Nucleic Acids Research, 1997, Vol. 25, No. 1 **151–157**

Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects

P. Hainaut, T. Soussi[1], B. Shomer[2], M. Hollstein[3], M. Greenblatt[4], E. Hovig[5], C. C. Harris[6] and R. Montesano*

International Agency for Research on Cancer, 150 cours Albert-Thomas, 69372 Lyon Cedex 08, France, [1]Unité 301 INSERM, 27 rue Juliette Dodu, 75010 Paris, France, [2]EMBL-Outstation-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, [3]German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany, [4]University of Vermont, College of Medicine, Health Sciences Complex, Burlington, VT 05405, USA, [5]Institute for Cancer Research, Norwegian Radium Hospital, 0310 Oslo, Norway and [6]Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD 20892, USA

Received October 11, 1996; Accepted October 15, 1996

ABSTRACT

In recent years, there has been an exponential increase in the number of p53 mutations identified in human cancers. The p53 mutation database consists of a list of point mutations in the *p*53 gene of human tumors and cell lines, compiled from the published literature and made available through electronic media. The database is now maintained at the International Agency for Research on Cancer (IARC) and is updated twice a year. The current version contains records on 5091 published mutations and is expected to surpass the 6000 mark in the January 1997 release. The database is available in various formats through the European Bioinformatics Institute (EBI) ftp server at: ftp://ftp.ebi.ac.uk/pub/databases/p53/ or by request from IARC (p53database@iarc.fr) and will be searchable through the SRS system in the near future. This report provides a description of the criteria for inclusion of data and of the current formats, a summary of the relevance of *p*53 mutation analysis to clinical and biological questions, and a brief discussion of the prospects for future developments.

INTRODUCTION

The p53 tumor suppressor gene encodes a nuclear phosphoprotein with cancer-inhibiting properties. The development of human cancer often involves inactivation of this suppressor function through various mechanisms, including gene deletions and point mutations. Since the identification of tumor-specific, missense p53 mutations in 1989, there has been a widespread interest in the possibility that the localisation and the characteristics of these mutations may reveal clues about the etiology and the molecular pathogenesis of human cancer (reviewed in 1–3). Point mutations are scattered over more than 250 codons and are common in many forms of human cancer. In this respect, the p53 gene differs from other tumor suppressor genes such as Rb, APC and $p16^{MTS-1}$ which are frequently inactivated by deletion or nonsense mutations, and from the oncogenes of the *ras* family, which are activated by mutation at a small number of well-defined codons.

The p53 protein is a multi-functional transcription factor involved in the control cell cycle progression, DNA integrity and cell survival in cells exposed to DNA-damaging agents. DNA-damage induces a transient nuclear accumulation and activation of the p53 protein, with transcriptional activation of target genes such as the cyclin kinase inhibitor p21 waf-1 (a negative regulator of cell-cycle) and the regulator of apoptosis bax-1 (a dominant-negative inhibitor of bcl-2). Most mutations impair the specific DNA-binding capacity of p53, therefore allowing cells to proliferate in conditions where cells with intact p53 function are suppressed or eliminated. Mutation of p53 may thus provide a selective advantage for the clonal expansion of pre-neoplastic or neoplastic cells. However, all mutations are not equivalent. Mutant proteins differ by the extent of their loss of suppressor function and by their capacity to inhibit wild-type p53 in a dominant-negative manner. In addition, some p53 mutants apparently exert an oncogenic activity of their own, but the molecular basis for this gain-of-function phenotype is still unclear (reviewed in 4).

The diversity of p53 mutations provide a valuable tool to identify important sources of cancer-causing agents in the human setting. Mutagens and carcinogens damage the genome in characteristic ways, leaving ‘mutagen fingerprints’ in DNA. Specific DNA changes can also be brought about by endogenous biological processes. DNA-repair and bioselection of mutants with specific properties act as additional ‘filters’ to generate the final mutation spectrum observed in any particular tumor type. Thus, p53 mutations can provide clues to the nature of exogenous

* To whom correspondence should be addressed. Tel: +33 472 73 85 32; Fax: +33 472 73 85 75; Email: hainaut@iarc.fr

agents or endogenous cellular events important in the natural history of cancer. They also reveal information on the nature of the mechanisms of mutation acquisition (5,6) and on the molecular biology of interactions of p53 with macromolecular partners and targets. Finally, identification of p53 mutations may provide useful information with respect to the detection, diagnosis and therapy of cancer.

SCOPE OF THE MUTATION COMPILATION AND CRITERIA FOR INCLUSION OF DATA

This database has been developed as a list of mutations associated with human cancers, based exclusively on published p53 mutations (7). Its scope is to retrieve and arrange data from the literature in an electronic format. This provides a powerful means for manipulation, comparison, search and retrieval of records describing the nature of *p*53 mutations in various cancers. Software which enables more sophisticated mutation spectra analysis is also available (8). Each record of the database lists published data on human somatic point mutations in cells lines, primary tumors, neoplastic and pre-neoplastic tissues. The database was last updated in January 1996 (5091 mutations) and the next update (January 1997) will contain more than 6000 entries. Germ line mutations, including those identified in families with Li–Fraumeni syndrome and known polymorphisms of the human *p*53 gene, are not included. Experimentally induced mutations in tumor cells or cell lines *in vitro*, unselected mutations in histologically normal tissue and p53 mutations in animal tumors are also beyond the present scope of our task.

The main criteria for inclusion are that the mutation are (i) reported in a peer-reviewed journal and (ii) identified by DNA sequencing of either PCR-amplified material or cloned PCR products. Mutations described in preliminary reports or in abstracts are not included. Each database entry is linked with a bibliographic reference. These references are given in a separate bibliographic list. Mutations detected by screening techniques such as the single strand conformation polymorphism (SSCP) and denaturing gradient gel electrophoresis (DGGE), but not confirmed by DNA sequencing, are not included. In many cases, the investigators have included such screening technique as a preliminary step for selecting material for sequencing, but information on these methodologies has not been entered in the present version of the database. Data on loss of heterozygosity at the p53 locus, gene rearrangements and immunohistochemical analysis with anti-p53 antibody are not part of this compilation. Note that many papers initially retrieved by common bibliographic search systems using ‘p53’ and ‘mutation’ as keywords will thus not figure in the reference list constituting the source of mutations in this database. Also, mutations identified by digestion of DNA with restriction enzymes and demonstration of an RFLP are not entered. Analysis is usually limited to exons 5–8, where most mutations cluster. A bias against identification of DNA sequence alterations outside this region can thus be expected. Tumors that were screened in only part of the mutation cluster in mid-region (exons 5–8) are excluded. If identical samples and mutations were published in more than one article, only one of the reports is referenced and the data are entered only once in the database. If the identical mutation was found in two separate samples from the same patient, for example in the primary tumor and in the metastatic tissue or in the cell line derived from the tumor or in two separate biopsies at the same site and from the same patient, the mutation is presumed to be a single event and is entered only once.

Mutations identified in tumors are presumed to be somatic unless: (i) analysis of normal tissue from the same patient demonstrated that the mutation was constitutional in that individual; (ii) the mutation corresponded to one of the known constitutional polymorphisms of the human p53 gene (at codons 21, 31, 47, 72 and 213), as these are unlikely to be mutations that arose in the tumors sequenced.

OBTAINING THE DATABASE

The data described are provided to the scientific community in several formats. The database is available as an Excel™ spreadsheet, which requires use of the Microsoft Excel™ program on either an MS-DOS™ or Windows™ systems or an Apple Macintosh™. The data have also been converted into a ‘flatfile’ format modeled according to the standard used by the EMBL nucleotide sequence database. In this format the data are stored in an ASCII text file with each column of the spreadsheet represented by a special tag. The flatfile format can be used on any computer system and with standard text editors. The data can be obtained from the EBI network server by using one of the following methods: (i) anonymous ftp://ftp.ebi.ac.uk/pub/databases/p53/ ; (ii) World Wide Web access using the URL http://www.eb.ac.uk/, selecting ‘services’ going to the databases selection.

The p53 directory contains the original spreadsheet file as a Macintosh binHex4.0 self-extracting archive. The release notes are included in the file p53.doc and the references are in p53.ref. Also include are the database in flatfile format (p53.dat) and the data in tab-delimited (data.tab) and comma-delimited (data.comma) formats for usage by other data management systems.

A version of the database which is searchable on-line has been developed at EBI using the Retrieval System (SRS; ref. 9), which allows a sophisticated combined search through the various database fields and records and linking the database with other databases. During this process, the references database has been reformatted and many references were automatically validated against the Medline™ database. We hope to complete this process for all the references during 1997. In the near future, searching the database will be made available through WWW using the SRS system at: http://www.ebi.ac.uk/srs/ . This interface will be available as soon as version 5.0 of SRS will be officially announced.

We have recently developed a more elaborated version of the database under FileMakerPro 3.0™. This requires the use Claris FileMakerPro 3.0™ software on Apple Macintosh™ or under Windows95™. This new version of the database contains searchable bibliographic records as well as calculation formulas to return mutation spectra, distribution of mutations in the codon sequence or distribution of substitutions at selected codons. Figure 2 in this article is assembled from printouts of this database version. The FileMakerPro 3.0™ version is not available at EBI but can be obtained on request at IARC. A full description of the database will be available on the IARC home page (http://www.iarc.fr/) at the beginning of 1997.

DESCRIPTION OF SPREADSHEET FORMAT

Each row (record) represents a single tumor mutation with arbitrarily assigned unique identity number. The spreadsheet columns contain the following information and abbreviations.

Column A

Unique mutation identity number. Tandem mutations (two adjacent base substitutions) are considered as one mutation event and are entered together, therefore tandem mutations have only one identity number and are a single record.

Column B

Codon number at which the mutation is located (1–393). If a tandem dinucleotide mutation spans two codons, both codons are entered. If other mutations span more than one codon, (e.g. deletion of several bases) only the first (5′) codon is entered (see note below regarding deletion, insertion and complex mutations). If the mutation is located in intron sequences this is indicated by 'intron' and the intron number.

Column C

Normal and mutated base sequence of the codon in which the mutation occurred. If the mutation is a base pair deletion or insertion this is indicated by 'del' or 'ins'.

Column D

Nucleotide position at which the mutation is located (1–1179), numbered from the ATG codon to the termination codon. This information is not entered in the present version for deletion, insertion, intron and complex mutations (see note below).

Column E

Base change, read from the coding strand by convention, for base substitutions. For deletions (indicated by '–') and insertions (indicated by '+') the number of bases deleted or inserted is given in parentheses.

Column F

The name or number given by the authors to the tumor sample or cell line is entered here. If the name is not distinctive, e.g., if the publication refers to samples as tumors 1, 2, 3, etc., then we have arbitrarily assigned a name, usually the first letters of the first author's name, followed by the numbers in the series. Thus entirely different tumors may have, by chance, a similar or identical trivial tumor sample name (see recommendations to authors below). If more than one mutation has been found in the same sample, the tumor name in column F is suffixed with an apostrophe.

Column G

Anatomic site or type of tumor as described in the publication cited. Abbreviations used in this column are: HCC (hepatocellular carcinoma); Leuk/Lym (leukemias and lymphomas); cholangioca (cholangiocarcinoma); choriocarc (choriocarcinoma); colon, (cancers of the colon or rectum). We are currently modifying these entries to adopt the WHO International Classification of Disease for Oncology. New definitions should be available in the database by the end of 1997.

Column H

Reference number indicating the publication in which the mutation is described. The full citation (authors, title, journal, pages, year) is given as a separate text file (see Notes to authors, below).

Column I

This is a column with heterogeneous notes, usually containing comments regarding the tumor or the patient, such as histological type of tumor or exposure history or other information emphasized by authors reporting the mutations. The terminology used by the authors has been retained and no attempt has been made to complete the data with unpublished information or to standardize the entries. Abbreviations of tumor subtype or cell type are as follows: SCLC, small cell lung cancer; adenoca, adenocarcinoma; adenosq, mixed adenosquamous carcinoma; medullobl, medulloblastoma; SCC, squamous cell carcinoma; TCC, transitional cell carcinoma; NPC, nasopharyngeal carcinoma. For abbreviations of sarcoma subtypes or leukemia and lymphoma subclassifications, e.g. ATL (adult T cell leukemia), refer to the cited reference. Uniformity of these abbreviations in the different reports has not been verified.

Other abbreviations: UC, ulcerative colitis; FAP, familial adenoma polyposis; XP, xeroderma pigmentosum; HPV+ or HPV–, tumor harboring or lacking human papilloma virus DNA; diff, differentiated tumor; undiff, undifferentiated tumor; CIS, carcinoma *in situ*; premal, premalignant.

Other information. (i) 'metastasis' specifies that the DNA analyzed for the mutation was obtained from metastatic tissue (the primary tumor location is in column G). (ii) Examples of exposure history: tobacco smoke; radon gas, etc.

The content of this column is also being reorganized towards greater uniformity. A new version of this column should be available by the end of 1997.

Column J

An entry 'L' indicates that the material examined was from a tumor cell line. If there is no entry the material is from tumor tissue or biopsy (most instances), xenograft or unspecified.

Column K

Mutations that are single base transitions at CpG dinucleotides, i.e. CpG→TpG or CpG→CpA are designated by 'yes'. If there is no entry the mutation does not fall into this category.

Column L

Chain terminating mutations due to single base substitutions are designated by '(three letter amino acid abbreviation) →stop'. Frame-shift mutations are designated by 'frameshift', whereas in-frame deletions and insertions are designated 'deletion' or 'insertion'. Mutations that do not results in an amino acid change are designated 'silent', while mutations that occurred in intron sequences are sometimes indicated by the term 'splicing', even though in most instances it was not determined whether splicing

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	143	GTG to GCG	428	T to C	Cx3	Colon	1				Val->Ala		
2	175	CGC to CAC	524	G to A	Cx1	Colon	1			yes	Arg->His		
3	132	AAG to CAG	394	A to C	BT 20	Breast	2		L		Lys->Gln		
4	249	AGG to AGC	747	G to C	BT 549	Breast	2		L		Arg->Ser		
5	280	AGA to AAA	839	G to A	MDA231	Breast	2		L		Arg->Lys		
6	285	GAG to AAG	853	G to A	BT 474	Breast	2		L		Glu->Lys		
7	157	GTC to TTC	469	G to T	OZ1	HCC	3				Val->Phe		
8	249	AGG to AGT	747	G to T	OZ2	HCC	3				Arg->Ser		
9	249	AGG to AGT	747	G to T	OZ3	HCC	3				Arg->Ser		
10	249	AGG to AGT	747	G to T	OZ4	HCC	3				Arg->Ser		

Figure 1. Section of the database in Excel spreadsheet format showing record 1 as an example.

errors did result from the mutation (some of these base changes are likely to be phenotypically silent).

Column M

If the information on the nature or location of the mutation in the reference is ambiguous or contradictory the letter ‘e’ appears in this column.

Column N

When a record requires correction or information is added the date of the revision will be noted in this column. Record revisions are due to inherent inconsistencies or ambiguities in the published report or omissions or errors during data entry.

Example: The information on the mutation identified as 1 on the spreadsheet (Fig. 1) is as follows:

Column A Unique mutation identity number is 1.
Column B Mutation is located in codon 143.
Column C The wild-type codon sequence is GTG and the mutated allele sequence is GCG.
Column D The nucleotide position of the base change is 428.
Column E The base change is T→C.
Column F The tumor sample name is Cx3.
Column G The DNA harboring the mutation was obtained from a tumor of the colon or rectum.
Column H The mutation is reported in reference 1 of the database.
Column I (Blank: no comments regarding histology or patient data were entered for this record).
Column J (Blank: the DNA analyzed was not obtained from a cell line).
Column K (Blank: the mutation is not a base transition substitution at a CpG dinucleotide).
Column L The amino acid change is from valine to alanine.
Column M (Blank: no errors to the published information have been found).
Column N (Blank: no corrections or additions to the record have been made since it was entered).

Comments regarding insertion, deletion and complex mutations

Since precise information describing these kinds of mutations requires a more expanded format than the current spreadsheet and requires conventions for presenting sequence changes that are inherently ambiguous, e.g. base deletions in repetitive sequences, our compilation presently gives minimal information on these mutations. A HUGO initiative is underway to standardize and establish conventions for specific locus mutation databases. We will attempt to follow the recommendations in revising our format and in providing additional sequence information on this class of mutation.

Notes to authors

When tumor mutations are reported for the second time in a new publication we recommend this be stated in a footnote to the table where the mutations are re-listed, also indicating which tumor mutations were reported previously. Providing tumor samples with unique case numbers would also help to avoid redundancies in the database.

The inherent inconsistencies we have detected (~2% of all records) in reported mutations were usually traceable to typographical errors in the publication, to reading the genetic code from the wrong DNA strand or to misnumbering the codons in sequencing film illustrations.

The electronic form of the database may be cited by referencing this *Nucleic Acids Research* article.

p53 MUTATIONS AND HUMAN CANCER

Classes of p53 mutants

The p53 protein comprises several domains, including an acidic N-terminal region containing the transactivation domain (residues 20–42), a core containing the sequence-specific DNA-binding domain (residues 100–293) and a complex C-terminal domain with multiple functions such as promotion of oligomerisation, interactions with heterologous proteins and non-sequence-specific binding to single-stranded nucleic acids and damaged DNA *in vitro*. The protein is phosphorylated on multiple sites in both N- and C-terminal portions and forms high molecular weight complexes with several cellular proteins, including components of the multi-molecular machines regulating gene transcription (TFIID), DNA replication (RPA) and DNA repair (ERCC3) (reviewed in 1–4 and references therein).

The exact role of each individual domain in the tumor suppressive properties of p53 is not fully clear. Functionally, the N- and C-termini are replaceable by heterologous transactivation and oligomerization domains, suggesting that the core, DNA-binding domain alone is sufficient to confer the specific suppressive properties. About 90% of mutations reported in the

THE p53 MUTATION DATABASE

Codon	substitution	Number	Total	Frequency (in %)
all cancers				
273	Arg->Cys	130	348	37,4
273	Arg->His	157	348	46.3
273	Arg->Gly	2	348	,6
273	Arg->Pro	8	348	2,3
273	Arg->Leu	41	348	11,8
273	Arg->Ser	4	348	1,1
273	Arg->Asn	1	348	,3
273	frameshift	1	348	,3
lung cancers				
273	Arg->Cys	7	41	17,1
273	Arg->His	7	41	17,1
273	Arg->Pro	5	41	12,2
273	Arg->Leu	18	41	43,9
273	Arg->Ser	3	41	7,3
273	Arg->Asn	1	41	2,4
Colon cancers				
273	Arg->Cys	20	48	41,7
273	Arg->His	28	48	58,3

Figure 2. Different substitutions at codon 273 according to tumor type. This figure shows a typical output of p53 mutation analysis using the version of the database under FileMakerPro 3.0™. Mutations at codon 273 have been selected, sorted according to the amino acid substitution and summarized. The entire database contains 348 mutations at codon 273 (Arg), the most frequent substitutions being to His (45.1%), Cys (37.4%) and Leu (11.8%). In lung cancers, 41 mutations at codon 273 are found, and the most frequent substitutions are to Leu (43.9%), Cys (17.1%) and His (17.1%). Note also the relatively high frequency of substitutions to Ser (7.3%) and Pro (12.2%). In colon cancers (48 mutations), only substitutions to His (58.3%) and Cys (41.7%) are found. This example shows that different tumor types may select for different mutations at codon 273.

database are found in the core domain (this proportion may be overestimated, since many investigators have limited their analysis to exons 5–8). Mutations at five 'hotspot' codons (175, 245, 248, 249 and 273) represent ~20% of all mutations found so far.

The DNA-binding domain is made of two antiparallel β sheets forming a 'scaffold' supporting a DNA-binding surface of non-contiguous loops and helixes (10). Mutations can be grouped in three broad classes according to their impact on the structure of the DNA-binding domain. Class I mutations affect residues of the DNA-binding surface, such as Arg 248 and Arg 273, and disrupt protein–DNA contact points. Class II affect residues crucial for the correct orientation of the DNA-binding surface (such as Arg175 and Arg249, which are involved in the connections between the scaffold and the binding surface). These mutations may disrupt the regulation of p53 protein flexibility. Class III mutations fall within the 'scaffold' and disrupt the tertiary structure of the whole DNA-binding domain. There is evidence, both *in vitro* and *in vivo*, that mutants corresponding to these categories, have distinct functional properties. However, the

properties of mutant p53 are also cell type-specific (11–13). Classification of p53 mutants into structural groups may provide a molecular basis to explain properties of mutant p53 such as dominant-negative activity, oncogenic potential or temperature-sensitivity. In addition, the functional properties of a particular mutant depends upon the nature of the amino-acid substitution at a given codon (11). Analysis of the database reveals that the nature of the substitutions found at the most frequently mutated codons may vary from one tumor type to the other (Fig. 2).

p53 mutations as 'mutagen fingerprints'

Differences in patterns of p53 mutations in several types of cancer reflect the effect of specific carcinogens (reviewed in 1–3). Well-characterised examples of such 'mutagen fingerprints' include G:C to T:A transversions in lung cancers in association with cigarette smoke, G:C to T:A transversions at codon 249 in liver cancers in association with dietary exposure to Aflatoxin B1 (AFB1) and CC:GG to TT:AA tandem dipyrimidine transitions in skin cancers in association with UVB exposure. In these examples, the nature of the mutations observed is consistent with the type of damage induced by these carcinogens, and experimental studies are also supportive of these associations. In the case of liver cancer and AFB1 exposure, this carcinogen has been shown to induce G to T transversions in cultured human liver cells and these mutations preferentially localized in the third nucleotide of codon 249 of the p53 gene (14). In skin cancers associated with UVB exposure, recent studies in basal cell carcinoma show that the tandem dipyrimidine transitions are preferentially detected in cancers developing from regions of the skin exposed to sunlight, whereas tumors arising from less exposed areas show mostly transversions (15). These data also implicates agents different than UVB in the etiology of skin cancers. In experimental skin cancers induced in mice, distinct types of p53 mutations are observed in tumors induced by UVB and by PUVA (8-methylpsoralen + UVA) (16).

Another characteristic 'mutagen fingerprint' is the high proportion of mutations at A:T base pairs in hemangiosarcomas of the liver (ASL) of workers occupationally exposed to vinyl chloride. These mutations are uncommon in sporadic or thorotrast-induced ASL (17) but occur at high frequency in ASL occurring in rats exposed to vinyl chloride (A. Barbin *et al*., unpublished results).

Controversial data have been reported regarding the possibility that very specific mutations in p53 may exist in lung cancers from uranium miners exposed to radon. An initial study by Taylor *et al*. (18) reported a high frequency of transversions at the second base of codon 249. However, this observation was not confirmed in other studies by Vähäkangas *et al*. (19) who found a heterogeneous mutation spectrum, and by Bartsch *et al*. (20).

p53 mutations as clues to the etiopathogenesis of cancer

Given the complexity of most exposures, it may not be surprising that we have only few examples of clear association between a mutagen and a mutation pattern (see comprehensive review in 1). Most cancers show a complex and heterogeneous mutation pattern, but the growing body of data available may help to reveal more subtle differences in the mutation patterns, in particular in studies with exposure cohorts matched for various parameters that could influence the mutation spectrum (such as age, sex, ethnic origin, etc.). The analysis of the mutation spectrum in breast cancers is a typical but controversial example of this situation. As primarily shown by Biggs *et al.* (21) and confirmed by further studies, the prevalence of G:C to T:A transversions and G:C to A:T transitions in breast cancer is intermediate between colon and lung cancer, implying that some environmental factor(s) may play a role in the genesis of p53 mutations in breast cancers. Recent studies have shown that the frequency and pattern of p53 mutations differ in populations from different geographic origin. In Japanese women, the prevalence of p53 mutation is high (59.3%), compared with US or European populations (25–35%) (22–23). Furthermore, there is a high frequency of microdeletions in the rural white US population (24–25). These observations further demonstrate that the etiology of breast cancer is highly complex but also imply that exogenous factors may contribute to this cancer.

Clinical implications of p53 mutations

A large proportion of the literature on p53 mutations addresses the usefulness of mutation analysis in the molecular pathology of cancer. Mutations may serve as molecular indicators of clonality or as early markers of relapse in a patient with a previously identified mutation in the primary tumor. There is a growing body of evidence that the prognosis of cancer may differ according to the presence and the nature of p53 mutations (26). As p53 plays an essential role in the cellular response to DNA-damage, the mutation status may be an important determinant of the tumor response to chemo- or radiotherapy. Finally, the presence of p53 antibodies in the serum of some cancer patients may provide an interesting tool for diagnosis and follow-up of cancer (27).

FUTURE PROSPECTS

The p53 mutation database initially started as a simple list of mutations and has proven to be an invaluable source of information on mutations in human cancer. As this database is rapidly developing, and as the secrets of p53 protein functions progressively unravel, new challenges are raised by the multiple implications of p53 mutations for human cancers and the demands of a growing community of users. In the future, the database should be capable of providing rapid and detailed information, as well as links with other related databases, in the following areas: (i) molecular epidemiology of cancer, (ii) molecular pathology and (iii) structural analysis of p53 protein. Meeting these challenges will require the re-evaluation of the existing p53 literature and to incorporate in the database more detailed and specific information on the techniques used for mutation analysis, on the identification of the tumor pathology and grade, on the individual characteristics of the tumors examined (age and ethnicity of the patients, prognosis, response to treatment), and on individual exposures to cancer risk factors. It will also be necessary to summarize in the database the existing information on the biological properties of mutant p53 proteins (dominant-negative activity, oncogenic activity, temperature-sensitivity, tumor suppressive potential). Finally, the availability of crystal structures of specific conformers of p53 will allow to analyse the structural effects mutations and to predict the functional properties of p53 mutants. Meeting these demands, within the scope of a database which is simple to access and to use and which meets the highest ethical standards for confidentiality of individual data, is a major challenge for the years to come and will necessitate to reconsider the structure and scope of this database.

ACKNOWLEDGEMENTS

We thank Mrs M. Wrisez for secretarial assistance. The partial support of EC grant 'Environment and Climate' EV5V-CT92-0199 is acknowledged.

REFERENCES

1 Greenblatt,M.S., Bennett,W.P., Hollstein,M. and Harris,C.C. (1994) *Cancer Res.*, **54**, 4855–4878.
2 Harris,C.C. (1996) *Br. J. Cancer*, **73**, 261–269.
3 Soussi,T. (1996) *Mol. Med. Today*, **2**, 32–37.
4 Ko,L.J. and Prives,C. (1996) *Genes Dev.*, **10**, 1054–1072.
5 Greenblatt,M.S., Grollman,A.P. and Harris,C.C. (1996) *Cancer Res*., **56**, 2130–2136.
6 Krawczak,M., Smith-Sorensen,B., Schmidtke,J., Kakkar,V.V., Cooper,D.N. and Hovig,E. (1995) *Hum. Mutat.*, **5**, 48–57.
7 Hollstein,M., Shomer,B., Greenblatt,M., Soussi,T., Hovig,E., Montesano,R. and Harris,C.C. (1996) *Nucleic Acids Res*., **24,** 141–146.
8 Cariello,N.F., Douglas,G.R. and Soussi,T. (1996) *Nucleic Acids Res*., **24,** 119–120.
9 Etzold,T. and Argos,P. (1993) *Comput. Appl. Biosci*., **9,** 49–57.
10 Cho,Y., Gorina,S., Jeffrey,P.D. and Pavletich,N.P. (1994) *Science*, **265**, 346–355.
11 Ory,K., Legros,Y., Auguin,C. and Soussi,T. (1994) *EMBO J*. **13**, 3496–3504.
12 Forrester,K., Lupold,S.E., Ott,V.L., Chay,C.H., Band,V., Wang,X.W. and Harris,C.C. (1995) *Oncogene*, **10**, 2103–2011.
13 Rolley,N., Butcher,S., Milner,J. (1995) *Oncogene*, **11,** 763–770.
14 Aguilar,F., Hussain,S.P., Cerutti,P. (1993) *Proc. Natl. Acad. Sci USA*, **90,** 8586–8590.
15 Gailani,M.R., Leffell,D.J., Ziegler,A.M., Gross,E.G., Brash,D.E. and Bale,A.E. (1996) *J. Natl. Cancer Inst*., **88**, 349–354.
16 Nataraj,A.J., Black,H. and Ananthaswamy,H.N. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 7961–7965.
17 Soini,Y., Welsh,J.A., Ishak,K.G. and Bennett,W.P. (1995) *Carcinogenesis*, **16,** 2879–2881.
18 Taylor,J.A., Watson,M.A., Devereux,T.R., Michels,R.Y., Saccomanno,G. and Anderson,M. (1994) *Lancet*, **343,** 86–87.
19 Vahakangas,K.H., Samet,J.M., Metcalf,R.A., Welsh,J.A., Bennett,W.P., Lane,D.P. and Harris,C.C. (1992) *Lancet*, **339,** 576–580.
20 Bartsch,H., Hollstein,M., Mustonen,R., Schmidt,J., Spiethoff.,A.,Wesch,H., Wiethege,T. and Muller,K.M. (1995) *Lancet*, **346,** 121.
21 Biggs,P.J., Warren,W.,Venitt,S. and Stratton,M.R. (1993) *Mutagenesis*, **8,** 275–283.
22 Hartmann,A., Rosanelli,G., Blaszyk,H., Cunningham,J.M., McGovern,R.M., Schroeder,J.J., Schaid,D.J., Kovach,J.S. and Sommer,S.S. (1995) *J. Clin. Invest*., **95,** 686–689.
23 Hartmann,A., Blaszyk,H., Saitoh,S., Tsushima,K., Tamura,Y., Cunningham,J.M., McGovern,R.M., Schroeder,J.J., Sommer,S.S. and Kovach,J.S. (1996*) Br. J. Cancer*, **73,** 896–901.
24 Saitoh,S., Cunningham,J., De Vries,E.M., McGovern,R.M., Schroeder,J.J., Hartmann,A., Blaszyk,H., Wold,L.E., Schaid,D., Sommer,S.S., *et al*. (1994) *Oncogene*, **9**, 2869–2875.
25 Sommer,S.S., Cunningham,J., McGovern,R.M., Saitoh,S., Schroeder,J.J., Wold,L.E. and Kovach,J.S. (1992) *J. Natl. Cancer Inst*., **84,** 246–252.
26 Goh,H.S., Yao,J. and Smith,D.R. (1995) *Cancer Res*., **55,** 5217–5221.
27 Soussi,T. (1996) *Immunol. Today*, **17**, 354–356.