

Relazione progetto LAB5

Laboratorio di modellizzazione e screening di molecole bioattive (A.1.5): Sviluppo di metodologie per la modellizzazione e lo studio di farmaci e biofarmaci.

Il programma di ricerca del laboratorio di Bioinformatica del CRS4 riguarda le attività di ricerca industriale (RI) dell'Obiettivo realizzativo 1 (OR1): Modelling molecolare e bioinformatica, diviso nelle seguenti fasi operative:

1. OR1A

- 1A.1. Individuazione delle possibili modificazioni aminoacidiche nelle citochine, chemochine e neutrofine.
- 1A.2. Valutazione della probabilità dell'interazione tra recettore e citochine mutate.
- 1A.3. Messa a punto e validazione di un metodo predittivo delle interazioni tra altre proteine e relativi recettori.
- 1A.4. Ottimizzazione di sistemi strutturali/enzimatici implicati nella produzione di biocomposti.

In questo rapporto delle attività sono riportate le attività di ricerca svolte nel periodo 1 Gennaio 2006-30 Giugno 2008.

Oggetto dell'indagine è la nuova metodologia enzimatica per la preparazione di derivati di proteine terapeutiche che prevede l'uso dell'enzima Microbial Transglutaminase (MTGase) da *Streptovorticillum Mobaerensis* per legare in modo covalente catene del polimero biocompatibile PEG a siti specifici delle proteine targets.

Le attività iniziali del progetto sono state volte alla definizione della relazione tra sequenza aminoacidica e individuazione di siti accessibili a modificazioni post-trascrizionali per le proteine della famiglia delle citochine, tramite l'analisi dettagliata dello stato dell'arte sia da un punto di vista sperimentale che modellistico. In particolare, in accordo con BIOKER che già aveva iniziato le analisi su tali proteine, il sistema modello da cui si è decisi di partire con l'analisi bioinformatica è quello dell'interazione tra l'enzima transglutaminasi e la proteina Granulocyte Colony Stimulating Factor (G-CSF).

1A.1. Individuazione delle possibili modificazioni aminoacidiche nelle citochine, chemochine e neutrofine

L'enzima MTGase (E.C.:2.3.2.13.), prodotto nello strain s-8112 del microrganismo *Streptovorticillum mobaraense*, è una transferase che catalizza la reazione di trasferimento di un gruppo acyl tra il gruppo γ -carbossamide dei residui di glutammia e i gruppi γ -amino dei residui di lisina formando una ϵ -(γ -glutammi)lisina. La sua struttura è stata risolta ai raggi X [1] (PDB code 1IU4) e come si può vedere in [Figura 1](#) ha una struttura a disco con una tasca contenente la triade catalitica (CYS64, HIS274, ASP255). Tale struttura tridimensionale è diversa da quella delle TGs di mammifero (rappresentativo è il fattore umano XIII della coagulazione, PDB code 1FIE), nonostante sia conservata la disposizione spaziale degli amminoacidi a livello del sito attivo. In particolare, la triade catalitica (C64, D255 e H274) risulta essere sovrapponibile, con un'inversione della posizione di His e Asp rispetto alla Cys.


```

51  GHSLGIPWAPLSSCPSQALQLAGCLSQLHSGFLYQGLLQALEGISPELG
    HHHHHT      TTTTT      HHHHHHHHHHHHHHHHHHHHTTTTTHH

101 PTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAGG
    HTHHHHHHHHHHHHHHHHHHHHHHT      SS      SHHHHHH

151 VLVASHLQSFLEVSYRVLRLHLAQF
    HHHHHHHHHHHHHHHHHHHHHHT

```

L'analisi degli intorni delle glutammine presenti nella sequenza aminoacidica del G-CSF identifica come possibili siti di transglutaminazione Q11, Q131, Q173 a causa della loro vicinanza ad aminoacidi positivi o carichi (rispettivamente K16, H79 ed R169, H170). Un altro fattore segnalato in letteratura è la maggiore mobilità o flessibilità del backbone vicino alla glutammina ed, in effetti, la predizione di struttura secondaria evidenzia un lungo loop "disordered" contenente Q131 e Q134. In totale quindi tramite questa prima analisi bioinformatica sono stati identificati 4 possibili siti di transglutaminazione del G-CSF (Q11, Q131, Q134, Q173) da sottoporre a più approfonditi studi sperimentali in laboratorio da parte di BIOKER. Inoltre l'ipotesi della flessibilità sarà studiata in modo più approfondito nel seguito della ricerca, tramite simulazioni di dinamica molecolare per identificare le regioni più flessibili del backbone del substrato.

1A.2. Valutazione della probabilità dell'interazione tra recettore e citochine mutate.

Nel caso sia presente nel Protein Data Bank la struttura della proteina in complesso con il recettore si possono avere indicazioni molto attendibili sui residui coinvolti nell'interazione con il recettore. E' questo il caso della proteina G-CSF per la quale esiste la struttura cristallizzata in forma complessata con il recettore(PDBid: 2d9q). Attraverso il calcolo della Solvent Accessible Surface (SAS) per le glutammine nella proteina isolata e nel complesso con il recettore sono state identificate le glutammine coinvolte nell'interazione con il recettore, i risultati sono riportati nella tabella 1. Il calcolo delle SAS è stato svolto utilizzando MOLMOL[9] e SCit [10].

La Q11 risulta essere coinvolta nel legame con il recettore e quindi non può essere sottoposta a Peghlylazione per evitare una perdita di funzionalità del G-CSF Pegylato.

1A.3. Messa a punto e validazione di un metodo predittivo delle interazioni tra altre proteine e relativi recettori.

I processi biologici sono governati da interazioni tra porzioni di superfici delle molecole partners che sono complementari per aspetti sterici, elettrostatici e idrofobici. In effetti l'effetto idrofobico è considerato essere uno dei principali determinante nella stabilità e funzione delle proteine e dei loro complessi. Per l'attuazione della fase operative 1A.3 si completeranno gli studi di SAS delineati nella sezione precedente con la ricerca di patches idrofobici sulla superficie molecolare A tale scopo è stato installato e testato sui computer del CRS4 il software QUILT [11, 12] che prevede la possibilità di analizzare le traiettorie risultanti da simulazioni di dinamica molecolare per la ricerca dei patches idrofobici temporalmente più stabili sulla superficie molecolare.

1A.4. Ottimizzazione di sistemi strutturali/enzimatici implicati nella produzione di biocomposti.

Per l'analisi e modellazione strutturale e funzionale di complessi proteici sono usate tecniche di Docking proteina-proteina nelle quali si ricerca la conformazione in cui le superfici delle due proteine sono complementari attraverso l'uso di algoritmi di ricerca grid-based o coarse-grained ed utilizzando metodi Monte-Carlo per esplorare lo spazio di tutte le possibili mutue conformazioni del complesso. Uno dei codici più noti ed efficienti di questo tipo, è RosettaDock

[13] che come parte di questo lavoro è stato installato e testato sul cluster di PC per calcoli ad alta prestazione del CRS4. Il sistema MTGASE/GCSF è stato studiato per cercare di mostrare le specificità di interazione tra l'enzima e la proteina tramite studi computazionali di docking su grande scala.

I risultati di tali analisi sono stati presentati alla conferenza ISMB/ECCB Luglio 2007 nel poster "Flexible Protein-protein docking : MTgase/G-CSF" [14].

Q	SAS G-CSF	SAS G-CSF/receptor	Δ SAS	Secondary Structure
11 ^o	44.18	29.92	14.26	Near to beginning of α 1 helix
20 ^o	15.66	0.82	14.84	α 1 helix
25	22.51	22.51	0	α 1 helix
32	22.34	22.34	0	α 1 helix
67	25.53	25.53	0	loop near disulphide bridge C64-C74
70 ^o	80.74	67.95	12.79	loop near disulphide bridge C64-C74
77	19.32	19.32	0	α 3 helix
86	12.81	12.81	0	α 3 helix
90	48.65	48.65	0	near to end of α 3 helix
107	14.34	14.34	0	α 4 helix
119 ^o	59.5	22.23	37.27	α 4 helix
120 ^o	14.62	9.33	5.29	α 4 helix
131	69.65	69.65	0	loop between α 4 and α 5 helix
134	32.57	32.57	0	loop between α 4 and α 5 helix
145	28.71	28.71	0	α 5 helix
158	19.5	19.5	0	α 5 helix
173	72.04	72.04	0	C-terminal

Tabella 1: Valori di Solvent Accessible Surface (SAS per le 17 glutammine presenti nel G-CSF nel caso di struttura isolate (SAS G-CSF) e nel caso di struttura complessata con il recettore (SAS G-CSF/RECEPTOR). Le glutammine che mostrano variazioni di SAS sono indicate con un cerchio rosso.

Bibliografia

1. Kashiwagi, T., et al., *Crystal structure of microbial transglutaminase from Streptovorticillium mobaraense*. J Biol Chem, 2002. **277**(46): p. 44252-44260.
2. Ohtsuka, T., et al., *Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors*. Biosci Biotechnol Biochem, 2000. **64**(12): p. 2608-13.
3. Tanaka, T., N. Kamiya, and T. Nagamune, *Peptidyl linkers for protein heterodimerization catalyzed by microbial transglutaminase*. Bioconjug Chem, 2004. **15**(3): p. 491-497.
4. Sato, H., et al., *Further studies on the site-specific protein modification by microbial transglutaminase*. Bioconjug Chem, 2001. **12**(5): p. 701-710.
5. Seitz, A., et al., *Enzymatic cross-linking of purple membranes catalyzed by bacterial transglutaminase*. Biomacromolecules, 2001. **2**(1): p. 233-8.
6. Lee, D.S., et al., *Identification of the epsilon-(gamma-glutamyl)lysine cross-linking sites in alpha-lactalbumin polymerized by mammalian and microbial transglutaminases*. J Agric Food Chem, 2002. **50**(25): p. 7412-9.

7. Nieuwenhuizen, W.F., et al., *Modification of glutamine and lysine residues in holo and apo alpha-lactalbumin with microbial transglutaminase*. J Agric Food Chem, 2003. **51**(24): p. 7132-9.
8. Hill, C., T. Osslund, and D. Eisenberg, *The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors*. Proc.Natl.Acad.Sci.USA, 1993. **90**: p. 5167-5171.
9. Koradi, R., M. Billeter, and K. Wuthrich, *MOLMOL: a program for display and analysis of macromolecular structures*. J Mol Graph, 1996. **14**(1): p. 51-5, 29-32.
10. Gautier, R., A.C. Camproux, and P. Tuffery, *SCit: web tools for protein side chain conformation analysis*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W508-11.
11. Lijnzaad, P., H.J. Berendsen, and P. Argos, *A method for detecting hydrophobic patches on protein surfaces*. Proteins, 1996. **26**(2): p. 192-203.
12. Lijnzaad, P., H.J. Berendsen, and P. Argos, *Hydrophobic patches on the surfaces of protein structures*. Proteins, 1996. **25**(3): p. 389-97.
13. Gray, J.J., et al., *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*. J Mol Biol, 2003. **331**(1): p. 281-299.
14. Valentini, M., et al., *Flexible protein-protein docking: MTgase/G-CSF*, in *ISMB/ECCB 07*. 2007: Wien.