Are Genes Units of Inheritance?

THOMAS FOGLE

Saint Mary's College
Notre Dame, IN 46556
U.S.A.

ABSTRACT: Definitions of the term 'gene' typically superimpose molecular genetics onto Mendelism. What emerges are persistent attempts to regard the gene as a 'unit' of structure and/or function, language that creates multiple meanings for the term and fails to acknowledge the diversity of gene architecture. I argue that coherence at the molecular level requires abandonment of the classical unit concept and recognition that a gene is constructed from an assemblage of domains. Hence, a domain set (1) conforms more closely to empirical evidence for genetic organization of DNA regions capable of transcription and (2) has ontological properties lacking in the traditional unit definition.

KEY WORDS: Genetics, gene structure, hereditary unit.

The 'gene' concept became a centerpiece for demonstrating Nagel's (1961) framework for theory reduction in the sciences when Schaffner (1967) attempted to sketch a reduction of Mendelian genetics (reduced theory) from molecular genetics (reducing theory) (see Rosenberg, 1985 for a review of the subsequent debate). Hull (1974) argued that many genes are needed to define an organismal phenotype and that a variety of phenotypic contexts can be influenced by a given gene. Because the relationship between genotype and phenotype is many-to-many, he claimed that the enormity of the premises needed to effect a formal reduction seriously undermined the Logical Positivist goal. Numerous arguments have been presented to obviate the need for reduction. Maull (1977) claimed that the relationship between fields (interfield theories), such as genetics and biochemistry, explains the transfer of terms better than reduction (absorption of one theory by another). Wimsatt (1976) questioned the virtue of rigorous deductive underpinnings stating that to attempt the reduction "is not even an effective *means* to some end because it is not the answer to a request for formal assistance which anyone has made or would be likely to make!" More recently, Kitcher (1982) argued that his perspective for conceptual change bypasses formal reductionism if it is acknowledged that a 'gene' has multiple referents sorted by mutually understood cues among members of a linguistic community (scientists). Lederman (1987) discusses one way in which juxtaposition of Mendelian and molecular referents successfully promoted detection of genetic diseases via DNA markers (segments of DNA positioned near a gene sorted by size after

Biology and Philosophy **5**: 349–371, 1990.

exposure to restriction enzymes). Implicit in both Kitcher's and Lederman's interpretation is the view that the juxtaposition of referents for the 'gene' has promoted research agendas. Kitcher states:

> We can *permit* 'gene' to have a highly heterogeneous reference potential. Those whose concern is with the continuation of the classical research of Sturtevant, Bridges, and Muller, the construction of gene maps for organisms, may reasonably employ the principles of segmentation used by their predecessors. Those who wish to resolve particularly difficult regions, determining genetic fine structure, will find Benzer's principle of segmentation helpful. Molecular biologists, when pressed to identify the referent 'gene', have adopted, almost universally, the view that genes (more exactly *structural* genes) are chromosomal segments which code for particular polypeptides.

Equally optimistic is Lederman's claim that "Biologists have internalised a great many concepts of 'gene' and can restrict the gene concepts they discuss or employ in a particular context." Despite proposed methodological advantages for the juxtaposition of 'gene' concepts it is also true, as the present paper demonstrates, that confusion and ontological consequences follow when the classical intention for 'gene' conjoins a molecular 'gene' with fluid meaning. The fact that geneticists unsuccessfully merge the two indicates that they fail to restrict meaning to a particular context as Lederman proposes. Therefore, the clear-cut sorting of meanings by empiricists, as portrayed by Kitcher, is not supported.

THE RELATIONSHIP BETWEEN THE MENDELIAN AND MOLECULAR GENE

Recent molecular genetics investigations on the physical organization of the gene reveals that the delicate bridge connecting the Mendelian and molecular points of view is near collapse. The conventional perspective for the molecular gene is a nucleotide sequence that manufactures RNA (whether tRNA, rRNA, or mRNA) as the primary level of expression. These are the regions of DNA that produce a tangible product (at the molecular level) and can most directly be associated with the sort of phenotypic variability that enabled geneticists of the pre-molecular era to postulate entities as causative. Both the Mendelian and molecular models are materially localized; the former being ill-defined as to physical characteristics but tandemly aligned through linkage analysis in the pre-molecular era and the latter bounded by a start and stop codon upon discovery of the genetic code. In this broad categorization of the material gene, the molecular model is a refined outgrowth of the older search to identify and explain deterministic factors. The early molecular gene had edges, a beginning and an end which were easily defined, and singular function — features that permitted the particulate model to be updated with detail while retaining intact the paradigm that genes are discrete sections of DNA with common organization.

If a gene is a unit of inheritance, as often defined, attempts by molecular biologists to find a coherent relationship to the Mendelian entity, at first successful, are currently strained by the robust architectural diversity of the molecular gene. While the Mendelian gene acts in transmission as a unit, there is no unifying characterization at the molecular level. The essential ingredients of the transcription and translation story remain the same as in previous decades, yet the boundaries constituting a gene are increasingly difficult to define. If genes are the bones that lie beneath the flesh, the skeleton is looking less familiar.

My intent is to examine the architecture of protein coding genes from eukaryotes to argue that a gene is not constructed as a unit. I demonstrate the diversity of organization within and among genetic systems and pursue features that deviate from the classical picture of the physical gene to support my contention that the molecular gene is more appropriately described as a set of embedded, tandem, and overlapping domains. For manageability, I exclude discussion of genes that produce tRNA, rRNA, and the immune system; each contains a unique assemblage of structural features that lacks unit organization as well.

GENETIC INFORMATION THEN AND NOW

The unit concept for the gene pre-dates the rediscovery of Mendelism in 1900 when a diverse array of particulate hypotheses, such as biophores and gemmules, were postulated. In the post-1900 era of classical genetics, Mendelism merged with the more general notion of particulate inheritance as the theory of heredity evolved into the gene being a factor, a unit of inheritance. The term "gene," when used to convey a Mendelian entity that acts to effect a phenotypic outcome, has much the same meaning today as it did then.

The Mendelians of the early part of this century were, of course, unaware of the mechanism of gene expression. The unit-characters described by Bateson, an early and ardent proponent of Mendelism, conflated particulate inheritance with the features that explained them. It was Johanssen (1909) who proposed the separation of the genetic contribution from the biological contribution yet resisted assigning a specific entity to explain unit-characters (Falk, 1986). Nevertheless, the gene as a hypothetical construct quickly emerged among the early practitioners of Mendelism. The parameters of that construct were not universally established. Castle's long running debate with his contemporaries is an example of how a gene that followed Mendel's laws could be interpreted in various ways. Castle believed the gene to be an entity that could change when selection acted on it. When he disproved his own hypothesis (Castle, 1919) with his classic experiments on hooded rats, he

left the factor school uncontested. Morgan and his students refined the gene concept, but it was Muller who suggested that factors (genes) were material entities, not just theoretical constructs. The contrast between these two views lies in the belief that "for the one [Morgan] the gene was *determined* through its phenotype, for the other [Muller] it *determined* the phenotype" (Falk, 1986).

The physical gene elaborated in the molecular era was something with shape, defined structure, and comprehensible mechanics. Following at the heels of the bead-on-a-string model, it was an extension of the particulate entity arrayed onto chromosomes. The physical gene was therefore a linearly organized set of instructions that gave rise to phenotypic expression. The triumph of this empirical reductionism was the interpretation established for a gene identified using classical methods. If a trait followed Mendel's laws, as for example, the demonstration of a 3:1 inheritance pattern from the cross of two hybrid parents, then one postulated the existence of a gene, presumably rooted in a DNA sequence coding for some primary product. The ontology established, Mendelian genes were material units of information defined by a particular sequence of nucleotides capable of transcription.

By the time the first details of gene regulation were established in the *lac* operon of *Escherichia coli* (Jacob and Monod, 1961), the boundaries and classification of genetic units in the prokaryotic system were dissolving. The three structural genes and the regulator gene conformed to the molecular model of a message unit that produces a product. The adjectives structural and regulatory neatly differentiate their functions, but it is their similarity in active transcription which unites them. What then is the ontological status of the operator and promoter regions? Positioned adjacent to the *z* structural gene, they are necessary sequences of nucleotides which, in the case of the promoter, permit the transcription enzyme RNA polymerase to attach, and in the case of the operator, regulate whether the *z*, *y*, and *a* genes can transcribe. Neither is itself transcribed. In the jargon of the geneticist these are regarded as the operator *region* and the promoter *site*. That is, neither are given status as a gene yet both are heritable, capable of mutating, and influence the phenotype. Clearly, the physical gene has restricted meaning to geneticists and does not apply equally to all regions of the genome that effect the phenotype.

Efforts to find a system comparable to the *lac* operon in eukaryotes were not successful. This does not mean that the genetic architecture of eukaryotic systems is more straightforward. Quite the contrary. The eukaryotic genome has, in many ways, a fundamentally different organization that is far more complex, as exemplified by discovery of split genes (see section on Gene Structure) (Breathnach and Chambon, 1981).

The prokaryotic and eukaryotic systems implicate a multitude of

formats for genetic information, whereas the current juxtaposition of classical and molecular perspectives treats hereditary information as if there exists a dyadic system of units (genes) and non-units (repetitive DNA, regulatory sequences).

STRUGGLING FOR A DEFINITION OF THE GENE

The highly regarded and comprehensive text on molecular genetics by Watson *et al.* (1987), now in its fourth edition, exemplifies the difficulty of defining a gene. The authors grapple uncomfortably with the problem and reach this conclusion for prokaryotes:

> . . . it may actually make more sense to restrict the term *gene* to those DNA sequences that code for amino acids (and polypeptide chain termination) or that code for functional RNA chains, treating all transcriptional control regions as extragenic elements.

Their conclusion for eukaryotes is similar:

> Here again, logic may tell us that if we wish to continue to designate the gene as the unit of chromosomal function, then it should refer to those DNA segments that become transcribed into discrete RNA chains rather than to those DNA segments that provide the genetic code for given polypeptide chains (or more directly transcribed into the sequence of mature rRNA or tRNA molecules).

These authors define the eukaryotic gene by both function and structure and specify the physical entity to be a segment that produces RNA. Both definitions recognize that expressed information within that structure is important for living systems, not just the structure itself. Although the definition for eukaryotes refers only to a functional unit, the larger agenda is to localize a sequence of nucleotides that specify that unit. There is, therefore, an attempt to delineate the edges for a physical unit as well. Watson *et al.* have gingerly superimposed a molecular gene defined through phenotypic expression onto Mendelism (a unit of structure and function).

What Watson *et al.* want is a definition of the gene that has at its heart the classical meaning of the term as a unit and at the same time incorporates current empirical evidence from molecular biology. This is no easy task. Is it possible to define the gene as a unit of function? Can a structural unit be described? Given the existence of a material gene, there should be properties and boundaries that define that unit at the molecular level.

GENE STRUCTURE

To begin, I characterize what might best be described as a generic protein

coding gene. It contains major features typically associated with eukaryotic gene structure and serves to introduce basic concepts and terms.

Transcription is the process of making a complementary copy of RNA nucleotides from one of the two strands of DNA. A region proximal to the site of initiation, termed the promoter, serves as a site of attachment for RNA polymerase, the enzyme that links the RNA nucleotides to form the complementary chain (mRNA). Within the promoter is a highly conserved sequence termed the TATA box. It positions the start site of transcription and typically lies 20 to 30 nucleotides away from that point. A second sequence, termed the CAAT region, resides at variable locations farther from the start site than the TATA box and plays a prominent role in the frequency of initiation (Melton, 1987). Likewise, a short sequence of guanines and cytosines (GC box) serves a similar role and is associated with genes that are constitutively expressed (housekeeping genes). A distinguishing feature of GC boxes is the ability to promote bidirectional transcription. Housekeeping genes, which are the major class of mammalian genes, code for proteins that perform tasks essential for metabolism.

One or more enhancers of regulated genes can modulate transcription and may be some distance (thousands of base pairs) from the transcribed region. While promoters are relatively position dependent, enhancers are relatively position independent. Enhancers most commonly flank the transcription start site, but can lie within the transcribed region (Lewin, 1987).

The newly formed transcript separates from DNA, is "capped" by the chemical addition of a methylated guanine at one end (5′) and clipped of a nucleotide sequence at the other end (3′). A short sequence of nucleotides proximal to the cut site specifies where to cut. A long chain of about 200 adenine nucleotides (poly A tail) is then bonded to the 3′ end. The cap and poly A tail (termed post-transcriptional modifications) are not coded by the region that produced the transcript. In 1977 an additional modification of the transcript was identified. Sequences of nucleotides are internally removed from the mRNA and broken down, while the remaining sections are spliced together. The size and number of removed sequences (intervening sequences or introns) is characteristic for a given gene. The sections that are spliced together (exons) contain the nucleotide sequence to code for a polypeptide. Translating the message on a ribosome occurs by recognizing a start codon (AUG) and proceeding to interpret each consecutive three nucleotides as an amino acid until a stop codon (UGA, UAG, UAA) is reached. Bordering the reading frame are non-coding nucleotide sequences termed the leader (5′ end) and trailer (3′ end).

Most, though not all, protein coding genes have introns and among those that do there can be as few as one or more than 50. As much as 90% of the primary transcript can be introns. The set of exons for a given

gene contains the leader, the coding region, and trailer sequence. A particular exon can be entirely non-coding or coding or it can straddle either the leader or the trailer and part of the coding sequence.

THE GENE AS A UNIT OF FUNCTION

The definition of the eukaryotic gene by Watson *et al.* describes it as a "unit of chromosomal function." If by function they mean chromosomal activity only, then it is hard to argue with their perspective. However, the nature of genetic expression, what gives it significance in the cell, is its activity (function) beyond just being something a chromosome does. Genes contain information for cellular structures and processes. To be a unit of function in this broader sense has long been part of the debate about the nature of the gene.

The classical model of the gene as a unit of function (a cistron in Benzer's (1957) language) suggests that an amino acid sequence has a singular function through the production of a polypeptide. From a strictly deterministic standpoint, the protein product effects some defined process in the cell thus initiating a sequence of causal events. Conversely, an organismal phenotype can be causally linked to a unit of inheritance by applying Mendel's laws. Because the classical gene is inferred from the phenotype, there is an apparent direct correspondence between the phenotype and the gene that codes for it. With the advent of more refined genetic analysis at the subcellular and molecular level following the discovery of the double helix, there began a stepwise bridging of the gap between gross phenotype and the molecular gene. Nowhere is this better epitomized than with the well known sickle cell anemia story. The altered DNA sequence substitutes a valine for a glutamic acid in the sixth position of the hemoglobin beta chain. The sickled cells resulting from this difference impede blood flow and deplete oxygen from localized sites of tissue causing clinical symptoms. The appeal of this example lies with the connectibility from one level of organization to another even up to the level of populations (from the evidence for heterozygote advantage in African regions with a high incidence of malaria).

Secondary effects of genes, known as pleiotropy, often take on lesser importance to the experimenter than the trait of interest, a reification of the unit concept — one genetic entity producing one primary genetic effect (Fogle, 1987). From the sickle cell example, it is customary to view the primary effect as a blood disorder and jaundice, kidney failure, bone damage, and paralysis as secondary effects. To declare a one-to-one relationship between gene and phenotype requires a narrowly defined context and denial that other contexts exist.

Benzer's introduction of the concept 'cistron' is an attempt to intercede

between the extremes of molecular genetics and the gross phenotype, with the hope of more directly linking cause and effect. The evidence from the hemoglobin gene, which emerged after Benzer's study, would seem to vindicate the one-to-one relationship between a gene and its function at the primary level of expression.

That this view is untenable is firmly established. Bonner (1965) found that tryptophan synthetase forms tryptophan from indole glyceryl phosphate and serine and also catalyzes two additional reactions. Using mutant strains that lost the ability to catalyze tryptophan, he concluded that the enzyme has two binding sites on a single polypeptide. Here, the primary product has two distinct catalytic functions.

Frezal *et al.* (1983) cite several additional cases of multifunctional proteins composed of one polypeptide including two in pyrimidine biosynthesis. One gene codes for enzymes necessary for steps one, two, and three in the pathway and the other for steps five and six. For both enzymes, a different active site catalyzes each reaction.

A different sort of primary product with multiple functions occurs during translation of secretory proteins. The first 15 to 20 amino acids allow the ribosome to bind and inject the developing polypeptide into the cavity of the endoplasmic reticulum. From here, the macromolecule is typically transported to the Golgi body and packaged into vesicles for secretion. The amino acid sequence that enables binding to the endoplasmic reticulum (termed the signal sequence) is removed and broken down. The primary polypeptide maintains two distinct though dependent functions, one related to intracellular transport and the other related to extracellular activity. The function of the signal sequence is common to many gene products that have very different extracellular functions.

No simple one-to-one correspondence exists between a primary gene product and function. The previous examples, as well as evidence that DNA loci can code for multiple polypeptide sequences by alternative splicing from a common transcript (discussed as part of Models B, C, and D) argues against a unit of function (referring to cellular activity) as *the* criterion for the material gene.

MODELS FOR THE STRUCTURAL GENE

If function creates ambiguities with the unit concept, perhaps the structural component of Watson's *et al.* (1987) definition, where transcription is a crucial ingredient, offers more hope. I apply their definition to protein coding genes by outlining four possible structural models (Figure 1). In the process of discussing each, I intend to demonstrate the enormous diversity of structure and the discontinuity between the classical and molecular gene.

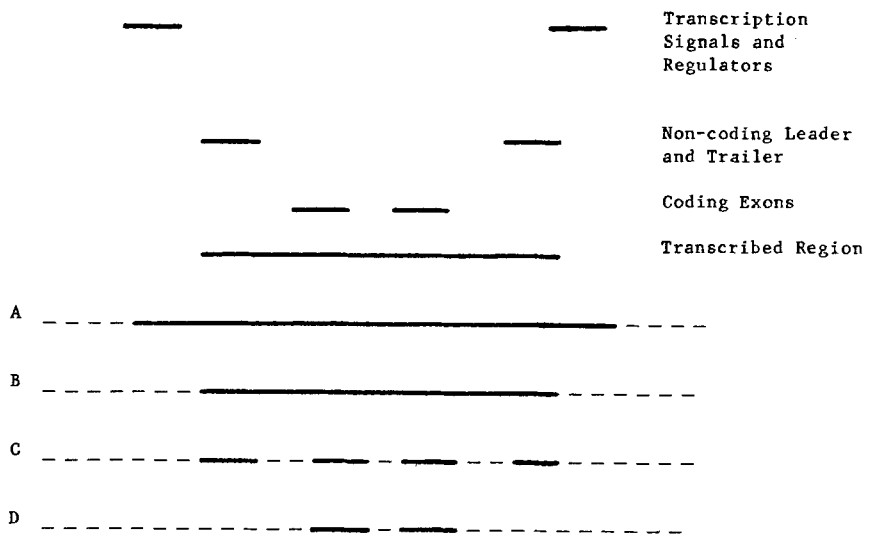

Fig. 1. Four structural models for the protein coding gene. The solid line symbolizes inclusive regions. Model A incorporates sequences of DNA in the microenvironment bordering the gene that influences expression. Model B restricts the definition to the transcribed region. Models C and D define a gene through transcribed exons that splice into a contiguous mRNA subsequent to processing, but differ by whether or not they include the non-coding leader and trailer.

Based on the generic gene discussed previously, and recognizing that both structure (a "DNA segment") and expression ("transcribed into discrete DNA chains") are essential to the definition of the gene, I characterize the structural unit as:

(A) The transcribed region plus all neighboring sequences that have some detectable influence on its expression.

(B) The transcribed region only.

(C) The set of exons derived from a primary transcript.

(D) The coding region only from the set of exons of a primary transcript.

MODEL A

Model A has the broadest boundary and includes the entire gamut of proposed *cis*-acting (neighboring) effectors on transcription including enhancers, promoters, terminators, regulators and the like. The boundary edge as to where such regions no longer influence the transcribed region is empirically difficult to assign and may overlap the borders of one or many other "genes" that depend on some of the same nucleotide sequences.

Overlapping muddies the waters but does not in and of itself refute the possible existence of units. Watson *et al.* specifically exclude "extragenic elements" from the gene definition for prokaryotes, but do not explicitly do so with respect to eukaryotes. Hence, Model A considers the implications for their inclusion.

Incorporating control elements in Model A is a feature fraught with problems. If promoters and enhancers are responsible for temporal and spatial activation of transcription then there is no good justification for excluding other regional effectors. Modulation of transcription, as influenced by *cis*-acting nucleotide sequences, need only be identified to require inclusion in the unit.

Levis *et al.* (1985) have studied the effect of relative position for the gene causing white eyes in *Drosophila* by using a P-transposable element as a vehicle to move the structural gene. P-elements randomly insert into the DNA and when they do so, insert the attached gene as well. At 20 different sites expression was normal, but at three sites, the eye had a mosaic pattern. In one such transduced strain (A4-3) the eyes were more darkly pigmented anteriorly than posteriorly. The gene for white eyes is normally on the X chromosome, but in this case was near the centromeric heterochromatin (tightly coiled repetitive DNA) of a non-sex chromosome. A second strain, A4-4, had a yellow eye with scattered flecks of red. The transduced gene was positioned near the end of a different non-sex chromosome. Were these both examples of position effects or had mutations occurred during the process of transduction? Levis *et al.* moved A4-3 to 13 new locations and A4-4 to 2 new locations. Twelve of the 13 A4-3 strains reverted as did both of the A4-4 strains. The remaining strain had a mosaic pattern different from the original strain. Subsequent transduction of that strain also produced revertants. There were also several revertants among strains that were not positioned in new locations. However, in every case there was evidence for structural changes in the gene or flanking region.

Position effects, particularly when genetic material is rearranged so as to place it near heterochromatin, are known for a variety of genetic systems (Lewis, 1950). The work by Levis *et al.* elegantly affirms this conclusion. Model A assumes that the modulating effectors are part of the gene structure. Position effects indicate that although the presumed unit appears to be independently acting for some positions in the genome, it is not independent of others and for those genes juxtaposed to heterochromatin would require inclusion of all (or at least a large section) of the structure into the unit. Expanding the gene border in this manner creates units composed of many millions of base pairs and means that a gene and a major segment of a whole chromosome can, at least in some situations, be synonymous — an outcome at odds with any mainstream definition of the gene.

The empirical evidence for position effects on the gene concept harkens

back to the debate between Richard Goldschmidt and, principally, Herman Muller, the ultimate materialist. For Goldschmidt (1938) the whole chromosome was the unit and genes as material entities did not exist. He argued forcefully against "hyperatomism" and referred to regional influences of the chromosome on the phenotype as "loci" rather than genes. One of his arguments against the atomism of his contemporaries stemmed from the work on Bar eye in *Drosophila*, a phenotype with a narrow compound eye. Using salivary gland chromosomes, Bridges (1936) found that the individuals with Bar eye had a small duplication of a specific region of the X chromosomes. Most interesting was the finding from those individuals with the duplication on both chromosomes as compared to those that contained a triplication on one chromosome and a single copy on the other. Although both kinds of individuals carry a total of four copies of the region, the eye shape of flies with unbalanced genetic copies on the two chromosomes were markedly narrower. Goldschmidt interpreted this as support for his conviction that the proper order or pattern of genetic information was critical for normal development. Muller was unphased (Raffel and Muller, 1940); he simply asserted that the units could mutually interact in some way or share some common controlling regions (*cis*-acting elements in modern jargon). Clearly it is Muller's vision that we retain today.

A second problem with Model A rests on defining the flanking borders of the microenvironment influencing gene expression. Surely this is partly an empirical problem quite apart from establishing the theoretical borders. However, analysis of promoter and enhancer regions from specific genes indicates both qualitative and quantitative effects of specific sequences as well as interactive responses among multiple *cis*-acting sequences. This diverges from the impression left by the generic gene previously described in which there seem to be discretely operating and precisely definable control elements for all genes. While a small number of specific sequences necessary and sufficient for maximal expression can be determined for some genes (see Sastry *et al.*, 1988 for an example involving the human apolipoprotein Al gene), this is not always true. Lee *et al.* (1987) investigated the human metallothionein gene and found that within the 760 base pairs upstream of the start of transcription resides a proximal TATA box, a GC box, two basal level enhancers, four sites that respond to induction by heavy metals, and a steroid hormone (glucocorticoid) induction site. *In vitro* splicing of partial sequences of this region or a mutated form onto a transcribable sequence produced a broad range of effects on the relative efficiency of transcription. The wildtype set of nine elements showed 30 times the rate of transcription compared to a control composed of a totally abnormal promoter. When the distal three elements were spliced onto the structural gene the rate was only 1.5 times base line. Using the distal six elements or eight elements (all but the TATA box) the rate was three times and twelve times, respectively. On the other hand, splicing the

proximal five elements generated 25 times baseline while a spliced region that was entirely normal except for a mutated fifth element (a basal level enhancer) resulted in only three times normal levels. In general, the more proximal elements had a stronger effect on transcription efficiency but not in any simple and additive manner.

Clear evidence that control elements can act in both a cooperative and also a cell specific manner comes from an experiment on the rat insulin gene (Edlund *et al.*, 1985). One element proximal to the start of transcription has little effect until coupled with its native upstream enhancer whereupon it activates transcription specifically in pancreatic cells. Likewise Evans and Scarpulla (1988) have found multiple transcriptional control elements in the rat somatic cytochrome *c* gene. Maximal expression requires three regions in the COS-1 cell line of kidney cells. One element is a CAAT box proximal to the start site of transcription. The other two sites show strong sequence similarity and compete for binding factors (proteins) that interact on these elements to influence transcription. One of the elements is located distal to the CAAT box and the other is positioned within the first intron.

Several points follow from these examples. First, there is considerable diversity in distribution, size, organization, and complexity of the promoter/enhancer component including regions that can lie both inside and outside of the transcriptional border. Second, cooperativity by three elements to maximize transcription of the cytochrome *c* and the nonadditive influence of nine elements on transcriptional efficiency of the metallothionein gene precludes any logical way to specify the upstream border by including some elements and not others. Finally, the metallothionein gene demonstrates greatly reduced effects on transcription efficiency for the distal as compared to the proximal elements. The most distal four elements barely produce an empirically detectable effect. It is as if the influence of the distal elements fades into the genetic horizon as one searches farther and farther upstream of the gene. Indeed there may be still more elements with very minor, as yet undetected, effect located beyond the most distal of the nine elements. For this gene, the upstream border has no clearcut point of termination.

The problem of defining how subtle the modulating effect must be before excluding neighboring sequences of DNA as part of the structure coupled with the regulatory effect of huge sections of DNA from experiments on position effects, means that a completely inclusive model is vacuous.

MODEL B

The structural boundaries for Model B are defined by the process of tran-

scription. If genetic information is to be utilized by the cell, transcription represents a mechanism for intracellular communication. That transcription has clean borders makes this model even more appealing. Is this the unit we seek?

In Model B, the transcript derives from a continuous section of the DNA and suggests a relationship between that proposed unit and the information it bears to make a polypeptide during translation. If the resultant polypeptide has multiple functions or is cleaved into fragments with different functions, the entire translational message is still collectively encoded by one transcript.

The rat tropinin T gene (Breitbart *et al.*, 1985) demonstrates that this linear correspondence between a gene and a unit of information through transcription does not always hold. During mRNA processing, intron removal produces two different coding messages that differ by thirteen of the 259 amino acids in the polypeptide. The transcript contains two choices for the "fourth" exon regulated by developmental cues and cell type. The inserted exon changes one portion of the coding region and hence alters the polypeptide also. All other exons are the same. Even though the cellular environment plays a role in regulating the system, the splicing sites are specified in the DNA. Because two different message units are constructed *prior* to product formation, and therefore prior to function, the DNA sequence transcribed acts as two units of structure and information.

In many cases (but not always) multiple gene products formed in this manner share similar properties and constitute a protein family. As another example, this same mechanism of alternative processing is utilized by the rat alpha-tropomyosin transcript to produce a minimum of six forms of the polypeptide (Wieczorek *et al.*, 1981)! Approximately 30% of the genes expressed in the nervous system invoke alternative processing (Sutcliffe and Milner, 1988). Surely this is not a rare phenomenon and has great utility for serving the needs of living systems, but does not support a gene concept centered on the transcript as a unit.

Model B is inappropriate because transcripts may bear more than one unit of information through combinatorial generation of message sequences. If so, then perhaps units are smaller in size.

MODEL C

Model C designates the exon as the structural unit bearing information. The alternative exons of the troponin gene represent alternative units of information spliced into the chain of exons comprising the processed mRNA. In this way, the unit concept can be salvaged by redefining the gene as an exon or a set of exons that share a common transcript.

Henikoff *et al.* (1987) sequenced the DNA from the Gart locus of *Drosophila* and identified two primary proteins encoded by seven exons. One product is a polyprotein that is cleaved after translation to form GAR synthetase, AIR synthetase, and GAR transformylase. The other product forms GAR synthetase only by transcribing a shorter transcript that uses the first four exons. Although this would appear to be a variant form of alternative splicing, the structural border of the fourth exon is different from that of the long transcript. The 3′ end of the fourth exon from the short transcript includes a stretch of non-coding nucleotides that is spliced out by the longer transcript. The coding regions of the short and long transcript are the same through all four exons differing only by the presence of a trailer on the short sequence. The fourth unit contains a different structure, yet codes for an identical polypeptide. Here the structural units responsible for a given product lack precise physical correspondence. That is, more than one splicing border is possible generating two different sized units for a common section of DNA. Which unit is the gene?

MODEL D

We can, however, rescue Model C by a slight modification. Model D presumes that our concern should focus strictly on the coding region. The unit problem vanishes if the fourth exons are made equivalent by considering the coding region only. Therefore, the difference in the length of the 3′ non-coding ends becomes irrelevant.

Unfortunately, alternative splicing can affect the size and coding region of exons as well. The *Drosophila* Eip 28/29 gene produces two primary polypeptide products that are induced by the hormone ecdysone (Schulz *et al.*, 1986). The gene, composed of four exons, encodes two processed mRNAs that differ in length by 12 internal nucleotides resulting in translated polypeptides that differ by 4 amino acids. Alternative splicing sites at the 3′ juncture of exon two and the 5′ end of intron two account for the two forms of mRNA. The 12 nucleotides affected are thus part of an intron with respect to the short mRNA and part of a coding exon for the long mRNA.

The previous example is not unique. The gene for human growth hormone shows alternative splicing between the end of the second intron and start of the third exon. The variant mRNAs differ in length by 45 nucleotides and the resultant polypeptides differ internally by 15 amino acids (see Hampson and Rottman, 1987).

The distinction between exon and intron further blurs in the bovine growth hormone gene (Hampson and Rottman, 1987). Five exons and four introns are present in one form of processing that synthesizes a 217

amino acid precursor polypeptide. In the other, alternative splicing fails to remove the fourth intron and the retained sequence becomes part of the coding message. Even more interesting is that the reading frame entering the fifth exon is shifted! Translation occurs by reading consecutive three nucleotide codons. Inserted nucleotides, if not in multiples of three, will generate an entirely different amino acid sequence from that point on. Further, the ultimate length of the message can change because the original stop codon is not read and another is needed in the new reading frame. The final length of the polypeptide can be longer or shorter than the original polypeptide dependent on the location of a new stop codon. Hampson and Rottman did not purify the protein, but did find evidence that the variant mRNA was bound to the translation machinery of the cell, a strong indication that a polypeptide is made. Based on the DNA sequence, the two polypeptides will have the first 125 amino acids in common followed by entirely different sequences that are either 108 amino acids (derived from the mRNA that retains the intron) or 66 amino acids long. The two polypeptides utilize different lengths of the fifth exon.

The bovine growth hormone gene, then, has a nucleotide sequence that is an intron for one mRNA and is part of an exon for another. Exon five, which they have in common, is translated entirely differently and the border between the coding and non-coding trailer at the end of the message differs as well. This is a gene which can be perceived as having five exons and four introns or four exons and three introns (depending on which processing scheme is considered) and two different lengths if only the coding regions are regarded as defining the structure.

These examples demonstrate that there is nothing sacred about the structure of an exon or intron nor are the 5′ and 3′ borders of the reading frame fixed. Shuffling of exons to assemble multiple translatable messages is one mechanism of alternative splicing but not the only one. Therefore, Model D is also inadequate.

THE PROBLEM OF OVERLAP

Having failed to define the structural unit with Models A through D, I investigate the possibility that the unit, though perhaps complexly organized, still represents a discrete region of DNA that is distinct from other such complexly organized units.

The various themes of alternative splicing document overlap of coded information; the alternative protein products have at least some common coding sequences in the DNA. In the previous examples, there is common transcriptional control and often a common primary transcript. This need not be the only form of overlap.

The two strands of the double helix are oriented in opposite directions

(antiparallel) and, for a given transcript, the message is contained on only one of the two complementary strands. The genetic message on the copied strand is always read in a 3′ to 5′ direction forming a complementary copy of mRNA that is antiparallel (formed 5′ to 3′). Different genes in the same DNA can use different strands.

Spencer *et al.* (1986) found that the end of the *Drosophila* gene for dopa decarboxylase overlaps another gene of unknown function. The two genes are on opposite strands and overlap tail-to-tail by 88 nucleotides at the 3′ end of the transcript. Analysis of the temporal and spatial presence of the transcripts revealed that high levels of both transcripts are never concordant. Developmentally, the dopa decarboxylase transcript is found at maximal levels in 18 hour embryos, during larval molts, during pupation, and adult emergence. The other transcript is maximally present in 1—4 hour embryos and the adult. During the temporal overlap in expression there are tissue specific differences — high levels of the dopa decarboxylase transcript in epidermis and high levels of the other transcript in testes. The authors speculate that the alternative expression may be mutually regulatory. The fact that a supposed genetic unit, either the dopa decarboxylase gene or its shared antiparallel mate, is temporally or spatially expressed comes as no surprise. Regulation at the level of transcription is regarded as an important mechanism of developmental regulation. Rather, the intriguing feature of this system is that the antiparallel sequences may act in a dual role as a transcriptionally based unit for a polypeptide and as a *cis*-acting regulator. Is this one large unit that produces two transcripts or two units each with a structural and regulatory role?

Gene overlap is a common form of genetic organization in viruses. Examples in eukaryotes have been documented only recently, but may not be rare. The gene for gonadotropin-releasing hormone (GnRH) has four exons and is similarly organized in both rats and humans. In humans, the gene undergoes alternative splicing in a tissue specific manner. Unlike the hypothalamic form, the placental form fails to remove the first intron and instead becomes part of the 5′ non-translated leader sequence. Remarkably, a second gene (termed SH) actively transcribes nearly the entire complementary strand (Adelman *et al.*, 1987). Further, these investigators found that the SH gene (whose function is unknown) produces three different forms of processed mRNA. Each form contained a common trailer sequence spliced with three different exons. The nature of the evidence did not enable determination of whether the mRNAs are translated nor could Adelman's group distinguish whether they originate from one transcript that is alternatively processed or three different transcripts each with its own promoter. In either case, the GnRH-SH locus is richly complex on *both* strands of the same DNA segment.

While dopa decarboxylase genes are slightly overlapping, those of

GnRH-SH are entirely overlapping. If the genetic unit of inheritance consists of a nucleotide sequence, then in these cases a double stranded DNA consists of two antiparallel units. Most amazingly, the GnRH-SH locus displays both forms of overlap-alternative splicing of GnRH (and perhaps SH also) and expression of both DNA strands. Recall that the definition by Watson *et al.* (1987) indicates that the gene should "refer to those DNA segments that become transcribed . . ." Since the GnRH-SH locus is a "segment of DNA" and is "transcribed", by that definition it contains just one gene, irrespective of the fact that the two strands produce very different transcripts.

The dunce locus of *Drosophila melanogaster* is still more unusual. Partial characterization of the molecular structure reveals 13 exons and strong evidence that it codes for the enzyme cAMP phosphodiesterase (Chen *et al.*, 1987). The gene extends over 90 kb, the second intron alone occupying 79 kb. Within the large intron are at least two nested genes. Pig-1 (preintermolt gene-1), expressed primarily in the larval salivary glands, has a transcript of less than 1 kb encoded on the opposite strand. Located just 840 base pairs away lies a similarly sized Sgs-4 gene that expresses larval glue protein in the salivary gland. Neither Pig-1 nor Sgs-4 bear introns. Further, there is indirect evidence for at least two other genes within the large intron. Multiple nesting of genes is not the only unusual feature of this system. In contrast with previous examples, Sge-4 is encoded on the same strand as the gene within which it is nested. Therefore, it is not possible to resolve the unit problem by simply distinguishing strands for overlapped genes.

TRANS-SPLICING

All mRNAs of the single celled protozoan *Trypanosoma brucei* are composed of two distinct components — an untranslated leader sequence of 35 nucleotides and the message sequence. The two components are encoded in different loci of the genome and spliced together, hence the term *trans*-splicing (to distinguish it from *cis*-splicing, the removal of introns). The leader, part of a repetitive DNA sequence of about 200 copies, is itself an exon of a primary transcript three times as large that, after processing, attaches to the 5′ end of a protein coding message unit. Functions for the leader sequence are unknown.

Both *cis*- and *trans*-splicing have consensus sequences at the junction of the splicing site yet introns are not found in trypanosome protein coding genes. This has led to speculation that trypanosomes, organisms that may have diverged from the eukaryotic line prior to the divergence of animals, plants, and fungi, displays an ancient form of splicing.

Speculation aside, the mechanism of *trans*-spliced RNA engenders an

entirely new wrinkle for the unit concept. Is the *trans*-spliced RNA the product of two units or is it a coding unit independent of its split origin in the DNA? The contiguity of a primary transcript located at one specific point (one locus) was a common feature for Models A through D. At least with conventional *cis*-splicing the structural integrity of the transcript regionalized the unit even if its borders were ill-defined. Now even that is lost.

Evidence that *trans*-splicing is not unique to trypanosomes comes from analysis of the four actin genes of the nematode *Caenorhabditis elegans* (Krause and Hirsh, 1987). Three of the four are attached to a 22 nucleotide leader coded elsewhere in the genome. Additional evidence suggests that other mRNAs of this species may *trans*-splice in a similar manner.

In both examples above, an untranslated leader is *trans*-spliced onto a translatable sequence. The protein coding sequence remains intact. In one known example, even this continuity is violated. The S12 ribosomal protein gene encoded by chloroplast DNA (Koller *et al.*, 1987) appears to be organized as three separately transcribed exons, two on one strand and the third on the opposite strand. *Trans*-splicing accounts for a contiguous message sequence for translation.

THE VANISHED UNIT

The present investigation of the unit concept finds no feature in common among described protein coding gene of eukaryotes that unites them as material units of structure or function. Falk (1986) argues that the current concept of the gene is pragmatically flexible, adjusted to the needs of the experimenter and that there resides an entrenched conviction for the existence of a genetic entity with more than one meaning required to justify it. Thus, my analysis supports Falk's contention that the contemporary view of the material gene is an instrumental construct, an abstraction. This abstraction, analogous to how 'factor' based models built the genetic theory of quantitative traits, successfully encompasses a patchwork of molecular anomalies. However, retreating to abstraction confounds meaning and encumbers specification of gene properties.

The generic gene masks a wealth of structural diversity. We have seen that structural regions can be shared, overlapping on opposite or like strands, nested, and even physically split. Introns may or may not be present, the same nucleotide sequence can be both coding and non-coding, alternative splicing can produce multiple products of translation, proteins from one transcript can have multiple functions as either an intact structure or be cleaved into separate polypeptides, frame shifting can create a second translatable product from one nucleotide sequence, information crucial to gene expression can reside externally or internally

with respect to a transcribed message . . . The list seems endless. To place all of these under the rubric of being a "unit of inheritance," in some material sense, strains credulity. Having initially defined the gene in this manner we are left with the hopeless task of explaining exactly what that means in anything other than pre-molecular terms. The gene may *act* like a unit from a Mendelian point of view, but it does not *look* like a unit at the molecular level. Genomic organization is vastly more complex than Mendelism ever revealed. The term unit, appropriate for the pre-molecular era, can not be reconciled with either multiple use of the same DNA sequence or dispersed structure. Long entrenched in genetic history, the unit concept derived from Mendelism has not disappeared from molecular genetics in the face of enormous difficulties with its current meaning.

In the succeeding section I attempt to outline an alternative framework for gene structure mindful that diversity and overlap are normal phenomena requiring integration into any such system.

GENES AS SETS OF DOMAINS

A gene looks less like a Mendelian entity — particulate and tandemly linked — than a collection of component entities that together define its structure and influence the phenotype. Many nucleotide sequences, some very short (a codon consists of three bases) and some very long (a single intron can be thousands of bases), are parts of the structural hierarchy essential for expression. My intent is to define a framework that disengages the unit concept of Mendelism from molecular genetics.

I propose that sets of domains more closely describe a gene, where domain is intended to mean a sequence of nucleotides identified by structural properties and/or activities which distinguish it from other such sequences. A given nucleotide sequence may be part of more than one domain in a manner that allows nesting and overlapping. Promoters, for example, are domains of activity, namely the start site for transcription, some of which contain a TATA box and a CAAT box as nested domains within them. Exons have the activity of splicing in common and hence are domains for that process. Alternative splicing can change the set of exons or the boundary of some exons while retaining the attendant activity (splicing). Processed mRNA, which results from the removal of introns, is coded by a DNA domain containing exons as a subset of domains. Set structure uncouples the need to find a single unit for a region of genetic information as is customary when thinking in Mendelian terms.

A material gene, then, lacks coherent meaning at the molecular level until a partial set of transcriptional domains is specified. Specification of all domains influencing expression may not be necessary for nominative or heuristic purposes and may be impossible to establish empirically. What is

needed is sufficient itemization to prescribe a set that has communal agreement. No one set of domains at the primary level of organization could descriptively fit all genes. Not all genes have introns or upstream enhancers, for example.

When molecular geneticists use terms such as enhancer, promoter, open reading frame, intron, exon, and the like they have in mind common properties of structure or activity among genes. That is, they are dissecting the gene into domains. There is loose application of terminology, however, caused by communication difficulties that arise during rapid expansion of empirical evidence. Terms like regulatory element, cis-acting element, regulatory sequence, and 5′ regulator share relatedness in meaning, yet lack of precise definition. Applying set structure does not resolve this problem. It does, however, promote clarification by demanding that the domain be specified by structure and/or activity.

Domains are also in current usage when classifying genes. For example, the 5′ flanking regions of housekeeping genes have GC rich sequences as compared to a TATA box and CAAT box of regulated genes. This domain difference promotes a descriptive encapsulation of thousands of genes with a property of major significance to the organism. In a similar fashion, subdivision of the genome could be established along structural, functional, or geneological lines. The latter is represented in current terminology as gene families, sequences of nucleotides that presumably duplicated and diverged during the course of evolution. Within that family are sequences that transcribe (genes) and those that have lost the capacity to do so (pseudogenes). Geneological domain sets are a natural grouping for genetic phenomena that have a recognizable historical context and alleviates the necessity for awkward terms such as pseudogenes. These are sequences that were identified by structural analysis of DNA not by activity as determined for a Mendelian unit. Consequently they do not fit into the traditional accounting of gene units. They are like genes but something different. Layering molecular genetics over classical genetics ignores much of the molecular complexity because only two conventional categories of phenomena are permitted; either something is a gene or it is something else (repetitive DNA, regulator sequences, pseudogene, etc.).

A domain model accepts that information resides within units smaller or larger than a conventionally defined gene and that sets of domains are needed to establish correspondence with Mendelian entities. A region that codes for the ability to alternatively process mRNA, whether temporally or in a tissue specific manner, represents a multiple Domain Set for Active Transcription (DSAT). Hence, the number of DSATs in the genome far exceeds the number of genes as defined by units and more accurately quantifies the number of primary polypeptide products.

While many domains comprise a gene, not all domains reside in the

microenvironment bounding any one gene. Some domains may have no empirically established relationship to any one gene (many repetitive DNA sequences, for example). Likewise, *trans*-splicing domains may be common to functional transcripts of many genes and physically coded elsewhere in the DNA.

DSATs, like the material gene, have elements of structure and expression which distinguish them from other regions of the DNA. Is a DSAT still a unit, albeit one which does not require localization and accepts organizational complexity? A 'unit' in genetic terminology is rooted in the Mendelian tradition of loci, factors, and particulate inheritance. To redefine a domain set as a 'unit' is to create a second referent for the term. And it is precisely such fluidity of meaning that has awkwardly fused Mendelian and molecular concepts. Still, one might think of domain sets as having properties united through structure and function. In this broader sense, one which has no previous commitment to historical context in genetics, domain sets have 'unit' properties. My introduction of the domain concept does not deny this. Instead, it divorces Mendelian imagery from the molecular gene and promotes an ontology about genetic information missing from classical units. Two examples of the ontological virtues for a domain model will be given.

Domain sets more closely mirror the structure of genetic information and call into question the validity of models that manipulate genetic parameters as if all genes had common properties of size, complexity, and/or organization. Models in population genetics and quantitative genetics commonly follow this tack and, in fact, owe their conceptual framework to the Mendelian 'factor' from the early 1900s. The extent to which factor (unit) based models generalize in the face of molecular diversity of the gene is unknown.

Second, what does it mean to say that an organism has x number of genes? Conventional accounting equates genomic information with gene number. Not all domains of information are encapsuled within genes and an enumeration of classical units underestimates novel information content caused by alternative processing. If molecular information is packaged in a dispersed and overlapping manner, then knowing the number of 'genes' is a less useful parameter than knowing some measure of the density, type, and number of domains present in a given organism.

Clearly there exist important levels of organization left unaccounted between the nucleotide and the hypothetical construct of the gene as a unit, levels of organization that need discarding of Mendelian baggage. The importation of a classical unit concept into molecular genetics, more than just being ambiguous, so oversimplifies genetic architecture that almost any statement about the structure of the gene is doomed to a litany of exceptions.

REFERENCES

Adelman, J. P., C. T. Bond, J. Douglass, and E. Herbert: 1987, 'Two Mammalian Genes Transcribed from Opposite Strands of the Same DNA Locus', *Science* **235**, 1514–1518.

Benzer, S.: 1957, 'The Elementary Units of Heredity', in *The Chemical Basis of Heredity*, Johns Hopkins Press, Baltimore, pp. 70–93.

Bonner, D. M.: 1965, 'Gene-enzyme Relationships', in S. J. Geerts (ed.), *Gene Today*, Vol. 2, Pergamon Press, Oxford, pp. 141–149.

Breathnach, R. and P. Chambon: 1981, 'Organization and Expression of Eucaryotic Split Genes Coding for Proteins', *Annual Review of Biochemistry* **50**, 349–383.

Bridges, C.: 1936, 'The Bar "Gene" a Duplication', *Science* **83**, 210–211.

Castle, W. E.: 1919, 'Piebald Rats and the Theory of Genes', *Proceedings of the National Academy of Science* **5**, 126–130.

Chen, C., T. Malone, S. K. Beckendorf, and R. L. Davis: 1987, 'At Least Two Genes Reside within a Large Intron of the *Dunce* Gene of *Drosophila*', *Nature* **329**, 721–724.

Edlund, T., M. D. Walker, P. J. Barr, and W. J. Rutter: 1985, 'Cell-Specific Expression of the Rat Insulin Gene: Evidence for Role of Two Distinct 5′ Flanking Elements', *Science* **230**, 912–916.

Evans, M. J. and R. C. Scarpulla: 1988, 'Both Upstream and Intron Sequence Elements Are Required for Elevated Expression of the Rat Somatic Cytochrome *c* Gene in COS-1 Cells', *Molecular and Cellular Biology* **8**, 35–41.

Falk, R.: 1986, 'What Is a Gene?', *Studies in History and Philosophy of Science* **17**, 133–173.

Fogle, T.: 1987, 'The Phenotypic Deception: Influences of Classical Genetics on Genetic Paradigms', *Perspectives in Biology and Medicine* **31**, 65–80.

Frezal, J., A. Munnich, and G. Mitchell: 1983, 'One Gene, Several Messages. From Multifunctional Proteins to Endogenous Opiates', *Human Genetics* **64**, 311–314.

Goldschmidt, R.: 1938, 'The Theory of the Gene', *The Scientific Monthly* **46**, 268–273.

Hampson, R. K. and F. M. Rottman: 1987, 'Alternative Processing of Bovine Growth Hormone mRNA: Nonsplicing of the Final Intron Predicts a High Molecular Weight Variant of Bovine Growth Hormone', *Proceedings of the National Academy of Science* **84**, 2673–2677.

Henikoff, S. and M. K. Eghtedarzadeh: 1987, 'Conserved Arrangements Nested Genes at the *Drosophila Gart* Locus', *Genetics* **117**, 711–725.

Hull, D. L.: 1974, *The Philosophy of Biological Science*, Prentice-Hall, Englewood Cliffs, N.J.

Jacob, F. and J. Monod: 1961, 'Genetic Regulatory Mechanisms in the Synthesis of Proteins', *Journal of Molecular Biology* **3**, 318–356.

Johanssen, W.: 1909, *Elemente der exakten Eblichkeitslehre*, Fisher, Jena.

Kitcher, P.: 1982, 'Genes', *British Journal for the Philosophy of Science* **33**, 337–359.

Koller, B., H. Fromm, E. Galun, and M. Edelman: 1987, 'Evidence for *in vivo Trans* Splicing of Pre-mRNAs in Tobacco Chloroplasts', *Cell* **48**, 111–119.

Krause, M. and D. Hirsh: 1987, 'A *Trans*-Spliced Leader Sequence on Actin mRNA in C. Elegans', *Cell* **49**, 753–761.

Lederman, M.: 1987, '"Genes" Amplified', *British Journal for the Philosophy of Science* **38**, 561–566.

Lee, W., A. Haslinger, M. Karin, and R. Tijan: 1987, 'Activation of Transcription by Two Factors That Bind Promoter and Enhancer Sequences of the Human Metallothionein Gene and SV40', *Nature* **325**, 368–372.

Levis, R., T. Hazelrigg, and G. M. Rubin: 1985, 'Effects of Genomic Position on the Expression of Transduced Copies of *White* Gene of *Drosophila*', *Science* **229**, 558–561.

Lewin, B.: 1987, *Genes III*, Wiley, New York.

Lewis, E. B.: 1950, 'The Phenomenon of Position Effect', *Advances in Genetics* **3**, 75–115.

Maull, N.: 1977, 'Unifying Science without Reduction', *Studies in History and Philosophy of Science* **8**, 143–162.

Melton, D. W.: 1987, 'Strategies and Mechanisms for the Control of Transcriptional Initiation of Mammalian Protein-Coding Genes', *Journal of Cell Science* **88**, 267–270.

Nagel, E.: 1961, *The Structure of Science*, Harcourt, Brace and World, New York.

Raffel, D. and H. J. Muller: 1940, 'Position Effect and Gene Divisibility Considered in Connection with Three Strikingly Similar Scute Mutations', *Genetics* **25**, 541–583.

Rosenberg, A.: 1985, *The Structure of Biological Science*, Cambridge University Press, Cambridge.

Sastry, K. N., U. Seedorf, and S. K. Karathanasis: 1988, 'Different *cis*-Acting DNA Elements Control Expression of the Human Apolipoprotein AI Gene in Different Cell Types', *Molecular and Cellular Biology* **8**, 605–614.

Schaffner, K. F.: 1967, 'Approaches to Reduction', *Philosophy of Science* **34**, 137–147.

Schulz, R. A., L. Cherbas, and P. Cherbas: 1986, 'Alternative Splicing Generates Two Distinct *Eip*28/29 Gene Transcripts in *Drosophila* Kc Cells', *Proceedings of the National Academy of Science* **83**, 9428–9432.

Spencer, C. A., R. D. Gietz, and R. B. Hodgetts: 1986, 'Overlapping Transcription Units in the Dopa Decarboxylase Region of *Drosophila*', *Nature* **322**, 279–281.

Sutcliffe, J. and R. Milner: 1988, 'Alternative mRNA Splicing: The *Shaker* Gene', *Trends in Genetics* **4**, 297–299.

Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner: 1987, *Molecular Biology of the gene*, 4th ed., Vol. 1, Benjamin/Cummings, Menlo Park, Ca.

Wieczorek, D. F., C. W. J. Smith, and B. Nadal-Ginard: 1988, 'The Rat a-Tropomyosin Gene Generates a Minimum of Six Different mRNAs Coding for Striated, Smooth, and Nonmuscle Isoforms by Alternative Splicing', *Molecular and Cellular Biology* **8**, 679–694.

Wimsatt, W. C.: 1976, 'Reductive Explanation: A Functional Account', in R. S. Cohen (ed.), *PSA 1974*, Reidel, Dordrecht, Holland, pp. 671–710.