# MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

*Garland*
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

*Cell Biology Interactive*
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

**Bruce Alberts** received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.



**Library of Congress Cataloging-in-Publicaton Data**
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. ]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

**Front cover** Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

**Back cover** In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila, Arabidopsis,* Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

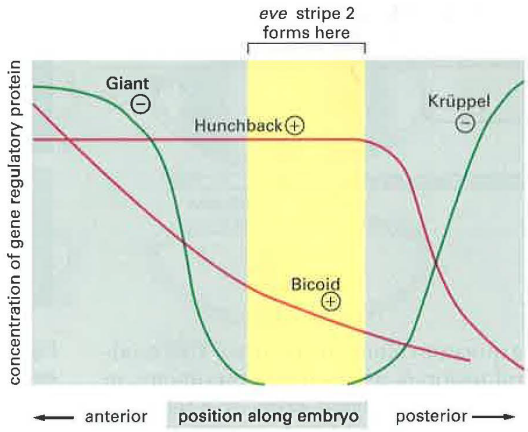



**Figure 7–56 Distribution of the gene regulatory proteins responsible for ensuring that *eve* is expressed in stripe 2.** The distributions of these proteins were visualized by staining a developing *Drosophila* embryo with antibodies directed against each of the four proteins (see Figures 7–52 and 7–53). The expression of *eve* in stripe 2 occurs only at the position where the two activators (Bicoid and Hunchback) are present and the two repressors (Giant and Krüppel) are absent. In fly embryos that lack Krüppel, for example, stripe 2 expands posteriorly. Likewise, stripe 2 expands posteriorly if the DNA-binding sites for Krüppel in the stripe 2 module (see Figure 7–55) are inactivated by mutation and this regulatory region is reintroduced into the genome.

The *eve* gene itself encodes a gene regulatory protein, which, after its pattern of expression is set up in seven stripes, regulates the expression of other *Drosophila* genes. As development proceeds, the embryo is thus subdivided into finer and finer regions that eventually give rise to the different body parts of the adult fly, as discussed in Chapter 21.

This example from *Drosophila* embryos is unusual in that the nuclei are exposed directly to positional cues in the form of concentrations of gene regulatory proteins. In embryos of most other organisms, individual nuclei are in separate cells, and extracellular positional information must either pass across the plasma membrane or, more usually, generate signals in the cytosol in order to influence the genome.

We have already discussed two mechanisms of combinatorial control of gene expression—heterodimerization of gene regulatory proteins in solution (see Figure 7–22) and the assembly of combinations of gene regulatory proteins into small complexes on DNA (see Figure 7–50). It is likely that both mechanisms participate in the complex regulation of *eve* expression. In addition, the regulation of stripe 2 just described illustrates a third type of combinatorial control. Because the individual regulatory sequences in the *eve* stripe 2 module are strung out along the DNA, many sets of gene regulatory proteins can be bound simultaneously and influence the promoter of a gene. The promoter integrates the transcriptional cues provided by all of the bound proteins (Figure 7–57).

The regulation of *eve* expression is an impressive example of combinatorial control. Seven combinations of gene regulatory proteins—one combination for each stripe—activate *eve* expression, while many other combinations (all those found in the interstripe regions of the embryo) keep the stripe elements silent. The other stripe regulatory modules are thought to be constructed along lines similar to those described for stripe 2, being designed to read positional information provided by other combinations of gene regulatory proteins. The entire gene control region, strung out over 20,000 nucleotide pairs of DNA, binds more than 20 different proteins. A large and complex control region is thereby built from a series of smaller modules, each of which consists of a unique arrangement of short DNA sequences recognized by specific gene regulatory proteins. Although the details are not yet understood, these gene regulatory proteins are thought to employ a number of the mechanisms previously described for activators and repressors. In this way, a single gene can respond to an enormous number of combinatorial inputs.

## Complex Mammalian Gene Control Regions Are Also Constructed from Simple Regulatory Modules

It has been estimated that 5–10% of the coding capacity of a mammalian genome is devoted to the synthesis of proteins that serve as regulators of gene

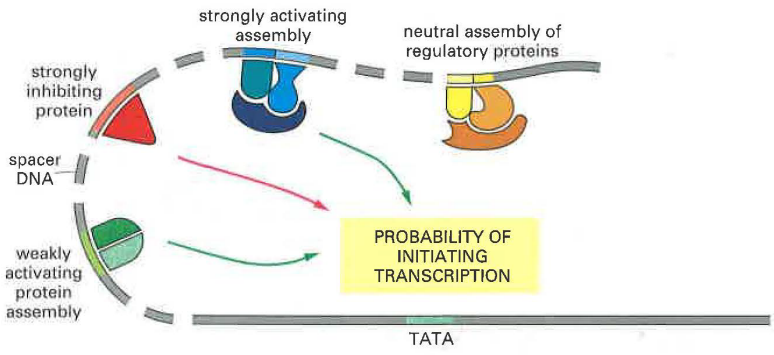



**Figure 7–57 Integration at a promoter.** Multiple sets of gene regulatory proteins can work together to influence transcription initiation at a promoter, as they do in the *eve* stripe 2 module illustrated previously in Figure 7–55. It is not yet understood in detail how the integration of multiple inputs is achieved, but it is likely that the final transcriptional activity of the gene results from a competition between activators and repressors that act by the mechanisms summarized in Figures 7–43, 7–44, 7–45, 7–46, and 7–49.

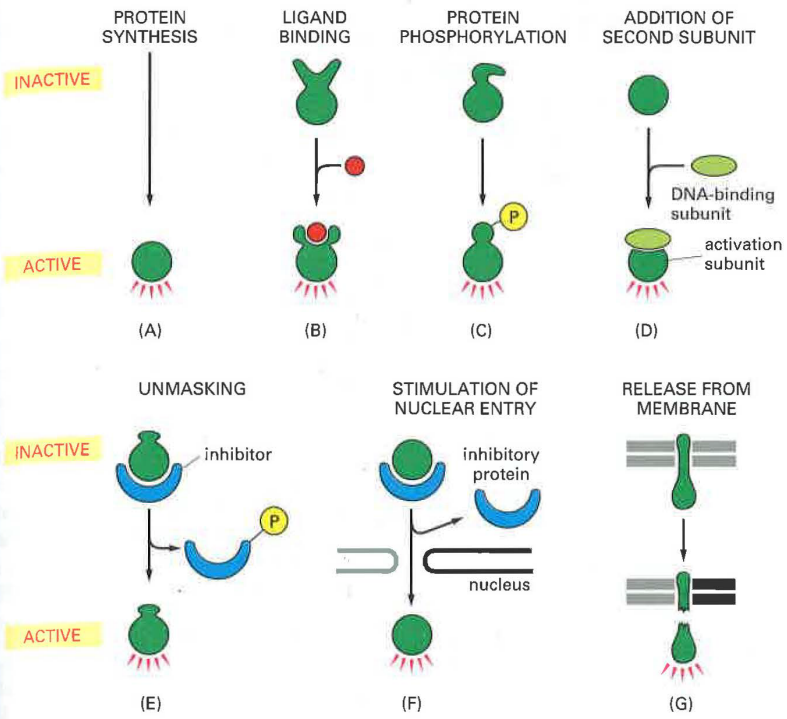
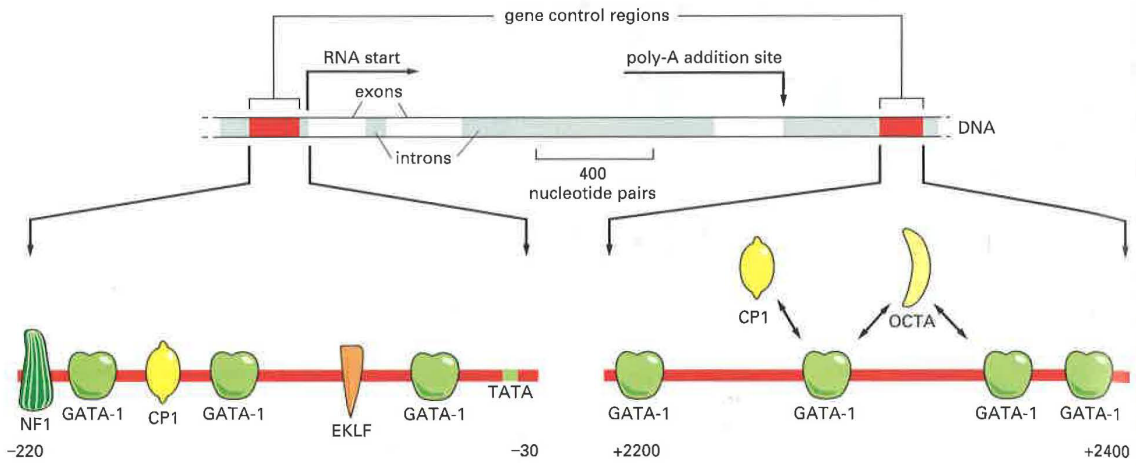



**Figure 7–58 Some ways in which the activity of gene regulatory proteins is regulated in eucaryotic cells.** (A) The protein is synthesized only when needed and is rapidly degraded by proteolysis so that it does not accumulate. (B) Activation by ligand binding. (C) Activation by phosphorylation. (D) Formation of a complex between a DNA-binding protein and a separate protein with a transcription-activating domain. (E) Unmasking of an activation domain by the phosphorylation of an inhibitor protein. (F) Stimulation of nuclear entry by removal of an inhibitory protein that otherwise keeps the regulatory protein from entering the nucleus. (G) Release of a gene regulatory protein from a membrane bilayer by regulated proteolysis.

Each of these mechanisms is typically controlled by extracellular signals which are communicated across the plasma membrane to the gene regulatory proteins in the cell. The ways in which this signaling occurs is discussed in Chapter 15. Mechanisms (A)–(F) are readily reversible and therefore also provide the means to selectively inactivate gene regulatory proteins.

transcription. This large number of genes reflects the exceedingly complex network of controls governing expression of mammalian genes. Each gene is regulated by a set of gene regulatory proteins; each of those proteins is the product of a gene that is in turn regulated by a whole set of other proteins, and so on. Moreover, the regulatory protein molecules are themselves influenced by signals from outside the cell, which can make them active or inactive in a whole variety of ways (Figure 7–58). Thus, pattern of gene expression in a cell can be viewed as the result of a complicated molecular computation that the intracellular gene control network performs in response to information from the cell's surroundings. We shall discuss this further in Chapter 21, dealing with multicellular development, but the complexity is remarkable even at the level of the individual genetic switch, regulating activity of a single gene. It is not unusual, for example, to find a mammalian gene with a control region that is 50,000 nucleotide pairs in length, in which many modules, each containing a number of regulatory sequences that bind gene regulatory proteins, are interspersed with long stretches of spacer DNA.

One of the best-understood examples of a complex mammalian regulatory region is found in the human β-globin gene, which is expressed exclusively in red blood cells and at a specific time in their development. A complex array of gene regulatory proteins controls the expression of the gene, some acting as activators and others as repressors (Figure 7–59). The concentrations (or activities) of many of these gene regulatory proteins are thought to change during development, and only a particular combination of all the proteins triggers transcription of the gene. The human β-globin gene is part of a cluster of globin genes (Figure 7–60A). The five genes of the cluster are transcribed exclusively in erythroid cells, that is, cells of the red blood cell lineage. Moreover, each gene is turned on at a different stage of development (see Figure 7–60B) and in different organs: the ε-globin gene is expressed in the embryonic yolk sac, γ in the yolk sac and the fetal liver, and δ and β primarily in the adult bone marrow. Each of the globin genes has its own set of regulatory proteins that are necessary to turn the gene on at the appropriate time and tissue. In addition to the individual regulation of each of the globin genes, the entire cluster appears to be subject to a shared control region called a *locus control region (LCR)*. The LCR lies far




upstream from the gene cluster (see Figure 7–60A), and we shall discuss its function next.

In cells where the globin genes are not expressed (such as brain or skin cells), the whole gene cluster appears tightly packaged into chromatin. In erythroid cells, by contrast, the entire gene cluster is still folded into nucleosomes, but the higher-order packing of the chromatin has become decondensed This change occurs even before the individual globin genes are transcribed, suggesting that there are two steps of regulation. In the first, the chromatin of the entire globin locus becomes decondensed, which is presumed to allow additional gene regulatory proteins access to the DNA. In the second step, the remaining gene regulatory proteins assemble on the DNA and direct the expression of individual genes.

The LCR appears to act by controlling chromatin condensation, and its importance can be seen in patients with a certain type of thalassemia, a severe inherited form of anemia. In these patients, the β-globin locus is found to have undergone deletions that remove all or part of the LCR, and although the β-globin gene and its nearby regulatory regions are intact, the gene remains transcriptionally silent even in erythroid cells. Moreover, the β-globin gene in the erythroid cells fails to undergo the normal chromatin decondensation step that occurs during erythroid cell development.

Many LCRs (that is, DNA regulatory sequences that regulate the accessibility and expression of distant genes or gene clusters) are present in the human genome, and they regulate a wide variety of cell-type specific genes. The way in which they function is not understood in detail, but several models have been proposed. The simplest is based on principles we have already discussed in this chapter: the gene regulatory proteins that bind to the LCR interact through DNA

**Figure 7–59 Model for the control of the human β-globin gene.** The diagram shows some of the gene regulatory proteins thought to control expression of the gene during red blood cell development (see Figure 7–60). Some of the gene regulatory proteins shown, such as CP1, are found in many types of cells, while others, such as GATA-1, are present in only a few types of cells—including red blood cells—and therefore are thought to contribute to the cell-type specificity of β-globin gene expression. As indicated by the *double-headed arrows*, several of the binding sites for GATA-1 overlap those of other gene regulatory proteins; it is thought that occupancy of these sites by GATA-1 excludes binding of other proteins. Once bound to DNA, the gene regulatory proteins recruit chromatin remodeling complexes, histone modifying enzymes, the general transcription factors and RNA polymerase to the promoter. (Adapted from B. Emerson, in Gene Expression: General and Cell-Type Specific [M. Karin, ed.], pp. 116–161. Boston: Birkhauser, 1993.)

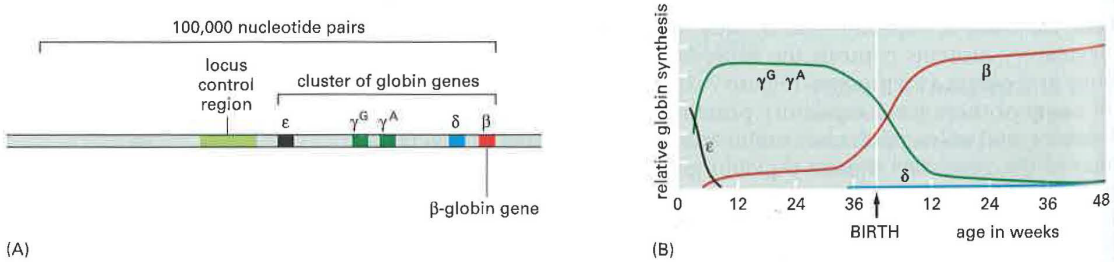



**Figure 7–60 The cluster of β-like globin genes in humans.** (A) The large chromosomal region shown spans 100,000 nucleotide pairs and contains the five globin genes and a locus control region (LCR). (B) Changes in the expression of the β-like globin genes at various stages of human development. Each of the globin chains encoded by these genes combines with an α-globin chain to form the hemoglobin in red blood cells (see Figure 7–115). (A, after F. Grosveld, G.B. van Assendelft, D.R. Greaves, and G. Kollias, *Cell* 51:975–985, 1987. © Elsevier.)