

Biomolecular Engineering

Zengyi Shao
Ee Lui Ang
Huimin Zhao

*Department of Chemical and Biomolecular Engineering, University of Illinois,
Urbana, Illinois, U.S.A.*

INTRODUCTION

Biomolecular engineering is an emerging field of academic research and industrial practice having the goal of engineering value-added biomolecules and biomolecular processes for applications in medical, chemical, agricultural, and food industries.^[1] This new subject is very broad and highly interdisciplinary, including, but is not limited to, protein engineering, metabolic engineering, bioinformatics, bioprocessing, gene therapy, drug design, discovery and delivery, biomaterials, and nano-biotechnology. In the past decades, biomolecules such as protein pharmaceuticals, DNA vaccines, monoclonal antibodies, and industrial enzymes have been increasingly commercialized. In this entry, the key concepts and technologies in biomolecular engineering and their applications in engineering proteins, pathways, and nucleic acids are introduced.

KEY CONCEPTS AND TECHNOLOGIES EMPLOYED IN BIOMOLECULAR ENGINEERING

Protein Engineering

Protein engineering refers to the ability to alter protein structure to achieve a desired protein function. Two main protein engineering approaches, rational design and directed evolution, have been developed in the past two decades (Fig. 1). The former involves alterations of selected residues in a protein via site-specific mutagenesis to achieve predicted changes in function. In comparison, directed evolution mimics the process of natural evolution in the test tube, involving repeated cycles of creating molecular diversity by random mutagenesis/gene recombination, followed by screening/selecting the functionally improved variants.

Metabolic Pathway Engineering

Metabolic pathway engineering involves the directed improvement of product formation or cellular properties through the modification of specific biochemical

reaction(s) or through the introduction of new one(s) by recombinant DNA technology.^[2] Specifically, metabolic engineering includes identification of metabolic pathways, elucidation of regulatory mechanisms, metabolic flux analysis, metabolic control analysis (MCA), identification of inter- or intra-cellular transport mechanisms, and discovery and manipulation of biosynthetic pathways. This area is particularly important to biotechnology because it offers ways for improving existing bioprocesses, designing new bioprocesses, as well as producing novel chemicals and pharmaceuticals.

Many molecular biology and analytical chemistry tools have been developed for metabolic engineering.^[3] For example, various vectors have been designed for the optimal expression of heterologous genes in industrial hosts. Several gene-cloning techniques have been developed to isolate novel genes from cultivable or uncultivable organisms, whereas RNA-antisense techniques have been used to silence gene expressions. In addition, directed evolution methods have been used to construct genes, pathways, or whole genomes with altered functions. As important as these genetic engineering tools, a number of powerful analytical techniques have also been developed for metabolic pathway analysis and analyses of cellular functions, such as gas chromatography and mass spectrometry (GC-MS), nuclear magnetic resonance (NMR), two-dimensional gel electrophoresis, DNA chips, and protein chips.

Because of the complexity of metabolic networks, rationally designed metabolic pathways often have undesired metabolic consequences on unrelated cellular properties. To address this limitation, a new metabolic engineering strategy of particular interest, “inverse metabolic engineering” (IME), was developed^[4] (Fig. 2), which integrates directed evolution principles with the “direct” classical metabolic engineering. The strategy begins with the construction and identification of a desired phenotype; then the genetic basis or environmental factor for the desired phenotypic characteristic is determined; finally, this phenotype is endowed on another strain or organism by genetic manipulation. The essential and challenging step of IME is to identify the genetic basis of the desired phenotype,^[5] which is illustrated later.

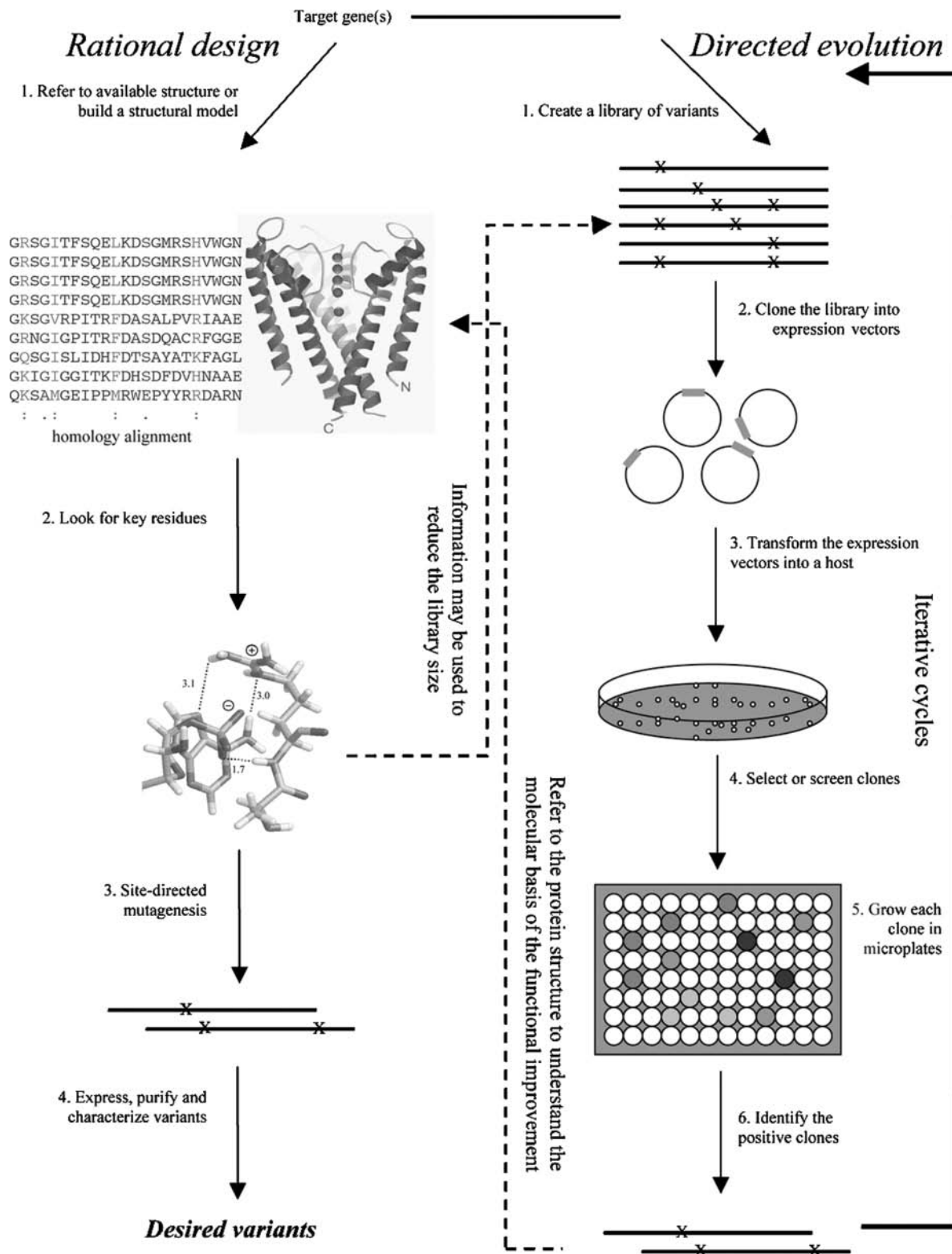


Fig. 1 Comparison between rational design and directed evolution. (View this art in color at www.dekker.com.)

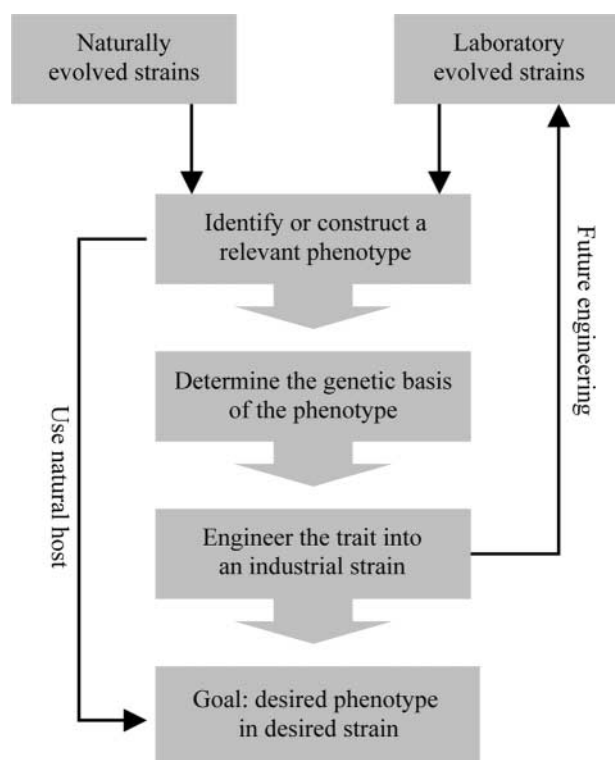


Fig. 2 The scheme of the inverse metabolic engineering (IME) approach. (From Ref.^[5].)

Another two key concepts in metabolic engineering are metabolic pathway analysis and metabolic pathway modeling. The former is used for assessing inherent network properties in the complete biochemical reaction networks. It involves identification of the metabolic network structure (or pathway topology), quantification of the fluxes through the branches of the metabolic network, and identification of the control structures within the metabolic network.^[3]

To identify metabolic network structure of common organisms, it is often helpful to do an extensive literature review. Also, a few pathway databases, such as BioCyc Knowledge Library (www.biocyc.org), may be used to identify specific metabolic pathways. However, in some cases, it is not easy to identify the complete metabolic network. This will cause a serious problem as some parts of the pathways that carry significant metabolic fluxes will possibly be ignored, resulting in only partial or even false estimation and analysis of the network structure. In these cases, enzyme assays and isotope-labeling techniques combined with either NMR or GC-MS are valuable tools for confirming the presence of specific enzymes and pathways. Once the whole metabolic network structure has been identified, quantification of the flux distribution through branches can be determined by solving a series of algebraic mass balance equations in terms of substrates, products, biomass constituents,

and intracellular metabolites. For example, Fig. 3 shows how the carbon fluxes are distributed through the individual branches of metabolic pathways of *Saccharomyces cerevisiae* under anaerobic growth. It offers insights into other important cell physiological characteristics that can be used for controlling and directing metabolic fluxes, identifying alternative pathways, and calculating the maximum theoretical yields of the products.^[6] Once the fluxes through the different branches are quantified, controlling the fluxes can be attempted. It is necessary and important to understand how the enzymes at the branch points are regulated. Various types of regulatory mechanisms have been identified, such as feedback inhibition, cooperativity, modification of covalent enzymes, and control of enzyme synthesis.^[6] Determining the metabolite concentrations is also important for understanding how metabolic fluxes are controlled. In this case, GC-MS and LC-MS-MS are good choices for measuring a large number of metabolites.^[3] Furthermore, to quantify the flux control, the concept of metabolic control analysis is commonly used with flux control coefficients (FCC), which is defined as the relative change in the steady-state flux, resulting from an infinitesimal change in the activity of an enzyme of the pathway, divided by the relative change of enzymatic activity.^[6]

For pathway modeling, the typical aims include exploration of the possible behavior of a system, interpretation and evaluation of measured data, systematic analysis of a metabolic network structure and its qualitative behavior, and designing/predicting/optimizing the outcome of future experiments. The different modeling approaches currently used, including structure model, stoichiometric model, carbon flux model, stationary and nonstationary mechanistic model, and gene regulation model, are discussed and evaluated by Wiechert.^[7]

Nucleic Acid Engineering

Previously, nucleic acids were considered as the carriers of genetic information in cells, with very few other functions. However, with the recent discovery of ribozymes (catalytic RNA), aptamers (binding RNA/DNA), and DNAzymes (catalytic DNA), nucleic acids have also been increasingly explored for diagnostic and therapeutic applications. As nucleic acids carry both structural (genotype) and functional (phenotype) information in a single molecule, they are particularly amenable to powerful *in vitro* selection methods such as SELEX (Systematic Evolution of Ligands by Exponential Enrichments) (Fig. 4).^[8] The principle behind these selection methods is the same as that of directed protein evolution, where the DNA or RNA molecules are put through iterative rounds of diversification and selection.

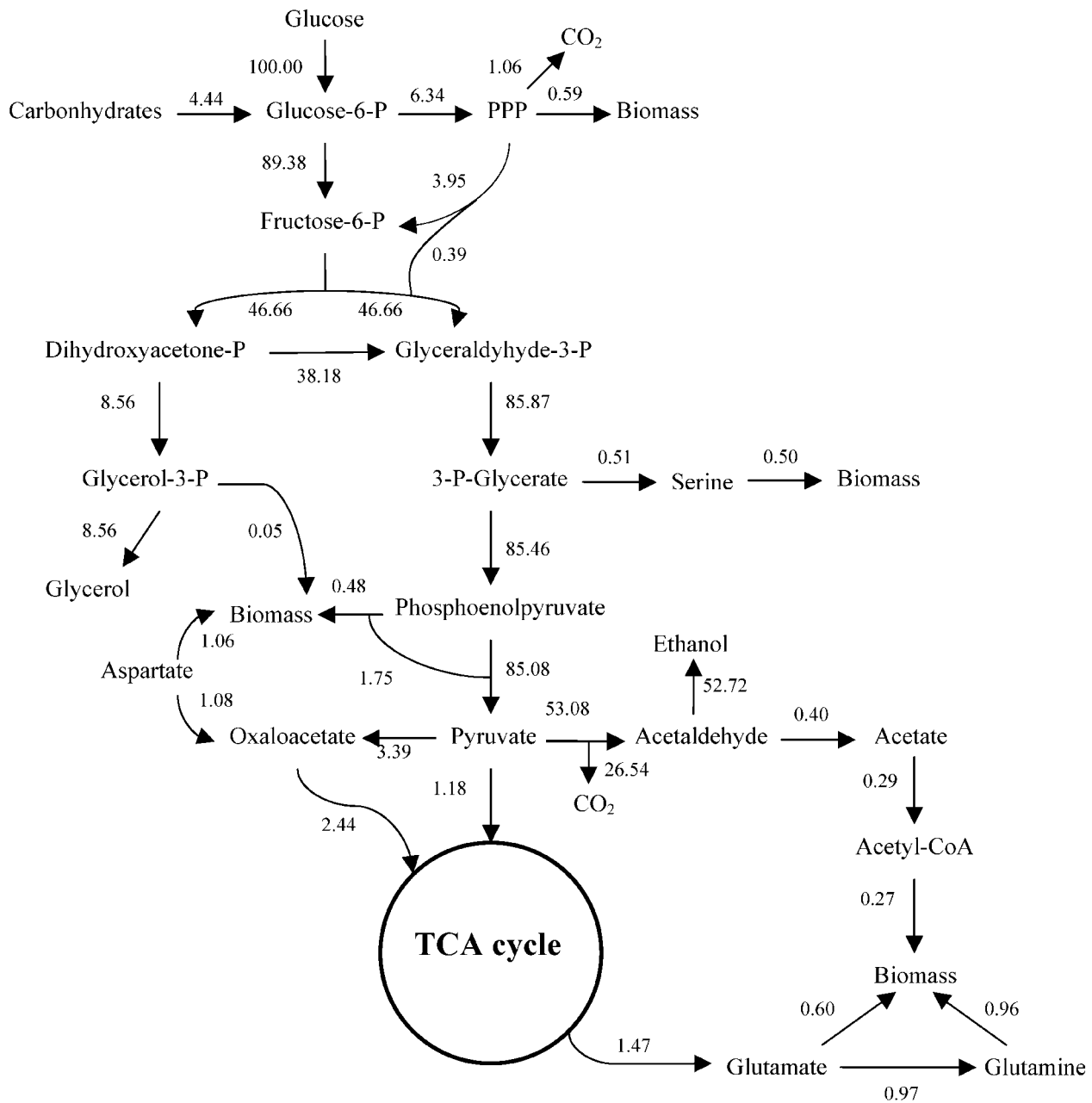


Fig. 3 An example of metabolic flux analysis showing fluxes through the different metabolic pathways of *S. cerevisiae* under anaerobic growth. (Adapted from Ref.^{[6].})

Although the different classes of therapeutic nucleic acids work via different mechanisms to treat diseases, the development of all forms of nucleic acid molecules into practical therapeutic agents faces a common obstacle—the omnipresence of nucleases in biological fluids. Nucleases degrade the DNA or RNA agents in the body, shortening the life span and consequently the efficiency of the nucleic acid therapeutic agents.^[9] As a result, larger doses are required to achieve desired results, making the treatment very expensive. There are several methods to overcome this limitation, first of

which is modification of bases. By replacing the 2'-OH-group in RNA with a host of different bases, such as 2'-NH₂- and 2'-F-pyrimidines, the stability of aptamers that were used as RNA-based therapeutic agents can be greatly increased.^[10] However, 2'-modified aptamers must be compatible with the overall SELEX protocol. Therefore, any modified nucleotides used must be recognizable by the various polymerases used in the amplification process. Alternatively, the use of "Transcription Free SELEX" bypasses the need for polymerase compatibility by allowing random RNA

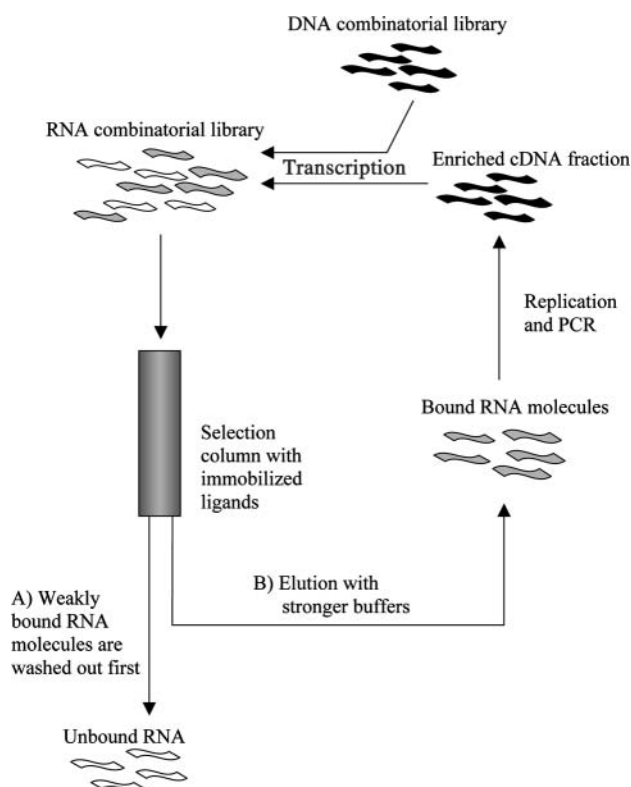


Fig. 4 The scheme of the SELEX method.

nucleotide fragments to bind to random DNA templates. Standard chemical reactions such as carbodiimide^[11] or cyanogen bromide condensations^[12] can then be used to ligate these fragments, and the RNA aptamers can then be recovered by melting the duplex.^[13]

The modification of bases has also been applied to another type of RNA therapeutic agent, the antisense oligonucleotides. Currently, the modifications can be classified into various generations. In the first generation, modifications were done on the backbone of the oligonucleotides, a common example is the replacement of one nonbridging oxygen atom with sulfur.^[14,15] These were known as phosphorothioate (PS) oligonucleotides. One successful example is VitraveneTM, which is currently the only FDA-approved antisense drug in the market. The second generation contains sugar modifications, with the substitution of the 2'-OH groups being the most common. Oligonucleotides with 2'-O-alkyl derivatives were found to be more resistant to DNA- or RNA-cleaving enzymes. Currently, 2'-O-methyl and 2'-O-alkyl have already been incorporated into a number of oligonucleotides in clinical trials.^[16] The third generation of antisense RNA consists of zwitterionic oligonucleotides.^[17] It was found that a modification of 2'-O-aminopropyl (AP)-RNA displayed a much higher resistance to snake venom

phosphodiester (SVPD) compared with a 2'-O-butyl-modified RNA.^[18] AP-RNAs were able to competitively inhibit the degradation of single-stranded DNA by *Escherichia coli* Klenow fragment (KF) 3'-5' exonuclease and SVPD. Crystal structure studies of AP-RNA revealed that the positively charged 2'-O-substituent is able to interfere with the metal-ion binding site B of the KF exonuclease, slowing down the degradation process.

Another method of improving nucleic acid stability is the use of mirror image molecules. Substrate recognition by natural nucleases is inherently stereospecific because these enzymes consist of only L-amino acids. Thus, only (D-)oligonucleotides can be recognized and degraded by the nucleases. To increase the stability of aptamers in biological fluids, nuclease resistant aptamers can be generated by using non-natural L-nucleotide aptamers, which are also known as spiegelmers ("spiegel" meaning mirror in German).^[19] However, because of the very reason that spiegelmers can escape nuclease recognition, no natural enzymes are able to recognize and amplify such nucleic acids. This greatly limits the screening of spiegelmers for potential drug agents as they cannot be directly screened using SELEX methods. To overcome this problem, unmodified D-RNA or D-DNA libraries were screened against the mirror image of the natural drug target instead.^[19] Following the rules of symmetry, the mirror image of the selected D-aptamers will in turn bind to the natural drug targets.

Bioinformatics

Bioinformatics is a rapidly expanding field, involving the application of computer technology to the management of biological information. Here, only two of the key bioinformatics components related to biomolecular engineering, database and computer modeling, are discussed.

The successful sequencing of the genomes from more than 100 organisms, including humans, has led to the increasing use of genomic databases such as GenBank, European Molecular Biology Laboratory Nucleotide Sequence Database (EMBL), and DNA Data Bank of Japan (DDBJ). Each of these three international databases collects and archives the sequencing data reported worldwide on a daily basis.^[20] One extremely powerful tool for searching these nucleotide and protein databases is the Basic Local Alignment Search Tool (BLAST),^[21] which compares a target DNA or protein sequence to all the existing sequences in the databases to find known sequences with high sequence similarities, and thereby obtain insights into the structure and function of the DNA or protein of interest. This tool is freely available to the public on the

National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov/BLAST). It should be noted that many powerful tools for web-based database searching and sequence analysis have been developed, such as the San Diego Supercomputer Center Biology Workbench (SDSC Biology Workbench, workbench.sdsc.edu). In the SDSC Biology Workbench, database searching is combined with access to a wide variety of sequence analysis and modeling tools, which makes it very convenient for researchers to analyze nucleic acids and proteins.

The availability of a protein crystal structure is important for understanding the molecular basis of protein functions as well as for engineering new or improved protein functions. However, the protein structures are typically solved by x-ray crystallography and nuclear magnetic resonance (NMR), which are very time-consuming and laborious processes. As a result, the number of solved protein structures is far fewer than that of known proteins. Within this context, the use of computer-modeling algorithms and programs to rapidly create a structural model of the target protein *in silico* is a very attractive alternative to experimental structure-determination methods. In general, there are three basic methods for structure prediction: *ab initio*, homology (comparative) modeling, and fold recognition (threading).

Ab initio methods rely on the fact that the folded protein is in a state of lowest free energy; hence, the predicted structures are created through energy minimization based solely on possible interactions between the residues. As *ab initio* methods do not refer to existing protein structures, they are typically computation intensive. However, theoretically these methods could be superior, as, unlike homology modeling and fold recognition, they are not biased by existing structural information. Many different types of empirical force fields and levels of structure description are used in *ab initio* methods. Homology modeling attempts to predict protein structure based on the similarity between the sequence of a protein and the sequences of other proteins of known structures, whereas fold recognition defines a database of known structures and then fits (threads) the sequence to that database, evaluating the score that assesses the suitability of each possible fit to a given fold. For high sequence identities (>30%), homology modeling usually gives a reasonably accurate model. In cases of low sequence identities (<30%), threading methods often yield more accurate models than does homology modeling,^[22] especially considering that there are some cases where folds are the same, yet sequences are very different (low identity).

As mentioned earlier, many bioinformatics tools were developed by individual research groups all around the world and are freely available on the

Internet. In addition, there are also a few commercial software programs, such as Insight II (Accelrys Inc., San Diego, California) and SYBYL[®]/Base (Tripos, Inc., St. Louis, Missouri). These programs integrate force fields, energy minimization, homology modeling, molecular dynamics simulation, and molecular visualization into a single platform, making them very powerful tools for protein analysis, modeling, and design.

EXAMPLES

Protein Engineering

There are many examples of protein engineering that use rational design and directed evolution.

Metabolic Pathway Engineering

Metabolic pathway engineering has numerous applications in food, agriculture, chemical, and pharmaceutical industries. Examples include, but are not limited to, increasing the yield of antibiotics, biosynthetic precursors, or polymers, expanding the metabolic capacity to degrade harmful compounds, or producing novel compounds that cannot be found in nature.

Biosynthesis

To establish metabolic pathways in a production host that are able to channel carbon flux to a desired product at a high yield requires careful consideration of the entire metabolic microenvironment in the host. These efforts normally include four key elements: 1) direct and optimize the primary metabolic pathway flux to the target product, including removal of rate-limiting steps, and transcriptional and allosteric regulation; 2) genetically block competing branch pathways; 3) modify secondary metabolic pathways to enhance energy metabolism and availability of required enzymatic cofactors; and 4) remove detrimental side-products.^[23] For example, the central glucose metabolic pathway of *E. coli* was engineered to achieve high recombinant protein production. The acetate accumulated at the end of the central glycolysis pathway affected both the recombinant protein production and the cell density detrimentally.^[24] Strategies investigated to reduce acetate accumulation included modification of glucose uptake rate by a glucose analog methyl α -glucoside,^[25] redirection of carbon flux toward a less inhibitory byproduct, acetoin, by introducing acetolactate synthase gene of *Bacillus subtilis* into *E. coli* to convert pyruvate to acetoin instead of acetate,^[26] and elimination of critical enzymes, including acetate kinase and phosphotransacetylase that are involved directly in

the formation of acetate.^[27] Another example is synthesis of hydrocortisone, an important starting material for steroidal drug synthesis, from glucose in yeast.^[28] The whole process involved the natural yeast biosynthetic pathways, a yeast pathway rerouted by one plant enzyme and other five enzyme steps catalyzed by eight mammalian proteins. Experimentally, recombinant *S. cerevisiae* was engineered to express 13 assembled genes, while several unwanted side reactions brought about by endogenous yeast genes were disrupted, efficiently directing endogenous carbon flux toward the target product, hydrocortisone.

Biodegradation

Some natural organisms can degrade harmful organic compounds, referred to as xenobiotics, such as aromatics, halogenated aliphatics, and pesticides. However, these naturally occurring degradation processes are extremely slow and inefficient. Thus, speeding up these processes by metabolic engineering has become an attractive strategy. Most work in this field has focused on biodegradation of aromatic hydrocarbons, among which the TOL pathway of *Pseudomonas putida* was studied extensively.^[29] The TOL pathway is a plasmid-encoded pathway and can utilize toluene, *m*-, and *p*-xylene as sole sources of carbon and energy. However, it cannot use benzene, which is often present with toluene and xylene as a mixture. Heterologous expression of a toluene dioxygenase in a *P. putida* strain carrying the TOL pathway resulted in a recombinant micro-organism that can degrade all these four aromatic compounds at a high rate.^[30]

Inverse metabolic engineering

The strategy of IME starts from variant construction, which can be classified into two categories: “exogenous” mutagenesis and “endogenous” mutagenesis. The former, which is usually plasmid based, involves directed evolution of proteins, whereas the latter mainly involves evolutionary engineering of host chromosome(s). An approach to create more robust strains by random mutagenesis of the *E. coli* chromosome was successfully carried out by transforming an exogenous plasmid pmut containing the *mutD* gene into the host.^[31] The *mutD* gene encodes the ϵ -subunit of DNA polymerase III, which is responsible for proof-reading of DNA replication. *mutD5*, which carries two amino acid substitutions, is non-functional, but still competes with the functional MutD protein produced from the chromosomal copy *mutD* gene, resulting in an increase of the mutation frequency of *E. coli*. In this way, a number of *E. coli* cells with a

broad spectrum of base substitutions and even frame-shift mutations were obtained. Among these, it was found that three bacterial strains can tolerate dimethylformamide up to 10–20 g/L. Another approach to create novel traits in microbes is genome shuffling. In this strategy, classical strain-improvement method is used to create a population (pool) of microbes with small functional improvements that are then shuffled by recursive pool-wise protoplast fusion.^[32] For example, it was reported that the acid tolerance of *Lactobacillus* was improved by genome shuffling.^[33] The lower pH is desirable because, at low pH, the fermentation product of *Lactobacillus*, lactic acid, is mostly in the free-acid form and can be purified much more easily than its lactate form (which is the predominant form at higher pH). The initial genetic diversity was first created in a fermenter by slowly decreasing the fermentation pH from 6.0 to 4.1 over a period of 1200 hr, resulting in a stable population of *Lactobacilli* variants growing at pH 4.1, a pH that severely inhibits growth of wild type strain. Then genome shuffling was carried out within the selected population by means of pool-wise recursive protoplast fusion, and finally a population of microbes that can grow at pH 3.8 was obtained. The yield of lactic acid of these improved microbes was threefold higher than that of the wild type microbes at pH 4.0.

Identifying the genetic basis for the desirable phenotype remains one of the central challenges in IME. Traditional approaches include DNA sequencing of the inserts after identifying desirable clones from the plasmid-based library and evaluation of gene disruption libraries to identify the genes essential for a desired phenotype.^[5] However, these approaches are quite time-consuming and expensive, because it is not easy to ensure that all the relevant genes have been analyzed or the same insert will not be repeatedly sequenced. Thus, new approaches such as DNA microarray combined with molecular bar codes^[34] and insertional mutagenesis^[35] were developed to address these limitations. For both approaches, a pool of cell variants was generated, with its specific gene interrupted by DNA fragments (molecular bar codes or insertional elements) that can facilitate later identification and quantification of each interrupted gene by DNA microarray. Cell variants, growing in different selective conditions competitively, were combined. Under a selective pressure, the genes involved in the biosynthesis of nutrients not provided in the media would be significantly enriched,^[35] while the sequence tags of the interrupted genes would diminish in the culture if the deleted genes were important for the growth.^[34] Compared with the wild type or unselected library profile, the relevant genes for a specific phenotype can be identified and quantified through DNA microarray.

Nucleic Acid Engineering

Ribozymes

Ribozymes are RNA molecules with catalytic activities. These molecules were discovered by Cech et al. in 1981,^[36] and have been found to catalyze a variety of reactions in the cell such as RNA splicing, RNA processing, the replication of RNA genomes, and peptide bond formation during translation.^[37] The main function of naturally occurring ribozymes is the sequence-specific cleavage of RNA molecules. Ribozymes can function either in the *cis* manner, in which they catalyze the splicing of their own RNA sequence, or in the *trans* manner, in which they catalyze the cleavage of other RNA molecules.

One of the most studied *trans*-acting ribozymes is the hammerhead ribozyme. It consists of three base-paired helices surrounding a “core” sequence (Fig. 5). Stems I and III bind to complementary sequences on the target RNA, and the central region catalyzes the cleavage of the RNA at the 3' end of a UH sequence (where H = U, A, or C). Thus the ribozyme can theoretically be designed to target any RNA molecule containing the UH sequence. Hammerhead ribozymes possess great therapeutic potential and have been targeted at numerous genes, ranging from viral disease genes, such as hepatitis B and HIV-1, to cancer related genes, such as multidrug resistance (MDR-1).^[37] The hammerhead ribozyme can be improved in a variety of ways such as *in vitro* selection to isolate ribozymes with higher activity and chemical modification of the nucleotide bases to improve nuclease resistance. These engineering methods have been covered in the review

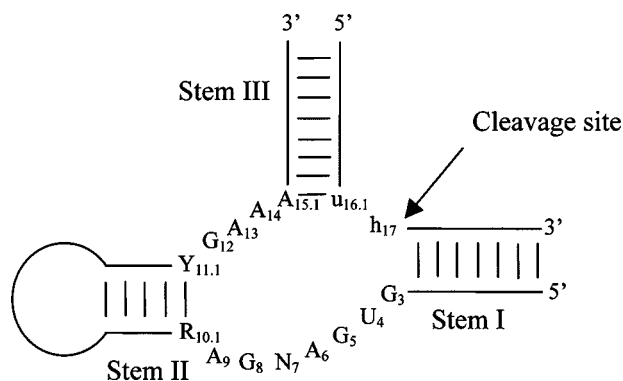


Fig. 5 The hammerhead ribozyme, showing the three non-conserved stems and the cleavage site. Ribozyme nucleotides are in uppercase letters, while substrate nucleotides are in lowercase. Y represents C or U; R represents A or G; N represents A, C, U, or G; h represents A, C, or U. (From Ref.^[48].)

by Usman et al.^[38] In a recent study, a hammerhead ribozyme (RzA) that cleaves the hepatitis B virus (HBV) poly A signal was connected to the viral encoded RNA involved in the packaging of bacterial virus Φ 23 DNA into procapsids (pRNA).^[39] The chimeric ribozyme (pRNA-RzA) was able to cleave the substrate HBV-polyA with nearly 100% efficiency. Furthermore, *e*-antigen assays and northern blot analysis showed that the chimeric ribozyme displayed better performance in inhibiting the HBV-polyA than the ribozyme alone in cell cultures. These findings suggest that pRNA can chaperone and escort the hammerhead ribozyme to function in the cell, enhancing the cleavage efficiency and inhibition effect of the ribozyme on HBV.

Another type of ribozyme is the Group I intron, which carries out a self-splicing reaction via a two-step *trans*-esterification reaction. As such, *trans*-splicing Group I ribozymes can be used as a form of treatment for genetic disorders as they can repair RNA transcripts from mutated genes. The Group I ribozyme recognizes the target mRNA by base pairing to an accessible region of the transcript upstream of a nonsense or mis-sense mutation via an internal guide sequence (IGS). After base pairing, the ribozyme cleaves off the downstream region and splices in a corrected transcript to restore the correct genetic information (Fig. 6).^[40] The Group I ribozyme has been adopted to mediate the repair transcripts of mutated p53 gene, a tumor suppressor gene that is mutated in many cancers.^[41] Using *in vitro* selection, two ribozymes, Rib41 and Rib65, which were able to cleave the majority of p53 transcripts and yield products of the correct size, were isolated from a ribozyme library containing randomized IGS, 5'-GNNNNN-3' (where N represents any of the four nucleotides). These ribozymes were able to repair the defective p53 RNA transcripts with high fidelity and specificity. In addition, the corrected transcripts were found to be functionally translated, resulting in a 23-fold induction of a p53 responsive promoter and a threefold reduction in the MDR-1 gene promoter.

DNA enzymes

DNA has long been regarded as a passive molecule that is ideal for carrying genetic information but is structurally monotonous and thus functionally impoverished. However, this notion was changed with the discovery of catalytic RNA.^[42] Since then, DNA enzymes, which are cation-dependent enzymatic molecules composed entirely of DNA, have been developed.^[43,44]

In the seminal work by Breaker and Joyce,^[43] a mixture of random N₅₀ DNA oligomers were tethered to a matrix by a RNA nucleotide-containing linker.

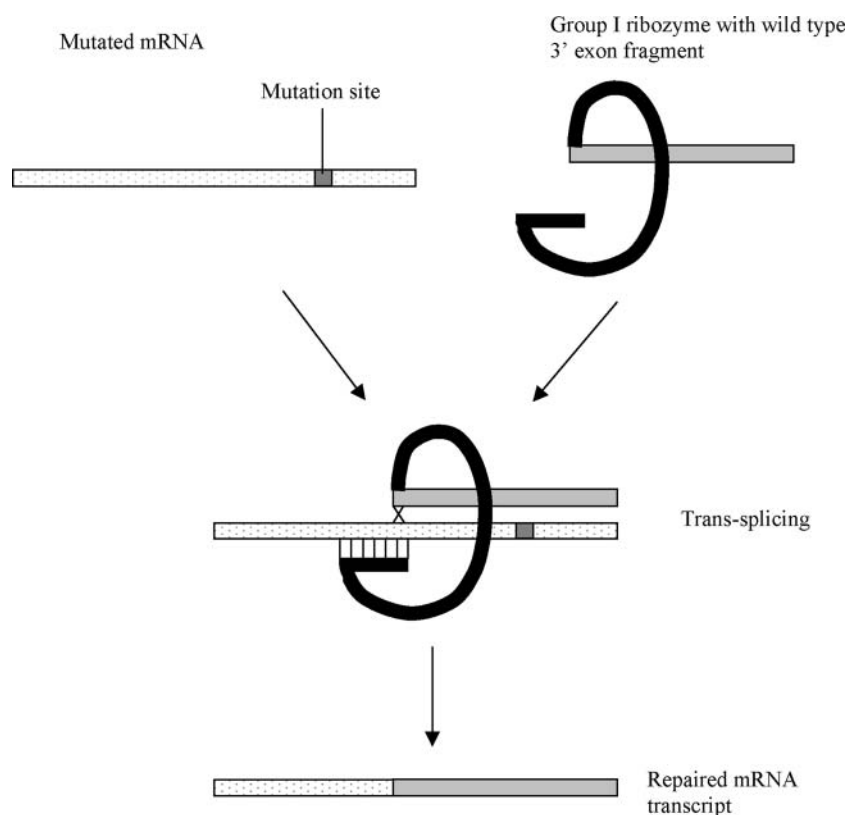


Fig. 6 The scheme of mRNA repair by Group I ribozyme.

When washed with lead solution, oligomers capable of utilizing lead ions to cleave the tethering RNA were eluted together with the solution. Hence the DNA enzymes were able to “select” themselves. It was also found that these enzymes can be engineered to cleave RNA substrates specifically and efficiently. DNA enzymes with 10–23 motifs or 8–17 motifs are prominent examples of such biomolecules. Both the DNA enzymes consist of a nucleotide catalytic core flanked by two substrate-binding arms of seven to eight bases each.^[44] By changing the substrate-binding domain sequence, the DNA enzyme can be designed to target the AU nucleotides of different mRNA substrates with high specificity.

One potential target for DNA enzyme therapy is angiogenesis of solid tumors. Although angiogenesis is an important step in tumor development, current understanding of key transcription factors regulating it is limited.^[45] However, it is known that one gene involved in angiogenesis is the early growth response (Egr-1) gene. A 10–23 motif DNA enzyme designed to target a specific motif in the 5'-untranslated region of Egr-1 mRNA was able to inhibit Egr-1 protein expression, microvascular endothelial cell replication and migration, and microtubule network formation on basement membrane matrices.^[45] In addition, the Egr-1 DNA enzymes suppressed tumor growth by

fivefold without influencing body weight, wound healing, blood coagulation, or other hematological parameters of mice. This study not only shed light on the understanding of key transcription factors regulating angiogenesis but also showed the huge potential of DNA enzyme as therapeutic drug in cancer treatments.

Aptamers

Aptamers are ligands of double-stranded DNA or single-stranded RNA that can be generated against amino acids, drugs, proteins, and other molecules. They have potential applications in analytical devices, such as biosensors, and as therapeutic agents. When combined with in vitro selection techniques,^[8,46] aptamers can become powerful screening tools for rapidly identifying targets for new pharmaceuticals.

A spiegelmer, NOX 1255, with a high affinity ($K_D = 20$ nM) for the peptide hormone, gonadotropin-releasing hormone (GnRH), was isolated by using the mirror image selection method.^[47] The stability of this molecule was further increased by adding a 40 kDa polyethylene glycol moiety onto its 5' end (NOX 1257). The spiegelmer was also found to be highly specific as it did not show any inhibition by buserelin, a peptide analog of GnRH. Furthermore, the anti-GnRH spiegelmer

exhibited a very low immunogenic potential in in vivo experiments with rabbits, suggesting that spiegelmers are not recognized by the immune system and can be administered repeatedly over long periods without adverse effects. This study demonstrates the potential of spiegelmers as a new pharmaceutical approach against GnRH and other targets.

Selected spiegelmers must be synthesized chemically, as no enzyme is able to synthesize L-DNA or L-RNA. However, with the current technology the efficient chemical synthesis of oligoribonucleotides is limited to 60 nt. Typical spiegelmers identified through the SELEX method are 60–90 nt long consisting of a 30–40 nt long randomized region and fixed primer sites of about 15–25 nt on each end. Truncation of these flanking primer regions from selected spiegelmers is required before further testing in biological systems. However, the flanking primers may affect the binding ability of the spiegelmer to its target and spiegelmers may not work well if they are truncated. To reduce this uncertainty, Vater et al. developed a strategy known as Tailored-SELEX to minimize the number of fixed primer regions in a spiegelmer.^[9] Using this strategy, a rat α -CGRP binding spiegelmer with a calculated affinity of $K_D = 2.5$ nM at 37°C, which is close to the binding constant of the neuropeptide to its receptor ($EC_{50} = 1$ nM), was isolated.

CONCLUSIONS

With the advances in high-throughput screening method for discovering target biomolecules and the accumulation of data in functional genomics and proteomics, the rate of designing and discovering valuable biomolecules will rapidly grow. Many enzymes and proteins have been engineered for process improvement and generation of high-value products at a low cost. A few biomolecular techniques have been utilized in disease diagnosis, and some engineered biomolecules have entered clinical trials for therapeutic uses. When the enabling technologies such as protein engineering, metabolic engineering, microarray, mass spectrometry, and bioinformatics grow more mature, and when our understanding of cells becomes more thorough, biomolecular engineering will no doubt become one of the most important research areas in both academia and industry.

ACKNOWLEDGMENT

We thank the Office of Naval Research and National Science Foundation for supporting our work in biomolecular engineering.

ARTICLES OF FURTHER INTEREST

Biocatalysis, p. 101.
Protein Design, p. 2467.

REFERENCES

- Ryu, D.D.; Nam, D.H. Recent progress in biomolecular engineering. *Biotechnol. Prog.* **2000**, *16* (1), 2–16.
- Stephanopoulos, G. Metabolic fluxes and metabolic engineering. *Metab. Eng.* **1999**, *1* (1), 1–11.
- Nielsen, J. Metabolic engineering. *Appl. Microbiol. Biotechnol.* **2001**, *55* (3), 263–283.
- Bailey, J.E.; Sburlati, A.; Hatzimanikatis, V.; Lee, K.; Renner, W.A.; Tsai, P.S. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol. Bioeng.* **2002**, *79* (5), 568–579.
- Gill, R.T. Enabling inverse metabolic engineering through genomics. *Curr. Opin. Biotechnol.* **2003**, *14* (5), 484–490.
- Stephanopoulos, G.N.; Aristidou, A.A.; Nielsen, J. *Metabolic Engineering: Principles and Methodologies*; Academic Press: London, U.K., 1998.
- Wiechert, W. Modeling and simulation: tools for metabolic engineering. *J. Biotechnol.* **2002**, *94* (1), 37–63.
- Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment—RNA ligands to bacteriophage-T4 DNA-polymerase. *Science* **1990**, *249* (4968), 505–510.
- Vater, A.; Jarosch, F.; Buchner, K.; Klussmann, S. Short bioactive spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: Tailored-SELEX. *Nucl. Acids Res.* **2003**, *31* (21), e130.
- Eulberg, D.; Klussmann, S. Spiegelmers: Biostable aptamers. *Chem. Biol. Chem.* **2003**, *4* (10), 979–983.
- Dolinnaya, N.G.; Sokolova, N.I.; Gryaznova, O.I.; Shabarova, Z.A. Site-directed modification of DNA duplexes by chemical ligation. *Nucl. Acids Res.* **1988**, *16* (9), 3721–3738.
- Dolinnaya, N.G.; Sokolova, N.I.; Ashirbekova, D.T.; Shabarova, Z.A. The use of Brcn for assembling modified DNA duplexes and DNA–RNA hybrids—Comparison with water-soluble carbodiimide. *Nucl. Acids Res.* **1991**, *19* (11), 3067–3072.
- Smith, J.D.; Gold, L. Transcription-Free SELEX US Patent 6,387,620, 2002.
- Agrawal, S.; Goodchild, J.; Civeira, M.P.; Thornton, A.H.; Sarin, P.S.; Zamecnik, P.C. Oligodeoxynucleoside phosphoramidates and

- phosphorothioates as inhibitors of human immunodeficiency virus. Proc. Natl. Acad. Sci. U.S.A. **1988**, *85* (19), 7079–7083.
15. Matsukura, M.; Shinozuka, K.; Zon, G.; Mitsuya, H.; Reitz, M.; Cohen, J.S.; Broder, S. Phosphorothioate analogs of oligodeoxynucleotides: Inhibitors of replication and cytopathic effects of human immunodeficiency virus. Proc. Natl. Acad. Sci. U.S.A. **1987**, *84* (21), 7706–7710.
 16. Corey, D.R. Telomerase inhibition, oligonucleotides, and clinical trials. Oncogene **2002**, *21* (4), 631–637.
 17. Urban, E.; Noe, C.R. Structural modifications of antisense oligonucleotides. Farmaco **2003**, *58* (3), 243–258.
 18. Teplova, M.; Wallace, S.T.; Tereshko, V.; Minasov, G.; Symons, A.M.; Cook, P.D.; Manoharan, M.; Egli, M. Structural origins of the exonuclease resistance of a zwitterionic RNA. Proc. Natl. Acad. Sci. U.S.A. **1999**, *96* (25), 14240–14245.
 19. Klusmann, S.; Nolte, A.; Bald, R.; Erdmann, V.A.; Furste, J.P. Mirror-image RNA that binds D-adenosine. Nat. Biotechnol. **1996**, *14* (9), 1112–1115.
 20. Stoesser, G.; Sterk, P.; Tuli, M.A.; Stoehr, P.J.; Cameron, G.N. The EMBL nucleotide sequence database. Nucl. Acids Res. **1997**, *25* (1), 7–14.
 21. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. **1990**, *215* (3), 403–410.
 22. Schoonman, M.J.; Knechtel, R.M.; Grootenhuis, P.D. Practical evaluation of comparative modeling and threading methods. Comput. Chem. **1998**, *22* (5), 369–375.
 23. Chotani, G.; Dodge, T.; Hsu, A.; Kumar, M.; LaDuca, R.; Trimbura, D.; Weyler, W.; Sanford, K. The commercial production of chemicals using pathway engineering. Biochim. Biophys. Acta **2000**, *1543* (2), 434–455.
 24. Yang, Y.; Bennett, G.N.; San, K.Y. Genetic and metabolic engineering. J. Biotechnol. **1998**, *1* (3), 134–141.
 25. Chou, C.H.; Bennett, G.N.; San, K.Y. Effect of modulated glucose uptake on high-level recombinant protein production in a dense *Escherichia coli* culture. Biotechnol. Prog. **1994**, *10* (6), 644–647.
 26. Aristidou, A.A.; San, K.Y.; Bennett, G.N. Modification of central metabolic pathway in *Escherichia coli* to reduce acetate accumulation by heterologous expression of the *Bacillus subtilis* acetolactate synthase gene. Biotechnol. Bioeng. **1994**, *44* (8), 944–951.
 27. Yang, Y.T.; Aristidou, A.A.; San, K.Y.; Bennett, G.N. Metabolic flux analysis of *Escherichia coli* deficient in the acetate production pathway and expressing the *Bacillus subtilis* acetolactate synthase. Metab. Eng. **1999**, *1* (1), 26–34.
 28. Szczebara, F.M.; Chandelier, C.; Villeret, C.; Masurel, A.; Bourot, S.; Duport, C.; Blanchard, S.; Groisillier, A.; Testet, E.; Costaglioli, P.; Cauet, G.; Degryse, E.; Balbuena, D.; Winter, J.; Achstetter, T.; Spagnoli, R.; Pompon, D.; Dumas, B. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. Nat. Biotechnol. **2003**, *21* (2), 143–149.
 29. Assinder, S.J.; Williams, P.A. The TOL plasmids: determinants of the catabolism of toluene and the xylenes. Adv. Microb. Physiol. **1990**, *31*, 1–69.
 30. Lee, J.Y.; Jung, K.H.; Kim, H.S. Amplification of toluene dioxygenase genes in a hybrid pseudomonas strain to enhance the biodegradation of benzene, toluene, and *p*-xylene mixture. Biotechnol. Bioeng. **1995**, *45* (6), 488–494.
 31. Selifonova, O.; Valle, F.; Schellenberger, V. Rapid evolution of novel traits in microorganisms. Appl. Environ. Microbiol. **2001**, *67* (8), 3645–3649.
 32. Zhang, Y.X.; Perry, K.; Vinci, V.A.; Powell, K.; Stemmer, W.P.; del Cardayre, S.B. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature **2002**, *415* (6872), 644–646.
 33. Patnaik, R.; Louie, S.; Gavrilovic, V.; Perry, K.; Stemmer, W.P.; Ryan, C.M.; del Cardayre, S. Genome shuffling of *Lactobacillus* for improved acid tolerance. Nat. Biotechnol. **2002**, *20* (7), 707–712.
 34. Giaever, G.; Chu, A.M.; Ni, L.; Connelly, C.; Riles, L.; Veronneau, S.; Dow, S.; Lucau-Danila, A.; Anderson, K.; Andre, B.; Arkin, A.P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachat, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deutschbauer, A.; Entian, K.D.; Flaherty, P.; Foury, F.; Garfinkel, D.J.; Gerstein, M.; Gotte, D.; Guldener, U.; Hegemann, J.H.; Hempel, S.; Herman, Z.; Jaramillo, D.F.; Kelly, D.E.; Kelly, S.L.; Kotter, P.; LaBonte, D.; Lamb, D.C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S.L.; Revuelta, J.L.; Roberts, C.J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D.D.; Sookhai-Mahadeo, S.; Storms, R.K.; Strathern, J.N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C.Y.; Ward, T.R.; Wilhelmy, J.; Winzeler, E.A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J.D.; Snyder, M.; Philippsen, P.; Davis, R.W.; Johnston, M. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **2002**, *418* (6896), 387–391.
 35. Badarinarayana, V.; Estep, P.W., III; Shendure, J.; Edwards, J.; Tavazoie, S.; Lam, F.; Church, G.M.

- Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat. Biotechnol.* **2001**, *19* (11), 1060–1065.
36. Cech, T.R.; Zaug, A.J.; Grabowski, P.J. In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981**, *27* (3 Pt 2), 487–496.
 37. Puerta-Fernandez, E.; Romero-Lopez, C.; Barroso-delJesus, A.; Berzal-Herranz, A. Ribozymes: Recent advances in the development of RNA tools. *FEMS Microbiol. Rev.* **2003**, *27* (1), 75–97.
 38. Usman, N.; Beigelman, L.; McSwiggen, J.A. Hammerhead ribozyme engineering. *Curr. Opin. Struct. Biol.* **1996**, *6* (4), 527–533.
 39. Hoepflich, S.; Zhou, Q.; Guo, S.; Shu, D.; Qi, G.; Wang, Y.; Guo, P. Bacterial virus phi29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus. *Gene Therapy* **2003**, *10* (15), 1258–1267.
 40. Sullenger, B.A.; Cech, T.R. Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* **1994**, *371* (6498), 619–622.
 41. Watanabe, T.; Sullenger, B.A. Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (15), 8490–8494.
 42. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA—Auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of *Tetrahymena*. *Cell* **1982**, *31* (1), 147–157.
 43. Breaker, R.R.; Joyce, G.F. A DNA enzyme with Mg^{2+} -dependent RNA phosphoesterase activity. *Chem. Biol.* **1995**, *2* (10), 655–660.
 44. Santoro, S.W.; Joyce, G.F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (9), 4262–4266.
 45. Fahmy, R.G.; Dass, C.R.; Sun, L.Q.; Chesterman, C.N.; Khachigian, L.M. Transcription factor Egr-1 supports FGF-dependent angiogenesis during neovascularization and tumor growth. *Nat. Med.* **2003**, *9* (8), 1026–1032.
 46. Ellington, A.D.; Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346* (6287), 818–822.
 47. Wlotzka, B.; Leva, S.; Eschgfäller, B.; Burmeister, J.; Kleinjung, F.; Kaduk, C.; Muhn, P.; Hess-Stumpff, H.; Klussmann, S. In vivo properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (13), 8898–8902.
 48. Hertel, K.J.; Pardi, A.; Uhlenbeck, O.C.; Koizumi, M.; Ohtsuka, E.; Uesugi, S.; Cedergren, R.; Eckstein, F.; Gerlach, W.L.; Hodgson, R. Numbering system for the hammerhead. *Nucl. Acids Res.* **1992**, *20* (12), 3252.