Molecular Microbiology (1997) **23**(6), 1133–1145

Different structures of selected and unselected *araB–lacZ* fusions

Geneviève Maenhaut-Michel,[1] Catherine E. Blake,[2†] David R. F. Leach[2] and James A. Shapiro[3*]
[1]*Université Libre de Bruxelles, Department of Molecular Biology, Laboratory of Prokaryotic Genetics, 67 rue des Chevaux, 1640 Rhode-St-Genése, Belgium.*
[2]*Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JR, UK.*
[3]*Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA.*

Summary

Formation of *araB–lacZ* coding-sequence fusions is a key adaptive mutation system. Eighty-four independent *araB–lacZ* fusions were sequenced. All fusions carried rearranged MuR linker sequences between the *araB* and *lacZ* domains indicating that they arose from the standard intermediate of the well-characterized Mu DNA rearrangement process, the strand transfer complex (STC). Five non-standard *araB–lacZ* fusions isolated after indirect sib selection had novel structures containing back-to-back inverted MuR linkers. The observation that different isolation procedures gave rise to standard and non-standard fusions indicates that cellular physiology can influence late steps in the multi-step biochemical sequence leading to *araB–lacZ* fusions. Each *araB–lacZ* fusion contained two novel DNA junctions. The MuR–*lacZ* junctions showed 'hot-spotting' according to established rules for Mu target selection. The *araB*–MuR and MuR–MuR junctions all involved exchanges at regions of short sequence homology. More extensive homology between MuR and *araB* sequences indicates potential STC isomerization into a resolvable four-way structure analogous to a Holliday junction. These results highlight the molecular complexity of *araB–lacZ* fusion formation, which may be thought of as a multi-step cell biological process rather than a unitary biochemical reaction.

Received 18 December, 1996; accepted 9 January, 1997. †Present address: Molecular Virology Laboratory, Department of Medical Microbiology, The Medical School, University of Edinburgh, Edinburgh EH8 9AG, UK. *For correspondence. E-mail jsha@midway.uchicago.edu; Tel. (773) 702 1625; Fax (773) 702 0439.

Introduction

Mu-mediated formation of *araB–lacZ* coding-sequence fusions was the first clearly documented example of what is now called 'adaptive mutation', i.e. mutations occurring more frequently when they are useful in overcoming a selective challenge (Shapiro, 1984). The idea that fusions were adaptive in this sense arose because they were initially observed to form only under selective conditions and were undetectable during normal growth of the prefusion culture (Shapiro, 1984; Cairns *et al.*, 1988). However, it was found that aerobic starvation stress in the absence of selective substrates was sufficient to induce the formation of *araB–lacZ* fusions (Mittler and Lenski, 1990; Maenhaut-Michel and Shapiro, 1994; Foster and Cairns, 1994; Sniegowski, 1995). Thus, the term 'adaptive' most properly refers to the physiological responsiveness of the fusion process (Shapiro, 1995). Each example of adaptive mutation has its own molecular characteristics. Consequently, it is important to work out the details of mechanism and regulation in several cases before attempting to draw general conclusions. The *araB–lacZ* fusion system illustrates how a complex series of biochemical events leading to a specific novel DNA structure can be influenced by different conditions, and we have argued why this complexity may be typical of adaptive mutation responses (Shapiro, 1997b). The results reported here extend our knowledge of the fusion process by indicating additional molecular intermediates and by suggesting that physiological influences operate at multiple steps.

Fusions creating a functional hybrid *araB–lacZ* cistron arose as a result of DNA rearrangements removing all blocks to transcription and translation between *araB* (interrupted by a Mu*cts62* insertion) and a decapitated *lacZYA* operon with no promoter and the *ochre U118* block to translation at codon 17 (Fig. 1; Casadaban, 1976). Fusion formation required the same gene products as Mu transposition (Shapiro and Leach, 1990). This observation, combined with sequence analysis of other Mu-mediated coding sequence fusions (Shapiro, 1987), led us to formulate a molecular mechanism for the fusion process in which a standard Mu strand transfer complex (STC) serves as an intermediate (Shapiro and Leach, 1990). The STC is the branched structure formed when the nucleoprotein complex (or 'transposasome') containing Mu termini, Mu A protein and the host HU protein attacks a target sequence and forms covalent single-strand linkages

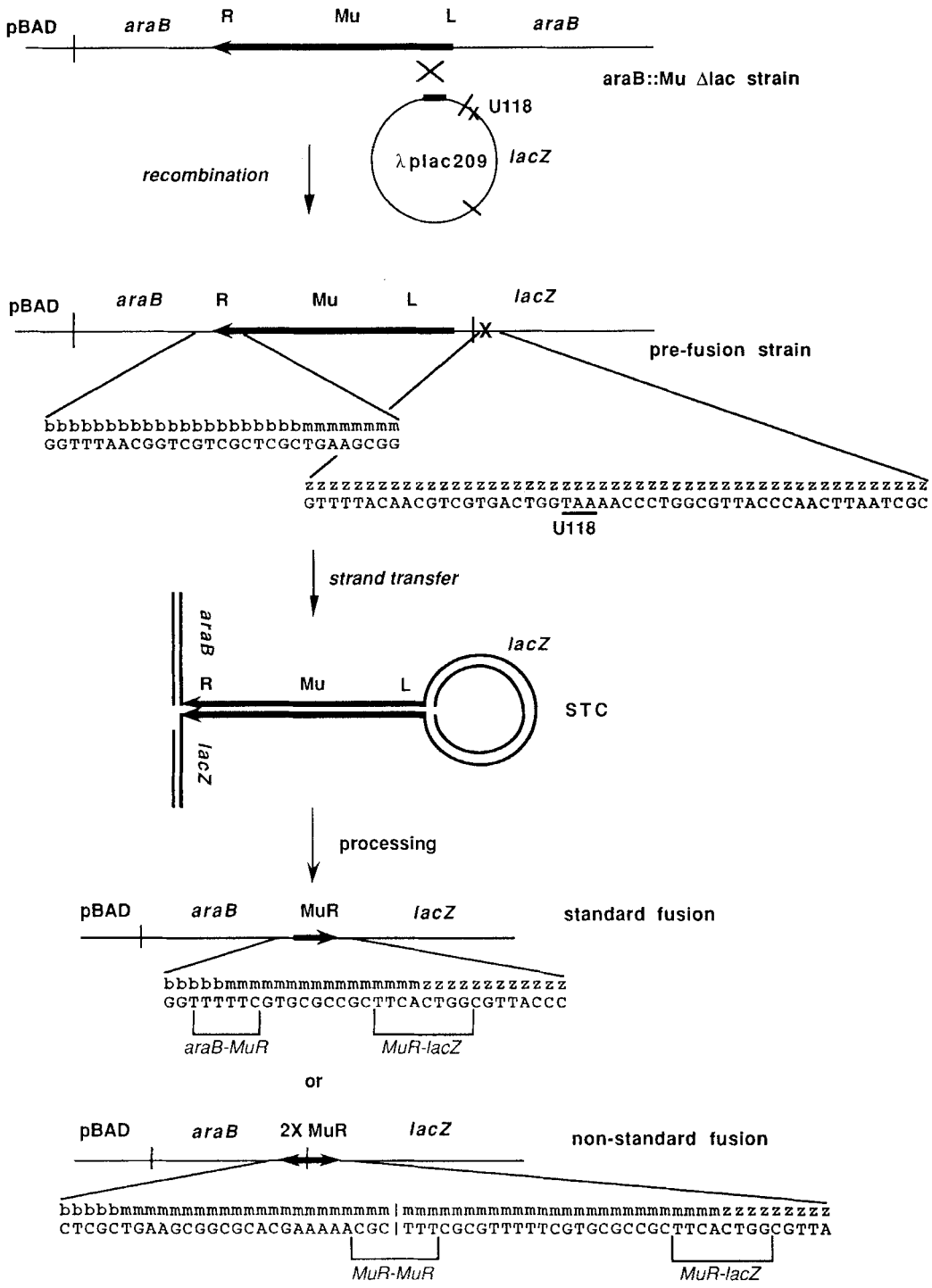


Fig. 1. Sequence components of the *araB–lacZ* fusion system. The series of steps leading from construction of the prefusion strain, in which an intact Mu prophage separates *araB* and *lacZ*, through generation of a strand transfer complex (STC) at a site downstream of the *lacZU118 ochre* mutation, to formation of standard and non-standard fusions is based on Casadaban (1976) and Shapiro and Leach (1990). We have expanded all of the relevant sequence junctions in the prefusion and fusion strains, using the most common junctions for illustration. The lowercase letters b, m or z above each nucleotide position indicate whether it corresponds to *araB* (b), MuR (m) or *lacZ* (z). Note that the *araB*–MuR junction in the prefusion strain and the MuR–*lacZ* junctions in the fusion strains were formed by the Mu transposasome independently of DNA sequence homology. The *araB*–MuR junctions in the standard fusions and MuR–MuR junctions in the non-standard fusion were formed by different processes at regions of short sequence homology.

between Mu and the target (Shapiro, 1979; Mizuuchi, 1992; Haniford and Chaconas, 1992). The initial steps of fusion formation in the model are identical to those for an adjacent inversion, one of the standard intramolecular Mu-mediated DNA rearrangements which is independent of Mu B protein and occurs at roughly the same frequency as Mu transposition to a new genomic location (Shapiro, 1979; Faelen and Toussaint, 1980). The parallel between the two classes of rearrangement is supported by experiments showing that prefusion strains carrying B^- mutant prophages have the same fusion kinetics at 32°C as the standard B^+ strain (our unpublished observations). The structure of a single sequenced *araB–lacZ* fusion agreed with the predictions of the model (Fig. 1), and polymerase chain reaction (PCR) analysis of fusions selected on arabinose plus lactose (AraLac) medium indicated that virtually all of them fell into a single structural class we called the 'standard fusion' (Maenhaut-Michel and Shapiro, 1994). However, it was a surprise result of the PCR analysis to discover that fusions isolated by indirect sib selection displayed a broader range of structures (generally larger), indicating a possible difference in fusion mechanism after liquid medium starvation in the absence of selective substrates (Maenhaut-Michel and Shapiro, 1994).

In order to investigate the possibility that novel fusion mechanisms were operating under non-selective stress or during AraLac selection of mutants strains, we sequenced the fusions arising under a variety of physiological and genetic conditions. All fusions contained rearranged MuR termini and had structures consistent with their derivation from an STC intermediate (Shapiro and Leach, 1990). The sib-selected non-standard fusions had two copies of MuR and retained the original *araB–MuR* junction (Fig. 1). This structure indicated a different process of DNA rearrangements occurring after glucose starvation in liquid medium as compared to what happens on the surface of AraLac selective plates. By examining the MuR–*araB* junctions in the standard fusions and the MuR–MuR junctions in the non-standard fusions, we were able to infer the existence of two different kinds of intermediates in the conversion of the STC to a complete duplex fusion.

Results

We sequenced *araB–lacZ* fusion junctions from a total of 86 different fusion clones. Twenty-eight of these clones were derived from the normal MCS2 parent (20 standard fusions selected on AraLac plates plus three standard and five non-standard fusions isolated by indirect sib-selection after glucose starvation in liquid medium). The other fusions were selected on AraLac plates from derivatives of MCS2 carrying various prophage or chromosomal mutations (Table 1). Mutant prefusion strains carrying the following lesions displayed dramatically reduced fusion yields: Mu*A*::miniTn*10* (Shapiro and Leach, 1990), *clpP*::*CM* (Shapiro, 1993), *clpX*::*KAN* (our unpublished results; cf. Mhammedi-Alaoui *et al.*, 1994), and *clpP*::*CM gidA2137*::miniTn*10* (our unpublished results). From these backgrounds, we picked rare fusions which arose late on the selected lawns. While we wanted to see if any fusions had novel structures indicating an alternative biochemical fusion mechanism, we knew that rare fusions could also arise from bypass of the ClpPX defects or from Mu*A*::miniTn*10* cells which had previously undergone a miniTn*10* excision. The kinetics of fusion colony appearance from Mu*A*::miniTn*10* cultures were consistent with such a two-step process. Only one or two colonies appeared on AraLac plates in the first three weeks of incubation, while control (MuA^+) cultures produced saturating levels of colonies (Fig. 2 of Shapiro and Leach, 1990). However, a number of fusions appeared in clusters around rare, early fusion colonies in MuA::miniTn*10* lawns during

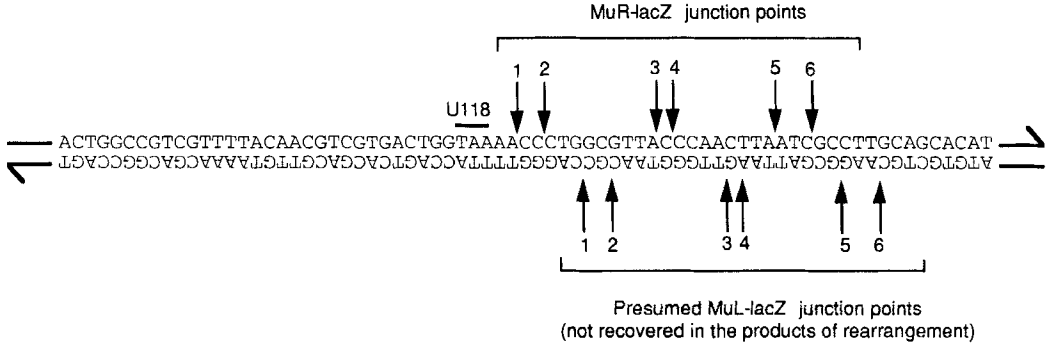


Fig. 2. Mu target sites in *lacZ*. Arrows above the duplex *lacZ* sequence indicate the six different junction positions where MuR is attached to *lacZ* in the 86 *araB–lacZ* fusions (Table 1). The Mu transposasome makes a five-nucleotide staggered break in the target sequence and ligates MuR and MuL to each target strand to generate the STC (Shapiro, 1979; Haniford and Chaconas, 1992; Mizuuchi, 1992). The positions of the MuR ligations correspond to the MuR–*lacZ* junctions recovered in the *araB–lacZ* fusions. The positions of the corresponding MuL ligations are indicated by the arrows below the duplex, even though these junctions are lost in processing of the STC intermediate to a fusion product (Shapiro and Leach, 1990).

Table 1. Sequences of novel junction regions in 86 *araB–lacZ* fusions.

A.

Number of isolates	Fusions isolated by direct selection on AraLac medium from MCS2-	Junction *araB*–MuR	Junction MuR–*lacZ*	Junction MuR–MuR
Total = 20	Wild type			
12	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTA	a	2	–
2	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATCGCCTTA	a	3	–
2	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCAGCCTTGCAGCACATC	a	6	–
2	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzzzzzzzz GGTTTAACGGTCGCCGCTTCACTGGCGTTACCCAAC	b	2	–
2	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzz GGTTTAACGGTCGTCGCTTCACCCAACTTA	c	3	–
Total = 5	Mu*A*::miniTn*10*			
2	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTT	a	2	–
1	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATCGCCT	a	3	–
1	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTAACGGTCGCCGCTTCACTGGCGTTACCCAACTT	b	2	–
1	bbbbmmmmmmmmmmmmmmmmmmzzzzzzzz GGTTTTTCGTGCGCCGCTTCAATCGCCTT	a	5	–
Total = 8	Mu*Bam* Δ7701::KAN			
8	bbbbbmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTT	a	2	–
Total = 10	*clpP*::CM			
7	bbbbmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTAA	a	2	–
1	bbbbmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATCGCCT	a	3	–
1	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzzzzzzzz GGTTTAACGGTCGCCGCTTCACTGGCGTTACCCAAC	b	2	–
1	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzzzzzzzzz GGTTTAACGGTCGCCGCTTCACCCAACTTAATCGCCT	b	3	–
Total = 10	*clpX*::KAN			
6	bbbbbmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTAA	a	2	–
3	bbbbbmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATCGCCT	a	3	–
1	bbbbbbbbbbbbbmmmmmmmmmzzzzzzzzzzzzzzz GGTTTAACGGTCGCCGCTTCACTGGCGTTACCCAAC	b	2	–
Total = 1	*clpP, gidA*2137::miniTn*10*			
1	bbbbbmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATCGCCT	a	3	–

the fourth week of incubation (our unpublished observations). Other mutant prefusion strains did not display significant quantitative effects on fusion kinetics but were examined to determine if the following prophage or chromosomal mutations influenced the fusion process in other detectable ways: B⁻KANΔ7701 prophage, *sbcCD*, *mutS*, or *sbcCD mutS*. SbcCD affects the replication of DNA palindromes (Chalker *et al.*, 1988), so *sbcCD* mutations could possibly affect the processing of the branched STC intermediate (Shapiro and Leach, 1990).

The 86 *araB–lacZ* fusion sequences are summarized in Table 1. Eighty-four of these fusions were of independent origin. In two cases, a pair of fusion clones was isolated from a single sib-selection (Maenhaut-Michel and Shapiro, 1994). Within these pairs, the two clones were identical and almost certainly arose as siblings rather than as

Table 1. Continued.

A.

Number of isolates	Fusions isolated by direct selection on AraLac medium from MCS2-	Junction *araB*–MuR	Junction MuR–*lacZ*	Junction MuR–MuR
Total = 4	Δ*sbcCD*::KAN			
3	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTA	a	2	–
1	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCAATCGCCTTGCAG**T**CACATC	a	5	–
Total = 15	*mutS*::miniTn*10*			
1	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTA	a	2	–
10	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCC**T**AACTTA	a	2	–
2	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACCCAACTTAATC**T**GA	a	3	–
2	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCAATCGCCTTGCAG**T**CACA	a	5	–
Total = 5	Δ*sbcCD*::KAN *mutS*::miniTn*10*			
4	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCACTGGCGTTACCC**T**AACTTA	a	2	–
1	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCAATCGCCTTGCAG**T**CACAT	a	5	–

B.

Number of isolates	Fusions isolated by sib-selection from cultures starved in MM+glucose			
Total = 8	Wild type			
3*	bbbbbmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzzzzz GGTTTTTCGTGCGCCGCTTCAATCGCCTTGCAGCACATCCCCC	a	5	–
1	bbbbbbbbbbbbbbbbbbbbbmmmmmmmmmmmmmmmmmmmmm**X** GGTTTAACGGTCGTCGCTCGCTGAAGCGGCGCACGAAAAACGC mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmzzzzzzzzzzzzzzzzzz TTTCGCGTTTTTCGTGCGCCGCTTCACTGGCGTTACCCAACTTA	Parental	2	X
2	bbbbbbbbbbbbbbbbbbbbbmmmmmmmmmmmmmmmmmmmm**Y** GGTTTAACGGTCGTCGCTCGCTGAAGCGGCGCACGAAAAACG mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm CGAAAGCTAAAGTTTTCGCATTTATCGTGAAACGCTTTCGCG mmmmmmmmmmmmmmmmmmmmmmzzzzzzzzzz TTTTTCGTGCGCCGCTTCACCCTGGCGTTA	Parental	1	Y
2*	bbbbbbbbbbbbbbbbbbbbbmmmmmmmmmmmmmmmmmmmm**Y** GGTTTAACGGTCGTCGCTCGCTGAAGCGGCGCACGAAAAACG mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm CGAAAGCTAAAGTTTTCGCATTTATCGTGAAACGCTTTCGCG mmmmmmmmmmmmmmmmmmmmmmzzzzzzzzz TTTTTCGTGCGCCGCTTCACCAACTTAA	Parental	4	Y

Fusions were obtained from the prefusion strain MCS2 and its derivatives (Table 3) as previously described (Shapiro, 1984; Maenhaut-Michel and Shapiro, 1994). The junctions were sequenced from PCR products as described in the *Experimental procedures*. The letters over the nucleotide designations indicate the presumed origin of each base pair from *araB* (b), MuR (m) or *lacZ* (z).The incorrect Ts in the *lacZ* portion of the *sbcCD* and *mutS* sequences are marked in bold face and are likely to be errors due to sequencing directly on the PCR products.
(*) These isolates contained two probable siblings obtained from the same sib selection.

independent isolates. We examined these pairs to verify that fusion structures remained constant during the sib selection procedure. All fusions contained rearranged MuR linkers joined to *lacZ* at six different sites labelled MuR–*lacZ* junctions 1 to 6 in the order 5′ to 3′ in the *lacZ* sequence (Fig. 2). As the MuR sequence was originally adjacent to *araB* in the prefusion strain (Fig. 1), the presence of a MuR–*lacZ* junction indicates the involvement of Mu-transposition functions in the formation of each fusion (Shapiro and Leach, 1990). Taking the five nucleotides adjacent to MuR in each junction as the target sequence for the Mu transposasome (Fig. 2), we find that the three most common junctions fit the established target consensus perfectly, and the other three targets differ at only a single position (Table 2; Mizuuchi and Mizuuchi, 1993).

Table 2. Specificity of *lacZ* target sequences among the 84 independent *araB–lacZ* fusions.

Target sequence	Junction no.	Position (*lacZ* nucleotide)	Frequency
NYG/CRN (=consensus)			
CCCT*G	1	57	2
CTGGC	2	59	59
CCCAA	3	68	13
CCA*AC	4	69	1
ATCGC	5	77	7
GCCT*T	6	80	2

Mu target sequences in *lacZ* were identified from the data in Table 1 as described in the legend to Fig. 2. The asterisks indicate deviations from the consensus established by Mizuuchi and Mizuuchi (1993).

Among the 80 independent standard fusions, there were only three different *araB*–MuR junctions labelled a, b and c (Table 1). As can be seen from aligning the junction sequences with the appropriately orientated *araB* and MuR parental sequences, the cross-over points occurred within regions of short homology (Fig. 3). The five non-standard fusions carried a second inverted copy of MuR arranged back-to-back with the *lacZ*-proximal MuR terminus. These secondary MuR copies all maintained the same *araB*–MuR junction as the prefusion strain in addition to the MuR–MuR junction (Table 1 and Fig. 1). There were two different MuR–MuR junctions, labelled x and y. Comparison of these junction sequences with MuR in both orientations showed that the cross-over points also

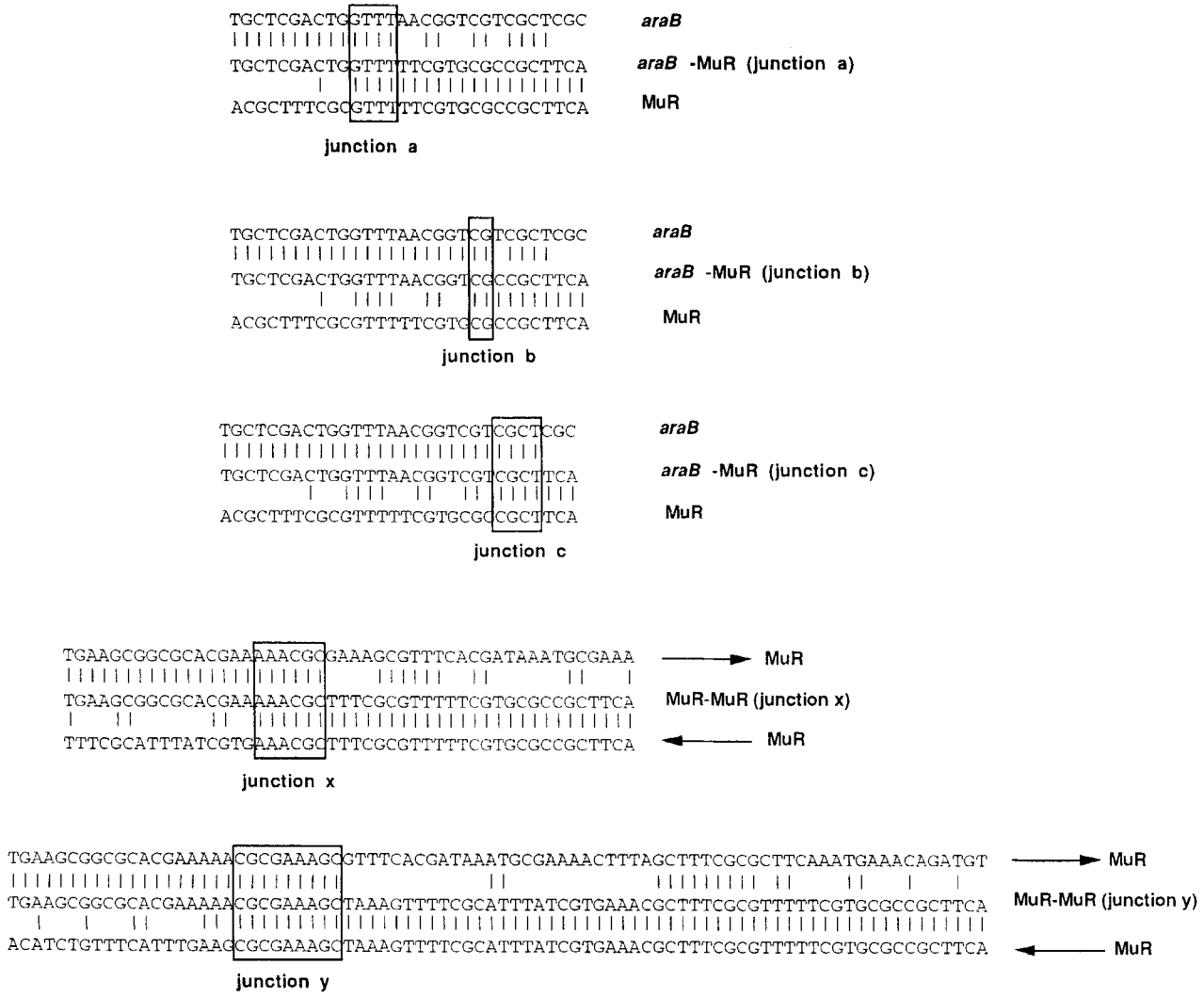


Fig. 3. Identification of the *araB*–MuR and MuR–MuR junction regions. Among the 86 fusions, there were only five different novel sequences upstream of *lacZ* (Table 1). These are represented by the middle line of each sequence alignment ending with TTCA, the last four nucleotides of MuR that were attached to different sites in *lacZ* (Table 1 and Fig. 2). The five novel fusion sequences were aligned at their 3′ ends with MuR (bottom line) and at their 5′ ends with either *araB* for the standard fusions or MuR in the opposite orientation for the non-standard fusions (top line). Sequence matches are indicated by vertical lines. The parental sources of segments in the fusions correspond to regions of continuous matches to *araB* and MuR. Genetic exchanges must have occurred within the boxed homologies where these continuous matches overlap. The positions of these homologies in the STC intermediate are shown in Figs 4D and 5.

occurred within regions of short homology (Fig. 3). The short homologies at the five junctions are similar to short oligonucleotide repeats found at other rearrangement hotspots (Albertini *et al.*, 1982; Allgood and Silhavy, 1991; Schofield *et al.*, 1992). As discussed below, the positions of these five repeats in the STC intermediate suggest mechanistic hypotheses for completion of the fusion process.

Discussion

Standard fusion structures from mutant cultures

All *araB–lacZ* fusions contained a MuR–*lacZ* junction fitting the established Mu target site preference, indicative of a role for Mu transposition functions and the STC intermediate (Tables 1 and 2; Shapiro and Leach, 1990; Mizuuchi and Mizuuchi, 1993). Thus, even in the mutant cultures deficient in MuA and ClpPX activity, it seems that there is only one biochemical mechanism for *araB–lacZ* fusion formation. In the ClpPX-deficient strains, the fusions may have arisen in cells where Mu*cts62* derepression occurred independently of Clp activity (Geuskens *et al.*, 1992; Shapiro, 1993) or where MuA protein was released from the STC independently of ClpX (Mhammedi-Alaoui *et al.*, 1994; Kruklitis *et al.*, 1996). The possibility of Clp-independent derepression is consistent with the observation that Mu*dlac* derepression in glucose-starved colonies can occur in *clpP::CM* strains if there is also a *gidA*::miniTn*10* mutation in the chromosome (Shapiro, 1997a). In addition, M.-J. Gama has observed Mu*cts62* derepression in stationary-phase *clpP*::CM cultures using *lacZ* connected to the Mu early promoter (personal communication). In the Mu*A*::miniTn*10* strains, the observation that the vast majority of colonies appeared very late after plating has led us to assume that fusion formation was preceded by miniTn*10* excision to restore MuA function. This assumption will be tested in the future by examining Mu*A* deletion mutants for their ability to produce *araB–lacZ* fusions. We saw no effects of *sbcCD* or *mutS* mutations on fusion kinetics or structure.

Specificity of Mu target selection in lacZ

The MuR–*lacZ* junctions were highly specific: only six different *lacZ* target sequences were found in 84 independent fusions, and there were only 4 different targets in the 78 fusions selected directly on AraLac plates (Table 1 and Fig. 2). At least two independent factors contributed to this 'hot-spotting': (i) the selection for growth on AraLac plates demanded in-frame fusions within a restricted sequence window of *lacZ*, and (ii) the Mu target site selection process is non-random. Of the six *lacZ* target sites, three accounted for 79/84 fusions (94%) and matched the NYG/CRN consensus defined previously for Mu target selection *in vitro* (Mizuuchi and Mizuuchi, 1993). Note that the previously published fusion sequence (Maenhaut-Michel and Shapiro, 1994) and 59/84 (70%) of the fusions presented here all shared a single *lacZ* consensus target – CTGGC at position 59 (Tables 1, 2 and Fig. 2). The three non-consensus *lacZ* targets were found in only five fusions, all but one of which was isolated by sib-selection. It is likely that broader target specificity resulted from relaxed demand for high β-galactosidase activity in the sib-selection procedure applied to glucose-starved liquid cultures when compared to selection from surface lawns for growth on AraLac medium. However, we cannot yet exclude a difference in the target selection process under these very distinct stress conditions.

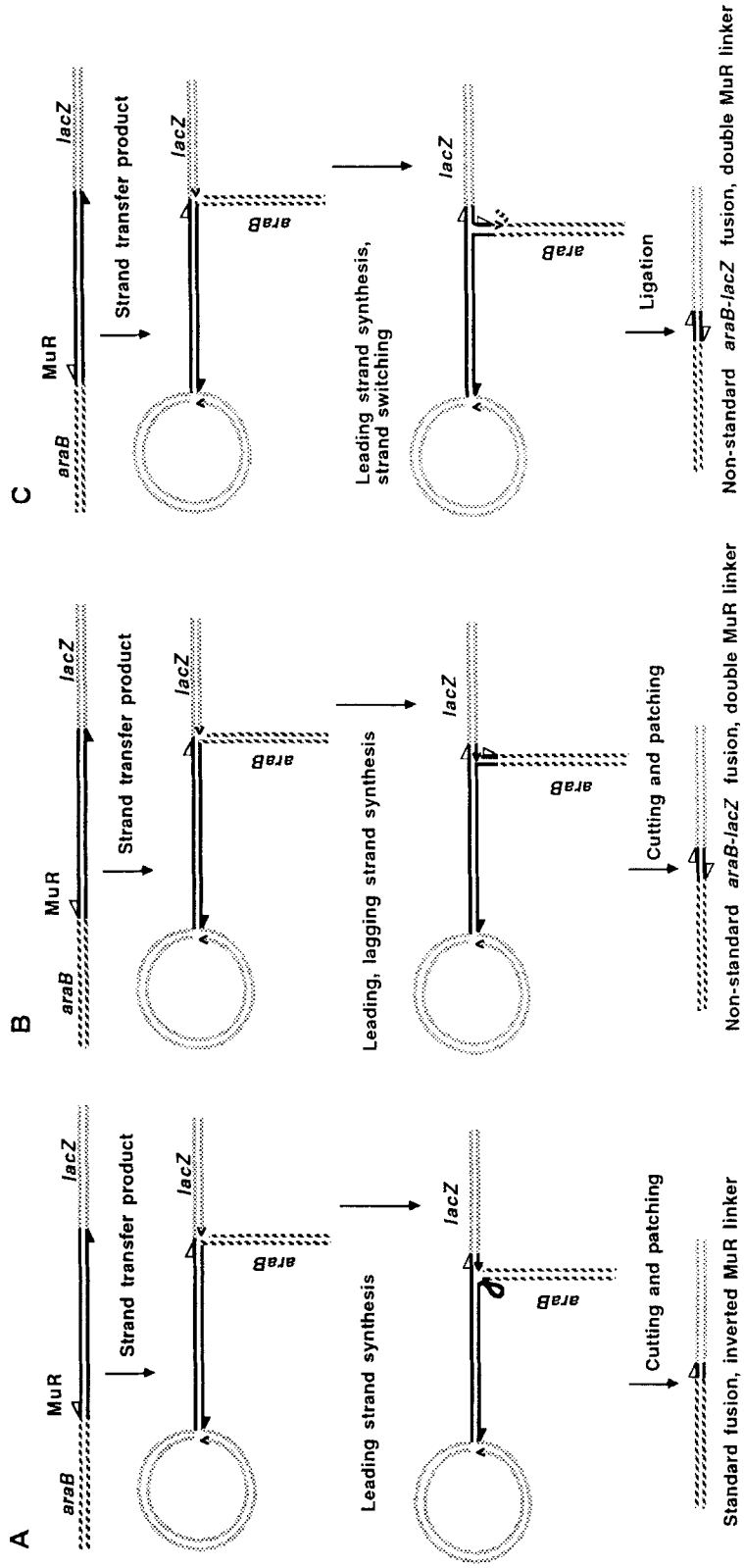
Non-standard fusion structures

Our most striking result was to find a major difference in the structure of non-standard fusions which could explain our previous PCR results (Maenhaut-Michel and Shapiro, 1994). This difference was the presence of a second copy of the MuR terminus in the linker region between the *araB* and *lacZ* domains of the fused coding sequence (Fig. 1, Table 1). The *araB*-proximal MuR sequence always maintained the prefusion strain *araB*–MuR junction sequence at position *araB*1236 (Fig. 1).

Like the standard fusion, the non-standard fusion structure appears to arise from a branched STC intermediate following limited DNA replication from the MuR terminus (Fig. 4; Shapiro and Leach, 1990). It should be recalled that the Mu strand transfer process forms a potential replication fork at each end of the Mu genome (Shapiro, 1979; Haniford and Chaconas, 1992; Mizuuchi, 1992). To account for the inverted copy of MuR in standard fusion structures, we postulated limited leading-strand synthesis at MuR followed by strand slippage or cutting and patching (Fig. 4A; Shapiro and Leach, 1990). To account for the back-to-back MuR copies and the maintenance of the original *araB*–MuR junction in non-standard fusion structures, we only need to invoke one of two mechanistically plausible variations of our original model: (i) leading and lagging strand synthesis followed by cutting and patching (Fig. 4B), or (ii) template strand switching at a repeated sequence, as is postulated to occur in DNA palindromes (Fig. 4, C and D) (Leach, 1994).

*Formation of MuR–*araB *and MuR–MuR junctions*

Our earlier molecular model for fusion formation did not address the mechanistic details of resolution of the branched STC intermediate (Shapiro and Leach, 1990). The sequence data provide some clues to that process in the form of the MuR–*araB* and MuR–MuR junctions in


A
B
C
araB
MuR
lacZ
araB
Strand transfer product
Leading strand synthesis
Cutting and patching
Standard fusion, inverted MuR linker
Leading, lagging strand synthesis
Non-standard araB-lacZ fusion, double MuR linker
Leading strand synthesis, strand switching
Ligation
Non-standard araB-lacZ fusion, double MuR linker

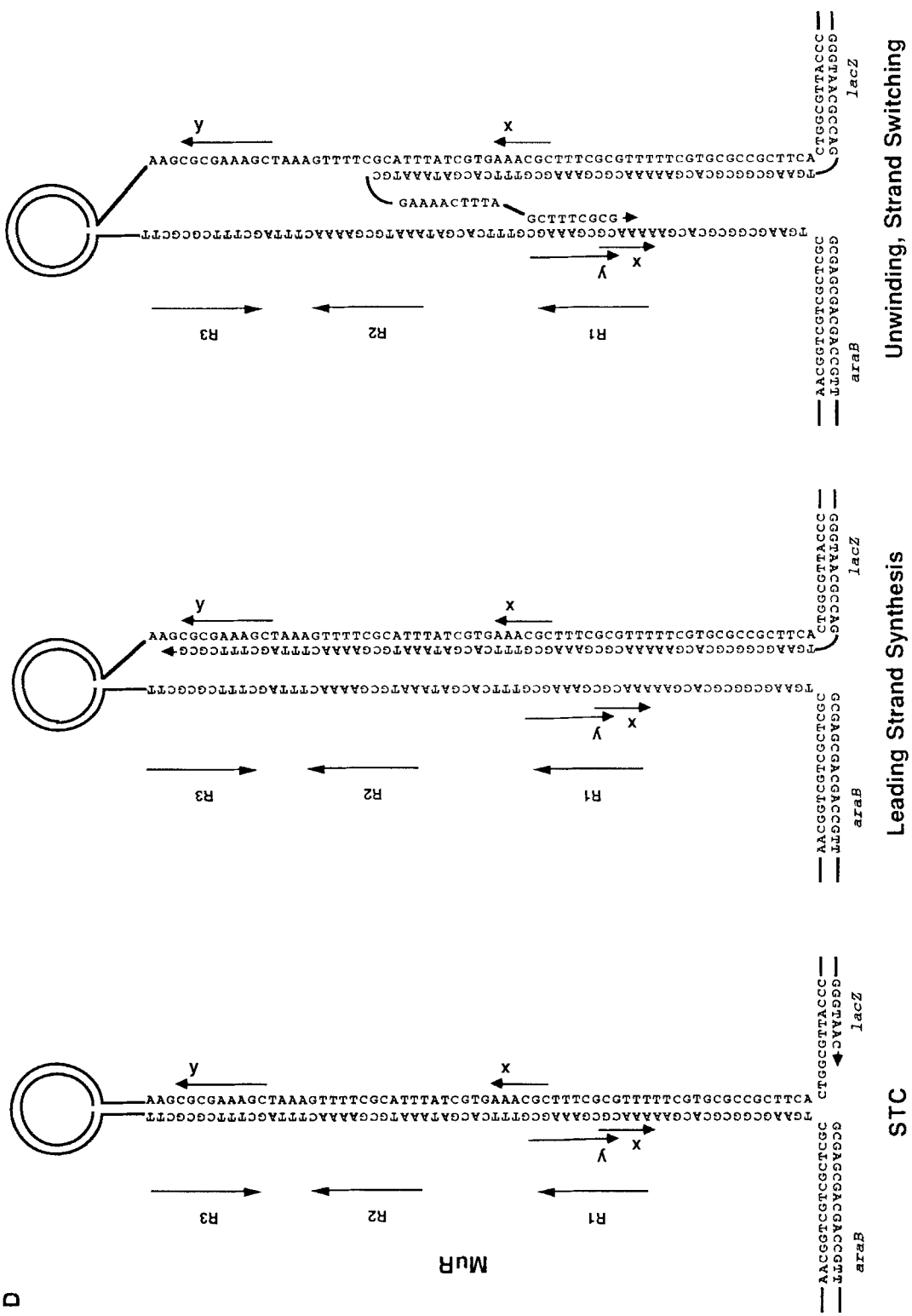


Fig. 4. The role of replication in processing of the STC intermediate.
A. Leading strand replication creating the substrates for standard fusion formation as postulated previously (Shapiro and Leach, 1990).
B. Leading and lagging strand synthesis creating the substrates for non-standard fusion formation by reciprocal exchange at homologies in newly replicated MuR sequences.
C. Leading strand synthesis and strand slippage creating a continuous MuR–MuR junction for non-standard fusion formation.
D. Detailed illustration of the possible role in strand slippage of the x and y inverted repeats. These repeats are located near the MuA-binding sites R1, R2 and R3 (Craigie *et al.*, 1984). Within the continuous strand of the STC, these inverted repeats are linked in direct orientation (as occurs in the branches of extruded palindromes) and can thus be used as donor and target sequences for strand slippage (Leach, 1994, 1996). The strand-slippage mechanism predicts that the MuR sequence attached to *lacZ* will be longer than that attached to *araB* (as observed) because leading strand synthesis from the free *lacZ* 3′ hydroxyl group (arrowhead) causes unwinding of the hairpin stem, uncovering an unpaired target sequence for strand switching closer to the base.

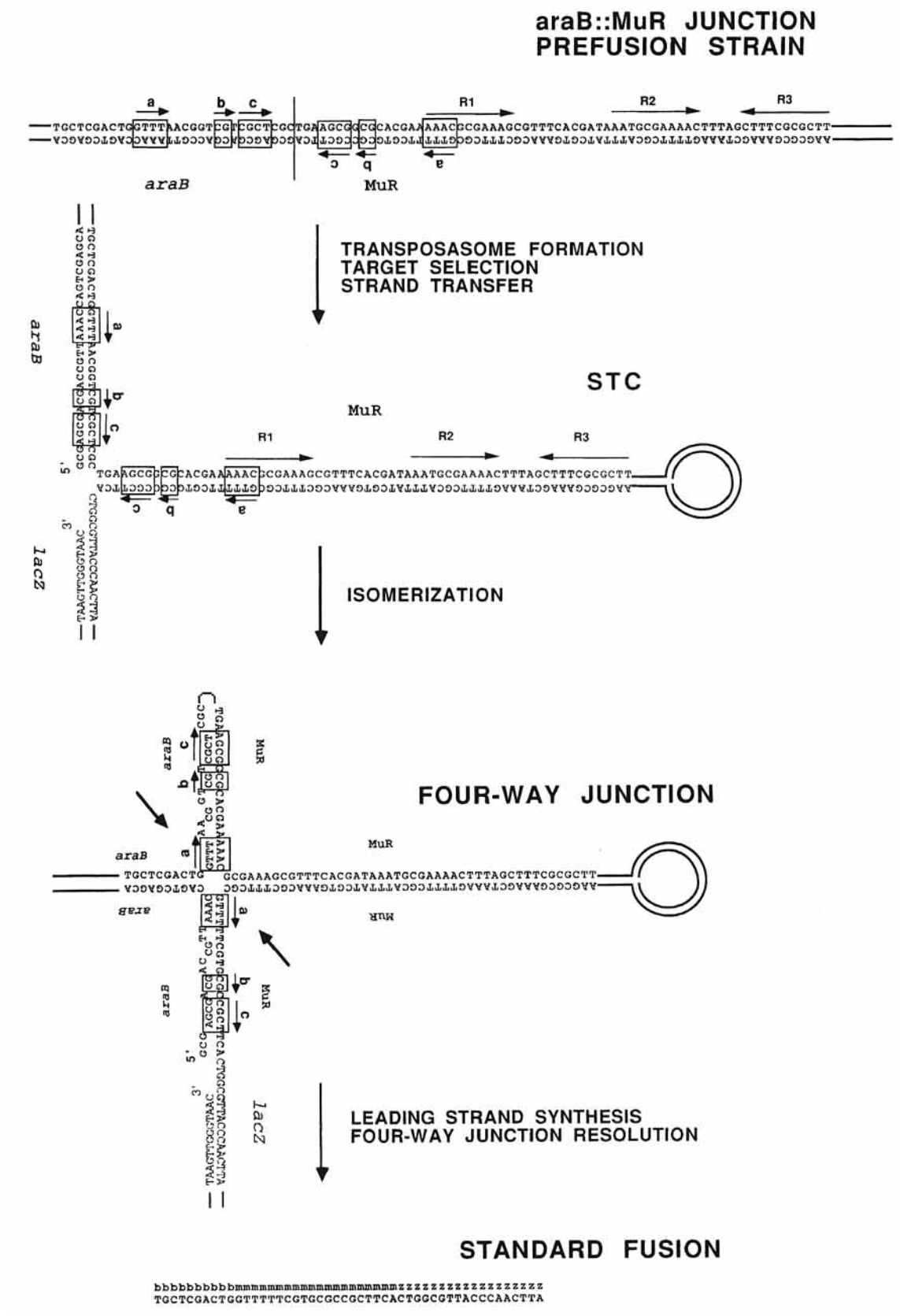
the standard and non-standard fusions (Table 1 and Fig. 3). Ninety per cent of the standard fusions share MuR–*araB* junction 'a' (Table 1, Figs 3 and 5). The cross-over point occurs within a 4 bp GTTT homology (Fig. 3). The other seven standard fusions have MuR–*araB* junctions of types 'b' or 'c' which have cross-over points within the directly repeated sequences CG and CGCT, respectively (Fig. 3). These short homologies may represent the reassociation targets in strand-slippage reactions (Leach, 1994) or may indicate that the cutting and patching process involves a staggered cleavage/religation reaction, such as the kind of two-site DNA gyrase reaction postulated to account for deletions and plasmid integrations in phage λ (Ikeda, 1994; Shapiro, 1985). However, the symmetrical positioning of the three cross-over points with respect to the original MuR–*araB* junction and the existence of significant inverted homology between the MuR and *araB* sequences suggest an alternative explanation for the origin of all the standard fusions. It is possible to envisage isomerization of the three-way branched STC to a four-way branched structure reminiscent of a Holliday junction (Fig. 5; Leach, 1994). This structure, when extruded to various extents, places the three cross-over points exactly at the branch points, where they could be cleaved and religated by a Holliday junction resolving nuclease such as RuvC and DNA ligase (Kowalczykowski *et al.*, 1994; West, 1994). Full extrusion of the structure places the most frequently used cross-over point 'a' at the junction point (Fig. 5). It will be interesting to test the effects of *ruvC* mutations to see whether they alter the frequency or structures of *araB–lacZ* fusions and to compare these effects with those observed in the *lac33* adaptive reversion system (Harris *et al.*, 1996; Foster *et al.*, 1996).

Among the five non-standard fusions, there are only two different MuR–MuR junction sequences. Junction x occurs within an inverted 6 bp homology of nucleotides 17–22 and 32–27, and junction y occurs within an inverted 9 bp homology at nucleotides 20–28 and 63–55 (Fig. 4D). Additional interrupted homologies flank the cross-over regions (Fig. 3). The homologies at positions 17–22 and 20–28 correspond to the R1 MuA protein binding site at nucleotides 17–27; the 32–27 homology is located within the region spanning the R1- and R2-binding sites; the 63–55 homology corresponds to the inverted R3 MuA protein-binding site at nucleotides 56–66 (Fig. 4D; Craigie *et al.*, 1984). While the structures of the two MuR–MuR fusion junctions are compatible with either of the two mechanisms postulated to explain the back-to-back structure (Fig. 4, B and C), it is worth noting that the non-standard fusion structures all fit the predictions of the strand-switching model proposed to explain partial deletions of extended DNA palindromes (Leach, 1996). This model predicts that the MuR arm connected to *lacZ* will always be longer than the arm connected to *araB* because strand switching occurs in the region already unwound by leading strand replication (Fig. 4, C and D), and this is exactly what we observe.

Significance of standard and non-standard fusions arising from different selection protocols

It is not obvious why non-standard fusion structures should only be detectable after sib-selection. Direct plating of glucose-starved cultures on AraLac medium produced an overwhelming majority of standard fusion structures. We considered the possibility that the sib selection protocol may require lower levels of hybrid *araB–lacZ* expression, but reconstruction experiments showed that many sib-selected non-standard fusions were sufficiently competitive in terms of their growth rates to have emerged from direct selection on plates (Maenhaut-Michel and Shapiro, 1994). The inverted MuR linkers can be expected to introduce inhibitory secondary structure into the mRNA, and we do find that many non-standard fusions have reduced (but still significant) levels of arabinose-induced β-galactosidase activity. The five non-standard fusions we sequenced (including two pairs of siblings) had induced levels of 839 and 727, 1333 and 1247, and 419 units compared to 1859 and 2019 units in the two standard fusions that were also detected by sib selection and sequenced (Maenhaut-Michel and Shapiro, 1994). Other non-standard fusions had induced β-galactosidase levels ranging from 306 to 3046 units (Maenhaut-Michel and Shapiro, 1994). With regard to the possibility of poor β-galactosidase expression from non-standard fusions, it should be noted that our earlier PCR data (Maenhaut-Michel and Shapiro, 1994) as well as the concordant sequencing of sibling clones reported here demonstrate that there is no change in the structure of non-standard fusions during growth on

Fig. 5. Mechanism for STC processing to standard fusions by a process similar to Holliday-junction resolution. The *araB*–MuR junction sequences a, b and c form part of an imperfect inverted repeat in the STC comprising MuR and the adjacent *araB* sequence. The symmetric disposition of these repeats suggests that the three-way STC junction may isomerize to a four-way junction as illustrated. Perfect homology at the base of the junction would allow it to fold into a stacked X configuration (Duckett *et al.*, 1988) allowing limited trans-junction DNA synthesis from the 3′ hydroxyl in *lacZ*. Such limited polymerization would favour stability of the four-way junction by removing mismatches in the arm attached to *lacZ* and ensure that no junctions could be resolved with *araB* sequences fused directly to *lacZ*. The long Mu arm and the short mismatched *araB*–MuR heteroduplex arm could then be removed by nuclease action, possibly by a Holliday-junction resolving enzyme that would cleave diagonally (as indicated by the large arrows), followed by ligation.


araB::MuR JUNCTION
PREFUSION STRAIN
a
b
c
R1
R2
R3
araB
MuR
TRANSPOSASOME FORMATION
TARGET SELECTION
STRAND TRANSFER
STC
ISOMERIZATION
FOUR-WAY JUNCTION
LEADING STRAND SYNTHESIS
FOUR-WAY JUNCTION RESOLUTION
STANDARD FUSION
lacZ

AraLac plates. Such changes would have been expected if strains carrying these DNA structures were disadvantaged on selective medium.

There are two key differences in the alternative selection procedures: one is that sib selection demands that fusions be completed in liquid culture, and the other is the presence of arabinose and lactose substrates in the direct selection on AraLac plates. It is therefore possible that some unknown physiological input either of surface growth or the presence of selective substrates (e.g. induction of *araB* transcription) could influence the post-STC biochemical events needed to complete construction of an active duplex *araB–lacZ* fusion. While it may seem unnecessarily complicated to postulate physiological regulation after STC formation, it is important to keep in mind that fusion formation is a complex cell-biological process of many steps, not a single biochemical reaction (Fig. 1). The biochemical complexity presents multiple potential targets for physiological regulation, and there is now evidence in the *araB–lacZ* system for physiological influences at more than one step of the DNA rearrangement process. We can demonstrate that Mu prophage derepression by itself is not a sufficient condition for fusion formation to occur and that one or more starvation-dependent events are required for derepressed prefusion cells to produce *araB–lacZ* clones (G. M.-M., M.-J. Gama, A. Toussaint and J. A. Shapiro, in preparation). Genetic results suggest that one of these events may be inhibition of Mu B function as a consequence of activating the cAMP catabolite response system. Inhibition of Mu B would prevent lethal prophage replication and facilitate processing of the STC intermediate to standard and non-standard fusions by the mechanisms indicated in Figs 4 and 5. The observation that non-standard fusions only appear under certain regimes is one piece of evidence showing that culture conditions can influence which DNA rearrangement mechanism is used. In this case, we postulate that strand-switching is favoured by extended starvation in liquid. These results fit the view that *Escherichia coli* possesses a wide repertoire of natural genetic engineering capabilities, that these capabilities can be combined in various sequences to produce particular DNA structures, and that they can be differentially activated in response to the many environmental and physiological situations encountered in bacterial life history (Shapiro, 1997b).

Experimental procedures

Bacterial strains and culture conditions

Microbiological and fusion selection methods have been described previously (Shapiro, 1984; Maenhaut-Michel and Shapiro, 1994). The bacterial strains from which *araB–lacZ* fusion clones were isolated are listed in Table 3.

Table 3. Bacterial strains.

Strain	Genotype	Reference
MCS2	F⁻ *thi* Δ(*lacIPOZYA-argF*)U169 *fla relA rpsLaraD139 araB*::+Mu*cts62*::λp209	Shapiro (1984)
MCS1330, MCS1366	MCS2 Mu*A2098*::miniTn*10*	Shapiro and Leach (1990)
MCS1731-34	MCS2 Δ*clpP*::CM	Shapiro (1993)
MCS1865	MCS1731 *gidA2137*::miniTn*10*	Shapiro (1997a)
MCS1888-98	MCS2 *clpX*::KAN	Mhammedi-Alaoui *et al.* (1994)
MCS1960-65	MCS2 Mu*Bam1066* Δ7701::KAN	Faelen (1987)
MCS1966-73	MCS2 Mu*Bam5176* Δ7701::KAN	Faelen (1987)
MCS1974-80	MCS2 Mu*Bam7154* Δ7701::KAN	Faelen (1987)
DL 943	MCS2 *sbcCD*::KAN	Blake (1996)
DL 944	MCS2 *mutS*::miniTn10	Blake (1996)
DL 945	MCS2 *sbcCD*::KAN *mutS*::miniTn*10*	Blake (1996)

These prefusion strains were all derived from MCS2 by P1 transduction and selection for the appropriate antibiotic-resistance marker.

DNA technology

DNA extractions and PCR amplifications of fusion junctions with primers *ara7* and *lac6* were performed as described previously (Maenhaut-Michel and Shapiro, 1994). With this pair of primers, DNA fragments between 0.5 and 0.65 kb are obtained on DNA containing *araB–lacZ* fusions (Maenhaut-Michel and Shapiro, 1994). PCR fragments were cloned in the Promega pGEM-T vector system. Fusion junctions were sequenced either directly off amplification products using the US Biochemical Sequenase PCR product sequencing kit or off cloned fragments by the Sanger method using *ara7* as the primer.

Acknowledgements

We are grateful to Marie-Jo Gama and Ariane Toussaint for discussions and for communicating unpublished results. We thank Pam Westfall and Rob Moore for their technical assistance. This research was supported, in part, by a grant from the National Science Foundation Microbial Genetics Programme to J.A.S. (NSF MCB-9220089). G.M.M. was supported by a research grant from the Belgian 'Fonds National de la Recherche Scientifique'. C.B. was supported by a studentship from the UK Medical Research Council.

References

Albertini, A.M. Hofer, M., Calos, M.P., and Miller, J.H. (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* **29:** 319–28.

Allgood, N.D., and Silhavy, T.J. (1991) *Escherichia coli xonA* (*sbcB*) mutants enhance illegitimate recombination. *Genetics* **127:** 671–680.

Blake, C. (1996) Illegitimate recombination in *Escherichia coli*. PhD Thesis, University of Edinburgh, Edinburgh, UK.
Cairns, J., Overbaugh, J., and Miller, S. (1988) The origin of mutants. *Nature* **335:** 142–145.
Casadaban, M.J. (1976) Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophages lambda and Mu.*J Mol Biol* **104:** 541–555.
Chalker, A.F., Leach, D.R., and Lloyd, R.G. (1988) *Escherichia coli sbcC* mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* **71:** 201–205.
Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984) Site-specific recognition of the bacteriophage Mu ends by the MuA protein. *Cell* **39:** 387–394.
Duckett, D.R., Murchie, A.I.H., Diekmann, S., von Kitzing, E., Kemper, B., and Lilley, D.M.J. (1988) The structure of the Holliday junction and its resolution. *Cell* **55:** 79–89.
Faelen, M. (1987) Appendix 3: useful Mu and mini-Mu derivatives. In *The Bacteriophage Mu*. Symonds, N., Toussaint, A., van de Putte, P., and Howe, M. (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 309–316.
Faelen, M., and Toussaint, A. (1980) Inversion induced by bacteriophage Mu-1 in the chromosome of *E. coli* K1. *Bacteriol* **142:** 391–399.
Foster, P.L., and Cairns, J. (1994) The occurrence of heritable Mu excisions in starving cells of *Escherichia coli*. *EMBO J* **13:** 5240–4.
Foster, P.L., Trimarchi, J.M., and Maurer, R. (1996) Two enzymes, both of which process recombination intermediates, have opposite effects on adaptive mutation in *Escherichia coli*. *Genetics* **142:** 25–37.
Geuskens, V., Mhammedi-Alaoui, A., Desmet, L., and Toussaint, A. (1992) Virulence in bacteriophage Mu: a case of trans-dominant proteolysis by the *E. coli* Clp serine protease. *EMBO J* **13:** 5121–5127.
Haniford, D.B., and Chaconas, G. (1992) Mechanistic aspects of DNA transposition. *Curr Opin Genet Devel* **2:** 698–704.
Harris, R.S., Ross, K.J., and Rosenberg, S.M. (1996) Opposing roles of the Holliday junction processing systems of *Escherichia coli* in recombination-dependent adaptive mutation. *Genetics* **142:** 681–691.
Ikeda, H. (1994) DNA topoisomerase-mediated illegitimate recombination. *Adv Pharmacol* **29:** 147–165.
Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* **58:** 401–465.
Kruklitis, R., Welty, D.J., and Nakai, H. (1996) ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *EMBO J* **15:** 935–944.
Leach, D.R.F. (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays* **16:** 893–900.
Leach, D.R.F. (1996) Cloning and characterization of DNAs with palindromic sequences. In *Genetic Engineering*. Vol. 18. Setlow, J.K. (ed.). New York: Plenum Press New York, pp. 1–11.
Maenhaut-Michel, G., and Shapiro, J.A. (1994) The roles of selection and starvation in the emergence of *araB–lacZ* fusion clones. *EMBO J* **13:** 5229–5239.
Mhammedi-Alaoui, A., Pato, M., Gama, M.-J., and Toussaint, A. (1994) A new component of bacteriophage Mu replicative transposition machinery: the *Escherichia coli* ClpX protein. *Mol Microbiol* **11:** 1109–16.
Miller, J.H. (1992) *A Short Course in Bacterial Genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
Mittler, J., and Lenski, R.E. (1990) Further experiments on excisions of Mu from Escherichia coli MCS2 cast doubt on directed mutation hypothesis. *Nature* **344:** 173–175.
Mizuuchi, K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu Rev Biochem* **61:** 1011–1051.
Mizuuchi, M., and Mizuuchi, K. (1993) Target site selection in transposition of phage Mu. *Cold Spring Harb Symp Quant Biol* **58:** 515–23.
Schofield, M.A., Agbunag, R., and Miller, J.H. (1992) DNA inversions.between short inverted repeats in *Escherichia coli*. *Genetics* **132:** 295–302.
Shapiro, J.A. (1979) A molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci USA* **76:** 1933–1937.
Shapiro, J.A. (1984) Observations on the formation of clones containing *araB–lacZ* cistron fusions. *Mol Gen Genet* **194:** 79–90.
Shapiro, J.A. (1985) Mechanisms of DNA reorganization in bacteria. *Int Rev Cytol* **93:** 25–56.
Shapiro, J.A. (1987) Some lessons of phage Mu. In *The Bacteriophage Mu*. Symonds, N., Toussaint, A., van de Putte, P., and Howe, M. (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 251–258.
Shapiro, J.A. (1993) A role for the Clp protease in activating Mu-mediated DNA rearrangements. *J Bacteriol* **175:** 2625–2631.
Shapiro, J.A. (1995) Adaptive mutation: who's really in the garden? *Science* **268:** 373–4.
Shapiro, J.A. (1997a) Multicellularity is the rule, not the exception: lessons from *E. coli* colonies. In *Bacteria as Multicellular Organisms*. Shapiro, J.A., and Dworkin, M. (eds). Oxford: Oxford University Press, pp. 14–49.
Shapiro, J.A. (1997b) Genome organization, natural genetic engineering, and adaptive mutation. *Trends Genet*, in press.
Shapiro, J.A., and Leach, D.R.F. (1990) Action of a transposable element in coding sequence fusions. *Genetics* **126:** 293–299.
Sniegowski, P.D. (1995) A test of the directed mutation hypothesis in *Escherichia coli* MCS2 using replica plating. *J Bacteriol* **177:** 1119–20.
West, S.C. (1994) The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. *Cell* **76:** 9–15.