# Protein Directed Evolution

## (Evolución Dirigida de Proteínas)

Roberto Laos[1,2]

[1]Department of Chemistry, University of Florida, Gainesville, FL 32611, USA

**Abstract**

Directed evolution allows us to explore protein functionalities not required in the natural environment. It mimics natural genetic processes and selective pressures. This approach is used when the molecular basis is not completely understood and rational design is a difficult task. This approach consists of serial cycles of consecutive diversification and selection which eventually lead to the accumulation of beneficial mutations. Here are presented two cases where directed evolution is used to modify two different proteins: *Taq* polymerase, enzyme used for DNA extension in PCR, and the LacI repressor protein which regulates gene expression on *E.coli*.

**Key words:** directed evolution, water in oil emulsions, Taq polymerase, lac repressor, PCR, error prone PCR, compartmentalized self replication (CSR), Artificially Expanded Genetic Information System (AEGIS).

**Resumen**

La evolución dirigida es una técnica que nos permite explorar funciones enzimáticas que no son requeridas en el ambiente natural. Esta técnica, simula procesos genéticos naturales y de selección. Esta estrategia se utiliza cuando un diseño racional es muy complicado. Consiste en una repetición de ciclos de diversificación y selección que llevan a la acumulación de mutaciones benéficas. Aquí se presenta dos ejemplos de evolución dirigida con los cuales se ha trabajado directamente: la ADN polimerasa del organismo *Thermus aquaticus* usada comúnmente en PCR, y la proteína LacI que regula la expresión de genes usados para el metabolismo de lactosa en *E. Coli*.

### Directed evolution and generation of a library of genes

Directed evolution allows us to explore protein functions not required in the natural environment or for which the molecular basis are not completely understood and rational design is a difficult task[1]. This approach mimics natural genetic processes and selective

[2] laos.r@pucp.edu.pe, rlaos@ufl.edu

pressures. It consists of serial cycles of consecutive diversification and selection which eventually lead to the accumulation of beneficial mutations.

The success of directed evolution is based on two steps: generating molecular diversity and identifying the improved variants. The size and variety of the library as well as the ability of an assay to meaningfully detect improvements in the desired activity are paramount for a successful in vitro directed evolution process. The most straightforward way to obtain a library is by having a collection of nucleic acid molecules from which the protein library can be translated.

Error prone PCR (ePCR) is a very simple and versatile way to generate a library. It is a modification of a standard PCR where the desired gene is amplified using *Taq* polymerase. Taq polymerase has an inherent propensity to introduce mistakes when copying the template gene. Its error rate can be increased by introducing $Mn^{2+}$ which is a known mutagenic agent for DNA polymerases[2]. However, this method has a significant limitation. The inherent propensity of the polymerase is to exchange purine for purine and pyrimidine for pyrimidine (transitions). This leads to a bias in the composition of the library. This bias can be corrected by altering the ePCR to favor transversions. The use of unbalanced dNTPs (deoxy nucleotide triphosphates) on ePCR (Polymerase chain reaction) has been reported as a method to increase the proportion of transversions [2] over transitions. Transversions can introduce more variety on the library. Although, even without this bias, due to the inherent characteristics of the genetic code, most codons cannot be changed to encode for certain amino acids with a single mutation. Let us use the valine codon as an example. Single point mutations are capable of encoding phenylalanine, leucine, isoleucine, aspartate, or glycine but in order to access to other amino acids two point mutations or even three point mutations are required. With mutation rates in the order of 4-6/1000 bp two or three consecutive mutations are very unlikely.

Another aspect that needs to be considered when choosing a method to generate a library is the protein sequence space that can be covered by the chosen method. Protein sequence space is immeasurable: with 20 letters in the protein alphabet, there are $20^{300}$ ways to arrange together 300 aminoacids. For each sequence there are 300 x 19 sequences that differ in just one amino acid. Any two points are at most 300 mutational steps away. This protein space is also mostly empty of function. The number of sequences expected to encode a solution to a specific problem is an infinitesimal fraction of this space. It is crucial then to create a library that covers as much as possible of the protein sequence space. The success of Directed evolution is based on two steps: generating molecular diversity and identifying the improved variants. Mutations, introduced by error-prone PCR, gene shuffling or other methods, allow the descendent molecules to “evolve”, improving their binding or catalytic activities through several selection cycles. In addition to its technological value, in vitro selection offers the opportunity to ask general questions about the relation between structure and function in organic molecules.

**Part I**

**Taq polymerase**

Nucleic acids are important in current biotechnology, due to its ease of amplification based on the simple complementarity rules between the A-T(U) and G-C base pairs. They present versatile functionalities which allow them to be used as catalysts, probes, aptamers, etc. However, having only four nucleotides in standard nucleic acids restricts their functions. There is a growing interest in expanding the nucleic acids repertorie [3,4]. The specific complementarity of an unnatural base pair allows the site-directed incorporation of extra nucleotide analogs into DNA and RNA by polymerases. An unnatural base pair that selectively functions in replication and transcription facilitates the PCR amplification of the DNA fragments and synthesis of RNA molecules containing nucleotide analogs.

The first unnatural base pairs to be developed were: isoguanine and its pair isocytosine, (isoG-isoC) and xanthosine which pairs with diaminopyrimidine. These have different hydrogen-bonding patterns from those of the natural base pairs[5,6]. DNA fragments containing these unnatural base pairs had been subjected to PCR amplification[7,8] and a modified isoG had been incorporated into RNA by transcription using isoC-containing templates[9]. The pursuit of an expanded genetic alphabet was begun in 1989 by Steven Benner (Fig. 1) who demonstrated that unnatural base pairs with altered hydrogen-bonding topologies could be synthesized by DNA polymerases[10]. In 1997, Eric Kool and coworkers made the remarkable observation that hydrogen bonds are not an absolute requirement for the pairing of nucleotides during DNA synthesis[11]. The landmark works of Benner and Kool suggested that hydrogen bonds between the nucleobases could be rearranged or even replaced with other types of complementary interactions. Since Benner's original work, more than 50 unnatural base pairs have been reported in the literature[12]. Although a majority of these pairs are stable in duplex DNA, the enzymes that synthesize DNA (DNApolymerases) rarely recognize the unnatural analogs with the efficiency and fidelity needed for practical tasks, such as PCR.

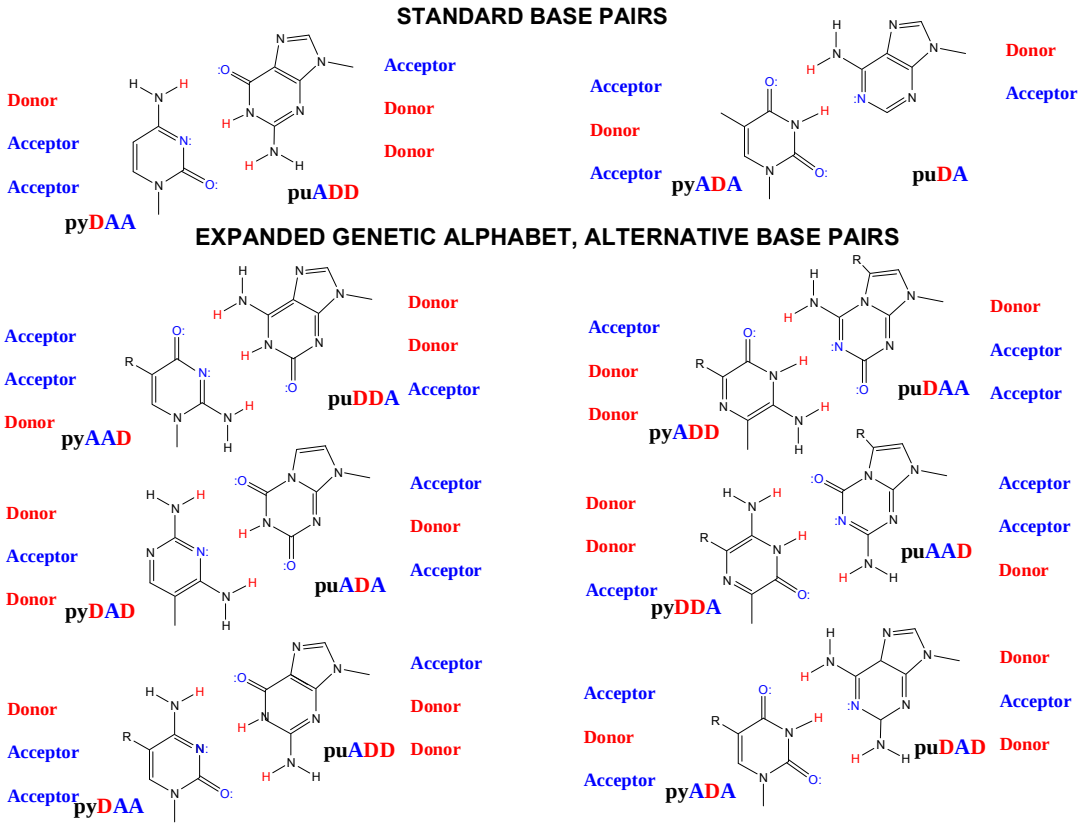


**Figure 1 The Artificially Expanded Genetic Information System (AEGIS).**

In 1999 Battersby[13] reported the first example of incorporation of a positively charged functional group into an in vitro selection experiment and the first quantitative comparison of a selection experiment with standard biopolymers and functionalized biopolymers of any kind. The results provide a remarkable example of convergent evolution of physical behavior in two molecular systems, as well as a view of the landscapes that relate structure and behavior in macromolecules. They noticed that when selection experiments were run with a uridine carrying a side chain, the oligonucleotide that emerged were substantially shorter than those obtained when standard nucleotides were used. These suggested that modest discrimination of the polymerase against the functionalized uridine, suggesting that there was room for improvement for polymerases used for in vitro selection.

In 1998 Tawfik and Griffiths[14] reported a method, compartmentalized self replication (CSR) for the cellular compartmentalization of bacteria taking advantage of the fact that the nature of cells makes them keep together both, the genes and proteins encoded by them, thus linking genotype to phenotype. These compartments can be used to select for genes. In 2001 Ghaddessy and coworkers[15] reported a method for doing such selections for polymerases. This method is based on a simple feedback loop consisting of a polymerase that encodes its own encoding gene. Each droplet will have one cell containing one copy of a variation of the *Taq* gene, during PCR the cell breaks and the primers can anneal the target *Taq* gene which encoded for the *Taq* polymerase (*Taq* pol). If this variant of *Taq* is active it can amplify its own

gene. These genes can be introduced in cells for another round of selection. This is illustrated in Figure 2.

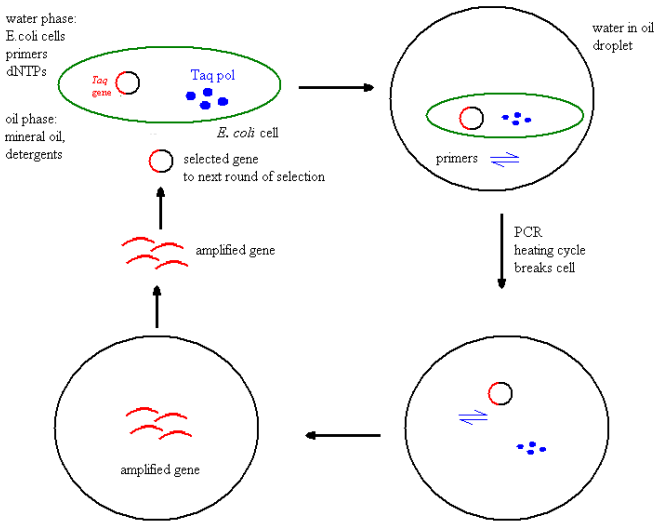


**Figure 2 Directed evolution scheme**

A library of genes is created and introduced in *E. coli* cells by electroporation. These genes express a mutant *Taq* pol enzyme (top left). Each cell is isolated in a water in oil emulsion (top right). The first heating cycle of PCR will break the cell wall leaving a droplet containing: primers, dNTPs, a mutated gene of *Taq* and the enzyme expressed by this gene (bottom right). After PCR if the mutated polymerase is active it will amplify its own gene enriching the pool of mutants with mutants of interest according to the selective pressure imposed by the experiment (bottom left). The genes are extracted and inserted in a plasmid vector (circular DNA) and then enter to the cycle of selection again. After repeating these cycles a enriched pool of variants of the original gene are produced.

The Holliger approach can be clearly adapted for polymerases that can be selected simply by presenting non standard nucleotides. In this work we aim to establish a robust system that allows us to evolve Taq polymerase in order to carry out selections for non standard nucleotides. Based on the CSR system developed by Tawfik and Griffiths we looked forward to develop a robust procedure that allows us to perform CSR cycles and select variants of Taq Pol under different selection pressures.

## Experimental

**Cell growth**.- 74 known codon optimized mutants provided by the company DNA 2.0 were growth by inoculating LB media (100mL). Cells were spinned down and washed with Tris buffer. **Emulsions**.- Emulsions were made by mixing 220 µL of aqueous phase with 400 µL of the arlacel solution. Emulsions were made by using a stirring bar in a polypropylene tube. Stirring plate was set to 1000 rpm. When using the stirring bar the aqueous phase was added

in 40uL aliquots every 5 seconds. **PCR**.- Once the emulsions were made they were transferred to PCR tubes and the following cycles were performed: 94 °C - 5 min; 94 °C - 1 min; 55 °C - 1 min; 72 °C, 3 min; go to step 2, 14 times; 72 °C - 7 min. **Extraction.-** Emulsions were broken with water saturated ether. DNA was purified using Promega kit, eluted with pure water. **Ligation**.- Vector was obtained by purifying the wild type plasmid and digesting it with NcoI and SacII. After digestion, the vector was separated in an agarose gel, cut under long wave UV light and treated with Antarctic phosphatase. The recovered DNA from emulsions was digested and ligated into the vector described above. The ligation product was purified with promega kit and used for electroporation. Colonies obtained from this transformation were used to grow cultures and obtain plasmids for sequencing.

**Results and discussion**

A preliminary experiment was designed using known mutants of *Taq* polymerase which are codon-optimized[3] by the company DNA2.0. This first cycle of directed evolution was intended to be a test run. It was expected to recover the known mutants that were introduced in the pool. There was no selective pressure other than to be able to produce copies of its genome by PCR in emulsions. One of the recovered sequences matches with mutant 5399, a mutant active at lower temperatures. Another sequence presents two mutations (575, 583) that are both present in mutant 5347 and the other three mutations (657, 664, 740) are present in mutant 5421. A third sequence presents the three mutations found in mutant 5347 and one that is not part of the pool: P250L. The fourth sequenced plasmid presents two mutations listed for mutant 5350 and three other mutations not listed in the original pool of mutants.

**Table 1**

**Mutations found in the samples sequenced. Mutations in red are present in the pool of mutants provided by DNA 2.0**

| Sample | Mutations: | | | | | |
|---|---|---|---|---|---|---|
| A | K537I | L606P | A740S | E742H | | |
| B | D575F | V583K | R657D | F664Y | A740R | |
| C | P250L | D575F | V583K | M670A | | |
| D | E260G | R428H | F595W | L601P | D622S | M643V |

These results suggest that during the annealing cycles of PCR more than one mutant (more than one cell) is present in the water droplet and they are acting as templates introducing

[3] A codon-optimized gene is one that utilizes codons that use the t-RNA more abundant in the organism that carries the gene. This codon optimization ultimately enhances the expression of the target protein.

mutations. That explains how some of these recovered mutants are chimeras of the known mutants. The new mutations not present in the pool of mutants can be product of the inherent mutational rate characteristic of *Taq* polymerase which can increased due to the components of the emulsifiers. Future work will explore the production of more robust emulsions to avoid leaking of plasmids into other droplets or avoiding having more than one cell per droplet. In any case we can predict that certain level of background and false positives will occur. The fact some mutants from the original pool can be recovered show that this system has a minimum requirement for future experiments.

**Part II**

**Lac repressor**

The *lac* repressor of *E. coli* regulates the bacterium's metabolism in response to environmental conditions. The *lac* operon is a set of genes and regulatory elements that controls the expression of the proteins needed in order to change the metabolism on *E. coli* to use lactose when glucose is absent in the environment. Gene expression is turned on when the *lac* repressor (LacI), a protein product of the *lacI*[♣] gene binds an inducer molecule, allolactose. This binding event triggers the expression of the genes necessary for lactose metabolism. The term operon was coined by Jacob and Monod who first studied it in 1961[16]. The operon consists of three genes: *lacZ*, *lacY*, and *lacA*, which encode for the proteins required for lactose metabolism. Upstream of these there is a promoter, an operator sequence and the *lacI* gene, which codes for the *lac* repressor protein (LacI) which is transcribed independently. The operator is the binding site for the Lac repressor. A byproduct of the lactose metabolism, allolactose, binds the *lac* repressor producing an allosteric change on its conformation. As a consequence, the repressor loses its affinity for the operator sequence. Then LacI leaves the operator, repression is relieved and RNA polymerase is free to transcribe the polycistronic *lacZYA* genes[17] (Fig. 3).

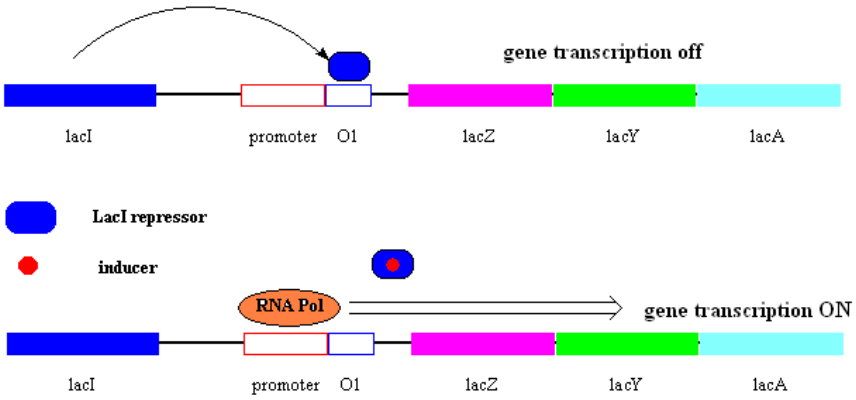


**Figure 3 Scheme of the lactose operon.**

♣ LacI refers to the protein expressed by the gene *lacI*, genes are written in italics.

The repressor protein (LacI) is encoded by the *lacI* gene. The repressor is transcribed independently and binds a region of the operator (O1) upstream of the genes *lacZYA.* Binding to the operator hinders the transcription by RNA polymerase. When glucose is depleted and lactose is available, allolactose binds the repressor producing an allosteric change which reduces its affinity for the operator. Then RNA polymerase is free to transcribe the polycistronic *lacZYA.*

One interesting trait of the *lac* repressor is that it is relatively flexible for accepting different molecules as inducers. It is known to accept other inducer molecules such as IPTG[4] as well as other sugars. In addition, LacI belongs to a larger family of regulators, the GalR and LacI family[18]. These proteins, which regulate transcription in prokaryotes, have homologous sequences[18] and accept a surprisingly broad range of functionalities as inducer molecules. These include sugars, nucleotides, phosphorylated sugars and even purines. Furthermore GalR and LacI have regions of homology with three sugar-binding periplasmic proteins[18]. The goal of this project is to produce a variant of the Lac repressor capable of recognize molecules such as 2-O-Methyl-α-D-N-acetylneuraminic acid, b-methyl mannopyranoside, deoxygalactonojirimycin and other glycons that require laborious procedure to be synthesized. Having a repressor capable of recognize one of this molecules with high specificity could be used as a tool for identifying enzymes that can synthesize these molecules. To take advantage of the fact that LacI shares homology with a series of homologous regulators one of the main goals of this project is to produce a chimera of genes by recombination techniques. One of these techniques is gene shuffling. It consists on assembling small fragments of genes from the different members of the family to produce a new gene. The expected outcome is a gene that will express a protein that contains motifs present in the different members of the family. One combination of these motifs might be able to recognize the target molecules. Part of the alignment of these protein sequences is presented on Fig.4. Screening for mutants use the enzyme β-galactosidase, expressed by *lacZ,* and detected by X-gal which decomposes giving an intense blue color.

[4] isopropyl β-D-thiogalactosidase

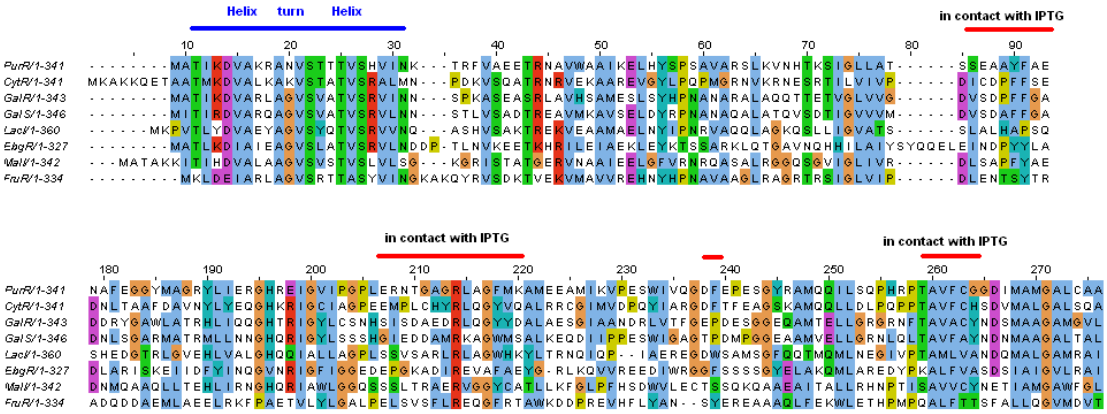


**Figure 4** Alignment of proteins of the GalR-LacI family and the inducers that recognize. Left panel: alignment of different repressors that belong to the LacI family. The proteins are present in *E. coli*. Protein sequences were obtained from EcoCyc, a database for the bacterium *Escherichia coli* K-12 MG1655. Multiple sequence alignment was done with ClustalW. These proteins are: PurR, purine repressor; CytR, cytidine repressor; GalR, galactose repressor; GalS, galactose isorepressor; LacI, lactose repressor; EbgR, evolved β–galactosidase repressor; MalI, maltose repressor and FruR, fructose repressor. Adapted from Weickert and Adhya[18].

## Experimental

**Error prone PCR (ePCR)** a basic reaction used: 50 mM KCl, 10 mM Tris HCl pH 8.3, 1.5 mM $MgCl_2$; 5ng of template (pet28); dNTPs: 0.2 mM each; primers: 10 µM each; *Taq* 10 U. Variations of this reaction had in addition: 150 µM $MnCl_2$ to induce a higher rate of mutations or 500 µM of $MnCl_2$ and a 10x excess of dTTP or dCTP to increase the transversions as discussed above. Primers were obtained from Integrated DNA Technologies. Enzymes were obtained from New England Biolabs (NEB). Sequencing was done at the UF sequencing core. **Electrocompetent cells** were made by growing XL10-Gold cells (Stratagene) in low salt (SOB) media until the cells reached $OD_{600}$ of 0.8.Cells then were harvested by centrifugation, washed with pure and sterile water and then transferred into a sterile dialysis membrane and dialyzed against 1 L of ice-cold ultra-pure water, changing the water every 30 min. for a total of 4 L in water changes. Cells were finally centrifuged at 2000 x g for 10 min at 0° C and resuspend in 6 mL 15% v/v glycerol, 2.5 % w/v sorbitol solution that has been sterile filtered with a 0.2 micron filter. Cells were stored at –80° C until use. **Gene shuffling:** to amplify the genes from the LacI family one colony of *E. coli* K-12 was resuspended in 20 µL of water and an aliquot was used as a template. PCR program:1. 94 ºC 1min; 2. 94 ºC 1min; 3. 51 ºC 1min; 4. 72 ºC 2min; 5. go to 2 x 29times; 6. 72 ºC 7min. **Controlled digestion:** a DNaseI solution was prepared: 450 µL water; 45 µL 10x NEB buffer for DNaseI; 5 µL MnCl2; 1 µL DNaseI (2U), NEB. Then 100 µL gene mix (aprox. DNA concentration: 50 µg/mL ) + 43 µL of DNaseI solution were mixed on ice. Timing started placed at 15 ºC. The reaction was

terminated by placing the tubes in the 90 ºC for 10 min. The digested DNA produces an smear on agarose gels as reported on the literature. Fragments between 750-400 bp and between 500-300 bp were sliced from the gel and purified using silica columns following the QUIAGEN kit protocol and eluted with pure water. **PCR without primers:** 50 µL of the gel purified fragments were mixed with 1.1 µL dNTPs (10 uM each), 5.8 µL of 10x PCR buffer and 1 µL of Taq polymerase and placed on the following heating cycle: 1. 96 ºC 90 sec; 2. 94 ºC 30 sec; 3. 65 ºC 90 sec; 4. 62 ºC 90 sec; 5. 59 ºC 90 sec; 6. 56 ºC 90 sec; 7. 53 ºC 90 sec; 8. 50 ºC 90 sec; 9. 47 ºC 90 sec; 10. 44 ºC 90 sec; 11. 41 ºC 90 sec; 12. 72 ºC 4min; 13. go to 2, 34 times; 14. 72 ºC 7min. **Amplification of full length sequences: a**t the end of the previous step 1 µL of a 10-fold dilution of the reassembly products were used to amplify a full length product using the CytR primers. For a 100 µL reaction: 10 µL of 10X PCR buffer; 4 µL of each primer 10 µM; 1 µL reassembly products (product of the PCR without primers); 2 µL dNTPs (10 µM each); 79 µL water.

## Results

### *E. coli* DH1 (Δ*lacI*) strain for blue-white screening.

Since screening experiments will try to identify mutants of the *lac* repressor protein, it is important to use a host that does not produce the wild type protein. Following the Wanner method for the inactivation of chromosomal genes [19], primers were designed with overhangs homologous to the flanking regions of the *lacI* gene. The other end of the primers anneal on plasmid pKD4 amplifying the KanR gene. Special care was taken when designing these primers to avoid disruption of operator regions which overlap with *lacI* and *lacZ*. Chemically competent DH1 cells were transformed with the pKD46 vector. Transformants carrying this vector were made electrocompetent and subsequently transformed with the PCR product of the primers mentioned above using the pKD4 plasmid as template. This PCR product contains the KanR[5] marker flanked by a region homologous to the bacteria's genome. After this last electroporation step a 4h recovery followed at 30 ºC. During the last 3h the cells were grown in presence of L-(+) arabinose. This compound induces the lambda Red recombination proteins encoded by plasmid pKD46. It was observed that this long recovery period was critical for obtaining viable cells when plated on agar/kanamycin plates. Since the *lac* repressor protein is absent, the *lacZ* gene can freely express the β–galactosidase enzyme. Figure 5 shows the result of this modification on the genome of DH1 *E. coli.*

[5] KanR Gene that confers resistance to the antibiotic kanamycin.

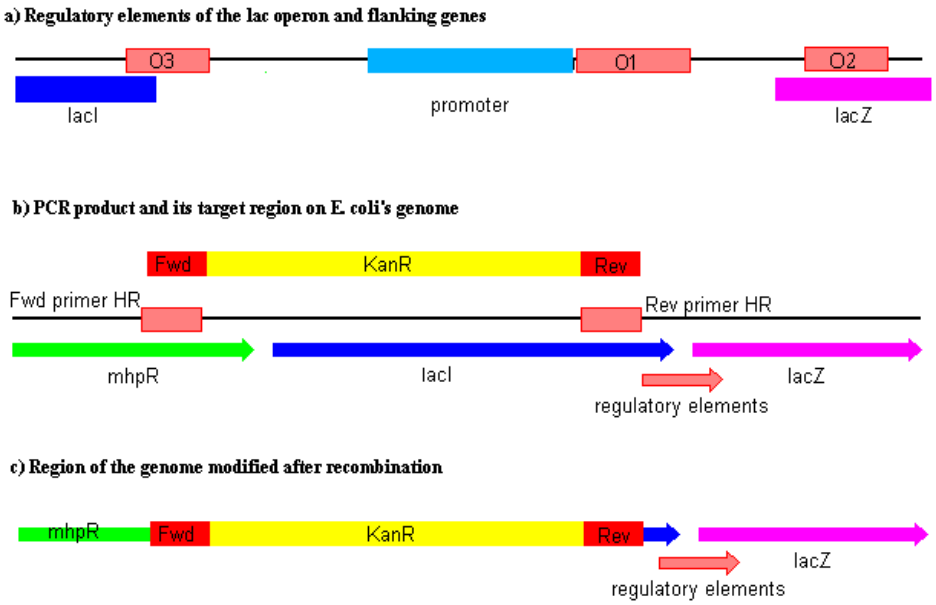


**Figure 5 Inactivation of the chromosomal lacI gene**

a) Some of the relevant regions from the lac operon. Operator 3 (O3) overlaps with *lacI* gene, O2 overlaps *lacZ*. The strongest binding occurs with O1. The primers were designed to inactivate the *lacI* gene without disrupting these elements. b) The kanamycin resistance gene (KanR) is produced by PCR amplification of plasmid pkD4. The primers are flanked by sequences homologous to certain sections of the genome (HR) shown in red boxes. The recombination proteins expressed by the plasmid pkD46 incorporate the PCR product into the chromosome. c) After the recombination step KanR gene is incorporated into the chromosome. It is shown that part of the gene *mhpR* is deleted, the region of the *lacI* gene which contains the O3 remains intact as well as the other regulatory elements and the *lacZ* gene.

The desired transformants were identified for producing blue colonies in presence of X-gal[6]. Finally, the plasmid pKD46 was cured by growing the cells at 37°C. When transformed with a high copy number vector encoding LacI, white colonies are produced. Sequencing of the altered region of the genome of the modified host shows that the operators and promoter binding regions are intact which supports the experimental results observed.

### Blue-white screening

Figure 6 (top) illustrates how a DH1 (Δ*lacI*) *E. coli* cell is used for the blue-white screening. This strain will be used to identify mutants sensitive to a new inducer molecule. After electroporation, each cell carries a plasmid vector with different mutations on the *lacI* gene. If the repressor protein encoded by the mutated gene is sensitive to a new inducer, it will produce the dissociation of the repressor from the operator.

[6] X-gal: 5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside. In presence of the enzyme β-galactosidase, expressed by *lacZ*, it decomposes giving an intense blue color.

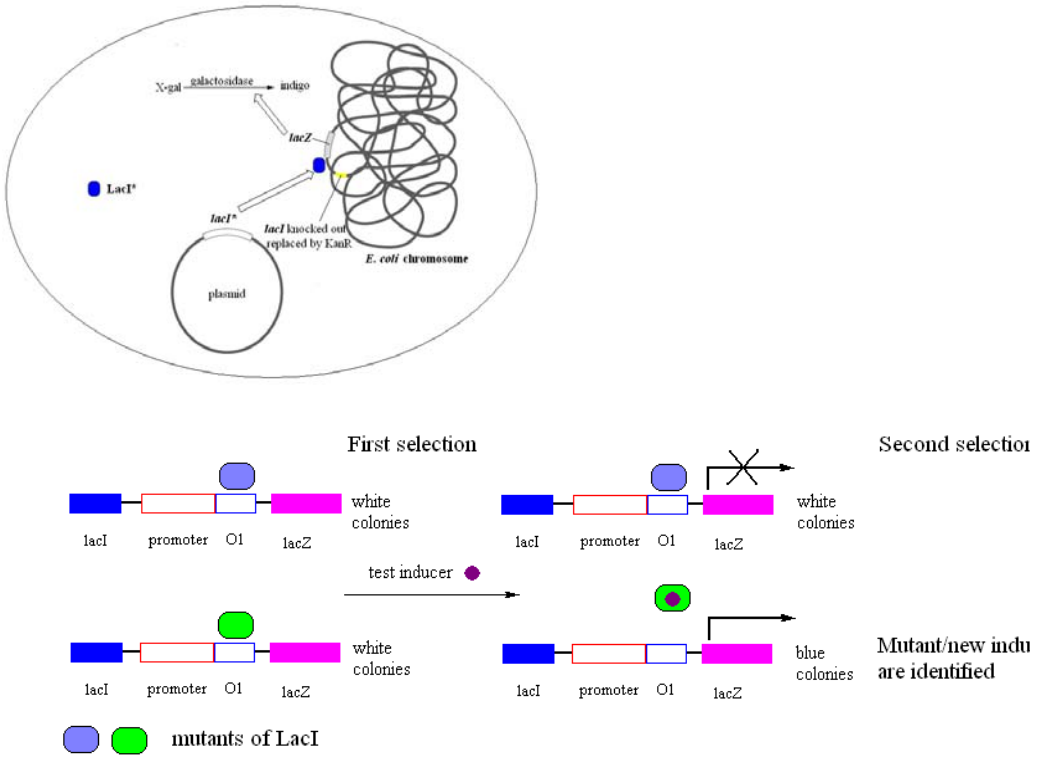


**Figure 6 Blue-white screening**

**Top**: Scheme of an *E. coli* cell were the chromosomal *lacI* gene has been deleted. LacI repressor encoded by a mutated gene (*lacI**) can interact with the lac operator. The nature of the interaction will produce either a blue or white colony. Bottom: Screening strategy. Transformants that contain a mutant of *lacI* that encodes for a repressor protein, sensitive to the inducer molecule tested, will dissociate from the operator and allow the expression of β-galactosidase. In presence of X-gal, colonies that carry such repressor generate blue colonies. If the inducer molecule has no effect on the affinity of LacI for the operator, β-galactosidase is not expressed and the colonies remain white.

### Gene Shuffling

Gene shuffling uses recombination to produce chimeras of genes. Starting from a collection of genes from the same family the genes are fragmented randomly and re-assembled by annealing progressively at different temperatures. This approach has an enormous advantage over other methods since it takes advantage of already evolved fragments that come from different genes. Figure 7 shows a successful sequence for gene shuffling: a mixture of genes from the LacI family were cut randomly with DNaseI and reassembled producing a chimera of genes.

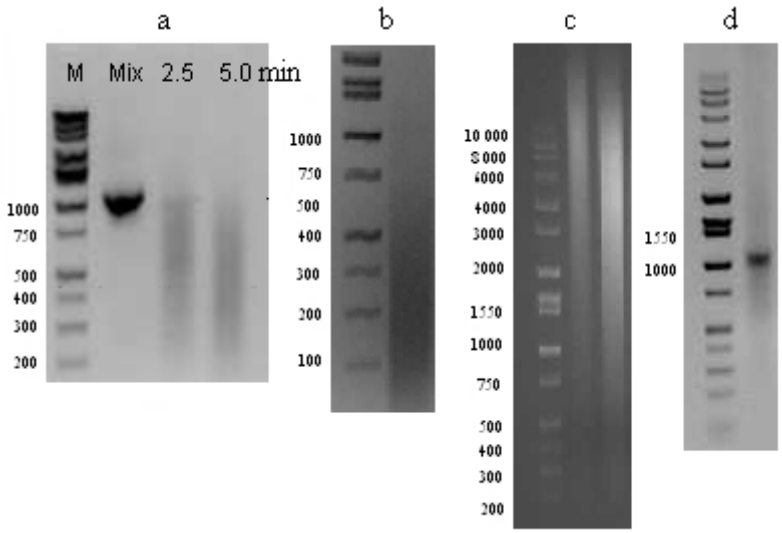


**Figure 7 Gene shuffling of the genes of the LacI family**

**a.** Mixture of genes of the LacI family, the two right hand lanes show the product of digestion with DNaseI after 2.5 and 5 minutes respectively, 2% agarose gel. **b.** Product of large scale digestion with DNaseI. **c.** Product of PCR without primers. Fragments anneal to each other based on complementary on their sequences. A smear including products of higher molecular weight than the parent gene are typical. **d.** Final extention. An aliquot of the previous step and primers for the parent gene are used on PCR to produce a full size product of length comparable to the parent gene.

**Variant of LacI found**

Using the tools presented above a variant of LacI was identified. Starting from a library of mutated *lacI* genes was created using error prone PCR (average 6 mutations per 1000bp). Which was electroporated into the *E. coli* DH1 (Δ*lacI*) strain. Screening of a few hundred colonies lead us to find a mutant of LacI which is not responsive to the anti-inducer phenyl β-D-galactoside. When LB-X-gal plates were coated with a 50 μM water solution of this anti-inducer, blue colonies were observed. Sequencing of this mutant showed three mutations: one silent mutation, one were Asp 275 is changed for Gly and another on Val 271 changed to Asp. These aminoacids have been previously reported as intolerant to substitutions by the comprehensive work of Suckow[20]. Moreover, the substitution on the aminoacid 275 is adjacent to Asp274 which has been reported to be in contact with the inducer well-studied inducer IPTG by Friedman[21]. This preliminary results show that the ongoing effort can lead us to find to identify mutants of the Lac repressor. However, this blue-white screening strategy does not allow us to explore a large number of mutants. A selection based on the viability of the cells where we would select for cells that survive the selection would speed up the process.

**Acknowledgments**

The author thanks Dr. Steven Benner from the Foundation For Applied Molecular Evolution (FFAME) for his continued guidance and support through these years with the polymerase project. For the Lac repressor project the author thanks the Chemistry Department at the University of Florida and to Dr. Nicole Horenstein for her guidance.

**References**

1. Yuan, L.; Kurek, I.; English, J.; Keenan, R. *Microbiology and Molecular Biology Reviews* **2005**, *69*, 373-+.
2. Vartanian, J. P.; Henry, M.; WainHobson, S. *Nucleic Acids Research* **1996**, *24*, 2627-2631.
3. Henry, A. A.; Romesberg, F. E. *Current Opinion in Chemical Biology* **2003**, *7*, 727-733.
4. Moser, M. J.; Prudent, J. R. *Nucleic Acids Research* **2003**, *31*, 5048-5053.
5. Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33-37.
6. Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489-10496.
7. Sismour, A. M.; Benner, S. A. *Nucleic Acids Research* **2005**, 33, 5640-5646.
8. Sismour, A. M.; Lutz, S.; Park, J. H.; Lutz, M. J.; Boyer, P. L.; Hughes, S. H.; Benner, S. A. *Nucleic Acids Research* **2004**, *32*, 728-735.
9. Tor, Y.; Dervan, P. B. *Journal of the American Chemical Society* **1993**, *115*, 4461-4467.
10. Switzer, C.; Moroney, S. E.; Benner, S. A. *Journal of the American Chemical Society* **1989**, *111*, 8322-8323.
11. Moran, S.; Ren, R. X. F.; Rumney, S.; Kool, E. T. *Journal of the American Chemical Society* **1997**, *119*, 2056-2057.
12. Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C. Z.; Geierstanger, B. H.; Romesberg, F. E. *Journal of the American Chemical Society* **2004**, *126*, 6923-6931.
13. Battersby, T. R.; Ang, D. N.; Burgstaller, P.; Jurczyk, S. C.; Bowser, M. T.; Buchanan, D. D.; Kennedy, R. T.; Benner, S. A. *Journal of the American Chemical Society* **1999**, *121*, 9781-9789.
14. Tawfik, D. S.; Griffiths, A. D. *Nature Biotechnology* **1998**, *16*, 652-656.
15. Ghadessy, F. J.; Ong, J. L.; Holliger, P. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, **98**, 4552-4557.
16. Jacob, F.; Monod, J. *Journal of Molecular Biology* **1961**, *3*, 318-&.
17. Hu, M. C. T.; Davidson, N. *Cell* **1987**, *48*, 555-566.
18. Weickert, M. J.; Adhya, S. *Journal of Biological Chemistry* **1992**, *267*, 15869-15874.
19. Datsenko, K. A.; Wanner, B. L. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 6640-6645.
20. Suckow, J.; Markiewicz, P.; Kleina, L. G.; Miller, J.; KistersWoike, B.; MullerHill, B. *Journal of Molecular Biology* **1996**, *261*, 509-523.
21. Friedman, A. M.; Fischmann, T. O.; Steitz, T. A. *Science* **1995**, *268*, 1721-1727.