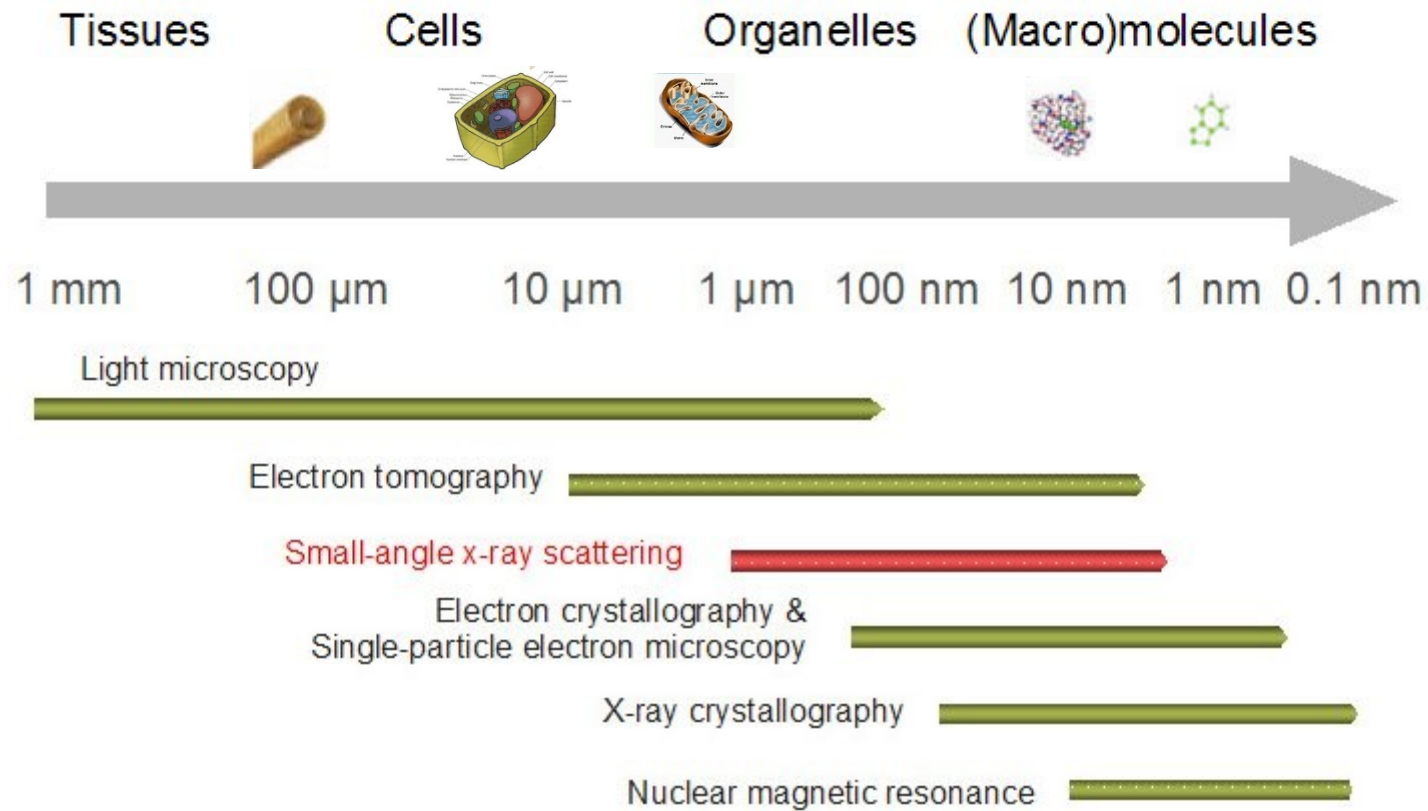


# Studi strutturali: BioSAXS

# BioSAXS:



**SAXS:** livello di risoluzione compreso tra circa 1 e 10 nm

# BioSAXS:

## Perché preoccuparsi della bassa risoluzione?

Relativamente veloce

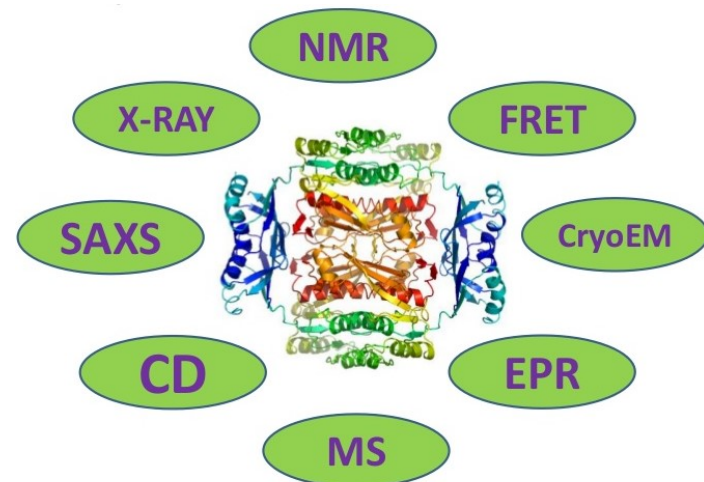
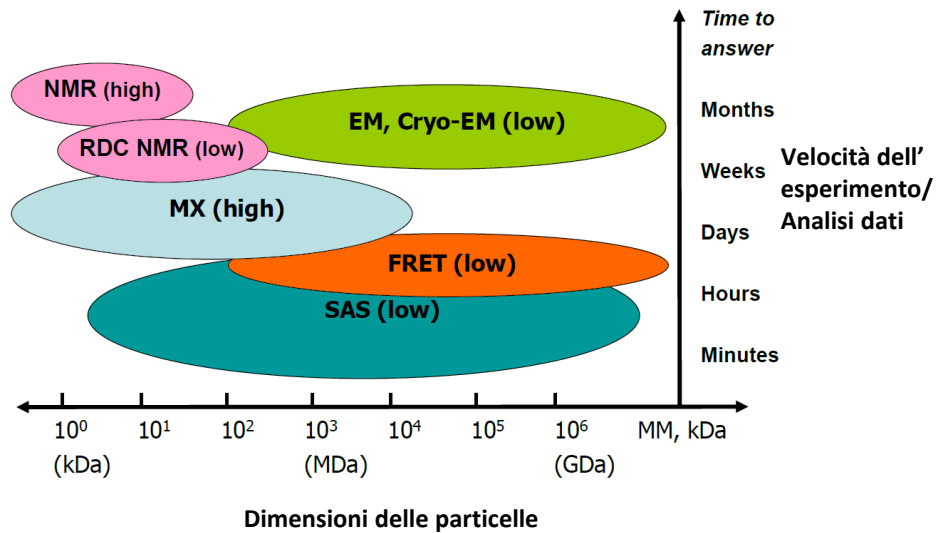
IN SOLUZIONE

- non richiede cristalli  
preparazione speciale del campione

non limitato dal MW

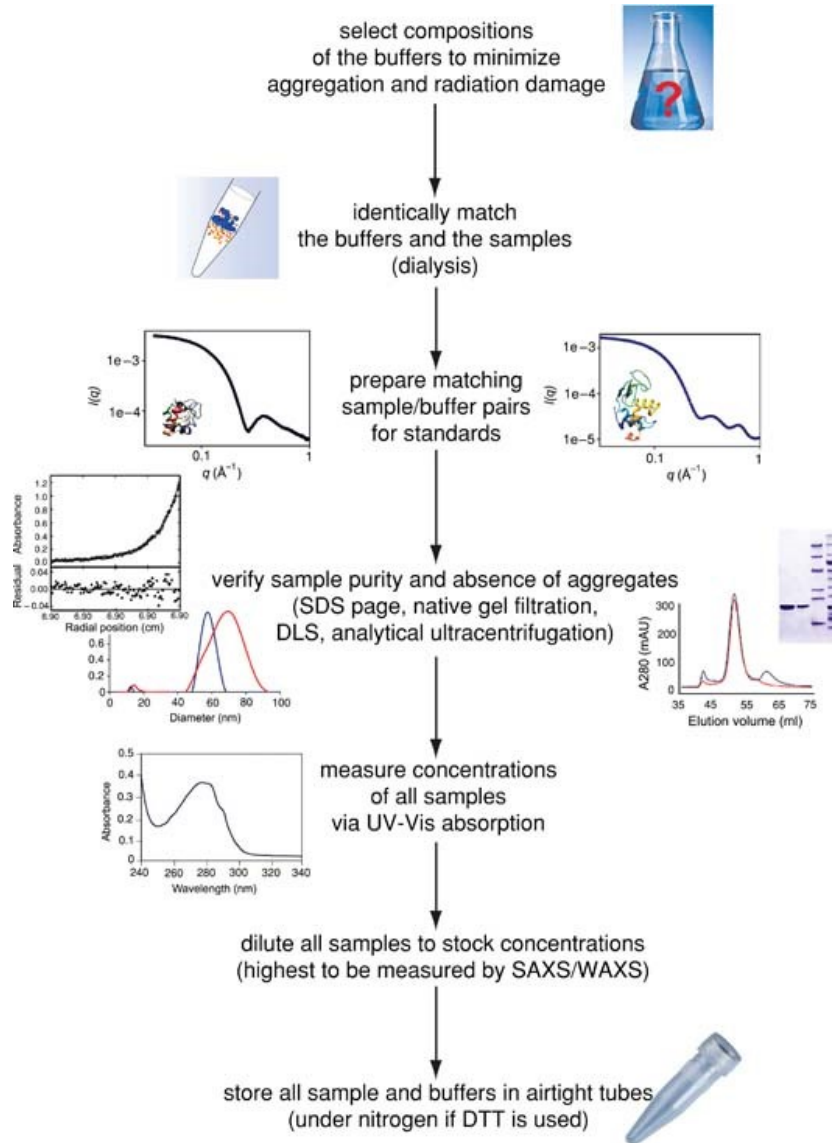
Applicabile in condizioni quasi fisiologiche  
a complessi/sistemi dinamici:  
studi **time resolved**

PIÙ INFORMATIVO SE COMBINATO CON  
**TECNICHE COMPLEMENTARI**



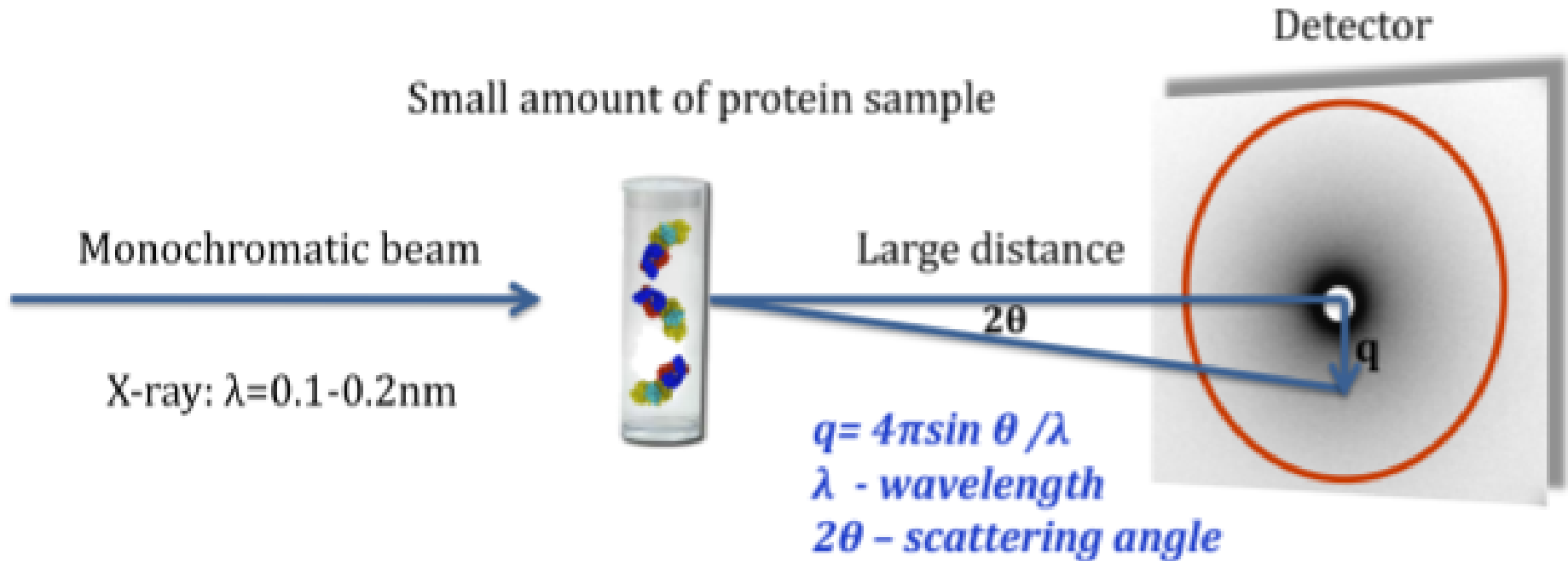
# BioSAXS:

## Preparazione del campione



# BioSAXS:

Set up sperimentali:



## Sorgenti:

Generatore di X-ray ( $\lambda = 0.1 - 0.2 \text{ nm}$ )

Sincrotrone ( $\lambda = 0.03 - 0.35 \text{ nm}$ )

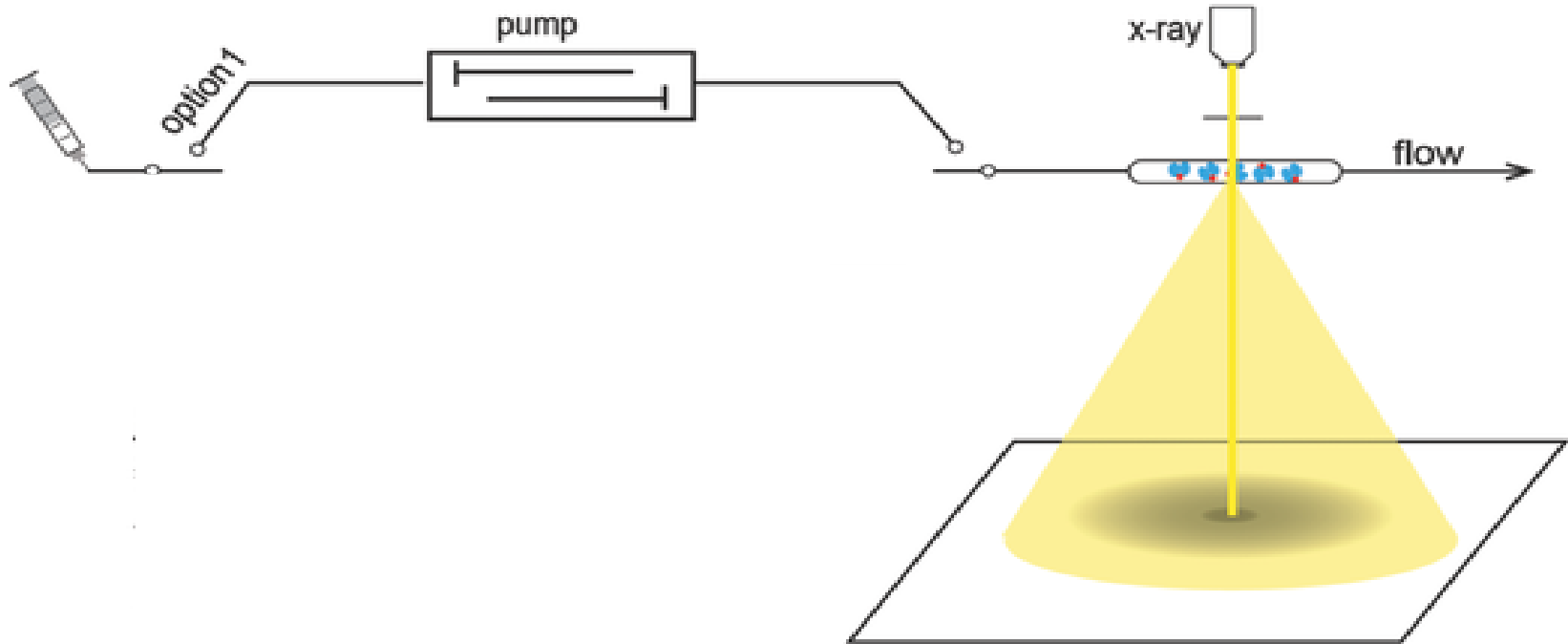
Neutroni ( $\lambda = 0.2 - 1 \text{ nm}$ )

$$s = 4\pi \sin \theta / \lambda = q$$

(anche  $S = 2 \sin \theta / \lambda = 2\pi s = s (!) = \mu \dots$ )

# BioSAXS:

Set up sperimentali: STANDARD BATCH MODE



Con un sample changer automatizzato (8–40°C )

Misurate serie di **concentrazioni**

**Centrifugazione**

**misura della CONCENTRAZIONE**

Tempo variabile della raccolta dati da **1 sec in a 5 minuti a seconda della sorgente di raggi X.**

# BioSAXS:

## Set up sperimentali: SEC-SAXS

Accoppiato anche con altri detectors (UV-MALLS-DLS-RI)

Raccolta dati contemporanea alla purificazione;  
Sorgenti molto intense → esposizioni molto brevi:  
molti frames a cavallo del profilo di eluizione  
con sufficiente qualità per essere  
analizzati individualmente

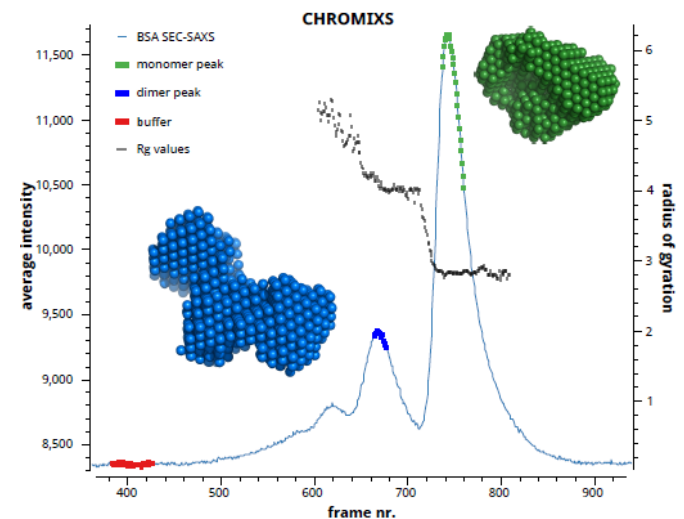
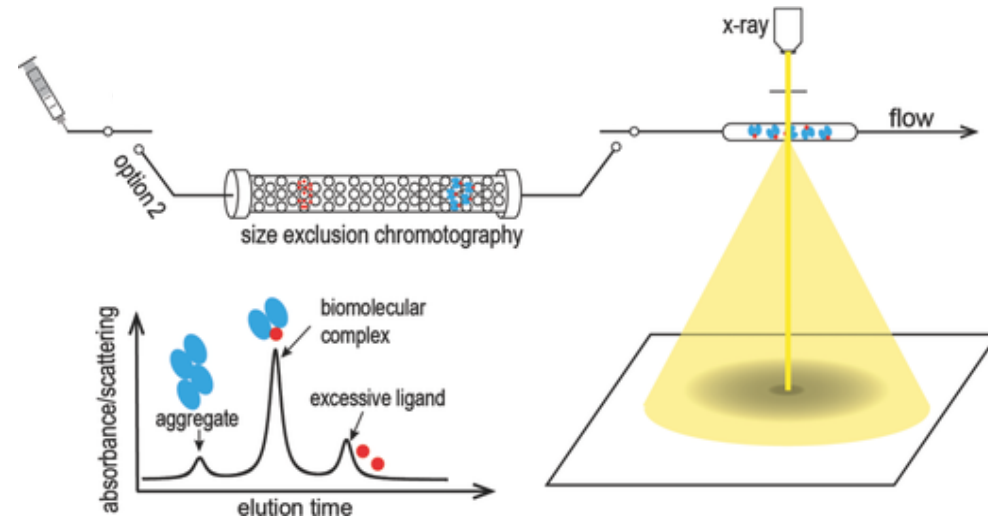
Vantaggi:

- **proteine prone ad aggregare**
- **formazione parziale di complessi**  
(complessi co-esistono con le subunità)
- **Proteine di membrana** (solubilizzate con detergenti

Miscela di complessi proteina:detergente, detergente solubile e micelle).

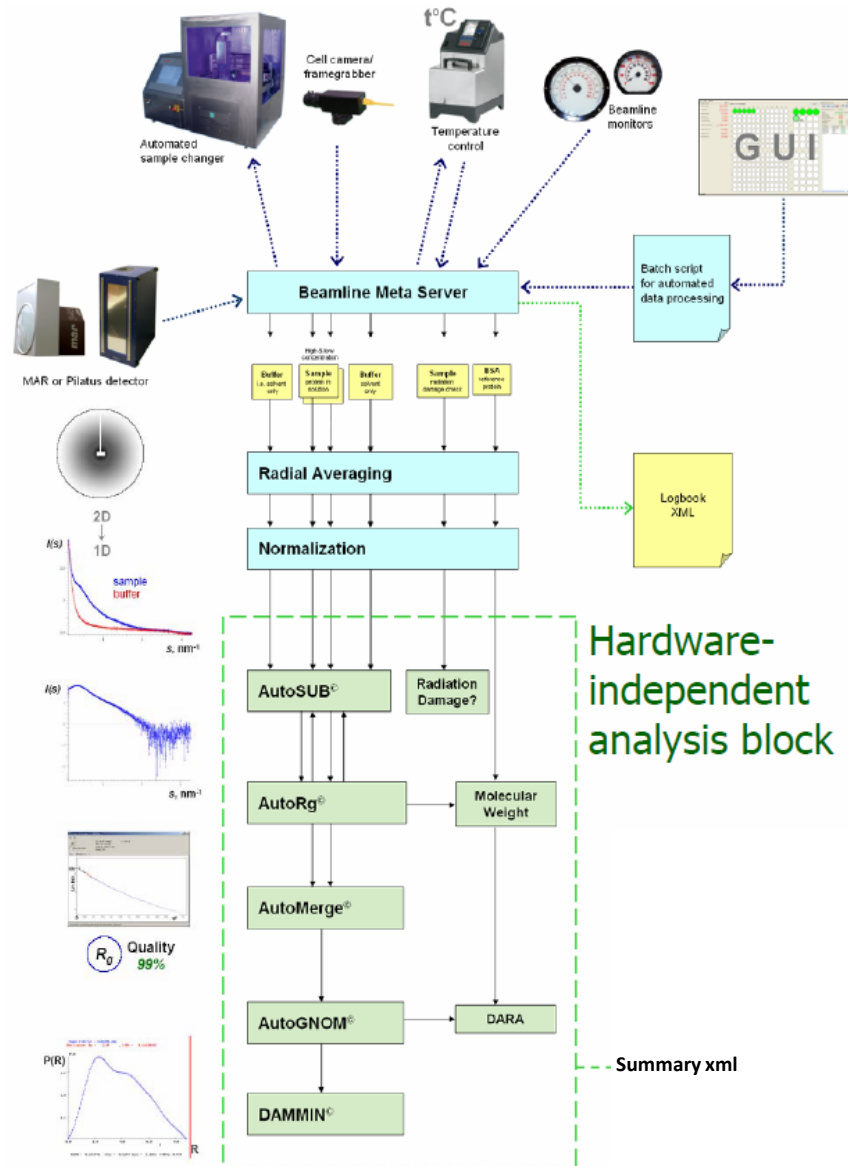
Problema di buffer subtraction: composizione esatta è sensibile alla presenza della proteina

In SEC-SAXS si sottrae il segnale del tampone che eluisce vicino al picco della proteina.



# BioSAXS:

## Set up sperimentale automatizzato





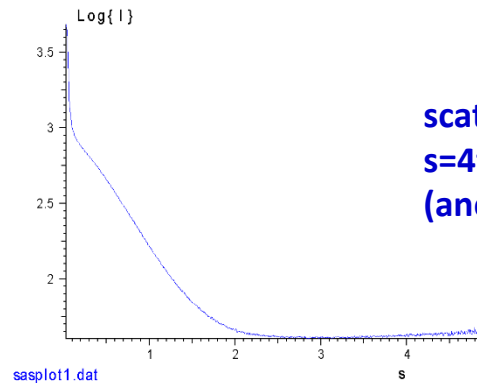
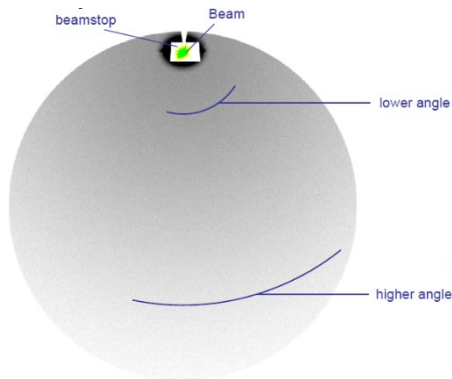
# BioSAXS:

## Analisi dei dati

### 1) Data reduction: integrazione radiale delle immagini 2D a una curva 1D

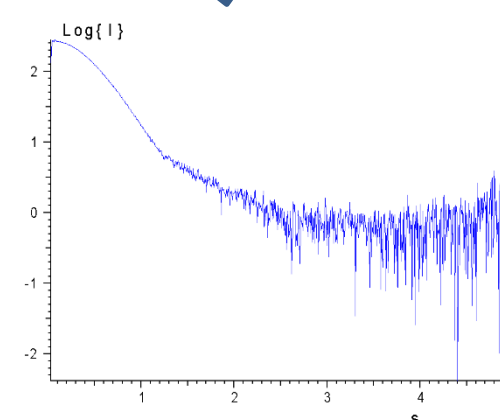
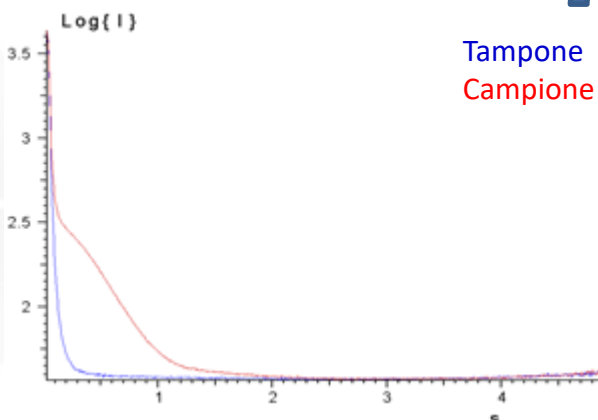
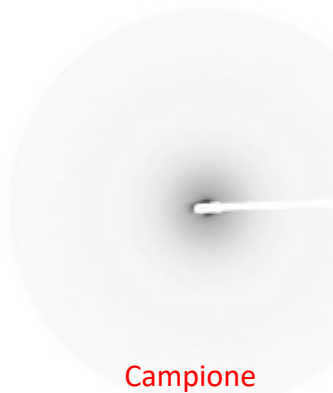
Coordinate esatte del centro del raggio

Dati normalizzati per I raggio trasmesso / t esposizione



scattering curve:  $s$  vs  $\text{Log}(s)$   
 $s = 4\pi \sin\theta / \lambda = q$   
(anche  $S = 2\sin\theta / \lambda = 2\pi s = s (!) = \mu \dots$ )

### 2) Averaging and Buffer subtraction



# BioSAXS:

## Qualità dei dati

### 1) Confronto diverse esposizioni e la media:

#### Danno da Radiazione

BM29, SOLIEL e SIBYLS hanno  $\sim 10^{12}$  -  $10^{13}$  fotoni/s,  
EMBL P12 arriva a  $\sim 10^{14}$  fotoni/s.

Ipotesi: la prima immagine (**esposizione**) non ha subito  
radiation damage

confrontata con tutte le altre

per sommare tutte quelle che sono consistenti con la prima  
Cambiamenti ascrivibili a radiation damage

che induce aggregazione

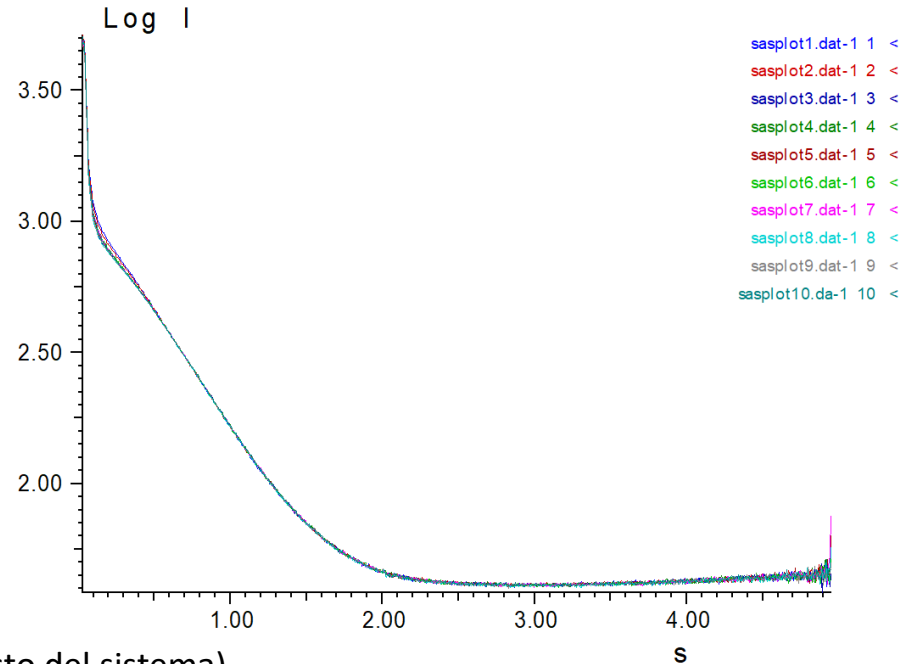
Riducibile mediante

- Additivi :
  - 3–5% v/v **glicerolo** ([ $\square$ ] più alte abbassano il contrasto del sistema).
  - agenti **riducenti** (1–2 mM DTT; 2–5 mM TCEP)
  - 1–2 mM **acido ascorbico**
- **Attenuazione del fascio** per ridurre il flusso di X-ray  
(comporta riduzione significativa nella qualità dei dati)

#### Average and Buffer subtraction:

Segnali delle proteine meno del **5% al di sopra del** livello del background; l'entità delle differenze campione/tampone modulabili cambiando

- 1) [ $\square$ ]
- 2) **densità elettronica media del solvente** (con saccarosio, glicerolo o sale)



# BioSAXS:

## SISTEMI IDEALI MONODISPERSI

**Monodispersione:** particelle identiche

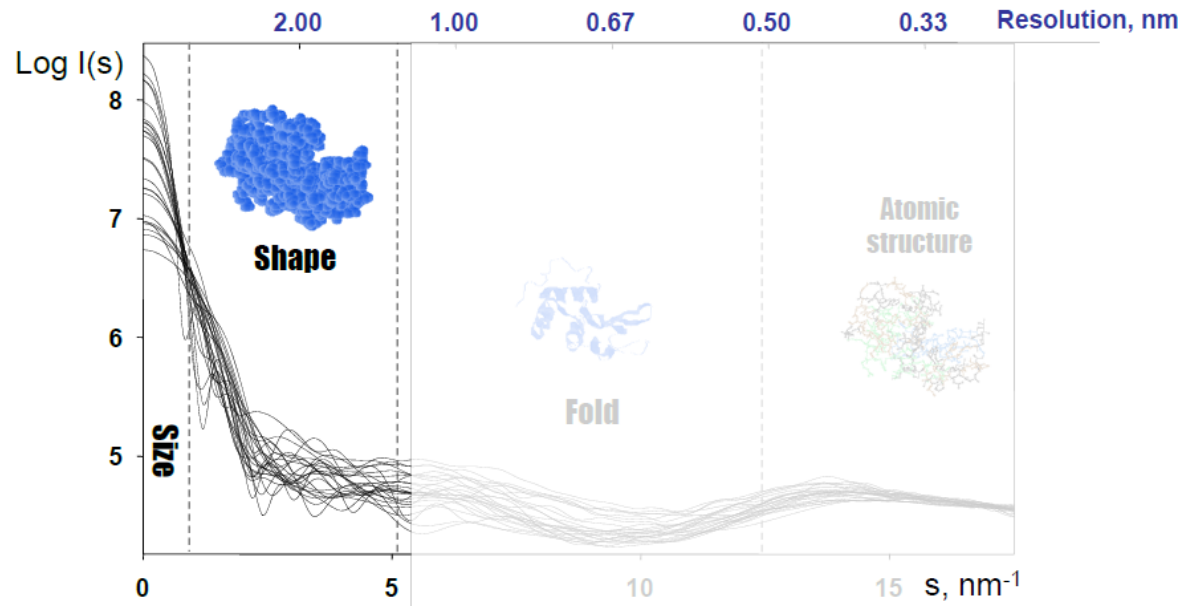
Deriva da protocollo di purificazione

Confermata da caratterizzazione biofisica (SEC, DLS, AUC, MALLS + RI, ecc.)

**Idealità:** nessuna interazione intermolecolare

raggiunta lavorando a diluizione infinita

In pratica: si eseguono misurazioni a concentrazioni decrescenti e si verifica che lo scattering è indipendente dalla concentrazione

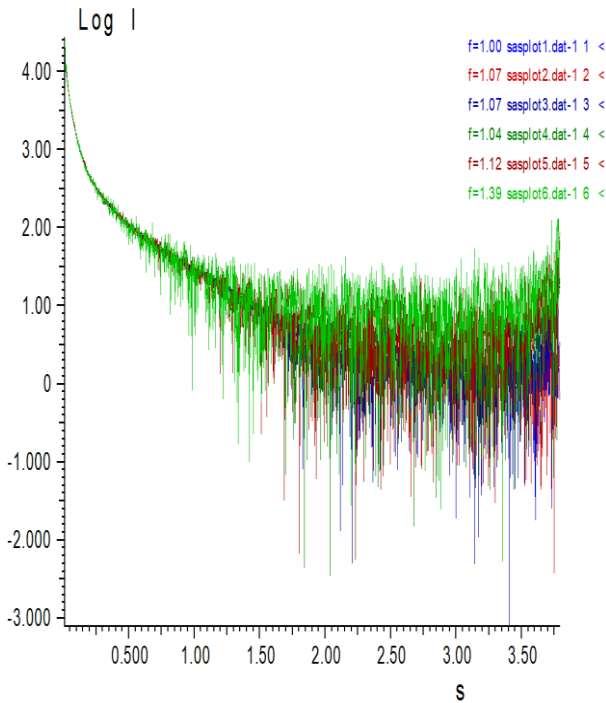


# BioSAXS:

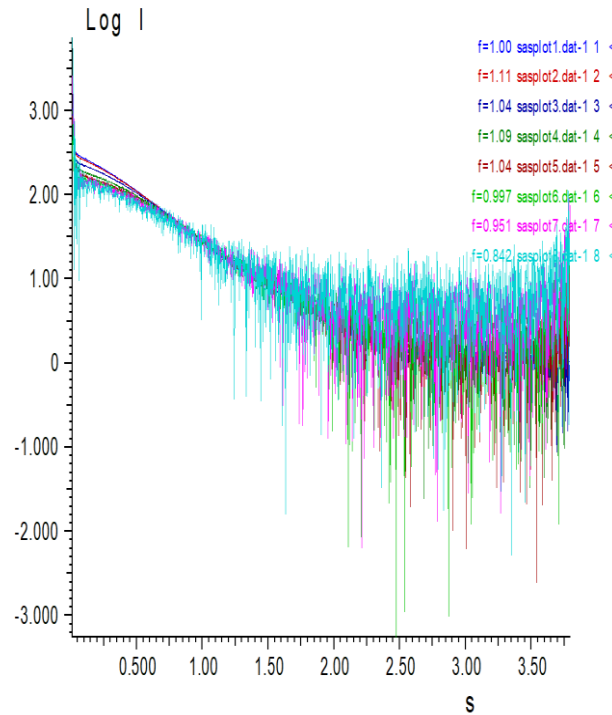
## Qualità dei dati

### 2) Confronto curve di scattering misurate a diverse []:

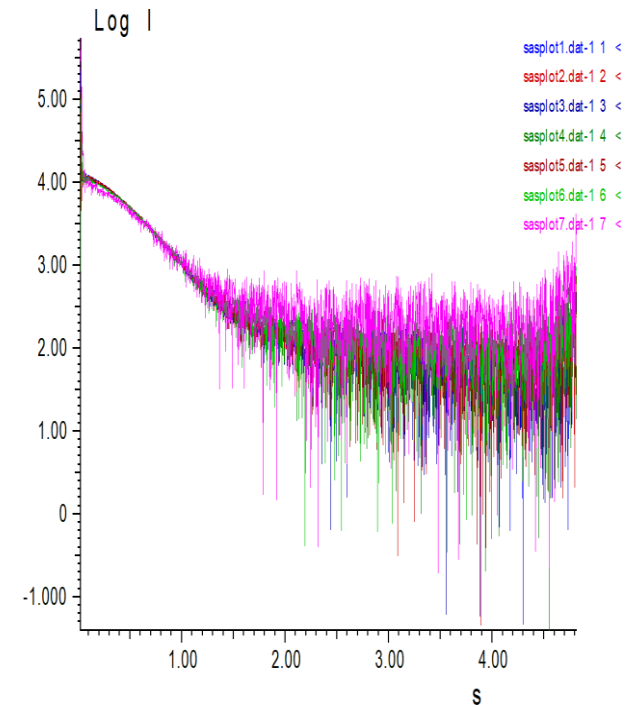
Aggregazione



Dipendenza da []



Interazioni fra particelle  
(attrattive/repulsive)



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 1) Informazioni su forma, dimensioni e conformazione

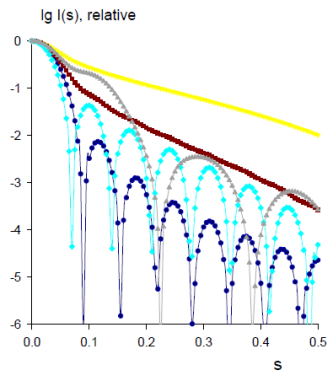
### Shape



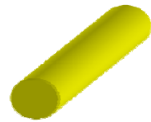
Solid sphere



Hollow sphere



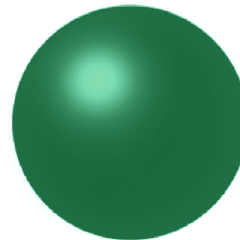
Dumbbell



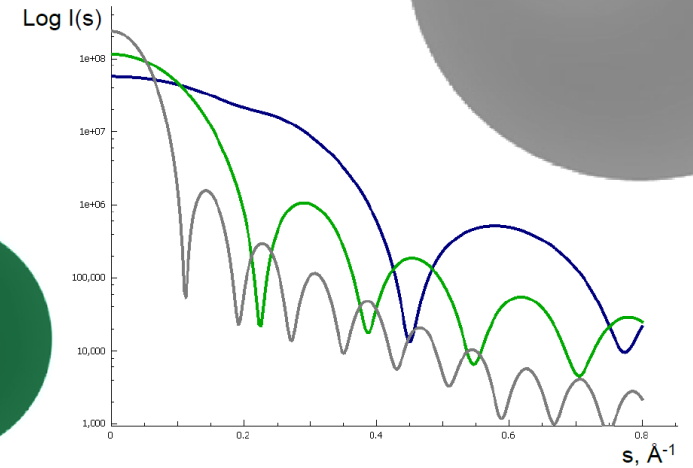
Long rod



Flat disc



### Size



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 1) Informazioni su forma, dimensioni e conformazione

Dalla **legge di Guinier**:  $\ln[I(s)] \approx \ln[I(0)] - s^2 R_G^2 / 3$

Intervallo di validità:  $0 < sR_G < 1.3$  per piccole particelle globulari

Deviazioni dalla linearità

→ interazione o aggregazione intermolecolare

### Raggio di girazione (Rg)

Descrive la distribuzione di massa di macromolecola attorno al suo centro di gravità

Misura la dimensione complessiva di una macromolecola

Indicativo per evidenziare modifiche conformazionali

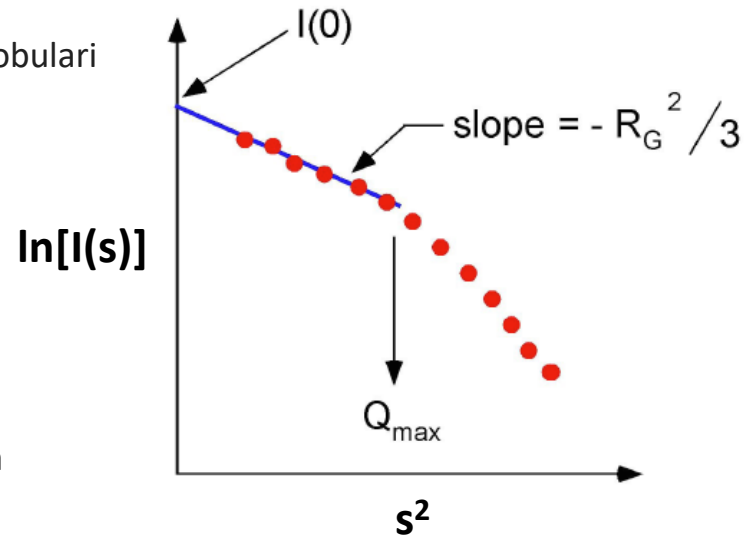
### Peso Molecolare (MM)

determinato dal valore dell'intensità all'origine  $I(0)$  conoscendo:

$[\eta]$ ,  
V parziale specifico,  
I assoluta

In pratica, determinato dai dati su scala relativa confrontando con  $I(0)$  di una proteina di riferimento (ad es. BSA, lisozima o citocromo C) con MM e stato di oligomerizzazione noti:

$$MM_{\text{sample}} = MM_{\text{reference}} * (I(0)_{\text{sample}} / I(0)_{\text{reference}})$$



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 1) Informazioni su forma, dimensioni e conformazione

Dalla **legge di Porod**: descrive il comportamento asintotico dell'I di scattering a grandi valori di s

$$I(s) \sim S s^{-4}$$

S area superficiale

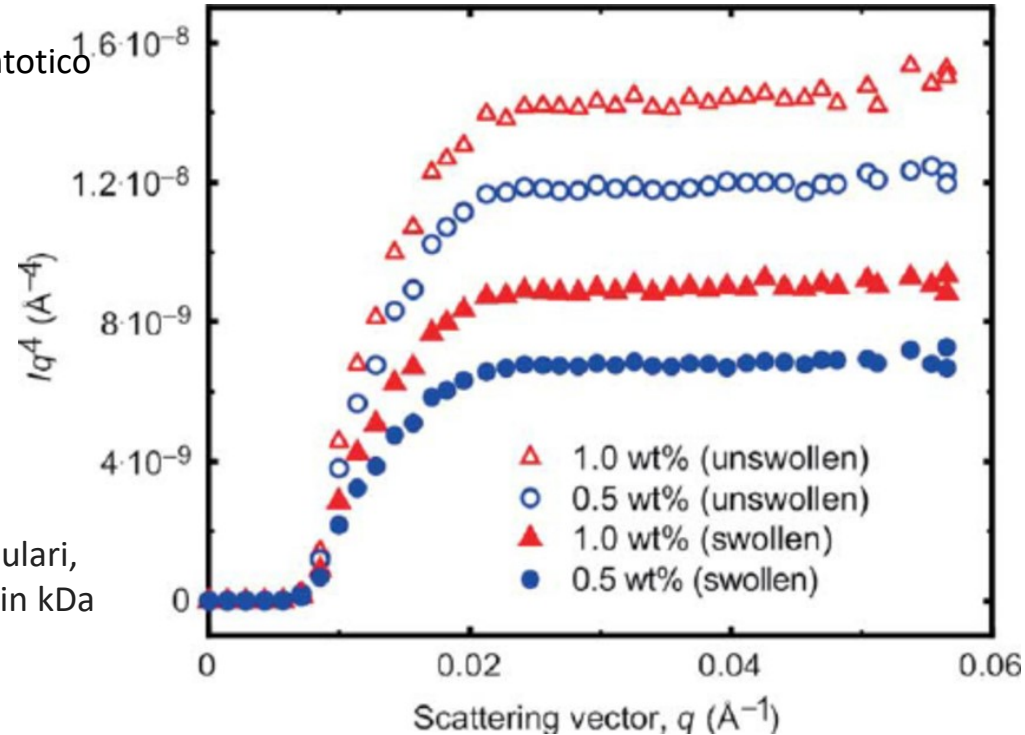
Volume escluso della particella idrata ( $V_p$ )

→ Peso molecolare (MM)

Secondo un risultato empirico per macromolecole globulari, il valore di  $V_p$  in  $\text{nm}^3$  è approssimativamente 2X il MM in kDa

$$V_p = \frac{2\pi^2 I(0)}{\int_0^\infty [I(s) - K_4] s^2 ds}$$

$K_4$  costante che assicura che il calo asintotico di I sia proporzionale a  $s^{-4}$  ad alti angoli



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

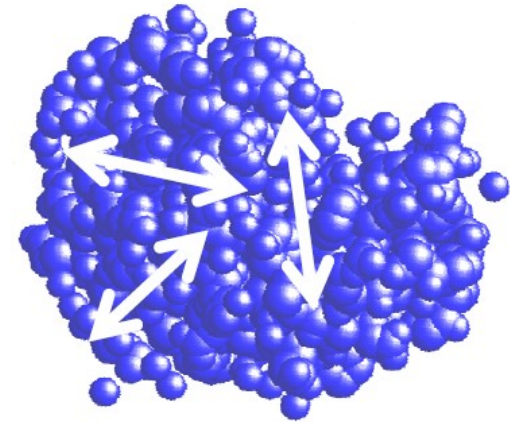
## 1) Informazioni su forma, dimensioni e conformazione

Dalla **Distance distribution function**  $p(r)=r^2\gamma(r)$

$$\gamma_0(r)=\gamma(r)/\gamma(0)$$

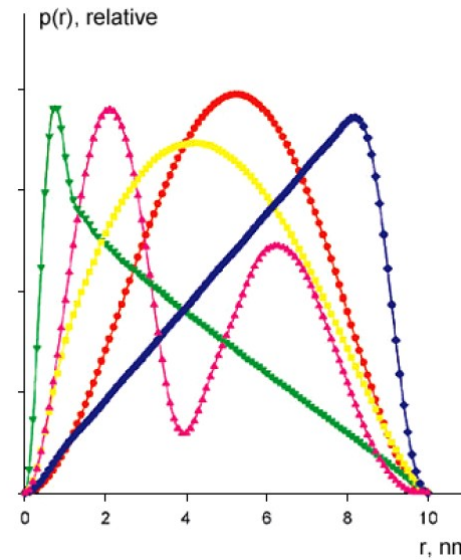
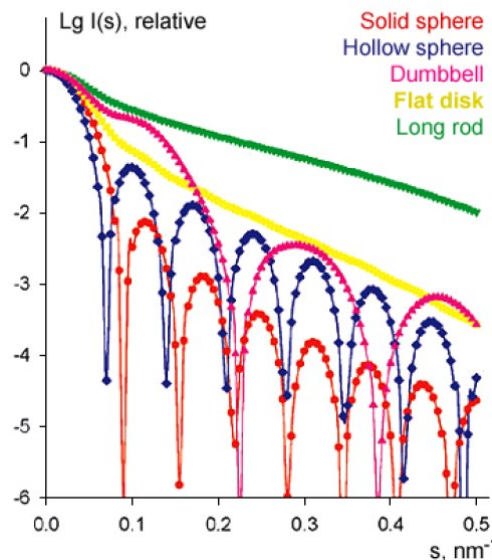
Probabilità di trovare un punto a distanza r da un dato punto all'interno della particella

$$I(s) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(sr)}{sr} dr \iff p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} s \frac{I(s) \sin(sr)}{sr} ds$$



Trasformata di Fourier indiretta

**Dmax = dimensione massima della particella**





# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

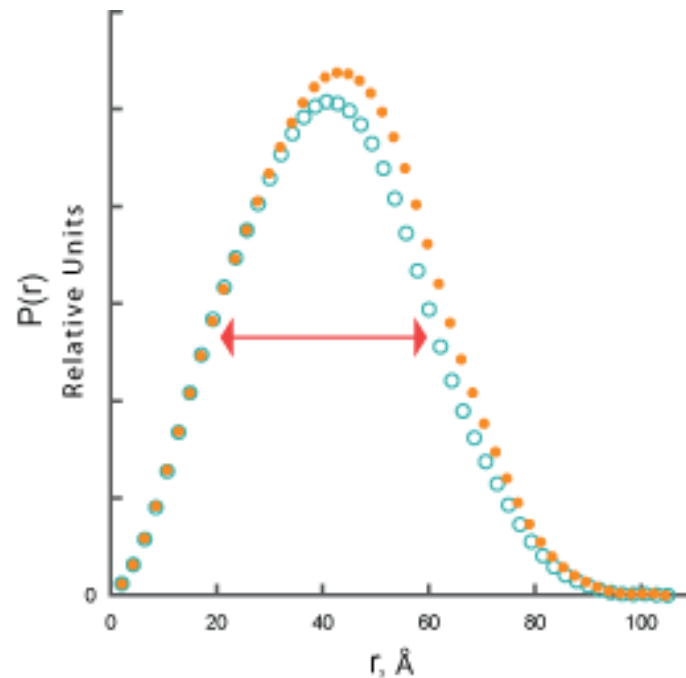
## 1) Informazioni su forma, dimensioni e conformazione

Dalla **Distance distribution function**  $p(r)=r^2\gamma(r)$

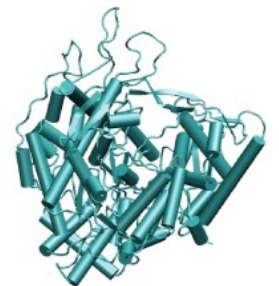
descrive l'insieme di tutte le **coppie di distanze** all'interno di una struttura

Cambiamenti nella posizione relativa di un dominio modificano la distribuzione di massa intorno al suo centro di gravità o

**Rg** («Real Space" Rg) ottenuto integrando la funzione con  $r^2$  su tutti i valori di  $r$



T7 Promoter Complex



T7 Promoter-Transition Complex

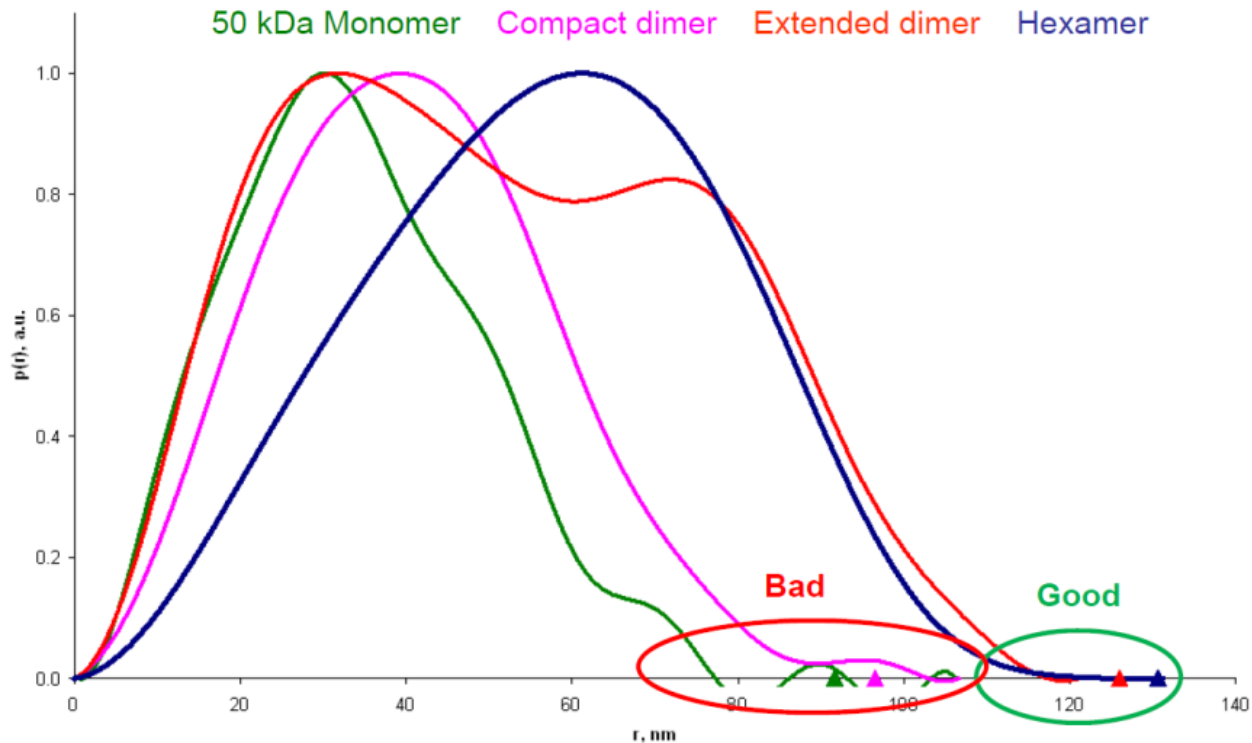


# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 1) Informazioni su forma, dimensioni e conformazione

Distance distribution function  $p(r)=r^2\gamma(r)$



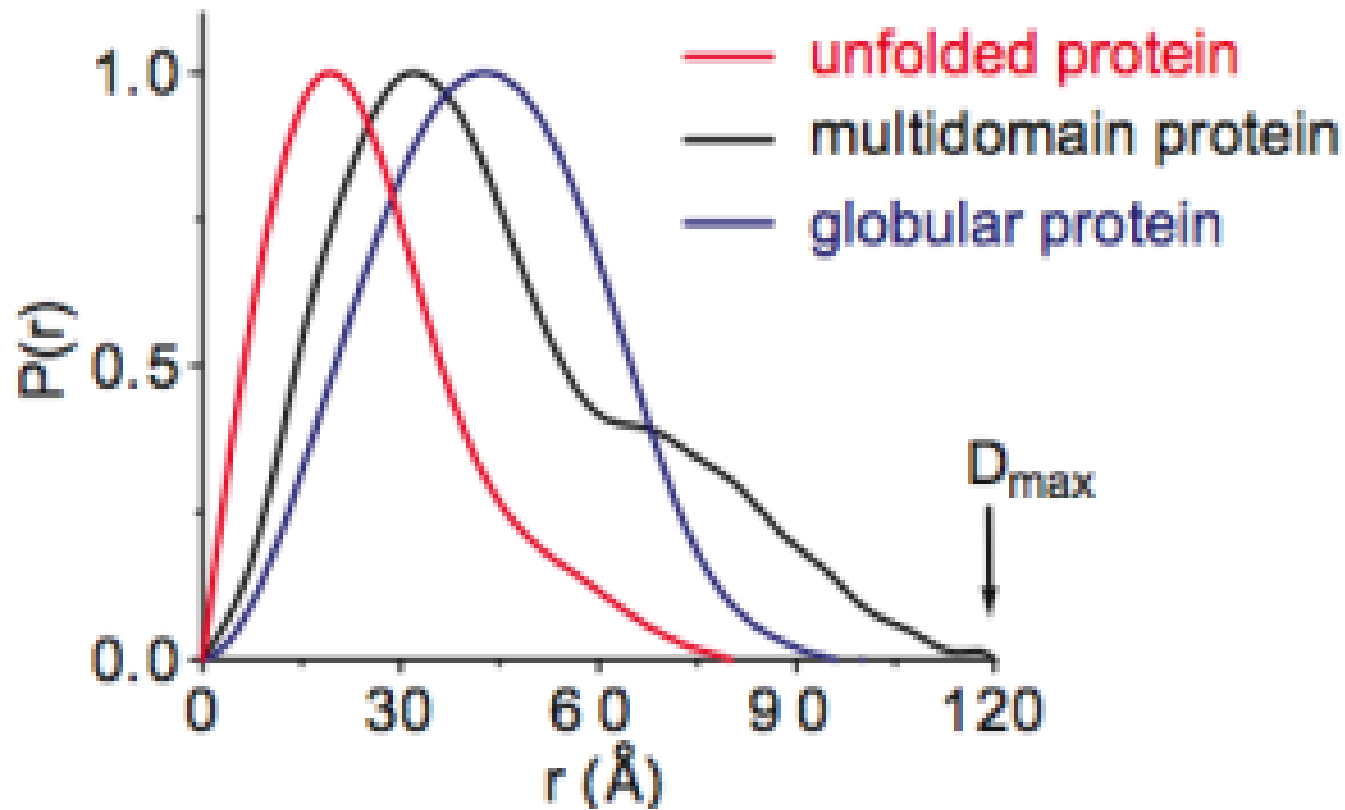
# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 1) Informazioni su forma, dimensioni e conformazione

Dalla **Distance distribution function**  $p(r)=r^2\gamma(r)$

caratteristiche qualitative all'interno di una distribuzione  $P(r)$  utili per comprendere il comportamento delle proteine



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

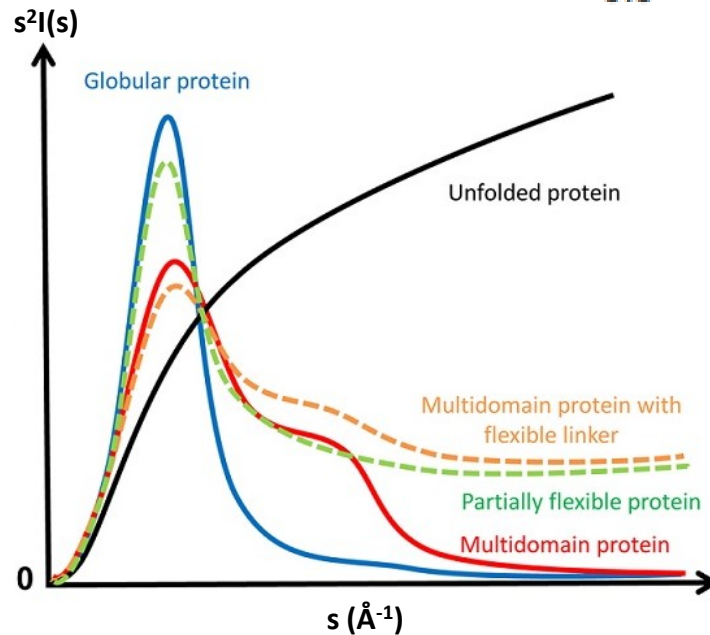
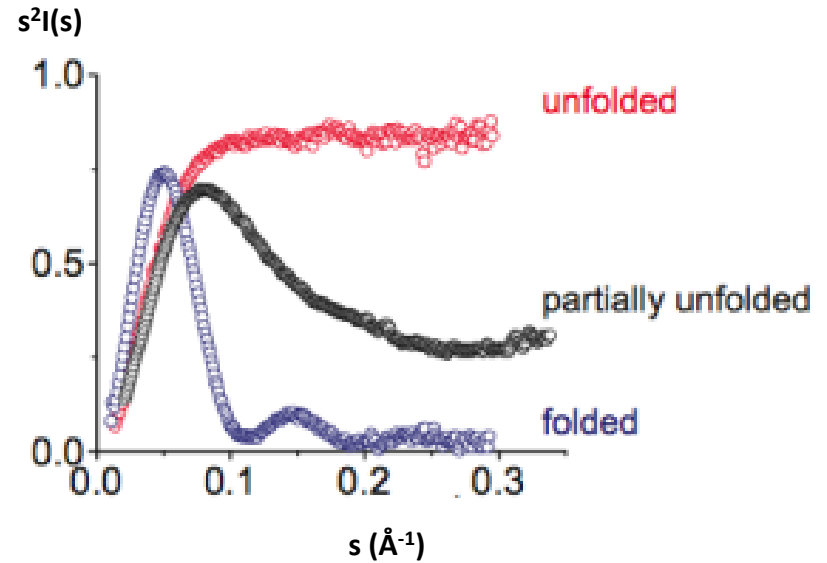
## 1) Informazioni su forma, dimensioni e conformazione

Dal **Kratky plot**:  $s$  vs  $s^2I(s)$

### Grado di compattezza / flessibilità

Il comportamento asintotico del decadimento dell'intensità nella regione di Porod dà informazioni sulla forma della particella (Legge di Porod)

### Stato del Folding



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

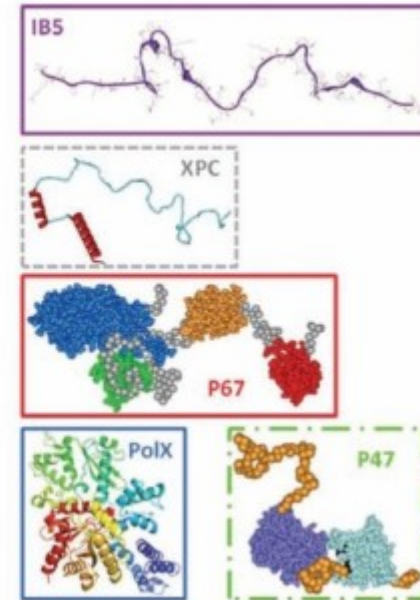
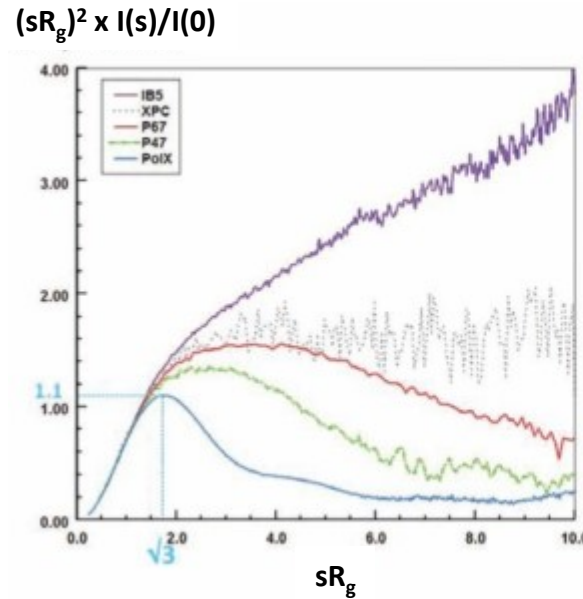
## 1) Informazioni su forma, dimensioni e conformazione

dal Dimension-less Kratky plot :  $(sR_g)^2 \times I(s)/I(0)$

$s R_g$  vs  $(s R_g)^2 I(s)/I(0)$

per confrontare macromolecole con MM  
e stati conformazionali diversi:  
senza dimensioni

→normalizzare per massa e concentrazione



(Receveur-Brechot V & Durand D (2012), *Curr Protein Pept. Sci.*, 13, 55-75)

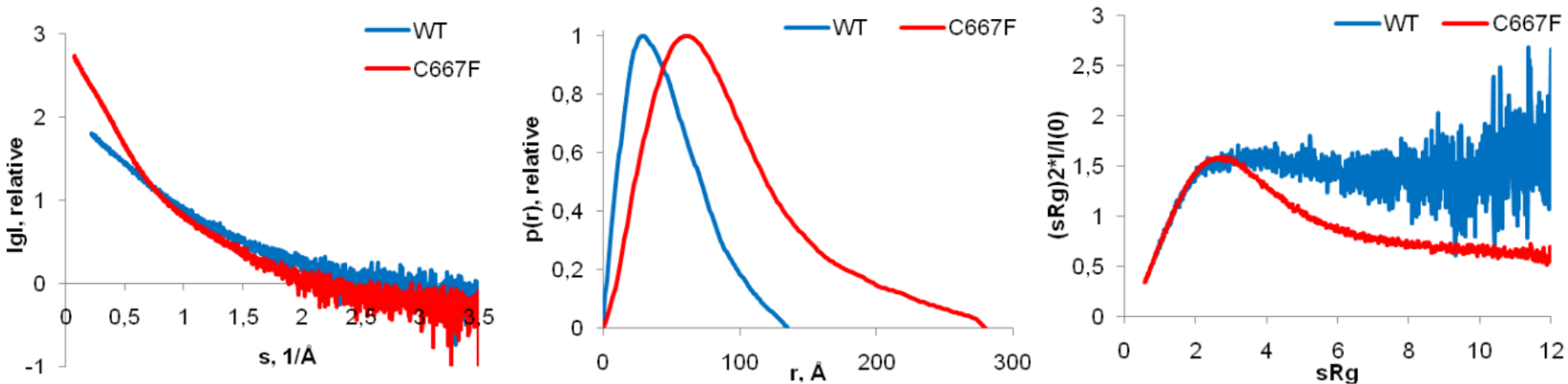
# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto:

Informazioni indipendenti dal modello:  
**Parametri strutturali**



Signorino G\*, Covaceuszach S\*, Bozzi M, Hübner W, Mönkemöller V, Konarev PV, Cassetta A, Brancaccio A, Sciandra F (2018) *Hum Mutat* 39(2):266-280



Oligomerization states of  $\beta$ -DG WT and C667F pathological mutant  
C667F missense mutation induces some degree of tertiary structures in  $\beta$ -DG.

# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 2) Informazioni su forma e modelli ab-initio

**Bead Modelling: DAMMIN** (Svergun (1999) *Biophys J.* 76, 2879-2886)

**DAMMIF** (Franke & Svergun (2009) *J. Appl. Cryst.* 42, 342–346)

simulated annealing

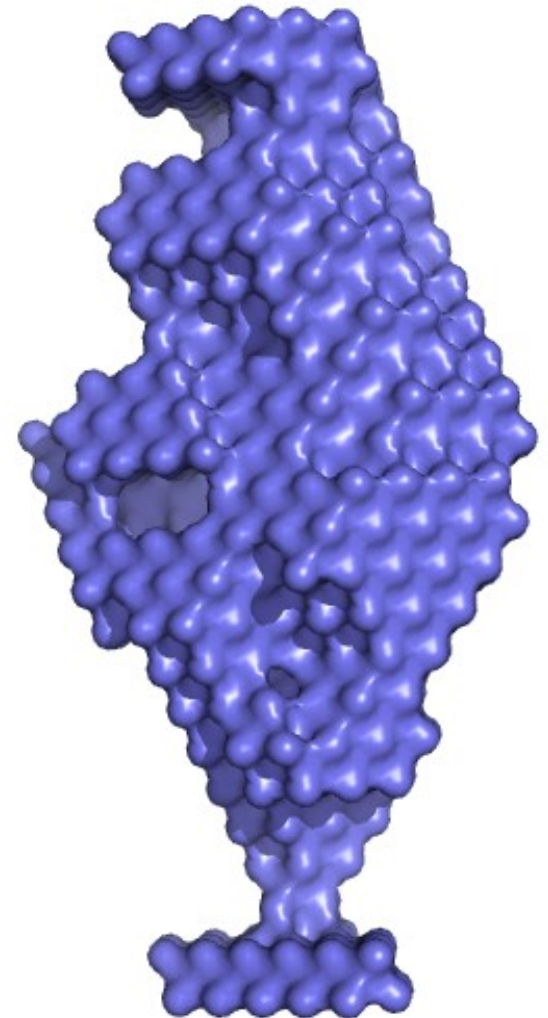
genera forme (**insiemi di beads**) con **proprietà protein-like** (in termini di **connettività, densità elettronica**)

Simmetria/anisometria

virtualmente a **qualsiasi complessità di forma,**

in **un intervallo molto ampio di dimensioni/MM**  
(*e.g. macromolecole, polimeri, nanoparticelle*)

Usa solo la parte **iniziale della curve di scattering**



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 2) Informazioni su forma e modelli ab-initio

### Bead Modelling: DAMMIN

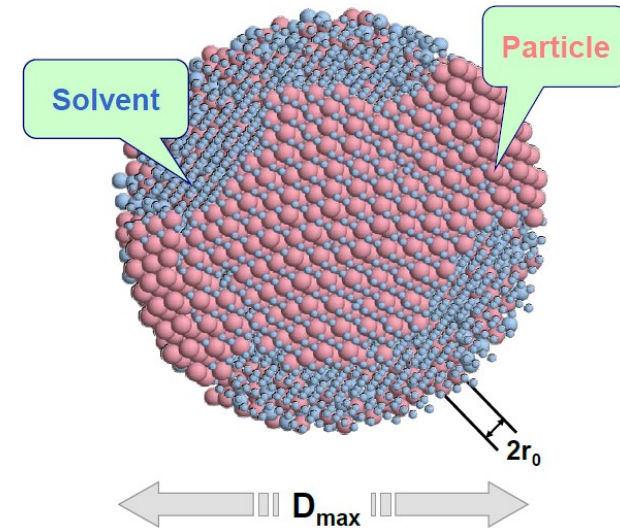
Una sfera di raggio  $D_{max}$  è riempita da beads densamente impaccate (SFERE DUMMY non corrispondenti ad atomi reali) di raggio  $r_0 \ll D_{max}$ : la FORMA del MODELLO DUMMY è descritta dal vettore dei parametri del modello:

Posizione (j) =  $X_j =$  1 se particella  
0 se solvente

$$\text{Target function} = \chi^2 [I_{exp}(s), I(X,s)] + \sum \alpha_i P_i(X)$$

Discrepanza dai dati sperimentali

Penalties per assicurare compattezza e connettività



Minimizzata per trovare una stringa binaria ottimale che descriva un assembly compatto e interconnesso di beads con un buon fitting dei dati.



# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 2) Informazioni su forma e modelli ab-initio

### Bead Modelling: DAMMIF

Completamente re-implementato (basato su un search volume adattivo)  
(25-40 volte più veloce)

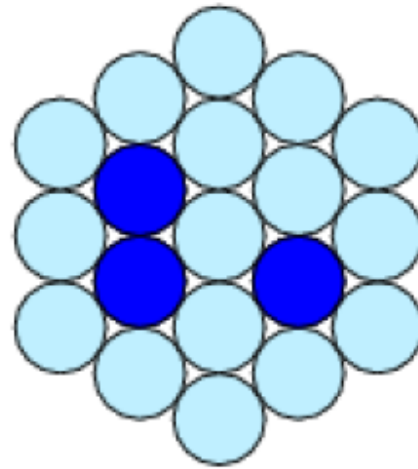
#### Single Phase Bead Models

A ogni singola iterazione:

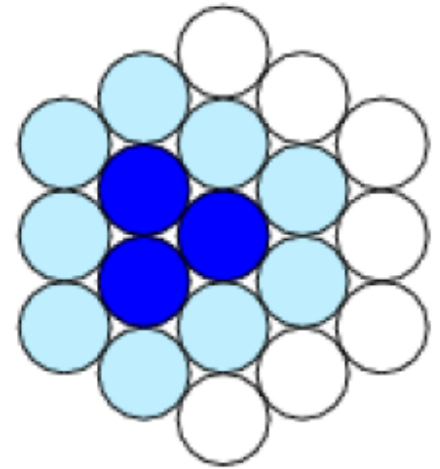
blu particella, può diventare solvente

Azzurro solvente, può diventare particella

bianco solvente, fisso



DAMMIN



DAMMIF

# BioSAXS:

Domande a cui SAXS può rispondere se NIENTE è noto

## 2) Informazioni su forma e modelli ab-initio

**Bead Modelling: GASBOR** (Svergun et al. (2001) *Biophys J.* 80, 2946-2953)

proteina  $\approx$  catena di **dummy residues**

cerca **chain-like arrangement** (contigui nello spazio con un determinato numero di vicini a brevi distanze per ogni atomo)

Questi residui sono "anonimi"

Simmetria/anisometria

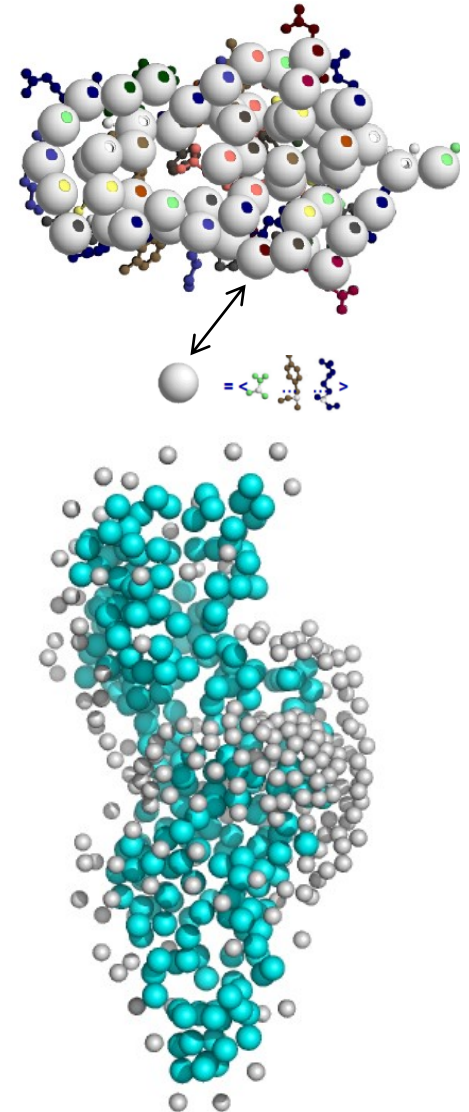
Opzione di fitting sia nello spazio reciproco che reale.

Applicabile **per catene polipeptidiche** (*i.e. proteine e loro assemblies*) fino a **5000aa**

Fitting delle curve di scattering ad angoli più alti (fino a risoluzione di 5Å)

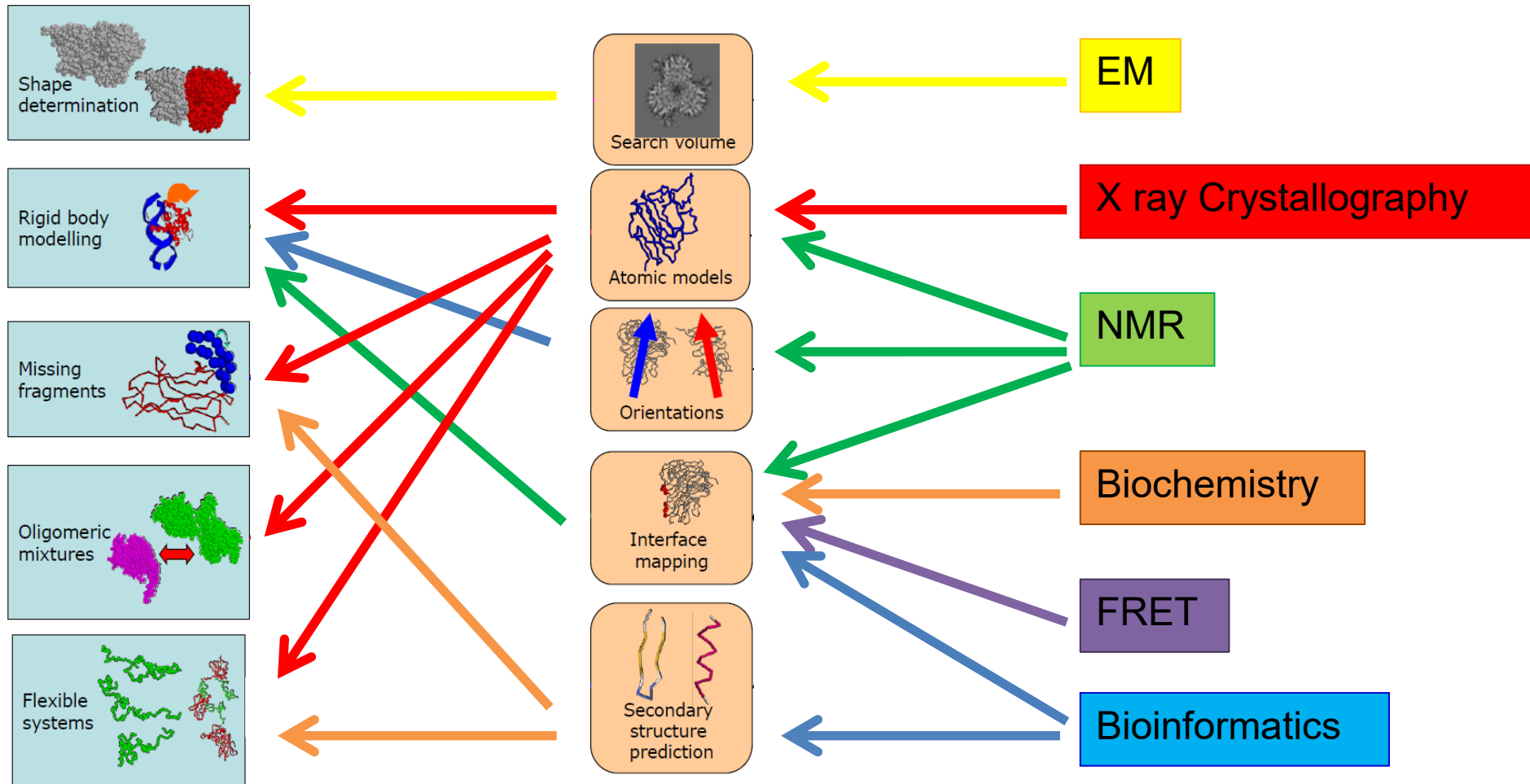
## 3) Caratterizzazione di sistemi polidispersi

equilibri dinamici monomeri/oligomeri (**dissociazione di multimeri**)



# BioSAXS:

Domande a cui SAXS può rispondere se sono disponibili **informazioni strutturali ad alta risoluzione INCOMPLETE/COMPLETE**



# BioSAXS:

Domande a cui SAXS può rispondere se sono disponibili **informazioni strutturali ad alta risoluzione**  
**INCOMPLETE/COMPLETE**

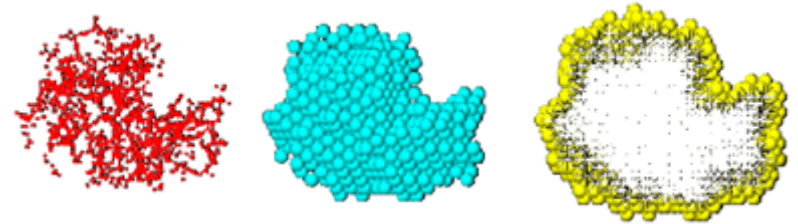
## 1) Validazione in soluzione di modelli cristallografici ad alta risoluzione /Identificazione di oligomeri biologicamente attivi

Intensità di Scattering della macromolecola in soluzione:

Scattering **Atomico in vacuo**

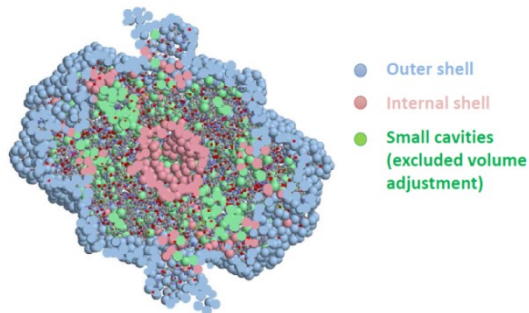
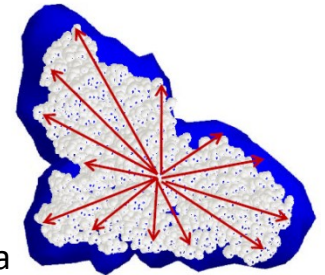
- Scattering dal **volume Escluso**

+ Scattering dalla **sfera di idratazione**



**CRYSOL:** (Svergun et al. (1995). J. Appl. Cryst.28, 768):

lo strato di idratazione è modellato come **strato implicito** usando una **envelope function** che si adatta ai dati sperimentali variando la densità dello strato di idratazione e il volume totale escluso



**CRYSOL 3.0:**

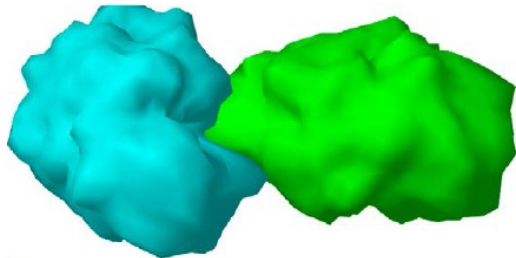
**per forme complesse**, l'envelope function non fornisce una descrizione adeguata

**Sfera di idratazione rappresentata con dummy beads di solvente classificati** come esterni, interni e piccole cavità.

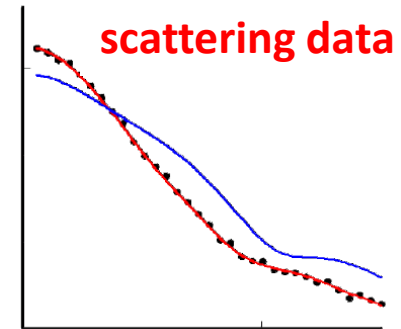
# BioSAXS:

Domande a cui SAXS può rispondere se sono disponibili **informazioni strutturali ad alta risoluzione INCOMPLETE/COMPLETE**

## 2) Strutture quaternarie mediante rigid body refinement



Trial and error



Ricerca non-lineare:

complesso arbitrario costruito **traslando/ruotando i componenti** (3 parametri rotazionali e 3 traslazionali).  
Informazioni aggiuntive da altri metodi complementari per risolvere/ridurre l'ambiguità

**SASREF:** (Petoukhov & Svergun (2005) *Biophys J.* 89, 1237; (2006) *Eur. Biophys. J.* 35, 567)

simulated annealing

**disposizione spaziale interconnessa senza clashes** con traslazioni / rotazioni casuali ad ogni passo, adattando la curva di scattering sperimentale.

Più set di dati da costruiti parziali per ottenere un adattamento simultaneo di più curve di scattering

Alcune subunità possono essere fissate nelle posizioni e negli orientamenti iniziali per mantenere la loro disposizione reciproca

Per le glicoproteine, strumento integrato per la generazione di glicani

# BioSAXS:

Domande a cui SAXS può rispondere se sono disponibili **informazioni strutturali ad alta risoluzione INCOMPLETE/COMPLETE**

## 3) Modelling di proteine multidominio / aggiunta di frammenti mancanti

probabile configurazione delle porzioni mancanti

**BUNCH:** (Petoukhov & Svergun (2005) *Biophys J.* 89, 1237)

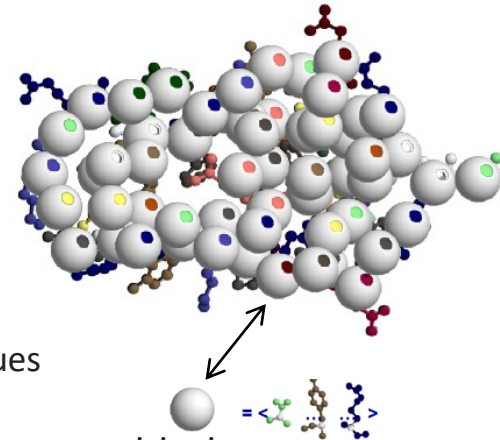
combina corpo rigido e ab initio modelling per trovare

le posizioni e gli orientamenti di domini rigidi e

le probabili conformazioni di linker flessibili rappresentati come catene di dummy residues

Necessari corretti legami e angoli diedri nelle catene dei dummy residues e assenza di sovrapposizioni

**Può trattare solo una catena singola**



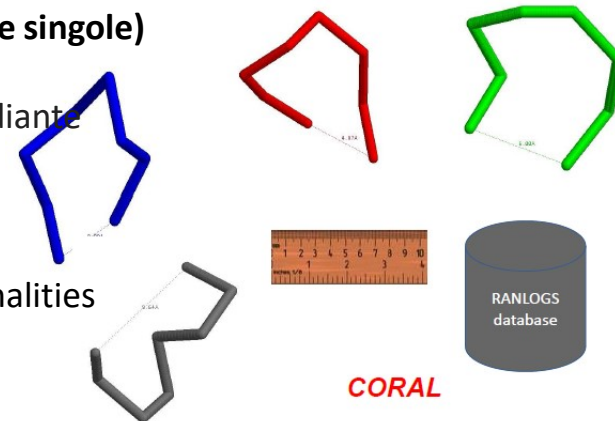
**CORAL (Complexes with Random Loops):** (Petoukhov et al. (2012) *J. Appl. Cryst.* 45, 342-350)

Combina SASREF (non ricostruisce regioni mancanti) e BUNCH (lavora su catene singole)

**Rigid body modelling** mantenendo vincolate le distanze tra N e C terminale mediante una libreria di loop casuali composta da dummy residues.

Alcune sottounità possono essere fissate

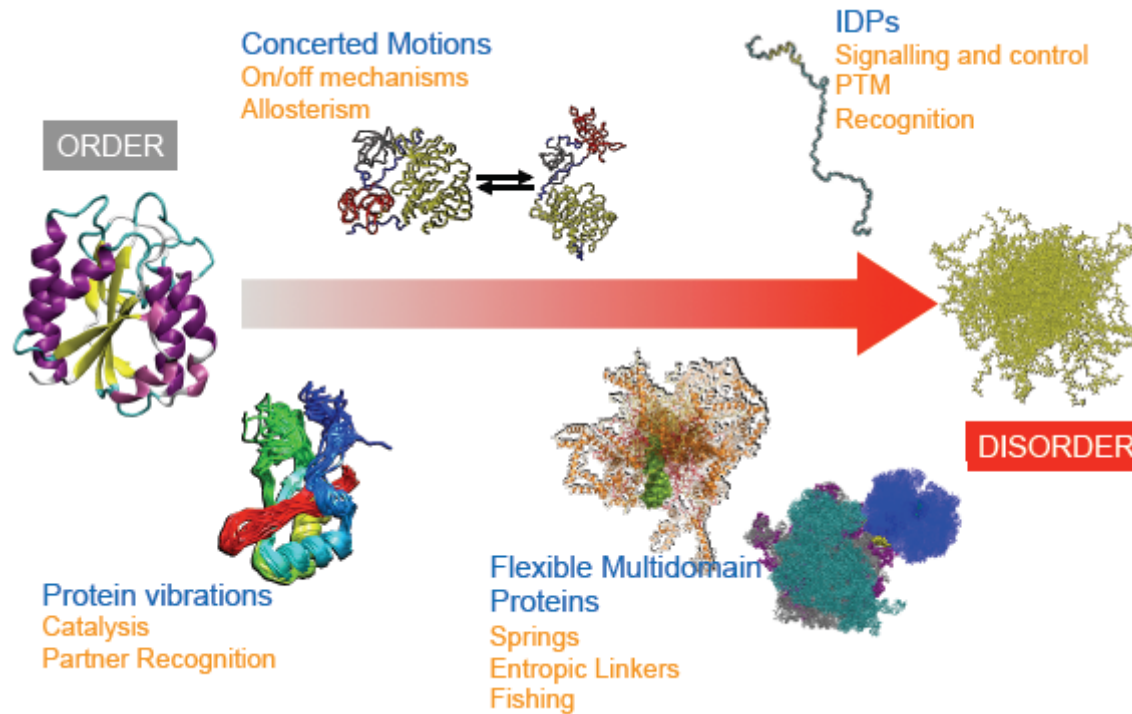
Informazioni aggiuntive da altri metodi complementari da aggiungere come penalties



# BioSAXS:

## SISTEMI POLIDISPERSI

### 1) Polidispersione di forma



La cellula è un ambiente affollato e in continua evoluzione in cui le biomolecole interagiscono. Ogni singola struttura proteica non è un'entità 3D rigida ma una distribuzione spazio-temporale di un numero indefinito di stati conformazionali

questo insieme di conformazioni definisce le funzioni biologiche

# BioSAXS:

## SISTEMI POLIDISPERSI

### 1) Polidispersione di forma

Caratterizzazione quantitativa di sistemi flessibili / proteine multidominio con linker flessibili/  
proteine intrinsecamente non strutturate

#### EOM (Ensemble optimization method):

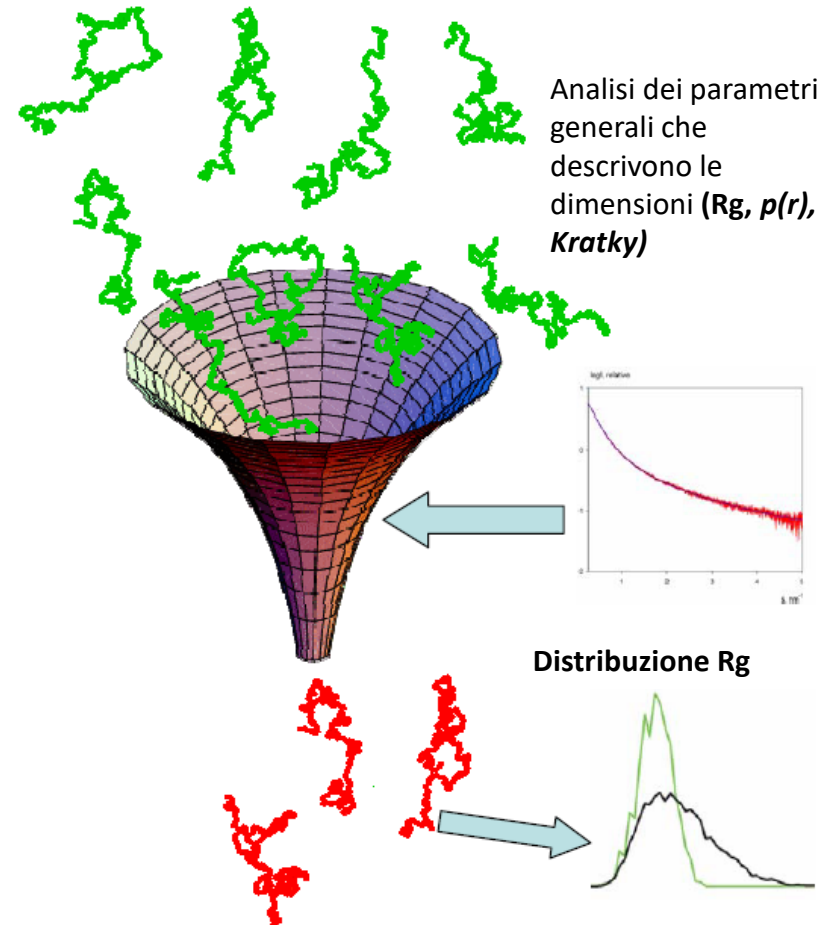
(Tria et al. (2015) *IUCr* **2**, 207-217)

Permette la **co-esistenza di conformazioni multiple** (generate da un algoritmo genetico), **selezionate da** un ampio random pool iniziale per trovare un insieme con le stesse caratteristiche del campione.

Simmetria

Assemblies Multimerici

Fitting di Curve Multiple





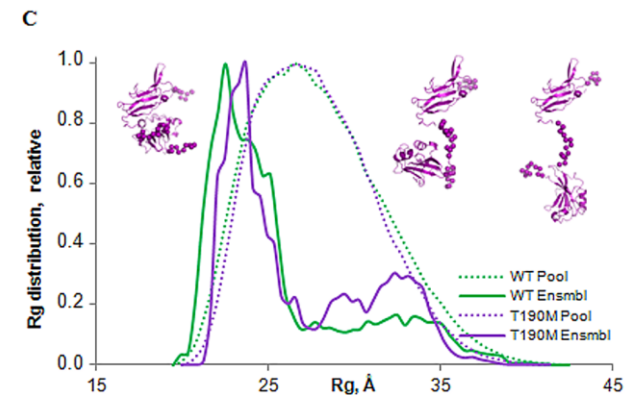
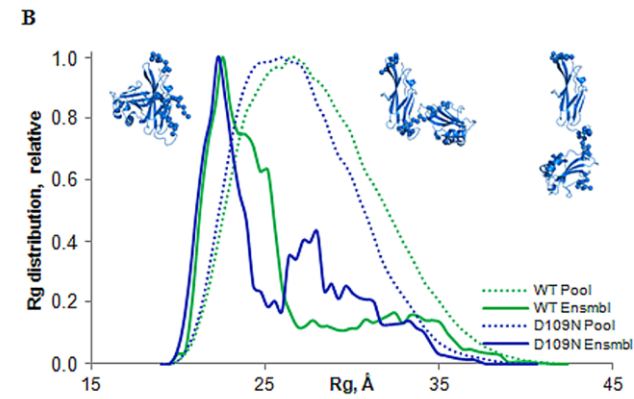
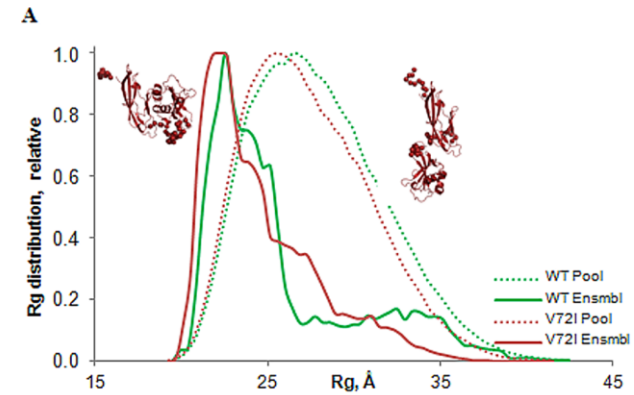
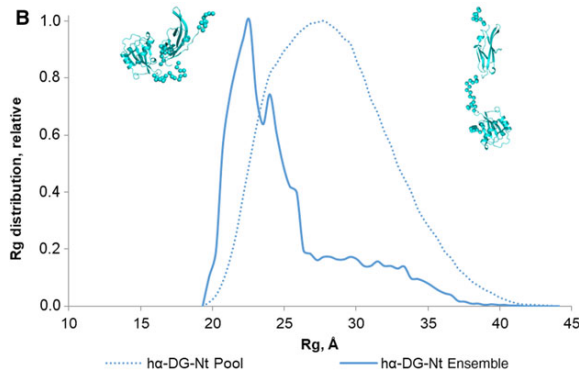
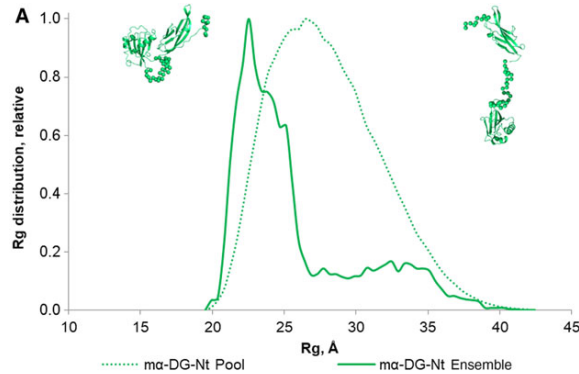
# BioSAXS:

## SISTEMI POLIDISPERSI

### 1) Polidispersione di forma

Covaceuszach S, Bozzi M, Bigotti MG, Sciandra F, Konarev PV, Brancaccio A, Cassetta A (2017)  
*PLoS One* 12(10):e0186110

Covaceuszach S, Bozzi M, Bigotti MG, Sciandra F, Konarev P, Brancaccio A, Cassetta A (2017)  
*FEBS Open Bio.* 17;7(8):1064



# BioSAXS:

## SISTEMI POLIDISPERSI

### 2) Polidispersione di dimensioni

#### Caratterizzazione quantitativa di assemblies/miscele:

stato di associazione in soluzione o all'interno di membrane  
impatto della variazione delle condizioni del solvente

#### Miscele oligomeriche all'equilibrio :

Il numero minimo di curve indipendenti richieste per rappresentare l'intero set di dati mediante le loro combinazioni lineari è derivato dal numero di vettori singolari significativi espresso dal parametro **SVD (decomposizione del valore singolare** ovvero le curve non casuali con valori singolari significativi)

#### SVDPLOT

calcola l'SVD e visualizza i vettori e i valori singolari,  
stima il **numero di componenti indipendenti**

**OLIGOMER** (Konarev et al. (2003) *J. Appl. Cryst.* 36, 1277) stima la **frazione dei componenti**

#### Fitting di Curve Multiple

# BioSAXS:

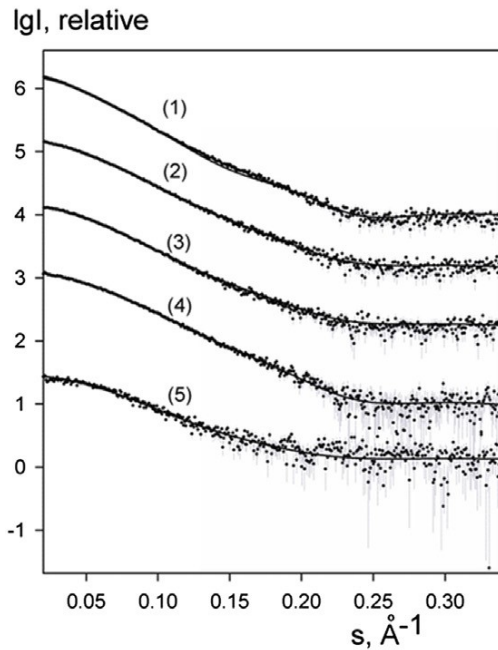
## SISTEMI POLIDISPERSI

### 2) Polidispersione di dimensioni

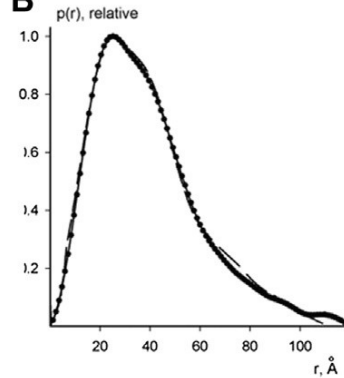
Covaceuszach S, Konarev PV, Cassetta A, Paoletti F, Svergun D., Lamba D, Cattaneo A (2015)

Biophysical Journal Volume 108 687–697

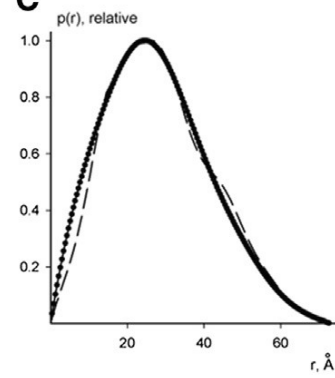
**A**



**B**



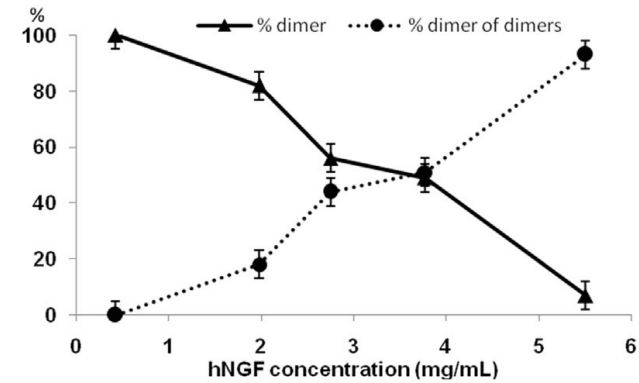
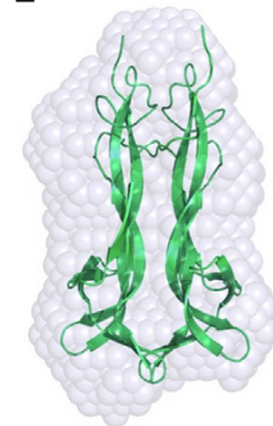
**C**



**D**



**E**



# BioSAXS:

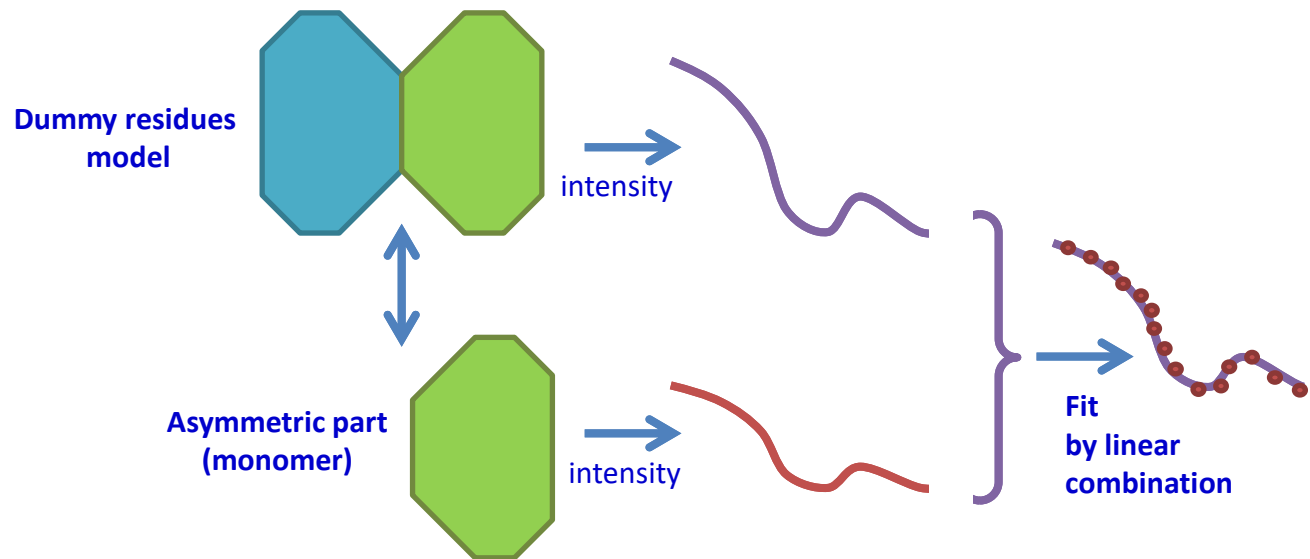
## SISTEMI POLIDISPERSI

### 2) Polidispersione di dimensioni

- **Equilibri dinamici fra monomeri/oligomeri (dissociazione di multimeri)**

**GASBORMX** (*Petoukhov et al. (2013) Biochemistry, 52, 6844*):

### Fitting simultaneo di Curve Multiple



# BioSAXS:

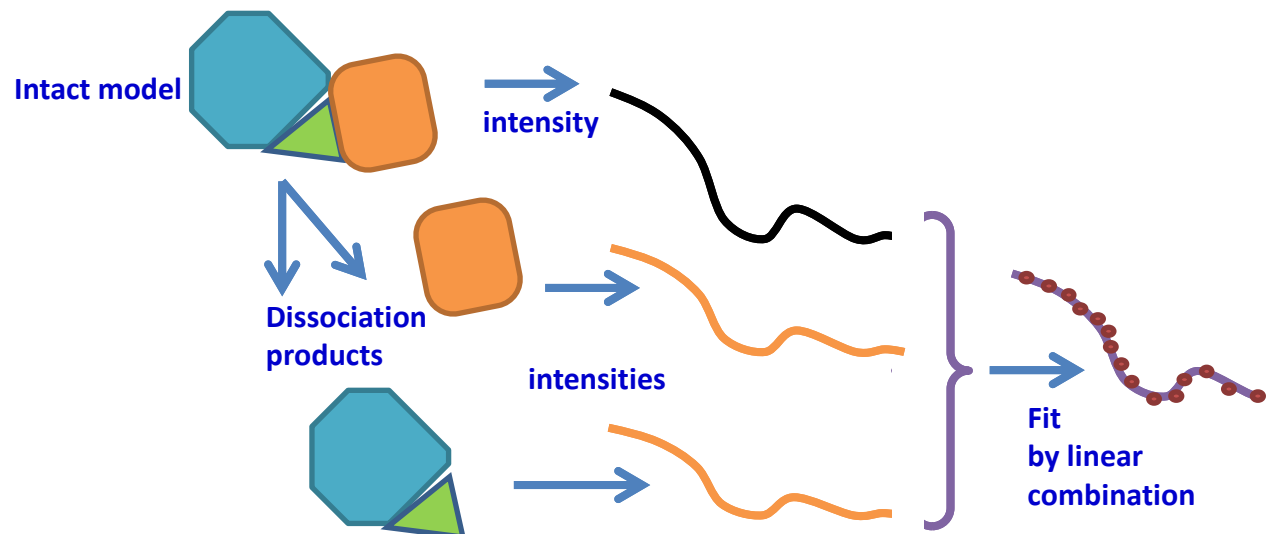
## SISTEMI POLIDISPERSI

### 2) Polidispersione di dimensioni

- **Equilibri dinamici** fra componenti liberi e in complesso (**complessi transienti a bassa affinità**)

**SASREFMX** (*rigid body modelling*) (Petoukhov et al. (2012) *J. Appl. Crystallogr.* 45,342)

**Rigid body modelling di assemblies transienti**



# BioSAXS:

## APPLICAZIONI

### IN ASSENZA DI DATI COMPLEMETARI

- **folding, unfolding, aggregazione**
- **Forma, conformazione e dimensioni**
- modelli a bassa risoluzione **ab initio**

### IN PRESENZA di dati strutturali **INCOMPLETI** :

- Configurazioni probabili delle regioni mancanti

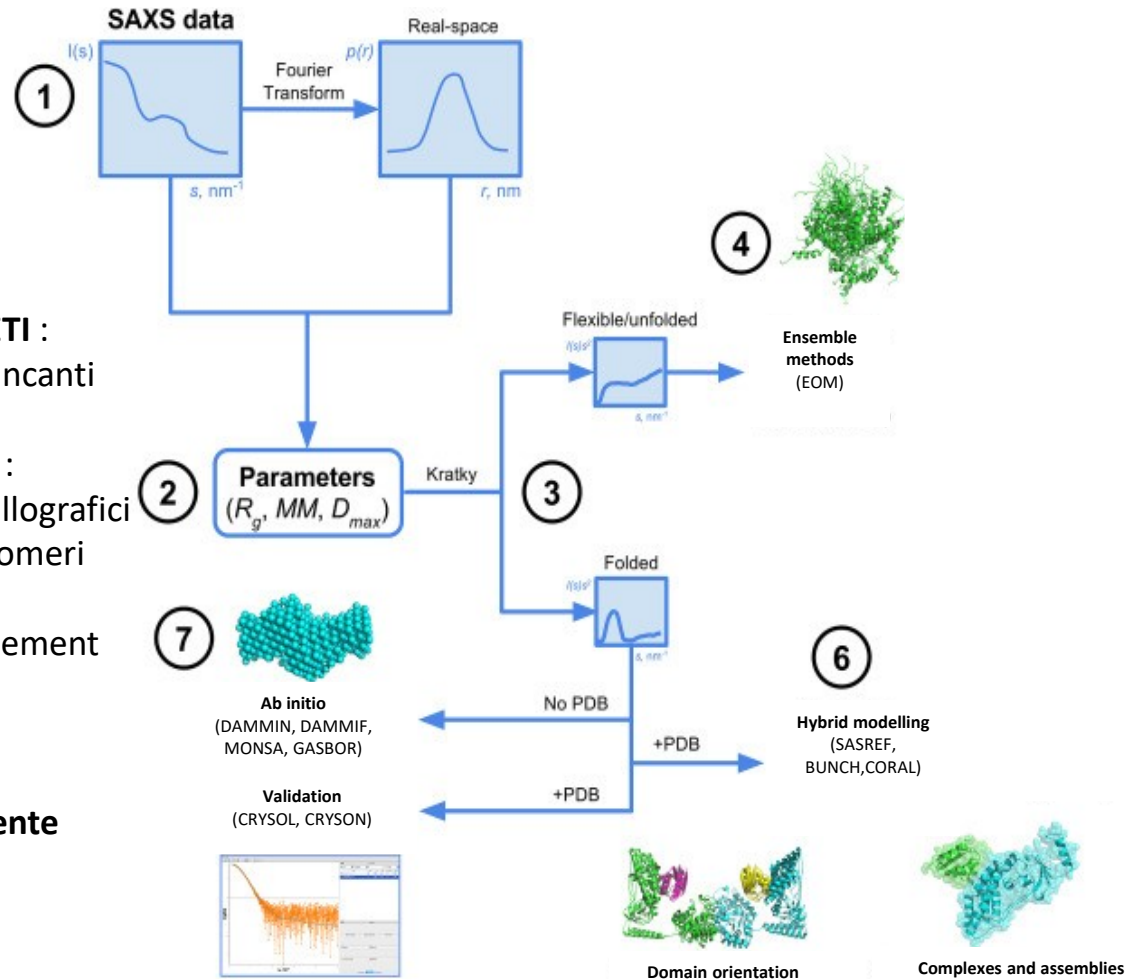
### IN PRESENZA di dati strutturali **COMPLETI** :

- **validazione** in soluzione di modelli cristallografici ad alta risoluzione /Identificazione di oligomeri biologicamente attivi

- **struttura quaternaria** da rigid body refinement

Sia in sistemi ideali monodispersi che in **sistemi polidispersi (Miscele/assemblies)**

- **Sistemi Flessibili /proteine intrinsecamente disordinate**



# BioSAXS:

## Validazione dei dati sperimentali

### 1) Qualità del campione

Purezza/idealità verificate con metodi biofisici e analitici (MALLS, SDS-PAGE, DLS, AUC)

### 2) Qualità dei dati

• **Linearità del Guinier plot** → campione monodisperso

Deviazioni:

aggregazione

interazioni repulsive

• **Confronto dei parametri indipendenti dal modello** ( $D_{max}$ ,  $MW$ ,  $R_g$ )

Confronto delle stime dei MM basate su parametri indipendenti

problemi nell'andamento della  $p(r)$  ad alte  $d$  → aggregazione

• **Dipendenza dal tempo** ( $R_g$  e  $I(0)$ )

→ aggregazione indotta da Radiazione

• **Dipendenza dalle []** ( $I(0)/c$  e  $R_g$  non costanti all'interno della serie di [])

→ processi di oligomerizzazione

# BioSAXS:

## Validazione del modello ab initio

ab initio modelling basato su dati unidimensionali è intrinsecamente ambiguo.

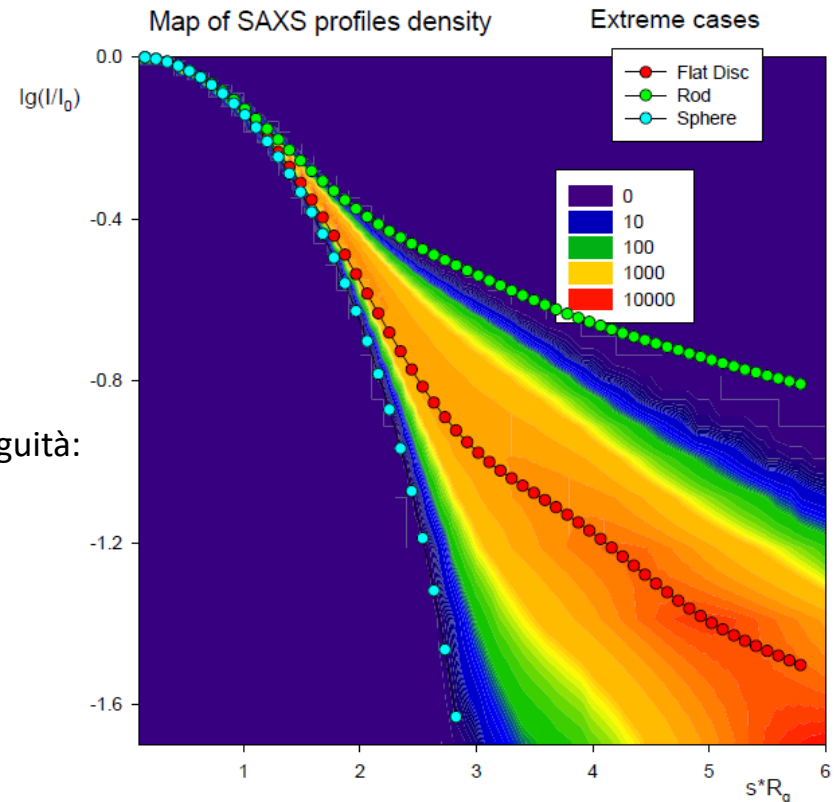
## Valutazione dell'ambiguità indipendente dal modello

**AMBIMETER** (Petoukhov et al. (2015). *Acta Cryst. D71*, 1051)

basato su una libreria di valori normalizzati profili di scattering (eliminando le informazioni sulla dimensione e mantenendo le informazioni sulla topologia) da oltre 14.000 topologie di forme.

La potenziale ambiguità = numero di topologie compatibili con i dati sperimentali, fornendo il punteggio di ambiguità:  
sotto 1,5 → forma ab initio potenzialmente unica  
sopra 2,5 → potenziale ambiguità delle forme

Come regola generale, le strutture altamente oblate sembrano essere le più difficile per la ricostruzione ab initio dai dati SAXS.



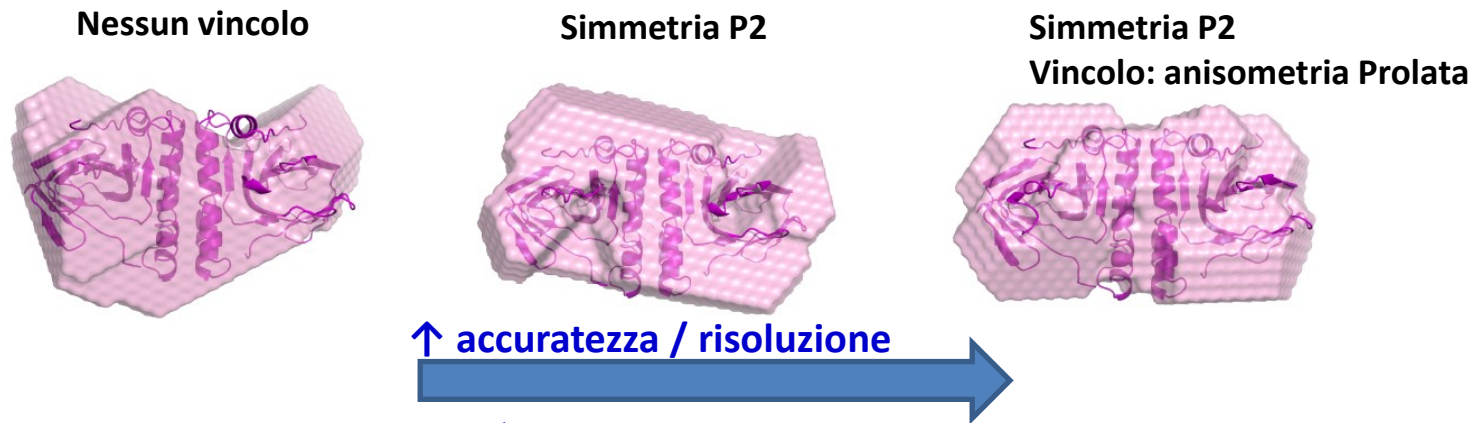


# BioSAXS:

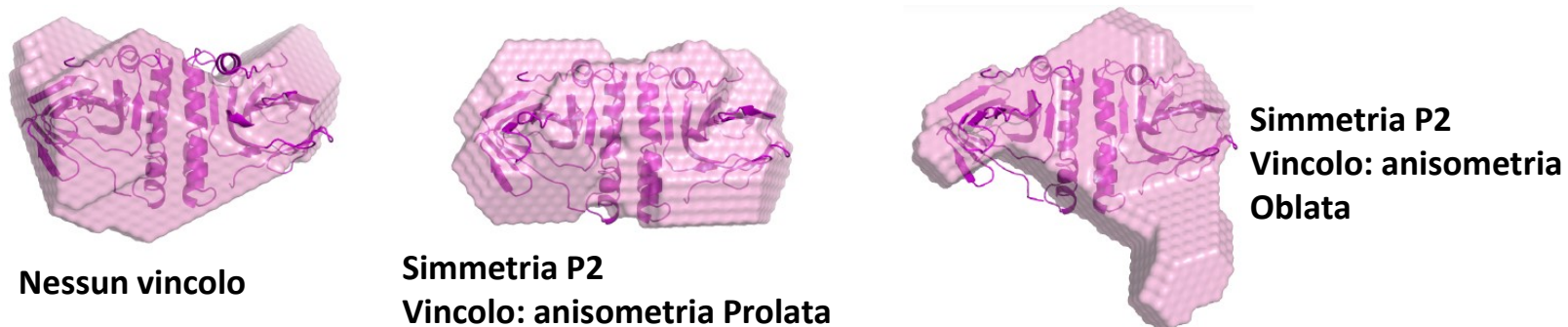
## Validazione del modello ab initio

ab initio modelling basato su dati unidimensionali è intrinsecamente ambiguo:

Ulteriori vincoli limitano lo spazio di ricerca, riducono l'ambiguità e accelerano i calcoli nell' ab initio modelling



CAVEAT usando i vincoli: restrizioni errate possono produrre modelli errati che si adattano ugualmente bene ai dati



# BioSAXS:

## Validazione del modello ab initio

ab initio modelling basato su dati unidimensionali è intrinsecamente ambiguo:

Per controllare la potenziale unicità del modello: ab initio modelling indipendenti per analizzare le variazioni nei modelli risultanti sovrapponendoli e utilizzando la **discrepanza spaziale normalizzata NSD** (Kozin, et al., (2001). J. Appl. Cryst. 34, 33) come misura della dissomiglianza tra i singoli modelli per classificare i modelli in cluster strutturali distinti e ottenerne la media

### Post processing: 1) Superimposition

**SUPCOMB** (Kozin & Svergun (2000) J. Appl. Cryst.34, 33-41)

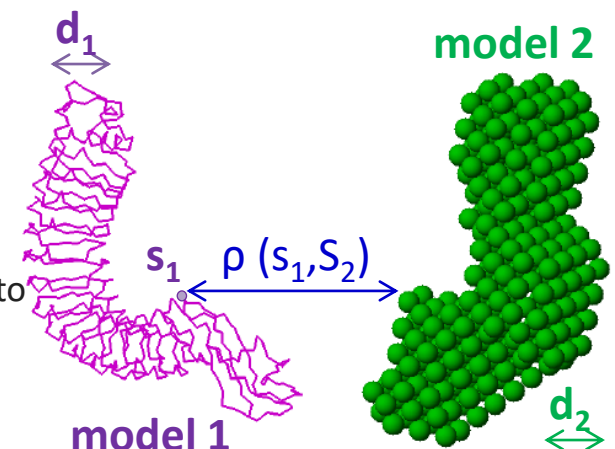
sovrappone strutture 3D (bead models a bassa risoluzione e/o strutture ad alta risoluzione).

Per trovare il miglior allineamento di due modelli, rappresenta ciascuna struttura come un insieme di punti e minimizza una discrepanza spaziale normalizzata

$$NSD(S_1, S_2) = \left[ \frac{1}{2} \left( \frac{1}{N_1 d_2^2} \sum_{i=1}^{M_1} \rho^2(s_{1i}, S_2) + \frac{1}{N_2 d_1^2} \sum_{i=1}^{N_2} \rho^2(s_{2i}, S_1) \right) \right]^{1/2}$$

misura di somiglianza quantitativa tra insiemi di punti tridimensionali

per ogni punto nel primo set (modello 1), trova il valore minimo tra le distanze tra questo punto e tutti i punti nel secondo set (modello 2) e lo stesso viene fatto per i punti nel secondo set. Queste distanze vengono sommate/normalizzate rispetto alle distanze medie tra i punti vicini per i due set.



# BioSAXS:

## Validazione del modello ab initio

### 2) Post processing: Averaging

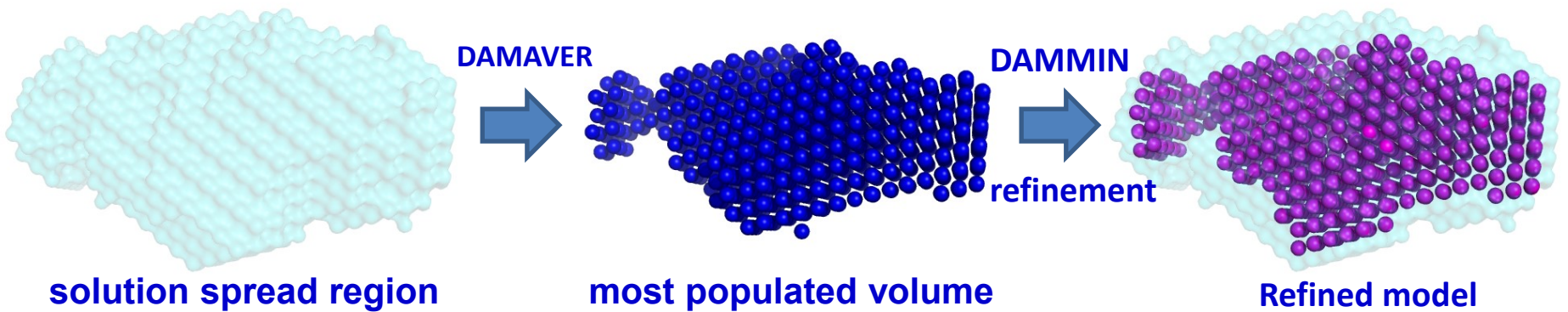
**DAMAVER** (Volkov & Svergun (2003) J. Appl. Cryst. 36, 860-864)

Sovrappone i modelli a coppie

Calcola le somiglianze tra i modelli: NSD < 1 per modelli simili

Media i modelli allineati

Calcola la densità di occupazione spaziale



Rifiutando i nodi con la minima occupazione per trovare il volume più popolato

# BioSAXS:

## Validazione del modello ab initio

### DAMMIF

**OPZIONE: Multiple runs + Run Clustering and Resolution Assessment Algorithms**

**POSTPROCESSING: SUPCOMB**

- **Sovrappone** i modelli a 2 a 2
- **Calcola** la somiglianza fra i modelli: **NSD < 1 per modelli simili**

### DAMAVER

- **Media** i modelli allineati
- Ottiene la **solution spread region** e il **volume più popolato**

### SASRES

- **Valuta la risoluzione** dei modelli *ab initio*

**SASRES (Tuukkanen et al. (2016) IUCrJ. 3, 440)**

**Valuta la risoluzione** dei modelli *ab initio* (sia bead models che and dummy-residue models) dalla **variabilità di un insieme di modelli stimata** attraverso il metodo della **Fourier Shell Correlation (FSC)** (usato in cryo-EM).

## Validazione del modello da rigid body modelling

Analogamente va ripetuto più volte e seguito da un raggruppamento strutturale

→ stima la stabilità delle soluzioni e rivela le caratteristiche comuni del modello

l'affidabilità migliora in modo significativo in combinazione con informazioni aggiuntive da tecniche complementari, in particolare con dati sperimentali e / o computazionali sui contatti tra domini o tra subunità