

1 Feb. 2010, aula Caldirola

1. Cristallografia di proteine: struttura 3D di macromolecole biologiche

3 Feb. 2010, aula I

2. Virus e proteine virali: dalla struttura alla funzione biologica

8 Feb. 2010, aula Caldirola

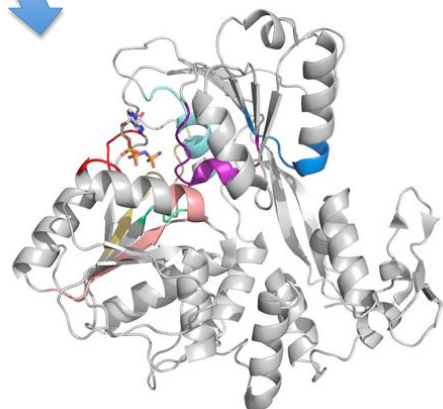
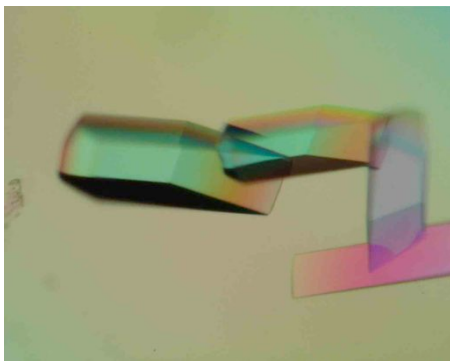
3. Proteine virali e ricerca razionale di farmaci antivirali

Cristallografia di proteine: struttura 3D di macromolecole biologiche

Mario Milani

mario.milani@mi.infm.it

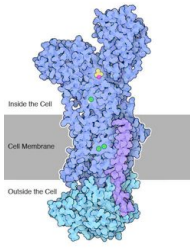
<http://digilander.libero.it/mario.milani/teaching.html>



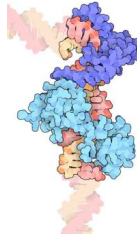
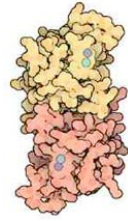
Cosa sono le proteine ?

<http://www.pdb.org/pdb/motm.do>

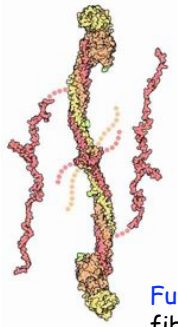
Funzioni di protezione:
detossificazione del superossido



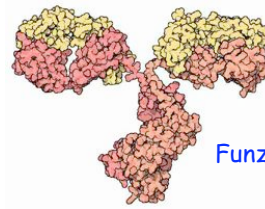
Funzioni di trasporto: pompa sodio fuori e potassio nella cellula



Funzioni di controllo: i fattori di trascrizione controllano l'espressioni dei geni

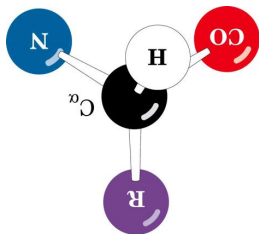


Funzioni strutturali: fibrina si assembla per bloccare il flusso sanguigno a seguito di un trauma



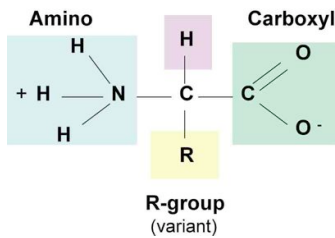
Funzioni di difesa: anticorpi

(...) e molto altro ancora !

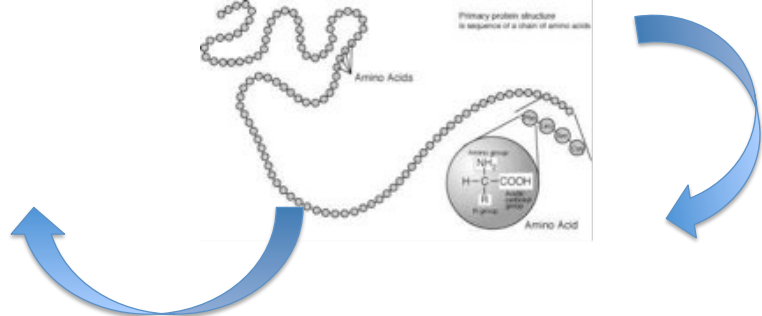
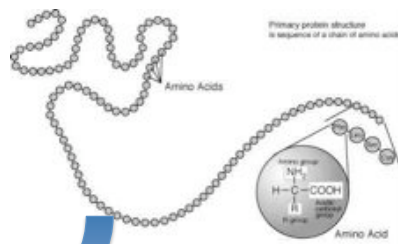
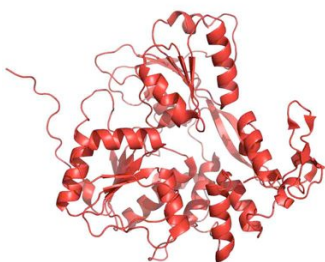
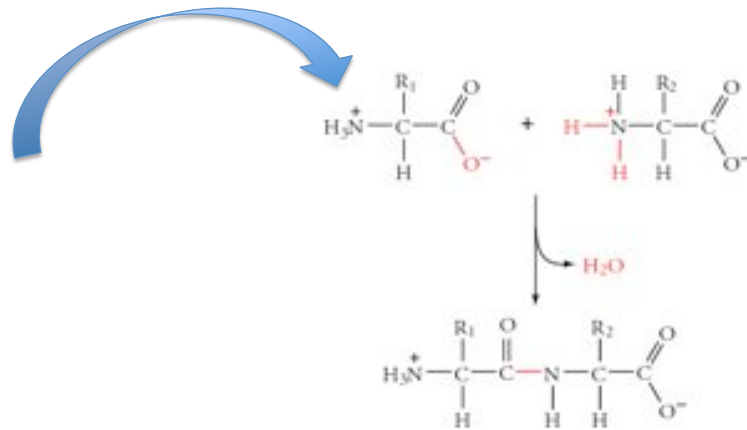


Amino Acid Structure

Hydrogen

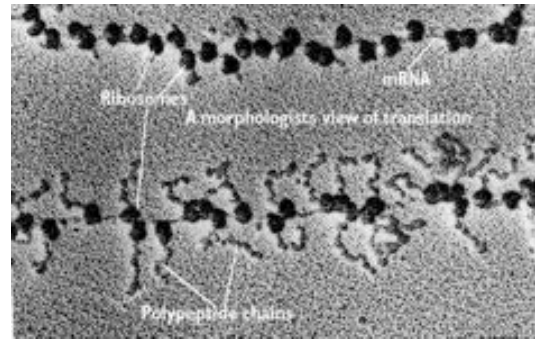
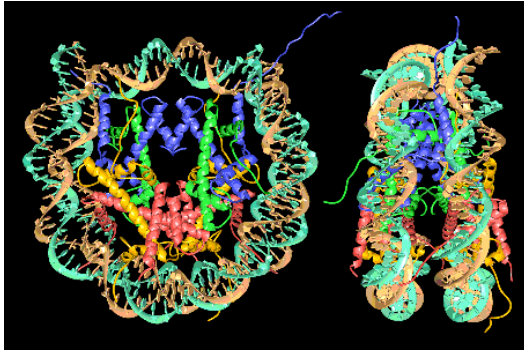


Ma in pratica come sono fatte ?

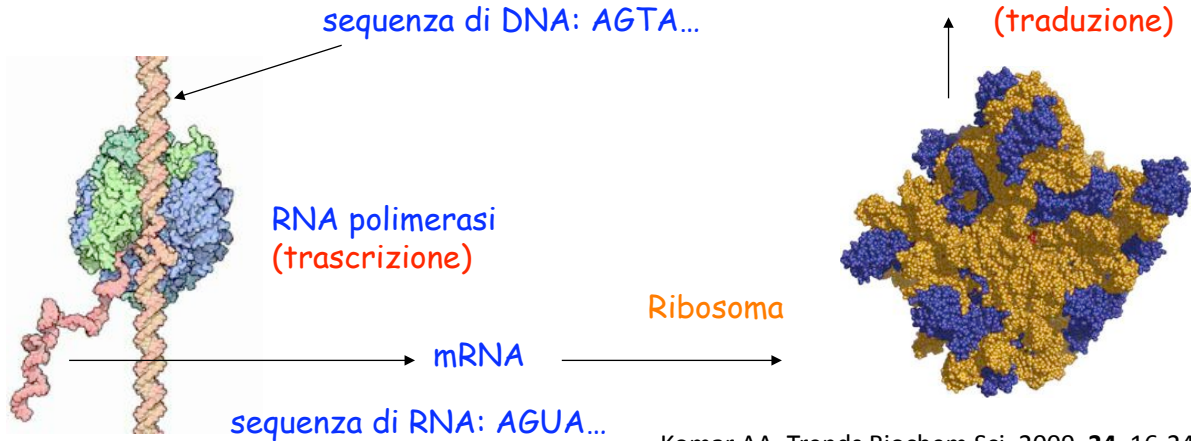


E il DNA cosa c'entra con le proteine?

DNA (nucleosomi)



sequenza di peptidi = proteine
(traduzione)



Komar AA. Trends Biochem Sci. 2009, **34**, 16-24.

Codice genetico

		Second Letter				
		T	C	A	G	
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA } Stop TAG } Stop	TGT } Cys TGC } TGA } Stop TGG } Trp	T C A G
	C	GTT } GTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	GAT } His CAC } CAA } Gln CAG }	CGT } CGC } Arg CGA } CGG }	T C A G
	A	ATT } ATC } Ile ATA } ATG } Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA } Lys AAG }	AGT } Ser AGC } AGA } Arg AGG }	T C A G
	G	GTT } GTC } Val GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA } Glu GAG }	GGT } GGC } Gly GGA } GGG }	T C A G

Cristallizzazione di proteine

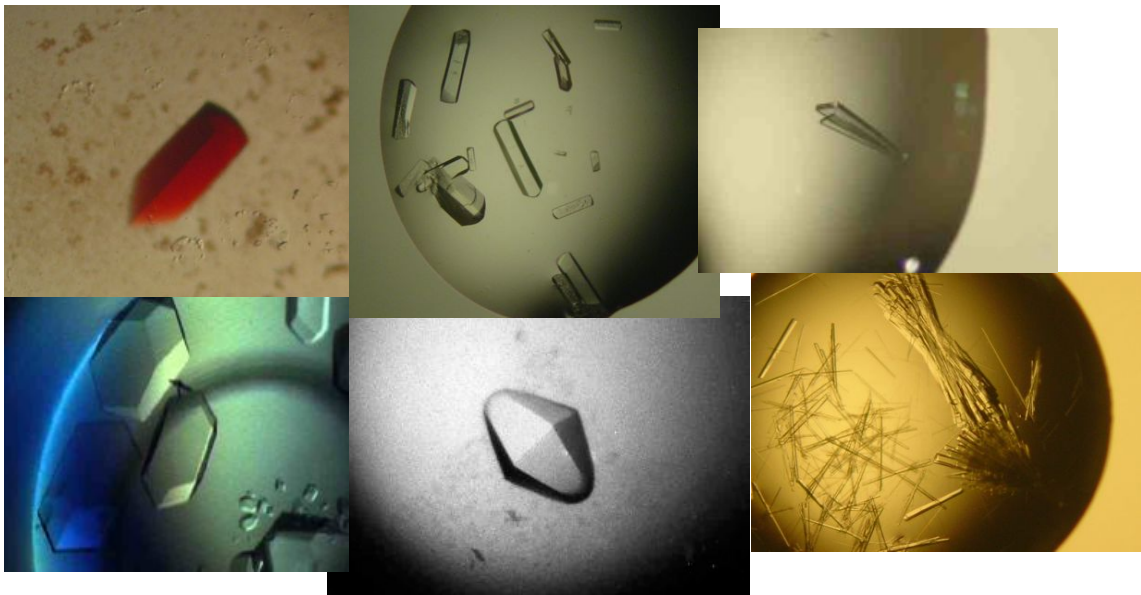
Esprimere e purificare la proteina

Scelta di un buffer in cui la proteina e' solubile e stabile (DLS, Termofluorimetria)

Portare la proteina in uno stato di *supersaturazione* (ridurre la solubilità)

Inizio della crescita di cristalli

Cristalli di proteine



volume $< 0.1 \text{ mm}^3$

periodicità del reticolo cristallino $> 50\text{-}100 \text{ \AA}$

contenuto solvente 30% - 80% v/v

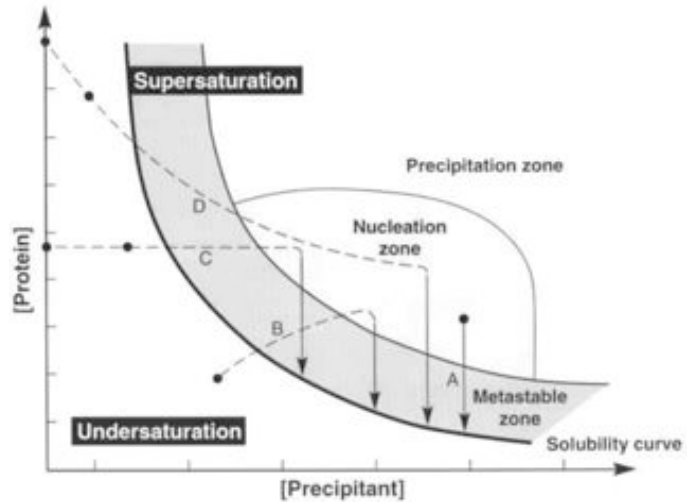
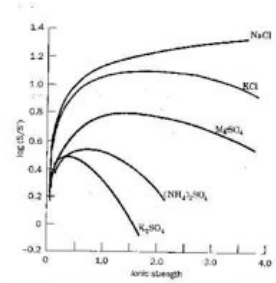
bassa stabilità meccanica ($E_{\text{stab.}} < 10 \text{ kcal/mol}$) ($E_{\text{H-bond}} \approx 1 \text{ kcal/mol}$)

Cristallizzazione di proteine

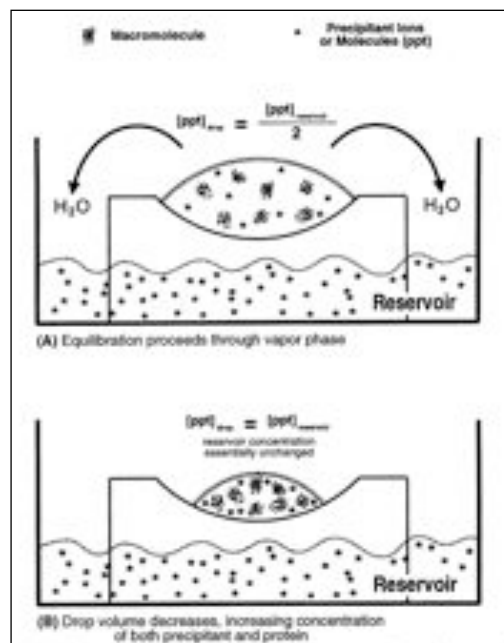
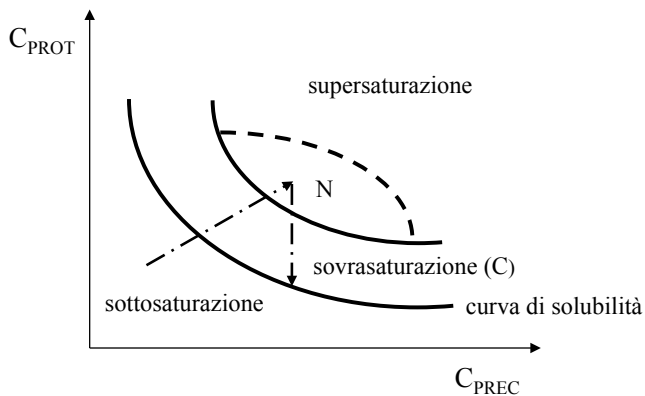
Cristallizzazione: procedura empirica in cui si riduce *lentamente* la solubilità di una proteina: essa tende dunque a "precipitare" formando una struttura cristallina (non amorfa)

La solubilità e' legata a varie proprietà chimico-fisiche :

- ✓ forza ionica ($I = 1/2 \sum_i c_i z_i^2$)
- ✓ pH
- ✓ temperatura
- ✓ concentrazione proteina / concentrazione reagenti
- ✓ costante dielettrica del solvente
- ✓ condizioni iniziali ($V_{in} \dots$)



Metodo della diffusione di vapore



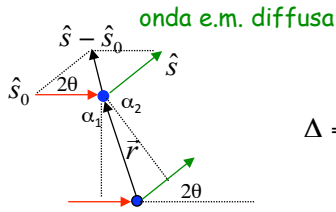
gocce 'a mano': 1 ul P + 1 ul R

gocce 'robot': 0.2 ul P + 0.1 ul R

sitting drop

atomo ●

Interferenza



onda e.m. diffusa

diff. di cammino ottico (Δ)

$$\Delta = r(\sin\alpha_1 + \sin\alpha_2) = \vec{r} \cdot (\hat{s} - \hat{s}_0) = \lambda n$$

$$|\hat{s} - \hat{s}_0| = 2 \sin \vartheta$$

Perché ci sia int. costruttiva la **diff. di cammino ottico** (Δ) deve essere pari ad un numero intero (n) di lunghezze d'onda (λ)

onda e.m. incidente

$$\Delta = \vec{r} \cdot \text{ver}(\hat{s} - \hat{s}_0) 2 \sin \vartheta = 2d \sin \vartheta = n\lambda \quad \text{Legge di Bragg}$$

\mathbf{S} = vettore di scattering

sfasamento, 2 punti:

$$\delta = \frac{2\pi}{\lambda} \Delta = \frac{2\pi}{\lambda} \vec{r} \cdot (\hat{s} - \hat{s}_0) = 2\pi \mathbf{r} \cdot \mathbf{S}$$

$$\mathbf{S} \equiv \frac{(\hat{s} - \hat{s}_0)}{\lambda}$$

generalizzazione N punti

$$A = \sum_{j=1}^N A_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S})$$

$$|\mathbf{S}| \equiv \frac{|(\hat{s} - \hat{s}_0)|}{\lambda} = \frac{2 \sin \theta}{\lambda}$$

fattore di scattering dell'atomo j

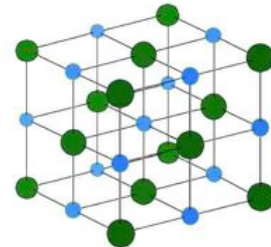
Fattore di scattering della **cella elementare**: **fattore di struttura**

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

Fattore di scattering del cristallo

Reticolo cristallino: $\mathbf{Q}_{uvw} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$

$$K(\mathbf{S}) = F(\mathbf{S}) \sum_{l=0}^{n_1} \exp(2\pi i l \mathbf{a} \cdot \mathbf{S}) \sum_{m=0}^{n_2} \exp(2\pi i m \mathbf{b} \cdot \mathbf{S}) \sum_{n=0}^{n_3} \exp(2\pi i n \mathbf{c} \cdot \mathbf{S})$$



Intensità diffusa in un esperimento di scattering

$$I \propto |K(\mathbf{S})|^2 = \left| \sum_{\text{cristallo}} e^{i2\pi \mathbf{S} \cdot \mathbf{Q}} F(\mathbf{S}) \right|^2$$

La presenza del reticolo cristallino (interferenza) seleziona solamente quei valori di \mathbf{S} che corrispondono a punti del reticolo reciproco (= \mathbf{S} deve essere un vettore di reticolo reciproco). Il **fattore di struttura** si misura solo su valori discreti !

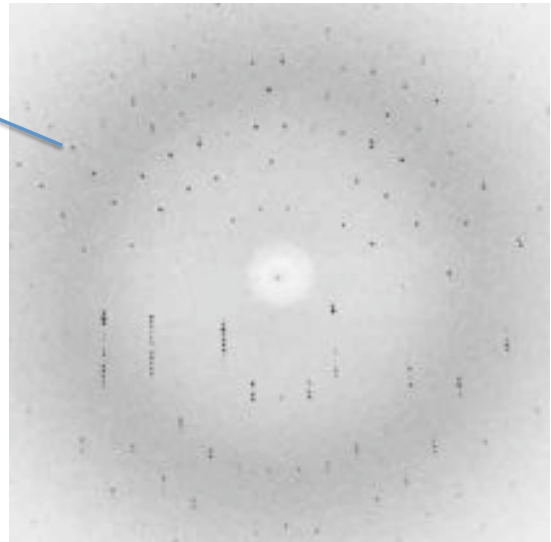
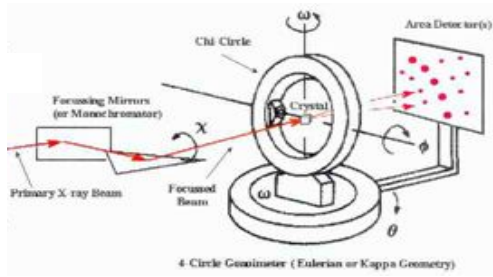
\mathbf{S} continuo

$$I \propto |F(\mathbf{S})|^2 \xrightarrow{\text{cristallo}} |F(h,k,l)|^2$$

\mathbf{S} discreto

Esperimento di diffrazione

$$I(h,k,l) \propto |F(h,k,l)|^2$$

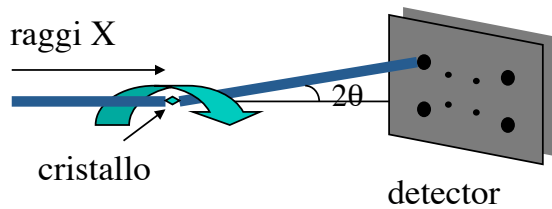


Corso di cristallografia

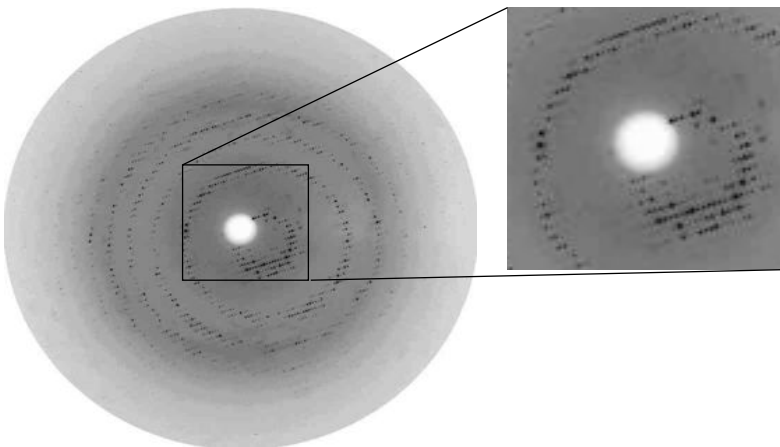
<http://escher.epfl.ch/eCrystallography/>

la sfera di Ewald

<http://escher.epfl.ch/x-ray/diff.mpeg>



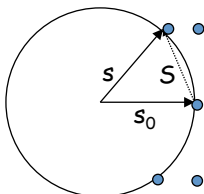
Risultato sperimentale



h	k	l	I	σ
0	0	18	5377.7	426.7
0	0	30	87315.1	7080.9
0	0	39	79150.9	5678.3
0	0	42	88255.3	6544.6
0	0	45	14582.6	1511.1
0	0	48	8125.2	596.7
0	0	51	46929.6	3740.0
0	0	54	79917.3	8107.1
0	0	57	2243.6	316.8
0	0	60	21097.8	1703.7
0	0	63	90391.1	6236.3
0	0	66	96333.2	9161.3
0	0	72	38851.1	2596.8
0	0	75	106827.0	11204.0
0	0	78	76020.8	6028.0
0	0	84	882.7	197.4
(...)				

Sfera di Ewald

Si ha diffrazione solamente quando la sfera di Ewald (selezionata dalla lunghezza d'onda λ) interseca un punto del reticolo reciproco



Ruotando il cristallo si ruota il reticolo reciproco

$$|s| = \frac{1}{\lambda} \text{ raggio della sfera}$$

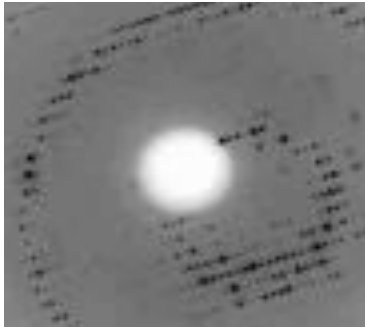
Il problema della fase

$$I(\mathbf{S}) \propto |K(\mathbf{S})|^2 = \left| \sum_{\text{cristallo}} e^{i2\pi\mathbf{S}\cdot\mathbf{Q}} F(\mathbf{S}) \right|^2 \longrightarrow I \propto |F(h,k,l)|^2$$

cristallo

$$F(h,k,l) = \int_{\text{cell}} \rho(x,y,z) \exp[2\pi i(hx + ky + lz)] dV \longleftarrow \rho(x,y,z) = \frac{1}{V} \sum_{h,k,l} F(h,k,l) \exp[-2\pi i(hx + ky + lz)]$$

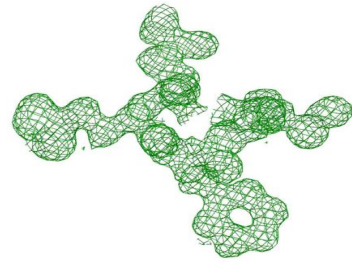
F.T.



misura sperimentale

$$I(h,k,l) \propto |F(h,k,l)|^2$$

→



informazione strutturale

Il problema della fase:

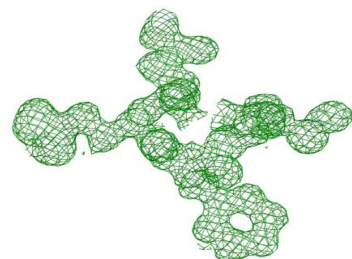
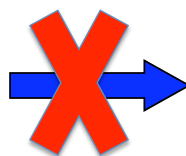
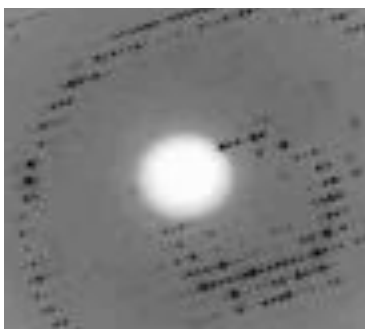
Il fattore di struttura e' un numero complesso

$$F = F' + iF'' = |F(h,k,l)| e^{i\alpha(h,k,l)}$$

$$\frac{1}{V} \sum_{h,k,l} |F(h,k,l)| e^{i\alpha(h,k,l)} \exp[-2\pi i(hx + ky + lz)] \neq \rho(x,y,z)$$



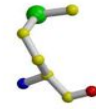
La fase non e' misurabile direttamente



Soluzione del problema della fase

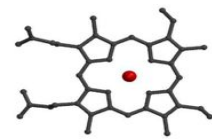
-Molecular Replacement (MR)

- Atomi pesanti: Se, Fe, Hg, Pt, Sm, ...:



MIR
Multiple Isomorphous Replacement

MAD
Multiple Anomalous Dispersion



- Metodi diretti (...)

Molecular Replacement

- Ricerca di un modello adeguato: allineamento di sequenza contro la banca dati PDB

proteina	... YKTQAGKTVVDYINAAIGG----SADGAGL ...	struttura ignota
modello(i)	... YQTQASKTVVDYITAA LAGSRNV SADAAGL ...	struttura nota
	* *	

- **orientare** il modello nella cella elementare del cristallo sperimentale così come è orientata la proteina (**ricerca di 3 parametri rotazionali**). Si utilizza la **funzione di Patterson**

$$P(\mathbf{u}) = P(u, v, w) \equiv \frac{1}{V} \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{u}) = \int_{\text{cella}} \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{u}) d^3 r$$

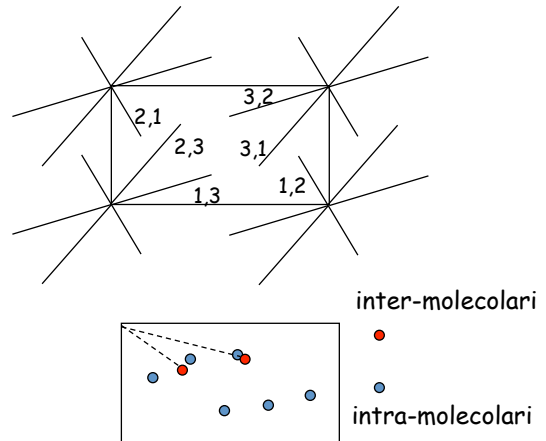
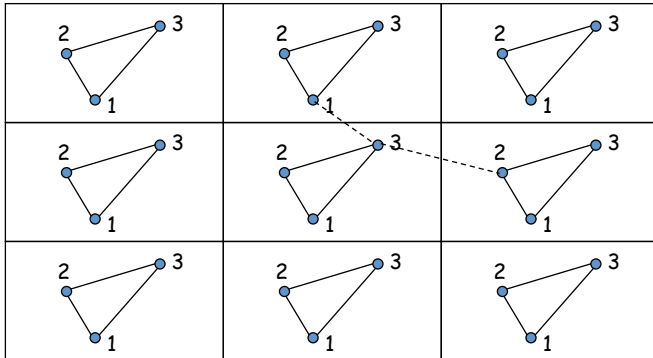
- **posizionare** il modello, correttamente orientato, nella posizione in cui si trova la proteina nella cella elementare del cristallo (**ricerca di 3 parametri traslazionali**)

$$R = \frac{\sum_{\mathbf{h}} \left| |F_{\mathbf{h}}(\text{obs})| - k |F_{\mathbf{h}}(\text{calc})| \right|}{\sum_{\mathbf{h}} |F_{\mathbf{h}}(\text{obs})|}$$

Esempio: funzione di Patterson

Spazio reale: 3 atomi

Spazio Patterson: $N^2 - N = N(N - 1)$ picchi
 prot. di 200 aa. ~ 2000 atomi ~ 4×10^6 picchi



Ricerca rotazionale: convoluzione di 2 funzioni di Patterson

$$R(\alpha, \beta, \gamma) = \int_U d^3 u P^{esp}(\mathbf{u}) P_R^{mod}(\mathbf{u}_R)$$

\swarrow volume di integrazione U

$$I \propto z^2$$

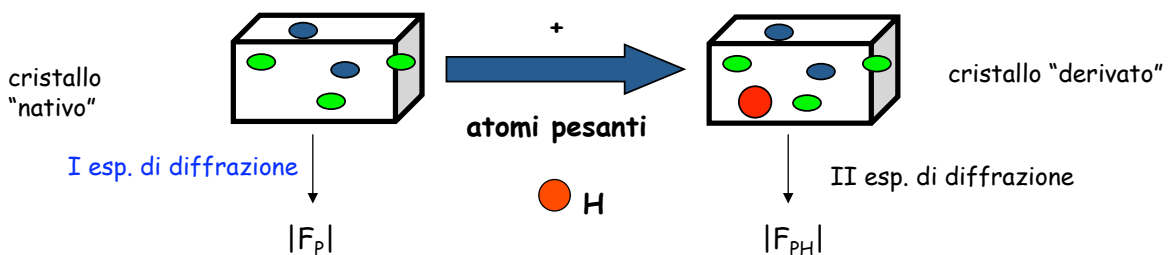
(Atomi pesanti)

Ottenere delle fasi mediante atomi pesanti: M.I.R.
 (Multiple Isomorphous Replacement)

$$F(h, k, l) = \int_{cell} \rho(x, y, z) \exp[2\pi i(hx + ky + lz)] dV = \sum_{j=1}^{N_{atomi}} f_j \exp[2\pi i(hx_j + ky_j + lz_j)]$$

Fattore di scattering atomico (atomo)

Proprietà strutturale (posizione)



$$F_{PH} = F_P + F_H$$

$$\forall (h,k,l) \equiv \mathbf{h}$$

$$F_{PH}(\mathbf{h}) = F_P(\mathbf{h}) + F_H(\mathbf{h})$$



Se troviamo la posizione degli atomi pesanti nella cella elementare abbiamo:

$|F_H|, \alpha_H$ (calcolati dalla posizione dell'atomo pesante)

$|F_P|, |F_{PH}|$ misurati con 2 esperimenti di diffrazione ($I \propto |F|^2$)

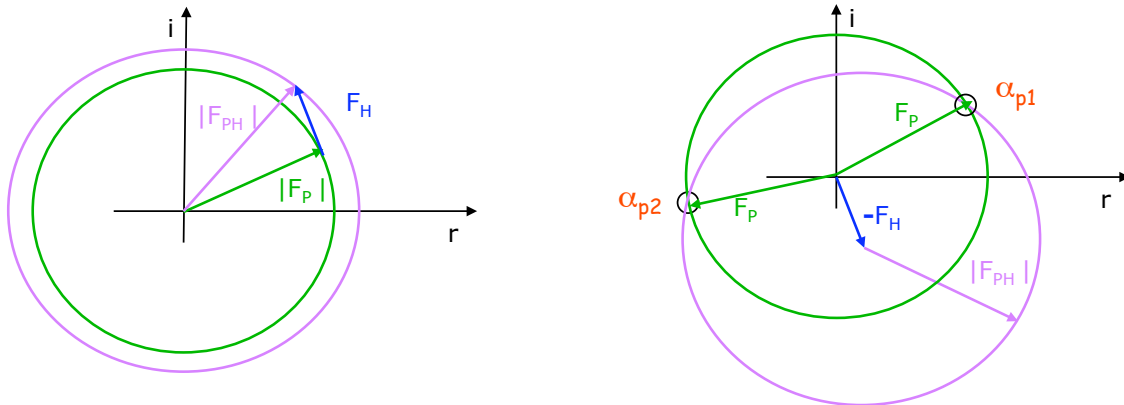
conosciamo F_H in modulo e fase

$$\forall (h,k,l) \equiv \mathbf{h}$$

Consideriamo il cerchio di raggio $|F_{PH}|$ centrato in $-F_H$ le 2 intersezioni con il cerchio di raggio $|F_P|$, danno 2 valori di fase possibile: α_{p1} e α_{p2}

$$F_P = -F_H + F_{PH}$$

conosciamo solamente i moduli $|F_P|, |F_{PH}|$
li rappresentiamo come cerchi nel piano complesso



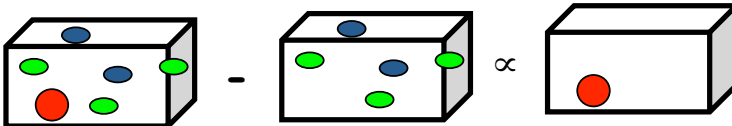
2 soluzioni per la fase di F_P : ambiguità di fase per ogni fattore di struttura $F_P(\mathbf{h})$

Come trovare $\alpha_H \Leftrightarrow$ posizione degli atomi pesanti ?

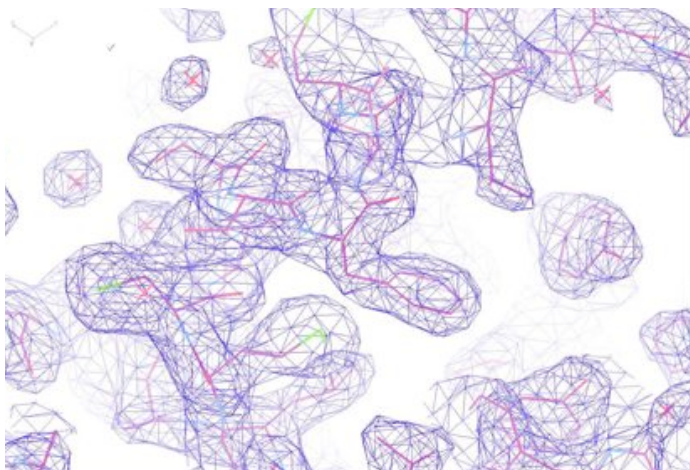
$$F_H(h,k,l) = \int_{cell} \rho_H(x,y,z) \exp[2\pi i(hx + ky + lz)] dV = \sum_{j=1}^{N \text{ atomi pesanti}} f_{Hj} \exp[2\pi i(hx_j + ky_j + lz_j)]$$

posizione degli atomi pesanti $\mathbf{x}_H \Leftrightarrow F_H = |F_H| e^{i\alpha_H}$ fattore di struttura

differenze isomorfe: si puo dimostrare che:

$$\Delta |F|_{iso} \equiv |F_{PH}| - |F_P| \propto |F_H|$$


fun. di Patterson delle diff. isomorfe = fun. di Patterson dei soli atomi pesanti!
 \Rightarrow posizioni degli atomi pesanti



densita' elettronica

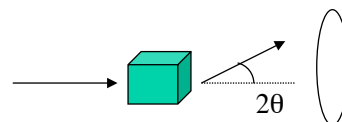
$$\rho(\vec{x}) = \frac{1}{V} \sum_{\vec{h}=\{\min\}}^{\{\max\}} F(\vec{h}) \exp[-2\pi i(\vec{h}\vec{x})]$$



modello 3D

Risoluzione (legge di Bragg)

$$d = \frac{\lambda}{2 \sin \vartheta_{\max}}$$



Validazione modello

Esame del modello finale raffinato:

- accordo fra modello e dati sperimentali (R-factor)

$$R = \frac{\sum_{\mathbf{h}} \left| |F_{\text{obs}}(\mathbf{h})| - k |F_{\text{cal}}(\mathbf{h})| \right|}{\sum_{\mathbf{h}} |F_{\text{obs}}(\mathbf{h})|}$$

- stereochimica delle strutture macromolecolari

- contatti/interazioni fra residui vicini spazialmente

—————> PDB

```

CRYST1   39.550   74.570   66.560   90.00   99.94   90.00 P 1 21 1 4
ATOM      1  CB  SER      1    30.854  -6.329   36.118   1.00  41.46   6
ATOM      2  OG  SER      1    31.600  -7.531   36.190   1.00  44.54   8
ATOM      3  C   SER      1    30.991  -3.833   36.183   1.00  40.24   6
ATOM      4  O   SER      1    30.868  -2.883   36.961   1.00  40.08   8
ATOM      5  N   SER      1    31.848  -5.217   38.114   1.00  39.55   7
ATOM      6  CA  SER      1    31.668  -5.130   36.630   1.00  40.83   6
ATOM      7  N   THR      2    30.605  -3.796   34.906   1.00  39.92   7
ATOM      8  CA  THR      2    29.920  -2.658   34.286   1.00  38.97   6
ATOM      9  CB  THR      2    30.734  -2.067   33.086   1.00  39.67   6
ATOM     10  OG1 THR      2    31.045  -3.101   32.139   1.00  38.83   8
ATOM     11  CG2 THR      2    32.020  -1.403   33.566   1.00  38.83   6
ATOM     12  C   THR      2    28.542  -3.116   33.777   1.00  38.40   6
ATOM     13  O   THR      2    27.815  -2.341   33.141   1.00  38.79   8
.
.
.
ATOM    2166  HOH  WAT     401    20.048   3.400   49.038   1.00  31.45   8
ATOM    2167  HOH  WAT     402    -9.403   0.553   32.633   1.00  44.09   8
ATOM    2168  HOH  WAT     403     3.928  -6.370   32.724   1.00  21.48   8
END
    
```

contenuto di solvente: 49.4%

cella elementare: a=44.3, b= 96.6, c=111.0; 90°, 90°, 90°; P2₁2₁2₁

