DOI 10.18699/SBB-2022-671

GeneCut – a software tool for oligonucleotide design, assembly and cloning of gene constructs

Sukhomlinov D.[1]*, Pereverzev I.[2], Bondiuk S.[1], Seryakov A.[1]

[1] *NCIT UNIPRO, LLC, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

* *dsukhomlinov@unipro.ru*

Key words: oligonucleotides, codon optimization, gene assembly, cloning

Motivation and Aim: The synthesis of long nucleic acids is of great importance for modern molecular biology and genetic engineering, because it allows one to artificially create fragments with a given nucleotide sequence without using a DNA template. However, at present, it is technologically possible to synthesize only short DNA fragments, called oligonucleotides, which are subsequently assembled into a long DNA chain. But, when performing assembly *in vitro*, oligonucleotides in a mixture can form not only the necessary target, but also numerous non-target complexes – hairpin loops, self- and heterodimers. Computer modeling of oligonucleotide complexes makes it possible to provide the absence of stable non-target complexes and choose the optimal set of input data without carrying out experiments on real molecules.

The aim was to develop a tool for *in silico* experiments on gene splitting and molecular cloning. The tool should allow one to split the input target sequence into a set of oligonucleotides, that could be used to the further assembly the original sequence using the Polymerase Chain Reaction technology.

Methods and Algorithms: Unipro UGENE [1] is a desktop multiplatform software package that integrates dozens of widely used bioinformatics tools. Its core has been used as the basis for the technical implementation of a new product called GeneCut.

To split into fragments, you need to specify one long sequence and from one to four steps of the experiment, which are executed sequentially.

In the first step, "Optimize codon context", the sequence is modified in such a way that the fraction of occurrence of each codon in this sequence is as close as possible to the codon fraction values of the target organism from the corresponding codon frequency table.

At the second step "Exclude restriction and splicing sites", the chosen sites are replaced by sequences of the same length that do not contain these sites. The important point of the first two steps is to preserve the resulting amino acid sequence.

The third step “Long fragment assembly” is intended for situations when the input nucleotide sequence is long enough (about several thousand bases) sequences. In this case, first, splitting into long blocks is carried out, and then each block is split into short oligonucleotides. This step models the splitting for subsequent assembly by *Gibson assembly (GA)* [2] or *Overlap extension polymerase chain reaction (OE-PCR)*[3] methods. The difference between them is that the splitting by the *GA* method is carried out into blocks of a given length, and by the *OE-PCR* method - into blocks with specified temperatures of the sticky ends.

At the fourth step, “Oligonucleotides assembly”, the splitting of the input sequence into a set of short oligonucleotides is modeled for subsequent assembly by *Polymerase*

cycling assembly (PCA) [4] or *Thermodynamically balanced inside-out (TBIO)* [5] methods. Their difference lies in the fact that when splitting by the *PCA* method, oligonucleotides are arranged in turns on opposite DNA strands, while when splitting by the *TBIO* method, they are located from the center on a direct strand towards the 5'end and at the reverse complementary towards the 3' end.

During cloning, the algorithm selects oligonucleotide primers corresponding to the specified melting temperature parameters, which allow completing and combining the specified vectors and/or fragments into a single construct. The DNA cloning methods Gibson Assembly and Gateway Technology [6] have been implemented.

Results: GeneCut has been used for *in silico* splitting and cloning of target genes. The "Codon optimization" feature has been used for: IL2 gene: wild type (403 nt),eukaryotic optimized (403 nt), prokaryotic optimized (426 nt) EPO gene: eukaryotic optimized (582 nt), prokaryotic optimized (603 nt).

The "Long fragments assembly" *OE-PCE* method has been used for splitting of bacteriophage MS2 RNA-dependent RNA polymerase gene (1323 nt). This gene has been split into 2 fragments of 668 nt and 699 nt.

The "Oligonucleotides assembly" Gibson assembly method has been used for splitting of the gene 5.5 kilobyte size (private data) into 8 fragments of 730 nt each of them.

The resulting oligonucleotides have been synthesized, gene constructs have been assembled, cloned into a plasmid, and sequenced.

Conclusion: A software tool GeneCut has been developed. This tool contains a set of algorithms that model the splitting of a DNA sequence into short oligonucleotides for their subsequent assembly. The efficiency of assembly from the obtained oligonucleotides was confirmed experimentally.

Acknowledgements: This project has been founded by the Candidate of chemistry Shevelev Georgy Yurevich, who tragically died in July 2021. The scientific component of this work is the result of his long work.

References

1. Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012;28:1166-1167. doi: 10.1093/bioinformatics/bts091.
2. Gibson D., Young L., Chuang R.Y. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 2009;6:343-345. doi: 10.1038/nmeth.1318.
3. Horton R.M., Ho N.S, Pullen J.K., Hunt H.D., Cai Z., Pease L.R. Gene splicing by overlap extension. *Methods Enzymol*. 1993;217:270-279. doi: 10.1016/0076-6879(93)17067-f.
4. Stemmer W.P.C., Crameri A., Ha K.D., Brennan T.M., Heyneker H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene.* 1995;164:49-53. doi: 10.1016/0378-1119(95)00511-4.
5. Gao X., Yo P., Keith A., Ragan T.J., Harris T.K. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res*. 2003;31(22):e143. doi: 10.1093/nar/gng143.
6. Katzen F. Gateway(®) recombinational cloning: a biological operating system. *Expert Opin Drug Discov*. 2007;2(4):571-589.