Deletions and Insertions in the p53 Tumor Suppressor Gene in Human Cancers: Confirmation of the DNA Polymerase Slippage/Misalignment Model

Marc S. Greenblatt, Arthur P. Grollman, and Curtis C. Harris

ABSTRACT

We analyzed all published deletions and insertions in the p53 gene to assess the relevance of mutagenesis models. Almost all deletions and insertions can be explained by one or more of the following DNA sequence features: monotonic base runs, adjacent or nonadjacent repeats of short tandem sequences, palindromes, and runs of purines or pyrimidines (homocopolymer runs). Increased length of monotonic runs correlates positively with increased frequency of events. Complex frameshift mutations can be explained by the formation of quasi-palindromes, with mismatch excision and replication using one strand of the palindrome as a template. Deletions and insertions in the p53 tumor suppressor gene may reflect both spontaneous and carcinogen-induced mutagenesis.

INTRODUCTION

The somatic mutations that accumulate during multistep carcinogenesis include major chromosomal events (such as translocations, allelic deletions, and gene rearrangements) and point mutations (such as base substitutions and small deletions or insertions; Ref. 1). The relative frequencies of different classes of mutations vary among cancer-related genes. For example, mutations found in the APC tumor suppressor gene are most often short deletions or insertions, whereas activating mutations in the ras proto-oncogene are base substitutions (1). The largest data base of somatic mutations in human cancers exists for the p53 tumor suppressor gene (2, 3). All classes of mutations are found. Point mutations, often accompanied by deletion of the second allele, are the most common event. The majority of p53 point mutations are base substitutions, but 10–15% of both somatic and germ line p53 mutations are small, intragenic deletions or insertions (1, 4). This data set is large enough to test hypotheses regarding the in vivo mechanisms of intragenic deletions and insertions.

In vitro mutagenesis studies have established that DNA sequence context is the most important factor in determining the specificity of deletions and insertions (reviewed in Refs. 5 and 6). Almost all small deletions and insertions in model systems occur at monotonic runs of two or more identical bases, or at repeats of 2–8-bp DNA motifs, either in tandem or separated by intervening nucleotides. Several mechanisms have been proposed. The most well-studied mechanism, first proposed by Streisinger et al. (7) in 1966, involves slippage or misalignment of the template DNA strands in iterated bases during replication. This leads to either base deletion (if the nucleotides excluded from pairing are on the template strand) or insertion (if they are on the primer strand); replication “fixes” these mutations in the daughter cells (5, 6). Misalignment can occur “spontaneously” in monotonic runs or can be initiated by a spontaneous or adduct-induced base misincorporation that creates a context for slippage (5, 8, 9).

Misalignment may also occur in repeat sequences that are not monotonic runs. When a repeat sequence mispairs with a tandem (adjacent) or nontandem (nonadjacent) complementary motif, the loop formed by the intervening oligonucleotide sequence may be deleted or duplicated (10–12). Such repeats have been identified at the sites of spontaneous and ionizing radiation-induced deletions (13, 14) and at the sites of deletions in synthetic DNA sequences with added bases (9). Loops may also form between nonadjacent inverted repeats (palindromes), leading to deletion of the loop and sometimes to complex frameshift mutations involving both deletion and base substitution (12, 15–17). A review of short (<20-bp) germ line deletions and insertions in genes associated with human genetic disease confirmed that these DNA sequence features are the most important factors determining the specificity of these events (11).

Quantitative mutational spectrum analysis in vitro shows that: (a) the frequency of frameshift mutations increases as the length of the monotonic run increases; and (b) single-base deletions are much more frequent in vitro than are single-base insertions (reviewed in Ref. 5). The proposed explanation is superior stability of the hydrogen bonding of extensive repeats, and G:C-rich regions, and intermediate DNA structures associated with deletions. DNA polymerase infidelity may lead to spontaneous frameshift errors; different polymerases vary in the frequency of errors and the ratio of deletions:insertions (5). The surrounding sequence context is another major factor affecting the likelihood of misalignment mutations (9, 18–20).

Limited analysis of smaller p53 mutation data bases has confirmed the presence of some of these high-risk sequence features, including monotonic and tandem repeats, at the sites of p53 somatic mutations (10). Long runs of purines or pyrimidines, known as homopolymers (e.g., AGAGAG, CCTTCC, and AAAGGG) have also been noted at sites of p53 deletions (21). We updated and reviewed the data base of all reported p53 mutations (2) and analyzed the subset of all deletions and insertions in more detail to assess: (a) the extent to which the mechanisms determined in experimental systems could account for somatic p53 deletions and insertions; (b) how patterns of observed p53 deletions and insertions might refine existing models; and (c) whether relative frequencies of deletions and insertions provide information to distinguish between spontaneous or carcinogen-induced events.

MATERIALS AND METHODS

The p53 mutation data base has been described and is available via E-mail from the European Molecular Biology Library (2). We updated the data base to include all p53 mutations published before September 1994, as identified by searches of MEDLINE and Current Contents. All intron and exon sequences at the sites of, adjacent to, and up to 40 bp away from insertions and deletions were examined, in some cases with the aid of the Gene Jockey sequence analysis program (Biosoft, Cambridge, United Kingdom). Regression and \( \chi^2 \) analyses were performed using standard statistical software.

RESULTS

Since the publication of our original p53 data base (2), the prevalence of deletions and insertions among all p53 mutations has decreased from 13 to 11%. This reduction probably reflects a detection
bias, because most recent reports have examined only exons 5–8, in which missense mutations are more common, and have neglected exons 2–4 and 9–11, in which the majority of mutations are deletions and insertions (1). The data base contains 370 deletions and insertions in the p53 gene. Of these, 21 detected by sequencing of cDNA involve entire exons or large deletions at splice sites, indicating aberrant mRNA splicing. Because splicing mutations often result from either point mutations in splice sequences at the intron-exon junction or dysregulation of the splicing mechanism and not from loss or gain of DNA, these mutations were not classified as DNA deletions or insertions. The patterns and DNA sequence features of the remaining 349 deletions and insertions (or “events”) were analyzed.

Sizes and Relative Frequencies of p53 Deletions and Insertions. Simple deletions constituted 74% (259 of 349) of all events (71% of one bp, 76% of two bp, and 82% of ≥3 bp). Simple insertions constituted 22% (78 of 349) of the events; the remaining 3% (12 of 349) were complex frameshift mutations that combined deletion, insertion, and base substitution. In 49% (166 of 349, 118 deletions and 48 insertions) only one bp was affected; 10% (34 of 349, 27 deletions and 7 insertions) involved 2 bp; and 39% (137 of 349, 113 deletions and 24 insertions) altered 3 or more bp. Only six intragenic deletions in p53 larger than 30 bp and unassociated with splice sites were reported; the largest was 137 bp.

DNA Sequence Features at Sites of Deletions and Insertions. We calculated the proportions of all DNA deletions and insertions attributable to each of the following mechanisms (Table 1 and Fig. 1): monotonie base runs, repeats of short tandem sequences, palindromes (inverted repeats of dyad symmetry), and runs of four or more purines or pyrimidines (homocopolymer runs).

Almost all [341 (98%) of 349] deletions and insertions can be explained by one or more of these DNA sequence features. The most common sequence motifs seen at the site of deletions or insertions are monotonie runs (two to five consecutive identical bases), present at the sites of 77% (269 of 349) of all events; only 46% of all p53 bp are within these runs (odds ratio for an event occurring at a monotonie run, 3.95; 95% confidence interval, 3.00–5.20; P < 0.0001). Repeated tandem sequences, either adjacent to or distant from each other, encompass or flank 30% of deletions or insertions, and palindromes flank another 5%. Homocopolymer runs of 4 or more bp occur at the sites of 56% of events, but in only 3% are they the only associated feature. Homocopolymer runs of 4 or more bp encompass 42% of all bp in p53-coding exons (odds ratio for an event occurring at a homocopolymer run, 1.75; 95% confidence interval, 1.38–2.23; P < 0.0001).

Table 1. DNA sequence features associated with deletions and insertions of p53 in human cancers

<table>
<thead>
<tr>
<th>Sequence features associated with events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class of event</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>All</td>
</tr>
<tr>
<td>All deletions</td>
</tr>
<tr>
<td>All insertions</td>
</tr>
<tr>
<td>Complex frameshifts</td>
</tr>
<tr>
<td>1-bp deletions</td>
</tr>
<tr>
<td>1-bp insertions</td>
</tr>
<tr>
<td>2-bp deletions</td>
</tr>
<tr>
<td>2-bp insertions</td>
</tr>
<tr>
<td>3+-bp deletions</td>
</tr>
<tr>
<td>3+-bp insertions</td>
</tr>
</tbody>
</table>

Fig. 1. Sequence features associated with p53 deletions and insertions. MONO RUN, monotonie base run; TAN RPT, tandem repeat sequence; HCP RUN, homocopolymer run of four or more bases; NONTAN RPT, nontandem repeated sequence (nonadjacent repeat sequence flanks event); PAL, palindrome.
Fig. 2. Relationship between length of monotonic repeat and frequency of deletions and insertions in p53.
A, deletions increase in a linear, positive relationship with length of repeat \(R = 0.930\) for 1 bp (○), \(0.864\) for all (●). B, insertions increase in frequency only at monotonic runs of five bases.
ty-five mutations occur at nonadjacent tandem repeats, and 18 occur complete sequence between the repeats. Those between inverted re-
tandem sequences usually include one full copy of the repeat and the
bp, and few are between 4 and 9 bp. Deletions or insertions flanked by
bp. Of sequences deleted or inserted between palindromes, however,
three-fourths of the sequences deleted or inserted between tandem
at inverted repeats (palindromes). In all but one of these events, the
bias in detection of deletions is unlikely to affect these results.

In the absence of exogenous mutagens, the prevalence of CpG site
transitions varies widely among tumor types, generally correlating nega-
tively with carcinogen association (1).

Because DNA polymerase infidelity alone can produce frameshift
mutations, deletions and insertions also have been suggested as a
marker of spontaneous mutation (10). However, chemical carcinogens
and their adducts can induce frameshift mutations in vitro (6, 18) and,
therefore, may contribute to frameshift mutations in human cancers.
We compared the prevalence of deletions and insertions among tu-
mors to determine whether they correlated either positively or nega-
tively with carcinogen exposure and could thus be used as an indicator
of endogenous or exogenous mutations.

Deletions and insertions are most common in tumors of the thyroid
(17%), head and neck (15%), breast (15%), and sarcomas (15%); they
are least common in melanoma (5%) and other skin cancers (8%) and
colon (8%), hepatocellular (8%), and cervical carcinomas (8%). Both
of these lists contain tumors that have been frequently (thyroid, head
and neck, skin, hepatocellular, and cervical) and infrequently (sarco-
mas and colon) associated with environmental mutagens, and both
contain tumors that frequently (thyroid and colon) and infrequently
(head and neck and hepatocellular) contain CpG site transitions,
markers of mutations unassociated with exogenous mutagens (1).

The prevalence of deletions and insertions varies in tobacco-assoc-
ated cancers (e.g., cervix, 8%; lung, 10%; bladder, 12%; and head
and neck, 15%). The hot spot region from codons 151–159 is affected
in 16% of 36 lung cancer deletions and insertions, including 3 around
codon 157. Codon 157 is a hot spot for base substitutions in lung
cancer (1); this sequence may be the target of an important promuta-
genic, tobacco-associated adduct.

DISCUSSION

This analysis confirms many of the prior observations of deletions
and insertions in human disease (10, 11, 21), highlights the relevance
of many of the mechanistic models for deletion and insertion derived
from in vitro mutagenesis studies (5, 6, 9), and suggests refinement of
some models.

The regression curve of the frequency of deletions and insertions
versus the length of a monoton linear run shows that as the length of
a run increases, the probability of deletions (especially 1-bp deletions)
increases in a linear relationship. However, a different relationship is
noted for insertions. A run of repeated bases must reach a threshold of
5 bp before insertions are likely to occur in the p53 gene. This
threshold pattern has been demonstrated for other systems in vitro,
such as deletions in DNA polymerase α replication of free DNA (5).
Sequence context surrounding runs of identical length can contribute
to differences in deletion and insertion frequency (5, 19, 20); there-
fore, a simple length-frequency relationship is difficult to interpret.
However, the same sequences in p53 are at risk for both deletions and
insertions; the fact that their incidence-versus-length curves differ

### Table 2 Features of deletions and insertions flanked by palindromes and tandems

<table>
<thead>
<tr>
<th>Palindromes</th>
<th>Tandems</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>18</td>
</tr>
<tr>
<td>Deletions</td>
<td>14</td>
</tr>
<tr>
<td>Insertions</td>
<td>4</td>
</tr>
<tr>
<td>Length of paired sequence (%)</td>
<td></td>
</tr>
<tr>
<td>2 bp</td>
<td>0</td>
</tr>
<tr>
<td>3 bp</td>
<td>17</td>
</tr>
<tr>
<td>4 bp</td>
<td>39</td>
</tr>
<tr>
<td>5 bp</td>
<td>22</td>
</tr>
<tr>
<td>&gt;5 bp</td>
<td>22</td>
</tr>
<tr>
<td>Length of deletion (%)</td>
<td></td>
</tr>
<tr>
<td>1–3 bp</td>
<td>7</td>
</tr>
<tr>
<td>4–9 bp</td>
<td>14</td>
</tr>
<tr>
<td>10–20 bp</td>
<td>43</td>
</tr>
<tr>
<td>&gt;20 bp</td>
<td>36</td>
</tr>
<tr>
<td>Length of insertion (%)</td>
<td></td>
</tr>
<tr>
<td>1–3 bp</td>
<td>25</td>
</tr>
<tr>
<td>4–9 bp</td>
<td>0</td>
</tr>
<tr>
<td>10–20 bp</td>
<td>0</td>
</tr>
<tr>
<td>&gt;20 bp</td>
<td>75</td>
</tr>
<tr>
<td>Segment deleted (%)</td>
<td></td>
</tr>
<tr>
<td>Both primers</td>
<td>14</td>
</tr>
<tr>
<td>Intervening loop only</td>
<td>21–29</td>
</tr>
<tr>
<td>One primer + loop</td>
<td>0–7</td>
</tr>
<tr>
<td>Partial primer + loop</td>
<td>57–71</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
<tr>
<td>Segment inserted (%)</td>
<td></td>
</tr>
<tr>
<td>Both primers</td>
<td>50</td>
</tr>
<tr>
<td>Intervening loop only</td>
<td>25</td>
</tr>
<tr>
<td>One primer + loop</td>
<td>0</td>
</tr>
<tr>
<td>Partial primer + loop</td>
<td>25</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
</tbody>
</table>
suggests that the intermediate events involved in their generation differ. One potential explanation is that repair of short heteroduplex DNA structures predisposing to deletions is less efficient than that of those leading to insertion (see below).

The proportion of p53 deletions and insertions that represent spontaneous events due to DNA replication infidelity versus mutagen-induced changes remains unclear. Frameshift mutations that seem to form spontaneously in genomic DNA may be generated by DNA polymerase-induced errors in replication or may arise as a result of endogenous or exogenous damage to template DNA. Endogenous DNA damage is generated by reactive oxygen species (22), products of lipid peroxidation (23, 24), or other metabolic reactions that alter DNA structure (25), thereby complicating the assignment of these mutations to specific environmental agents.

We compared the prevalence of frameshift mutations in tumor types with epidemiological and mutational spectrum evidence supporting carcinogen-induced and endogenous mutagenesis (1); association with cancers linked to carcinogens would strengthen the inference that frameshifts were induced by exogenous damage. Deletions and insertions did not correlate well with these indices. This suggests that differences among tumor types in their prevalence may reflect both endogenous and carcinogen-induced mutagenesis, such as organ-specific carcinogens or tissue-specific properties that regulate replication fidelity. The presence in lung cancers of a hot spot for frameshift mutations in the G:C-rich sequence at codons 151–159, near a base substitution hot spot (1), suggests that carcinogen-mediated frameshifts may occur in vivo.

Several models involving DNA damage-induced template misalignment or slippage during DNA replication have been proposed as mechanisms of frameshift mutagenesis (5, 6, 8, 9, 17–19, 26–29). The following sequence of reactions may lead to base deletions: DNA polymerase encounters a damaged or adducted base at the replication fork, and DNA synthesis is blocked, often transiently, by helix deformation or adduct bulk. If chain extension is delayed, the nucleotide(s) at the 3' primer terminus may pair with a complementary base(s) 5' to the lesion. This misaligned template generates a replication product with deleted bases. The complementary base pairing at the primer-template junction facilitates extension of the misaligned chain. Base substitutions occur when insertion of an "incorrect" base opposite the lesion is followed rapidly by chain extension. Adduct-induced base misincorporation may also increase the stability of a misaligned intermediate.

Frameshift mutations are induced in prokaryotic models by some bulky chemical mutagens, including aflatoxin B1 and arylamines. The ability of a damaged nucleotide to generate frameshift mutations is determined by the nature of the base inserted opposite the adduct, the surrounding sequence context (adduct-induced frameshift mutations occur more frequently in monotonous or tandem base runs), and the rate of translesional DNA synthesis estimated from kinetic analysis (9, 19, 20, 30–32). Studies with chemically defined DNA adducts have established the mutational specificity of some carcinogen-induced lesions in vitro (31, 33). Although the frequency and sites of mutations are not always predictable, the relative propensity for some DNA lesions to promote frameshift deletions in vitro can be estimated (9).
In addition to the effects of specific mutagens, DNA repair capacity is an important factor affecting mutagenesis. The rates and types of spontaneous and induced mutations in *Salmonella typhimurium* can vary dramatically depending on the DNA repair capacity of the strain (17, 34). In eukaryotes, mismatch repair involves a complex of several repair enzymes. One of these, MSH2, binds to palindromes with high affinity, and to large and small insertions and G/T mismatches with moderate affinity (35). MSH2 mutants are defective in the repair of mismatches and small insertion/deletion mispairs (36); thus, this system is likely responsible for repair of the misalignments leading to frameshifts. Germ line mutations in hMSH2 (37, 38) and other mismatch repair enzymes have been linked to hereditary nonpolyposis colon cancer (39–41). Individuals who inherit these mismatch repair abnormalities demonstrate genome-wide insertions and deletions in repetitive DNA sequences (replication errors). This "mutator phenotype" may contribute to intragenic deletions in cancer-related genes (42). However, replication errors and the mismatch repair status in cases of colon or other tumors containing p53 mutations have not yet been reported.

These mismatch repair enzymes also are involved in homologous recombination, as are other proteins, including RecA and its eukaryotic homologues, which bind to homologous DNA regions and promote pairing and strand exchange (43–46). The heteroduplex DNA sequences formed during this process are generally longer than the intermediate structures proposed for the mispairs that lead to frameshifts (47, 48), and it is uncertain how these recognition and repair systems are involved in mutagenesis.

Palindromes and nonadjacent repeat sequences often flank deletions and insertions larger than 3 bp in vitro (5, 6, 16) and in this report. The patterns we observed suggest that the intermediate pathways differ for events initiated by these two sequence motifs. Palindromes generally flank larger deletions, suggesting that short loops between palindromic dyad elements are less likely to be deleted than are longer loops. The presence of short loops in the dyads associated with complex frameshifts (Fig. 3) supports this explanation. The data also indicate that, although direct repeats of 2-bp motifs can predispose to short deletions, more than 2-bp homology is required for both palindromes and nontandem repeat elements to form stable structures. Longer dyad sequences are required for palindromes than for tandems. The deletions associated with tandem sequences usually exactly incorporate the intervening loop and one copy of the repeat, confirming the existing model (11, 13, 14, 49). However, the p53 deletions associated with palindromes usually include portions of both flanking motifs; the exact deletion generally cannot be predicted. This suggests that different deletion mechanisms occur, perhaps involving variable actions of DNA repair proteins (see above). The specific system or mutagen also may influence repair. For example, the spectrum of spontaneous and induced mutations in *S. typhimurium* usually changes when a gene for an error-prone DNA repair enzyme is inserted, but no change is seen in mutations induced by the heterocyclic amine Glu-P-1 (17, 26).

Analysis of complex frameshift mutations confirms that DNA replication and repair mechanisms may be involved in their genesis. Prior studies of complex frameshift mutations induced in *S. typhimurium* generated a model of adduct-mediated mispairing, quasi-palindromes, and DNA replication in palindromes (17). In this system, the presence of an error-prone DNA repair gene enhances the formation of complex frameshifts. We show that complex mutations in p53 are consistent with DNA excision repair and synthesis in palindromes, with one strand of the palindrom used as a template (Fig. 3). This may be triggered by an adduct-induced substitution mutation that increases the stability of a palindrom (17).

Germ line and somatic mutation have been reported in other cancer-related genes, including *APC* in colon cancers (50) and p16
cdkn2 (CDKN2 and MTS-1) in a variety of tumors (51–54). Interestingly, 1-bp deletions in *APC* are less often associated with monotonous runs but are often associated with tandem repeats, perhaps reflecting the influence of impaired mismatch repair due to abnormalities in the *hMLH1* and *hMSH2* genes in colon tissue. Of 24 deletions and five insertions described in p16
cdkn2, only 38% (11 of 29) affect 1 bp, and 48% (14 of 29) affect 3 or more bp. DNA sequence features in this preliminary data base are similar to those seen in p53. Monotonous runs (86%) and tandem sequences (72%) are the most frequent motifs present at the sites of deletions.

Deletions and insertions in p53 are likely underreported, because many investigators concentrate on exons 5–8, the highly conserved DNA-binding region in which most missense mutations occur. More complete analysis of all p53-coding exons would give a more thorough picture of mutational patterns. Missense mutations constitute 77% of mutations in exons 5–8 but less than half of the mutations outside of these regions (1). This is probably due to less frequent selection of missense mutants in exons 2–4 and 9–11. Single-base substitutions in the DNA-binding domain often alter p53-induced transactivation (1). However, single-amino acid substitutions in the transactivation domain (amino acids 20–42) do not affect transactivation; some double substitutions do (55). Frameshifts occurring within or 5' to the DNA-binding domain eliminate the p53 DNA-binding function and thus would alter p53 function and cell cycle kinetics every time they occur. The carboxyl terminus (exons 9–11) of p53 also is important in the oligomerization and nuclear localization of the p53 protein (reviewed in Ref. 56) and regulates its sequence-specific DNA binding (57–60).

ACKNOWLEDGMENTS

We are grateful to Dorothea Dudek for excellent editorial assistance, Peter Shields for assistance with statistical analysis, and Thomas Kunkel for critical review of the manuscript.

REFERENCES


20. Parris, C. N., Levy, D. D., Jessee, J., and Seidman, M. M. Proximal and distal effects


Deletions and Insertions in the $p53$ Tumor Suppressor Gene in Human Cancers: Confirmation of the DNA Polymerase Slippage/Misalignment Model

Marc S. Greenblatt, Arthur P. Grollman and Curtis C. Harris


Updated version

Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/56/9/2130

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.