Introduction of a lacZ reporter gene into the mouse int-2 locus by homologous recombination

(Suzanne L. Mansour, Kirk R. Thomas, Chuxia Deng, and Mario R. Capecchi)

Howard Hughes Medical Institute, Department of Biology, University of Utah, Salt Lake City, UT 84112

Communicated by John R. Roth, June 29, 1990 (received for review May 1, 1990)

ABSTRACT We demonstrate that the frequency of gene targeting is unaffected by the length of nonhomologous DNA transferred to a target chromosomal sequence. A result of this finding is that a much wider spectrum of designed genomic alterations is now feasible. As a first application, we inserted a 5.4-kilobase cassette of nonhomologous DNA into the int-2 locus in mouse embryo-derived stem cells by gene targeting. The inserted DNA contained a lacZ gene positioned to create an in-frame fusion with the int-2 protein-coding region. Upon differentiation of these cells to embryoid bodies, the int-2–lacZ fusion faithfully reproduced the expression pattern of int-2 RNA. This ability to target reporter genes, such as lacZ, to specific mouse loci, combined with the ability to move the tagged gene into different mutant backgrounds, may provide an ideal approach for analyzing interactions among genes that participate in a developmental network.

Homologous recombination between an exogenously introduced DNA molecule and its endogenous chromosomal counterpart provides a means to transfer specific sequence modifications, created in vitro, into the genome of an intact cell. Furthermore, if the recipient cell is a pluripotent, mouse embryo-derived stem (ES) cell, then it is possible to transfer that modified gene to the genome of a living mouse. In this way, the effects of designed genome alterations can be monitored in the context of the entire mammalian developmental program (for reviews, see refs. 1–3).

It is commonly believed that the frequency of gene targeting events in mammalian cells is inversely proportional to the length of nonhomologous DNA sequences that are transferred to the chromosomal target (4–7). This belief is based in part on the observation that in cultured mammalian cells, the frequency of intrachromosomal gene conversion is inversely proportional to the length of nonhomologous DNA converted in the recipient DNA sequence (8). However, the recombination substrates that participate in gene targeting and in intrachromosomal gene conversion are different, and therefore the above assumption warranted direct investigation.

In this paper we demonstrate that the frequency of gene targeting at the mouse hypoxanthine phosphoribosyltransferase (hppt) locus is insensitive to the length of nonhomologous DNA transferred to the target chromosomal sequence. This finding vastly increases the spectrum of potential genomic alterations that can be generated by gene targeting. As an initial test of this unexpected observation, we introduced 5.4 kilobases (kb) of nonhomologous DNA containing the genes which encode β-galactosidase (β-gal) and neomycin resistance (neo') into one of the endogenous int-2 alleles in an ES cell line.

int-2 is a protooncogene identified by its activation in mouse mammary tumors due to the nearby insertion of mouse mammary tumor virus DNA (9, 10). int-2 is a member of the fibroblast growth factor gene family (11), and analysis of the pattern of int-2 expression suggests several potential roles for this gene in cell–cell interactions during normal mouse development (12, 13). ES cells provide an in vitro model for some of the early events in mammalian development, and this property facilitates the analysis of certain aspects of the int-2 expression pattern. Upon in vitro differentiation of ES cells, int-2 mRNAs are induced dramatically and are restricted to a subpopulation of extraembryonic-like endodermal cells (12, 14). We show here that the int-2–lacZ fusion faithfully displays this expected int-2 expression pattern.

MATERIALS AND METHODS

Vector Construction. The hppt vectors were constructed by starting with a 9.1-kb fragment of hppt (4) that extends from a Bgl II site in intron 5 to a Bgl II site 3’ of hppt coding sequences. To create pHPPT+8, an 8-base-pair (bp) Xho I linker was inserted into the Scu I site in exon 8. pHPPT+8 and pHPPT+18 were created by inserting a 1-kb fragment from pMC1-NEO (4) and a 3.4-kb fragment from pC4Bgal (15), respectively, into the Xho I linker in pHPPT+8.

The int-2–lacZ replacement vector, pINT-2-LACZ/TK, was constructed by starting with pINT-2-NPa, which contains a 10-kb EcoRI fragment of int-2 genomic DNA (10) disrupted at the Apa I site in exon 1b by the insertion of the polyadenylation signal-containing neo' gene from pMC1-NEO. A DNA fragment from pC4Bgal (15) that contains the Escherichia coli lacZ gene, missing the first seven codons, and followed by the simian virus 40 small t splice and early polyadenylation signals, was inserted into pINT-2-NPa within exon 1b such that the reading frame between the first 21 int-2 codons and the lacZ sequences was preserved, creating pINT-2-LACZ. The herpes simplex virus thymidine kinase (HSV-tk) gene (16) was inserted into pINT-2-LACZ on the 5’ side of the int-2 sequences, creating pINT-2-LACZ/TK.

Electroporation and Isolation of Cell Lines. Linearized plasmid DNAs were introduced by electroporation into the ES cells as described previously (4). Isolation of 6-thioguanine-resistant (6-TG') and G418' plus gancyclovir-resistant (GANC') cell lines was as described previously (4, 16) except for the following changes in the isolation of 6-TG' cell lines. After electroporation of ES cells with the hppt targeting vectors, 2 × 10⁴ ES cells were plated on 100-mm culture plates that contained feeder STO cells in nonselective medium. After 4 days of incubation, ½ of the cells on a single

Abbreviations: ES, embryo-derived stem; β-gal, β-galactosidase; HSV-tk, herpes simplex virus thymidine kinase; GANC, gancyclovir; 6-TG, 6-thioguanine; ', resistant; FDG, fluoroscine β-D-galactopyranoside.
plate were transferred to a fresh feeder layer in growth medium that contained 6-TG at 1 µg/mL. Three days later, the cells from a single plate were again transferred to a new feeder layer in 6-TG medium. Under these conditions, at least 33% of the resulting plates had no 6-TG' colonies, consistent with the Poisson prediction of a single mutational event per plate.

Embryoid Body Formation and Fluorescence Imaging of β-gal Activity and SPARC Antigen. Embryoid body formation was initiated according to Martin et al. (17). Four days after the initial cell clumps were put into suspension culture, the embryoid bodies were assayed in vivo for β-gal activity with fluorescein β-D-galactopyranoside (FDG) as the substrate. Reagents for the assay were purchased from Molecular Probes. The assay was performed as described by the supplier and by Nolan et al. (18). The embryoid bodies were observed in microconcavity slides by using a Bio-Rad confocal laser scanning microscope equipped for simultaneous phase-contrast and fluorescence observation (fluorescein filters). The microscope stage was cooled to 4°C to prevent leakage of fluorescein from cells.

SPARC antigen (19) was detected by using a whole-mount indirect immunofluorescence protocol (C. Davis, D. Holm‐yard, K. Millen, and A. Joyner, personal communication) based on refs. 20 and 33. The primary rabbit anti-SPARC serum (a gift of B. Hogan) was used at a 1:200 dilution. The secondary Texas red-labeled donkey anti-rabbit IgG (Jackson ImmunoResearch) was used at a 1:100 dilution. The stained embryoid bodies were observed by using the confocal microscope equipped with rhodamine filters.

RESULTS

Disruption of the hprt Gene with Inserts of Different Size. To assess the effect of sequence heterology on the frequency of gene targeting, insertions of various lengths were targeted to the X chromosome-linked hprt locus in male ES cells. The targeting vectors illustrated in Fig. 1 each contained a 9.1-kb fragment of the hprt gene that was disrupted by heterologous DNA inserts of 8 bp, 1 kb, or 3.4 kb located within the eighth exon. Homologous recombination between the targeting vectors and the endogenous gene would disrupt the hprt coding sequences, rendering the cells hprt' and thus resistant to the purine analog 6-TG. The linearized vector DNAs were introduced separately into ES cells by electroporation. These cells were then subjected to selection in medium containing 6-TG. We found the frequency of targeted disruption of hprt to be the same irrespective of the length of the heterologous DNA inserted into the endogenous gene (Table 1). In addition to this study of heterologous inserts in the eighth exon, we have also examined the effect of heterologous inserts of 4.3 kb and 12 kb in the third exon on the hprt targeting efficiency. As can be seen from Table 1, comparable targeted disruption frequencies were also obtained with these hprt vectors.

DNA from each independent 6-TG' colony was subjected to Southern blot hybridization analysis to determine whether resistance to 6-TG resulted from targeted disruption of the hprt gene or from spontaneous mutation. Approximately 50% of the 6-TG' colonies arose as a result of targeted disruption of the hprt gene, regardless of which targeting vector was used. The remaining 6-TG' colonies arose by spontaneous mutation of the hprt gene (Table 1). Even though the spontaneous mutation rate at the hprt locus is approximately 1/20th of the gene targeting rate, the spontaneous hprt' mutants accumulate during the selection protocol, whereas gene targeting events occur only over a short period following the introduction of the targeting vector into the ES cells (1). This accounts for the eventual accumulation of approximately equal numbers of targeted and spontaneous hprt' mutations in these experiments.

Southern blot hybridization analysis of the hprt' cell lines with targeted insertions in exon 8 is shown in Fig. 2. Digestion of the parental hprt+ ES DNA with the enzyme Xba I generated an 8.4-kb fragment which hybridized to the hprt probe. Because the three targeted hprt' cell lines analyzed here contain insertions of DNA in the eighth exon, the length of this Xba I fragment increased by the length of the insertion [i.e., 8 bp (not discernible), 1 kb, and 3.4 kb]. The insertion mutations were also characterized by the presence of an Xho I site, not found in the eighth exon of the wild-type gene. The presence of this site was demonstrated by digestion of the mutant DNAs with the enzymes Xho I and Xba I, which generate a 1.8-kb fragment common to the hprt' cell lines (Fig. 2). A similar analysis of the cell lines derived by targeting with the vectors that contain insertions in exon 3 confirmed the presence of the expected disruption in each case (data not shown).

Insertion of a lacZ Reporter Gene into the int-2 Locus. Lactose. 3A illustrates the int-2 targeting vector, pNT-2-LACZNT. This vector contains an int-2-lacZ fusion, along with the neo' and HSV-tk genes. The lacZ coding sequences, excluding the first seven codons, were fused in-frame with int-2 coding sequences. The downstream neo' gene, in the presence of G418, provides a positive selection for insertion of the targeting vector sequences. Similarly, the HSV-tk gene, in the presence of GANC, provides a negative selection to reduce the number of transfected ES cells that contain
nontargeted insertions of the vector DNA (16). Homologous recombination between the int-2 targeting vector and the endogenous int-2 gene would precisely position the lacZ reporter gene within the int-2 cis-regulatory environment and was therefore expected to place lacZ under int-2 transcriptional and translational control (Fig. 3B).

The int-2 targeting vector was linearized as shown and introduced into ES cells by electroporation, and the treated cells were grown in medium containing G418 and GANC. DNA from the doubly resistant cell lines was analyzed by restriction digestion and Southern blot hybridization, using enzymes and probes that distinguish random from homologous integration events. Of 89 cell lines examined, 9 were found to contain an insertion of the lacZ and neo' DNAs in one of the two int-2 alleles (Table 2).

Southern blot hybridization analysis of one of the targeted cell lines, 62-8K, is shown in Fig. 4. Parental ES DNA digested with HindIII and Xho I, or Xmn I alone, and probed with flanking int-2 sequences, not present in the targeting vector (probe A), reveals fragments of 11.2 kb and 6.2 kb, respectively. DNA from the mutant cell line, 62-8K, digested with the same enzymes, reveals fragments of the same size as those from the parental cell line. These are derived from the unmutagenized copy of the int-2 gene. Probe A also detects new fragments of 6.7 kb (HindIII and Xho I) and 11.6 kb (Xmn I) derived from the mutated copy of int-2. The new fragments correspond in length to those predicted from homologous recombination between the targeting vector DNA and the endogenous int-2 gene. As expected, these fragments both hybridize with the neo' probe (probe B). In addition, the lacZ probe (probe C) hybridizes with a 10.2-kb fragment (HindIII and Xho I) and with an 11.6-kb fragment (Xmn I) in 62-8K DNA, confirming the presence of lacZ sequences at the expected site in the mutant allele.

**Table 2.** Gene targeting into the int-2 locus by pINT2-LACZN/TK

<table>
<thead>
<tr>
<th>Exp. electroporation</th>
<th>G418' colonies</th>
<th>G418'+GANC' colonies</th>
<th>G418'+GANC' colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. cells surviving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 x 10^6</td>
<td>5.2 x 10^4</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>5 x 10^6</td>
<td>4.9 x 10^4</td>
<td>43</td>
</tr>
</tbody>
</table>

Aliquots of cells treated with the int-2 vector (Fig. 3A) were subjected to one of three growth conditions: nonselective medium to assess the total number of cells surviving electroporation (approximately 50%), G418 medium to assess the fraction of survivors transformed by the lacZ-neo' vector, or G418 plus GANC medium to enrich for cells containing a targeted disruption of the int-2 gene. Among the G418'+GANC' colonies, targeting events were identified by Southern transfer analysis (see Fig. 4).
entiated cells surrounded by a layer of extraembryonic-like endoderm (17). During this differentiation, \( \text{int}-2 \) expression is induced and its RNA is localized in the outer, endodermal, cell layer (12, 14). Due to the low level of \( \text{int}-2 \) protein product present in mouse embryos and embryo bodies, as well as the difficulty of raising high-affinity antibodies directed against undenatured protein, \( \text{int}-2 \) protein has not been detected directly in these sources (ref. 24 and unpublished observations).

To determine the pattern of \( \text{E. coli} \) \( \beta \)-gal activity in 62-8K cells as well as in the parental cell line, CC1.2, we prepared undifferentiated cultures and simple embryoid bodies from these cell lines and stained them with 5-bromo-4-chloro-3-indolyl \( \beta \)-d-galactoside (X-Gal), a chromogenic substrate of \( \beta \)-gal (25). Consistent with the low level of \( \text{int}-2 \) RNA found in undifferentiated cultures of ES cells (12, 14), very little \( \beta \)-gal activity was detectable in undifferentiated cultures of either cell line (data not shown). In contrast, \( \beta \)-gal activity was readily detectable in embryoid bodies derived from 62-8K cells but not in CC1.2 embryoid bodies (data not shown).

To localize the source of \( \beta \)-gal activity in the embryoid bodies, we stained unfixed 62-8K (Fig. 5 a, b, and c) and control (Fig. 5 d, e, and f) embryoid bodies with a fluorogenic substrate of \( \beta \)-gal, FDG, and examined optical sections of these embryoid bodies for the fluorescent product, fluorescein, using confocal laser scanning microscopy. Fig. 5 shows surface (b) and central (c) optical sections through a typical 62-8K embryoid body. Only the outer, endodermal, cell layer was labeled, although not every cell in that layer showed activity. \textbf{In situ} hybridization to \( \text{int}-2 \) RNA in embryoid bodies shows a similarly patchy pattern (12). The control (CC1.2) embryoid bodies were not significantly labeled with FDG. Fig. 5 shows surface (e) and central (f) sections of a typical CC1.2 embryoid body which had only two or three interior cells that stained positive for \( \beta \)-gal activity. This labeling may reflect endogenous lysosomal \( \beta \)-gal activity in a few disrupted cells.

The endodermal nature of the outer cell layer was confirmed by whole-mount indirect immunofluorescence against SPARC, an endodermal marker (19), followed by confocal laser scanning microscopy. As expected, the labeled cells were confined to the outer surface of the embryoid bodies (Fig. 5 f and g). Thus the pattern of \( \text{int}-2 \)-lacZ expression described above accurately reflects the distribution of normal \( \text{int}-2 \) RNA in ES cell-derived embryoid bodies.

**DISCUSSION**

We have demonstrated that the frequency of gene targeting is not influenced by the length of nonhomologous DNA sequences transferred to the target chromosomal locus. The
size of the heterologous insertions was varied over a range covering three orders of magnitude. As a result of this finding, a wider spectrum of designed genomic alterations is now feasible. Thus, it would be reasonable to attempt substitution of one controlling region of a gene for another, or, as shown here, to replace entire coding regions with another. Also, with respect to homologous recombination, introducing large insertions and generating large deletions are nearly equivalent operations. Therefore, our results suggest that generation of large genomic deletions should also be possible. Overlapping sets of deletions could be used to locate genes within mapping intervals.

Encouraged by our observations on the disruption of the hprt gene, we showed that a lacZ reporter gene/neo cassette 5.4 kb in length could be efficiently introduced by homologous recombination into the endogenous int-2 locus. Direct comparison of these results (Table 2) with previously reported results (table 1 of ref. 16) obtained with an int-2 targeting vector that contained only the neo' gene (1.1 kb), shows that the targeted disruption frequencies using these two vectors are essentially equivalent.

Following in vitro differentiation, the β-gal activity displayed by cells that contain the fusion construct was consistent with the pattern of int-2 RNA observed in situ by Wilkinson et al. (12). int-2 RNA in the latter study and β-gal activity in this study were up-regulated in endodermal cells. In both cases, not all endodermal cells expressed the gene product in question. Further studies will be necessary to determine whether this result is due to heterogeneity in the in vitro produced endoderm as suggested by Wilkinson et al. (12). A determination of the relationship between the marker gene's expression and that of int-2 RNA during actual embryonic development will be feasible when germ-line chimeras that carry the int-2-lacZ fusion are produced.

lacZ fusion genes have been introduced as transgenes at random locations within the mouse genome (see for example, ref. 26). Introduction of the lacZ reporter gene into endodermal loci by homologous recombination may obviate two serious limitations imposed by the standard transgenic technology. First, homologous recombination places the reporter gene among all of the cis-acting control elements that normally mediate expression of the endogenous gene. Often, the locations of these elements are not defined or they may be highly dispersed within a locus, making construction of a transgene that mimics the expression of the endogenous gene difficult or impossible. Second, the expression pattern of the reporter gene at the endogenous locus is not subject to the chromosomal position effects that can result from integration of the transgene at ectopic sites. As a consequence of these effects, independent transgenic animals that contain the same lacZ fusion construct integrated at different sites may exhibit different patterns of β-gal activity, even when the known cis-acting elements are included in the construct. A consensus expression pattern must be inferred from comparisons of the patterns observed in a number of independent transgenic animals.

As discussed above, the pattern of β-gal activity in het-

erozygous int-2-lacZ mice may provide a sensitive tool for following int-2 expression in the developing embryo. A unique aspect of this experimental design, which does not apply to conventional transgenes, is that the lacZ-neo cassette activates the int-2 gene. Therefore, homozygous int-2-lacZ mice will exhibit the phenotype normally associated with the absence of the int-2 gene product. As a result, the same cells responsible for the int-2′ phenotype should also be labeled with the lacZ gene product. This property may facilitate the interpretation of the resultant phenotype at the cellular level.

lacZ reporter genes have provided an important tool for the analysis of early development in Drosophila and Caenorhabditis elegans (27–32). Such reporter genes have been particularly instructive in the analyses of interactions among genes that participate in developmental regulatory networks. By moving the recombinant construct into different mutant backgrounds, the effects of other genes on the expression of the fusion gene can be easily assayed. Comparable molecular genetic analysis of mouse development is just beginning. It is our hope that the ability to target reporter genes to loci that mediate developmental decisions will contribute to this emerging analysis.

We thank Carol Lenz, Marjorie Allen, and Susan Tamowski for their expert technical assistance with the culturing of the ES cells, Carl Thummel for pC49gal, Brigid Hogan for the anti-SPARC serum, and David Gard for instruction and assistance in using the confocal microscope. S.L.M. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation.