Characterization of the recombinant joints formed by single-strand annealing reactions in vaccinia virus-infected cells

Xiao-Dan Yao and David H. Evans*

Department of Molecular Biology and Genetics, The University of Guelph, Guelph, Ontario, N1G 2W1 Canada

Received 29 July 2002; accepted 8 November 2002

Abstract

Poxviruses appear to use single-strand annealing reactions to recombine linear molecules sharing short (<20 bp) regions of end homology. We have examined the effect of base mismatches and base insertions on the reaction efficiency and used mismatch-containing DNAs to further characterize the polarity of the exonuclease postulated to catalyze these reactions in vivo. Incorporating one or two base substitutions within the 20-bp segment of end homology had little effect on virus-promoted recombination, reducing the frequency of recombinational repair of transfected plasmids only 10–20%. Base insertions were more destabilizing and their presence inhibited recombination 40% (with one insertion) and 75% (with two). The sequence of the recombinants recovered from virus-infected and transfected cells suggested that hybrid DNA is usually formed and then resolved by replication without repair. However, a few of the joints retained sequences suggestive of more complex enzymatic processing in vivo. We also used transfection studies to examine the fate of each of the four strands processed by the vaccinia recombination machinery. The preferential retention of base substitutions located near each of the 5'-ended strands confirmed that virus single-strand annealing reactions are catalyzed primarily by a 3'- to 5'-exonuclease. Other studies showed that mismatch repair reactions do not invalidate these conclusions, even though base excision repair systems are seemingly active and preferentially convert T·G and C·A mismatches to CG base pairs in vaccinia-infected cells.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Vaccinia virus; Mismatch repair; Single-strand annealing; Recombination; DNA polymerase; GT glycosylase, Exonuclease

Introduction

The Poxviridae comprise a family of large double-stranded DNA viruses that replicate in the cytoplasm of infected cells (Moss, 1996). The prototype, vaccinia virus, encodes ~260 proteins within a 190-kbp genome (Goebel et al., 1990) and this genome is thought to encode most, if not all, of the gene products required for viral replication, repair, and recombination. A characteristic feature of poxvirus-infected cells is that the viral replicative machinery can function in trans to replicate and recombine any DNA-transfected into virus-infected cells (DeLange and McFadden, 1986). We, and others, have exploited this biological feature to show that replication and recombination are intimately linked processes that promote extraordinarily high frequencies of genetic recombination.

The production of recombinant molecules in virus-infected cells is linked to the formation and disappearance of abundant heteroduplex DNA (Fisher et al., 1991), and we have suggested that single-strand annealing (SSA) recombination models can explain this and other features of poxvirus recombination reactions (Yao and Evans, 2001). In a SSA recombination scheme, recombinants are created by enzyme-catalyzed or spontaneous annealing of single strands. These molecules could be displaced during DNA replication, generated by DNA helicases, or produced by exonuclease processing of double-stranded breaks. Exonuclease processing permits recombinant formation if the exposed single-stranded ends share complementary sequence homology, and can be modeled in vitro by incubating linear DNA molecules with purified vaccinia virus DNA polymerase (Willer et al., 2000). This in vitro reaction depends upon

* Corresponding author. Fax: +1-519-837-2075.
E-mail address: dhevans@uoguelph.ca (D.H. Evans).
the polymerase-encoded 3′- to 5′-proofreading exonuclease activity to expose complementary single-stranded sequences in the duplex termini, which then anneal to form noncovalently linked recombinants. Genetic studies have long suggested that the DNA polymerase plays some role in promoting poxvirus recombination (Merchlinsky, 1989; Willer et al., 1999) and there are many similarities between the in vitro reaction and the in vivo behavior of transfected DNA (Willer et al., 2000). In particular, analysis of the DNA recovered from transfected cells showed that the recombinant junctions were predominantly processed by a 3′- to 5′-exonuclease as would be predicted for a reaction dependent upon the activity of the proofreading exonuclease (Yao and Evans, 2001). Collectively these data suggest that poxviral DNA polymerases may well function as viral “recombinases” in vivo.

Several features of these SSA processes remain to be more firmly characterized, particularly the effects of reduced homology on the efficiency of this reaction in vivo. This is an important parameter, since this reaction feature determines the fidelity of virus recombination and double-strand break repair reactions and might ultimately affect the evolution of poxvirus genomes (Yao and Evans, 2001). In this paper, we show that viral SSA reactions exhibit the capacity to accommodate significant amounts of imperfect homology between the ends of recombining molecules. We also show that some of the hybrid molecules used in these studies are subjected to a process resembling mismatch repair and that, when this phenomenon is taken into account, the case for a 3′- to 5′-exonuclease playing a role in catalyzing poxvirus SSA reactions becomes highly compelling.

Results and discussion

Effect of sequence heterology upon the efficiency of recombination in vaccinia-infected cells

Previous experiments have shown that only 14–18 bp of perfect end-sequence homology still suffices to permit recombination between linear DNAs in vaccinia-infected cells (Yao and Evans, 2001). What effect do small sequence differences have on the efficiency of these reactions, if such sequence differences were incorporated into these short regions of homology? To examine this question we used the PCR and a series of oligonucleotide primers to create a set of linear substrates encoding the middle portion of the luciferase open reading frame (Fig. 1A). These substrates shared approximately 20 bp of sequence overlap, on either side of a gap in the luciferase gene created by digesting it with BstEII and PacI in plasmid pRP406. Recombination between the two DNAs will repair the double-stranded break and the presence of a reconstructed luciferase gene can be measured in virus-infected cells because a P11 promoter regulates its expression.

Interestingly, incorporating base substitutions or base insertions into the homologous portions of these recombination substrates had only a rather limited effect upon the efficiency of recombination in vivo. Introducing one (LucM20-S1) or two (LucM20-S2) genetically silent base substitutions into the left end of the linear insert (Fig. 1), inhibited recombination only 10–20% in transfected cells (Fig. 2, upper panel). Introducing one (LucM20-I1) or two (LucM20-I2) single-base insertions into the right end of the linear insert (Fig. 1) had a seemingly greater effect on recombination frequency with two single-base insertions inhibiting recombination more than did a single-base insert (Fig. 2). However, it should be noted that these insertions create frameshift mutations that reduce the luciferase activity of either mutant recombinant polypeptide >95% (data not shown) and this correction should be taken into account if the data are to be compared between substrates (Fig. 2, dotted lines). If we assume that approximately equal numbers of mutant and wild-type genes are reconstructed by recombination (see below) and that only one of the two polypeptides is catalytically genes, then the presence of one and two base insertion reduces recombination frequencies about 40 and 75%, respectively, relative to the perfectly homologous LucM20 substrate.

To explore the physical basis of this phenomenon we incubated the same substrates with purified vaccinia DNA polymerase and single-strand DNA-binding protein, in an in vitro reaction that replicates features of in vivo recombination reactions (Willer et al., 2000). In most cases the efficiency of joint molecule formation was not greatly affected by the presence of sequence heterologies, as judged by the relative yield of recombinant molecules comigrating with nicked circular joint molecules on agarose gels (Fig. 2, lower panel). However, there was a small (~2-fold) reduction in the yield of this species when the LucM20-I2 substrate was used and this is consistent with the observation that this was the least efficient of the five different insert substrates in vivo (Fig. 2, lane 5). Any bulge loops that might be formed on hybridizing the ends of vector and insert DNAs would be significantly more destabilizing than would simple base mismatches (ΔG° ≈ +3 kcal mol⁻¹ versus about 0 kcal mol⁻¹ (Cantor and Schimmel, 1980)). Consequently joint molecules incorporating LucM20-I2, and thus incorporating two such loops, are predicted to be the least thermodynamically stable of the five possible joint molecules.

Sequence analysis of joint molecules

We used DNA sequencing to investigate what happens to these molecules when base substitutions and insertions interrupt a region of overlapping sequence homology. Cells were cotransfected with one of each of the five LucM20 inserts, plus BstEII- and PacI-cut pRP406, and the concatamers formed by viral replication and recombination reactions recovered from infected cells at 24 h posttransfection.
This DNA was digested with DpnI to degrade any input DNA, then converted into monomeric circles using SacI and T4 DNA ligase, and used to transform a mismatch-repair-deficient strain of E. coli to ampicillin resistance. The results of these analyses are shown in Table 1. Table 1 also shows the predicted structures of the intermediate joint molecules. These structures are drawn on the assumption that DNA ends are processed by a 3'- to 5'-exonuclease both in vivo and in vitro. This hypothesis is discussed in more detail below.

Analysis of the DNA sequence spanning the junction between vector and insert molecules, in plasmids assembled through recombination in vivo, showed that most of the molecules encoded only the sequences expected of simple
segregation products. Plasmids incorporating all of the eight possible recombinant junctions were recovered from transfected and virus-infected cells, and these plasmids encoded approximately equal numbers of vector- or insert-derived junction sequences. The simplest way to explain these in vivo observations is to suppose that if hybrid junctions are stable enough to form covalently joined recombinants in virus-infected cells, subsequent enzymatic processing mostly involves additional rounds of replication without further DNA repair. Still, two recombinant molecules were recovered whose structure cannot be explained by such a simple recombination and replication scheme.

The first of these molecules was formed in vivo in cells transfected with the LUCM20-I insert plus BstEII/PacI-cut pRP406 vector and consisted of a junction sequence encoding information derived from both vector and insert sequences (Table 1, sequence -T--T-A--T-). In this molecule, the “cross-over” site was formally located between the two base substitutions. Such molecules could not arise by a mechanism involving the formation of a hybrid joint spanning both base substitutions (Fig. 3, molecule 1A), followed by resolution of this intermediate through replication. Instead, replication and repair systems may be processing intermediates that are more complex, in which hybrid DNA perhaps spans only one of the two base substitutions (Figs. 3, 1B, and 1C). Alternatively, something resembling “very short patch” mismatch repair systems could target molecules like “1A” and, through gene conversion, create the appearance of crossing-over between very close markers. The possible impact of mismatch repair systems on recombinant structure is investigated in a later section.

The second of these molecules was formed in vivo in cells transfected with the LUCM20-I insert plus BstEII/ PacI-cut pRP406 vector and consisted of a junction sequence encoding information derived from neither vector or insert sequences (Table 1, sequence -CAA--GTT-). This molecule would presumably derive from processing of a hybrid intermediate encoding a bulged or extrahelical cytosine (Fig. 3, molecule 2A), but the exact process by which it might arise is obscure. Poxvirus replication systems are prone to making frameshift errors in sequence repeats (Weir and Moss, 1983) and it is notable that the extrahelical base is located next to a trio of thymidines. A slipped replication error, of the type first suggested by Streisinger et al. (1966) and perhaps triggered by the mismatched base, would most readily explain the addition of an adenosine opposite this thymidine-rich template and thus the structure of this particular recombination product.

**Evidence for 3’- to 5’-exonuclease processing of recombinants in vivo**

The joint molecules sketched in Table 1 and in Fig. 3 presuppose that hybrid molecules are formed in vivo by the action of 3’- to 5’-exonucleases. biochemical studies leave

<table>
<thead>
<tr>
<th>Insert</th>
<th>Hybrid structure</th>
<th>Sequence recovered (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LucM20-S1</td>
<td>-a-</td>
<td>-A-</td>
</tr>
<tr>
<td>LucM20-S2</td>
<td>-a--t-</td>
<td>-A--T-</td>
</tr>
<tr>
<td>LucM20-I1</td>
<td>-C A-</td>
<td>-CA-</td>
</tr>
<tr>
<td>LucM20-I2</td>
<td>-A TAC A-</td>
<td>-ATACA-</td>
</tr>
</tbody>
</table>

* The hybrid structure assumes that joint molecules arise through processing by a 3’- to 5’-exonuclease. Insert- and vector-derived bases are shown in lower case and capitals, respectively.

| N, the number of independently isolated plasmids with the given sequence. |
little doubt that the in vitro strand-joining reactions shown in Fig. 2 are catalyzed by the 3′- to 5′-proof-reading exonuclease activity of vaccinia DNA polymerase (Willer et al., 2000). However, the in vivo evidence is more indirect and is based upon our previous transfection studies involving a single set of mismatch-tagged DNA substrates (Yao and Evans, 2001). These substrates encoded base mismatches in the middle of the regions of overlapping sequence homology and, because the four base substitutions differentiate the four strands involved in strand joining, allowed the unambiguous identification of the origin of a strand after it has been replicated and recombined. These studies led to the conclusion that 75% of recombinant junction sequences were created in vivo, by reactions involving a 3′- to 5′-exonuclease.

A confounding problem with these experiments is mismatch repair, which could seem to bias the recovery of particular strands if such systems are also active in virus-infected cells. We were concerned that such activities might have compromised our earlier investigations and therefore constructed a second set of substrates that would address this concern. These new substrates were identical to the previous set in encoding G·T, C·A, C·T, and G·A mismatches, except that the oligonucleotides ligated onto each end of the two DNA substrates were designed to transverse the base mismatches across the helix (Fig. 1B).

The underlying logic was that if such a rearrangement altered the putative reaction polarity, then something such as mismatch repair (and not end-directed-processing reactions) was mostly influencing the recovery of particular strands.

These new sequence-tagged DNAs were transfected into vaccinia-infected cells, and recombinant molecules were then recovered in Escherichia coli and sequenced across the two junctions in each plasmid. The results of these investigations are presented in Table 2. To facilitate comparison we also present the data acquired previously using the original sequence-tagged substrates (Yao and Evans, 2001). Producing recombinants by this strategy is difficult because the oligonucleotide-tagged substrates are hard to prepare in large quantities and thus the yield of recombinants is low. However, the sequencing studies revealed that among 24 sequence-tagged junctions, in the 12 recombinant plasmids we were able to recover >90% of the junction sequences retained the base characteristic of processing by a 3′- to 5′-excision reaction (Table 2). Separately or collectively the preferential loss of bases from the 3′-ended strand is highly significant since a χ² test of the hypothesis that there is no bias can be rejected with a probability of \( P < 0.01 \). Thus, inverting the orientation of the base mismatches does not alter the apparent reaction polarity and this suggested that mismatch repair reactions are not major factors influencing...
the outcome of these experiments. If anything, these studies further reinforced the conclusion that the homologous ends of linear DNAs are preferentially excised during poxviral recombination in vivo by an exonuclease with 3′- to 5′- polarity.

Mismatch repair in vaccinia virus-infected cells

If the majority of the recombinants recovered from poxvirus-infected cells seem to have been subjected to 3′- to 5′- exonucleolytic processing reactions, the question arises as to what might be the process giving rise to the remaining minority of molecules? We wondered whether some of these molecules could be the products of competing mismatch repair reactions. For this to happen, one would need evidence that mismatch repair systems are active in poxvirus-infected cells and can target the DNAs replicated by viral enzymes. We had previously not been able to detect such activities in rabbit SIRC cells, using biochemical methods (Fisher et al., 1991), but these assays are not very sensitive and always fraught with uncertainties relating to protein stability and extraction efficiency. We decided to look again for mismatch repair activities in BSC40 cells using a more sensitive and genetics-based transfection assay.

To do this, we prepared circular DNA substrates encoding two of the mismatches found in the substrates shown in Table 2 (Materials and methods), but located at a distance from the termini of the molecules to avoid any bias caused by end effects. T · G mismpairs were of particular interest because such mismatches are efficiently repaired in a variety of cells (Hoeijmakers, 2001; Scharer and Jiricny, 2001). These molecules were prepared by cutting pBSII-BclI with ScaI and pBSII-NruI with AlfIII, boiling and annealing the DNAs, and then gel-purifying the resulting nicked circular plasmids containing T · G and, by default, C · A mismatches. The DNA mixture was then transfected into vaccinia virus-infected cells, recovered, circularized, and used to transform E. coli. Because the base substitutions were designed to affect α-complementation, the ratio of lacZ+ to lacZ− colonies provided a rapid way of measuring the numbers of each allele recovered and DNA sequencing confirmed the accuracy of the approach. Using this assay, we recovered significantly more dark blue lacZ+ transformed bacteria (P < 0.001) than white lacZ transforms, from the DpnI-resistant DNA isolated from cells transfected with heteroduplex DNA (Fig. 4). Control experiments showed that this was not due to a replication bias. This is because equal numbers of the two alleles were recovered from cells that had been transfected with identical amounts of the two parental plasmids either mixed together and cotransfected (134 lacZ+: 144 lacZ− colonies) or transfected separately and then mixed prior to purifying the DNA and transforming bacteria (171 lacZ+:162 lacZ). The eight-fold excess of lacZ+ transforms suggested that BSC40 mismatch repair systems could modify the DNAs replicated by vaccinia virus, even though these viral replication systems are located in the cytoplasm of infected cells. These repair enzymes would appear to excise T from T · G mismatches and/or A from C · A mismatches. The net effect is to favor the recovery of a CG base pair and thus plasmids encoding the lacZ+ phenotype.

The substrates used in this study did not identify which of these events was biasing the recovery of lacZ+ plasmids.
since the DNA encoded a mixture of both T · G and C · A mismatches. To address this concern, oligonucleotide linkers were used to create plasmids encoding one or the other of the two mismatched DNA substrates (designated pBSII-T · G and pBSII-C · A) and then these DNAs were transfected into virus-infected cells. This analysis showed a 4.4-fold preference for converting T · G mismatches to CG basepairs and a 3.4-fold preference for converting C · A mismatches to CG basepairs (Fig. 5). Some evidence of repair failure was also noted using these substrates, as judged by the recovery of light blue colonies which DNA sequencing showed carried both lacZ+ and lacZ− plasmids (10 and 13% of transformants recovered from T · G- and C · A-transfected cells, respectively). The repair bias is still separately and collectively significant although not as dramatic as the 8-fold bias detected using mixtures of heteroduplex molecules. (This is perhaps because these more complex substrates are likely subject to other illegitimate repair reactions and these may contribute additional lacZ− alleles to the pool of rescued molecules.) Nevertheless, it is clear that substrates encoding T · G and C · A mismatches are preferentially converted to CG base pairs and this suggested a scheme by which at least some of the supposedly 5′- to 3′-excision events listed in Table 2 could have been created by a process involving biased mismatch repair. How might such a scheme work?

Mammalian mismatch repair is accomplished using either a postreplicative process, which is initiated by the binding of MutSα or MutSβ heterodimers to mismatched DNA or by base-excision repair initiated by DNA glycosylases (Hoeijmakers, 2001; Scharer and Jiricny, 2001). The nuclear distribution and complexity of the former system render it unlikely that DNAs being replicated by cytoplasmic viruses would be attacked by these proteins, nor is it clear what substrate feature could provide the necessary strand bias. However, a simple enzyme like the cellular mismatch-specific thymine DNA glycosylase might well attack the T · G heteroduplex used in these studies. This enzyme would have the requisite specificity, since it excises the thymine from a T · G mismatch, and this process would thus favor the recovery of CG base pairs from cells transfected with plasmids bearing these T · G mismatches (Figs. 4 and 5). If such a glycosylase were also to attack the thymidine in a T · G mispair, but a thymidine located near the 5′-end of a mismatched-tagged recombination substrate, subsequent attack by an apurinic endonuclease would likely release the upstream oligonucleotide and create the appearance of processing by a 5′- to 3′-exonuclease. This process might explain how a cross involving T · G and A · C-containing substrates could give rise to a few recombinants encoding cytosine in the top strand (Table 2, top row, 2/16).

How this process could bias the recovery of GC base pairs from cells transfected with circular substrates encoding A · C mismatches is less clear. One of the mammalian MutY glycosylase homologs has been reported to slowly cleave the adenine from an A · C mismatch (McGoldrick et al., 1995) and, if such an activity were present in BSC40 cells, it would again favor recovery of a GC base pair. However, neither this activity nor a G · T glycosylase activity could explain how cells transfected with DNAs bearing T · G and A · C mismatches could give rise to molecules encoding adenosine in the top strand (Table 2,
top and third rows). This suggests that even though mismatch repair systems are seemingly active in these cells, some (or even all) of the minor class of SSA events detected in poxvirus-infected cells might really be catalyzed by 5′ to 3′-exonucleases.

Conclusions

We have confirmed that, when linear DNAs are combined by vaccinia virus’s recombination systems, they are mostly processed by enzymes (like the viral DNA polymerase) which encode 3′- to 5′-exonuclease activities. We have also shown that some of the less abundant molecules bearing putative evidence of attack by 5′- to 3′-exonucleases could possibly have been the products of processing by mismatch-repair enzymes. Few reports are found in the literature as to the effect of cellular mismatch repair systems upon replicating viruses, although inefficient repair of a T · G mismatch in the primer-binding site of Moloney leukemia virus has been described (Tang and Zhang, 2000). The proteins involved in this process are probably components of base excision repair pathways, because some of these enzymes have the requisite specificity to explain the biased conversion patterns we have observed. Base excision repair pathways minimize the mutagenesis that can arise through errors in replication or through DNA damage (Scharer and Jiricny, 2001). This process is probably of little biological relevance for poxviruses because there is no obvious way of ensuring strand specificity and thus it is difficult to see how such systems could contribute anything to the fidelity of DNA synthesis. T · G mismatch repair systems are also primarily responsible for repairing deaminated 5-methylcytosine, a base not reportedly found in poxvirus genomes. However, depending upon how active these systems might be relative to competing processes like DNA replication and packaging, their existence could have some impact upon the base composition of poxviruses. Repeated passage of a virus under conditions promoting the conversion of T · G and C · A mismatches into CG base pairs would favor the gradual accumulation of a high G + C content as is seen in molluscum contagiosum virus.

We have also shown that vaccinia virus infection leads to the expression of a single-strand annealing pathway for recombination which exhibits a remarkable capacity for recombining molecules, even if these molecules share only short regions of homology interrupted by base mismatches or additional nucleotides. This observation is of some practical utility, because it further defines the minimal homology that is needed to construct recombinant viruses using recombination reactions, but of more interest is the evolutionary significance of this observation. A recombination system that can tolerate imperfect homology would facilitate the acquisition of new genes, and this may account for why poxviruses encode so many orthologs of host genes. In essence, such a system ensures that each infection provides a pool of replicating viruses with an opportunity to sample, at low frequencies, some of the host genes arrayed against them.

Materials and methods

Bacteria, cells, and virus culture

Vaccinia strain WR was propagated on BSC40 cells in minimal essential media supplemented with 5% fetal bovine serum, 1% glutamine, 1% amino acids, and antibiotic plus antimycotic at 37°C in 5% CO2. Media were purchased from Gibco. E. coli strain BMH71-18 was obtained from BD/Clontech and these mismatch repair-deficient (mutS) cells were used for all analysis of DNA recovered from virus-infected cells.

Linear DNA substrates

To investigate what effect sequence heterology has on poxvirus-mediated recombination reactions, we used the combinations of PCR-amplified and -restricted DNA illustrated in Fig. 1A. One of these substrates consisted of plasmid pRP406, which had been digested with BstEII and PacI and gel-purified to excise the middle portion of the luciferase gene. The other substrates consisted of five different luciferase gene fragments that spanned this gap. These DNAs were prepared using the PCR, a pRP406 template, and oligonucleotides that introduced zero, one, or two base substitutions within the two 20-bp homologous ends common to both the BstEII/PacI-cut pRP406 and the PCR-amplified DNAs (boxed, Fig. 1A). The mutations introduced into LucM20-S1 and LucM20-S2 were silent substitutions that would not be expected to alter luciferase activity regardless of their recombinational fate. The LucM20-I1 and LucM20-I2 insertion mutations unavoidably created one and two base frame shifts that interfere with activity measurements.

To further investigate the fate of DNA ends in vivo, we constructed a set of sequence-tagged DNA substrates similar to those described previously (Yao and Evans, 2001), except that the telomeric mismatched bases were transposed across the helix (Fig. 1b). One of these substrate pairs consisted of a 514-kb fragment of phage λ DNA that was excised from plasmid pBluescript II-λ with XhoI and HindIII, gel-purified, and ligated to two phosphorylated left and right oligomer adaptor duplexes containing G · T and C · T mismatches (bold):

GCTTGGACC GCCACAGCCCT
CGAACCTGG TGGTCTGGGGAGCTp
pAGCTGGGGGACGGGGCCGGTCACT
ccctgcccctcgccagtagc.

The ligation products were then gel-purified. The other substrate in these experiments consisted of plasmid pBlue-
script IISK that had been cut with XhoI plus XbaI and then ligated to an adapter duplex containing C · A and G · A mismatches:

\[
\begin{align*}
\text{pTCGAAGCTTGACACAGCCGCGGAGCCGCCTGCTCCT} \\
\text{TCAGAGCTTGACACAGCCGCGGAGCCGCCTGCTCCT} \\
\text{GCTGCTGACAGATCTAAGCT} \quad (\text{Linker A, strand A1}) \\
\text{CATCGGAGCAGCTAGTTCTAGA} \quad (\text{Linker A, strand A2}).
\end{align*}
\]

The circular ligation products were gel-purified and then linearized with SmaI (underlined above) to create pBSII-AdM.

**Circular DNA substrates**

Circular DNA substrates were prepared containing base mismatches located in the lacZ α-complementing region of pBluescript IISK. The starting point involved digesting pBluescript IISK with \(KpnI\) and \(SacI\) and inserting linker A:

\[
\begin{align*}
\text{GCTGCTGACAGATCTAAGCT} \quad (\text{Linker A, strand A1}) \\
\text{CATCGGAGCAGCTAGTTCTAGA} \quad (\text{Linker A, strand A2}).
\end{align*}
\]

This produced plasmid pBSII-BclI (lacZ) in which a single base substitution (bold) simultaneously created a BclI site (underlined) and a stop codon that disrupts \(α\)-complementation. A second plasmid was constructed by inserting linker B into \(KpnI-\) and \(SacI\)-cut pBluescript:

\[
\begin{align*}
\text{GCTGCGATCAAGATCTAGCT} \quad (\text{Linker B, strand B1}) \\
\text{CATCGGAGCAGCTAGTTCTAGA} \quad (\text{Linker B, strand B2}).
\end{align*}
\]

This produced plasmid pBSII-NruI encoding an \(NruI\) site (underlined) and a T-to-C substitution (bold) that maintains the open-reading frame and lacZ\(^+\) phenotype.

Circular substrates containing a mix of \(T\cdot G\) and \(A\cdot C\) mispairs were prepared by digesting pBSII-BclI and pBSII-NruI with \(SacI\) and \(A\|III\), respectively, denaturing and annealing equimolar amounts of the two DNAs, digesting with \(BclI\) and \(NruI\) to eliminate homoduplex molecules, and gel-purifying the resultant nicked heteroduplex circles. Alternatively, mismatched linkers were prepared by mixing strands A1 and B2 (\(T\cdot G\) mismatch) or strands B1 and A2 (\(C\cdot A\) mismatch), ligating the phosphorylated linkers to \(KpnI/SacI\)-cut pBSKII, gel-purifying the covalently closed circular products, cleaving with \(BglII\) to eliminate any extra copies of the linkers, religating, and repurifying the products.

**In vitro joining reactions**

Twenty microliter reactions contained vaccinia virus DNA polymerase (12 ng), vaccinia single-strand DNA-binding protein (gp31L, 30 ng), 0.1 pmol of \(BsaEII/PacI\)-digested pRP406, 0.4 pmol of PCR-amplified DNA insert, 30 mM Tris·HCl (pH 7.9), 5 mM MgCl\(_2\), 70 mM NaCl, 1.8 mM dithiothreitol, and 80 \(\mu\)g/mL of bovine serum albumin (Willer et al., 2000). The reactions were incubated at 37°C for 20 min and then deproteinized by adding 0.1% sodium dodecyl sulfate, 0.2 mg/mL proteinase K, and 50 mM EDTA and incubating for a further 20 min at 37°C. Reaction products were fractionated by agarose gel electrophoresis and visualized by staining with ethidium bromide.

**Transfection-based recombination assays**

The luciferase-based recombination assay has been described previously (Yao and Evans, 2001). Briefly, 60-mm dishes of BSC40 cells was infected with vaccinia virus at a multiplicity of infection of 2, and then calcium phosphate used to transfect these cells with a mixture of luciferase- and \(β\)-galactosidase-encoding substrates (50 ng of each luciferase-encoding recombination substrate plus 50 ng of plasmid pRP7.5lacZ). Luciferase and \(β\)-galactosidase activities were measured 5 h posttransfection and the luciferase activity, normalized for variations in transfection efficiency by \(β\)-galactosidase activity, was used to calculate a recombination frequency. Each measurement was an average calculated from duplicate dishes.

**Recovery of recombinants in E. coli and mismatch repair analysis**

A 60-mm dish of BSC40 cells was infected with vaccinia virus and transfected with pairs of linear sequence-tagged DNAs as described above (Fig. 1a). The next day (24 h post-infection), the DNA was recovered by lysis with a solution containing 0.2 mg/mL proteinase K, 1.2% sodium dodecyl sulfate, 50 mM Tris·HCl (pH 8), 4 mM CaCl\(_2\), and 4 mM EDTA and purified by phenol extraction and ethanol precipitation (Evans et al., 1988). The DNA was digested with \(DpnI\) and \(SacI\), phenol-extracted, and ethanol-precipitated, circularized with \(T4\) ligase, and used to transform the \(E. coli\) strain BMH71-18 to ampicillin resistance. Plasmids recovered through this strategy were purified and sequenced to determine the junction sequence.

Another strategy was used to genotype the plasmids recovered from cells transfected with circular mismatched substrates. The assay is based upon the fact that bacteria transformed with pBSII-BclI produce lacZ white colonies while pBSII-NruI-transformed cells exhibit a lacZ\(^+\) blue colony phenotype. DNA was recovered from infected and transfected cells as described above, except that \(A\|III\) (rather than \(SacI\)) was used to linearize the concatemers. The DNA was circularized with \(T4\) ligase and \(E. coli\) BMH71-18 transformants plated on LB agar containing ampicillin, X-gal, and IPTG. DNA sequencing confirmed that dark blue colonies arose from transformation by pBSII-NruI and white colonies from transformation by pBSII-BclI. Light blue colonies comprised a few percent of transformants and contained a mix of the two plasmids.

**Acknowledgments**

This work was supported by operating grant funds provided to D.H.E. by the Canadian Institutes for Health Re-
search. Xiao-Dan Yao is a recipient of an NSERC Postdoctoral Fellowship.

References


