ACS central science

FIRST REACTIONS

Mysteries in a Minimal Genome

Maxwell Coyle, Jennifer Hu and Zev Gartner

Pharmaceutical Chemistry, University of California San Francisco, 600 16th Street N512E, San Francisco, California 94143, United States

The history of the field of total synthesis demonstrates that there is value in constructing novel products from simpler building blocks, a value that extends beyond that of the products themselves. "New synthetic methods are often incorporated into the synthetic schemes towards the target, and the exercise of the total synthesis becomes an opportunity for the invention and discovery of new chemistry," writes K. C. Nicolaou and colleagues.[1] The maturation of the field of total synthesis has provided a bounty of useful medicines and materials. But in tandem with the achievement of progressively more elaborate syntheses came paradigm-shifting insight into the nature of chemical bonds and bond-forming reactions.

Many see synthetic biology as poised to follow a similar path as synthetic chemistry.[2] Building living things from simple and well-characterized parts can provide useful technologies such as microbial fuel factories,[3] immune cells engineered to selectively attack cancer,[4] living tissues,[5] and novel materials.[6] However, synthetic biology also promises to yield new and fundamental biological insights. Through their efforts to build new life forms, synthetic biologists will reveal what works, what does not work, and in the process ask, "why?"

The synthesis of a minimal functional genome from simple chemical precursors (rather than copied from an existing biological template) has stood as a practical and symbolic milestone in synthetic biology, since the genome encodes the complete instruction set for life. The first synthetic genome was viral—that of the hepatitis C virus (HCV), encompassing 9600 bases and completed in 2000.[7] In 2008, Craig Venter and colleagues synthesized the first *cellular* genome, that of *Mycoplasma genitalium*, spanning 580,000 bases and encoding 525 genes.[8] In 2010, Venter's team synthesized an even larger cellular genome, that of *Mycoplasma mycoides*, and proved its functionality by transplanting it into the cell body of *Mycoplasma capricolum*, thereby morphing the cell into the *M. mycoides* species as it read and rebooted itself from the new genetic instructions.[9] This genome, JCVI-syn1.0, also contained watermark sequences, modifications to its genetic code that spelled out names and quotations, but were biologically inert. However, this genome, from a functional perspective, was identical to the genome from which it was derived.

In a feat of synthetic biology, Craig Venter and team construct a minimal bacterial genome from scratch.

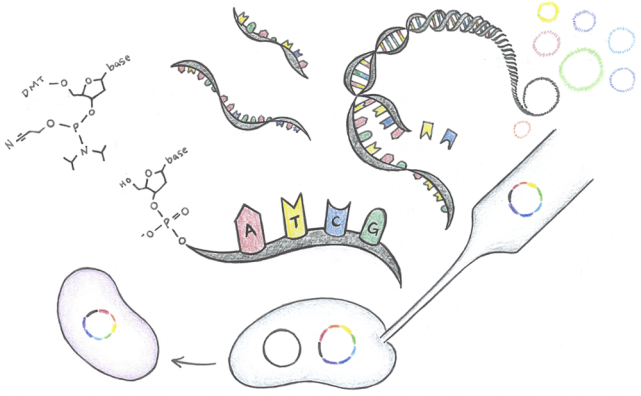

Figure 1. Beginning with solid-phase DNA synthesis from phosphoramidite precursors, and continuing with a series of PCR, in vitro assembly, and cloning steps, Venter and colleagues developed a workflow for synthesizing synthetic cellular genomes in ∼3 weeks. Breaking the genome into one-eighth pieces allowed them to minimize each segment independently and test it against a seven-eights wild-type background. Artwork credit: Jennifer Hu.

In an important recent milestone for the field of synthetic biology, Venter and colleagues completed the synthesis of another synthetic genome called JCVI-syn3.0, which is JCVI-syn1.0 distilled to a minimal necessary set of components.[10] The goal of identifying a minimal genome emerged as genome sequencing technologies came online in the 1990s. A minimal genome promises to provide a genetic chassis upon which more complex cell behaviors and functions can be constructed, for example, the capacity to produce fuels.

Published: May 2, 2016

DOI: 10.1021/acscentsci.6b00110
ACS Cent. Sci. 2016, 2, 274−277

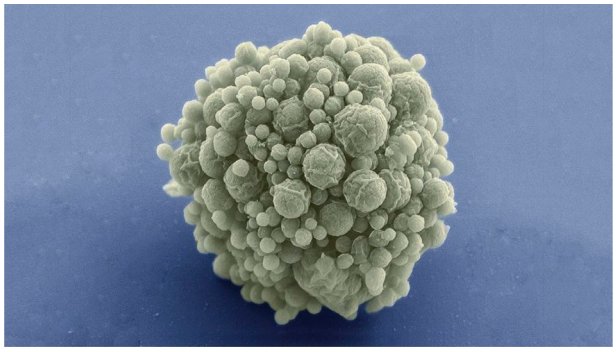

Figure 2. Once the genome of JCVI-syn3.0 was "booted up" in the cytoplasm of wild-type *M. capricolum*, the cells were able to divide themselves, with a doubling time of approximately 3 h. Reproduced with permission from ref 8. Copyright 2016 American Association for the Advancement of Science.

How do you design a minimal genome? While viral genomes are the smallest in nature, and genes are regularly stripped away from them for laboratory purposes (e.g., second and third generation lentivirus from HIV), little work has been done to define the minimal components. This is probably because naturally occurring viruses already have as few as four genes (bacteriophage MS2),[11] and viruses require cellular hosts, and the replication machinery specified by cellular genomes, to propagate. Designing a minimal genome for an autonomously replicating cell is more challenging. Proposals have included both top-down approaches—minimizing the smallest known cellular genome, *M. genitalium* at 580 kb (525 genes)—and bottom-up approaches: asking which enzymes are necessary and sufficient for macromolecule polymerization and genome replication.[12−15]

Top-down approaches have included comparative genomics of *M. genitalium* and *Haemophilus influnezae*, the first two cellular genomes to be sequenced, showing that 256 genes are conserved between them.[12] Another top-down approach (Venter's team) used single gene mutations to identify 387 protein coding genes essential for viability in *M. genitalium*.[13] Finally, whole-cell *in silico* models of *M. genitalium*, based on cumulative knowledge of its molecular biology (900+ publications), suggested 284 genes essential for simulated growth and division.[14]

Each of these methods has fundamental limitations. Comparative genomics can miss divergent genes that accomplish similar essential functions. Deletion data ignore synthetic lethality, where deleting one of a pair of genes may have no adverse affect, while deleting both is lethal. Finally, computer modeling is limited to known molecular biology, and *M. genitalium* is a relatively poorly studied organism with dozens of genes of unknown function.

In contrast to these top-down approaches, bottom-up approaches have also been theorized, most notably by Anthony Forster and George Church. Using cumulative biochemical data from *in vitro* polymerization of DNA, RNA, and polypeptides (mostly with recombinant *E. coli* proteins), they propose a minimal genome of 166 genes.[15] While promising theoretically, this approach has yet to be validated.

JCVI-syn3.0, enabled by advances in genome synthesis, is the product of a herculean empirical effort. It represents the first approximation of a minimal cellular genome that has also been shown experimentally to be functional.

JCVI-syn3.0, enabled by advances in genome synthesis, is the product of a herculean empirical effort.

JCVI-syn3.0 has a genome of 473 genes, fewer than the genome of *M. genitalium* (525), and when transplanted into *M. capricolum*, provides a complete set of instructions to maintain cellular viability and propagation. Rather than minimizing the *M. genitalium* genome, Venter's team, led by Clyde Hutchison III and Ray-Yuan Chuang, began with the 901-gene *M. mycoides* genome, a faster growing species that allowed for more rapid experimentation.

Overall, they took a deletion approach, using genome-wide Tn5 transposon mutagenesis screens to classify genes as essential, nonessential, and quasi-essential (necessary for growth but not viability). However, the authors found that deleting all the "non-essential" genes led to a nonviable genome, largely because of the previously mentioned issues with synthetic lethality; the essentiality of a gene is a function of the genomic background. For example, deleting one gene may turn a second from being nonessential to being essential, or even vice versa.

Venter's team battled this obstacle by breaking the genome into eight pieces and testing the modifications in each segment when combined with a 7/8 wild-type background. Enabling this approach were significant technical advances in their ability to synthesize whole genomes from scratch. They can now synthesize an entire bacterial genome in ∼3 weeks, 2 orders of magnitude faster than their first synthetic bacterial genome in 2008. This speed enhancement meant that rapid cycles of design, assembly, and testing of genomes were possible for the first time, allowing iteration to drive sequential rounds of minimization. At various stages along the optimization route, they repeated the Tn5 transposon screen to see which essential/nonessential/quasi-essential designations were changing as the genetic background for these mutations evolved during the minimization process.

DOI: 10.1021/acscentsci.6b00110
ACS Cent. Sci. 2016, 2, 274−277

Overall, this work demonstrates technical advances to whole-genome synthesis and yields the symbolic victory of a minimal functional cellular genome smaller than anything found in nature. However, the process of synthesizing JCVI-syn3.0 has raised several important questions about our incomplete understanding of what comprises a minimal set of genes. Of the 473 genes in the reduced set, no specific biological function could be ascribed for 149, an astonishing 31.5% of the genome. This could be partly due to *M. mycoides* being a relatively poorly studied model organism, but some of these mysterious essential genes are conserved across other species, including *Homo sapiens*. The authors point to these conserved genes as promising candidates for future study. The most lasting importance of this work may be in its prompting the synthetic biology community to ask what these genes do and why they are essential. Mirroring the history of total synthesis, Venter and colleagues have provided insights about fundamental biological processes, in addition to synthesizing the first minimal cellular genome.

Of the 473 genes in the reduced set, no specific biological function could be ascribed for 149, an astonishing 31.5% of the genome.

Among the annotated genes in the *M. mycoides* genome, the study also provided new insight into which classes of proteins were and were not essential. Seventy-three out of 87 lipoproteins could be deleted, as well as every single one of the 72 genes involved in mobile elements and DNA restriction. Of the genes that were required, 48% played a role in DNA replication and DNA expression, while roughly equal numbers were involved in cytosolic metabolism (17%) and the constitution/organization of the cell membrane (18%). The authors note an evolutionary trade-off between these two categories, where more metabolic enzymes might necessitate fewer membrane transporters and vice versa.

This trade-off speaks to a larger issue that must be addressed when considering a minimal genome. Note that the authors reference *a* minimal genome, and not *the* minimal genome, in their title. Minimal genomes are environment-dependent. For instance, *Mycoplasma* genomes are among the smallest known because these microbes have evolved to become more reliant on animal hosts and have lost adaptability for other environments. A genome is only one component of a living system, and "austere" environments likely require more complicated genetic toolkits.

The minimization of genomes, especially through the iteration of deletion events, may also be pathway-dependent. As the essentially of each gene depends on the surrounding genetic context, this landscape likely contains local minima that may foil many optimization algorithms. Therefore, identifying a global minimum may require a considerably more involved effort.

The most lasting importance of this work may be in its prompting the synthetic biology community to ask what these genes do and why they are essential.

Further work on minimal synthetic genomics ought to help describe how both the complexities of the environment and genetic context relate to the essentiality of any given combination of genes. Furthermore, elucidation of the unknown essential genes identified by this study may eventually yield whole-cell *in silico* models of unprecedented complexity.

A gulf still exists between the minimal synthetic genome described here and a minimal synthetic cell. Hutchison, Chuang, and colleagues "booted up" their genome in the cytoplasm of wild-type *M. capricolum*, thus giving it access to a nonreduced set of gene products. Further work might clarify which gene products are required for booting up a synthetic genome, perhaps through an entirely reconstituted system involving recombinant proteins and a synthetic genome bound within a synthetic lipid bilayer.

Author Information
E-mail: Zev.gartner@ucsf.edu.

REFERENCES

(1) Nicolaou, K.; Vourloumis, D.; Winssinger, N.; Baran, P. The Art and Science of Total Synthesis at the Dawn of the Twenty-First Century. *Angew. Chem., Int. Ed.* **2000**, *39*, 44−122.
(2) Yeh, B. J.; Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **2007**, *3*, 521−525.
(3) Steen, E. J.; et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **2010**, *463*, 559−562.
(4) Roybal, K. T.; et al. Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell* **2016**, *164*, 770−779.
(5) Todhunter, M. E.; et al. Programmed synthesis of three-dimensional tissues. *Nat. Methods* **2015**, *12*, 975−981.
(6) Widmaier, D. M.; et al. Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **2009**, *5*, 309.
(7) Blight, K. J.; Kolykhalov, A. A.; Rice, C. M. Efficient Initiation of HCV RNA Replication in Cell Culture. *Science (Washington, DC, U. S.)* **2000**, *290*, 1972−1974.
(8) Gibson, D. G.; et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **2008**, *319*, 1215−1220.
(9) Gibson, D. G.; et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **2010**, *329*, 52−56.
(10) Hutchison, C. A.; et al. Design and synthesis of a minimal bacterial genome. *Science (Washington, DC, U. S.)* **2016**, *351*, aad6253−aad6253.

DOI: 10.1021/acscentsci.6b00110
ACS Cent. Sci. 2016, 2, 274−277

(11) Fiers, W.; et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **1976**, *260*, 500−507.
(12) Mushegian, A. R.; Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 10268−10273.
(13) Glass, J. I.; et al. Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 425−430.
(14) Karr, J. R.; et al. A whole-cell computational model predicts phenotype from genotype. *Cell* **2012**, *150*, 389−401.
(15) Forster, A. C.; Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2006**, *2*, 45.

DOI: 10.1021/acscentsci.6b00110
ACS Cent. Sci. 2016, 2, 274−277