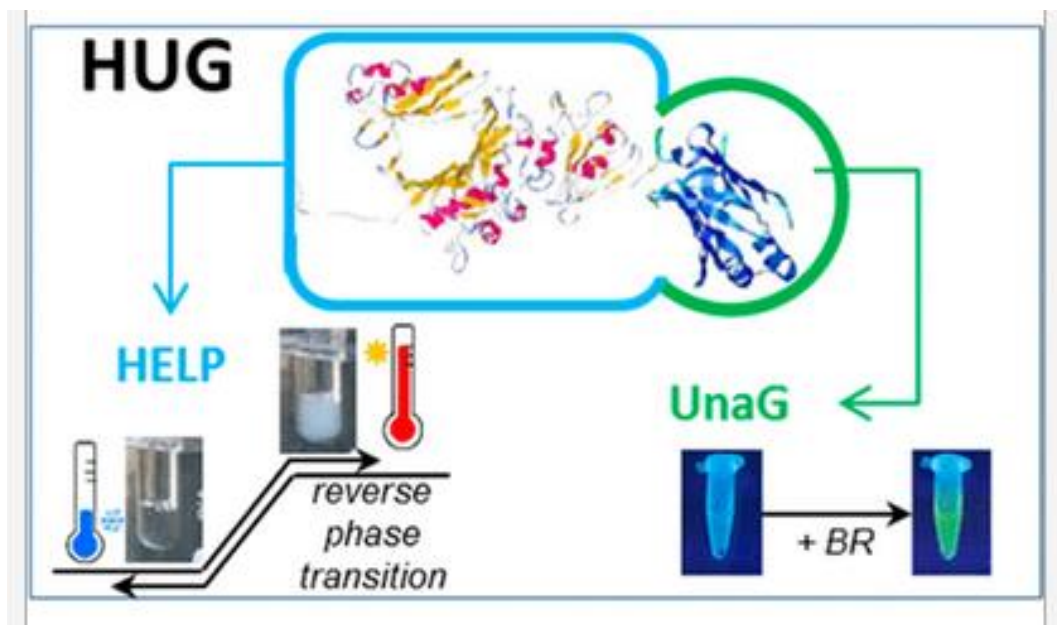


Macromolecular and Solution Properties of the Recombinant Fusion Protein HUG

Paola Sist, Antonella Bandiera, Ranieri Urbani,* and Sabina Passamonti



Ricerche svolte dal 2010 ad oggi

I polimeri proteici rappresentano una promettente alternativa ai composti sintetizzati chimicamente, ancora comunemente utilizzati in molte applicazioni biotecnologiche e biomediche.

Tali materiali proteici hanno un grande potenziale in molte applicazioni biotecnologiche e biomediche come componenti di base di biomateriali e matrici, vettori per la somministrazione controllata di farmaci, rivestimenti bioattivi e molte altre applicazioni.

Una strategia molto interessante per la progettazione di materiali avanzati con funzionalità migliorate è quella di imitare la struttura delle proteine presenti in natura per produrre polimeri proteici artificiali con le proprietà desiderate.

La costruzione di geni sintetici ripetitivi che codificano per proteine con proprietà desiderate può essere ottenuta con la tecnologia del DNA ricombinante (isolare e tagliare brevi sequenze di DNA per trasferirle e inserirle nel genoma di altre cellule, in modo da modificarne uno o più geni).

I geni possono essere ulteriormente adattati per includere altri domini che aggiungono diverse funzionalità.

Le macromolecole risultanti non esistono in natura, ma mantengono la loro origine biotica e possono essere prodotte da microrganismi ricombinanti.

In questo contesto, un modello molto interessante per la produzione di strutture desiderate è rappresentato dai **polipeptidi elastinici (ELP)**, che si basano su motivi ripetuti di elastina.

Gli ELP sono modellati sul motivo pentapeptidico regolarmente ripetuto che si trova nel dominio idrofobico della tropoelastina bovina, ossia **VPGVG**

Sono stati progettati e costruiti geni sintetici basati sulla tropoelastina umana per ottenere biopolimeri ingegnerizzati sul motivo esapeptidico ripetuto **VAPGVG**.

Il lavoro è finalizzato alla produzione di una famiglia di **polipeptidi simili all'elastina umana (HELP)** che presentano proprietà fisico-chimiche diverse.

Le cellule di *Escherichia coli* sono state trasformate con i costrutti di espressione recanti i geni sintetici HELP.

HUG

```
10      20      30      40      50      60      70
MRGSHHHHHH GSAAAAAAA KAAAKAAQFG LVPGVGVAPG VGVAPGVGVÄ PGVGLAPGVG VAPGVGVAPG

80      90      100     110     120     130     140
VGVAPGIAPÄ AAAAAKAAAK AAQFGLVPGV GVAPGVGVAP GVGVAPGVGL APGVGVAPGV VAPGVGVAP

150     160     170     180     190     200     210
GIAPAAAAAA KAAAKAAQFG LVPGVGVAPG VGVAPGVGVÄ PGVGLAPGVG VAPGVGVAPG VGVAPGIAPÄ

220     230     240     250     260     270     280
AAAAAKAAAK AAQFGLVPGV GVAPGVGVAP GVGVAPGVGL APGVGVAPGV VAPGVGVAPG GIAPAAAAAA

290     300     310     320     330     340     350
KAAAKAAQFG LVPGVGVAPG VGVAPGVGVÄ PGVGLAPGVG VAPGVGVAPG VGVAPGIAPÄ AAAAAKAAAK

360     370     380     390     400     410     420
AAQFGLVPGV GVAPGVGVAP GVGVAPGVGL APGVGVAPGV VAPGVGVAPG GIAPAAAAAA KAAAKAAQFG

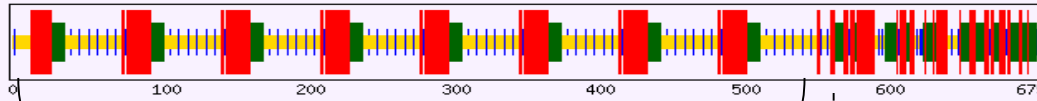
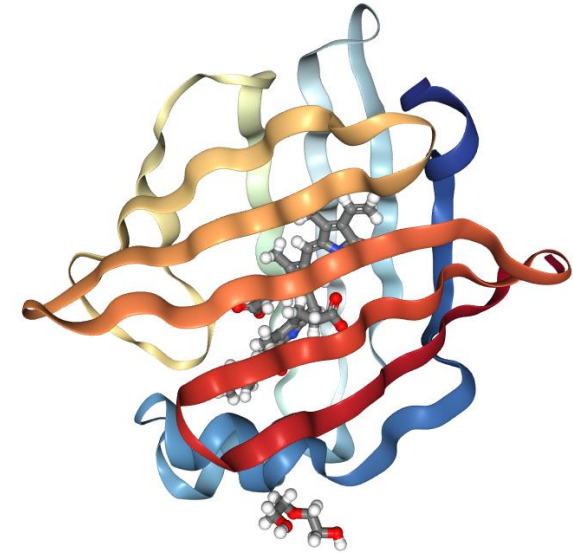
430     440     450     460     470     480     490
LVPGVGVAPG VGVAPGVGVÄ PGVGLAPGVG VAPGVGVAPG VGVAPGIAPÄ AAAAAKAAAK AAQFGLVPGV

500     510     520     530     540     550     560
GVAPGVGVAP GVGVAPGVGL APGVGVAPGV VAPGVGVAPG GIAPGGMVEK FVGWTKIADS HNFGEYLKAI

570     580     590     600     610     620     630
GAPKELSDGG DATPTLYIS QKDGDKMIVK IENGPTFLD TQVKFKLGEÄ FDEFPSDRRK GVKSVVNLVÄ

640     650     660     670
EKLVYVQKWD GKETTYVREI KDGKLVVILT MGDVVAVRSY RRATE
```

UnaG



HELP

Kyte & Doolittle Hydrophobicity Plot

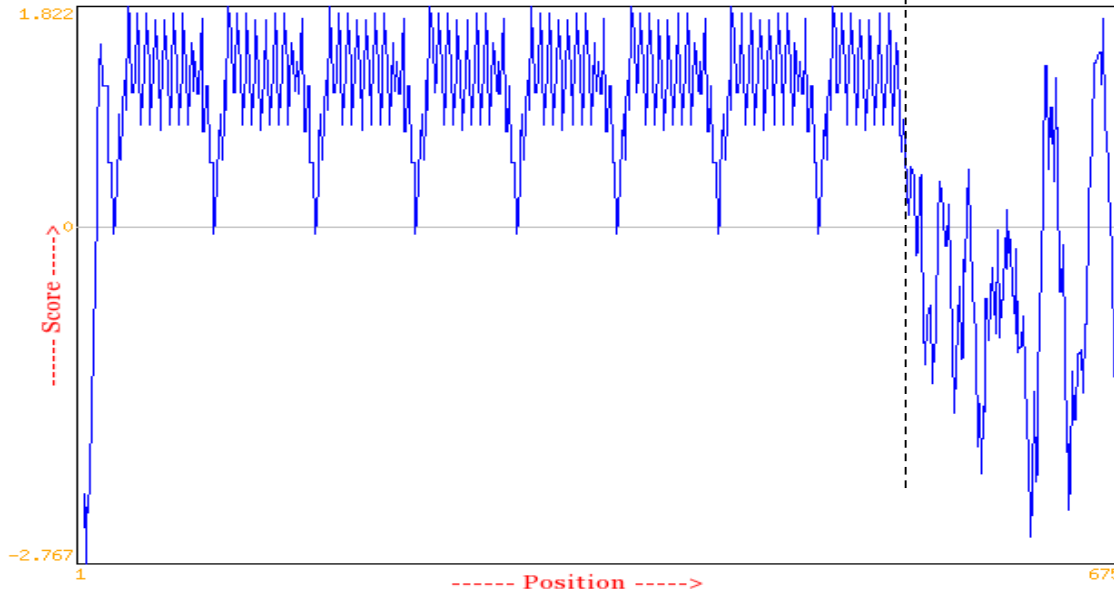


Table 3. Prediction of secondary structure using GOR IV (%)

	number of a.a.	α -helix %	β -strand %	random coil + β -turn, %
HUG	675	22	10	68
HELP	536	26	4	70
Human ELP	757	23	12	65
<u>UnaG</u>	141	8	35	57

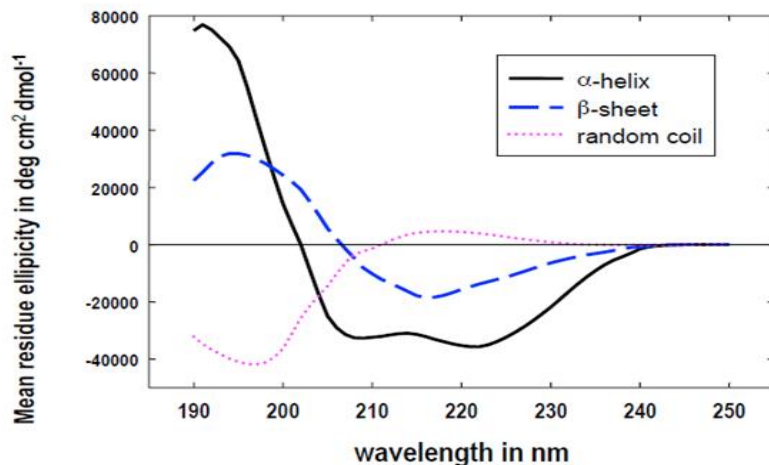
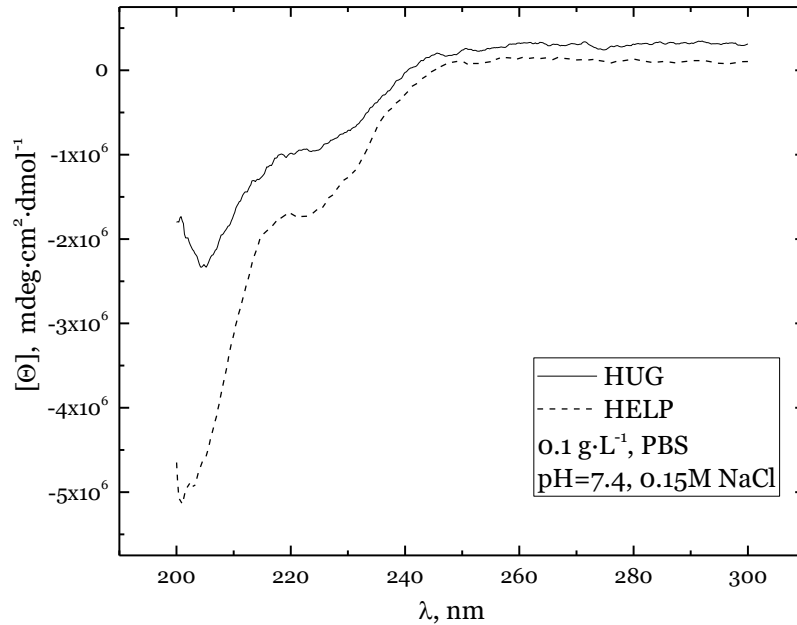
CHARACTERIZATION OF HUG

- **Molecular mechanics and dynamic simulation**
- **Circular Dichroism (CD)**
- **Differential Scanning Calorimetry (DSC)**
- **Potentiometric titration of HUG and HELP**
- **Binding of bilirubin by fluorescence**

The physical and chemical characterization of the new fluorescent biosensor HUG. This step is very important for optimization of liquid assay of free bilirubin in biological samples

CIRCULAR DICHROISM

Il dicroismo circolare (CD) è un metodo eccellente per valutare rapidamente la struttura secondaria, il ripiegamento e le proprietà di binding delle proteine.



Spettri CD ottenuti ad una temperatura inferiore all'ITT ($T = 25 \text{ }^\circ\text{C}$) mostrano il tipico profilo riportato per i polipeptidi simili all'elastina, con una banda negativa intorno ai 200 nm (transizione $\pi\pi^*$), associata alla coesistenza di diverse popolazioni di conformeri che variano tra un'alta proporzione di *coil* e una bassa proporzione di strutture secondarie β -turn.

La differenza di segnale CD a 200 nm tra HUG e HELP è dovuta al **contributo positivo della struttura β** attribuito al dominio HUG.

La presenza di ulteriori bande negative e di un minimo meno pronunciato a 222 nm (transizione $n\pi^*$) a lunghezze d'onda maggiori è stata attribuita sia ai domini α -elica (222 nm) sia alle conformazioni β -turn (225 nm).

Negli ultimi trent'anni, sono stati sviluppati diversi metodi per l'analisi degli spettri CD delle proteine, basati sulle caratteristiche spettrali delle strutture secondarie.

In questi metodi vengono utilizzati gli spettri CD di **polipeptidi modello** o di un **insieme di proteine di riferimento con struttura cristallina nota**.

In questo modo lo spettro CD di una determinata proteina viene analizzato come una combinazione lineare di spettri di struttura secondaria nota.

$$\theta_t = x_\alpha \theta_\alpha + x_\beta \theta_\beta + x_c \theta_c$$

Si ottiene una serie di equazioni lineari che vengono risolte con minimi quadrati o con metodi di riconoscimento di pattern ossia del contenuto di struttura secondaria corrispondente a un dato spettro CD.

Le informazioni derivate da queste analisi sono state in gran parte limitate alla stima della frazione di strutture α -elica, β -sheet, β -turn e poly(Pro)-II.

Lo spettro CD della proteina osservata è una media dei segnali CD di tutte le conformazioni e riflette la variabilità geometrica della struttura secondaria.

Un'altra variabile strutturale nelle proteine è la **lunghezza media delle α -eliche**; ad esempio, in generale, le proteine ricche di α -eliche hanno α -eliche più lunghe e quelle ricche di β -turn hanno α -eliche più corte.

Raggi-X di 29 proteine di riferimento per la deconvoluzione degli spettri CD

- | | |
|--|---|
| 1. myoglobin, | 15. Papain |
| 2. <u>Hemoglobin</u> | 16. <u>subtilisin novo</u> |
| 3. <u>Hemerythrin</u> | 17. <u>Ribonuclease A</u> |
| 4. T4-lysozyme | 18. <u>Pepsinogen</u> |
| 5. triose-phosphate isomerase | 19. <u>β-lactoglobulin</u> |
| 6. lactate dehydrogenase | 20. <u>α-chymotrypsin</u> |
| 7. Lysozyme | 21. <u>Azurin</u> |
| 8. <u>Thermolysin</u> | 22. <u>Elastase</u> |
| 9. cytochrome c | 23. <u>γ-crystallin</u> |
| 10. Phosphoglycerate kinase | 24. <u>Prealbumin</u> |
| 11. Eco R1 endonuclease | 25. <u>concanavalin A</u> |
| 12. <u>flavodoxin</u> | 26. <u>Bence-Jones protein</u> |
| 13. <u>subtilisin BPN9</u> | 27. <u>tumor necrosis factor</u> |
| 14. glyceraldehyde 3-phosphate dehydrogenase | 28. <u>superoxide dismutase</u> |
| | 29. <u>α-bungaratoxin</u> |

In ordine decrescente di contenuto di α -eliche

In generale, esistono **due metodi** per valutare la conformazione delle proteine.

- Il primo utilizza **standard di polipeptidi**, con composizioni definite di conformazioni, che sono state determinate mediante scattering di raggi X su film o mediante IR in soluzione.
- Il secondo utilizza **come standard gli spettri di proteine** caratterizzate da cristallografia a raggi X. Questi sono poi confrontati con gli spettri di proteine sconosciute mediante minimi quadrati.

L'analisi va a calcolare la RMSD (δ) ed i coefficienti di correlazione (r) tra le stime delle frazioni di struttura secondaria dai raggi X e dai CD per le diverse assegnazioni di struttura secondaria delle proteine di riferimento.

La regressione lineare è anche utile per valutare gli effetti di mutazioni, di ligandi e solventi sulla conformazione delle proteine.



		α %	β %	rc %
HELP	theor	26	4	70
	CD	29	10	61

DSC SETARAM

I metodi calorimetrici presentano un vantaggio rispetto ad altri metodi per la loro capacità di fornire un quadro relativamente completo dei parametri termodinamici durante le reazioni di legame, comprese le costanti di legame all'equilibrio, l'entalpia delle reazioni di legame (ΔH) e la variazione di entropia (ΔS).

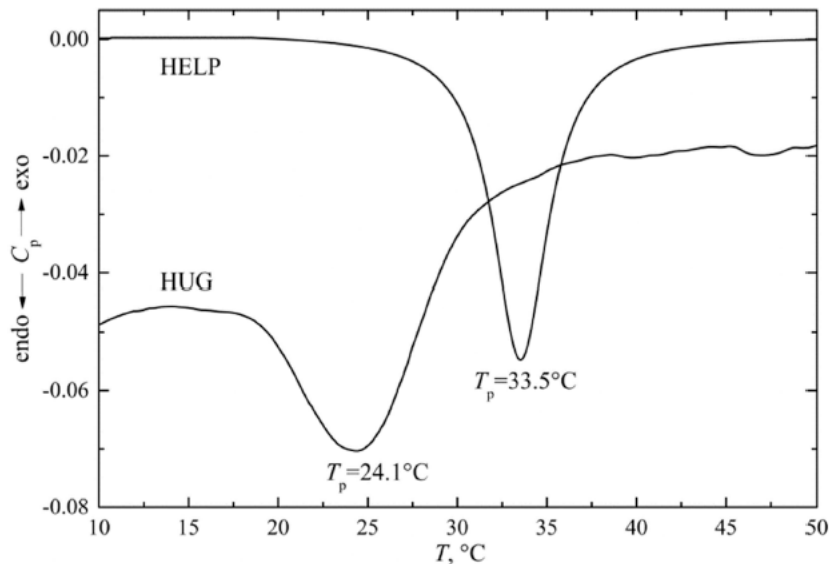


Figure 6. DSC thermograms of HELP and HUG. Solutions (4 g L^{-1} , in PBS, $\text{pH} = 7.4$) were analyzed at scan rate of $0.5 \text{ }^\circ\text{C min}^{-1}$.

Table 4. Onset and Peak Temperature, Enthalpy, and Entropy Variation for HUG and HELP Proteins From DSC Measurements (HUG and HELP 4 g L^{-1} in PBS, $\text{pH} = 7.4$)

	$T_{\text{ons}}, \text{ }^\circ\text{C}$	$T_p, \text{ }^\circ\text{C}$	$\Delta H_{\text{tr}}, \text{ kJ}\cdot\text{mol}^{-1}$	$\Delta S_{\text{tr}}, \text{ kJ}\cdot\text{mol}^{-1}\text{K}^{-1}$
HUG	19.0	24.1	158	12.4
HELP	29.4	33.5	216	12.8

Questi dati mostrano una differenza significativa nelle temperature ed entalpie tra i biopolimeri HUG e HELP, principalmente a causa del diverso contributo della distruzione dell'acqua strutturata associata al dominio idrofobico della catena, che è maggiore per la proteina HELP rispetto alla HUG.

La variazione di entropia durante la transizione inversa è simile e ciò suggerisce che le sfere idrofobiche di solvatazione molto estese di questi due biopolimeri sono simili e contribuiscono in modo simile all'entropia di transizione.

Potentiometric titration of HUG and HELP

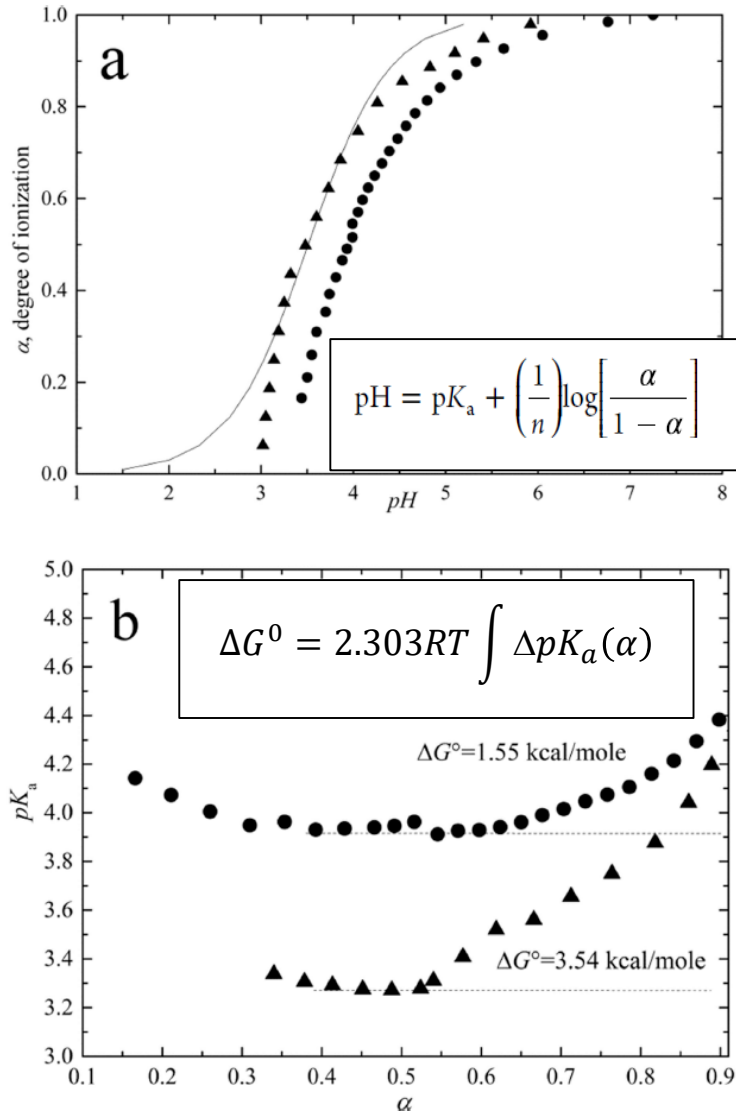


Figure 4. Potentiometric analysis of HUG and HELP. (a) Titration curves for acid-dialyzed proteins HUG (triangle) and HELP (circles). The solid line represents the Henderson–Hasselbalch equation. (b) Dependence of pK_a on the degree of protonation (α).

L'aumento della pK_a è una conseguenza della formazione di specie cariche durante la titolazione, che è responsabile della destrutturazione delle molecole d'acqua nella struttura pentagonale della struttura idrofobica di idratazione.

In altre parole, i gruppi COO^- di nuova formazione, spostano l'acqua dai gusci di idratazione idrofobici nei propri gusci di idratazione.

L'aumento della densità di carica durante la titolazione causa un crescente effetto di repulsione carica-carica, che aumenta l'energia libera del sistema e si osserva una continua variazione del valore pK (cooperatività negativa) per $\alpha > 0,3$.

In presenza di repulsione carica-carica, la pendenza della curva di titolazione (-) è molto più netta di quella data dall'equazione di Henderson-Hasselbalch.

Il ΔG è pari a 1,55 e 3,54 $\text{kcal}\cdot\text{mol}^{-1}$ per i biopolimeri HELP e HUG, (Figura 4b), maggiore per HUG dato il maggior contenuto di residui carichi (10%) rispetto al biopolimero HELP (3,2%).

La deviazione dall'equazione di Henderson-Hasselbalch è di solito valutata mediante l'introduzione del coefficiente di Hill, n_H

$$\text{pH} = \text{p}K_a + \left(\frac{1}{n_H} \right) \log \left[\frac{\alpha}{1 - \alpha} \right]$$

che introduce la **cooperatività** nella teoria della titolazione dovuta sia all'azione cooperativa del residuo acido che della significativa idratazione idrofobica.

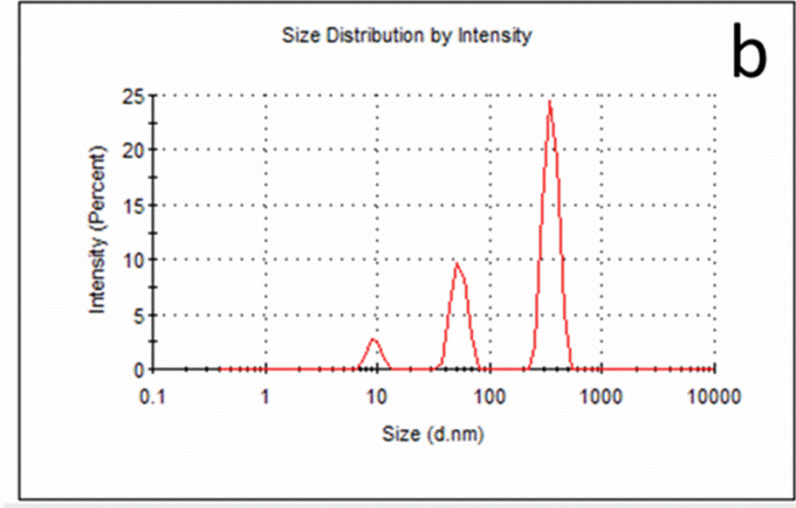
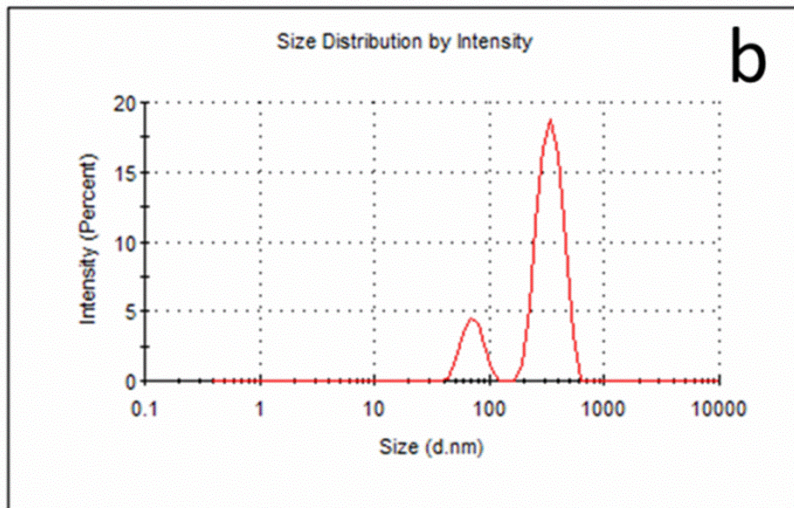
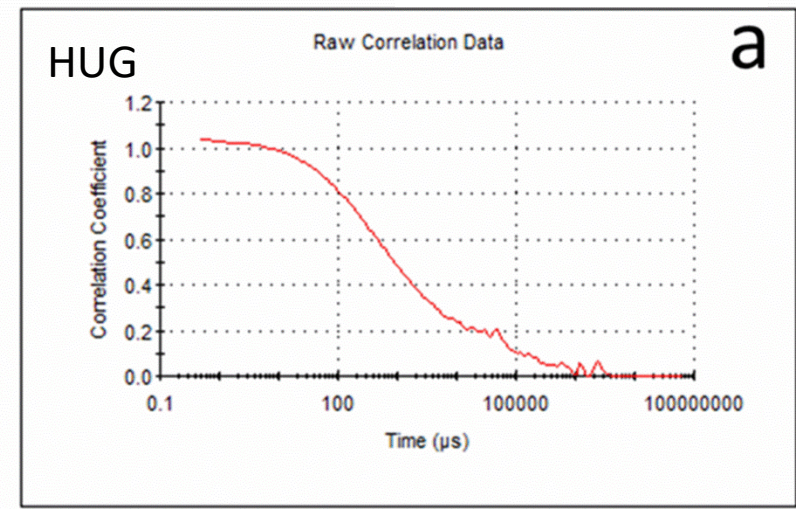
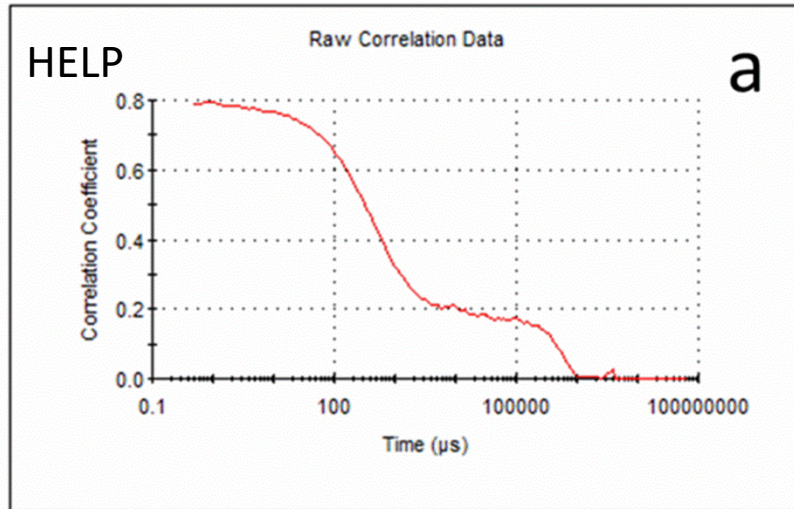
L'aumento di $\Delta\text{p}K$ dovuto alla repulsione carica-carica (Figura 4b) allarga la curva di titolazione (cooperatività negativa) di entrambe le proteine HUG e HELP in soluzione, a una temperatura inferiore della T di transizione inversa in modo che rimangano in soluzione per tutta la durata della titolazione.

In queste condizioni, entrambe le proteine non mostrano alcuna transizione di ripiegamento e assemblaggio idrofobico quindi lo spostamento di $\Delta\text{p}K$ è dovuto principalmente al lavoro necessario per disgregare la struttura idrofobica di idratazione.

Dynamic Light Scattering

$$G(\tau) = A[1 + Be^{-2\Gamma\tau}]$$

$$\Gamma = Dq^2 \quad \Rightarrow \quad d_h = \frac{kT}{3\pi\eta D}$$



Dynamic Light Scattering

Utilizzando la tecnica DLS siamo in grado di misurare il **raggio idrodinamico** delle macromolecole in soluzione e le dimensioni di aggregati di diverse dimensioni a diverse temperature e concentrazioni.

Sono state osservate distribuzioni dimensionali trimodali per la proteina HUG nell'intervallo di temperatura 10-30 °C con un valore medio di R_h di 62, 250 e 1523-2497 Å.

Queste tre distribuzioni differiscono per la percentuale di area di picco, indicando che le particelle con le **dimensioni più elevate** in questa soluzione sono la frazione principale fino a 30 °C.

Con l'aumento della temperatura la proteina mostra una transizione di fase inversa e i picchi si risolvono in un'unica distribuzione, formando enormi aggregati complessi con raggi idrodinamici sempre più grandi, fino a fino a 6000 Å.

Questo comportamento è simile a quello osservato per il biopolimero HELP

In questo caso, la frazione polipeptidica più importante nell'intervallo di temperatura 10-30 °C è quella con $R_h = 50$ Å circa (78%) insieme a quella di circa $R_h = 1500$ Å (22%)

Aumentando la temperatura, la proteina mostra un'aggregazione iniziale formando particelle con un valore medio di R_h superiore a 1000 Å.

La coacervazione completa è poi raggiunta a $T > 50$ °C, per HELP e $T > 30$ °C per HUG dove si osservano singole distribuzioni (monomodali).

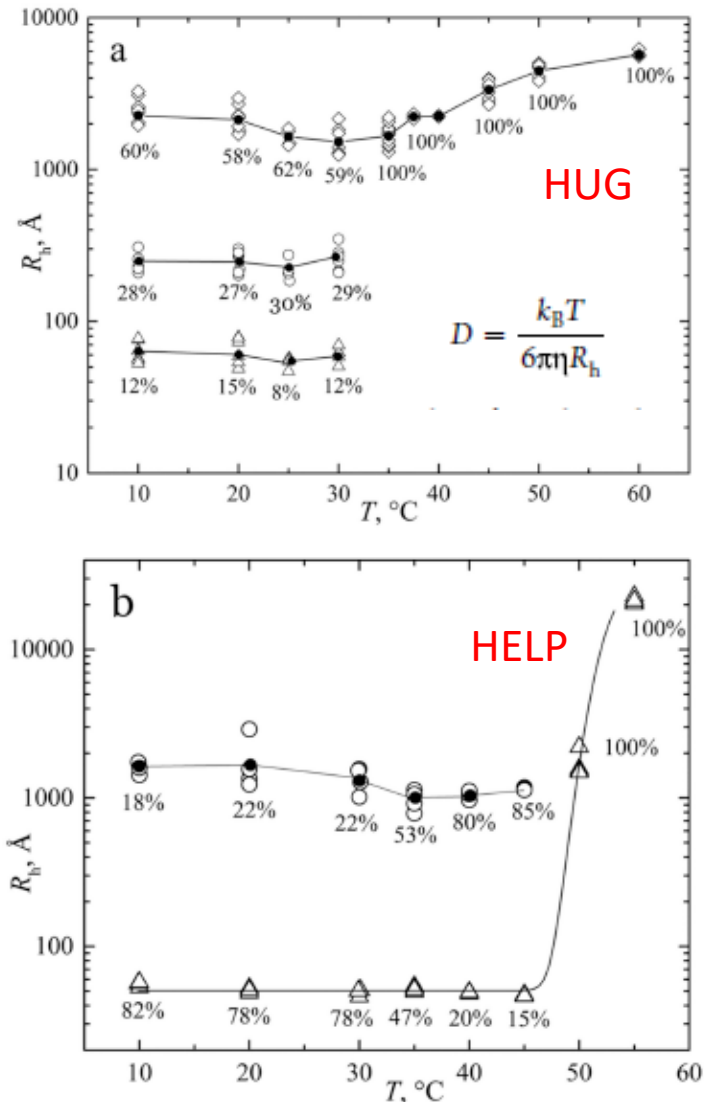


Figure 7. Temperature dependence of the hydrodynamic radii of HUG and HELP. (a) HUG values: (Δ) $R_h = 62$ Å, (\circ) $R_h = 250$ Å, (\diamond) $R_h = 1523-2497$ Å. (b) HELP average values: (Δ) $R_h = 50$ Å, (\circ) $R_h = 1350$ Å. Filled circles are average values.

Static Light Scattering

Utilizzando l'equazione di Debye, possiamo determinare il peso molecolare (M_w) e il secondo coefficiente viriale (A_2) del polimero in un determinato mezzo solvente dall'equazione

$$\frac{KC}{R_\theta} = \frac{1}{M_w} + 2A_2C$$

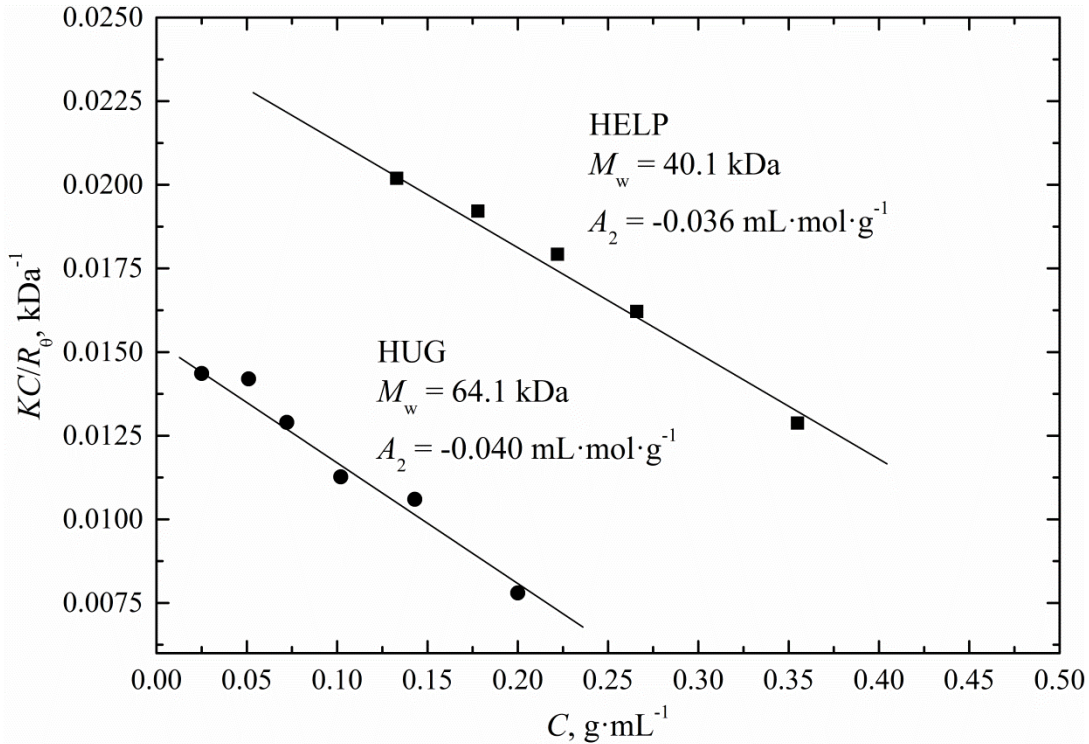


Table 2. Static light scattering results at 25°C in 0.15 M NaCl solution

	<i>Theoretical M_w</i> <i><u>kDa</u></i>	<i>M_w, <u>kDa</u></i>	<i>A_2, $mL \cdot mol \cdot g^{-2}$</i>
HUG	60.4	65.3 ± 5.7	-0.049 ± 0.024
HELP	44.9	40.5 ± 0.46	-0.065 ± 0.053

Questi risultati confermano che i polimeri espressi dai batteri corrispondono alla sequenza primaria dedotta dalle sequenze geniche e che non si è verificata alcuna degradazione delle proteine durante il processo di purificazione.

Il secondo coefficiente viriale A_2 è sempre negativo, indicando una debole interazione tra le catene polimeriche e le molecole di solvente (solvente povero), motivo per cui questi polimeri tendono ad aggregarsi.

Predizione di struttura

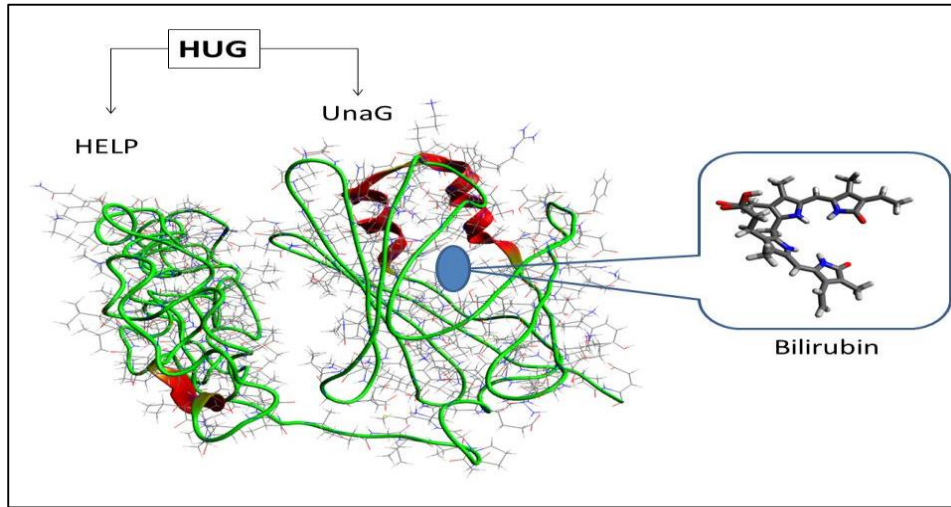


Figure 1. I-TASSER minimized structure of HUG fragment

Piattaforma *on-line* I-TESSER

La predizione della struttura e della funzione.

I-TASSER è un approccio gerarchico che permette la predizione di modelli 3D e funzioni biologiche di alta qualità partendo dalla struttura primaria attraverso un metodo di *threading* multiplo.

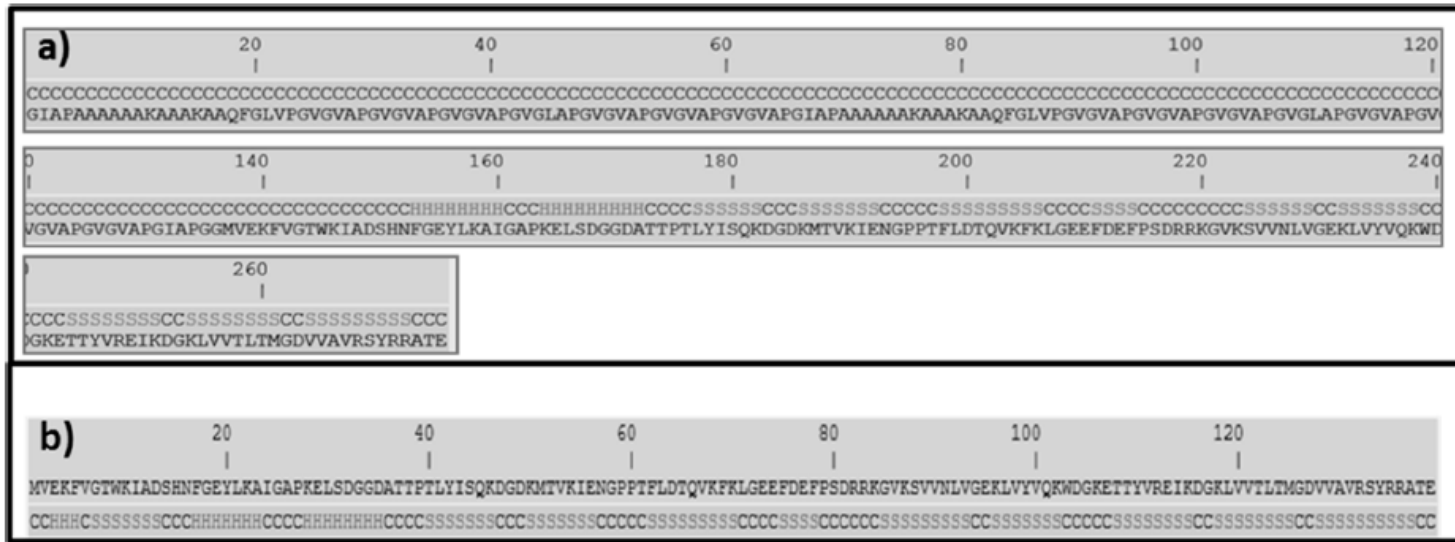
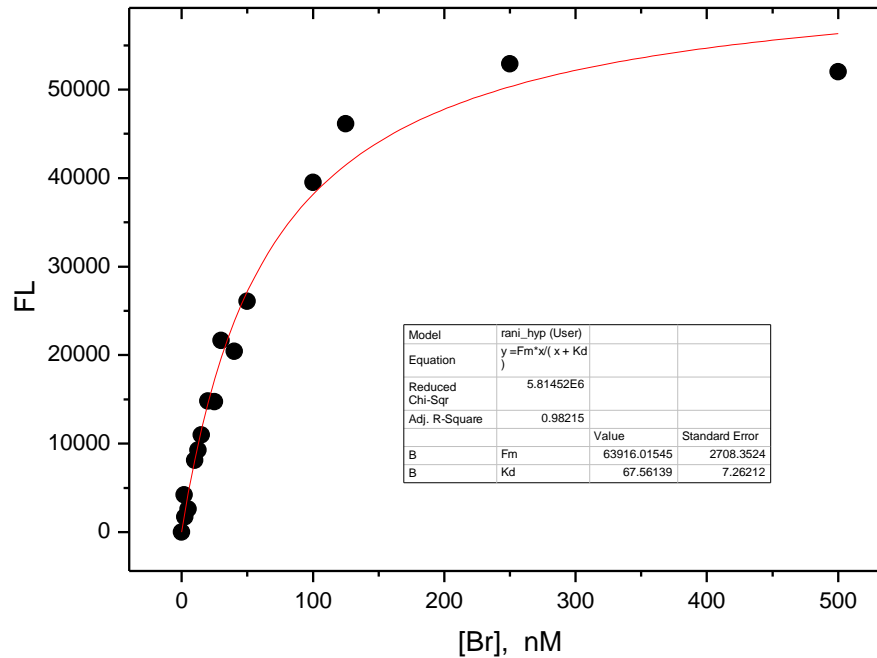


Figure 2. Prediction of the secondary structure of a portion of HUG and of UnaG. (a) Primary structure of HUG starting from a.a. 400 and (b) primary structure of UnaG. Both are associated to the secondary structures (C = coil, H = α -helix, S = β -strand), obtained by I-TASSER simulation.

È interessante notare che la simulazione ha mostrato che entrambe le sequenze, HELP e UnaG, mantengono le proprie strutture secondarie, anche nella proteina di fusione HUG.

La regione HELP adotta la conformazione coil (100%), mentre il dominio UnaG mostra il β -sheet (46%), il random coil (41%) e la conformazione a elica (12%).

Binding HUG / Bilirubina



For each fluorescence value (F), the fractional enhancement (Y) was computed by the equation:

$$Y = \frac{F}{F_0}$$

where Y is the fractional saturation related to the extent of binding and F_0 the fluorescence intensity at the Br/HUG ratio greater than 1 (the asymptotic value).

$$K_d = \frac{[\text{HUG}][\text{BR}]}{[\text{BR-HUG}]}$$

Durante la titolazione, la concentrazione di BR aumenta e la saturazione Y è espressa in termini di concentrazione del complesso BR-HUG

$$Y = \frac{[\text{BR-HUG}]}{[\text{HUG}]_T}$$

$$[\text{BR-HUG}] = \frac{[\text{HUG}][\text{BR}]}{K_d}$$

$$[\text{HUG}]_T = [\text{HUG}] + [\text{BR-HUG}] = P_T$$

$$Y = \frac{[\text{HUG}][\text{BR}]/K_d}{[\text{HUG}] + [\text{HUG}][\text{BR}]/K_d}$$

$$Y = \frac{[\text{BR}]}{K_d + [\text{BR}]}$$

Poiché nell'esperimento di binding, invece delle concentrazioni delle specie libere, sono note le concentrazioni **totali** di proteina e ligando, l'espressione di **Y** in funzione delle quantità totali viene ottenuta come segue

$$[\text{BR}] = [\text{BR}]_T - [\text{BR-HUG}] = L_T - [\text{BR-HUG}]$$

$$K_d = \frac{(L_T - [\text{BR-HUG}])(P_T - [\text{BR-HUG}])}{[\text{BR-HUG}]}$$

$$K_d[\text{BR-HUG}] = P_T L_T - (P_T + L_T)[\text{BR-HUG}] + [\text{BR-HUG}]^2$$

$$[\text{BR-HUG}]^2 - (P_T + L_T + K_d)[\text{BR-HUG}] + P_T L_T = 0$$

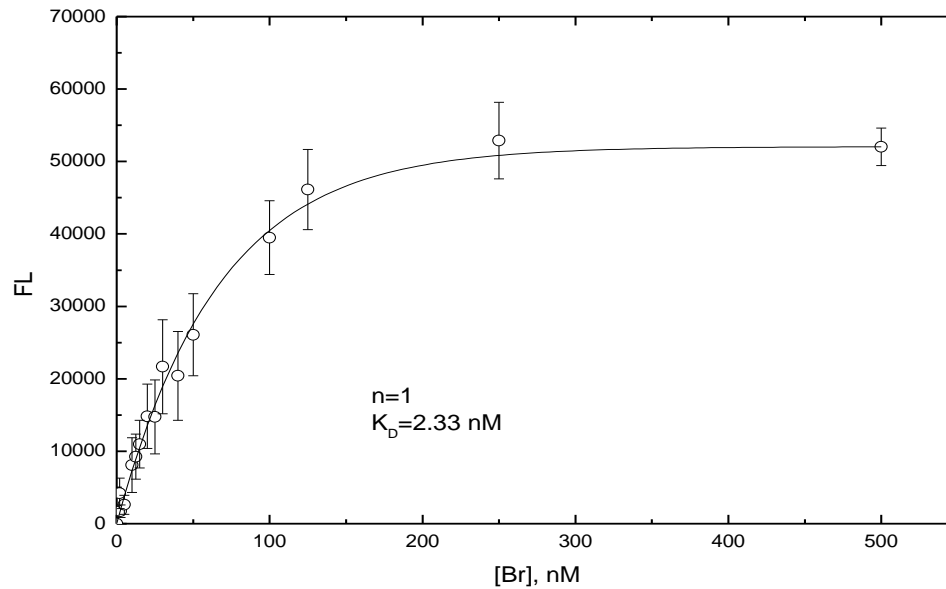
Le radici dell'equazione di secondo grado è:

$$[\text{BR-HUG}] = \frac{(P_T + L_T + K_d) - \sqrt{\{[-(P_T + L_T + K_d)]^2 - 4P_T L_T\}}}{2}$$

Che sostituita nell'equazione di **Y**

$$Y = \frac{(P_T + L_T + K_d) - \sqrt{\{[-(P_T + L_T + K_d)]^2 - 4P_T L_T\}}}{2P_T}$$

Il grafico di $F=F_oY$ rispetto a $L_T = [Br]$ è mostrato in Figura, dove sono riportati i valori medi di tutti i risultati ottenuti da diversi esperimenti.

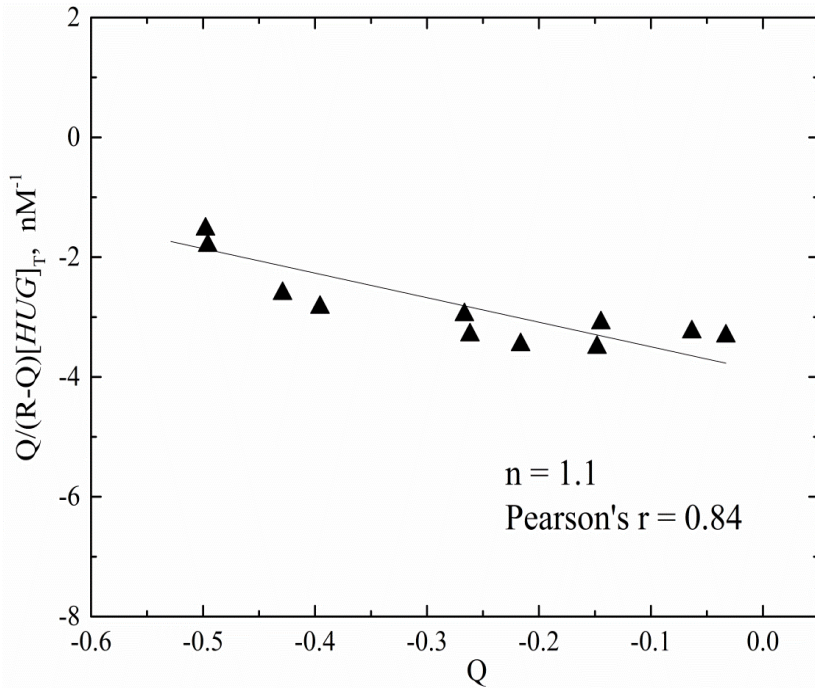


Mediante fitting non lineare della curva iperbolica si ottengono i valori delle costanti di dissociazione e di binding ($K_D=1.1 \text{ nM}$, $K_A=0,91 \times 10^9 \text{ M}^{-1}$) e del valore massimo di fluorescenza della proteina legata al ligando ($F_o=65600$).

L'assunzione del modello con $n=1$ è confermata dal diagramma di Scatchard.

Il valore della costante $K_a = 0,91 \times 10^9 \text{ M}^{-1}$ per il complesso BR-HUG ha dato un valore maggiore di quello del complesso BR-Albumina ($K_a \approx 10^7$), che determina il sequestro da parte di HUG di BR dall'albumina.

Scatchard plot for BR and HUG binding.



The Scatchard plot was obtained by using equation (Levine, 1977):

$$v/C_f = \frac{Q}{(R-Q)[HUG]_T} = nK_a - K_a Q$$

where:

$$Q = \frac{[BR-HUG]}{[HUG]_T} = v$$

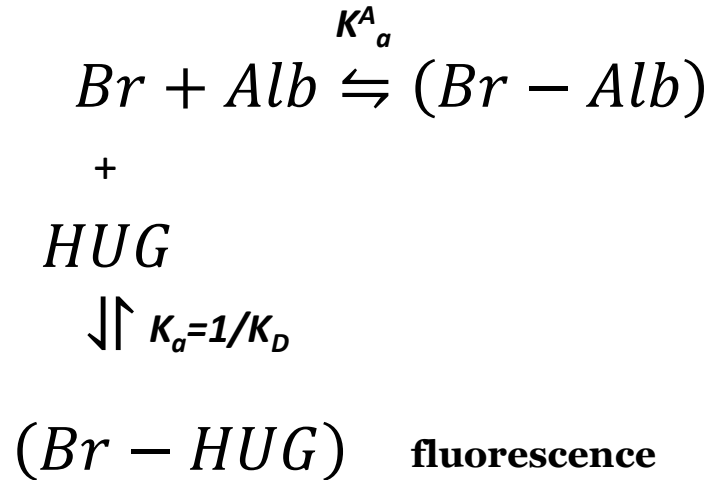
$$R = \frac{[BR]_T}{[HUG]_T}$$

with $[HUG]_T$ and $[BR]_T$ the total concentration of HUG and BR, respectively.

Q is related to the fractional saturation $Y = F/F_o$ as defined above.

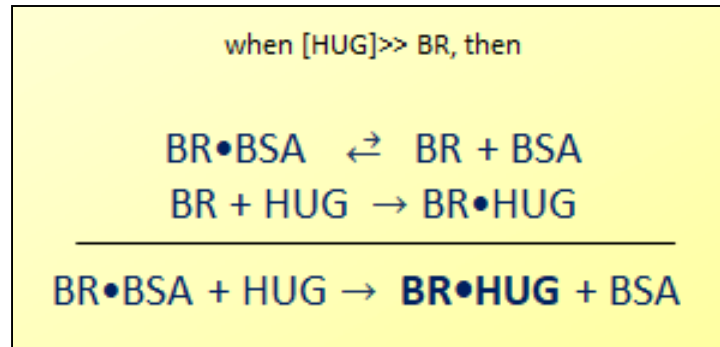
The linear fitting of data in Figure gives

$n = 1.1$.



$$K_a = 1/K_D = 0,91 \times 10^9 \text{ M}^{-1} \text{ per } Br-HUG$$

$$K_a^A = 1/K_D = \text{ca } 10^7 \text{ M}^{-1} \text{ per } Br-Alb$$



Alb = BSA