Nucleic Acids Research, 1994, Vol. 22, No. 17 3547–3548

Database and software for the analysis of mutations at the human *hprt* gene

Neal F.Cariello
Glaxo, Inc., Drug Safety Assessment, 5 Moore Drive, 9.2011, Research Triangle Park, NC 27709, USA

ABSTRACT

A computerized database containing DNA sequence information regarding human HPRT mutants has been created. The database itself is in the dBASE format and contains information on about 1500 mutants. In addition, an IBM PC compatible software package to analyze the information in the database has been developed. Both the database and software are freely available via the Internet.

INTRODUCTION

The hypoxanthine guanine phosphoribosyl transferase (*hprt*) gene codes for an enzyme that functions in the purine salvage pathway. Human germinal mutations at this X-linked locus can produce offspring with the Lesch–Nyhan syndrome or gouty arthritis (for review see reference 1). The sequence of the cDNA has been known for over a decade (2) and the entire 45 kb gene has been sequenced (3).

The selection system for *hprt* mutants is phenotypic, cells with a poorly-functioning enzyme will be resistant to the toxic effects of purine analogs. Cells bearing a mutation in the *hprt* gene can be selected and cloned from tissue culture experiments and from T-cells isolated from rodents (4), primates (5) and humans (6, 7). Thus somatic mutations arising *in vivo* in humans can be studied.

In vivo human mutations at the *hprt* gene are currently of great interest. Mutagens typically induce a unique pattern of mutational changes in a given gene and it may be possible to use the pattern of mutations found at the *hprt* locus as a biomonitor of mutagenic exposure. Several studies have shown an increase in *hprt* mutant fraction as a function of a known mutagenic exposure (for review see reference 8).

There is a considerable body of information about the *in vitro* mutational spectra of different mutagens and carcinogens at the *hprt* locus. It may be possible to relate the mutations in the *hprt* gene resulting from a human *in vivo* mutagenic exposure to a specific *in vitro* mutational spectrum.

DATABASE AND SOFTWARE

In order to facilitate the mutational analysis of the *hprt* gene, we have created a computerized database containing DNA sequence information on human mutants (9). The database itself is in the dBASE format. Information currently exists on about 1500 mutants; 550 of these mutants were isolated from humans *in vivo*, and 950 mutants were derived from tissue culture experiments.

The information in the *hprt* database has been used to produce a very detailed description of mutations at this locus (10); such an extensive analysis was feasible through the development of numerous computer programs to query the *hprt* database. A single stand-alone executable program is available that incorporates many of the individual routines that were developed in conjunction with the *hprt* database (11).

Numerous routines have been developed for the analysis of single base substitutions, including programs to (i) determine if two mutational spectra are different, (ii) display the number of mutations and mutable sites in each exon, (iii) determine if mutations show a DNA strand bias, (iv) determine the frequency of transitions and transversions, (v) display the number and kind of mutations observed at each base in the coding region, (vi) perform nearest neighbor analysis, and (vii) display mutable amino acids in the *hprt* protein.

An IBM-compatible personal computer running MS-DOS is required to run the stand-alone executable. A 80386 or greater processor is absolutely required, the program will not load on a machine with a 80286 processor. 4 MB of RAM is necessary to run all modules. A hard disk is required, the programs occupy about 3 MB of disk space. A standard VGA color monitor and video adapter is required (640×480 resolution). The program is designed for use with a mouse; users without a mouse will be disappointed.

AVAILABILITY

The database and software program are freely available to the scientific community so other experimenters can use these tools to pursue their individual research interests.

The preferred distribution method is remote file transfer using INTERNET. The node name for remote file transfer is UNCVX1.OIT.UNC.EDU (152.2.21.17). When prompted for a Username, enter ANONYMOUS, when prompted for a Password, enter any characters. Upon login, users will be at the top-level directory, PUB. The files of interest can be found below the subdirectory CARIELLO \ HPRT.

Two files are present for remote file transfer in the subdirectory CARIELLO \ HPRT \ ANALYZE, HH_READ.ME and HH_UNPAK.EXE. The HH_READ.ME file is a standard text file which contains information for remote file transfer. The HH_UNPAK.EXE file is a self-extracting compressed file which contains the executable program, associated files, and instructions for use. It is expected that almost all experimenters do have remote file transfer capability at their sites. Electronic

mail should be sent to CARIELLOHPRT@UNCVX1.OIT.UNC.EDU.

The program and database will be mailed to users without access to remote file transfer. Users must send a self-addressed, postage-paid, floppy disk mailer and a blank 1.44 M floppy disk, and the files will be transferred and returned by mail.

CONCLUSION

The database and software program simplifies the analysis of the rapidly increasing information about *hprt* mutation. The programs permit the facile comparison between *in vitro* and *in vivo* data, as well as the identification of mutational patterns that may be of importance to experimenters using *hprt* as a biomonitor and of importance to researchers studying mechanisms of mutation.

REFERENCES

1. Caskey, C.T. and Kruh, G.D. (1979): The HPRT locus. Cell, 16:1–9.
2. Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T., Friedmann, T. (1983) Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyl transferase. Proc. Natl. Acad. Sci. (USA), 80, 477–481.
3. Edwards, A,. Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C.T., Ansorge, W. (1990): Automated DNA sequencing of the human HPRT Locus. Genomics 6: 593–608.
4. Jones, I.M., Burkhart-Schultz, K., and T. L. Crippen (1987) Cloned mouse lymphocytes permit analysis of somatic mutations that occur in vivo, Som. Cell Mol. Genet., 13, 325–333
5. Harbach, P.R., Filipunas, A.L., Wang,Y., Aaron,C.S. (1992) DNA sequence analysis of spontaneous and N-ethyl-N-nitrosourea-induced hprt mutations arising in vivo in cynomolgus monkey T-lymphocytes. Environ. Mol. Mutagen., 20, 96–105.
6. Albertini, R.J., O'Neill, J.P., Nicklas, J.A. Heinz, N.H., and P.C. Kelleher (1985) Alterations of the *hprt* gene in human in vivo-derived 6-thioguanine-resistant T lymphocytes, Nature, 316, 369–371.
7. Turner, D.R., Morley, A.A, Haliandros, M., Kutlaca, R., and B.J. Sanderson (1985) In vivo somatic mutations in human lymphocytes frequently result from major gene alterations, Nature, 315, 343–345.
8. Cole, J. and Skopek, T. R. (1994) Mutant frequency, mutation rate and mutational spectra in the human population *in vivo*. Mutat. Res., 304, 33 - 105.
9. Cariello, N.F., Craft, T.R., Vrieling, H., van Zeeland, A.A., Adams, W.T., and T.R. Skopek (1992) Human HPRT mutant database: software for data entry and retrieval, Environ. Mol. Mutagen., 20, 81–83.
10. Cariello, N.F. and T.R. Skopek (1993) Analysis of mutations occurring at the human *hprt* locus, J. Mol. Biol., 231, 41–57.
11. Cariello, N.F. (1994) Software for the analysis of mutation at the human *hprt* locus. Mutat. Res., 312, 173–195.