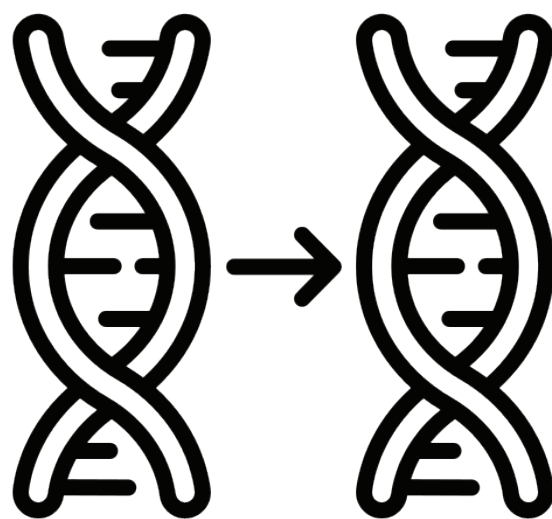


FUNDAMENTAL PROCESSES

MOLECULAR BIOLOGY

7TH EDITION



Dr. KAILASH CHOUDHARY

Copyright 2026, IFAS Publications

All rights reserved. No part of this publication or the information contained herein may be reproduced, adapted, abridged, translated, stored in a retrieval system, computer system, photographic or other systems or transmitted in any form or by any means, electronic, mechanical, optical, digital, by photocopying, recording or otherwise, without written prior permission from the publisher. Any breach will attract legal action and prosecution without further notice.

Disclaimer: While every effort has been made to avoid errors and omissions, this publication is being sold and marketed on the understanding and presumption that neither the editors (or authors) nor the publishers nor the printers would be liable in any manner whatsoever, to any person either for an error or for an omission in this publication, or for any action to be taken on the basis of this work. Any inadvertent discrepancy noted may be brought to the attention of the publisher, for rectifying it in future editions, if published.

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the editors and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The editors and publisher have attempted to trace and acknowledge the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission and acknowledgement to publish in this form have not been obtained. If any copyright material has not been acknowledged please write and let us know so that we may rectify it.

Trademark Notice: Publications or corporate names may be trademarks, and are used only for identification and explanation in bonafide intent without intent to infringe.

Design & Published by:
IFAS Publications

Address:
**Shivranjan Complex,
Someshwar Wadi, Near
Rajwada Hotel, Baner Road,
Pune 411 008.**

Contact No.: **+91-960 489 7388**

Website: **<https://ifasonline.com>**

7th Edition: **2026**

Price: **₹ 895.00**

ISBN: **978-81-69023-05-4**

PREFACE

Every cell carries a library it must copy without error, read without confusion, and repair without losing the sense of the text.

Molecular biology is the branch of life science that studies the processes essential to a cell's functioning and maintenance — replication, transcription and translation. It is the level at which the abstract idea of a gene becomes an actual chemical event, and it has had a profound impact on the way life-science investigation is now carried out.

Major developments in molecular biology over the last four decades have revolutionised life science. Innovations in next-generation sequencing, CRISPR/Cas, epigenetics, meganucleases, genome editing and recombinant DNA technology have made it possible to divulge the high complexity of the genome and to reveal the precise mechanisms by which genetic information is transmitted from one generation to the next.

It is now established that gene expression, DNA replication and DNA repair are processes far more intricate than was formerly assumed. That complexity includes — but is not limited to — interconnected regulatory pathways involving previously unforeseen actors such as non-coding RNA; the importance of epigenetic mechanisms and post-transcriptional modifications; the correct execution, timing, sequence and coordination of every cell-cycle event; the pleiotropic functions of the players operating in these processes; and the influence of energetic metabolism and of signals arriving from the environment.

Given the rapidly changing and continuously evolving nature of the field, this book is an attempt to place updated information about the molecular processes in the reader's hands. It describes, systematically, the important concepts of replication, DNA damage and repair, transcription, translation, the regulation of gene expression, and the molecular techniques through which all of this became known.

Numerous books on molecular biology are available, but there has been a scarcity of books that generate the quest to learn the subject. Emphasis has therefore been laid on beginning every topic with the basic question that first comes to a student's mind, and on presenting the detailed, updated information that answers it in the simplest manner possible — so that reading becomes an interest rather than an obligation.

Although every effort has been made to make this book error free, I am under no illusion. I welcome comments, criticism and suggestions from readers, and they will shape the contents of the next edition.

Kailash Choudhary

ACKNOWLEDGEMENT

A book carries one name on its spine and the fingerprints of many hands on every page.

First of all, I would like to thank all our students at the Institute for Advanced Studies (IFAS), who have helped me to learn and to practise both the art and the science of molecular biology. Their questions decided, more than anything else, what this book had to explain and how plainly it had to be said.

I would like to thank Er. Radheshyam Choudhary (CEO) and Dadasaheb Sondge (COO) IFAS Publications, for being an incessant might behind this project and for making this book happen.

My special thanks to Pranshu Dwivedi and Pratiksha Sen for reviewing the manuscript and helping me prepare its systematic layout; their comments were invaluable in improving the text, and I could not have done this without their help and their passion for molecular biology. Furthermore, my special thanks to Vishal Bhujbal and Digambar Jagtap for their help with the layout of the book. An additional thanks to Lalji Kannoujiya and Rakesh Kumawat for their creative suggestions.

This book is a team effort, and producing it would have been impossible without the outstanding people of IFAS Publications. It has been a pleasure to work with the many dedicated and creative people of IFAS during its production, for the regular reminders that brought the project to timely completion. Thanks are due to Deepak Namabani, who drew my ideas in the form of beautiful diagrams. Special thanks to Vikendra Metha for the wonderful design of the cover page, and to Kuldeep Singh for the typesetting.

And finally, my humble greetings to all those who put in significant effort and remain unmentioned here. Their work is in this book even if their names are not.

Kailash Choudhary

CONTENTS

1	DNA Replication.....	1
1.1	Introduction: Why, When and How DNA Is Copied.....	1
1.2	Basic Concepts of Replication.....	2
1.3	The Proteins and Enzymes of Replication	5
1.4	DNA Replication in Bacteria	11
1.5	Nuclear Replication in Eukaryotes	14
1.6	The End-Replication Problem	19
1.7	Chromatin Replication.....	22
1.8	Extra-Chromosomal Replicons.....	23
2	DNA Repair & Recombination.....	29
2.1	Mutations and the DNA-Damage Response	29
2.2	Endogenous Sources of DNA Damage	30
2.3	Exogenous Sources of DNA Damage.....	32
2.4	Direct Reversal and Mismatch Repair	34
2.5	Base- and Nucleotide-Excision Repair	37
2.6	Crosslink Repair and Translesion (SOS) Synthesis.....	40
2.7	Repair of Strand Breaks: HR and NHEJ	42
2.8	Homologous and Site-Specific Recombination	43
3	Transcription.....	45
3.1	Introduction: Transcription Compared with Replication.....	45
3.2	The Many Types of RNA	46
3.3	RNA Polymerases	47
3.4	Transcription in Bacteria	48
3.5	Transcription in Eukaryotes	52
3.6	Co-transcriptional Capping, Polyadenylation and Termination.....	54
3.7	Transcription of rRNA Genes by RNA Polymerase I	57
3.8	Transcription of tRNA and 5S Genes by RNA Polymerase III.....	57
4	Post-Transcriptional Processing	59
4.1	The Split Gene: Exons and Introns.....	59
4.2	RNA Splicing and the Spliceosome	60
4.3	Self-splicing Introns and Ribozymes	61
4.4	The Fidelity of Splicing	63
4.5	Alternative Splicing	64
4.6	Trans-splicing and the Origin of Introns.....	68
4.7	Alternative Polyadenylation and RNA Editing	69
4.8	RNA Transport, Stability and rRNA/tRNA Processing.....	71
4.9	Ribozymes	74

5	Translation.....	75
5.1	Messenger RNA	75
5.2	The Genetic Code	76
5.3	Transfer RNA and Aminoacyl-tRNA Synthetases	79
5.4	The Ribosome	81
5.5	Translation in Prokaryotes	83
5.6	Translation in Eukaryotes	88
5.7	mRNA Surveillance Pathways.....	93
5.8	Translational Inhibitors	95
5.9	Post-Translational Modification and Protein Degradation	95
5.10	Protein Targeting and Trafficking.....	96
6	Prokaryotic Gene Regulation	99
6.1	Levels of Regulation.....	99
6.2	How Activators and Repressors Work	99
6.3	The Operon and the lac Operon	101
6.4	The ara Operon and Alternative Sigma Factors	107
6.5	Regulation After Transcription Initiation	109
6.6	Antisense RNA and the Stringent Response	111
6.7	CRISPR-Cas Adaptive Immunity	112
6.8	Gene Regulation in Bacteriophage λ	113
7	Eukaryotic Gene Regulation	119
7.1	Overview of Eukaryotic Gene Regulation	119
7.2	Eukaryotic Activators	120
7.3	How Eukaryotic Activators Work.....	121
7.4	Eukaryotic Repressors and Signal Transduction	125
7.5	Transcriptional Silencing and Epigenetics	126
7.6	RNA Interference and Regulatory RNA	128
7.7	Regulation at the Translational Level	132
8	Molecular Techniques	135
8.1	Analysing Promoters and Regulatory Elements	135
8.2	Analysing DNA-Protein Interactions.....	137
8.3	Analysing Protein-Protein Interactions	140

DNA Replication

LEARNING OBJECTIVES

- | | |
|--|--|
| 1.1 Need of replication | 1.5 Eukaryotic origins |
| 1.2 Modes of replication | 1.6 The end-replication problem |
| 1.3 Enzymes & proteins of the replication fork | 1.7 Replication of chromatin |
| 1.4 Initiation and its control in bacteria (oriC) | 1.8 Extra-chromosomal replicons |

Every time a cell divides, it must hand each daughter cell a complete and accurate copy of its genome. DNA replication is the process that makes this copy. Because the two strands of the double helix are complementary — an A on one strand always faces a T on the other, and a G always faces a C — each strand already carries all the information needed to rebuild its partner. During replication the two strands are pulled apart, and each old strand acts as a template on which a new, complementary strand is built. The result is two double helices; each made of one old strand and one brand-new one.

This chapter follows that process from the simple chemistry of copying a strand, through the enzymes that run the replication fork, to the special problems posed by circular bacterial chromosomes, linear eukaryotic chromosomes, plasmids and mitochondrial DNA.

1.1 Introduction: Why, When and How DNA Is Copied

DNA carries the complete set of genetic instructions required to build, maintain, and reproduce a living organism. In humans, each chromosome consists of a single, extraordinarily long DNA molecule packaged with proteins into chromatin. Before a cell divides, it must duplicate its entire genome so that each daughter cell inherits a complete set of genetic information. DNA replication is therefore an essential prerequisite for cell division.

When does it happen?

In bacteria (prokaryotes), DNA replication occurs during the C period of the bacterial cell cycle. In eukaryotes, replication takes place during the tightly regulated S (synthesis) phase of the cell cycle, which lies between the G1 and G2 phases in the sequence G1 → S → G2 → M.

Although the details differ between bacteria and eukaryotes, the fundamental principles of DNA replication are highly conserved. The shared features are:

- Replication begins at a defined DNA sequence called the origin of replication (ori). Origins are often A–T rich, although they are not exclusively composed of A–T base pairs.
- Replication is bidirectional—two replication forks usually move away from each origin in opposite directions.
- Replication is semi-conservative—each daughter DNA molecule contains one parental strand and one newly synthesized strand.
- DNA is synthesized only in the 5'→3' direction, with the template strand read in the 3'→5' direction.
- Replication is semi-discontinuous—the leading strand is synthesized continuously, whereas the lagging strand is synthesized as Okazaki fragments.
- DNA polymerases cannot initiate DNA synthesis de novo. They require a pre-existing 3'-OH group, provided by an RNA primer synthesized by primase.

The principal differences between bacteria and eukaryotes arise from chromosome organization and genome size. In bacteria, DNA is associated with nucleoid-associated proteins but lacks the nucleosome-based chromatin found in eukaryotes, making it generally more accessible to the replication machinery. In contrast, eukaryotic DNA is wrapped around histone proteins to form nucleosomes, which must be disassembled ahead of the replication fork and reassembled behind it.

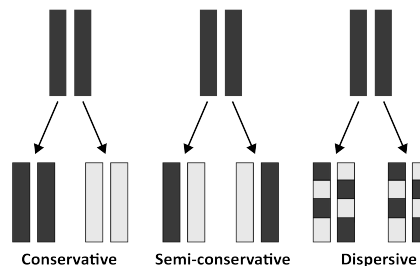
Most bacterial chromosomes contain a **single origin of replication**, so each chromosome forms one replicon, whereas eukaryotic chromosomes contain **multiple origins**, allowing large genomes to be copied efficiently. Rapidly growing bacteria can initiate a new round of DNA replication before the previous round has been completed, whereas eukaryotic cells normally ensure that each origin fires only **once during each cell cycle**. Finally, the **linear chromosomes** of eukaryotes possess specialized ends called **telomeres**, which create the end-replication problem discussed in Section 1.6. This problem does not arise in the **circular chromosomes** of most bacteria.

1.2 Basic Concepts of Replication

Three possible modes — and the one nature actually uses

In principle, template-directed copying could work in three ways (Figure 1.1). In semi-conservative replication the two parental strands separate and each templates a new partner, so every daughter molecule is half old and half new. In conservative replication the two parental strands would stay together and an entirely new double helix would be made alongside them. In dispersive replication each strand would end up a patchwork of old and new stretches. Only the semi-conservative mode actually occurs in living cells; the other two are useful mainly as ideas to test against experiment.

Figure 1.1. The three theoretically possible modes of DNA replication: semi-conservative, conservative and dispersive.

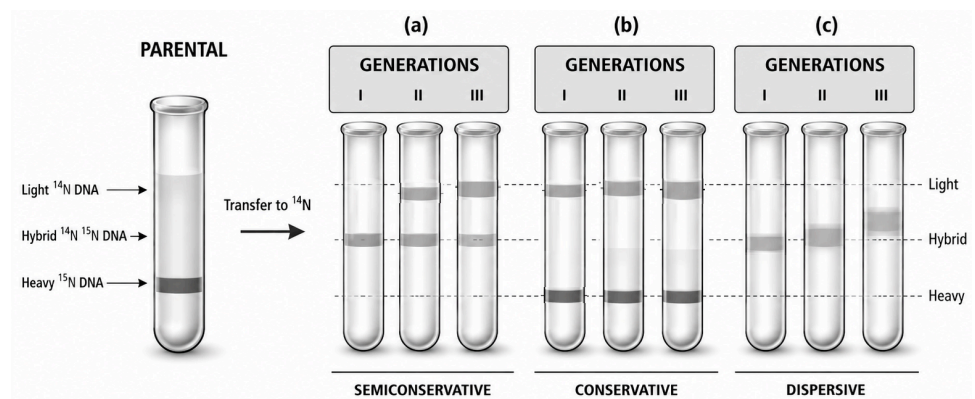


Proving replication is semi-conservative: Meselson and Stahl

The classic proof came from Matthew Meselson and Franklin Stahl (Figure 1.2). They first grew *E. coli* for many generations in a medium whose only nitrogen source was the heavy isotope ^{15}N , so that all the DNA became 'heavy'. They then switched the cells to ordinary ^{14}N ('light') medium and let them divide. At intervals they extracted the DNA and spun it in a caesium-chloride density gradient — a technique in which DNA settles at the point in the tube where its own density matches that of the CsCl solution, detected by its absorbance of ultraviolet light at 260 nm.

The results could only fit the semi-conservative model. After exactly one round of replication in ^{14}N , all the DNA had a single intermediate density — each molecule was a hybrid of one heavy old strand and one light new strand. After a second round, half the DNA was intermediate and half was fully light. Had replication been conservative, the original heavy band would have persisted throughout; had it been dispersive, the bands would have smeared toward lighter densities in a different pattern. Neither was seen.

Figure 1.2. The Meselson–Stahl experiment. (a) The banding pattern actually seen, which fits semi-conservative replication; (b, c) the patterns expected if replication were conservative or dispersive.



J. Herbert Taylor and colleagues demonstrated semi-conservative DNA replication in chromosomes using **tritiated thymidine (^3H -thymidine)**, which is incorporated into newly synthesized DNA and detected by autoradiography in *Vicia faba* (broad bean) root tip cells. After one round of replication, both sister chromatids of each chromosome were labeled because each contained one newly synthesized radioactive strand. After a second round of replication in unlabeled medium, only one chromatid of each chromosome remained labeled, while the other was unlabeled, exactly as predicted by the semi-conservative model.

The same conclusion has also been demonstrated in eukaryotes using the thymidine analogue **bromodeoxyuridine (BrdU)**, which is incorporated into newly synthesized DNA in place of thymidine. After one round of replication in BrdU, both sister chromatids of each chromosome contained one BrdU-labeled strand. After a second round of replication in BrdU, one chromatid contained BrdU in both DNA strands, whereas the other contained one BrdU-labeled strand and one original thymidine-containing parental strand, exactly as predicted by the semi-conservative model (Figure 1.3).

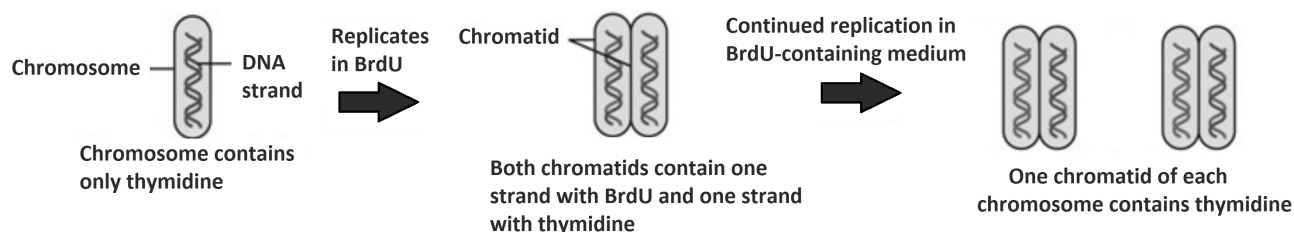
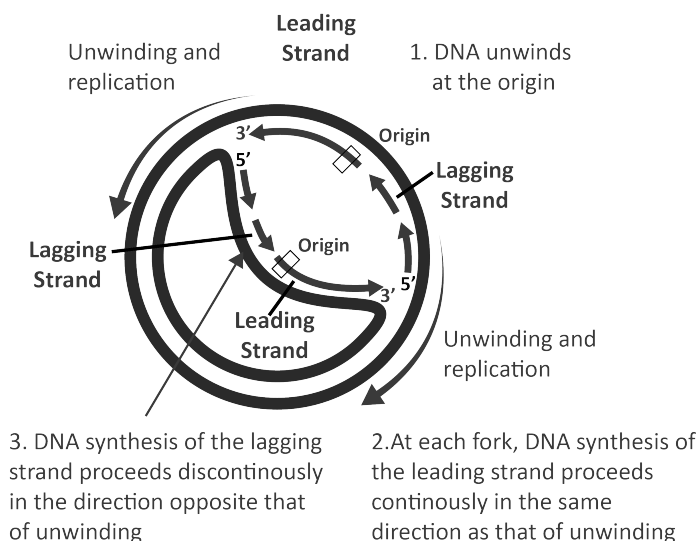


Figure 1.3: Demonstration, using BrdU labelling of sister chromatids, that replication is semi-conservative in eukaryotes.

Replication is bidirectional

Working with *Escherichia coli* in 1963, John Cairns used autoradiography by growing cells in medium containing **tritiated thymidine (^3H -thymidine)** and exposing the isolated DNA to photographic film, allowing him to visualize the entire circular chromosome during replication. The replicating DNA resembled the Greek letter theta (θ): a replication bubble, or 'replication eye', formed at the origin and expanded as two replication forks moved in opposite directions around the chromosome (Figure 1.4). These images gave rise to the term 'theta replication', and subsequent studies confirmed that DNA replication proceeds bidirectionally.

Figure 1.4. Theta (θ) replication of the circular *E. coli* chromosome, showing a bubble with two forks moving in opposite directions.



Direction and substrate of synthesis

A new DNA strand always grows in the 5'→3' direction: each incoming nucleotide is added to the free 3'-hydroxyl group at the growing end, forming a new phosphodiester bond, while the template strand is read in the 3'→5' direction. The substrates for DNA synthesis are the four deoxyribonucleoside 5'-triphosphates (dATP, dGTP, dCTP, and dTTP). As each nucleotide is incorporated, its two terminal phosphate groups are released as pyrophosphate (PP_i), and the subsequent hydrolysis of pyrophosphate drives the reaction forward, making DNA synthesis effectively irreversible.

Why a polymerase needs a primed template

DNA polymerases are the enzymes that build new DNA; the first, DNA polymerase I, was purified from *E. coli* by Arthur Kornberg. Every DNA polymerase, whether bacterial or eukaryotic, has two non-negotiable requirements: a single-stranded template to copy, and a primer—a short pre-existing strand that supplies a free 3'-OH end onto which the next nucleotide can be added. This is why neither a fully double-stranded molecule (no exposed template) nor a bare single strand (a template but no 3'-OH to extend) can be copied on its own. Only a partly double-stranded, 'primed' template satisfies both needs and supports synthesis (Figure 1.5).

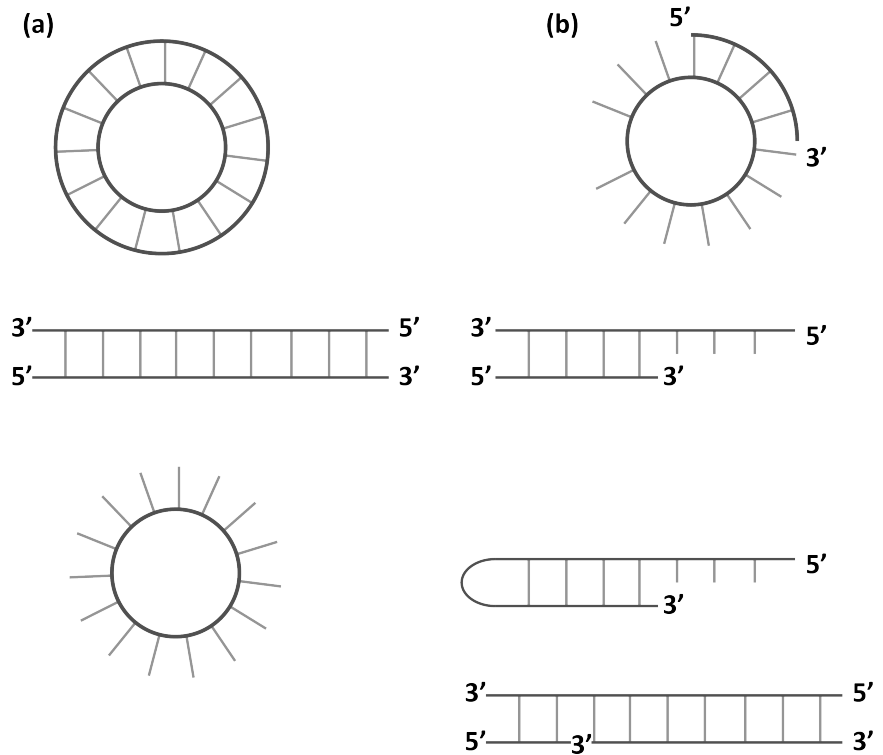


Figure 1.5. *Templates that do and do not support DNA synthesis. A usable substrate has a single-stranded template plus a primer strand with a free 3'-OH.*

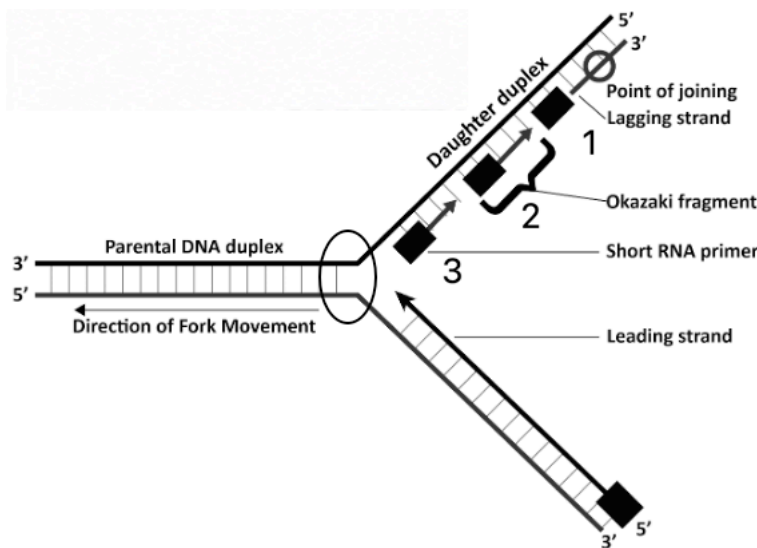
Replication is semi-discontinuous: leading and lagging strands

Here is the geometric challenge at the heart of DNA replication. The two strands of DNA are antiparallel—they run in opposite directions—yet DNA polymerases can synthesize DNA only in the 5'→3' direction. As the replication fork opens, one template runs 3'→5' in the direction of fork movement, allowing its complementary strand to be synthesized continuously toward the fork.

This is the **leading strand**, and its polymerase remains highly processive, synthesizing long stretches of DNA without dissociating. The other template runs 5'→3' relative to fork movement, so its complementary strand must be synthesized discontinuously as short segments away from the fork. This is the **lagging strand**, and its short DNA segments are called **Okazaki fragments**—typically about 1000–2000 nucleotides long in bacteria and 100–200 nucleotides long in eukaryotes. These fragments were discovered by **Reiji and Tsuneko Okazaki** in 1968. Because one strand is synthesized continuously while the other is synthesized discontinuously, DNA replication is described as **semi-discontinuous** (Figure 1.6).

Synthesis of the lagging strand in bacteria involves four repeating steps for each Okazaki fragment: **primase** synthesizes a short RNA primer, **DNA polymerase III** extends the primer, **DNA polymerase I** removes the RNA primer while replacing it with DNA, and **DNA ligase** seals the remaining nick by forming a phosphodiester bond. Despite these additional steps, the leading and lagging strands are synthesized simultaneously at the replication fork by coordinated DNA polymerases that progress at the same overall rate.

Figure 1.6. Semi-discontinuous synthesis: the leading strand is made continuously while the lagging strand is built as Okazaki fragments. 1, 2 and 3 represents order of synthesis of Okazaki fragments on lagging strand.



1.3 The Proteins and Enzymes of Replication

A replication fork is a small molecular machine — a 'replisome' — built from many proteins that act in a strict order. The same jobs must be done in every organism, but the specific proteins differ between bacteria and eukaryotes. Table 1.1 lines up the two sets side by side; keep it handy as we meet each protein in turn.

Table 1.1. The main proteins of the replication fork in *E. coli* and in eukaryotes, matched by the job they do.

E. coli protein	Eukaryotic protein	Function
DnaA	ORC (Orc1–6)	Recognises the origin of replication
DnaB	Mcm2–7	Helicase that unwinds the parental duplex
DnaC	Cdc6, Cdt1	Loads the helicase onto the origin
Gyrase	Topoisomerase II	Relieves positive supercoils ahead of the fork
SSB	RPA	Keeps separated strands single-stranded
Primase (DnaG)	Primase (in Pol α)	Makes the short RNA primer
—	Pol α	Extends the RNA primer with a short DNA stretch
γ -complex	RFC	Clamp loader; places the sliding clamp on DNA
β -clamp	PCNA	Sliding clamp; makes the polymerase processive
Pol III core	Pol δ / Pol ϵ	Main replicative polymerases (lagging / leading)
Pol I	FEN1 (+ RNase H2)	Removes RNA primers and fills the gap
DNA ligase	DNA ligase	Seals Okazaki fragments into a continuous strand

Getting started: initiator proteins and helicase loaders

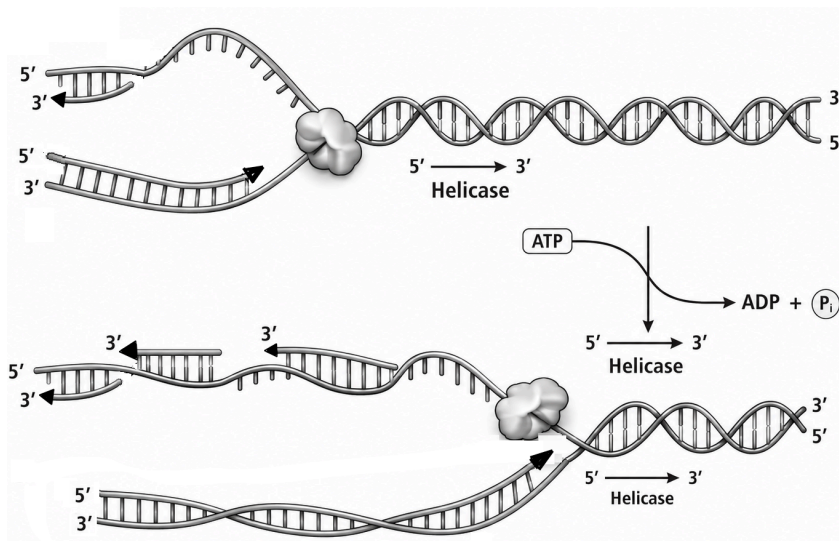
Initiator proteins recognise the origin and set up the machinery. In *E. coli* the origin (*oriC*) is recognised by DnaA; in yeast the origins (called ARS) are recognised by the six-protein origin recognition complex, ORC. All initiators share a common ATP-using AAA+ motif.

Helicase loaders hand the helicase to the origin. They recognise the initiator already bound there and load the helicase around the DNA. In bacteria this is done by DnaC; in eukaryotes the job is shared by Cdc6 and Cdt1.

Unwinding the helix: helicases and their polarity

A helicase is the motor that pulls the double helix apart, using the energy of ATP to melt the duplex into two single strands just ahead of the fork. *E. coli* uses DnaB; eukaryotes use the Mcm2–7 complex. An important detail is polarity — the direction a helicase travels and which strand it rides on. DnaB moves 5'→3' along the lagging-strand template, whereas the eukaryotic Mcm2–7 (and the SV40 T antigen) move 3'→5' along the leading-strand template (Figure 1.7).

Figure 1.7. Helicase polarity: *DnaB* tracks 5'→3' on the lagging-strand template, while eukaryotic Mcm2–7 tracks 3'→5' on the leading-strand template.



Keeping strands apart: single-stranded DNA-binding proteins

Freshly separated strands are highly complementary and would happily snap back together. To prevent this, single-stranded DNA-binding proteins (SSB in bacteria, RPA in eukaryotes) coat the exposed strands the instant they appear. Their binding is cooperative — each bound protein helps the next one bind alongside it — so a strand is quickly wrapped as it emerges from the helicase. They bind the phosphate backbone regardless of sequence, hold the strand in an extended shape ready for copying, and are displaced as the polymerase moves through.

The DNA polymerases

Bacteria have five DNA polymerases and eukaryotes at least fifteen, but they all share one chemistry: they add nucleotides only to a free 3'–OH, building 5'→3'. Table 1.2 summarises them. A few deserve special mention.

Table 1.2. The DNA polymerases of *E. coli* and of eukaryotes, with their main roles.

E. coli polymerase	Main role	Eukaryotic polymerase	Main role
Pol I	Primer removal, gap filling, repair	Pol α	Primer synthesis (with primase)
Pol II	DNA repair	Pol β	Base-excision repair
Pol III (core)	Chromosome replication	Pol γ	Mitochondrial DNA replication
Pol III (holoenzyme)	Chromosome replication	Pol δ	Lagging-strand synthesis
Pol IV (DinB)	Translesion synthesis	Pol ε	Leading-strand synthesis
Pol V (UmuD' ₂ C)	Translesion synthesis	Pol η, ι, κ, ζ, Rev1	Translesion synthesis / repair

DNA polymerase I was the first ever purified. It is the most abundant polymerase in *E. coli* but not the main replicative one; its job is to remove RNA primers and fill the gaps between Okazaki fragments, and to help in repair. Cutting it with a protease yields the large Klenow fragment (5'→3' polymerase plus 3'→5' proofreading) and a small fragment carrying the 5'→3' exonuclease used for primer removal. The Klenow fragment is a laboratory workhorse, used to label DNA by nick translation and random priming.

DNA polymerase III is the true replicase of *E. coli*. It is a holoenzyme of ten different subunits; its α-subunit does the 5'→3' polymerisation and its ε-subunit provides 3'→5' proofreading.

Polymerases II, IV and V handle repair and 'translesion' synthesis — copying past damage that has stalled Pol III. Pol IV and Pol V belong to the error-prone Y-family and can introduce mutations, which is the price of getting past a lesion.

Eukaryotic replicative polymerases divide the labour: Pol α starts a strand (its primase makes the RNA primer, which Pol α then extends with a short DNA stretch), Pol ε copies the leading strand and Pol δ the lagging strand, while Pol γ copies mitochondrial DNA. Bacteria copy DNA at about 500–1000 nucleotides per second; mammalian forks run slower, roughly 50–100 per second, largely because the DNA must be unpacked from histones as it is read.

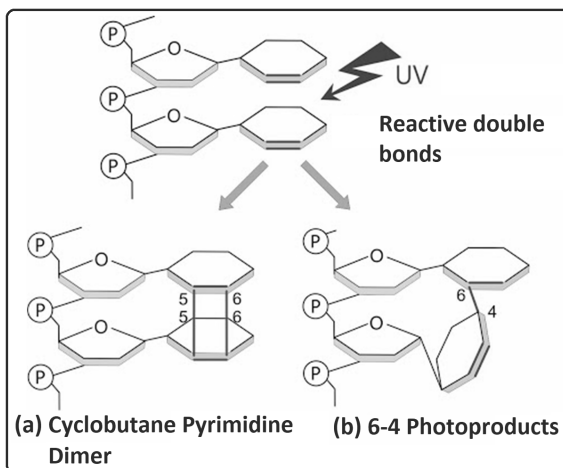
Base analogues and intercalators. Base analogues such as 5-bromouracil resemble a normal base closely enough to be incorporated during replication, but then mispair (5-bromouracil pairs with guanine via its enol form). Intercalating agents — flat molecules such as proflavin, acridine and ethidium bromide — slot between the stacked bases and cause insertions or deletions, hence frameshift mutations.

Physical agents

Ionizing radiation (α , β , γ and X-rays) breaks the phosphodiester backbone directly, causing single- and double-strand breaks — the double-strand breaks give serious chromosomal rearrangements (duplications, deletions, inversions, translocations). Its breaks often carry unusual 3'-phosphate or 3'-phosphoglycolate ends, and it also damages DNA indirectly by generating ROS. This lethality is harnessed in radiotherapy.

Ultraviolet (UV) radiation is the leading cause of skin cancer. DNA absorbs UV maximally at 260 nm; the most harmful band, UV-C, is largely filtered by the ozone layer. UV makes adjacent pyrimidines on the same strand fuse into cyclobutane pyrimidine dimers (about 75% of lesions, commonest between two thymines) and (6–4) photoproducts between adjacent pyrimidines (Figure 2.12). UV-A acts more indirectly, through photo-oxidation and photosensitisers.

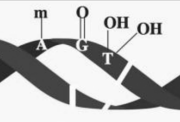
Figure 2.12. UV light causes (a) cyclobutane pyrimidine dimers and (b) (6–4) photoproducts between adjacent pyrimidines.



2.4 Direct Reversal and Mismatch Repair

A cell that could not repair the daily load of damage would die within hours, so it deploys a toolkit of pathways, each matched to a class of lesion (Figure 2.13). The simplest simply reverse the chemical change; most others cut the damage out and resynthesise the missing DNA using the intact complementary strand as a guide. The observed mutation rate is the balance between how much damage occurs and how much is correctly repaired.

Figure 2.13. An overview of the common types of DNA damage and the repair pathways that deal with them.

DNA damaging agents	Toxins Alkylating agents Base deamination Replication errors	Oxidative damage Electrophiles	Ionizing radiation UV radiation Crosslinking agent Aromatic compounds Heat Cold Hypoxia
Damaged DNA	 Mismatches Uracil Abasic sites Adducts	 Lesions Single strand break Double strand break	 Bulky lesions Intra- & Interstrand crosslink Single strand break Double strand break
DNA repair pathways	Mismatch repair Base excision repair	Base excision repair Single strand break repair Double strand break repair	Nucleotide excision repair Interstrand crosslink Repair Single strand break repair Double strand break repair Translesion synthesis

In eukaryotes (Figures 2.16–2.17), the logic is the same but the players differ. Slippage at microsatellite repeats produces insertion–deletion loops (Figure 2.16). The MutS α heterodimer (MSH2/MSH6) recognises base mismatches and small loops, while MutS β (MSH2/MSH3) handles larger loops; the MutL α heterodimer (MLH1/PMS2) is then recruited, and its endonuclease directs Exonuclease 1 to remove the error in a PCNA/RFC-dependent way, after which polymerase δ and Ligase 1 finish the job (Figure 2.17). Eukaryotes do not use methylation to pick the new strand — PCNA left at the growing end appears to mark it. Because MMR also guards microsatellite stability, inherited defects in these genes cause Lynch syndrome (hereditary non-polyposis colorectal cancer, HNPCC).

Figure 2.16. Strand slippage at a repetitive sequence produces an insertion–deletion loop.

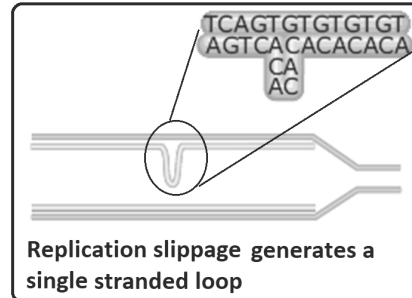


Figure 2.17. The human mismatch-repair system: MutS α/β recognition and MutL α -directed excision.

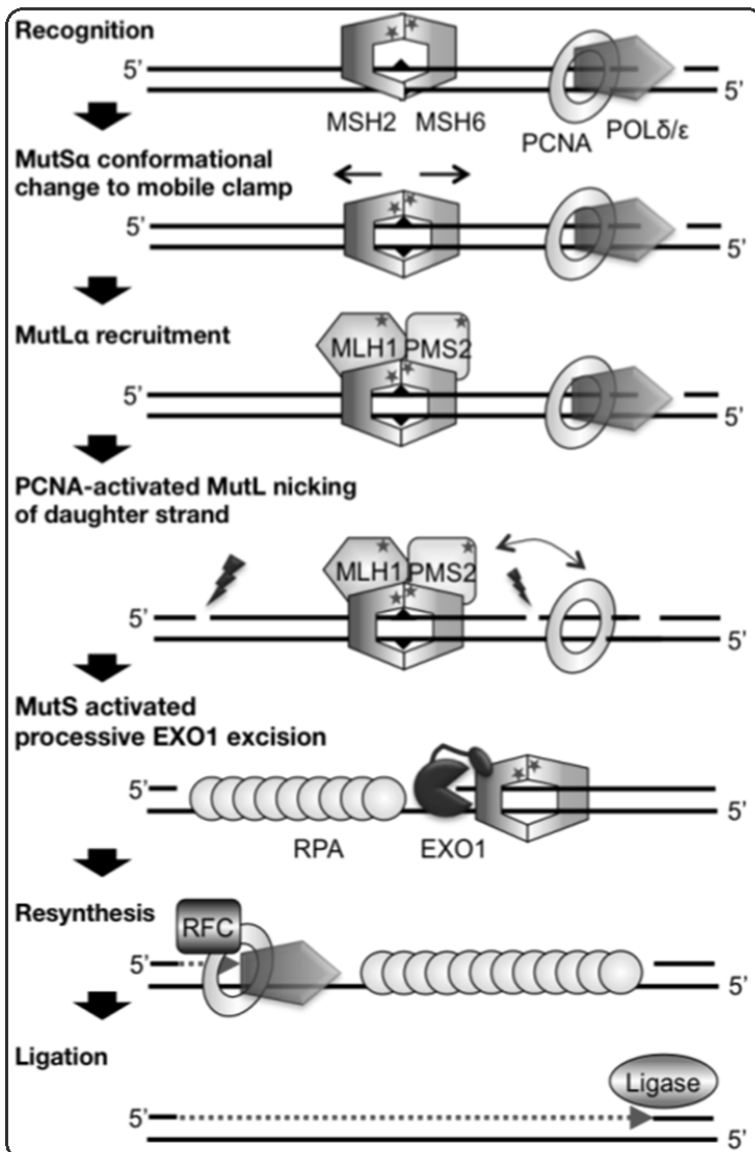
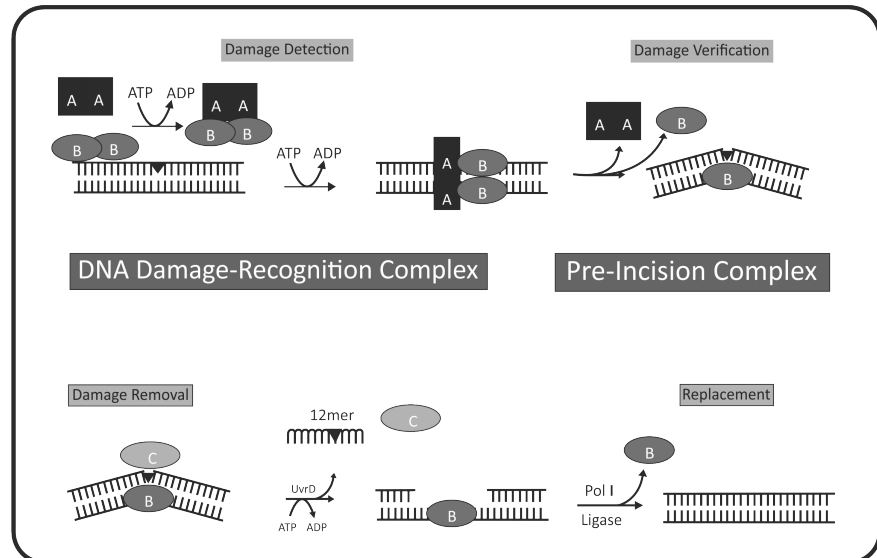


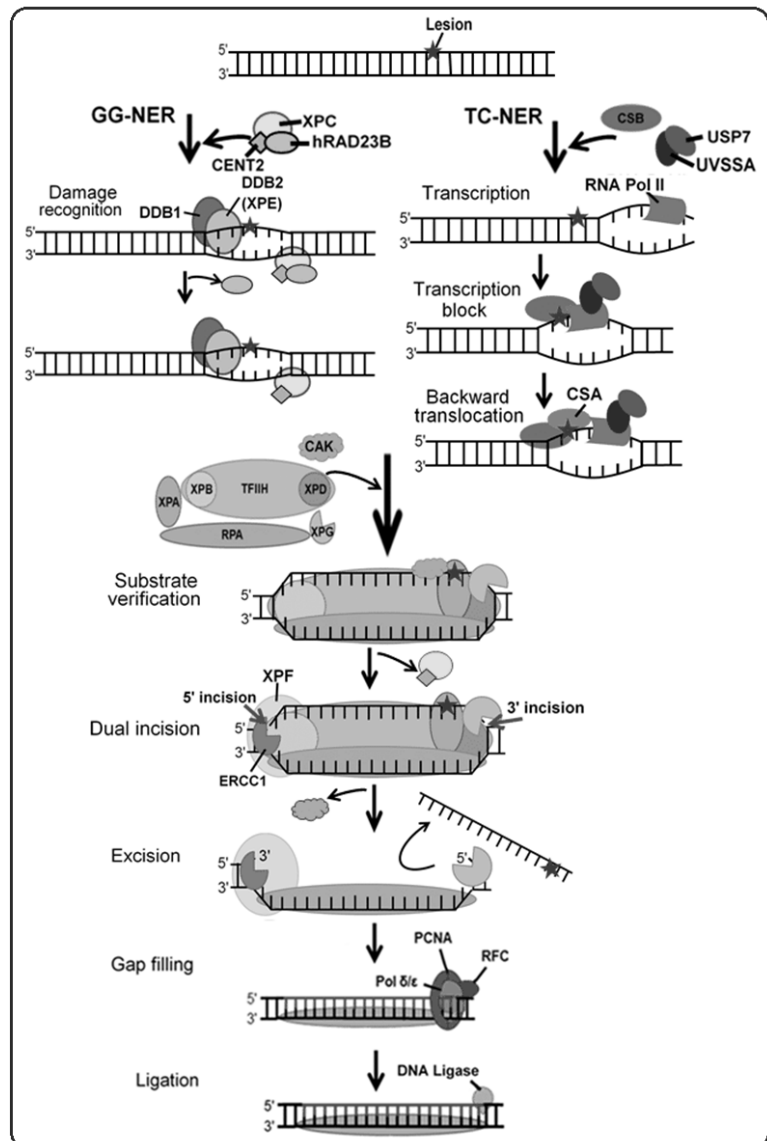
Figure 2.22. Bacterial nucleotide-excision repair by the UvrABC excinuclease and UvrD helicase.



NER mechanism in bacteria

In eukaryotes (Figure 2.23, Table 2.2), NER is more elaborate and, when defective, causes xeroderma pigmentosum (XP) — extreme UV sensitivity and skin cancer — as well as Cockayne syndrome. It runs in two modes. Global-genome NER (GG-NER) uses the XPC–RAD23B complex (helped by the UV-DDB complex for UV dimers) to detect distortions anywhere in the genome. Transcription-coupled NER (TC-NER) is triggered when RNA polymerase II stalls at a lesion in an active gene, and uses the CSA and CSB proteins. Both modes then converge: the TFIIH complex opens the DNA, and the endonucleases XPF–ERCC1 (5') and XPG (3') excise the damaged patch, which PCNA, RFC, polymerase $\delta/\epsilon/\kappa$ and ligase fill and seal.

Figure 2.23. Eukaryotic nucleotide-excision repair, showing the GG-NER and TC-NER branches converging on TFIIH and dual incision.



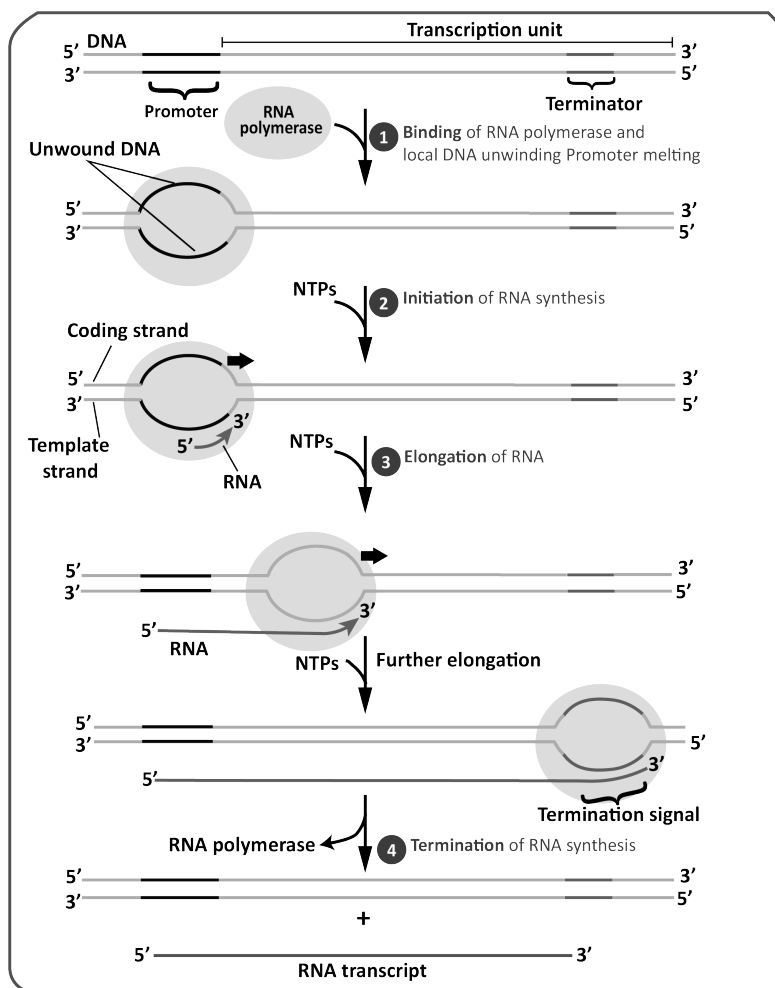
Fidelity, and why it matters less than in replication

Transcription is accurate but less so than replication — about one error per 10^4 nucleotides, versus one in 10^9 for replication — because its proofreading is more limited (two forms exist, pyrophosphate editing and hydrolytic editing, which we meet in Section 3.4). This lower fidelity is tolerable: a replication error becomes a permanent, heritable mutation, whereas a transcription error spoils only one of many short-lived RNA copies and is soon gone.

The transcription cycle

Transcription proceeds in three stages (Figure 3.2). In initiation, RNA polymerase binds a promoter, opens the DNA into a 'bubble', and starts joining ribonucleotides. In elongation, the enzyme moves along the template, unwinding the DNA ahead and resealing it behind, adding one nucleotide at a time and peeling the RNA away. In termination, the polymerase reaches a stop signal (a terminator), releases the RNA, and lets go of the DNA.

Figure 3.2. *The transcription cycle: initiation, elongation and termination.*



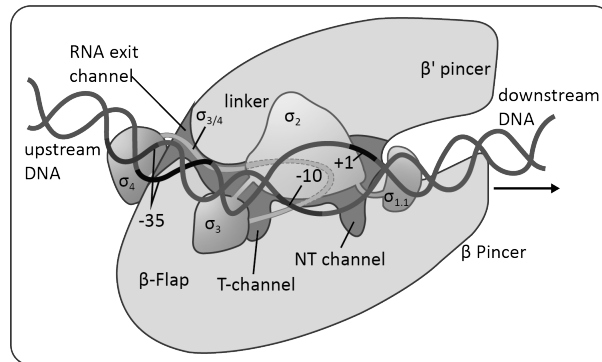
3.2 The Many Types of RNA

The central dogma once cast RNA as a passive middleman — mRNA carrying the message, tRNA bringing amino acids, and rRNA building the ribosome. We now know RNA does far more: many RNAs are catalytic or regulatory, and their variety hints that an 'RNA world' may have preceded DNA and protein. RNAs that are not translated are called non-coding RNAs (ncRNAs).

Small nuclear RNAs (snRNAs) work in splicing. They sit in the nucleus bound to proteins as snRNPs ('snurps'); the U1, U2, U5 and U4/U6 particles carry out the removal of introns from pre-mRNA.

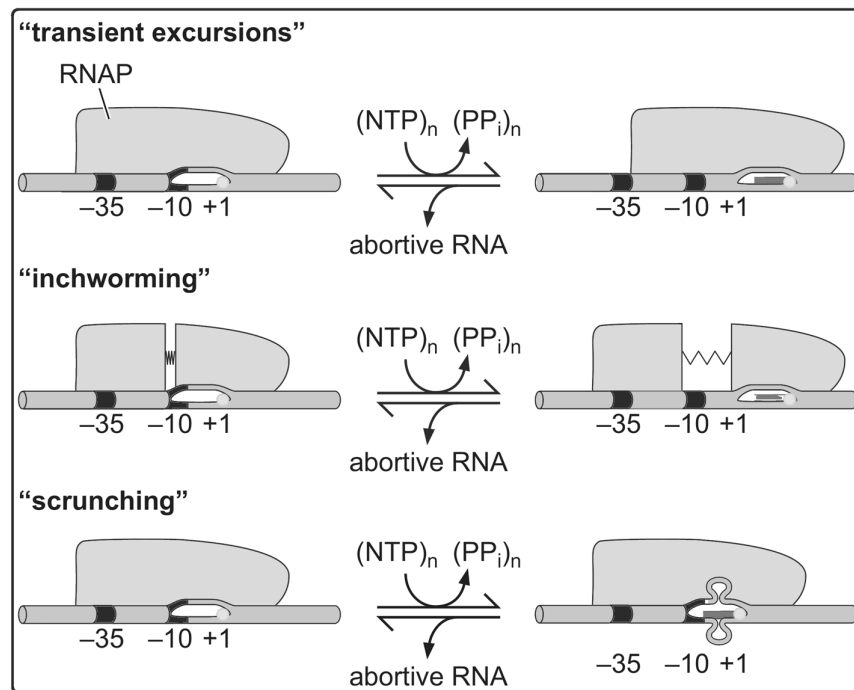
Small nucleolar RNAs (snoRNAs) guide the chemical modification of ribosomal RNA in the nucleolus. The C/D box family directs 2'-O-methylation and the H/ACA box family directs pseudouridylation. Small Cajal-body RNAs (scaRNAs) do the same job for the spliceosomal snRNAs.

Figure 3.8. The RNA polymerase–DNA open complex, with the template strand in the active-centre cleft.



Then comes a curious hurdle. Before it commits to the full transcript, the polymerase makes and releases many abortive short RNAs (~10 nucleotides). Single-molecule experiments show this happens by 'scrunching': the enzyme stays put on the promoter and pulls downstream DNA into itself, storing energy as single-stranded bulges (Figure 3.9). Once a transcript exceeds ~10 nucleotides, the stored energy is released, σ is displaced (its region 3/4 linker is ejected from the RNA-exit channel), and the polymerase escapes the promoter as a stable ternary complex of enzyme, DNA and RNA — the start of elongation.

Figure 3.9. Models of abortive initiation; single-molecule data favour the 'scrunching' model.



Elongation and proofreading

During elongation the DNA enters between the pincers, the strands part in the cleft, and the RNA stays paired to the template for only ~8–9 nucleotides before exiting; the enzyme steps forward one base pair per nucleotide added, like a molecular motor. It proofreads in two ways. In pyrophosphorolytic editing it backtracks by one nucleotide and removes a mis-inserted base by re-adding pyrophosphate (it dwells longer over mismatches, so removes them preferentially). In hydrolytic editing it backtracks further and cleaves off the error-containing end, a reaction stimulated by the Gre factors (the eukaryotic equivalent is TFIIIS). Elongation factors such as NusA and NusG keep the processive enzyme from stalling; NusG also links transcription to termination and, in bacteria, to the ribosome.

Transcription–translation coupling. Because bacteria have no nuclear membrane, a ribosome can begin translating an mRNA while it is still being transcribed. RNA polymerase and the leading ribosome are physically connected — a NusG-family protein can serve as the adapter — so the two processes influence each other directly.

Clearing stalled polymerase. If the enzyme arrests at a DNA lesion it blocks both gene expression and other polymerases. The transcription-repair coupling factor Mfd (TRCF), a DNA translocase, pushes the stalled enzyme off and recruits the NER protein UvrA to the exposed lesion — the basis of transcription-coupled repair.

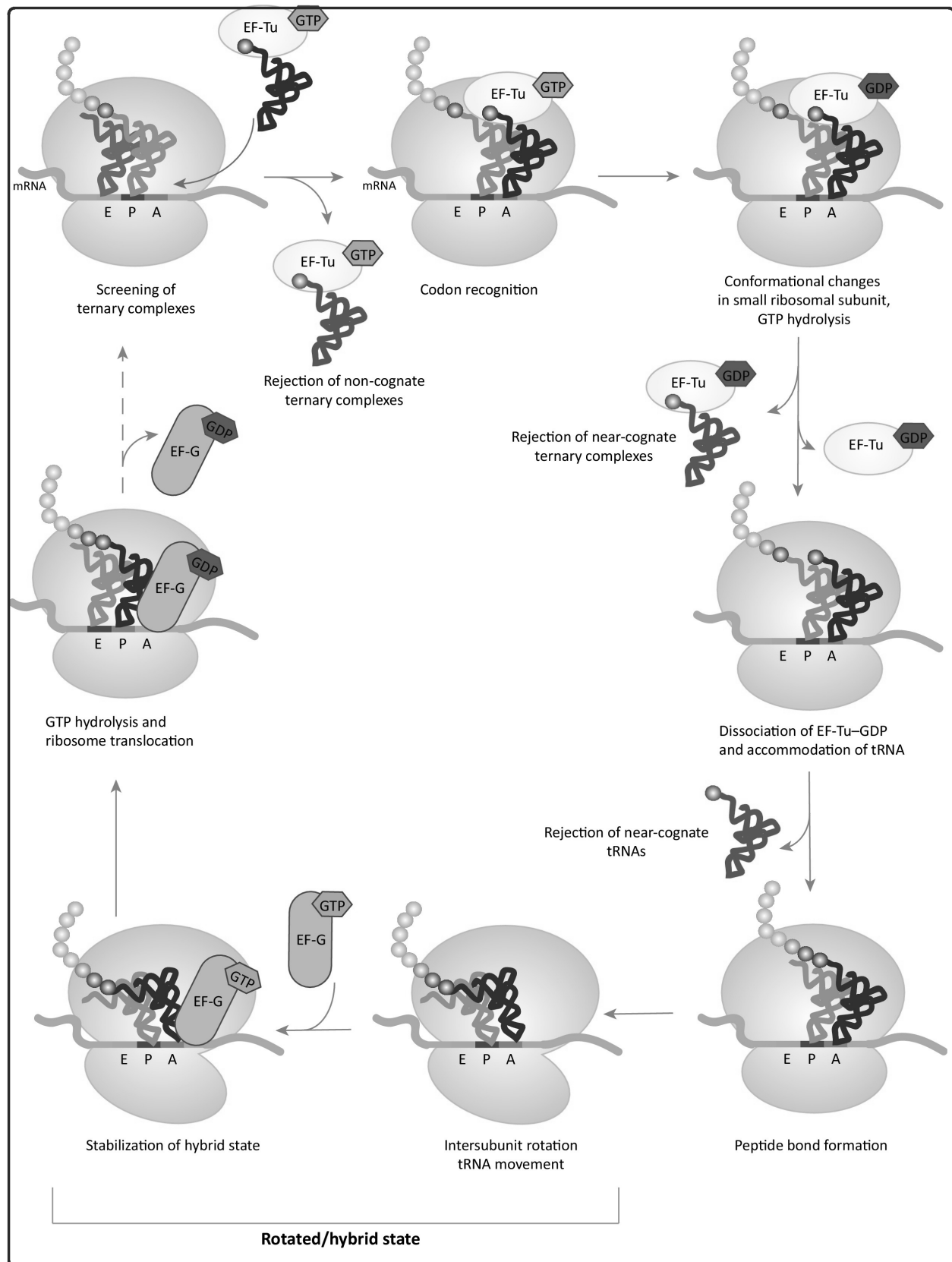
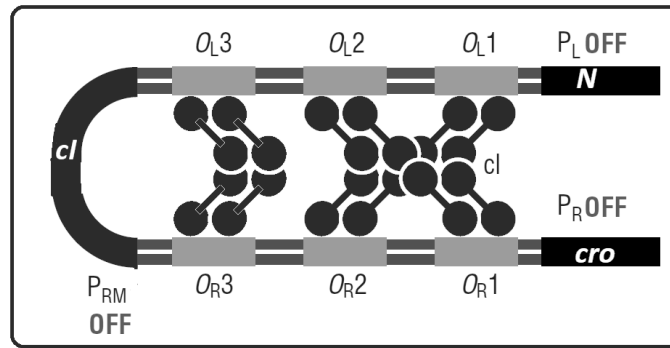


Figure 5.29. *The detailed steps of prokaryotic translation elongation.*

Cost and unusual amino acids. Each peptide bond costs one ATP (charging) plus two GTP (EF-Tu delivery and EF-G translocation) — only the ATP is used for the bond itself; the GTPs buy accuracy and order. Pyrrolysine, the 22nd amino acid, is inserted at a UAG codon by a dedicated tRNA and synthetase (Figure 5.30).

Figure 6.41. At high concentration, λ repressor binds O_R3 and turns off its own gene (negative autoregulation).



Inducing the switch to lysis

DNA damage flips a lysogen into lysis. Damage activates RecA, which triggers self-cleavage of the SOS repressor LexA — and λ repressor has evolved to mimic LexA (molecular mimicry), so it too is cleaved. Losing its C-terminal domain, the repressor can no longer dimerise cooperatively and falls off O_{R1}/O_{R2} ; P_R and P_L fire, Cro is made and takes O_{R3} to block any new repressor, and the decision to go lytic becomes irreversible.

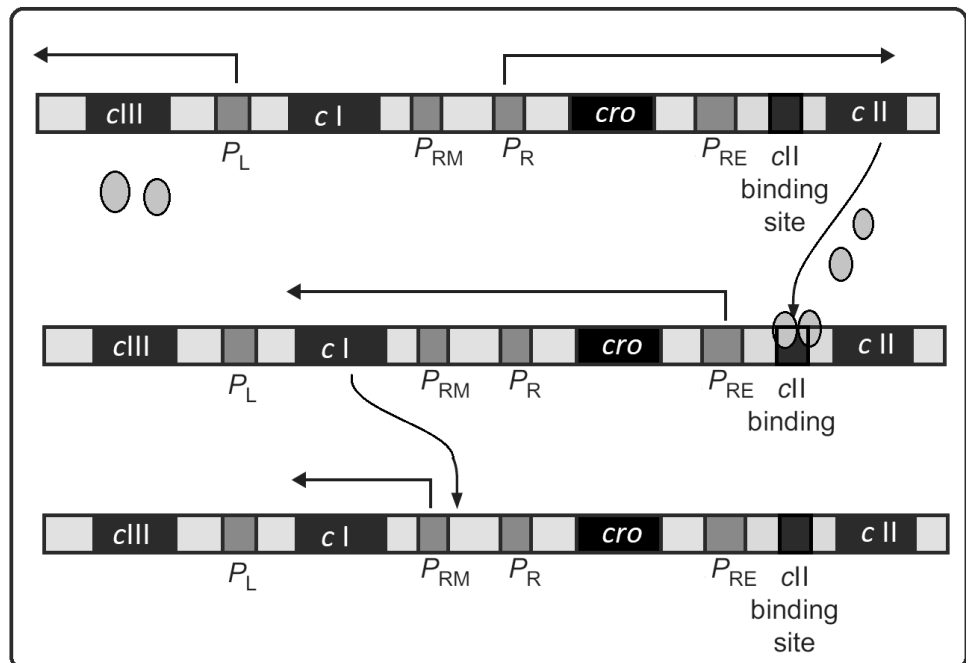
The lysis/lysogeny decision on first infection

On a fresh infection the outcome hinges on the CII protein (Figures 6.42–6.43). P_R and P_L fire immediately, making Cro (favouring lysis) and CII/CIII (favouring lysogeny). CII is an activator that switches on P_{RE} to make a burst of λ repressor, and on P_{int} to make integrase for prophage integration. The balance is set by how long CII survives: it is rapidly destroyed by the host protease FtsH (HflB), and CIII protects it. When the host grows well FtsH is active, CII is destroyed, and the phage goes lytic; when the host grows poorly FtsH is sluggish, CII persists, repressor accumulates, and the phage lysogenises. A high multiplicity of infection (MOI — several phages per cell) also favours lysogeny.

Figure 6.42. The genes and promoters that decide between lysis and lysogeny.



Figure 6.43. Establishment of lysogeny: CII activates P_{RE} to make repressor, which then maintains itself from P_{RM} .



Anti-termination by N and Q

Two λ proteins, N and Q, are anti-terminators — they let RNA polymerase read through terminators it would otherwise stop at (Figures 6.44–6.45).

Transcripts from P_R and P_L terminate early unless N modifies the polymerase; N binds not DNA but the nascent RNA at nut (N-utilization) sites, then loads onto the polymerase (with host Nus factors including NusA and the ribosomal protein S10/NusE), making it terminator-resistant so the early genes are fully transcribed.

Q acts later: it binds a DNA site (QBE) near the late promoter $P_{R'}$ and, once polymerase pauses just after starting, transfers onto it and lets it transcribe the large (~26 kb) late operon. (The pause depends on σ recognising a -10-like sequence.)

Figure 6.44. The recognition sites and action of the N and Q anti-terminators.

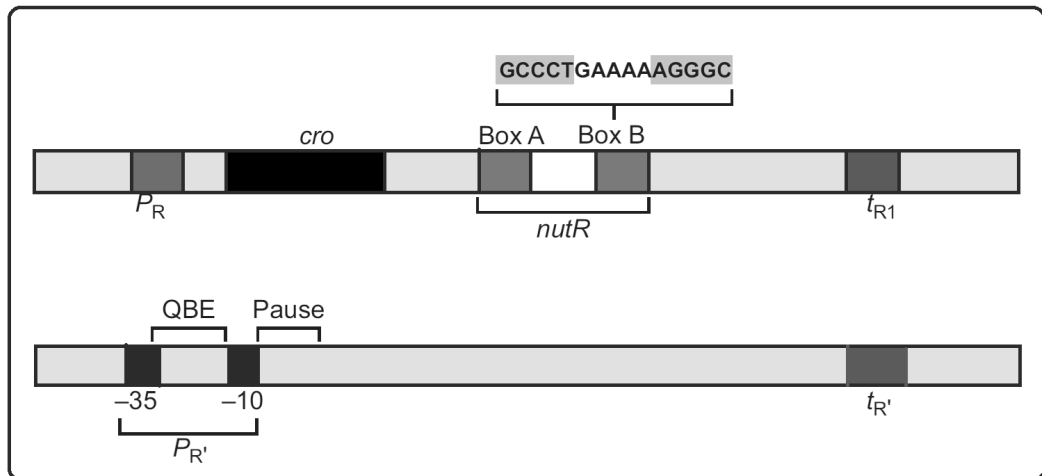
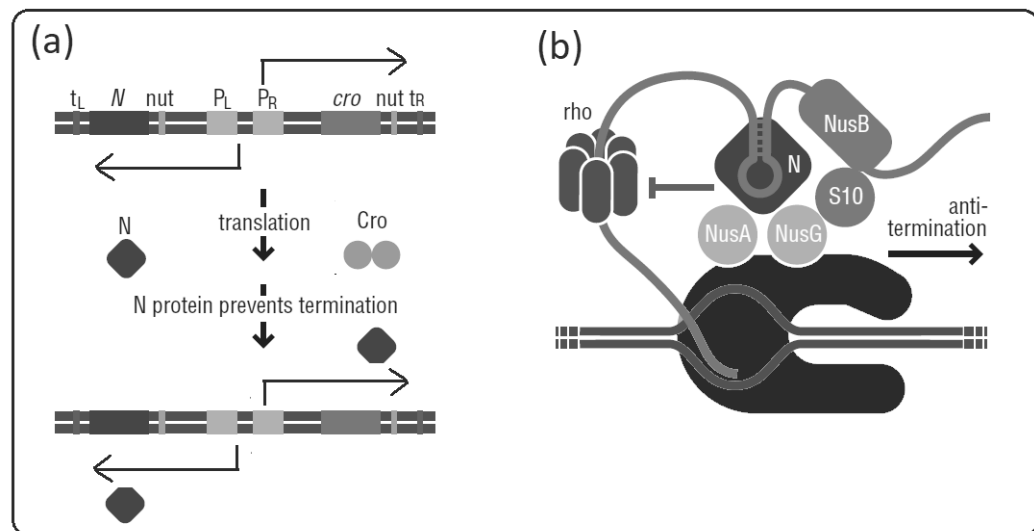


Figure 6.45. Anti-termination by the N protein loaded onto RNA polymerase at a nut site.



Retro-regulation of int. The *int* gene (integrase) is transcribed from both P_L and its own CII-activated promoter P_{int} , yet integrase is made efficiently only from P_{int} (Figure 6.46).

The reason is the 3' end of the mRNA: the P_{int} transcript ends at a stem-loop terminator and is stable, but the P_L transcript — made terminator-resistant by N — runs past that point, forming a longer 3' structure that host nucleases attack, degrading the message backward through the gene.

Because this control site lies downstream of the gene, it is called retro-regulation — ensuring integrase is made for lysogeny but not wasted during lysis.