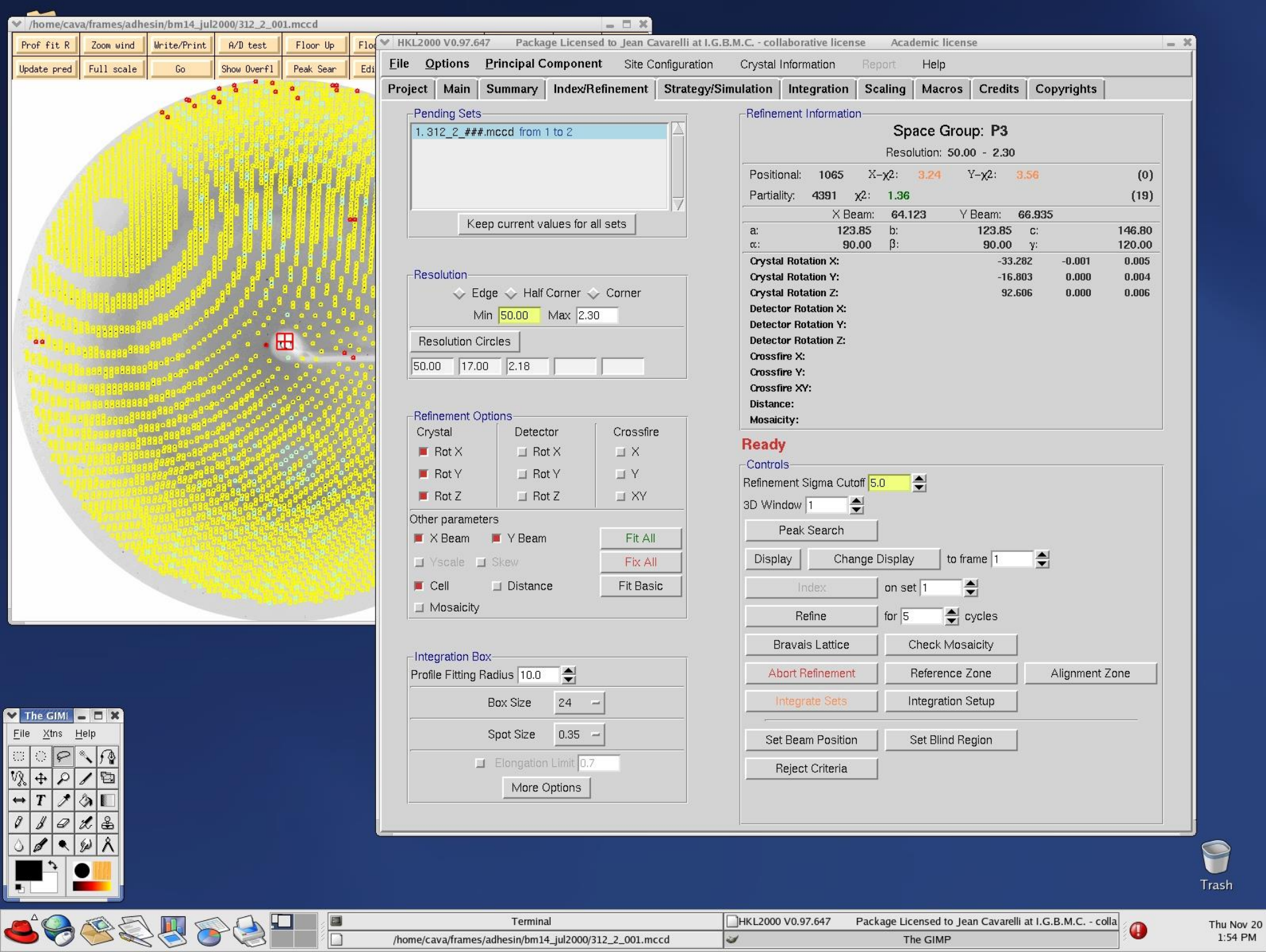
HKL2000 V0.97.647 Package Licensed to Jean Cavarelli at I.G.B.M.C. - colla

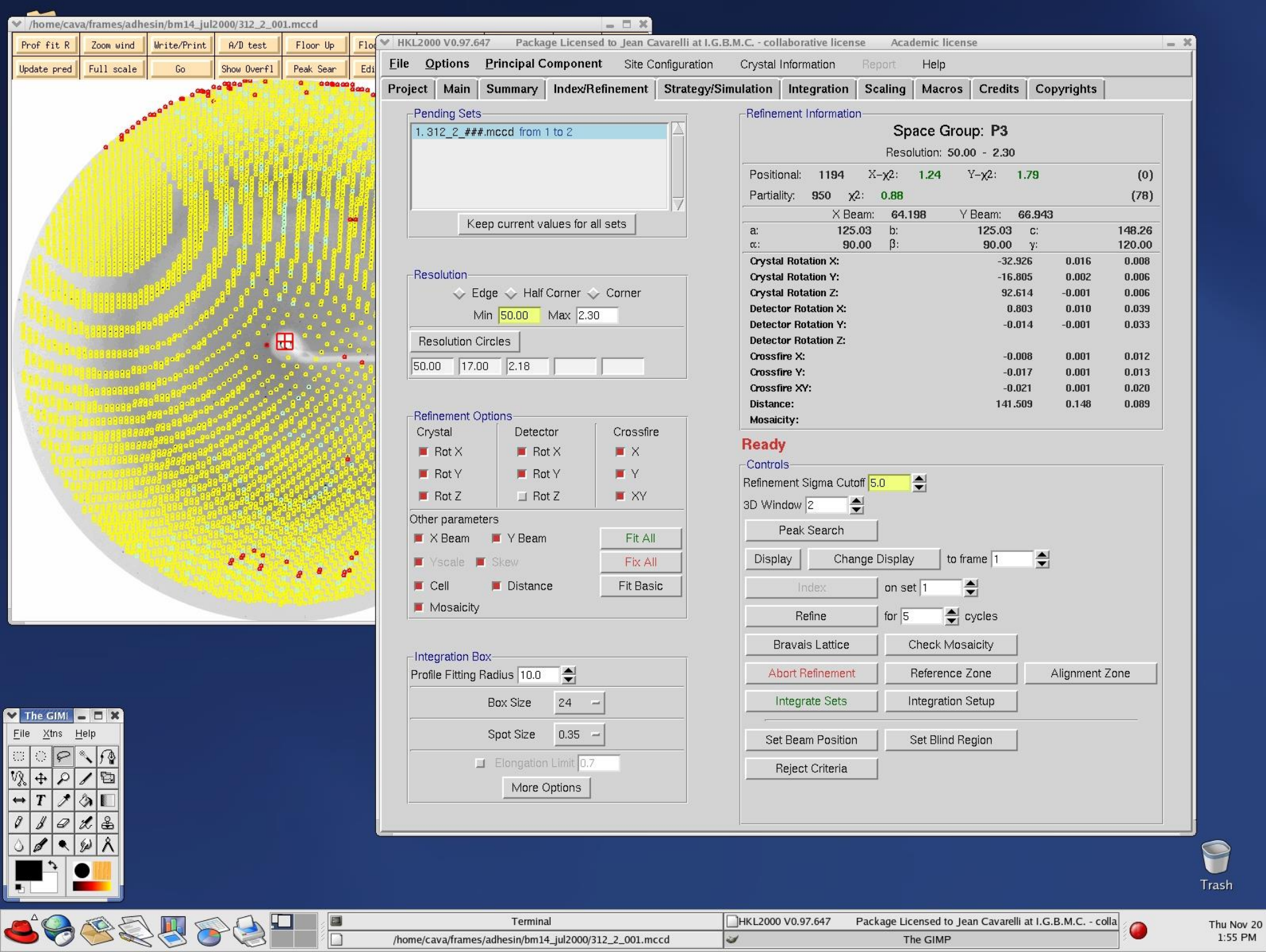
The GIMP

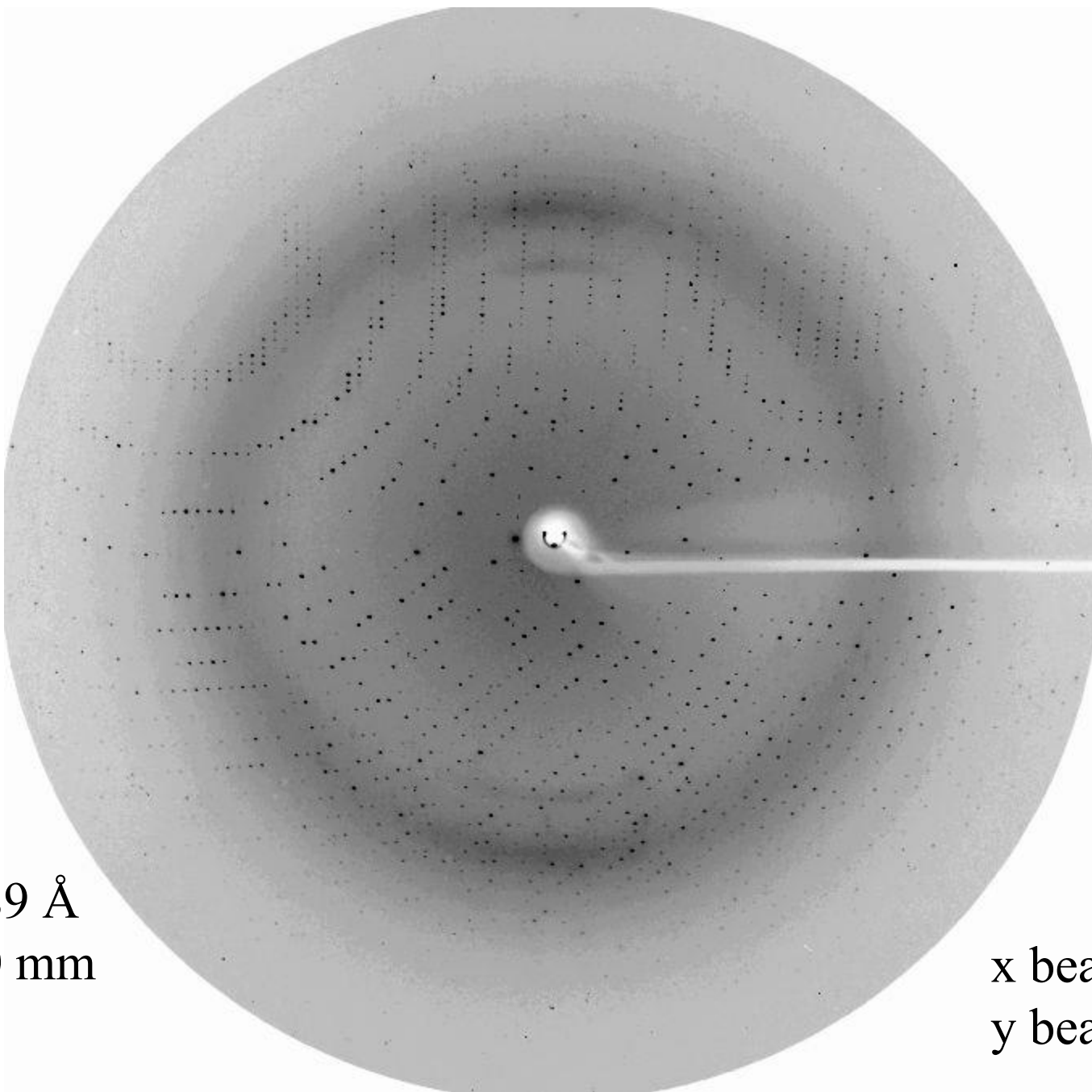


Thu Nov 20
1:52 PM



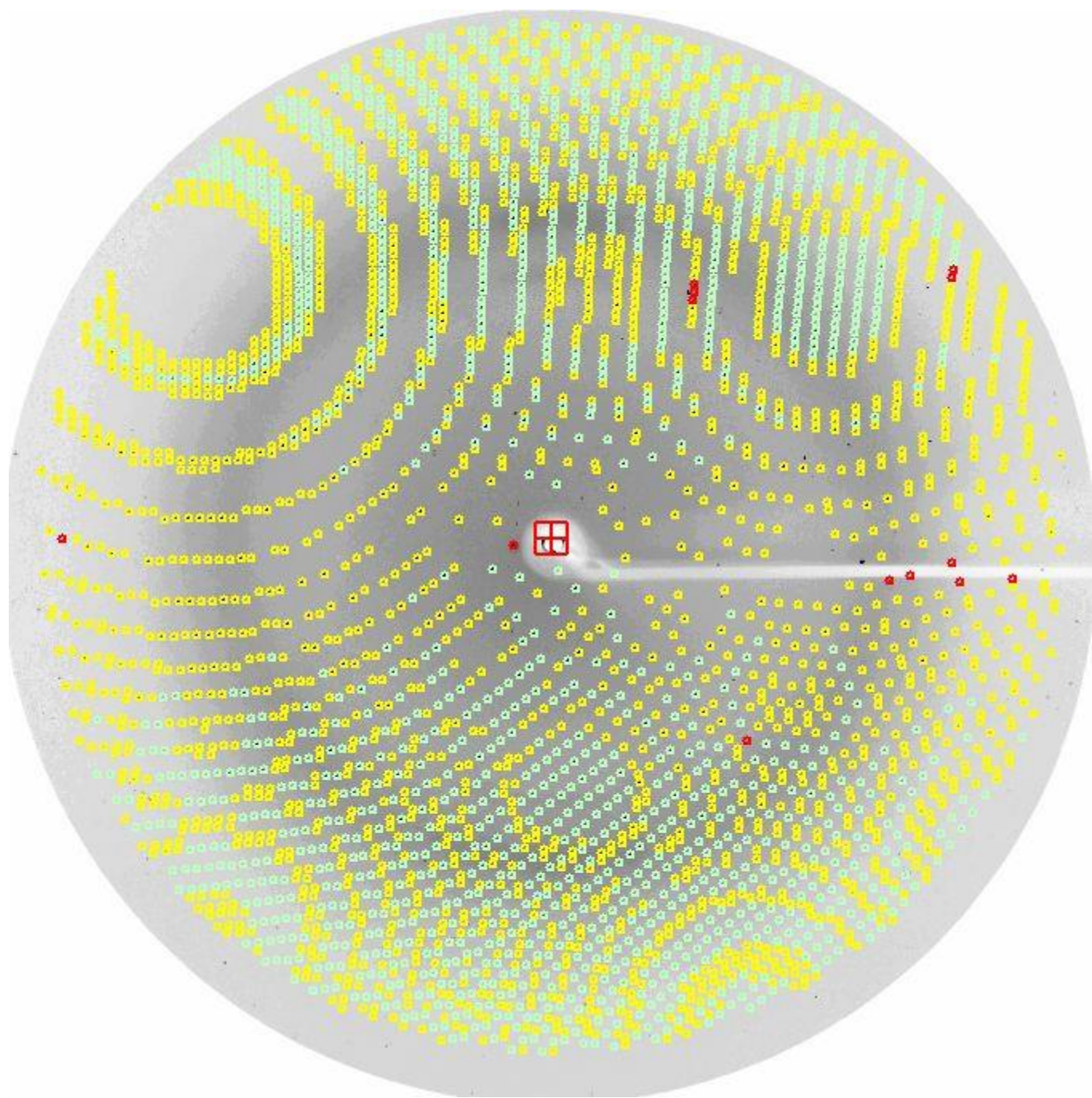




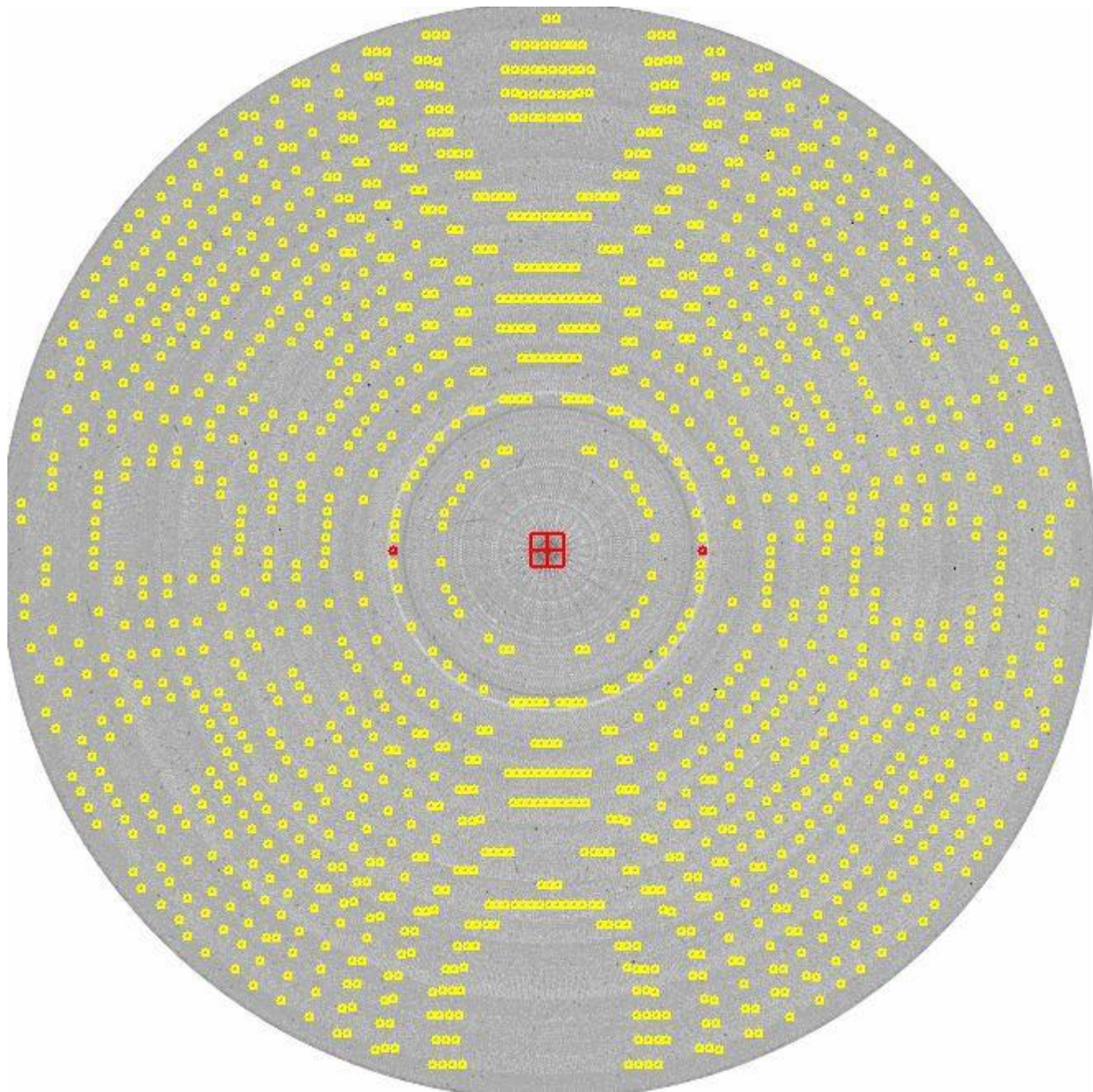


$\lambda = 0.9789 \text{ \AA}$
DF = 140 mm
 $\Delta\phi = 0.7^\circ$

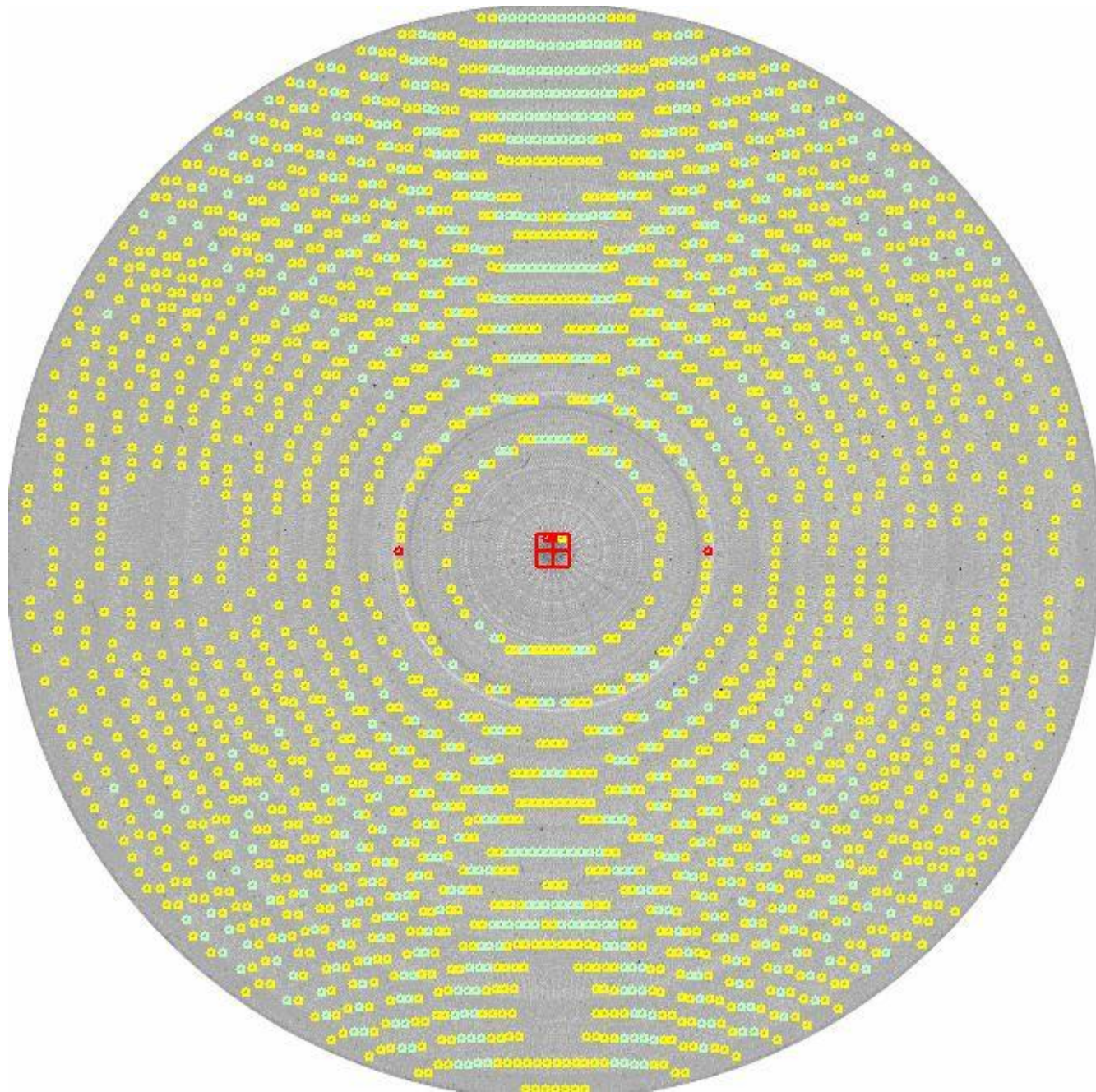
x beam	64.1
y beam	66.9



$$\Delta\phi = 0.2^\circ$$

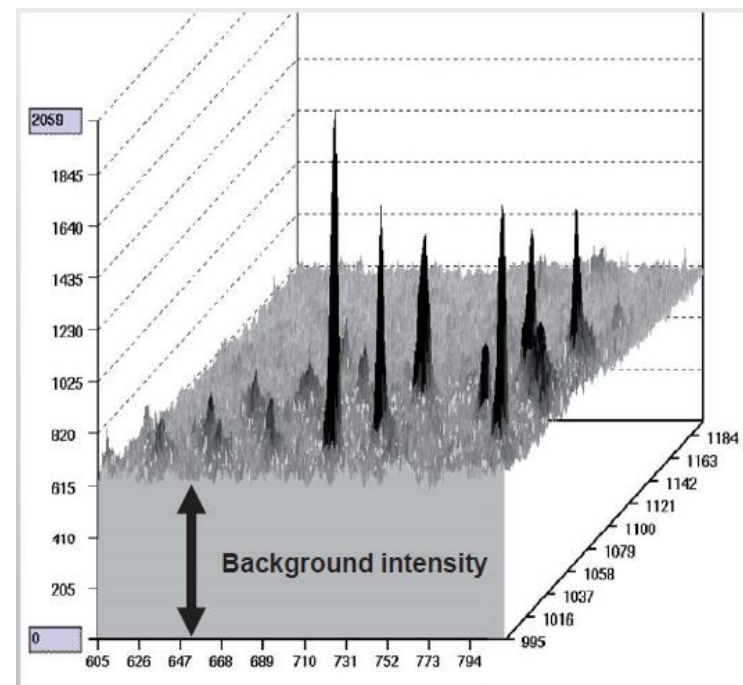
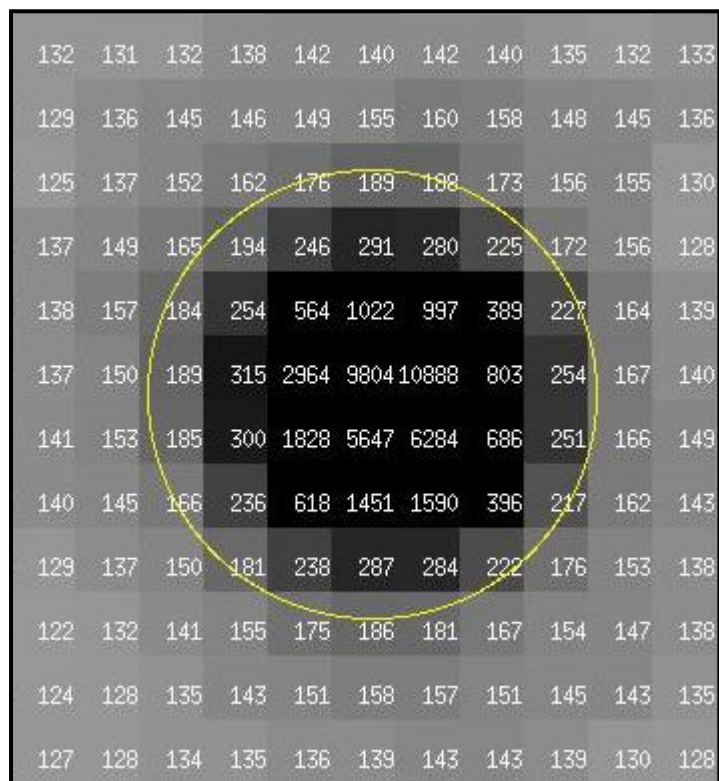


$$\Delta\phi = 0.7^\circ$$

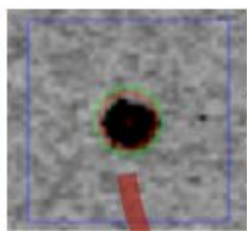
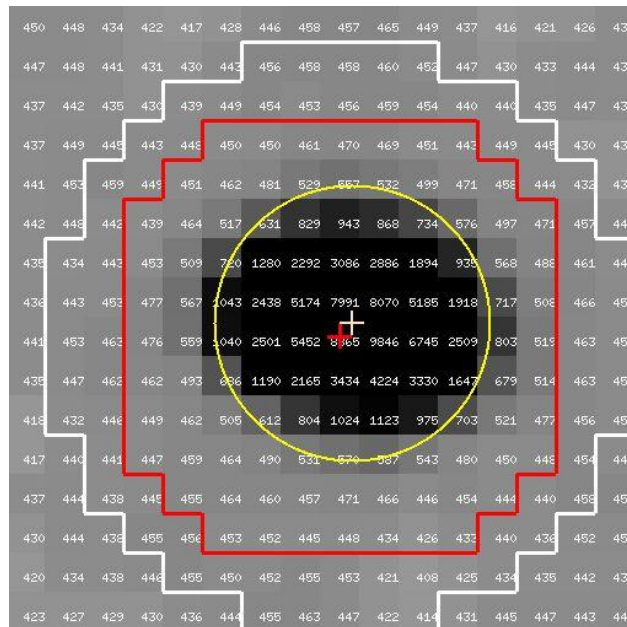
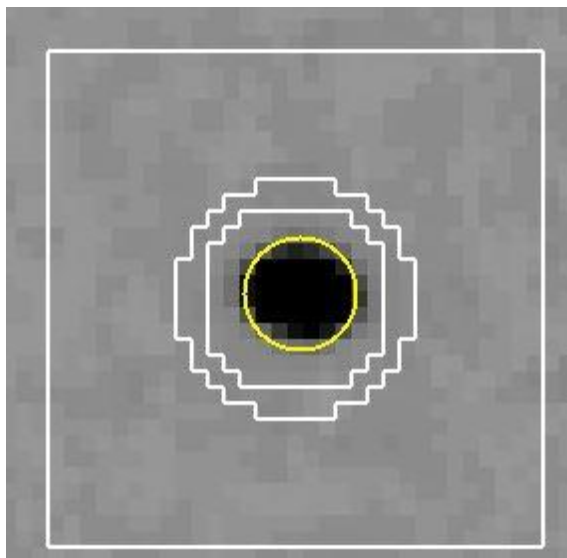


Traitements des données de diffraction

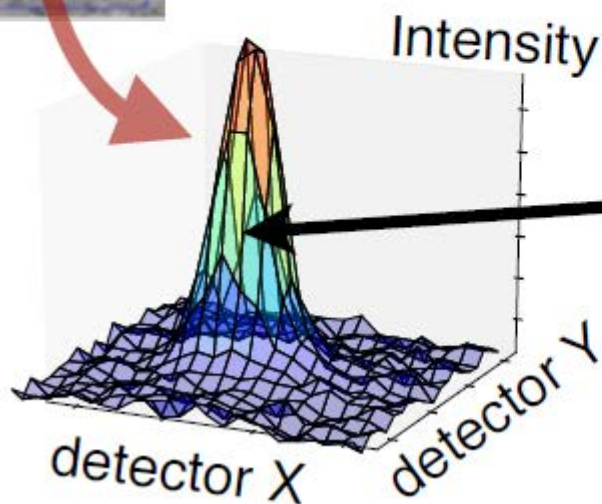
1. Inspection visuelle d'une image de diffraction
2. Indexation initiale
3. Choix des paramètres optimaux d'enregistrement
 - Domaine angulaire à couvrir
 - Angle d'oscillation
 - Distance cristal détecteur
 - Temps d'exposition
4. Collecte d'un espace complet de diffraction
5. Affinement des paramètres
$$\sum ((P,Q)^{hkl}_{calc} - (P,Q)^{hkl}_{obs}) = f(\text{param}(i), i=1,n)$$
4. mesures des intensités par cliché (Intégration 2D ou 3D)
 - Méthode du profil "Profil fitting"
5. Mises à l'échelle
6. Post-Affinement
7. Réduction des données, statistiques, erreurs
8. Résultats



$$|\mathbf{F}|^2 \pm \sigma_{|\mathbf{F}|^2}$$

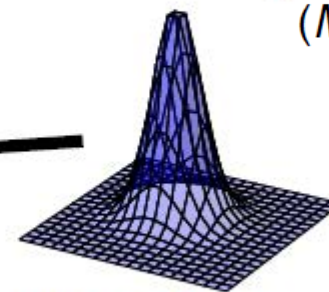


Bragg peak



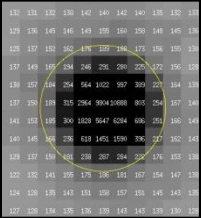
Fit

2D profile
(MOSFLM, DENZO)

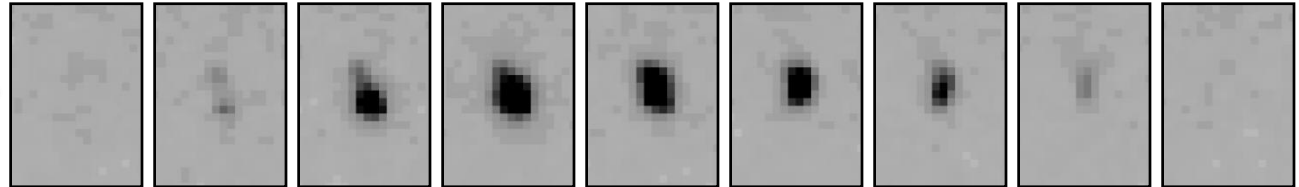


Reference
profile

Consecutive diffraction images

137	135	142	138	142	140	142	140	138	137	141
135	136	145	146	143	135	140	139	149	145	135
133	137	150	152	145	133	137	146	152	150	133
137	149	149	154	146	135	140	135	142	149	137
138	151	154	154	154	142	137	133	144	139	138
137	150	159	155	136	14	1022	997	303	223	149
141	153	159	155	136	14	1022	997	303	223	149
149	145	146	146	136	131	1405	1506	936	527	149
129	137	145	145	145	139	139	139	139	139	129
127	132	141	135	135	135	135	135	135	135	127
129	130	130	141	131	129	127	131	145	142	129
127	128	128	128	128	127	127	127	127	127	127



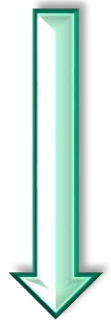
Indexing =
Assigning hkl
index to each
reflection

hkl

Integration = Calculating intensities and their
uncertainties and adding intensities between
images

$$I_{hkl,1} \pm \sigma_{I_{hkl,1}}$$

$$I_{hkl,2} \pm \sigma_{I_{hkl,2}}$$



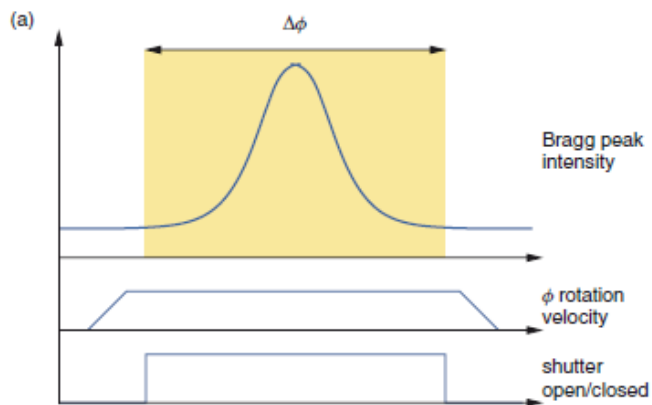
Merging = Combining intensities
for symmetrically equivalent
reflections

$$\langle I_{hkl} \rangle \pm \sigma_{\langle I_{hkl} \rangle}$$

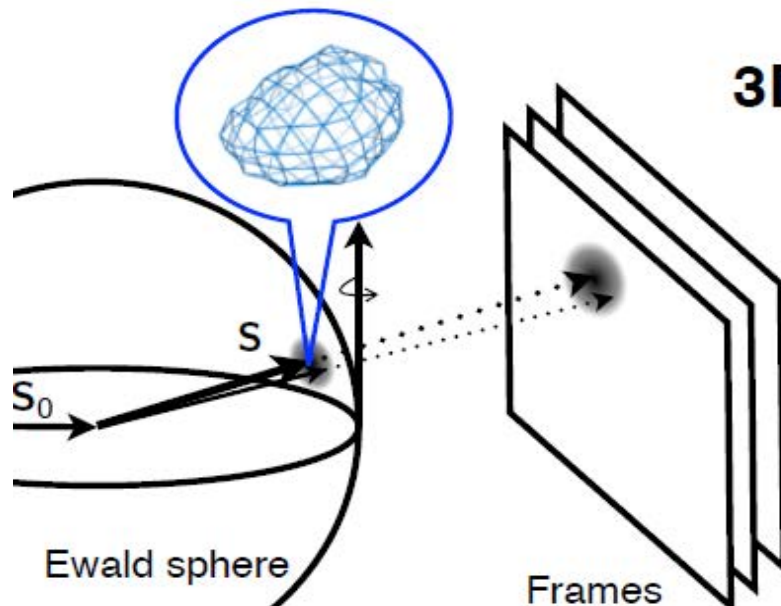
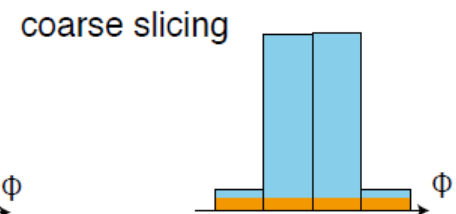
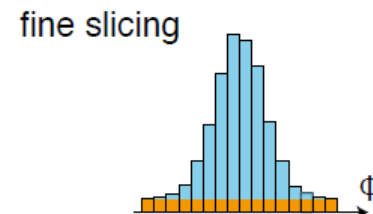
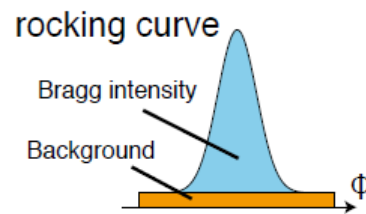
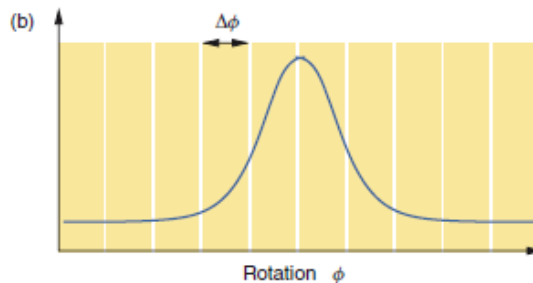


Scaling = Applying scale factor to
integrated intensities to correct
for various experimental factors
changing between diffraction
images

$$I_{hkl} \pm \sigma_{I_{hkl}}$$

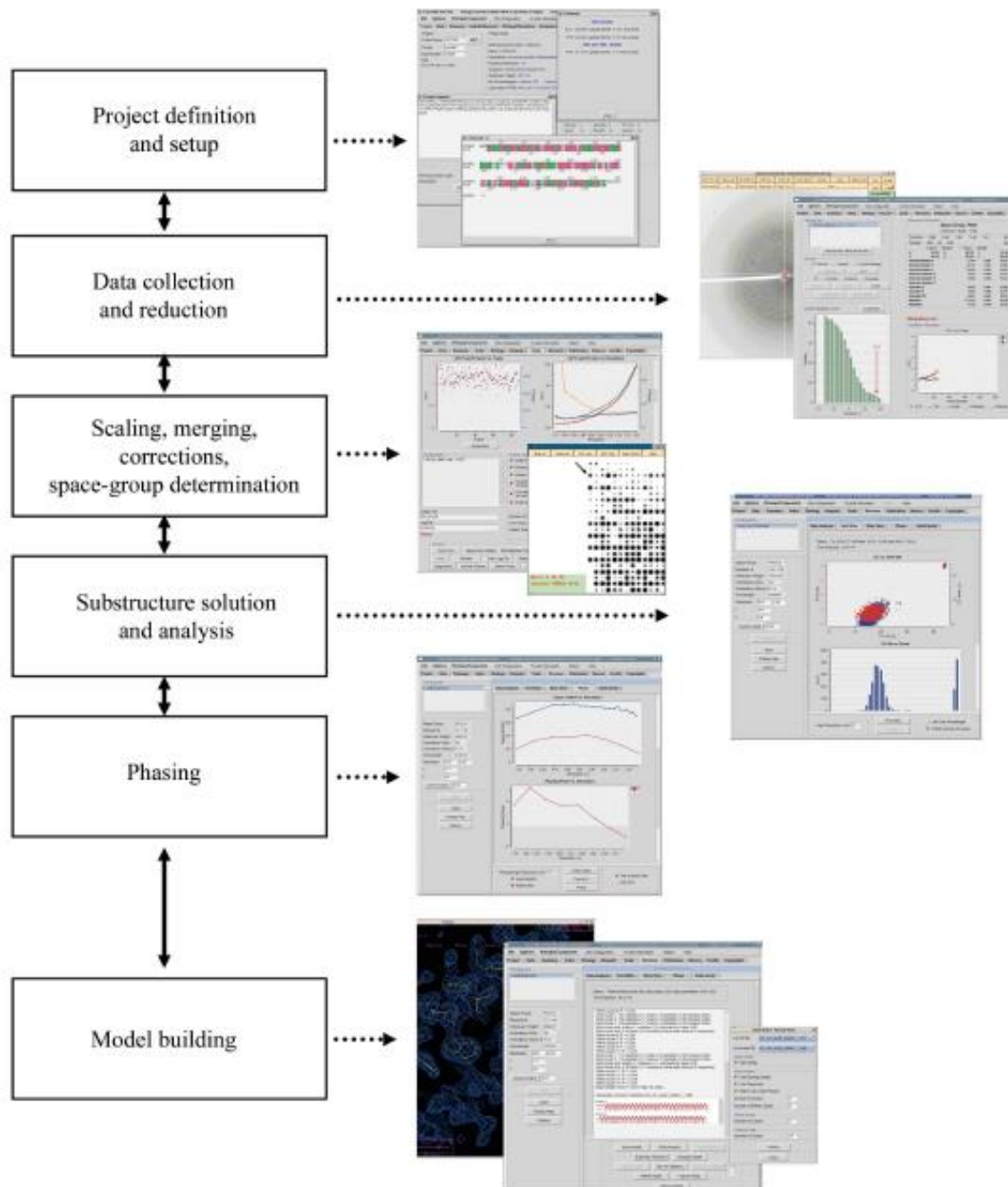


Recording single-crystal diffraction data. (a) In the 'oscillation' method, signal is integrated over an angle $\Delta\phi$, chosen to be large enough to include entire Bragg peaks, but not so large so as to make the recorded image overcrowded. After the exposure, the image is read out, which typically takes a second or more. This process is repeated until the entire required ϕ -range is covered. (b) In the fine phi-slicing mode using modern photon-counting detectors, the readout time is insignificant compared to the typical times the ϕ -motor needs to move over a small $\Delta\phi$, in which case no x-ray shutter is needed, the ϕ -motor runs constantly, and data acquisition is quasi-continuous, with only ms-dead times between exposures.



3D profile (*XDS*, *d*TREK*)

- Consider angular direction as well as detector surface (1+2D)
- Partial reflections are summed during integration.



Traitements des données de diffraction

1. Inspection visuelle d'une image de diffraction
2. Indexation initiale
3. Choix des paramètres optimaux d'enregistrement
4. Collecte d'un espace complet de diffraction
5. Affinement des paramètres
4. Mesures des intensités par cliché

5. Mises à l'échelle entre tous les clichés
6. Réduction des données, statistiques, erreurs
7. Résultats
 - Maille, groupe(s) d'espace(s)
 - N réflexions
 - $h, k, l, I_{hkl}, \sigma(I_{hkl})$

Pending Sets

1. lyso_a_0###.img from 1 to 300
2. lyso_b_0###.img from 1 to 300
3. lyso_c_0###.img from 1 to 300
4. lyso_d_0###.img from 1 to 300
5. lyso_e_0###.img from 1 to 300
6. lyso_f_0###.img from 1 to 300

Output File

output.sca

Log File

scale.log

Scaling

Scaling Options

☒ Scale Restrain 0.01

☐ B Restrain 0.02

☒ Absorption Correction Low

☐ Use Auto Corrections

☐ Use rejections on next run

☒ Write rejection file

☒ Fix B

☒ Anomalous

☐ Scale Anomalous

☒ Ignore Overloads

☐ Direction Cosines

☐ Fit Goniostat

Number of Zones 20

Error Scale Factor 1.3

Default Scale 10

Resolution

Min

Max

Space Group: P43212

Global Refinement

☒ Non-Slipping Crystal Perfect Goniostat

☒ Non-Slipping Crystal Imperfect Goniostat

☒ Small Slippage Imperfect Goniostat

☐ Custom Postrefinement

☐ No Postrefinement

Error Scale Factor 10

Scalepack Version

Controls

Scale Sets

Adjust Error Model

Edit Rejection File

Show Redundancies

Display Frame 1

Load Output File

Abort

Reindex

Show Log File

Reciprocal Lattice

Reprocess

Delete Reject File

Diagnostics

Exclude Frames

Check Space Group

Space Group Diagram

Check in PDB

Système	Groupe Ponctuel	Groupes chiraux
Triclinique	1	P1
Monoclinique	2	P2 P2 ₁ C2
Orthorhombique	222	P222 P222 ₁ P2 ₁ 2 ₁ 2 P2 ₁ 2 ₁ 2 ₁ C222 C222 ₁ I222 I2 ₁ 2 ₁ 2 ₁ F222
Quadratique	4, 422	P4 P4 ₁ P4 ₂ P4 ₃ P422 P42 ₁ 2 P4 ₁ 22 P4 ₃ 22 P4 ₂ 22 P4 ₂ 2 ₁ 2 P4 ₁ 2 ₁ 2 P4 ₃ 2 ₁ 2 I4 I4 ₁ I422 I4 ₁ 22
Trigonal / Rhomboédrique	3, 32	P3 P3 ₁ P3 ₂ P312 P3 ₁ 12 P3 ₂ 12 P321 P3 ₁ 21 P3 ₂ 21 R3 R32
Hexagonal	6, 622	P6 P6 ₁ P6 ₂ P6 ₃ P6 ₄ P6 ₅ P622 P6 ₁ 22 P6 ₂ 22 P6 ₃ 22 P6 ₄ 22 P6 ₅ 22
Cubique	23, 432	P23 P2 ₁ 3 P432 P4 ₂ 32 P4 ₃ 32 P4 ₁ 32 I23 I2 ₁ 3 I432 I4 ₁ 32 F23 F432 F4 ₁ 32

Pour un espace complet et redondant de diffraction pour une forme cristalline donnée :

1. A partir des positions géométriques des réflexions

- Forte présomption $a=b, c$, $\alpha=\beta=90^\circ$ $\gamma=120^\circ$
- Système hexagonal si la symétrie d'ordre 6 est présente
- Système trigonal si seulement une symétrie d'ordre 3 est présente

2. A partir des intensités des réflexions

- Validation (ou non) de l'axe d'ordre 6
 - Validation (ou non) de l'axe d'ordre 3
- Recherche d'axe d'ordre 2

3. A partir des extinctions systématiques

- 6 , 6_1 (ou 6_5), 6_2 (ou 6_4), 6_3 selon c

4. Pas de possibilités à ce stade de distinguer les groupes ayant même symétries et même extinctions

- $P6_1/P6_5$, $P6_2/P6_4$, $P6_122/P6_522$, $P6_222/P6_422$

" Rsym" classique

$$R_{\text{sym}}(I) = \frac{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} |\langle I_{hkl} \rangle - I_{hkl,j}|}{\sum_{hkl} \langle I_{hkl} \rangle} \quad \langle I_{hkl} \rangle = \frac{1}{n_{hkl}} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}$$

" Rrim ", "Redundancy-independent merging R factor "

$$R_{r.i.m}(I) = \frac{\sum_{hkl} \sqrt{\frac{n_{hkl}}{n_{hkl}-1}} \sum_{j=1}^{j=n_{hkl}} |\langle I_{hkl} \rangle - I_{hkl,j}|}{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}}$$

$$R_{\text{rim}} = R_{\text{meas}}$$

The redundancy-independent merging R factor gives the precision of individual intensity measurements, independent of multiplicity – hence, shows how broad the distribution is. This is particularly useful to evaluate low-resolution data quality, for example in experimental phasing.

" Rpim" , "Precision-indicating merging R factor "

$$R_{p.i.m}(I) = \frac{\sum_{hkl} \sqrt{\frac{1}{n_{hkl}-1}} \sum_{j=1}^{j=n_{hkl}} |\langle I_{hkl} \rangle - I_{hkl,j}|}{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}}$$

$$R_{\text{pim}}$$

The precision indicating merging R factor describes the precision of the **averaged** intensity measurements: it gives you the standard error of the mean - How precisely the center of the distribution is known. Like R_{meas} , it needs to be calculated before symmetry equivalents are merged.

Weiss, (2001), J. Apply. Cryst., 34, 130-135.

Data collection	
X-ray source	ESRF ID29
Wavelength (Å)	0.97625
Data collection temperature (K)	100
Detector	Pilatus 6MF
Crystal-detector distance (mm)	383.18
Total rotation range (°)	180
Exposure range (°) and time (s) per image	0.1, 0.04
Mosaicity (°)	0.474
Cell parameters (Å)	a = 100.36, b = 100.36, c = 105.04, $\alpha = \beta = \gamma = 90^\circ$
Space group	P4 ₁ 2 ₁ 2
Resolution range (outer shell) (Å)	46.50-2.10 (2.23-2.10)
Total number of reflections	404573 (61254)
Number of unique reflections	31973 (4976)
Completeness (%)	99.7 (98.4)
Multiplicity	12.7 (12.2)
$\langle I/\sigma(I) \rangle$	19.2 (1.1)
Rmerge	0.074 (2.2)
R _{meas}	0.077 (2.320)
CC _{1/2}	0.999 (0.436)

Linking Crystallographic Model and Data Quality

P. Andrew Karplus¹ and Kay Diederichs^{2*}

In macromolecular x-ray crystallography, refinement R values measure the agreement between observed and calculated data. Analogously, R_{merge} values reporting on the agreement between multiple measurements of a given reflection are used to assess data quality. Here, we show that despite their widespread use, R_{merge} values are poorly suited for determining the high-resolution limit and that current standard protocols discard much useful data. We introduce a statistic that estimates the correlation of an observed data set with the underlying (not measurable) true signal; this quantity, CC^* , provides a single statistically valid guide for deciding which data are useful. CC^* also can be used to assess model and data quality on the same scale, and this reveals when data quality is limiting model improvement.

25 MAY 2012 VOL 336 SCIENCE www.sciencemag.org

Cysteine DiOxygenase (CDO; PDB 3ELN) re-refined against 15-fold weaker data

Karplus, Diederichs (2012). Science, 336, 1030-1033.

"Coefficients de corrélation" $CC_{\frac{1}{2}}$, CC^*

To determine where to cut the data, results from Student's t-test can indicate where statistical significance ends (as a rule-of-thumb data are usually insignificant if $CC_{1/2} < 0.15$ or lower).

$$CC = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^n (x_i - \langle x \rangle)^2 \sum_{i=1}^n (y_i - \langle y \rangle)^2}}$$

$$CC^* = \sqrt{\frac{2 CC_{\frac{1}{2}}}{1 + CC_{\frac{1}{2}}}}$$

CC of the full dataset against the true Intensities (usually unmeasurable)

- Data should be used to higher resolution than the conservative cutoff suggests
- $CC_{\frac{1}{2}}$ directly assesses significance of data
- CC^* predicts the agreement of data and truth
- CC^* provides a statistic that not only assesses data quality but also allows direct comparison of crystallographic model quality and data quality on the same scale
- Crystallographic models at high resolution have F_{calc}^2 closer to the true data than the experimental data are.

Traitements des données de diffraction

1. Inspection visuelle d'une image de diffraction
2. Indexation initiale
3. Choix des paramètres optimaux d'enregistrement
4. Collecte d'un espace complet de diffraction
5. Affinement des paramètres
4. Mesures des intensités par cliché
5. Mises à l'échelle
6. Réduction des données, statistiques, erreurs
 - Réflexions équivalentes
 - Extinctions
7. Résultats
 - Maille, groupe(s) d'espace(s)
 - N réflexions
 - $h, k, l, I_{hkl}, \sigma(I_{hkl})$
 - $h, k, l, |F_{hkl}|, \sigma(|F_{hkl}|)$



Traitements des données de diffraction

Résultats. Un fichier de réflexions avec les mesures différents formats selon les programmes

- Maille, groupe(s) d'espace(s)
- N réflexions : $h, k, l, I_{hkl}, \sigma(I_{hkl})$
- Pour la suite du processus, $h, k, l, |F_{hkl}|, \sigma(|F_{hkl}|)$
- Calcul des modules des facteurs de structures
 - Approche simpliste
 - Approches Statistiques
- Analyses des données
- Fichier final : $h, k, l, |F_{hkl}|, \sigma(|F_{hkl}|)$ différents formats
 - Une norme : format CCP4

Calcul des modules des facteurs de structures

Expérience de diffraction = Mesure de I_{hkl} et calcul $\sigma(I_{hkl})$

$$I_{hkl} \propto \begin{cases} |F_{hkl}|^2 = F_{hkl} \overline{F_{hkl}} & \text{cristal "idéalement imparfait"} \\ & \text{("parfaitement mosaïque")} \\ |F_{hkl}| & \text{cristal parfait} \end{cases}$$

A partir des données, $|F_{hkl}| = \sqrt{I_{hkl}}$ $\sigma(|F_{hkl}|) = \frac{1}{2} \frac{\sigma(I_{hkl})}{\sqrt{I_{hkl}}}$

Problèmes :

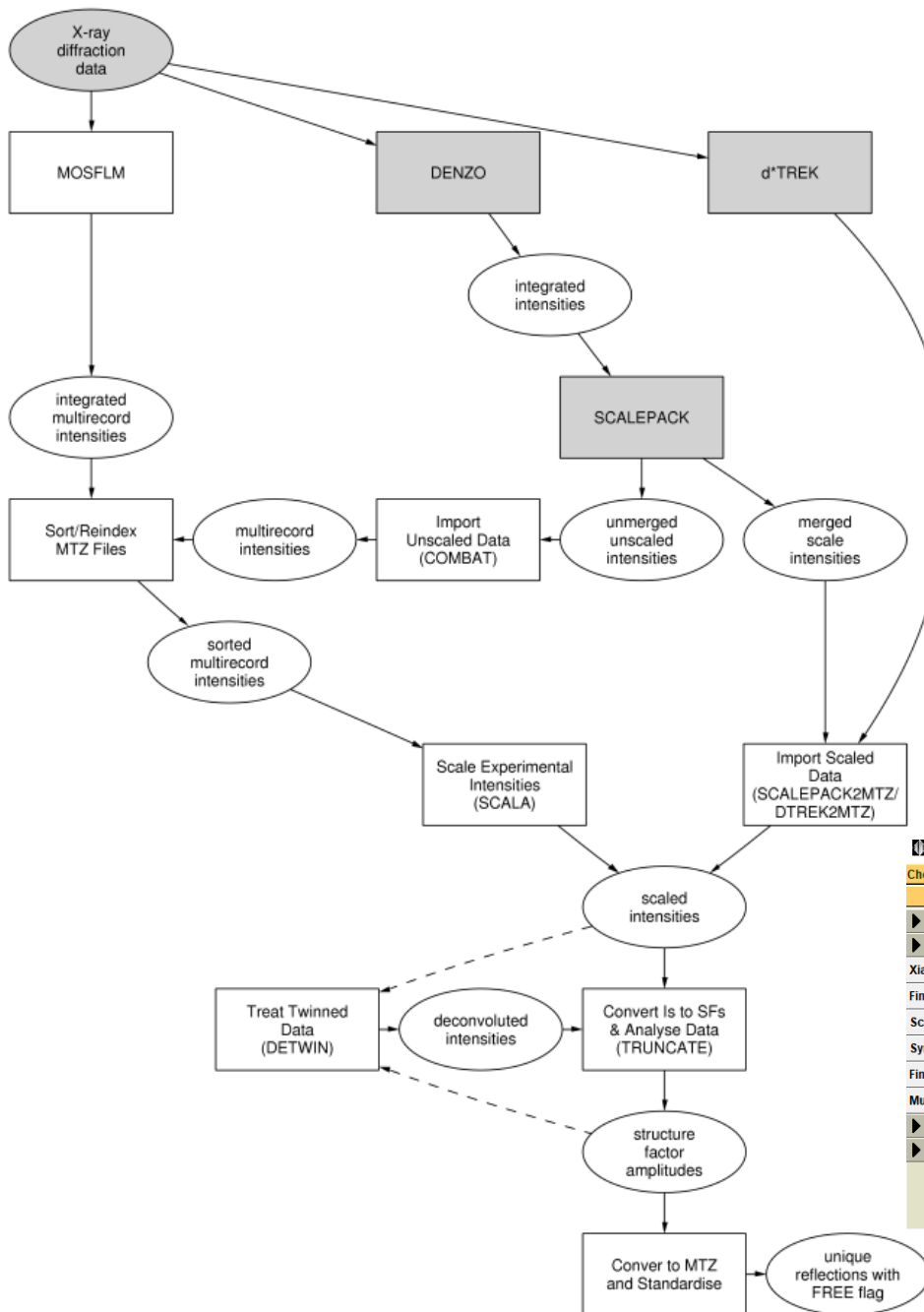
Calcul de $|F_{hkl}|$ pour les réflexions mesurées négatives

Calcul de $\sigma(|F_{hkl}|)$ pour les réflexions faibles

Approche statistique Bayésienne :

- French, Wilson, (1978), Acta Cryst. A34, 517-525. Program Truncate (CCP4)
- Sivia and David, (1994), Acta Cryst. A50, 703-714.

$$|F_{hkl}| = \frac{1}{2} \sqrt{2I_{hkl} + \sqrt{4I_{hkl}^2 + 8\sigma_{I_{hkl}}^2}} \quad \sigma(|F_{hkl}|) = \left(\frac{1}{F_{hkl}^2} + \frac{2(3F_{hkl}^2 - I_{hkl})}{\sigma_{I_{hkl}}^2} \right)^{-\frac{1}{2}}$$



CCP4Interface 7.0.024 running on JCAVA-HPZ Project: PROJ

Message line help

Data Reduction and Analysis		1	28	Nc
...	Start iMosflm			
...	Run Mosflm in batch			
▼ Import Integrated Data				
...	Import Unmerged Data (Pointless)			
...	Import Unmerged Data (Combat)			
...	Import Merged Data			
...	Import Merged Data from Quicksc			
Xia2 - automatic dataprocessing				
Find or Match Laue Group				
Scale and Merge Intensities				
Symmetry, Scale, Merge (Aimless)				
Find Symmetry, Scale & Merge (Scala)				

CCP4Interface 7.0.024 running on JCAVA-HPZ Project: PROJECT

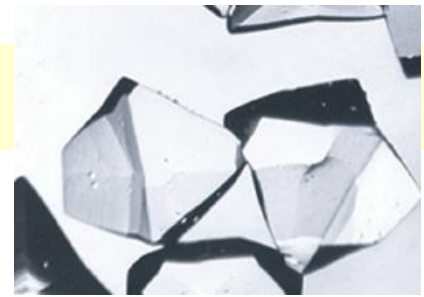
Choose module		Change Project	Help
Data Reduction and Analysis		1	28 Nov 16 FINISHED refmac5 Restrained
▶	Data Processing using Mosflm		
▶	Import Integrated Data		
Xia2 - automatic dataprocessing			
Find or Match Laue Group			
Scale and Merge Intensities			
Symmetry, Scale, Merge (Aimless)			
Find Symmetry, Scale & Merge (Scala)			
Multiple dataset analysis (Blend)			
▶	Utilities		
▶	Check Data Quality		
Directories&ProjectDir			
View Any File			
View Files from Job			
Search/Sort Database..			
Graphical View of Project			
Delete/Archive Files..			
Kill Job			
ReRun Job..			
Edit Job Data			
Preferences			
System Administration			
Exit			

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr
	OBSERVED	UNIQUE	POSSIBLE								
3.39	43959	7862	7879	99.8%	3.3%	3.3%	43901	45.93	3.7%	99.9*	15*
2.40	78904	14040	14053	99.9%	4.1%	3.9%	78859	35.17	4.5%	99.8*	13*
1.96	103441	18058	18058	100.0%	5.5%	5.3%	103418	25.92	6.0%	99.7*	10
1.69	120608	21306	21319	99.9%	9.2%	9.0%	120560	15.15	10.1%	99.5*	6
1.52	133927	24143	24147	100.0%	16.5%	16.8%	133866	8.69	18.2%	98.3*	3
1.38	147353	26645	26649	100.0%	28.7%	29.5%	147313	5.14	31.7%	95.7*	2
1.28	157651	28885	28893	100.0%	48.4%	50.0%	157573	3.09	53.6%	88.1*	2
1.20	152989	31082	31095	100.0%	74.2%	76.2%	152814	1.88	83.2%	73.5*	0
1.13	87972	31089	33018	94.2%	90.9%	95.4%	83292	1.01	110.5%	50.5*	-2
total	1026804	203110	205111	99.0%	6.7%	6.6%	1021596	10.69	7.4%	99.9*	4

	Overall	InnerShell	OuterShell
Low resolution limit	44.22	44.22	1.84
High resolution limit	1.80	9.00	1.80
Rmerge (within I+/I-)	0.082	0.032	1.672
Rmerge (all I+ and I-)	0.088	0.033	1.816
Rmeas (within I+/I-)	0.097	0.037	1.972
Rmeas (all I+ & I-)	0.095	0.036	1.967
Rpim (within I+/I-)	0.051	0.020	1.029
Rpim (all I+ & I-)	0.037	0.015	0.752
Rmerge in top intensity bin	0.036	-	-
Total number of observations	308208	2484	18911
Total number unique	47105	440	2780
Mean((I)/sd(I))	11.7	37.8	1.0
Mn(I) half-set correlation CC(1/2)	0.999	0.998	0.470
Completeness	100.0	98.9	100.0
Multiplicity	6.5	5.6	6.8

Analyse des données de diffraction

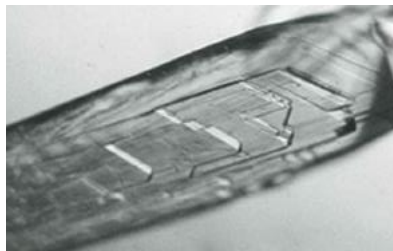


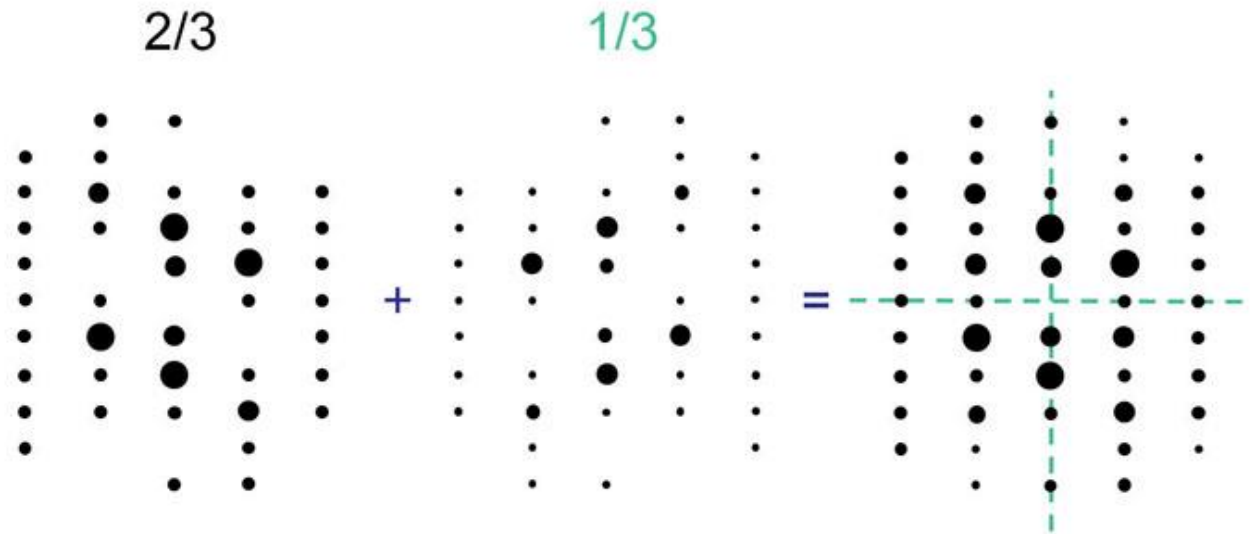
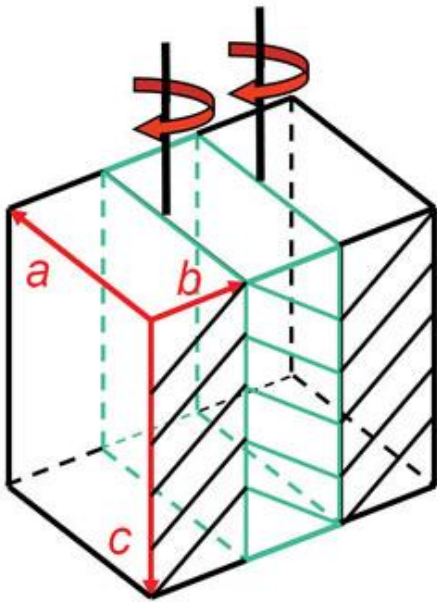
Analyse statistique des données

- Un contrôle qualité
- Éventuellement mise en évidence de défauts cristallins.

Défauts cristallins :

- Échantillon polycristallin. Identifier avant.
- Les défauts peuvent ne pas être visibles et l'aspect extérieur peut apparaître comme monocristallin " twinning "
- Un échantillon cristallin consistant en N domaines qui ont une orientation mutuelle donnée par une opération de symétrie qui n'appartient pas au groupe ponctuel du cristal mais qui est liée à la symétrie du réseau correspondant.





L'analyse statistique de la distribution des intensités permet de mettre en évidence et de tenir compte du phénomène

Moments de la distribution des intensités, " Britton plot ", " Rees $N(z)$ plot ", "Yeates $S(H)$ plot " , " L-function"

Voir Truncate, CCP4, phenix.xtriage

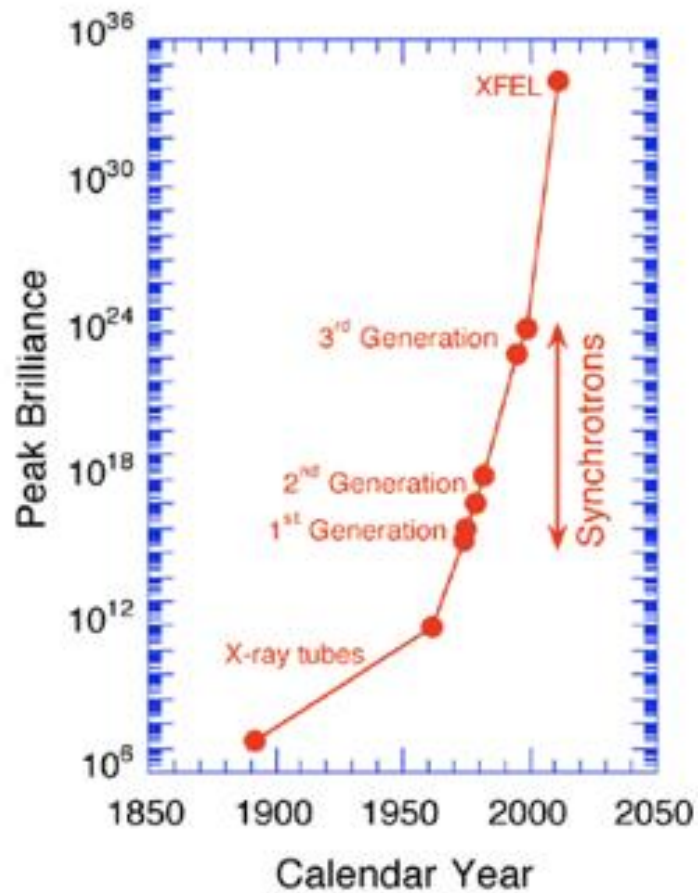
Collecte des données de diffraction

Collecte des données de diffraction

- Un processus crucial
- Un processus de plus en plus automatisé
- Un processus rapide : un espace complet en quelques minutes
 - Matériel : Utilisation de robots
 - Logiciels
 - Des algorithmes complexes avec une utilisation de plus en plus simple
 - Un traitement des données " synchronisé " avec la collecte.
- "Collecte MAD/SAD. Une carte peut être obtenue rapidement sur le site

Des progrès énormes au cours des dernières années

- Obtenir une structure à partir de "cristaux inutilisables" hier
- De nouvelles possibilités

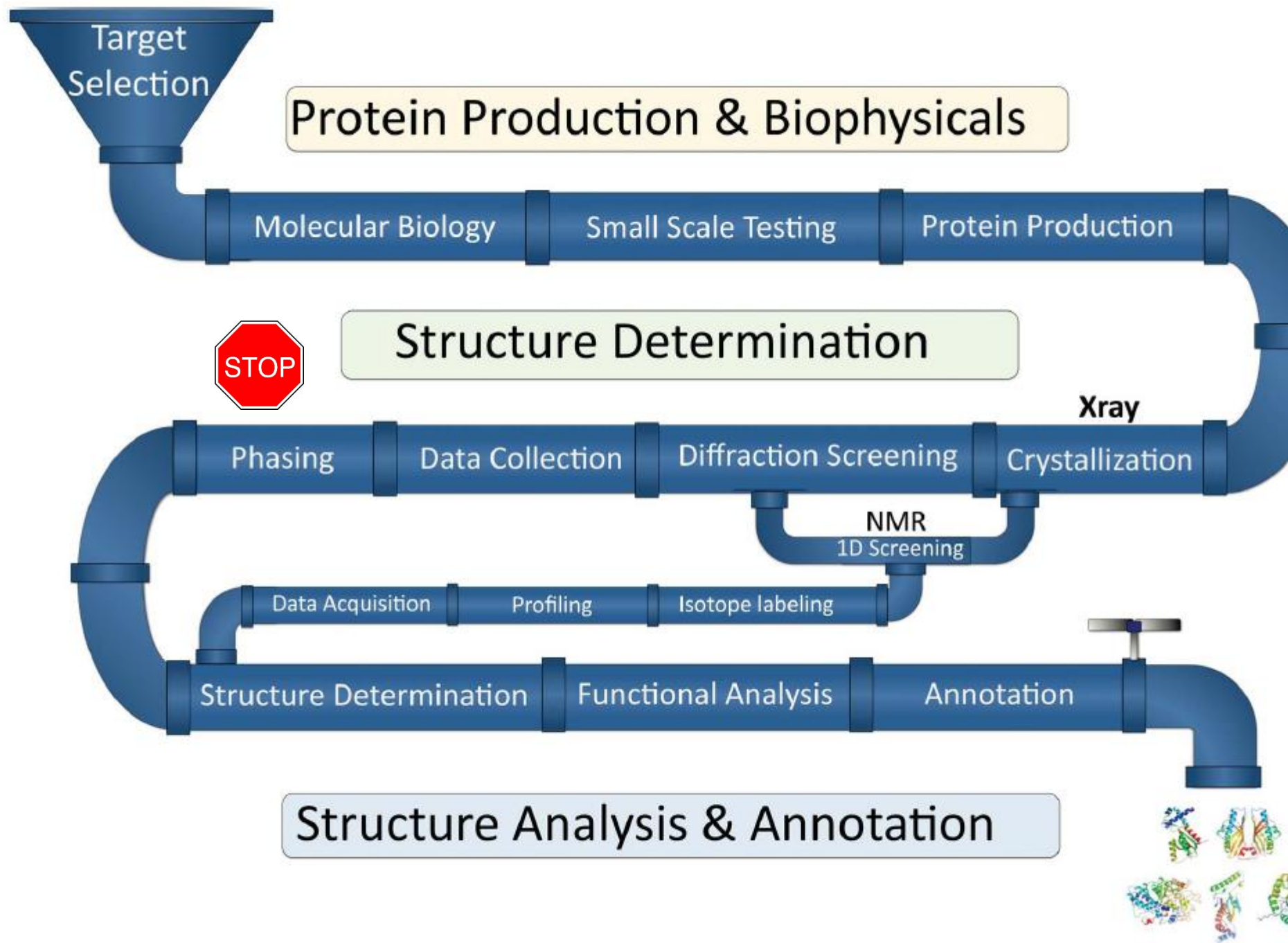


Serial synchrotron crystallography

- From **micro** to **nano**-crystallography

XFELs

- From **nano** to **femtosecond** crystallography
- **structure determination without crystallization**



Méthodes de détermination des phases

- Utiliser des contraintes entre facteurs de structure
 - Méthodes directes.
1934 Fonction de Patterson, 1948 inégalités Harker-Kasper, 1950 déterminants de Karle-Hauptman, 1951 équation de Sayre, 1953 invariants et semi-invariants de Hauptman et Karle, 1956-1958 Hauptman et Karle Formule de la tangente et théorie probabilistique, 1960s "symbolic addition" Karle and Karle, ..., 1985 Hauptman and Karle Prix Nobel Chimie , ...
- Utiliser des variations de modules de facteurs de structure (pour chaque réflexion)
 - Remplacement isomorphe multiple, "Multiple Isomorphous Replacement" (MIR)
 - Utilisation de la diffusion anormale,
 - Travail à plusieurs longueurs d'onde, "Multi-wavelength Anomalous Scattering" (MAD)
 - Travail à une longueur d'onde, "Single wavelength Anomalous Scattering" (SAD)
 - Utilisation de la dégradation de la structure.
 - "(site-specific) Radiation damage-Induced Phasing" (RIP)
Quelques exemples de modifications de structures. "X-Ray RIP" : ruptures de points di-sulfures, decarboxylations Asp, Glu ; pertes de OH de Tyr ; pertes de S-CH₃ de Met ; pertes de Se de SeMet ; pertes de métaux, de Br-AN.
- Utiliser un modèle de molécule parente correctement orientée et positionnée
 - Remplacement Moléculaire

Obtention d'un jeu de phases initiales qui seront le plus souvent améliorées (pour chaque réflexion, diminution de l'erreur par rapport à la phase exacte)

- Méthodes de modifications de densité

Les méthodes "expérimentales" de phasage. Utilisation de variations de modules de chaque facteur de structure

Pour chaque réflexion "indépendante" notée H,

N_H mesures, $|F_{H,j}|, j = 1 \rightarrow N_H$

$$F_{H,j} = P_H + V_{H,j}$$

- P_H une partie constante, inconnue, la même pour toutes les sources de données j
- $V_{H,j}$ une partie variable, due à un sous-groupe d'atomes donnés

$V_{H,j}$ reflète **une structure atomique** (positions atomiques, facteurs de déplacements atomiques, ...), qui peut être représenté par une variable vectorielle θ ,

$$|F_{H,j}(\theta)| = |P_H + V_{H,j}(\theta)|, \quad j = 1 \rightarrow N_H$$

$$|F_{H,j}(\theta)| = |P_H + V_{H,j}(\theta)|, \quad j = 1 \rightarrow N_H$$

en théorie, chaque équation peut être résolue très simplement dans le plan complexe.

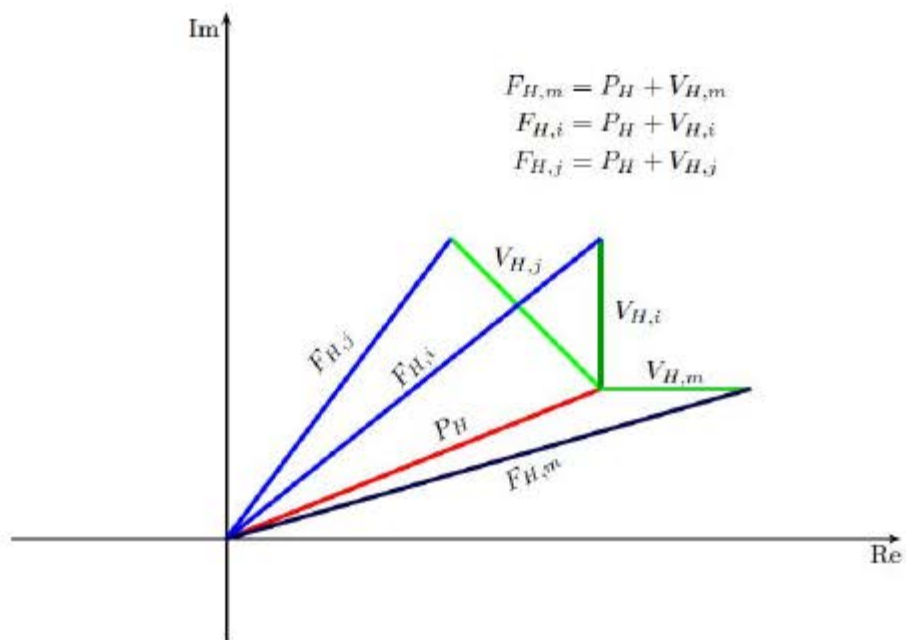
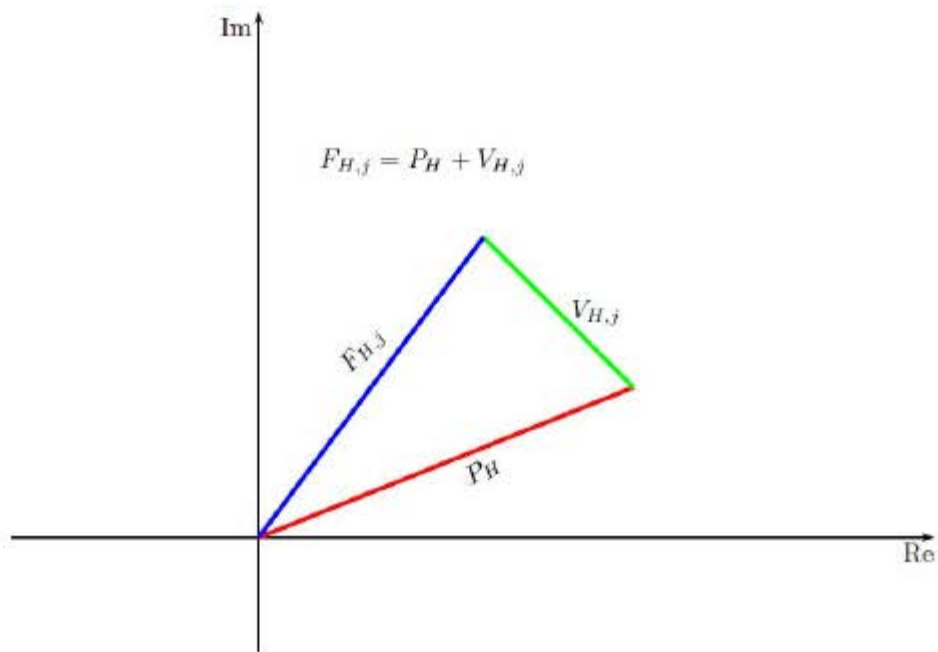
Construction de Harker

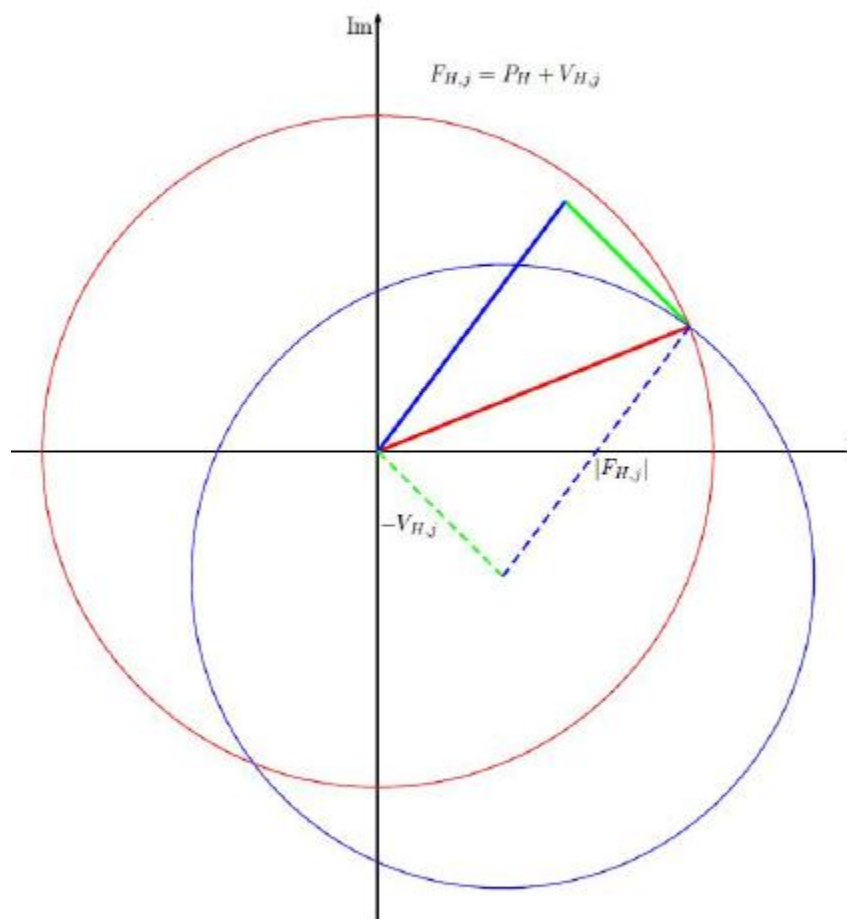
Intersection de N_H cercles centrés en $-V_{H,j}$ et de rayon $|F_{H,j}|$

Si les $V_{H,j}$ ne sont pas colinéaires, Les systèmes d'équations sont résolus si :

- réflexions centriques, $N_H > 1$
- réflexions non-centriques, $N_H > 2$

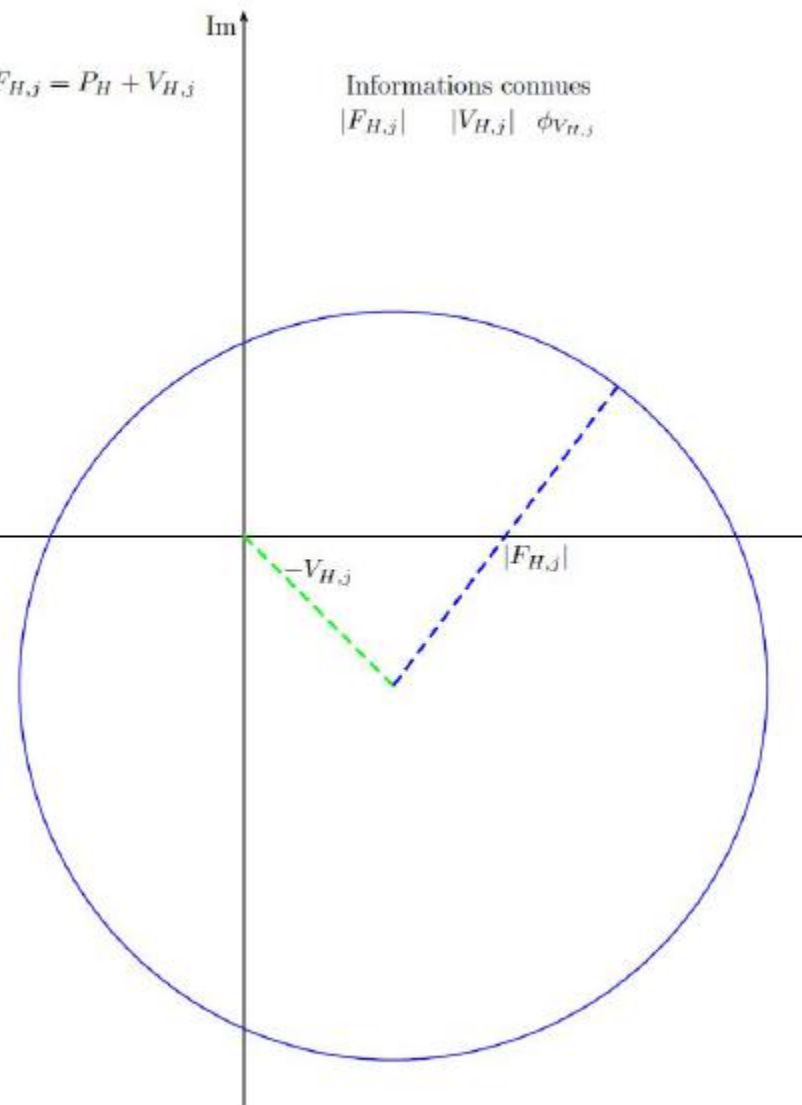
Dans le cas réel, plusieurs catégories d'erreurs vont polluer la simplicité des équations et nécessiter un traitement statistique complexe.

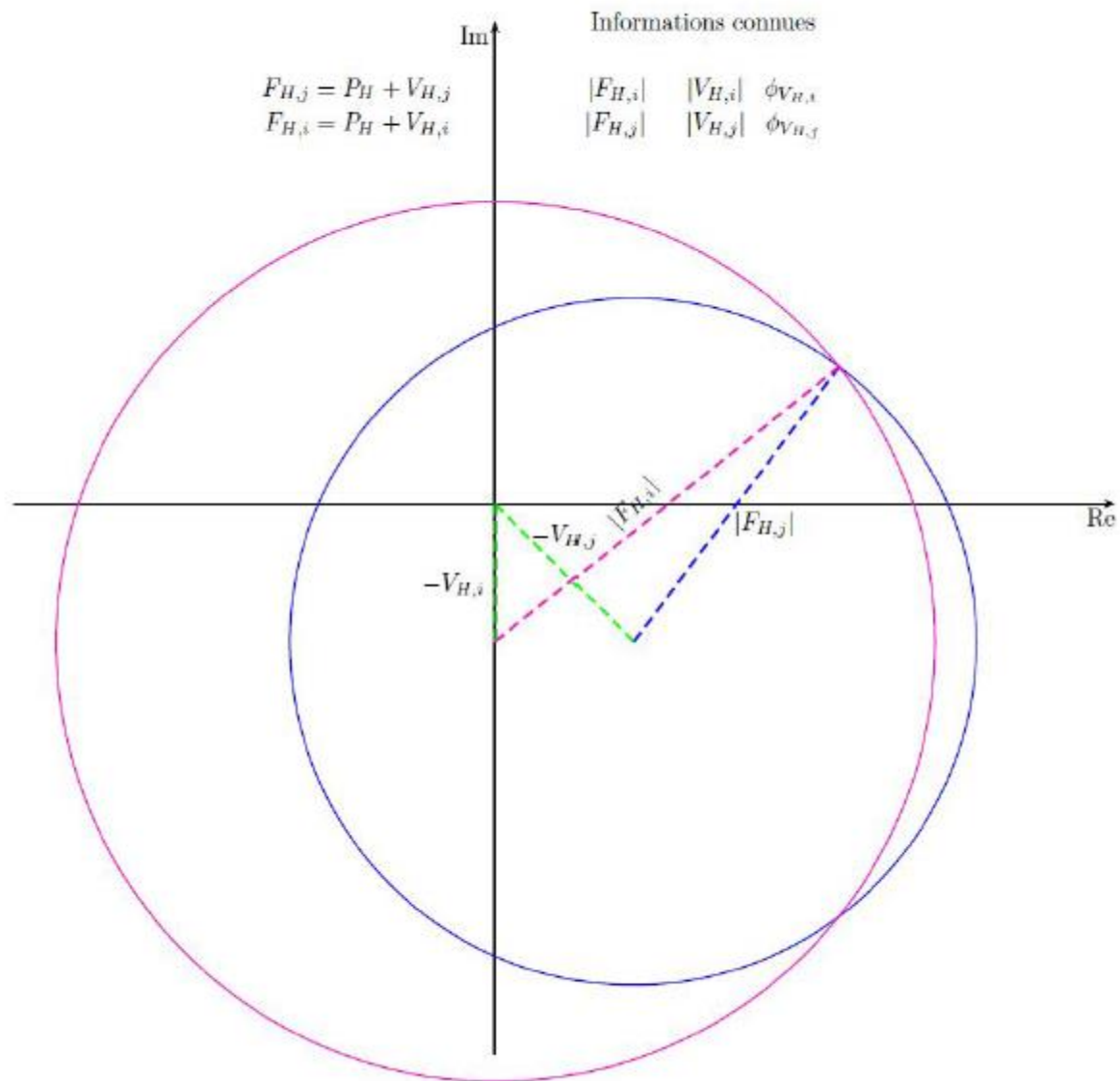


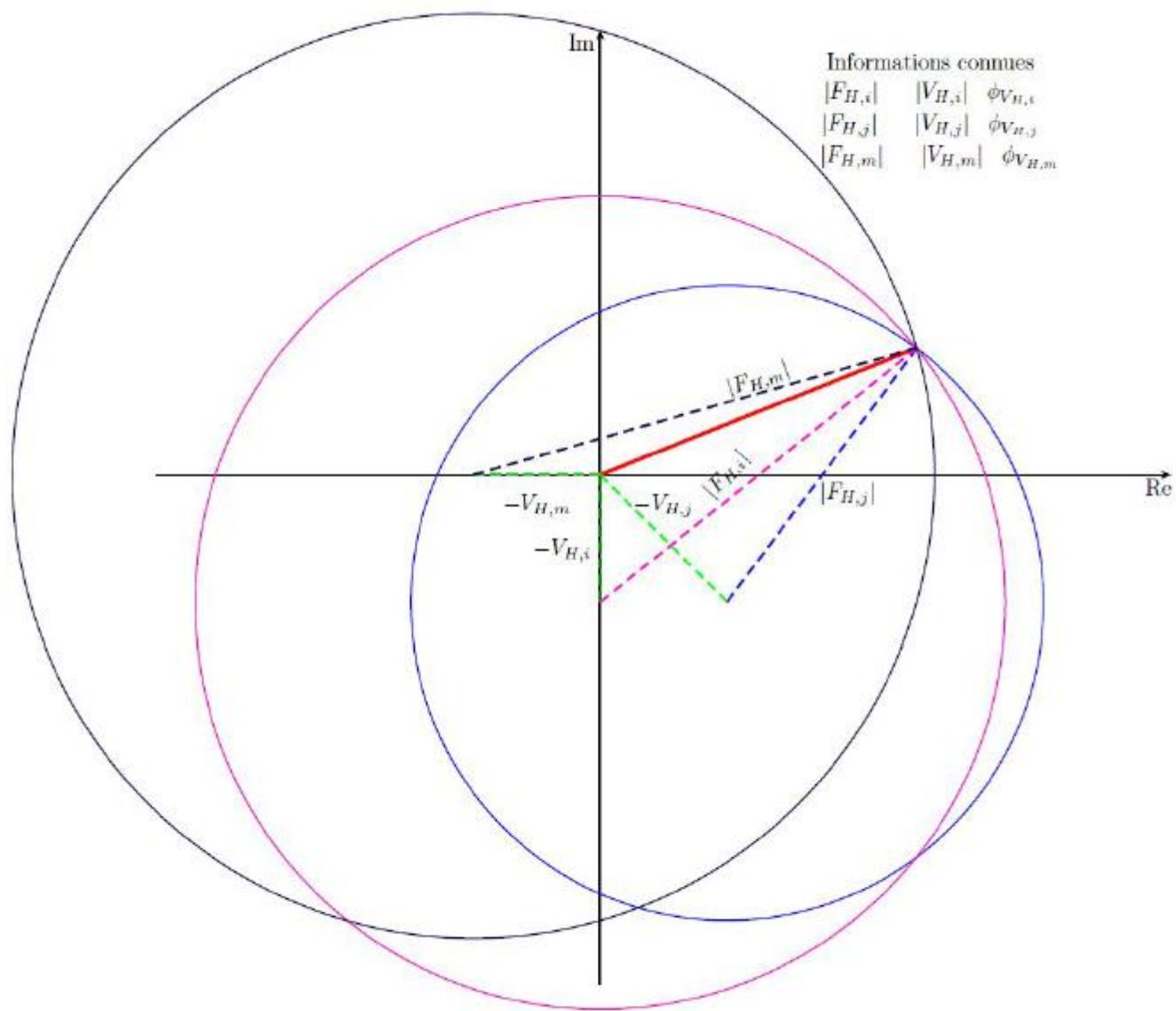


$$F_{H,j} = P_H + V_{H,j}$$

Informations connues
 $|F_{H,j}|$ $|V_{H,j}|$ $\phi_{V_{H,j}}$







$$|F_{H,j}(\theta)| = |P_H + V_{H,j}(\theta)|, \quad j = 1 \rightarrow N_H$$

Dans le cas général, le problème des phases est résolu si $N_H > 2$.

Le processus de phasage :

- ① Collecte des N_H $j = 1 \rightarrow N_H$ jeux de diffraction
- ② Détermination des sous-structures $V_{H,j}$, en particulier les positions atomiques,
- ③ Détermination de la "main" des sous-structures,
- ④ Calculs et interprétation des cartes de densité électronique : est-ce interprétable ?
 - Méthodes de modifications de densité,
- ⑤ Construction, Affinement de la structure,
- ⑥ Validation et contrôle qualité.

Des programmes de plus en plus performants et automatisés

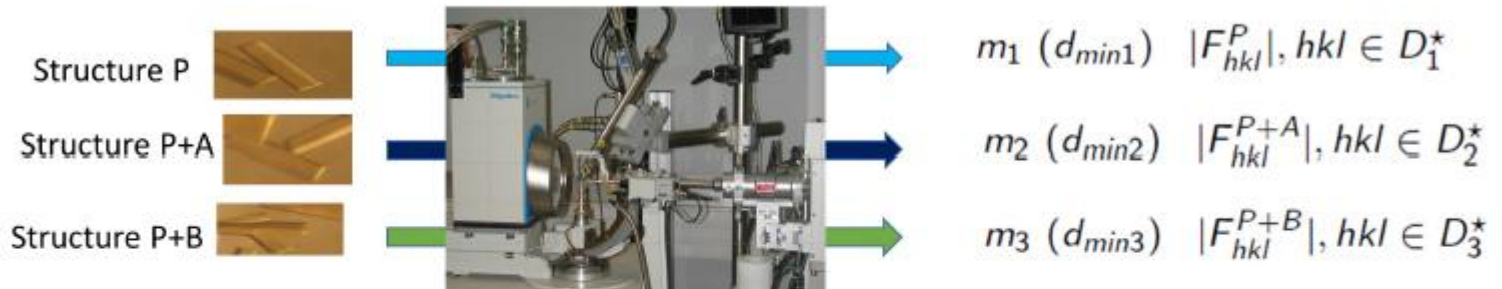
- Des données de diffraction à un premier modèle construit dans une carte de densité électronique modifiée.
- Quelques exemples : CCP4, PHENIX, HKL3000, CNS

$$|F_{H,j}(\theta)| = |P_H + V_{H,j}(\theta)|, \quad j = 1 \rightarrow N_H$$

- "Single Isomorphous Replacement (SIR), Multiple Isomorphous Replacement (MIR)"
j fait référence aux différents dérivés lourds
- "Single-wavelength anomalous diffraction (SAD), Multi-wavelength anomalous diffraction (MAD)"
j fait référence aux différences de "Bijvoet"
- "(site specific) Radiation damage-Induced phasing (RIP)"
j fait référence aux degrés d'irradiations

Méthode du remplacement isomorphe

$$F_{hkl} = \sum_{j=1}^{N+m} q_j f_j w_j e^{2i\pi(hx_j+ky_j+lz_j)} = \sum_{j=1}^N q_j f_j w_j e^{2i\pi(hx_j+ky_j+lz_j)} + \sum_{j=1}^m q_j f_j w_j e^{2i\pi(hx_j+ky_j+lz_j)}$$



Méthode de la diffusion anormale

$$F_{hkl} = \sum_{j=1}^N ({}^0f_j + {}^\lambda f_j' + i {}^\lambda f_j'') q_j w_j^{hkl} e^{2i\pi(hx_j+ky_j+lz_j)}$$



$$\lambda_1 \quad |F_1^+| \quad |F_1^-| \quad F_1' \quad F_1''$$

$$\lambda_2 \quad |F_2^+| \quad |F_2^-| \quad F_2' \quad F_2''$$

(cas 1 seul type de diffuseur anormal)

Les principaux programmes en bio-cristallographie

- Des programmes de plus en plus performants et automatisés

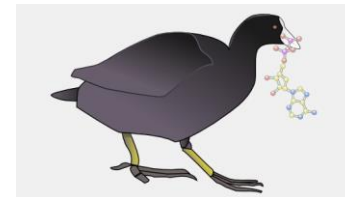
- Quelques exemples : CCP4, PHENIX, HKL3000, CNS

Quelques références actuelles :

- Collecte de données
 - HKL2000, MOSFLM, XDS
- Phasage MIR, MAD, SAD
 - SHARP, SOLVE, SHELX, CRANK,
- Remplacement Moléculaire
 - PHASER, MOLREP, AMoRe, BALBES, MR-ROSETTA
- Modification de densité (avec parfois constructions automatiques)
 - RESOLVE, SOLOMON, ARP/WARP, DM, SHELX
- Construction des structures
 - Automatique
 - Buccaneer, Resolve, Arp/wArp
 - Manuelle
 - Coot, O
- Affinement
 - REFMAC, PHENIX, SHELX, BUSTER, CNS



Phenix





CCP4 Tutorials

ccp4i2 tutorials

- + [ccp4i2 video tutorials](#)
- + [Data processing and SAD phasing](#)
- + [Molecular replacement](#)
- + [Model building and refinement](#)

ccp4i tutorials

- + [Basic phasing tutorials](#)
- + [Integration and data processing](#)
- + [Multi-crystal processing](#)
- + [Refinement with ligands and covalent links](#)

External tutorials

- + [More MX tutorials](#)

PHENIX home

Quit

Preferences

Help

Citations

Reload last job

Other tools

Ask for help

ActionsJob history

Projects

Show group:All groupsManage...

SelectDeleteNew projectSettings

ID	Last modified	# of jobs	R-free
✓ p9-sad	Mar 13 2014 03:0...	34	0.3035

Data analysis

Experimental phasing

Molecular replacement

Model building

Refinement

Validation

Ligands

Reflection tools

Maps

Model tools

Other utilities

Current directory:/Users/nat/Documents/tutorials/p9-sadBrowse...

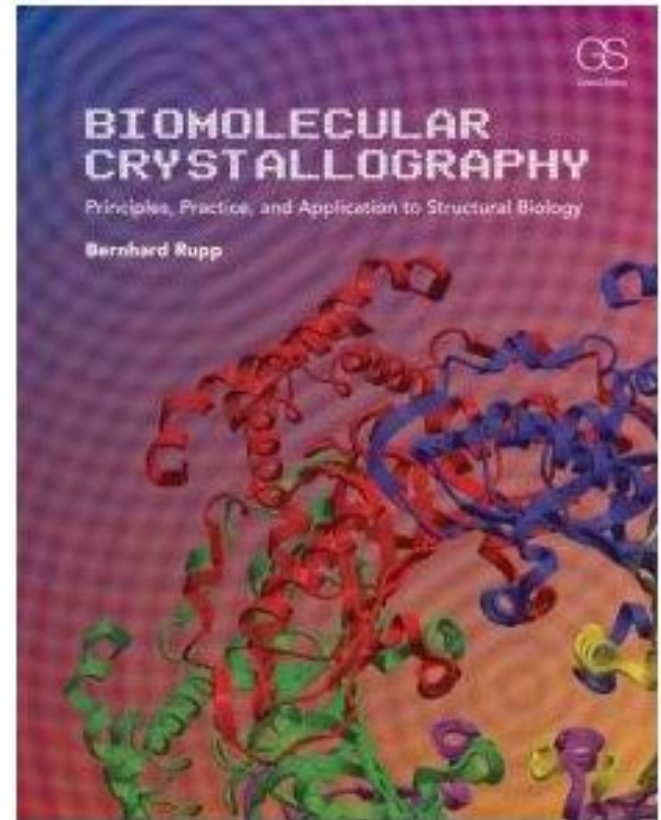
PHENIX version dev-1549Project: p9-sad

Bio - Cristallographie

Déterminations des structures 3D des macromolécules
biologiques par diffraction des rayons X

<https://www.renafobis.fr/>

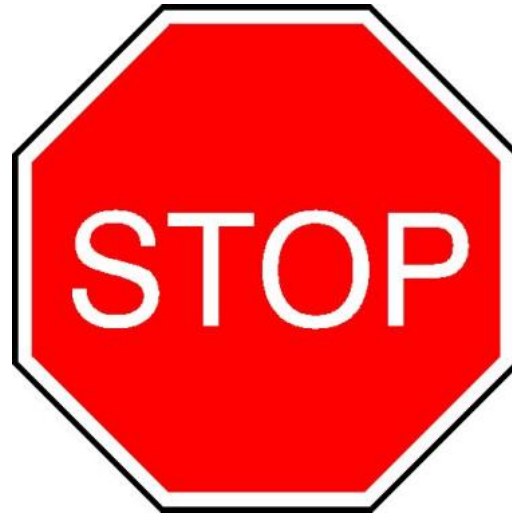
JEAN CAVARELLI
Université de Strasbourg,
Centre de Biologie Structurale Intégrative,
Institut de Génétique et de Biologie Moléculaire et Cellulaire
1, rue Laurent Fries, - 67404 Illkirch-Grafenstaden Cedex
Tél : 03 69 48 52 74 - Fax : 03 88 65 32 76 - e-mail : jean.cavarelli@unistra.fr



http://jean.cavarelli.free.fr/igbmc/labo_services/xray_progs.html

http://jean.cavarelli.free.fr/unistra/Public/biblio/biblio_cristallographie.html

Take-home messages



$$\lambda F(H) = {}^0F_T(H) + \lambda F'_A(H) + i \lambda F''_A(H)$$

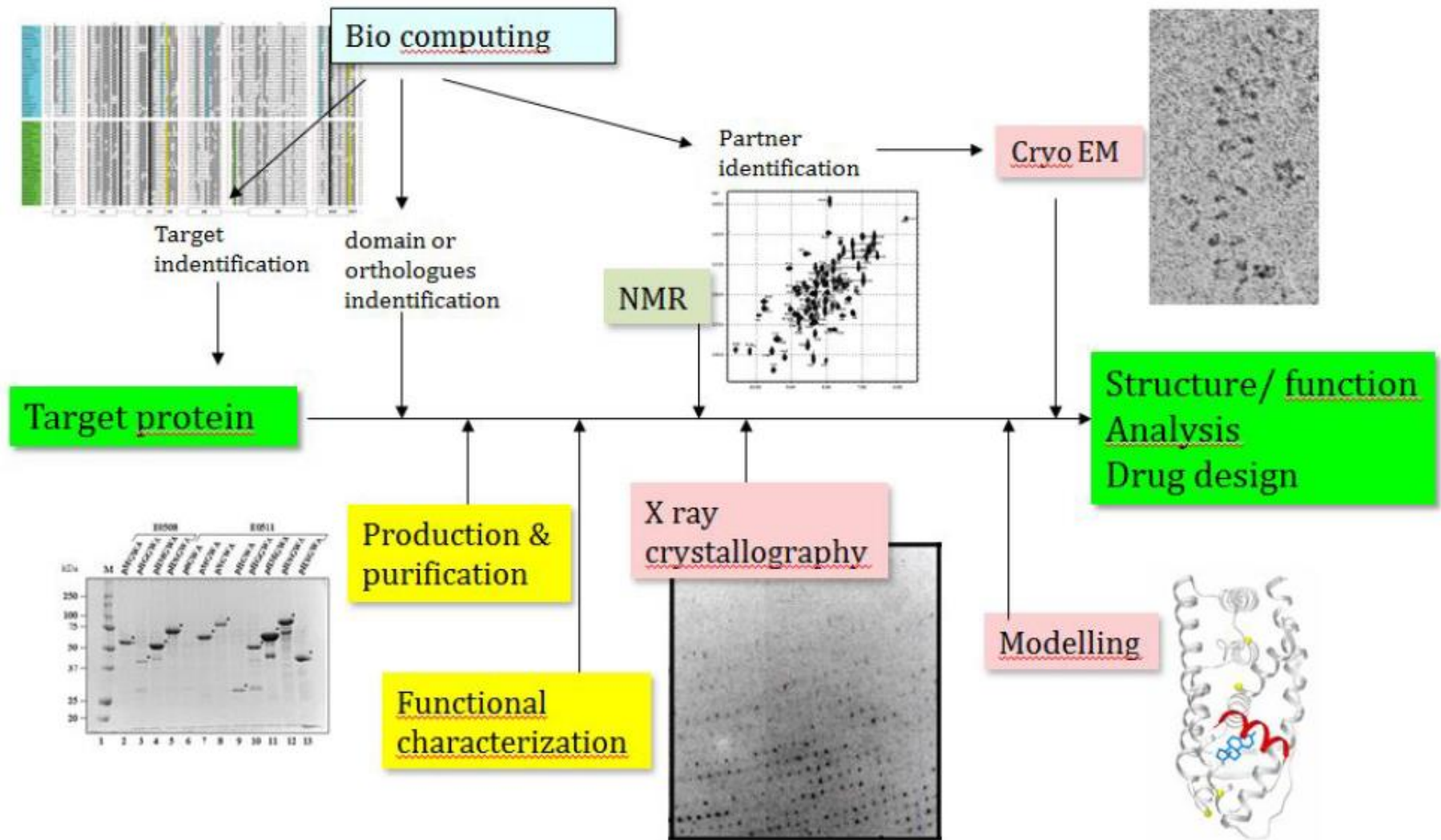
Phasage SAD avec 1 seul type de diffuseur anomaux

- 2 solutions pour ${}^0\phi_T$, symétriques par rapport à ${}^0\phi_A + 90^\circ$
- ${}^0\phi_T^{Vrai} = 2{}^0\phi_A + \pi - {}^0\phi_T^{Vrai}$

en notant $\overline{F}$ le complexe conjugué du facteur de structure F

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h,k,l} ({}^0F_T^{hkl} - \overline{{}^0F_T^{hkl}} e^{2i\phi_A^{hkl}}) e^{-2i\pi(hx+ky+lz)}$$

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h,k,l} {}^0F_T^{hkl} e^{-2i\pi(hx+ky+lz)} - \left(\frac{1}{V}\right) \sum_{h,k,l=-\infty}^{+\infty} \overline{{}^0F_T^{hkl}} e^{2i\phi_A^{hkl}} e^{-2i\pi(hx+ky+lz)}$$



Paramètres de déplacements atomiques

Modèle simple isotrope

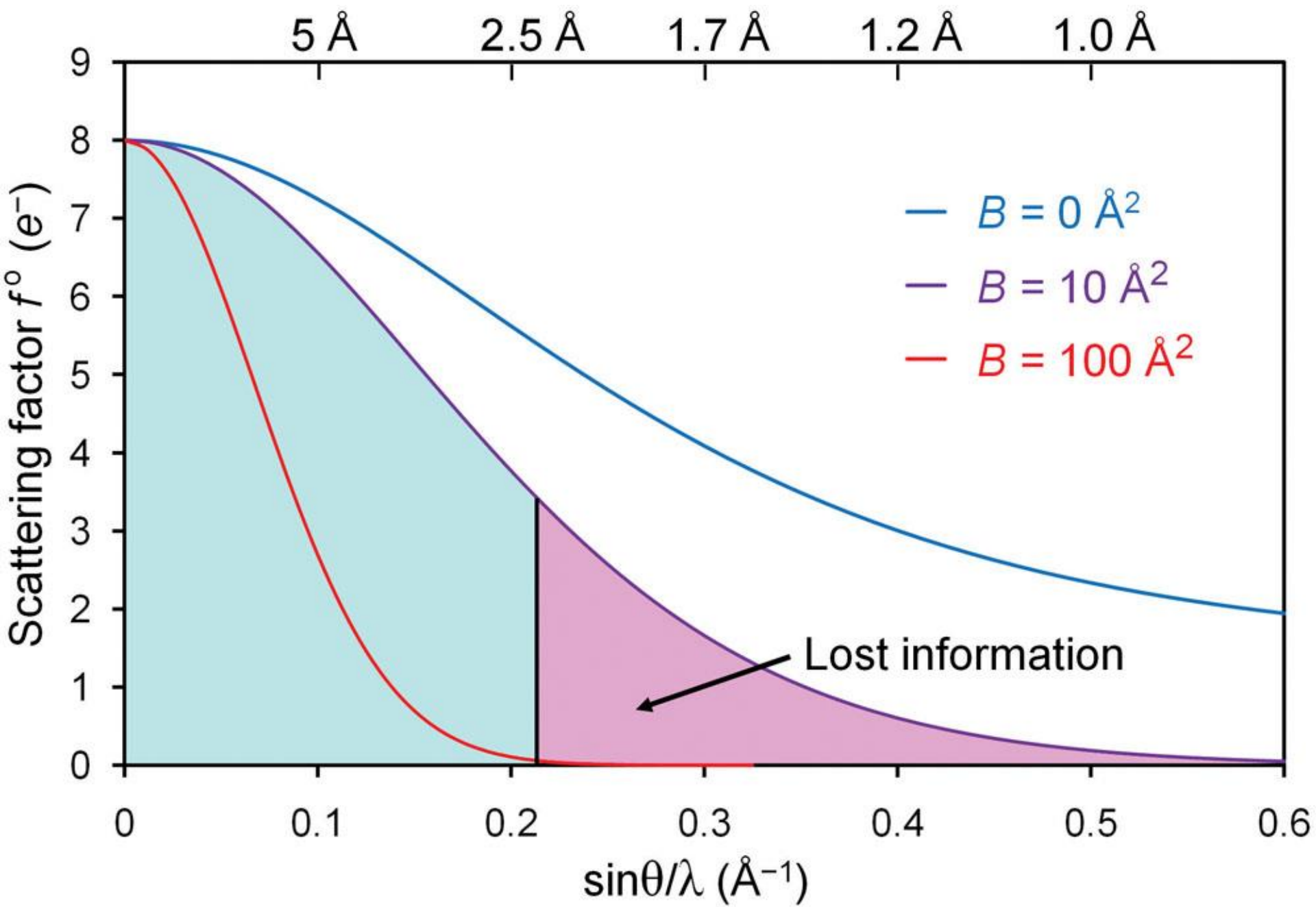
$$w_j^{hkl} = e^{-B_j \left(\frac{1}{2d_{hkl}}\right)^2} = e^{-B_j \left(\frac{\sin \theta}{\lambda}\right)^2}$$

$B_j = \frac{8\pi^2}{3} \bar{U}_j^2$	B (Å) ²	U (Å)
	5	0.25
$\bar{U}_j^2 = \bar{U}_{jx}^2 + \bar{U}_{jy}^2 + \bar{U}_{jz}^2$	80	1.0
	120	1.5

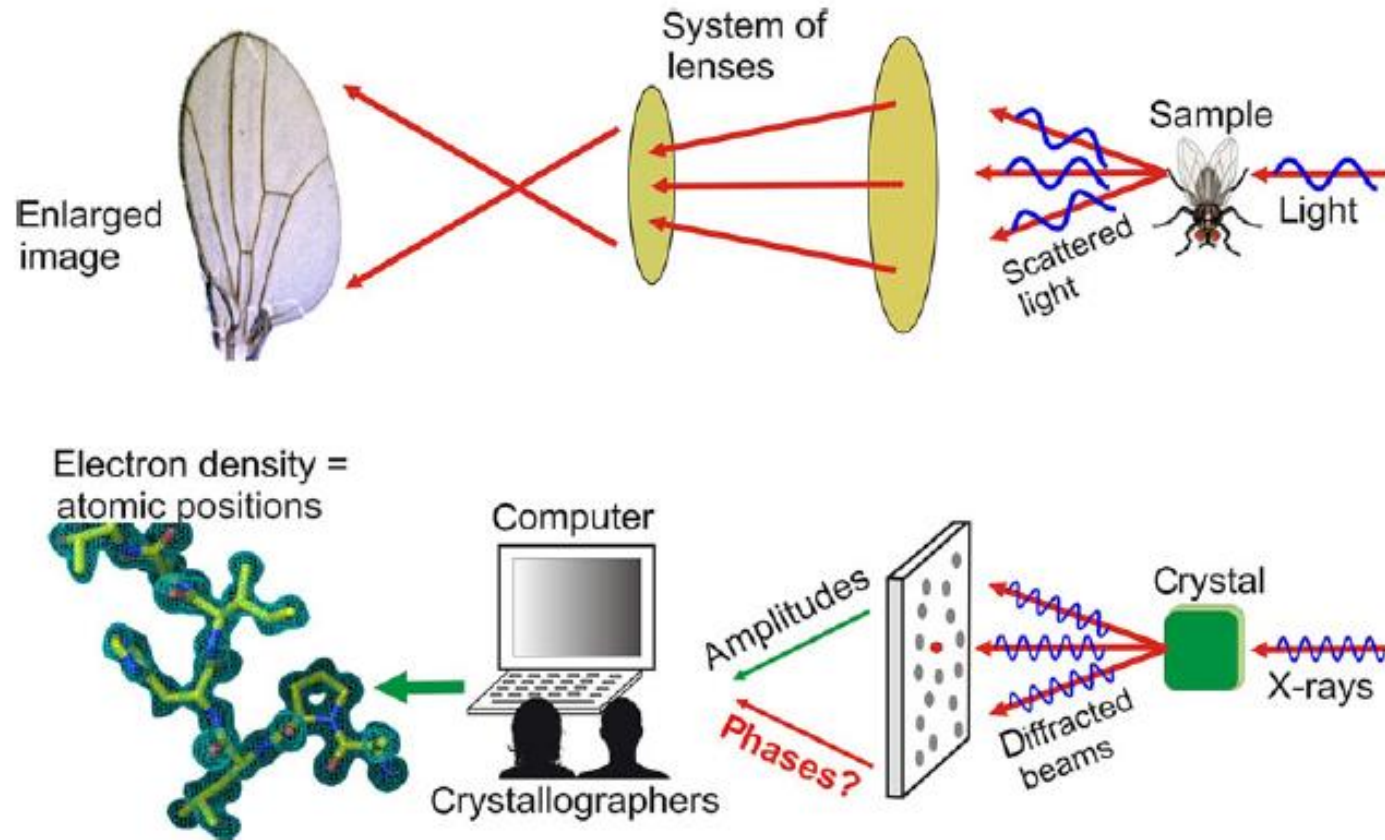
Modèle plus sophistiqué

- matrice de tenseur de déplacement atomique

$$w_j^H = e^{-(H^T \beta_j H)}$$



Comparison of optical microscope (top) and X-ray diffraction (bottom) experiments



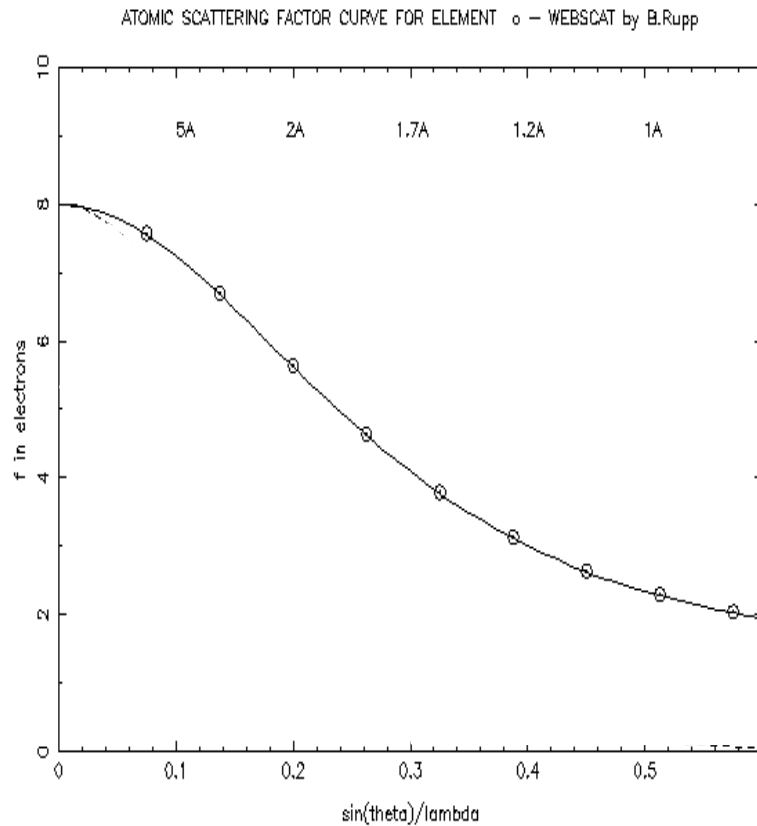
Bio - Cristallographie

Biol. Cell (2014) 106, 323–345

Déterminations des structures 3D des macromolécules biologiques par diffraction des rayons X

Reproduction ULP Strasbourg. Autorisation CFC-Paris

Diffusion par un atome. Symétrie sphérique

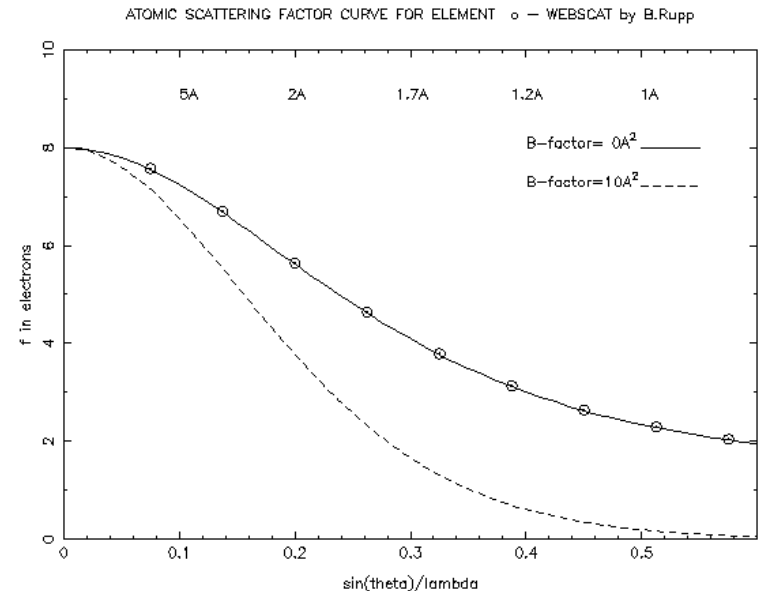


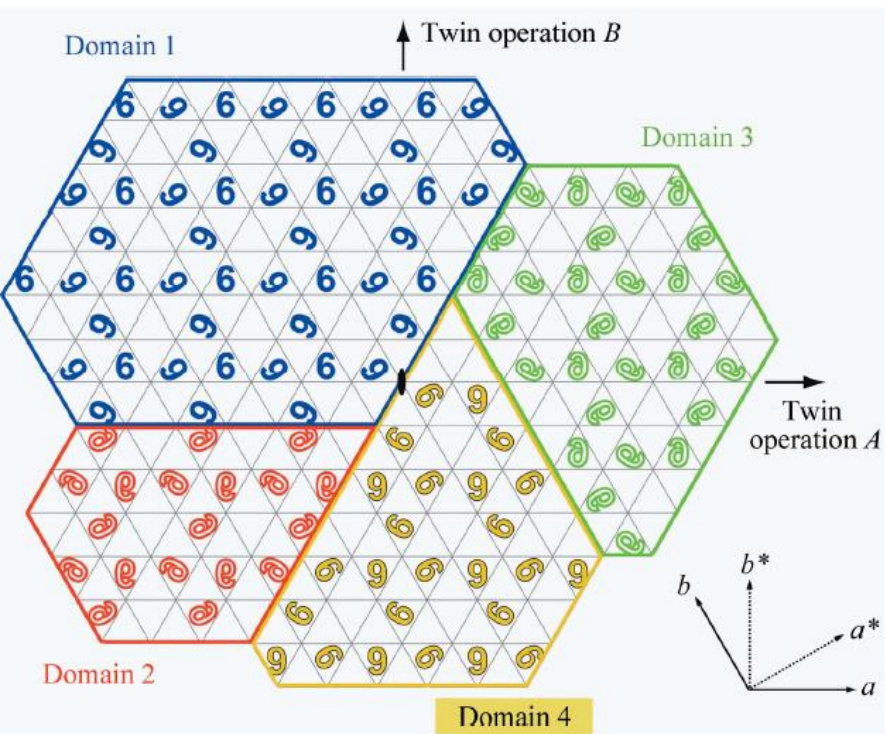
pour $\frac{\sin \theta}{\lambda} = 0, \quad f_i^0 \cong Z_i$

Approximation polynomiale (Cromer, Waber 1968)

$$f^0\left(\frac{\sin \theta}{\lambda}\right) = \sum_{i=1}^{i=4} a_i e^{-b_i \left(\frac{\sin \theta}{\lambda}\right)^2} + c$$

$$f_B = f \cdot e^{-B(\sin \theta / \lambda)^2}$$





Acta Cryst. (2008). D64, 1158–1164

Yeates & Yu • Tetartohedral twin fractions

An illustration of a hypothetical tetartohedrally twinned crystal specimen. The true space group is $P3$. The figure '6' illustrates a single molecule. The four differently oriented twin domains are shown in different colors; in a real crystal these domains might be interspersed throughout the specimen. The red and green molecules are drawn as hollow outlines to emphasize their upside-down orientation in the layer. Three distinct twin operations are present. They constitute symmetry elements of the hexagonal lattice, but not of the $P3$ space-group symmetry. The two twin operations that lie perpendicular to the threefold axis of symmetry are shown as arrows. The third twin operation is parallel to the threefold axis and is indicated by the black symbol in the center. Each of the twin operations exchanges the four twin domains in pairs. For example, operation A exchanges twin domain 1 with 2 and 3 with 4. Operation B exchanges domain 1 with 3 and 2 with 4. Operation C exchanges domain 1 with 4 and 2 with 3. The diffraction patterns from the four domains overlap, so that each observed intensity is the sum of four distinct crystallographic intensities weighted according to the twin fractions (or fractional volumes) of the four domains. With the choice of unit-cell directions shown, the three twin operators A , B and C would correspond to exchanging reflection hkl with $(h, -h - k, -l)$, $(-h, h + k, l)$, and $(-h, -k, l)$, respectively.

L'analyse statistique de la distribution des intensités permet de mettre en évidence et de tenir compte du phénomène

Moments de la distribution des intensités, " Britton plot ", " Rees $N(z)$ plot ",
 "Yeates $S(H)$ plot " , " L-function"
 Voir Truncate, CCP4, phenix.xtriage