Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 124 (2006) 206–212
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

Review

# Another side of genomics: Synthetic biology as a means for the exploitation of whole-genome sequence information

Peer Stähler[a,*], Markus Beier[a], Xiaolian Gao[b], Jörg D. Hoheisel[c]

[a] *Febit Biotech GmbH, Im Neuenheimer Feld 515, 69120 Heidelberg, Germany*
[b] *Department of Biology & Biochemistry, University of Houston, Houston, TX 77004-5001, USA*
[c] *Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany*

Received 6 September 2005; accepted 1 December 2005

**Abstract**

The successful completion of the Human Genome Project and other sequencing projects opened the door for another quantum jump in science advancement. The most important public sequence databases are doubling in size every 18 months. By revealing the genetic program of many organisms, these efforts endow biologists with the ability to study the basic information of life in toto as an initial step toward a comprehensive understanding of the complexity of entire organisms. We review the area of synthetic biology, defined as the making and use of biosystems founded on the chemical synthesis of the coding DNA (and potentially RNA). The recent developments discussed here introduce a rich source of oligonucleotides to the field: in situ synthesised microarrays, which in fact represent nothing else but matrix nucleic acid synthesisers. With this new way of producing the oligonucleotides used in the making of synthetic genes in a very cost-effective manner, the field of synthetic biology can be expected to change dramatically in the next decade. Synthetic genes will then be the tools of choice to obtain any sequence at any time in any laboratory.


*Keywords:* Synthetic biology; Synthetic gene; In situ synthesised microarrays; Matrix nucleic acid synthesiser; Array-derived oligonucleotide

## Contents

1. Introduction ..... 207
2. A definition of synthetic biology ..... 207
3. New sources for oligonucleotides ..... 208
4. Conclusions ..... 210
References ..... 211

* Corresponding author. Tel.: +49 622 1873 3440; fax: +49 622 1873 3442.
*E-mail address:* peer.staehler@febit.de (P. Stähler).

0168-1656/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2005.12.011

## 1. Introduction

The successful completion of the Human Genome Project and many other sequencing projects opened the door for another quantum jump in science advancement. New genome sequences are being finished almost daily, and many more will be forthcoming as we continue to add species from the evolutionary tree. This dynamic development already reminds some (Myhrvold, 2005) of Moore's law governing information technology. Moore's law is the empirical observation that at our rate of technological development, the complexity of an integrated circuit with respect to minimum component cost will double in about 24 months.

The most important public sequence databases are doubling in size every 18 months.

With all this information in hand, researchers have only just begun to think about new biological questions they can ask. By revealing the genetic program of many organisms, biomedical science has been transformed, endowing biologists with the ability to study the basic information of life in toto as an initial step toward a comprehensive understanding of the complexity of entire organisms. To work on the systems biology level will lead to a deeper understanding of life and have many important implications for the health care sector, the biotechnology industry and our environment.

## 2. A definition of synthetic biology

The purpose of this review is a discussion of another side of genomics: synthetic biology or "Synbio", with special emphasis on recent developments that promise a major breakthrough. We define synthetic biology as the making and use of biosystems founded on the synthesis of the coding DNA (and potentially RNA) based upon chemically synthesised nucleic acids oligonucleotides of programmable sequences. The recent developments discussed here make use of a rich source of oligonucleotides: in situ synthesised microarrays, which in fact represent nothing else but matrix nucleic acid synthesisers.

Synthetic biology is older than recombination technology. While the latter has been developed in the early 1970s (Smith and Wilcox, 1970; Jackson et al., 1972) after the successful discovery and use of sequence specific restriction enzymes and ligases, the making of synthetic genes goes back another decade. The first functional synthetic gene was made and published in 1964 by a research team led by H.G. Khorana, as part of their work on the elucidation of the genetic code (Khorana, 1968; Agarwal et al., 1974; Khorana, 1965). The gene carried the code for the tyrosine transfer RNA and was successfully tested in bacteria after being built from basic chemicals.

Others have later optimised and perfected the chemical synthesis of nucleic acids (Beaucage, 1993) and automated the complete process (Lashkari et al., 1995). For the two decades from 1980 to 2000 the predominant applications of the vast majority of the synthesised oligonucleotides were use as primers for PCR (Mullis et al., 1986) and sequencing (Sanger et al., 1977) as well as site-directed mutagenesis (Flavell et al., 1975; Gillam and Smith, 1979). During the 1990s, a growing number of biochip applications was added, (Schena et al., 1995; Fodor et al., 1991).

DNA recombination technology, at some point increasingly supported by PCR, mutagenesis and related methods, was faster and more cost effective then the creation of synthetic genes, and shaped the field of molecular biology. Synthetic genes played a minor role in the molecular biology laboratories, though they were always recognised as valuable tools. They sometimes might be the only accessible source for the desired DNA, as in cases where the natural DNA may be unavailable to the experimenter or the desirable DNA has never existed. Also, sequences that are deduced rather than experimental (e.g. resulting from protein sequencing or representing a new fusion of gene domains) can be made that way. Generally, synthetic genes are especially useful, if it is desired to re-engineer the target sequence, either the coding region or regulatory signals of importance (protein initiation, ribosome binding sites, promoters, and the like), and to alter the codon usage for a particular host or model organisms. Last but not least, synthetic genes allow the efficient construction of a family of related, but different constructs, with permutations at special regions of interest. Ultimately, synthetic genes allow complete flexibility of target sequence design and obviate intermediate steps often needed to get to a desired sequence by cloning and recombination technology.

The creation of synthetic genes is based on a simple concept: the target DNA is produced by programmed chemical synthesis of short oligonucleotides (typically

25–70 bases long): a combination of these oligonucleotides is dictated by specific hybridisation, from which longer DNA molecules are created by ligation, PCR, and a combination of the two methods (LCR). Using the ligation experimental scheme, there are two different approaches for the assembly of synthetic genes:

1. Synthesising the complete sequence (both strands) and ligating the backbone.
2. Synthesising part of the sequence while leaving gaps in the fragile hybrid, and filling the gaps by employing a DNA polymerase enzyme.

In the first approach, already used and optimised by Khorana and his team, a series of sequentially overlapping oligonucleotides are synthesised (Au et al., 1998). The sequences are selected with the gap in one strand bridged by an oligonucleotide of the other strand, so by simple hybridisation the precursor of the double stranded target DNA molecule containing nicks on both strands is formed. The nicks are then sealed in a reaction with DNA ligase to catalyse the formation of the phosphodiester bonds between the 5′-phosphate of one oligonucleotide fragment and the 3′-hydroxyl terminus of the adjacent oligonucleotide fragment.

The second approach (Dillon and Rosen, 1990) makes use of a template-directed and primer-dependent 5′- to 3′-synthesis reaction of a DNA-polymerase. The set of oligonucleotides is chosen such that after end annealing of the oligonucleotides, the polymerase uses deoxynucleotide triphosphates to fill the gaps. Any nick in the resulting, double-stranded DNA is later sealed by DNA-ligase. The length of the oligonucleotides used tends to be longer than in the Khorana method, and as an advantage one does not have to cover the complete sequence by the synthetic oligonucleotides, thus reducing cost. On the other hand, the frequency of mutations can be higher using this approach, due to the DNA polymerase.

Several modifications and improvements of these basic strategies have been tested and published to date (Stemmer et al., 1995). In all methods, the total length of the genetic element was restricted by the production costs of the necessary oligomers when made using the standard automated synthesisers, and by the logistics accompanied with the handling of the multitude of different molecules.

Two recent projects are of special significance. One is the chemical synthesis of poliovirus cDNA that directly led to the generation of infectious virus in the absence of natural template (Cello et al., 2002). Although the basic methodology is the same as the ones described above, it is the scope of the work that has created some discussion in the field. Simply by using a sequence stored in silico, Cello et al. were able to re-create an infectious virus. In principle, this method could be used to synthesise other viruses with similarly short DNA sequences. This includes viruses that could be considered to be potential bio-warfare agents, like Ebola virus. The poliovirus itself is not an effective biological weapon, but the experiment shows the potential of synthetic genetic engineering and also highlights one of its problems.

The other important project – or series of projects – is the creation of a minimal genome, ultimately made fully synthetically (Hutchison et al., 1999; Smith et al., 2003). A minimal genome is the smallest set of genes an organism needs to live in a particular environment. While most microbes have hundreds or thousands of genes, some use only a fraction of these at any one time, depending on their surroundings. The smallest genomes known only contain a few hundred genes. It is hoped that once the smallest set of genes is defined – as somewhat like a bottom line – one has a powerful model system at hand to create new functional microorganisms, for purposes such as energy generation or bioremediation.

The field is advancing in other aspects as well. Several sources are available for gene design software (Hoover and Lubkowski, 2002; Rouillard et al., 2004; Jayaraj et al., 2005), so the groundwork of disassembling a particular sequence to optimised short oligonucleotides is no bottleneck and should foster the spread of synthetic gene production.

## 3. New sources for oligonucleotides

One major bottleneck, however, remains to be solved: a source for the necessary oligonucleotides that is quicker and cheaper than the current standard DNA synthesisers. Several promising concepts emerged recently, all based on an especially rich source for the large number of different molecules that is required for synthetic genes (Weiler and Hoheisel,

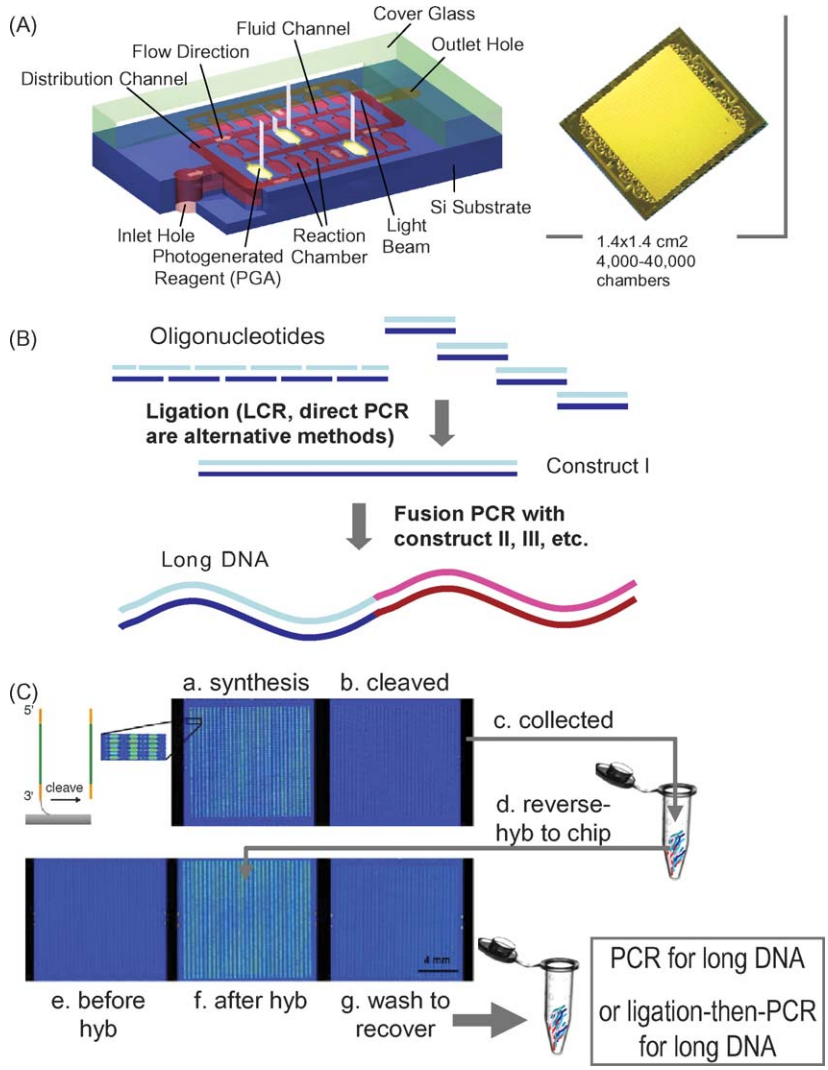


Fig. 1. Schematic illustration of microarray-based DNA synthesis. (A) The major features of the microfluidic chip. used in Tian et al. (2004) and Zhou et al. (2004). (B) Ligation or PCR approaches for long DNA synthesis. (C) The procedure used for synthesis of long DNA using array-derived oligos.

1996; Richmond et al., 2004; Zhou et al., 2004; Staehler et al., 2003; Tian et al., 2004): in situ microarray synthesis. With this new way of producing oligonucleotides in a cost-effective manner, the field of synthetic biology can be expected to change dramatically in the next decade (Fig. 1).

In a series of recent papers, several groups have published the use of array-derived oligonucleotides for synthetic biology purposes. Richmond et al. (2004) combined the Maskless Array Synthesiser (MAS) concept (Singh-Gasson et al., 1999) with a chemically labile linker. The oligonucleotides (up to 70 mers) created in situ on the microarray could be eluted by adding a cleaving reagent. To cope with the small amount of each oligonucleotide, they employed an amplification step that was made possible by adding to each oligonucleotide sequence a flanking generic primer sites of 15 bases. A simple PCR of the eluted oligonucleotides led to sufficient amounts of material for subsequent steps. The primer sequences were cleaved of by includ-

ing a restriction enzyme site. By such means, a set of 40 mer oligonucleotides was produced as an ingredient for a gene assembly. A first proof of principle of the complete process of amplification and assembly of chip-eluted DNA (AACED) resulted in a 180 bp DNA sequence that was verified qualitatively by restriction analysis. If the AACED will be extended to the full potential of the MAS platform, a single chip could potentially lead to the assembly of 15 Mb of DNA.

A second paper presents the use of a microfluidic synthesiser platform to make the oligonucleotides (Zhou et al., 2004). The microarrays were created in situ within small reaction chambers by combination of standard DNA synthesis chemistry and photo-generated acid activation (Gao et al., 2001), termed PicoArray synthesis. The authors synthesised 30 or 45 mer oligonucleotides and constructed several genes by using an approach of ligation followed by PCR with primers specific for the respective ends of designated sequence stretches. Interestingly, Gao et al. already employed the microarray for more then the synthesis alone. They additionally introduced a point check hybridisation to a complementary second microarray as quality control method. The genes were made in two steps: first, a set of intermediate sequence stretches of several hundred base pairs was assembled. Then several of these fragments were joined to form the target sequence. As an especially convenient functional assay, they chose the green fluorescent protein gene (742 bp) and report fluorescence of *E. coli* clones that were transfected with the synthetic gene.

Tian et al. (2004) showed the assembling of all 21 genes (also in a form of 14.6 kb operon) that encode the proteins of the *E. coli* 30 s ribosomal subunit, taking the system to the next level. They employed the PicoArray synthesis platform and announced to test the MAS platform in ongoing work, therefore using both array-derived oligonucleotide sources established in the previously discussed publications. Their process was similar to AACED and included a PCR step based on common flanking primer sites. As control experiments, they showed a chip with covalently labelled oligonucleotides still on the chip, the chip after cleavage and elution, and a hybridisation of the eluted oligonucleotides to a second, complementary chip (similar to the methods in Zhou et al., 2004). To reduce the rate of mutations in the final target sequence, the authors developed a method using array-derived oligonucleotides again: in addition to the set of oligomers coding the target sequence, called construction oligonucleotides, they synthesised and eluted a second set of so called selection oligonucleotides. Each of these molecules was half of the full-length of the respective construction oligonucleotide and linked to a biotin moiety to support known extraction protocols. Construction and selection oligonucleotides were used for a downstream selections step. The two sets were hybridised to each other and perfectly matching pair purified using the biotin moiety. This new approach even outperformed PAGE purification, giving a lower mutation rate at one error per 1394 bp.

## 4. Conclusions

With the use of array-derived oligonucleotides the field of synthetic biology has discovered a new source for the most important ingredient, the coding oligonucleotides. As a result of this, development costs of synthetic genes will drop significantly and their use spread in consequence. Also, one can then expect new applications to emerge. Church and co-workers already mentioned new biochemical activities of the engineered proteins and novel self-replicating systems (Tian et al., 2004).

At the same time, complementary DNA technologies will be developed or employed for synthetic biology applications. One important example is sequencing. With the cost of the oligonucleotides being reduced by three orders of magnitude, the downstream steps such as cloning, quality control and sequencing will become the major cost factors for synthetic genes. This might fit very well with recent developments in the field of high-performance sequencing (Shendure et al., 2004) that promise to lower the cost per base by a factor of 10 in the mid term range. If one of the successful novel sequencing tools is combined with the array-derived oligonucleotides, one can envision such miniaturising technologies will drive another dramatic decrease in price and increase in speed and turnover. As some of the advanced sequencing methods include the use of microarrays themselves, hybridisation based or via sequencing by synthesis, the array is used at more then one point of the process chain.

The downstream processing will benefit from optimisation and fine tuning work. A promising approach includes the use of mutation detection proteins (MutS from *T. aquaticus*) (Carr et al., 2004; Binkowski et al., 2005) to deplete hybridised oligomer sets of mismatched pairs. Error rate reduction of more than 15-fold was reported.

The further development of synthetic biology will most likely see two trends: for once, longer sequences will be synthesised and utilised in innovative applications. A small community of advanced academic laboratories are likely to lead, among them the ones referred to in this review. As a second trend, however, we expect that the array-derived oligonucleotides will continue to be optimised in quality and in the formats for use to become standard reagents in the daily work of life science researchers around the globe. Given the unique properties of oligonucleotides, i.e. they can be easily duplicated, amplified, extended in length and detected at low abundance or even at single molecule level, new applications will emerge in areas which have not seen the role of oligonucleotides. Within 10–20 years, synthetic genes will be the tools of choice to obtain any sequence at any time in any laboratory. This evolution will be very beneficial for all areas of molecular biology. Eventually, synthetic genes could replace recombination technology for most purposes. We witness a huge asymmetry between the amount of sequence information available in databases and its utilisation in practical experiments. Investigating a gene's function requires a physical copy of that gene's DNA and derivatives thereof. The same is true for sets of genes, chromosomes or a complete genome. This simple fact represents a major bottleneck in the path of advancement for making use of the wealth of sequence information stored. Synthetic biology will eventually balance the growing amount of sequence information with equal numbers of synthetic genes.

## References


Agarwal, K.L., Buchi, H., Caruthers, M.H., Gupta, N., Khorana, H.G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U.L., Van de Sande, J.H., 1974. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. Nature 227, 27–34.

Au, L.-C., Yang, F.-Y., Yang, W.-J., Lo, S.-H., Kao, C.-F., 1998. Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli*. Biochem. Biophys. Res. Commun. 248, 200–203.

Beaucage, S.L., 1993. Oligodeoxyribonucleotide synthesis: phosphoramidite approach. In: Agrawal, S. (Ed.), Protocols for Oligonucleotides and Analogs. Methods in Molecular Biology series, vol. 20. Humana Press, New Jersey, pp. 33–62.

Binkowski, B.F., Richmond, K.E., Kaysen, J., Sussman, M.R., Belshaw, P.J., 2005. Correcting errors in synthetic DNA through consensus shuffling. Nucleic Acids Res. 33, e55.

Carr, P., Park, J.S., Lee, Y.-J., Yu, T., Zhang, S., Joseph, M., Jacobson, J.M., 2004. Protein-mediated error correction for de novo DNA synthesis. Nucleic Acid Res. 23, 32 (20), e162.

Cello, J., Paul, A.V., Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297, 1016–1018.

Dillon, P.J., Rosen, C.A., 1990. A rapid method for the construction of synthetic genes using the polymerase chain reaction. Biotechniques 9, 298–300.

Flavell, R.A., Sabo, D.L., Bandle, E.F., Weissmann, C., 1975. Site-directed mutagenesis: effect of an extracistronic mutation on the in vitro propagation of bacteriophage Qbeta RNA. Proc. Natl. Acad. Sci. U.S.A. 72, 367–371.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D., 1991. Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767–773.

Gao, X., LeProust, E., Zhang, H., Srivannavit, O., Gulari, E., Yu, P., Nishiguchi, C., Xiang, Q., Zhou, X., 2001. Flexible DNA chip synthesis gated by deprotection using solution photogenerated acids. Nucleic Acids Res. 29, 4744–4750.

Gillam, S., Smith, M., 1979. Site-specific mutagenesis using synthetic oligodeoxyribonucleotide primers: II. In vitro selection of mutant DNA. Gene 8, 99–106.

Hoover, D.M., Lubkowski, J., 2002. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 30, e43.

Hutchison III, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C., 1999. Global transposon mutagenesis and a minimal mycoplasma genome. Science 286 (5447), 2165–2169.

Jackson, D.A., Symons, R.H., Berg, P., 1972. Biochemical method for inserting new genetic information into DNA of Simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 69 (10), 2904–2909.

Jayaraj, S., Reid, R., Santi, D.V., 2005. GeMS: an advanced software package for designing synthetic genes. Nucleic Acids Res. 33 (9), 3011–3016.

Khorana, H.G., 1965. Polynucleotide synthesis and the genetic code. Fed. Proc. 24 (6), 1473–1487.

Khorana, H.G., 1968. Nucleic acid synthesis in the study of the genetic code. In: Nobel Lectures: Physiology or Medicin (1963–1970), vol. 5. Elsevier Science Ltd., Amsterdam, pp. 341–369.

Lashkari, D.A., Hunicke-Smith, S.P., Norgren, R.M., Davis, R.W., Brennan, T., 1995. An automated multiplex oligonucleotide synthesizer: development of high-throughput, low-cost DNA synthesis. Proc. Natl. Acad. Sci. U.S.A. 92, 7912–7915.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the poly-

merase chain reaction. Cold Spring Harb. Symp. Quant. Biol. LI, 263–273.

Myhrvold, N., 2005. The Institute for Systems Biology's Computational Challenges in Systems Biology Symposium.

Richmond, K.E., Li, M.H., Rodesch, M.J., Patel, M., Lowe, A.M., Kim, C., Chu, L.L., Venkataramaian, N., Flickinger, S.F., Kaysen, J., Belshaw, P.J., Sussman, M.R., Cerrina, F., 2004. Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. Nucleic Acids Res. 32, 5011–5018.

Rouillard, J.M., Lee, W., Truan, G., Gao, X., Zhou, X., Gulari, E., 2004. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. Nucleic Acids Res. 32, 176–180.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74 (12), 5463–5467.

Schena, M., Shalon, D., Davis, R., Brown, P., 1995. Science 270, 467–470.

Shendure, J., Mitra, R.D., Varma, C., Church, G.M., 2004. Advanced sequencing technologies: methods and goals. Nat. Rev. Genetics 5, 335–344.

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F., 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat. Biotechnol. 17 (10), 974–978.

Smith, H.O., Hutchison III, C.A., Pfannkoch, C., Venter, J.C., 2003. Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. 100, 15440–15445.

Smith, H.O., Wilcox, K.W., 1970. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. J. Mol. Biol. 51 (2), 379–391.

Staehler, P., Staehler, C., Mueller, M., 2003. Method for Producing Polymers. US Patent no. 6,586,211.

Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164, 49–53.

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., Church, G., 2004. Accurate multiplex gene synthesis from programmable DNA microchips. Nature 23, 432 (7020), 1050–1054.

Weiler, J., Hoheisel, J.D., 1996. Combining the preparation of oligonucleotide arrays and synthesis of high quality primers. Anal. Biochem. 243, 218–227.

Zhou, X., Cai, S., Hong, A., You, Q., Yu, P., Sheng, N., Srivannavit, O., Muranjan, S., Rouillard, J.M., Xia, Y., Zhang, X., Xiang, Q., Ganesh, R., Zhu, Q., Matejko, A., Gulari, E., Gao, X., 2004. Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. Nucleic Acids Res. 32, 5409–5417.