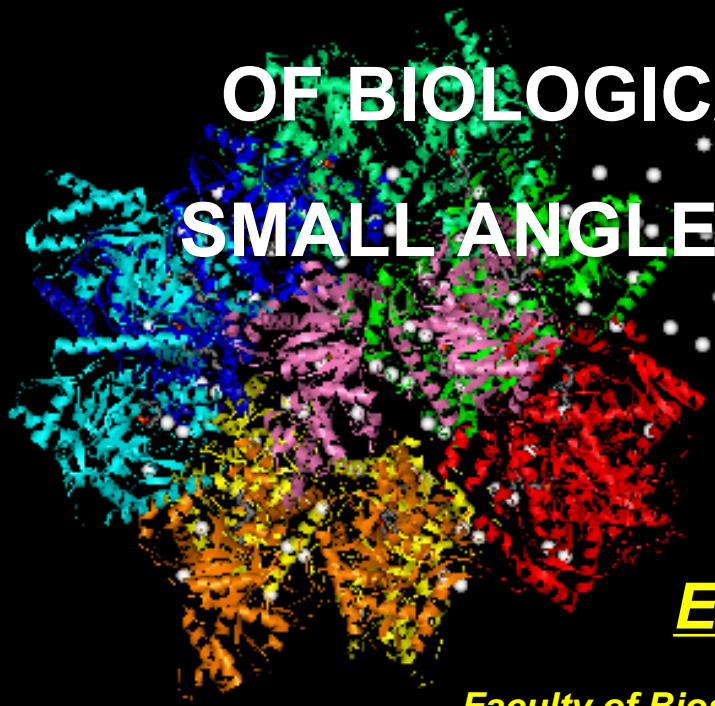
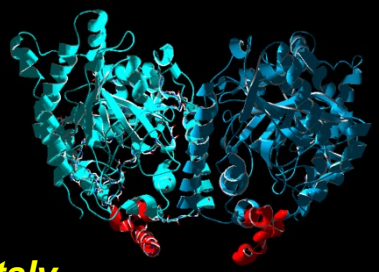
UNIVERSITÀ DEGLI STUDI DI TERAMO

UNITE

STRUCTURAL AND FUNCTIONAL ANALYSIS OF BIOLOGICAL MACROMOLECULES BY SMALL ANGLE X-RAY SCATTERING (SAXS)





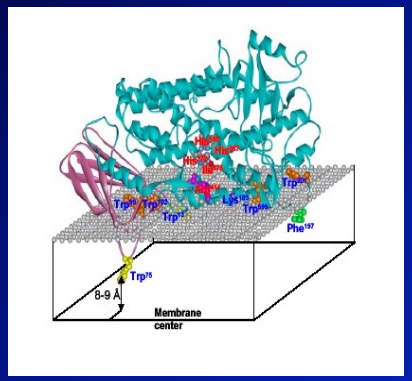
Enrico Dainese

Faculty of Biosciences, University of Teramo, Italy

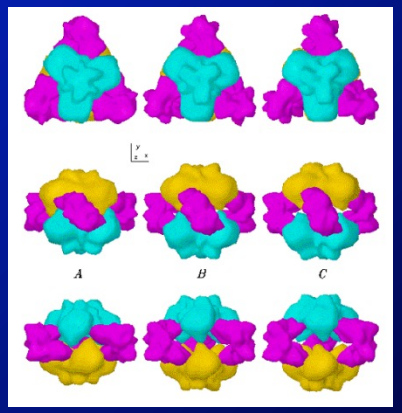
Studies by SAXS: conformational changes of biological macromolecules

Applications to proteins

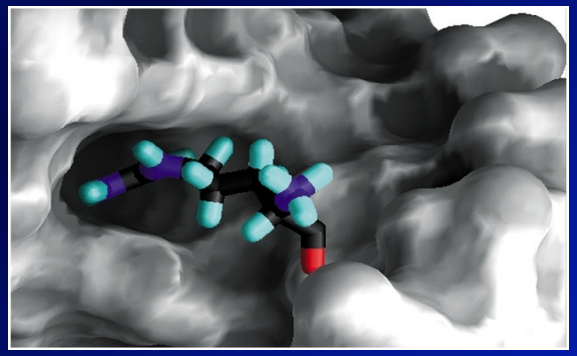
folding-unfolding



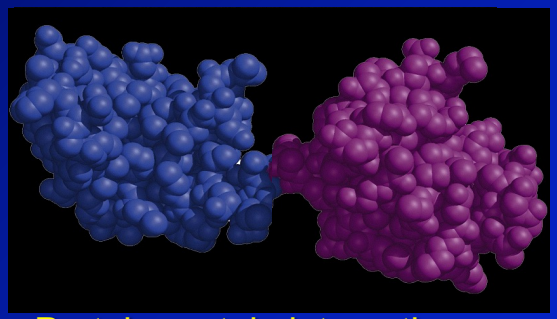
Protein-lipids interactions



Oligomeric organization



Protein-ligand interactions



Protein-protein interactions

Some requirements for useful measurements of biological samples in solution by SAXS (bio-SAXS)

Limits:

- The solutions of biological macromolecules, essentially made out of light atoms in water, display a weak contrast of electron density with the solvent and therefore a weak scattering power;
- The scattering signal of the biological macromolecules may be less than 1% above the scattering of the solvent.

Requirements:

- X-ray source with collimated beam and a flux like the synchrotron radiation from storage rings;
- Beam stability (no beam movements that may produce background changes), low background scattering;
- Detector systems with low-readout noise.

Some requirements for useful measurements of biological samples in solution by SAXS (bio-SAXS)

Since SAXS experiments give information in the reciprocal space, the larger the scattering particle, the smaller the scattering angle. Therefore, the scale of observable particles depends strongly on the quality of the beam (low divergence, high collimation).

Typical protein dimensions: in the range of 10-1000 Å

Need for reliable beamline operating at a fixed wavelength and giving access to a large range of scattering vector:

at least $0.006\ Å^{-1} \leq q \leq 0.6\ Å^{-1}$

Depending on available wavelength, it could take different sample to detector distances.

The experimental set-up should preferably allow for a fast reproducible and automated change of sample-detector distance.

Some requirements for useful measurements of biological samples in solution by SAXS (bio-SAXS)

Limits:
biological samples are highly sensitive to X-ray radiation damage.

Requirements:
- Short measuring times and/or use of attenuators (as in e.g. ID2@ESRF or X33@EMBL-Hamburg) may be required;
-Possibility to slowly circulate the solution through the capillary to always expose fresh sample to the beam;

Sample environment requirements:
- **Low sample volume (20 – 50 μl);**
- **vacuum thermostated capillary;**
- **easy visualization of the capillary;**
- **easily washable (samples can stick on the walls of the sample holder);**

Supplementary facilities
- **stopped flow apparatus for kinetic measurements;**
- **LS apparatus and HPLC on line;**
- **calorimeter (coupling DSC with SAXS);**
- **high-pressure cell;**
- **support hutch (for sample handling).**

What is a SAXS signal?

Data quality

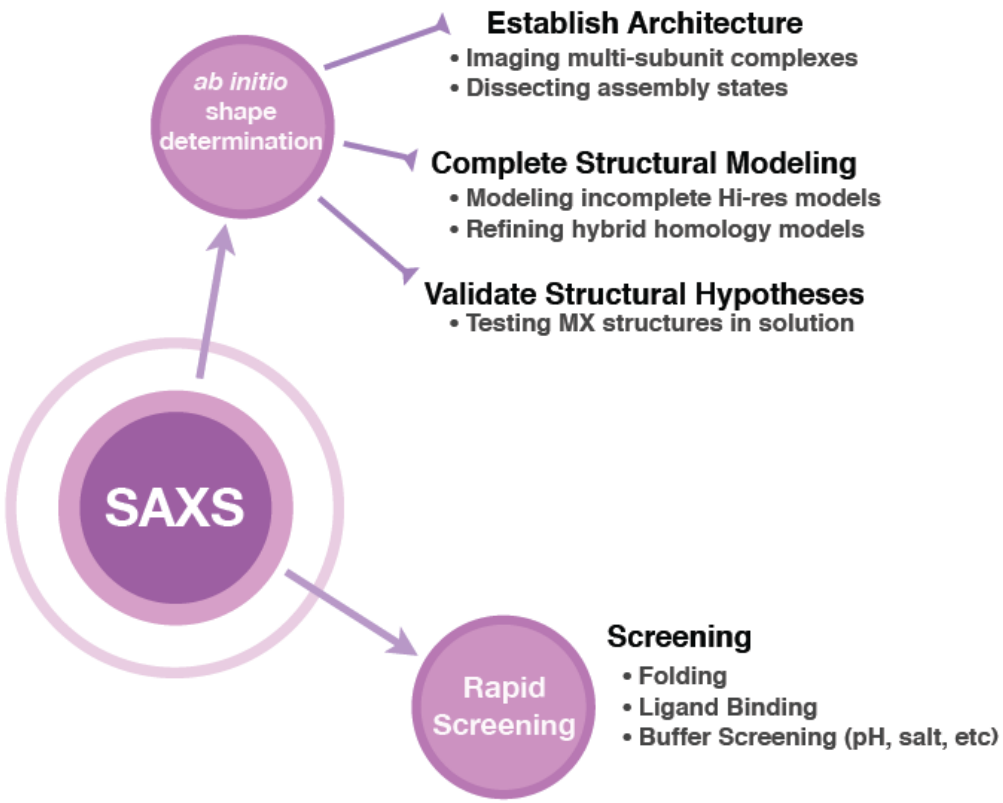
- instrumentation background
- sample cell
- radiation damage

Sample quality

- heterogeneity (aggregation)
- buffer matching
- concentration dependent scattering
  - ★interparticle interference or multimerization

How to tell good from bad?

- instrumentation background (smiling in Guinier residuals)
- sample cell (positive contribution to high-q)
- radiation damage (smiling in Guinier residuals)



Quality sets limits

↓

How useful is the data? (not all data is useable)

↓

Need to assess sample quality!

How?

What do you want from SAXS?

CAPITOLO 24

Metodica SAXS

di Enrico Dainese

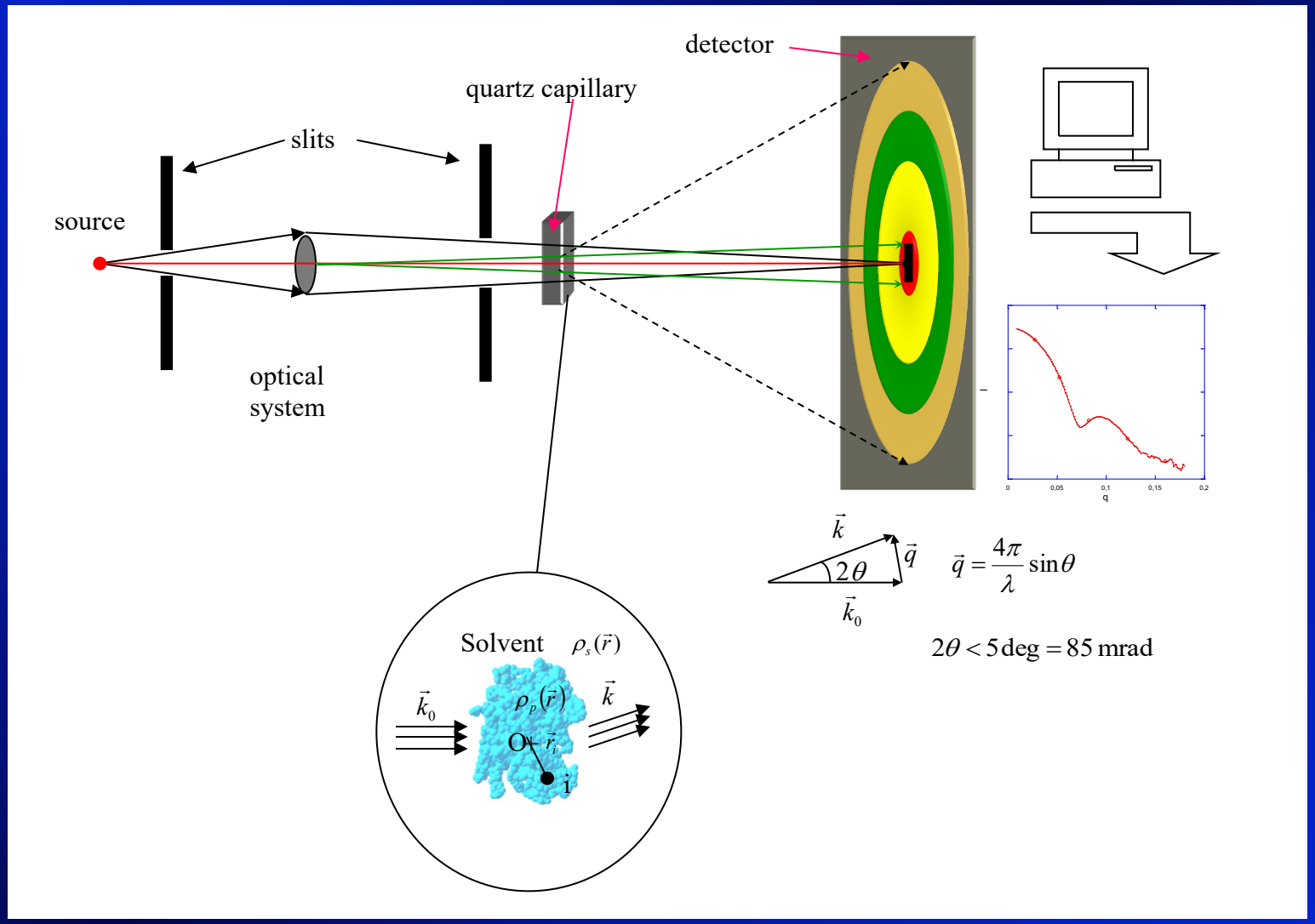
24.1 Scopo della metodica

La **diffusione di raggi-X a piccolo angolo** (*Small Angle X-ray Scattering*, **SAXS**) è una metodica sperimentale che rappresenta uno strumento molto potente nell'ambito degli studi strutturali e funzionali di **macromolecole biologiche** in soluzione. Questi studi stanno attualmente subendo un'enorme espansione, associata, da un lato, all'incremento delle capacità di calcolo e miglioramento dei programmi per l'analisi dei dati e, dall'altro, al miglioramento della qualità dei dati ottenuti. Ciò anche grazie all'avanzamento della tecnologia dei rivelatori (*detector*), nonché agli sviluppi di hardware e software presso le grandi installazioni sperimentali (*large facilities*) di **luce di sincrotrone**.

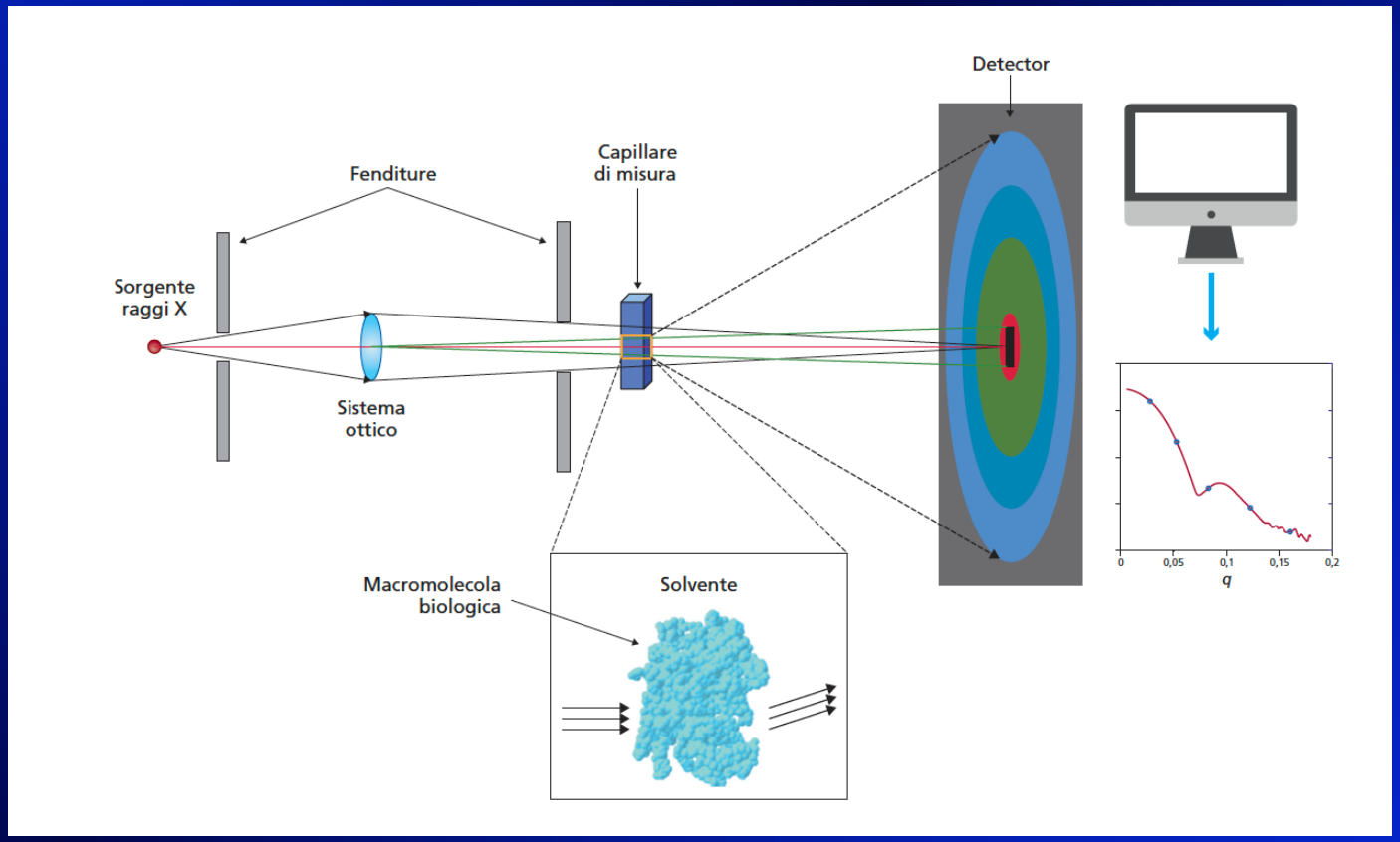
La **struttura tridimensionale** (**3D**) delle macromolecole biologiche in soluzione dipende dalle loro interazioni con il solvente acquoso e da interazioni interne alla macromolecola stessa (idrofobiche, elettrostatiche, legami a idrogeno, interazioni di van der Waals ecc.). La dinamica molecolare che controlla la funzione delle macromolecole biologiche è influenzata proprio dal sottile equilibrio tra le interazioni esterne dei gruppi di superficie con il solvente e le forze intramolecolari. Si ritiene che l'**effetto idrofobico** sia uno dei principali fattori chimico-fisici che regolano le interazioni in soluzione acquosa. Inoltre, la conformazione di una macromolecola associata alla funzione biologica della stessa dipende molto da come questa risponde all'interazione con altri partner (lipidi di membrana, ligandi, altre macromolecole ecc.), che inducono specifici cambiamenti conformazionali.

Metodologie biochimiche e biomolecolari. Strumenti e tecniche per il laboratorio del nuovo millennio.
Zanichelli Editore 2019 .

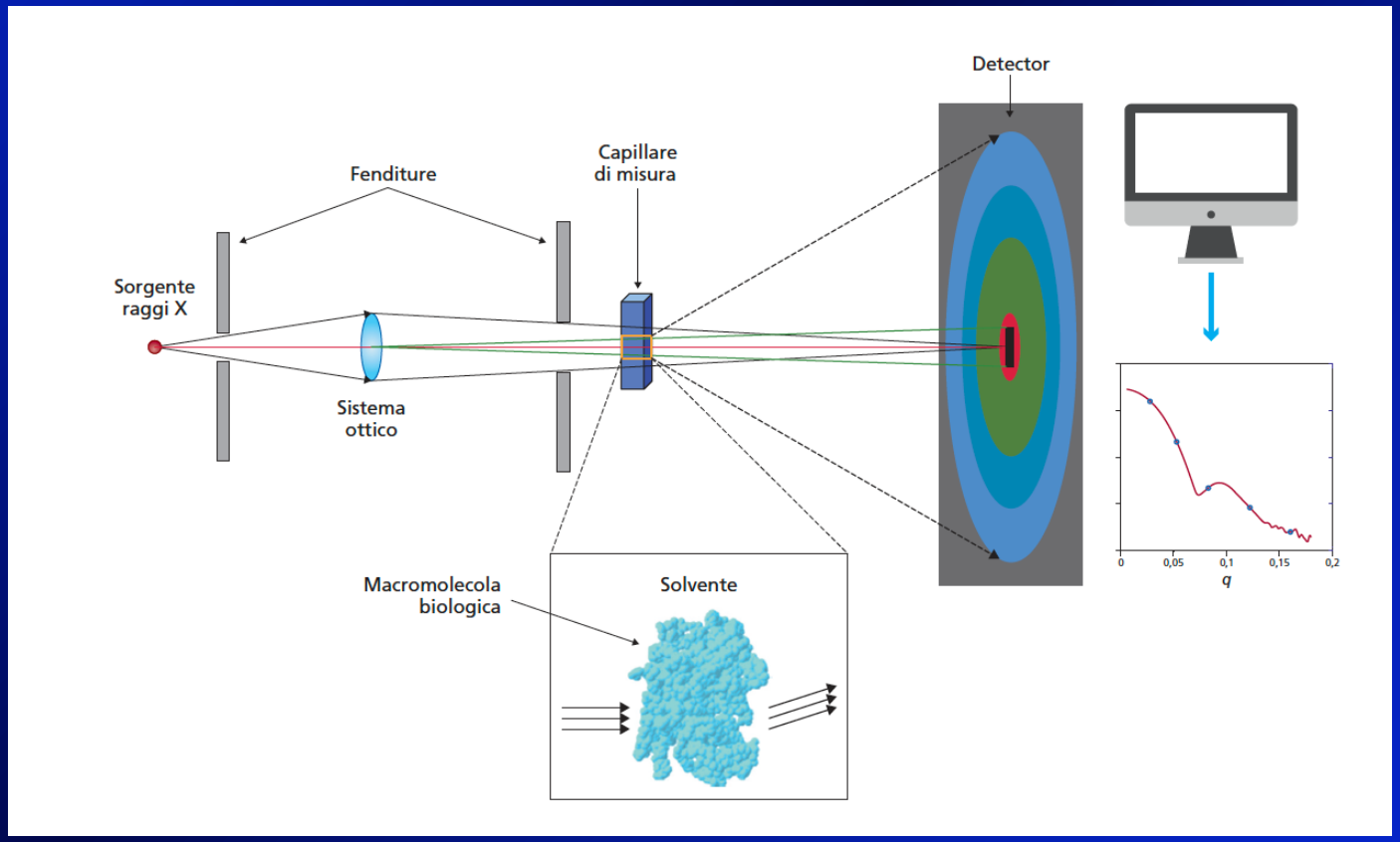
SMALL ANGLE X-RAY SCATTERING



Schema generale di un esperimento SAXS. Il campione biologico viene irradiato all'interno di un capillare di quarzo a temperatura controllata. L'intensità di diffusione dei raggi X a basso angolo viene registrata da un rilevatore bidimensionale, mentre un beam-stop evita che il fascio colpisca direttamente il detector. I dati vengono registrati come I(q) in funzione di q.

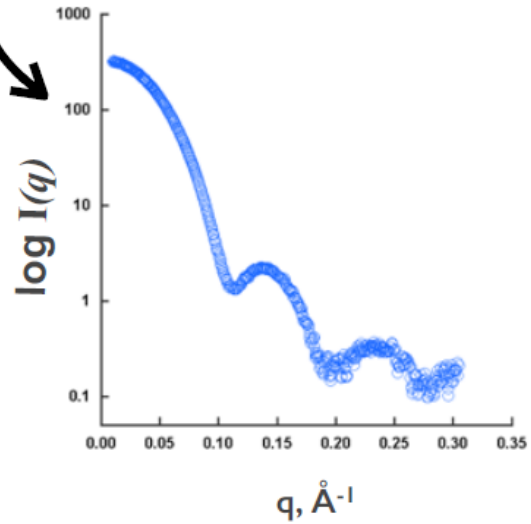


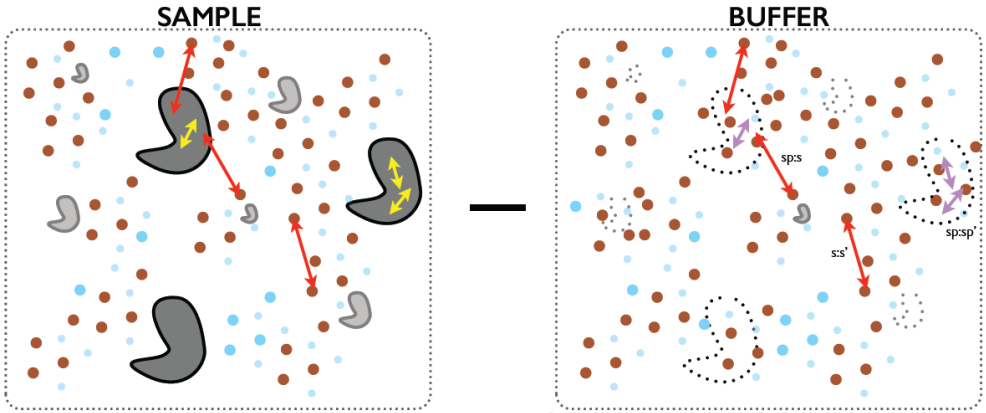
Lo studio della diffusione dei raggi X a piccoli angoli fornisce informazioni sulle dimensioni e sulla conformazione dovuta alle eterogeneità elettroniche delle particelle in esame (macromolecole biologiche) rispetto a un ambiente omogeneo (solvente), nonché sulla loro distribuzione dimensionale e sulle loro mutue interazioni. Per esempio, nel caso di soluzioni di proteine, la particella è la proteina stessa con una media ponderata di densità elettronica p (r), circondata dal solvente (soluzione tampone).



Basics of a SAXS curve

- I(q) measured intensities





$$I_{particle}(q) = I_{sample}(q) - I_{buffer}(q)$$

low resolution ⟷ high resolution

- Features throughout the curve relate to shape
- At low resolution, can approximate particle as a homogenous body of electron density
- larger the object, the faster the I(q) decay

Analisi dei dati SAXS

Nella tecnica SAXS vengono considerati solo i fenomeni di **diffusione** (***scattering***) **elastica**, in cui non vi è scambio di energia tra i raggi X incidenti e gli elettroni degli atomi che compongono la macromolecola biologica. Solo gli elettroni diffonderanno i raggi X a causa della loro massa molto piccola. In pratica, poiché il numero di atomi che compongono una macromolecola è molto alto e la risoluzione della tecnica non è a livello atomico, la somma delle lunghezze di diffusione a bassi angoli della macromolecola viene descritta utilizzando una distribuzione continua di densità elettronica.

La radiazione diffusa viene misurata da un rilevatore posto a una specifica distanza dal campione biologico che genera il fascio di raggi trasmessi. Nella Figura 24.1 il **vettore di diffusione** $\vec{q}$ è definito come $\vec{k}_1 - \vec{k}_0$, dove k_1 è il vettore caratteristico dell'onda diffusa e $\vec{k}_0$ è il vettore caratteristico del fascio incidente ($\vec{k}_0 = 1/\lambda$, dove λ è la lunghezza d'onda dei raggi X). Il modulo del vettore di diffusione è:

$$|\vec{q}| \cong \frac{4\pi}{\lambda}\sin\theta \qquad [24.1]$$

dove
2θ = angolo di diffusione;
λ = lunghezza d'onda dei raggi X.

Tale approssimazione vale solo per piccoli valori di angoli di diffusione, cioè per $2\theta < 5$ gradi.

La Regione di Guinier

Nella tecnica SAXS vengono considerati solo i fenomeni di **diffusione** (***scattering***) **elastica**, in cui non vi è scambio di energia tra i raggi X incidenti e gli elettroni degli atomi che compongono la macromolecola biologica. Solo gli elettroni diffonderanno i raggi X a causa della loro massa molto piccola. In pratica, poiché il numero di atomi che compongono una macromolecola è molto alto e la risoluzione della tecnica non è a livello atomico, la somma delle lunghezze di diffusione a bassi angoli della macromolecola viene descritta utilizzando una distribuzione continua di densità elettronica.

La radiazione diffusa viene misurata da un rilevatore posto a una specifica distanza dal campione biologico che genera il fascio di raggi trasmessi. Nella Figura 24.1 il **vettore di diffusione** $\vec{q}$ è definito come $\vec{k}_1 - \vec{k}_0$, dove k_1 è il vettore caratteristico dell'onda diffusa e $\vec{k}_0$ è il vettore caratteristico del fascio incidente ($\vec{k}_0 = 1/\lambda$, dove λ è la lunghezza d'onda dei raggi X). Il modulo del vettore di diffusione è:

$$|\vec{q}| \cong \frac{4\pi}{\lambda}\sin\theta \qquad [24.1]$$

dove
2θ = angolo di diffusione;
λ = lunghezza d'onda dei raggi X.

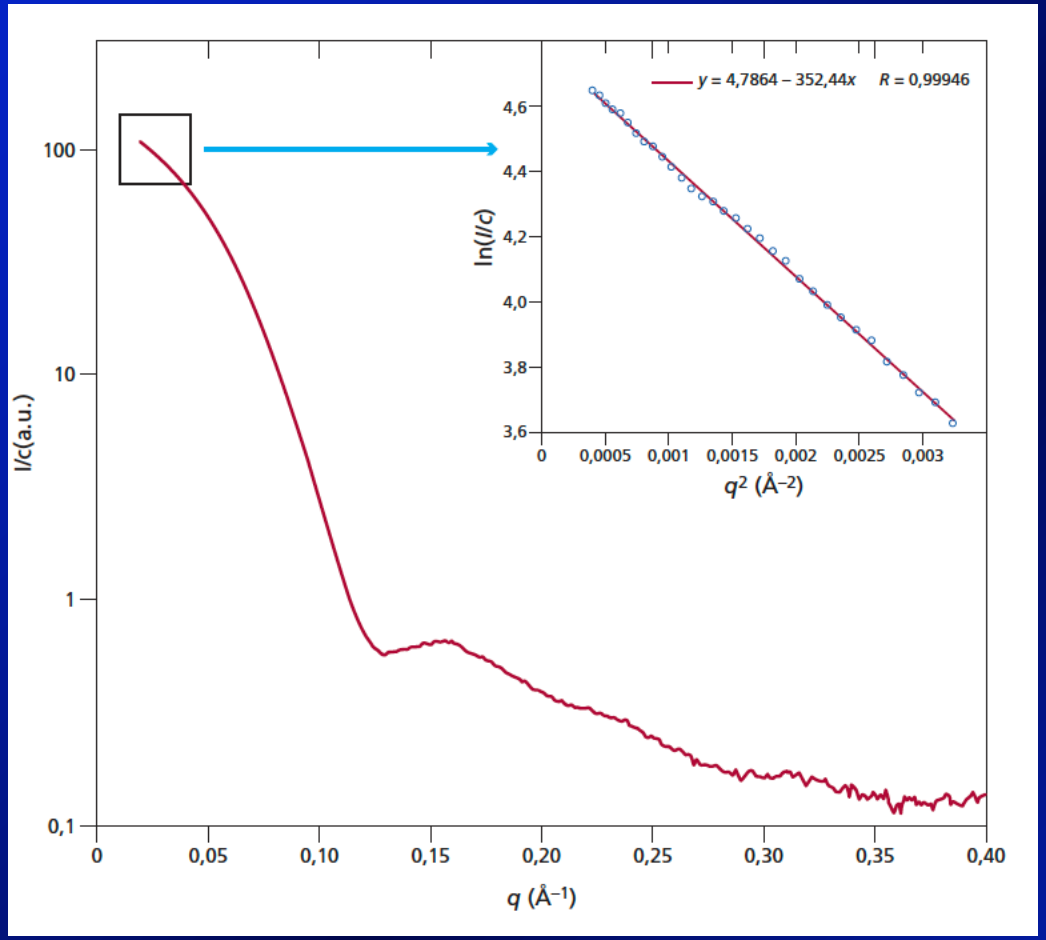
Tale approssimazione vale solo per piccoli valori di angoli di diffusione, cioè per $2\theta < 5$ gradi.

I dati iniziali di una misura SAXS ottenuti dopo opportuna sottrazione del solvente sono espressi come intensità di diffusione $I(q)$ in funzione di q (**Figura 24.2**). L'analisi dei dati consente la determinazione di alcuni parametri strutturali rilevanti. Per piccoli valori di q, il **raggio di girazione** (R_g) della particella in soluzione può essere determinato interpolando i dati SAXS sulla base dell'**approssimazione di Guinier**:

$$I(q) \cong n \cdot (\rho_c V)^2 \exp\left(-\frac{R_g^2 q^2}{3}\right) \qquad [24.2]$$

dove
n = numero di particelle (molecole biologiche in soluzione, per esempio molecole di proteina) che diffondono i raggi X,
ρ_c = contrasto c della densità elettronica tra la proteina e il solvente $r_p - r_s$ (Figura 24.1, pannello inferiore),
V = volume della proteina,
$n \cdot (\rho_c V)^2 = I(0)$ è l'intensità di diffusione ad angolo zero.

La Regione di Guinier

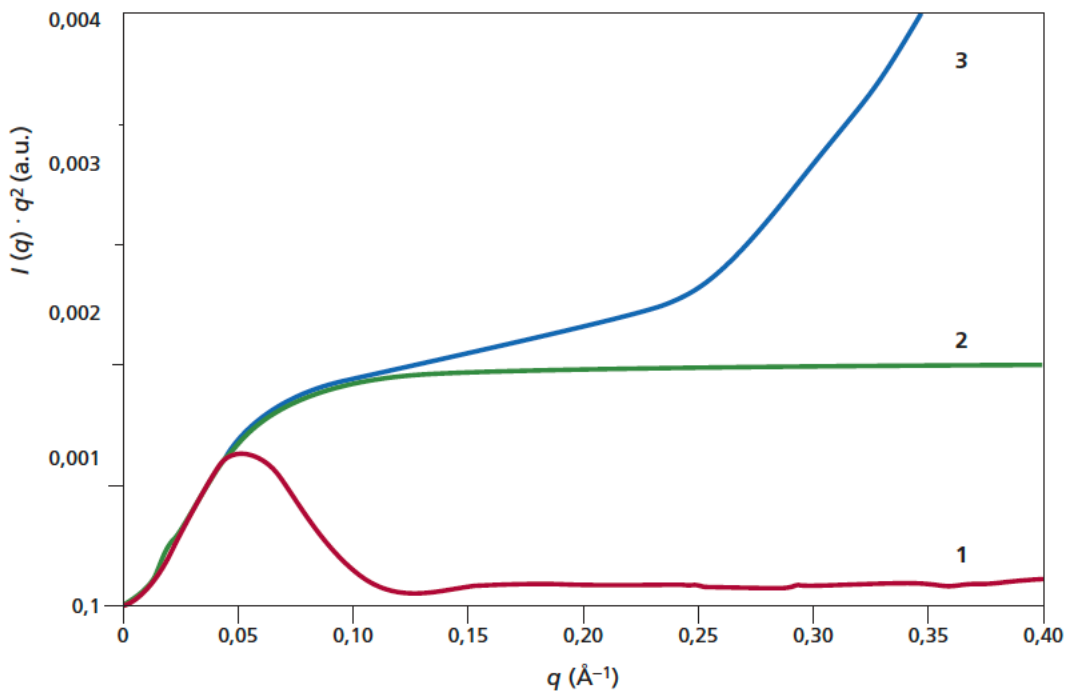


Nell'analisi dei dati viene utilizzata una rappresentazione linearizzata, in cui è riportato in grafico il $\ln I(q)$ in funzione di q^2 (**diagramma di Guinier**). Una **regressione lineare** di tale grafico (**Figura 24.2**, inserto) a bassi valori di q consente di ottenere il raggio di girazione dalla pendenza della retta, unitamente all'intensità di scattering ad angolo zero $I(0)$. Va tenuto presente che l'approssimazione di Guinier si basa sull'assunzione di una soluzione ideale e monodispersa. Se, invece, sono presenti diverse molecole biologiche con differenti dimensioni, la loro diffusione si somma e, nella maggior parte dei casi, il valore di R_g non può essere determinato in modo univoco. Inoltre, la determinazione del raggio di girazione sarà errata se le interazioni tra le molecole biologiche in soluzione non sono trascurabili. Per essere nelle condizioni di minimizzare tali interazioni, le misure verranno eseguite a diverse concentrazioni. Quando i valori di raggio di girazione e di intensità ad angolo zero $I(0)/c$ non variano con la concentrazione, è possibile utilizzare correttamente questi dati per calcolare la regione di Guinier ($R_g \cdot q \leq 1{,}2$) senza errori dovuti a eventuali interazioni attrattive o repulsive tra le molecole biologiche oggetto di analisi.

Kratky Plot

Un'altra rappresentazione utile dei dati SAXS è il **grafico di Kratky** (***Kratky plot***), in cui si riportano i valori di $I(q) \cdot q^2$ in funzione di q (**Figura 24.3**). Tale grafico è molto sensibile alla compattezza della macromolecola biologica in studio, per esempio nel caso del ***folding/unfolding*** di proteine. Infatti, il profilo del *Kratky plot* per proteine globulari mostra la presenza di un chiaro picco, la cui posizione dipende del raggio di girazione R_g (Figura 24.3, curva 1); invece, nel caso di una conformazione con struttura a catena globulare spiralata (per esempio, lo stato a globulo fuso delle proteine), la curva raggiunge un plateau (Figura 24.3, curva 2). Infine, per una catena polimerica disordinata (*random coil*), come può accadere per una proteina in condizioni fortemente denaturanti, la curva prima cresce, poi raggiunge un plateau e, a valori elevati di q, aumenta in modo lineare (Figura 24.3, curva 3). Questa marcata differenza nel comportamento asintotico fornisce un modo facile e sensibile per monitorare il grado di folding e la compattezza di una macromolecola biologica al variare di un dato parametro (temperatura, pressione, pH, agenti denaturanti ecc.). In questo modo, usando i diagrammi di Kratky, si può distinguere facilmente tra conformazioni denaturate e native delle proteine, ma si può anche evidenziare l'eventuale presenza di proteine con strutture native intrinsecamente disordinate.

Figura 24.3
Esempio di *Kratky plot* per una macromolecola biologica in tre diverse conformazioni: 1) globulare; 2) con una conformazione estesa simile a un globulo fuso; 3) completamente disordinata (si veda il testo per i dettagli).

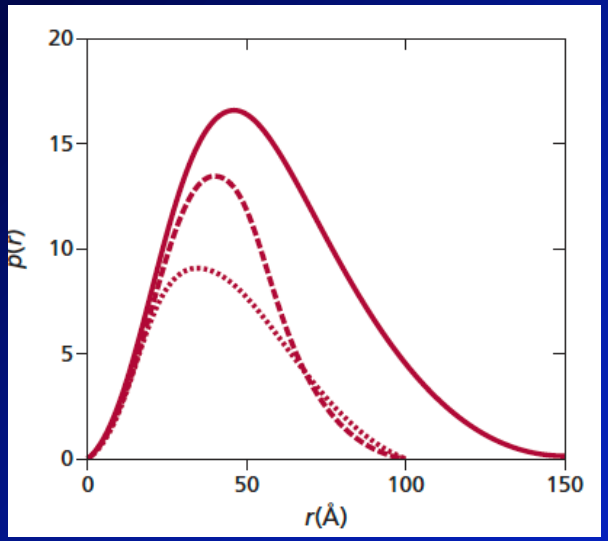


P(r) function

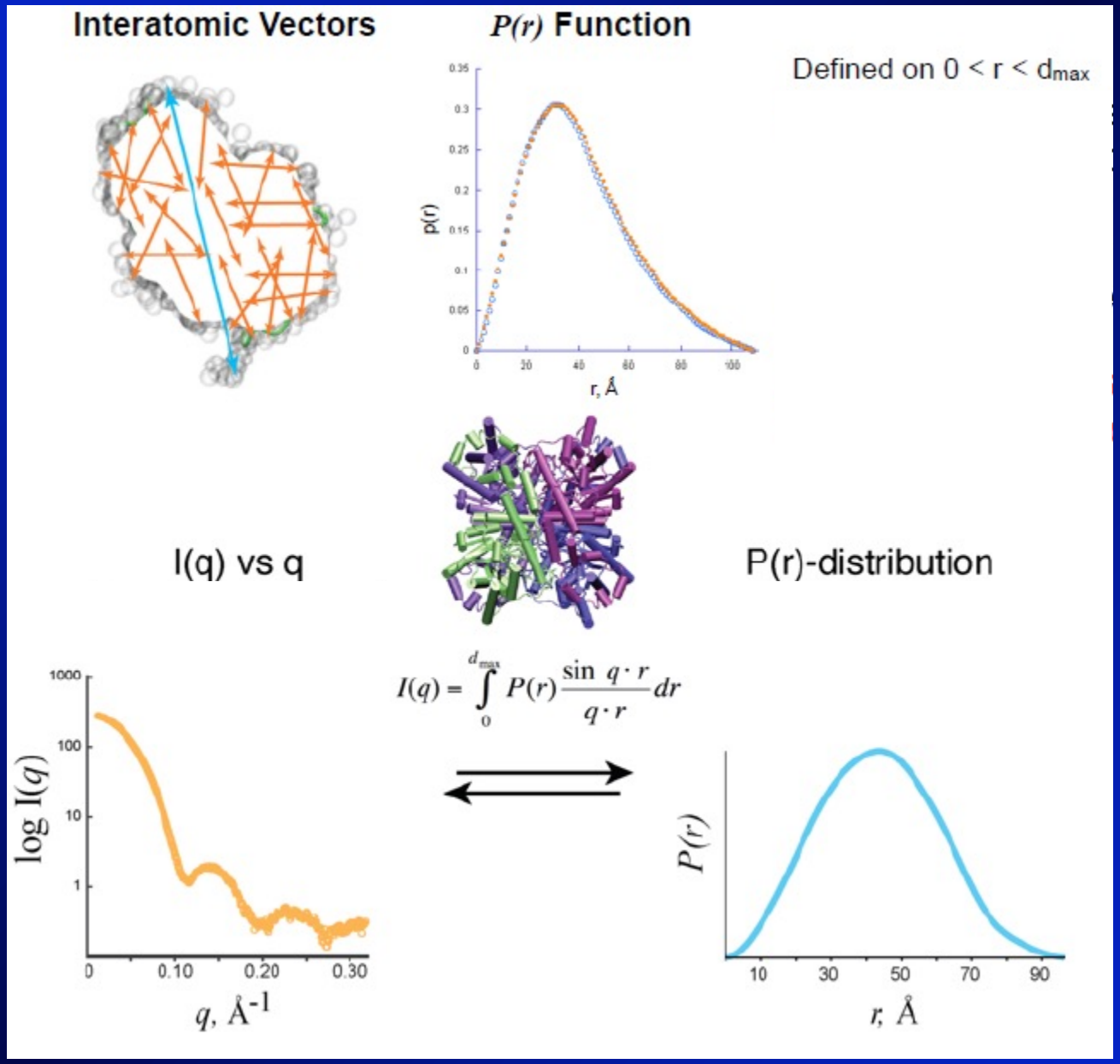
Un altro strumento molto utile nell'analisi dei dati SAXS è lo studio della **funzione *p*(*r*)**, proveniente dalla trasformata di Fourier indiretta della curva sperimentale. Per una particella omogenea, come la maggior parte delle macromolecole biologiche purificate oggetto di analisi SAXS, questa funzione rappresenta l'istogramma delle distanze tra coppie di punti all'interno della particella stessa. Il valore della funzione $p(r)$ è pari a zero oltre il valore di D_{max}, ovvero la dimensione massima della macromolecola biologica in esame. Quindi, a differenza dell'intensità di *scattering* in funzione di q, la funzione $p(r)$ è descritta in una regione limitata dello spazio e può essere definita come:

$$I(q) = 4\pi \int_0^{D_{max}} p(R) \frac{\sin(qR)}{qR} dR \qquad (24.3)$$

L'analisi della funzione $p(r)$ può essere eseguita con il **programma GNOM** e fornisce i valori di R_g e D_{max} (ma anche di intensità ad angolo zero), che sono meno influenzati dagli effetti di interazione residua o dalla presenza di piccoli aggregati rispetto a quelli ottenuti dai calcoli di Guinier. Infine, il programma GNOM fornisce anche la trasformata inversa della funzione $p(r)$, che può essere impiegata nelle successive analisi per lo sviluppo di **modelli strutturali *ab initio*** a partire dai dati SAXS. Tale analisi risulta particolarmente utile nel caso di validazione della formazione di complessi funzionali tra proteine diverse, per esempio nell'interazione tra ceruloplasmina e lattoferrina esemplificata nella **Figura 24.4a**.

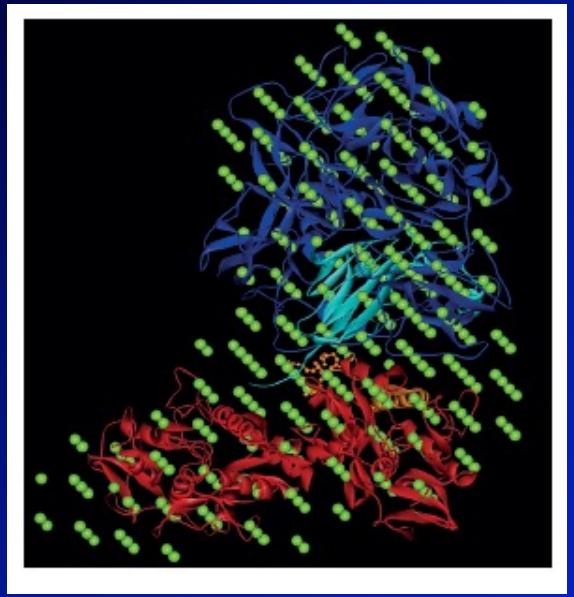


P(r) function


Interatomic Vectors
P(r) Function
Defined on 0 < r < dmax
p(r)
r, Å
I(q) vs q
P(r)-distribution
$I(q) = \int_0^{d_{max}} P(r) \frac{\sin q \cdot r}{q \cdot r} dr$
log I(q)
q, Å-1
P(r)
r, Å

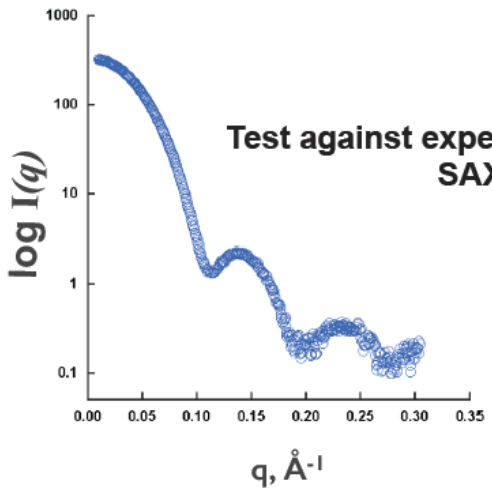
DAMMY ATOMS MODELING

Uno dei metodi maggiormente diffusi per ottenere modelli *ab initio* dalle curve SAXS consiste nella modellazione mediante impiego di piccole sfere (*bead modeling*). Questo approccio è stato impiegato da Svergun e collaboratori nel programma di analisi SAXS noto come DAMMIN. Tale software riempie una sfera di diametro D_{max} con piccolissime sfere (atomi fittizi) di raggio $r_0 << D_{max}$. Il modello strutturale che si ottiene (*Dummy Atom Model*, DAM) è definito da un vettore di configurazione X che assegna un indice (0) a ciascuna sfera corrispondente al solvente e (1) quando la sfera si trova all'interno della macromolecola biologica oggetto di studio. In linea con la bassa risoluzione dei dati di diffusione di una soluzione di biomolecole, il metodo parte da una configurazione casuale e cerca una configurazione X minimizzando $f(X) = \chi^2 + \alpha P(X)$, dove α è un parametro positivo e il termine di penalità $P(X)$ assicura che il modello strutturale ottenuto (***DAM model***) abbia una risoluzione bassa rispetto al raggio r_0 degli atomi fittizi. La minimizzazione viene eseguita utilizzando un metodo di *simulated annealing*, tramite il quale si possono ottenere diversi modelli partendo dall'analisi degli stessi dati SAXS sperimentali. Il confronto tra questi consente di verificare la stabilità della soluzione e la bontà del modello strutturale ottenuto. La modellazione *ab initio* può essere eseguita principalmente utilizzando dati SAXS di campioni **monodispersi**, poiché il programma cercherà un modello per descrivere la curva di diffusione, che, quindi, deve anche derivare da una soluzione contenente una singola specie di macromolecola biologica che causa diffusione di raggi X (particella di *scattering*).

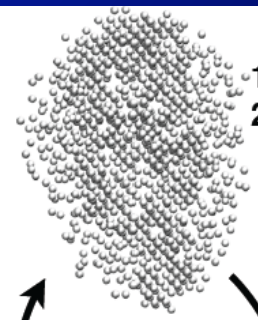


DAMMY ATOMS MODELING

ab initio modeling



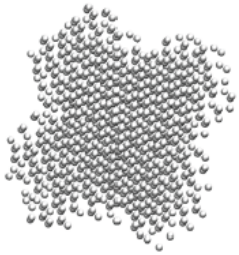
Test against experimental SAXS curve



1. Define Search Volume
2. Fill with beads (dummy atoms)
- **proportional to mass**

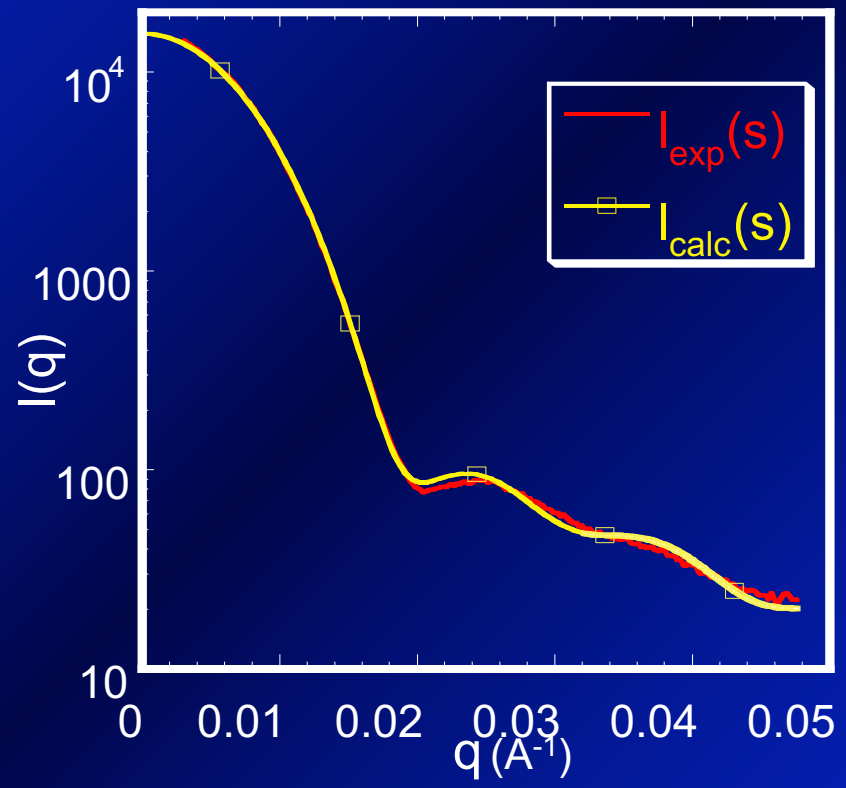
Sample different bead arrangements
- Simulated Annealing
- Genetic Algorithm

Repeat

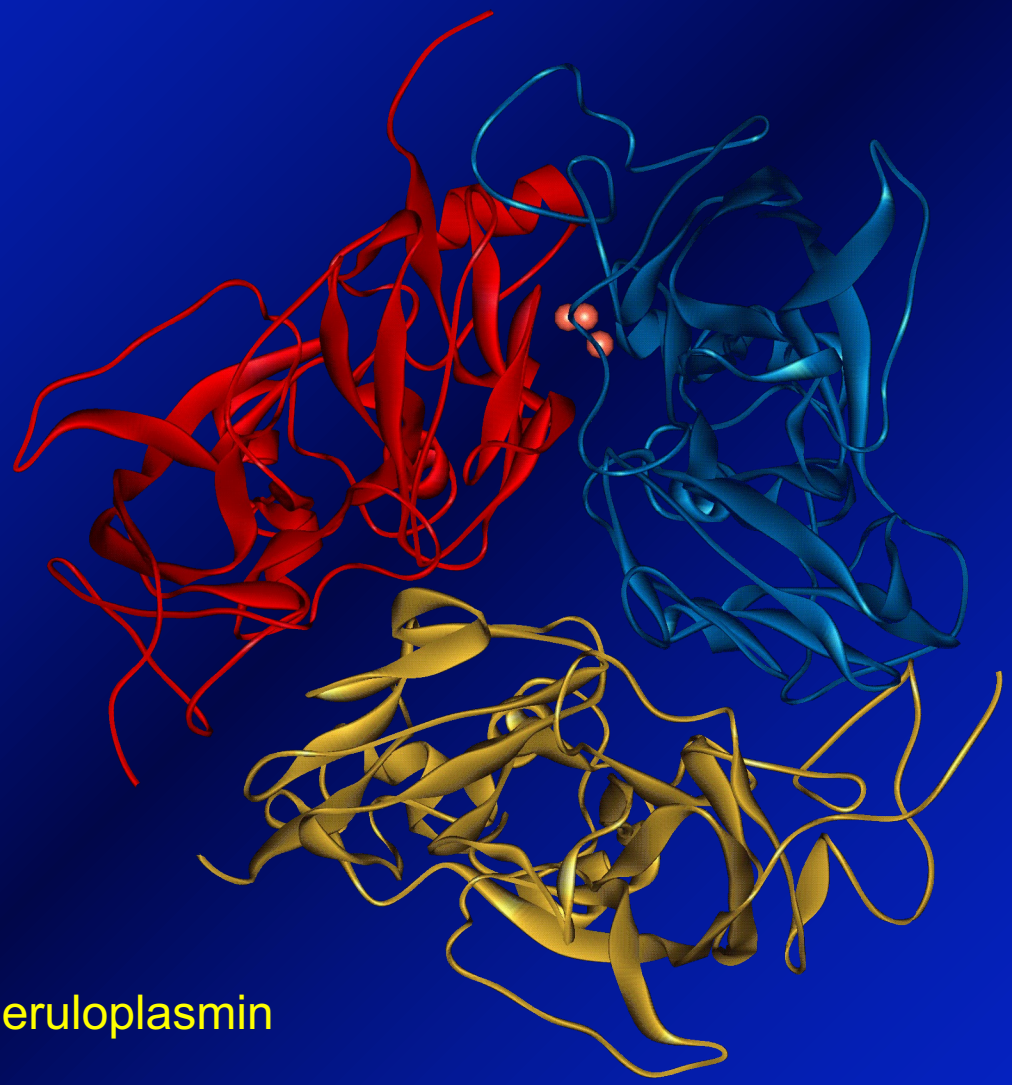


Modeling depends greatly on
1. high quality SAXS data
2. accurate mass constraints

La sovrapposizione con i dati cristallografici CRYSOL

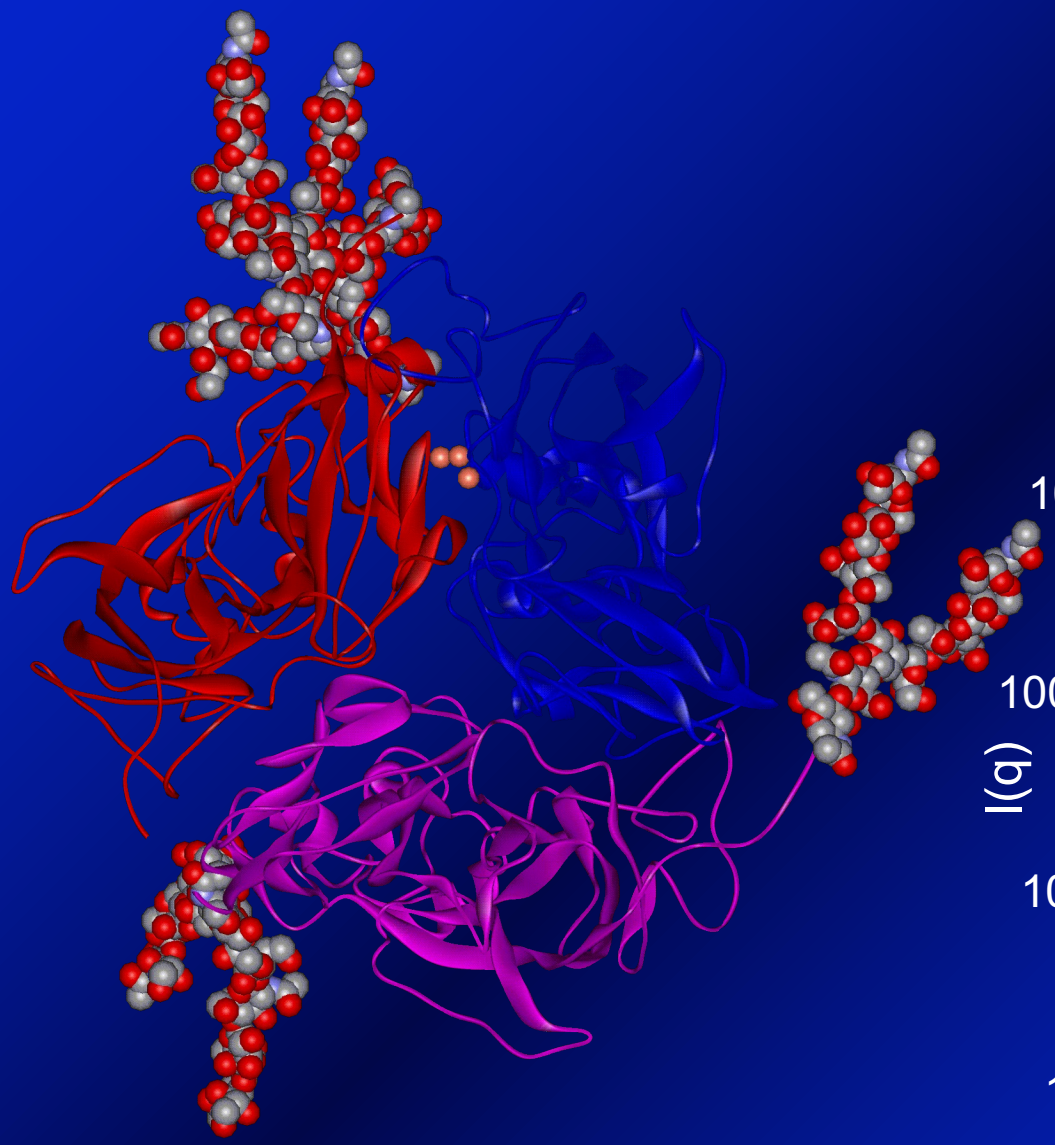


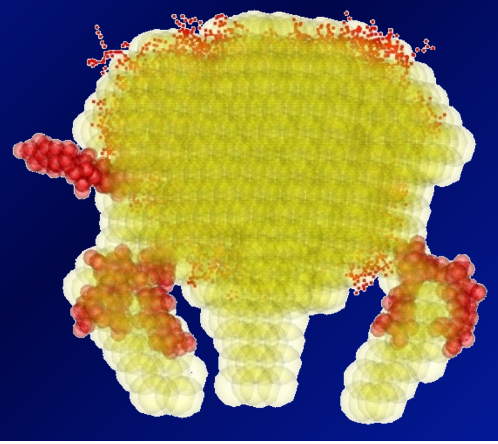
Structural and functional role of the copper ligand in human ceruloplasmin

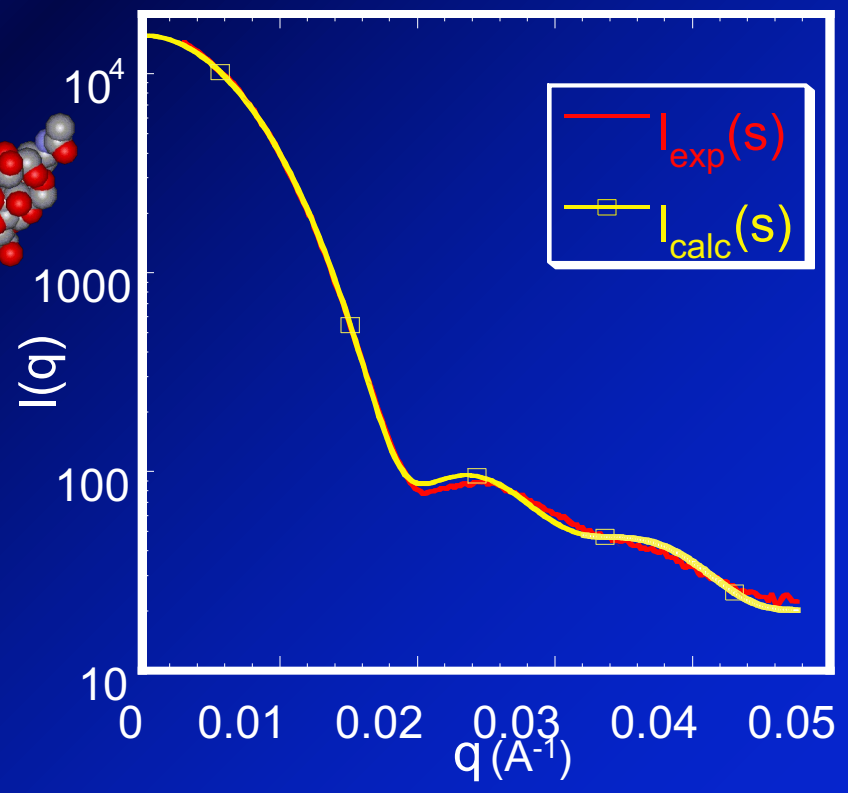


Human Ceruloplasmin

Crystallographic data versus SAXS data

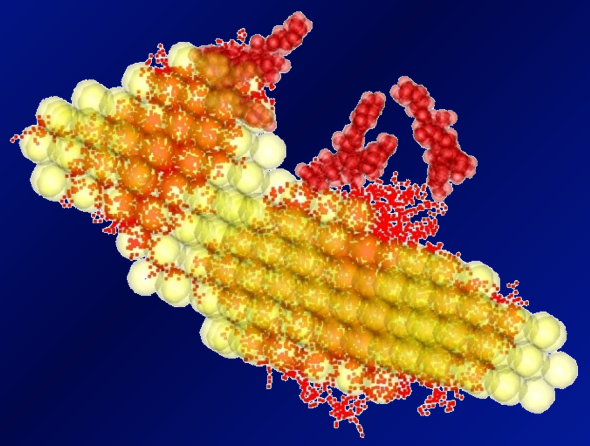


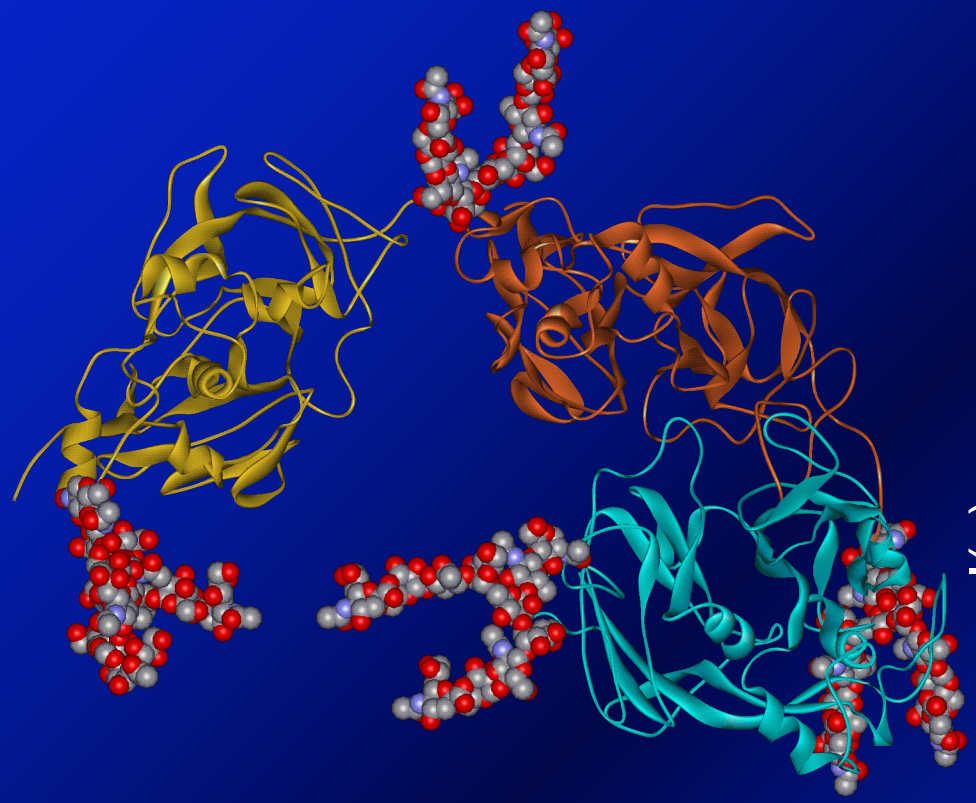


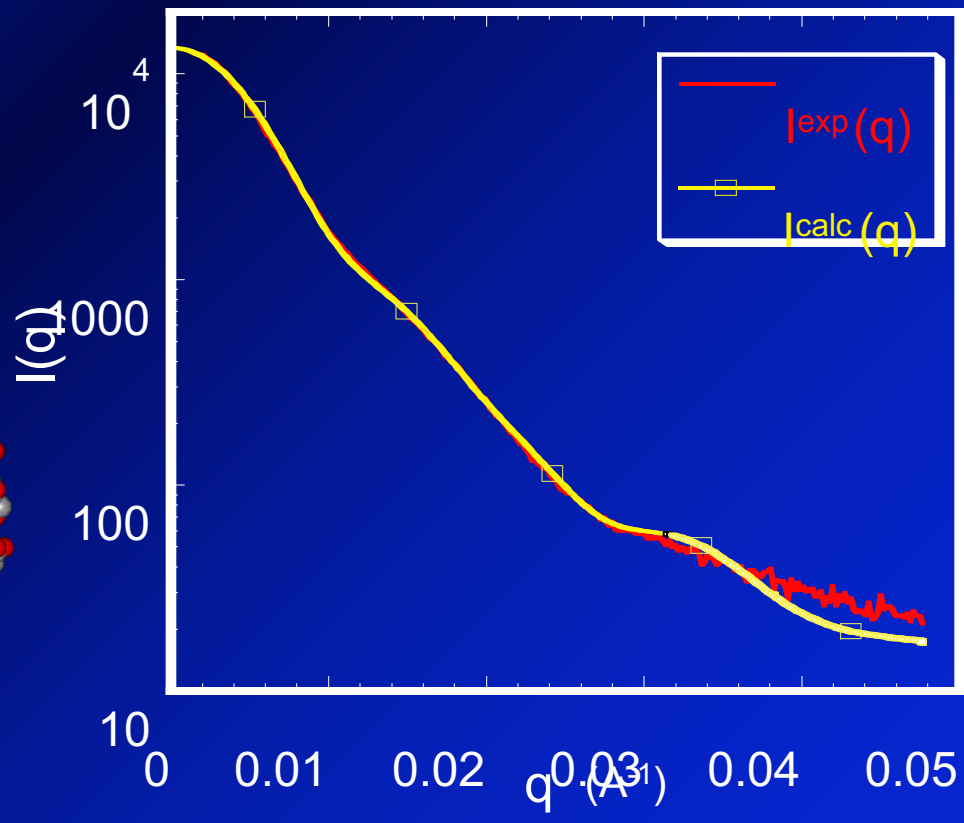


Vachette & Dainese et al., (2002) JBC. 277, 40823.

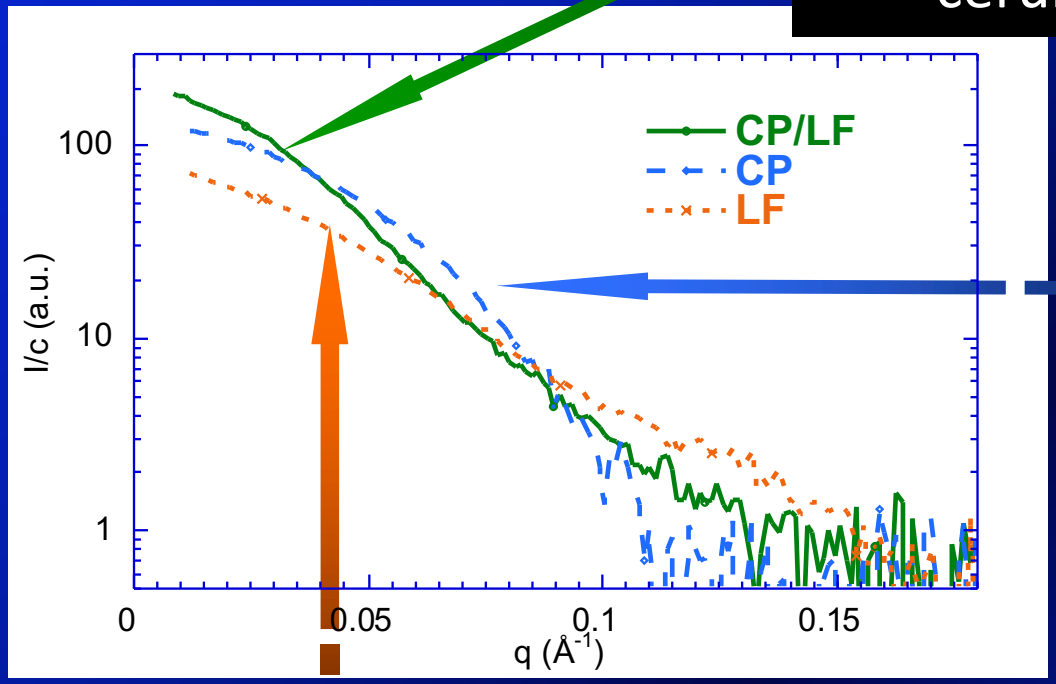
Structure of the apo-ceruloplasmin by SAXS

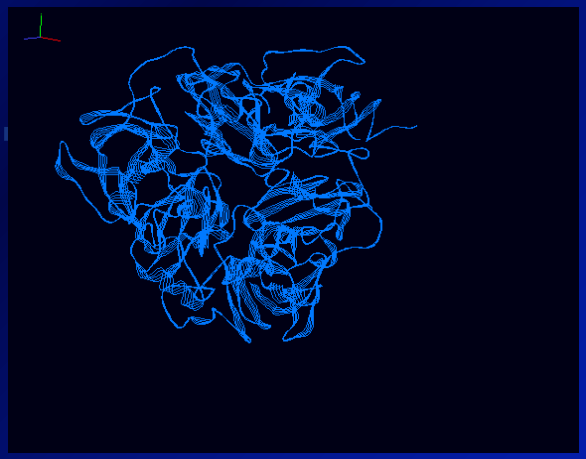


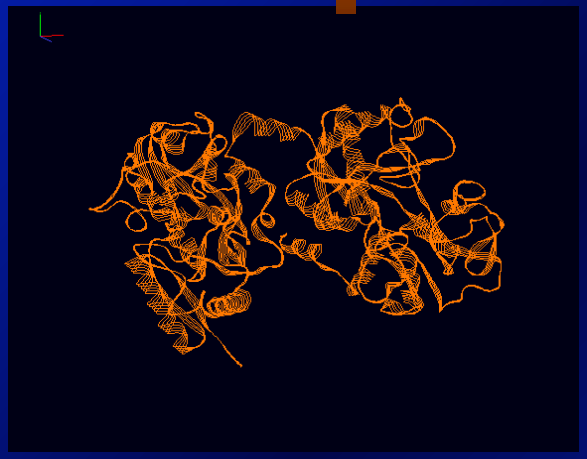




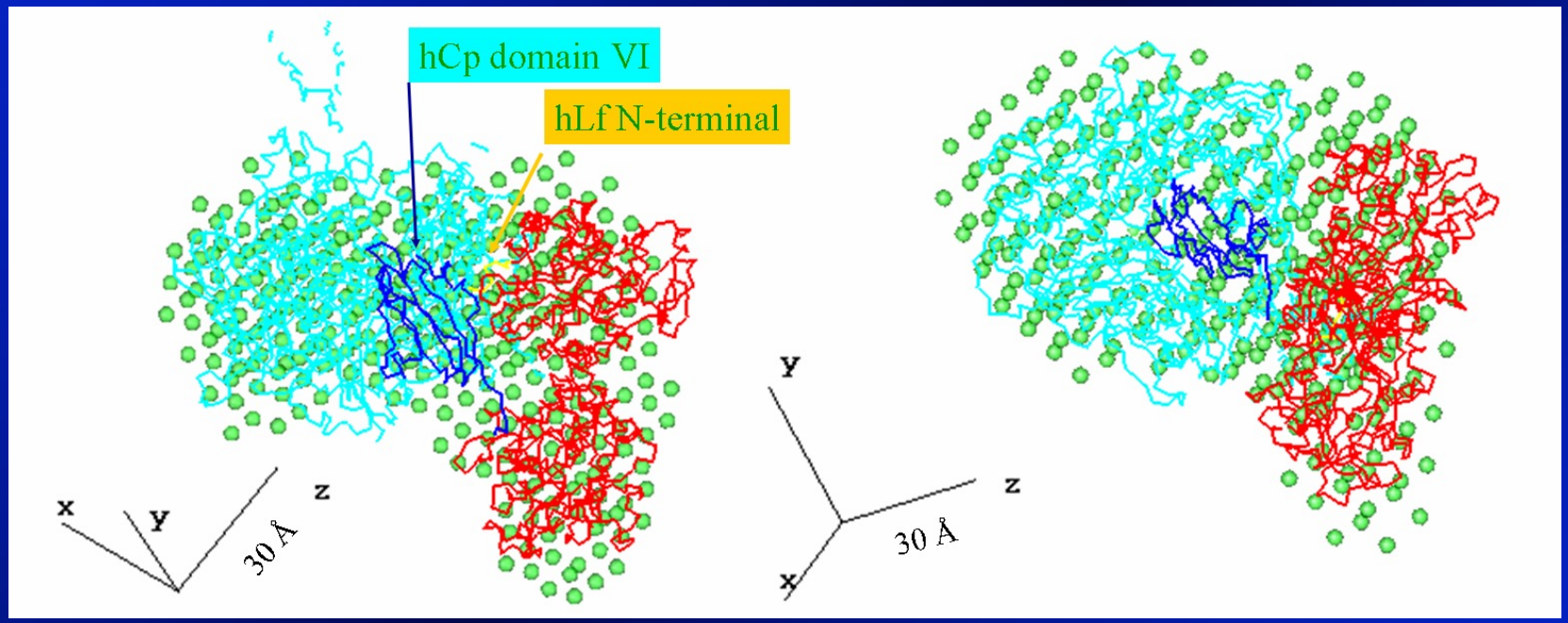
Interaction between human ceruloplasmin and lactoferrin







Ceruloplasmin has the correct orientation to facilitate the iron incorporation into lactoferrin



doi:10.1016/j.jmb.2007.05.089 J. Mol. Biol. (2007) 371, 1038–1046

JMB

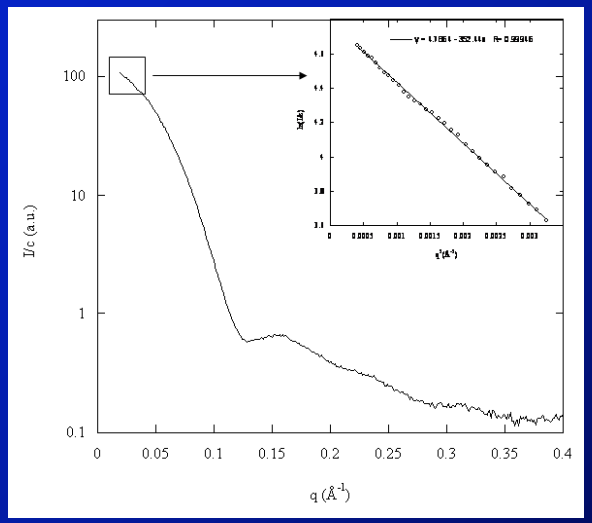
Available online at www.sciencedirect.com
ScienceDirect

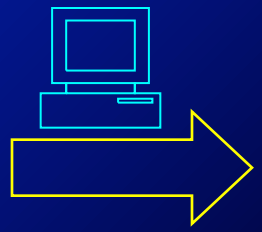
ELSEVIER

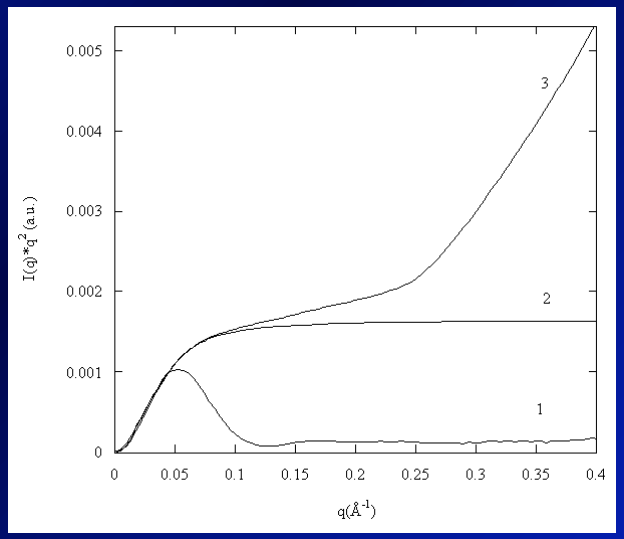
Structural Characterization of the Ceruloplasmin: Lactoferrin Complex in Solution

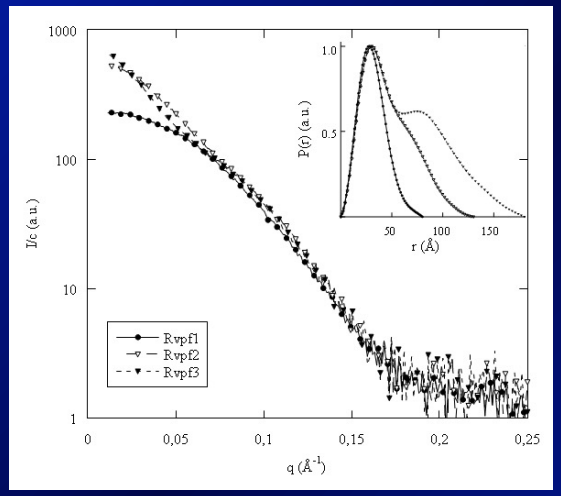
Annalaura Sabatucci[1], Patrice Vachette[3,4], Vadim B. Vasilyev[5]
Mariano Beltramini[2], Alexey Sokolov[5], Maria Pulina[5]
Benedetto Salvato[2], Clotilde B. Angelucci[1], Mauro Maccarrone[1]
Ivo Cozzani[1] and Enrico Dainese[1]*

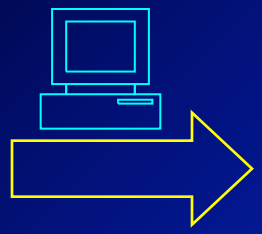
Direct structural parameters and 3D modeling

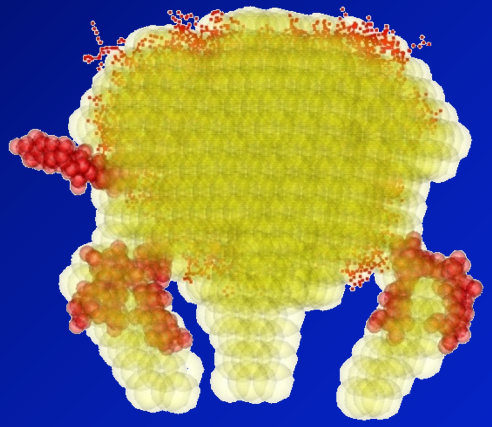




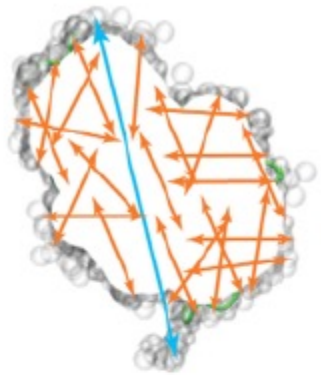






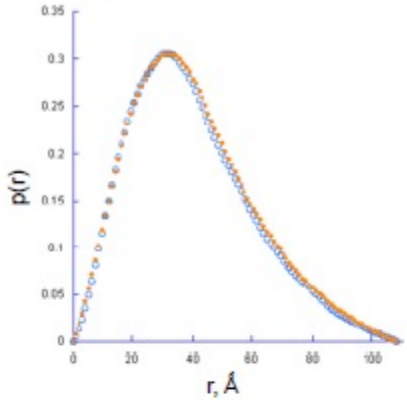


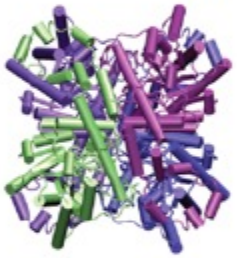
Interatomic Vectors



P(r) Function

Defined on $0 < r < d_{max}$

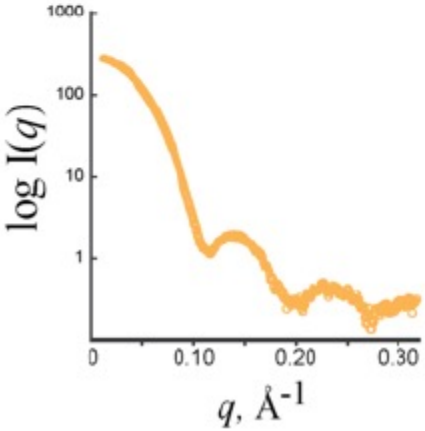


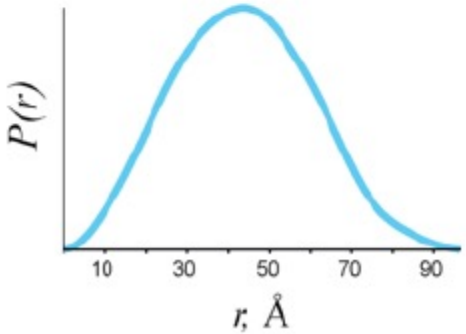


I(q) vs q

P(r)-distribution

$$I(q) = \int_0^{d_{max}} P(r) \frac{\sin\ q \cdot r}{q \cdot r} dr$$





TAKE HOME MESSAGES FOR THE SAXS APPROACH

- Complementarity between the crystallographic NMR and Cryo-EM studies and the solution work in the comprehension of structure-function relationships in proteins and the effect of new molecules in Drug Design;

- The increasing interest in the structural analysis of specific functional complexes between different proteins combined with the availability of molecular biology methods to obtain highly purified and concentrated preparations of recombinant proteins, allows to depict a scenario where SAXS will be one of the most promising and emerging methods in the post-genome structural studies;

- MD simulations can be driven by SAXS data helping to understand the dynamics of a biological process.

Applicazioni in ambito biomedico della metodica SAXS

Alcuni esempi di ambito biomedico in cui la metodica SAXS è particolarmente utile sono: l'analisi di modificazioni conformazionali di proteine allosteriche, lo studio strutturale di proteine multidominio i cui domini sono legati da linker flessibili, l'analisi di **proteine intrinsecamente disordinate** (*Intrinsically Disordered Proteins*, IDP), l'analisi di oligomeri iniziali responsabili di processi che portano alla formazione di **fibrille di tipo amiloide**, lo studio di proteine multimeriche, l'analisi delle **proteine di membrana** nonché delle interazioni proteine-lipidi.

Pertanto, la metodica SAXS consente anche lo studio di transizioni strutturali delle macromolecole biologiche, che vanno da cambiamenti conformazionali a processi di assemblaggio che, a loro volta, possono essere modulati in modo dinamico da perturbazioni dovute a interazioni con specifici **ligandi**, oppure da una varietà di modifiche chimico-fisiche del solvente. Per evidenziare la variazione conformazionale della macromolecola biologica oggetto di studio, la modifica delle variabili può essere effettuata gradualmente o mediante induzione di una rapida perturbazione con tecniche di mescolamento rapido (apparati a flusso interrotto, o ***stopped-flow***); ciò al fine di portare la dinamica del rilassamento del sistema verso il nuovo stato di equilibrio. Tali apparecchi

Journal of Medicinal Chemistry

This article is made available via the ACS COVID-19 subset for unrestricted RESEARCH re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for the duration of the World Health Organization (WHO) declaration of COVID-19 as a global pandemic.

pubs.acs.org/jmc

Article

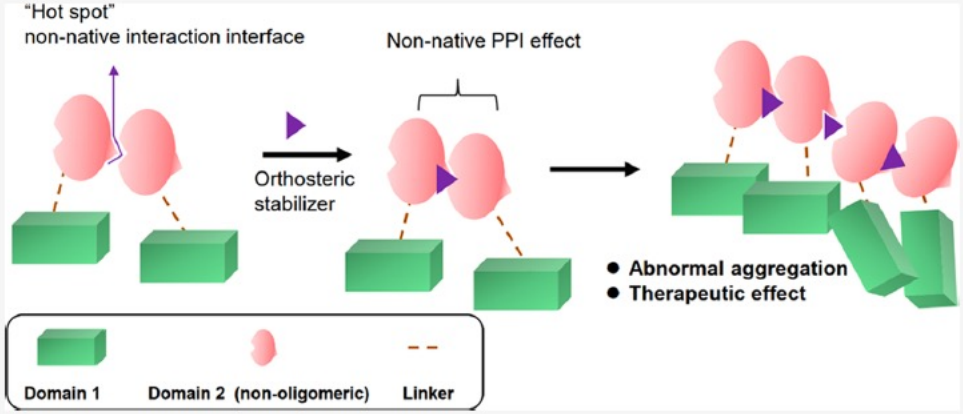
Structure-Based Stabilization of Non-native Protein–Protein Interactions of Coronavirus Nucleocapsid Proteins in Antiviral Drug Design

Shan-Meng Lin,[○] Shih-Chao Lin,[○] Jia-Ning Hsu,[○] Chung-ke Chang,[○] Ching-Ming Chien, Yong-Sheng Wang, Hung-Yi Wu, U-Ser Jeng, Kylene Kehn-Hall, and Ming-Hon Hou*

Cite This: *J. Med. Chem.* 2020, 63, 3131–3141

Read Online

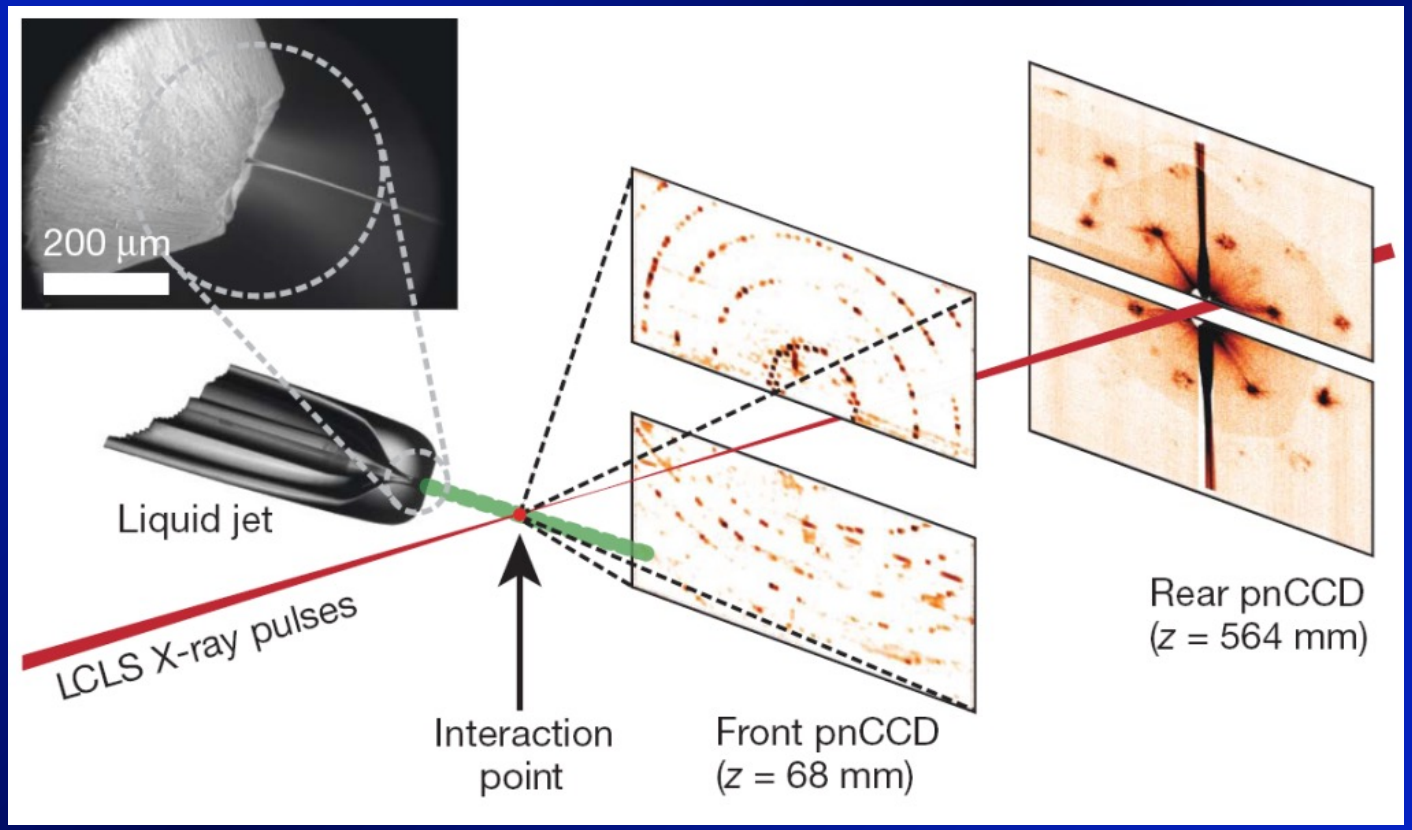
ACCESS | Metrics & More | Article Recommendations | Supporting Information



FUTURE DIRECTIONS

"FREE ELECTRON LASER" FEL

Image before destroy snapshots
femtosecond protein structure



X-Ray FEL video