

DISS. ETH NO. 18153

# **DNA-Encoded Chemical Libraries**

A dissertation submitted to the

ETH Zurich

For the degree of

Doctor of Sciences

Presented by

**Luca Mannocci**

Dott. Chim. Università degli Studi di Pisa

Born September 7, 1979

Citizen of Pisa (Italy)

Accepted on the recommendation of

Prof. Dr. Dario Neri, examiner

Prof. Dr. Karl-Heinz Altmann, co-examiner

Zurich, 2009

## 1. SUMMARY

The isolation of small organic molecules capable of specific binding to biological targets is a central problem in chemistry, biology and pharmaceutical sciences. Consequently, there is a considerable interest in the development of powerful and convenient technologies for the construction of large sets (“libraries”) of chemical compounds and of novel screening methodologies for the identification of binding molecules. DNA-encoded chemical libraries represent an innovative approach to the construction and screening of libraries of unprecedented dimension and quality. Such libraries consist of a collection of chemical compounds, each individually coupled to a distinctive DNA fragment which serves as identification bar code. DNA-encoded chemical libraries can be “panned” on a target protein immobilized on a solid support. Typically, high-throughput sequencing reveals the different composition of the library before and after panning, thus allowing the identification of binding molecules to the target protein of interest. In this respect, DNA-encoded chemical libraries bear a logical similarity to phage display libraries of proteins and peptides, in which the binding specifically displayed on the tip of the phage surface (“phenotype”) is physically linked to the gene coding for the polypeptide (“genotype”).

In the first part of this thesis, I present a general strategy for the stepwise coupling of coding DNA fragments to nascent organic molecules following individual reaction steps, as well as the implementation of high-throughput sequencing for the identification and relative quantification of library members. The methodology was exemplified in the construction of a DNA-encoded chemical library containing 4’000 compounds (DEL4000) covalently attached to unique DNA-fragments serving as amplifiable identification bar-codes. We have also assessed the relative composition of the new library and its functionality by performing selection experiments on sepharose resin coated with streptavidin. This study has led to the identification of novel chemical compounds with submicromolar dissociation constants towards streptavidin. Moreover we have found that selections can conveniently be decoded using a recently described high throughput DNA sequencing technology (termed “454 technology”), originally developed for genome sequencing,

In a second selection experiment binding molecules to polyclonal human IgG were identified. I could show that, upon coupling to resin, these compounds could be used for the affinity purification of human IgG from culture supernatants.

Furthermore we also carried out a selection against the catalytic domain of human matrix metalloproteinase 3 (MMP3). Matrix metalloproteinases (MMPs) are zinc-dependent proteases which are involved in tissue remodelling of a variety of physiological and pathological processes. The selection facilitated the identification of a binding compound with dissociation constant in the low  $\mu\text{M}$  range.

Encouraged by these results we investigated methodologies for the construction of very large DNA-encoded chemical libraries, featuring the stepwise addition of at least three independent sets of chemical moieties onto an initial scaffold, using suitable orthogonal chemical reactions and/or protecting strategies, followed by the sequential addition of the corresponding DNA codes. Our experiments have shown that it should be possible to construct DNA-encoded libraries containing over one million individual chemical compounds. The construction of such libraries is currently in progress.

## RIASSUNTO

L'isolamento di sostanze organiche in grado di interagire specificamente con target biologici è un problema cruciale sia in chimica, biologia che in campo farmaceutico. Di conseguenza sta emergendo un crescente interesse in sviluppare nuove rapide ed efficienti tecnologie per la costruzione e lo screening di ampie raccolte ("librerie") di composti organici. Un'innovativa e brillante soluzione a questo problema è rappresentato dalle librerie chimiche "DNA-encoded". Essenzialmente queste tecniche prevedono la costruzione di librerie di composti organici in cui ciascun membro è covalentemente coniugato a uno specifico frammento di DNA che "codifica" inequivocabilmente la sua natura. Per tanto, la selezione di composti d'interesse con specifiche attività biologiche ("screening") utilizzando librerie "DNA-encoded" può essere facilmente eseguita incubando ad esempio la libreria con l'opportuno target biologico immobilizzato su un supporto solido. Dopo aver escluso i composti non-leganti, attraverso appropriati lavaggi del supporto, le moderne tecniche di "high-throughput sequencing" permettono di sequenziare gli specifici codici di DNA, di determinare la composizione della libreria prima e dopo la selezione e di conseguenza di identificare i composti effettivamente interagenti con il target biologico d'interesse. Da questo punto di vista le librerie chimiche "DNA-encoded" racchiudono un'intrinseca analogia con le librerie di fagi utilizzate nella "phage display", in cui ciascuna proteina o peptide ("fenotipo") è fisicamente associata al corrispondente gene codificante ("genotipo").

Nella prima parte di questa Tesi è descritta una strategia generale per la costruzione di librerie chimiche "DNA-encoded" e l'implementazione delle tecniche di "high-throughput sequencing" per l'identificazione e la relativa quantificazione dei membri della libreria prima e dopo la selezione. La metodologia è qui esemplificata nella costruzione di una libreria chimica "DNA-encoded" contenente 4'000 composti (DEL4000) ciascuno univocamente identificato tramite specifici DNA-oligonucleotidi covalentemente coniugati. In seguito è stata determinata la relativa composizione della libreria e la sua funzionalità eseguendo esperimenti di selezione impiegando streptavidina immobilizzata su resina di sefariosio. Questi studi hanno condotto all'identificazione di nuovi composti chimici con costanti di dissociazione sub-

micromolare verso la streptavidina e hanno inoltre dimostrato che le tecniche di “high-throughput sequencing” (denominate “tecnologie 454”), originariamente sviluppate per la sequenziazione del genoma, possono essere efficacemente impiegate nel processo di decodifica delle selezioni.

In una seconda selezione utilizzando DEL4000 sono stati identificati composti specifici per “polyclonal human IgG”. E’ stato quindi dimostrato che tali composti, in seguito a immobilizzazione su resina cromatografica, possono essere utilizzati nella purificazione per affinità di IgG umani da supernatanti derivanti da colture cellulari. Infine è stata eseguita una selezione per l’identificazione di nuovi composti specifici per il dominio catalitico del “human matrix metalloproteinase 3” (MMP3). Le “matrix metalloproteinases” (MMPs) sono una famiglia di proteasi zinco-dipendenti coinvolte nel rimodellamento del tessuto in una varietà di processi fisiologici e patologici. La selezione ha permesso l’identificazione di un composto con costante di dissociazione micromolare.

Incoraggiati da questi risultati, abbiamo deciso di approfondire le ricerche per la costruzione di una libreria chimica “DNA-encoded” di dimensioni superiori prevedendo la congiunzione sequenziale di almeno tre serie indipendenti di composti, utilizzando reazioni chimiche ortogonali e/o strategie di protezione/deprotezione di gruppi funzionali, seguita dall’introduzione di corrispondenti codici di DNA. E’ stata quindi dimostrata la possibilità di costruire una libreria chimica “DNA-encoded” contenente oltre un milione di composti. La costruzione di questa libreria (DEL10e6) è attualmente in corso.