THE GENE

ITS STRUCTURE, FUNCTION, AND EVOLUTION

LAWRENCE S. DILLON

The Gene

Its Structure, Function, and Evolution

LAWRENCE S. DILLON

Texas A&M University
College Station, Texas

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Dillon, Lawrence S.
The gene: its structure, function, and evolution.

Bibliography: p.
Includes index.
1. Genetics. 2. Molecular genetics. I. Title.
QH430.D539 1986 575.1 86-25291
ISBN 0-306-42319-7

© 1987 Plenum Press, New York
A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013

Printed in the United States of America

Preface

For a long period in the early years of genetics, the gene was viewed as a hypothetical carrier, located in the cell, of a given hereditary trait. Then, with progress in cytogenetics, it became a specific locus on a certain chromosome, and still later, a sequence in the DNA molecule, after the chemical composition of chromosomes had been established. But that was back in a period of relative naïveté, before advanced technology permitted the rapid determination of the precise macromolecular structure of almost any gene. Yet, ironically, as understanding of chemical structure has become increasingly concrete, ideas of how the molecular unit induces given phenotypic traits—particularly those of multicellular plants and animals—have decreased in lucidity. Only in the most recent literature has any light been thrown on basic changes that can lead to mutations in the expression of, for example, eye color or wing length in *Drosophila.* All that is known is that a gene encoding a certain enzyme is modified, either directly or in the flanking regions, and that in specific cells that enzyme somehow leads to the production of white instead of the usual reddish eye color or to vestigial instead of normally functional wings.

The flanking regions just mentioned are assuming far greater functional importance in current molecular biological thought than was the case previously. Indeed, when the vast distances in the DNA between actual coding areas were first discovered, those ''spacers'' were often referred to as ''junk DNA,'' being deemed to have no, or virtually no, active function in the cell. Gradually, however, the pendulum has swung toward the opposite extreme, so that much of the regions around a gene as far as 350, or even 1000, nucleotide base pairs from either end of the coding sector has often been implicated in the expression of some gene in one organism or another. ''Boxes'' of innumerable types have been described as active in triggering initiation or termination of transcription or other molecular activity. These signals, although supposedly universal, too often tend to be confused or absent when sought in other genes in the given organism. So many generalizations based on limited data have been made that future generations may well view the present period of exploration into the nature of the gene as a new age of naïveté which sought quick solutions to a problem that by then will have been most thoroughly demonstrated to be one of infinite complexity.

It is the author's aim to assist in the taking of the first steps across the threshold into understanding this most vital of all biological problems by thoroughly analyzing the genes and their surrounding sequences from a broad spectrum of organisms and the eukaryotic organelles. As many types as feasible are examined, particularly those found in both prokaryotes and eukaryotes, such as the transfer and ribosomal RNAs and cytochromes, along with others that are widespread and sufficiently characterized. No preconceptions are held as these gene structures are analyzed, in order to avoid bias, for the statement made by a geologist, "I will see it, if I believe it," unfortunately applies equally in all scientific investigations. Support of some particular dogma is not what is sought here, only an insight into how genes function insofar as the data permit. In short, the author attempts to follow the guidelines laid down by Thomas Henry Huxley: "Sit down before fact as a little child, be prepared to give up every preconceived notion, follow humbly wherever and to whatever abysses nature leads, or you shall learn nothing."

This being the first unified analysis of the gene on a broadly comparative basis, it is inevitable that new general characteristics and interrelationships among genes as entities in their own right should be disclosed. Genes, for example, have organizational properties unrelated directly to their encoded products that show them to fall into several major classes, each with its own distinctive properties. Indeed, on more than several occasions, the data indicate a need for hierarchical arrangements to be made, whether for the genes themselves, for their organization in the genome, or for the manners in which they at times overlap one another. At other points, the information does not permit firm conclusions to be drawn, but is sufficiently substantial to disclose new questions of extensive impact to be posed. Although evolutionary matters necessarily are interwoven with all aspects of the subject, the time is not yet here when it will be possible to demonstrate a mechanism for direct environmental influence upon the course of genetic changes, although faint hints at its existence do come to light.

Acknowledgment of the assistance of a number of persons is gladly made. In the first place, no study of this type could have even been begun without the innumerable researchers whose brilliant technological achievements patiently applied to solving particular problems in the laboratory have supplied the mass of detail needed for the present analysis. To all those persons, cited in the literature references or not, the author expresses his greatest admiration and deepest gratitude. Others have contributed in a more direct manner, particularly H. Faaren of Agrigenetics, who generously supplied information regarding certain seed proteins. Discussions with T. M. Hall and H. W. Sauers of Texas A&M University also have been of great value, and the author greatly appreciates their assistance, as he does that of many others who must remain unnamed here. Likewise, many persons have assisted with the preparation of the numerous tables, but Molly Allen and Esse Bakor have been especially helpful. Finally, my wife, as always, has been an indispensable collaborator throughout all the preparatory stages of the manuscript, literature search, preparation of tables, and interpretation of data.

LAWRENCE S. DILLON

The Gene

Its Structure, Function, and Evolution

Contents

1

Major Features of the Gene

Over the years, the gene has received numerous definitions, many of which became widely accepted. It has long been known as the unit of inheritance, or the factor that results in a given hereditary trait, definitions that continue in use today; it was also considered a point, or locus, on a chromosome that serves in the foregoing capacity. Later, as the molecular aspects of cell function began to unfold, a better understanding of the nature of the gene was gained, and the concept of ''one gene, one protein'' came into existence. Then, when proteins were perceived as being constructed of several subunits, each the product of a separate locus, that view became modified to ''one gene, one polypeptide.'' All these ideas are both sound and unsound; even the most recent, which defines a gene as a sequence in the nucleic acid of the genome, contains an element of weakness, which will become apparent as the discussion of gene structure proceeds immediately below.

1.1. STRUCTURAL FEATURES OF THE GENE

The gene, like all biological units, has been found to be extremely diversified, even in its basic construction, but the numerous species can be perceived to fall into three major classes: simple, compound, and complex. The members of the last category, which according to present knowledge are confined to advanced metazoans, are extremely intricate, as is disclosed in a subsequent section. But even the first of these types is soon found to be simple only by comparison.

1.1.1. The Simple Gene

Far back in primordial times, the gene in the archetypal biological forms may have been merely a continuous sequence in a nucleic acid that specified a particular polypeptide or functional RNA, but that condition would have been confined to primitive organisms in which the genetic coding system was still undergoing completion. Perhaps that condition persists in some of the early descendants of those archaic living things, such as the simple RNA viruses (Chapter 10), but even in the better known bacteriophages of today, a gene is far more than just a continuous sequence of nucleotides encoding a particular macromolecule, as it certainly is also in cellular organisms.

The Coding Sequence. The primary functional portion of the gene obviously is the sequence of nucleotides that encodes a particular macromolecule which plays a role in the structure or metabolic processes of a cell. Thus the coding strand of this region of DNA is read by an enzyme that copies it in complementary fashion into RNA strands. These products then may be used, usually after further treatment called processing, either directly, as with the ribosomal and transfer RNAs, or, as in the case of messenger RNAs, after translation on ribosomes into proteins or their polypeptide subunits. Because the first products of the gene are in sequences of RNA that are complementary to the coding strands of DNA, it is customary in studies of the gene at the molecular level to present the complementary strand sequence of the duplex DNA, rather than the actual active one. Since that procedure simplifies comparison of the gene and its product, it is followed throughout the present study (Sharp, 1985).

Intergenic Spacers. It may be readily imagined that in the earliest protobionts that had genes coded in nucleic acid the latter consisted solely of coding sequences placed one after another. Perhaps these were arranged in uninterrupted series, for occasionally even in modern organisms two or more genes may abut against each other in this fashion. Indeed, as shown shortly, they sometimes even overlap. More typically, however, genomic regions that are not directly involved in the final product are found between the actual coding sectors. These portions, called intergenic spacers in a general sense, vary greatly in length, from a single base pair to many thousands (Federoff, 1979). Also in a nonspecific manner, the parts of these spacers adjacent to the mature gene are frequently referred to as flanking regions.

Since certain parts of these spacers exhibit specialization of function, separate terms are applied. That sector of the noncoding strand preceding the 5′ end of the gene proper (hereafter referred to as the mature gene) is called the leader (Baralle, 1983), while that following the 3′ end is the train (Figure 1.1A). If the habit of viewing the complementary strand is abandoned temporarily, it may be seen that the leader is actually located before the 3′ end of the functional region of the coding strand and that the train follows the 5′ end, because of the antiparallel arrangement of DNA strands (Figure 1.1A). This reversed situation should be borne in mind in order that the real nature of gene structure and function may be understood.

It also should be realized that one given strand of DNA does not necessarily bear the coding sequences of all the genes located in any single sector of the genome; sometimes some genes are located on one strand while others in the same series are on the opposite one. Since genes (or cistrons as they are frequently termed) are always read (transcribed) in the 3′ to 5′ direction on the coding strand, those on opposite strands also differ in polarity. Hence, it may be perceived that in instances where a number of genes form a cluster, as is the frequent case in both prokaryotes and eukaryotes, the train of one gene may be located under a mature gene of opposite polarity or even below its leader (Figure 1.2A). This topic receives full attention in Chapter 10.

Promoters. Before a gene can be read by the responsible enzyme, the correct coding region obviously must be located and its polarity determined. Since the nucleoids of prokaryotes and the chromatin of eukaryotes consist of many millions of nucleotides of only four major varieties, the task of locating the proper sector, which is comprised of but

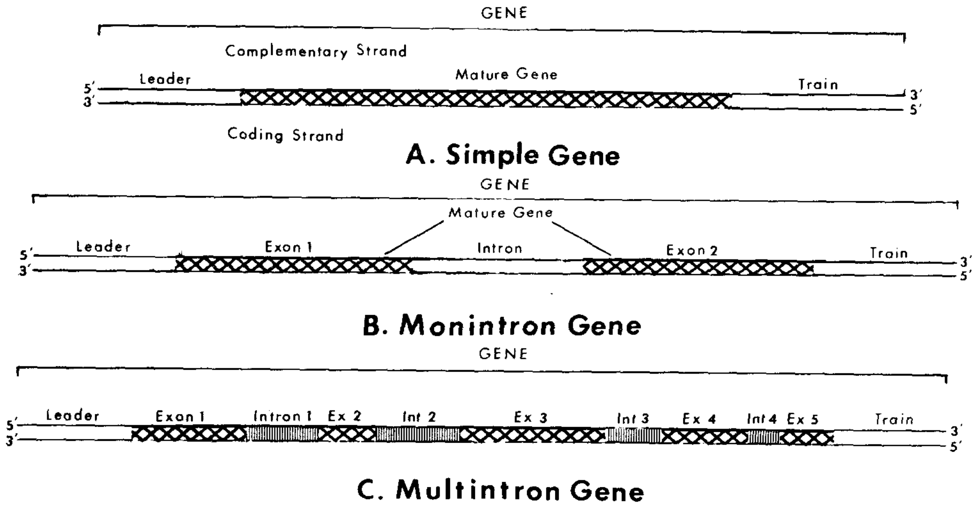

Figure 1.1. Varieties of simple genes. (A) Those genes that consist of the mature coding region plus essential parts of the flanks are considered members of the simple class. (B, C) Sometimes this basic pattern is modified by the presence of one or more introns.

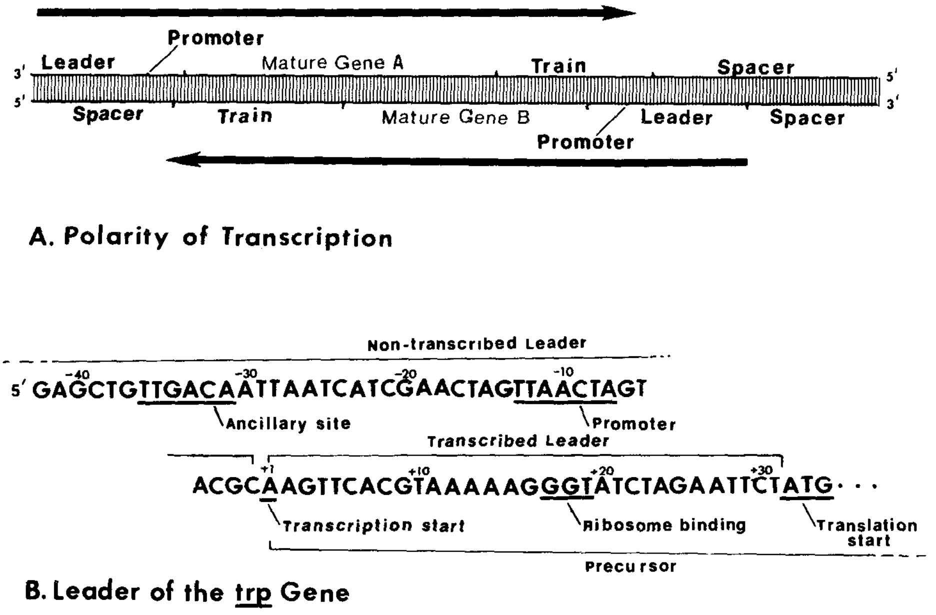

Figure 1.2. Principal features of genes. (A) Not all genes are necessarily confined to a single strand of DNA, but may occur on either strand. Because transcription is always in the 3' → 5' direction, those on opposite strands have different polarities. (B) Structural relations at the 5' end of a gene.

a few hundred nucleotides of the same four basic types, can be quite formidable. To assist in locating, reading, and processing the gene and its product, a number of signaling devices appear to exist. To delineate the nature of these signals, when they are actually present, is one main goal of this book. It will be found that, although the principle of having such devices present seems a simple necessity, in practice their identification by cellular processes is without equal insofar as complexity is concerned.

In the literature a number of different "boxes" have been proposed to serve in recognition purposes. The Pribnow or −10 box, of varying constitution, is often cited in studies of prokaryotic genes, while in those on eukaryotic cistrons the term Goldberg–Hogness box, of equally variable constitution, applies to a somewhat corresponding location. Various other names, including CAT-TA, ACT-TA, and TA-TA boxes, also have been in vogue. Since bacteriologists had named such functional sectors many years before sequencing DNA had become readily feasible, their term, promoters, is uniformly applied here to such signals to permit freer comparison from gene to gene and organism to organism. Promoters may be viewed as the specific site or sites of a given gene that serve in the latter's recognition and subsequent attachment by the transcribing enzyme (Figure 1.2B). Transcription usually begins a variable number of sites downstream (in the 3′ direction on the complementary strand), in *Escherichia coli* typically at a purine residue, but it varies widely from organism to organism.

Ancillary Sites. Another active site on the leader has frequently been reported to which several terms have been applied, as in the case of the promoter. Since this is frequently located between 30 and 40 sites upstream (in the 5′ direction on the complementary strand) from the transcriptional initiation site, it often is referred to as the −35 sequence, particularly with prokaryotic genes (Figure 1.2B). However, it is also called the CAP site in certain bacterial coding sequences, a topic investigated further in Chapter 2. For the sake of uniformity demanded in studies comparing prokaryotes, eukaryotes, and organellar structures, the name ancillary site is employed here. It is at such points that enzymes (ancillary proteins) that assist in the location and transcription of the gene may attach, as shown in the next chapter. This avoidance of the use of the bacteriological term CAP (referring to a cAMP-activated protein) has the additional advantage of reducing confusion, because many eukaryotic transcripts receive a cap in quite a different sense. Characteristically eukaryotic messengers receive a highly modified nucleotide on the 5′ terminus, usually methylated and linked to the main chain by 5′ to 5′ bonds (Adams and Cory, 1975; Perry and Scherrer, 1975). One of the commoner caps is m^{7}GpppNm- (Plotch *et al.*, 1981), often referred to as cap 1.

Enhancers. Still another site on the leader, quite recently discovered, is the enhancer, the prototype of which was found to be a 72-base-pair, tandemly repeated sequence (Benoist and Chambon, 1981; Gruss *et al.*, 1981; Weber *et al.*, 1984). This sequence was located 100 nucleotides upstream of the ancillary site in the DNA of simian virus (SV) 40. Mutational removal of this element reportedly reduced expression of early genes by a factor of at least 100, resulting in the loss of viability of the virus. A similar enhancer region has been described from other viral sources; a feature common to all of these is the series GGTGTGGAAAG, a sequence that occurs frequently in modified form (Khoury and Gruss, 1983). In actuality this series in SV40 is part of a 72-base-pair repeated sequence (Figure 1.3), but in other viruses it may be a portion of 50- to 100-base-pair repeated elements, only one of which is essential for enhancement.

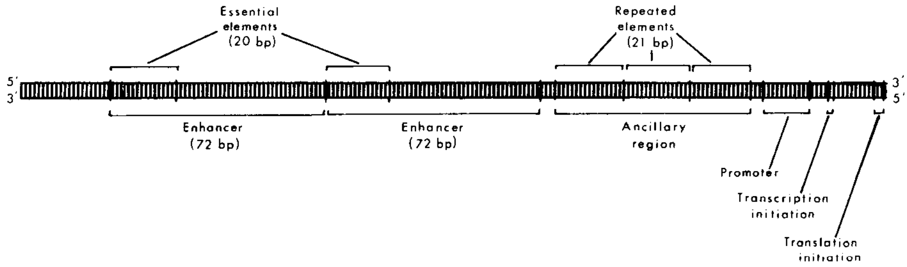

Figure 1.3. An enhancer region from SV40. (After Benoist and Chambon, 1981).

Enhancers have been detected associated with eukaryotic genes also, as shown frequently in subsequent chapters, studies on immunoglobulin genes proving to be especially productive (Ephrussi *et al.*, 1985; J. O. Mason *et al.*, 1985a; Mercola *et al.*, 1985).

Whether enhancers represent a single type or an entire class of related substances is a debatable point at this time, for they have been found to display a number of distinctive traits. Among those that indicate them to be a class of substances are the following characteristics: (1) They act most frequently when located on the coding strand, that is, in the *cis* configuration, but *trans*-acting ones also are known (Chapter 11). (2) They are effective when located either upstream or downstream from the promoter. (3) They are active whether arranged in the same or opposite polarity as the mature gene. (4) They are equally effective regardless of the organism from which the gene is derived when attached to foreign DNA. An example of the latter property is the finding that certain regions from a mouse immunoglobulin gene served as an enhancer when cloned to the SV40 early promoter (Mercola *et al.*, 1983). Other eukaryotic genes, however, have been shown to resist enhancement, including the α-globin gene of man.

Other Leader Signals. With the exception of the initiation site, all of the foregoing signals are frequently located in the nontranscribed region of the leader and thus are not part of the nascent RNA molecule (primary transcript) transcribed by the enzyme. Usually only one further signal follows the initiation site on the transcribed leader of messengers, those RNAs (mRNAs) that are subsequently translated into proteins (Figure 1.2B). Typically this feature, which clues the initiation point for translation, is the codon for methionine (ATG), but rarely certain codons for other amino acids can serve in this same capacity. As pointed out in greater detail later, in prokaryotes an additional signal is present here, because translation of mRNAs is required to occur concurrently with transcription. Accordingly, bacterial transcripts have a short sequence (the Shine–Dalgarno box) that provides a point of attachment for the ribosome, hereafter referred to merely as the ribosome binding site (Figure 1.2B).

The position and reading frame orientation of the ATG codon (AUG in the messenger) are apparently not the sole determinants for initiation of translation, as effectively demonstrated by a recent study (Johansen *et al.*, 1984). When the AUG was experimentally inserted into the leader of an mRNA, it had various effects on translation, depending upon the sequences that flanked it, the combination A(or G)CC before the AUG and a G after it being the most effective. It is pertinent here also to note that GUG or UUG may sometimes be substituted for the standard signal, as is seen in several instances in later chapters. However, these are uniformly translated as methionine, although the first alternate usually encodes valine and the second leucine. Rarely AUU (usually for isoleucine) may serve in this capacity, as in the carbamoyl phosphate synthetase of *Escherichia coli* (Piette *et al.*, 1984). Here, then, is the first example of a condition that recurs repeatedly in later chapters, the nonmechanistic behavior of mechanistic molecules, which has on earlier occasions provided the basis for a more realistic philosophy of living matter, referred to as the biomechanistic point of view (Dillon, 1978, p. 427; 1983, pp. 400–410).

Promoters of Prokaryotes. While the several chapters devoted to transcription provide an in-depth analysis of the nature of promoters from individual classes of genes, a preliminary examination of several examples at this point is essential to a more concrete understanding of the gene's structural features. Accordingly, Table 1.1 lists a number of leader sectors from the well-known bacterium *E. coli,* showing their promoters located about ten sites from the actual start of transcription indicated by +1. The predominant

Table 1.1
Promoters of Sets of Genes from E. coli[a]

		Ancillary Site			Promoter	
		−35	−30	−20	−10	+1
trp[b]	GAG-CTG	*TTGACA*	-ATTAATCA	TCGAACTAG-	*T-TAACT*	AGTACGCA
lacUV5[b]	CAGGC--	*TTTACA*	CTTTATGCT	TCCGGCTCG-	*TATAATG*	TGT-GG-A
pheU[c]	--TTAGG	*TTGACG*	-AG-ATGTG	CAGATTACGG	*TTTAATG*	CG-CCC-G
leuV[c]	---ACTA	*TTGACG*	AAA-A-GCT	GAAAACCAC-	*TAGAATG*	CGCCTCCG
tyrT[c]	--AACAC	*TTTACA*	-GCGGCGCG	CGTCATTTGA	*TATGATG*	CGCCCC-G
tyrT[d]	GTAACAC	*TTTACA*	-GCGG--CG	CGTCATTTGA	*TATGATG*	CGCCCC-G
rrnA[c]	---CCTC	*TTGTCA*	GGC--CGGA	ATAACTCCC-	*TATAATG*	CGCCACCA
rrnD[c]	---ATAC	*TTGTGC*	AAAAAATTG	GGA--TCCC-	*TATAATG*	CGCCTCCG
str[e]	ATATTTC	*TTGACA*	CCTTTTCGG	CATCG-CCC-	*TAAAATT*	CGGGC--G
rpoB[f]	ATATACT	*GCGACA*	GGACGTC--	CGTTCTGTG-	*TAAATCG*	CAATGA-A
glnS[g]	CTAACAG	*TTGTCA*	GCCTGTC-C	CGCTT-ATA-	*AGATCAT*	AC-CCC-G
rpsT$_1$[h]	GGAAAAG	*CTGTAT*	TCA-CACCC	GCAAGC-TGG	*TAGAAT-*	CCTGC--G
rpsT$_2$[h]	AAATCCA	*TTGACA*	AAAGAAGGC	TAAAA--GGG	*CATA-TT*	CCTCGG-C

[a]Promoters and ancillary sites are italicized; +1 marks the actual transcriptional starting site.
[b]De Boer *et al.* (1983).
[c]Schwartz *et al.* (1983).
[d]Sekiya *et al.* (1976); Berman and Landy (1979).
[e]Post *et al.* (1978).
[f]Post *et al.* (1979); An and Friesen (1980).
[g]Hawley and McClure (1983).
[h]Mackie and Parsons (1983).

Table 1.2
Promoters of Gene Sets of Various Prokaryotes[a]

		Ancillary Site				Promoter	
		-35	-30	-20		-10	+1
rrnA$_1$[b]	TATTATGTA	*TTGACT*T	AGACAA	CTAAAGC-T	GT-	*TATTCT*	AATATAC-G
rrnD$_1$[b]	TCATAACCC	*TTTACA-*	-GTCAT	AAAAATTAT	GG-	*TATAAT*	CATTT-C-G
rrnA[b]	AAAAAGTTG	*TTGACA-*	GTAGCG	GCCGGTAAAT	GT-	*TATGAT*	AATAAA--G
rrnD[b]	AAAAAAGTA	*TTGACCT*	AGTTAA	CTAAA-AAT	GT-	*TACTAT*	TAAGTA--G
spoVG$_L$[c]	GGATTTCAG	AAAAAAT	CGT-*GG*	*AATTGATA-*	CA-	*-CTAAT*	*GCTTT*T--A
spoVG$_E$[c]	TTAAAAACG	AGCAGGA	TTTCAG	-AAAAAATC	GT-	*GGAATT*	*GATA*CA--C
spoVC[c]	CATTTTTCG	AGGTTTA	AATCCT	TATCGTTAT	GG*G*	*TATTGT*	*TTGT*AAT-A
penP[d]	AAAAAACGG	*TTGCAT*T	AAAATC	TTAC-ATAT	GT-	*AATACT*	TTCAAA--G
malX[e]	AAAAAATAC	*TTGCAA*C	CGTTTT	CTAT-TTGT	GC-	*TATACT*	AAGCTC--A
malM[e]	TTAAAACGC	*TTGCAA*T	TATGCG	TTGAAAAG-	GAG	*TATACT*	TATAAGT-A
nifH[f]	ATACATAAA	CAGGCAC	GG*CTGG*	-TATGT-TC	C*C*-	*TGCA*CT	TCTCTGC-T
nifE[f]	AAAATCAAG	GCTCCGC	TT*CTGG*	-AGCGC-GA	A*T*-	*TGCA*TC	TTCCCCC-T
nifL[f]	CTGCACATC	ACGCCGA	TA*AGGG*	-CGCACCGG	T*T*-	*TGCA*TG	GTTATCACC

[a]All experimentally established signals are italicized. +1, Actual transcriptional start site.
[b]*Bacillus subtilis;* Ogasawara *et al.* (1983).
[c]*Bacillus thuringiensis;* Wong *et al.* (1983).
[d]*Bacillus licheniformis;* McLaughlin *et al.* (1982).
[e]*Streptococcus pneumoniae;* Stassi *et al.* (1982).
[f]*Klebsiella pneumoniae;* Beynon *et al.* (1983).

base at each given site in these signals can be selected to give a "consensus" sequence, in the present case TATAATG. But such average compositions tend to give a distorted picture of how genes are recognized, for obviously the RNA polymerase does not react with a consensus through a wide spectrum of types, but with the particular combination that exists in each actual gene. For example, the cited consensus sequence is found in only two of the 13 that are listed, those for the two sets of rRNA genes, *rrnA* and *rrnD*. Such strong deviations from the norm as AGATCAT can be noted, in this case in the *ginS* gene, in which none of the bases corresponds to the consensus. While it is evident that transcription in *E. coli* usually is initiated with a purine nucleotide, a cytosine residue is employed in one of these examples.

Although the composition of the ancillary signal and promoter may thus vary over broad ranges, they are spaced at a markedly uniform distance from the initiation sites in *E. coli*. When samples of structure are compared from other bacteria, however, as in Table 1.2, variations in size and location as well as in composition are found to be rampant. This lack of uniformity is apparent even within a single genus, as among the three species of *Bacillus*. Whereas the signals of *B. subtilis* and *B. licheniformis* are identically located and similarly constructed, those of *B. thuringiensis* are widely disparate. In the first place, the promoters of this third species consist of 10 or 11 nucleotide residues, compared to six in the other two and to seven in *E. coli,* so that their 3′ ends are only one to three sites removed from the start of transcription (indicated by +1 in the table). Second, upstream signals are often lacking, being present only in the late form of *spoVG,* distinguished by the subscript L. During the early stages of sporulation this identical sector serves as the promoter, as shown in *spoVG*$_E$ in the table.

The two maltosaccharide utilization genes of *Streptococcus, malX* and *malM,* have signals quite comparable to *B. subtilis* and *B. licheniformis,* so that these three form a cluster of interrelated species not too remote in kinship from *E. coli.* On the other hand, *Klebsiella pneumoniae* is seen to be distinct from all the others given here in having both the promoter and the ancillary site shorter and of distinctive base composition. Moreover, the latter are situated downstream of those from bacilli and coliform bacteria. Whether these points of departure are suggestive of an advanced or more primitive condition cannot be made clear until the structures of protein-specifying genes of Archaebacteria and such primitive forms as *Clostridium, Beggiatoa,* and blue-green algae have been established in sufficient numbers.

Signals on the Trains. As might be expected, the chief signals on the 3′ trains of genes are those that are concerned with the cessation of transcription (Rosenberg and Court, 1979). Typically there exists a series of T–A base pairs, in prokaryotes usually six or more in number, while in metazoans four sometimes seem to suffice. Occasionally these series are accompanied on the upstream side by regions of dyad symmetry, so that a stem-and-loop structure could conceivably form in the transcript (Young, 1979; Holmes *et al.,* 1983). In some instances a combination of the stem-and-loop and the sequence of Ts seem to be requisite or at least most effective (Figure 1.4; Yanofski, 1981; Lau *et al.,* 1982)—the loop, of course, would form on the transcript, not in the DNA double strand (Platt, 1981), as illustrated by a *Bacillus subtilis* cluster of tRNA genes (Green and Vold, 1983). Such regions are rich in guanosine and cytidine residues.

But even in prokaryotes, termination is not always signaled in such a simple fashion, but often is dependent upon the presence of additional factors. While full discussion is more appropriate to the topic of polymerase activities, it can be mentioned now that two such factors are known, rho and NusA (Andrew and Richardson, 1985; Barik *et al.,* 1985; Bear *et al.,* 1985). Definite sequence requirements for attachment of these proteins have not been established, but sometimes are quite complex, being a series of tandem sectors

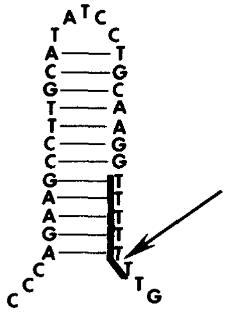

Figure 1.4. A possible terminator in a tRNA operon of *Bacillus subtilis.* (Green and Vold, 1983.)