Genome organization, natural genetic engineering and adaptive mutation

JAMES A. SHAPIRO (jsha@midway.uchicago.edu)

Bacterial evolution is considered in the light of molecular discoveries about genome organization, biochemical mechanisms of genetic change, and cellular control networks. Prokaryotic genetic determinants are organized as modular composites of coding sequences and protein-factor binding sites joined together during evolution. Studies of genetic change have revealed the existence of biochemical functions capable of restructuring the bacterial genome at various levels and joining together different sequence elements. These natural genetic engineering systems can be subject to regulation by signal transduction networks conveying information about the extracellular and intracellular environments. Mu-mediated araB–lacZ coding sequence fusions provide one example of adaptive mutation (increased formation of useful mutations under selection) and illustrate how physiological regulation can modulate the activity of a natural genetic engineering system under specific conditions.

Molecular genetics has greatly extended our knowledge of how genomic systems are organized, how they change and how cellular complexes acting on the genome respond to regulatory circuits. This new understanding is fundamental to thinking about the genomic reorganization processes that are at the heart of evolutionary change. Here I discuss a specific example of the adaptive mutation phenomenon and attempt to integrate research on organization patterns of genetic systems, molecular mechanisms of genetic change and signal transduction networks. While the discussion focuses on work with bacteria, the general principles exemplified by prokaryotic systems are applicable to all organisms.

The term 'adaptive mutation' has been applied to mutational events that occur more frequently under selective conditions, when the resulting genetic changes are adaptively useful, than during normal growth. Suggestions that mutations can be directed by selective conditions to produce only certain beneficial changes have not been substantiated. In this article, 'adaptive mutation' is understood to mean two things: (1) mutagenesis increases in response (or adaptation) to new physiological conditions during selection for a particular trait; and (2) the ability to increase the frequency of potentially useful mutations is beneficial (adaptive) for the bacterial population undergoing selection. It is worth noting that observations on the influence of environmental conditions on the levels of hereditary variation are hardly novel and go back at least as far as the opening chapter of Darwin's *Origin of Species*.

What has molecular genetics taught us about genomic organization?

One of the earliest and most intensively studied genetic systems is the *Escherichia coli lac* operon. A historical survey of changes in how we have depicted *lac* over the past 50 years reveals the revolutionary impact of molecular genetics on our conceptions of genetic elements[1]. The *lac* operon has been deconstructed from a dimensionless point on a genetic map into an intricate complex of polypeptide coding sequences and DNA binding sites for transcription factors (Fig. 1). All the basic elements of 'gene structure' as we understand them today are found in the *lac* operon. Of particular importance were the realizations that *lac* encoded four different proteins, that one of these proteins was a regulatory DNA-binding molecule, and that the operon contained several different classes of *cis*-acting regulatory sites that are fundamentally different from polypeptide-encoding sequences. The discovery of the promoter and the *CAP*-binding site for the catabolite activator protein (cAMP–Crp) complex are additionally important because they represent repetitive genetic elements dispersed at many genetic loci. These repetitive elements constitute the physical basis for integrating distributive genomic networks, such as the family of catabolite-sensitive functions in *E. coli*[2].

Beyond what is illustrated in Fig. 1, we know that each component of the *lac* operon has its own internal organization composed of smaller elements. For example, *lacI* contains sequences encoding protein domains involved in DNA binding, inducer binding and oligomerization[3]. Sequence database comparisons indicate that this modular organization into separate domains is the rule among coding sequences in all organisms[4]. Similarly, each promoter contains –10 and –35 consensus elements as well as spacer regions of defined length[5]. In other words, the *lac* operon, like all expressed genetic loci, is organized hierarchically in a modular fashion out of a series of DNA sequence elements. Other kinds of genetic determinants, such as replication origins and replicon partitioning sites, display analogous but distinct patterns of modular organization. At higher levels, plasmids, viruses, multilocus regulons (e.g. for chemotaxis, symbiosis, protein secretion), and even whole genomes all display their own characteristic architectures integrating smaller and smaller subsystems[6]. Molecular taxonomy of *E. coli* plasmids provides additional evidence for such evolutionary mosaicism[7].

In thinking about how such determinants and higher-order systems could have evolved by reassortment of genetic motifs, it is hard to avoid postulating a role for cut-and-splice processes to stitch different sequence elements together[8,9]. Following this line of thought one step further, we need to ask whether the DNA rearrangements, which used to be called 'illegitimate recombination', might not ultimately prove to be key genetic events in generating functionally significant evolutionary variability. The conventional alternative is to assume that repeated motifs have all evolved independently by random mutation. One would expect this process to be relatively slow. The existence of well-documented molecular mechanisms for amplifying and dispersing repetitive elements throughout the genome means that DNA rearrangement processes offer a biochemically plausible and more rapid alternative[8,9].

 0168-9525/97/$17.00
PII: S0168-9525(97)01058-5

What has molecular genetics taught us about the mechanisms of mutational change?

The molecular analysis of bacterial mutagenesis has taught us two fundamental and complementary lessons. The first lesson is that cells possess a wide range of repair and proofreading functions to remove accidental changes in DNA sequence information[10]. It has been surprising to learn how much of the bacterial genome is dedicated to encoding functions that correct stochastic genetic changes resulting from replication errors and physicochemical insults, suggesting that bacteria have little tolerance for purely random variability. The complex biochemistries of repair and proofreading systems provide many opportunities for regulation. Regulation operates in response to damage itself, such as UV induction of the SOS system[11], but can also respond to other physiological inputs, such as cAMP control of SOS (Ref. 12).

The second lesson of molecular studies of mutagenesis is that cells possess numerous biochemical systems capable of changing and reorganizing DNA sequence information (Table 1). These biochemical complexes can be characterized collectively as natural genetic engineering systems[9]. The discovery of cellular systems for genome restructuring dates back to the 1940s with McClintock's work on transposable elements in maize and the study of episomes (temperate bacteriophages and sex factor plasmids) by the pioneers of bacterial genetics (summarized in Refs 13, 14). They are the tools cells can use to modify their genomes in ways that resemble our own genetic engineering.

The ecological and evolutionary relevance of natural genetic engineering can be seen most clearly in the emergence of bacterial antibiotic-resistance systems based on plasmids, transposons and gene cassette/integron systems for exchanging specific resistance determinants[13,15,16]. In the case of operons encoding multiple antibiotic resistance functions, we have explicit examples of their origin by a genetic engineering process involving successive site-specific recombination events[16]. In thinking about possible theories of evolutionary change, it is useful to remember that the discovery of plasmids encoding multiple antibiotic resistances in the 1950s (Ref. 15) was a surprise that contradicted the established view that antibiotic resistance in bacteria would emerge by the accumulation of point mutations modifying cellular targets. Antibiotic resistance transposons (Tn*3*, Tn*5*, Tn*10* and so on) further exemplify the natural genetic engineering concept because they not only mobilize their own resistance determinants but also generate deletions, inversions and translocations of adjacent sequences[17–19].

Adaptive mutation as an example of natural genetic engineering systems at work: *araB–lacZ* fusions

It has been known for many years that mutations occur at low levels in starved bacterial cultures[20]. The distinct phenomenon we now call adaptive mutation (increased mutagenesis activated by selective conditions) was first clearly documented in the *araB–lacZ* fusion system (Fig. 2)[21]. I focus on this system as a paradigm because of my own familiarity with it. The depth of understanding of the *araB–lacZ* fusion system is built upon a foundation of elegant studies of *Mu* genetics, biochemistry and regulation from many laboratories[22].

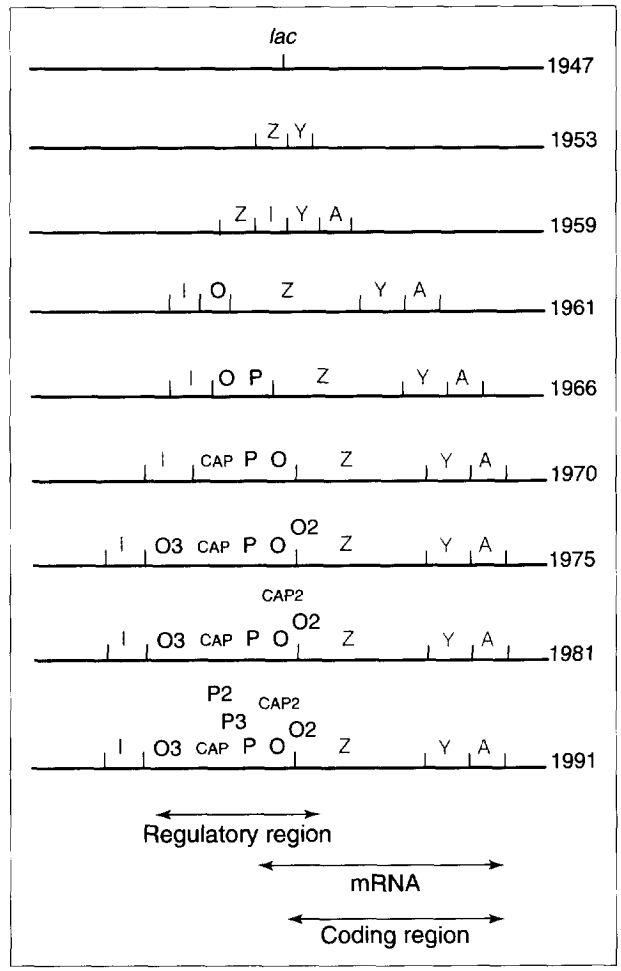

FIGURE 1. Historical deconstruction of the *lac* operon (expanded from Ref. 1). In the late 1940s, Lederberg and Tatum used lactose utilization (*lac*) as a marker to construct genetic maps where each locus was a single dimensionless point. In the 1950s, Monod and his colleagues discovered first that *lac* consisted of two coding units for β-galactosidase (Z) and permease (Y), and then found an additional enzymatic activity (A for transacetylase) as well as a protein governing inducibility (I). With the operon theory in 1961, a new class of genetic element appeared, the operator (O), a *cis*-acting site in the DNA that did not encode a protein product. Later in the same decade, the repressor binding and transcriptional initiation functions were found to be separable and a new *cis*-acting site, the promoter (P), was defined as the site for binding of RNA polymerase and the start of transcription. In 1970, a third *cis*-acting site was identified (CAP) as the locus of catabolite repression. Over the following two decades, it was realized that there are multiple copies of each of these *cis*-acting sites and that effective regulation of the *lac* operon involves cooperative interactions between these sites and the proteins that bind to them. In this figure, which is not drawn to scale, designations for *cis*-acting sites are green and protein encoding sequences are purple. Note that there are some regions where *cis*-acting sites and coding sequences overlap.

Casadaban had devised a way to fuse the 5′ region of any cistron in *E. coli* to *lacZ*, thereby creating a model system for studying the formation of two-domain hybrid protein coding sequences[23]. He used the transposable phage *Mu* as a region of portable genetic homology to construct the prefusion strain, aligning a promoterless *lacZ* (carrying an *ochre* triplet at codon 17) with the other coding sequence (*araB* in the present

TABLE 1. Some natural genetic engineering systems[a]

System	DNA substrate	DNA rearrangement	Functional consequence	Refs
SOS mutagenesis (UmuCD′–RecA complex)	Non-specific, UV-damaged DNA	Point mutations	Loss, alteration, gain of function	11
Homologous recombination	*Neisseria gonorrhoeae* silent pilin cassettes and the pilin expression locus	Gene conversion in the pilin coding sequence	Alteration in pilin structure (antigenic variation), gain or loss of function	67
	F plasmid and chromosomal IS elements	Reciprocal crossing over	Hfr formation, F′ excision	68
	Plasmid IS elements or tandem r-det repeats	Unequal crossing over	Amplification of resistance determinants	13, 52
Site-specific recombination	Antibiotic resistance cassettes, integrons	Integration	Generation of multiple resistance operons	16
	Salmonella typhimurium H2 flagellar region	Promoter inversion	Flagellar phase variation	51
	Phage *Mu* G region (encoding tail-fiber proteins)	Partial coding sequence inversion	Host-range variation	51
Transposition	Plasmid IS*1* elements	Replicon fusion	Formation of cointegrate R plasmids	13, 52
	IS*1* or IS*5* termini, *bgl* operon promoter region	Transposition/insertion	Silent promoter activation	69
	IS*2* termini, silent chromosomal *ampC* locus	Transposition/insertion	Creation of a new promoter	70

[a] Except where indicated, these cases are documented in *Escherichia coli*. (Additional examples of functionally significant biochemically mediated DNA rearrangements can be found in Refs 9, 13, 52, 67.) The inclusion of point mutations in the Table deserves special comment. While the role of protein factors, such as the UmuCD′–RecA complex, in SOS point mutagenesis is usually interpreted as a passive one, merely creating the necessary conditions (bypass replication) that allow mutagenesis to occur, the full details of this process have yet to be elucidated, and it remains true that point mutations are formed under SOS conditions by these protein factors. Many kinds of point mutation (such as the *lac33* frameshifts discussed in the text) require biochemical input in the form of glycosylases, nucleases, helicases, polymerases, ligases and DNA-binding proteins. The thesis of this Perspective is that all these inputs can be considered to represent active cellular participation in the mutagenic process and to present opportunities for regulation. As the example of pilin antigenic variation in *Neisseria gonorrhoeae* further demonstrates, Rec-mediated gene conversion can introduce nucleotide substitutions within a specific small region of the genome. Thus, bacterial cells may be thought to have the biochemical tools needed to carry out localized mutagenesis.

case) so that the appropriate in-frame deletion events could form an active genetic fusion leading to growth on selective medium with lactose as the sole carbon source (Fig. 2). The original idea was that the *Mu* prophage would be a passive source of homology and that 'spontaneous' breakage-rejoining events would generate the actual fusions by removing all blocks to transcription and translation between *araB* and a site in *lacZ* downstream of the U118 *ochre* triplet at codon 17. However, detailed study of the fusion process showed that the *Mu* prophage played an active role in generating *araB–lacZ* fusions and did so in a way that was regulated by cellular physiology.

It was common laboratory lore among workers who used the Casadaban technique that thick plates were needed to obtain fusions because long incubations were necessary. Detailed examination of fusion kinetics on arabinose–lactose plates confirmed these anecdotes[21]. Fusion colonies only appeared after a delay of 4–19 days, and then the number of new colonies per day increased until the plate was saturated. Appropriate control experiments demonstrated that any *araB–lacZ* fusions formed before plating onto selective medium would have supported visible colony growth within two or three days. Thus, surprisingly, fusion events were undetectable during normal growth (at a frequency of less than one fusion per 3×10^{10} cells). Although it was suggested that selective substrates might play a 'directing' role in the fusion process[24], this notion was undermined by the finding that long-term aerobic starvation in glucose medium led to fusion colony appearance within two days of plating[25]. The subsequent demonstration that *araB–lacZ* fusion clones could be purified from glucose-starved cultures that had never been exposed to arabinose or lactose provided definitive evidence that selective substrates played no essential role at any stage of fusion formation[26]. Although the idea of 'directed mutation' was excluded in this case, the remarkable effect of prolonged selection or aerobic starvation in turning on the fusion process remained to be explained. Clearly, the ability to trigger mutagenic events under stress conditions could be extremely useful to organisms at key moments in their evolutionary history when genetic change was essential to survival.

Initial understanding of how cellular physiology could affect the formation of *araB–lacZ* fusions came from molecular mechanistic studies. Examination of a number of hybrid *lacZ* coding sequences derived by the Casadaban technique revealed the presence of *Mu* linkers[27]. These linkers came from the *Mu* extremity adjacent to the 5′ domain in the prefusion strain, but were found inverted adjacent to the 3′ *lacZ* domain in the fusion, indicating a rearrangement more complex than a simple two-site deletion (Fig. 2). The additional finding that MuA transposase activity was needed for *araB–lacZ* fusions led to the formulation of a molecular model in which the *Mu* strand-transfer complex (STC) served as a precursor both to non-replicative fusion events and to *Mu* replication (Fig. 3)[28]. In fusion formation, virtually all of the *Mu* genome is lost, and only short linkers composed of several MuR nucleotides remain (Fig. 2). The fusion process is, therefore, advantageous from the point of view of the host cell, but disadvantageous for *Mu* considered as an autonomous genetic element. The results of sequencing 84 independent *araB–lacZ* fusions, including unselected fusions with a distinct structure, are consistent with this model[29]. The distinct structures of selected and unselected fusions appear to reflect the influence on the molecular details of STC processing of physiological factors, such as *araB* transcription and/or differences between solid and liquid growth. Experiments are in progress to identify these factors.

The mechanistic scheme in Fig. 3 shows that many different proteins and DNA sequences have to come together in an intricate sequence of events involving the formation and rearrangement of nucleoprotein complexes necessary to direct precise phosphodiester bond cleavages and ligations[30–32]. Thus, each *araB–lacZ* fusion results from a multicomponent-cell-biological process, not just a simple biochemical reaction. The distinct biochemical steps constitute targets for physiological regulation. To date, genetic studies have demonstrated the involvement of the following *E. coli* regulatory factors: IHF (integration host factor) and HU (Ref. 28), ClpPX protease (Ref. 33; G. Maenhaut-Michel, M-J. Gama, A. Toussaint and J.A. Shapiro, unpublished), adenylate cyclase and catabolite activator protein (G. Maenhaut-Michel, M-J. Gama, A. Toussaint and J.A. Shapiro, unpublished), and the growth-phase specific regulators HNS and RpoS (J.M. Gomez, J. Blazquez, F. Baquerol and J.L. Martinez, pers. commun.).

The initial expectation was that all regulation would occur at the level of derepressing the *Mu* prophage, and some results supported this idea. For example, ClpPX protease is known to function at the level of repressor inactivation[34], and IHF is needed for *Mu* early protein expression[35]. However, using *Mu* genetics to separate the derepression step from the rest of the fusion process suggests that significant physiological regulation operates at stages other than derepression (Fig. 3; G. Maenhaut-Michel, M-J. Gama, A. Toussaint and J.A. Shapiro, unpublished). We can derepress prefusion strains carrying mutations in the *Mu* prophage *B* and *kil* loci without provoking cell death. Derepression can be accomplished either by thermal inactivation of the *Mu cts62* repressor or by aerobic growth to saturation in broth. (Static broth cultures are not derepressed.) Once induced, *B⁻kil⁻* lysogens can remain derepressed for many generations during normal growth at low temperature. While derepression is necessary for fusions, it is not sufficient. Derepressed bacteria do not immediately make fusions; they must first spend some time under starvation conditions on selection plates.

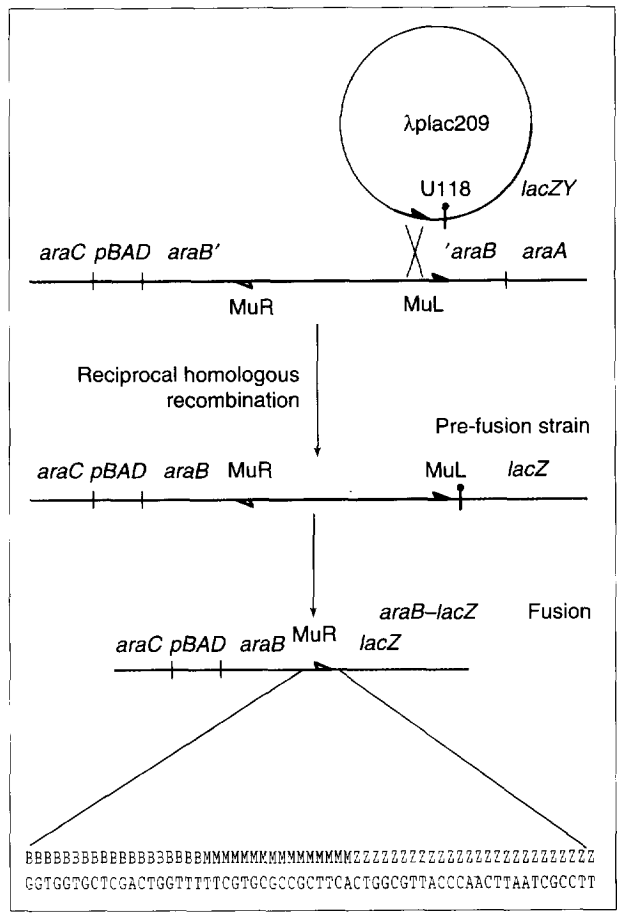

FIGURE 2. Schematic outline of the Casadaban scheme for generating a hybrid *araB–lacZ* fusion coding sequence[23]. The pre-fusion strain is formed by homologous recombination between an *araB::Mucts62* prophage insertion and a *Mu* fragment in the λ*plac*209 phage to align a promoterless *lac* operon downstream of the 5′ portion of *araB*. Note that the *araB* and *lacZ* sequences to be fused are separated by a complete *Mucts62* prophage, and that both *lacY* and a λ prophage are located downstream of *lacZ* (not shown here). The *lacZ* in the pre-fusion strain cannot be expressed because it lacks a functional promoter and because it has the U118 *ochre* mutation at codon 17. In order to recover the ability to grow on lactose when arabinose is present as inducer, the bacteria must undergo an in-frame deletion distal to U118 to create a hybrid *araB–lacZ* cistron. The fused coding sequence can be transcribed from the *araBAD* operon promoter, *pBAD*, which is activated in the presence of arabinose by the cAMP–Crp complex and the AraC transcription factor encoded by *araC*. When *araB–lacZ* hybrid cistrons are sequenced (Refs 26, 29), the fusion joints invariably carry the MuR extremity (nucleotides labelled M) between nucleotides derived from *araB* (B) and from *lacZ* (Z). Note that the orientation of MuR adjacent to *lacZ* nucleotides in the *araB–lacZ* fusion is inverted from its original orientation adjacent to *araB* in the prefusion strain.

A key component of the *E. coli* response to starvation is Crp, the cAMP receptor protein, which is activated under carbon substrate limitation[2]. Mutations inactivating Crp block fusion formation when *araB* transcription has been rendered independent of catabolite activation

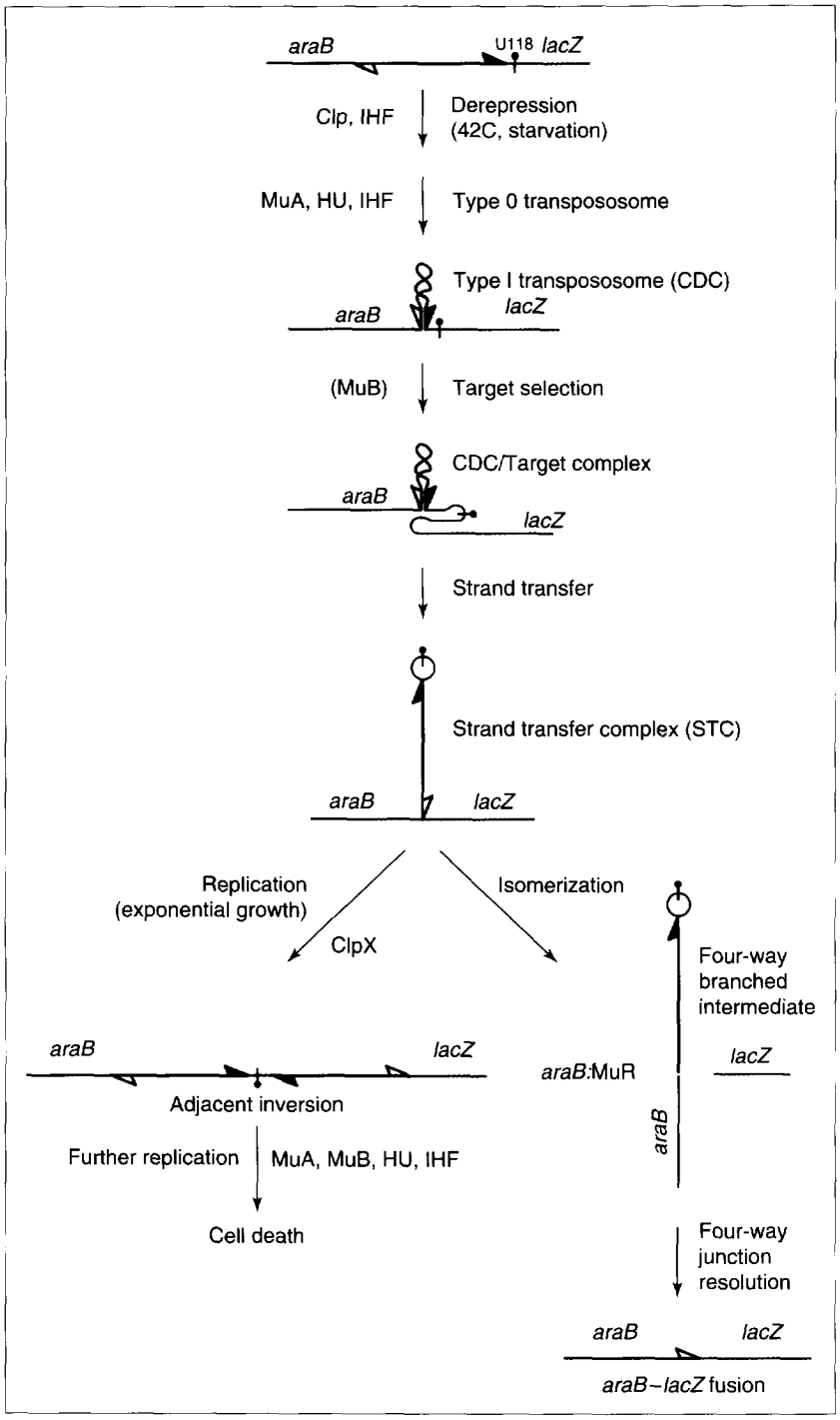

FIGURE 3. Some steps involved in *araB–lacZ* fusion formation. The steps leading to the strand-transfer complex (STC) are based on *Mu* biochemistry[31,32]. The type 0 transpososome is the complex between *Mu* ends and MuA protein; the type I transpososome or cleaved donor complex (CDC) is formed by cleavage of the prophage 3′ termini. The MuA–*Mu* termini complex maintains the prophage as a separate supercoiled domain. The CDC selects a 5 bp genomic target just downstream of the U118 *ochre* triplet, cleaves on each side of the target, and ligates the exposed 5′ termini to the 3′ *Mu* prophage termini to complete strand transfer and generate the STC (Refs 31, 32, 64). (The details of these cleavages and ligations are in Ref. 28.) MuB protein participates in target sequence selection during active *Mu* replication, but DNA rearrangements, transpositions and fusions can occur in the absence of B protein. During exponential growth with MuB protein, the *Mu* prophage in the STC replicates. The first round produces an adjacent inversion of the proximal *lacZ* segment carrying the U118 mutation between inverted *Mu* copies[64]; subsequent rounds of replication are lethal. Under starvation on arabinose–lactose plates, isomerization of the STC to a four-way branched structure containing short *araB*:MuR heteroduplexes (yellow) and subsequent resolution by an activity similar to a Holliday junction resolvase have been postulated[29]. *Mu* or host gene products known to be involved at particular steps are indicated.

by an *araC*^{c} plasmid. This, too, was initially thought to be an effect on derepression. However, it has been shown that Crp is not required for *Mu* derepression under starvation conditions, and the *B*$^{-}$*kil*$^{-}$ prophage mutations create a partial bypass of the *crp*$^{-}$ block to fusion formation (G. Maenhaut-Michel, M-J. Gama, A. Toussaint and J.A. Shapiro, unpublished). As the *B*$^{-}$ mutation prevents prophage replication, we hypothesize that some cAMP+Crp-dependent function prevents replication of the STC intermediate under starvation conditions. Because it is lethal, replication would pre-empt fusion formation from the STC intermediate (Fig. 3). Choosing between replicative and non-replicative STC processing is proposed to constitute a second decision point in the fusion process. The need for starvation-dependent events in at least two distinct steps of the fusion process would explain one of the most striking features of the *araB–lacZ* system, namely that fusions are never recovered after normal growth, but only after prolonged aerobic starvation.

How general are the conclusions drawn from the *araB–lacZ* system?

Mobile genetic elements like *Mu* are found in all organisms. It, thus, seems reasonable to hypothesize that the regulatory aspects of the mutational process exemplified by the *araB–lacZ* system might have some general applicability. Are there other examples of mutagenesis that illustrate physiological regulation?

One other adaptive mutation system has been the object of careful genetic and molecular scrutiny. Selection greatly stimulates pseudoreversion of a plasmid-borne *lac33* mutation, a *lacI–Z* fusion containing a +1 frameshift in the *lacI* portion of the hybrid coding sequence which prevents β-galactosidase expression[36,37]. The mutations stimulated by selection are a specific class of −1 frameshifts that restore the proper reading frame to encode an enzymatically active LacI–LacZ fusion peptide, and these frameshifts are different from the ones that occur during normal growth[38,39]. Genetic studies have revealed that adaptive *lac33* pseudoreversion requires homologous recombination[40,42] and F′*lac* conjugation (Tra) functions[43,44]. Mutations or culture conditions that

block F′*lac* transfer eliminate 90% of the adaptive class of frameshifts, and a chromosomal *lac33* allele shows at least a 50-fold decrease in adaptive mutation compared with a plasmid allele. The role of homologous recombination functions is particularly intriguing. Mutations in *recA*, *recBC* and *ruvAB* severely inhibit the adaptive class of −1 frameshifts, while *recD* and *recG* mutations increase them. The genetic results thus indicate an unexpectedly large number of molecular players, and multistep biochemical models have been proposed to explain the mechanisms that generate adaptive frameshifts[41,42,45].

Conjugation is a regulated process and, therefore, the discovery that conjugation functions are needed for selection-induced *lac33* pseudoreversion might be sufficient to explain the adaptive nature of the system. It has been found that prolonged aerobic incubation stimulates homosexual F′*lac* transfer[46,47] and leads to an increase in F′*lac* copy number (S. Benson, pers. commun.). The fundamental parallel with the *araB–lacZ* system is that both systems involve physiologically regulated systems that can be stimulated to action under stressful conditions characteristic of selection for novel carbon source utilization.

The depth of regulatory interactions between cellular signal transduction networks and natural genetic engineering systems is likely to prove typical rather than exceptional. For example, the λ integration/excision complex is controlled by the host RecA, IHF, Hfl and FIS proteins[48]. FIS, which is a growth-stage-specific regulatory factor, is also required in DNA inversion systems based on TnpR-like site-specific recombinases[49]. The DNA rearrangement activities of insertion sequences and transposons are known to depend on many different cellular regulatory functions, including IHF, DnaA, Dam methylation, programmed frameshifting, and differential utilization of internal promoters[17–19,50,51].

What implications do natural genetic engineering and adaptive mutation have for evolutionary theory?

Thinking about genetic change as a regulated biological function is fundamentally different from thinking about genetic change as the stochastic, accidental result of replication errors and physicochemical insults. The idea is that genomes can change because of the actions of built-in molecular machines subject to biological and environmental feedback. For theories of evolutionary variability, this unconventional viewpoint has at least two basic implications. One implication is that large-scale, coordinated changes within the genomes of single cells are mechanistically plausible because a particular natural genetic engineering system can be activated to operate at multiple sites in the genome. Such rapid restructurings in response to 'genome shock' have been described in maize[52,53], and they are a regular part of ciliate macronuclear development[54]. In sexually reproducing diploid organisms, genome-wide events can occur premeiotically in germ cell lineages that will ultimately produce many gametes sharing novel chromosome structures, as happens during hybrid dysgenesis in *Drosophila melanogaster*[55,56]. Thus, molecular genetics and developmental biology show how interbreeding populations with new genome architectures can emerge rapidly in evolution.

The second basic implication for evolutionary theory is the opportunity for adaptive feedback onto the process of genetic change. This feedback can be temporal and quantitative. The cellular ability to activate natural genetic engineering functions under stress can significantly accelerate evolutionary change in episodes of crisis without threatening genome stability under ordinary circumstances. Control of the timing and extent of genetic change, coupled with regulatory selection of the natural genetic system(s) to be activated, introduce important specificity elements into the mutational process. Despite occasional claims to the contrary, there is not yet clear evidence for environmental or cellular inputs targeting the mutations to particular places in the genome, although it might not be unreasonable to predict that further study of mobile genetic elements and DNA rearrangements will ultimately uncover true directed mutation. This prediction is based upon recent discoveries demonstrating that protein–protein interactions can couple the transcriptional apparatus to biochemical functions that are able to alter DNA sequence information. One case of transcriptional coupling involves repair of pre-mutagenic lesions[57], and another involves targeting the insertion of a yeast Ty retrotransposon to tRNA promoters by RNA polymerase III-associated transcription factors[58]. These observations serve as precedents for the possible coupling of natural genetic engineering functions to transcription complexes. Because signal transduction systems can direct the transcriptional apparatus to specific genomic locations in response to physiological inducers and signalling molecules, there exists a plausible molecular mechanism for similarly directing the mutational apparatus within the genome.

The experimental agenda for testing the natural genetic engineering concept will include a more careful investigation of how mutation frequencies are affected by different conditions (e.g. liquid versus surface growth, exponential versus stationary phase cultures, anaerobic versus aerobic incubation, saturating nutrients versus starvation regimes). It has been well-documented for many decades that culture ageing has a profound influence on the occurrence of highly pleiotropic changes in bacterial heredity[59], and these early studies need to be extended using molecular techniques. It will be important in these tests to choose systems that detect the formation of new DNA structures, such as gene activation by IS element transposition (Table 1). Such creative molecular events are likely to be better models for evolutionarily significant change than restoration of a previously evolved sequence by reversion of a point mutation. Once examples of significant physiological responses in mutagenesis have been identified (and once proper control experiments have excluded trivial explanations[60,61]), then it will be important to work out the molecular mechanism of DNA change in each system.

Knowing about mechanisms for multiple coordinated changes in the genome, and about the potential for biological feedback onto genome restructuring, forces us to think of evolutionary genetic change in cell biological terms, not as fundamentally different from other kinds of cellular biochemistry. Contemporary cell biology increasingly focuses on the operation of signal transduction networks, which are, in fact, molecular parallel computers of considerable sophistication[62,63]. We are just beginning to appreciate the power of cellular capacities for information processing. Because the genome is basically a

central information repository, we should be ready for some big surprises as we explore more deeply the notion that signal transduction networks can guide the abilities of cells to engineer their own DNA.

Acknowledgements

I thank my collaborators, G. Maenhaut-Michel, A. Toussaint, D. Leach, P. Higgins and M-J. Gamar for their stimulation and support on *Mu*-mediated DNA rearrangements, and also J.M. Gomez and J. Blazquez for sharing unpublished results. My research on the *araB–lacZ* system was supported by the National Science Foundation.

References

1 Reznikoff, W.S. (1992) *Mol. Microbiol.* 6, 2419–2422
2 Saier, M.H., Jr, Ramseier, T.M. and Reizer, J. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (2nd edn) (Neidhardt, F.C. *et al.*, eds), pp. 1325–1343, ASM Press
3 Pace, H.C., Lu, P. and Lewis, M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 1870–1873
4 Doolittle, R.F. (1995) *Annu. Rev. Biochem.* 64, 287–314
5 Gralla, J.D. and Collado-Vides, J. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (2nd edn) (Neidhardt, F.C. *et al.*, eds), pp. 1232–1245, ASM Press
6 Shapiro, J.A. (1991) *Genetica* 84, 3–4
7 Boyd, E.F., Hill, C.W., Rich, S.M. and Hartl, D.L. (1996) *Genetics* 143, 1091–1100
8 Shapiro, J.A. (1977) *Trends Biochem. Sci.* 2, 622–627
9 Shapiro, J.A. (1992) *Genetica* 86, 99–111
10 Rupp, W.D. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (2nd edn) (Neidhardt, F.C. *et al.*, eds), pp. 2277–2294, ASM Press
11 Walker, G. (1996) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (2nd edn) (Neidhardt, F.C. *et al.*, eds), pp. 1400–1416, ASM Press
12 Taddei, F., Matic, I. and Radman, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 11736–11740
13 Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (eds) (1977) *DNA Insertion Elements, Episomes and Plasmids*, Cold Spring Harbor Laboratory Press
14 Shapiro, J.A. (1995) in *Mobile Genetic Elements – Frontiers in Molecular Biology* (Sherratt, D.J., ed.), pp. 1–17, IRL Press
15 Watanabe, T. (1963) *Bacteriol. Rev.* 27, 87–115
16 Recchia, G.D. and Hall, R. (1995) *Microbiology* 141, 3015–3027
17 Sherratt, D. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 163–184, ASM Press
18 Berg, D.E. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 185–210, ASM Press
19 Kleckner, N. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 227–268, ASM Press
20 Ryan, F.J. (1955) *Genetics* 40, 726–738
21 Shapiro, J.A. (1984) *Mol. Gen. Genet.* 194, 79–90
22 Symonds, N., Toussaint, A., van de Putte, P. and Howe, M. (eds) (1987) *The Bacteriophage Mu*, Cold Spring Harbor Laboratory Press
23 Casadaban, M.J. (1976) *J. Mol. Biol.* 104, 541–555
24 Cairns, J., Overbaugh, J. and Miller, S. (1989) *Nature* 335, 142–145
25 Mittler, J. and Lenski, R.E. (1990) *Nature* 344, 173–175
26 Maenhaut-Michel, G. and Shapiro, J.A. (1994) *EMBO J.* 13, 5229–5239
27 Shapiro, J.A. (1987) in *The Bacteriophage Mu* (Symonds, N., Toussaint, A., van de Putte, P. and Howe, M., eds), pp. 251–258, Cold Spring Harbor Laboratory Press
28 Shapiro, J.A. and Leach, D. (1990) *Genetics* 126, 293–299
29 Maenhaut-Michel, G., Blake, C., Leach, D. and Shapiro, J.A. *Mol. Microbiol.* (in press)
30 Echols, H. (1986) *Science* 233, 1050–1053
31 Haniford, D.B. and Chaconas, G. (1992) *Curr. Opin. Genet. Dev.* 2, 698–704
32 Mizuuchi, K. and Craigie, R. (1986) *Annu. Rev. Genet.* 20, 385–429
33 Shapiro, J.A. (1993) *J. Bacteriol.* 175, 2625–2631
34 Geuskens, V., Mhammedi-Alaoui, A., Desmet, L. and Toussaint, A. (1992) *EMBO J.* 13, 5121–5127
35 Krause, H.M. and Higgins, N.P. (1986) *J. Biol. Chem.* 261, 3744–3752
36 Cairns, J. and Foster, P.L. (1991) *Genetics* 128, 695–701
37 Foster, P.L. and Cairns, J. (1992) *Genetics* 131, 783–789
38 Rosenberg, S.M., Longerich, S., Gee, P. and Harris, R.S. (1994) *Science* 265, 405–407
39 Foster, P.L. and Trimarchi, J.M. (1994) *Science* 265, 407–409
40 Harris, R.S., Longerich, S. and Rosenberg, S.M. (1994) *Science* 264, 258–260
41 Harris, R.S., Ross, K.J. and Rosenberg, S.M. (1996) *Genetics* 142, 681–691
42 Foster, P.L., Trimarchi, J.M. and Maurer, R. (1996) *Genetics* 142, 25–37
43 Galitski, T. and Roth, J.R. (1995) *Science* 268, 421–423
44 Radicella, J.P., Park, P.U. and Fox, M.S. (1995) *Science* 268, 418–420
45 Rosenberg, S.M. (1994) *Genome* 37, 893–899
46 Peters, J.E. and Benson, S.A. (1995) *J. Bacteriol.* 177, 847–849
47 Peters, J.E., Bartoszyk, I.M., Dheer, S. and Benson, S.A. (1996) *J. Bacteriol.* 178, 3037–3043
48 Landy, A. (1990) *Annu. Rev. Biochem.* 58, 913–949
49 van de Putte, P. and Goosen, N. (1992) *Trends Genet.* 8, 457–462
50 Galas, D.J. and Chandler, M. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 109–162, ASM Press
51 Chandler, M. and Fayet, O. (1993) *Mol. Microbiol.* 7, 497–503
52 McClintock, B. (1978) *Stadler Genetics Symp.* 10, 25–48
53 McClintock, B. (1984) *Science* 226, 792–801
54 Prescott, D.M. (1992) *Trends Genet.* 8, 439–445
55 Bregliano, J-C. and Kidwell, M.G. (1983) in *Mobile Genetic Elements* (Shapiro, J.A., ed.), pp. 363–410, Academic Press
56 Engels, W.R. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 437–484, ASM Press
57 Selby, C.P. and Sancar, A. (1994) *Microbiol. Rev.* 58, 317–329
58 Kirchner, J., Connolly, C.M. and Sandmeyer, S.B. (1995) *Science* 267, 1488–1491
59 Hadley, P. (1927) *J. Infect. Dis.* 40, 1–312
60 Lenski, R.E. and Mittler, J.E. (1993) *Science* 259, 188–194
61 Galitski, T. and Roth, J.R. (1996) *Genetics* 143, 645–659
62 Bray, D. (1990) *J. Theor. Biol.* 143, 215–231
63 Bray, D. (1995) *Nature* 376, 307–312
64 Shapiro, J.A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 1933–1937
65 Swanson, J. and Koomey, M.J. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 743–761, ASM Press
66 Deonier, R.C. (1987) in *Escherichia coli and Salmonella Cellular and Molecular Biology* (2nd edn) (Neidhardt, F.C. *et al.*, eds), pp. 982–989, ASM Press
67 Reynolds, A.E., Felton, J. and Wright, A. (1981) *Nature* 293, 625–629
68 Jaurin, B. and Normark, S. (1983) *Cell* 32, 809–816

J.A. Shapiro *is in the Department of Biochemistry and Molecular Biology, University of Chicago, 920 E. 58th Street, Chicago, IL 60637, USA.*