High-Throughput synthesis and cloning of genes encoding venom peptides: developing a platform for the discovery of novel therapeutic molecules

Ana F. Sequeira[1,2], Joana L. A. Brás[1,2], Catarina I.P.D. Guerreiro[2], Renaud Vincentelli[3], Carlos M.G.A. Fontes[1,2]

1. CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; 2. NZYTech Lda, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E-R/C, 1649-038 Lisboa, Portugal; 3. Architecture et Fonction des Macromolécules Biologiques (A.F.M.B), UMR7257 CNRS, Université Aix-Marseille,Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex, France.

Background

Venomous animals have developed an arsenal of small-reticulated proteins (peptides), generally termed venom peptides, which target with high selectivity and efficacy a variety of biological molecules thus playing a critical role in defense and predation. The remarkable potency and pharmacological diversity of animal venoms has made them an increasingly valuable source of lead molecules for drug and insecticide discovery. Nevertheless, the majority of the structural diversity harnessed by these venoms remains uncharacterized, in part due to difficulties in producing these molecules recombinantly. It is estimated that in the world of venoms and toxins, there are more than 40,000,000 molecules awaiting discovery and characterization while only 3000 venom peptides are known presently.

Motivated by the advancement of genomics and the urgent need of discovery of new molecules with therapeutic interest, as are venom peptides, high-throughput pipelines (HTP) have been developed to replace traditional approaches and protocols for synthetic gene synthesis and protein production. Several high throughput platforms have been used in the past decade to identify conditions for soluble protein expression or for synthetic DNA production in large scale*. As a member of the FP7 European Venomics project (www.venomics.eu), our challenge was to develop a roboust gene synthesis automated method with the aim of producing thousands of synthetic genes encoding novel venom proteins. Here we present the most important characteristics of this innovative platform. The platform was designed to synthetize and cloning multiple genes into a prokaryotic expression vector through a ligation-independent cloning protocol. A total of 4992 genes that represent untapped venomic diversity were synthesized using oligonucleotides assembled by PCR. Synthetic genes were cloned into a prokaryotic expression vector, pHTP4, which contains the fusion tag DsbC to promote peptide folding. The encoded recombinant peptides contain an internal His tag and a TEV (Tobacco Etch Virus) recognition cleavage site to allow removal of appended tags from the toxins after recombinant expression in. NZYTech automated platform for gene synthesis allows obtaining thousands of genes in weeks with a 99,6% by screening a maximum of three clones per gene.

I. Developing of HTP gene synthesis platform

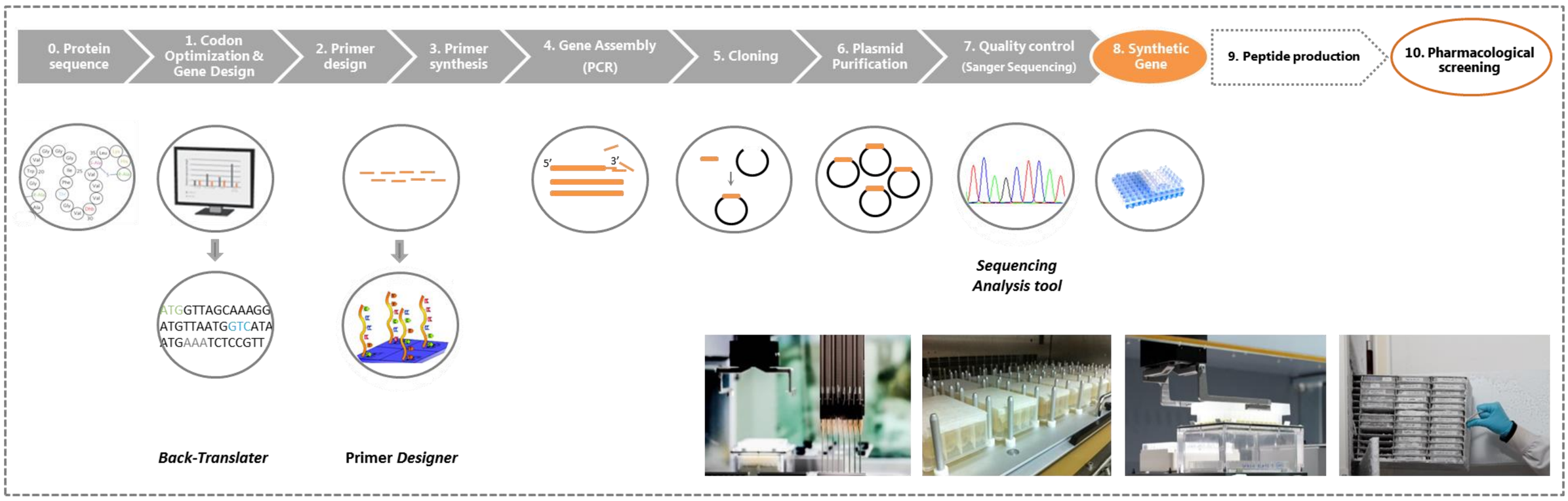

Figure 1. An HTP gene synthesis platform was developed to produce multiples of 96 synthetic genes enconding venomic peptides. This pipeline includes 7 steps that allow the successful synthesis of multiples of 96 genes. The first step corresponds to Codon Optimization and gene design ; from protein sequences multiples DNA sequences are designed and optimized for expression in *E. coli*, using NZYTech algorithm of codon optimization software (*BackTranslater*). In steps 2, 3 and 4 oligonucleotides required for gene assembly are designed ,synthesised and assembled by PCR using optimal conditions. Synthetic genes are cloned using NZYTech LIC protocol into the *E. coli* expression vector pHTP4. Bacterial transformation and DNA preparations are accomplished using high throughput protocols. DNA sequences are checked for the presence of sequence errors using the Sequencing Analysis tool. All steps are automated using a liquid handling robot.

II. HTP Gene synthesis of 5,000 genes encoding venomic peptides

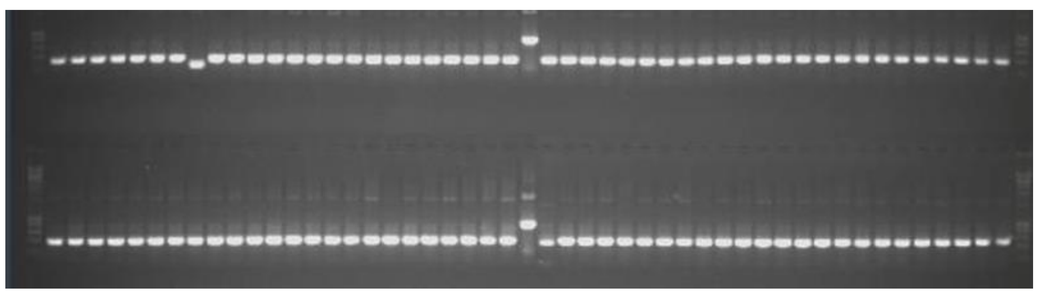

Figure 3. Agarose gel electrophoresis of genes synthesized following NZYTech HTP gene synthesis protocol. Multiples of 96 synthetic genes were synthesized by assembly of unpurified oligonucleotides, using optimal PCR conditions.

Figure 2. Construction of the expression plasmids by ligation-independent cloning. All plasmids carry a T7 promoter/terminator, a 6xHis tag (blue) for nickel affinity purification, a *Tobacco Etch Virus* (TEV) cleavage site, a DsbC fusion partner and a gene encoding a venom protein (VP). The cloning region are represented in green box.

Table 1. A total of 4,992 genes were designed from 4,992 venomic peptides sequences. The DNA sequences were optimized for recombinant expression in *Escherichia coli*, using the software *BackTranslater*. Oligonucleotides for gene assembly were designed with *Primer designer* tool.

Properties of Genes Designed				
	Length (bp)	GC content (%)	Number of primers	Codon Adaptation Index (CAI)
Mean (± SD)	220 ± 54	49 ± 4	6 ± 1,6	0,92 ± 0,04
Maximum	413	58	10	0,94
Minimum	137	42	4	0,8

Figure 4. Efficiency of HTP gene synthesis platform for production of genes encoding venom proteins. We have selected 4,992 genes representing the animal venom diversity. After gene synthesis and cloning into expression vector, we checked three clones of each gene to detected the presence of errors in DNA sequences. From initial sample, 3,941 genes are positives when was screened only 1 clone, for 809 genes were required 2 clones and for 242 genes were screened 3 clones.

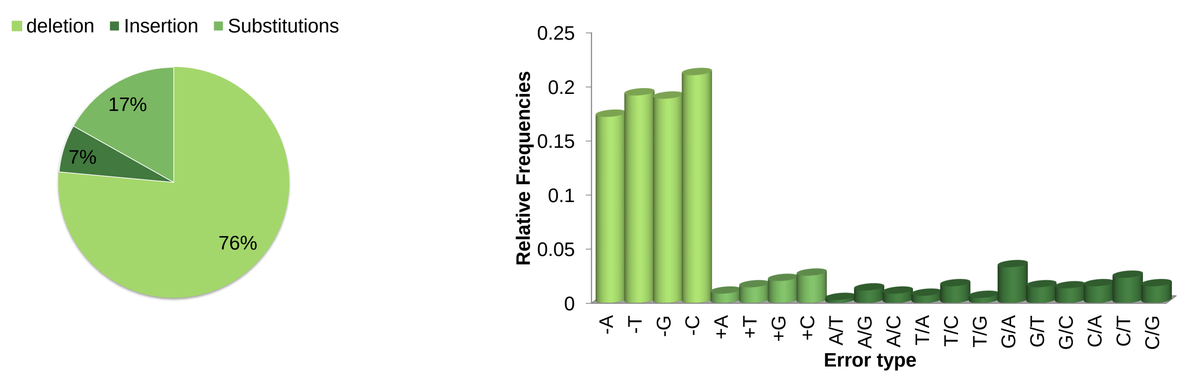

Figure 5. Distribution of observed errors in synthetic genes. The deletion is the most common error observed in genes with an mutation event.

Conclusions

- NZYTech HTP gene synthesis pipeline provides a quick and accurate alternative for the rapid design and assembly of synthetic genes. This HTP platform is suitable for diverse applications in synthetic biology.
- The most common type of sequence errors identified in synthetic genes were deletions and insertions, being the deletion of one nucleotide C the most frequent.
- The presence of sequence errors might be linked to oligonucleotides quality; primers used for gene assembly were not purified and may include truncated versions that reduce the efficacy of the method.
- NZYTech HTP gene synthesis protocols was used to produce 4,992 genes enconding venomic peptides with a success rate of 1,3, meaning that 1,3 clones need to be screened to obtain a positive one.

REFERENCES:
[1] Escoubas P, King GF (2009) Venomics as a drug discovery platform. Expert Rev Proteomics 6 (3), 221-224.
[2] Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutiérrrez JM (2009) Venoms, venomics, antivenomics. FEBS Letters 583:1736-1743.
[3] Escoubas P, Sollod B, King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon 46 (6), 650-663.
[4] Nozach H *et al*. (2013) High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. Microbial Cell Factories, 12:37.

ACKNOWLEDGEMENTS:

This work was supported by the Seventh Framework Program (FP7 HEALTH):2011-2015.