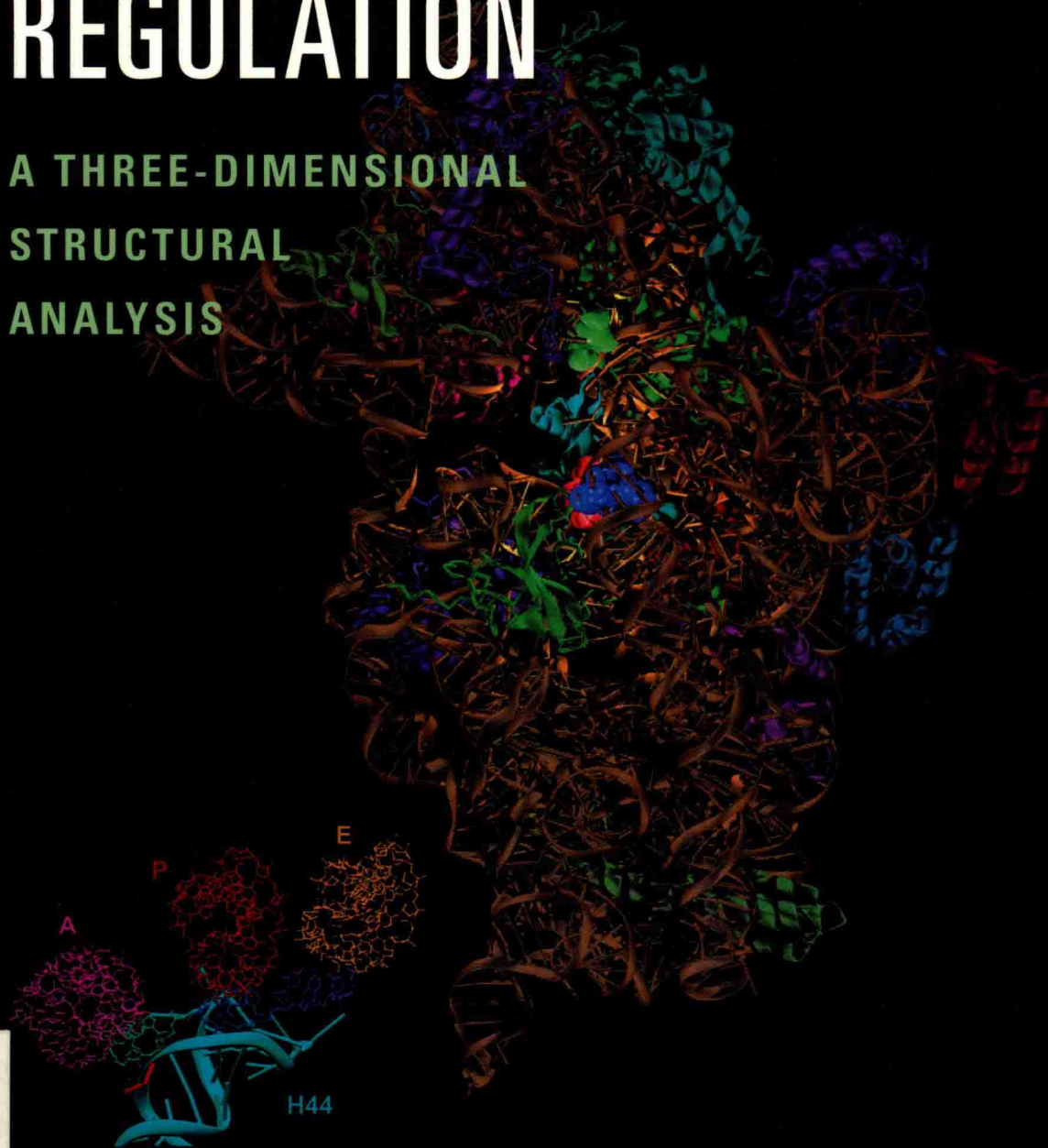


ANATOMY OF GENE REGULATION

A THREE-DIMENSIONAL
STRUCTURAL
ANALYSIS



PANAGIOTIS A. TSONIS

ANATOMY OF GENE REGULATION

A Three-Dimensional Structural Analysis

PANAGIOTIS A. TSONIS

University of Dayton



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Cambridge University Press 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typefaces Times Ten 10/13 pt. and Formata *System* L^AT_EX 2_ε [TB]

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

Library of Congress Cataloging in Publication Data

Tsonis, Panagiotis A.

Anatomy of gene regulation : a three-dimensional structural analysis / Panagiotis A. Tsonis.
p. cm.

Includes bibliographical references and index.

ISBN 0-521-80030-7 – ISBN 0-521-80474-4 (pb.)

1. Genetic regulation. 2. Nucleic acids – Structure. 3. Proteins – Structure.

4. Three-dimensional imaging in biology. I. Title.

QH450 .T78 2002

572.8'65 – dc21 2001037406

ISBN 0 521 80030 7 hardback

ISBN 0 521 80474 4 paperback

ANATOMY OF GENE REGULATION

A Three-Dimensional Structural Analysis

Determination of three-dimensional structures has revealed astonishing snapshots of molecules in action. No longer simple line drawings on a page, molecular structures can now be viewed in full-figured glory, often in color and even with interactive possibilities.

Anatomy of Gene Regulation is the first book to present the parts and processes of gene regulation at the three-dimensional level. Vivid structures of nucleic acids and their companion proteins are revealed in full-color, three-dimensional form. Beginning with a general introduction to three-dimensional structures, the book looks at the organization of the genome, the structure of DNA, DNA replication and transcription, splicing, protein synthesis, and ultimate protein death. The text discusses genetics and structural mechanics throughout. The concise and unique synthesis of information offers insight into gene regulation and into the development of methods to interfere with regulation at diseased states.

This textbook is appropriate for both undergraduate and graduate students in genetics, molecular biology, structural biology, and biochemistry courses.

Panagiotis A. Tsonis is Professor in the Department of Biology at the University of Dayton.

To my daughters
Isidora and Sol

Preface

What is in the structure? It is, of course, quality!

The determination of the three-dimensional (3-D) structure of DNA in 1953 heralded the beginning of molecular biology. At the same time, we saw one of the first examples of how the 3-D structure of a biomolecule reveals its function. The 3-D structure of DNA immediately suggested how the genetic information is passed to the progeny. Eventually, the discovery of DNA structure led to the understanding of how genetic information accounts for the final product, which is protein synthesis. For the past 40 years, research in molecular biology has led to the identification of a cascade in gene regulation from its packing into chromosomes to transcription, splicing, modifications, protein synthesis, and, finally, the death of proteins. Eventually, knowledge of the mechanisms involved in these events led to manipulation of genes, recombinant DNA, and cloning technology, all of which helped us grasp the function of genes and their role in the study of differentiation, development, and diseases.

As the major players at all the different levels of gene regulation were discovered, it became apparent that the final mechanisms will be best revealed when we can observe the action of enzymes and genes at the 3-D level. Information gathered by biochemical and molecular experiments could identify the function of an enzyme, say the role of DNA polymerase in replication or the role of helicase in unwinding DNA. However, only by observing the process at the 3-D level can we visualize and fathom the mechanism by which the enzyme possesses such activity. In fact, we can visualize enzyme action not only at the molecular level but also at the atomic level.

All this was possible because of the development of technology to determine the 3-D structure of biological molecules. In this sense, molecular biology has given rise to atomic biology where interactions between

biomolecules and mechanisms can be resolved at the atomic level. These techniques have become better over the years and undoubtedly will become even better. So far, the determination of 3-D structures has revealed astonishing snapshots of molecules in action and has made secrets known. We are able now to virtually study the anatomy of molecular events.

In this book, I present the story of gene regulation with emphasis at the 3-D level, thus the title “Anatomy of Gene Regulation.” Like surgeons (molecular surgeons), we can now open the nucleus and other organelles and, with extraordinary glasses, see the mechanisms unfolding at the atomic level. I have elected to begin our discussion with the packing (organization) of DNA in the nucleus and then to explore DNA replication, transcription, and splicing, followed by RNA modification. Next, we will follow the path of mRNA to the cytoplasm and its decoding during protein synthesis. The final chapter will consider the birth and death of proteins, which is the end of the regulation process.

Why did I undertake such a task? Even though I am not a structural biologist, I have been teaching molecular biology for nearly 15 years. Throughout my teaching career, I realized that most of the molecular biology textbooks were incredibly massive, full of more information than can be covered in one semester. Furthermore, textbooks in different disciplines overlap considerably. Most textbooks on microbiology, genetics, molecular or cellular biology, and even developmental biology contain chapters on the basic aspects of gene regulation. Students, therefore, are exposed to the same contents throughout their education. I believe that, at least in the study of molecular biology, some changes are warranted. Consequently, I slowly started incorporating the 3-D aspects. I soon realized that my course was becoming more unique and exciting to the students. Finally, I decided that a molecular biology course that emphasized 3-D aspects must be beneficial to both undergraduate and graduate students. Determining the 3-D structure of a mechanism is the ultimate level that we should go to if we are to understand the mechanism. Therefore, all disciplines dealing with cellular and molecular events will eventually need to study them at the 3-D level.

Producing this text was very challenging. For every structure presented in the book, I read the corresponding paper(s) very carefully and extracted only the information that would produce a sequence that flows well and is accurate. Therefore, my book is by design short and concise. It focuses on the 3-D aspects of gene regulation only. However, it does not leave out information pertinent to molecular biology when structures are not involved. I intend to focus on the 3-D aspects of gene regulation and not to overwhelm the reader with details that can be found in many other general textbooks and courses. Each chapter is preceded by a section, which I call primer. This section outlines the general plan of the sequence that I follow

in the chapter. It should help keep the reader focused on the main aim of the text.

Even a book on 3-D structures can be overwhelming. For example, the structure of many different DNA polymerases must be solved. Likewise, we must know the 3-D structure of numerous transcriptional factors. Obviously, reporting all of them in a book would be tedious for the reader. Therefore, I have filtered the information and concentrated on a series of structures that best represent the desired style of the book and its aim, which is to acquaint the readers with the 3-D aspects of gene regulation. At the same time, I have tried not to compromise quality and accuracy. The structures are presented in different ways using different models. This technique is deliberate because some aspects of a 3-D structure are better depicted by one model than by others.

Such a synthesis has not been attempted before, so I am sure that there will be supporters and critics. I do count on both to provide me with valuable comments on this project so that can improve it in the future.

Finally, I am grateful to numerous colleagues who have furnished me with the figures that are included in the book. Without their help and co-operation, my book would have been impossible to produce. The reference list is not extensive and is concentrated mainly on the papers that deal with the 3-D structures. General references on standard molecular biology can be found in several excellent textbooks, which are cited in the reference section.

CREDITS

Several databases, which have cataloged available three-dimensional structures of proteins, were a big help. Throughout the book numerous images have been used from these databases. These databases follow:

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demesy, T., Hsieh, S.-H., Srinivasan, A. R., and Schneider, B. (1992). The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids (NDB). *Biophys. J.* 63: 751–9. <http://ndbserver.rutgers.edu/NDB/NDBATLAS/>

Berman, H. M., Westbrook, J., Zeng, Z., Gilliland, G., Bhat, T. N., Weissing, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank (PDB). *Nucleic Acids Res.* 28: 235–42. www.rcsb.org/pdb

Reichert, J., Jabs, P., Slickers, J., and Suhnel, J. (2000). The IMB Jena Image Library of Biological Macromolecules. *Nucleic Acids Res.* 28: 246–9. www.imb-jena.de

Contents

	Preface	<i>page xi</i>
1	A General Introduction to 3-D Structures	1
2	The Higher Organization of the Genome	6
	Packing DNA in the Chromosomes	6
	The Three-Dimensional Structure of the Nucleosome	8
	Other Proteins	10
3	Structure of DNA and Telomeres	14
	The Three-Dimensional Structure of DNA	14
	Telomeres	17
	Structure of OnTEBP and Interaction with Telomeric Sequences	19
	Structure of Yeast RAP1	20
4	DNA Replication	26
	The Prereplicative Complex	28
	Replication Origins in Prokaryotes	28
	Replication Origins in Eukaryotes	28
	Elongation and the Synthesis of New DNA Strands	30
	DNA Polymerases	30
	The Three-Dimensional Structure of DNA pol I	31
	The Incorporation of the Incoming Nucleotide	33
	The Three-Dimensional Structure of the DNA Polymerase from Phage RB69	36
	The Three-Dimensional Structure of Human DNA Polymerase Beta	36
	Editing	38

Processivity of DNA Polymerases and the Structure of the Clamp	39
The Loading of the Clamp	41
Processivity Factors of Eukaryotic DNA Polymerases	42
Synthesis of the Primers in the Lagging Strand	42
Processing of the Okazaki Fragments into a Continuous Strand	45
The Single-Stranded DNA-Binding Proteins	48
Viewing the Replication Complex	48
Helicases	50
The <i>E. coli</i> Rep Helicase	51
The <i>Bacillus stearothermophilus</i> PcrA Helicase	53
The Bacteriophage T7 Helicase Domain	53
Topoisomerases	58
The 3-D Structure of Human Topo I	61
Mechanism of Cleavage and Relaxation of DNA	61
The Structure of Yeast Topoisomerase II	65
Termination of Replication	67
Box: How Drugs Induce Mutations	71
5 Transcription in Prokaryotes	73
Structure and Function of Prokaryotic RNA Polymerase	74
The 3-D Structure of the <i>Thermus Aquaticus</i> RNA Pol	77
Regulation in Prokaryotes	86
The Helix-Turn-Helix Motif in Prokaryotic Gene Regulation	87
The Lac Operon	87
The Pur Repressor	89
The Trp Repressor	89
Catabolite Activator Protein	90
Regulation in Phages	90
The 434 Phage	92
The Lambda Phage	95
Termination of Prokaryotic Transcription	98
The Attenuator	101
6 Transcription in Eukaryotes	104
Structure and Function of the Yeast RNA Pol II	105
Comparison of Prokaryotic and Eukaryotic RNA Polymerases with the T7 RNA Polymerase	108
Regulation in Eukaryotes	111
Regulation in Class I Genes	111
Regulation in Class III Genes	111
Structure of TFIIB	113
Regulation in Class II Genes	115

The Basal Transcriptional Apparatus in Class II Genes	115
Structure and Function of TFIID.	116
Structure of TFIIA	118
Structure of TFIIB	119
Structure of TFIIF	120
Transcriptional Activation in Eukaryotes	120
Liver-Specific Transcription	121
The GAL Regulatory System	122
The Steroid Hormone Receptors	123
The Homeodomain-Containing Genes	124
The Mediator	127
Chromatin Structure and Gene Regulation	128
Regulation of Cell Fate by Homeodomains: The Yeast Mat Genes	130
Structural Features of Class II DNA-Binding Domains	132
The Helix-Turn-Helix Motif	133
The Zinc-Binding Domains	135
The Class 1 Zinc-Binding Domains	136
The Class 2 Zinc-Binding Domains	140
The Class 3 Zinc-Binding Domains	142
Other DNA-Binding Domains	143
The Leucine Zipper	144
The Immunoglobulin Fold in NFAT	145
The T Domain	148
Signal Transducers and Activators of Transcription Proteins	148
Box: Most Frequent Base-Amino Acid Interactions	152
7 Splicing	153
Splicing of the Nuclear Pre-mRNA Introns	155
The First Transesterification	155
The Second Transesterification	155
The Spliceosome	156
Structure and Assembly of snRNPs	156
Other Splicing Factors	157
The Events of Splicing	159
Splicing of Group II Introns	160
Splicing of Group I Introns	163
Other Ribozymes	167
Structure of the Hammerhead Ribozyme	167
The Interesting Case of Group I Splicing by a Protein, Tyrosyl-tRNA Synthetase	170
8 Modifications of mRNA	172
The Capping of the 5' Ends	172
Structure of the Yeast RNA 5'-Triphosphatase	173
Structure of the Guanylyltransferase	173

Maturation at The 3' End: Addition of polyA	178
3-D Structure of the polyA-Binding Protein	179
Maturation of mRNA Lacking polyA Tails	180
RNA Editing	181
RNA Interference	182
9 Compartmentalization of Transcription	184
Splicing Factors	185
Visualization of Single Transcripts	185
The Moving of Transcripts Out of the Nucleus	187
Shuttling mRNA-Binding Proteins	190
Segregation of Replication and Transcription	190
10 Protein Synthesis	191
The Major Players in Protein Synthesis	193
The 3-D Structure of tRNA	193
The Charging of tRNA with Amino Acid	194
The 3-D Structure of <i>E. coli</i> GlutaminyI-tRNA	
Synthetase and Its Interactions with tRNA ^{Gln}	195
The Ribosome	200
Initiation	202
Initiation in Prokaryotes	203
Initiation in Eukaryotes	203
The Structure of IF-1	205
The Structure of IF-3	205
The Structure of IF-2/eIF-5B	207
The Structure of eIF-4E and eIF-4G	207
The Structure of eIF-1 and eIF-1A	209
Elongation	210
The Structure of Elongation Factors EF-Tu and EF-G	212
Movement of tRNAs During Translocation	214
Movement of EF-G and Conformational Changes of the Ribosome During Translocation	217
Interactions Between the 30S and 50S Subunits	219
Interactions Between tRNA and rRNAs	221
Interactions Between tRNA and 16S rRNA	222
Interactions Between tRNA and 23S rRNA	223
Interactions Between tRNA and 5S rRNA	230
The Decoding Center of the Ribosome	232
The Peptide Bond Center	236
Termination of Protein Synthesis	237
The Three-Dimensional Structure of Human eRF1	237
Mechanism of Stop Codon Recognition	238
The Ribosome Recycling Factor	239
Box: How Do Antibiotics Work? A Lesson from the 3-D Structure of the Ribosome	239

11	The Birth and Death of Proteins	241
	The Signal Peptide and Its Recognition	242
	Structure and Function of the Signal Recognition Particle	243
	The Barrel of Birth ...	247
	The 3-D Structure of DnaK and Hsp70	249
	The 3-D Structure of DnaJ	251
	The 3-D Structure of GrpE	251
	The 3-D Structure of Prefoldin	252
	The 3-D Structure of GroEL and GroEs	252
	... And The Barrel of Death	255
	Further Reading and References	259
	Index	271

A General Introduction to 3-D Structures

PRIMER The three-dimensional structure of nucleic acids and proteins as it pertains to the mechanisms involved in gene regulation is the major focus of this book. Therefore, the reader will encounter many 3-D structures. The first chapter of the book presents the very basic ideas behind the three-dimensional aspects of biomolecules. The first part deals with the techniques used to determine 3-D structures. The presentation is virtually for the layperson. Then the basic structural elements found in proteins are examined. Having done this, we examine a particular 3-D structure (that includes both DNA and protein) presented with different modeling. This exercise will help you to become familiar with the different ways that scientists present their 3-D structures. We use different models because one aspect of structure and function can be better represented with one model, whereas another aspect is more suited to a different model.

This book deals with the three-dimensional aspects of gene regulation. The reader will encounter numerous three-dimensional structures, but this should not scare anybody away. Unfamiliar readers might think that interpreting these structures is difficult, but this is not true. All we need is a basic introduction into the three-dimensional aspects of proteins and nucleic acids and the way that it can be represented. The basic 3-D structure of a protein can be reduced to two elements: the alpha helix and the beta strand (and loops that connect them). The complicated 3-D structure of a protein is a combination of several of these elements. Also, depending on the presentation, the alpha helix or the beta strand might be shown with different styles. To get started, let us review the main elements of the 3-D structures, the different representation

styles, and the basic methods used in the determination of the three-dimensional structures.

Two methods are generally used to determine the 3-D structure of a biomolecule (nucleic acid or protein). One method is Nuclear Magnetic Resonance (NMR) spectroscopy, and the other is X-ray diffraction (or X-ray crystallography). NMR uses properties of the atomic nuclei to determine how closely they are positioned. The so-called nuclear Overhauser effect (NOE) is a nuclear relaxation effect. This intensity is a measure of the distance between two nuclei that are close together. The two nuclei might be far apart in the primary sequence, but they could be close in 3-D because the protein is folded. Gathering data from all atoms enables the researcher to create a 3-D model of the molecule under investigation. NOE(s) are detected by NOE spectroscopy (NOESY) NMR experiments. The intensity of NOESY determines the actual distance between two nuclei. A strong intensity indicates that the two nuclei are 3 Å apart, a medium intensity measures less than 4 Å, and weak intensity is less than 6 Å. Because structure determination by NMR is in solution, ends or loops of proteins, which are flexible, are sometimes not solved well. For this, more than 20 calculated structures should be received and superimposed. At this point, we should be able to see the regions that are not defined well. Finally, based on all calculated structures, an average structural model can be produced. These superimposed structures appear throughout the text. One limitation of this method is that it can resolve structures of small proteins (about 30 kDa). However, a few proteins of about 50 kDa have also been solved, and future developments might push these limits. Also, because NMR determines structures in solution, the protein should be stable in solution.

The other method, using X-ray diffraction, can be applied to large molecules or even complexes of them. When using X-ray diffraction, the protein must be crystallized. The crystal is then exposed to X-rays, and a picture is received on a film where the diffracted light from the crystal produces patterns, depending on the 3-D structure of the protein. For example, the celebrated 3-D structure of DNA, which is a periodic pattern, produced spots on the film, which were symmetrically arranged. This symmetry led Watson and Crick to deduce that the DNA must have two periodicities, one from base to base and the other every helical turn (nearly every 10 bases; see Chapter 2). Obviously, most complicated 3-D structures, such as the ones found in proteins, would produce a more elaborate pattern on the film, but algorithms and techniques have been developed to put these patterns into a 3-D structure. X-ray diffraction would provide very clear 3-D solutions and does not have the limitations with the flexible regions as in NMR. The only limitation is that not all proteins can be crystallized efficiently. When the same structure has been solved with both NMR and X-ray diffraction, the results usually match very well, indicating that both methods are quite reliable.

Let us now familiarize ourselves with the basic structures in a protein. As noted earlier, the primary amino acid sequence can assume either a helical or a

beta strand conformation. Some amino acids are more likely than others to be in an alpha helix, and the same is true for amino acids found in beta strands. First, we will examine the basic structure of an amino acid and the peptide bond. All amino acids have a central carbon, C_α , to which a hydrogen atom, NH_2 (amino group), and $COOH$ (carboxyl group) are attached. What discriminates the 20 different amino acids is the side chain, R , which is attached to the central carbon atom (Figure 1.1A). Amino acids are joined via the peptide bond to create polypeptides (Figure 1.1B). When amino acids are arranged in an alpha helix, there is hydrogen bonding between the $C=O$ of a residue and

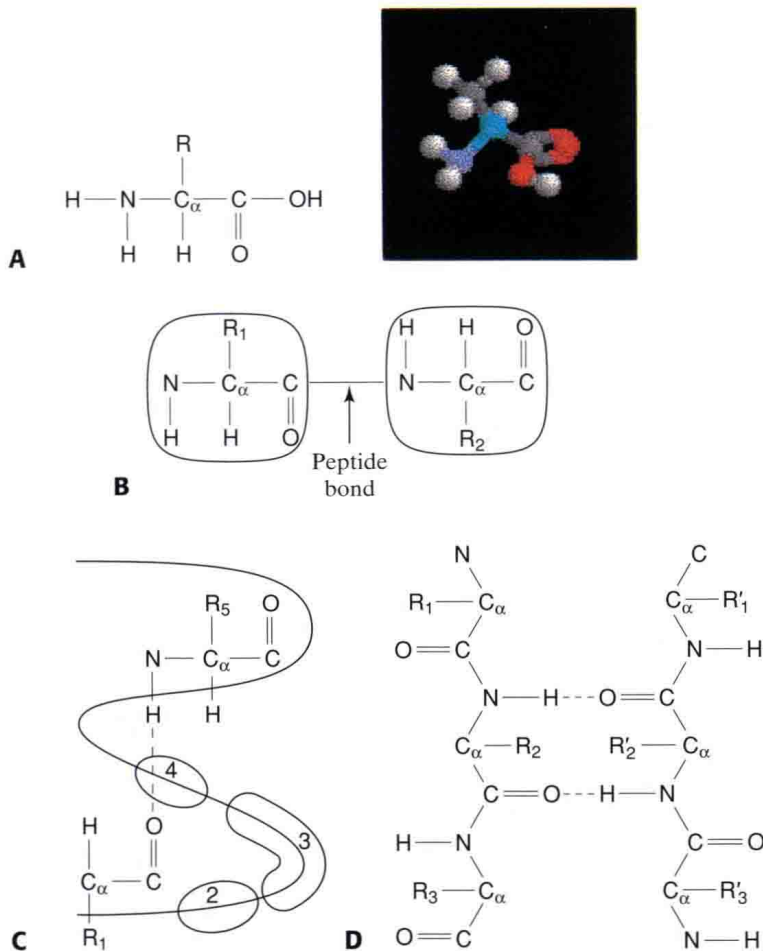


Figure 1.1. **A:** The basic chemical structure of an amino acid, indicating the standard H , NH_2 , and $COOH$ groups. R is the side chain that can vary in different amino acids. In the ball-and-stick representation, R is a CH_3 group and the amino acid is alanine. C is gray with the C_α cyan, N is blue, oxygen is red, and H is white. **B:** A dipeptide showing the creation of the peptide bond. **C:** An illustration of alpha helix. Note that residue 1 and residue 5 interact via hydrogen bonding (dashed line) using their $C=O$ and NH groups, respectively. **D:** Two antiparallel beta strands creating a beta sheet via hydrogen bonding (dashed line) using an NH group from one strand and a $C=O$ group from another. From F. R. Gorga, Protein Data Bank (PDB), Nucleic Acids Res. 28: 235–42.

the NH of another residue four positions away. In other words, there would be hydrogen bonds between residue 1 and 5, 2 and 6, and so on (Figure 1.1C). The alpha helix has 3.6 residues per turn, but variations exist with hydrogen bonds to residue $n + 5$ (pi helix) or $n + 3$ (3_{10} helix). Most of these helices are found at the end of alpha helices. The alpha helices are usually depicted as ribbons or cylinders in the 3-D structure of a protein.

The beta strand, and the resulting beta sheets from their interaction, is the second major element found in proteins. A beta strand contains 5 to 10 amino acids, which are in almost fully extended conformation. Interactions with adjacent beta strands can form a beta sheet. These interactions involve hydrogen bonding between the C=O of one strand and the NH group of another (Figure 1.1D). From such a configuration, we can see that the beta strands are pleated with C_{α} atoms successively above or below the plane of a sheet. The side chains follow this pattern as well. A beta sheet is called parallel when the strands run in the same direction or antiparallel when they do not. The example in Figure 1.1D is an antiparallel beta sheet. The beta strands are usually represented as arrows in the 3-D structure of a protein with the arrowhead pointing to the direction ($N \rightarrow C$).

This book contains numerous structures that are represented as different models. This approach is deliberate because some models can show a particular feature much better than others. To illustrate, Figure 1.2 presents a particular structure using four different models. The structure shows the interaction between the paired domain of the activator pax-6 with DNA. It is a good example

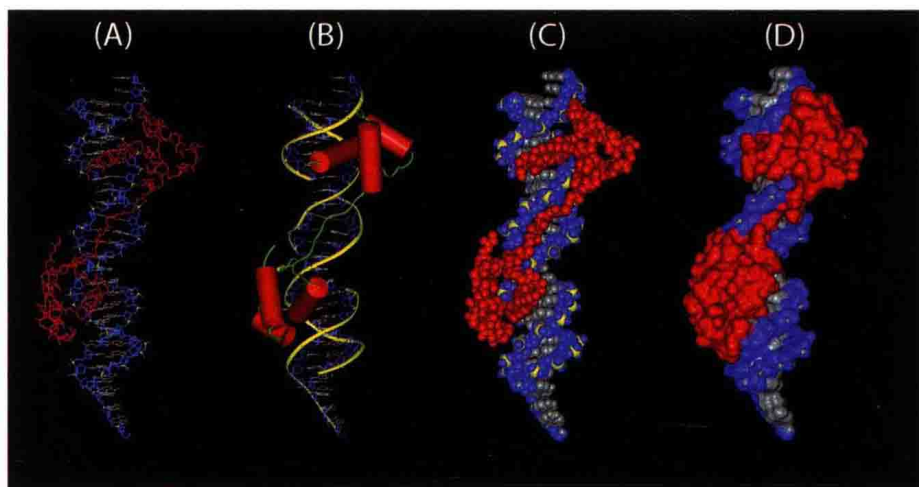


Figure 1.2. Different models of pax-6 bound to DNA. **A:** Ball-and-stick diagram. Phosphates are yellow, the sugar moiety is blue, nucleotide bases are gray, and pax-6 is red. **B:** Same as in A, but the helices of pax-6 are shown as red cylinders, and the connecting parts, as green strings. The DNA phosphates are yellow and have been traced to highlight the DNA. **C:** CPK (space-filling) model with same colors as in A. **D:** Surface representation with the same colors as in A. Images generated by E. Fuentes; Xu et al., *Genes Develop.* 13: 1263–75 (1999).