Vol. 141, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 26, 1986 Pages 229-237

HUMAN LIVER GLUTATHIONE S-TRANSFERASES: COMPLETE PRIMARY SEQUENCE OF AN H_a SUBUNIT cDNA

Chen-Pei D. Tu† and Biao Qian

Department of Molecular and Cell Biology,
The Pennsylvania State University,
University Park, PA 16802

Received October 13, 1986

SUMMARY: Multiple human liver GSH S-transferases (GST) with overlapping substrate specificities may be essential to their multiple roles in xenobiotics metabolism, drug biotransformation, and protection against peroxidative damage. Human liver GSTs are composed of at least two classes of subunits, H_a (M_r=26,000) and H_b (M_r=27,500). Immunological cross-reactivity and nucleic acid hybridization studies revealed a close relationship between the human H_a subunit and rat Y_a, Y_c subunits and their cDNAs. We have determined the nucleotide sequence of the H_a subunit 1 cDNA, pGTH1. The alignments of its coding sequence with the rat Y_a and Y_c cDNAs indicate that they are ~80% identical base-for-base without any deletion or insertion. Regions of sequence homology (>50%) have also been found between pGTH1 and a corn GST cDNA and rat GST cDNAs of the Y_b and Y_p subunits. Among the 62 highly conserved amino acid residues of the rat GST supergene family, 56 of them are preserved in the H_a subunit 1 coding sequences. Comparison of amino-acid replacement mutations in these coding sequences revealed that the percentage divergence between the rat Y_a and Y_c genes is more than that between the H_a and Y_a or H_a and Y_c genes.

Multiple GSH S-transferases (GST, EC 2.5.1.18) have been purified from human liver cytosol (1-8) and thirteen isozymes were resolved by chromatofocusing (1). They are composed of at least two classes of subunits, H_a (M_r=26,000) and H_b (M_r=27,500), according to their relative electrophoretic mobility to rat GSTs (1,9). We have demonstrated earlier a definitive immunological cross-reactivity between the human liver and various rat GSTs and nucleotide sequence homology between their respective cDNAs (9). Human liver GSTs lack a mobility class equivalent to the rat liver GST Y_c subunits (M_r=28,000) yet Y_c cDNA (pGTR262) (9,10) hybridized to the H_a subunit cDNA (pGTH1) as strongly as did the Y_a cDNA (pGTR261) (9,11). Furthermore, the H_a subunit 1 cDNA, pGTH1, selected rat Y_a and Y_c subunit mRNAs with almost equal efficiency in hybrid-selected *in vitro* translation (9). In this communication, we provide a molecular basis to these observations by determining the nucleotide sequence of the H_a subunit 1 cDNA, pGTH1.

†To whom correspondence should be addressed.

0006-291X/86 $1.50

EXPERIMENTAL PROCEDURES

Antibiotics, nucleotides, and enzyme reagents: Tetracycline, chloramphenicol and ampicilln were purchased from Sigma (St. Louis, MO). The dNTPs and ddNTPs were obtained from pharmacia - P.L. Biochemicals (Milwaukee, WI). [α^{32}-P] dNTPs and [α-S^{35}] dATP were products of ICN pharmaceuticals (Irvine, CA) and New England Nuclear (Boston, MA) respectively. Restriction endonucleases were products of New England Biolabs (Beverly, MA). DNA polymerase I and its Klenow fragment were products of Boehringer Mannheim and New England Biolabs (Beverly, MA), respectively.

DNA purification and sequence analysis: Plasmid DNA isolation (12), DNA fragment purification (13), restriction endonuclease digestion, 3'-end labelling of DNAs (13) were carried out according to published procedures (14). DNA sequence analysis was carried out by the chemical method of Maxam and Gilbert (15) or by the chain-termination method (16) after subcloning the EcoRI insert of pGTH1 into the M13 mp18 sequencing vector (17).

Computer analysis of DNA sequence homology: DNA sequence homology among various GST cDNAs and divergence for amino-acid replacement sites were measured by the Dot Matrix Program of Zweig (18) and the Diverge Program of Perler et al. (19), respectively, with an IBM PC/AT personal computer.

RESULTS AND DISCUSSION

Nucleotide sequence of the pGTH1 cDNA insert.

We have isolated several cDNA clones from a human liver λgt11 cDNA expression library with the anti-human liver GST antiserum. One of them, designated λGTH1, has been characterized after subcloning into pBR325 by hybrid-selected in vitro translation and DNA blot hybridization to be a cDNA for the H_a subunit (9). The cDNA sequence of pGTH1 was determined by a combination of the chemical method (15) and the chain-termination method after subcloning into the EcoRI site of the M13 mp18 vector (16,17) as outlined in the sequencing strategy (Figure 1).

The cDNA is 810 nucleotides long, containing 66 nucleotides in the 5' noncoding region, a 222 amino acid open reading frame (M_r=25,629) and a complete (78bp) 3' noncoding sequence including the poly(A) addition signal

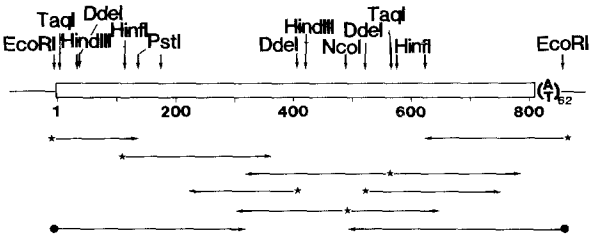

Figure 1. Sequencing strategy for the H_a subunit 1 cDNA, pGTH1.

The EcoRI insert was purified before restriction digestion and 3' end-labelling (*) and before cloning into the EcoRI site of the M13 mp18 sequencing vector. Both orientations of the EcoRI fragment subclone were used for sequence analysis (●). The Maxam-Gilbert chemical reactions were repeated at least once on each of the 3' end-labelled DNA fragments.

```
          10          20          30          40          50          60
AGT TGT CGA GCC AGG ACG GTG ACA GCG TTT AAC AAA GCT TAG AGA AAC CTC CAG GAG ACT

          70          80          90          100         110         120
GCT ATC ATG GCA GAG AAG CCC AAG CTC CAC TAC TTC AAT GCA CGG GGC AGA ATG GAG TCC
        MET Ala Glu Lys Pro Lys Leu His Tyr Phe Asn Ala Arg Gly Arg Met Glu Ser

          130         140         150         160         170         180
ACC CGG TGG CTC CTG GCT GCA GCT GGA GTA GAG TTT GAA GAG AAA TTT ATA AAA TCT GCA
Thr Arg Trp Leu Leu Ala Ala Ala Gly Val Glu Phe Glu Glu Lys Phe Ile Lys Ser Ala

          190         200         210         220         230         240
GAA GAT TTG GAC AAG TTA AGA AAT GAT GGA TAT TTG ATG TTC CAG CAA GTG CCA ATG GTT
Glu Asp Leu Asp Lys Leu Arg Asn Asp Gly Tyr Leu Met Phe Gln Gln Val Pro Met Val

          250         260         270         280         290         300
GAG ATT GAT GGG ATG AAG CTG GTG CAG ACC AGA GCC ATT CTC AAC TAC ATT GCC AGC AAA
Glu Ile Asp Gly Met Lys Leu Val Gln Thr Arg Ala Ile Leu Asn Tyr Ile Ala Ser Lys

          310         320         330         340         350         360
TAC AAC CTC TAT GGG AAA GAC ATA AAG GAG AGA GCC CTG ATT GAT ATG TAT ATA GAA GGT
Tyr Asn Leu Tyr Gly Lys Asp Ile Lys Glu Arg Ala Leu Ile Asp Met Tyr Ile Glu Gly

          370         380         390         400         410         420
ATA GCA GAT TTG GGT GAA ATG ATC CTC CTT CTG CCC GTA TGT CCA CCT GAG GAA AAA GAT
Ile Ala Asp Leu Gly Glu Met Ile Leu Leu Leu Pro Val Cys Pro Pro Glu Glu Lys Asp

          430         440         450         460         470         480
GCC AAG CTT GCC TTG ATC AAG GAG AAA ATA AAA AAT CGC TAC TTC CCT GCC TTT GAA AAA
Ala Lys Leu Ala Leu Ile Lys Glu Lys Ile Lys Asn Arg Tyr Phe Pro Ala Phe Glu Lys

          490         500         510         520         530         540
GTC TTA AAG AGC CAT GGA CAA GAC TAC CTT GTT GGC AAC AAG CTG AGC CGG GCT GAC ATT
Val Leu Lys Ser His Gly Gln Asp Tyr Leu Val Gly Asn Lys Leu Ser Arg Ala Asp Ile

          550         560         570         580         590         600
CAT CTG GTG GAA CTT CTC TAC TAC GTC GAG GAG CTT GAC TCC AGT CTT ATC TCC AGC TTC
His Leu Val Glu Leu Leu Tyr Tyr Val Glu Glu Leu Asp Ser Ser Leu Ile Ser Ser Phe

          610         620         630         640         650         660
CCT CTG CTG AAG GCC CTG AAA ACC AGA ATC AGC AAC CTG CCC ACA GTG AAG AAG TTT CTA
Pro Leu Leu Lys Ala Leu Lys Thr Arg Ile Ser Asn Leu Pro Thr Val Lys Lys Phe Leu

          670         680         690         700         710         720
CAG CCT GGC AGC CCA AGG AAG CCT CCC ATG GAT GAG AAA TCT TTA GAA GAA GCA AGG AAG
Gln Pro Gly Ser Pro Arg Lys Pro Pro Met Asp Glu Lys Ser Leu Glu Glu Ala Arg Lys

          730         740         750         760         770         780
ATT TTC AGG TTT TAA TAA CGC AGT CAT GGA GGC CAA GAA CTT GCA ATA CCA ATG TTC TAA
Ile Phe Arg Phe

          790         800         810
AGT TTT GCA ACA ATA AAG TAC TTT ACC TAA (A60)
```

Figure 2. Nucleotide sequences of pGTH1 cDNA amino acid sequences of the H_a subunit 1. The putative initiation codon ATG is boxed and the poly(A) addition signal is underlined.

AATAAA (Figure 2). The coding region nucleotide sequence is approximately ~80% identical base-for-base to the rat liver Y_a (pGTR261) subunit sequence (79.6%) (11) and to the rat liver Y_c (pGTB42) subunit sequence (79.9%) (20). The predicted amino acid sequence of this human H_a subunit is as different from the rat Y_a subunit as is from the Y_c subunit of rat liver GSTs, containing 55 substitutions among the 222 residues (75.2%) (10,11,20).

The 3' noncoding region of pGTH1 cDNA (78bp) is shorter than the corresponding region (101bp) of the Y_a cDNA pGTR261 (11) and much shorter than that in the Y_c cDNAs pGTB42 (20) and pGTR262 (10) (255bp). They share short, dispersed sequence homology, however. For example, nucleotides 756 to 797 of pGTH1 (Figure 2) are ~63% identical to the pGTR261 sequence (nucleotides 734 to 775) base-for-base without any deletion or insertion and nucleotides 761 to 794 of pGTH1 are ~56% identical to the pGTB42 sequence (nucleotides 807 to 840) base-for-base without any deletion or insertion.

Nucleotide sequence homologies between pGTH1 and other GST cDNAs (Figure 3).

We have made a comparison of the pGTH1 cDNA sequence with all the GST cDNA sequences available in the literature from rat and corn origins

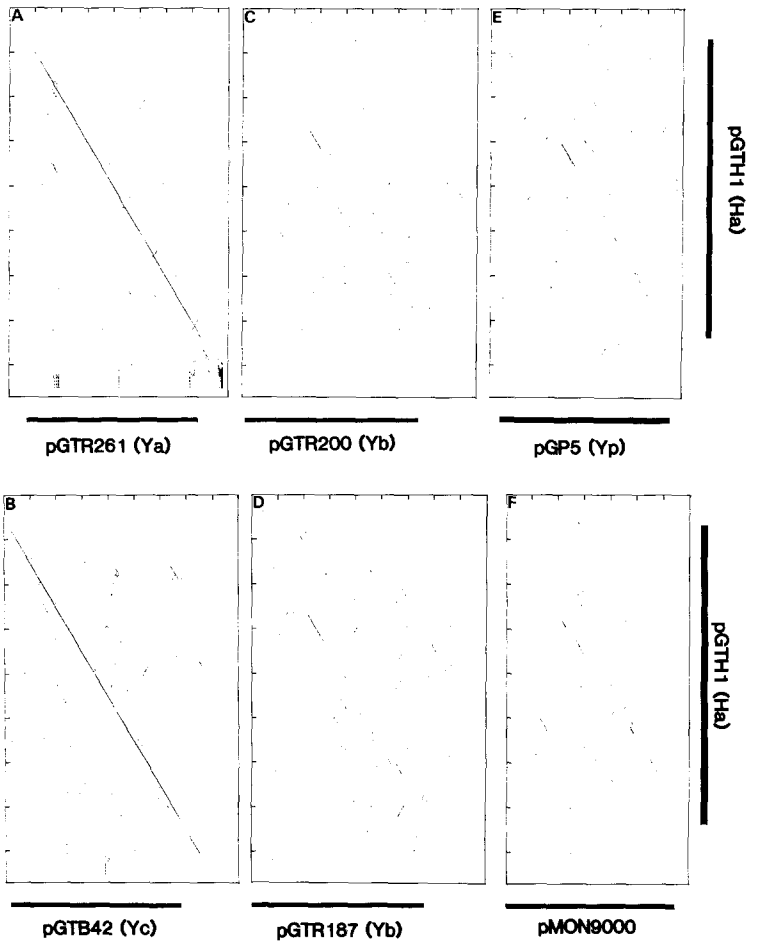

Figure 3. Nucleotide sequence homology between pGTH1 cDNA and GST cDNAs from rat and corn origins. A: pGTR261, Y_a (11); B: pGTB42, Y_c (20); C: pGTR200, Y_b (22); D: pGTR187, Y_b (21); E: pGP5, Y_p (26); and F: pMON9000, a GST cDNA sequence from corn (28). Analyses were carried out by a personal computer using the Dot Matrix program of Zweig (18) with a parameter of 18 matches in a span of 24 nucleotides. The solid bars on the X and Y axis represent coding region sequences of each cDNA.

(10,11,20-22,26-28). In addition to the homology with Y_a and Y_c subunit cDNA of rat liver GSTs the H_a subunit cDNA showed a region of considerable homology with the anionic (pGTR187) and a basic Y_b subunit (pGTR200) near the N-terminal region with 62% and 69% homology, respectively (Figure 4). This is consistent with the DNA blot hybridization signals we observed before (9). This conserved region happens to encode amino acid residues 70 to 95, which are highly conserved throughout the GST supergene family (Figure 5). This region of pGTH1 cDNA has almost equal extent of homologies (59%) with the rat placental subunit Y_p cDNA pGP5 (Figure 4) (26). More interestingly, the human GST sequence has detectable homology (>50%) to a corn GST cDNA sequence

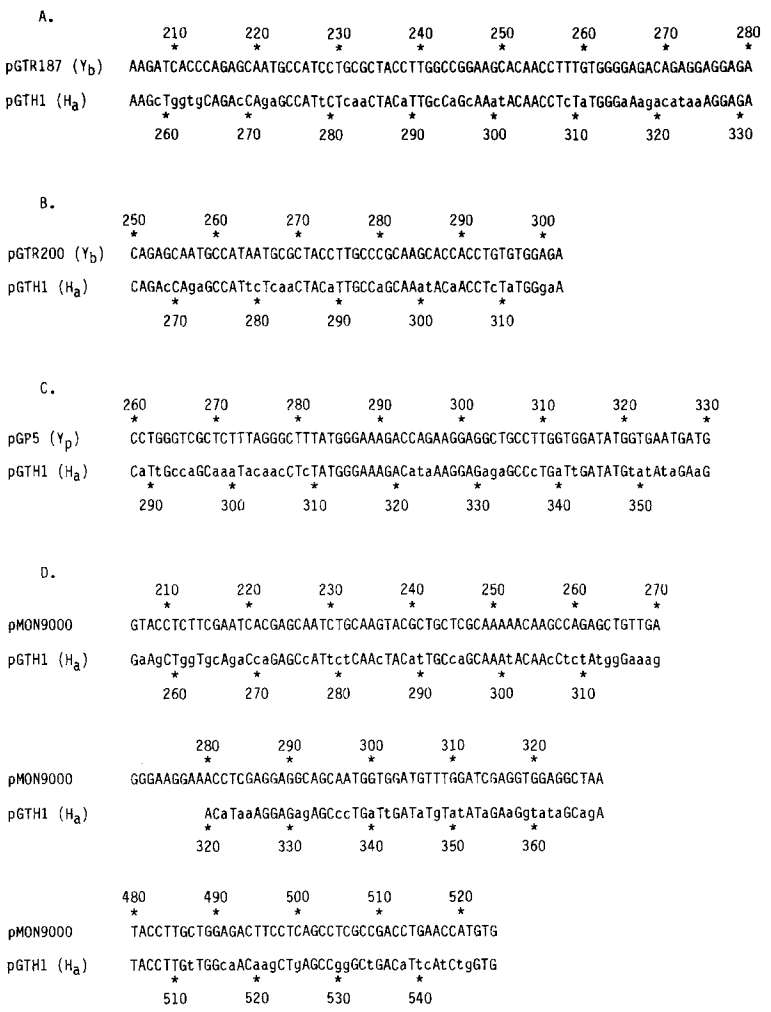

Figure 4. Nucleotide sequence match between pGTH1 cDNA and rat Y_b, Y_p and corn GST cDNA sequences. Regions listed here have more than 50% sequence homologies. Nucleotides are numbered according to Figure 2 (for pGTH1) and the original publications: A, pGTR187 (21); B, pGTR200 (22); C, pGP5 (26); D, pMON9000 (28). Non-homologous nucleotides in pGTH1 are represented in lower case letters.

pMON9000 (28) in three regions (Figure 4, 52%, 55% and 64% respectively), possibly reflecting the common recognition of the essential substrate GSH and/or the hydrophobic nature of some part of the xenobiotic substrate.

Conserved amino acids between the H_a subunit 1 and rat GST subunits.

The rat GSTs are encoded by a supergene family (21,22). The Y_b subunit cDNAs (e.g. pGTR187 and pGTR200) did not hybridize to the Y_a or Y_c subunit cDNAs in Southern hybridizations (21,22). They do have overlapping substrate specificities against some common substrates such as 1-chloro-2, 4-dinitrobenzene (CDNB), however (23). After careful comparison of their coding sequences, significant conservation of the amino acid residues were

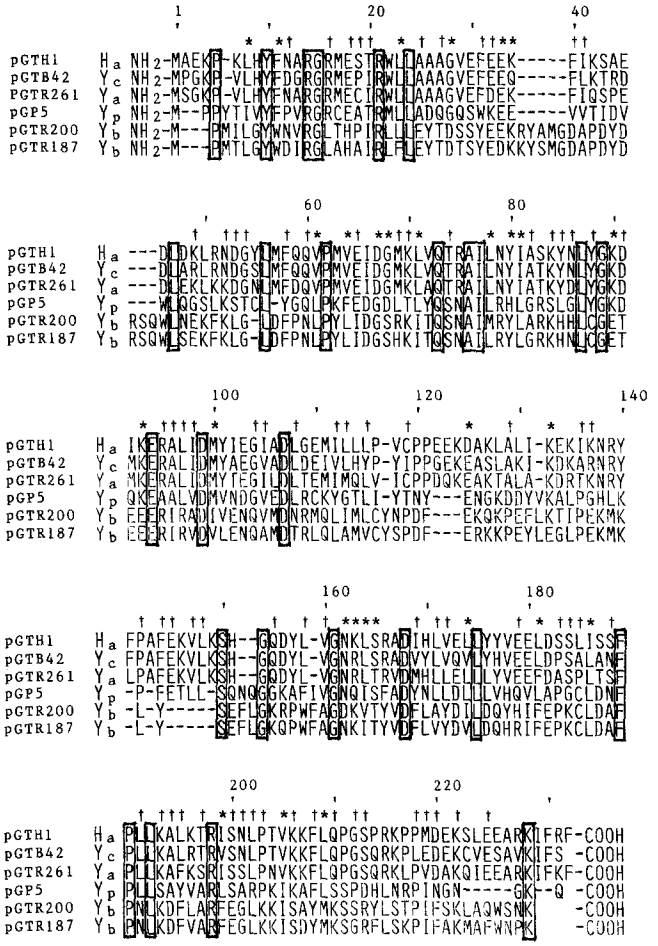

Figure 5. Amino acid sequence conservation between the human GST H_a subunit 1 and the rat GST supergene family. Identical amino acids are in boxes. Amino acids of the same groups (i.e., small polar, S, G, D, N; large polar, E, Q, K, R; intermediate polarity, Y, H, W; large nonpolar, F, M. L, I, V; and small nonpolar, C, P, A, T) are labelled with the asterisks, *. The daggers, †, denote amino acids of the same group in three out of the five subunit classes, Y_a, Y_b, Y_c, Y_p, and H_a.

identified (~29%) (21,22). We found that the H_a subunit 1 sequence has the majority of these conserved amino acids in rat GSTs. The tabulated GST sequences in Figure 5 showed 27 identical amino acid residues and 29 others in the same side chain groupings (24,25) for a total of 25.7% homology between this H_a subunit and various rat GST subunits.

The amino acid positions 70 to 95 in the GST supergene family (Figure 5) may be of unusual signficance since they are conserved exceedingly well throughout evolution. It is very likely that they are important residues for GSH binding and/or substrate binding and catalysis. They should provide a reasonable target for mutational studies in vitro.

Table 1. Divergence for amino-acid replacement sites in human and rat GSTS

Pairwise comparison	% Divergence replacement sites
Human H_a (pGTH1) - rat Y_a (pGTR261)	14.7
Human H_a (pGTH1) - rat Y_c (pGTB42)	15.1
Rat Y_a (pGTR261) - rat Y_c (pGTB42)	20.7
Rat anionic Y_b (pGTR187) - rat basic Y_b (pGTR200)	12.1

Calculations were based on the method of Perler et al. (19) and data are from references 11, 20, 21, and 22.

The evolutionary relationship between pGTH1 (H_a), pGTR261 (Y_a) and pGTB42 (Y_c) is an intriguing one because the H_a cDNA sequence is approximately equally homologous to the rat Y_a and Y_c sequences (Figure 2). We did a quantitative comparison of these sequences with the method of Perler et al. (19) to analyze percentages of amino-acid replacement mutations in these three coding sequences. The result is presented in Table 1. It is surprising to find that the Y_a (pGTR261) and Y_c (pGTB42) sequences are more distantly related to each other (20.7% divergence) than are the two interspecies subunit comparisons (14.7% for H_a vs Y_a and 15.1% for H_a vs Y_c). As a reference, we compared the percentage divergence for amino-acid replacement sites between two rat Y_b subunit sequences, pGTR200 and pGTR187 (21,22). The result showed a 12.1% divergence, which is comparable to the divergence of rabbit and mouse β globins, but well beyond the allelic replacements in β globins and preproinsulins (19).

The higher percentage of replacement site divergence between the rat Y_a and Y_c genes is consistent with the notion that rates of nucleotide substitutions is higher in rodents than in man (29). The calculated values in Table 1 suggest that there is likely a duplication of the Y_a/Y_c ancestral gene in the rat genome that did not occur in the human genome. The duplicated genes of rats have since diverged from each other at a rate 1.3 times faster than in man (30). We cannot rule out the possibility that other H_a cDNAs from liver or other human tissues may be more homologous to either the Y_a or the Y_c sequences, however. A more thorough analysis of GST gene evolution, however, awaits more sequence information from many other species and gene families.

ACKNOWLEDGEMENTS

We thank David Rhoads for technical assistance in the initial characterization of pGTH1, Jean Lai and Jeff DeJong for helping in the computer analyses and Eileen McConnell for typing the manuscript. This

research project has been supported by a United States Public Health Service Grant ES02678. C.-P. D. Tu is the recipient of a Research Career Development Award (K04 ES00140) from the United Stated Public Health Service.

REFERENCES

1. Vander Jagt, D. L, Hunsaker, L. A., Garcia, K. B., and Royer, R. E. (1985) J. Biol. Chem. 260, 11603-11610.

2. Stockman, P. K., Beckett, G. J., and Hayes, J. D. (1985) Biochem. J. 227, 457-465.

3. Dao, D. D., Partridge, C. A., Kurosky, A., and Awasthi, Y. C. (1984) Biochem. J. 221, 33-41.

4. Warholm, M., Guthenberg, C., and Mannervik, B. (1983) Biochemistry 22, 3610-3617.

5. Awasthi, Y. C., Dao, D. D., and Saneto, R. P. (1980) Biochem. J. 191, 1-10.

6. Polidoro, G., Di Ilio, C., Del Boccio, G., Zulli, P., and Federici, G. (1980) Biochem. Pharmacol. 29, 1677-1680.

7. Marcus, C. J., Habig, W. H., and Jakoby, W. B. (1978) Arch. Biochem. Biophys. 188, 287-293.

8. Kamisaka, K., Habig, W. H., Ketley, J. N., Arias, I. M., and Jakoby, W. B. (1975) Eur. J. Biochem. 60, 153-161.

9. Tu, C.-P. D., Matsushima, A., Li, N., Rhoads, D. M., Srikumar, K., Reddy, A. P., and Reddy, C. C. (1986) J. Biol. Chem. 261, 9540-9545.

10. Tu, C.-P. D., Lai, H.-C. J., Li, N., Weiss, M. J., and Reddy, C. C. (1984) J. Biol. Chem. 259, 9434-9439.

11. Lai, H.-C. J., Li, N., Weiss, M. J., Reddy, C. C., and Tu, C.-P. D. (1984) J. Biol. Chem. 259, 5536-5542.

12. Kupersztock, Y. M. and Helinski, D. R. (1973) Biochem. Biophys. Res. Commun. 54, 1451-1459.

13. Wu, R., Jay, E., and Roychoudhury, R. (1976) Methods Cancer Res. 12, 87-176.

14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

15. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.

16. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. (USA) 74, 5463-5467.

17. Messing, J. (1983) Methods Enzymol. 101, 20-78.

18. Zweig, S. E. (1984) Nucleic Acids Res. 12, 767-776.

19. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell 20, 555-566.

20. Telakowski-Hopkins, C. A., Rodkey, J. A., Bennett, C. D., Lu, A. Y. H., and Pickett, C. B. (1985) J. Biol. Chem. 260, 5820-5825.

21. Lai, H.-C. J. and Tu, C.-P. D. (1986) J. Biol. Chem. 261, 13793-13799.

22. Lai, H.-C. J., Grove, G., and Tu, C.-P. D. (1986) Nucleic Acids Res. 14, 6101-6114.

23. Tu, C.-P. D. and Reddy, C. C. (1985) J. Biol. Chem. 260, 9961-9964.

24. Swanson, R. (1984) Bull. Math. Biol. 42, 187-203.

25. Doolitle, R. F. (1985) Sci. Am. 253, 88-99.

26. Suguoka, Y., Kano, T., Okuda, A., Sakai, M., Kitagawa, T., and Muramatsu, M. (1985) Nucleic Acids Res. 13, 6049-6057.

27. Ding, G. J.-F., Lu, A. Y. H., and Pickett, C. B. (1985) J. Biol. Chem. 260, 13268-13271.

28. Shah, D. M., Hironaka, C. M., Wiegand, R. C., Harding, E. I., Krivi, G. G., and Tiemeier, D. C. (1986) Plant Mol. Biol. 6, 203-211.

29. Wu, C.-I. and Li, W.-H. (1985) Proc. Natl. Acad. Sci. (USA) 82, 1741-1745.