CURRENT HEMATOLOGY AND ONCOLOGY

VIRGIL F. FAIRBANKS

VOLUME 4

CURRENT

HEMATOLOGY AND ONCOLOGY

VOLUME 4

Edited by

Virgil F. Fairbanks, M.D.
Professor of Medicine and Laboratory Medicine
Mayo Medical School
Consultant, Mayo Clinic
Rochester, Minnesota

YEAR BOOK MEDICAL PUBLISHERS, INC.
Chicago • London

International Standard Serial Number: 0739-4810
International Standard Book Number: 0-8151-3208-5

Sponsoring Editor: Elizabeth Sugg
Manager, Copyediting Services: Frances M. Perveiler
Production Project Manager: Etta Worthington
Proofroom Supervisor: Shirley E. Taylor

CURRENT HEMATOLOGY AND ONCOLOGY VOLUME 4

Contributors

Andrew Arnold, M.D.
Metabolic Branch
National Cancer Institute
National Institutes of Health
Bethesda, Maryland

Ajay Bakhshi, M.D.
Metabolic Branch
National Cancer Institute
National Institutes of Health
Bethesda, Maryland

Edward J. Benz, Jr., M.D.
Associate Professor of Medicine and Human Genetics
Yale University School of Medicine
New Haven, Connecticut

Joseph H. Butterfield, M.D.
Consultant, Division of Allergic Diseases and Internal Medicine
Mayo Clinic and Mayo Foundation
Assistant Professor of Medicine
Mayo Medical School
Rochester, Minnesota

Oswaldo Castro, M.D.
Deputy Director
Center for Sickle Cell Disease
Associate Professor of Medicine and Pediatrics
Howard University College of Medicine
Washington, D.C.

Herbert A. Cooper, M.D.
Professor, Departments of Pediatrics and Pathology
University of North Carolina School of Medicine
Chapel Hill, North Carolina

Gerald J. Gleich, M.D.
Chairman, Department of Immunology
Mayo Clinic and Mayo Foundation
Professor of Immunology and Medicine
Mayo Medical School
Rochester, Minnesota

Alois Gratwohl, M.D.
Division of Hematology
Department of Medicine
Kantonsspital, Basel, Switzerland

John A. Koepke, M.D.
Medical Director
Clinical Hematology Laboratories
Duke University Medical Center
Durham, North Carolina

Stanley J. Korsmeyer, M.D.
Metabolic Branch
National Cancer Institute
National Institutes of Health
Bethesda, Maryland

Harry J. Long III, M.D.
Consultant
Department of Medical Oncology
Mayo Clinic and Mayo Foundation
Assistant Professor
Mayo Medical School
Rochester, Minnesota

H. Bryan Neel III, M.D., Ph.D.
Chairman
Department of Otolaryngology
Mayo Clinic and Mayo Foundation
Professor of Otolaryngology and
Associate Professor of Microbiology
Mayo Medical School
Rochester, Minnesota

Arthur W. Nienhuis, M.D.
Chief, Clinical Hematology Branch
National Heart, Lung, and Blood Institute
National Institutes of Health
Bethesda, Maryland

Catherine Nissen, M.D.
Department of Research
Kantonsspital, Basel, Switzerland

Gary R. Pearson, Ph.D.
Professor and Head
Department of Microbiology
Georgetown University Medical Center
Washington, D.C.

Peter T. Rowley, M.D.
Professor of Medicine, Pediatrics, Genetics, and Microbiology
University of Rochester School of Medicine
Rochester, New York

Katherine Siminovitch, M.D.
Metabolic Branch
National Cancer Institute
National Institutes of Health
Bethesda, Maryland

Bruno Speck, M.D.
Division of Hematology
Department of Medicine
Kantonsspital, Basel, Switzerland

Thomas Waldmann, M.D.
Metabolic Branch
National Cancer Institute
National Institutes of Health
Bethesda, Maryland

Louis H. Weiland, M.D.
Consultant, Department of Surgical Pathology
Mayo Clinic and Mayo Foundation
Professor of Pathology
Mayo Medical School
Rochester, Minnesota

R. K. Winkelmann, M.D., Ph.D.
Professor of Dermatology and Anatomy
Mayo Medical School
Consultant, Department of Dermatology
Mayo Clinic and Mayo Foundation
Rochester, Minnesota

William P. Winter, Ph.D.
Senior Biochemist
Center for Sickle Cell Disease
Associate Professor of Medicine
Howard University College of Medicine
Associate Professor of Genetics and Human Genetics
Howard University Graduate School of the Arts and Sciences
Washington, D.C.

Preface

Readers of the earlier volumes of *Current Hematology and Oncology* have responded favorably to the approach we have taken of intermingling reviews of broad areas of hematology with more narrowly focused chapters on special topics, particularly on topics that reflect major advances in our field. We have continued this approach in Volume 4 by including chapters on molecular genetics, the eosinophil, the structure of the immunoglobulin genes, Epstein-Barr virus-related disorders, and on immune therapy of aplastic anemia. Volume 4 also continues to reflect progress in hematology and medical oncology, with chapters on nasopharyngeal cancer and on germ cell tumors. Thus, the contents of Volume 4 are weighted nearly even for hematology and for medical oncology.

For this volume we have also taken cognizance of the frequency of cutaneous expressions of hematologic and neoplastic disorders by including a chapter by a distinguished dermatologist who has had a long interest in the interplay of these disorders.

In view of the extensive coverage of leukemias and lymphomas provided in the previous volumes of this series, we have elected to defer a review of these topics to Volume 5.

A new publisher brings not only new enthusiasm and a new and brighter format, but a new schedule as well. Beginning with Volume 5, now in preparation, we anticipate that the appearance of *Current Hematology and Oncology* will become an annual event.

I wish to thank all the contributors for the excellent quality of their chapters and for the generous expenditure of their time and energy. I also wish to thank my family for their patience, my secretary, Ms. Sara Brackett, for her careful work, and Year Book Medical Publishers for the excellent preparation of Volume 4.

VIRGIL F. FAIRBANKS

Contents

CHAPTER 1

Molecular Genetics for the Hematologist*

Peter T. Rowley, M.D.

Professor of Medicine, Pediatrics, Genetics, and Microbiology, University of Rochester School of Medicine, Rochester, New York

Edward J. Benz, Jr., M.D.

Associate Professor of Medicine and Human Genetics, Yale University School of Medicine, New Haven, Connecticut

Arthur W. Nienhuis, M.D.

Chief, Clinical Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland

"Molecular genetics" is the study of the chemical basis for the inheritance and regulated expression of genetic information. This discipline combines the techniques of the geneticist, the biochemist, the cell biologist, and the microbiologist in an effort to define the structure and the functional behavior of the macromolecules that control gene expression: DNA, RNA, and protein.

We first review the basic structural features of DNA, RNA, and proteins, and

*This chapter is based on an Educational Session entitled "ABCs of Molecular Genetics," presented at annual meetings of the American Society of Hematology. This work was supported by ACS grant CH-250, N.I.H. grant HL29231, and N.Y. State Contracts C-00308 and 20097-02 (P.T.R.); also, grants HL24385 and AM28076 from the NIH, and a grant from the Cooley's Anemia Foundation (E.J.B.).

0739-4810/86/0004-0001-0038-$04.00

then consider the basic experimental method (molecular hybridization) for the analysis of specific genes and their RNA products.

THE CHEMICAL BASIS OF INHERITANCE

Virtually all species except a few RNA viruses utilize DNA for the storage, expression, and transmission to later generations of genetic information. In other words, genes consist of DNA. DNA is a linear (unbranched) polymer consisting of a sugar-phosphate backbone and four types of nitrogenous (i.e., purine or pyrimidine) bases that protrude from the backbone of the polymer (Fig 1).

The sugar portion of DNA is deoxyribose, and the four bases are adenine, guanine, cytosine, and thymidine. The fundamental unit of the DNA polymer is the nucleotide, which consists of one molecule of deoxyribose bound to a phosphate group at its 5′-carbon position, and one of the nitrogenous bases bound at the 1′-carbon. Adjacent nucleotides are linked together by phosphodiester bonds between the 5′ carbon of the deoxyribose moiety of one nucleotide and the 3′ carbon of the next. Thus, a DNA polymer linked in this fashion has polarity, i.e., the two ends are recognizably different. At the 5′-end of DNA the terminal nucleotide has a phosphate monoester at the 5′ carbon of deoxyribose; at the 3′ end of DNA, there is a phosphate monoester at the 3′ carbon. All other phosphates in the molecule are bound as diesters to both the 5′ and 3′ carbons of deoxyribose. This is important, because it allows the cell (and the molecular biologist!) to identify the direction of a particular DNA strand.

The length of a DNA molecule may be described in terms of the number of

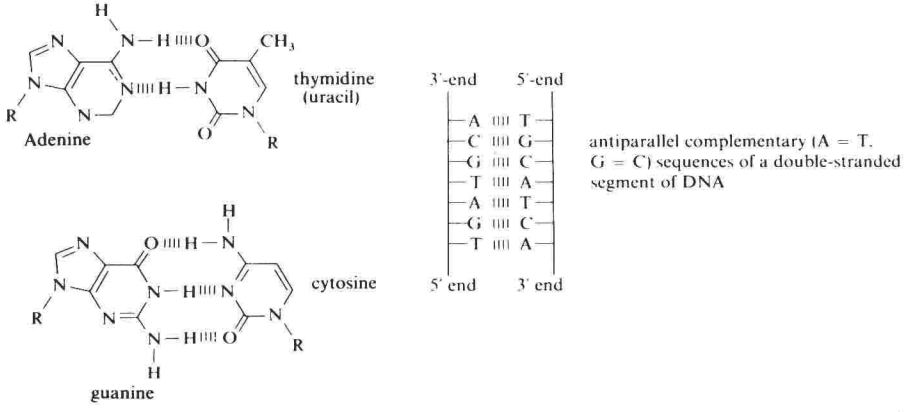

Fig 1.—Structural components of DNA. Shown are the structures of the four bases and the rules of hydrogen bonded base pairing (A-T and G-C only). A short double-stranded DNA molecule model illustrates the antiparallel complementarity by base pairing.

nucleotides contained in a linear chain. Since the genetic code is determined by the sequence of nitrogenous bases, the length of the DNA chain may also be described in terms of its number of bases. For example, a single-stranded DNA molecule containing 120 nucleotides is said to be 120 bases long. DNA is most frequently a double-stranded molecule. For this reason, chain lengths of double-stranded DNA are expressed as number of base pairs (bp). Since many DNA molecules are thousands of bases or base pairs long, a shorthand designation, "kilobases" or "kilobase paris," is often used, abbreviated "kb" or "kbp." Thus, a single-stranded DNA molecule 12,000 bases long is said to be 12 kb long, while a double-stranded molecule of the same length is 12 kbp long.

The "base sequence" of a DNA strand is the order in which the nitrogenous bases occur in the strand of DNA. In Figure 1, for example, the base sequence of the left-hand strand of the DNA molecule shown is 3′-ACGTAGT-5′, and the right strand complementary sequence is 5′TGCATCA-3′. The polarity of DNA molecules renders the sequences stereochemically (and genetically) unique. Thus, these strands have entirely different biological meanings dependent on whether they are read from 3′ to 5′ or from 5′ to 3′.

The four bases, thymine, adenine, guanine, and cytosine (T, A, G, and C, respectively), compose the genetic alphabet. Sequences of these bases form the words and punctuation signals. The polarity of the sequence determines whether one is reading the language in a forward or backward direction.

The chemistry of DNA molecules just outlined gives rise to two fundamental properties of DNA that account for its biological capacity to store and express information. First, for a variety of physical reasons, double-stranded DNA molecules are the most stable form in most organisms. Second, the stable double-stranded forms of the molecule can exist only if the sequence of bases on one strand is "matched" by a complementary sequence of bases on the opposite strand. In other words, each of the four bases form stable hydrogen bonds to only one of the four bases on an opposite strand. Thus, adenine will form hydrogen bonds only to thymidine, while guanine will bond only to cytosine. These bonds result in base pairs. There are A-T and G-C base pairs, but under ordinary circumstances base pairs do not form from A-C, A-G, G-T, etc. The biological consequences of these base pairing rules are immediately apparent: the sequence of bases on one strand immediately dictates the sequence of bases that can occur on the opposite or complementary strand. The sequence 5′-GTACCGGAT-3′ on one strand of DNA will form double-stranded structures only with a complementary DNA strand containing the sequence 3′-CATGGCCTA-5′. The two strands form a stable double-stranded DNA helix only when they are arranged in an antiparallel fashion. In other words, one strand in the 5′-3′ polarity is bound to its complementary strand with a 3′-5′ polarity.

Given these chemical rules for formation of double-stranded DNA molecules, the ability of DNA to carry information in the form of a base sequence and transmit it to subsequent generations of cells becomes clear. If the two strands of DNA having complementary base sequences are separated, and each strand is "copied"

by the appropriate polymerase enzymes, the rules for stable base pairing dictate that each of the newly formed daughter strands be complementary in base sequence to the parent strands used as template. Since the two parent strands were originally complementary to each other, the result is two double-stranded DNA molecules, each pair of strands identical to the other pair. If each molecule is partitioned into a daughter cell, as occurs during mitosis and cell division, each of the daughter cells then contains the exact genetic information that was present in the parental cells.

For this chemical information to be converted into phenotypically effective biochemical properties of the cell, there must be additional reagents (macromolecules) and rules for decoding the information in biologically relevant ways.

CONVERSION OF GENETIC INFORMATION IN DNA INTO USEFUL BIOCHEMICAL PROPERTIES

The genetic information contained in the base sequence of a strand of DNA is expressed first by synthesis (transcription) of a molecule of RNA called messenger RNA. RNA is also a linear polymer containing nucleotide subunits composed of sugar, phosphate, and nitrogenous bases. Its structure is similar to that of DNA, except that the sugar is ribose rather than deoxyribose, and the nitrogenous pyrimidine base uracil is used in RNA in place of the thymine base in DNA. For our purposes, the three biologically important forms of RNA that occur in cells are messenger RNAs (mRNA), transfer RNAs (tRNA), and ribosomal RNAs (rRNA). The rRNAs form part of the structural scaffolding of ribosomes utilized for translation of genetic information into the amino acid sequence of proteins. The tRNA molecules serve as the "adaptors," which recognize particular sequences within a messenger RNA molecule and insert corresponding amino acids in the proper sequence of a protein being synthesized on ribosomes. The mRNA serves as the intracellular copy of the gene that carries the information from the nucleus, where its production is regulated, to the cytoplasm, where its information is converted into protein. Thus, the ability of a base sequence of DNA to affect the properties of the cell is based on what has come to be known as the "central dogma" of molecular biology. That is, a DNA molecule is copied into an RNA molecule (mRNA), and the base sequence in the mRNA molecule is translated into the amino acid sequence of a protein through the mediation of rRNA and tRNA molecules. The proteins thus produced possess the enzymatic or structural capabilities that determine the phenotype of a cell. This central dogma of genetic information flow or gene expression can be summarized as DNA → RNA → protein.

The rules by which the base sequence of a strand of mRNA is converted into the amino acid sequence of the protein are summarized by the term "genetic code." The base sequence of the mRNA copy, or transcript, of the DNA coding strand is read as a series of consecutive, non-overlapping triplets (group of three

bases). Thus, the mRNA sequence 5′-AUGUGGUUU-3′ specifies the incorporation of three amino acids: N-methionine-tryptophan-phenylalanine-C. The specificity that allows a particular triplet, called a codon, to direct incorporation of only one type of amino acid at that position is mediated by the tRNA molecule. For 61 of the 64 (from 4^3) possible codons that can be formed from triplets of the four nitrogenous bases, there is a tRNA molecule which, at the appropriate position along its nucleotide sequence, contains the complementary triplet called the anticodon. For three codons, UAG, UAA, and UGA, there is no corresponding anticodon; these three codons serve as "stop" or "termination" signals used to designate the position at which assembly of the protein chain stops. These three codons are sometimes called "nonsense" codons.

The anticodons present in tRNA bind to the mRNA codon as each codon is exposed within the ribosome. Thus, for example, when the codon 5′-AUG-3′ is exposed, the only tRNA anticodon that can bind at that point is 3′-UAC-5′. The adaptor function of the tRNA molecule arises from the fact that each tRNA species contains a site capable of covalent binding (activation) to only one of the 21 amino acids utilized in proteins. Thus, the tRNA with the anticodon 3′-TAC-5′ can be activated only with methionine. The codon to anticodon binding at the appropriate position thus causes methionine to be brought into that position of the growing polypeptide chain, where it is donated from the tRNA molecule to the nascent polypeptide. In this manner, the code in the mRNA molecule is translated into the amino acid sequence of the protein as the ribosome reads the mRNA like ticker-tape. Note the antiparallel nature of the polarity as one proceeds from the coding strand of DNA (3′-5′) to mRNA (5′-3′) to tRNA anticodon (3′-5′) to amino acid sequence of the protein (N-C). Figure 2 shows a diagram of the flow of genetic information from DNA to RNA to protein.

There are 64 possible codons and only 21 amino acids utilized for protein synthesis; thus, for some amino acids there are several codons. In this sense the code is degenerate. For example, there are six codons specifying the incorporation of leucine, but only one specifying the incorporation of methionine. Therefore, knowledge of the amino acid sequence of the protein does not immediately reveal the nucleotide sequence of the DNA and RNA molecules responsible for its production. However, knowledge of the nucleotide sequence of the relevant DNA and RNA molecules immediately predicts the amino acid sequence, since in no case does a single codon specify the incorporation of more than one type of amino acid. The genetic code is thus degenerate but not ambiguous.

The codon AUG has a special meaning in gene expression. In eukaryotic cells (the cells of all plants and animals except fungi and microorganisms), all proteins begin with methionine. The AUG codon serves as the start, or initiation, signal. AUG codons utilized to initiate protein synthesis are recognized by a specialized methionine tRNA that has the TAC anticodon but is unique elsewhere in its nucleotide sequence. Presumably, the secondary structure of this molecule is particularly suited for initiating protein synthesis. Methionine incorporated at internal positions also utilizes the AUG codon, but a different methionine tRNA is employed. This tRNA also has the TAC anticodon but differs from initiator tRNA

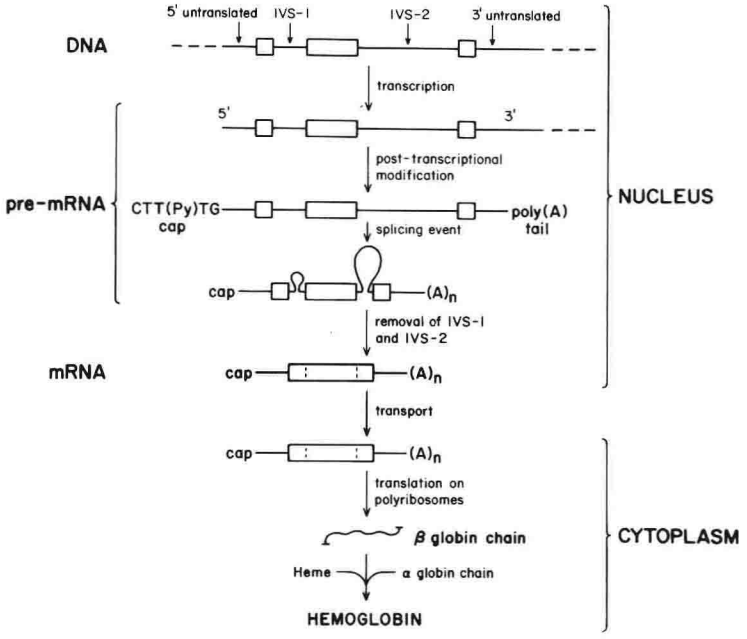

Fig 2.—The structure of a typical gene, the human β-globin gene. *Black bars* are mRNA coding regions (exons). *IVS-1,* intron 1; *IVS-2,* intron 2.

elsewhere in its sequence. Very few mature eukaryotic proteins retain methionine as the N-terminal amino acid. Rather, the methionine is cleaved from the nascent polypeptide chain at a later phase of protein synthesis.

We have reviewed the basic tenets of genetic information storage and flow in the form of chemical information stored in DNA base sequences. We have introduced the essential vocabulary by which nucleotide sequences are converted into amino acid sequence of proteins. In this regard, virtually all species are identical. However, eukaryotic and prokaryotic (microbial) species differ significantly in the manner in which DNA sequences are organized into the functional units that we call genes. Hereafter we shall consider only that system which applies to most eukaryotic genes.

STRUCTURAL ORGANIZATION AND EXPRESSION OF A TYPICAL EUKARYOTIC GENE: THE HUMAN β GLOBIN GENE

Eukaryotic cells contain their DNA within the nucleus in the form of nucleoprotein complexes called chromosomes. For example, human cells contain 46 chromosomes. Each chromosome consists of a single, incredibly long molecule of