Modern approaches to artificial gene synthesis: aspects of oligonucleotide synthesis, enzymatic assembly, sequence verification and error correction

G.Y. Shevelev¹,², D.V. Pyshnyi¹,²

¹ Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
² Novosibirsk State University, Novosibirsk, Russia

Synthetic biology is a rapidly developing field aimed at engineering of biological systems with predictable properties. Synthetic biology accumulates the achievements of modern biological sciences, programming and computational modeling as well as engineering technologies for creation of biological objects with user-defined properties. Evolution of synthetic biology has been marked by a number of technological developments in each of the mentioned fields. Thus, significant reduction in cost of DNA sequencing has provided an easy access to large amounts of data on the genetic sequences of various organisms, and decreased the price of the DNA sequence synthesis, which, analogous to Moore’s law, resulted in an opportunity to create a lot of potential genes without the time-consuming and labor-intensive traditional methods of molecular biology. Development of system biology has allowed forming a deeper understanding of the functions and relationship of natural biological models, as well as of the computational models describing processes at the cell and system levels. Combination of these factors has created an opportunity for conscious changes of natural biological systems. In this review the modern approaches to oligonucleotide gene assembly synthesis are discussed, including such aspects as protocols for gene assembly, sequence verification, error correction and further applications of synthesized genes.

Key words: oligonucleotide synthesis; artificial gene synthesis; enzymatic gene assembly; polymerase cycling assembly (PCA); ligase cycling assembly (LCA); error correction; sequence verification.

HOW TO CITE THIS ARTICLE:

Современные подходы к синтезу генов: аспекты синтеза олигонуклеотидов, ферментативной сборки, проверки последовательностей и коррекции ошибок

Г.Ю. Шелев, Д.В. Пышный

¹ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
² Новосибирский государственный университет, Новосибирск, Россия

Синтетическая биология – быстро развивающаяся отрасль науки, нацеленная на создание биологических систем с предсказанными свойствами. При этом она использует достижения современной биологии, программирования и компьютерного моделирования, а также инженерной отрасли для создания биологических объектов, обладающих набором заранее заданных пользовательских свойств. Развитие синтетической биологии было обусловлено множеством технологических разработок в каждой из упомянутых отраслей. Так, значительное снижение стоимости технологии секвенирования ДНК привело к наработке больших объемов данных о генетических последовательностях различных организмов, и снижение стоимости синтеза генов, аналогичное закону Мура, позволило создавать библиотеки потенциально интересных генов без необходимости их создания традиционными методами. Развитие системной биологии позволило получить более глубокое понимание взаимосвязей и функций природных биологических моделей, а также построить прогностические модели, описывающие молекулярные процессы на клеточном и системном уровнях. Комбинированный подход позволяет создавать библиотеки синтетических генов, представляющих потенциальный интерес в работе генетических инженеров.

Ключевые слова: синтез олигонуклеотидов; синтез искусственных генов; ферментативная сборка генов; полимеразная циклическая сборка; лигазная циклическая сборка; коррекция ошибок; проверка последовательности.
Since the first gene was synthesized from scratch in the 70’s (Agrawal et al., 1970) and it’s in vivo activity was tested (Ryan et al., 1979) around 50 years has passed, and the technology of artificial gene synthesis has made a significant leap beyond the approaches implemented in the classical works. Nowadays, it relies on the modern methods of oligonucleotide synthesis, the variants of Polymerase Chain Reaction (PCR) (Saiki et al., 1985; Mullis et al., 1986) with high fidelity DNA polymerases (Böhlke et al., 2000), and the sequence verification step using the Sanger sequencing (Sanger, Coulson, 1975; Smith et al., 1986) or high-throughput DNA sequencing (Church, 2006; Hall, 2007; Bosch, Grody, 2008; Schuster, 2008; Shendure, Ji, 2008; Tucker et al., 2009). During the last decade modern approaches to assemble large (up to dozens and hundreds kb) DNA molecules (Gibson et al., 2009; Gibson, 2011) and error correction within synthesized gene have been introduced (Ma et al., 2012b).

**Oligonucleotide synthesis**

All the synthesis approaches rely on usage of chemically synthesized oligonucleotides as building blocks for an enzyme-mediated assembly (Ma et al., 2012b; Kosuri, Church, 2014).

**Solid-phase oligonucleotide synthesis**

The most widely used approach to oligonucleotide synthesis has been the solid-phase phosphoramidite method developed in the 80’s (Beaucage, Caruthers, 1981; Matteucci, Caruthers, 1981). The DNA oligonucleotides are synthesized from the 3’ to 5’ end by consecutive coupling of activated building deoxynucleoside phosphoramidites to an initial deoxynucleoside attached to a solid support (usually the support is a controlled pore glass (CPG) or highly cross-linked polystyrene beads) by its 3′-OH group (Ellington, Pollard, 2000). A solid support matrix is placed into several or several hundred parallel flow-through individual reaction columns, and while synthesis all the reagents necessary for the synthesis cycle flow through the solid support column. A full cycle to add a single nucleotide consists of several stages: (1) deprotection: acid is used to remove a Dimethoxytrityl (DMT) group from the 5′-end of a growing oligonucleotide chain; (2) coupling: the solid-phase 5′-OH group of an attached oligonucleotide reacts with tetrazole – activated nucleoside phosphoramidite producing the elongated product of reaction; (3) capping: treatment of the uncoupled 5′-OH groups by acetic anhydride to minimize the products of deletion mutations; (4) oxidation: conversion of an unstable phosphate triester into a stable phosphate prior to the next step of detritylation. The cycle repeats until a full-length oligonucleotide is obtained. The synthesized oligonucleotide is then cleaved from solid support by treatment with a strong base such as ammonia in aqueous solution and the remaining protection groups are removed during the cleavage. For gene synthesis application high quality and homogeneity of synthesized oligonucleotides is critically important (Xiong et al., 2006) since the synthesized oligonucleotides need to be purified either by the high performance liquid chromatography (HPLC) or by polyacrylamide gel electrophoresis (PAGE) methods.

Synthesis of oligonucleotides with the length up to 100 nt using solid-phase phosphoramidite method is a standard for gene synthesis industry and other applications. The limiting factor for overcoming barrier over 100 nt during oligonucleotide synthesis is stochastic depurination side reactions which accompany acid deprotection step (Hall et al., 2009). Vendors such as Integrated DNA Technologies (IDT), however, declare possibility to synthesize high quality oligonucleotides up to 200 nt using a special support and synthetic protocol for low-yield and long-oligonucleotide synthesis.

As an alternative to chemical oligonucleotide synthesis, T4 RNA ligase-mediated solid-phase enzymatic oligonucleotide synthesis approach was proposed (Schmitz, Reetz, 1999). The approach is based on coupling reaction between a solid-phase attached oligonucleotide primer and the mononucleoside 3′,5′-biphosphate mediated by T4 RNA ligase in a water solution. The terminal phosphate blocking group is removed enzymatically by alkaline phosphatase, and all excess reagents are simply washed off from the resin. The next nucleoside 3′,5′-biphosphate can then bind to the end of the chain. Following NH$_4$OH-induced cleavage from the resin gives the desired elongation of the product. The drawback of this approach is relatively low rate of elongation (48 h) due to the kinetic properties of the enzyme. However, possible automation and molecular biological optimization of enzyme using directed evolution can make the enzymatic approach a viable alternative to today’s chemical synthesis approach.

**Parallel oligonucleotide synthesis using microchips**

The oligonucleotide amounts necessary for gene synthesis application are relatively small (picomols) while modern DNA synthesizers can produce oligonucleotides in a higher excess (nanomolar and micromolar) making this approach redundant. Significant improvements in performance and cost-reduction may be achieved by parallelization and miniaturization of oligonucleotide synthesis platforms. Thus, microchip-based approaches previously used for DNA diagnostics and sequencing have been adopted for multi-parallel synthesis of oligonucleotide arrays. Several constructive implementations such as ink-jet DNA printing (Lausted et al., 2004), microfluidic devices (Zhou et al., 2004; Huang et al., 2009; Lee et al., 2010), light-directed (Richmond et al., 2004) and electrochemical (Egeland, Southern, 2005; Chow et al., 2009) microarray synthesis and LED-controlled capillaries (Blair et al., 2006) synthesis have been independently developed. A conditions modification during the detritylation step of chip-based oligonucleotide synthesis lead to improved quality of oligonucleotides with the possible length up to 150 nt (LeProust et al., 2010). However, the limitation of all microarray synthesis methods is the attomoles (10^-18 mol) of an individual oligonucleotide that is insufficient for a gene assembly. PCR of selected subpools allows amplifying synthesized oligos at acceptable quantities (Kosuri et al., 2010; Schmidt et al., 2015) for a subsequent multiple gene assembly using barcoded magnetic beads (Plesa et al., 2018).

**Gene assembly**

Enzymatic assembly of synthetic oligonucleotides into long double-stranded DNA fragments with the length up to several hundred bp or even several kb (Ma et al., 2012b) is the next step of gene synthesis with numerous approaches developed since this technology was implemented. In this paper we will discuss several protocols for gene assembly and their parameters.
Ligase cycling assembly (LCA)

LCA utilizes the properties of *Taq* DNA ligase to repair single stranded 5′-P, 3′-OH breaks (nicks) within a double stranded DNA at increased temperature (50–60 °C). Conjunction of two oligonucleotides (one 5′-phosphorilated and another with 3′-OH) takes place when they are annealed to a third template oligonucleotide. After joining of oligonucleotides the reaction mixture contains multiple oligonucleotides heated to 95 °C until all the complexes are denatured and further ligation of longer fragments takes place. These steps (annealing, ligation and denaturation) repeat a number of cycles and after that the reaction mixture is amplified with terminal primers and DNA polymerase until the full-length product is obtained (Fig. 1).

One of the alternative implementation of LCA is blunt-end ligation of double stranded DNA blocks using streptavidin-coated magnetic beads and biotin bonded oligonucleotides joined with subsequently added 5′-phosphorylated DNA duplexes (Dietrich et al., 1998). The latest modifications of this approach utilize 5′-phosphorilated single-stranded oligonucleotides (Pengpumkiat et al., 2016).

Polymerase cycling assembly (PCA)

PCA is the most widely used method of gene synthesis based on the polymerase chain reaction (PCR) simultaneously carried out in multiple places within the set of short complementary overlapping oligonucleotides forming paired complexes with sticky ends (Fig. 2).

Each step of filling the gaps by polymerase is followed by denaturation and annealing steps and leads to formation of another set of elongated complexes with sticky ends. The process repeats until a full-size product appears (with some side products formed) in the mixture and the final product is amplified with conventional PCR using terminal primers. In the earlier experiments using PCA, in the process of gene amplification with terminal primers gene assembly was performed in a separate tube. Later on this approach was adopted for one-tube assembly (Wu et al., 2006, TerMaat et al., 2009). The desired fragment assembled in a “single-pot” cycled enzymatic reaction or in a several step assembly of separately preassembled structural blocks of the final genetic circuit. Similarity with recursion process gave it its name, so
it is known as recursive PCR. The approach was proposed in 1992 and used for the synthesis of human lysozyme gene (Prdromou, Pearl, 1992). The structure of oligonucleotides overlapping ends was specially designed to obtain the melting temperature within the range 52–56 °C to minimize heterodimers and the oligonucleotides were purified with PAGE. Gapless implementation of recursive PCR was used for 2703-bp plasmid assembly using 134 unpurified 40nt oligonucleotides with 20nt sticky ends (Stemmer et al., 1995).

Since the first usage, PCA approach has been thoroughly investigated and several improvements of the approach have been developed, e.g. primeSTAR HS DNA polymerase was proposed as the best-choice polymerase for efficient high-fidelity gene synthesis (Cherry et al., 2008). Utilization of real-time PCR with intercalating dyes has given the insight information concerning the optimal amount of PCR cycles during gene assembly: the authors declared the amount of PCR cycles of about two times more than the theoretically calculated minimal amount of cycles necessary for assembling a full-length product (Ye et al., 2009). Also, the optimal oligonucleotide concentrations while gene assembly was found to be in a range 10–60 nM, external primer concentration – 0.2–1 µM (Wu et al., 2006) and dNTP increased concentration – up to 1 mM each (Ye et al., 2009).

Modification of PCA approach known as TopDown one-step gene synthesis was proposed as advanced technique (Ye et al., 2009). The essential idea of the approach is in a difference between the \( T_m \) values of complexes corresponding to the overlapping sites of oligonucleotides and terminal primers. The outer primers \( (T_m \sim 50 \, ^\circ C) \) and inner oligonucleotides \( (T_m \sim 65 \, ^\circ C) \) were designed with a melting temperature difference of \( \sim 15 \, \text{°C} \). A higher annealing temperature \( 65 \, ^\circ C \) was used for the first 20 cycles of product assembly and after that the annealing temperature was lowered to 50 °C for the next cycles of full-length product amplification using outer primers. Further improvement of the TopDown synthesis led to the automatic kinetics switch approach (Cheong et al., 2010). The additional feature of this kinetic approach is flanking tail at the outer primer sequence. Primers have two regions with melting temperatures \( T_{m_1} \) and \( T_{m_2} \) where \( T_{m_1} \) is the temperature of a gene-specific region without a flanking tail, and \( T_{m_2} \) is the melting temperature of a primer including a flanking tail. \( T_{m_1} \) is lower than the melting temperature of the oligonucleotides assembly and \( T_{m_2} \) is \( \sim 72 \, ^\circ C \). When the full-length template appears, the outer primers first created full-length DNA with flanked tails, causing the shift in melting temperatures of outer flanked template to \( T_{m_2} \). The approach provided kinetic switch from inner oligonucleotides assembly to full-length template amplification during one-step gene assembly.

**Thermodynamically balanced inside-out (TBIO) approach**

Another PCR-based approach is the TBIO gene synthesis (Gao et al., 2003). The approach utilizes the process of primer dimers extension under PCR conditions in the absence of template. The process of gene synthesis starts in the center of gene sequence between two overlapping oligonucleotides (Fig. 3).

After extension by PCR, the annealing and extension of the next pair of primers between previously synthesized DNA duplex hybridized through specially designed overlapped sequences with adjusted melting temperatures (i.e. thermodynamically balanced) begins. The next iteration with the third pair of primers leads to further elongation of the fragment. The primers in a TBIO assembly are added with gradient concentration: the most internal primers are added in the lowest concentration and the most external primers – in the highest concentration facilitating reaction in the direction inside out. Several gene fragments up to 400–500 bp are independently assembled using a TBIO approach and gel-purified to remove PCR byproducts. Then, these fragments are assembled in one larger gene circuit.

Another strategy for gene synthesis utilizes a combination of Dual Asymmetric (DA) and Overlap Extension PCR (OE-PCR) (Young, Dong, 2004). The oligonucleotides are designed in the same manner as it is done in the PCA approach to contain overlapping regions with or without a gap. Two pairs of complementary oligonucleotides are mixed in one tube and extended by \( Pfu \) polymerase to double-stranded blunt-ended blocks (Fig. 4).

Several blocks are assembled independently and then mixed with each other to make a whole gene assembly. Reduction of the gap size or its total elimination leads to higher fidelity of the gene assembly during PCR process. To further decrease the mutation product ratio, the full length PCR products are denatured, reannealed and cleaved by T7 endonuclease I. In this manner several genes from 470 bp to 1.2 kb are synthesized. The approach is relatively low-cost because it does not require gel purification of enzymatic phosphorylation.

**Error correction**

The source of the errors presented in a synthetic gene can be either the product of oligonucleotide synthesis or of enzymatic gene assembly. Various strategies can be utilized to reduce the error occurrence during different stages of gene synthesis.
Correction of errors in synthetic oligonucleotides

Since the efficacy of solid-phase oligonucleotide synthesis is less than 100%, the most common types of errors are deletions and insertions. The deletions occur due to an incomplete capping step of solid-phase oligonucleotide synthesis with the rate up to 0.5% per nucleotide, while the insertions happen due to tetrazole cleavage of DMT group with the rate below 0.4% per nucleotide. Another error, depurination (cleavage of the N-glycosidic bond), occurs under acidic conditions during detritylation and causes apurinic sites within an oligonucleotide chain (Ellington, Pollard, 2000). Such modifications of synthesis protocol as increased coupling time of the nucleotide chain (Ellington, Pollard, 2000). A combination of mismatch-specific endonucleases such as single-strand-specific nucleases S1 and P1, mung bean nuclease and CEL I nuclease (Desai, Vepatu, 2003), mismatch repair endonuclease Muth (Smith, Modrich, 1997), and resolvases, such as T7 endonuclease I, E. coli endonuclease V and T4 endonuclease VII (Ma et al., 2012a) can be used for error correction in both column and microarray synthesized oligonucleotides using Surveyor nuclease (Saaem et al., 2012; Currin et al., 2014). However, it is unclear which type of mismatch-specific endonucleases is the most effective for error removal (Sequeira et al., 2016).

Another approach for correct gene harvesting is functional selection. The gene of purpose could be fused to a selection marker, such as luciferase gene (Yarimizu et al., 2015). In the case when a marker gene is expressed after cloning, the luciferin-producing clones could be sequenced further. This approach, however, works only for protein-coding genes and is not effective for non-functional mutations.

Blocks assembly

The typical length of a gene fragment that can be synthesized using the above-mentioned methods is about 1 kb. The
synthesized fragments need to be cloned into a plasmid for sequence verification, the verified fragments combined into a larger gene circuit.

One of the approaches to combine several genetic fragments is the BioBrick strategy allowing assembling a standardized DNA fragment flanked with restriction sites for restriction/ligation reactions (Knight, 2003). Different BioBrick fragments represent various biologically valuable modular parts such as promoters, ribosome-binding sites, coding sequences and transcriptional terminators that can be assembled in a combinatorial way (Shetty et al., 2008). The approach allows assembling several genes and regulatory elements. The method’s drawback is the “scars” between modules caused by restriction/ligation assembly.

Utilization of the type II restriction enzymes can be a solution for scarless assembly of gene clusters with the length of about 32 kb (Kodumal et al., 2004). This family of enzymes cut outside of a non-palindromic recognition site has a 4nt overhang, which can be specially chosen for assembling of neighboring fragments without scars (Szybalski et al., 1991). This method is called Golden Gate and has been utilized for assembly of multiple inserts (Engler et al., 2008).

An alternative to restriction/ligation methods is a group of overlap assembly methods (Quan et al., 2009). Similarly to OE-PCR, the overlap methods calls for homologous DNA fragments between blocks to be joined by in vitro recombination. There are several commercial cloning kits such as Gateway (Thermo) using proprietary recombinase based on site-specific recombination system of bacteriophage lambda to shuttle sequences between plasmids bearing specific flanking compatible recombination attachment (att) sites (Liang et al., 2013). However, carrying out of a multiple assembly projects with this technique is not convenient because it requires specific sites to be present in each act of assembly. Another kit for cloning into plasmids is In-Fusion (Clontech) which is more promising since it only needs a 15nt sequence overlap between digested plasmid and gene fragment or between gene fragments, which significantly expands the possibility of its application for large circuits assembly (Slight et al., 2010). Sequence and ligation independent cloning (SLIC) approach utilizes 3'-5' proofreading exonuclease activity of the T4 DNA polymerase in the absence of dNTP to create overhangs of 30 nt after 30 min incubation with double stranded DNA fragments. DNA overhangs anneal with homologous excised fragments. Recombinase RecA is used to insert the fragment into a plasmid. As a demonstration of applicability of this approach for parallel assembly, 10 fragments with the size ranges between 275 and 980 bp with 40 bp overlaps were inserted into a 3.1 kb vector (Li, Elledge, 2007).

The recombinase-free approach named Gibson Assembly is quite useful for multiple assembly of dsDNA fragments with overlapping ends (Gibson et al., 2009). The approach utilizes a cocktail of three enzymes: (1) exonuclease that chews back the end of the fragments and exposes ssDNA overhangs that are specifically annealed; (2) polymerase that fills the gaps within the annealed products; (3) ligase that covalently sews the fragments together. The authors proposed one-step implementation of this technique using exonuclease III preserving its functional properties in the presence of dNTP, antibody-bound Taq polymerase preventing competition with exonuclease for 3' ends of DNA and Taq ligase placed in a single tube in a thermocycler. Three 5-kb DNA fragments with 40 bp overlaps have been assembled and cloned into E. coli given 15 kb insert. Moreover, an assembly of 25 % M. genitalium genome using two over-100 kb DNA fragments with 257 bp overlap, and its further complete genome has made the Gibson assembly the most universal and convenient method for large scale gene assembly. It has also been adopted for oligonucleotide-scale to gene-scale assembly: the mouse mitochondrial genome was assembled from 60-mer oligonucleotides using this approach (Gibson et al., 2010).

Oligonucleotides design software
One of the main limitations in gene synthesis is designing a set of oligonucleotides that should meet several criteria: (1) similar values of $T_m$ between overlapping fragments; (2) the length of oligonucleotides within the set should not exceed 50–60 nt; (3) oligonucleotides must be annealed only with their neighboring oligonucleotides in the set and all abnormal intra and intermolecular annealing must be excluded by a splitting algorithm. Designing oligonucleotides for gene synthesis purposes manually is time-consuming because many factors such as GC content, restriction sites, overlapping fragments, codon frequency, and so on need to be taken into account. Various different software packages have been developed in order to optimize the gene design and synthesis process.

One of the most convenient programs used for these purposes is DNAWorks (https://bpcwebapps.cit.nih.gov/dnaworks/) (Hoover, Lubkowski, 2002). This program allows designing oligonucleotides for gapped and gapless overlap extension PCR assembly as well as TBIO assemblies with specified $T_m$ range and oligonucleotide length. It is also capable to transfer from amino acid to nucleotide sequence using about 10 different codon frequency tables as well as custom user defined table.

Gene2Oligo (http://berry.engin.umich.edu/gene2oligo/) (Rouillard et al., 2004) allows one to design oligonucleotides for LCA and gapless PCA assemblies. GeneDesign (http://54.235.254.95/gd/index.html) (Louw et al., 2011) provides oligo design for gapped PCA assembly.

GeneDesign (http://54.235.254.95/gd/index.html) (Villalobos et al., 2006; Richardson et al., 2010) is a program enabling one to divide long DNA sequence into ~500 bp fragments joined by restriction sites and to chop this fragments into sets of overlapping oligonucleotides. This approach can be successfully applied for very long DNA sequences up to several kb. GeneDesign also provides the “Codon Juggling” feature which is capable to produce sequence that is codon-optimized for expression and, as different as possible from the original sequence (and still coding the same protein). Either, addition and removing of restriction sites within sequence is possible.

TmPrime (http://prime.ibn.a-star.edu.sg) (Bode et al., 2009) and Genecomposer (http://www.genecomposer.net) (Lorimer et al., 2009) are other software solutions for gene design with expanded set of functions analogous to GeneDesign. Currently, not available on the Internet.

Another algorithm based on reliable gradient optimization and the derivative objective function approximated with a central difference operator was proposed for gene synthesis.
application by Louw et al. (2011). The authors compared the results of their gradient optimization approach against the one using Gene2Oligo and concluded their approach produces higher purity and yield of assembled genes.

DNASynth (http://dnasynth.sourceforge.net/) (Nowak et al., 2015) is a relatively new algorithm that designs the whole artificial gene synthesis process, developing the optimal nucleotide sequence encoding a given peptide for a given host organism and determining the best long DNA LCA-based assembly protocol.

Microchip-based gene assembly protocols needs special software for oligonucleotide arrays design. PICKY (Birla, Chou, 2015) is software that can be applied for multiple one-pot gene assemblies from dsDNA fragments using the Gibson Assembly protocol. It provides thermodynamic analysis that identifies all unique junctions in gene, where consecutive DNA fragments are specially designed to connect to each other.

### Perspectives and applications

Current applications of synthetic DNA combine synthesis of genetic constructs and metabolic pathways as well as synthetic of artificial genomes and production of artificial genome-based microorganisms (Gibson et al., 2008). Last improvements of long DNA assembly approaches lead to the possibility of creation artificial living system based on *M. genitalium* with minimized functional genome (Hutchinson et al., 2016). Recently, the Human Genome Wright project has been announced by a group of US scientists (Boeke et al., 2016). The goal of the project is to synthesize a human genome to improve understanding of the interconnections between genes. For this sake, the additional aim of the project is to catalyze the reduction in price per base, which is now floats around 0.2 $ per 1 bp, to as low as a few cents per bp. Further improvements in microchip-based DNA synthesis technologies will help to drop down the current price per 1 bp.

Another significant application of synthetic biology is metabolic engineering. By designing metabolic systems working in parallel with host metabolic pathways, scientists can program cells for practical applications such as bioremediation (Wasiklowski et al., 2012; Garibus et al., 2017) and production of industrial compounds (Adrio, Demain, 2010). For example, engineering a pathway for biosynthesis of antimarial compound Artemisinin in yeast has lowered its price by about 10 times if compared to the previously used method (Paddon, Keasling, 2014).

Application of next generation sequencing technologies (NGS) for synthesis and amplification of error-free oligonucleotides (Schwartz et al., 2012; Lim et al., 2018; Plesa et al., 2018) leads to expanding the application of chip-synthesized oligonucleotide pools into *in vitro* directed evolution methods. For instance, development of multiplex automated genome engineering (MAGE) platform for large-scale programming and evolution of cells and its application to optimization of 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway in *E. coli* for overproduction of the industrially important isoprenoid lycopene has allowed for fivefold increase in lycopene amounts which is quite remarkable achievement (Wang et al., 2009). Recent developments of several chip-based oligonucleotide synthesis platforms will enable us to amplify error-free subpools of oligos with further independent assembly of homologous gene libraries allowed authors to rationally explore sequence-function relationships at unprecedented scale (Plesa et al., 2018).

### Acknowledgments

This work has been supported by Russian Science Foundation (no. 17-74-10157).

### Conflict of interest

The authors declare to have no conflict of interest.

### References


large number of overlapping oligodeoxyribonucleotides. J. Bio-

Chow B.Y., Emig C.J., Jacobson J.M. Photocyclephotochemical synthesis of
15224. DOI 10.1073/pnas.0813011106.


Curran A., Swainston N., Day P.J., Kell D.B. SpeedyGenes: An im-
proved gene synthesis method for the efficient production of error-
corrected, synthetic protein libraries for directed evolution. Protein

Desai N.A., Vepatsu S. Single-strand-specific nucleases. FEMS Micro-

Dietrich R., Wirsching F., Opitz T., Schwienhorst A. Gene assembly
by use of synthetic biology in pharmaceutical development. Nat.

Lausted C., Dahl T., Warren C., King K., Smith K., Johnson M., Sale-
8-R58.

Lee C.C., Snyder T.M., Quake S.R. A microfluidic oligonucleotide syn-
genetkdgkq092.

LeProust E.M., Peck B.J., Spirin K., McCuen H.B., Moore B., Nam-
raeve E., Caruthers M.H. Synthesis of high-quality libraries of long
(150mer) oligonucleotides by a novel depurination controlled pro-
gkl163.

Li M.Z., Elledge S.J. Harnessing homologous recombination in vitro
to generate recombinant DNA via SLIC. Nat. Methods. 2007;4(3):251-
256. DOI 10.1038/nmeth1010.

Liang X., Peng L., Baek C.-H., Katzen F. Single step BP/ LR com-
10.2144/000114101.

Lim H., Cho N., Ahn J., Park S., Jang H., Kim H., Han H., Ji H.L.,
Bang D. Highly selective retrieval of accurate DNA utilizing a pool of
in situ-replicated DNA from multiple next-generation sequencing
gkq016.

Lorimer D., Raymond A., Walchli J., Mison J., Barrow A., Wallace E.,
Grice R., Burgin A., Stewart L. Gene composer: database software
for protein construct design, codon engineering, and gene synthesis.

Low T.M., Whitney S.E., Tornatt J.R., Pienna E., Viljoen H.J. Oli-
gonucleotide optimization for DNA synthesis. AIChE J. 2011;57(7):

Ma S., Saem I., Tian J. Error correction in gene synthesis technology.

Ma S., Tang N., Tian J. DNA synthesis, assembly and applications in
10.1016/j.cub.2012.05.001.

Matteucci M.D., Caruthers M.H. Synthesis of deoxyoligonucleotides
DOI 10.1021/ja00401a041.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Echberg H. Specific
enzymatic amplification of DNA in vitro: the polymerase chain reac-
10.1101/SQB.1986.051.01.032.

Nowak R.M., Wojcieszek-Krawiec A., Puimentiaccia A. DNASynth: 
a computer program for assembly of artificial gene parts in decreasing

Paddon C.J., Keasling J.D. Semi-synthetic artemisinin: a model for the
Modern approaches to artificial gene synthesis: oligo synthesis, enzymatic assembly, sequence verification and error correction

G.Y. Shevelev
D.V. Pyshnyi


