

WILEY-VCH

Edited by Christina Smolke

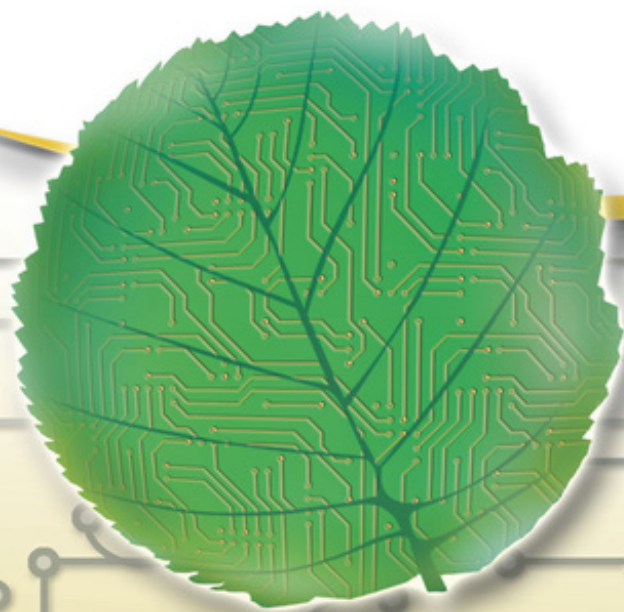
Synthetic Biology

Parts, Devices and Applications

Volume 8

Series Editors:

S. Y. Lee, J. Nielsen,
G. Stephanopoulos



Advanced Biotechnology

Synthetic Biology: Parts, Devices and Applications

Related Titles

Luisi, P.P. (ed.)

Chemical Synthetic Biology

2011

Print ISBN: 978-0-470-71397-6

Further Volumes of the "Advanced Biotechnology" Series:

Published:

Villadsen, J. (ed.)

Fundamental Bioengineering

2016

Print ISBN: 978-3-527-33674-6

Love, J. Ch. (ed.)

Micro- and Nanosystems for Biotechnology

2016

Print ISBN: 978-3-527-33281-6

Wittmann, Ch., Liao, J.C. (eds.)

Industrial Biotechnology Microorganisms (2 Volumes)

2017

Print ISBN: 978-3-527-34179-5

Wittmann, Ch., Liao, J.C. (eds.)

Industrial Biotechnology Products and Processes

2017

Print ISBN: 978-3-527-34181-8

Yoshida, T. (ed.)

Applied Bioengineering

2017

Print ISBN: 978-3-527-34075-0

Nielsen, J., Hohmann, S. (eds)

Systems Biology

2017

Print ISBN: 978-3-527-33558-9

Coming soon:

Chang, H. N.

Emerging Areas in Bioengineering

2018

Print ISBN: 978-3-527-34088-0

Planned:

Lee, G. M., Fastrup Kildegaard, H.

Cell Culture Engineering Recombinant Protein Production

2018

Print ISBN: 9783527343348

Synthetic Biology: Parts, Devices and Applications

Edited by Christina Smolke

WILEY-VCH

Volume Editor

Christina Smolke

Stanford School of Medicine
Y2E2 Building, Room #B07
473 Via Ortega
Stanford, CA 94305
United States

Series Editors

Sang Yup Lee

KAIST
Department of Chemical & Biomolecular
Engineering
291 Daehak-ro, Yuseong-gu
34141 Daejeon
Republic of Korea

Jens Nielsen

Chalmers University of Technology
Department of Biology and Biological
Engineering
Kemivägen 10
41296 Göteborg
Sweden

Gregory Stephanopoulos

Massachusetts Institute of Technology
Department of Chemical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139
USA

Cover

fotolia_yurakp and fotolia_EvgeniyBobrov

■ All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Print ISBN: 978-3-527-33075-1

ePDF ISBN: 978-3-527-68808-1

ePub ISBN: 978-3-527-68809-8

Mobi ISBN: 978-3-527-68807-4

oBook ISBN: 978-3-527-68810-4

Cover Design Adam-Design, Weinheim, Germany

Typesetting SPi Global, Chennai, India

Printing and Binding

Printed on acid-free paper

Contents

About the Series Editors xv

Part I DNA Synthesis and Genome Engineering 1

- 1 Competition and the Future of Reading and Writing DNA 3**
Robert Carlson
- 1.1 Productivity Improvements in Biological Technologies 3
 1.2 The Origin of Moore's Law and Its Implications for Biological Technologies 5
 1.3 Lessons from Other Technologies 6
 1.4 Pricing Improvements in Biological Technologies 7
 1.5 Prospects for New Assembly Technologies 8
 1.6 Beyond Programming Genetic Instruction Sets 10
 1.7 Future Prospects 10
 References 11
- 2 Trackable Multiplex Recombineering (TRMR) and Next-Generation Genome Design Technologies: Modifying Gene Expression in *E. coli* by Inserting Synthetic DNA Cassettes and Molecular Barcodes 15**
Emily F. Freed, Gur Pines, Carrie A. Eckert, and Ryan T. Gill
- 2.1 Introduction 15
 2.2 Current Recombineering Techniques 16
 2.2.1 Recombineering Systems 17
 2.2.2 Current Model of Recombination 17
 2.3 Trackable Multiplex Recombineering 19
 2.3.1 TRMR and T²RMR Library Design and Construction 19
 2.3.2 Experimental Procedure 23
 2.3.3 Analysis of Results 24
 2.4 Current Challenges 25
 2.4.1 TRMR and T²RMR are Currently Not Recursive 26
 2.4.2 Need for More Predictable Models 26
 2.5 Complementing Technologies 27
 2.5.1 MAGE 27
 2.5.2 CREATE 27

2.6	Conclusions	28
	Definitions	28
	References	29
3	Site-Directed Genome Modification with Engineered Zinc Finger Proteins	33
	<i>Lauren E. Woodard, Daniel L. Galvan, and Matthew H. Wilson</i>	
3.1	Introduction to Zinc Finger DNA-Binding Domains and Cellular Repair Mechanisms	33
3.1.1	Zinc Finger Proteins	33
3.1.2	Homologous Recombination	34
3.1.3	Non-homologous End Joining	35
3.2	Approaches for Engineering or Acquiring Zinc Finger Proteins	36
3.2.1	Modular Assembly	37
3.2.2	OPEN and CoDA Selection Systems	37
3.2.3	Purchase via Commercial Avenues	38
3.3	Genome Modification with Zinc Finger Nucleases	38
3.4	Validating Zinc Finger Nuclease-Induced Genome Alteration and Specificity	40
3.5	Methods for Delivering Engineered Zinc Finger Nucleases into Cells	41
3.6	Zinc Finger Fusions to Transposases and Recombinases	41
3.7	Conclusions	42
	References	43
4	Rational Efforts to Streamline the <i>Escherichia coli</i> Genome	49
	<i>Gabriella Balikó, Viktor Vernyik, Ildikó Karcagi, Zsuzsanna Györfy, Gábor Draskovits, Tamás Fehér, and György Pósfai</i>	
4.1	Introduction	49
4.2	The Concept of a Streamlined Chassis	50
4.3	The <i>E. coli</i> Genome	51
4.4	Random versus Targeted Streamlining	54
4.5	Selecting Deletion Targets	55
4.5.1	General Considerations	55
4.5.1.1	Naturally Evolved Minimal Genomes	55
4.5.1.2	Gene Essentiality Studies	55
4.5.1.3	Comparative Genomics	56
4.5.1.4	<i>In silico</i> Models	56
4.5.1.5	Architectural Studies	56
4.5.2	Primary Deletion Targets	57
4.5.2.1	Prophages	57
4.5.2.2	Insertion Sequences (ISs)	57
4.5.2.3	Defense Systems	57

4.5.2.4	Genes of Unknown and Exotic Functions	58
4.5.2.5	Repeat Sequences	58
4.5.2.6	Virulence Factors and Surface Structures	58
4.5.2.7	Genetic Diversity-Generating Factors	59
4.5.2.8	Redundant and Overlapping Functions	59
4.6	Targeted Deletion Techniques	59
4.6.1	General Considerations	59
4.6.2	Basic Methods and Strategies	60
4.6.2.1	Circular DNA-Based Method	60
4.6.2.2	Linear DNA-Based Method	62
4.6.2.3	Strategy for Piling Deletions	62
4.6.2.4	New Variations on Deletion Construction	63
4.7	Genome-Reducing Efforts and the Impact of Streamlining	64
4.7.1	Comparative Genomics-Based Genome Stabilization and Improvement	64
4.7.2	Genome Reduction Based on Gene Essentiality	66
4.7.3	Complex Streamlining Efforts Based on Growth Properties	67
4.7.4	Additional Genome Reduction Studies	68
4.8	Selected Research Applications of Streamlined-Genome <i>E. coli</i>	68
4.8.1	Testing Genome Streamlining Hypotheses	68
4.8.2	Mobile Genetic Elements, Mutations, and Evolution	69
4.8.3	Gene Function and Network Regulation	69
4.8.4	Codon Reassignment	70
4.8.5	Genome Architecture	70
4.9	Concluding Remarks, Challenges, and Future Directions	71
	References	73
5	Functional Requirements in the Program and the Cell Chassis for Next-Generation Synthetic Biology	81
	<i>Antoine Danchin, Agnieszka Sekowska, and Stanislas Noria</i>	
5.1	A Prerequisite to Synthetic Biology: An Engineering Definition of What Life Is	81
5.2	Functional Analysis: Master Function and Helper Functions	83
5.3	A Life-Specific Master Function: Building Up a Progeny	85
5.4	Helper Functions	86
5.4.1	Matter: Building Blocks and Structures (with Emphasis on DNA)	87
5.4.2	Energy	91
5.4.3	Managing Space	92
5.4.4	Time	95
5.4.5	Information	96
5.5	Conclusion	97
	Acknowledgments	98
	References	98

Part II Parts and Devices Supporting Control of Protein Expression and Activity 107

6	Constitutive and Regulated Promoters in Yeast: How to Design and Make Use of Promoters in <i>S. cerevisiae</i> 109
	<i>Diana S. M. Ottoz and Fabian Rudolf</i>
6.1	Introduction 109
6.2	Yeast Promoters 110
6.3	Natural Yeast Promoters 113
6.3.1	Regulated Promoters 113
6.3.2	Constitutive Promoters 115
6.4	Synthetic Yeast Promoters 116
6.4.1	Modified Natural Promoters 116
6.4.2	Synthetic Hybrid Promoters 117
6.5	Conclusions 121
	Definitions 122
	References 122
7	Splicing and Alternative Splicing Impact on Gene Design 131
	<i>Beatrix Suess, Katrin Kemmerer, and Julia E. Weigand</i>
7.1	The Discovery of “Split Genes” 131
7.2	Nuclear Pre-mRNA Splicing in Mammals 132
7.2.1	Introns and Exons: A Definition 132
7.2.2	The Catalytic Mechanism of Splicing 132
7.2.3	A Complex Machinery to Remove Nuclear Introns: The Spliceosome 132
7.2.4	Exon Definition 134
7.3	Splicing in Yeast 135
7.3.1	Organization and Distribution of Yeast Introns 135
7.4	Splicing without the Spliceosome 136
7.4.1	Group I and Group II Self-Splicing Introns 136
7.4.2	tRNA Splicing 137
7.5	Alternative Splicing in Mammals 137
7.5.1	Different Mechanisms of Alternative Splicing 137
7.5.2	Auxiliary Regulatory Elements 139
7.5.3	Mechanisms of Splicing Regulation 140
7.5.4	Transcription-Coupled Alternative Splicing 142
7.5.5	Alternative Splicing and Nonsense-Mediated Decay 143
7.5.6	Alternative Splicing and Disease 144
7.6	Controlled Splicing in <i>S. cerevisiae</i> 145
7.6.1	Alternative Splicing 145
7.6.2	Regulated Splicing 146
7.6.3	Function of Splicing in <i>S. cerevisiae</i> 147
7.7	Splicing Regulation by Riboswitches 147
7.7.1	Regulation of Group I Intron Splicing in Bacteria 148
7.7.2	Regulation of Alternative Splicing by Riboswitches in Eukaryotes 148

7.8	Splicing and Synthetic Biology	150
7.8.1	Impact of Introns on Gene Expression	150
7.8.2	Control of Splicing by Engineered RNA-Based Devices	151
7.9	Conclusion	153
	Acknowledgments	153
	Definitions	153
	References	153
8	Design of Ligand-Controlled Genetic Switches Based on RNA Interference	169
	<i>Shunnichi Kashida and Hirohide Saito</i>	
8.1	Utility of the RNAi Pathway for Application in Mammalian Cells	169
8.2	Development of RNAi Switches that Respond to Trigger Molecules	170
8.2.1	Small Molecule-Triggered RNAi Switches	171
8.2.2	Oligonucleotide-Triggered RNAi Switches	173
8.2.3	Protein-Triggered RNAi Switches	174
8.3	Rational Design of Functional RNAi Switches	174
8.4	Application of the RNAi Switches	175
8.5	Future Perspectives	177
	Definitions	178
	References	178
9	Small Molecule-Responsive RNA Switches (Bacteria): Important Element of Programming Gene Expression in Response to Environmental Signals in Bacteria	181
	<i>Yohei Yokobayashi</i>	
9.1	Introduction	181
9.2	Design Strategies	181
9.2.1	Aptamers	181
9.2.2	Screening and Genetic Selection	182
9.2.3	Rational Design	183
9.3	Mechanisms	183
9.3.1	Translational Regulation	183
9.3.2	Transcriptional Regulation	184
9.4	Complex Riboswitches	185
9.5	Conclusions	185
	Keywords with Definitions	185
	References	186
10	Programming Gene Expression by Engineering Transcript Stability Control and Processing in Bacteria	189
	<i>Jason T. Stevens and James M. Carothers</i>	
10.1	An Introduction to Transcript Control	189
10.1.1	Why Consider Transcript Control?	189
10.1.2	The RNA Degradation Process in <i>E. coli</i>	190

10.1.3	The Effects of Translation on Transcript Stability	192
10.1.4	Structural and Noncoding RNA-Mediated Transcript Control	193
10.1.5	Polyadenylation and Transcript Stability	195
10.2	Synthetic Control of Transcript Stability	195
10.2.1	Transcript Stability Control as a “Tuning Knob”	195
10.2.2	Secondary Structure at the 5′ and 3′ Ends	196
10.2.3	Noncoding RNA-Mediated	197
10.2.4	Model-Driven Transcript Stability Control for Metabolic Pathway Engineering	198
10.3	Managing Transcript Stability	201
10.3.1	Transcript Stability as a Confounding Factor	201
10.3.2	Anticipating Transcript Stability Issues	201
10.3.3	Uniformity of 5′ and 3′ Ends	202
10.3.4	RBS Sequestration by Riboregulators and Riboswitches	203
10.3.5	Experimentally Probing Transcript Stability	204
10.4	Potential Mechanisms for Transcript Control	205
10.4.1	Leveraging New Tools	205
10.4.2	Unused Mechanisms Found in Nature	206
10.5	Conclusions and Discussion	207
	Acknowledgments	208
	Definitions	208
	References	209
11	Small Functional Peptides and Their Application in Superfunctionalizing Proteins	217
	<i>Sonja Billerbeck</i>	
11.1	Introduction	217
11.2	Permissive Sites and Their Identification in a Protein	218
11.3	Functional Peptides	220
11.3.1	Functional Peptides that Act as Binders	220
11.3.2	Peptide Motifs that are Recognized by Labeling Enzymes	221
11.3.3	Peptides as Protease Cleavage Sites	222
11.3.4	Reactive Peptides	223
11.3.5	Pharmaceutically Relevant Peptides: Peptide Epitopes, Sugar Epitope Mimics, and Antimicrobial Peptides	223
11.3.5.1	Peptide Epitopes	224
11.3.5.2	Peptide Mimotopes	224
11.3.5.3	Antimicrobial Peptides	225
11.4	Conclusions	227
	Definitions	228
	Abbreviations	228
	Acknowledgment	229
	References	229

Part III Parts and Devices Supporting Spatial Engineering 237

- 12 Metabolic Channeling Using DNA as a Scaffold 239**
Moja Benčina, Jerneja Mori, Rok Gaber, and Roman Jerala
- 12.1 Introduction 239
- 12.2 Biosynthetic Applications of DNA Scaffold 242
- 12.2.1 L-Threonine 242
- 12.2.2 *trans*-Resveratrol 245
- 12.2.3 1,2-Propanediol 246
- 12.2.4 Mevalonate 246
- 12.3 Design of DNA-Binding Proteins and Target Sites 247
- 12.3.1 Zinc Finger Domains 248
- 12.3.2 TAL-DNA Binding Domains 249
- 12.3.3 Other DNA-Binding Proteins 250
- 12.4 DNA Program 250
- 12.4.1 Spacers between DNA-Target Sites 250
- 12.4.2 Number of DNA Scaffold Repeats 252
- 12.4.3 DNA-Target Site Arrangement 253
- 12.5 Applications of DNA-Guided Programming 254
- Definitions 255
- References 256
- 13 Synthetic RNA Scaffolds for Spatial Engineering in Cells 261**
Gairik Sachdeva, Cameron Myhrvold, Peng Yin, and Pamela A. Silver
- 13.1 Introduction 261
- 13.2 Structural Roles of Natural RNA 261
- 13.2.1 RNA as a Natural Catalyst 262
- 13.2.2 RNA Scaffolds in Nature 263
- 13.3 Design Principles for RNA Are Well Understood 263
- 13.3.1 RNA Secondary Structure is Predictable 264
- 13.3.2 RNA can Self-Assemble into Structures 265
- 13.3.3 Dynamic RNAs can be Rationally Designed 265
- 13.3.4 RNA can be Selected *in vitro* to Enhance Its Function 266
- 13.4 Applications of Designed RNA Scaffolds 266
- 13.4.1 Tools for RNA Research 266
- 13.4.2 Localizing Metabolic Enzymes on RNA 267
- 13.4.3 Packaging Therapeutics on RNA Scaffolds 269
- 13.4.4 Recombinant RNA Technology 269
- 13.5 Conclusion 270
- 13.5.1 New Applications 270
- 13.5.2 Technological Advances 270
- Definitions 271
- References 271

14	Sequestered: Design and Construction of Synthetic Organelles	279
	<i>Thawatchai Chaijarasphong and David F. Savage</i>	
14.1	Introduction	279
14.2	On Organelles	281
14.3	Protein-Based Organelles	283
14.3.1	Bacterial Microcompartments	283
14.3.1.1	Targeting	285
14.3.1.2	Permeability	287
14.3.1.3	Chemical Environment	288
14.3.1.4	Biogenesis	289
14.3.2	Alternative Protein Organelles: A Minimal System	290
14.4	Lipid-Based Organelles	292
14.4.1	Repurposing Existing Organelles	293
14.4.1.1	The Mitochondrion	293
14.4.1.2	The Vacuole	294
14.5	<i>De novo</i> Organelle Construction and Future Directions	295
	Acknowledgments	297
	References	297

**Part IV Early Applications of Synthetic Biology: Pathways, Therapies,
and Cell-Free Synthesis 307**

15	Cell-Free Protein Synthesis: An Emerging Technology for Understanding, Harnessing, and Expanding the Capabilities of Biological Systems	309
	<i>Jennifer A. Schoborg and Michael C. Jewett</i>	
15.1	Introduction	309
15.2	Background/Current Status	311
15.2.1	Platforms	311
15.2.1.1	Prokaryotic Platforms	311
15.2.1.2	Eukaryotic Platforms	312
15.2.2	Trends	314
15.3	Products	316
15.3.1	Noncanonical Amino Acids	316
15.3.2	Glycosylation	316
15.3.3	Antibodies	318
15.3.4	Membrane Proteins	318
15.4	High-Throughput Applications	320
15.4.1	Protein Production and Screening	320
15.4.2	Genetic Circuit Optimization	321
15.5	Future of the Field	321
	Definitions	322
	Acknowledgments	322
	References	323

16	Applying Advanced DNA Assembly Methods to Generate Pathway Libraries	331
	<i>Dawn T. Eriksen, Ran Chao, and Huimin Zhao</i>	
16.1	Introduction	331
16.2	Advanced DNA Assembly Methods	333
16.3	Generation of Pathway Libraries	334
16.3.1	<i>In vitro</i> Assembly Methods	335
16.3.2	<i>In vivo</i> Assembly Methods	339
16.3.2.1	<i>In vivo</i> Chromosomal Integration	339
16.3.2.2	<i>In vivo</i> Plasmid Assembly and One-Step Optimization Libraries	340
16.3.2.3	<i>In vivo</i> Plasmid Assembly and Iterative Multi-step Optimization Libraries	341
16.4	Conclusions and Prospects	343
	Definitions	343
	References	344
17	Synthetic Biology in Immunotherapy	349
	<i>Patrick Ho and Yvonne Y. Chen</i>	
17.1	The Need for a New Therapeutic Paradigm	349
17.2	Rationale for Cellular Therapies	350
17.3	Synthetic Biology Approaches to Cellular Immunotherapy Engineering	351
17.3.1	CAR Engineering for Adoptive T-Cell Therapy	352
17.3.2	Genetic Engineering to Enhance T-Cell Therapeutic Function	357
17.3.3	Generating Safer T-Cell Therapeutics with Synthetic Biology	359
17.4	Challenges and Future Outlook	362
	Acknowledgment	364
	Definitions	364
	References	365
	Part V Societal Ramifications of Synthetic Biology	373
18	Synthetic Biology: From Genetic Engineering 2.0 to Responsible Research and Innovation	375
	<i>Lei Pei and Markus Schmidt</i>	
18.1	Introduction	375
18.2	Public Perception of the Nascent Field of Synthetic Biology	376
18.2.1	Perception of Synthetic Biology in the United States	377
18.2.2	Perception of Synthetic Biology in Europe	379
18.2.2.1	European Union	379
18.2.2.2	Austria	379

18.2.2.3	Germany	381
18.2.2.4	Netherlands	382
18.2.2.5	United Kingdom	383
18.2.3	Opinions from Concerned Civil Society Groups	384
18.3	Frames and Comparators	384
18.3.1	Genetic Engineering: Technology as Conflict	386
18.3.2	Nanotechnology: Technology as Progress	387
18.3.3	Information Technology: Technology as Gadget	387
18.3.4	SB: Which Debate to Come?	388
18.4	Toward Responsible Research and Innovation (RRI) in Synthetic Biology	389
18.4.1	Engagement of All Societal Actors – Researchers, Industry, Policy Makers, and Civil Society – and Their Joint Participation in the Research and Innovation	390
18.4.2	Gender Equality	391
18.4.3	Science Education	392
18.4.4	Open Access	392
18.4.5	Ethics	394
18.4.6	Governance	395
18.5	Conclusion	396
	Acknowledgments	397
	References	397

Index	403
--------------	------------

About the Series Editors



Sang Yup Lee is distinguished Professor at the Department of Chemical and Biomolecular Engineering at the Korea Advanced Institute of Science and Technology. At present, Prof. Lee is the Director of the Center for Systems and Synthetic Biotechnology, Director of the BioProcess Engineering Research Center, and Director of the Bioinformatics Research Center. He has published more than 500 journal papers, 64 books, and book chapters, and has more than 580 patents (either registered or applied) to his credit. He has received numerous awards, including the National Order of Merit, the Merck Metabolic Engineering Award, the ACS Marvin Johnson Award, Charles Thom Award, Amgen Biochemical Engineering Award, Elmer Gaden Award, POSCO TJ Park Prize, and HoAm Prize. He is Fellow of American Association for the Advancement of Science, the American Academy of Microbiology, American Institute of Chemical Engineers, Society for Industrial Microbiology and Biotechnology, American Institute of Medical and Biological Engineering, the World Academy of Science, the Korean Academy of Science and Technology, and the National Academy of Engineering of Korea. He is also Foreign Member of National Academy of Engineering, USA. In addition, he is honorary professor of the University of Queensland (Australia), honorary professor of the Chinese Academy of Sciences, honorary professor of Wuhan University (China), honorary professor of Hubei University of Technology (China), honorary professor of Beijing University of Chemical Technology (China), and advisory professor of the Shanghai Jiaotong University (China). Apart from his academic associations, Prof. Lee is the editor-in-chief of the *Biotechnology Journal* and is also contributing to numerous other journals as associate editor and board member. Prof. Lee is serving as a member of Presidential Advisory Committee on Science and Technology (South Korea).



Jens Nielsen is Professor and Director to Chalmers University of Technology (Sweden) since 2008. He obtained an MSc degree in chemical engineering and a PhD degree (1989) in biochemical engineering from the Technical University of Denmark (DTU) and after that established his independent research group and was appointed full professor there in 1998. He was Fulbright visiting professor at MIT in 1995–1996. At DTU, he founded and directed the Center for Microbial Biotechnology. Prof. Nielsen has published more than 350 research papers and coauthored more than 40 books, and he is inventor of more than 50 patents. He has founded several companies that have raised more than 20 million in venture capital. He has received numerous Danish and international awards and is member of the Academy of Technical Sciences (Denmark), the National Academy of Engineering (USA), the Royal Danish Academy of Science and Letters, the American Institute for Medical and Biological Engineering and the Royal Swedish Academy of Engineering Sciences.



Gregory Stephanopoulos is the W.H. Dow Professor of Chemical Engineering at the Massachusetts Institute of Technology (MIT, USA) and Director of the MIT Metabolic Engineering Laboratory. He is also Instructor of Bioengineering at Harvard Medical School (since 1997). He received his BS degree from the National Technical University of Athens and PhD from the University of Minnesota (USA). He has coauthored about 400 research papers and 50 patents, along with the first textbook on metabolic engineering. He has been recognized by numerous awards from the American Institute of Chemical Engineers (AIChE) (Wilhelm, Walker and Founders awards), American Chemical Society (ACS), Society of Industrial Microbiology (SIM), BIO (Washington Carver Award), the John Fritz Medal of the American Association of Engineering Societies, and others. In 2003, he was elected member of the National Academy of Engineering (USA) and in 2014 President of AIChE.

Part I

DNA Synthesis and Genome Engineering

1

Competition and the Future of Reading and Writing DNA

Robert Carlson

Biodesic and Bioeconomy Capital, 3417 Evanston Ave N, Ste 329, Seattle, WA 98103, USA

Constructing arbitrary genetic instruction sets is a core technology for biological engineering. Biologists and engineers are pursuing even better methods to assemble these arbitrary sequences from synthetic oligonucleotides (oligos) [1]. These new assembly methods in principle reduce costs, improve access, and result in long sequences of error-free DNA that can be used to construct entire microbial genomes [2]. However, an increasing diversity of assembly methods is not matched by any obvious corresponding innovation in producing oligos. Commercial oligo production employs a very narrow technology base that is many decades old. Consequently, there is only minimal price and product differentiation among corporations that produce oligos. Prices have stagnated, which in turn limits the economic potential of new assembly methods that rely on oligos. Improvements may come via recently demonstrated assembly methods that are capable of using oligos of lower quality and lower cost as feedstocks. However, while these new methods may substantially lower the cost of gene-length double-stranded DNA (dsDNA), they also may be economically viable only when producing many orders of magnitude with more dsDNA than what is now used by the market. The commercial success of these methods, and the broader access to dsDNA they enable, may therefore depend on structural changes in the market that are yet to emerge.

1.1 Productivity Improvements in Biological Technologies

In considering the larger impact of technological monoculture in DNA synthesis, it is useful to contrast DNA synthesis and assembly with DNA sequencing. In particular, it is instructive to compare productivity estimates of commercially available sequencing and synthesis instruments (Figure 1.1). Reading DNA is as crucial as writing DNA to the future of biological engineering. Due to not just commercial competition but also competition between sequencing technologies, both

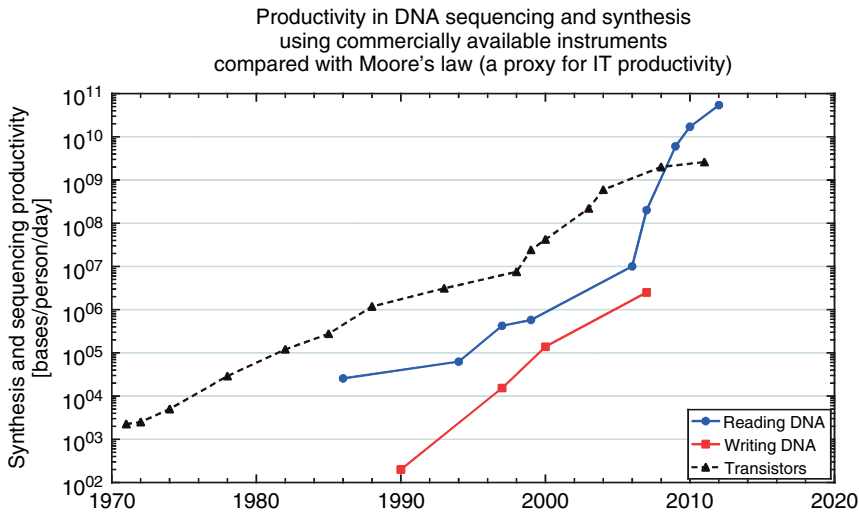


Figure 1.1 Estimates of the maximum productivity of DNA synthesis and sequencing enabled by commercially available instruments. Productivity of DNA synthesis is shown only for column-based synthesis instruments, as data for sDNA fabricated on commercially available DNA arrays is unavailable; exceptions are discussed in the text. Shown for comparison is Moore's law, the number of transistors per chip. (Intel; Carlson, 2010 [3]; Loman *et al.* 2012 [4]; Quail *et al.* 2012 [5]; Liu, 2012 [6].)

prices and instrument capabilities are improving rapidly. The technological diversity responsible for these improvements poses challenges in making quantitative comparisons. As in previous discussions of these trends, in what follows I rely on the metrics of price [\$/base] and productivity [bases/person/day].

Figure 1.1 also directly compares the productivity enabled by commercially available sequencing and synthesis instruments to Moore's law, which describes the exponential increase in transistor counts in CPUs over time. Readers new to this discussion are referred to References 3 and 4 for in-depth descriptions of the development of these metrics and the utility of a comparison with Moore's law [3, 7]. Very briefly, Moore's law is a proxy for productivity; more transistors enable greater computational capability, which putatively equates to greater productivity.

Visual inspection of Figure 1.1 reveals several interesting features. First, general synthesis productivity has not improved for several years because no new instruments have been released publicly since about 2008. Productivity estimates for instruments developed and run by oligo and gene synthesis service providers are not publicly available.¹

¹ It is likely that array-based DNA synthesis used to supply gene assembly operates at a much higher productivity than column-based synthesis. For example, Agilent reportedly produces and ships in excess of 30 billion bases of ssDNA a day, the equivalent of more than 10 human genomes, on an undisclosed number of arrays (Darlene Solomon, Personal Communication).

Second, it is clear that DNA sequencing platforms are improving very rapidly, now much faster than Moore's law.

Moore's law and its economic and social consequences are often used to benchmark our expectations of other technologies. Therefore, developing an understanding of this "law" provides a means to compare and contrast it with other technological trends.

1.2 The Origin of Moore's Law and Its Implications for Biological Technologies

Moore's law is often mistakenly described as a technological inevitability or is assumed to be some sort of physical phenomenon. It is neither; Moore's law is a business plan, and as such it is based on economics and planning. Gordon Moore's somewhat opaque original statement of what became the "law" was a prediction concerning economically viable transistor yields [8]. Over time, Moore's economic observation became an operational model based on monopoly pricing, and it eventually enabled Intel to outcompete all other manufacturers of general CPUs. Two important features distinguish CPUs from other technologies and provide insight into the future of trends in biological technologies: the first is the cost of production, and the second is the monopoly pricing structure.

Early on Intel recognized the utility of exploiting Moore's law as a business plan. A simple scaling argument reveals the details of the plan. While transistor counts increased exponentially, Intel correspondingly reduced the price per transistor at a similar rate. In order to maintain revenues, the company needed to ship proportionally more transistors every quarter; in fact, the company increased its shipping numbers faster than prices fell, enabling consistent revenue to grow for several decades. This explains why Intel former CEO Andy Grove reportedly constantly pushed for an even greater scale [9].

In this sense, Moore's law was always about economics and planning in a multibillion-dollar industry. In the year 2000, a new chip fab cost about \$1 billion; in 2009, it cost about \$3 billion. Now, according to *The Economist*, Intel estimates that a new chip fab costs about \$10 billion [9]. This apparent exponential increase in the cost of semiconductor processing is known as Rock's law. It is often argued that Moore's law will eventually expire due to the physical constraints of fabricating transistors at small length scales, but it is more likely to become difficult to economically justify constructing fabrication facilities at the cost of tens to hundreds of billions of dollars. Even through the next several iterations, these construction costs will dictate careful planning that spans many years. No business spends \$10 billion without a great deal of planning, and, more directly, no business finances a manufacturing plant that expensive without demonstrating a long-term plan to repay the financiers. Moreover, Intel must coordinate the manufacturing and delivery of very expensive, very complex semiconductor processing instruments made by other companies. Thus Intel's planning and finance cycles explicitly extend many years into the future. New technology has certainly been required to achieve each planning goal, but this is part of the ongoing research, development, and planning process for Intel.

Moore's law served a second purpose for Intel and one that is less well recognized but arguably more important; it was a pace selected to enable Intel to win. Intel successfully organized an entire industry to move at a pace only it could survive. And only Intel did survive. While Intel still has competitors in products such as memory or GPUs, companies that produced high volume, general CPUs have all succumbed to the pace of Moore's law. The final component of this argument is that, according to Gordon Moore, Intel could have increased transistor counts faster than the historical rate.² In fact, Intel ran on a faster internal innovation clock than it admitted publicly, which means that Moore's law was, as one Intel executive put it, a "marketing head fake" [10]. The inescapable conclusion of this argument is that the management of Intel made a very careful calculation; they evaluated product rollouts to consumers – the rate of new product adoption, the rate of semiconductor processing improvements, and the financial requirements for building the next chip fab line – and then set a pace that nobody else could match but that left Intel plenty of headroom for future products. In effect, if not intent, Intel executed a strategy that enabled it to set CPU prices and then to reduce those prices at a rate no other company could match.

This long-term planning, pricing structure, and the resulting lack of competition contrasts quite strongly with the commercial landscape for biological technologies. Whereas the exponential pace of doubling of transistor counts was controlled by just one company, productivity in DNA sequencing has recently improved faster than Moore's law due to competition not just among companies but also among technologies. Conversely, the lack of improvement in synthesis productivity suggests that the narrow technology base for writing DNA has reached technical and, therefore, economic limits. Nonetheless, while Figure 1.1 may suggest a temporary slowdown in the rate of improvement for sequencing, and in effect shows zero recent improvement for synthesis, new technologies will inevitably facilitate continued competition and, therefore, continued productivity improvement.

1.3 Lessons from Other Technologies

Compared with that in other industries, the financial barrier to entry in biological technologies is quite low. Unlike chip manufacturing, there is nothing in biology with a commercial development price tag of \$10 billion. The Boeing 787 reportedly cost \$32 billion to develop as of 2011 and is on top of a century of multibillion-dollar aviation projects that preceded it [11]. Better Place, an electric car company, declared bankruptcy after receiving \$850 million in investment [12]. Tesla Motors has reported only one profitable quarter since 2003 and continues to operate in the red while working to achieve manufacturing scale-up [13, 14].

There are two kinds of costs that are important to distinguish here. The first is the cost of developing and commercializing a particular product. Based on the

2 Gordon Moore to Danny Hillis, as related by Danny Hillis, Personal Communication.

money reportedly raised and spent by Illumina, Pacific Biosciences, Oxford Nanopore, Life, Ion Torrent, and Complete Genomics (the latter three before acquisition), it appears that developing and marketing a second-generation sequencing technology can cost more than \$100 million. Substantially more money gets spent, and lost, in operations before any of these product lines is revenue positive. Nonetheless, relatively low development costs have enabled a number of companies to enter the market for DNA sequencing, resulting in a healthy competition in a market that is presently modest in size but that is expected to grow rapidly over the coming decades.

1.4 Pricing Improvements in Biological Technologies

The second kind of cost to keep in mind is the use of new technologies to produce an object or produce data. Figure 1.2 is a plot of commercial prices for column-synthesized oligos, gene-length synthetic DNA (sDNA), and DNA sequencing. Prior to 2006, the sequencing market was dominated by Sanger-based capillary instruments produced by Applied Biosystems, in effect another pricing monopoly. After 2006, the market saw a rapid proliferation of not just commercial but also technological competition with the launch of next-generation systems from 454, Illumina, Ion Torrent, Pacific Biosciences, and Oxford Nanopore based on a diversity of chemical and physical detection methodologies [15]. Illumina presently dominates the market for sequencing instruments but is facing competition from Oxford Nanopore and various Chinese insurgents. There also remains technological diversity between companies, which contributes to competitive

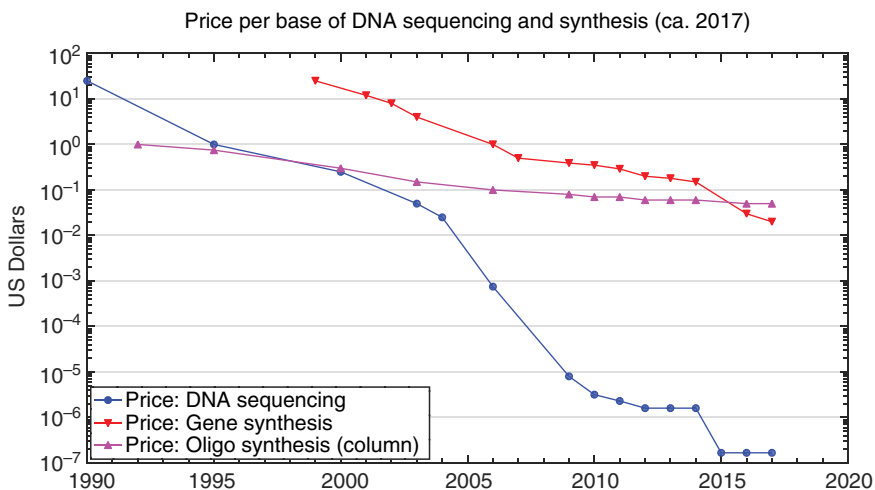


Figure 1.2 Commercial prices per base for DNA sequencing, column-synthesized oligonucleotides, and gene-length sDNA. Reported prices for array-synthesized oligos vary widely, and no time series is available. Market pricing for genes can vary by up to an order of magnitude, depending on sequencing composition and complexity. (Carlson (2010), Commercial price quotes.)

pressures. An important consequence of the emergence of technological competition in the DNA sequencing market is a rapid price decrease. The NIH maintains a version of this plot that compares sequencing prices with cost per megabyte for memory, another form of Moore's law [16]. Both Figure 1.2 and the NIH plot show that sequencing costs kept pace with Moore's law while a pricing monopoly was in effect. The emergence of technological competition produced both productivity improvements and price changes that outpaced Moore's law.

In contrast, despite modest commercial competition in the DNA synthesis market, the lack of technological competition has limited price decreases in the last 5 years. The industry as it exists today is based on chemistry that is several decades old, in which oligos are synthesized step by step on an immobilized substrate. Using array-synthesized oligos for gene assembly appears to be lowering the market price, though quality and delivery time are reportedly inconsistent across the industry. Improved error correction and removal technologies may further reduce the assembly cost for genes and thereby improve the profit margins [17]. My informal conversations with industry insiders suggest that oligo producers may no longer include the cost of goods in calculating prices; that is, oligo prices are evidently determined largely by the cost of capital rather than the cost of raw materials. This suggests that very little pricing improvement can be expected for genes produced from standard oligo synthesis.

1.5 Prospects for New Assembly Technologies

Array synthesis has the advantage of a low volume production of oligos with high library diversity [18]. Gene assembly based on array synthesis has proved difficult to commercialize. At least three companies in this space, Codon Devices, Gen9, and Cambrian Genomics, have gone bankrupt or been acquired in recent years. Twist, a more recent entrant, now quotes prices in the neighborhood of \$10 per base and publicly asserts it will push prices much lower in the coming years.

With prices potentially soon falling by orders of magnitude, one must ask about the subsequent impact on the market for synthetic genes. New firms entering the market are implicitly working on the hypothesis that supply-side price reductions will drive increased demand. The most obvious source of that demand would be forward design of genetic circuits based on rational models. Yet the most sophisticated synthetic genetic circuits being constructed in industrial settings are designed largely using heuristic models rather than quantitative design tools [19]. Moreover, these circuits contain only a handful of components, which stand as a substantial bottleneck for demand. Alternatively, customers may employ less up-front predictive design and instead rely on high-throughput screening of pathway variants; screening libraries of pathways has the potential to create substantial demand for synthetic genes [20].

Considering the interplay between market size and price reveals challenges for companies entering the gene synthesis industry. Recalling the lessons of Moore's law, a relatively simple scaling argument will reveal the performance constraints

of the gene synthesis industry. Intel knew that it could grow financially in the context of exponentially falling transistor costs by shipping exponentially more transistors every quarter – that is, the business model of Moore’s law. But that was in the context of an effective pricing monopoly, and Intel’s success required a market that grew exponentially for decades. The question for synthetic gene companies is whether the market will grow fast enough to provide adequate revenues when prices fall. For every order of magnitude drop in the price of synthetic genes, the industry will have to ship an order of magnitude of more DNA just to maintain constant revenues. More broadly, in order for the industry to grow, synthesis companies must find a way to expand their market at a rate faster than when prices fall. Unfortunately, as best as I can tell, despite falling prices and putative increases in demand, the global gene synthesis industry generated only about \$150 million in 2015 [21]. The total size of the industry appears to have been static, or even to have decreased, over the prior decade.

Ultimately, for a new wave of gene synthesis companies to be successful, they have to provide their customers with something of value. Academic customers are likely to become more plentiful as it becomes even more obvious that ordering genes is cheaper than cloning genes, even with graduate student labor costs. Gene synthesis pioneer John Mulligan used to cite NIH expenditures on cloning – approximately \$3 billion annually – as a potential market size for gene synthesis [22]. This is certainly an attractive potential market. However, with the price per base potentially falling dramatically in the near term, the comparison to cloning must focus on the total number of cloned bases replaced by synthesis and at what exact price.

For commercial customers, it is less obvious that lower prices will equate to substantial increases in demand. The cost of sDNA is always going to be a small cost of developing a product, and it is not obvious that making a small cost even smaller will affect the operations of an average corporate lab. In general, research only accounts for 1–10% of the cost of the final product [23]. The vast majority of development costs are in scaling up production and in polishing the product into something customers will actually buy. For the sake of argument, assume that the total metabolic engineering development costs for a new product are in the neighborhood of \$50–100 million, a reasonable estimate given the amounts that companies such as Gevo and Amyris have reportedly spent. In that context, reducing the cost of sDNA from \$50 000 to \$500 may be useful, but the corporate scientist-customer will be more concerned about reducing the \$50 million overall costs by a factor of two, or even an order of magnitude, a decrease that would drive the cost of sDNA into the noise. Thus, in order to increase demand adequately, the production of radically cheaper sDNA must be coupled with innovations that reduce the overall the product development costs. As suggested above, forward design of complex circuits is unlikely to provide adequate innovation anytime soon. An alternative may be high-throughput screening operations that enable testing many variant pathways simultaneously. But note that this is not just another hypothesis about how the immediate future of engineering biology will change but also another generally unacknowledged hypothesis. It might turn out to be wrong, and elucidating one final difference between transistors and DNA may explain why.

The global market for transistors has grown consistently for decades, driven by an insatiable demand for more computational power and digital storage. Every new product must contain more transistors than the model it replaces. In contrast, while the demand for biological products is also growing, every new biological product is made using, in principle, just one DNA sequence. In practice, while many different DNA sequences may be constructed and tested in developing a new product, these many sequences are still winnowed down to only one sequence that defines a microbial, plant, or mammalian production strain. Nevertheless, this fundamental difference in use between transistors and DNA reveals the gene synthesis industry as the provider of engineering prototypes rather than as a large volume manufacturer of consumer goods. Consequently, while high-throughput synthetic biology companies such as Amyris, Ginkgo Bioworks, and Zymergen may place relatively large orders for sDNA, the price and volume of that sDNA will never have much impact on the final products produced by those companies.

1.6 Beyond Programming Genetic Instruction Sets

At present, the cost of purifying oligos and short dsDNA can exceed the cost of the DNA itself by as much a factor of three. The availability of lower cost, high quality dsDNA may therefore enable applications that are presently not economically viable at large scale. Beyond its utility in programming biological systems, dsDNA can be used as nanoscale structural or functional components [24]. The future of these applications is difficult to predict but could include circuitry assembled from DNA that is modified using proteins and chemistry to create conductive and semiconductive regions useful for computation [25]. It is unclear what sDNA market size these applications may support. Recent progress suggests that new demand might emerge from the use of DNA as a digital information storage medium [26]. Even a single, modestly size data center would consume many orders of magnitude of more sDNA than any prospective use of sDNA in biological contexts [27].

1.7 Future Prospects

Regardless of the particular course of companies entering the gene synthesis market, it appears that prices are likely to fall, potentially fueling an increase in demand. That demand may come in part from customers who fall outside the usual academic and corporate classifications; start-up companies, community labs, and individual, independent entrepreneurs and scientists are likely to use sDNA in new and interesting ways. The standing biosecurity strategy of the United States is to explicitly engage and encourage this innovation, including in contexts such as “garages and basements” [28]. This strategy recognizes the important role of entrepreneurs in innovation and job creation and also recognizes the difficulty of preventing access to biological technologies through regulations or restrictions.

Complementing the engagement strategy is an effort to prevent accidentally synthesizing and shipping potentially hazardous sequences. Most gene synthesis companies have voluntarily signed onto international agreements to screen orders against lists of pathogens and toxins such as the Harmonized Screening Protocol of the International Gene Synthesis Consortium (IGSC) [29].

The technical potential of new sDNA production methods may provide an opportunity to build and test far more genetic circuit designs than what is now feasible. The economic demand for biological production is enormous and is growing rapidly [30, 31]. Whether newly emerging sDNA companies survive economically depends in large part on their ability to increase total market demand sufficiently to offset falling prices. The size of that market, in turn, largely depends on whether less expensive dsDNA enables customers to reduce research and development costs and to create more products. The fundamental problem for the synthesis industry is that, however valuable sDNA is substantively to biological engineering in practice, the monetary value of that DNA is small compared with total development costs and has been falling, at times very rapidly, for decades. Falling prices limit both the maximum profit margin and the incentive to invest in new technology. Any new technology that does enter the market will inevitably drive competition, further depressing prices and margins. Going forward, productivity and prices are likely to display step changes resulting from the emergence of new technology and competition rather than display smooth long-term changes. Finally, given the relatively low barriers to entry for biological technologies and the consequent inevitable competition, it is worth asking whether centralized production is the future of the industry. As with printing documents, it may be that the economics of printing and using DNA favor distributed production, perhaps even a desktop model. There is no fundamental barrier to integrating any demonstrated synthesis and assembly technologies into a desktop gene printer. Ultimately, over the long term, a globally expanding customer base will ultimately determine how sDNA is produced and used. Regardless of how current technology specifically impacts supply and prices, that customer base is increasing, and it is likely that the trends displayed in Figures 1.1 and 1.2 will continue for many years to come.

References

- 1 Ellis, T. *et al.* (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.*, **3** (2), 109–118.
- 2 Baker, M. (2012) De novo genome assembly: what every biologist should know. *Nat. Methods*, **9**, 333–337.
- 3 Carlson, R. (2010) *Biology is Technology: The Promise, Peril, and New Business of Engineering Life*, Harvard University Press.
- 4 Loman, N. *et al.* (2010) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.*, **30**, 434–439.
- 5 Quail, M. *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

- 6 Liu, L. (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.*, Article ID 251364.
- 7 Carlson, R. (2004) The pace and proliferation of biological technologies. *Biosecur. Bioterror.*, **1** (3), 203–214.
- 8 Moore, G.E. (1965) Cramming more components onto integrated circuits. *Electronics*, 114–117.
- 9 Under new management, *The Economist*, 2 April 2009.
- 10 Hutcheson, G.D. (2009) The economic implications of Moore's Law, in *Into the Nano Era: Moore's Law Beyond Planar Silicon CMOS* (ed. H. Huff), Springer.
- 11 Wikipedia Boeing 787 Dreamliner, http://en.wikipedia.org/wiki/Boeing_787_Dreamliner (accessed 28 May 2013).
- 12 Wired Better Place Runs Out of Juice, Reportedly Plans Bankruptcy, <http://www.wired.com/autopia/2013/05/better-place-bankruptcy-report/> (accessed 10 January 2018).
- 13 Higgins, T. (2016) Tesla Generates a Profit. *The Wall Street Journal* (Oct. 26).
- 14 Reuters (2016) Tesla Just Raised \$1.5 Billion to Finance Model 3 Production (May 20).
- 15 See Carlson, R. (2013) How Competition Improves DNA Sequencing, http://www.synthesis.cc/synthesis/2013/04/how_competition_improves_reading_dna (accessed 3 January 2018).
- 16 National Institutes of Health DNA Sequencing Costs, <http://www.genome.gov/sequencingcosts/> (accessed 10 January 2018).
- 17 Carlson, R. (2009) The changing economics of DNA synthesis. *Nat. Biotechnol.*, **27** (12), 1091–1094.
- 18 Tian, J. *et al.* (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432** (7020), 23.
- 19 Carlson (2009, 2010); Paddon, C.J. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496** (7446).
- 20 See, for example, Integrated DNA Technologies Building Biological Factories for Renewable and Sustainable Products, IDT Decoded Newsletter, <http://www.idtdna.com/pages/decoded/decoded-articles/your-research/decoded/2013/01/18/building-biological-factories-for-renewable-and-sustainable-products> (accessed 10 January 2018).
- 21 Carlson, R. (2016) On DNA and Transistors http://www.synthesis.cc/synthesis/2016/03/on_dna_and_transistors (accessed 3 January 2018).
- 22 See, for example, Carlson, R. (2009) On the Demise of Codon Devices, http://www.synthesis.cc/synthesis/2009/04/on_the_demise_of_condon_devices (accessed 3 January 2018).
- 23 See, for example, National Research Council (1999) *Funding a Revolution: Government Support for Computing Research*, National Academies Press, Washington, DC.
- 24 (a) Tørring, T. *et al.* (2011) DNA origami: a quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.*, **12**, 5636; (b) Venkataraman, S. *et al.* (2007) An autonomous polymerization motor powered by DNA hybridization. *Nat. Nanotechnol.*, **2**, 490; (c) Gu, H. (2010) A proximity-based programmable DNA nanoscale assembly line. *Nature*, **465**, 202–205.