

## Phasage - Aspects pratiques

Stéphane Réty, LBMC, ENS Lyon

# Calcul de la densité électronique

- Somme des densités électroniques sur l'ensemble du volume de la maille

$$F(\vec{s}) = F_{hkl} = V \iiint_V \rho(\vec{x}) e^{2\pi i \vec{s} \cdot \vec{x}} d_3x$$

$$F_{hkl} = V \iiint_V \rho(\vec{x}) e^{2\pi i (hx + ky + lz)} d_3x$$

déphasage  
Densité électronique  
au point  $\{x, y, z\}$

- Somme des contributions électroniques des  $n$  atomes de coordonnées fractionnaires  $\{x, y, z\}$  :

$$F_{hkl} = \sum_j f_j e^{2\pi i (hx_j + ky_j + lz_j)}$$

Facteur de diffusion  
Du  $j^{\text{eme}}$  atome

Coordonnées du  
 $j^{\text{eme}}$  atome

# Transformée de Fourier

$$f(x) \xrightarrow{\mathcal{F}} F(s)$$

$$F(s) = \int_{-\infty}^{+\infty} f(x) e^{-2\pi i s x} dx$$

# Transformée de Fourier inverse

$$f(x) \xrightarrow{\overline{\mathcal{F}}} F(s)$$

$$F(s) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{2\pi i s x} dx$$

$$F(s) \xrightarrow{\overline{\mathcal{F}}} f(x)$$

# Produit de convolution

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-u)g(u)du$$

$$f * g = g * f$$

## Produit de convolution et transformée de Fourier

$$f \xrightarrow{\mathcal{F}} F$$

$$g \xrightarrow{\mathcal{F}} G$$

$$f \times g \xrightarrow{\mathcal{F}} F * G$$

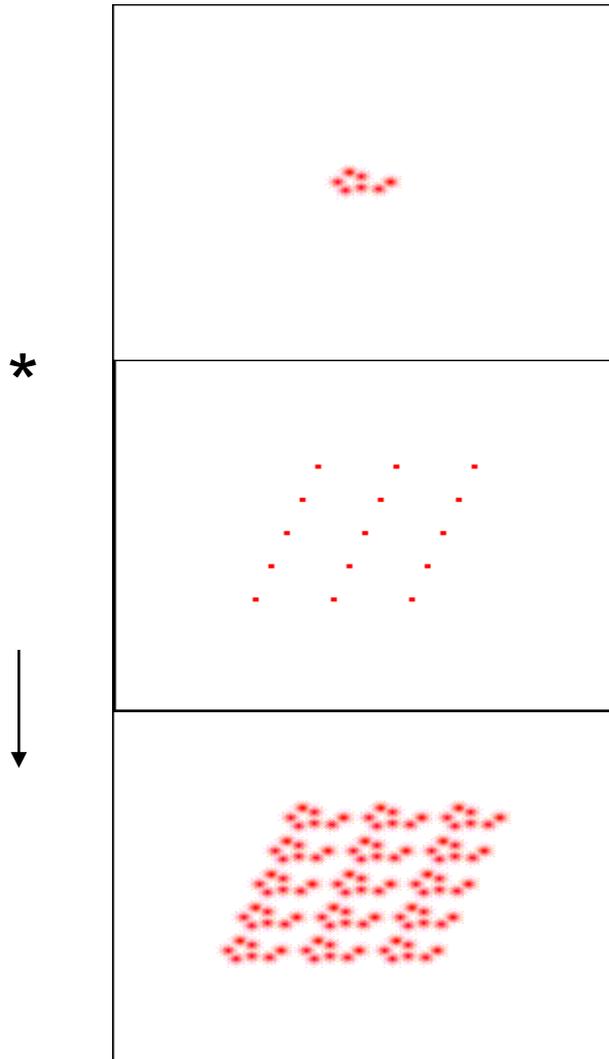
$$f * g \xrightarrow{\mathcal{F}} F \times G$$

# Principes de la cristallographie des rayons X

Espace direct  
Densité électronique

$$\begin{array}{c} \xrightarrow{FT} \\ \xleftarrow{FT^{-1}} \end{array}$$

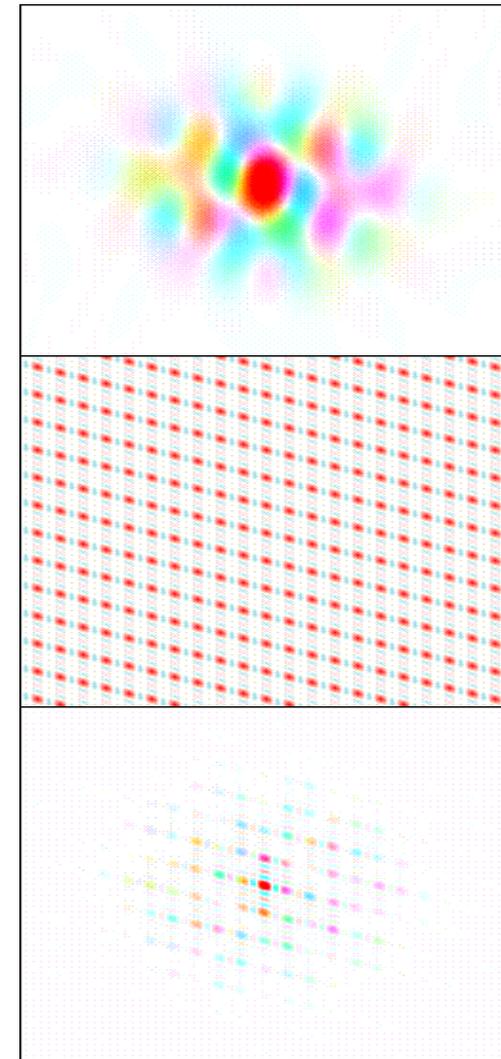
Espace réciproque  
Facteur de structure



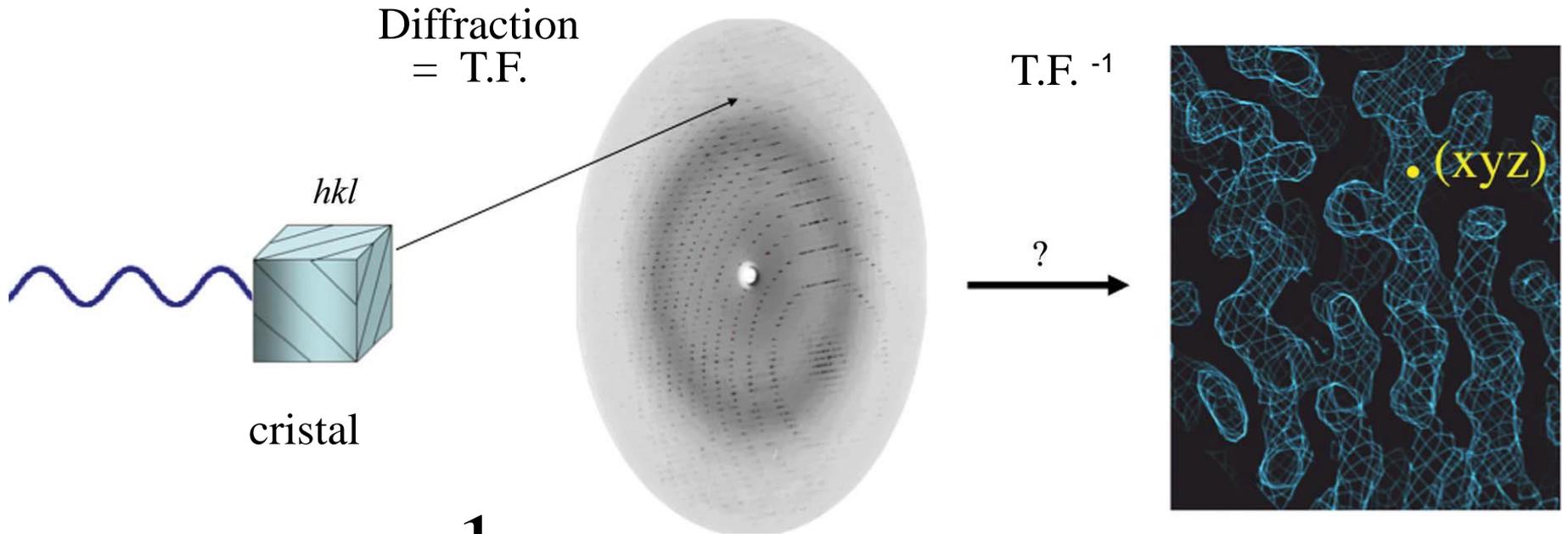
molécule

réseau

crystal



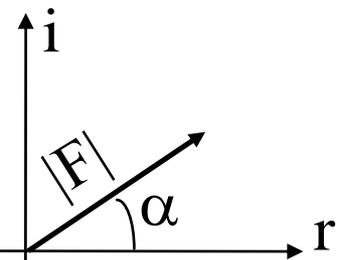
X



$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

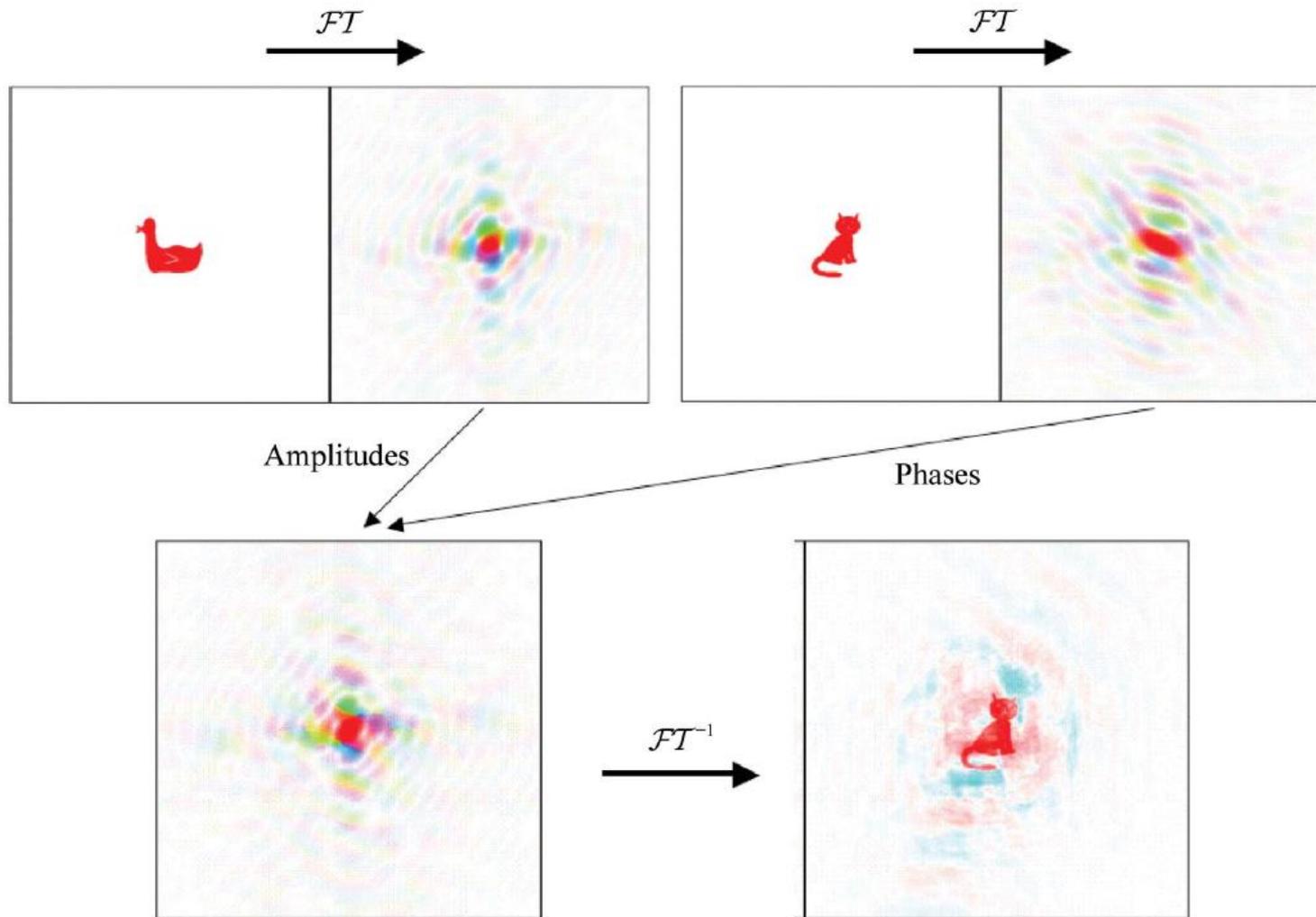
Coefficients = **facteurs de structure** = nombres complexes.

(**modules**  $|F|$  connus à partir des intensités, mais **les phases**  $\alpha$  restent inconnues)



**C'est le problème des phases !**

# L'importance des phases



# Retrouver les phases

---

- Méthodes directes
- Remplacement isomorphe
- Utilisation de la diffusion anormale
- Remplacement moléculaire

# Méthodes directes

- Résolution des données  $< 1,2\text{\AA}$
- Basées sur la positivité et l'atomicité de la densité électronique  
→ relations de phases entre les facteurs de structure
- Relation entre phases de triplet de réflexions

$$\alpha_{-\mathbf{h}} + \alpha_{\mathbf{h}'} + \alpha_{\mathbf{h}-\mathbf{h}'} \simeq 0$$

- Formule de la tangente (affinement)

$$\tan \alpha_{\mathbf{h}} = \frac{\langle E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} \sin(\alpha_{\mathbf{h}'} + \alpha_{\mathbf{h}-\mathbf{h}'}) \rangle_{\mathbf{h}'}}{\langle E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} \cos(\alpha_{\mathbf{h}'} + \alpha_{\mathbf{h}-\mathbf{h}'}) \rangle_{\mathbf{h}'}}$$

Programmes:  
SHELXD  
SnB  
ACORN

# Fonction de Patterson

---

- En 1934, Patterson (1902-1966) a introduit une fonction qui porte désormais son nom:

$$P(\mathbf{u}) = V \int_V \rho(\mathbf{u}) \rho(\mathbf{x} + \mathbf{u}) d\mathbf{x}$$

$$P(\mathbf{u}) = V \rho(\mathbf{x}) * \rho(-\mathbf{x})$$

- Les maxima de la fonction de Patterson correspondent aux vecteurs interatomique

# Fonction de Patterson dans l'espace réciproque

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)}$$

$$\rho(\mathbf{x}) = TF[F(\mathbf{h})]$$

$$\rho(-\mathbf{x}) = TF[F(-\mathbf{h})]$$

$$F(-\mathbf{h}) = \overline{F(\mathbf{h})}$$

$$P(\mathbf{u}) = V \cdot \rho(\mathbf{x}) * \rho(-\mathbf{x})$$

$$P(\mathbf{u}) = V \cdot TF[F(\mathbf{h})] * TF[\overline{F(\mathbf{h})}]$$

$$P(\mathbf{u}) = V \cdot TF[F(\mathbf{h}) \cdot \overline{F(\mathbf{h})}]$$

$$P(\mathbf{u}) = V \cdot TF[I(\mathbf{h})]$$

# Fonction de Patterson

- Si  $\rho(x,y,z)$ , calculée avec  $F$ , dépend de la position des atomes, la même fonction, calculée avec  $F^2$  dépendra des différences entre atomes. On l'appelle  $P(u,v,w)$
- Si  $\rho(x,y,z)$  dépend d'une origine qui doit être connue ( vecteurs 0 -> atomes),  $P(u,v,w)$  ne dépend plus que des vecteurs différence  $\{u,v,w\}$ , donc pas de l'origine.
- Ce Fourier (noté  $\{|F|^2,0\}$  ) est appelé Patterson

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 e^{-2\pi i(hu + kv + lw)}$$

$$P(u, v, w) = \frac{F_{000}^2}{V} + \frac{2}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos 2\pi(hu + kv + lw)$$

- $P(u,v,w)$  calculable en tout point car indépendante des phases
- $P(u,v,w)$  a la même périodicité que  $\rho(x,y,z)$

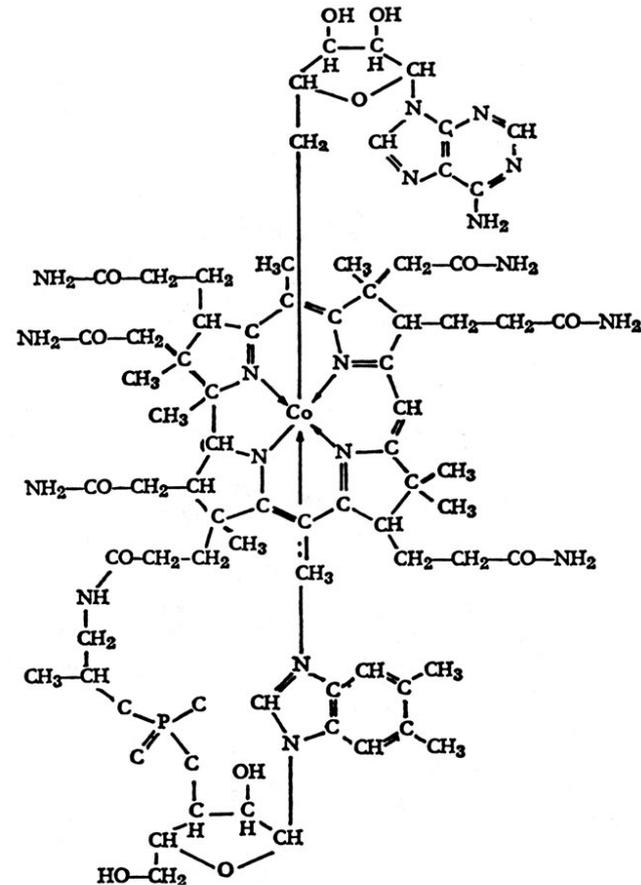
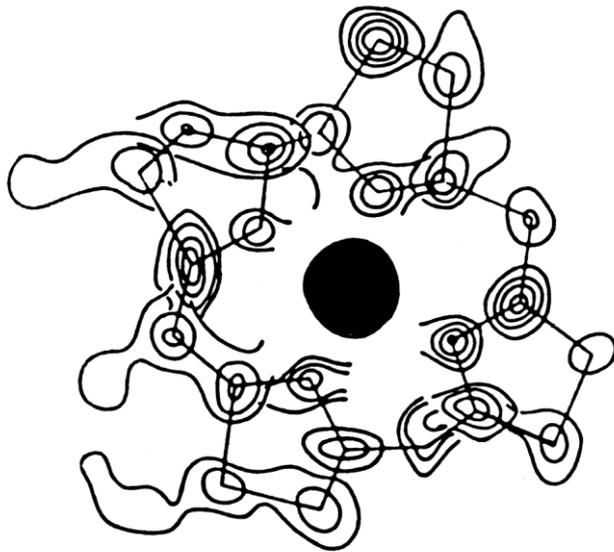
# Fonction de Patterson

	densité électronique	Patterson
maxima	$r_i$	$r_{ij}=r_i-r_j$
nombre de pics	$N$	$N^2$
hauteur	$Z_i$	$Z_i Z_j$

La fonction de Patterson en cristallographie est l'équivalent de la Pair Distance Distribution Function en SAXS

# Structure de la vitamine B<sub>12</sub>

Utilisation de la fonction de Patterson pour résoudre des structures de quelques dizaines d'atomes



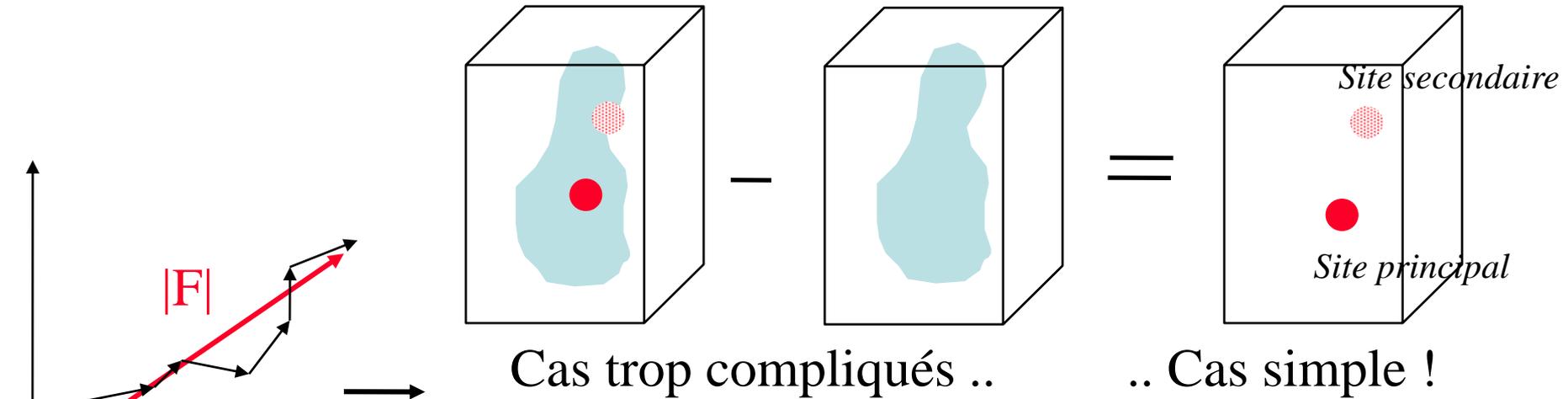
Hodgkin. Nature 176, 325 (1955)

Comment résoudre des structures  
de plusieurs milliers d'atomes?

... en simplifiant le problème

# Résolution du problème des phases: Méthode M.I.R.

(*Molecular Isomorphous Replacement, Remplacement Isomorphe*)



... on a toujours :  $\mathbf{F}_{\text{total}} = \mathbf{F}_{\text{atomes légers}} + \mathbf{F}_a$

Ou pour les protéines :  $\mathbf{F}_d = \mathbf{F}_p + \mathbf{F}_a$        $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$

Isomorphe = même maille, même position

Utilisation d'un "signal"  
de référence facile à  
localiser

Rappel sur les indices :

*a = atome lourd (chez les anglo-saxons, h pour heavy)*

*p = protéine (native)*

*d = dérivé, (PH)*

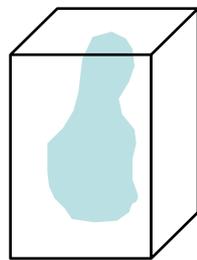
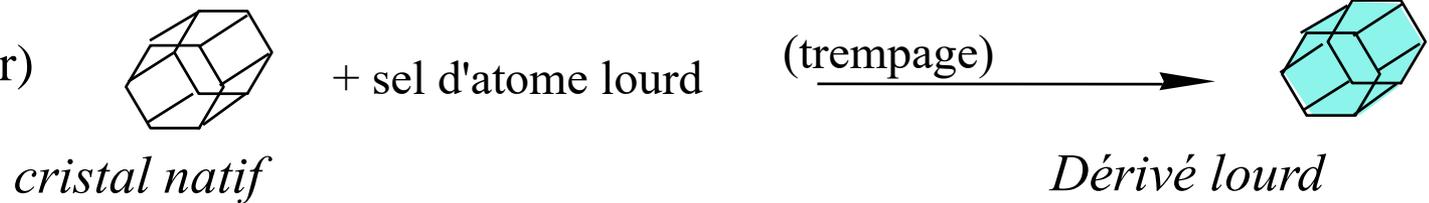
# Comment obtenir la référence en M.I.R. ?

-On fait diffuser un atome lourd dans un cristal pré existant

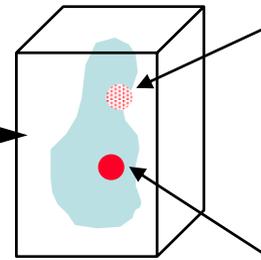
-Co-cristallisation

-Halogénures (I, Br)

-Gaz rare (Xe, Kr)



*Sel*



Site minoritaire

Site majoritaire

*ion*

*Z=*

Acétate d'uranyle



92

Fluoro-uranate de potassium



92

Sodium hexachloroplatinate



78

Chlorure de Mercure



80

triméthyl acétate de plomb



82

Chlorure de plomb



82

Xénon



54

KI



53

Seureate



34

# La référence des phases

Atome  
lourd :

$(x_H, y_H, z_H)$



Phases  $\alpha_H$

$$F_H(hkl) = f_H e^{\underbrace{2\pi i(hx_H + ky_H + lz_H)}_{\alpha_H}}$$

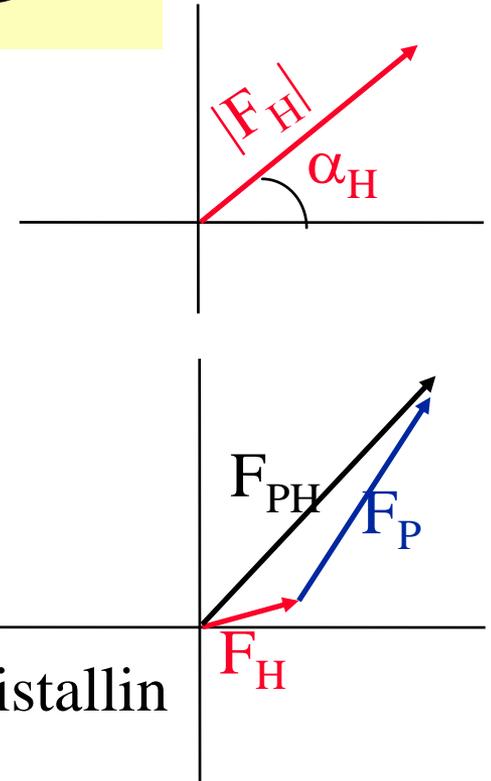
Facteur de diffusion  
de l'atome lourd :

$$f_H = Z_H e^{\frac{-B \sin^2 \theta}{\lambda^2}}$$

condition d'isomorphisme :

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

⇒ L'atome lourd ne doit pas perturber le réseau cristallin



- $|F_H|$  connu à partir des intensités : variation  $F_{PH} - F_P$
- $\alpha_H$  connue à partir de la position  $(x_H, y_H, z_H)$  de l'atome lourd

# Détermination de l'isomorphisme

$$\left\langle \frac{\Delta I}{I} \right\rangle = \sqrt{\frac{N_H}{2N_P} \frac{f_H}{f_P}}$$

$$R_{iso} = \frac{\langle \left| |F_{PH}| - |F_P| \right| \rangle}{\langle |F_P| \rangle} = \sqrt{\frac{N_H}{2} \frac{f_H}{\langle |F_P| \rangle}}$$

Pour 1000 atomes, l'incorporation d'un atome Hg change l'intensité de 25%  
Mais un changement d'isomorphe par un changement de 0.5% des paramètres de mailles ou de la rotation de 0.5° de la molécule dans l'unité asymétrique change les intensités de 15% ...

Pour une protéine, une valeur approchée de  $\langle |F_P| \rangle$  en e<sup>-</sup> peut être calculée par:

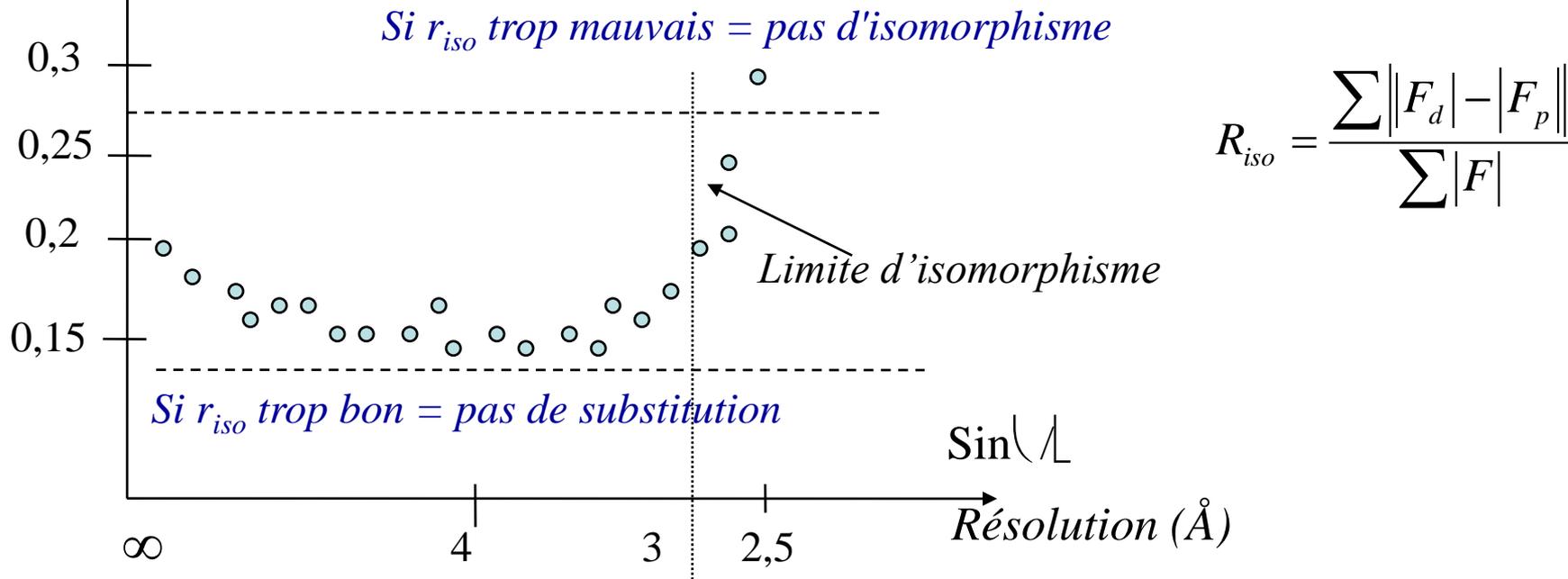
$$\frac{6,7 \times \sqrt{NA}}{\sqrt{346 \times NR}} \\ \sqrt{3,14 \times MM}$$

NA: nombre atomes

NR: nombre de résidus

MM: masse moléculaire

# Détermination de l'isomorphisme



Il est nécessaire de bien connaître la sous-structure des atomes lourds avant de procéder à l'étape de phasage (application de l'égalité  $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$ ).

Qualité des atomes lourds

- Xe, Kr « excellents » dérivés isomorphes
- Pt, Au, Pb... assez bons
- Hg : dérivés peu isomorphes

# Localisation des atomes lourds

---

- Méthodes de Patterson

La **Fonction de Patterson** est une carte de tous les **vecteurs interatomiques** dans la maille

Elle est calculée directement à partir des **intensités diffractées**

- Méthodes directes

Programmes:  
ShelxC+ShelxD  
HySS

# Patterson $(|F_{PH}| - |F_P|)^2$

- Patterson calculée avec comme coefficients  $(|F_{PH}| - |F_P|)^2$
- Estimation de  $F_H^2$  : contribution de l'atome lourd

$$\Delta|F|_{iso} = |F_{PH}| - |F_P| = |F_H| \cos(\alpha_{PH} - \alpha_H) - |F_H| \sin(\alpha_{PH} - \alpha_H) \tan \frac{1}{2}(\alpha_P - \alpha_{PH})$$

Comme  $\alpha_P - \alpha_{PH}$  est petit du fait que  $|F_H| \ll |F_P|$  et  $|F_{PH}|$ , on peut faire l'approximation

$$\Delta|F|_{iso} \approx |F_H| \cos(\alpha_{PH} - \alpha_H)$$

La fonction de Patterson calculée avec  $(\Delta|F|_{iso})^2$  aura donc comme coefficients

$$|F_H|^2 \cos^2(\alpha_{PH} - \alpha_H) = \frac{1}{2}|F_H|^2 + \frac{1}{2}|F_H|^2 \cos 2(\alpha_{PH} - \alpha_H)$$

Le second terme contribue au bruit car les angles  $\alpha_{PH}$  et  $\alpha_P$  ne sont pas corrélés.

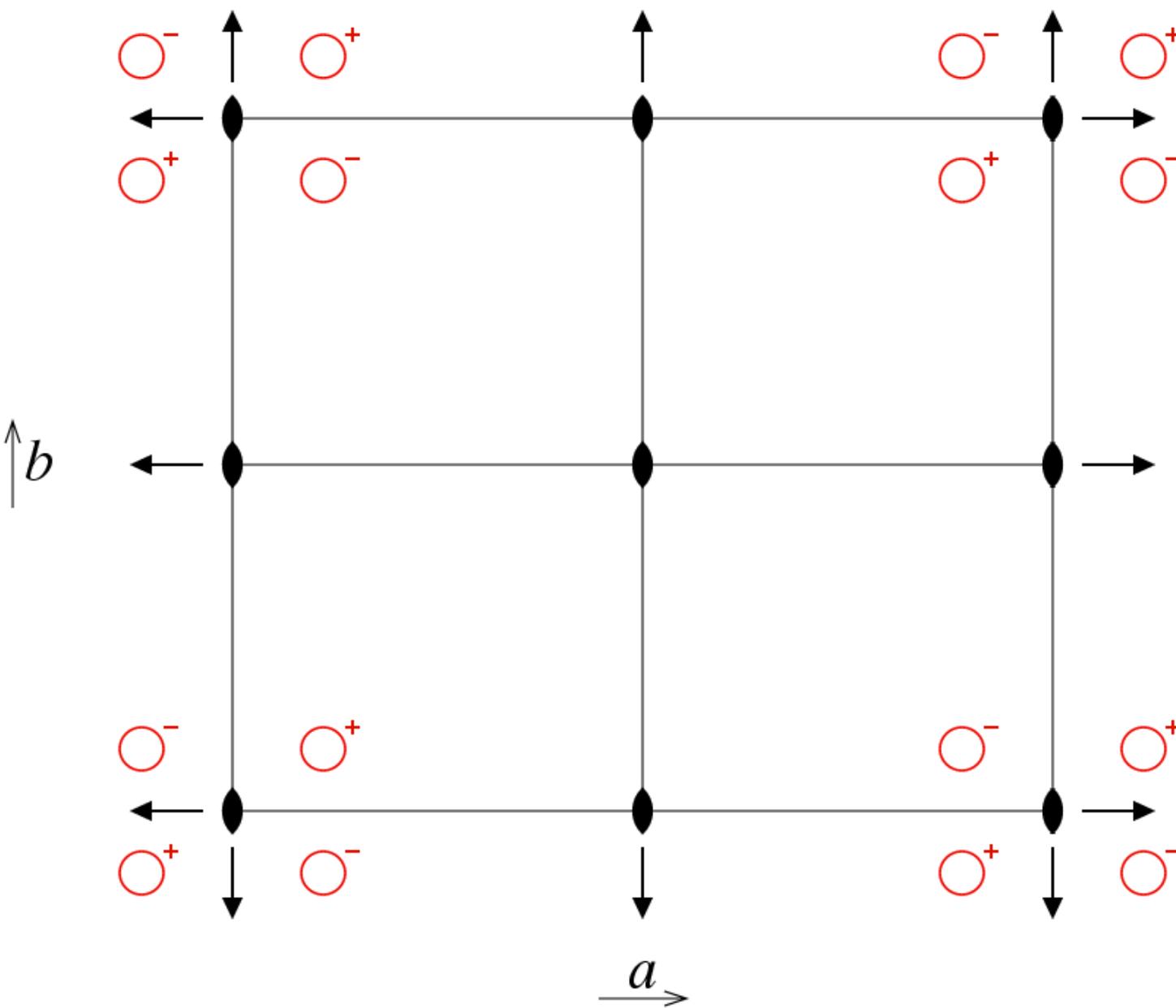
$(\Delta|F|_{iso})^2$  est donc une estimation de  $\frac{1}{2}|F_H|^2$

*P222*

*P 2 2 2*

*222*

No. 16



Origine fixe

# Vecteurs interatomiques entre positions équivalentes

	$x, y, z$	$-x, -y, z$	$x, -y, -z$	$-x, y, -z$
$x, y, z$	0	$2x, 2y, 0$	$0, 2y, 2z$	$2x, 0, 2z$
$-x, -y, z$	$-2x, -2y, 0$	0	$-2x, 0, 2z$	$0, -2y, 2z$
$x, -y, -z$	$0, -2y, -2z$	$2x, 0, -2z$	0	$2x, -2y, 0$
$-x, y, -z$	$-2x, 0, -2z$	$0, 2y, -2z$	$-2x, 2y, 0$	0

# Section w=0

Pics (u, v, 0)

2x, 2y, 0

2x, -2y, 0

-2x, -2y, 0

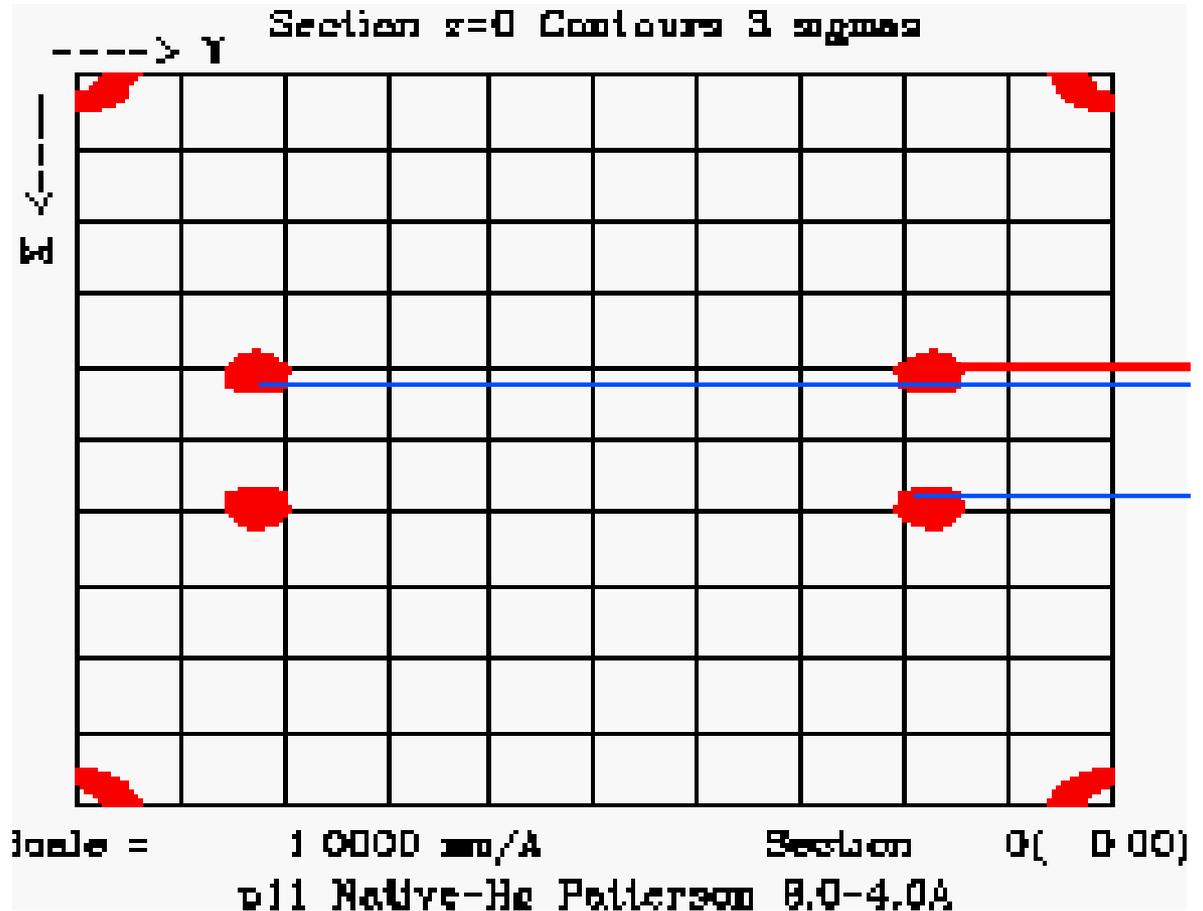
-2x, -2y, 0

0,4 0,175 0

0,6 0,175 0

0,4 0,825 0

0,6 0,825 0



Solutions du type  $u/2 \pm 1, v/2 \pm 1, w/2 \pm 1$

2x=0.4 ou 2x=0.6

x=0.2 ou x=0.3 ou x=0.8 ou x=0.7

2y=0.175 ou 2y=0.825

y=0,0875 ou y=0,9125 ou y=0.4125 ou y=0,5875

# Section $u=0$

Pics (0, v, z)

0, 2y, 2z

0, -2y, 2z

0, 2y, -2z

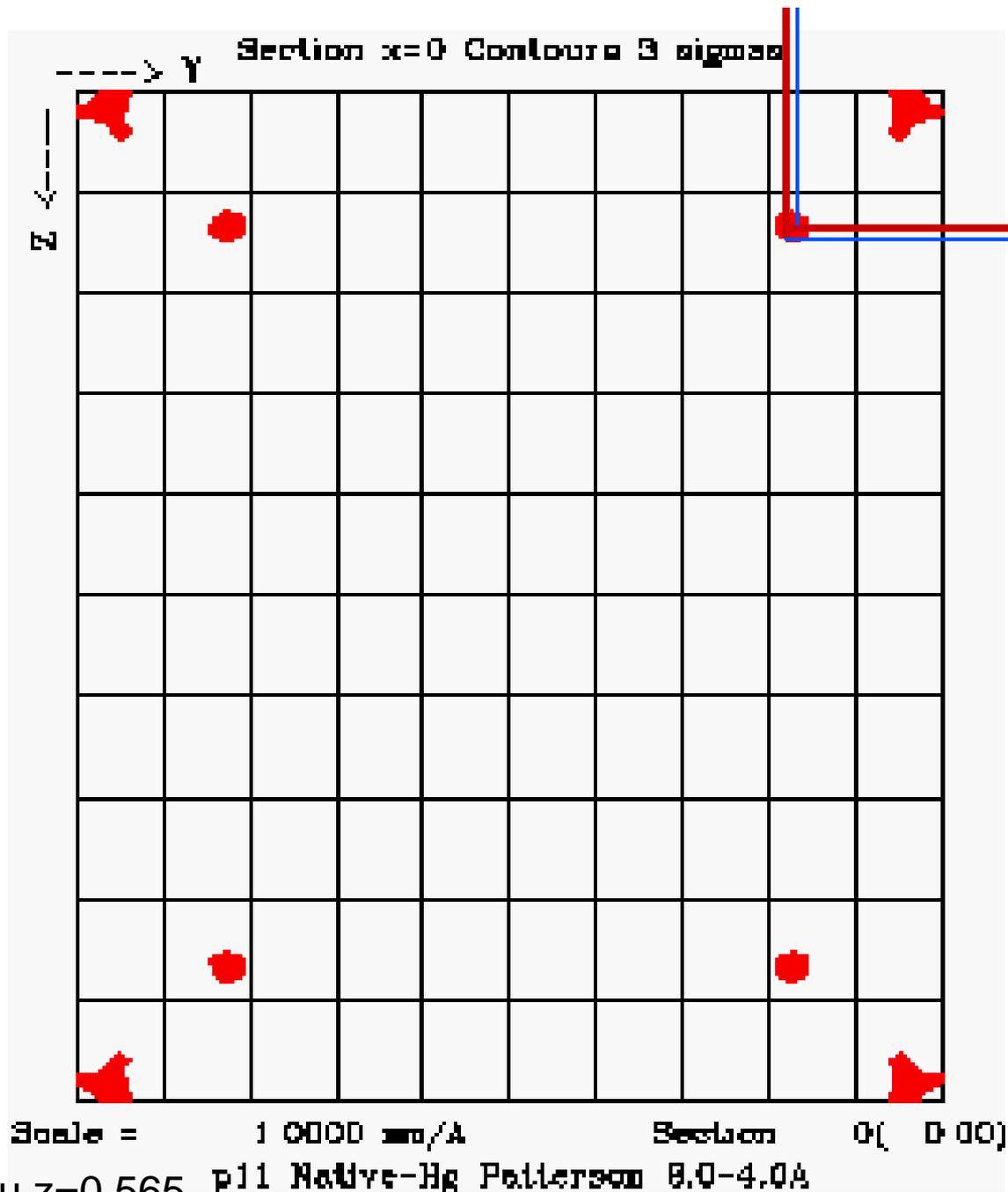
0, -2y, -2z

0 0,175 0,13

0 0,825 0,13

0 0,175 0,87

0 0,825 0,87



$z=0,065$  ou  $z=0,935$  ou  $z=0,435$  ou  $z=0,565$

# Section $v=0$

Pics (u, 0, z)

2x, 0, 2z

-2x, 0, 2z

2x, 0, -2z

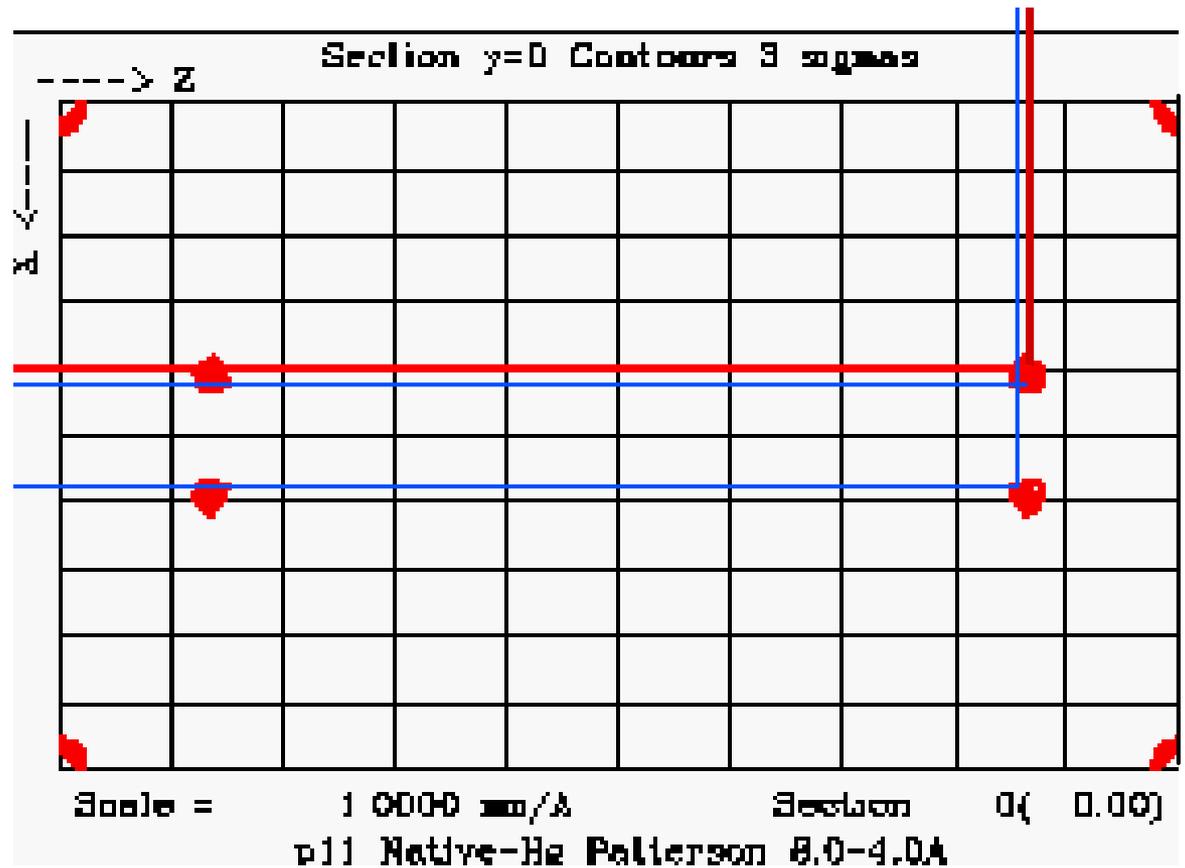
-2x, 0, -2z

0,4 0 0,13

0,6 0 0,13

0,4 0 0,87

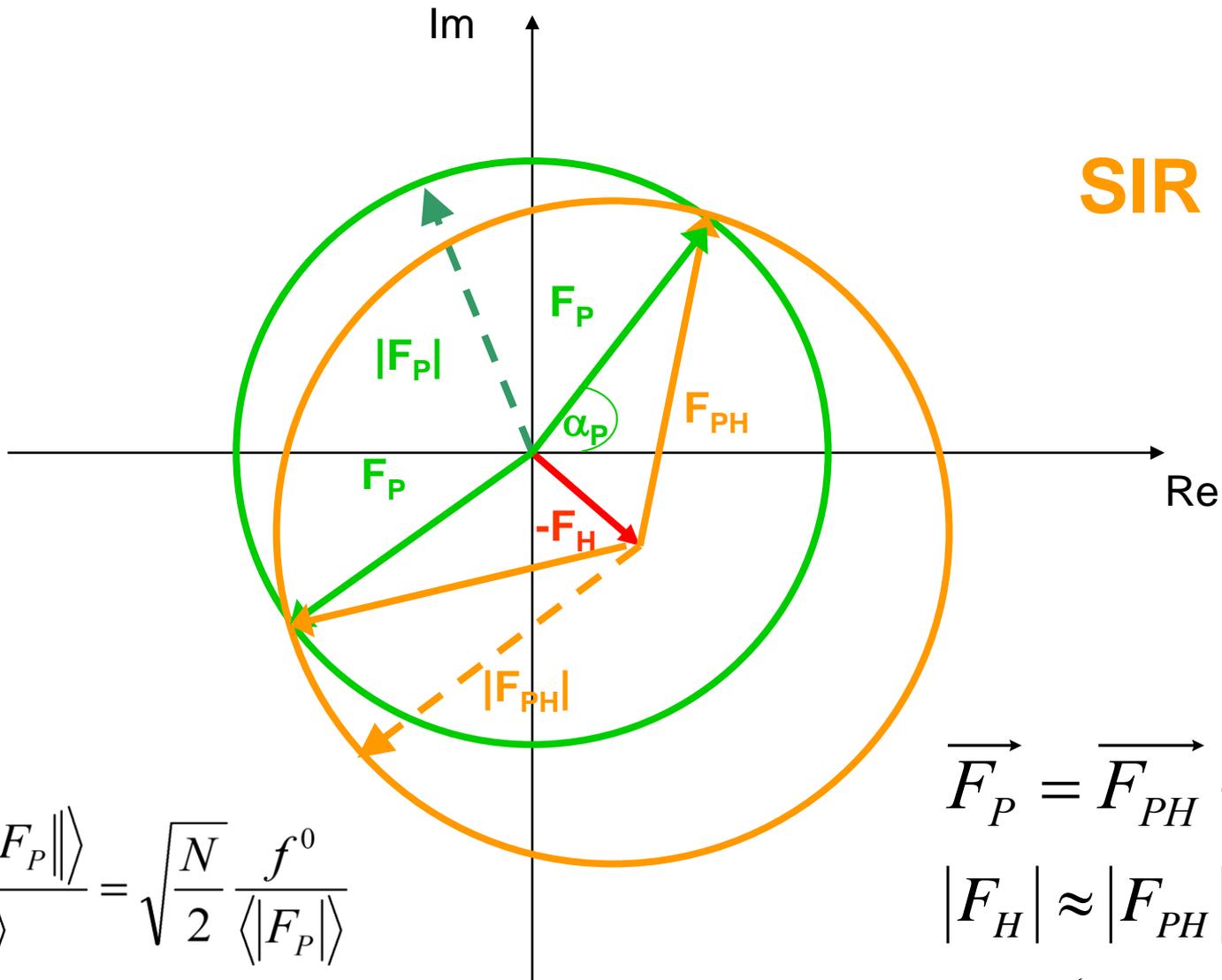
0,6 0 0,87



$x=0.2, y=0,0875, z=0.0625$

# Construction de Harker

SIR



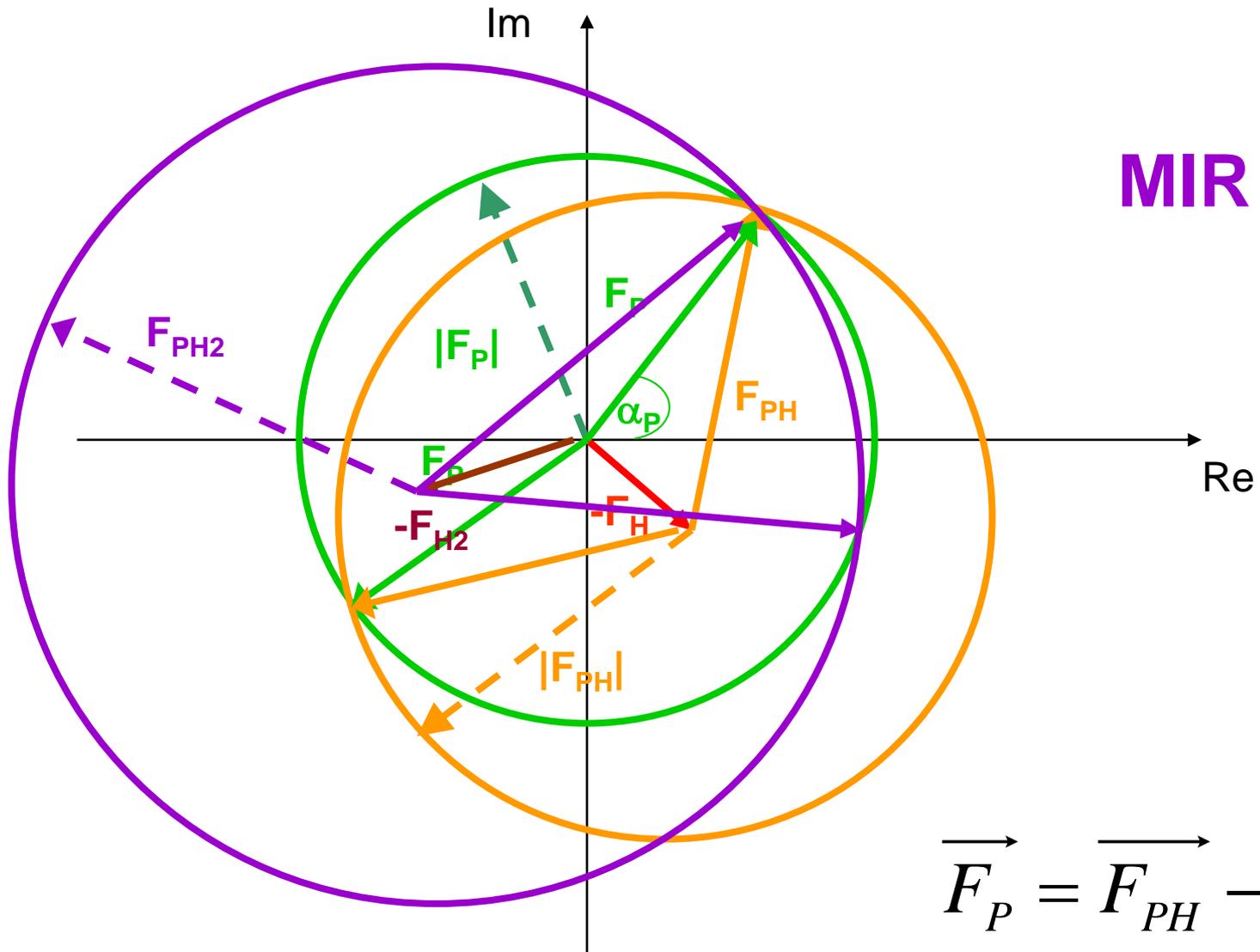
$$R_{iso} = \frac{\langle \left| |F_{PH}| - |F_P| \right| \rangle}{\langle |F_P| \rangle} = \sqrt{\frac{N}{2}} \frac{f^0}{\langle |F_P| \rangle}$$

$$\vec{F}_P = \vec{F}_{PH} - \vec{F}_H$$

$$|F_H| \approx |F_{PH}| - |F_P|$$

$$\alpha_P = \alpha_H \pm \cos^{-1} \left( \frac{F_{PH}^2 - F_P^2 - F_H^2}{2F_P F_H} \right)$$

# Construction de Harker



$$\vec{F}_P = \vec{F}_{PH} - \vec{F}_H$$

$$\vec{F}_P = \vec{F}_{PH2} - \vec{F}_{H2}$$

# Construction de Harker

---

⇒ Problème du défaut d'isomorphisme

C'est l'écart de structure de la protéine entre les formes cristallines native et dérivé.  
La relation  $F_{PH} = F_P + F_H$  n'est plus vraie.

- La diffraction d'un dérivé est en général moindre que celle d'un cristal natif (réseau perturbé par l'atome lourd)
- Les paramètres de maille doivent être identiques ( $\pm 5\%$  max.)
- Il faut des  $F$  communs entre dérivé et native : la limite pratique de résolution est donc dictée par le plus mauvais des deux !
- Mise à l'échelle précise des jeux de données  $F_{PH}$  et  $F_P$  pour limiter les erreurs systématiques

# Dans la pratique ...

Erreurs de :

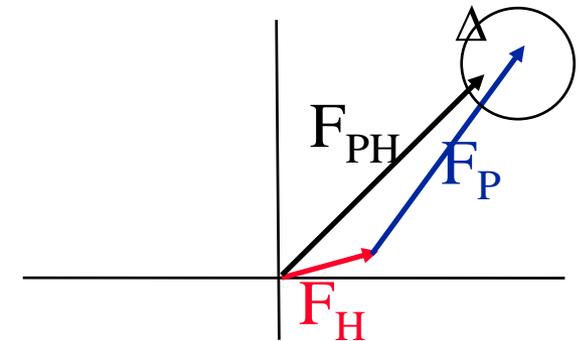
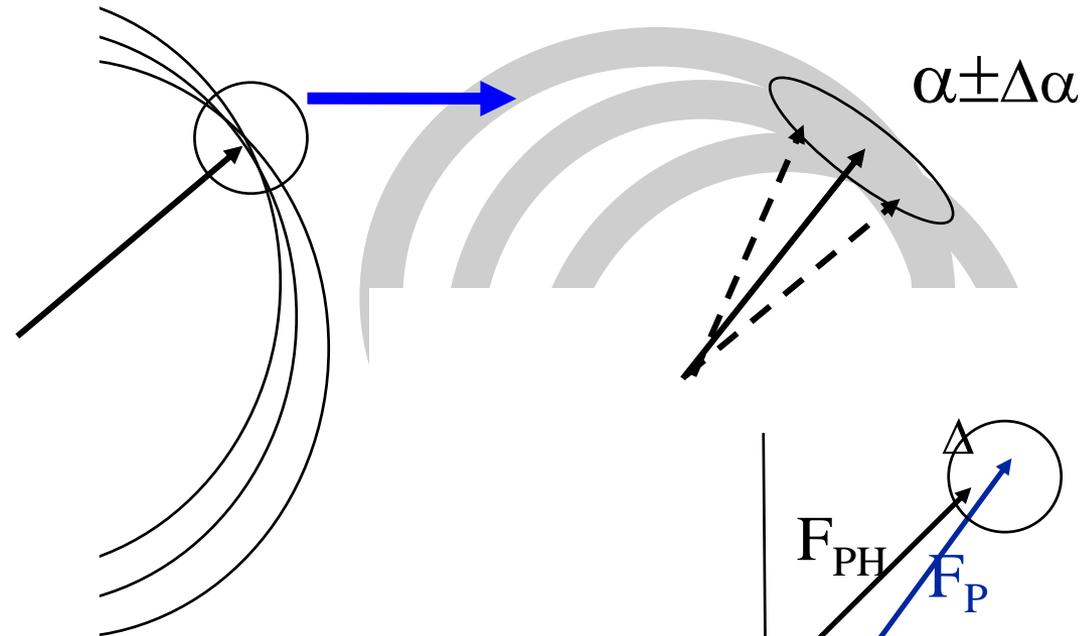
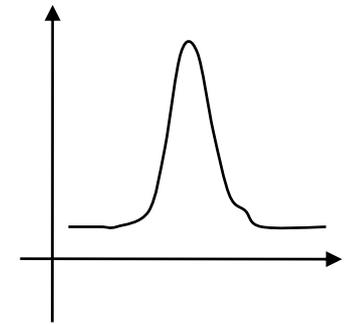
défaut d'isomorphisme

de mesure

de mise à l'échelle

...

*proba*



Cela s'appelle aussi le défaut de fermeture

(du triangle d'isomorphisme) :  $\varepsilon = \langle |F_{PH}|_{\text{calc}} - |F_{PH}|_{\text{obs}} \rangle$

$|F_{PH}|_{\text{calc}}$  est obtenu par trigonométrie :

$$= |F_P|^2 + |F_H|^2 + 2 |F_P| |F_H| \cos(\alpha_P - \alpha_H)$$

*indices :*

*H = atome lourd*

*P = protéine (native)*

*PH = dérivé*

# Cas du SIR

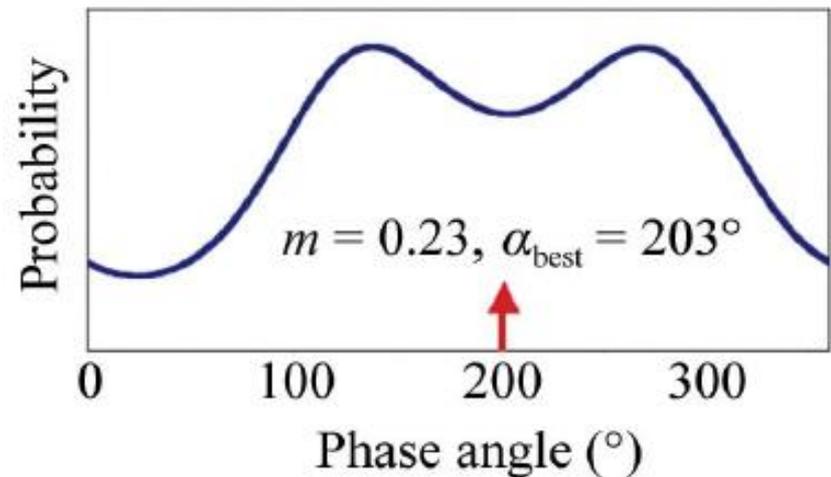
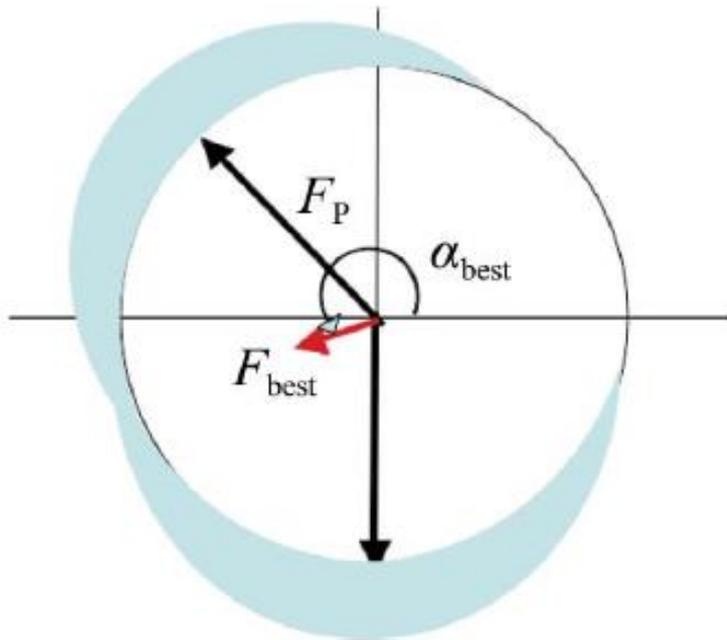
Défaut de fermeture

$$\varepsilon = \left| F_{PH(obs)} \right| - \left| F_{PH(calc)} \right|$$

$$\varepsilon = \left| F_{PH(obs)} \right| - \left| \left[ \left| F_P \right| e^{i\alpha_P} + \left| F_H \right| e^{i\alpha_H} \right] \right|$$

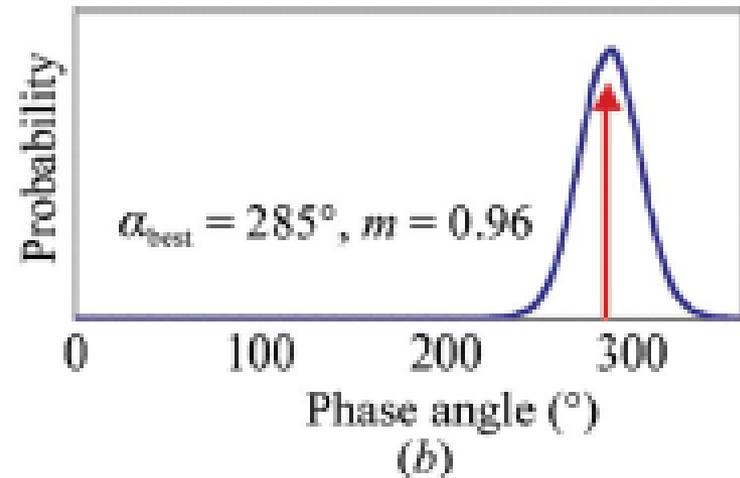
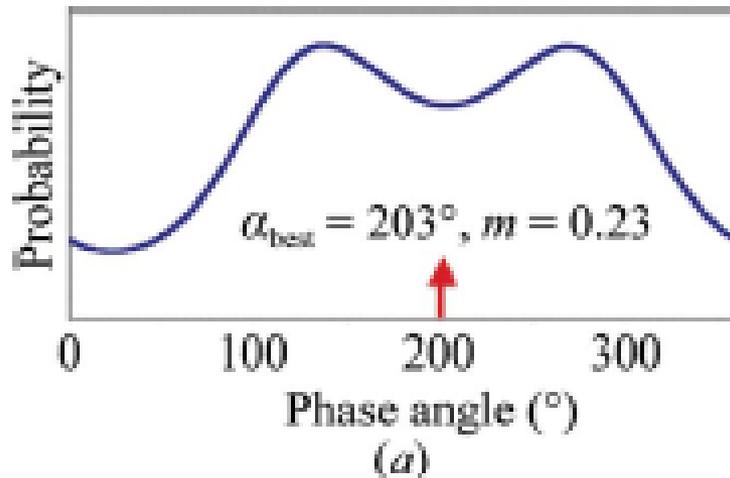
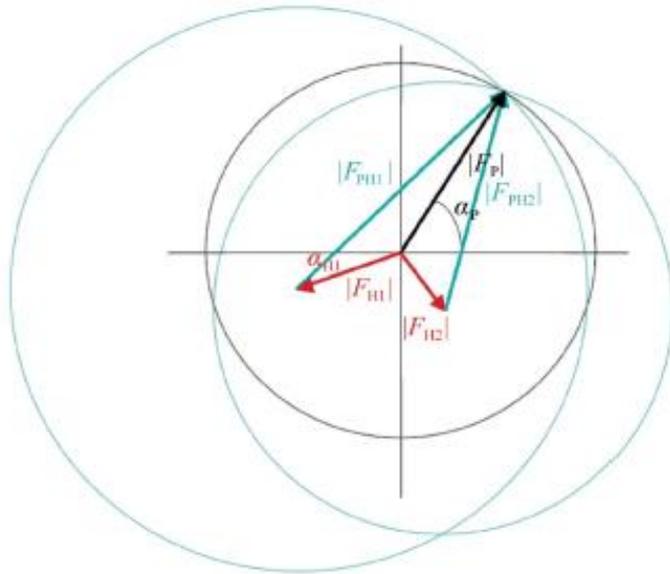
$$P(\alpha) = N e^{\frac{-\varepsilon^2}{2E^2}}$$

$$E = \left\langle \left| F_{PH(obs)} - F_{PH(calc)} \right|^2 \right\rangle$$

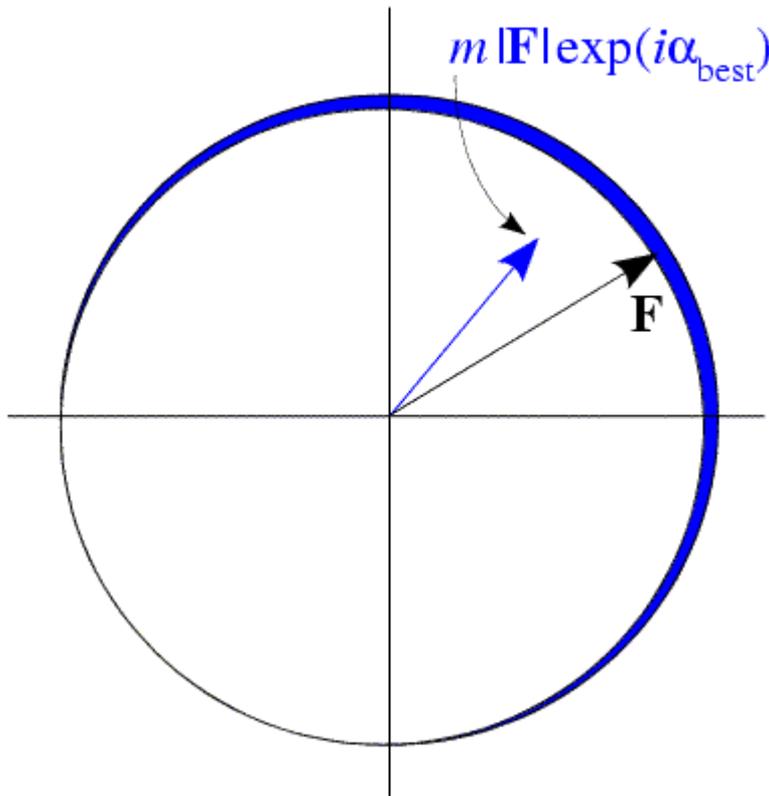


$\alpha_{best}$ : centroide de la distribution de phase

# Cas du MIR



# Figure of merit (FOM)



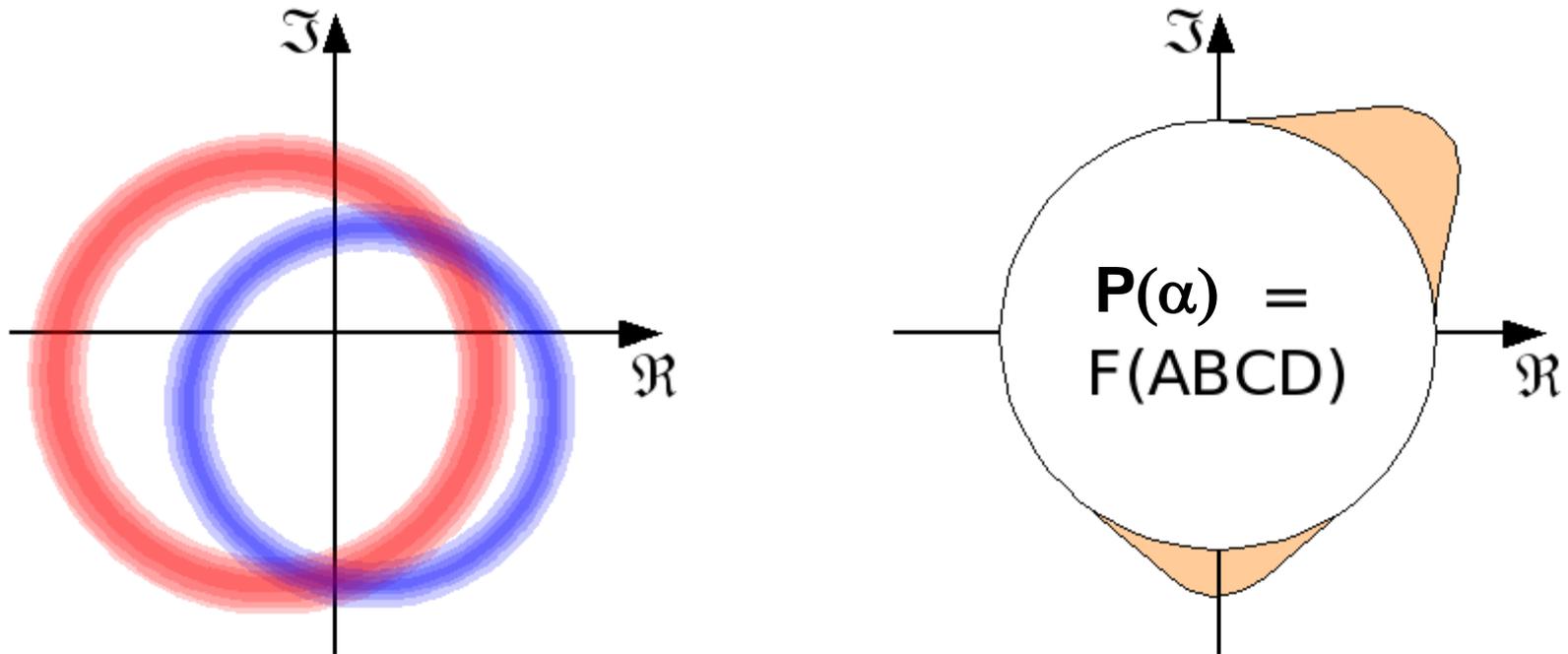
$$m = \frac{|F_{hkl}(best)|}{|F_{hkl}|}$$

$$m \cos \alpha_B = \frac{\sum_i P(\alpha_i) \cos \alpha_i}{\sum_i P(\alpha_i)}$$

$$m \sin \alpha_B = \frac{\sum_i P(\alpha_i) \sin \alpha_i}{\sum_i P(\alpha_i)}$$

Best Fourier avec  $m|\mathbf{F}|e^{i\alpha_B}$

# Coefficients de Hendrickson et Lattman



Description de la répartition bimodale

$$P(\alpha) = N \exp( A \cos(\alpha) + B \sin(\alpha) + C \cos(2\alpha) + D \sin(2\alpha) )$$

Les coefficients A et B représentent une distribution unimodale centrée sur  $\alpha = \text{atan}(B/A)$ .  
Les coefficients C et D modulent cette distribution par une seconde distribution unimodale symétrique.

Si C et D sont nuls, la distribution est celle représentée par  $\alpha_{\text{best}}$  et la figure of merit

$$A = X \cos(\alpha_{\text{best}}), \quad B = X \sin(\alpha_{\text{best}}),$$

$$FOM = \tanh(X/2) \text{ (centric) and } FOM = I1(X)/I0(X) \text{ (acentric).}$$

# Cas des réflexions centrées

La projection d'une structure selon un axe d'ordre pair est toujours centro-symétrique

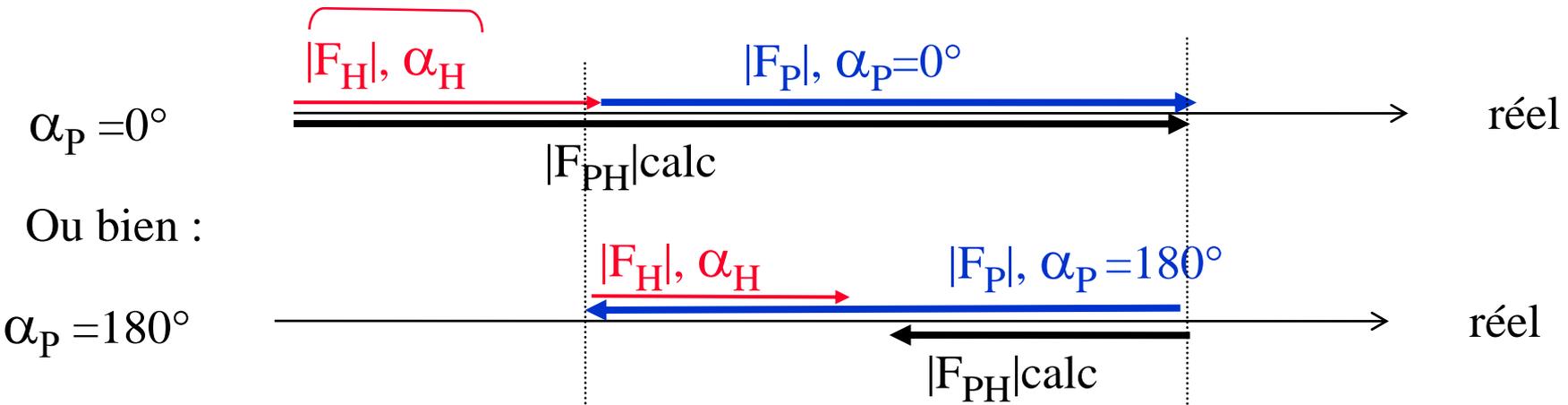
Ex  $P2_1 = 1$  zone centrée : réflexions  $h0l$

$P2_12_12_1 = 3$  zones centrées :  $h0l$ ,  $0kl$  et  $hk0$

P1 ou R3 : pas de refl. centrée

Avantage : 2 phases possibles pour  $\alpha_p = 0/180^\circ$

*Connus (Patterson)*



➔ On a juste à comparer  $|F_P|_{\text{obs}}$  avec les 2 valeurs de  $|F_P|_{\text{calc}}$  obtenus en considérant  $\alpha_p = 0$  ou  $180^\circ$

...Et à choisir le meilleur ➔  $\alpha_p$  déterminée !

# Statistiques et validité du phasage

Critères d'un bon phasage :

- Pics élevé dans la Patterson ( $\geq 10 \sigma$  en MIR ;  $5 \sigma$  en anomal)

Figures de mérite :

◇ Pouvoir de phasage (du dérivé) :

$$P = \langle F_H \rangle / \langle \Delta E \rangle$$

$$\begin{cases} P > 1,5 & \text{excellent} \\ P > 1 & \text{bon -} \\ P > 0,5 & \text{médiocre} \end{cases}$$

◇ R cullis

$$R_{cullis} = \frac{\sum \left| |F_{d,obs} \pm F_{p,obs}| - F_{a,calc} \right|}{\sum |F_{d,obs} - F_{p,obs}|}$$

*Normalement défini pour les réflexions centriques mais utile si étendu aux acentriques*

$$\begin{cases} R_{cullis} < 0,6 & \text{(centriques) excellent} \\ 0,6 < R_{cullis} < 0,9 & \text{( id° ) médiocre} \\ R_{cullis \text{ anomal}} < 1 & = \text{contribution} \\ & \text{anomale significative} \end{cases}$$

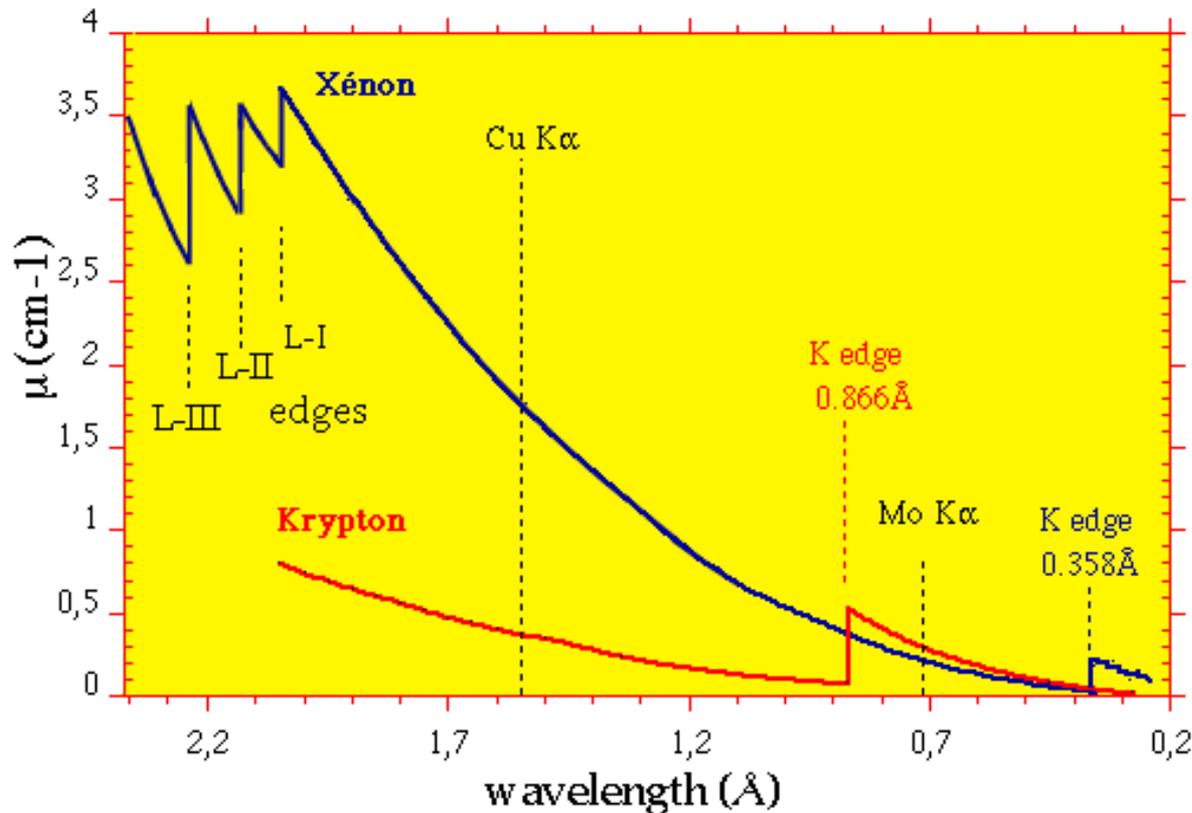
Ces indicateurs sont calculés par tranches de résolution

# La diffusion anormale

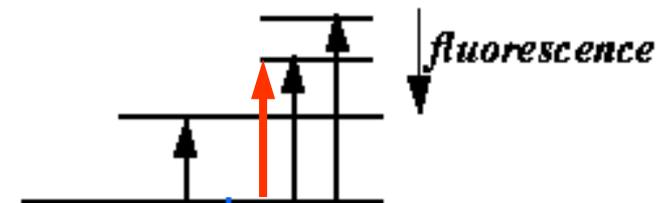
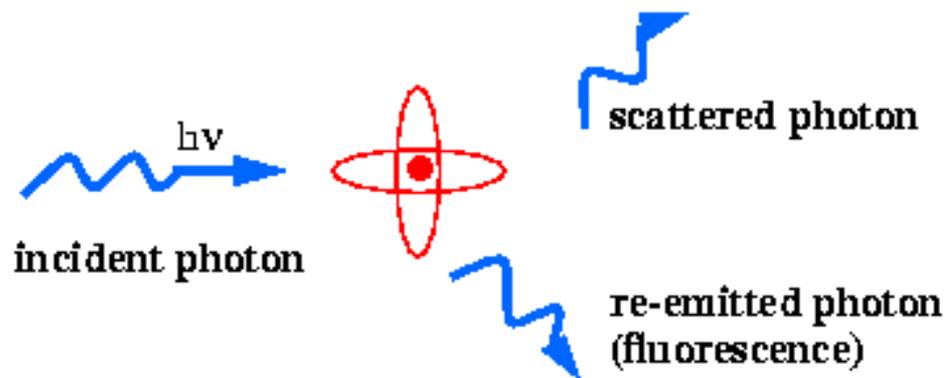
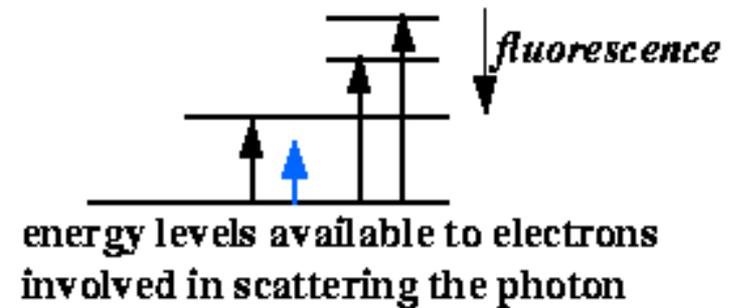
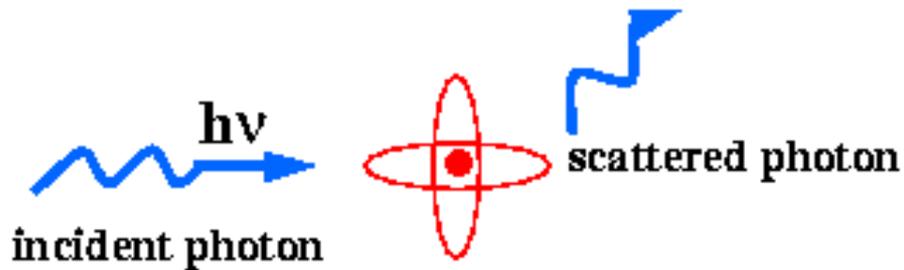
- Discontinuités d'absorption des rayons- X si  $\lambda$  près d'une transition électronique interne

transitions  $\Rightarrow$  *K, L or M*

- Coefficient d'absorption  $\mu$  pour Xe et Kr fonction de la longueur d'onde :



# Diffusion anomale

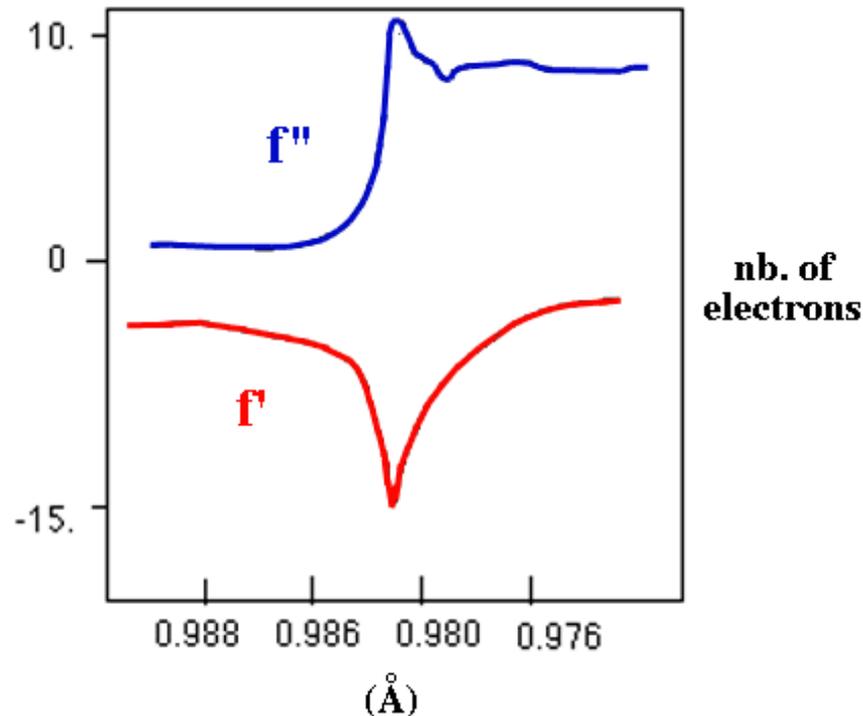


- Si  $\lambda$  est proche d'une **transition électronique**, le facteur de diffusion  $f$  de l'atome (lourd) n'est plus **constant** et égal à  $Z$ .

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

partie constante  
( indép. de  $\lambda$  )

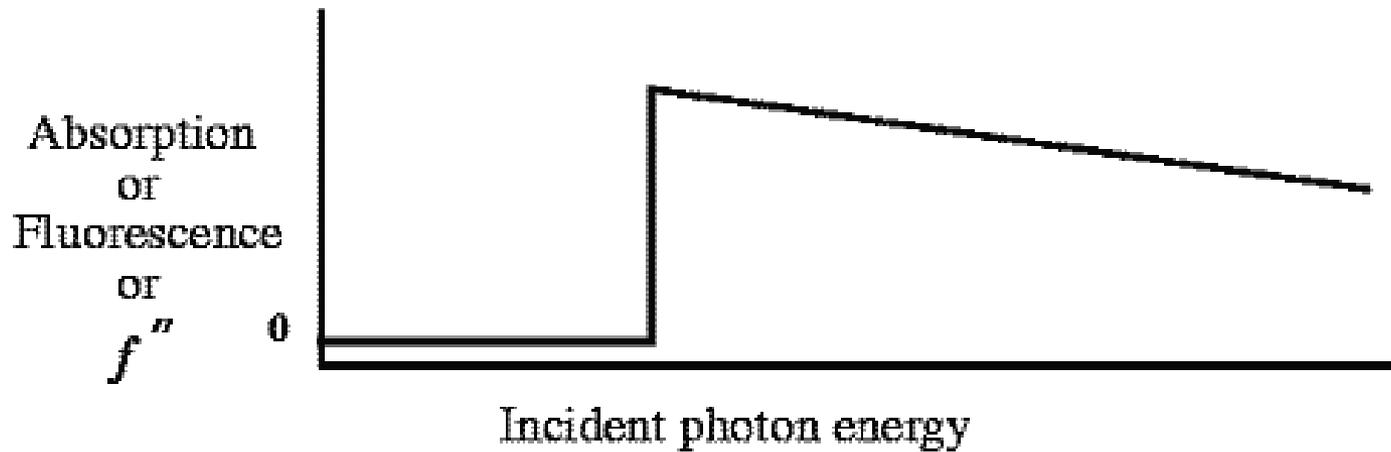
corrections **réelle et**  
**imaginaire**



(Ex : Se)

= comportement « anormal » des atomes lourds

# Relation de Kramers-Kronig



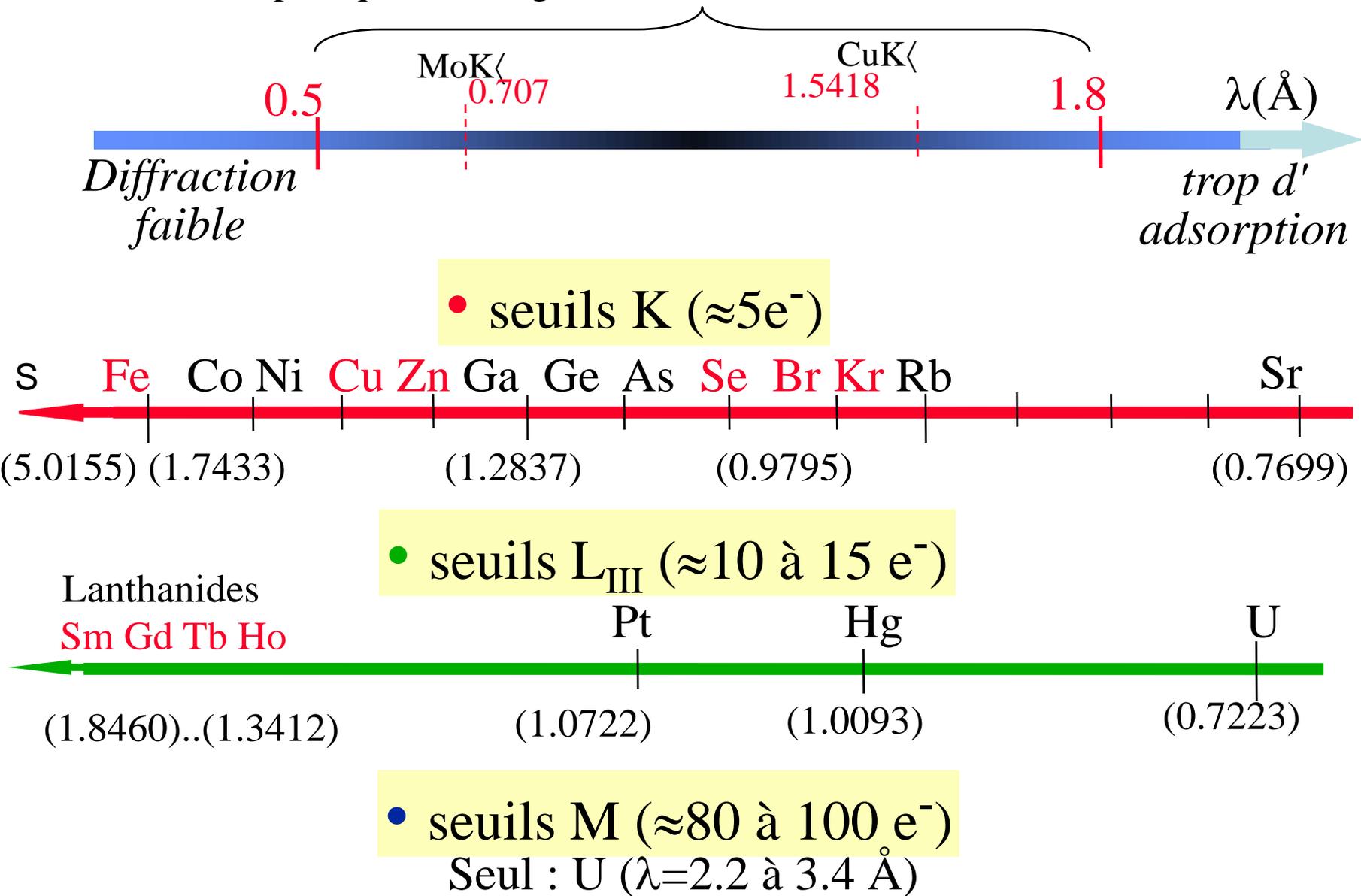
$$f'(\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' f''(\omega')}{\omega^2 - \omega'^2} \partial \omega'$$

Tables de Sasaki:

<http://www.tsurumi.yokohama-cu.ac.jp/xtal/Sasaki-table/>

# Atomes possédant un signal anomal utilisable

Domaine pratique des longueur d'onde en diffraction X



# Loi de Friedel

L'intensité de réflexion est identique d'un côté ou de l'autre du plan  $hkl$

$$I_{h,k,l} = I_{-h,-k,-l}$$

$$|\mathbf{F}_{hkl}| \sim \sqrt{I_{hkl}}$$

$$F_{hkl} = A + iB$$

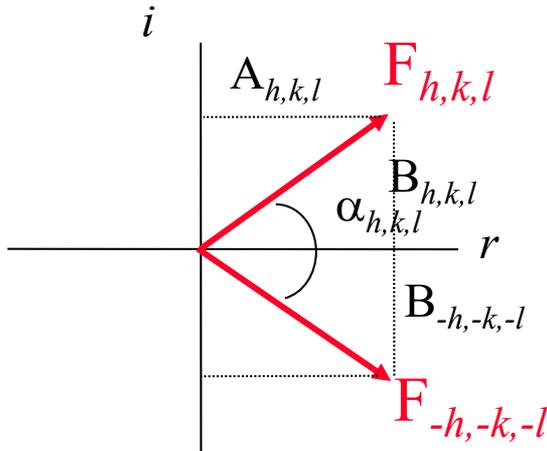
$$A_{h,k,l} = A_{-h,-k,-l} \quad \text{et} \quad B_{h,k,l} = -B_{-h,-k,-l}$$

Comme :  $I_{h,k,l} = A_{h,k,l}^2 + B_{h,k,l}^2$



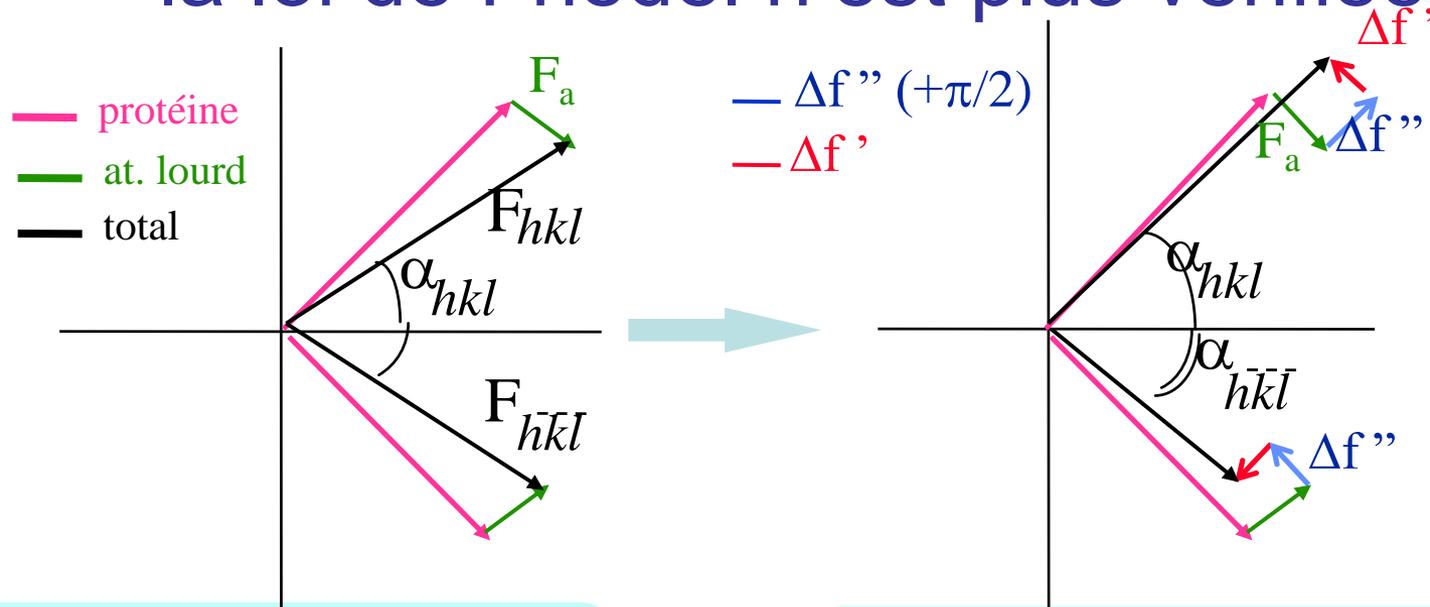
$$\alpha_{h,k,l} = -\alpha_{-h,-k,-l}$$

$$|F_{h,k,l}| = |F_{-h,-k,-l}|$$



→ Mesure de 1/2 sphère de diffraction suffirait

# Conséquence sur les facteurs de structure : la loi de Friedel n'est plus vérifiée



sans diffuseur

anomal :

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

$$\alpha_{hkl} = -\alpha_{\bar{h}\bar{k}\bar{l}}$$

avec diffuseur

anomal :

$$|F|_{hkl} \neq |F|_{\bar{h}\bar{k}\bar{l}}$$

$$\alpha_{hkl} \neq -\alpha_{\bar{h}\bar{k}\bar{l}}$$

$$\Delta F_{\pm hkl} = |F(\lambda)|_{hkl} - |F(\lambda)|_{\bar{h}\bar{k}\bar{l}}$$

$F_+$

$F_-$

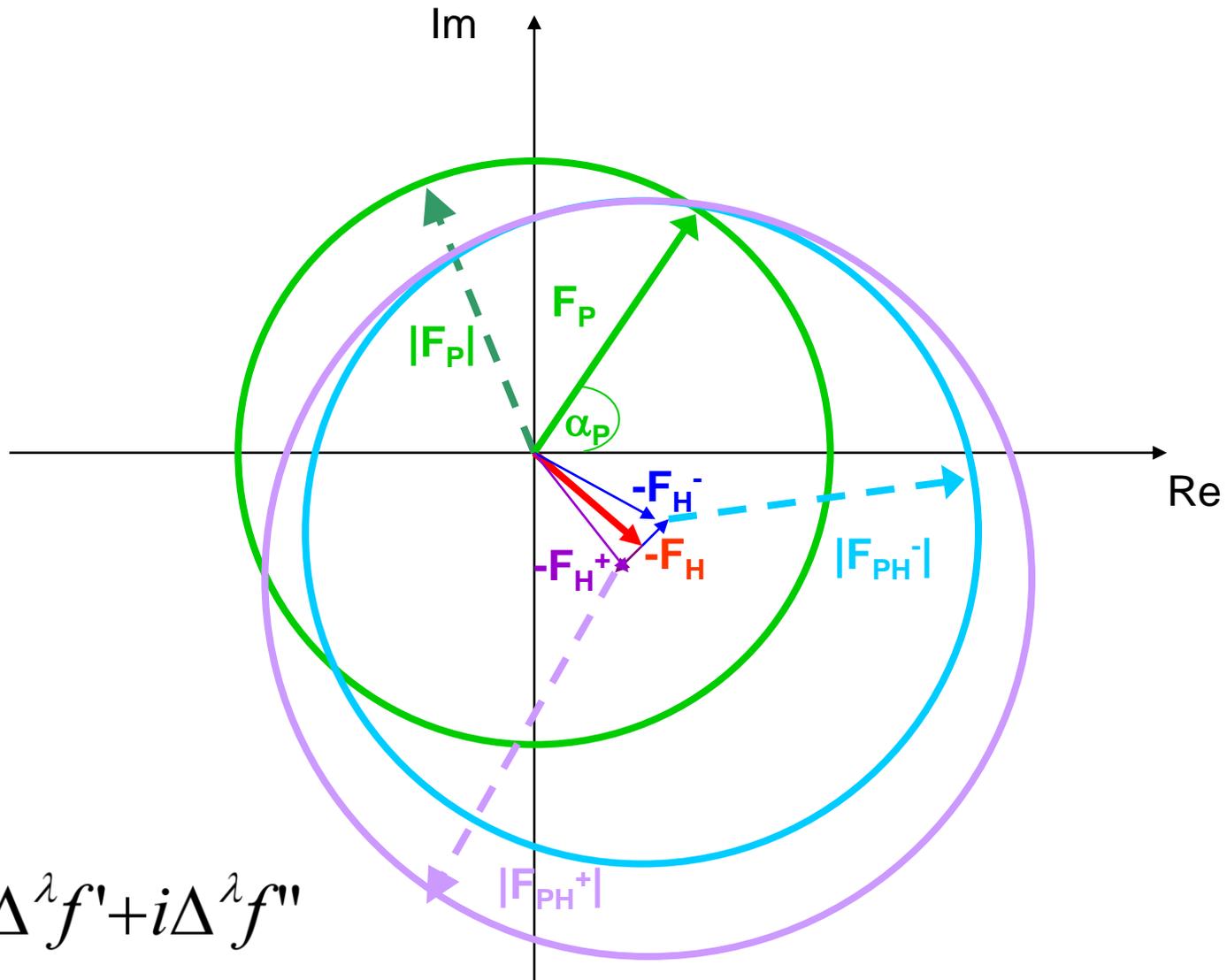
$\Delta F_{hkl} = \text{différence anormale}$

# Collecte des données anormales

---

- Mesurer avec précision les paires de Bijvoet ( $F_+$  et  $F_-$ , c.à.d  $F_{hkl}$  et  $F_{-h-k-l}$ )
- Inverse beam (rotation de  $180^\circ$  entre chaque image)  $\Rightarrow$  paires mesurées sur images consécutives
- Orienter le cristal  $\Rightarrow$  paires sur la même image

# SIRAS



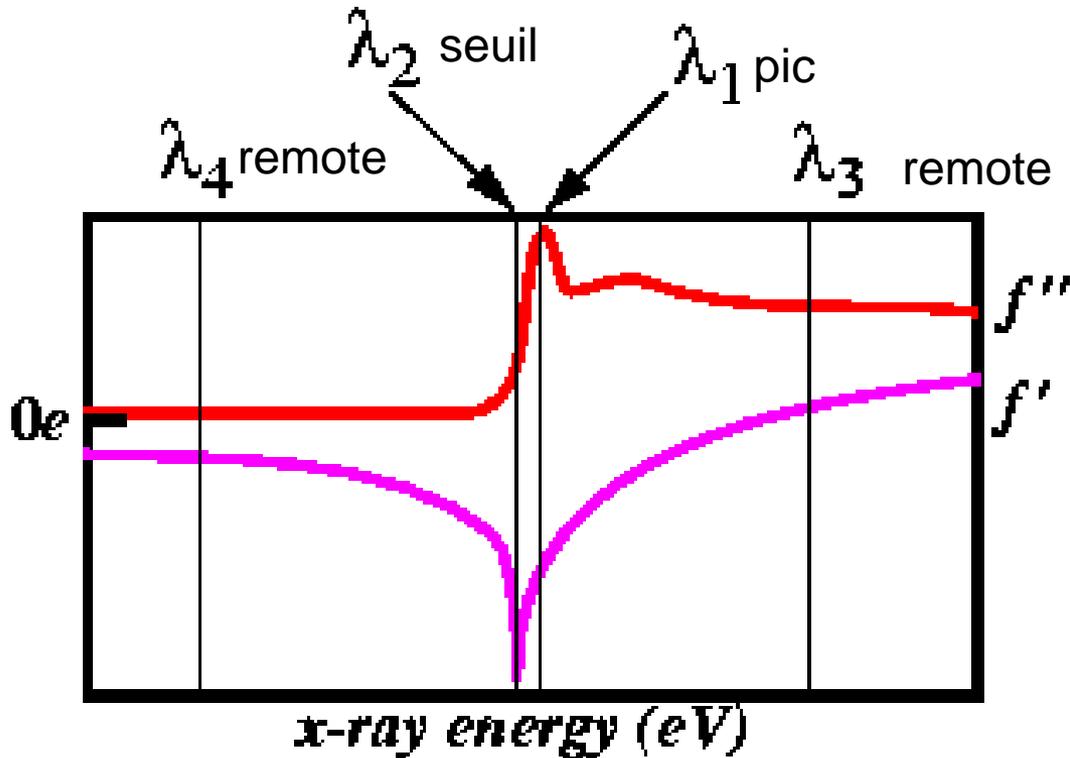
$${}^\lambda f = {}^0 f + \Delta {}^\lambda f' + i \Delta {}^\lambda f''$$

$$F_{hkl} \neq \overline{F_{hkl}}$$

Utilisation de  $F^+$  et  $F^-$  pour le phasage

# Méthode MAD

## (Multiwavelength Anomalous Diffraction)



Rapport anomal:

$$\frac{\langle \| \mathbf{F}_{\lambda_i}^+ \| - \| \mathbf{F}_{\lambda_i}^- \| \rangle}{\langle \| \mathbf{F}_T \| \rangle} = \sqrt{\frac{N}{2}} \frac{2f_{\lambda_1}''}{\langle \| \mathbf{F}_T \| \rangle}$$

Rapport dispersif

$$\frac{\langle \| F_{\lambda_i} \| - \| F_{\lambda_j} \| \rangle}{\langle \| F_T \| \rangle} = \sqrt{\frac{N}{2}} \frac{|f'_{\lambda_1} - f'_{\lambda_2}|}{\langle \| F_T \| \rangle}$$

## Le problème des phase a une solution analytique !

- Les termes dépendants et indépendants de la longueur d'onde peuvent être séparés (*Karle, Hendrickson*):

sous forme d'une fonction :  $G_T(\lambda) = F_T(\mathbf{h}) + D_A(\lambda)$

(  $\mathbf{h}$  : vecteur  $hkl$  )

*T pour Total*

*A pour Anomal*

*bleu =  $\lambda$  dépendant*

*noir = indépendant*

$$\begin{aligned} |G_T(\pm\mathbf{h},\lambda)|^2 &= |F_T(\mathbf{h})|^2 + a(\lambda) |F_A(\mathbf{h})|^2 \\ &+ b(\lambda) |F_T(\mathbf{h})| |F_A(\mathbf{h})| \cos(\phi_T - \phi_A) \\ &\pm c(\lambda) |F_T(\mathbf{h})| |F_A(\mathbf{h})| \sin(\phi_T - \phi_A) \end{aligned}$$

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

# Méthode MAD

$${}^{\lambda}\mathbf{F}_{hkl}^{+} = {}^0\mathbf{F}_T + \left( \frac{{}^{\lambda}f'}{{}^0f} + i \frac{{}^{\lambda}f''}{{}^0f} \right) {}^0\mathbf{F}_A$$

$${}^{\lambda}\mathbf{F}_{hkl}^{-} = {}^0\mathbf{F}_T + \left( \frac{{}^{\lambda}f'}{{}^0f} - i \frac{{}^{\lambda}f''}{{}^0f} \right) {}^0\mathbf{F}_A$$

$$I_{hkl}^{+} = \mathbf{F}^{+} \overline{\mathbf{F}^{+}}$$

$$I_{hkl}^{+} = |{}^0\mathbf{F}_T|^2 + {}^{\lambda}a |{}^0\mathbf{F}_A|^2 + {}^{\lambda}b |{}^0\mathbf{F}_T| |{}^0\mathbf{F}_A| \cos \Delta\phi \pm {}^{\lambda}c |{}^0\mathbf{F}_T| |{}^0\mathbf{F}_A| \sin \Delta\phi$$

déterminés  
à 3 long. d'ondes  
différentes

$${}^{\lambda}a = \frac{{}^{\lambda}f'^2 + {}^{\lambda}f''^2}{{}^0f^2} \quad {}^{\lambda}b = \frac{2{}^{\lambda}f'}{{}^0f} \quad {}^{\lambda}c = \frac{2{}^{\lambda}f''}{{}^0f} \quad \Delta\phi = {}^0\phi_T - {}^0\phi_A$$

$$I_{hkl}^{+} = x + {}^{\lambda}ay + {}^{\lambda}bz + {}^{\lambda}ct \quad \text{avec} \quad z^2 + t^2 = xy \quad \textit{inconnus} \quad (\Delta f' \text{ et } \Delta f'' \text{ connus})$$

Pratiquement : ceci est un jeu de 4 équations linéaires à 4 inconnues.  
x,y,z,t peuvent être calculés, puis  $|F_T|$ ,  $|F_A|$  et finalement  $\Delta\phi$ .

# Méthode SAD

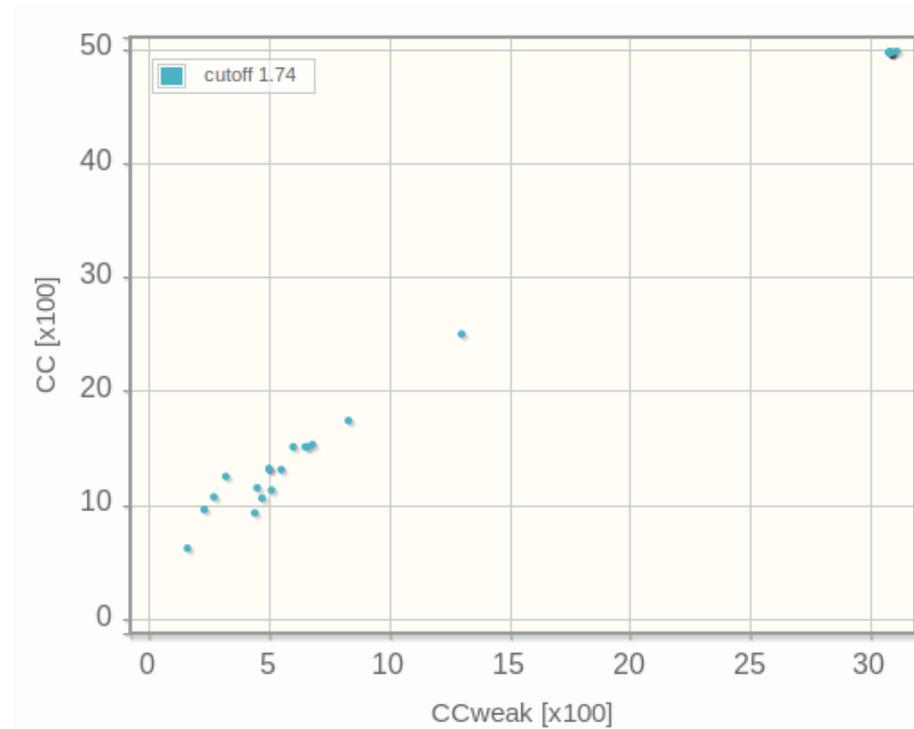
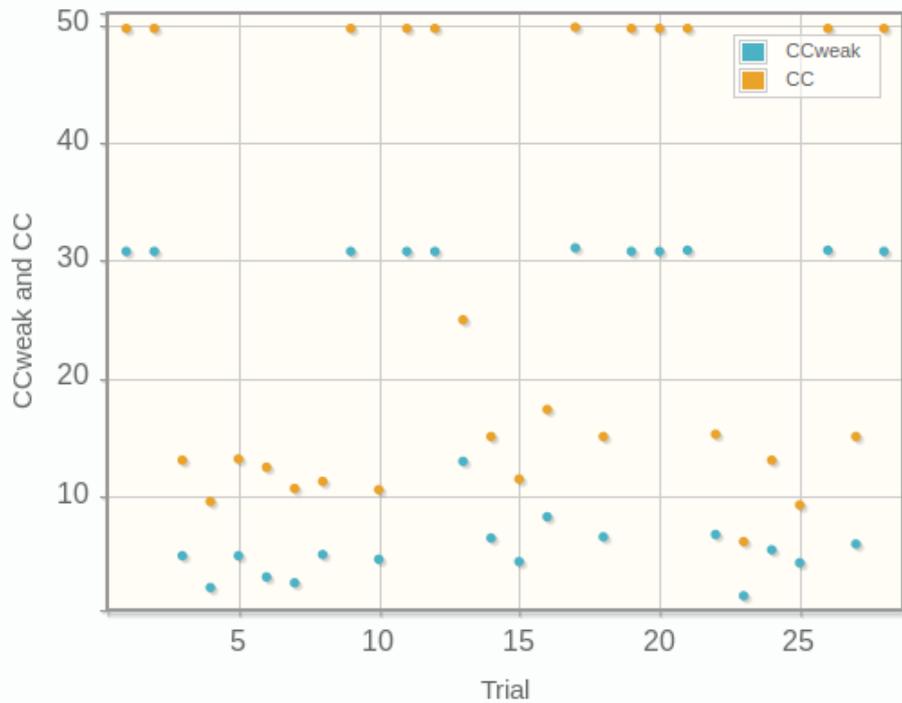
- Un seul enregistrement au pic
- Estimation des  $F_A$

$$|F_+| - |F_-| = c |F_A| \sin \phi$$

Les différences  $||F_+| - |F_-||$  les plus grandes sont sélectionnées avec  $\phi \sim 90^\circ$  quand  $I_+ \gg I_-$  et  $\phi \sim 270^\circ$  quand  $I_+ \ll I_-$ .

- Localisation des diffuseurs anomaux
- Phasage ( $\sim$ SIR)
- Nivellement de solvant / choix de la main

# Localisation des diffuseurs



Trial: essai-erreur pour localiser les diffuseurs  
CC entre  $F_{Acalc}$  et  $F_{Aobs}$  avec  $|+-|-$  les plus grandes

Maximiser CC et CCweak

CC weak calculés avec tous les  $|+-|-$

# En résumé

## MIR

Trois enregistrements :

- Une native
- 2 dérivés lourds
- longueur d'onde fixe

-les atomes lourds doivent être “lourds”

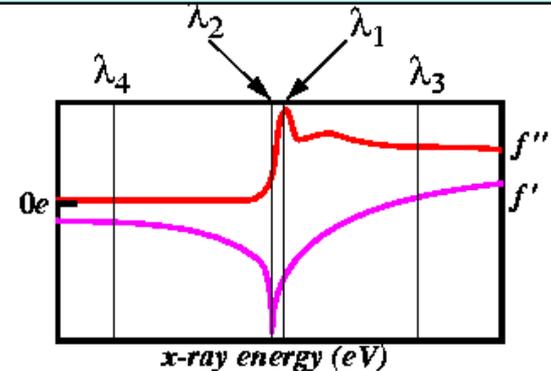
Ex : Pt, Hg, U

-Sources de RX:  
Générateurs de  
laboratoire

## SAD / MAD

Un à trois enregistrements :

- Un seul cristal  
 (“native” incluant un  
 diffuseur anomal)
- 1 à 3 longueur d'ondes



- Utile si metallo-enzymes (Fe, Mn, Cu, Zn...)  
- ou protéines modifiées: Se-Met, Se-Cys  
- Source de RX très stable dont la  $\lambda$  peut  
être ajustée: synchrotron

# Bioincorporation de la selenométhionine

- Ne change pas la structure de la protéine
- Conditions de cristallisation sont proches de celles de la protéine native (en général)
- Incorporation de SeMet proche de 100%
- Nombre de Met modéré (2,4% des acides aminés)
- Phasage efficace

# SAD sur le soufre

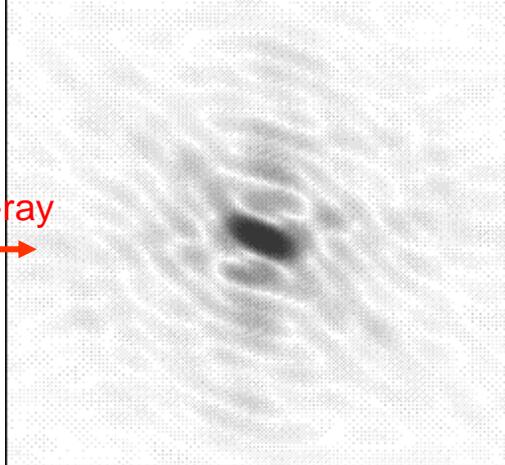
- Abondance naturelle
- Faible signal anomal
- Grande longueur d'onde

# Remplacement moléculaire

cristal



X-ray  
→

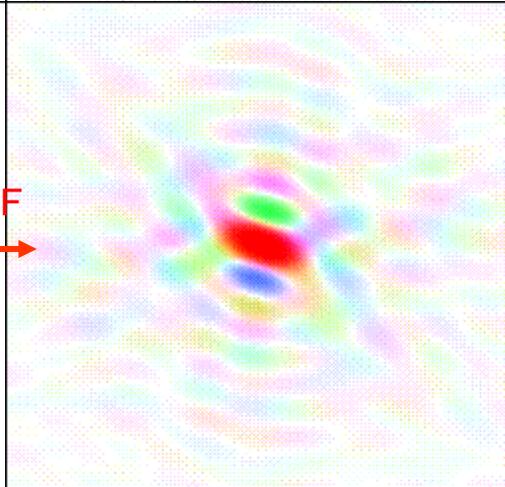


$F_{obs}$  experimental

modèle PDB  
atomes  $x_i, y_i, z_i$



TF  
→

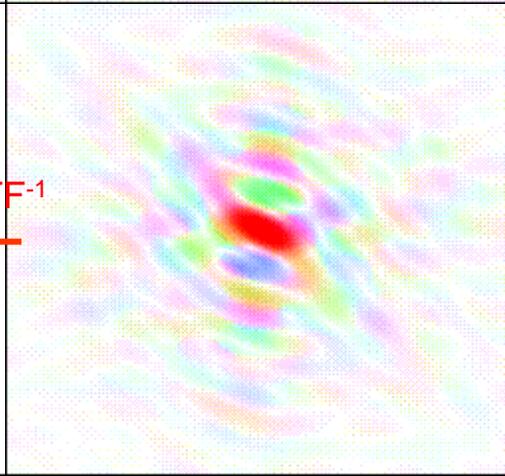


$F_{calc}, \phi_{calc}$  Calculé

Carte de densité  
électronique



TF<sup>-1</sup>  
←



Calcul de  $TF^{-1}(F_{obs}, \phi_{calc})$

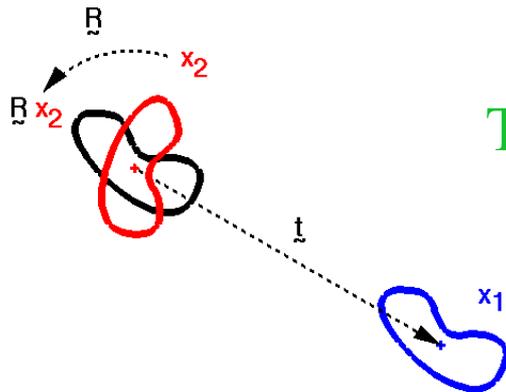
Validation avec  $R_{factor} (<50\%)$

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

# Remplacement moléculaire

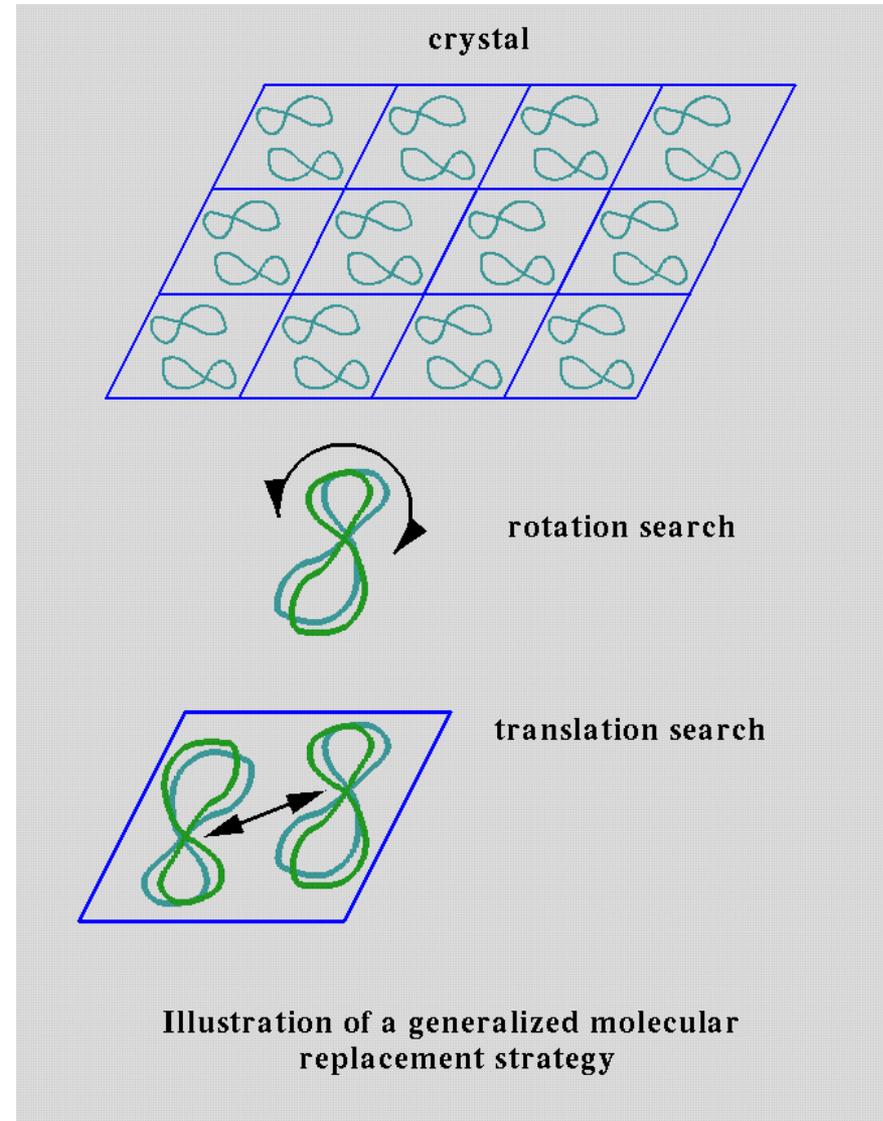
- Structure homologue connue (modèle):
- Superposer le modèle à la cible dans le cristal
- Calcul des phases à partir du modèle initial

Rotation



Translation

$$x_1 = B x_2 + t$$



# Problème

- Il faut chercher l'orientation (rotation) et la position (translation) du modèle dans le cristal
- Recherche à six dimensions
  - 6 rotations
  - 3 translations

$$x' = c_{11}x + c_{21}y + c_{31}z + v_x$$

$$y' = c_{12}x + c_{22}y + c_{32}z + v_y$$

$$z' = c_{13}x + c_{23}y + c_{33}z + v_z$$

- Impossible car nécessite trop de temps de calcul (tous les 5 deg en rotation et qq Å en translation)

# Exemple

- 6 paramètres à déterminer ( $\alpha, \beta, \gamma, t_x, t_y, t_z$ )
- N molécules dans l'unité asymétrique
- 6N paramètres à déterminer

- Rotation

0-360°, 0-180°, 0-360° par pas de 2,5°  $N_{\text{rotation}} = 1,5 \cdot 10^6$

- Translation

100 x 100 x 100Å par pas de 1Å  $N_{\text{translation}} = 10^6$

- Recherche exhaustive à 6 dimensions

$\Rightarrow N_{\text{search}} = 1,5 \cdot 10^{12}$

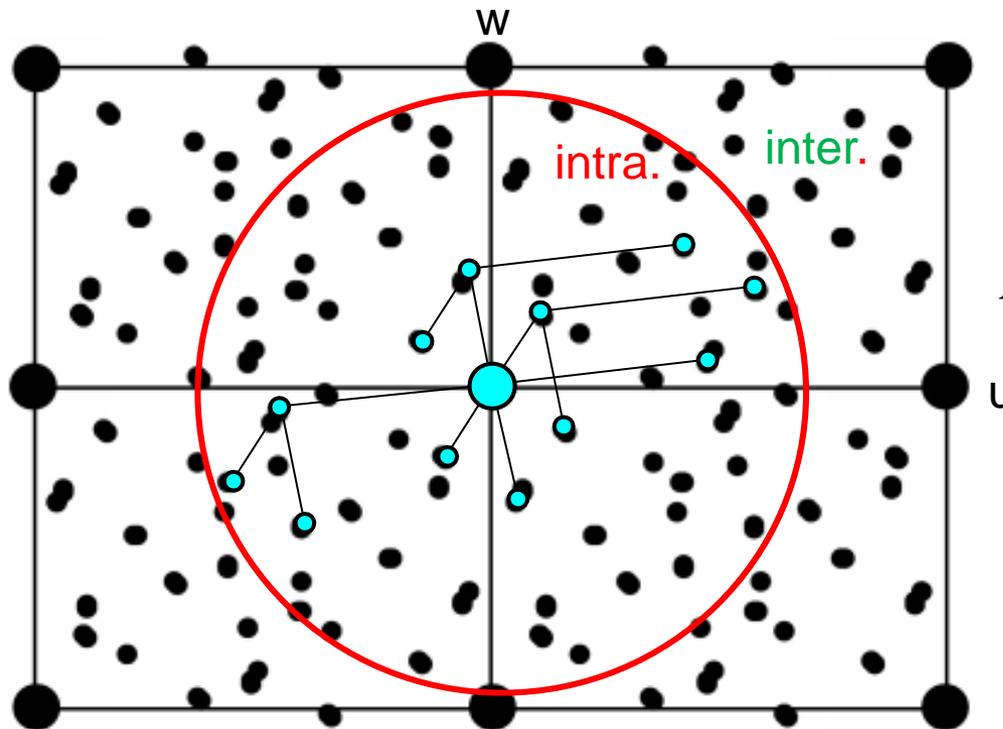
- Recherche séparée (translation recherchée uniquement pour les meilleures orientations)

$\Rightarrow N_{\text{search}} = 1,5 \cdot 10^6 + 10^6 = 2,5 \cdot 10^6$

# I. Fonction de rotation

Superposition de  $P_A$  native et  $P_B$  model dans une région de rayon R

Utilisation des vecteurs **intra.** Exclusion des vecteurs **inter.**



$$R([\mathbf{R}]) = \int P_A(u) P_B([\mathbf{C}]u) du$$

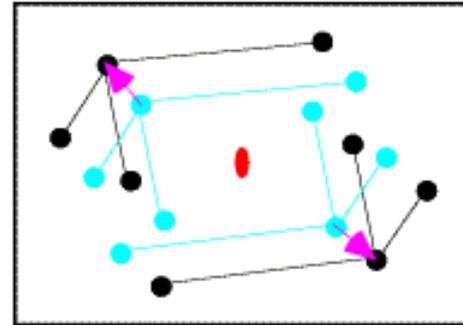
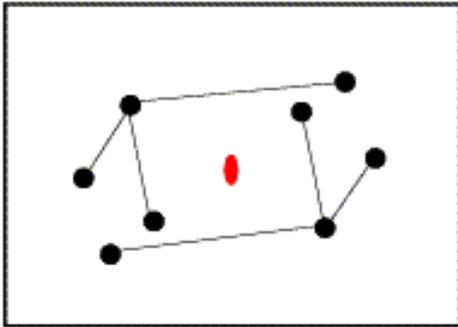
$U$  native      model

↑  
TF<sup>-1</sup>(I(hkl))

Recherche du maximum de la fonction de rotation (corrélation maximale entre les deux cartes de Patterson)

## II. Fonction de translation

Superposition entre les vecteurs intermoléculaires observés et les vecteurs intermoléculaires calculés en déplaçant le modèle orienté selon les axes u, v, w



$$T(\mathbf{t}) = \int_V P_{1,2}(\mathbf{u}, \mathbf{t}) \times P(\mathbf{u}) d\mathbf{u}$$

calculated Patterson observed Patterson

La solution correspond au maximum de la fonction de translation

Programme: MOLREP

Pipelines: BALBES, MrBUMP...

# Fonction de self-rotation

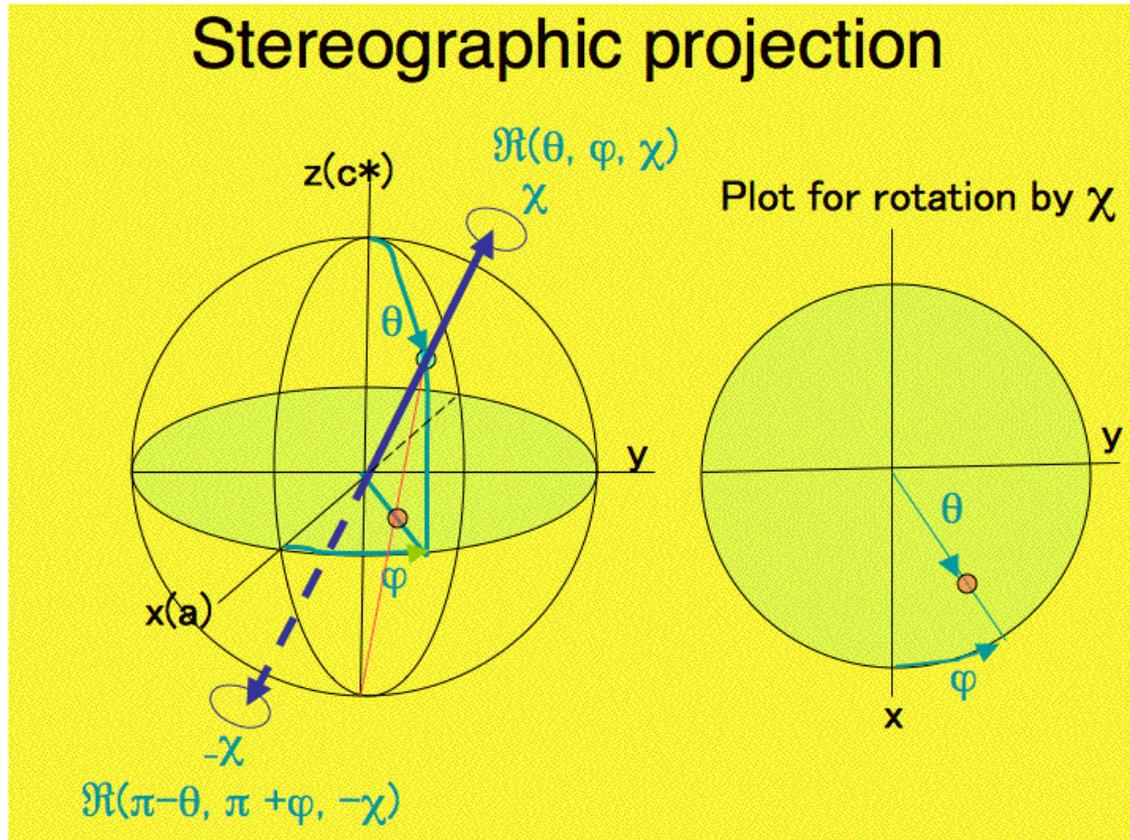
---

- La Patterson expérimentale est tournée et superposée *sur elle-même*
- Elle permet de détecter la présence de symétrie non-cristallographique (NCS) reliant plusieurs protéines dans l'unité asymétrique
- Devrait toujours être calculée, avant toute technique de phasage: la NCS permet également d'améliorer la qualité des cartes de densité électronique

# Symétries non cristallographiques

- Quand il y a plusieurs molécules dans l'unité asymétrique...
- Opération de symétrie reliant des molécules dans l'unité asymétriques
- $\neq$  symétries cristallographiques
- les axes peuvent être parallèles mais les positions sont différentes

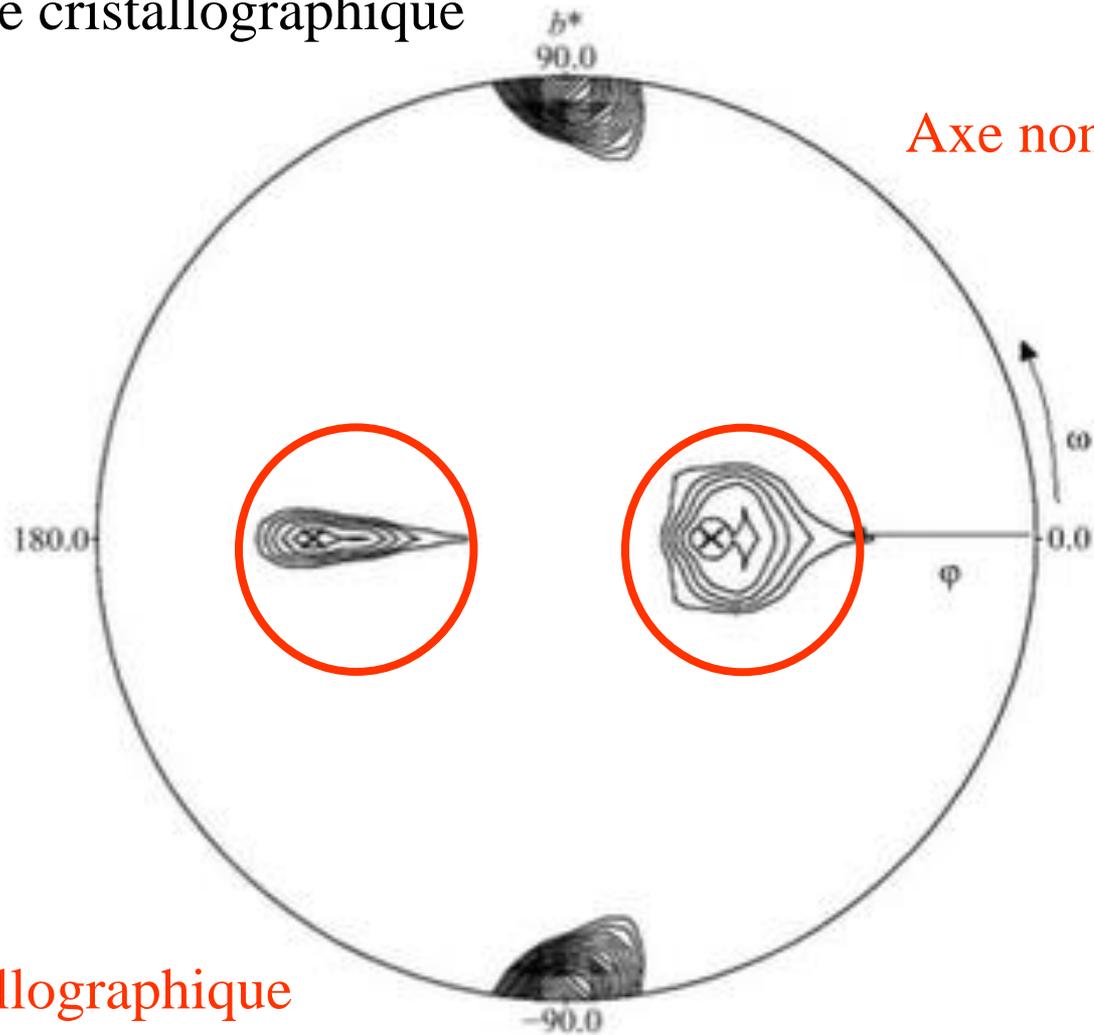
# Projection stereographic



projections pour différents  $\chi$

$$\kappa=180^\circ$$

Axe cristallographique



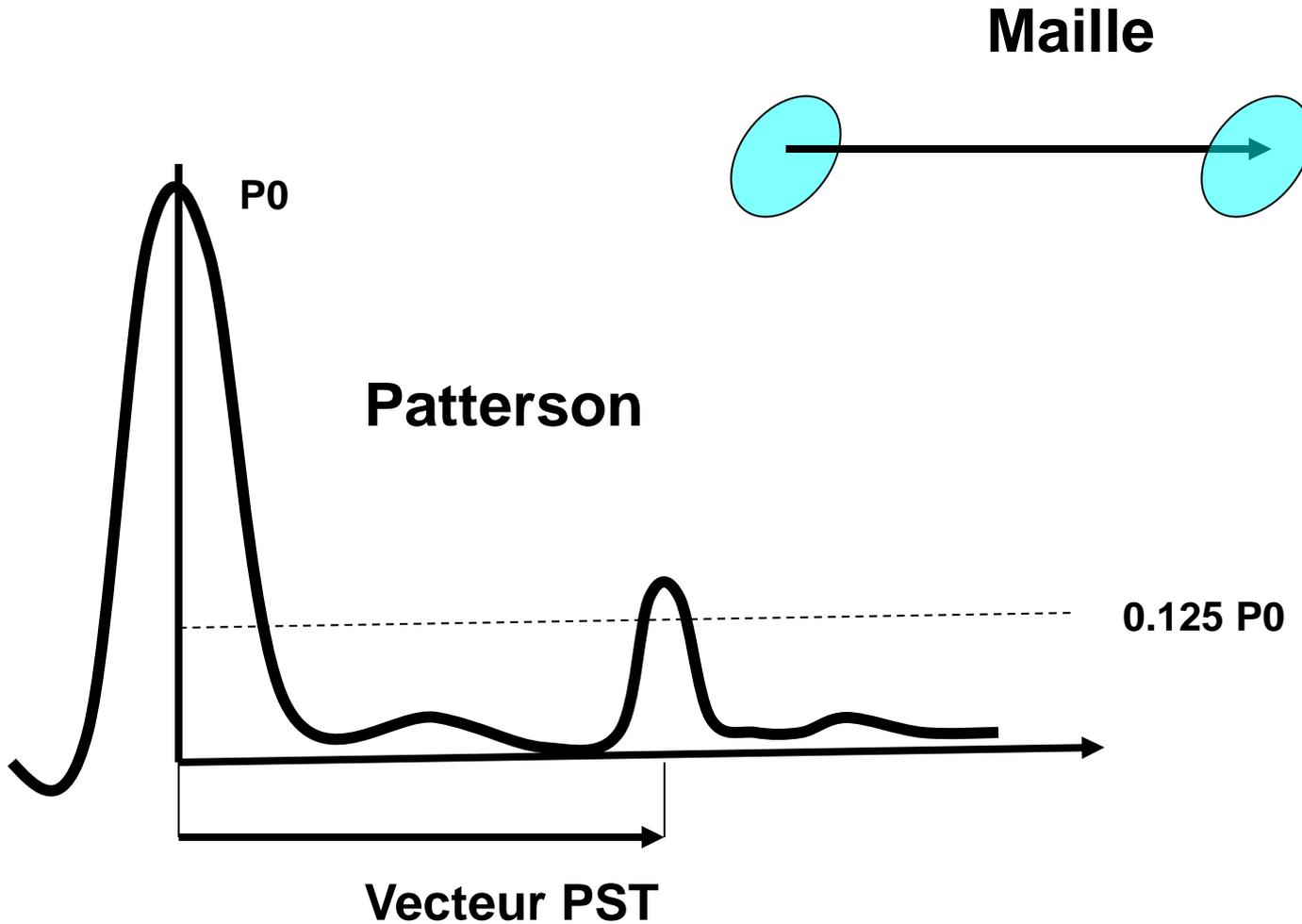
Axe non cristallographique

Axe non cristallographique

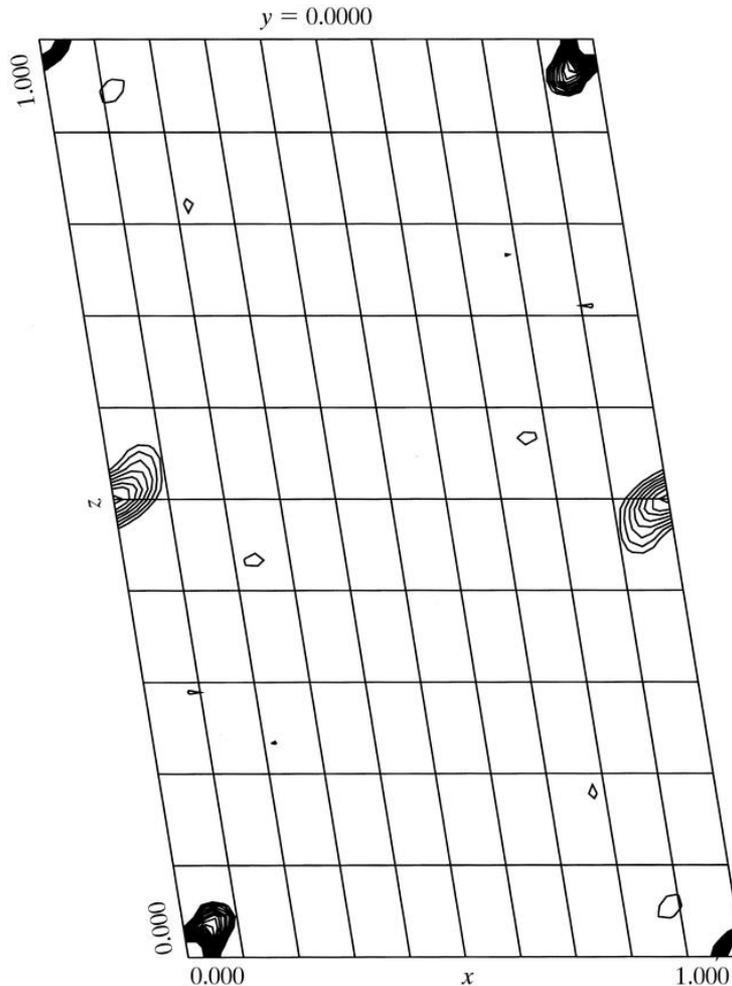
Axe cristallographique

C2? 2 molécules dans l'unité asymétrique

# Pseudo-translation



# Pseudo-translation



Patterson native

⇒ Rend le remplacement moléculaire difficile

P21,  $a = 66.218$ ,  $b = 46.113$ ,  $c = 112.079$  Å,  $\beta = 99.30^\circ$

# Probabilité



- 1 dé à 4 faces, 1 dé à 6 faces, 1 dé à 8 faces, 1 dé à 10 faces dans une boîte
- Un dé tiré au hasard
- Probabilité de deviner le bon dé:

$$1/4$$

# Maximum de vraisemblance



- 1 dé à 4 faces, 1 dé à 6 faces, 1 dé à 8 faces, 1 dé à 10 faces dans une boîte
- Un dé tiré au hasard et jeté. On note le résultat.
- Probabilité de deviner le bon dé
  
- La probabilité d'obtenir les données observées (le résultat du dé) étant donné le modèle (le nombre de faces sur le dé) est la vraisemblance
  - Si on obtient 10  $P(10;10)= 1/10$ ,  $P(10;8)=0...$
  - Si on obtient 7  $P(7;10)=1/10$ ,  $P(7;8)=1/8...$
  - Si on obtient 1  $P(1;4)=1/4$ ,  $P(1;6)=1/6...$
  - Le dé le plus probable est celui ayant la plus grande vraisemblance de générer le résultat: c'est le principe du maximum de vraisemblance

Le quotient entre deux vraisemblances est une mesure de la certitude

$$\frac{P(7;8)}{P(7;10)} = \frac{\frac{1}{8}}{\frac{1}{10}} = 1.25$$

# Maximum de vraisemblance

- Principe: quel est le modèle qui rend compte des données avec la plus grande probabilité
- On choisit la solution avec la plus grande probabilité
- Théorème de Bayes

$$P(model; data) = \frac{P(model)}{P(data)} \times P(data; model)$$

$$P(model; data) = P(model) \times P(data; model)$$

$P(data)$  est constante

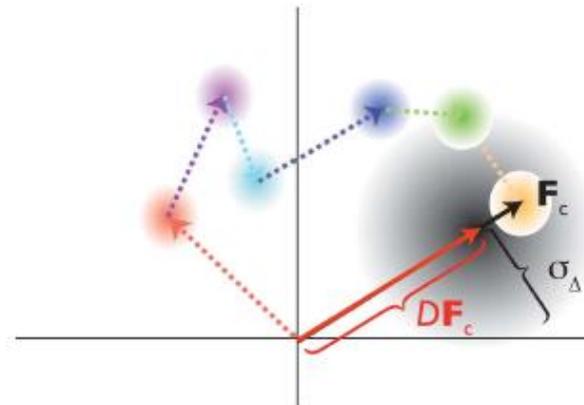
$P(model; data)$  est la probabilité a posteriori

$P(data; model)$  est la vraisemblance

Utilisation de ce qui est connu sur la structure (stéréochimie, % identité, rmsd...)

# Maximum de vraisemblance

- Pour chaque orientation et position du modèle  $(R,t)$ , quelle est la probabilité d'obtenir l'amplitude du facteur de structure observé?
- Erreur provenant du modèle et de la position
- Erreur totale est une gaussienne à deux dimensions centrée sur  $DF_c$  et de variance  $\sigma_{\Delta}^2$



$$0 < D < 1$$

Probabilité d'observer un  $F_o$  particulier:  $P(F_o|F_c)$

# Avantage du maximum de vraisemblance sur la Patterson

---

- Modélisation des erreurs expérimentales ( $\sigma_F$ ) et du modèle (r.m.s)
- Fonction de rotation de Patterson est une approximation de la fonction de vraisemblance de rotation (1<sup>er</sup> terme du développement en série de Taylor)
- Les fragments correctement placés augmentent le signal de la fonction de vraisemblance de rotation

# Autres méthodes

---

- RIP (utilisation des dommages dues aux radiations)
- ...
- ARCIMBOLDO (utilisation d'hélices pour le remplacement moléculaire)
- Utilisation de modèles *ab initio* générés par Rosetta (AMPLE)
- Utilisation de la dynamique moléculaire / modes normaux... (xMDFF...)

# Amélioration des phases par nivellement de solvant

Calcul d'un masque où la zone de protéine vaut 1 et le solvant 0

```
00000000000000000000
000111110000011000
001111111111111100
00111111111111000
00001111111110000
00000000000000000000
```

masque

```
11212321112212112
21245475122256232
12367867784567631
32867559877685432
12325677987754322
12234221221132141
```

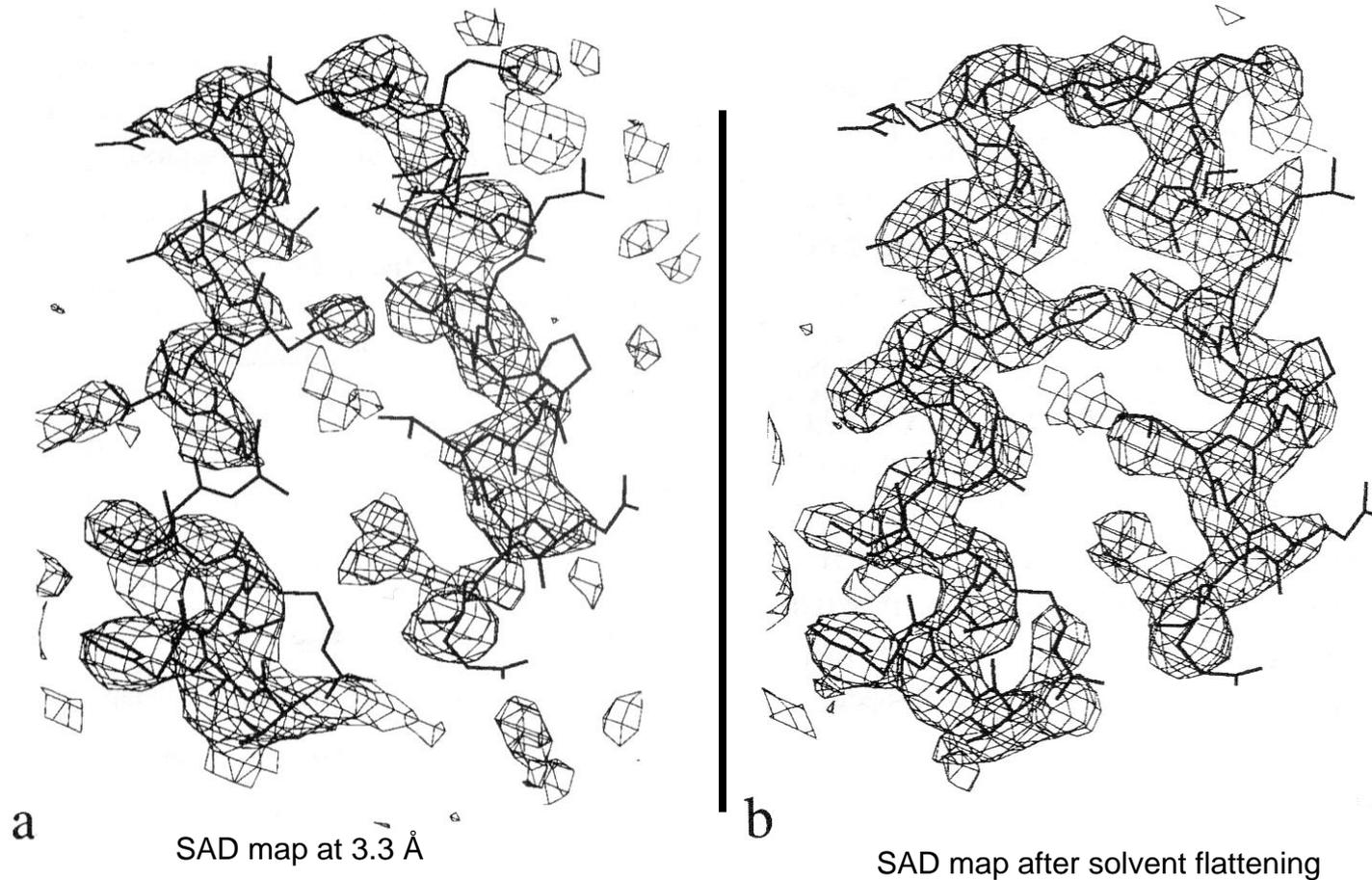
carte de densité électronique

Détermination de la densité moyenne dans la région de solvant ( $\rho_s = 2$ ).

```
2222222222222222222
222454752222256222
22367867784567622
22867559877685222
22225677987752222
2222222222222222222
```

Programmes:  
DM  
Solomon  
Parrot...

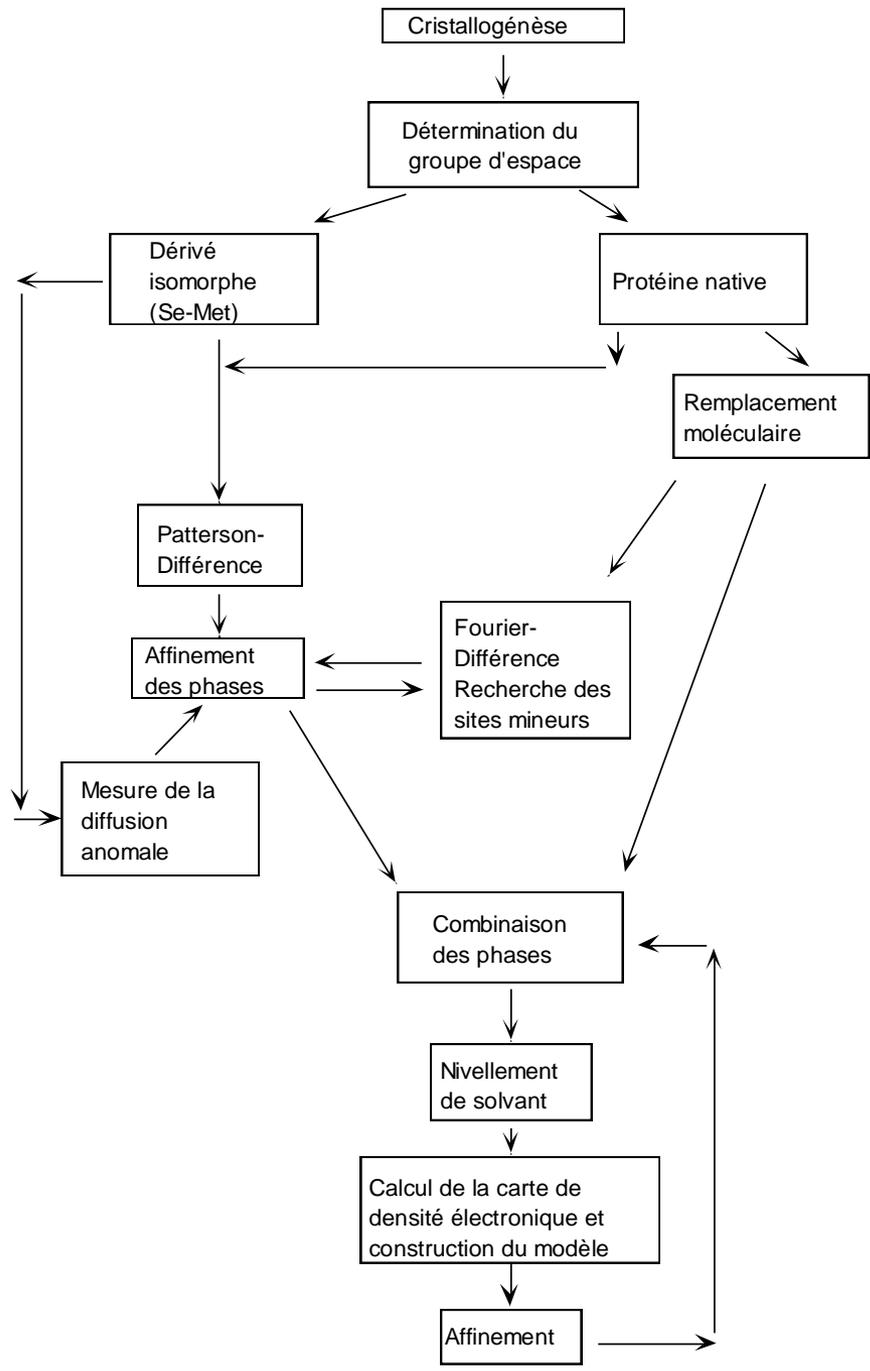
De meilleurs phases sont obtenues par calcul des F par transformée de Fourier .  
Plusieurs cycles de nivellement.



## Utilisation de la moyennation des symétries non cristallographiques (NCS)

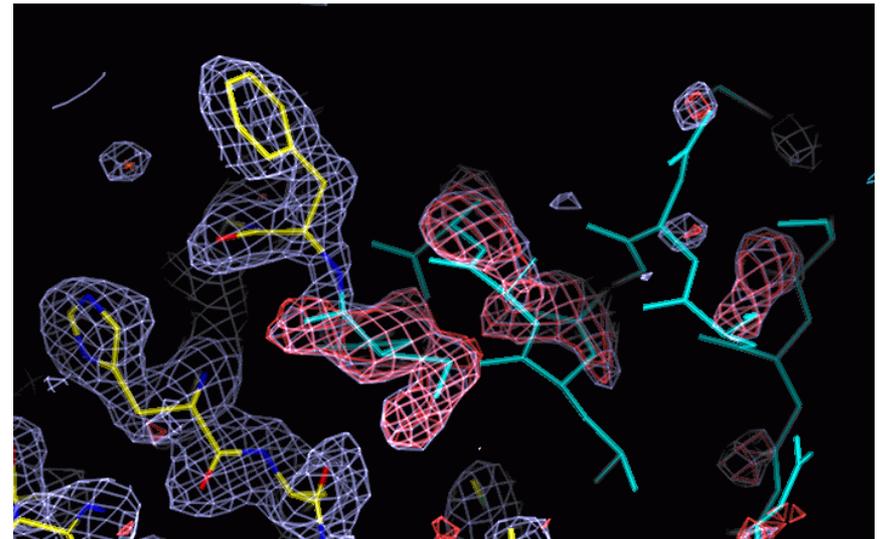
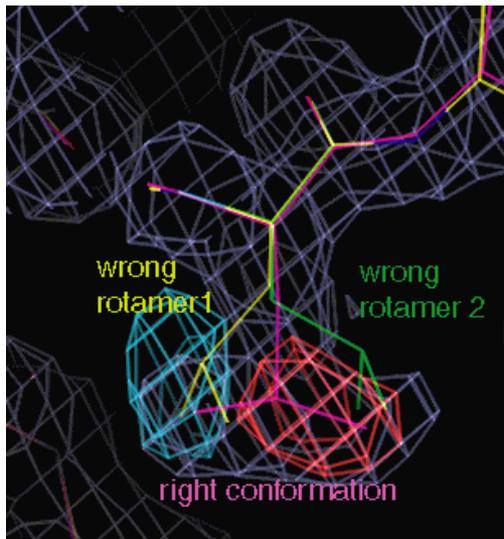
La densité d'une protéine est considérée comme identique dans les régions reliées par NCS. Un masque est appliqué sur ces régions. Les densités sont moyennées dans les masques. La technique est efficace pour des structures ayant beaucoup de NCS (virus icosaédrique, SIR+NCS averaging)

$$\rho_{avg}(x) = \sum_{i=1}^N \frac{1}{N} M_i(x) \sum_{j=1}^N \rho(x_{ij})$$



# Interprétation de la densité électronique

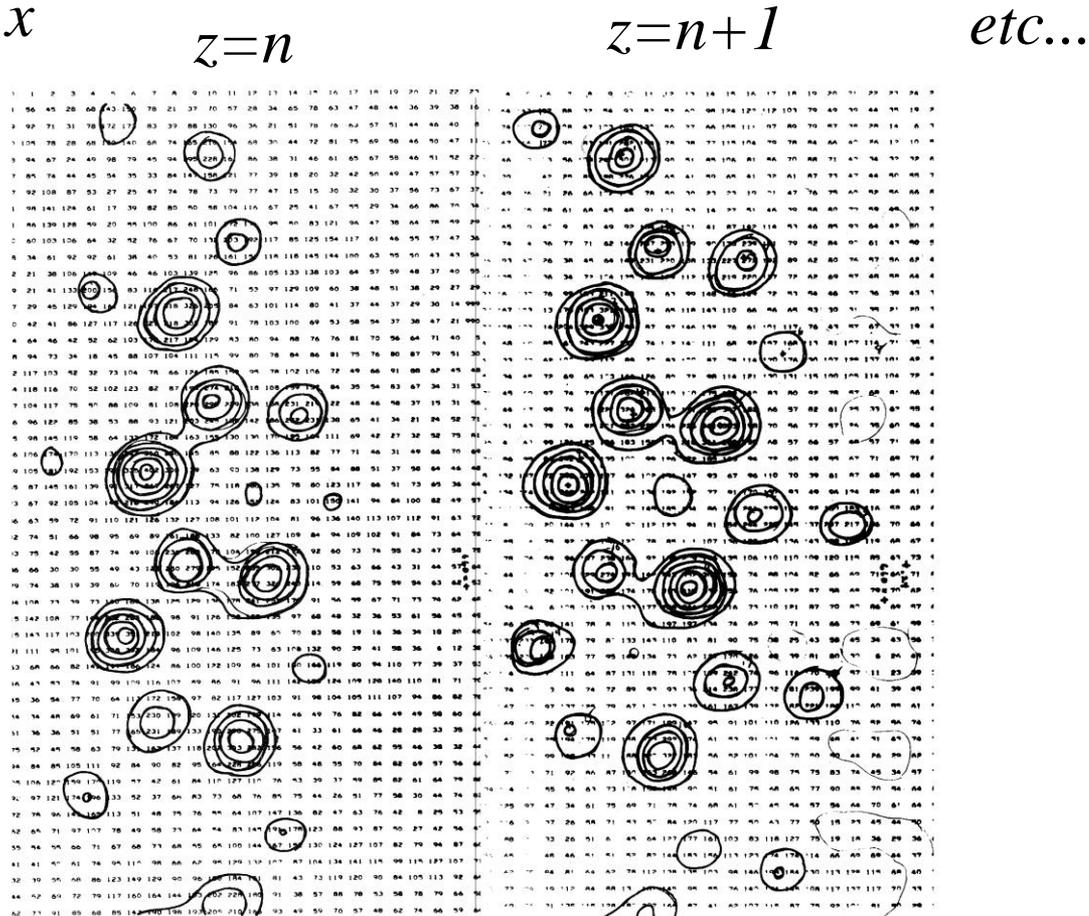
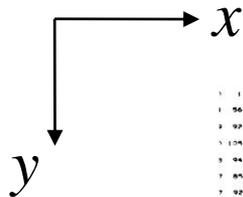
- Limiter les biais introduits par le modèle
  - Omit map
  - Carte de Fourier différence
    - $F_{\text{obs}} - F_{\text{calc}}$   
( $mF_{\text{obs}} - DF_{\text{calc}}$ )
    - $2F_{\text{obs}} - F_{\text{calc}}$   
(ou  $2mF_{\text{obs}} - DF_{\text{calc}}$ , m figure of merit et D dérivé de pondération par SigmaA)



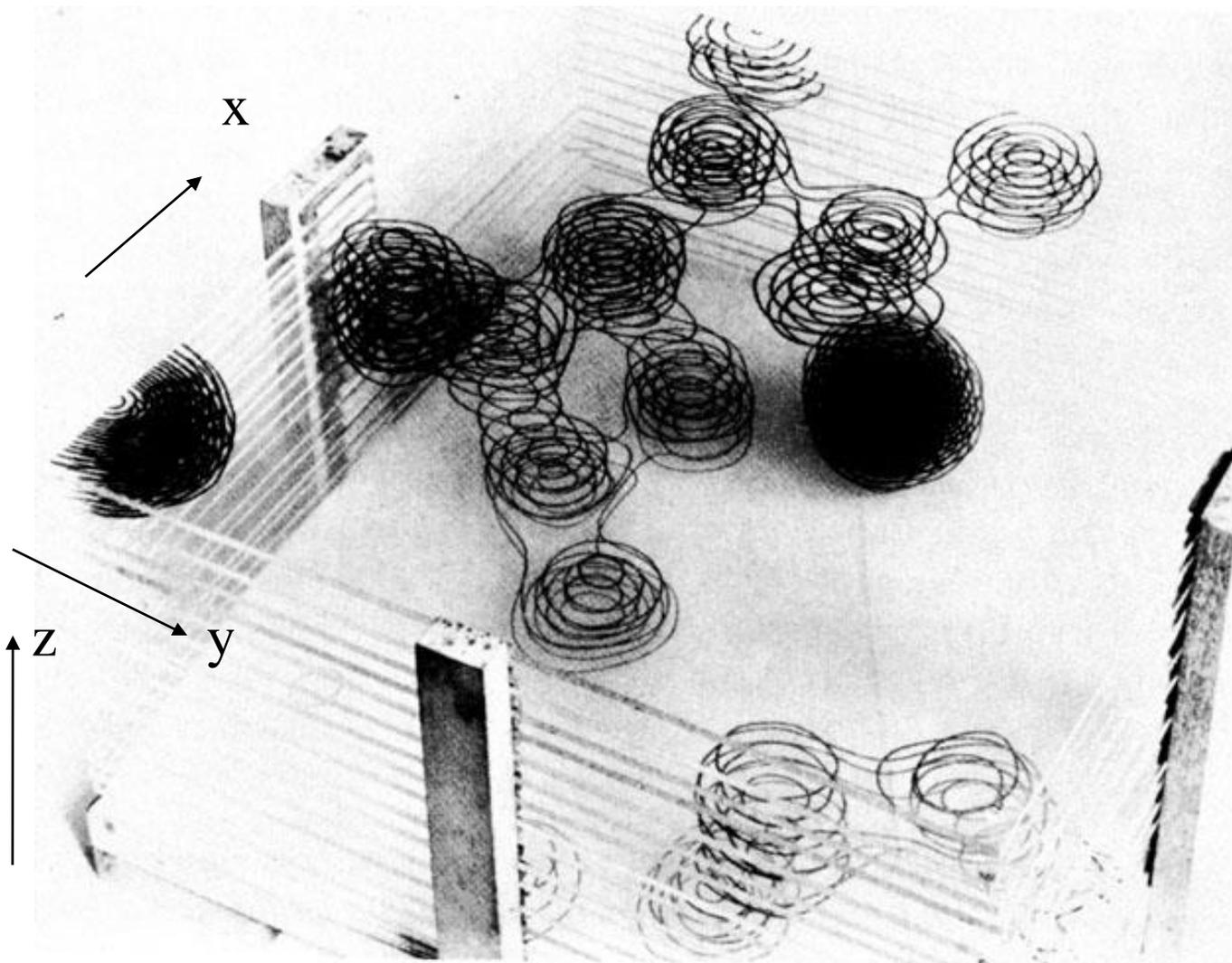
# Application et visualisation des densités électroniques

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

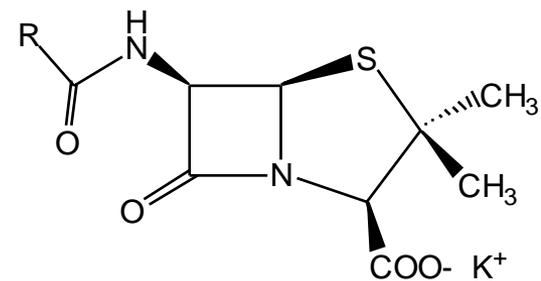
En 2D :



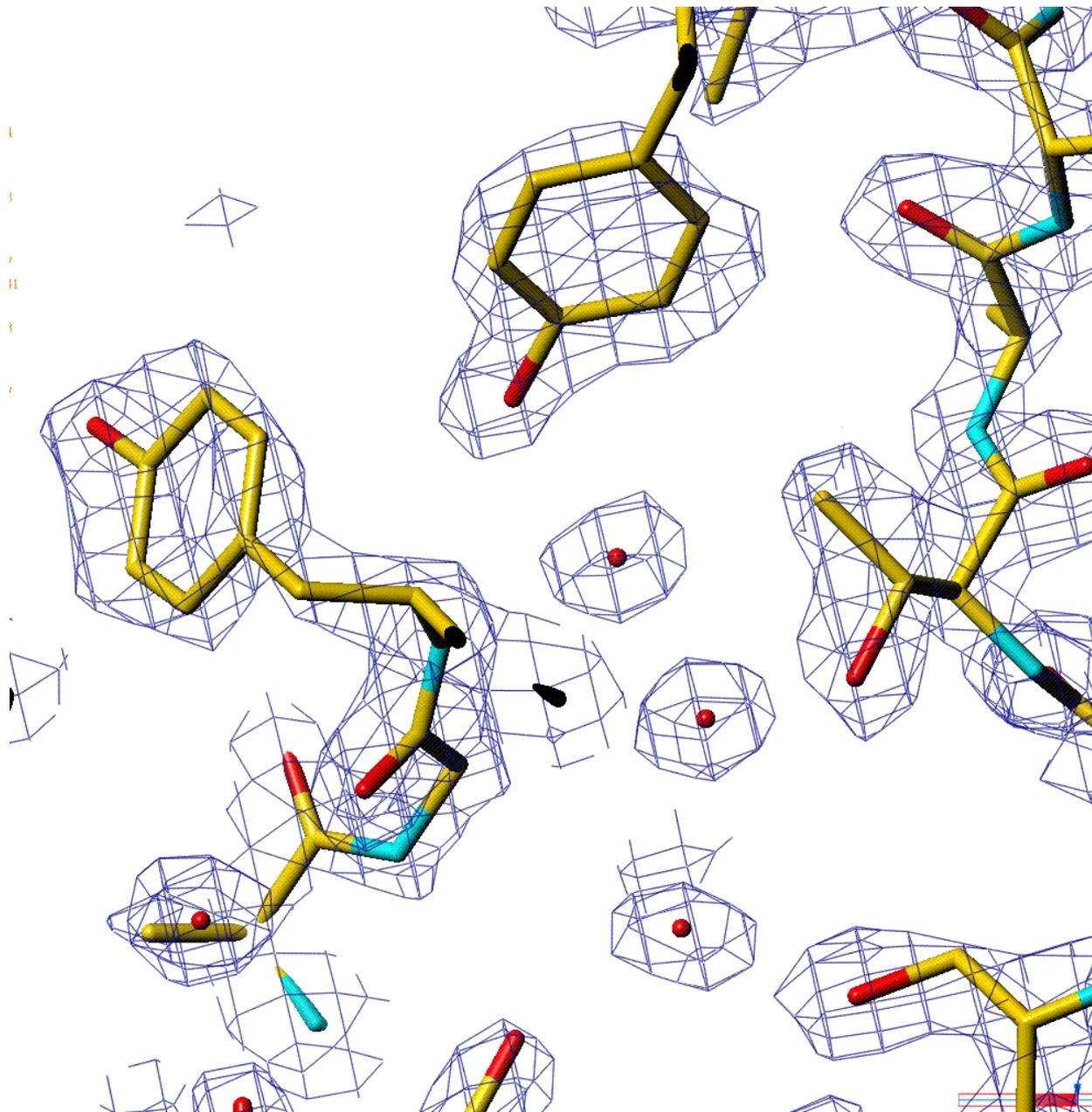
*En 3D :*      Structure de la pénicilline (sel de K<sup>+</sup>)



*Dorothy Crowfoot-Hodgkin, 1949*

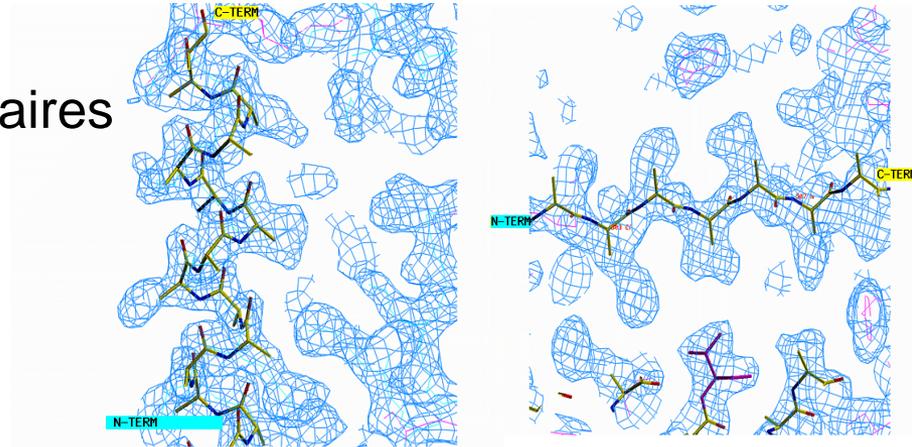


*Autre présentation en 3D (O, TURBO-FRODO, COOT...)*

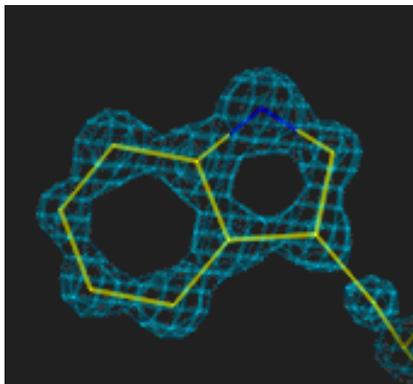


# Interprétation de la densité électronique

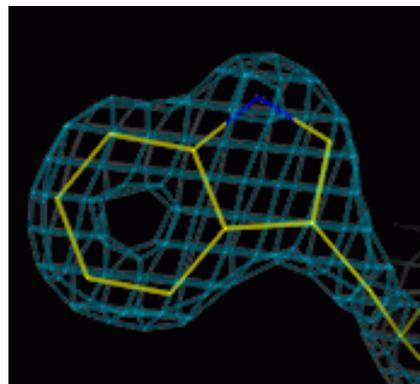
- Tracé de la chaîne principale  $C\alpha$
- Reconnaissance des structures secondaires
- Directionnalité de la chaîne
- Reconnaissance de la séquence
- Addition de molécule d'eau, ligand...



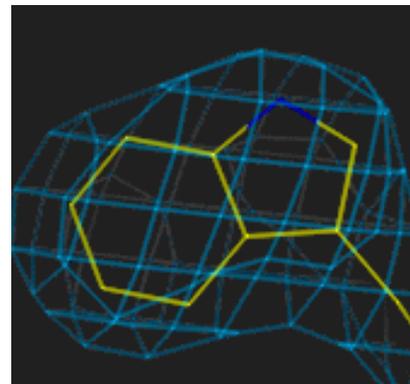
- Le S des Se-méthionines aide à l'attribution de la séquence à partir des phases Se-SAD (ou Se-SIRAS, MAD...)
- La chaîne latérale du Trp est facilement reconnaissable.



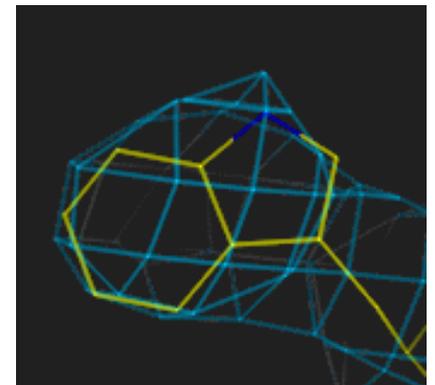
1.0 Å



1.8 Å



3.0 Å



4.0 Å

# Construction automatique

- Buccaneer (basse résolution)
- SOLVE
- SHELXE
- Arp / Warp

# Affinement cristallographique

Affinement des paramètres atomiques du modèle (x,y,z, B)

Le facteur d'accord cristi

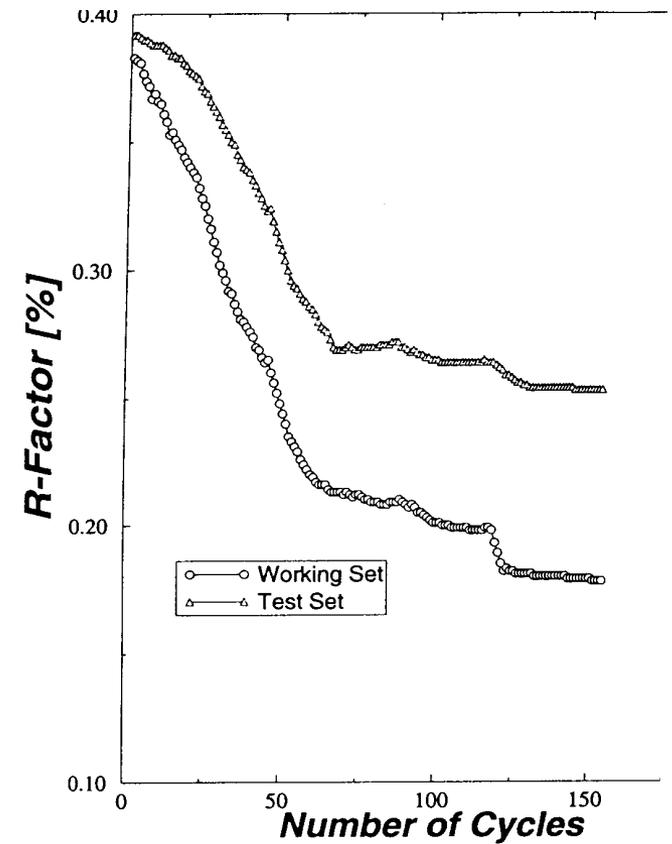
$$R = \frac{\sum_{hkl} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$$

Le facteur,  $R_{\text{free}}$ , calculé avec 5% des données non incluses dans l'affinement n'est pas biaisé par le modèle

$$R_T^{\text{free}} = \frac{\sum_{hkl \in T} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \in T} |F_{\text{obs}}|}$$

Lors de l'affinement  $R_{\text{free}}$  doit décroître avec  $R_{\text{factor}}$ .

- Utilisation de la dynamique moléculaire (simulated annealing): XPLOR, CNS, Phenix refine
- Maximum de vraisemblance (ML): Refmac, Buster, Phenix refine



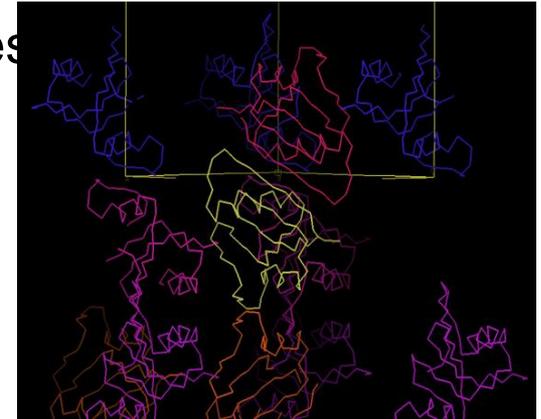
# Validation et Dépôt des Structures

Vérifier les contacts entre molécules (et symétriques)

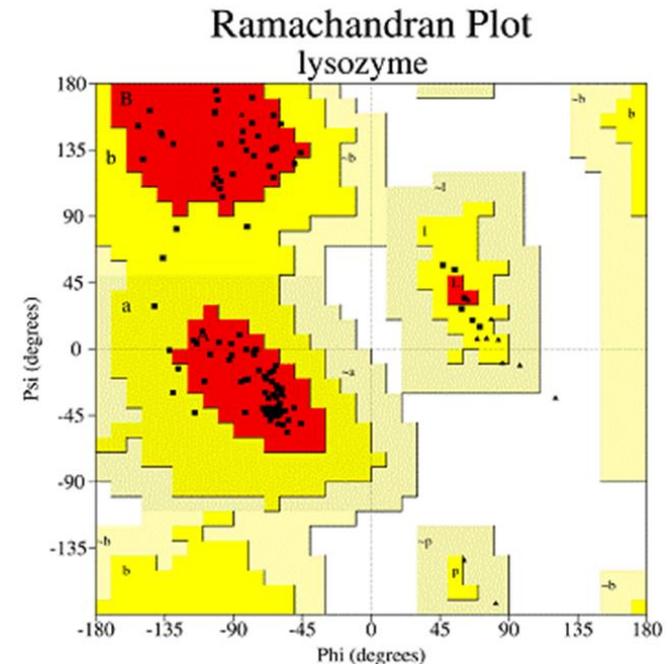
Interpréter toute la densité

La stéréochimie (angles, longueur de liaisons...) doit être dans les limites standards

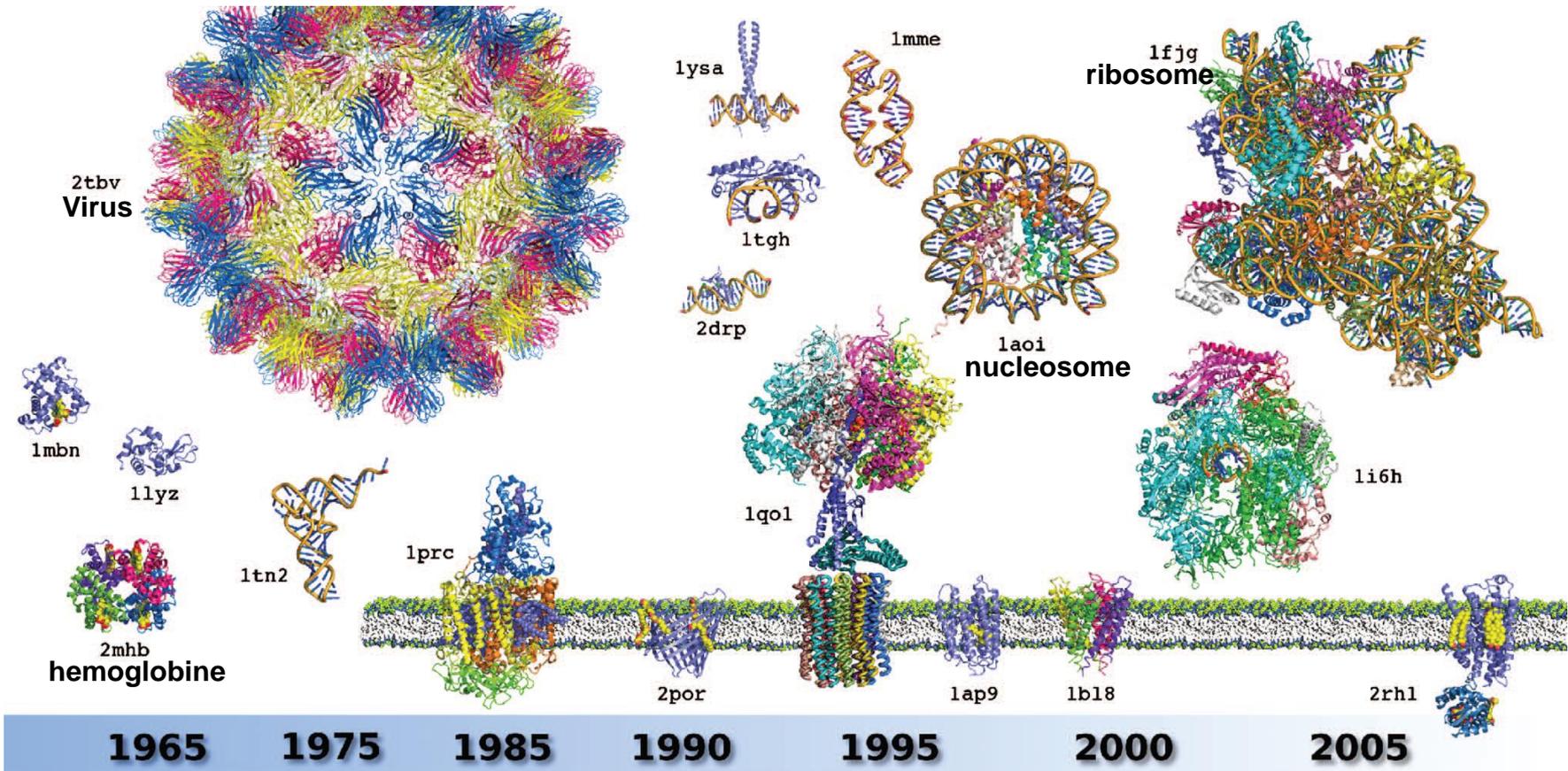
Le diagramme de Ramachandran doit être respecté



**==> WHATCHECK, MOLPROBITY, ...**



# Protein Data Bank, > 100 000 atomic structures



*Why water boils at 100°C and methane at -161°C, why blood is red and grass is green, why diamond is hard and wax is soft... The answers to all these problems have come from structural analysis.*

**Max Perutz**, July 1996, Churchill College, Cambridge