

# Enabling Life with a Molecular Toolkit of Genes and Proteins Designed *De Novo*

**Michael H. Hecht**

*Department of Chemistry, Princeton University, Princeton, NJ 08540*

The entire collection of genes and proteins in all the living systems comprises a minuscule fraction of sequence space. From the enormous diversity of possible gene and protein sequences, billions of years of evolution have selected only a very small collection of 'molecular parts' that sustain living organisms (only ~4,000 genes in *E. coli* and ~20,000 in humans.) These considerations might lead to an assumption that the sequences that enable life are unusual or somehow special. Is this true? Or can sequences designed from scratch sustain the growth of living cells?

To address these questions, we designed and constructed a collection containing millions of artificial proteins (a model 'proteome') encoded by a library of synthetic genes (an artificial 'genome'). Structural studies show that many (perhaps most) of our novel proteins fold into stable 3-dimensional structures. Next, we used genetic selections to demonstrate that several of these novel proteins provide biochemical functions that are essential for the growth of *E. coli*. Thus, artificial sequences, which had never before existed on earth, possess activities that sustain life.

This initial foray into artificial genomics suggests (i) the molecular toolkit for life need not be limited to genes and proteins that already exist on earth; (ii) the construction of artificial genomes composed of non-natural sequences is within reach; and (iii) it may be possible to devise synthetic organisms that are sustained by *de novo* designed proteins encoded by novel genomes.