**Gene Synthesis on Microchips**

Joachim W. Engels

With gene synthesis we will in future be able to synthesize complete genomes. From there, we will be able to answer essential biological questions, for example, how chemical reactions proceed within a cell. The first step, genome sequencing, especially of the human genome, has created a huge amount of data. It is currently estimated that humans have about 25,000 genes. Chemically speaking, genes are merely defined sequences of long DNA stretches. In the last 40 years, methods for the chemical synthesis of DNA have improved drastically.\[1–3\] These processes are based on the stepwise coupling of individual nucleotide building blocks to yield oligonucleotides. These in turn, arrange themselves into double-stranded DNA owing to their base complementarity. Hydrogen bonding followed by chemical or chemoenzymatic ligation of the phosphoric esters results in the stable duplexes. By exploiting this principle, Khorana and co-workers were able to synthesize the first gene in 1970, the yeast tRNAAla gene.\[3\] Since then, synthetic methods have been refined, most noteworthy by the application of the polymerase chain reaction (PCR), and the goals have become more ambitious.\[4–6\] The subsequent years witnessed the syntheses of the following genes: α-interferon (1985),\[7\] the interleukin-2 receptor α (1987),\[8\] and a plasmid (1990).\[9\] The year 2002 saw the synthesis of a polio virus DNA\[10\] with around 7500 base pairs. Its biological activity was proven by its infectivity. The synthesis of the bacteriophage \(\Phi X 174\) with 5386 base pairs was reported in 2003;\[11a\] this molecule was first sequenced by Fred Sanger with the newly developed dideoxy sequencing method.\[11b\] The longest recorded gene

---

**Highlights**

- **Synthetic Biology**
- **Gene Synthesis on Microchips**
- **Joachim W. Engels**

**Keywords:**
- chemical biology
- hybridization
- microchips
- oligonucleotides
- polymerase chain reaction

The synthesis of the bacteriophage \(\Phi X 174\) with 5386 base pairs was reported in 2003;\[11a\] this molecule was first sequenced by Fred Sanger with the newly developed dideoxy sequencing method.\[11b\] The longest recorded gene...
synthesized to date is the 32,000 base pair polyketide synthase gene cluster (2004).\[12\] This gene was also biologically active and yielded 6-deoxyerythronolide after transformation in E. coli.

Combinatorial or parallel syntheses yield a variety of chemical structures, whose functions can be analyzed through appropriate assays. Genes on the megabase scale (e.g. human DNA) can, in principle, be synthesized chemically or chemoenzymatically. Oligonucleotides in the range of 20 to 100 mers are now commercially available for about 0.10 euro per nucleotide and have error rates of around 1:100 to 1:400. These errors are mainly due to problems that occur during the chemical synthesis, such as incomplete reactions during coupling and deprotection and side reactions caused by the reagents used (e.g. Michael addition of acrylonitrile to thymine). High-throughput gene synthesis is limited by economic factors and error rates; cloning and sequencing raise the price about tenfold. Consequently, parallel syntheses on microchips were introduced to reduce the costs. The yield of the oligonucleotides turned out to be problematically low (atto- and femtomolar range). These 10^3–10^5 molecules must then be amplified by PCR to 10^5–10^7 molecules. Gao, Church, and co-workers have recently showed\[14\] how approximately 4,000 oligonucleotides can be synthesized simultaneously on light programmable microarrays (Figures 1a and b).\[15–17\] Photolabile protecting groups or photoliberated protons and the dimethoxytrityl (DMTr) protecting group allow on-chip synthesis according to the phosphoramidite method of Caruthers and Beaucage.\[1\] As mentioned above, errors appear during the chemical synthesis on chip and hence Gao, Church, and co-workers chose stringent hybridization as a cheap means of quality control, utilizing so-called quality assessment chips synthesized with the complementary oligonucleotides. Error-containing oligonucleotides, for example, oligonucleotides with the incorporation of wrong bases (point mutations) were purified by affinity chromatography on immobilized short complementary oligonucleotides. In this step it is important to synchronize the melting points (T_m values) isothermally by choosing oligonucleotides of different lengths.

An important point in gene design is the selection of the restriction enzymes to be used. Restriction enzymes of the type IIS were chosen which cut the DNA outside the recognition sequence and thus liberate internal sequences at will—in this case gene sequences (Figure 2). The authors used the enzymes BsaI with the recognition sequence 5'-GGTCTC(1/5) and a four-base 5'-overlap and BseRI with the recognition sequence 5'-GGGAG(10/8) and a two-base 3'-overlap of undefined sequence. In addition to the gene sequences, selection oligonucleotides immobilized on streptavidin beads through a biotin linker (Figure 3) were used for separation. Only the unmodified, error-free oligonucleotides (in this case 50-mers) were selected! The exactly fitting oligonucleotides were filtered out in two steps with the help of 25-mer complementary oligonucleotides R and L (so-called capture probes). Following this purification by selection, the gene-assembly experiment is developed on microtiter plates (e.g. 96 or 384 wells) by the polymerase assembly multiplexing (PAM) reaction. In this case the dsDNA, which is held together by hydrogen-bonding interactions, is assembled into a gene by the action of polymerases, without a primer. The quality of the oligonucleotides used for genesynthesis was tested in three different formats: 1) unpurified (as synthesized), 2) purified by gel electrophoresis, and 3) purified through hybridization on affinity beads.\[18\] The latter method showed the lowest error rate (1:1394 bp), which is an improvement
E. coli the small 30S ribosome subunit of codon-varied genes for 21 proteins of efficiency. It was hoped that the forma-
discussed as the cause of low synthetic structures of the mRNA have been ing the appropriate codons. Secondary efficiency would be increased by choos-
and it was hoped that the translation rate of these 21 proteins in vitro is low
Church, and co-workers[14] synthesized this method in synthetic biology, Gao,
polymer support followed by enzymatic assembly by chemical synthesis on a
(roughly by a factor of 10) over the assembly by chemical synthesis on a polymer support followed by enzymatic ligation.
As an example for the application of this method in synthetic biology, Gao, Church, and co-workers[14] synthesized codon-varied genes for 21 proteins of the small 30S ribosome subunit of E. coli.[19] It is known that the expression rate of these 21 proteins in vitro is low and it was hoped that the translation efficiency would be increased by choosing the appropriate codons. Secondary structures of the mRNA have been discussed as the cause of low synthetic efficiency. It was hoped that the forma-
nucleotides was estimated to be 1:7300 bp! This clearly proves that the error rate of synthetic oligonucleotides is the limiting step in chemoenzymatic gene synthesis and not the precision of the polymerases, which have an error rate of about 1:100000 bp.
This experiment clearly demonstrates the potential of on-chip oligonu-
ucleotide synthesis. Given the present capacity of microchips, large-scale gene-synthesis experiments are possible. A rough calculation indicates the possi-
bility of lowering the costs for gene synthesis from 10 bp per euro to 20 kbp per euro. This is possible by using, for example, microchips from NimbleGen (www.nimblegen.com) with an oligonucleotide density of 95000–382000, to give a final sum of 2–18 Mbp DNA.
In an effort to reduce the error rate for the synthetic gene even further, a correction of the assembled oligonucleotides or gene fragments by a protein from the DNA-repair machinery (e.g. mutS) is being pursued.[20] Thus, Jacob-
son and co-workers corrected a synthetic DNA with the fixed mismatch binding protein mutS from Thermus aquaticus. They eliminated mismatch products and obtained DNA with an error rate of around 1:10000. Another alternative to obtain completely error-free DNA is a cheap high-throughput-sequencing method in which thousands of sequences can be read simultaneously.[21]
In summary, the methods and procedures for the synthesis of genes were mostly known, but they have now been combined in a new way to allow more-efficient and economic access to error-
free genes in the 10-kb range. The individual steps include (estimated times in brackets):[14]
a) Design and modeling of the appro-
riate oligonucleotide building blocks for the gene synthesis (2 h).
b) Oligonucleotide synthesis in micro-
fluidic devices (8 h).
c) Amplification of the synthetic, on-
chip oligonucleotides by PCR (14 h).
d) Cutting the amplified oligonucleoti-
des by type IIS restriction enzymes to the appropriate size (50-mers).
e) Selection of the correct sequences through hybridization (30 h).
f) Assembly of the final gene by poly-
merase fill-in followed by PCR (4 h).
g) In vitro transcription/translation and detection of mRNAs and proteins (1 h).
As potential application one can envisage the synthesis of gene clusters for natural products such as polyketides or the resynthesis of bacterial cell ge-
nones(e.g. mycoplasma). Thus gene synthesis will be a way to evolve new secondary metabolites such as poly-
ketides or peptide antibiotics to obtain non-natural isomers of natural products. Gene synthesis is key to functional
genomics and will result in the assignment of all chemical reactions within a cell. The ultimate goal of biological synthesis will be to find new arrangements of genes from a gene pool and hence to search for new solutions for new products.


