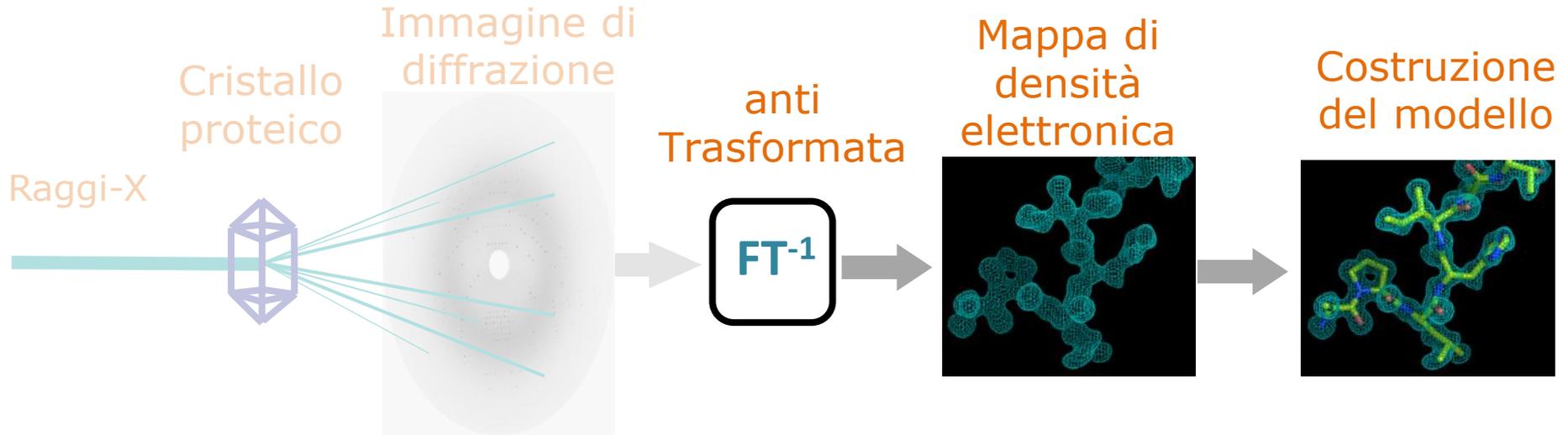


Lezione 5  
Andrea Ilari

IL PROBLEMA DELLA FASE

IBPM-CNR

# L'esperimento di diffrazione



Expression  
Purification  
Crystallization



X-ray diffraction  
(data collection)

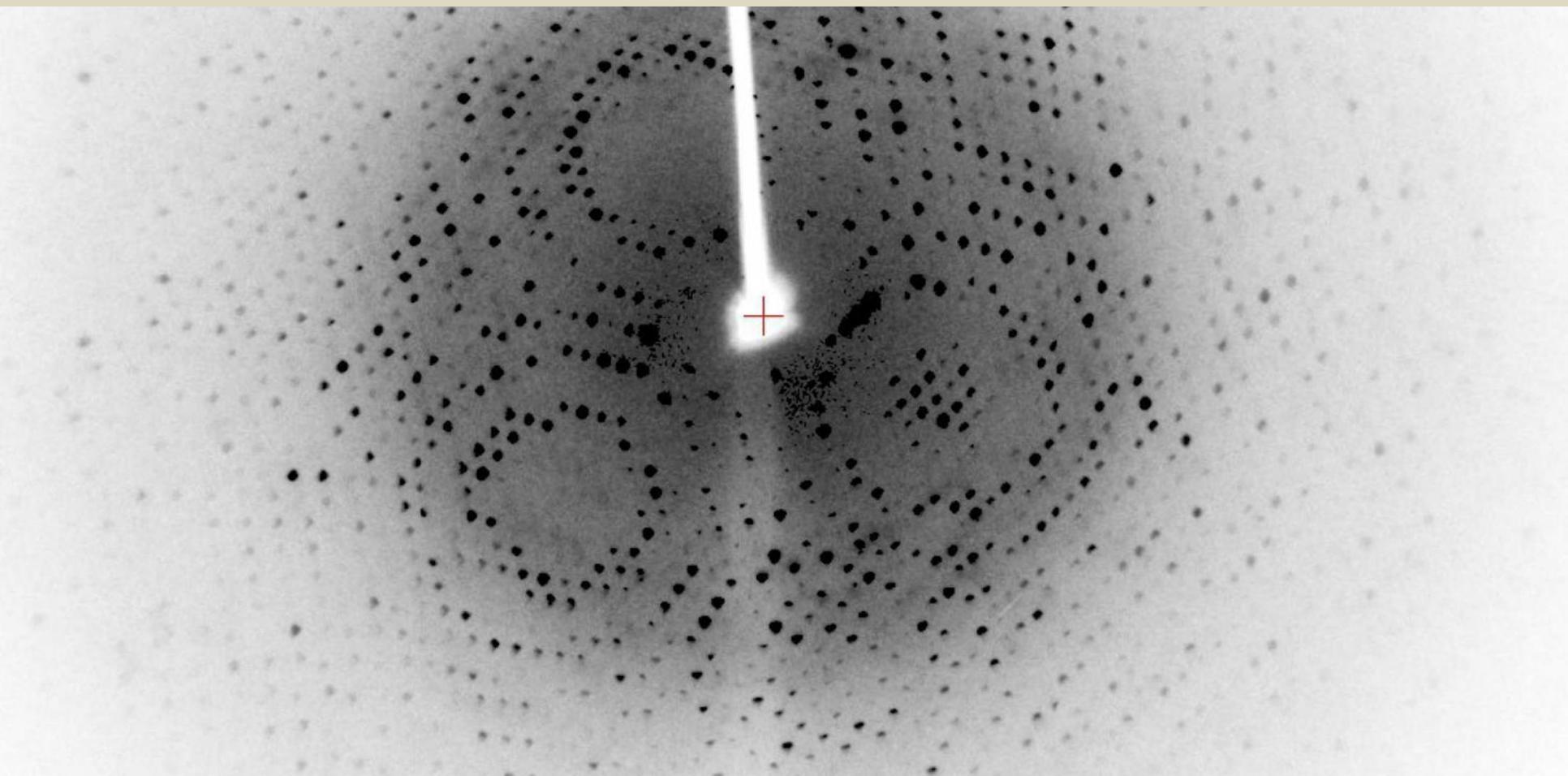


Phasing and model building



# ANALISI DEI DATI

Indicizzazione – Integrazione - Scalaggio



Con la **RACCOLTA DATI** i programmi di analisi **fanno 3 cose:**

## INDICIZZAZIONE

Analizzo le prime immagini di diffrazione:

- misurano **le posizioni** dei picchi



- ricavano **il gruppo spaziale**

## INTEGRAZIONE

Analizzo tutte le immagini di diffrazione:

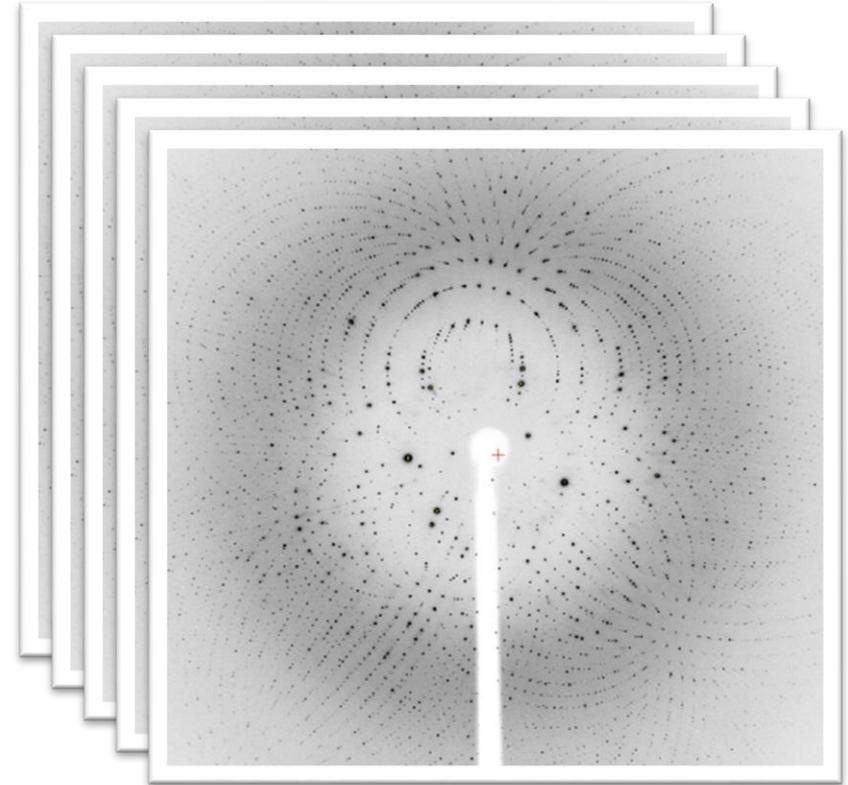
- misurano le **Intensità di ogni picco (spot)**

## SCALAGGIO

Unisco tutti i dati ottenuti:

- scrivono in un **output le Intensità relative ad ogni posizione nello spazio reciproco (spot) a cui assegnano gli indici  $h, k, l$**

- **R merge**
- **mosaicità** (disordine intrinseco)
- **completezza**, ridondanza
- **$I/\sigma(I)$**  (segnale/rumore)



## RACCOLTA DATI:

Una serie di immagini di diffrazione dei raggi-X ottenute facendo ruotare il cristallo

# INDICIZZAZIONE (indexing)

## 1. INDICIZZAZIONE

Analizzo le prime immagini di diffrazione:

- misurano **le posizioni** dei picchi → ricavano **il gruppo spaziale e i parametri della cella**

iMosflm Nov 2013 (using Mosflm 7.1.0)

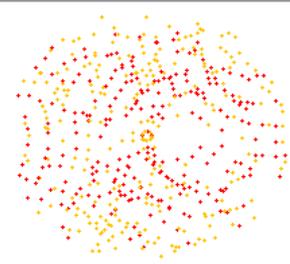
Session Settings Help

119.55 120.09 250.00 5.00 10.0 1.05 1.05 0.00 20 278 2.50

### Autoindexing

hg\_###.mar1600 :1, 84

Image	$\phi$ range	Auto	Man	Del	> I/ $\sigma$ (I)	Find	Use
1	0.00 - 1.00	294	0	0	148	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
84	83.00 - 84.00	222	0	0	105	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Total</b>		<b>516</b>	<b>0</b>	<b>0</b>	<b>253</b>		

Index 

### Lattice 1

Solution	Lat.	Pen.	a	b	c	$\alpha$	$\beta$	$\gamma$	$\sigma$ (xy)	$\sigma$ ( $\phi$ )	$\delta$ beam
1 (ref)	aP	0	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.44 ( 0.1)
2 (ref)	aP	0	58.5	58.6	62.1	61.9	62.0	60.0	0.21	0.35	0.44 ( 0.1)
3 (ref)	mC	1	101.3	58.6	62.2	90.0	123.0	90.0	0.22	0.34	0.44 ( 0.2)
4 (ref)	mC	1	101.5	58.5	62.1	90.0	122.9	90.0	0.20	0.39	0.45 ( 0.2)
5 (ref)	mC	2	101.5	58.5	62.1	90.0	122.9	90.0	0.20	0.36	0.45 ( 0.2)
6 (ref)	hR	2	58.6	58.6	156.4	90.0	90.0	120.0	0.21	0.54	0.46 ( 0.3)
7 (ref)	mC	2	101.5	58.5	62.1	90.0	122.9	90.0	0.20	0.36	0.45 ( 0.2)
8 (reg)	hR	60	60.4	60.4	144.9	90.0	90.0	120.0	-	-	-
9 (reg)	mC	60	103.5	62.2	58.5	90.0	124.4	90.0	-	-	-
10 (reg)	mC	60	85.3	85.4	58.5	90.0	133.2	90.0	-	-	-
11 (reg)	oI	60	58.6	62.2	85.3	90.0	90.0	90.0	-	-	-
12 (reg)	mC	61	101.4	58.6	62.1	90.0	122.9	90.0	-	-	-
13 (reg)	tI	61	60.4	60.4	85.3	90.0	90.0	90.0	-	-	-
14 (reg)	oI	61	58.5	62.2	85.4	90.0	90.0	90.0	-	-	-
15 (reg)	mC	61	101.5	58.5	62.2	90.0	122.9	90.0	-	-	-

Show lattices summary [+]

Spacegroup: h3

Mosaicity: 0.00 Estimate

Search beam-centre [+]

Green warnings 1

Cella più probabile

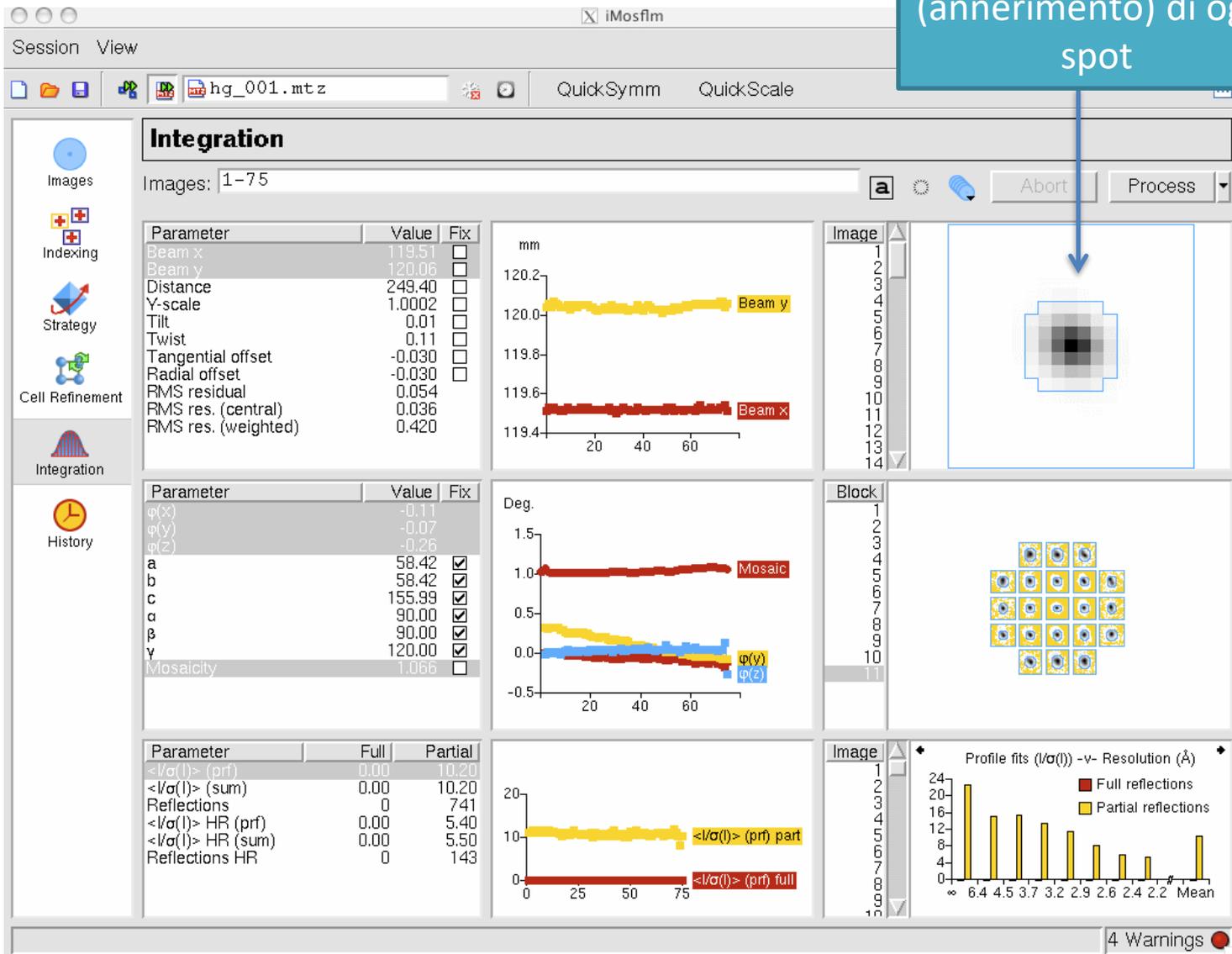
# INTEGRAZIONE (integration)

## 1. INTEGRAZIONE

Analizzo tutte le immagini di diffrazione:

- misurano le **INTENSITÀ di ogni picco (spot)**

Misuro l'intensità  
(annerimento) di ogni  
spot



# SCALAGGIO (scaling)

## 1. SCALAGGIO

Unisco e confronto tutti i dati ottenuti:

- scrivono in un **output le Intensità relative ad ogni posizione nello spazio reciproco (spot) a cui assegnano gli indici  $h, k, l$**  (indici di Miller)

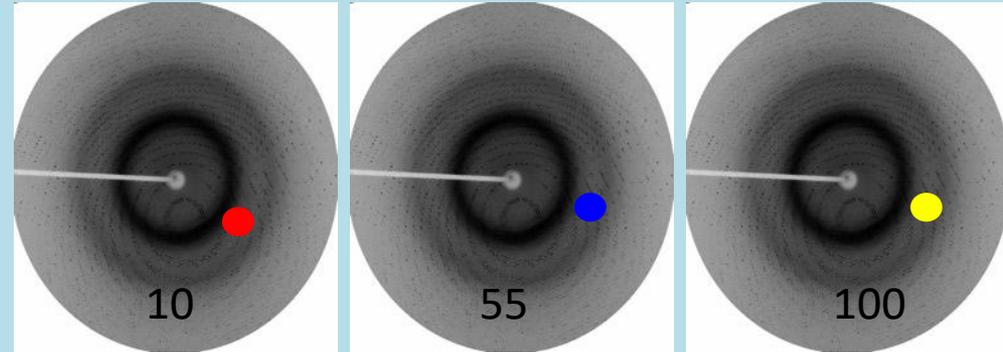
### STATISTICHE

( indicano la bontà del dato sperimentale)

- **mosaicità** (disordine intrinseco del cristallo)



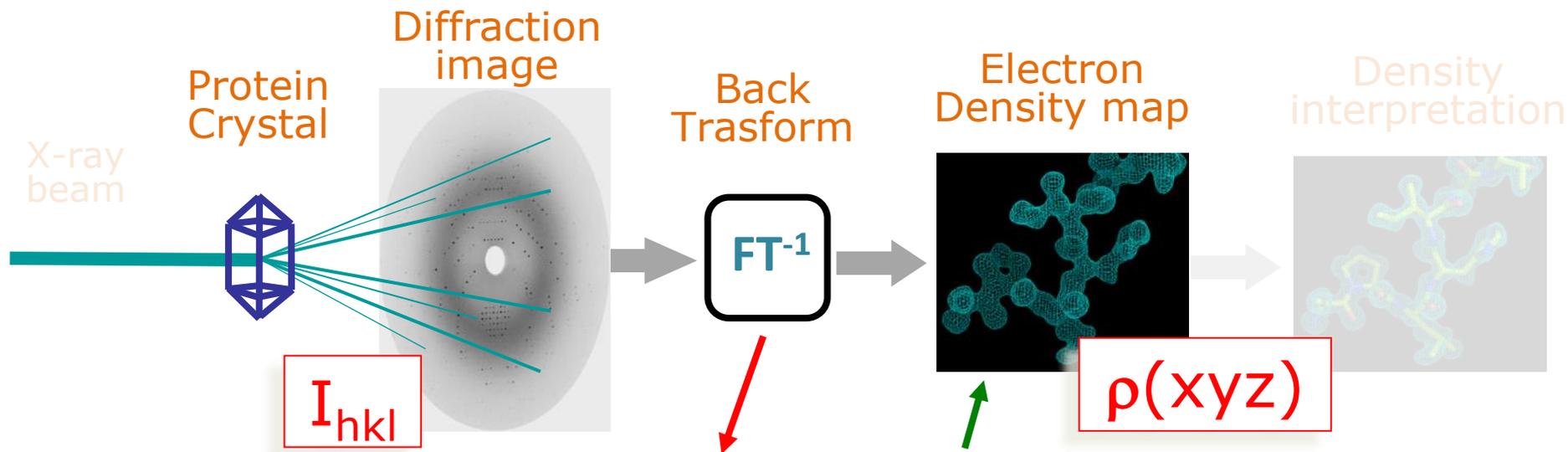
- **R merge**
- **completezza**, ridondanza
- **$I/\sigma(I)$**  (segnale/rumore)
- Etc...



$$I_1(234) \quad I_2(234) \quad I_3(234)$$

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^n |I(hkl) - I_i(hkl)|}{\sum_{hkl} N \times I(hkl)}$$

# L'ESPERIMENTO DI DIFFRAZIONE



$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

$\alpha_{hkl}$  is lost in the experiment - the phase problem!

structure factor amplitude  $|F_{hkl}| \propto (I_{hkl})^{1/2}$

OTTENUTI DAI DATI DI DIFFRAZIONE

PERSA



$$F(hkl) = |F_{hkl}| \exp(i\alpha_{hkl})$$

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

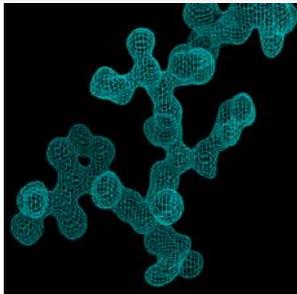
$\alpha_{hkl}$  is lost in the experiment -  
the phase problem!

structure factor amplitude  $|F_{hkl}| \propto (I_{hkl})^{1/2}$

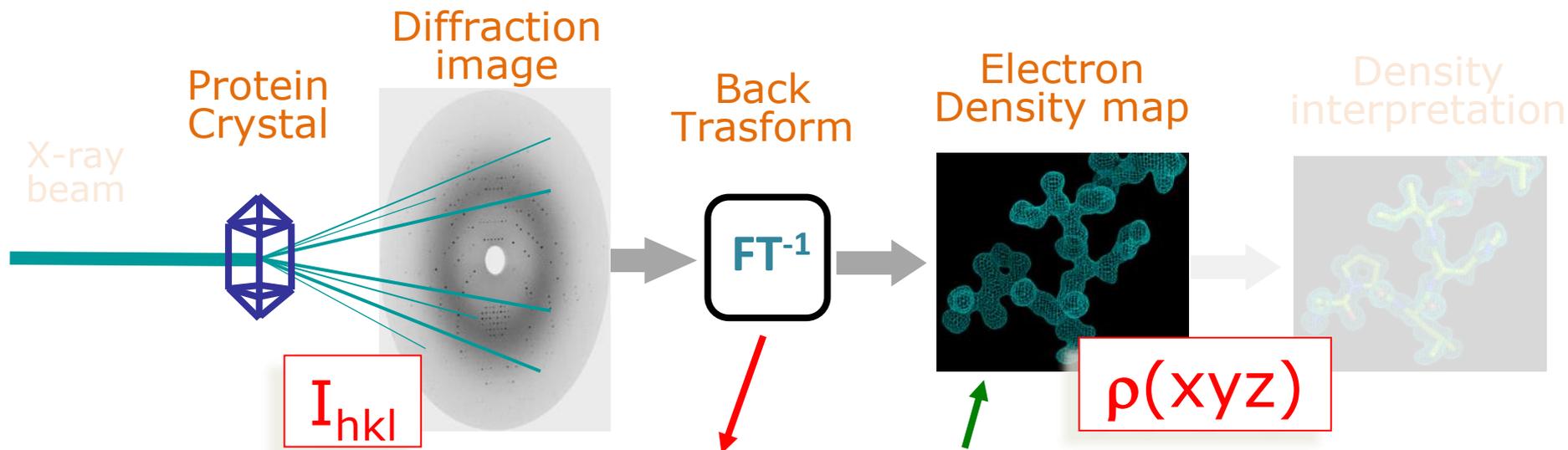
$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

Eq DENSITÀ ELETTRONICA

SAPZIO REALE



# L'ESPERIMENTO DI DIFFRAZIONE



$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

**AMPIEZZA**

structure factor amplitude  $|F_{hkl}| \propto (I_{hkl})^{1/2}$

**FASE**

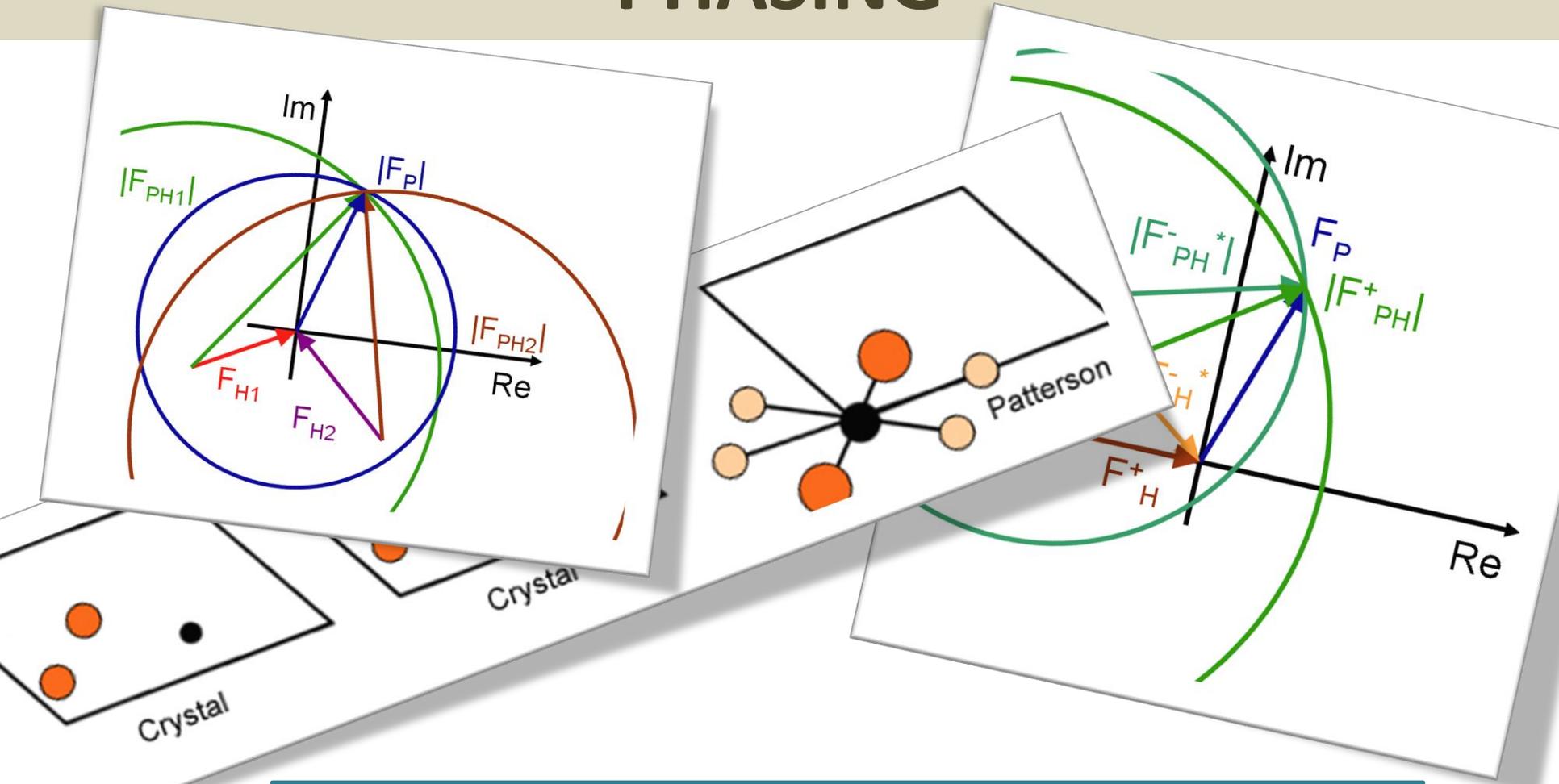
$\alpha_{hkl}$  is lost in the experiment - the phase problem!

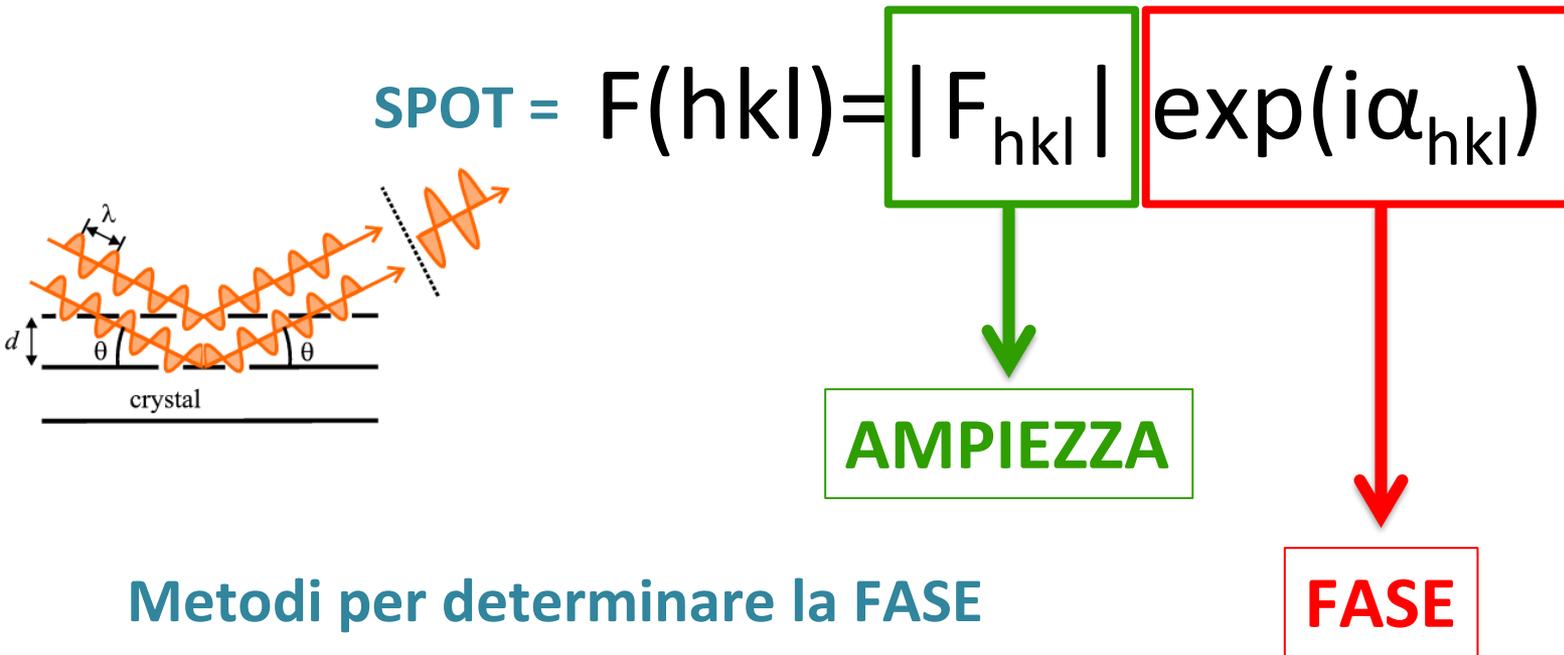
**PERSA**

**OTTENUTI DAI DATI DI DIFFRAZIONE**



# PHASING





## Metodi per determinare la FASE

- Metodi Diretti (solo per piccole molecole)
- Molecular Replacement
- Metodi basati su atomi pesanti
  - Sostituzione isomorfa (SIR – MIR)
  - Diffrazione Anomala (SAD - MAD)
  - Metodi combinati (SIRAS – MIRAS)

## PHASING: la funzione di PATTERSON

$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

MAPPA DI  
PATTERSON

USO LE  
AMPIEZZE  
SPERIMENTALI  
AL QUADRATO

PONGO TUTTE  
LE FASI = 0

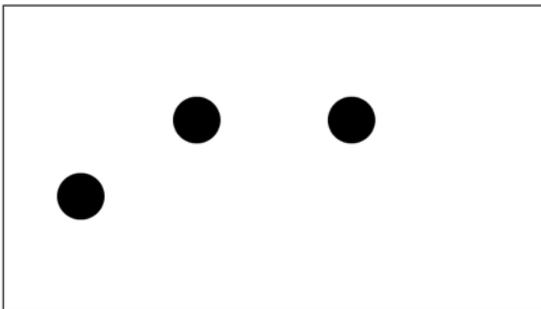
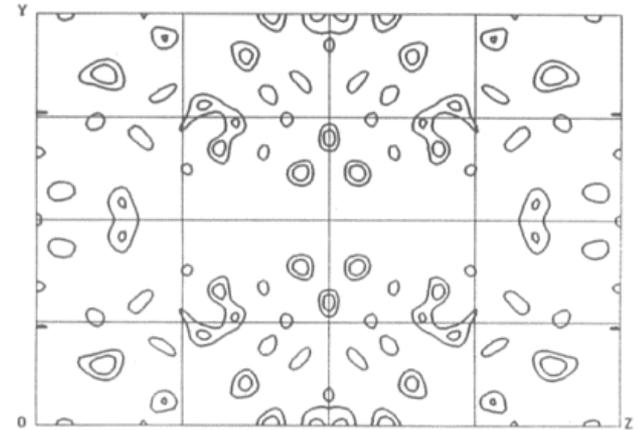
$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu+kv+lw)]$$

POSSO CALCOLARE LA MAPPA DI PATTERSON DAI DATI SPERIMENTALI, PERCHÉ NON DIPENDE PIÙ DALLA FASE

MAPPA DI  
PATTERSON

$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu+kv+lz)]$$

LA **MAPPA DI PATTERSON**  
MAPPA DI VETTORI INTERATOMICI:  
CONTIENE **INFORMAZIONI SULLE DISTANZE**  
**INTERATOMICHE**  
QUINDI INFORMAZIONI STRUTTURALI

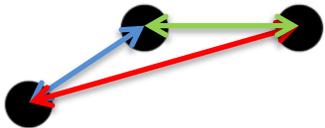
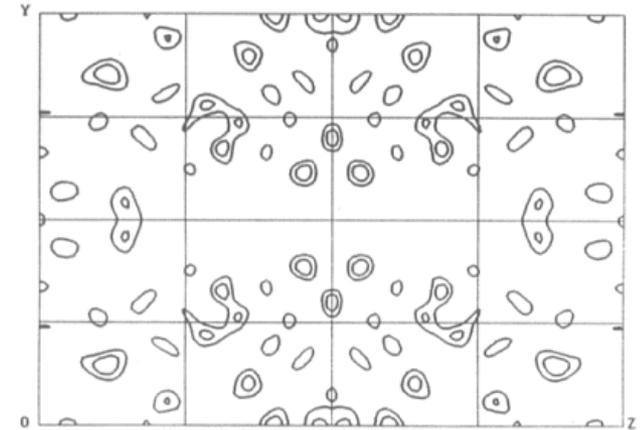


Disposizione di 3 Atomi  
nel cristallo

MAPPA DI  
PATTERSON

$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu+kv+lz)]$$

LA **MAPPA DI PATTERSON**  
MAPPA DI VETTORI INTERATOMICI:  
CONTIENE **INFORMAZIONI SULLE DISTANZE**  
**INTERATOMICHE**  
QUINDI INFORMAZIONI STRUTTURALI

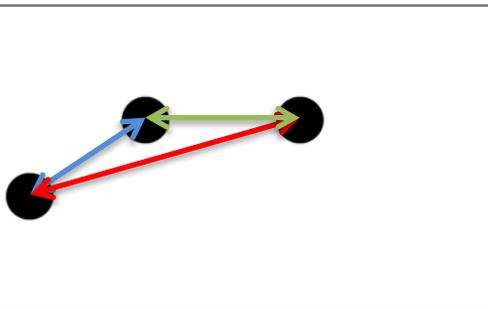
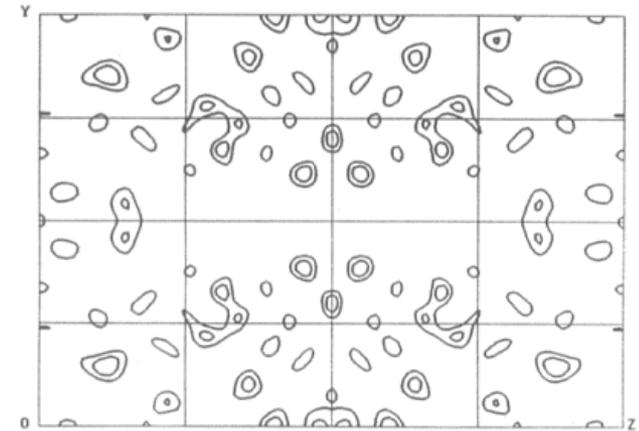


Disposizione di 3 Atomi  
nel cristallo

## MAPPA DI PATTERSON

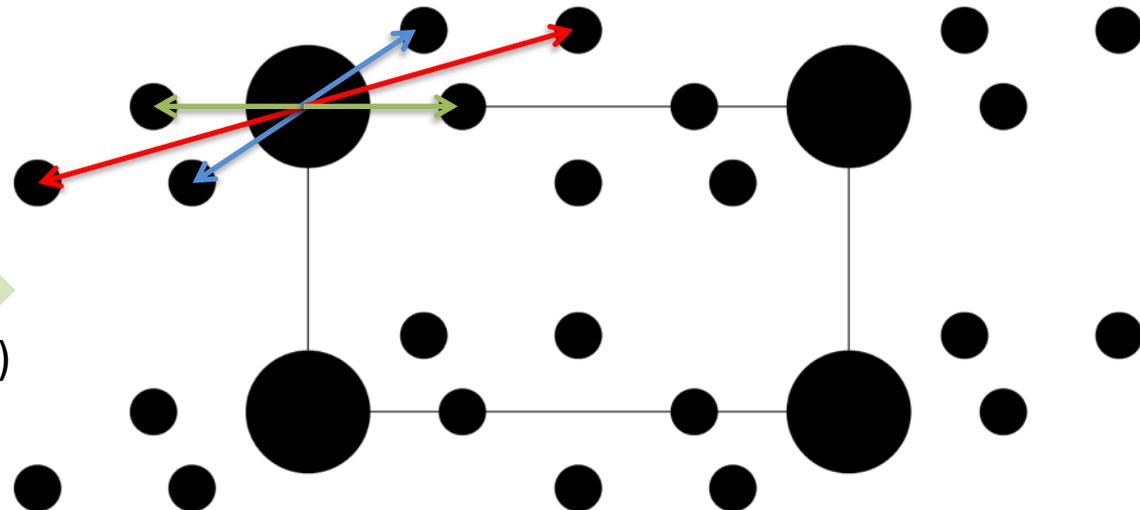
$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu + kv + lz)]$$

LA **MAPPA DI PATTERSON**  
 MAPPA DI VETTORI INTERATOMICI:  
 CONTIENE **INFORMAZIONI SULLE DISTANZE**  
**INTERATOMICHE**  
 QUINDI INFORMAZIONI STRUTTURALI



Disposizione di 3 Atomi  
 nel cristallo

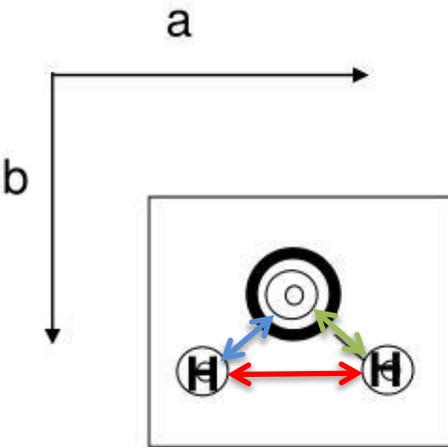
→  
 $P(u,v,w)$



Picchi della mappa di Patterson corrispondente

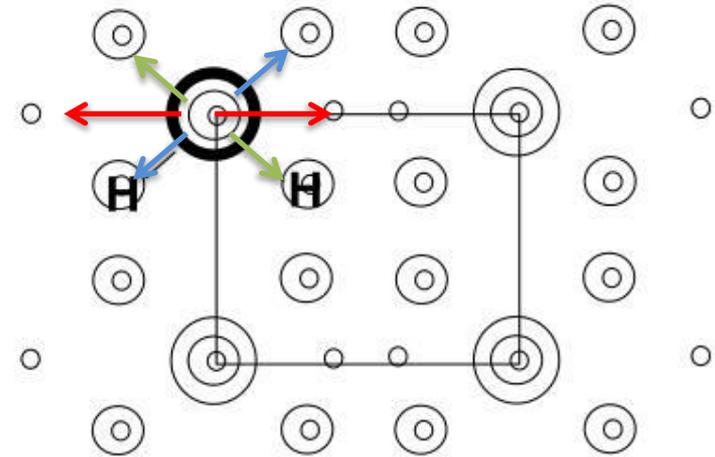
# Patterson maps are more complicated than electron density maps.

Imagine the complexity of a Patterson map of a protein



**Electron Density Map**  
single water molecule in  
the unit cell

Unit cell repeats fill  
out rest of cell with  
peaks

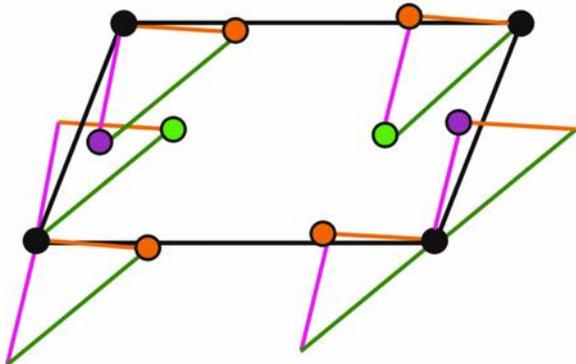
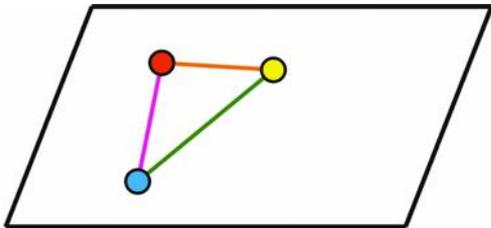


**Patterson Density Map**  
single water molecule  
convoluted with its  
inverted image.

MAPPA DI  
PATTERSON

$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu+kv+lz)]$$

## METODI DIRETTI

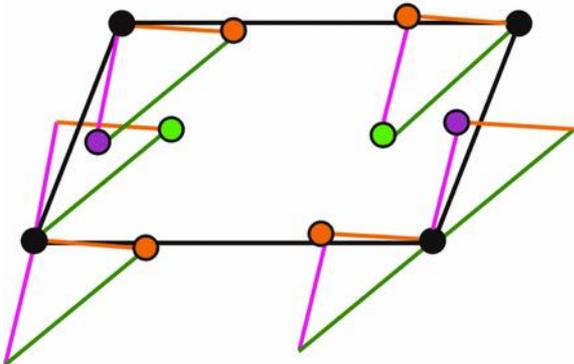
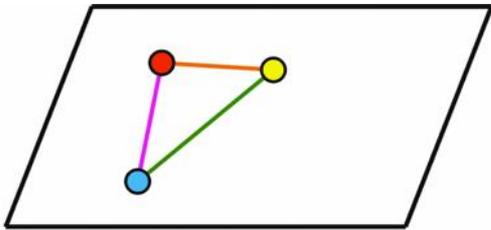


Dalla mappa di Patterson  
per molecole con < 100 atomi  
**Posso provare tutte le posizioni  
atomiche possibili**  
E determinare le fasi

MAPPA DI  
PATTERSON

$$P(u,v,w) = 1/V \sum |F_{hkl}^2| \exp[-2\pi i(hu+kv+lz)]$$

## METODI DIRETTI



Dalla mappa di Patterson  
per molecole con < 100 atomi  
**Posso provare tutte le posizioni  
atomiche possibili**  
E determinare le fasi

SOLO  
Piccole  
molecole

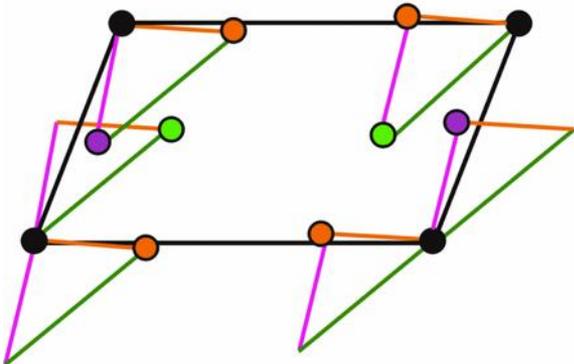
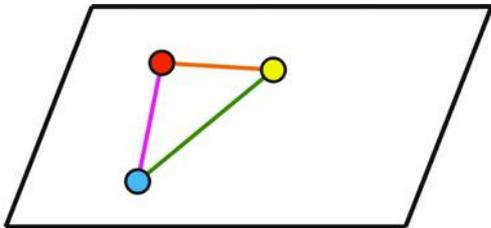
# PHASING: MOLECULAR REPLACEMENT

$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

MAPPA DI  
PATTERSON

AMPIEZZE  
SPERIMENTALI

PONGO TUTTE  
LE FASI = 0



**PER LE PROTEINE?**

NON POSSO USARE I METODI DIRETTI

perché ho troppi atomi, ma posso partire dalla disposizione atomica di proteine omologhe → **elevata identità di sequenza (> 30-50%)** → **simile struttura 3D**

# Molecular Replacement (MR)

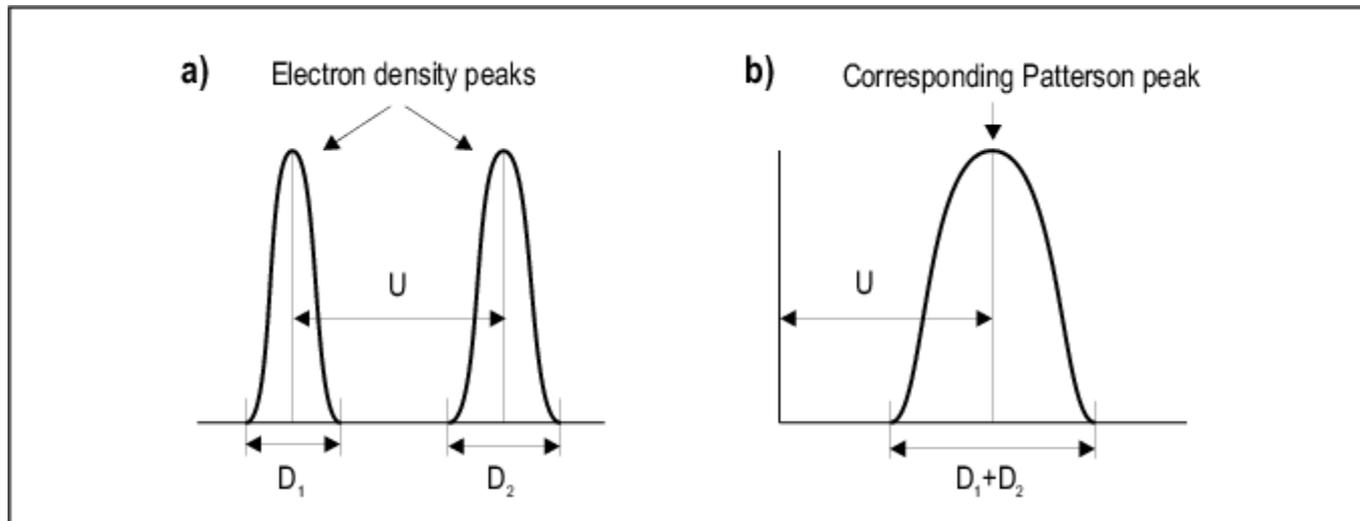
- La proteina di interesse deve avere una proteina omologa di cui sia presente la struttura nel PDB (**search model**)
- L'identità di sequenza deve essere maggiore del 30 %
- Si confrontano le due Patterson, quella ottenuta dal modello e quella ottenuta dai dati sperimentali

$$P(\vec{u}) = \int_V \rho(\vec{x})\rho(\vec{x} + \vec{u})dV$$

**model**

$$P(\vec{u}) = \frac{1}{V} \sum_{\vec{h}} \sum_{\vec{h}'} F(\vec{h})F(\vec{h}') e^{-2\pi i\vec{h}'\vec{u}} = \frac{1}{V} \sum_{\vec{h}} |F(\vec{h})|^2 e^{-2\pi i\vec{h}\vec{u}}$$

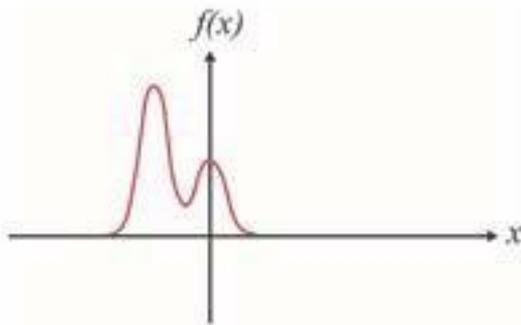
**experimental**



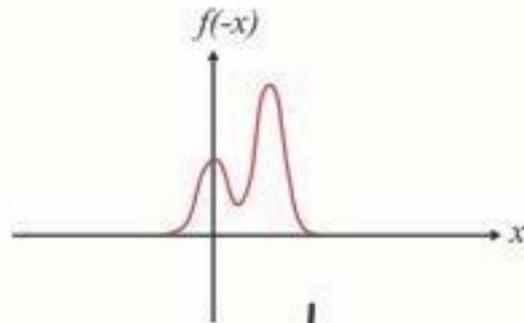
# Convolution operation

La Patterson del modello  
Scaricato dal pdb è calcolata  
applicando l'autoconvoluzione alla  
funzione della densità elettronica

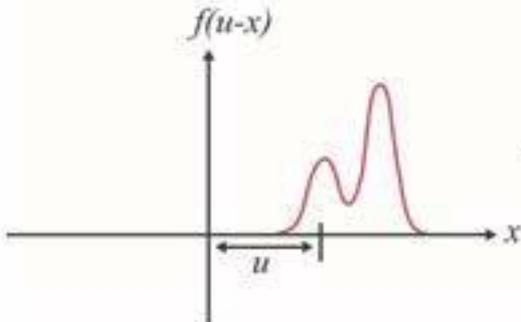
$$f(x) * g(x) = \int f(x) g(u-x) dx$$



generate  $f(-x)$

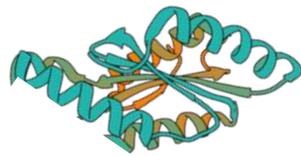


translate  $f(-x)$  by  $u$



form  
integral

# PHASING: MOLECULAR REPLACEMENT

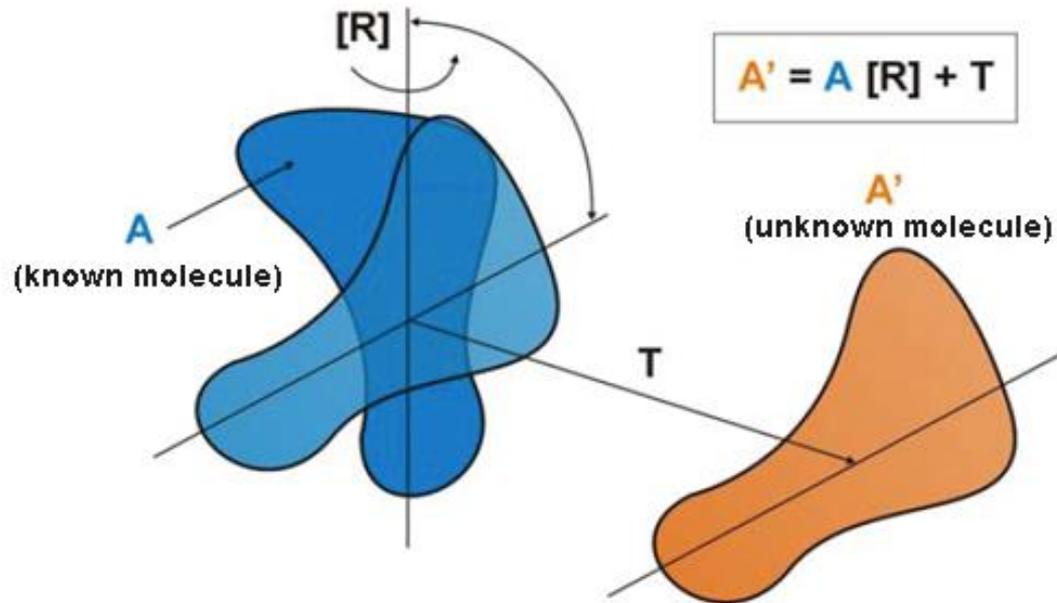


**MAPPA DI  
PATTERSON  
MODELLO**

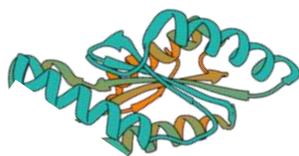
Ruoto e Traslo la mappa del  
modello per farla coincidere  
con quella della proteina



**MAPPA DI  
PATTERSON  
PROTEINA**



# PHASING: MOLECULAR REPLACEMENT



**MAPPA DI  
PATTERSON  
MODELLO**

Ruoto e Traslo la mappa del  
modello per farla coincidere  
con quella della proteina

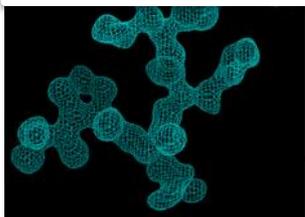


**MAPPA DI  
PATTERSON  
PROTEINA**



**FASI DEL  
MODELLO orientato  
Come nel cristallo reale**

**MAPPA DI  
DENSITÀ ELETTRONICA**



**AMPIEZZE  
SPERIMENTALI**



$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$



## METODI BASATI SUGLI ATOMI PESANTI

### Principio di base:

se ad un'onda incognita aggiungo **un'onda nota**, l'onda risultante mi permetterà di risalire all'incognita

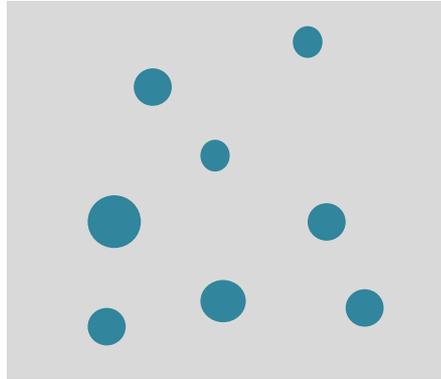
### Mezzi sperimentali utilizzati:

- **Atomi pesanti** (alto numero di elettroni)
- **Atomi che a  $\lambda$  diverse diffrangono in modo diverso** perché assorbono una parte dei raggi X (diffusori anomali)

### Metodologia:

- **Il cristallo viene messo in contatto con una soluzione di un sale di atomi pesanti oppure gli atomi pesanti vengono incorporati nelle proteine (Selenio-metionine)**
- Ottengo che pochi atomi si posizionano in punti discreti e uguali in tutte le molecole
- Questa aggiunta **non deve variare in alcun modo (<0.5%) i parametri della cella** o l'orientazione della molecola nella stessa (isomorfismo)
- **L'immagine di diffrazione non cambia, cambia solo l'intensità di alcuni riflessi**

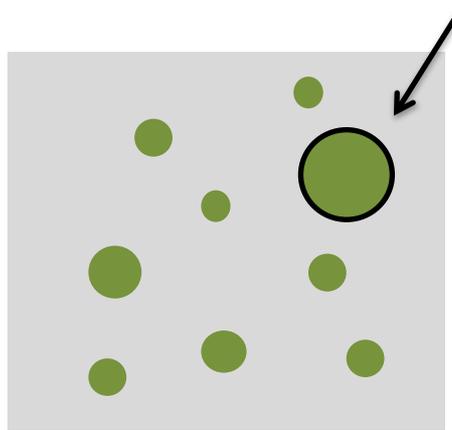
SOSTITUZIONE ISOMORFA



$|F_p|$  Proteina Nativa

$$F_p = |F_p| \exp(i\alpha_p)$$

???



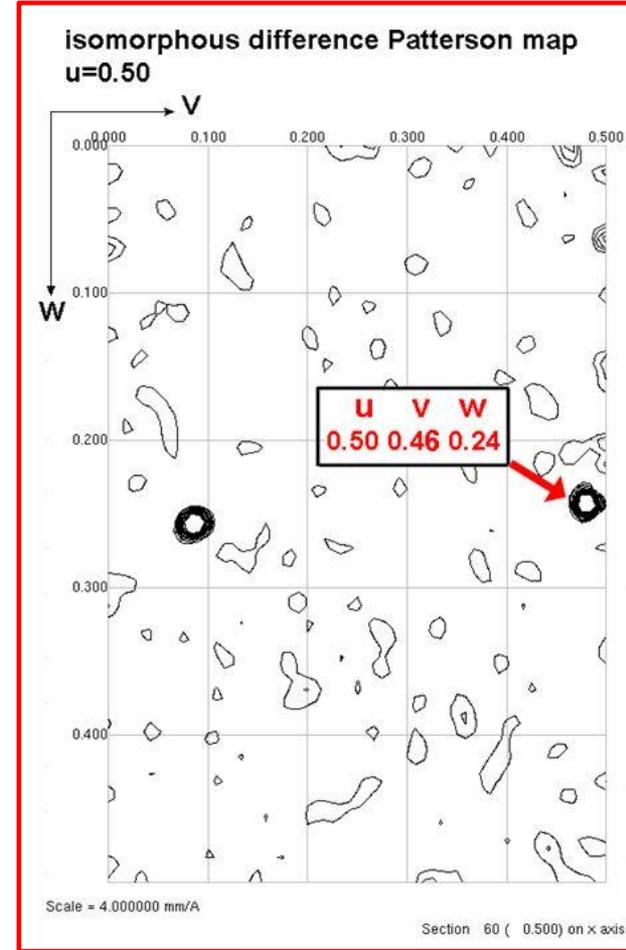
Heavy Atom

$|F_{pH}|$  Proteina + Heavy Atom

$$F_{pH} = |F_{pH}| \exp(i\alpha_{pH})$$

???

**SIR**  
Single  
Isomorphous  
Replacement

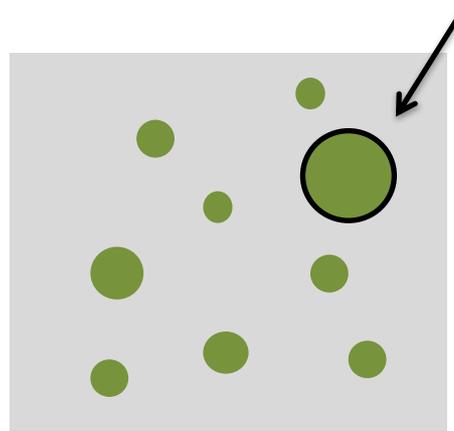


Posso fare la differenza  
delle mappe di  
Patterson

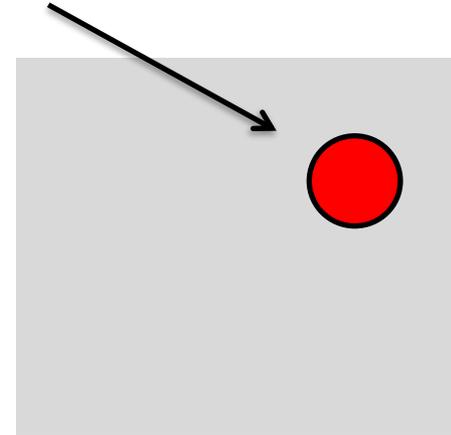
SOSTITUZIONE ISOMORFA



$|F_P|$  Proteina Nativa



$|F_{PH}|$  Proteina + Heavy Atom



$F_H$  Heavy Atom

$$F_P = |F_P| \exp(i\alpha_P)$$

???

$$F_{PH} = |F_{PH}| \exp(i\alpha_{PH})$$

???

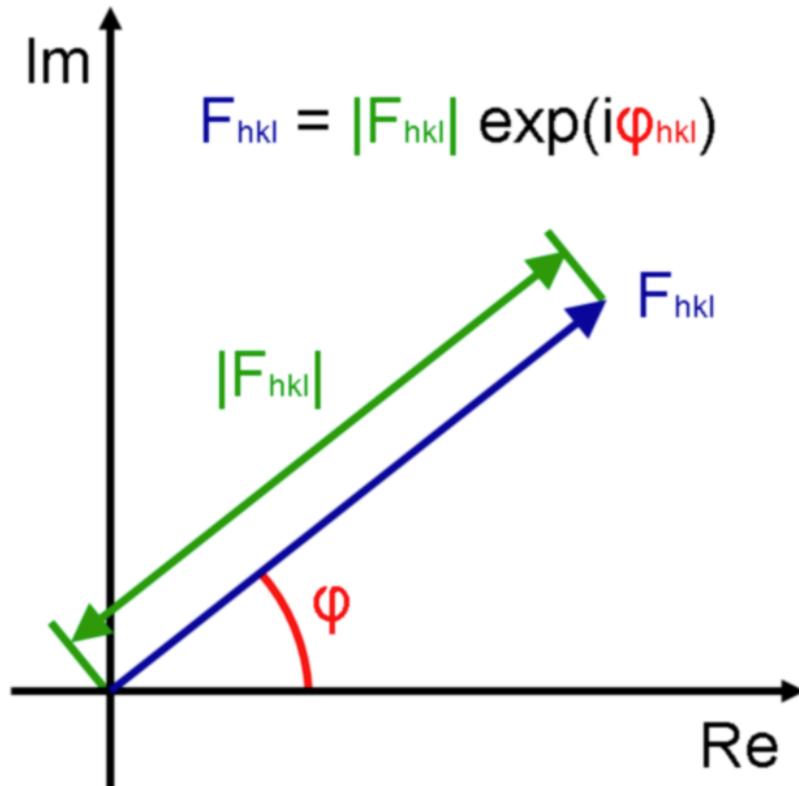
$$F_H = |F_H| \exp(i\alpha_H)$$

**SIR**  
Single  
Isomorphous  
Replacement

**USO I METODI DIRETTI**  
Per calcolare la  
Heavy Atom  
sub-structure

## METODI BASATI SUGLI ATOMI PESANTI

$$F(hkl) = |F_{hkl}| \exp(i\alpha_{hkl})$$



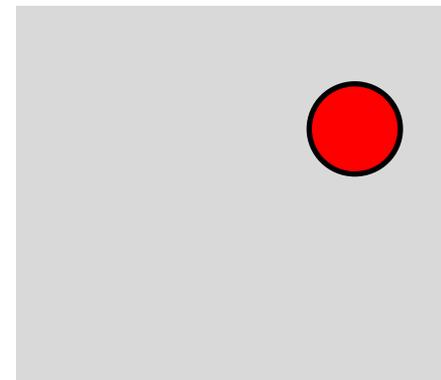
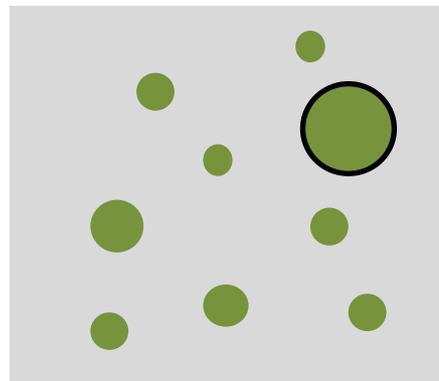
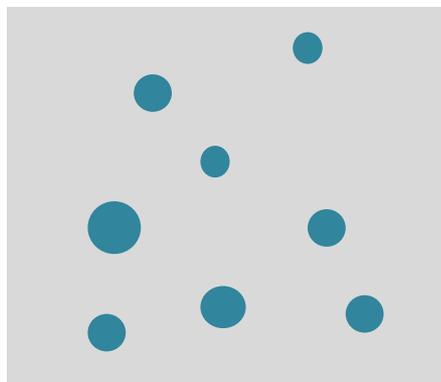
AMPIEZZA

FASE

...UN ALTRO MODO DI  
RAPPRESENTARE I  
FATTORI DI STRUTTURA

SOSTITUZIONE ISOMORFA

$$F_{hkl} = |F_{hkl}| \exp(i\alpha_{hkl})$$

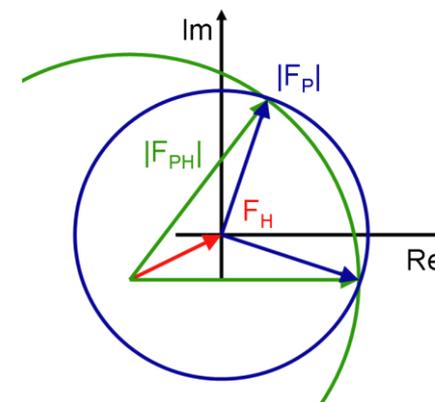
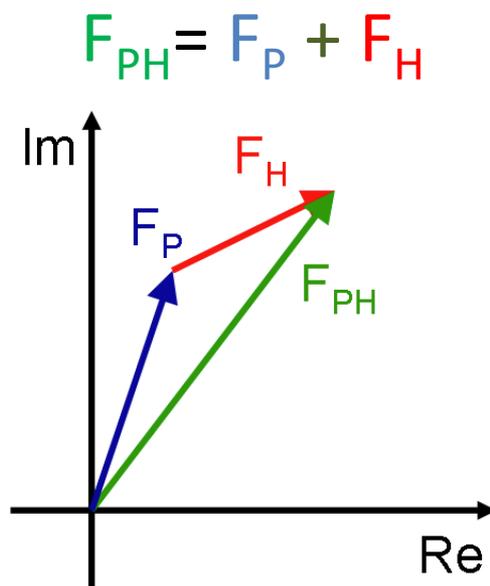


$|F_P|$  Proteina Nativa

$|F_{PH}|$  Proteina + Heavy Atom

$F_H$  Heavy Atom

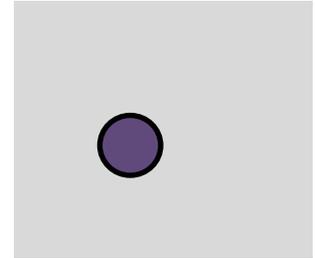
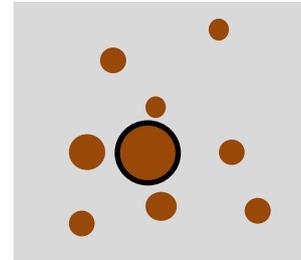
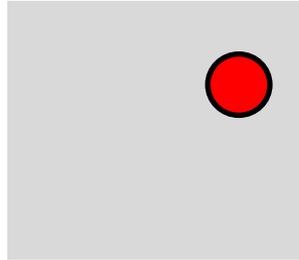
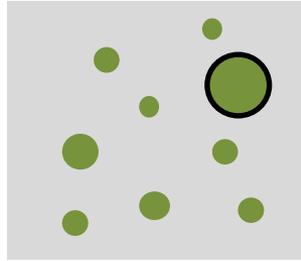
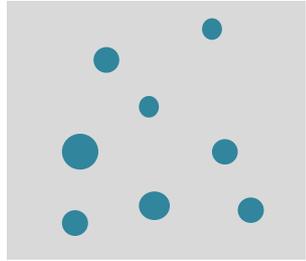
**SIR**  
Single  
Isomorphous  
Replacement



2 possibili soluzioni per la fase di  $F_p$  (phase ambiguity)  
Controllo le mappe per capire quella corretta

SOSTITUZIONE ISOMORFA

$$F_{hkl} = |F_{hkl}| \exp(i\alpha_{hkl})$$



$$|F_P|$$

Diff. Proteina Nativa

$$|F_{PH1}|$$

Diff. Prot + HA 1

$$F_{H1}$$

Diff. HA 1

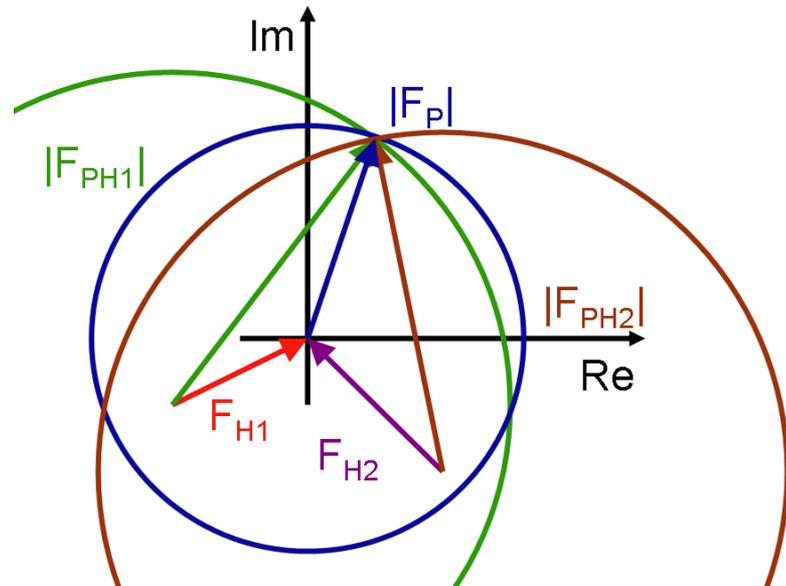
$$|F_{PH2}|$$

Diff. Prot + HA 1

$$F_{H2}$$

Diff. HA 1

**MIR**  
Multiple  
Isomorphous  
Replacement



AGGIUNGO UN ALTRO  
ATOMO PESANTE  
Non ho più ambiguità

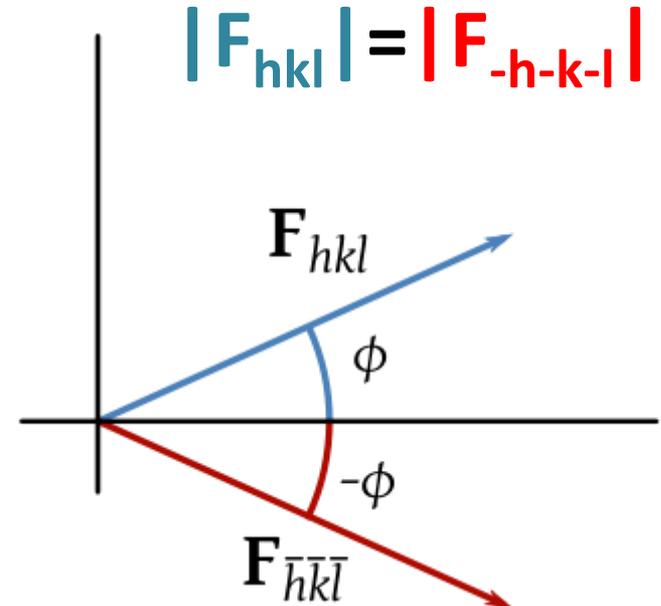
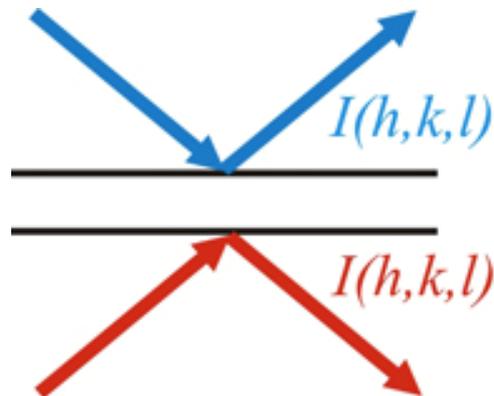
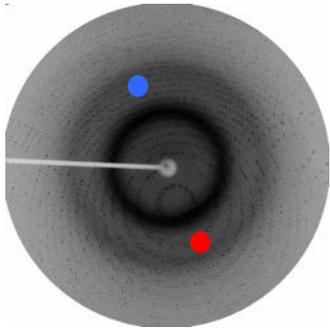
## Classi di atomi pesanti (H, heavy atoms)

- Ioni metallici legati elettrostaticamente alla molecola
- Metalli endogeni che possono essere facilmente sostituiti (es.  $\text{Sr}^{2+}$  per  $\text{Ca}^{2+}$ )
- **Selenio (es. Se-Met)**
- Composti metallici legati covalentemente (reazione chimica su cristallo)
- Agglomerati multimetallici per molecole  $> 500$  Kda (es. cluster di oro o di tantalio)
- Gas nobili come Xe e Kr
- Alogenuri (es. tri-ioduri e basi brominate)

## DIFFRAZIONE NORMALE

$$F_{hkl} = |F_{hkl}| \exp(i\alpha_{hkl})$$

## LEGGI DI FRIEDEL



In condizioni di diffrazione non anomala 2 riflessi simmetrici (coppie di Friedel) hanno stessa ampiezza e fase opposta

DIFFRAZIONE ANOMALA

$$f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$$

**A DETERMINATE  $\lambda$  GLI ATOMI PESANTI POSSONO ASSORBIRE I RAGGI X**

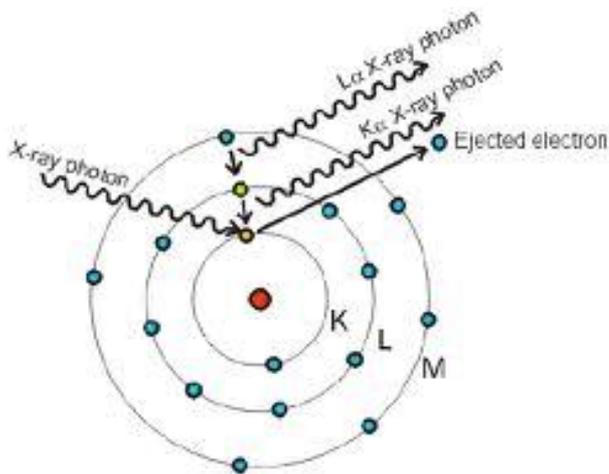
Estraendo un  $e^-$  da un orbitale più interno.

**Il segnale dell'onda diffratta viene modificato.**

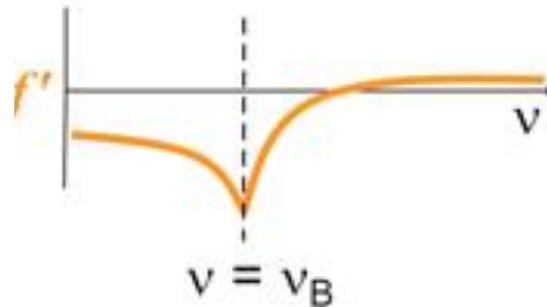
$$f = f_0 + f' + if''$$

$f_0$  = normal atomic scattering factor (real number)

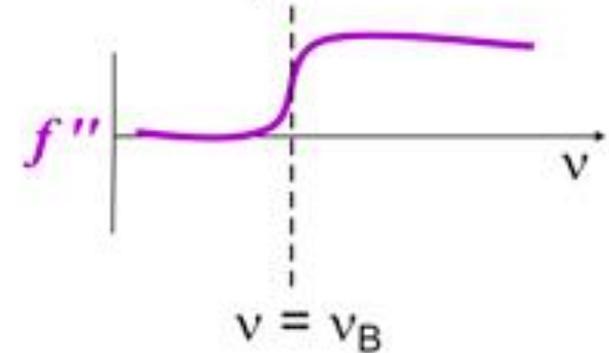
$f'$  and  $f''$  = real and imaginary correction factors



$$f' = g \left( \frac{v_B}{v} \right)^2 \log_c \left| \left( \frac{v}{v_B} \right)^2 - 1 \right|$$



$$f'' = g \frac{\pi v_B^2}{v^2}$$



## DIFFRAZIONE ANOMALA

$$f(\lambda) = f^o + f'(\lambda) + if''(\lambda)$$

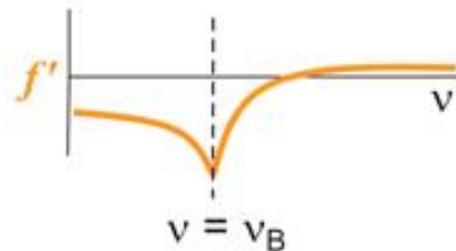
- **A particolari  $\lambda$  (edges, cuspidi)** i termini  $f'(\lambda)$  e  $f''(\lambda)$  mostrano, rispettivamente, un **minimo** ed un **massimo**
- **Il termine dispersivo  $[f'(\lambda)]$  diminuisce** il fattore di diffrazione normale perché parte dell'energia viene spesa per far avvenire la transizione
- **Il termine di assorbimento  $[f''(\lambda)]$  subisce una variazione di fase di  $\pi/2$**  perché l'energia di eccitazione viene riemessa (diffratta) con un certo ritardo

$$f = f_o + f' + if''$$

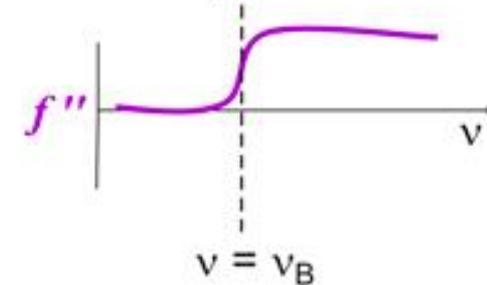
$f_o$  = normal atomic scattering factor (real number)

$f'$  and  $f''$  = real and imaginary correction factors

$$f' = g \left( \frac{v_B}{v} \right)^2 \log_e \left| \left( \frac{v}{v_B} \right)^2 - 1 \right|$$



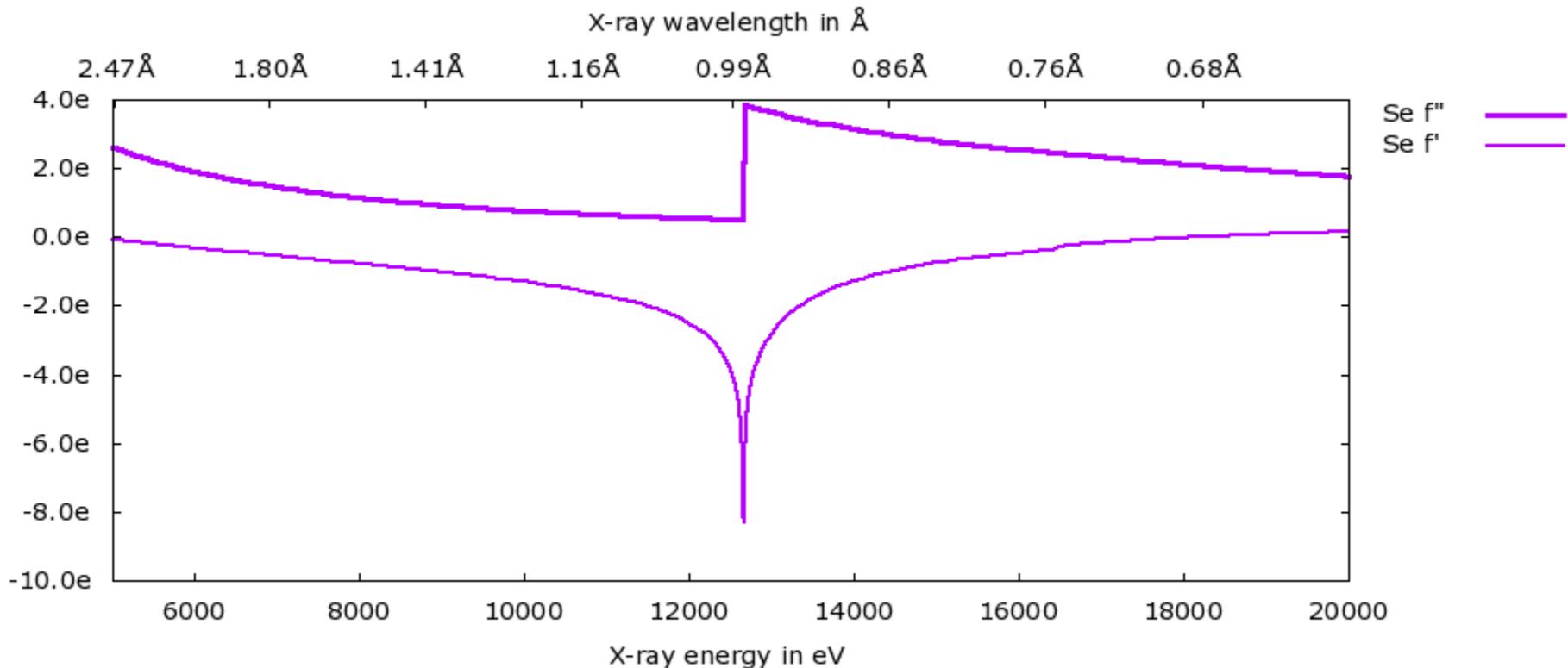
$$f'' = g \frac{\pi v_B^2}{v^2}$$



## DIFFRAZIONE ANOMALA

$$f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$$

Solitamente si usano Proteine espresse in presenza di  
**Selenio-Metionine** (automazione)



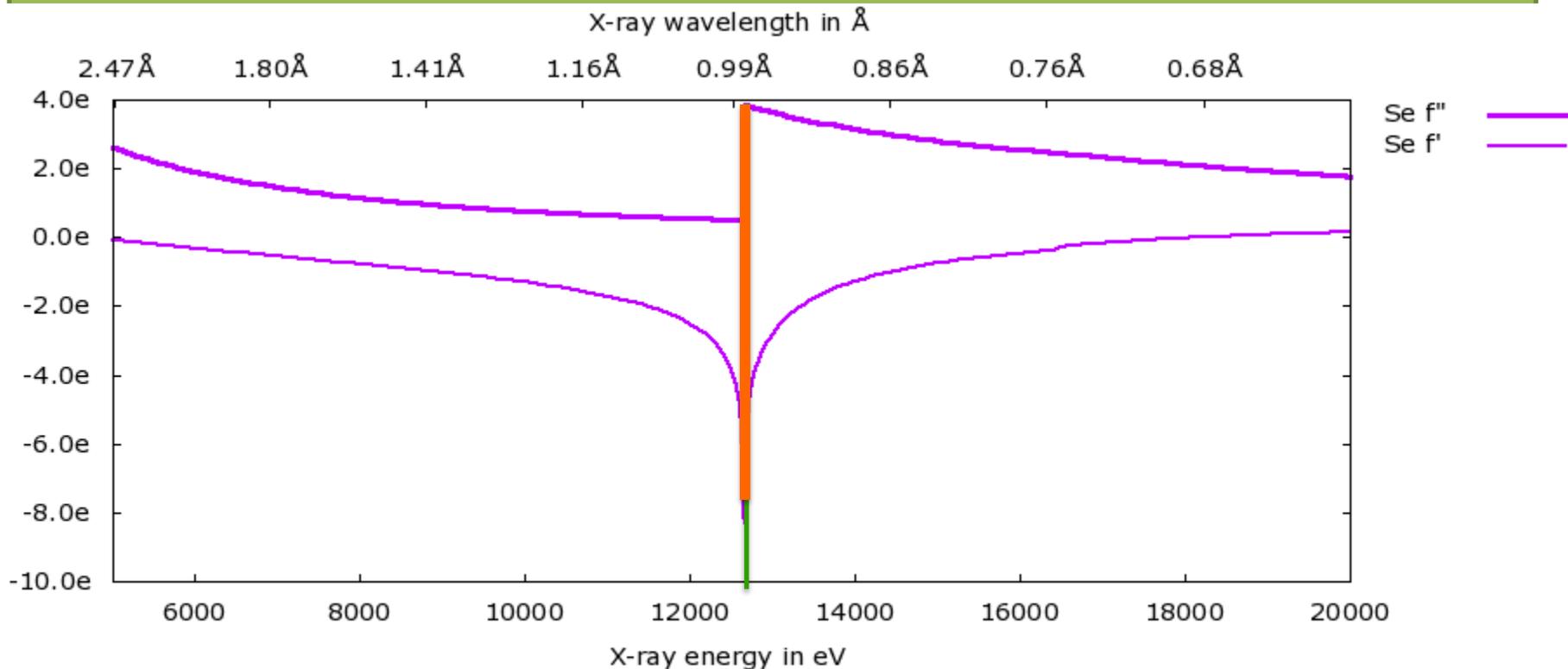
DIFFRAZIONE ANOMALA

$$f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$$

Single Anomalous Dispersion

**SAD**

1 raccolte dati a 1 sola lunghezza d'onda (edge)



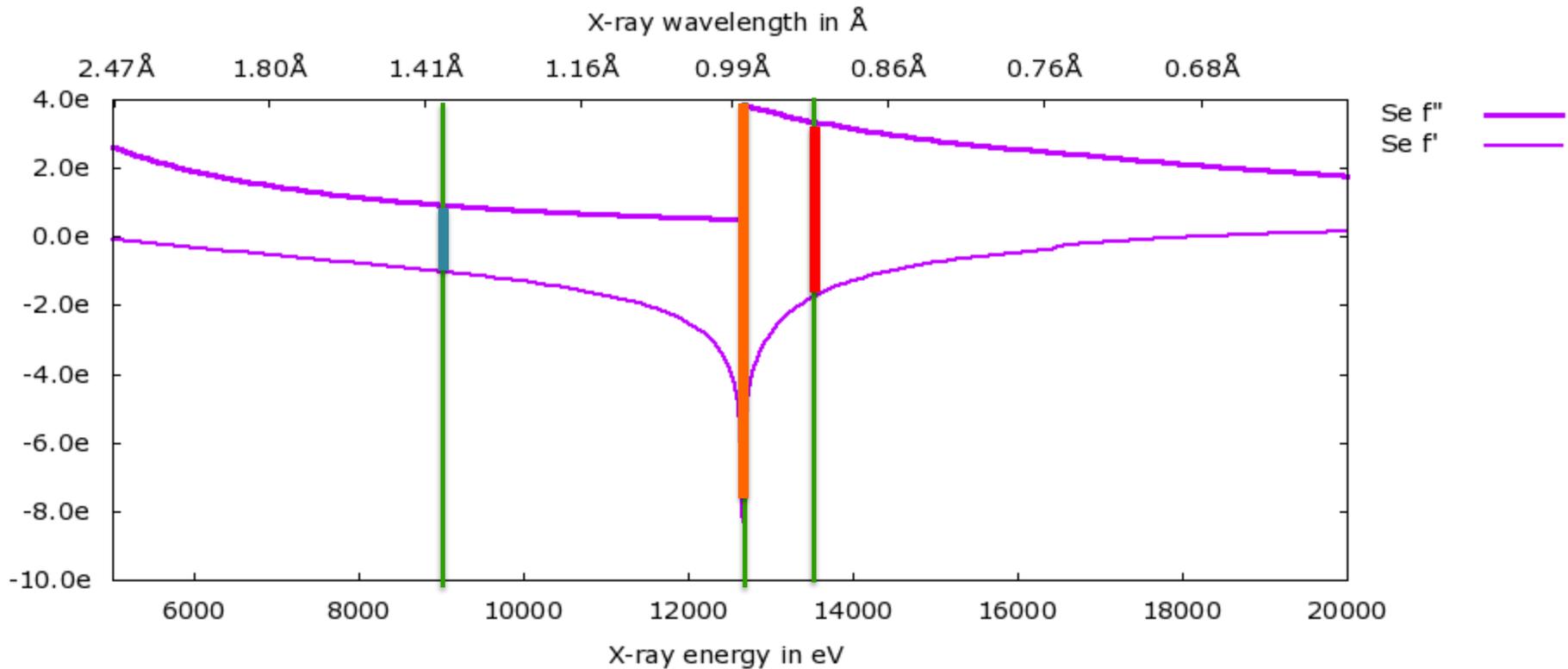
DIFFRAZIONE ANOMALA

$$f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$$

Multiple Anomalous Dispersion

**MAD**

3 raccolte dati a 3 diverse lunghezze d'onda



DIFFRAZIONE ANOMALA

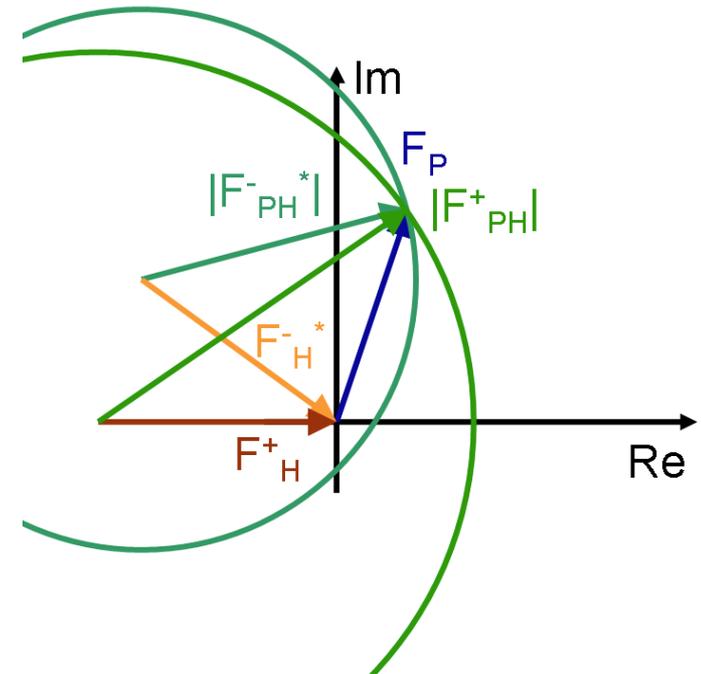
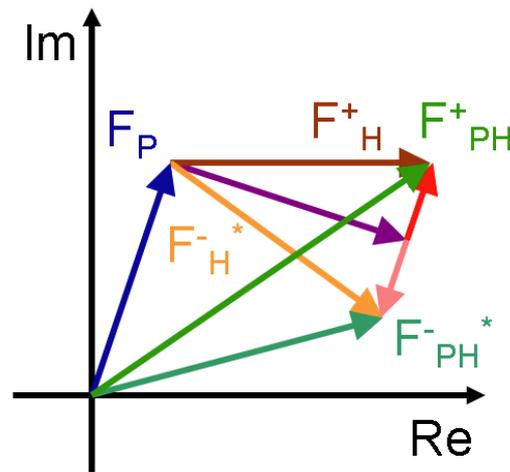
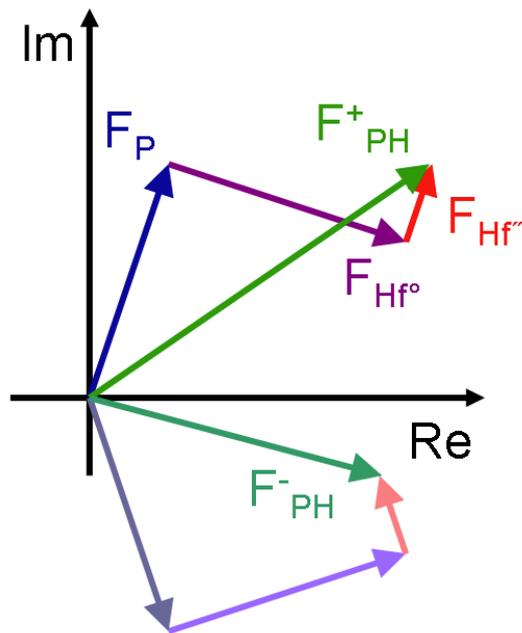
$$f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$$

In presenza di diffrazione anomala

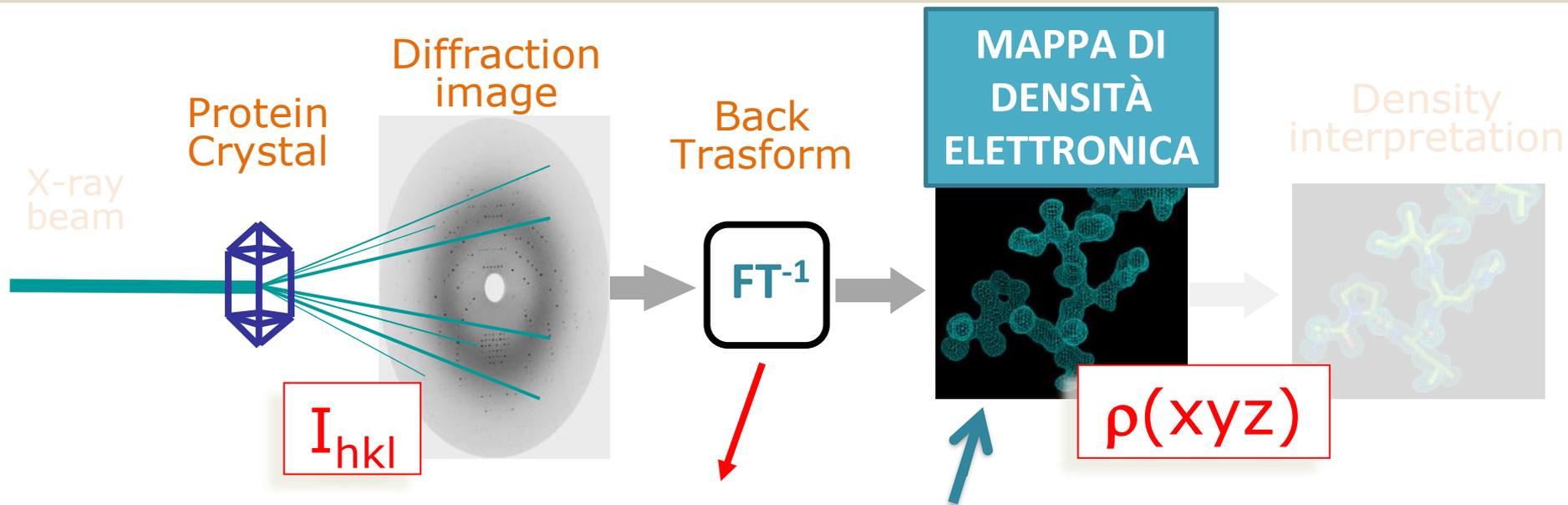
**non vale più la legge di Friedel**

$$|F_{PH}^{+}| \neq |F_{PH}^{-}|$$

Questa anomalia consente di calcolare la posizione degli atomi pesanti mediante metodi diretti e poi di determinare la fase di  $F_P$



# L'ESPERIMENTO DI DIFFRAZIONE



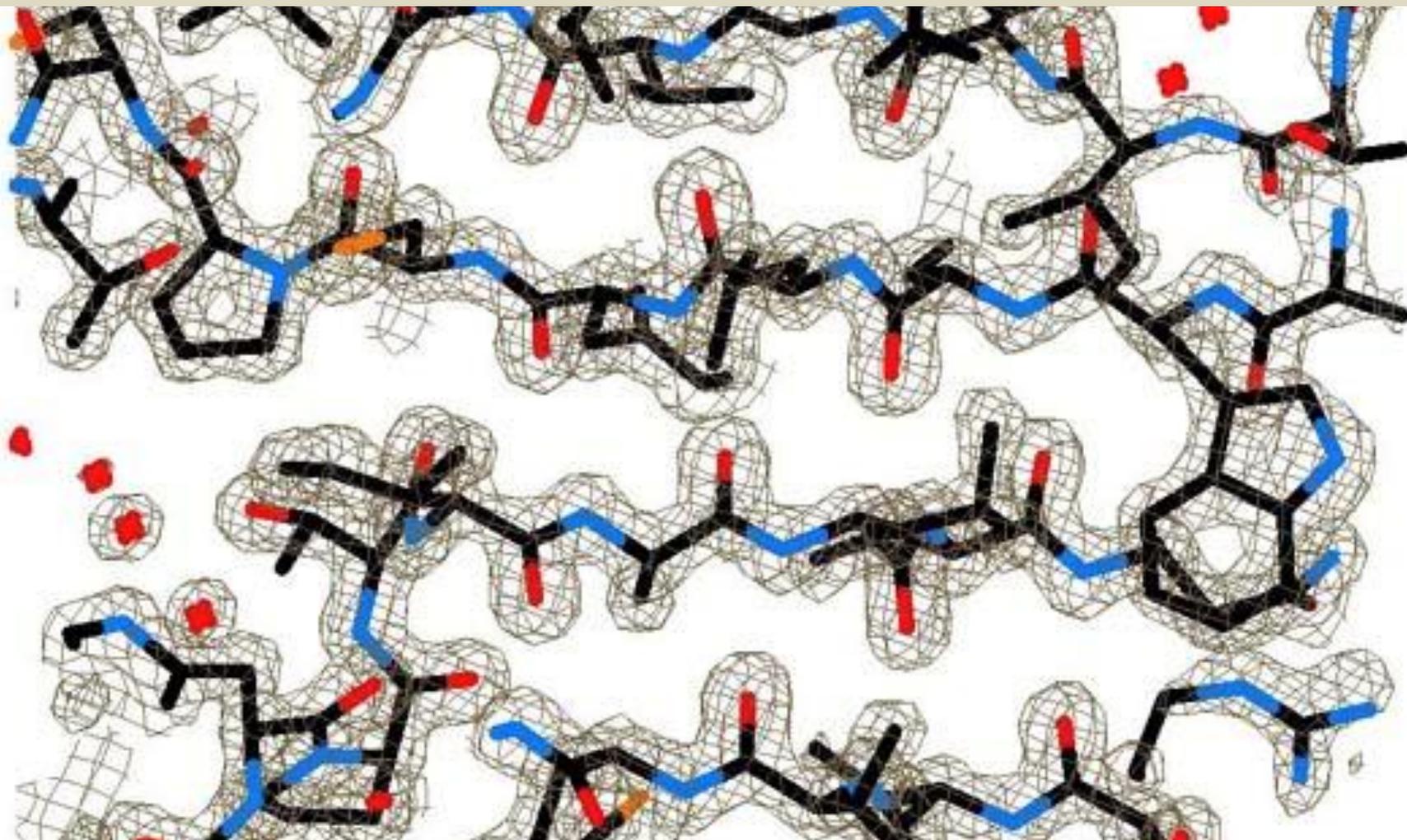
$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

USO LE  
AMPIEZZE  
SPERIMENTALI

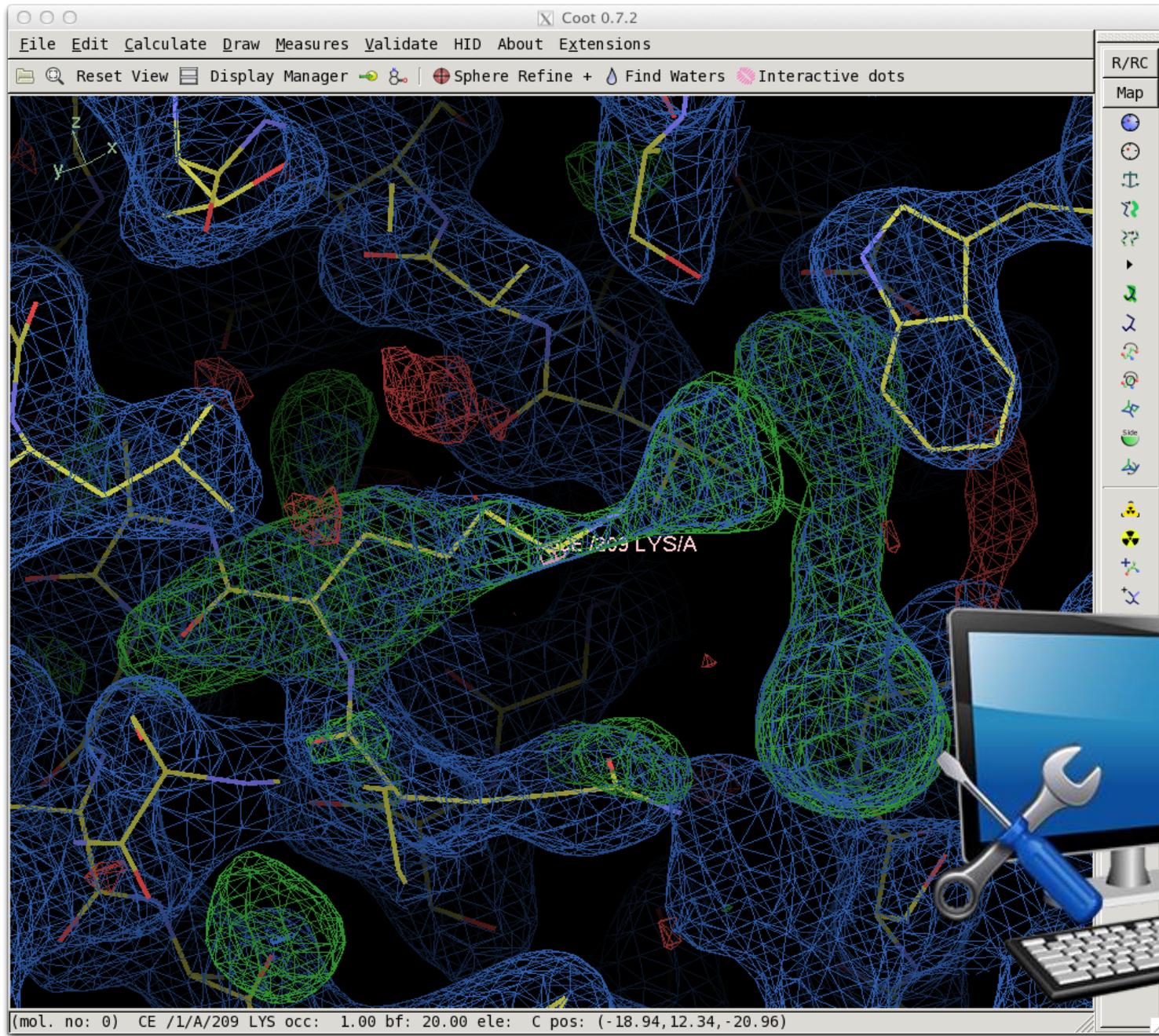
OTTENUTO  
CON UNO DEI  
METODI DESCRITTI



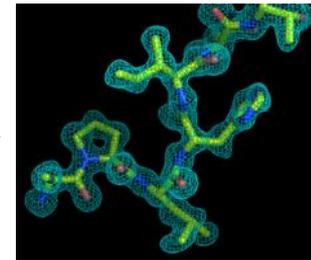
# MODEL BUILDING AND REFINEMENT



# MODEL BUILDING



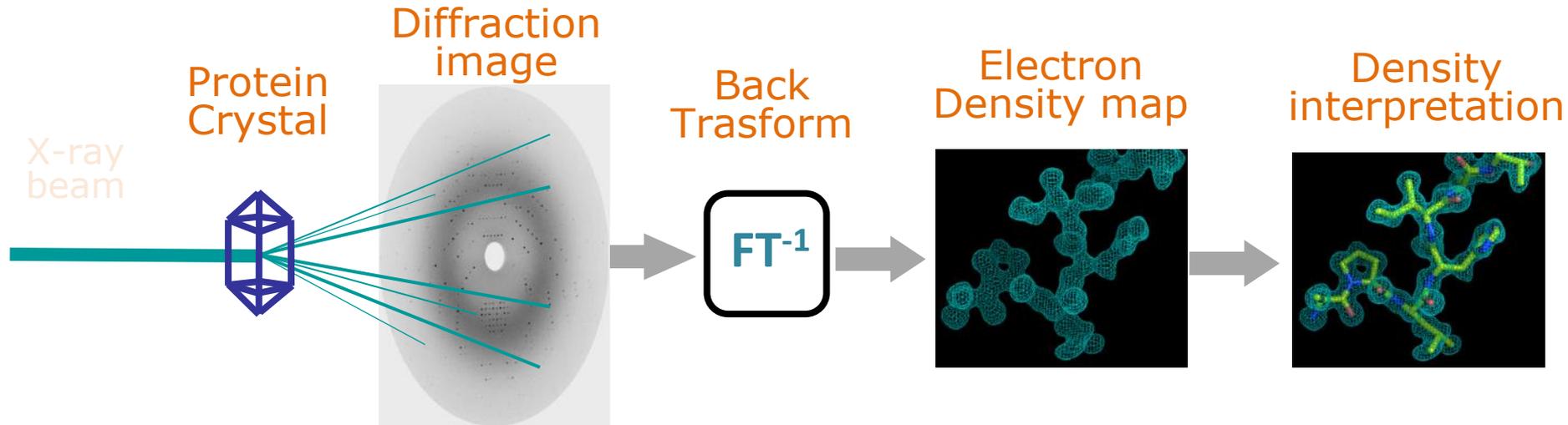
Density interpretation



IS FUN



# MODEL BUILDING AND REFINEMENT CYCLE



$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

$|F_{obs}|$

Amplitudes from experimental data

$|F_{cal}|$

$\alpha_{cal}$

Phases from model

1. Build a model and Calculate new maps
2. Refine the model and get better maps



# Refinement Cycle

**Refinement:** Improving the agreement between the model and the experimental density.

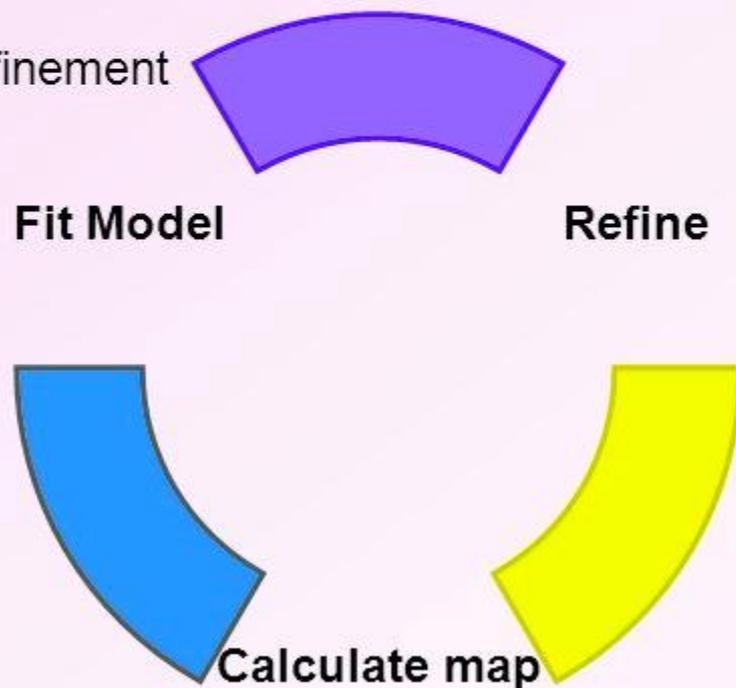
Compare  $F_{obs}$  (From reflection Intensities) to  $F_{calc}$  (Calculated from the model)

Least squares minimization

Simulated Annealing / Molecular dynamics

$R_{factor}$  = numerical indicator to follow progress of refinement  
agreement between data and model

$$R = \frac{\sum \left\| \overset{\substack{\text{data} \\ \downarrow}}{F_{obs}} - \overset{\substack{\text{model} \\ \downarrow}}{F_{calc}} \right\|}{\sum \underset{\substack{\uparrow \\ \text{data}}}{|F_{obs}|}}$$



# MODEL BUILDING AND REFINEMENT CYCLE

In realtà io calcolo 2 tipi di mappe in cui al posto di  $|F_{hkl}|$  metto:

$2|F_{obs}| - |F_{cal}|$  = mappa di densità elettronica (solitamente visualizzata in **blu**)

Oppure

$|F_{obs}| - |F_{cal}|$  = mappa DIFFERENZA (visualizzata in **verde** se  $> 0$ , in **rosso** se  $< 0$ )

$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

$$|F_{obs}|$$

Amplitudes from  
experimental data

$$|F_{cal}|$$

$$\alpha_{cal}$$

Phases from  
model

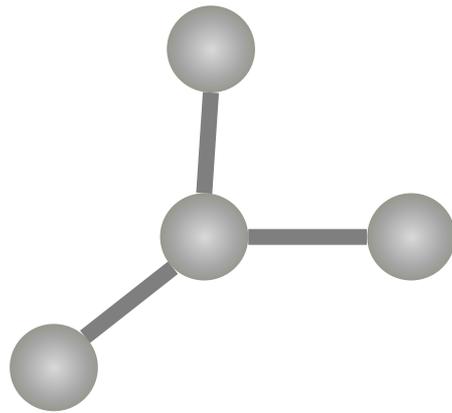
# MODEL BUILDING AND REFINEMENT CYCLE

In realtà io calcolo 2 tipi di mappe in cui al posto di  $|F_{hkl}|$  metto:

$2|F_{obs}| - |F_{cal}|$  = mappa di densità elettronica (solitamente visualizzata in **blu**)

Oppure

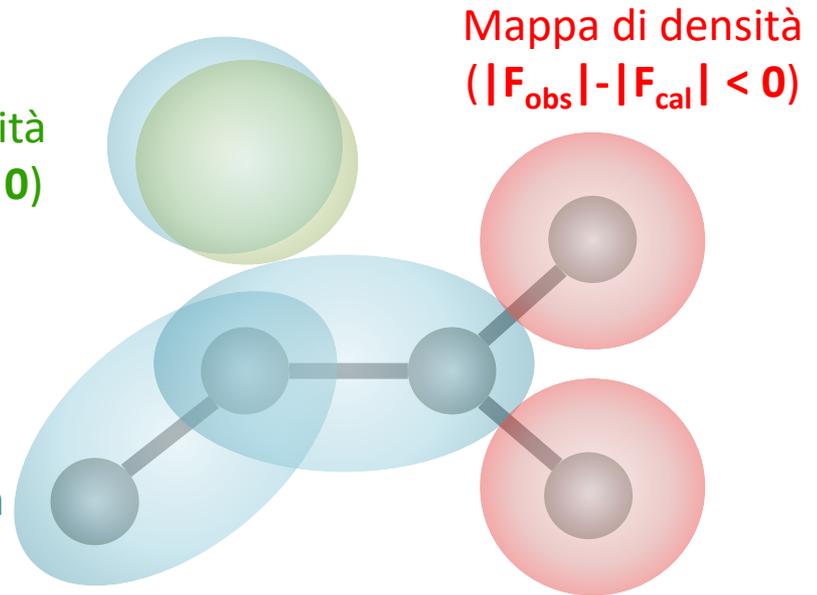
$|F_{obs}| - |F_{cal}|$  = mappa DIFFERENZA (visualizzata in **verde** se  $> 0$ , in **rosso** se  $< 0$ )



Proteina  $|F_{obs}|$

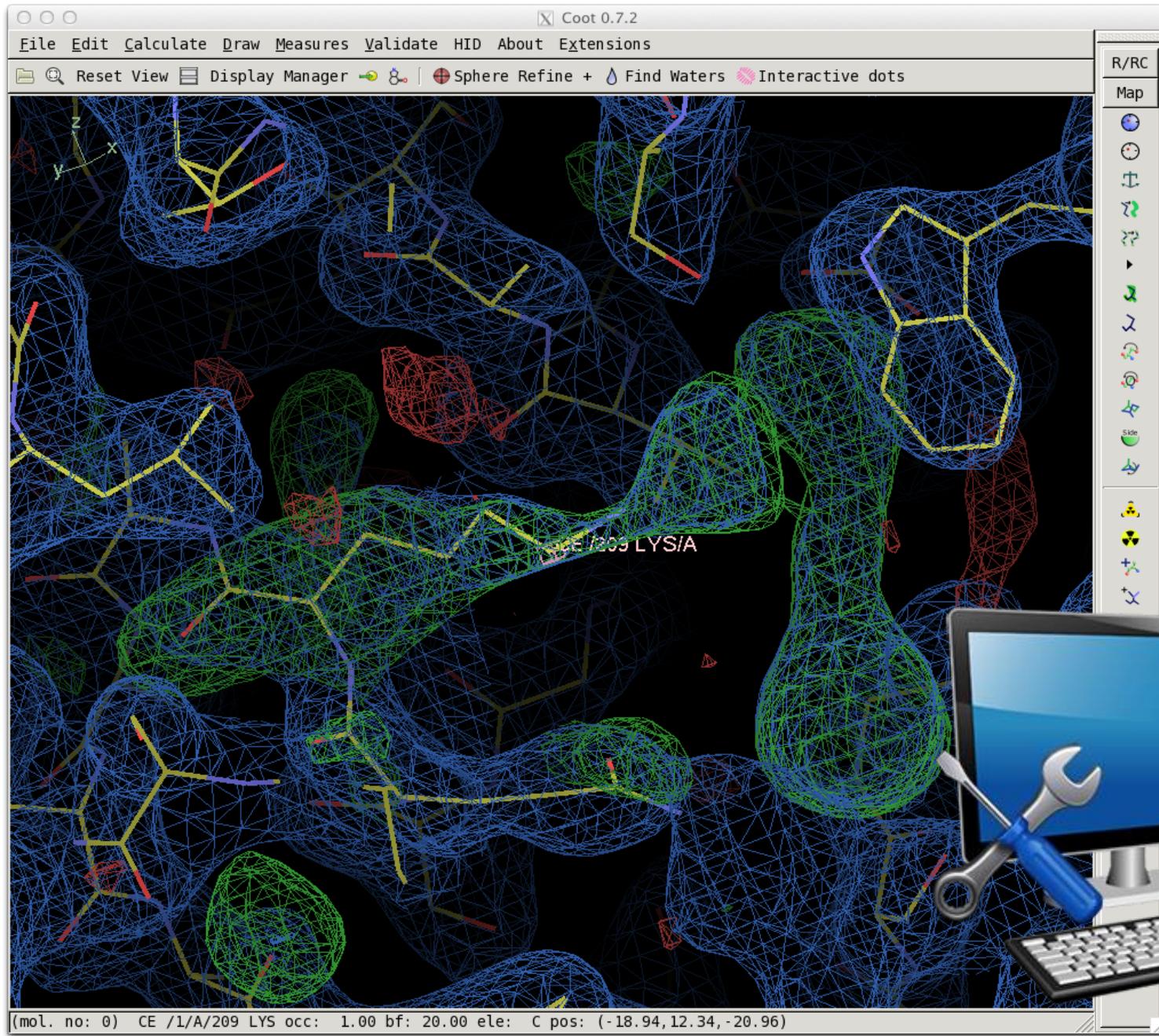
Mappa di densità  
( $|F_{obs}| - |F_{cal}| > 0$ )

Mappa di densità  
( $2|F_{obs}| - |F_{cal}|$ )

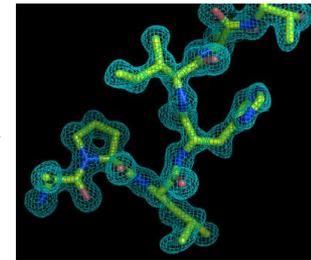


Modello  $|F_{cal}|$

# MODEL BUILDING



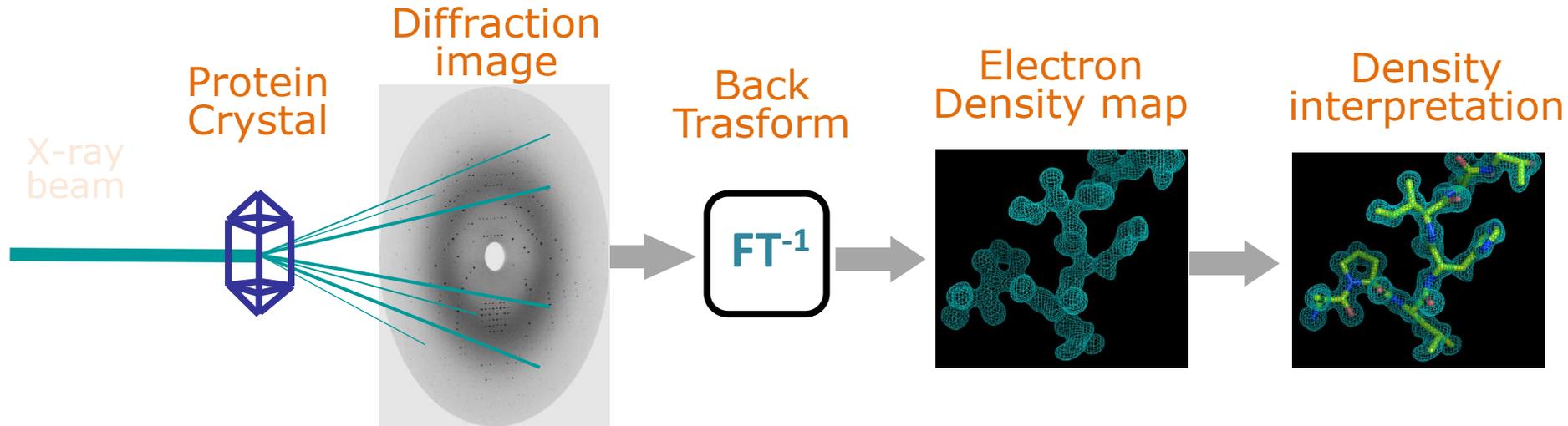
Density interpretation



IS FUN



# MODEL BUILDING AND MAPS REFINEMENT



**MODEL BUILDING** may be done automatically but in most cases is **done by HUMANS**.



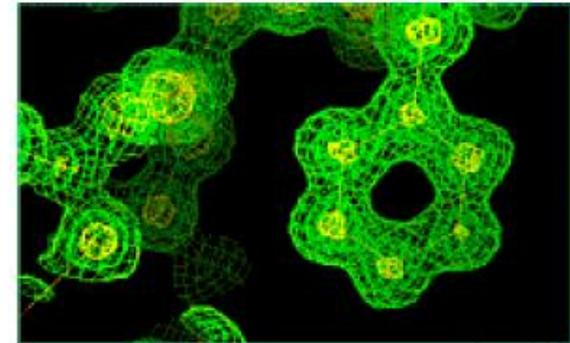
1. Build a model and Calculate new maps
2. Refine the model and get better maps

Errors in map interpretation depend on data **resolution** and **quality**

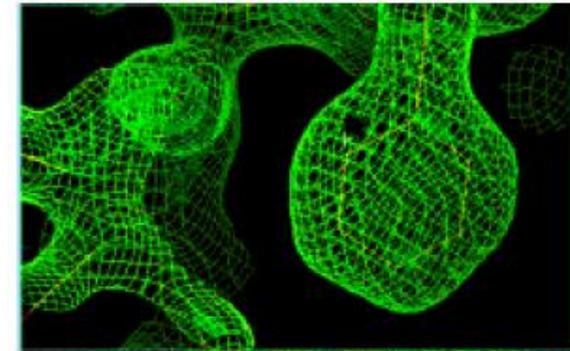


ALTA RISOLUZIONE →  
MAPPE MIGLIORI

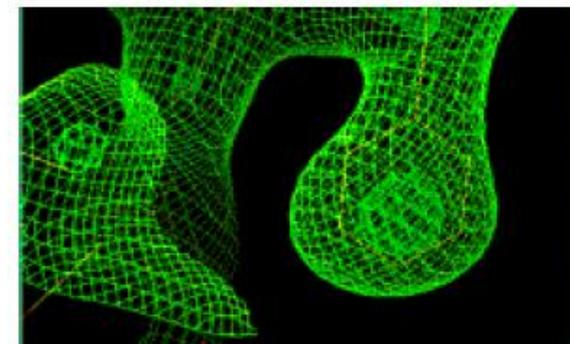
PROTEIN FEATURES	RES. (Å)
Alpha-helix	9
Beta-sheet	4
'random' main-chain	3.7
Aromatic side-chains	3.5
Interpretable conformation of side-chains	2.9
Ordered water molecules	2.7
Correct stereochemistry at Cbeta of isoleucine	2.2
Individual atoms	1.5



1.2 Å



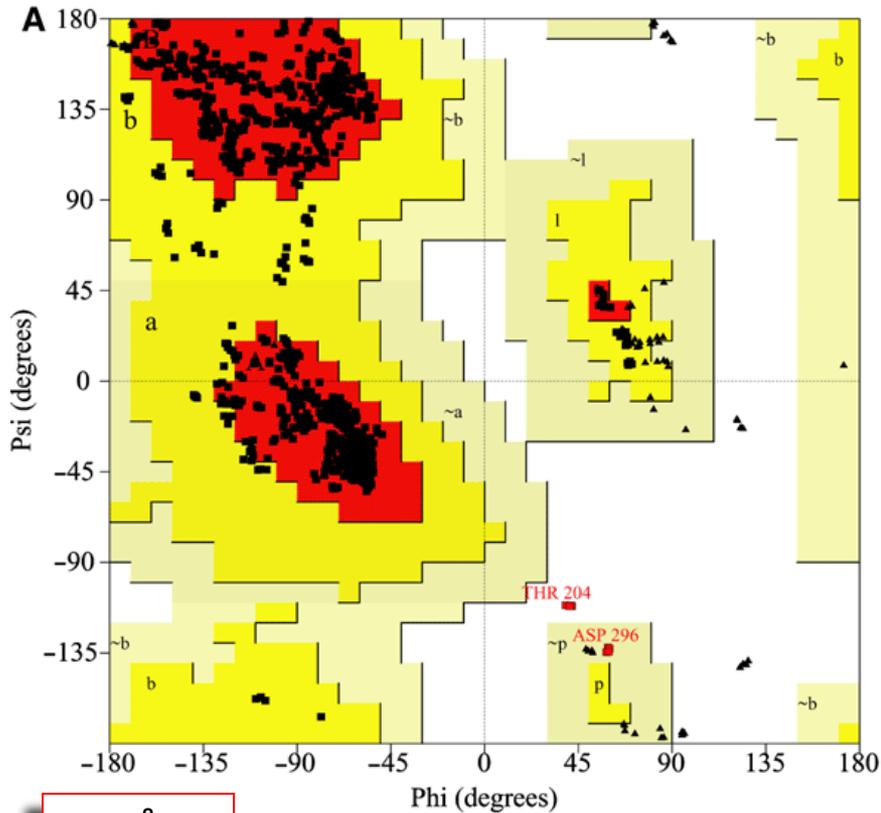
2.0 Å



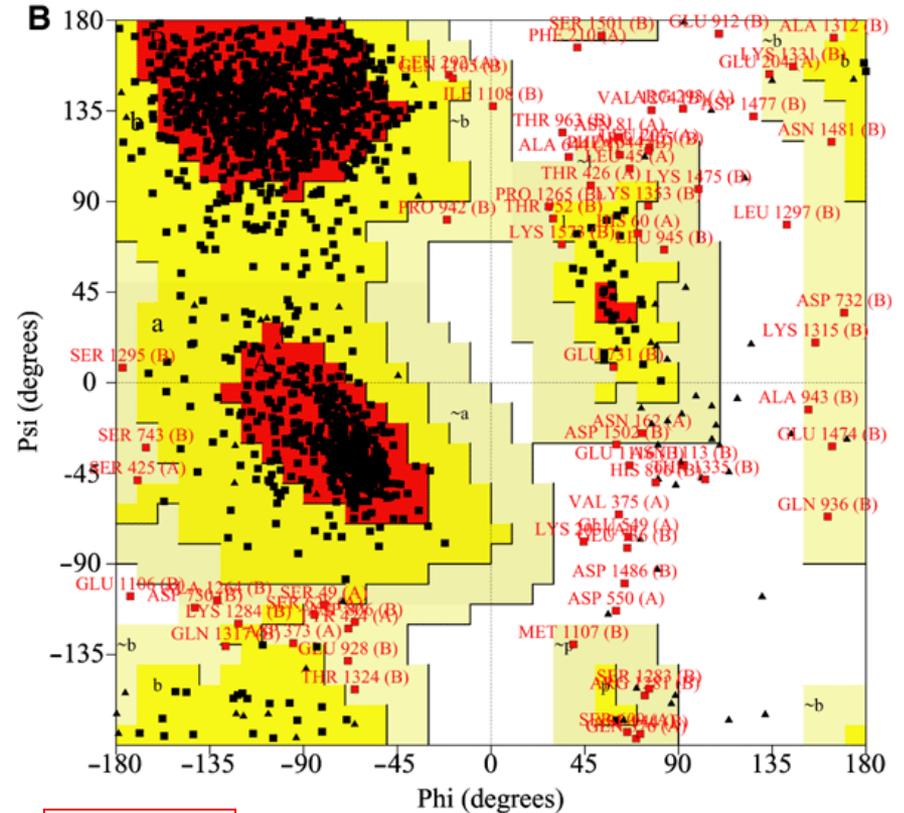
3.0 Å

# VALIDAZIONE DELLA STRUTTURA

## RAMACHANDRAN PLOT

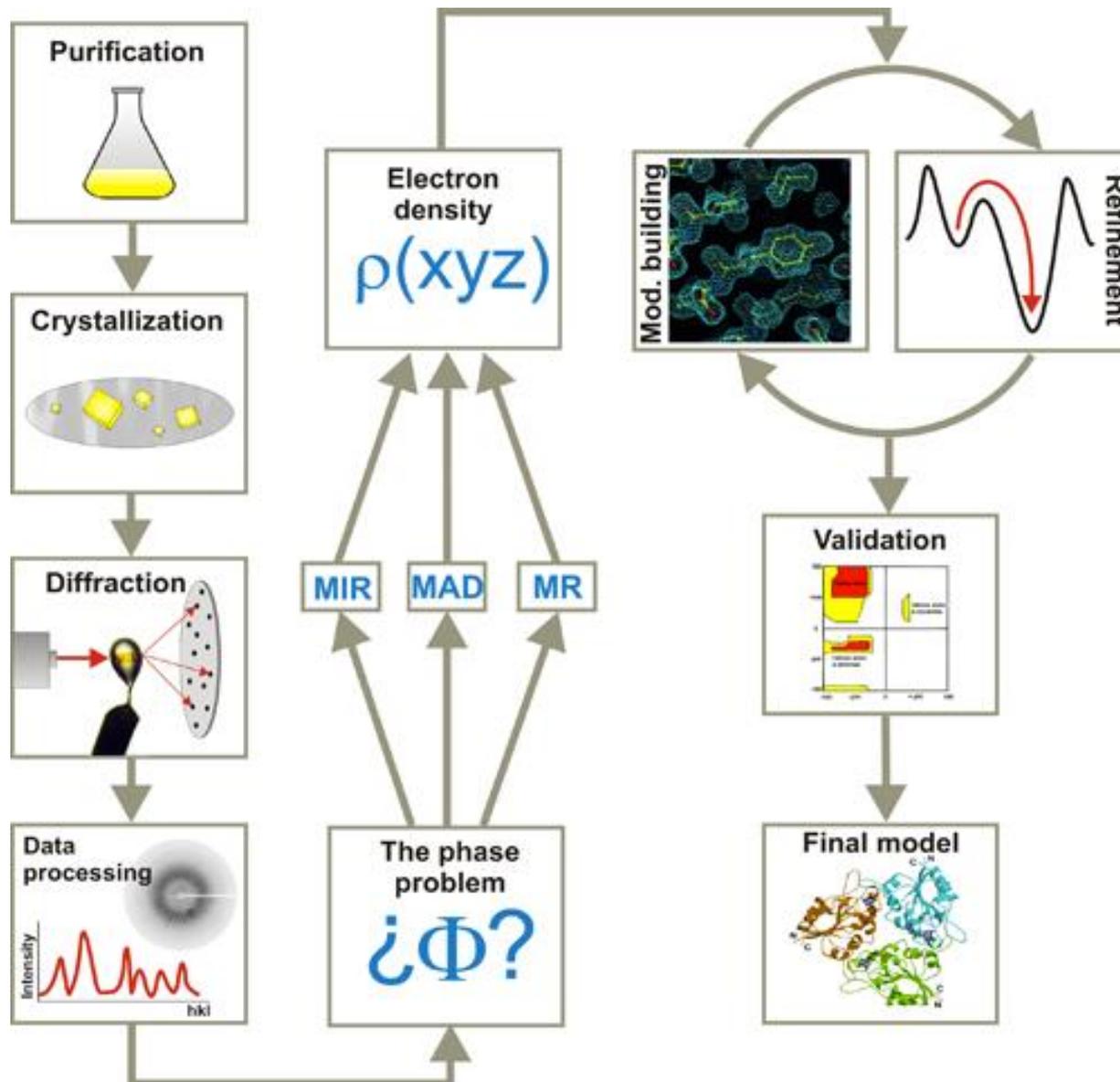


1.2 Å



2.2 Å

# CRISTALLOGRAFIA: THE SUPER SUMMARY



**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

Data Collection	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	128.9, 132.6, 88.4
Resolution (Å)	30–2.71 (2.85–2.71)
R <sub>merge</sub> (%)	9.8 (43.6)
I/σ(I)	10.2 (1.9)
Completeness (%)	95.0 (78.0)
Redundancy	2.9 (2.4)
Refinement	
Number of reflections	39,043
R <sub>work</sub> /R <sub>free</sub>	21.7/25.4
Number of atoms	
Protein	7,012
Ligands	272
Water	14
B factors (Å <sup>2</sup> )	
Protein	41.0
Ligands	43.7
Water	29.2
Rmsds	
Bond lengths (Å)	0.011
Bond angles (°)	1.5

Data in parentheses belong to the outer resolution shell.

QUALITÀ  
DATI SPERIMENTALI



QUALITÀ  
AFFIDABILITÀ  
MODELLO

**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

Data Collection

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	128.9, 132.6, 88.4

Resolution (Å) 30-2.71 (2.83-2.71)

R<sub>merge</sub> (%) 9.8 (43.6)

I/ $\sigma$ (I) 10.2 (1.9)

Completeness (%) 95.0 (78.0)

Redundancy 2.9 (2.4)

Refinement

Number of reflections 39,043

R<sub>work</sub>/R<sub>free</sub> 21.7/25.4

Number of atoms

Protein 7,012

Ligands 272

Water 14

B factors (Å<sup>2</sup>)

Protein 41.0

Ligands 43.7

Water 29.2

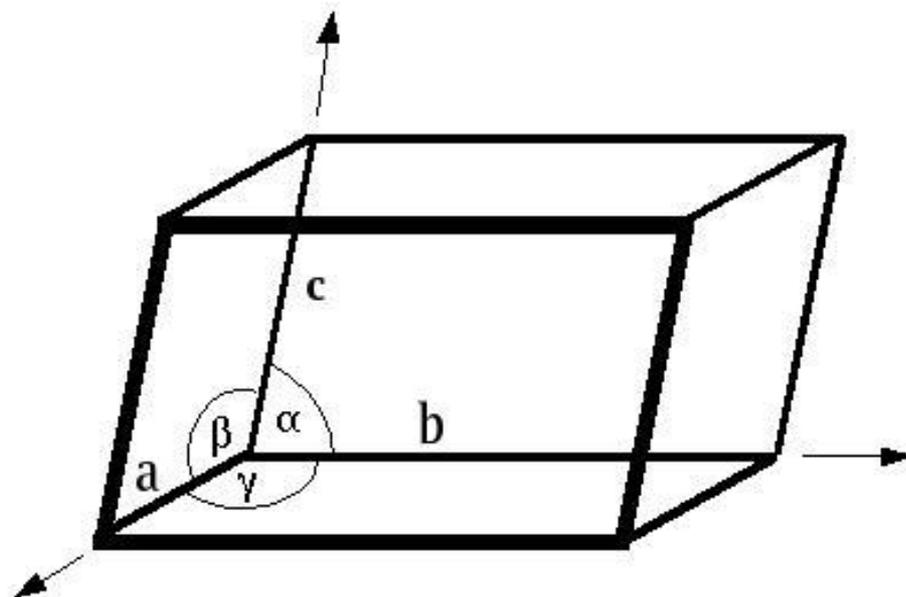
Rmsds

Bond lengths (Å) 0.011

Bond angles (°) 1.5

Data in parentheses belong to the outer resolution shell.

## QUALITÀ DATI SPERIMENTALI



**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

## Data Collection

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	128.9, 132.6, 88.4
Resolution (Å)	30–2.71 (2.85–2.71)
R <sub>merge</sub> (%)	9.8 (43.6)
I/σ(I)	10.2 (1.9)
Completeness (%)	95.0 (78.0)
Redundancy	2.9 (2.4)

## Refinement

Number of reflections	39,043
R <sub>work</sub> /R <sub>free</sub>	21.7/25.4
Number of atoms	
Protein	7,012
Ligands	272
Water	14
B factors (Å <sup>2</sup> )	
Protein	41.0
Ligands	43.7
Water	29.2
Rmsds	
Bond lengths (Å)	0.011
Bond angles (°)	1.5

Data in parentheses belong to the outer resolution shell.

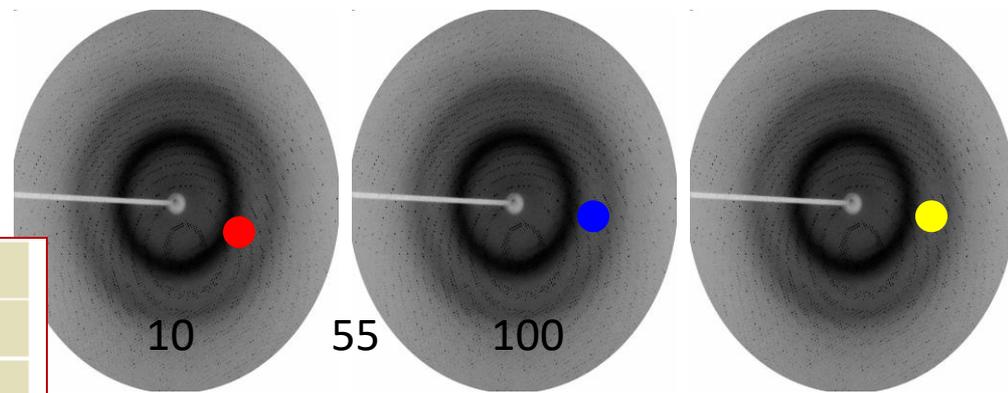
PROTEIN FEATURES	RES. (Å)
Alpha-helix	9
Beta-sheet	4
‘random’ main-chain	3.7
Aromatic side-chains	3.3
Interpretable conformation of side-chains	2.9
Ordered water molecules	2.5
Correct stereochemistry at Cbeta of isoleucine	2.2
Individual atoms	1.5

**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

Data Collection	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	128.9, 132.6, 88.4
Resolution (Å)	30–2.71 (2.85–2.71)
R <sub>merge</sub> (%)	9.8 (43.6)
I/σ(I)	10.2 (1.9)
Completeness (%)	95.0 (78.0)
Redundancy	2.9 (2.4)
Number of reflections	39,043
R <sub>work</sub> /R <sub>free</sub>	21.7/25.4
Number of atoms	
Protein	7,012
Ligands	272
Water	14
B factors (Å <sup>2</sup> )	
Protein	41.0
Ligands	43.7
Water	29.2
Rmsds	
Bond lengths (Å)	0.011
Bond angles (°)	1.5

Data in parentheses belong to the outer resolution shell.

## QUALITÀ DATI SPERIMENTALI



$$I_1(234) \quad I_2(234) \quad I_3(234)$$

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^n |I(hkl) - I_i(hkl)|}{\sum_{hkl} N \times I(hkl)}$$

**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

Data Collection	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	
Resolution (Å)	
R <sub>merge</sub> (%)	
I/σ(I)	
Completeness (%)	95.0 (78.0)
Redundancy	2.9 (2.4)
Refinement	
Number of reflections	39,043
R <sub>work</sub> /R <sub>free</sub>	21.7/25.4
Number of atoms	
Protein	7,012
Ligands	272
Water	14
B factors (Å <sup>2</sup> )	
Protein	41.0
Ligands	43.7
Water	29.2
Rmsds	
Bond lengths (Å)	0.011
Bond angles (°)	1.5

Data in parentheses belong to the outer resolution shell.

$$\rho(xyz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

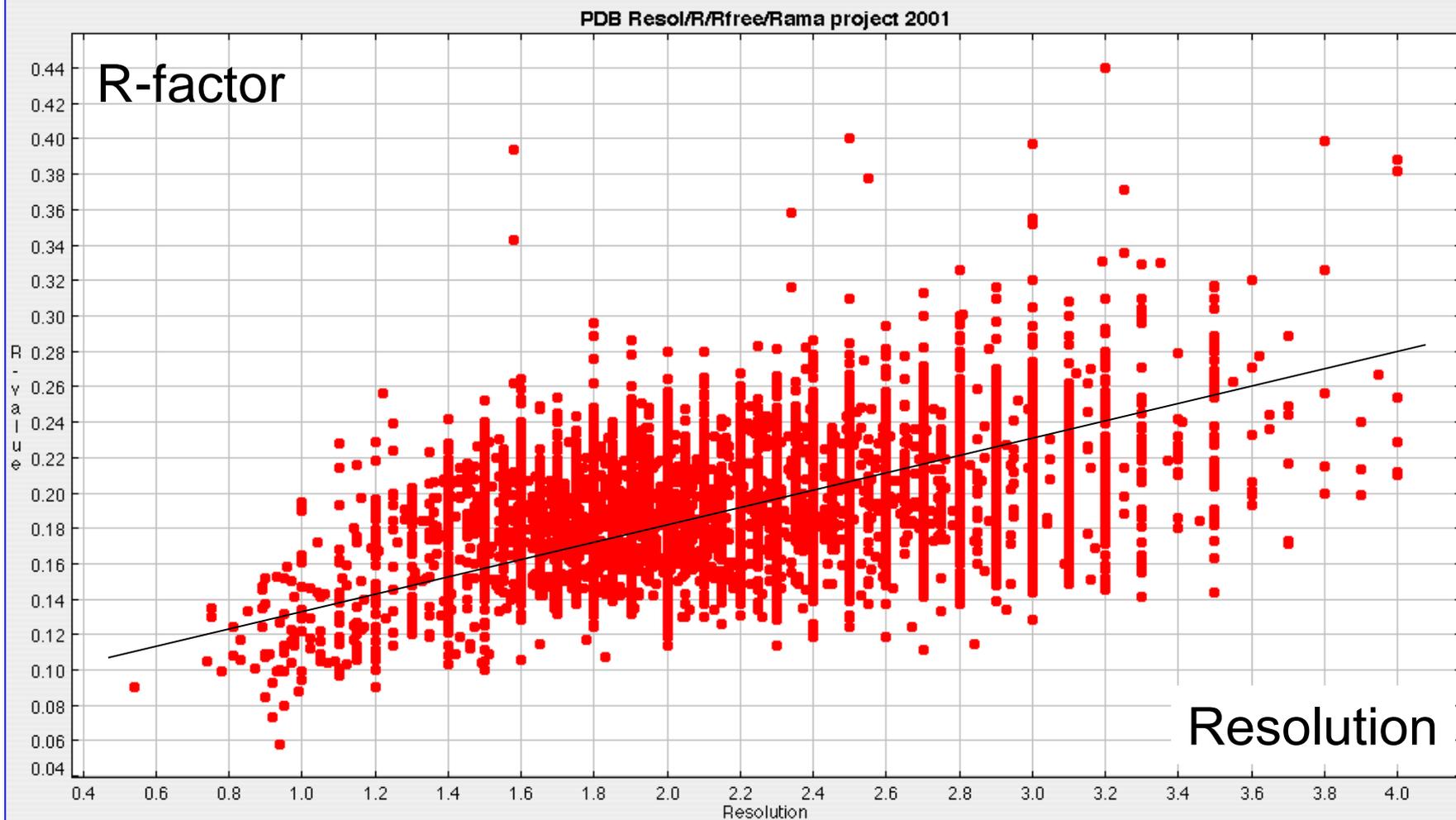
$|F_{obs}|$

DATI  
sperimentali

$|F_{cal}|$

$\alpha_{cal}$  MODELLO

$$R_F = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|}$$



X-parameter: Resolution

Y-parameter: R-value

Plotype: Scatter-plot

Version:  
Marks:  
Enable links:

Harry Plotter v.1.0

dots

Plotter

Reset

Current Mousepointer Position:

Resolution: 1,8759  
R-value: 0,0535

Help

Edit

Print

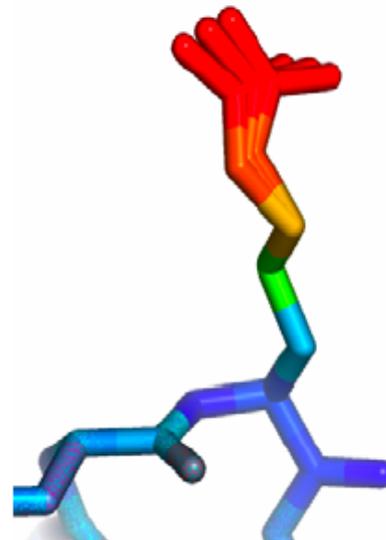
Clear

**Table 1. Data Collection and Refinement Statistics of Activated PleD in Complex with c-di-GMP and GTP $\alpha$ S**

Data Collection	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
a, b, c (Å)	128.9, 132.6, 88.0
Resolution (Å)	30–2.71 (2.85–2.60)
R <sub>merge</sub> (%)	9.8 (43.6)
I/σ(I)	10.2 (1.9)
Completeness (%)	95.0 (78.0)
Redundancy	2.9 (2.4)
Refinement	
Number of reflections	39,043
R <sub>work</sub> /R <sub>free</sub>	21.7/25.4
Number of atoms	
Protein	7,012
Ligands	272
Water	14
B factors (Å <sup>2</sup> )	
Protein	41.0
Ligands	43.7
Water	29.2
Rmsds	
Bond lengths (Å)	0.011
Bond angles (°)	1.5

Data in parentheses belong to the outer resolution shell.

Temperature Factor is a measure of motion or disorder



Colour atoms by B-factor

low amount of motion  
high amount of order  
low B-factor (5-10)



high amount of motion  
low amount of order  
high B-factor (50-100)

# PDB ENTRY

Structure Summary

3D View

Annotations

Sequence

Sequence Similarity

Structure Similarity

Experiment

Literature

Biological Assembly 1 ?



View in 3D: [NGL](#) or [JSmol](#) or [PV](#) (in Browser)

### Standalone Viewers

[Simple Viewer](#) [Protein Workshop](#)  
[Ligand Explorer](#) [Kiosk Viewer](#)

## 2WB4

### ACTIVATED DIGUANYLATE CYCLASE PLED IN COMPLEX WITH C-DI-GMP

DOI: [10.2210/pdb2wb4/pdb](https://doi.org/10.2210/pdb2wb4/pdb)

Classification: [TRANSFERASE](#)

Deposited: 2009-02-20 Released: 2010-04-21

Deposition author(s): [Wassmann, P.](#), [Schirmer, T.](#)

Organism: [Caulobacter vibrioides](#)

Expression System: ESCHERICHIA COLI

Structural Biology Knowledgebase: [2WB4 \(52 models >21 annotations\)](#) [SBKB.org](#)

### Experimental Data Snapshot

Method: X-RAY DIFFRACTION

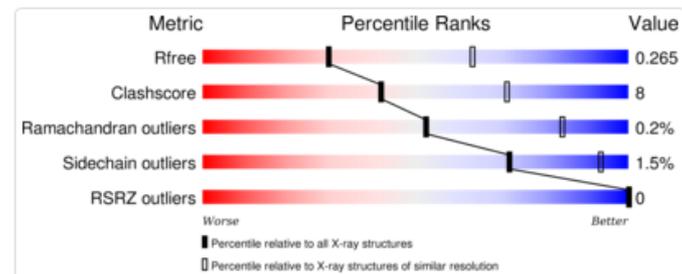
Resolution: 2.8 Å

R-Value Free: 0.268

R-Value Work: 0.239

### wwPDB Validation

[3D Report](#) [Full Report](#)



# PDB ENTRY

Sequence

Sequence Similarity

Structure Similarity

Experiment

Literature

## 2WB4

### ACTIVATED DIGUANYLATE CYCLASE PLED IN COMPLEX WITH

DOI: [10.2210/pdb2wb4/pdb](https://doi.org/10.2210/pdb2wb4/pdb)

Classification: [TRANSFERASE](#)

Deposited: 2009-02-20 Released: 2010-04-21

Deposition author(s): [Wassmann, P.](#), [Schirmer, T.](#)

Organism: [Caulobacter vibrioides](#)

Expression System: ESCHERICHIA COLI

Structural Biology Knowledgebase: [2WB4 \(52 models >21 annotations\)](#) [SBKB.org](#)

#### Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.8 Å

R-Value Free: 0.268

R-Value Work: 0.239

Display Files ▾

Download Files ▾

FASTA Sequence

PDB Format

PDB Format (gz)

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

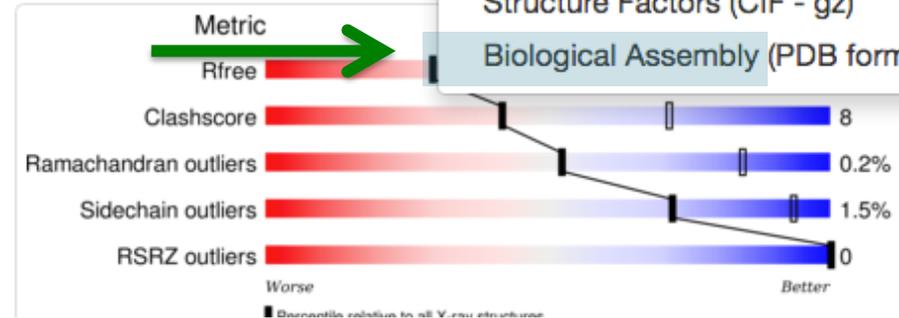
PDBML/XML Format (gz)

Structure Factors (CIF)

Structure Factors (CIF - gz)

Biological Assembly (PDB format - gz) (A+S)

#### wwPDB Validation



# PDB ENTRY

## STRUCTURE QUALITY AT A GLANCE

Structure Summary

3D View

Annotations

Sequence

Sequence Similarity

Structure Similarity

Experiment

Literature

Biological Assembly 1 ?

### 2WB4

ACTIVATED DIGUANYLATE CYCLASE PLED IN COMPLEX WITH G-DI-GMP

Display Files

Download Files

### Experimental Data Snapshot

**Method:** X-RAY DIFFRACTION

**Resolution:** 2.8 Å

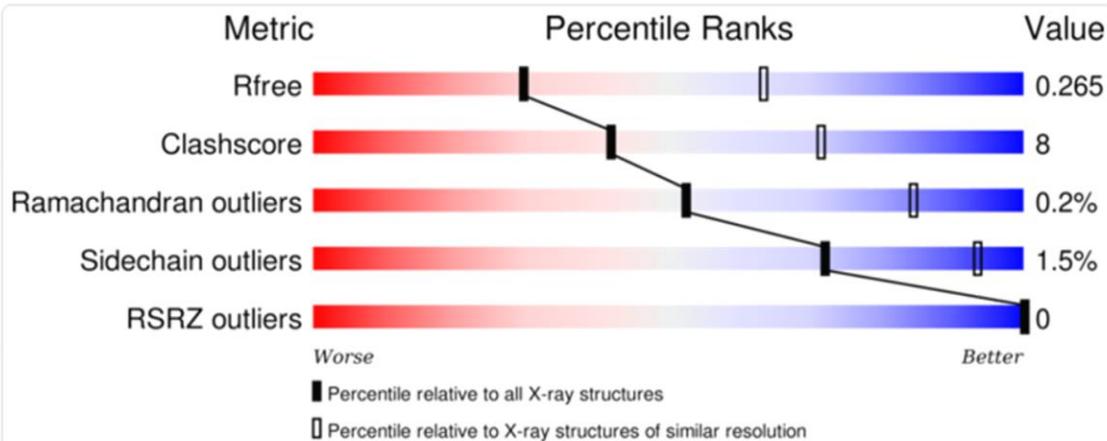
**R-Value Free:** 0.268

**R-Value Work:** 0.239

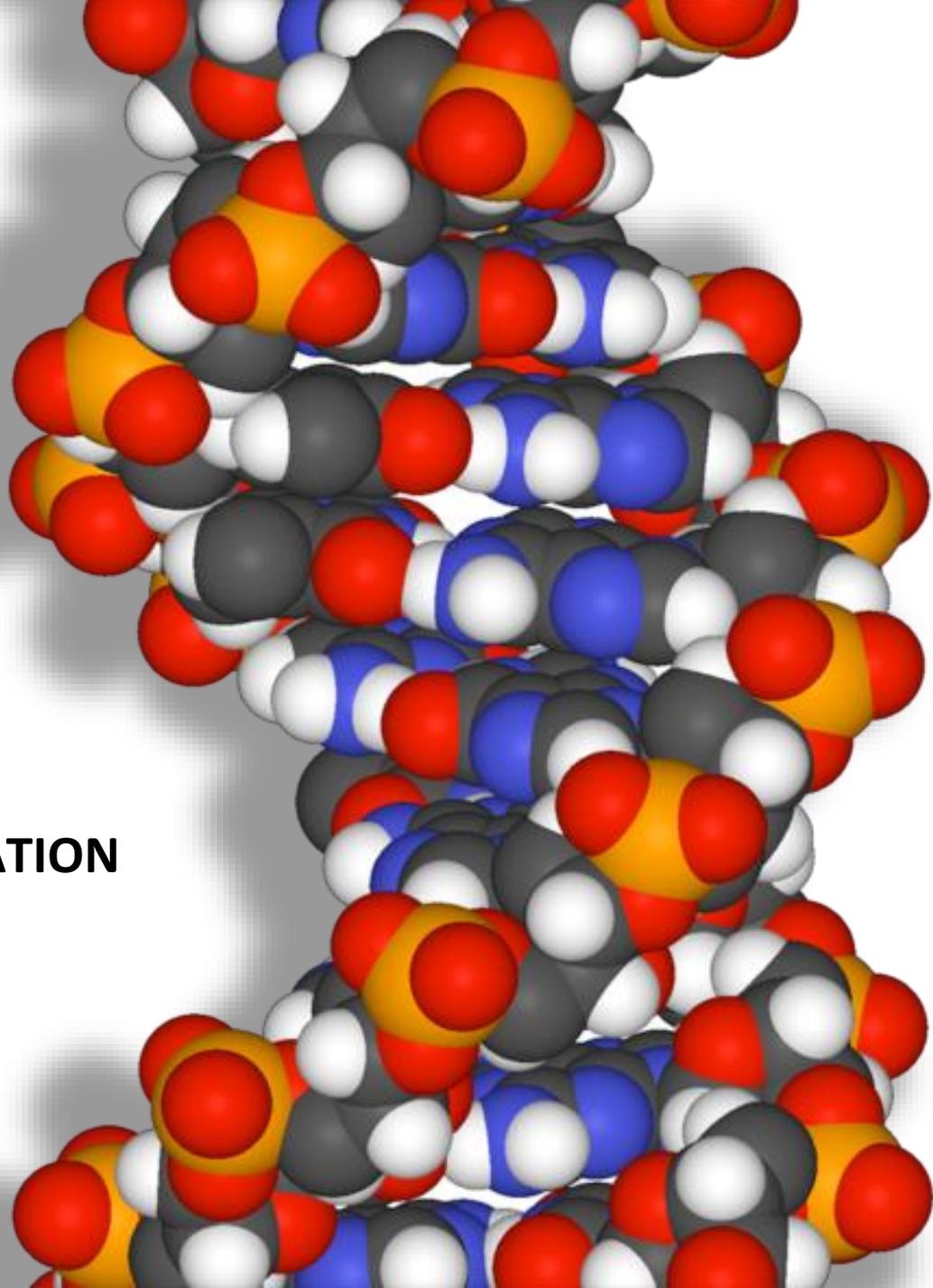
### wwPDB Validation

3D Report

Full Report

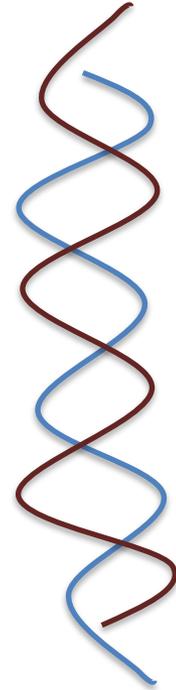
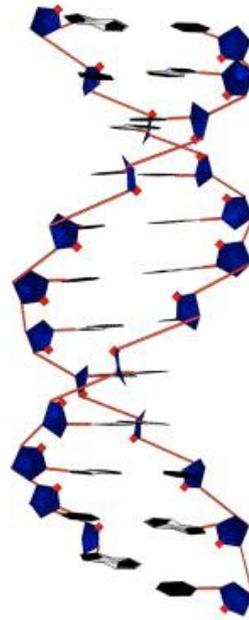
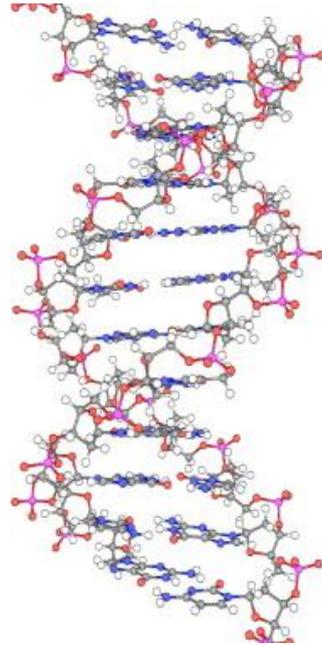
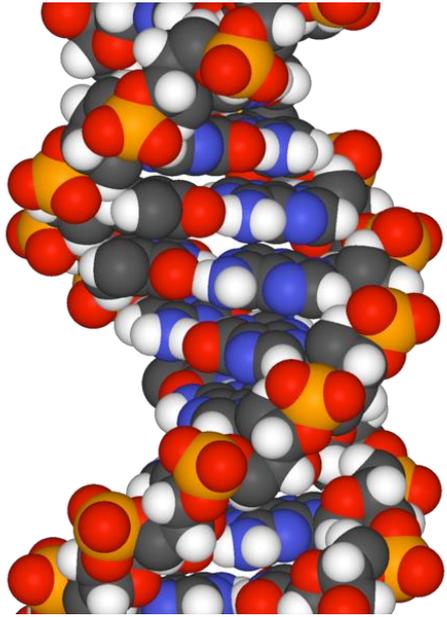


## STRUCTURE VISUALIZATION



# STRUCTURE VISUALIZATION

## Models and Conventions



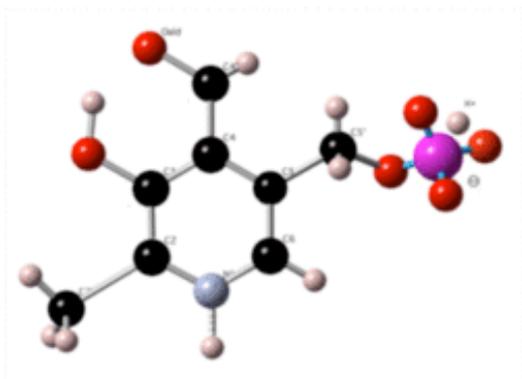
# STRUCTURE VISUALIZATION

## Structure file:

Is a text file containing

**3 coordinates (x,y,z) per atom,**

And information on chemical nature  
and connectivity



-4.4457	-1.0032	0.4113	O
-2.4005	-1.9114	0.0531	O
3.9024	-0.4411	0.0362	C
3.1486	0.8194	-0.4313	C
1.6528	0.7393	-0.2472	C
-1.0946	0.5935	0.0901	C
-2.5728	0.5152	0.2719	C
3.6076	-0.7760	1.5005	C
3.5239	-1.6299	-0.8505	C
0.8690	0.2281	-1.2695	C
1.0897	1.1791	0.9405	C
-0.5135	0.1549	-1.0998	C
-0.2929	1.1057	1.1102	C
-3.3736	1.3495	-0.7486	C
-3.1004	-0.9233	0.2332	C
4.9775	-0.2527	-0.0735	H
3.3789	1.0242	-1.4854	H
3.5298	1.6907	0.1182	H

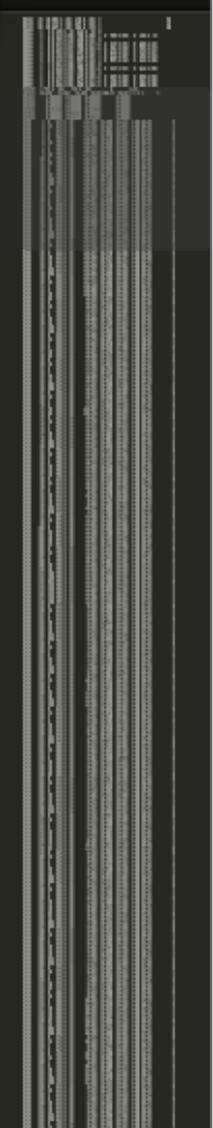
# STRUCTURE VISUALIZATION

3O05-dimer.pdb

REGISTERED

3O05-dimer.pdb

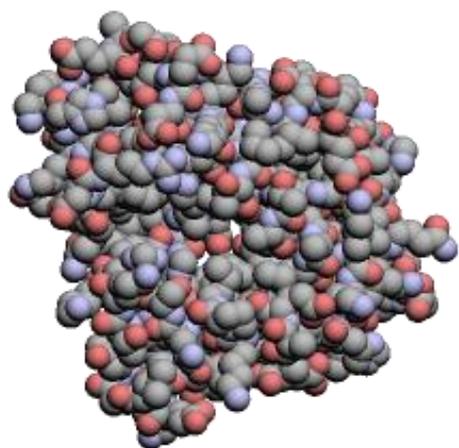
	CISPEP	1	PRE A	304	PRO A	305	0	4.53			
500	CRYST1	95.034	75.029	75.199	90.00	108.77	90.00	C 1 2 1	4		
501	ORIGX1	1.000000	0.000000	0.000000		0.000000					
502	ORIGX2	0.000000	1.000000	0.000000		0.000000					
503	ORIGX3	0.000000	0.000000	1.000000		0.000000					
504	SCALE1	0.010523	0.000000	0.003576		0.000000					
505	SCALE2	0.000000	0.013328	0.000000		0.000000					
506	SCALE3	0.000000	0.000000	0.014045		0.000000					
508	ATOM	1	N	GLY A	26	-16.197	55.295	51.112	1.00	59.88	N
509	ATOM	2	CA	GLY A	26	-15.704	56.425	51.955	1.00	60.48	C
510	ATOM	3	C	GLY A	26	-16.776	56.901	52.917	1.00	61.38	C
511	ATOM	4	O	GLY A	26	-17.477	56.088	53.519	1.00	62.82	O
512	ATOM	5	N	TRP A	27	-16.880	58.218	53.091	1.00	61.20	N
513	ATOM	6	CA	TRP A	27	-18.088	58.838	53.640	1.00	58.93	C
514	ATOM	7	C	TRP A	27	-19.173	58.964	52.562	1.00	57.15	C
515	ATOM	8	O	TRP A	27	-20.354	58.761	52.845	1.00	58.29	O
516	ATOM	9	CB	TRP A	27	-17.777	60.215	54.245	1.00	60.92	C
517	ATOM	10	N	THR A	28	-18.769	59.241	51.319	1.00	53.08	N
518	ATOM	11	CA	THR A	28	-19.714	59.248	50.206	1.00	47.95	C
519	ATOM	12	C	THR A	28	-19.586	58.045	49.260	1.00	46.26	C
520	ATOM	13	O	THR A	28	-20.582	57.534	48.757	1.00	44.20	O
521	ATOM	14	CB	THR A	28	-19.658	60.574	49.411	1.00	50.52	C
522	ATOM	15	OG1	THR A	28	-18.354	60.743	48.839	1.00	52.00	O
523	ATOM	16	CG2	THR A	28	-19.961	61.766	50.330	1.00	46.17	C
524	ATOM	17	N	GLY A	29	-18.362	57.569	49.047	1.00	42.87	N
525	ATOM	18	CA	GLY A	29	-18.110	56.599	47.993	1.00	37.98	C
526	ATOM	19	C	GLY A	29	-17.918	57.272	46.660	1.00	34.46	C
527	ATOM	20	O	GLY A	29	-18.037	56.650	45.608	1.00	35.58	O
528	ATOM	21	N	GLN A	30	-17.625	58.555	46.699	1.00	36.53	N
529	ATOM	22	CA	GLN A	30	-17.443	59.332	45.485	1.00	37.71	C
530	ATOM	23	C	GLN A	30	-16.119	60.067	45.539	1.00	37.58	C
531	ATOM	24	O	GLN A	30	-15.763	60.800	44.610	1.00	37.55	O
532	ATOM	25	CB	GLN A	30	-18.608	60.335	45.313	1.00	39.39	C
533	ATOM	26	CG	GLN A	30	-19.990	59.637	45.368	1.00	40.75	C
534	ATOM	27	CD	GLN A	30	-21.122	60.595	45.612	1.00	49.35	C
535	ATOM	28	OE1	GLN A	30	-21.611	60.727	46.742	1.00	53.43	O
536	ATOM	29	NE2	GLN A	30	-21.553	61.271	44.560	1.00	40.06	N
537	ATOM	30	N	GLU A	31	-15.431	59.929	46.670	1.00	36.94	N
538	ATOM	31	CA	GLU A	31	-14.101	60.512	46.840	1.00	37.62	C



# STRUCTURE VISUALIZATION = MOLECULAR GRAPHICS SOFTWARES

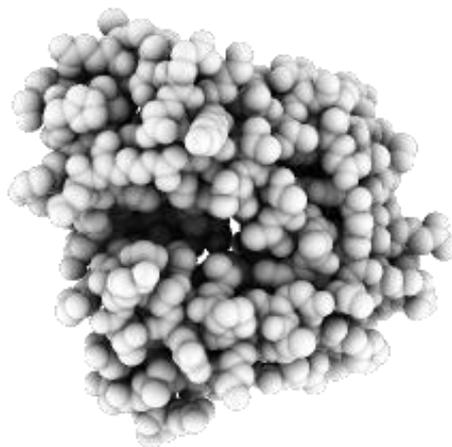
We can use the same RENDERING of non scientific 3D softwares to **VISUALIZE** our structure

-4.4457	-1.0032	0.4113	O
-2.4005	-1.9114	0.0531	O
3.9024	-0.4411	0.0362	C
3.1486	0.8194	-0.4313	C
1.6528	0.7393	-0.2472	C
-1.0946	0.5935	0.0901	C
-2.5728	0.5152	0.2719	C
3.6076	-0.7760	1.5005	C
3.5239	-1.6299	-0.8505	C
0.8690	0.2281	-1.2695	C
1.0897	1.1791	0.9405	C
-0.5135	0.1549	-1.0998	C
-0.2929	1.1057	1.1102	C
-3.3736	1.3495	-0.7486	C
-3.1004	-0.9233	0.2332	C
4.9775	-0.2527	-0.0735	H
3.3789	1.0242	-1.4854	H
3.5298	1.6907	0.1182	H



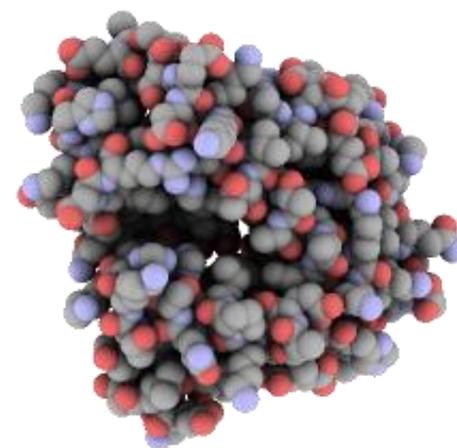
Illuminazione classica

+



Ambient occlusion

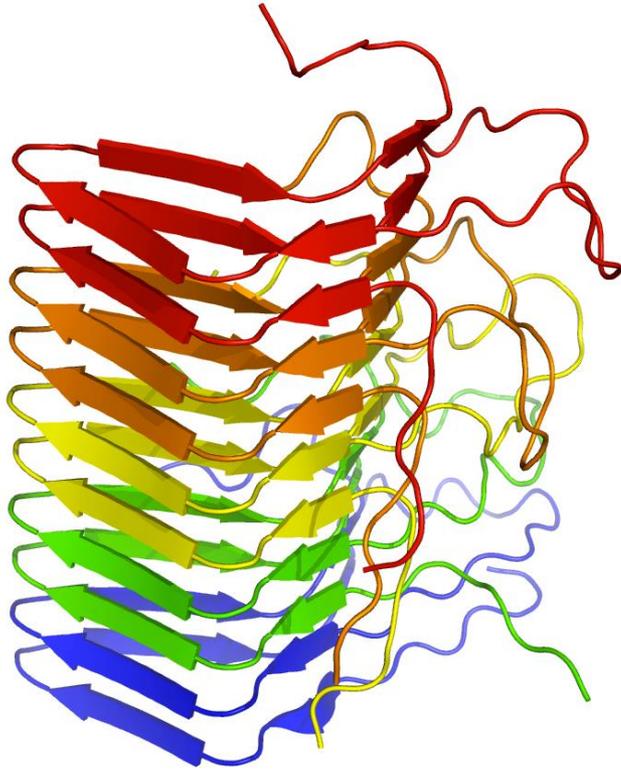
=



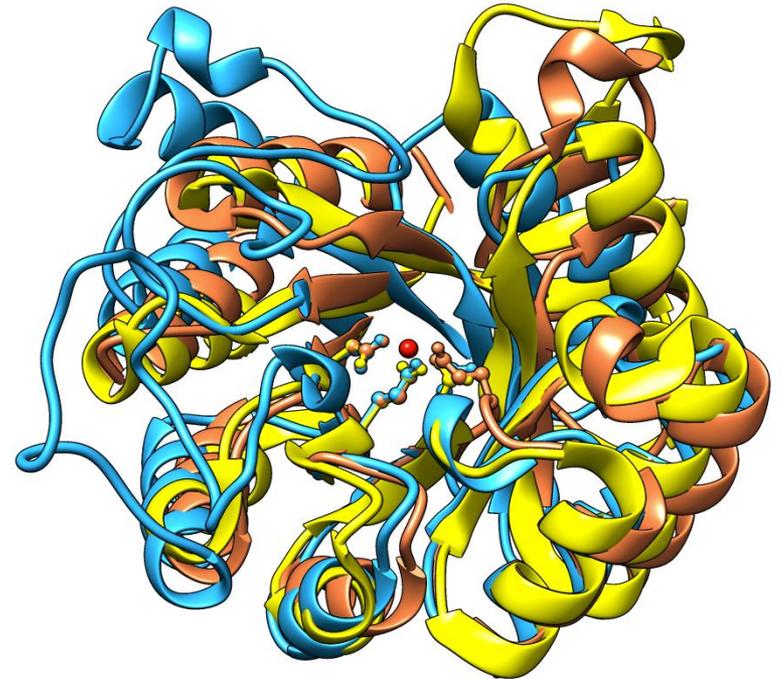
Migliore percezione della forma

# STRUCTURE VISUALIZATION = MOLECULAR GRAPHIC SOFTWARES

PyMol



CHIMERA



But also

CueMol

VMD

QuteMol

# STRUCTURE VISUALIZATION = MOLECULAR GRAPHIC SOFTWARES

PyMol

codice

Viewport

The PyMOL Molecular Graphics System <2>

File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help

```
Ray: render time: 5.10 sec. = 706.2 frames/hour (29.44 sec. accum.).
PyMOL>set ambient, 0.35
Setting: ambient set to 0.35000.
PyMOL>orient
PyMOL>ray
Ray: render time: 5.47 sec. = 658.4 frames/hour (34.91 sec. accum.).
PyMOL>ray
Ray: render time: 5.39 sec. = 668.1 frames/hour (40.30 sec. accum.).
PyMOL>
```

PyMOL Viewer <2>

all A S H L C  
1wld A S H L C  
apbs\_map A S H L C  
e\_lvl\_0\_1 A S H L C

Mouse Mode 3-Button Viewing  
Buttons L M R Wheel  
& Keys Rota Move MovZ Slab  
Shft +Box -Box Clip MovS  
Ctrl +/- PkAt Pk1 MvSZ  
CtSh Sele Orig Clip MovZ  
SngClk +/- Cent Menu  
DblClk Menu - PkAt  
Selecting Residues  
State 1/ 1

-1.000 1.000

PyMOL>\_

Fast keys

Molecule  
Caricature

**A= actions**  
**S>Show**  
**H=hide**  
**L=label**  
**C=color**

# STRUCTURE VISUALIZATION = MOLECULAR GRAPHIC SOFTWARES

CHIMERA

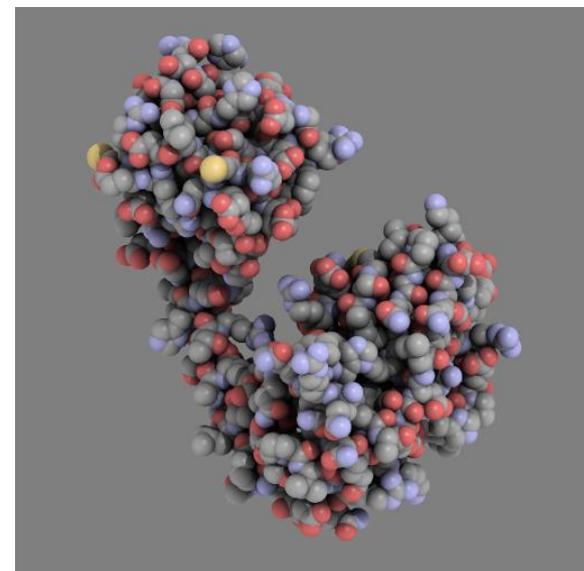
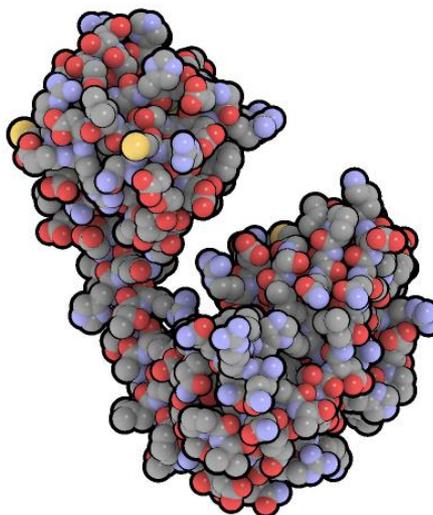
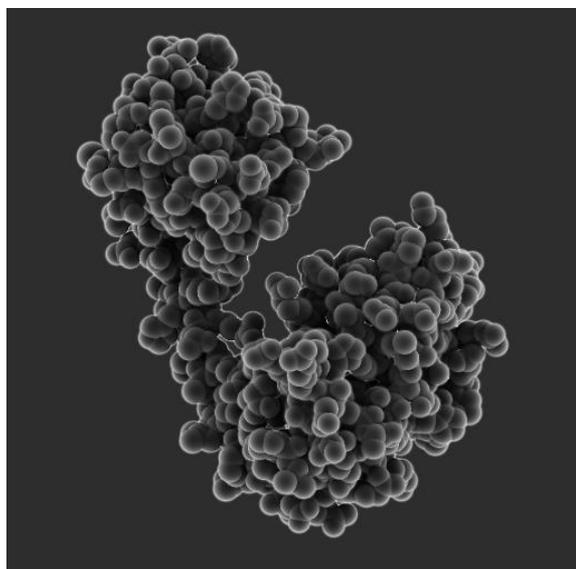
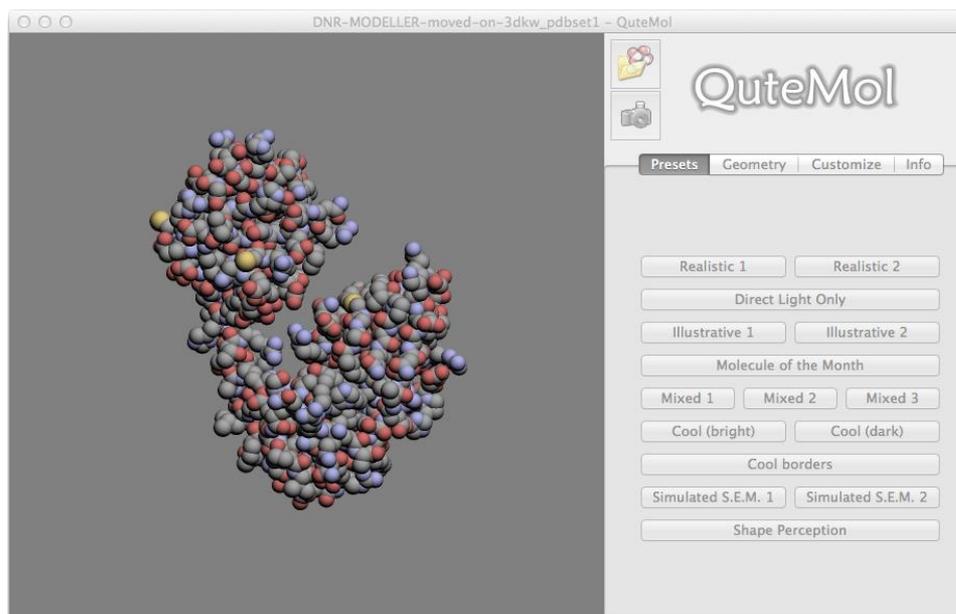
FREE



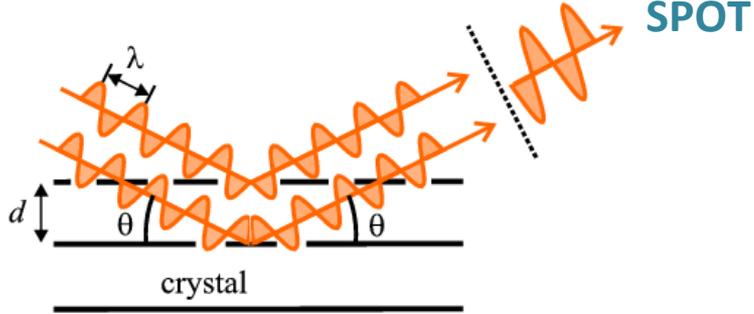
<http://www.cgl.ucsf.edu/chimera/download.html>

# STRUCTURE VISUALIZATION = MOLECULAR GRAPHIC SOFTWARES

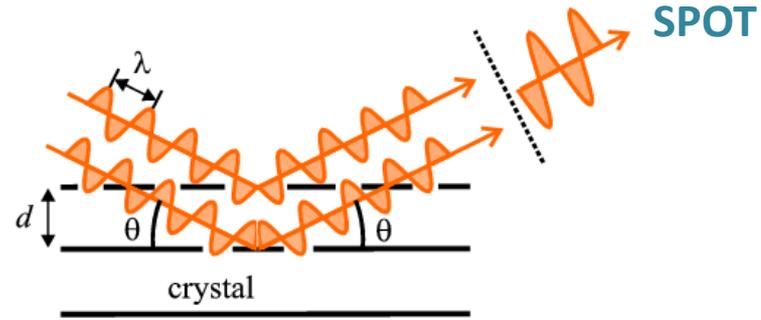
QuteMol



# ORDER VS DISORDER



# ORDER VS DISORDER



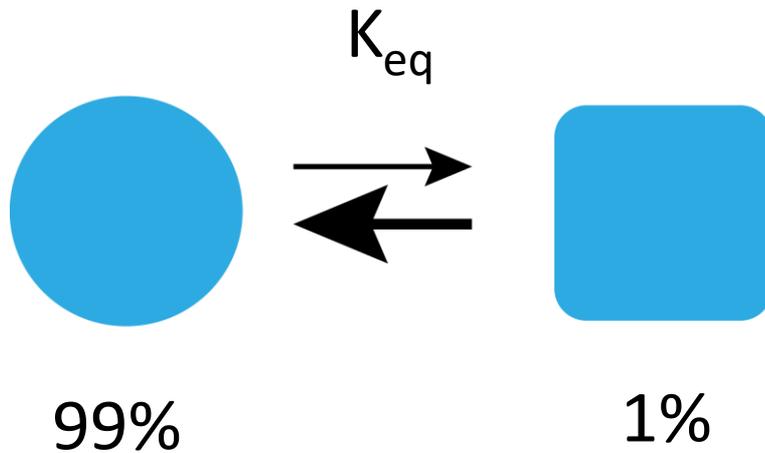
NO **O**RDER, NO **D**ENSITY!



WE MISS  
**D**YNAMIC  
PARTS!

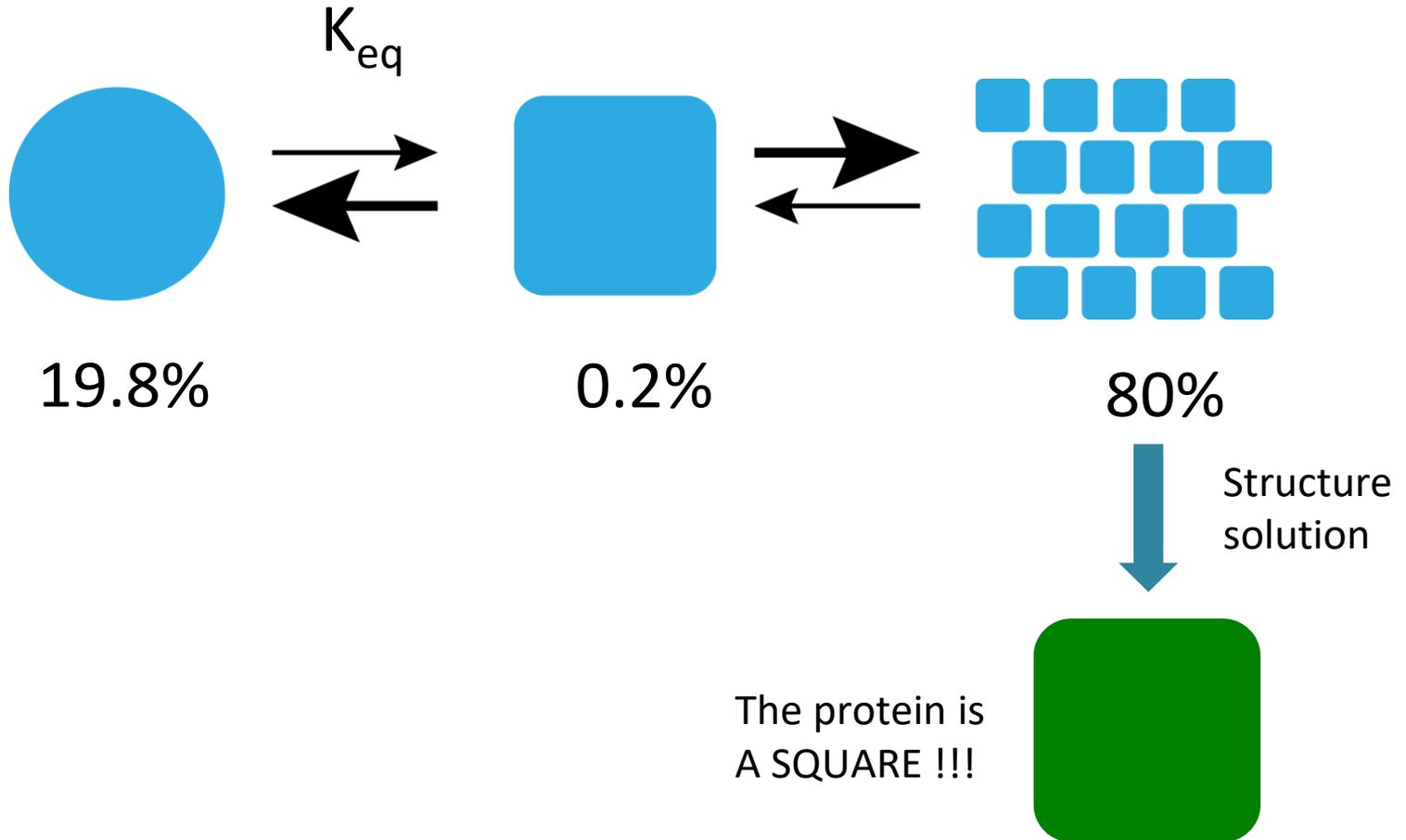
# STRUCTURE ANALYSIS – COMMON ERRORS

The conformation observed in the crystal structure is  
ONE POSSIBLE conformation and not THE conformation of the protein



# STRUCTURE ANALYSIS – COMMON ERRORS

The conformation observed in the crystal structure is  
ONE POSSIBLE conformation and not THE conformation of the protein



## VANTAGGI



- **Risoluzione atomica**
- Non ci sono limitazioni nelle dimensioni delle macromolecole

## SVANTAGGI



- **Cristalli singoli**
- Altamente ordinati
- ragionevolmente grandi (30-100  $\mu\text{m}$ )