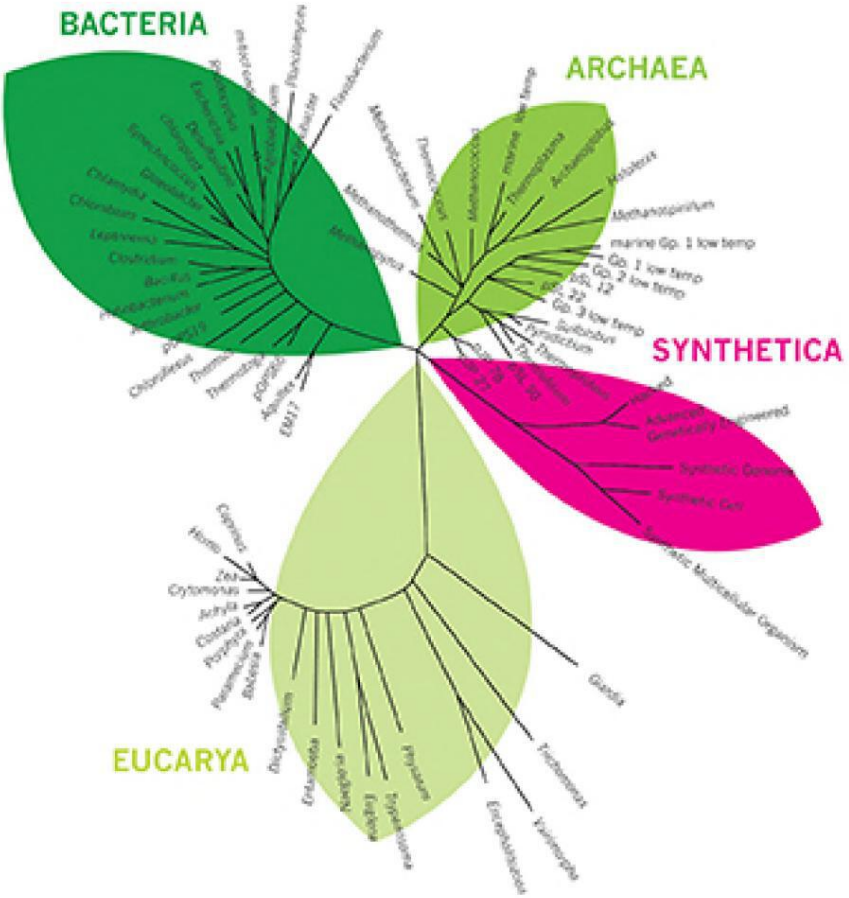


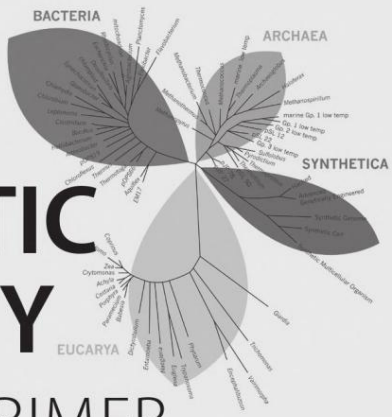
SYNTHETIC BIOLOGY

A PRIMER

Revised Edition

Baldwin • Bayer • Dickinson • Ellis
Freemont • Kitney • Polizzi • Stan

Imperial College Press



SYNTHETIC BIOLOGY

A PRIMER

Revised Edition

Geoff Baldwin
Travis Bayer
Robert Dickinson
Tom Ellis
Paul S Freemont
Richard I Kitney
Karen Polizzi
Guy-Bart Stan

Imperial College, UK

ICP

Imperial College Press

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Synthetic biology (Freemont)

Synthetic biology : a primer / [edited by] Paul Simon Freemont, Richard I. Kitney. --
Revised edition.

p. ; cm.

Preceded by: Synthetic biology / Geoff Baldwin ... [and 7 others]. [2012].

Includes bibliographical references and index.

ISBN 978-1-78326-878-8 (hardcover : alk. paper) -- ISBN 978-1-78326-879-5 (pbk. : alk. paper)

I. Freemont, Paul S., editor. II. Kitney, R. I., editor. III. Title.

[DNLM: 1. Synthetic Biology. QT 36]

TA164

660.6--dc23

2015026135

British Library Cataloguing-in-Publication Data

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

Typeset by Stallion Press

Email: enquiries@stallionpress.com

Printed in Singapore

Contents

This page intentionally left blank

List of Contributors

Editors and Contributors:

Paul Freemont (p.freemont@imperial.ac.uk)

Professor Paul Freemont is co-director and co-founder of the EPSRC Centre for Synthetic Biology and Innovation and the National UK Innovation and Knowledge Centre for Synthetic Biology (SynbiCITE) at Imperial College London. He is also currently the head of the Section of Structural Biology in the Department of Medicine at Imperial. His research interests span from understanding the molecular mechanisms of human diseases to the development of synthetic biology platform technologies and biosensors. The author of over 170 scientific publications, Professor Freemont has also been active in a number of public engagement activities on synthetic biology, including acting as co-curator of 'Grow Your Own' synthetic biology exhibition at Dublin Science Gallery (2013), co-organiser of Imperial Fringe Event 'Life as we know it' London (2013), a participant in the RCUK Public Dialogue on Synthetic Biology (2010) and the Royal Academy of Engineering public engagement on Synthetic Biology (2008). He has appeared regularly on radio and television broadcasts on the subject of synthetic biology and has successfully co-supervised with Professor Richard Kitney Imperial undergraduate iGEM teams since 2006 and acted as an iGEM judge.

Richard Kitney (r.kitney@imperial.ac.uk)

Richard Kitney is Professor of BioMedical Systems Engineering; Chairman of the Institute of Systems and Synthetic Biology; and Co-director of the EPSRC Centre for Synthetic Biology and Innovation and the UK's national industrial translation centre for synthetic biology (SynBiCITE). He was the co-chair of the joint Inquiry by The Royal Academy of Engineering and the Academy of Medical Sciences on Systems Biology. The report of the inquiry — *Systems Biology: a vision for engineering and medicine* — was published in February 2007. He was also Chair of The Royal Academy of Engineering Inquiry into Synthetic Biology, for which the report *Synthetic Biology: scope, applications and implications* was published in May 2009. Professor Kitney is a member of The Royal Society Working Party on Synthetic Biology. He has worked extensively in the United States and has been a Visiting Professor at MIT since 1991. He is a co-director of the Imperial College-MIT International Consortium for Medical Information Technology. Professor Kitney works extensively in Synthetic Biology and is heading Imperial College's initiative in this area with

Professor Paul Freemont. Together they have been responsible for three highly successful Imperial College iGEM teams.

Contributors:

Geoff Baldwin (g.baldwin@imperial.ac.uk)

Dr. Geoff Baldwin is a Reader in Biochemistry at Imperial College London. He has spent his career at the interface of the physical and life sciences. Having obtained an undergraduate degree in Chemistry, he then moved into biochemistry for his PhD, working on DNA-protein interactions. Dr. Baldwin has continued to cross boundaries with other disciplines and his quantitative approach to enzyme characterisation and knowledge of DNA mediated transactions have seen him move into engineering in the field of synthetic biology. This began with involvement supervising an iGEM team in 2008 and has continued as Dr Baldwin advised several of the very successful Imperial College iGEM teams. He has also been responsible for developing the training pipeline of synthetic biologists at Imperial College, having established the final year undergraduate module that is taught across the Life Science and Bioengineering Departments, and was Director of the MRes in Systems and Synthetic Biology until 2013. He has been involved with establishing the Centre for Synthetic Biology and Innovation where he has an active programme of research in DNA assembly, part characterization, and in vivo approaches to directed evolution for the creation of new bioparts.

Travis Bayer (tsbayer@gmail.com)

Dr Travis Bayer joined Imperial College in 2010 as an investigator at the Centre for Synthetic Biology and Innovation and as a lecturer in the Division of Molecular Biosciences, Department of Life Sciences. Dr Bayer moved to the University of Oxford in 2013 and then to San Francisco in 2014 where he co-founded Asilomar Bio Inc. Dr Bayer obtained his undergraduate degree at the University of Texas at Austin and his PhD at Caltech working with Professor Christina Smolke. He followed his PhD studies with a postdoctoral fellowship at UCSF working with Professor Chris Voigt on metabolic pathway engineering using synthetic biology.

Robert Dickinson (robert.dickinson@imperial.ac.uk)

Dr Robert Dickinson is a lecturer at the Department of Bio-engineering, Imperial College. He graduated with a degree in physics from Cambridge University and received his PhD in Biophysics, specifically ultrasound signal processing, from the University of London. Dr Dickinson has extensive experience in medical imaging, in both hospital and industrial environments. He worked on MRI coil development and system integration at Picker International Ltd and ultrasound imaging in a small start-up company where he developed a sub-1mm intravascular ultrasound imaging catheter. This device is used for imaging coronary arteries, and some images are shown below. He has substantial experience in the bio-compatibility and other issues of invasive medical devices, together with commercialisation and IP transfer. He has filed over 12 patents, and has CE marked

a number of medical devices. He has also worked with Emcision Ltd on their range of electrosurgical devices.

Tom Ellis (t.ellis@imperial.ac.uk)

Dr Tom Ellis is a senior lecturer in the Department of Bioengineering at Imperial College London and leads development of foundational synthetic biology at the Centre for Synthetic Biology and Innovation. His research focuses on advancing the tools and methods to assemble and rewire gene expression and to enable the future design and assembly of synthetic microbial genomes. This research is being applied to a variety of projects for the cellular synthesis of therapeutics, biofuels and biomaterials. Ellis was an undergraduate at Oxford University and received his PhD from the University of Cambridge. Following two years at a drug development start-up, he became a synthetic biology postdoctoral fellow with Professor James J. Collins at Boston University. He now leads the UK contribution to the Sc2.0 Project, an international project to rewrite, synthesise and assemble a synthetic version of the yeast genome. At Imperial College London, he also co-organises undergraduate synthetic biology teaching and supervision of the university's iGEM teams.

Karen Polizzi (k.polizzi@imperial.ac.uk)

Dr Karen Polizzi joined Imperial College in September 2008 as an RCUK Fellow in Biopharmaceutical Processing within the Division of Molecular Biosciences, Department of Life Sciences. She obtained her PhD in Chemical and Biomolecular Engineering at the Georgia Institute of Technology working with Professor Andreas Bommarius on protein engineering of biocatalysis. Dr Polizzi's laboratory is interested in applying synthetic biology to bioprocess engineering with a particular focus on the production of therapeutic proteins and the use of *in vivo* biosensors for developing and controlling metabolic pathways related to protein production and glycosylation. She is also interested in enzymatic processing related to the production of therapeutics.

Guy-Bart Stan (g.stan@imperial.ac.uk)

Dr Guy-Bart Stan is a Reader (Associate Professor) in Engineering Design for Synthetic Biology and is the head of the Control Engineering Synthetic Biology group at the Centre for Synthetic Biology and Innovation (CSynBI) and the Department of Bioengineering of Imperial College London. He is the recipient of the very prestigious UK EPSRC Fellowship for Growth in Synthetic Biology. He is also a PI member of CSynBI, the Flowers Consortium, and SynBiCITE. Dr Stan teaches various courses at the intersection of systems and control engineering and biological systems analysis, design, and control. He obtained his PhD in Applied Sciences (nonlinear dynamical systems and control) from the University of Liège, Belgium, in 2005. Dr Stan then worked as postdoctoral research associate in the Control Group at Department of Engineering, University of Cambridge before joining Imperial College in 2009. Dr Stan's main research interests are in the areas of Nonlinear Dynamical Systems Design, Analysis and Control, Synthetic Biology, and Systems Biology with a particular focus on the study of core engineering design principles of biological systems and their applications for synthetic biology, biotechnology, and cell-based medicine.

This page intentionally left blank

Preface

Synthetic Biology is an exciting and rapidly evolving interdisciplinary research field which aims to provide a systematic framework for the engineering of biological systems and cells at the genetic level. What makes synthetic biology so unique? In part, the application of engineering principles and systematic design tools involving computational modeling, modular parts and standardised measurements. These principles are completely novel in the field of reprogramming cellular systems at the genetic level for desired functional outputs. This powerful engineering approach has the potential to transform both our fundamental understanding of biological systems and our ability to manipulate them for societal benefit. Why now? The origins of Synthetic Biology can be traced back to the early 1960s when developments in our understanding of the genetic code in the form of DNA led to the central dogma of molecular biology: DNA encodes RNA, which encodes protein macromolecules. This powerful dogma, combined with techniques first developed in the 1970s that allowed the manipulation, transfer and cloning of DNA, has underpinned much of our understanding of molecular biosciences and cell biology and has quickly led to the genome revolution which we are currently experiencing. This rapid growth in molecular and genomic understanding was enabled by the parallel developments of new technologies and capabilities including increasing computing power and the establishment of sophisticated information systems, and the ability to rapidly sequence DNA. Together all of these components culminated in 2000 to a landmark moment in human scientific understanding with the completion of the human genome sequence. In 2015, we now have complete genome sequences for nearly every major class of organism on Earth. For the first time, we have complete lists of the basic components encoded at the genetic level that constitute living systems, accessible from any web browser in the world including a mobile phone. This rapid accumulation of biological information is correlated with the increasingly interdisciplinary nature of biological research. One unique aspect of Synthetic Biology is therefore the fusion of engineers and engineering principles with molecular cellular and systems biology thinking.

One of the driving forces of Synthetic Biology is the goal of making systematic design and construction of novel biological systems and cells more easy, repeatable, and predictable. The natural outcome of this will be to accelerate the development of novel applications in biotechnology. The conceptual engineering framework that underpins Synthetic Biology research translates the techniques used in engineering design like specifications, mathematical modeling, standardisation, prototyping and the design of

biological systems at the genetic level. One could view this as utilising the genetic code as a truly programmable material, which can result in novel and robust genetic designs. The idea for this textbook came from the establishment of a Master's training program at Imperial College London within the Centre of Synthetic Biology and Innovation. As our new students began their interdisciplinary training in Synthetic Biology we realised there was no source textbook we could recommend — which has now led to this primer text.

Synthetic Biology — A Primer (*Revised Edition*) aims to give a broad overview of Synthetic Biology and the foundational concepts on which it is built. It will be primarily of interest to final year undergraduates, postgraduates, and established researchers who are interested in learning more about this exciting new field. The book introduces readers to fundamental concepts in molecular biology and engineering and then explores the two major themes for synthetic biology, namely 'bottom-up' and 'top-down' engineering approaches. 'Top-down' engineering utilises a conceptual framework of engineering and systematic design to build new biological systems by integrating robustly characterised biological parts into an existing system through the use of extensive mathematical modeling. The 'bottom-up' approach involves the design and building of synthetic protocells using basic chemical and biochemical building blocks from scratch. Exemplars of cutting-edge applications designed using synthetic biology principles are presented including the production of novel biofuels from renewable feedstocks, microbial synthesis of pharmaceuticals and fine chemicals, and the design and implementation of biosensors to detect infections and environmental waste. The book also uses the Internationally Genetically Engineered Machine (iGEM) competition to illustrate the power of Synthetic Biology as an innovative research and training science. Finally, the primer includes a chapter on the ethical, legal and societal issues surrounding synthetic biology, illustrating the legitimate integration of social sciences in synthetic biology research. This chapter is a joint contribution from our social science colleagues at Kings College London, including postgraduate students and senior researchers, and thus all the authors are cited at the start of this chapter.

In this Revised Edition we have revised the text in line with the accelerating nature of the synthetic biology field without altering the basic concepts, which we believe readers who are not aware of Synthetic Biology need to engage with the field. We hope that students and professional researchers will find this revised text useful as an introduction to Synthetic Biology and if readers become inspired as we have by the potential of Synthetic Biology to tackle the pressing problems of planet Earth via biological solutions combined with a responsible overview then we will be delighted. The enthusiasm, energy, excitement and commitment shown by our iGEM team, junior researchers and students, colleagues in the Royal College of Art, Kings College, University College and Imperial College London as well as the growing international research community suggests that Synthetic Biology is a truly transformative field that crosses many disciplinary boundaries as well as international borders.

Professors Paul Freemont and Richard Kitney
Centre for Synthetic Biology and Innovation, Imperial College London
Summer 2015.

CHAPTER 1

Introduction to Biology

1.1 Introduction

This introduction is intended to provide a brief overview of the basic principles that lie at the heart of biology. It is intended for engineers and physical scientists who do not have any knowledge of biology but who are interested in synthetic biology. Having an understanding of the basic principles will aid the physical scientist and engineer to model biological systems and also to communicate with biologists. There are of course many excellent textbooks dedicated to detailed descriptions of the molecular processes that occur within biology, and readers who wish to extend their knowledge beyond this brief survey are encouraged to seek out more details in the relevant texts.

There are two basic concepts that you need to understand if you wish to be able to engineer biological systems: how information flows in biological systems and how this information flow is controlled. With an understanding of these concepts one can, in principle, apply engineering principles to the design and building of new biological systems: what we call synthetic biology.

Biology is, of course, highly complex and there are important differences that distinguish it from other engineering disciplines. Firstly, biology is not programmed on a printed circuit board, so interactions cannot be programmed by their physical position; rather interactions are based on interactions between molecules that occur in the complex milieu of the cell. Secondly, biology is subject to natural selection, so that modifications which are deleterious to the cell will be selected against and competed out of the population. These evolutionary pressures are not applicable when building an aircraft, and so new definitions of robustness are relevant to biology. Other concepts such as complexity and emergent behaviour may be familiar to engineers, but one must be aware of how they can arise in biology and what their effects may be.

1.2 Information Storage in Biology

1.2.1 *DNA structure*

It is widely known by even non-specialists that DNA (deoxyribonucleic acid) is the molecule that stores the information within biological systems. The double-helical structure of the

DNA molecule has also entered the wider consciousness and provides a totemic symbol of the genomic age in which we live. To fully appreciate how this molecule functions as a dynamic information store, we must explore its structure in more detail.

The double-helical structure of DNA arises because it is composed of two separate strands. Each strand is a long molecule formed by the linking together of the individual building blocks, the nucleotides. Each nucleotide has three important parts: the phosphate, the deoxyribose sugar and the base (Fig. 1.1). The phosphate and sugar always remain the same, but there are four different bases: A, G, C and T (adenine, guanine, cytosine and thymine). The phosphate and the sugar link together to form the backbone of each DNA strand, while the bases decorate the backbone and provide variability. It is the sequence of these nucleotides that provides the basis of information coding in DNA.

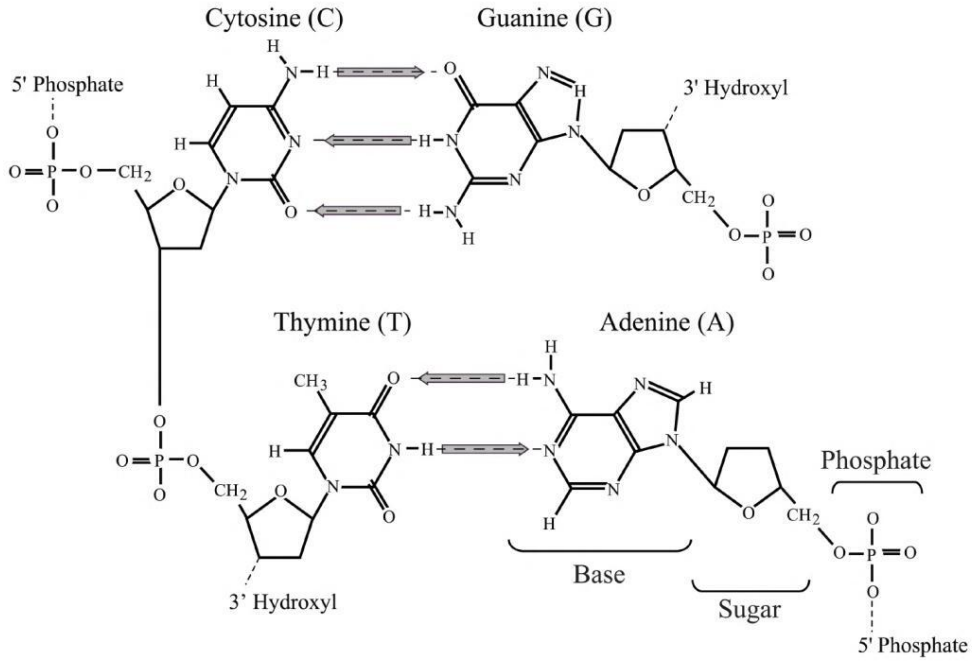
A key feature of DNA, which leads to the double-helix structure, is how the two strands interact. Each of the four bases can only pair with one of the other bases; A can only pair with T, and G with C. A cannot pair with G, C or A and so on. This complementarity means that the sequence of one strand defines the sequence of the other: a G in one strand means that the other strand will always have a C in that position, etc. The consequence of this, from the cell's perspective, is that it will always have two copies of the information encoded by the DNA sequence. This is important because if one strand is damaged, the other can direct the repair and no information is lost.

The complementarity arises through hydrogen bond interactions between the bases. Looking at the structure of the paired bases reveals that the positions of these hydrogen bonds occur at the same relative positions between the functional groups of the different bases (Fig. 1.1). However, the pattern of hydrogen bond donors (the group with the H-atom) and acceptors (the interacting partner with no H) is different, so that if the bases are swapped around, H-bonds cannot be made (H-bonds cannot form between 2 donors or 2 acceptors). There are 2 H-bonds between A and T, and 3 H-bonds between G and C. This means that GC base pairs are more stable and thus harder to separate.

As two complementary strands of DNA come together the base pairs will interact with each other to form hydrogen bonds. The planar structure of the aromatic bases means that they will 'stack' on top of each other, each phosphate on the backbone has a negative charge and they are thus repelled from each other, while the sugar provides a linking group of a fixed length. The attraction of the stacking bases in each strand is balanced by the repulsion of the phosphates within the limit of the deoxyribose link. The net result is that each base is offset by around 34° from the previous one and the two strands twist into a double helix (Fig. 1.1).

1.2.2 *DNA replication*

The ability to reproduce is one of the key features of life and DNA replication lies at the heart of this process. The complementarity of the two DNA strands means that it is possible for the strands to separate and for each one to act as a template in the synthesis of a new DNA strand, so that two new daughter DNA helices are produced. Each one of these daughter DNA molecules contains one of the strands of the parental DNA double helix.



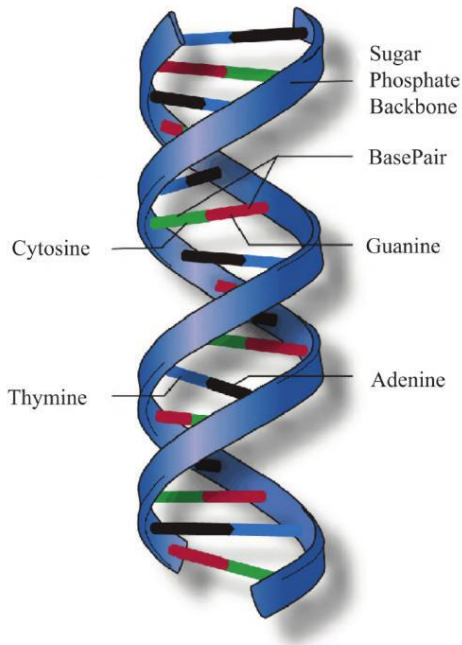


Fig. 1.1 The chemical structure of the nucleotide base pairs and the double-helical structure of DNA. (Courtesy: National Human Genome Research Institute)

DNA is synthesised by enzymes called DNA polymerases from component deoxyribonucleotide triphosphates (dNTPs). The triphosphate is a chemical group that effectively acts as an energy source for the reaction. This is a common feature in nature, and a different nucleotide — ATP — functions as a near universal energy source within living systems. DNA polymerases are able to select the correct nucleotide by its ability to base pair with the template strand. It then catalyses its incorporation into the growing strand, releasing a diphosphate (Fig. 1.2).

Another key feature of DNA structure is that the strands have a polarity as the chemistry of the backbone means that they are not the same at each end; they are denoted as the 5′ and 3′ ends. Double-stranded DNA always forms an anti-parallel configuration. This has very important implications for DNA replication, since DNA polymerases can only add on to the 3′ end. Furthermore, DNA polymerases can only extend an existing chain; they cannot initiate synthesis on single-stranded DNA. In a cell, replication is initiated by an enzyme called primase, which synthesises a short RNA primer *in situ* so that DNA polymerase can begin the process. This also enables a greater degree of control on DNA replication. *In vitro*, we also use short sequences to initiate DNA synthesis, but we use synthetic DNA oligonucleotides. It turns out that this is a very useful feature of DNA polymerases as it means we can direct the synthesis of DNA very precisely.

1.2.3 *PCR*

The polymerase chain reaction (PCR) is one of the most useful techniques in molecular biology and it is worth studying since understanding it requires understanding all of the key features of DNA structure and replication. PCR is a method for the exponential amplification of DNA. Its value lies in its speed and specificity; one can readily make large quantities of a specific DNA fragment which have been isolated from vanishingly small quantities of the original template sequence. It relies on DNA polymerases that have been isolated from bacteria that live in hot springs or ocean sea vents; these extremophile organisms can live in temperatures in excess of 100°C. As a consequence their DNA polymerases are also stable at high temperatures, a key feature that is required due to the thermal cycling of the reaction.

The specificity comes from the oligonucleotides that are used to prime the DNA polymerase. Two primers are required, one for each end of the DNA. The reaction operates through temperature cycling: the first stage requires heat denaturation of the DNA; as it is then cooled the primers will anneal to their complementary sequence on the DNA template. DNA synthesis is then directed from the 3′ end of the primers and the reaction is heated to 72°C, the optimal temperature of the thermostable DNA polymerases. The reaction is then heated to 95°C to melt the newly synthesised double-stranded DNA back to single strands, allowing for another round of priming and synthesis. The doubling nature of the reaction means that the specific DNA fragments increase in an exponential order so after 25–30 cycles a very large quantity of DNA has been synthesised (Fig. 1.3 and Appendix 1).

The general applicability of PCR has increased as new DNA polymerases have been isolated and engineered. The first thermostable DNA polymerase, *Taq*, had relatively

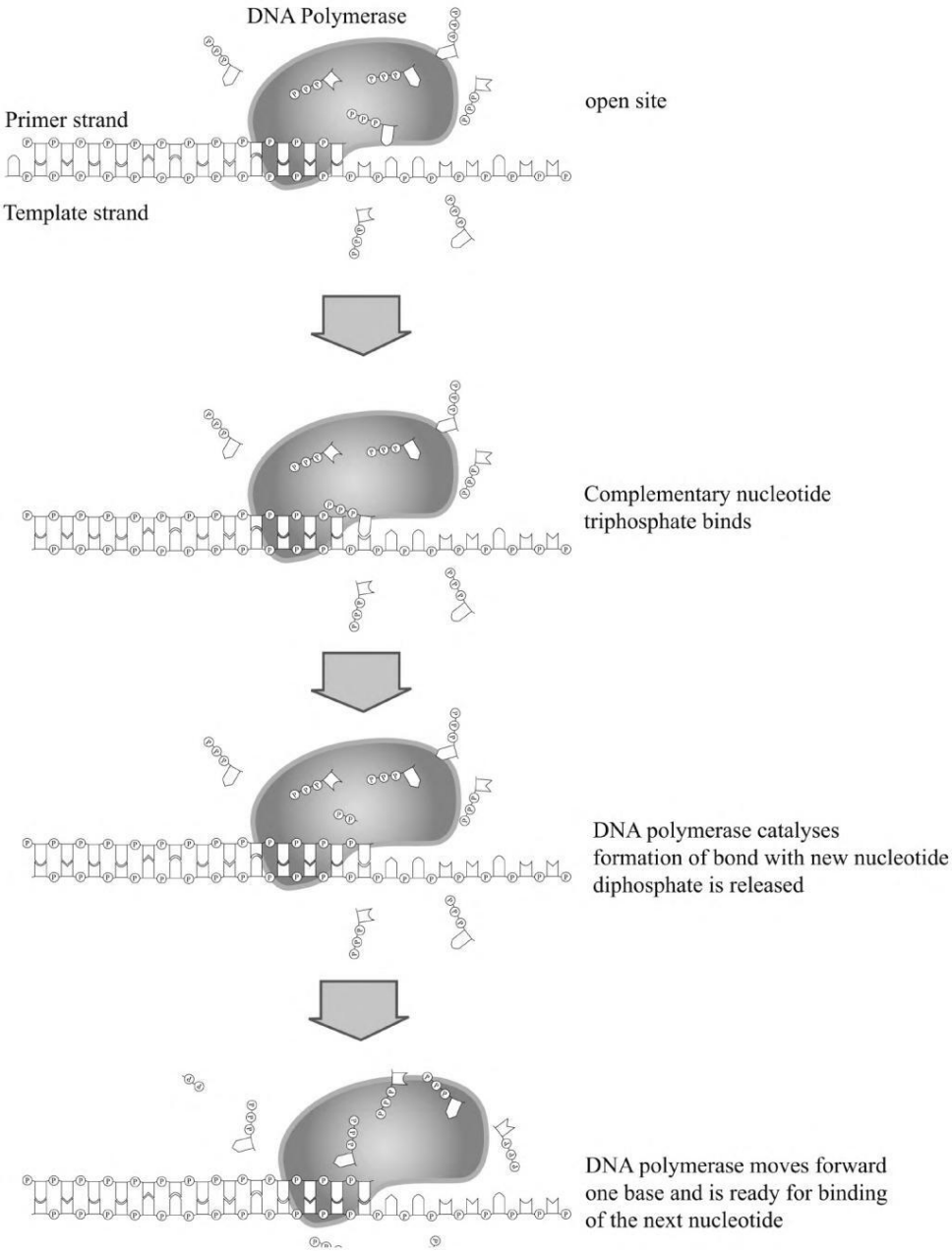


Fig. 1.2 DNA replication and the role of DNA polymerases in synthesising DNA.

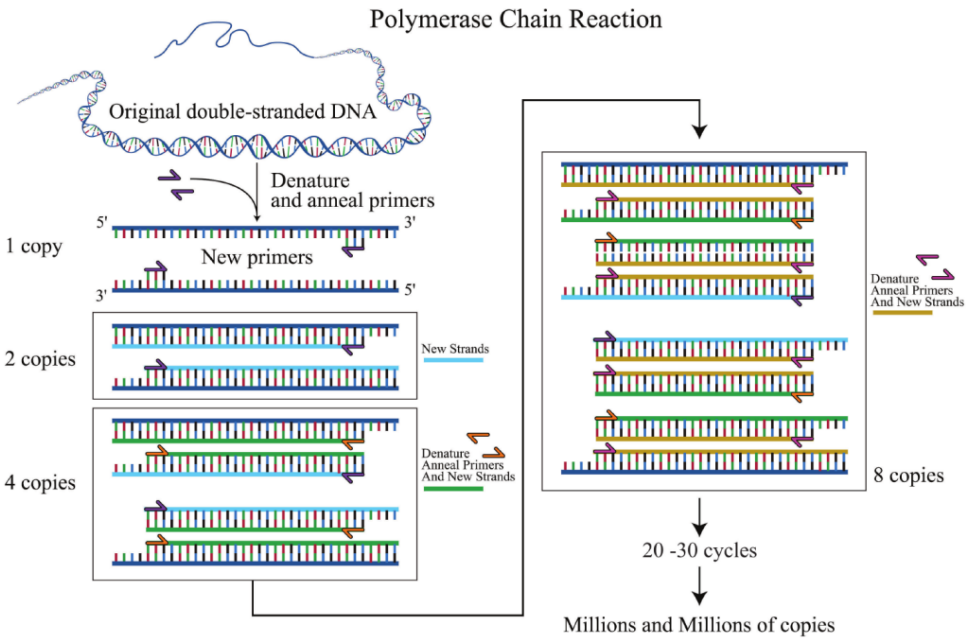


Fig. 1.3 The polymerase chain reaction. (Courtesy: National Human Genome Research Institute)

low fidelity, so amplification of sequences frequently led to mutations in the products. New DNA polymerases have been isolated from other sources that contain proofreading domains (a 3′–5′ exonuclease activity that is able to remove mismatched bases), which has significantly increased the accuracy with which DNA can be amplified. Other polymerases such as Phusion™ have been engineered to contain a DNA binding domain that increases the processivity of the polymerase, increasing the length of DNA that can readily be made by PCR. This increased accuracy means that PCR can now be adapted to many more cloning techniques that require amplification of the vector backbone (Ellis, Adie and Baldwin, 2011).

1.3 Information Flow in Biology

DNA is the information store, but the majority of functions within the cell are performed by proteins. The question thus arises of how information leads to function. The flow of information within biological systems forms what is known as the 'central dogma' (Fig. 1.4). Information is stored in the form of DNA. When required, the message within DNA is transcribed into an intermediate messenger molecule, RNA, before being translated into the final product, protein. The fundamental unit of hereditary information is known as a gene. Most genes are a stretch of DNA that codes for a protein, although some genes will produce functional RNA molecules rather than proteins. A gene should also be considered to include the regulatory elements required to control it. The basic structure of a gene is depicted in

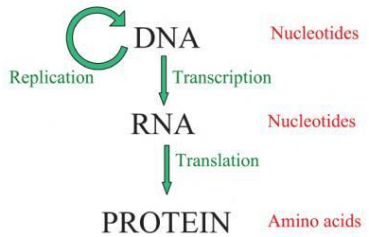


Fig. 1.4 The central dogma of molecular biology.

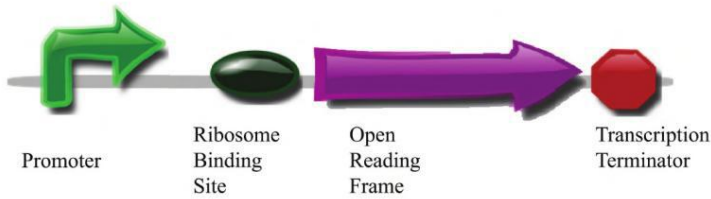


Fig. 1.5 The basic structure of a protein-coding gene.

Fig. 1.5. When a gene is turned on, the protein (or other functional molecule) is produced or expressed; when it is turned off it is not produced. More detail on these control mechanisms is provided in the next section.

The messenger molecule is composed of nucleotides similar to those found in DNA, called RNA (ribonucleic acid). The chemical structures of these two molecules are almost identical, the key difference being the 2'OH on the sugar moiety of RNA; the lack of this group in DNA is what gives the deoxy-designation. The presence of this extra OH means that RNA can form more complex structures; it is also less stable. In addition, RNA uses U (uracil) in place of T, the difference being the presence of a methyl group in thymine.

The conserved nature of the bases between RNA and DNA means that they can base pair in an identical way to DNA bases. This feature is used when transcribing the information in DNA into the messenger RNA (mRNA) molecule. During transcription a stretch of DNA known as a gene is transcribed into mRNA. That is, a mRNA molecule is synthesised *in situ* based on complementary base pairing between the RNA and DNA nucleotides. The key feature of a gene is that the sequence of DNA bases code for a protein. The complementary base pairing between RNA and DNA means that this code is retained in mRNA.

Once mRNA has been synthesised it fulfills its role as a messenger by taking the 'message' to the ribosome. This is a very large macromolecular complex composed of proteins and RNA molecules and it is here that the message is translated from nucleotides into amino acids, the building blocks of proteins. The end result is a protein where the amino acids have been put together in a specific sequence designated by the DNA sequence. Thus DNA codes for proteins.

RNA codon table

1st position	2nd position U	2nd position C	2nd position A	2nd position G	3rd position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G
	Amino Acids				

Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp:Aspartic acid
Cys:Cysteine
Gln: Glutamine
Glu: Glutamic acid
Gly: Glycine
His: Histidine
Ile: Isoleucine
Leu: Leucine
Lys: Lysine
Met: Methionine
Phe: Phenylalanine
Pro: Proline
Ser: Serine
Thr: Threonine
Trp: Tryptophan
Tyr: Tyrosisne
Val: Valine

Fig. 1.6 The genetic code table. Amino acids are coded by three nucleotides (codons). The information is carried to the ribosome, which synthesises the protein encoded by the messenger RNA (mRNA). The codons are therefore usually shown in RNA nucleotides, where Uracil (U) is used in place of Thymine (T). (Courtesy: National Human Genome Research Institute)

1.3.1 *The genetic code*

The sequence of bases A, G, C and T provides the basis for information storage within DNA. This DNA sequence codes for protein sequence, but proteins are composed of 20 different amino acids, therefore the coding from nucleotides to amino acids is not one-to-one. It is, in fact, a triplet code, so three nucleotides code for a single amino acid (Fig. 1.6). The triplet code means that DNA sequence can be read in three different 'frames' on each strand (six in total; Fig. 1.7).

The triplet code means that there is redundancy, and all but two amino acids are coded for by multiple three nucleotide sequences (codons; Fig. 1.6). The exceptions are tryptophan and methionine, which have unique codons and this reflects their reduced frequency of occurrence in proteins.

Methionine is also the initiating amino acid in all bacterial proteins. The start signal for protein synthesis comprises the position where the ribosome binds the mRNA (ribosome binding site, RBS or Shine-Dalgarno sequence) and the first ATG codon, which must be close to the RBS. Reading the first codon establishes the reading frame of the mRNA, so having a unique codon for the initiating amino acid reduces ambiguity.

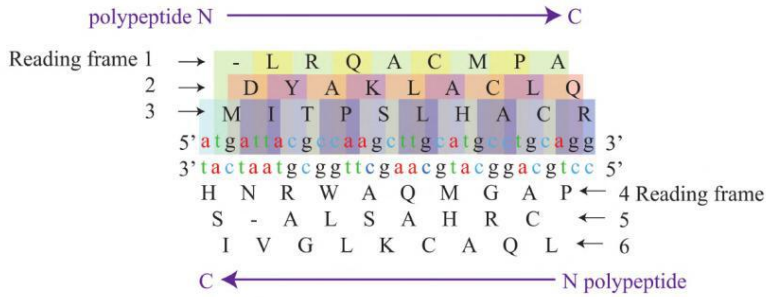


Fig. 1.7 A short section of double-stranded DNA. There are six possible coding frames on double-stranded DNA. The top strand codes for a protein. The beginning of the protein sequence is the ATG Methionine codon in reading frame 3. By shifting the reading frame by 1 bp steps, it can be seen that there are three other possible reading frames for the same piece of DNA. In reality these do not code for a protein because they do not satisfy the requirements of an open reading frame (they are not preceded by a promoter, are not uninterrupted by stop codons and do not finish with a stop codon). The other DNA strand also has three other possible reading frames. As the polarity of the DNA is reversed on the partner strand, these read in the opposite direction.

The degeneracy in the genetic code also leads to differences between different organisms. Each organism has evolved within its own niche and its translation machinery has typically developed a narrower set of codons that are regularly used, or are used under particular growth conditions. Since this codon usage varies from organism to organism, a gene that originates from one organism and is transcribed in another may be poorly translated as the codons of the gene are not regularly used in the new host.

The process by which the nucleotide sequence of the mRNA molecule is translated into protein (shown in Fig. 1.8) again requires intermediate molecules, this time to deconvolute the nucleotide sequence into amino acid sequence. These intermediaries are called transfer RNA molecules (tRNA) because they transfer amino acids onto the growing protein. As the name suggests, these molecules are composed of RNA, but are linked to amino acids. Cells contain different types of tRNA molecules, and each type can only be linked with a single amino acid. This highly specific reaction is catalysed by tRNA synthetase enzymes.

Because tRNAs are composed of nucleotides, they are able to base pair with the nucleotides of the mRNA, in the same way that DNA bases pair. The tRNAs are structured in such a way that a triplet of RNA bases is presented to the mRNA template, thus reading the genetic code and presenting the correct amino acid to the ribosome for incorporation.

1.3.2 *Proteins*

The great majority of functional and structural roles within organisms are carried out by proteins. They have a truly remarkable repertoire of properties and functions. They can form structural materials such as spider silk and keratin, which makes up hair and nails; they can act as motors and transduce chemical energy into mechanical energy, like myosin in muscles; there are many enzymes that catalyse highly specific chemical reactions and there are a large number of regulatory proteins. Regulation can occur at many levels from