Genome System Architecture and Natural Genetic Engineering in Evolution

JAMES A. SHAPIRO[a]

Department of Biochemistry and Molecular Biology, University of Chicago, 920 E. 58th Street, Chicago, Illinois 60637, USA

ABSTRACT: Molecular genetics teaches three lessons relevant to the nature of genetic change during evolution: (1) Genomes are organized as hierarchies of composite systems (multidomain protein-coding sequences; functional loci made up of regulatory, coding, processing, and intervening sequences; and multilocus regulons and replicons) interconnected and organized into specific "system architectures" by repetitive DNA elements. (2) Genetic change often occurs via natural genetic engineering systems (cellular biochemical functions, such as recombination complexes, topoisomerases, and mobile elements, capable of altering DNA sequence information and joining together different genomic components). (3) The activity of natural genetic systems is regulated by cellular control circuits with respect to the timing, activity levels, and specificities of DNA rearrangements (e.g., adaptive mutation, Ty element mobility, and P factor insertions). These three lessons provide plausible molecular explanations for the episodic, multiple, nonrandom DNA rearrangements needed to account for the evolution of novel genomic system architectures and complex multilocus adaptations. This molecular genetic perspective places evolutionary change in the biologically responsive context of cellular biochemistry.

A "QUANTUM REVOLUTION" IN BIOLOGY

Most of the basic concepts in conventional evolutionary theory predate 1953 when virtually nothing was known about DNA. In the first half of the 20th century, mathematical treatments of the evolutionary process were elaborated using terms such as genes, alleles, dominance, penetrance, mutation, epistasis, fitness, and selection. Since 1953, tremendous progress has been made in identifying the molecular components of genomes, defining their evolutionary relationships, and tracing the dynamics of how these components act and interact during cellular proliferation and multicellular development. Although molecular geneticists still use much of the old language (as when they want to "clone the gene" for some function), they actually operate in a distinct conceptual universe (e.g., by cloning a cDNA into an expression vector). The conceptual universe of molecular genetics is as different from classical genetics and evolutionary theory as quantum physics is from classical mechanics. Thus, if we try to formulate basic genetic and evolutionary concepts using molecular knowledge of cell function plus the information in sequence databases, we are obliged to come up with a radically different picture of genome organization and reorganization. The objective of this chapter is to explain why and how evolution must be viewed afresh at the end of the 20th century.

[a]Address for correspondence: James A. Shapiro, Department of Biochemistry and Molecular Biology, University of Chicago, 920 E. 58th Street, Chicago, IL 60637. Phone, 773 702-1625; fax, 773 702-0439; e-mail, jsha@midway.uchicago.edu

GENOME ORGANIZATION: A HIERARCHY OF NESTED SYSTEMS

Four decades of dissecting genome function at the molecular level have brought many insights that were not anticipated in 1953. Two of the most far reaching are: 1. Many different genetic codes exist in addition to the triplet code for amino acids (ref. 80; Trifonov, this volume). These codes affect many diverse aspects of genome function, such as replication, transcription, recombination, DNA packaging and chromatin organization, imprinting, RNA and protein processing, and chromosome localization, pairing, and movement. 2. There do *not* exist fundamental genomic units larger than the individual codons in the various functional codes.

In other words, what used to be considered basic genetic elements actually function as systems composed of multiple codons. For example, we can take the *Escherichia coli lac* operon as the paradigm of a gene and see that it is actually a mosaic of transcription factor binding sites and protein coding sequences (FIG. 1).[67] Even the operon's individual open reading frames (ORFs) are systems rather than elementary units, because they can be divided into distinct domains capable of encoding genetically separable functions (e.g., α and ω domains of *lacZ*, DNA-binding and inducer-binding domains of *lacI*).[62]

The *lac* operon also illustrates the hierarchical nature of genomic systems. At a lower level, the σ70 promoter can be broken down into –10, –35, and spacer regions,[36] and the

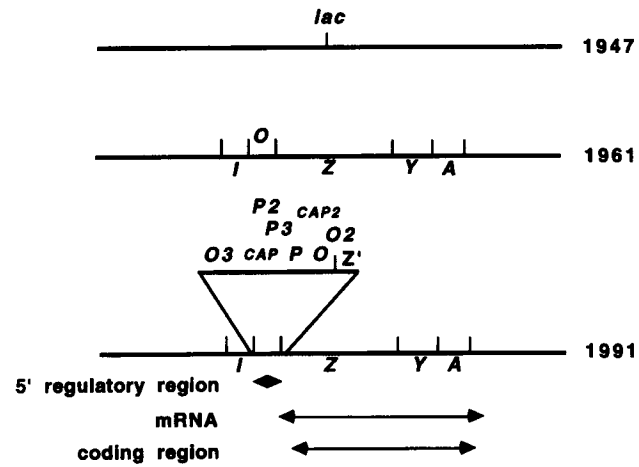

FIGURE 1. Molecular genetic dissection of the *lac* operon. Progress in our understanding of the *E. coli* determinants for lactose utilization over the last half century are illustrated, starting with *lac* as a single point marker on the genetic map (1947) and continuing through the operon theory (1961), which introduced *lacO*, the operator, as a *cis*-acting site in addition to the "structural genes" for repressor (*I*), beta-galactosidase (*Z*), lactose permease (*Y*), and galactoside transacetylase (*A*), to our more recent understanding of the 5′ regulatory region as containing three operators, one of which (*O2*) overlaps with the *lacZ* coding sequence, three promoters, and two *CAP* binding sites for the cAMP-Crp activator complex (see ref. 67 for further details).

DNA binding domain of *lacI* can be broken down into its helix-loop-helix subdomains.[62] At higher levels, the promoter integrates *lac* into the σ70 transcriptional regimen of *E. coli* functions active during exponential growth conditions, whereas the *CAP* transcription factor binding site integrates *lac* into the regulon of catabolite-repressed functions that are not expressed when glucose or other optimal carbohydrate growth substrates are available.[68]

The integrative aspect of genomic systems has the additional consequence that it is generally not possible to make unitary genotype-phenotype assignments. To transcribe *lac* operon protein coding sequences and be lactose-positive in phenotype, an *E. coli* cell must also express proteins encoded by the *cya* (adenylate cyclase) and *crp* (cAMP receptor protein) loci, each of which has its own composite structure.[68] The complexity of genomic integration into multilocus systems becomes even greater when we consider the biogenesis and functioning of cellular organelles and the processes of multicellular development, where suites of loci throughout the genome respond to intra- and intercellular signals.[1] In many cases, it is really impossible to assign a specific organismal phenotype to a particular locus, because its gene product(s) can participate in the execution of multiple cellular or developmental programs.[26]

GENOME SYSTEM ARCHITECTURE: ORGANIZED BY REPETITIVE ELEMENTS

As illustrated by the *lac* operon example, much attention has been devoted to the study of transcription initiation control. We now understand this process to be largely the result of transcription factor complexes interacting with RNA polymerase to inhibit or stimulate transcription from a given promoter. Proper timing and level of transcription are controlled by interaction of one or more transcription factors with external or internal signals, sometimes directly (as lac repressor with IPTG) or indirectly by means of a signal transduction cascade (as in yeast mating pheromone response).[4] Multiple loci can be connected in coordinately regulated networks because, the genomic binding sites recognized by the transcription factor/RNA polymerase/initiation factor complexes are dispersed repetitive DNA elements, such as the *CAP* site in *lac*, and are present near different protein coding sequences. Being a codon rather than a complete functional unit, the regulatory effect of any particular binding site and its cognate transcription factor can be positive or negative depending on the context of other transcriptional codons. This functional flexibility of transcriptional codons was first realized in studies of the phage λ control region where cooperative repressor binding to the O_1 and O_2 operators acts negatively on one promoter, P_R, but positively on another, PRM.[66]

Gene expression is regulated by other features of genome organization in addition to promoter and enhancer elements. In some complex loci, such as the *Drosophila bithorax* region[56] and the mammalian globin determinants,[40] the linear arrangement of transcriptional determinants and protein coding sequences along the chromosome plays a role in determining their time of expression in development. The proximity of genetic loci to blocks of simple sequence repetitive DNA (heterochromatin) is likewise an important factor in gene expression.[45,48,82] Some loci are expressed specifically in heterochromatin, whereas the general rule is that expression of most loci is inhibited by the proximity to heterochromatin. This inhibition is exemplified by the phenomenon known as "position effect variegation" (PEV) in *Drosophila*. Even relatively small repetitive arrays, such as the trip-

let repeats found in coding sequences or introns of some human disease loci, can amplify to block transcription in their immediate neighborhood.[30,50] Both PEV and triplet repeat effects on gene expression are believed to result from the binding of repeat-specific proteins to create an altered chromatin structure that is unfavorable for access by RNA polymerase complexes.

From a theoretical perspective, the functional significance of nonprotein coding repetitive DNA elements in transcriptional regulation should not be surprising. Biologically appropriate gene expression is a process of reading those texts (protein coding sequences) suitable for one set of conditions while ignoring the other texts stored in the genome. To accomplish this task, the transcriptional apparatus requires a hierarchical addressing system, which is provided by the right combination of repetitive elements determining both chromatin organization and transcription factor binding. Controlling transcription initiation is only one aspect of the regulation of gene expression. Transcription termination, mRNA processing, and protein processing all play equally important roles in ensuring that each cell contains the right constellation of proteins, and each of these processes has its cognate codons in the genome.[1] These codons are also dispersed repetitive elements because they must be recognized the same way at different loci.

As an information storage organelle, the genome must undergo duplication and accurate transmission to progeny cells. These essential processes also depend on dispersed repetitive elements that do not encode proteins, such as the codons that make up telomeres, centromeres, and replication origins.[9] There is rarely a single way to accomplish a particular functional goal in genome maintenance, but the alternatives usually involve repetitive DNA elements. For example, most eukaryotes from yeast and ciliated protozoa through mammals use telomerase-generated repeats for maintaining the ends of their chromosomes,[8] but *Drosophila* and other *Diptera* use recombinational mechanisms, such as retroposons, for this purpose.[7,55,63] In meiotic divisions, clustered repetitive satellite DNA arrays serve to ensure proper chromosome disjunction. Recent work on *Drosophila* has shown that rDNA spacer repeats are essential in sex chromosome pairing necessary for accurate male meiosis[61] and that heterochromatin is necessary for pairing in achiasmate (nonexchange) disjunction in females.[20,49]

Based on their fundamental roles in genome transmission and in determining patterns of gene expression, it can be proposed that repetitive DNA elements set the "system architecture" of each species.[74] The term "system architecture" is used to draw the analogy with computers, where programs with the same functionality (e.g., Microsoft Word©) are encoded differently according to the requirements of the underlying hardware and operating system (e.g., MacOS© or Windows©). From the system architecture perspective, what makes each species unique is not the nature of its proteins (a Windows desktop resembles a Macintosh desktop) but rather a distinct "specific" organization of the repetitive DNA elements that must be recognized by nuclear replication, segregation, and transcription functions. In other words, resetting the genome system architecture through reorganization of the repetitive DNA content is a fundamental aspect of evolutionary change.

The resetting process can occur without major changes in the protein coding sequences. On the one hand, many organisms within a phylum have functionally interchangeable proteins but differ critically in how protein synthesis is regulated during development (e.g., mice and elephants). On the other hand, there exist many cases of so-called "sibling species" pairs, such as *D. melanogaster* and *D. simulans*, that have no significant

phenotypic differences (or even changes in the order of genetic loci on the chromosomes), and yet the two species can easily be distinguished by their repetitive DNAs.[5,25] Sibling species can often interbreed to produce viable but sterile progeny. The viability of the hybrid progeny shows that protein coding sequences and their regulation are effectively interchangeable in the two species, whereas the sterility indicates that gamete production (i.e., genome maintenance in the germline) has gone awry.

If the system architecture resetting view of speciation has merit, then we would expect to find the repetitive DNA elements that help determine patterns of gene expression and genome maintenance to be the most taxonomically specific components of the genome. They are.[24] As long ago as 1979, it was possible to construct a better primate phylogeny based on restriction site polymorphisms in the tandem repeat arrays of alphoid DNA segments which surround the centromeres of all primate chromosomes than could be generated by comparing protein coding sequences, which are often highly conserved across species boundaries.[22] Fingerprinting of simple tandem sequence repeats and microsatellite arrays can be used in all biologic taxa to identify individuals, families, races, species, and genera.[27,31,37] The specificity of primate alphoid DNA even extends to the level of individual chromosomes.[47]

NATURAL GENETIC ENGINEERING: CELLULAR TOOLS FOR GENOME REORGANIZATION

The idea that genomes are built up Lego-like out of codons specifying protein domains, regulation of gene expression, and genome maintenance dates back to McClintock's observations that developmental patterns in maize varied by the insertion and excision of chromosome bits that she called "controlling elements".[59,60] The idea of exon shuffling as the origin of different protein structures is another variant of the same theme.[11,23,34] The view that evolutionary genetic change is largely a process of codon reorganization requires that cells contain the biochemical activities needed to carry out processes equivalent to the kind of genetic engineering practiced in biotechnology. They do.

Another major molecular genetic lesson in the last four decades has been that all cells contain "natural genetic engineering" capacities (refs. 3, 6, 15, 70, 71, 73, 75, and 77; Arber, this volume). From its earliest prokaryotic days, molecular genetics has focused on the ability of cells to carry out recombination between homologous DNA segments, to integrate exogenous DNA (transformation), to transfer DNA from one cell to another (plasmids and phages), to insert and excise episomes (F and λ), to mobilize defined DNA segments from one location to another (transposons), and to join DNA segments that do not share sequence homology (so-called "illegitimate recombination" involving topoisomerase and transposase activities).[73,76] Recent results on bacterial antibiotic resistance and virulence determinants have shown that conservative site-specific recombination events can integrate single-ORF "cassettes" to build up multicistronic operons in plasmids, transposons, and chromosomes (refs. 42 and 57; Hall, this volume).

Studies of eukaryotes have greatly extended the range of well documented *in vivo* cellular DNA manipulations to include homology-independent integration of transfected DNA into mammalian chromosomes; regular site-specific nonhomologous V(D)J joining, class switching and somatic hypermutation during development of the vertebrate immune

system (ref. 2; Lewis, Kenter, this volume); massive genome cleavage and restructuring in the ciliated protozoa (refs. 39, 65, and 87; Prescott, this volume); and reverse transcription coupled with the genomic insertion of pseudogenes, proretroviruses, and retrotransposable elements.[13,83]

NATURAL GENETIC ENGINEERING AND SIGNAL TRANSDUCTION: GENETIC CHANGE SUBJECT TO CELLULAR CONTROL

It is widely recognized that a great deal of genetic variation results from the action of cellular biochemical functions. For example, Green has estimated that about 80% of all "spontaneous" visible mutations in *D. melanogaster* result from insertion of only four different retrotransposons.[38] Nonetheless, the conceptual implications of natural genetic engineering are not widely recognized, and most evolutionists try (unrealistically) to model the action of these cellular functions to resemble the random mutational events of conventional evolutionary theory.[18] What distinguishes cellular biochemistry from chemical events outside the living cell is that cellular events are subject to biological regulation by signal transduction networks.[1] There is abundant evidence that DNA biochemistry is no exception to this rule and that natural genetic engineering is also subject to biological regulation controlling both the timing and the localization of changes, as seen in the following cases.

Regular DNA Rearrangements

The clearest instances of regulated natural genetic engineering are the systems that operate in a developmentally specific fashion. The fact that the DNA rearrangement activities only operate at specific stages of the organismal life-cycle shows that expression and/or activity is subject to the developmental control network. Examples include (1) the excision, cleavage, and rejoining activities in macronuclear development following mating of ciliated protozoa (refs. 39, 65, and 87; Prescott, this volume); (2) the removal of heterochromatin blocks from the chromosomes of sibling Cyclops species at specific cleavage divisions in early embryonic development;[5] and (3) V(D)J joining, class switching, and somatic hypermutation, each of which only occurs at specific stages of lymphocyte maturation in the vertebrate immune system (ref. 10; Lewis, Kenter, this volume). Regulatory phenomena linked to the life-cycle are also seen with yeast retrotransposons. Ty1 transposition is inhibited posttranscriptionally in haploid cells by mating pheromone.[86] By contrast, Ty3 transcription is activated by the pheromone-response pathway and (like Ty1) repressed by the diploid MATa/α transcription factors, thereby programming Ty3 retrotransposition to occur specifically during mating.[52]

In addition to programmed DNA rearrangements, there are cases in which natural genetic engineering systems are normally inactive in the organismal life cycle but remain subject to derepression/activation under particular circumstances. These systems exemplify the kind of episodic temporal specificity needed to generate high levels of genome reorganization during brief periods of evolutionary change. Two broad categories of such episodically active systems are:

Stress-Induced Mutations

This category includes mutations stimulated by cycles of chromosome breakage that activate transposable elements,[58] by radiation or chemical DNA damage that induces SOS mutagenesis,[81] by passage through tissue culture conditions that activates transposable elements and produces somaclonal variation in plants,[64] by wounding or pathogen attack that activates plant retrotransposons,[84] and by aerobic starvation that activates many mutational systems in bacteria, such as Mu-mediated coding sequence fusions,[72,77] IS excisions,[41] and recombination-dependent frameshifts (ref. 16; Foster, this volume; Rosenberg, this volume).

Hybrid Dysgenesis

This term refers to major germline instabilities, including insertion mutations and chromosome rearrangements, which follow intraspecific matings between two distinct populations, one of which harbors a family of transposable elements.[12] The best-studied examples of hybrid dysgenesis involve P factor DNA transposons[29] and I factor retroposons[14] in *Drosophila*. Hybrid dysgenesis is a particularly relevant model for evolutionary change, because it occurs in natural populations,[35] results in the amplification and dispersion of mobile elements throughout the genome, and generates genomic changes during *mitotic* germline development, so that subsequent meioses produce clusters of gametes (hence progeny) which all share the same newly configured genome. Mating between previously separated populations is also a very attractive candidate for the kind of abnormal event that could trigger widespread genome reorganization, leading to speciation at times of environmental crisis (i.e., when few mates from the same population would be available).

NATURAL GENETIC ENGINEERING AND SIGNAL TRANSDUCTION: TARGET SELECTION

A distinct mode of regulation of natural genetic engineering activities produces nonrandom target selection within the genome. A high degree of sequence specificity is inherent in many natural genetic engineering systems, because they use recognition codons, such as transposase binding sites at the termini of transposable elements (ref. 6; Fedoroff, Iida, this volume) or the joining signals that direct V(D)J recombination and class switching in immunoglobulin loci (ref. 2; Lewis, Kenter, this volume). However, natural genetic engineering also produces changes that involve "target" sites in the genome which have no recognition codons. These targets can vary widely in sequence (e.g., regions subject to transposable element insertion). The potential for action throughout the genome does not mean, however, that target selection is a random process. A growing body of evidence indicates that cellular signal transduction networks can guide natural genetic engineering systems to preferred locations, as seen in the following cases.

Somatic Hypermutation

This is a late process in immunoglobulin (Ig) genetic engineering that creates clustered base substitutions in the 5′ part of the Ig coding sequence and thus produces antibodies for selection of higher affinity variants.[2,10,79] Hypermutation occurs 1–2 kb downstream of Ig promoters and will alter non-Ig sequences linked to the appropriate transcription initiation signals. Although it had been assumed that hypermutation only occurred in Ig determinants, a recent paper reports that certain select non-Ig loci also undergo hypermutation in B lymphocytes (*BCL6* but not *c-MYC* or the *S14* ribosomal locus).[78] Thus, lymphocytes (and potentially other cell types as well) can use transcriptional control signals to direct mutagenic activities within the genome.

Ty Element Targeting

The various classes of Ty elements have preferential insertion specificities within the genome. Ty1 elements avoid coding sequences and insert preferentially 5′ to PolIII-transcribed tRNA loci or PolII-transcribed protein coding sequences, indicating that they are attracted to DNase hypersensitive sites.[21,28,46,69] Ty3 is more highly specific and inserts within 4 bp of PolIII start sites.[17] It has also been demonstrated that PolIII transcription factors target Ty3 integration in vitro.[53] By contrast, Ty5 elements show specificity for silent chromatin regions at mating-type loci and telomeres.[88]

Telomere-Associated Retrotransposon Targeting

As just mentioned, *Drosophila* uses retrotransposons rather than telomerase to extend the ends of its chromosomes[7,55,63] This process involves a high rate of retrotransposition (~1% per generation) which is limited to movement from centromeric heterochromatin to the telomeres. Because the Het-A and TART elements can also cap broken chromosome ends, it appears that this example of specific insertion involves recognition of chromosome ends.

P Element Targeting

The insertional specificity of synthetic P element transformation vectors depends on the transcriptional signals (codons) contained within. The original studies of elements carrying the *White* locus showed a higher-than-expected frequency of inserts near *White* on the X chromosome.[44] Later studies have shown that inclusion of a small *engrailed* fragment preferentially targeted P elements near loci expressed in stripes during development,[51] whereas inclusion of a larger *engrailed* fragment gave high-frequency homing to *engrailed* (7 of 20 inserts in ref. 43; see also ref. 85). Similar biasing of insertional specificity of a P element construct containing regulatory segments from the *polyhomeotic* (*ph*) locus to sites corresponding to binding sites for the Polyhomeotic and Polycomb proteins was found.[33] These results indicate that P factors are similar to Ty3 in that transcription factors can direct insertion to regions where they also direct RNA polymerase.

NATURAL GENETIC ENGINEERING AND EVOLUTION OF COMPLEX ADAPTATIONS

Although new species can evolve without significant phenotypic change, most important evolutionary events are believed to be accompanied by the appearance of new adaptive phenotypes. One of the most important questions in evolution is: How can new adaptations originate? This is a difficult question, because most evolutionary novelties, such as the eye or the wing, involve the orchestrated expression of many different loci, a number of which act in the expression of multiple phenotypes.[1,26] Conventional explanations that randomly generated advantageous changes in complex characters accumulate one locus at a time are unconvincing on both functional and probabilistic grounds, because there is too much interconnectivity and too many degrees of mutational freedom. The genomic reorganization perspective, however, allows us to restate the question of adaptive novelties as: How can a complex multicomponent genomic system be assembled before screening by selection?

The outlines of an answer to the foregoing question lie in the demonstrated ability of natural genetic engineering systems to operate nonrandomly at multiple loci. It is well documented that natural genetic engineering systems can be activated to work at many sites in the genome within a single cell or organismal generation, as they do in hybrid dysgenesis[12,14,29] and macronuclear development (refs. 65 and 87; Prescott, this volume). Thus, it is mechanistically plausible to postulate that major changes can occur rapidly in the repetitive DNA content of the genome during speciation. Over three decades ago, McClintock demonstrated the potential for coordinated multilocus changes by creating systems where expression in maize kernels of unlinked endosperm and aleurone loci were brought under common control of an autonomous transposable element (refs. 59 and 60; Fedoroff, this volume). Similarly, multiple loci in the yeast genome can be brought under common mating-type control by insertions of Ty1 retrotransposons.[32] The documented preference of Ty1 for insertion in 5′ regulatory regions[28] dramatically raises the probability that multiple insertions will lead to alterations of transcriptional regulation without deleterious coding sequence mutations.

Established natural genetic engineering capabilities thus provide a plausible molecular mechanism for the generation of novel coordinately regulated genomic systems. In its simplest form, this mechanism depends only on activation of one or more mobile element systems that can rapidly insert regulatory motifs into appropriate sites in multiple genetic loci, leaving selection the task of picking out variants with new functionalities. However, it is not necessary to postulate that there will be a completely "blind" choice of insertion targets. The genome is dynamically organized by the transcriptional regulatory apparatus into suites of functionally integrated loci.[1,54] It is likely that these suites will constitute preferred coordinate insertion targets, because they will be in similar physical states (e.g., open chromatin, undergoing active transcription[19]) and because (as just discussed) target choice by mobile elements can be influenced by interaction with the transcription apparatus. In other words, the cellular networks that interpret the status of each cell and adjust gene expression accordingly are also likely to play an important role in determining patterns of genome reorganization.

CONCLUSION

In summary, as we approach a new century, molecular genetics has provided us a more detailed view of the genome and revealed previously unsuspected cellular capabilities for genome restructuring. These molecular insights lead to new concepts of how genomes are organized and reorganized, opening a range of possibilities for thinking about evolution. Rather than being restricted to contemplating a slow process depending on random (i.e., blind) genetic variation and gradual phenotypic change, we are now free to think in realistic molecular ways about rapid genome restructuring guided by biological feedback networks.

ACKNOWLEDGMENTS

I am particularly grateful to the late Barbara McClintock, who made me open my eyes to new ways of looking at evolutionary change and genome function, and to Phillip Harriman and the National Science Foundation, who have supported my research on natural genetic engineering for over two decades. I also thank my colleagues Sankar Adhya, Jon Beckwith, the late Karen Ippen-Ihler, Lorne MacHattie, Pat Higgins, Genevieve Maenhaut-Michel, and David Leach for helping me realize how skillful the *E. coli* K12 cell can be as a genetic engineer.

REFERENCES

1. ALBERTS, B., D. BRAY, J. LEWIS, M. RAFF, K. ROBERTS & J.D. WATSON. 1994. The Molecular Biology of the Cell, 3rd Ed. Garland. New York.
2. ALT, F.W., T.K. BLACKWELL & G.D. YANCOPOULOS. 1987. Development of the primary antibody repertoire. Science **238:** 1079–1087.
3. ARBER, W. 1993. Evolution of prokaryotic genomes. Gene **135:** 49–56.
4. BARDWELL, L., J.G. COOK, C.J. INOUYE & J. THORNER. 1994. Signal propagation and regulation in the mating pheromone response pathway of the yeast *Saccharomyces cerevisiae*. Dev. Biol. **166:** 363–379.
5. BEERMAN, S. 1977. The diminution of heterochromatic chromosomal segments in Cyclops (Crustacea, Copepoda). Chromosoma **60:** 297–344.
6. BERG, D.E. & M.M HOWE, Eds. 1989. Mobile DNA. ASM Press, Washington, D.C.
7. BIESSMANN, H. & J.M. MASON. 1997. Telomere maintenance without telomerase. Chromosoma **106:** 63–69.
8. BLACKBURN, E.H. 1991. Structure and function of telomeres. Nature **350:** 569–573.
9. BLACKBURN, E.H. & J.W. SZOSTAK. 1984. The molecular structure of centromeres and telomeres. Ann. Rev. Biochem. **53:** 163–194.
10. BLACKWELL, T.K. & F.W. ALT. 1989. Mechanism and developmental program ofimmunoglobulin gene rearrangement in mammals. Ann. Rev. Genet. **23:** 605–636.
11. BLAKE, C.C.F. 1985. Exons and the evolution of proteins. Int. Rev. Cytol. **93:** 149–185.
12. BREGLIANO, J.-C. & M.G. KIDWELL. 1983. Hybrid dysgenesis determinants. In ref. 71: 363–410.
13. BROSIUS, J. 1991. Retroposons—Seeds of evolution. Science **251:** 753.
14. BUCHETON, A. 1990. I transposable elements and I-R hybrid dysgenesis in Drosophila. Trends Genet. **6:** 16–21.
15. BUKHARI A.I., J.A. SHAPIRO & S.L. ADHYA. 1977. DNA Insertion Elements, Episomes and Plasmids. Cold Spring Harbor Press. Cold Spring Harbor, NY.
16. CAIRNS, J. & P. L. FOSTER. 1991. Adaptive reversion of a frameshift mutation in Escherichia coli. Genetics **128:** 695–701.

17. CHALKER, D.L. & S.B. SANDMEYER. 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. **6:** 117–128.
18. CHARLESWORTH, B., P. SNIEGOWSKI & W. STEPHAN. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **317:** 215–220.
19. CURCIO, M.J. & R.H. MORSE. 1996. Tying together integration and chromatin. Trends Genet. **12:** 436–438.
20. DEMBURG, A.F., J.W. SEDAT & R.S. HAWLEY. 1996. Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell **86:** 135–146.
21. DEVINE, S.E. & J.D. BOEKE. 1996. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes & Dev. **10:** 620–633.
22. DONEHOWER, L. & D. GILLESPIE. 1979. Restriction site periodicities in highly repetitive DNA of primates. J. Mol. Biol. **134:** 805–834.
23. DOOLITTLE, R.F. 1995. The multiplicity of domains in proteins. Ann. Rev. Biochem. **64:** 287–314.
24. DOVER, G.A. 1982. Molecular drive: A cohesive mode of species evolution. *Nature* **299**, 111–117.
25. DOWSETT, A.P. 1983. Closely related species of *Drosophila* can contain different libraries of middle repetitive DNA sequences. Chromosoma **88:** 104–108.
26. DUBOULE, D. & A.S. WILKINS. 1998. The evolution of 'bricolage.' Trends Genet. **14:** 54–59.
27. ELDER, J.F. & B.J. TURNER. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Quart. Rev. Biol. **70:** 297–320.
28. EIBEL, H. & P. PHILIPPSEN. 1984. Preferential integration of yeast transposable element Ty into a promoter region. Nature **307:** 386–388.
29. ENGELS, W.R. 1989. P elements in *Drosophila melanogaster.* In ref. 6: 437–484 .
30. EPPLEN, C., E.J. SANTOS, W. MAUELER, P. VAN HELDEN & J.T. EPPLEN. 1997. On simple repetitive DNA sequences and complex diseases. Electrophoresis **18:** 1577–1585.
31. EPPLEN, J.T., W. MAUELER & C. EPPLEN. 1994. Exploiting the informativity of "meaningless" repetitive DNA from indirect gene diagnosis to multilocus genome scanning. Biol. Chem. Hoppe Seyler **375:** 795–801.
32. ERREDE, B., T.S. CARDILLO, G. WEVER & F. SHERMAN. 1981. ROAM mutations causing increased expression of yeast genes: Their activation by signals directed toward conjugation functions and their formation by insertions of Ty1 repetitive elements. Cold Spring Harbor Symp. Quant. Biol. **45:** 593–607.
33. FAUVARQUE, M.O. & J.M. DURA. 1993. Polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila.* Genes Dev. **7:** 1508–1520.
34. GILBERT, W. 1978. Why genes in pieces? Nature 271: 501.
35. GOLUBOVSKY, M.D. 1980. Mutational process and microevolution. Genetika **52/53:** 139.
36. GRALLA, J.D. & J. COLLADO-VIDES. 1996. Organization and function of transcription regulatory elements. In *Escherichia coli* and Salmonella Cellular and Molecular Biology, 2nd Ed. F.C. Neidhardt *et al.*, Eds.: 1232–1245. ASM Press. Washington, D.C .
37. GRECHKO, V.V. *et al.* 1997. Restriction endonuclease analysis of highly repetitive DNA as a phylogenetic tool. J. Mol. Evol. **45:** 332–336.
38. GREEN, M.M. 1988. Mobile DNA elements and spontaneous gene mutation. In Eukaryotic transposable elements as mutagenic agents. Banbury Rep. **30:** 41–50.
39. GRESLIN, A.F., D.M. PRESCOTT, Y. OKA, S.H. LOUKIN & J.C. CHAPPELL. 1989. Reordering of nine exons is necessary to form a functional actin gene in *Oxytricha nova.* Proc. Natl. Acad. Sci. USA **86:** 6264–6268.
40. GROSVELD, F. *et al.* 1993. The regulation of human globin gene switching. Phil. Trans. Roy. Soc. Lond. Series B. **339:** 183–191.
41. HALL, B.G. 1988. Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence. Genetics **120:** 887–897.
42. HALL, R.M. 1997. Mobile gene cassettes and integrons: Moving antibiotic resistance genes in gram-negative bacteria. Ciba Found. Symp. **207:** 192–202.
43. HAMA, C., Z. ALI & T.B. KORNBERG. 1990. Region-specific recombination and expression are directed by portions of the *Drosophila* engrailed promoter. Genes Dev. **4:** 1079–1093.
44. HAZELRIGG, T., R. LEVIS & G.M. RUBIN. 1984. Transformation of white locus DNA in *Drosophila*: Dosage compensation, zeste interaction, and position effects. Cell **36:** 469–481.
45. HENIKOFF, S. 1992. Position effect and related phenomena. Curr. Opin. Genet. Dev. **2:** 907–912.

46. JI, H., D.P. MOORE, M.A. BLOMBERG, L.T. BRAITERMAN, D.F. VOYTAS, G. NATSOULIS & J.D. BOEKE. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell **73:** 1007–1018.
47. JORGENSEN, A.L. 1997. Alphoid repetitive DNA in human chromosomes. Dan. Med. Bull. **44:** 522–534.
48. KARPEN, G.H. 1994. Position-effect variegation and the new biology of heterochromatin. Curr. Opin. Genet. Dev. **4:** 281–291.
49. KARPEN, G.H., M.H. LE & H. LE. 1996. Centric heterochromatin and the efficiency of achiasmate disjunction in *Drosophila* female meiosis. Science **273:** 118–122.
50. KASHI, Y., D. KING & M.S. DOLLER. 1997. Simple sequence repeats as a source of quantitative variation. Trends Genet. **13:** 74–78.
51. KASSIS, J.A., E. NOLL, E.P. VANSICKLE, W.F. ODENWALD & N. PERRIMON. 1992. Altering the insertional specificity of a Drosophila transposable element. Proc. Natl. Acad. Sci. USA **89:** 1919–1923.
52. KINSEY, P.T. & S.B. SANDMEYER. 1995. Ty3 transposes in mating populations of yeast: A novel transposition assay for Ty3. Genetics **139:** 81–94.
53. KIRCHNER, J., C.M. CONNOLLY & S.B. SANDMEYER. 1995. Requirement of RNA polymerase III transcription factors for *in vitro* position-specific integration of a retroviruslike element. Science **267:** 1488–1491.
54. LAMOND, A.I. & W.C. EARNSHAW. 1998. Structure and function in the nucleus. Science **280:** 547–553.
55. LEVIS, R.W., R. GANESAN, K. HOUTCHENS, L.A. TOLAR & F.M. SHEEN. 1993. Transposons in place of telomeric repeats at a Drosophila telomere. Cell **75:** 1083–1093.
56. LEWIS, E.B. 1992. Clusters of master control genes regulate the development of higher organisms. J.A.M.A. **267:** 1524–1531.
57. MAZEL, D., B. DYCHINCO, V.A. WEBB & J. DAVIES. 1998. A distinctive class of integron in the Vibrio cholerae genome. Science **280:** 605–608.
58. MCCLINTOCK, B. 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA **36:** 344–355.
59. MCCLINTOCK, B. 1956. Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol. **21:** 197–216.
60. MCCLINTOCK, B. 1965. The control of gene action in maize. Brookhaven Symp. Biol. **18:** 162–184.
61. MCKEE, B.D., L. HABERA & J.A. VRANA. 1992. Evidence that intergenic spacer repeats of *Drosopila melanogaster* rRNA genes function as X-Y pairing sites in male meiosis, and a general model for achiasmatic pairing. Genetics **132:** 529–544.
62. MILLER, J.H. 1996. Structure of a paradigm. Nat. Struct. Biol. **3:** 310–312.
63. PARDUE, M.L., O.N. DANILEVSKAYA, K. LOWENHAUPT, F. SLOT & K.L. TRAVERSE. 1996. Drosophila telomeres: New views on chromosome evolution. Trends Genet. **12:** 48–52.
64. PESCHKE, V.M., R.L. PHILLIPS & B.G. GENGENBACH. 1987. Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science **238:** 804–807.
65. PRESCOTT, D.M. 1992. The unusual organization and processing of genomic DNA in hypotrichous ciliates. Trends Genet. **8:** 439–445.
66. PTASHNE, M. 1986. A Genetic Switch. Cell/Blackwell. Cambridge, MA; Palo Alto, CA.
67. REZNIKOFF, W.S. 1992. The lactose operon-controlling elements: A complex paradigm. Molec. Microbiol. **6:** 2419–2422.
68. SAIER, M.H., JR., T.M. RAMSEIER & J. REIZER. 1996. Regulation of carbon utilization. In *Escherichia coli* and Salmonella Cellular and Molecular Biology, 2nd Ed. F.C. Neidhardt *et al.*, Eds.: 1325–1343. ASM Press. Washington, D.C .
69. SANDMEYER, S.B., L.J. HANSEN & D.L. CHALKER. 1990. Integration specificity of retrotransposons and retroviruses. Ann. Rev. Genet. **24:** 491–518.
70. SHAPIRO, J.A. 1977. DNA insertion elements and the evolution of chromosome primary structure. Trends Biochem. Sci. **2:** 622–627.
71. SHAPIRO, J.A. 1983. Mobile Genetic Elements. Academic Press. New York.
72. SHAPIRO, J. 1984. Observations on the formation of clones containing araB-lacZ cistron fusions. Molec. Gen. Genet. **194:** 79–90.
73. SHAPIRO, J.A. 1985. Mechanisms of DNA reorganization in bacteria. Int. Rev. Cytol. **93:** 25–56.

74. SHAPIRO, J.A. 1991. Genomes as smart systems. Genetica **84:** 3–4.
75. SHAPIRO, J.A. 1992. Natural genetic engineering in evolution. Genetica **86:** 99–111.
76. SHAPIRO, J.A. 1995. The discovery and significance of mobile genetic elements. In Mobile Genetic Elements—Frontiers in Molecular Biology. D.J. Sherratt, Ed.: 1–17. IRL Press. Oxford, U.K.
77. SHAPIRO, J.A. 1997. Genome organization, natural genetic engineering, and adaptive mutation. Trends Genet. **13:** 98–104.
78. SHEN, H.M., A. PETERS, B. BARON, X. ZHU & U. STORB. 1998. Mutation of the BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science **280:** 1750–1752.
79. STORB, U. 1996. Molecular mechanism of somatic hypermutation of immunoglobulin genes. Curr. Op. Immunol. **8:** 206–214.
80. TRIFONOV, E.N. & V. BRENDEL. 1986. GNOMIC: A Dictionary of Genetics Codes. Balaban. Philadelphia.
81. WALKER, G.C. 1987. The SOS response of *Escherichia coli.* In *Escherichia coli* and *Salmonella typhimurium.* F. C. Neidhardt *et al.*, Eds.: 1346–1357. ASM Press, Washington, D.C.
82. WEILER, K.S. & B.T. WAKIMOTO. 1995. Heterochromatin and gene expression in Drosophila. Ann. Rev. Genet. **29:** 577–605.
83. WEINER, A.M., P.L. DEININGER & A. EFSTRATIADIS. 1986. Nonviral retroposons: Genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. Ann. Rev. Biochem. **55:** 631–661.
84. WESSLER, S.R. 1996. Turned on by stress. Plant retrotransposons. Curr. Biol. **6:** 959–961.
85. WHITELEY, M. & J.A. KASSIS. 1997. Rescue of Drosophila engrailed mutants with a highly divergent mosquito engrailed cDNA using a homing, enhancer-trapping transposon. Development **124:** 1531–1541.
86. XU, H. & J.D. BOEKE. 1991. Inhibition of Ty1 transposition by mating pheromones in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **11:** 2736–2743.
87. YAO, M-C. 1989. Site-specific chromosome breakage and DNA deletion in ciliates. In ref. 6: 715–734.
88. ZOU, S., N. KE, J.M. KIM & D.F. VOYTAS. 1996. The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev. **10:** 634–645.