Hypermutable Bases in the p53 Cancer Gene Are at Vulnerable Positions in DNA Secondary Structures

Barbara E. Wright, Jacqueline M. Reimers, Karen H. Schmidt, and Dennis K. Reschke
Division of Biological Sciences University of Montana, Missoula, Montana 59812

Abstract

A DNA folding analysis indicates that the most hypermutable bases in exons 5, 7, and 8 of the p53 tumor suppressor gene are located immediately next to stems in stable DNA stem-loop structures. On the basis of the highest negative energy (−ΔG) value of the structures containing each mutable base and on the extent to which each base is unpaired during transcription, their relative mutabilities are calculated using a new computer algorithm. These predicted mutation frequencies correlate well with those observed in 14,000 human cancers (R² = 0.76), whereas there is no such correlation (R² = 0.0005) for nearby control bases. The correlation of hypermutable base frequencies with −ΔG values is poor (R² = 0.19), indicating that the extent to which a base is unpaired during transcription is a significant contribution to predicting mutation frequencies.

Introduction

Chromosomal DNA is negatively supercoiled, and supercoiling creates and stabilizes SLSs that are sequence determined, i.e., they form as a consequence of inverted complementary sequences in ssDNA (1, 2). These SLSs contain regions of ssDNA with unpaired and mispaired bases vulnerable to mutation (3, 4). SLSs not only expose specific bases to mutation but also increase the time of their exposure by periodically blocking the progress of transcribing RNA polymerase complexes during transcript elongation. Such pausing is also known to be sequence specific and to promote mutagenesis (5, 6). Transcription enhances supercoiling, which is highly localized (7–9). The direct effects of various kinds of stress on supercoiling have also been shown to be quite specific within the genome and to affect primarily those genes related to the stress involved. Examples are starvation (10), reactive oxygen species (11), metal toxicity (12), virulence (13), and osmotic shock (14). Transcription-enhanced mutations would therefore occur preferentially at pause sites in SLSs of supercoiled DNA formed in the wake of the advancing transcription bubble. If specific types of stress directly target particular areas of the genome for enhanced supercoiling, the formation of secondary structures will be facilitated; if stress first activates transcription, this will in turn cause localized supercoiling. By either scenario, the resulting increase in concentration of SLSs results in localized hypermutation. A mechanism that targets increased mutation rates to genes that must mutate to overcome prevailing conditions of stress has likely been selected during evolution (15). Direct evidence linking increased rates of transcription with localized hypermutation in unpaired bases of SLSs has been obtained recently, in Escherichia coli (4). Here, evidence is presented suggesting that the hypermutable bases in p53 are also uniquely vulnerable to mutation because of their location and position in high negative energy SLSs.

Results and Discussion

Stress-induced Mutations in E. coli. Localized supercoiling and hypermutation should occur as a consequence of starvation-induced derepression and transcriptional activation (15). To test this hypothesis, mutations in derepressed genes of starving E. coli cells were analyzed and compared with mutations in cells not under stress. Zuker’s DNA folding computer program (16) was used to reveal the most thermodynamically stable SLSs and their negative-free energies of folding (−ΔG) for each nucleotide sequence analyzed. The higher the −ΔG value, the more likely a stable SLS will form and result in a pause site. Using Zuber’s program, 30, 40, and 50 nt segments were folded, and the 30 nt folds showed the most specific correlation of maximum peak −ΔG values with the presence of each mutant nt in the derepressed genes. A high −ΔG per se is due primarily to the stability of the stem(s) and does not necessarily indicate the presence of a mutable base. To predict the mutability of a base, it is also necessary to know the extent to which the base is unpaired in all its foldings during transcription (3, 4, 15). We therefore developed an interfacing program that calculates: (a) the percentage of folds in which each base is unpaired; and (b) a Mutability Index for each base, which is expressed as an absolute value and defined as follows: MI = (percentage of folds in which the base is unpaired) × (highest −ΔG of all folds in which it is unpaired) / http://biology.dbs.umt.edu/wright/upload/mfg.html).

Thus, MIs indicate the relative mutation frequencies of bases in predicted SLSs. The highest −ΔG term gives the most weight to the most stable fold containing the unpaired base. That is, primary importance is given to the structure in which the base is exposed for the longest period of time and therefore has the highest probability of mutation. Mutatable bases in derepressed genes are preferentially located in high −ΔG SLSs compared with the unстressed mutable bases. In the mutable bases of transcribing genes, there is a good correlation between relative MIs and mutation frequencies determined experimentally, whereas there is no correlation for mutable bases in unstressed cells (4). The results of these investigations in prokaryotes are consistent with the location of hypermutable bases in high −ΔG SLSs of transcribing genes, suggesting that environmental stressors may cause mutations in such genes. The relevance of these findings to higher organisms has now been investigated in p53. A database of sequences for point mutations in the human p53 gene is available (17) for analyzing DNA secondary structure as a possible cause of hypermutation and cancer.
Hypermutable p53 Codons in High −ΔG SLSs. Exons 5–8 in p53 were analyzed with the DNA folding program to determine whether the hypermutable bases are located in high −ΔG regions of the gene. The size of the folding window used appeared to be less critical than in prokaryotes because folding sequences between 35 and 50 nts gave essentially the same results. About 100 nts containing hypermutable bases (exons 5, 7, and 8) or not containing hypermutable bases (exon 6) were folded in successive, overlapping 43 nt windows, beginning at each fifth nt (data not shown). The average −ΔGs for the regions, including exons 5, 7, and 8, with the hypermutable bases are 6.2, 5.6, and 4.4, respectively, as compared with 3.0 for exon 6 with no hypermutable bases.

The individual −ΔGs for the SLS containing the hypermutable bases in the first and second positions of codons 175, 245, 248, 249, 273, and 282 are given in Table 1 (the third position is silent, except for 249). Also shown in this table are the predicted MIs for each base compared with its observed mutation frequency recorded from 14,000 sequences of the database. As seen in Fig. 1, the correlation between observed and predicted mutation frequencies is significant (R^2 = 0.40) using a 35 nt window (C) and highly significant (R^2 = 0.76 and 0.63) for 43 nt (A) and 50 nt (D) windows, respectively. These results suggest that, in vivo, the average number of nts folded during transcription is in the range of 40–50 nts. The correlation between MIs and mutation frequencies presumably depends upon both of the variables involved, i.e., the −ΔGs of the relevant SLSs, as well as the extent to which the bases are unpaired during transcription. To test this assumption, mutation frequencies were plotted against −ΔG using a 43 nt window (Fig. 1E). The R^2 value of 0.19 is poor compared with 0.76, indicating the contribution of both variables to the predictive value of MIs. The highest −ΔG SLS containing the most mutable base in each of the six codons are shown in Fig. 2. In each case, the unpaired base is immediately next to a stem. Proximity to a stem also characterized the more mutable bases in a similar analysis of prokaryotes.

The last column of Table 1 indicates that 9 of 12 hypermutable bases are located immediately next to a stem, compared with 5 of 24 control mutable bases that were arbitrarily chosen based on their proximity to the hypermutable codons (Table 2). No correlation is seen (R^2 = 0.0005) between predicted and observed mutation frequencies in these control mutable bases (Fig. 1B).

The Mechanism and Consequences of Mutability. The process of creating SLSs during transcription is robust. Regardless of probable continuous variations in the length of the nt segments folded, the location of mutable bases in these structures generally coincide with peak −ΔG values. In this dynamic and ever-changing scene, the stems, which create SLSs, are the invariant part. Thus, the most reliable location for an unpaired base that evolved to be mutable is immediately next to a stem, as seen in most of the hypermutable bases (Table 1). Except for one or perhaps two nts next to a stem, ssDNA segments at the beginning or end of a SLS may, in reality, be paired to their normal complement in double-stranded DNA. As this is not recognized by the computer program, MI values may not be valid for such bases. This would be the case for the C of codon 175 and the G of codon 282 (SLS not shown) but would not change the total mutation frequencies of these codons significantly. Ten of the control bases would be affected in this manner.

The data suggest that exposure time of the hypermutable bases during transcription is a critical variable in the complex of events causing human cancers. Proximity to a stem in a high −ΔG SLS may

Table 1  Comparison of mutation frequencies, MIs, and −ΔGs in p53 hypermutable codons

<table>
<thead>
<tr>
<th>Codon</th>
<th>nt</th>
<th>No. of mutations</th>
<th>MI</th>
<th>−ΔG (kcal/mole)</th>
<th>nts from stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>175</td>
<td>C</td>
<td>47</td>
<td>2.3</td>
<td>7.2</td>
<td>9</td>
</tr>
<tr>
<td>(Arg)</td>
<td>G</td>
<td>558</td>
<td>5.7</td>
<td>8.1</td>
<td>1</td>
</tr>
<tr>
<td>245</td>
<td>G</td>
<td>275</td>
<td>5.1</td>
<td>6.8</td>
<td>1</td>
</tr>
<tr>
<td>(Gly)</td>
<td>G</td>
<td>134</td>
<td>3.7</td>
<td>6.3</td>
<td>1</td>
</tr>
<tr>
<td>248</td>
<td>C</td>
<td>382</td>
<td>5.1</td>
<td>7.3</td>
<td>2</td>
</tr>
<tr>
<td>(Arg)</td>
<td>G</td>
<td>537</td>
<td>5.9</td>
<td>7.3</td>
<td>1</td>
</tr>
<tr>
<td>249</td>
<td>A</td>
<td>44</td>
<td>1.5</td>
<td>5.4</td>
<td>1</td>
</tr>
<tr>
<td>(Arg)</td>
<td>G</td>
<td>74</td>
<td>0.9</td>
<td>5.3</td>
<td>1</td>
</tr>
<tr>
<td>273</td>
<td>C</td>
<td>353</td>
<td>3.4</td>
<td>4.9</td>
<td>0*</td>
</tr>
<tr>
<td>(Arg)</td>
<td>G</td>
<td>507</td>
<td>4.3</td>
<td>5.0</td>
<td>1</td>
</tr>
<tr>
<td>282</td>
<td>C</td>
<td>309</td>
<td>3.6</td>
<td>5.0</td>
<td>1</td>
</tr>
<tr>
<td>(Arg)</td>
<td>G</td>
<td>33</td>
<td>1.3</td>
<td>3.0</td>
<td>8</td>
</tr>
</tbody>
</table>

* The C in codon 273 (next to the red G in Fig. 2) may be essentially unpaired in that two Gs compete to pair with it: one at the end of the lower stem and one at the other end of the loop.

Fig. 1. Linear regression analyses of the data for hypermutable bases (Table 1) and control bases (Table 2) are shown. MIs determined by folding 43 nt windows are plotted against mutation frequencies of hypermutable bases (A) and control bases (B). Windows of 35 nts (C) and 50 nts (D) are also shown for the hypermutable bases. The mutation frequency is also plotted against −ΔG alone (E) to contrast with the fit obtained in (A).
help explain why the hypermutable bases in p53 are so mutable and why they are targets for various kinds of environmental stress such as nitric oxide and oxyradicals (18, 19). Mutagens are known to stimulate p53 transcription (20, 21) as well as supercoiling (11, 12), and the MI of a base may dictate the extent to which a toxin can affect mutation frequency. Because stressors cannot affect MIs directly, their effects on mutation frequency may also be indirect by enhancing supercoiling and SLSs directly or indirectly via transcriptional activation (10–15). By the same token, suppressing such mutations may require inhibition of the supercoiling and/or transcription involved.

Whereas MIs determine relative mutation frequencies, the spectrum of mutations produced is apparently determined by the chemical mechanisms involved. This conclusion is substantiated by the fact that the spectrum of p53 mutations that occurs is similar to that of background E. coli lacI mutations (22): C mutates to T much more frequently than to A or G, and G mutates to A much more frequently than to C or T. The thermodynamic properties of hydrolytic reactions that occur in nucleic acids under physiological conditions are such that the deamination of C is much greater than A, and depurination of G or A is much greater than depyrimidation of C or T (reviewed in Ref. 23). Also, the oxidation potential of a G located 5' to another G (codons 245, 248, and 249) is greater than it is next to a C or T (reviewed in Ref. 24). Because of its size, an A is more likely than a C or T to replace a G. The total G and C content of the SLS containing the bases in Table 1 is 32 and 30%, respectively (data not shown). In Table 1, 91% of the bases are Gs or Cs compared with 50% of the control bases (Table 2). In contrast to the hypermutable bases, no Gs and only one C in the control bases are immediately next to a stem (codon 250, Table 2). The data suggest that while mutagens increase mutation frequency, their specificity for targeting Gs or Cs cancel out in a database of 14,000 mutations such that the spectrum is similar to that of background mutations. The spectrum of mutants produced in p53 or lacI is independent of MI. Whether or not a mutagen is present, the underlying cause of relative mutation frequencies in p53 appears to be the inevitable consequence of folding the sequences that constitute this gene. Most, if not all, mutagens are more reactive on ssDNA, and all of the mechanisms by which they are presumed to cause mutation occur at lesser (but significant) levels in an untreated genome because of the inherent nature of the bond involved. Thus, it is more appropriate to use the term background rather than “spontaneous” mutation (23).

In conclusion, it appears as though GC richness, thermodynamic potential of depurination, deamination, and oxidation, as well as proximity to a stem conspired during evolution to create the hypermutable bases of the p53 gene. In unicellular organisms, adverse environmental conditions can bestow mutability on those genes regulating the cell’s response to each type of stress, thus minimizing genome-wide genetic damage while creating the most appropriate variants for selection of the fittest in response to stress (15). In multicellular organisms, however, variants are not independent cells but part of a whole, and selection of the fittest results in abnormal growth. This would be an ironic legacy for mechanisms that were selected to accelerate evolution.

Acknowledgments

We thank Drs. Mike Minnick, Steve Lodmell, Frank Rosenzweig, Brooke Martin, and Kent Sugden for valuable suggestions and discussions.
References

Hypermutable Bases in the p53 Cancer Gene Are at Vulnerable Positions in DNA Secondary Structures

Barbara E. Wright, Jacqueline M. Reimers, Karen H. Schmidt, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/20/5641

Cited articles
This article cites 21 articles, 9 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/20/5641.full#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/20/5641.full#related-urls

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.