Il Problema della Fase

Laurea Magistrale in Biotecnologie Mediche

Curriculum Nanobiotecnologie

A.A. 2020-21

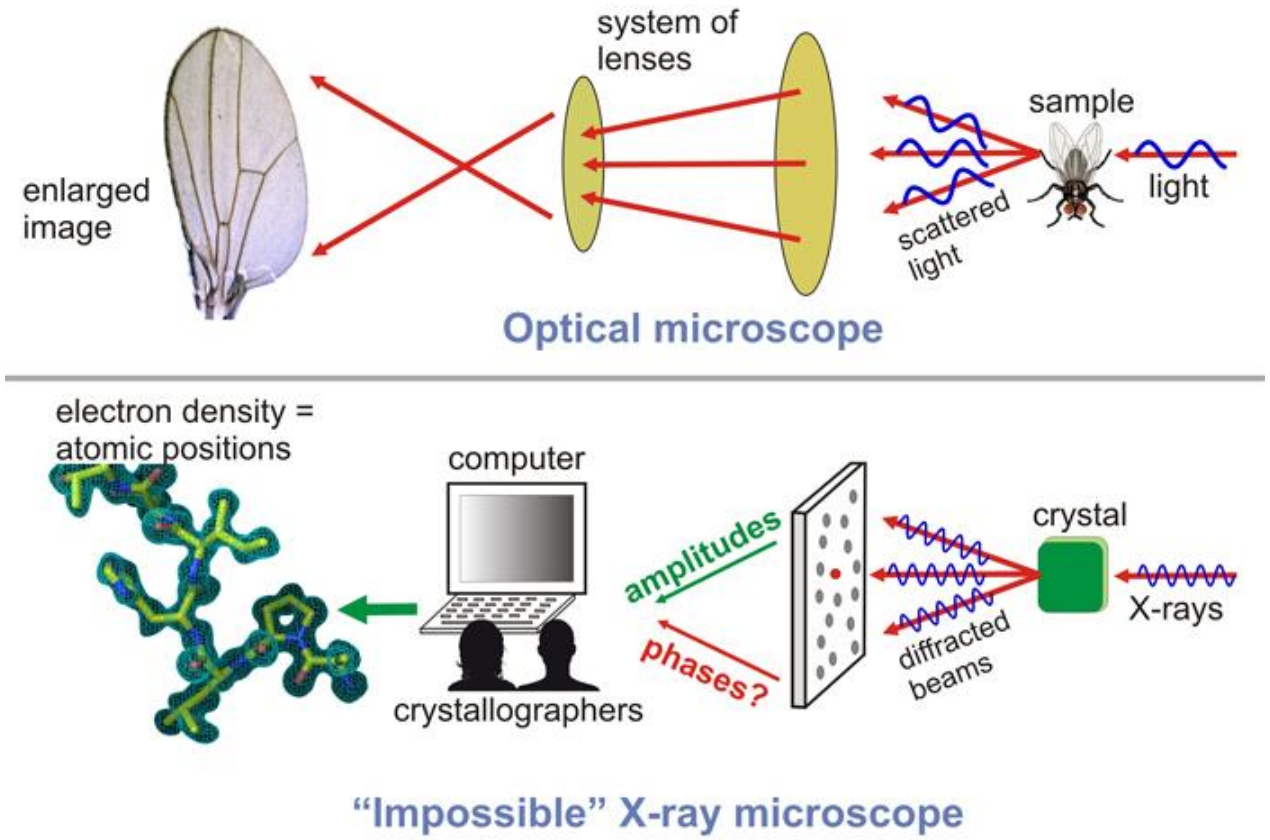
Introduzione al problema della fase

Il problema della fase

Per ricostruire l'immagine (la struttura molecolare nel cristallo) abbiamo bisogno di **Ampiezze** e **Fasi** delle onde diffratte dal cristallo.

In cristallografia non abbiamo un sistema di lenti.

L'informazione di fase va 'recuperata' in qualche altro modo.



Il problema della Fase in cristallografia

Un esperimento di diffrazione permette di ottenere le **intensità** e quindi il **modulo** dei fattori di struttura $|F_{hkl}|$

$$\rho_{cell}(r) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} \boldsymbol{F_H} \exp[-2\pi i\,(hx + ky + lz)]$$

Per la ricostruzione della densità elettronica dobbiamo conoscere i fattori di struttura in **modulo** e **fase**

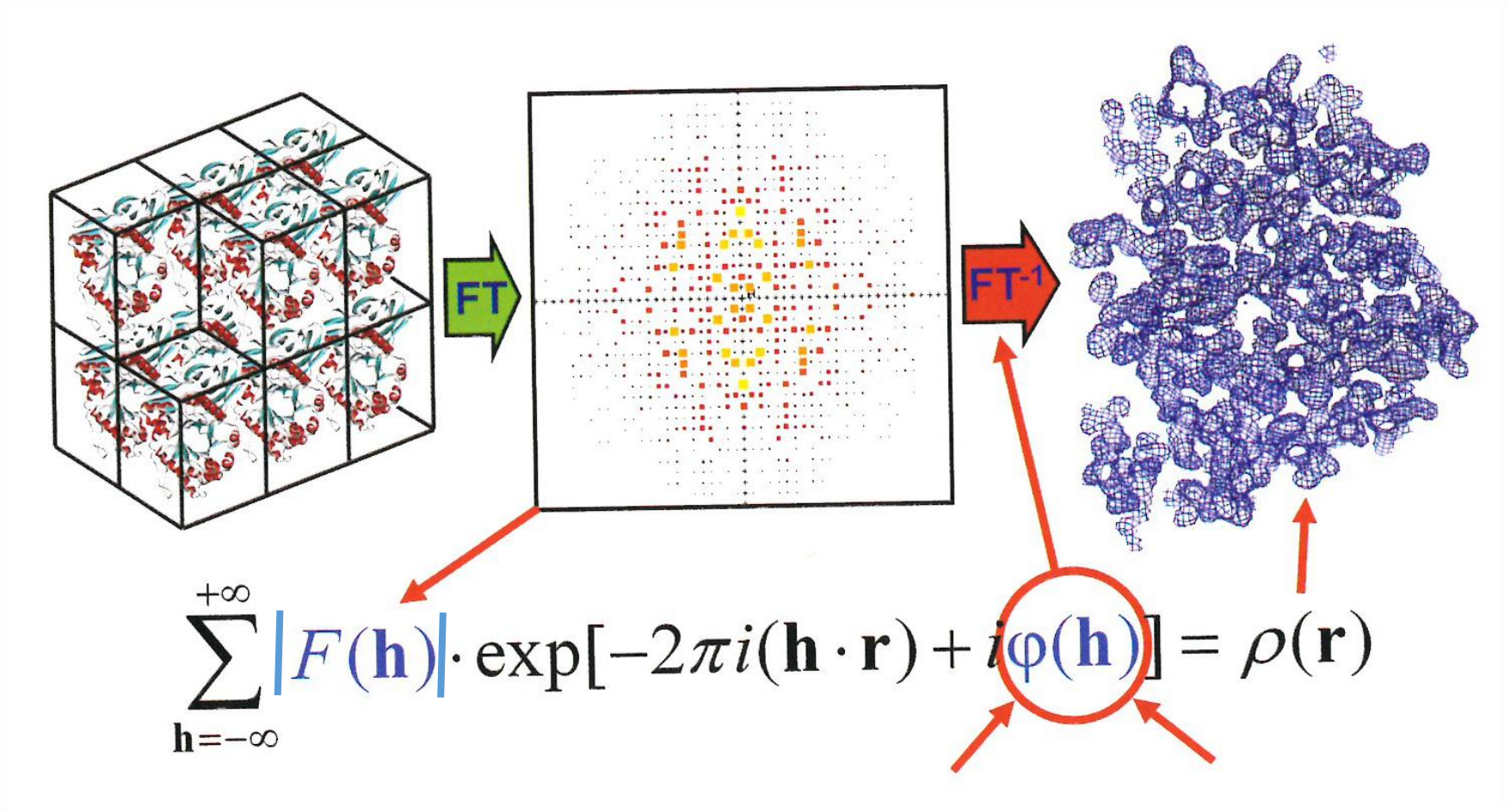
$$\rho_{cell}(r) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |\boldsymbol{F_{hkl}}| \cdot \exp[-2\pi i\,(hx + ky + lz) + i\boldsymbol{\varphi_{hkl}}]$$

Determinato dall'esperimento

Non è nota!

L'assenza di informazione sulle fasi delle onde diffratte costituisce il **problema della fase** in cristallografia

Problema della fase



L’importanza della fase

J. Karle
(intensità Karle, fase Karle)

H. Hauptman
(intensità Hauptman, fase Hauptman)

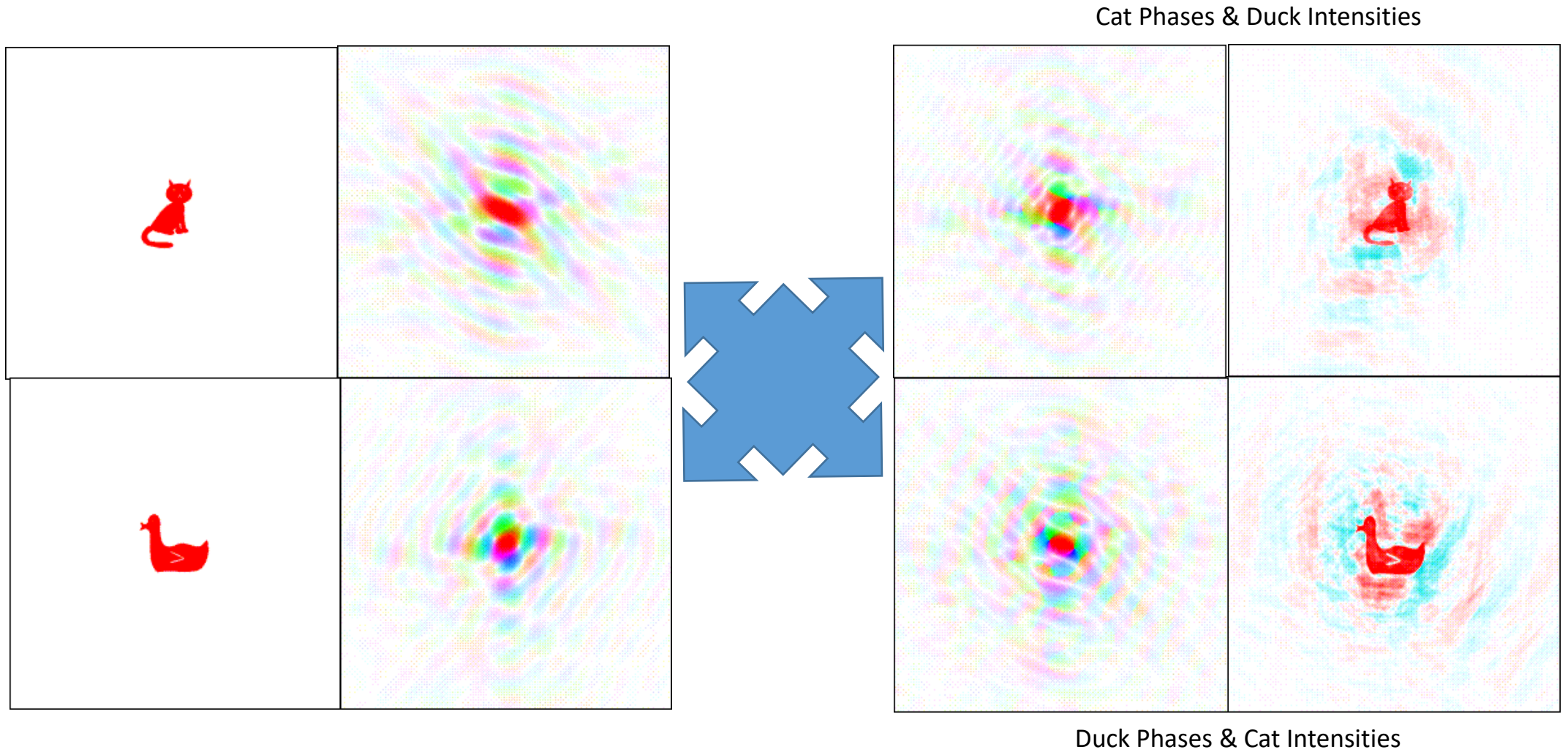
Karle/Hauptman
(intensità Karle, fase Hauptamn)

Hauptman/Karle
(intensità Hauptman, fase Karle)

Fasi corrette con intensità completamente sbagliate, permettono comunque una ricostruzione dell’immagine (da cui provengono le fasi) accettabile.

L’informazione di fase è fondamentale ed è il termine dominante nel processo di ricostruzione dell’immagine

Duck & Cat



Intensità, fase e struttura molecolare

Fase di un fattore di struttura

$$F_{hkl} = \sum_{j=1}^{N} f_j(r^*)\, exp\left(2\pi i \vec{r_j} \cdot \overrightarrow{r^*_{h,k,l}}\right)$$

Il fattore di struttura F_{hkl} può essere visto come la somma di tanti contributi atomici, ognuno con una sua fase definita dalle coordinate dell'atomo, all'interno della cella unitaria.

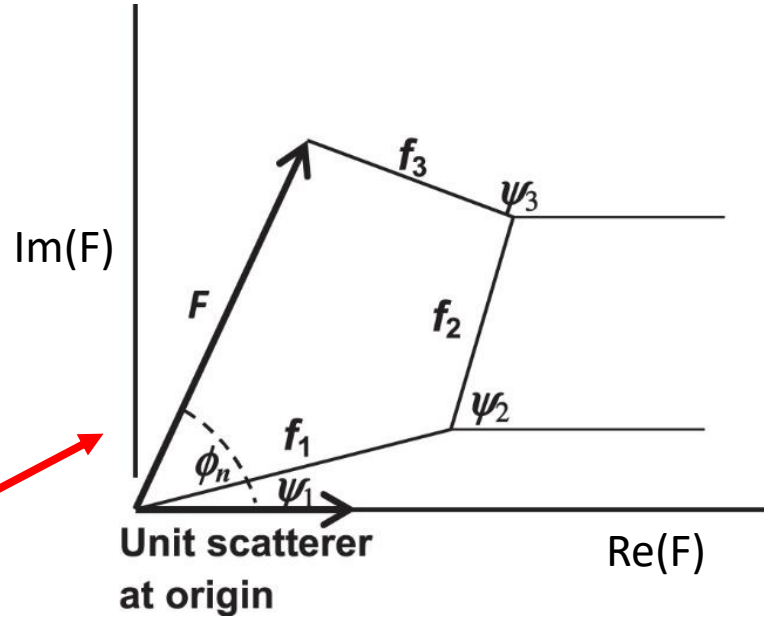
$$F_{hkl} = f_1\, exp[2\pi i(hx_1 + ky_1 + lz_1)] + \ \ldots + f_N\, exp[2\pi i(hx_N + ky_N + lz_N)]$$

Che può essere anche scritto come

$$F_{hkl} = |F_{hkl}| exp(i\phi_{hkl}) = f_1\, exp(i\psi_1) + \ \ldots + f_N\, exp(i\psi_N)$$

Che per tre atomi può essere scritto (e rappresentato sul piano di Argand)

$$F_{hkl} = |F_{hkl}| exp(i\phi_{hkl}) = f_1\, exp(i\psi_1) + f_2\, exp(i\psi_2) + f_3\, exp(i\psi_3)$$

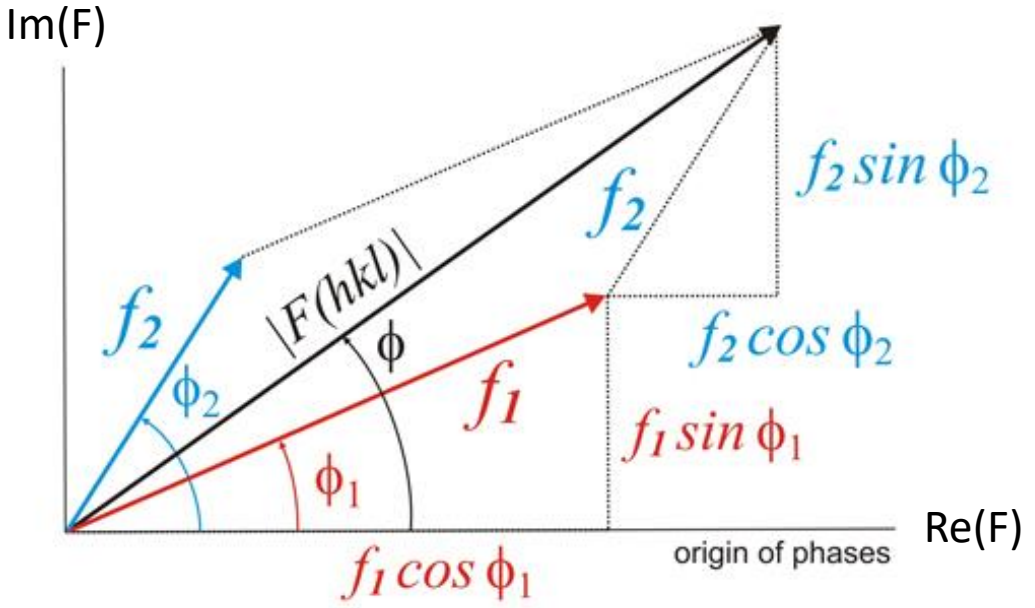


Fase di un fattore di struttura

$$F_H = |F_{hkl}|exp(i\phi_{hkl}) = \sum_{j=1}^{N} f_j(r^*)\cos[2\pi(hx_j + ky_j + lz_j)] + i\sum_{j=1}^{N} f_j(r^*)sin[2\pi(hx_j + ky_j + lz_j)]$$

La fase di un fattore di struttura è quindi determinata dalla posizione degli atomi nella cella unitaria.

Conoscendo le posizioni degli atomi nella cella unitaria e il loro tipo (C, N, Fe, …), il fattore di struttura è completamente determinato



Intensità e struttura cristallina

$$I_{hkl} = |F_{hkl}|^2 = [\sum_{i=1}^{N} f_j(r^*)\, exp\left(2\pi i \overrightarrow{r_j} \cdot \overrightarrow{r^*_{h,k,l}}\right)] \cdot [\sum_{j=1}^{N} f_j(r^*)\, exp\left(-2\pi i \overrightarrow{r_j} \cdot \overrightarrow{r^*_{h,k,l}}\right)]$$

Da cui si ottiene:

$$|F_{hkl}|^2 = F_{hkl} \cdot {F_{hkl}}^*$$

$$|F_{hkl}|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i f_j exp[2\pi i r^*_{hkl}(r_i - r_j)]$$

Le intensità diffratte contengono una informazione sulla struttura cristallina, in particolare, **le intensità diffratte dipendono dalle distanze tra gli atomi nella cella unitaria** (vettori interatomici)

Assioma fondamentale per la soluzione del problema della fase

Per una qualsiasi struttura cristallina il problema della fase è risolto quando sono determinate le fasi ϕ_{hkl} di tutti i fattori di struttura F_{hkl}, ovvero quando sono note le posizioni degli atomi nella cella unitaria (cella elementare).

La soluzione dell'equazione che lega intensità e posizioni atomiche **non è necessariamente univoca**. In teoria esiste più di una soluzione che lega intensità a posizione atomica (l'intensità dipende da vettori interatomici, non dalle posizioni 'assolute' dei singoli atomi!).

Tuttavia nella risoluzione del problema della fase si ammette e si considera sempre valido il presente assioma:

Esiste una sola soluzione al problema della fase compatibile con una struttura molecolare chimicamente corretta

Questo vuol dire che esiste una sola soluzione che permetta di ricostruire delle molecole dotate di senso chimico. Il 'senso' chimico della struttura ottenuta (lunghezze, angoli di legame, distanze interatomiche, stereochimica), ci permette di distinguere tra una soluzione corretta e una sbagliata.

Introduzione ai metodi per la risoluzione del problema della fase

La risoluzione del problema della fase

Il problema della fase è stato chiaro fin dagli inizi della cristallografia
Nel corso degli anni sono stati sviluppati metodi teorici e sperimentali per affrontare questo problema.

Inizialmente si procedeva 'trial and error', un approccio possibile con strutture semplici, ma sicuramente non adeguato a strutture (relativamente) complesse.

Metodi Sperimentali:
Richiedono ulteriori esperimenti (di diffrazione).

Metodi Computazionali:
Non richiedono ulteriori esperimenti di diffrazione, ma eventualmente delle informazioni aggiuntive

Metodi Sperimentali

- Single/**Multiple Isomorphous Replacement** (SIR/MIR)
- Multiwavelength Anomalous Diffraction (MAD)
- **Single Wavelength Anomalous Diffraction** (SAD)

Metodi Computazionali

- **Molecular Replacement**
- Metodi Diretti

Metodi Sperimentali

I metodi sperimentali per la soluzione del problema della fase si rendono necessari quando l'esperimento di diffrazione eseguito sul cristallo di cui vogliamo conoscere la struttura (detta **struttura o cristallo nativo**) non è sufficiente a determinare (completamente) le informazioni di fase necessarie al calcolo della densità elettronica.
E' quindi necessario eseguire uno o più esperimenti di diffrazione, in cui si sono modificate le condizioni dell'esperimento, in aggiunta all'esperimento sul cristallo nativo.

- **Introducendo nuovi atomi nel cristallo** (struttura o **cristallo derivato**) che modificando la diffrazione del cristallo permettono di ricavare le fasi del **cristallo nativo** (**metodi della sostituzione isomorfa**: SIR/MIR).

- **Modificando la lunghezza d'onda dei raggi-X** in modo da realizzare le condizioni di **scattering anomalo** con conseguente modifica del fattore di struttura (metodi basati sullo scattering anomalo: MAD e SAD).

I due approcci possono essere utilmente combinati.

Metodi computazionali

In questo caso le fasi dei fattori di struttura possono essere ottenute senza fare ricorso ad ulteriori esperimenti, quindi ci troviamo di fronte a **situazioni sperimentalmente più semplici, poiché basta il solo esperimento di diffrazione sul cristallo nativo.**
I due metodi di tipo non-sperimentale si basano su due approcci molto diversi:

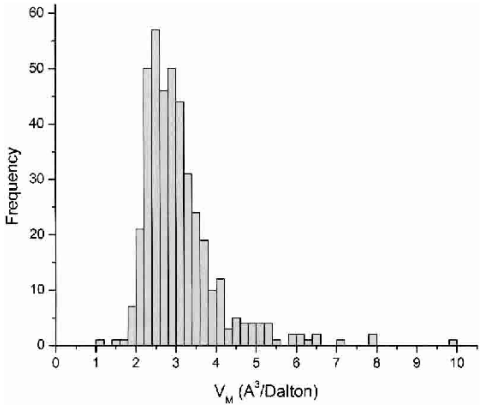
- **Conosco la struttura cristallografica di una macromolecola 'simile'** (omologia di sequenza) alla macromolecola che sto studiando (metodo della sostituzione molecolare o **Molecular Replacement**).

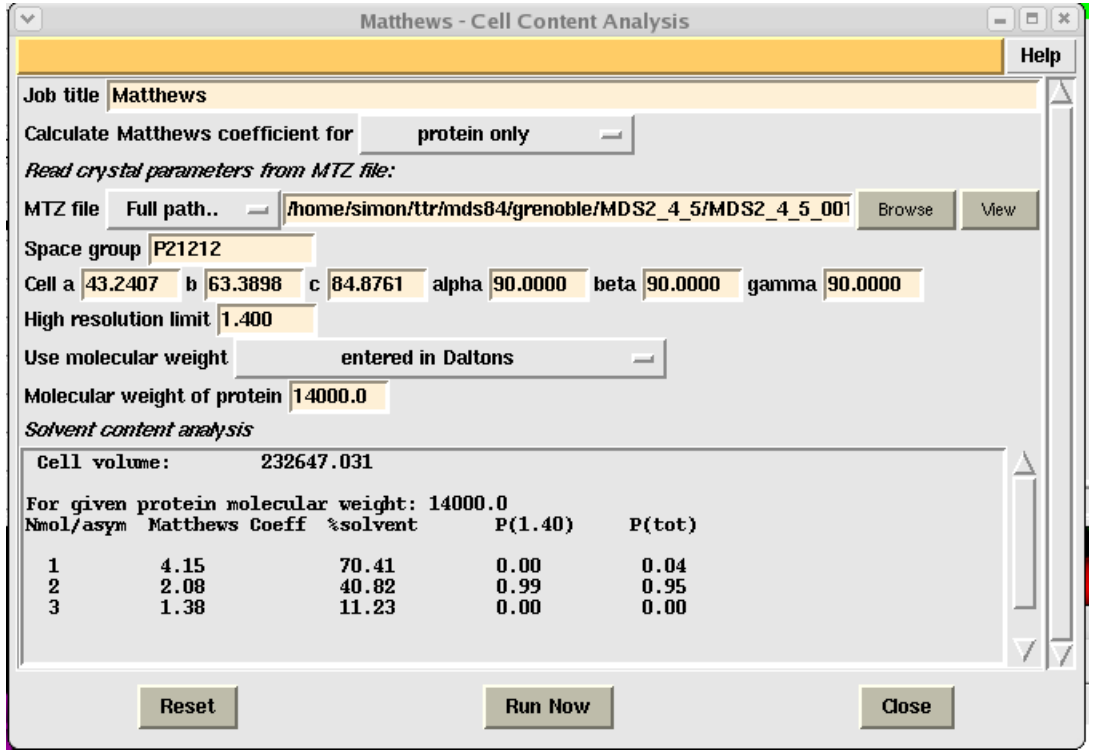
- **La struttura cristallina diffrange ad altissima risoluzione** per cui **posso determinare le fasi dei fattori di struttura per via puramente matematica (probabilistica)**. **Non ho bisogno di nessuna informazione** eccetto i fattori di struttura ottenuti dalla diffrazione del cristallo nativo e la conoscenza (anche approssimata) del contenuto, inteso come tipo e numero di atomi, della cella unitaria (**metodi diretti**).

Contenuto della Cella

E’ stato determinato che in media il rapporto tra volume della cella in $Å^3$ e peso molecolare in Da del suo contenuto definito dalla quantità di macromolecole, ha un valore intorno ai 2.15 $Å^3$/Da. Questo rapporto è noto come **coefficiente di Matthews** (V_m)

Noto il peso molecolare della molecola, il gruppo spaziale e il volume della cella elementare, è possibile stimare il numero di molecole presenti nella unità asimmetrica che danno un valore di V_m vicino al valore aspettato





Funzione di Patterson

Introduzione alla funzione di Patterson

La funzione di Patterson è la **Trasformata di Fourier delle intensità diffratte** ed è molto utile nella risoluzione del problema della fase.

L'analisi della funzione di Patterson, combinata con le informazioni sulla simmetria del cristallo (il suo gruppo spaziale), permette la determinazione della posizione degli atomi di peso atomico più elevato, presenti all'interno della cella unitaria.

La funzione di Patterson è di enorme importanza per la localizzazione dei derivati isomorfi nei metodi basati sulla sostituzione isomorfa e nella localizzazione dei diffusori anomali nei metodi basati sullo scattering anomalo.

La funzione di Patterson è inoltre largamente usata nel metodo della Sostituzione Molecolare.

Significato della funzione di Patterson

La funzione di Patterson è l'autocorrelazione della densità elettronica calcolata per una struttura cristallina

Per il caso unidimensionale:

$$P(\vec{u}) = \frac{1}{V}\int_{o}^{a} \rho(\vec{r})\rho(\vec{r}+\vec{u})\,dx$$

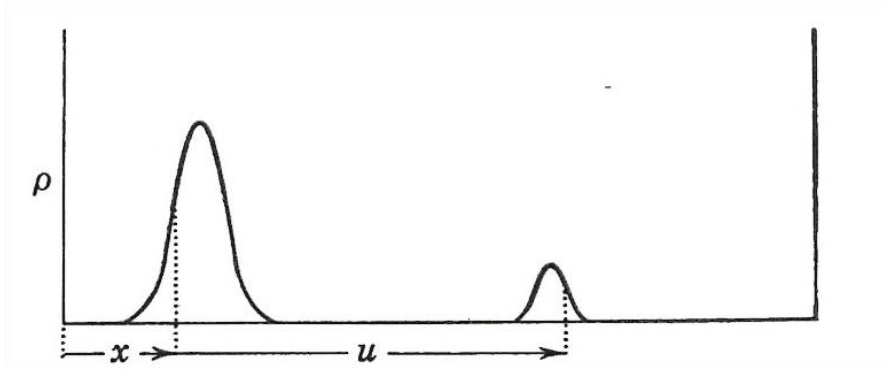
Questa funzione è già stata incontrata nella discussione dello scattering in soluzione nel calcolo della p(r).

In cristallografia, in virtù della simmetria cristallina, questa funzione, che assume il nome di funzione di Patterson, assume delle proprietà particolari, sia per il calcolo che per il suo significato, più quantitativo.

Derivazione della funzione di Patterson - 1

Caso unidimensionale, atomo in posizione *x*:

$$\rho(x) = \frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp(-2\pi i h x)$$



E per un atomo in posizione *x+u*:

$$\rho(x+u) = \frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp[-2\pi i h (x+u)]$$

Il loro prodotto sarà:

$$\rho(x)\rho(x+u) = \left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp(-2\pi i h x)\right\}\left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp[-2\pi i h (x+u)]\right\}$$

Questo non è altro che il prodotto di due densità elettroniche calcolate una nel punto *x* e l'altra a distanza *u* dal punto *x*.

Derivazione della funzione di Patterson - 2

Ricordando che il valore medio di una funzione continua è (passo dalla somma all'integrazione) :

$$\langle Y(x)\rangle = \frac{1}{a}\int_0^a Y(x)dx$$

Possiamo calcolare il valor medio del prodotto di ρ(x) ρ(x+u) al variare di *x* con *u* costante.

$$A(u) = \frac{1}{a}\int_o^a \rho(x)\rho(x+u)\,dx$$

Valor medio del prodotto delle densità che distano u

Se andiamo a sostituire quanto determinato in precedenza:

$$A(u) = \frac{1}{a}\int_o^a \left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp(-2\pi i h x)\right\}\left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp[-2\pi i h(x+u)]\right\}dx$$

Una relazione apparentemente complessa, ma che si semplifica molto…

Derivazione della funzione di Patterson - 3

$$A(U) = \frac{1}{a}\int_o^a \left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp(-2\pi i h x)\right\}\left\{\frac{1}{a}\sum_{h=-\infty}^{\infty} F_h \exp[-2\pi i h(x+U)]\right\} dx$$

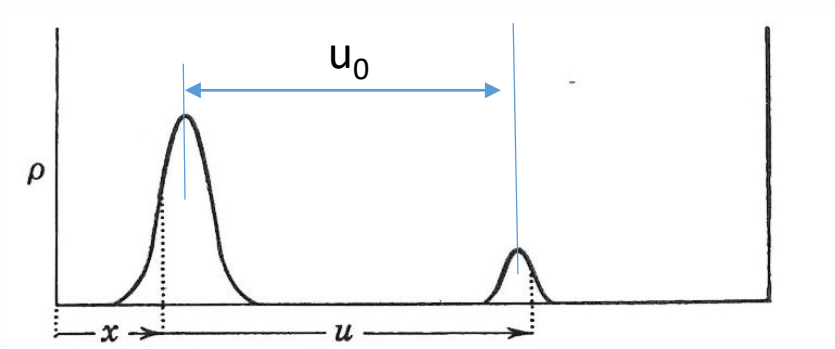
Si può sviluppare e si semplifica in:

Funzione di Patterson
(in una dimensione)

$$A(U) = \frac{1}{a^2}\sum_{h=-\infty}^{\infty} |F_h|^2 \exp(2\pi i h U)$$

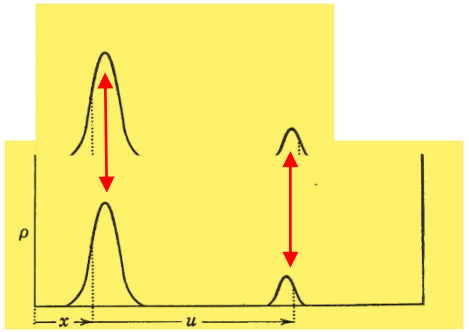
Questa funzione ci dice quale è il valor medio del prodotto delle densità elettroniche al variare del vettore u.

Il valore di questa funzione avrà valori massimi per:
U=0
U=u_0 (distanza tra i massimi dei picchi)



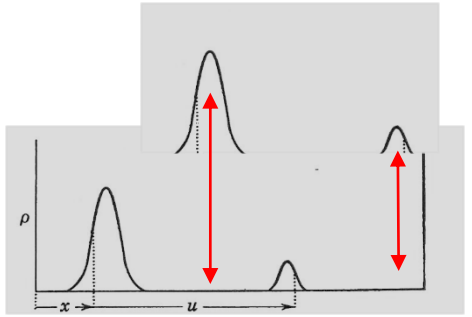
Signfificato della funzione di Patterson - 1

$U = 0 \rightarrow A(U) > 0\ (valore\ massimo)$



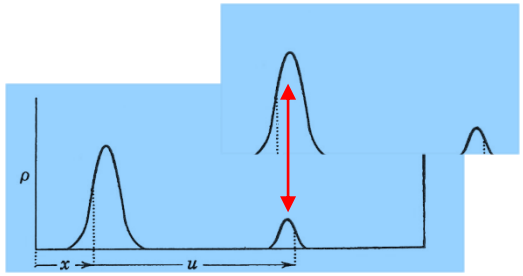
Sovrapposizione Completa

$U = v\ (diverso\ da\ u) \rightarrow A(U) = 0$



Sovrapposizione Assente

$U = u \rightarrow A(U) > 0\ [ma < A(0)]$



Sovrapposizione Parziale

Significato della funzione di Patterson - 2

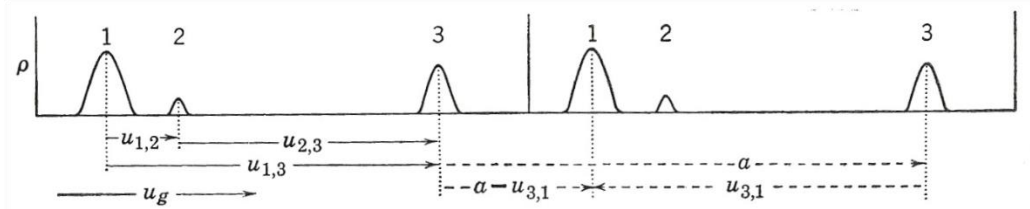
La funzione di Patterson dipende da:

- **Le distanze tra i picchi** (atomi, vettori interatomici)
- **L'intensità del picco** (altezza del picco, Z dell'atomo)

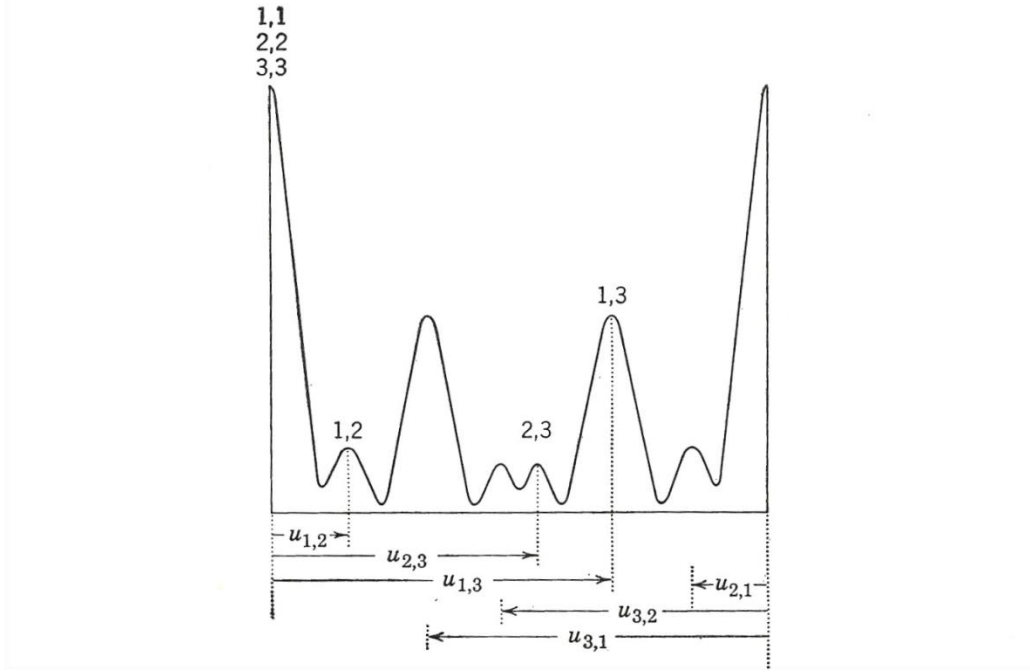
Per U coincidente con distanze tra picchi intensi, avrò valori elevati della funzione di Patterson
$[A(u_{1,3}) > A(u_{1,2}) > A(u_{2,3})$

Notare che la funzione di Patterson è **intrinsecamente centrosimmetrica** poiché la distanza interatomica non ha una direzione preferenziale $[\rho(\mathbf{x}) \cdot \rho(\mathbf{y}) = \rho(\mathbf{y}) \cdot \rho(\mathbf{x})]$

$$u_{1,2} = u_{2,1} \Rightarrow A(u_{1,2}) = A(u_{2,1})$$

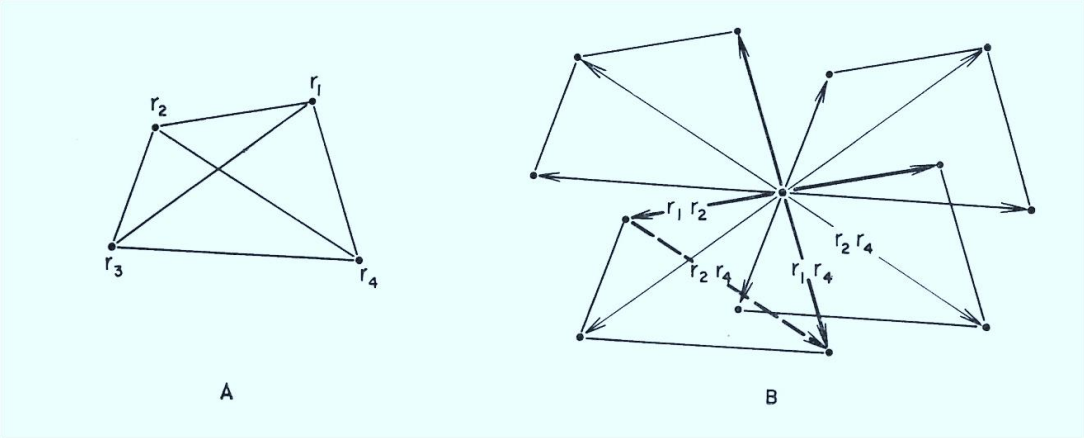


ρ(x)



A(U)

Significato geometrico della funzione di Patterson



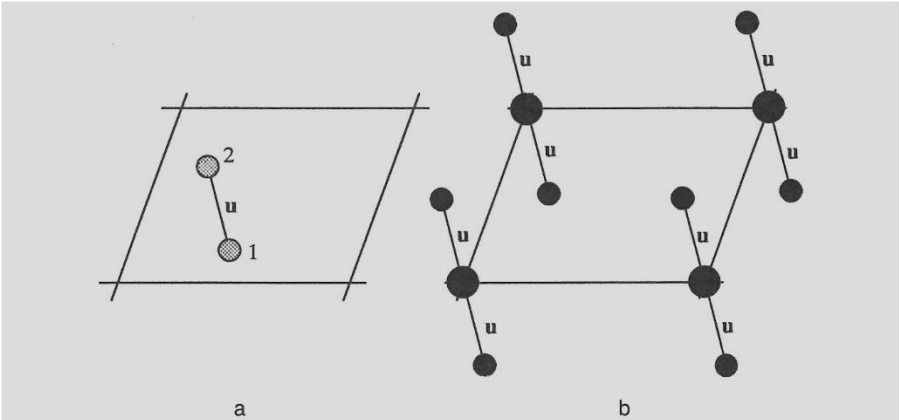
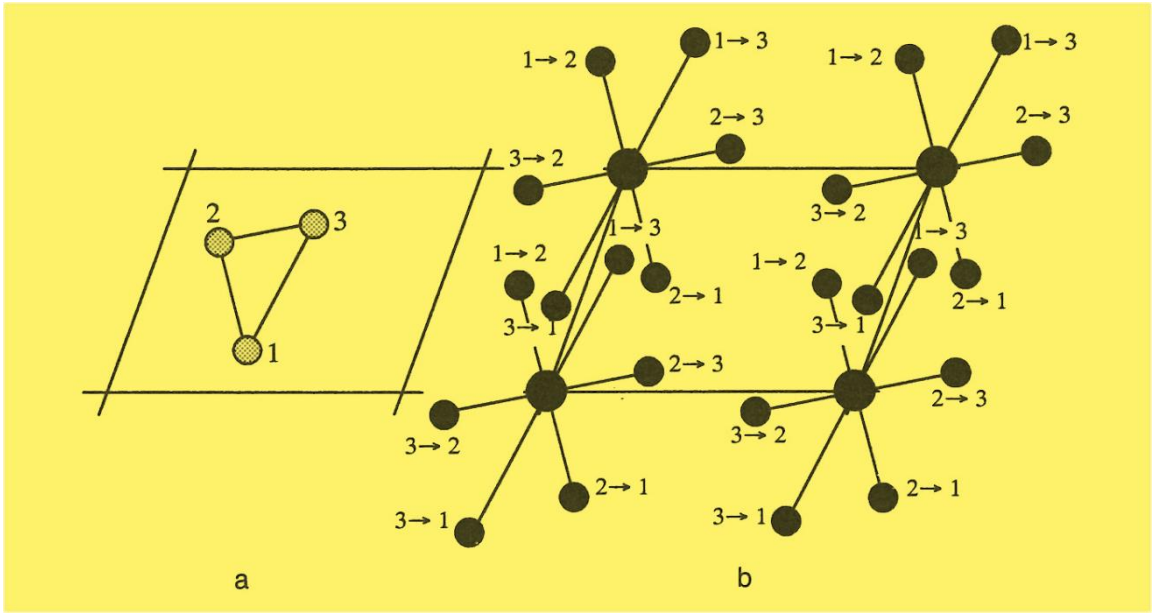


Figure 7.1. (a) A two-dimensional unit cell with only two atoms. (b) The corresponding Patterson cell.



Man mano che la struttura diventa più complessa la funzione di Patterson diventa a sua volta molto complessa, con molti picchi.

Dati N atomi abbiamo N(N-1) picchi.

Funzione di Patterson e sua interpretazione

Per una struttura molecolare in una cella unitaria, considerando gli atomi come punti discreti di valore Z (numero di elettroni dell'atomo), la funzione di Patterson descrive una mappa tridimensionale delle distanze interatomiche.

La funzione di Patterson ideale avrà valori diversi da zero solo in corrispondenza di vettori interatomici.

Ogni massimo della funzione di Patterson corrisponderà ad una distanza tra atomi nella cella unitaria.

L'intensità di ogni punto diverso da zero presente nella mappa di Patterson sarà uguale al prodotto dei numeri atomici degli atomi separati dal vettore interatomico opportuno

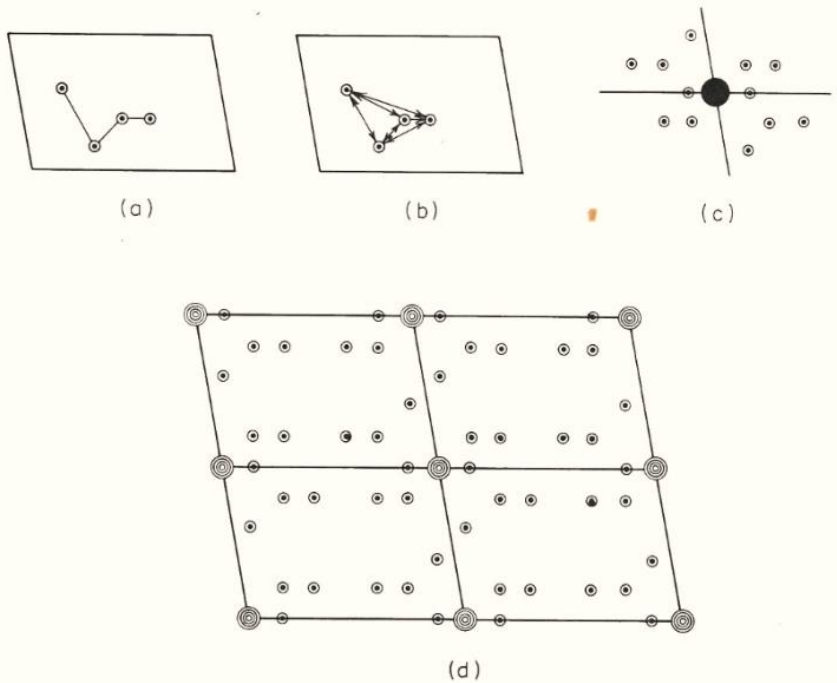


Figure 11.1. (a) Set of points. (b) Interatomic vectors. (c) Patterson peaks about the origin from one set; ●, origin peak. (d) Patterson peaks in four unit cells.

Funzione di Patterson in 3 dimensioni

La funzione di Patterson **in tre dimensioni** può essere scritta:

$$P(\vec{u}) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |F_h|^2 \exp[2\pi i(hv + kw + lz)]$$

($\vec{\boldsymbol{u}}$ è $un\ vettore\ di\ elementi\ v, w\ e\ z; V$ è $il\ volume\ della\ cella\ unitaria$)

Che può essere ulteriormente semplificata, tenendo conto della formula di Eulero e della legge di Fiedel ($|F_{h,k,l}| = |F_{-h,-k,-l}|$). La funzione di Patterson è una quantità reale.

$$\boldsymbol{P(\vec{u}) = \frac{2}{V} \sum_{h=0;\ k,l=\infty}^{\infty} |F_h|^2 \cos[2\pi(hv + kw + lz)]}$$

$|F_h|^2$ è di fatto l'intensità dell'onda diffratta dal piano di Bragg di indici (h k l)
La funzione di Patterson è la trasformata di Fourier calcolata con le intensità diffratte

Funzione di Patterson, autocorrelazione e convoluzione

In tre dimensioni la funzione di Patterson viene scritta:

$$P(\vec{u}) = \frac{1}{V} \sum_{h,k,l=0}^{\infty} |F_h|^2 \cos[2\pi(hv + kw + lz)\,]$$

Che altro non è che:

$$P(\vec{u}) = \frac{1}{V} \int_o^a \rho(\vec{r})\rho(\vec{r} + \vec{u})\, dx$$

Ma questa non è altro che la funzione di Autocorrelazione della densità elettronica $\rho(\vec{r})$, ovvero la convoluzione della $\rho(\vec{r})$ con la $\rho(-\vec{r})$, cioè la densità elettronica della struttura 'invertita'.

$$P(\vec{u}) = \rho(\vec{r}) * \rho(-\vec{r})$$

Funzione di Patterson e simmetria

La funzione di Patterson contiene informazioni sulle distanze interatomiche, però è in genere molto complessa.

La funzione di Patterson è calcolata su tutto il volume della cella elementare e quindi **abbiamo picchi dovuti ad atomi in relazione di simmetria tra di loro**.

Questa relazione di simmetria può essere sfruttata per determinare le coordinate degli atomi nella cella

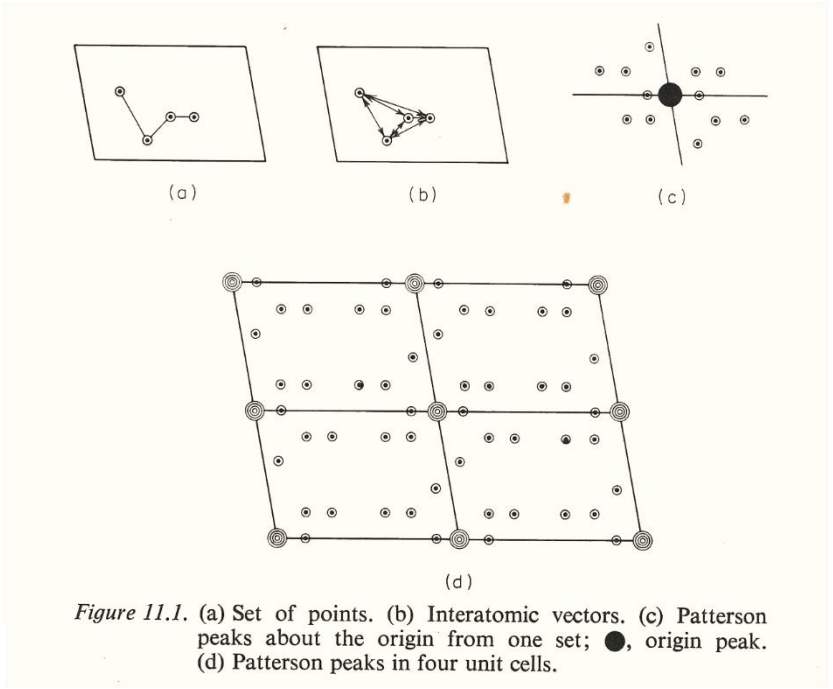
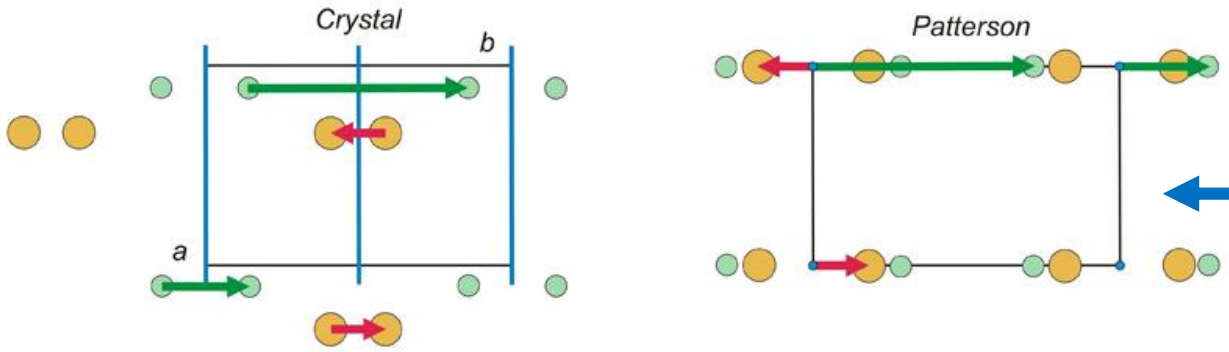
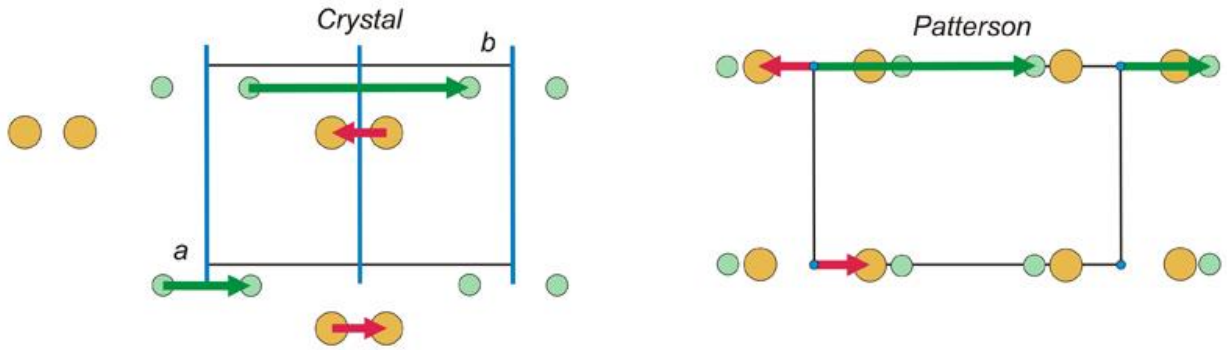


Figure 11.1. (a) Set of points. (b) Interatomic vectors. (c) Patterson peaks about the origin from one set; ●, origin peak. (d) Patterson peaks in four unit cells.



Attenzione! Mancano dei picchi!!

Funzione di Patterson e coordinate atomiche



A partire dalle relazioni di simmetria descritte dal gruppo spaziale della struttura cristallina, nei casi più favorevoli, l'interpretazione delle distanze interatomiche ottenute dalla funzione di Patterson permette di determinare le coordinate degli atomi presenti nella cella unitaria.

Nella pratica questo è possibile solo per strutture piuttosto semplici e praticamente impossibile per le strutture di macromolecole biologiche.

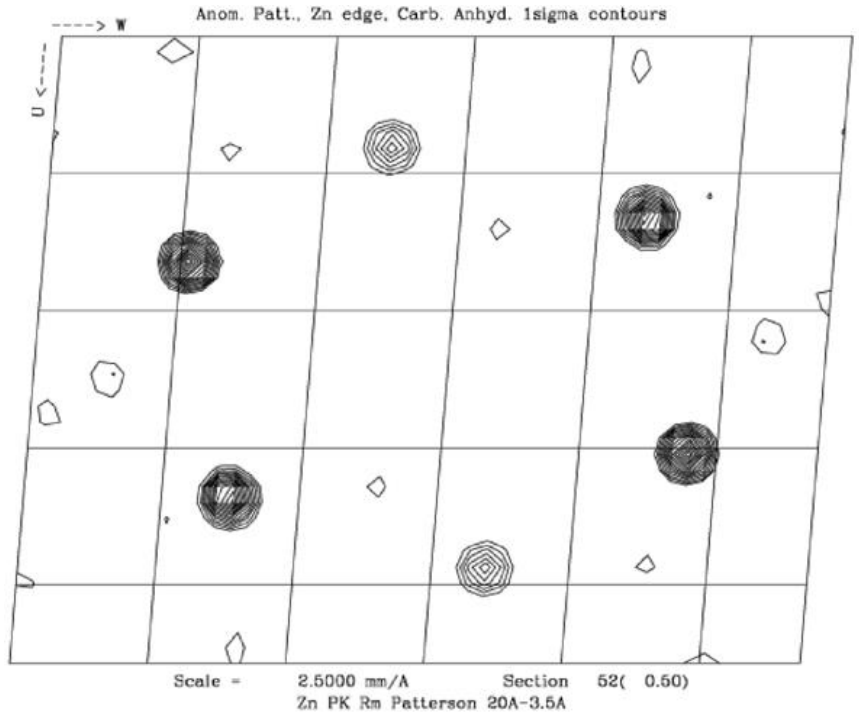
Tuttavia, **tramite la funzione di Patterson è possibile determinare le coordinate di atomi con elevato peso atomico, se presenti nella cella unitaria**.

Metodo dell'Atomo Pesante

Se nella nostra molecola esiste un atomo con Z molto alto (**Atomo Pesante**), nella funzione di Patterson i picchi che lo coinvolgono, ovvero le distanze interatomiche che lo coinvolgono avranno intensità molto alta in quanto proporzionale al prodotto dei numeri atomici (Z) degli elementi coinvolti.
In particolare:

Vettore interatomico	**Intensità del picco**
At. Pesante – At. Pesante	**Molto Alta ($Z_{HA} \cdot Z_{HA}$)**
At. Pesante – At. Leggero	Media ($Z_{HA} \cdot Z_{LA}$)
At Leggero – At. Leggero	Molto Bassa ($Z_{LA} \cdot Z_{LA}$)

Sfruttando le relazioni di simmetria (che sono note), in presenza di un atomo pesante è possibile determinare la posizione dell'atomo pesante medesimo.



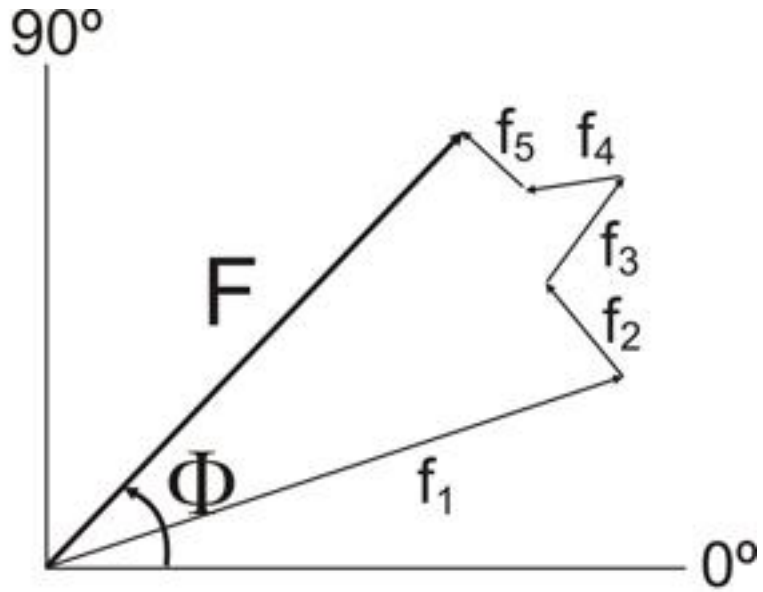
Metodo dell'Atomo Pesante

La posizione dell'atomo pesante consente di ottenere delle fasi approssimate dei fattori di struttura, in genere sufficienti come inizio, per la risoluzione del problema della fase, poiché **l'atomo pesante 'domina' nella sommatoria del fattore di scattering.**

Il risultato della funzione di Patterson è in genere piuttosto complesso, più è elevato Z dell'atomo pesante più i picchi che lo riguardano sono facilmente identificabili.

Il numero degli atomi pesanti non deve essere eccessivo per evitare che l'analisi dei picchi diventi eccessivamente complessa (capire chi è in relazione con chi).

Anche una simmetria elevata può complicare l'analisi della funzione di Patterson (molti picchi dovuti a molte relazioni di simmetria)



f_1 è il contributo dell'atomo pesante

Funzione di Patterson: riassumendo…

- La funzione di Patterson è la **trasformata di Fourier delle intensità** diffratte da un cristallo
- La funzione di Patterson, per atomi puntiformi, avrà v**alori diversi da zero solo in corrispondenza di vettori interatomici**
- La **posizione** dei picchi della funzione di Patterson indica la **distanza interatomica tra due atomi nella cella unitaria**
- L'**intensità** di un picco della funzione di Patterson è **proporzionale al peso atomico dei due atomi** a cui si riferisce la distanza interatomica
- La conoscenza della **simmetria cristallina** permette, in linea di principio, di ottenere le **coordinate degli atomi** a partire dal calcolo della funzione di Patterson
- La funzione di Patterson si può calcolare facilmente a partire dalle intensità diffratte, ma la sua complessità aumenta notevolmente con l'aumentare del numero degli atomi
- Se nella mia struttura cristallina sono presenti un **piccolo numero di atomi ad elevato peso atomico**, tramite il calcolo della funzione di Patterson e **considerando le relazioni di simmetria, posso determinare le coordinate di questi atomi**
- **Atomi di elevato peso atomico contribuiscono maggiormente al fattore di struttura, quindi le fasi calcolate a partire dalle loro coordinate costituiscono un'approssimazione accettabile**

Sostituzione Isomorfa: Introduzione

Metodo del derivato isomorfo (SIR/MIR)

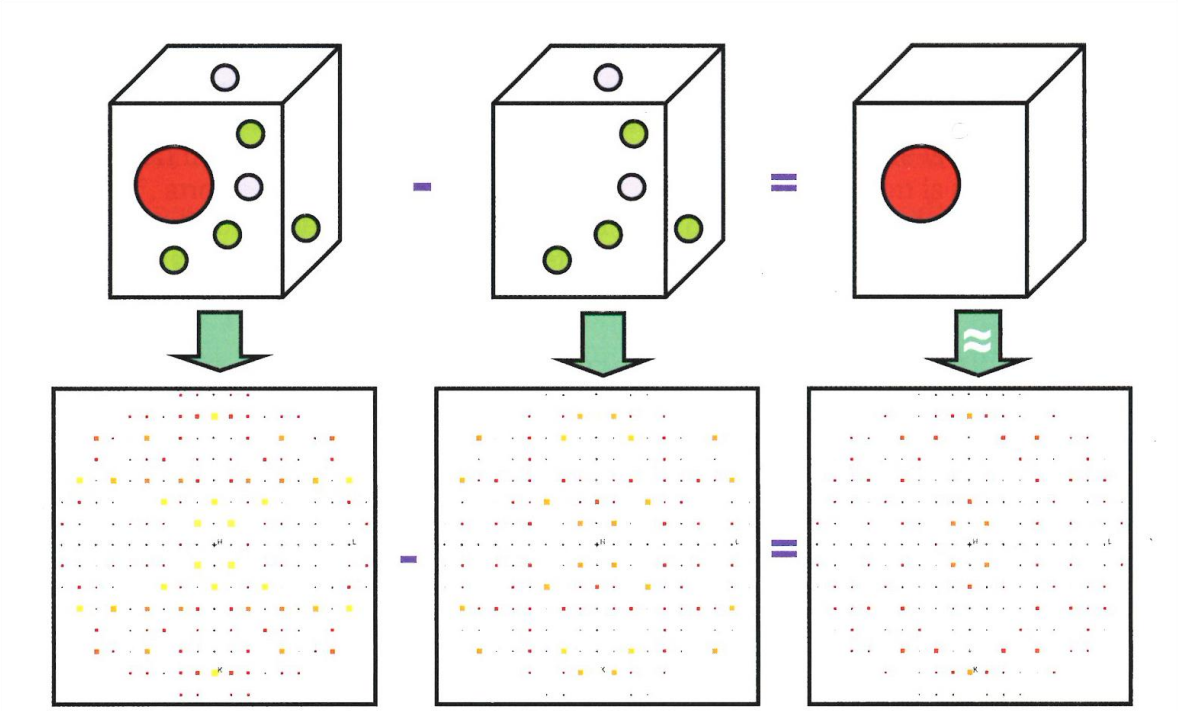
Il metodo del Derivato Isomorfo,Single (SIR) or Multiple Isomorphous Replacement (MIR), è concettualmente molto semplice e può essere descritto nel seguente modo.

Sia dato un cristallo di proteina, di cui vogliamo determinare la struttura (Cristallo **Nativo**)

1. Del cristallo nativo misuriamo le intensità diffratte e avremo alla fine un insieme di fattori di struttura $\boldsymbol{F_P}$
2. In un secondo cristallo introduciamo un **atomo pesante** (Numero atomico Z elevato) di qualche tipo (Hg, Pb, Pt, Au) ottenendo un **Derivato Isomorfo**, quindi ne misuriamo le intensità diffratte, ottenendo così i suoi fattori di struttura $\boldsymbol{F_{PA}}$
3. Dalla differenza tra i fattori di struttura si ricava la posizione degli atomi pesanti e quindi delle fasi iniziali $\boldsymbol{\varphi_P}$ da utilizzare con i fattori di struttura nativi $\boldsymbol{F_P}$, sperabilmente sufficienti per ottenere una prima mappa di densità elettronica adeguata

$$\rho(r) \sim \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |\boldsymbol{F_P}| \cdot \exp[-2\pi i\,(hx + ky + lz) + i\boldsymbol{\varphi_P}]$$

Metodo del Derivato Isomorfo



$$\boldsymbol{F_{PA} = F_P + F_A}$$

Sostituzione Isomorfa Singola - SIR

Sperimentalmente io misuro solo $|F_{PH}|$ e $|F_P|$ e non conosco né $|F_A|$ né la sua fase φ_A
Inoltre:

$$F_{PA} = F_P + F_H \neq |F_{PA}| = |F_P| + |F_A|$$

Infatti se consideriamo il modulo al quadrato (intensità)

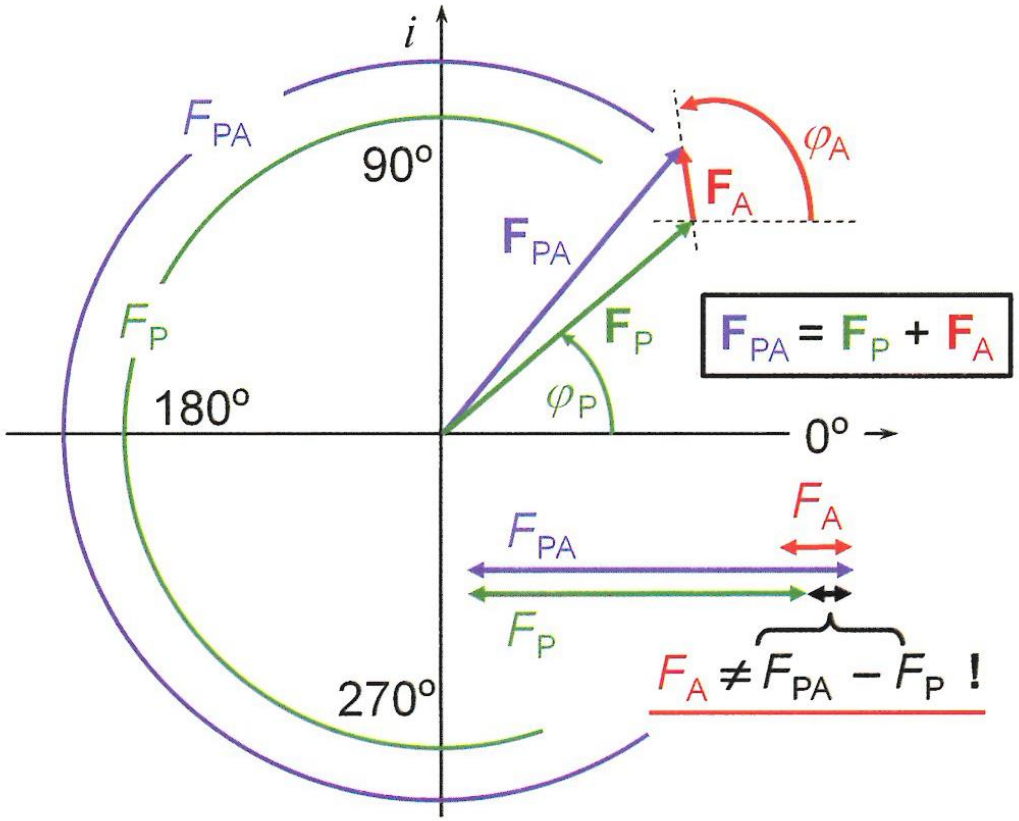
$$|F_{PA}|^2 = |F_P + F_A|^2$$

Sviluppando e usando il teorema dei coseni, otteniamo:

$$|F_{PA}|^2 = |F_P|^2 + |F_A|^2 - 2|F_P||F_A| \cos[(\varphi_P + (\pi - \varphi_A)]$$

Quindi:

$$|F_{PA}|^2 = |F_P|^2 + |F_A|^2 + 2|F_P||F_A| \cos(\varphi_P - \varphi_A)$$



Sostituzione Isomorfa Singola - SIR

$$|F_{PA}|^2 = |F_P|^2 + |F_A|^2 + 2|F_P||F_A|\cos(\varphi_P - \varphi_A)$$

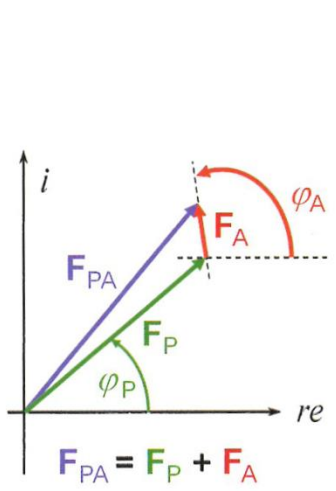
Che possiamo trasformare in:

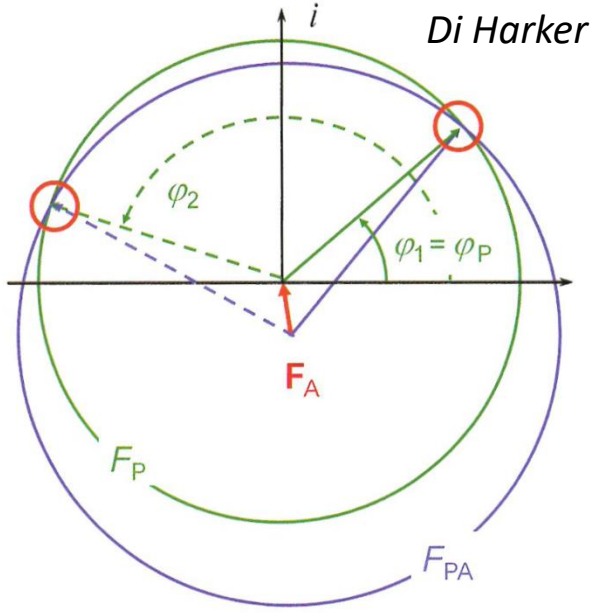
$$\boldsymbol{\varphi_P = \varphi_A \pm \cos^{-1} \frac{|F_{PA}|^2 - |F_P|^2 - |F_A|^2}{2|F_P||F_A|}}$$

Noi conosciamo solo $|\boldsymbol{F_{PA}}|^2 e |\boldsymbol{F_P}|^2$ (intensità misurate)

$|\boldsymbol{F_A}|^2$ deve essere stimato come anche $\boldsymbol{\varphi_A}$

Diagramma Di Harker





<u>Ho un'incertezza sul segno del coseno</u>, un solo derivato isomorfo non determina le fasi in modo univoco, esiste un'**ambiguità nelle fasi determinate con un solo derivato.**

La sostituzione Isomorfa Singola (**SIR**, **Single Isomorphous Replacement**), non risolve completamente il problema della fase

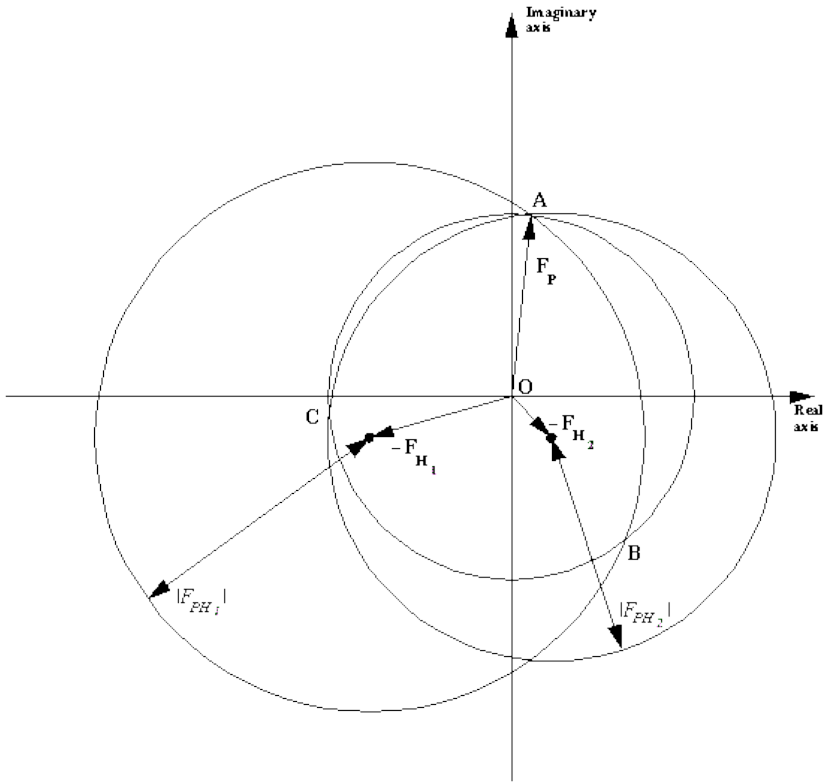
Sostituzione Isomorfa Multipla: MIR

Per poter determinare senza ambiguità le fasi di un insieme di fattori di struttura, abbiamo bisogno di **almeno due derivati isomorfi**: Metodo della sostituzione isomorfa multipla (**MIR**, **Multiple Isomorphous Replacement**).

In linea teorica con due derivati isomorfi è possibile assegnare la fase corretta ad ogni fattore di struttura.

In pratica il numero di derivati isomorfi necessario può essere maggiore…

Per esempio un derivato isomorfo in cui un atomo pesante occupi la stessa posizione occupata da un atomo pesante in un altro derivato isomorfo, non da alcun nuovo contributo alla determinazione della fase



Sostituzione Isomorfa: Aspetti sperimentali

Isomorfosimo

Condizione fondamentale **perché il metodo MIR funzioni è che il derivato ottenuto in seguito all'introduzione dell'atomo pesante, sia quanto più isomorfo possibile alla struttura nativa.**

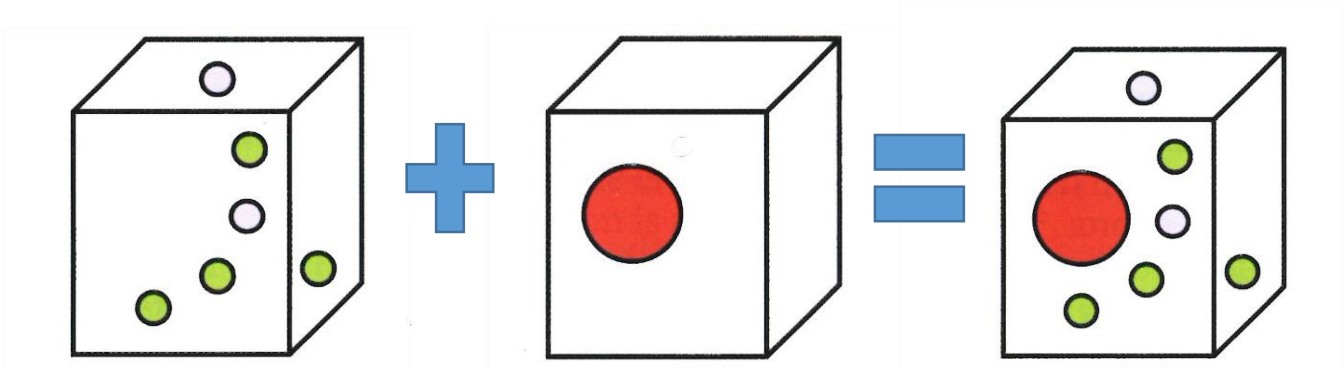
Isomorfo vuol dire che:

- la cella unitaria della struttura nativa e del derivato **devono essere il più possibile uguali**
- La **struttura della proteina nel cristallo non venga perturbata** dall'introduzione dell'atomo pesante

Queste condizioni sono difficili da ottenere, nei fatti l'isomorfismo non è mai perfetto.

La mancanza di isomorfismo può rendere il derivato inutile.

In generale il non perfetto isomorfismo introduce degli errori che rendono la determinazione della fase più difficile



Atomi Pesanti

Nel SIR/MIR devono essere introdotti atomi con numero atomico Z elevato: **Au, Pt, Hg, Pb**, [U] e altri ancora. Questi atomi vengono introdotti come composti inorganici o organometallici, e interagiscono con gruppi specifici:
Tra i più usati (sono quasi tutti **MOLTO TOSSICI**):

$HgCl_2$, K_2HgI_4	Interagisce con **Cys**, His
K_2PtCl_4, K_2PtBr_6	Interagisce con Cys, **His**, Met
$KAu(CN)_2$, $AuCl_3$	Interagisce con Cys, **His**
$Pb(NO_3)_2$, $Pb(CH_3COO)_2$	Interagisce con Glu, Asp

Esistono molti altri preparati per le più diverse condizioni, dallo Xe (gassoso) a grandi cluster metallici del Tantalio e Tungsteno e Bromo per grandi strutture proteiche.

Negli oligonucleotidi (RNA/DNA) si utilizzano nucleotidi bromurati

Atomi Pesanti

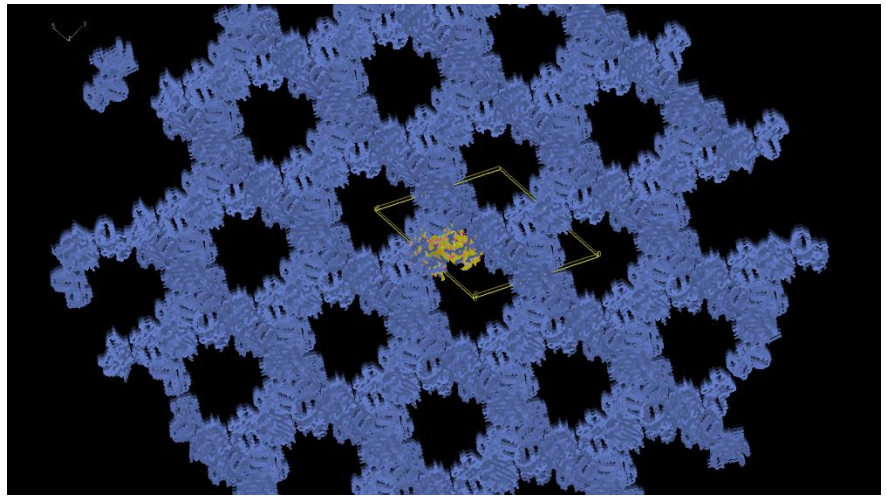
Nel considerare i potenziali derivati isomorfi, bisogna fare alcune considerazioni sulla chimica del sistema:

- Il pH del Buffer
- La presenza di eventuali complessanti (es: ammine) o chelanti (es: EDTA)
- Il tipo di crioprotettore (potrebbe complessare l'atomo)

Gli atomi pesanti sono introdotti in genere per 'soaking', i cristalli nativi sono trasferiti per un certo tempo (dai giorni alle decine di minuti) in una goccia contenente l'atomo pesante ad una concentrazione data (dai 100 μM alle decine mM).

L'introduzione dell'atomo pesante in genere 'deteriora' il cristallo che in genere diffrangerà ad una peggiore risoluzione rispetto al cristallo nativo (raramente i derivati isomorfi diffrangono bene).

Inoltre, la presenza di atomi pesanti aumenta il problema del danno da radiazione.



Gli atomi pesanti penetrano nella struttura cristallina per mezzo dei canali presenti nel cristallo

Determinazione della posizione dell'atomo pesante A

Per una corretta 'fasatura' è cruciale la determinazione della, posizione degli atomi pesanti ($|\boldsymbol{F_A}|$, $\boldsymbol{\varphi_A}$)

Come metodo di uso generale si utilizza la funzione di Patterson con coefficenti

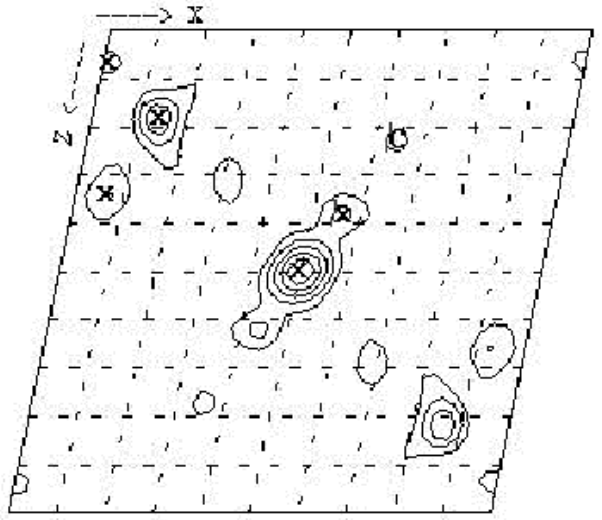
$$(|F_{PH}| - |F_P|)^2$$

Dove:

$$\Delta F_{ISO} = |F_{PH}| - |F_P|$$

E

$$P(\vec{u}) = \frac{1}{V} \sum_{h,k,l} {\Delta F_{ISO}}^2 \cos[2\pi(hv + kw + lz)]$$



E' la funzione di **Patterson** calcolata con le **differenze isomorfe** (**Isomorphous difference Patterson map**)

Allo stato attuale esistono algoritmi più complessi che fanno anche uso dei *metodi diretti*

Fasi con il metodo MIR

Una volta note le coordinate dell'atomo pesante si possono assegnare le fasi iniziali.

Nella determinazione delle posizioni degli atomi pesanti va considerato che gli atomi pesanti possono legare in più posizioni e non sempre con occupazione piena (**occupanza** = percentuale di posizioni occupate rispetto alle teoriche totali nel cristallo), ovvero non tutte le celle unitarie del cristallo contengono atomi pesanti.

Tutti questi parametri (posizione, occupanza…) vengono migliorati nel corso del processo di 'fasatura' (*phasing*).

Esistono diversi indicatori per la qualità delle fasi ottenute, l'indicatore migliore è però la qualità della mappa di densità elettronica ottenuta $\rho(r)$.

Riassumendo SIR/MIR

- Nel metodo del derivato isomorfo, viene introdotto un *atomo pesante* nella struttura cristallina
- La differenza tra i fattori di struttura del cristallo derivatizzato e quelli della struttura nativa permettono di identificare la posizione dell'atomo pesante
- Il metodo del derivato isomorfo funziona se esiste isomorfismo tra cristallo nativo e cristallo derivato (l'atomo pesante introdotto non modifica la struttura cristallina preesistente)
- Le coordinate dell'atomo pesante permettono di calcolare delle fasi iniziali approssimate che permettono un primo calcolo della densità elettronica
- Sono necessari ALMENO due derivati isomorfi perché le fasi siano determinate in modo univoco
- Gli errori sperimentali (assenza di isomorfismo soprattutto) o la bassa affinità dell'atomo pesante (occupazione parziale e bassa) possono determinare la necessità di più di due derivati isomorfi per la risoluzione del problema della fase
- I derivati isomorfi sono preparati a partire da molecole contenenti atomi ad alto peso atomico (Hg, Pb, Pt, Au)
- La posizione degli atomi pesanti è determinata utilizzando la funzione di Patterson calcolata con la differenza tra moduli dei fattori di struttura del derivato isomorfo e del cristallo nativo

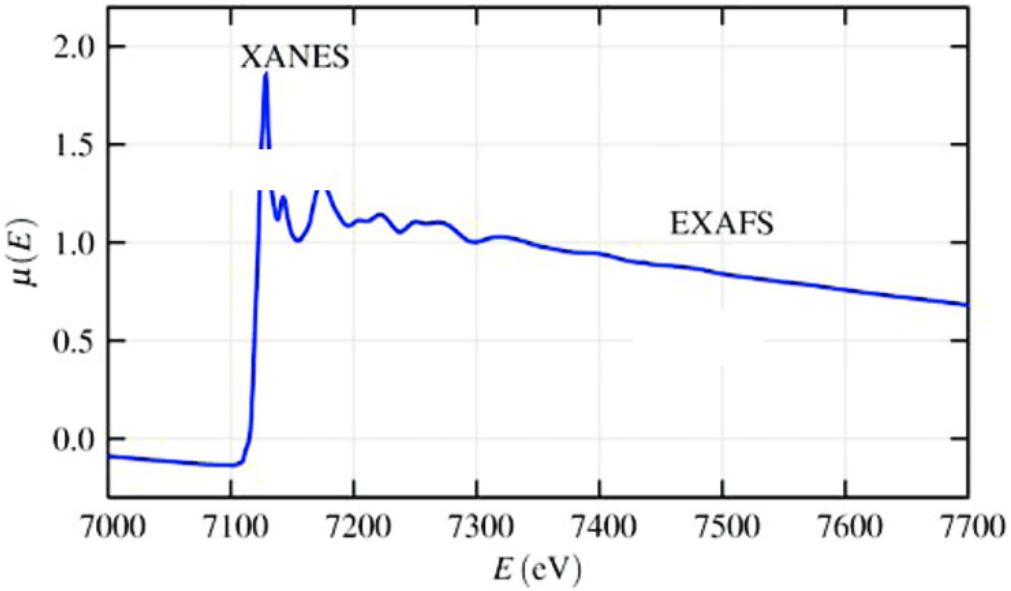
Uso dello Scattering Anomalo per la determinazione delle fasi

Scattering Anomalo

La legge di Friedel stabilisce che le intensità diffratte dai piani di Bragg con indici hkl opposti, sono tra loro uguali

Se però la lunghezza d'onda della radiazione incidente è vicina o coincidente a quella di una soglia di assorbimento di un elemento chimico contenuto nel cristallo, il fattore di struttura atomico viene modificato con due termini aggiuntivi:

$$f(r^*) = f_{at} + \Delta f' + if''$$



La presenza di un termine immaginario causa la 'rottura' della legge di Friedel, per cui se la lunghezza d'onda è vicina ad una soglia di assorbimento le intensità diffratte da piani di Bragg centrosimmetrici **non sono più uguali**.

$$\boldsymbol{I(h\,k\,l) \neq I(-h\,-k\,-l)}$$

Incidenza dello Scattering Anomalo sui F_{hkl}

Per un atomo che si trova in condizioni di scattering anomalo, il fattore di struttura atomico viene modificato dalla presenza di una parte reale ($\Delta f'$) e di una parte immaginaria (f'') il valore di di questi due termini dipende dal tipo di atomo e dalla lunghezza d'onda del fascio di raggi-X incidente

$$\boldsymbol{f(r^*)_\lambda = f_{at} + \Delta f'_\lambda + i f''_\lambda}$$

Se consideriamo la formula del fattore di struttura nel caso in cui ci sia un atomo in condizioni di scattering anomalo

$$F^{\lambda}_{hkl} = \sum_{j=1}^{N} f_j(r^*)\, exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right] + \left(\Delta f'_A + i f''_A\right) exp[2\pi i(hx_A + ky_A + lz_A)]$$

$$\boldsymbol{F^{\lambda}_{hkl} = F_{hkl} + F^{Anom}_{hkl}}$$

Il fattore di struttura avrà quindi una componente aggiuntiva (termine anomalo) dipendente dall'atomo (diffusore anomalo) e dalla lunghezza d'onda

Incidenza dello Scattering Anomalo sui F_{hkl}

l termine immaginario di f'' introduce uno 'sfasamento' di 90° rispetto a f_{at} e f'

$$if''_{\lambda} = |f''_{\lambda}|\exp(i\frac{\pi}{2})$$

Possiamo scrivere, riunendo le parti reali, l'equazione di f come:

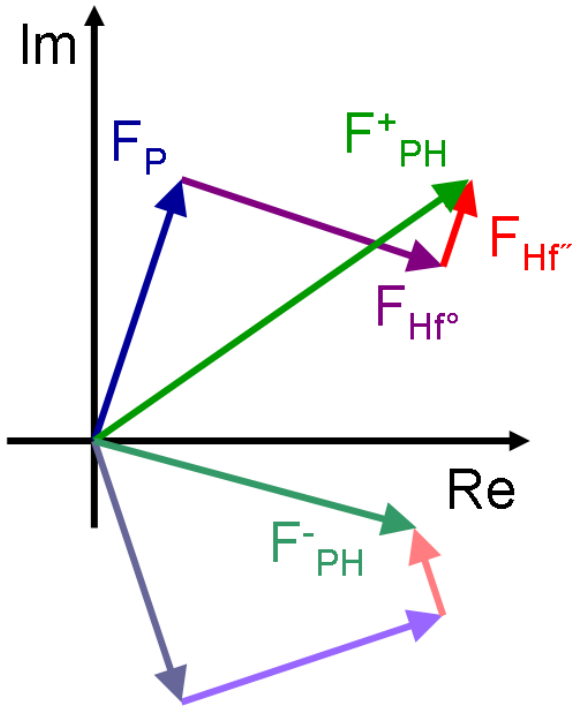
$$f(r^*)_{\lambda} = f'_{\lambda} + if''_{\lambda}$$

Se costruiamo il fattore di struttura

$$\boldsymbol{F^{\lambda}_{hkl} = F_{hkl} + F^{Anom}_{hkl}}$$

Considerando che il termine f'' e sempre perpendicolare al termine f' (dovuto allo sfasamento di $\pi/2$), avremo che i fattori di struttura $\boldsymbol{F^{\lambda}_{hkl}}$ per (h, k, l) e (-h, -k, -l) non sono più uguali (rottura della legge di Friedel).
I fattori di struttura centrosimmentrici vengono indicate come:

$$\boldsymbol{F^{+}_{hkl}; F^{-}_{hkl}}$$



In condizioni di scattering anomalo $\boldsymbol{F^{+}_{hkl}}$ $\boldsymbol{e}$ $\boldsymbol{F^{-}_{hkl}}$saranno in generale diversi tra loro

Utilizzo dello scattering anomalo nella sostituzione isomorfa

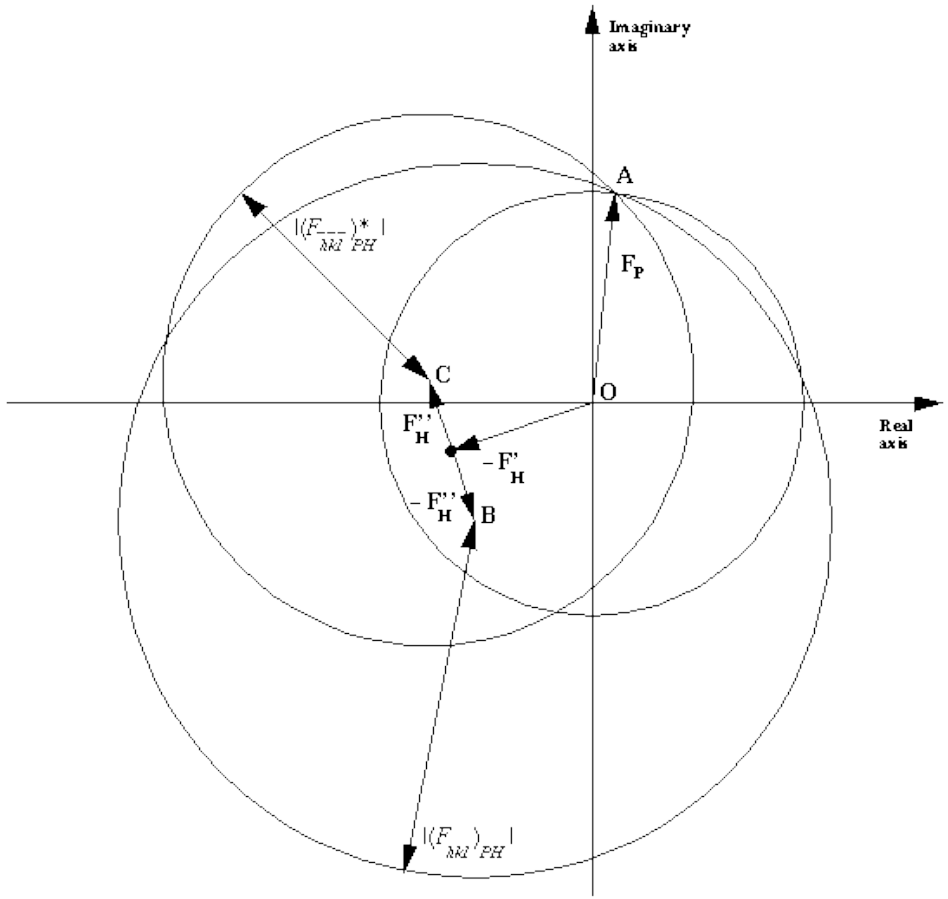
Utilizzo dello Scattering Anomalo: SIRAS

La presenza di un contributo di scattering anomalo può facilitare la risoluzione del problema della fase.

Nel caso di un singolo derivato isomorfo, sappiamo che esiste un'ambiguità nella determinazione della fase.
La presenza del contributo di scattering anomalo da parte dell'atomo pesante permette di risolvere questa ambiguità.

In presenza di scattering anomalo da parte dell'atomo pesante, è come se avessimo due derivati isomorfi invece di uno solo.

Lo scattering anomalo introduce una informazione complementare a quella dovuta al solo atomo pesante



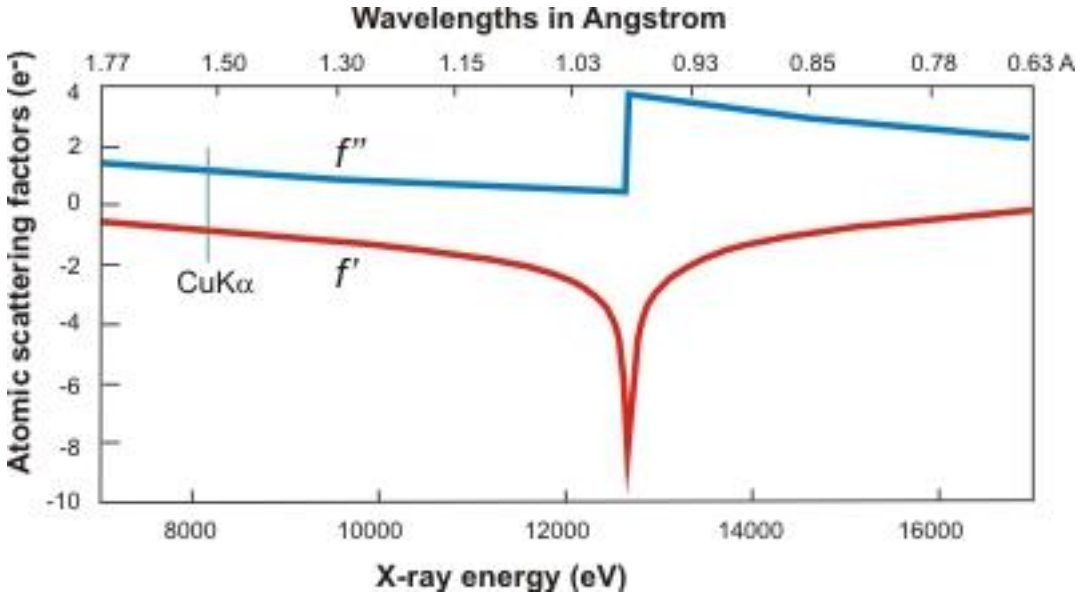
SIR Optimized Anomalous Scattering (SIROAS)

Gli atomi utilizzati per la preparazione dei derivati isomorfi hanno soglie di assorbimento nella regione 0.8-1.2 Å, utilizzando una lunghezza d'onda opportuna (facile per un **sincrotrone**) è possibile sfruttare il fenomeno dello scattering anomalo per la risoluzione del problema della fase.

I valori di f' e f'' dipendono dalla lunghezza d'onda e sono massimi in prossimità della soglia di assorbimento dell'atomo.

Poiché il valore assoluto del contributo anomalo non è molto grande, **è bene scegliere opportunamente la lunghezza d'onda, in modo da massimizzare f' e f''.**

Parliamo quindi di **Single Isomorphous Replacement with Optimized Anomalous Scattering** (**SIROAS**).



Atomo Pesante e Scattering Anomalo

L'impatto di un atomo pesante o del termine di scattering anomalo può essere valutato dai rispettivi valori di f_{at} o $\Delta f'$ e f''

Hg:	$f_{at}(0)$=**80** e;	$\Delta f'$= **-20** e;	f''=**10** e
Br:	$f_{at}(0)$=**35** e;	$\Delta f'$= **-8** e;	f''= **4** e

Il contributo al fattore di struttura apportato da un atomo pesante è nettamente maggiore di quello apportato dal termine di scattering anomalo, che può essere anche piccolo per atomi come Br, Fe etc...

A differenza del fattore atomico di scattering, che decade in funzione di $\sin\vartheta/\lambda$ (angolo di scattering) **$\Delta f'$ e f'' hanno valore costante ad ogni angolo di scattering**, questo è un grande vantaggio dello scattering anomalo, non varia con la risoluzione.

MIROAS

Per quanto il termine di scattering anomalo introduca un'informazione di fase supplementare e la teoria ci dica che l'uso di un singolo derivato isomorfo unito all'utilizzo dello scattering anomalo è sufficiente alla risoluzione del problema della fase, non sempre le cose sono così semplici.

La presenza di errori dovuti a:

- Isomorfismo non perfetto
- Scarsa occupazione da parte dell'atomo pesante
- Disordine locale
- Bassa risoluzione
- Danno da radiazione (particolarmente forte con atomi pesanti e in prossimità di soglie di assorbimento)

Può complicare notevolmente la risoluzione del problema della fase

Di conseguenza può essere necessario l'uso di più derivati isomorfi, anche in presenza di scattering anomalo: MIR con Ottimizzazione dello Scattering Anomalo: MIROAS.

Il termine di scattering anomalo costituisce sempre un'informazione ulteriore e complementare a quella del derivato isomorfo e aumenta quindi le probabilità di determinare le fasi corrette.

Scattering Anomalo: Summary

- Lo scattering anomalo (o diffusione anomala) è un fenomeno che si manifesta quando la lunghezza d'onda dei raggi-X incidenti è molto vicina alla lunghezza d'onda di una soglia di assorbimento di un atomo contenuto nel cristallo
- Il fenomeno dello scattering anomalo modifica il fattore di struttura atomico, introducendo un termine reale e un termine immaginario
- La presenza di un termine immaginario, modifica le intensità diffratte in modo tale che la legge di Friedel, che impone che l'uguaglianza delle intensità diffratte da piani centrosimmetrici. non sia più valida
- Nel caso del metodo del derivato isomorfo, il contributo dello scattering anomalo introduce una informazione complementare a quella introdotta dall'atomo pesante
- Se la misura di diffrazione è eseguita in condizioni di scattering anomalo, ovvero ad una lunghezza d'onda vicina alla soglia di assorbimento di un atomo, combinando le informazioni di fase tra loro complementari derivanti dall'atomo pesante e dallo scattering anomalo, teoricamente posso determinare in modo univoco le fasi dei fattori di struttura con un solo derivato isomorfo (SIROAS)
- Nella pratica sperimentale un solo derivato isomorfo può comunque non bastare e sono richiesti più derivati. Lo scattering anomalo introduce comunque un'informazione complementare che aiuta la risoluzione del problema della fase

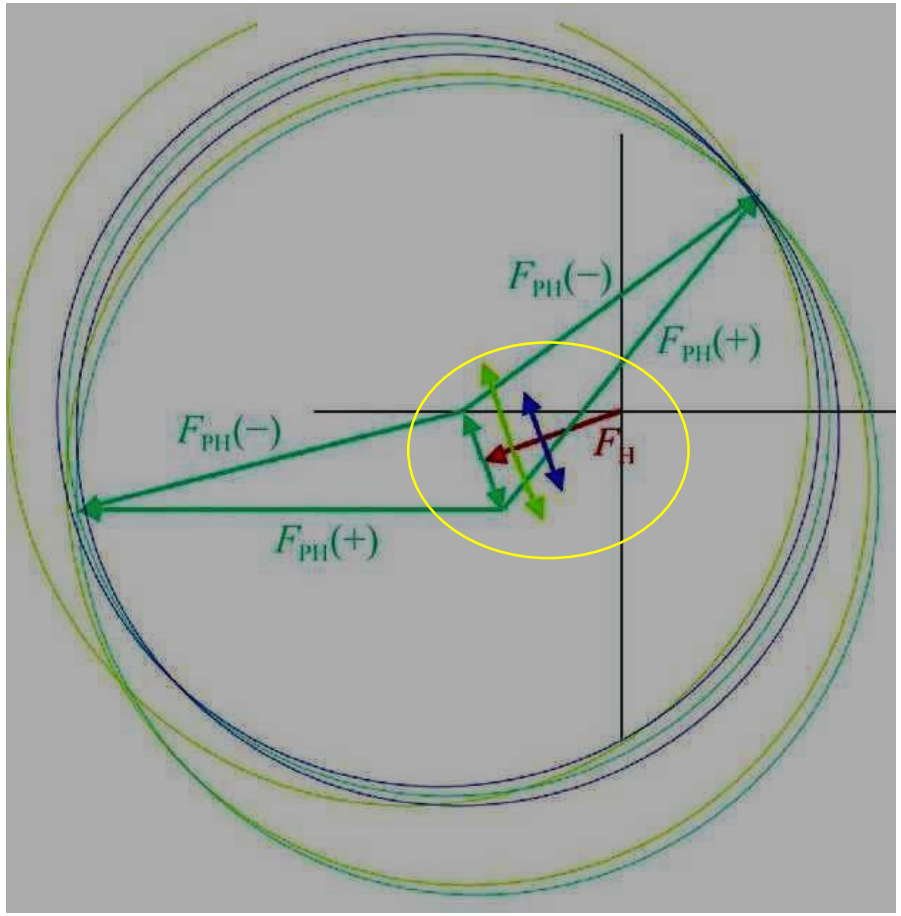
Multiwavelength Anomalous Diffraction

Multiwavelength Anomalous Diffraction (MAD)

Abbiamo visto come l'utilizzo dello scattering anomalo sia equivalente all'introduzione di un ulteriore derivato isomorfo.

Se acquisiamo i dati a lunghezze d'onda differenti e quindi con valori di $\Delta f'$ e f'' differenti, visto che il loro valore dipende dalla lunghezza d'onda, l'effetto complessivo sarà equivalente ad avere un numero N di derivati isomorfi quante sono le N lunghezze d'onda a cui abbiamo acquisito i nostri dati.

Questa è l'idea alla base del **Multiwavelength Anomalous Diffraction** (**MAD**)



MAD – il formalismo

Avendo a disposizione N insiemi di fattori di struttura, acquisiti a lunghezze d'onda diverse, **con diverso contributo di Δf' e f''** , è possibile scrivere un sistema di equazioni dalla cui risoluzione si possono ottenere le fasi della struttura nativa.

$$|^{\lambda}F_{obs}^{\pm}|^2 = |F_T|^2 + a_\lambda|F_A|^2 + b_\lambda|F_T||F_A|\cos(\Phi_T-\Phi_A) \pm c_\lambda|F_T||F_A|\sin(\Phi_T-\Phi_A)$$

In genere vengono utilizzate almeno 3 lunghezze d'onda:

Remota: lunghezza d'onda distante dalla soglia di assorbimento

Flesso: punto di flesso della soglia di assorbimento, massimo valore di f'

Picco: picco della soglia di assorbimento, massimo valore di f''

$$a_\lambda = (f''_\lambda{}^2 + f'_\lambda{}^2)/(f(0))^2,$$
$$b_\lambda = 2f'_\lambda / f(0)$$
$$c_\lambda = 2f''_\lambda/f(0).$$

MAD come caso speciale del MIR

Nella pratica, l'approccio più usato è quello di utilizzare lo stesso formalismo del MIR, ma considerando che ho solo i termini di scattering anomalo; ogni insieme di dati ad una certa lunghezza d'onda equivale ad un insieme di dati da un singolo derivato isomorfo).

$$\Delta F_{dis} = |F_{hkl}|^{\lambda 1} - |F_{hkl}|^{\lambda 2} \quad Differenza\ dispersiva \ (Dispersive\ difference)$$

$$\boldsymbol{\Delta F_{anom}} = |\boldsymbol{F_{hkl}}|^{+} - |\boldsymbol{F_{\overline{hkl}}}|^{-} \quad Differenza\ anomala \ (Anomalous\ difference)$$

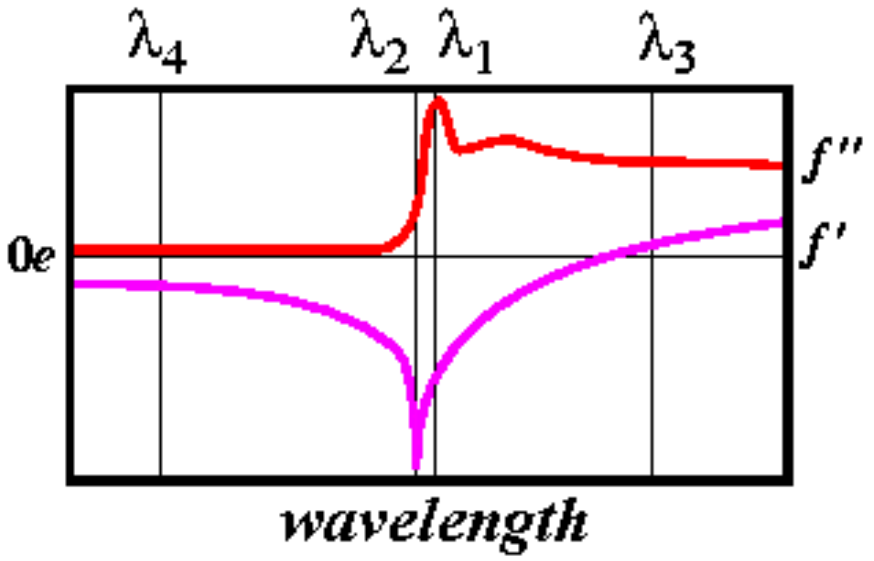
Dal calcolo della funzione di Patterson con le differenze anomale e dispersive, calcolate separatamente e in modo analogo alla funzione di Patterson con le differenze isomorfe, riesco a determinare le posizioni degli atomi che danno luogo allo scattering anomalo (diffusori anomali).

E' usata soprattutto la Patterson calcolata con le differenze anomale perché meno soggetta ad errori sperimentali.

MAD – in partica

In pratica si acquisiscono i dati di diffrazione a 3, massimo 4 lunghezze d'onda distinte.
Queste sono scelte con il seguente criterio:

- Lunghezza d'onda relativa al **picco** dello spettro di assorbimento (corrisponde al **massimo di f''**)
- Lunghezza d'onda relativa al punto di **flesso** dello spettro di assorbimento (corrisponde al **minimo di Δf'**, Δf' è sempre negativo)
- Lunghezza d'onda distante dalle prime due (**remota**) a lunghezza d'onda più lunga
- Eventualmente, lunghezza d'onda distante dalle prime due (remota) a lunghezza d'onda più corta



Tre lunghezze d'onda distinte (lunghezze d'onda remota, picco e flesso) equivalgono ai dati ottenuti dalla proteina nativa e due derivati nel MIR

Vantaggi del MAD

In linea di principio, ed è auspicabile che sia così, **tutti i dati necessari alla risoluzione del problema della fase con il MAD possono essere acquisiti da un singolo cristallo**. **Non ho problemi di isomorfismo!**

Non ho la necessità di preparare più derivati diversi con cristalli diversi.

Lo scattering anomalo non dipende dalla risoluzione per cui il contributo non diminuisce con l'aumentare della risoluzione (ma c'è sempre il moto termico!)

MAD vs MIR

Nel MAD, in linea di principio non ho problemi di non-isomorfismo, quindi gli errori sono inferiori.

La variazione di intensità diffratta dovuta al segnale anomalo è inferiore al segnale introdotto da un atomo pesante, quindi **le misure devono essere più accurate**.

Per poter usare il metodo MAD ho assolutamente bisogno di un sincrotrone e una beamline con lunghezza d'onda variabile.
Posso applicare il metodo MIR anche con una sorgente di laboratorio a lunghezza d'onda fissa.

Applicazione del MAD

Il metodo MAD non richiede una distinzione tra struttura nativa e derivato isomorfo, richiede solo **uno o più atomi che diano luogo al fenomeno dello scattering anomalo nelle condizioni dell'esperimento**.

Proteine che contengono diffusori anomali come atomi costitutivi, come le **metalloproteine**, sono dei naturali candidati per l'uso del MAD

Fe: $\Delta f' = -8$ e; $f'' = 4$ e

L'atomo metallico ha tipicamente occupazione intera, essendo costitutivo della proteina nativa.

Tuttavia il contributo di un singolo atomo di ferro è piccolo, possono essere richiesti dati molto accurati

Nel caso di **oligonucleotidi si utilizzano nucleotidi brominati**

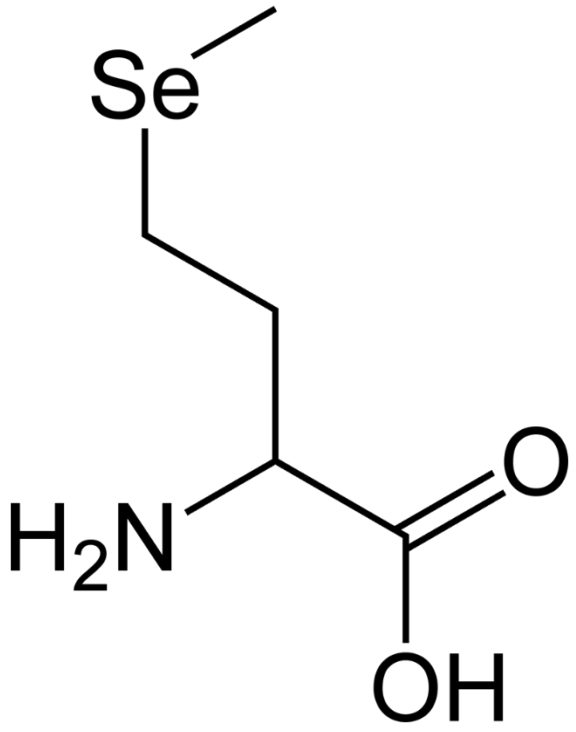
Seleno-Metionina e MAD

Un passo avanti per molti versi rivoluzionario è stata l'introduzione dell'uso della Selenio-Metionina accoppiata con il MAD.

Il Selenio ha una soglia di assorbimento K, vicino a 0.98Å (Δ**f'= -8** e **f''= 4** e).

Con tecniche di Biologia Molecolare è possibile introdurre la Seleno-Metionina in sostituzione della metionina.

In questo modo il diffusore anomalo viene incorporato direttamente nella proteina con una percentuale di occupazione elevata. Utilizzando ceppi di E. coli opportuni si può arrivare ad una occupazione del 100%).



E' consigliata la presenza di una Se-Met ogni ~100 aminoacidi (in media) per riuscire a ottenere delle fasi corrette.

Preparazione di proteine contenti Se-metionina

L'introduzione della Se-Met non altera le proprietà strutturali della proteina; la struttura nativa è conservata anche nella proteina marcata con Selenio.
Può accadere le condizioni di cristallizzazione e la simmetria del cristallo differisca da quella del cristallo nativo (contenente metionina)

L'espressione di proteine con selenometionina in batteri può essre fatta in più modi:

1. Crescita con linee cellulari batteriche standard ma in presenza di L-SeMetionina
2. Utilizzo di particolari linee cellulari batteriche (es: E. coli B834) che utilizzano solo L-SeMetionina

Nel caso 1 ho una resa di proteina più elevata, ma può accadere che non tutti le metionine siano sostiuite da Se-metionine.
Nel caso 2 ho rese più basse ma ho proteina contenenti esclusivamente Se-metionina

E' possibile esprimere proteine contenenti SeMet anche in eucarioti, per quanto più complesso

Impatto dell'uso della Se-Metionina

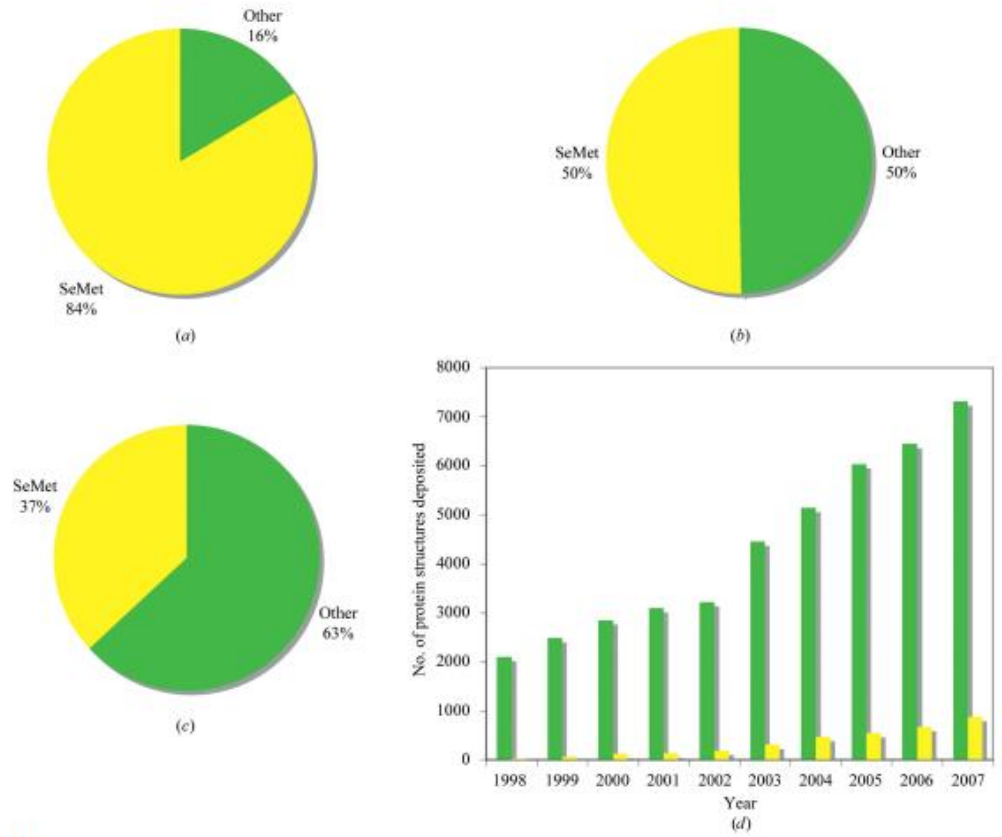


Figure 1
Incidences of SeMet use for phasing in the Protein Data Bank (PDB). (*a*) Pie chart showing the proportion of MAD-phased structures using SeMet (yellow) as a percentage of the total structures phased using MAD. (*b*) Proportion of SAD-phased structures using SeMet (yellow) as a percentage of the total SAD-phased structures. (*c*) Proportion of structural genomics (SGX) derived crystal structures solved using SeMet (yellow) as a percentage of the total number of SGX-derived crystal structures. (*d*) Total number of crystal structures deposited each year (green), with the number solved using SeMet shown in yellow. Numbers were correct at the end of 2008 for (*a*)–(*c*) and the end of 2007 for (*d*).

Walden; Acta Cryst. (2010). D66, 352–357

Localizzazione dei diffusori anomali

Come per il metodo del derivato isomorfo, la localizzazione dei diffusori anomali viene fatta per mezzo della funzione di Patterson: ***funzione di Patterson anomala.***

Si utilizzano come coefficienti, **le differenze tra un fattore di struttura F(hkl) e il suo centrosimmetrico F(-h –k –l)** , talvolta di trovano anche come F^+ e F^-

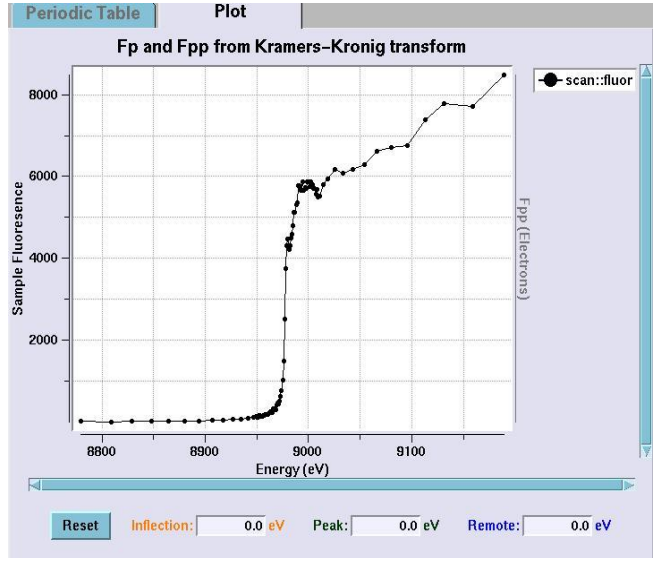
$$P_{anom}(\vec{u}) = \frac{1}{V}\sum_{h,k,l} \Delta F_{ANOM}{}^2 \cos[2\pi(hv + kw + lz)\,]$$

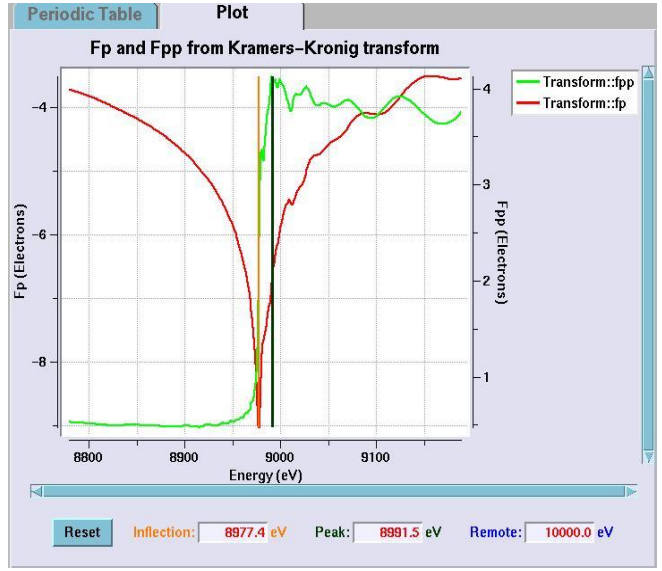
Si può anche utilizzare la funzione di Patterson che usa le differenze tra gli stessi fattori di struttura misurati a lunghezza d'onda diverse, ma in genere è meno accurata

$$P_{disp}(\vec{u}) = \frac{1}{V}\sum_{h,k,l} \Delta F_{disp}{}^2 \cos[2\pi(hv + kw + lz)\,]$$

$$\Delta F_{dis} = |F_{hkl}|^{\lambda 1} - |F_{hkl}|^{\lambda 2} \qquad \Delta F_{anom} = |F_{hkl}|^{+} - |F_{\bar{h}\bar{k}\bar{l}}|^{-}$$

MAD - esperimento





Per prima cosa viene effettuato uno spettro di fluorescenza in corrispondenza della soglia di assorbimento teorico.

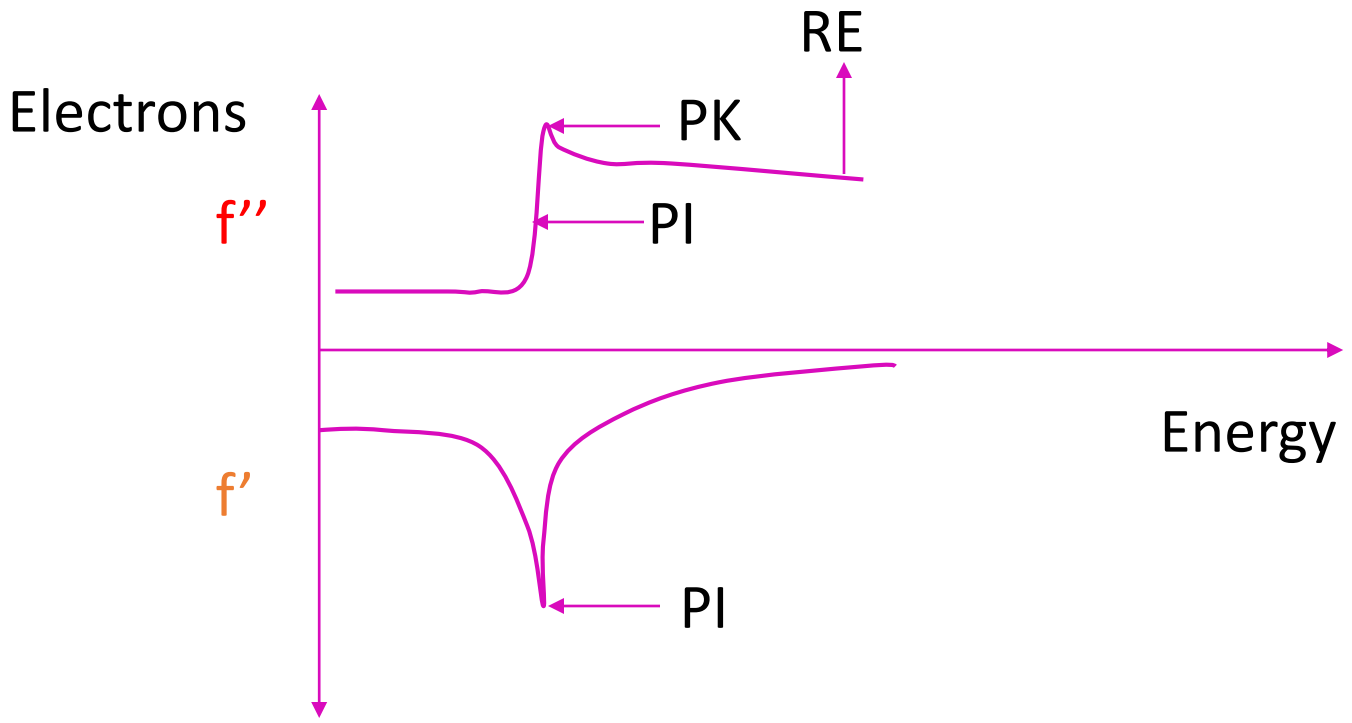
La soglia di assorbimento può essere leggermente diversa da quella prevista teoricamente (tabulata) per via dell'intorno chimico (*chemical shift*) **del diffusore anomalo nella proteina**.

Inoltre possono esserci problemi di taratura del monocromatore.

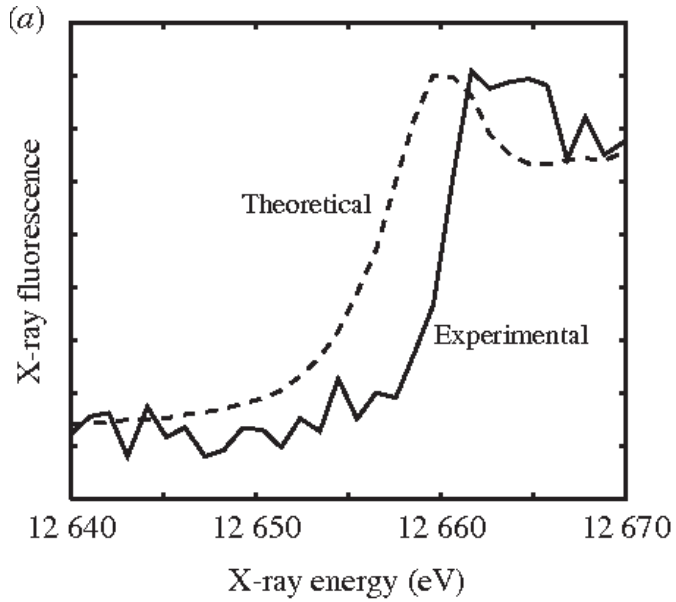
MAD - esperimento

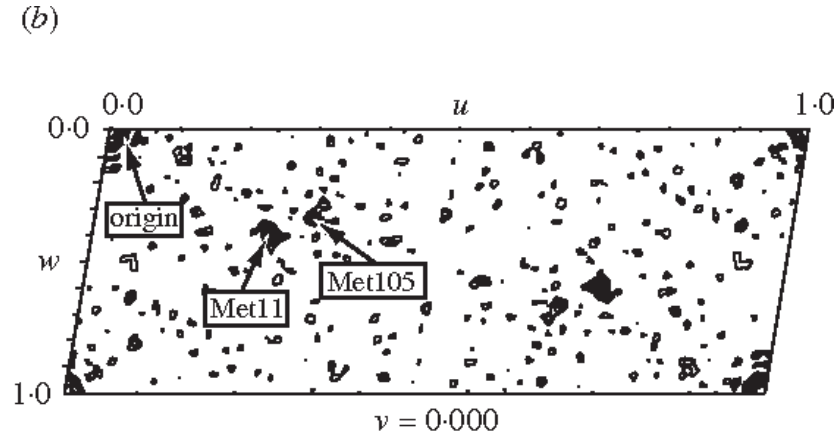
Vengono quindi scelte le lunghezze d'onda opportune a massimizzare $\Delta f'$ e f'' (punti di flesso e picco).
Infine almeno una lunghezza d'onda remota.

Come sempre bisogna tenere conto del **danno da radiazione**, **particolarmente presente in prossimità della soglia di assorbimento**

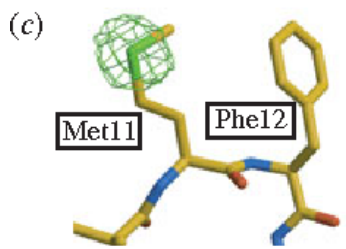


MAD - pratica

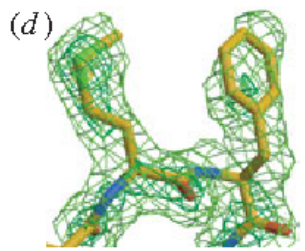




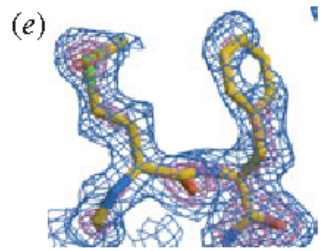
Anomalous difference Patterson



Anomalous difference density



Experimental density



Final refined structure

MAD: Summary

- Nel metodo del Multiwavelength Anomalous Diffraction (MAD) è prevista la presenza di almeno un diffusore anomalo che possa dare luogo al fenomeno dello scattering anomalo.
- Acquisendo dati di diffrazione a più lunghezze d'onda è possibile costruire un sistema di equazioni con cui si può risolvere il problema della fase (metodo analitico)
- In alternativa il MAD può essere trattato come un metodo MIR i dati a lunghezze d'onda diversa sono assimilabili a tanti derivati isomorfi (quante sono le lunghezze d'onda diverse)
- Il metodo MAD non presenta problemi di isomorfismo, quindi ho non ho errori di questo tipo
- Il contributo dello scattering anomalo è debole, ho bisogno di misure più accurate
- Vengono misurati dati di diffrazione ad almeno tre lunghezze d'onda diverse (remota, picco e flesso)
- Se ho metalloproteine posso utilizzare il metallo nativo come diffusore anomalo
- Negli oligonucleotidi posso utilizzare nucleotidi derivatizzati con Bromo
- Nelle proteine posso esprimere molecole contenenti Se-metionine al posto delle metionine, ed eseguire il MAD sul Selenio

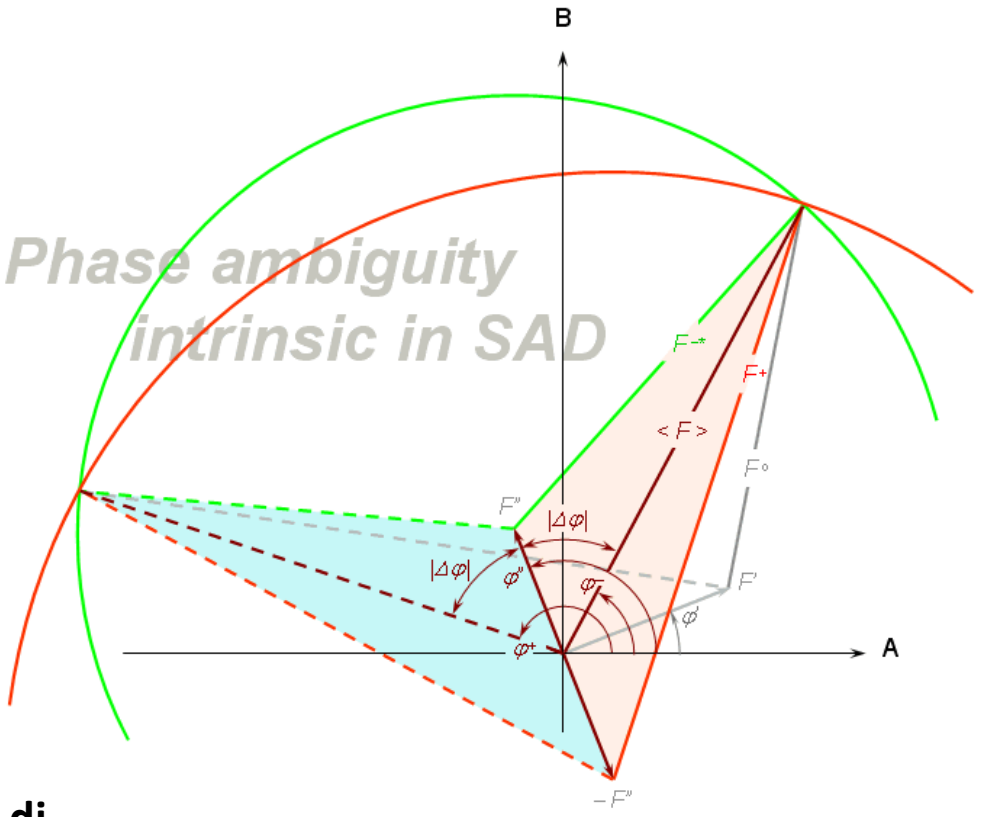
Single Wavelength Anomalous Diffraction (SAD)

Single-Wavelength Anomalous Diffraction (SAD)

Come per il caso del SIR, uso dati ad una sola lunghezza d'onda, purché tale da dar luogo al fenomeno dello scattering anomalo, ma ho comunque una ambiguità sulla fase.

Delle due soluzioni possibili una sola è quella valida.

L'approccio seguito nel Single wavelength Anomalous Diffraction: SAD è proprio quello di acquisire dati ad una sola lunghezza d'onda e di stabilire (per tentativi) a posteriori quale delle due è quella valida.



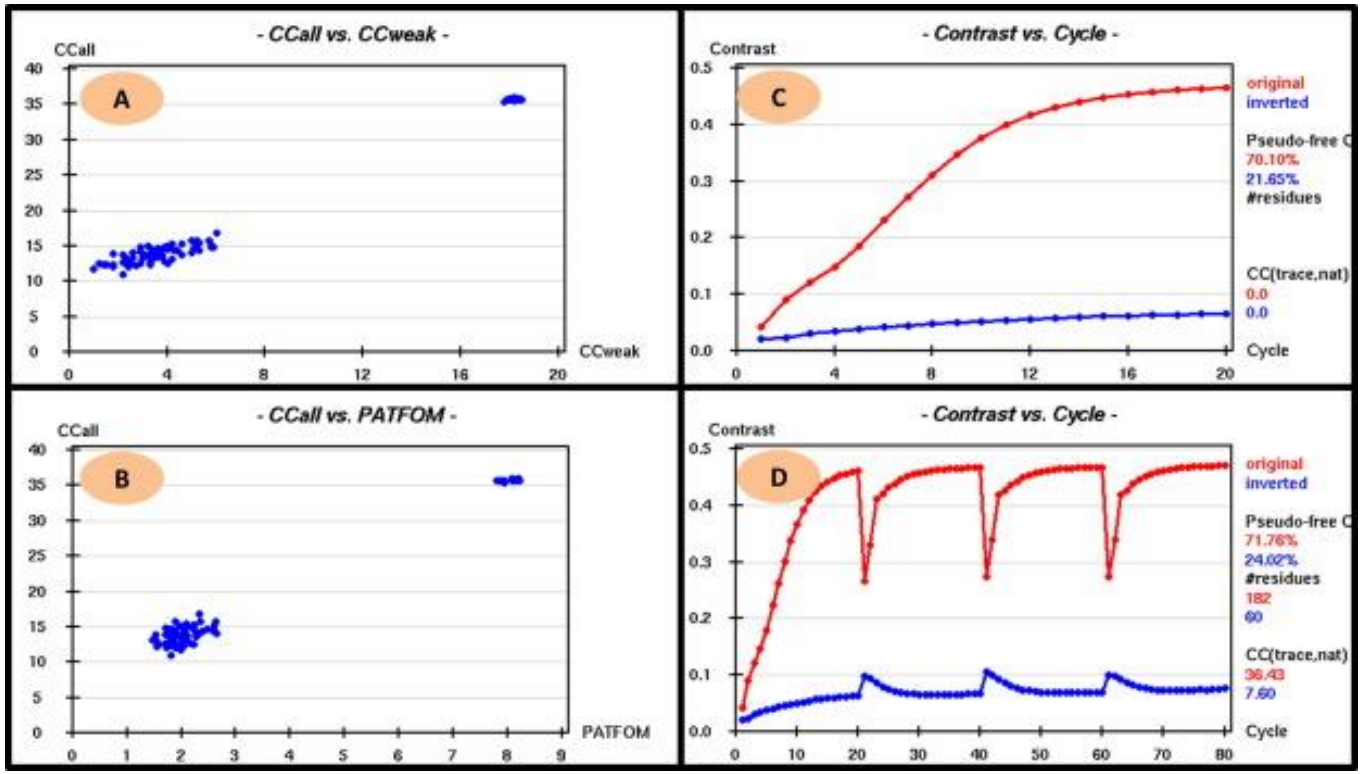
Il SAD contiene intrinsecamente una minor informazione di fase rispetto al MAD

SAD

In pratica le due soluzioni generano due strutture di diffusori anomali (per esempio gli atomi di Selenio) **tra loro enantiomorfi** (centrosimmetrici).

Solo una struttura sarà compatibile con delle fasi corrette e in grado di generare una mappa di densità elettronica e quindi una struttura corretta, ovvero con tutte le caratteristiche di connettività, contrasto dal solvente proprie di una struttura chimicamente corretta.

Queste proprietà emergono dopo ripetuti cicli di ***density modification***



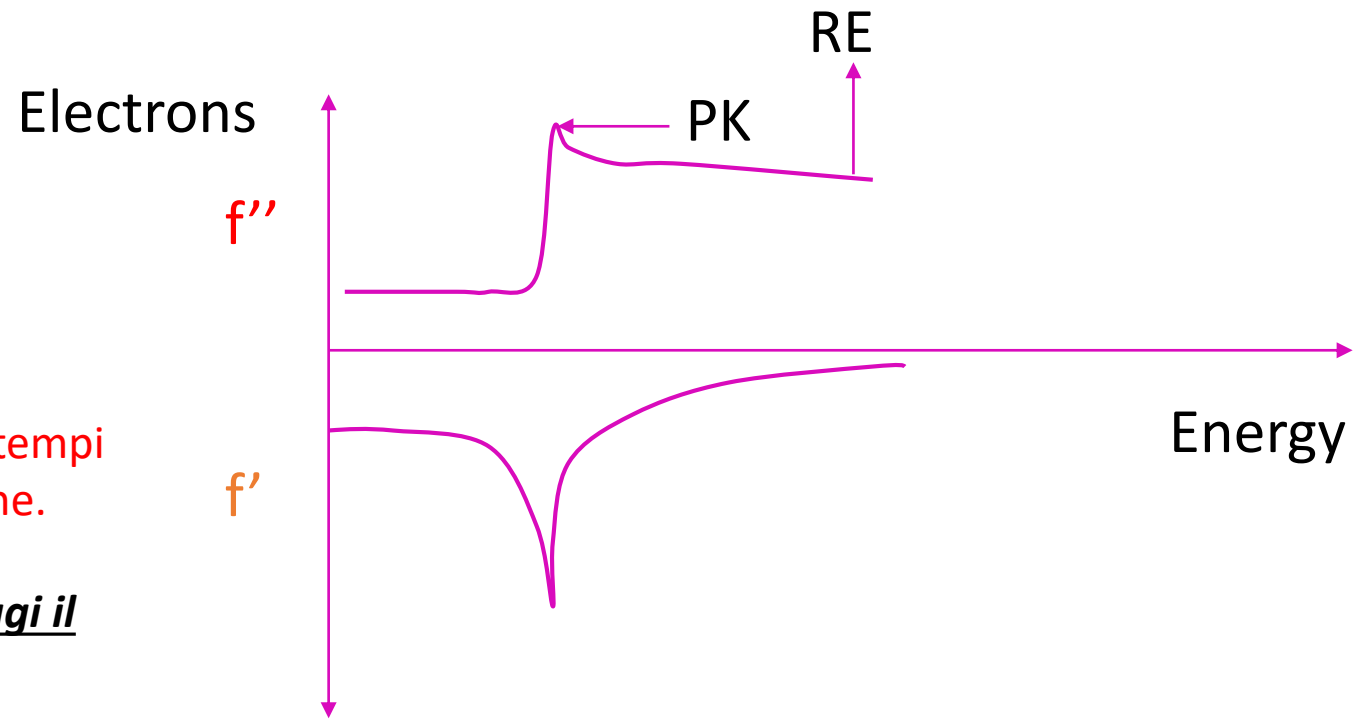
SAD - esperimento

Sperimentalmente il SAD non si differenzia molto da un MAD, ma poiché vengono sfruttate solo le differenze anomale, ovvero quelle tra fattori di struttura di indici di Miller opposti, si utilizza solo la lunghezza d'onda del picco di assorbimento.

In molti casi non si procede nemmeno all'esecuzione dello spettro di assorbimento, ma si utilizza una lunghezza d'onda un poco più corta di quella teorica di assorbimento del diffusore anomalo.

Il vantaggio del SAD è che riduco di molto i tempi di acquisizione e limito il danno da radiazione.

Per la sua semplicità il SAD è al giorno d'oggi il metodo più utilizzato per la risoluzione del problema della fase



Diffusori anomali per il SAD

Qualsiasi diffusore anomalo incorporabile nel cristallo è potenzialmente utilizzabile per il SAD.

Ad esempio si possono 'pressurizzare' i cristalli in un'atmosfera di Xe, i cui atomi possono occupare cavità idrofobiche nella proteina.

Un altro metodo molto usato è quello di **immergere brevemente i cristalli (*soaking*) in una soluzione di alogenuri di NaI o NaBr**. Gli alogenuri possono coordinare con gruppi carichi positivamente.

A differenza delle proteine conteneneti Se-Met, in cui conosco con precisione quanti diffusori anomali sono presenti (circa), nel caso di soaking con alogenuri il numero è assolutamente incognito.

In ogni caso si può fare ricorso a proteine contenenti selenometionine.

S-SAD

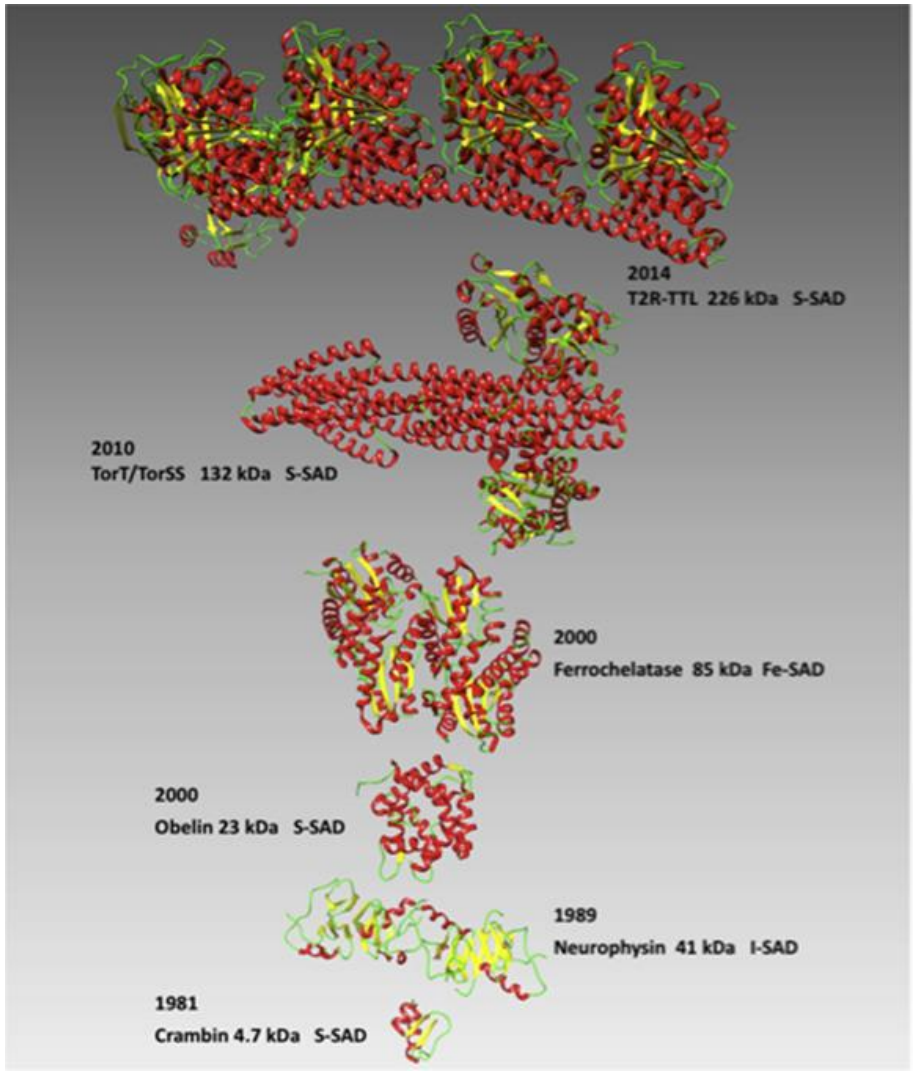
Lo zolfo è naturalmente presente nelle proteine (cisteine, metionine) ed è quindi potenzialmente un diffusore anomalo molto attraente, tuttavia ha una soglia di assorbimento a circa 5 Å.
Però a lunghezze d'onda comprese tra 1.5 e 2.0 Å, il valore di f'' è diverso da zero anche se molto piccolo.

S (1.5 Å): $\Delta f' = -0.38$ e $f'' = 0.53$ e

Acquisendo dati **MOLTO** accurati, è possibile applicare il metodo SAD (S-SAD) sfruttando gli atomi di zolfo nativi.
L'accuratezza richiesta si ottiene:

- Aumentando la ridondanza dei dati (20-30 volte!)
- Limitando il più possibile il background nella misura

Nel caso di nucleotidi si può utilizzare il **fosforo** nativo come diffusore anomalo



Sostituzione Molecolare

Molecular Replacement

Il Molecular Replacement è un metodo utilizzabile laddove esista una proteina di struttura sperimentalmente nota e sufficientemente omologa alla proteina di cui vogliamo determinare la struttura

Le uniche informazioni necessarie sono

- **I fattori di struttura della proteina studiata (target)**
- **Le coordinate della proteina modello (model)**

E' un metodo di risoluzione della fase molto attraente in quanto non richiede ulteriori esperimenti, a differenza del SIR/MIR e del SAD/MAD.

E' chiaramente la prima opzione da provare nel processo di determinazione strutturale

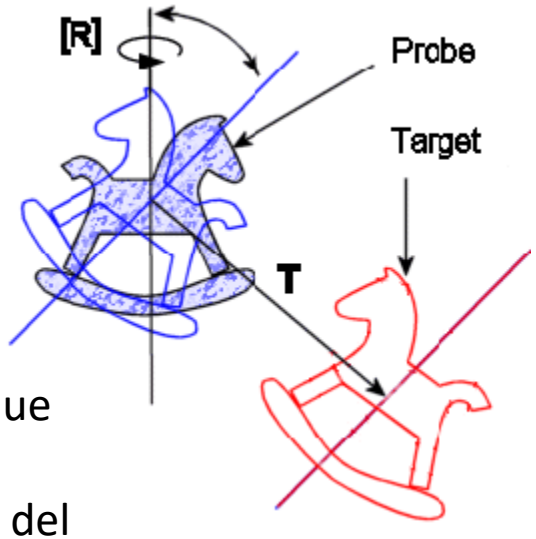
Molecular Replacement

Concettualmente il Molecular Replacement è molto semplice e parte dal presupposto che la nostra proteina target sia posizionata in qualche punto (non noto a priori) della cella unitaria.

La proteina modello, che supponiamo essere simile se non uguale alla proteina target, viene posizionata e orientata in vari modi nella cella unitaria, per ogni posizione e orientazione vengono calcolati i fattori di struttura e confrontati con i valori sperimentali.

La soluzione sarà ottenuta per quella posizione del modello che da il migliore accordo tra dati calcolati e dati osservati.

$$F_H = \sum_{j=1}^{N} f_j(r^*)\, exp[2\pi i(hx_j + ky_j + lz_j)]$$



Per quanto concettualmente semplice la soluzione può essere complessa e comunque computazionalmente intensiva poiché si tratta di determinare **3 angoli di rotazione** (corretta orientazione del modello) e **un vettore di traslazione** (le 3 coordinate x,y,z del centro di massa del modello). Si tratta di un problema numerico a 6 dimensioni!

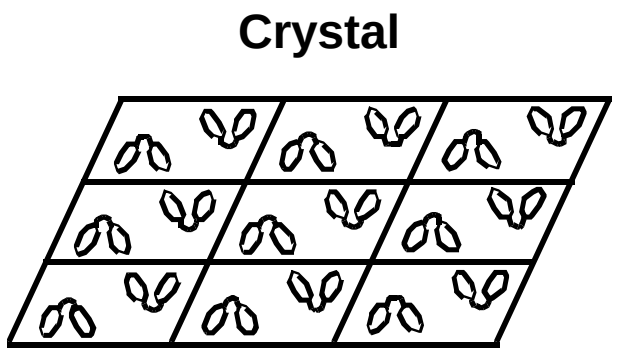
Molecular Replacement

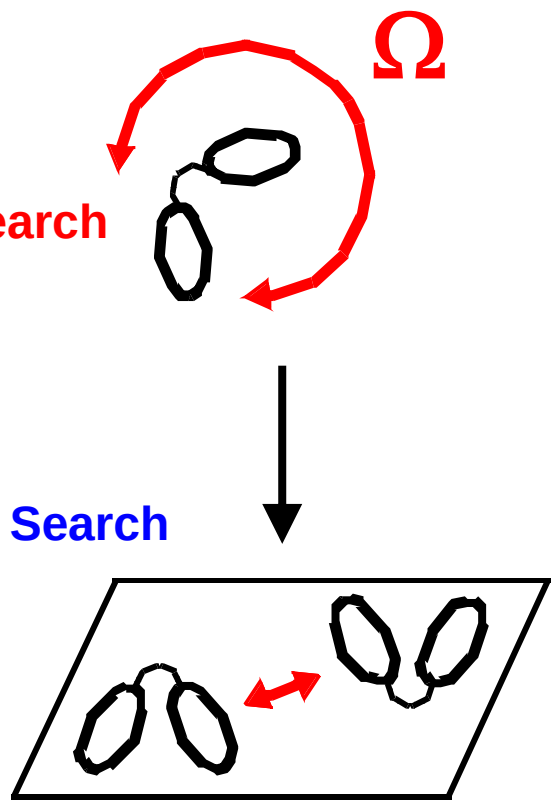
Una prima grande semplificazione deriva dal fatto che il problema a 6 dimensioni può essere semplificato dividendo il problema in una parte **Rotazionale** e una **Traslazionale** (due problemi a 3 dimensioni).
Matematicamente

$$P(x_f, y_f, z_f) = R(\Omega)M(x_0, y_0, z_0) + T(X, Y, Z)$$

Modello

Si tratta di trovare gli **angoli di rotazione** e **il vettore di traslazione** che permettano di portare il nostro modello nella posizione corretta.





Funzione Rotazionale

Se sono note le coordinate di una molecola è sempre possibile calcolarne la funzione di Patterson. La funzione di Patterson dipende solo dai vettori interatomici e non dalle coordinate assolute.
D'altra parte la funzione di Patterson può essere calcolata anche con le sole intensità.

$$P(\vec{u}) = \frac{1}{V}\int_o^a \rho(\vec{r})\rho(\vec{r}+\vec{u})\,dx; \qquad P(\vec{u}) = \frac{1}{V}\sum_{h,k,l=0}^{\infty} |F_h|^2 \cos[2\pi(hv+kw+lz)\,];$$

La funzione di Patterson non dipende dalle posizioni assolute degli atomi, ma solo dalle distanze tra atomi

La **funzione rotazionale** è così definita:

$$Rot(M) = \frac{1}{V}\int_{rmin}^{rMax} P(r)_{target} P[Rot(\Omega)M(x_0, y_0, z_0)]_{model} d^3r$$

Dove P sono le funzioni di Patterson e R definisce la rotazione

Voglimo determinare il valore degli angoli Ω per cui Rot(M) è massimo, che indica il migliore accordo tra la funzione di Patterson calcolata con le coordinate del modello e funzione di Patterson calcolata con i dai dati sperimentali. Il valore massimo di Rot(M) indica l'orientazione corretta del modello.

Funzione Traslazionale

La funzione traslazionale non è molto diversa, solo che viene cercata una traslazione e non una rotazione

$$Tra(M) = \frac{1}{V}\int P(r)_{target} P(r-T)_{model} d^3r$$

Dove **T è un opportuno vettore di traslazione**

Le condizioni per cui Tra(M) è massimo indicano il vettore di traslazione T necessario a posizionare correttamente il modello correttamente orientato a partire dalla funzione rotazionale .

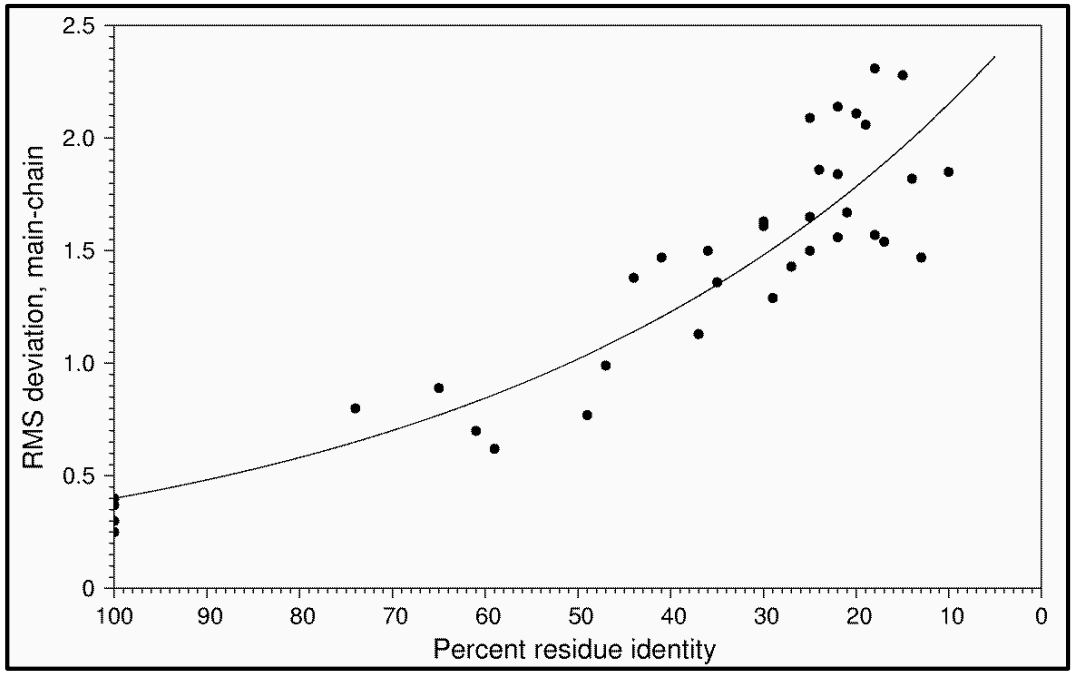
Una volta determinate la rotazione e la traslazione la soluzione ottenuta **deve essere validata**, ovvero va verificata **la compatibilità con la simmetria del cristallo**.

Questa compatibilità è assicurata dal calcolo di una **funzione di Packing**, che ci assicura che la soluzione non generi sovrapposizioni (clash) tra molecule, nel cristallo.

Sostituzione Molecolare in pratica

Omologia e rmsd

Perché il modello scelto possa funzionare per il Molecular Replacement, deve essere **strutturalmente simile alla proteina target.**



La similitudine, espressa dalla r.m.s.d sulle posizioni dei Cα delle proteine sovrapposte, **è legata alla omologia delle proteine** (identità di sequenza)

Una relazione empirica, stabilisce una relazione

$$\boldsymbol{rmsd = 0.4\exp[1.87(1 - s)]}$$

Dove la r.m.s.d. è espressa in Å e s è l'identità di sequenza

Limiti del Molecular Replacement

Una regola empiricamente accettata è che il Molecular Replacement fallisce quando la rmsd tra proteina modello e proteina target, opportunamente sovrapposte, è superiore a circa 1.5 Å, ovvero quando l'identità di sequenza è inferiore al 30%

In generale valgono le seguenti indicazioni (empiriche):

- Identità > 40% Quasi sicuramente il Molecular Replacement funzionerà
- Identità 30-40% Probabilmente il Molecular Replacement funziona, ma in alcuni casi la soluzione presenta difficoltà
- Identità 20-30% Difficile che il Molecular Replacement funzioni, ma potrebbe…
- Identità < 20% Quasi sicuramente il Molecular Replacement non funziona (ma provare è gratis!!)

Nei casi in cui l'identità non sia molto alta, si possono provare vari approcci, rivolti essenzialmente ad eliminare dal modello le parti della molecola che possono 'interferire' nel corso della soluzione, ovvero parti del modello potenzialmente strutturalmente diverse nel target.

Modifica del modello

Al fine di eliminare le parti del modello interferenti si possono provare diversi approcci più o meno sofisticati:

1) **Eliminare tutte le catene laterali**. Il modello è costituito solo dal backbone (poly-Alanina)

2) **Eliminare** dal modello le parti **probabilmente flessibili**, come i loops

3) **Eliminare dal modello le parti non comuni della sequenza** (Model Trimming)

4) **Effettuare un vero e proprio morphing** del modello, con un approccio molto simile a quello dell'homology modelling, sulla base di più modelli con diverso grado di similitudine con la proteina target.

Molecular Replacement: Summary

- Il metodo del Molecular Replacement (Sostituzione Molecolare) prevede l'esistenza di una molecola simile (modello) la cui struttura molecolare sia sperimentalmente nota (target)
- Il Molecular Replacement non richiede ulteriori esperimenti di diffrazione
- La molecola modello viene posizionata (ruotando e traslando) nella cella unitaria e le intensità calcolate vengono confrontate con le intensità sperimentali
- Poiché la posizione del modello è definita da 6 gradi di libertà (3 rotazionali e 3 traslazionali) posso dividere il problema in due: una parte rotazionale (3 angoli) e una parte traslazionale (3 coordinate)
- I valori degli angoli e delle coordinate sono determinati facendo ricorso alla funzione di Patterson
- La posizione della molecola modello deve essere alla fine coerente con l'impacchettamento del cristallo
- Il successo del metodo del molecular replacement dipende dalla similitudine tra struttura modello e struttura target
- Se l'identità tra le due molecole (modello e target) è inferiore al 30%, difficilmente il Molecular Replacement avrà successo
- Per eliminare possibili interferenze, posso modificare il modello (catene laterali, parti flessibili, parti non comuni tra modello e target)

Metodi Diretti

I metodi diretti sono usati per determinare le fase a partire dai soli Fattori di Struttura e dalla conoscenza del contenuto della cella unitaria (Peso Formula della cella unitaria), per mezzo di relazioni matematiche.

I metodi diretti fanno largo uso delle proprietà statistiche dei fattori di struttura e della teoria della probabilità. Nei metodi diretti sono stabilite delle fasi, la cui probabilità di essere valide è legata alle intensità diffratte.
Ciò è possibile partendo dai seguenti assunti:

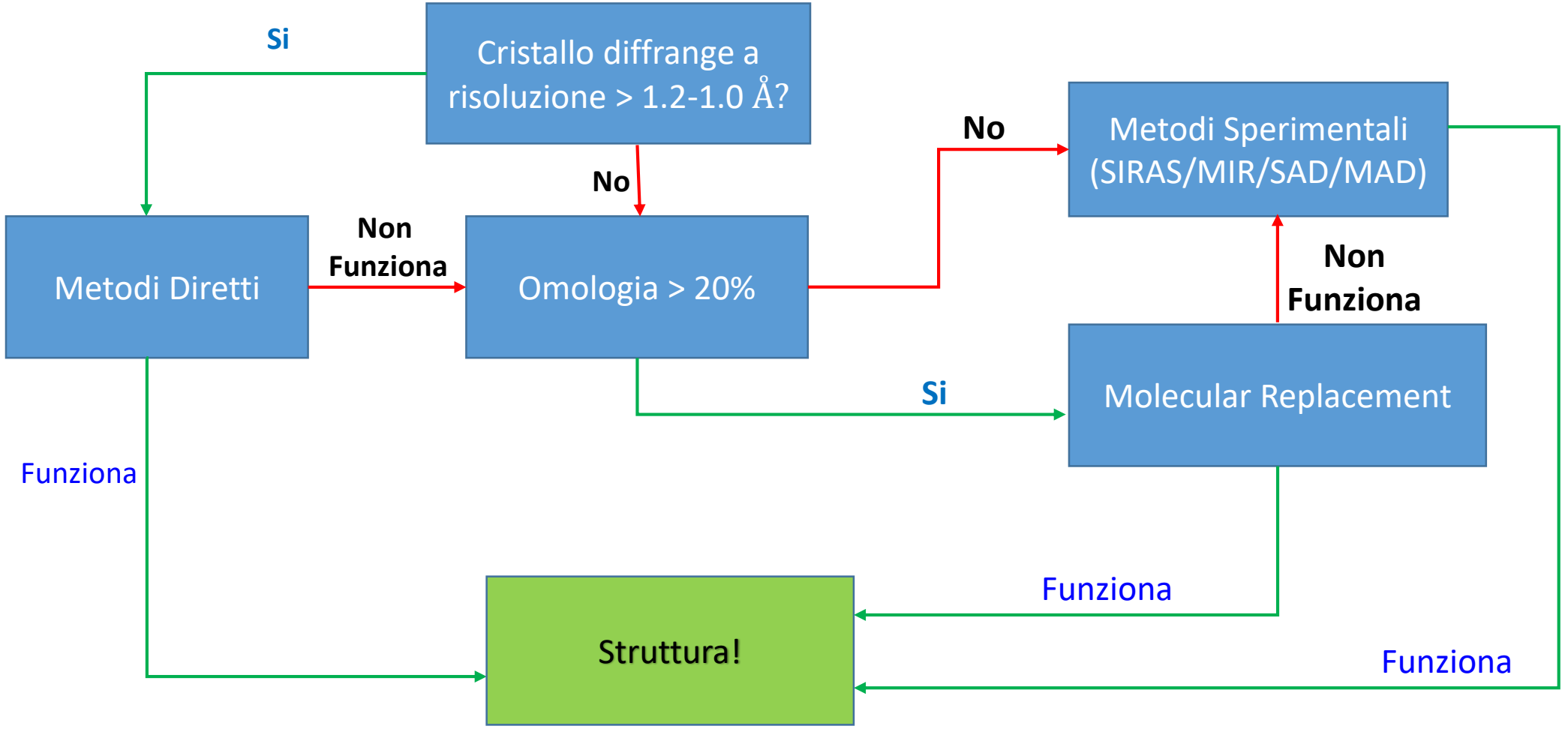
- La $\rho(r)$ è ovunque positiva
- Alla $\rho(r)$ contribuiscono atomi discreti (atomicità)

I Metodi Diretti funzionano bene solo per strutture ad alta risoluzione (d < 1.2 Å) e con un numero di atomi non troppo elevato, due condizioni raramente incontrate nelle strutture macromolecolari.

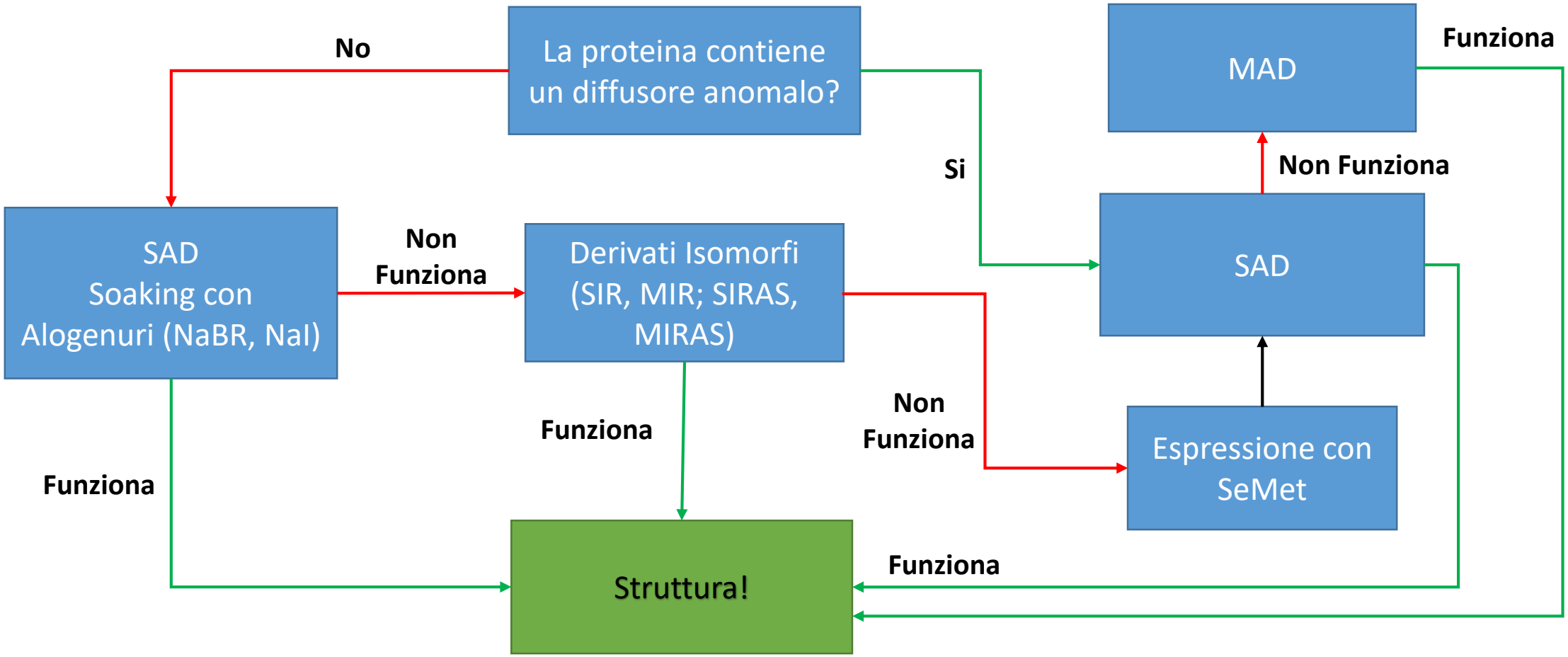
I metodi diretti trovano scarso impiego nella determinazione strutturale delle proteine, solo in pochi casi ad altissima risoluzione e per proteine piccole. Al contrario, i metodi diretti sono oramai pressoché l'unico metodo utilizzato nella determinazione strutturale delle molecole organiche (small molecules).
In alcuni casi sono utilizzati nel procedimento di determinazione della fase degli 'atomi pesanti' nei metodi sperimentali (MIR/SIR e MAD/SAD; localizzazione degli 'atomi pesanti').

Problema della fase in pratica

Risoluzione del problema della fase in pratica

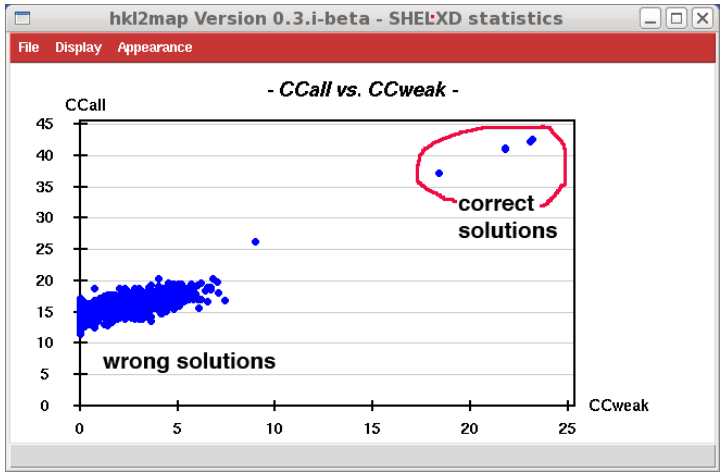


Risoluzione sperimentale del problema della fase

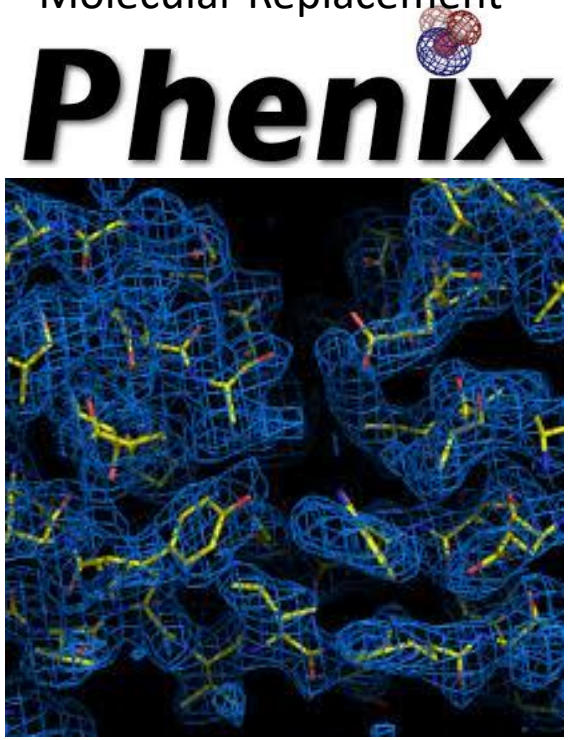


Software

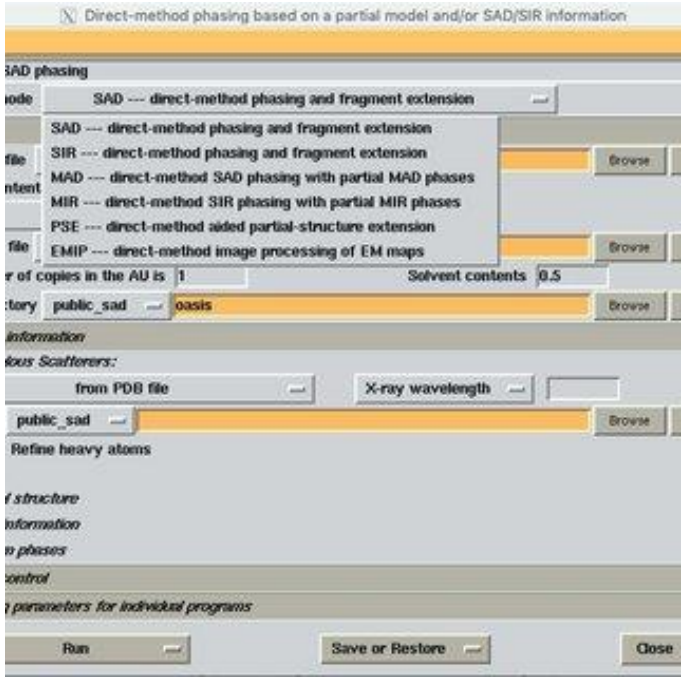
SHELX-C / **SHELX-D**
(SIRAS, MIR, SAD,MAD)



Phenix
SIRAS, MIR, SAD, MAD,
Molecular Replacement



CCP4
SIRAS, MIR, SAD, MAD,
Molecular Replacement



Phasing method	Phasing marker	Remarks	Available data	Marker	Difference data	Data to be phased
SAD via sulfur atoms (S-SAD)	S in Met, Cys residues	Highly redundant data collection, must be combined with density modification	Anomalous pairs F_{PS}^{+}, F_{PS}^{-} also serve as native data	Sulfur positions, $\mathbf{F}_S$	ΔF_{ano} from $\lvert F_{PS}^{+} - F_{PS}^{-} \rvert$	Merged F_{PS}
SAD via naturally bound metals	Naturally bound metal ion, cofactor	Requires density modification for resolution of phase ambiguity	Anomalous pairs F_{PA}^{+}, F_{PA}^{-} also serve as native F_P	Anomalous scatterer positions, $\mathbf{F}_A$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$	Merged F_{PA}
SIR(AS) via isomorphous metals	Heavy atom ion, specifically bound anions Br^-, I^-, I^{3-}, also Xe	Isomorphous phasing power proportional to $z_{(H)}$, anomalous signal or density modification needed to break phase ambiguity	Native data F_P, isomorphous data F_{PA}, in pairs F_{PA}^{+}, F_{PA}^{-} for SIRAS	Isomorphous/ anomalous scatterer positions, $\mathbf{F}_A$	ΔF_{iso} from $\lvert F_{PA} - F_P \rvert$ and ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$	Native F_P
MIR(AS) via isomorphous metals	Heavy atom ions, clusters, specifically bound anions Br^-, I^-, I^{3-}	As above, except multiple derivatives or anomalous signal break phase ambiguity. Hg, Pt, Au, etc. phase several hundred residues, heavy atom clusters more.	Native data F_P, isomorphous data $n\cdot(F_{PA})$ in pairs $n\cdot(F_{PA}^{+}, F_{PA}^{-})$ for MIRAS	Isomorphous/ anomalous scatterer positions, $n\cdot(\mathbf{F}_A)$	$n\cdot\Delta F_{iso}$ from $n\cdot\lvert F_{PA} - F_P \rvert$, and ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$ pairs	Native F_P
MAD via Se	Se in Se-Met residues	1 Se phases 100–200 residues, introduced by expression host	Bijvoet pairs at n wavelengths, $n\cdot(F_{PA}^{+}, F_{PA}^{-})_{\lambda n}$, optional native data	Anomalous scatterer positions $\mathbf{F}_{Se}$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert_{\lambda}$ pairs ΔF_{λ} from $\lvert F_{\lambda i} - F_{\lambda j} \rvert$ pairs	Best merged data F_{PA}, optional native F_P
MAD via isomorphous metals	Heavy atom, specifically bound	Strong signal on XAS "white lines," particularly at L-edges, can phase several hundred residues	Bijvoet pairs at n wavelengths $n\cdot(F_{PA}^{+}, F_{PA}^{-})_{\lambda n}$, native data F_P (not needed for phasing)	Anomalous scatterer positions $\mathbf{F}_A$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert_{\lambda}$ pairs ΔF_{λ} from $\lvert F_{\lambda i} - F_{\lambda j} \rvert$ pairs	Native F_P, optional best merged F_{PA}
Direct methods	None	Near atomic resolution data (1.2 Å or better), relatively small proteins	F_P	All non-H atom positions *ab initio*	N/A	Native F_P
Density modification	None	Needs multiple copies of motif in asymmetric unit for *ab initio* phasing	F_P	Multiple copies of a subunit	N/A	Native F_P
MR via model structure	Positioned search model	Needs search model with structural similarity, subject to model bias, particularly at low resolution	Native data F_P and search model structure factors F_C	Entire model serves as search probe	N/A	Native F_P