[CANCER RESEARCH 51, 3617–3620, July 1, 1991]

Advances in Brief

From Gene to Carcinogen: A Rapidly Evolving Field in Molecular Epidemiology

Peter A. Jones,[1] Jonathan D. Buckley, Brian E. Henderson, Ronald K. Ross, and Malcolm C. Pike

Kenneth Norris Jr. Comprehensive Cancer Center, University of Southern California, Los Angeles, California 90033

Abstract

Chemical and physical carcinogens leave footprints of their activities on DNA because of the patterns of base changes they induce. Additionally, the conversion of 5-methylcytosine to thymine in CpG sequences leads to a characteristic mutation which can be used to estimate the contribution of endogenous processes to human mutations. Knowledge of the pattern mutations found in genes commonly mutated in human cancer, such as the *p53* tumor suppressor gene, allows for predictions to be made on the likelihood of an exogenous DNA-damaging agent being involved. Working from gene to carcinogen is likely to have a profound impact on our understanding of the origins of human cancer.

The discovery by Percival Pott of the high risk of scrotal cancer in chimney sweeps was the first demonstration of an association between a human cancer and an environmental chemical exposure. The epidemiological approach has since been successful in identifying a wide range of human carcinogens. Extraction, purification, and identification of pure chemical constituents from environmental carcinogenic mixtures and the demonstration that these molecules could induce carcinogenic changes in appropriate experimental systems has been a major achievement of the last 60 years. In all of these cases, epidemiologists have elucidated particular patterns of tumor distribution and exposure to various agents, and laboratory workers have frequently been able to use this information to identify the precise causative agent.

It is now axiomatic that many or even most "carcinogens" induce carcinogenic changes as a result of direct interaction with DNA, frequently after some chemical transformation of the carcinogen. A considerable amount is known about the interaction of ultimate chemical carcinogens and ionizing radiation with specific bases and DNA sequences and, in particular, the end results of these interactions in terms of permanent DNA changes recognized as mutations. For example, benzo(*a*)pyrene exposure often results in transversions (defined as changes of a purine to a pyrimidine or *vice versa*) of G to T (1), and melphalan induces predominantly A to T transversions (2), whereas the binding of alkylating agents to the O^6 position of guanine causes an alteration of hydrogen-binding properties and causes predominantly transitions (defined as changes of a purine to a purine or a pyrimidine to a pyrimidine) of G to A (3). Ionizing radiation on the other hand often causes deletions of DNA sequences.

The relevance of such observations has been demonstrated in animal carcinogenesis experiments in which mutations induced *in vivo* are precisely those predicted for the carcinogens used. For example, mammary carcinomas induced by nitrosomethylurea in rats contain G to A transitions (4), whereas skin tumors induced by dimethylbenzanthracene in mice contain A to T transversions in their Ha-*ras* genes (5); the type of mutation was as expected in both experiments. A similar correspondence between the type of mutation seen in a tumor and that expected based on laboratory studies of the putative carcinogen has recently been described in humans. Twelve of 13 point mutations reported in hepatocellular carcinomas from patients living in regions where aflatoxins are known risk factors for this disease were substitutions of T for G (6, 7). The importance of this observation is apparent when it is noted that the effect of aflatoxin B_1 in experimental systems is to induce G to T transversions (8).

These results suggest that knowledge of the site and nature of DNA changes in particular tumors should be useful in eliminating certain agents as major "causes" of the tumor and may direct attention to the classes of chemicals, or even to specific chemicals, whose effects are consistent with the mutations actually observed. In this paper we suggest, in particular, that one can probably distinguish between mutations caused by direct-acting carcinogens and tumors caused by "spontaneous" mutations by noting the frequency of one particular type of mutation, that arising from CpG dinucleotides, in the tumors. This distinction is clearly of the utmost importance since it directs the search for the causal agent either to a direct-acting mutagen or to a "promoter" such as an agent which led to increased cell division (9).

p53 as a Common Target in Human Carcinogenesis

The *p53* gene codes for a protein which appears to function as a cell cycle-regulatory molecule. It is located on chromosome 17p, a region often reduced to homozygosity in common cancers of Western societies, and is the most frequently altered gene in human cancers (10). The fact that as many as half of these common cancers contain *p53* mutations (10) supports a causative role for them in tumorigenesis. Direct evidence for the regulatory function of *p53* comes from experiments showing that wild-type *p53* can suppress the growth of colorectal carcinoma cell lines containing mutant alleles (11). A central role a *p53* in human cancers is therefore well established.

An important feature of the *p53* mutations is that they are scattered over a wide area of the gene and encompass several kinds of damage, including transitions, transversions, and deletions. Apparently there are many ways by which the function of the gene can be altered, reflecting the relative ease with which gene function can be destroyed through mutation. In contrast, activation of a protooncogene may require more specific changes that confer new properties to the gene product; activating *ras* gene mutations, for example, tend to cluster tightly in three places (codons 12, 13, and 61; Ref. 12). Thus while the mutation pattern seen in either a tumor suppressor gene or a protooncogene can be useful in identifying possible carcinogens, the former may be more informative since a wider range of changes is possible.

It is worth noting that in at least one tumor *p53* appears to be acting more like a protooncogene than a suppressor gene. It

Received 5/2/91; accepted 5/16/91.

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] To whom requests for reprints should be addressed, at Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, United Kingdom.

has recently been reported (6, 7) that 11 of 13 mutations in *p53* in hepatocellular carcinomas in patients from Africa and China have resulted in an arginine to serine substitution in codon 249 of *p53*. Furthermore, 12 of 13 point mutations found in these patients were G to T transversions. Clearly the specificity of these mutations indicates that the resultant *p53* protein has not simply been inactivated but has apparently conferred a selective advantage on the cell. One suggestion is that since hepatitis B is a risk factor for hepatocellular carcinomas in Africa and China, the specific mutant *p53* may interact with a hepatitis B protein to provide a growth advantage in hepatomas (13). The most likely cause of the mutations is exposure to aflatoxin B_1, a food contaminant in both Africa and China, which is a known risk factor for hepatocellular carcinomas. Aflatoxin B_1 binds preferentially to G residues in GC-rich regions and induces G to T transversions almost exclusively.

The wide scale of involvement of *p53* in human cancer, together with the broad spectrum of observed mutations, make it an attractive gene for molecular epidemiology studies.

A "Baseline" Mutation Pattern

Comparison of mutational spectra between genes is complicated by several parameters, including codon usage, differential susceptibilities of particular sites to alteration, and ascertainment bias because of severity of disease. Nevertheless, analysis of hundreds of mutations in several genes has led to some important generalizations regarding mutation frequencies in humans. These are: (*a*) the dinucleotide CpG, although underrepresented in DNA, is a hot spot and accounts for fully 30–40% of all human germ-line mutations (14); and (*b*) the most common mutations are transitions followed by transversions and deletions (15, 16).

The role of CpG in human mutation is thought to result from the frequent methylation of the cytosine residue at such sites. The resulting 5-methylcytosine residues can deaminate spontaneously to thymine resulting in G:T mismatches which are not always repaired accurately and give diagnostic C to T transitions at CpG sites (17). Since both cytosines in the CpG palindromic sequence in double-stranded DNA are usually methylated, there is also the possibility of corresponding G to A transitions occurring at CpGs if the 5-methylcytosine residue in the noncoding sequence changes to a T. Support for this mechanism has come from recent experiments demonstrating that a CpG in the low-density lipoprotein receptor gene known to have undergone a transition to TpG is indeed methylated in human sperm (18). Transitions from CpG to TpG or CpA appear therefore to be characteristic of "DNA-methylation-induced" mutations.

The methylation of cytosine residues in CpG sites is clearly an endogenous process, and no exogenous factor has yet been found that alters the frequency of deamination of methylated cytosine in CpG or the efficiency of repair of such mutations. It is thus reasonable to hypothesize that the rate of such mutations may be essentially constant, be due simply to the very nature of the methylated CpG nucleotide, and be legitimately termed "spontaneous." The frequency of transitions at CpG locations thus appears to give a measure of the rate of endogenous mutations in a particular gene. A lessening of the proportion of mutations occurring at CpGs (specifically, CpG to TpG or CpA transitions) would be indicative that a chemical or physical agent was having a direct effect on DNA. The strong potential for 5-methylcytosine to act as an endogenous mutagen also means that transitions occurring at CpG should be considered separately from those occurring at non-CpG sites. CpG mutations are thus a highly significant contributor to point mutations in human genes and can be used, it appears, to give an estimate of the rate of spontaneous *versus* induced base alterations.

An examination of patterns of mutations in the gene responsible for Factor IX deficiency hemophilia provides support for this concept of a baseline spontaneous pattern of mutation (19, 20). The proportions of each type of mutation are remarkably constant in different populations studied (15). Since these populations are likely to be subject to quite different environmental exposures, this constancy suggests that the pattern seen is independent of external factors and represents a baseline spectrum of mutations. Analysis of the Factor IX mutations clearly shows the prominent role of CpG in causing this disease (Table 1). The remaining mutations are relatively evenly distributed between G:C and A:T base pairs with a slight preference for transitions at A:T base pairs. Fully 72% of mutations are base transitions with transversions responsible for 20% of alterations and 8% being caused by insertions or deletions. The non-CpG mutations in the baseline spectrum may reflect the intrinsic DNA replication error rate and/or the effects of background radiation exposure.

Analysis of *p53* Mutations in Human Cancers

Table 1 summarizes the mutations which have recently been reported in the *p53* gene in uncultured human tumors or human tumor cell lines. The spectrum of mutations found in the Factor IX gene in patients with hemophilia B is included for comparison.

The mutations found in the *p53* gene in human colon cancers, leukemias, and sarcomas are similar to each other and, like the Factor IX gene, include a high proportion of point mutations at CpG residues. Although it is difficult to compare directly the patterns between two different genes (*p53* and the Factor IX

Table 1 *Distribution of site and type of mutations in the p53 gene, by tumor type*

The distribution of mutations in the Factor IX gene of hemophilia patients is shown for comparison.

Gene (*n*)	Tissue	Mutations at indicated base: mC[a] (%)	G or C[b] (%)	A or T[b] (%)	Deletions (%)	Transitions (%)	Transversions (%)	Ref.
Factor IX (194)	Germ-line	36	28	28	8	72	20	19
p53 (40)	Colon cancer	63	20	18	0	95	5	26–29
p53 (23)	Leukemia and sarcoma	48	30	22	0	78	22	30–32
p53 (12)	Bladder cancer	33	50	17	0	83	17	22
p53 (24)	Lung cancer	13	75	4	8	25	67	16

[a] Mutations at 5-methylcytosine are scored as transitions occurring at CpG which are consistent with deamination-induced endogenous mutations (*i.e.*, CpG to TpG or CpA). It is assumed that all CpGs in the *p53* gene are methylated, although this has been demonstrated only for a subset of CpGs (14).

[b] Mutations at C:G or A:T base pairs are scored as indicated because it is not possible to determine which base in the pair sustained a mutational event. Those involving C:G pairs have been scored after removal of mutations occurring at CpG.

gene) for the reasons enumerated earlier, several trends are apparent. The prominent role of CpG is clear, as is the prevalence of transitions over transversions. The presence of 5-methylcytosine at several of the CpG sites known to have undergone mutation in these tumors has recently been demonstrated directly by genomic sequencing (18). These data are therefore consistent with the idea that 5-methylcytosine, although underrepresented in the *p53* gene, plays a dominant role in inducing mutations in these tumor types. The data also strongly suggest that the mechanisms for the induction of mutations in this tumor suppressor gene are similar in colon cancer, leukemias, and sarcomas and may not involve the direct interaction of an exogenous agent with DNA in the majority of instances. The proportions of mutations at CpG, G or C, and A or T (Table 1) were not significantly different for colon cancers and leukemia and sarcoma ($P = 0.49$; exact contingency table analysis).

The data for small-cell lung carcinoma are dramatically different, as has been pointed out previously (16, 21). The role of 5-methylcytosine is reduced considerably, and mutations at G and C residues at non-CpG sites now predominate and account for 75% of mutations. This is also reflected in the fact that transversions account for 67% of mutations in lung cancer but only 5–22% in colon cancer, leukemias, and sarcomas. The distribution of bases affected in lung tumors (the first three percentages in Table 1) was significantly different from that seen in colon, leukemia, and sarcoma ($P = 0.00001$; exact analysis). These data are therefore consistent with the *p53* mutations in lung cancer having been induced by interaction of a carcinogen, presumably cigarette smoke, directly with DNA (21).

The results for bladder cancer, although based on a limited number of cases, suggest a pattern that is intermediate between that seen for colon cancer and lung cancer (22). This is particularly interesting since bladder cancer, like lung cancer, has been etiologically associated with smoking, but the attributable risk for smoking in bladder cancer, the proportion of cases that are thought to be due to exposure to cigarette smoke, is estimated to be 50% (23). Due to the small sample, the distribution of bases affected for bladder tumors did not differ significantly from either the lung cancers ($P = 0.14$) or the combined colon, leukemia, and sarcoma data ($P = 0.17$).

Another feature of these comparisons, not evident from Table 1, is the information which can be gained from mutations at a single site within the *p53* gene. For example, codon 273 (CGT) is a hot spot for mutation in both lung cancer and tumors of other sites, suggesting that the arginine residue in this position is essential for proper *p53* function. All four reported mutations at this site in colon, brain, and breast cancers are consistent with 5-methylcytosine-induced transitions (*i.e.*, transitions from CGT to CAT or TGT). In contrast, only one in four mutations at this site in lung cancer was induced by this mechanism, and the remaining three involve transversions of the G to a C or T which are clearly not the result of the spontaneous deamination of 5-methylcytosine. The difference in mutational spectra in the *p53* gene between lung and other tumor types is therefore evident even at the level of a single codon.

The *p53* gene has also been implicated in the Li-Fraumeni syndrome, and 2 of 6 germ-line mutations reported in *p53* in this syndrome occurred at 5-methylcytosine residues (24, 25). These data are consistent with the "spontaneous" pattern of mutations seen in the Factor IX gene and the non-smoking-related malignancies, but clearly a much larger series of families will need to be studied to establish the distribution with any certainty.

Summary

The finding that mutations in the *p53* gene are a common feature of a large number of human tumor types opens the door to studies on the precise nature of the carcinogenic damage. This analysis is facilitated considerably by the examination of the same gene and, in some cases, the same codon in tumors arising in different tissues presumably as a result of different carcinogenic insults. Analysis of the mutational spectra occurring in *p53* in several different tumor types allows for simple and direct comparisons which are uncomplicated by the problems associated with comparisons between different genes. Preliminary evidence suggests that *p53* mutations in colon cancer, leukemias, and sarcomas are not induced by direct interaction of carcinogens with DNA. Rather, they are caused by endogenous processes with 5-methylcytosine playing a dominant role. On the other hand small-cell carcinomas of the lung show patterns of mutations consistent with direct DNA damage induced by carcinogen exposure.

Clearly the observations presented in Table 1 need to be extended, both to increase the sample size for each tumor type and to extend the comparisons to include other *p53*-related tumors. Perhaps even more interesting will be comparisons, restricted to a single tumor type, that attempt to correlate specific mutational patterns with known or presumed environmental exposures. Examples would include a comparison of the patterns of *p53* changes in lung and bladder cancers from smokers and nonsmokers. Will the smokers and nonsmokers exhibit a similar pattern of *p53* changes? If so, would this suggest that they have been exposed to carcinogens the same as or similar to those that smokers have (as, for example, through passive smoking or air pollution)? Alternatively, a mutational spectrum which includes an increase in deletions would be consistent with another environmental factor that has been suggested as being responsible for a significant proportion of lung cancer in the nonsmoker, *i.e.*, radon. It is also possible that the mutational spectrum would match closely the "baseline" pattern, which would raise the question of whether any of the above environmental factors had been important in this group of patients.

References

1. Mazur, M., and Glickman, B. Sequence specificity of mutations induced by benzo[*a*]pyrene-7,8-diol-9,10-epoxide at endogenous *APRT* gene in CHO cells. Somat. Cell Mol. Genet., *14:* 393–400, 1988.
2. Wang, P., Bennett, R. A. O., and Povirk, L. F. Melphalan-induced mutagenesis in an SV40 based shuttle vector: predominance of A.T → T.A transversions. Cancer Res., *50:* 7527–7531, 1990.
3. Loechler, E., Green, C., and Essigmon, J. *In vivo* mutagenesis by O^6-methylguanine built into a unique site in viral genome. Proc. Natl. Acad. Sci. USA, *81:* 6271–6275, 1984.
4. Zarbl, M., Sukumar, S., Arthur, A. V., Martin-Zanca, D., and Barbacid, M. Direct mutagenesis of Ha-*ras*-1 oncogenes by *N*-nitroso-*N*-methylurea during initiation of mammary carcinogenesis in rats. Nature (Lond.), *315:* 382–385, 1985.
5. Quintanilla, M., Brown, K., Ramsden, M., and Balmain, A. Carcinogen specific mutation and amplification of Ha-*ras* during mouse skin carcinogenesis. Nature (Lond.), *322:* 78–80, 1986.
6. Bressac, B., Kew, M., Wands, J., and Ozturk, M. Selective G to T mutations of *p53* gene in hepatocellular carcinoma from southern Africa. Nature (Lond.), *350:* 429–431, 1991.
7. Hsu, I. C., Metcalf, R. A., Sun, T., Welsh, J. A., Wang, N. J., and Harris, C. C. Mutational hotspot in the *p53* gene in human hepatocellular carcinomas. Nature (Lond.), *350:* 427–428, 1991.

8. Foster, P. L., Eisenstadt, E., and Miller, J. H. Base substitution mutations induced by metabolically activated aflatoxin B1. Proc. Natl. Acad. Sci. USA, *80:* 2695–2698, 1983.
9. Preston-Martin, S., Pike, M. C., Ross, R. K., Jones, P. A., and Henderson, B. E. Increased cell division as a cause of human cancer. Cancer Res., *50:* 7415–7421, 1990.
10. Vogelstein, B. A deadly inheritance. Nature (Lond.), *348:* 681–682, 1990.
11. Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. V., and Vogelstein, B. Suppression of human colorectal carcinoma cell growth by wild-type p53. Science (Washington DC), *249:* 912–915, 1990.
12. Barbacid, M. Mutagens, oncogenes and cancer. Trends Genet., *2:* 188–192, 1986.
13. Harris, A. L. Telling changes of base. Nature (Lond.), *350:* 377–378, 1991.
14. Cooper, D. N., and Youssoufian, M. The CpG dinucleotide and human genetic disease. Hum. Genet., *78:* 151–155, 1988.
15. Bottema, C. D., Ketterling, R. P., Yoon, M. S., and Somer, S. S. The pattern of factor IX germ-line mutation in Asians is similar to that of Caucasians. Am. J. Hum. Genet., *47:* 835–841, 1990.
16. Somer, S. S. Mutagen test. Nature (Lond.), *346:* 22–23, 1990.
17. Coulondre, C., Miller, J. M., Farabaugh, P. J., and Gilbert, W. Molecular basis of base substitution hotspots in *Escherichia coli.* Nature (Lond.), *274:* 775–780.
18. Rideout, W. M., Coetzee, G. A., Olumi, A. F., and Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and *p53* genes. Science (Washington DC), *249:* 1288–1290.
19. Giannelli, F., Green, P. M., High, K. A., Lozier, J. N., Lillicrap, D. P., Ludwig, M., Olak, K., Reitsma, P. M., Goossens, M., Yoshioka, A., Sommer, S., and Brownlee, G. G. Haemophilia B: database of point mutations and short additions and deletions. Nucleic Acids Res., *18:* 4053–4059, 1990.
20. Koeberl, D. D., Bottema, C. D. K., Ketterling, R. P., Bridge, P. J., Lillicrap, D. P., and Sommer, S. S. Mutations causing hemophilia B: direct estimate of the underlying rates of spontaneous germ-line transitions, transversions, and deletions in a human gene. Am. J. Hum. Genet., *47:* 202–217, 1990.
21. Chiba, I., Takahashi, T., Nau, M. M., D'Amico, D., Curel, D. T., Mitsudomi, T., Buchhagen, D. L., Carbone, D., Piantadosi, S., Koga, M., Reissman, P. T., Slamon, D. J., Holmes, E. C., and Minna, J. D. Mutations in the *p53* gene are frequent in primary resected non-small cell lung cancer. Oncogene, *5:* 1603–1610, 1990.
22. Sidransky, D., Van Eschenbach, A., Isai, Y. C., Jones, P. A., Summerhayes, I., Marshall, F., Meera, P., Green, P., Manilton, S. R., Frost, P., and Vogelstein, B. The *p53* gene is frequently altered in primary invasive bladder carcinoma and can be identified in urine sediment. Science, in press, 1991.
23. Wynder, E. L., and Goldsmith, R. The epidemiology of bladder cancer. A second look. Cancer (Phila.), *40:* 1246–1268, 1977.
24. Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Nelson, C. E., Kim, D. M., Kessel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A., and Friend, S. M. Germ-line *p53* mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science (Washington DC), *250:* 1233–1238, 1990.
25. Srivastava, S., Zou, Z., Pirollo, K., Blattner, W., and Chang, E. H. Germ line transmission of a mutated *p53* gene in a cancer-prone family with Li-Fraumeni syndrome. Nature (Lond.), *348:* 747–749, 1990.
26. Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., Van Tuinen, P., Ledbetter, D. M., Baker, D. F., Nakamura, Y., White, R., and Vogelstein, B. Chromosome 17 deletions and *p53* mutation in colorectal carcinomas. Science (Washington DC), *244:* 217–221, 1989.
27. Nigro, J. M., Baker, S. J., Preisinger, A. C., Jessup, J. M., Mostetter, R., Cleary, K., Bigner, S. M., Davidson, N., Baylin, S., Devilee, P., Glover, T., Collins, F. S., Weston, A., Modali, R., Marris, C. C., and Vogelstein, B. Mutations in the *p53* gene occur in diverse human tumor types. Nature (Lond.), *342:* 705–708, 1989.
28. Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K. V., Hamilton, S., and Vogelstein, B. *p53* gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. Cancer Res., *50:* 7717–7722, 1990.
29. Rodrigues, N. R., Rowan, A., Smith, M. E. F., Kerr, I. B., Bodmer, W. F., Ganon, J. V., and Lane, D. P. *p53* mutations in colorectal cancer. Proc. Natl. Acad. Sci. USA, *87:* 7555–7559, 1990.
30. Cheng, J., and Haas, M. Frequent mutations in the *p53* tumor suppressor gene in human leukemia T-cell lines. Mol. Cell. Biol., *10:* 5502–5509, 1990.
31. Diller, L., Kassel, J., Nelson, C. E., Gryka, M. A., Litwak, G., Gebhardt, M., Bressac, B., Ozturk, M., Baker, S. J., Vogelstein, B., and Friend, S. M. *p53* functions as a cell cycle control protein in osteosarcomas. Mol. Cell. Biol., *10:* 5772–5781, 1990.
32. Menon, A. G., Anderson, K. M., Riccardi, V. M., Chung, R. Y., Whaley, J. M., Yandell, D. W., Farmer, G. E., Freiman, R. N., Lee, J. K., Li, F. P., Barker, D. F., Ledbetter, D. M., Kleider, A., Martuza, R. L., Gusella, J. F., and Seizinger, B. R. Chromosome 17p deletions and *p53* gene mutations associated with the formation of malignant neurofibrosarcomas in von Recklinghausen neurofibromatosis. Proc. Natl. Acad. Sci. USA, *87:* 5435–5439, 1990.

From Gene to Carcinogen: A Rapidly Evolving Field in Molecular Epidemiology

Peter A. Jones, Jonathan D. Buckley, Brian E. Henderson, et al.

Cancer Res 1991;51:3617-3620.