

## Small Angle X-Ray Scattering

Scattering:  $I \propto \frac{d\sigma}{d\omega} = |F|^2 \propto \left| \int d\vec{x}' e^{i\vec{q}\cdot\vec{x}'} V(\vec{x}') \right|^2$

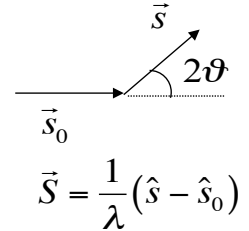
diluizione: interferenza tra le mol. trascurabile

Fattore di struttura del sistema:  $F(\vec{S}) = \sum_{k=1}^N e^{2\pi i \vec{r}_k \cdot \vec{S}} f_k(\vec{S})$

$N$  = numero di molecole nel sistema,

$f_k$  = fattore di struttura della molecola  $k$ :

$$f_k(\vec{S}) = \sum_{n=1}^{N_{el}^k} e^{2\pi i \vec{r}_n \cdot \vec{S}} f_{el}(\vec{S})$$



L'intensità diffusa è proporzionale al modulo quadro del fattore di struttura

$$I(\vec{S}) \propto |F|^2 = \left| f_{el}(\vec{S}) \right|^2 \sum_{n=1}^{N_{el}} e^{-2\pi i \vec{r}_n \cdot \vec{S}} \sum_{m=1}^{N_{el}} e^{2\pi i \vec{r}_m \cdot \vec{S}} = \left| f_{el}(\vec{S}) \right|^2 \sum_{n,m=1}^{N_{el}} e^{-2\pi i \vec{r}_{nm} \cdot \vec{S}}$$

$$\vec{r}_{nm} = \vec{r}_n - \vec{r}_m$$

1

$$I(\vec{S}) \propto \left| f_{el}(\vec{S}) \right|^2 \sum_{n,m=1}^{N_{el}} e^{-2\pi i \vec{r}_{nm} \cdot \vec{S}}$$

Media su tutte le possibili orientazioni della molecola: integrazione sull'angolo solido  $d\omega$

$$\langle I(\vec{S}) \rangle \propto \int_0^{2\pi} d\varphi \int_0^\pi d\vartheta \sin\vartheta \left| f_{el}(\vec{S}) \right|^2 \sum_{n,m=1}^{N_{el}} e^{-2\pi i \vec{r}_{nm} \cdot \vec{S}}$$

$$\int_0^{2\pi} d\varphi \int_0^\pi \sin\vartheta d\vartheta e^{-2\pi i \vec{r} \cdot \vec{S}} = - \int_0^{2\pi} d\varphi \int_1^{-1} d\cos\vartheta e^{-2\pi i r S \cos\vartheta} = 4\pi \frac{\sin(2\pi r S)}{2\pi r S}$$

$$\int_1^{-1} dx e^{-iax} = -\frac{1}{ia} (e^{ia} - e^{-ia}) = -\frac{2}{a} \sin a$$

$$\langle I(\vec{S}) \rangle \propto 4\pi \left| f_{el}(\vec{S}) \right|^2 \sum_{m,n=1}^{n_{el}} \frac{\sin(2\pi S r_{nm})}{2\pi S r_{nm}}$$

2

$$f_k(\vec{S}) = \sum_{n=1}^{N_{el}} e^{2\pi i \vec{r}_n \cdot \vec{S}} f_{el}(\vec{S}) \quad \text{rappresentazione discreta}$$

**rappresentazione continua:**  $\rho_k$  densità elettronica della mol.  $k$

$$\begin{aligned} f_k(\vec{S}) &= \int_0^{2\pi} d\varphi \int_0^\pi \sin \vartheta d\vartheta \int_0^\infty r^2 dr \rho_k(r) e^{2\pi i \vec{r} \cdot \vec{S}} = - \int_0^{2\pi} d\varphi \int_0^\pi d \cos \vartheta \int_0^\infty r^2 dr \rho_k(r) e^{2\pi i r S \cos \vartheta} = \\ &= 4\pi \int_0^\infty r^2 dr \rho_k(r) \frac{\sin(2\pi S r)}{2\pi S r} \end{aligned}$$

3

**rappresentazione della mol. in termini di elettroni**

$$\langle I(\vec{S}) \rangle \propto 4\pi |f_{el}(\vec{S})|^2 \sum_{m,n=1}^{N_{el}} \frac{\sin(2\pi S r_{nm})}{2\pi S r_{nm}}$$

**rappresentazione della mol. in termini di atomi**

$$\langle I(\vec{S}) \rangle \propto 4\pi \sum_{k,h=1}^N f_k(\vec{S}) f_h^*(\vec{S}) \frac{\sin(2\pi S r_{hk})}{2\pi S r_{hk}}$$

4

## Small Angle X-Ray Scattering

$$\langle I(\vec{S}) \rangle \propto 4\pi |f_{el}(\vec{S})|^2 \sum_{m,n=1}^{n_{el}} \frac{\sin(2\pi S r_{nm})}{2\pi S r_{nm}}$$

$S r_{nm} \rightarrow 0$

$$\frac{\sin x}{x} = 1 - \frac{x^2}{6} + o(x^4)$$

per piccoli angoli di diffrazione:

$$\langle I(\vec{S}) \rangle \approx 4\pi |f_{el}(0)|^2 \sum_{m,n=1}^{n_{el}} \left( 1 - \frac{(2\pi S r_{nm})^2}{6} \right)$$

Definiamo il quadrato del *raggio di girazione*:

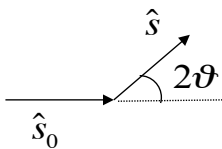
$$R_G^2 \equiv \frac{1}{(n+1)^2} \sum_{i < j} |\vec{r}_i - \vec{r}_j|^2 = \frac{1}{(n+1)^2} \sum_{i < j} r_{ij}^2 = \frac{1}{2n_{el}^2} \sum_{i,j} r_{ij}^2$$

5

$$\langle I(\vec{S}) \rangle \approx 4\pi |f_{el}(\vec{S})|^2 n_{el}^2 \left( 1 - \frac{(2\pi S)^2 R_G^2}{3} \right)$$

vettore di scattering:  $S \rightarrow 0$

$$\langle I(\vec{S}) \rangle \xrightarrow{S \rightarrow 0} 4\pi |f_{el}(\vec{S})|^2 n_{el}^2 \left( 1 - \frac{(2\pi S)^2 R_G^2}{3} \right)$$

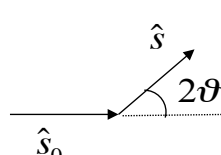


$$2\vartheta \leq 5^\circ$$

$$\vec{S} = \frac{\hat{s}}{\lambda} - \frac{\hat{s}_0}{\lambda} \quad |\vec{S}|^2 = (s^2 + s_0^2 - 2ss_0 \cos 2\vartheta) = \frac{2}{\lambda^2} (1 - \cos 2\vartheta)$$

$$\cos 2\vartheta = \cos^2 \vartheta - \sin^2 \vartheta \quad |\vec{S}|^2 = \frac{4}{\lambda^2} (\sin^2 \vartheta) \rightarrow |\vec{S}| = \frac{2}{\lambda} (\sin \vartheta)$$

6



$$\langle I(\vec{S}) \rangle \rightarrow \langle I(\vartheta) \rangle \propto I(0)n_{el}^2 \left( 1 - \frac{4\pi^2 R_G^2 \sin^2 \vartheta}{3\lambda^2} \right)$$

$e^{-\frac{x^2}{a}} \xrightarrow{x \rightarrow 0} 1 - \frac{x^2}{a}$ 
rappresentazione come decadimento esp.

$$\left\langle \frac{I(\vartheta)}{I(0)} \right\rangle \propto \left( 1 - \frac{16\pi^2 R_G^2 \sin^2 \vartheta}{3\lambda^2} \right) \approx e^{-\frac{x^2}{a}}$$

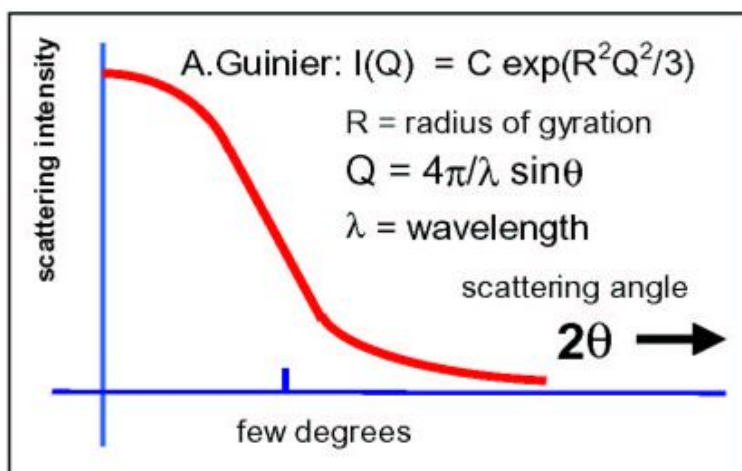
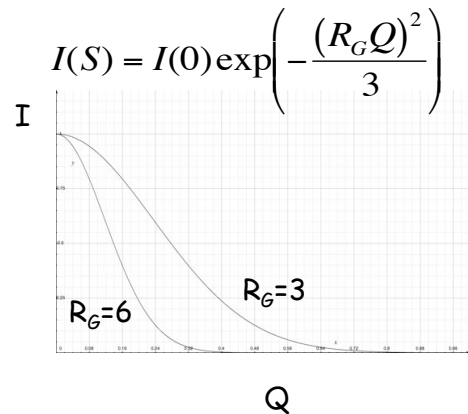
$$\left\langle \frac{I(\vartheta)}{I(0)} \right\rangle \xrightarrow{s \rightarrow 0} \exp\left(-\frac{16\pi^2 R_G^2 \sin^2 \vartheta}{3\lambda^2}\right)$$

7

## Guinier Plot

$$\left\langle \frac{I(\vartheta)}{I(0)} \right\rangle \xrightarrow{s \rightarrow 0} \exp\left(-\frac{16\pi^2 R_G^2 \sin^2 \vartheta}{3\lambda^2}\right)$$

$$\ln\left(\left\langle \frac{I(\vartheta)}{I(0)} \right\rangle\right) \approx -\left(\frac{4\pi R_G \sin \vartheta}{\lambda}\right)^2 \frac{1}{3}$$



8

Facciamo un passo indietro: consideriamo il fattore di struttura di una molecola in soluzione

$$F(\vec{S}) = \int d\vec{r} \rho(\vec{r}) e^{2\pi i \vec{r} \cdot \vec{S}} = \int_0^{2\pi} d\phi \int_0^{\pi} \sin \theta d\theta \int_0^{\infty} dr r^2 \rho(r) e^{2\pi i r S \cos \theta} =$$

$$= 2\pi \int_0^{\infty} dr r^2 \rho(r) \frac{e^{2\pi i r S} - e^{-2\pi i r S}}{2\pi i r S} = 4\pi \int_0^{\infty} dr r^2 \rho(r) \frac{\sin 2\pi r S}{2\pi r S} = 4\pi \rho_0 \int_0^R dr r^2 \frac{\sin 2\pi r S}{2\pi r S}$$

Se la molecola è sferica e la sua dens. el. e' cost.  $\rho$  seleziona il raggio di integrazione =  $R$  = raggio della sfera

Integriamo per parti:

$$\int \frac{\sin ar}{a} r dr = \frac{1}{a^2} \int (-d \cos ar) r = \frac{1}{a^2} \left( \int (\cos ar) dr - r \cos ar \Big|_0^R \right) =$$

$$= \frac{1}{a^2} \left( \frac{\sin aR}{a} - R \cos aR \right)$$

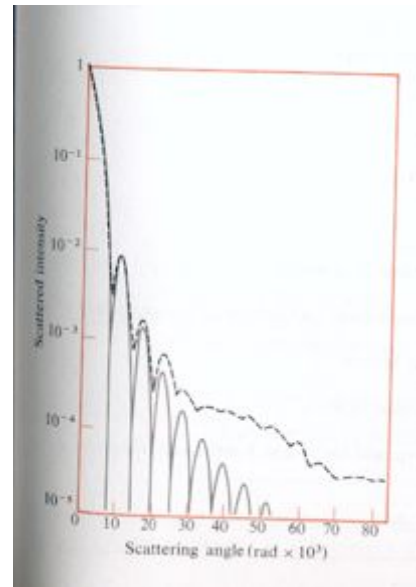
Se la particella è sferica

$$F(S) = 4\pi \rho_0 R^3 \left( \frac{\sin(2\pi R S) - (2\pi R S) \cos(2\pi R S)}{(2\pi R S)^3} \right)$$

9

$$F(S) = 4\pi \rho_0 R^3 \left( \frac{\sin(2\pi R S) - (2\pi R S) \cos(2\pi R S)}{(2\pi R S)^3} \right)$$

$$S = \frac{2}{\lambda} (\sin \vartheta)$$



Per ottenere il Guinier plot abbiamo fatto 2 approssimazioni successive arrivando al decadimento esponenziale.

In realtà partendo da  $\langle I(\vec{S}) \rangle \propto \sum_{m,n=1}^{n_{el}} \frac{\sin(2\pi S r_{nm})}{2\pi S r_{nm}} \rightarrow \frac{\sin x}{x}$

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{\frac{\partial}{\partial x} \sin x}{\frac{\partial}{\partial x} x} = \lim_{x \rightarrow 0} \cos x = 1$$

10

## Stima dell'errore nell'approssimazione di Guinier

funzione pari

$$\frac{\sin x}{x} = 1 + \cancel{f'(0)x} + f''(0)\frac{x^2}{2} + \cancel{f'''(0)\frac{x^3}{6}} + f^{IV}(0)\frac{x^4}{24} + o(x^6)$$

$$= 1 + f''(0)\frac{x^2}{2} + f^{IV}(0)\frac{x^4}{24} = 1 - \frac{x^2}{6} + \frac{x^4}{120}$$

$$e^{-\frac{y}{a}} \approx 1 - \frac{y}{a} + \frac{y^2}{2a^2} \rightarrow 1 - \frac{x^2}{6} + \frac{x^4}{72}$$

$$\begin{cases} y = \frac{x^2}{2} \\ a = 3 \end{cases}$$

$$\boxed{\frac{\sin x}{x} \approx e^{-\frac{x^2}{6}} + o(x^6)}$$

11

## Intensità diffusa

$$F(\vec{S}) = \int d\vec{r} \rho(\vec{r}) e^{2\pi i \vec{r} \cdot \vec{S}}$$

$$\begin{aligned} I(\vec{S}) &\propto |F(\vec{S})|^2 = F(\vec{S})F(\vec{S})^* = \int d\vec{r} \rho(\vec{r}) e^{2\pi i \vec{r} \cdot \vec{S}} \int d\vec{r}' \rho(\vec{r}') e^{-2\pi i \vec{r}' \cdot \vec{S}} = \\ &= \int d\vec{r} d\vec{r}' \rho(\vec{r}) \rho(\vec{r}') e^{2\pi i \vec{S} \cdot (\vec{r} - \vec{r}')} \end{aligned}$$

$$\vec{r} - \vec{r}' \equiv \vec{u}$$

$$d\vec{r} = d\vec{u}$$

trasf. di Fourier

$$I(\vec{S}) \propto \int d\vec{u} d\vec{r}' \rho(\vec{r}' + \vec{u}) \rho(\vec{r}') e^{2\pi i \vec{S} \cdot \vec{u}} = \int d\vec{u} \gamma(\vec{u}) e^{2\pi i \vec{S} \cdot \vec{u}} = F.T.(\gamma)$$

$$\gamma(\vec{u}) = \int d\vec{r}' \rho(\vec{r}' + \vec{u}) \rho(\vec{r}') = \begin{cases} = 0 & \text{Se } \vec{u} \text{ non unisce 2 atomi della proteina} \\ \neq 0 & \text{Se } \vec{u} \text{ unisce 2 atomi della proteina} \end{cases}$$

$\gamma$  = funzione di autocorrelazione della densità elettronica (Patterson)

12

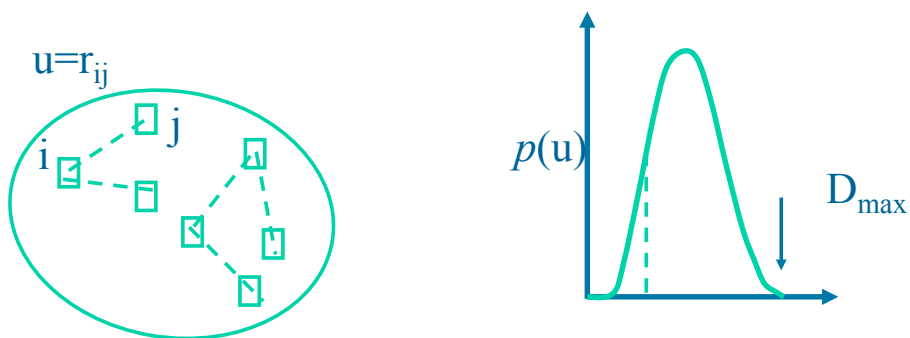
$$I(\vec{S}) \propto \int d\vec{u} d\vec{r}' \rho(\vec{r}' + \vec{u}) \rho(\vec{r}') e^{2\pi i \vec{S} \cdot \vec{u}} = \int d\vec{u} \gamma(\vec{u}) e^{2\pi i \vec{S} \cdot \vec{u}} = F.T.(\gamma)$$

Mediando su tutte le possibili orientazioni =  $\gamma(\vec{u}) = \gamma(u, \vartheta, \varphi) \xrightarrow{\langle \rangle} \gamma(u)$   
 $\gamma$  dipende solo dal modulo di  $u$

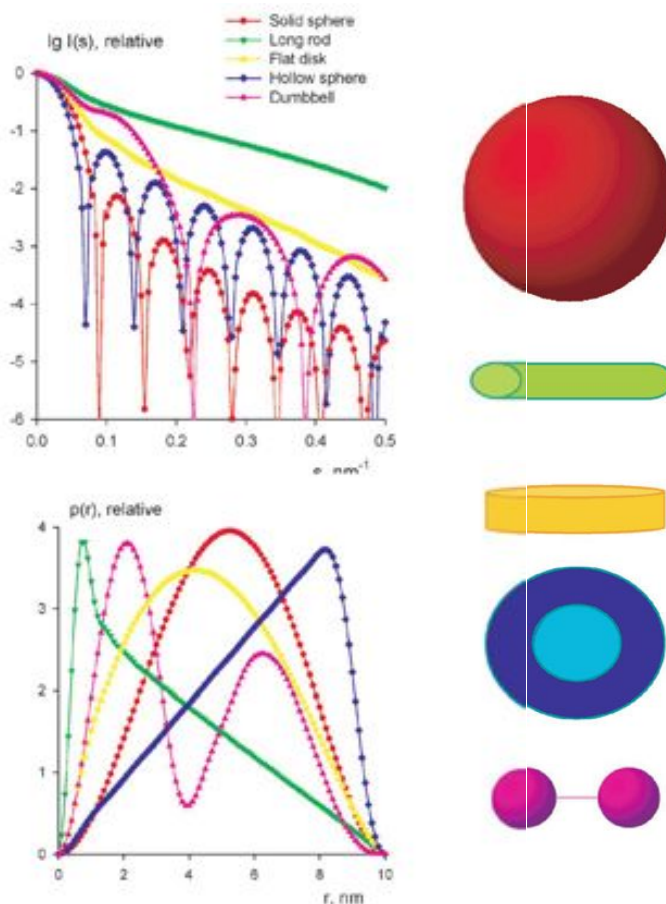
$$\langle I(S) \rangle \propto 4\pi \int du u^2 \gamma(u) \frac{\sin(2\pi u S)}{2\pi u S} = 4\pi \int du p(u) \frac{\sin(2\pi u S)}{2\pi u S} = \langle F.T.(\gamma) \rangle$$

### Funzione di distribuzione delle distanze

$$p(u) = u^2 \gamma(u) = u^2 \int d\vec{r}' \rho(\vec{r}' + \vec{u}) \rho(\vec{r}')$$



13

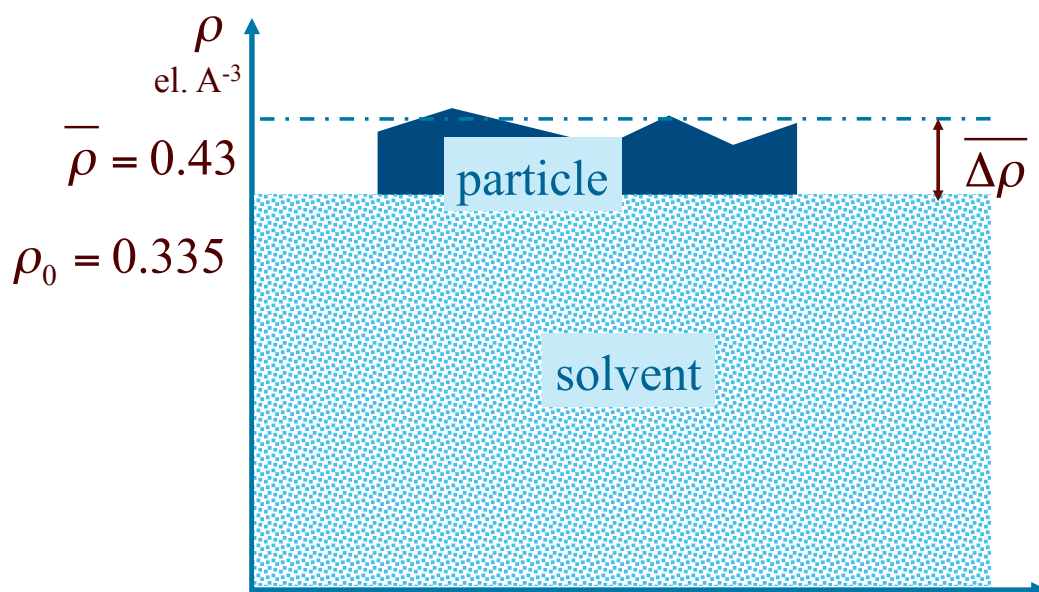


14

## Contrasto

La radiazione incidente interagisce con tutte le molecole in soluzione: solvente e soluto

$$\rho \rightarrow \Delta\rho = \rho_{proteina} - \rho_{solvente} \approx (0.43 - 0.34) \frac{el}{\text{\AA}^3}$$



15

## Intensità all'origine (S=0)

singola proteina

peso della proteina

$$i(0) \propto \left| \int d\vec{r} \Delta\rho(\vec{r}) \right|^2 = (v\Delta\rho)^2 = \left( \frac{Mw}{N_A} \bar{v}\Delta\rho \right)^2$$

numero di moli

volume specifico 0.71-0.73 ml/gr

$$c = \frac{NMw}{N_A V}$$

concentrazione in mg/ml

$$I(0) = ni(0) = NN_A \left( \frac{Mw}{N_A} \right)^2 (\bar{v}\Delta\rho)^2 = \frac{NMw}{N_A} Mw(\bar{v}\Delta\rho)^2 = cVMw(\bar{v}\Delta\rho)^2$$

$$I(0) \propto cMw$$

Se la concentrazione è nota l'intensità all'origine ci da il peso molecolare e dunque lo stato di aggregazione (struttura quaternaria) della proteina

Intensità per grandi angoli: **legge di Porod**:

$$I(S) \xrightarrow{S \rightarrow \infty} S^{-4}$$

16



## Tecnica sperimentale

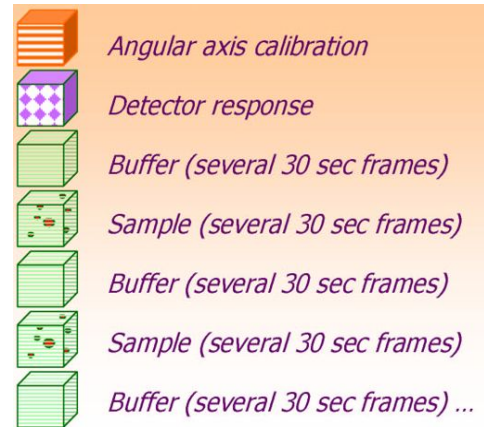
Di solito si fanno 3-4 esperimenti: in ognuno di essi la proteina da analizzare viene diluita nel medesimo buffer in concentrazioni differenti.

Le curve sperimentali vengono scalate tra loro secondo la:

$$\frac{I(0)}{M_{WC}} = \text{cost.}$$

Le 3 curve scalate vengono mediate per produrre un'unica curva sperimentale

L'apparato sperimentale misura i conteggi (n. di fotoni che arrivano) in funzione dell'angolo di diffusione.



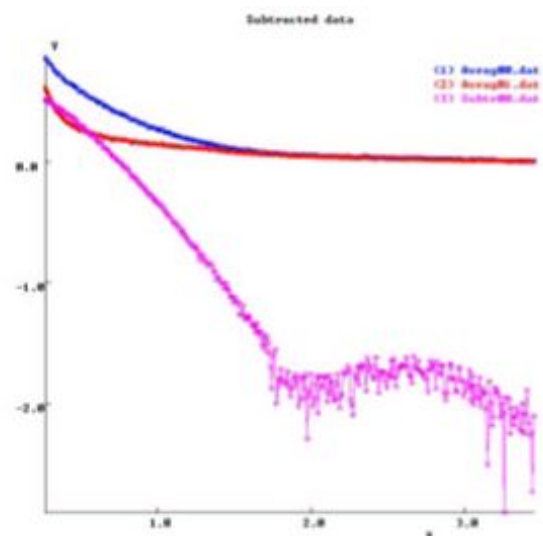
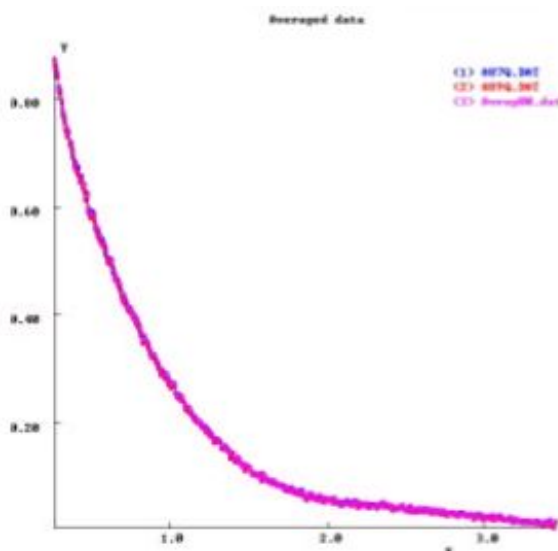
17

## Analisi dati

viene prima sottratto il contributo del solvente

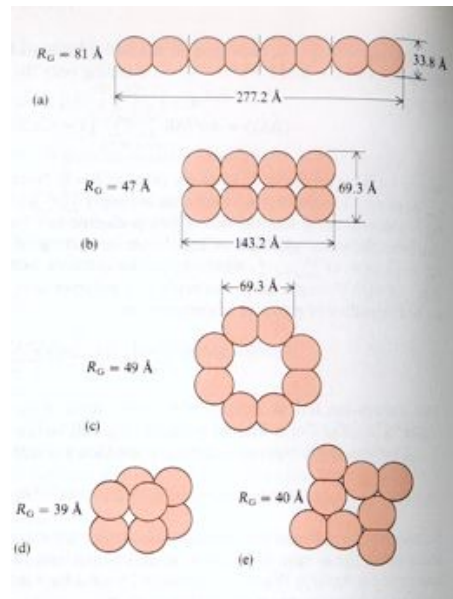
Poi le 3-4 curve sperimentali vengono scalate tra loro secondo la:

$$\frac{I(0)}{M_{WC}} = \text{cost.}$$



## Informazioni che possono essere ricavate sulla proteina in soluzione

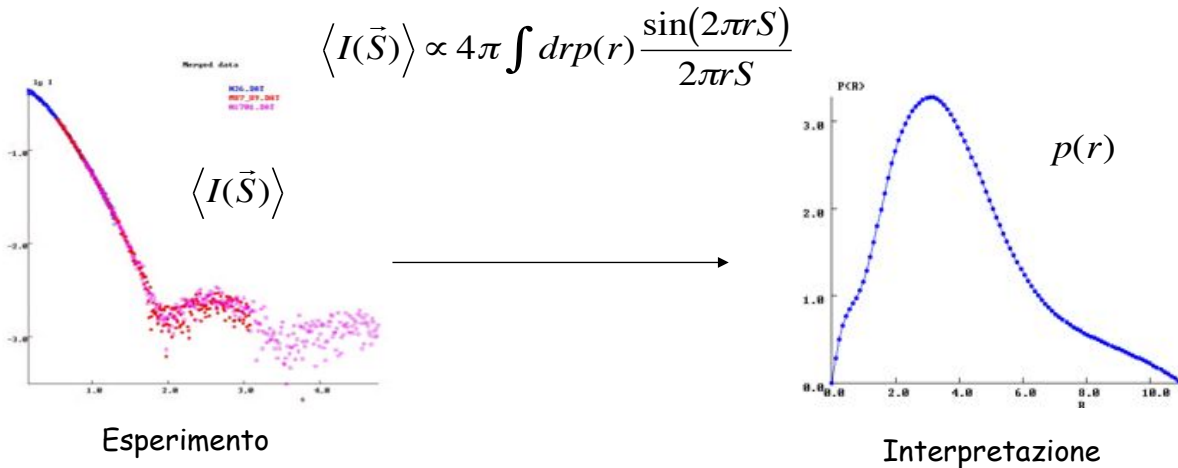
1. raggio di girazione, dal Guinier plot
2. massa della proteina dalla misura di  $I_0$ , ovvero stato di aggregazione proteica



Se si considera il contributo dello scattering a medio angolo si possono costruire modelli strutturali della proteina a bassa risoluzione

19

## Funzione di distribuzione delle distanze



In theory, the calculation of  $p(r)$  from  $I(s)$  is simple.

**Problem** :  $I(s)$  - is only known over  $[s_{\min}, s_{\max}]$  : truncation  
 - is affected by experimental errors

⇒ Calculation of the Fourier transform of *incomplete and noisy data*, which is an *ill-posed problem*.

**Solution** : Indirect Fourier Transform. First proposed by O. Glatter in 1977.

20

Si ipotizza che le particelle abbiano dimensione finita  $R_{\max}$

$$\langle I(\vec{S}) \rangle \propto 4\pi \int_0^{R_{\max}} dr p(r) \frac{\sin(2\pi r S)}{2\pi r S}$$

$p(r)$  is parameterized on  $[0, R_{\max}]$   
by a linear combination  
of **orthogonal functions**

$$p(r) = \sum_{n=1}^M c_n \varphi_n(r)$$

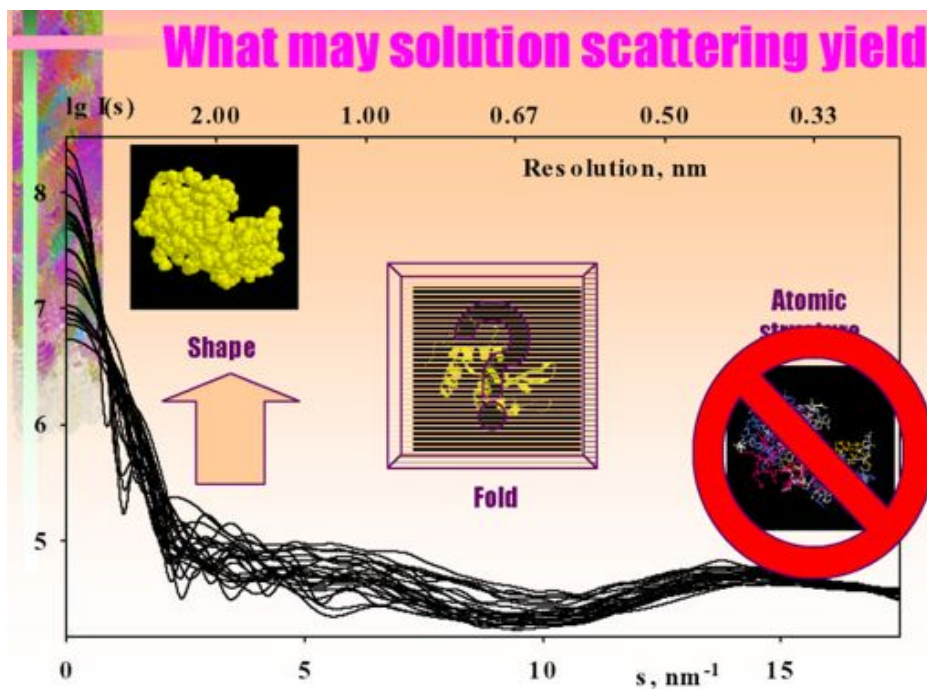
The radius of gyration and the intensity at the origin can be derived from  $p(r)$  using the following expressions :

$$\langle I(0) \rangle = 4\pi \int_0^{R_{\max}} dr p(r)$$

$$R_G^2 = \frac{\int_0^{R_{\max}} dr r^2 p(r)}{2 \int_0^{R_{\max}} dr p(r)}$$

This alternative estimate of  $R_g$  makes use of the **whole scattering curve**, and is much less sensitive to interactions or to the presence of a small fraction of oligomers.

21



Se il fit dei dati sperimentali è cattivo, il modello è scorretto  
Se il fit dei dati sperimentali è buono, il modello è corretto ????

22

## Intensita' diffusa (esperimento)

$$F(\vec{s}) = \sum_{l,m} F_{l,m}(s) Y_{l,m}(\Omega) \quad \text{3D}$$

$$I(\vec{s}) = |F(\vec{s})|^2 = \sum_{l,m} \sum_{l',m'} F_{l,m}(s) Y_{l,m}(\Omega) F_{l',m'}^*(s) Y_{l',m'}^*(\Omega)$$

$$\langle I(\vec{s}) \rangle = \int d\Omega |A(\vec{s})|^2 = \sum_{l,m} |F_{l,m}(s)|^2 = \sum_L |F_L(s)|^2 \quad \text{1D}$$

$L = l, m$

## Densità elettronica (dato che si vuole ricavare)

$$\rho(\vec{r}) = \sum_L \rho_L(r) Y_L(\omega) \quad \text{3D}$$

$$I(\vec{s}) = \int d\vec{r} d\vec{r}' \rho(\vec{r}) \rho(\vec{r}') e^{2\pi i \vec{s} \cdot (\vec{r} - \vec{r}')} \Rightarrow F_{l,m}(s) \propto \int_0^\infty r^2 dr \rho_{l,m}(r) j_l(sr)$$

↓  
Funzioni di Bessel <sup>23</sup>

Struttura della proteina: funzione di forma angolare  $f$   
(descrive la sup. della proteina nello spazio 3D)

$$\rho(\vec{r}) = \begin{cases} 1 & \text{se } 0 \leq r \leq f(\omega) \\ 0 & \text{se } r > f(\omega) \end{cases}$$

superficie nello spazio 3D

$$f(\omega) = \sum_{l,m} f_{l,m} Y_{l,m}(\omega)$$

ricostruzione

$$I(s) = \sum_L |F_L(s)|^2 = \sum_{n=0}^{\infty} s^{2n} a_n = \sum_{n=0}^{\infty} s^{2n} \mathfrak{S}(f_{l,m}) \quad \text{1D}$$

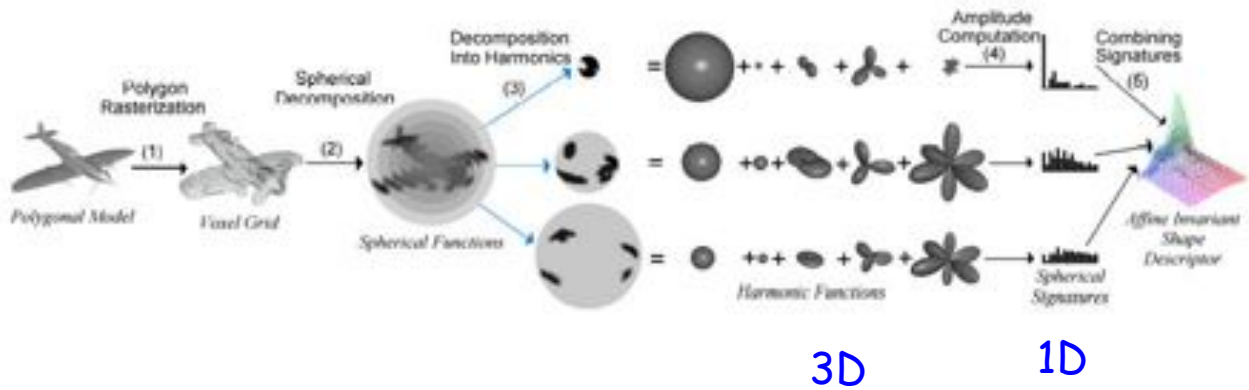
esperimento

forma della proteina

La relazione sperimentale in 1D permette di ricostruire la superficie della proteina in 3D

$$A_{l,m}(s) \leftrightarrow f_{l,m}$$

## Decomposizione in armoniche sferiche



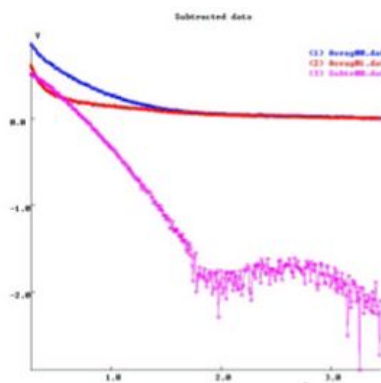
8 armoniche

16 armoniche

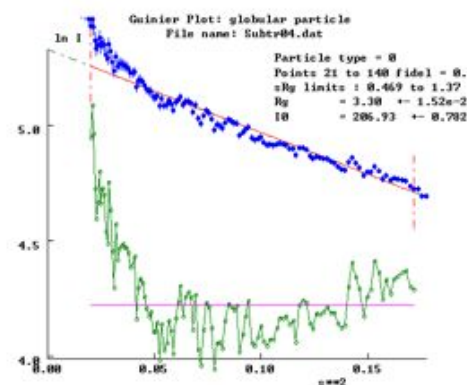
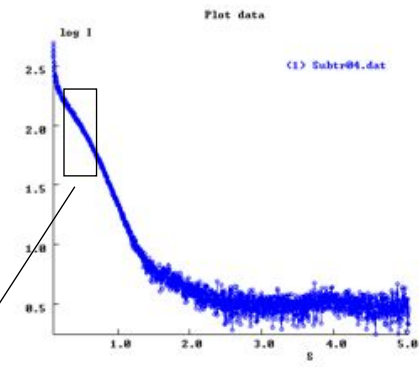
24 armoniche

25

## Esempio: la struttura di NS3



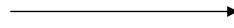
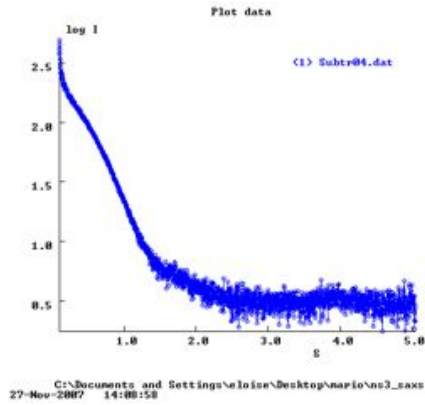
3 curve sperimentali: 2 per il solvente, 1 per la proteina. Si mediano le 2 curve del solvente e si sottraggono alla curva della proteina



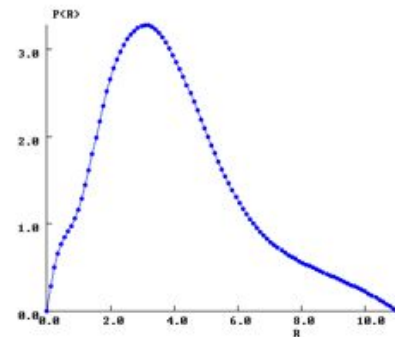
Analisi del raggio di girazione con il Guinier Plot  $R_g=3.30$  nm

26

## Analisi di $p(r)$

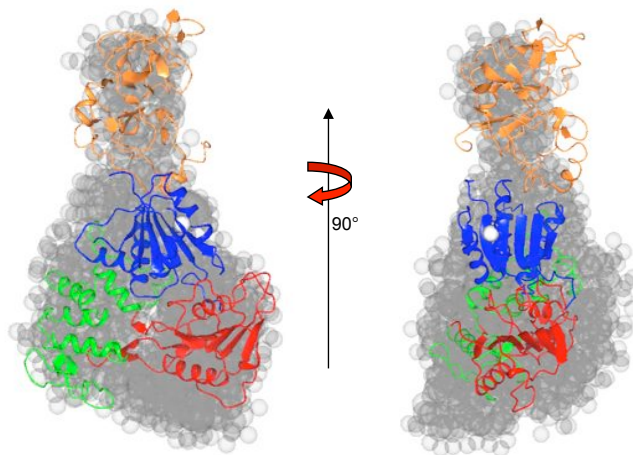
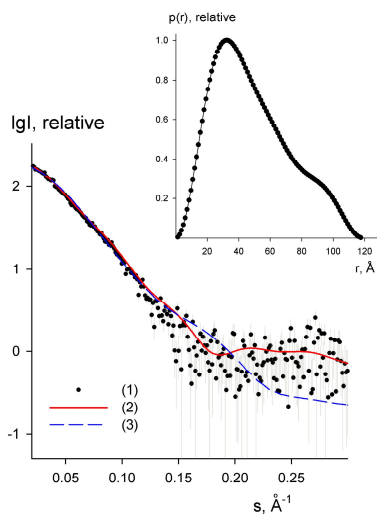


$R_{\max}=11$  nm



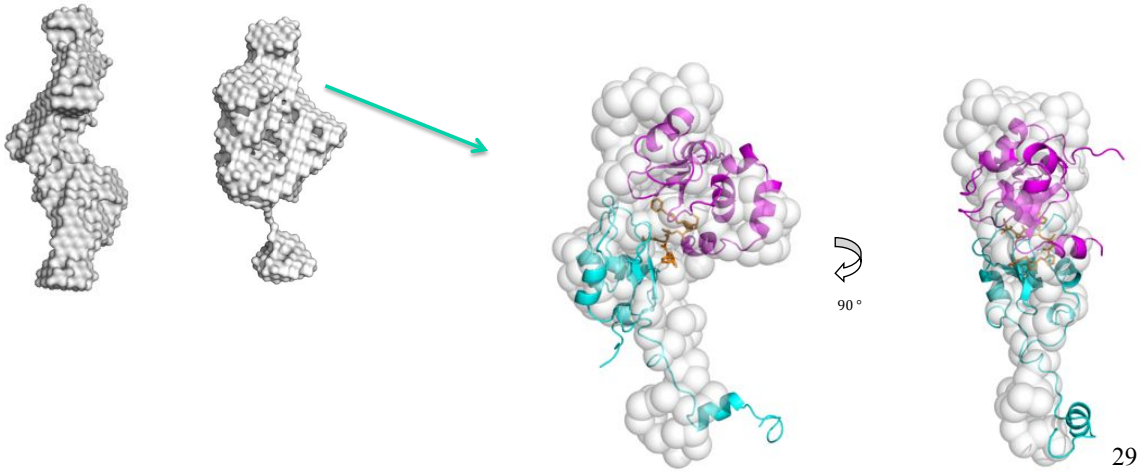
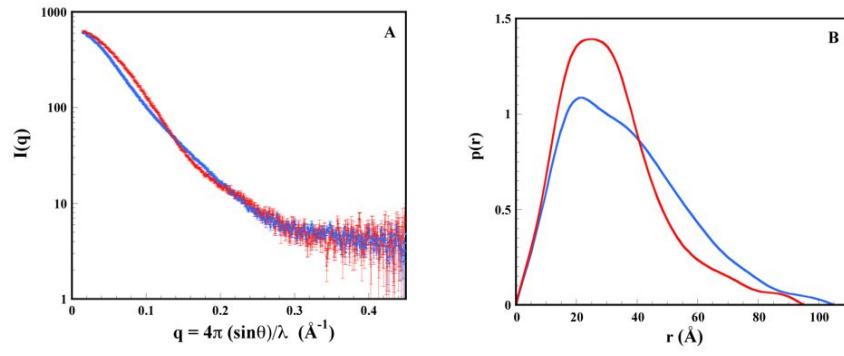
27

## Fit dei dati con il programma GASBOR





# Bir3 + Smac7



Cossu F., Milani M. et al., 2009, *J. Mol. Biol.* **392**, 630-44.

# NS5

