ROMA TRE
UNIVERSITÀ DEGLI STUDI

Scuola Dottorale in Biologia
Sezione "Biologia Cellulare e Molecolare"
Ciclo di Dottorato XXIII

"Systems optimization for the selection of phage display random peptide libraries"

"Ottimizzazione di sistemi per la selezione di librerie peptidiche a sequenza casuale espresse tramite Phage Display"

A.A. 2010/2011

Candidato:
Dott.ssa Anna Quintarelli

Docente Guida:
Prof. Pier Luigi Luisi

Coordinatore:
Prof. Paolo Mariottini

ABSTRACT

This doctoral work is part of the Never Born Protein project. It is based on idea that the proteins existing in our Earth are only an infinitesimal fraction of the all possible sequences. To use an analogy to explain the relationship between existing and possible proteins, we can say that the ratio between them corresponds roughly to the ratio between the Sahara desert and a single grain of sand.

Given the vast number of possible sequences, Nature could not have explored all possible amino acid combinations. Therefore, we are faced with the problem of how the "few" extant proteins were produced and/or selected: do extant proteins possess any special chemico-physical properties (such as solubility, fold, functionality, thermodynamic stability) that made their selection inevitable? Or rather are they the result of contingency, a frozen accident? If they are the result of contingency, there exists a universe of proteins which properties have never been sampled by nature: the Never Born Proteins – NBP.

To investigate the NBP world, random peptide 50 amino acid long libraries have been designed and selected for several characteristics by phage display technique.

To display the protein libraries we created 3 different phagemid systems, improving vectors used in previous experiments. The main improvements have been in the increase of efficiency in the library cloning and in the phage production steps. In addition, in the new vectors a His-tag replaced the old tag (a c-myc tag recognized by MAb 9E10 antibody). This involved a change of the purification system, in fact His-tagged proteins can be identified by Immobilized metal affinity chromatography (IMAC), whose efficiency is higher than antibody purification.

The complexity degree of phagemid libraries was estimated at 10^7 different sequences per ml. This complexity degree is a good compromise between a wide space of sequences and a good number of copies of each sequence.

According to the production of phage libraries, by a 3+3 monovalent phage system we obtained phage particles with the phagemid incapsidated and a single NBP fused to the minor capsid protein (pIII protein).

The first selection aim has been to look for three-dimensional structure using the resistance to thrombin digestion as folding criterion.

All random sequences have in the middle a PRG (proline-arginine-glycine) site, corresponding to a thrombin cleavage site. Folded sequences preserve PRG during thrombin digestion, while un-folded sequences are cut since

PRG is exposed to enzyme cleavage. To recover His-tagged sequences un-digested by thrombin we chose the Ni-affinity purification, using ELISA Ni-coated plates. Results obtained after the second biopanning cycle showed a high proportion of resistant sequences (around 10^3 cfu/ml, the starting phage titre of 10^8 cfu/ml), indicating a good chance to select folded NBPs. Performed controls confirmed the results excluding un-specific bonds.

In parallel, a probable catalytic activity was tested, screening phage libraries against a TSA target, which mimics the transition state of an amide bond hydrolysis.

The aim of this selection is that a molecule which binds the transition state during a reaction can catalyze this reaction, lowering the activation energy. Therefore, selected NBPs could catalyze amide bond hydrolysis reaction. The procedure implied the selection of the phage library on a solid surface coated with the TSA. TSA was immobilized on the plate by a covalent bond between the acid group of TSA and secondary amine on the plate. In addition, we tested the influence of a metallic cofactor (Zinc) on affinity between NBPs and the target.

A good amount of phage library was recovered from wells coated with TSA (around10^5 cfu/ml, the starting phage titre of 10^9 cfu/ml). The metal cofactor did not seem to affect the affinity for the target, in fact the phage quantity recovered from the plate was similar both with and without Zn.

Performed controls excluded un-specific bonds between capsid proteins and TSA and between NBPs and reagents used for TSA coupling. Un-specific bonds with plate were excluded too.

The last selection was to test possible interactions between NBPs and other proteins and investigate their possible use like inhibitors/activators of the target proteins.

The basic idea of this selection is that a ligand, to act as an inhibitor/ activator of a protein, first of all, has to bind such protein.

Chosen targets were 4 proteases: papain, pepsin, trypsin and α-chymotrypsin. They were immobilized on a polystyrene plate by adsorption. To avoid possible bounds between phage and plate surface, wells were incubated also with BSA.

Already after the second cycle, the distribution of selected phage is different for the 4 proteases: the number of phage from papain- and pepsin- wells is higher than those for the other proteases, in particular trypsin.

Performed controls excluded un-specific bonds between capsid proteins and each protease. Un-specific bond with BSA were excluded too.

The possibility of finding folded random proteins might suggest that folding is not a special property to be selected by Nature. This point could be taken as favouring the contingency hypothesis. Also the probability to find new sequences with a potential activity supports the idea that there may be an entire universe of possible proteins, with unknown properties.

RIASSUNTO

Il mio lavoro di dottorato fa parte del progetto *Never Born Proteins*. Questo progetto si basa sul concetto che le proteine esistenti sulla Terra sono solo una frazione infinitesimale di tutte le sequenze possibili. Per spiegare il rapporto tra le proteine esistenti e possibili, possiamo dire che il rapporto tra di esse corrisponde al rapporto tra il deserto del Sahara e un singolo granello di sabbia.
Dato il gran numero di possibili sequenze, la Natura non avrebbe potuto esplorare tutte le possibili combinazioni di amminoacidi. Pertanto, siamo di fronte al problema di come le "poche" proteine esistenti sono state prodotte e/o selezionate: le proteine esistenti hanno particolari proprietà chimico-fisiche (come solubilità, struttura terziaria, funzionalità, stabilità termodinamica) che hanno reso inevitabile la loro selezione? O piuttosto sono il frutto della contingenza? Se fossero il risultato della contingenza, esisterebbe allora un universo di proteine con proprietà mai testate dalla natura: le *Never Born Proteins - NBP*.
Per esplorare il mondo delle NBPs, librerie codificanti peptidi di 50 aminoacidi a sequenza casuale sono state progettate e poi selezionate per diverse caratteristiche con la tecnica del *phage display*.
Per esprimere le librerie abbiamo creato 3 diversi sistemi fagemidici, migliorando alcuni vettori utilizzati in precedenti esperimenti. I principali miglioramenti hanno riguardato un aumento dell'efficienza nel passaggio di clonazione della libreria e nella fase di produzione dei fagi. Inoltre, i nuovi vettori possiedono un His-tag al posto del vecchio *tag* (un c-myc tag riconosciuto dall'anticorpo MAb9E10). Ciò ha comportato un cambiamento del sistema di purificazione, infatti le proteine legate ad un His-tag possono essere identificate anche tramite cromatografia di affinità (*Immobilized metal affinity chromatography - IMAC*) la cui efficienza è superiore alla purificazione con anticorpi.
Il grado di complessità delle librerie fagemidiche è stato stimato 10^7 sequenze diverse per ml. Questo grado di complessità è un buon compromesso tra un ampio spazio delle sequenze e un buon numero di copie di ciascuna sequenza.
Per quanto riguarda la produzione delle librerie fagiche, tramite un sistema fagico monovalente 3+3 abbiamo ottenuto particelle fagiche con il fagemide incapsidato e una singola NBP fusa alla proteina minore del capside (la proteina pIII).

L'obiettivo della prima selezione è stato quello di cercare strutture tridimensionali utilizzando come criterio di *folding* la resistenza alla digestione da parte della trombina.
Tutte le sequenze casuali possiedono nella regione centrale un sito PRG (prolina-arginina-glicina), corrispondente ad un sito di taglio per trombina. Le sequenze strutturate proteggono il PRG dal taglio della trombina, mentre in quelle non strutturate il PRG è esposto alla digestione dell'enzima.
Per recuperare le sequenze legate all'His-tag e non tagliate dalla trombina abbiamo scelto la purificazione mediante affinità al Ni, utilizzando piastre ELISA ricoperte di Ni. I risultati ottenuti dopo il secondo ciclo di biopanning hanno mostrato un'elevata quantità di sequenze resistenti (circa 10^3 cfu/ml con un titolo di incubazione iniziale pari a 10^8 cfu/ml), ciò indica una buona possibilità di selezionare NBPs foldate. I controlli fatti hanno confermato i risultati escludendo legami aspecifici.
In parallelo le NBPs sono state testate anche per una possibile attività catalitica, selezionando le librerie fagiche contro un TSA, che mima lo stato di transizione della reazione di idrolisi del legame ammidico.
L'idea di base di questa selezione è che una molecola che lega lo stato di transizione durante una reazione può catalizzare questa reazione, abbassando l'energia di attivazione. Pertanto, le NBPs selezionate potrebbero catalizzare l'idrolisi del legame ammidico. La procedura implica la selezione della libreria dei fagi su una superficie solida rivestita con il TSA. Il TSA è immobilizzato su una piastra mediante un legame covalente tra il suo gruppo acido e l'ammina secondaria della piastra. Abbiamo anche verificato l'influenza di un cofattore metallico (zinco) sull'affinità di legame tra le NBPs e il TSA. Una quantità importante della libreria fagica è stata recuperata dai pozzi rivestiti con il TSA (circa 10^5 cfu/ml con un titolo di incubazione iniziale pari a 10^9 cfu/ml). Il cofattore metallico non sembra influenzare l'affinità per il target, infatti la quantità dei fagi recuperati dalla piastra è simile sia con che senza Zn. I controlli fatti hanno escluso legami aspecifici tra le proteine del capside e il TSA e tra le NBPs e i reagenti utilizzati per il *coupling* del TSA. Sono stati esclusi anche legami aspecifici con la piastra.
L'ultima selezione è stata effettuata per verificare le possibili interazioni tra le NBPs e altre proteine ed indagare il loro possibile uso come inibitori/attivatori delle proteine bersaglio.
L'idea di base di questa selezione è che un ligando per agire come inibitore/attivatore di una proteina, prima di tutto, deve legare tale proteina.
I *targets* scelti sono stati 4 proteasi: papaina, pepsina, tripsina e α-chimotripsina. Esse sono stati immobilizzati su una piastra di polistirene

tramite adsorbimento. Per evitare possibili legami tra fagi e la superficie della piastra, i pozzetti sono stati incubati anche con BSA.
Già dopo il secondo ciclo, la distribuzione dei fagi selezionati è diversa per le 4 proteasi: il numero di fagi recuperati dai pozzetti con papaina e pepsina è superiore a quello per le altre proteasi, in particolare la tripsina. I controlli fatti hanno escluso legami aspecifici tra le proteine del capside e ogni proteasi. È stato escluso anche il legame aspecifico con la BSA.
La probabilità di trovare proteine a sequenza casuale con una struttura terziaria potrebbe suggerire che il *folding* non è una proprietà necessaria per essere selezionato dalla Natura. Questo punto potrebbe andare a favore della teoria della contingenza. Anche la probabilità di trovare nuove sequenze con una potenziale attività sostiene l'idea che ci possa essere un intero universo di possibili proteine, con proprietà sconosciute.

INDEX

1. INTRODUCTION

Science assumes that life on Earth originated from inanimate matter by a gradual and spontaneous increase of molecular complexity. Several different theoretical frameworks have been proposed to account for the spontaneous emergence of life. Wächtershäuser [Wächtershäuser, 1988] identifies enzyme-free metabolic cycles as the pivotal system underpinning life's origin; Kauffmann proposes auto-catalytic peptide cycles as the primary motor [Kauffmann, 1996], whereas Cech fostered the idea that RNA was the scaffold of the first living system [Cech, 1993]. Conversely, Luisi emphasises the autopoietic nature of life [Luisi, 2003]; whereas Lancet proposes composition inheritance as the foundation of life [Segre, 2000]. Despite the differences, all theories must confront the same fundamental question: is life an obligatory pathway given certain initial conditions (i.e. determinism)? Or it is rather the result of the simultaneous interplay of different contingent factors (i.e. contingency)?
Within this debate, one of the most interesting questions in modern life science is how prebiotic evolution of the first biopolymers occurred and so how functional known biopolymers were selected.

It is common knowledge that for specific functionality, such as binding or catalysis, specific sequences which fold specifically are necessary. In fact, protein activity is the result of a specific three-dimensional structure, which, in turn, is determined by the amino acid primary sequence. Therefore, based on this premise, the idea of the present work is to investigate the folding frequency of random polypeptide sequences, their potential catalytic activity and finally their possible use like inhibitors/activators for several proteases.
The random polypeptide library has been designed with no sequence or structural constrains so it can be reasonably considered as a mirror image of a peptide population produced under plausible prebiotic conditions. The selection for folding is based on the concept that folded polypeptides are more protected against digestion by a protease than unfolded ones, while the ability to bind a transition state analogue of the ester and amide bond hydrolysis is at the bottom of catalytic selection.

1.1 ABOUT ORIGIN OF LIFE

The Oparin-Haldane theory [Oparin, 1954 (I - II), 1924; Haldane, 1954, 1929] about the origin of life on Earth, stating that life is the result of a series of spontaneous events which produced the first self-reproducing protocells starting from inanimate matter, is one of the most recognized. By definition, this transition to life via prebiotic molecular evolution excludes panspermia, (the idea that life on earth comes from space), and divine intervention.
Life is based on the action of proteins and nucleic acids whose functions are the result of a specific sequence that produces their structural fold and so the activity. These amino acid long sequences are co-polymeric structures, formed by macromolecules having chemically different monomer units in the same chain.

The assumption that life derives from inanimate matter bears the simple consequence that it could be reproduced in laboratory, a basic concept that is the linchpin of origin of life experimental research. Even assuming that this is possible, a big question remains unanswered as clearly formulated by Eschenmoser and Kisakürek [Eschenmoser, 1996]: “Is the pathway that goes from inanimate to animate matter determined by the laws of physics and chemistry? Or is it due to a unique event due to the contingent parameters operating in a particular time/space situation – something that in the old nomenclature would be called chance?”.

From a deterministic point of view, life arose from inanimate matter through a series of causally linked events ruled by the laws of physics and chemistry. Conversely, contingency may be defined as the outcome of a particular set of concomitant effects that apply in a particular space-time situation and thus determine the outcome of a given event. In addition, in a strictly deterministic situation, the state of a system at any point in time determines the future behaviour of the system – with no random influences. By contrast, in a non-deterministic (i.e. stochastic system) it is not generally possible to predict the time evolution of the system.
The deterministic view gained a broad support by de Duve and Morowitz [de Duve, 2002; Morowitz, 1993]. De Duve [de Duve, 1995] wrote: “[…] given the suitable initial conditions, the emergence of life is highly probable and governed by the laws of chemistry and physics [...]”, suggesting the idea that life on Earth was inevitable.

To support the contingency theory one may cite Jacques Monod's *Chance and Necessity* [Monod, 1971], his colleague François Jacob [Jacob, 1982] and Stephen Jay Gould [Gould, 1989]. Contingency, in this particular context, can be defined as the simultaneous interplay of several concomitant effects to shape an event in a given space/time situation. In this scenario any change in the starting contingent conditions would dramatically affect the final result. The implications of contingency are dramatic, as Stephen Jay Gould illustrates when he says: "[...] run the tape again, and the first step from prokaryotic to eukaryotic cell may take twelve billions years instead of two" [Gould, 1989].

The first scientist who consciously did chemistry in pursuit of the origin of life was the graduate student Stanley Miller, fascinated by Oparin's idea [Oparin, 1961]. Miller obtained the formation of several α-amino acids and other relatively complex substances of biological importance by passing electrical discharges through a flask filled with hydrogen, ammonia, methane, and water vapour (assumed by Oparin to be the constituents of prebiotic reductive atmosphere) (Figure 1.1) [Miller, 1953]. Although the nature of the primitive atmosphere is strongly controversial, the key point of Miller's experiments is that relatively complex biochemicals can be formed from a mixture of very simple gaseous components in a plausible prebiotic way.

Following 1953 Miller's experiment, many chemists successfully attempted the synthesis of other compounds of biochemical relevance under prebiotic conditions. For instance Shen and Orò reported the prebiotic synthesis of histidyl-histidine [Shen, 1990 (I – II)] and Plankensteiner and co-workers investigated the possible catalytic activity of short oligopeptides synthesized under prebiotic evolution [Plankensteiner, 2002].

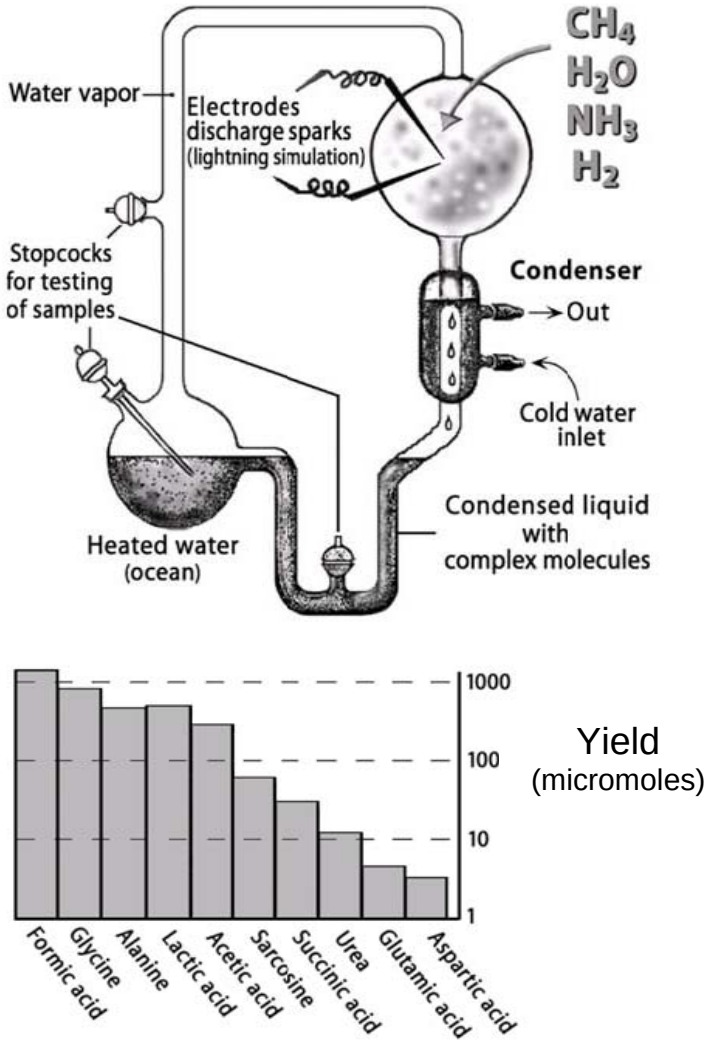


Figure 1.1 ***Stanley Miller's experiment.*** *The assumption was a strongly reduced atmosphere consisting of the four gaseous components assumed by Oparin to be the constituents of a prebiotic atmosphere (hydrogen, ammonia, methane and water vapour), and electric discharges as energy source. The graph represents the products and the relative yields obtained (Adapted from Luisi 2006).*

The sequences, or primary structures, of existing biopolymers are believed to be a product of evolution. From the molecular point of view, biological evolution can be viewed as a random walk and optimisation through the sequence space. This space is astronomically big because the number of all possible sequences is exponentially proportional to the length of the polymer. Indeed, it is commonly accepted that the proteins existing in our Earth are only an infinitesimal fraction of the possible sequences [de Duve, 2002; Levinthal, 1968].

To use an analogy to explain the relationship between existing and possible proteins, we can say that the ratio between them corresponds roughly to the ratio between the Sahara desert and a single grain of sand [Luisi, 2006].

Given the vast number of possible sequences, Nature could not have explored all possible amino acid combinations and, therefore, many proteins with interesting new properties may have never been sampled by Nature - this is true even for relatively short peptides. For example, to summarize some well known calculations, taking a polypeptide of 50 residues there are 20^{50} ($\sim 10^{65}$) possible 50mer products. If only one molecule of each of these peptides were to be synthesized, approximately $1.8 \cdot 10^{41}$ moles of material with a total mass of $\sim 10^{42}$ kg would be produced. This quantity corresponds to $\sim 10^{18}$ times the weight of the earth. Moreover, if this set of peptides could be synthesized at a rate of 10^6 molecules per second, it would take $\sim 3 \cdot 10^{51}$ years to complete their synthesis.

In light of these figures, we are faced with the problem of how the “few” extant proteins were produced and/or selected during prebiotic molecular evolution and subsequently, through a series of spontaneous steps of increasing molecular complexity, they could produce life in terms of self-reproducing protocells.

In other words, do extant proteins possess any special chemico-physical properties (such as solubility, fold, functionality, thermodynamic stability) that made their selection inevitable? Or rather are they the result of contingency, a frozen accident? If they are the result of contingency, there exists a universe of proteins which properties have never been sampled by nature: the Never Born Proteins - NBP [Luisi, 2006].

1.2 PEPTIDE FOLDING

It is common knowledge that protein activity is the result of a specific three-dimensional structure, specific to its function, such as binding or catalysis, and that this structure by the primary amino acid sequence.
Although some polypeptide chains fold spontaneously into the native state, others require the assistance of enzymes. For example, some polypeptide chains require enzymes to catalyse the formation and exchange of disulfide bonds and many others require the assistance of *chaperones*. These last are proteins that promote the folding of polypeptide chains and at the same time prevent the formation of illicit associations between such chains and other proteins.

A protein in its native state is not static. The secondary structural elements of the domains, as well as the entire domains, continually undergo small movements in space. The energy difference between the native state and the denatured one, in physiological conditions, is quite small, about 5-15 kcal/mol, not much more than the energy contribution of a single hydrogen bond, which is of the order of 2-5 kcal/mol [Branden, 1998]. From a biological point of view, it is important that this free energy difference is small because cells must be able to degrade proteins as well as synthesize them, and the functions of many proteins require structural flexibility.

When a fully extended unfolded polypeptide chain begins to fold, hydrophobic residues tend to be buried in the interior, greatly restricting the number of possible conformations the chain can assume, and therefore allowing proteins to fold in seconds rather then years.
The first observable event in the folding pathway of at least some proteins is a collapse of the flexible disordered unfolded polypeptide chain into a partly organized globular state, the *molten globule*. The *molten globule* has most of the secondary structure of the native state and in some cases even native-like positions of the α helices and β strands. However, it is less compact than the native structure and the proper packing interactions in the interior of the protein have not been formed. Instead, the interior side chains may be mobile, more closely resembling a liquid than the solid-like interior of the native state. In addition, loops and other elements of surface structure remain largely unfolded, with different conformations. For the single native form to be reached in the final stage of folding, the formation of native interactions throughout the protein, including hydrophobic packing in the interior as well as the fixation of surface loops must take place.

Since the fundamental prerequisite for a polypeptide to serve as a protein is a three-dimensional structure, the question whether a random sequence is likely to fold into a native state is basic to understand the possibility of the emergence of functional proteins in the "primordial soup".
To tackle the highly challenging problem of protein folding, a number of models have been developed; one of these is the lattice Monte Carlo simulation [Hilhorst, 1975; Verdier, 1973]. This simulation showed that the necessary condition for a polypeptide to have a kinetically reachable stable conformation is that it has a pronounced global energy minimum [Shakhnovich, 1993; Gutin, 1993; Goldstein, 1992]. It was estimated [Gutin, 1995] that only a small fraction of random sequences satisfies this requirement for chains of realistic lengths. The implication of this finding is that it is unlikely that folded protein will be found as a result of random exploration of the sequence space.
Although stable proteins might be only a tiny fraction of all possible sequences, stable polypeptides were not hydrolyzed and accumulated over time, whereas more frequently occurring unstable sequences underwent hydrolysis.
To experimentally address these hypothesises it is necessary to produce large random libraries of polypeptides. The construction and study of such libraries is discussed in the following chapters.

1.3 PEPTIDES SYNTHESIS UNDER PREBIOTIC CONDITIONS

Unfortunately, we do not know how the first proteins were formed in the prebiotic Earth. The characteristics of such polymers are so unique that it is only possible to speculate about their formation and, of course, try different solutions to reproduce them under prebiotic conditions; but, even knowing a useful method to produce such polymers, the problem would not be solved. In fact, as discussed previously, by synthesizing a random 50-mer chain using all 20 different amino acids it is theoretically possible to produce about 10^{65} different sequences and the probability to sample two identical chains is approximately equal to zero.

Furthermore, it is very difficult to predict the formation of a specific sequence due to the different elements playing a role in structure's formation. It is useful, to consider the dynamics behind the synthesis of a copolymer formed by two monomers, A and B (Luisi 2006) (Figure 1.2).

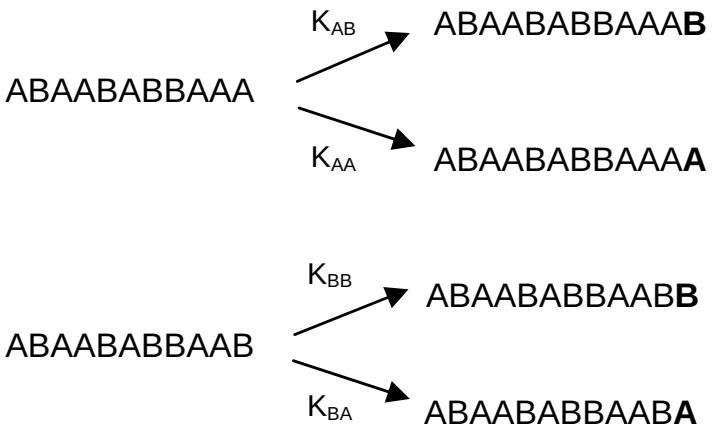


Figure 1.2 ***Growing of copolymers constituted by only two monomers A and B (Adapted from Luisi 2006)***

The growing chain can terminate with A or B and in the next step for each of the two sequences there is the same theoretical possibility to acquire either A or B (Figure 1.2). Four kinetic constants (k_{AA}, k_{BB}, k_{AB}, and k_{BA}) define the probability that a monomer is incorporated into the chain in a classic linear polymerization, while $r_A=k_{AA}/k_{AB}$ and $r_B=k_{BB}/k_{BA}$ the tendency of the chain to assume a certain composition and sequence. When r_A and r_B are both significantly larger than one, any polymer chain will tend to grow by incorporating a monomer equal to the last added, which will result in long stretches of homo-polymers. Conversely, when r_A and r_B are both significantly smaller than one the two monomers will be alternated in the sequence. In both cases the composition of the sequences could be roughly predicted knowing the monomers initial concentration. However, when r_A and r_B are both close to one the polymerization will proceed in a pure random fashion with an unpredictable final result. In the case of proteins or nucleic acids the situation is even more complicated because they are formed by several different monomers.

Furthermore, since the final result could be influenced by kinetic and thermodynamic parameters, the composition of prebiotic biopolymers could not be directly deduced even if we had known the amino acid composition of the "prebiotic soup". Indeed, neither the high or low abundance in the starting solution of a specific amino acid expresses its relative frequency in a poly-condensed product. Taking all these factors into account, it is reasonable to assume that the prebiotic synthesis of biopolymers occurred in an unpredictable random way.

The creation of polypeptides has also been explored under prebiotic conditions. For example, Fox and Dose [Fox, 1977] show that *proteinoids* (bodies containing polymerized amino acids) can be formed by heating mixtures of amino acids (containing a 10 fold excess of residues with reactive side chains, such as glutamic acid, aspartic acid or lysine) at 180° for a few hours. However, this procedure results in a high abundance of branched products and therefore is not considered a reliable method to produce biopolymers. Nevertheless, it was however reported that when using amides, the presence of clay increases the yields during repeated drying and heating, and Ito and co-workers [Yanagawa, 1990] reported a substantial arrays of polypeptides prepared in this way. But generally, one can accept the statement that no reliable method is known to produce high molecular weight of linear co-polypeptides under prebiotic conditions.
It is also interesting to mention the condensation of N-carboxyanhydrides (known as Leuch's anhydrides). The relevance of this reaction lies in the fact that NCA-amino acid derivatives are supposed to be prebiotic compounds [Taillades, 1999]. As noticed by Ferris [Ferris, 2002], this synthetic route has proved advantageous with regard to other synthetic paths, in fact, the synthesis can occur in water solution. Since the polymerization is faster than the hydrolysis rate, there is no racemization and therefore the synthesis is specific for α-amino acids. Oligomers up to 10-mers can be obtained in one single step, but this is the limit of the method, as it appears impossible to reach significantly higher polymerization degrees. However, decamers can be used to synthesise longer polypeptides by means of fragments condensation.
Finally, a new interesting development is offered by Orgel and Ghadiri's groups [Leman, 2004]; they have shown that carbonyl sulfide (COS), a simple volcanic gas, brings about the formation of peptides from amino acids under mild conditions in aqueous solution, reaching high yields (around 80%) at room temperature. Following this procedure, dipeptides and tripeptides have been successfully synthesised.

Alongside the chemical approach producing long peptides, there is the biological approach. In this case, it is necessary to put aside any focus on the mechanism and assume that the polymeric sequences were formed in a wide distribution of randomly produced polypeptide chains in which there is a specific percentage of folded chains which may have catalytic activity.
The aim of my work is to use the phage display method for obtaining this broad distribution of chains and investigate the existence of a stable fold and then their potential catalytic properties.

1.4 PHAGE DISPLAY TECHNOLOGY

Phage Display technology, introduced by G. Smith [Smith, 1985], is a good method for selecting specific molecules from large peptide or proteins libraries [Siegel, 2001; Ladner, 2000; Hoogenboom, 1998; Burton, 1995; Ladner, 1995; Neri, 1995 (II); Winter, 1994; Griffiths, 1993; Barbas, 1993]. Phage display has been successfully applied to a wide range of different purposes such studying:

- protein-protein interactions [Hertveldt, 2009; Sidhu, 2007; Cesareni, 1992];
- receptor and antibody-binding sites [Winter, 1994; Griffiths, 1993; Better, 1988; Skerra, 1988];
- protein stability [Kotz, 2004; Hoess, 2001; Forrer, 1999],
- new enzyme substrates and inhibitors [Hawinkels, 2007; Sedlacek, 2005; Deperthes, 2002; Kay, 2001; Hyde-DeRuyscher, 2000];
- the design of catalytic antibodies and enzymes with novel specificities [Fernandez-Gacio, 2003].

This technology utilizes the ability to express foreign proteins on the surface of phage particles which are fused to the coat proteins. The DNA encoding the displayed protein of interest is inserted into the single-stranded genome of filamentous phage, providing a physical link between genotype and phenotype. Instead of having to genetically engineer proteins or peptide variants one-by-one and subsequently express, purify, and analyze them, phage display enables the construction of large libraries of protein. Specific clones can then be selected and their sequence easily determined by sequencing the DNA contained in the phage particle. The selection format is usually based on biopanning of the library on solid phase surfaces carrying immobilized ligands. In addition, the high *in vitro* stability of the phage particle permits the use of a wide range of selection conditions, such as high temperature, denaturants, pH and ion concentration.

1.4.1 Biology of filamentous phage

Filamentous phage have a single-stranded DNA genome which is encased in a long cylinder approximately 6 nm wide by 900 to 2000 nm in length (Figure 1.3) and they are able to infect a variety of gram negative bacteria. Three bacteriophage, M13, fl, and fd, are the most able to infect *E.Coli* cells, containing the F conjugative plasmid. Their genome has been

completely sequenced resulting 98% homologous DNA sequences [Hill, 1982; Beck, 1981;Van Wezenbeek, 1980].

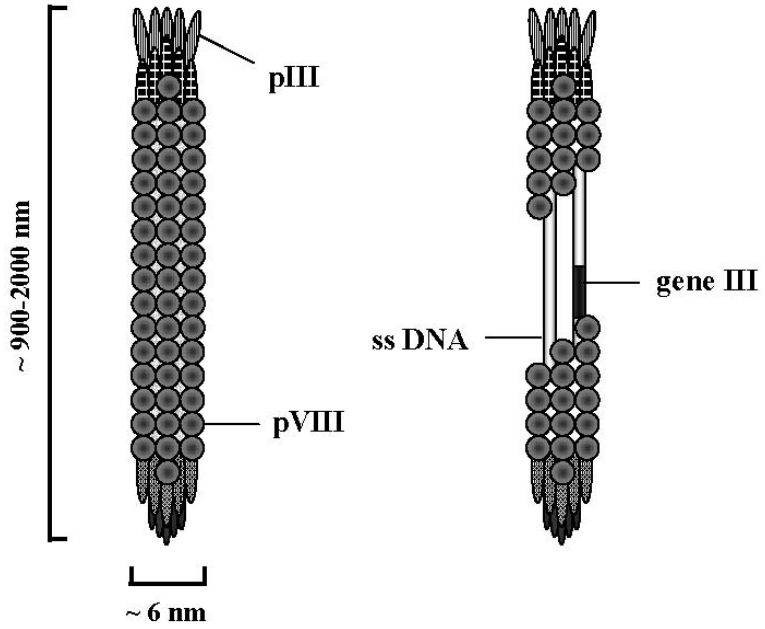
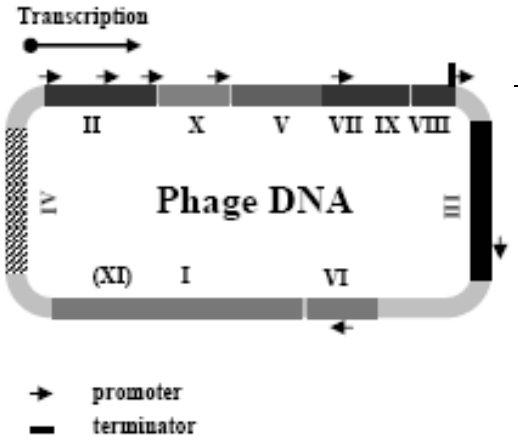


Figure 1.3 Schematic drawing of phage particles. *pIII is the minor coat protein three, while pVIII represents the major coat protein eight. The phage genome is constituted by a single stranded DNA (ss DNA) encased into a cylindrical phage particle.*

The entire genome of these phage consists of 11 genes (Figure 1.4). Two of these genes, X and XI, overlap and are in-frame with the larger genes II and I [Rapoza, 1995; Model, 1988]. The arrangement of the genes on DNA is based on their functions in the life cycle of the bacteriophage. Two genes (gII and gX) encode proteins required for DNA replication while a third one (gV) encodes for a protein necessary both at the assembly and DNA level; a group of three genes (gI, gIV and gXI) is involved in the phage assembly process at membrane level, while a last group encodes the capsid proteins. In addition to the regions which encode proteins, is the "Intergenic Region" which contains the sites of origin for the synthesis of the (+) strand (phage DNA) or (-) strand as well as a hairpin region which is the site of initiation for the assembly of the phage particles (packaging signal) (Figure 1.4). A phage expresses about 2700 copies of the major coat protein (pVIII, 50 aa long), and 3 to 5 copies of the minor coat protein (pIII, a 406 aa long) [Russel, 1991].



Gene	aa	MW	Function
I	348	39,502	Assembly
II	410	46,137	DNA replication
III	406	42,522	Minor capsid protein
IV	405	43,476	Assembly
V	87	9,682	Binding ssDNA
VI	112	12,342	Minor capsid protein
VII	33	3,599	Minor capsid protein
VIII	50	5,235	Major capsid protein
IX	32	3,650	Minor capsid protein
X	111	12,672	DNA replication
XI	108	12,424	Assembly

Figure 1.4 Genome and gene products of the fl bacteriophage. *The single-stranded DNA contains 11 genes, listed in the table. It has 6407 nucleotides which are numbered from the unique HindII site located in gene II.*

Infection of *E.Coli* by the bacteriophage is a two-step process. In the initial phase the pIII end of the phage particle interacts with the tip of the F conjugative pilus (Figure 1.5). The pilus is retracted so the tip of the phage is moved to the membrane surface. The retraction of the pilus presumably occurs the normal polymerization-depolymerization cycles inherent to the pili or may be the attachment of the phage can trigger pilus retraction [Frost, 1993].
The integration of the pVIII major capsid proteins and perhaps the other capsid proteins into the inner membrane together with the translocation of the DNA into the cytoplasm is the last step of infection. After that, the host DNA replication machinery converts the single-stranded phage DNA into the double-stranded plasmid like replicative form (RF). The RF serves as template for expression of the phage proteins and to produce new ssDNA. Phage progeny are assembled by packaging of ssDNA into protein coats and extruded through the bacterial membrane into the medium (Figure 1.5).

The assembly of filamentous phage begins where the inner and outer membranes of the *E.Coli* cells are in close contact [Lopez, 1985]. These assembly sites may be the result of specific interactions between pI, pIV

and pXI (Figure 1.5). The event that initiates the assembly of the particle is probably the interaction of the pV-phage DNA complex with proteins in the assembly site. The single-stranded DNA binding protein pV is 87 amino acids long and its biological functional entity is a homodimer; it is a multifunctional protein that not only is implicated as a scaffolding and/or chaperone during the phage assembly process, but regulates viral DNA replication and gene expression at the level of mRNA translation.
Subsequently the pV dimers are displaced from, and the capsid proteins added to the DNA during the extrusion through the bacterial envelope into the media.
The assembly process is conveniently divided into three parts, initiation, elongation and termination, reflecting the events required for packaging the different ends and the long cylinder of the phage [Endeman, 1995; Russel, 1991; Russel, 1989; Lim, 1985] (Figure 1.5).
Prior to assembly the coat proteins are imbedded in the inner membrane with the C termini in the cytoplasm. pVII and pIX are firstly incorporated at one end of the particle and then pVIII molecules are added along the length of the particle in thousands of copies.
The conclusion of the phage assembly process occurs when the end of the DNA is reached, and protein pVI and pIII added. In the absence of either of these proteins, assembly goes on with pVIII continuing to encapsulate another DNA producing polyphage containing multiple copies of the genome. The assembled phage particle is then released from the bacterial envelope into the extracellular environment.
Since the capsids are assembled around the DNA, there are no constraints on the length of DNA packaged. This property led to its use as a cloning vehicle. On the other hand, the membrane-associated assembly properties of the capsid proteins allow the packaging of chimeric proteins into the phage particle. The flexibility of the assembly process has led to an impressive array of applications for the use of phage display.
Filamentous phage does not produce a lytic infection in *E.Coli*, but rather induces a state in which the infected bacteria produce and secrete phage particles without undergoing lysis and the bacteria continue to grow and divide.

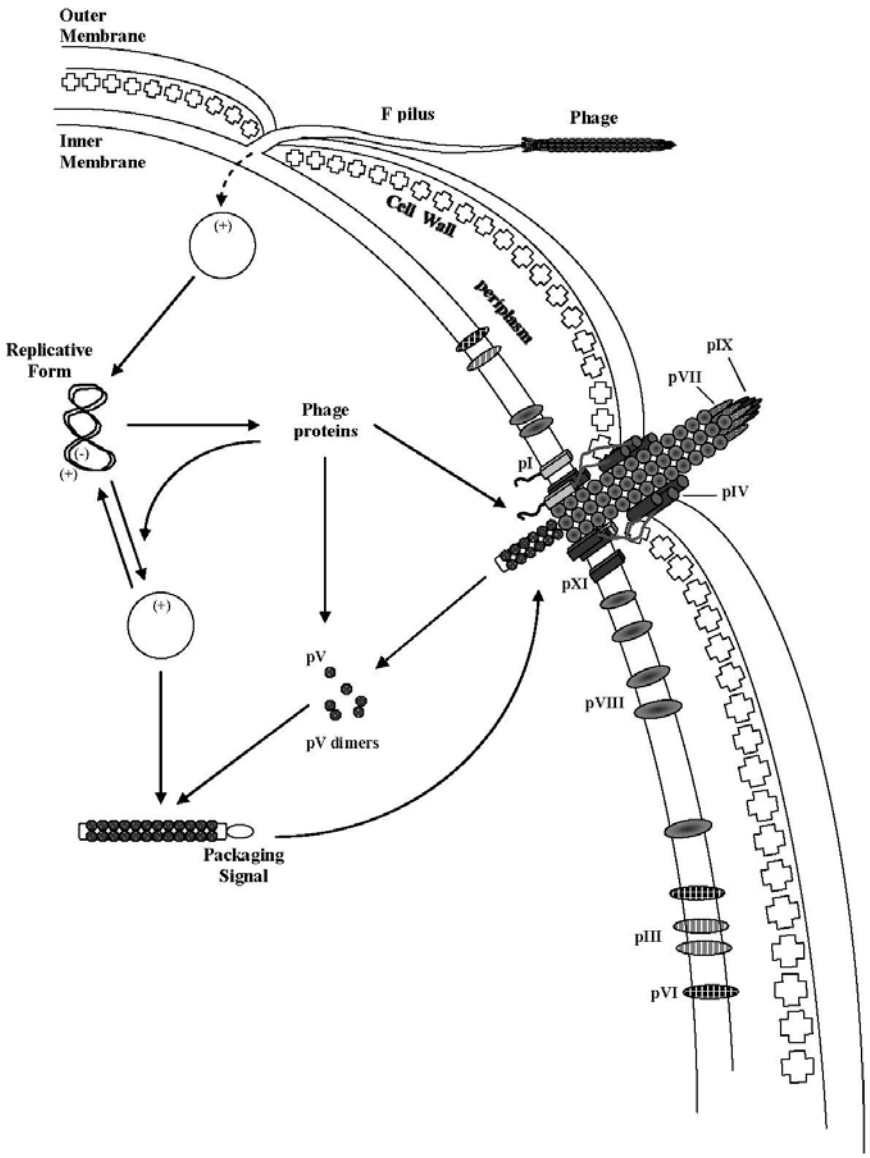


Figure 1.5 Schematic representation of the bacteriophage life cycle. *The phage starts the infection of E.Coli by the specific interaction with the tip of the F pilus. (+), the bacteriophage single-stranded DNA; (-), the complementary DNA strand; pV, the product of the bacteriophage gene V essential for the assembly process; pIII, pVI, pVII and pIX, the minor coat proteins; pVIII, the major coat protein; pI, pIV and pXI, the assembly proteins, the phage is extruded through an assembly site formed by these proteins.*

1.4.2 Phage display vector

According to Smith's classification [Smith, 1997, 1993], there are different formats for Phage Display (Figure 1.6):

– type 3 / type 8:

In these formats, the exogenous peptide or protein is expressed on the virion capsid as fusion protein to either the pIII or the pVIII protein. In particular, the corresponding encoding gene is cloned upstream of the gIII or gVIII into a wild-type viral vector. The result is that every copy of the capsid protein displays the fusion producing a multivalent display.

– type 33 / type 88:

In these systems, the vectors are phage that carry two copies of gene gIII or gVIII, one of which is fused with exogenous gene; bacterial cells infected with these phage incorporate both wild-type and fusion copies of pIII or pVIII into the same viral particles.

– type 3+3 / type 8+8:

In these formats, the protein fusion gene can be placed in a phagemid vector. Phagemids are hybrids of phage and plasmid vectors, usually containing an M13 origin of replication, the packaging signal site, multiple cloning sites and an antibiotic-resistance gene in addition to the elements required for plasmid propagation in *E.Coli* cells [Mead, 1988]. The phagemid replicates in *E.Coli* as a double-stranded plasmid and co-infection occurs with a helper phage, resulting in the production of single-stranded phagemid DNA, which is packaged into phage particles. The helper phage provides all the proteins necessary for phage assembly, including wild-type copies of all the coat proteins. In general, for these display formats are mainly used pIII as well as pVIII. In the case of using pIII, the resulting phage particles may incorporate either pIII derived from the helper phage or the pIII fusion protein, encoded by the phagemid, producing a monovalent display.

The level of display for different polypeptides varies greatly. The ratio between fusion proteins and wild type pIII may range between 1:9 and 1:1000, depending on both the length and the sequence of the displayed peptide and the growth conditions.

Using a 3+3 system the infectivity is not compromised because wild-type copies of pIII are also provided. Fused protein domains are more accessible when linked to domain 2 of pIII. Most protein display experiments have

utilized this truncated form of pIII for efficient display. However, replacement of domain I (aa 1-198) of pIII with an exogenous sequence leads to non-infectious phage when in multivalent phage display (Figure 1.7).

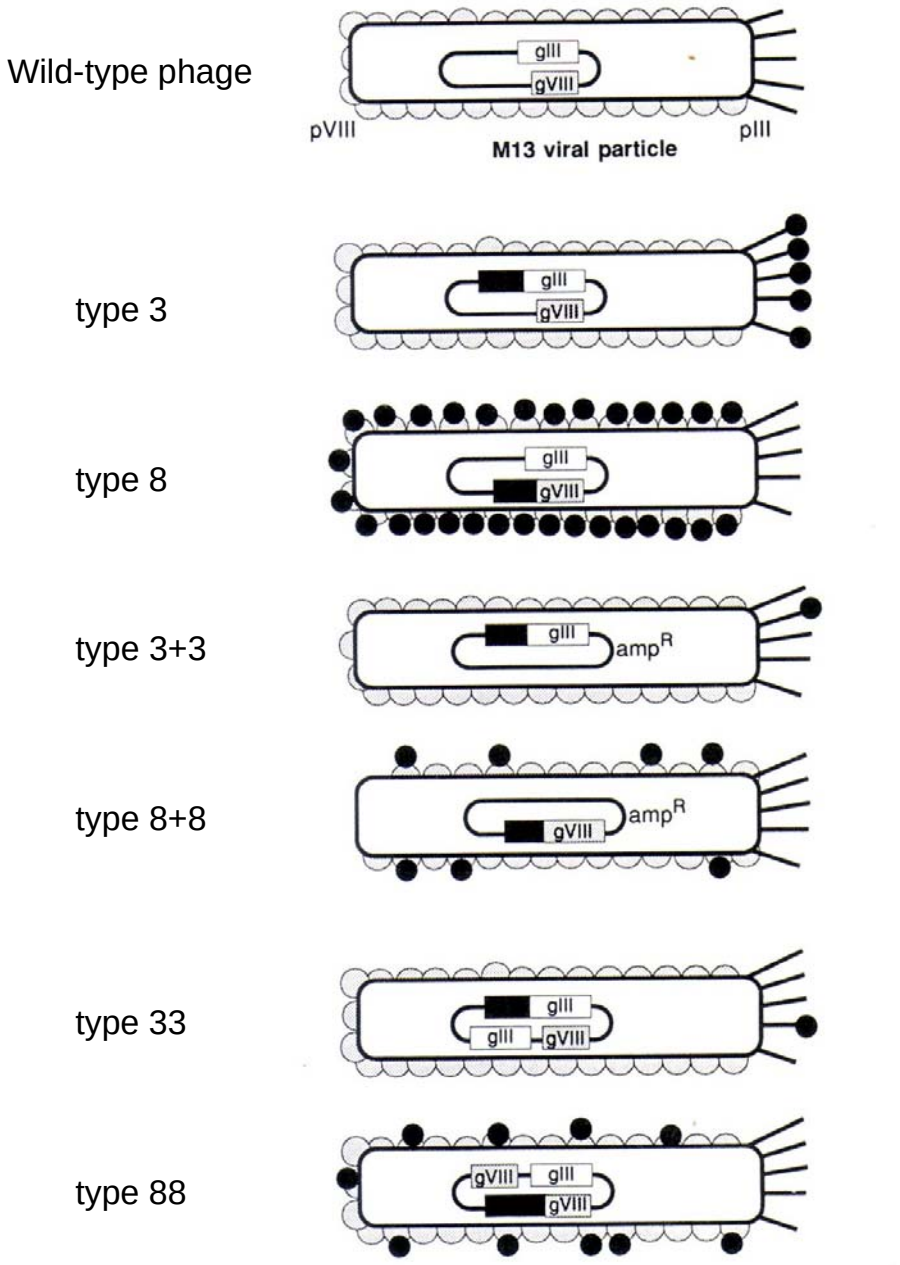
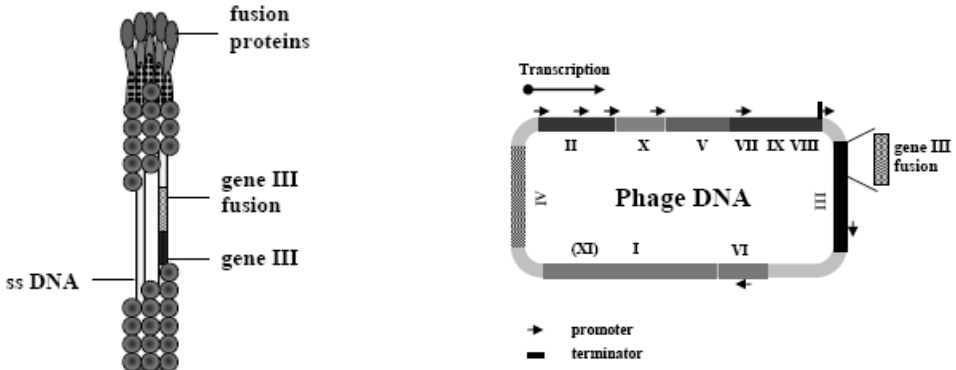


Figure 1.6 Classification of phage display vectors. *The black boxes and spheres correspond to the foreign genetic elements and their encoded peptides, respectively.*

3 MULTIVALENT PHAGE DISPLAY



3+3 MONOVALENT PHAGE DISPLAY

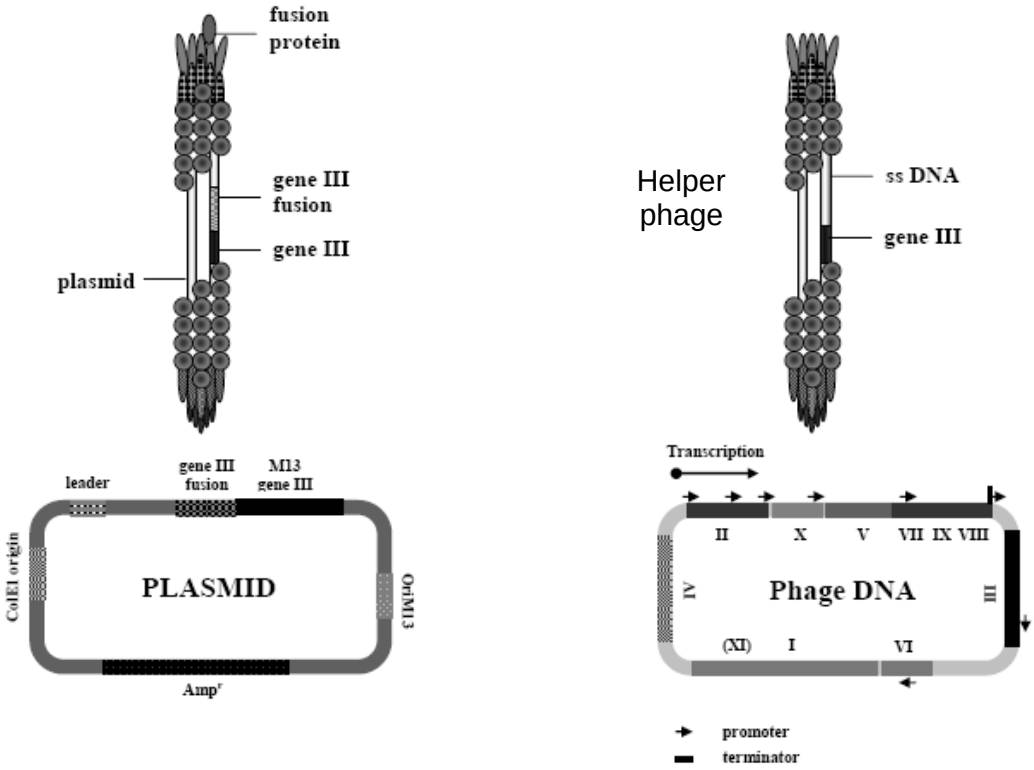


Figure 1.7 Phage system and Phagemid system in comparison.

In designing a vector for phage display, a series of consideration has to be taken into account, such as the insertion of antibiotic resistance genes, linkers, tags and stop codons. For example, a displayed element can be separated from pIII by short linkers. While there have been no formal experiments on the best linker sequences to use, many vectors use some variation of the sequence GGGGS. It is also possible to include a proteolytic cleavage site between the displayed peptide/protein and the capsid protein. While to reduce the level of non-recombinants in phage display libraries, a stop codon is inserted in the fragments of gene III. The stop codon, TAG, can be suppressed efficiently in bacteria containing *sup*E or *sup*F; when these vectors are propagated in such bacterial strains either a glutamine (Q) or tyrosine (Y) is inserted at the TAG codon, respectively. However, when TAG containing gene III is in bacterial strains that lack either *sup*E or *sup*F, no full-length pIII accumulates and therefore no virus particles are generated.
Furthermore, the vectors can be engineered to express short peptide "tags", recognized by specific antibodies, in the N terminal region of fused protein to immunologically discriminate between parental and recombinant phage particles. One such sequence is the c-myc epitope [Evan, 1985], which is recognized by MAb 9E10, or a hexa-histidine tag [Hochuli, 1987], which is useful for a recovery of a specific protein using by Ni-column affinity chromatography.

1.5 PEPTIDE LIBRARIES BY PHAGE DISPLAY – THE AIM OF WORK

It is noteworthy that the synthesis of peptide libraries with random amino acid sequences, followed by a procedure that selects new peptides with desirable properties, has become a relatively widely exploited area of research. Many investigators, however, have a biotechnological or pharmaceutical background and focus on the isolation of functional proteins [Siegel, 2001; Ladner, 2000, 1995; Neri, 1995 (I)]. Their main interest is the isolation of proteins with improved stability, new or optimized catalytic properties [Dumon, 2008; Seelig, 2007; Lingen, 2002; Meyer, 2002] or proteins that bind target molecules with enhanced affinity [Ho, 2006; Doorbar, 1994]. Generally, however, such investigations have been done starting from selected extant protein scaffolds and randomizing either restricted regions or the entire gene [Cirino, 2003; Murakami, 2003; Murakami, 2002; Shao, 1998; Cadwell, 1994]. Alternatively, using

recombination techniques, DNA fragments have been mixed to obtain novel combinations [Aguinaldo, 2003; Coco, 2003; Joern, 2003; Lutz, 2003; O'Maille, 2002]. These approaches can be defined as "directed randomizations" [Neylon, 2004], in the sense that randomization is performed in order to achieve certain desired properties.
Others investigations have been made in *de novo* sequence libraries with different levels in terms of randomization. For example, using a binary pattern of polar and non-polar amino acids [Bradley, 2006; Wei, 2003 (I-II); Moffet, 2001; West, 1999; Cordes, 1996; Kamtekar, 1993]. Nevertheless, using this method to construct the library, proteins were forced to assume a specific secondary structure [Hecht, 2004; Moffet, 2001]. An even more reductively methodology, in terms of sequence, has been used in Sauer's group [Davidson, 1995, 1994]. Here the idea has been to investigate the possibility to obtain stable three-dimensional structures into random sequence libraries prepared using 3 amino acids only: glutamine, leucine and arginine.
In contrast to these approaches the rationale of this work is completely different. It is aimed at investigating the frequency of folding and of catalytic activity in a totally random library where is no bias towards any given sequence or structural feature. This approach to create *de novo* libraries can be defined as a "total randomization".
We created 50 amino acid long random peptide libraries encoding proteins non extant in nature. As discussed previously, by synthesizing a random 50-mer chain using all 20 different amino acids it is theoretically possible to produce about 10^{65} different sequences and the probability to sample two identical chains is approximately equal to zero.
We chose the 50 amino acid length because it was quite short to resemble simple prebiotic oligopeptides but sufficient long to be folded. Indeed, in nature there are small proteins (around 30 amino acids long) with a stable three-dimensional structure, as the APP (Avian Pancreatic Peptide). On the contrary, smaller peptides are unstable since their mobility is increased by thermic agitation. Therefore, the 50 amino acid length seemed to be a good compromise. In addition, the structure of 50 amino acid long peptide can be investigated by NMR spectroscopy easily.
The random sequences were first cloned in phagemid vectors and then the resulted phagemid libraries were converted in phage libraries by the phage display technique.
The phage libraries were selected to look for NBPs with a folded structure and, parallel to this screening, NBPs were investigated for potential catalytic activities and possible interaction with other proteins.

According to the investigation of folded proteins, we used the resistance to proteolytic digestion as folding criterion. We chose thrombin as digestion enzyme because it is more selective than other proteases and not affected by flanking residues. On the contrary, other enzymes are inhibited when specific residues are near the cleavage site. As target site of thrombin we chose the PRG (proline-arginine-glycine) site which was placed in the middle of random sequence.
The base of this selection is that in a folded sequence the PRG is preserved during thrombin digestion while in un-folded sequence the PRG is exposed to enzyme cleavage.

The next step of this work concerns the selection of NBPs for their potential catalytic activity. Our screening strategy is based on the transition state theory [Jencks, 1969; Glasstone, 1941]. According to this theory, a catalyst enhances the reaction rate stabilizing high-energy transition state structures that are formed during the reaction. Therefore, any protein that stabilizes the transition state is a potential candidate for catalyzing that reaction.
In regard to the catalytic reaction, we focused on amide bond hydrolysis. This reaction spontaneously occurs in aqueous environment. As reported by Luisi and co-workers [Gorlero, 2008], under anhydrous conditions, or following the precipitation of the product, the entire equilibrium of the reaction can be shifted towards the synthesis. Within this framework, peptides capable of hydrolysing the amide bond may catalyse the reverse reaction, synthesizing short peptides. Accordingly, amide-hydrolytic peptides might have triggered off the emergence of primordial metabolism [Nakashima, 1980] or auto-catalytic peptides cycles [Kaufamann, 1996].
In addition, amide hydrolysis involves the formation of a high-energy tetrahedral intermediate [Tanaka, 2002] which can be mimicked by a phosphonates or phosphomamidates TSA (Figure 1.8).

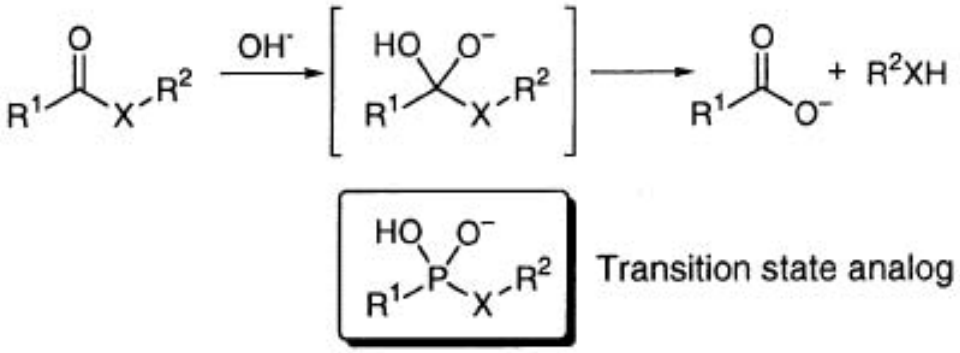


Figure 1.8 Molecular structure of TSA. *The upper part of figure represents the mechanism of amide hydrolysis, in the lower there is the structure of transition state analog (TSA), that can be a phosphonate (X=O) or a phosphonamidate (X=NH). (Adapted from Tanaka, 2002).*

The last part of this work explores the possible interactions between NBPs and other proteins, in particular proteases. In theory, peptides are able to bind to a protein anywhere on its solvent-exposed surface. However, most peptides bind at sites that coincide with natural ligand-binding sites. In fact, this site seems to have features that predispose it for ligand binding [Sidhu, 2000]. Consequently, peptides bound to the natural ligand-binding site could act as inhibitors/activators like the natural ligand, even with better results. Therefore, the aim of our work is to select NBPs against four proteases immobilized by adsorption on a plate to have potential new inhibitors/ activators of target-proteases. Positive selected sequences, with a demonstrated activity, could be used in pharmaceutical and biotechnological field.

2. RESULTS

The first part of this work describes the development of 3 different phagemid vectors employed to display a totally *de novo* random library encoding totally new proteins (the Never Born Proteins).
The 150 bp long DNA library is totally random except for tripeptide proline-arginine-glycine (PRG) which is the substrate for thrombin used for folding selection. Another shared characteristic is the hexa-histidine tag (His-tag) at the N-terminal of all sequences, useful for the recovery of NBPs. These sequences were inserted into each phagemid vector described hereafter, obtaining 3 different phagemid libraries. Each library was expressed by the phage display technique. Using the 3+3 monovalent phage system (described in section 1.4.2), a single NBP-pIII fusion protein was displayed on each phage (Figure 2.1). The other pIII proteins were wild-type and thus phage infectivity was not compromised.

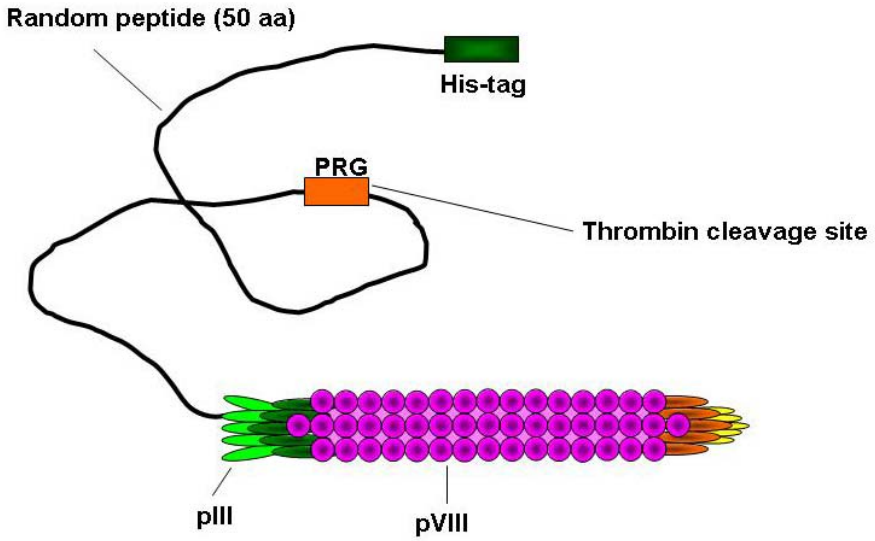


Figure 2.1 Schematic drawing of random peptide fused with pIII. *The random peptide is exposed as fusion protein to pIII (minor coat protein) on the top of the phage while pVIII represents the major coat protein. The PRG (proline-arginine-glycine) site is in the middle of random peptide, while the His-tag is exposed at the N-terminal of the random sequence.*

Phage libraries were selected for several characteristics using the ELISA panning technique. The first selection was to look for three-dimensional structures using the resistance to thrombin digestion as folding criterion.
In parallel, a probable catalytic activity was tested, screening phage libraries against a TSA target, which mimics the transition state of an amide bond hydrolysis. Finally, NBPs were selected to analyse the possible interaction with other proteins, in particular 4 different proteases: papain, pepsin, trypsin and α-chymotrypsin.

2.1 CONSTRUCTION OF PHAGEMID VECTORS

2.1.1 pIII-DUMMY-His-tag

The pIII-DUMMY-His-tag comes from pOCI1050 c-myc (courtesy of Dr. Chiarabelli, University of Roma Tre) that allows in-frame expression of:

- the pelB signal sequence, to ensure a reliable membrane translocation;
- the c-myc tag sequence (16 residues long), to recognize by MAb 9E10 antibody;
- a linker;
- a gene of interest;
- the C-terminal part of gene III (sequence D197-S406).

The signal sequence is cleaved when the protein is exported to the periplasm. A gene of interest can be cloned into the unique *NotI, XbaI* sites and an amber codon is located between the gene of interest and the C-terminally truncated gene III (Figure 2.2).

One of the main differences between pIII-DUMMY-His-tag and pOCI1050 c-myc is the His-tag replaces the c-myc tag. This involves a change of the purification system: His-tag fusion proteins can be identified by Immobilized metal affinity chromatography (IMAC), whose efficiency is higher than antibody purification.
In addition, the new vector has the “Dummy” sequence and the Xa site. The first is an insertion sequence (738 bp long) harbouring 3 stop codons in the 3 different reading frames in order to allow visual control during the cloning step and minimise the production of non-recombinant fusion proteins in case of phagemid self-ligation. The Xa site is the cleavage site of a protease,

the activated coagulation Factor X, to allow the purification of exogenous proteins separately from the tag.

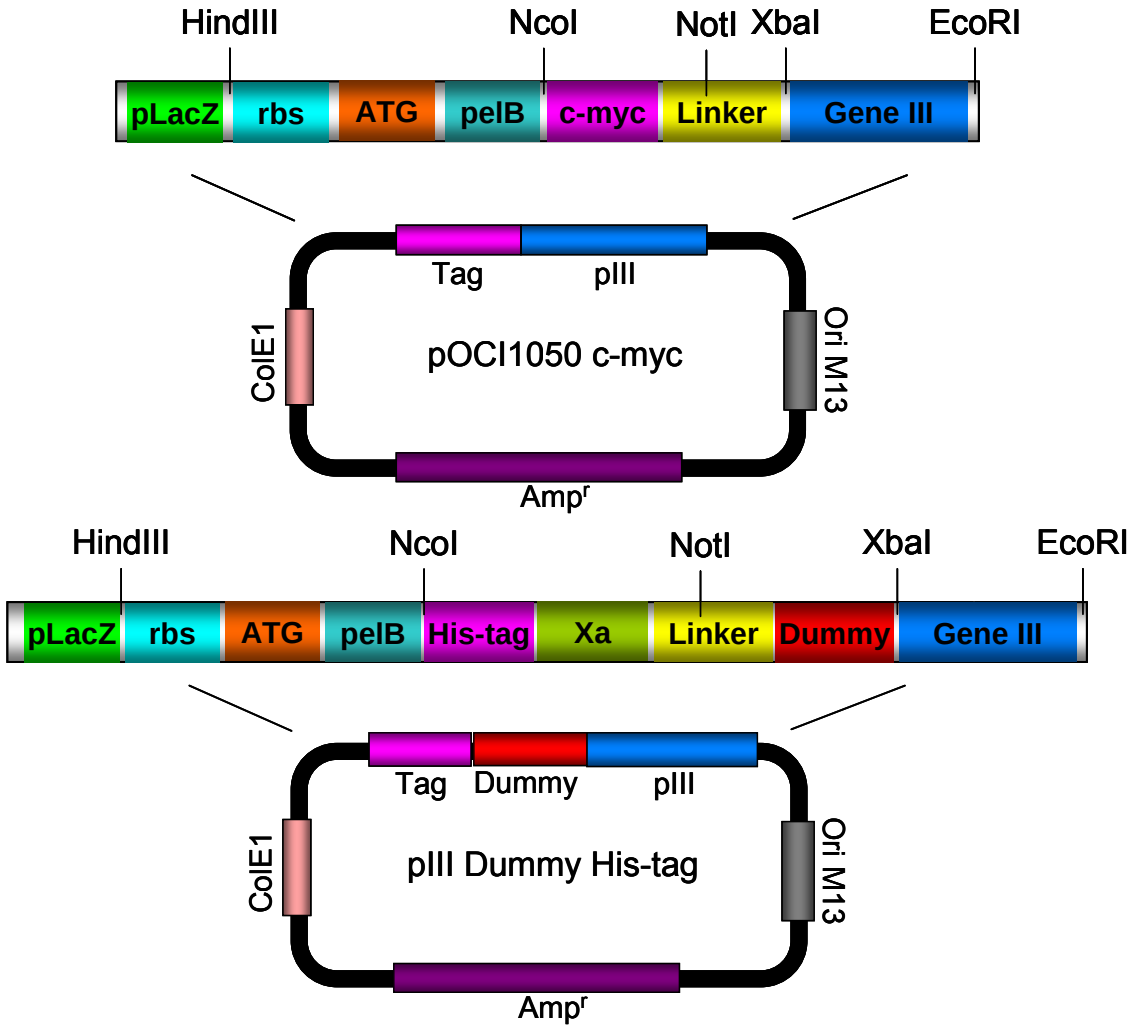


Figure 2.2 Maps of phagemid vectors pOCI1050 and pIII-DUMMY-His-tag . *Both vectors have a ribosome binding sequence (RBS) upstream pelB and a p-Lac promoter (pLacZ gene) to control the expression of foreign proteins. These last can be inserted using NotI and XbaI cloning sites.*

2.1.2 pOCI1050 His-tag

This vector derives from pOCI1050 c-myc described before. The c-myc tag is replaced with a His-tag from pET14b (Novagen) cloning into the pOCI1050 c-myc between *NcoI* and *NotI* sites (upstream and downstream of c-myc sequence respectively) (Figure 2.3).

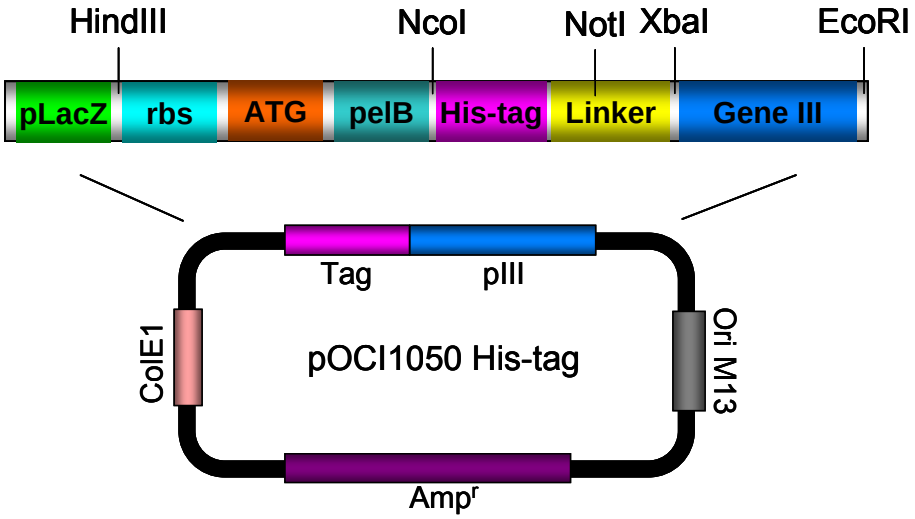
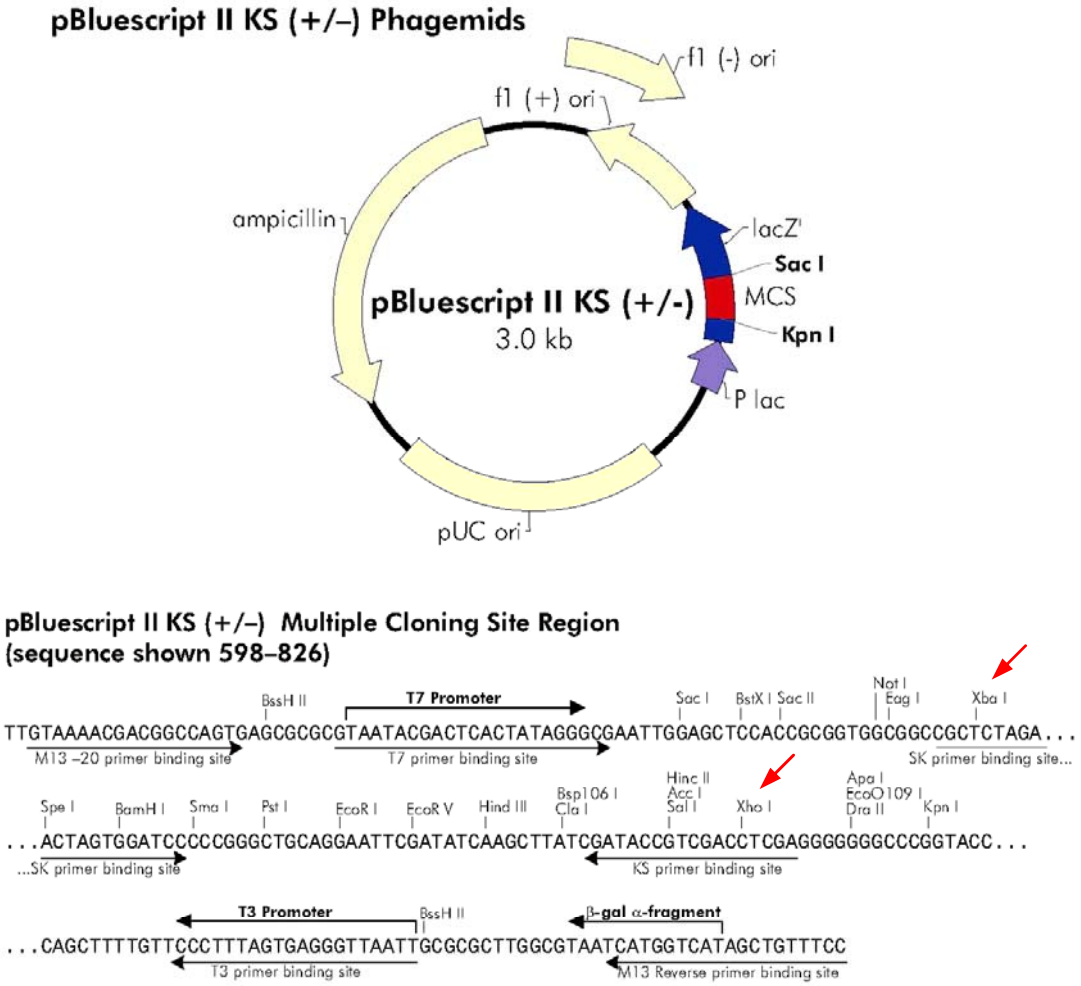


Figure 2.3 Map of pOCI1050 His-tag. *The vector has a ribosome binding sequence (RBS) upstream pelB and a p-Lac promoter (pLacZ gene) to control the expression of foreign proteins. These last can be inserted using NotI and XbaI cloning sites.*

2.1.3 pBluescript II KS His-tag

The pBluescript II KS His-tag derives from a commercial vector, the pBluescript II KS.
The pBluescript II phagemids have an extensive polylinker with 21 unique restriction enzyme recognition sites. The polylinker and T7 and T3 RNA polymerase promoter sequences are present in the N-terminal portion of a lacZ gene fragment. A total of 131 amino acids of β-galactosidase coding sequence are present in the pBluescript II phagemid, but the coding sequence is interrupted by the large polylinker (Figure 2.4).



Feature	Nucleotide Position
f1 (+) origin of ss-DNA replication [pBluescript KS (+) only]	135–441
f1 (–) origin of ss-DNA replication [pBluescript KS (–) only]	21–327
β-galactosidase α-fragment coding sequence (lacZ')	460–816
multiple cloning site	653–760
T7 promoter transcription initiation site	643
T3 promoter transcription initiation site	774
lac promoter	817–938
pUC origin of replication	1158–1825
ampicillin resistance (*bla*) ORF	1976–2833

Figure 2.4 Circular map and list of features for the pBluescript II KS (wild-type). *Red arrows indicate plasmid restriction sites eliminated.*

The pBluescript II KS wild-type was modified to insert a His-tag and change its restriction system. The first modification consisted of the inclusion of a new fragment, corresponding to the region in pOCI 1050 His-tag between the rbs and gIII (~790 bp long) (Figure 2.3), which was cloned into the pBluescript II KS between *HindIII* and *EcoRI* sites (Figure 2.5).
After insertion of the His-tag, the new plasmid was modified in its restriction system to allow cloning of the library. The *XhoI* site in the pBluescript II KS (Figure 2.4) was removed to avoid problems during the last step of the library cloning. In fact, to obtain the DNA sequence coding for the PRG site it is necessary to digest the DNA annealing region with the *XhoI* enzyme, as described in the next section. To modify this plasmid region, pBluescript His-tag was digested with *SalI* and *XhoI*. The small fragment between the two enzyme sites was then eliminated by purification and plasmid reclosed.
The *XbaI* site in the pBluescript II KS (Figure 2.4) was also removed because it is present a second *XbaI* site in the new fragment inserted into the pBluescript II KS during the first modification. To do this, the plasmid was digested with *SmaI* and *SacI* and the small fragment between the two enzyme sites was eliminated by purification. After incubation with Klenow polymerase to remove the 5' protruding, the pBluescript His-tag blunt ends were ligated.

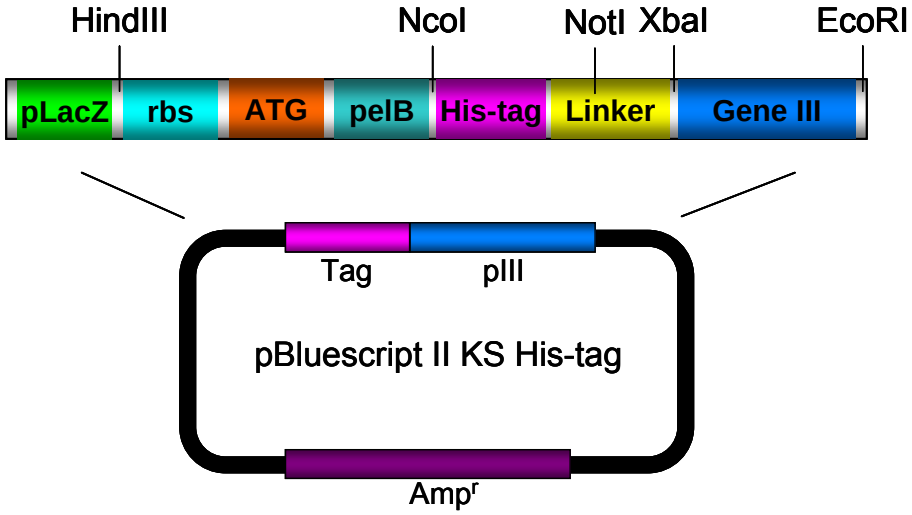


Figure 2.5 pBluescript II KS His-tag map. *The vector has a ribosome binding sequence (RBS) upstream pelB and a p-Lac promoter (pLacZ gene) to control the expression of foreign proteins. These last can be inserted using NotI and XbaI cloning sites.*

2.2 RANDOM LIBRARY

2.2.1 Construction of phagemid library

The DNA library cloned in the phagemid systems was synthesized using two groups of oligonucleotides (forward and reverse) with codon schemes NNK and NNM, where N is an equimolar mixture of all four bases and where K is either G or T and M either C or A. These schemes use 32 codons to encode all 20 amino acids and 1 stop codon (TAG), yielding an acceptably low frequency of stop codons when used to encode short polypeptides.

The oligonucleotides consist of random nucleotides, encoding 23 amino acids in the forward group and 24 in the reverse, flanked by 11-18 fixed residues that are necessary for annealing and cloning (Figure 2.6). After annealing and incubation with RedTaq polimerase, the DNA library was cloned in all described vectors using the unique *NotI*, *XbaI* sites. The bases in excess in the annealing region were subsequently removed by cleavage with *XhoI* restriction enzyme and in this way PRG site was formed.

All the vectors were transformed into *E.Coli* cells (XL1Blue MRF' strain) by electroporation. This technique was chosen because produces higher efficiency than the best chemical methods and this is crucial in order to achieve large, representative primary libraries.
The final number of different random sequences present in each library in the bacteria was estimated to be about 10^7 per ml.

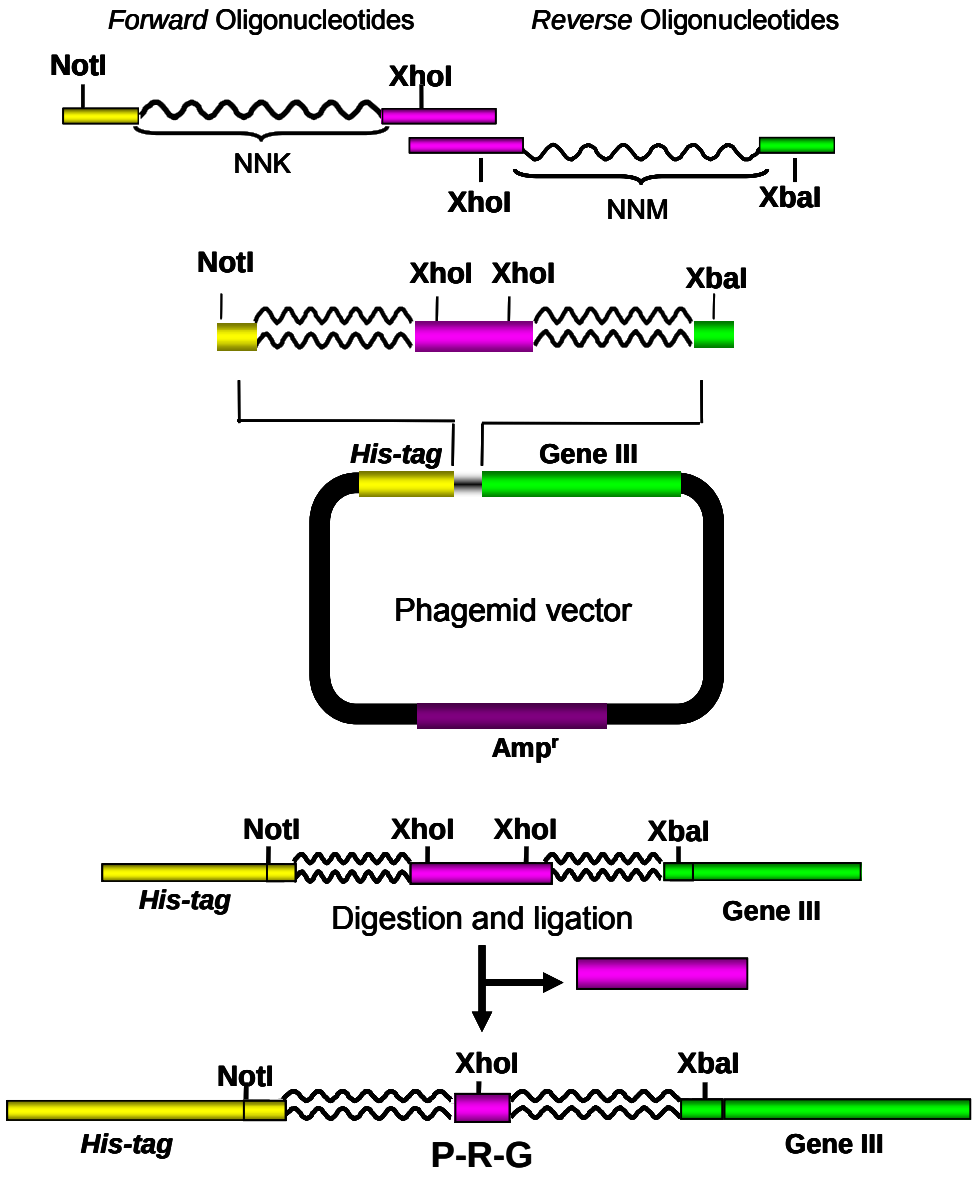


Figure 2.6 Library construction.

2.2.2 Computational analysis of the random sequences

From each phagemid library, we analyzed DNA sequence of 10 samples at random to verify their correctness and randomness. The majority of sequences, about 80%, were found to be correct with random sequence in-frame in the vector. The remaining part of the sequences consisted of either not in-frame sequences or vectors lacking random sequences.

Of all the correct sequences, randomness at the amino acid level was analysed comparing experimental and expected data (Figure 2.7). The graph shows that the random sequences have an amino acid frequency in line with the expected data. The experimental value of glutamine residues (Q) is the result of both, the codons for Q in the random sequences, and the translation of TAG codons into Q residues due to the *E.Coli* Amber mutation strain used.

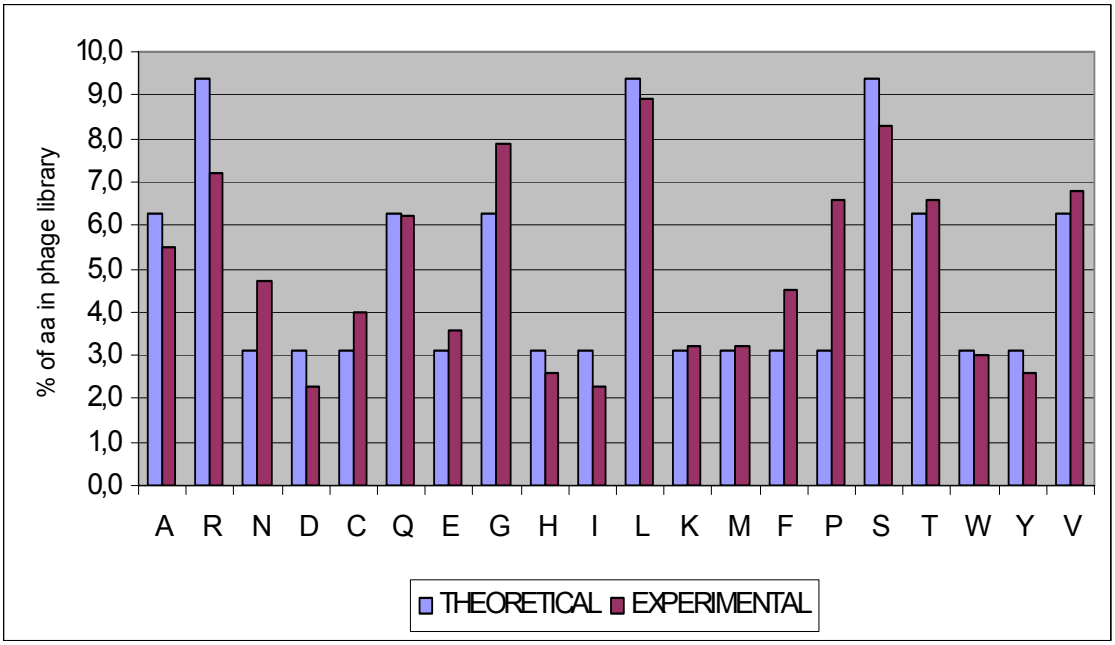


Figure 2.7 Randomness of phage peptide library at amino acid level

Randomness at a nucleic acid level of the correct clones was also investigated (Figure 2.8) taking into account the NNK scheme used in the design of the library. Also in this analysis experimental values are in line with theoretical ones.

	1st position		2nd position		3rd position	
	%Theor.	%Exp.	%Theor.	%Exp.	%Theor.	%Exp.
T	25	24.7	25	25.5	50	49.8
C	25	23.3	25	24.4	0	2.2
A	25	25.7	25	24.3	0	0
G	25	26.3	25	25.8	50	47.9

Figure 2.8 Randomness of phage peptide library at nucleic acid level.

2.3 PHAGE PRODUCTION

E.Coli cells, electroporated with all phagemid libraries, are used to produce 3 different phage libraries.
As mentioned in the introduction, a co-infection with a helper phage is necessary for the production of phage particles, with the phagemid incapsidated and a single NBP-pIII fusion protein on the surface.
Helper phage M13K07 used in this step is an M13 derivative. It is able to replicate itself in the absence of phagemid, but, in the presence of a phagemid bearing a wild-type M13 origin, single-stranded phagemid is packaged preferentially.
Electroporated cells need also of IPTG, since fusion gene (random sequence-gIII) is controlled by a pLac promoter.
Four different types of phage come from co-infected cells (Figure 2.9), but only one is the recombinant phage with exogenous peptide on the capsid and the corresponding exogenous gene in its genome. Only the recombinant phage is suitable for biopanning selections.

The phage libraries were recovered by precipitation with PEG 6000 + NaCl and solubilized in PBS buffer. The phage titre was estimated to be about 10^{11} cfu/ml (Colony Forming Unit per millilitre) for all phage libraries.

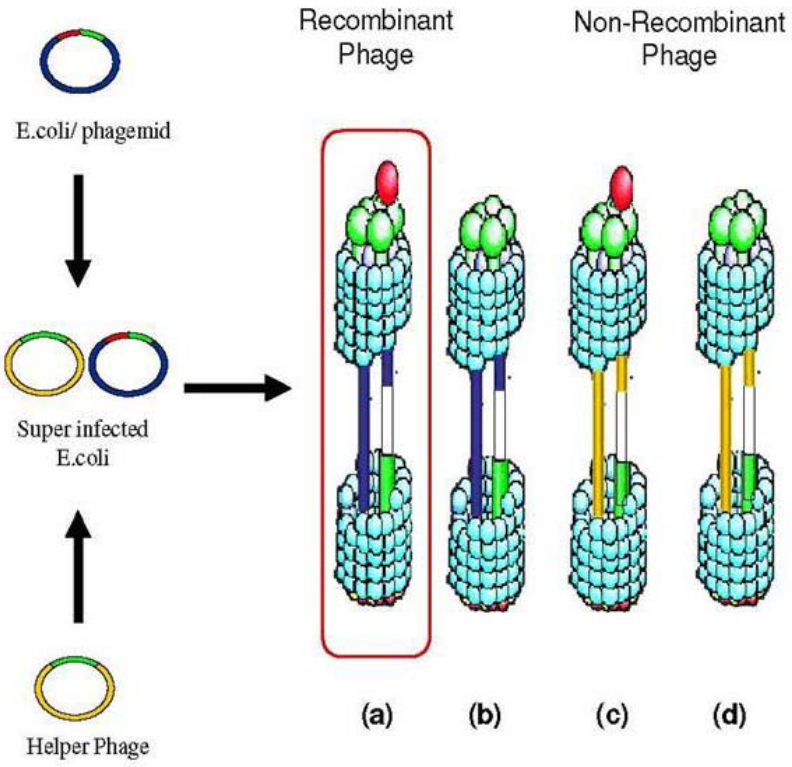


Figure 2.9 Overview of the possible phage types produced during E. Coli infection. *(a) Recombinant phage with exogenous peptide (phenotype) and the corresponding encoding gene (genotype); (b) non-recombinant phage lacking the phenotype; (c) non-recombinant phage lacking the genotype; (d) non-recombinant phage lacking both phenotype and genotype, i.e. helper phage.*

2.4 FOLDING SELECTION BY THROMBIN DIGESTION

Resistance to thrombin digestion was used as a first, rudimentary folding criterion.

Formed from prothrombin, thrombin is a serine protease: a protein-cutting enzyme that uses a serine amino acid to perform the cleavage (Figure 2.10). Other examples of serine proteases are trypsin and chymotrypsin, enzymes involved in digestion. Thrombin, however, is more specific than these gastrointestinal cleavage machines. It selectively cleaves Arg-Gly bonds in fibrinogen to form fibrin and release fibrinopeptides A and B.

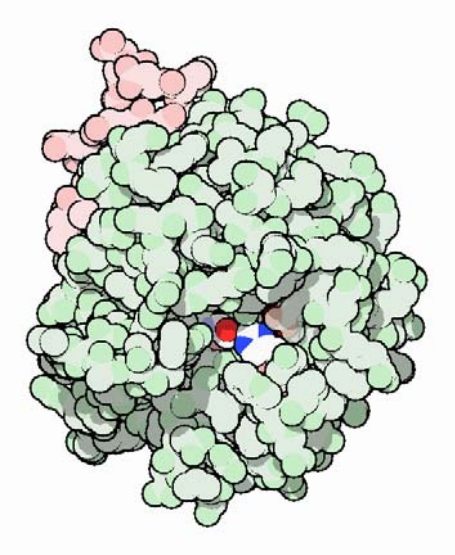


Figure 2.10 Thrombin structure. *Thrombin is a serine protease, the active site is localized in the structure of activated thrombin at the base of a deep groove. The oxygen atom of the key serine amino acid is shown in bright red, and the two bright blue nitrogen atoms are part of a histidine that activates the serine.*

All random sequences have in the middle a PRG (proline-arginine-glycine) site, corresponding to a thrombin cleavage site. Folded sequences preserve PRG during thrombin digestion, while un-folded sequences are cut since PRG is exposed to enzyme cleavage.

2.4.1 Validation of selection method

Immobilized metal affinity chromatography (IMAC) is widely used for the purification and identification of recombinant fusion proteins with histidine tags. The affinity of the His-tag for the nickel chelate is sequence dependent but generally very high. This allows the histidine-containing protein to be captured on a solid support (agarose, multiwell plate, magnetic beads, etc.) that contains a chelated nickel ion.
In this work, the HIS-Select® High Capacity (HC) Nickel Coated Plates were chosen for two main reasons: they can recover little amount of proteins with His-tag and their 96 wells allow to screen several samples and/or several conditions at once. A preliminary incubation of phage library resulted in a capacity affinity binding of $7 \cdot 10^5$ phage per µl.

2.4.2 Selection cycles

All 3 phage libraries were separately loaded into the HIS-Select® High Capacity (HC) Nickel Coated Plate separately at a final concentration in each well of 10^8 cfu/ml.
The exposed His-tags were bound by Ni. The un-bound phage were removed by washing; then thrombin was added into wells and sequences with exposed PRG were cut (Figure 2.11).
Un-bound phage were rid of by washing; un-cut phage were eluted by an Imidazole buffer, titrated and used to infect *E. Coli* cells, producing a new phage population enriched with positively selected sequences. This population was used to do the second selection cycle, using the same criterion of the first.

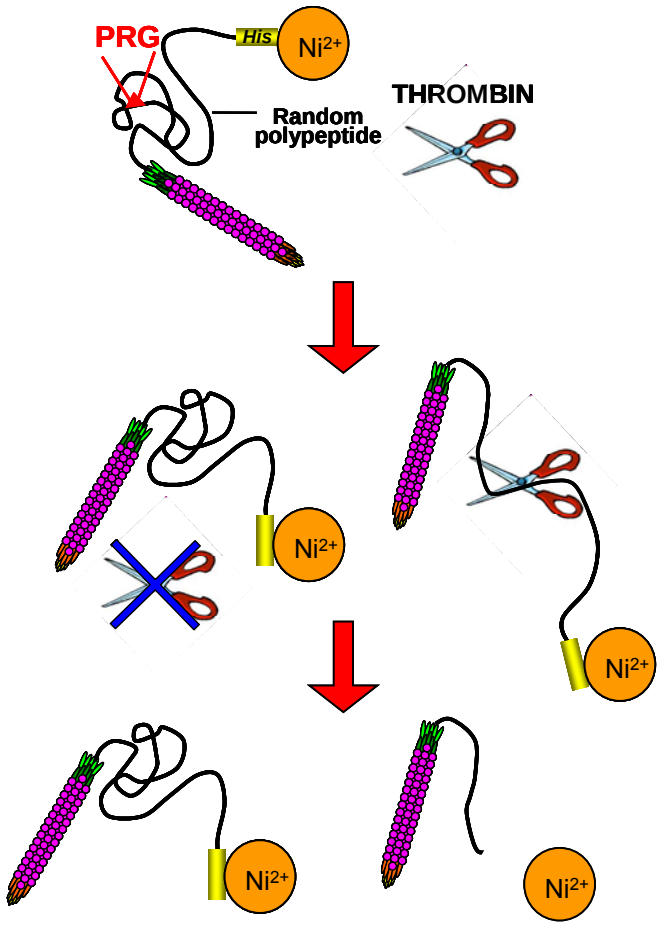


Figure 2.11 Scheme of fold selection by thrombin. *After incubation with thrombin, the un-fold peptide is cleaved and phage cannot be recognized by Nickel ion. On the contrary, folded peptide is not digested and phage remains bound on the plate.*

Three types of control were done: two negative controls (C_1- and C_2-), consisting of helper phage M13K07 added, respectively, into wells without thrombin, to test the affinity between capsid phage and Ni, and into wells with thrombin, to monitor possible cleavages of capsid proteins. One positive control (C+) was also performed, consisting of phage libraries into wells without thrombin, to test the total amount of phage that could be recovered.
Figure 2.12 shows the results after the second bio-panning cycle. No difference in the yield of the 3 phagemid libraries was found, therefore only one graph for all 3 selections is shown.
The efficiency of the method used was supported by the large number of recovered phage from C+, such number was very similar to the starting phage titre. The small amount of phage from C_1- resulted in a low affinity between capsid proteins and Ni. The phage titre from C_2- was comparable to C_1- indicating that thrombin did not digest capsid proteins at all.
An important amount of phage library was recovered (indicated with L in the graph), around 10^3 cfu/ml, indicating a good chance of selecting folded NBPs.

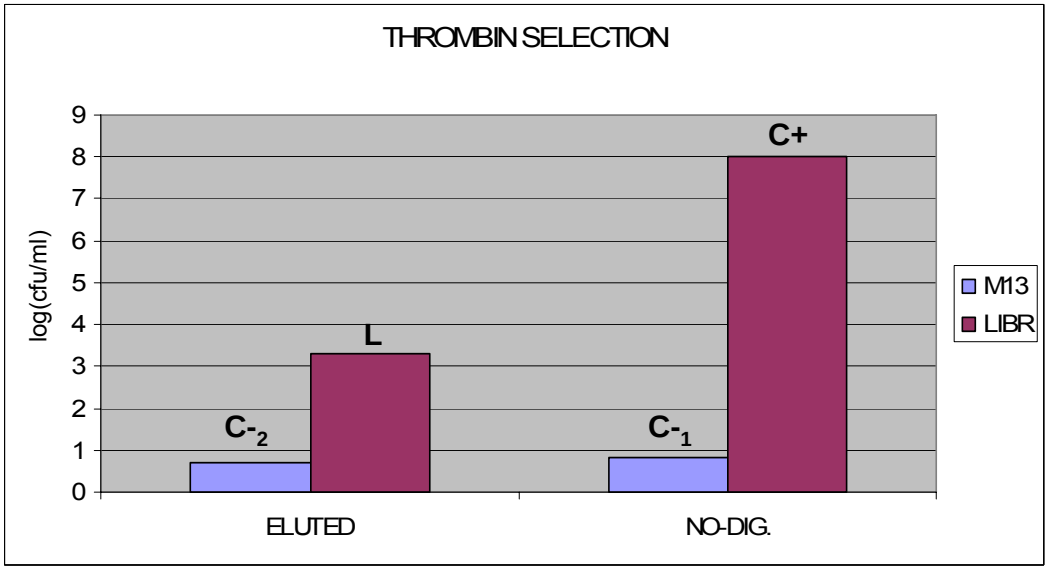


Figure 2.12 Selection of libraries by thrombin digestion. *This graph shows the selection results after the second selection cycle for all 3 phagemidic systems. C+ represents recovered phage libraries without thrombin digestion; C_1- corresponds to recovered helper phage without thrombin digestion; C_2- is recovered helper phage after thrombin digestion and L indicates recovered phage libraries after thrombin digestion.*

2.5 SELECTION AGAINST TSA

The chosen selection criterion to discover functional peptides in a completely *de novo* random library of 50 amino acid long peptides relies on the capability of screened peptides to bind to a transition state analogue (TSA) for the hydrolysis of amide bond.
The procedure involves the biopanning of the phage library on to a solid surface coated with the TSA.
The coupling reaction occurs between the reactive acid group of the TSA (p-nitrophenyl hydrogen-4-[hydroxycarbonyl]butylphosphonate) (Figure 2.13) and the secondary amine on the functionalised plate. This bond is promoted by sulfo-NH (N-hydroxysulfosuccinimide sodium salt) that activates TSA's carboxyl and EDC (N-[3-Dimethylaminopropyl]-N′-ethylcarbodiimide hydrochloride) that acts as a condensing agent.

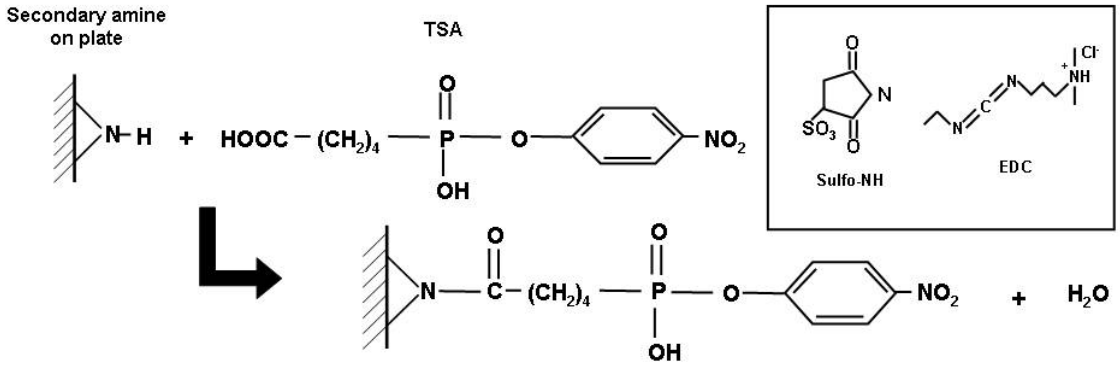


Figure 2.13 Reaction mechanism for the coupling of TSA to amine-functionalised solid surface

2.5.1 TSA stability check

The stability of the TSA was tested at different pH conditions in order to evaluate the possibility of screening the peptide library on different pH. The pH-dependent hydrolysis of the TSA was measured by absorbance at 405 nm of the released p-nitro phenolate group (Figure 2.14).
The results obtained showed that TSA molecule was stable up to pH 8, so the selection of NBPs against TSA were made at pH 7.

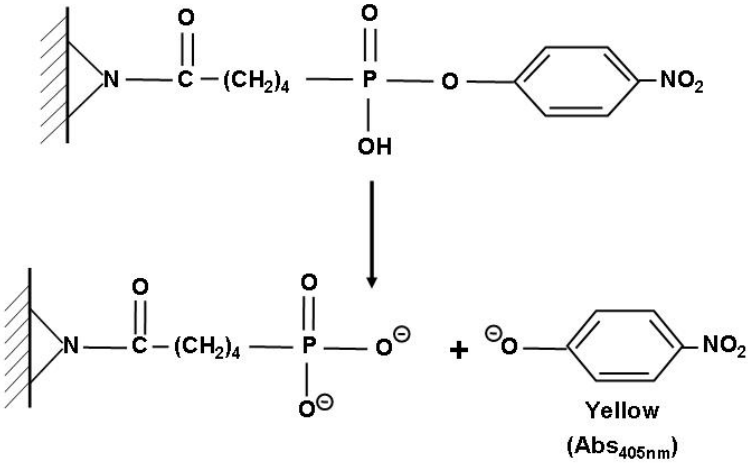


Figure 2.14 Hydrolysis of p-nitro phenolate group of TSA.

2.5.2 Validation of TSA coupling

The possibility of monitoring the TSA hydrolysis by absorbance was used to evaluate TSA coupling on the plate.
In practice, TSA pre-incubated wells were treated with a basic buffer and their absorbance value compared with a blank (the basic buffer). The occurred coupling was indicated by a greater absorbance value of samples taken from TSA pre-incubated wells. In fact the hydrolysis of TSA with the released of the phenolate group was the cause of the increase in Absorbance at 405 nm.

2.5.3 Selection cycles

The selection was made at pH7; in addition the influence of a metallic cofactor, zinc, was tested.
All the 3 phage libraries were loaded separately into the plate CovaLink™ NH MicroWells™, in particular in 3 kind of wells: TSA-wells (with TSA coupled), NH-wells (with sulfo-NH and EDC) and empty wells. The final phage concentration in each well was 10^9 cfu/ml.
After the incubation period, un-bound phage were removed by washing and bound phage were eluted by a high TSA concentration, titrated and used to infect *E.Coli*. The new phage population, enriched with selected sequences, was used in the second panning cycle. In this step, more washes were done to increase the specificity of selected NBPs (Figure 2.15).

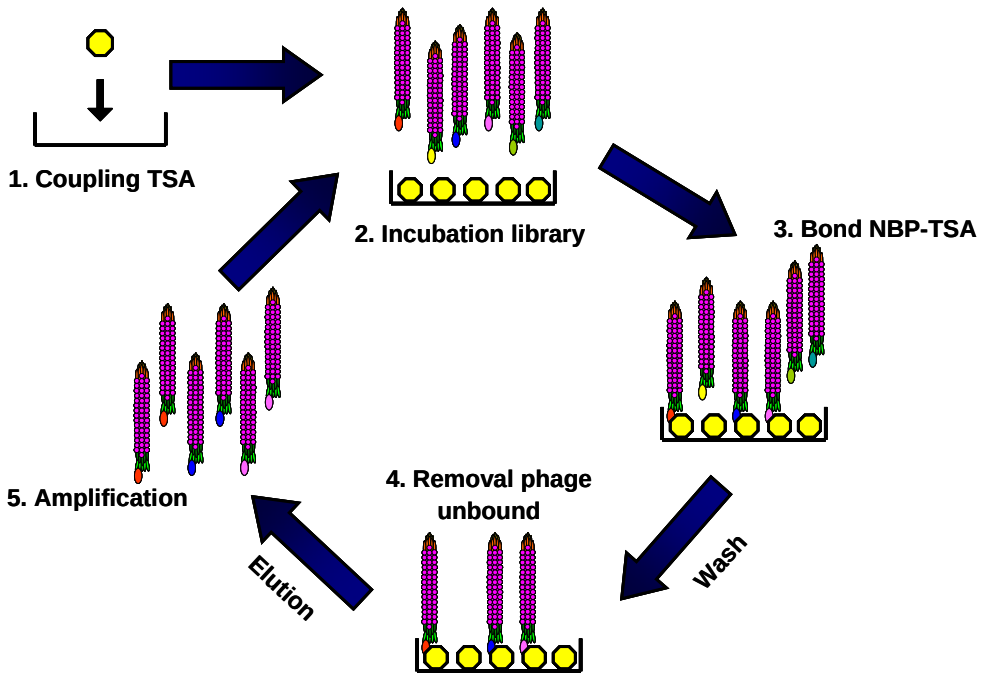


Figure 2.15 Scheme of TSA selection cycle.

One negative control was done: helper phage M13K07 was incubated and processed in the same way of the phage libraries, using the 3 types of well, to monitor the possible affinity between capsid proteins and TSA (in TSA-wells) / coupling reagents (in NH-wells) / surface plate (in empty wells). No phage were recovered from any wells, so capsid proteins do not have affinity with any components into the well.

Figure 2.16 shows the results after the second cycle of bio-panning. The affinity between NBPs and coupling reagents as well as the plate surface was very low, in fact the phage amount recovered from NH-wells and empty wells was very low. On the contrary, a significant phage titre resulted from TSA-wells, around 10^5 cfu/ml. The metallic cofactor does not seem to affect the affinity for the target, in fact the phage quantity recovered from the plate is similar both with and without Zn.

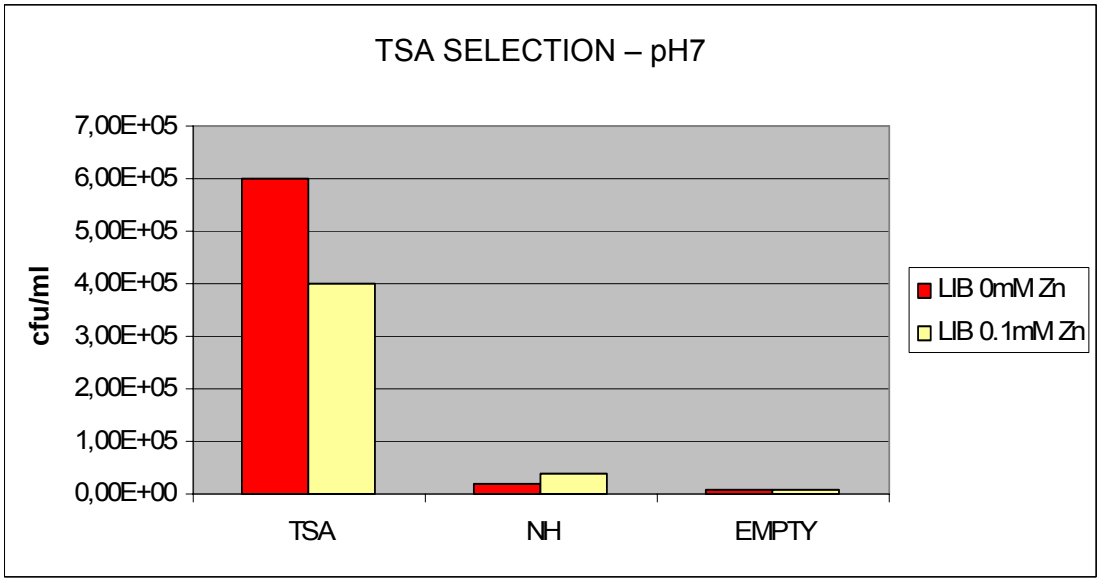


Figure 2.16 Selection of de novo library against TSA. *This graph shows the selection results after the second selection cycle for all 3 phagemidic systems. "TSA" indicates eluted phage from wells pre-incubated with TSA; "NH" those from wells pre-incubated with only NH-sulfo and EDC; "EMPTY" phage from bare wells.*

2.6 SELECTION AGAINST PROTEASES

The last selection was done to test possible interactions between NBPs and other proteins and investigate their possible use as inhibitors/activators of the target proteins.
The basic idea of this selection is that a ligand, to act as an inhibitor/ activator of a protein, first of all, has to bind such protein.
Chosen targets were 4 proteases: papain, pepsin, trypsin and α-chymotrypsin.

2.6.1 Selection cycles

Each protease was immobilized on a polystyrene plate (Maxisorp Immuno-Plate) by adsorption, after incubation with BSA 1% and washing, phage libraries were loaded separately in two kind of wells: one with a single protease and one with BSA 1%. The final phage concentration in each well was 10^9 cfu/ml.

After the incubation at pH7, un-bound phage were removed by washing; bound phage were eluted by a glycine buffer, titrated and used to infect *E.Coli* cells. The new phage population, enriched with selected sequences, was used to do the second panning cycle. In this step, more washes were done to increase the specificity of selected NBPs (Figure 2.17).

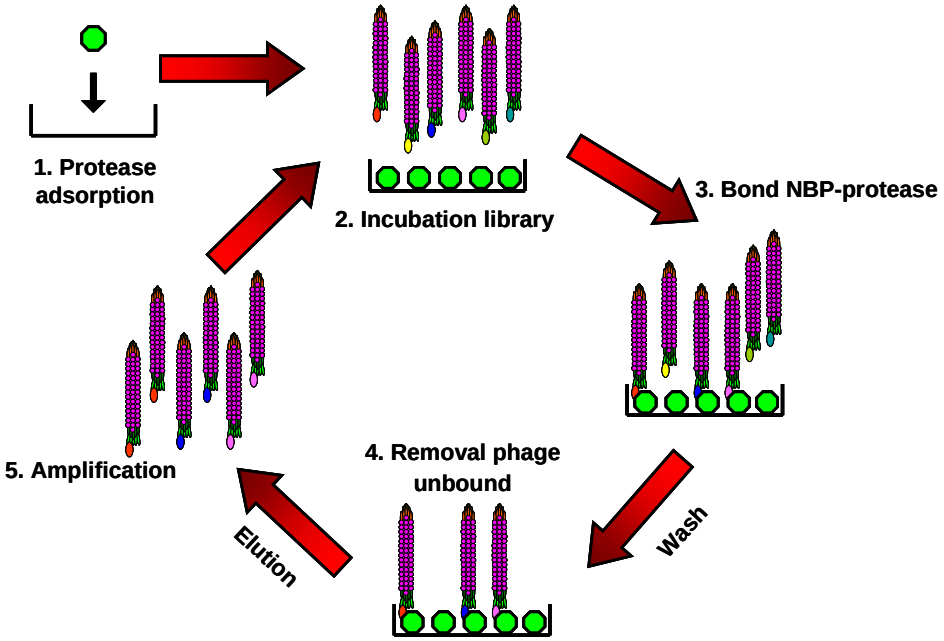


Figure 2.17 Scheme of the protease selection cycle.

As negative control helper phage M13K07 was incubated into wells with a single protease to test the possible affinity between capsid proteins and the protease. In addition, helper phage M13K07 was incubated into wells with BSA to test the possible affinity between capsid proteins and BSA.
The un-specific bonds (helper phage with proteases, helper phage and BSA, phage libraries and BSA) decreased during the second cycle, because of the increased washing.
Figure 2.18 shows the results after the second cycle of bio-panning. Already after the second cycle, the distribution of eluted phage was different for the 4 proteases: the numbers of eluted phage from papain- and pepsin- wells were higher than those for the other proteases, in particular trypsin.

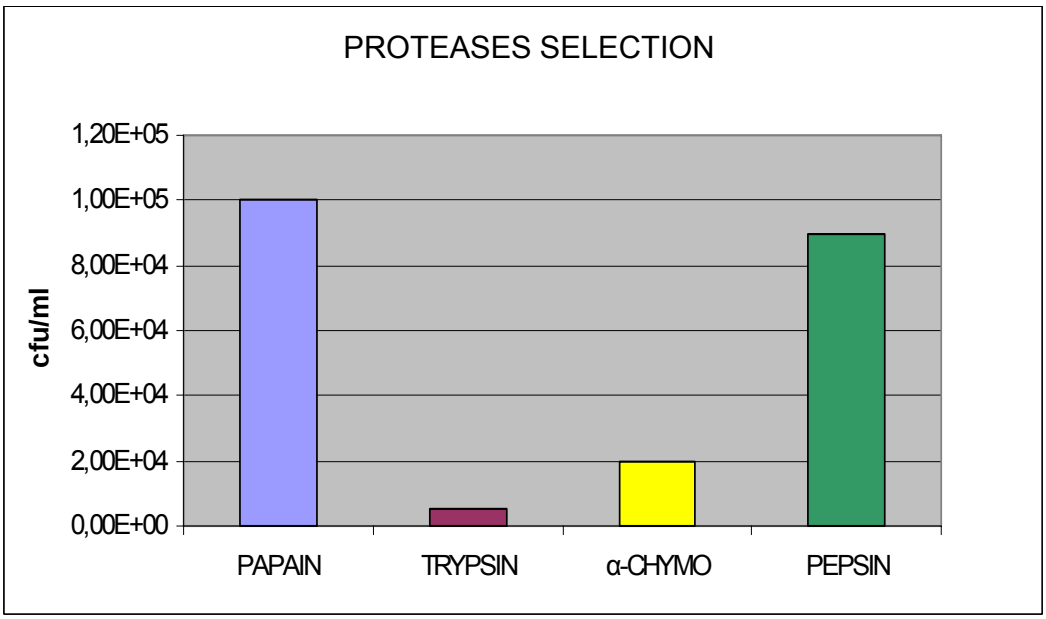


Figure 2.18 Selection of de novo library against proteases. *This graph shows the selection results after the second selection cycle for all 3 phagemidic systems.*

3. DISCUSSION

The first part of this doctoral thesis concerns the design and synthesis of 50 amino acid long random peptide libraries. To display the protein libraries we created 3 different phagemid systems, improving vectors used in previous experiments by Luisi's group.
After that, random sequences were expressed by the phage display technique in order to have a phage population with the phagemid incapsidated and a single NBP-pIII fusion protein on the surface. The resulted phage were finally employed in selection cycles to look for folded structures, potential catalytic activities and possible interactions with other proteins.

3.1 CONSTRUCTION OF PHAGEMID VECTORS

Three different phagemid vectors were created: pIII-Dummy-His-tag, pOCI1050 His-tag and pBluescript II KS His-tag.

The first two vectors derive from pOCI1050 c-myc (courtesy of Dr. Chiarabelli, University of Roma Tre). Their common characteristics are:

- the pelB signal sequence, to ensure a reliable membrane translocation;
- a Linker region;
- an affinity tag (c-myc tag in pOCI1050 c-myc and His-tag in the others);
- the C-terminal part of gene III (sequence D197-S406), encoding the capsid protein pIII;
- a TAG stop codon upstream of gIII to reduce the level of non-recombinant phage. In particular, the TAG can be suppressed expressing fusion protein in *Amber* mutation *E.Coli* (*supE44*) which translate the TAG into a glutamine (Q);
- a pLac promoter upstream of the tag which allows fusion gene expression in the presence of IPTG.

In addition, the pIII-Dummy-His-tag has the "Dummy" sequence and the Xa site. The first is an insertion sequence (738 bp long), which does not encode any proteins but allows visual control during the cloning step. The

Xa site is the cleavage site of a protease, the activated coagulation Factor X, to allow the purification of exogenous proteins without the tag.

The pBluescript II KS His-tag comes from a commercial vector, the pBluescript II KS which has a T7 polymerase promoter which is more efficient than pLac promoter. The first modification of the wild-type vector consisted in the inclusion of a fragment from the pOCI His-tag containing: pLac, pelB, His-tag and gIII. Then the vector restriction system was modified to allow library cloning and creation of the PRG site.

The 3 new vectors have the His-tag replacing the c-myc tag. This involved a change of the purification system, in fact His-tag tagged proteins can be identified by Immobilized metal affinity chromatography (IMAC), whose efficiency is higher than antibody purification previously used to recognize c-myc tagged proteins (obtained with the old phagemid system pOCI1050 c-myc tag).

The random sequence libraries were cloned into the 3 vectors and the efficiency of the cloning step was improved, resulting in a library complexity degree higher than previous libraries cloned into pOCI1050 c-myc (around 10^7 different sequences per ml).
In addition, the new phagemids were more efficient in the phage production step, described afterwards.
Also in the selection steps these vectors proved to be functional since they did not affect results, in fact we have not found any difference in terms of yield for the 3 phagemids.

3.2 RANDOM LIBRARIES

Random libraries were constructed using two oligonucleotides: forward and reverse. Both oligos were designed with a codon scheme NNK (where N can be one of all bases while K can be either G or T). In this way we obtained 31 codons encoding all the amino acids and only one stop codon, the TAG. So the possibility to have truncated proteins was reduced.
Both oligos had a cloning restriction site at the 5' end (*NotI* for the forward and *XbaI* for the reverse) and an annealing region at the 3'end with a *XhoI* restriction site.

Random libraries were cloned into the 3 vectors mentioned before by digestion in *NotI*, *XbaI* sites. The following cut with *XhoI* restriction enzyme caused the PRG site, which is necessary in the fold selection step.

As mentioned above, the complexity degree of phagemid libraries was estimated at 10^7 different sequences per ml. This complexity degree is a good compromise between a wide space of sequences and a good number of copies of each sequence. In fact, the space of sequences has to be quite wide to allow a screening of many different proteins, but a too wide space involves too few copies of each sequence and this could alter results during a selection.

From each phagemid library, we analyzed DNA sequence of 10 samples to verify their correctness and randomness. The majority of sequences, about 80%, were found to be correct with random sequence in-frame in the vector. The remaining part of the sequences consisted of either not in-frame sequences or vectors lacking in the random sequences.
Of all the correct sequences, randomness at the amino acid level was analysed by comparing experimental and expected data. We found an amino acid frequency in line with the expected data.
Randomness at a nucleic acid level was also investigated, taking into account the NNK scheme used in the design of the library. The experimental data are perfectly in line with the expected one. According the high experimental value of the cytosine in 3rd position it is due to the design of the reverse primer. The last random triplet in the reverse primer is not corresponding to the NNK scheme, because it contains part of the *XbaI* site, necessary for the cloning in the phagemid. In fact, such last codon has the sequence NTC causing in the randomness of phage peptide library at nucleic acid level table a value in percentage of 2.2 for C in 3rd position.

3.3 PHAGE PRODUCTION

As described in the introduction, there are different formats for the phage display technique. We chose the 3+3 monovalent Phage Display. In this system the phagemid electroporated into *E.Coli* can replicate as a double-stranded plasmid. A co-infection with helper phage M13K07 is necessary for the production of phage particles with the phagemid incapsidated and a single NBP-pIII fusion protein on the surface. The other pIII proteins are wild-type, in fact they are provided by the helper with all the proteins

necessary for phage assembly. In this way the phage infectivity is not compromised.
Four different types of phage come from co-infected cells, only one is the recombinant phage with exogenous peptide on the capsid and the corresponding exogenous gene in its genome.
To maximise the fraction of recombinant phage it is necessary to use a mutated helper phage. For example, helper phage M13K07 is able to replicate itself in the absence of phagemid, but, in the presence of a phagemid bearing a wild-type M13 origin, single-stranded phagemid is packaged preferentially.
In addition, the amount of recombinant phage can be affected by phage production protocol. After several tests, we obtained an increase of the recombinant phage titre when the helper co-infection occurs with a cell density equals to $OD_{600} \approx 0.8$. The same cell density is favourable for the IPTG induction, necessary because fusion gene (random sequence-gIII) is controlled by a pLac promoter.
All these improvements resulted in a recovered phage titre of around 10^{11}cfu/ml (Colony Forming Unit per millilitre) for all phage libraries.

3.4 FOLDING SELECTION BY THROMBIN DIGESTION

The resistance to thrombin digestion was used as a rudimentary folding criterion. Thrombin was chosen because it is more selective than other proteases and not affected by flanking residues. On the contrary, other enzymes are inhibited when specific residues are near the cleavage site: for example, trypsin is inhibited when a proline is near the cleavage site.
As target site of thrombin we chose the PRG (proline-arginine-glycine) site. It is in the middle of random sequence and it is created following the cut of the annealing region by *XhoI* enzyme.
The PRG site was chosen because it is the shortest cleavage site digested by thrombin. In fact, previous tests on a small protein, the APP (Avian Pancreatic Peptide), demonstrated that the PRG site did not affect protein structure, even if it was inserted in different regions of the protein [Chiarabelli, 2006].
The base of this selection is that in a folded sequence the PRG is preserved during thrombin digestion while in un-folded sequence the PRG is exposed to enzyme cleavage.
To recover His-tagged sequences un-digested by thrombin we chose the Ni-affinity purification, using ELISA Ni-coated plates. Results obtained after

the second biopanning cycle showed a high proportion of resistant sequences, indicating a good chance to select folded NBPs. Performed controls confirmed the results excluding un-specific bonds between Ni and capsid proteins.
The chance to find folded random proteins (unselected by evolutionary pressure) might suggest that folding is not a special property to be selected by Nature. This point could be taken as favouring the contingency hypothesis.
However, these results are preliminary because, first of all, they derive from two selection cycles; in addition, our folding criterion is valid only as a first, general and qualitative screening method.

3.5 SELECTION AGAINST TSA

The selection criteria to discover NBPs with possible catalytic activities is the capability to bind a transition state analogue (TSA).
This selection is supported by the transition state theory. According to this theory, during a reaction, the reagents have to go through a transition state with the highest level of energy. A molecule that binds this transition state can catalyze this reaction, lowering the activation energy. Therefore, using a TSA which mimics the transition state of an amide bond hydrolysis, selected NBPs could catalyze this reaction and also the contrary reaction.
The procedure implies the selection of the phage library on a solid surface coated with the TSA. TSA is immobilized on the plate by a covalent bond between the acid group of TSA and secondary amine on the plate.
Since coupling reaction is fundamental for the selection, we optimized an easy but efficient test to validate the occurred coupling. It is based on the instability of TSA at basic pH. In this condition, in fact, the hydrolysis of the p-nitro phenolate group takes place and can be measured by absorbance at 405 nm. In practice, TSA pre-incubated wells were treated with a basic buffer and their absorbance value compared with a blank (the basic buffer). The occurred coupling was indicated by a greater absorbance value of samples taken from TSA pre-incubated wells.
The selection was made at pH7 since TSA is unstable at greater pH. In addition, we tested the influence of a metallic cofactor (Zinc) on affinity between NBPs and the target.
An important amount of phage library was recovered from wells coated with TSA. The metal cofactor does not seem to affect the affinity for the

target, in fact the phage quantity recovered from the plate is similar both with and without Zn.
Performed controls excluded un-specific bonds between capsid proteins and TSA and between NBPs and coupling reagents (sulfo-NH and EDC). Un-specific bonds with plate were excluded too.
Also in this selection, the promising results are preliminary because they derive from two biopanning cycles.
Since the TSA is an analogue of the transition state, NBP can bind it, but cannot hydrolyze it. Therefore, it is not sure if a NBP which bind the TSA can really catalyse the corresponding reaction. Of course, the selected NBPs are good potential candidates.

3.6 SELECTION AGAINST PROTEASES

The last part of my doctoral work explores the possible interactions between NBPs and other proteins, in particular proteases. The aim is to find potential new inhibitors/activators of target-proteases.
The basic idea of this selection is that a ligand, to act as an inhibitor/activator of a protein, first of all, has to bind this protein. In theory, peptides are able to bind a protein anywhere on its solvent-exposed surface. However, most peptides bind at sites coinciding with natural ligand-binding sites. Consequently, these peptides could act as inhibitors/activators like the natural ligands.
Our target proteases were: papain, pepsin, trypsin and α-chymotrypsin. They were immobilized on a polystyrene plate by adsorption. To avoid possible binds between phage and plate surface, wells were incubated also with BSA.
Already after the second cycle, the distribution of selected phage is different for the 4 proteases: the number of phage from papain- and pepsin- wells is higher than those for the other proteases, in particular trypsin.
Performed controls excluded un-specific bonds between capsid proteins and each protease. Un-specific bond with BSA were excluded too.
As for the previous selections, obtained results are preliminary, so in this condition it is very difficult to speculate about the preference of NBPs for a protease and not for another one.

4. CONCLUSIONS AND OUTLOOK

The Never Born Protein project is based on idea that the fraction of proteins existing in nature is only a minimal part of all possible theoretical amino acid sequences. This observation induces important questions: do extant proteins possess any special chemico-physical properties (such as solubility, fold, functionality, thermodynamic stability) that made their selection inevitable? Or rather are they the result of contingency? If they are the result of contingency, a universe of unknown proteins could exist.
To approach these questions, completely *de novo* random libraries encoding 50 amino acid long peptides were designed. To express these libraries, 3 different phagemid vectors were created. With the new phagemids the efficiency of the cloning step was improved, resulting in a library complexity degree of 10^7 different sequences per ml, which is higher than previous experiments. In the new vectors an His-tag was inserted, this involves a change of the purification system: His-tagged proteins can be identified by Immobilized metal affinity chromatography (IMAC), whose efficiency is higher than previous purification methods.
The next step was the production of the phage libraries by phage display technique. The improvement of previous protocols resulted in an important production of recombinant phage with the phagemid incapsidated and a single NBP-pIII fusion protein on the surface.
The recombinant phage were used to select NBPs for several characteristics. The first selection was to look for sequences with a three-dimensional structure. It resulted in a good chance of having folded structures in the random sequence libraries.
Results obtained from the selection against a Transition State Analogue (TSA) showed that it is possible to have *de novo* proteins with a potential catalytic activity. Similar conclusions can be drawn from the selection against proteases: *de novo* proteins could act as inhibitors/activators of bound proteases.
The possibility of finding folded random proteins might suggest that folding is not a special property to be selected by Nature. This point could be taken as favouring the contingency hypothesis. Also the probability to find new sequences with a potential activity supports the idea that there may be an entire universe of possible proteins, with unknown properties.
Of course, the reported data are preliminary, coming from two cycles of selections, however they are an encouraging starting points.

Further selection cycles with more limiting screening conditions are necessary to raise the specificity of selected sequences. For example the more limiting conditions could be the time and the temperature of the incubation steps, in fact higher temperatures are more restrictive than low one increasing the reaction kinetic; in addition, the washing can be increased to reduce un-specific interactions.
Finally, the selected sequences could be synthesized with reasonable amounts to permit the study of their structural and kinetic properties.

5. REFERENCES

Aguinaldo AM, Arnold FH. Staggered extension process (StEP) *in vitro* recombination. Methods Mol Biol 2003, 231, 105-110

Barbas CF 3rd. Recent advances in phage display. Curr Opin Biotechnol 1993, 4, 526–530

Beck E, Zink B. Nucleotide sequence and genome organization of filamentous bacteriophages fl and fd. Gene 1981, 16, 35-58

Better M, Chang CP, Robinson RR, Horwitz AH. *Escherichia coli* secretion of an active chimeric antibody fragment. Science 1988, 240, 1041–1043

Bradley LH, Thumfort PP, Hecht MH. De novo proteins from binary-patterned combinatorial libraries. Meth Mol Biol 2006, 340, 53-70

Branden C, Tooze J. Introduction to protein structure second edition. Garland Publishing, New York 1998

Burton DR. Phage display. Immunotechnology 1995, 1, 87–94

Cadwell RC, Joyce GF. Mutagenic PCR. PCR Methods Appl 1994, 3, S136–40

Cech TR. In The RNA World (eds Gesteland RF & Atkins JF). Cold Spring Harbor Laboratory Press, Plainview, NY 1993**,** 239-269

Cesareni G. Peptide display on filamentous phage capsids. A new powerful tool to study protein-ligand interaction. FEBS Lett 1992, 307, 66–70

Chiarabelli C, Vrijbloed JW, Thomas RM, Luisi PL. Investigation of de novo totally random biosequences, Part I: A general method for in vitro selection of folded domains from a random polypeptide library displayed on phage. Chem Biodivers 2006, 3(8), 827-839

Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. Methods Mol Biol 2003, 231, 3-9

Coco WM. RACHITT: gene family shuffling by random chimeragesesis on transient templates. Methods Mol Biol 2003, 231, 111-127

Cordes MHJ, Davidson AR, Sauer RT. Sequence space, folding and protein design. Curr Opin Struct Biol 1996, 6, 3-10

Davidson AR, Lumb KJ, Sauer RT. Cooperatively folded proteins in random sequence libraries. Nat Struct Biol 1995, 2, 856-864

Davidson AR, Sauer RT. Folded proteins occur frequently in libraries of random amino acid sequences. Proc Natl Acad Sci USA 1994, 91, 2146-2150

Deperthes D. Phage display substrate: a blind method for determining protease specificity. Biol Chem 2002, 383, 1107–1112

Doorbar J, Winter G. Isolation of a peptide antagonist to the thrombin receptor using phage display. J Mol Biol 1994, 244, 361-369

Dumon C, Varvak A, Wall MA, Flint JE, Lewis RJ, Lakey JH, Morland C, Luginbühl P, Healey S, Todaro T, De Santis G, Sun M, Parra-Gessert L, Tan X, Weiner DP, Gilbert HJ. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. J Biol Chem 2008, 283, 22557-22564

Duve CD. Life Evolving: Molecules, Mind, and Meaning by Christian de Duve. Oxford University Press 2002

Duve CD. Vital dust: life as a cosmic imperative. Ed. Basic Book, NY 1995

Endemann H, Model P. Location of filamentous phage minor coat proteins in phage and in infected bacteria. J Mol Biol 1995, 250, 496-506

Eschenmoser A, Kisakurek M. Chemistry and the origin of life. Helv. Chim. Acta 1996, 79, 1249-1259

Evan GI, Lewis GK, Ramsay G, Bishop JM. Isolation of monoclonal-antibodies specific for human c-myc proto-oncogene product. Mol Cell Biol 1985, 5, 3610-3616

Fernandez-Gacio A, Uguen M, Fastrez J. Phage display as a tool for the directed evolution of enzymes. Trends Biotechnol 2003, 21, 408–414

Ferris JP. From building blocks to the polymers of life (In: Life's Origin, The Beginning of Biological Evolution). JW Schopf Ed, California University Press, CA 2002

Forrer P, Jung S, Plückthun A. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. Curr Opin Struct Biol 1999, 9, 514-520

Fox SW, Dose K. Molecular Evolution and the Origin of Life. Marcel Dekker Publisher, New York and Basel 1977

Frost LS. Conjugative pili and pilus-specific phages. In *Bacterial Conjugation.* D. B. Clewell Ed., Plenum NY 1993

Glasstone S, Laidler K, Eyring H. The Theory of Rate Processes. McGraw-Hill, New York 1941

Goldstein RA, Luthey-Schulten ZA, Wolynes PG. Protein tertiary structure recognition using optimized Hamiltonians with local interactions. Proc Natl Acad Sci USA 1992, 89, 9029-9033

Gorlero M, Wieczorek R, Adamala K, Giorgi A, Schininà ME, Stano P, Luisi PL. Ser-His catalyses the formation of peptides and PNAs. FEBS Lett 2008, 583, 153-156

Gould SJ. Wonderful Life: the burgess shale and the nature of history. Penguin, London 1989

Griffiths AD. Production of human antibodies using bacteriophage. Curr Opin Immunol 1993, 5, 263–267

Gutin AM, Abkevich VI, Shakhnovich EI. Evolution-like selection of fast-folding model proteins. Proc Natl Acad Sci USA 1995, 92, 1282-1286

Gutin AM, Shakhnovich EI. Ground-state of random copolymers and the discrete random energy-model. J Chem Phys 1993, 98, 8174-8177

Haldane J. The origin of life. Rationalist Annual 1929, 148, 3-10

Haldane JBS. The origin of life. New Biology 1954, 16, 12-27

Hawinkels LJ, van Rossenberg SM, de Jonge-Muller ES, Molenaar TJ, Appeldoorn CC, van Berkel TJ, Sier CF, Biessen EA. Efficient degradation-aided selection of protease inhibitors by phage display. Biochem Biophys Res Commun 2007, 364, 549–555

Hecht MH, Das A, Go A, Bradley LH, Wei Y. De novo proteins from designed combinatorial libraries. Protein Science 2004, 13, 1711-1723

Hertveldt K, Belien T, Volckaert G. General M13 phage display: M13 phage display in identification and characterization of protein-protein interactions. Methods Mol Biol 2009, 502, 321–339

Hilhorst HJ, Deutch JM. Analysis of Monte Carlo results on the kinetics of lattice polymer chains with excluded volume. J Chem Phys 1975, 63, 5153-5161

Hill DF, Patersen GB. Nucleotide sequence of bacteriophage fl DNA. J Virol 1982, 44, 32-46

Ho M, Nagata S, Pastan I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. Proc Natl Acad Sci USA 2006, 103, 9637-9642

Hochuli E, Döbeli H, Schacher A. New metal chelate adsorbent selective for proteins and peptide containing neighbouring histidine residues. J Chromatogr 1987, 411, 177–184

Hoess RH. Protein design and phage display. Chem Rev 2001, 101, 3205-3218

Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC. Antibody phage display technology and its applications. Immunotechnology 1998, 4, 1–20

Hyde-DeRuyscher R, Paige LA, Christensen DJ, Hyde-DeRuyscher N, Lim A, Fredericks ZL, Kranz J, Gallant P, Zhang J, Rocklage SM, Fowlkes DM, Wendler PA, Hamilton PT. Detection of small-molecule enzyme inhibitors with peptides isolated from phage-displayed combinatorial peptide libraries. Chem Biol 2000, 7, 17–25

Jacob F. The possible and the actual. University of Washington Press 1982

Jencks, W. Catalysis in chemistry and enzymology. McGraw-Hill, New York 1969

Joern JM. DNA shuffling. Methods Mol Biol 2003, 231, 85-89

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. Protein design by binary patterning of polar and nonpolar amino acids. Science 1993, 262, 1680-1685

Kauffman S. At Home in the Universe: The Search for the Laws of Self-Organization and Complexity. Oxford University Press 1996

Kay BK, Hamilton PT. Identification of enzyme inhibitors from phage-displayed combinatorial peptide libraries. Comb Chem High Throughput Screen 2001, 4, 535–543

Kotz JD, Bong CJ, Cochran AG. Phage-display as a tool for quantifying protein stability determinats. Eur J Biochem 2004, 271, 1623-1629

Ladner RC. Constrained peptides as binding entities. Trends Biotechnol 1995, 13, 426–430

Ladner RC. Phage display and pharmacogenomics. Pharmacogenomics 2000, 1, 199–202

Leman L, Orgel L, Ghadiri MR. Carbonyl sulfide-mediated prebiotic formation of peptides. Science 2004, 306, 283-286

Levinthal C. Are there pathways for protein folding. J Chim Phys 1968, 65, 44-45

Lim CJ, Haller B, Fuchs JA. Thioredoxin is the bacterial protein encoded by fip that is required for filamentous bacteriophage fl assembly. J Bacteriol 1985, 161, 799-802

Lingen B, Grötzinger J, Kolter D, Kula MR, Pohl M. Improving the carboligase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis. Protein Engineering 2002, 7, 585-593

Lopez J, Webster RE. Assembly site of bacteriophage fl corresponds to adhesion zones between the inner and outer membranes of the host cell. J Bacteriol 1985, 163, 1270-1274

Luisi PL. Autopoiesis: a review and a reappraisal. Naturwissenschaften 2003, 90, 49-59

Luisi PL. The Emergence of Life. From Chemical Orgins to Synthetic Biology. Cambridge University Press, London 2006

Lutz S, Ostermeier M. Preparation of SCRATCHY hybrid protein libraries: size- and in-frame selection of nucleic acid sequences. Methods Mol Biol 2003, 231, 143-151

Mead DA, Kemper B. Chimeric single-stranded DNA phage-plasmid cloning vectors. Biotechnology 1988, 10, 85–102

Meyer A, Schmid A, Held M, Westphal AH, Röthlisberger M, Kohler HPE, van Berkel WJH, Witholt B. Changing the substrate reactivity of 2-hydroxybiphenyl 3-monooxygenase from *Pseudomonas azelaica* HBP1 by directed evolution. J Biol Chem 2002, 7, 5575-5582

Miller SL. Production of amino acids under possible primitive Earth conditions. Science 1953, 117, 2351-2361

Model P, Russel M. Filamentous bacteriophage. In *The Bacteriophages.* R. Calendar Ed., Plenum NY 1988

Moffet DA, Hecht MH. De novo proteins from combinatorial libraries. Chem Rev 2001, 101, 3191-3203

Monod J. Chance and Necessity: an essay on the natural philosophy of modern biology. A.A. Knopf, New York 1971

Morowitz H. Origin of cellular life: Metabolism recapitulates biogenesis. Yale University Press 1993

Murakami H, Hohsaka T, Sisido M. Random insertion and deletion mutagenesis. Methods Mol Biol 2003, 231, 53-64

Murakami H, Hohsaka T, Sisido M. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat Biotechnol 2002, 20, 76-81

Nakashima T, Fox SW. Synthesis of peptides from amino acids and ATP with lysine-rich proteinoid. J Mol Evol 1980, 15, 161-168

Neri D, Momo M, Prospero T, Winter G. High-affinity antigen binding by chelating recombinant antibodies (CRAbs). J Mol Biol 1995, 246, 367–373 [I]

Neri D, Petrul H, Roncucci G. Engineering recombinant antibodies for immunotherapy. Cell Biophys 1995, 27, 47–61 [II]

Neylon C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res 2004, 4, 1448-1459

O'Maille PE, Bakhtina M, Tsai MD. Structure-based combinatorial protein engineering (SCOPE). J Mol Biol 2002, 321, 677-691

Oparin AI. Life: Its Nature, Origin and Development. Oliver and Boyd, London, 1961
Oparin AI. Problem of the origin of life according to modern biological concepts. Cas Lek Cesk 1954, 93, 889-894 [II]

Oparin AI. Problem of the origin of life according to the progress of natural science. Izv Akad Nauk SSSR Biol 1954, 2, 5-13 [I]

Plankensteiner K, Righi A, Rode BM. Glycine and diglycine as possible catalytic factors in the prebiotic evolution of peptides. Orig Life Evol Biosph 2002, 32, 225-236

Rapoza MP, Webster RE. The products of gene I and the overlapping in-frame gene XI are required for filamentous phage assembly. J Mol Biol 1995, 248, 627-638

Russel M, Model P. Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. J Virol 1989, 63, 3284-3295

Russel M. Filamentous phage assembly. Mol Microbiol 1991, 5, 1607–1613

Sedlacek R, Chen E. Screening for protease substrate by polyvalent phage display. Comb Chem High Throughput Screen 2005, 8, 197–203

Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature, 2007, 448, 828-831

Segre D, Lancet D. Composing life. EMBO Rep 2000, 1, 217-222

Shakhnovich EI, Gutin AM. Engineering of stable and fast-folding sequences of model proteins. Proc Natl Acad Sci USA 1993, 90, 7195-7199

Shao Z, Zhao H, Giver L, Arnold FH. Random-priming *in vitro* recombination: an effective tool for directed evolution. Nucleic Acid Res 1998, 2, 681-683

Shen C, Lazcano A, Oró J. The enhancement activities of histidyl-histidine in someprebiotic reactions. J Mol Evol 1990, 31, 445-452 [II]

Shen C, Mills T, Oró J. Prebiotic synthesis of histidyl-histidine. J Mol Evol 1990, 31, 175-179 [I]

Sidhu SS, Koide S. Phage display for engineering and analyzing protein interaction interfaces. Curr Opin Struct Biol 2007, 17, 481–487

Sidhu SS. Phage display in pharmaceutical biotechnology. Curr Opin Biotechnol 2000, 11, 610-616

Siegel DL. Research and clinical applications of antibody phage display in transfusion medicine. Transfus Med Rev 2001, 15, 35–52

Skerra A, Plückthun A. Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. Science 1988, 240, 1038–1041

Smith GP, Petrenko VA. Phage display. Chem Rev 1997, 97, 391–410

Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 1985, 228, 1315–1317

Smith GP. Preface. Surface display and peptide libraries. Gene 1993, 128, 1–2

Taillades J, Cottet H, Garrel L, Beuzelin I, Boiteau L, Choukroun H, Commeyras A. *N*-Carbamoyl Amino Acid Solid–Gas Nitrosation by NO/NOx: A New Route to Oligopeptides via α-Amino Acid *N*-Carboxyanhydride. Prebiotic Implications. J Mol Evol 1999, 48, 638-645

Tanaka F, Barbas CF. Reactive immunization: a unique approach to catalytic antibodies. J Immunol Methods 2002, 269, 67-79

Van Wezenbeek PMGF, Hulsebos TJM, Schoenmakers JGG. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: Comparison with phage fd. Gene 1980, 11, 129-148

Verdier PH. Monte Carlo studies of lattice-model polymer chains. III. Relaxation of Rouse coordinates. J Chem Phys 1973, 59, 6119-6126

Wachtershauser G. Before Enzyme and templates: theory of surface metabolism. Microbiol Rev 1988, 52, 452 -484

Wei Y, Kim S, Fela D, Baum J, Hecht MH. Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci USA 2003, 100, 13270-13273 [II]

Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH. Stably folded de novo proteins from a designed combinatorial library. Protein Science 2003, 12, 92-102 [I]

West MW, Wang W, Patterson J, Mancias JD, Beasley JR, Hecht MH. De novo amyloid proteins from designed combinatorial libraries. Proc Natl Acad Sci USA 1999, 96, 11211-11216

Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR. Making antibodies by phage display technology. Annu Rev Immunol 1994, 12, 433–455

Yanagawa H, Kojima K, Ito M, Handa N. Synthesis of polypeptides by microwave heating I. Formation of polypeptides during repeated hydration-dehydration cycles and their characterization. J Mol Evol 1990, 31, 180-186

MATERIALS AND METHODS

1. Primers

Primers	Sequence
R-AMPLI-HIS-NOT	TTTTCCTTTTGCGGCCGCCCCGCTGTGATGATGATGATGATGGCT
T7-PROMOTER	TTAATACGACTCACTATAGGG
F-LIBRARY	GGTGGGCGGCCGCGNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKCCTCGAGGATAACTCGAG
LIBR-REVERS	CCACCTCTAGANMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNACCTCGAGTTATCCTCGAG
DUMMY-FW1	CTGAATATCGAAGGCCGCGGGGCGGCCGCATAAATAAATAAATGGCCTTGGTTGACGGT
DUMMY-FseI-RV	CGGGATCCGGCCGGCCTTTCTAGATTGAGAGTAATAAGACCCAAATT
HisXa-DUMMY-FW2	AGAGAGCATGCCATGGGCTCTCATCATCATCATCATCATTCTGGCCTGAATATCGAAGGCCGC

2. Plasmids

Plasmids	Reference
pET14b	Novagen
pOCI c-myc	Courtesy of Dr. Chiarabelli, University of RomaTre - Italy
pOCI1050 His-tag	This work
pBluescript II KS	Stratagene
PAO1	Courtesy of Dr. Leoni, University of Roma Tre

3. Cell strains

Cell strain	Genotype	Reference
XL1Blue MRF'	Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene
XL1Blue	supE44 hsdR17(rK-, mK+) recA1 endA1 gyrA96 thi-1 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene

4. Phage

Phage	Description	Reference
M13K07	Helper phage	Amersham

5. Buffers and Media

BR Buffer: 0.04 M H_3BO_3 0.04 M H_3PO_4 0.04 M CH_3COOH
Cova Buffer: 116,9 g NaCl 10 g $MgSO_4 \cdot 7H_2O$ 500 µl Tween 20 Adjust volume to 1 L with PBS 1x
DIGESTION BUFFER: NaH_2PO_4 50mM, pH 7.4
EDC working solution: 5.8 mg EDC Adjust volume to 10 ml with H_2O
ELUTION BUFFER: NaH_2PO_4 50mM NaCl 300mM Imidazole 125mM pH 7.4

NHSS working solution: 1.83 mg Sulfo-NH Adjust volume to 10 ml with H_2O
PBS: (solution 10x) 80 g NaCl 2 g KCl 14,14 g Na_2HPO_4 2,4 g KH_2PO_4 Add d.d. water up to 1L. pH 7.4
TE Buffer: Tris-HCl 1 mM pH8 + EDTA 0.5mM ph8
TSA stock solution (23mM): 7 mg TSA 400 µl H_2O 600 µl DMSO
TSA/NH, working solution: 1.83 mg Sulfo-NH 500 µl TSA solution stock Adjust volume to 10 ml with H_2O

6. Chemicals

Chemical	**Reference**
EcoRI	NEB
HindIII	NEB
NcoI – Fast Digest	Fermentas
NotI – Fast Digest	Fermentas
SacI	NEB
SalI	Fermentas
SmaI	NEB
XbaI - FastDigest	Fermentas
XhoI	Fermentas
XhoI	NEB
Alkaline phosphatase	Roche
Ampicillin	Sigma Aldrich
BSA	Sigma Aldrich

CH_3COOH	Sigma Aldrich
DMSO	Sigma Aldrich
DNA Polymerase I - Klenow	NEB
dNTPs	NEB
EDTA	Fluka
Glucose	Sigma Aldrich
H_3BO_3	Sigma Aldrich
H_3PO_4	Sigma Aldrich
Imidazole	Fluka
IPTG	Fluka
Kanamycin	Sigma Aldrich
KCl	Sigma Aldrich
KH_2PO_4	Sigma Aldrich
LB AGAR	Sigma Aldrich
LB BROTH	Sigma Aldrich
$MgSO_4 \cdot 7H_2O$	Fluka
N-[3-Dimethylaminopropyl]-N′-ethylcarbodiimide hydrochloride (EDC)	Fluka
Na_2HPO_4	Sigma Aldrich
NaCl	Sigma Aldrich
N-hydroxysulfosuccinimide sodium salt (sulfo-NH)	Fluka
Papain	Sigma Aldrich
PEG 6000	Sigma Aldrich
Pepsin	Sigma Aldrich
RedTaq - Taq DNA Polymerase	Sigma Aldrich
T4 DNA Ligase	NEB
Tetracycline	Sigma Aldrich
Thrombin	Sigma Aldrich
Tris-HCl	Sigma Aldrich
Trypsin	Sigma Aldrich
TSA	ProtEra
TWEEN 20	Sigma Aldrich
ZnCl	Sigma Aldrich
α-chymotrypsin	Sigma Aldrich

7. Kits and ELISA plates

Kit	**Reference**
Wizard SV Gel and PCR Clean-Up System	Promega
ELISA plates	
CovaLink™ NH MicroWells™	NUNC
HIS-Select® High Capacity (HC) Nickel Coated Plate	Sigma Aldrich
Maxisorp Immuno-plate	NUNC

8. Construction of phagemid pIII-DUMMY His-tag

Dummy fragment comes from the LasR gene, into plasmid PAO1 (courtesy of Dr. Leoni, University of Roma Tre). This last was amplified using primers DUMMY-FW1 and DUMMY-FseI-RV (see PCR1 conditions below). The FW1 primer introduces a NotI restriction site and a Xa site codifing DNA upstream the LasR gene. The RV primer introduces a XbaI restriction site.
The resulting DNA fragment was amplified with primers HisXa-DUMMY-FW2 and DUMMY-FseI-RV (see PCR2 conditions below). The FW2 primer introduces a hexa-histidine tag and a NcoI restriction site at the 5' edge. This last fragment and pOCI c-myc were digested with NcoI and XbaI. The products of both digestions were purified with kit Wizard SV Gel and PCR Clean-Up System and ligated together.

PCR 1 conditions: (V_f= 100 µl)

DNA PAO1 plasmid	1µl
primer DUMMY-FW1 (16 pmol/µl)	4µl
primer DUMMY-FseI-RV (16 pmol/µl)	4µl
dNTPs 10mM	2µl
Buffer Red Taq 10X	10µl
Red Taq	4µl
H_20	75µl

PCR 2 conditions: (V_f= 100 µl)

DNA PCR1	1µl
primer HisXa-DUMMY-FW2 (16 pmol/µl)	4µl
primer DUMMY-FseI-RV (16 pmol/µl)	4µl
dNTPs 10mM	2µl
Buffer Red Taq 10X	10µl
Red Taq	4µl
H_20	75µl

Thermocycle profile for both PCR:

N° Step	**Temperature**	**Time**
1	95°C	5’
2	95°C	30’’
3	55°C	30’’
4	75°C	1’
5	Repeat steps 2-3-4 for 24 times	
6	74°C	7’
7	4°C	For ever

Restriction enzymes digestion conditions:

- for fragment:

PCR 2 product	70µl
NcoI	1µl
XbaI	1µl
Buffer Fast-Digest 10X	8µl

- for pOCI1050 c-myc:

plasmid	30µl
NcoI	0.5µl
XbaI	0.5µl
Buffer Fast-Digest 10X	4µl

37°C, 1h

Ligation conditions:

pOCI1050 c-myc	20µl
DUMMY fragment	30µl
T4 DNA Ligase Buffer 10x	5µl
T4 DNA Ligase	2µl

16°C, over night

9. Construction of phagemid pOCI1050 His-tag

The hexa-histidine tag derived from a pET14b plasmid and primers R-AMPLI-HIS-NOT and T7-PROMOTER. The resulting DNA fragment was 134 bp long and digested with NcoI and NotI. The pOCI c-myc was digested with the same enzymes to remove the c-myc tag. The products of both digestions were purified with kit Wizard SV Gel and PCR Clean-Up System and ligated together. The resulting plasmid was 3924 bp and presented a His-tag upstream cloning site.

PCR conditions: (V_f= 100 µl)

DNA pEt14b	1µl
primer T7 promoter (16 pmol/µl)	4µl
primer R-AMPLI-HIS-NOT (16 pmol/µl)	4µl
dNTPs 10mM	2µl
Buffer Red Taq 10X	10µl
Red Taq	4µl
H20	75µl

Thermocycle profile:

N° Step	**Temperature**	**Time**
1	95°C	5’
2	95°C	30’’
3	55°C	30’’
4	75°C	1’
5	Repeat steps 2-3-4 for 24 times	
6	74°C	7’
7	4°C	For ever

Restriction enzymes digestion conditions:

- for His-tag fragment:

PCR product	70µl
NcoI	1µl
NotI	1µl
Buffer Fast-Digest 10X	8µl

- for pOCI1050 c-myc:

plasmid	30µl
NcoI	0.5µl
NotI	0.5µl
Buffer Fast-Digest 10X	4µl

37°C, 1h

Ligation conditions:

pOCI1050 c-myc	20µl
His-tag fragment	30µl
T4 DNA Ligase Buffer 10x	5µl
T4 DNA Ligase	2µl

16°C, over night

10. Construction of phagemid pBluescript II KS His-tag

a) His-tag Insertion:

The pOCI1050 His-tag was digested with EcoRI and HindIII to obtain a DNA fragment containing His-tag, Linker and gene III (sequence D197-S406). Also the pBluescript II KS wild-type was digested with the same enzymes. The products of both digestions were purified with kit Wizard SV Gel and PCR Clean-Up System and ligated together.

Restriction enzymes digestion conditions:

- for His-tag fragment:

pOCI1050 His-tag	70µl
EcoRI	1µl
HindIII	1µl
Buffer NEB 2 10X	8µl

- for pBluescript II KS:

plasmid	30µl
EcoRI	0.5µl
HindIII	0.5µl
Buffer NEB 2 10X	4µl

37°C, 1h

Ligation conditions: same of construction of phagemid pOCI1050 His-tag

b) Changing restriction system:

To eliminate plasmid XhoI site, pBluescript II KS His-tag was digested with SalI and XhoI. The digested plasmid was purified with Wizard SV Gel and PCR Clean-Up System, finally reclosed through ligation.

Restriction enzymes digestion conditions:

pBluescript II KS His-tag	40µl
Buffer O^+ 10x	5µl
SalI	1µl
XhoI	1µl

37°C, 1h

Ligation conditions:

Plasmid	20µl
T4 DNA Ligase Buffer 10x	2,5µl
T4 DNA Ligase	2µl

16°C, over night

To eliminate plasmid XbaI site, pBluescript II KS His-tag was digested with SmaI and SacI. Into the digestion mix was present the Klenow (DNA Polimerase I Large Fragment) that, with its 5'→3' exonuclease activity, removed bare nucleotides in the SacI cut site. After this, the digested plasmid was purified with Wizard SV Gel and PCR Clean-Up System, finally reclosed by ligation.

Restriction enzymes digestion conditions:

pBluescript II KS His-tag	40µl
Buffer NEB 4 10x	5µl
SacI	2µl
SmaI	2µl
Klenow	1µl

37°C, 1h

Ligation conditions:

Plasmid	20µl
T4 DNA Ligase Buffer 10x	2,5µl
T4 DNA Ligase	2µl

16°C, over night

11. Preparation of duplex DNA library from oligonucleotides

For the preparation of duplex DNA-LIBRARY, two oligonucleotides (f-library and libr-revers) were mixed, heated up 95 °C for 10 min, and annealed by slowly cooling the mixture to 4°C for 5 minutes. The single stranded regions were converted to duplex DNA by continuing the incubation at 72°C for 1 hour in the presence of dNTPs and Red Taq Polimerase. Following the reaction, double stranded DNA (184bp long) was recovered after purification using Wizard SV Gel and PCR Clean-Up System.

Reaction mix: (V_f= 100 µl)

primer F-LIBRARY (16pmol/µl)	25µl
primer LIBR-REVERS (16pmoli/µl)	25µl
dNTPs 10mM	6µl
Red Taq Buffer 10x	30µl
Red Taq Polimerasi	12µl
H2O	4µl

12. Construction of the phagemid library

The resulting duplex DNA library fragments and three phagemid vector (pIII-Dummy-His-tag, pOCI1050 His-tag, pBluescript II KS His-tag) were digested separately with NotI and XbaI enzymes. The products of both digestions were purified with kit Wizard SV Gel and PCR Clean-Up System and ligated together.

Restriction enzymes digestion conditions:

- for library fragment:

fragment	50µl
NotI	1.5µl
XbaI	1.5µl
Fast-Digest Buffer 10X	5µl

- for phagemide vectors:

plasmid	100µl
NotI	2µl
XbaI	2µl
Fast-Digest Buffer 10X	10µl
Alkaline phosphatase	2µl

37°C, 2h

Ligation conditions:

purified and concentrated plasmid	total
fragment	50µl
T4 DNA Ligase Buffer 10x	6µl
T4 DNA Ligase	4µl

16°C, over night

13. Removing of annealing region and creation of the PRG site

The phagemid libraries were digested with XhoI to remove the annealing region. The product of digestion was purified with kit Wizard SV Gel and PCR Clean-Up System and ligated. In this way the PRG site was create in the middle of random sequence.

Restriction enzymes digestion conditions:

library	100µl
XhoI (NEB)	5µl
Buffer 2 10x (NEB)	10µl

37°C, 1h

Ligation conditions:

library	100µl
T4 DNA Ligase Buffer 10x	11µl
T4 DNA Ligase	8µl

16°C, over night

14. Phage production and precipitation

Stocks of XL1Blue MRF' containing the plasmid library were used to inoculate LB broth containing 1 % glucose, tetracycline (100 mg/ml) and ampicillin (100 mg/ml). The culture was incubated and shaken at 37 °C over night.

To remove glucose, a little volume of culture was pelleted by centrifugation and washed twice with LB, then the pellet was resuspended in LB, tetracycline (100 mg/ml) and ampicillin (100 mg/ml). The culture was incubated and shaken at 37 °C up to OD_{600} about 0.8, at this point M13K07 helper phage was added. After 15 min. at 37 °C with no shaking, the incubation was continued for 1 h at 37 °C shaking. The culture was diluted 20-fold in fresh LB containing tetracycline (100 mg/ml), ampicillin (100 mg/ ml), kanamycin (20 mg/ml) and 0.01 mM IPTG. The culture was incubated and shaken at 37 °C over night.

Phagemid particles from liquid cultures were obtained by centrifugation (20 min at 12000g). The phage were precipitated from the supernatant by addition of 0.1 volume NaCl (5M) and 0.1 volume PEG 6000 (40%). The phage were harvested by centrifugation (20 min at 12000g), the phage pellet was resuspended in 8 ml TE and reprecipitated as described above. Finally, the pellet was resuspended in PBS.

15. Thrombin digestion and biopanning

The phage libraries, resuspended in PBS buffer, were added into each well of HIS-Select® High Capacity (HC) Nickel Coated Plate (V_f/well = 100µl; C_f phage/well =10^8 cfu/ml) incubated 1 h at room temperature. The wells were washed 10 times with PBS to eliminate un-bound phage. Thrombin, resuspended in Buffer Digestion at a concentration of 0.1 u/µl, was added into the well and incubated 1 h at 37°C. The supernatant was thrown away. The wells were washed 10 times with PBS + TWEEN 20 and then 10 times with PBS. After this, bound phage were eluted with 100µl Elution Buffer. At this point 10-fold dilution series of eluted phage were made. These dilutions were used to infect E. coli XL1-blue cells. The cells were pre-grown on LB containing 1% glucose up to OD_{600} 0.8. After 15 min incubation at 37 °C without shaking, the culture was incubated at 37 °C shaking 1 h. Then aliquots of the bacterial solutions were plated on LB agar containing 1 % glucose, tetracycline (100 mg/ml) and ampicillin (100 mg/ml). Colonies appeared on the plates the following day were counted to calculate the total number of phage that bounded the Ni^{2+} into the wells.

Negative control: M13K07 helper phage was incubated like phage library with and without thrombin and the eluted material was titrated as described before for the library but using agar plates were with kanamycin.

Positive control: phage library was incubated without thrombin and the eluted material was titrated as described before.

Second cycle of selection: The eluted phage in the first cycle were used to infect E. coli XL1-blue cells, pre-grown on LB containing 1 % glucose up to an OD_{600} of 0.8. Phage production and precipitation was the same described before. Screening conditions were the same of the first selection.

16. TSA Coupling and NH-wells

a) Coupling TSA:
Each well in the CovaLink™ NH MicroWells™ plate was incubated for 90', room temperature, with:

H_2O	50µl
TSA/NH working solution	100µl
EDC working solution	50µl

Then, wells were washed 3 times with 200µl of Cova Buffer but, before the last wash, the plate was incubated with the buffer for 15', room temperature, shaking.

b) NH-well:
More wells were incubated 90' at room temperature, with:

H_2O	50µl
NHSS working solution	100µl
EDC working solution	50µl

Then, wells were washed as before.

17. pH Check

90µl of BR (Britton-Robinson) Buffer at different pH from 6 to 10 was added to 10µl of TSA stock solution. Abs 405 nm of each solution was measured. Blank was 100µl of BR buffer.

18. TSA coupling Validation

200 µl of BR buffer pH10 were added to one well pre-incubated with TSA as described before. The buffer in the first well moved to the next and so on eighth wells in a column. Abs 405 nm of the final solution was measured and compared with 200µl of BR buffer pH10.

19. TSA biopanning

The phage library, resuspended in PBS buffer, was added into each well of the CovaLink™ NH MicroWells™ plate (V_f/well = 100µl; C_f phage/well $=10^9$ cfu/ml). Three kind of wells were used: TSA coupled, NH-wells (only with sulfo-NH and no TSA) and empty. The selection was made at pH7, with and without Zinc, using two different buffers: BR pH7-ZnCl 0mM

Buffer and BR pH7-ZnCl 0.1mM Buffer. The incubation was at room temperature for 2h. The wells were washed 5 times, respectively, with BR pH7-ZnCl 0mM Buffer + 200mM NaCl and BR pH7-ZnCl 0.1mM Buffer + 200mM NaCl. Then, bound phage were eluted, respectively, with 100µl of BR pH7-ZnCl 0mM Buffer + 250µM TSA and BR pH7-ZnCl 0.1mM Buffer + 250µM TSA.
At this point eluted phage were used to infect E. coli XL1-blue cells. The cells were pre-grown on LB containing 1 % glucose up to an OD_{600} of 0.8. After 15 min incubation at 37 °C with no shaking, incubation was continued at 37 °C shaking for 1 h. Then aliquots of the bacterial solutions were plated on LB agar containing 1 % glucose, tetracycline (100 mg/ml) and ampicillin (100 mg/ml). Colonies that appeared on the plates the following day were counted to calculate the total number of selected phage.

Negative control: M13KO7 helper phage was incubated as phage library and the eluted material was titrated as described before for the library but using agar plates were with kanamycin.

Second cycle of selection: The eluted phage from the first cycle were used to infect E. coli XL1-blue cells, pre-grown on LB containing 1 % glucose up to an OD_{600} of 0.8. Phage production and precipitation were the same described before. Screening conditions were the same of the first selection, but washes rose to 10 times.

20. Immobilization of proteases on the plate

In each Maxisorp Immuno-plate well (NUNC) 200µl of protease solutions 0.5mg/ml (papain, pepsin, trypsin and α-chymotrypsin) in PBS were added. The incubation prolonged over night at 4°C. Then, the plate was washed 3 times with PBS and then 200 µl of BSA 1% were added to the well and incubated over night, at 4°C. Finally, the plate was washed 6 times with PBS.

21. Proteases biopanning

The phage library, resuspended in PBS buffer, was added into each well of the Maxisorp Immuno-plate (V_f/well = 100µl; C_f phage/well =10^9 cfu/ml). Two kind of wells were used: with protease and only with BSA 1%.

The incubation was prolonged for 2h at room temperature. The wells were washed 5 times with PBS. Then, bound phage were eluted with 100 µl of glycine 50 mM, pH 2.5 and then neutralized with NaH_2PO_4 2 M, pH 7.5.
At this point eluted phage were used to infect E. coli XL1-blue cells. The cells were pre-grown on LB containing 1 % glucose up to an OD_{600} of 0.8. After 15 min incubation at 37 °C with no shaking, incubation was continued at 37 °C with shaking for 1 h. Then aliquots of the bacterial solutions were plated on LB agar containing 1 % glucose, tetracycline (100 mg/ml) and ampicillin (100 mg/ml). Colonies that appeared on the plates the following day were counted to calculate the total number of selected phage.

Negative control: M13KO7 helper phage was incubated as phage library and the eluted material was titrated as described before for the library but using agar plates were with kanamycin.

Second cycle of selection: The eluted phage from the first cycle of selection were used to infect E. coli XL1-blue cells, pre-grown on LB, containing 1 % glucose, up to an OD_{600} of 0.8. Phage production and precipitation were the same described before. Screening conditions were the same of the first selection, but washes rose to 10 times.