ARTIFICIAL GENE SYNTHESIS *IN VITRO*

[1]Aamir Javed, [2]Muralidhar, T.S., [3]Manish Jaiswal & [1]Faiz Ahmad
[1]Centre for R &D in Life Sciences; Biotechnology Research Laboratory, #Post Graduate-Department of Biotechnology Dayananda Sagar Institutions, Dr.C.D.SagarCentre for Life Sciences, Shavige Malleshwara Hills,Kumaraswamy Layout,Bangalore-560078, India.
[3]Patna Science College, Biotechnology Research Laboratory, #Post Graduate-Department of Biotechnology, Patna University, Ashok Raj path, Patna-800006, India.

ABSTRACT
Gene of interest is rare in nature, or the genetic material containing such genes is difficult to obtain or isolate, *de nova* chemical synthesis of complete genes is a preferred method of cloning. One can now envisage synthesizing not only a new gene, but also a whole library of genes or mutants of a single gene for bioengineering applications, structural studies, drug development, and combinatorial biology and so on. The major technological driving force of this field is to make increasingly long stretches of error free DNA at reduced cost and also on improving the design of genes for specific purposes. This can be achieved by synthesizing oligos with overlapping ends, then annealing these oligos and extending to form a full-length gene. But due to the inherent nature of oligo synthesis, mutations, especially deletions occur frequently during the synthesis process. This situation dramatically limits the gene size and increase the cost of gene synthesis. To overcome this problem, a novel gene synthesis platform has been developed to efficiently synthesize the gene of larger size with no mutation. The synthetic gene can be optimized for expression and constructed for easy mutational manipulation without regard to the parent genome. Our program requires simple input information *i.e.* amino acid sequence of the target protein and melting temperature (needed for the gene assembly) of synthetic oligonucleotides. The program output a series of oligonucleotides sequences with codon optimized for expression in an organism of choice. Those oligonucleotides are characterized by highly homogenous melting temperatures and a minimize tendency for hairpin formation. With the help of this program and a two step PCR method, synthetic genes more than 1000bp can constructed. The approach present here simplifies the production of proteins from a wide variety of organisms for genomics-based studies.

KEY WORDS: Oligonucleotides synthesis, low-cost gene synthesis, bioengineering applications, structural studies, drug development, combinatorial biology, Error free synthesis, PCR method, ligation and easy mutational manipulation.

INTRODUCTION

Concept and Scientific Rationale of the idea

Artificial gene construction is the process of synthesizing a gene in vitro without the need for initial template DNA samples. Gene cloning and expression are schedule techniques used by molecular biologists. However, the PCR cloning step in general requires the existence of template DNA, which is not always readily available (Lei Young *et al.*, 2004). In enzyme engineering applications in particular, the desired DNA sequence is nearly always nonexistent. Furthermore, the natural DNA sequence may not be optimally expressed in a different organism, thus requiring codon optimization to achieve efficient expression. While site directed mutagenesis is expected to solve some of these problems, the process may become tedious and costly if too many nucleotides need to be changed (Stemmer *et al.*,1995). As an alternative, total gene synthesis is rapidly becoming the preferred method for applications requiring the assembly of DNA sequences, both natural and engineered(Picksley *et al.*,1990). To date, several methods for gene synthesis have been described, such as the ligation of preformed duplexes of phosphorylated overlapping oligonucleotides, the Fok I method and a modified form of ligase chain reaction for gene synthesis. However, all these methods require phosphorylated, polyacrylamide gel (PAGE) purified oligonucleotides for best result. The preparation of such oligonucleotides is labour intensive, and is therefore a major deterrent for researchers to pursue such a route.

Oligonucleotide synthesis

The oligonucleotide synthesis was developed by Beaucage & Caruthers (1981) and the method was known as phosphoramidite method or the solid phase method. DNA synthesis cycle involves four steps such as Detritylation, Coupling, Capping and Oxidation (Fig.1). The oligonucleotide synthesis is done by using a solid support of control pour glass (CPG) of 50 nm or 100 nm size to which dNTPs are attached covalently. This method is easier compared to traditional methods due to the fact that the removal of chemical reagent is easy and the washing procedure is simple. For synthesis, low quantity of starting material is required since the reaction can be carried in a single vessel and also can be automated fully. After the synthesis, cleavage of the solid phase oligo has to be transformed into liquid phase by using the 32% liquid ammonia under the pressure. Once the cleavage is done, the liquid oligonucleotide has to go for the deprotection 2 hrs at 75 degree Celsius to remove the excess of ammonia.

FIGURE-1

The oligonucleotides have to be purified due to follwing reasons:

- By-products from synthesis and cleavage
- Free protecting groups (salts)
- Residual solvents
- Truncated sequences (CE < 100%)
- Deletion mutants (capping reaction is not 100 % efficient)
- Base modifications, depurinations (difficult to remove)
- Unlabeled product

PCR Assembly- Annealing of oligonucleotides

Once the desired oligonucleotide sequenced is achieved, the PCR based ligation needs to be done. Now the respective oligonucleotide is kept in the PCR 96 well plate along with the PCR components such as Pyrococcus furiosus (Pfu) DNA polymerase, dNTPs and BTD Mix (Xiong et al., 2008). The F1 primer will bind to the R1 primer and similarly the F2 primer will go and bind to R2, similarly the whole chopped primer will finally gets ligated to each other. Once the ligation is done, the full sequence is achieved. Then, normal PCR will be carried out with the forward and reverse primer to achieve the desired concentration. For example in 20 cycles 106 sequence can be obtained and for 30 cycles109 sequence can be achieved (Fig-2).

FIGURE-2 Protocol for Gene Synthesis

1. The desired gene sequence needs to be synthesized is screened from the source database.
   Ex: NCBI, EMBL, UCSC Genome, Ensebbl Genome, NCGR.
2. The selected gene sequence needs to be chopped into small oligonucleotides with the help of GENE Cutters programs into desired length (30-45 nt base).
   Ex: DNA Works at Helix system, Gene to Oligo and GENEIUS.
3. Once the desired oligonucleotides are obtained, then the PCR based ligation will be carried out
   Ex: Closed chamber PCR.
4. Successful ligation of all the oligonucleotides will give rise to a copy of the desired gene.
5. Once the desired gene is synthesized, it has to be checked for the expected size by Agarose gel electrophoresis.
6. After determining the correct size, the gene has to be sequenced for exact position of bases as in the original gene sequence.
7. Once the error free gene is ready, that will be amplified by using forward and reverse primer for complete amplification of gene by PCR.

Example discussed for Artificial Gene Synthesis. Mouse genomic DNA for 5' region of cellular tumor antigen p53
LOCUS X01235 =554 bp DNA linear ROD 07-NOV-1985

Sequence

Fasta format

1 Ggcttacaaa gactctgtct taaaaatcca aaaagatggc tatgactatc tagctggata ggaaagagca cagagctcag aacagtggcg gtccacttac gataaaaact taattctttc cactctttat acttgacaca gaggcaggag tcctccgaat cggtttccac ccattttgcc ctcacagctc tatatcttag acgacttttc acaaagcgtt cctgctgagg gcaacatctc agggagaatc ctgactctgc aagtccccgc ctccatttct taccctcaac ccacggaagg acttgccctt acttgttatg gcgactatcc agctttgtgc caggagtctc gcgggggttg ctgggattgg gactttcccc tcccacgtgc tcaccctggc taaagttctg tagcttcagt tcattgggac catcctggct gtaggtagcg actacagtta gggggcacct agcattcagg ccctcatcct cctccttccc agcagggtgt cacgcttctc cgaagactgg gtaagtaatt gatgagcgtg acga 554

The Chopping needs to be done to achieve small oligonucleotide by GENE cutters and the parameter is detailed in **Table-1** Ex: DNA works at helix system.

TABLE 1. Parameters for trial 1

Sl.No	Parameters	Measures
1	Total Size of Gene	554nt
2	Protein Residues	0
3	Mutatable Residues	0
4	Fixed Nucleotides	554nt
5	Oligo Size	28-45nt
6	Total Oligo	30
7	Annealing Temp	58 +/- 1*C
8	Oligo Concentration	1.00E-7 M
9	Sodium Concentration	5.00E-2 M
10	Mg2+ Concentration	2.00E-3 M

30 Oligonucleotides are synthesized: For the complete gene synthesis

1	ggcttacaaagactctgtcttaaaaatc	28	
2	gatagtcatagccatctttttggatttttaagacagagtctttgt	45	
3	caaaaagatggctatgactatctagctggataggaaagagc	41	
4	cactgttctgagctctgtgctctttcctatccagcta	37	
5	acagagctcagaacagtggcggtccacttacgataa	36	
6	ataaagagtggaaagaattaagtttttatcgtaagtggaccgc	43	
7	aaacttaattctttccactctttatacttgacacagaggcagg	43	
8	aaccgattcggaggactcctgcctctgtgtcaagt	35	
9	agtcctccgaatcggtttccacccattttgccc	33	
10	cgtctaagatatagagctgtgagggcaaaatgggtgga	38	
11	tcacagctctatatcttagacgacttttcacaaagcgttcc	41	
12	agatgttgccctcagcaggaacgctttgtgaaaagt	36	
13	tgctgagggcaacatctcagggagaatcctgactct	36	
14	aggcggggacttgcagagtcaggattctccctg	33	
15	gcaagtccccgcctccatttcttaccctcaacc	33	
16	ggcaagtccttccgtgggttgagggtaagaaatgg	35	
17	cacggaaggacttgcccttacttgttatggcgactat	37	
18	cctggcacaaagctggatagtcgccataacaagtaag	37	
19	ccagctttgtgccaggagtctcgcgggggt	30	
20	aagtcccaatcccagcaacccccgcgagact	31	
21	tgctgggattgggactttcccctcccacgtgc	32	
22	gaactttagccagggtgagcacgtgggagggga	33	
23	tcaccctggctaaagttctgtagcttcagttcattggg	38	
24	cctacagccaggatggtcccaatgaactgaagctaca	37	
25	accatcctggctgtaggtagcgactacagttagggg	36	
26	gcctgaatgctaggtgccccctaactgtagtcgcta	36	
27	gcacctagcattcaggccctcatcctcctccttcc	35	
28	cgtgacaccctgctgggaaggaggaggatgagg	33	
29	cagcagggtgtcacgcttctccgaagactgggt	33	
30	tcgtcacgctcatcaattacttacccagtcttcggagaag	40	

Once the small oligonucleotide is achieved the PCR ligation will be followed Fig-3

FIGURE-3

MATERIALS AND METHODS

Gene synthesis can be done either by purchasing oligonucleotides from a commercial company or it can be synthesized in the laboratory depending on the facilities available. Oligonucleotides needs to be synthesized on a 50 nmol scale and stock solutions has to be prepared at a desired concentration in water without additional purification. All the PCR components such as Pyrococcus furiosus (Pfu) DNA polymerase, dNTPs and buffers of highest quality needs to be used (Gao *et al*., 2003).

Gene-Assembly-and Amplification

The assembly of the synthetic gene from component oligonucleotides could be performed accordingly. In brief, equal volumes of oligonucleotide solutions (each at a concentration of ~1 mg/ml) were mixed together and diluted with water to a final concentration of ~1 ng/µl for each oligonucleotide. The oligonucleotide mixture was diluted 5-fold with the PCR solution. The final concentrations of components were 0.2 ng/µl for each oligonucleotide, 20 mM for Tris–HCl (pH 8.8), 10 mM for KCl, 10 mM for (NH4)2SO4, 6 mM for MgSO4, 0.1% (v/v) for Triton X-100, 0.1 mg/ml for bovine serum albumin, 0.2 mM for each dNTP and 2.5 U for Pfu polymerase (Sandhu *et al.,* 1992). The PCR protocol for gene assembly began with one 5 min denaturation step of 95°C, during which the polymerase was added to avoid any possible mispriming ('hot start' PCR). This step was followed by 25 cycles of a denaturation temperature 95°C for 30 seconds, a variable annealing temperature (dependent on the melting temperature chosen in the program) for 30 seconds and an extension temperature of 72°C for 1.5 min (Mehta *et al.*, 1990). The last step in this protocol was an incubation cycle at 72°C for 10 min. For gene amplification, 1 µl of the mixture resulting from the gene assembly reaction was used as the template, with the outermost oligonucleotides used as primers (Innis *et al.,* 1990). The PCR protocol for gene amplification was essentially the same as gene assembly, except that the annealing temperature was raised to 62°C.

Program outline

DNAWorks was created with the idea of simplicity of use and moderate flexibility necessary to suit most common scenarios of designing the synthetic genes. For optimization several parameters like Tm, sodium ion concentarion, oligo length, codon optimization scheme etc has to be standardized.

An input of protein sequence is given to DNAWorks, which then reverse-translates the protein sequence into a series of highest-frequency codons for the organism chosen. The sequence can be entered either manually or as a file in FASTA or simple text format. Any desired flanking sequences can then be provided and incorporated into the synthetic sequence. At the beginning of its course, the back-translated initial sequence of the synthetic gene is divided into an odd number of contiguous sections, which are characterized by near-equal melting temperatures (Hoover *et al*., 2005). Algorithms used to calculate the oligonucleotide melting temperatures are based on the nearest-neighbor model. A score is then assigned to each section on the basis of codon frequency, hairpin formation (within each oligonucleotide), and deviations from the desired melting temperature and size. The possibility of hairpin formation is determined by comparison of oligonucleotide sequences for individual sections and the sequences of reverse complements for adjacent sections. The number, length, composition (G+C content) and relative position of matches between those sequences is used for calculating the 'hairpin-formation' score. The scores associated with deviations of melting temperatures and sizes of individual sections beyond the input tolerances are calculated using a parabolic function. This approach restrains the values for those parameters from drifting too far from the numbers requested in the input. Because of the high number of possible sequences (assuming just two possible codons per amino acid residue, a protein consisting of 100 residues can be encoded by 2^{100} different genes), we employed a stochastic method of optimization (a variant of a simulated annealing algorithm), rather than a deterministic one like steepest gradient. Two primary benefits of this strategy are its robustness against premature termination as a result of entrapment in local minima and time efficiency. During each optimization step, the boundaries of sections are also redefined. The initial and final synthetic gene sequences, a list of optimized oligonucleotide sequences, along with the scores for each section from both the initial and final sequence are generated and written out.

Codon Optimization

Codon optimization is a technique recently used by many scientists to improve the protein expression in living organism by increasing the translational efficiency of gene of interest; Optimizing codon for the custom gene will be done using standard programs to increase the functionality of gene (Lin *et al*., 2002).

Error Correction

To overcome problems associated with oligonucleotide quality, several elaborate strategies have been developed employing either separately prepared finishing oligonucleotides or mismatch binding enzymes of the mutS family or specific endonucleases from bacteria's. (Qiu *et al*., 2001).

Application of Artificial genes

To clone genes without template.

To improve protein expression.

To synthesize gene variants.

To make mutation libraries.

To freely design and make new vectors or complex constructs.

Construction of Hybrid Genes (Meyer et *al*., 2006).

Market Potential

At present, the synthesis of virtually all genes involves at least some manual steps and economies of scale are modest. This allows some gene synthesis companies to survive on revenues of just €1 – 1.5 million per year. On the other hand, automation will only cover its expected investment and maintenance costs if it is done on a very large scale. This suggests that the industry will be dominated by one or two very large, low cost firms by the middle of the next decade. There will probably also be five to ten smaller firms that specialize in technologies (including traditional manual methods) optimized around specific classes of genes that the dominant firms cannot readily mass produce. On the other hand, their economies of scale and survival will depend on attracting and retaining a critical mass of customers. We therefore expect large customers to exert substantial leverage over their suppliers' prices and service terms. Indeed, most large and mid-sized customers have already developed elaborate strategies to exploit their buying power. Most Western gene synthesis companies still export less than five to ten percent of all orders outside the US and Europe. Furthermore, most of these sales are focused on Asia. By contrast, sales to the Middle East account for less than one percent of all revenues and most of these come from Israel. This situation undoubtedly reflects biotechnology's still-incomplete penetration in many parts of the world. On the other hand, low demand cannot be the whole Story. Indeed, even the US and European markets are still significantly segmented with US firms doing about two-thirds of their business in North America and Europeans doing about 85% of their business in EU.Some of the companies like Genscript, Entelechon Geneart charges $0.29/bp to $0.35/bp. But with our phosphoramadite gene synthesis method, the cost of gene synthesis will be comparatively cheaper than that of other Asian and Europe companies which are in the market. Large number of metegenomic projects is happening throughout the world where in sequence of unknown organisms is being done leading to discovery of new genes of interest. Since only the digital sequences are available but not the template, scientists has to depend wholly on artificial gene synthesis for obtaining their gene of interest. The metagenomics is developing into a potential field of growth; the demand for artificial gene synthesis will also grow concomitantly offering bigger business opportunities in terms of volume and cost and becoming an important component of molecular biology products. In addition, *in vitro* gene synthesis will also have greater potency to bring in higher business opportunities in the area of cDNA construction, heterogenous protein expression, creation of gene variants, recombinant antibodies and synthetic biology.

Commercialization Prospects and Scalability

Synthesized gene will be comparatively cheaper than obtained from other methods for many projects.

Academic groups have lower budgets but still find *in vitro* synthesis more economical for many projects

Cost of synthesis are expected to decline rapidly

Complete control of sequence allows improved experimental design and new experimental approaches to modify genes

Current market estimate of *in vitro* gene synthesis is $20 million to $30 million a year

Revenues for *in vitro* gene synthesis are growing at 30% to 50% a year

Still a tiny fraction of the overall molecular biology market and there is lot of scope for improvement.

We expect it to grow rapidly but it could take 5-10 years to reach a significant fraction of the molecular biology market.

Current challenges and roadblocks

Error rate: ~1/300

Mismatched hybridization can lead to scrambled order

Reliability impacts speed and cost

Some sequences are difficult or impossible to PCR

Difficult sequences can add to the cost and delivery time

Novelty

- The rapid improvements in DNA synthesis technology hold the potential to revolutionize biosciences in the near future (Adleman *et al*., 2003). Traditional genetic engineering methods are template dependent and make extensive but laborious use of site-directed mutagenesis to explore the impact of small variations on an existing sequence "theme." *De novo* gene and genome synthesis frees the investigator from the restrictions of the pre-existing template and allows for the rational design of any conceivable new sequence theme.
- By using the newly developed method of PCR-based gene synthesis, error-free synthetic genes for the human protein kinases PKB2, S6K1 and PDK1 can be obtained with little or no corrective mutagenesis.
- Artificial gene synthesis can be used to customise services to improve antibody affinity and/or specificity -Domain Engineering and Comprehensive Mutagenesis. Domain engineering simplifies the process of making a series of directed changes to an antibody (Benner *et al*., 2005 and Donald *et al*., 2007).

RESULT

Sequence of the mature protein were obtained from the different protein banks and used as input for DNA works .For each single protein the program was executed with the various combination of the melting temperature, in order to get the artificial sequence characterized by the lowest score. The set of the calculated oligonucleotide were cross verified using the program GCG (26) to verify the consistency of the melting temperature of all the oligonucleotide .The calculation time for a 250bp sequence is 5-10 minutes, while for the sequence 1000bp and above is approximately 15-20 minutes, the program is executed on a PC with a2.00 GHz Intel(R) Core(TM) 2CPU processor.

The procedure for synthesizing genes is relatively fast and rationalized. With the conventional old solubilization of lyophilized oligonucleotides .The final gene amplification reaction gave rise to a single dominant band of correct size as shown by the gel electrophoresis (Fig-4).

FIGURE-4

a) 1 is a 2kb Ladder
b) Ladder marked
c) 3, 4&5 PCR gene products are all of the expected size (~554 bp).

CONCLUSION AND DISCUSSION

Artificial gene synthesis is the process of synthesizing a gene *in vitro* without the need for initial template DNA samples. The main method is currently by oligonucleotide synthesis (also used for other applications) from digital genetic sequences and subsequent annealing of the resultant fragments. In contrast, natural DNA replication requires existing DNA templates for synthesizing new DNA (Gibson, D.G et al., 2008). If your gene of interest is rare in nature, or the genetic material containing the gene of interest is difficult to obtain or isolate, then the gene of interest can be synthesized *de novo* by *in vitro* artificial synthesis.By synthesizing oligos with overlapping ends; these oligos can be annealed and extended to form a full-length gene. Due to the nature of oligo synthesis, mutations, especially deletions occur frequently during the synthesis process. This situation dramatically limits the gene size and increase the cost of gene synthesis. To overcome this issue, a novel gene synthesis platform has been developed to efficiently synthesize the gene of larger size with no mutation (.Neylon, C et al.,2004).

ACKNOWLEDGEMENT

The authors are extremely grateful to Dr. Premchandra Sagar, Vice Chairman, Dayananda Sagar Institutions and Dr. Krishna Gowda, Director, Dayananda Sagar College of Biological Sciences, Bangalore-560078, INDIA, for their immense guidance and support for this project.

REFERENCES

Adleman, L. M. (1994) Molecular computation of solutions to combinatorial problems ,Science, pp-1021±1024

Benner , S . A . and Sismour, M . A . (2005) Synthetic biology . Nature Reviews Genetics pp- : 533 – 543.

Donald J. Quigg Extreme Genetic Engineering, An Introduction to Synthetic Biology, ETC Group, January, 2007 . Animal Legal Defense Fund v., pp-920-930

Gao, X., Yo, P., Keith, A., Ragan, T.J. and Harris,T. K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-®delity assembly of longer gene sequences. Nucleic Acids Res., pp-31-143

Gibson, D. G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science, pp- 1215-1220.

Hoover, D. M. and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis Nucleic Acids Res., pp-30-43.

Innis, M. A. and Gelfand, D.H. (1990) In Innis, M.A., Gelfand, D.H., Snindky, J.J. and White, T.J. (eds) PCR Protocols. Academic Press, New York, pp. 3–12.

Lei Young and Qihan Dong Two-step total gene synthesis method, Nucleic Acids Research, 2004, Vol. 32, No. pp-1058.

Lin,Y., Cheng, G., Wang, X. and Clark, T.G. (2002) The use of synthetic genes for the expression of ciliate proteins in heterologous systems. Gene, pp- 85±94.

Mehta, R. K. and Singh, J. (1999) Bridge-overlap-extension PCR method for constructing chimeric genes. Biotechniques, pp- 1082±1086.

Meyer, M., Wagner, E. (2006) Recent developments in the application of plasmid DNA-based vectors and small interfering RNA therapeutics for cancer. Hum. Gene Ther., pp- 1062-1076.

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

Picksley, S. M., Parsons, C. A., Kemper, B. and West, S. C. (1990) Cleavage speci®city of bacteriophage T4 endonuclease VII and bacteriophage T7 endonuclease I on synthetic branch migratable Holliday junctions. J. Mol. Biol., pp-723±735.

Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A.V., Tiedje, J. M. and Zhou, J. (2001) Evaluation of PCR-generated chimeras, mutations and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol., pp-, 880±887.

Sandhu, G. S., Aleff, R. A. and Kline, B. C. (1992) Dual asymmetric PCR one-step construction of synthetic genes. Biotechniques, pp- 14±16.

Stemmer,W.P., Crameri, A., Ha, K. D., Brennan, T. M. and Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene, pp- 49±53.

Xiong, A.S., Peng, R.H., Zhuang, J., Liu, J.G., Gao, F., Chen, J.M., Cheng, Z.M., Yao, Q.H. (2008) Non-polymerase cycling-assembly-based chemical gene strategies, methods, and progress. Biotechnol. Adv., pp-121-134.