Università degli Studi "Roma Tre"-Dipartimento di Biologia

Scuola Dottorale in Biologia
Sezione "Biologia Cellulare e Molecolare"
Tesi di Dottorato XXI° ciclo

"Investigation of the functional properties of a completely de novo random library of peptides. Implication for the origin of life research ".

"Studio delle proprietà funzionali di una libreria peptidica a sequenza casuale. Implicazioni per lo studio dell'origine della vita".

Candidato:
Dott. Davide De Lucrezia

Docente Guida:
Prof. Pier Luigi Luisi

RIASSUNTO

La scienza assume che la vita sulla Terra abbia avuto origine attraverso un processo spontaneo di autorganizzazione ed aumento di complessità a partire dalla materia inanimata. Diverse teorie sono state proposte per spiegare questo spontaneo aumento di complessità. Wächtershäuser identifica cicli metabolici privi di enzimi come il cardine alla base dell' origine della vita; Kauffmann propone cicli di peptide auto-catalitici come il principale motore, mentre Cech individuò in molecole auto replicanti di RNA la struttura fondamentale del primo sistema con proprietà "viventi". Al contrario, Luisi identifica nella struttura autopoietica dei sistemi viventi il principio primo, mentre Lancet enfatizza l'ereditarietà composizionale come caratteristica saliente dei sistemi viventi. Nonostante le differenze, tutte le teorie devono affrontare la stessa questione fondamentale: è il percorso di transizione alla vita univocamente determinato dalla legge della fisica e della chimica? Oppure è piuttosto il risultato della simultanea interazione di diversi fattori contingenti?

La controversia tra determinismo e contingenza emerge con forza se si considera l'emergere di biopolimeri funzionale nel quadro di origine della vita. Infatti, l'intera architettura della vita si basa su biopolimeri così che l'eziologia dei biopolimeri rappresenta un problema cruciale nel settore della ricerca sull'origine della vita. La questione principale è come biopolimeri funzionale siano stati selezionati in condizioni pre- o protobiotiche. Infatti, il numero di sequenze teoricamente possibile cresce esponenzialmente al crescere della lunghezza raggiungendo rapidamente cifre astronomiche così che non è ragionevole assumere che l'intero spazio delle sequenze sia stato esplorato durante l'evoluzione prebiotica. Da questa osservazione nasce l'interrogativo di come siano stati selezionati biopolimeri funzionali in un contesto prebiotico. Vi sono forse principi chimico-fisici particolari che sottendono alla selezione di biopolimeri funzionali? O piuttosto l'evoluzione molecolare è il risultato dell'interazione di fattori contingenti?

Nel tentativo di rispondere a queste domande una libreria di sequenze peptide completamente casuale è stata sintetizzata e studiata per le sue proprietà funzionali al fine di verificare la possibilità di isolare nuove sequenze funzionali che non presentassero omologie con quelle presenti in natura. La libreria è stata progettata senza vincoli strutturali o sequenza in modo che possa essere ragionevolmente considerata come omologa ad una popolazione di peptidi prodotta in condizioni prebiotiche.

La strategia sperimentale adottata ha utilizzato la tecnica dell'evoluzione *in vitro* che permette di testare simultaneamente un numero consistente di sequenze diverse al fine di individuare quelle che soddisfano un particolare criterio di selezione. L'evoluzione *in vitro* mima l'evoluzione naturale tramite cicli iterativi di mutazione-selezione-amplificazione. Ci sono due requisiti

fondamentali per applicare l'evoluzione *in vitro*: il primo è la possibilità di creare un legame diretto tra genotipo e fenotipo. Il secondo si basa sulla disponibilità di una idonea procedura di screening per selezionare le sequenze che soddisfano i criteri di selezione. In questo progetto di dottorato la connessione tra genotipo e fenotipo è stata ottenuta utilizzando la tecnica del *phage display*, mentre la capacità di legare un analogo dello stato di transizione (TSA) per la reazione di idrolisi di esteri ed ammidi è stata utilizzata come criterio di selezione.
La probabilità di successo in un esperimento di evoluzione *in vitro* è direttamente correlata alla complessità totale della libreria ed alla robustezza del processo di selezione. Conseguentemente, la prima parte di questo progetto di dottorato è stata dedicata alla costruzione di librerie di DNA codificanti peptidi a sequenza casuale di 20 residui e all'ottimizzazione della tecnica del *phage display*. Infine si è proceduto allo screening di suddetta libreria per la capacità di legare il TSA vincolante peptidi in diverse condizioni chimiche.
La costruzione di librerie peptidiche pone una serie di problemi tecnici che limitano la loro applicabilità, per ovviare a queste limitazioni un nuove vettore fagemidico è stato sviluppato per consentire la costruzione di librerie altamente degeneri e diverse. I risultati ottenuti sono riassunti qui di seguito:

- Complessità della libreria $\cong 10^8$
- Diversità della libreria > 70%
- Copertura dello Sequenza delle spazio $\cong 10^{-56}$

Un ulteriore problema inerente la tecnologia del *phage display* risiede nella eterogeneità della popolazione fagica. Tale eterogeneità limita severamente la porzione di spazio delle sequenze effettivamente esplorabile e riduce drasticamente la possibilità di trovare funzioni rare come quella catalitica. Per ovviare a questi limiti è stata condotta un'ottimizzazione del processo di produzione e purificazione dei fagi che garantisse l'omogeneità del prodotto. I risultati ottenuti sono riassunti qui di seguito:

- Rapporto tra fagi ricombinante e non ricombinante aumentato di un fattore 10 rispetto agli standard commerciali e di letteratura.
- Rapporto tra genotipo wild-type e genotipo ricombinante ridotto del 33%

La libreria è stata quindi sottosta a screening per il legame al TSA in diverse condizioni chimica ed in particolare a diversi pH (da 4 a 10) e concentrazione di zinco (0 mm a 10 mm). per mezzo di 4 cicli iterativi di selzione-amplificazione (biopanning). Al contempo, le condizioni di selezione, forza ionica e concentrazione di detergente, sono state modulate per minimizzare il binding aspecifico. Infine, i tempi di incubazione ed eluizioni sono stati

modificati nei diversi cicli di biopanning per favorire il ricupero di peptidi con costanti di affinità basse.
Il sequenziamento di cloni selezionati ha rivelato la presenza di pattern conservati al N- e C-terminale e la presenza di residui acidi a valle della regione N-terminale. L'analisi delle sequenze non hanno evidenziato altri pattern conservati ad eccezion fatta di quelli sopramenzionati. L'allineamento delle sequenze ha evidenziato una distribuzione isotropica delle stesse nello spazio delle sequenze. Inoltre, i risultati mostrano come il legame all'aptene sia favorito a bassi pH e non sia influenzato dalla concentrazione di zinco.
Questi risultati mostrano che è possibile recuperare selettivamente peptidi funzionali capaci di legare il TSA in diverse condizioni chimiche anche se emerge chiaramente che alcune condizioni (i.e. pH>4) risultano in una selezione non ottimale. Inoltre, l'analisi delle sequenze mostra una notevole eterogeneità dei peptidi selezionati che suggerisce che vi siano diverse famiglie di sequenze non omologhe capaci di legare il TSA. Nel loro complesso i risultati suggeriscono che biopolimeri funzionali diversi da quelli presenti in natura sono uniformemente distribuiti nello spazio delle sequenze a sostegno della teoria della contingenza.

ABSTRACT

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 identifies enzyme-free metabolic cycles as the pivotal system underpinning life's origin; Kauffmann proposes auto-catalytic peptide cycles as the primary motor, whereas Cech fostered the idea that RNA was the scaffold of the first living system. Conversely, Luisi emphases the autopoietic nature of life; whereas Lancet proposes composition inheritance as the foundation of life. Despite the differences, all theories must confront the same fundamental question: is the transition to life pathway determined univocally by the law of physic and chemistry? Or it is rather the result of the simultaneous interplay of different contingent factors?

The controversy between determinism and contingency emerges forcefully when one considers the emergence of functional biopolymers in the framework of the origin of life. Indeed, the entire architecture of life relies on biopolymers so that the aetiology of biopolymers represents a major issue in the field of origin of life research. The main question is how functional biopolymers have been selected under pre- or protobiotic conditions. Indeed, the number of theoretically possible sequences quickly reaches astronomic figures as the length increases so that the correspondent sequence space could not have been sampled exhaustively during natural evolution even for short biopolymers. A straightforward question arises: how were biopolymers selected? Were these biopolymers the best ones ever possible? Or were they simply the outcome of contingency shaped by natural evolution?

To tackle this question, a completely de novo random library of short peptides has been designed and tested for potential catalytic activity. The 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 criterion was based on the ability to bind a transition state analogue of the ester and amide bond hydrolysis in order to asses the frequency and distribution in sequence space of functional peptides.

The ultimate objective of this doctoral project was to investigate the catalytic properties of a random library of peptides in order to assess whether and to what extend functional biopolymers display catalytic function and how functional peptides are distributed in sequence space.

Accordingly, the envisaged experimental strategy had to ensure the screening of a vast library of random peptide in order to explore effectively the sequence space. The choice fell on *in vitro* evolution which allows the simultaneous screening of a vast library of candidate peptides for *a priori* defined function without a prior knowledge. *In vitro* evolution mimicries natural evolution in a test tube by means of iterative cycles of mutation-selection-amplification.

There are two fundamental requirements to carry out directed evolution: the first is the availability of physical link between the genotype and the phenotype. The second relies on the availability of a suitable screening procedure to enrich the initial peptide population of those sequences satisfying the selection criteria. In this doctoral project the phage display technique has been employed to ensure a physical link between the genotype and the phenotype, whereas the binding to a transition state analogue (TSA) for the ester and amide bond hydrolysis has been used as selection criterion.
The likelihood of success in a *in vitro* evolution experiment is directly related to the total size library, as evaluating more sequences increases the chances of finding one with the desired properties. Accordingly, the first part of this doctoral project has been devoted to the construction of a DNA library encoding for random 20mer peptides and to the optimization of the screening technique. Subsequently, the designed library and the optimised screening technique have been used to select TSA-binding peptides under different chemical conditions.
The selection of catalytic peptides from a random library of sequences has never been attempted before and poses a number of technical challenges due to technical difficulties related to short fragment cloning, short peptide expression and purification. To tackle these problems a novel multipurpose vector has been developed that allowed the cloning of short DNA fragment with >90% efficiency. The results obtained are summarised hereafter:

- Library complexity $\cong 10^8$
- Library diversity: >70%
- Sequence space coverage: 10^{-56}

Phage display relies on the presentation of foreign peptides/protein on M13 phage capside whereas the correspondent gene is encapsulated within. One of the major bottleneck of phage display is the inefficient display of foreign peptides on the virion capside and the encapsulation of wild-type phage genome into the recombinant viral particles. These limitations severely affect the screening efficiency and ultimately reduced the sequence space coverage. To overcome these limitations, the phage display techniques has been optimised in order increase the ratio of phage displaying foreign peptides and minimize the ratio of wild-type genome encapsulation. The results obtained are summarised hereafter:

- recombinant : non-recombinant phage ratio increased 10-fold
- wild-type : phagemid ratio decreased 33%

The random DNA library encoding for random peptides has been screened for binding to the TSA under different chemical condition with respect to pH (4 to 10) and zinc concentration (0 mM to 10 mM). Phage library has been subjected

to 4 selections round of increasing stringency with respect to incubation time, wash condition (ion strength and surfactant concentration) and elution time.
Sequencing of selected clones retrieved a highly conserved consensus sequence at the N- and C-terminus of selected peptide, no consensus sequence could be identified except for a conserved acid residues downstream the conserved N-terminal region. Sequences alignment shows that selected peptides are isotropically distributed in sequence space. In addition, recovery yield greatly changed under different chemical condition with highest recovery yield at pH 4 with suboptimal recovery at different pH. Finally, zinc concentration does not seem to affect TSA binding irrespective of the pH.
These results show that is possible to selectively recover peptides binding to the transition state analogue (TSA) for the ester and amide bond hydrolysis reaction under different chemical environment although this results in a suboptimal selection under certain conditions. In addition, sequence analysis shows a remarkable heterogeneity of selected peptides that may suggest that multiple sequences are capable to perform binding. Although an enzymatic validation of results is required, results suggest that potentially functional sequences are evenly distributed in sequence space supporting the contingency theory.

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 emphases the autopoietic nature of life [Luisi, 2003 (I)]; whereas Lancet proposes composition inheritance as the foundation of life [Segre and Lancet, 2000].

Despite the differences, all theories must confront the same fundamental question: is the transition to life deterministically specified by initial condition? In other words, 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)?

The controversy between determinism and contingency emerges forcefully when one considers the emergence of functional biopolymers in the framework of the origin of life [Luisi, 2003]. Indeed, the entire architecture of life relies on biopolymers so that the aetiology of biopolymers represents a major issue in the field of origin of life research. The main question is how functional biopolymers have been selected under pre- or protobiotic conditions. Indeed, the number of theoretically possible sequences quickly reaches astronomic figures as the length increases so that the correspondent sequence space could not have been sampled exhaustively during natural evolution even for short biopolymers. A straightforward question arises: how were biopolymers selected? Were these biopolymers the best ones ever possible? Or were they simply the outcome of contingency shaped by natural evolution, namely completely different functional biopolymers could have been possible?

To tackle this question, a completely de novo random library of short peptides has been designed and tested for potential catalytic activity. The 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 criterion was based on the ability to bind a transition state analogue of the ester and amide bond hydrolysis in order to asses the frequency and distribution in sequence space of functional peptides.

1.1 Origin of Life and the contingency vs. determinism debate

The Oparin-Haldane theory [Oparin, 1954a;1954b; 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 as we know it is based on the action of proteins and nucleic acids, whose functions are only due to a specific sequence, the one that produce the folding and the activity. These long sequences are co-polymeric structures, namely they are formed by macromolecules having chemically different monomer units in the same chain.

The assumption that life derives from inanimate matter bears a profound enigma about the very nature of evolution as authorevously conceptualized 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 determines the outcome of a given event.

In principle, 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. In 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 several authors [Morowitz, 1993; de Duve, 2002], like Christian de Duve [de Duve, 1995] who wrote: "given the suitable initial conditions, the emergence of life is highly probable and governed by the laws of chemistry and physics [...]", which suggests the idea that life on Earth was inescapable as restated recently [de Duve, 2002].

To support the contingency theory one may cite Jacques Monod with his *Chance and Necessity* [Monod, 1971], and 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. Any change in the starting contingent conditions would dramatically affect the final result. The implications are severe, in fact citing Stephen Jay Gould, "[...] run the tape

again, and the first step from prokaryotic to eukaryotic cell may take twelve billions years instead of two" [Gould 1989].
Within this framework, one of the most interesting questions in modern life science is how prebiotic evolution of the first biopolymers occurred. The sequences, or primary structures, of existing biopolymers are believed to be a product of evolution (following life onset) or prebiotic molecular evolution (prior to life onset). 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 grows exponentially as the length of the polymer increases. For this reason, Nature could not perform an exhaustive search in sequence space for functional biopolymers.
Focussing on polypeptides, taking a polypeptide of 20 residues there are 20^{20} ($\sim 10^{26}$) possible 20mer products. If only one molecule of each of these peptides were to be synthesized, approximately $1.8 \bullet 10^{41}$ 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 $\sim 3 \bullet 10^{51}$ years to complete their synthesis. By and large, we are faced with the problem of how nature explored the sequenced space in order to find functional polypeptides instrumental to life.
The important question is whether or not it is likely that functional sequences appeared in the prebiotic scenario. Ascribing the origin of polypeptides to nucleic acid mediated mechanisms is not a satisfactory answer, as it simply shifts the question of the aetiology of specific sequences from proteins to polynucleotides. This is all concerned with the aetiology of proteins, the question being whether their selection is based primarily on deterministic principles, or governed chiefly by contingency. 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 this is the case, there exists an universe of proteins which properties have never been sampled by nature: the Never Born Proteins - NBP [Luisi, 2003 (II)].
The questions we wish to address are: which is the fraction of all possible sequences displays a function? Are those sequences isotropically distributed in the sequence space? Do functional peptides resemble extant enzyme sequences?
To tackle these questions, the possibility of selecting functional peptides from a completely random library without any sequence or structural constrain has been carried out.

1.2 Peptide synthesis under prebiotic conditions

Peptide synthesis under plausible prebiotic conditions have long challenged organic chemists. In the 70' Fox and coworkers [Brooke and Fox, 1977; Nakashima and Fox, 1980] shows that *proteinoids* (bodies containing polymerized amino acids) can be formed 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. It was however reported that when using amides in presence of clay increases the yields during repeated drying and heating, and Ito and coworkers [Yanagawa, 1990] reported a substantial arrays of polypeptides prepared in this way.

Limtrakul and Rode [Limtrakul, 1985] proposed that under high local ions concentrations -promoted by tidal evaporation- incompletely hydrated ions of metal ions may activate a dehydration leading to peptide condensation. From this, the technique of the Salt Induced Peptide Condensation (SIPC) has been developed [Rode, 1999; Saetia, 1993; Oie, 1983].

There is some claim that long polypeptide chains may derive not so much from the condensation of aminoacids, but from the polymerization of HC≡N followed by simple prebiotic chemistry to mould the side chains [Matthews, 1975] – but this theory has not yet found a great support in the field.

A new interesting development is offered by Orgel and Ghadiri's groups [Leman, 2004]; they could show 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.

Finally it is interesting to mention the condensation of N-carboxyanhydrides (also 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 been 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 the synthesis is specific for α-amino acids. Oligomers up to 10-mers can be obtained in one single step. Subsequently, decamers could further react to yield longer polypeptides by means of fragments condensation [Chessari, 2006].

Accordingly, the present doctoral project will focus on short peptides 20 residues long in order to investigate the potential catalytic properties of

peptides resemble those that might have been synthesised under plausible prebiotic conditions.

1.3 The quest for catalytic peptides

The elegance and efficiency of enzymatic catalysis have long tempted chemists and biochemists with reductionist leanings to try to mimic the functions of natural enzymes in much smaller peptide. Accordingly, catalytic peptides have been the object of numerous researches due to their relevance for the origin of life research as primordial catalysts and for their potential application in the biotechnological sector for highly selective biotransformation processes. Enzymology and protein biochemistry over the last decades have elucidated several features of natural enzymes that seem to account for the enzymes' catalytic proficiency and versatility:

1. Entropic effects of enzyme-substrate association and substrate immobilization;
2. Exclusion of solvent from the active site cavity;
3. Transduction of binding energy;
4. Arrangement of specific chemical groups at fixed locations in the enzyme;
5. Electronic "tuning" of the active site.

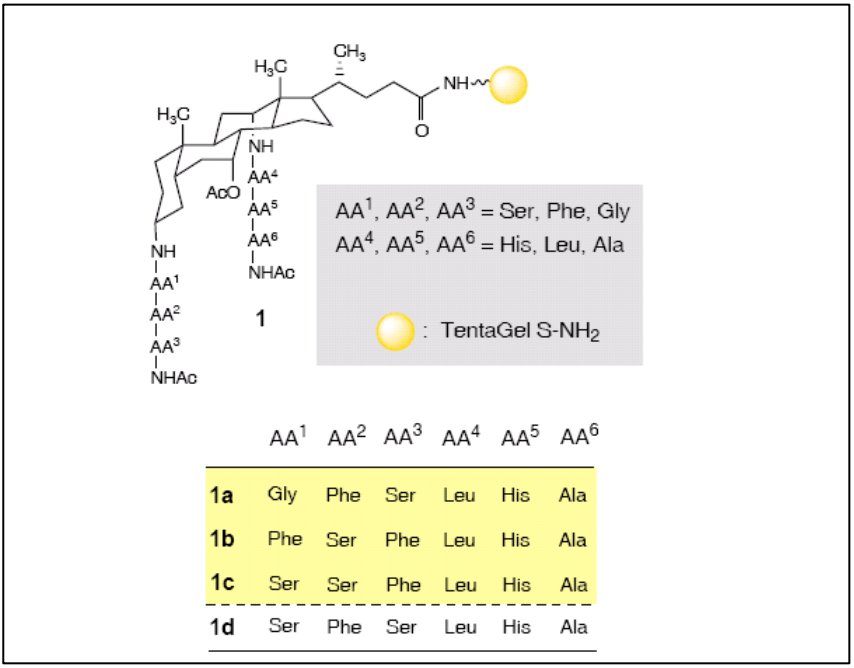
These categories are somewhat arbitrary and do not represent an holistic view of the catalytic capabilities of enzymes. "Entropic Effect" includes both conformational freezing and proximity effects. The power of an enzyme to immobilize substrates is also important in bringing reactants into proximity in multi-substrate reactions [Bruice and Pandit, 1960]. "Solvent Exclusion" embodies both exclusion of water and other solvent molecules as competitive reagents, and development of a hydrophobic cavity in which the dielectric constant may be vastly different from that observed in solution, thereby enhancing dipole and charge-charge interactions. "Binding Energy Transduction" relates to the deformation of substrate geometrical and electronic configuration to facilitate reaching the transition state [Jencks, 1975; Baumann et al., 1970; Thompson and Blout, 1973; Sanchdev and Fruton, 1970; Moussaoui et al., 1996]. The full enthalpy attributable to the sum of the favourable interactions in the enzyme-substrate complex is generally not observed in the ΔH of binding, because a portion of the enthalpy of association is reserved as potential energy in the form of conformational strain in the enzyme and/or substrate; this strain may be released as the transition state is approached, or new favourable interactions may be realized. These effects, which collectively reduce the energetic barrier to be overcome in catalysis, are generally known as "transition state stabilization" or "transition state complementarity," depending on whether one is referring to the energetics or

the interactions [Jecks, 1969; Fersht, 1977; Pauling 1948]. “Reactive groups arrangement” is the most obvious mechanism by which enzymes promote chemical transformation, but it is closely related to “Entropic effect”, in that immobilization of the chemical groups in proximity to the substrate is actually an entropic form of rate enhancement promoting a local concentration effect. "Electronic Tuning” depends on both the active site environment, which often has a higher dieletric constant than bulk solvent, and the positioning of polar side chains or dipoles composed of multiple residues in the protein structure. The concept of orbital steering [Storm and Koshland, 1970] holds that the orientations of the molecular orbitals of the chemically active groups of the enzyme and substrate are both fixed in position and electronically tuned by the active site.

Authoritative authors argue that none of these features of enzymes and enzyme-substrate complexes are fully available to peptides without fixed conformations and conclude that scientists will unlikely be able to reproduce the marvellous properties of enzyme in smaller systems [Corey and Corey, 1996]. In particular, Corey observes that “Entropic Effect” is virtually ruled out, since stable, many-point contacts between a peptide and another small molecule, neither of which has a fixed three-dimensional structure. “Solvent Exclusion” and “Binding Energy Transduction” would be highly improbable in the absence of a stable binding cavity. However Corey states that “Reactive groups arrangement” is not manifestly impossible for peptides, however this feature is severely limited by the difficulty to deliver appropriate three-dimensional scaffolds for proper placement of the tools of catalysis. Some authors argue that smaller, disordered peptides may contain all the functional groups of enzyme active sites, but even if the peptide is capable of assuming a catalytic conformation, it is conceivable, given the enormous magnitude of shape space accessible to a peptide of even 30 amino acid residues, that the age of the planet would elapse before the desired conformation were to be assumed [Levinthal, 1969]. Finally, while limited provision of polar groups and dipoles necessary for “Electronic tuning” may be possible in a peptide, the fixed alignment and hydrophobic cavity, which aim and tune these electronic effects, are likely to be inadequate or absent. In short, it is believed that *“[…] all the major features of enzymatic catalysis are likely to be deeply flawed or entirely missing in peptides without fixed conformations […]”* [Corey, cit.].

This common knowledge has deeply affected the research on catalytic peptides over the last decades to such extent that few attempts have been made to investigate the catalytic properties of biological, non-constrained peptides.

For example, De Clercq and co-workers [De Muynck et al., 2000; Madder et al., 2002] have synthesized a non-biological peptidosteroid library (Figure 1.1) and screened it for esterase activity.



	AA^1	AA^2	AA^3	AA^4	AA^5	AA^6
1a	Gly	Phe	Ser	Leu	His	Ala
1b	Phe	Ser	Phe	Leu	His	Ala
1c	Ser	Ser	Phe	Leu	His	Ala
1d	Ser	Phe	Ser	Leu	His	Ala

Figure 1.1
Non-biological peptidosteroid library for the identification of esterases. Peptidosteroid as model for serine esterase. The table show the amino acid composition of members of peptidosteroid library. Active residues are highlighted in yellow. Adapted from Berkessel, 2003.

Authors performed a bulk screening of the library employing red p-nitrophenyl acetate to identify by deconvolution active member. Control experiments revealed that the cooperative action of Ser and His accounts for the acceleration observed.
Berkessel and Herault [Berkessel and Herault, 1999] prepared a solid matrix-constrained combinatorial peptide library of 625 members to indentify catalytic peptides capable to hydrolyse the phosphodiesteric bound. Authors were able to identify active sequences characterised by a well defined serine-histidine-arginine pattern capable to hydrolyse indolyl phosphate.
Berkessel and Riedl [Berkessel and Riedl, 2000] extended their combinatorial search for N-acylated solid-phase-bound hydrolase models. Authors identified 2 peptides (Figure 1.2) capable to cleave bis(p-nitrophenyl)phosphate in the presence of Zn or Eu.

Figure 1.2
Representation of the "minimal catalophore" identified by Berkessel and co-workers. Adapted from Berkessel, 2003

Authors analyzed their hit structures to identify the minimal subunit necessary for catalytic activity by mutations and deletions. TentaGel-bound histidine turned out to be the only crucial structural element. In addition, since neither PEG-bound histidine nor histidine amide were active [Berkessel and Riedl, 2000], this observation stresses the importance of the solid support as an instrumental part to the overall catalytic property.
In the approach by Miller et al. [Miller et al., 1998], short oligopeptides (3-mer and 4-mer) incorporating N-alkyl histidine derivatives were employed as enantioselective acylation catalysts. The design of the tripeptide and the tetrapeptide [Copeland et al., 1998] incorporates the N-alkylated and catalytically active His-derivative, a Pro-Aib sequence, to induce proper folding-back of the catalyst, and further elements of chirality such as phenethylamine.
Results revealed that, in particular, acyl amino alcohols were efficiently distinguished, whereas both enantiomers of 1-(1-naphthyl)ethanol were acetylated at identical rates. The activity of all peptides presented in the context of acylation catalysis was based on non-biological amino acid derivates such as N-imidazolyl alanine or p-methylated histidine.
Jacobsen et al. [Sigman and Jacobsen, 1998] were able to identify and optimize metal-free 'peptoid' catalysts (Figure 1.3) for the addition of HCN to imines affords α-aminonitriles, which upon hydrolysis give a-amino acids, a reaction known as Strecker reaction.
The (thio)urea derivatives afford enantiomeric excesses >90% for a wide variety of substrate imines and they are excellently suited for immobilization on polystyrene supports [Sigman et al., 2000]. In addition, Jacobsen's peptoids catalyze the addition of HCN to ketoimines with enantiomeric excesses up to 95% [Vachal and Jacobesen, 2000].

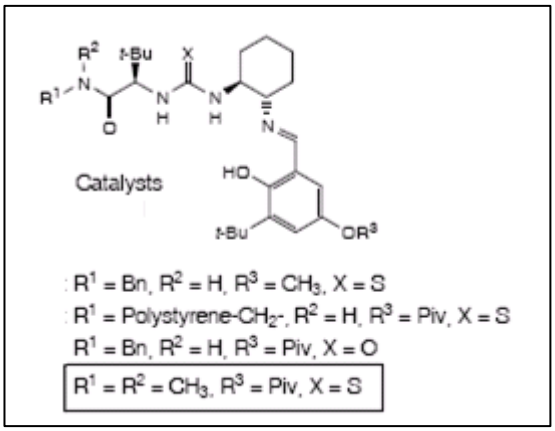


Figure 1.3
Structure of peptoids catalysing asymetric addition of HCN to imenes. Adapted from Sigman and Jacobsen, 1998.

These examples highlight the catalytic potential of short peptides, however it has to be stressed that those peptides are not composed of proteinogenic amino acids and their activity strongly depends on organocatalytic groups or non-biological supports.
Some research groups have focused on proteinogenic peptides, most notably the work reported by Johnsson [Johnsson et al, 1993] and Tanaka [Tanaka, 2005].
Johnsson reported the rational design of polypeptides that catalysed the decarboxylation of oxoaloacetate via an imine intermediate. The polypeptide exploit mechanism-based design in order to rationally engineer an amphiphilic α-helix scaffold capable of building helix bundles in aqueous.
Positively charged residues were appropriately located along the α-helix sequence in order to bind the dianionic carbinolamine by coulombic interaction. To achieve the envisaged goal the pK_a of the amino groups had to be finely tuned and this was achieved either by stabilising the terminal amino group of the polypeptide chain by interaction with the helix dipole or by blocking the n-terminus amine and lowering the pK_a of the ε-amino group of lysine side chain by interaction with neighbour protoneted lysines. This remarkable results of biochemical engineering could catalyse the decarboxylation of oxaloacetate with Michaelis-Menten saturation kinetics (K_{cat} = 0.4 min^{-1}, K_m= 14 mM).
Tanaka and co-workers reported the selection of peptides possessing aldolase activity by screening a phage display library for enaminone formation, a covalent intermediate in the enamine-based aldolase reaction. The starting

library results from the addiction of a random stretch of 6 amino acids to a conformationally defined 18-mer capable to adopt a stable α-helix and possessing oxaloacetic acid decarboxylation activity [Perez-Paya, 1995]. Selected peptides showed Michaelis-Menten kinetics with $K_{cat} = 2,1 \bullet 10^{-4}$ min^{-1}, $K_m = 1,8$ mM.

Those two examples clearly demonstrate the catalytic potential of peptides made of proteinogenic amino acids, however they heavily relay on *a priori* well-defined secondary structure that serve as stable scaffold for further engineering processes.

By and large, the aforementioned examples describes peptides which catalytic properties result from either extensive chemical modification of complex *a priori* defined secondary structure.

To this regard, whether and to what extend completely random sequences -with neither sequence nor structural constrains such those synthesised under plausible prebiotic conditions- may display a catalytic function remains an open question.

1.4 Experimental strategy

The ultimate objective of this project was to investigate the catalytic properties of a random library of peptides in order to assess whether and to what extend functional biopolymers display catalytic function and how functional peptides are distributed in sequence space.
Accordingly, the envisaged experimental strategy had to ensure the screening of a vast library of random peptides in order to explore effectively the sequence space. The choice fell on *in vitro* evolution which allows the simultaneous screening of a vast library of peptide candidates for *a priori* defined function without a prior knowledge. *In vitro* evolution mimicries natural evolution in a test tube by means of iterative cycles of mutation-selection-amplification (Figure 1.4).

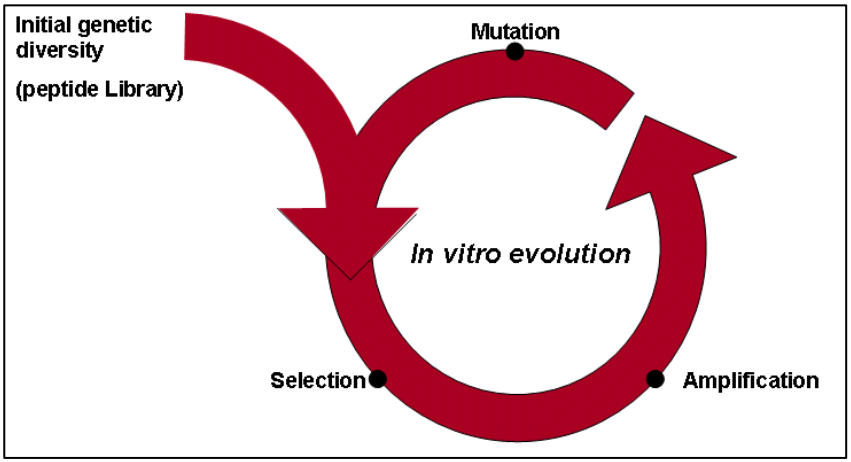


Figure 1.4
Schematic representation of an *in vitro* evolution cycle.

There are two fundamental requirements to carry out directed evolution: the first is the availability of physical link between the genotype and the phenotype (see chapter 1.4.1). The second relies on the availability of a suitable screening procedure to enrich the initial peptide population of those sequences satisfying the selection criteria (see chapter 1.4.2). The envisaged methodology has to allow the screening under different chemical conditions in order to explore the catalytic potential under a wide range of plausible prebiotic conditions and in particular over a broad range of pH and zinc concentration (see chapter 2.6).

1.4.1 Linking genotype to phenotype: The Phage display technology

There are a number of techniques that allow to physically link the genotype to phenotype and therefore enable in vitro evolution, most notably:

- In vivo selection technologies
  - *direct immunization*
    Direct immunization (DI) allows the isolation of antibody against a variety of targets. When coupled with transition state analogue or bait-and-switch, DI allows the isolation of catalytic antibodies. The methodology exploit the vast repertoire of natural immune system eliciting target-specific antibody by means of direct inoculation of the target, fused to immunogenic protein carrier, into mammalian hosts. Although DI bears enormous potential, this technique is limited to the discovery of catalytic antibody [Tanaka and Barbas, 2002].

  - *yeast or bacterial display*
    Bacterial [Freudl et al., 1986; Fuchs, 1991; Agterberg et al., 1990] and yeast [Boder and Wittrup 1998, 2000; Kieke er al 1997] display is a protein engineering technique used for in vitro protein evolution. Libraries of polypeptides displayed on the surface of bacteria can be screened using flow cytometry or iterative selection procedures (biopanning). In yeast display, for example, a protein of interest is displayed as a fusion to the Aga2p protein on the surface of yeast. The Aga2p protein is naturally used by yeast to mediate cell-cell contacts during yeast cell mating. As such, display of a protein via Aga2p expose the protein on the outer surface of the cell surface, minimizing potential interactions with other molecules on the yeast cell wall. The use of magnetic separation and flow cytometry in conjunction with a yeast display library is a highly effective method to isolate high affinity protein ligands against nearly any receptor through directed evolution. Advantages of yeast display over other in vitro evolution methods include eukaryotic expression and processing, quality control mechanisms of the eukaryotic secretory pathway, minimal avidity effects, and quantitative library screening through fluorescent-activated cell sorting (FACS). Disadvantages include smaller mutagenic library sizes compared to alternative methods [].

- In vitro selection technologies
  - *mRNA display*
    The mRNA display was developed by Roberts and Szoztak [Roberts and Szostak, 1997] and results in translated peptides or proteins that are associated with their mRNA progenitor via a puromycin linkage. The complex then binds to an immobilized target in a selection step (biopanning or affinity chromatography). The mRNA-protein fusions that bind well are then reverse transcribed to cDNA and their sequence amplified via a polymerase chain reaction. The end result is a nucleotide sequence that encodes a peptide with high affinity for the molecule of interest. Puromycin is an analogue of the 3' end of a tyrosyl-tRNA with a part of its structure mimics a molecule of adenosine, and the other part mimics a molecule of tyrosine. Compared to the cleavable ester bond in a tyrosyl-tRNA, puromycin has a non-hydrolysable amide bond. As a result, puromycin interferes with translation, and causes premature release of translation products. This enables the efficient production of high-quality, full-length mRNA-polypeptide fusion. After each round of selection, those library members that stay bound to the immobilized target are PCR amplified, and non-binders are washed off. The major drawback of mRNA display is the chemical sensitivity of the mRNA-puromycin-peptide complex to chemical conditions.

  - *Ribosome display*
    Ribosomal display was first described by Dower and coworkers in 1994 [Mattheakis et al., 1994] and successively optimised by Plückthun's group [Hanes et al., 1997]. The process results in translated proteins that are associated with their mRNA progenitor as a non-covalent trimolecular complex with the ribosome. In turn, this complex is used to bind to an immobilized ligand in a selection step. The mRNA-protein hybrids that bind well are then reverse transcribed to cDNA and their sequence amplified via PCR. The end result is a nucleotide sequence that can be used to create tightly binding proteins. Ribosome display either begins with a DNA sequence or naive library of sequences coding for a specific protein. The sequence is transcribed, and then translated in vitro into protein. However, the DNA library coding for a particular library of binding proteins is genetically fused to a spacer sequence lacking a stop codon. This spacer sequence, when

translated, is still attached to the peptidyl tRNA and occupies the ribosomal tunnel, and thus allows the protein of interest to protrude out of the ribosome and fold. the results is a complex of mRNA, ribosome, and protein which can bind to surface-bound ligand. This complex is stabilized with the lowering of temperature and the addition of cations such as Mg^{2+}. During the subsequent panning, the complex is introduced to surface-bound ligand. The complexes that bind well are immobilized. Subsequent elution of the binders allows dissociation of the mRNA. The mRNA can then be reverse transcribed back into cDNA, undergo mutagenesis, and iteratively fed into the process with greater selective pressure to isolate even better binders. The main limitation of ribosome display is the non-covalent link between the genotype and the phenotype which narrows the application of this technique.

- *Phage display*
  Phage display technology, introduced by G. Smith [Smith,1985], is a good method for selecting functional specific molecules from large peptide or proteins libraries [Siegel, 2001; Ladner, 2000; Hoogenboom, 1998; Burton, 1995; Ladner, 1995; Neri, 1995; Winter, 1994; Griffiths, 1993; Barbas 1993]. Phage display has been successfully applied to a wide series of different purposes such as to study protein-protein interactions [Cesareni, 1992], receptor and antibody-binding sites [Winter, 1994; Griffiths, 1993; Better, 1988; Skerra, 1988], to investigate protein stability [Kotz, 2004; Hoess, 2001; Forrer, 1999], for the directed evolution of enzymes [Fernandez-Gacio, 2003] and to improve or modify the affinity of proteins for their binding partners [Burton, 1995]. This technology utilizes the ability to express foreign proteins on the surface of phage particles as fusion 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.

In order to pursue the project objective, the phage display technology has been chosen since it allows the selection of large and diverse peptide libraries under a broad range of chemical conditions. To this regard, phage display is superior to in vivo technologies since the latter allow the screening of only modest library (up to 10^6 clones simultaneously). In addition, phage display provides superior performance with respect to other in vitro selection methodologies, in fact both mRNA and ribosome display are severely impaired by extreme pH and ion concentration and therefore do not suit the experimental conditions chosen for the selection (see chapter 2.5).

Biology of Filamentous Phage

The 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 1.5) and they are able to infect a variety of gram negative bacteria. Three bacteriophage, M13, fl, and fd, are the best characterized of those infecting *E. coli* cells containing the F conjugative plasmid. Their genome has been completely sequenced resulting 98% homologous [Van Wezenbeek, 1980; Beck, 1981; Hill, 1982].

The entire genome of these phage consists of 11 genes (Figure 1.6). Two of these genes, X and XI, overlap and are in-frame with the larger genes II and I [Model, 1988; Rapoza, 1995]. 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) for a protein necessary both at 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 capside proteins. The “Intergenic Region” is a short stretch of DNA which encodes no proteins. It 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 1.6). 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].

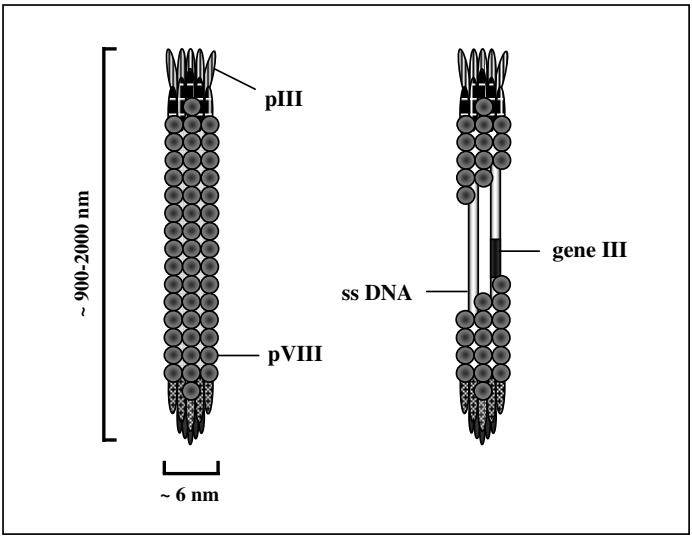


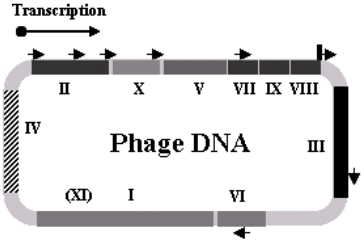
Figure 1.5
Schematic drawing of phage particles. pIII is the minor coat protein three, while pVIII represents the major coat protein eight. The phage genome is constituted by a single stranded DNA (ss DNA) encased into a cylindrical phage particle.

Infection of *E. coli* by the bacteriophage is at least 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 1.7). The pilus is retracted, presumably by depolymerization of the pili subunits into the inner membrane, so the tip of the phage is moved to the membrane surface. This event could be due to 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. Subsequently 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 1.7).
Where the inner and outer membranes of the *E. coli* cells are in close contact starts the assembly of the filamentous phage [Lopez, 1985]. These assembly sites may be the result of specific interactions between pI, pIV and pXI (Figure

1.7). The initiating event for assembly of the particle is probably the interaction of the pV-phage DNA complex with proteins in the assembly site. 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 single-stranded DNA binding protein pV 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. pV is 87 amino acids long and its biological functional entity is a homodimer. 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. 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 1.7).

Infection of *E. coli* by the bacteriophage is at least 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 1.7). The pilus is retracted, presumably by depolymerization of the pili subunits into the inner membrane, so the tip of the phage is moved to the membrane surface. This event could be due to 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. Subsequently 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 1.7).
Where the inner and outer membranes of the *E. coli* cells are in close contact starts the assembly of the filamentous phage [Lopez, 1985]. These assembly sites may be the result of specific interactions between pI, pIV and pXI (Figure 1.7). The initiating event for assembly of the particle is probably the interaction of the pV-phage DNA complex with proteins in the assembly site. 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 single-stranded DNA binding protein pV 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. pV is 87 amino acids long and its biological

functional entity is a homodimer. 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. 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 1.7).



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 1.6
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.

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 while pVIII molecules are following added along the length of the particle in thousands 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.

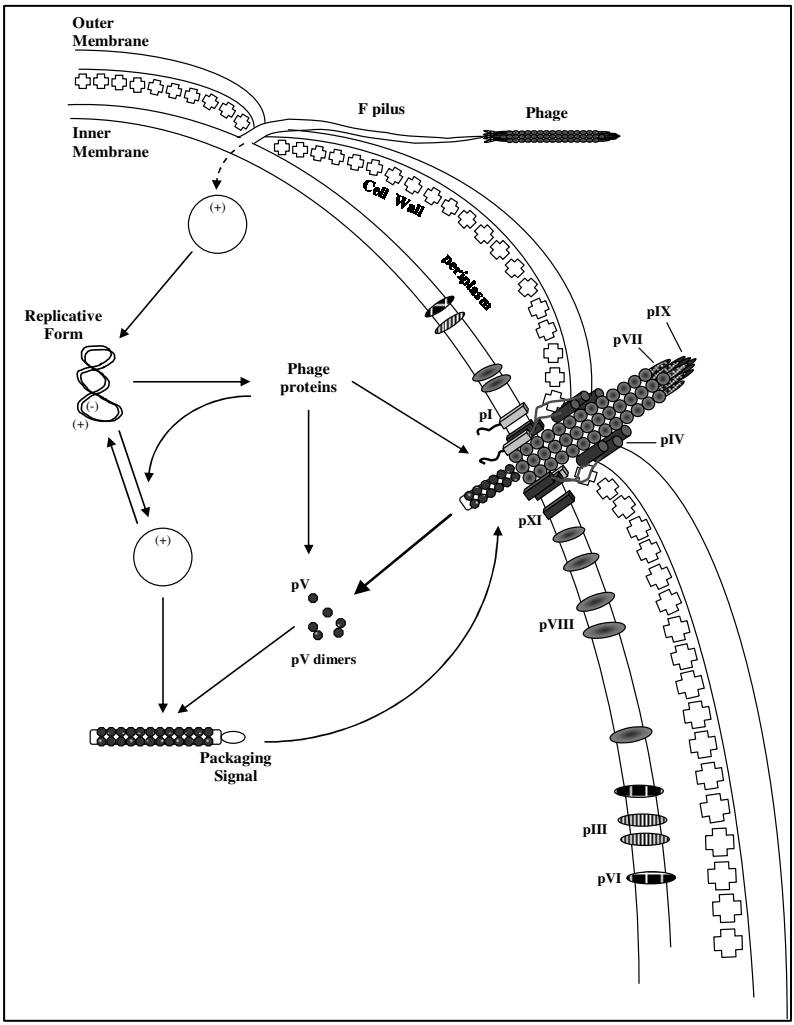


Figure 1

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.

Phage display vectors

There are different formats for the Phage display that either exploiting wild-type viral genome of engineered phagemid DNA:

- *3 multivalent Phage display*

  In this format the exogenous peptide or protein is expressed on the virion capside as N-terminal fusion protein to minor coat protein. In this format the corresponding encoding gene is cloned upstream the gIII into a wild-type viral vector. The result is that every copy of the capside protein displays the fusion producing a multivalent display (Figure 1.8Figure 2).

- *3 + 3 monovalent Phage display*

  In this format 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, but co-infection with a helper phage results 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 vary greatly and 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 also the growth conditions. In this type of display the infectivity is not compromised since 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 noninfectious phage when in multivalent phage display (Figure 1.9).

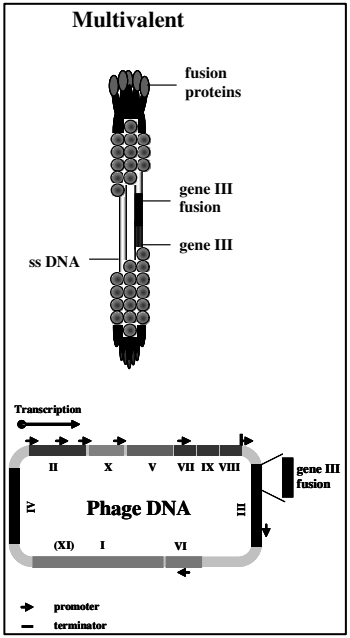


Figure 2
Schematic representation of the 3 multivalent phage display format

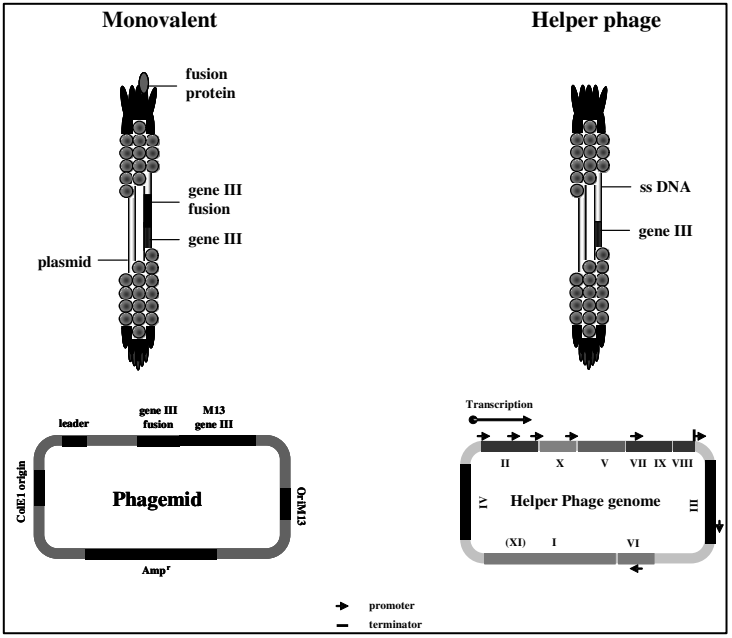


Figure 3
Schematic representation of the 3+3 phage display format.

The two formats described above differ for the number of protocols details; however the major difference lays in the selection output. On one hand the "3 multivalent phage display" allows the selection of peptides binding to the target with a high K_d exploiting the high avidity due to the display of multiple copies of the exogenous peptide on the virion capside. Conversely, the "3+3 monovalent phage display" allows the selection of peptides with low K_d exploiting the high affinity due to the monovalent copy of foreign peptide display on the virion capside. Due to the lack of any previous knowledge on the frequency and affinity of catalytic peptides in a completely random library both formats have been employed during the screening procedure (see chapter 2.3 and 2.4).

It is noteworthy that the synthesis of libraries of peptides 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 [Ladner, 1995; Neri, 1995; Ladner, 2000; Siegel, 2001]. Their main interest is in the isolation of proteins with, for example, improved stability, new or altered catalytic properties or in proteins that bind target molecules with enhanced affinity [Doorbar and Winter, 1994; Ho et al., 2006]. In contrast, the present work is aimed to investigate the frequency of potentially catalytic peptides in a completely random library without any constrain on sequence and secondary structure. To this regard, random protein space has been explored several times to search for optimized enzymatic functionalities [Mordukhova et al., 2008; Vazquez-Figueroa et al., 2008; Dumon et al., 2008, Seelig and Szostak, 2007], but generally such investigations, defined as directed molecular evolution, have been carried out starting from selected extant protein scaffolds and randomizing either restricted regions or the entire gene. Those approaches can be defined as "directed randomizations", in the sense that randomization is performed in order to achieve certain desired properties. The rationale of the approach described in this work is completely different, as there is no bias towards any given sequence or structural feature. It is a "total randomization" approach that leads almost necessarily to explore the sequence space toward regions that have not been sampled by nature during the natural evolution – the notion of "Never Born Peptides, NBP".

There are of course many unknown proteins in nature that have not been isolated and studied as yet. However, the NBPs, being products of an *ab initio* random choice, are characterized by a lack of any evolutionary pressure. In this sense they may open new avenues of investigation in not only basic science but also biotechnology.

1.4.2 Selection of biocatalyst by biopanning: binding to transition state analogue

Selection of novel or improved biocatalysts has been a long-term goal of biochemist and protein engineers. To this regard several experimental strategies have been reported in literature, among which:

1. *Screening by bait-and-switch.*
   Small compounds that carry positive or negative charges can be used for the selection of potential catalyst having complementary charged amino acid residues in the catalytic site. Based on this fact, a new approach for hapten design has been developed to expand the repertoire of screening assays for catalysis. This strategy involves the placement of a point charge on the hapten in close proximity to, or in direct substitution for, a chemical functional group that is expected to transform the corresponding substrate. This strategy sacrifices transition state mimicry for charge density and it is particularly effective for general-acid/base or nucleophilic catalysts selection [Xu et al., 2004; Janda, 1990]. Since the haptens designed according to this strategy serve as "bait" for eliciting catalytic functions, which is then "switched" for the substrate, the strategy has been named "bait-and-switch".

2. *Screening against transition state analogues*.
   According to the transition state theory [Glasstone et al., 1941; Jencks, 1969] a catalyst enhances the reaction rate by means of stabilizing high-energy transition state structures as they are formed during the reaction. Accordingly, any protein that stabilizes the transition state is a potential candidate for that reaction. This strategy has been exploited to evolve catalytic antibodies that bind to transition state analogues (TSA) compounds for a number of reactions. A number of different TSAs have been designed to mimic the transition state and related high-energy intermediates with regard to fractional bond orders, lengths, angle, charges and overall geometry. To date, there are TSAs available for numerous reactions [Tanaka, 2004; Xu et al., 2004]

3. *Screening by reactive haptens.*
   Both TSA and bait-and-switch strategy rely on almost inert compounds, an alternative strategy is to select potential catalysts by reactive substrate that undergo a chemical reaction with residues

within the reaction site, thus covalently labelling the active peptides [Wirsching et al., 1995]. Following the reaction, labelled peptides can be recovered and selected out from the initial pool and used for subsequent cycles of amplification and selection.

Among the different selection procedures, screening against TSA perfectly suits the experimental requirements forasmuch as it allows screening the peptide library under a wide range of chemical conditions (most notably pH, see chapter 2.5). In fact, both the "bait-and-switch" and the "reactive haptens" strategies strongly rely on molecule net charge and therefore are severely affected by pH conditions.

Concerning the catalytic reaction, we focused on ester and amide hydrolysis due to their relevance in the prebiotic scenario. Ester hydrolysis might have played a major role in the peptide-assisted prebiotic synthesis of surfactants which led to the formation of the first cell-like compartments. Following the hydrolysis of ester precursors of fatty acid, the latter are released in the aqueous environment and spontaneously form closed compartments called vesicles [Luisi, 2004]. Accordingly, the emergence of ester-hydrolytic peptides might have triggered the emergence of a virtual cycle where peptides catalyse the formation of vesicles and vesicles entrap peptides preventing their dilution and degradation.

As concern amide bond hydrolysis, it spontaneously occurs in aqueous environment. However, the free energy balance of the hydrolysis reaction is modestly negative (2,4 Kcal/mol) [Stiger, 1988] and is mainly driven by the overwhelming abundance of one of the reactants (i.e. water). However, under anhydrous conditions or following the precipitation of the product the entire equilibrium can be shifted toward the synthesis as reported by Luisi and coworkers [Gorlero et al., 2008]. Within this framework, peptides capable to hydrolyse the amide bond may as well catalyse the reverse reaction, namely the synthesis of short peptides. Accordingly, amide-hydrolytic peptides might have triggered the emergence of primordial metabolism [Nakashima and Fox, 1980] or auto-catalytic peptides cycles [Kaufamann, 1996]. In addition, ester and amide hydrolysis involve the formation of the same high-energy tetrahedral intermediate [Tanaka, 2002] that subsequently decomposes to the corresponding acid and alcohol or amine depeding on the pK_a of the leaving group (Figure 1.9a).Accordingly, this high-energy intermediate can be mimicked by a phosphonates or phosphomamidates TSA for both reactions (Figure 1.9b).

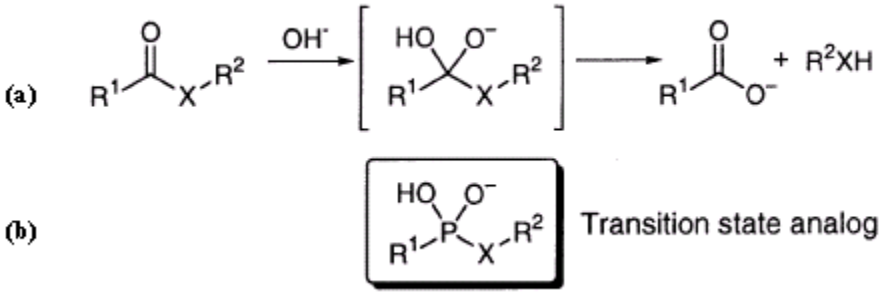


Figure 4
(a) Mechanism of ester and amide hydrolysis. (b) Corresponding phosphonate (X=O) or phosphonamidate (X=NH) Transition State Analogue. Adapted from Tanaka, 2002.

2. Results

The ultimate objective of this doctoral thesis is the screening of a completely de novo random library of 20 residues long peptides to discover functional molecules to assess the likeliness of the emergence of functional biopolymer under simulated prebiotic conditions. To tackle this question the phage display technique (see chapter 1.4.1) has been employed to screen a diverse library of random peptides for the biding to a transition state analogue (see chapter 1.4.2) for the ester and amide bond hydrolysis.

The likelihood of success in an in vitro evolution experiment is directly related to the total library complexity, degeneracy and quality and to the availability of a robust and reproducible selection method. The phage display technique and the screening against the transition state analogue have been widely employed, however both procedures suffer a number of limitations that must be overcame in order to increase the chances of finding functional peptides in a random library. Accordingly, the first part of this doctoral project has been devoted to optimise the experimental procedures and in particular the construction of a novel phagemid vector specifically designed to display short random sequences (see chapter 2.1), the optimization of phage production (see chapter 2.2) and the optimization of the coupling reaction of the transition state analogue to a solid surface (see chapter 2.3). The novel phagemid and the optimised techniques have been used to construct 2 DNA libraries encoding for random 20mer peptides (see chapter 2.4 and 2.5) and the subsequent screening of these libraries under different chemical conditions (see chapter 2.6).

2.1 Design and construction of a novel phagemid vector

M13 Phagemid vectors are widely employed to display peptides and short proteins (up to 200 residues) to enrich a random population of sequences with the desired properties such as binding to a specific target, enhanced thermostability and to redesign substrate specificity. M13 Phagemid vectors harbour a plasmid origin of replication (ColE1) and a phage origin of replication (OriM13) driving the in vivo synthesis of phagemid single strand genome to be encapsulated in the virion particles. In addition, M13 phagemid vectors encode for a truncated version of the minor capside protein lacking of Domain I encompassing amino acids from 1 to 198 that can be replaced by the exogenous sequence which is displayed as N-terminal fusion protein to minor coat protein (see chapter 1.4.1). Finally, M13 phagemid vectors usually harbour an antibiotic gene cassette as selective marker.

In designing a vector for phage display, a series of considerations have to be taken into account; such as:

1. signal peptide sequence for the membrane translocation of the exogenous sequence to be displayed; there exist a number of different signal peptides sequences either derived from bacteria (pelB, ompA, StiI) or from phages. The first are well defined sequences with a precise cleavage point, whereas the latter are less defined and identify a consensus sequence that may vary.
2. the multiple cloning site MCS which as to allow the in-frame cloning of the exogenous sequence in between the leader peptide and the truncated minor coat protein. In addition, several MCS harbour either a "dummy" insert (e.g beta-lactamase) or a killer gene (e.g. barnase) to enhanced cloning efficiency and reduce vector self-ligation background.
3. peptides linker between the exogenous sequences and the truncated version of the minor coat protein to enhance the folding of the displayed sequence and minimise deleterious interaction of the latter with the truncated minor coat protein. For example, displayed elements 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 (e.g TEV, Thr, Xa proteases) between the displayed peptide/protein and the capside protein.
4. affinity tags for phage and fusion protein purification and detection. For example, 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), EQKLISEEDLN, which is recognized by MAb 9E10. Alternatively, poly-histidine (e.g. 6xHis) stretch can be placed either at the N- or C-terminal to favour purification by metal affinity chromatography.
5. stop codons between upstream the minor coat protein sequence to reduce the level of non-recombinant phages. To this end a stop codon is inserted in the N-terminal fragments of gene III. The stop codon should be tuneable in order to allow the selective suppression/expression of the full-length minor coat protein.

A number of commercial and custom-made phagemid vectors are available today, however none of these possesses every and all the features mentioned above. Accordingly a novel phagemid vector has been designed and constructed to full fill all the above requirements, in particular:

1. The pelB leader peptide has been chosen to ensure a reliable membrane translocation,

2. A multiple cloning site MCS has been designed within a "dummy" sequences harbouring 3 stop codons in the 3 different reading frame in order to allow a visual control during step and minimise the production of non-recombinant fusion protein in case of phagemid self-ligation,

3. 2 different peptides linker have been designed to insulate the exogenous sequence from the minor coat protein followed by a protease Xa cleavage site to allow the recovery of purified phage,

4. 2 different affinity tags (c-myc and 6xHis) have been engineered at the N- and C-terminal of the exogenous sequence to allow the purification of the recombinant phage by means of both immuno-affinity chromatography and metal affinity chromatography,

5. a TAG stop codons has been cloned between upstream the minor coat protein sequence to reduce the level of non-recombinant phages. To this end a TAG stop codon is inserted at the N-terminal region of gene III fragment. 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 no virus particles are generated.

The overall structure of the novel phagemid vector, named pIII dummy, is depicted below (Figure 2.1).

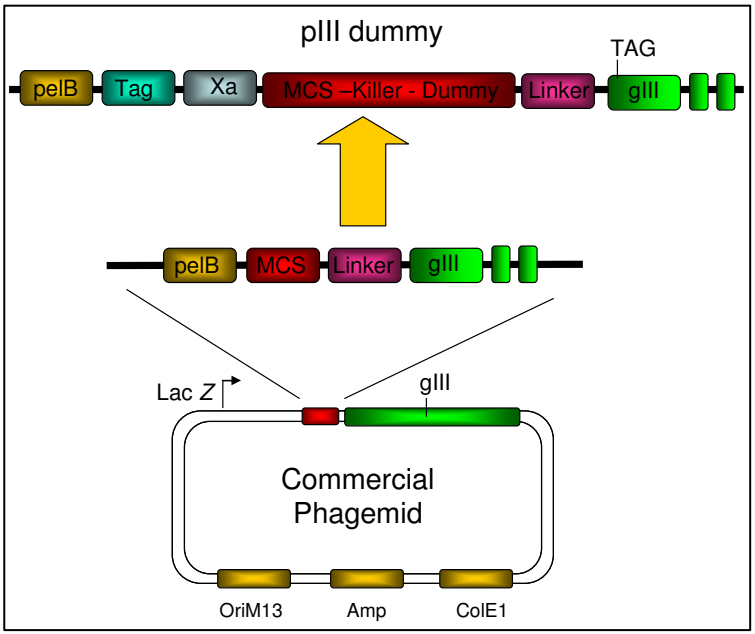


Figure 2.1
Structure of the pIII dummy phagemid vector representing the major features.

The novel phagemid vector has been constructed from pOCI 1050 (courtesy of Dr. Chiarabelli, University of Roma Tre). The synthesis of the DNA construct was performed by step-wise PCR. In the first step the LasR gene (courtesy of Dr. Leoni, University of Roma Tre) has been amplified using FW1 and RV primers (see Material and Methods, table 1). The FW1 primer introduces a NotI restriction site, a translational killer and a Xa site codifing DNA upstream the LasR gene. The RV primer introduces a XbaI restriction site. The PCR product will be further amplified by PCR using FW2 and RW primer couple. The FW2 primer will introduce either a c-myc or a hexahistidine tag and a NcoI restriction site at the 5' edge. An overview of the cloning procedure is given below (Figure 2.2).

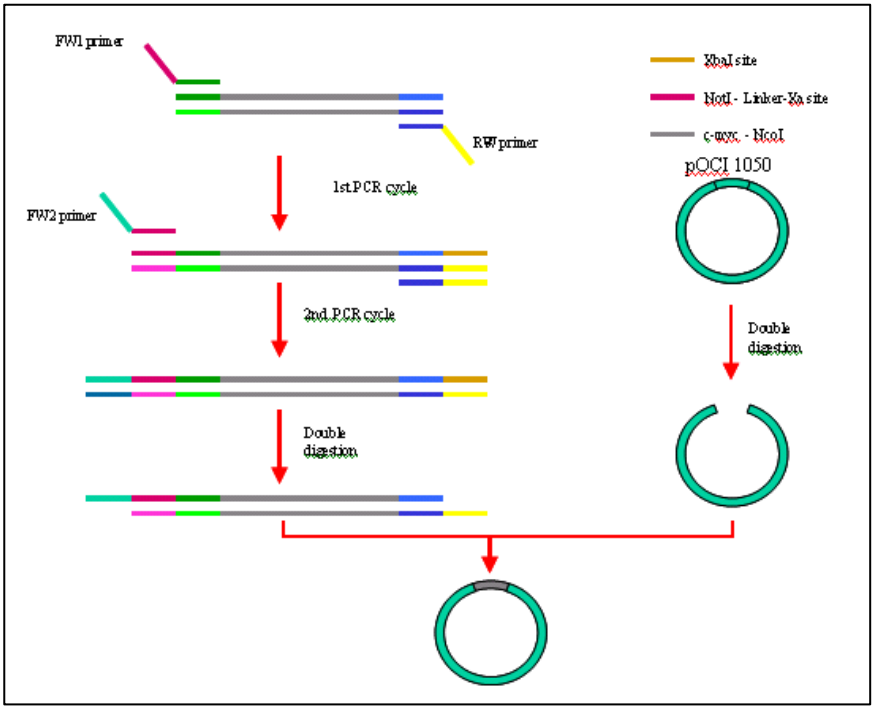


Figure 2.2
Construction strategy of the pIII dummy by step-wise PCR of the N-terminal region upstream gene III.

The novel phagemid has been tested for cloning efficiency and phage particle production for gene fragment of different size. Results show that the novel phagemid allows the efficient cloning of fragments as short as 60 residues

without detectable self-ligation (Figure 2.3) and represents a substantial improvement with respect to commercial phagemids.

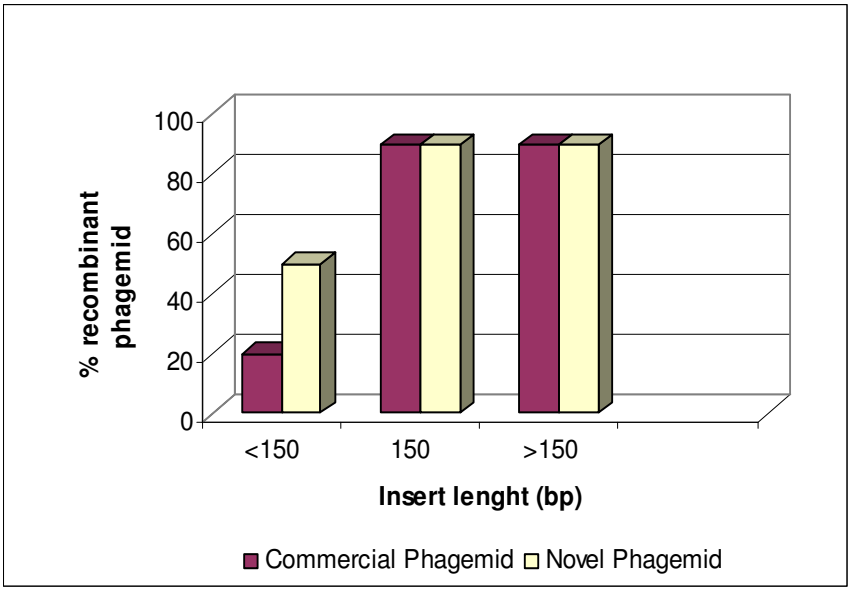


Figure 2.3
Cloning efficiency and background ligation plotted against exogenous gene length.

2.2 Optimization of phage display

There are different format for the Phage display that either exploiting wild-type viral genome of engineered phagemid DNA as described in chapter 1.4.1. In the *3 multivalent Phage display* the exogenous peptide or protein is expressed on the virion capside as N-terminal fusion protein to minor coat protein. In this format the corresponding encoding gene is cloned upstream the gIII into a wild-type viral vector. The result is that every copy of the capside protein displays the fusion producing a multivalent display. In the *3 + 3 monovalent Phage display* the protein fusion gene can be placed in a phagemid vector. The phagemid replicates in *E. coli* as a double-stranded plasmid, but co-infection with a helper phage results 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. 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 vary greatly and the ratio between fusion proteins and wild type pIII may range between 1:9 and 1:1000. In addition, a minor fraction of virion particle encapsulate the helper phage genome. To this regard, the *3 + 3 monovalent Phage display* produces a heterogeneous phage population as depicted in figure 2.4. Only the recombinant phage displaying the exogenous sequence and encapsulating the phagemid genome (Figure 2.4a) is suitable for in vitro evolution since it carries both the phenotype and the genotype. To this regard, it is of paramount importance to optimise the phage production in order to maximise the fraction of recombinant phage with respect to the overall population.

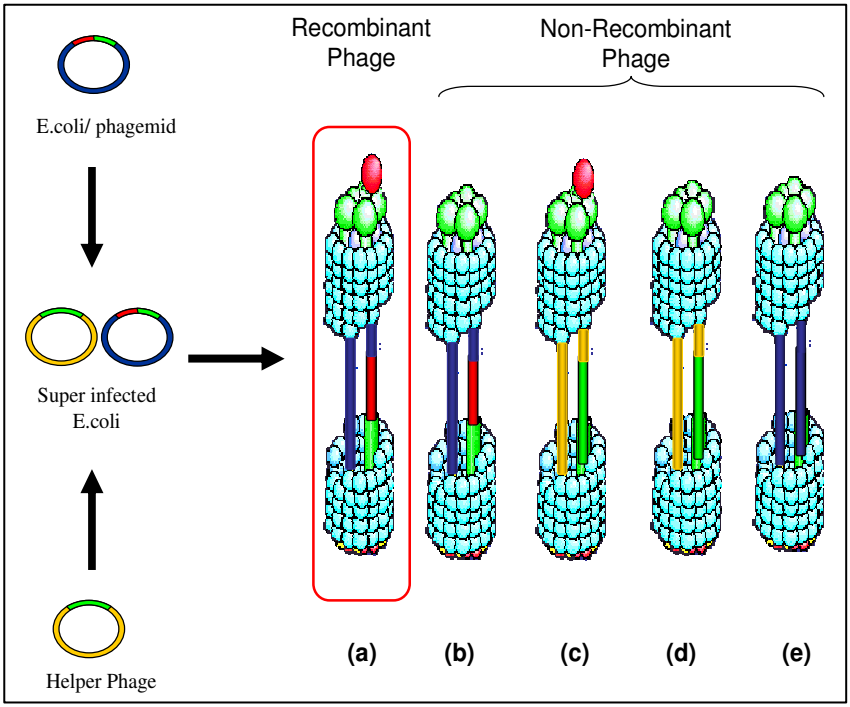


Figure 2.4
Overview of the phage production process. (a) recombinant phage with exogenous peptide (phenotype) and the corresponding encoding gene (genotype). (b) non-recobinant phage lacking the phenotype. (c) non-recobinant phage lacking the genotype. (d) non-recobinant phage lacking both phenotype and genotype, i.e. helper phage. (e) non-recobinant phage deriving from the self-ligation of the phagemid.

The recombinant to non-recombinant phage ratio is directly affected by the multiplicity of infection (MOI) and cell density during the superinfection of phagemid harbouring *E.coli* with the helper phage. In order to optimise the production of phage encapsulating the phagemid genome, a factorial design of experiments has been employing testing different conditions with regard to cell density and MOI employing test phagemid harbouring the avian pancreatic polypeptide (APP) gene (1PPT_A; GI:157833555). Phagemid genome (harbouring an ampicillin gene cassette) to helper phage genome (harbouring a kanamycine gene cassette) ratio has been evaluated by differential selective plating on ampicillin and kanamycine LB Agar respectively. Results show that the phagemid production is miximised when *E.coli* are infeceted during early lag-phase (OD_{600} = 0,2) with a MOI ranging between 1 and 0,5 (Figure 2.5) reducing the fraction of non-recombinant phage to 0,1% from 0,3% obtained with standard protocols.

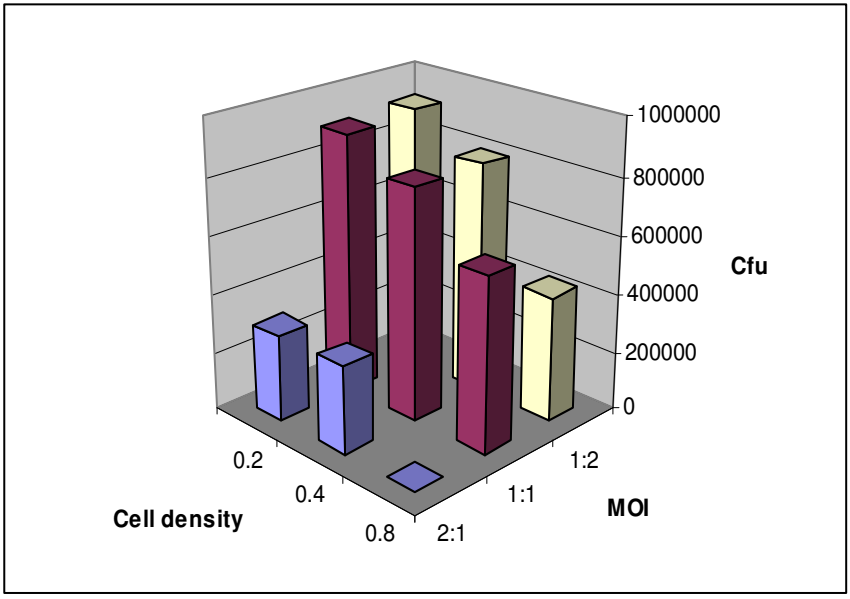


Figure 2.5
Recombinant phage production as dependent variable of cell density and multiplicity of infection.

Next, the purification of the phage particle has been optimised in order to maximise the recovery of phages displaying the exogenous sequences. To date,

phages are recovered by precipitation following flocculation triggered by NaCl and PEG. This method, although effective, does not allow the selective recovery of phages displaying the exogenous sequence. To circumvent this problem, the hexahistidine tag upstream the foreign sequence has been employed to recover recombinant phages by metal affinity precipitation employing super paramagnetic beads. Results show that phage displaying foreign peptides as N-terminal fusion protein to an hexahistidine tag can be recovered 10 times more effectively that non-tagged ones. However, test control of His-tagged phages spiked with c-myc tagged ones show an unspecific recovery as high as 30%.

2.3 Optimization of TSA coupling

The chosen selection criteria to discover functional peptide in a completely de novo random library of 20 residues long peptides relies on the capability of screened peptides to tightly bind a transition state analogue (TSA) for the hydrolysis of ester and amide bond under a broad range of chemical conditions (see chapter 1.4.2). The procedure implies the biopanning of the phage library onto a solid surface coated with the TSA. To this regard, the reproducibility of the coupling reaction between the functionalised solid surface and the TSA is of paramount importance to ensure reproducibility of the screening procedure. The coupling reaction occurs between the reactive acid group of the TSA (Figure 2.6) and the secondary amine of the solid surface.

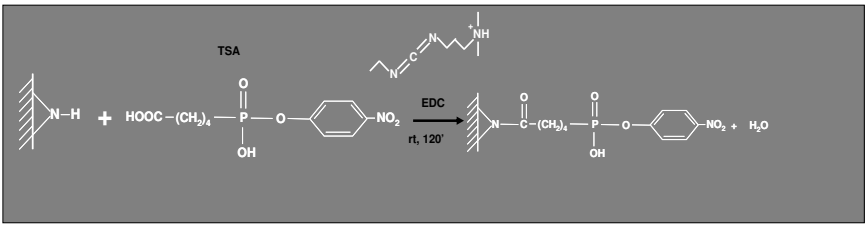


Figure 2.6
Semplified reaction mechanism for the coupling of TSA to amine-functionalised solid surface

The condensation reaction is promoted by dehydration reagent such as carbodiimide as illustrated in Figure .
The formation of an amide using a carbodiimide is straightforward, but with several side reactions. Reaction proceeds through O-acylisourea intermediate (2), which can be considered as a carboxylic ester with an activated leaving group. The O-acylisourea further reacts with amines to give the desired amide (3) and urea (4). The O-acylisourea can also react with an addition carboxylic acid (1) to give a acid anhydride (5), which can react further to give the desired amide (3). The main undesired reaction pathway involves the rearrangement of the O-acylisourea to the stable N-acylurea (6). In order to minimise side reactions, the use of solvents with low-dielectric constants (e.g. dichloromethane, DMSO or chloroform) and additives (eg. N-hydroxysuccinimide and sulfo- N-hydroxysuccinimide) have been investigated. In particular, the use of sulfo-NHS is thought to prevent the formation of the stable N-acylurea and to increase coupling efficiency creating a stable amine-reactive product (Figure 2.8).

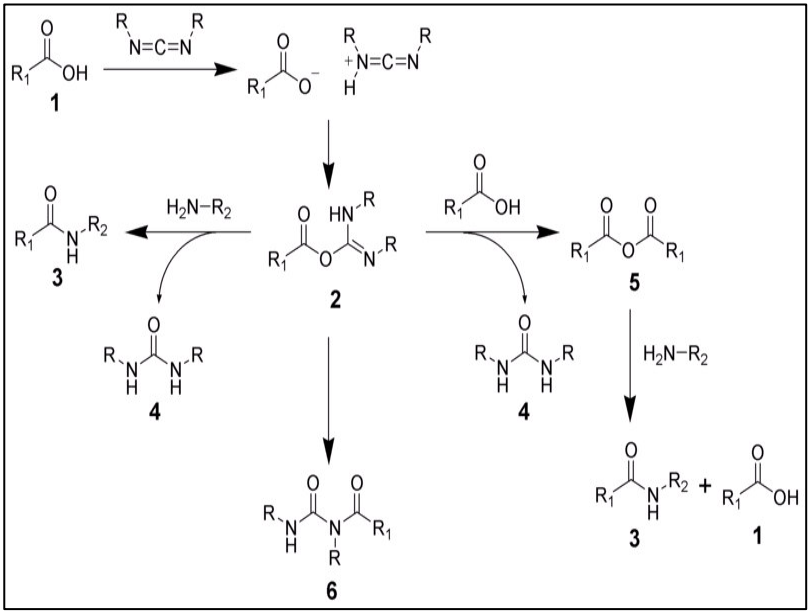


Figure 2.7
Reaction mechanism illustrating intermediates and side-reactions.

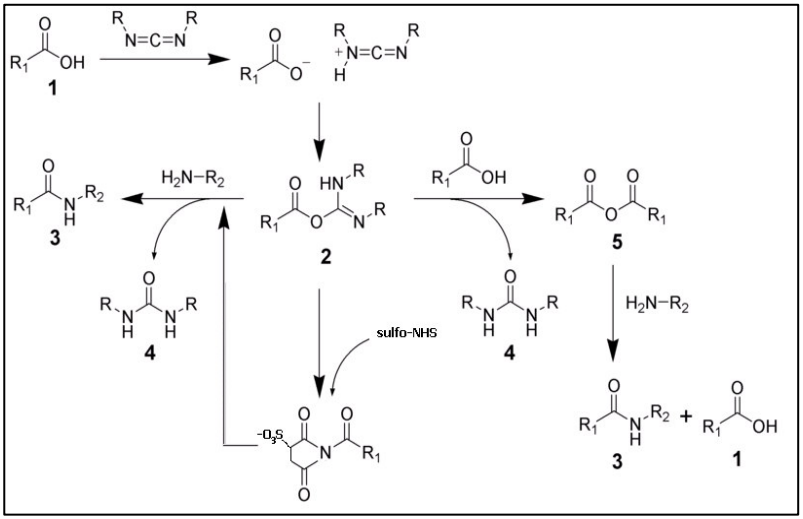


Figure 2.8
Reaction mechanism of optimised condensation reaction.

The efficiency of the optimised reaction has been evaluated by differential coupling, namely the same surface has been coupled to TSA first and biotin subsequently and compared to the biotin-reacted surface. Read-out has been performed by time resolved fluorescence measurements employing Europium labelled streptavidin.
Results indicate that the reaction occur almost to saturation as shown by comparing the read-out of the biotin-reacted surface to TSA-reacted one (Figure 2.9).
In particular, TSA-reacted surface shows a signal close to the negative control (non-treated surface) indicating a reaction yield close to 100%.
Finally, the stability of the TSA has been tested against different pH conditions in order to evaluate the possibility to employ TSA-functionalised surface to screen the peptide library under different pH. pH-dependent hydrolysis of the TSA has been measured by absorbance of the released p-nitro phenolate group (Figure 2.10).
By and large, these results indicate that the optimised procedure is suitable for the robust and reliable coupling of TSA to amine-functionalised surface and that TSA is stable under a wide range of pH up to 10.

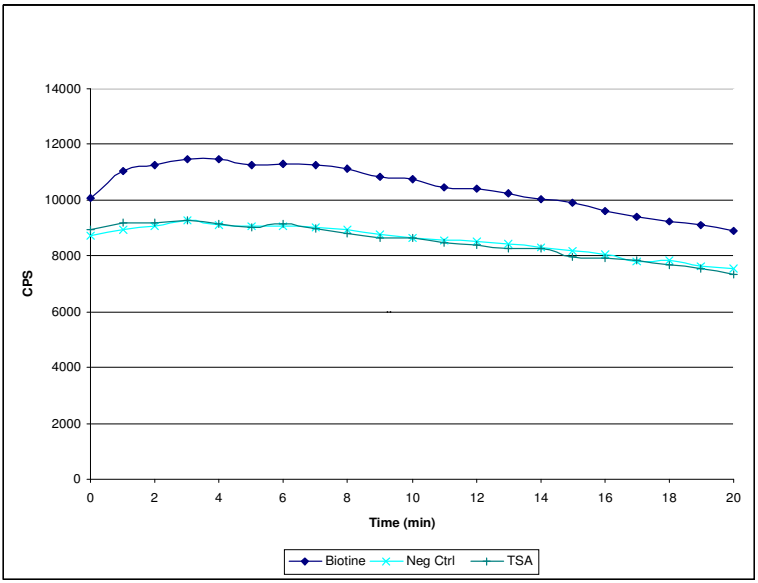


Figure 2.9
Count per seconds (CPS) plotted vs. time of differently treated surfaces.

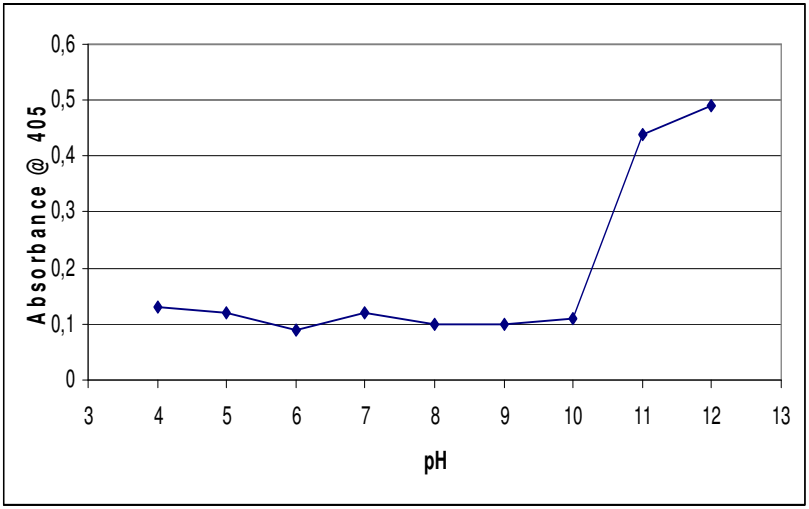


Figure 2.10
Absorbance at 405nm plotted vs. pH

2.4 Design and construction of a 3+3 monovalent phagemid library

A DNA library encoding for 20 residues long random peptides was cloned into the novel phagemid vector to be displayed as monovalent fusion protein to the minor coat protein. DNA library was synthesized using two groups of oligonucleotides (ProtLib20-NNK_FW and ProtLib20-NNK_RV) with the first encoding for the stretch of 20 random residues encoded by NNK schemes, where N is an equimolar mixture of all four bases and where K is either G or T. 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 forward oligonucleotide consist of random nucleotides encoding 20 amino acids followed by an PAAA linker and Xa protease cleavage site flanked by 30 fixed residues at the 5'-end necessary for amplification and 36 fixed residues for annealing with the reverse oligonucleotide. The latter oligo consists in 82 nucleotide encoding for an hexahistidine stretch with an upstream annealing region and a downstream amplification region.

After annealing and incubation with klenow, the DNA library was cloned and amplified using CutEnhFW and CutEnhRV primers. Purified product was cloned in pIIIdummy vector using the unique *Nco*I, *BamHI* sites (Figure 2.11).

The resulting phagemid library was transformed into electrocompetent *E.coli* cells to yield $1{,}5 \bullet 10^6$ transformants.

Ten colonies were individually growth and submitted to miniprep and sequenced to verify library cloning accuracy, randomness, degeneracy and any eventual homology to extant sequences.

Sequence analysis indicates that 7 clones out of 10 have properly received the DNA insert with a cloning accuracy of 70% (Table2.1).

Analysis of the DNA composition is shown in figure 2.13a, a remarkable agreement between expected randomness and experimental is found on the second and third position of each codon; whereas a substantial over representation of G is found in the first position. Sequence alignment indicates that the DNA library is characterised by high degeneracy and is isotropically distributed in the sequence space (Figure 2.13b). Randomness at the amino acid level -and corresponding distribution in sequence space- is shown in figure 2.14. Amino acid composition shows a significant over expression of glycine, valine and alanine and a substantial lack of arginine. Sequence alignment confirms the isotropic distribution of the sequences in the sequence space. Sequence BLAST did not reveal any significant homology with extant sequences neither at DNA level nor at amino acid level even employing permissive E-values.

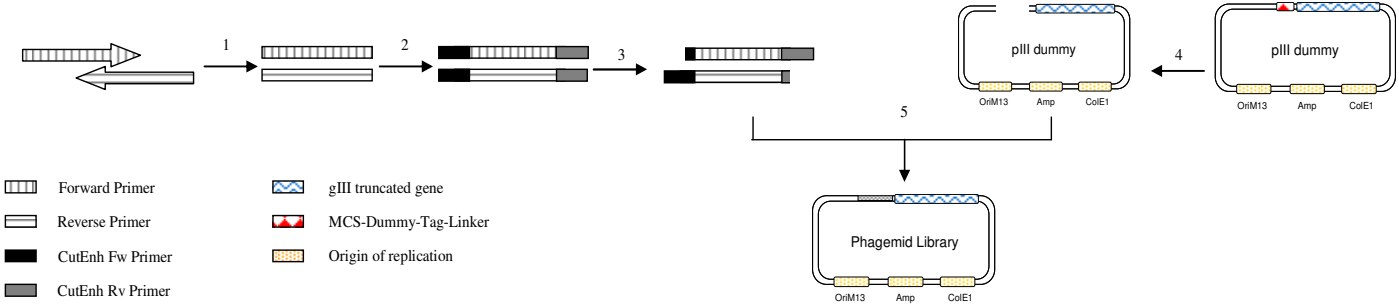


Figure 2.11
Overview of the cloning strategy and final phagemid library. (1) Fill-in reaction, (2) PCR, (3 and 4) Insert and vector double-digestion, (5) Ligation.

Table A

Exp24_S10	AGTGTTGAGCCTCTGTTGCCGACTCTGTTGCATTATCAGGTTTTTTAGTATGTGGGGGGG
Exp24_S2	AAGCATGCGGGGCGGAATGCTGGGAATGGGGCGCATTTGAGGAGTATTGTGTTTTATGTG
Exp24_S3	AAGGGGTCAGGAGAGGTCTACGCAGTGTGGGAGTTTGTATCTGATGACGAATCGTGGAAG
Exp24_S7	AAGCATGCGGGGCGGAATGCTGGGAATGGGGCGCATTTGAGGAGTATTGTGTTTTATGTG
Exp24_S4	GCTGATGATTATCGGGCGATGGTGCCGTTTTGGCTTAGTTAGTATGCTTTTGGTGCTGTG
Exp24_S9	CTTGACCGGGATGCTGCTATGGATGCGCCTGGTGTGCTGGGGAGTCAGAAGACGTTTGTG
Exp24_S6	GCGAAGGAGTCTGGGGTGTGGATTTTTGGTCTTCGGGGTCCGTGTTCGGTGTGGCAGTTT

Table B

Exp24_S10	SVEPLLPTLLHYQVF*YVGG
Exp24_S2	WGWADTRCGCVGMW*RCTRG
Exp24_S3	RGQERSTQCGSLYLMTNRGR
Exp24_S7	KHAGRNAGNGAHLRSIVFYV
Exp24_S4	ADDYRAMVPFWLS*YAFGAV
Exp24_S9	LDRDAAMDAPGVLGSQKTFV
Exp24_S6	AKESGVWIFGLRGPCSVWQF

Table 2.1. DNA sequences from the monovalent phagemid library (a) and corresponding amino acid translation (b).

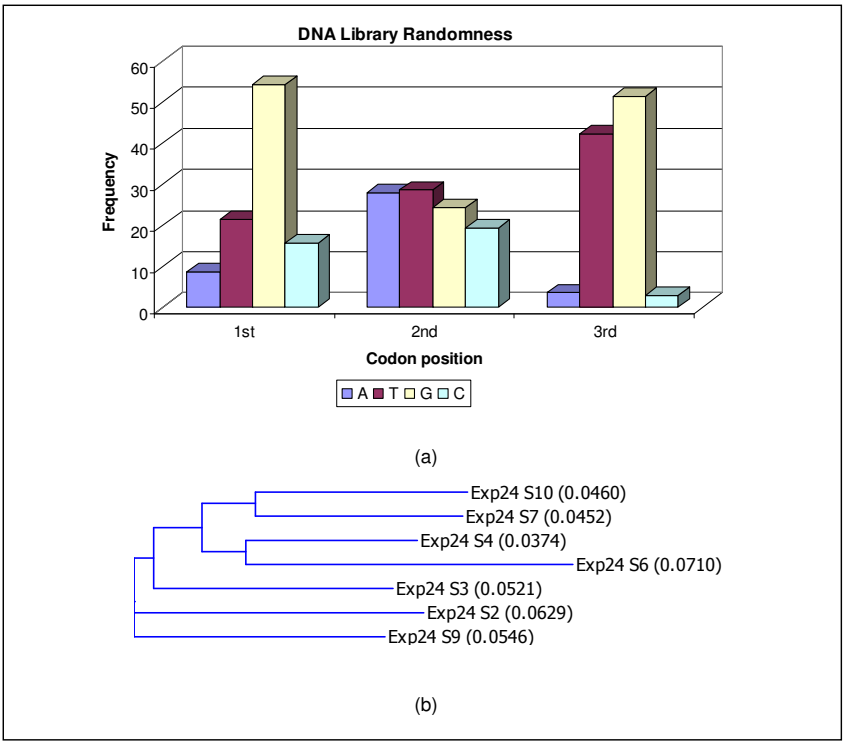


Figure 2.12
Analysis of DNA library randomness. (a) nucleotide randomeness. (b) Sequence alignment.

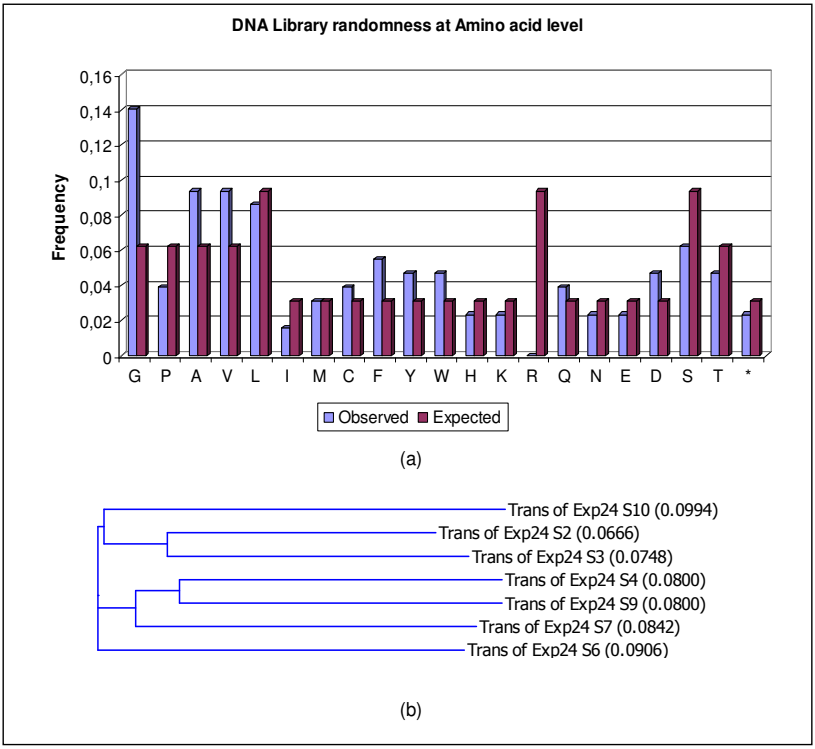


Figure 2.13
Analysis of DNA library randomness at amino acid level. (a) amino acid randomeness. (b) Sequence alignment.

2.5 Design and construction of a polyvalent phage library

A DNA library encoding for 20 residues long random peptides was cloned into a wild-type phage genome to be displayed as polyvalent fusion protein to the minor coat protein. DNA library was synthesized using two groups of oligonucleotides (M13ProtLib20-NNK_FW and M13ProtLib20-NNK_RV, the latter serves as template) with the first encoding for the stretch of 20 random residues encoded by NNK schemes. The forward oligonucleotide consist of random nucleotides encoding 20 amino acids followed by an GAAA linker and Xa protease cleavage site flanked by 30 fixed residues at the 5'-end necessary for amplification and 36 fixed residues for annealing with the reverse oligonucleotide. The latter oligo encodes for an hexahistidine stretch with an upstream annealing region and a downstream amplification region.

After annealing and incubation with klenow, the DNA library was cloned and amplified using ScLeft and GP3-ClaI primers. Purified product was cloned in M13m19phage_1 genome using the unique *KpnI* and *EagI* sites (Figure 2.14).

The resulting phagemid library was transformed into chemical competent *E.coli* cells to yield $2{,}1 \cdot 10^8$ transformants.

Twelve colonies were individually growth and submitted to miniprep and sequenced to verify library cloning accuracy, randomness, degeneracy and any eventual homology to extant sequences.

Sequence analysis indicates that 9 clones out of 10 have properly received the DNA insert with a cloning accuracy of 90% (Table 2.2).

Analysis of the DNA composition show a remarkable agreement between expected randomness and experimental one as shown in figure 2.15; whereas randomness at the amino acid level is shown in Figure 2.16. Results indicate that the DNA library is characterised by high degeneracy and is isotropically distributed in the sequence space (Figure 2.15b). The analysis of correspondent amino acid sequences indicate an overall agreement between expected amino acid frequencies and experimental ones with the exception of alanine (over represented) and cysteine and phenylalanine (under represented). The latter two being not statistically significant as tested by χ^2 hypothesis test. Sequence BLAST did not reveal any significant homology with extant sequences neither at DNA level nor at amino acid level even employing permissive E-values.

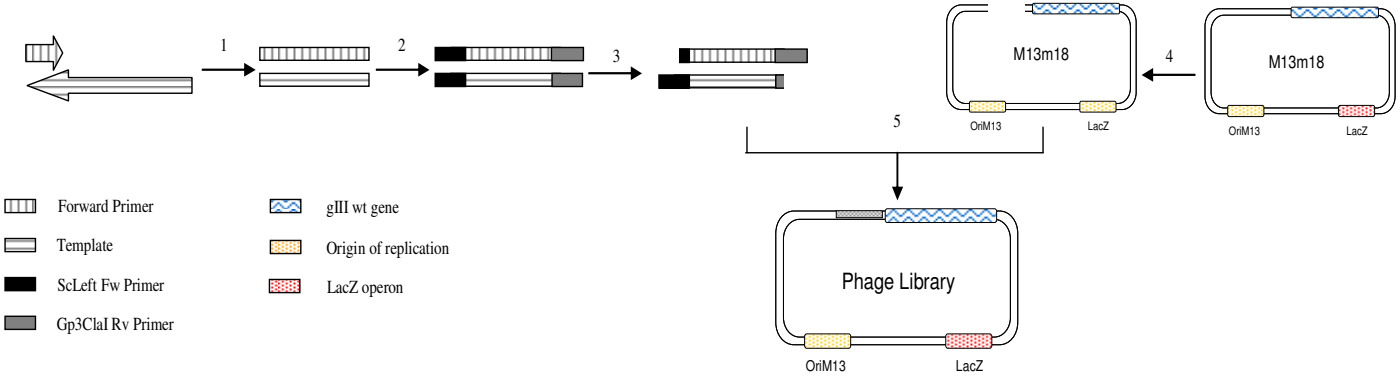


Figure 2.14
Overview of the cloning strategy and final phagemid library. (1) Fill-in reaction, (2) PCR, (3 and 4) Insert and vector double-digestion, (5) Ligation.

Table A

B0108_Exp4_0001	TGGTAGCATGATAGGATGCTAGCTACGATGTGTCGGCTTTATCTGGCTACGGATATTGAG
B0108_Exp4_0002	CATGCTAAGGACCAGGCTAGGCAGCGTAAGCTTGCTCTGGCTTCGATCGCTCGGTCGATG
B0108_Exp4_0003	GAGTGGCTAGCGGTTGGGCTTGCTAGTGTGTCGATTGGGGAGCGTATTGCTATGGCGATG
B0108_Exp4_0004	TCTTGGTAGGGTTATGGGTATGCGATTGCTCTATCGCATGCTACGCGATTGGAGGCGATT
B0108_Exp4_0005	ACTCAGCGTAAGGCGCATGCGGCGATGCTTGATGTTACGCGCATTTAGCGGCTTTATTAG
B0108_Exp4_0006	CAGAATAATAAGCTTAGTGCGCGTTGTCGGGCTTTGGGGGCTATTTGGGCTCATGATCCG
B0108_Exp4_0007	AGGCATGGTACGCATACGTCGCGGTCTGCAATGGTTACGGATCGGATGGTTGTGCCTGGG
B0108_Exp4_0008	GAGTAGAAGGTTGCTCATTTTACTAAGGTGAATCTTAGGACTCGGATGCCGGGGCTGGGT
B0108_Exp4_0010	TATTAGATGCACAGGCTTGGATTGGCTATTTCGCTGGGTCCGGATCGGCTTCCTAATGGG

Table B

B0108_Exp4_0001	W * H D R M L A T M C R L Y L A T D I E
B0108_Exp4_0002	M H A K D Q A R Q R K L A L A S I A R S
B0108_Exp4_0003	E W L A V G L A S V S I G E R I A M A M
B0108_Exp4_0004	S W * G Y G Y A I A L S H A T R L E A I
B0108_Exp4_0005	T Q R K A H A A M L D V T R I * R L Y *
B0108_Exp4_0006	Q N N K L S A R C R A L G A I W A H D P
B0108_Exp4_0007	R H G T H T S R S A M V T D R M V V P G
B0108_Exp4_0008	E * K V A H F T K V N L R T R M P G L G
B0108_Exp4_0010	Y * M H R L G L A I S L G P D R L P N G

Table 2.2

DNA sequences from the polyvalent phage library (a) DNA sequences and (b) corresponding amino acid translation.

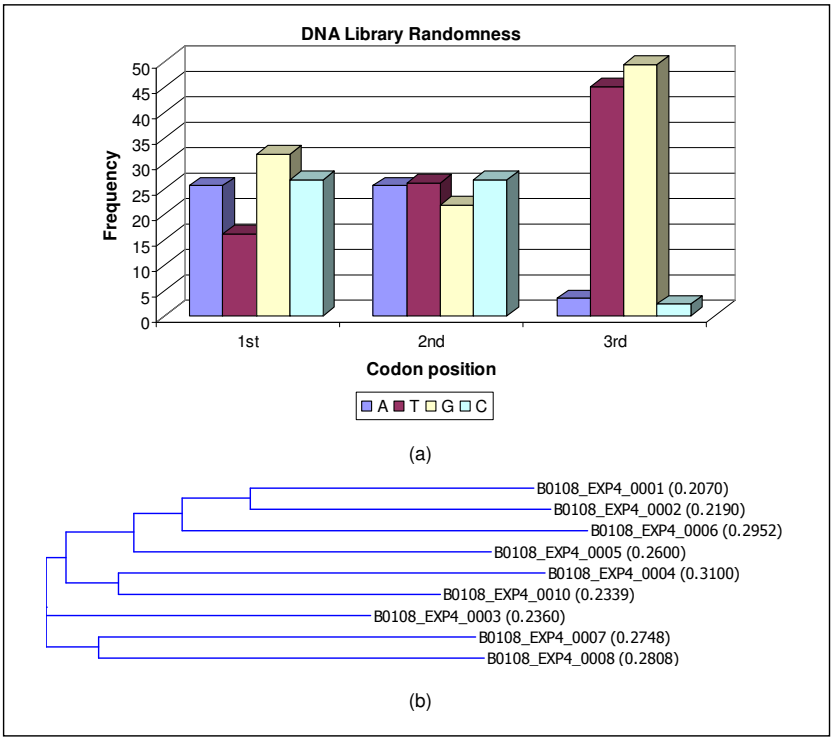


Figure 2.15
Analysis of DNA library randomness: nucleotide composition (a) and Sequence alignment (b).

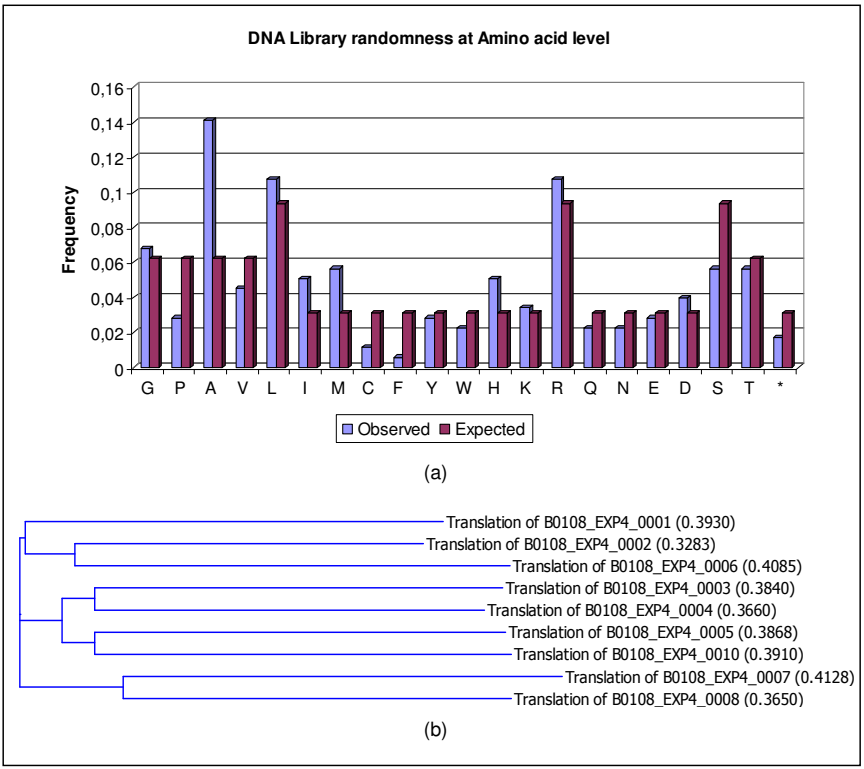


Figure 2.16
Analysis of DNA library randomness at amino acid level. (a) amino acid randomeness. (b) Sequence alignment.

2.6 Biopanning against transition state analogue

The 3+3 monovalent phage library described in chapter 2.2 was biopanned against the transition state analogue under different chemical conditions. In particular, the library was screened for binding to TSA under pH from 4 to 10 and zinc concentration from 0 to 10mM (Figure 2.17) under standard condition with appropriate buffer without neither salt nor detergent.

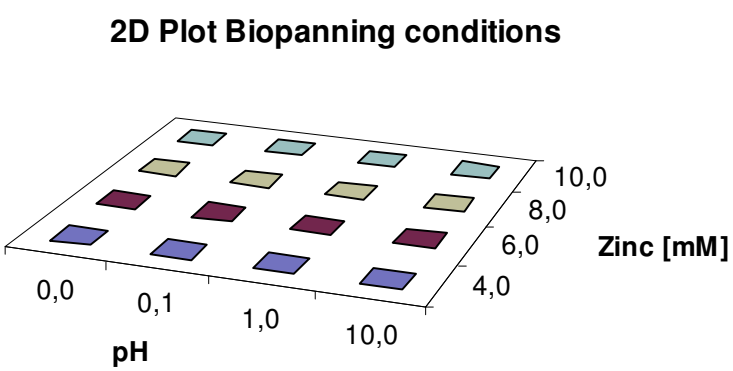


Figure 2.17
2D plot of the chemical conditions employed during phage biopanning against TSA of the monovalent phage library

The first round of biopanning round yields a bell-shape distribution with regard zinc concentration (Figure 2.18) and a complex distribution with regard pH with maximum recovery under acid pH conditions (Figure 2.19). A 3D plot is represented in Figure 2.20.

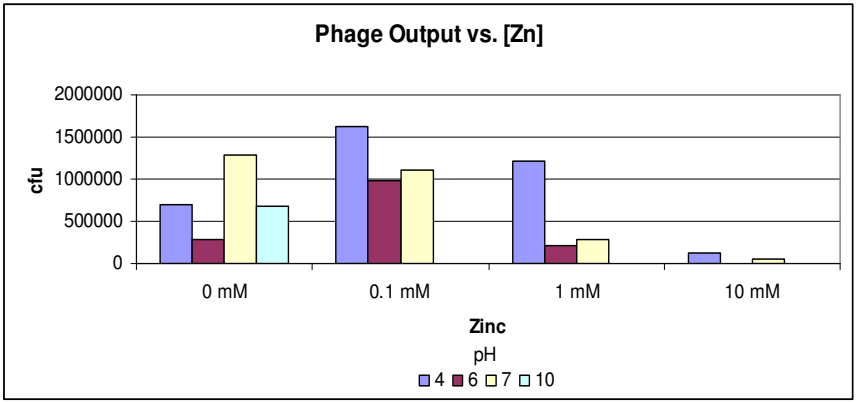


Figure 2.18
1D plot of phage output vs. [Zn] after the first biopanning round of the monovalent phage library.

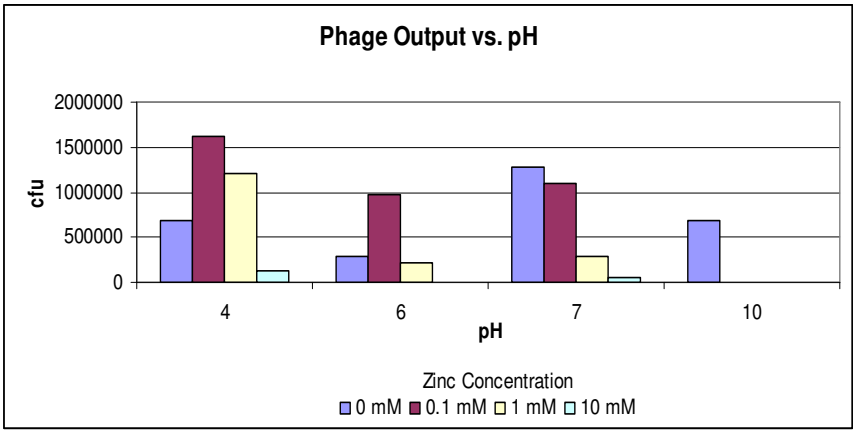


Figure 2.19
1D plot of phage output vs. pH after the first biopanning round of the monovalent phage library.

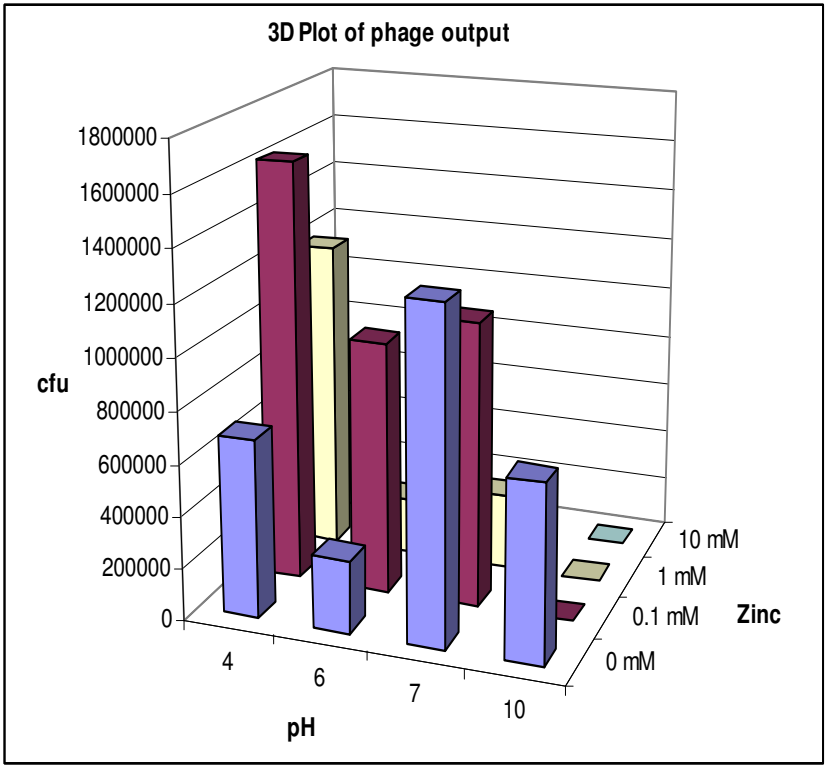


Figure 2.20
2D plot of phage output vs. pH and Zinc concentration after the first biopanning round of the monovalent phage library.

Results show that under basic pH even minute amount of zinc precipitate as white powder (most likely zinc hydroxide) so that this condition was no further employed for the subsequent biopanning rounds. In addition, the titer of phage output for the 3+3 monovalent phage library showed a remarkable inter- and intra-assay variability (CV> 10% on triplicate, data not shown). This could negatively affect the entire biopanning procedure and consequently this library was no further employed.
The polyvalent library described in chapter 2.3 was biopanned against TSA under different pH and zinc concentration as depicted in Figure 2.21

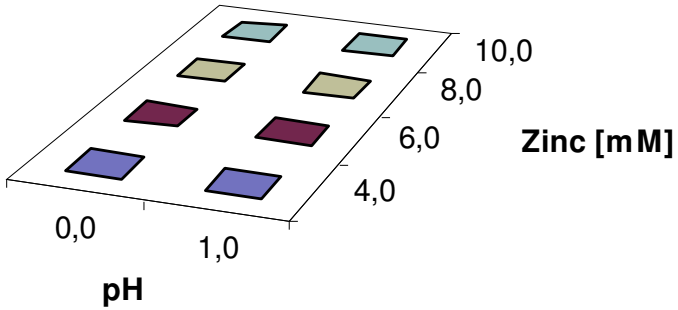


Figure 2.21
2D plot of the chemical conditions employed during phage biopanning against TSA of the polyvalent phage library

The first round of biopanning yields a decreasing monotone distribution with regard pH and zinc concentration with maximum recovery under acid pH 4 (Figure 2.22). Inter- and intra-assay variability showed acceptable results (CV < 5%, data not shown).

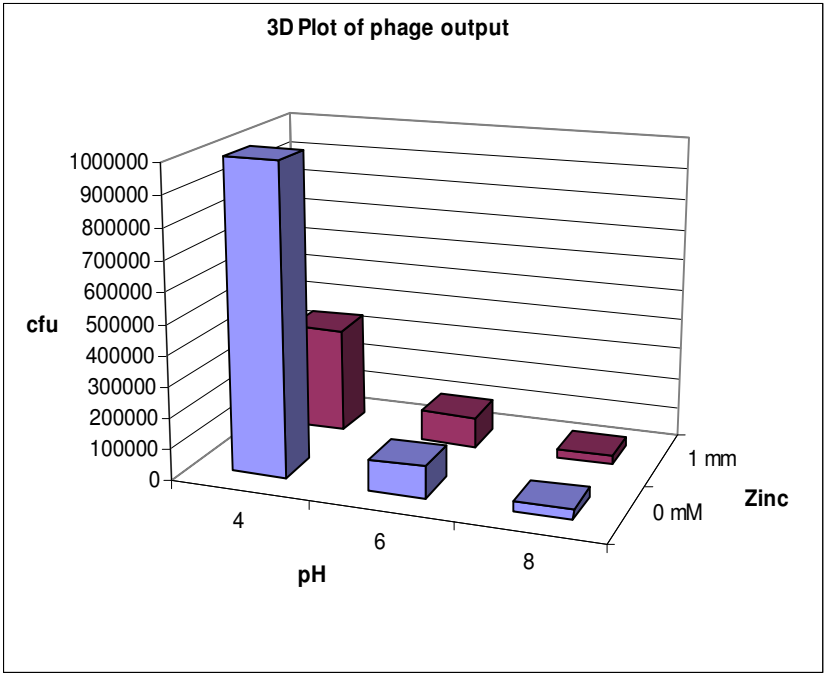


Figure 2.22
2D plot of phage output vs. pH and Zinc concentration after the first biopanning round of the polyvalent phage library.

Bound phages were recovered by competitive elution with free TSA and amplified by infecting early log-phase E.coli cells. Phage were precipitated and titer by serial dilution on selective medium. Purified phage library was further submitted to 4 biopanning rounds under increasingly stringent conditions with regard ionic strength and detergent concentration (Figure 2.25a and 2.25b) to minimise unspecific binding. In addition the incubation time and the elution time have been decreased during successive biopanning rounds in order to selectively recover peptides with low K_{on} and K_{off} (Table 2.3).

	Biopanning Rounds			
	1st	2nd	3rd	4th
NaCl (mM)	200	200	250	300
Tween 20 (%)	0	0	0,025	0,05
Incubation Time (min)	120	120	60	30
Elution Time (min)	180	180	120	60

Table 2.3
Selection condition during biopanning: (a) salt concentration, (b) detergent concentration, (c) Incubation time and elution time.

Figure 2.26 shows the phage output yield plotted as against biopanning rounds for each different chemical condition adopted. Each different chemical condition identifies a phage population which is represented with a different color according to the following table.

Population ID No.	pH	Zinc (mM)
1	4	0
2	4	1
3	6	0
4	6	1
5	8	0
6	8	1

Results show a maximum recovery at pH4 irrespective of zinc concentration. Phage output increases over biopanning rounds as the population is enriched of binding peptides. Most notably, phage output tends asymptotically to phage input only for phage biopanned at pH4, whereas phages biopanned under different conditions do not reach the expected maximum not even after 4 biopanning rounds. Finally, phages biopanned at pH 4 are only minimally affected by increased stringency conditions (round 3), whereas phages biopanned at different pH are strongly affected by increased salt and detergent concentration.

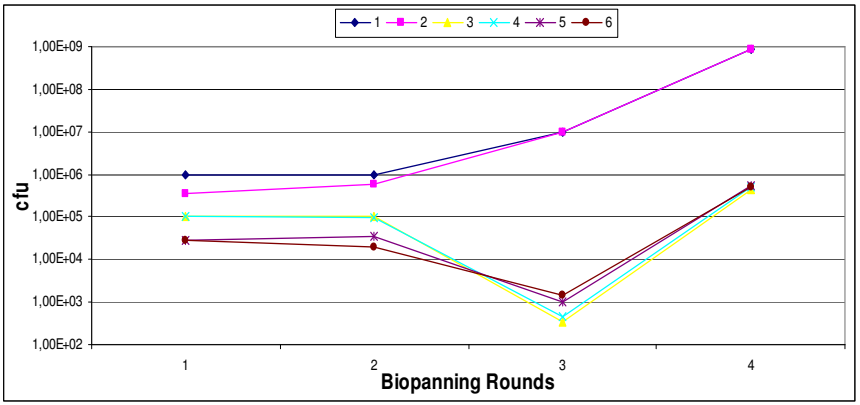


Figure 2.23
Phage output during selection process. For details refer to text.

Following the 4th biopanning round, 12 clones for each phage population biopanned at pH4 and 6 clones for each other phage population (48 clones total) were sequenced and analysed. The alignment analysis of the 48 clones revealed 2 major clusters (
Figure 2.24). The first includes the majority of clones selected at pH 4 (either with or without zinc), whereas the second includes the rest of the clones.

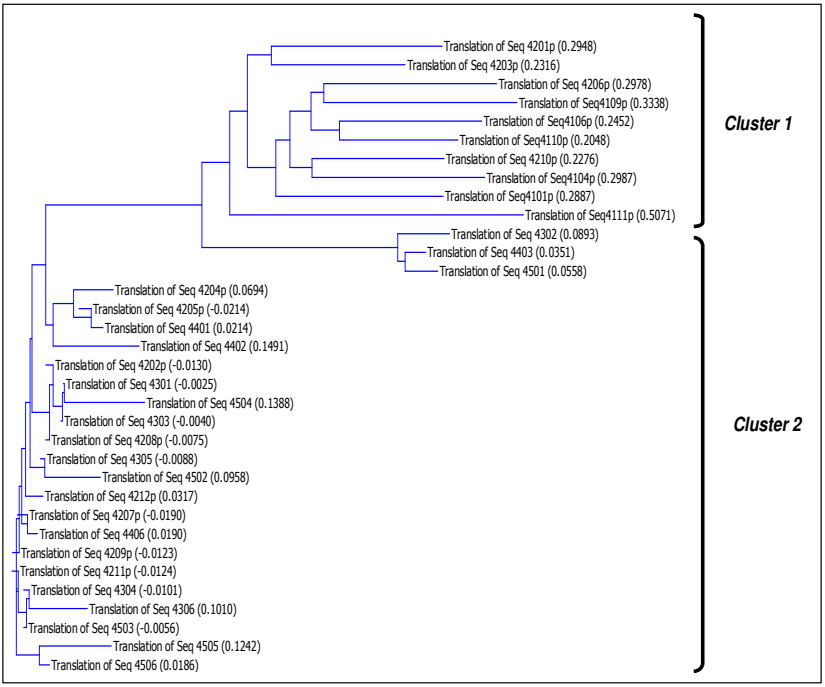


Figure 2.24

Sequence alignment of 34 out of 48 clones after the 4th round of biopanning and selected under different chemical conditions.

A detailed analysis of clones revealed that sequences belonging to *Cluster 2* are characterised by extensive deletions in the N-terminal region so that the foreign peptide is usually absent. Consequently, no further analyses have been carried out on these clones. Conversely, sequences belonging to *Cluster 1* show an intact N-terminal region and a consensus sequences characterised by a histidine-serine cluster at N-terminal and a poly-glycine at C-terminal. Figure 2.28 shows the in depth analysis of 12 clones randomly chosen within *Cluster 1* using the web WebLogo software which calculates the information content in bits for each residues from a multiple alignment analysis using the Shannon Entropy parameter [Crooks et al, 2004].

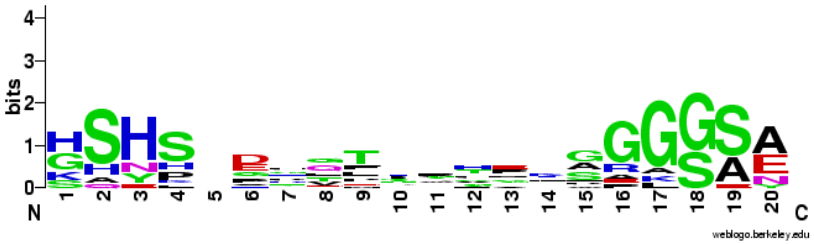


Figure 2.25
Sequence alignment of 12 clones after the 4th round of chosen within *Cluster 1*.

3. Discussion

The ultimate objective of this doctoral thesis is the screening of a completely de novo random library of 20 residues long peptides to discover functional molecules to assess the likeliness of the emergence of functional biopolymer under simulated prebiotic conditions. To tackle this question the phage display technique (see chapter 1.4.1) has been employed to screen a diverse library of random peptides for the biding to a transition state analogue (see chapter 1.4.2) for the ester and amide bond hydrolysis.

The first part of this doctoral project has been devoted to optimise the experimental procedures and in particular the construction of a novel phagemid vector specifically designed to display short random sequences (see chapter 2.1), the optimization of phage production (see chapter 2.2) and the optimization of the coupling reaction of the transition state analogue to a solid surface (see chapter 2.3). The novel phagemid and the optimised techniques have been used to construct 2 DNA libraries encoding for random 20mer peptides (see chapter 2.4 and 2.5) and the subsequent screening of these libraries under different chemical conditions (see chapter 2.6).

3.1 Design and construction of a novel phagemid vector

The likelihood of success in an in vitro evolution experiment is directly related to the total library complexity, degeneracy and quality and to the availability of a robust and reproducible selection method. The phage display technique and the screening against the transition state analogue have been widely employed, however both procedures suffer a number of limitations that must be overcame in order to increase the chances of finding functional peptides in a random library. The novel phagemid vector has been designed with a number of features that empower the effective and robust selection of short peptides, among which:

1. The presence of a leader peptide sequence *pel*B to ensure a reliable membrane translocation,
2. A multiple cloning site MCS has been designed within a “dummy” sequences harbouring 3 stop codons in the 3 different reading frame in order to allow a visual control during step and minimise the production of non-recombinant fusion protein in case of phagemid self-ligation,
3. 2 different peptides linker have been designed to insulate the exogenous sequence from the minor coat protein followed by a protease Xa cleavage site to allow the recovery of purified phage,
4. 2 different affinity tags (c-myc and 6xHis) have been engineered at the N- and C-terminal of the exogenous sequence to allow the purification

of the recombinant phage by means of both immuno-affinity chromatography and metal affinity chromatography,

5. a TAG stop codons has been cloned between upstream the minor coat protein sequence to reduce the level of non-recombinant phages. To this end a TAG stop codon is inserted at the N-terminal region of gene III fragment. The stop codon, TAG, can be suppressed efficiently in bacteria harbouring the *sup*E mutation which ensure that the stop codon is translated as glutamine (Q) when this vector is propagated in an appropriate bacterial strains. In addition, the propagation of the phagemid library in *sup*E^+ *E.coli* strain allows the use of degenerate libraries with NNK randomization scheme. This randomization scheme encodes for all amino acids and reduce the numner of stop codons to one (i.e. TAG) which is translated as Q.

The novel phagemid has been tested for cloning efficiency and phage particle production for gene fragment of different size. Results show that the novel phagemid allows the efficient cloning of fragments as short as 60 residues without detectable self-ligation and represents a substantial improvement with respect to commercial phagemids. To this regard, the novel phagemid vector represent a substantial improvement with respect to the state of the art and can be widely employed for the *in vitro* evolution of short peptides and proteins up to 200 amino acids.

3.2 Optimization of phage display

There are different format for the Phage display that either exploiting wild-type viral genome of engineered phagemid DNA as described in chapter 1.4.1. The *3 multivalent Phage display* format represents the most widely employed format, however it results in the pentavalent display of the exogenous peptide on the virion capside. As a consequence, the selection procedure tends to enrich the initial population of peptides with high avidity to the detriment of affinity. To this regard, the *3 + 3 monovalent Phage display* format still represents a valid alternative since it allows the selection of peptides and proteins with a high affinity. In this format the phagemid replicates in *E. coli* following a co-infection with a helper phage which provides all the proteins necessary for phage assembly, including wild-type copies of all the coat proteins. 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 vary greatly and the ratio between fusion proteins and wild type pIII may range between 1:9 and 1:1000. In addition, a minor fraction of virion particle encapsulate the helper phage genome. To this regard, the *3 + 3 monovalent Phage display* produces a heterogeneous phage population where only the recombinant phage displaying the exogenous sequence and encapsulating the phagemid genome is suitable for in vitro evolution since it carries both the phenotype and the genotype. In order to overcome this limitation the phage production process has been optimised in order to maximise the fraction of recombinant phage with respect to the overall population. The optimisation procedure relied on factorial design of experiments where the main variables investigated were the multiplicity of infection (MOI) and cell density during the superinfection of phagemid harbouring *E.coli* with the helper phage. Results show that the phagemid production is miximised when *E.coli* are infeceted during early lag-phase (OD_{600} = 0,2) with an MOI ranging between 1 and 0,5 reducing the fraction of non-recombinant phage to 0,1% from 0,3% obtained with standard protocols.

In addition, the presence of a peptide tag (i.e. 6xHis) has been exploited in order to maximised the recovery of phages displaying the exogenous sequences by metal affinity chromatography employing super paramagnetic beads charged with nickel. Results show that phage displaying foreign peptides as N-terminal fusion protein to a hexahistidine tag can be recovered 10 times more effectively that non-tagged ones. However, test control of His-tagged phages spiked with c-myc tagged ones show an unspecific recovery as high as 30%. Although promising, the metal affinity purification procedure proposed requires further optimization with regard to binding and washing conditions to minimise unspecific recovery.

The optimised procedure allows the production of more homogeneous phage library which directly affect the robustness of the phage display process. It is noteworthy that the optimised procedure has been developed with test peptides (i.e. APP) and employed with a library of completely random peptides, to this regard the optimised procedure is widely applicable and represents an improvement with respect to standards procedures reported in literature.

3.3 Optimization of TSA coupling

The chosen selection criteria to discover functional peptide relied on the capability of screened peptides to tightly bind a transition state analogue (TSA) for the hydrolysis of ester and amide bond under a broad range of chemical conditions (see chapter 1.4.2). The procedure implies the biopanning of the phage library onto a solid surface coated with the TSA. To this regard, the reproducibility of the coupling reaction between the functionalised solid surface and the TSA is of paramount importance to ensure reproducibility of the screening procedure. The condensation reaction is promoted by dehydration reagent such as carbodiimide and proceeds through O-acylisourea intermediate which further reacts with amines to give the desired products. Although straightforward, the reaction yields several by-products which severely impar the robusteness and reproducibility of the coupling reaction and uòtemetely the effectiveness of the biopanning procedure. In order to minimise side reactions, the use of solvents with low-dielectric constants (e.g. dichloromethane, DMSO or chloroform) and additives (eg. N-hydroxybenzotriazole, N-hydroxysuccinimide and sulfo- N-hydroxysuccinimide) have been investigated by factorial design of experiments. Results indicate that the optimised reaction occurs almost to saturation and show a consistent reproducibility with low inter- and intra-assay variability. By and large, these results indicate that the optimised procedure is suitable for the robust and reliable coupling of TSA to amine-functionalised surface.

Although the conditions employed cannot be extended to other reactions, since each reaction is deeply affected by the nature of the hapten to be coupled, the optimization procedure may represent a valid methodology to approach other coupling reaction. In particular, the detection method employed, based on competitive coupling detected by time-resolved fluorescence, may be extended to other coupling reaction and employed in the broad filed of chemical genetics and affinity proteomics.

3.4 Design and construction of degenerate DNA libraries

Two degenerate DNA libraries encoding for 20 residues long random peptides were deisgned and constructed using the NNK randomization schemes, where N is an equimolar mixture of all four bases and where K is either G or T. 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 (λ_{STOP} = 1/32 • 20 = 0,625 Stop codons per sequence). The randomization code employed is substantially different from the natural genetic code, however the normalised frequencies of each amino acid are consistent in the two different genetic codes with the exception of tryptophane and metionine which are over represented in the NNK genetic code and isoleucine and the stop codons which are under represented (Figure 3.1).

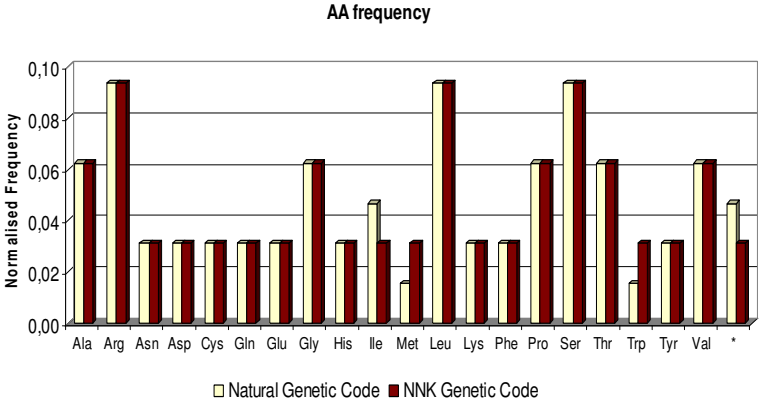


Figure 3.1
Normalised frequencies of amino acid in the natural (beige) and NNK genetic code (dark red).

The NNK genetic code encodes for $(4•4•2)^{20} = 10^{30}$ theoretically different codons for 20^{20} theoretically different sequences of 20 residues, the two library consist of $1,5 • 10^6$ and $2,1 • 10^8$ transformants respectively with a cloning efficiency of 70% and 90% respectively.
Analysis of the DNA composition showed a remarkable agreement between expected randomness and experimental one for both nucleic acid composition with the exception of a over representation of guanine in the first codon position in the first library.
Sequence alignment indicates that the DNA library is characterised by high degeneracy and is isotropically distributed in the sequence space.

With regard to amino acid composition and randomness, results show a significant over expression of glycine, valine and alanine and a substantial lack of arginine in the first library, whereas the second library show a remarkable agreement between expected randomness and experiential one with the exception of alanine (over represented) and cysteine and phenylalanine (under represented). The latter two being not statistically significant as tested by χ^2 hypothesis test. Both protein libraries are highly isotropic and evenly distribution in the sequence space. Sequence BLAST did not reveal any significant homology with extant sequences neither at DNA level nor at amino acid level even employing permissive E-values.
Although an experimental test of the uniqueness of the library is not possible, the degeneracy of the library can be inferred theoretically assuming an *a priori* probability for each sequence $p_i = 0{,}05^{20}$ and an expected number $\lambda = p_i \bullet N = 1{,}80 \bullet 10^{-18}$. Assuming a Poisson distribution [$P(k) = (e^{-\lambda} \bullet \lambda^k)/ k!$] it is possible to demonstrate that the probability to have more than one time a certain sequence within the same library is approximately zero:

$$\text{for } k=0 \Rightarrow P(0) = e^{-\lambda}$$
$$\text{for } k=1 \Rightarrow P(k=1) = e^{-\lambda} \bullet \lambda$$
$$\text{for } k>1 \Rightarrow P(k>1) = 1 - P(0) - P(1) = 1 - e^{-\lambda}(1+\lambda) \cong 0$$

By and large, DNA libraries show a substantial randomness without any significant bias; are highly degenerated and show no homology to extant sequences.
To this regard, DNA libraries can be considered as a collection of completely de novo random sequences that may mimicry the pool of sequences synthesised under proto- and prebiotic conditions.

3.5 Biopanning against transition state analogue

The 3+3 monovalent phage library described in chapter 2.2 was biopanned against the transition state analogue under different chemical conditions. In particular, the library was screened for binding to TSA under pH from 4 to 10 and zinc concentration from 0 to 10mM under standard condition with appropriate buffer without neither salt nor detergent. The first round of biopanning round yields a bell-shape distribution with regard zinc concentration and a complex distribution with regard pH with maximum recovery under acid pH conditions.

Results show that under basic pH even minute amount of zinc precipitate as white powder (most likely zinc hydroxide) so that this condition was no further employed for the subsequent biopanning rounds. In addition, the titer of phage output for the 3+3 monovalent phage library showed a remarkable inter- and intra-assay variability (CV> 10% on triplicate, data not shown). This could negatively affect the entire biopanning procedure and consequently this library was no further employed so that further biopanning rounds were performed only with the polyvalent library described in chapter 2.3.

Results showed a decreasing monotone distribution with regard pH and zinc concentration with maximum recovery under acid pH 4 with low inter- and intra-assay variability.

This result comes unexpected since most of extant enzymes show optimal catalysis -and substrate binding- at pH 7. This results cannot be explained by a mere experimental artefact as confirmed by sequence analysis which showed that sequences selected under condition different from pH 4 are characterised by extensive deletions in the N-terminal region so that the foreign peptide is usually absent. In addition, experiments performed on a completely de novo random library of 50 residues showed the same recovery profile with maximum recovery at pH 4 (unpublished data by Dr. Quintarelli, University of Roma Tre).

This result can be explained considering the high net negative charged of the hapten employed (a phosphonate derivate) during the biopanning process which are partially neutralised at low pH which, in turn, may favour binding. Alternatively, one may argue that low pH induces a positively net charge on the entire virion particle which may lead to unspecific binding of the M13 to the negatively charged hapten. This conjecture was not confirmed by control experiments with non-recombinant M13 (unpublished data by Dr. Quintarelli, University of Roma Tre). Accordingly, one may conclude that short peptides preferentially bind the transition state analogue at low pH, although this does not necessarily imply that catalysis occur optimally under the same conditions.

Another striking feature is that different population dynamic between phages selected at different pH. In general, phage output increases over biopanning rounds as the population is enriched of binding peptides irrespective of the

chemical conditions. However, phage output tends asymptotically to phage input only for phage biopanned at pH4, whereas phages biopanned under different conditions do not reach the expected maximum not even after 4 biopanning rounds. This result suggests that selection at pH different than 4 do not allow optimal recovery not even after successive biopanning rounds, namely there are conditions under which selective enrichment (i.e. simulate evolution) is prohibited. It is noteworthy that phages biopanned at pH 4 are only minimally affected by increased stringency conditions, whereas phages biopanned at different pH are strongly affected by increased salt and detergent concentration. To this regard, one may argue that the phage population selected under "extreme conditions" (i.e. low pH) reaches an optimal state within few selection rounds and are characterised by a certain degree of robustness (i.e. evolutive resilience) against chemical changes in the selection condition (i.e environmental perturbations).

A detailed sequence analysis has been performed on sequences recovered under pH 4. Results show an intact N-terminal region and a consensus sequences characterised by a histidine-serine cluster at N-terminal and a poly-glycine at C-terminal. It is noteworthy that acid residues are consistently conserved downstream the histidine-serine cluster and that no other sequence pattern can be found. This sequence pattern recalls the catalytic triad (Ser-His-Asp) of serine-histidine proteases such as Chymotrypsin and Subtilisin. This family of enzymes employs an acid-basid catalysis which proceeds through an acyl-enzyme intermediate stabilised by an oxyanion pocket. Although appealing, this reaction mechanism is unlikely in short peptides such those investigated in this doctoral project due to the impossibility to form a stable oxyanion pocket.

The histidine-serine cluster found at the N-terminal region of selected peptides also recalls metalloproteases such as Thermolysin and Carboxypeptidase A which contains a zinc ion in the active site coordinated by histidine residues which polarise a water molecule that acts as nucleophilic group. However this reaction mechanism is not consistent with the observation that zinc concentration seems uninfluential for the selection process.

By and large, results show that is possible to select functional peptide capable of binding a transition state analogue for the amide and ester bond hydrolysis reactions from a completely de novo random library of unstructured peptides characterized by a clear consensus sequence characterised by a histidine-serine cluster followed by an acid residues at the N-terminal region and a glycine-rich cluster at the C-terminal region.

4. Conclusions

Science assumes that life on Earth originated from inanimate matter by a gradual and spontaneous increase of molecular complexity. The assumption that life derives from inanimate matter bears a profound enigma about the very nature of evolution which is known as the contingency versus determinism debate. 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 determines the outcome of a given event.

The controversy between determinism and contingency emerges forcefully when one considers the emergence of functional biopolymers in the framework of the origin of life. Indeed, the entire architecture of life relies on biopolymers so that the aetiology of biopolymers represents a major issue in the field of origin of life research. Within this framework, one of the most interesting questions in modern life science is how prebiotic evolution of the first biopolymers occurred. 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 grows exponentially as the length of the polymer increases. For this reason, Nature could not perform an exhaustive search in sequence space for functional biopolymers.

This observation poses a number of intriguing questions: which is the fraction of all possible sequences displays a catalytic function? Are those sequences isotropically distributed in the sequence space? Do functional peptides resemble extant enzyme sequences?

To tackle this question, a completely de novo random library of short peptides has been designed and tested for potential catalytic activity. The 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 criterion was based on the ability to bind a transition state analogue of the ester and amide bond hydrolysis in order to asses the frequency and distribution in sequence space of functional peptides.

Results proved that is indeed possible to selectively recover peptides from a completely random library capable to recognise and bind the transition state analogue. Surprisingly, peptides preferentially bind the transition state analogue at low pH, although this does not necessarily imply that catalysis occur optimally under the same conditions. In addition, results suggest that there are conditions under which selective enrichment (i.e. simulate evolution) is prohibited. It is noteworthy that peptides selected under extreme conditions (i.e. low pH) are unaffected by high salt and detergent concentration displaying a

surprisingly degree of robustness (i.e. evolutive resilience) against chemical changes in the selection condition (i.e environmental perturbations).
Sequence analysis showed a histidine-serine cluster at N-terminal followed by an acid residues and a poly-glycine at C-terminal. This sequence pattern recalls the features of extant protease such as serine proteases and metallo-proteases. The presence of a conserved pattern similar to extant enzyme in an otherwise random sequence suggests that catalytic function may be found in sequences completely unrelated to existing ones and thus support the contingency theory. In addition, the possibility to identify functional peptides in a relatively small random library of 10^8 different sequences suggests that functional peptides are more common and evenly distributed in the sequence space that previously thought. These results must be interpreted within the framework of a broader research project [Luisi, 2003] that includes the investigation of the structural and functional activity of random biopolymers including longer proteins and RNA molecules. Previous works have demonstrated that random protein of 50 amino acid display a surprisingly high frequency of folding [Chiarabelli et al., 2006; Chiarabelli and De Lucrezia, 2007 and Minervini et al., submitted] even in absence of any evolutive pressure. Likewise, random RNA sequences of 150 nucleotides -a mirror image of the sequence space defined by the 50mer library- are characterised by a compact fold stable up to 60°C [De Lucrezia et al., 2006].
The result of this doctoral thesis together with this body of works suggests that random biopolymers are embedded with the same structural and functional potential of extant ones, thus supporting the contingency theory to explain the emergence of functional proteins and RNAs during the first steps of evolution. As a corollary, one may argue that sequence space is densely populated of functional biopolymers so that emergence of functional peptides or proteins during early evolution was an event more probable that commonly thought.

5. References

Agterberg, M.; Adriaanse, H.; van Bruggen, A.; Karperien, M. & Tommassen, J.
Outer-membrane PhoE protein of Escherichia coli K-12 as an exposure vector: possibilities and limitations.
Gene, **1990***, 88*, 37-45

Barbas, C. F.
Recent advances in phage display.
Curr Opin Biotechnol, **1993***, 4*, 526-530

Baumann, W.; Bizzozero, S. & Dutler, H.
Specificity of alpha-chymotrypsin. Dipeptide substrates.
FEBS Lett, **1970***, 8*, 257-260

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

Berkessel & Riedl
Combinatorial de novo synthesis of catalysts: how much of a hit-structure is needed for activity?
J Comb Chem, **2000***, 2*, 215-219

Berkessel, A. & Herault, D.
Discovery of peptide-zirconium complex that mediate phosphate hydrolysis by bacth screening of combinatorial undecapeptide library
Angew Chem Int Ed Engl, **1999***, 38*, 102-105

Better, M.; Chang, C. P.; Robinson, R. R. & Horwitz, A. H.
Escherichia coli secretion of an active chimeric antibody fragment.
Science, **1988***, 240*, 1041-1043

Boder, E. T. & Wittrup, K. D.
Yeast surface display for directed evolution of protein expression, affinity, and stability.
Methods Enzymol, **2000***, 328*, 430-444

Boder, E. T. & Wittrup, K. D.
Optimal screening of surface-displayed polypeptide libraries.

Biotechnol Prog, **1998***, 14*, 55-62

Brooke, S. & Fox, S. W.
Compartmentalization in proteinoid microspheres.
Biosystems, **1977***, 9*, 1-22

Burton, D. R.
Phage display.
Immunotechnology, **1995***, 1*, 87-94

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

Cech, T.R.
in The RNA World (eds Gesteland, R.F. & Atkins, J.F.) pp 239-269,
Cold Spring Harbor Laboratory Press, Plainview, NY. **1993**

Chiarabelli, C. & DeLucrezia, D.
Question 3: the worlds of the prebiotic and never born proteins.
Orig Life Evol Biosph, **2007***, 37*, 357-361

Chiarabelli, C.; Vrijbloed, J. W.; Lucrezia, D. D.; Thomas, R. M.; Stano, P.; Polticelli, F.; Ottone, T.; Papa, E. & Luisi, P. L.
Investigation of de novo totally random biosequences, Part II: On the folding frequency in a totally random library of de novo proteins obtained by phage display.
Chem Biodivers, **2006***, 3*, 840-859

Copeland, G.; Jarvo, E. & Miller, S.
Minimal Acylase-Like Peptides. Conformational Control of Absolute Stereospecificity.
J Org Chem, **1998***, 63*, 6784-6785

Corey, M. J. & Corey, E.
On the failure of de novo-designed peptides as biocatalysts.
Proc Natl Acad Sci U S A, **1996***, 93*, 11428-11434

Cropp, T. A.; Anderson, J. C. & Chin, J. W.
Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells.
Nat Protoc, **2007**, 2, 2590-2600

DeLucrezia, D.; Anella, F. & Chiarabelli, C.
Question 5: on the chemical reality of the RNA world.
Orig Life Evol Biosph, **2007***, 37*, 379-385

DeLucrezia, D.; Franchi, M.; Chiarabelli, C.; Gallori, E. & Luisi, P. L.
Investigation of de novo totally random biosequences, Part IV: Folding Properties of de novo, totally random RNAs.
Chem Biodivers, , **2006***, 3*, 869-877

DeMuynck, H.; Madder; Farcy; PJ, D. C.; Pérez-Payán; Öhberg & Davis
Application of Combinatorial Procedures in the Search for Serine-Protease-Like Activity with Focus on the Acyl Transfer Step.
Angew Chem Int Ed Engl, **2000***, 39*, 145-148

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, M. A.; Flint, J. E.; Lewis, R. J.; Lakey, J. H.; Morland, C.; Luginbühl, P.; Healey, S.; Todaro, T.; DeSantis, G.; Sun, M.; Parra-Gessert, L.; Tan, X.; Weiner, D. P. & Gilbert, H. J.
Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure.
J Biol Chem, **2008***, 283*, 22557-22564

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

Duve, C. D.
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 cells.
J Mol Biol, **1995***, 250*, 496-506

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

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, J.
Schopf, J. *(ed.)*
From building blocks to the polymers of life. In Life's Origin
California University Press, **2002**

Fersht, A.
Enzyme structure and mechanism
Freeman, New York, **1977**

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

Freudl, R.; MacIntyre, S.; Degen, M. & Henning, U.
Cell surface exposure of the outer membrane protein OmpA of Escherichia coli K-12.
J Mol Biol, **1986***, 188*, 491-494

Frost, L.
Conjugative pili and pilus-specific phages. In Bacterial Conjugation.
D. B. Clewell, **1993**

Fuchs, P.; Breitling, F.; Dübel, S.; Seehaus, T. & Little, M.
Targeting recombinant antibodies to the surface of Escherichia coli: fusion to a peptidoglycan associated lipoprotein.
Biotechnology, **1991***, 9*, 1369-1372

Glasstone; Laidler & Eyring
The Theory of Rate Processes
McGraw-Hill, New York, **1941**

Gorlero, M.; Wieczorek, R.; Adamala, K.; Giorgi, A.; Schininà, M. E.; Stano, P. & Luisi, P. L.
Ser-His catalyses the formation of peptides and PNAs.
FEBS Lett, **2009***, 583*, 153-156

Griffiths, A. D.
Production of human antibodies using bacteriophage.

Curr Opin Immunol, **1993***, 5*, 263-267

Haidane, J.
The origin of life.
New Biology, **1954***, 16*, 12-27

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

Hanes, J.; Jermutus, L.; Weber-Bornhauser, S.; Bosshard, H. R. & Plückthun, A.
Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries.
Proc Natl Acad Sci U S A, **1998***, 95*, 14130-14135

Hill & Patersen
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 U S A, **2006***, 103*, 9637-9642

Hoess, R. H.
Protein design and phage display.
Chem Rev, **2001***, 101*, 3205-3218

Hoogenboom, H. R.; de Bruïne, A. P.; Hufton, S. E.; Hoet, R. M.; Arends, J. W. & Roovers, R. C.
Antibody phage display technology and its applications.
Immunotechnology, **1998***, 4*, 1-20

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

Janda, K. D.
New strategies for the design of catalytic antibodies.
Biotechnol Prog, **1990***, 6*, 178-181

Jencks, W.
Catalysis in chemistry and enzymology

McGraw-Hill, New York, **1969**

Jencks, W. P.
Binding energy, specificity, and enzymic catalysis: the circe effect.
Adv Enzymol Relat Areas Mol Biol, **1975***, 43*, 219-410

Johnsson, K.; Allemann, R. K.; Widmer, H. & Benner, S. A.
Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides.
Nature, **1993***, 365*, 530-532

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

Kieke, M. C.; Cho, B. K.; Boder, E. T.; Kranz, D. M. & Wittrup, K. D.
Isolation of anti-T cell receptor scFv mutants by yeast surface display.
Protein Eng, **1997***, 10*, 1303-1310

Kotz, J. D.; Bond, C. J. & Cochran, A. G.
Phage-display as a tool for quantifying protein stability determinants.
Eur J Biochem, **2004***, 271*, 1623-1629

Ladner, R. C.
Phage display and pharmacogenomics.
Pharmacogenomics, **2000***, 1*, 199-202

Ladner, R. C.
Constrained peptides as binding entities.
Trends Biotechnol, **1995***, 13*, 426-430

Leman, L.; Orgel, L. & Ghadiri, M. R.
Carbonyl sulfide-mediated prebiotic formation of peptides.
Science, **2004***, 306*, 283-286

Levinthal, C.
De Brunner P, T. J. & E, M. *(ed.)*
Mossbauer spectroscopy in biological system
Univ. of Illinis Press, Urbana IL, **1969**, 22-24

Lim, C. J.; Haller, B. & Fuchs, J. A.

Thioredoxin is the bacterial protein encoded by fip that is required for filamentous bacteriophage f1 assembly.
J Bacteriol, **1985***, 161*, 799-802

Limtrakul, J.; Kokpol, S. & Rode, B.
Quantum chemical investigation on ion-dipeptide complex-formation
J Sei Soc Thailand, **1985***, 11*, 129-133

Lopez, J. & Webster, R. E.
Assembly site of bacteriophage f1 corresponds to adhesion zones between the inner and outer membranes of the host cell.
J Bacteriol, **1985***, 163*, 1270-1274

Luisi, P. L. [I]
Autopoiesis: a review and a reappraisal.
Naturwissenschaften, **2003**, 90, 49-59

Luisi, P. L. [II]
Contingency and determinism.
Philos Transact A Math Phys Eng Sci, **2003***, 361*, 1141-1147

Madder, A.; Li, L.; Muynck, H. D.; Farcy, N.; Haver, D. V.; Fant, F.; Vanhoenacker, G.; Sandra, P.; Davis, A. P. & Clercq, P. J. D.
Evaluation of a two-stage screening procedure in the combinatorial search for serine protease-like activity.
J Comb Chem, **2002***, 4*, 552-562

Mattheakis, L. C.; Bhatt, R. R. & Dower, W. J.
An in vitro polysome display system for identifying ligands from very large peptide libraries.
Proc Natl Acad Sci U S A, **1994***, 91*, 9022-9026

Matthews, C. N.
The origin of proteins: Heteropolypeptides from hydrogen cyanide and water.
Orig Life, **1975***, 6*, 155-162

Mead, D. A. & Kemper, B.
Chimeric single-stranded DNA phage-plasmid cloning vectors.
Biotechnology, **1988***, 10*, 85-102

Miller; Copeland; Papaioannou; Horstmann & Ruel
Kinetic Resolution of Alcohols Catalyzed by Tripeptides Containing the N-Alkylimidazole Substructure

J. Am. Chem. Soc, **1998***, 120*, 1629-1630

Model & Russel
Calendar, R. *(ed.)*
Filamentous bacteriophage. In The Bacteriophages
Plenum NY, **1988**

Mordukhova, E. A.; Lee, H. & Pan, J.
Improved thermostability and acetic acid tolerance of Escherichia coli via directed evolution of homoserine o-succinyltransferase.
Appl Environ Microbiol, **2008***, 74*, 7660-7668

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

Moussaoui, M.; Guasch, A.; Boix, E.; Cuchillo, C. & Nogués, M.
The role of non-catalytic binding subsites in the endonuclease activity of bovine pancreatic ribonuclease A.
J Biol Chem, **1996***, 271*, 4687-4692

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

Neri, D.; Petrul, H. & Roncucci, G.
Engineering recombinant antibodies for immunotherapy.
Cell Biophys, **1995***, 27*, 47-61

Oie, T.; Loew, G. H.; Burt, S. K. & MacElroy, R. D.
Quantum chemical studies of a model for peptide bond formation. 3. Role of magnesium cation in formation of amide and water from ammonia and glycine.
J Am Chem Soc, **1984***, 106*, 8007-8013

Oparin
A.I., O. *(ed.)*
The Origin of Life
Moscow Worker publisher, **1924**

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

Oparin, A. I.
[Problem of the origin of life according to modern biological concepts.]
Cas Lek Cesk, **1954***, 93*, 889-894

Pauling, L.
Chemical achievement and hope for the future.
Am Sci, **1948***, 36*, 51-58

Pérez-Payá, E.; Houghten, R. A. & Blondelle, S. E.
Functionalized protein-like structures from conformationally defined synthetic combinatorial libraries.
J Biol Chem, **1996***, 271*, 4120-4126

Rapoza, M. P. & Webster, R. E.
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

Roberts, R. W. & Szostak, J. W.
RNA-peptide fusions for the in vitro selection of peptides and proteins.
Proc Natl Acad Sci U S A, **1997***, 94*, 12297-12302

Rode, B. M.; Son, H. L. & Suwannachot, Y.
The combination of salt induced peptide formation reaction and clay catalysis: a way to higher peptides under primitive earth conditions.
Orig Life Evol Biosph, **1999***, 29*, 273-286

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

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

Sachdev, G. P. & Fruton, J. S.
Secondary enzyme-substrate interactions and the specificity of pepsin.
Biochemistry, **1970***, 9*, 4465-4470

Saetia, S.; Liedl, K. R.; Eder, A. H. & Rode, B. M.
Evaporation cycle experiments--a simulation of salt-induced peptide synthesis under possible prebiotic conditions.

Orig Life Evol Biosph, **1993***, 23*, 167-176

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

Segré, D. & Lancet, D.
Composing life.
EMBO Rep, **2000**, 1, 217-222

Siegel, D. L.
Research and clinical applications of antibody phage display in transfusion medicine.
Transfus Med Rev, **2001***, 15*, 35-52

Sigman & Jacobsen
Schiff Base Catalysts for the Asymmetric Strecker Reaction Identified and Optimized from Parallel Synthetic Libraries
J. Am. Chem. Soc., **1998***, 120*, 4901-4902

Sigman; Vachal & Jacobsen
A General Catalyst for the Asymmetric Strecker Reaction This work was supported by the NIH (GM-43214). A postdoctoral fellowship to M.S.S. (NIH), and a predoctoral fellowship to P.V. sponsored by Alfred Bader are gratefully acknowledged.
Angew Chem Int Ed Engl, **2000***, 39*, 1279-1281

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

Smith, G. P.
Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.
Science, **1985***, 228*, 1315-1317

Storm, D. R. & Koshland, D. E.
A Source for the Special Catalytic Power of Enzymes: Orbital Steering.
Proc Natl Acad Sci U S A, **1970***, 66*, 445-452

Stryer, L.
Biochemistry

Freeman, New York, **1988**

Taillades, J.; Collet, 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 alpha-amino acid N-carboxyanhydride. Prebiotic implications.
J Mol Evol, **1999***, 48*, 638-645

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

Tanaka, F.; Fuller, R. & Barbas, C. F.
Development of small designer aldolase enzymes: catalytic activity, folding, and substrate specificity.
Biochemistry, **2005***, 44*, 7583-7592

Thompson, R. C. & Blout, E. R.
Restrictions on the binding of proline-containing peptides to elastase.
Biochemistry, **1973***, 12*, 51-57

Vachal, P. & Jacobsen, E. N.
Enantioselective catalytic addition of HCN to ketoimines. Catalytic synthesis of quaternary amino acids.
Org Lett, **2000***, 2*, 867-870

Vazquez-Figueroa, E.; Yeh, V.; Broering, J. M.; Chaparro-Riggers, J. F. & Bommarius, A. S.
Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media.
Protein Eng Des Sel, **2008***, 21*, 673-680

Wachtershauser, G.
Before Enzyme and templates: theory of surface metabolism.
*Microbiol. Rev., **1988**, 52,* 452 -84

van Wezenbeek, P. M.; Hulsebos, T. J. & Schoenmakers, J. G.
Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd.
Gene, **1980***, 11*, 129-148

Winter, G.; Griffiths, A. D.; Hawkins, R. E. & Hoogenboom, H. R.
Making antibodies by phage display technology.
Annu Rev Immunol, **1994***, 12*, 433-455

Wirsching, P.; Ashley, J. A.; Lo, C. H.; Janda, K. D. & Lerner, R. A.
Reactive immunization.
Science, **1995***, 270*, 1775-1782

Xu, Y.; Yamamoto, N. & Janda, K. D.
Catalytic antibodies: hapten design strategies and screening methods.
Bioorg Med Chem, **2004***, 12*, 5247-5268

Università degli Studi "Roma Tre"-Dipartimento di Biologia

Scuola Dottorale in Biologia
Sezione "Biologia Cellulare e Molecolare"
Tesi di Dottorato XXI° ciclo

"Investigation of the functional properties of a completely de novo random library of peptides. Implication for the origin of life research ".

"Studio delle proprietà funzionali di una libreria peptidica a sequenza casuale. Implicazioni per lo studio dell'origine della vita".

Material and Methods

Candidato:
Dott. Davide De Lucrezia

Docente Guida:
Prof. Pier Luigi Luisi

Index

1. Material

1.1 Primers and Oligos

Oligo Name	Sequence
HisXa-Dummy FW2	agagagcatgccatgggctctcatcatcatcatcatcattctggcctgaata tcgaaggccgc
Dummy-FseI RV	cgggatccggccggcctttctagattgagagtaataagacccaaatt
cMycXa-Dummy FW2	agagagcatgccatggctgaacaaaaactcatctcagaagaggatctga atatcgaaggccgc
Dummy FW1	ctgaatatcgaaggccgcggggcggccgcataaataaataaatggcctt ggttgacggt
pUC 18rev	ggataacaatttcacacagg
ProtLib01 NNK-20_FW	ccgacgacgacgccatggcannknnknnknnknnknnknnknnk nnknnknnknnknnknnknnknnknnknnknnkgggccgg ccgcggc
ProtLib01 NNK-20_RV	tccaggcatgccccatcccctctagagccagaatgatgatgatgatgatg agagccgcggccttcgatagcggccgcggccggccc
LCupro RV	tccaggcatgccggatccggccggcctttctagatatggcaatgtggatg ccacgggcaatggcctgcggccgccgtcgtcgtcgg
LCupro FW	ccgacgacgacggcggccgcaggccattgcccgtggcatccacattgc catatctagaaaggccggccggatccggcatgcctgga
CutEnh-RV	tccaggcatgcc
CutEnh-FW	ccgacgacgacg
SC-Left	ggattgttattactcgcggc
GP3-ClaI	caccagtagcaccattacc
-96 gIII seq primer	ccctcatagttagcgtaagc

1.2 Plasmids

Plasmid	Supplier
pOCI 1050	Courtesy of Dr. Chiarabelli, University of RomaTre - Italy
pLasR	Courtesy of Dr. Leone, University of RomaTre - Italy
M13m19_1	Courtesy of Dr. Franceschini, Explora Srl - Italy

1.3 Cell Strain & Phage

Cell Strain	Genotype
E.coli XL1-Blue Electroporation-Competent Cells	XL1-Blue strain: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacIqZDM15 Tn10 (Tetr)]

E.coli XL10-Gold Ultracompetent Cells	Tetr D(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]a
E.coli Jm 109	F¢ (traD36, proAB+ lacIq, D(lacZ)M15) endA1 recA1
E.coli DH-5a	supE44 hsdR17 recA1 gyrA96 thi-1 relA1
E.coli ER2738	F´proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10(TetR)/ fhuA2 glnV Δ(lac-proAB) thi-1 Δ(hsdS-mcrB)5
M13K07 (Cat # N0315S)	NEB - England

1.4 Enzymes & Kits

Enzymes	**Supplier**
RedTaq - Taq DNA Polymerase (Cat # 28304)	Sigma Aldrich - Italy
GreenTaq - Taq DNA Polymerase (Cat # FD0694)	Promega - Italy
XhoI-FastDigest (Cat # 28304)	Fermentas - England
NcoI-FastDigest (Cat # ER0571)	Fermentas - England
XbaI-FastDigest (Cat # ER0681)	Fermentas - England
NdeI-FastDigest (Cat # ER0581)	Fermentas - England
BamHI-FastDigest (Cat # FD0054)	Fermentas - England
NotI-FastDigest (Cat # ER0591)	Fermentas - England
Klenow Fragment (Cat #EP0421)	Fermentas - England
T4 DNA Ligase (Cat # EL0015)	Fermentas - England
EagI (Cat # R3505S)	NEB - England
KpnI (Cat # R0142S)	NEB - England
T4 DNA Ligase (Cat # M0202L)	NEB - England
Eu-labelled Streptavidin (Cat # 1244-360)	PerkinElmer - Italy

Kits	Supplier
QIAquick Nucleotide Removal Kit (Cat # 28304)	Qiagen - germany
Wizard SV Gel and PCR Clean-Up System (Cat # A9281)	Promega - Italy
MagneHis Protein Purification System (Cat #V8500)	Promega - Italy
Sigma GenElute Plasmid Miniprep Kit (Cat # PLN70)	Sigma Aldrich - Italy

1.5 Buffers, Media, Chemicals & Compounds

Compounds
TSA: p-nitrophenyl hydrogen 4-(hydroxycarbonyl)butylphosphonat) (Protera Srl, Cat # not available, compound custom made)

Buffers & Media
10% Ammonium persulfate Ammonium persulfate 1g DEPC-treated water to 10 ml Store at 4°C
100X Salts Stock solution 1 M NaCl 0,25 M KCl 1 M MgCl2 1 M MgSO4
1X TBE solution 89 mM Tris-Borate 2 mM EDTA
30% Acrylamide/Bisacrylamide solution Acrylamide 29g N,N'-methylenbisacrylamide 1g DEPC-treated water to 100 ml

50X TAE Buffer 242 g Tris base 57 g ml glacial acetic acid 100 ml 0.5 M EDTA(pH 8)
Ampicillin stock 1000x Ampicillin (20 mg/ml)
Annealing buffer 100 mM Tris-HCL pH 7,5 500 mM NaCl 10 mM EDTA
Biotin Stock Solution: 6 mg Biotin 400 uL d.d. water 600 uL DMSO
Biotin/NH, working solution: 1.83 mg sulfo-NHS 500 µl Biotin stock solution d.d.water up to 10 ml
CovaBuffer: 116,9 g NaCl 10 g $MgSO_4$ 500 µl Tween 20 PBS 1X up tp 1L
DELFIA Assay Buffer Concentrate (PerkinElmer, Cat # 1244-111)
DELFIA Wash Buffer Concentrate (PerkinElmer, Cat # 1244-114)
EDC, working solution: 5.8 mg EDC d.d.water up to 10 ml
Ethidium-bromide stock solution Ethidium Bromide 10 mg/ml in water
Kanayicin stock 1000x Kanayicin (10 mg/ml)

LB Agar plates add 15 g /liter of Bacto-agar to LB medium. Sterilize by autoclaving. Allow the solution to cool down to 50°C, add the required antibiotic and poor the solution into petri-plates.
LB medium Per liter: bacto-tryptone 10 g bacto-yeast extract 5 g NaCl 10 g add deionized water to 1 liter adjust to pH 7.0, sterilize by autoclavinig.
Loading Dye 2X Solution 95% (w/v) deionized formamide 0.5 mM EDTA 0.025% xylen cyanolo 0.025% bromophenol blue 0.025% SDS
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. Adjust pH to 7.4
SOC Medium 2 g Bacto Tryptone 0.5 g Bacto Yest extract 1 ml 100X salts stock solution Sterilize by autoclaving 2 ml 1M Glucose
Tetracycline stock 1000x Tetrayicline (10 mg/ml)
TSA, soluzione stock: 7 mg TSA 400 µl d.d. water 600 µl DMSO

TSA/NH, working solution: 1.83 mg sulfo-NHS 500 µl TSA soluzione stock d.d.water up to 10 ml
Wash Buffer: 5 ml TWEEN 20 PBS 1X up tp 1L

Chemical	**Supplier**
Acrilammide/bis-acrilammide (29:1) 30%	Sigma Aldrich - Italy
Agarose	Sigma Aldrich
Ammonium persulphate	Sigma Aldrich
Ampicillin	Sigma Aldrich
Bacto-agar	Sigma Aldrich
Bacto-tryptone	Sigma Aldrich
Bacto-yeast extract	Sigma Aldrich
Biotin	Sigma Aldrich
DMSO	Sigma Aldrich
DTT	Sigma Aldrich
EDC	Pierce - England
EDTA	Sigma Aldrich
Ethanol	Sigma Aldrich
Glicerol	Sigma Aldrich
Glucose	Sigma Aldrich
HCl	Sigma Aldrich
Kanamycine	Sigma Aldrich
KCl	Sigma Aldrich
KH_2PO_4	Sigma Aldrich
$MgCl_2$	Sigma Aldrich
$MgSO_4$	Sigma Aldrich
NaCl	Sigma Aldrich
Na_2HPO_4	Sigma Aldrich
Sulfo - NHS	Pierce - England
NHS	Sigma Aldrich
SDS	Sigma Aldrich
Tetracycline	Sigma Aldrich
Trizma base	Sigma Aldrich
Tween 20	Sigma Aldrich
Covalink NH_2 -MP	Nunc - Denmark

Chemical	Supplier
DMSO	Sigma Aldrich
Chloroform	Sigma Aldrich
DMF	Sigma Aldrich

2. Methods

2.1 Construction of the phagemid vector

The following PCR reactions are performed:

Component	Concentration	A	B	C	- CTRL
DNA template Genomic PAO1	25mg/µl	1µl	1µl	1µl	--
Buffer mix	46.5	46.5	46.5	46.5	46.5
cMycXa-Dummy FW2	16pmoles/µl	--	2.5	--	--
HisXa-Dummy FW2	16 pmoles/µl	--	--	2.5	--
Water		2.5	--	--	3.5
	50µL				

Buffer Mix:

Component	Concentration	For 1 reaction	For 4 reactions
RedTaq Buffer	10X	5µl	20µl
dNTPs	10 mM	1	5
Dummy FW1	16pmoles/µl	2.5	10
Dummy-FseI RV	16 pmoles/µl	2.5	10
Water		33	132
RedTaq polymerase	2U/µl	2.5	11
		46.5µl	186µl

Thermocycle profile:

94°C/5’	
94°C/30’’-50°C/45’’-72°C/1’	10 times
72°C/7’	
94°C/30’’-68°C/45’’-72°C/1’	25 times
4°/20h	

The following PCR reactions are performed:

Component	Concentration	A1	A2	- CTRL	
DNA template PCR A	25mg/µl	3µl	3µl	--	
Buffer mix	46.5	44.5	44.5	44.5	
cMycXa-Dummy FW2	16pmoles/µl	2.5	--	--	
HisXa-Dummy FW2	16 pmoles/µl	--	2.5	2.5	
Water		--	--	3	
	50µL				

Buffer Mix:

Component	Concentration	For 1 reaction	For 3 reactions
RedTaq Buffer	10X	5µl	15µl
dNTPs	10 mM	1	3
Dummy-FseI RV	16 pmoles/µl	2.5	7.5
Water		33	100.5
RedTaq polymerase	2U/µl	2.5	7.5
		46.5µl	133.5µl

Thermocycle profile:

94°C/5’	
94°C/30’’-55°C/45’’-72°C/1’	10 times
72°C/7’	
94°C/30’’-68°C/45’’-72°C/1’	25 times
4°/20h	

Parent plasmid (pOCI 1050) is transformed into electrocompetent *E.coli* DH5α cells according to [Maniatis et al., 2004]
One single colony is inoculated into 10ml LB o.n. shaking 37°C.
Cell culture is minipreped (Sigma). Final elution volume: 100 ul

The following restriction reactions are arranged:

Component	Concentration	1	A2	A2
DNA template	25mg/µl	pOCI1050 24ul	PCR A1 16ul	PCR A1 16ul
BamHI Buffer	10X	3	2	2
BSA	10x	3	2	2
BamHI	1 Unit/µl	0.5	0.5	0.5
Water		--	--	--
		30.5	20.5	20.5
Incubation temperature	37°C			
Incubation Time	o.n.			

The following restriction reactions are arranged:

Component	Conc.	1	A2	A2	+ ctrl 1	+ ctrl 2
DNA template	25mg/µl	Reac. 1 30ul	Reac A1 20ul	Reac A1 20ul	pOCI 1050 8ul	pOCI10 50 8ul
BamHI Buffer	10X	--	--	--	1	1
BSA	10x	3	2	2	1	1
BamHI	1 Unit/µl	0.5	0.5	0.5	--	0.5
NcoI Buffer	10X	6	4	4	1	--
NcoI		1	1	1	0.5	--
Water		23	15	15	--	--
		60	40	40	10	10
Incubation temperature	37°C					
Incubation Time	o.n.					

The following ligation reactions are arreanged:

Component	Concentration	pIII dummy/cMyc	pIII dummy/His tag
Double digested vector	25mg/µl	10	10
Insert	10X	7	7
T4 DNA Ligase Buffer	10x	2	2
T4 DNA ligase	1 Unit/µl	1	1
Water		--	--
Incubated o.n. at 16°C			

The following ligation are performed

Component	A1L	A2L	+ ctrl
Ligated Product	pIII dummy/cMyc (5ul)	pIII dummy/His tag (5ul)	pOCI 1050 (5ul)

Ligation products are transformed into chemical-competent *E.coli* DH5α cells
Cloning products are checked by PCR using Sc-Left and GP3-ClaI primers with the following thermocycle profile

94°C/5’	
94°C/30’’-52°C/45’’-72°C/1’	25 times
72°C/7’	
4°/20h	

10 clones from each reaction A1L and A2L are individually cultured o.n. at 37°C.
Clones are submitted to miniprep and corresponded plasmid are sequenced using Sc-Left and GP3-ClaI primers.

2.2 Testing cloning of efficiency of the new vector

The following reactions are arranged:

Components	- ctrl	1: APP (ca 133 bp)	2: LasR (ca 717 bp)	3: L-Cupro (ca 30 bp)	4: pOCI 1050
Template	2 ul H2O	2 ul	2 ul	2 ul	2 ul
FW Primer	Sc-Left 2.5 ul	Sc-Left 2.5 ul	GIORD FW 2.5 ul	LCUPRO FV 2.5 ul	Sc-Left 2.5 ul
RV Primer	GP3-ClaI 2.5 ul	GP3-ClaI 2.5 ul	GIORD RV 2.5 ul	LCUPRO RV 2.5 ul	GP3-Cla 2.5 ul I
Buffer Mix	42.5 ul	42.5 ul	42.5 ul	42.5 ul	42.5 ul

Buffer Mix	Components
dNTPs 10 mM	12 ul
RedTaq Buffer 10X	30 ul
Taq 2 EU/ul	12 ul
d.d. Water	to 258 ul

Thermocycler Profile (PetPCR2)

94°C/5’	
94°C/30’’-52°C/45’’-72°C/1’	25 times
72°C/7’	
4°/20h	

The following restriction reaction is arranged:

Components	Inserts (1 to 3)	Phagemid Vector
DNA	PCR products (1 to 3) 24 ul	24 ul
BamHI Buffer 10X	3 ul	3 ul
BSA 10X	3 ul	3 ul
BamHI (NEB)	1 ul	1 ul
Incubated o.n. at 37°C		

The following restriction reaction is arranged:

Components	Inserts (1 to 3)	Phagemid Vector	+ Ctrl
DNA	Insert reactions 30 ul (1 to 3)	Vector reaction 30 ul	Vector 12 ul
BamHI Buffer 10X	-- ul	-- ul	1 ul
BSA 10X	6 ul	6 ul	1 ul
BamHI	-- ul	-- ul	-- ul
Not I Buffer O+ 10X	6 ul	6 ul	2 ul
Not I (Fermentas)	1 ul	1 ul	0.5 ul
H2O	to 60 ul		to 20 ul
Incubated o.n. at 37°C			

Double-digested inserts are purified by spin-column procedure (Sigma). Purified inserts are lyophilised and resuspended in appropriate volume to get a final concentration of ca 15 pmoles/ul
Double-digested vector is gel excised and purified by spin-column procedure (Promega)

The following ligation reactions are arranged:

Components	L1	L2	L3
Double-digested and purified insert	10 ul	10 ul	10 ul
Double-digested and purified vector	7 ul	7 ul	7 ul
T4 Ligase Buffer 10X	2 ul	2 ul	2 ul
T4 Ligase	1 ul	1 ul	1 ul
H2O	To 20 ul	To 20 ul	To 20 ul

Ligation products are transformed into chemical-competent *E.coli* DH5ά cells Transformants are verified by PCR using Sc-Left and GP3-ClaI primers with the following thermocycle profile

94°C/5'	
94°C/30''-52°C/45''-72°C/1'	25 times
72°C/7'	
4°/20h	

10 clones from each reaction are individually cultured o.n. at 37°C.
Clones are submitted to miniprep and corresponded plasmid are sequenced using Sc-Left and GP3-ClaI primers.

2.3 Optimization of the phage propagation protocol

2.3.1 Optimization of the phenotype recovery

Phagemid harbouring the APP gene downstream either the 6xHis tag or the cMyc tag (pIII-His-APP and pIII-cMyc-APP respectively) are propagated in XL-1 Blue DMSO-competent according to Maniatis et al., 2004

Each individual phages population is recovered according to the standard protocol [McCafferty et al., 1996].

Each individual phages population is recovered exploiting superparamagnetic beads according to manufacturer protocol (Promega)

The four phage populations are titer according to the standard protocol [McCafferty et al., 1996] and their titer compared in order to determined recovery yield.

Additionally, phage deriving from the pIII-His-APP phagemid are spiked with 1:1 solution of M13K07 (NEB) and purified by IMAC using superparamagnetic beads according to manufacturer protocol (Promega)

Eluted phages are serially diluted and plated on LB Agar plates containing either ampicillin (selection marker for pIII-His-APP phagemid) or kanamycin (selection marker for M13K07 phage)

Titer comparison allows the to calculate the unspecific background recovery when mixed phage solution are recovered by IMAC

2.3.2 Optimization of the genotype recovery

Phagemid pIII-His-APP is propagated in XL-1 Blue DMSO-competent according to McCafferty [McCafferty et al., 1996] varying the following parameters:

Multiplicy of Infection (MOI): 0,5 - 1 - 2
Cell density upon infection as Abs @ 460: 0,2 - 0,4 - 0,8

Each individual phages population is recovered according to the standard protocol [McCafferty et al., 1996].

The nine phage populations are titer according to the standard protocol [McCafferty et al., 1996] by plating them on LB Agar plates containing either

ampicillin (selection marker for pIII-His-APP phagemid) or kanamycin (selection marker for M13K07 phage)

Titer comparison allows the to calculate the ratio of M13K07 genome encapsulated during phage production and determine optimal condition.
unspecific background recovery when mixed phage solution are recovered by IMAC

2.4 Construction of a monovalent phagemid library

Double-strand DNA library is synthetised as follows:

1. Kleenow reaction	
Oligo FW: ProtLib01 NNK-20_FW (10 uM)	5 ul
Oligo RV: ProtLib01 NNK-20_RV (10 uM)	5 ul
NEBuffer 2 10X	2.5 ul
d.d. H_2O	10 ul
Incubation at 95°C/5'	Y
Cool down slowly at r.t.	N
Incubation at 4°C	Y
dNTPs 10 mM	2.5 ul
Klenow Fragment	0.5 ul
Incubation temperature	25°
Incubation time	15'
Reaction Purification by spin column procedure (Qiagen)	
2. PCR amplification	
Purified reaction	2.5 ul
CutEnh FW 16 pmol/ul	1.5 ul
CutEnh RV 16 pmol/ul	1.5 ul
RedTaq Buffer 10X	5 ul
dNTPs 10 mM	1.5 ul
RedTaq Enzyme (2 EU/ul)	1 ul
d.d. H2O	To 50 ul
Reaction Purification by spin column procedure (Qiagen)	

The following digestion reaction is arranged:

	I
Vector	15 ul
NcoI Fastzyme (Fermentas)	1 ul
BamHI Fastzyme (Fermentas)	1 ul
Fastzymes Buffer 10X	3ul

d.d. H2O	To 30 ul
37°C/ 45’	

Reaction Purification by spin column procedure (Qiagen).

Monovalent phagemid constructed as described in chapter 2.1 is propagated in chemical-competent *E.coli* DH5α cells, minipreped by spin-column procedure according to manufacturer’s instruction (Sigma) and double digested as follows:

	pIII-Dummy-His
Vector	15 ul
NcoI Fastzyme (Fermentas)	1 ul
BamHI Fastzyme (Fermentas)	1 ul
Fastzymes Buffer 10X	2ul
d.d. H2O	to 20 ul
37°C/ 2 hours	

Double-digested vector is gel excised and purified by spin-column procedure (Promega).

The following ligation reaction is arranged:

Vector	8 ul
Insert	10 ul
T4 DNA ligase (NEB)	1 ul
Buffer 10X	2 ul
d.d. H2O	to 20 ul
25°C/ 20’	

Ligation Product (5 ul) is transformed in electrocompetent *E.coli Ten Gold* cell (45 ul)
Electroporation parameters:

Set voltage: 1.8 kV
Actual voltage: 1.8 kV
V/2 Time: 3.9 msec

Cell are rescued according to the standard protocol [Maniatis et al., 2004]

10 colonies are culture, minipreped (Sigma) and sequenced using pUC 18 rev primer

2.5 Construction of a polyvalent phagemid library

Double-strand DNA library is synthetised as follows:

1. Kleenow reaction	
Oligo FW: M13ProtLib20-NNK_FW (10 uM)	5 ul
Oligo RV: M13ProtLib20-NNK_RV (10 uM)	5 ul
NEBuffer 2 10X	2.5 ul
d.d. H_2O	10 ul
Incubation at 95°C/5'	Y
Cool down slowly at r.t.	N
Incubation at 4°C	Y
dNTPs 10 mM	2.5 ul
Klenow Fragment	0.5 ul
Incubation temperature	25°
Incubation time	15'
Reaction Purification by spin column procedure (Qiagen)	
2. PCR amplification	
Purified reaction	2.5 ul
ScLeft FW 10 pmol/ul	1.5 ul
GP3-ClaI 10 pmol/ul	1.5 ul
Green Taq Buffer 10X	5 ul
dNTPs 10 mM	1.5 ul
GreenTaq Enzyme (2 EU/ul)	1 ul
d.d. H2O	To 50 ul
Reaction Purification by spin column procedure (Qiagen)	

The following digestion reaction is arranged:

	I
Vector	15 ul
EagI (NEB)	1 ul
KpnI (NEB)	1 ul
Buffer 10X	3ul
d.d. H2O	To 30 ul
37°C/ 45'	

Reaction Purification by spin column procedure (Qiagen).

M13m19_1 (Explora) is propagated in chemical-competent *E.coli* ER2738 cells, minipreped by spin-column procedure according to manufacturer's instruction (Sigma) and double digested as follows:

	pIII-Dummy-His
Vector	15 ul
EagI (NEB)	1 ul
KpnI (NEB)	1 ul
Buffer 10X	2ul
d.d. H2O	to 20 ul
37°C/ 2 hours	

Double-digested vector is gel excised and purified by spin-column procedure (Promega).

The following ligation reaction is arranged:

Vector	8 ul
Insert	10 ul
T4 DNA ligase (NEB)	1 ul
Buffer 10X	2 ul
d.d. H2O	to 20 ul
25°C/ 20'	

Ligation Product (5 ul) is transformed in chemical-competent *E.coli* XL10-Gold Ultracompetent cell (Sigma) (45 ul)
Cell are rescued according to manufacturer's instruction (Sigma)

12 colonies are culture, minipreped (Sigma) and sequenced using -96 gIII seq primer

2.6 Optimization of the coupling reaction

2.6.1 Coupling TSA to NH_2-functionalised surface

In order to optimize the coupling of an apten to a NH_2-functionalised 96-wells microplate (Nunc) the following reagents are added in the specified order:

d.d. water	50 µl
Apten	*x* µl
EDC	*y* µl
Additive	*r* µl
Solvent	*z* µl

where:

$x + y + r + z = 150$ µl
$x + r = 100$ µl

and:

10 µl < x < 100 µl
10 µl < y < 100 µl
0 µl < r < 50 µl
0 µl < z < 15 µl

and:

"Solvent" is either DMSO, DMF or chloroform
"Additive" is either N-hydroxysuccinimide or sulfo- N-hydroxysuccinimide

Incubate at r.t. for 90'

Wash with 200µl of CovaBuffer for 3 times and incubate for 15' the last wash prior to discard the content

For negative controls the following reagents are added in the specified order:

d.d. water	50 µl
Apten	*0* µl
EDC	*50* µl
Additive	*0* µl
Solvent	*0* µl

For positive controls the following reagents are added in the specified order:

d.d. water	50 µl
Biotin/S-NHS	*100* µl
EDC	*50* µl
Additive	*0* µl
Solvent	*0* µl

In order to evaluate reaction yield, apten treated surface are further reacted with biotin.

Sample, negative and positive control and double-treated wells are incubated with Eu-labelled Streptavidin (PerkinElmer) as follows:

- wash 5 times each well with 200 µl Delfia Wash Buffer (PerkinElmer) 1:20 diluted
- each well is incubated for 1 hour with 100 µl Eu-labelled Streptavidin diluted down to 100 ng/ml with DELFIA Assay Buffer (PerkinElmer)
- wash 5 times each well with 200 µl Delfia Wash Buffer (PerkinElmer) 1:100 diluted
- add to each well 200 µl Delfia Enhance Buffer (PerkinElmer), shake for 5 minutes and read fluorescence output (Victor 3V -PerkinElmer).

2.7 Biopanning

2.7.1 Coupling TSA to NH_2-functionalised surface

A NH_2-functionalised 96-wells microplate (Nunc) is treated as follows:

For all well but negative controls the following reagents are added in the specified order:
50µl of d.d.H2O
100µl of TSA/NH solution
to every well 50µl of EDC Solution.
Incubate at r.t. for 90'

Wash with 200µl of CovaBuffer for 3 times and incubate for 15' the last wash prior to discard the content

For negative controls the following reagents are added in the specified order:
Add 150µl of H2O
Add to every well 50µl of EDC Solution.
Incubate at r.t. for 90'

Wash with 200µl of CovaBuffer for 3 times and incubate for 15' the last wash prior to discard the content

2.7.2 Biopanning of monovalent phagemid library

The following Microplate is arranged:

	1	2	3	4	9	10	11	12
A								
B								
C								
D								
E								
F								
G								
H								

TSA-loaded wells
Negative Controls

Propagate phagemid library constructed as described in chapter 2.4 according to protocol developed in chater 2.3

For wells A1:D1
BR buffer pH4: 90µl
ClZn 100 mM: 0 µl
Phage sol 7•108: 10 µl
For wells A2:D2
BR buffer pH6: 90µl
ClZn 100 mM: 0 µl
Phage sol 7•108: 10 µl
For wells A3:D3
BR buffer pH7: 90µl
ClZn 100 mM: 0 µl
Phage sol 7•108: 10 µl
For wells A4:D4
BR buffer pH10: 90µl
ClZn 100 mM: 0 µl
Phage sol 7•108: 10 µl

For wells A9:D9
BR buffer pH4: 80µl
ClZn 100 mM: 10 µl
Phage sol 7•108: 10 µl
For wells A10:D10
BR buffer pH4: 80µl
ClZn 100 mM: 10 µl
Phage sol 7•108: 10 µl
For wells A11:D11
BR buffer pH4: 80µl
ClZn 100 mM: 10 µl
Phage sol 7•108: 10 µl
For wells A11:D11
BR buffer pH4: 80µl
ClZn 100 mM: 10 µl
Phage sol 7•108: 10 µl

For wells E1:H1
BR buffer pH4: 80µl
ClZn 10 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E2:H2
BR buffer pH4: 80µl
ClZn 10 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E3:H3
BR buffer pH4: 80µl
ClZn 10 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E4:H4
BR buffer pH4: 80µl
ClZn 10 mM: 10 µl
Phage sol 7•108: 10 µl

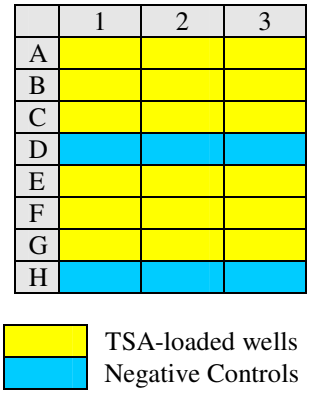
For wells E9:H9
BR buffer pH4: 80µl
ClZn 1 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E10:H10
BR buffer pH4: 80µl
ClZn 1 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E11:H11
BR buffer pH4: 80µl
ClZn 1 mM: 10 µl
Phage sol 7•108: 10 µl
For wells E11:H11
BR buffer pH4: 80µl
ClZn 1 mM: 10 µl
Phage sol 7•108: 10 µl

Incubate for 2 hour at r.t.
Wash 15 times with 200µl BR Buffer/NaCl 150 mM. Incubate 1 hours after the 15th wash with 200µl BR Buffer/NaCl 150 mM
Wash 5 times with 200µl BR Buffer/NaCl 150 mM. Incubate 30'. Discard the solution
Add 200 µl 250µM TSA, incubate 2 hours and recover eluted phages.

Phage solutions are stored at 4°C until titering. Titering is carried out according to the standard protocol [McCafferty et al., 1996]

2.7.3 Biopanning of polyvalent phagemid library

The following Microplate is arranged and prepared as in 2.6.1:



Propagate phagemid library constructed as described in chapter 2.5.

1st Biopanning cycle

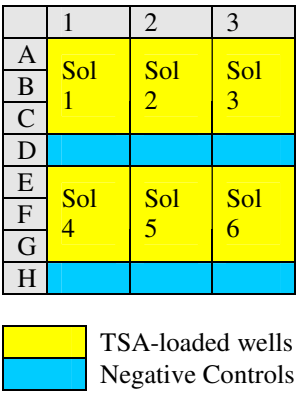
The following buffer solutions are prepared:

Solution 1: BR buffer pH4: 270µl ClZn 10 mM: 0 µl Phage sol 1012: 30 µl	Wash Buffer 1: BR buffer pH4: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl
Solution 2: BR buffer pH4: 240µl ClZn 10 mM: 30 µl Phage sol 1012: 30 µl	Wash Buffer 2: BR buffer pH4: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl
Solution 3: BR buffer pH6: 270µl ClZn 10 mM: 0 µl Phage sol 1012: 30 µl	Wash Buffer 3: BR buffer pH6: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl

Solution 4: BR buffer pH6: 240µl ClZn 10 mM: 30 µl Phage sol 1012: 30 µl	Wash Buffer 4: BR buffer pH6: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl
Solution 5: BR buffer pH8: 280µl ClZn 10 mM: 0 µl Phage sol 1012: 30 µl	Wash Buffer 5: BR buffer pH8: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl
Solution 6: BR buffer pH8: 240µl ClZn 10 mM: 30 µl Phage sol 1012: 30 µl	Wash Buffer 6: BR buffer pH8: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl

Phage solution: 1012 cfu/ml of NNK20-M13KOE

Add 50µl PBS to each well and then add 100µl of phage solution to the microplate according to the following scheme where the number identifies the selection conditions according to the solution above:



Incubate for 2 hours at r.t.
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 10 times.
Incubate 1 hours after the 10th wash.
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 5 times.
Remove wash solution and firmly slap the plate face down onto a clean towel
Add 100 µl of TSA (100µg/ml), incubate 3 hours and recover eluted phages.

Serial dilutions of each phage population are prepared in PBS down to 10^-3. Each phage population is tittered by infecting 200µl early-log phase *E.coli* ER2738 cells with 10µl phage solution. The rest of the eluate is stored at 4°C.

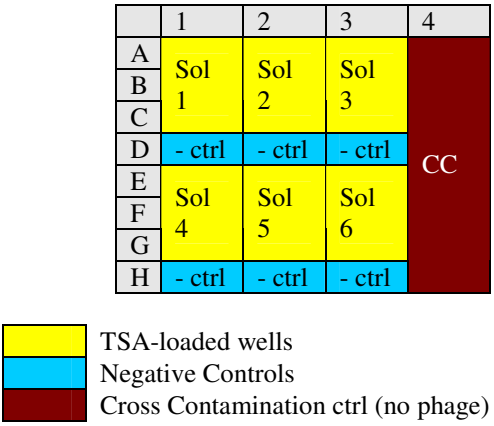
2nd Biopanning cycle
Each phage population eluted in the previous biopanning step is propagated in ER2738 *E.coli*

The following buffer solutions are prepared:

Solution 1: BR buffer pH4: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 1: BR buffer pH4: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl
Solution 2: BR buffer pH4: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 2: BR buffer pH4: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl
Solution 3: BR buffer pH6: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 3: BR buffer pH6: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl
Solution 4: BR buffer pH6: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 4: BR buffer pH6: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl
Solution 5: BR buffer pH8: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 5: BR buffer pH8: 9000µl ClZn 10 mM: 0 µl NaCl 2M : 1000µl
Solution 6: BR buffer pH8: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 6: BR buffer pH8: 8000µl ClZn 10 mM: 1000 µl NaCl 2M : 1000µl

* phage solution are not tittered since previous experiments showed that the present amplification protocol yields ca 10^13 cfu/ml

100µl of phage solution is added to each well to the microplate according to the scheme outlined bellow except for column 4:



Incubate for 2 hour at r.t. Add 100µl of PBS to every well, but in column 4 where 200 µl of PBS are added. No phage are added to wells in column 4, those wells are treated as the others (including elution and tittering) to verify that the automated protocol does not cause cross-contamination (Protocol Name: "Davide_IVE").
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 10 times.
Incubate 1 hours after the 10th wash.
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 5 times.
Remove wash solution and firmly slap the plate face down onto a clean towel
Add 100 µl of TSA (100µg/ml), incubate 3 hours and recover eluted phages.
Phage solutions are stored at 4°C until titering.

Serial dilutions of each phage population are prepared in PBS down to 10^-3. Each phage population is tittered by infecting 200µl early-log phase *E.coli* ER2738 cells with 10µl phage solution. The rest of the eluate is stored at 4°C.

3rd Biopanning Cycle

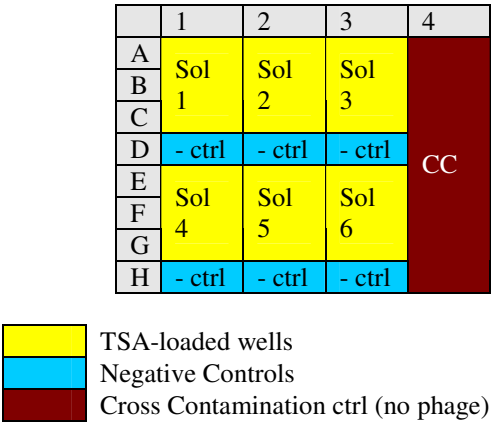
Each phage population eluted in the previous biopanning step is propagated in ER2738 *E.coli*

The following buffer solutions are prepared:

Solution 1: BR buffer pH4: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 1: BR buffer pH4: 8750µl ClZn 10 mM: 0 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)
Solution 2: BR buffer pH4: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 2: BR buffer pH4: 7750µl ClZn 10 mM: 1000 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)
Solution 3: BR buffer pH6: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 3: BR buffer pH6: 8750µl ClZn 10 mM: 0 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)
Solution 4: BR buffer pH6: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 4: BR buffer pH6: 7750µl ClZn 10 mM: 1000 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)
Solution 5: BR buffer pH8: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 5: BR buffer pH8: 8750µl ClZn 10 mM: 0 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)
Solution 6: BR buffer pH8: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 6: BR buffer pH8: 7750µl ClZn 10 mM: 1000 µl NaCl 2,5M : 1000µl Tween 20: 250 µl (0,025%)

* phage solution are not tittered since previous experiments showed that the present amplification protocol yields ca 1013 cfu/ml

100µl of phage solution is added to each well to the microplate according to the scheme outlined bellow except for column 4:



Incubate for 1 hour at r.t. Add 100µl of PBS to every well, but in column 4 where 200 µl of PBS are added. No phage are added to wells in column 4, those wells are treated as the others (including elution and tittering) to verify that the automated protocol does not cause cross-contamination (Protocol Name: "Davide_IVE").
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 15 times.
Incubate 1 hours, shaking after the 10th wash.
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 15 times.
Remove wash solution and firmly slap the plate face down onto a clean towel
Add 100 µl of TSA (100µg/ml), incubate 2 hour and recover eluted phages.
Phage solutions are stored at 4°C until titering.

Serial dilutions of each phage population are prepared in PBS down to 10^-3. Each phage population is tittered by infecting 200µl early-log phase *E.coli* ER2738 cells with 10µl phage solution. The rest of the eluate is stored at 4°C.

4th Biopanning Cycle

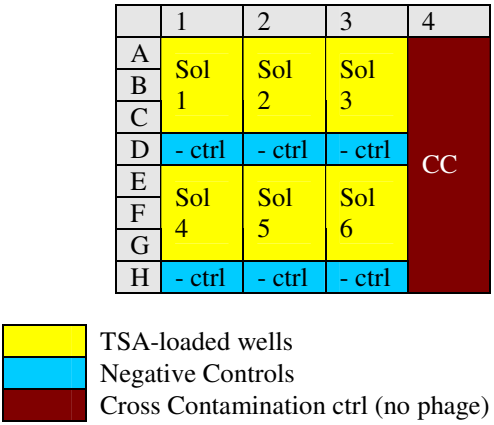
Each phage population eluted in the previous biopanning step is propagated in ER2738 *E.coli*

The following buffer solutions are prepared:

Solution 1: BR buffer pH4: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 1: BR buffer pH4: 8500µl ClZn 10 mM: 0 µl NaCl 2,5M : 1000µl Tween 20: 500 µl (0,05%)
Solution 2: BR buffer pH4: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 2: BR buffer pH4: 7500µl ClZn 10 mM: 1000 µl NaCl 3M : 1000µl Tween 20: 500 µl (0,05%)
Solution 3: BR buffer pH6: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 3: BR buffer pH6: 8500 µl ClZn 10 mM: 0 µl NaCl 3M : 1000µl Tween 20: 500 µl (0,05%)
Solution 4: BR buffer pH6: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 4: BR buffer pH6: 7500 µl ClZn 10 mM: 1000 µl NaCl 3M : 1000µl Tween 20: 500 µl (0,05%)
Solution 5: BR buffer pH8: 360µl ClZn 10 mM: 0 µl *Phage sol: 40 µl	Wash Buffer 5: BR buffer pH8: 8500 µl ClZn 10 mM: 0 µl NaCl 3M : 1000µl Tween 20: 500 µl (0,05%)
Solution 6: BR buffer pH8: 320µl ClZn 10 mM: 40 µl *Phage sol: 40 µl	Wash Buffer 6: BR buffer pH8: 7500 µl ClZn 10 mM: 1000 µl NaCl 3M : 1000µl Tween 20: 500 µl (0,05%)

* phage solution are not tittered since previous experiments showed that the present amplification protocol yields ca 1013 cfu/ml

100µl of phage solution is added to each well to the microplate according to the scheme outlined bellow except for column 4:



Incubate for 0,5 hour at r.t. Add 100µl of PBS to every well, but in column 4 where 200 µl of PBS are added. No phage are added to wells in column 4, those wells are treated as the others (including elution and tittering) to verify that the automated protocol does not cause cross-contamination (Protocol Name: "Davide_IVE").
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 20 times.
Incubate 1 hours, shaking after the 10th wash.
Remove 100µl from each well and add 100µl of the corresponding wash solution. Perform wash 20 times.
Remove wash solution and firmly slap the plate face down onto a clean towel
Add 100 µl of TSA (100µg/ml), incubate 1 hour and recover eluted phages.
Phage solutions are stored at 4°C until titering.

Serial dilutions of each phage population are prepared in PBS down to 10^-3. Each phage population is tittered by infecting 200µl early-log phase *E.coli* ER2738 cells with 10µl phage solution. The rest of the eluate is stored at 4°C.

12 colonies for each population are picked, resuspended in 1,5 LB Tet and grown o.n. at 37° shaking in a deep-well plate.

The 72 clones are minipreped according to manufacturer instruction (Millipore - HT Miniprep). Resuspended in 50 µl d.d.H2O.

12 clones for each phage population biopanned at pH4 and 6 clones for each other phage population (48 clones total) were sequenced and analysed using the -96 gIII seq primer.