Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage ϕ C31 Integrase

BHASKAR THYAGARAJAN, ERIC C. OLIVARES, ROGER P. HOLLIS, DANIEL S. GINSBURG, AND MICHELE P. CALOS*

Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120

Received 12 January 2001/Returned for modification 21 February 2001/Accepted 23 March 2001

We previously established that the phage ϕ C31 integrase, a site-specific recombinase, mediates efficient integration in the human cell environment at *attB* and *attP* phage attachment sites on extrachromosomal vectors. We show here that phage *attP* sites inserted at various locations in human and mouse chromosomes serve as efficient targets for precise site-specific integration. Moreover, we characterize native “pseudo” *attP* sites in the human and mouse genomes that also mediate efficient integrase-mediated integration. These sites have partial sequence identity to *attP*. Such sites form naturally occurring targets for integration. This phage integrase-mediated reaction represents an effective site-specific integration system for higher cells and may be of value in gene therapy and other chromosome engineering strategies.

For the past 25 years, it has been possible to construct precisely designed DNA molecules in the test tube thanks to the techniques of recombinant DNA. In contrast, the ability to make controlled and efficient alterations in the genomes of living higher cells has been limited. The use of site-specific recombinases such as Cre and FLP provided an important advance (17), but because of the reversibility of these enzyme reactions, their main utility has been for creating deletions. For the integration of new material into the genome, fortuitous integration of transfected DNA is most often used, and it produces integration at random locations at low frequency. Homologous recombination provides site specificity, but at very low efficiency (26).

We began working with another site-specific recombinase, the phage ϕ C31 integrase, because it offered the potential for unidirectional integration that would therefore occur at higher net frequencies than the reversible integration directed by recombinases, such as Cre. Cre recombines two identical *loxP* sites, recreating two identical sites after recombination that can undergo a subsequent round of recombination. In contrast, the *attB* and *attP* recognition sites recognized by the ϕ C31 integrase are dissimilar in sequence (15). After reaction, the recombined *att* sites differ from *attB* and *attP* and are refractory to further synapsis by the integrase, thus locking in integration reactions (23, 24). We demonstrated that this enzyme, derived from a *Streptomyces* phage (9), worked well in the human cell environment (7), consistent with its lack of cofactor requirements (23). This feature distinguishes it from better known phage integrases, such as that of phage λ , which does require cofactors (10). The λ integrase is in the family of recombinases that includes Cre and FLP and carries out a tyrosine-mediated strand exchange (4, 13). The ϕ C31 integrase is in the other major family of site-specific recombinases that includes many resolvases and invertases and uses a serine-catalyzed reaction

mechanism (20). The two site-specific recombinase families are unrelated. The ϕ C31 integrase is a member of a recently discovered subclass of the serine recombinase family whose members are especially long and function as phage integrases (9, 15, 23).

By using extrachromosomal plasmids in human cells, we documented that the ϕ C31 integrase mediates highly efficient intramolecular integration reactions (>50%) and also efficient intermolecular integration into an Epstein-Barr virus model chromosome (7). These results suggest that the ϕ C31 integrase would be useful for mediating integration into mammalian chromosomes, which forms the subject of this study. We created human and mouse cell lines containing *att* sites inserted at random locations in the chromosomes. These lines were tested for the efficiency of integration of incoming plasmids bearing marker genes and *att* sites when cotransfected with the ϕ C31 integrase gene. We describe *attP*-containing cell lines that exhibit site-specific integration at appreciable frequencies.

Another opportunity afforded by this integrase system is the possibility of accessing integration at naturally occurring chromosomal sequences. Such reactions would obviate the need to first place a target *att* site in the genome. This strategy is of particular relevance in applications such as in vivo gene therapy, where a high frequency of integration in unaltered patient tissue is desired. We previously showed that the sizes of the *attB* and *attP* target sites recognized by the ϕ C31 integrase are 34 and 39 bp, respectively (7). This size range makes it statistically feasible that “pseudo” *att* sites, sites with degenerate *att* identity that is still recognizable by the enzyme, may be present in large genomes, such as those of mammals (25). We document here the presence of active pseudo *attP* sites in the human and mouse genomes that are recognized at significant frequencies by the ϕ C31 integrase. This phenomenon gives rise to a strategy for the efficient and precise alteration of the genomes of living cells at predetermined sites.

* Corresponding author. Mailing address: Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5120. Phone: (650) 723-5558. Fax: (650) 725-1534. E-mail: calos@stanford.edu.

MATERIALS AND METHODS

Plasmids. Plasmids used for the measurement of luciferase activity were made as follows. A fragment carrying the cytomegalovirus (CMV) immediate-early

FIG. 1. Plasmids used to detect chromosomal integration. (A) The pL-attB plasmid carried the ϕ C31 *attB* site and the luciferase gene and was used to measure the level and persistence of luciferase expression in the presence and absence of the ϕ C31 integrase, reflecting chromosomal integration at pseudo *attP* sites. A similar plasmid, pL-attP, carried the *attP* site and was used to monitor integration at pseudo *attB* sites. (B) pHZ-attP was used to place *attP* sites into the genome by using the hygromycin resistance selectable marker. pHZ-attB is a similar plasmid used for placement of *attB* sites in the genome. (C) pNC-attB, which carried an *attB* site and the neomycin resistance marker, was used as the incoming plasmid for identification of clones that underwent integration at an *attP* site or a pseudo *attP* site. pNC-attP was the parallel plasmid used to detect integration at *attB* or pseudo *attB* sites. (D) Zeocin selection scheme. From the top is shown a host chromosome (wavy lines) in which we integrated plasmid pHZ-attP (straight lines) bearing the hygromycin resistance gene for selection, an *attP* site for integrase-mediated recombination with *attB*, and a zeocin resistance gene lacking a promoter. The donor plasmid pNC-attB carries a neomycin resistance gene for selection, a green fluorescent protein (GFP) gene for analysis of transfection efficiency, and an *attB* site for integrase-mediated recombination with *attP*. Adjacent to *attB* is the strong promoter from CMV. Upon cotransfection of pNC-attB with pCMV-Int as a source of integrase, site-specific integration leads to the potential for expression of the zeocin resistance gene from the CMV promoter introduced on pNC-attB. The same donor plasmid was used in unmodified 293 and 3T3 cells, where it was selected with neomycin alone.

promoter was cloned into the *Sma*I site upstream of the firefly luciferase gene of pGL3-basic (Promega Corporation, Madison, Wis.) to create pL. Plasmid pL-attB (Fig. 1A) was generated by cloning a 307-bp *Eco*RI fragment from pTA-attB (7) containing the minimal length ϕ C31 *attB* site and ~270 bp of surrounding ϕ C31 sequence into the *Bam*HI site of pL. Plasmid pL-attP was generated by cloning a 250-bp *Eco*RI fragment from pTA-attP (7) containing the minimal *attP* site and ~210 bp of surrounding ϕ C31 sequence into the *Bam*HI site of pL.

Plasmid pHZ-attP (Fig. 1B) was used to generate cell lines containing wild-type *attP*, while pHZ-attB was used to generate *attB* cell lines and in plasmid rescue experiments. To generate pHZ-attB, a multiple cloning site was first added into the *Hind*III site of pTK-Hyg (Clontech, Palo Alto, Calif.) to facilitate

further cloning, generating pMSE. A 1-kbp *Bam*HI fragment containing the upstream mouse sequence transcriptional terminator sequence from pUCE (27) was cloned downstream of the hygromycin resistance gene in pMSE to generate pMSEU. The UMS terminator prevents read-through transcription from the herpes simplex virus-thymidine kinase (HSV-TK) promoter. A 673-bp *Rsv*II-*Sph*I fragment from pZeoSV2(+) (Invitrogen, Carlsbad, Calif.) containing the EM-7 bacterial promoter, zeocin open reading frame, and simian virus 40 poly(A) region was cloned into the *Sma*I site of pMSEU to generate pMSEUZ1. Plasmid pHZ-attB was then generated by cloning a 307-bp *Eco*RI fragment from pTA-attB (7) containing the ϕ C31 *attB* site into the *Bam*HI site of pMSEUZ1. Plasmid pHZ-attP was constructed similarly, except that a 250-bp *Eco*RI frag-

ment from pTA-attP (7) containing the ϕ C31 *attP* site was cloned into the *Bam*HI site of pMSEUZ1.

Plasmid pEGFP-C1 was obtained from Clontech and was used to measure transfection efficiency by observing transfected cells under a UV microscope 48 h after transfection and counting bright cells. Plasmids pNC-attB (Fig. 1C) and pNC-attP were used as incoming donor plasmids for integration and were made as follows. A 307-bp *Eco*RI fragment from pTA-attB (7) containing ϕ C31 *attB* was cloned into the *Bgl*II site of pL so that the *attB* site was 3' of the CMV promoter. A 1.2-kb *Mlu*I-*Hind*III fragment containing the CMV-*attB* fragment was then cloned into the *Mlu*I site of pEGFP-C1 to generate pNC-attB. For pNC-attP, a 630-bp *Bgl*II-*Rsr*II fragment from pZeoSV2(+) containing the CMV promoter was cloned into the *Bam*HI site of pTA-attP (7) to generate pCRPCMV1. A 960-bp *Hind*III-*Xho*I fragment from pCRPCMV1 containing the CMV promoter and the *attP* site was cloned into the *Mlu*I site of pEGFP-C1 to generate pNC-attP.

pCMVSPORT β Gal (Life Technologies, Gaithersburg, Md.) was used as carrier DNA, and pCMV-Int was used for ϕ C31 integrase expression (7).

Cell culture and luciferase assays. 293 human embryonic kidney cells (6) and mouse NIH 3T3 cells (American Type Culture Collection, Manassas, Va.) were grown in Dulbecco's modified Eagle medium (Life Technologies) supplemented with 110 mg of sodium pyruvate/liter and 9% fetal bovine serum.

To assay luciferase expression, 293 or 3T3 cells that had reached 50 to 80% confluency in a 60-mm-diameter dish were transfected with 50 ng of the donor plasmid containing a luciferase expression cassette and either no *att* site (pL), one *attB* site (pL-attB), or one *attP* site (pL-attP), and 5 μ g of either pCMV-Int or pCMVSPORT β Gal carrier, by using Lipofectamine (Life Technologies). At 24 h after transfection (day 1), the cells were transferred onto 100-mm-diameter plates at an appropriate dilution. Seventy-two hours after transfection (day 3), two-thirds of the cells were harvested and a crude protein extract was prepared from them as described below. The remaining cells were replated onto 100-mm-diameter plates. This process was repeated every 2 to 4 days, depending on the confluency of the cells. Three such experiments were performed with 293 cells, and two experiments were performed with NIH 3T3 cells.

Harvested cells (approximately 10^7) were washed three times with ice-cold phosphate-buffered saline. The cells were then resuspended in 400 μ l of cold lysis buffer (25 mM Tris-HCl [pH 7.8], 2 mM EDTA, 0.5% Triton X-100, 5% glycerol) and incubated on ice for 5 min. The lysed cell suspension was centrifuged for 5 min in a microcentrifuge. The supernatant was carefully removed, transferred into aliquots, and stored at -80°C . Luciferase Assay Reagent (Promega Corporation) was used to determine luciferase activity in crude protein extracts by using a TD-20e luminometer (Turner Designs, Sunnyvale, Calif.). The relative luciferase activity in the crude protein extracts was standardized with respect to Quantilum recombinant luciferase (Promega). The luciferase activity in the protein extracts was further normalized with respect to the protein concentration. Protein concentration of the extracts was determined using the DC protein assay (Bio-Rad Laboratories, Hercules, Calif.). Results were expressed as a percentage of day 3 luciferase values.

Construction of *att* cell lines. Plasmids pHZ-attB and pHZ-attP were treated with *Hpa*I to generate linear molecules. Either 5 or 10 μ g of linearized plasmid DNA were electroporated into 293 and NIH 3T3 cells using a Bio-Rad Gene Pulser according to the manufacturer's recommendations. The cells were then plated onto nonselective medium and allowed to recover. After 24 h, selection was started using medium containing 200 μ g of hygromycin B (Calbiochem, La Jolla, Calif.)/ml. Single, well-isolated colonies were picked 12 to 14 days after the start of selection and expanded until ready for analysis. Cell lines were screened for the presence of the *att* site by PCR and Southern analysis. Four *attP*-containing 293 cell lines (293P1, 293P2, 293P3, and 293P4) and three *attP*-containing 3T3 cell lines (3T3P1, 3T3P2, and 3T3P3) were selected for further analysis, along with several *attB* cell lines of each cell type.

Selection assays. Unmodified 293 and 293 *att*-containing cell lines were grown to 50 to 80% confluency in 60-mm-diameter dishes and transfected with 50 ng of the donor plasmid and 5 μ g of either pCMV-Int or pCMVSPORT β Gal carrier by using Lipofectamine (Life Technologies). At 24 h after transfection, the cells were transferred onto 100-mm-diameter dishes at an appropriate dilution. We found that 5 μ g of DNA is near the upper limit for transfection of 293 cells on 60-mm-diameter dishes without appreciable toxicity. At this point, the total number of transfected cells was determined by counting the number of enhanced green fluorescent protein (EGFP)-expressing cells. The transfection frequency ranged from 4 to 7% for 293 cells. A further 24 h after expansion, selection was started with medium containing either 350 μ g of Geneticin (G418, a neomycin analog; Life Technologies)/ml or a combination of Geneticin and 200 μ g of zeocin (Invitrogen)/ml. Selection was continued for 14 days, and individual colonies were counted. The integration frequency was calculated as the ratio of

the number of colonies obtained to the total number of transfected cells and was expressed as a percentage. Similar experiments were performed with NIH 3T3 cells and 3T3 *att*-containing cell lines, with the following changes: 50 ng of donor plasmid was transfected along with 2 μ g of either pCMV-Int or pCMVSPORT β Gal carrier by using Lipofectamine Plus (Life Technologies). The transfection frequency ranged from 1 to 5% for 3T3. Selection was performed with medium containing either 650 μ g of Geneticin/ml or 650 μ g of Geneticin/ml and 200 μ g of zeocin/ml.

Plasmid rescue and sequence analysis of integrants at pseudo *att* sites. Unmodified human 293 and mouse 3T3 cells were cotransfected with *attB* donor plasmid pHZ-attB and ϕ C31 integrase expression plasmid pCMV-Int. Transfections were split to three 100-mm-diameter tissue culture dishes 24 h after transfection, and selection with hygromycin (200 μ g/ml) was begun 48 h after transfection. After 2 weeks of selection, colonies were trypsinized and redistributed over the plates to generate pools of hygromycin-resistant integrant clones. The pools were grown to confluency, and genomic DNA was prepared using a Blood and Cell Culture DNA Maxi kit (Qiagen, Valencia, Calif.).

To recover integrated pHZ-attB plasmids along with flanking genomic sequences, genomic DNA was linearized with two sets of restriction enzymes with compatible ends that did not cleave within the plasmid (*Bam*HI and *Bgl*II; *Xba*I, *Spe*I, and *Nhe*I). Digests were ligated with T4 DNA ligase under dilute conditions favoring monomer circularization. Ligations were extracted with phenol:chloroform and were ethanol precipitated, and a fraction (25%) was electroporated into competent DH10B *Escherichia coli* cells. Bacteria were plated on Luria-Bertani agar containing 50 μ g of zeocin/ml and 100 μ g of ampicillin/ml to select for two of the resistance genes contained by pHZ-attB. Plasmid DNA was prepared from single colonies and subjected to restriction mapping and DNA sequencing. Primers used for sequencing were attB-F (5'-TACCGTCGACGATGTAGGTCACGGTC-3') and attB-R (5'-GTCGACATGCCCGCCGTGACC G-3').

Analysis of distribution of integration sites. pNC-attB was transfected into the four 293 *attP*-containing cell lines as described above. After 14 days of neomycin selection, individual colonies were picked and expanded for further analysis of integration. A total of 24 neomycin-resistant clones and 24 neomycin- and zeocin-resistant clones were picked from each cell line, for a total of 96 neomycin-resistant clones and 96 neomycin- and zeocin-resistant clones. Genomic DNA was prepared from each clone using the DNeasy 96 Tissue kit (Qiagen) and screened for the presence of site-specific recombination junctions. These genomic DNA samples were amplified by PCR using primers specific for a junction generated by a site-specific recombination reaction between *attP* and *attB*. The remaining neomycin-resistant clones were similarly screened for integration into pseudo *attP* site human ψ A by using primers specific for that site.

RESULTS

Genomic integration detected by stable luciferase expression. We wished to determine whether the ϕ C31 integrase could direct measurable integration into the chromosomes of unmodified human and mouse cells. As a first approach, we placed the luciferase gene on an incoming *attP* or *attB* plasmid construct and cotransfected with a plasmid expressing the integrase. We then measured the expression of luciferase over a time course. These results revealed a statistically significant increase in long-term luciferase expression in human 293 cells when the gene was transfected on a plasmid bearing *attB* in the presence of integrase (Fig. 2). The stability of luciferase expression over the 4-week time course, in contrast with its rapid extinction in the absence of integrase and an *attB* site, was consistent with integrase-mediated integration of the luciferase gene into the chromosomes. The dependence of the reaction on integrase and *attB* suggested that native sites similar to those of *attP* were being recognized and accessed in the human genome. A statistically significant increase in luciferase expression was not observed when the incoming plasmid carried *attP* (Fig. 2). Similar results were seen when these experiments were carried out with mouse 3T3 cells. These data suggested that sites similar to *attP* were also being recognized in the mouse genome, as confirmed and examined in detail below.

FIG. 2. Luciferase expression after transfection of *att*-luciferase plasmids into unmodified 293 cells. Shown is the time course of luciferase expression for constructs that carried the luciferase gene and either no *att* site (pL), *attB* (pL-*attB*), or *attP* (pL-*attP*) and which were transfected with and without the integrase-expressing plasmid pCMV-Int. Each data point represents the average of three independent experiments, with appropriate standard error bars. A greatly elevated level of luciferase expression indicative of efficient chromosomal integration was seen only when the incoming plasmid carried an *attB* site and integrase was present. Similar data were obtained for mouse 3T3 cells.

Genomic integration detected by selection. The luciferase studies suggested that *attP* sites recognized by the ϕ C31 integrase were naturally resident in the genomes of human and mouse cells. To examine this possibility more closely, we co-transfected unmodified human 293 and mouse 3T3 cells with a plasmid bearing the neomycin resistance marker and either *attB* or *attP*, with and without a plasmid expressing the ϕ C31 integrase. Negative controls included parallel transfections with a non-*att*-bearing plasmid.

The data in Table 1 that were obtained with 293 and 3T3

cells indicate that an elevated frequency of neomycin-resistant colonies was observed in the presence of the *attB*-neo plasmid pNC-*attB* and pCMV-Int. Increases in the range of 5- to 10-fold for colony numbers of the mouse and human genomes suggested that one or more native genomic sites that can interact with integrase and *attB* were present in these genomes and that the frequency of integration at these sites competed favorably with the background of random integration. For the pNC-*attP* plasmid and pCMV-Int, only a slightly elevated frequency of neomycin-resistant colonies, less than twofold, was

TABLE 1. Integration frequency at native and inserted *attP* sites^a

Cell type	% G418 ^r integrants			% G418 ^r and Zeo ^r integrants
	(−) Int (<i>n</i>)	(+) Int (<i>n</i>)	Fold increase	
Human cells				
293	0.008 ± 0.0001 (172)	0.069 ± 0.002 (1,306)	8.6	NA
293P1	0.013 ± 0.0007 (20)	0.18 ± 0.01 (190)	13.8	0.004 ± 0.0003 (32)
293P2	0.013 ± 0.003 (17)	0.15 ± 0.04 (191)	11.5	0.009 ± 0.001 (88)
293P3	0.02 ± 0.002 (48)	0.25 ± 0.02 (386)	12.5	0.011 ± 0.002 (135)
293P4	0.014 ± 0.002 (21)	0.24 ± 0.01 (269)	17.1	0.006 ± 0.001 (50)
Mouse cells				
3T3	0.25 ± 0.005 (136)	1.2 ± 0.2 (253)	4.8	NA
3T3P1	0.27 ± 0.04 (191)	2.6 ± 0.5 (331)	9.5	ND
3T3P2	0.19 ± 0.03 (162)	1.7 ± 0.3 (428)	9.1	ND
3T3P3	0.34 ± 0.01 (74)	4.6 ± 1 (476)	13.5	ND

^a Percent integration was the ratio of the number of colonies obtained to the total number of cells transfected in the absence [(-) Int] or the presence [(+) Int] of an integrase-expressing plasmid. Each value (expressed as a percentage) is the average of at least three independent experiments, with the standard error. The numbers in parentheses represent the actual numbers of colonies obtained. NA, not applicable; ND, none detected.

seen (data not shown), consistent with the results obtained with luciferase.

Characterization of pseudo *attP* sites. Efficient integration of plasmids bearing *attB* in unmodified human and mouse cells in the presence of ϕ C31 integrase suggested that sites with significant identity to *attP* were present in these genomes. In order to test this prediction, we cloned the integration sites by plasmid rescue. Total DNA was prepared from pools of hygromycin-resistant 293 and 3T3 cells that had received pCMV-Int and pHZ-*attB*, which carried *attB* and the genes for hygromycin, zeocin, and ampicillin resistance. From the elevations in integration frequency observed above, ~80 to 90% of such integrations were expected to occur at native sites resembling *attP*. The genomic DNA was cut with sets of restriction enzymes with compatible ends that do not cut in pHZ-*attB*, was self-ligated, was transformed into *E. coli* cells, and was selected for the ampicillin and zeocin markers on pHZ-*attB*. Selection for two intact selectable markers on the plasmid served to limit the recovery of random integrants among the rescued plasmids. Sixty-seven colonies rescued from two independent pools of transfected human cells and 120 colonies rescued from four independent pools of transfected mouse cells were analyzed. Restriction enzyme digests and agarose gel analysis of the rescued plasmids displayed fragments expected for pHZ-*attB*, as well as various fragments contributed by the flanking genomic DNA for each integration event. This analysis showed that some of the rescued plasmids had the same fragment patterns within a species, consistent with repeated integrations into the same sites, while other plasmids were present as single occurrences.

In order to determine whether the integration events were integrase mediated, whether they were precise, and whether the integration sites possessed DNA sequence similarity to *attP*, we sequenced both of the *attB*-genome junctions of all 187 of the rescued plasmids. We used primers from within *attB* to sequence outward on both sides. In all cases, a junction was detected that fused half of *attB* with non-plasmid sequences. By colocalizing the crossover junctions with *attB*, these results confirmed that the integration events were mediated by integrase and were not random.

This DNA sequence information also allowed us to evaluate how many different genomic integration sites were present in our collection and the distribution of integrants among them. In the case of the human cells, one of the sites, which we designated human pseudo *attP* site A or ψ A, received 32 of the 67 integration events analyzed at the sequence level. Because integration events from the same pool were not necessarily independent, some of the multiple occurrences could have come from the same mammalian clone. However, as noted below, the exact integration junctions at pseudo *attP* sites often differed slightly at the base pair level, so many integrants at ψ A were demonstrably independent. Four other human sites, ψ B, ψ C, ψ D, and ψ E, occurred two or three times each. Another 26 sites were seen once each. Therefore, at least 31 locations in the human genome could be accessed by the ϕ C31 integrase, although one of the sites, ψ A, was used preferentially. Because of the large number of single occurrences, it is unlikely that we saturated the number of potential integration sites.

In mouse cells, of the 120 sets of integration junctions sequenced, 12 occurred at the same genomic site, which we

designated mouse ψ A. Another 20 sites received two to nine integration events each, and of these, at least four sites were recovered between independent pools of cells. Thirty-six sites received a single integration. Therefore, we identified 57 pseudo *attP* integration sites accessed by the ϕ C31 integrase in the mouse genome, with some sites more favorable than others. As in human cells, the large number of single events suggested that we did not saturate the number of potential integration sites. However, in both species, the fraction of multiple occurrences suggested that we were beginning to approach the total number of integration sites.

The same *attB* primers used to identify the integration junctions allowed us to sequence 100 to 200 bp into the genomic flanking sequences, which we undertook for four of the most favored integration sites. In the cases of three human and one mouse integration sites, we used the genomic sequence on both sides of the integrated plasmid to develop appropriate PCR primers to retrieve the corresponding intact genomic fragment from the genome. The DNA sequences of these four regions and their GenBank accession numbers are reported in Fig. 3A. The sequence of human ψ A was also determined for a 254-bp region of DNA prepared from genomic DNA derived from human diploid fibroblasts and was found to be identical to the corresponding sequence we obtained from 293 cells. All four genomic sequences encompassing these pseudo *attP* sites were present in the human and mouse databases.

A comparison of these sequences in the region of the crossover point with that of *attP* allowed evaluation of the level of identity. For both the human and mouse ψ A sites, over the 25-bp region centered over the 3-bp TTG core of the minimal *attP* site (7), the identity was 56% (14 out of 25), while it was 40% in this region for human ψ C and 24% for ψ D. We designate these sites as pseudo *attP* sites, meaning that while they differ in sequence from wild-type *attP*, they apparently possess enough sequence identity to trigger ϕ C31 integrase-mediated pairing and reaction with *attB*. The 56% identity at human ψ A, possibly in combination with favorable context features that are currently undefined, may lead to preferential use of this site in human cells by the ϕ C31 integrase.

For those integrants at the four pseudo *attP* sites for which we determined the sequence of the unmodified genomic region, the crossover sequences on both sides of the integrant enabled us to define the precision of the recombination event. Several crossover junctions at human ψ A are reported in Fig. 3B and are representative of the many examples we sequenced. In some cases, the recombination junction between the incoming *attB* and the genomic pseudo *attP* was completely precise, with no loss or gain of bases. However, in most cases, a small deletion of 1 to 11 bp was present at the junction of *attB* and genomic pseudo *attP* sequences. The small deletions affected bases from the genome, *attB*, or both, varying for different integrants. Thus, integration events at pseudo *attP* sites were not completely precise at the sequence level, differing slightly between individual rescue events. The slightly different sequences present at the integration junctions confirmed the independence of many of the integration events cloned. The slight imprecision suggested that the relatively poor match of the pseudo *attP* site with wild-type *attP* impeded the ability of the ϕ C31 integrase to complete the integration reaction. This

FIG. 3. DNA sequence characterization of pseudo *attP* sites and recombination junctions in human and mouse cells. (A) The wild-type ϕ C31 *attP* site is shown on top. The 3-bp TTG crossover region is marked, and arrows delineate the imperfect palindromes. Beneath are the DNA sequences for four pseudo *attP* sites. The GenBank accession numbers for these pseudo *attP* sites are in parentheses below. Human ψ A (AF333429) is the primary site of ϕ C31 integrase-mediated integration in unmodified human cells, and human ψ C (AF333430) and ψ D (AF333431) are two secondary human integration sites. Mouse ψ A (AF333432) is a site used repeatedly in mouse cells. A 100-bp DNA sequence centered on the *attP* core is presented here, with identities with *attP* shown filled in. (B) The sequences of the crossover region between the human ψ A pseudo *attP* site (lowercase letters) and the *attB* site on the incoming plasmid (uppercase letters) for eight representative rescued events from neomycin-resistant pools of 293 cells. Colons represent bases missing in the novel joint. These small deletions ranged from 0 to 11 bp among 120 junctions sequenced. The bases affected by deletion in one or more junctions are highlighted in the composite junction sequence at the bottom of the diagram. (C) The sequences of the crossover region between an inserted wild-type *attP* site and the *attB* site on the incoming plasmid, which were derived from representative neomycin-resistant or neomycin- and zeocin-resistant colonies that were shown by PCR to have integrated at inserted *attP* site. The contribution from *attP* is shown in lowercase letters, while the contribution from *attB* and the 3-bp crossover region are in uppercase letters. The crossover junctions were perfect in each case (8 out of 8).

result contrasted with recombination between wild-type *attB* and *attP* sites, which was always precise to the base (see below).

Integration at inserted *att* sites. In order to determine whether the ϕ C31 integrase could also mediate efficient integration at wild-type *attB* and *attP* sites placed into the context of mammalian chromosomes, we created human and mouse cell lines carrying *attB* or *attP* recognition sites and then co-transfected the lines with two plasmids, one expressing the ϕ C31 integrase and one carrying the complementary *att* site along with the neomycin resistance selectable marker.

Plasmids containing either *attB* or *attP* and the hygromycin resistance marker (pHZ-*attB* and pHZ-*attP*; Fig. 1B) were transfected into human 293 cells and mouse 3T3 cells, and hygromycin-resistant colonies were picked and expanded. We verified that most of the clonal lines carried an integrated *att* site by the presence of an appropriate PCR band. Southern blottings on four of the 293 *attP* lines showed that pHZ-*attP* appeared to be integrated in a single copy at one location in

each case. Human and mouse *attP*- and *attB*-containing 293 and 3T3 cell lines were transfected with a plasmid bearing the complementary *att* site and the neomycin resistance gene (pNC-*attB* or pNC-*attP*; Fig. 1C), with and without pCMV-Int. The integrated plasmid also contained a promoterless zeocin resistance gene that would have been activated by insertion of the incoming plasmid at the *att* site by virtue of the CMV promoter on the incoming pNC-*attB* plasmid (Fig. 1D).

For the four 293 and three 3T3 lines carrying an integrated *attP* site that we analyzed, we observed an elevated number of colonies when both the integrase and an *attB* plasmid were introduced and neomycin selection was carried out (Table 1). The increases were in the range of 10- to 20-fold for four human and three mouse *attP* cell lines. In the case of cell lines carrying an inserted *attB* site, only a modest (approximately twofold) increase in integration frequency was observed (data not shown), consistent with the results for pseudo *attB* sites.

On replicate plates, selection was carried out with both neo-

mycin and zeocin. The double selection should have ensured that integrants were located at the inserted *attP* site, where provision of a promoter by the incoming pNC-*attB* activated the zeocin resistance gene. Modest numbers of such colonies were observed in the case of the human cells and were not detected in mouse cells (Table 1). However, evidence about the location of the integrants, described below, indicated that the zeocin selection underestimated the number of integrants at the inserted *attP* sites, probably due to poor expression of the zeocin resistance gene under control of the CMV promoter, especially in mouse cells. In an integrated position, the CMV promoter can be silenced over time. This effect is less severe in 293 cells, which express the adenovirus E1A protein (6). Integrations at *attP* in mouse cells were detected by PCR.

Distribution of integration events. Because the overall integration frequency in human cell lines in which we had inserted an *attP* site was within twofold of the frequency in unmodified 293 cells, we expected that the integration events would be distributed between the *attP* site and the pseudo *attP* sites. To examine this point, we determined where the incoming *attB* plasmid was integrated by PCR analysis of DNA extracted from independent colonies derived from each of the four *attP* 293 cell lines. We examined 96 of the colonies selected with neomycin alone and 96 of the colonies selected with both neomycin and zeocin. In each case, 24 colonies were chosen at random from each of the four *attP* cell lines. The neomycin- and zeocin-resistant group represented 2 to 6% of the number of colonies seen with neomycin selection. They were expected to be located at the inserted *attP* sites because the incoming *attB* plasmid provides a promoter for the zeocin resistance gene (Fig. 1D). As expected, essentially all (95 out of 96) neomycin- and zeocin-resistant colonies showed a PCR band indicative of site-specific integration into the inserted *attP* site.

By the same assay, 14 out of 96 (14.6%) of the colonies selected with neomycin alone were found to be integrated at the inserted *attP* site. In eight of these integrants where a PCR fragment was detected that indicated site-specific integration at *attP* (four neomycin resistant only and four neomycin and zeocin resistant), these fragments were sequenced to assess the precision of the integration events. The results showed that integrase-mediated recombination that was exact to the base had occurred between *attP* and *attB* in all cases (Fig. 3C). The remaining 82 out of 96 of the neomycin-resistant colonies were not at the inserted *attP* site. Approximately 5 to 10% of these were expected to be random integrants, because the integrase-mediated reactions were approximately 10- to 20-fold above background. The rest of the integrants were presumably located at pseudo *attP* sites. As described above, the human ψ A *attP* pseudo site is preferentially used. We analyzed the 82 non-*attP* integrants with PCR primers that detect integration at ψ A and determined that 5 (5.2%) of the integrations occurred at this *attP* pseudo site. The remaining integrase-mediated integrants were expected to be distributed among the other pseudo *attP* sites.

DISCUSSION

This study demonstrates that site-specific integration into the chromosomes of living mammalian cells can be obtained by using the ϕ C31 phage integrase, which carries out precise

recombination between *attP* and *attB* recognition sites of minimum lengths of 30 to 40 bp (7). In cell lines containing an inserted *attP* site, we detected integration into the genome at frequencies approximately 10- to 20-fold above the spontaneous background frequency of random integration in mouse and human cells. Furthermore, in the absence of an inserted *attP* site, we detected integrase-mediated recombination at endogenous sites in the genome at frequencies $\sim$ 5- to 10-fold above the background of random integration. These integration events were shown to occur at sets of native sequences having partial sequence identity to *attP*, which were termed pseudo *attP* sites.

These integration frequencies compare favorably to other site-specific integration systems described to date. Our frequencies are $\sim$ 2 to 3 orders of magnitude higher than those typically involved in targeting by using homologous recombination. The frequency of integration by homologous recombination appears to be in the vicinity of 10^{-6} for most mammalian cells (2, 26), though it has been reported that this frequency can be increased up to 20-fold by using completely isogenic DNA (22). By making a double-strand break at the target site, homologous recombination efficiency can be improved by $\sim$ 100-fold or more (3, 14). However, a means to generate such a break in endogenous sequences is currently lacking. The integration frequency mediated by the ϕ C31 integrase is approximately 10- to 100-fold higher than Cre-mediated integration at an inserted wild-type *loxP* site (18) or integration mediated by FLP at an inserted FRT site (12). Use of specially designed *loxP* cassettes designed to limit the reverse reaction has resulted in higher integration frequencies (5, 16, 17). However, this strategy is not applicable when using endogenous sequences as targets. The ϕ C31 integrase appears to be more efficient in mammalian cells than phage integrases of the tyrosine-catalyzed site-specific recombinase family, such as integrases from phages λ and HK022 (8, 11). Integration mediated by retroviruses and some transposases can be efficient (1, 28), but it takes place at random, leading to mutagenesis and inconsistency of gene expression.

This study demonstrated that in the presence of the ϕ C31 integrase, a plasmid bearing *attB* will be efficiently integrated into mammalian genomes. In unmodified human cells, nearly 90% of the integration events will be integrase mediated and will be distributed among a set of pseudo *attP* sites. By sequencing the junctions at 67 rescued integration events, we identified a hierarchy of pseudo *attP* sites, some of which were used repeatedly. The pattern of single and recurring sites in this collection of 31 different pseudo *attP* sites suggests that the total number of pseudo *attP* sites may be between 10^2 and 10^3 , with some sites significantly preferred over others. While integration is occurring at multiple sites, the level of specificity is still dramatically increased over that of random integration. Since the genome contains approximately 3×10^9 bp, many of which are presumably available for random integration, restriction to approximately 10^2 integration sites represents a gain of several orders of magnitude in specificity.

In mouse cells, where 120 integration events were sequenced, we identified 57 pseudo *attP* sites, 21 of which were recurrent, creating a similar picture of pseudo *attP* site frequency. Since the identity of these sites to the wild-type *attP* was $<60\%$, one would expect a similar number of sites in any

mammalian genome. The number of pseudo *attP* sites found in mammalian genomes suggests that pseudo *attP* sites for the ϕ C31 integrase also exist in smaller genomes, such as those of important model organisms like *Caenorhabditis elegans*, *Drosophila*, and *Arabidopsis*. This integrase may become a valuable tool for genetic manipulation of those organisms. Because the ϕ C31 integrase requires no cofactors, it is expected to work well in a broad range of species, including plants, mammals, and other vertebrates and invertebrates, such as insects and worms. Indeed, it is likely that endogenous sites exist for many recombinases. We have documented the occurrence of pseudo *loxP* sites for Cre in mammalian cells (25), and it appears that these sites can be used in vivo in mice, at least under conditions of continuous high-level expression of Cre (19).

In human and mouse cell lines in which we inserted a wild-type *attP* site, the randomly placed *attP* sites competed with the set of native pseudo *attP* sites. From a sample of 96 colonies from four human cell lines carrying *attP*, we found that about 15% of the integrations occurred at the inserted *attP* site, compared to 5% at the predominant human ψ A pseudo *attP* site. Most of the other integrations were expected to be distributed over the other ~ 100 pseudo *attP* sites. The integration frequency at a given *attP* site is presumably the result of the DNA sequence of the site and its chromosomal context, which may influence gene expression and integrase access. It will be interesting to measure the relative integration frequency at *attP* and human ψ A when the chromosomal context of both sites is kept constant. We did not observe any multiple integrations in the same cell in this study, but we are not in a position to rule it out.

We have found that integration of a plasmid bearing *attB* into a chromosomally placed *attP* site is invariably precise, yielding the expected recombination event at the DNA sequence level. This precision reflects effective operation of the enzyme at its *attB* and *attP* recognition sites in many different contexts in mammalian chromosomes. Integration at the pseudo *attP* sites is, in contrast, slightly imprecise at the sequence level, presumably reflecting a less exact reaction when one of the *att* sites differs from the wild-type sequence. This result is expected from what is known of the interactions between the ϕ C31 integrase and its *att* recognition sites (24). The sequence of the *att* site seems to influence the nature of the complex formed, which in turn determines whether or not the reaction will proceed. It is conceivable that with imperfect sites, mammalian repair enzymes may participate in completing the reaction. We observed that when the incoming plasmid bore an *attP* site instead of an *attB* site, we detected less-efficient reactions, both with pseudo *attB* sites or with an inserted *attB* site. We do not have an explanation for this lack of symmetry. It may reflect the order of formation of the complex and the ability of a complex to form on *att* sites located on abundant and exposed incoming plasmid DNA versus *att* sites buried in the chromosomes.

The reaction involving integration of a vector bearing *attB* into a wild-type *attP* site previously inserted in the genome has many applications in genetics and biotechnology, such as positioning incoming genes at the same integration site repeatedly in a given cell line. If a selection scheme is employed, such as the zeocin selection used here, then close to 100% of the selected events will be precisely positioned at the chromosomal

attP site. For species that possess endogenous sequences that are recognized at appreciable frequencies by the ϕ C31 integrase, integration into unmodified genomes is also feasible. This reaction may be valuable in in vivo gene therapy and other applications involving unmodified cells and tissues. This integration reaction would be more valuable if it occurred at an even higher frequency and if the number of target sites were more limited. More stringent target sequence requirements may be characteristic of related integrases found in nature. Additionally, it may be possible to increase enzyme efficiency and change target specificity by directed evolution (21).

ACKNOWLEDGMENTS

Bhaskar Thyagarajan and Eric C. Olivares contributed equally to this work.

We thank Eddie Baba and Andrew Neviaser for technical support and Man-Wah Tan for comments on the manuscript.

National Institutes of Health grants DK55569 and DK58187 provided support to the Calos lab. E.C.O. was supported by a graduate fellowship from the Ford Foundation, D.G. was supported by a graduate training grant from the NIH, and B.T. was partially supported by PHS grant CA09302 from the National Cancer Institute.

REFERENCES

- Anderson, W. F. 1998. Human gene therapy. *Nature* **392**:25–30.
- Capecchi, M. R. 1989. Altering the genome by homologous recombination. *Science* **244**:1288–1292.
- Donoho, G. P., M. Jasin, and P. Berg. 1998. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol. Cell. Biol.* **18**:4070–4078.
- Esposito, D., and J. J. Scooca. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* **25**:3605–3614.
- Feng, Y. Q., J. Seibler, R. Alami, A. Eisen, K. A. Westerman, P. Leblouch, S. Fiering, and E. Bouhassira. 1999. Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J. Mol. Biol.* **292**:779–785.
- Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**:59–72.
- Groth, A. C., E. C. Olivares, B. Thyagarajan, and M. P. Calos. 2000. A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. USA* **97**:5995–6000.
- Kolot, M., N. Silberstein, and E. Yagil. 1999. Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Mol. Biol. Rep.* **26**:207–213.
- Kuhstoss, S., and R. N. Rao. 1991. Analysis of the integration function of the Streptomyces bacteriophage ϕ C31. *J. Mol. Biol.* **222**:897–908.
- Landy, A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58**:913–949.
- Lorbach, E., N. Christ, M. Schwikardi, and P. Droge. 2000. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. *J. Mol. Biol.* **296**:1175–1181.
- Merrihew, R. V., R. G. Sargent, and J. H. Wilson. 1995. Efficient modification of the APRT gene by FLP/FRT site-specific targeting. *Somat. Cell Mol. Genet.* **21**:299–307.
- Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**:391–406.
- Phillips, J. E., and M. P. Calos. 1999. Effects of homology length and donor vector arrangement on the efficiency of double-strand break-mediated recombination in human cells. *Somat. Cell Mol. Genet.* **25**:91–100.
- Rausch, H., and M. Lehmann. 1991. Structural analysis of the actinophage ϕ C31 attachment site. *Nucleic Acids Res.* **19**:5187–5189.
- Sauer, B. 1996. Multiplex Cre/*lox* recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. *Nucleic Acids Res.* **24**:4608–4613.
- Sauer, B. 1994. Site-specific recombination: developments and applications. *Curr. Opin. Biotechnol.* **5**:521–527.
- Sauer, B., and N. Henderson. 1990. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol.* **2**:441–449.
- Schmidt, E. E., D. S. Taylor, J. R. Prigge, S. Barnett, and M. R. Capecchi.

2000. Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. USA* **97**:13702–13707.
20. Stark, W. M., M. R. Boocock, and D. J. Sherratt. 1992. Catalysis by site-specific recombinases. *Trends Genet.* **8**:432–439.
21. Stemmer, W. P. C. 1994. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751.
22. te Riele, H., E. R. Maandag, and A. Berns. 1992. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA* **89**:5128–5132.
23. Thorpe, H. M., and M. C. M. Smith. 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.
24. Thorpe, H. M., S. E. Wilson, and M. C. Smith. 2000. Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕ C31. *Mol. Microbiol.* **38**:232–241.
25. Thyagarajan, B., M. J. Guimaraes, A. C. Groth, and M. P. Calos. 2000. Mammalian genomes contain active recombinase recognition sites. *Gene* **244**:47–54.
26. Vega, M. A. 1991. Prospects for homologous recombination in human gene therapy. *Hum. Genet.* **87**:245–253.
27. Wohlgemuth, J. G., S. H. Kang, G. H. Bulboaca, K. A. Nawotka, and M. P. Calos. 1996. Long-term gene expression from autonomously replicating vectors in mammalian cells. *Gene Ther.* **3**:503–512.
28. Yant, S., L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak, and M. A. Kay. 2000. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat. Genet.* **25**:35–41.