10-2011

The roles of transcription and genotoxins underlying p53 mutagenesis in vivo

Barbara E. Wright
University of Montana, Missoula

Karen H. Schmidt
University of Montana, Missoula

Aaron T. Hunt
University of Montana, Missoula

J Stephen Lodmell
University of Montana, Missoula

Michael F. Minnick
University of Montana, Missoula

See next page for additional authors

Carcinogenesis vol.32 no.10 pp.1559–1567, 2011
doi:10.1093/carcin/bgr177
Advance Access publication July 29, 2011

The roles of transcription and genotoxins underlying *p53* mutagenesis *in vivo*

Barbara E.Wright*, Karen H.Schmidt, Aaron T.Hunt, J.Stephen Lodmell, Michael F.Minnick and Dennis K.Reschke

Division of Biological Sciences, The University of Montana, 32 Campus Drive, Missoula, MT 59812, USA

*To whom correspondence should be addressed. Tel: +1 011 1 406 243 6676; Fax: +1 011 1 406 243 4184;
Email: barbara.wright@mso.umt.edu

Transcription drives supercoiling which forms and stabilizes single-stranded (ss) DNA secondary structures with loops exposing G and C bases that are intrinsically mutable and vulnerable to non-enzymatic hydrolytic reactions. Since many studies in prokaryotes have shown direct correlations between the frequencies of transcription and mutation, we conducted *in silico* analyses using the computer program, *mfg*, which simulates transcription and predicts the location of known mutable bases in loops of high-stability secondary structures. *Mfg* analyses of the *p53* tumor suppressor gene predicted the location of mutable bases and mutation frequencies correlated with the extent to which these mutable bases were exposed in secondary structures. *In vitro* analyses have now confirmed that the 12 most mutable bases in *p53* are in fact located in predicted ssDNA loops of these structures. Data show that genotoxins have two independent effects on mutagenesis and the incidence of cancer: Firstly, they activate *p53* transcription, which increases the number of exposed mutable bases and also increases mutation frequency. Secondly, genotoxins increase the frequency of G-to-T transversions resulting in a decrease in G-to-A and C mutations. This precise compensatory shift in the 'fate' of G mutations has no impact on mutation frequency. Moreover, it is consistent with our proposed mechanism of mutagenesis in which the frequency of G exposure in ssDNA via transcription is rate limiting for mutation frequency *in vivo*.

Introduction

More than half of all cancers are associated with mutations in the *p53* gene. Since the p53 protein coordinates complex mechanisms of DNA repair and regulates cell growth, these relationships are disrupted when the *p53* gene is inactivated, allowing for the selection of cells with higher rates of division that frequently lead to cancer. Transcription of *p53* is very low in normal tissue (1), and genotoxic agents induce *p53* expression (2–5). Genotoxic activity also forms oxyradicals resulting in base damage to G, which has a high oxidation potential (6,7). Earlier work with *Escherichia coli* (reviewed in refs 8,9) has shown that stress-related mutations are frequently located in single-stranded loops of predicted secondary structures, and this is also observed in hypermutable bases of *p53* (10). Analyses in the present study clarify the roles of genotoxins, transcription and intrinsically mutable bases in *p53* mutagenesis *in vivo*.

Secondary structures in *p53* form primarily in the non-transcribed strand from segments of ~35–45 nts (10). Point mutations in G and C bases of *p53* are concentrated in several codons, which correspond to specific critical interaction domains in the p53 protein. Transcription drives supercoiling (11–14) which stabilizes the formation of secondary structures, and mutations occur in ssDNA or 'hot spot' loops of secondary structures (15,16). Since the early 1980s (17), evidence has indicated that these structures represent the precursors for endogenous 'background' mutations (reviewed in refs 8,9). Hydrolytic, endogenous mutations (especially C-to-T and G-to-A) occur at significant frequencies under physiological conditions (7,17–25). The ssDNA in which such mutations occur probably arises during maintenance levels of replication (26), transcription and/or supercoiling, allowing double-stranded DNA strands to separate sufficiently for stem–loop structure (SLS) formation within ssDNA regions. In double-stranded DNA, for example, N3 of cytosine is H-bonded to the N1 of guanine, and ssDNA (via transcription and/or negative supercoiling) is required for protonation at N3 and deamination to uracil. Thus, it is widely recognized that endogenous mutations of C and G are the result of their intrinsic thermodynamic instability, suggesting that ssDNA availability (via transcription) may be rate limiting for mutation frequency under the steady-state conditions existing *in vivo*.

Direct correlations between transcription and mutation frequency have been observed in both eukaryotes (27–29) and prokaryotes (30,31). These considerations and earlier work demonstrating the predictive value of kinetic models (32,33) prompted the construction of *mfg*, a computer algorithm (34) that interfaces with the *mfold* program (35). *Mfg* simulates transcription in order to locate known mutable bases in secondary structures (34). Earlier work had shown correlations between starvation-induced transcription and mutation frequency (31), and the predictive value of the *mfg* program has since been demonstrated by showing correlations between: (i) transcription and reversion frequency in *E.coli* (30), (ii) reversion frequency and promoter strength (36,37), (iii) SLS stability (ΔG) and mutation rate (34), (iv) protein sequence diversity and increased levels of transcription (38) and (v) base exposure and mutation frequency in *p53* (10) and, in the present study, *in vitro* analyses have confirmed the *mfg*-predicted locations of *p53* hypermutable bases in loops of secondary structures.

Abbreviation: SLS; structure.

Materials and methods

Mutation analysis

The data in Table II, Supplementary Tables 1 and 2 and Figures 1 and 2 (available at *Carcinogenesis* Online) pertaining to the *p53* gene were obtained from the IARC TP53 database version R14 (39). This database is maintained by the International Agency for Research on Cancer. The somatic mutation dataset which contains 26 597 mutations from 2179 references was downloaded into Excel, sorted and analyzed. The data were used to examine the fate of G and C mutations in different categories of cancers and were derived by including or excluding data from original column headings as follows: the twelve most hypermutable bases from codons 175, 245, 248, 249, 273 and 282 were selected. Only 'missense', 'nonsense' or 'silent' mutations were included. Tumor types included were 'primary' and 'NA' (not assigned) and excluded were 'metastasis', 'recurrent' and 'secondary'. Cancers were sorted by organ affected, and bases were sorted by fates of mutations (i.e. G-to-T, C-to-A).

Data in Table I pertaining to *VH* genes were obtained from the dataset in (40), Supplementary Table 1. This dataset contains 28 307 mutations in different *VH* genes and referenced sequences. We examined the mutations in three genes: *VH*5 (GenBank accession numbers X92278 and M99684), *VH*94 (GenBank accession number L10094) and *VH*186.2 (GenBank accession number L10088). The data were derived by sorting columns as follows: data under column headings: 'Region', CDR1 and CDR2 or FW1, FW2 and FW3 were sorted independently. For the column 'NCBI Accession #_Vgene', only X92278_IGHV-34*01, M99864_IGHV4-59*01, L10088_IGHV4-59*01 and L10094_IGVH-39*01 were analyzed. In column 'R-or-S', R refers to missense mutations, S refers to silent mutations, and the mutations were sorted accordingly. The 'Exchange' column allows for sorting by fate of mutations (i.e. G-to-T, C-to-A).

The mfg program

This program is described (Figure 1) and directions for downloading and using *mfg* are freely available online: http://www.dbs.umt.edu/research_labs/wright-lab/upload/mfg.html. *Mfg* is an open source desktop application that runs on Windows, Linux and Mac operating systems.

Construction of p53-containing plasmids for generation of double-stranded DNA for S1 endonuclease cleavage assays

A 1.6 kb fragment from a clone containing the entire *p53* gene (Spectra Genetics, Pittsburgh, PA) was amplified by polymerase chain reaction using specific primers P5 (5′-ACCCGCGTCCGCGCCA-3′) and 3′p53ex9

Table I. Frequency and fates of G and C intrinsic instability mutations in ssDNA

	G-to-A	G-to-T	G-to-C	C-to-T	C-to-A	C-to-G
A: Thermodynamic characteristics of nucleosides leading to mutation in ssDNA						
Type of base substitution	Transitions[a]	Depurination, (G → T), (A → T)	Depurination (G → C), (A → C)	Deamination[b] (C → U → T)	Depyrimidation (C → A), (T → A)	Depyrimidation (C → G), (T → G)
Mutations/day/10^{10} bases/cell (% of total)	(∼95%)	(n.a.)	(n.a.)	(95.6%)	(∼2.2%)	(∼2.2%)
B: Missense mutations in eight codons of *p53*						
Number of mutations (% of total)	3017 (75.9%)	798 (20.1%)	158 (4.0%)	1728 (93.0%)	28 (1.5%)	103 (5.5%)
C: Silent mutations in all exons of *p53*[c]						
Number of mutations (% of total)	239 (78.4%)	35 (11.5%)	31 (10.1%)	453 (79.3%)	78 (13.7%)	40 (7.0%)
D: Silent mutations in framework regions of *VH* genes[d]						
Number of mutations (% of total)	1156 (67.5%)	274 (16.0%)	283 (16.5%)	1066 (80.4%)	103 (7.8%)	156 (11.8%)
E: Missense mutations in germ line light chain genes[e]						
Number of mutations (% of total)	48 (72.7%)	5 (7.6%)	13 (19.7%)	27 (81.8%)	5 (15.2%)	1 (3.0%)

[a]About 5000 purine bases turn over each day due to hydrolytic depurination, with G being replaced by A because of its size (7, 22).
[b]In *p53*, hypermutable Cs are methylated and mutate directly to T.
[c]Silent mutations are presumed to be underreported in the IARC TP53 database because they do not result in cancer.
[d]Data were obtained from the *VH*5, *VH*94 and *VH*186.2 gene segments (see Materials and methods).
[e]Missense mutations in germ line Gs and Cs in V_kO_x1 light chain genes (53). Msh2−/− & Msh2+/+ values are similar.

Table II. Comparing the frequency and fates of *p53* mutations leading to cancer in the most hypermutable G bases in 13 categories of cancer[a]

	Number of mutations and % of total		
Category of cancer	G-to-A, intrinsic	G-to-T	G-to-C
Total Mutations (all)	2181 (74.0%)	664 (22.5%)	102 (3.5%)
Liver[b]	**40 (12.6%)**	**272 (85.8%)**	**5 (1.6%)**
Lung[c]	**121 (33.9%)**	**206 (57.7%)**	**30 (8.4%)**
Head and neck	45 (60.0%)	27 (36.0%)	3 (4.0%)
Bladder	112 (79.4%)	18 (12.8%)	11 (7.8%)
Pancreas	43 (81.1%)	7 (13.2%)	3 (5.7%)
Prostate	20 (83.3%)	2 (8.3%)	2 (8.3%)
Breast	353 (86.1%)	40 (9.8%)	17 (4.1%)
Esophagus	231 (85.9%)	31 (11.5%)	7 (2.6%)
Colon	245 (91.8%)	16 (6.0%)	6 (2.2%)
Stomach	159 (90.3%)	13 (7.4%)	4 (2.3%)
Rectum	173 (92.5%)	8 (4.3%)	6 (3.2%)
Brain	221 (94.0%)	10 (4.3%)	4 (1.7%)
Colorectum	418 (95.9%)	14 (3.2%)	4 (0.9%)

Bold type depicts tissues most prone to genotoxic stressors.
[a]Data from the eight most hypermutable bases in codons 175, 245, 248, 249, 273 and 282 are included (Source: IARC TP53 database, version R14; see Materials and methods).
[b]Full description is liver and intrahepatic bile ducts.
[c]Full description is bronchus and lung.

(5′-CTGAAGGGTGAAATATTCTCC-3′) (Integrated DNA Technologies, Coralville, IA). This segment includes the latter half of exon 5 through the end of exon 9 and was cloned into pCR2.1-TOPO (Invitrogen, Carlsbad, CA) in both orientations and named 1.6 + TOPO and 1.6 − TOPO, respectively. Exon 5 alone was subcloned similarly using primers Ex5p53us (5′-GCTGTGGGTGTATTC-3′) and 3′p53ex5i (5′-CTAAGAGCAATCAGTG-3′) and named ex5TOPO. Exons 5, 6 and 7 were subcloned together using primers P5 and 3′p53ex7i (5′-TGTGCAGGGTGGCAAG-3′) and named Ex5,6,7TOPO. An additional 1.6 kb fragment with the complete exon 5 segment at the 3′ end was cloned using primers 5′p53ex2 (5′-AGCCGCAGTCAGATC-3′) and 3′p53ex5i and named 1.6 ± ex5TOPO. Plasmid midipreps (Promega, Madison, WI) were used to purify double-stranded DNA for cleavage assays.

S1 endonuclease analysis for loop structures

A digest of 10–30 μg of double-stranded plasmid DNA was performed with 10U or 20U of S1 endonuclease (Promega) for 1 h at 25°C. DNA was then purified with a Wizard PCR clean-up system (Promega). Eluate was digested with XhoI (NEB, Ipswich, MA) for 2 h at 37°C and heat inactivated for 20 min at 65°C. Non-radioactive deoxycytidine triphosphate/deoxythymidine triphosphate/deoxyguanosine triphosphate was added together with 10 μCi of dATP α-^{32}P (6000 Ci/mmol; MP Biomedicals, Solon, OH) and 5U Klenow exo- (NEB) to fill in and label the 5′ overhang. This procedure labels only the transcribed or non-transcribed strand of the fragment, depending on the construct. After incubation for 1 h at 37°C, the enzyme was heat inactivated for 20 min at 75°C. A 1.6 kb fragment was generated by SpeI digestion, separated by agarose gel electrophoresis and purified by gel extraction columns (Omega Bio-tek, Norcross, GA). Products were separated on a urea containing 7% polyacrylamide gel with an appropriate sequencing ladder. The primer XhoIMSCrev2 (5′-CTCGAGCGGCCGCCAGTG-3′) was used for sequencing ladders. The control omitting S1 digestion was treated in parallel. Mung bean endonuclease (NEB) was used for 1 h at 30°C according to manufacturer's instructions with similar results (data not shown). Sequence ladders (Sequenase 2.0; USB/Affymetrix, SantaClara, CA) were generated using the same template and primers, labeled with ^{32}P- αATP. Sequencing ladders are complementary to the fragments. Each reaction result was confirmed by repeating the procedure at least three times using at least two pools of DNA.

Results

Using mfg to predict the location of hypermutable bases in loops of secondary structures

Figure 1A shows an *mfg* computer output for a given sequence. The sequence analyzed was folded in 40 nt segments or 'windows', and the C at nt 233 was unpaired in 77% (% Unp) of its total folds during simulated transcription. *Mfg* interfaces with the *mfold* program which is used by *mfg* for folding successive single-stranded segments of a specified length for a given sequence. A window size (number of nucleotides) is chosen, and for each successive base in the sequence analyzed, *mfg* reports the stability (ΔG) of the most stable secondary structure (among many) in which that base is unpaired and also reports the percent of total folds in which it is unpaired. The mutability index (MI) of each unpaired base is the product of these two variables. *Mfg* simulates transcription *in vivo* by showing progressive 'snap shot' views of the most stable 40 nt SLSs in which each base is unpaired, similar to RNA polymerase progression along the template strand during transcription. The window size chosen relates to the rate of transcription and the amount of ssDNA folded. These simulations involve a running competition for shared nucleotides between successive interconverting secondary structures of different stabilities. Patterns emerge of secondary structures that predominate by forming repeatedly. In general, the highest stability SLSs form most frequently compared with those of lower stability.

In Figure 1B, transcription and associated supercoiling are modeled using a 40 nt window generating an SLS exposing a hypermutable C in exon 8 of *p53*. This structure (named SLS 5.0 for its ΔG) is

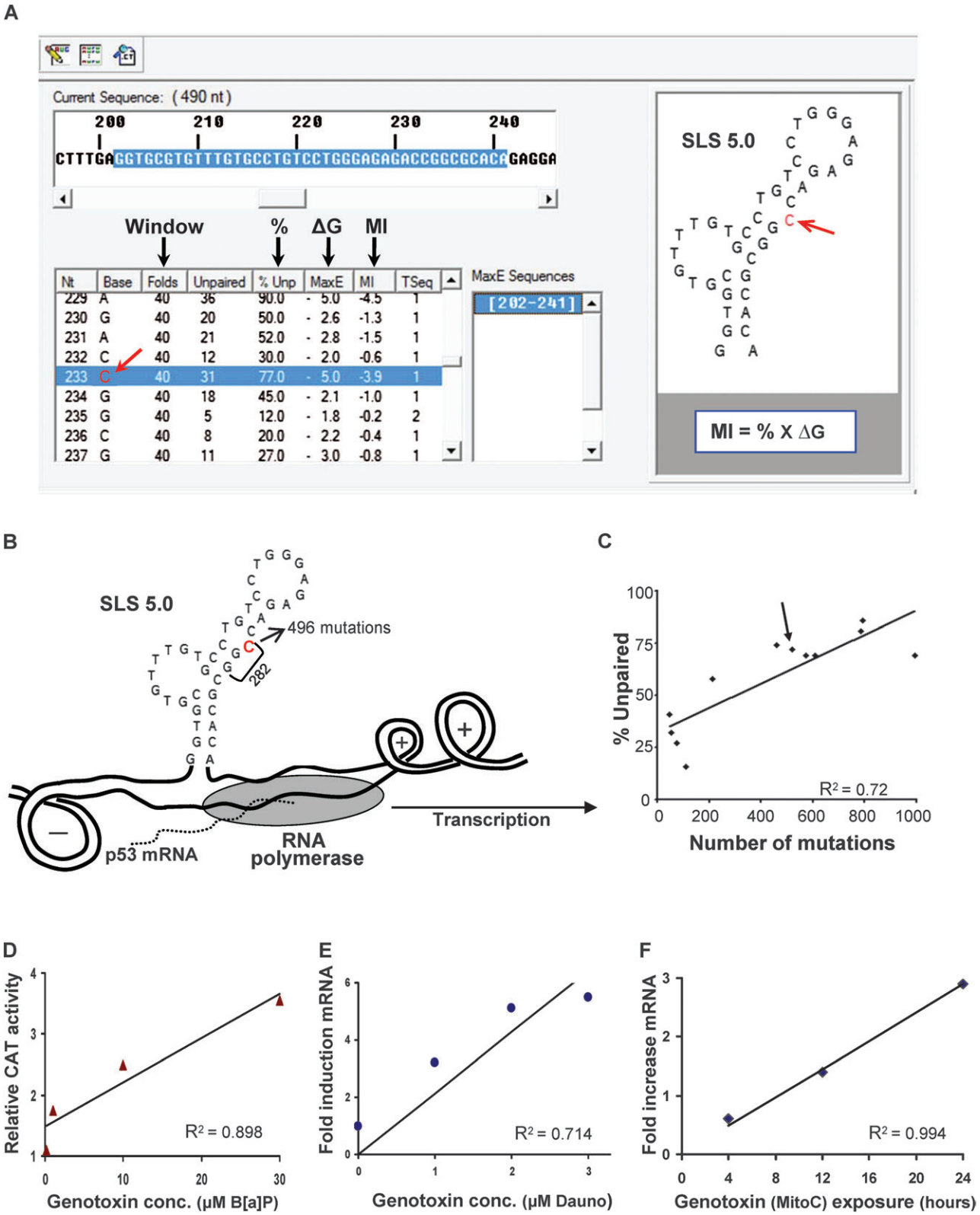

Fig. 1. (**A**) An example of *mfg* computer output. The C nucleotide (red) at *mfg* nt 233 was selected and the most stable SLS (ΔG, −5.0) in which that base is unpaired is shown. The output also provides the window size, percent unpaired, ΔG and mutability index. (**B**) Transcription of the *p53* gene, resulting in negative supercoiling in the wake of the transcription complex. The SLS formed (with a ΔG of −5.0) contains hypermutable codon 282. (**C**) Correlation between mutation frequencies of the twelve most hypermutable bases in *p53* (39) and the *mfg*-predicted percent unpaired for each base, using a window size of 43 nts. The hypermutable base of codon 282 is arrowed. Excel was used to calculate R^2 values, which are statistically significant when $P \leq 0.05$. (**D**) Activation of the p53 promoter as measured by Chloramphenicol AcetylTransferase reporter activity as benzo[*a*] pyrene (B[*a*]P) concentration increases (4). (**E**) Induction of the p53 promoter as measured by fold increase in luciferase reporter activity with increasing concentrations of the genotoxin daunomycin (2). (**F**) Increase in p53 messenger RNA over time as a function of increased exposure time to genotoxin mitomycin C (5). For (C–F) *n* values are 13, 4, 4 and 3, respectively.

presumably formed in negative supercoils in the wake of the transcription complex and is the most stable structure in which this hypermutable C in *p53* codon 282 is exposed. The mutable base is arrowed in Figure 1C which shows the correlation between mutable base exposure during transcription and the number of mutations for the 12 most mutable bases in the current IARC TP53 somatic human cancer database (39). Separate analyses determined the optimal window size for each hypermutable base (supplementary Figure 1A is available at *Carcinogenesis* Online), which averaged about 43 nt (10). An *mfg* analysis of exons 5–9 also shows that exons 5, 7 and 8 (containing the five most hypermutable *p53* codons) form higher stability structures than those in exons 6 and 9 which have no hypermutable codons (supplementary Figure 1B is available at *Carcinogenesis* Online).

Transcription induction of the *p53* gene in response to genotoxins has previously been reported (2–5). Figure 1D and E demonstrate induction of *p53* with increasing levels of genotoxin, and Figure 1F shows messenger RNA levels increasing over time as cells are exposed to genotoxin, indicating the activation of transcription. There is a strong correlation between mutation and transcription frequency in both prokaryotes (Figure 2A and B) and eukaryotes, for example, in the pre-B cell line 18–81 (Figure 2C) and in the *VHB1-8* antibody gene (Figure 2D) using different strength promoters in mice (29). Thus, in living systems, a correlation exists between transcription and mutation, which underlies the fundamental mechanism of mutagenesis essential to evolution. These correlations therefore represent a newly recognized key contribution to the 'Unity in Biochemistry', which demonstrates that metabolic mechanisms and molecules have been preserved essentially intact through billions of years of Darwinian evolution (41,42).

The G and C nucleotides in ssDNA are thermodynamically unstable and intrinsically mutable

The overall thermodynamic characteristics of nucleotides in ssDNA underlying the majority of background or 'endogenous' mutations are shown in Table IA. *In vivo*, the ssDNA in which unpaired bases are exposed and mutable is presumably due to maintenance level metabolic activities such as transcription and replication. The hydrolytic deamination of 5-methyl C and C occur ~300 times/day/human cell, which is more than sufficient to explain the observed endogenous mutation frequency (43,44). About 5000 purine bases turn over each day due to hydrolytic depurination, with G being replaced by A at those apurinic sites because of its similar size (20). In the *p53* gene, the relative mutation frequencies of Cs and Gs in missense mutations (Table IB) and of germ line light chain antibody genes (Table IE) are presumably also due to base exposure during transcription and intrinsic chemical instability in ssDNA. In silent mutations of *p53* (Table IC) and framework regions of antibody *VH* genes (Table ID), G-to-A and C-to-T transitions also predominate. Perhaps, the most compelling evidence for a common mechanism of mutability in unpaired bases is seen in silent mutations, where the outcome of a mutation does not alter the encoded amino acid and is not selected, reflecting only intrinsic base instabilities (Table IC and D).

Analyses of G and C mutations in the current p53 mutation database

The frequency and three possible fates of 2947 missense mutations of G-to-A, T and C in 13 different types or categories of cancer in the eight most mutable codons are summarized in Table II. Supplementary Table 3, available at *Carcinogenesis* Online, provides comparable data for mutations of C. Note that, on average for all categories of cancer, the fates of G and C mutations are similar to those for background or 'intrinsic' mutations (Table I), with the exception of somewhat higher numbers of G-to-T transversions associated with genotoxin-prone tissues such as liver and lung. Genotoxins activate *p53* transcription (2–5), as shown in Figure 1D–F, and transcription frequency is directly correlated with mutation frequency in both eukaryotes and prokaryotes (Figure 2). There is also a direct correlation between transcription and mutation in antibody genes during

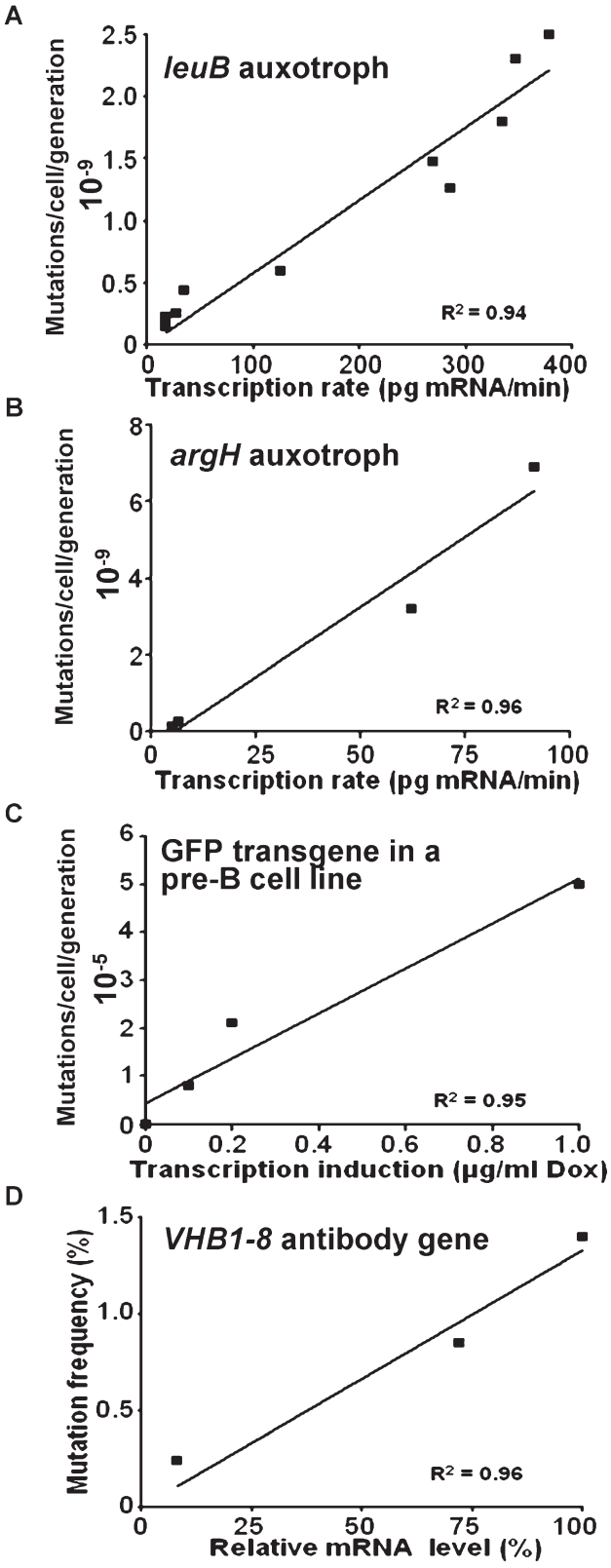

Fig. 2. (**A** and **B**) Correlation between rates of transcription and mutation in *Escherichia coli* auxotrophs *leuB* (30) and *argH* (31). (**C** and **D**) Correlation between mutation rates and transcription in a Green Fluorescent Protein transgene in the human pre-B cell line 18–81, due to the activation of transcription by doxycycline (dox) (27) and in the *VHB1-8* antibody gene (29). Linear regression values were calculated in Excel. R^2 values are statistically significant when $P \leq 0.05$. For (A–D) *n* values are 10, 4, 4 and 3, respectively.

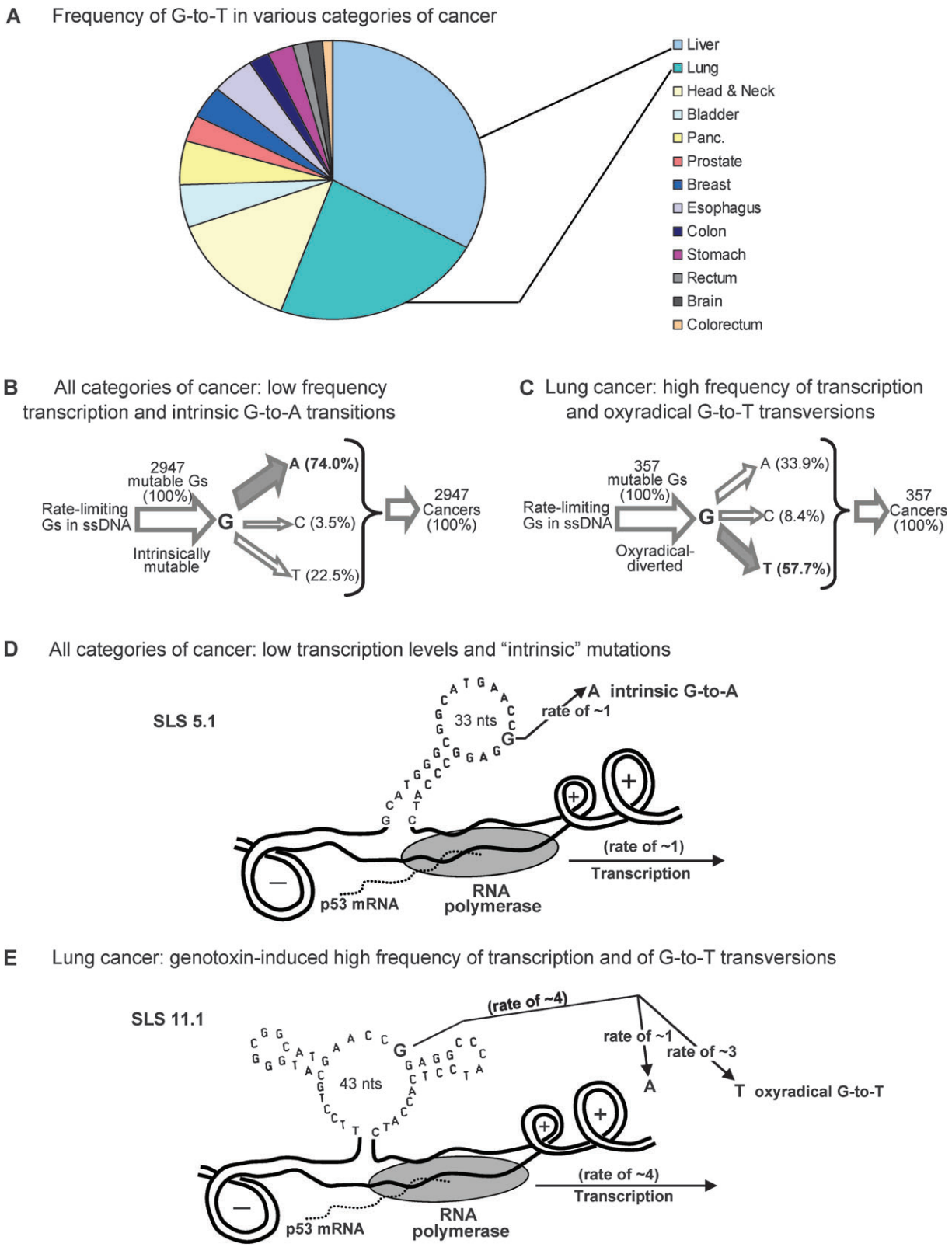

Fig. 3. Examination of the fate of mutable Gs in *p53* in all cancers compared with lung cancer. (**A**) The relative frequencies of *p53* G-to-T mutations by tissue type. (**B**) The fate of G mutations in all cancers in which a majority of mutations (74.0%) are G-to-A (intrinsic). (**C**) The fate of G mutations in lung cancers in which a majority (57.7%) are oxyradical-induced G-to-T transversions. Mutation data are derived from missense mutations in the eight most mutable codons in the IARC TP53 database. Data for B and C are taken from Table II. (**D**) In all (pooled) cancers, the relative rate of G-to-A intrinsic mutations occurring in a specific G base at low levels of transcription is depicted by a 33 nt SLS and (**E**) in lung cancers at high levels of transcription (depicted by a 45 nt structure containing the same hypermutable G) with the approximate relative rates of G-to-A intrinsic mutations (rate of ∼1) and genotoxins-oxidized G-to-T transversions (rate of ∼3).

somatic hypermutation (Figure 2C and D). However, in that system, the predominant transcription-driven mutation is C-to-U due to Activation induced cytidine deaminase (AID) activation (see Discussion). As indicated in Table II (detailed in Supplementary Table 2 available at *Carcinogenesis* Online), the number of G-to-T transversions in liver and lung cancers is exceptionally high compared with the all cancer category, consistent with the direct exposure of these tissues to genotoxins, which increase the rate of transcription. Toxins would be expected to accumulate preferentially in the liver, as a detoxifying organ, and in lungs, which are exposed to polluted aerosols. The distribution of G-to-T transversions in various categories of cancer is shown in Figure 3A.

The dual role of genotoxins in mutagenesis is complex since genotoxins both activate transcription (exposing mutable Gs in ssDNA) and cause base damage (G-to-T transversions) primarily in lung and liver tissues (45). Figure 3B and C documents the effects of G availability and genotoxins on mutation frequency in all categories of cancer versus lung cancer. In all categories of cancer (Figure 3B), low rates of transcription and ssDNA availability are the primary sources of intrinsically mutable G-to-A mutations, similar to 'background' mutations (Table I), whereas at higher (genotoxin-induced) levels of transcription (Figure 3C), many Gs are oxidized to Ts, thus diverting the usual fate of G mutations away from A and to T. Since G availability via transcription is rate limiting, these compensatory shifts in the fate of G mutations have no effect on the incidence of cancer. This is also the case in Figure 3C, i.e. the incidence of cancer is the same as the number of available mutable Gs. These data indicate that the rate of availability of unpaired mutable Gs in ssDNA during *p53* transcription is rate limiting for and determines the frequency of G mutations leading to the inactivation of *p53*. Dynamic models help to clarify these relationships. For example, in Figure 3D in all categories of cancer, at low levels of transcription, the mutable G is depicted to be in a single-stranded region of a 33 nt SLS of low stability (SLS 5.1). In lung cancers (Figure 3E), at higher levels of genotoxin-induced transcription (2–5), the same G is shown to be exposed in a more stable SLS at a window size of 43 nt. As discussed above, genotoxins activate *p53* transcription, which increases base exposure (Figure 1C) and mutation frequency (Figure 2). Although the rate of transcription and mutable G availability determine mutation frequency, other variables such as genotoxins can affect the distribution or relative 'fates' of those G mutations to A, T and C as modeled in Figure 3E. However, the total number of cancers is not affected by these compensatory changes in mutation fate and is only determined by the rate of availability of intrinsically mutable Gs (Figure 3B and C). The most important variables that determine the frequency at which a mutable G is available *in vivo* are: (i) the frequency of transcription and unpaired base availability in ssDNA, (ii) the inherent instability and mutability of the unpaired G or C base and (iii) the extent of base exposure during transcription (Figure 1C). The type or fate of a G mutation is determined by oxyradical-induced mutations in *p53* and in somatic hypermutation; a parallel may be seen in enzyme-induced AID mutations (see Discussion).

In vitro confirmation of the predicted location of loops in SLSs-containing p53 codons 245 and 248

To determine whether loops in secondary structures predicted by *mfg* to form could be verified experimentally, we examined supercoiled plasmid DNA, which contained relevant *p53* exons that had been digested with S1 endonuclease. We hypothesized that, while static, this method would provide a model to detect ssDNA exposure in secondary structure. Figure 4A shows the *mfg* analysis of nts 120–165 in the non-transcribed strand of exon 7, which contains hypermutable codons 245 and 248. The first three columns show the number of mutations reported in the *p53* mutation database for those bases, the codons of interest and the *mfg* nucleotide number and the fourth column shows the stabilities (ΔG) of those structures. The four hypermutable bases are circled in column 6 and in SLS 9.3, SLS 12.8 and SLS 11.1, which are color coded with their location in Figure 4B and C. The location of codons 245 and 248 are bracketed black in Figure 4B and C. (see figure legends for further details). Nine S1 endonuclease cleavage sites are arrowed red in Figure 4A–D. Note the color coordination between each secondary structure and its predicted repetition (e.g. SLS 12.8 is seen 21 times in column 4, Figure 4A). The four hypermutable bases in codons 245 and 248 are unpaired and mutable (circled red) except for one G in codon 245 (SLS 12.8). Cleavage sites are immediately next to the hypermutable Gs in codon 245 in two structures (SLS 9.3 and SLS 11.1) and two cleavage sites occur in the same loop of SLS 12.8 and SLS 11.1. The two most mutable of the four hypermutable bases in codons 245 and 248 are exposed in all three structures, and the location of all cleavage sites observed experimentally (Figure 4D) correlates with their positions in loops of the relevant *mfg*-predicted secondary structures (Figure 4A–C). For analysis of codons 245 and 248, the structures depicted in Figure 4C represent 65% of those listed in the *mfg* analysis. To examine codon 175 and to verify the assumption that secondary structures formed in the two strands of a codon are similar, both strands of codon 175 were analyzed (supplementary Figure 3A and B is available at *Carcinogenesis* Online). Codons 273 and 282 of exon 8 were also analyzed in the non-transcribed strand (supplementary Figure 3C and Table 1C are available at *Carcinogenesis* Online). Exon 8 has lower stability SLSs than exons 5 and 7; thus, the most stable SLSs observed were SLS 5.0 and 4.9 (supplementary Figure 3C is available at *Carcinogenesis* Online). Gels of S1 cleavage products (data not shown) revealed many faint bands, consistent with the formation of several relatively low stability structures formed from ssDNA.

In summary, separating the variables that determine the frequency of a mutation from those that determine its fate clarifies the mechanism of mutagenesis. The availability of mutable Gs in ssDNA via transcription is apparently rate limiting for, and determines the frequency of mutation, whereas the increase in oxyradical-induced base damage (G-to-T transversions), *per se*, has no effect on the incidence of cancer (Figure 3B and C). Given the proposed model of mutagenesis in *p53*, oxyradical-associated increases in the rate of transcription play a critical role in mutation frequency, and any method of downregulating the rate of *p53* transcription to baseline levels (1) under stressful conditions may lower the incidence of liver and lung cancers.

Over the past 30 years, many reports on mutable bases in both prokaryotes and eukaryotes have concluded that such bases are unpaired in ssDNA loops of secondary structures, that exposed bases are chemically unstable and intrinsically mutable and that there is a direct correlation between mutation and transcription frequency. It is also known that guanine has a high oxidation potential and that exposure to oxyradicals frequently results in G-to-T transversions. The proposed mechanism of mutagenesis is also supported by additional lines of evidence. *In vitro* analyses and structure profiling of *p53* have confirmed, for the first time, that hypermutable bases in *p53* are indeed located in predicted loops of high-stability secondary structures. Moreover, all 22 S1 endonuclease cleavage sites are located in loops of predicted structures, consistent with the observation that the extent of base exposure in *p53* correlates directly with mutation frequency during simulated transcription (Figure 1C). Chemical analyses and silent mutations (that are not selected) have also shown that C and G nucleotides are chemically unstable and intrinsically mutable. These observations indicate that the frequency of transcription, or more specifically, the frequency of unpaired Gs and Cs in ssDNA during transcription, is rate limiting for mutation frequency *in vivo.*

Discussion

This work addresses a longstanding and serious problem, namely, the relevance of *in vitro* data to the dynamics of metabolism under steady-state conditions *in vivo*. Differences between metabolism *in vitro* versus *in vivo* have been the concern of a number of investigators in the past (32,33,46–51), who point out that substrates are typically used at saturating levels *in vitro* and enzymes at low rate-limiting concentrations in order to measure the rate of a reaction. *In vivo*, however (except for a few 'pacemakers' such as glycogen phosphorylase), enzyme concentrations under steady-state conditions usually far exceed substrate concentrations and substrate availability is

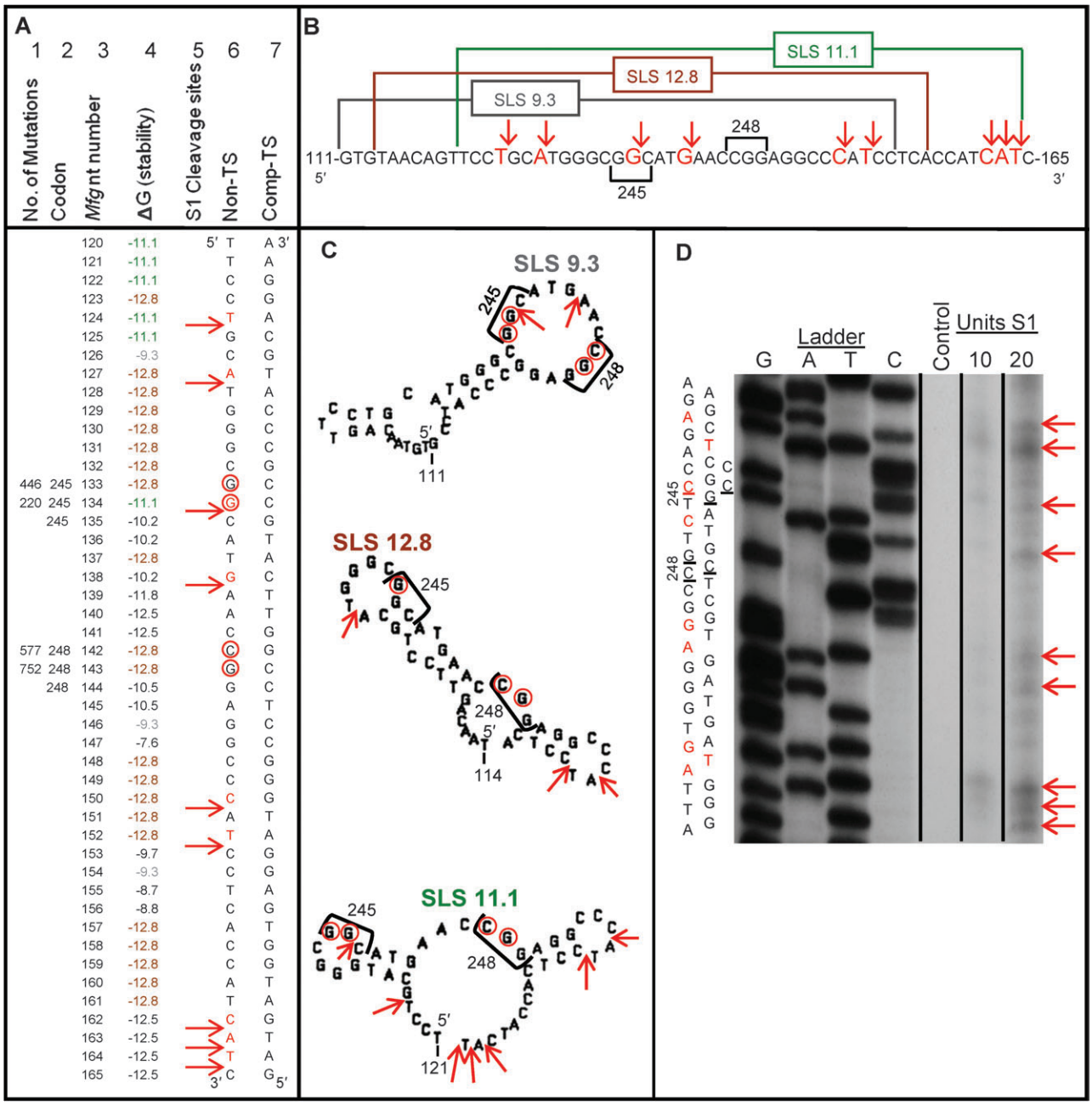

Fig. 4. S1 cleavage analysis of exon 7 of *p53*. Plasmid DNA containing the non-transcribed strand of exon 7 and hypermutable codons 245 and 248 was analyzed by S1 endonuclease. (**A**) Selected *mfg* computer output (15°C, 44 nt window). Column 1 shows mutation frequencies of the hypermutable bases; column 2 identifies the hypermutable codon; column 3, the *mfg* assigned nt number; column 4, the color-coded stability (ΔG) of the predicted highest stability SLSs containing the mutable unpaired base; column 5 contains red arrows at S1 cleavage sites; column 6 shows the sequence of the non-transcribed strand; column 7 shows the complementary transcribed strand. (**B**) A segment of the *p53* sequence identifying the relative locations of predicted color-coded SLSs in relation to the locations of cleavage sites (red font and arrows) and the hypermutable codons (black bracket). (**C**) Three *mfg*-predicted secondary structures containing codons 245 and 248 showing S1 cleavage sites in loops (red arrows). Hypermutable bases are circled (red). (**D**) Polyacrylamide gel showing S1 cleavage sites (red arrowed) next to the transcribed strand sequence ladder.

frequently more rate limiting than enzyme concentration in the intact cell. The work of Kascer (46) is an elegant example of experimental approaches that can reveal, under steady-state conditions *in vivo*, which enzymes are, and which are not rate controlling, depending upon their 'kinetic position' in the metabolic sequence. Thus, there is precedence for our conclusion that intrinsically mutable Gs and Cs are the rate-limiting substrates for mutagenesis in *p53*. Furthermore, available Gs and Cs in single-stranded loops of *VH* genes may also mutate because of intrinsic mutability and because ssDNA sites are available for mutation of C-to-U by AID-based diversification (52).

Dynamic *in vivo* models of mutagenesis (Figure 3D and E) show the origin and fate of transcription driven unpaired intrinsically mutable bases exposed in loops of secondary structures. The availability of these bases via transcription is apparently the rate-limiting variable that determines mutation frequency, which is also the case for mutagenesis in somatic hypermutation. Thus, the model for background mutations in *p53* (Figure 3D) shows G-to-A as the intrinsically mutable rate-limiting base, many of which are usurped by genotoxins which convert them to T. In somatic hypermutation, antigen challenge increases transcription, ssDNA and mutable C availability. In this

case, AID activation takes most of these Cs and mutates them to Us, initiating high frequency enzyme diversification.

In both prokaryotes and eukaryotes, correlations are seen between the frequency of ssDNA availability (via transcription) and mutation frequency (Figure 2), consistent with the conclusion that substrate availability is rate limiting *in vivo* for mutagenesis. Mutation frequency in *p53* is determined by the level of transcription, which dictates the rate of availability of exposed intrinsically mutable Gs and Cs in loops of secondary structures, and the intrinsic mutability and exposure of the relevant unpaired base. *In vitro* analyses in this study substantiate the existence of predicted SLSs in *p53* and confirm the location of hypermutable bases in loops of those structures. We propose that the intrinsic mutability of unpaired Gs and Cs in *p53*, together with the rate of transcription and availability of ssDNA, determines mutation frequency *in vivo*. The type or fate of a mutation can be intrinsic (primarily C-to-T and G-to-A) or altered by mutagens or by enzymes.

Supplementary material

Supplementary Tables 1–3 and Figures 1–3 can be found at http://carcin.oxfordjournals.org/

Funding

National Institutes of Health (RO1CA099242) and the Stella Duncan Memorial Research Institute.

Acknowledgements

We thank Drs Ralph Judd and George Card for critical reading of this paper.

Conflict of Interest Statement: None declared.

References

1. Rogel,A. *et al*. (1985) p53 cellular tumor antigen: analysis of mRNA levels in normal adult tissues, embryos, and tumors. *Mol. Cell. Bio.*, **5**, 2851–2855.
2. Hellin,A. *et al*. (1998) Nuclear factor—κB-dependent regulation of *p53* gene expression induced by daunomycin genotoxic drug. *Oncoene*, **16**, 1187–1195.
3. Li,H. *et al*. (2007) p53 promoter-based reporter gene in vitro assays for quick assessment of agents with genotoxic potential. *Acta Biochim. Biophys. Sin. (Shanghai)*, **39**, 181–186.
4. Pe,X.H. *et al*. (1999) Benzo[*a*]pyrene activates the human *p53* gene through induction of nuclear factor kappaB activity. *J. Biol. Chem.*, **274**, 35240–35246.
5. Sun,X. *et al*. (1995) Identification of a novel *p53* promoter element involved in genotoxic stress-inducible *p53* gene expression. *Mol. Cell. Biol.*, **15**, 4489–4496.
6. Hussain,S.P. *et al*. (1994) Mutagenesis of codon 248 of the human *p53* tumor suppressor gene by N-ethyl-N-nitrosourea. *Oncogene*, **9**, 13–18.
7. Singer,B. *et al*. (1982) Chemical mutagenesis. *Annu. Rev. Biochem.*, **51**, 655–693.
8. Wright,B.E. (2000) A biochemical mechanism for nonrandom mutations and evolution. *J. Bacteriol.*, **182**, 2993–3001, Mini Review.
9. Wright,B.E. (2004) Stress-directed adaptive mutations and evolution. *Mol. Microbiol.*, **52**, 643–650, Micro Review.
10. Wright,B.E. *et al*. (2002) Hypermutable bases in the *p53* cancer gene are at vulnerable positions in DNA secondary structures. *Cancer Res.*, **62**, 5641–5644.
11. Balke,V.L. *et al*. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.*, **169**, 4499–4506.
12. Krasilnikov,A.S. *et al*. (1999) Large-scale effects of transcriptional DNA supercoiling *in vivo*. *J. Mol. Biol.*, **292**, 1149–1160.
13. Liu,L.F. *et al*. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA*, **84**, 7024–7027.
14. Pruss,G.J. *et al*. (1989) DNA supercoiling and prokaryotic transcription. *Cell*, **56**, 521–523.
15. Dayn,A. *et al*. (1992) Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res.*, **20**, 5991–5997.
16. Rahmouni,A.R. *et al*. (1992) Direct evidence for the effect of transcription on local DNA supercoiling *in vivo*. *J. Mol. Biol.*, **223**, 131–144.
17. Ripley,L.S. *et al*. (1983) Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation. *Cold Spring Harb. Symp. Quant. Biol.*, **47** Pt 2, 851–861.
18. Amosova,O. *et al*. (2006) Self-catalyzed site-specific depurination of guanine residues within gene sequences. *Proc. Natl Acad. Sci. USA*, **103**, 4392–4397.
19. Bebenek,K. *et al*. (1993) Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J. Biol. Chem.*, **268**, 10324–10334.
20. Cooper,D.N. *et al*. (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum. Genet.*, **85**, 55–74.
21. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
22. Lindahl,T. *et al*. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405–3410.
23. Skandalis,A. *et al*. (1994) Strand bias in mutation involving 5-methylcytosine deamination in the human *hprt* gene. *Mutat. Res.*, **314**, 21–26.
24. Smith,K.C. (1992) Spontaneous mutagenesis: experimental, genetic and other factors. *Mutat. Res.*, **277**, 139–162.
25. Todd,P.A. *et al*. (1982) Mutational specificity of UV light in *Escherichia coli*: indications for a role of DNA secondary structure. *Proc. Natl Acad. Sci. USA*, **79**, 4123–4127.
26. Pereira,F. *et al*. (2008) Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol. Biol. Evol.*, **25**, 2759–2770.
27. Bachl,J. *et al*. (2001) Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.*, **166**, 5051–5057.
28. Datta,A. *et al*. (1995) Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science*, **268**, 1616–1619.
29. Fukita,Y. *et al*. (1998) Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity*, **9**, 105–114.
30. Reimers,J.M. *et al*. (2004) Increased transcription rates correlate with increased reversion rates in *leuB* and *argH Escherichia coli* auxotrophs. *Microbiology*, **150**, 1457–1466.
31. Wright,B.E. *et al*. (1999) Hypermutation in derepressed operons of *Escherichia coli* K-12. *Proc. Natl Acad. Sci. USA*, **96**, 5089–5094.
32. Wright,B.E. *et al*. (1981) Kinetic models of metabolism in intact cells, tissues, and organisms. *Curr. Top. Cell. Regul.*, **19**, 103–158.
33. Wright,B.E. (1986) Measuring metabolic control with kinetic models. *Trends Biochem. Sci.*, **11**, 164–165.
34. Wright,B.E. *et al*. (2003) Predicting mutation frequencies in stem-loop structures of derepressed genes: implications for evolution. *Mol. Microbiol.*, **48**, 429–441.
35. Markham,N.R. *et al*. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, W577–W581.
36. Burkala,E. *et al*. (2007) Secondary structures as predictors of mutation potential in the *lacZ* gene of *Escherichia coli*. *Microbiology*, **153**, 2180–2189.
37. Schmidt,K.H. *et al*. (2006) The effect of promoter strength, supercoiling and secondary structure on mutation rates in *Escherichia coli*. *Mol. Microbiol.*, **60**, 1251–1261.
38. Kim,H. *et al*. (2010) Transcription-associated mutagenesis increases protein sequence diversity more effectively than does random mutagenesis in *Escherichia coli*. *PLoS One*, **5**, e10567.
39. Petitjean,A. *et al*. (2007) Impact of mutant *p53* functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.*, **28**, 622–629.
40. Zheng,N.Y. *et al*. (2005) Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *J. Exp. Med.*, **201**, 1467–1478.
41. Kluyver,A.J. (1926) Die Einheit in der Biochemie in der Biochie. *Chem. D. Zelle U Gew.*, **13**, 134–190.
42. Kornberg,A. (2000) Ten commandments: lessons from the enzymology of DNA replication. *J. Bacteriol.*, **182**, 3613–3618.
43. Frederico,L. *et al*. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, **29**, 2532–2537.
44. Shen,J.C. *et al*. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.*, **22**, 972–976.

45. Cheng,K.C. *et al*. (1992) 8-Hyrdoxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. *J. Biol. Chem.*, **267**, 166–217.
46. Kascer,H. (1963) In Harris,R.J.C. (ed.) *Biological Organization at the Cellular and Super-cellular Level*. Academic Press Inc, New York, NY, pp. 25–41.
47. Krebs,H.A. (1969) The role of equilibria in the regulation of metabolism. In Horecker,B.L. and Stadtman,E.R. (eds.) *Current Topics in Cellular Regulation*. Academic Press, New York, NY, pp. 45–55.
48. Sauer,F. *et al*. (1970) Turnover rates and intracellular pool size distribution of citrate cycle intermediates in normal, diabetic and fat-fed rats estimated by computer analysis from specific activity decay data of 14C-labeled citrate cycle acids. *Eur. J. Biochem.*, **17**, 350–363.
49. Srere,P.A. (1967) Enzyme concentrations in tissues. *Science*, **158**, 936–937.
50. Vegotsky,A. *et al*. (1958) The significance of substrate concentration to *in vivo* enzyme reactions. *Enzymologia*, **19**, 143–150.
51. Wright,B.E. *et al*. (1973) Glycogen metabolism during differentiation in *Dictyostelium discoideum*. *Ann. N. Y. Acad. Sci.*, **210**, 51–63.
52. Di Noia,J.M. *et al*. (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.*, **76**, 1–22.
53. Phung,Q. *et al*. (1998) Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J. Exp. Med.*, **187**, 1745–1751.

Received February 22, 2011; revised July 13, 2011; accepted July 25, 2011