

the book is central to life sciences, and that it should be accessible to students at introductory levels, the book has been entirely freshly illustrated; and indeed the illustrations extend now to overviews and summaries that depict virtually every major point of the book (conceptual as well as factual) in diagrammatic form.

It is as always a pleasure to thank colleagues who have read chapters or discussed individual topics (in many cases stimulating new directions of thought and much improving the book), including particularly Tania Baker, Michael Chamberlin, Albert Dahlberg, Douglas Engel, Martin Gellert, Tony Hunter, Michael Levine, Richard Losick, Ira Mellman, Paul Nurse, Paul Schimmel, Matthew Scott, Philip Sharp, Robert Thach, and Andrew Travers.

GENES retains the aim of providing a *tour d'horizon* of the current state of the art, and also identifying the questions that need now to be answered. By keeping the current version of the text up to date with the latest facts and thoughts in the field, I hope to make it possible for both teachers and students to keep abreast of modern genetics.

Cambridge, Massachusetts
May 1993

B.L.

Outline

INTRODUCTION

Cells as macromolecular assemblies 1

- 1 Cells obey the laws of physics and chemistry 3
- 2 Cells are organized into compartments 29

PART 1

DNA as a store of information 57

- 3 Genes are mutable units 59
- 4 DNA is the genetic material 81
- 5 The topology of nucleic acids 109
- 6 Isolating the gene 127

PART 2

Translation: expressing genes as proteins 161

- 7 The assembly line for protein synthesis 163
- 8 Transfer RNA is the translational adaptor 197
- 9 Ribosomes provide a translation factory 233
- 10 Messenger RNA is the template 253

PART 3

Constructing the cell 277

- 11 The apparatus for protein localization 279
- 12 Receptors and signal transduction: channels and ion uptake 319
- 13 Cell cycle and growth regulation 349

PART 4

Control of prokaryotic gene expression 375

- 14 Control at initiation: RNA polymerase-promoter interactions 377
- 15 A panoply of operons: the lactose paradigm and others 413
- 16 Control by RNA structure: termination and antitermination 457
- 17 Phage strategies: lytic cascades and lysogenic repression 491

PART 5

Perpetuation of DNA 525

- 18 The replicon: unit of replication 527
- 19 Primosomes and replisomes: the apparatus for DNA replication 571
- 20 Systems that safeguard DNA 605

PART 6

Organization of the eukaryotic genome 631

- 21 The extraordinary power of DNA technology 633
- 22 Genome size and genetic content 657
- 23 The eukaryotic gene: conserved exons and unique introns 677
- 24 Gene numbers: repetition and redundancy 703
- 25 Genomes sequestered in organelles 733
- 26 Organization of simple sequence DNA 749
- 27 The genome is packaged into chromosomes 767
- 28 Chromosomes consist of nucleosomes 797

Outline

PART 7 Eukaryotic transcription and RNA processing 845

- 29 Building the transcription complex: promoters, factors, and RNA polymerases 847
- 30 Regulation of transcription: factors that activate the basal apparatus 879
- 31 The apparatus for nuclear splicing 911
- 32 RNA as catalyst: changing the informational content of RNA 941

PART 8 The dynamic genome: DNA in flux 965

- 33 Recombination of DNA 967
- 34 Transposons that mobilize via DNA 999
- 35 Retroviruses and retroposons 1033
- 36 Rearrangement and amplification in the genome 107

PART 9 Genes in development 1101

- 37 Generation of immune diversity by gene reorganization 1103
- 38 Gene regulation in development: gradients and cascades 1141
- 39 Oncogenes: gene expression and cancer 1181

EPILOGUE Landmark shifts in perspectives 1231

Contents

INTRODUCTION

Cells as macromolecular assemblies 1

CHAPTER 1 3

Cells obey the laws of physics and chemistry

- Macromolecules are assembled by polymerizing small molecules 6
- Proteins consist of chains of amino acids 9
- Protein conformation depends upon the aqueous environment 14
- Protein structures are extremely versatile 18
- How do proteins fold into the correct conformation? 21

CHAPTER 2 29

Cells are organized into compartments

- Cellular compartments are bounded by membranes 31
- The cytoplasm contains networks of membranes 37
- Cell shape is determined by the cytoskeleton 40
- Some organelles are surrounded by an envelope 43
- The environment of the nucleus and its reorganization 46
- The role of chromosomes in heredity 48

PART 1	DNA as a store of information	57
CHAPTER 3		59
Genes are mutable units	Discovery of the gene	62
	Genes lie in a linear array on chromosomes	65
	One gene—one protein: the basic paradigm	70
	A modern definition: the cistron	72
	Mapping mutations at the molecular level	74
	The nature of multiple alleles	76
CHAPTER 4		81
DNA is the genetic material	The discovery of DNA	82
	DNA is the (almost) universal genetic material	84
	The components of DNA	87
	DNA is a double helix	91
	DNA replication is semiconservative	94
	The genetic code is read in triplets	98
	Mutations change the sequence of DNA	101
	Mutations are concentrated at hotspots	105
	The rate of mutation	106
CHAPTER 5		109
The topology of nucleic acids	DNA can be denatured and renatured	110
	Nucleic acids hybridize by base pairing	111
	Single-stranded nucleic acids may have secondary structure	115
	Inverted repeats and secondary structure	117
	Duplex DNA has alternative double-helical structures	119
	Closed DNA can be supercoiled	122
	Supercoiling influences the structure of the double helix	124
CHAPTER 6		127
Isolating the gene	A restriction map can be constructed by cleaving DNA into specific fragments	129
	Restriction sites can be used as genetic markers	134
	Obtaining the sequence of DNA	142
	Prokaryotic genes and proteins are colinear	146
	Eukaryotic genes can be interrupted	149
	Some DNA sequences code for more than one protein	151
	Genetic information can be provided by DNA or RNA	154
	The scope of the paradigm	157

PART 2	Translation: expressing genes as proteins	161
CHAPTER 7		163
The assembly line for protein synthesis	Transfer RNA is the adaptor	165
	Messenger RNA is translated by ribosomes	167
	The meaning of the genetic code	171
	The ribosomal sites of action	174
	Initiation in bacteria needs 30S subunits and accessory factors	176
	A special initiator tRNA starts the polypeptide chain	179
	Eukaryotic initiation involves many factors	183
	Elongation factor T brings aminoacyl-tRNA into the A site	185
	Translocation moves the ribosome	188
	Finishing off: three codons terminate protein synthesis	192
CHAPTER 8		197
Transfer RNA is the translational adaptor	The universal cloverleaf	198
	The tertiary structure is L-shaped	200
	Synthetases fall into two classes that recognize similar features in tRNA	202
	Discrimination in the charging step	207
	Codon-anticodon recognition involves wobbling	211
	tRNA contains many modified bases	213
	Base modification may control codon recognition	215
	The genetic code is altered in ciliates and mitochondria	217
	Suppressor tRNAs have mutated anticodons that read new codons	219
	tRNA may influence the reading frame	224
	tRNA transcripts are cut and trimmed from clusters by several enzymes	227
CHAPTER 9		233
Ribosomes provide a translation factory	Ribosomes are compact particles in which most proteins interact with rRNA	234
	Subunit assembly is linked to topology	239
	The role of ribosomal RNA in protein synthesis	242
	Ribosomes have several active centers	246
	The accuracy of translation	249
CHAPTER 10		253
Messenger RNA is the template	The lifecycle of messenger RNA	254
	Most bacterial genes are expressed via polycistronic messengers	258

The translation of eukaryotic mRNA	262
Most eukaryotic mRNAs are polyadenylated at the 3' end	264
All eukaryotic mRNAs have a methylated cap at the 5' end	266
Initiation involves base pairing between mRNA and rRNA	268
Small subunits migrate to initiation sites on eukaryotic mRNA	270
Processing is necessary to produce some RNAs	271
Stability of mRNA is determined by particular sequences	273

PART 3

CHAPTER 11
The apparatus for protein localization

Constructing the cell 277

Post-translational membrane insertion depends on leader sequences	282
Leaders determine protein location within mitochondria and chloroplasts	285
Signal sequences link protein synthesis to membranes during co-translational transfer	289
Anchor sequences cause proteins to be retained in membranes	294
Bacterial proteins are transported by both co-translational and post-translational mechanisms	298
Oligosaccharides are added to proteins in the endoplasmic reticulum and Golgi	300
Coated vesicles transport both exported and imported proteins	305
Protein localization depends on further signals	314

CHAPTER 12
Receptors and signal transduction: channels and ion uptake

A more detailed view of the plasma membrane	323
Receptors recycle via endocytosis	326
Protein tyrosine kinases induce phosphorylation cascades	331
G proteins may activate or inhibit target proteins	334
Carriers and channels form water-soluble paths through the membrane	336
Pores control nuclear ingress and egress	342

CHAPTER 13
Cell cycle and growth regulation

Replication and mass cycles are coordinated	351
Regulatory activities are found at S phase and at M phase	353
M-phase kinase is a dimer that regulates entry into mitosis	355
Protein phosphorylation and dephosphorylation control the cell cycle	359
p34 is the key regulator in yeasts	361
CDC28 acts at both START and mitosis in <i>S. cerevisiae</i>	367
What controls the G1-S transition?	369

PART 4**CHAPTER 14****Control at initiation: RNA polymerase–promoter interactions****Control of prokaryotic gene expression 375****377**

Transcription is catalyzed by RNA polymerase	379
Bacterial RNA polymerase consists of core enzyme and sigma factor	383
Sigma factor controls binding to DNA	387
Promoter recognition depends on consensus sequences	393
RNA polymerase binds to one face of DNA	396
Substitution of sigma factors may control initiation	401
Sporulation utilizes a cascade of many sigma factors	404

CHAPTER 15**A panoply of operons: the lactose paradigm and others****413**

Structural gene clusters are coordinately controlled	415
The activity of repressor protein is controlled by a small molecule inducer	418
Mutations identify the operator and the regulator gene	421
Repressor protein binds to the operator and is released by inducer	427
The specificity of protein–DNA interactions	432
Repression can occur at multiple loci	435
Distinguishing positive and negative control	437
Catabolite repression involves positive regulation at the promoter	439
Autogenous control may occur at the level of translation	445
Hard times provoke the stringent response	450

CHAPTER 16**Control by RNA structure: termination and antitermination****457**

Bacterial RNA polymerase has two modes of termination	460
How does rho factor work?	462
Antitermination depends on specific sites	465
More subunits for RNA polymerase	470
Alternative secondary structures control attenuation	473
Small RNA molecules can regulate translation	479
Regulation by cleavage of mRNA	482
Cleavages are needed to release prokaryotic and eukaryotic rRNAs	484

CHAPTER 17**Phage strategies: lytic cascades and lysogenic repression****491**

Lytic development is controlled by a cascade	494
Functional clustering in phages T7 and T4	496
The lambda lytic cascade relies on antitermination	499
Lysogeny is maintained by an autogenous circuit	503

The DNA-binding form of repressor is a dimer	506
Repressor binds cooperatively at each operator using a helix-turn-helix motif	508
How is repressor synthesis established?	515
A second repressor is needed for lytic infection	519
A delicate balance: lysogeny versus lysis	521

PART 5**Perpetuation of DNA****525****CHAPTER 18****The replicon: unit of replication**

Origins can be mapped by autoradiography and electrophoresis	529
The bacterial genome is a single replicon	532
Each eukaryotic chromosome contains many replicons	535
Isolating the origins of yeast replicons	537
D loops may be maintained at mitochondrial origins	539
The problem of linear replicons	540
Rolling circles produce multimers of a replicon	544
Single-stranded genomes are generated for bacterial conjugation	549
Connecting bacterial replication to the cell cycle	553
Cell division and chromosome segregation	555
Multiple systems ensure plasmid survival in bacterial populations	560
Plasmid incompatibility is connected with copy number	563

CHAPTER 19**Primosomes and replisomes: the apparatus for DNA replication**

DNA polymerases: the enzymes that make DNA	572
DNA synthesis is semidiscontinuous and primed by RNA	579
The primosome initiates synthesis of Okazaki fragments	582
Coordinating synthesis of the lagging and leading strands	586
The replication apparatus of phage T4	592
Creating the replication forks at an origin	594
Common events in priming replication at the origin	597
Does methylation at the origin regulate initiation?	600

CHAPTER 20**Systems that safeguard DNA**

The consequences of modification and restriction	606
Type II restriction enzymes are common	608
The alternative activities of type I enzymes	609
The dual activities of type III enzymes	613
Dealing with injuries in DNA	614
Excision-repair systems in <i>E. coli</i>	618
Controlling the direction of mismatch repair	621
Retrieval systems in <i>E. coli</i>	623
An SOS system of many genes	625
Mammalian repair systems	628

PART 6**Organization of the eukaryotic genome****631****CHAPTER 21****The extraordinary power of DNA technology**

Any DNA sequence can be cloned in bacteria or yeast	634
Constructing the chimeric DNA	636
Copying mRNA into cDNA	640
Isolating individual genes from the genome	642
Walking along the chromosome	647
Eukaryotic genes can be expressed in prokaryotic systems	652

CHAPTER 22**Genome size and genetic content**

The C-value paradox describes variations in genome size	658
Reassociation kinetics depend on sequence complexity	660
Eukaryotic genomes contain several sequence components	663
Nonrepetitive DNA complexity can estimate genome size	664
Eukaryotic genomes contain repetitive sequences that are related but not identical	666
Most structural genes lie in nonrepetitive DNA	668
How many nonrepetitive genes are expressed?	671
Genes are expressed at widely varying levels	674

CHAPTER 23**The eukaryotic gene: conserved exons and unique introns**

Organization of interrupted genes may be conserved	679
Genes show a wide distribution of sizes	682
One DNA sequence may code for multiple proteins	688
Exon sequences are conserved but introns vary	690
Genes can be isolated by the conservation of exons	691
How do interrupted genes evolve?	695

CHAPTER 24**Gene numbers: repetition and redundancy**

Essential genes and total gene number	705
Globin genes are organized in two clusters	709
Unequal crossing-over rearranges gene clusters	711
Gene clusters suffer continual reorganization	715
Sequence divergence distinguishes two types of sites in DNA	717
The evolutionary clock traces the development of globin genes	718
Pseudogenes are dead ends of evolution	721
Genes for rRNA comprise a repeated tandem unit	723
An evolutionary dilemma: how are multiple active copies maintained?	729

CHAPTER 25	733
Genomes sequestered in organelles	
Organelle genomes are circular DNA molecules that code for organelle protein	735
The chloroplast genome has similarities to both prokaryotic and eukaryotic DNA	739
The mitochondrial genome is large in yeast but small in mammals	741
Recombination and rearrangement of organelle DNA	745
CHAPTER 26	749
Organization of simple sequence DNA	
Highly repetitive DNA forms satellites	750
Satellite DNAs often lie in heterochromatin	751
Arthropod satellites have very short identical repeats	752
Mammalian satellites consist of hierarchical repeats	754
Evolution of hierarchical variations in the satellite	758
The consequences of unequal crossing-over	760
Crossover fixation could maintain identical repeats	762
Minisatellites are useful for genetic mapping	763
CHAPTER 27	767
The genome is packaged into chromosomes	
Condensing viral genomes into their coats	768
The bacterial genome is a nucleoid with many supercoiled loops	772
Loops, domains, and scaffolds in eukaryotic DNA	776
The contrast between interphase chromatin and mitotic chromosomes	779
The extended state of lampbrush chromosomes	782
Transcription disrupts the structure of polytene chromosomes	784
The eukaryotic chromosome as a segregation device	788
Chromosome ends are special	791
CHAPTER 28	797
Chromosomes consist of nucleosomes	
The nucleosome is the subunit of all chromatin	798
The core particle is highly conserved	802
DNA is coiled around the histone octamer	804
Supercoiling and the periodicity of DNA	809
The path of nucleosomes in the chromatin fiber	810
Organization of the histone octamer	813
Reproduction of chromatin requires assembly of nucleosomes	815
Are nucleosomes arranged in phase?	819
Are transcribed genes organized in nucleosomes?	822
DNAase hypersensitive sites change chromatin structure	827
Regulation of domains	831
Gene expression is associated with demethylation	835
Methylation is responsible for imprinting	839

PART 7

19 May

CHAPTER 29

Building the transcription complex: promoters, factors, and RNA polymerases

Eukaryotic transcription and RNA processing	845
Eukaryotic RNA polymerases consist of many subunits	849
Promoter elements are defined by deletions, point mutations, and footprinting	851
RNA polymerase I has a bipartite promoter	853
RNA polymerase III uses both downstream and upstream promoters	857
The basal transcription apparatus consists of RNA polymerase II and general factors	860
Promoters for RNA polymerase II promoters contain elements consisting of short sequences	864
Enhancers contain bidirectional elements that assist initiation	869
3' ends are generated by termination and by cleavage reactions	873

CHAPTER 30

Regulation of transcription: factors that activate the basal apparatus

CHAPTER 30	879
Response elements identify genes under common regulation	880
Transcription factors bind DNA and activate transcription through independent domains	882
There are many types of DNA-binding domains	887
A zinc finger motif may provide a DNA-binding domain	889
Steroid receptors have domains for DNA binding, hormone binding, and activating transcription	893
Homeo domains may bind related targets in DNA	897
Helix-loop-helix proteins interact by combinatorial association	899
Leucine zippers may be involved in dimer formation	902
Speculations about the nature of gene activation	904

CHAPTER 31

The apparatus for nuclear splicing

CHAPTER 31	911
Nuclear splicing junctions are interchangeable but are read in pairs	913
Nuclear splicing proceeds through a lariat	916
Small RNAs are required for splicing and form a spliceosome	919
Alternative splicing involves differential use of splicing junctions	929
<i>Cis</i> -splicing and <i>trans</i> -splicing reactions	932
Yeast tRNA splicing involves cutting and rejoining	935

CHAPTER 32	941
Changing the informational content of RNA	
Group I introns undertake self-splicing by transesterification	942
Group I introns form a characteristic secondary structure	947
Ribozymes have various catalytic activities	949
Introns may code for endonucleases that sponsor mobility	954
RNA can have ribonuclease activities	956
RNA editing utilizes information from several sources	958

PART 8	The dynamic genome: DNA in flux	965
CHAPTER 33	Recombination of DNA	967
	Breakage and reunion involves heteroduplex DNA	970
	Do double-strand breaks initiate recombination?	973
	Double strand breaks may initiate synapsis	976
	Bacterial recombination involves single strand assimilation	978
	Gene conversion accounts for interallelic recombination	984
	Topological manipulation of DNA	986
	Gyrase introduces negative supercoils in DNA	989
	Specialized recombination involves breakage and reunion at specific sites	991
CHAPTER 34	Transposons that mobilize via DNA	999
	Insertion sequences are simple transposition modules	1001
	Composite transposons have IS modules	1003
	Transposition occurs by both replicative and nonreplicative mechanisms	1006
	Conservative and replicative transposition may pass through common intermediates	1010
	TnA transposition requires transposase and resolvase	1013
	Transposition of Tn10 is subject to multiple controls	1015
	Controlling elements in maize are transposable	1018
	Autonomous elements give rise to nonautonomous elements	1020
	<i>Ds</i> may transpose or cause chromosome breakage	1022
	The suppressor-mutator family also consists of autonomous and nonautonomous elements	1024
	The role of transposable elements in hybrid dysgenesis	1026
CHAPTER 35	Retroviruses and retroposons	1033
	The retrovirus life cycle involves transposition-like events	1035
	Retroviruses may transduce cellular sequences	1043
	Yeast Ty elements resemble retroviruses	1045
	Many transposable elements reside in <i>D. melanogaster</i>	1049
	Retroposons fall into two classes	1059
CHAPTER 36	Rearrangement and amplification in the genome	1057
	The mating type pathway is triggered by signal transduction	1059
	Yeast can switch silent and active loci for mating type	1062

Silent cassettes at <i>HML</i> and <i>HMR</i> are repressed	1067
Unidirectional transposition is initiated by the recipient <i>MAT</i> locus	1069
Regulation of <i>HO</i> expression	1071
Trypanosomes rearrange DNA to express new surface antigens	1074
Interaction of Ti plasmid DNA with the plant genome	1079
Selection of amplified genomic sequences	1087
Exogenous sequences can be introduced into cells and animals by transfection	1091

PART 9**CHAPTER 37****Generation of immune
diversity by gene
reorganization**

Genes in development	1101
CHAPTER 37	1103
Clonal selection amplifies lymphocytes that respond to individual antigens	1106
Immunoglobulin genes are assembled from their parts in lymphocytes	1108
The diversity of germ-line information	1115
Recombination between V and C genes generates deletions and rearrangements	1118
Allelic exclusion is triggered by productive rearrangement	1122
DNA recombination causes class switching	1124
Early heavy-chain expression can be changed by RNA processing	1127
Somatic mutation generates additional diversity	1128
T-cell receptors are related to immunoglobulins	1131
The major histocompatibility locus codes for many genes of the immune system	1135

CHAPTER 38**Gene regulation in
development: gradients
and cascades**

CHAPTER 38	1141
A gradient must be converted into discrete compartments	1143
Maternal gene products establish gradients in early embryogenesis	1146
Cell fate is determined by compartments that form by the blastoderm stage	1157
Complex loci are extremely large and involved in regulation	1166
The homeobox is a common coding motif in homeotic genes	1173

CHAPTER 39**Oncogenes: gene
expression and cancer**

CHAPTER 39	1181
Transforming viruses may carry oncogenes	1185
Retroviral oncogenes have cellular counterparts	1190
Ras proto-oncogenes can be activated by mutation	1193
Insertion, translocation, or amplification may activate proto-oncogenes	1196
Loss of tumor suppressors causes tumor formation	1201
Immortalization and transformation	1205
Oncogenes code for components of signal transduction cascades	1208
Oncogenic variants of ras proteins are constitutively active	1213
Growth factor receptor kinases and cytoplasmic tyrosine kinases	1216
Oncoproteins may regulate gene expression	1223

Epilogue	Landmark shifts in perspectives	1231
Glossary		1235
Index		1259

Cells as macromolecular assemblies

INTRODUCTION

We shall assume the structure of the gene to be that of a huge molecule, capable only of discontinuous change, which consists in a rearrangement of the atoms and leads to an isomeric molecule. The rearrangement [mutation] may affect only a small region of the gene, and a vast number of different rearrangements may be possible.

Erwin Schrödinger, 1945