Interstrand crosslink-induced homologous recombination carries an increased risk of deletions and insertions
Vidya S. Jonnalagadda, Tetsuya Matsuguchi, Bevin P. Engelward

Biological Engineering Division, Massachusetts Institute of Technology, 77 Massachusetts Ave., 56-631, Cambridge, MA 02139, USA

Received 4 February 2005; accepted 4 February 2005

Abstract
Homology directed repair (HDR) defends cells against the toxic effects of two-ended double strand breaks (DSBs) and one-ended DSBs that arise when replication progression is inhibited, for example by encounter with DNA lesions such as interstrand crosslinks (ICLs). HDR can occur via various mechanisms, some of which are associated with an increased risk of concurrent sequence rearrangements that can lead to deletions, insertions, translocations and loss of heterozygosity. Here, we compared the risk of HDR-associated sequence rearrangements that occur spontaneously versus in response to exposure to an agent that induces ICLs. We describe the creation of two fluorescence-based direct repeat recombination substrates that have been targeted to the ROSA26 locus of embryonic stem cells, and that detect the major pathways of homologous recombination events, e.g., gene conversions with or without crossing over, repair of broken replication forks, and single strand annealing (SSA). SSA can be distinguished from other pathways by application of a matched pair of site-specifically integrated substrates, one of which allows detection of SSA, and one that does not. We show that SSA is responsible for a significant proportion of spontaneous homologous recombination events at these substrates, suggesting that two-ended DSBs are a common spontaneous recombinogenic lesion. Interestingly, exposure to mitomycin C (an agent that induces ICLs) increases the proportion of HDR events associated with deletions and insertions. Given that many chemotherapeutics induce ICLs, these results have important implications in terms of the risk of chemotherapy-induced deleterious sequence rearrangements that could potentially contribute to secondary tumors.

Keywords: Mitotic homologous recombination; Interstrand crosslinks; Single strand annealing; Direct repeat; Mouse ES cells

1. Introduction
Endogenous and exogenous DNA damaging agents create thousands of DNA lesions per cell each day. Lesions that affect just one strand of the DNA duplex are generally repaired by excision of the damaged nucleotide(s) and replacement using the intact complementary strand as a template [1]. However, if both strands are damaged (e.g., by a double strand break [DSB]), another source of sequence information is required for accurate repair. Homology directed repair (HDR) allows cells to use the undamaged sister chromatid or the homologous chromosome as a template for repair. In addition, HDR can reconstitute broken replication forks by accurately reinserting the broken DNA end into the sister chromatid [2–6].

HDR often occurs via a relatively conservative pathway in which there is a non-reciprocal transfer of sequence information at the site of the original damage, without exchange of flanking sequences (e.g., gene conversion without crossing over). However, some HDR events are associated with crossovers, and crossovers between misaligned sister chromatids inevitably lead to gains and losses of sequence information. In cases where HDR occurs between homologous chromosomes, even if there is accurate alignment, crossovers can lead to loss of heterozygosity along vast stretches of the chromosome (from the exchange point to the telomere).

It is thought that most homologous recombination events are initiated by two-ended DSBs and one-ended DSBs. Two-ended DSBs can be created at any time during the cell cycle by agents that destroy the integrity of the sugar-phosphate
backbone or by agents that induce damage that is subsequen-
tly enzymatically cleaved (e.g., base excision repair of
closely opposed lesions can lead to DSBs [7]). On the other
hand, one-ended DSBs are completely replication dependent,
arising when replication forks encounter lesions that inhibit
progression. While some lesions can directly disintegrate the
fork (e.g., encounter with single strand nicks or gaps), other
lesions can indirectly cause replication forks to breakdown
by stalling progression, which can lead to formation of a
Holliday junction that is prone to enzymatic cleavage (for
excellent recent reviews on this subject, the reader is referred
to [4,5,8]).

For all DSBs, one of the earliest steps of HDR is resec-
tion of the DNA end(s) to create a 3′ overhang that becomes
a nucleo-
protein filament capable of homology searching [2–6]. This
filament can invade homologous sequences and undergo re-
pair synthesis using a homologous duplex DNA as a tem-
plate. In the prototypic break repair model, as first proposed
by Szostak et al., both ends of a two-ended DSB invade
the homologous sequence, forming two Holliday junctions that,
when cleaved, lead to exchange of flanking sequences 50% of
the time [9]. Alternatively, Holliday junction cleavage can be
avoided via synthesis dependent strand annealing (SDSA),
wherein translocation of a Holliday junction releases a DNA
end, which can then anneal to the other end of the break
[10,11]. In addition to the prototypic break-repair and SDSA
models, DSBs at repeated sequences can also be repaired by
single strand annealing (SSA), a subpathway of homologous
recombination where the resected single stranded regions
simply anneal to one another, without invading a homologous
duplex DNA [12,13]. This pathway inevitably leads to loss of
one copy of the repeated sequences. We define a homologous
recombination event in which there is an associated gain or
loss of sequences to be non-conservative, since the original
arrangement of the DNA has been lost.

It is well established that conditions that lead to increased
levels of HDR are associated with an increased risk of cancer
[14,15]. Indeed, many known carcinogens are potent recom-
binogens, including oxidizing agents, alkylating agents, UV
light, and ionizing radiation [16–18]. Although the mech-
nanisms of recombination initiated by homing endonucle-
ases that introduce two-ended DSBs are well studied (e.g.,
[19–23]), to our knowledge, there are only two studies report-
ing the mechanisms of recombination induced by other types
of DNA damaging agents in mammalian cells [24–25]. We are
particularly interested in the possibility that carcinogens not
only can induce HDR, but that exposure may skew the distri-
bution of events in favor of non-conservative HDR events that
are associated with deletions and insertions. Among known
carcinogens, agents that induce interstrand crosslinks (ICLs)
are highly recombinogenic in mammalian cells [26–29]. ICLs
interfere with normal DNA replication by preventing strand
separation, and it is thought that ICLs induce HDR primarily
by inducing replication fork breakdown, resulting in the for-
mation of one-ended DSBs [30]. This model is supported by
two studies showing that the majority of ICL-induced DSBs
arise in a replication-dependent fashion [31–33]. Thus, by
investigating the types of sequence rearrangements induced
by ICLs, we can learn about the relative risk of deleterious
sequence rearrangements induced by an agent that is thought
to primarily cause replication-dependent one-ended DSBs.

A common approach for studying HDR is to use direct
repeat substrates in which recombination restores expression
of a selectable marker. Many such substrates detect SSA, but
SSA events are not generally discernable from other classes
of recombination events, such as unequal sister chromatid
exchanges. Here we describe the creation and application
of two site-specifically integrated direct repeat substrates to
study both spontaneous and damage-induced recombination
in non-transformed mammalian cells. Both substrates de-
tect gene conversions (with and without crossing over) and
HDR events associated with repair of broken fork. How-
ever, only one of the substrates detects SSA, thus providing
a strategy to estimate the contribution of SSA. We found that
~50–65% of spontaneous recombination events are due to
conservative HDR (e.g., gene conversions without crossing
over), which is consistent with previously published reports
indicating that most spontaneous HDR events are conserva-
tive gene conversions [28,29,34–36]. In addition, we provide
some of the first direct evidence that SSA is a frequent spon-
taneous homologous recombination event at direct repeats
in mammalian cells, which suggests that endogenously in-
duced two-ended DSBs are the underlying cause of a sig-
nificant portion of spontaneous homologous recombination
events. When comparing spontaneous recombination events
to those induced by exposure to an agent that forms ICLs, we
find that most ICL-induced recombination events are non-
conservative, thus putting cells at an increased risk of dele-
terious rearrangements. These results have important impli-
cations to cancer patients who are frequently treated with
chemicals that induce ICLs.

2. Materials and methods

2.1. Enzymes, oligonucleotides, plasmids

Restriction enzymes were from New England Biolabs (Beverly, MA). Hotstart polymerase (Eppendorf Interna-
tional) was used for diagnostic PCR, and Advantage2 poly-
merase (BD Biosciences, CA) was used in construction. Oligonucleotides were from Ambion Inc. (Allston, MA). Tis-
sue culture reagents were from Gibco/BRL. Plasmid pCX-
EGFP was a gift from Okabe et al. [37], and plasmids pBigT
and pROSA26PA were gifts from Soriano [38] and Srinivas et
al. [39]. Oligonucleotide and vector sequences are available
upon request.

2.2. Construction of recombination substrates

The Prf–BamHI fragment in pCX-EGFP was replaced
with a synthetic adaptor carrying NsiI, NotI and XhoI sites,
and the coding sequences for enhanced green fluorescent protein (EGFP) removed by digestion with EcoRI to obtain pCX-NNX. Full length and truncated coding sequences (\(\Delta 3\text{egfp}\), lacking sequences at the 5’ end, and \(\Delta 5\text{egfp}\), lacking sequences from the 3’ end) for EGFP were amplified by PCR from pCX-EGFP using primers carrying Apol sites. In \(\Delta 5\text{egfp}\), this resulted in a removal of 27 bp (of which 15 were in the coding region), which were replaced with 31 bp of unique sequence. In \(\Delta 3\text{egfp}\), this resulted in a removal of 93 bp (of which 81 were in coding sequence), which were replaced with 62 bp of unique sequence. Apol digested PCR products were subcloned into EcoRI digested pCX-NNX to obtain pCX-NNX-EGFP, pCX-NNX-\(\Delta 3\text{egfp}\) and pCX-NNX-\(\Delta 5\text{egfp}\). The expression cassettes (promoter, enhancer, intron, coding and polyadenylation sequences) from pCX-NNX-\(\Delta 5\text{egfp}\) and pCX-NNX-\(\Delta 3\text{egfp}\) were released with Sall and PstI. The cassette were cloned into NotI and XhoI-digested pCX-NNX-\(\Delta 3\text{egfp}\) and pCX-NNX-\(\Delta 5\text{egfp}\), respectively, to obtain pCX-NNX-\(\Delta GI\) and pCX-NNX-\(\Delta GF\). Plasmid pBigT-TpA [39] was derived from pBigT by removal of the 1 kb \(\text{HindIII/NotI}\) fragment. Expression cassettes from plasmids pCX-NNX-EGFP, pCX-NNX-\(\Delta 3\text{egfp}\), pCX-NNX-\(\Delta 5\text{egfp}\), pCX-NNX-\(\Delta 5\text{egfp}\) and pCX-NNX-\(\Delta GF\) were released with Sall and NotI and cloned into pBigT-TpA digested with Sall and NotI. Each of these plasmids was digested with Ascl and PciI to release the corresponding expression cassettes, which were then cloned into pROSA26PA digested with Ascl and PciI. Restriction enzymes were used to linearize all five plasmids pROSA26-EGFP (KpdI), pROSA26-\(\Delta 3\text{egfp}\) (KpdI), pROSA26-\(\Delta 5\text{egfp}\) (KpdI), pROSA26-\(\Delta GI\) (XhoI) and pROSA26-\(\Delta GF\) (XhoI). Linearized plasmids were electroporated into mouse embryonic stem (ES) J1 cells (gift of R. Jaenisch, MIT). Clones selected for resistance to neomycin (DMEM containing 15% fetal bovine serum, penicillin, streptomycin and glutamine, LIC, containing G418) were selected using PCR and Southern blotting for correct targeting of the expression cassettes into the ROSA26 locus.

2.3. Southern blotting and PCR analysis

For Southern blotting, genomic DNA was digested with EcoRV and probed using a previously described small fragment that lies outside of the targeting vector [38,39]. Correctly targeted clones were further analyzed by Southern blotting using at least two independent diagnostic digestions followed by probing with EGFP sequences. Only clones that showed both correct targeting and a single integration event were used in subsequent studies. PCR testing of targeting of insert to the ROSA26 locus was done using primers that yield a product of 1.2 kb exclusively at the targeted allele [38] (see also http://www.fhcrc.org/lab/soriano/rosa26.htm). To analyze recombinant clones, PCR primers were designed to exclusively amplify EGFP, \(\Delta 3\text{egfp}\) and \(\Delta 5\text{egfp}\), as described in Section 3 (primer sequences are available upon request). Template was added to two reactions that were always performed in parallel, one to assay for EGFP, and the other to assay for \(\Delta 3\text{egfp}\) and \(\Delta 5\text{egfp}\).

2.4. Flow cytometry for analysis and isolation of recombinant cells

Trypsinated cells were resuspended in OptiMEM and analyzed on a Becton Dickenson FACScan flow cytometer (excitation 488 nm, argon laser). Fluorescent cells were sorted using a MoFlo cytometer (Cytometry Inc.; excitation 488 nm, argon laser; emission 580/30). Live cells were gated using forward and side scatter. Recombinant cells were sorted into gelatin coated tissue culture plates and single clones expanded for PCR and Southern blotting.

2.5. Rate of spontaneous recombination

Independent cultures were initiated with \(\sim 2 \times 10^4\) cells (or fewer) and the frequency of fluorescent recombinant cells was determined by flow cytometry after 48–72 h. Samples with frequencies of recombinant cells consistent with the presence of a fluorescent cell at the time of plating were excluded from subsequent analysis, and the rate of recombination and standard deviations were calculated as previously described using the MSS Method of Maximum Likelihood [40].

2.6. Toxicity and frequency of recombinant cells in ICL-exposed cultures

Cells (\(\sim 0.5-2 \times 10^4\)) were plated in gelatinized six-well plates. After 24 h, the cells were exposed to Mitomycin C (MMC) in DMEM for 1 h. Samples were analyzed by flow cytometry after 48–72 h (\(\sim 3–4\) population doublings in control cells). Population growth relative to untreated control cells was estimated by Coulter Counting.

2.7. Sister chromatid exchange analysis

Sister chromatid exchanges were counted in metaphase spreads as previously described [41]. Briefly, cells were grown for 24 h in medium containing BrdU, and colcemid was added for the last 3 h. Trypsinated cells were suspended in hypotonic solution, fixed in Carnoy’s solution, and dropped onto slides. Metaphase spreads were stained with Hoechst 33258 and Giemsa. Over 20 metaphase spreads were analyzed for each sample.

3. Results

3.1. Recombination substrate construction and targeted integration

To create homologous recombination substrates, we modified an expression cassette for EGFP [37] (Fig. 1A)
Fig. 1. Construction of direct repeat recombination substrates. (A) Features of the EGFP expression cassette. Expression is driven by the chicken β-actin promoter with the cytomegalovirus enhancer; pA, polyadenylation sequence. Drawn to scale where the total length = ~3 kb. (B) Expression cassettes targeted to the ROSA26 locus. Arrows indicate the cassette, gray regions indicate coding sequences and black boxes indicate deletions (deletions not to scale). (C) Diagram showing the targeting approach for the ROSA26 locus, as has been previously described [39]. S.A. and L.A., short arm and long arm of ROSA26 genomic sequences; Neo, neomycin; DTA, diphtheria toxin A; insert, each of the five different inserts is shown in ‘B’. P1 and P2, primers that yield a 1.2 kb product exclusively for the targeted allele. The probe used for Southern blotting is indicated, and falls outside of the targeting vector, yielding a 3.8 kb EcoRV fragment unique to the correctly targeted allele. (D) Examples of PCR and Southern blotting results for correctly targeted clones. Left: PCR identification of targeted clones showing a 1.2 kb product diagnostic of the targeted allele. Right: Representative Southern blotting showing EcoRV digested genomic DNA from a correctly targeted clone. (E) Images show phase contrast and fluorescence microscopy. Flow cytometry plots indicate the relative fluorescence intensity for 515–545 nm (FL1) vs. 562–588 (FL2). The R2 region was delineated to capture most fluorescent cells while excluding non-fluorescent cells.

to create two different truncated egfp expression cassettes. DNA sequences for amino acid residues shown to be essential for fluorescence [42] were deleted, such that the coding sequence lacks 15 bp at the 5' end in ΔSegfp (including the start codon), and 81 bp at the 3' end in ΔSegfp (Fig. 1(B)). The truncated coding sequences are flanked by >500 bp of identical sequences that include an intron, polyadenylation signal sequences, and a promoter known to yield high levels of expression in mice. These deleted cassettes were then ligated in tandem in two arrangements. In ΔGI (deleted green internal), ΔSegfp is downstream of ΔSegfp such that the deletions are positioned internally, relative to the coding sequences (Fig. 1(B)). In ΔGF (deleted green flanking), ΔSegfp is placed upstream of ΔSegfp, such that the deletions are flanking the coding sequences (Fig. 1(B)). The rationale behind this approach is that SSA can restore a full length EGFP in ΔGI, but not in ΔGF (Fig. 2(D)). Therefore, by comparing the rates of recombination in this pair of substrates, we can estimate the rate of SSA. It is well established that the spontaneous rate of recombination depends upon the locus of integration [36]. Therefore, in order to reveal the extent of SSA, ΔGI and ΔGF needed to be integrated into an identical locus. Using a previously described targeting system [38,39], each of the three control cassettes and the two recombination substrates were individually ligated into a ROSA26 targeting vector (Fig. 1(C)) and electroporated into mouse embryonic stem (ES) cells. Correctly targeted clones were identified by a diagnostic PCR reaction that yields a 1.2 kb fragment uniquely present in the targeted alleles (Fig. 1(C) and (D), left). Correct targeting in all five clones was also confirmed by the appearance of a 3.8 kb EcoRV fragment by Southern blotting (a representative clone is shown in Fig. 1(D), right), and by additional blotting with EGFP sequences to identify clones carrying a single integration event (data not shown).

Expression of EGFP was assessed in the five targeted clones by fluorescence microscopy and by flow cytometry. Microscopic examination of wild type cells shows that none of the cells are significantly fluorescent, whereas nearly 100% of the cells expressing the EGFP coding sequence are brightly fluorescent (Fig. 1(E), top images). When analyzed by flow cytometry, a plot of relative fluorescence intensity per cell shows a wide range of natural fluorescence for wild type ES cells (Fig. 1(E), top left). Cells expressing EGFP fluoresce
GENE CONVERSIONS WITHOUT CROSSOVERS

(A) ΔGI
- Δsegfp as donor
- Δsegfp as donor

(B) ΔGI
- Δsegfp as donor
- USCE
  (GC-CO)

(C) ΔGI
- Δsegfp as donor
- RFR

(D) ΔGI
- Δsegfp as donor
- SSA

Fluorescent Recombinant Clones

(F) 740 bp EGFP
- 415 bp Δsegfp
- 250 bp Δsegfp

(E) Targeted Clones

(G) Number of Recombinant Clones

48% GC w/o CO

68% GC w/o CO
The fluorescence-positive R2 region was delineated to exclude all non-fluorescent cells while capturing most of the fluorescent cells. No fluorescent cells were observed in populations of either ΔSegfp and ΔSegf, as expected (Fig. 1(E), middle). However, rare fluorescent cells were detected among ΔGI ΔGF cells, both by flow cytometry and microscopy (Fig. 1(E), bottom), indicating that recombination of full length EGFP occurs only when mitotic homologous recombination is possible between the ΔSegfp and ΔSegf cassettes.

To test the stability of EGFP expression at the ROSA26 locus, a positive control EGFP expressing clone was maintained in culture and periodically analyzed by flow cytometry. We did not observe any significant change in the percentage of fluorescent cells over the course of more than 3 weeks of continuous culturing (>30 doublings; data not shown), indicating that EGFP is stably expressed at this locus.

### 3.2. Distribution of spontaneous homologous recombination events in ΔGI and ΔGF

Restoration of full length EGFP in ΔGI or ΔGF cells can occur via various mechanisms of homologous recombination, and the outcomes can be classified broadly into two groups: conservative gene conversions (in which there is no associated crossing over), and non-conservative recombination events that are inevitably associated with deletions and insertions (Fig. 2). In conservative recombination at the ΔGI and ΔGF substrates, sequence information is transferred in a non-reciprocal fashion (gene conversion) and the overall arrangement of the DNA is conserved (Fig. 2(A)). The two major HDR pathways thought to cause gene conversions are SDSA and the prototypic break-repair model (if the Holliday junctions are resolved symmetrically). In contrast, alternative classes of recombination events result in a loss or a gain of one of the repeats in ΔGI and ΔGF, respectively. For example, EGFP can be restored by crossover of flanking sequences during unequal sister chromatid exchange (also called gene conversion with crossovers), resulting in a deletion in ΔGI and an expansion in ΔGF (Fig. 2(B)). In addition, replication fork repair (e.g., reconstitution of a broken replication fork in a homologous-recombination dependent fashion) or long tract gene conversion similarly results in a deletion in ΔGI or an expansion in ΔGF (Fig. 2(C)). Finally, fluorescent recombinant cells that result from SSA can only arise in the ΔGI clone, since SSA in ΔGF produces a doubly mutant expression cassette (Fig. 2(D)).

In order to classify recombination events, PCR primers were designed to specifically detect full-length EGFP, Δsegfp and Δsegf. The presence of a full-length EGFP coding sequence was ascertained using primers that anneal to the undeleted 5′ and 3′ ends (Fig. 2(E)). During construction of the recombination substrates, unique sequences were inserted in place of each of the deleted regions within both Δsegfp and Δsegf. Using primers that anneal to these unique sequences, combined with primers that anneal to shared coding sequences, PCR products specific to Δsegfp and Δsegf can be obtained (Fig. 1(E)). Each of the five targeted clones was then analyzed in parallel reactions, one to test for the presence of full length EGFP, and the other to assay for Δsegfp and Δsegf. As can be seen in Fig. 2(E) (lower panel), the positive control clone carries only full length EGFP (lane 1), each of the negative controls carries either Δsegfp or Δsegf (lanes 2 and 3), and populations of unrecombined ΔGI and ΔGF cells show the presence of both Δsegfp and Δsegf, but not full length EGFP sequence, as expected (lanes 4 and 5).

To broadly classify clones as having undergone rearrangements that appear to be either gene conversions without crossovers or deletion/insertion events, we isolated spontaneous recombinants from multiple independent cultures by fluorescence activated cell sorting. Cells were expanded in culture, and analyzed for the presence of full length EGFP, Δsegfp, and Δsegf coding sequences. Examples of representative recombinant clones are shown in Fig. 2(F). For ΔGI, clones that have undergone a gene conversion without crossover have both full length EGFP and either Δsegfp or Δsegf (Fig. 2(A) and (F), lanes 1 and 2). Alternative recombination pathways in ΔGI result in a deletion, detectable by the presence of full length EGFP coding sequences and the absence of both Δsegfp and Δsegf (Fig. 2(B)–(D) and (F), lane 3). For ΔGF, gene conversions without crossovers are classified using the same approach as for ΔGI (Fig. 2(A) and (F), lanes 4 and 5), however alternative pathways result in...
an expansion, revealed by the presence of full length EGFP and both Δ35gfP and Δ35gfP (Fig. 2B, C) and (F), lane 6). Clones that did not fit these characteristics were classified as complex events. Eight independent clones were also analyzed by Southern blotting, and the arrangement of the DNA was exactly as expected based on the results of the PCR analysis (data not shown). This PCR approach thus provides an effective strategy for delineating the major classes of homologous recombination events that reconstitute EGFP coding sequences, making it possible to assess the proportion of recombination events due to gene conversion without crossovers versus other classes of events under spontaneous and damage-induced conditions.

To obtain a random sampling of spontaneous recombinant clones, ≥ 80 independent cultures of AGI and AGF were created by plating a low numbers of cells (∼20,000) and allowing cells to expand over the course of approximately 1 week. To assure that only recombinants that arose during expansion were assessed, we eliminated cultures in which the frequency of recombinant cells indicated the existence of a pre-existing recombinant in the original inoculum (e.g., cultures with a frequency of 1/20,000). Spontaneous recombinant fluorescent cells were isolated from each culture by flow cytometry, and up to four independent clones were expanded in culture, and analyzed by PCR, as described above. If two recombinant clones from the same culture showed the same PCR result (which was often the case), it was counted as a single event, since recombinant cells in the same culture could have been derived from the same lineage. We tallied 33 and 20 independent recombination events in AGI and AGF, respectively. Note that we found that this diagnostic approach yielded highly consistent results, such that there was perfect concordance in over a dozen clones that were analyzed independently in a blinded fashion.

For AGI we found that 48% of recombinant clones had undergone gene conversions without crossovers, in which Δ35gfP donated sequences to Δ35gfP in a non-reciprocal fashion, or vice versa (Fig. 2G left; Al and A2). Similarly, we found that a significant proportion (65%) of the spontaneous ΔGF recombinants had undergone gene conversions without crossovers (Fig. 2G right; Al and A2). Note that 15% of the spontaneous recombinant ΔGF clones had undergone complex recombination events (Fig. 2G right; X). Thus, taken together, we found that ∼50–65% of the recombination events were gene conversions without crossovers in AGI and AGF, respectively, which is consistent with previous studies in which it has been reported that 50–100% spontaneous HDR events are gene conversions (the range in frequencies likely depends on intrinsic features of the recombination substrates and the cell type being studied) [28,29,34–36].

For both clones, gene conversions in which Δ35gfP acted as the donor were detected somewhat more frequently than those in which Δ35gfP acted as the donor (Fig. 2G). This result is likely due to the fact that a longer stretch of non-homologous sequence needs to participate in gene conversion events in which Δ35gfP acts as the donor, which is consistent with previous studies showing that gene conversions are increasingly suppressed by lengthening inserted non-homologous sequences [43]. In the context of these studies, the preference for Δ35gfP over Δ35gfP as the donor in gene conversions in both AGI and AGF indicates that sequences upstream and downstream of the recombination substrates (e.g., promoter sequences for nearby genes) do not impose a significant position effect (e.g., we did not see a preference for the upstream cassette acting as the donor, for example).

3.3. SSA is a commonly detected spontaneous recombination event in mouse ES cells

SSA events will only be detected in the AGI substrate (Fig. 2D). On the other hand, unequal sister chromatid exchanges and replication fork repair (Fig. 2, classes B and C) are detectable in both AGI and AGF. Therefore, the extent of SSA can be estimated by comparing the proportion of non-conservative recombination events that appear as either deletions (AGI) or expansions (ΔGF). We found that deletions and expansions accounted for 52% of the recombinants isolated from AGI and 20% of the recombinants from AGF. Therefore, by subtraction, we estimate that ∼30% of spontaneous recombination events are due to SSA (Fig. 2G; subtraction of the BC classes of AGF from the BCD classes in AGI).

As an independent measure of the frequency of SSA, we compared the spontaneous rates of homologous recombination at the AGI and ΔGF substrates. We found that the rate of spontaneous HDR for two independent AGI clones was significantly higher than that of two independent ΔGF clones (Fig. 3A); p < 0.05. This result is consistent with SSA events that are detectable for AGI, but not for ΔGF. To assess the global frequency of spontaneous homologous recombination, we assessed sister chromatid exchanges in metaphase spreads from one of each of the ΔGF and ΔGI clones (Fig. 3B).
We found that there was no significant difference between these clones, indicating that a higher rate in the ∆GI clone is not due to intrinsic hyper-recombination. Taking the average spontaneous rates for each pair of independent clones, we found that there is a ∼45% excess rate of spontaneous recombination in the ∆GI clones, which is attributable to SSA.

3.4. Mechanisms of ICL-induced homologous recombination

To explore the spectrum of recombination events induced by exposure to an agent that is thought to primarily induce one-ended DSBs, we exposed cells to MMC, an agent that creates ICLs. With increasing doses of MMC, we observed increasing toxicity (Fig. 4(A)). At 500 ng/ml, we observed a ∼2.1- and a ∼2.4-fold increase in the frequency of recombination events in ∆GF and ∆GI, respectively (Fig. 4(B)), thus demonstrating that exposure to MMC induces recombination in these clones. To determine the mechanisms of damage induced HDR, fluorescent recombinant cells were isolated from multiple independent MMC-exposed cultures, and the arrangement of the recombined DNA was analyzed by PCR, as described above.

In clones derived from MMC-exposed populations, the proportion of insertions in ∆GF increased significantly, from 35% to 60% (from 7/20 to 18/30) (Fig. 4(C) and (D)). These non-conservative events include complex events, which rose from 15% to 30% (from 3/20 to 9/30). Among ∆GF clones that had undergone gene conversions without crossovers, the frequency of gene conversions in which ∆egfp acted as the donor cassette was somewhat higher, which was also the case for spontaneous events. Among recombinants isolated from MMC-exposed ∆GI cultures, we observed a significant increase in the proportion of recombination events associated with deletions, rising from 52% to 90% (from 17/33 to 17/19) (Fig. 4(C) and (D)); note that there were no complex events among the ∆GI recombinant clones.

It is noteworthy that it is possible that not all of the recombinant clones isolated from MMC-exposed cultures resulted from MMC induction. Based upon the frequency of recombinant cells present prior to plating (and taking into consideration the plating efficiency and the degree of MMC-induced toxicity), the expected proportion of MMC-exposed cultures carrying pre-existing recombinant cells is <1/10, which is not likely to significantly affect the results of these studies. However, it is possible that spontaneous recombinants arose during expansion of the MMC-exposed cultures, after the drug had been removed from the media. Consequently, if anything, we have underestimated the extent to which the spectrum of recombination events shifts following MMC exposure. Given that in both ∆GF and ∆GI we observed a statistically significant shift from gene conversions without crossovers to alternative pathways associated with insertions and deletions (p < 0.05; Fig. 4(D)), we conclude that MMC exposure increases the risk of deleterious recombination events, not just by causing...
a general increase in the frequency of recombination, but also by increasing the proportion of recombination events that are associated with deletions and insertions.

4. Discussion

Although homologous recombination is known to cause sequence rearrangements that contribute to cancer, it is not yet known what causes spontaneous homologous recombination in mammals. The two major classes of lesions thought to initiate HDR events are one-ended DSBs, which arise as a consequence of replication fork breakdown, and two-ended DSBs, which can potentially arise in a replication-fork independent fashion. We are interested both in what causes spontaneous HDR, and in the possibility that exposure to exogenous DNA damaging agents might alter the proportion of HDR events associated with large scale sequence rearrangements. To explore these issues, we created recombination substrates that reveal the relative proportions of the major classes of recombination events. Using these substrates, we delineated the spectra of spontaneous events and events induced by exposure to an agent that induces ICLs, which inhibit replication fork progression. Two of the major findings of these studies are that (1) a significant proportion of spontaneous recombination events are consistent with induction by two-ended DSBs, and (2) exposure to an agent that induces ICLs significantly alters the spectrum of recombination events, causing a shift toward events that are associated with deletions and insertions.

In order to delineate the major classes of spontaneous and damage-induced recombination events, we created a pair of matched direct repeat recombination substrates. It is noteworthy that the design of a recombination substrate can profoundly influence the spectrum of observed events. For example, some substrates cannot detect short-tract gene conversions (e.g., [44]), while others have been designed to exclude SSA (e.g., [36]). Here, we describe the creation of substrates that detect all of the major classes of HDR events. Furthermore, by site specifically integrating a matched pair of substrates (wherein one detects SSA and the other does not), the contribution of SSA to spontaneous non-conservative recombination events can be revealed. The use of fluorescence as a marker for recombinant cells makes it possible to use flow cytometry to rapidly quantify recombinant cells in an automated fashion, and by inserting unique sequences into the deletion sites, the major classes of HDR can be delineated by PCR. Finally, by performing these studies in mouse ES cells, we can learn about recombination in untransformed mammalian cells.

SSA was first proposed as a model for repairing DSBs based on studies of extrachromosomal recombination products in mammalian cells [45], and it is now well established that SSA is a kinetically and genetically separable subpathway of HDR in S. cerevisiae [13,46]. Although it is widely held that SSA is also a subpathway of HDR in mammalian cells, the frequency of spontaneous SSA in mammalian cells had not previously been reported, most likely because application of a single recombination reporter cannot distinguish SSA from other types of non-conservative HDR events. In previous studies by Bollag and Liskay, a matched pair of direct repeat substrates were created wherein one substrate detects SSA and the other does not [36]. Although the results of this previous study are consistent with the results presented here, showing an excess rate of ~50% due to SSA, their findings were inconclusive due to noise associated with random integration [36]. Additional substrates have been designed to detect SSA in mammalian cells, however, they have not been used to study spontaneous recombination events (i.e., these studies have focused on DSBs induced by a homing endonuclease) (e.g., [23]). Here, we have used a pair of targeted substrates to estimate the prevalence of spontaneous SSA events, and we found that ~30–50% of spontaneous recombination at these direct repeat substrates is due to SSA. It is expected that the frequency of SSA at a direct repeat will be highly variable, depending on both intrinsic factors (e.g., the length of the repeats) and extrinsic factors (e.g., the locus of integration). Nevertheless, these studies are valuable because they show that SSA can occur quite frequently in mammalian cells, which is of particular interest due to the fact that all models of SSA show that these events are initiated by one-ended DSBs at replication forks. We reasoned that if most spontaneous homologous recombination events are stimulated by one-ended DSBs at replication forks, then exposure to an agent that creates ICLs would yield a spectrum similar to that of spontaneous HDR. However, the spectrum of spontaneous and ICL induced recombination events were significantly different from one another, which suggests that spontaneous recombination is often initiated by two-ended, rather than one-ended DSBs.

ICLs are thought to induce recombination by inhibiting replication fork progression. One possibility is that forks breakdown when single strand breaks are introduced at ICLs by ERCC1-XPF [47,48]. Another possibility is that DSBs arise when fork progression is inhibited [49], presumably due to fork regression leading to formation of Holliday junctions that are subsequently cleaved by resolvases, as has been shown to be the case in Escherichia coli [50]. The observations that most ICL-induced DSBs occur in a replication-dependent fashion [31–33] are consistent with induction of one-ended DSBs at replication forks. We reasoned that if most spontaneous homologous recombination events are stimulated by one-ended DSBs at replication forks, then exposure to an agent that creates ICLs would yield a spectrum similar to that of spontaneous HDR. However, the spectrum of spontaneous and ICL induced recombination events were significantly different from one another, which suggests that spontaneous recombination is often initiated by two-ended, rather than one-ended DSBs.

Intriguingly, we found that there was a modest but reproducible increase in the susceptibility of the ΔGI clone to MMC-induced recombination compared to the ΔGFI clone, suggesting that exposure to MMC increases the frequency SSA. This was unexpected, since ICLs are not thought to...
directly induce two-ended DSBs. One possibility is that some ICLs indeed lead to two-ended DSBs in a replication-independent fashion, which is supported by previous work in which pulse-field gel analysis shows a small but significant proportion of ICL-induced DSBs arising in arrested cells [31]. Another possibility is that if two forks converge at the same ICL, both can breakdown to create a two-ended DSB. Finally, it is also possible that some of the recombination events observed in the MMC-exposed cultures were not induced directly by ICLs, but instead indirectly resulted from either the associated redox potential of MMC which can lead to oxidative damage [51], or from a change in the state of the cells that renders them prone to HDR (perhaps due to the associated toxicity of the exposure) [52]. The latter is consistent with studies showing that thymidine induces HDR in the absence of detectable DSBs [53].

Several previous studies have investigated the effects of exogenous DNA damage on the spectrum of recombination events in mammalian cells. Two of particular relevance to this study investigated the effects of UV damage and phenytoin exposure, a chemical that induces reactive oxygen species [24,25]. In the UV study, it was found that most of the UV-induced recombination events were non-conservative [24]. Given that UV lesions are thought to induce HDR by inhibiting replication fork progression [54], these results are consistent with the work reported here. However, it should be noted that most of the spontaneous recombination events at this particular substrate were also non-conservative, so it did not appear that UV exposure significantly changed the proportion of non-conservative recombination events [24]. An additional study also supports the possibility that HDR induced by inhibition of replication forks is associated with exchanges [55], however the substrate used in this study is not amenable to comparisons between conservative short-tract gene conversions (without associated crossing), and other types of non-conservative HDR events. In a separate study of phenytoin-induced HDR, most of the induced events were conservative [25]. These results are consistent with the fact that oxidative damage can induce two-ended DSBs, and that enzymatically induced two-ended DSBs have been shown in several studies to primarily induce conservative gene conversion events [19–21,24]. Thus, together with the results presented here, it may be the case that exposure to agents that inhibit replication fork progress are associated with an increased risk of deleterious sequence rearrangements, though additional studies are clearly necessary to fully explore this possibility.

In this work, we have used direct repeat substrates to study the proportions of conservative and non-conservative recombination events. Although most HDR events in normal mammalian cells are not likely to occur at repeated sequences, these classifications nevertheless shed light on the prevalence of various mechanisms of recombination, some of which are more likely to be associated with deleterious rearrangements than others. For example, it is likely that SDSA is the underlying mechanism of a significant proportion of the conservative recombination events in these studies [3]. SDSA is a relatively safe mechanism of HDR, since SDSA prevents crossovers, and crossovers between homologous chromosomes can cause LOH from the exchange point all the way to the telomere. In contrast, the non-conservative recombination events that appear as insertions and deletions at AGF and ΔGI are always associated with either crossovers or SSA, the former being associated with a risk of LOH, and the latter inevitably leading to loss of sequence information. Therefore, knowledge about the relative proportions of gene conversions without crossovers and recombination events associated with deletions and insertions yields valuable information about the extent to which the cell is performing HDR in a fashion that could lead to deleterious sequence rearrangements.

It is now broadly accepted that HDR processes are critical for maintenance of genomic integrity. Although HDR is essential for correct reinsertion of broken ends that arise during DNA replication, it is not known how often broken forks are the cause of spontaneous recombination events. Here, we show for the first time that SSA, a subpathway of HDR that is initiated by two-ended DSBs, is a common spontaneous recombination event in mammalian cells, which suggests that two-ended DSBs are the underlying cause of a significant proportion of spontaneous homologous recombination events. These events provide a valuable framework for future studies aimed at identifying the specific classes of DNA lesions that are likely to cause spontaneous two-ended DSBs. In addition, in this work we have shown that ICL exposure causes a significant shift in the spectrum of homologous recombination events, favoring non-conservative pathways that are associated with deletions and insertions. Loss of heterozygosity, deletions, insertions and translocations all can arise as a consequence of sequence exchanges during HDR. The results presented here therefore provide new information about how exposure to DNA damage can contribute to tumorigenic sequence rearrangements that potentially contribute to secondary tumors.

Acknowledgements

We are grateful to Dr. Andrew Engelward for help with statistical analysis for rate calculations, and to Dr. J. Nickoloff for helpful discussions and careful critique of this work. We thank Dr. E. Spek for suggestions in the design of the substrates, and to Jeffrey Loh and Jessica Lee for their valuable technical support. We thank the MIT Cancer Center for access to their flow cytometry core. We thank M. Okabe, P. Soriano, S. Strainas, and F. Costantini for DNA vectors, with special thanks to the Soriano laboratory for exceptional technical resources made available for ROSA26 targeting. We thank Dr. R. Jaenisch for the J1 ES cells. This work was supported by R01CA79827 and R33 CA84740 with partial support from ES02109, P01-CA26735, and grants from Burroughs Wellcome Fund.

D.G. Taghian, J.A. Nickoloff, Chromosomal double-strand breaks


