Mutations in the \( p53 \) Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular Pathogenesis

M. S. Greenblatt, W. P. Bennett, M. Hollstein, and C. C. Harris

Laboratory of Human Carcinogenesis, Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20892; and Division of Toxicology, German Cancer Research Center, D-69120 Heidelberg, Germany [M. H.]

I. Introduction

The \( p53 \) tumor suppressor gene has come to the forefront of cancer research because it is commonly mutated in human cancer and the spectrum of \( p53 \) mutations in these cancers is providing clues to the etiology and molecular pathogenesis of neoplasia (1–3). Detection of \( p53 \) abnormalities may have diagnostic, prognostic, and therapeutic implications (4).

The 15-year history of \( p53 \) investigations is a paradigm in cancer research, illustrating the convergence of previously parallel lines of basic, clinical, and epidemiological investigation and the rapid transfer of research findings from the laboratory to the clinic. \( p53 \) is clearly a component in biochemical pathways central to human carcinogenesis; \( p53 \) protein alterations due to missense mutations and loss of \( p53 \) protein by nonsense or frameshift mutations provide a selective advantage for clonal expansion of preneoplastic and neoplastic cells (5). The potential for a missense mutation to cause loss of tumor suppressor function and gain of oncogenic activity, i.e., to transform cells by two mechanisms, is one explanation for the commonality of \( p53 \) mutations in human cancer. Recent studies investigating the mechanisms underlying the biological activity of \( p53 \) indicate that the protein is involved in gene transcription, DNA synthesis and repair, genomic plasticity, and programmed cell death (1–6). These complex biochemical processes are performed by multicomponent protein machines; therefore, it is not surprising that the \( p53 \) protein forms complexes with other cellular proteins (Fig. 1) and that some viral oncoproteins alter the functions of these machines by binding to \( p53 \) and perturbing its interaction with other cellular protein components.

In this Perspective, we will focus on the origin of \( p53 \) mutations, the mutational spectrum of \( p53 \) in human cancers, and the hypotheses generated by the analysis of \( p53 \) mutations in premalignant and malignant cells. The interpretation of \( p53 \) mutations in human cancers is based on observations of the patterns of DNA damage induced by chemical and physical mutagens in model systems. In this Introduction, we will review these data, which provide the background for many of the inferences drawn from \( p53 \) mutational analysis.

A. Origins of Mutation

Both exogenous carcinogens and endogenous biological processes are known to cause mutations (7–9). Classes of DNA damage include deletion, insertion, and base substitution, either transition (change of a pyrimidine to a purine or vice versa; or a purine to another purine) or transversion (change of a pyrimidine to another pyrimidine or a purine to another purine) or transversion (change of a pyrimidine to another pyrimidine or a purine to another purine). Important sources of the spontaneous generation of point mutations in human cells include deamination of cytosine and 5-methylcytosine (11), DNA polymerase infidelity (12), depurination (13), and oxidative damage from free radicals generated by biological processes (14, 15).

The most well-studied endogenous mechanism of DNA damage at this time is the phenomenon of deamination of 5-methylcytosine. Methylation of DNA is one epigenetic mechanism involved in regulating gene expression. Cytosine and 5-methylcytosine residues can spontaneously deaminate to uracil and thymine, respectively, which if not repaired will result in G:C→A:T transitions. These mutations occur most frequently at CpG dinucleotides (a cytosine followed by a guanine), which are frequently methylated (9, 11, 16).

DNA replication in humans is performed by multicomponent protein machines that include several DNA polymerases, which vary in characteristics such as exonuclease activity and fidelity of replication. Error rates vary from 1/2,000 for DNA polymerase-\( \beta \), which repairs only short gaps, to 1/50,000–1,000,000 for other polymerases which contain exonuclease activity (17). The most common base substitutions observed are G:C→A:T and A:T→G:C transitions and G:C→T:A transversions. Imbalances in deoxynucleoside triphosphate pools (18), mutations in DNA polymerase-\( \alpha \) (19), and slippage of DNA polymerase at nucleotide repeats (Ref. 8; section II.B.2) are examples of mechanisms contributing to fidelity of DNA synthesis. Considering the multiplicity of proteins involved in DNA synthesis, DNA repair, mitosis, and the cell cycle, the number of potential gene targets which could interfere with replication fidelity in somatic cells is large.

Abnormalities in some of these processes could result in the mutator phenotype (20, 21), which could increase the probabilities of both neoplastic transformation and the generation of increasingly malignant subclones during tumor progression (22, 23). If such mutations occur in germ cells and are nonlethal, they could lead to heritable syndromes of increased cancer risk. The hypothesis that inherited genomic instability may be associated with increased cancer risk is supported by studies of two diseases. Children with Bloom's syndrome have an increased risk of childhood cancers, and their cells show an increase in cytogenetic abnormalities (24). Germline mutations in two DNA mismatch repair genes, \( hMSH2 \) (25, 26) and \( hMLH1 \) (27, 28), recently have been identified as the genetic abnormalities responsible for HNPCC.3 Mutations of these genes are

---

3 The abbreviations used are: HNPCC, hereditary nonpolyposis colorectal carcinoma; aprt, adenine phosphoribosyl transferase; dfr, dihydrofolate reductase; hprt, hypoxanthine-guanine phosphoribosyl transferase; CHO, Chinese hamster ovary; BPDE, 7,8-dihydro-9,10-epoxynitro-7,8,9,10-tetrahydrobenzo(a)pyrene; BP, benzo(a)pyrene; IHC, immunohistochemistry; SSCP, single-stranded conformation polymorphism; PCR, polymerase chain reaction; XPC, xeroderma pigmentosum, complementation group C; PAH, polycyclic aromatic hydrocarbon; AFB, aflatoxin; HCC, hepatocellular carcinoma; HBV, hepatitis B virus; SCLC, small cell lung carcinoma; NSCLC, non-small cell lung carcinoma; NPC, nasopharyngeal carcinoma; EBV, Epstein-Barr virus; HPV, human papilloma virus; CAT, chloramphenicol acetyltransferase; LFS, Li-Fraumeni Syndrome; MP, mean pairwise.

Received 4/27/94; accepted 7/20/94.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Due to space limitations and the breadth of this topic, comprehensive reviews which are the source of additional citations will be cited frequently.

2 To whom requests for reprints should be addressed, at Laboratory of Human Carcinogenesis, Building 37, Room 2C05, National Cancer Institute, Bethesda, MD 20892.
Fig. 1. Schematic of the p53 molecule. Vertical lines, distribution of mutations; horizontal bars, regions of protein binding. The p53 protein consists of 393 amino acids with functional domains, evolutionarily conserved domains, and regions designated as mutational hotspots. Functional domains include the transactivation region (amino acids 20—42, orange area), sequence-specific DNA binding region (amino acids 100—293, dark green area), and oligomerization region (amino acids 319—360, light green area). Cellular or oncoviral proteins bind to specific areas of the p53 protein. Evolutionarily conserved domains (amino acids 17—29, 97—292, and 324—352; rose-colored area) were determined using the MACAW program as described in the text. Seven mutational hotspot regions within the large conserved domain are identified: amino acids 130—142, 151—164, 171—181, 193—200, 213—223, 234—258, and 270—286. Functional domains and protein binding sites were compiled from references in text and from X. Wang and C. Harris (unpublished data). Vertical lines above the schematic, missense mutations; lines below schematic, nonmissense mutations. The majority of missense mutations are in the conserved hydrophobic midregion, while nonmissense mutations (nonsense, frameshift, splicing, and silent mutations) are distributed throughout the protein, determined primarily by sequence context, as described in the text.

Fig. 2. A-G. Spectrum of p53 mutations in various tumor types. (n) the number of mutations for which exact base change or deletion was confirmed by DNA sequencing. A, the mutational spectrum of p53 in all reported tumors and cell lines (n = 2567) in the database. B, squamous (n = 32, including Bowen’s disease) and C, basal cell (n = 52) carcinomas of the skin exhibit frequent G:C—÷A:T transitions and occasional tandem mutations at dipynmidine sites (usually CC:GG—÷TT:AA), both characteristic of UV-induced damage. Squamous cell carcinomas have been linked with occupational exposure to PAHs, and one-fourth contain G:C—÷T:A transversions, which are consistent with the mutations caused by these carcinogens in model systems. The mutational spectra differ by site of origin of head and neck squamous carcinomas [(E) nasopharynx, n = 17; (F) oral, n = 69; (G) pharynx/larynx, n = 42].
Fig. 2. (continued) H-O, spectrum of p53 mutations in various tumor types. The mutational spectra in lung cancers differ among histological types [(H) adenocarcinoma, n = 61; (I) squamous, n = 75; (J) small cell (SCLC), n = 92; (K) large cell, n = 27]. Adenocarcinomas display a lower mutation prevalence, a lower percentage of G:C→T:A transversions, and a higher percentage of G:C→A:T transitions. The mutational spectra in lung cancer vary by smoking history [(L) smokers, n = 169; (M) non-smokers, n = 16]. The most common mutations in (N) colon carcinoma (n = 324) and (O) LFS (germline p53 mutations, n = 47) are G:C→A:T transitions at CpG dinucleotides, which may be the result of spontaneous deamination of 5-methylcytosine.
associated with deletions and insertions in microsatellite sequences of somatic cells and a genetic predisposition to tumors of the colon and other tissues (29). Somatic mutations in hMLH2 have also been observed. Therefore, the mutator phenotype can be both a predisposing inherited factor initiating carcinogenesis or a secondary somatic change produced during this multistage process.

A well-studied mechanism by which chemical carcinogens and their ultimate carcinogenic metabolites cause mutations is by forming covalent adducts with the nucleotides in DNA, increasing the probability of errors during DNA replication (30). Some small carcinogen-DNA adducts, such as O6-methylguanine resulting from alkylating agents, may cause DNA polymerase to misread the base pairing due to the altered hydrogen bonding properties of a base which contains an additional methyl or ethyl group. Laboratory studies have established that the most common mutations caused by alkylating agents are G:C→A:T transitions, consistent with O6-methylguanine mispairing with thymine (31). Bulky carcinogen-DNA adducts may render the bases unreadable and stall the replication machinery. DNA polymerases may fill these noninstructive sites preferentially with adenine. If this occurs opposite to a guanine (a common target for bulky carcinogens), a thymine will pair with the adenine in the next round of DNA synthesis, resulting in a G:C→A:T transversion (32). Studies using prokaryotic and eukaryotic cells and site-specific mutagenesis assays (13, 33) have shown that some carcinogenic agents produce a "fingerprint" (1, 34, 35); specific types and locations of DNA adducts have been linked with the mutational spectrum of an agent. These patterns have been studied in eukaryotes using both exogenous genes introduced into the appropriate host cell by a shuttle vector (33, 36) and endogenous genetic loci (34) such as aprt, dhfr, and hprt.

B. Mutations in Model Systems

The carcinogen-induced mutational spectra in prokaryotes and in target genes in eukaryotes can be compared to spontaneous mutational spectra, although uncertainty exists in the extrapolation of in vitro data to human carcinogenesis. Confounding variables which could affect experimental systems include differences in the genes and cell types studied, interspecies variability of DNA repair, and the potential mutagenicity of transfection and shuttle vectors. The interpretation of mutational spectra is complicated, and observed mutations may not reflect all events at the DNA level. The DNA sequence can influence both carcinogen-DNA adduct formation and DNA repair, and some mutations may be either more readily detected due to clonal expansion or less readily detected due to inhibition of the mutant cell population.

1. Mutational Spectra in Housekeeping Genes

The majority of spontaneous mutations in the model systems used are single base substitutions. Studies in prokaryotes indicate that carcinogen-induced mutations occur at nucleotides which are targets for carcinogen-DNA interactions and alterations (37). Differences have been observed between the spontaneous mutational spectra of the endogenous aprt gene of CHO cells and both bacterial genes and exogenous genes carried by shuttle vectors in mammalian cells (38). This may reflect either cell-specific or gene-specific biological properties.

A comparison of the spontaneous mutational spectrum at the aprt locus in CHO cells with the spectra induced by UV radiation, ionizing radiation, or BPDE illustrates carcinogen-specific damage (Table 1). The mutational spectrum of UV light is consistent with the mutagenesis models of the promutagenic cyclobutane and pyrimidine-pyrimidine (6-4) photoproducts. Ionizing radiation often causes chromosomal abnormalities, deletions, and losses (39, 40). BPDE, which primarily binds to the 2-amino group of deoxyguanine, produces predominantly the G:C→A:T transversions predicted of a bulky carcinogen-DNA adduct; a similar mutational spectrum for BPDE has been observed in human cell assays (41). The hprt locus also allows comparison of mutational spectra obtained by exposing cultured cells to carcinogens with those observed in human donor lymphocytes exposed in vivo to environmental carcinogens (42). A recent comprehensive database has been developed which contains more than

Table 1  Spontaneous and induced mutational spectra: in vitro and animal models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO aprt gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous (30)</td>
<td>71</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>UV light (31)</td>
<td>52</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Ionizing radiation</td>
<td>19</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>13</td>
<td>19</td>
<td>0</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>BPDE (21)</td>
<td>5</td>
<td>62</td>
<td>14</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Human aprt gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous (74)</td>
<td>28</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>NR</td>
<td>15</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>UV light (55)</td>
<td>47</td>
<td>13</td>
<td>7</td>
<td>15</td>
<td>9</td>
<td>7</td>
<td>NR</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>BPDE (102)</td>
<td>14</td>
<td>69</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>NR</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MNNG (103)</td>
<td>63</td>
<td>13</td>
<td>2</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>NR</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ENU (81)</td>
<td>31</td>
<td>9</td>
<td>5</td>
<td>17</td>
<td>25</td>
<td>9</td>
<td>NR</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Animal p53 gene†‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse skin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UV light (33)</td>
<td>42</td>
<td>18</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BP (7)</td>
<td>14</td>
<td>71</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Butadiene (10)†</td>
<td>10</td>
<td>40</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>DMBA (18)</td>
<td>22</td>
<td>17</td>
<td>11</td>
<td>11</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>MCA (7)</td>
<td>29</td>
<td>57</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rat esophagus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMBA (10)</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Rat liver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choline-deficient diet (9)</td>
<td>55</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
</tbody>
</table>

* Values, percentages of each class of mutation; MNNG, N-methyl-N'-nitro-N-nitrosoguanidine; ENU, N-ethyl-N'-nitrosourea; DMBA, dimethylnitrosamine; MCA, methylcholanthrene; NMBA, nitrosomethylbenzylamine; CHOe Chinese Hamster Ovary; Ref., reference no; NR, not reported.

† Number of mutations sequenced.
‡ Deletions.
§ Insertions.
R. Wiseman, personal communication.
Examples of mutation prevalence may be found in Refs. 287 and 288.

4858
1000 human hprt mutations (43). The mutational spectra found at the aprt locus of CHO cells and the hprt locus of human cells are detailed in Table 1. The most striking similarities are the high percentages of G:C—A:T transitions in BDE and high percentages of G:C—T:A transversions in UV-exposed cells, consistent with models predicted for these agents. The largest difference between the two systems is in the spontaneous mutational spectrum. In the aprt locus, G:C—A:T transitions account for 71% of the 30 spontaneous alterations, whereas in the hprt locus, only 28% of the 74 mutations are G:C—A:T (P = 0.001, Fisher’s exact test). Thus, cell-specific or gene-specific features may affect spontaneous mutation patterns.

2. Mutational Spectra in Animal Models

Animal models provide opportunities for mutational spectrum studies that cannot be performed with human tumors. For example, dose response experiments can define the mutational spectra of carcinogens in vertebrate neoplasms. The spectrum of carcinogen-induced mutations in cancer-related and housekeeping genes can be compared in rodent and human cells, assessing the relevance of these model systems to human carcinogenesis. Furthermore, transgenic animals can provide mechanistic insights into the interplay between environmental carcinogens and host susceptibility factors such as carcinogen metabolism and alterations of oncogenes and tumor suppressor genes.

The frequency of p53 alterations in murine primary tumors is generally lower than in human tumors. p53 mutations have been found in tumors of the breast (42%), lung (8%), skin (30%), and other tissues (Table 1). Alterations generally have not been found in colon or liver tumors. The low prevalences in lung and colon tumors are surprising for several reasons: (a) the gene is highly conserved between mice and humans; (b) p53 mutation frequency is high in human lung and colon cancers; (c) the morphological features of lung and colon adenocarcinomas are similar in humans and mice; and (d) lung and colon tumors from both species frequently contain ras mutations. The mutation database in laboratory animals is still limited (Table 1), and more studies of different carcinogens and dose regimens are needed.

The most well-studied system is skin carcinogenesis in the mouse. Several carcinogens have been examined in rat, hamster, and nonhuman primate models, but mutation rates are generally too low to generate meaningful spectra.

UV exposure efficiently induces squamous skin carcinomas in several strains of inbred mice (44, 45). In these studies, as in human tumors, there is a striking preference for single base transitions and tandem mutations at dipyrimidine sites (30 of 33; 91% of mutations). BP, a carcinogen found in tobacco smoke and burning fossil fuels, induces papillomas and squamous carcinomas in Sencar mice when applied topically. p53 mutations were identified in 1 of 8 papillomas and 6 of 12 carcinomas (46) with 5 G:C—T:A transversions, 1 G:C—G:C transversion, and 1 G:C—A:T transversion (Table 1). These data provide the first characterization of the p53 mutation spectrum of BP-induced tumors in vertebrates. Skin tumors induced by 7,12-dimethylbenz(a)anthracene provide a contrasting mutation spectrum. Eighteen p53 mutations have been reported (46–48), and the mutation pattern is notable for a high frequency (40%) of mutations at A:T base pairs and a low frequency of G:C—T:A transversions. This is consistent with the mutagenic specificity of 7,12-dimethylbenz(a)anthracene (48).

These mutational spectra of aprt, dbsf, and hprt genes and animal models serve as a foundation with which to compare the spectra observed in cancer-related genes in humans. Differences in spontaneous mutation patterns may reflect differences in inherent characteristics of various cell types (e.g., DNA repair capacities), whereas induced mutation patterns that override the spontaneous profiles are more likely to reflect mutagen-DNA interactions.

II. p53 Mutational Spectrum Analysis in Human Cancers

One of the goals of carcinogenesis research is to identify the precise molecular alterations responsible for neoplastic transformation, abnormal differentiation, and growth control. Molecular epidemiology has developed to attempt to integrate traditional epidemiological investigation of cancer risk factors with the substantial expansion of knowledge of the molecular mechanisms of cellular processes (49).

Mutational spectrum analysis, the study of the types and locations of DNA alterations, describes the often characteristic patterns of DNA changes caused by endogenous and exogenous mutagens. Alterations of cancer-related genes found in tumors not only represent the interactions of carcinogens with DNA and cellular DNA repair processes but also reflect the selection of those mutations which provide pre-malignant and malignant cells with a growth advantage. Study of the frequency, timing, and mutational spectra of abnormalities of p53 and other cancer-related genes can thus generate and test a variety of hypotheses in carcinogenesis. These include questions regarding carcinogen-DNA interactions, functions of the affected gene products, mechanisms of carcinogenesis in specific organs or tissues, and features of general cell biology such as DNA replication and repair.

The information derived from this type of mutational spectrum analysis is most relevant when it addresses alterations of genes vital to carcinogenesis. A burgeoning literature (reviewed in Refs. 1, 2, 5, and 6) testifies to the important role of p53 protein in normal cell function and neoplastic transformation. Mutations of p53 are the most common genetic abnormality yet found in human cancers. The prevalence of p53 mutations varies among tumor types, ranging from 0 to 60% in major cancers (Table 2), and is over 80% in some histological subtypes. All classes of mutations occur in the p53 gene (Fig. 2A); three-fourths of substitutions occur at G:C base pairs.

The p53 gene is well-suited to mutational spectrum analysis for several reasons. First, since p53 mutations are common in many human cancers, a sizable database has accrued whose analysis can yield statistically valid conclusions (50). Its modest size (11 exons, 393 amino acids) permits study of the entire coding region, and it is highly conserved in vertebrates, allowing extrapolation of data from animal models (51). Point mutations which alter p53 function are distributed over a large region of the molecule, especially in the hydrophobic midportion (1), where many base substitutions alter p53 conformation and sequence-specific transactivation activity (sec. II.C.1); thus, correlations between distinct mutants and functional changes are possible. Frameshift and nonsense mutations which truncate the protein can be located outside of these regions, so evaluation of the entire DNA sequence yields relevant data. This situation differs from that of the ras oncogenes. Transforming mutations of ras occur primarily in three codons and are thus limited to a few sequence motifs which identify a critical functional domain (52). The diversity of p53 mutational events permits more extensive inferences regarding mechanisms of DNA damage and mutation.

For this Perspective, we compiled 2567 mutations in human carcinomas or cell lines from over 300 papers published or in press through January 1, 1994, as identified by searches of Medline and Current Contents in January, 1994. Only mutations confirmed by DNA sequencing were entered into the database. Selected additional reports were added to illuminate specific discussions. An updated version of the database is available from the European Molecular Biology Laboratory Data Library through e-mail; details of its compilation have been reported separately (50). Here we analyze these data with attention to: (a) questions regarding carcinogenesis which have been...
Table 2: Prevalence and spectra of p53 mutations in human cancers

|--------------|------------------|------------|------------|------------|------------|------------|------------|--------------|-------|-----------
| All tumors   | 37               | 41         | 17         | 8          | 11         | 6          | 4          | 13           | 24    | 175 R>H   |
|              |                  |            |            |            |            |            |            |              |       | 245 G>S,D,C,V |
|              |                  |            |            |            |            |            |            |              |       | 248 R>W,Q   |
|              |                  |            |            |            |            |            |            |              |       | 249 R>S,M   |
|              |                  |            |            |            |            |            |            |              |       | 273 R>H,C   |
|              |                  |            |            |            |            |            |            |              |       | 157 V>F     |
|              |                  |            |            |            |            |            |            |              |       | 248 R>W,L,Q |
|              |                  |            |            |            |            |            |            |              |       | 249 R>M,S   |
|              |                  |            |            |            |            |            |            |              |       | 273 R>L,H,C |
|              |                  |            |            |            |            |            |            |              |       | 175 R>H     |
|              |                  |            |            |            |            |            |            |              |       | 245 G>S,D,V |
|              |                  |            |            |            |            |            |            |              |       | 248 R>W,Q   |
|              |                  |            |            |            |            |            |            |              |       | 273 R>H,C   |
|              |                  |            |            |            |            |            |            |              |       | 282 R>W     |
|              |                  |            |            |            |            |            |            |              |       | None       |
| Lung         | 56               | 24         | 40         | 9          | 7          | 5          | 3          | 11           | 9     | 175 R>H     |
|              |                  |            |            |            |            |            |            |              |       | 245 G>S,D,C,V |
|              |                  |            |            |            |            |            |            |              |       | 248 R>W,Q   |
| Colon        | 50               | 63         | 9          | 3          | 11         | 4          | 1          | 8            | 47    | 248 R>W     |
|              |                  |            |            |            |            |            |            |              |       | 278 P>S,F   |
|              |                  |            |            |            |            |            |            |              |       | None       |
| Esophagus    | 45               | 38         | 16         | 3          | 15         | 14         | 3          | 12           | 18    | 273 R>CH    |
| Ovary        | 44               | 42         | 9          | 12         | 18         | 3          | 6          | 9            | 23    | 273 R>H     |
| Pancreas     | 44               | 41         | 13         | 6          | 17         | 8          | 5          | 10           | 21    | 273 R>H     |
| Skin         | 44               | 41         | 12         | 6          | 4          | 4          | 2          | 31           | 17    | 248 R>W     |
| Gastric      | 41               | 47         | 6          | 3          | 17         | 5          | 6          | 16           | 35    | 278 P>S,F   |
| Head and neck| 37               | 31         | 18         | 11         | 14         | 5          | 3          | 19           | 13    | 248 R>Q,W,L |
| Bladder      | 34               | 37         | 14         | 21         | 9          | 3          | 2          | 14           | 16    | 280 R>T     |
| Sarcoma      | 31               | 47         | 14         | 6          | 7          | 6          | 1          | 19           | 21    | None       |
| Prostate     | 30               | 44         | 13         | 6          | 16         | 0          | 6          | 16           | 28    | None       |
| Hepatocellular| 29              | 20         | 37         | 8          | 11         | 12         | 3          | 9            | 9     | 249 R>S,M,W |
| Brain        | 25               | 50         | 11         | 5          | 14         | 1          | 3          | 15           | 30    | 175 R>H     |
| Ade n al     | 23               | 36         | 13         | 8          | 11         | 7          | 6          | 16           | 23    | 248 R>Q,W   |
| Breast       | 22               | 51         | 10         | 4          | 14         | 0          | 6          | 14           | 37    | 273 R>HC    |
| Endometrium  | 22               | 22         | 13         | 8          | 11         | 7          | 6          | 16           | 23    | 175 R>H     |
| Mesothelioma | 22               | 13         | 9          | 13         | 9          | 0          | 4          | 17           | 28    | 248 R>Q,G,W |
| Renal        | 19               | 13         | 9          | 13         | 9          | 0          | 4          | 17           | 28    | 248 R>Q,G,W |
| Thyroid      | 13               | 48         | 9          | 13         | 9          | 0          | 4          | 17           | 28    | 248 R>Q,G,W |
| Hematological| 12               | 49         | 7          | 7          | 12         | 4          | 5          | 15           | 31    | 175 R>H     |
| Carcinoid    | 11               | 36         | 13         | 8          | 11         | 7          | 6          | 16           | 23    | 248 R>Q,W   |
| Melanoma     | 9                | 35         | 10         | 4          | 14         | 0          | 6          | 14           | 37    | 273 R>C     |
| Parathyroid  | 8                | 35         | 10         | 4          | 14         | 0          | 6          | 14           | 37    | 273 R>C     |
| Cervix       | 7                | 35         | 10         | 4          | 14         | 0          | 6          | 14           | 37    | 273 R>C     |
| Neuroblastoma| 1                | 0          | 0          | 0          | 10         | 0          | 0          | 10           | 25    | 273 R>C     |
| Wilms        | 0                | 0          | 0          | 0          | 10         | 0          | 0          | 10           | 25    | 273 R>C     |
| Testes       | 0                | 0          | 0          | 0          | 10         | 0          | 0          | 10           | 25    | 273 R>C     |
| Pituitary    | 0                | 0          | 0          | 0          | 10         | 0          | 0          | 10           | 25    | 273 R>C     |
| Pheochromocytoma| 0        | 0          | 0          | 0          | 10         | 0          | 0          | 10           | 25    | 273 R>C     |

* Mutational spectrum of a tumor type was calculated only if ≥20 mutations were reported.
* a Number of tumors of each cell type evaluated for p53 mutation by PCR-based techniques as compiled in the p53 mutation database (50).
* b Prevalence of p53 mutations in cancers from each organ/tissue, as detected by PCR-based techniques, screening at least exons 5–8. Premalignant lesions such as adenomas or dysplastic lesions are not included in prevalence data. Prevalence in all cancers is a crude estimate derived from statistics on worldwide cancer incidence and the prevalence of p53 mutations for each type as calculated from the reports in the mutation database (see text).
* c Del, deletion; Ins, insertions.
* d % of mutations which are G:C→A:T transitions at CpG dinucleotides.
* e Hotspots for the entire database are defined as codons at which at least 50 mutations have been reported. Hotspots for individual tumor types include codons at which at least 10 mutations, or 10% of all mutations for that tumor, have been reported. Letters refer to amino acid change, using standard one letter code. Amino acid substitutions which occur more than once are listed in descending order of frequency at each codon.
* f CpG site.

4860
addressed by mutation analysis, and conclusions reached to date; (b) how mutational spectrum analysis and in vitro experiments provide complementary information on carcinogenesis; and (c) new hypotheses generated from the accumulated information to be specifically addressed by future studies of p53 mutations.

A. Analytical/Technical Considerations

1. Methods of Analysis for p53 Abnormalities

The methods of analysis for mutations can bias conclusions regarding the prevalence and nature of p53 mutations. A few caveats concerning techniques and experimental design are warranted before examining the data in detail.

Solid tumors are heterogeneous, consisting of tumor cells mixed with nonneoplastic stromal cells and necrotic debris. Since tumors undergo clonal evolution, different sites within a malignancy may also vary in genotype and phenotype (21); a false negative result might arise from sampling error. Microscopic assessment is important to ensure that tissue samples contain adequate amounts of the desired histological features and to avoid analysis of necrotic debris. Samples may be enriched for tumor content by several methods, such as manual microdissection of slides (removing only the tumor tissue for study) and cell sorting using DNA flow cytometry (53, 54).

DNA sequencing is required for precise identification of mutations, but an altered p53 gene or protein can be detected by other immunological or chemical tests (55–58). Because DNA sequencing is time and labor intensive, studies evaluating human cancers for p53 mutations often use screening tests, performing DNA sequencing only on tumors which are likely to contain mutations. The sensitivity and specificity of these techniques are critical factors which could introduce bias into p53 mutation prevalence data.

The wild-type p53 protein usually resides in the cell nucleus, but has a very short half-life and is present in such small quantities that it cannot be detected by routine IHC. Missense mutations often increase the half-life and quantity of the p53 protein, allowing its detection by IHC (58). Many investigators have used IHC as a screening test prior to DNA sequencing, and staining is often considered a surrogate marker for gene mutation even in the absence of confirmatory DNA studies. In the 84 studies we identified which report IHC and sequencing, in the same tumor sets, positive staining was found in 44% of tumors by IHC, while 36% contained mutations. However, the sensitivity of IHC in these studies was only 75% (range, 36–100%), and the positive predictive value was only 63% (range, 8–100%), with considerable variation among tumor types. Thus, the status of the p53 protein by IHC should not be equated with wild-type or mutant genotype.

Discrepancies between IHC and sequencing results may be due to false negative and positive results from technical aspects of sample preparation and analysis or the result of biological phenomena responsible for the discordance (reviewed in Ref. 58). Mutations which result in deletion or truncation of the protein (nonsense and frameshift) do not cause protein accumulation. Therefore, IHC will be less sensitive in detecting mutations in tumor types with high proportions of these nonmissense mutations. Some tumors, including melanoma (59, 60) and testicular carcinoma (61, 62), are frequently immuno-positive for p53 in the absence of mutation. Potential mechanisms for positive IHC due to accumulation and staining of nonmutant p53 protein include stabilization of p53 protein by binding to viral or cellular proteins and DNA damage from chemical or physical genotoxic agents (58). Another hypothesis is that mutations are occurring in a gene downstream of p53, such as WAF1/Cip1, a Mr 21,000 protein which is transactivated by the p53 protein and is an inhibitor of cyclin-dependent kinases (63, 64). Interruption of the p53 pathway could interfere with a feedback loop which regulates expression of wild-type p53. This can be tested by searching for mutations in genes downstream in the p53 biochemical pathway in the cancers in which staining is frequently detected but which rarely contain p53 mutations.

Other methods of screening tumors to identify those likely to contain mutations involve detection of altered electrophoretic mobility patterns of DNA sequences containing point mutations. The most common techniques, SSCP, constant denaturant gradient gel electrophoresis, and constant denaturant gel electrophoresis (68, 69), have a sensitivity and specificity of approximately 90% in detecting mutations confirmed by DNA sequencing (67, 69, 70). Sequencing should be done to confirm that positive results represent mutations and not technical artifacts or germline polymorphisms, which have been described at codons 72 and 213 (71, 72).

PCR-based techniques, including SSCP, constant denaturant gel electrophoresis, and DNA sequencing, are subject to false negative and false positive results. The polymerases used in PCR to amplify specific gene sequences are subject to error rates from 1/10,000 to more than 1/500 base pairs, depending on reaction conditions (73). Large deletions might not be detected if the missing segment includes primer sites, and the wild-type gene of stromal tissue is amplified. Contamination of a PCR reaction with other airborne PCR products may also lead to false positive or negative sequencing results (74). Therefore, it is imperative that all mutations found by direct sequencing of PCR products be confirmed by sequencing of a second, independent PCR product. This practice has not been followed uniformly; hence, some mutations in the database may represent false positives. There also is considerable inconsistency among papers in the extent of analysis. Mutation prevalence data are included in this report only if a minimum of exons 5–8 was sequenced. Due to underdetection, mutation prevalence rates are likely underestimated in the database, although reported rates for some tumors were probably inflated due to preselection of samples by screening tests. Thus, absolute mutation prevalence rates are difficult to assess, given the heterogeneity of techniques in cited reports.

2. Extent of p53 Gene Analysis

The p53 gene consists of 11 exons; exons 2–11 code for the protein of 393 amino acids. Early studies noted that most p53 mutations occurred in the regions of the gene which are highly conserved through evolution and presumably of functional importance, primarily in exons 5–8 (codons 126–306) (55). Based on these preliminary data, most investigators have confined their analyses to exons 5–8 (and sometimes exon 4). Partly because of this bias in searching, 95% of the reported mutations have been found in exons 5–8 and their intervening introns (mutations in introns which affect splicing of the next exon have been classified as mutations of the deleted exon). We have identified 50 studies in which sequencing of the entire coding region of p53 was reported. Of the 560 mutations reported in these papers, 87% were in exons 5–8, and most of the others were in exons 4 (8%) and 10 (4%). The distribution varied among tumor types. Mutations outside exons 5–8 were most frequent in bladder (28%) and hepatocellular (24%) carcinomas and least frequent in head and neck and hematopoietic malignancies (2% each). Over 70% of these mutations generated stop codons or frameshifts; therefore, they likely would be missed by IHC screening.

Thus, evaluation of only exons 5–8 is likely to underestimate the prevalence of p53 mutations, especially nonmissense mutations, by over 20% in some tumors. Analysis of exons 4 and 10 is therefore warranted when mutations are not found in exons 5–8. The tissue-specific differences in the prevalence of mutations in various exons...
are clues that detailed p53 mutational spectrum analysis can reveal subtle patterns of genetic damage.

Another potential source of underestimating the prevalence of p53 mutations might be the presence of mutations in unevulated DNA sequences, such as its 2 promoters or its 10 introns. Although the intron bases adjacent to exons, which are responsible for splicing, are often examined, complete evaluation of introns is not common. Reverse transcription-PCR, a technique using reverse transcriptase to amplify mRNA and sequence the resulting complementary DNA, cannot evaluate introns, which have been removed by splicing. Mutations in intronic regulatory regions may be prevalent and undetected by screening with IHC (if protein does not accumulate) and PCR-based techniques (if primers do not encompass the relevant intron sequence).

3. Epidemiological Considerations

Specific features of populations evaluated for p53 mutations may also create bias in the conclusions. For some cancer types, differences in frequency or spectrum have been detected between groups which differ in ethnicity or nationality, possibly due to specific carcinogenic exposures or inherited features in these populations (sec. II.D.3.). If atypical subgroups are overrepresented, features of the mutational spectrum of other groups may be obscured.

In order to yield valid conclusions, molecular epidemiology must adhere to standard epidemiological principles of statistical analysis: i.e., selection of appropriate control and study groups, assessment for potential confounding variables (which may be different for each tumor type), and adequate sample sizes to minimize type I errors (e.g., the premature assignment of relationships between putative carcinogens and cancers) and provide the statistical power to detect important differences and avoid type II errors (the erroneous conclusion that no difference exists). Although these admonitions may seem elemental, the literature contains many examples of preliminary conclusions based on small sample sizes, which often pass into general opinion without sufficient critique. The database provides a powerful tool to highlight the areas in which data are inadequate and focus future endeavors.

B. Patterns of Mutation: Insights into DNA Replication and Repair

1. DNA Repair and Strand Bias of Mutations

Since all organisms are constantly exposed to exogenous and endogenous mutagens, sophisticated systems of DNA repair to maintain the integrity of the genome have evolved. DNA repair may be defined as a cellular response which restores the normal nucleotide sequence and stereochemistry of DNA following damage. DNA damage often causes a physical block to transcription; therefore, repair is generally faster in actively transcribed genes and is closely linked to other cellular pathways which control gene expression (75). Several recently described genes code for enzymes exhibiting both DNA repair and gene transcription activities, suggesting a potential molecular mechanism for control of cellular proliferation until damage is corrected (76, 77). In addition, the p53 gene is repaired faster than the dhfr gene, and much more efficiently than the inactive genomic regions (78). In vitro studies and observations in repair-deficient organisms indicate that repair is important in determining mutation patterns.

Importantly, repair is faster on the transcribed (template) strand of DNA than on the nontranscribed (coding) strand (75). Experiments in bacteria, yeast, and mammalian cells and in vivo mouse skin carcinogenesis models have shown that mutations induced by exogenous carcinogens preferentially occur on the nontranscribed strand of genes such as dhfr, hprt (79), and p53 (46, 75). That is, the mutations induced on the nontranscribed (coding) strand by UV light or BP are much more likely to be CC→TT than GG→AA and G→T than C→A, respectively. This difference is thought to be an evolutionary adaptation which ensures that structure and function of essential proteins with short half-lives are maintained, increasing the likelihood of cell survival. Repair of the nontranscribed strand can occur later and still preserve the fidelity of the gene in the daughter cells. If repair of the nontranscribed strand does not take place, then daughter cells might arise with mutations that are lethal or that confer a survival advantage on the mutant clone.

The hypothesis that deficient DNA repair can affect mutation patterns in humans was tested by measuring the repair of UV-induced cyclobutane pyrimidine dimers in the individual DNA strands of p53 in a normal, repair-proficient human fibroblast strain and in fibroblasts from a patient with the repair-deficient disorder XP-C (78). Selective repair of the transcribed DNA strand of p53 is observed in both normal and XP-C fibroblasts, and the strand bias of repair is more pronounced in XP-C cells. Since mutations may occur due to replication errors at the sites of unrepaired DNA damage, these results predict that the majority of p53 mutations in skin cancers, especially those from patients with XP-C, would occur on the nontranscribed strand. Indeed, 100% of tandem mutations in skin cancers from XP-C patients and four of seven from non-XP patients exhibit such a strand bias (80, 81). More data in non-XP cancers are needed. Tumors from XP-C patients are much more likely than sporadic cancers to contain tandem pyrimidine transitions (58 versus 11%), suggesting a differential ability to repair this type of DNA damage. This sequence of studies is an example of testing in the laboratory of a hypothesis generated by analysis of the p53 mutation spectrum and then predicting the result of further molecular epidemiological investigation of human tumors.

Some tumor types have been linked by epidemiological and laboratory evidence to specific carcinogens with known mutational spectra, such as lung cancer and exposure to cigarette smoke (section II.D.5.C). Ninety-one percent of G:C→T:A transitions in lung cancer occur on the nontranscribed DNA strand (i.e., G→T rather than C→A), supporting the hypothesis that DNA repair is an important determinant of mutation specificity of cancer-related genes in human cancer. The hypothesis may be extended to postulate that a nontranscribed strand bias in the mutation pattern of a specific base substitution or tumor type implies that an exogenous carcinogen may be responsible. Evaluating the entire p53 database for strand bias may therefore provide clues to suspected and unrecognized carcinogens.

Four patterns of base substitution exhibit strand bias: G→T (87% on the nontranscribed strand) and T→G (67%) transversions, and A→G (70%) and non-CpG site G→A (65%) transitions (Table 3). G:C→A:T transitions occurring at CpG dinucleotides are reported separately from other G:C→A:T transitions because of differences in mechanism discussed in section I. Bulky carcinogens such as BP, other PAHs (82), and AFB (82–84) have been associated with transitions and transversions at both G:C and A:T base pairs (85). The relative contributions of these and other agents to human somatic mutations are unknown.

In support of the hypothesis that tumors which are associated with environmental carcinogens should contain a high proportion of base substitutions with bias toward the nontranscribed strand, Table 3 lists all tumor types in descending frequency of the prevalence of the base substitution types which exhibit strand bias (G→T, non-CpG site G→A, A→G, and T→G). Among the tumors which most frequently contain these types of mutations are tobacco-associated cancers of the lung (most frequent), cervix, esophagus, head and neck (fourth, fifth, and sixth most frequent, respectively), and hepatocellular carcinoma (second most frequent), which is associated with dietary AFB exposure in certain geographic regions (section II.D.5.B). Most of these
Table 3 DNA strand bias by tumor

<table>
<thead>
<tr>
<th>Tumor type (n)</th>
<th>Total % of Muts at strand bias site</th>
<th>% G→T via strand bias</th>
<th>% G→A non-CpG</th>
<th>% A→G</th>
<th>% T→C</th>
<th>% G→T on nontranscribed strand</th>
<th>% G→A on nontranscribed strand</th>
<th>% A→G on nontranscribed strand</th>
<th>% T→G on nontranscribed strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tumors (2567)</td>
<td>50</td>
<td>17</td>
<td>17</td>
<td>11</td>
<td>4</td>
<td>87</td>
<td>65</td>
<td>70</td>
<td>67</td>
</tr>
<tr>
<td>Lung (328)</td>
<td>66</td>
<td>40</td>
<td>15</td>
<td>7</td>
<td>3</td>
<td>95</td>
<td>52</td>
<td>91</td>
<td>64</td>
</tr>
<tr>
<td>HCC (224)</td>
<td>61</td>
<td>38</td>
<td>10</td>
<td>11</td>
<td>3</td>
<td>89</td>
<td>77</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>Pancreas (63)</td>
<td>56</td>
<td>13</td>
<td>21</td>
<td>17</td>
<td>5</td>
<td>100</td>
<td>69</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>Cervix (20)</td>
<td>55</td>
<td>35</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>86</td>
<td>100</td>
<td>NA*</td>
<td>50</td>
</tr>
<tr>
<td>Esophagus (115)</td>
<td>53</td>
<td>16</td>
<td>19</td>
<td>15</td>
<td>3</td>
<td>89</td>
<td>55</td>
<td>76</td>
<td>100</td>
</tr>
<tr>
<td>Head and neck (148)</td>
<td>52</td>
<td>18</td>
<td>18</td>
<td>14</td>
<td>3</td>
<td>85</td>
<td>78</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td>Brain (104)</td>
<td>50</td>
<td>12</td>
<td>20</td>
<td>14</td>
<td>3</td>
<td>50</td>
<td>71</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>Ovary (159)</td>
<td>50</td>
<td>9</td>
<td>17</td>
<td>17</td>
<td>6</td>
<td>86</td>
<td>74</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>Prostate (32)</td>
<td>50</td>
<td>9</td>
<td>17</td>
<td>17</td>
<td>6</td>
<td>86</td>
<td>74</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>Sarcoma (85)</td>
<td>48</td>
<td>14</td>
<td>26</td>
<td>7</td>
<td>1</td>
<td>67</td>
<td>73</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>Breast (223)</td>
<td>46</td>
<td>13</td>
<td>16</td>
<td>16</td>
<td>62</td>
<td>100</td>
<td>68</td>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>Bladder (90)</td>
<td>45</td>
<td>14</td>
<td>21</td>
<td>90</td>
<td>1</td>
<td>86</td>
<td>81</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>Endometrial (49)</td>
<td>43</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td>100</td>
<td>17</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>Hematological (236)</td>
<td>43</td>
<td>7</td>
<td>19</td>
<td>12</td>
<td>6</td>
<td>88</td>
<td>86</td>
<td>69</td>
<td>62</td>
</tr>
<tr>
<td>Skin (110)</td>
<td>40</td>
<td>12</td>
<td>22</td>
<td>5</td>
<td>2</td>
<td>77</td>
<td>42*</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>Gastric (121)</td>
<td>40</td>
<td>6</td>
<td>11</td>
<td>17</td>
<td>6</td>
<td>71</td>
<td>69</td>
<td>38</td>
<td>57</td>
</tr>
<tr>
<td>Thyroid (46)</td>
<td>39</td>
<td>9</td>
<td>17</td>
<td>17</td>
<td>4</td>
<td>100</td>
<td>38</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Colon (324)</td>
<td>36</td>
<td>9</td>
<td>15</td>
<td>11</td>
<td>2</td>
<td>68</td>
<td>64</td>
<td>66</td>
<td>50</td>
</tr>
</tbody>
</table>

*a number of mutations detected in benign and malignant tumors at each site. These numbers differ from values obtained from Table 2 by multiplying the total number tested × prevalence rates because: (a) some tumors contain multiple mutations; (b) mutations in nonmalignant precursors are included here but not in Table 2 prevalence data; and (c) only sequenced mutations are included here, whereas mutations determined by other techniques may be included in prevalence statistics.

*b Percentage of p53 mutations that are one of the indicated base substitutions associated with strand bias.

*c Percentage of p53 mutations of indicated base substitution on both strands.

*d Percentage of indicated base substitution on nontranscribed strand.

*e NA, not applicable.

* UV light-induced transitions at tandem pyrimidine sites are a cause of C→T transitions with nontranscribed strand bias in experimental models. Therefore, in skin cancer, C→T transitions are considered to be carcinoinduced nontranscribed strand mutations.

Substitutions exhibit marked strand bias. When strand bias is not apparent, it may reflect a mechanism of mutation which does not exhibit strand bias [such as non-bulky DNA adducts of alkylating agents, which cause mainly G:C→A:T transitions due to mispairing (31, 86)]. Alternatively, multiple carcinogens could be involved, and the bias patterns could negate each other (e.g., in esophageal and lung cancer, one tobacco-associated carcinogen might cause G→A and another, C→T transitions). In vitro studies which establish the mutational spectrum of a mutagen can be complementary to analyses of clinical specimens for strand bias in base substitutions and can implicate potential tumor-specific carcinogens (section II.D.5). In the lower half of Table 3 (i.e., tumors which exhibit few mutations with strand bias) are cancers which have not been strongly associated with environmental carcinogens (87), supporting the nontranscribed strand bias-exogenous carcinogen hypothesis.

2. Deletions and Insertions: DNA Replication and the Mispairing Hypothesis

Previous analyses of short deletions and insertions in p53 (88) and other genes associated with genetic disease (89) have concluded that the DNA sequence context is the most important factor determining these events. Almost all short deletions and insertions occur at monotonic runs of two or more identical bases or at repeats of 2- to 8-base pair DNA motifs, either in tandem or separated by a short intervening sequence. Several mechanisms are probably involved (90). The most well-studied is called slipped mispairing, a misalignment of the template DNA strands during replication that leads to either deletion, if the nucleotides excluded from pairing are on the template strand, or insertion, if they are on the primer strand. When direct repeat sequences mispair with a complementary motif nearby, the intervening oligonucleotide sequence may form a loop between the two repeat motifs and be deleted (88, 89). More lengthy runs and sequence repeats are more likely to generate frameshift mutations. In vitro research which detects errors in replication of reporter genes has helped quantify this phenomenon (reviewed in Ref. 17).

The database contains 285 deletions or insertions of ≤30 base pairs, including 98 single base pair deletions, 40 single base pair insertions, and 217 total events of ≤6 base pairs. Almost all (265; 90%) are at monotonic runs and short direct repeats, supporting the prior observations that mispairing is the primary mechanism of short deletions.
and insertions in human genes. Both deletions and insertions increase in frequency as the number of repeated bases in a run increases. The region of p53 which most frequently contains deletions or insertions is codons 151–159. Twenty-seven (9% of all deletions and insertions and 1% of all p53 mutations) have been reported at this G:C-rich sequence with multiple runs and direct repeats, confirming in vitro observations regarding sequence motifs which predispose to mispairing and other mechanisms of frameshift mutations.

Much remains unknown regarding the cellular phenomena which influence the risk of mispairing. Factors which affect DNA polymerase function, e.g., binding to polymerase accessory proteins, are areas of interest for mechanistic study (89). Some bulky chemical mutagens induce frameshift mutations in prokaryotic models through mechanisms related to adduct-induced deformation of the DNA helix. These adduct-induced mutations occur more frequently in monotonic or tandem base runs and at G:C-rich sequences (82, 90, 91). The contribution of chemical carcinogens to frameshift mutations of p53 in eukaryotic cells is potentially significant. Tumor-related differences in deletion prevalence may reflect organ-specific carcinogens or tissue-specific properties that regulate replication fidelity.

It is uncertain whether the recently described human mismatch repair genes hMSH2 and hMLH1 (25–28), which are responsible for causing mutations at tandem repeat DNA sequences (microsatellites) in families with HNPPC (section I.A; Refs. 92 and 93), play a role in p53 deletions and insertions. Most events in colon and other cancers occur at consecutive base runs, but 40% are at short tandem repeats. None of the seven deletions at the most mutable tandem repeat sequence in the p53 gene (GTG GTG GTG at codons 216–218) are found in colon cancers. Future studies might include evaluation of HNPPC tumors for slippage mutations at repeat sequences, assessment of the frequency of microsatellite changes in cancers which exhibit frequent deletions and insertions, and evaluation of mutations in hMLH1, hMSH2, and other mismatch repair genes.

C. Determination of Functions and Properties of p53

Mutational spectrum analysis complements in vitro studies of the relationship between p53 structure and function and suggests many experiments which can test the hypothesis that some mutants of p53 alter its function and confer upon cells a growth advantage. In other tumor suppressor genes (e.g., RB and APC), the most common abnormalities found are deletions or nonsense mutations which eliminate protein expression, indicating that loss of their function promotes tumorigenesis (94, 95). The presence of mutated p53 protein in tumors suggests that cells acquire a selective growth advantage from not only loss but also missense alteration of p53, frequently accompanied by loss or mutation of the second allele. Analysis of all p53 mutations in human cancers reveals a nonrandom distribution within the p53 molecule, implying that specific mutant proteins possess different growth stimulatory properties (Fig. 1). Identification of these locations has focused research on p53 structure and function toward specific regions of the protein, such as exons 5–8. What is the functional importance of these regions? Are mutations more frequent here because of the gain and loss of specific functions? Are the types of p53 mutations seen in other regions of the molecule different than those in the conserved regions?

1. Structure-Function Relationships of p53 Protein

Homologies have been noted between p53 domains and other proteins, providing clues to both structure and function. The conformations of wild-type and some mutant p53 proteins and their relationship to the critical functions of sequence-specific DNA binding and transactivation have been examined using full-length and truncated proteins. Some of these inferences have been corroborated by recent evidence describing the nuclear magnetic resonance analysis and the crystal structure of several p53 domains (96, 97).

A schematic of the p53 molecule, indicating the functional regions most frequently evaluated, is shown in Fig. 1. The NH2-terminal of p53 contains an acidic domain similar to those found in other transcription factors (5). Chimeric proteins containing amino acids 20–42 fused to the DNA binding site of Gal4 will transactivate reporter genes linked to Gal4-binding DNA sequences (98). Interestingly, this region also is responsible for binding to cellular protein Mdm2, and direct inhibition of transactivation is the likely mechanism by which Mdm2 interferes with p53 function (99–102). A proline-rich linking region between amino acids 60 and 92 begins the hydrophobic midsection of the protein. Between amino acids 100 and 295 lies the critical region (delineated by studies of truncated and chimeric proteins) which determines p53 conformation, specific binding to DNA consensus sequences, and sequence-specific transcriptional activity which may be essential for growth suppression (103). This region also binds to SV40 T-antigen (104). The recently described crystal structure of the DNA-binding domain identifies several motifs which directly contact DNA, including nine amino acids which abut the p53 consensus binding sequence (97). The positively charged carboxy terminus contains the basic nuclear localization sequence (amino acids 316–325) and oligomerization domain (amino acids 319–360; Refs. 5, 96, 105, and 106). Oligomerization into dimers or tetramers appears to be required for DNA binding and transactivation function (97, 107, 108). The observation that the nonsense mutation CCG→TGA at codon 342 has been reported 10 times in human tumors supports the experimental evidence that truncation of the last 50 amino acids does lead to a selection advantage in vivo.

Many mutations in the hydrophobic midregion change the conformation of the molecule, perhaps by partial denaturation. This exposes an epitope between amino acids 213–217, which is the target for mAb240, a monoclonal antibody that binds specifically to conformationally altered or denatured p53 (97). This antibody has been used as a marker of altered p53 conformation in immunoprecipitation and immunoblotting studies. The importance of conformation in p53 function has been demonstrated using truncated and mutant p53 molecules (including temperature-sensitive mutants whose conformation changes from wild-type at 32°C to mutant at 37°C) and by modulating wild-type conformation by oxidation. Molecules whose conformation is altered by mutation or oxidation (109) lose sequence-specific DNA binding and transactivation activity; temperature-sensitive mutants exhibit this loss only at 37°C, when they exist in mutant conformation (110, 111). Chimeric proteins, with the central p53 conformation region linked to foreign nuclear localization and transactivation regions, maintain sufficient p53 transactivating properties to suppress growth of lung and colon carcinoma cells (103, 112), indicating that the amino and carboxy termini contribute generic functions but not specificity.

Humans and laboratory animals which develop tumors carrying p53 mutations may also develop antibodies against the protein, and this phenomenon is now being investigated as a possible marker for diagnosis and prognosis. Anti-p53 antibodies occur in patients whose tumors contain missense mutations, but they are usually directed against epitopes in the amino and carboxy termini and not the conformation-determining midregion, suggesting that the native protein can be immunogenic. Antibodies have been detected in the serum of patients with carcinomas of the bladder (28%), lung (15%), breast (12%), esophagus (16%), liver (25%), pancreas (19%), and lymphomas (5%) and at least seven other types of cancer, as well as in patients with precursor lesions such as Barrett’s esophagus (5%; Refs. 113–118). In several anecdotal cases, the antibodies have been de-
ected in serum samples in high-risk patients (Barrett’s esophagus, tobacco smokers, and occupational vinyl chloride exposure) several years before the clinical detection of cancer.4 Hypotheses regarding anti-p53 antibodies include: (a) does the immune response affect carcinogenesis, e.g., does presence of anti-p53 antibody affect rates of progression to malignancy of preneoplastic lesions? (b) can the detection of anti-p53 antibodies be used as a screening test in patients at high risk for specific cancers (e.g., smokers, Barrett’s esophagus, and occupational carcinogens)? and (c) can the presence of serum anti-p53 antibodies identify individuals most responsive to immunotherapy?

2. Distribution of Mutations and the Relationship to p53 Conservation through Evolution

The clonal selection theory of carcinogenesis posits a growth advantage for clones containing mutant p53; therefore, sites where missense mutations frequently occur presumably identify amino acids whose replacement leads to loss of p53 suppressor function and perhaps gain of oncogenic function. Previous reviews have noted that major mutational “hotspots” occur in four of five domains labeled I–V (I, amino acids 13–19; II, amino acids 117–142; III, amino acids 171–181; IV, amino acids 234–258; and V, amino acids 270–286), which are highly conserved in evolution and are presumed to contain amino acids essential for p53 function (Fig. 1; Ref. 51). Of the 393 amino acids in human p53, 96 (24%) are conserved in 8 species for which the p53 sequence is reported. By comparing all hotspot sites to the pattern of AA conservation in all species, we can address the hypothesis that common mutations affect conserved amino acids important to p53 function.

Fig. 1 demonstrates the distribution of missense and nonmissense mutations in the database. Missense mutations are the most common type (79% of mutations) in the conserved midregion but are uncommon (23%) in the amino and carboxy termini, where nonmissense mutations (nonsense, frameshift, splicing, and silent mutations) predominate. The rarity of missense mutations in the amino and carboxy termini suggests that in these regions the integrity of a single amino acid residue is not essential to function or conformation. Of note, no mutations have been reported in conserved domain I. Nonconserved residues within exons 5–8 also show a low proportion of missense mutations, supporting the hypothesis that substitutions of nonconserved amino acids do not provide cells with a positive selection advantage. The most commonly mutated amino acids in the DNA-binding region are close to the protein-DNA interface (97). The distribution of deletions and insertions in p53 appears to be related primarily to DNA sequence repeats (Fig. 1; section II.B.2) and not to characteristics of the protein; as predicted, they occur frequently at both conserved and nonconserved codons.

We analyzed patterns of evolutionary conservation using the Multiple Alignment Construction & Analysis Workbench (MACAW) program (119). This method identifies the most highly conserved portion as amino acids 97–292 (MP-scores, 359.4 and 471.2 for two contiguous conserved regions). This area corresponds almost exactly to the region which determines sequence-specific DNA binding and transactivation (section II.C.1). When lower stringency for sequence similarity is allowed, secondary conserved blocks are noted between amino acids 17–29 (MP-score, 40.4), corresponding to the transactivation region identified through Gal4 binding experiments, and amino acids 324–352 (MP-score 114.3), encompassing the oligomerization domain. Higher stringencies fail to identify subregions within amino acids 97–292, including domains II–V, that are significantly more homologous than the sequence as a whole. Detailed analysis of the patterns of mutation clusters in conserved domains II–V and elsewhere adds to the evolutionary data and provides clues to amino acids of greater and lesser importance.

The predominance of missense over nonmissense mutations begins and ends abruptly at codons 130 and 286 (Fig. 1), suggesting that these are the boundaries of the conformation-determining region. This is shorter than the span identified by in vitro and evolutionary studies as important for DNA binding specificity (amino acids 93–293) and suggests that changes to amino acids 93–130 may not alter p53 conformation significantly. This region contains three short β-sheets and an unstructured loop, of which only one amino acid (Lys 120) interacts with DNA (97). Forty-four codons are relative hotspots for missense mutation, arbitrarily defined as more than 12 mutations reported in the database (>0.5% of all mutations); 40 of the 44 hotspot amino acids (91%) are conserved in all 8 species. This represents only 42% of the 96 conserved AAAs, suggesting that mutations at many other conserved sites confer little or no growth advantage; this hypothesis can be tested by functional assays of mutants created by site-directed mutagenesis.

Some amino acids in the previously identified hotspot domains are affected much more frequently than others, and within each domain are some highly conserved amino acids which are rarely mutated. For example, only 3 of the 25 amino acids in domain II (132 Lys, 135 Cys, and 141 Cys) are relative hotspots. All three of these residues begin or end β-strands, although other amino acids which flank β-strands are not hotspots (97). Of the six hotspot codons which contain 23% of reported p53 mutations (Fig. 1), five contain CpG dinucleotides (codons 175, 245, 248, 273, and 282 but not codon 249), suggesting a shared mechanism (deamination of 5-methylcytosine) for the prevalent G:C→A:T transitions reported at these sites. The resulting mutants at some of these codons have been shown to alter p53 transcriptional activity and growth suppression (section II.D.5.F). Only two of these amino acids (Arg 248 and 273) directly contact DNA, but all appear to be critical in stabilizing the protein-DNA interface (97).

Some relative hotspots for missense mutations occur outside the previously recognized conserved domains, marking three unrecognized regions of potential functional significance (Fig. 1). Between domains II and III lies a stretch of 14 amino acids, 151–164, of which one-half are conserved and at which 167 missense mutations (and 33 nonmissense mutations) have been reported. This region forms a β-strand which is not adjacent to DNA (97). Its DNA sequence is rich in G:C base pairs, often repetitive. No codons in this span are mutated as frequently as the six hotspots noted above. Despite the fact that five CpG sites occur here (section II.C.2), G:C→A:T transitions are rare. This suggests that these sites may not be methylated or that the AA substitutions resulting from deamination (three of which are silent) offer little growth advantage, perhaps because they are not adjacent to DNA (97). These hypotheses also are testable using in vitro mutagenesis, evaluating the mutants for transactivation, growth suppression, and other properties, as has been done for other CpG site mutations (section II.D.5.F).

Two other highly mutated areas lie between domains III and IV. Amino acids 193–200 are almost completely conserved; missense mutations are common at the first three codons (His-Leu-Ile). Codons 213–223 also define a region of high conservation, with missense mutational hotspots at codons 216 (Val) and 220 (Tyr). Two other conserved moieties which are frequently altered are 205 Tyr and 266 Gly. Of these mutational hotspots, five contain a single favored mutant: 157 G:C→T:A, Val→Phe; 194 G:C→T:A, Lys→Arg; 195 T:A→C,G, Ile→Thr; 205 A:G→G:C, Tyr→Cys; and 220 A:T→G:C, Tyr→Cys. Codon 157 mutations, especially G:C→T:A transversion, are especially frequent in lung cancer, perhaps indicating a tobacco-
specific carcinogen (section II.D.5.C) or a growth advantage peculiar to bronchial cells.

Taken together, these specificities do not exclude that certain regions of DNA may be susceptible to specific mutagenic events, with only favored mutants providing a selection advantage. The finding that tyrosine residues are affected at several hotspots suggests that they are important to p53 function. Potential reasons include contribution to p53 conformation or modulation of phosphorylation status, although a mechanism has not yet been demonstrated.

Some other specific residues with potential importance are serines 9, 15, 37, 315, and 392, targets for phosphorylation by DNA-activated protein kinase, p34\(^{\text{cdc}}\), kinase, and casein kinase II. *In vitro* mutagenesis studies creating alanine residues which are not phosphorylated have implicated serine 15 in human p53 as playing a role in suppressor function; conflicting results have been found by altering serine 392 (120–122). That no mutations of these amino acids have been reported is inconsistent with a strong selective advantage and argues against the hypothesis that these residues contribute significantly to suppressor function.

Our conclusions regarding evolutionary aspects of p53 are: (a) the most highly conserved region corresponds to the area of p53 responsible for sequence-specific DNA binding functions (amino acids 97–292); (b) the transactivation, nuclear localization, and oligomerization domains are only moderately conserved, provide generic functions which are not p53-specific, and are not significantly functionally altered by single amino acid substitutions; and (c) based on mutational spectrum analysis, essential conserved residues and favored mutants occur at previously described domains II–V and three additional regions (1) codons 151–164, especially codons 151–152 (Pro-Pro), 157–158 (Val-Arg), and 163 (Tyr); (2) codons 193–195, especially the codon 194 Lys>Arg and 195 Ile>Thr mutants; and (3) codons 213–220, especially the Tyr>Cys mutant at codon 220. Two other residues which are both conserved and frequently mutated are codon 205, another tyrosine which often mutates to a cysteine, and glycine at codon 266. A search of the SwissProt, PIR International, and translated Genbank files found no other reported sequences homologous to these three regions; therefore, their specific properties remain obscure.

D. Clues to Carcinogenesis

1. p53 in the Context of Multistep Carcinogenesis

Is alteration of function of the p53 pathway a prerequisite for neoplastic transformation? The frequency of p53 mutations argues for a critical role, but for most tumor types mutations have been found in only 20–50% of cases, with a maximum of 81% in oral mucosa squamous cell, 70% in small cell lung cancers, and 68% in anaplastic/undifferentiated thyroid cancer. Current techniques are unlikely to miss more than 15–20% of coding sequence abnormalities. How frequent and important to functional abrogation are mutations in the p53 promoter or unevaluated introns, epigenetic inactivation of p53, and/or alteration of genes or proteins downstream in the p53 pathway? Is p53 only one of several pathways whose disruption is sufficient for neoplastic transformation? Support for this concept comes from the observations that both p53 and ras mutations are common events but often occur independently (67, 123–127), and either may be associated with aggressive tumor behavior (4, 128, 129). Does neoplastic transformation require a critical mass of genetic injury, in which p53 is a frequent but not essential casualty? In this scenario, loss of p53 genomic stabilization properties would predispose a cell to an acceleration in the rate of genetic damage and greatly increase the likelihood of neoplastic transformation and/or malignant progression. Research which attempts to unify molecular and cellular theories of carcinogenesis should address these and other questions. Evaluation of mutation patterns of p53 and other genes in clinical tumors can be an important part of this process.

Some cancers are thought to arise from a single cell, while for others a “field defect” of the entire mucosal surface is thought to result in multiple clones of transformed cells and multiple primary neoplasms (130). Field carcinization is thought to occur in epithelia of the upper aerodigestive tract, bladder, and ulcerative colitis-associated colon (131–133). Careful identification of the stages of this process and molecular analysis at each stage can help illuminate the processes of clonal evolution in specific tumors. The multiplicity of p53 mutational events indicates that identical mutations are unlikely to occur independently, and specific mutants may be used as clonal markers. For example, the field carcinization hypothesis for head and neck squamous carcinoma is supported by molecular analyses which have found discordant mutations among nonneoplastic respiratory mucosa, primary cancers, and second upper aerodigestive primary lesions, suggesting that p53 is an important early target for mutations which can occur at multiple mucosal sites (132, 133).

2. Timing of p53 Mutations in Human Cancer

Several human cancers develop through multiple, morphologically defined stages, such as the adenoma–carcinoma sequence in the colon and the dysplasia progression in the oral and bronchial mucosae. The anatomically defined stages provide milestones for marking the timing of p53 mutation, and there is evidence that some tumor types undergo early and others late p53 mutation. The issue has both mechanistic and practical implications, including the use of p53 as a potential marker for early diagnosis and as an intermediate biomarker for chemoprevention studies. This section reviews current data assessing early and late alterations in human tumors. Early is defined here as more than one morphological stage before invasion of the basement membrane, which is the traditional hallmark of malignancy. Late means just before, or after, crossing the basement membrane. Alterations in the p53 pathway (mutation or epigenetic inactivation) are thought to be generally early events in cancers of the lung, esophagus, head and neck, breast, cervix, and stomach and generally late events in cancers of the brain, thyroid, liver, and ovary (134–139). Cancers of the colon, bladder, liver, and perhaps other organs may develop through multiple pathways; p53 alterations may be early events in some pathways and late events in others for these tumor types. Some of the better characterized examples are discussed here.

Squamous cell carcinomas account for 25% of lung cancers and the large majority of cancers of the upper aerodigestive tract. These tumors usually progress through multiple, histologically recognizable stages of dysplasia (usually subdivided as mild, moderate, and severe), carcinoma *in situ*, and transgression of the basement membrane, which marks the beginning of invasive cancer. Despite the small size of preneoplastic lesions, which frequently measure less than 100 \(\mu\)m, some genetic studies have been done, including allelic deletion, SSCP, and sequence analysis. However, the bulk of the data come from IHC using p53 protein accumulation as a surrogate marker for missense mutation (section II.A.1).

Genetic and cytogenetic analyses of preinvasive bronchial lesions have shown that p53 mutations can precede invasion. Three reports (140–142) document missense mutations, p53 allelic deletion, and chromosome 17p deletion in dysplastic mucosa associated with carcinoma. However, composite IHC data show p53 protein accumulation in 0% of normal bronchial mucosae, 7% of squamous metaplasias, 25% of mild dysplasias, 32% of moderate dysplasias, 69% of severe dysplasias, 57% of carcinomas *in situ*, 70% of microinvasive carcinomas, and 76% of fully invasive lung carcinomas (140–144). These results support a multistage model for squamous lung carcinoma in...
which p53 mutation occurs in 25% of the earliest neoplastic lesions, and the large majority of mutations occur before invasion develops.

The timing of p53 alterations in the oral cavity, oropharynx, and larynx is addressed by at least 10 published reports, usually using IHC as a marker for putative missense mutation in small progenitor lesions. Several groups have reported p53 alterations (positive IHC or mutation) at or before the stage of severe dysplasia, including some histologically normal mucosa (132, 145-149). Future studies could contribute to this area by providing additional histological detail to IHC studies, further molecular data, or correlating additional factors (e.g., DNA content and other genetic changes) with p53 alterations. The timing of p53 mutation in squamous neoplasms in the esophagus appears similar to those in the oropharynx (150-152).

Sporadic colorectal cancers generally develop through an adenoma-carcinoma sequence which has been extensively analyzed for alterations in oncogenes and tumor suppressor genes (reviewed in Ref. 153). Allelic deletion and p53 point mutation rarely occur until the late adenoma/carcinoma boundary (154; reviewed in Ref. 153). In contrast, in ulcerative colitis, an inflammatory condition predisposing to colorectal cancer in non-polypoid, non-adenomatous, dysplastic mucosa, p53 mutation and allelic deletion are found in the earliest recognized dysplastic lesions (131, 155, 156). It is notable that several IHC studies find protein accumulation at earlier stages of sporadic colon cancer, suggesting a nonmutational mechanism of protein accumulation during incipient neoplasia (157, 158).

3. p53 Genotype and Cancer Phenotype

The conventional method for describing the phenotype and predicting the clinical behavior of a cancer has been histological grading as assessed by light microscopy of tumor cells. IHC, which identifies specific cellular antigens, and other specialized staining techniques can yield more detailed phenotypes and sometimes clinically useful information, but these traditional measures are merely crude descriptions of essential tumor behaviors. Correlation of phenotypes, such as histological grade or proliferative rate, with precise molecular events, such as alterations of p53 or other genes (e.g., metastasis-related gene nm23, RB, and ras), promises to expand our understanding of tumor behavior and perhaps foster new paradigms of risk assessment, prevention, early detection, determination of prognosis, and antineoplastic therapies.

The relationship of p53 mutations to tumor grade and stage has been evaluated in many tumor types. Associations of p53 mutation or positive IHC staining with higher grade and more advanced stage are frequent, although sample sizes are often too small to draw statistically valid conclusions from individual studies. A relationship between p53 mutation and advanced tumor stage has been noted for cancers of the endometrium (159), cervix (159), ovary (138, 139, 160), liver (136, 137), prostate (161), and bladder (162, 163), indicating that for these tumors p53 mutation may be a late event contributing to tumor progression, but not for lung (67, 164, 165), head and neck (132, 149, 166), and breast (134, 135), in which mutations are known to occur in early stage tumors and precursor lesions.

Studies of brain neoplasms are archetypes for establishing the role of p53 in tumor progression and clonal dedifferentiation. Low grade gliomas, which are indolent and often curable by resection, may transform to high grade, aggressive, rapidly lethal glioblastomas. Two studies evaluating both low grade tumors and recurrent high grade tumors document the acquisition of p53 mutations in five of six tumors after transformation from low to high grade, although low grade tumors may also contain p53 mutations (167, 168). The p53 mutational spectra of low grade astrocytomas versus high grade glioblastomas are similar, supporting the concept that the high grade tumors evolve from the low grade precursors and are subject to the same mutations. As 45% of mutations are G:C→A:T transitions at CpG dinucleotides plus short deletions, the mutagens responsible for progressive genetic damage in these tumors may be endogenous cellular processes (section II.D.5.F).

Recent studies have correlated p53 mutation with metastasis and other important but less well-studied tumor phenotypes, such as angiogenesis (169). The regulation of specific genes which contribute to these behaviors [such as MDR-1 (170) and thrombospondin (171)] is one area in which the known transcriptional activating function of p53 could contribute to acquisition of the metastatic phenotype. Since p53 regulates genomic stability, and prevents cell cycle entry in response to DNA damage, loss of p53 function would be expected to correlate with aneuploidy and increased proliferative rates, two features often associated with tumor progression and poor prognosis. Preliminary studies have noted a statistically significant association between aneuploidy and p53 mutation for ovarian (172), colorectal (173), and gastric (54) cancers and a positive trend in esophageal cancer (53). Other properties which are being studied for a relationship to p53 function include radiosensitivity and hormone responsiveness (161, 174-176).

These data support a critical role for p53 in tumor progression. The results to date of p53 mutation analysis indicate that the details of this process are complex and probably mutant- and tissue-specific. Specific p53 functions potentially involved include the pathways of gene transactivation, cell cycle checkpoint control, and programmed cell death and suggest unbounded opportunities for in vitro testing. Future research might address p53 mutation and mechanisms of tumor invasion, metastatic properties such as angiogenesis, and drug resistance. Such studies could clarify questions generated by clinicopathological observations, which are often inconclusive due to unavoidable confounding variables or sample size limitations.

4. Immortalization and Neoplastic Transformation: In Vitro Features

Much in vitro carcinogenesis research uses cultured human cells derived from normal or malignant tissues (177). Tumor cell lines often differ from the corresponding in vivo normal tissues in properties such as protein or marker expression, response to growth modulators, and oncogene abnormalities. p53 mutations are found twice as frequently in all cell lines compared with all clinical tumor specimens (53% vs 26%); 15 of 20 tumor types have an excess of mutations in cell lines. Potential explanations for this observation include: (a) mutations might be underdetected in clinical specimens due to technical reasons; (b) mutations could arise frequently during the process of establishing a cell line and, therefore, reflect in vitro mutagenic events which are not relevant to in vivo carcinogenesis; (c) clones with mutated p53 which comprise a minority of the in situ tumor could have an in vitro growth advantage over clones without such mutations. According to this hypothesis, the mutations would reflect in vivo mutagens, but p53 mutation would not be an important early event. The genetic instability triggered by loss of p53 function would lead eventually to clonal evolution and tumor progression.

Mutational spectrum analysis supports the hypothesis that cell line mutations do reflect in vivo mutagens. Distribution of mutation sites and percentages of each type of base substitution, frame-shift mutations, and CpG site mutations are similar in clinical tumors and cell lines. Experience in the establishment of cell lines from cervical (178-180), thyroid (181, 182), nasopharyngeal (183, 184), and other cancers strongly suggests that cells with abnormal p53 function have a selective growth advantage in vitro. This advantage also may exist in athymic nude mice, as suggested by the higher prevalence of p53
5. Carcinogens and Specific Organ Carcinogenesis

The narrow mutational spectra exhibited by some mutagens has popularized the idea that each agent might leave a specific identifying “fingerprint” of site and type of DNA damage (35). It is probably more realistic to expect that carcinogens will produce mutation patterns which are characteristic and instructive but not as unique as fingerprints. Table 2 contains the mutational spectra of all cancers for which more than 20 mutations were reported at the time of analysis of the database, in descending order of mutation prevalence.

A) Skin Carcinoma and UV Light. A clear example of the value of mutational spectrum analysis in identifying carcinogen-specific mutations is seen in skin carcinomas, the most common type of human cancer, in which the role of UV light as the major carcinogen is unquestioned. Exposure to UV light increases the risk of both basal cell and squamous cell carcinoma, as well as the less common but more lethal melanoma (186, 187). This physical mutagen produces distinctive pyrimidine dimers that, if un repaired, can produce tandem mutations, most characteristically CC→TT transitions. Tandem dipyrimidine mutations are infrequently caused by mutagens other than UV light (188, 189) and have rarely been observed in noncutaneous malignancies, i.e., 10 in over 2400 tumors with p53 mutation. Thus, the observation that these tandem mutations are common in squamous cell (9%) and basal cell (12%) carcinomas of the skin (Fig. 2, B-D) directly incriminates both exposure to UV light as the cause of damage to the p53 gene and the loss of its tumor-suppressor function in the development of the cancers (190). The detection of rare cells with p53 mutations at dipyrimidine sites in sun-exposed nonmalignant skin indicates that UV light initiated these mutations at the earliest stage of skin carcinogenesis (191).

Mutations at dipyrimidine sites in skin carcinomas show a nonrandom distribution among sites within the p53 gene. Of the 30 CC:GG dinucleotides (18 on the nontranscribed strand) in conserved domains, transitions are common (5 or more) at only five (codons 151-152, 247-248, 178-179, 278, and 281-282); frequency of mutations also varies at other dipyrimidine sites (CT:GA, TT:AA, and TC:AG). Recent work shows that rates of cyclobutane dimer repair vary among codons within p53 (192). Comparison of the frequency of transition mutations to rates of repair shows that 94% of C-G→T:A or CC:GG→TT:AA transitions occur at sites where more than 10% of cyclobutane pyrimidine dimers remain 24 h after UV treatment, suggesting that slow repair of these dimers is responsible for some of the site specificity of UV-induced mutations. The dipyrimidine sites which are frequently mutated in XP-C patients were not evaluated in this study. It remains unknown whether variations in site-specific DNA repair in XP patients follow the same patterns as normal individuals and whether these patterns explain the different dipyrimidine mutation site specificity seen in XP-C. Mutant selection due to functional changes of specific amino acid substitutions also is likely to play a role in determining this specificity. To further define the mechanisms responsible for site specificity of mutations, study of the functional consequences (e.g., transactivation) of the commonly found mutants in skin and other cell types would be of interest.

Whereas p53 mutations are common in basal cell carcinoma and squamous cell carcinoma, they are infrequent in melanomas (Table 2). Since sunlight also is a risk factor for melanomas (reviewed in Ref. 193), the rarity of p53 mutations is surprising, although the frequent accumulation of p53 in the nuclei of these tumor cells (59, 186, 194) supports a role for the p53 biological pathway in tumorigenesis. Evaluation in melanomas of other genes and proteins which interact with and stabilize p53 (e.g., Mdm2) or act downstream in the p53 pathway [e.g., WAF1/Cip1 (63)] would be informative. The p16INK4 gene, which like WAF1/Cip1 encodes an inhibitor of cyclin-dependent kinases, is another candidate tumor suppressor gene involved in melanomas and other cancers (122, 195, 196).

B) Hepatocellular Carcinoma, Aflatoxin, and Hepatitis. The viral-chemical etiology and multistage pathogenesis of HCC are archetypal of human carcinogenesis. It is estimated that 75–90% of HCC cases are attributable to HBV (197). The relative risk of HCC is elevated in viral hepatitis carriers with chronic active hepatitis (198). These findings suggest that cell proliferation and/or inflammatory response associated with chronic active hepatitis are the critical factors responsible for the increased probability of neoplastic transformation of the precursor cells of HCC.

HBV DNA integrates into HCC cells at random sites in the genome. It contains the X gene, which codes for a protein (HBX) that modulates the transactivation of many cellular genes and is a candidate viral oncoprotein (199). Transgenic mice containing HBX gene in their germline have an increased frequency of HCC (200). HBX protein binds with p53 in vitro and in vivo (201), inhibits p53 sequence-specific DNA binding and transactivation activities, partially disrupts p53 oligomerization, and prevents p53 binding to transcription-repair coupling factor ERCC3 (202). These mechanisms of epigenetic p53 inactivation could account for the transactivation properties of HBX. Correlation of p53 staining and mutation data with HBX protein expression in HCCs could help assess the interactions between HBV and p53.

AFB also is considered to be a significant etiological factor in certain geographic areas (e.g., southern Africa and Asia), where this mycotoxin is consumed in food contaminated by Aspergillus flavus (reviewed in Ref. 203). Epidemiological studies conducted in the 1970s and 1980s provided statistical evidence that dietary consumption of AFB was positively correlated with incidence of HCC and suggested a synergistic interaction with alcoholic beverage consumption (204) and chronic active viral hepatitis (205–207).

Many investigators have reported data on the presence of chronic HBV infection and p53 mutations in the same tumors (208–216). Their results indicate that HBV infection alone does not influence the rate of p53 mutation and that AFB exposure is the most important influence on mutation prevalence (28 versus 29% mutation prevalence in HBV-positive versus HBV-negative cases in low AFB exposure areas; 54% prevalence in high AFB exposure areas, where almost all cases were HBV positive). The combination of these exposures may be synergistic in causing mutation, but data are insufficient for definitive conclusions (28 of 49 AFB positive/ HBV-positive cases versus 1 of 5 AFB positive/HBV-negative cases contain mutations; Refs. 208, 211, and 216).

The results from a recent prospective cohort study of 18,244 people provide convincing evidence that AFB has an etiological role in hepatocellular carcinogenesis and indicate a synergy between HBV and AFB (217, 218). This nested case-control analysis shows statistically significant associations among the presence of AFB and its metabolites in urine specimens, serum HBV surface antigen positivity, and HCC risk. The finding of the promutagenic AFB-N7-guanine adduct in the urine is further evidence that AFB has been activated to its electrophilic ultimate carcinogenic metabolite, AFB 8,9-oxide. Human hepatocytes in vitro can enzymatically activate AFB to its 8,9-oxide (219); the interindividual variation in forming the AFB-N7-guanine adduct may be 10-fold or greater (220). Since several isoforms of cytochrome P450 enzymes can activate AFB (221), the interpretation of pharmacokinetics data and its use in risk assessment
will be complicated. The mutational spectrum of AFB (222, 223) has been established in experimental systems; G:C→T:A transversions are the most common base substitutions (224–226). The high frequency of AGG→AGT transversions on the nontranscribed strand at p53 codon 249 in HCCs from areas of China and Mozambique with a high HCC incidence could be due to the high mutability of the third base of codon 249 by AFB and/or a selective growth advantage of hepatocyte clones carrying this specific 249ser mutant in liver chronically infected by HBV.

The preferential mutability hypothesis has been tested by Aguilar et al. (227), who found that in a human liver cell line exposed to AFB, the third base in codon 249 is preferentially but not exclusively mutated compared to immediately adjacent codons, suggesting that both preferential mutability and clonal selection are involved in human hepatocellular carcinogenesis. This highly sensitive and specific genotypic mutation assay also can be used to determine the p53 mutation load in nontumorous liver tissue from donors living in geographic areas of high HCC incidence in which urinary AFB and/or macromolecular-AFB adducts have been detected. A pilot study using a genotypic mutation assay has compared the load of codon 249ser (AGG→AGT) mutant cells in nontumorous liver tissues from HCC patients from Qidong, People’s Republic of China, urban Thailand, and the United States (228). The loads of codon 249ser mutations were elevated in specimens from Qidong, slightly elevated in one biopsy from Thailand, but did not exceed background levels in specimens from the United States. These results indicate a positive association between loads of codon 249ser mutation and dietary exposure to AFB. They also suggest that, in some populations, p53 mutation occurs early in hepatocarcinogenesis, and clonal expansion of these mutant cells occurs in the nonmalignant tissue. A number of hypotheses could account for a selection advantage for these mutant cells, including altered responses to growth factors or resistance to programmed cell death induced by hepatitis viral infection. These and other hypotheses can be tested by in vitro models using human hepatocytes (219) and clinicopathological studies, such as correlation of aflatoxin-DNA adducts with mutations in codon 249 G:C→T:A transversions in individual patients. The genotypic mutation assay also might be applied to other tumor-specific hotspots.

C) Tobacco: Lung and Other Cancers. Tobacco smoking is the leading preventable cause of cancer mortality, accounting for 40% of cancer deaths in men and 20% in women. Cigarette smoke is a complex mixture with thousands of constituents including tumor initiators, promoters, complete carcinogens, and cocarcinogens (229). The contributions of many of these agents to carcinogenesis in the lung and other organs are not known. Some of the compounds found in tobacco smoke, most notably N-nitrosamines and PAHs, have been purified, tested in model systems, and found to be in vitro carcinogens which induce characteristic DNA abnormalities (section I.B). This finding has encouraged mutational spectrum analysis in tobacco-associated cancers to test the hypothesis that patterns of mutation can implicate specific agents in tobacco responsible for inducing human cancers. Compounds of interest include N-nitrosamines, PAHs (e.g., BP), active oxygen species, aldehydes, and metals. The p53 mutational spectra in tobacco-associated cancers show surprising intertumor variation, suggesting that many different tobacco-related carcinogens may be germane to human cancers.

1) Lung Carcinoma. Cigarette smoking is now thought to be responsible for 90% of lung carcinomas in men and 78% in women (230). A large body of epidemiological evidence has confirmed a dose-response relationship proportional to duration and amount of smoking (231, 232). How does mutational spectrum analysis further define this association?

p53 mutations are common in lung cancer, with the highest prevalence (70%) in SCLC and the lowest (33%) in adenocarcinomas. The prevalent mutation is G:C→T:A transversion with a predominance of guanine residues on the nontranscribed DNA strand (Fig. 2, H-K), and the frequency of transitions at CpG sites (9%) is lower than in almost all other cancers. This spectrum is consistent with data for several different types of chemical carcinogens found in tobacco smoke. The highly mutagenic metabolites of PAHs such as BP preferentially attack deoxyguanines and lead to mutations (Table 1), and one of the quantitatively minor adducts from the tobacco-specific N-nitrosamine 4-(methylnitrosamo)1-(3-pyridyl)-1-butanone also leads to G:C→T:A transversions (233).

Analysis of p53 mutations in the database in relation to smoking history yields statistically significant conclusions consistent with the experimental data. The frequency of p53 mutation and of G:C→T:A transversion on the nontranscribed DNA strand are positively correlated with lifetime cigarette consumption (234). Compared with nonsmokers, smokers have significantly lower frequencies of all transitions (35 versus 69%; P < 0.05), G:C→A:T transitions (21 versus 69%; P < 0.001), and higher rates of G:C→T:A transversions with a DNA nontranscribed strand bias (29 versus 0%; P < 0.01; Fig. 2, L-M; Ref. 235). These observations are consistent with the model of preferential repair of carcinogen-induced damage on the transcribed DNA strand (75, 236). Cigarette smoke and some of its components have been shown in vitro (229, 237) to increase mutation frequency and cause defects in DNA repair, which may contribute to the p53 mutation spectrum. These results also have implications in the current controversy concerning the relative risk of lung cancer in nonsmokers exposed to environmental (passive) tobacco smoke. Molecular analysis would contribute important data to the debate by determining whether critical mutations in lung cancers from these individuals show a spectrum similar to smokers (a dose-response increase in G:C→T:A transversions with exposure to environmental tobacco smoke), although a large sample set may be required to detect changes in the spectrum.

Evidence exists of potential differences in pathways of carcinogenesis among histological types of lung cancer. In vitro studies of lung cancer differentiation support a pluripotent cell of origin common to all types, or multiple cells which can differentiate by variable, often overlapping pathways (238–240). Sufficient data exist to analyze the mutational spectra of squamous cell, large cell, adenocarcinoma, and SCLC (Fig. 2, H-K). p53 mutations are found in 70% of SCLC and 47% of NSCLC, including 65% of squamous and 60% of large cell, but only 33% of adenocarcinomas (140, 234, 241–246), a prevalence which more closely approximates that of other adenocarcinomas than of other lung cancer histologies. The mutational spectra are notable for an excess of A:T→G:C mutations in SCLC (11 versus 6%) and an excess of deletions and insertions in NSCLC (14 versus 4%). Subdividing NSCLC types reveals that the prevalence of G:C→T:A transversions in large cell, squamous, and small cell carcinomas is similar (43–49%), but they are less common in adenocarcinoma. This finding is consistent with the weaker association of adenocarcinoma with cigarette smoking (247–249), since G:C→T:A transversions are thought to result from bulky molecules found in tobacco smoke. In adenocarcinomas, higher proportions of G:C→A:T (non-CpG site) and G:C→C:G mutations are seen, suggesting the presence of other carcinogens. Since adenocarcinomas are the most common cell type in women and squamous cell is less frequent, analysis of mutational spectrum by gender (controlling for cell type and smoking history) may provide clues to different carcinogenesis pathways, hormonal influences, or differing susceptibilities to tobacco-associated carcinogens.

2) Other Tobacco-associated Cancers. Epidemiological studies have linked cigarette smoking to other cancers with attributable risks ranging from 20 to 90% (250). Mutational spectrum analysis of some of these tumors should reflect the effects of relevant tobacco-associ-
ated carcinogens. Despite the similarity of risk factors, the mutational spectra of these tumors differ from each other and differ significantly from the lung cancer spectrum. This may be due to different constituents of tobacco or nontobacco-related carcinogens.

Head and Neck and Esophageal Carcinomas: The combination of tobacco use and drinking alcoholic beverages is responsible for approximately 75% of esophageal and 90% of head and neck cancers, most of them squamous cell (250). Two lines of evidence derived from the database implicate environmental carcinogens in head and neck carcinogenesis: (a) over one-half of all mutations in head and neck cancers are substitutions at base pairs which usually demonstrate nontranscribed strand bias (Table 3; section II.B.1), and 80% of these occur on the nontranscribed strand; and (b) CpG transitions (suggestive of endogenous events; section II.D.5.F) are uncommon (13%). Compared with lung cancers, G:C→T:A transversions are less common (18 versus 40%), while A:T→G:C (14%) and non-CpG G:C→A:T (18%) transitions are prominent, as are deletions and insertions (19%). Of 18 deletions and insertions, 12 can be localized to repetitive DNA sequences, consistent with the hypothesis that mispairing during replication causes frameshifts (section II.B.2), and two are large deletions at splice site codons, suggesting that the actual DNA alteration is an intrinsic point mutation. Chemical carcinogens such as PAHs in tobacco smoke may contribute to these frameshift mutations (section II.B.2).

Subdivision of head and neck carcinomas by anatomic region of origin shows differing patterns of p53 mutations among the various primary mucosal sites (Fig. 2, E-G). In laryngeal and pharyngeal primaries, the mutation prevalence is relatively low (34%), but the spectrum resembles that of lung cancer. G:C→T:A transversions, primarily on the nontranscribed strand, are the most common mutation (34%), and non-CpG G:C→A:T transitions also are common. Primary tumors in the oral cavity have a high mutation prevalence (81%) and exhibit a spectrum in which non-CpG site G:C→A:T transitions, A:T→G:C transitions, deletions, and insertions predominate. Thus, in this class of cancers, subgroups which are often grouped together when addressing carcinogenesis and therapy demonstrate mutational differences, which could be due to differences in carcinogen exposure (e.g., soluble tobacco-specific N-nitrosamines in oral cancers versus tobacco combustion products in laryngeal primaries) or variation in carcinogen activation or DNA repair between distinct mucosal sites. These hypotheses can be tested in the laboratory and by molecular epidemiological studies of mutational spectra of well-characterized tumor types in populations with well-defined exposures.

The epidemiology of primary NPC differs from other head and neck sites; these tumors are associated with diet N-nitrosamines, EBV, and perhaps genetic factors (183). Clinical NPCs rarely contain mutated p53 (4%). One hypothesis for this low frequency of mutation could be the epigenetic inactivation of p53 by EBV proteins (251). The spectrum of the few p53 mutations in NPC tumors and lines is unusual, reflecting a lack of mutations known to be associated with tobacco-related carcinogens or N-nitrosamines in oral cancers versus tobacco combustion products in laryngeal primaries) or variation in carcinogen activation or DNA repair between distinct mucosal sites. These hypotheses can be tested in the laboratory and by molecular epidemiological studies of mutational spectra of well-characterized tumor types in populations with well-defined exposures.

The spectrum in esophageal squamous cell carcinomas also shows less frequent G:C→T:A transversions than in lung carcinomas (16 versus 40%), with corresponding increases in G:C→A:T (non-CpG site 22 versus 16%) and A:T→G:C (14 versus 7%) transitions, as well as A:T→T:A transversions (15 versus 5%). As discussed (section I.B; Table 1), PAHs, N-nitrosamines, and alkylating agents are among the compounds which cause these substitutions in experimental systems. Risk factors for squamous cell carcinoma in addition to tobacco smoking are alcoholic beverage consumption, and in China, dietary N-nitrosamine ingestion (which could account for non-CpG site G:C→A:T transitions; Ref. 252). Future studies could address the carcinogenicity and mutational spectra of these suspected carcinogens in cultured human esophageal and upper aerodigestive cells or in animal models.

As in the lung, adenocarcinoma of the esophagus is increasing in frequency relative to squamous cell; the last decade has seen a 3-fold increase in the incidence of adenocarcinoma and a 23% decrease in squamous cell carcinoma (253, 254). Adenocarcinoma of the esophagus resembles carcinoma of the gastric cardia and usually arises within the intestinal metaplasia of Barrett's esophagus. Chronic gastric reflux is considered a risk factor for this condition, but risk factors for neoplastic transformation are unknown. Few studies of the mutations in esophageal adenocarcinomas or Barrett's esophagus have been performed, but they suggest differences in mutations from squamous carcinoma. Of the mutations reported, 6 of 9 (67%) have been G:C→A:T transitions at CpG sites, compared with a 14% incidence of CpG site mutations in squamous cell esophageal cancer. Evaluation of more adenocarcinomas will be important to generate statistically valid comparisons with squamous carcinomas of the esophagus and with gastric cardia cancers.

Bladder Carcinoma: One-half of bladder carcinomas in men and one-third in women are attributable to smoking (255). A significant association has been found in superficial bladder cancers between p53 nuclear staining by IHC and the number of cigarettes smoked (256). The p53 mutational spectrum in bladder cancer is unusual with a high frequency of G:C→C:G transversions (21 versus 7% for all tumors) with a nontranscribed DNA strand bias. G:C→A:T transitions at both CpG and non-CpG sites are the most common mutation, with nontranscribed DNA strand bias at non-CpG sites. One study has found a hotspot G:C→C:G (Tyr>Lys) mutation at codon 280, but this has not been confirmed. Two reports have compared the spectra in a total of 34 smokers and 23 nonsmokers (163, 257) with similar spectra found in the two groups. Further epidemiological studies, including quantitative smoking histories, are warranted to evaluate the data for more subtle patterns.

D) Radiation. 222Radon is a colorless, odorless, inert gas, a naturally occurring decay product of 238uranium whose decay emits α-particle radiation. Conclusive epidemiological evidence implicates occupational exposure to radon in uranium miners as a cause of lung cancer (reviewed in Refs. 258 and 259). Early epidemiological studies of domestic radon and lung cancer risk exposure suggest that the dose-response and relative risk for residential radon exposure is the same as for occupational exposure (260–262). Radon exposure and cigarette smoking increase risk synergistically, and analysis of human tumors may reflect interactions between radon and tobacco carcinogens (258, 259, 262).

Two studies report the molecular analysis of the p53 gene in lung cancers from uranium miners (140, 263). Missense p53 mutations were found frequently in both studies. A mutational hotspot (16 of 30 mutations) at codon 249, AAG→ATG Arg→Met, was observed in miners from the Colorado Plateau who were exposed to high amounts of radon. Three of the codon 249mutations occurred in lung cancers from never smokers, implicating a non-tobacco-associated carcinogen. Since these miners also were heavily exposed to other carcinogens found in the mines, the hypothesis that radon was responsible for the specific base substitution requires further testing.

Other sources of ionizing radiation which have been associated with the development of hematological and solid malignancies include therapeutic radiation, atomic bomb explosions, and the radio-

4870

Downloaded from cancerres.aacrjournals.org on November 13, 2017. © 1994 American Association for Cancer Research.
logical contrast agent Thorotrast. Mutational spectrum analysis of cancers in these patients could indicate radiation-induced events. p53 mutations in lung cancers from nine nonsmoking atomic bomb survivors from Hiroshima, Japan, were compared to eight nonexposed controls; no differences in prevalence or mutational spectrum were noted in this small study (235). The study of larger numbers and a greater variety of radiation-related cancers would contribute more statistically meaningful data to the issue of radiation-induced mutagenesis in human cancer.

E) Cervical Carcinoma and Human Papilloma Virus. Epidemiological studies established the link between cervical carcinoma and HPV types 16 and 18, and the molecular basis for this association emerging from laboratory research features the p53 pathway. HPV types 16 and 18 both produce E6 protein, which complexes with cellular protein AP6 and degrades p53 in the cytoplasm (264, 265), an example of epigenetic p53 inactivation. Mutation data would support this hypothesis if p53 mutations were found in HPV-negative more than in HPV-positive cancers, and IHC would be expected to demonstrate no p53 staining in HPV-positive cases and nuclear staining in HPV-negative cases with missense mutations. Early reports found that mutation frequency was 100% in HPV-negative and 0% in HPV-positive tumors and cell lines (179, 266, 267), but this dramatic difference has not been validated in further studies of clinical tumors (159, 178, 268–271). These later studies confirmed the rarity of mutations in HPV-positive tumors (4%) and lines (7%); mutations have been found in 7 of 39 of HPV-negative carcinomas now reported (18%); \( P \approx 0.0007 \) versus HPV-positive tumors.

These data indicate that dysfunction of the p53 pathway contributes to cervical carcinogenesis, but unless mutations or other alterations in the p53 pathway are demonstrated in all HPV-negative tumors, it cannot be considered essential. Future studies might address the low rate of mutations in HPV-negative tumors by searching for other p53 pathway abnormalities, such as occult mutations, downstream gene abnormalities, or other epigenetic mechanisms of inactivation.

F) Endogenous Sources of Mutation. As discussed in section I.A, not all mutations are caused by exogenous carcinogens. Normal cellular events which are mutagenic include deamination of 5-methylcytosine at CpG dinucleotides, DNA polymerase infidelity due to mispairing at repeat sequences or other mechanisms, and oxygen radical damage (reviewed in section I.A. and Refs. 14 and 272). Studies of constitutive cellular genes to determine baseline mutation rates in human and other mammalian cells have found rates of between 10^{-10} and 10^{-12} mutations/gene/cell generation (10, 85). However, these may underestimate the true mutation rate because of insensitivity of the assays to some classes of mutations (e.g., only hemoglobin variants with point mutations were detectable) or failure of an assay’s cell culture conditions to select clones with mutations but no growth advantage (10, 85). This background level of mutations implies that populations will always exhibit a baseline rate of cancer, even in the absence of carcinogen exposure. The rates of endogenous mutations may also be affected by exogenous events; future studies on the effects of carcinogens and cocarcinogens on deamination and deletion are warranted. Mutational spectrum analysis may be able to define these baseline mutation frequencies by identifying the proportions of mutations in cancer-related genes which are caused by endogenous processes.

1) \( CpG \) Deamination. It cannot be proven that an individual G:C→A:T transition found at a CpG dinucleotide in a tumor results from deamination of 5-methylcytosine (as opposed to mispairing of an \( O^6 \)-methylguanine adduct with thymine, for example). However, the in vitro rates of spontaneous deamination, the clustering of spontaneous G:C→A:T mutations in experimental systems at sites where cytosine is methylated, and the large database of observed mutations suggests that this is the usual mechanism by which G:C→A:T transitions arise at CpG sites in human tumors. Examination of the p53 database for associated features such as amino acid selection and strand bias of CpG transitions may yield further insights into this process in tumor cells.

The p53 coding region contains 39 CpG dinucleotides (i.e., 78 base pairs) which are potential sites for G:C→A:T transition from deamination. Transitions at these sites account for 24% of all p53 mutations described, with large variations among tumors. Seven sites (codons 175, 196, 213, 245, 248, 273, and 282) account for 90% of CpG transitions, or 21% of all p53 mutations. These mutations occur equally on both DNA strands; strand bias in recognition and repair of G:T mismatches is thus unlikely to contribute to the selection of mutants.

Some mechanisms by which cells with some CpG transitions acquire a selective growth advantage have been studied. At codons 248 and 273, both C:G→T:A and G:C→A:T transitions commonly occur, implying that both amino acid changes (Arg>Trp and >Gln at 248, Arg>Cys and >His at 273) result in a selection advantage. At codons 175 and 282, only one of the two mutations is selected (175, Arg>His but not >Cys; 282, Arg>Trp but not >Gln). Soussi and colleagues have constructed all potential transition mutants at CpG hotspot codons 175, 248, and 273 and assessed conformation and in vitro transcriptional activation in the sarcoma cell line SAOS, which contains no endogenous p53 protein.\(^5\) Transactivation, which is thought to contribute to p53 tumor suppressor function, was assessed by transcriptional assays using the reporter gene CAT linked to an upstream p53 binding region. Wild-type and mutant p53 constructs cotransfected into target cells with the reporter constructs were compared in their transcriptional activating abilities. The favored Arg>His mutant at codon 175 assumes the mutant conformation and transactivation properties, whereas the rare Arg>Cys mutant retains wild-type conformation and activity. Interestingly, both mutants at codons 248 (Arg>Trp and >Gln) and 273 (Arg>Cys and >His) maintain apparent wild-type conformation by antibody analysis but acquire mutant transactivation function. These two amino acids are in direct contact with DNA (97). This evidence supports the studies (section II.C.1) which connect p53 transactivation activity with growth suppression, but they imply that conformational changes as detected by these methods are not necessary for modulation of this function (at least in SAOS sarcoma cells). Future studies could address other features of CpG site mutants which could eliminate p53 suppressor activity. It is possible that the more infrequent mutants are functionally silent, and their presence in cancer cells does not contribute to the cancer phenotype.

Thus, observational and experimental analysis supports the hypothesis that selection advantage due to altered transactivation and perhaps other functions is responsible for the hotspot mutations at CpG dinucleotides. Evidence argues against strand bias in mismatch repair of CpG transitions.

2) Cancers with High Rates of CpG Mutations, Deletions, and Insertions. The reasons discussed above argue strongly that deamination of 5-methylcytosine, an endogenous mechanism of mutation, is primarily responsible for G:C→A:T transitions at CpG dinucleotides (273). Similarly, deletions and insertions may result from misalignment due to polymerase infidelity during replication, another endogenous mechanism. Transitions and frameshifts also may be induced by carcinogen-DNA adducts and other mechanisms (section II.B.1.A). We hypothesize, however, that the sum of CpG transitions, deletions, and insertions, and potentially of other mutations which may eventually be linked specifically to endogenous processes, may be used as a

\(^5\) K. Ory and T. Soussi, personal communication.
Transitions at CpG sites, deletions, and insertions comprise 37% of all reported p53 mutations but ≥45% of mutations of brain, colon (Fig. 2W), stomach, endometrial, thyroid, and hematological cancers. This suggests that endogenous cellular processes may contribute to a significant fraction of these tumors. The mutational spectrum of LFS implicates 5-methylcytosine deamination as the primary event in the formation of p53 germline mutations; 53% of germline mutations are G:C→A:T transitions at CpG sites (Fig. 2O). The second most frequent class of mutations (13%) are deletions and insertions, mostly of a single base pair at repeat sequences, a pattern which is consistent with the slipped mispairing mechanism. Thus, two-thirds of LFS cases can be attributed to endogenous-type mutational events, compared with slightly more than one-third of p53 somatic mutations found in human cancers. Mutations known or suspected to be associated with carcinogen exposure (A:T→G:C and non-CpG site G:C→A:T transitions, G:C→T:A transversions, and tandem mutations) represent only 23% of LFS mutations versus 46% of somatic mutations (Fig. 2, A and O).

Comparison to germline mutations of other genes associated with cancer or other diseases supports the concept that these putative mechanisms of endogenous mutations are important sources of germ-line mutation, but mutation patterns are not consistent and probably reflect specific functional characteristics of the gene and protein. For example, germline mutations in the APC gene cause familial adenomatous polyposis, a syndrome in which multiple colon polyps, adenomas, and eventually carcinomas develop in all affected persons. Germline and somatic mutations in APC result in truncation or complete loss of the protein in 96% of cases and in an altered protein in only 4% (274). Thus, frameshift mutations are prominent (48% of mutations), and CpG site G:C→A:T transitions, although the most frequent base substitution, comprise only 14% of all mutations. Similar to p53 germline mutations, however, three-quarters of APC germline mutations are consistent with the endogenous processes of deamination and slipped mispairing. The spectrum of mutations in the factor IX gene, which causes hemophilia B, contains similarities and differences from the spectra in the tumor suppressor genes (273, 275). Approximately two-thirds are missense substitutions, less than 10% are deletions and insertions, and the remaining one-quarter are splice site alterations and promoter mutations. Transitions at CpG sites are common (approximately 36%), but other base substitutions are more prevalent in factor IX mutations than in p53 germline mutations. Expansion of the data sets on germline mutations in other genes such as HPRT, APC, DCC, hMLH1, and hMSH2 may suggest patterns which are now inapparent.

3) Mutational Spectrum Differences among Populations. Although the focus of mutational spectrum analysis in HCC has been on the role of codon 249 mutations and AFB exposure, deeper scrutiny of the data produces additional hypotheses. Since the spectrum in HCC is particularly dependent on a population's environmental exposure, subclassification of the data by geographic region may reveal patterns hidden by the AFB-related tumors. Fig. 3 depicts the mutational spectra of tumors from Qidong, other regions of mainland China, Taiwan, and Japan. While tumors from Qidong contain an overwhelming preponderance of G:C→T:A transversions (95%) due to association with AFB, Japanese and Taiwanese tumors harbor this type in only 26% of cases each. That both of these latter groups exhibit some codon 249 mutations suggests that AFB may be a risk factor even in populations not normally considered to be at high risk for dietary contamination. The remainder of the spectrum differs between the groups, however, and suggests other potential carcinogens. In Taiwanese patients, the most frequent base change for all HCCs is A:T→G:C (41%); the strand bias (all 11 mutations are A→T) suggests that an exogenous carcinogen is responsible. For Japanese patients, G:C→A:T (at CpG and non-CpG sites) and A:T→G:C transitions (also with nontranscribed strand bias) are both common, suggesting environmental carcinogens; deletions and insertions, linked to both endogenous and exogenous mutational mechanisms, are also frequent. The identification of relevant carcinogens with these spectra will be important. Only nine mutations have been reported in HCC from Western countries; five are G:C→A:T transitions, three of which are at CpG sites. Thus, no pattern suggesting carcinogen exposure is yet discernible.

Environmental exposures or inherited characteristics which vary among populations may be important determinants of mutational spectra in other tumors as well. Sufficient data are available to evaluate lung, gastric, and breast cancers for differences in spectrum between Japanese and Western patients. Compared to Western patients, Japanese lung tumors show a higher frequency of G:C→A:T transitions (29 versus 20%, P = 0.13; most are at non-CpG sites) and a lower frequency of G:C→T:A transversions (30 versus 43%, P = 0.01), both with a nontranscribed strand bias. These differences
are not found in SCLC but are seen in both adenocarcinomas (G:C→A:T transitions: 48 versus 20%, P = 0.02; G:C→T:A transversions: 14 versus 27%, P = 0.21) and non-adenos NSCLC (24 versus 20%, P = 0.004; 29 versus 55%, P = 0.0003, respectively). Deletions and insertions are more frequent in Japanese tumors (18 versus 10%, P = 0.22). These results suggest differences between Western and Japanese individuals, either in exogenous carcinogens or in cancer susceptibility factors such as germline polymorphisms in genes involved in carcinogen metabolism (49).

Most of the gastric tumors analyzed have been from Japan, where incidence rates are much higher than in Western countries. Epidemiological studies in Japan have found an association between gastric cancer and dietary salt and N-nitrosamines (276). Comparison of the mutational spectra reveals that endogenous-type mutations are seen in 72% of Western and 55% of Japanese tumors. More Western tumors must be analyzed to provide a more statistically valid spectrum. The major type of p53 mutation seen in Japanese but not Western tumors is A:T→G:C transition (20 of 69 versus 1 of 18, P = 0.03), with strand bias favoring T→C on the nontranscribed strand, the opposite of the usual A:T→G:C strand bias. Perhaps this represents the consequences of alkylthyme adducts from N-nitrosamines (277-279) in the Japanese diet.

No specific exogenous carcinogens have been conclusively linked with development of human breast carcinoma. Although only 25% of breast cancers contain p53 mutations, 15% of these are G:C→T:A transversions, 16% are G:C→A:T transitions at non-CpG sites, and 11% are A:T→G:C transitions, with nontranscribed strand bias of 89, 63, and 75%, respectively, suggesting that exogenous bulky carcinogens may play a role. The responsible carcinogens may vary among populations. Two reports of breast cancers from Japanese women (135, 280) indicate a mutational spectrum different from tumors in Western women. G:C→T:A transversions are uncommon in Japanese cases (1 of 48 versus 25 of 175, P = 0.02), while A:T→G:C transitions are more common (8 of 48 versus 10 of 175, P = 0.01). Thus, A:T→G:C transitions with nontranscribed strand bias are prominent in the spectra of three different types of tumors in Japanese patients (HCC, gastric, and breast). Whether this represents the effect of a regional environmental agent(s) or population differences in carcinogen metabolism or DNA repair is an area for future investigation. Transitions at CpG sites, deletions, and insertions are common in both populations, suggesting a similar contribution of endogenous mutations.

These analyses suggest that molecular studies with careful identification of geographic and ethnic features of populations can make a significant contribution to carcinogenesis research. Future research should evaluate statistically valid sample sizes in order to address specific hypotheses of differences in carcinogenesis among populations. For example, studies of Japanese populations in different geographic regions (Japan, Hawaii, and mainland United States) have demonstrated differences in cancer rates. Molecular studies on these cancers should be done to search for changes in mutational spectra, which would suggest that environmental carcinogens differ among geographic regions for several tumor types.

4) p53 Mutation and Prognosis. Since tumors with p53 mutations often behave more aggressively, the value of p53 as a prognostic factor has been addressed. Studies have implicated p53 protein expression as an independent prognostic factor in carcinomas of the breast, stomach, colon/rectum, bladder, and NSCLC (reviewed in Ref. 129).

Future studies might integrate hypotheses of carcinogenesis into studies evaluating clinical outcomes. Since p53 mutation may correlate with other indicators of poor prognosis, such as aneuploidy, gene amplification, and high proliferative fraction (section II.D.3), it will be important to adhere to careful epidemiological methodologies to assure that p53 status adds information to established prognostic markers. This can be accomplished by performing prospective studies in conjunction with large clinical trials to ensure statistical power, facilitate multivariate analyses to control for other established prognostic variables, and assess the responses to therapies in relation to p53 status. The ability to perform IHC and PCR-based diagnostic studies from archived tissue means that some of these studies could be accomplished on materials from completed clinical trials which may already have reached their end points, expediting the answers to some questions. Comparison of multiple molecular markers in the same tumors can help to determine the relative importance to carcinogenesis of abnormalities in various cellular pathways.

Such studies can address both specific and general hypotheses regarding carcinogenesis, for example: that tumors with p53 mutations might be more resistant to therapy-induced apoptosis; that tumors with mutations are more likely to be associated with amplification of other cancer-related genes which influence therapy; whether all cancers have abnormalities of the p53 pathway, or of a specific combination of pathways; the relationship of response to therapies to molecularly determined tumor phenotypes and genotypes; identification of subsets of patients with favorable and unfavorable prognoses that should be analyzed separately and excluded from standard treatment protocols; and whether innovative therapies (e.g., antisense oligonucleotides and gene therapy) can be tailored to specific molecular features of a tumor to improve the probability of successful outcome.

III. Conclusions

p53 mutational spectrum analysis contributes to a variety of mechanistic questions in carcinogenesis, including integration of molecular and cellular hypotheses. Examples which are of ongoing interest include: what is the timing and sequence of the genetic changes contributing to neoplastic transformation, progression, and metastasis? Does cancer always arise from a single transformed clone, or can multiple transforming events occur in some tissues? How do carcinogen-host interactions, such as individual differences in carcinogen metabolism and DNA repair, influence an individual’s risk of developing cancer? Do different histological types of cancer contain characteristic mutations reflecting a pattern of mutagen exposure, and do cancers of similar histological types in different organs share these patterns? Are abnormalities of the p53 pathway critical for all tumors? In cancers with low rates of p53 mutation, is p53 inactivated by epigenetic events, or is its function interrupted by changes in other genes which participate in p53-mediated pathways?

The emerging field of molecular epidemiology promises important contributions to cancer research, but full realization of its potential will require careful adherence to the tenets of both laboratory science and epidemiology. This will require increased interaction between laboratory researchers, clinicians, and epidemiologists in order to design studies which can adequately address important hypotheses. For example, published reports of the molecular features of cancers (and other diseases) should include important epidemiological data, such as gender, age, ethnicity, geographic origin, and, when appropriate, associated medical conditions, tobacco smoking, and occupational or other exposures. Large scale clinical trials which include data on the relevant variables can provide tumor tissue for molecular study. This may be accomplished using archived materials from past studies; future trials could collect and preserve tissue specifically for molecular analyses. Attention to specimen handling at all times would minimize tissue degradation and maximize the reliability of laboratory findings.

Mutational spectrum analysis and molecular epidemiology can be applied in innovative ways to both generate and address basic and practical hypotheses including the areas of chemical, physical, and bio-

4873
logical carcinogens, risk assessment, host factors, control of cell and tissue behaviors, mechanisms of subcellular communication, protein function, DNA replication, transcription, translation, and evolution.

Acknowledgments

We appreciate discussion and comments from our colleagues, including Bettie Appella, Mark Boguski, Kathleen Forrester, Stephen Friend, S. Perwez Hussain, David Lane, W. Edward Mercer, Ruggero Montesano, Moshe Oren, Jennifer Pietinen, Varda Rotter, Thierry Soussi, Glen Trivers, Bert Vogelstein, Xin Wang, and Stuart Yuspa. The expert graphic and editorial assistance of Dorothea Dudek and the assistance of Peter Shields with statistical analyses are also appreciated.

References


p53 MUTATIONS AND CANCER


p53 MUTATIONS AND CANCER


Downloaded from cancerres.aacrjournals.org on November 13, 2017. © 1994 American Association for Cancer Research.
Mutations in the \textit{p53} Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular Pathogenesis

M. S. Greenblatt, W. P. Bennett, M. Hollstein, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/54/18/4855.citation

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, use this link
http://cancerres.aacrjournals.org/content/54/18/4855.citation.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.