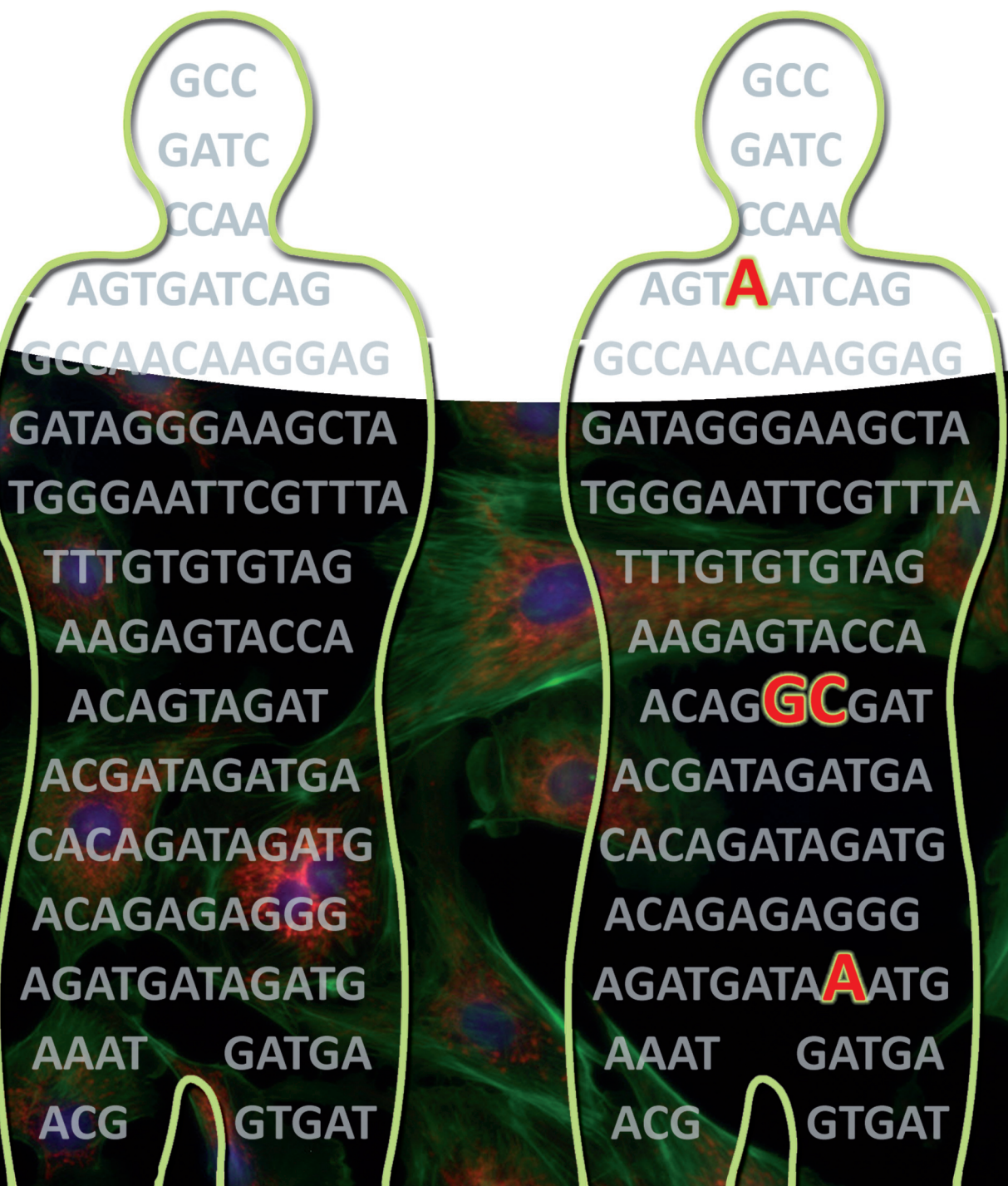


Jens Kurreck and Cy A. Stein

Molecular Medicine

An Introduction



Jens Kurreck and Cy Aaron Stein

Molecular Medicine

Related Titles

Meyers, Robert A. (eds.)

Translational Medicine

Cancer

2015

Print ISBN: 978-3-527-33569-5

Meyers, Robert A. (eds.)

Synthetic Biology

2015

Print ISBN: 978-3-527-33482-7

Vertes, Alain / Qureshi, Nasib / Caplan, Arnold I. / Babiss, Lee (eds.)

Stem Cells in Regenerative Medicine

Science, Regulation and Business Strategies

2015

Print ISBN: 978-1-119-97139-9

Pelengaris, S., Khan, M. (eds.)

The Molecular Biology of Cancer

A Bridge from Bench to Bedside

2 Edition

2013

Print ISBN: 978-1-118-00881-2; also available in electronic formats

Karp, G.

Cell and Molecular Biology

7 Edition

2013

Print ISBN: 978-1-118-20673-7; also available in electronic formats

Dickenson, J., Lloyd Mills, C., Freeman, F., Thode, C., Sivasubramaniam, S.

Molecular Pharmacology - From DNA to DrugDiscovery

2012

Print ISBN: 978-0-470-68443-6; also available in electronic formats

MacPherson, G., Austyn, J.

Exploring Immunology

Concepts and Evidence

2012

Print ISBN: 978-3-527-32412-5; also available in electronic formats

Acheson, N.H.

Fundamentals of Molecular Virology, Second Edition

2 Edition

2011

Print ISBN: 978-0-470-90059-8; also available in electronic formats

Pasternak, J.J.

An Introduction to Human Molecular Genetics

Mechanisms of Inherited Diseases, Second Edition

2 Edition

2005

Print ISBN: 978-0-471-47426-5; also available in electronic formats

Jens Kurreck and Cy Aaron Stein

Molecular Medicine

An Introduction

WILEY-VCH
Verlag GmbH & Co. KGaA

Authors

Jens Kurreck

Berlin University of Technology
Institute of Biotechnology
Department of Applied Biochemistry, TIB 4/3-2
Gustav-Meyer-Allee 25
13355 Berlin
Germany

Cy Aaron Stein

City of Hope Medical Center
Department of Medical Oncology and Experimental Therapeutics
1500 E. Duarte Road
Duarte, CA 91010
USA

Illustrator

Anke Wagner

Berlin University of Technology
Institute of Biotechnology
Department of Bioprocess Engineering, ACK24
Ackerstr. 76
13355 Berlin
Germany

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>>.

© 2016 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: 978-3-527-33189-5

ePDF ISBN: 978-3-527-67508-1

ePub ISBN: 978-3-527-67507-4

Mobi ISBN: 978-3-527-67511-1

Cover Design Adam Design, Weinheim

Typesetting Thomson Digital, Noida, India

Printed on acid-free paper

*Dedicated to Ferdinand Hucho and John J. Rossi for their contributions to the
life sciences, and most importantly, for being extraordinarily well-respected,
decent, honorable human beings.*

Contents

Preface xiii

1

Introduction 1

1.1 The Basics of Molecular Medicine 2

1.1.1 Topics of Molecular Medicine 2

1.1.2 Stages of Drug Development 3

1.2 The Human Cell 4

1.2.1 Organelles 4

1.2.1.1 The Nucleus 4

1.2.1.2 Mitochondria 6

1.2.1.3 Endoplasmic Reticulum and Golgi Apparatus 7

1.2.1.4 Peroxisome and Lysosome 8

1.2.2 Cell Cycle 8

1.2.3 Apoptosis 9

1.3 DNA Replication and Gene Expression 10

1.3.1 DNA Replication 11

1.3.2 Mutations 13

1.3.3 Transcription 14

1.3.4 Epigenetic Regulation of Gene Expression 19

1.3.5 Translation 21

1.3.6 Protein Degradation 24

1.4 Biological Communication 25

1.4.1 Neurotransmitters 26

1.4.2 Hormones 27

1.4.3 Signal Transduction 28

1.5 The Immune System 30

1.5.1 The Innate Immune System 30

1.5.1.1 The Complement System 31

1.5.2 The Adaptive Immune System 33

1.5.2.1 Cellular Immunity 33

1.5.2.2 Humoral Immunity 34

References 36

2

Methods in Molecular Medicine 37

2.1 DNA Microarrays 38

2.2 Quantitative Polymerase Chain Reaction 40

2.3 Next-Generation Sequencing 45

2.4 Animal Models in Biomedical Research 51

2.5 Additional Methods 56

2.5.1 Fluorescence Microscopy 56

2.5.2 Flow Cytometry and Fluorescence-Activated Cell Sorting 58

2.5.3 Surface Plasmon Resonance 59

References 59

3

Genetic Disorders 61

3.1 Single-Gene Disorders 62

3.1.1 Autosomal Dominant Disorders 64

3.1.1.1 Familial Hypercholesterolemia 65

3.1.1.2 Polycystic Kidney Disease 67

3.1.1.3 Marfan's Syndrome 67

3.1.1.4 Huntington's Disease 68

3.1.2 Autosomal Recessive Disorders 69

3.1.2.1 Cystic Fibrosis 70

3.1.2.2 Tay-Sachs Disease 71

3.1.2.3 Phenylketonuria 72

3.1.2.4 Xeroderma Pigmentosum 73

3.1.3 X-Linked Recessive Disorders 74

3.1.3.1 Red-Green Color Blindness 75

3.1.3.2 Duchenne and Becker Muscular Dystrophy 75

3.1.4 Mitochondriopathies 77

3.2 Polygenic Disorders 80

3.2.1 Asthma 80

3.2.2 Diabetes Mellitus 81

References 83

4

Molecular Oncology 85**4.1 Molecular Biology of Breast Cancer and Its Clinical Implications 88**

- 4.1.1 Intrinsic Subtypes of Breast Cancer 88
 - 4.1.1.1 Luminal 88
 - 4.1.1.2 Subclassification of TNBC 89
- 4.1.2 Molecular Profiling of Breast Cancer 89
- 4.1.3 Signaling Pathways 89
 - 4.1.3.1 The Role of the Estrogen Pathway in Breast Cancer 90
 - 4.1.3.2 Endocrine Therapy Resistance 90
 - 4.1.3.3 The mTOR/PI3K Pathway and Endocrine Resistance 90
 - 4.1.3.4 The CDK 4/6 Pathway 90
 - 4.1.3.5 HER2 Pathway and HER2 Targeted Therapy 91
- 4.1.4 Angiogenesis Pathway 92
 - 4.1.4.1 PARP Inhibitors 92
- 4.1.5 Other Biological Therapies/Approaches 93

4.2 Lung Cancer 93

- 4.2.1 Genetic Alterations in Non-Small Cell Lung Cancer 93
 - 4.2.1.1 Epidermal Growth Factor Receptor 93
 - 4.2.1.2 Anaplastic Lymphoma Kinase 94
 - 4.2.1.3 Kirsten Rat Sarcoma (KRAS) 94
 - 4.2.1.4 The Proto-Oncogene ROS1 95
 - 4.2.1.5 The Proto-Oncogene BRAF 95
 - 4.2.1.6 The Human Epidermal Growth Factor Receptor 2 (HER2) 95
 - 4.2.1.7 The RET Proto-Oncogene 95
 - 4.2.1.8 The MET Proto-Oncogene 95
 - 4.2.1.9 Phosphatidylinositol-3-Kinase (PI3K) 95
 - 4.2.1.10 Immune Checkpoint Inhibition 96

4.3 Hepatocellular Carcinoma 96

- 4.3.1 Risk Factors for Hepatocellular Carcinoma 96
- 4.3.2 Molecular Biology of Hepatocellular Carcinoma 97
- 4.3.3 Development of Sorafenib for the Treatment of Hepatocellular Carcinoma 97
- 4.3.4 Complexity of Cancer 98

4.4 Molecular Biology of Colorectal Cancer and Its Clinical Implications 99

- 4.4.1 Colorectal Cancer Carcinogenesis 99
 - 4.4.1.1 Chromosomal Instability Pathway 100
 - 4.4.1.2 Microsatellite Instability Pathway 100
 - 4.4.1.3 CpG Island Methylator Phenotype (CIMP) Pathway 101
- 4.4.2 Hereditary Colorectal Cancers 101

- 4.4.2.1 Familial Adenomatous Polyposis 101

- 4.4.2.2 Management of FAP Patients 101

- 4.4.2.3 Hereditary Non-Polyposis Colorectal Cancer 102

- 4.4.2.4 Management of HNPCC-Associated Germline Mutation Carriers 103

- 4.4.2.5 MUTYH-Associated Colorectal Cancer 103

- 4.4.2.6 Management of MAP Patients 103

- 4.4.3 Clinical Impact of Molecular Markers on the Management of Colorectal Cancer 103

- 4.4.3.1 MSI-H Status and Colorectal Cancer 103

- 4.4.3.2 Epidermal Growth Factor Receptor Pathway Targeting and Colorectal Cancer 103

- 4.4.3.3 RAS Mutations and Response to Anti-EGFR Therapy 104

- 4.4.3.4 BRAF Mutations and Colorectal Cancer 104

4.5 Molecular Biology of Renal Cell Carcinoma 105

- 4.5.1 Biology of Clear Cell Renal Cell Carcinoma 105
- 4.5.2 Approved Drugs for the Treatment of Clear Cell Renal Cell Carcinoma 106
- 4.5.3 Investigational Approaches for the Treatment of Clear Cell Renal Cell Carcinoma 107
- 4.5.4 Biology and Treatment of Papillary Renal Cell Carcinoma 108
- 4.5.5 Biology and Treatment of Chromophobe Renal Cell Carcinoma 108
- 4.5.6 Further Subtypes of Renal Cell Carcinoma 108

4.6 Molecular Biology of Prostate Cancer 109

- 4.6.1 Genes Associated with Hereditary Prostate Cancer 109
- 4.6.2 Tumor Suppressor Genes in Sporadic Prostate Cancer 110
- 4.6.3 Oncogenes 111

4.7 Molecular Biology of Hematological Malignancies 114

- 4.7.1 The Importance of Cytogenetics in Diagnosis and Treatment Decision-Making 115
- 4.7.2 Recognition of a Genetic Basis for the Hematological Malignancies 117
- 4.7.3 Targeted Therapeutics for Hematological Malignancies 119
- 4.7.4 Risk-Adapted Therapies 120
- 4.7.5 Epigenetics and Hematological Malignancies 120
- 4.7.6 The Unknown Unknowns – The Future of Molecular Oncology 120

References 121

5**Molecular Virology 123****5.1 The Basics of Virology 124**

- 5.1.1 Human Immunodeficiency Virus 127
- 5.1.2 Hepatitis B Virus 130
- 5.1.3 Influenza Virus 130

5.2 Vaccination 132

- 5.2.1 Live Vaccines 133
- 5.2.2 Recombinant Virus Vaccines 136
- 5.2.3 Inactivated Virus Vaccines 136
- 5.2.4 Subunit Vaccines 137
- 5.2.5 DNA Vaccines 139
- 5.2.6 HIV Vaccines 139

5.3 Detection of Viruses 139

- 5.3.1 Cytopathic Effects 139
- 5.3.2 Electron Microscopy 140
- 5.3.3 Hemagglutination Assay 140
- 5.3.4 Enzyme-Linked Immunosorbent Assay (ELISA) 140
- 5.3.5 Indirect ELISA 140
- 5.3.6 Polymerase Chain Reaction (PCR) 140
- 5.3.7 Antiviral Susceptibility Testing 142

5.4 Antiviral Therapy 142

- 5.4.1 Human Immunodeficiency Virus (HIV) 145
- 5.4.2 Hepatitis C Virus 149
- 5.4.3 Influenza Virus 149
- 5.4.4 Other Viruses 149

5.5 Prions 151**References 151****6****Bacteria and Eukaryotic Pathogens 153****6.1 Bacteria 154**

- 6.1.1 Pathogenic Bacteria 155
- 6.1.2 Bacterial Vaccines, Diagnostic, and Antibiotics 160
 - 6.1.2.1 Vaccines 160
 - 6.1.2.2 Diagnostic 160
 - 6.1.2.3 Antibiotics 160

6.2 Eukaryotic Pathogens 166**References 168****7****Genomics and Proteomics 169****7.1 Whole Genome Sequencing 170**

- 7.1.1 Cloning of a Genome 170
- 7.1.2 Mapping and Assembly of the Genome 172
- 7.1.3 Sequencing of a Large Genome 173

7.2 The Human Genome 174

- 7.2.1 Sequencing of the Human Genome 174
- 7.2.2 The International HapMap Project 183
- 7.2.3 The 1000 Genomes Project and the Personal Genome Project 184
- 7.2.4 Encyclopedia of DNA Elements (ENCODE) 186

7.3 Proteomics 188

- 7.3.1 Two-Dimensional Gel Electrophoresis and Mass Spectrometry 189
- 7.3.2 Quantitative and Shotgun Proteomics 192
- 7.3.3 Structural Proteomics 194

References 194**8****Genetic Testing 197****8.1 Types of Genetic Tests 198**

- 8.1.1 Postnatal Genetic Tests 198
- 8.1.2 Prenatal Genetic Tests 200

8.2 Chromosome Abnormalities 202

- 8.2.1 Conventional Karyotyping 203
- 8.2.2 Fluorescence In Situ Hybridization 203
- 8.2.3 Comparative Genomic Hybridization 205

8.3 Molecular Diagnosis 207

- 8.3.1 PCR-Based Methods 207
- 8.3.2 DNA Sequencing 209
- 8.3.3 DNA Microarray-Based Methods 212

References 213**9****Pharmacogenetics/Pharmacogenomics 215****9.1 Uptake and Transport of Drugs 217****9.2 Drug Metabolism 218**

- 9.2.1 Cytochrome P450 Enzymes 218
- 9.2.2 Other Drug Metabolizing Enzymes 220

9.3 Drug Targeting 222**9.4 Drug Toxicity and Hypersensitivity 226****9.5 Drug Development and Individual Pharmacotherapy 226****References 227****10****Recombinant Protein Drugs 229****10.1 Production of Recombinant Proteins 232**

- 10.1.1 Bacteria 233
- 10.1.2 Yeast and Other Fungi 234

10.1.3 Insect Cells 235
10.1.4 Mammalian Cells 235
10.1.5 Transgenic Animals and Plants 236
10.2 Classes of Recombinant Drugs 238
10.2.1 Monoclonal Antibodies 239
10.2.2 Hormones 245
10.2.3 Growth Factors 247
10.2.4 Fusion Proteins 249
10.2.5 Cytokines 250
10.2.6 Blood Coagulation Factors: Anticoagulants and Thrombolytics 251
10.2.7 Therapeutic Enzymes 254
10.2.8 Recombinant Vaccines 254
References 255

11

Gene Therapy 257

11.1 Types of Gene Therapy 258
11.2 Methods of Gene Transfer 259
11.2.1 Retroviral Vectors 260
11.2.2 Adenoviral Vectors 262
11.2.3 Adeno-Associated Virus Vectors 264
11.2.4 Nonviral Gene Transfer 266
11.3 Tissue Specificity of Gene Transfer and Gene Expression 267
11.4 Applications of Gene Therapy 270
11.4.1 Gene Therapy of Monogenic Diseases 271
11.4.2 Gene Therapy of Cancer 272
11.4.3 Other Diseases 273
11.5 Future Prospects 275
References 276

12

Stem Cells 277

12.1 Embryonic Stem Cells 279
12.1.1 Generation and Properties of Embryonic Stem Cells 279
12.1.2 Therapeutic Cloning 281
12.2 Adult Stem Cells 282
12.3 Induced Pluripotent Stem Cells 286
12.3.1 Generation of Induced Pluripotent Stem Cells 286
12.3.2 Properties of Induced Pluripotent Stem Cells 288
12.4 Transdifferentiation and Direct Reprogramming 289
12.5 Differentiation of Stem Cells 291

12.6 Medical Applications of Stem Cells 293
12.6.1 Adult Stem Cell Therapies 293
12.6.2 Pluripotent Stem Cells for Biomedical Research 296
12.6.3 Therapeutic Applications of Pluripotent Stem Cells 299
12.6.3.1 Diabetes 299
12.6.3.2 Heart Disease 300
12.6.3.3 Neurodegenerative Diseases 300
12.6.3.4 Combinations of Stem Cell and Gene Therapy 301
12.6.3.5 Clinical Trials 301
References 302

13

Antisense, Ribozyme, and RNA Interference Strategies 303

13.1 Antisense Oligonucleotides 305
13.1.1 Mechanism of Action of Antisense Oligonucleotides 305
13.1.2 Development and Stabilization of Antisense Oligonucleotides 306
13.1.3 Clinical Applications 308
13.2 Ribozymes 311
13.2.1 Classification of Ribozymes 311
13.2.2 Development of Ribozymes for Medical Applications 312
13.2.3 Clinical Applications of Ribozymes 314
13.3 RNA Interference 315
13.3.1 Mechanism of RNA Interference 316
13.3.2 Nonspecific Side Effects 319
13.3.3 Delivery 320
13.3.4 Preclinical Applications of RNA Interference 321
13.3.5 Clinical Trials 322
13.4 MicroRNAs 325
13.4.1 The Biology of MicroRNAs 325
13.4.2 MicroRNAs and Disease 327
References 330

14

Aptamers 333

14.1 Selection of Aptamers 335
14.2 Modifications of Aptamers 337
14.3 Clinical Development of Aptamers 339
14.4 Decoy and Immunostimulatory Oligonucleotides 342
References 344

15
Ethics in Molecular Medicine 345

15.1 **The Basis of Bioethics** 346

15.2 **Fields of Application** 348

15.2.1 Genetic Testing and the Right
“Not to Know” 348

15.2.2 Stem Cell Research 350

15.2.3 Preimplantation Genetic Diagnosis 352

Abbreviations 355

Glossary 363

Index 367

Preface

Molecular medicine is one of the hottest, most dynamic areas in all the life sciences and combines the disciplines of biology, biochemistry, human biology, and pharmacology, in addition to basic and clinical medicine. Knowledge of the molecular sources of disease has and will continue to pave the way for the development of novel diagnostic procedures and therapeutic strategies to treat inherited or acquired diseases. The expanding repertory of therapeutic approaches finally makes it possible to address diseases that hitherto have been untreatable. Molecular medicine will without doubt change our disease treatment paradigms; rather than employing standard therapies for every individual and adjusting treatment based on observed efficacy and side effects, personalized approaches based on specific constitutional features of the distressed individual will become possible.

In recent years, many universities worldwide have established bachelors and masters degree programs to study molecular medicine. In addition, molecular medicine has become an indispensable component of well-established curriculums. Courses of study in medicine, (molecular) biology, biochemistry, human biology, and biotechnology have frequently been offered. We thus had the sense that it was time to assemble an up-to-date introduction to the entire spectrum of molecular medicine, one written in a clear and uniform style. Numerous color figures and expositions of clinical relevance will, we hope, also ensure a systematic introduction to the topic.

This textbook primarily addresses undergraduates and lecturers. It is based on our experience in teaching courses in molecular medicine for many years, in laboratory research, and in medical practice. We assume that students who use this textbook have at least some basic knowledge of molecular and cell biology. If not, we recommend that students first consult appropriate introductory textbooks such as *Principles of Biochemistry* by Voet *et al.* or *Molecular Biology of the Cell* by Alberts *et al.* Although this textbook is primarily written for undergraduates, we also hope that medical and graduate

students and researchers in the field of molecular medicine will also benefit from it.

For many reasons, writing a book on molecular medicine is a challenging task. First, there is no common consent on just what molecular medicine actually is. We view molecular medicine as a discipline that investigates normal and pathological cellular processes at the molecular level. Such in-depth analysis will help us to better understand the causes of disease, improve diagnosis, and develop novel therapeutic strategies. This understanding of molecular medicine has led us to the selection of the topics that are covered in this book.

The broad range of topics poses a second challenge, as no two-author team can possibly be experts in all areas of molecular medicine. However, we chose not to edit a book composed of individual specific articles, each written by an expert in a particular field in his or her own style. Instead, we have attempted to present a homogeneous textbook in a uniform and consistent style.

Numerous cross-references will help the reader to understand the complex interdependencies between the different fields of molecular medicine. For example, the monoclonal antibody Herceptin not only provides a comparatively new treatment option for aggressive breast cancer but is also a recombinant protein drug that is challenging to produce. Its successful therapeutic use represents one of the paradigmatic applications of pharmacogenetic testing.

Finally, the dynamics of the field of molecular medicine required us to assemble a textbook with the most up-to-date information possible. An accompanying Web site – http://www.wileyvch.de/home/molecular_medicine – will contain figures and figure legends from the book, in addition to regular updates on the latest developments in the field.

This textbook is divided into chapters, each of which can be viewed as one lecture of a one semester course on molecular medicine. The sections are structured in a logical order. However, each chapter can also stand alone as an introduction to a single topic, for example,

gene therapy or stem cell technology. Cross-references will help the reader to find sections in the book that should be consulted for an in-depth understanding of the topic.

The textbook commences with a short general introduction to molecular and cellular biology and then describes some selected methods widely used in modern life science research. Several chapters deal with the molecular causes of disease, and established as well as new diagnostic approaches are also described. Viral, bacterial, or eukaryotic pathogen infections are also covered, as these are a major cause of suffering and death worldwide, and require continuous improvements in therapeutics for the management of emerging pathogens and drug resistance. The outline of novel therapeutic approaches for the treatment of cancer and genetic disorders developed in the era of molecular medicine will certainly leave the reader impressed by the dynamism of this field: Drugs based on recombinant proteins, particularly monoclonal antibodies, have already become an important element in clinical practice. Timely topics such as the newest advances in gene therapy, stem cell research, and RNA technologies will be introduced with numerous helpful figures. The book will conclude with a chapter on the ethical dimensions of molecular medicine.

Every chapter is accompanied by recommendations for further reading. Rather than providing an exhaustive list of bibliographical references that would be of limited use for most students, we selected educationally valuable review articles for each topic. While we initially intended to choose current articles, we also found older reviews that can provide excellent introductions to various subjects. In addition, the reference lists are intended to help students to begin navigating through the jungle known as the modern scientific literature.

We hope that the readers will enjoy our journey through the field of molecular medicine and share our enthusiasm for this fascinating subject. We would be very pleased if our textbook helps prepare you for working in this exciting field.

Acknowledgments

This book would not have been possible without the help of many people. We are particularly thankful to Erik Wade for reading all the chapters and providing valuable suggestions. We also thank Anke Wagner for her comments on several chapters. We also want to thank our (JK) student Derya Günes for her chapter revisions. Additional thanks goes to Mary Houlemarde and Tonya Nickens for their support, and to Daniela Castanotto for her unfailing good humor, common sense, and help in probing some of the more complex issues of this narrative. A big thanks goes to Harry Kurreck for proofreading all the chapters and giving valuable comments.

We want to thank the team at Wiley, Waltraud Wüst, Anne Chassin du Guerny, Gregor Cicchetti, and Andreas Sendtko, for their advice and assistance. A special thank goes to our graphic designer Anke Wagner for converting our quirky sketches into illustrative figures.

It is impossible for two scientists to cover the whole field of molecular medicine in depth. We are, therefore, thankful to our expert colleagues for their valuable comments on specific topics: Thomas Bock, Toni Cathomen, Henry Fechner, Anja Pöhlein, Roland Lauster, and Daniela Castanotto. In addition, we want to thank Tatjana Schütze for support with the graphical representation of protein structures. We also want to thank our research groups for their enthusiasm about molecular medicine.

Last but not least, we express our most heartfelt appreciation and thanks to our families for their patience when we spent far too much time writing this book. Therefore, this book is for Harry, Marianne, Malte, Paul, and Anke; for Myra, Allison and Warren; Lauren and Brian; and Lily Leigh and Margot Nicole.

May 2015

Jens Kurreck, Berlin
Cy Aaron Stein, Duarte

Introduction

Summary

- Molecular medicine is a highly dynamic field of life science research that uses interdisciplinary approaches to understand normal and pathological cellular processes at the molecular level. The findings of basic research have entered clinical practice, as new diagnostic assays and novel therapeutic strategies focus not only on the symptoms but also on the causes of disease.
- The development of drugs is a long-term and expensive process that starts with basic and pre-clinical research. A candidate drug must then successfully pass through three types of clinical trial in humans before a novel agent can be approved for therapeutic purposes.
- The eukaryotic cell is compartmentalized into several cellular organelles by intracellular membranes. The nucleus harbors the genetic material, mitochondria are the cellular power plants, and the endoplasmic reticulum and the Golgi apparatus are responsible for the glycosylation and sorting of proteins.
- Cells follow a tightly regulated cycle of four phases. These include the two gap phases G_1 and G_2 , the S phase in which new DNA is synthesized, and mitosis, during which the cell divides.
- Apoptosis is the process of programmed cell death, which is important as a normal physiological mechanism and for protection against infections and cancer. Apoptosis can be triggered by extrinsic or intrinsic signals.
- Genomic DNA is amplified by DNA polymerases in a process known as replication. The synthesis occurs in a semiconservative and semidiscontinuous way.
- Expression of genes requires two steps. In the first step, the DNA is transcribed into RNA. Most primary transcripts are posttranscriptionally processed. For mRNAs, this step includes the addition of a cap at

Contents List

The Basics of Molecular Medicine

- Topics of Molecular Medicine
- Stages of Drug Development

The Human Cell

- Organelles
- Cell Cycle
- Apoptosis

DNA Replication and Gene Expression

- DNA Replication
- Mutations
- Transcription
- Epigenetic Regulation of Gene Expression
- Translation
- Protein Degradation

Biological Communication

- Neurotransmitters
- Hormones
- Signal Transduction

The Immune System

- The Innate Immune System
- The Adaptive Immune System

the 5' end and of a poly(A) tail at the 3' end. Introns are spliced out to link the exons together. Several bases are modified in various types of RNAs. The second step in gene expression is the translation of the genetic information into proteins. This process is carried out by ribosomes. Posttranslational modifications of proteins include activation by proteolytic cleavage and covalent modification of amino

acid side chains. This can occur, for example, by glycosylation or reversible phosphorylation.

- Sophisticated communication between cells is essential for the functioning of a multicellular organism. Neurons transmit signals at synapses. Hormones are molecules that induce physiological responses over a long distance or in adjacent cells. The extracellular signals are transmitted into the cell by cell surface receptors and induce a signaling cascade that leads to a biological response.
- The immune system protects an organism against (infectious) disease. The innate immune response recognizes general patterns of pathogens, while the adaptive immune system is directed against specific targets. The adaptive immune system involves a cellular immune response (T cells) and a humoral immune response (B cells that produce antibodies).

1.1 The Basics of Molecular Medicine

1.1.1 Topics of Molecular Medicine

Molecular medicine is a discipline dedicated to understanding normal and pathological cellular processes at the molecular level. This approach requires the use of many physical, chemical, biological, biochemical, and medical techniques (some of which are introduced in Chapter 2) to understand fundamental molecular mechanisms and how they go awry in disease. Molecular medicine combines classical disciplines such as cell and molecular biology, biochemistry, and medicine. Knowledge is often acquired via interdisciplinary investigation and can be used to develop new forms of molecular diagnosis and therapeutic intervention.

Molecular medicine can be divided into a basic research and an applied clinical discipline. The basic research component investigates molecular and genetic mechanisms of cellular function and identifies pathological processes. In many cases, this addresses a specific question with a hypothesis-driven approach, and can lead to large-scale investigations of whole genomes and proteomes (Chapter 7). The discipline known as translational research then tries to apply the findings from basic science to the clinic, where it may provide new forms of diagnosis and therapy.

A report published by Linus Pauling in 1949 laid the basis for the establishment of the field of molecular medicine. In his seminal paper, he showed that hemoglobin from patients suffering from sickle cell anemia

had a different electrical charge than hemoglobin from healthy individuals. This study demonstrated that a disease could be traced to an alteration in the molecular structure of a protein. This novel perspective opened the possibility of establishing novel forms of diagnosis and therapy at the molecular level. Sickle cell anemia is not the only case in which a detailed understanding of the molecular etiology of the disease (e.g., of inherited genetic disorders, Chapter 3) has led to new diagnostic options (Chapter 8), although with only a modestly improved therapeutic outcome.

The field of oncology also illustrates the paradigm shift caused by a molecular perspective. While cancer treatment is still largely based on removal of the tumor by surgery (followed by chemotherapy and/or radiation therapy), molecular oncology (Chapter 4) tries to elucidate those pathways that lead to cellular transformation. This knowledge helps to produce a comprehensive molecular diagnosis of the disease basis in a single patient so that the treatment can be adjusted accordingly, an approach that has come to be known as “personalized medicine.” Many modern anti-cancer drugs block specific pathways that lead to uncontrolled cellular proliferation. Similarly, elucidation of the life cycles of pathogens has helped develop new drugs for the treatment of infectious diseases (Chapters 5 and 6). For example, advancements in virus biology have led to the identification of novel targets for antiviral agents.

Most drugs belong to the class of small molecular compounds. To achieve oral bioavailability and to promote rapid diffusion across cell membranes and intracellular trafficking to their sites of action, the majority of (oral) drugs have molecular weights below 550 Da (although some antibacterial agents fall in the 700–900 Da range). A prominent example is acetylsalicylic acid (trade name Aspirin, Figure 1.1a), a drug mainly used as an analgesic.

Molecular medicine has broadened the spectrum of entities used as drugs. New medications are now often based on large molecules of biological origin (known as “biologics”). These include, for example, recombinant proteins such as monoclonal antibodies (Figure 1.1b) (Chapter 10), short pieces of DNA or RNA (Chapters 13 and 14), entire genes that can be delivered by viral vectors (Chapter 11), or even complete cells (Chapter 12). Pharmacogenetic investigations aim at discovering why the efficacy and toxic side effects of a drug at a given dose vary between individuals (Chapter 9). However, molecular medicine not only develops new diagnostic and therapeutic approaches but also poses heretofore unknown ethical issues, some of which will be introduced in Chapter 15.

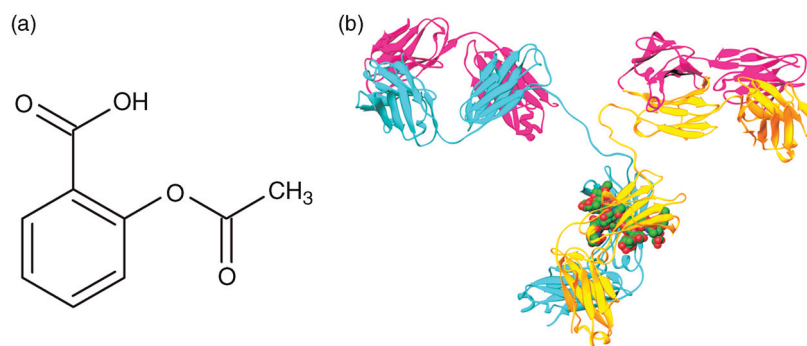


Fig. 1.1 Small molecular drugs and biologics. (a) The chemical structures of acetylsalicylic acid (Aspirin) and (b) the crystal structure of an antibody, shown for comparison. The two structures are not drawn to scale. (Part (b) adapted from Ref. [1] with kind permission from John Wiley & Sons, Inc.)

1.1.2 Stages of Drug Development

The development of a new drug is a time-consuming and expensive process (Figure 1.2) that may take 12–15 years (and in some cases even longer). The cost calculus of developing a new drug is complex and controversial, but the average cost to bring a new molecular entity (NME) to the market has been estimated to be as high as \$1.8 billion. Drug development usually starts with the identification of a new target, which, for example, may be a proliferative factor that causes tumor growth. The next step in the process is to characterize the target, its location (extracellular, membrane-bound, cytosolic, and nuclear), and its function. Confirmation that the potential drug target fulfills the expected function is known as *target qualification* or *validation*. One way to identify a new compound is to perform what is known as a high-throughput screen (HTS). This approach allows testing a large number (up to millions) of compounds to identify an active molecule that modulates a particular target

(e.g., inhibits a proliferative factor). In almost all cases, the primary hit must be optimized by chemical modification to obtain higher binding affinities, better solubility, and so on. The efficacy and toxicology of the substance are then investigated in *in vitro* studies and animal experiments. The process of drug development may deviate substantially from this path depending on the type of drug being developed. Biologics, for example, are usually not obtained by HTS. Any substance will only be tested in humans after having passed extensive toxicological examination.

Clinical research is usually divided into three main phases. However, these phases are sometimes preceded by an exploratory trial (frequently called as phase 0) in a small number of subjects with a very small, subtherapeutic dose designed to gather data on the agent's basic properties in humans. This trial does not produce data about safety or efficacy. The actual clinical research starts with a phase I trial, usually carried out with a small number (20–100) of subjects. The main purpose of a phase I trial is to assess the safety of the drug. The

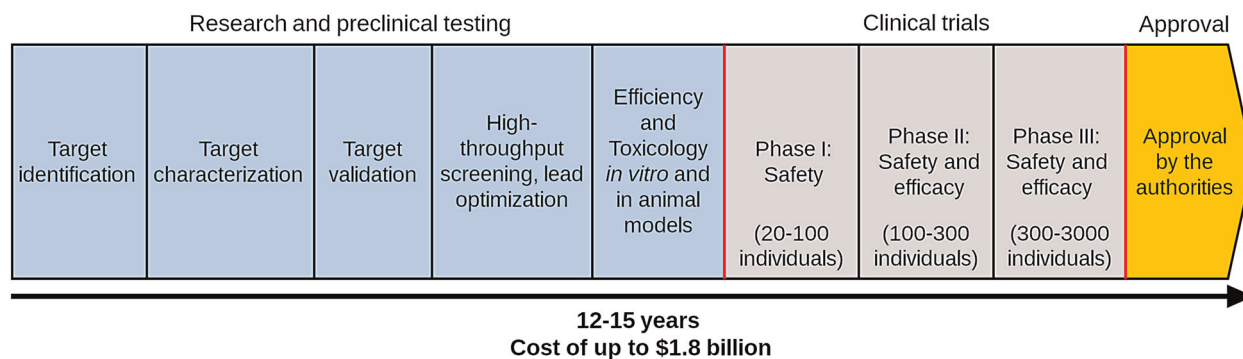


Fig. 1.2 Stages of drug development. The preclinical stages comprise the identification, characterization, and validation of a target and the identification and optimization of a compound, in addition to toxicological evaluation. The drug then undergoes three main phases of clinical testing before it is approved by the regulatory authorities.

trial is frequently designed to include a dose escalation to determine the optimal dose and the dose at which unacceptable toxicity supervenes. Although phase I trials are often carried out with healthy volunteers, under some circumstances, ill patients are enrolled. This is done most often with cancer patients, as the drugs to be evaluated are likely to make healthy individuals ill or may carry a significant risk of long-term toxicity.

Phase II trials are carried out with a larger number (100–300) of individuals. The central aim of the phase II study is to evaluate the efficacy of the drug. The trial is, therefore, usually performed in sick patients. Phase II studies are sometimes divided into phase IIA and phase IIB. While a phase IIA trial assesses the optimal dosing of the drug, a phase IIB trial is designed to study the efficacy of the drug at the prescribed dose. Another important goal of a phase II trial is to assess drug safety in a larger group of individuals.

Phase III trials are designed to assess the effectiveness and safety of a new drug in clinical practice in a large patient group (300–3000 or more individuals). These studies are carried out in randomized, controlled, multicenter trials. Phase III trials are usually designed as double blind studies, that is, the patients are randomly assigned to an experimental and a control group (in some trials, the control group may either receive a placebo or standard of care treatment). Neither the patients nor the physicians monitoring the outcome know which patient is receiving which treatment. Phase III trials aim at assessing the efficacy of a drug in comparison to placebo or the current standard of care treatment.

If drug safety and efficacy have been demonstrated in multiple phase III trial(s), approval for marketing can be applied for from the appropriate regulatory agency such as the Food and Drug Administration (FDA) in the United States or the European Medicines Agency (EMA) in the European Union. These agencies may request postmarketing monitoring, which is sometimes referred to as a phase IV trial. A phase IV trial involves safety surveillance after the drug has received permission to be marketed. In principle, it is designed to detect rare or long-term adverse effects in a much larger patient population and over a longer time period than was possible during the earlier phase of clinical trials. A phase IV trial may also identify interactions with other marketed drugs. Even after marketing, if harmful effects are discovered, any drug may be withdrawn at any time or its use restricted only to certain conditions.

The term phase V is increasingly used to describe studies that determine whether the therapeutic effect of a new drug is realized in day-to-day clinical practice. Community-based research is employed to survey whether the effects under typical (and somewhat

variable) clinical contexts are similar to those that were found in the controlled efficacy studies. A phase V trial may also analyze the cost-benefit ratio of a drug or therapeutic intervention.

1.2 The Human Cell

Despite the extreme complexity of living organisms and the myriad number of functions that each constituent organ must carry on, only a surprisingly limited set of molecules are commonly employed. Typical biomolecules found in living organisms include nucleic acids, proteins, polysaccharides, and lipids. These macromolecules are composed of a relatively limited number of monomeric building blocks such as DNA and RNA nucleotides, amino acids, monosaccharides, and fatty acids. In addition, inorganic ions, organic acids, and a variety of metabolites are also important constituents of cells. While the basic features of biomolecules are extensively covered in the general textbooks of biochemistry listed at the end of this chapter, the cell as the basic functional unit of an organism and the major intracellular processes relevant to human physiology and pathology will be outlined here.

The human body consists of approximately 100 trillion (10^{14}) cells. Although all cells of a given organism carry the same genome, these cells have different functions and are highly specialized. A typical eukaryotic (animal) cell is illustrated in Figure 1.3. The most prominent characteristic that distinguishes eukaryotic cells from prokaryotes is its compartmentalization. The main membrane-bound organelles of animal cells are the nucleus, the endoplasmic reticulum (ER), the Golgi apparatus, the mitochondrion, the lysosome, and the peroxisome. Each of these organelles contains a specific set of proteins that fulfill a specific function. These organelles are embedded in a gelatinous fluid called the cytosol.

1.2.1 Organelles

1.2.1.1 The Nucleus

The central organelle of a eukaryotic cell is the nucleus, which contains the genetic material. The nucleus is surrounded by the nuclear envelope consisting of two membranes: the inner and the outer nuclear membranes (Figure 1.4a). The outer nuclear membrane is continuous with the rough endoplasmic reticulum (RER). Proteins referred to as nucleoporins form aqueous channels, called nuclear pores, through the envelope. These pores allow small water-soluble molecules to pass into and out

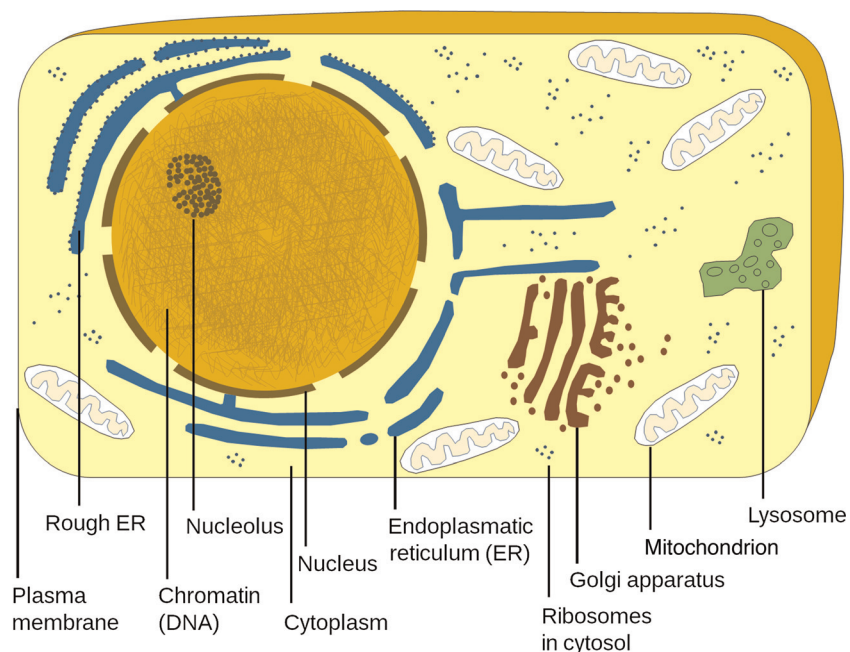


Fig. 1.3 Diagram of a typical eukaryotic cell. Membrane-bound organelles include the nucleus, endoplasmic reticulum (ER), Golgi apparatus, mitochondrion, lysosome, and peroxisome. (Reproduced with permission from Ref. [2]. Copyright 2008, Garland Science/Taylor & Francis LLC.)

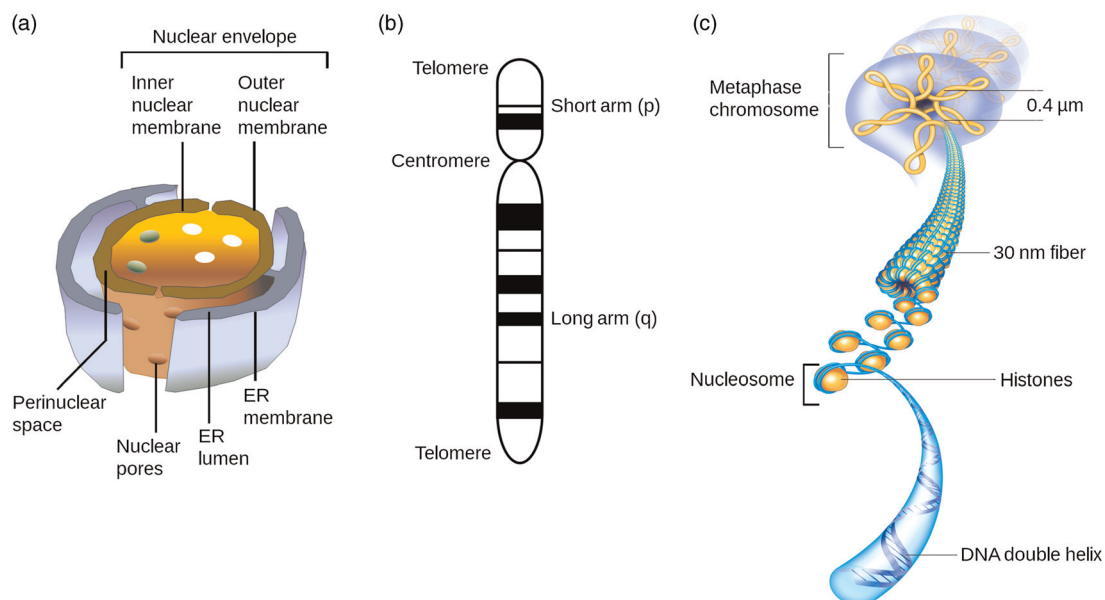


Fig. 1.4 Cell nucleus and chromatin organization. (a) The nucleus is surrounded by the nuclear envelope consisting of two membranes. Pores in the envelope allow the exchange of small water-soluble molecules. The outer nuclear membrane is continuous with the membrane of the endoplasmic reticulum (ER). (b) Schematic representation of a eukaryotic metaphase chromosome. Each chromosome has a specific banding pattern after staining. (c) DNA winds around histone proteins to form nucleosomes that fold into a 30 nm fiber. Loops of chromatin are then attached to a protein scaffold to form the metaphase chromosome. (Part (a) reproduced with permission from Ref. [2]. Copyright 2008, Garland Science/Taylor & Francis LLC. Part (c) adapted from Ref. [1] with kind permission from John Wiley & Sons, Inc.)

of the nucleus, while larger molecules must be actively transported in or out. A filamentous network of lamin proteins provides mechanical support, but is also involved in the regulation of replication and cell division. A distinct structure found in the cell nucleus is the nucleolus (indicated in Figure 1.3). The nucleolus occupies up to one fourth of the volume of the nucleus and forms around specific chromosomal regions. In the nucleolus, ribosomal RNA is transcribed and assembled with proteins to give (incomplete) ribosomes.

The DNA together with DNA-binding proteins and RNA molecules is organized in chromosomes. The human genome consists of 46 chromosomes that can be identified based on the specific banding pattern after staining. Figure 1.4b shows a schematic representation of a condensed metaphase chromosome. The ends of the chromosome are called *telomeres*. The *centromere* is the point where the two identical chromosomes touch and the microtubules attach for separation during mitosis. Most chromosomes are asymmetric; the centromere separates a short arm (p for the French word *petit*) and a long arm (q, chosen as next letter in the alphabet after p). The position of a gene or a DNA sequence in the human genome (denoted as the *locus*) is indicated by the chromosome number, the arm, and three numbers that refer to the region, the band and the subband. For example, the locus 11p15.3 indicates that a DNA sequence is located on the short arm of chromosome 11, in region 1, band 5 and subband 3.

If the DNA of every cell in the human body were lined up end to end, it would stretch from the Earth to the Sun and back 100 times. This means that a single cell must package DNA strands with a combined length of approximately 2 m into a nucleus with a volume in the cubic micrometer (μm^3) range.

Chromosomes consist of the DNA complexed with proteins, together known as chromatin (Figure 1.4c). The first step in chromatin condensation is achieved by winding the DNA around so-called histone proteins. These proteins are positively charged (i.e., contain many arginine and lysine residues) and can bind negatively charged DNA by electrostatic interactions. Two copies of each of the histones H2A, H2B, H3, and H4 form a core around which the DNA is wound twice. The fifth histone protein, the linker histone H1, binds to the middle of the DNA and to its two ends, locking the DNA into place. This structural unit, called the nucleosome, allows packaging in higher ordered structures, such as the so-called structural element “30 nm fiber.”

The interaction between proteins and DNA has a major impact on gene expression (Section 1.3.3). Much of a cell’s DNA is not expressed, as it does not encode a gene product. In addition, the expression of a particular

gene may be inappropriate for a certain cell type. Non-expressed DNA is typically highly condensed; it is called *heterochromatin*. In contrast, lightly packed DNA (*euchromatin*) facilitates access to those enzymes required for gene expression and is, therefore, transcriptionally active. Nucleosomes also undergo remodeling as proteins required for transcription must bind to their target DNA. The *chromatin remodeling complex* disrupts the interaction between histones and DNA in the nucleosomes in an ATP-dependent manner so that the DNA becomes accessible. Histones undergo extensive covalent posttranslational modification (Section 1.3.5), which, together with DNA methylation, are two of the main mechanisms of the epigenetic regulation of gene activity (Section 1.3.4).

1.2.1.2 Mitochondria

Mitochondria (Figure 1.5) are organelles with double membranes that function as “cellular power plants” since they are the site of the oxidative metabolism that leads to the production of ATP. A eukaryotic cell typically contains 800–2000 mitochondria. In the internal matrix, they contain all the enzymes required for the citric acid cycle and for fatty acid oxidation. Redox proteins involved in the electron transport chain and oxidative phosphorylation are embedded in the inner membrane, which invaginates into cristae to expand its surface area.

A special feature of mitochondria is that they carry a genome that encodes several (but not all) mitochondrial proteins. The human mitochondrial genome is a circular molecule of approximately 17 000 base pairs encoding 13 proteins and 2 ribosomal and 22 transfer RNAs. Each mitochondrion contains multiple copies of the mitochondrial genome. Additional mitochondrial proteins are encoded in the nucleus. The mitochondria and their

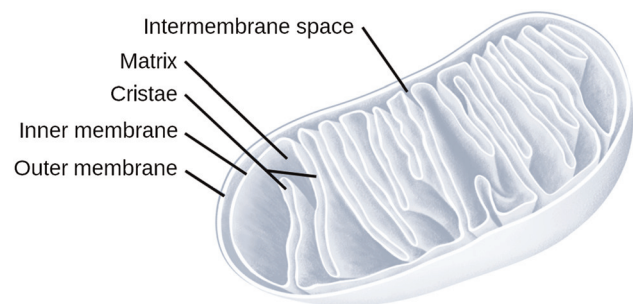


Fig. 1.5 The mitochondrion. The mitochondrion is bounded by two membranes. The inner membrane harbors the electron transport chain bearing the respiratory enzymes of the cell. It is extensively invaginated, forming the so-called cristae. The internal matrix contains high levels of enzymes involved in oxidative metabolism, and also holds the mitochondrial DNA. (Adapted from Ref. [1] with kind permission from John Wiley & Sons, Inc.)

genetic material are exclusively inherited from the mother. The *endosymbiotic hypothesis* is an attempt to explain the origin of mitochondria (and plastids such as chloroplasts in plants). According to this theory, mitochondria and plastids evolved from formerly free-living bacteria that were engulfed by another cell as an endosymbiont through endophagocytosis. The mitochondrial DNA (mtDNA) represents the remnant of the bacterial genome, while the remaining genetic material moved into the nuclear genome.

Mitochondrial dysfunction may result in mitochondrial diseases, which may be caused either by mutations in mtDNA or in nuclear genes that encode mitochondrial proteins (Section 3.1.4). Mitochondrial dysfunction may have severe consequences, due to their critical function in cellular energy supply. These effects usually vary from organ to organ and also depend on other variations in the genome. Mitochondrial diseases often lead to neuromuscular disease symptoms (myopathy), but also include diabetes mellitus, deafness, visual loss, and dementia.

1.2.1.3 Endoplasmic Reticulum and Golgi Apparatus

The endoplasmic reticulum and the Golgi apparatus play an important role in the posttranslational modification and sorting of proteins (Figure 1.6). The ER is a network of interconnected, flattened, membrane-enclosed sacs known as cisternae. The enclosed space is called the

lumen (or cisternal space). The main function of the ER is to support the synthesis and export of proteins and membrane lipids. Two types of ER can be distinguished, the rough ER (RER) and the smooth ER (SER). While the SER is involved in lipid and carbohydrate metabolism and in detoxification, the RER plays an important role in posttranslational modification of proteins. The rough appearance is caused by ribosomes that bind to the cytosolic side of the ER membrane and translocate newly synthesized proteins into the lumen of the ER during the translation process. The main functions of the RER include the support and control of protein glycosylation and of correct protein folding, as will be described in more detail in Section 1.3.5. The RER carries out initial N-linked glycosylation, which is terminated in the Golgi apparatus. Proteins are shuttled from the RER to the *trans* Golgi network in membrane-bound vesicles.

The main functions of the Golgi apparatus include the termination of N-glycosylation and the initiation of O-linked protein glycosylation. Vesicles from the RER carrying proteins fuse with the membrane on the *cis* face of the Golgi apparatus and release their cargo. Proteins are modified in the lumen and leave the Golgi apparatus from the *trans* Golgi network, which sorts proteins for extracellular release or for transport to the lysosome, according to the markers they carry.

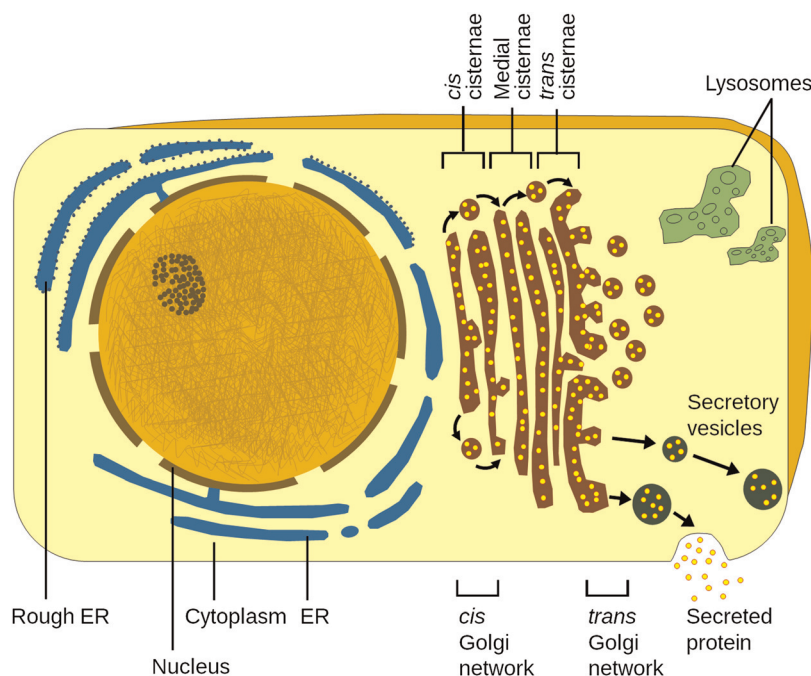


Fig. 1.6 Endoplasmic reticulum (ER) and Golgi apparatus. The endoplasmic reticulum and the Golgi apparatus are two organelles that are important for the correct folding, posttranslational modification, and sorting of proteins. N-Glycosylation of proteins is initiated in the ER and is terminated in the Golgi apparatus, while O-glycosylation exclusively occurs in the Golgi apparatus.

The importance of correct protein sorting is illustrated by the hereditary disease mucopolidosis II, which is characterized by psychomotor retardation, skeletal deformities, and death in childhood. The molecular cause of the disease is the lack of a phosphotransferase that normally phosphorylates mannose residues to mannose-6-phosphate on *N*-linked glycoproteins. Without this marker, enzymes that normally degrade glycosaminoglycans are transported from the Golgi apparatus into the extracellular space, rather than into the lysosome. This defect in protein sorting eventually leads to the accumulation of glycosaminoglycans and glycolipids in large inclusions in lysosomes that cannot be degraded. The disease is thus also known as Inclusion-cell (I-cell) disease.

1.2.1.4 Peroxisome and Lysosome

Peroxisomes and lysosomes are eukaryotic organelles with important metabolic functions, which consist mainly of degradative processes. Peroxisomes are involved in the catabolism of very long-chain and branched fatty acids and in the oxidative catabolism of amino acids. In addition, they contain enzymes of the pentose phosphate pathway. Peroxisomal disorders typically affect the nervous system, since peroxisomes also play a role in the biosynthesis of ether phospholipids, which are critical for the normal functioning of the mammalian brain.

Lysosomes are the cell's waste disposal and recycling system, as they contain enzymes that digest dispensable cellular components and ingested material. Lysosomes employ approximately 50 hydrolytic enzymes to break down cellular waste products, fats, carbohydrates, proteins, and other macromolecules into simpler compounds, which are returned to the cytoplasm as building blocks for biosynthetic processes. A variety of proteases, known as *cathepsins*, degrade proteins in a nonselective manner. Lysosomes maintain an interior pH of approximately 5, and their enzymes function optimally at acidic pH. The cytosol, with its higher pH (slightly above 7), is thus protected from the degradative activity of lysosomal enzymes that may leak into the cytosol, since they are inactive in this more alkaline environment.

Approximately 50 rare inherited metabolic disorders are known to result from malfunction of the lysosome. They are collectively referred to as *lysosomal storage diseases* (LSDs). The lack of a particular lysosomal enzyme may result in accumulation of substances destined for breakdown. LSDs have a variety of symptoms, but typically affect children, who die at a young age. Lysosomes also have a major influence on the pharmacokinetics of several drugs. Uncharged, weak bases may cross lysosomal membranes and accumulate in lysosomes. This phenomenon is called

lysosomotropism and may explain the high tissue concentration and long elimination half-lives of some drugs. Accumulation in lysosomes may result in the inhibition of the activity of lysosomal enzymes.

1.2.2 Cell Cycle

The cell cycle is the sequence of events that leads to cell division into two daughter cells. It can be divided into three main periods: the interphase, the mitotic phase, and cytokinesis (Figure 1.7). Mitosis and cytokinesis together define the mitotic (M) phase of the cell cycle. A complete cell cycle in cultured cells takes approximately 16–24 h. In contrast, the doubling times of a cell in a multicellular organism can vary from 8 h to >100 days. Some terminally differentiated cells, such as muscle cells or neurons, never divide. This quiescent state is called the G_0 phase.

Interphase is characterized by cell growth, the accumulation of nutrients, preparation for mitosis, and the replication of DNA. It can be subdivided into three phases: The longest segment of interphase is the G_1 phase (G stands for gap), in which cells grow in size, have elevated biosynthetic activity, and produce enzymes required for DNA replication. DNA replication itself takes place in the synthesis phase (S phase), during which time all the chromosomes are replicated. S phase is followed by the second gap phase (G_2) during which time cells continue to grow. Several control mechanisms ensure that the cell is ready to enter mitosis.

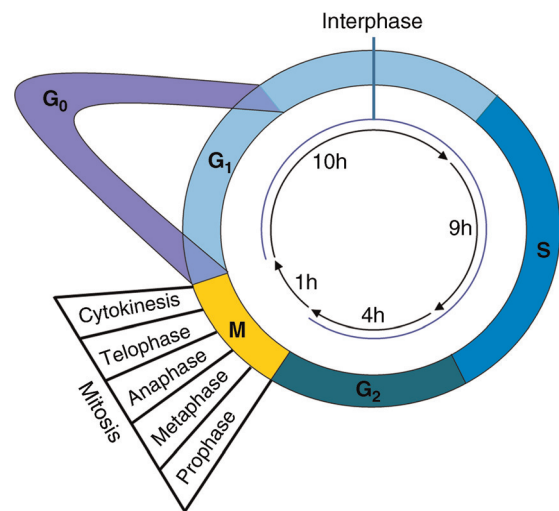


Fig. 1.7 The cell cycle. The eukaryotic cell cycle is divided into three main phases: (1) interphase consists of the G_1 and G_2 gap phases as well as the synthesis phase (S); (2) mitosis comprises pro-, meta-, ana-, and telophase; (3) in the final phase, cytokinesis, the cell divides.

Mitosis starts with *prophase*, which is characterized by spindle formation and condensation of the duplicated chromosomes. The nuclear membrane then breaks apart and the nucleus disappears (this process is sometimes called prometaphase). In the subsequent *metaphase*, the duplicated chromosomes move to positions midway between the spindle poles. The next step is the separation of the sister chromatids of the duplicated chromosomes and their movement to opposite poles of the cell. This occurs in *anaphase*. The last phase of mitosis is called *telophase* and involves decondensation of the chromosomes and restoration of the organelles and the nucleus. Although cytokinesis is not part of mitosis in the strict sense, it directly follows mitosis and involves division of the cytoplasm into two daughter cells.

As uncontrolled proliferation is a main characteristic of cancer, strict control of cell cycle is essential for the organism. The cell cycle has several checkpoints that monitor its progress and arrest the cell cycle if certain conditions have not been fulfilled. For example, the cell cycle only progresses to mitosis after the cell's DNA has been completely replicated. Another checkpoint ensures that all chromosomes are properly attached to the mitotic spindle.

Two classes of proteins, known as cyclins and cyclin-dependent kinases (Cdks), are the main regulators of progression through the cell cycle. Different types of cyclins are specific for each phase of the cell cycle. Cyclins are synthesized during one cell cycle phase and subsequently degraded during the succeeding phase. A cyclin forms a complex with its corresponding Cdk, which leads to the activation of the Cdk. Cdks are serine/threonine protein kinases that phosphorylate nuclear target proteins involved in the various processes of the cell cycle. Cdks are also regulated by cyclin-dependent kinase inhibitors, which arrest the cell cycle under certain conditions. These conditions include contact with adjacent cells, DNA damage, terminal cell differentiation, and cell senescence. Alterations of the inhibitory activity of Cdks are frequently found in cancer. Leland H. Hartwell, R. Timothy Hunt, and Paul M. Nurse were awarded the 2001 Nobel Prize in Physiology or Medicine for their discovery of the cyclin/Cdk system.

1.2.3 Apoptosis

Apoptosis (the Greek word for “falling off”, for example, of leaves from a tree) is the process of programmed cell death. Apoptosis is a normal, physiological mechanism found in multicellular organisms that is important for the development and maintenance of proper physiological functioning and for the protection against cancer and

other diseases. An example of the essential function of apoptosis in developmental processes is the formation of fingers in vertebrates. Initially, the digits are fully connected by webbing. During development, the web is eliminated by programmed cell death. An estimated 10^{11} of the $\sim 10^{14}$ cells of an adult human die every day through apoptosis and are replaced by new ones. Apoptosis is also important for the elimination of virus-infected or malignantly transformed cells. Abnormal apoptosis may result in neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's diseases, as well as in damage caused by stroke or heart attack.

The apoptotic program includes a sequence of cellular morphological changes: As a first step, the cell begins to shrink. Its chromatin becomes condensed at the nuclear periphery and the cytoskeleton collapses. Subsequently, the nuclear envelope disappears, the DNA is fragmented, and the plasma membrane forms irregular bulges (blebs). Eventually the cell disintegrates into numerous membrane-enclosed vesicles called apoptotic bodies. Apoptotic bodies are phagocytosed by neighboring cells and macrophages. Apoptosis should be differentiated from necrosis, which is a form of cell death that results from trauma, such as acute cellular injury. In this case, water rushes into necrotic cell due to the loss of cell membrane integrity. The cell swells, its organelles lose their function, and the cell eventually bursts, spilling its contents into the extracellular space.

The processes of apoptosis are mainly controlled by *caspases*, a family of cysteinyl aspartate-specific proteases. Caspases are heterotetramers, consisting of two α - and two β -subunits, which are expressed as inactive, single-chained zymogens (procaspases). Upon activation, the N-terminal prodomain is excised and the α - and β -subunits are proteolytically separated to yield the active $\alpha_2\beta_2$ caspase. Two types of apoptotic caspases can be distinguished: initiator and effector (executioner) caspases. Initiator caspases cleave inactive zymogen precursors of effector caspases, thereby activating them. The effector caspases then cleave other cellular proteins, triggering the apoptotic process.

Apoptosis may be induced in two ways (Figure 1.8): External signals trigger the *extrinsic pathway* (death by commission), while the absence of external signals that inhibit apoptosis can lead to the activation of the *intrinsic pathway* (death by omission). The extrinsic pathway is induced by binding of a so-called death ligand (e.g., the Fas ligand) on an inducing cell to the death receptor (Fas receptor) on the surface of the cell destined to undergo apoptosis. This leads to activation of the initiator caspase-8 (and possibly caspase-10), which triggers executioner caspases to cleave their substrate proteins, driving the cell into apoptosis. The internal pathway is

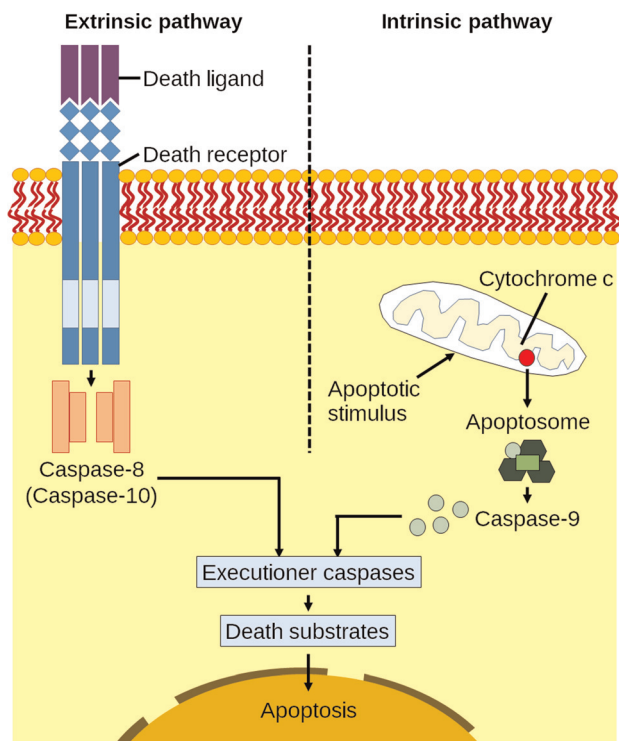


Fig. 1.8 Extrinsic and intrinsic pathways of apoptosis. Programmed cell death can be induced by extrinsic signals from a death ligand or by an apoptotic stimulus that leads to the release of cytochrome c from mitochondria. In both cases, an initiator caspase is activated that activates executioner caspases, which cleave their substrates. This program eventually leads to cell death. (Adapted from Ref. [3] with kind permission from Macmillan Publishers Ltd., Copyright 2002.)

induced in the absence of signals from the environment, for example, neighboring cells, which normally prevent the cell from committing suicide. Proapoptotic members of the Bcl-2 protein family are important mediators of these signals. They activate mitochondria to release cytochrome c into the cytosol, where it binds the apoptotic protease activating factor 1 (APAF1) to form the *apoptosome*. In the apoptosome, the initiator caspase-9 is

activated, which then triggers executioner caspases as per the program described for the extrinsic pathway.

In 2002, the Nobel Prize in Medicine was awarded to Sydney Brenner, Robert Horvitz and John E. Sulston for their discoveries concerning the genetic regulation of organ development and programmed cell death.

1.3 DNA Replication and Gene Expression

(Anke Wagner, Berlin University of Technology, Berlin, Germany)

Cellular DNA must be copied before a cell can divide into two daughter cells. This occurs via a process known as *replication*. In addition, the genetic information stored in the DNA must be transferred into proteins, the actual primary effectors of cell function. This process of gene expression comprises two steps: the *transcription* of DNA into RNA and the *translation* of the RNA sequence into the corresponding amino acid sequence necessary for protein formation. This flow of genetic information, formulated in brief as “DNA makes RNA makes protein,” was summarized by Francis Crick in 1958 in what is known as the *central dogma of molecular biology* (Figure 1.9). Crick himself acknowledged that the term dogma may be misunderstood as a belief that cannot be doubted. In fact, we now recognize several exceptions to the original model. The process of reverse transcription from RNA to DNA in retroviruses, or the replication of RNA molecules into new RNA molecules by other viruses (Section 5.1.1) are such examples. The dogma also does not cover processes such as epigenetic regulation of gene expression by DNA methylation (Section 1.3.4), posttranscriptional gene silencing by the RNA interference pathway (Section 13.3), or post-translational protein modification. However, the assumption that proteins cannot serve as the template for the

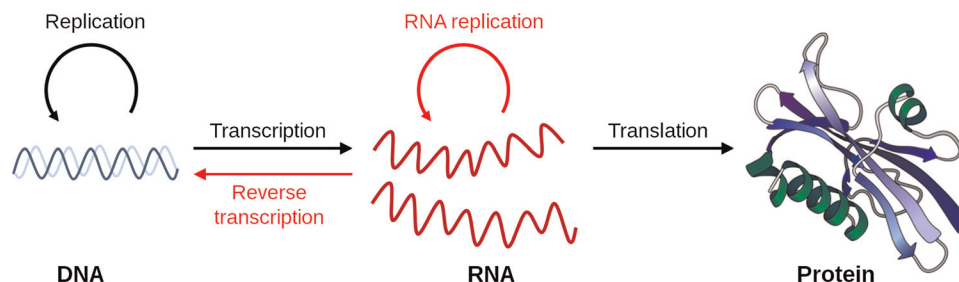


Fig. 1.9 The central dogma of molecular biology. DNA is copied in a process known as replication. The genetic information stored in the DNA is expressed by transcription into RNA, followed by translation into an encoded protein. Several exceptions to the model (marked in red) are known, including the replication of RNA into RNA and the reverse transcription of RNA into DNA.

synthesis of new proteins or for the transfer of sequence information into RNA or DNA is still valid.

1.3.1 DNA Replication

Replication is the process of copying the genetic information contained in cellular DNA. The double-helix model of DNA first proposed by Watson and Crick suggested a molecular mechanism for the transmission of hereditary information: Each strand of the double-stranded DNA molecule can serve as a template for the synthesis of a complementary daughter strand. Since the newly synthesized DNA double strands consist of one parental strand and one daughter strand, this mode of replication is termed *semiconservative*.

The process of prokaryotic and eukaryotic replication is similar, although it is more complex in the eukaryotes. Altogether, replication involves the concerted action of more than 25 proteins. DNA synthesis occurs at the so-called replication fork, a simplified scheme of which is depicted in Figure 1.10.

In eukaryotic cells, the initial step of replication is the separation of the two strands of the DNA by a heterohexameric helicase called minichromosome maintenance (MCM). The unwinding process, which is driven by the hydrolysis of ATP (or another NTP), causes supercoiling

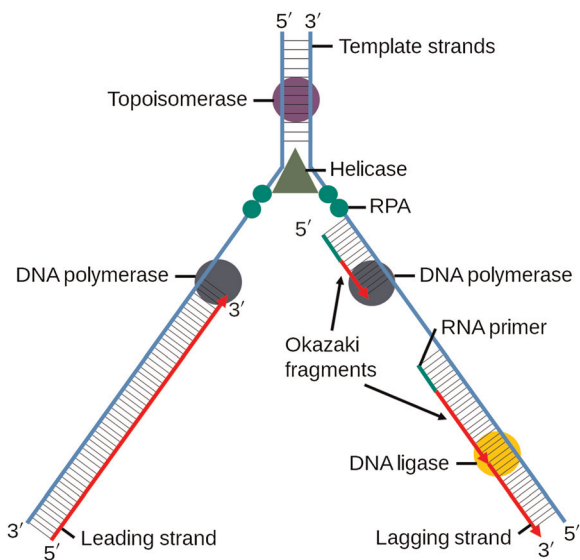


Fig. 1.10 Replication fork. Two DNA polymerase enzymes synthesize the new DNA strands, which are called the leading and lagging strands, respectively. A helicase unwinds the double-stranded template molecule and topoisomerases prevent supercoiling of the DNA. The replication protein A (RPA) stabilizes the unwound DNA single strands. DNA polymerase requires RNA primers to initiate synthesis of the Okazaki fragments, which are finally joined together by DNA ligase.

of the DNA ahead of the replication fork, which would eventually halt replication. To overcome these topological problems, the enzyme topoisomerase catalyzes the relaxation of DNA by cutting one or both strands of the DNA, rotating it, and reannealing the cleaved strand(s).

The two strands separated by the helicase have a strong tendency to reanneal to form a DNA duplex again. To prevent reannealing, the replication protein A (RPA) coats the single-stranded DNA. The central enzymes of DNA replication are the DNA-dependent DNA polymerases, simply known as DNA polymerases. Animal cells contain at least 13 distinct types of DNA polymerases. Like bacterial DNA polymerases, they can extend DNA strands only in the $5' \rightarrow 3'$ direction. The two strands of parental DNA, therefore, have to be replicated in different ways. While the *leading strand* is synthesized continuously in the $5' \rightarrow 3'$ direction of the replication fork movement, the other strand, known as the *lagging strand*, is also synthesized in the $5' \rightarrow 3'$ direction, but can only be made in small segments known as Okazaki fragments. The overall process is, therefore, described as a *semidiscontinuous* mode of replication.

Another common feature shared by all known DNA polymerases is that they cannot synthesize a new strand *de novo*, but can only add nucleotides to the free $3'$ -OH group of a base-paired polynucleotide. The replication of DNA is, therefore, initiated by the synthesis of a short RNA primer. In eukaryotes, a primase synthesizes a 7- to 10-nt RNA primer. DNA polymerase α is associated with the primase and extends the primer by approximately 15 DNA nucleotides. The chain is then further extended by DNA polymerases δ and ϵ . Both polymerases are highly processive, meaning they can polymerize long stretches of the DNA without dissociating from the strand. DNA polymerase δ needs a sliding clamp protein named proliferating cell nuclear antigen (PCNA) to maintain its processivity. PCNA is also required for the synthesis of the lagging strand. In contrast, DNA polymerase ϵ is highly processive in the absence of PCNA and is involved in leading strand synthesis.

Both polymerases δ and ϵ carry out an additional *proofreading* step, for which they possess $3' \rightarrow 5'$ exonuclease activity: In case the DNA polymerase incorporates a mismatched nucleotide into the newly synthesized strand, its polymerase activity is halted and the mismatched nucleotide is hydrolytically excised by the exonuclease activity. The enzyme can then incorporate the correct nucleotide and resume its normal polymerase activity. DNA polymerases with proofreading activity are characterized by high fidelity, that is, their error rates are only about one in every 10^6 – 10^{11} nucleotides, depending on the type of polymerase.

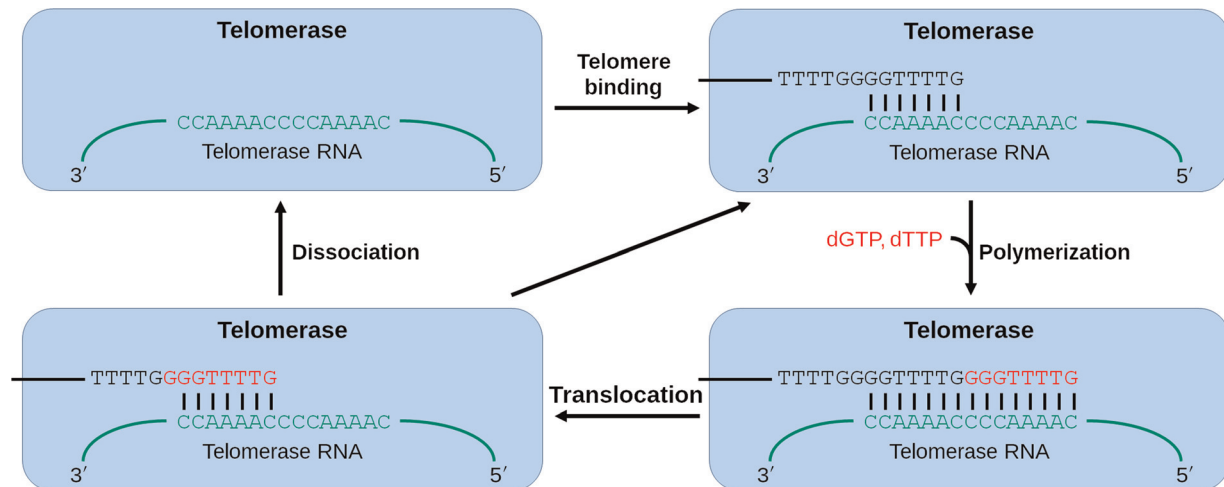


Fig. 1.11 Telomerase. Telomerase is a cellular reverse transcriptase that extends the ends of linear chromosomes (telomeres). It contains an RNA component to which the single-stranded 3' end of the chromosome hybridizes. Telomerase then adds several nucleotides. The telomere can translocate and undergo several rounds of extension before it dissociates from telomerase.

Eventually, the RNA primers must be removed from the DNA and the Okazaki fragments must be joined together. In eukaryotes, this step is initiated by RNase H1 (H stands for hybrid, since the enzyme recognizes a DNA–RNA hybrid and hydrolyzes the RNA; this catalytic activity is also used in antisense technologies, as will be described in Section 13.1). The RNase removes most of the RNA and leaves only one RNA nucleotide adjacent to the DNA, which is then removed by the flap endonuclease-1. The excised nucleotides are replaced by DNA polymerase δ and the nick is eventually sealed by DNA ligase.

Eukaryotic chromosomes are large: The longest human chromosomes consist of more than 200 million base pairs. To achieve replication in a few hours, eukaryotic chromosomes contain multiple origins of replication, one every 3–300 kb depending on the species and tissue. Clusters of 20–80 adjacent replicating units (*replicons*) are activated simultaneously. Several sets of these clusters must be activated so that the entire chromosome can be replicated.

The linear ends of eukaryotic chromosomes, called telomeres, present a particular problem for the replication machinery. As already outlined, an RNA primer is required for the initiation of DNA replication. This primer can be removed but not replaced by the DNA polymerase. To prevent shortening of the linear chromosomes with each round of replication, eukaryotes have developed a mechanism to maintain the telomeres. The enzyme telomerase is a ribonucleoprotein that consists of a protein and an RNA component (of 451 nucleotides in humans). Since it functions as a

reverse transcriptase, as illustrated in Figure 1.11, the protein component is called telomerase reverse transcriptase (TERT). The RNA component of telomerase is a complementary sequence to the repeating telomeric sequence and directs the addition of nucleotides to the 3' end of the DNA. This process of polymerization followed by translocation can be repeated several times before the telomere dissociates from the enzyme. The complementary strand can then be synthesized by the normal cellular machinery for lagging strand synthesis.

Somatic cells of multicellular organisms lack telomerase activity; thus, the telomeres shorten with each round of replication. Cells in culture can only undergo a certain number of cell divisions before they reach senescence and die. The absence of telomerase in somatic cells is at least part of the basis for aging in multicellular organisms. The presence of shortened telomeres in somatic cells used for the cloning of animals is also believed to be one of the reasons for the health problems found in these clones (Section 10.1.5, Box 10.1). Telomerase is active mainly in two types of cells: germ cells that must maintain intact telomeres and cancer cells that divide rapidly and would stop growing without a mechanism to prevent shortening of the telomeres. The senescence of somatic cells and the process of aging due to lack of telomerase activity can be viewed as mechanisms that protect multicellular organisms from cancer. Telomerase is an attractive target for new anticancer drugs, since its inhibition should prevent the uncontrolled growth of tumor cells.

1.3.2 Mutations

Mutations are changes in the genetic information. They may cause cancer or genetic disorders. A large number of human cancers (estimates reach up to 80%) arise from substances referred to as *carcinogens*, which induce mutations. Over 70 000 man-made chemicals are currently of commercial importance and 1000 new ones are introduced every year.

Germinal mutations occur in germline cells and are transmitted through the gametes to the progeny. In contrast, *somatic mutations* occur in somatic cells; thus, the mutant phenotype will occur only in the descendants of that cell and will not be transmitted to the progeny. While *spontaneous mutations* occur without any known cause, perhaps due to inherent metabolic errors or to unknown agents in the environment, *induced mutations* result from exposure of organisms to mutagens. The degeneracy and order in the genetic code outlined below help prevent many mutations from affecting the phenotype of the organism. These changes are called *neutral* or *silent mutations*.

Several types of mutations are known: These include substitutions of one base pair for another and deletions and insertions of one or more base pairs (Figure 1.12). The most common type of mutation is the substitution of one base pair for another. A *transition* replaces a pyrimidine with another pyrimidine or a purine with another purine. A *transversion* replaces a pyrimidine with a purine or a purine with a pyrimidine. Altogether, four different types of transitions and eight different types of transversions are possible in DNA.

An example of a germline mutation is a single base substitution in the β -globin gene that results in sickle-cell anemia in homozygous individuals (Sections 3.1.2

and 6.2). This disorder is caused by an A to T substitution, causing the hydrophilic amino acid glutamic acid to be replaced with the hydrophobic amino acid valine at the sixth position of the β -globin protein. The exchange promotes the noncovalent polymerization (aggregation) of hemoglobin, which distorts red blood cells into a sickle shape and decreases their elasticity.

Base pair insertions or deletions within the coding regions of genes usually lead to so-called *frameshift mutations* (except if three nucleotides or multiples thereof are added or removed), because they alter the reading frame of all the base pair triplets. Since the triplets specify codons in mRNA and amino acids in the encoded protein (Figure 1.12), frameshift mutations mostly result in the synthesis of nonfunctional protein products.

Mutations can be induced physically (e.g., by UV radiation), chemically, or biologically (e.g., by transposons or viruses). Chemical mutagens may be mutagenic for replicating DNA only (e.g., acridine dyes) or mutagenic for both replicating and nonreplicating DNAs (e.g., alkylating agents and nitrous acid). For example, treatment of DNA with nitrous acid causes point mutations by oxidative deamination. Thus, cytosine and adenine can be converted to uracil and hypoxanthine, respectively.

Insertions and deletions can arise from the treatment of DNA with intercalating compounds such as acridine orange. Intercalation of such an agent into DNA almost doubles the distance between two consecutive base pairs. As a consequence, the replication of the distorted DNA occasionally results in the insertion or deletion of one or more nucleotides.

Environmental factors, such as UV or ionizing radiation, and some chemical compounds can also cause physical damage to the DNA. For example, UV light promotes the formation of a cyclobutyl ring between adjacent thymine residues that leads to the formation of

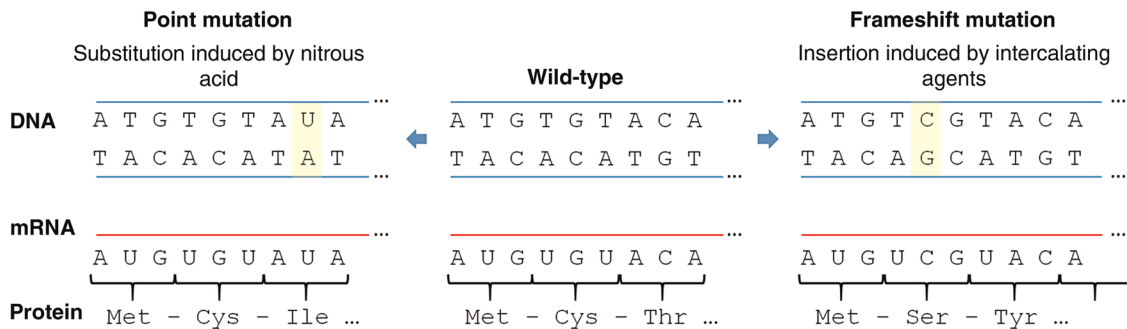


Fig. 1.12 Types of mutations. Substances such as nitrous acid can induce point mutations, that is, the substitution of one base by another. This may lead to a change of the codon and the incorporation of an incorrect amino acid into the protein. Frameshift mutations, induced, for example, by intercalating agents, occur by the introduction (or deletion) of one or multiple nucleotide(s). A reading frameshift changes all the codons and results in the synthesis of a nonsense protein.

intrastrand thymine dimers. Cytosine or cytosine–thymine dimers can also be formed, but at lower rate. These pyrimidine dimers all result in errors during transcription and replication. Treatment of DNA with ionizing radiation may lead to strand breakage. DNA damage may either be caused directly by radiation or by the formation of free radicals in the surrounding aqueous medium. Cells have developed several very effective mechanisms for the repair of DNA damage. Clinically, individuals with inherited disorders in the repair systems can develop disease such as *xeroderma pigmentosum*, which can lead to the early development of cancer through exposure to sunlight (Section 3.1.2).

The *Ames test* is a widely employed method to investigate the mutagenic potential of a substance. It is a quick and convenient bacterial test that serves as an alternative to time-consuming carcinogenesis tests on mice and rats; such evaluation in animals may take up to 3 years. In the assay, a special tester strain of *Salmonella typhimurium* that cannot synthesize histidine (denoted as *his*[−]) is used. These bacteria are unable to grow in the absence of the amino acid. Bacteria of the tester strain are spread on a culture plate lacking histidine. In the presence of a mutagen, some bacteria revert to the *his*⁺ phenotype, that is, they regain the ability to synthesize histidine and can grow and form visible colonies on the histidine-deficient medium. When this happens unusually frequently, it is an indication that the tested material is mutagenic.

1.3.3 Transcription

Transcription is the process of copying of DNA into RNA by the enzyme RNA polymerase (RNAP). Similar to replication, transcription relies on the transmission of genetic information by the pairing of complementary bases; however, transcription usually involves only copying of one of the two DNA strands, while both strands are doubled during replication. The DNA strand that serves as a template for the synthesis of RNA is called the noncoding or antisense strand, while the strand with the same sequence as the RNA is known as the coding or sense strand. As will be outlined in various chapters of this book, it has become apparent only recently that a great diversity of RNA molecules exists (Sections 7.2.4 and 13.4.2), all of which are generated by RNAPs. Animals contain three distinct types of RNAPs in the cell nucleus (as well as a separate mitochondrial RNAP), which transcribe different types of RNAs (Table 1.1).

Eukaryotic RNAPs are large complexes consisting of up to 14 subunits. They possess DNA unwinding activity, synthesize the new RNA molecule in the 5′ → 3′

Table 1.1 Eukaryotic RNA polymerases.

| <i>Polymerase</i> | <i>Location</i> | <i>Transcribed RNAs</i> |
|--------------------|-----------------|---|
| RNA polymerase I | Nucleolus | rRNA precursors |
| RNA polymerase II | Nucleoplasm | mRNA precursors; U1-, U2-, U4-, and U5-snRNAs |
| RNA polymerase III | Nucleoplasm | 5S-rRNA, tRNAs, U6-snRNA, most miRNAs, and various other small RNAs |

The table summarizes the location of the three eukaryotic RNA polymerases in the nucleus and the types of RNAs they transcribe.

direction, and carry out a proofreading step similar to the one described for DNA polymerases. The mode of action of the three RNAPs is somewhat different, and the following paragraphs will focus on the synthesis of mRNA precursors by RNAP II.

Transcription is initiated at specific sites of the genome known as promoters. RNAP II promoters contain certain sequence motifs, sometimes referred to as core promoter elements (Figure 1.13), such as a GC-rich region or the so-called TATA box (thymidine–adenosine–thymidine–adenosine) located 25–31 nucleotides upstream of the transcription start site (it is often overlooked that approximately two thirds of protein-encoding genes lack the TATA box!). The promoters of housekeeping genes that are expressed in all tissues have a GC box upstream of their transcription start site. However, this GC box is missing in structural genes selectively expressed in certain cell types. About half of the protein-coding genes have a conserved 7-nucleotide initiator element (Inr) that includes the first nucleotide of the transcribed RNA (+1). None of the core promoter elements are present in all promoters. Further elements, such as the CCAAT box located in the region −50 to −110 to the transcriptional start site, are indicated in Figure 1.13.

Additional elements, called enhancers, are required for the full activity of their cognate promoters. These control elements do not have fixed positions and orientations relative to their corresponding transcribed sequences. Enhancers are recognized by specific transcription factors that stimulate RNAP II to bind to the corresponding, but often distant, promoter. This requires that the DNA between the enhancer and the promoter forms a loop that permits simultaneous interaction of the transcription factors with the enhancer and the RNAP at the promoter site.

While a large number of promoters are ubiquitously active in all cell types, others are tissue specific. For example, α_1 -antitrypsin is a protease inhibitor that is expressed only in the liver. Its promoter, abbreviated