Nucleic Acids Research, 2005, Vol. 33, No. 19 ***6101–6113***
doi:10.1093/nar/gki922

Rearranging the centromere of the human Y chromosome with ϕC31 integrase

Sunir Malla, Felix Dafhnis-Calas, John F. Y. Brookfield, Margaret C. M. Smith and William R. A. Brown*

Institute of Genetics, Queen's Medical Centre, Nottingham NG7 2UH, UK

Received September 4, 2005; Revised and Accepted October 6, 2005

ABSTRACT

We have investigated the ability of the integrase from the *Streptomyces* ϕC31 'phage to either delete or invert 1 Mb of DNA around the centromere of the human Y chromosome in chicken DT40 hybrid somatic cells. Reciprocal and conservative site-specific recombination was observed in 54% of cells expressing the integrase. The sites failed to recombine in the remaining cells because the sites had been damaged. The sequences of the damaged sites indicated that the damage arose as a result of repair of recombination intermediates by host cell pathways. The liability of recombination intermediates to damage is consistent with what is known about the mechanism of serine recombinase reactions. The structures of the products of the chromosome rearrangements were consistent with the published sequence of the Y chromosome indicating that the assembly of the highly repeated region between the sites is accurate to a resolution of about 50 kb. Mini-chromosomes lacking a centromere were not recovered which also suggested that neocentromere formation occurs infrequently in vertebrate somatic cells. No ectopic recombination was observed between a ϕC31 integrase *attB* site and the chicken genome.

INTRODUCTION

Manipulation of large scale chromosome structure, often referred to as chromosome engineering, has become an important tool in experimental genetics. In mice, deletions and inversions are being used in screens for recessive mutations (1) and translocations or inversions (2) are being exploited as models for the mutations underlying various human genetic diseases, particularly those involved in somatic malignancies. In *Drosophila* somatically induced mitotic recombination is used in mosaic analysis of the effects of mutations that are lethal or pleiotropic (3). Construction of such rearrangements requires the use of site-specific recombinases. In all cases, of which we are aware, chromosome rearrangements in metazoa have employed either the Cre recombinase isolated from the bacteriophage P1 or the FLP recombinase isolated from the 2 μ plasmid of *Saccharomyces cerevisiae*.

The Cre and the FLP recombinases are members of the integrase family of site-specific recombinases. These proteins use tyrosine as their active site nucleophile and the recombination reaction proceeds through a Holliday junction intermediate (4). Both proteins catalyse reversible reactions between identical sites of 34 bp in length, termed *loxP* sites in the case of Cre. The reversibility of the reactions however limits the utility of these enzymes. In particular it is difficult to use either Cre or FLP to promote inversions or translocations irreversibly except by supplying the recombinase transiently and selecting for the desired product. It is technically straightforward to do this in tissue culture but impractical in whole organisms. Although some have suggested that this type of problem may be overcome using mutant target sites we are unaware of any reports of the successful use of such sites in the engineering of large scale chromosome rearrangements *in vivo*. A second, more practical problem is posed by the toxicity and mutagenicity of Cre (5). Although widely ignored it is now clear that Cre has growth inhibitory effects and the ability to cause widespread DNA damage when expressed at high levels in mouse cells (5). The damage is thought to arise as a result of interactions between Cre and sequences similar to *loxP*; termed *loxP* pseudo-sites, in the mouse genome. These undesirable properties mean that Cre expression may have side effects in some live animal studies and in general that Cre will only be able to be used at levels which are toxicity limited. A unidirectional, non-toxic alternative to Cre would therefore be generally useful.

*To whom correspondence should be addressed. Tel: +441158493244; Fax: +441159709906; Email: william.brown@nottingham.ac.uk
Present address:
Margaret C. M. Smith, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK

A second family of site-specific recombinases uses serine as the active site nucleophile (6). The best characterized members of the serine recombinase family are the γδ resolvase and Hin invertase. These proteins bring two recombination sites together to form a synapse in which all four DNA strands are cleaved and in which phosphor di-ester links are formed between active site serine residues and each strand of the participating DNA molecules. Strands are exchanged as a result of a ligation reaction that occurs after relative rotation of the two halves of the protein–DNA complex (7). The serine recombinase family has been divided into at least three different classes of protein that differ in their domain organization and requirements for accessory proteins (6). The large serine recombinase class includes members that promote site-specific recombination without requiring any accessory proteins (8). The integrases encoded by the *Streptomyces* bacteriophage ϕC31 (9) and *Mycobacterium* 'phage BxbI (10) integrate and excise the respective viral DNA molecules into and from the host chromosome and are the best understood examples of this class of protein. The integration reaction promoted by each of these proteins occurs between two sites, termed attachment (*att*) sites; att*B* and att*P*, about 40 bp in length to yield product sites; *attL* and *attR*, and is promoted by the integrase acting alone. The excision reactions which use *attL* and *attR* to reform *attP* and *attB* are thought to require the integrases and accessory proteins. Thus these integrases acting alone promote unidirectional site-specific recombination reactions and are, thus, potentially ideal for promoting precisely those reactions for which the current set of reagents; Cre and FLP, are unsuited. In addition there are many members of the serine integrase family (6) and they probably have similar general properties. Several, including ϕC31 integrase, have been shown to be able to promote site-specific integration in eukaryotic cells (11–14). However it has been reported that vertebrate genomes contain large numbers of pseudo attachment sites for ϕC31 integrase (15). Pseudo-sites would have the potential to compromise the specificity of a desired manipulation. As in the case of Cre, pseudo-sites could also act as targets for the ectopic activity of the integrase and thus could give rise to a mutagenic background. Such concerns may have discouraged others from using the ϕC31 integrase and other serine recombinases for genome engineering.

The centromeres of the human sex chromosomes have been extensively mapped and sequenced and thus are structurally the best characterized vertebrate centromeres (16,17). Here experiments are described in which the integrase of the *Streptomyces* ϕC31 'phage has been used to manipulate the long range structure of the centromere of the human Y chromosome. Chicken DT40 cells were used to engineer a mini-chromosome derivative of the human Y chromosome to contain *attB* and *attP* sites flanking the centromeric alphoid DNA. The sites, separated by ~1 Mb, were introduced in either respective orientation and this enabled investigation of the ability of ϕC31 integrase to promote either deletion or inversion of the centromeric interval. It does so accurately and efficiently in about half of the reactions but in the rest of the sites are damaged during the rearrangements and either fail to complete the reaction or rearrange unpredictably. Pseudo-sites were not detected in the chicken genome.

MATERIALS AND METHODS

Plasmid construction

Plasmids were constructed by standard methods. In each case the *att* sites were engineered into plasmids using synthetic DNA provided by Invitrogen. The plasmids were checked by restriction site mapping and in all cases by sequencing across the *att* sites. The hygromycin resistance gene used was present in the counter selectable hygromycin–thymidine kinase fusion (HyTk) (18). The puromycin resistance gene (19) was a gift of Jean-Marie Buerstedde. The CCAG promoter (20) was a gift of Ian Chambers (Edinburgh University). The nuclear localization signal used to tag the ϕC31 integrase (8) was derived from the large T antigen of SV40 virus. It included the residues MPKKKRKV and was placed at the caroboxy terminus.

Cell culture

The starting mini-chromosome Δ2.5 was originally described elsewhere (21). It was moved from Chinese hamster ovary cells, in which it was isolated, into DT40 cells by microcell fusion (22). DT40 cells and DT40 somatic cell hybrids were maintained and electroporated as also described previously (22) except that the medium used was RPMI 1640 including 446 mg/l L-alanyl-L-glutamine with 10% foetal bovine serum (FBS), 1% chicken serum, 10-5 M 2-mercaptoethanol, 10 U/ml penicillin and 10 μg/ml streptomycin. This medium gave cleaner selection after addition of antibiotics than our earlier DMEM based medium. Following electroporation with the targeting or ϕC31 integrase expression constructs the cells were plated out in 96-well dishes. Selection was applied 18 h later. Colonies were isolated after 12–14 days. The proportion of cells expressing eGFP was determined using a fluorescence activated cell sorter.

PCR, fluorescent *in situ* and filter hybridization analysis

Conventional agarose, pulsed field gels and filter hybridization were as described previously (22). The sequences of the primers used in the PCR to check the breakpoints in the ϕC31 mediated rearrangements are given in Supplementary Table 1. Fluorescence *in situ* hybridization (FISH) was by standard methods as described (22) except that the alphoid DNA was directly labelled with Alexafluor-594 dUTP and the second sequence was labelled with biotin and detected using Streptavidin conjugated to Alexafluor-488. The Alexafluor conjugates were from Molecular Probes. These improvements led to a technique that was faster, more sensitive and gave less background than before. Restriction enzymes used in the eGFP analysis of the *attP* and *attL* sites were Asp718 and BglII and in the puromycin analysis of the *attB* and *attR* sites were NcoI, Asp718 and BglII. Targeting sequences were constructed by PCR from cloned DNA and the respective primers are given in Supplementary Table 1.

Estimation of reaction rates

We can estimate the combined rate of the two possible reactions by considering the reaction as follows.

Let us consider an initial cell P and two daughter cells I and II. There are three things that can happen to the cell as it goes

from P to I or from P to II. It can either reach final state A, or reach final state B, or can remain uncommitted. Let us call the probabilities of these three outcomes P_A, P_B and $1 - P_A - P_B$. We are interested in the state of the final descendants of the population of cells started from P. In particular, we are interested in the mean and variance of the proportion of cells that are of type A. The mean is obviously $P_A/(P_A + P_B)$. Let us suppose that the expected variance between replications is V. What is the contribution of cell I to this variance? If the cell is of type A, then all its descendants will be of type A. If I is A, then half the population will inevitably be of type A, whereas, on average, a proportion of $P_A/(P_A + P_B)$ of I's descendants would be expected to be type A. Thus, I being A contributes $\{0.5\ [1 - P_A/(P_A + P_B)]\}^2$ to the overall variance, or $[0.5\ P_B/(P_A + P_B)]^2$. But the probability of cell I is of type A is P_A. Thus, the contribution to the overall variance created by cell I changing to A is $0.25\ P_A P_B^2/(P_A + P_B)^2$. By symmetry, the contribution to the variance of cell I changing to B is $0.25\ P_B P_A^2/(P_A + P_B)^2$. If, however, I has remained undifferentiated its expected contribution to the total variance is $V/4$. Thus, in total, the contribution of cell I to the total variance is

$$0.25[(P_A + P_B)(P_A + P_B)/(P_A + P_B)^2 + (1 - P_A - P_B)V] = 0.25[(P_A P_B/(P_A + P_B) + (1 - P_A - P_B)V]$$

The contribution from both cell I and cell II is thus $0.5\ [(P_A P_B/(P_A + P_B) + (1 - P_A - P_B)V]$.

But, by definition, this is also V, so

$$0.5[(P_A P_B/(P_A + P_B) + (1 - P_A - P_B)V] = V$$

Call $(P_A + P_B)$, x, so that mean A is P_A/x and mean B is P_B/x.

Now $P_A P_B/x + (1 - x)V = 2V$ and so $V(1 + x) = P_A P_B/x$.

However, if we sample binomially N events where the probability of A is P_A/x and probability of B is P_B/x, the expected variance is $P_A P_B/Nx^2$. Thus, in our distribution by comparing the mean with the variance we can estimate N, the effective sample size given x. And N can tell us what x is in the following way.

Since $V = P_A P_B/[x(1 + x)] = P_A P_B/Nx^2$, it follows that $N = (1 + x)/x$.

For example, if the change to the final state was always in the first division), then $x = 1$ and N is 2, the variance would be determined by binomial sampling of the two cells I and II. Any lower rate of resolution (x) will increase N and thus decrease the variance.

RESULTS

Strategy

We used sequence targeting and telomere directed chromosome breakage in DT40 hybrid somatic cells to create two mini-chromosome derivatives of the human Y chromosome in each of which the centromere was flanked by *attB* and *attP* sites (Figure 1A and B). In one: CA4 (Figure 1C), the sites were placed in an opposite orientation with respect to one another and in the other: XP4 (Figure 1D), the sites were placed in a parallel orientation. The *attB* sites were first targeted into the centromeric array of alphoid DNA by telomere directed chromosome breakage of a mini-chromosome called Δ2.5 (21) to generate a mini-chromosome called ΔΔ16 (Figure 1B). The *attP* sites were then targeted into the DYZ5 array using replacement constructs (Figure 1B). Restriction analysis established that the *attB* site was at the end of a 160 kb alphoid DNA fragment while the *attP* site in CA4 was 420 kb from the centromeric end of the DYZ5 array and the *attP* site in XP4 was only 40 kb from the centromeric end of the array. The distance between the alphoid and DYZ5 arrays is 700 kb and thus the distance between the *attB* and *attP* sites in both CA4 and XP4 is ~1 Mb. Details of the routine manipulations and mapping are described in Supplementary Data.

In order to help identify any chromosomes that had been rearranged as a result of site-specific recombination we placed a promoterless *eGFP* gene adjacent to the *attP* site such that site-specific recombination would place the *eGFP* gene (Figure 1B) under the control of the β-actin promoter originally upstream of the puromycin gene in the *attB* containing construct. We used a hygromycin resistance gene fused to a herpes simplex virus thymidine kinase (*HyTk*) gene driven by a CCAG promoter (20) as the marker to select for stable transfectants containing the *attP* site. We ultimately intend to use the engineered mini-chromosomes for experimental substitution of centromeric sequences so we flanked the coding region of this marker with a *loxP* site and an *attP* site for the ϕBT1 integrase (23). As we will describe elsewhere the presence of the *loxP* and ϕBT1 *att* sites in the targeting construct should allow us to use Cre and the ϕBT1 integrase to introduce additional sequences into the engineered chromosomes independently of any rearrangement that the ϕC31 integrase may induce.

Expression of the ϕC31 integrase in cells containing a single chromosome with the sites oriented in an opposite orientation would be expected to invert the centromeric interval between the two sites (Figure 1C). The action of the ϕC31 integrase in G2 cells with two sister chromatids containing sites in the opposite orientation has the potential however to generate more complex outcomes. A single site-specific recombination event between two sister chromatids would generate a larger metacentric chromatid (referred to in the text as an inter-chromatid maxi) containing *attP* and *attR* sites and a mini-chromatid containing *attB* and *attL* sites and only the intervening DNA. Both would contain a single centromere and would be expected to be stable (Figure 1C). Of the three types of rearranged chromosome generated by the integrase acting on the sites in opposite orientations only the inter-chromatid mini retains a un-rearranged puromycin resistance gene and so cells with this mini-chromosome would, uniquely, be expected to retain resistance to this antibiotic.

Intra-chromatid site-specific recombination between *attP* and *attB* sites oriented in the same orientation would be expected to generate a circular centric fragment and an acentric linear fragment (Figure 1D). However a single exchange event between sister chromatids containing sites in the same orientation, would generate the acentric fragment produced by the intra-chromatid event but would also generate a linear dicentric fragment containing *attB*, *attP* and *attR* sites. Such a chromosome could be rearranged by a second site-specific recombination event to yield a circular dicentric chromatid (Figure 1D). The fate of these two types of dicentric

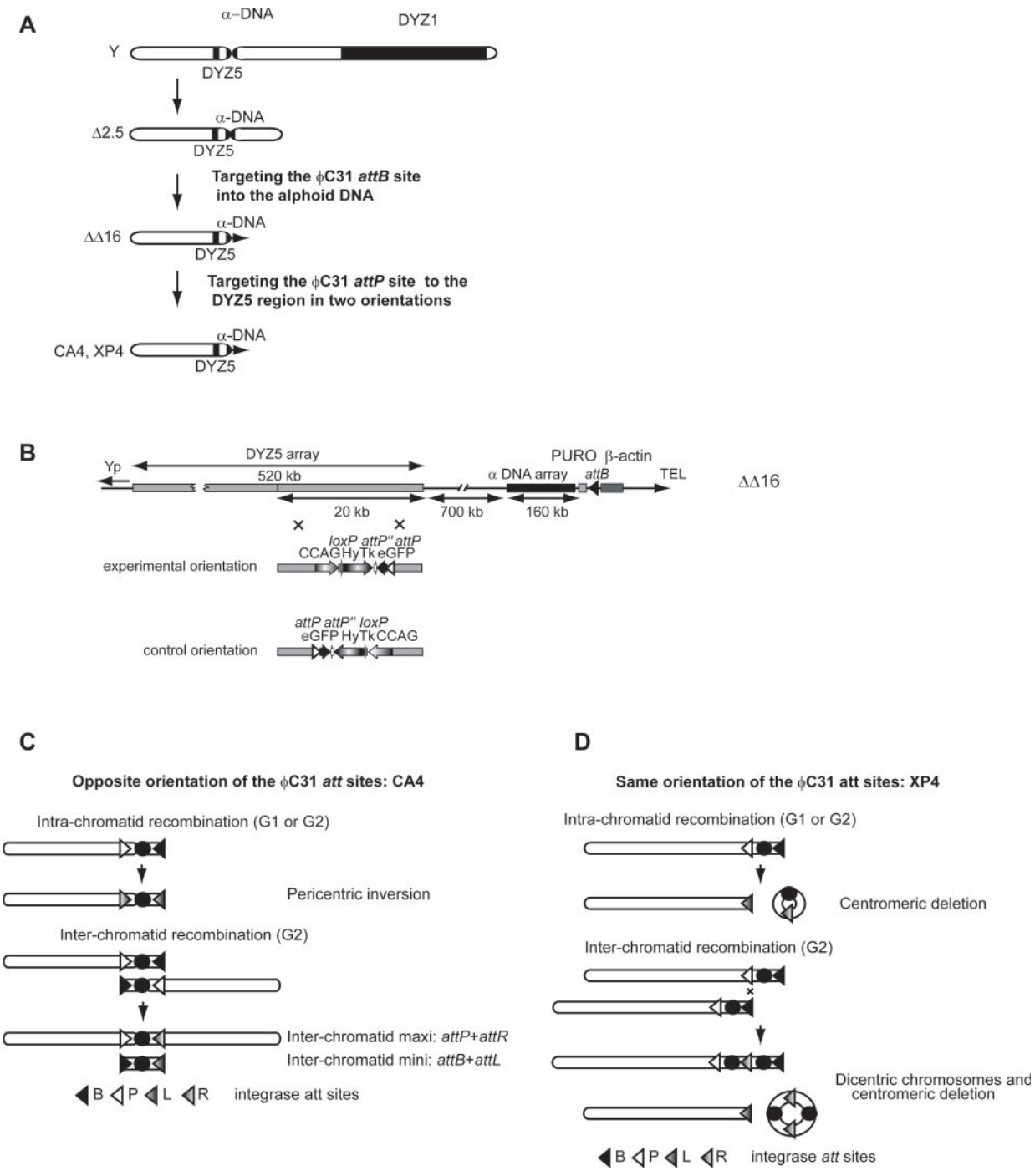

Figure 1. Engineering a human mini-chromosome with *attB* and *attP* sites that flank the centromere and the products of site-specific recombination between these sites. (**A**) Outline of scheme for flanking the centromere of a human mini-chromosome with *attB* and *attP* sites. The mini-chromosome Δ2.5 was derived from the human Y chromosome by telomere directed chromosome breakage in Chinese hamster ovary cells (21) and then transferred into chicken DT40 cells by somatic cell fusion. In the experiments described in this article an *attB* site was introduced into the centromeric alphoid DNA array by telomere directed chromosome breakage to generate the mini-chromosome ΔΔ16 and then an *attP* site was targeted to the centromere proximal DYZ5 array in each of the two possible orientations to generate the mini-chromosomes used in this work termed CA4 and XP4. Details are described in the text and in Figure 1 of the Supplementary Data. (**B**) Details of the constructs used to target the DYZ5 array. A ϕC31 integrase *attP* site was introduced into the DYZ5 array in each of the two possible orientations using constructs containing a hygromycin–thymidine kinase fusion resistance gene, flanked by a *loxP* site for the Cre recombinase and an *attP* site (indicated as *attP″*) for the ϕBT1 integrase. Immediately adjacent to the ϕC31 *attP* site on each construct was a promoterless *eGFP* gene that could act as an indicator for the site-specific recombination event. The DYZ5 unit repeat is 20 kb in length and the array is composed of ∼25 such repeats. In the CA4 chromosome the construct targeted 420 kb from the centromeric end of the array while in the XP4 chromosome the construct targeted 40 kb from the centromeric end. (**C**) Possible products of site-specific recombination between attachment sites oriented in opposite orientations. The upper section of this illustration indicates the product of an intra-chromatid recombination event. The lower section indicates the possible products that are unique to a single inter-chromatid recombination event. (**D**) Possible products of site-specific recombination between attachment sites oriented in the same orientation. The upper section of this illustration indicates the products of an intra-chromatid recombination event. The lower section indicates the possible products that are produced by inter-chromatid recombination events.

rearranged chromatids in the next division cannot be predicted with certainty but studies in budding yeast [reviewed in (24)] suggest that kinetochore polarity is established during S-phase of the cell cycle and thus it would seem likely that these fragments would be effectively dicentric for at least one cycle and would be structurally unstable in the next mitosis.

Site-specific recombination between sites flanking the centromere in an opposite orientation

Clone CA4, containing the mini-chromosome targeted with the ϕC31 *att* sites in an opposite relative orientation, was transfected with a construct that allowed expression of a derivative of the ϕC31 integrase tagged at the carboxy terminus with a nuclear localization signal (Figure 2A). Stably transfected clones were analysed for site-specific recombination first of all by using a fluorescent activated cell sorter to measure the proportion of cells that expressed *eGFP*. This varied (Figure 2B) between no detectable expression and almost complete expression. Western blotting was used to measure the amount of integrase expressed by the individual clones (Figure 2C). This indicated that there was no relationship between the levels of integrase expression and of induced fluorescence. Six clones that expressed different levels of *eGFP* were analysed using direct techniques. Filter hybridization after gel electrophoresis using probes for either the *eGFP* gene or the puromycin resistance genes that respectively flank the *attP* or *attB* sites in the un-rearranged CA4 mini-chromosome (Figure 2D) indicated an explanation for the *eGFP* fluorescence data. The simplest of the six clones was number 15 in which 100% of the cells were expressing *eGFP*. This clone contained restriction fragments consistent with intra-chromatid inversion with no residual un-rearranged mini-chromosome. PCR (Figure 2E) confirmed the presence of *attL* and *attR* and absence of *attB* and *attP* sites indicated by the blot. However, clones 6 and 10 expressed no *eGFP* and contained *attP* and *attR* sites but no *attB* and *attL* sites at the level of sensitivity afforded by filter hybridization. This combination of sites suggested that these clones arose as a result of an inter-chromatid rearrangement but that they had lost the cells containing the inter-chromatid mini-chromosome product and instead contained only the reciprocal inter-chromatid maxi (Figure 1C). For this explanation to be valid the cells containing the inter-chromatid mini-chromosome that would have also been generated by inter-chromatid recombination must have been lost from the population. One explanation for this loss was that the cells containing this inter-chromatid mini-chromosome grew more slowly than the cells containing the inter-chromatid maxi as a result of the small size of the inter-chromatid mini-chromosome activating a cell division checkpoint that delayed the cell cycle of this lineage within what would have been a mosaic cell line. Clones 2 and 8 contain a mixture of sites consistent with their containing a mixture of the intra-chromatid inversion and inter-chromatid maxi. However, Clone 2 contain no *eGFP* bright cells and this may reflect the instability of expression of the subtelomeric β-actin promoter. Clone 4 uniquely contained *attB*, *attP*, *attL* and *attR* sites and could have been explained by the presence of either incomplete rearrangement, a mixture of inter-chromatid mini- and maxi-chromosomes or by a combination of each of these types of chromosomes. However the alphoid DNA in this chromosome was found by long range mapping to be rearranged (data not shown) and so the chromosome was not studied further.

Cytogenetic or long range mapping studies was carried out in order to confirm the results of the blotting and PCR analysis. The structure of the hypothetical inter-chromatid maxi in clone 6 was confirmed by FISH (Figure 3A) using probes for the alphoid DNA (red) and for a subtelomeric sequence, cY29 (green) (25) on the short arm of the human Y chromosome. Conventional FISH had insufficient resolution to provide unambiguous evidence for the inversion in clone 15 and so pulsed field gels and restriction enzyme mapping with an enzyme, PmeI, that does not cut in either the alphoid DNA, the DYZ5 array or in the DYZ5 targeting construct enzyme was used to confirm the presence of the DYZ5-alphoid DNA and DYZ5-telomere junction fragments predicted by the maps of the un-arranged chromosome (Figure 3E and F).

The second type of experiment designed to investigate the site-specific recombination between sites with an opposite relative orientation was carried out under conditions where a unique product would be predicted: we investigated whether we could recover the inter-chromatid mini-chromosome that was missing in the products of the above experiment. Accordingly clone CA4 was transfected with the ϕC31 integrase expression construct and stably transfected clones were isolated in the presence of puromycin. We checked these by PFGE and confirmed the presence of a mini-chromosome of about 1 Mb in size. Detailed analysis of three such clones (mini-4, mini-56 and mini-81) by gel electrophoresis and filter hybridization and by PCR is shown in Figure 2F, G and H. At the level of sensitivity afforded by filter hybridization none of these three clones contained the *attR* site indicative of either an intra-chromatid inversion or an inter-chromatid maxi and all contained an *attL* site characteristic of the inter-chromatid mini. Clones mini-56 and mini-81 also contained an *attP* site as detected by gel electrophoresis and filter hybridization which indicated that the rearrangement was incomplete. Quantitation of the signal strength in the *eGFP* cognate fragments indicated that the rearrangement was 70 and 74% complete respectively. The mini-chromosome detected by FISH in the inter-chromatid mini-4 cells (Figure 3B) using an alphoid (red) and DYZ5 probe (green) was, as anticipated, smaller than the starting mini-chromosome (Figure 3C).

Incomplete site-specific recombination is associated with mutation of recombination sites

The DNA used in the analysis illustrated in Figure 2G and 2H was extracted from the cloned cells approximately one month after transfection with the integrase expression construct. This data indicated that in the clones mini-56 and mini-81 the rearrangement reaction had not occurred in all of the cells in the population. In order to test the idea that the rate of rearrangement was slow and had not yet continued to completion for kinetic reasons the cells were cultured for another month in the presence of zeocin. Analysis at the end of the incubation (data not shown) was not significantly different from that shown in Figure 2G and H. These results indicated that the reaction had proceeded to completion but that a fraction of the cells in the population were failing to rearrange. We were using an internal ribosome entry site construct

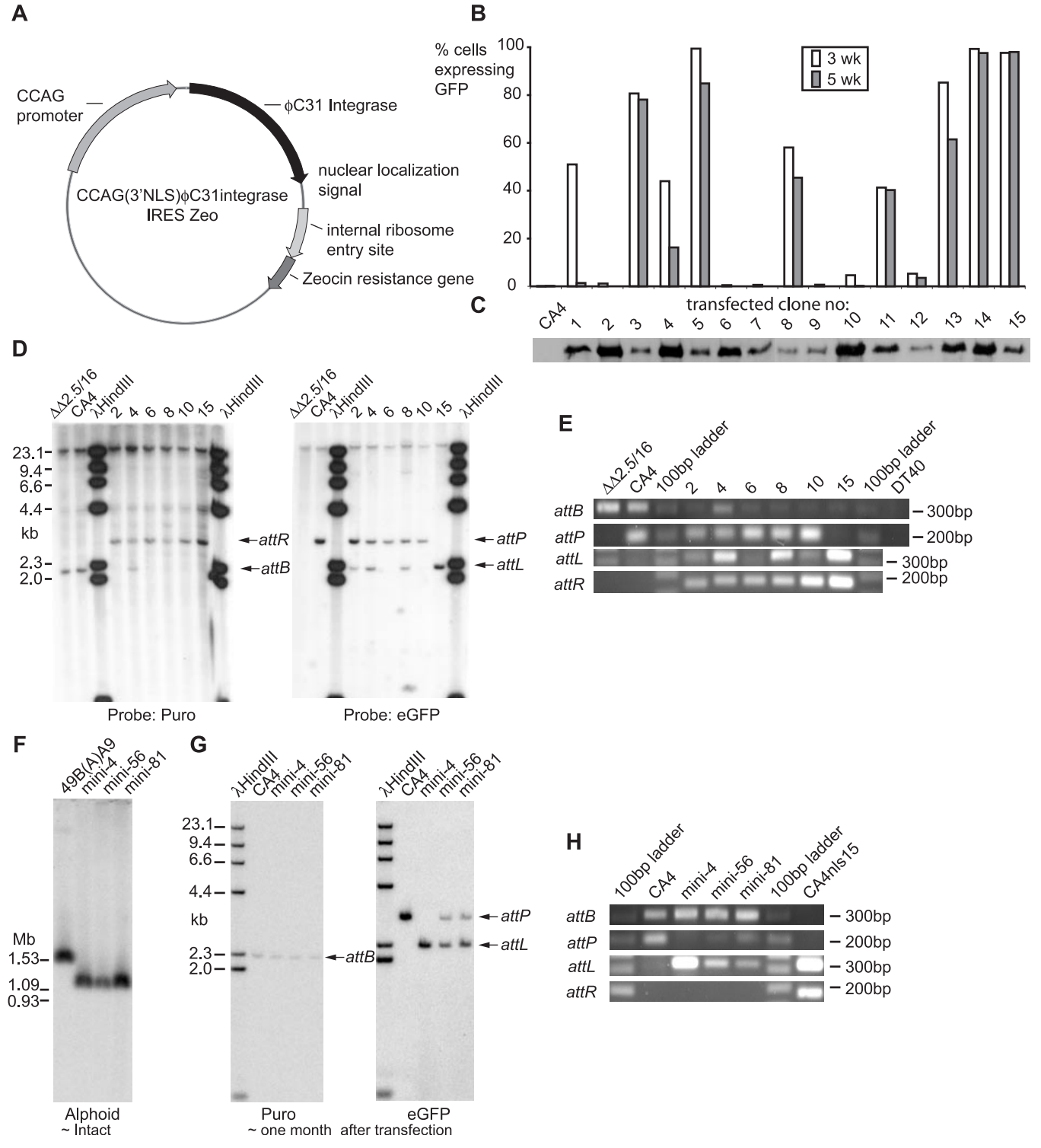

Figure 2. φC31 integrase mediated rearrangement of the *attB* and *attP* targeted human mini-chromosome CA4. (**A**) CCAG 3′-NLS φC31 integrase IRES Zeo; the plasmid used to direct expression of the NLS tagged φC31 integrase. (**B**) eGFP expression in clones derived following stable integration of CCAG 3′-NLS φC31 integrase IRES Zeo into DT40 cells containing the CA4 mini-chromosome. The bars indicate the proportion of eGFP bright cells at three and five weeks following transfection as measured by FACS. (**C**) 3′-NLS φC31 integrase expression in the clones shown in B determined by western blotting. (**D**) Assay of site-specific recombination around the puromycin resistance and eGFP genes by restriction enzyme digestion, gel electrophoresis and filter hybridization following stable expression of 3′-NLS φC31 integrase in a selection of the clones that were analysed in B and C. (**E**) Assay by PCR of *attB*, *P*, *L* and *R* sites in the clones analysed in D. (**F**) PFGE of three inter-chromatid mini-chromosome containing clones isolated from CA4 cells after stable expression of 3′-NLS φC31 integrase in the presence of puromycin. 49B(A)A9 is a mini-chromosome described in an earlier publication (22) and included as a molecular weight marker. (**G**) Assay of site-specific recombination around the puromycin resistance and eGFP genes, in the clones that were analysed in F, by restriction enzyme digestion, gel electrophoresis and filter hybridization following stable expression of 3′-NLS φC31 integrase in the presence of puromycin. (**H**) Assay by PCR of *attB*, *attP*, *attL* and *attR* sites in the clones analysed in F and G.

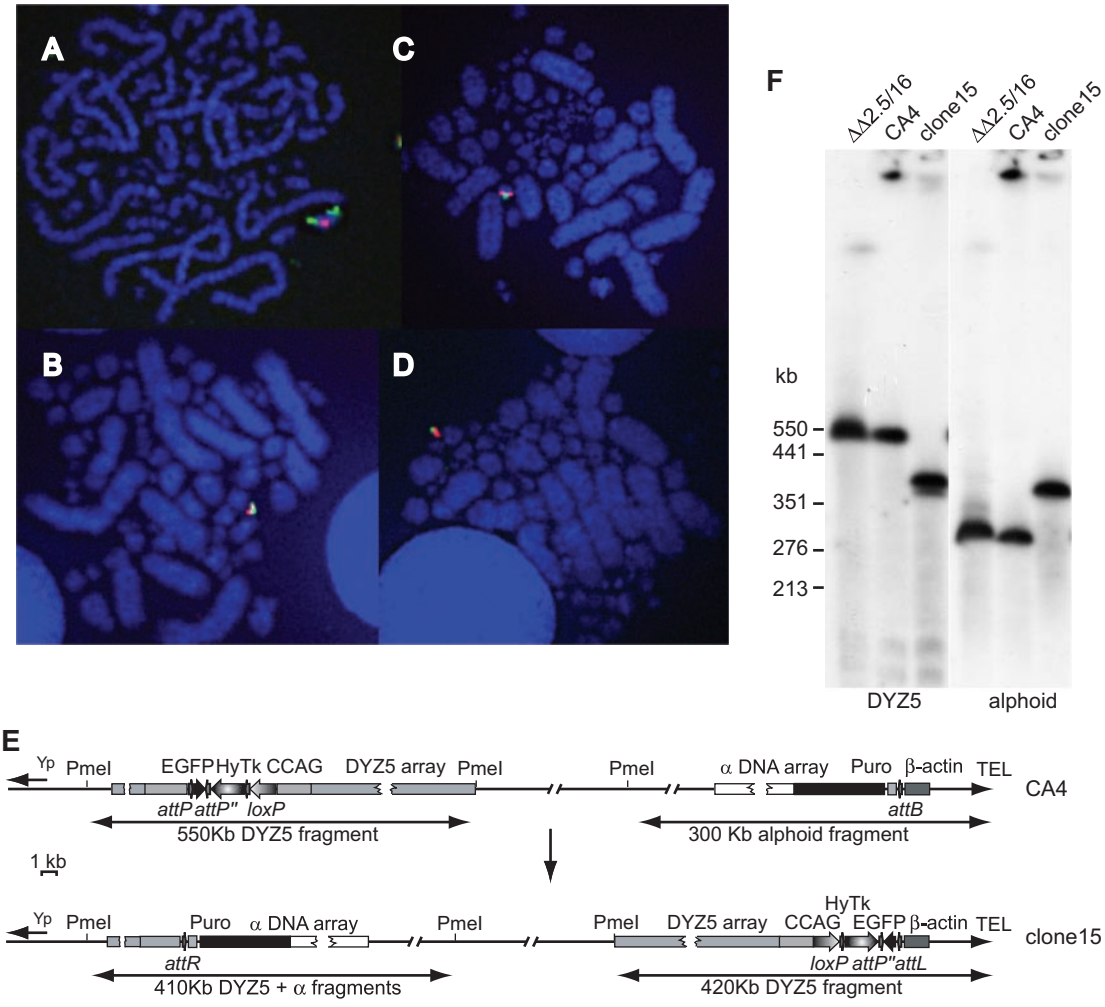

Figure 3. Establishing the identity of the ϕC31 integrase mediated chromosome rearrangements by FISH and by restriction site mapping. (**A**) FISH analysis of the inter-chromatid maxi-chromosome in the cell line CA4 clone 6 using probes against the alphoid DNA (red) and the telomere of the short arm of the human Y chromosome (25) (cY29) (green). (**B**) FISH analysis of the inter-chromatid mini-chromosome in the puromycin resistant cell line mini-4 containing an inter-chromatid mini-chromosome using probes against the alphoid DNA (red) and the DYZ5 sequence (green). (**C**) FISH analysis of an un-rearranged chromosome (in the cell line XP4) using probes against the alphoid DNA (red) and the DYZ5 sequence (green). The chromosomes in the CA4 cell line are similar. The image of the mini-chromosome is slightly confusing because the two alphoid signals (red), deriving from the sister chromatids, appear to be more centrally located than the DYZ5 sequence and to be a more obvious doublet. The well resolved doublet is a characteristic of the telomeric position of the sequences but the fact that they appear in a sub-telomeric position may reflect the way the chromosome is folded. This distribution of the two sequences was observed consistently. (**D**) FISH analysis of the circular deletion mini-chromosome in the cell line XP4 DEL 14 using probes against the alphoid DNA (red) and the DYZ5 sequence (green) of chromosomes. (**E**) Arrangement of PmeI sites around the alphoid and DYZ5 sequences in the cell lines, CA4 and CA4 clone 15 following ϕC31 integrase mediated intra-chromatid inversion of the interval between the *attB* and *attP* sites. (**F**) Filter hybridization and restriction enzyme analysis of the arrangement of the PmeI sites around the alphoid and DYZ5 sequences in the cell lines ΔΔ16, CA4 and CA4 clone 15 that was isolated after introduction of a ϕC31 integrase expression construct in the absence of puromycin.

(Figure 2A) conferring resistance to zeocin to drive expression of the integrase gene and had maintained the cells under continuous selection for integrase expression. Therefore, it seemed unlikely that variegated integrase expression was a cause of the incomplete reaction and so we wondered whether all of the *attB* and *attP* sites were intact. Of these sites the *attP* sites would be present exclusively on any un-rearranged starting mini-chromosome while the *attB* sites would be present on both the un-rearranged and inter-chromatid mini-chromosomes. Accordingly ten copies of both the *attP* and *attB* sites from each of clones mini-56 and mini-81 were sequenced. Each collection of *attB* and *attP* sites contained copies that had been mutated by deletions and substitutions (Table 1). The level of substitutions was such as to be consistent with PCR error. Taken together these observations suggested that the clones mini-56 and mini-81 consisted of a mixture of recombined and un-recombined mini-chromosomes and that the incomplete reaction was caused by the presence of either a mutated *attB* or *attP* site on the un-rearranged mini-chromosome. In order to test this idea the cell line containing the inter-chromatid mini-chromosome mini-81 was sub-cloned and six sub-clones that contained *attB* and *attL* sites (Figure 4; clones 7–12) were identified. The structure of the mini-chromosomes was checked in these sub-clones using sequence tagged sites (STSs) (Figure 5) (26) from the short arm of the Y chromosome and this showed that, as predicted, the Y chromosome distal short arm as detected by sY20 and sY37 was absent. These clones contained the STS

Table 1. Sequences of attachment sites recovered from cell lines described in this study

Cell line	Site sequenced	Sequence detected	Frequency or source of sequence
ΔΔ16φC31	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	9/9
CA4	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	20/20
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	10/10
Mini-4	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	7/10
		CCGCGGTGCGGGTGCCAGGGCGT GGCTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGC CCCCGGGCGCGTACTCCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGT GCTCCCCGGGCGCGTACTCC	1/10
	attP	Not detected	
	attL	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGT	PCR product
	attR	Not detected	
Mini-56	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTG GGCTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTGCCC CGGGCGCGTACTCC	8/10
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	8/10
		TAGTAGTGCCCCAACTGGGGTAACCT TTCTCTCAGTTGGGGGCGTA	1/10
		TAGTAGTGtCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	1/10
	attL	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGT	PCR product
	attR	Not detected	
Mini-81	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	2/10
		CCGCG GCTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTGCCC CCGGGCGCGTACTCC	2/10
		CCGCGGTGCGGGTGCCAGGGCGTGC TCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTGCC CTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGC CCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTGC CTCCCCGGGCGCGTACTCC	1/10
		CCGCGGTGCGGGTGCCAGGGCGTGCC CCCCGGGCGCGTACTCC	1/10
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	8/10
		TAGTAGTGCCCCAACaGGGGTAACCT TTCTCTCAGTTGGGGGCGTA	1/10
		TAGTAGTGCCCCgACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	1/10
	attL	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGT	PCR product
	attR	Not detected	
XP4nls 10	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGT GGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGTG CCCCGGGCGCGTACTCC	5/9
		CCGCGGTGCGGGTGCCAGGGCGT GCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGTGCC CCCCGGGCGCGTACTCC	1/9
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	8/9
		TAGTAGTGCCCCAACT GGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	1/9
	attL	Not detected	
	attR	TAGTAGTGCCCCAACTGGGGTAACCTTTGGGCTCCCCGGGCGCGTACTCCA	PCR Product
XP4nls 11	*attB*	CCGCGGTGCGGGTGCCAGGGC TCCCCGGGCGCGTACTCC	9/9
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	8/9
		TAGTAGTGCCCCAACTGGGGTAACCTTTGAGcTCTCTCAGTTGGGGGCGTA	1/9
	attL	Not detected	
	attR	TAGTAGTGCCCCAACTGGGGTAACCTTTGGGCTCCCCGGGCGCGTACTCCA	PCR product
XP4nls 14	*attB*	CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGG TTGGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAG GGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCG CCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGTGCCCTT GGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGTGCC CTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCA GGGCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGCGTG GCTCCCCGGGCGCGTACTCC	1/9
		CCGCGGTGCGGGTGCCAGGGC TCCCCGGGCGCGTACTCC	1/9
	attP	TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTA	7/10
		TAGTAGTGCCCCAACTGGGGTAACC TTGAGTTCTCTCAGTTGGGGGCGTA	1/10
		TAGTAGTGCCCCAACTGGGGTAACCTTTGAGTcCTCTCAGTTGGGGGCGTA	1/10
		TAGTAGTGCCCCAACTG TTGAGTTCTCTCAGTTGGGGGCGTA	1/10
	attL	Not detected	
	attR	TAGTAGTGCCCCAACTGGGGTAACCTTTGGGCTCCCCGGGCGCGTACTCCA	PCR product

Each row in the table represents a sequence of either an attachment site sequenced as a PCR product or as cloned DNA and this is indicated in the right hand column. This column also indicates the frequency with which the respective sequence was recovered when cloned. Deletions are indicated by gaps in the sequence and substitutions by lower case letters. Thus a continuous sequence of upper case letters indicates that the sequence was recovered intact.

detecting the DYZ5 sequence sY61 and therefore contained a mini-chromosome with a structure consistent with an inter-chromatd mini-chromosome. We also identified six clones that contained the STS's from the distal short arm of the chromosome (Figure 4; clones 1–6) and which were predicted to contain mini-chromosomes that had not rearranged. These were typed for *attB*, *attP* and *attL* and in all of them the *attL* site was absent and an *attB* site was present. However in two of

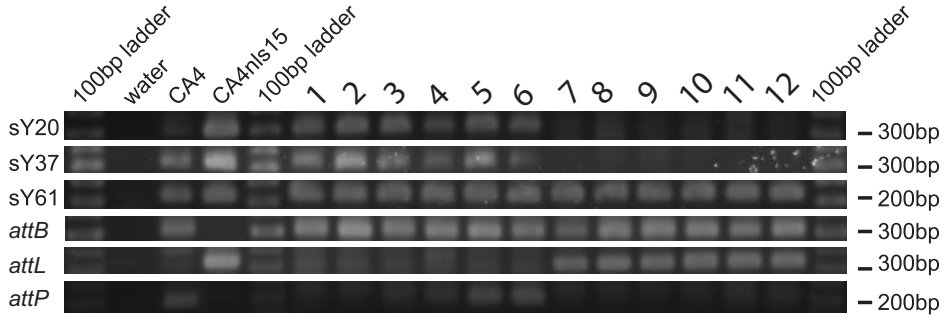

Figure 4. Sub-cloning of the cell line inter-chromatid mini-81 establishes that it is a mixture of cells containing rearranged and un-rearranged mini-chromosomes. This figure shows PCR analysis of 12 sub-clones of the cell line containing the inter-chromatid mini-chromosome mini-81 analysed in Figure 3. STS markers sY20, 37, 61 STS's from the short arm of the Y chromosome and the attachment sites, *attB*, *attL* and *attP* were analysed. The starting clone CA4 and the cell line with the intra-chromatid inversion CA4 clone 15 were analysed as controls.

the clones the *attP* site was present and in four was undetectable. We suggest that the four clones in which the *attP* site was undetectable included a deletion extending from the site which had destroyed one or both of the PCR primer binding sites. The *attB* sites and, where possible, the *attP* sites were sequenced in the six sub-clones containing the un-rearranged mini-chromosome and in every sub-clone the *attB* site had been damaged by one of the deletions seen in the PCR products from uncloned cells that had been subsequently isolated in bacterial clones. The *attL* sites in each of the six sub-clones containing the inter-chromatid mini-chromosome were also sequenced and in every case the sequence was consistent with accurate and conservative site-specific recombination.

Damage to recombination sites is associated with failed site-specific recombination reactions

These observation that the recombination sites were found to be mutated in chromosomes that had failed to recombine raised four questions, firstly whether the sites were mutated before introduction of the integrase? We sequenced *attB* and *attP* sites from clone CA4 before the introduction of the integrase. These results (Table 1) demonstrated that both sites were intact in the starting cells. The second question was whether the sites were mutated when exposed to integrase alone or did the mutations only arise when both the *attB* and *attP* sites were exposed to integrase in the same cell? We addressed this question by introducing an integrase expression construct into a ΔΔ16 clone that contains only the *attB* site and sequencing ten copies of the *attB* site after one month in culture. The sequence was not detectably mutated (Table 1). The third question was whether mutations were seen only in the substrate sites or were also present in the product sites. Accordingly we sequenced PCR products containing *attL* sites isolated from the inter-chromatid mini-chromosome containing clones mini-56 and mini-81 and in each case the sequence was consistent with accurate site-specific recombination between undamaged sites. Similarly we sequenced *attL* and *attR* sites from the intra-chromatid inversion clone 15 described above and showed that both were as predicted by accurate site-specific recombination.

These results thus indicate that the substrates were being damaged when exposed to the integrase only when both were present in the same cell and that damage was specific to the substrate sites. We interpret these observations mechanistically to suggest that the intermediate of the ϕC31 integrase reaction is a substrate for the host DNA damage response. Our observations suggest that repair of the double-strand break introduced by the integrase destroys the ability of the sites to participate in subsequent site-specific recombination reactions. The final question was whether the presence of mutated recombination sites was specific to the clones that were failing to rearrange completely? Therefore we sequenced ten copies of the *attB* site in the mini-4 clone and of these, three sequences included a deletion (Table 1). We also sequenced the *attL* site in this clone and, consistent with the results for clones mini-56 and mini-81, found this to be intact. These results suggest that the sequence of reactions leading to mini-4 was as follows. One sister undertook an abortive attempt at recombination and this left it with damaged *attB* sites and an intact *attP* site. The other sister, containing an intact *attB* site necessary for site-specific recombination then participated in a productive inter-chromatid site-specific recombination reaction with the intact *attP* site present on its sister which led to the mini-4 with sequences that we detect.

No detectable ectopic interaction between a native *attB* site on a mini-chromosome and pseudo-*attP* sites present in the chicken genome

The inter-chromatid mini-chromosomes, containing *attB* sites, were present in cells expressing the ϕC31 integrase and so we could use such chromosomes to test the possibility that there might be pseudo-*attP* sites in the chicken genome available for recombination with the *attB* site on the mini-chromosome. Such ectopic recombination would be predicted to lead to integration of the mini-chromosome into the host genome or to mini-chromosome loss. Therefore we cultured cells containing inter-chromatid mini-4 for one month and measured both copy number by FISH (Figure 5A) and structural stability by PFGE (Figure 5B). This analysis provided no evidence for any rearrangement or structural instability of the mini-chromosome and was thus consistent with an absence of ectopic site-specific recombination. However in order to interpret these observations we needed to show that the *attB* site present on the mini-chromosome and the integrase present in the cell were both active. Therefore we assembled

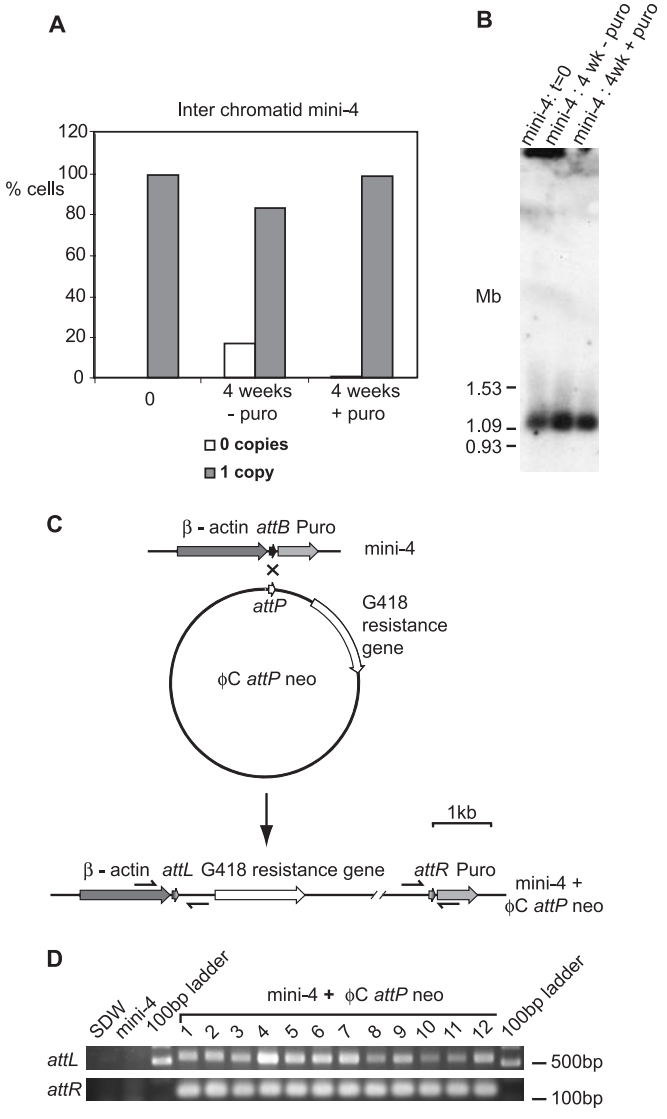

Figure 5. A mini-chromosome with a functional *attB* site is stable in DT40 cells in the presence of φC31 integrase. (**A**) Stability of mini-chromosomes in clone CA4 inter-chromatid mini-chromosome clone 4 analysed using FISH with probes for the alphoid DNA and the DYZ5 sequence after the indicated time in culture in the presence and absence of selection with puromycin. Zeocin, selecting for integrase expression, was present in both of the cultures. The filled bars indicate the proportion of cells containing a mini-chromosome and the empty bars indicate the proportion of cells lacking a mini-chromosome. (**B**) Mini-chromosomes in clone 4 remained structurally intact after four weeks in culture as judged by pulsed field electrophoresis. (**C**) Diagrammatic representation of the site-specific recombination reaction between the *attB* site in the mini-chromosomes in clone number 4 and a plasmid containing *attP* and a promoterless gene conferring resistance to the antibiotic G418; φC*attP*neo. (**D**) *attB* sites on mini-chromosomes in clone number 4 are able to undergo site-specific recombination after one month in culture as judged by PCR. Cells from clone 4 containing the inter-chromatid mini-chromosome were electroporated with the plasmid φC*attP*neo and G418 resistant clones isolated. DNA extracted from 12 such clones was analysed by PCR across the products of the site-specific recombination reaction between φC*attP*neo and the resident *attB* site.

a plasmid *attP-neo* (Figure 5C) that contained a promoterless G418 resistance gene and introduced it into the clone 4 containing the inter-chromatid mini-chromosome and selected for G418 resistant cells. We were able to recover many thousands of G418 resistant clones. We checked 12 of these for site-specific recombination using specific amplification across the recombinant *attR* and *attL* sites and found that each was a site-specific recombinant (Figure 5D). The behaviour of the chromosome thus provides no evidence for the idea that there are pseudo-*attP* sites in the chicken genome.

Site-specific recombination between sites flanking the centromere generates a circular mini-chromosome, is often incomplete and is also associated with mutation of *attB* or *attP* sites

The results of the previous experiments demonstrated that the φC31 integrase functioned with useful efficiency in promoting site-specific recombination across distances as far as 1 Mb in vertebrate cells. We also wanted to demonstrate that it could be used to promote centromere deletion and thereby provide the basis of an assay for centromeric DNA. Therefore we took cells containing the XP4 engineered mini-chromosome and transfected them with the φC31 integrase expression construct, selected for stably transfected clones in the absence of puromycin selection and analysed them by filter hybridization analysis and PCR (Figure 6A and B). The results were more complicated than one might have initially predicted but in large part this could be explained by ascribing the complications to the behaviour of the dicentric chromosomes produced by an inter-chromatid site-specific recombination (Figure 1D). Clones XP4 DEL 10, 12 and 14 were the most straightforward. The puromycin probe detected an *attR* fragment and a small amount of un-rearranged *attB* fragment. The *eGFP* probe detected residual un-rearranged *attP*. This pattern of fragments suggests that these clones contained the circular centric fragment, a variable fraction of cells with the un-recombined chromosome but lacked the linear acentric fragment predicted to contain the *attL* site. We analysed clones DEL 12 and 14 by FISH and the results bore this out (Figure 3D). In each of these clones we could see a small fragment that hybridized with alphoid DNA and weakly with DYZ5. We could also see variable amounts of both the starting chromosome and cells lacking any hybridizing material (Figure 6C). Clone 8 lacked hybridizing material completely. The presence of cells lacking hybridizing mini-chromosome in these clones could be explained in two ways. One was that the original site-specific recombination event was exclusively intra-chromatid, that this went almost to completion but that the circular centric mini-chromosome was mitotically unstable. A second explanation was that the circular mini-chromosome was mitotically stable and that the cells lacking any hybridizing material arose as a result of a different mechanism. We investigated which of these two explanations were correct by culturing clones DEL 12 and 14 for one month in the absence of selection; the results shown for clone DEL 14 (Figure 6C) shows that there is no detectable loss of the centric fragment. There is also no evidence for any further recombination indicating that as with the inversions analysed in Figure 3 the reaction went to completion by the time of the first analysis. Sequencing (Table 1) across the *attB* and *attP* sites in the clones XP4 DEL 10 and 14 detected damaged sites consistent with the notion that, as with the inversions, damage to the substrate sites was preventing

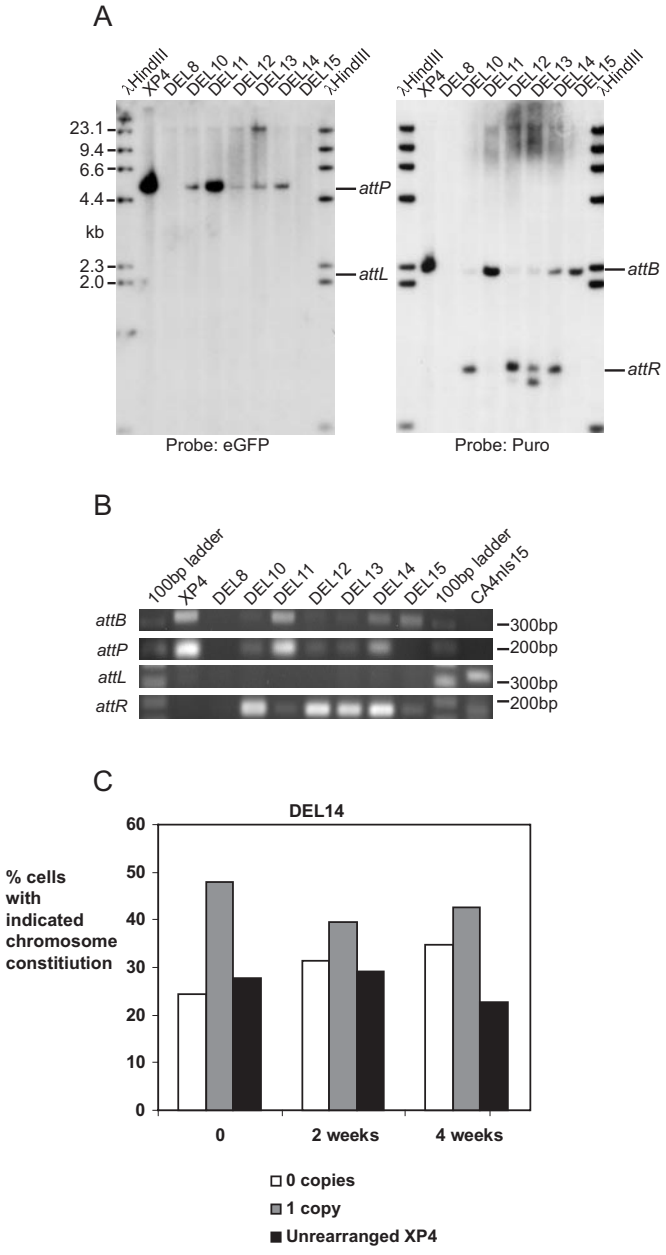

Figure 6. Deleting the centromere of the human mini-chromosome XP4 with the ϕC31 integrase. (**A**) Assay of site-specific recombination around the puromycin resistance and *eGFP* genes by restriction enzyme digestion, gel electrophoresis and filter hybridization following stable expression of 3′-NLS ϕC31 integrase. (**B**) Assay by PCR of *attB*, *attP*, *attL* and *attR* sites in the clones analysed in A. (**C**) Stability of mini-chromosomes in clone XP4 derived deletion clone DEL 14 analysed using FISH with probes for the alphoid DNA and the DYZ5 sequence after the indicated time in culture. The circular mini-chromosome could be readily discriminated from the un-rearranged chromosome on the basis of size and more easily by the intensity of staining with the DYZ5 probe (as indicated in Figure 3).

complete reaction. In order to account for the cells that lack any detectable hybridizing material we need to consider the consequences of an inter-chromatid recombination (Figure 1D). The inter-chromatid recombination event gives rise to at least two types of predictably unstable and dicentric fragments. If the broken fragments had been degraded or had failed to be incorporated into the reforming nucleus at telophase then we could account for the presence of the cells lacking any hybridizing material. Broken fragment are recombinogenic and would also be expected to integrate into the chicken genome and such a combination of processes explains the existence of a clone such as XP4 DEL 15 which contains an *attB* site and no *eGFP* cognate DNA. These considerations leave us with clones DEL 11 and 13. Clone DEL 11 has not undergone any detectable site-specific recombination. We sequenced both the *attB* and *attP* sites in this clone (Table 1) and discovered that all of the recovered *attB* sites had been mutated thus providing an explanation for the failure of detectable site-specific recombination. Clone DEL 13 contains what appeared to be an aberrantly rearranged *attR* fragment. One explanation for this reaction was that it arose as a result of faulty site-specific recombination reaction in which damage to one or both of the sites had occurred before inter-strand ligation. In order to investigate this possibility we used inverse PCR to recover the fragment. Sequence analysis however showed that this was not an inter-strand ligation product but that the *attB* site was intact. This fragment may therefore have arisen either as a result of repair of the *attB* component of an integrase reaction intermediate into the host genome or as a result of damage of an inter-chromatid dicentric derived fragment.

DISCUSSION

The results of our investigations may be stated qualitatively as demonstrating that the ϕC31 integrase promotes efficient long range chromosome rearrangements but that the reaction fails to go to completion because the sites are sometimes mutated in the course of the attempt at recombination. The mutations are mainly deletions but sometimes we detect base substitutions (Table 1). However the substitutions are detected rarely in the cloned products of the PCR, are not present in directly sequenced attachment sites and are therefore unlikely to represent mutations introduced by the repair of the attachment sites. In addition to mutations occurring within the attachment sites we have also detected larger scale rearrangements as judged by the presence of aberrantly sized restriction enzyme fragments in the products of the rearrangements. We have detected no evidence of ectopic recombination between an engineered *attB* site and ectopic pseudo-sites resident in the host genome suggesting that the protein is not reacting with pseudo-sites. Exclusive damage to the participating sites detracts from the potential utility of the ϕC31 integrase because it reduces the efficiency of the process. A more serious concern is the extent to which the DNA damage response may cause intermediates in the reaction to recombine non-homologously with the host cell genome. We detected one such potential rearrangement in the studies of the inversion reaction (Clone CA4 clone 4). However, such mutagenic events are a small minority of those that we observe.

In order to consider the results quantitatively we turn to the data from the analysis of fifteen clones isolated from the clone XP4 deletion experiments. Of these; three contained rearranged fragments and one lacked any detectable mini-chromosome DNA. These four clones could all be composed predominantly of derivatives of the inter-chromatid events discussed with reference to Figure 1D. Alternatively they could be derived by chromosome rearrangements originating from repair of the ϕC31 integrase reaction intermediates.

This first alternative seems likely to hold for at least some clones but cannot be proven and so these clones cannot be used in any quantitative discussion. In the remaining clones we estimated the extent of rearrangement by comparing the intensities of the *attB* and *attR* containing fragments recognized by hybridization with the puromycin probe (Figure 6A). This indicated that the average extent of rearrangement was 54% although this varied widely between clones. The variation in the extent of rearrangement arises in two ways. These can be easily understood by considering the site-specific recombination reaction. This may be productive or unproductive but it is irreversible. If the integrase is introduced into a G1 cell and the reaction proceeds before the subsequent S-phase then all of the cells in the clone derived from this cell will be of one type or other. However if the integrase reacts slowly with the target sites and the reaction takes many cell divisions to go to completion then the clone will be a mixture of both outcomes. Drift in the composition of the population arising from differences in growth rate between sub-clones however may also give rise to differences between the clones in the relative proportions of the productive and unproductive events. The long term culture experiments indicate that differences in growth rate are small and for the purposes of a crude estimate may be excluded. If we make this simplification then we can use the variance of the proportion of productively recombined *attB* site in the deletion reaction to estimate the rate of reaction per cell division as 0.86 per cell division (see Materials and Methods for a discussion of the calculation). This is an upper estimate but despite the small sample size and systematic imprecision of the calculation would seem to justify more elaborate experiments aimed at measuring the rate of site-specific inter-chromosomal translocations in this as well as other systems.

Attachment site mutation arises only when the two sites are present in the same cell as the integrase. This observation suggests that the attachment sites are mutated as a result of the host cell DNA repair machinery recognizing the ϕC31 integrase nucleoprotein complex as it undergoes recombination. A likely stage in recombination that could be targeted is the cleaved DNA intermediate undergoing strand exchange. Such repair is consistent with the structure of the intermediates and the kinetics of serine recombinases promoted reactions (7,27,28). Furthermore both the BxbI (10) and ϕC31 (9) integrases will relax a supercoiled plasmid containing an attachment site in the presence of a short linear double-stranded DNA molecule containing the reciprocal site. These results indicate that the DNA bound integrase subunits whilst covalently linked to the cleaved DNA can iterate strand exchange before effecting ligation. However it is important to point out that we are studying reactions between substrates separated by ~1 Mb. Further work will be required to establish how the level of attachment site mutation varies with the distance between the two sites and the experimental cell line.

A third variable is the site-specific recombinase itself. The ϕC31 integrase is a member of a large class of serine recombinases (6) and it would seem reasonable to imagine that small differences between the kinetics of the reactions promoted by different members of this class of protein and between the structures of the different recombination synapses would give rise to differences in the susceptibility of the different intermediates to the reactions that we suggest are promoted by the cellular repair pathways.

Our results thus raise the obvious question: what utility does the ϕC31 integrase have for engineering chromosome rearrangements in vertebrate cells? This question cannot be addressed in isolation but only by consideration of the alternatives. As discussed in the introduction Cre is a powerful mutagen and at levels that are not detectably mutagenic promotes translocations in vertebrates only slowly (29). We are not aware of any study that has subject Cre to the scrutiny that our work has applied to the ϕC31 integrase. If the ϕC31 integrase were to be more efficient and less damaging mutagen than Cre then ϕC31 integrase would be a general alternative to Cre. Cre and Flp are also reversible and so are unsuited to experiments where the rearrangement needs to be irreversible. Our results prove that ϕC31 integrase can be used for such experiments.

Our work was initiated as part of an ongoing project aimed at investigating vertebrate centromeres and the results also have implications for the current attempts to close the human genome sequence and for the study of centromeres. Firstly the sizes of the inter-chromatid mini-chromosomes are consistent with the current assembly of the sequence of the proximal short arm of the human Y chromosome. Such extensively repeated regions of human chromosomes are difficult to assemble into reliable contigs and our observations provide confidence in the accuracy of such assemblies. The resolution of the pulsed field gels is only about 50 kb and we cannot exclude errors in assembly or gaps smaller than this. Second the loss of the acentric fragments generated in the centromere deletion reactions formally demonstrates that DT40 cells do not form neo-centromeres on linear chromosome fragments with two telomeres at a high frequency.

In summary we have shown that that the ϕC31 integrase promotes efficient, irreversible, site-specific long range chromosome rearrangement in vertebrate cells. No evidence of ectopic recombination between a ϕC31 integrase attachment site and genomic pseudo-sites was found. The ϕC31 integrase is therefore unique in its proven ability to promote irreversible chromosome engineering reactions. However further work is required to establish the full range of chromosome engineering reactions for which it is the most suitable reagent.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the BBSRC under the Exgen program, the EU (FP5 Contract QLK3-CT-2000-00785 and FP6 Proposal GENINTEG) and a Nottingham University studentship to SKM. We thank Alistair Chambers for comments on the manuscript and Jean-Marie Buerstedde for support. Funding to pay the Open Access publication charges for this article was provided by JISC.

Conflict of interest statement. None declared.

REFERENCES

1. Yu,Y. and Bradley,A. (2001) Engineering chromosomal rearrangements in mice. *Nature Rev. Genet*, **2**, 780–790.

2. Forster,A., Pannell,R., Drynan,L.F., Codrington,R., Daser,A., Metzler,M., Lobato,M.N. and Rabbitts,T.H. (2005) The invertor knock-in conditional chromosomal translocation mimic. *Nature Methods*., **2**, 27–30.
3. Lee,T. and Luo,L. (2001) Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci*., **24**, 251–254.
4. Guo,F., Gopaul,D.N. and van Duyne,G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*., **389**, 40–46.
5. Loonstra,A., Vooijs,M., Beverloo,H.B., Allak,B.A., van Drunen,E., Kanaar,R., Berns,A. and Jonkers,J. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl Acad. Sci. USA*., **98**, 9209–9214.
6. Smith,M.C.M. and Thorpe,H.M. (2002) Diversity in the serine recombinases. *Mol. Microbiol*., **44**, 299–307.
7. Li,W., Kamtekar,S., Xiong,Y., Sarkis,G.J., Grindley,N.D. and Steitz,T.A. (2005) Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs. *Science*., **309**, 1210–1215.
8. Thorpe,H.M. and Smith,M.C.M. (1998) *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl Acad. Sci. USA*., **95**, 5505–5510.
9. Smith,M.C., Till,R. and Smith,M.C.M. (2004) Switching the polarity of a bacteriophage integration system. *Mol. Microbiol*., **51**, 1719–1728.
10. Ghosh,P., Kim,A.I. and Hatfull,G.F. (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol. Cell*., **12**, 1101–1111.
11. Groth,A.C., Olivares,E.C., Thyagarajan,B. and Calos,M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl Acad. Sci. USA*., **97**, 5995–6000.
12. Thyagarajan,B., Olivares,E.C., Hollis,R.P., Ginsburg,D.S. and Calos,M.P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell Biol*., **21**, 3926–3934.
13. Olivares,E.C., Hollis,R.P. and Calos,M.P. (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene*., **278**, 167–176.
14. Stoll,S.M. and Calos,M.P. (2002) Extrachromosomal plasmid vectors for gene therapy. *Curr. Opin. Mol. Ther*., **4**, 299–305.
15. Thyagarajan,B., Guimaraes,M.J., Groth,A.C. and Calos,M.P. (2000) Mammalian genomes contain active recombinase recognition sites. *Gene*., **244**, 47–54.
16. Schueler,M.G., Higgins,A.W., Rudd,M.K., Gustashaw,K. and Willard,H.F. (2001) Genomic and genetic definition of a functional human centromere. *Science*., **294**, 109–115.
17. Skaletsky,H., Kuroda-Kawaguchi,T., Minx,P.J., Cordum,H.S., Hillier,L., Brown,L.G., Repping,S., Pyntikova,T., Ali,J., Bieri,T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*., **423**, 825–837.
18. Lupton,S.D., Brunton,L.L., Kalberg,V.A. and Overell,R.W. (1991) Dominant positive and negative selection using a hygromycin phosphotransferase-thymidine kinase fusion gene. *Mol. Cell Biol*., **11**, 3374–3378.
19. de la Luna,S., Soria,I., Pulido,D., Ortin,J. and Jimenez,A. (1988) Efficient transformation of mammalian cells with constructs containing a puromycin-resistance marker. *Gene*., **62**, 121–126.
20. Niwa,H., Yamammura,K. and Miyazaki,J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*., **108**, 193–201.
21. Shen,M.H., Yang,J., Loupart,M.L., Smith,A. and Brown,W.R.A. (1997) Human mini-chromosomes in mouse embryonal stem cells. *Hum. Mol. Genet*., **6**, 1375–1382.
22. Yang,J.W., Pendon,C., Yang,J., Haywood,N., Chand,A. and Brown,W.R.A. (2000) Human mini-chromosomes with minimal centromeres. *Hum. Mol. Genet*., **9**, 1891–1902.
23. Gregory,M.A., Till,R. and Smith,M.C.M. (2003) Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *J. Bacteriol*., **185**, 5320–5323.
24. Nasmyth,K. (2002) Segregating sister genomes: the molecular biology of chromosome separation. *Science*., **297**, 559–565.
25. Cooke,H.J., Brown,W.R.A. and Rappold,G.A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature*., **317**, 687–692.
26. Vollrath,D., Foote,S., Hilton,A., Brown,L.G., Beer-Romero,P., Bogan,J.S. and Page,D.C. (1992) The human Y chromosome—a 43-interval map based on naturally-occurring deletions. *Science*., **258**, 52–59.
27. Dhar,G., Sanders,E.R. and Johnson,R.C. (2004) Architecture of the hin synaptic complex during recombination: the recombinase subunits translocate with the DNA strands. *Cell*., **119**, 33–45.
28. Leschziner,A.E. and Grindley,N.D. (2003) The architecture of the gammadelta resolvase crossover site synaptic complex revealed by using constrained DNA substrates. *Mol. Cell*., **12**, 775–781.
29. Zong,H., Espinosa,J.S., Su,H.H., Muzumdar,M.D. and Luo,L. (2005) Mosaic analysis with double markers in mice. *Cell*., **121**, 479–492.