Biology as a Basis for Biochemical Engineering

Sujata K. Bhatia, M.D., P.E.
Harvard Univ.

Biochemical engineering demonstrates the versatility of the living cell as a miniature reactor. Living cells can be genetically engineered to produce specific proteins, tuned to manufacture novel enzymes, and optimized to guarantee genetic stability and consistent production of desired products.

Biochemical engineering harnesses living cells as miniature chemical reactors, enabling the production of designer molecules ranging from pharmaceuticals to plastics to biofuels. Discoveries in molecular and cellular biology have driven the development of biochemical engineering techniques, and the subsequent rise of the modern biotechnology industry.

Live cells possess unique capabilities to manufacture complex chemical entities. Even a relatively simple bacterial cell performs highly complex synthesis, fueling, and polymerization reactions through native pathways. Such reactions — occurring readily in organisms — would be very difficult to accomplish via traditional synthetic chemistry or inorganic catalysts.

Yet living cells also introduce unique challenges and tradeoffs for manufacturing and process engineering. The demands of large-scale production can place unnatural stresses on cells, affecting both cellular growth and the quality and reliability of cellular output.

The biochemical engineer must successfully balance the capacities and limitations of these cellular reactors. Designing and optimizing cellular manufacturing processes, therefore, requires an understanding of the biological underpinnings of biochemical engineering.

This article describes the fundamental principles of molecular and cellular biology, and explains the flow of genetic information within biological systems. It then relates core biological phenomena to genetic engineering and metabolic engineering. Finally, it discusses important engineering considerations in the design of biochemical manufacturing processes.

Major models of molecular and cellular biology

A living cell is an elegant reactor, containing the instructions for chemical synthesis and breakdown reactions, process regulation and control, and quality assurance, all within the core nucleus of the cell. The basic instructions governing all processes within the cell are encoded in deoxyribonucleic acid (DNA), the genetic material of the cell. DNA is a long polymer composed of repeating monomer subunits called nucleotides. The nucleotide subunits are adenine (A), cytosine (C), guanine (G), and thymine (T), all arranged in a specific sequence that codes for all structure and function within the cell. The living cell operates based on sequences of A, C, G, and T, much the same way that computers operate based on sequences of 0s and 1s. The structure of DNA was first presented by Watson and Crick in a now classic work of molecular biology *(1)*; it is difficult to overestimate the impact of this work on biochemical engineering and industrial biotechnology.

In living organisms, DNA typically does not exist as a single strand, but rather as two long polymer strands twisted around one another to form a double-helix structure. The two strands are held together by hydrogen bonds, and the two strands are completely complementary to one another. Adenine (A) preferentially binds to thymine (T), and cytosine (C) preferentially binds to guanine (G). Thus the structure and nucleotide sequence of a single DNA strand completely defines the structure and nucleotide sequence of its complementary DNA strand *(2)* (Figure 1).

Every cell within an organism contains the same DNA double-helix structure and sequence. Accurate communication of the information encoded within DNA is absolutely

essential for the organism to survive. Cells have evolved an ingenious mechanism for ensuring consistent and error-free transmission of DNA-encoded information.

Transcription. The first step of the information transmission process is known as transcription. During transcription, the double helix unzips into two single DNA strands. Nucleotides line up along one of the strands to form a new single-stranded polymeric molecule called ribonucleic acid (RNA). RNA is composed of a sequence of the nucleotide subunits adenine (A), cytosine (C), guanine (G), and uracil (U) — uracil in RNA fulfills an analogous role to that of thymine in DNA. Because the sequence of nucleotides in a single DNA strand completely defines the sequence of nucleotides in its complementary DNA strand, it is also true that the nucleotide sequence of a single DNA strand defines the sequence of nucleotides in its complementary RNA strand (Figure 2).

Translation. Just as DNA codes for RNA during transcription, RNA in turn codes for the synthesis of proteins, as well as the structural and functional molecules of the cell and the entire organism. RNA is an intermediary between the DNA genome of the cellular nucleus, and the synthesis of proteins in the cytoplasm outside the cellular nucleus. While DNA and RNA are macromolecules composed of nucleotide subunits, proteins are macromolecules composed of amino acid subunits. Each three-nucleotide sequence of RNA (for instance, ATC, CUG, GGG, etc.) is known as a codon and codes for a specific amino acid *(3)*. The process of protein synthesis based on the RNA code is called translation.

During translation (which takes place inside the ribosome), amino acids line up along the single-stranded RNA with corresponding codon triplets, and the amino acids bind to one another to form a protein. The sequence of the RNA strand completely defines the sequence of amino acids in the protein (Figure 3). Just as DNA completely defines RNA, RNA completely defines a protein. A protein's composition and sequence can be predicted from its corresponding DNA sequence, and the reverse is also true, so that a DNA sequence can be derived from a known protein sequence.

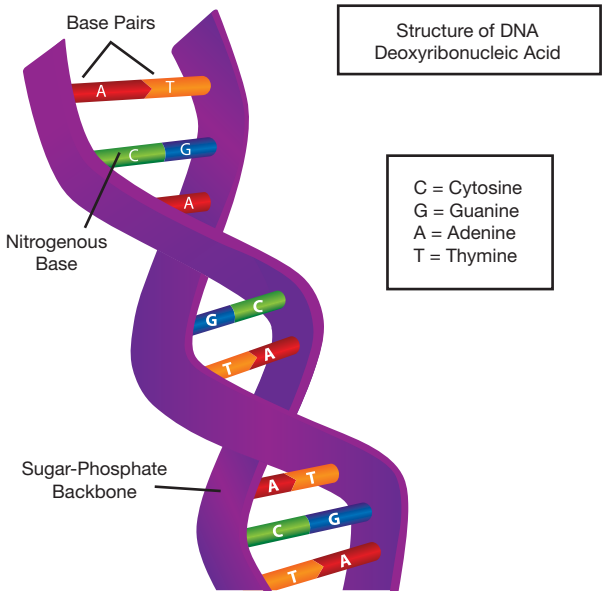
The main sources of information in the cell — DNA and RNA — as well as the main sources of structure and function in the cell — proteins — are all polymers.

Genetic engineering: Cells as reactors

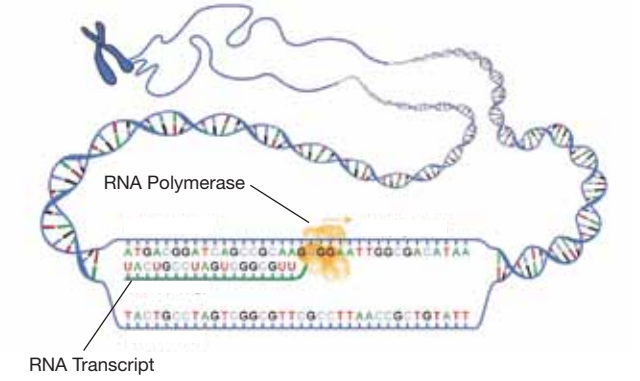
Genetic engineering, the foundation of the modern biotechnology industry, is based on the following key insight: Since a DNA sequence completely defines an RNA sequence through transcription, and an RNA sequence completely defines a protein sequence through translation, then any desired protein can be manufactured by cellular machinery, provided that the cell possesses the DNA template corresponding to the protein.

Cells containing the DNA template for a desired protein can be cultured in large fermentation batches and the protein can be produced at a large scale. For instance, suppose that you want to produce human insulin protein at a large scale for therapeutic purposes. If the DNA coding for human insulin can be inserted into a robust cellular host such as a yeast cell, and yeast cells subsequently grown in large fermentation batches (much the same way as beer is brewed), then large amounts of human insulin can be readily produced. The biggest challenge of this process is inserting the human insulin gene into the yeast cellular host.

Article continues on next page



▲ **Figure 1.** In a DNA strand, adenine (A) preferentially binds to thymine (T), and cytosine (C) preferentially binds to guanine (G).



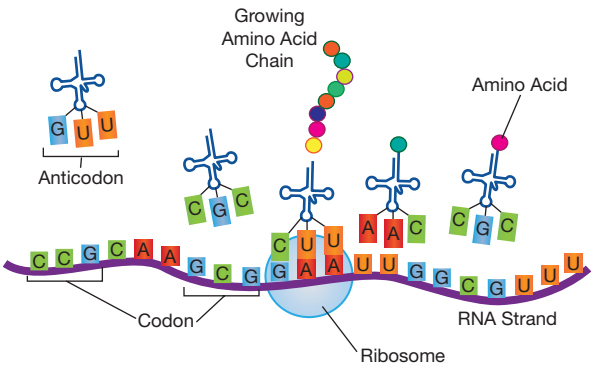
▲ **Figure 2.** During transcription, the DNA double-helix unwinds. Nucleotides line up along one of the strands to form an RNA transcript. The nucleotide sequence of a single DNA strand completely defines the sequence of nucleotides in its complementary RNA strand. Source: National Human Genome Research Institute.

Cohen and Boyer devised such a method for reliably introducing any given DNA sequence into a chosen cellular host *(4)*. The method relies on catalysts known as restriction enzymes or restriction endonucleases, which precisely cut DNA sequences at specific sites based on the nucleotide sequence. Different restriction enzymes correspond to different sequences of DNA. For example, the EcoRI (eco-R-one) enzyme cuts DNA only at the sequence GAATTC.

If two different pieces of DNA are cut with the same restriction enzyme, then the ends of these two pieces of DNA will be complementary to one another. The two pieces of DNA can join together in a ligation reaction catalyzed by DNA ligase enzyme to form a recombinant DNA molecule.

Figure 4 illustrates a common technique for creating recombinant DNA. A desired DNA fragment is inserted into a plasmid vector (*i.e.,* a circular piece of DNA that replicates in the cellular cytoplasm, independently of the chromosomal DNA in the cell nucleus). The recombinant DNA molecule can then be inserted into a host cell, such as a bacterium, yeast, or mammalian cell, through a process known as transformation. A typical method for cellular transformation is to apply an electrical field, which causes the cell membrane to become more porous to allow the recombinant DNA vector to enter the cell.

So, to make human insulin, start by choosing an appropriate restriction enzyme to cut both a plasmid DNA vector and the ends of a DNA fragment coding for human insulin. Then join together these two pieces of DNA using DNA ligase, and insert the resulting recombinant DNA plasmid into a host cell. If the host cells replicate in a fermentation bioreactor, human insulin will be produced at a large scale. In fact, in 1982, human insulin was the first genetically engineered human therapeutic to gain U.S. Food and Drug Administration (FDA) approval — ushering in the age of biotechnology.

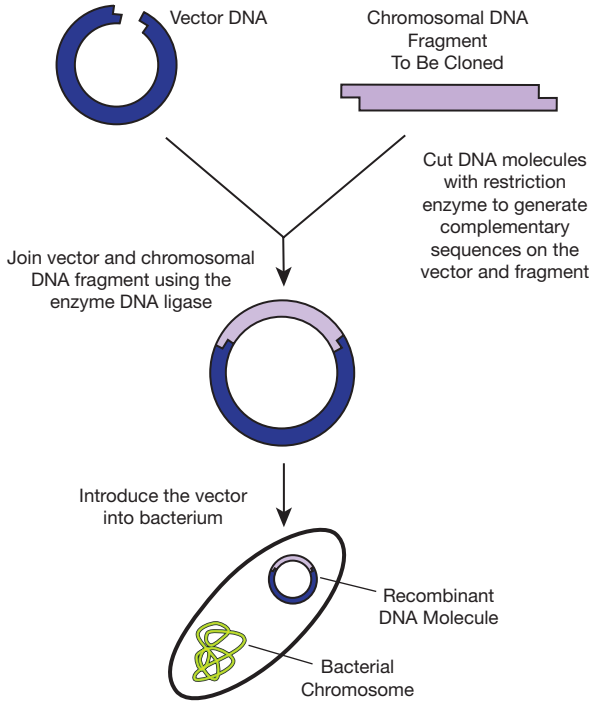


▲ **Figure 3.** During translation, amino acids line up along the single-stranded RNA with corresponding codon triplets, and the amino acids bind to one another to form a protein. The sequence of the RNA strand completely defines the sequence of amino acids in the protein. Genetic engineering leverages this relationship to manufacture desired proteins in live cells. Metabolic engineering further leverages the fact that all enzymes are proteins.

Metabolic engineering: Cells as highly tunable reactors

Biochemical engineering views the cell as the vehicle for producing a desired compound, and its subdiscipline of metabolic engineering aims to make the cell more productive and more versatile. In essence, metabolic engineering modifies and exploits the cell's innate pathways to manufacture useful compounds. Metabolic engineering is a powerful technique that takes advantage of the fact that virtually all enzymes, the natural catalysts of biochemical reactions, are proteins.

While genetic engineering leverages the catalytic machinery of the cell to manufacture proteins, metabolic engineering goes one step further, enabling the cell to manufacture enzymes to tune the cell's own catalytic machinery. Furthermore, metabolic engineering empowers the cell to carry out a wider variety of chemical reactions. Whereas genetic engineering enables the cell to make proteins,



▲ **Figure 4.** During the production of recombinant DNA, both a DNA vector (such as a plasmid) and a desired DNA fragment are cut with the same restriction enzyme, creating complementary sticky ends. The DNA vector and the DNA fragment are joined using DNA ligase enzyme. The recombinant DNA sequence can then be introduced into a host cell such as a bacterium. Source: U.S. Dept. of Energy Genome Program.

metabolic engineering enables the cell to make vitamins *(5)*, antibiotics, chemotherapeutics *(6)*, other small-molecule drugs, industrial polymers, dyes, fuels *(7)*, and other specialty chemicals. All that is required is a host cell with the necessary starting materials and the necessary DNA genetic material coding for the pathway enzymes.

For example, suppose that you want to manufacture indigo dye from the starting material tryptophan, a natural amino acid. The enzymes tryptophanase and naphthalene dioxygenase catalyze the chemical reaction pathway from tryptophan to indigo. You can first introduce a recombinant plasmid DNA vector coding for the tryptophanase and naphthalene dioxygenase enzymes into a bacterial host cell using the methods of genetic engineering described in the previous section. Then grow the transformed bacterial cells in a fermentation bioreactor and supply the cells with tryptophan as a starting material. Indigo dye will be manufactured in the bioreactor.

Metabolic engineering turns the cell into a powerhouse chemical reactor with tunable control systems. If a reaction is deemed productive, then the DNA that encodes its enzyme catalyst can be introduced (*i.e.,* up-regulated) within the cell. Conversely, if a reaction is deemed counter-productive, the DNA that encodes its enzyme catalyst can be deleted, interrupted, or down-regulated. The intracellular concentration of any targeted enzyme can be increased or decreased by changing either the number of copies of DNA encoding the enzyme, or the rate at which the DNA is transcribed and translated into the enzyme. The rates of transcription and translation can be modified by altering the balance of molecular activators and repressors in the cell; these regulatory molecules are generally proteins.

There is no doubt that the ready production of therapeutic proteins and enzymes via genetic engineering has made a dramatic impact on medicine. Yet metabolic engineering has the potential to impact all industries by enabling the ready production of any chemical species.

Reactor optimization: Complexities of cellular manufacturing

The ability of living cells to adapt to environmental changes in real time makes these cellular reactors highly desirable for manufacturing complex molecules. At the same time, the natural adaptations of cells introduce real challenges for biochemical engineering.

First and foremost, cellular systems have evolved a high degree of redundancy in pathways. Any given intracellular reaction may be catalyzed by more than one enzyme. The reverse is also true: Any given enzyme may catalyze multiple intracellular reactions that share identical reaction mechanisms. This redundancy makes it difficult for the biochemical engineer to target a pathway enzyme in isolation.

In addition, cellular systems have innate mechanisms for maintaining homeostasis via feedback control loops. As a result, an engineer may design an up-regulated pathway within the cell, only to find later that the pathway has been down-regulated by cellular feedback-control mechanisms. Toxicity of reaction intermediates and final products can also limit throughput of metabolically engineered pathways, and must be considered. Detailed insights and mathematical models of metabolic networks are necessary to overcome these challenges; computational simulations and analyses of metabolic fluxes are critical for metabolic engineering.

The interrelated biological phenomena of metabolic flux, metabolic burden, and genetic instability also complicate and frustrate metabolic engineering efforts.

Metabolic flux. Metabolic flux is the rate of turnover of molecules through a metabolic pathway. Engineered metabolic pathways can suffer from flux imbalances, as these pathways lack the regulatory mechanisms present in innate metabolic pathways. Metabolically engineered pathways often exhibit bottlenecks as a result of flux imbalances, leading to the diversion of molecules away from the desired product, as well as the accumulation of toxic intermediates. This is particularly problematic when several different enzymes from different organisms are introduced into one cellular host.

Flux balancing can address rate-limiting steps in production to solve these metabolic flux issues. Flux balancing increases product titers (*i.e.,* concentrations) and decreases the buildup of intermediates.

A major strategy for balancing metabolic flux is to modulate the expression levels of individual enzymes (*i.e.,* modulating the processing of DNA into RNA, and RNA into protein). Another strategy for addressing flux imbalance is to improve the activities of rate-limiting enzymes by directed evolution. An emerging strategy for improving metabolic flux is to control the spatial organization of successive enzymes in a given metabolic pathway; this approach mimics natural cellular mechanisms and places sequential enzymes close to one another within the cellular cytoplasm *(8)*.

Metabolic burden. Problems arise because expression levels of enzymes are often dramatically higher in a metabolically engineered pathway than would be typically found in any natural metabolic pathway. The metabolic burden of an engineered pathway robs the cell of its building blocks, including amino acids and nucleotide monomers; these building blocks are necessary for the cell to maintain function and remain viable. When novel pathways are introduced into a cell, essential building blocks can be diverted into those nonessential pathways as the cell ramps up production of the engineered protein. The overproduction of nonessential proteins triggers stress responses within the cell and slows its growth.

Article continues on next page

Ultimately, metabolic burden can limit product concentrations. Thus, the engineer must consider the tradeoff between cellular growth rate and engineered protein expression level, and determine the optimum protein production level for a sustained process. It is necessary to weigh the cell's basic need for survival and growth against efficient production of the desired product.

Genetic instability. Because the metabolic burden shouldered by a plasmid-bearing producer cell places it at a disadvantage relative to a plasmid-free nonproducer cell, genetic instability can occur in engineered cells. Genetic instabilities may be present if product titers decrease while cell growth rates increase. Mutations in plasmid DNA vectors *(9)* as well as losses of plasmid DNA vectors *(10)* can create genetic instability.

Recognizing the tradeoff between engineered protein expression levels and genetic stability is the first step toward managing genetic instability. Then, the engineer must find the optimal protein production level that guarantees genetic stability and reproducible production of the desired product in the cellular reactor.

The future of biochemical engineering

Biochemical engineering demonstrates the versatility of the living cell as a mini-reactor, and leverages the innate information pathway of the cell: DNA is transcribed to RNA, and RNA is translated into protein. While genetic engineering has established that individual proteins can be produced in cells, and metabolic engineering has now established that individual metabolic pathways can be introduced in cells, an important step in the future of biochemical engineering will be to engineer entirely new cells. This emerging field is known as synthetic biology.

SUJATA K. BHATIA, M.D., P.E., is a physician and bioengineer who serves on the biomedical engineering teaching faculty at Harvard Univ. (Phone: (617) 496-2840; Email: sbhatia@seas.harvard.edu). She is the Assistant Director for Undergraduate Studies in Biomedical Engineering at Harvard, and an Assistant Dean for Harvard Summer School. She is also an Associate of the Harvard Kennedy School of Government, for the Science, Technology, and Globalization Project, as well as a faculty member in the Harvard Kennedy School Executive Education program on Innovation for Economic Development. She received bachelor's degrees in biology, biochemistry, and chemical engineering, and a master's degree in chemical engineering, from the Univ. of Delaware, and she received an M.D. and a PhD in bioengineering, both from the Univ. of Pennsylvania. Prior to joining Harvard, she was a principal investigator at the DuPont Co., where her projects included the development of bioadhesives for wound closure and the development of minimally invasive medical devices. She has written two books, *Biomaterials for Clinical Applications* (a textbook that discusses opportunities for both biomaterials scientists and physicians to alleviate diseases worldwide) and *Engineering Biomaterials for Regenerative Medicine*. She received an award from the Harvard Univ. President's Innovation Fund for Faculty in recognition of her innovative approaches to biomedical engineering education, the John R. Marquand Award for Exceptional Advising and Counseling of Harvard Students, and the Capers and Marion McDonald Award for Excellence in Mentoring and Advising. She is a member of AIChE and is a registered P.E. in the state of Massachusetts.

In 2010, the trailblazing bioengineer J. Craig Venter created the first cell with a synthetic genome, thereby creating synthetic life *(11)*. Venter's team demonstrated that a complete genetic system can be reproduced by chemical synthesis, starting with only the digitized DNA sequence contained in a computer *(11)*. If DNA is viewed as the software of life as Venter has suggested, then biochemical engineers can program and build cells from scratch to carry out multiple complex functions. For instance, engineered cells may be designed for water purification, environmental remediation, food production, and cost-effective pharmaceutical manufacturing. Synthetic cells may even serve as novel sources of energy.

Biochemical engineering of synthetic life will thus enable engineers to address global challenges in healthcare, energy, and sustainability. This evolving area of cellular engineering will require a deep understanding of cellular pathways and control mechanisms, mathematical models of regulatory networks, and optimization of cellular reactors. Chemical engineers, already experts in reactor design and process control, are uniquely poised to contribute to the new wave of synthetic cellular bioengineering. CEP

Literature Cited

1. **Watson, J. D., and F. H. C. Crick,** "A Structure for Deoxyribose Nucleic Acid," *Nature,* **171,** pp. 737–738 (Apr. 25, 1953).
2. **Watson, J. D., and F. H. C. Crick,** "Genetical Implications of the Structure of Deoxyribonucleic Acid," *Nature,* **171,** pp. 964–967 (May 30, 1953).
3. **Crick, F. H. C., *et al.,*** "General Nature of the Genetic Code for Proteins," *Nature,* **192,** pp. 1227–1232 (Dec. 30, 1961).
4. **Cohen, S. N., *et al.,*** "Construction of Biologically Functional Bacterial Plastids *in Vitro*," *Proceedings of the National Academy of Sciences USA,* **70,** pp. 3240–3244 (Nov. 1973).
5. **Ye, V. M., and S. K. Bhatia,** "Pathway Engineering Strategies for Production of Beneficial Carotenoids in Microbial Hosts," *Biotechnology Letters,* **34,** pp. 1405–1414 (Aug. 2012).
6. **Ye, V. M., and S. K. Bhatia,** "Metabolic Engineering for the Production of Clinically Important Molecules: Omega-3 Fatty Acids, Artemisinin, and Taxol," *Biotechnology Journal,* **7,** pp. 20–33 (Jan. 2012).
7. **Keasling, J. D., and H. Chou,** "Metabolic Engineering Delivers Next-Generation Biofuels," *Nature Biotechnology,* **26,** pp. 298–299 (Mar. 2008).
8. **Dueber, J. E., *et al.,*** "Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux," *Nature Biotechnology,* **27,** pp. 753–759 (Aug. 2009).
9. **Kachroo, A. H., *et al.,*** "Metabolic Engineering without Plasmids," *Nature Biotechnology,* **27,** pp. 729–731 (Aug. 2009).
10. **Tyo, K. E. J., *et al.,*** "Stabilized Gene Duplication Enables Long-Term Selection-Free Heterologous Pathway Expression," *Nature Biotechnology,* **27,** pp. 760–765 (Aug. 2009).
11. **Gibson, D. G., *et al.,*** "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome," *Science,* **329,** pp. 52–56 (July 2, 2010).