

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Sede Amministrativa del Dottorato di Ricerca

XXI CICLO DEL DOTTORATO DI RICERCA IN NANOTECNOLOGIE

Engineering of coiled-coil protein scaffolds as innovative tools for biosensing applications

(Settore scientifico-disciplinare: Anatomia Comparata e Citologia – BIO/06)

Dottorando:

Tiziano Gaiotto

Coordinatore del collegio dei docenti:

Chiar.mo prof. Maurizio Fermeglia

Università degli Studi di Trieste

Supervisore-tutore:

Chiar.mo prof. Roberto Marzari

Università degli Studi di Trieste

Co-tutore:

Chiar.mo prof. Daniele Sblattero

Università degli Studi del Piemonte Orientale

ANNO ACCADEMICO 2007/2008

Engineering of coiled-coil protein scaffolds as innovative tools for biosensing applications

Abstract	pag. 4
Riassunto	pag. 5
Introduction	pag. 9
<i>1. Biosensors</i>	pag. 9
1.1 - Introduction to biosensors	
1.2 - Nanotechnology meets biosensors	
1.3 - From life science to proteomics: the nanobiotechnology	
1.4 - Nanobiotechnological applications	
1.4.1 - Nanoparticles	
1.4.2 - Nanosensors	
1.4.3 - Nanosurfaces	
1.5 - Micro and nanoarrays	
1.6 - <i>In situ</i> methods for protein arraying	
1.6.1 - Protein <i>in situ</i> arrays (PISA)	
1.6.2 - Nucleic acid programmable protein array (NAPPA)	
1.6.3 - Puromycin-capture from mRNA arrays	
1.6.4 - DNA array to protein array (DAPA)	
<i>2. Protein scaffolds</i>	pag. 20
2.1 - What are scaffolds?	
2.1.1 - Scaffolds with α -helical domains	
2.1.2 - Scaffolds with β -sheet domains	
2.1.3 - Scaffolds with irregular secondary structures	
2.2 - Coiled-coil domains	
2.2.1 - The structure	
2.2.2 - The hydrophobic core	
2.2.3 - Electrostatic interaction	
2.2.4 - Number of helices, orientation and oligomerization	
<i>3. E/K: a de novo designed coiled-coil</i>	pag. 30
3.1 - Introduction	

3.2 - The structure	
3.3 - Characterization of dimeric E/K-coiled-coil	
3.4 - E/K applications	
4. <i>The analysis of protein interactions</i>	pag. 35
4.1 - The detection of protein-protein interactions: an overview	
4.2 - The two-hybrid system	
4.3 - Bacterial two-hybrid systems	
4.3.1 - Complementation of enzyme fragments: the protein complementation assay (PCA)	
4.3.2 - The β -lactamase-based protein complement assay	
4.3.3 - Selection of scaffold libraries by use of PCA	
4.4 - The display systems	
4.4.1 - The phage display technology	
4.5 - Applications of polypeptide libraries	
Aim of the research	pag. 48
Results	pag. 49
1. <i>The E/K coiled-coil domain: in vivo validation</i>	pag. 49
1.1 - The protein complementation assay to validate E/K dimer	
1.2 - Setup of PCA: p α and p ω expression vectors	
1.3 - Cloning and validation of wild type E and K-coils	
1.4 - The PCA of wild type E/K coiled-coil domain	
2. <i>The wild type E/K domain as a scaffold</i>	pag. 56
2.1 - Engineering of a coiled-coil domain	
2.2 - Construction of E and K mutant coils	
2.3 - Validation of E and K mutant coils	
3. <i>Improving the wild type E/K dimer: the K random coils</i>	pag. 60
3.1 - Introduction	
3.2 - Construction of the p ω -K random library	
3.3 - The primers design and construction of p ω -Kran library	
3.4 - p ω -Kran library characterization	
3.5 - Selection of the p ω -Kran library by PCA	

3.5.1 - PCA-screening of p ω -Kran library	
3.5.2 - PCA-validation of selected p α -Ewt/p ω -Kran clones	
3.6 - Analysis of the expression of selected Kran coils	
4. <i>Improving the wild type E/K dimer: the K rational design coils</i>	pag. 68
4.1 - Construction of the p ω -K rational design library	
4.2 - The primers design and construction of p ω -Krd library	
4.3 - p ω -Krd library characterization	
4.4 - Selection of the p ω -Krd library by PCA	
4.4.1 - PCA-screening of p ω -Krd library	
4.4.2 - PCA-validation of selected p α -Ewt/ p ω -Krd clones	
4.5 - Analysis of the expression of selected Krd coils	
5. <i>In vitro validation of new Kran and Krd coils</i>	pag. 76
5.1 - Setup and vector engineering	
5.2 - Validation of the system	
6. <i>The single chain E-K</i>	pag. 81
6.1 - Setup of a biosensor unit: the single chain E-K	
6.2 - The pB vector	
6.3 - Construction of pB scE-K library	
6.4 - pB scE-K library characterization	
6.5 - Selection of the phage display library	
6.5.1 - Description of selection approach	
6.5.2 - Selection of the pB scE-K library by magnetic beads	
6.6 - Validation of selection strategy	
Discussion	pag.90
Materials and methods	pag. 95
References	pag. 109

Abstract

A new generation of protein scaffolds is becoming a valid alternative tool to recombinant antibodies of biotechnological, medical and pharmaceutical applications, where strong affinity and specificity are required. They share with antibodies important features (target affinity and specificity), but they have also some improvements (smaller size of molecule, tolerance to modification of the framework and the recognition site restricted to few residues), that can be exploited for biosensing application in nanotechnological platforms. Nanotechnology has been played an increasingly important role in the development of biosensors, improving the intrinsic features of biodevices.

In this thesis work, we analyzed the coiled-coil domain, a widely spread dimerization domain shared by several protein scaffolds, and involved in protein-protein interaction in both eukaryotic and prokaryotic cells. The analysis of the coiled-coil structure allows a *de novo* design of new peptides, namely E and K, that can dimerize as a E/K coiled-coil system: the dimerization feature and the stability of the interaction makes this system an ideal platform to build up functional and customizable biosensors.

A characterization of the E/K interaction was performed by using the protein complementation assay (PCA), a useful biological method to investigate the interaction between protein partners. With this *in vivo* method, we corroborate the interaction features determinate with circular dichroism, and we demonstrated that E and K coils effectively represent a protein scaffold, able to tolerate amino acid substitutions without altering its main structure. In addition, we create two libraries of K mutant coils, randomizing the peptide sequence, and with PCA we selected new K binders (Kran 5.17 and Krd F8) that showed a comparable interaction activity with the E-coil in preliminary *in vitro* tests.

In the last part of this work, we generate a library of a new scaffold molecule (the single chain E-K) capable to bind small molecules as a single protein product containing both domains. Using the phage display selection system, we isolated scsE-K that can bind our analyte (the caffeine) with high specificity. This new molecules can be a powerful tool for analytical and biomedical applications.

Riassunto

INTRODUZIONE.

La richiesta di dispositivi bioanalitici è marcatamente incrementata negli ultimi anni da parte di diversi settori, quali ad esempio quello biomedicale (diagnostico e terapeutico), commerciale, per il controllo ambientale e per la biodifesa. La richiesta è costantemente quella di specifici sensori che permettano misure e analisi in maniera attendibile, veloce ed efficiente.

Le nanotecnologie hanno dato un grande contributo per sviluppo di biosensori sempre più efficaci e in scala miniaturizzata. L'utilizzo di nuovi materiali e la scoperta di nuove proprietà fisico-chimiche, che vengono evidenziate in nano-scala, hanno permesso lo sviluppo di applicazioni biosensoristiche sempre più sensibili e configurabili, basate su piattaforme nanotecnologiche: nanoparticelle magnetiche, *quantum dots*, *nanocantilever* e *nanoarray* hanno trovato un ampio utilizzo, soprattutto dopo il raggiungimento di importanti obiettivi come il sequenziamento del genoma umano e lo studio dell'intero profilo proteico (proteoma).

Alcune di queste piattaforme biosensoristiche possono essere costruite utilizzando le proprietà d'interazione fra proteine, ad esempio fra un anticorpo e un suo antigene. Gli anticorpi ingegnerizzati sono utilizzati in molte applicazioni, ma hanno delle limitazioni riconducibili alla loro struttura e alle loro proprietà fisico-chimiche. Una nuova generazione di *scaffold* proteici ha suscitato un grande interesse per un utilizzo come elemento sensibile in applicazioni biotecnologiche e biosensoristiche. Queste proteine, spesso di origine naturale e coinvolte in attività biologiche, possiedono la stessa capacità di legame ad alta affinità e specificità con una molecola complementare, ma possiedono caratteristiche vantaggiose quali le minori dimensioni, il sito di riconoscimento ristretto a pochi residui amminoacidici e la tolleranza alle modifiche. La caratteristica fondamentale degli *scaffold* è la loro capacità di tollerare modifiche alla sequenza, non alterando il *folding* della proteina risultante. Le modifiche che possono essere apportate sono molte, e spesso volte a cambiare la specificità di legame. Considerando la struttura, si possono apportare specifiche modifiche: per esempio, possono essere rese accessibili delle regioni precise a degli interattori, formando vere e proprie tasche. Un aspetto fondamentale da considerare è la struttura dello *scaffold*, che può presentare solo domini ad α -elica (ad esempio, *affibodies* e domini a cerniere di leucine), β -foglietto (lipocaline e *knottins*) o una combinazione dei due (defensine e domini PDZ).

I domini ad α -elica sono molto comuni nei sistemi di interazione proteina-proteina. Un dominio molto comune, coinvolto in molti processi cellulari di dimerizzazione, viene chiamato *coiled-coil*. Sono stati inizialmente studiati domini dimerici, formati da due eliche destrorse che avvolgendosi formano un'unica elica sinistrorsa (es. α -cheratina); successivamente, analisi approfondite ai raggi X hanno permesso di scoprire domini *coiled-coil* multimerici. Le sequenze delle proteine che formano l'interazione hanno mostrato una presenza costante di determinati residui amminoacidici, fondamentali per l'instaurarsi dell'interazione stessa. Residui idrofobici formano un *core* apolare, fondamentale per la stabilità e per la specificità di legame, mentre residui carichi o polari formano una regione carica esterna che guida la dimerizzazione stessa e permette di isolare il *core* apolare, aumentando la forza d'interazione.

La stessa natura aminoacidica, in termini di carica posseduta, è stata mantenuta per la progettazione *ex novo* di un dimero *coiled-coil*, chiamato E/K. E' un dominio costituito da due peptidi distinti (E e K), formati da un'unica eptade ripetuta 5 volte; ognuno dei peptidi ha un'eptade con una sequenza propria. L'interazione fra i due peptidi è stata caratterizzata a fondo: l'elevata stabilità e affinità del sistema permettono di ipotizzare l'uso del dominio E/K come *scaffold* proteico per applicazioni biosensoristiche. E' possibile pensare di incrementare la qualità d'interazione fra i due peptidi, permettendo un miglioramento dell'efficienza del biosensore; inoltre è possibile creare delle nuove strutture che permettono il legame di piccole molecole, strutture che possono essere anch'esse applicate su dei biosensori.

RISULTATI E DISCUSSIONE.

Il dominio E/K possiede delle caratteristiche adatte per un suo utilizzo in biosensoristica come *scaffold* molecolare: è piccolo (costituito da 35 aminoacidi per elica), stabile e con grande affinità e specificità di interazione. Le sue proprietà d'interazione sono state determinate mediante metodi fisico-chimici, come il dicroismo circolare, ma nella prima parte di questo lavoro l'interazione è stata testata mediante un sistema *in vivo*, ovvero il saggio di complementazione proteica (PCA, *protein complementation assay*).

Sviluppato come sistema di studio delle interazioni proteina-proteina, è stato successivamente migliorato per l'analisi del proteoma: la tecnica si basa sull'utilizzo di una proteina *reporter* divisa in due domini funzionali, e prodotti in fusione con degli interattori. Il folding della proteina *reporter* dai suoi frammenti è catalizzata dal legame delle proteine interattrici, e l'attività dell'enzima integro è ricostituita. Ogni proteina *reporter* ha un suo specifico fenotipo rilevabile, ad esempio l'emissione di fluorescenza (GFP) o la crescita in terreno con antibiotico (β -lattamasi). L'enzima *reporter* utilizzato per questo lavoro è stato la β -lattamasi, un enzima che idrolizza l'anello β -lattamico di molte penicilline. E' stato diviso in due frammenti, chiamati α e ω : entrambi i frammenti sono stati espressi in batteri *E. coli* in fusione con uno dei peptidi del dominio *coiled-coil* (E con il frammento α , K con il frammento ω). Usando gli opportuni controlli, dall'analisi della capacità di crescita dei batteri in presenza di ampicillina è stata determinata la qualità d'interazione fra i peptidi: questi esperimenti preliminari hanno dimostrato come l'interazione fra E e K possa avvenire anche in un sistema *in vivo*, confermando i dati ottenuti dal dicroismo circolare, e come la PCA basata sulla β -lattamasi sia un metodo affidabile per valutare le interazioni proteina-proteina.

Lo sviluppo di un sistema basato sul dimero E/K, ma più evoluto dal punto di vista della stabilità, è stato un altro obiettivo di questo lavoro. Migliorare la stabilità di questo dominio proteico significa modificare opportunamente la struttura, senza intaccare il folding dell'intero complesso: pertanto abbiamo verificato che i peptidi E e K possano essere considerati uno *scaffold* proteico, ovvero che tollera cambiamenti della sequenza aminoacidica. Modifiche alla struttura *wild type* sono state apportate sostituendo i residui di entrambe le catene in specifiche posizioni nell'eptade centrale con l'aminoacido glicina: i risultati ottenuti tramite esperimenti di PCA hanno mostrato una capacità del sistema E/K di tollerare le modifiche, mantenendo il folding delle nuove catene, e la qualità della loro interazione. Nonostante le sostituzioni effettuate siano state le stesse su entrambe le catene, l'elica K ha dimostrato di avere una struttura meno sensibile alle mutazioni, esibendo una capacità di interazione anche superiore alla coppia di *coiled-coil wild type*.

Verificata la capacità di tolleranza migliore dell'elica K rispetto all'elica E, abbiamo cercato di migliorare il sistema d'interazione, modificandone puntualmente la struttura. Esiste un numero buon numero di lavori in letteratura che caratterizza proteine *coiled-coil* e le modifica per migliorarne l'interazione con diversi scopi: la stessa strategia è stata adottata anche in questa tesi per il dimero E/K. La PCA è stata utilizzata per vagliare due diverse librerie di interattori con lo scopo di identificare delle varianti con una migliore stabilità e affinità di interazione. Sono state adottate due strategie di mutagenesi, e visti i risultati preliminari tutte rivolte a modificare la sequenza del peptide K: la modifica è stata effettuata mediante tecniche di mutagenesi, ed è stata applicata una completa randomizzazione dell'eptade centrale (*K random coils*, o Kran) e una mutagenesi razionale dell'intera elica (*K rational design*, o Krd), considerando solo i residui fondamentali per la dimerizzazione del peptide e un numero ristretto di amminoacidi per le possibili sostituzioni. Da queste librerie sono state isolate delle varianti di K che in PCA hanno mostrato capacità di interazione con il peptide Ewt superiori all'elica Kwt; tutti i peptidi selezionati hanno permesso ai batteri testati in PCA di sopravvivere a concentrazioni molto elevate di antibiotico, in alcuni casi proibitive per il dimero originale.

I peptidi isolati sono stati successivamente prodotti come proteine ricombinanti e purificati, al fine di saggiare *in vitro*, mediante dei test ELISA, la capacità di interazione dimostrata *in vivo*. I dati raccolti con il saggio di complementazione proteica sono stati confermati, seppure con qualche differenza dovuta probabilmente alle diverse condizioni sperimentali. I test hanno comunque dimostrato che i peptidi K mutanti isolati sono in grado di legare con alta affinità il peptide E *wild type*, ed è stata anche dimostrata l'alta specificità dell'interazione, per l'assenza di formazione di omodimeri.

Nell'ultima parte di questo lavoro sono state costruite delle nuove strutture *scaffold*, chiamate *single chain* E-K (scE-K), in grado di legare delle piccole molecole. Si tratta di un'unica molecola contenente il peptide E, un *linker* e il peptide K; due cisteine alle estremità del polipeptide permettono la formazione di un ponte disolfuro e di mantenere chiusa l'intera struttura. Da una libreria di *single-chain*, che presentavano le eptadi centrali randomizzate in entrambi i peptidi, sono stati selezionati dei scsE-K in grado di legare con una buona affinità delle molecole di caffeina. La tecnica di selezione utilizzata per isolare i *single chain* reattivi contro l'analita utilizzato, la caffeina, è stata il *phage display*: sviluppato anch'esso come metodo di analisi delle proteine, risulta essere molto versatile per la possibilità di selezionare peptidi con opportune caratteristiche, tra cui la capacità di legame. I cloni isolati hanno mostrato una valida reattività contro la caffeina, e ulteriori test con il mentone (una molecola con una struttura simile alla caffeina) hanno dimostrato la selettività dei *single chain* isolati nel riconoscere la molecola di caffeina.

Pertanto, le tecniche di selezione utilizzate si sono dimostrate valide ed efficienti per valutare le interazioni molecolari, ed isolare interattori con caratteristiche vantaggiose per un utilizzo in biosensoristica. Sia la struttura E/K *single chain*, sia gli interattori mutanti K possono essere utilizzati come elementi molecolari sensibili in nano-piattaforme come i *nanoarray*, e la loro produzione può essere facilmente effettuata tramite nuove tecniche di co-trascrizione/traduzione proteica *in situ*, come le tecniche PISA (*protein in situ array*), NAPPA (*nucleic acid programmable protein array*) e DAPA (*DNA array to protein array*).

List of abbreviations

AD, activation domain	ORF, open reading frame
AP, alkaline phosphatase	PCA, protein complementation assay
AP-1, activator protein 1	PCR, polymerase chain reaction
APTES, aminopropyltriethoxysilane	PISA, protein in situ array
BD, binding domain	PPI, protein-protein interaction
BRET, bioluminescence resonance energy transfer	PSA, prostate specific antigen
BTH, bacterial two hybrid	QD, quantum dot
DAPA, DNA array to protein array	RT, room temperature
DARPs, designed ankyrin repeat proteins	SPR, surface plasmon resonance
dsDNA, double strand DNA	ssDNA, single strand DNA
ELISA, enzyme linked immunosorbent assay	STM, scanning tunneling microscopy
Ewt, E-coil wild type	THS, two-hybrid system
Fc, fragment crystallizable region	tTG, tissue transglutaminase
FCS, fluorescent correlation spectroscopy	wt, wild type
FRET, fluorescent resonance energy transfer	
GFP, green fluorescent protein,	G, glycine (Gly)
GST, glutathione S-transferase	P, proline (Pro)
HRP, horseradish peroxidase	A, alanine (Ala)
Ig, immunoglobulin	V, valine (Val)
IPTG, isopropyl β -D-1-thiogalactopyranoside	L, leucine (Leu)
Kran, K-coil random	I, isoleucine (Ile)
Krd, K-coil rational design	M, methionine (Met)
Kwt, K-coil wild type	C, cysteine (Cys)
LMWP, low molecular weight proteins	F, phenylalanine (Phe)
MALDI-TOF, matrix assisted laser desorption ionization-time of flight	Y, tyrosine (Tyr)
mDHFR, mouse dihydrofolate reductase	W, tryptophan (Trp)
MS, mass spectroscopy	H, histidine (His)
MW, molecular weights	K, lysine (Lys)
NAPPA, nucleic acid programmable protein array	R, arginine (Arg)
Ni-NTA, nickel nitrilo-triacetic acid	Q, glutamine (Gln)
NP, nanoparticle	N, asparagine (Asn)
O/N, over night	E, glutamic acid (Glu)
	D, aspartic acid (Asp)
	S, serine (Ser)
	T, threonine (Thr)

Introduction

Chapter 1

Biosensors

1.1 Introduction to biosensors

The demand for bioanalytical devices has increased markedly in recent years, driven by need in many commercial and research sectors for specific sensors that are capable of rapid, reliable measurements [1]. Biosensors are powerful tools that are able to provide accurate detection of target analytes; they have always been important in the fields of biomedical diagnostics (e.g., bacteria, virus tissue components), drug discovery, biodefense, and environmental monitoring (e.g., air, soil, water) [2], to more straightforward analyses such as fermentation, process monitoring, and quality control of foods and beverages [3].

Traditionally, a biosensor is defined as an analytical device which converts a biological response into an electrical signal: biological material is directly associated with a physicochemical transducer. Only certain substances, that are supposed to be interactors of biological part, will also result in an optical or electrical signal of physical transducer. Today, the term “biosensor” is used to cover sensors used to determine the concentration of molecules and other biological parameters, even if the sensor does not utilize a biological component.

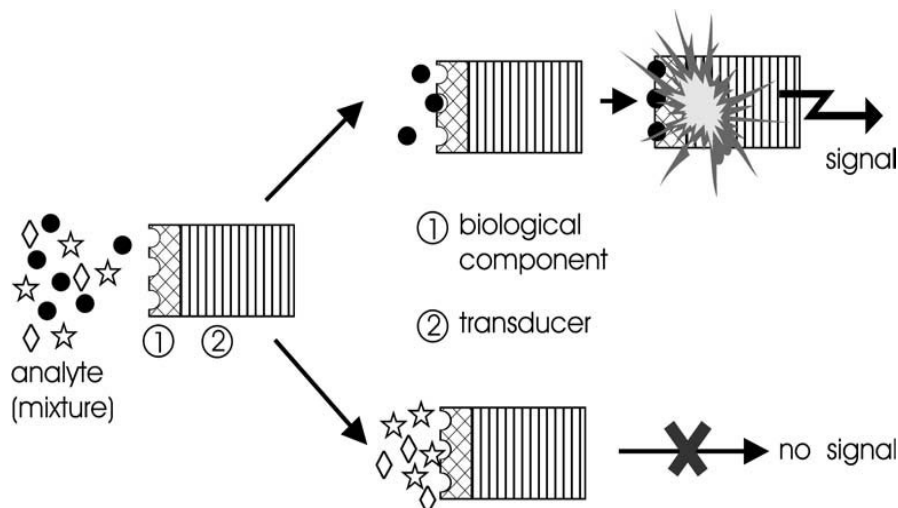


Fig. 1.1 – Principle of function of a biosensor. The interaction between a compound in a mixture and the biological component of the sensor is converted in a physical signal. Molecules which are not able to interact will not produce signal [4].

The biosensing research is multidisciplinary, involving biochemistry, bioreactor science, physical chemistry, electrochemistry and software engineering: the increasing interest for accurate and fast identification systems allows the production of suitable platforms to detect any type of substances.

A successful biosensor must possess at least some of the following beneficial features:

1. the biological unit onto the biosensor must be highly specific and stable under normal storage conditions;
2. the reaction should be independent of such physical parameters as pH and temperature;
3. each component of the biosensor should have minimal pre-treatment;
4. the results should be accurate, reproducible and easy to understand;
5. if the biosensor is used for invasive monitoring in biomedical analysis, the probe must be tiny and constructed with biocompatible materials;
6. the biosensor should be cheap, small and easy to hand; it may be used by completely non-skilled operators, on applications of personal care, health, food and environmental monitoring, therefore, the complete devices must be “user-friendly”

Biosensors were initially developed for clinical diagnosis, in the determination of blood glucose or the protein concentration in biological samples [5]. In recent years, biosensors have developed increasing sensitivity, which is an important feature, e.g., to detect tumor biomarkers that appear at very low levels in early stages, especially when the patient is still asymptomatic, or the disease is still in latent- or silent-stage. Some of these devices have been also miniaturized in order to obtain a chip-based sensory system.

Nanotechnology has played an increasingly important role in the development of biosensors, improving the intrinsic features of biodevices: the use of new material, their sub-micron size and their important new physical proprieties at nano-level, has revolutionized the field of chemical and biological analysis, enabling rapid and accurate analysis [6-10].

1.2 Nanotechnology meets biosensors

Nanotechnology is the creation, the control and the utilization of material, devices and systems through the control of nanoliter (1 billionth of a liter) amounts of fluids as well as control of matter on a nanometer (1 billionth of a meter) length scale.

An important aspect of nanotechnology is that if the size of the whole system decreases, the number of physical phenomena become pronounced, as a mechanical effects. In fact, the study of suitable materials for nanotechnological applications is one of the most important fields of research for these new devices: various nanostructures have been investigated to determine their proprieties and possible applications in biosensors, and these structures include nanotubes, nanofibers magnetic and semiconductor nanoparticles, plasmonic metallic nanostructures and thin films [8, 11].

Some nanotechnological improvements permit the miniaturization of several systems, materials and devices with advantages for several aims, from diagnostics to high throughput systems, resulting from the new proprieties of nano-size materials [10]: 1) Increment of available surface due to increased surface-volume, 2) shortened transport time of molecules by short distance, 3) the decreased needed sample to analyze allow a high linear flow rate acquirement, 4) in the case of arrays, almost infinite expansion of detected spot.

1.3 From life science to proteomics: the nanobiotechnology

Given the inherent nanoscale functions of the biological components of living cells, it was inevitable that nanotechnology would be applied to the life science (nanobiotechnology). Analyses of signaling pathways by nanotechnology techniques may provide new insights into disease process, indentifying more efficiently new biomarkers and understanding the mechanisms of action of drugs [12].

According to the current data, the human genome consists of about 25000 genes, and it has been completely sequenced. Now, the attention of researches is focused on the product of these genes: the proteins and the enzymes, that determine the cellular function and the architecture. These set of proteins, specifically expressed by a cell under defined conditions, is the “proteome”, for analogy with the genome [13]. In fact, the genome is a static information because it is a constant entity of cells. On the contrary, the proteome is a dynamic information, because it depends on the gene products and their interactions.

The proteome is a complex molecular system, and it is larger then proteome: there are more proteins than genes, and all it depends on differential splicing of respective mRNA, temporal and functional regulation of gene expression, formation of multimeric complexes, and post-translational modifications. The definition of the whole human-proteome will be the challenge of the next years, and it will be a more difficult task than the definition of the genome.

To this purpose, proteomics is playing an important role in the study of methodologies allowing the correlation of the huge quantity of data about genes and the analysis of a larger number of components, e.g. with throughput systems [14]. Some nanotechnologies are refining the application of proteomics: they are providing nanofabricated devices that are small, sensitive and less expensive (in terms of time of analysis and money) than the standard protocols,

allowing to overcome problems respect to sensitivity, speed and the requirement of large amounts of starting material [8, 10, 15].

1.4 Nanobiotechnological applications

Most of the nanobiotechnology application, for example, in molecular diagnostics falls under the wide category of biochips and microarrays, more correctly termed nanochips and nanoarrays. Currently, the advances in this field open new opportunities and provide other powerful tools in genomics, proteomics, molecular diagnostics and high throughput screenings [12].

To enhance the signal-to-noise ratio and to increase the features of biodevices, several types of nanotechnologies were directly incorporated to biosensing applications. The nanotechnologies applied to biodevices can be categorized into three potential fields (some examples for each category) [10]:

- 1) technologies that involve nanoparticles (NP): gold NP, magnetic NP, bio barcodes, and quantum dots (QDs);
- 2) technologies for developing nanosensors: PEBBLE sensors, electronic biosensors, SPR sensors, nanocantilevers, nanowires and nanotubes;
- 3) technologies for nanochips and nanoarrays, and nanostructured surfaces for analysis via mass spectroscopy.

In the last years, a lot of different variants of nanodevices have been developed especially for medical purposes (viral detections [16, 17], detection of bacteria [18, 19], detection of disease genes [20, 21], imaging [22]), but there are many potential applications in other fields.

1.4.1 Nanoparticles

The nanoparticles are the most used technologies at nano-size for diagnostics and several biomedical applications. The particles are sized between 1 and 100 nanometers, and they are often composed by material that at nano scale show unique and advantageous physical proprieties (e.g., a magnetic or electronic behavior) [11, 23, 24].

The Quantum Dots (QDs) are semi-conducting crystals that show a particular optical propriety that is absent in the bulk material: they emit fluorescent light when excited by a light source such a laser. The size and the shape of QDs can be easily controlled by parameters used in the synthesis [25]; the emitted spectra is wide, from ultraviolet to red, and it can be tuned by the control of the size and the compositions of the nanoparticles. The fluorescent features make the QDs powerful tools for several aims, from the cancer diagnosis and therapy [20] to in vivo molecular and cellular imaging [26-29].

The surface can be functionalized with antibodies or aptamers, creating fluorescent-immunoassay and –immunoistochemistry probes. Their use was demonstrated for long-term multicolor imaging and to label live cells [30]. The surface is completely configurable by using avidin: the interaction with biotinylated molecules permits a stable conjugation of the nanoparticles with a wide range of ligands. Other experiments demonstrated that QDs – probes can be used in an unique *in vitro* assay to simultaneously detect several different toxins from a little amount of sample [31].

Metal NPs, in particular gold, have been used to improve the performances of biosensing: they were initially used for first Point-of-Care diagnostics, in pregnancy tests [11]. As for QDs, the surface can be functionalized with antibodies, DNA and other ligands: recently, the electric and magnetic features were exploited for electrical detection in arrays, by using nanoparticle coated with DNA [32], and for a rapid screening of telomerase activity in tumor samples [33]. Gold-NP can be detected with a lot of different techniques (optical absorption, Surface Plasmon Resonance, Raman scattering, atomic and magnetic force), make them good labels for sensors [15, 34].

A new powerful detection system, based on metal nanoparticles, was developed to detect specific molecules with attomolar sensitivity: the bio-barcode assay [35]. It utilizes two types of metal particles: a magnetic microparticle coated with the recognition agent (oligonucleotides to detect nucleic acids, antibodies to detect proteins) and a gold nanoparticle coated with the same recognition agent and with hundred of DNA copies (the “bar-codes”) that encode for the target of interest. When the particles both recognize the target molecule, a magnetic field is used to separate the complexes from the solution. The DNA copies are detected with high sensitivity detection systems, by using amplification cycles of PCR or with modified-DNA chips. The lowest limits of detection achieved are zeptomolar (10^{-21}) for DNA and attomolar (10^{-18}) for proteins.

Regarding the high sensitivity of the system, this assay can be used to reveal the low levels of disease biomarkers in early stages, when it is still asymptomatic. It was used to detect the presence of prostate specific antigen (PSA) in 1 ul biological sample [36], and the amyloid-derived diffusible ligand in cerebral spinal fluids samples [37].

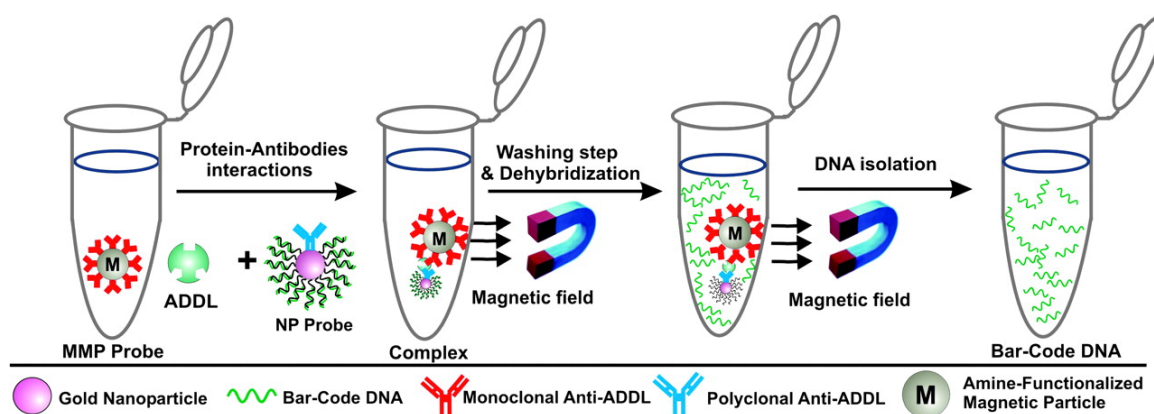


Fig. 1.2 – The bio-barcode assay. The assay uses a magnetic particle functionalized with antibodies, that recognize a protein (ADDL, amyloid- β -derived diffusible ligands, in the picture). The same protein is recognized by a nanoparticle that is functionalized with the antibodies and labeled with dsDNA. After washing steps and the dehybridization, the specific barcode-DNA is released [37].

1.4.1 Nanosensors

The most promising devices in biosensors are the nanowires. These sensors possess the ability to transduce a chemical binding in an electric signal: the conductance of the nanowire change after a binding event (by protein, DNA or other ligands). The big advantage of this nanobiosensor is the directly label-free detection signal of interaction. As nanoparticles, the

nanowires can be coated with DNA [38] or proteins [39] to detect DNA-DNA, protein-DNA or protein-protein interactions.

The Biodetect[®] system exploits the features of both nanowires and nanoparticles [40]. Two wires are separated on an array, but they can be linked by a molecule of DNA which interacts at the ends with the wires: then, the DNA is covered by metallic nanoparticles. The DNA molecules form a “electrical bridge” between the wires, and the passage of the current is easily detected. Several chips have been developed, with specific wires for different DNA target, from the same or a different organism.

Also nanotubes can be used in electrochemical biosensors, functionalized with proteins or nucleic acids [41]. Binding events change the electric and chemical proprieties of the nanotubes: attomolar limits of detection were achieved [42].

A nanocantilever is a small beams, similar to those used in the atomic force microscopy to screen biological samples for screening of genetic sequences. A label-free sensor, the nanocantilever amplifies the absorption process of a molecule on the surface, resulting in a measurable deflection due to changes in surface stress. The surface does not possess an intrinsic ability to bind any molecule: wanted binding events occur only if the surface is coated with adapt layers, antibodies, DNA probes or peptides. These nanodevices have been used to detect PSA biomarker [43], biowarfare agents [44] and the envelope protein of HIV [45]. The necessity to configure each cantilever allows the construction of multiplexed biochips and arrays to detect a lot of proteins simultaneously.

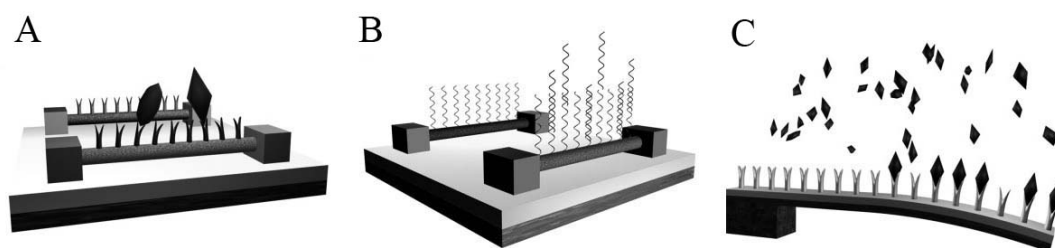


Fig. 1.3 – Three examples of biosensors for the detection of small molecules: (A) nanowires functionalized with antibodies, (B) nanowires functionalized with ssDNA, (C) nanocantilever functionalized with antibodies Adapted from [11].

1.4.3 Nanosurfaces

The matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) is the most used method for protein identification, exploiting organic matrixes: with the mass spectroscopy we can analyze the profile of protein, peptides, polymers and macromolecules in biological samples [46], and pesticides in water and food [47]. The suitable matrixes must be specific for each type of sample, and several kind of metal and metal oxides particle have been used [48]. The MALDI-TOF was used to analyze the low molecular weight proteins (LMWPs), which include at low concentration proteolytic fragments and potential disease biomarker [49].

The human plasma is composed by more than 300000 biomolecular species, with almost 8 orders of magnitude relative abundances; the presence of proteins at high concentration and with high molecular weight is a limitation of this approach, because they interfere with the

absorption of the protein at low concentration. For the analysis of these biomolecular species, an enrichment method was developed, which permit subsequent analysis with mass spectroscopy [50]. A silica-based nanoporous surfaces were constructed to capture LMWPs: with a size-exclusion based method, peptides can be selected, isolated and harvested by using pores with a estimated size about 7 nm. A similar type of nanoporous material (silica and glass beads) was used to selectively separate proteins from biological samples, analyzing their proprieties as disease biomarkers [51].

1.5 Micro and nanoarrays

Most of the nanotechnologies have been applied to construct useful devices for biomedical applications, generally for the health care: the extraordinary multiplexing capability of such devices allows to envision powerful solution also for Point-of-Care home-test, not only for biomedical aims but also for every field in which a cheaper, fast and sensitive test is needed. Regarding this purpose, the micro- and nano-arrays are the most promising tools.

The nanoarrays are the next stage in the evolution of the miniaturization of microarrays, but the basics are the same: to immobilize molecules at defined locations on a surface with micrometer or nanometer resolution. Currently, the molecular resolution on the surfaces of the arrays is higher than that of the first devices: the first protein array was described by Fodor and his coworkers [52], developed with a light-directed synthesis and composed by 1024 protein spots. New printing mechanisms allow to print 55000 molecules or nanoparticles on 1 cm² area [53].

Basically, the construction methods of micro- or nano-arrays are similar: the molecule (e.g. nucleic acids or proteins) must be immobilized on the surface. The deposition and the absorption is made by selecting adapt materials for the surfaces. The proteins, instead of nucleic acids, are chemically more variable; therefore, the necessity to choose the appropriate substrate for the application is required by users [54, 55] and by picking very small amount of samples (picoliters). The array can be often coated with a thin layer of chemical compounds (gold, glass, silicon), to facilitate the absorption of different inorganic or organic molecular species.

The last findings on array printing permit to obtain devices with resolution of about 100 nanometers or less, and positional accuracy of 1.5 nm [56]. With the positive-charged tip of scanning tunneling microscope (STM), DNA molecules were picked and immobilized on a surface, obtaining arrays with high resolution. Other commons micro patterning techniques are photolithografy, electrospray, ink-jet printing and microcontact printing. The last mentioned technique is very powerful method for surface structuring, useful to deposit unlimited patterns and to print simultaneously different kind of biomolecular species. Currently, the dip-pen nanolithography is the printing technique which deposits molecules with the highest resolution [53].

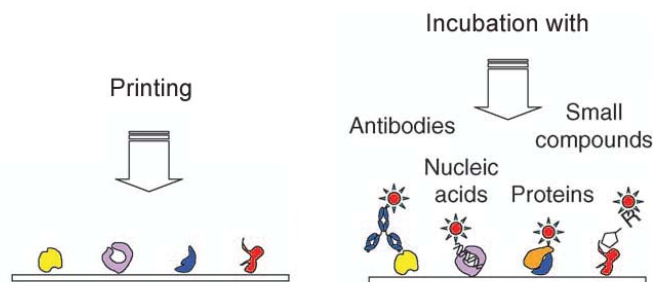


Fig. 1.4 – Schematic representation of an array slide printed with different types of proteins, and common reagents used for the array analysis.

About the detection of interaction onto the arrays, the system is completely configurable, by using fluorescent (fluorochromes Cy3 and Cy5), chemiluminescent (alkaline phosphatase and horseradish peroxidase) and electrochemical readout systems [57]. Fluorescence is preferred for the high sensitivity and dynamic range: the area of the array is excited by a laser beam to fluoresce line-by-line [58]. Recently, some label-free readout systems have also been developed [59], as ViriChip™ System, a virus detection platform based on an optimized atomic force microscope [60]. These platforms do not need secondary reporter systems, which sometimes can interfere with the assay.

1.6 *In situ* methods for protein arraying

A huge quantity of data is now available, after the completion of the human genome sequencing: since proteins rather than DNA carry out the biomolecular functions, there is large of interest to analyze proteins and their entirety, the proteome. The protein array was developed and applied in some fields of applications, such as the characterization of protein-protein interactions [61], the high-throughput screening of cellular processes and pathways [62], the drug discovery research [63], the investigation of potential diagnostic markers in bacterial [64] and autoimmune diseases [65].

Several recent progresses have been developed especially for nucleic acid printing onto the arrays, with lithographic methods, as yet explained in the previous paragraph. For the protein arraying, there is some drawback:

- 1) expression clones must be generated, and the soluble proteins must be expressed and purified, to be subsequently attached onto suitable surfaces. Human or recombinant proteins can be poorly expressed in bacterial system for insolubility, disulphide bonds forming, degradation or toxicity in cellular hosts. Often, proteins sequences must be sub-clones into an expression system, and all processes are time- and cost-intensive.
- 2) proteins, rather than DNA, are more involved in degradation events because they are chemically more heterogeneous: therefore, the long-term storage of printed proteins and the integrity of the arrays may be difficult to maintain in a functional state over long periods of time.

Some improvements have been applied to overcome these drawbacks: protein arrays are rapidly generated in one step by using cell-free lysate as *in situ* production method. The basic

idea is to use cell lysate as a coupled transcription/translation system, that allow the production and direct immobilization of any protein onto a functionalized surface. Currently, three different systems utilize this process: protein *in situ* array (PISA), nucleic acid programmable protein array (NAPPA), puromycin-capture from mRNA arrays and DNA array to protein array (DAPA).

1.6.1 Protein *in situ* array (PISA)

Originally described by He and Taussing [66-69] to generate proteins array from PCR DNA, the method is also known as DiscernArray™ technology [67]. The proteins are generally produced from PCR products, which include T7 promoter and sequences for *in vitro* transcription/translation procedure, included a tag sequence for immobilization (Figure 1.5).

The key feature of the system is the presence of a tag at the C- or N-term of the expressed protein sequence: the tag interact with an immobilization agent, pre-coated onto the surface of the array. If the protein are His-tagged, the surface must be Ni-NTA (nickel nitrilo-triacetic acid) or Ni-beads coated: when the PCR fragments are translated by using rabbit reticulocyte or *E. coli* S30 system, the proteins can specifically and rapidly interact with the Ni-surface.

In recent works, some improved double and triple His tags were used to bind the GFP (green fluorescent protein) and antibodies to pre-treated surfaces [70, 71], but also other interaction-systems can be used (e.g. biotin-streptavidin).

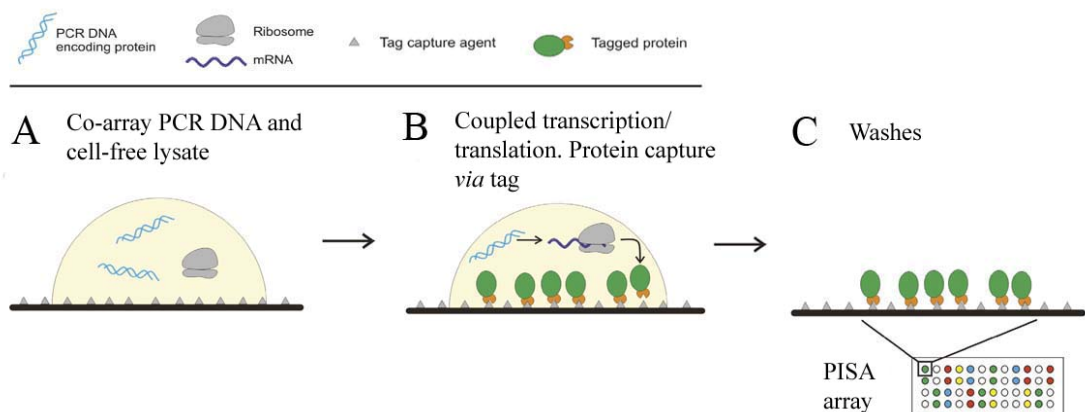


Fig. 1.5 – Schematic of PISA process. (A) The PCR fragments and cell-free lysate are co-arrayed on the slide surface, pre-coated with an immobilizing agent. (B) The proteins are transcribed and translated by the components of the lysate, fused with a tag peptide, that can bind the immobilizing agent. (C) Extensive washes clean the slide: the tag-proteins remain immobilized on the array. Adapted from [69].

1.6.2 Nucleic acid programmable protein array (NAPPA)

In the PISA system, PCR fragments and cell-free lysate are co-arrayed onto the pre-coated surface. The NAPPA method provide for the DNA immobilization onto the array, developing a technology for a directly “DNA-protein” array conversion from the same DNA pre-coated surface [69, 72].

Biotinylated plasmids, encoding for a recombinant protein in fusion with GST, and anti-GST antibodies are printed onto an APTES (aminopropyltrimethoxysilane) slide (figure 1.6). When the cell-free lysate is added, the proteins are translated from the plasmids: the GST tag is recognized by the antibody, and the protein is blocked near the encoding-plasmid, generating a

spot array with a DNA-protein co-localization. In the first experiment a 512 spot slide was generated, and each spot was spaced about 0,9 mm, containing an average of 675 pg of protein [73]. In a recent work, the entire genome of bacterium *Vibrio cholerae* was analyzed: about 3880 open reading frames were captured onto the array [74].

As for PISA, the method does not need an external production, purification and printing of proteins; moreover, it is possible to generate an immobilized “ready-to-use” DNA array, which can be produced, stored and used as required. Therefore, there are some drawbacks:

- the proteins must sub-cloned into the plasmid that carries the GST sequence,
- and plasmids need to be biotinylated, preventing as far as possible the dissociation from the array and the lateral diffusion.
- the produced array is not “pure”, because some molecules are printed: biotinylated plasmids, streptavidin, GST-proteins and antibodies. That may cause lower sensitivity and specificity.

In a recent work it has been demonstrated the directly absorption of DNA (cDNA or PCR fragments) onto amine-coated glass surfaces [72]: the positively charged amine bind the DNA, avoiding the biotinylation step.

The need for antibody as capture system for the expressed proteins was overcome in a system where only the expression vector is arrayed [75]. In this method, the plasmid directs the synthesis of the proteins and at the same time it is the capture agent. An high-affinity binding ($3\text{-}7 \times 10^{-13}$ M) was exploited: the binding between the *E. coli* protein Tus to Ter sequence, a 20 bp sequence involved in replication processes of bacteria. The proteins of interest are synthesized as Tus fusion proteins with a standard coupled transcription/translation, and they interact with Ter sequence, embedded in the vector.

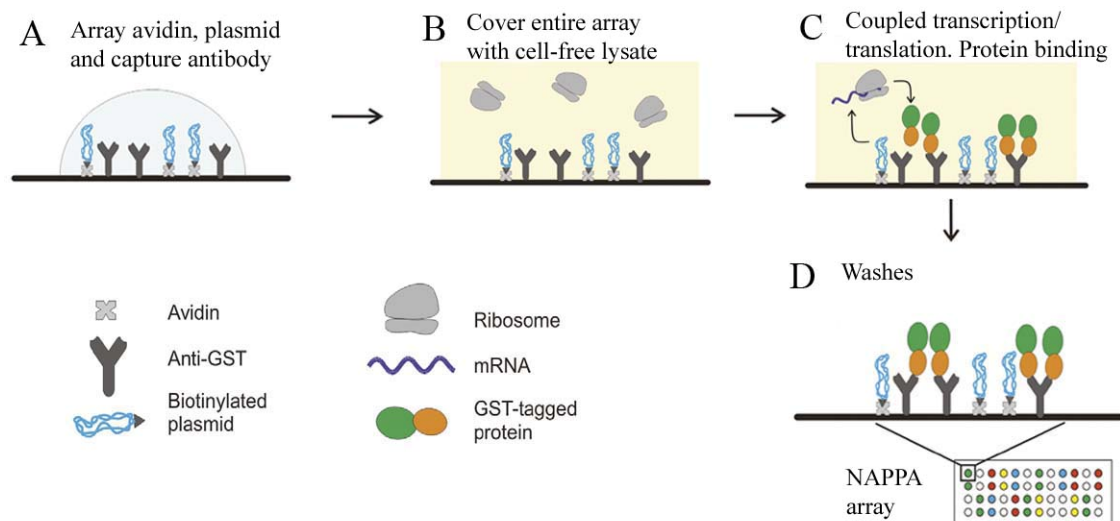


Fig. 1.6 – Schematic of NAPPA process. (A) The surface of the array is pre-coated with avidin, the biotinylated plasmid and the antibody anti-tag (e.g. GST). (B) The array is entirely covered with the cell-free lysate, and (C) the components of lysate transcript and translate the gene embedded into the plasmid. The antibody binds the tag expressed in fusion with the protein, and (D) after extensive washes the array is ready. Adapted from [69].

1.6.3 Puromycin-capture from mRNA array

Tao and Zhu [76] described a new strategy to obtain protein arrays, avoiding the problem of local diffusion of the just produced protein.

The PCR products are retro-transcribed in mRNA, and they are linked with a biotinylated-ssDNA, modified with puromycin. The mRNAs are arrayed onto the surface of streptavidin-coated slides; the ribosomes contained in the cell-free lysate translate them. When the ribosome reaches puromycin, the nascent polypeptide is captured by the puromycin itself; the RNA is removed by using the RNase, leaving a pure-protein array.

1.6.4 DNA array to protein array (DAPA)

Every array fabricated with one of the previously explained techniques can be used only for one experiment: the DAPA system allow the “on-demand” printing of multiple protein arrays from a single DNA array, as required [68].

PCR fragments are covalently immobilized onto a slide that is assembled with a second slide, coated with a tag-capturing substrate. A permeable membrane carrying a cell-free lysate is inserted between the two surfaces: the tagged-proteins are translated from the immobilized DNA, and quickly captured by the second slide. A protein array is created corresponding to the DNA array. The effectiveness of the method was demonstrated by producing 20 times a protein array of His-tagged GFP or single-chain antibodies, on Ni-NTA-coated slides: all proteins retained the protein proprieties (fluorescence for GFP, binding activity for single-chains).

The method must be improved for the drawback of protein diffusion, but the possibility to generate repeated pure-protein arrays from one DNA array template is a considerable advantage.

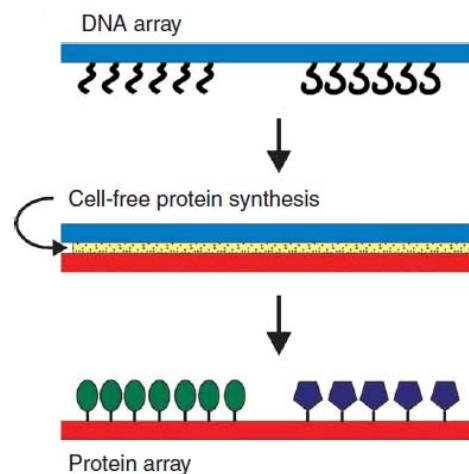


Fig. 1.7 – The DNA is spotted and immobilized onto a glass slide. Simultaneously, a second slide is coated with a protein-capture reagent. The slides are then assembled face-to-face with a membrane soaked with cell-free lysate. The proteins are transcribed and translated, and the newly-synthesized tagged-protein are captured and immobilized onto the second slide Adapted from [68]

Chapter 2

Protein scaffolds

2.1 What are scaffolds?

Currently, the antibodies are the most important and useful binding molecules in a wide range of biotechnological, medical and pharmaceutical applications. This success can be ascribed to the fact that the immune system was the only source for scaffold molecules with a huge diversity, that can interact with other molecules with high specificity and affinity. The antibody engineering allowed to modify their structure, dissecting in minimal binding fragments and enhancing the clinical efficiency for disease treatment [77]. However, immunoglobulins showed some limitations, based on biophysical proprieties and their structure: the whole IgG molecule needs to form disulfide bonds, which can limit the activity of the antibody in non-appropriate conditions. Smaller antibody fragments were used for both therapeutic and non-therapeutic applications, but the fragments tend to aggregate and to degrade. Moreover, some of these antibody domain must be expressed in eukaryotic cells because they fold poorly in bacteria: therefore, additional purification steps are required.

These drawbacks have been overcome by developing a new generation of protein scaffolds. They possess the same features of antibodies (target affinity and specificity), but with some improvements: smaller size of molecule, improved stability and the recognition site restricted to few residues [78, 79]. There are some definitions to generally describe a scaffold, but we can define it as “*a protein framework that can carry altered amino acids or sequence insertion, that confer on protein variants different functions, usually for binding specific targets*” [80]. A fundamental feature of scaffolds is that the native protein folding is not altered by the insertion of residues in the sequence of the protein.

Most of these are involved in natural binding events and have various origins, frameworks and functions. Original peptides can be structural proteins, sub-unit peptides from a multimeric structure and also enzymes [81], but it is also possible to design *de novo* structures [82, 83]. Non-immunoglobulinic scaffolds should improve some features respect to antibodies, also if affinity proprieties for binding targets are comparable: improving features don't compromise target affinity and specificity.

Modified scaffolds can be selected with *in vitro* and *in vivo* technique, isolating new molecules with better affinity against specific ligand [84]. There are different types of natural scaffolds that can be considered to modify, but several issues regarding the proprieties of a candidate protein should be considered before starting selection, in particular the structure, the topology and the size of the scaffold in conjunction with the intended modification strategy will be of concern [85]. These selected binding reagents may become tools which can be used in a wide range of biotechnological and biomedical application, like affinity purification [86],

protein microarray technology [87], bio-imaging [88], enzyme inhibition [89] and targeted drug delivery [90].

On the protein surface there are exposed loops that facilitate the binding activities; the loops link structural elements like α -helices and β -sheets of the scaffold. Not only the loops permit the binding activity with the ligand, but also the α -helices [91] and internal cavities [92]. However loops are the most accessible scaffold region for modification of the sequence, for grafting or integrating an affinity function onto or into the structural framework, without changing the overall structure affinity of the protein and generating a stable folded protein. The following figure shows possible modifications.

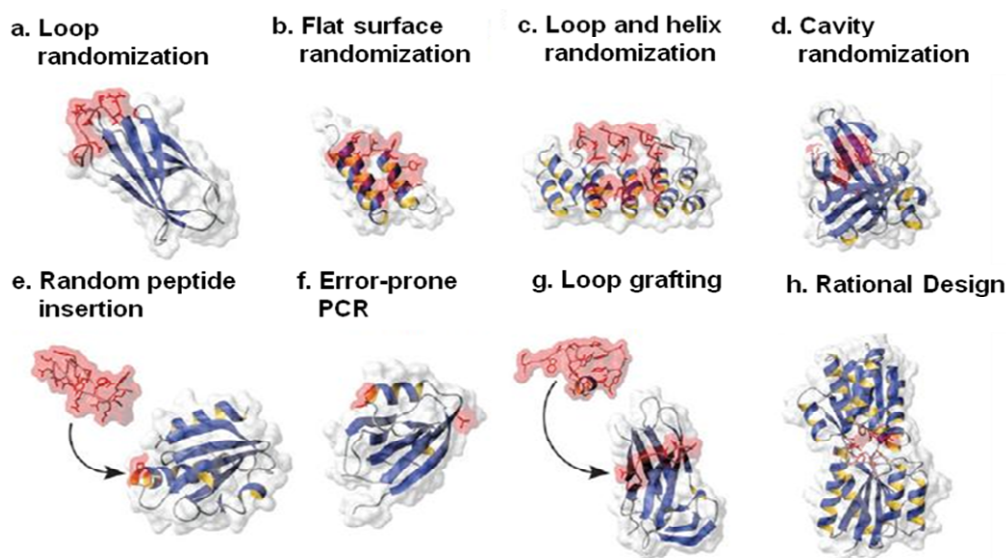


Fig. 1.8 – Engineering strategies to modify the binding site of protein scaffolds. The binding site has been altered, introducing the diversification with suitable strategies, as the randomization of specific domains (a, B, c, d). A random sequence has been inserted into the scaffold (e), or modified by error-prone PCR (f). Moreover, pre-existing binding sites can be grafted into a novel scaffold (g), or the binding site can be *de-novo* designed (h) [84].

It is important to choose carefully the residues to be randomized because some modifications may cause misfolding or instability of the scaffold itself [91].

Protein scaffolds can be characterized by several features, i.g. the presence of disulfide bonds or presence of enzymatic activity. Based on the architecture of their backbone, they can be assigned to one of these groups:

- scaffolds with α -helices;
- scaffolds with β -sheets;
- small scaffolds with few secondary structures or with an irregular architecture of α -helices and β -sheets.

2.1.1 Scaffolds with α -helical domains

Although helical coiled coils belong to the most abundant structural protein motifs, the number of library display scaffolds based on an α -helix architecture is limited compared to the numerous examples for β -sheet frameworks.

Nevertheless in the past years several α -helices scaffolds have been applied for libraries construction and for interaction validation.

AFFIBODIES: first non- β -sheet proteins used for libraries constructions [93], affibodies are small proteins derived from an engineered version (Z domain) of the five stable three α -helices bundle domains from the immunoglobulin Fc-binding region of staphylococcal protein A [94]. These scaffolds are soluble, proteolytically and thermally stable and does not contain disulfide bonds. They have been used for construction of phage libraries and selection against several proteins and enzymes, like *Taq* DNA polymerase, human insulin and human apolipoprotein. The small dimension and the strong protein structure are fundamental proprieties of these molecules for a broad range of biotechnologies applications: affibodies have been described for purification, detection and targeting, also suggesting the therapeutic/diagnostic potential of this protein scaffold [87].

REPEAT PROTEINS: the main members of this category are proteins with ankyrin repeat domains. These domains are one of the most well-studied and common modular protein-protein interaction motifs; the ankyrin repeat is a 33-residue motif in proteins consisting of two alpha helices separated by loops. They appear in bacterial, archaeal, and eukaryotic proteins, but are far more common in eukaryotes.

Designed ankyrin repeat proteins (DARPin) [95] are versatile binding proteins which display great advantages over other binding protein. These repeat-motifs protein offer an advantage for libraries construction because the binding interface depends on the number of randomized repeats. In a recent application libraries of ankyrin repeat proteins have been used to select high affinity inhibitors of aminoglycoside phosphotransferase (3')-IIIa (APH) [96]. The selected proteins showed the great DARPin ability for modulation of intracellular enzyme activity, with implication for therapeutic applications.

GCN4: GCN4, a basic leucine-zipper protein, is the primary regulator of the transcriptional pathway of amino acid synthesis. It binds specifically to the promoters of amino acid biosynthetic genes. As most proteins that bind DNA, GCN4 contain a leucine zipper sequence to bind and activate genes [97].

This is the well-studied coiled-coil protein for roles of trigger sequences in coiled-coil folding. The GCN4 α -helices interact, allowing the protein folding and monomers dimerization. Several libraries has been constructed to evaluating relevance of residues in the sequence.

Other examples of α -helix scaffolds are immunity proteins and the cytochrome b_{562} .

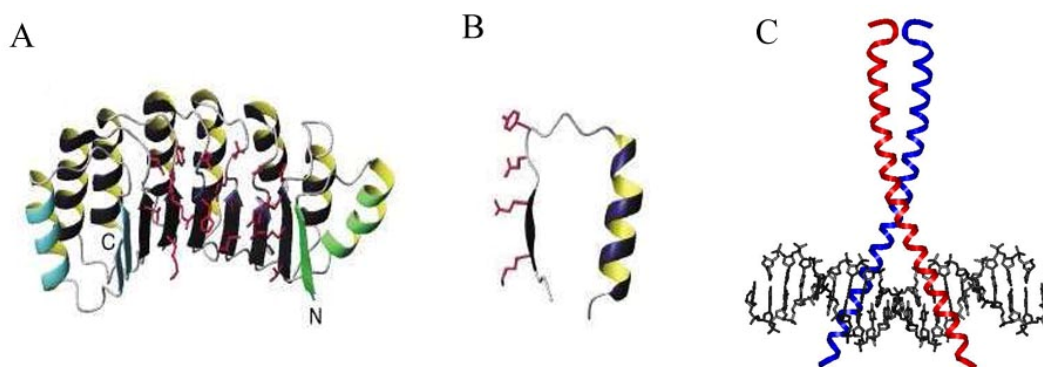


Fig. 1.9 – (A) Molecular model of a DARPin: the side-chains of randomized residues are labeled in red. (B) Molecular model of the single domain. (C) The dimerization region of the transcription factor GCN4 bound to DNA.

2.1.2 Scaffolds with β -sheet domains

Antibodies are indisputably the most characterized β -sheet protein with binding properties. The specificity is mediated through complementary determining regions (CDR), hypervariable loop regions that mediate binding to the antigen.

Immunoglobulins can bind proteins, sugars and diverse small proteins, but there are some limitations for biotechnology applications, for example due to the dimensions of antibodies. These are motivations to find alternative binding molecule.

KNOTTINS: knottins are small disulfide-rich proteins characterized by a special “disulfide through disulfide knot”. The knottin motif is found in some plant-protease inhibitors, toxins and antimicrobial peptides. A typical “knotted topology” unit consist in a small triple-stranded β -sheet with at least 3 disulfide bridges and interspersed variable peptide loops.

Residues in the loops can be mutagenized, and with affinity selection methods a better binding activity can be obtained [98]. In effect the knot proteins are an attractive scaffold for drug design, and they have been considered a good starting point for grafting recognition sites or active sites onto it for the production of the new bioactive molecules [99].

LIPOCALINS: the lipocalins are a family of proteins which transport small hydrophobic molecules such as steroids, bilins, retinoids, and lipids. They are 160- to 180-residues polypeptides that share common tertiary structure architecture, consisting in a beta-barrel of six-or eight anti-parallel β -strands, connected by four hyper-variable loops.

The conical structure encloses an internal ligand binding site which is ideal to randomize residues, allowing the generation of new protein binders with high affinity and specificity. Some new binders have been isolated from a bilin-binding-protein library and selected against fluorescein and digoxigenin with nanomolar K_d [78].

Other examples of β -sheet proteins are neocarzinostatin, T-cell receptors, 10th fibronectin type III domain, CTLA-4 and green fluorescent protein.

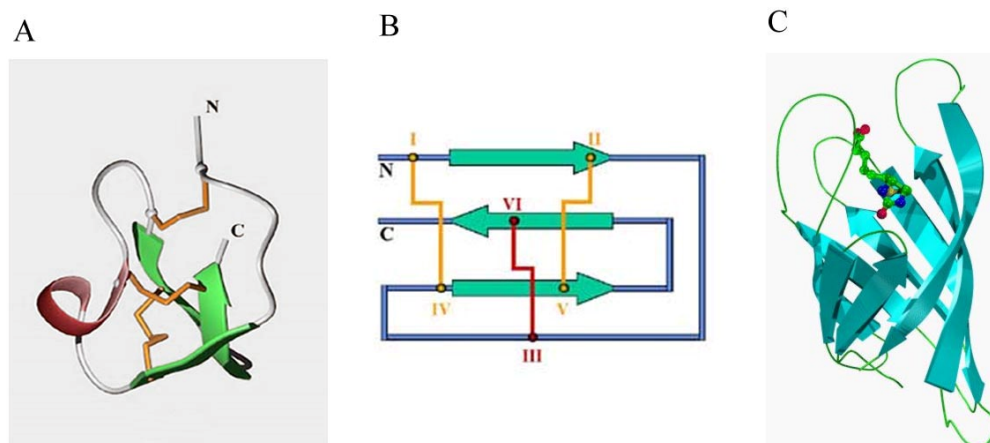


Fig 1.10 – (A) Molecular model of a typical knottin motif: this domain is also termed CSB (Cysteine Stabilized β -sheet) disulfide bridges. (B) A typical “knotted topology” consists in a small triple-stranded β -sheet with 3 disulfide bridges. (C) Molecular model of a lipocalin, with the beta-barrel tertiary structure composed by anti-parallel β -strands.

2.1.3 Scaffolds with irregular secondary structures

In this group there are some proteins that fold in structures with α -helices and β -sheets.

DEFENSINS: defensins are small (15-20 residue) cysteine-rich cationic proteins found in both vertebrates and invertebrates, they play important roles against invading microbes (bacteria, fungi and viruses). The group comprises proteins with an irregular architecture of α -coils and β -sheets, or proteins with few secondary structures.

Insect defensin A has been proposed like a new scaffold for the construction of a mutants peptide library: a total of seven residues were randomized within the two loops of the peptide [100]. The library was screened against tumor necrosis factor A, TNF receptor 1 and TNF receptor 2.

KUNITZ DOMAIN INHIBITORS: they are a group of small and irregular proteins of about 60 amino acids, stabilized by three disulfide bonds. They act as reversible inhibitors of serine proteases.

They possess high stability and, most important for library generation, a loop that can be mutated without destabilizing the structural folding. Therefore, they have been used for modeling specificity to novel protease targets and entered clinical trials [101].

Other scaffolds in this category are plant homeodomain (PHD) finger protein, PDZ domains, charybdotoxins and TEM-1 β -lactamase.

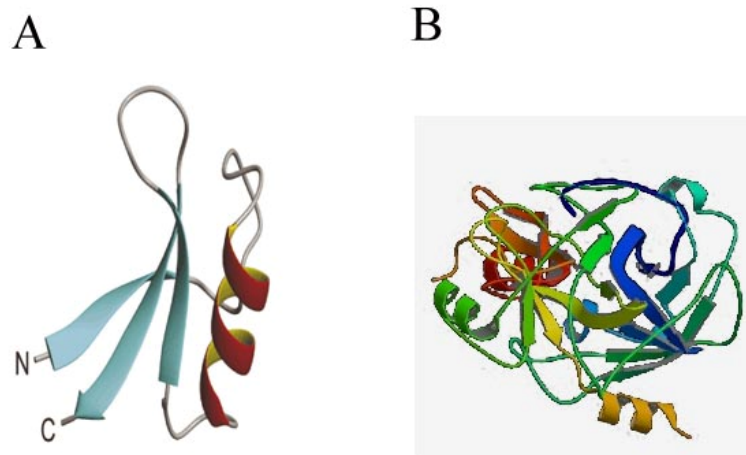


Fig 1.11 – (A) Three-dimensional structure of insect defensin A. (B) Three-dimensional model of the catalytic domain of coagulation factor XI in complex with kunitz protease inhibitor domain of protease nexin II.

2.2 Coiled-coil domains

In the previous paragraph, several protein scaffolds were listed and described, with particular attention for library construction and selection against various molecules. All these scaffolds present α -helices or β -sheets in the secondary polypeptide structure. The presence of these secondary structures is important for protein folding and function (ex. enzyme activity, binding function, etc). The α -helix domain is involved in the folding of another important protein structure: the coiled-coil domain [102].

The coiled-coil is an oligomerization domain found in a wide variety of proteins, including transcription factors, virus proteins and fibrous proteins (keratin, myosin, fibrinogen). Several surveys of genomic data bases suggest that up to 10% of eukaryotic proteins contain sequences predicted to be coiled-coils [103], confirming that they are the most common domain for interaction and dimerization. Despite their abundance, similarity in sequence and simple domain structure, they are highly specific, enabling them to form fine-tuned networks of homo- and heterotypic interactions.

2.2.1 The structure

Francis Crick was the first that proposed in 1952 the possibility of coiled coil architecture for α -keratin [104], suggested in 1951 by Linus Pauling and co-workers [105].

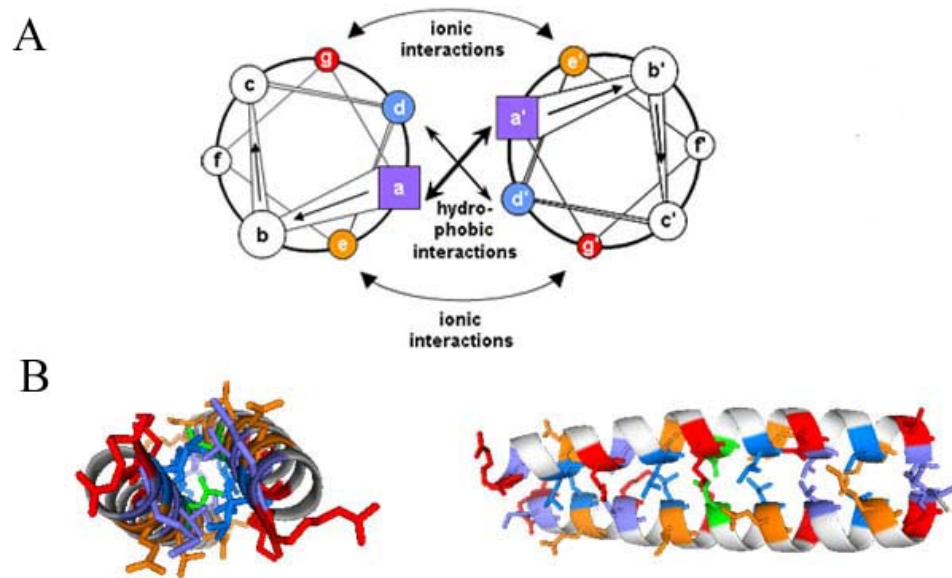


Fig. 1.12 – (A) Helical wheel representation of an heterodimeric coiled-coil domain, viewed from the N-terminus: the fundamental residue for the interaction are highlighted. The hydrophobic residues in positions *a*, *d*, *a'* and *d'* interact, allowing the formation of the apolar core; charged or polar residues in positions *e*, *g*, *e'* and *g'* permit the formation of potential ionic or electrostatic interchain interaction. (B) Two different views of a coiled-coil molecular model. Adapted from [106].

It is a structure composed by two amphipathic and right-handed α -helices that usually adopt a left-handed super-coil. The first protein described with this domain was tropomyosin: the analysis of its sequence revealed a repeated seven amino acid residue pattern called heptad repeats.

The structural features of coiled-coil have been reviewed extensively: their sequence is characterized by heptad, denoted “*a-b-c-d-e-f-g*” and shows a heptad repeat in the chemical nature of side chains. Positions *a* *e* *d* are occupied by hydrophobic residues and are important for interhelical contacts, which contribute to the largest part of the interaction surface. Typically, polar or charged amino acids occupy the *e* and *g* positions and aid in driving stability and specificity by increasing the solubility of the protein. Furthermore, these *e/g* edge position can shield the hydrophobic core from aqueous surroundings. The characteristic hydrophobic repeat is strongly maintained, it is a fundamental feature for structural reasons, for folding and stability of the protein [97].

2.2.2 The hydrophobic core

The analysis of several natural and designed coiled-coils showed the relevance of apolar amino acids in the core: in *de novo* designed peptides, leucine and isoleucine in *a* and *d* position showed better stability [107-109]. *d* position is typically leucine or alanine, contributing to dimerization stability but not to specificity [110]. In contrast the *a* position is more variable (always with apolar residues) and, thus, may contribute to both dimerization stability and specificity. In the context of the GCN4 leucine zipper sequence, ideal dimmers contain leucine

at the *d*, isoleucine at the *a*, and a single *a*-position with asparagine [111, 112] which can form a buried, interhelical hydrogen bond.

The amino acid variability can be explained by side-chain residues orientation relative to the axis formed by two α -helices (for multi-stranded coiled-coils, variability is less clear). The side-chains of *a* residues point towards the external super-coil surface, favoring β -branched residues for tight packing (ex. isoleucine, valine). The side-chains of *d* residues point straight away the hydrophobic interface, favoring unbranched amino acids [113, 114].

The characteristic packing of the hydrophobic core was first described as “knob-into-holes” packing by Crick [104].

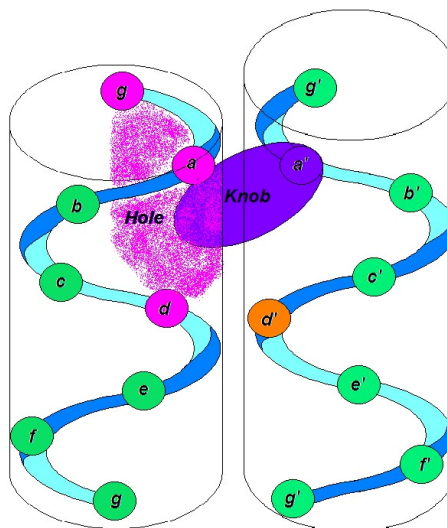


Fig. 1.13 – Schematic representation of “knob-into-hole” model proposed by Crick for two parallel right-handed helices.

The hallmark of coiled-coils is the distinctive packing of amino acids side-chains in the core of the bundle, in which a residue from one helix (knob) packs into a space surrounded by four side-chains of the facing helix. If in the *a/d* position there are big residues, they can not permit correct packing required in the core; if the amino acids are polar or charged, they can destabilize the entire super-coiled structure [108]. Therefore it is possible to modulate the matching of two helices with local destabilization by polar or charged residue, also allowing formation of internal cavities as a scaffold for small molecules [108, 112, 115].

2.2.3 Electrostatic interaction

The hydrophobic effect is accepted as a major driving force in protein interaction and folding of coiled-coils. Ionic interactions and other polar interactions are destabilizing or contribute little to protein folding [116, 117]. It is widely believed that charged and neutral polar side-chains play a significant role in conferring conformational specificity on protein folding and dimerization. It is clear that the charged residues of proteins are arranged so that attractions between opposite charges are favored over interactions between like charged.

Electrostatic interactions contribute in different ways in terms of stability for each considered coiled-coil protein, and this is due to the chemical characteristics of environment where protein folding takes place. Interactions directly exposed to the external surface, and so in

contact with solvent, are weak and they don't add stability to the protein, whereas interactions that take place in a protected surface contribute with a strong effect to structure stability [113].

In the coiled-coils the *e* and *g* position in the heptad repeat lie alongside the hydrophobic core and often contains charged side-chains, which have long been predicted to participate in interhelical interactions. The interactions also permit to isolate the core from the solvent, stabilizing the folding of the entire structure. They also play an important role specifying homo- and hetero-association in native proteins [118]. Some studies have shown that the balance between *g-e'* attraction (*e'* is *e* position of second helix) and repulsion may be critical for specifying homo *versus* hetero-oligomerization [113, 115, 118, 119].

In addition to inter-helical interactions, the coiled-coil can be stabilized by intra-helical interactions between charged residues [118]. The length of side-chains in residue in *e/g* positions is another important feature of the coiled-coil: there is no clear optimal total chain length but some works show the capability to tune dimer stability with specific residues, also with unnatural side-chains [115].

2.2.4 Number of helices, orientation and oligomerization

We described features of coiled-coil with two α -helices that form a super-coil dimer, but natural proteins with coiled-coil domains can be interact as dimers, trimers, tetramers, and also more complicated structures like tubes (tetrabrachion [120]), sheets (Small Conductance Potassium Channel [121]), spirals (flagellin [122]), funnels (bacteriophage Φ 29 [123]) and rings (Apolipoprotein A-I [124]).

The oligomeric state of coiled-coil is determined by packing interactions, and thus depends primarily on the nature of residues in positions *a*, *d*, *e* and *g*. The geometry of side-chains in positions *a* and *d* differs systemically between two-, three-, and four-stranded coiled-coils [108, 125]. In experiments with protein GCN4, Harbury showed that the structure can be switched between two-, three-, four-stranded states by changes of residues in hydrophobic core [126].

Much less is known about the role of residues in positions *e* and *g* in specifying oligomer state, but ionic and polar interactions between residues in these positions are determinant for orientation (parallel or anti-parallel) and preference for homo- and hetero-oligomeric association of helices. Most interactions occur between residues in *g* of one heptad and *e* of the next heptad; interactions between *e/g* residues of the same heptad are often prevented by the shape of the core. Opposite *e/g-e'/g* charges in two facing heptads are often necessary for the formation of stable heterodimeric coiled-coils, but the effect of charged pairs depends on sequence context, as Arndt and colleagues showed [106]. They isolated a stable peptide that contains two repulsive *e/g* interaction from a coiled-coil library. It is clear that coiled-coil stability is provided by number of heptads, hydrophobicity of the core, tightness of the core packing, shielding of the core from solvent and favorable polar and ionic interactions [108, 127].

It is not yet possible to systemically predict the structure that form, or the changes in structure that occur after modification of the protein sequence. Predicting structures is complicated by an enormous diversity of possible topologies. Coiled-coil structures are not determined by simple rules for hydrophobic packing and electrostatic interactions. They can react to various stimuli. Ambroggio and coworkers designed a sequence that form a coiled-coil homotrimer which switch to a zinc finger in the presence of ion metals [128]. Some peptides

form coiled-coil fibers only with specific conditions of pH or temperature, or simply when monomers are mixed [129]. Some softwares are available to predict and identify a coiled-coil structure from a sequence, like Coils, ParCoil, MultiCoil and MarCoil [130-132]. With a protein sequence database and residue probabilities, they analyze the sequence and predict the structure with high sensitivity and specificity. In a recent work, some of these softwares has been compared to predict coiled-coil motifs, evaluating the detecting accuracy of everyone [133].

Chapter 3

E/K: a de novo designed coiled-coil

3.1 Introduction

Coiled-coil peptides have been used in numerous applications, including affinity purification, direct assembly of extracellular receptor domains, creation of miniaturized antibodies, microarrays and peptide libraries [134].

Already in 1993 O'Shea et al. described the design of two peptides that are predominantly unfolded in isolation but, when mixed, they preferentially associate to form a stable, parallel, coiled-coil heterodimer [83]. They chose peptide sequences that would be predicted to give destabilizing electrostatic interactions in the homodimers that would be relieved in the heterodimer.

Based on typical coiled-coil features described above, in 1996 Chao and Tripet. characterized E/K, a *de novo* heterodimeric coiled-coil domain, for biosensor-based affinity and chromatographic applications [114, 135, 136]. They analyzed the structure of several natural coiled-coils, and some requirements were established to permit the correct folding of the new protein [137]:

- to select amino acids that are non-disruptive for the helical structure;
- the placement of non-polar residues in positions *a* and *d* for tightest packing in the hydrophobic core;
- to chose appropriately polar or charged residues in positions *e* and *g* that will promote the helices associations (heterodimeric form) [138];
- the inclusion of polar or charged residues in position *f*, to permit the protein solubilization in aqueous solution;
- the polar or charged residue in position *f* must be opposite charge respect to amino acids in position *e* and *g*, to balance the total charge on individual α -helices;
- to select an optimal length of α -helices to permit the protein production with synthetic means, and to avoid alteration of structure or stability of coiled-coil [139];
- balance between stability of hydrophobic core (that permit homodimer folding) and electrostatic repulsion (that prevent the assembly of homodimers)

Several designed peptides were considered before selecting the E/K as the best coiled-coil protein for biological application.

3.2 The structure

The E/K coiled-coil is a new domain made of two distinct peptides, E and K: each peptide is composed by an unique heptad repeated five times. The heptad sequence of helix E is E-V-S-A-L-E-K (position *g-a-b-c-d-e-f*), and the sequence of helix K is K-V-S-A-L-K-E.

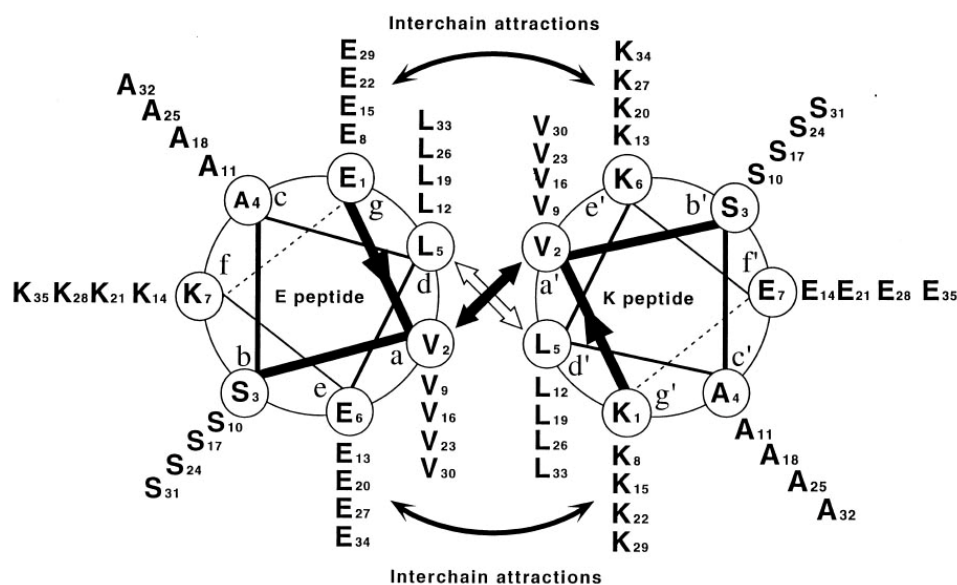


Fig. 1.14 – Helical wheel representation of E/K coiled-coil system, viewed from the N-terminus. Two right-handed helices (E and K) interact as an unique heterodimeric supercoil, and each coil with an unique five-times repeating heptad sequence. The hydrophobic residues valine and leucine interact in the apolar core, and they are mainly responsible for the folding and the stability of the coiled-coil. The charged residues glutamate and lysine contribute to the stability of the structure with interchain interactions, that shield the apolar core from the solvent and specify the heterodimerization of the coils [114].

Neither peptide adopted significant structure on its own, in fact E or K in solution are not correctly folded as a α -helix: circular dichroism analysis show that they exhibit a random coil spectra [114]. In contrast, a solution containing both peptides shows typical spectra of a helical structure: probably, peptides fold and interact as an unique super-coil protein only when they are co-present. They associate as a heterodimeric protein of 70 amino acids, with a molecular mass of 7682 Da. Each peptide is amidated and acetylated at the C and N termini, respectively, to increase the propensity to helix-forming and to reduce the likelihood of proteolysis.

For the hydrophobic core, the choice was valine for position *a* and leucine for position *d*. These β -branched amino acids are commonly found in natural and synthetic coiled-coil peptides; for the packing geometry of side-chains, coiled-coils with these residues tend to favor the dimeric parallel conformation [108, 134, 140]. Also if a core of valine/leucine is less stable than a all-leucine core (or a leucine/isoleucine core), the association in unwanted structure and high order aggregates is avoided [113, 126].

For the *e* and *g* positions, the select residues are glutamate and lysine: these are logical candidates because they frequently occupy these positions in natural coiled-coils [141]. They can also form favorable *e/g'* interhelical salt bridges [140]. In addition, at neutral pH, the repulsive interaction between the charged side-chains discourage formation of homodimeric E/E

and K/K coiled-coils, promoting formation of heterodimeric E/K [108, 140, 142, 143]. Moreover, the interactions also permit to isolate the core from the solvent, stabilizing the folding of the entire structure. Alanine was chosen for position *c* for its high intrinsic helical propensity, serine was introduced into position *b* to increase peptide solubility, and glutamate (K-coil) or lysine (E-coil) was introduced into position *f* to balance the total charge.

3.3 Characterization of dimeric E/K coiled-coil

Several coiled-coil structure exist yet in nature, but they can't be used for biotechnological application because these proteins present some drawbacks:

- low dimerization affinity: the natural α -helices dimerize with an affinity around 10^5 - 10^8 ;
- low dimerization specificity: natural proteins hardly present only one interactor;
- low stability in non-physiological condition: proteins could be degrade.

Main features of *de novo* designed coiled-coil proteins (like E/K) are high stability in denaturant conditions and with different pH values [135, 144]. Stability of this structure was studied with chemical and temperature denaturation experiments, monitored with CD spectroscopy [108, 135].

It is resistant to heat denaturation up to 85°C, and with 5M denaturant conditions. Also combination of 85°C and urea 5M is unable to denature the protein. Currently only guanidine HCl 3,9 M can denaturate the E/K coiled-coil. This high association strength is highlighted by the affinity constant for this dimer: 63 picomolar [145].

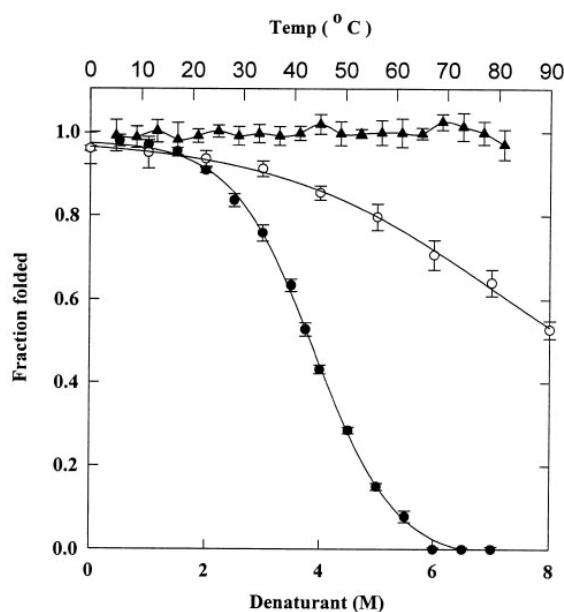


Fig. 1.15 – The graph resumes the denaturation experiments with urea (○), GdnHCl(●) and temperature (▲). Guanidine HCl is the only agent which can completely denaturate the E/K folding [114].

The ability to remain stable was tested also in biological medium, like human serum. E and K single peptides and folded dimer E/K were incubated for 50 h at 37°C with non significant degradation .

3.4 E/K applications

The high stability and fast association and slow dissociation kinetic make E/K an ideal protein for biological applications, that require a stable dimerization and capture domain. In an ideal biosensor applications, one of two α -helices is covalently immobilized on the surface sensor. The second peptide is conjugated with whatever protein, like antibodies, and injected over the sensor surface. The interaction of anchored peptide and fused protein generates an antibody surface for biosensing. In this way the surface is completely configurable with any combination of interactors, choosing the correct “fused” peptide. The sensor surface so assembled can be reused repeatedly as long as it is charged with a sufficient amount of the coiled-coil conjugated.

An example of this application is a particular type of nanoarray, named Nano-Capture Kcoil, developed by Plexera Bioscience. K α -helix can be covalently linked with disulfide bond on the surface, which is coated with a thin gold layer. Then, the array can be charged with the E-coil, conjugated with suitable interactor (protein A, protein G, antibodies, ecc) for the biosensor applications.

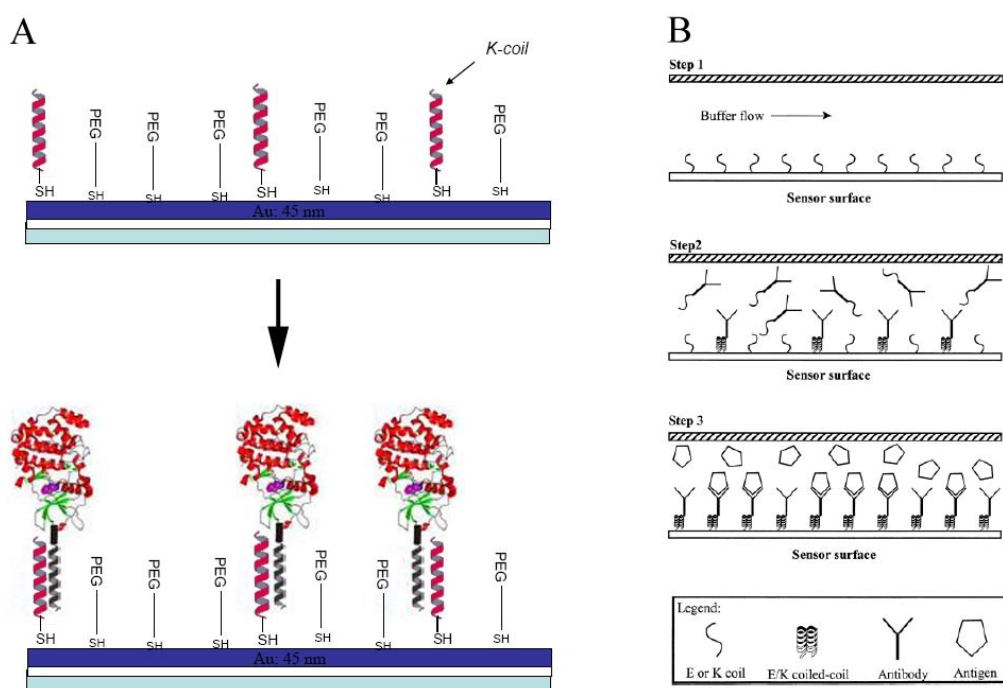


Fig. 1.16 – Examples of E/K applications. (A) K-coils are immobilized with thiol bonds onto a glass slide, pre-coated with a gold layer. E-coil-fusion proteins are injected over the slide surface, and, exploiting the E/K dimerization properties, the slide is functionalized with any desirable protein. (B) The surface can be also functionalized with antibodies fused with the E-coil, for antigen-binding biosensing platforms [114].

The versatility of the E/K heterodimeric coiled-coil as a dimerization and capture domain led us to explore the possibility of using this protein as an affinity purification tag [136]. E/K can be used 1) for detection of recombinant peptides or proteins and 2) for purification by selective dimerization. The first system can be used in assays like western blot: a protein can be produced by bacteria conjugated with a coil (e.g. E) and detected using labeled K-coil. The K-coil can be

labeled with biotin, peroxidase, alkaline phosphatase or other labeling systems. In the second system, the purification is carried out passing the cell-extract through a column containing the immobilized K-coil peptide. E-coil fusion proteins dimerize with K-coil in the column and are retained. After washing, the E-protein is eluted; then with chemical or enzymatic processes the tag can be removed.

The use of E/K fusion proteins is one of most promising techniques, achieving diverse experimental aims, like studies of mutations in multimeric complexes [146]. Bivalent and bispecific ScFv-based molecules can be obtained from *E. coli* encoding two ScFv fragment in as single peptide (coils have the advantage that disulfide bonds are not required for folding). A helix-loop-helix motif has been used like dimerization domain for the assembly of dimeric bispecific miniantibody, replacing the natural immunoglobulin domains [147, 148].

Fusions between coiled coils and other proteins also have been used to study the coiled-coil motif itself, to understand interactions between α -helices [138, 149-151].

Chapter 4

The analysis of protein interactions

4.1 The detection of protein-protein interactions: an overview

Proteins interact each other in a high specific manner, and protein interactions play a important role in many cellular processes, ranging from the formation of cellular structures and enzymatic complexes to the regulation of signaling pathway. Establishing networks of PPIs will eventually lead to a better understanding of cellular pathways and functional associations.

Rapid progress in genome projects generated an extensive list of DNA sequences, coding for proteins with unknown function: only a fraction of gene functions can be inferred from sequences. Therefore it is necessary to develop strategies to define gene function on a large scale. The exploration of protein-protein interaction (PPI) on a large-scale or global level is referred to interactive proteomics, cell-map or interaction proteomics [152].

A first step in defining the function of a gene is to determine the interactions with other proteins. Most of modern research is concerned with how, where and when proteins interact with other proteins (and nucleic acids), defining the interactoma. To understand the mechanisms of protein recognition at the molecular level and to reveal the entire network of protein interactions in the cell, different experimental techniques have been developed. Some methods characterize the single protein interaction, while other need for screening of interactions on a wide-scale genome.

Some of these techniques are:

- *Fluorescent resonance energy transfer (FRET)* [153]: this is a technique for measuring of interaction between two proteins. Two different fluorescent molecules (fluorophores) are genetically fused with two proteins of interest. Upon excitation of the donor, energy is transferred from the donor to the acceptor in a non-radiative manner via dipole-dipole coupling.
- *Bioluminescence resonance energy transfer (BRET)* [154]: one limitation of FRET is the requirement of external source of light to initiate the energy transfer. BRET uses a bioluminescent luciferase (typically the luciferase from *Renilla reniformis*) which catalyze the oxidation of luciferin to emit light.
- *Fluorescent correlation spectroscopy (FCS)* [155]: FCS is a typical technique used in microscopy, to characterize the dynamics of fluorescent species. Light is focused in a sample and a sensor measures fluctuations in fluorescence intensity caused by the diffusion or conformational changes of fluorescently labeled molecules in a small volume using temporal correlation. FCS is used to know average concentrations or kinetic rates.

- *Mass spectroscopy (MS)* [156, 157]: it is a powerful method to studying macromolecular interactions *in vitro*, identifying the chemical composition of a compound based on a mass-to-charge ratio of charged particles. The sample undergoes chemical fragmentation forming ions: they pass through magnetic and electric field and the ratio of charge to mass of the particles is calculated.
- *Surface plasmon resonance (SPR)* [158]: SPR is a technique in which biomolecules capable of binding to specific analytes or ligands are first immobilized on one side of a metallic film. Light is then focused on the opposite side of the film to excite the surface plasmons. The refractive index of light reflecting off this surface is measured. When the immobilized biomolecules are bound by their ligands, an alteration in surface plasmons on the opposite side of the film is created which is directly proportional to the change in bound, or absorbed, mass. Binding is measured by changes in the refractive index.
- *Two-hybrid system (THS)* [159-161]: this method is the most high-throughput interactomic tool, it has accelerated the screening of protein interactions *in vivo*. THS has been developed both in yeast and in bacteria, and it will be exhaustively elucidated in the next chapter.

4.2 The two-hybrid system

The first THS was developed in yeast by Song and Fields, as an assay to study binary protein interactions between proteins already know or those strongly suspected to associate [162]. The system has become the most widely used method to assess both individual PPIs and entire interactomes.

Yeast two-hybrid system is based on the fact many eukaryotic transcription activators have at least two distinct domains: a binding domain (BD) that bind the DNA promoter in the DNA sequence, and an activator domain (AD) that activates transcription. If these domains are physically separated, they can't refold in the active protein, producing the gene product after activation of promoter binding. Song and Fields showed that the transcriptional activity can be restored if both the domains are expressed with proteins which interact each other

A known protein (bait) is fused to the DNA binding domain and a interacting protein (prey) is fused to the activation domain of a transcriptional activator. Both fusion proteins are co-expressed in a appropriate yeast strain that contains sequences for the reporter genes with transcription depending on the association of two domains. Numerous variations of the THS has been developed, including using of several reporter genes and three-hybrid systems for the identification of proteins interactions with DNA or RNA [163].

Only recently the two-hybrid system has been developed in bacteria (BTH). The original idea of Song and Fields is based on modular structure of transcriptional factors, but only few bacterial transcriptional factors can be divided on two separated domains. Although this initial limitation, several systems of BTH have been developed.

4.3 Bacterial two-hybrid systems

The bacterial two-hybrid system presents some advantages comparing the yeast two-hybrid system:

- rapid rate growth of cells;
- high transformation efficiency;
- molecular biology methods more suitable for bacteria than yeast;
- nuclear localization of factors is not required;
- eukaryotic domains does not activate *E. coli* genes;
- lower false-positive interactions.

Although with different advantages, BTH is not better than yeast two-hybrid. The two systems are complementary: each one generate different data which is useful to understand in a better way the interaction under consideration [164].

Bacterial assays can be divided into 4 categories [165]:

- *based on hybrid transcriptional activators* [166]: in *E. coli*, the activation of transcription of numerous genes involves specific interactions between a protein, which bind a specific sequence of DNA (operator), and the C-terminus domain of the α -subunit of RNA polymerase. Dove and his coworkers engineered this system to develop a protein-protein assay. They constructed two chimeric proteins: the operator-binding protein fused with the first interactor (X), and the α -subunit of RNA polymerase fused with second interactor (Y). The chimeric operator-binding protein recognizes the specific activator sequence. When X and Y interact, the α -subunit of RNA polymerase can recruit other subunits, activating the reporter gene transcription. As in the case of the yeast, the transcription machinery of bacteria can be engineered to detect protein-protein interactions.
- *hybrid bacterial repressors as reporters of interaction*: LexA in *E. coli* inhibits the transcription of SOS proteins, which are related to DNA repair and cell division. It recognize a specific DNA sequence. LexA protein has a modular structure, consisting in a DNA-binding domain (with no intrinsic basal activity) and in a dimerization domain [167]. Dimerization domain can be replaced by other dimerization motifs or by supposed interactor proteins, preserving the repressor activity and developing a genetic system to study protein interactions.
- *based on the protein splicing*: protein splicing is a *in vivo* post-translational process, where polypeptides are removed from the entire protein. The process is autocatalytic and with multiple steps: internal fragments of the protein are removed and the flanking sequences are ligated, generating a functional protein [168]. Enhanced green fluorescence protein (EGFP) has been used for this method. One interactor (protein A) is expressed fused with both N-terminus fragments of EGFP and VDE (VMA1-derived endonuclease). An other interactor (protein B) is expressed fused with C-terminus fragments of EGFP and VDE. When protein A and B interact, the VDE fragments move closer, they refold in the active enzyme and the splicing of EGFP starts [169]. The interaction can be measured as fluorescence emitted by EGFP.
- *based on a complementation of enzyme fragments*: this method is thoroughly explained in the next chapter.

4.3.1 Complementation of enzyme fragments: the protein complementation assay (PCA)

In protein fragment complementation assay (PCA) strategy, a reporter protein is rationally dissected into two fragments, and the fragments fused to two proteins that are thought to bind to each other [170]. The refolding of the reporter protein from its fragments is catalyzed by the binding of the supposed interactor proteins, and is detected as reconstitution of enzyme activity. A lot of enzymes can be dissected in two functional domains, but the good candidate should have some features:

- relatively small and monomeric protein;
- information exists on its structure and function;
- assay must be simple in *in vitro* and *in vivo* condition;
- protein over-expression in both eukaryotic and prokaryotic cells;
- the activity of reporter protein must be detectable *in vivo*.

A fundamental feature of protein fragments is that they cannot fold spontaneously: if this occurred, PCA simply not works correctly, because it leads to a false-positive signal.

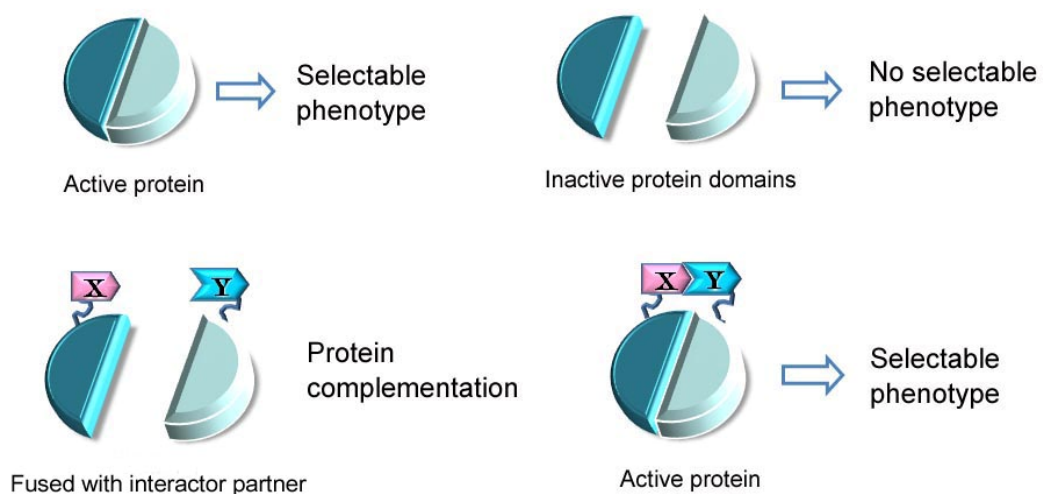


Fig. 1.17 – Representation of PCA strategy. The reporter protein is divided in two inactive domains that are not able to refold spontaneously. The enzyme activity is restored if the domains are expressed fused to interactor proteins: the interaction is necessary for detecting re-assembly of catalytically active enzyme from its fragments.

The design of PCA with the selected reporter starts with the study of three-dimensional structure of the protein [159]. The critical issue is where to cut: it is important to avoid re-folding problems, to preserve integrity of the structure and functionality of the refolded reporter. These problems can be solved studying three-dimensional data about the selected reporter protein, and also using algorithms which can generate ipotetical fragments analyzing the residues of the sequence [171, 172].

Another crucial issue is on which terminus to fuse the proteins of interest to the fragments: in each case, one configuration is favored over any other. Knowing the three-dimensional structure of the enzyme, it is possible to predict how close both fragments should be to assure that the domains fold correctly and have a measurable activity. A peptidic linker can be added

between fragment and interactor to increase the flexibility of entire complex. The linker length can be adjusted to allow only certain types of orientation of fused protein interaction to occur [159].

The molecular interaction are detected directly, as reconstitution of enzyme activity, and not through secondary events such as transcription activation [170]. Moreover, assays based on enzyme complementation can be performed in any cells types of interest or in diverse cellular compartments such as the nucleus, cytoplasm or plasma membrane. Different types of PCA in mammalian cells have been demonstrated [173, 174], and recently also in yeast cells [175].

The bacterial protein complementation assay is a useful method for several aims:

- we can understand how well-know proteins in a biochemical network interact, observing the role in a specific pathway [170].
- we can also detect new interactors from genomic libraries, or characterize a protein with unknown function [176].
- for pharmacological aims, it is possible to screen molecules that promote or disrupt some important interactions, which could be applied to drug discovery [174].
- the system is also useful for high-throughput screening of libraries [138, 149, 177, 178], but the interactors must be confirmed with others different techniques, avoiding problems of false-positive and false-negative results [179].
- additionally, PCA has been used to isolate antibody-antigen interaction and to select antibodies from a library [177, 178, 180, 181].

Several proteins can be used as a reporter for PCA. The following proteins have been used with good results:

- **mouse dihydrofolate reductase (mDHFR):** functional mDHFR confers on bacteria the ability to grow in presence of antibiotic: it can be dissected in two domains, fragment I and fragment II. The two divided fragments can not fold in the active protein. The folding happens only if they are expressed in framework with two interacting partners, allowing to the two halves of mDHFR to come into close contact, thus restoring its enzymatic activity. The *E. coli* DHFR is inhibited by the antibiotic trimethoprim, whereas the mDHFR does not show inhibition [177, 178].
- **adenylate cyclase:** the adenylate cyclase calmodulin-dependent (CyaA) is an essential toxin for the virulence of *B. pertussis*. When the bacteria infect eukaryotic cells, the enzyme is stimulated by endogenous calmodulin, activating the production of cAMP at non-physiological levels [182]. CyaA can be dissected proteolitically in two domains, generating the fragments T25 and T18. Two supposed interactors are expressed like fusion proteins in framework with fragments: the association of the hybrid proteins permit functional complementation of the active enzyme. This method are based on reconstitution of a signal pathway with the involvement of cAMP, in a *cya⁻ E. coil* strain [176].
- **green fluorescent protein (GFP):** isolated from jellyfish *A. Victoria*, GFP fluoresces green when exposed to blue light. It was shown that the protein can be splitted on two fragments: fluorescence of active GFP are achieved when the fragments reassemble, and only if they are fused with interacting peptides. This method was used to test the

strongly association of anti-parallel leucine zippers [183] and to isolate novel substrates for kinases [184].

- **β -lactamase:** the BTH system based on this enzyme is thoroughly explained in the following chapter.

4.3.2 The β -lactamase-based protein complement assay

β -lactamase is an hydrolase secreted in the periplasmic space by gram negative bacteria. It is a protein responsible for the resistance of the β -lactam antibiotics like penicillin, cephalosporins and cephamycins. The β -lactamase cuts β -lactam ring, deactivating it. The β -lactam antibiotics interfere in a specific way with the biosynthesis of peptidoglycan, the major constituent of bacterial cell wall, essential also for the shape and rigidity of the cell [185].

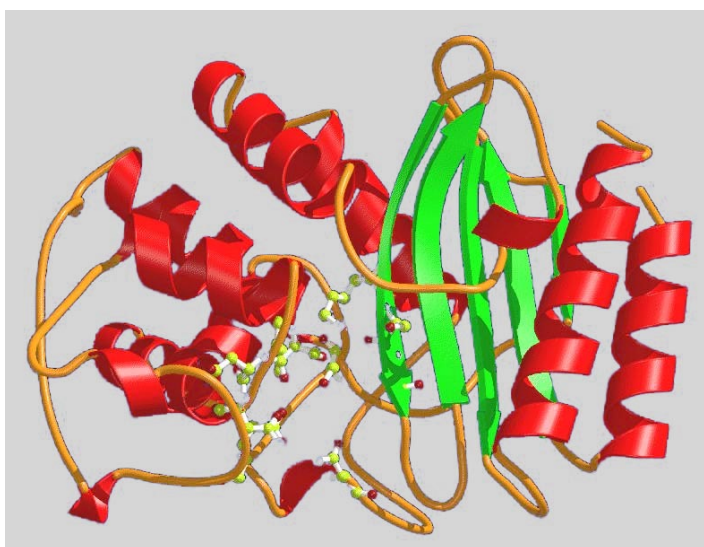


Fig. 1.18 – Three-dimensional model of TEM-1 β -lactamase.

β -lactamase is expressed in several isoforms: TEM-1, from *E. coli*, is the most commonly-encountered enzyme. It is a particular attractive candidate for an assay based on enzyme fragment complementation because it is monomeric, with relatively small size (29 kDa protein), well characterized concerning structure and function. In addition, β -lactamase is expressed successfully in prokaryotic and eukaryotic cells, making this system applicable in both classes of cells, and no orthologs exist in mammalian cells [186, 187]. It is a good reporter protein for PCA because permit the generation of signal as a direct result of interaction: the functional enzyme confers on bacteria the ability to grow in presence of antibiotic. Additionally, it has been used to set up a model systems for fluorescence-based monitoring of β -lactamase activity in living bacterial or eukaryotic cell [188].

The PCA strategy requires the design of optimal dissected fragments, to permit their assembling and folding after that proteins of interest have interact. The TEM-1 β -lactamase can be splitted, generating two polypeptides of 10 kDa and 19 kDa, namely the α -peptide and ω -peptide respectively, that are unable to refold spontaneously. The site used to split the enzyme in these two polypeptides (G195/L196) is not located close to the catalytic site. A point

mutations on the residue 182 (substitution of Met with a Thr) allows a better folding and increase the stability of the reconstituted enzyme [189].

Like other PCAs, the protein complementation assay based on TEM-1 β -lactamase needs of two different vectors for the expression of reporter fragments. The $p\alpha$ vector contains the α -peptide, corresponding to residues 1-195, whereas the $p\omega$ vector contains the ω -peptide, corresponding to residues 196-286. An interactor protein can be cloned with the correct framework in each vector. Only bacteria cells that express interactor proteins can complement β -lactamase fragments and survive in ampicillin plates.

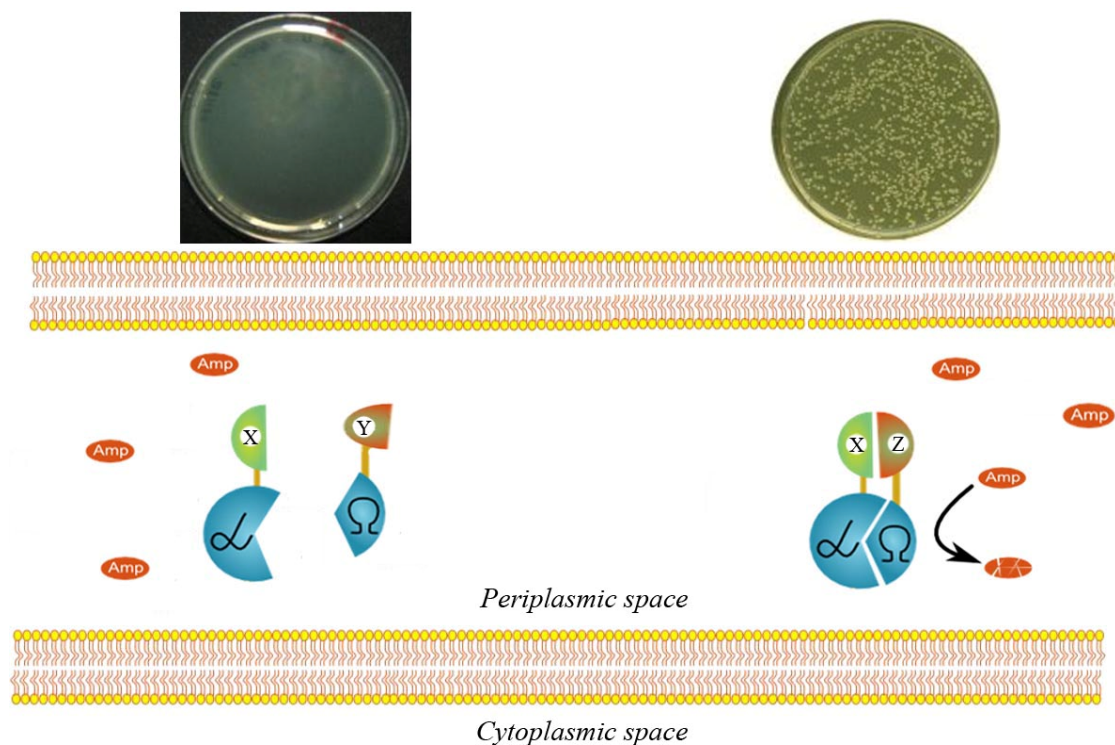


Fig. 1.19 – β -lactamase-based PCA strategy. Two potentially interacting protein are expressed in fusion with the α and ω fragments of the β -lactamase: the fragments complementation upon interaction leads to restored enzyme activity. The ampicillin is hydrolysed by the lactamase and bacterial growth is observed (right). If the in fusion peptides are non-interactor proteins, the lactamase is not complemented and the bacterial growth is impeded by the intact antibiotic (left). The presence of colonies onto the plates is the direct signal of the complementation.

A periplasmic secretory signal of 23 residues is included at the N-terminus of each recombinant protein: so, the fusion polypeptides can translocate and interact into the periplasmic space of the cells [181]. If recombinant proteins lack the periplasmic signal, the complementation takes place only at the cytoplasmic level

The beta-lactamase PCA can be useful for several aims:

- to construct interaction maps of several proteomes;
- to identify interactions between proteins of two different proteomes;
- to understand the role of a protein in a specific biochemical network;
- to identify molecules that inhibit interaction between well-known proteins, like peptides in signal transduction pathways [190];

- as consequence of previous point, to test molecules (drugs) that can interfere with protein interaction: this can be useful to understand;
- to select antibodies against specific protein and to complete an epitope mapping of interaction [181];

4.3.3 Selection of scaffold libraries by use of PCA

The protein complementation assay is a powerful method for measuring protein-protein interactions in biological systems. It can be exploited to understand the role of a protein in a specific biochemical network or to verify how some molecules (e.g. drugs) can interfere in protein interaction. Protein interaction are involved in all biochemical pathways, and the PCA can be useful to characterize a protein with unknown function [170].

The system was used to demonstrate that single-chain antibodies libraries can be screened to antibodies that recognize specific antigens [177]. In this assay, the antibody library and the antigen were expressed genetically fused to the fragments of mDHFR enzyme. If an antibody binds the antigen, the binding event can be detected or selected for by the reconstitution of enzyme activity. By using and optimizing the length of a linker between the protein and the fragment, the interaction of proteins with steric hindrance can be improved.

The PCA was already used for the selection of new scaffold proteins: Amstutz and colleagues described a method to select specific MAP-kinase binders from a library of DARPins [191]. The application was improved when a pre-selection step was introduced using ribosome display. In other works, the stability and affinity of some protein scaffolds were improved: scaffold libraries of coiled-coil proteins were generated by mutating or randomizing specific residues into the sequence, and using the PCA as selection method to isolate improved-coiled peptides [106, 138, 149-151, 179, 180].

The interaction of activator protein-1 (AP-1) factors c-Fos and c-Jun was deeply analyzed. AP-1 is a transcription factor involved in numerous types of cancer: the leucine-zipper region of c-Jun and c-Fos components allow the dimerization and the complex activity. Using these leucine-zipper regions as a template, the sequences were modified, generating libraries and testing them to find improved binders [106, 138, 149, 150]. The interesting idea is the concept of modifications: in this approach, the libraries were constructed by mutating rationally the residues in specific positions, (the apolar core and the charged external regions), optimizing the core packing, α -helical propensity and electrostatics. With the PCA, new interactors with biomedical potential were isolated and characterized (wJun and wFos) [192]. Hagemann and coworkers compared the ability of these peptides to disrupt the cFos/cJun interaction to new binders selected with a different selection technique, the phage display [179].

In a recent work, a similar approach was performed to disrupt the interaction between Myc and Max proteins, due to the presence of a coiled-coil dimerization domain [151]. Myc is involved in several pathways, as cell proliferation and apoptosis; it is frequently deregulated in human cancers, but for the carcinogenic function the dimerization with Max is crucial. Jouaux and coworkers applied a rational mutagenesis to the Max sequence, to generate a library of binder variants. They randomized residues important for interaction, and the selection was carried out using PCA. Two peptides were isolated because they disrupt the Myc-Max dimer formation: consequently, inactivation of Myc provides a new therapeutic opportunity.

4.4 The display systems

More than 20 years have passed since the first display techniques were developed and the relevance of this method allowed to expand the principle to different systems [193]. The principle underlying display technologies is the physical association between phenotype of polypeptides under analysis and the genotype. This means that the identification of displayed protein lead to the isolation of sequence encoding for that protein. The sequence can be immediately characterized and modified with molecular biology tools and genetic engineering methods.

From the first display method, one concept is strengthened: it is possible to create a collection of billions of different displayed particles displaying different polypeptides, named “display libraries”. From these huge collections we can isolate the protein of interest: since the methodology provides a link between the phenotype and the genotype, the isolate particle keep also the genetic information.

The display technologies can be divided in 5 major categories, according to the platform used for the specific method [194]:

- **Ribosome display**: this method was the first cell-free *in vitro* display assay developed [195]. The link between phenotype and genotype is accomplished during *in vitro* translation by stabilizing the complex consisting of the ribosome, the mRNA and the nascent, correctly folded polypeptide. The entire complex is used for selection steps. Avoiding the transformation step, the diversity of the library is not limited by the transformation efficiency of bacterial cells, but only by the number of ribosomes and different mRNA molecules;
- **DNA display**: it is a selection strategy in which individual proteins are covalently linked to the cognate encoding DNA template in separate droplets of a water-in-oil emulsion [196]. It exploits the unusual property of endonuclease P2A of *E. coli* bacteriophage P2: the endonuclease P2A initiates the DNA replication of the bacteriophage P2 by making a covalent bond with its own phosphate backbone. This enzyme has been exploited as a new *in vitro* display tool for antibody fragments, constructing hybrid protein with P2A protein.
- **mRNA display** [197]: the translated peptides are covalently linked with their coding mRNA with the puromycin, a peptidyl acceptor antibiotic. Ribosome moves along the mRNA template: when ribosome reaches the end of template, the puromycin enters in the ribosome and is incorporated into the polypeptide. Some unique applications are possible with mRNA display, including library construction with unnatural amino acids [198]. Like the ribosome display, the major advantage of this method is that it is entirely performed *in vitro*, avoiding the transformation step in cells;
- **Cell display**: in this display method, the chimeric protein is expressed and exposed by surface of bacteria, yeast or mammalian cells. Systems based on bacteria cells has been tested to display functional enzymes, vaccine antigens and polypeptide libraries, also with tolerated insertions of more than 100 residues [199]. A large number of membrane proteins is useful for this approach, like *OmpA*, *OmpC*, *LamB* and *BtuB*, extensively modified and improved for more efficient system. Yeast display provides an advantage over bacteria system, for secreted proteins that require post-translational modifications

for correct folding and activity. For biomedical application, the use of mammalian cells has been improved, avoiding bacteria and yeast limitations like the different codon usage and some restriction on post-translational modification [200].

- **Phage display:** phage display is the most well-studied high-throughput screening method to study protein interactions, based on displaying peptides or proteins on the surface of a bacteriophage. The system will be thoroughly explained in the next section.

4.4.1 The phage display technology

The origin of the phage display dates to the mid-80s when Smith first expressed a foreign segment of protein on the surface of bacteriophage M13 virus particle [193]. Smith demonstrated that a fusion phage can be isolate from a background of phage particles displaying no antigen with more than 1000-fold enrichment, allowing phages displaying the correct antigen to be affinity purified against the cognate antibody.

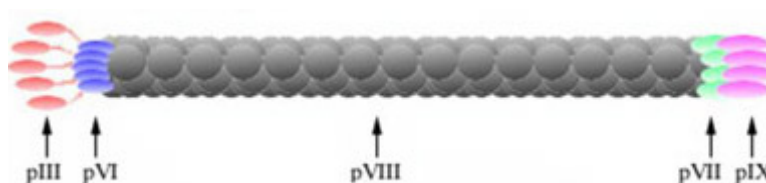


Fig. 1.20 – M13 bacteriophage structure: proteins which compose the capsid are indicated.

Filamentous bacteriophages are a group of related virus that infect only gram-negative bacteria. The most investigated are fF phages, including fd, f1 and M13 [201]. The genetic material consist on a single-stranded circular DNA molecule (6407 nucleotide long), encapsulated in a protein capsid. The genome of fF phages encodes 11 genes grouped according to their function: DNA replication, capsid packing and encoding proteins involved in phage membrane assembly [202]. The DNA is encased in a capsid composed by approximately 2700 copies of the major coat protein P8, and capped with 5 copies of four different minor proteins (P9, P7, P6, P3). The minor coat protein P3 binds to the receptor at the tip of the F-pilus of the host *E. coli*. For phage display system, the most important proteins are P3 and P8.

Thanks to their structure, their replication and packing processes (not limited by DNA size), these phages are useful tools to express recombinant proteins. With molecular methods, the DNA sequences of interest can be embedded in the phage genome with the correct framework of the genes coding for P3 or P8. The exogenous peptide are displayed on the phage surface in fusion with capsid proteins [203].

The principles of phage display are that the coat proteins of fF phages might tolerate being fused to foreign polypeptides without losing their function, and that the fused foreign polypeptide can be displayed on the surface of the filamentous phage. The general concept is that a phage encoding a specific fusion protein on its surface, could be isolated for its binding property to a given protein from a collection of billions of phages with different exposed proteins. As for other display systems, phage display provides a physical link between the phenotype and the genotype; the fact that the gene encoding a protein, selected for a specific property, is found within the phage displaying the protein itself, means that cloning of the gene

occurs simultaneously with the selection of the phage. The ease and rapidity of DNA sequence analysis permit to identify quickly the selected molecules.

Generally, the phage display cycle starts with the creation of diversity at the DNA level, creating a library of different variants. The DNA fragments are usually cloned upstream the gene encoding the protein 3 or the protein 8 of the phage: in the first case, there are 3 to 5 recombinant proteins at one end of the phage; in the second, all 2700 copies of the major coat protein are recombinant. Excluding the display of short peptides, the protein 3 based system is generally preferred. Although, P3 and P8 have been used to display proteins, P3 is the display protein par excellence, having been used to display large numbers of different proteins.

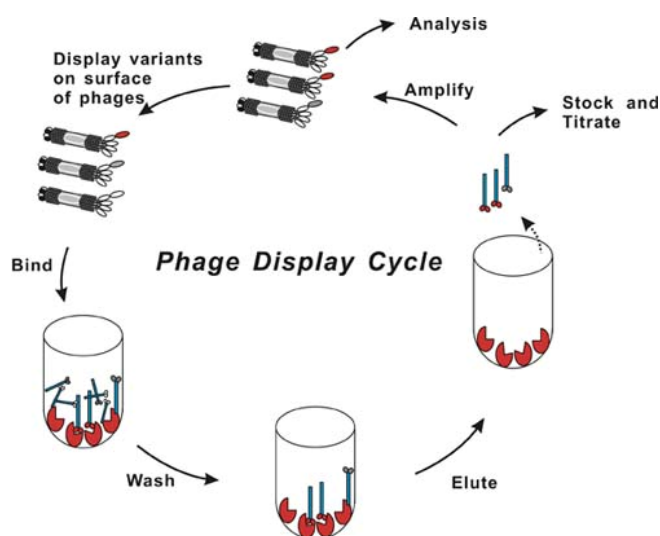


Fig. 1.21 – Schematic representation of a typical round selection: the elute, enriched in phages that bind the target molecule, is used for the next selection round. After 3-5 rounds, the library is enriched in phages that clearly bind the target material.

A typical selection round is illustrated in figure 1.21. In the phage display cycle, the phage displaying the protein of interest P is retained on a surface, coated with an antigen or an antibody recognizing P, while non-adherent phages are washed away. Bound phages can be removed from the surface, re-infected into bacteria and re-grow for further enrichment and eventually for binding analysis. The possibility to perform successive round of selection permits the isolation of proteins present in very low number in a population of billions of different phages. Usually, three or five rounds of panning are sufficient to enrich for binding peptide sequence.

Phagemid vectors have been developed as it is extremely difficult to work with phage genomes. These vectors, containing an Ff origin of replication and the gene encoding protein 3, can be manipulated and propagated as plasmids, and when display is required bacteria containing the vectors are infected with a helper phage, which provides all other components necessary for the creation of recombinant, selectable, phage particles.

4.5 Applications of polypeptide libraries

One of the most impressive aspects of the phage display is the variety of uses of this technology. Smith and his group have first used libraries displaying short peptides on antibodies for mapping antibody epitopes with that technique [193]. The increasing availability of structural data and genomic sequences of proteins provides new opportunity for the construction of improved phage libraries [204, 205].

The phage display technique allows the construction of protein libraries for several aims, like to find “consensus” sequences responsible of the protein-protein interaction, for epitope mapping or to find ligands of specific receptors, and for different final applications like protein microarrays [91], drug targeting and delivery and environmental monitoring [206].

A huge number of applications has been proposed for phage display techniques:

Combinatorial peptide libraries: random linear and cyclic peptide libraries and antigen- or gene fragment libraries have been used for the epitope mapping of monoclonal antibodies. Combinatorial peptide libraries, which can be constructed to contain billions of unique peptide sequences, have been a powerful source of molecules with different binding potential [207]. These libraries have served to identify protein binding sites, new substrates for enzymes, or peptide ligands for proteins involved in interactions. Peptide libraries have also been selected for motifs mimicking epitopes of antibodies produced in diseases [208].

Immunological peptide libraries: molecules naturally involved in immune recognition have been heavily exploited in display technologies. The two basic types of immunological interactions include (1) binding of antibody to an antigen and (2) binding of a T cell receptor (TCR) to a major histocompatibility complex (MHC)-presented antigen-derived peptide. The structural diversity between antibodies and TCRs has turned them into suitable platforms for construction of display libraries and as sources of specific and functional antigen-targeting molecules.

The most important and widely used application for immunological libraries is represented by antibody phage display [209, 210]. The entire antibody repertoire of the immune system could be transferred into this *in vitro* system, and used to select antibodies with a given specificity. Several monoclonal antibodies have been isolated and approved by the US Food and Drug Administration (FDA), and have been used as therapeutics in the clinic for several years [211]. Screening of antibody libraries aimed at the isolation of prototype drugs targeting cancer, infections and other diseases is one of the main approaches to target discovery and validation [212].

cDNA libraries: Display methods in general are limited in terms of the length of polypeptide translated from the library clones. For phage display system, longer polypeptides have a higher chance of disrupting virion functions, such as assembly and infectivity. In addition full-length cDNAs generally contain several stop codons near their 3' end.

These problems impeded the progress of display systems based on full-length cDNAs libraries. Initial breakthroughs were achieved with libraries displaying protein fragments. Such proteins domain libraries have been developed for

antibody epitope mapping to directly and unequivocally identify the antigen. Procedures for display cloning of cDNAs have evolved significantly since then, and monovalent phage display relying on helper phage has made whole protein presentation possible [213].

Scaffold libraries: antibodies are predominantly used to display permuted polypeptide sequence into a preserved protein domains. There are many other novel library frameworks that may be adopted for the presentation of polypeptides in display libraries [207]. The exposed loops that link structural elements of the protein (α -helices and β -sheets) are accessible for ligand binding and can accommodate sequence variation without changing overall structure and stability of the underlying scaffold.

In the last years many non-conventional scaffolds have been developed and used in protein display systems, like a good alternative to antibodies. Small dimensions, tolerance to multiple substitution and great stability are some features of these new tools.

Aim of the research

The advancements in biotechnology research allow the development of new biosensors, improved by nanotechnology knowledge: the use of new materials, or the exploiting of new physical-chemical proprieties emphasizing at nano-scale, allows the production of micro- and nano-devices with high biosensing potential.

Generally, a biological component is needed in bioanalytical device: several biological molecules have been considered and used for different purposes, such as antibodies, enzymes or oligonucleotides (DNA). Recently, a new generation of protein scaffolds has been developed: these structures, like antibodies, can recognize and bind specific molecules with high affinity, but possess smaller size, an important feature for a potential nanotechnological application.

The peculiar dimerization tendency of some proteins can be employed to build up specific biosensors. For example, the coiled-coil domain -a widely spread dimerization domain- is involved in protein-protein interaction in both eukaryotic and prokaryotic cells. The analysis of the sequence and 3D structure of these domains allows an *ex novo* design of new peptides, named E and K, which can interact with each other with high affinity and specificity. A deep characterization of their physical-chemical features has been performed. The stability of this interaction makes the coiled-coil domain an ideal platform to produce a fully functional and customized biosensor.

The aim of the present work is the study of the E/K structure, in order to understand if the dimerization of this coiled-coil domain can be used for biosensing applications, further improved with biological techniques.

Therefore, the main objectives of this research are:

- 1) the characterization of the E/K interaction in an *in vivo* systems;
- 2) to verify if E and K peptides of the coiled-coil domain possess a backbone that can tolerate sequence modifications, so that they can be considered scaffold proteins;
- 3) the design of strategical modification, allowing the isolation of new coils, with improved stability and affinity respect to E and K wild type coils;
- 4) the validation of dimerization capability of these new peptides with adequate *in vitro* tests, in order to verify the quality of interaction improvements;
- 5) the development of new scaffold protein, that can recognize and bind small molecules. Such new structure can be used as a biosensor units for nanotechnological platforms, e.g. micro- and nano-array.

The outcome of this research could allow the future development of arrays with co-transcriptional/translational techniques.

Results

Chapter 1

The E/K coiled-coil domain: in vivo validation

1.1 The protein complementation assay to validate E/K dimer

The E and K coils are a result of *de novo* peptide design, based on the knowledge that natural proteins can dimerize using α -helical domains. The physical and chemical properties characterizing this kind of interaction have been deeply characterized using circular dichroism, a technique that allows to resolve the structure of biological macromolecules, in particular the secondary structure of proteins. .

In order to validate the data obtained with physical methods, an *in vivo* assay is necessary to corroborate the dimerization capability and effective functionality of the coiled-coil domain. To this purpose, the protein complementation assay was used to validate the interaction between these peptides.

With the aim to develop a strategy for *in vivo* protein-protein interaction analysis and for high throughput screening of libraries, we employed a PCA based on the ampicillin resistance protein TEM-1 β -lactamase. The method has been extensively described in the introduction (paragraph 4.4.2). Briefly, a reporter protein is rationally divided in two fragments, unable to refold spontaneously, that are in turn fused to two proteins supposed to bind to each other. The refolding of the reporter protein, detected as reconstitution of its enzymatic activity, takes place only if the two proteins effectively interact.

1.2 Setup of PCA: $p\alpha$ and $p\omega$ expression vectors

The first step of this work was to setup suitable expression vectors to perform a TEM-1 β -lactamase-based PCA. Like other PCAs, the protein complementation assay based on this enzyme needs two different vectors for the expression of reporter fragments: the $p\alpha$ vector contains the α domain, whereas the $p\omega$ vector contains the second subunit, the ω domain. Both

plasmids are derived from the pBluescript [214] vector, whose beta-lactamase gene has been substituted by the kanamycin (kan) or chloramphenicol (Chl) resistance genes in $p\alpha$ and $p\omega$ respectively. The β -lactamase coding sequence of pUC119 plasmid was used as template to generate fragments by PCR [181].

The pAlfa ($p\alpha$) vector is constituted by the following cassettes:

- the kanamycin resistance cassette;
- the inducible lacZ promoter (Plac), to allow the control of the expression of the recombinant protein;
- the leader sequence proper of TEM-1 β -lactamase protein, to guide the proper secretion into the periplasmic space;
- the α domain, encompassing residues 1-195 of the β -lactamase ORF;
- a tripeptide sequence NGR, able to improve β -lactamase complementation;
- a glycine-serine (GS) linker;
- the wild type sequence of E-coil, with flanking 5' EcoRI and 3' HindIII restriction sites;
- an histidine tag (HIS), to detect the recombinant protein..

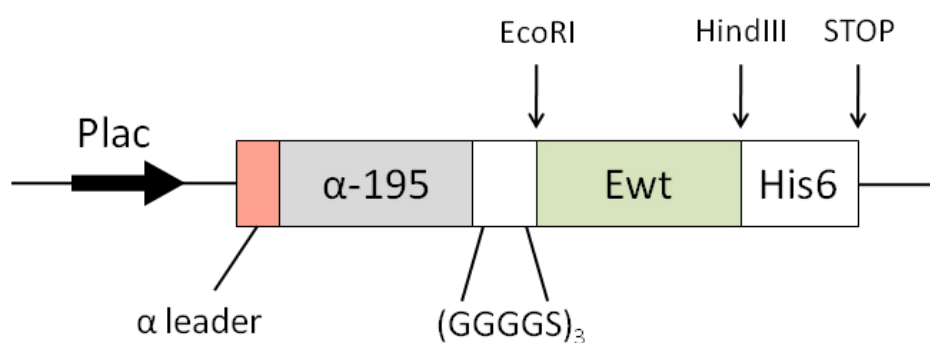


Fig. 2.1 – Schematic representation of $p\alpha$ vector. The expression of the recombinant protein is under the control of a Lac promoter, and can be monitored by using the His tag; the translocation into the periplasmic space is under the control of the lactamase leader sequence. The Ewt sequence is subcloned in fusion with the α -fragment of lactamase, the glycine-serine linker and the NGR sequence, able to improve enzyme complementation.

$p\Omega$ vector is constituted by the following cassettes:

- the chloramphenicol resistance cassette;
- the inducible lacZ promoter (Plac), to allow the control of the expression of the recombinant protein;
- a leader sequence to guide the correct secretion into the periplasmic space;
- the wild type sequence of K-coil with flanking 5' BssHIII and 3' NheI restriction sites;
- an SV5 tag, to detect the recombinant protein;
- the ω domain, encompassing residues 196-286 of the β -lactamase ORF.

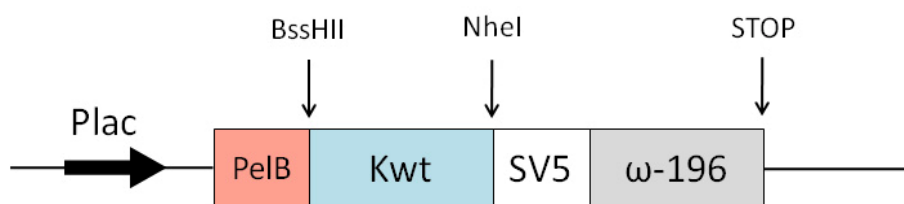


Fig. 2.2 - Schematic representation of the $p\omega$ vector. The Kwt sequence is subcloned in fusion with the ω -fragment of lactamase. The expression of the recombinant protein is under the control of a Lac promoter, and the production can be detected by using the SV5 tag; the translocation into the periplasmic space is under the control of the bacterial PelB leader.

The 3D folding of these chimeras is shown in figure 2.3.

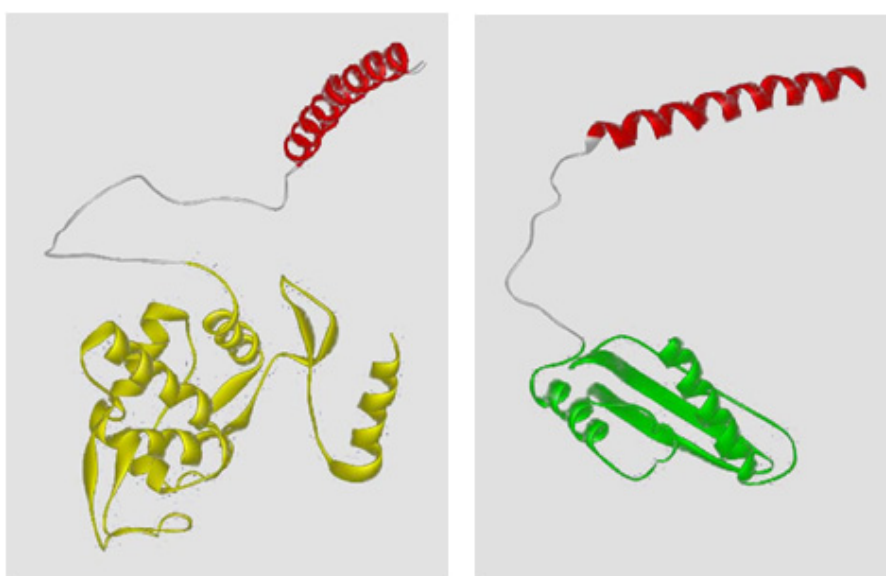


Fig. 2.3 – Three-dimensional molecular models of α -Ewt and ω -Kwt: the coils are labeled in red, the α fragment is labeled in yellow, the ω fragment in green.

In this thesis, the coils described by Tripet and Chao are named “wild type E” (Ewt) and “wild type K” (Kwt).

1.3 Cloning and validation of wild type E and K-coils

The E and K sequences were amplified through polymerase chain reaction (PCR). The first step was the design of primers, whose sequence has been derived from the amino acid sequence. The strategy to obtain the full fragments was based on the self-annealing of complementary 3' ends and extension of primers.

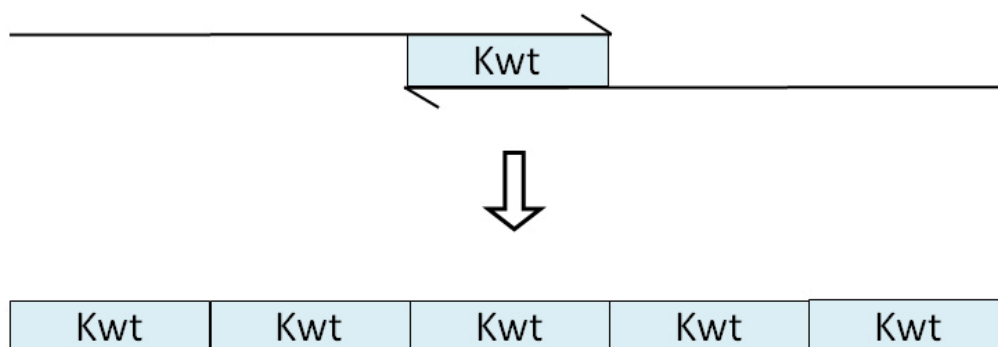


Fig. 2.4 – Scheme of the adopted strategy to obtain the full sequence of K wild type coil. The same strategy was applied to the Ewt coil.

Because each of the five heptads constituting E or K coils has the very same amino acid sequence, in order to guarantee the correct primer annealing in following PCR amplifications and avoid undesired base pairing that would lead to aberrant PCR products, it was necessary to design appropriate primers. Exploiting the redundancy of the genetic code, the same residue in different heptads was encoded by different codons, so that the nucleotide sequence was as more variable as possible, avoiding internal self annealing and allowing correct base-pairing during subsequent rounds of PCR ends.

Ewt							
Position	<i>g</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
Amino acid	E	V	S	A	L	E	K
Heptad I	gaa	gtt	agt	gcg	ctg	gag	aag
Heptad II	gag	gtt	agc	gct	ctt	gag	aaa
Heptad III	gag	gta	tcc	gcg	ctc	gaa	aaa
Heptad IV	gag	gtt	tct	gcg	ctt	gaa	aag
Heptad V	gaa	gtg	agc	gct	ctt	gag	aag

Kwt							
Positions	<i>g</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
Amino acid	K	V	S	A	L	K	E
Heptad I	aaa	gta	agc	gct	ctc	aag	gaa
Heptad II	aag	gta	tcc	gcg	ctt	aaa	gag
Heptad III	aag	gta	tct	gca	ctc	aag	gag
Heptad IV	aaa	gtt	tct	gct	ctc	aaa	gaa
Heptad V	aaa	gtg	agc	gca	ctg	aaa	gag

Table 2.5 – The nucleotide sequence of Ewt and Kwt, derived from the amino acid sequence: in each coil, the residues in the same position were encoded by different codons in different heptads. The nucleotide sequence was the basis for the primer design.

On the basis of these sequences, a pair of forward and reverse primers for each coil was designed, with a 3' terminal sequence homology to allow self-assembly and 5' terminal restriction sites (BssHIII/NheI for Kwt, EcoRI/HindIII for Ewt). PCR products were subsequently purified, digested and ligated with compatible digested vectors, generating p α -Ewt and p ω -Kwt plasmids in transformed *E. coli* DH5 α competent cells.

Positive colonies were sequenced to verify the correctness of the peptide sequences. The expression of α -Ewt and ω -Kwt fusion proteins and their effective export into the periplasmic space was induced treating bacteria with IPTG, that activates Lac promoter and downstream cassettes, and loading samples on SDS-PAGE to detect recombinant proteins with anti-tag antibodies.

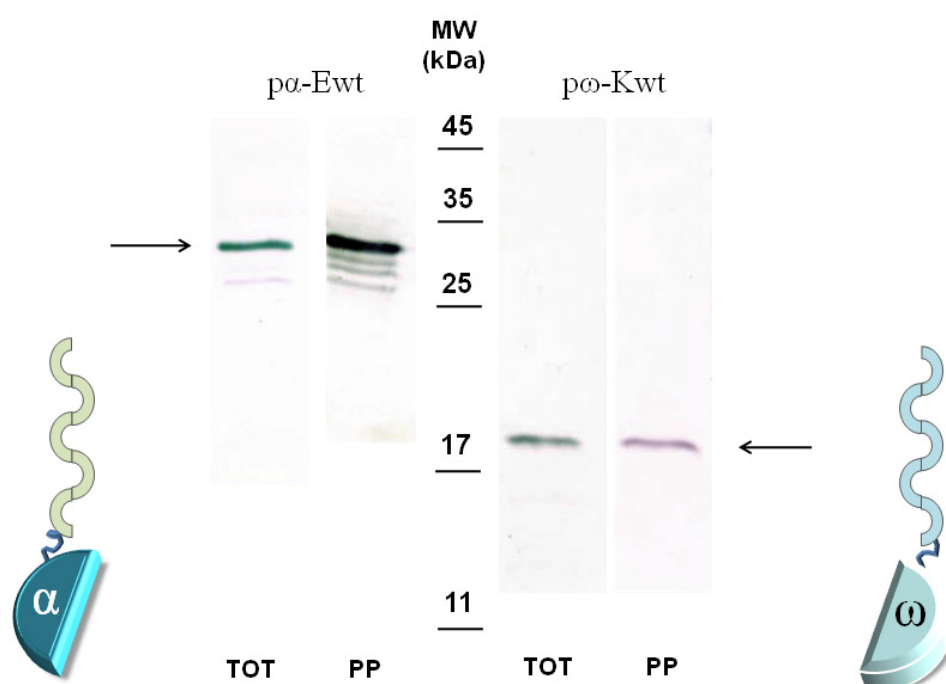


Fig. 2.6 – Western blot assay to check the size and the expression level of fusion proteins α -Ewt and ω -Kwt. The SDS-PAGE was loaded with total (TOT) and periplasmic (PP) fractions of DH5 α induced for the expression of both Ewt in fusion with the α fragment and the Kwt in fusion with ω fragment of the β -lactamase. The arrows indicate the predicted molecular weights of the fusion proteins. Primary antibody: mouse anti-His 1:5000 (α -Ewt), mouse anti-SV5 1:2000 (ω -Kwt). Secondary antibody: goat anti-mouse AP conjugated 1:2000.

The immunoblot shows that the proteins were correctly processed: the arrows indicate the predicted molecular weight, corresponding to about 30 kDa for α -Ewt and 19kDa for ω -Kwt.

1.4 The PCA of wild type E/K coiled-coil domain

The PCA was performed to determine if the coils are able to interact in a *in vivo* system. The goal was to verify the reconstitution of β -lactamase activity and, consequently, the bacterial growth on plates supplemented with ampicillin. As control plasmids, in order to test the interaction specificity between E and K coils, a scFv (Δ G2) specific for the *V. cholerae* toxin

ΔG was cloned into the $p\omega$ expression vector, and the tissue transglutaminase (tTG), an enzyme that catalyzes the cross-link between glutamine residues and polyamines, was cloned in $p\alpha$.

E. coli DH5 α cells were co-transformed with the following $p\alpha$ and $p\omega$ constructs:

- $p\alpha$ -Ewt / $p\omega$ -Kwt;
- $p\alpha$ -Ewt / $p\omega$ - $\Delta G2$;
- $p\alpha$ -tTG / $p\omega$ -Kwt.

Bacteria transformed with only one vector did not grow in presence of ampicillin, ruling out any possible intrinsic antibiotic resistance (data not show). DH5 α cells were also transformed with pUC119 (Invitrogen) to verify the bacterial growth of cells with an ampicillin resistance given by an intact and non-fragmented β -lactamase. All the co-transformed bacteria were plated onto agar plates containing either chloramphenicol/kanamycin, to confirm the presence of both vectors, or increasing amounts of ampicillin to assess the sensibility of the complementation system.

As yet explained in the paragraph 1.2, the $p\alpha$ vector contains the gene for kanamycin resistance, $p\omega$ the gene for chloramphenicol resistance. In the chloramphenicol/kanamycin plate the growth is allowed only for the bacteria that contain both vectors, $p\alpha$ and $p\omega$. Therefore, with this plate it is not possible to know something about interaction among the peptides. That information can be obtained from ampicillin plates, because this antibiotic selects bacteria producing two interactor peptides. Only interactors can reconstitute the β -lactamase: the active enzyme can break the β -lactam ring of antibiotic, and the growth of bacteria is permitted.

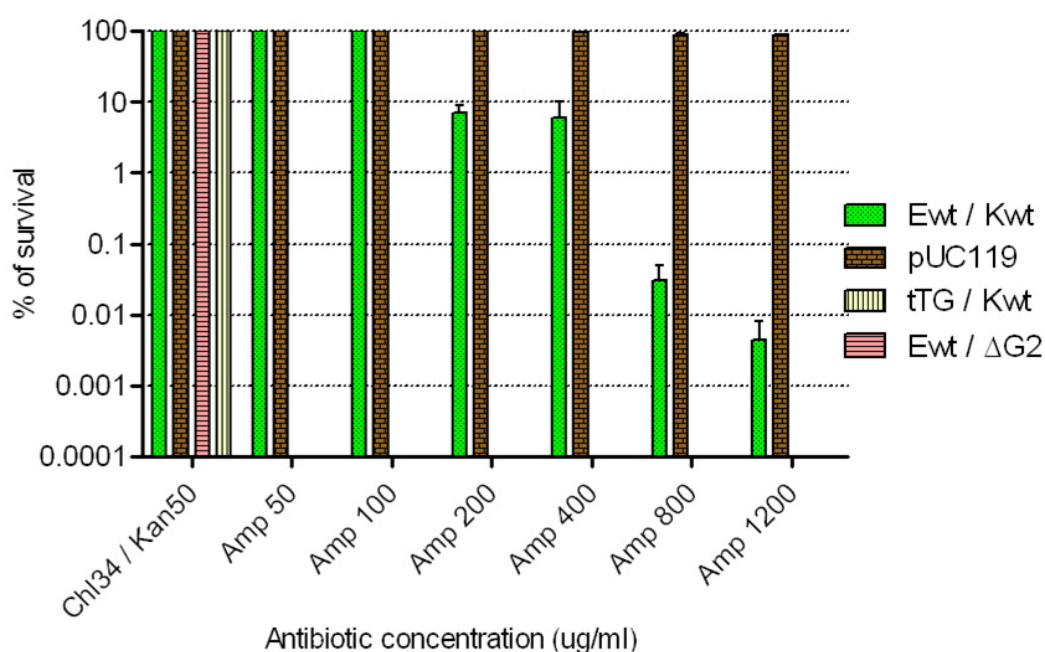


Fig. 2.7 – Percentage of bacterial survival of DH5 α transformed with pUC119, and co-transformed with $p\alpha$ -Ewt and $p\omega$ -Kwt, $p\alpha$ -tTG and $p\omega$ -Kwt, $p\alpha$ -Ewt and $p\omega$ - $\Delta G2$: the number of colonies growing on chloramphenicol/kanamycin plates, indicating the number of co-transformants, were set to 100%. The number of colonies grown on different ampicillin concentrations were compared with 100%.

The formation of ampicillin resistant colonies was observed only in bacteria co-transformed with $p\alpha$ -Ewt and $p\omega$ -Kwt, and transformed with pUC119. From graph in fig 2.7, one can

appreciate how the number of colonies was inversely correlated to the concentration of ampicillin, indicating PCA is a system sensible to antibiotic toxicity; whereas, bacteria with an intact β -lactamase grown without a detectable decrease. The graph confirms that wild type E and K coils strongly interact *in vivo*: the binding of the two coils reassembles the lactamase and reconstitutes its activity, corroborating the interaction data obtained with circular dichroism. A 3D-interaction model of the wild type coiled-coil domain with complemented α and ω domains is reported in fig 2.8.

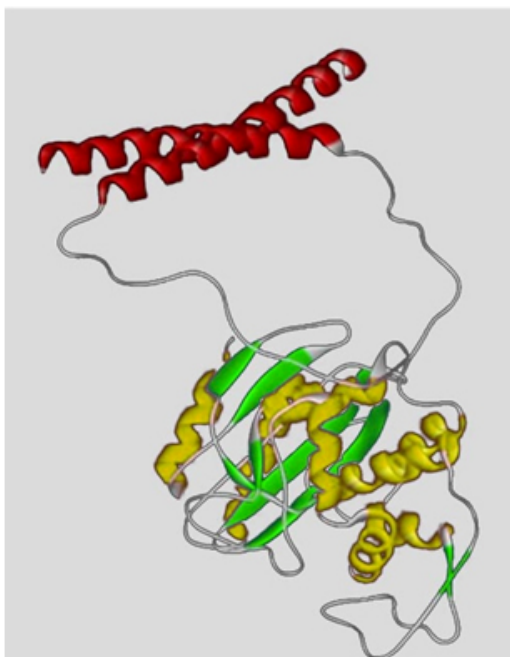


Fig. 2.8 – Three-dimensional molecular model of the interaction between α -Ewt and ω -Kwt the fusion proteins (α fragment in yellow, ω fragments in green).

Moreover, it was demonstrated that the activity of the complemented lactamase is more sensible to antibiotic than the non-fragmented enzyme. No colonies were detected onto the plates with the other combinations of co-transformed bacteria, even at low ampicillin concentrations; these observations confirm that the complementation of β -lactamase and its activity are strictly dependent on E and K interaction.

Chapter 2

The wild type E/K domain as a scaffold

2.1 Engineering of a coiled-coil domain

As described in details in chapter 2, scaffolds are binding reagents which may be employed in a wide range of biotechnological and biomedical applications. Modified scaffolds can be selected with both *in vitro* and *in vivo* techniques, allowing the isolation of new molecules with improved affinity against specific ligands [84].

The second aim of this project was to understand if wild type E and K peptides can be considered scaffold proteins. Generally, to use a protein as a scaffold some basic requirements should be considered:

- high stability;
- high interaction affinity;
- small size;
- easy expression in bacteria and eukaryotic cells;
- the folded protein (and the overall stability of the whole molecule) must be tolerant to sequence modification;

The last point is a fundamental feature of scaffolds: the native protein folding must not be altered by the modification or insertion of additional residues, reflecting the elasticity of supporting structures (or frameworks) which can exhibit a novel binding capability. Moreover, these structures may be obtained by rational, or most commonly, combinatorial protein engineering techniques prone to further improvements and modifications [85].

In this section all these requirements have been taken in consideration, according to previous reports [114, 135-137, 144, 145]. Moreover, we previously obtained good results about the ease of coil expression in bacteria, even if fused to β -lactamase fragments.

In order to assess the tolerance to modifications of the sequence, some mutations were inserted in the wild type E and K sequence. The resulting interactions were analyzed and validated.

2.2 Construction of E and K mutant coils

Considering the structure of E and K coils, the mutations were inserted only into the central heptad of the peptides. To this purpose, glycine was considered because of the small size and the absence of charge.

Two different strategies were applied to verify potential alterations of stability:

With the first cycle of PCR, the mutagenized region inserted into the forward primer was already introduced into the peptide. With the second and third amplification rounds, the entire peptide was generated. The PCR products with proper 5' and 3' restriction sites were cloned into the correct vector. Emut and Egly were cloned into p α vector, Kmut and Kgly into p ω vector, respectively.

The detailed cloning procedure is reported in the material and methods section.

2.3 Validation of E and K mutant coils

The first validation of new peptides was given by sequencing, to verify the correct cloned sequence of the central heptad.

The next step was to assess if the mutate coils can effectively interact as a stable heterodimer: to this purpose the protein complementation assay was performed. The aim was to compare the bacterial growth on plates with increasing concentration of ampicillin. *E. coli* DH5 α cells were co-transformed with each p α and p ω construct. As explained, the heterodimers tested for interactions were:

- p α -Ewt / p ω -Kwt, as positive control;
- p α -Emut / p ω -Kmut;
- p α -Egly / p ω -Kwt;
- p α -Ewt / p ω -Kgly.

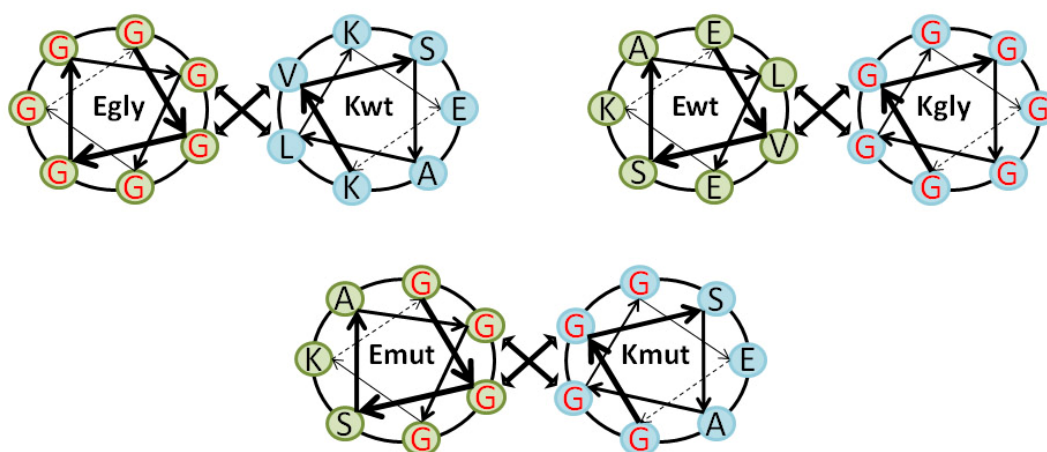


Fig. 2.11 – Helical wheel representation of the three new heterodimers, tested in PCA. The glycine-mutagenized residues are labeled in red.

Bacteria were plated onto agar plates containing either chloramphenicol/kanamycin or increasing amounts of ampicillin (from 50 μ g/ml to 1200 μ g/ml) and 1mM IPTG; the plates were incubated for 48 hours at 28°C. The survival ratio of each clone was calculated, comparing the growth rate on chloramphenicol/kanamycin and on ampicillin plates. The antibiotic sensitivity could be a simple means to select clones which express stable recombinant proteins among genetically homogenous cells.

Bacteria were grown on chloramphenicol/kanamycin plates to confirm the effective co-transformation, without giving any useful information about the E/K interaction. The latter was obtained plating cells on increasing ampicillin concentrations, in order to select functional interactions between coils. As show in figure 2.12, the formation of ampicillin-resistant colonies was observed in all co-transformed cells, but with different trends for each new heterodimer.

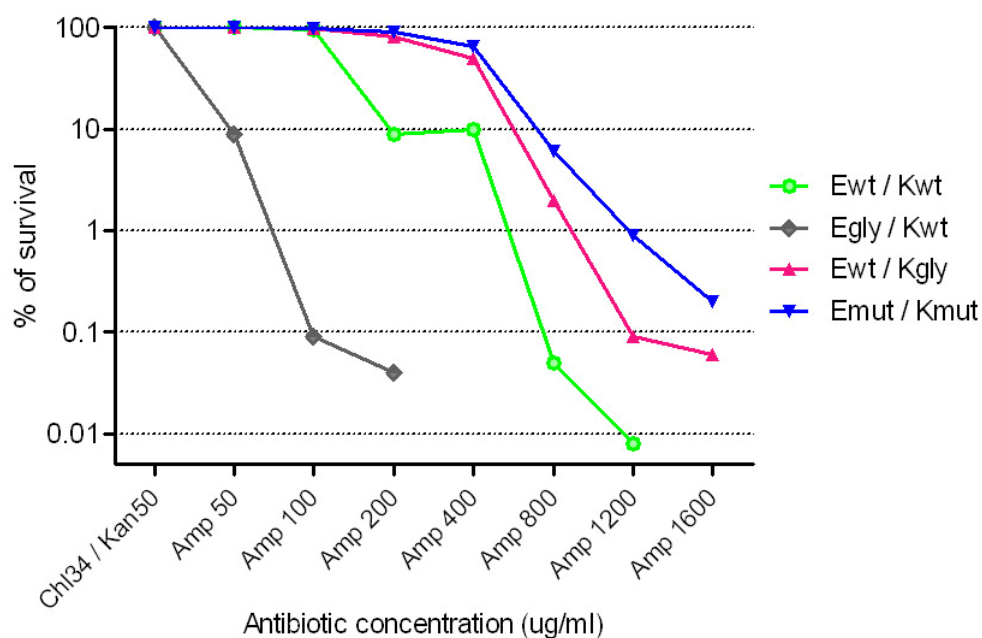


Fig. 2.12 – Bacterial growth of DH5 α co-transformed with $p\alpha$ -Ewt and $p\omega$ -Kwt, $p\alpha$ -Egly and $p\omega$ -Kwt, $p\alpha$ -Ewt and $p\omega$ -Kgly, $p\alpha$ -Emut and $p\omega$ -Kmut. The number of colonies growing on chloramphenicol/kanamycin plates, indicating the number of co-transformed bacteria, were set to 100%. The number of colonies grew on different ampicillin concentrations were compared with 100%.

The results clearly show that some modification compromised the stability of coils interaction. The Emut/Kmut dimer possess the best stability and affinity, reflecting increased antibiotic resistance, even if both coils were modified; interestingly, the mono-mutated dimer Ewt/Kgly is significantly more stable than Egly/Kwt. Therefore, even if with different activities, both mutated E and K peptides can fold to form an α -helix domain and interact with each other. According to Tripet and Chao [114, 136], each coil alone should present a random structure, that spontaneously folds into ordered α -helical motif to form a right-handed supercoil in presence of the complementary coil. We demonstrated that the wild type K-coil is more tolerant to sequence modification than the E-coil to allow this folding.

Therefore we can conclude that wild type E and K coils can be classified as scaffold proteins because they retain basic requirements even in presence of sequence alterations. Concerning the modifications, it seems that the K-coil is more resistant and stable than the E-coil, at least for the central heptad repeat mutation. Given these results, we focused our attention on the K-coil, in order to improve *in vivo* and *in vitro* performances of this heterodimeric coiled-coil complex

Chapter 3

Improving the wild type E/K dimer: the K random coils.

3.1 Introduction

Chao and coworkers designed a *de novo* heterodimeric coiled-coil domain after having analyzed the structure of natural coiled-coils [135, 136]. This wt E/K system was employed in purification systems and for biosensing applications, in virtue of the extremely stable and specific dimerization of E and K peptides. However, some experimental procedures often require extreme pH, redox and ionic strength conditions that could interfere or even disrupt the interactions between the coils and therefore alter the results of the assay. For this reason the development of improved coils with increased stability and specificity is needed.

Likely to previous studies, the interaction quality of protein scaffolds was ameliorated mutating specific residues into the sequence and selecting new peptides with improved features with [106, 138, 149-151, 179, 180]. In the previous section it was demonstrated that the K-coils is more tolerant to sequence modification than the E-coil. Then, the wild type sequence of K-coil was modified, and a set of new heterodimerization domains was obtained with different interaction abilities, that were selected through β -lactamase-based PCA.

Two different strategies were tested to improve the previously selected heterodimeric coiled-coil domain, optimizing the core packaging, α -helical structure propensity and electrostatic features :

- 1) in the first approach, the whole central heptad was randomly mutagenized (**a-b-c-d-e-f-g**), substituting each residue with one of 20 natural available amino acids. Any type of arrangement can be obtained, and a collection of different peptides was created: the Krandom library (Kran);
- 2) in the second approach, the mutagenesis was applied on the whole sequence of the peptide, not only in the central heptad. Here, only the residues fundamental for the interaction and dimerization were mutated, i.e. the amino acids in the apolar core (*a* and *d*) and in the charged external region (*e* and *g*). Secondly, the mutated positions were substituted not with any of the 20 natural amino acids, but only with ones sharing common charge and polarity (apolar residues for the hydrophobic core and charged or polar residues at the external surface).

Since the library was constructed with a rational design based on K-coil sequence, it was identified as a library of K rational design peptides (Krd).

3.2 Construction of the ρ -K random library

As already explained in the previous paragraph, the ρ -K random library was generated randomizing the central heptad of the K helix by means of mutated primers. The other four heptads of K peptide maintained the wild type sequence, as done in the case of glycine mutations, in order to avoid an excessive variability of the whole K helix.



Fig. 2.13 – Scheme of the K-coil randomization: only the central heptad was randomized, while other heptads retained the wild type sequence.

Due to the degeneration of the genetic code, all amino acids except methionine and tryptophan are encoded by more than one codon. For example, leucine, serine and arginine are encoded by six different codons. A degenerated triplet is represented as NNN, where N indicates any of the four nucleotides.

I Position: A or T or G or C = N

II Position: A or T or G or C = N

III Position: A or T or G or C = N

In order to insert random mutations, the use of degenerated random primers was necessary. However, with the NNN triplet three stop codons can be casually inserted (TAA, TAG, TGA), giving a premature arrest of the coil sequence. The codon NNY prevents the selection of nucleotides A and G, but it excludes three amino acids (methionine, glutamate and glutamine). This drawback was overcome choosing codon NNK as random triplet, where K stands for the nucleotides G or T: NNK allows to encode any of the 20 amino acids, avoiding the stop codons, except for TAG (amber stop). This problem was solved using the *E. coli* DH5 α , a suppressor bacteria strain that can translate TAG in a glutamine residue.

I Position A or T or G or C = N

II Position A or T or G or C = N

III Position T or G = K

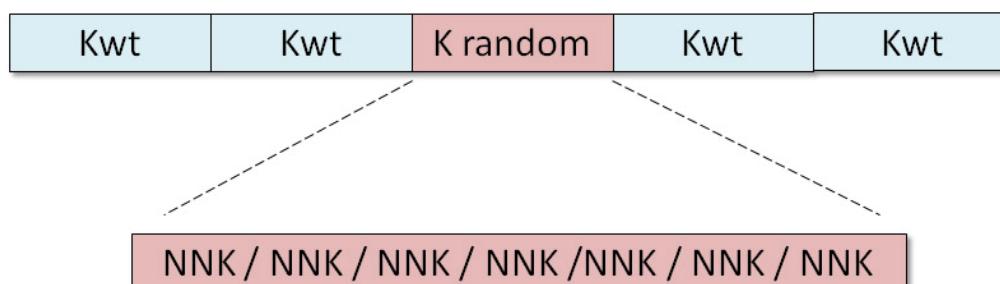


Fig. 2.14 – To insert amino acids randomly, the central heptad sequence was degenerated by using the NNK codon, that allows to encode any of the 20 natural amino acids avoiding the stop codons, except for TAG.

Considering the extension and the arrangement of the randomized sequence, the theoretic number of possible scaffolds is $1,28 \times 10^9$.

3.3 The primers design and construction of p ω -Kran library

The PCR was used to generate the library of K random fragments. Given the sequence of Kwt peptide, primers were designed to obtain a set of K fragments with the central heptad randomized: as for the construction of glycine-mutated coils, three subsequent cycles of PCR were performed (see figure 2.15).

With the first cycle of PCR, a mutagenized region was already introduced in the K fragments. With the second and third amplification reactions, the full K randomized fragments were generated. Detailed cloning procedure and the primers used are reported in the material and methods section.

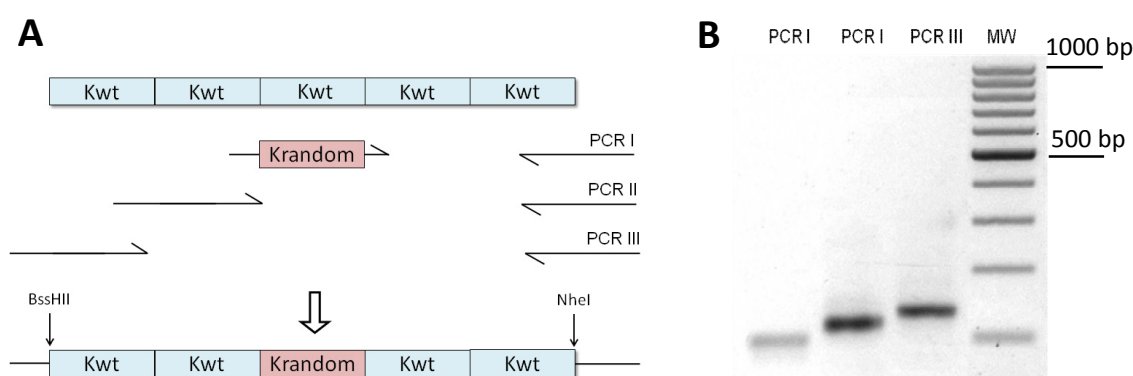


Fig. 2.15 – (A) Schematic representation of the three subsequent rounds of PCR necessary for the construction of the K random coils. (B) Electrophoresis of PCR products of three amplification rounds: in the last lane DNA marker, ranging from 100 to 1000 base pairs, is reported.

The figure 2.14-B shows the amplification product for each step. The final step yielded a PCR product of about 120 bp.

For the construction of the library, both p ω vector and K random fragments were digested with BssHII and NheI restriction endonuclease. Vector and inserts were ligated, and the ligation product was used to electroporate DH5 α electro-competent cells. Bacteria were subsequently plated onto six plates with 34 μ g/ml chloramphenicol, and a little amount was plated on ampicillin plates to verify the absence of contamination. The ampicillin check is fundamental, because the presence of contaminants at this level could cause false-positives during PCA. Dilution were plated onto a small plate with chloramphenicol to assess the size of the library.

After 15 hours of incubation at 30°C, cells were harvested and stocked with 20% of glycerol at -80°C. The number of independent clones was 10^8 : the result was considered valid and the library was characterized.

3.4 p ω -Kran library characterization

The presence of K random fragments into bacteria was assessed by PCR, performed on ten clones picked randomly. Generic primers were used, external to the cloning site of p ω vector (VLPT2/VHPT2)



Fig. 2.16 – Electrophoresis of PCR-amplified insert DNA of 10 randomly picked clones of p ω -Kran library. Each clone, expressing a different cloned Kran fragment, was amplified with VLPT2 and VHPT2 primers. 100bp DNA marker is reported in the last lane.

All clones presented a PCR product of 150 bp. The result here reported confirmed the integration of K fragments into the p ω vector, and the presence into each bacteria of one p ω -Kran plasmid.

To check the diversity of clones represented in the library, some of them were sequenced to verify the randomization of the central heptad. Table 2.17 shows that all the sequenced clones have a different central mutagenized heptad.

Clone	Sequence
Kwt	KVSELKE
1	AFRRFLV
2	LVFRAHL
3	RKILVVE
4	SERLAVH

Table 2.17 – Four clones of the library were sequenced to check the randomization. In this table the sequences are compared with Kwt sequence: the residues are completely randomized.

3.5 Selection of the p ω -Kran library by PCA

The obtained library is a huge collection of different K fragments with a short randomized sequence of seven residues. These K random fragments are characterized by a variable stability when interacting with the wild type E-coil. Again, β -lactamase-based PCA was performed to select better binding partners to Ewt than Kwt peptides. In this case, the E-coil represents the wild type peptide, while the K-coil is the new randomized binder. The identification of new binders took place with two subsequent screenings of the library.

3.5.1 PCA-screening of p ω -Kran library

In the first screening, 500 ng of p ω -K random library DNA was used to transform p α -Ewt electrocompetent *E. coli* DH5 α , generating co-transformed bacteria, containing both p α and p ω vectors. All co-transformed bacteria were equally plated on six plates with 1 mM IPTG and increasing concentration of ampicillin:

- ampicillin 200 μ g/ml;
- ampicillin 400 μ g/ml;
- ampicillin 800 μ g/ml;
- ampicillin 1200 μ g/ml;
- ampicillin 2000 μ g/ml;
- 34 μ g/ml chloramphenicol, 50 μ g/ml kanamycin.

The plates were incubated for 48 hours at 28°C.

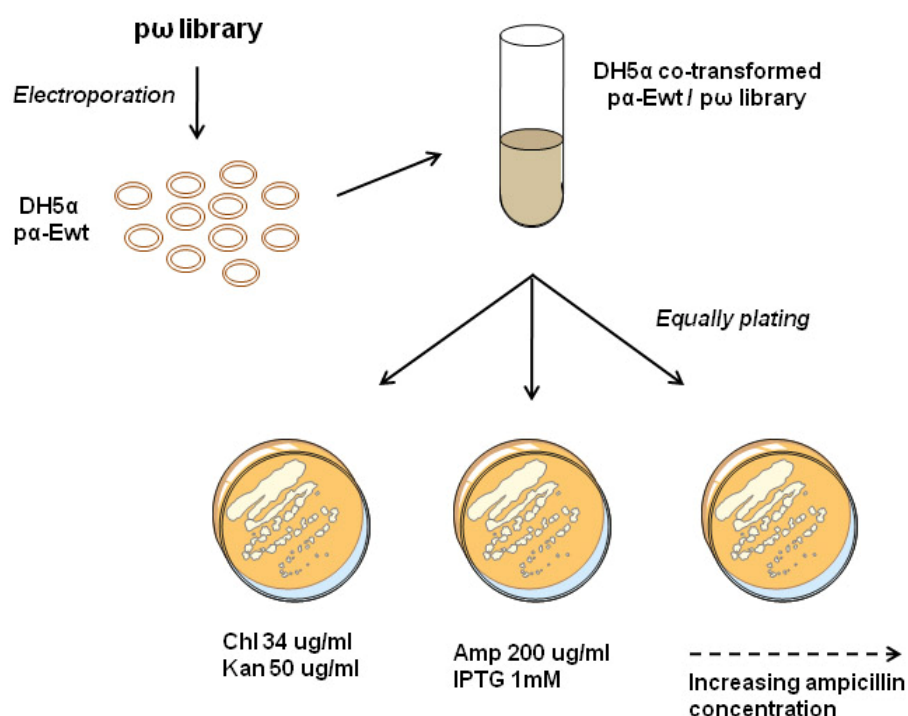


Fig. 2.18 – Summary of the first screening strategy: the p ω library (p ω -Kran in this case) was electroporated into *E. coli* DH5 α electrocompetent cells. The co-transformed bacteria (8×10^8) were equally plated onto agar plates supplemented with antibiotics (chloramphenicol/kanamycin and increasing concentration of ampicillin).

8×10^8 p ω -Kran/p α -Ewt co-transformed bacteria were obtained. This means that on each plate more than 10^8 bacteria were plated. Therefore it was possible to evaluate interactions of all Kran fragments. Bacteria grown on chloramphenicol/kanamycin plate confirmed the co-transformation efficiency, whereas on ampicillin substrate bacterial growth was a specific result of the interaction between Ewt coil and Kran binders.

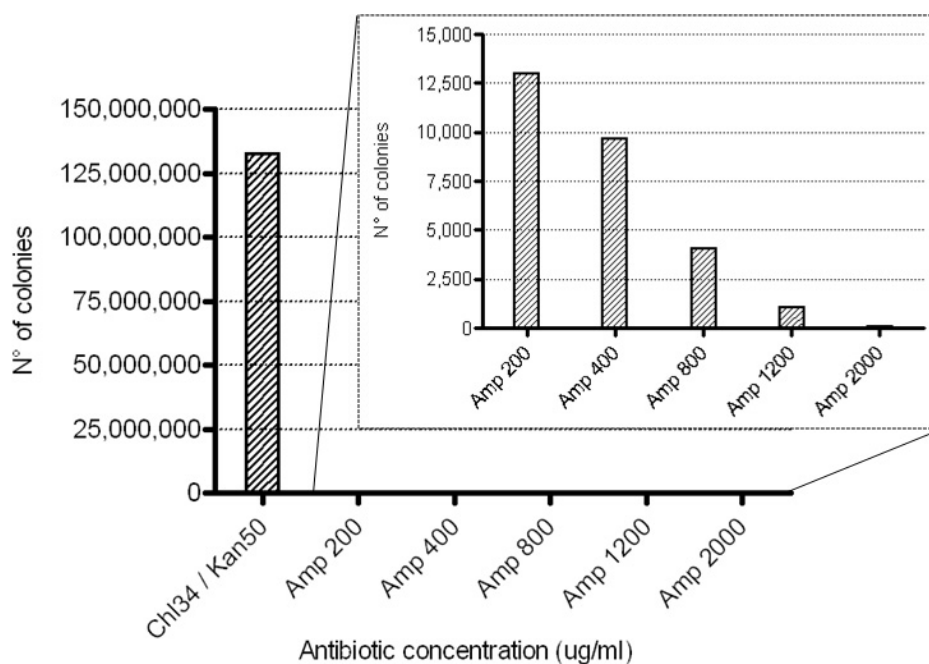


Fig. 2.19 – Results of the first-PCA screening of $\rho\omega$ -Kran: the graph shows the number of colonies that grew onto each plate. Any co-transformed bacteria grew onto the chloramphenicol/kanamycin plate; the selective pressure of increasing ampicillin concentrations allows the growth of the strongest Ewt/Kran interactor coils.

A decrease in the bacterial growth was observed at increasing concentrations of ampicillin, that confirms the selective pressure of antibiotic. The number decreased to 26 clones at 2000 $\mu\text{g/ml}$ of ampicillin. The higher the antibiotic concentration, the strongest the E/K affinity was. Out of 26 isolated clones, six were chosen for a second screening round.

3.5.2 PCA-validation of selected $\rho\alpha$ -Ewt/ $\rho\omega$ -Kran clones

With a second screening, the above selected clones were analyzed and compared with the curve of Ewt/Kwt growth. Our goal was to select the bacterial clone containing the best pair of interactors. The following $\rho\alpha$ / $\rho\omega$ clones were tested:

- Ewt / Kran 5.11;
- Ewt / Kran 5.12;
- Ewt / Kran 5.17;
- Ewt / Kran 5.32;
- Ewt / Kran 5.40;
- Ewt / Kran 5.44;
- Ewt / Kwt.

The survival ratio of each clone was calculated, comparing the growth rate on chloramphenicol/kanamycin and on ampicillin plates (from 200 $\mu\text{g/ml}$ to 200 $\mu\text{g/ml}$). Details of this method are reported in materials and methods section.

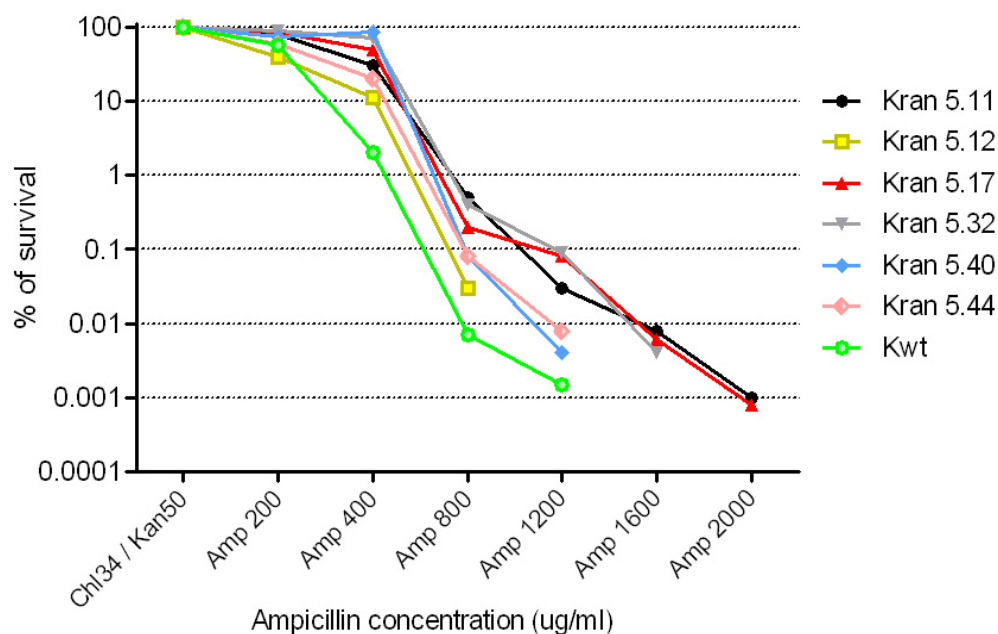


Fig. 2.20 – Results of the PCA-screening of six clones isolated in the first screening. bacterial growth is compared with the one of Ewt/Kwt (growth curves actually refer to co-transformant Ewt/Kwt or Ewt/Kran bacterial clones).

The results reported show that all of these seven clones displayed a better growth curve respect to the E/K wild type system, according to the data obtained in the first round of selection. These clones showed a better resistance to ampicillin than bacteria expressing wild type peptides, with the best performances registered for clones 5.11 and 5.17 (growth on plates with 2000 $\mu\text{g/ml}$ ampicillin).

This result confirmed the ability of PCA technology to select new coils with improved *in vivo* features, in terms of interaction and dimerization.

3.6 Analysis of the expression of selected Kran coils

The production of K random peptides was assessed by western blot assay. The production was verified:

- in bacteria containing only the isolated $p\omega$ -Kran: they are named “single clones”, because they lack of the $p\alpha$ vector;
- in co-transformed bacteria, containing both vectors.

Protein expression was induced in bacteria with 1 mM IPTG at 28°-30°C over night.

As already explained, the recombinant protein “ ω domain-Kran” is first produced in the cytoplasm, and then is translocated into the periplasmic space. The production and the translocation were verified using both the whole bacterial and the periplasmic lysate. All samples were loaded in a SDS-PAGE gel, followed by a immunoblot. The recombinant proteins were analyzed for reactivity to SV5 tag.

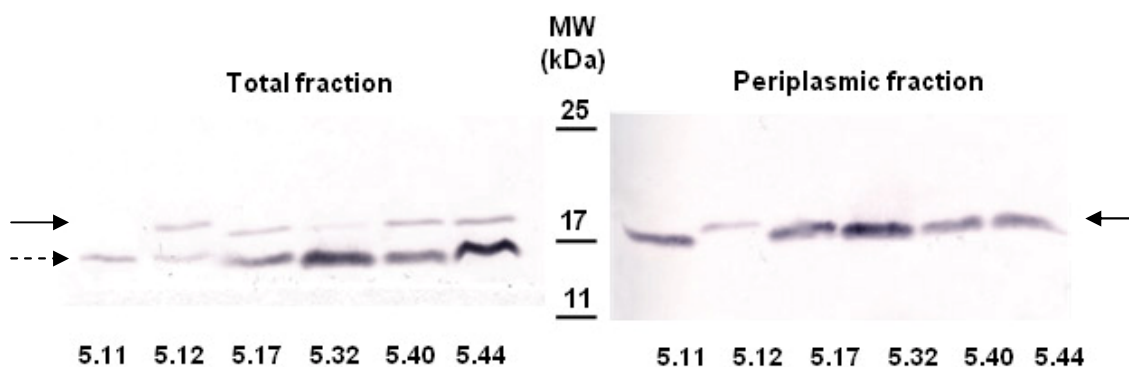


Fig. 2.21 - Western blot assay to check the size and the expression level of randomized K-coils in $p\omega$ -Kran single-transformed bacteria. The SDS-PAGE was loaded with both total (TOT) and periplasmic (PP) fractions of DH5 α induced for the expression of the recombinant proteins. The arrows indicate the predicted molecular weights of the fusion proteins, the dash marks the degradation product. Primary antibody: mouse anti-SV5 1:2000. Secondary antibody: goat anti-mouse AP conjugated 1:2000.

The pictures show the results for single-clones western blot assay: recombinant proteins were clearly detected, indicated by arrows. The molecular weight of 19 kDa is the expected one for Kran/ ω -domain chimera (arrows). Two bands were observed for some clones: the upper band is the correct fusion-protein product, showing the predicted molecular weight, whereas the lower band is a degradation product (dash arrows).

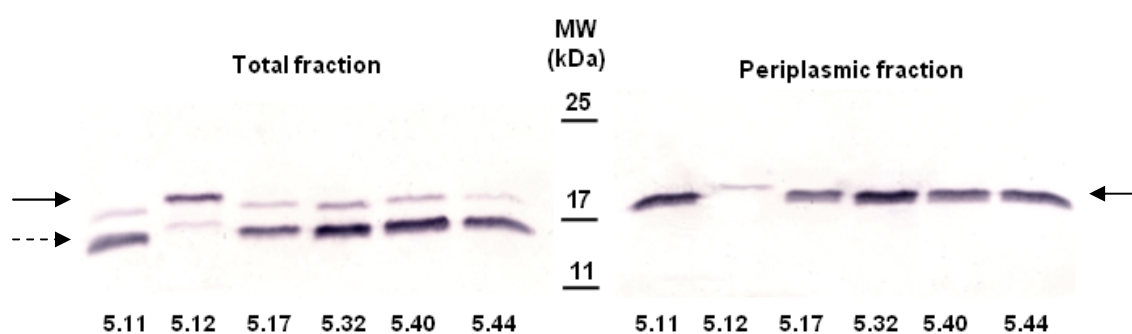


Fig. 2.22 - Western blot assay to check the size and the expression level of randomized K-coils in $p\alpha$ -Ewt/ $p\omega$ -Kran co-transformed bacteria. The SDS-PAGE was loaded with total (TOT) and periplasmic (PP) fractions of DH5 α induced for the expression of recombinant proteins. The arrows indicate the predicted molecular weights of the fusion proteins, the dash marks the degradation product. Primary antibody: mouse anti-SV5 1:2000. Secondary antibody: goat anti-mouse AP conjugated 1:2000.

In figure 2.22 it has been confirmed that also co-transformant cells correctly express and process K random coils.

Chapter 4

Improving the wild type E/K dimer: the K rational design coils

4.1 Construction of the p ω -K rational design library

A totally different approach was adopted to build up the K rational design library, derived from wild type K-coil. Analyzing the sequence of Kwt, a rational mutagenesis was developed considering:

- 1) to mutate amino acids only in specific heptad positions;
- 2) to employ only few amino acids for the randomization process;

The modifications were performed on all the five heptads, and not only in the central one as done before. Since the library was constructed with a rational design of K-coil sequence, it was identified as a library of “K rational design” peptides (Krd).

To this purpose, the p ω -Krd library was developed mutating the most important positions of Kwt heptads for the coiled-coil interaction: *a* and *d* (into the apolar core), necessary for packaging of hydrophobic surfaces, and *e* and *g* (external charged residues), important to drive specificity and stability during heterodimerization. In recent works [106, 138, 149, 179, 215] a similar rational design approach was adopted, but mutating only one position in the apolar core, either *a* or *d*. In the same way, in this work we kept wild type the residues in the position *a* (figure 2.24)

Only amino acids with similar polarity and charge respect to the original residue were inserted: apolar residues for the hydrophobic core and charged or polar residues for external surfaces.

- for *d*: valine, leucine, isoleucine, alanine, methionine and proline;
- for *e* and *g*: glutamine, glutamate, arginine, lysine, aspartate, histidine, serine and asparagine.

Analyzing each codon that translates for desired residues, two degenerate codons were designed: VYN and VRN.

Position	Codon
<i>d</i>	G or C or A = V
	T or C = Y
	A or T or G or C = N
<i>e/g</i>	G or C or A = V
	A or G = R
	A or T or G or C = N

Table 2.23 – Degenerate codons designed for the mutation of *d*, *e* and *g* positions.

In this way, each codon combination can translate for any residue on the considered position. However, with this degenerated codons unwanted amino acids could be produced: glycine (apolar) in *e* and *g*, threonine (polar) in *a*. Even if these residues carry unsuitable characteristics for the position they have to cover, these amino acid-encoding codons were retained.

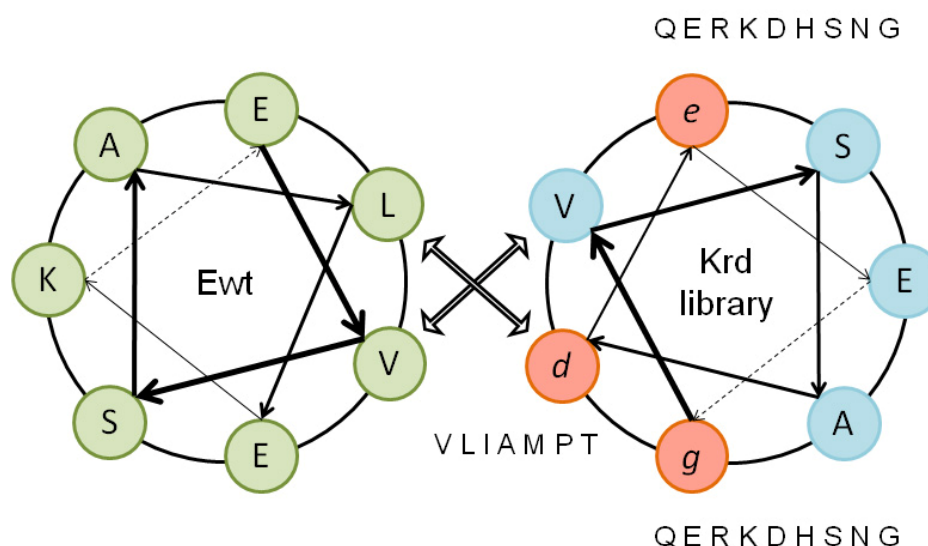


Fig. 2.24 – Schematical representation of the strategy adopted for the rational design of the new library. All possible residues are listed for red-highlighted *d*, *e* and *g* positions.

The first step for generating the Krd library was the design the suitable primer pairs. Unlike the construction of Kran library, only a pair of long primers (megaprimers) was used for Krd fragments. Both primers possess a 3' sequence homology that allows self-assembly (figure 2.25). To this purpose, it was necessary to conserv the central heptad in order to generate a primer homology region. In this way, the fragments of the library were obtained keeping the wild-type central heptad, and randomizing the four lateral heptads.

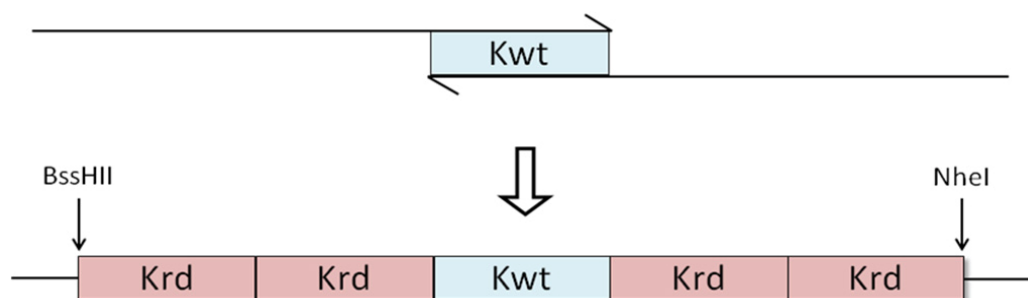


Fig. 2.25 - Scheme of the adopted strategy to obtain the full sequence of K randomized fragments: the sequence of the central heptad was conserved in the wild type form, to allow the annealing between the 3' ends of the designed primers. BssHIII and NheI restriction endonucleases sites are included at the 5' and 3' ends of PCR product respectively.

Considering the positions to mutate and the limited degeneration, the theoretic number of different fragments is about 10^{11} , but only $2,17 \times 10^{10}$ are the real appropriate scaffolds without unwanted residues (threonine and glycine).

4.2 The primers design and construction of $\rho\omega$ -Krd library

The primers design was performed studying the whole sequence of the Krd peptide and using degenerated codons. We considered the nucleotide sequence of Kwt (table 2.4) substituting the codon in the chosen positions with the degenerated codons previously designed.

	<i>g</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
Heptad I	<i>vrn</i>	gta	agc	gct	<i>vyn</i>	<i>vrn</i>	gaa
Heptad II	<i>vrn</i>	gta	tcc	gcg	<i>vyn</i>	<i>vrn</i>	gag
Heptad III	aag	gta	tct	gca	ctc	aag	gag
Heptad IV	<i>vrn</i>	gtt	tct	gct	<i>vyn</i>	<i>vrn</i>	gaa
Heptad V	<i>vrn</i>	gtg	agc	gca	<i>vyn</i>	<i>vrn</i>	gag

Table 2.26. – Sequence of the degenerated K rational design coil: the degenerated codons are labeled in red.

Likely to $\rho\omega$ -Kran library primers, 5' BssHIII and NheI sites were inserted in the oligonucleotide sequence.

Figure 2.27 shows the obtained K fragment (about 120 bp): a strong band is visible at the right size, proving that primers designed were correctly designed ensuring the desired amplification of Krd fragments.

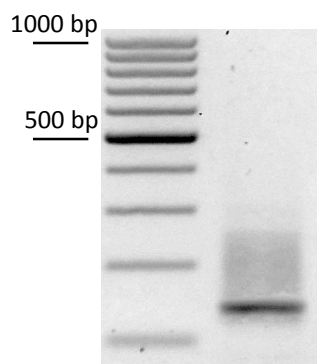


Fig. 2.27 - Electrophoresis of the PCR-assembly product of K rational design fragments: 100bp DNA marker, ranging from 100 to 1000 base pairs, is reported.

The library of Krd fragments and the $p\omega$ vector were both digested with BssHIII and NheI restriction endonucleases. Vector and inserts were ligated, and *E. coli* DH5 α competent cells were electroporated with the ligation product. Bacteria were then plated onto six plates containing 34 μ g/ml chloramphenicol. Also for this library a contamination check was performed, plating a little amount of cells onto an ampicillin plate. Dilution were plated onto a small plate with chloramphenicol to assess the size of the library. After 15 hours of incubation at 30°C, bacteria were harvested and stocked with 20% glycerol at -80°C.

The number of obtained clones was about 10^7 : the library was characterized and tested.

4.3 $p\omega$ -Krd library characterization

The presence of Krd fragments in bacteria, cloned into $p\omega$ vector, was assessed performing a PCR with generic primers (M13-20/VHPT2) on ten randomly picked clones (figure 2.28)

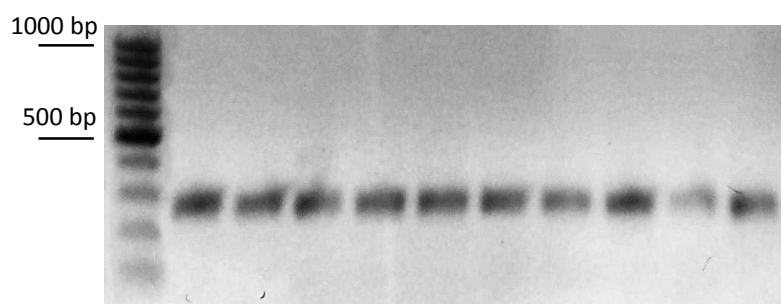


Fig. 2.28 - Electrophoresis of PCR-amplified insert DNA of 10 randomly picked clones of $p\omega$ -Krd library. Each clone, expressing a different Krd fragment, was amplified with M13-20 and VHPT2 primers. 100bp DNA marker is reported in the first lane.

All clones contained a fragment of the expected size, about 300 bp: the larger PCR size than Krd peptides was due to external position of the annealing primers employed for the screening. The results here reported confirmed the integration of Krd fragments into the $p\omega$ vector, and the presence into each bacteria of one $p\omega$ -Krd plasmid.

To check the real diversity of the clones of the library, some of them were sequenced, verifying the randomization of selected residues. The results are reported in table 2.29.

Clone	Heptad 1	Heptad 2	Heptad 3	Heptad 4	Heptad 5
Kwt	KVSALKE	KVSALKE	KVSALKE	KVSALKE	KVSALKE
1	EVSAIDE	RVSAVHE	KVSALKE	SVSAMKE	RVSAIEE
2	DVSAADE	DVSAASE	KVSALKE	DVSAISE	KVSATGE
3	DVSATGE	RVSAIHE	KVSALKE	KVSAVDE	RVSAPGE
4	HVSAPRE	GVSAVRE	KVSALKE	KVSAPSE	NVSALSE

Fig. 2.29 – The sequence of four Kran clones compared with the sequence of Kwt coil. The randomized residues are labeled in red. The central heptad (heptad 3) retains the wild type sequence in all the sequenced clones.

4.4 Selection of the p ω -Krd library by PCA

As done for the K random library, PCA was performed also on Krd library to validate new improved interaction between coils. In this case, the E-coil represents the wild type peptide, while the K-coil is the new rationally randomized binder. The identification of new binders took place with two subsequent screenings of the library.

4.4.1 PCA-screening of p ω -Krd library

In the first screening, 500 ng of p ω -Krd library were electroporated in p α -Ewt *E. coli* DH5 α competent cells. All transformed bacteria were equally plated onto six plates with 1 mM IPTG and increasing concentration of ampicillin:

- ampicillin 200 μ g/ml;
- ampicillin 400 μ g/ml;
- ampicillin 800 μ g/ml;
- ampicillin 1200 μ g/ml;
- ampicillin 2000 μ g/ml;
- 34 μ g/ml chloramphenicol, 50 μ g/ml kanamycin.

Bacteria were plated on allowed to grow 48 hours at 28°C. $7,2 \times 10^7$ p α -Ewt/p ω -Krd co-transformed bacteria were obtained. This means that on each plate about $1,2 \times 10^7$ bacteria were plated. Bacteria grown on chloramphenicol/kanamycin plate confirmed the co-transformation efficiency, whereas on ampicillin substrate bacterial growth was a specific result of the interaction between Krd and Ewt coils.

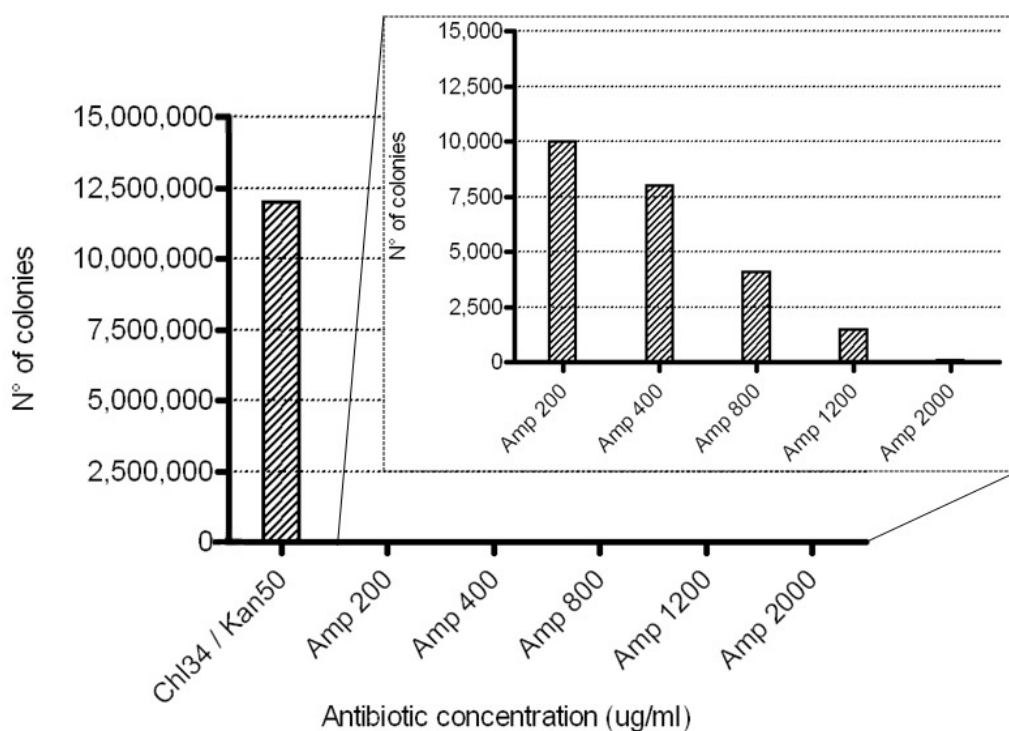


Fig. 2.30 - Results of the first-PCA screening of $p\omega$ -Kran library: the graph shows the bacterial growth onto each plate. Any co-transformed bacteria grew onto the chloramphenicol/kanamycin plate; the selective pressure of increasing ampicillin concentrations allows the growth of the strongest Ewt/Krd interactor coils.

The results on the graph show that all co-transformed bacteria grew in presence of chloramphenicol/kanamycin, confirming the presence of both vectors; a decrease in the bacterial growth was observed at increasing concentrations of ampicillin. At ampicillin 200 $\mu\text{g/ml}$, 10^4 clones grew, while the number decreased to 40 clones with 2000 $\mu\text{g/ml}$ of ampicillin.

Out of 40 isolated clones, seven were chosen for a second round of screening

4.4.2 PCA-validation of selected $p\alpha$ -Ewt/ $p\omega$ -Krd clones

The seven best clones from first screening were analyzed and compared with the growth ability of Ewt/Kwt clones, with the aim to select the best clone with the best pair of interactors coils. The following $p\alpha/p\omega$ clones were tested:

- Ewt / Krd F7;
- Ewt / Krd F8;
- Ewt / Krd F13;
- Ewt / Krd F14;
- Ewt / Krd F24;
- Ewt / Krd F25;
- Ewt / Krd F26;
- Kwt / Ewt.

The survival ratio of each clone was calculated, comparing the growth curve on chloramphenicol/kanamycin and on ampicillin (from 400 $\mu\text{g/ml}$ to 3000 $\mu\text{g/ml}$). The data obtained are reported in the following graph.

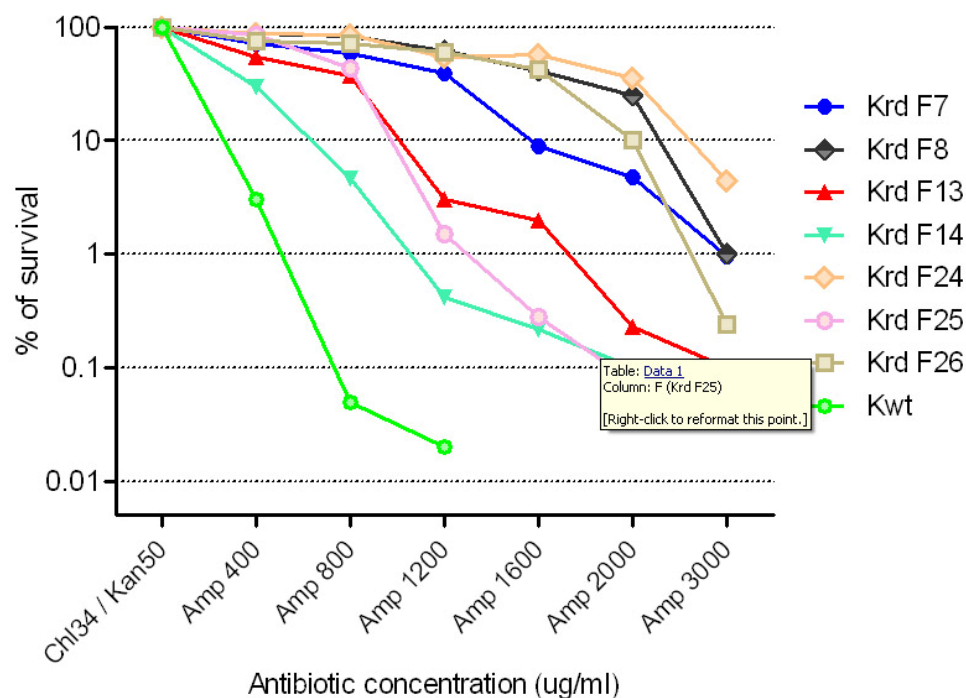


Fig. 2.31 - Results of the PCA-screening of seven clones isolated in the first screening. The bacterial growth is compared with the one of Ewt/Kwt (growth curves actually refer to co-transformant Ewt/K bacterial clones).

The reported results show that all seven clones survived to ampicillin 3000 $\mu\text{g/ml}$ with a better survival ratio than Ewt/Kwt, which were not able to grow at ampicillin 1600 $\mu\text{g/ml}$. This final PCA test allow to select four clones: Krd F7, Krd F8, Krd F24 and Krd F26. These K-coils interact very strongly with Ewt, allowing a stable reconstitution of β -lactamase with a consequent improved resistance to ampicillin than wild type cognate coils.

4.5 Analysis of the expression of selected Krd coils

The production of K rational design peptides was assessed by western blot. The production was verified:

- in bacteria containing only the isolated $p\omega$ -Krd: they are named “single clones”, because they lack of the $p\alpha$ vector;
- in co-transformed bacteria, containing both $p\omega$ -Krd/ $p\alpha$ -Ewt vectors.

Protein expression was induced in bacteria with 1 mM IPTG at 28°-30°C over night.

As for Kran library, peptides production and translocation were assessed using whole bacterial and periplasmic lysates. All samples were loaded in a SDS-PAGE gel, followed by a immunoblot. The recombinant proteins were analyzed for reactivity to SV5 tag, followed by anti-mouse AP conjugated antibody recognition.

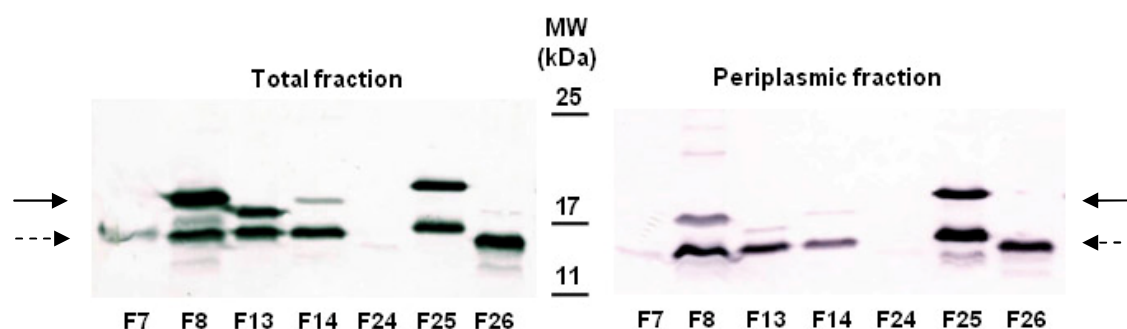


Fig. 2.32 - Western blot assay to check the size and the expression level of randomized K-coils in $p\omega$ -Krd single-transformed bacteria. The SDS-PAGE was loaded with both total (TOT) and periplasmic (PP) fractions of DH5 α induced for the expression of the recombinant proteins. The arrows indicate the predicted molecular weights of the fusion proteins, the dash marks the degradation product. Primary antibody: mouse anti-SV5 1:2000. Secondary antibody: goat anti-mouse AP conjugated 1:2000.

Recombinant proteins were clearly detected for almost every clone, except for clones F7 and F24, that showed a lower expression level. Two bands were pointed out for some clones: the upper band is the correct fusion-protein product, showing the predicted molecular weight (about 19 kDa), whereas the lower band is a degradation product. Some proteins could show an apparent different molecular weight after SDS-PAGE: this can be ascribed to different charge/mass ratio of the coils, because the amino acids in *d*, *e* and *g* position of each Krd fragment are randomized, and they could have different physical characteristics (polarity or charge).

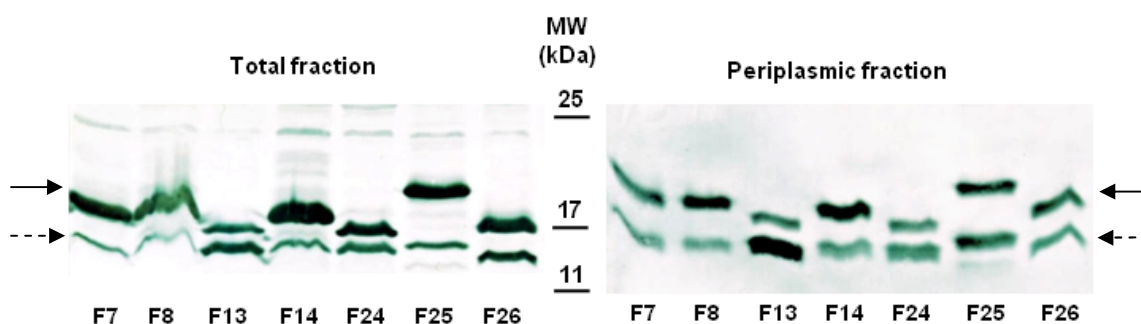


Fig. 2.33 - Western blot assay to check the size and the expression level of randomized K-coils in $p\alpha$ -Ewt/ $p\omega$ -Krd co-transformed bacteria. The SDS-PAGE was loaded with total (TOT) and periplasmic (PP) fractions of DH5 α induced for the expression of recombinant proteins. The arrows indicate the predicted molecular weights of the fusion proteins, the dash marks the degradation product. Primary antibody: mouse anti-SV5 1:2000. Secondary antibody: goat anti-mouse AP conjugated 1:2000.

The results obtained confirm that also co-transformed bacteria express correctly and efficiently the new coils. Interestingly, Krd F7 and Krd F24 clones correctly produced Krd peptides. It is likely that these Krd proteins expressed in the single-clones bacteria, that can not interact with Ewt, are highly unstable and rapidly degraded. On the other hand, if the Ewt coil is present, the interaction stabilizes efficiently avoiding the degradation.

Chapter 5

In vitro validation of new Kran and Krd coils

5.1 Setup and vector engineering

A fundamental condition for PCA is that the fragments of the reporter protein must not be able to refold spontaneously. As a consequence, the interactions between the expressed fusion proteins guide the complementation of the reporter-protein domains, which take place in periplasmic space. Sometimes a low background signal is present, due to an incorrect protein folding, to the formation of inclusion bodies and to a low amount of self-complementation of β -lactamase fragments. These drawbacks decrease the specificity and efficiency of our system.

To check that the interaction between isolated K mutant coils with Ewt coil is not catalyzed by the two fragments of β -lactamase (but the opposite), mutant and wild-type coils were produced and purified without the α and ω fragments. Then, *in vitro* tests were performed to verify that different conditions than the oxidative environment of the periplasm could allow the interaction between coils as well. These preliminary tests are fundamental to understand if the isolated proteins can be used for biosensing applications, where the experimental conditions can be quite different compared with the *in vivo* cellular environment.

To this purpose, two different expression vectors were used to produce the mutant coils and to perform multiple crossed-interaction assays: pMAL-ASC2 and pGEX. Each vector can express high levels of recombinant protein, which were further purified to decrease the background due to unwanted protein contaminants.

The pMAL-ASC2 vector is a previously modified version of pMAL-X plasmid, purchased from New England Biolabs. In the pMAL protein fusion system, the sequence of the protein of interest is inserted downstream of the MBP (maltose binding protein) cassette, resulting in a MBP-fusion protein that can be easily purified by affinity purification with maltose or maltodextrins. The vector includes a sequence recognized by a specific protease, in order to cleave and separate the protein of interest from MBP after purification, without any further vector-derived protein residues.

The pMAL-ACS2 vector was previously engineered inserting an *AscI* restriction site (compatible with *BssHIII* overhangs), allowing a quick sub-cloning of DNA fragments when the restriction endonucleases *BssHIII* and *NheI* are used. The pMAL-ASC2 carries a SV5 tag for recognition by anti-SV5 monoclonal antibody, followed by stop codon.

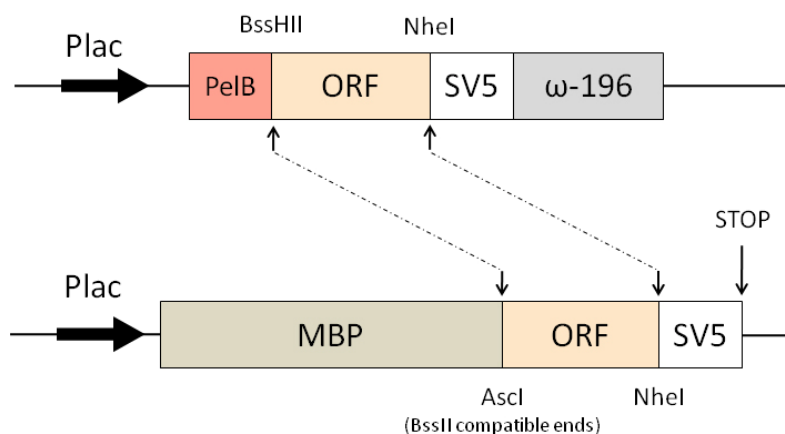


Fig. 2.34 – Sub-cloning of selected ORF sequence from the p α vector into the expression vector pMAL-ASC2.

For the same reason (sub-cloning compatibility), the pGEX-4T-1 was previously modified to obtain a vector with BssHIII and NheI sites. The pGEX vectors are members of the GST (glutathione-S-transferase) fusion system: these plasmids allow the production of recombinant protein in fusion with GST, which permits a fast and easy affinity purification with glutathione. Also pGEX vector includes a recognition site for a specific protease (thrombin).

Comparing to the original vector, the polylinker of pGEX contains only BssHIII and NheI sites for restriction endonucleases and the FLAG-tag for the recognition with a monoclonal antibody.

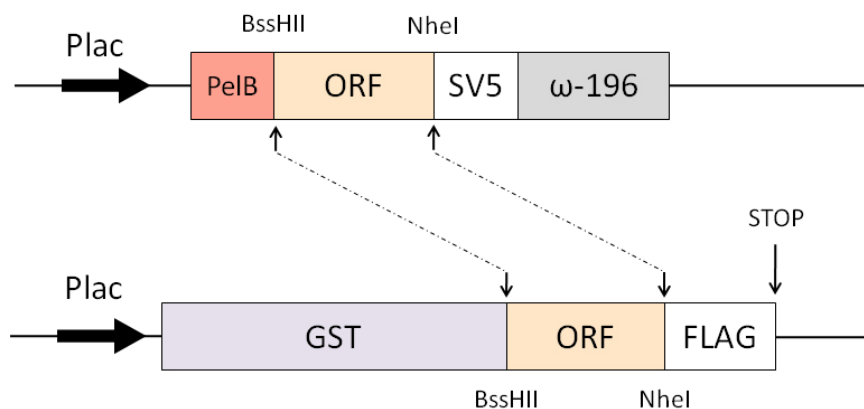


Fig. 2.35 – Sub-cloning of selected ORF sequence into the expression vector pMAL-ASC2

Into both pMAL-ASC2 and pGEX vectors, the expression of fusion peptides is under the control of an IPTG inducible promoter, and the presence of a leader sequence, like in p α and p ω vectors, allows the translocation of fusion proteins into the periplasmic space that can be quickly extracted from the soluble periplasmic fraction.

The sub-cloning step into these expression vector was performed for the following coils: Ewt, Kwt, Kran 5.11, Kran 5.17, Krd F8 and Krd F26 (wild type coils plus the best clones of two libraries PCA-selection). The Kwt and mutant K-coils were readily subcloned, because the fragments carry the BssHIII and NheI restriction sites. The Ewt coil needed an extra engineering step to insert via primer amplification the appropriate restriction endonuclease sites at the ends

of Ewt sequence. The coils selected for the production were subcloned and sequenced to check for the absence of mutations.

Every pGEX and pMAL-ASC2 coil protein was produced and purified with glutathione or amylose system purification, from 100ml of bacterial culture induced O/N with IPTG. Purified proteins were subsequently quantified by using both Coomassie gel staining and Bradford assay: the concentration varied between 0,3 mg/ml and 4 mg/ml. The results of purification procedure and Coomassie staining are reported in the next picture, where 1 μ l of each purified protein was loaded in a 12% polyacrylamide gel (figure 2.36).

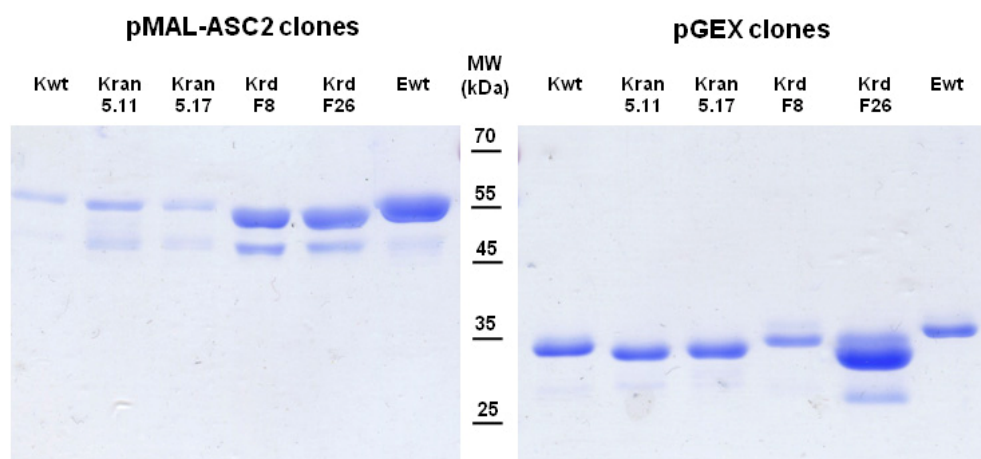


Fig. 2.36 – Two 12% polyacrylamide gels loaded with the MBP- and GST-fusion proteins after purification: loaded 1 μ l of each purified protein.

Interestingly, the coils cloned into pMAL-ASC2 and pGEX vectors showed a apparently little differences in molecular weight: as for the p ω proteins, this feature can be ascribed to differences in peptides sequences and therefore total charge.

To verify whether the ORFs of mutant coils were correctly maintained after the subcloning step, purified products were analyzed for reactivity to SV5 and FLAG tags in western blot assay, for MBP and GST fusion proteins respectively. The same amount (1 μ l) of the previous Coomassie was loaded for each protein..

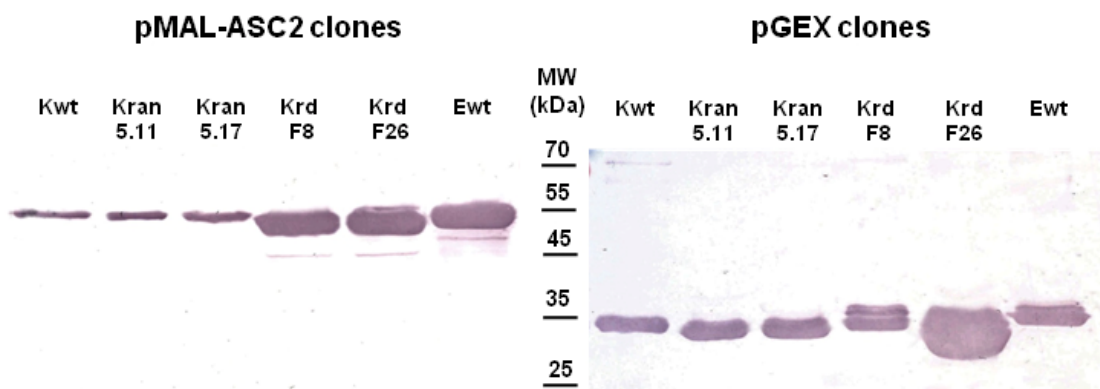


Fig. 2.37 – Western blot assays to assess the correct ORF of MBP- and GST-fusion proteins, subcloned into the new expression vectors. Primary antibody: mouse anti-SV5 1:2000 (pMAL-ASC2 clones), mouse anti-FLAG 1:5000 (pGEX clones). Secondary antibody: goat anti-mouse AP conjugated 1:2000.

5.2 Validation of the system

The validation of the interaction was made through ELISA experiments. Briefly, the GST-fusion molecules were immobilized on the surface of 96-multiwell plates; after a blocking step to avoid unspecific interactions, the MBP-fusion molecules were added and the reactivity to SV5 tag was analyzed after extensive washes. With this assay, only the dimerizing coil pairs can be detected. The results are reported in the figure 2.38.

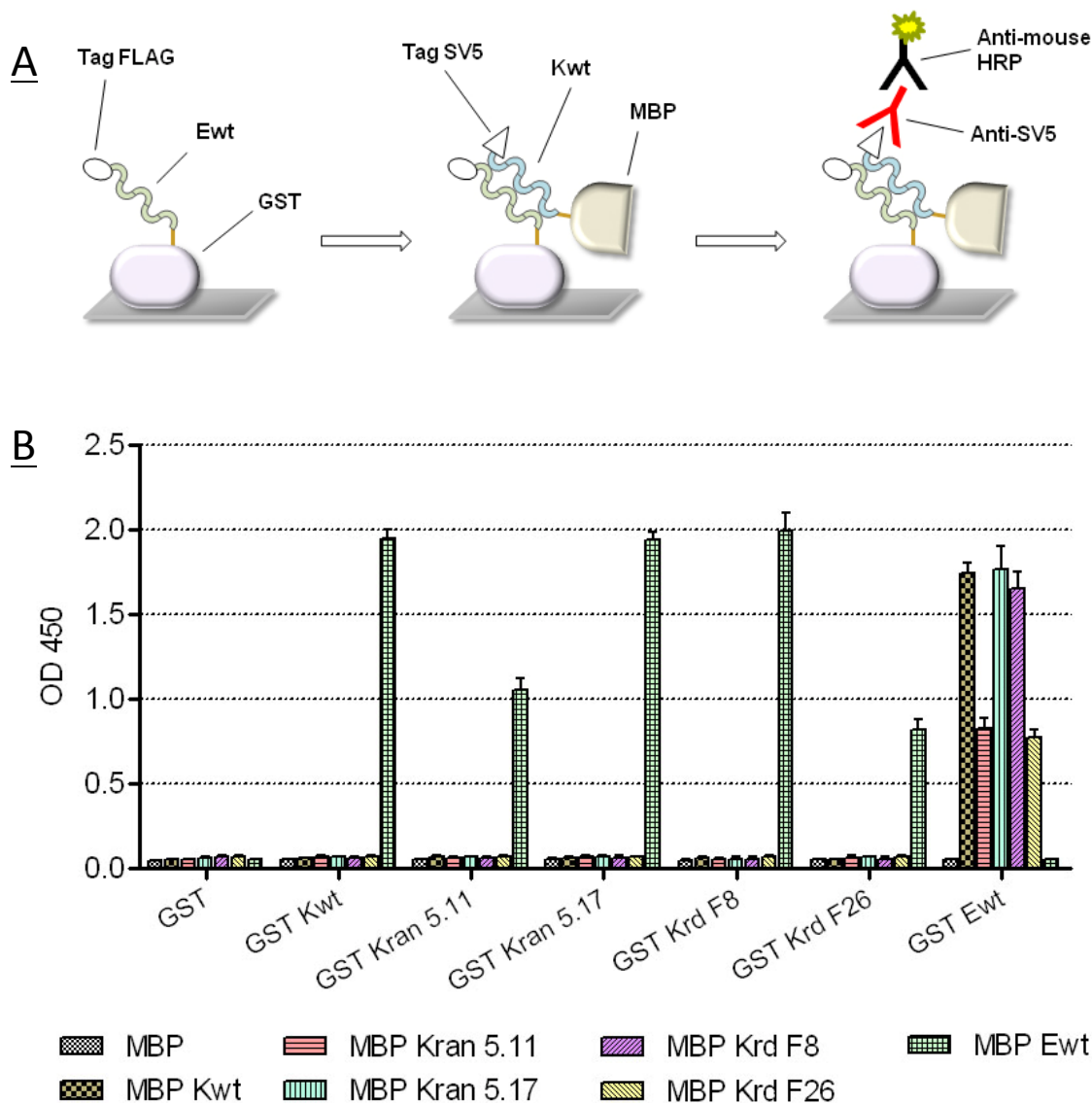


Fig. 2.38 – (A) Schematic representation of the ELISA assay: the GST-fusion proteins were firstly coated, the MBP-fusion proteins were added and the reactivity to SC5 was assessed. (B) The dimers were assayed for reactivity to SV5 tag: only the dimerizing coil pairs can be detected.

	GST Ewt	GST Kwt	GST Kran 5.11	GST Kran 5.17	GST Krd F8	GST Krd F26	GST
MBP	-	-	-	-	-	-	-
MBP Kwt	++	-	-	-	-	-	-
MBP Kran 5.11	+	-	-	-	-	-	-
MBP Kran 5.17	++	-	-	-	-	-	-
MBP Krd F8	++	-	-	-	-	-	-
MBP Krd F26	+	-	-	-	-	-	-
MBP Ewt	-	++	+	++	++	+	-

Fig. 2.38 - (C) Summary of the interaction: (++) strong interaction, (+) interaction and (-) no interaction. The red labeled cells are the ones supposed to be positive. The Ewt/Kwt dimer value was set to (++)

The results reported here confirm the reactivity between the selected coils, isolated with the previous complementation assays. The graph clearly demonstrates how the reactivity was limited exclusively to the E/K heterodimerization, because the E/E or K/K combinations gave no signal, ruling out any possible homodimerization process. The Kran 5.17 and Krd F8 coils showed a comparable reactivity respect Kwt to the Ewt coil, whereas the interaction of Kran 5.11 and Krd F26 to Ewt is lower. MBP and GST proteins were considered as negative control.

In conclusion, the mutated coils represent valid scaffolds for biosensing applications because they confirmed their specific affinity both *in vivo* and *in vitro*; and also that that the dimerization between the interactors is an active process independent from β -lactamase-fragments. The lower values obtained in the ELISA of some coils, that in the PCA showed a good dimerization activity, can be ascribed to the presence of the fusion protein MBP and GST, and their steric hindrance: further analysis using the coils alone must be accomplished, for example using the Biacore system.

Chapter 6

The single chain E-K

6.1 Setup of a biosensor unit: the single chain E-K

In the previous chapters, the structure of E and K coils was analyzed and modified, and the interaction activity of new peptides was verified with *in vivo* and *in vitro* methods (PCA and ELISA). The interaction was improved, and this allows to conceive customizable biosensors based on the dimerization of a coiled-coil domain.

The next step of this project was the generation of a new protein scaffold, that can be used as biosensor unit in many applications. Modifying the sequence of the coils, protein variants can be generated to bind different specific targets. The main features of a new scaffold must be:

- the capability to bind small molecules;
- small size, for an easy production and use on a biosensor;
- the maintenance of stability in several conditions of pH, ionic force and temperature;
- the construction of a small binding site, composed by few amino acids.

The modification of the central heptad to generate glycine-mutants and the p ω -K random library allowed to isolate and select new peptides with improved binding sites. The next passage will be the creation of a new E/K coiled-coil domain as a unique fusion protein, with the two coils covalently linked to each other. The central heptad of both coils will be completely modified so that the apolar core of the central heptads will become a binding pocket to accommodate small molecules or compounds.

Starting from the structure of the coiled-coil domain proposed by Chao and coworkers [135, 136], some modification were performed:

- 1) a linker was inserted between the coils: it is composed by eight residues (seven glycines and one threonine), that confer a marked flexibility to the whole molecule;
- 2) the K sequence was inverted, to allow the interaction through a parallel association as in the original model;
- 3) one cysteine was inserted in each of the peptide: these residues allow the formation of a intrachain disulfide bonds, which facilitate the association between coils and stabilize the scaffold after the modification of the central heptads.

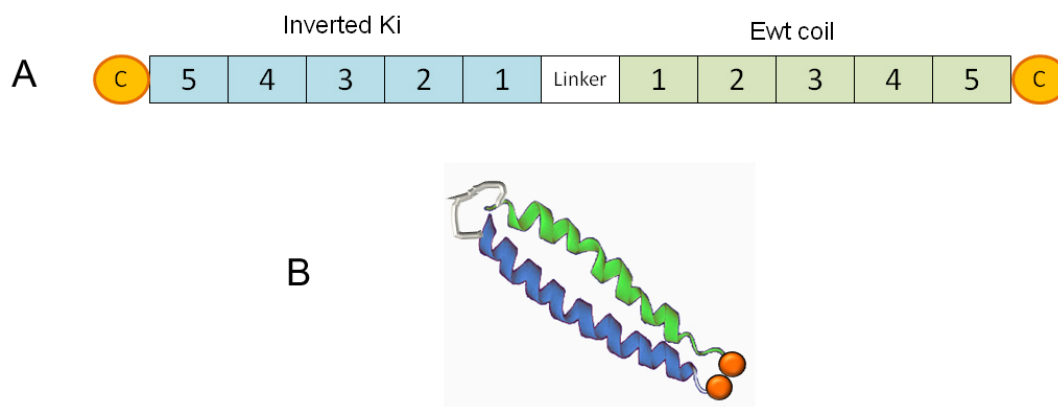


Fig. 2.39 – (A) Schematic linear representation and (B) three-dimensional model of the single chain E-K (C = cysteine; 1-5 = coil heptads, which for K are inverted).

The E and K coils, connected by a loop, are able to fold into helical domains and spontaneously dimerize to form a left-handed super-coil, the single chain E-K (scE-K). This structure was used to construct a library with a randomization-based mutagenesis of the central heptads of both peptides, as done for the $\rho\omega$ -Krandom library. This randomization was introduced in order to obtain a huge collection of scaffolds with several different pockets able to bind different ligands. The randomized library of the scE-K was subsequently cloned into an appropriate vector to perform a selection based on the phage display technique. The single chains are presented on the surface of a filamentous bacteriophage, and the selection is carried out against specific molecules. The main molecule considered in this study was caffeine.

Caffeine is a xanthine alkaloid, that acts as a psychoactive stimulant drug and a mild diuretic (figure 2.47-A) [216]. It is a stimulant drug of the central nervous system, but if consumed excessively it could be teratogenic because causes the inhibition of DNA repair and activity of cyclic AMP phosphodiesterase [217]. Considering the potential harmful effects of caffeine, the ability to dose it accurately and rapidly could be of particular interest for industrial processes, or in food and clinical chemistry.

6.2 The pB vector

The first step was the set up of a proper phage display vector for the construction of large and functional libraries. Generally, phagemid vectors are used for the generation of huge collections of genes, each of them is under control of an inducible promoter. The presentation of the peptide of interest on the surface of the phage is guided by its expression in conjunction with the g3p surface protein, in the form of a recombinant polypeptide. The vector used for the scE-K library was the phagemid vector pB, consisting of the following cassettes:

- the origin of replication (ORI);
- a polylinker site at the upstream of the g3p phage ORF;
- the inducible lacZ promoter, to control the expression of the scE-K;
- the ORF encoding for the g3p;
- the leader sequence to control the correct secretion into the periplasmic space;

- the ORF encoding β -lactamase for antibiotic selection;
- the wild type sequences of E and inverted K-coils (see above paragraph);

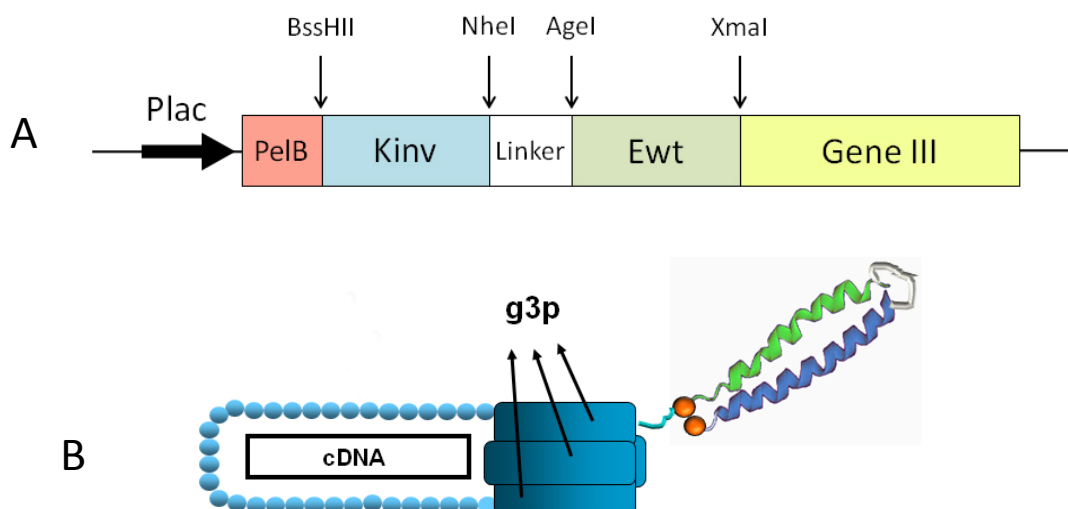


Fig. 2.40 – (A) Schematic representation of pB vector. The expression of single chain E-K is under the control of a Lac promoter; the translocation into the periplasmic space is under the control of the PelB leader sequence. (B) Phage particle displaying the single chain E-K in fusion with the gene 3 protein (g3p).

6.3 Construction of pB scE-K library

On the basis of the advantages deriving from the use of phage display, an single chain E-K library was created using a phagemidic vector. The randomization of the central heptad of each coil allowed to create a collection of single chains able to bind different ligands. In fact, each single chain has a different binding pocket with different ligand affinities.

Considering the extension and the available residues for randomization, the number of possible scaffolds was $1,63 \times 10^{18}$. In fact, one heptad was randomized in each coil, and two modified coils had to match each other. However, this variability is too much high to be obtained with this technique.

Once having detected the vector, the next step was the design of correct primers to generate the library by PCR amplification of fragments. As explained in paragraph 6.1, the target regions for randomization were the central heptads of both E and Kinv coils.

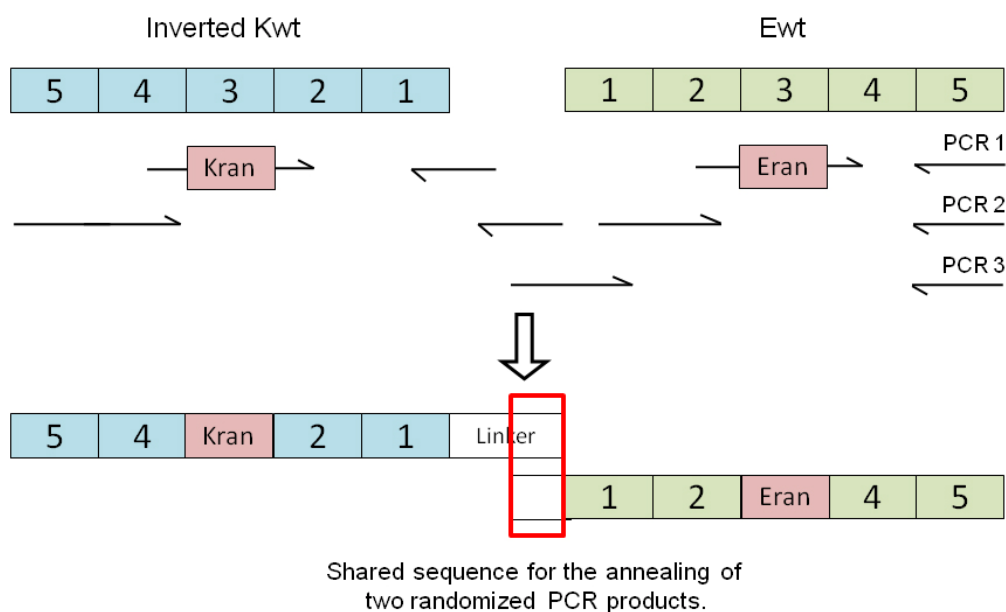


Fig. 2.41 - Schematic representation of the PCR rounds necessary for the construction of inverted Kwt and Ewt randomized coils. The red rectangle highlights the shared sequence included with the last round of amplification: this shared sequence was exploited to assemble the PCR products, generating scE-K library.

The modification was introduced using the same strategy applied to obtain the $\rho\omega$ -K random library, mutating randomly each position into the central heptad (**a-b-c-d-e-f-g**) with the aid of degenerated primers (NNK triplets). The library was created by assembling Kinv and E randomized coils in the single chain format.

The pool of PCR fragments and the pB vector were both digested with BssHIII and XmaI restriction enzymes. After purification, fragments and vector were ligated, and the ligation product was used to electroporate *E. coli* DH5 α competent cells. Bacteria were subsequently plated onto six ampicillin plates. After 15 hours of incubation at 37°C, cells were harvested and stocked with 20% glycerol at -80°C.

The extension of the library was 2×10^8 independent clones. This result did not assure the presence of all possible single chains, but the size was considered satisfying for this technique, and it was analyzed for subsequent rounds of selection.

6.4 pB scE-K library characterization

The diversity and integrity of this library was tested in several ways. First of all, more than 20 single chain E-Ks were tested by PCR from randomly selected bacterial clones. This showed that 21 clones contained a full length insert, while 5 of them were deleted.



Fig. 2.42 - Electrophoresis of PCR-amplified scE-K DNA of 26 randomly picked clones. Each clone, expressing a different scE-K fragment, was amplified with K2-PCR-sense and E1-PCR-anti primers. 100bp DNA markers, ranging from 100 to 1000 bp, are reported.

Furthermore a sequencing of 12 clones was necessary to check the diversity of the library, because a fingerprinting test was impossible to perform for the small size of randomized heptad.

Clone	Kinv (central heptad)	E (central heptad)	K sequence	E sequence
1	WT	Ran	E K L A S V K	V L T W W P L
2	Ran	Ran	L L L G R R S	L R N S R Y Y
3	Ran	Ran	P G G L R R W	R A F P S I R
4	WT	Ran	E K L A S V K	S F L Y V F W
5	Ran	Ran	F L T L A R I	S W K S Y V Y
6	Ran	Ran	L L I C L G G	A V V W V S E
7	WT	Ran	E K L A S V K	R S G M K L K
8	Ran	Ran	S L V G G N W	Q C G V F V F
9	Ran	Ran	P S S V G F C	C V V K M V F
10	Ran	Ran	C F R W M V F	D L L G G I I
11	Ran	Ran	V L L F D M R	F V C F G T W
12	Ran	Ran	V F Q C V L G	R A C S L V C

Table 2.43. – Summary of the randomized-central heptad sequencing of 12 scE-K clones. Three single chains carry the wild type sequence of the inverted K-coil (in bold).

The table above shows that 3 clones out of 12 have a wild type K-coil and a E-coil with the central heptad correctly randomized. The other clones possess both randomized peptides. This library was then used for standard phage display selection steps.

6.5 Selection of the phage display library

6.5.1 Description of selection approach

To select and isolate a single chain-expressing clone, the ligand of interest must be first coated on surface. The coating can be realized in two different ways::

- 1) using the biotin-streptavidin method;
- 2) with direct immobilization on an active surface able to bind amino groups.

The molecules used for the selection carry a linker sequence as a spacer between the coating surface and the real ligand. The linker carries a biotin molecule or a free amino group at one extremity, depending on the selected coating method. The first method needs a preliminary step for streptavidin immobilization onto a surface, after which, the biotinylated-ligand tightly binds to immobilized streptavidin.

Several cycles of selection with different protocols were performed to determine the best selection method (data not shown). On the basis of the results obtained in the preliminary experiments, a series of selections were programmed with an additional pre-incubation step. The library was incubated with the non-coated surface and binding buffer without the ligand of

interest: this passage was introduced to avoid the selection of unspecific clones, removing them from the phage population.

6.5.2 Selection of the pB scE-K library by magnetic beads

Several selection strategies were tested: the use of paramagnetic beads allowed to isolate positive clones during preliminary selections, demonstrating a good experimental performance. The selection strategy is depicted in figure 2.44.

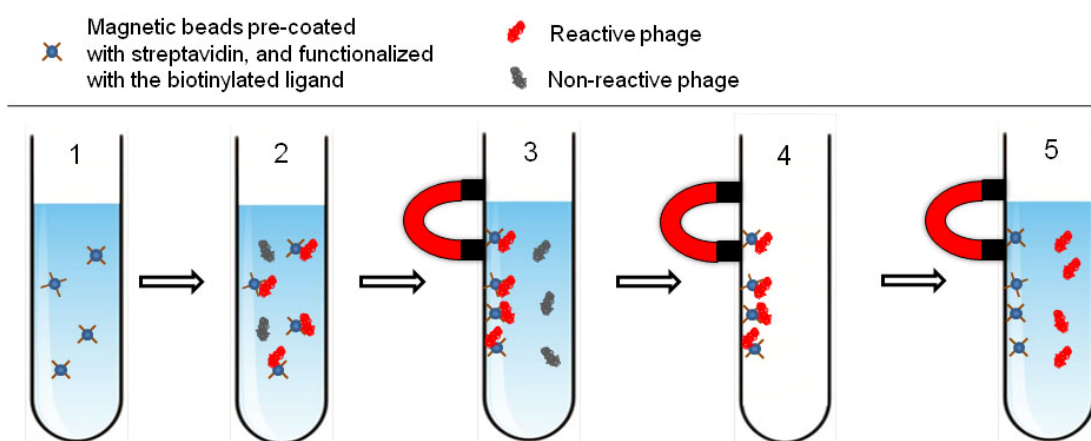


Fig. 2.44 - Selection strategy used to select phages reactive to biotinylated caffeine: (1) functionalization of magnetic beads with the biotinylated caffeine; (2) incubation with phages: reactive phages (labeled in red) bind the caffeine; (3) application a magnetic field to separate the non-reactive phages; (4) extensive washes to remove unbound phages; (5) elution of the reactive phages.

This selection methods can be reassumed in 5 steps:

1. pre-coating of magnetic beads with streptavidin, and incubation with biotinylated-ligand;
2. incubation of phages with the beads;
3. application of a magnetic field to collect beads bound phages;
4. extensive wash passages to remove unbound phages;
5. elution of bound phages by competition with pure, non-biotinylated ligand.

For each round of selection, 10^{12} phages were resuspended in 250 μ l of PBS. Phages were then pre-incubated with 2% non-fat milk (MPBS, the buffer used for the selections) to remove non-specific single chains expressing phages. Then, they were incubated in a mixture containing biotinylated caffeine bound to paramagnetic streptavidin-coated beads and shaken for 30'. After extensive washes with PBS / 0,1% Tween (PBST) and PBS (to remove all unbound and unspecific phages), elution was performed incubating beads for 30' with a 100 μ g/ml solution of caffeine without linker nor biotin.

After this selection, 9 ml of DH5 α at OD_{600nm} 0.5 were infected with the recovered phages, at 37°C for 45'. Bacteria were then plated and allowed to grow over night. The day after, bacteria were harvested and stocked at -80°C in glycerol. A small aliquot was immediately let to grow in order to produce new phages for a second round of selection. Four passages of selections were carried out, with increasing stringency of washing steps.

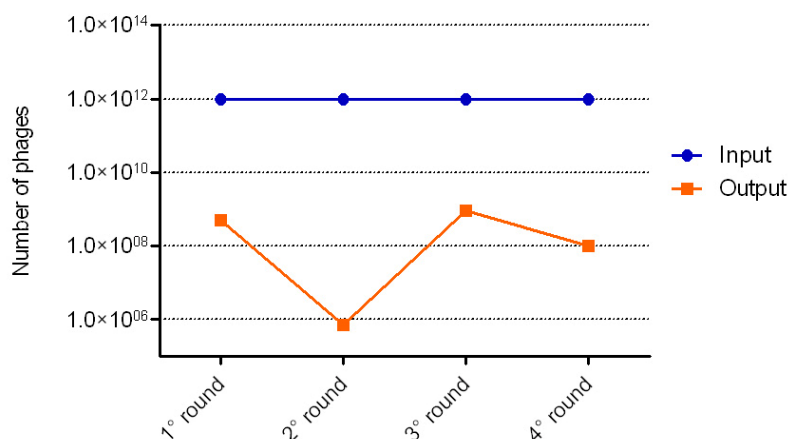


Fig. 2.45 – Input and output values for each round of selection.

For each step, an input of approximately 10^{12} phages was used. However, the final output of phages during the second round of selection was remarkably lower than the first one. This decreased number of recovered phages is a direct result of the increased stringency of the selection procedure. The third and fourth cycles maintained a high yield of recovered phages, probably because of the amplification of the previously selected clones.

6.6 Validation of selection strategy

After the fourth round of selection, 86 individual phage clones from the selected population were tested by phage ELISA against their specific antigen (biotinylated caffeine), and streptavidin, excluding any non specific reactivity. Out of these 86 clones, at least six showed a good reactivity to caffeine, but with a detectable reactivity with streptavidin. In order to reduce the background due to unspecific interactions, the PBS Tween 0,3% was used as buffer (figure 2.46).

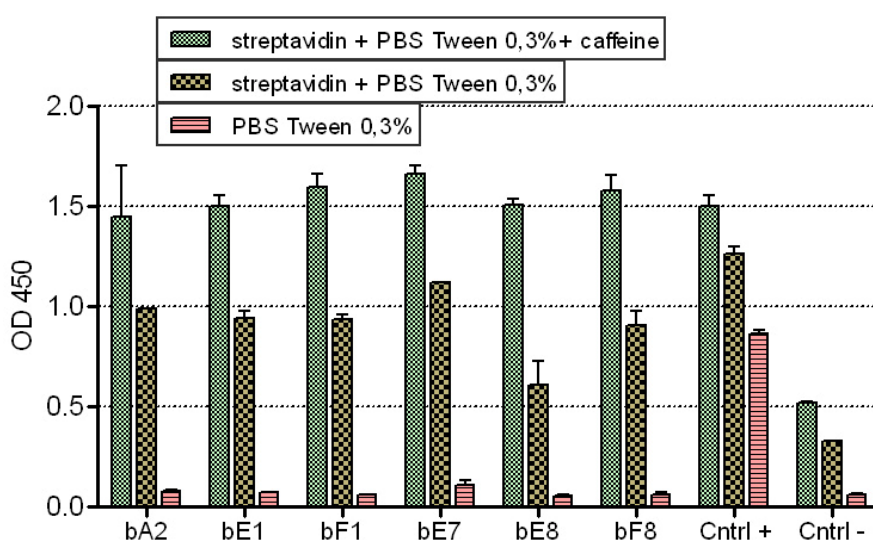


Fig. 2.46 – Reactivity of the previously isolated clones to caffeine, streptavidin and PBST 0.3%. The positive control was a reactive single chain, previously isolated; the negative control was the single chain E-K with the wild type coil sequence. Secondary antibody: anti-M13 HRP conjugated 1:2000.

These experiments demonstrated that there was specific interaction between the single chains in fusion with g3p protein of phages and caffeine, even if there is a still low detectable reactivity against streptavidin. To verify that the six selected clones were specific only for caffeine molecule, they were analyzed searching for a putative reactivity against menthone (a structurally related molecule) or to the biotinylated linker alone, to rule out its possible involvement in ligand-coil interaction. :

The results obtained with this assay confirmed that the isolated single chains are merely specific for caffeine: they interact with high affinity and high specificity with the ligand of interest, excluding any unwanted cross-reaction (figure 2.47).

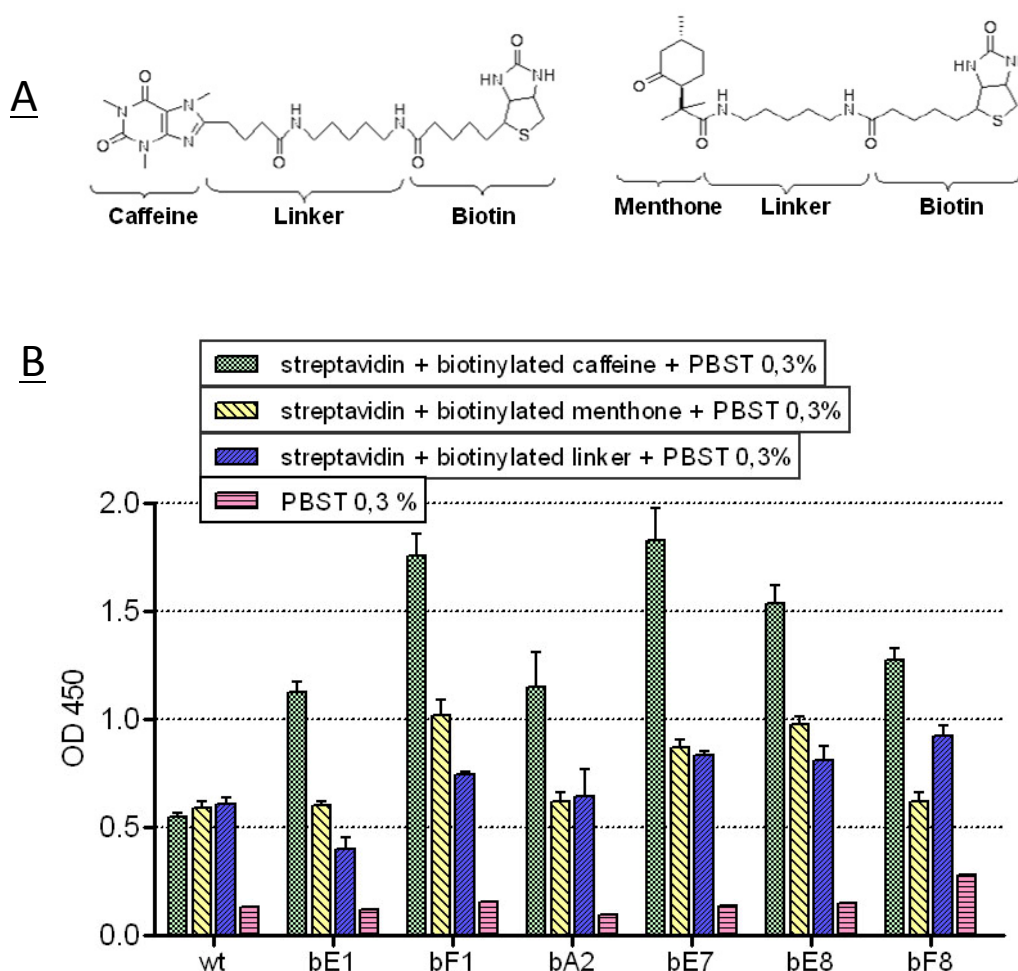


Fig. 2.47 – (A) Structure of biotinylated caffeine and menthone. (B) Reactivity of the isolated clones to caffeine, menthone, biotinylated linker and PBST 0.3%. The single chain E-K with the wild type coil sequence was considered as negative control. Secondary antibody: anti-M13 HRP conjugated 1:2000.

A second further validation to ensure the specificity of interaction between single chains and caffeine was done by using NUNC Immobilizer wells. These plates possess surfaces coated with a photoprobe consisting of electrophilic groups that can rapidly form a covalent bond with nucleophilic groups, like amines: their density is optimized for immobilizing peptides and proteins. NUNC plates allow the direct immobilization of proteins on ELISA wells, avoiding the use of streptavidin and also minimizing steric hindrance to optimize binding interactions.

The molecule of caffeine was modified to facilitate a covalent bond on NUNC Immobilizer wells and biotin was replaced by a reactive amino group. The buffer used for each step was PBS Tween 0,3%.

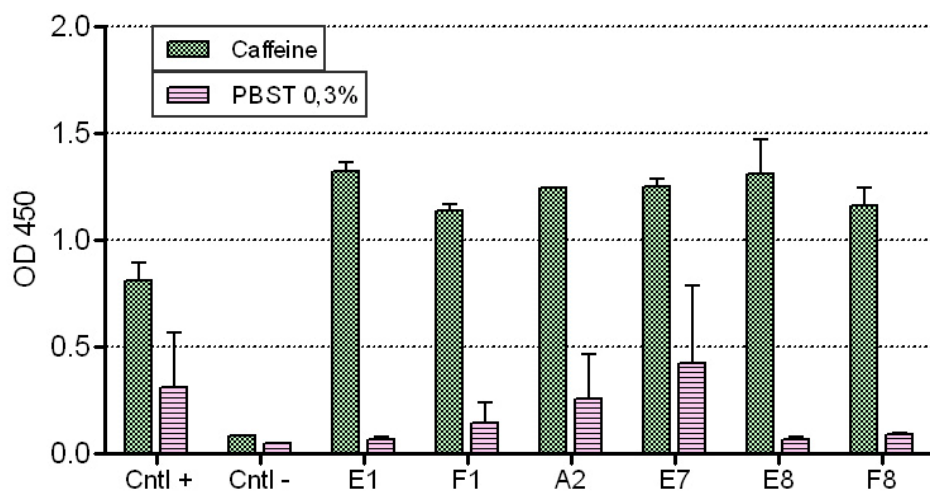


Fig. 2.48 - Reactivity of the isolated clones to caffeine, immobilized onto NUNC Immobilizer wells by using a reactive amino group. The positive control was a reactive single chain, previously isolated; the negative control was the single chain E-K with the wild type coil sequence. Secondary antibody: anti-M13 HRP conjugated 1:2000.

The results reported here finally show that values obtained with biotin-ligand molecule were independent from the linker sequence and from the streptavidin used for immobilization of biotinylated caffeine.

Discussion

In the last years the development of highly sensitive biosensors has considerably increased. In fact, the need of bioanalytical devices has gained importance for several applications, such as healthcare, diagnostics, food quality, drug delivery, environmental monitoring and many other research fields [2, 4, 218]. The demand is centered on small-size biosensors characterized by high sensitivity, high specificity and good accuracy for fast and reliable results.

Nanotechnology is revolutionizing the development of new biosensors, based either on the improved proprieties of new materials, or on the proprieties that standard materials show only at nanoscale. Miniaturized platforms allow to broaden the potential applications of bioanalytical devices, providing new reliable instruments that are cheaper, easier to use and, above all, smaller [8, 15]. The production of systems with high multiplexing potential is one of the most important aims in bio-nanotechnology research, allowing the integration of many sensors in one single platform. Nanotechnologies can help to produce biosensors with great advantages, such as the need of small sample volumes, that allows in the detection of biomolecules even at very low concentrations.

Engineered antibodies are currently the most widespread detecting tool for biotechnological, medical and pharmaceutical applications, because of the deep knowledge of their structure, binding capabilities and specificity of action. However, the emerging field of protein engineering allowed the production of a wide range of scaffolds different than antibodies, with various origins and characteristics, for protein library construction and for grafting experiments of binding sites [79, 80]. The proprieties of these new scaffolds are quite similar to the ones characterizing antibodies, i.e. specificity and affinity, but some features are improved, like a reduced size (the smaller scaffold is the knottin Min-23, only 23 amino acids), a binding pocket restricted to few residues and a good stability even without disulfide bonds. All these characteristics are suitable to create a customizable nanotechnological device with enhanced spatial resolution, that can be further improved to construct biosensors-based detecting system for any application.

A biosensor is defined as an analytical device that converts the energy released during binding between molecules into an electrical signal. The binding substrate is often of biological nature, therefore the choice of proper molecules to be used as recognition unit is fundamental. The miniaturization of the detecting system requires the employment of small molecules, and new scaffold proteins are the most promising tool for this purpose: the increasing knowledge of

their structure, and the wide potential of site-specific modifications on demand, allows to develop platforms of high scientific and commercial impact.

The aim of this work was the development and the improvement of new scaffold proteins for biosensing applications, and to this purpose two biological methods have been adopted: the protein complementation assay (PCA) and the phage display. In fact, we have demonstrated how they could represent a powerful selection tool and could offer a reliable help for the production of novel biosensors for environmental survey, biomedical applications and for every case in which high sensitivity and specificity are requested.

In the last five years many scaffolds have been taken in consideration for different purposes, sometimes replacing antibodies [85]. Each scaffold has well-defined features and properties that can be easily modified or even improved, in order to confer it an adequate flexibility in terms of application potential. For all these reasons, coiled-coil based scaffolds are recognized among the most promising instruments in the biosensor field [137]. This domain is present in a variety of proteins involved in different cellular processes, about in the 10% of the total proteins expressed by a single cell. Its peculiar structure has been extensively studied to characterize the dimerization dynamics and overall the tolerance to sequence modification, a fundamental pre-requirement to use coiled-coils as biosensor units.

In fact, the possibility to modify the amino acid sequence allows the improvement of binding features. Considering their characteristics, coiled-coils have been used as affinity purification tags, library presentation scaffolds and as bases for the creation of miniantibodies [114, 148]

To this purpose, this work has focused the attention on a *de novo* heterodimeric coiled-coil domain, named E/K. Starting from typical features of coiled-coil containing proteins, Tripet and Chao [114, 136] were able to design *de novo* this heterodimeric system. E and K peptides consist in a single amino acid heptad repeating five times: the hydrophobic residues in the apolar core and the charged residues on the external surface allow the strong interaction between them to form a unique right-handed super-coil. The physical-chemical properties of the domain have been deeply analyzed with circular dichroism to assess the high binding affinity and the remarkable thermal and chemical stabilities. All these properties render the E/K system an ideal base to create functional biosensors. Moreover, the hetero-dimerization capability is a positive feature that allows to combine different E and K peptide to produce customizable platforms.

In the present work, the stability of E/K coiled-coil domain was tested with protein complementation assay, in order to validate *in vivo* the data obtained with circular dichroism. Additional modifications and experiments were performed with the final aim of improving the interaction between the coils, in order to create E/K pairs with increased binding affinities or recognition of new specific ligands.

The first step was the *in vivo* validation of the E/K interaction system by using the cprotein complementation assay [170]. As previously described, PCA is based on a reporter protein rationally dissected into two subunits, unable to refold spontaneously, that are fused to two proteins that are thought to bind to each other. The refolding of the reporter protein from these fragments is catalyzed by the binding of the two supposed interactor proteins, and is detected as

reconstitution of its enzymatic activity. In this work, the use of PCA based on β -lactamase was described and extensively used, and the interaction among the proteins was easily detected by the bacterial growth on ampicillin-supplemented broth. The $p\alpha$ and $p\omega$ vectors were used to allow the expression of proteins fused to α and ω fragments of the β -lactamase.

The Ewt and Kwt coils were expressed by the $p\alpha$ and $p\omega$ vector, respectively. The growth rate of bacteria expressing these peptides was compared to control cells deficient for antibiotic resistance (negative control) and to cells constitutively expressing a wild-type β -lactamase (positive control), in order to evaluate any differences in enzymatic activity between wild-type and reconstituted β -lactamase. We first confirmed that only bacteria containing interactors can grow on selective pressure conditions, and no growth was observed for cells expressing proteins that were not able to interact. Even if the reconstituted enzyme in Ewt/Kwt expressing bacteria displayed a reduced hydrolytic activity than intact-enzyme strains, the growth rate on different antibiotic concentrations was anyhow considered acceptable.

The results obtained in these assays provide the demonstration that β -lactamase-based PCA is a reliable method to evaluate protein-protein interactions: the results derived are in good agreement with previous results obtained with circular dichroism.

When developing a biosensor, one has to take into consideration the use of highly stable molecules that do not change their properties in presence of extreme conditions, such as low/high pH, temperature and ionic strength, conditions that are often required for the storage or usage of a biosensor unit. This is why a major goal of this work was the selection of resistant E/K interactions able to tolerate extreme experimental settings. Any alteration of stability requires an alteration of the coil domain, that, however, needs to maintain its basic folding. Therefore, before the selection of new coiled-coil interactors we verified that this E/K domain effectively represents a protein scaffold, able to tolerate amino acid substitutions without altering its main structure.

First of all, glycine substitutions were performed, and the ability of mutated peptides to form helical coils as well as the quality of these novel interactions was verified through PCA. Theoretically, glycine insertion would lead to the severe impairment of the propensity to form helical coils, resulting in loss of E/K interaction. Our results clearly demonstrate not only that the E/K system tolerates sequence modifications, but also that the K-coil is more resistant to mutations showing an interaction capability even superior to wild type peptides. Therefore, this coil can be fully qualified as protein scaffold and PCA could be considered a valid methodology to select and validate proteins with scaffolding characteristics.

Once having demonstrated that the K-coil is more prone and resistant to sequence modifications than E-coil, we created two K-mutant libraries and selected mutants displaying improved interaction properties, as done in several other studies concerning different proteins [106, 149, 180]. The mutations were introduced by randomization either on the central heptad of K-coil, to generate K random library (Kran), or on the whole helix with a rational approach (i.e. mutation of crucial residues for dimerization, introducing selected amino acids), to generate a K rational design library (Krd). From these two libraries we selected and isolated bacterial clones expressing mutated K-coils displaying improved binding affinity to Ewt than the Kwt peptide.

All these interactions allowed bacterial survival even at high antibiotic concentrations, in some cases in prohibitive conditions for Ewt and Kwt expressing cells

The isolation of coil variants with better affinity and stability than the original ones is a clear demonstration of how the combination between structural modification of proteins and PCA allows a fast and effective way to select interactors with desired properties, but still retaining the original, native coiled-coil structure. The same strategy could be applied also for E-coil, even if it has demonstrated to be more unstable when glycine residues were inserted into the sequence. In this way we can progressively create a new coiled-coil dimer with improved interaction capability or ligand affinity for different applications.

Moreover, PCA and site directed mutagenesis can be rationally employed to characterize protein-protein interactions and map the domains involved in this phenomenon. In this way, not only a map of protein networks can be created (interactome), but also a fine characterization of the subdomains involved.

These coils will be finally employed on bioanalytical platforms, therefore outside the natural cellular environment: for this reason is necessary to evaluate the activity of E and K coils in an isolated *in vitro* environment. Chimeric coils have been purified from selected bacterial clones, and used for ELISA tests. The high interaction selectivity seen *in vivo* has been confirmed also *in vitro*: on one side, homodimerizations were not detected, on the other, the specific interactions between E and K peptides were measured as well. However, none of the interactions between mutated K-coils and wt E-coil displayed better interaction ability respect to wt peptides. This result is in conflict with *in vivo* evidences, because the same mutated K-coils conferred a better antibiotic resistance than Kwt-expressing bacteria (i.e. better β -lactamase activity deriving from better coiled-coil interaction).

These contrasting data can be ascribed to a different stability or impaired binding propensity between coils because of the experimental conditions used during ELISA tests (for example ionic strength or pH). Moreover, for *in vitro* assays, E and K peptides were purified as GST and MBP fusion chimeras: these tags might potentially interfere with the correct folding of E and K coils, resulting in a loss of signal.

Finally, to correctly evaluate the improvements of the interaction affinity of this system, other methods such as Biacore. The Biacore biosensor can measure protein-protein interaction and binding affinity of unlabeled interactors in real time, obtaining the individual association and dissociation rate constants.

The dimerization propensity and tolerance to sequence modifications of a coiled-coil system allowed us the construction of a new scaffold structure able to bind small molecules, with several analytical implications. This new structure based on E and K coils has been engineered to generate a single molecule containing both peptides, the single chain E-K (scE-K).

During this project we created an scE-K library, based on the randomization of the central heptad to allow the recognition and the binding of small molecules. The library contamination with Kwt coil was not considered critical, because the totality of E-coils were mutated, still allowing the creation of a binding pocket.

Phage display methodology [193] was used to isolate single chains reactive with the desired analyte, caffeine. Thanks to this technique, the single chain library was expressed in fusion with a surface capsid protein, and after several selection cycles we isolated single chain mutated peptides carrying a binding pocket highly specific for caffeine. We chose this molecule because it is small, easy to handle and of particular interest for industrial processes, or in food and clinical chemistry [216].

The selection process was carried out on magnetic beads that displayed a good functionality in preliminary experiments (data not shown): a pre-selection step was performed to clean out phages that unspecifically cross-reacted with the magnetic support or the buffer employed for the selection. Clones were then selected for a specific reactivity against caffeine, and the interaction was validated using a strictly related molecule (menthone) that was not recognized by caffeine reactive phages. This technique offers a high potential of selection of clones reactive against any desired molecule or compound, from toxins to heavy metals to diagnostic markers, because each library virtually contains any possible combination of mutated E and K peptides able to bind a wide range of molecules. One other advantage is that with the aid of phage display one can directly select protein scaffolds able to bind the analyte of interest bypassing time-consuming 3D modeling of interactions between coils and ligand. Moreover, this technique could be potentially applied to any other protein scaffolds based library.

The selection techniques employed in this research project have demonstrated to be valid and effective to evaluate molecular interactions and isolate protein interactors with features suitable for biosensing applications. In fact, an analytical platform containing a biological component is highly reliable because it is based on the extreme specificity of interaction between the “sensor” and the compound of interest (for example between antibody and antigen). In our case, both E/K single chains and K mutant coils can be immobilized on nano-platforms like nanoarrays for the employment in all those biotechnological applications that require miniaturization and high sensitivity and specificity. Moreover, the advances in the field of array printing (see chapter 1.6) have notably ameliorated the production of nano-platforms: the recent techniques (e.g. PISA, NAPPA and DAPA) allow the *in situ* co-transcription/translation of DNA products in proteins [69]. This feature enables the rapid, economic and “on demand” creation of biosensors, maintaining their properties unaltered, without needing to purify the molecules to be immobilized, as the ones isolated and characterized in the present work.

Materials and methods

Abbreviations

AP, alkaline phosphatase
APS, ammonium persulfate
BCIP, 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
B-PER, bacterial protein extraction reagent
BSA, bovine serum albumine
DMSO, dimethylsulfoxide
dNTPs, deoxynucleotides
DTT, dithiothreitol
HRP, horseradish peroxidase
IPTG, isopropyl β -D-1-thiogalactopyranoside
MW, molecular weights
NBT, nitro-blue tetrazolium
O/N, over night
PEG, polyethylenglicole
PPB, periplasmic buffer
RT, room temperature
SDS-PAGE, sodium dodecyl sulphate - polyacrylamide gel electrophoresis
TMB, tetramethylbenzidine
tTG, tissue transglutaminase

Solutions and buffers

- **Phosphate buffered saline (PBS)**
8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, 0.24 g KH₂PO₄ in 1000 ml H₂O, final pH 7.4.
- **Phosphate buffered saline Tween 0.1% (PBST)**
PBS added with 0.1% Tween 20
- **Milk Phosphate buffered saline (MPBS)**
PBS added with 2% non-fat milk powder.

- **2XTY liquid broth for bacteria**
6.4g Bacto-tryptone, 4g Extract Yeast, 2g NaCl, H₂O in 400 ml H₂O. If required ampicillin 0.1 mg/mL, kanamycin 0.05 mg/m, 0.034 mg/ml chloramphenicol.
- **2XTY Agar plates**
6.4g Bacto-tryptone, 4g Extract Yeast, 2g NaCl, 6g Bacto-Agar in 400 ml H₂O. If required ampicillin 0.1 mg/mL, kanamycin 0.05 mg/m, 0.034 mg/ml chloramphenicol.
- **TAE buffer for DNA electrophoresis on agarose gels**
0.04 M Tris-acetate, 0.001 M EDTA.
- **Loading buffer 6x for DNA samples (agarose gel)**
40% glycerol, 60% H₂O, 0.10% (w/v) bromophenol blue.
- **SDS Running Buffer for protein electrophoresis on acrylamide gels**
25 mM Tris, 250 mM glicin, 0.1% SDS pH 8.3.
- **Loading buffer 2 for protein samples**
100mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blu, 2% β-mercaptoethanol, 20% glycerol
- **Running gel for SDS-polyacrylamide gel**
12% polyacrylamide mix (29% acrylamide, 1% bisacrylamide), 0.375M Tris pH 8.8, 0.1% SDS, 0.1% APS, 2 μl TEMED (N,N,N',N'-Tetramethylethylenediamine), H₂O to 10 ml.
- **Stacking gel for SDS-polyacrylamide gel**
5% polyacrylamide mix (29% acrylamide, 1% bisacrylamide), 0.125M Tris pH 6.8, 0.1% SDS, 0.1% APS, 1 μl TEMED (N,N,N',N'-Tetramethylethylenediamine), H₂O to 4 ml.
- **Loading buffer 2x for protein samples (acrylamide gels)**
100 mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blue, 2% β-mercaptoethanol, 20% glycerol.
- **AP buffer**
Tris 1M pH9,5, NaCl 100mM, MgCl₂ 5mM
- **Lysis buffer**
20 mM Tris pH 8.00, 500mM NaCl, 0.1% Triton x100
- **TSS for preparation of competent E. coli cells**
85% 2xTY, 10% PEG mw 8000, 5% DMSO, 50 mM MgCl₂. Filtration with 0.2 μm filter
- **Glycerol buffer for preparation of electrocompetent E. coli cells**
10% glycerol in milli-Q H₂O. Autoclave before use.
- **PPB for extraction of periplasm fraction**
200 mg/mL Sucrose, 1mM EDTA, 30 mM TrisHCl pH 8.0.

Bacterial strains

The bacterial strains used in this study were:

- *Escherichia coli* DH5αF' (Gibco BRL), F'/endA1 hsd17 (rK-mK+) supE44 thi-1 recA1 gyrA (Nalr) relA1 _(lacZYA-argF) U169 deoR (F80dlacD-(lacZ)M15);

- *Escherichia coli* BL21-CodonPlus-RP strain: B F- *ompT hsdS*(rB- mB-) *dcm+* *Tetr gal* *endA Hte* (*rgU proL Camr*);

Oligonucleotides

All primers were purchased from Tib Molbiol – Roche.

<i>Name</i>	<i>Sequence</i>
Kwt-sense	ctc tag cgc gca tgc caa agt aag cgc tct caa gga aaa ggt atc cgc gct taa aga gaa ggt atc tgc act caa gga g
Kwt-anti	tga ctg cta gcc tct ttc agt gcg ctc act ttc tct ttg aga gca gaa act ttc tcc ttg agt gca gat ac
Ewt-sense	Ctc tag aat tcg aag tta gtg cgc tgg aga agg agg tta gcg ctc ttg aga aag agg tat ccg cgc tcg aaa aag ag
Ewt-anti	tga ctg cta gcc ttc tca aga gcg ctc act tcc ttt tca agc gca gaa acc tct ttt tcg agc gcg gat acc tc
E-mutgly-sense	gcg ctt aaa gag ggc ggc tct gca ggt ggt gag aaa gtt tct gct ctc
E-mutgly2-sense	gaa gtt agt gcg ctg gag aag gag gtt agc gct ctt gag
E-EcoRI-sense	agt cga att cga agt tag tgc gct gga g
E-HindIII-anti	agc tca agc ttt act tct caa gag cgc tca c
K-mutgly-sense	gcg ctt aaa gag ggc ggc tct gca ggt ggt gag aaa gtt tct gct ctc
K-mutgly2-sense	aaa gta agc gct ctc aag gaa aag gta tcc gcg ctt aaa gag
K-BssHIII-sense	ctc tag cgc gca tgc caa agt aag cgc tct caa g
K-NheI-anti	tga ctg cta gcc tct ttc agt gcg ctc ac
Egly-sense	gcg ctt aaa gag gga gga ggt gga ggt gga ggt gag aaa ggt tct gct ctc
Kgly-sense	gcg ctt aaa gag gga ggt gga ggt gga ggt gag aaa gtt tct gct ctc
Kmut1-sense	gta tcc gcg ctt aaa gag nnk nnk nnk nnk nnk nnk nnk aaa gtt tct gct ctc
VHPT2	tgg tga tgg tga gta cta tcc agg ccc agc agt ggg ttt g
VLPT2	tac cta ttg cct acg gca gcc gct gga ttg tta tta ctc
K-deg sense	act ggc gcg cat gcc vrn gta agc gct vyn vrn gaa vrn gta tcc gcg vyn vrn gag aag gta tct gca ctc aag gag
K-deg-anti	atg cgc tag cct cny bnr btg cgc tca cny btt cny bnr bag cag aaa cny bct cct tga gtg cag ata c
M13-20	gta aaa cga cgg cca gtg
E-BssHIII-sense	act ggc gcg cat gcc gaa gtt agt gcg ctg gag
E-NheI-anti	ctg agc tag cct tct caa gag cgc tcac
K-inv-sense	agt cgc gcg cat gcc gag aaa ctg gca agc gtg aaa gag aaa ctc gct tct gtt aaa gag aag ctc gca tct gta aag gag
K-inv-anti	atg cgc tag ctt tta cgc tag cga gct ttt cct tta cgg acg caa gtt tct cct tta cag atg cga gct tct c
K1-PCR-sense	aaa gag aaa ctc gct tct gtt aaa nnk nnk nnk nnk nnk

	nnk nnk gag aaa ctt gcg tcc gta aag
K1-PCR-anti	tga cta ccg gta ccg cct ccg cct ttt acg cta gcg ag
K2-PCR-sense	ctc tag cgc gca tgc cgc gtg cgg cgg aac tag tga gaa act ggc aag c
K2-PCR-anti	cgc act aac ttc acc ccc acc ggt acc gcc tcc gcc
E1-PCR-sense	gtt agc gct ctt gag aaa nnk nnk nnk nnk nnk nnk nnk gag gtt tct gcg ctt gaa
E1-PCR-anti	tga ctt cta gac ccg ggg cag ccg cca cca ccc ttc tca aga gcg ctc ac
E3-PCR-sense	acc ggt ggg ggt gaa gtt agt gcg

Table 3.1 – Sequences of all oligonucleotides used in this work

Experimental methods

PCR (Polymerase Chain Reaction)

The proof-reading Phusion Pyrococcus-like DNA polymerase (Finnzymes) was used for construction of coils sequences and the libraries.

Reaction mixture:

- Template DNA 0.01-1 ng (plasmidic DNA)
- Sense primer 0.5 pmol/μl
- Antisense primer 0.5 pmol/μl
- Phusion HF Buffer 5x
- dNTPs (Sigma) 0.25 mM
- Polymerase 0.025 units/μl
- H₂O to 50 μl

The following cycles were performed:

- Denaturation step, 30' at 98°C.
- 31 cycles of: denaturation, 15'' at 98°C; annealing, 15'' at 60°C; elongation, 30'' every 1000bp at 72°C.
- Final elongation step: 10' at 72°C.

Thermus termophilus DNA polymerase (Biotools) was used to validate the cloning steps.

Reaction mixture:

- Template DNA 0.01-1 ng (plasmidic DNA)
- Sense primer 0.5 pmol/μl
- Antisense primer 0.5 pmol/μl
- Biotools Buffer 10x
- dNTPs (Sigma) 0.25 mM
- MgCl₂ 2 mM
- Polymerase 0.025 units/μl

- H₂O to 20 µl

The following cycles were performed:

- Denaturation step, 5' at 94°C.
- 31 cycles of: denaturation, 45'' at 94°C; annealing, 45'' at 60°C; elongation, 1' every 1000bp at 72°C.
- Final elongation step: 10' at 72°C.

DNA electrophoresis on agarose gel

Agarose (Sigma) gels with a concentration of 2-1,5% in TAE buffer were used to separate PCR products; 0,8% agarose gels were used to separate plasmidic DNA preparations. 1µl of ethidium bromide (Sigma) was added to 30 ml of solution. 100 and 1000 base-pairs molecular weights were purchased from Fermentas.

DNA purification

The GenElute Gel Extraction Kit and GenElute PCR Clean-up kit (Sigma) were used for purification of DNA agarose gel and reaction mixtures respectively, following the instruction of the manufacturer.

The NucleoTrap CR (Machery-Nagel) was used for purification of mixtures of libraries ligations.

DNA digestion with restriction endonucleases

All restrictions endonucleases (BssHII, NheI, EcorI, HindIII, XmaI, AscI) were purchased from New England Biolabs.

Reaction mixture:

- DNA;
- NEB buffer 10x;
- BSA 100x (if necessary);
- Restriction endonuclease, 1unit/ug of DNA;
- H₂O to 50 µl.

The incubation was performed at the temperature required by the specific enzyme.

DNA Ligation

Vector and insert DNA were mixed at the experimental-required ratio, as number of molecules (1:3 – 1:5). T4 ligase from New England Biolabs was used:

Reaction Mixture:

- DNA
- T4 ligase buffer (NEB) 10x;
- T4 ligase (NEB), 1 unit;
- H₂O to 15 ul;

Incubation O/N at 16°C.

Plasmidic DNA extraction

The GenElute Plasmid Miniprep (Sigma) was used for plasmidic DNA mini-preparations, starting from 1-5 ml of bacteria culture, and following the instruction of the manufacturer.

DNA sequencing

PCR products were purified with Eppendorf Perfectprep Gel Cleanup kit. Reaction mixtures for sequencing was composed as follows:

- 50-100 ng of purified PCR product;
- 1 μ l primer (3,2pmol/ μ l);
- 2 μ l Terminator Mix (Applied Biosystems, BigDye Terminator v1.1 Cycle Sequencing Kit);
- 2 μ l buffer 5x;
- H₂O to 10 μ l;

and the sequencing program was:

- 1' at 96° ;
- 25 cycles: 15'' at 96°; 5'' at 50°; 4' at 60° ;

Reactions were afterwards purified with CENTRI SEP Spin Columns, following manufacturer instructions. 5 μ l of purified sequences were loaded on sequencing plates with 10 μ l of formamide, denaturated for 2' at 96° and analyzed with 3100 Genetic Analyzer sequencer (ABI PRISM-HITACHI).

Preparation of competent *E. coli* cells

50 ml of *E. coli* cells were grown at 37°C in 2XTY liquid broth to OD₆₀₀ 0.5. The bacteria were immediately chilled in ice for at least 20', then centrifuged at 4°C for 5' at 1200 rpm. The supernatant was discarded and the bacterial pellet resuspended in 5 ml of sterile TSS. The cells were immediately used or stocked at -80°C for two months.

Bacterial transformation

A 100 μ l vial of competent cells was incubated for 20' with 5 μ l of ligation reaction mixture. Heat shock was applied at 42°C for 1'15''; after 2' in ice, the bacteria were resuspended in 1 ml of liquid 2XTY and allowed to grow for 1 hour. No antibiotic was added. Bacteria were then plated on antibiotic-containing agar plates and grown O/N at 28-30°C.

Preparation of electrocompetent cells

For the preparation of electrocompetent cells, the following protocol was applied:

- from an O/N culture of bacteria, 200 μ l of culture was added to 200 ml of 2XTY liquid broth at 37°C to OD₆₀₀ 0,8;
- the whole culture was splitted in 4 50ml-tubes, and chilled for 30';
- centrifugation at 5000 rpm, at 4°C for 8';
- the supernatant was discarded, and pellet was resuspended with 5 ml of glycerol buffer;
- glycerol buffer was added to 50 ml and the tubes were chilled for 10';
- centrifugation at 5000 rpm, at 4°C for 10';

- the supernatant was discarded, the pellet resuspended with 5 ml of glycerol buffer and the suspensions collected in two 50ml-tubes;
- glycerol buffer was added to 50 ml and the tubes were chilled for 10’;
- centrifugation at 5000 rpm, at 4°C for 10’;
- the supernatant was discarded, the pellet resuspended with 5 ml of glycerol buffer and the suspensions collected in one 50ml-tubes;
- glycerol buffer was added to 50 ml and the tubes were chilled for 10’;
- centrifugation at 5000 rpm, at 4°C for 10’;
- the supernatant was discarded and the pellet resuspended with 125 µl of glycerol buffer for every 100 ml of bacterial culture;
- cells were immediately used or stocked at -80°C for 2 months.

10 pg of purified pUC119 vector (Invitrogen) was used to test the efficiency of cells, generally about 10^{10} transformed clones/µg DNA

Bacterial transformation of electrocompetent cells

The ligation mixture were purified and used to transform electrocompetent bacteria. A Eppendorf 2510 electroporator were used. 1 µl of purified ligation product was added to one aliquot of cells (35-40 µl), and chilled in ice for 30’’. The bacteria were transferred into a cuvette (Eppendorf): after the electroporation cycle (at 1800 volts), the bacteria were resuspended with 1ml of 2XTY, allowed to grow at 37°C for 1 hour and plated on antibiotic-containing agar plates and grown at 28-30°C.

Preparation of bacteria for -80°C stock

The bacteria were collected and resuspended with a solution of 80% 2XTY and 20% glycerol.

Protein electrophoresis on polyacrylamide gel and western blot

Protein samples, diluted 1:1 in loading buffer 2x for proteins, were loaded on 12% polyacrylamide gel, composed by a stacking and running gel. The proteins were separated at 18-20 mA.

Coomassie staining

The SDS-polyacrylamide gel was removed from the glass and stained with Coomassie solution, under gentle shacking for 1 hour at 37°C. Then, Coomassie solution must be removed (it can be reused many times) and the gel rinsed with destaining solution on a slow shaker for 6-12 hours.

Western blot

Proteins were separated by SDS-PAGE and transferred onto nitrocellulose membrane (Schleicher-Schuell) by semidry blotting using the Trans-Blot SD Transfer Cell (Biorad). The membrane was treated as follows:

- blocked with MPRB 1 hour at RT (20°C);

- primary antibody: mouse anti-FLAG 1:5000 (Sigma), mouse anti-His 1:5000 (Abcam), mouse anti-SV5 mAb 1:2000 (1mg/ml); 1 hour at RT;
- two washes with PBST and two washes with PBS;
- secondary antibody: goat anti-mouse AP conjugated (Jackson Immunoresearch), 1 hour at RT;
- two washes with PBST and two washes with PBS;

Immunocomplexes were revealed by adding the chromogenic substrate NBT-BCIP (Sigma) in AP buffer, and the plate read at OD₄₅₀.

Bradford protein assay

The following protocol was used:

- add 30 sample volumes of Bradford reagent (Sigma) to protein sample: mix the solution and incubate at RT for 20’;
- make the “blank”, adding the Bradford reagent to the buffer used for the solubilization of proteins;
- make a dilution series of BSA (bovine serum albumin, Sigma) in the range 50 ng-10µg;
- cool samples to RT and measure the absorbance at 562 nm;
- plot the average-blank corrected measurement for each standard vs. concentration. Use the slope of this standard curve to estimate the protein concentration of the unknown samples.

Expression of recombinant proteins

Clones were picked from agar plates and grown at 37° in 2 ml of 2XTY. At OD₆₀₀ 0.5, IPTG 0,2 mM was added to induce the production of the recombinant protein, allowing growth O/N at 28°C.

To obtain the total fraction, 100 µl of bacterial culture were centrifuged at 5000 rpm for 5’, the supernatant discarded, and the pellet resuspended with 50 µl of PBS.

To obtain the periplasmic fraction, the remaining bacterial broth were centrifuged at 5000 rpm for 5’, and the supernatant discarded. The pellet was resuspended in 1:10 of culture volume of PPB, chilled in ice for 1 hour, vortexing every 10’. The mixture was centrifuged at 11000 rpm for 15’. The supernatant containing the periplasmic proteins (periplasmic fraction) was collected.

Ewt, Kwt and glycine-mutant coils construction

All primers are reported in table 3.1.

Two pairs of primers were used to construct the Ewt and Kwt sequences: Kwt-sense/Kwt-anti, and Ewt-sense/Ewt-anti. The sense and antisense primers were annealed together by their complementary sequences; the resulting amplicates were cloned as BssHII/NheI (Kwt) and EcoRI/HindIII (Ewt) fragments into the p ω and p α vector respectively.

The glycine mutant coils (Emut, Kmut, Egly and Kgly) were generated with three subsequent cycles of amplification, where the product of the previous amplification is the template of the subsequent cycle. The templates for the first amplification were p α -Ewt and p ω -Kwt. The following combination of primers was used:

- Emut: Emutgly-sense, Emutgly2-sense and E-EcoRI-sense (first, second and third PCR amplification); E-HindIII-anti as antisense primer for all three cycles;
- Kmut: Kmutgly-sense, Kmutgly2-sense and K-BssHII-sense (first, second and third PCR amplification); K-NheI-anti as antisense primer for all three cycles;

The first primer sense was replaced by Egly-sense and Kgly-sense for the construction of Egly and Kgly coils respectively. Each final amplification step produced fragments with the appropriate restriction sites at the ends: BssHII/NheI for K-coils, EcoRI/HindIII for E-coils. The resulting amplicates were cloned into the p ω and p α vector respectively.

Each vector was transformed in *E. coli* competent cells, and the bacteria plated onto plates containing the antibiotic (chloramphenicol 34 μ g/ml for p ω constructs, kanamycin 50 μ g/ml for p α constructs). After an incubation O/N at 30°C, the bacteria were picked, checked and stocked at -80°C with 20% glycerol.

Preparation and transformation of p ω -Kran library

To produce the Krandom fragments, three subsequent cycles of PCR were performed, using the following combination of primers: Kran1-sense, Kmutgly2-sense and K-BssHII-sense (first, second and third PCR amplification), and K-NheI-anti as antisense primer for all three cycles. The template for the first amplification was p ω -Kwt vector. Final amplification step produced fragments with BssHII/NheI restriction endonucleases sites respectively at 5' and 3' ends.

p ω vector (2 μ g) and K random fragments (800 μ g) were digested using BssHII and NheI restriction endonuclease. After purification, they were ligated with T4 DNA ligase in a mixture at a 1:4 ratio (number of molecules). The ligation was performed O/N at 16°C. The enzyme was heat-inactivated at 60°C for 10', and the ligation product was purified using NucleoTraP CR Kit (Machery-Nagel) and resuspended in 20 μ l water. The transformation into *E. coli* cells was performed with the 2510 Electroporator (Eppendorf), by transforming 15-20 aliquots of electrocompetent cells, each one with about 1 μ l of DNA. After electroporation cells were pooled, allowed to grow in 5 ml of 2XTY liquid broth for 1 hour at 37°C, and then plated onto 6 25cm²-plates with 34 μ g/ml chloramphenicol. Dilution were plated onto a small plate with 34 μ g/ml chloramphenicol to assess the size of the library. To verify the absence of contamination for the subsequent PCA test, a little amount of bacteria was plated on plates containing 100 μ g/ml ampicillin.

After 15 hour of incubation at 30°C, libraries were inspected by PCR harvested and stocked in small aliquots at -80°C with 20% glycerol.

Preparation and transformation of p ω -Krd library

Two pairs of degenerated primers were used to construct the K rational design sequences: K-deg-sense and K-deg-anti. The sense primer shares a sequences with the complementary antisense primer, that allows an assembly-PCR step. The produced fragments carried BssHII/NheI restriction endonucleases sites, respectively at 5' and 3' ends.

p ω vector (2 μ g) and Krd fragments (800 μ g) were digested using BssHII and NheI restriction endonuclease. The same ligation-transformation procedure of p ω -Kran library was applied for this library.

After electroporation cells were pooled and grown in 5 ml of 2XTY liquid broth, for 1 hour at 37°C, and then plated onto 6 25cm²-plates with 34 µg/ml chloramphenicol. Dilution were plated onto a small plate with chloramphenicol to assess the size of the library. To verify the absence of contamination for the subsequent PCA test, a little amount of bacteria was plated on plates containing 100 µg/ml ampicillin.

After 15 hour of incubation at 30°C, libraries were inspected by PCR, harvested and stocked in small aliquots at -80°C with 20% glycerol.

PCA-validation of Ewt, Kwt and glycine-mutant coils.

The PCA was performed to test the interaction activity of the wild type and glycine-mutant coils. DH5α were co-transformed as listed in the table.

DH5α carrying vector:	Co-transformed with:
pα-Ewt	- pω-Kwt - pω-Kgly - pω-2.8
pω-Kwt	- pα-Egly - pα-tTG
pω-Kmut	- pα-Emut

Table 3.2 – The table shows the different combination of DH5α co-transformants cells.

After incubation for 1 hour at 37°C without selective pressure, the co-transformed cells were plated onto 2XTY agar plates supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol, and incubated O/N at 30°C.

Single clones of each co-transformed bacteria were grown in 2XTY broth, supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol, at 37°C to OD₆₀₀ 0.5. Serial 1:10 dilutions of the original bacterial culture were done: 20 µl of 10⁻³, 10⁻⁴, 10⁻⁵, 10⁻⁶ and 10⁻⁷ dilutions were plated, obtaining spots with distinguishable single colonies, either onto 2xTY agar plates supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol for titering, or onto plates supplemented with increasing concentrations of ampicillin (50–100–200–400–800–1200–1600 mg/ml) and 1 mM IPTG. The plates were incubated at 28°C for 48 hours. The bacterial growth of each dilution was analyzed and compared in all the plates: the comparison of the bacteria growth on agar plates allowed to assess the ratio “co-transformed/interactors” for each ampicillin concentration. The results of the bacteria plated on chloramphenicol/kanamycin plates represented 100% of growth. The data can be reported in a graph.

PCA-validation of K random and K rational design coils.

Electrocompetent DH5α cells, carrying the vector pα-Ewt, were electroporated with 500 ng of either pω-Kran or pω-Krd: the bacterial culture were equally plated onto a 25cm²-agar plate supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol, and onto 6 25cm²-agar

plates supplemented with increasing concentrations of ampicillin (200-400-800-1200-2000-3000 µg/ml) and 1 mM IPTG. The plates were incubated at 28°C for 48 hour. Dilutions were also plated onto a small 2xTY agar plate (25cm²) containing kanamycin and chloramphenicol for titering.

Selected bacteria, grown at the highest concentration of ampicillin, were picked, checked by PCR and grown to OD₆₀₀ 0.5. Serial 1:10 dilutions of the bacterial culture were done: 20 µl of 10⁻³, 10⁻⁴, 10⁻⁵, 10⁻⁶ and 10⁻⁷ dilutions were plated obtaining spots with distinguishable single colonies, either onto agar plates supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol for titering, or onto plates supplemented with increasing concentrations of ampicillin (200-400-800-1200-2000-3000 mg/ml) and 1 mM IPTG. The plates were incubated at 28°C for 48 hours.

The bacterial growth of each dilution was analyzed and compared in all the plates: the comparison of the bacteria growth on agar plates allow to assess the ratio “co-transformed/interactors” for each ampicillin concentration. The results of the bacteria plated on chloramphenicol/kanamycin plates represent 100% of growth. The data can be reported in a graph.

pMAL-ASC2 and pGEX expression vectors sub-cloning

The AscI endonuclease site of pMAL-ASC2 is compatible with the overhangs of BssHII site. Therefore, the fragments Kwt, Kran 5.11, Kran 5.17, Krd F8 and Krd F26 were directly subcloned as BssHII/HindIII fragments from the p ω vectors into pMAL-ASC2 vector, previously digested with AscI/NheI. Primers were designed to introduce the appropriate restriction endonuclease sites at the ends of Ewt sequence: the p α -Ewt was amplified with E-BssHII-sense and E-NheI-anti.

Fragments were digested with BssHII/NheI and subcloned into pMAL-ASC2 and pGEX vectors, ligated, purified and transformed into *E.coli* competent cells. The obtained clones were inspected by PCR and stocked at -80°C.

Production and purification of GST- and MBP-fusion proteins

Bacterial clones were grown in 100 ml of 2XTY liquid broth, added with 100 µg/ml ampicillin, at 37°C to OD₆₀₀ 0.5; the expression of the recombinant protein was induced by adding 0.2 mM IPTG, and grown O/N at 28°C. The next day, periplasmic fractions containing the recombinant proteins were extracted as follows:

- the bacterial culture was centrifuged for 20' at 4000 rpm (4°C), the supernatant discarded and the pellet resuspended with 10 ml of lysis buffer/gram of bacteria;
- the mixture was incubated in ice, in gentle shaking, for 15';
- the lysozyme was added at the mixture, 400 µg/ml of lysate, and the solution was incubate in ice, in gentle shaking, for 30';
- DNase (50-100 ug/mL of lysate), PMSF (1mM) and EDTA (1mM) were added to lysate, and incubated in ice, in gentle agitation for 45';
- Sonication (Bandelin Sonoplus) was performed: 3 cycles, 1-2 minutes for cycle;
- After centrifugation for 20' at 11000 rpm, the supernatant was collected and filtered with 0.45 µm filter.

To purify recombinant proteins, amylose resin (NEB) for MBP-fusion proteins and glutathione Sepharose (GE Helthcare) for GST-fusion proteins were used following the instruction of the manufacturer to perform affinity chromatography. Briefly:

	MBP purification	GST purification
Binding	1 ml of amylose resin was added to the lysate, and incubated for 1 hour in rotation	400 μ l of glutathione sepharose was added to the lysate, and incubated for 1 hour in rotation;
Packing	The resin was packed in a 10 ml-column, and extensively washed with a solution of 20 mM TrisHCl pH 8.0, 500 mM NaCl	The resin was packed in a 10 ml-column, and extensively washed with PBST and PBS pH 8.0
Elution	The proteins were eluted with a solution of 10 mM (Sigma) (0.5 ml each elution)	The proteins were eluted with a solution of 50mM GSH (Sigma), 100mM NaCl, in PBS (0.250 ml each elution).

Each fraction were dialyzed O/N to PBS.

MBP- and GST-fusion proteins ELISA assay

The ELISA assay with purified proteins was performed as follows:

- purified GST proteins was diluted to 10 μ g/ml in PBS and 100 μ l were coated in ELISA wells (Costar), O/N at 4°C;
- the wells were extensively washed with PBS and 120 μ l of blocking solution (MPBS) were added to each well, incubating at RT for 1 hour;
- purified MBP proteins were diluted 10 μ g/ml in PBS, and added 100 μ l to each well;
- extensive washes were performed with PBST and PBS;
- primary antibody: mouse anti-SV5, diluted 1:2000 in MPBS. 100 μ l to each well, at RT for 1 hour;
- extensive washes were performed with PBST and PBS;
- secondary antibody: goat anti-mouse HRP conjugated (Jackson Immunoresearch), diluted 1:2000 in MPBS. 100 μ l to each well, at RT for 1 hour;
- after extensive washing with PBST and PBS, the immunocomplexes were revealed with the chromogenic substrate TMB (Sigma), and the plate read at OD₄₅₀.

Preparation and transformation of pB scE-K library

Each coil was separately constructed and randomized.

Kinv-sense and Kinv-anti primers were used to construct the inverted K-coil. The sense and antisense primers were annealed together by their complementary sequences; the amplified fragment was BssHII/NheI cloned into the p ω vector.

By using the p ω -Kinv as template, two pairs of primers were used to produce the K inverted and randomized fragments: K1-PCR-sense/K1-PCR-anti and K2-PCR-sense/K2-PCR-anti. Two subsequent cycles were performed. The E randomized fragments were separately produced,

performing three subsequent cycles of PCR amplification, and using only one antisense primer: E1-PCR-sense, E-mutgly2-sense, E3-PCR-sense and E1-PCR-anti.

After purification of final PCR products, the library was created by assembling Kinv randomized fragments with E randomized fragments. The assembly was performed using specific cycles of PCR amplification:

- 9 cycles with Kinv and E randomized fragments mixed at 1:1 ratio (number of molecules): this step allowed the assortment of fragments and the assembly;
- 21 cycles: using external primers (K2-PCR-sense /E1-PCR-anti), the single chain sequences were re-amplified, and the restriction endonucleases sites were included at the ends (BssHII/XmaI).

The scE-K library (320 µg) and the pB vector (1500 µg) were both digested with BssHII and XmaI. After purification, they were mixed in a ligation mixture at a 1:3 ratio. The ligation was performed O/N at 16°C, and purified using NucleoTraP CR Kit (Machery-Nagel) and resuspended in 20 µl water. 15-20 aliquots of electrocompetent cells were transformed, each one with about 1 µl of DNA. After electroporation cells were pooled and grown in 5 ml of 2XTY liquid broth, for 1 hour at 37°C, and then plated onto 6 25cm²-plates with ampicillin 100µg/ml. Dilution were plated onto a small plate with ampicillin to assess the size of the library.

After 15 hour of incubation at 30°C, the library was inspected by PCR, harvested and stocked in small aliquots at -80°C with 20% glycerol.

Production and PEG-precipitation of phages

A small aliquot of bacterial stock of the library was grown into 10 ml of 2XTY, at 37°C to OD₆₀₀ 0.5: ampicillin was added. In these conditions, the bacteria express the *pilus* necessary for phage infection. Bacteria were infected with the helper phage, that carries all the genes necessary for the phage replication: 10 µl of helper phage (10⁹ viral particles/ml) were added each 10 ml of bacterial culture. The infection was performed at 37°C for 45'. After centrifugation at 4000 rpm for 10', the supernatant was discarded, and the pellet resuspended in 10 ml of 2XTY broth added with ampicillin (the resistance of phagemid) and kanamycin (the resistance of helper phage), to select the bacteria containing both the phagemid and the phage genome. Bacteria were grown shaking O/N at 30°C, to produce recombinant phages. To collect them, the bacterial broth was centrifuged at 7000 rpm for 20' and the supernatant recovered.

1:5 volume of PEG-NaCl (20% NaCl, 2.5 M NaCl) was added to supernatant. The mixtures was chilled for 45', shaking by inversion every 10', and then centrifuged at 7000 rpm for 20' to pellet the phages. The supernatant was discarded and the pellet resuspended in PBS.

Library selection with magnetic beads

The phages were produced, PEG-precipitated and resuspended in 250 µl of PBS (about 10¹² phages). Phages were then incubated in a well of ELISA plates (Sarstedt), previously pre-incubated with MPBS (the buffer used for the selection), to remove non-specific phages. The phage solution was moved into another well, still pre-incubated with MPBS: this step was repeated 10 times.

The phage solution was moved in a new well, previously pre-incubated with streptavidin (New England Biolabs) 10 µg/ml, and blocked with MPBS at RT for 10'. The solution was

moved into a new well, treated in the same way: this step was repeated 5 times. 250 μ l of MPBS and biotinylated caffeine (10 μ g for the first round, 50 μ g for the following rounds) were added. To the phage solution: the mixture was incubated and gently shaken at RT for 1 hour. Paramagnetic streptavidin-coated beads were added and shaken for 1 hour. After extensive washes with PBS/0,1% Tween (PBST) and PBS (to remove all unbound and unspecific phages), elution was performed incubating beads at RT for 30' with a 100 μ g/ml solution of caffeine without linker nor biotin..

After this selection, 9 ml of DH5 α at OD_{600nm} 0.5 were infected with the recovered phages, at 37°C for 45'. Then, bacteria were centrifuged at 2500 rpm for 15' (15°C), the supernatant discarded, the pellet the resuspended with 1 ml of 2XTY broth and plated, allowing the growth O/N at 37°C. The day after, bacteria were harvested and stocked at -80°C in glycerol. A small aliquot was immediately let to grow in order to produce new phages for a second round of selection. Four passages of selections were carried out, with increasing stringency of washing steps.

Phage ELISA

The phage ELISA assays, by using the selected phage library, were performed as follow:

- coating: each well was incubated with 100 μ l of coating solutions, containing the ligand (see table 3.3), O/N at 4°C;
- blocking: after extensive washes with PBS, the wells were blocked with either 140 μ l of MPBS or PBS Tween 0.3%, at RT for 1 hour;
- phage incubation: after extensive washes with PBS, 50 μ l of phage solution and 50 μ l of blocking solution were added, at RT for 1 hour;
- extensive washes with PBST and PBS;
- secondary antibody: mouse antiM13 HRP conjugated (Amersham, Pharmacia) was diluted 1:2000 in the blocking solution and added to the wells at RT for 1 hour.
- extensive washes with PBST and PBS;
- Immunocomplexes were revealed with the chromogenic substrate TMB (Sigma), and the plate read at OD₄₅₀.

	Nunc Immobilizer™ Amino NUNC Plates	Sarstedt plates
Coating solutions	10 μ g/mL of ligand solution in 100 mM pH 9.6 sodium carbonate buffer.	1) 10 μ g/mL of streptavidin solution (<i>New England Biolabs</i>), diluted in PBS. 100 μ l each well, O/N at 4°C. 2) 20 μ g/mL of biotinylated ligand, diluted in PBS. 100 μ l each well, 1 hour and 30' a RT.

Table 3.3 – Different coating solutions for Amino NUNC and Sarstedt plates.

References

1. Babu, V.R., et al., *Development of a biosensor for caffeine*. Anal Chim Acta, 2007. **582**(2): p. 329-34.
2. Andreescu, S., et al., *JEM Spotlight: Applications of advanced nanomaterials for environmental monitoring*. J Environ Monit, 2009. **11**(1): p. 27-40.
3. Chaudhry, Q., et al., *Applications and implications of nanotechnologies for the food sector*. Food Addit Contam Part A Chem Anal Control Expo Risk Assess, 2008. **25**(3): p. 241-58.
4. Keusgen, M., *Biosensors: new approaches in drug discovery*. Naturwissenschaften, 2002. **89**(10): p. 433-44.
5. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects*. Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
6. Vo-Dinh, T. and B. Cullum, *Biosensors and biochips: advances in biological and medical diagnostics*. Fresenius J Anal Chem, 2000. **366**(6-7): p. 540-51.
7. Vo-Dinh, T., *Nanobiosensors: probing the sanctuary of individual living cells*. J Cell Biochem Suppl, 2002. **39**: p. 154-61.
8. Jianrong, C., et al., *Nanotechnology and biosensors*. Biotechnol Adv, 2004. **22**(7): p. 505-18.
9. Jain, K.K., *Nanobiotechnology: applications, markets and companies*. Jain PharmaBiotech Publications, 2005.
10. Choi, J.W., et al., *Nanotechnology in biodevices*. J Microbiol Biotechnol, 2007. **17**(1): p. 5-14.
11. Johnson, C.J., et al., *Proteomics, nanotechnology and molecular diagnostics*. Proteomics, 2008. **8**(4): p. 715-30.
12. Ferrari, M., *Cancer nanotechnology: opportunities and challenges*. Nat Rev Cancer, 2005. **5**(3): p. 161-71.
13. Wasinger, V.C., et al., *Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium*. Electrophoresis, 1995. **16**(7): p. 1090-4.
14. Zhu, H., et al., *Global analysis of protein activities using proteome chips*. Science, 2001. **293**(5537): p. 2101-5.

15. Jain, K.K., *Nanotechnology in clinical laboratory diagnostics*. Clin Chim Acta, 2005. **358**(1-2): p. 37-54.
16. Abraham, A.M., et al., *Nanotechnology: a new frontier in virus detection in clinical practice*. Indian J Med Microbiol, 2008. **26**(4): p. 297-301.
17. Cheng, X., et al., *Micro- and nanotechnology for viral detection*. Anal Bioanal Chem, 2009. **393**(2): p. 487-501.
18. Zhao, X., et al., *A rapid bioassay for single bacterial cell quantitation using bioconjugated nanoparticles*. Proc Natl Acad Sci U S A, 2004. **101**(42): p. 15027-32.
19. Yang, H., et al., *Application of atomic force microscopy as a nanotechnology tool in food science*. J Food Sci, 2007. **72**(4): p. R65-75.
20. Zhang, H., et al., *Quantum dots for cancer diagnosis and therapy: biological and clinical perspectives*. Nanomed, 2008. **3**(1): p. 83-91.
21. Son, A., et al., *Rapid and quantitative DNA analysis of genetic mutations for polycystic kidney disease (PKD) using magnetic/luminescent nanoparticles*. Anal Bioanal Chem, 2008. **390**(7): p. 1829-35.
22. Tallury, P., et al., *Silica-based multimodal/multifunctional nanoparticles for bioimaging and biosensing applications*. Nanomed, 2008. **3**(4): p. 579-92.
23. Namdeo, M., et al., *Magnetic nanoparticles for drug delivery applications*. J Nanosci Nanotechnol, 2008. **8**(7): p. 3247-71.
24. Singh, R. and J.W. Lillard, Jr., *Nanoparticle-based targeted drug delivery*. Exp Mol Pathol, 2009.
25. Lacoste, T.D., et al., *Ultrahigh-resolution multicolor colocalization of single fluorescent probes*. Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9461-6.
26. Gao, X., et al., *In vivo cancer targeting and imaging with semiconductor quantum dots*. Nat Biotechnol, 2004. **22**(8): p. 969-76.
27. Pinaud, F., et al., *Advances in fluorescence imaging with quantum dot bio-probes*. Biomaterials, 2006. **27**(9): p. 1679-87.
28. Gao, X. and S.R. Dave, *Quantum dots for cancer molecular imaging*. Adv Exp Med Biol, 2007. **620**: p. 57-73.
29. Xing, Y. and J. Rao, *Quantum dot bioconjugates for in vitro diagnostics & in vivo imaging*. Cancer Biomark, 2008. **4**(6): p. 307-19.
30. Michalet, X., et al., *Quantum dots for live cells, in vivo imaging, and diagnostics*. Science, 2005. **307**(5709): p. 538-44.
31. Goldman, E.R., et al., *Multiplexed toxin analysis using four colors of quantum dot fluororeagents*. Anal Chem, 2004. **76**(3): p. 684-8.
32. Park, S.J., et al., *Array-based electrical detection of DNA with nanoparticle probes*. Science, 2002. **295**(5559): p. 1503-6.
33. Grimm, J., et al., *Novel nanosensors for rapid analysis of telomerase activity*. Cancer Res, 2004. **64**(2): p. 639-43.

34. Taton, T.A., *Preparation of gold nanoparticle-DNA conjugates*. Curr Protoc Nucleic Acid Chem, 2002. **Chapter 12**: p. Unit 12 2.
35. Nam, J.M., et al., *Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins*. Science, 2003. **301**(5641): p. 1884-6.
36. Goluch, E.D., et al., *A bio-barcode assay for on-chip attomolar-sensitivity protein detection*. Lab Chip, 2006. **6**(10): p. 1293-9.
37. Georganopoulou, D.G., et al., *Nanoparticle-based detection in cerebral spinal fluid of a soluble pathogenic biomarker for Alzheimer's disease*. Proc Natl Acad Sci U S A, 2005. **102**(7): p. 2273-6.
38. Morrow, T.J., et al., *Programmed assembly of DNA-coated nanowire devices*. Science, 2009. **323**(5912): p. 352.
39. Wang, W.U., et al., *Label-free detection of small-molecule-protein interactions by using nanowire nanosensors*. Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3208-12.
40. Nano, I., <http://www.integratednano.com>.
41. Fortina, P., et al., *Nanobiotechnology: the promise and reality of new approaches to molecular recognition*. Trends Biotechnol, 2005. **23**(4): p. 168-73.
42. Barone, P.W., et al., *Near-infrared optical sensors based on single-walled carbon nanotubes*. Nat Mater, 2005. **4**(1): p. 86-92.
43. Wu, G., et al., *Bioassay of prostate-specific antigen (PSA) using microcantilevers*. Nat Biotechnol, 2001. **19**(9): p. 856-60.
44. Ji, H.F., et al., *Molecular recognition of biowarfare agents using micromechanical sensors*. Expert Rev Mol Diagn, 2004. **4**(6): p. 859-66.
45. Lam, Y., et al., *Using microcantilever deflection to detect HIV-1 envelope glycoprotein gp120*. Nanomedicine, 2006. **2**(4): p. 222-9.
46. Laurell, T., et al., *Microfluidic components for protein characterization*. J Biotechnol, 2001. **82**(2): p. 161-75.
47. Lacorte, S. and A.R. Fernandez-Alba, *Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food*. Mass Spectrom Rev, 2006. **25**(6): p. 866-80.
48. Pan, C., et al., *Recent developments in methods and technology for analysis of biological samples by MALDI-TOF-MS*. Anal Bioanal Chem, 2007. **387**(1): p. 193-204.
49. Terracciano, R., et al., *Selective binding and enrichment for low-molecular weight biomarker molecules in human plasma after exposure to nanoporous silica particles*. Proteomics, 2006. **6**(11): p. 3243-50.
50. Gaspari, M., et al., *Nanoporous surfaces as harvesting agents for mass spectrometric analysis of peptides in human plasma*. J Proteome Res, 2006. **5**(5): p. 1261-6.
51. Geho, D., et al., *Fractionation of serum components using nanoporous substrates*. Bioconjug Chem, 2006. **17**(3): p. 654-61.

52. Fodor, S.P., et al., *Light-directed, spatially addressable parallel chemical synthesis*. Science, 1991. **251**(4995): p. 767-73.
53. Salaita, K., et al., *Massively parallel dip-pen nanolithography with 55 000-pen two-dimensional arrays*. Angew Chem Int Ed Engl, 2006. **45**(43): p. 7220-3.
54. Kusnezow, W. and J.D. Hoheisel, *Solid supports for microarray immunoassays*. J Mol Recognit, 2003. **16**(4): p. 165-76.
55. Kusnezow, W., et al., *Antibody microarray-based profiling of complex specimens: systematic evaluation of labeling strategies*. Proteomics, 2007. **7**(11): p. 1786-99.
56. Tang, J., et al., *Recognition Imaging and Highly Ordered Molecular Templating of Bacterial S-Layer Nanoarrays Containing Affinity-Tags*. Nano Lett, 2008.
57. Seidel, M. and R. Niessner, *Automated analytical microarrays: a critical review*. Anal Bioanal Chem, 2008. **391**(5): p. 1521-44.
58. Yu, F., et al., *Surface plasmon field-enhanced fluorescence spectroscopy studies of the interaction between an antibody and its surface-coupled antigen*. Anal Chem, 2003. **75**(11): p. 2610-7.
59. Yu, X., et al., *Label-free detection methods for protein microarrays*. Proteomics, 2006. **6**(20): p. 5493-503.
60. BioForceNano, <http://www.bioforcenano.com>.
61. Schweitzer, B., et al., *Microarrays to characterize protein interactions on a whole-proteome scale*. Proteomics, 2003. **3**(11): p. 2190-9.
62. Boutell, J.M., et al., *Functional protein microarrays for parallel characterisation of p53 mutants*. Proteomics, 2004. **4**(7): p. 1950-8.
63. Jain, K.K., *The role of nanobiotechnology in drug discovery*. Drug Discov Today, 2005. **10**(21): p. 1435-42.
64. Steller, S., et al., *Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections*. Proteomics, 2005. **5**(8): p. 2048-55.
65. Fritzler, M.J., *Advances and applications of multiplexed diagnostic technologies in autoimmune diseases*. Lupus, 2006. **15**(7): p. 422-7.
66. He, M. and M.J. Taussig, *Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method)*. Nucleic Acids Res, 2001. **29**(15): p. E73-3.
67. He, M. and M.J. Taussig, *DiscernArray technology: a cell-free method for the generation of protein arrays from PCR DNA*. J Immunol Methods, 2003. **274**(1-2): p. 265-70.
68. He, M., et al., *Printing protein arrays from DNA arrays*. Nat Methods, 2008. **5**(2): p. 175-7.
69. He, M., et al., *In situ synthesis of protein arrays*. Curr Opin Biotechnol, 2008. **19**(1): p. 4-9.

70. Khan, F., et al., *Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces*. *Anal Chem*, 2006. **78**(9): p. 3072-9.
71. Steinhauer, C., et al., *Improved affinity coupling for antibody microarrays: engineering of double-(His)6-tagged single framework recombinant antibody fragments*. *Proteomics*, 2006. **6**(15): p. 4227-34.
72. Ramachandran, N., et al., *Next-generation high-density self-assembling functional protein arrays*. *Nat Methods*, 2008. **5**(6): p. 535-8.
73. Ramachandran, N., et al., *Self-assembling protein microarrays*. *Science*, 2004. **305**(5680): p. 86-90.
74. Rolfs, A., et al., *Production and sequence validation of a complete full length ORF collection for the pathogenic bacterium *Vibrio cholerae**. *Proc Natl Acad Sci U S A*, 2008. **105**(11): p. 4364-9.
75. Chatterjee, D.K., et al., *Protein microarray on-demand: a novel protein microarray system*. *PLoS ONE*, 2008. **3**(9): p. e3265.
76. Tao, S.C. and H. Zhu, *Protein chip fabrication by capture of nascent polypeptides*. *Nat Biotechnol*, 2006. **24**(10): p. 1253-4.
77. Holliger, P. and P.J. Hudson, *Engineered antibody fragments and the rise of single domains*. *Nat Biotechnol*, 2005. **23**(9): p. 1126-36.
78. Schlehuber, S. and A. Skerra, *Anticalins as an alternative to antibody technology*. *Expert Opin Biol Ther*, 2005. **5**(11): p. 1453-62.
79. Hosse, R.J., et al., *A new generation of protein display scaffolds for molecular recognition*. *Protein Sci*, 2006. **15**(1): p. 14-27.
80. Binz, H.K., et al., *Engineering novel binding proteins from nonimmunoglobulin domains*. *Nat Biotechnol*, 2005. **23**(10): p. 1257-68.
81. Vita, C., *Engineering novel proteins by transfer of active sites to natural scaffolds*. *Curr Opin Biotechnol*, 1997. **8**(4): p. 429-34.
82. Causton, A.S. and J.C. Sherman, *Design of proteins using rigid organic macrocycles as scaffolds*. *Bioorg Med Chem*, 1999. **7**(1): p. 23-7.
83. O'Shea, E.K., et al., *Peptide 'Velcro': design of a heterodimeric coiled coil*. *Curr Biol*, 1993. **3**(10): p. 658-67.
84. Binz, H.K. and A. Pluckthun, *Engineered proteins as specific binding reagents*. *Curr Opin Biotechnol*, 2005. **16**(4): p. 459-69.
85. Nygren, P.A. and A. Skerra, *Binding proteins from alternative scaffolds*. *J Immunol Methods*, 2004. **290**(1-2): p. 3-28.
86. Lamla, T. and V.A. Erdmann, *The Nano-tag, a streptavidin-binding peptide for the purification and detection of recombinant proteins*. *Protein Expr Purif*, 2004. **33**(1): p. 39-47.

87. Renberg, B., et al., *Affibody protein capture microarrays: synthesis and evaluation of random and directed immobilization of affibody molecules*. Anal Biochem, 2005. **341**(2): p. 334-43.
88. Wikman, M., et al., *Selection and characterization of HER2/neu-binding affibody ligands*. Protein Eng Des Sel, 2004. **17**(5): p. 455-62.
89. Amstutz, P., et al., *Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins*. J Biol Chem, 2005. **280**(26): p. 24715-22.
90. Nicaise, M., et al., *Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold*. Protein Sci, 2004. **13**(7): p. 1882-91.
91. Enander, K., et al., *Designed, functionalized helix-loop-helix motifs that bind human carbonic anhydrase II: a new class of synthetic receptor molecules*. J Am Chem Soc, 2004. **126**(14): p. 4464-5.
92. Lee, M., et al., *Shapes of antibody binding sites: qualitative and quantitative analyses based on a geomorphic classification scheme*. J Org Chem, 2006. **71**(14): p. 5082-92.
93. Nord, K., et al., *A combinatorial library of an alpha-helical bacterial receptor domain*. Protein Eng, 1995. **8**(6): p. 601-8.
94. Nilsson, B., et al., *A synthetic IgG-binding domain based on staphylococcal protein A*. Protein Eng, 1987. **1**(2): p. 107-13.
95. Binz, H.K., et al., *Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins*. J Mol Biol, 2003. **332**(2): p. 489-503.
96. Kohl, A., et al., *Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein*. Structure, 2005. **13**(8): p. 1131-41.
97. Steinmetz, M.O., et al., *Molecular basis of coiled-coil formation*. Proc Natl Acad Sci U S A, 2007. **104**(17): p. 7062-7.
98. Baggio, R., et al., *Identification of epitope-like consensus motifs using mRNA display*. J Mol Recognit, 2002. **15**(3): p. 126-34.
99. Souriau, C., et al., *New binding specificities derived from Min-23, a small cysteine-stabilized peptidic scaffold*. Biochemistry, 2005. **44**(19): p. 7143-55.
100. Zhao, A., et al., *A conformation-constrained peptide library based on insect defensin A*. Peptides, 2004. **25**(4): p. 629-35.
101. Williams, A. and L.G. Baird, *DX-88 and HAE: a developmental perspective*. Transfus Apher Sci, 2003. **29**(3): p. 255-8.
102. Lupas, A.N. and M. Gruber, *The structure of alpha-helical coiled coils*. Adv Protein Chem, 2005. **70**: p. 37-78.
103. Liu, J. and B. Rost, *Comparing function and structure between entire proteomes*. Protein Sci, 2001. **10**(10): p. 1970-9.
104. Crick, F.H., *Is alpha-keratin a coiled coil?* Nature, 1952. **170**(4334): p. 882-3.

105. Pauling, L. and R.B. Corey, *The structure of feather rachis keratin*. Proc Natl Acad Sci U S A, 1951. **37**(5): p. 256-61.
106. Arndt, K.M., et al., *Comparison of in vivo selection and rational design of heterodimeric coiled coils*. Structure, 2002. **10**(9): p. 1235-48.
107. Litowski, J.R. and R.S. Hodges, *Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity*. J Pept Res, 2001. **58**(6): p. 477-92.
108. Litowski, J.R. and R.S. Hodges, *Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity*. J Biol Chem, 2002. **277**(40): p. 37272-9.
109. Acharya, A., et al., *Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R)*. Biochemistry, 2006. **45**(38): p. 11324-32.
110. Moitra, J., et al., *Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil*. Biochemistry, 1997. **36**(41): p. 12567-73.
111. Malashkevich, V.N., et al., *The crystal structure of a five-stranded coiled coil in COMP: a prototypic ion channel? Science*, 1996. **274**(5288): p. 761-5.
112. Yadav, M.K., et al., *Structure-based engineering of internal cavities in coiled-coil peptides*. Biochemistry, 2005. **44**(28): p. 9723-32.
113. Zhu, B.Y., et al., *Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded alpha-helical coiled coils/leucine zippers*. Protein Sci, 1993. **2**(3): p. 383-94.
114. Chao, H., et al., *Use of a heterodimeric coiled-coil system for biosensor application and affinity purification*. J Chromatogr B Biomed Sci Appl, 1998. **715**(1): p. 307-29.
115. Ryan, S.J. and A.J. Kennan, *Variable stability heterodimeric coiled-coils from manipulation of electrostatic interface residue chain length*. J Am Chem Soc, 2007. **129**(33): p. 10255-60.
116. Pace, C.N., *Contribution of the hydrophobic effect to globular protein stability*. J Mol Biol, 1992. **226**(1): p. 29-35.
117. Pace, C.N., et al., *Forces contributing to the conformational stability of proteins*. Faseb J, 1996. **10**(1): p. 75-83.
118. Kohn, W.D., et al., *Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils*. J Mol Biol, 1998. **283**(5): p. 993-1012.
119. Hodges, R.S., et al., *Reversed-phase liquid chromatography as a useful probe of hydrophobic interactions involved in protein folding and protein stability*. J Chromatogr A, 1994. **676**(1): p. 3-15.
120. Peters, J., et al., *Tetrabrachion: a filamentous archaeobacterial surface protein assembly of unusual structure and extreme stability*. J Mol Biol, 1995. **245**(4): p. 385-401.
121. Roncarati, R., et al., *Assembly and trafficking of human small conductance Ca²⁺-activated K⁺ channel SK3 are governed by different molecular domains*. Mol Cell Neurosci, 2005. **28**(2): p. 314-25.

122. Yonekura, K., et al., *Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy*. *Nature*, 2003. **424**(6949): p. 643-50.
123. Serna-Rico, A., et al., *The putative coiled coil domain of the phi 29 terminal protein is a major determinant involved in recognition of the origin of replication*. *J Biol Chem*, 2000. **275**(51): p. 40529-38.
124. Mangione, P., et al., *Amyloid fibrils derived from the apolipoprotein A1 Leu174Ser variant contain elements of ordered helical structure*. *Protein Sci*, 2001. **10**(1): p. 187-99.
125. Hadley, E.B., et al., *Preferred side-chain constellations at antiparallel coiled-coil interfaces*. *Proc Natl Acad Sci U S A*, 2008. **105**(2): p. 530-5.
126. Harbury, P.B., et al., *A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants*. *Science*, 1993. **262**(5138): p. 1401-7.
127. De Crescenzo, G., et al., *Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: effect of chain length on the kinetic and thermodynamic constants of binding*. *Biochemistry*, 2003. **42**(6): p. 1754-63.
128. Ambroggio, X.I. and B. Kuhlman, *Computational design of a single amino acid sequence that can switch between two distinct protein folds*. *J Am Chem Soc*, 2006. **128**(4): p. 1154-61.
129. Papapostolou, D., et al., *Engineering nanoscale order into a designed protein fiber*. *Proc Natl Acad Sci U S A*, 2007. **104**(26): p. 10853-8.
130. Berger, B., et al., *Predicting coiled coils by use of pairwise residue correlations*. *Proc Natl Acad Sci U S A*, 1995. **92**(18): p. 8259-63.
131. Wolf, E., et al., *MultiCoil: a program for predicting two- and three-stranded coiled coils*. *Protein Sci*, 1997. **6**(6): p. 1179-89.
132. McDonnell, A.V., et al., *Paircoil2: improved prediction of coiled coils from sequence*. *Bioinformatics*, 2006. **22**(3): p. 356-8.
133. Gruber, M., et al., *Comparative analysis of coiled-coil prediction methods*. *J Struct Biol*, 2006. **155**(2): p. 140-5.
134. Lupas, A., *Coiled coils: new structures and new functions*. *Trends Biochem Sci*, 1996. **21**(10): p. 375-82.
135. Chao, H., et al., *Kinetic study on the formation of a de novo designed heterodimeric coiled-coil: use of surface plasmon resonance to monitor the association and dissociation of polypeptide chains*. *Biochemistry*, 1996. **35**(37): p. 12175-85.
136. Tripet, B., et al., *Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins*. *Protein Eng*, 1996. **9**(11): p. 1029-42.
137. Woolfson, D.N., *The design of coiled-coil structures and assemblies*. *Adv Protein Chem*, 2005. **70**: p. 79-112.
138. Arndt, K.M., et al., *A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble*. *J Mol Biol*, 2000. **295**(3): p. 627-39.

139. Kwok, S.C. and R.S. Hodges, *Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length*. Biopolymers, 2004. **76**(5): p. 378-90.
140. Kohn, W.D., et al., *Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper*. Protein Sci, 1995. **4**(2): p. 237-50.
141. Zhou, N.E., et al., *The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability*. Protein Eng, 1994. **7**(11): p. 1365-72.
142. Kohn, W.D., et al., *The effects of interhelical electrostatic repulsions between glutamic acid residues in controlling the dimerization and stability of two-stranded alpha-helical coiled-coils*. J Biol Chem, 1995. **270**(43): p. 25495-506.
143. Yu, Y., et al., *Ion pairs significantly stabilize coiled-coils in the absence of electrolyte*. J Mol Biol, 1996. **255**(3): p. 367-72.
144. Krylov, D., et al., *Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids*. J Mol Biol, 1998. **279**(4): p. 959-72.
145. Ayriss, J., et al., *High-throughput screening of single-chain antibodies using multiplexed flow cytometry*. J Proteome Res, 2007. **6**(3): p. 1072-82.
146. Muller, K.M., et al., *Protein fusions to coiled-coil domains*. Methods Enzymol, 2000. **328**: p. 261-82.
147. Muller, K.M., et al., *A dimeric bispecific miniantibody combines two specificities with avidity*. FEBS Lett, 1998. **432**(1-2): p. 45-9.
148. Arndt, K.M., et al., *Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain*. J Mol Biol, 2001. **312**(1): p. 221-8.
149. Mason, J.M., et al., *Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design*. Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8989-94.
150. Mason, J.M., et al., *Positive aspects of negative design: simultaneous selection of specificity and interaction stability*. Biochemistry, 2007. **46**(16): p. 4804-14.
151. Jouaux, E.M., et al., *Targeting the c-Myc coiled coil with interfering peptides*. J Pept Sci, 2008. **14**(9): p. 1022-31.
152. Blackstock, W.P. and M.P. Weir, *Proteomics: quantitative and physical mapping of cellular proteins*. Trends Biotechnol, 1999. **17**(3): p. 121-7.
153. Siegel, R.M., et al., *Measurement of molecular interactions in living cells by fluorescence resonance energy transfer between variants of the green fluorescent protein*. Sci STKE, 2000. **2000**(38): p. PL1.
154. Xu, Y., et al., *A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins*. Proc Natl Acad Sci U S A, 1999. **96**(1): p. 151-6.
155. Medina, M.A. and P. Schuille, *Fluorescence correlation spectroscopy for the detection and study of single molecules in biology*. Bioessays, 2002. **24**(8): p. 758-64.

156. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
157. Di Tullio, A., et al., *Molecular recognition by mass spectrometry*. J Mass Spectrom, 2005. **40**(7): p. 845-65.
158. McDonnell, J.M., *Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition*. Curr Opin Chem Biol, 2001. **5**(5): p. 572-7.
159. Michnick, S.W., *Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments*. Curr Opin Struct Biol, 2001. **11**(4): p. 472-7.
160. Causier, B. and B. Davies, *Analysing protein-protein interactions with the yeast two-hybrid system*. Plant Mol Biol, 2002. **50**(6): p. 855-70.
161. Ratushny, V. and E. Golemis, *Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system*. Biotechniques, 2008. **44**(5): p. 655-62.
162. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions*. Nature, 1989. **340**(6230): p. 245-6.
163. Auerbach, D., et al., *The post-genomic era of interactive proteomics: facts and perspectives*. Proteomics, 2002. **2**(6): p. 611-23.
164. Hu, J.C., *Model systems: Studying molecular recognition using bacterial n-hybrid systems*. Trends Microbiol, 2001. **9**(5): p. 219-22.
165. Ladant, D. and G. Karimova, *Genetic systems for analyzing protein-protein interactions in bacteria*. Res Microbiol, 2000. **151**(9): p. 711-20.
166. Dove, S.L. and A. Hochschild, *Use of artificial activators to define a role for protein-protein and protein-DNA contacts in transcriptional activation*. Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 173-80.
167. Hu, J.C., et al., *Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions*. Science, 1990. **250**(4986): p. 1400-3.
168. Ozawa, T. and Y. Umezawa, *Detection of protein-protein interactions in vivo based on protein splicing*. Curr Opin Chem Biol, 2001. **5**(5): p. 578-83.
169. Ozawa, T., et al., *A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing*. Anal Chem, 2000. **72**(21): p. 5151-7.
170. Michnick, S.W., *Protein fragment complementation strategies for biochemical network mapping*. Curr Opin Biotechnol, 2003. **14**(6): p. 610-7.
171. Voigt, C.A., et al., *Protein building blocks preserved by recombination*. Nat Struct Biol, 2002. **9**(7): p. 553-8.
172. Galzitskaya, O.V. and B.S. Melnik, *Prediction of protein domain boundaries from sequence alone*. Protein Sci, 2003. **12**(4): p. 696-701.
173. Remy, I. and S.W. Michnick, *Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays*. Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5394-9.

174. Wehrman, T., et al., *Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments*. Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3469-74.
175. Shibasaki, S., et al., *Development of a yeast protein fragment complementation assay (PCA) system using dihydrofolate reductase (DHFR) with specific additives*. Appl Microbiol Biotechnol, 2008. **80**(4): p. 735-43.
176. Karimova, G., et al., *A bacterial two-hybrid system based on a reconstituted signal transduction pathway*. Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5752-6.
177. Mossner, E., et al., *Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs*. J Mol Biol, 2001. **308**(2): p. 115-22.
178. Koch, H., et al., *Direct selection of antibodies from complex libraries with the protein fragment complementation assay*. J Mol Biol, 2006. **357**(2): p. 427-41.
179. Hagemann, U.B., et al., *Selectional and mutational scope of peptides sequestering the Jun-Fos coiled-coil domain*. J Mol Biol, 2008. **381**(1): p. 73-88.
180. Pelletier, J.N., et al., *Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12141-6.
181. Secco, P., et al., *Antibody library selection by the {beta}-lactamase protein fragment complementation assay*. Protein Eng Des Sel, 2009. **22**(3): p. 149-58.
182. Glaser, P., et al., *The calmodulin-sensitive adenylate cyclase of Bordetella pertussis: cloning and expression in Escherichia coli*. Mol Microbiol, 1988. **2**(1): p. 19-30.
183. Magliery, T.J. and L. Regan, *Reassembled GFP: detecting protein-protein interactions and protein expression patterns*. Methods Biochem Anal, 2006. **47**: p. 391-405.
184. Remy, I. and S.W. Michnick, *A cDNA library functional screening strategy based on fluorescent protein complementation assays to identify novel components of signaling pathways*. Methods, 2004. **32**(4): p. 381-8.
185. Matagne, A., et al., *Catalytic properties of class A beta-lactamases: efficiency and diversity*. Biochem J, 1998. **330** (Pt 2): p. 581-98.
186. Moore, J.T., et al., *The development of beta-lactamase as a highly versatile genetic reporter for eukaryotic cells*. Anal Biochem, 1997. **247**(2): p. 203-9.
187. Galarneau, A., et al., *Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions*. Nat Biotechnol, 2002. **20**(6): p. 619-22.
188. Nord, O., et al., *Fluorescent detection of beta-lactamase activity in living Escherichia coli cells via esterase supplementation*. FEMS Microbiol Lett, 2005. **242**(1): p. 73-9.
189. Huang, W. and T. Palzkill, *A natural polymorphism in beta-lactamase is a global suppressor*. Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8801-6.
190. Remy, I. and S.W. Michnick, *Dynamic visualization of expressed gene networks*. J Cell Physiol, 2003. **196**(3): p. 419-29.

191. Amstutz, P., et al., *Rapid selection of specific MAP kinase-binders from designed ankyrin repeat protein libraries*. Protein Eng Des Sel, 2006. **19**(5): p. 219-29.
192. Mason, J.M., et al., *iPEP: peptides designed and selected for interfering with protein interaction and function*. Biochem Soc Trans, 2008. **36**(Pt 6): p. 1442-7.
193. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface*. Science, 1985. **228**(4705): p. 1315-7.
194. Sergeeva, A., et al., *Display technologies: application for the discovery of drug and gene delivery agents*. Adv Drug Deliv Rev, 2006. **58**(15): p. 1622-54.
195. Mattheakis, L.C., et al., *An in vitro polysome display system for identifying ligands from very large peptide libraries*. Proc Natl Acad Sci U S A, 1994. **91**(19): p. 9022-6.
196. FitzGerald, K., *In vitro display technologies - new tools for drug discovery*. Drug Discov Today, 2000. **5**(6): p. 253-258.
197. Liu, R., et al., *Optimized synthesis of RNA-protein fusions for in vitro protein selection*. Methods Enzymol, 2000. **318**: p. 268-93.
198. Josephson, K., et al., *Ribosomal synthesis of unnatural peptides*. J Am Chem Soc, 2005. **127**(33): p. 11727-35.
199. Samuelson, P., et al., *Display of proteins on bacteria*. J Biotechnol, 2002. **96**(2): p. 129-54.
200. Ho, M., et al., *Isolation of anti-CD22 Fv with high affinity by Fv display on human cells*. Proc Natl Acad Sci U S A, 2006. **103**(25): p. 9637-42.
201. Graus, Y.F., et al., *Selection of recombinant anti-HuD Fab fragments from a phage display antibody library of a lung cancer patient with paraneoplastic encephalomyelitis*. J Neuroimmunol, 1998. **82**(2): p. 200-9.
202. Burritt, J.B., et al., *Filamentous phage display of oligopeptide libraries*. Anal Biochem, 1996. **238**(1): p. 1-13.
203. Azzazy, H.M. and W.E. Highsmith, Jr., *Phage display technology: clinical applications and recent innovations*. Clin Biochem, 2002. **35**(6): p. 425-45.
204. Bai, Y. and H. Feng, *Selection of stably folded proteins by phage-display with proteolysis*. Eur J Biochem, 2004. **271**(9): p. 1609-14.
205. Srinivasan, N. and J.D. Kilburn, *Combinatorial approaches to synthetic receptors*. Curr Opin Chem Biol, 2004. **8**(3): p. 305-10.
206. Kramer, K., *Evolutionary affinity and selectivity optimization of a pesticide-selective antibody utilizing a hapten-selective immunoglobulin repertoire*. Environ Sci Technol, 2002. **36**(22): p. 4892-8.
207. Uchiyama, F., et al., *Designing scaffolds of peptides for phage display libraries*. J Biosci Bioeng, 2005. **99**(5): p. 448-56.
208. Mintz, P.J., et al., *Fingerprinting the circulating repertoire of antibodies from cancer patients*. Nat Biotechnol, 2003. **21**(1): p. 57-63.

-
209. McCafferty, J., et al., *Phage antibodies: filamentous phage displaying antibody variable domains*. Nature, 1990. **348**(6301): p. 552-4.
 210. Clackson, T., et al., *Making antibody fragments using phage display libraries*. Nature, 1991. **352**(6336): p. 624-8.
 211. Kim, S.J., et al., *Antibody engineering for the development of therapeutic antibodies*. Mol Cells, 2005. **20**(1): p. 17-29.
 212. Laffly, E. and R. Sodoyer, *Monoclonal and recombinant antibodies, 30 years after*. Hum Antibodies, 2005. **14**(1-2): p. 33-55.
 213. Santi, E., et al., *Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics*. J Mol Biol, 2000. **296**(2): p. 497-508.
 214. Mayer, M.P., *A new set of useful cloning and expression vectors derived from pBlueScript*. Gene, 1995. **163**(1): p. 41-6.
 215. Mason, J.M., et al., *Improved stability of the Jun-Fos Activator Protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis*. J Biol Chem, 2007. **282**(32): p. 23015-24.
 216. Nawrot, P., et al., *Effects of caffeine on human health*. Food Addit Contam, 2003. **20**(1): p. 1-30.
 217. Christian, M.S. and R.L. Brent, *Teratogen update: evaluation of the reproductive and developmental risks of caffeine*. Teratology, 2001. **64**(1): p. 51-78.
 218. Soper, S.A., et al., *Point-of-care biosensor systems for cancer diagnostics/prognostics*. Biosens Bioelectron, 2006. **21**(10): p. 1932-42.