Volume 12 Number 14 1984 Nucleic Acids Research

Human purine nucleoside phosphorylase cDNA sequence and genomic clone characterization

Steven R.Williams*, Judy M.Goddard** and David W.Martin,Jr.*

Howard Hughes Medical Institute, Department of Medicine and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, and Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

Received 29 May 1984; Accepted 17 June 1984

ABSTRACT

The isolation of a cDNA clone containing the complete coding region for human purine nucleoside phosphorylase (PNP) has been described previously. In this report we present the nucleotide sequence of this cDNA clone and compare the derived amino acid sequence, encoding a protein of 32 kilodaltons, with the published amino acid composition.

Using a fragment of the cDNA clone as a probe, human PNP genomic clones from a bacteriophage lambda library have been isolated and the structural organization of the wild type PNP gene determined.

INTRODUCTION

Purine nucleoside phosphorylase (PNP; EC 2.4.2.1) catalyzes the reversible phosphorolysis of the purine nucleosides and deoxynucleosides inosine, guanosine, deoxyinosine and deoxyguanosine. Homozygous deficiency of the autosomal gene for PNP results in a severe T-cell immunodeficiency in children (1,2). Affected individuals accumulate all four substrates of PNP in their plasma and urine (3), but only one of these, deoxyguanosine, can be metabolized intracellularly where it accumulates as the triphosphate. Accumulated dGTP has been shown to exhibit toxicity in a cultured murine T-lymphoma model system by feedback inhibition of ribonucleotide reductase (4). This inhibition hinders the reduction of CDP, resulting in depleted intracellular pools of dCTP (4). The depletion of dCTP inhibits DNA synthesis and could thereby account for the lymphocyte defect.

Human erythrocyte PNP has been demonstrated to be a trimer composed of identical subunits each with a molecular weight of approximately 30,000 (5,6). We have previously described the cloning of cDNA sequences containing the entire coding region of PNP from cultured human cells (7). In this report we present the nucleotide sequence of one of these clones and the translated amino acid sequence. The deduced amino acid composition compares favorably to the previously published amino acid content analysis (8).

Using a fragment of this cDNA clone as a probe we have isolated bacteriophage lambda clones containing human genomic PNP sequences. We present a map of the restriction enzyme cleavage sites in these clones, showing the organization of the major exons.

MATERIALS AND METHODS

DNA Sequencing

The isolation of clone pPNP1 was previously described (7). Restriction endonuclease cleavage sites in this plasmid were mapped and appropriate fragments were cloned into bacteriophage M13 vectors mp8 or mp9 (9). DNA sequencing was carried out by the dideoxynucleotide chain termination method (10). Sequence analysis was aided by computer manipulation of data.

Screening of the Human Genomic Library

The human genome library produced by partial digestion with AluI and HaeIII and cloned into bacteriophage lambda vector Charon 4A (11) was a generous gift of Dr. Richard Lawn. The library was screened using the *in situ* plaque hybridization technique (12).

Restriction Fragment Analysis

Phage DNA was isolated from 40 ml liquid cultures, digested with various restriction endonucleases and electrophoresed in 0.8% agarose gels. After transfer of DNA to nitrocellulose by blotting (13), hybridization was carried out with nick-translated ^{32}P-labeled probes (14).

Hybridization Conditions

Hybridizations for screening of the genomic library and for blots of restriction fragments electrophoresed in agarose gels were carried out under identical conditions. Filters were prehybridized at 42°C for 2 - 17 hours in 50% formamide, 0.1 g/L denatured salmon sperm DNA, 5 X SSC (1 X SSC = 0.15 M NaCl, 0.015 M Na citrate), 0.05 M sodium phosphate pH 6.8 and 5 X Denhardt's solution (15). Hybridizations were performed with the nick-translated ^{32}P-labeled probes at a concentration of 1 - 10 X 10^5 cpm/ml in prehybridization mix plus 10% dextran sulfate for 14 - 24 hours. The filters were then washed at 55°C - 60°C in 3 changes of 0.2 X SSC, 0.1% SDS for 30 minutes each and autoradiographed for 0.5 - 24 hours at -70°C with an intensifying screen.

Subcloning into Plasmids

DNA from lambda clones was digested with EcoRI and cloned into EcoRI-digested, dephosphorylated pUC13 (16). Miniscreen DNA was then

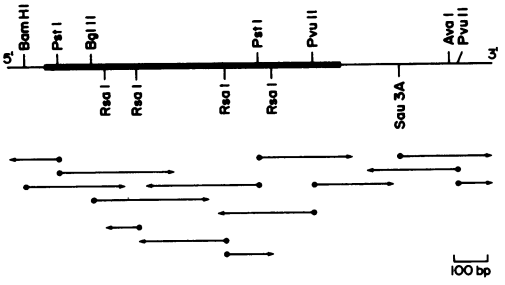

Figure 1. Map of restriction enzyme cleavage sites and sequencing strategy of pPNP1. Arrows indicate fragments sequenced and direction and extent of sequencing.

prepared from the transformants as described (17).

RESULTS

cDNA Sequence Analysis

Plasmid clone pPNP1 consists of a 1418 bp cDNA insert in the PstI site of pBR322. Each end of the insert is flanked by 20 - 22 dC/dG base pairs added during the cDNA cloning procedure (7). The sequencing strategy employed is illustrated in Figure 1. Arrows denote the fragments which were cloned and sequenced in the M13 system. Consensus sequences were obtained from overlapping fragments and both strands were sequenced over most of the molecule. The resulting sequence is shown in Figure 2. All three possible translational reading frames were examined and a single long open reading frame of 289 amino acid codons, corresponding to a molecular weight of 32,153 daltons, was found.

The sequence AATAAA is present 208 bp from the 3' end of the insert (nucleotides 1205-1210, Figure 2). Although this sequence is the same as the canonical polyadenylation signal (18), such signals are normally found 10 - 30 bp upstream from the poly(A) tract. No polyadenylate tract is present in pPNP1. Thus the authentic polyadenylation signal of the PNP message is probably not represented in pPNP1.

Table 1 compares the amino acid composition determined from the cDNA sequence with that determined from amino acid analysis of PNP protein purified from human erythrocytes (8). The amino acid composition data determined from the two methods are in close agreement, with the exception of the methionyl residues, which differ by a factor of two.

```
                                                                                                                    met glu
AACTGTGCGAACCAGACCCGGCAGCCTTGCTCAGTTCAGCATAGCGGAGCGGATCCGATCGGATCGGAGCACACCGGAGCAGGCTCATCGAGAAGGCGTCTGCGAGACC   ATG GAG
                                30                                                  60                                  90

asn gly tyr thr tyr glu asp tyr lys asn thr ala glu trp leu leu ser his thr lys his arg pro gln val ala ile ile cys gly
AAC GGA TAC ACC TAT GAA GAT TAT AAG AAC ACT GCA GAA TGG CTT CTG TCT CAT ACT AAG CAC CGA CCT CAA GTT GCA ATA ATC TGT GGT
   120                                 150                                 180

ser gly leu gly gly leu thr asp lys leu thr gln ala gln ile phe asp tyr ser glu ile pro asn phe pro arg ser thr val pro
TCT GGA TTA GGA GGT CTG ACT GAT AAA TTA ACT CAG GCC CAG ATC TTT GAC TAC AGT GAA ATC CCC AAC TTT CCT CGA AGT ACA GTG CCA
   210                                 240                                 270

gly his ala gly arg leu val phe gly phe leu asn gly arg ala cys val met met gln gly arg phe his met tyr glu gly tyr pro
GGT CAT GCT GGC CGA CTG GTG TTT GGG TTC CTG AAT GGC AGG GCC TGT GTG ATG ATG CAG GGC AGG TTC CAC ATG TAT GAA GGG TAC CCA
   300                                 330                                 360

leu trp lys val thr phe pro val arg val phe his leu leu gly val asp thr leu val val thr asn ala ala gly gly leu asn pro
CTC TGG AAG GTG ACA TTC CCA GTG AGG GTT TTC CAC CTT CTG GGT GTG GAC ACC CTG GTA GTC ACC AAT GCA GCA GGA GGG CTG AAC CCC
   390                                 420                                 450

lys phe glu val gly asp ile met leu ile arg asp his ile asn leu pro gly phe ser gly gln asn pro leu arg gly pro asn asp
AAG TTT GAG GTT GGA GAT ATC ATG CTG ATC CGT GAC CAT ATC AAC CTA CCT GGT TTC AGT GGT CAG AAC CCT CTC AGA GGG CCC AAT GAT
   480                                 510                                 540

glu arg phe gly asp arg phe pro ala met ser asp ala tyr asp arg thr met arg gln arg ala leu ser thr trp lys gln met gly
GAA AGG TTT GGA GAT CGT TTC CCT GCC ATG TCT GAT GCC TAC GAC CGG ACT ATG AGG CAG AGG GCT CTC AGT ACC TGG AAA CAA ATG GGG
   570                                 600                                 630

glu gln arg glu leu gln glu gly thr tyr val met val ala gly pro ser phe glu thr val ala glu cys arg val leu gln lys leu
GAG CAA CGT GAG CTA CAG GAA GGC ACC TAT GTG ATG GTG GCA GGC CCC AGC TTT GAG ACT GTG GCA GAA TGT CGT GTG CTG CAG AAG CTG
   660                                 690                                 720

gly ala asp ala val gly met ser thr val pro glu val ile val ala arg his cys gly leu arg val phe gly phe ser leu ile thr
GGA GCA GAC GCT GTT GGC ATG AGT ACA GTA CCA GAA GTT ATC GTT GCA CGG CAC TGT GGA CTT CGA GTC TTT GGC TTC TCA CTC ATC ACT
   750                                 780                                 810

asn lys val ile met asp tyr glu ser leu glu lys ala asn his glu glu val leu ala ala gly lys gln ala ala gln lys leu glu
AAC AAG GTC ATC ATG GAT TAT GAA AGC CTG GAG AAG GCC AAC CAT GAA GAA GTC TTA GCA GCT GGC AAA CAA GCT GCA CAG AAA TTG GAA
   840                                 870                                 900

gln phe val ser ile leu met ala ser ile pro leu pro asp lys ala ser OP
CAG TTT GTC TCC ATT CTT ATG GCC AGC ATT CCA CTC CCT GAC AAA GCC AGT TGA CCTGCCTTGGAGTCGTCTGGCATCTCCCACACAAGACCCAAGTAGCTG
   930                                 960                                 990                                1020

CTACCTTCTTTGGCCCCTTGCTGGAGTCATGTGCCTCTGTCCTTAGGTTGTAGCAGAAAGGAAAAGATTCCTGTCCTTCACCTTTCCCACTTTCTTCTACCAGACCCTTCTGGTGCCAGA
                  1050                                1080                                1110                               1140

TCCTCTTCTCAAAGCTGGGATTACAGGTGTGAGCATAGTGAGACCTTGGCGCTACAAAATAAAGCTGTTCTCATTCCTGTTCTTTCTTACACAAGAGCTGGAGCCCGTGCCCTACCACAC
                  1170                                1200                                1230                               1260

ATCTGTGGAGATGCCCAGGATTTGACTCGGGCCTTAGAACTTTGCATAGCAGCTGCTACTAGCTCTTTGAGATAATACATTCCGAGGGGCTCAGTTCTGCCTTATCTAAATCACCAGAGA
                  1290                                1320                                1350                               1380

CCAAACAAGGACTAATCCAATACCTCTTGGA
                  1410
```

Figure 2. Nucleotide sequence of the message strand of pPNP1. The numbers refer to the nucleotide above the last digit of the number. The deduced amino acid sequence is indicated.

Characterization of Genomic Clones

Two million plaques from the amplified human library were screened with the 1242 bp BamHI-AvaI probe fragment from pPNP1 (probe 7, Figure 3). Thirty-eight clones giving positive signals on duplicate filters were identified. Fourteen of these were plaque-purified and subjected to further analysis. EcoRI digestion of DNA from these fourteen clones revealed three distinct digestion patterns, and clones representing these different patterns were designated λPNP1, λPNP2 and λPNP4.

A map of restriction enzyme cleavage sites in these clones (Figure 3) was determined using the following strategy. EcoRI, BamHI and HindIII

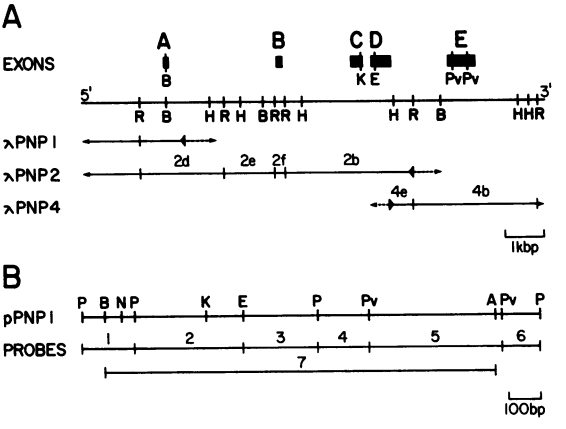

Figure 3.
(A) Restriction enzyme cleavage sites in cloned human genomic PNP. Exons are indicated by solid blocks on the top line. Restriction sites in these exons which correspond to sites in the cDNA are indicated. The second line presents a complete map of EcoRI, BamHI and HindIII sites in this region of the genome. The next three lines show the boundaries of lambda clones λPNP1, λPNP2 and λPNP4. Solid lines indicate inserted genomic sequences while dashed lines indicate lambda vector sequences, vertical slashes indicate EcoRI sites and wedges denote the EcoRI boundaries between vector and insert. Subcloned EcoRI fragments are indicated. These fragments are named according to their size; i.e., 2b is the second largest EcoRI fragment of λPNP2.
(B) cDNA probes used to analyze Southern blots of digested genomic clones. Restriction sites in pPNP1 used to isolate probes are shown, as well as those sites shown in the exons in Figure 3 (A) or discussed in the text. A = AvaI; B = BamHI; E = EcoRV; H = HindIII; K = KpnI; P = PstI; Pv = PvuII; N = NcoI; R = EcoRI.

digestion patterns were obtained for all three clones. These enzymes digest Charon 4A DNA in sites convenient for mapping inserted sequences. EcoRI fragments from λPNP2 and λPNP4 were subcloned into plasmid pUC13 and identified by restriction mapping and hybridization. Southern blots of the digested lambda clones and plasmid subclones were prepared and hybridized with the nick-translated 1242 bp cDNA probe (Probe 7, Figure 3) or with shorter probes consisting of small discrete regions of the cDNA (Probes 1-6, Figure 3). An example is shown in Figure 4.

Exons were mapped by digesting subcloned EcoRI fragments with various restriction enzymes, especially those enzymes which have one or two sites in the cDNA. These digested DNAs were separated by agarose gel electrophoresis, blotted, and hybridized with the 1242 bp cDNA probe or with one of the short probes. In this way, exon-containing sequences were mapped

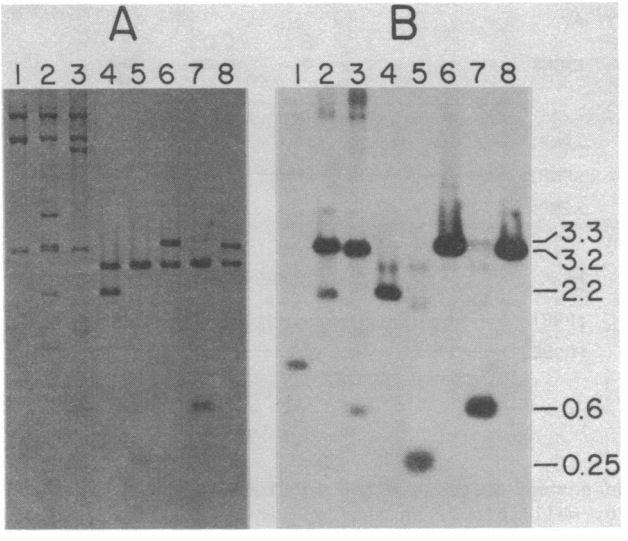

Figure 4. Southern blot analysis of PNP lambda clones and plasmid subclones.
DNA from PNP lambda clones (lanes 1-3) and plasmid subclones (lanes 4-8) was digested with EcoRI, electrophoresed in a 0.8% agarose gel, stained with ethidium bromide and photographed on a UV transilluminator. Panel A shows a negative image of the ethidium bromide fluorescence. DNA fragments were transferred to nitrocellulose, hybridized with the ^{32}P-labeled 1242 bp cDNA probe (probe 7, Figure 3) and autoradiographed (panel B). (1) λPNP1; (2) λPNP2; (3) λPNP4; (4) p2d; (5) p2f; (6) p2b; (7) p4e; (8) p4b. Size of DNA fragments (in kbp) is indicated.

to within a few hundred basepairs. Some restriction sites in the genomic clones were identified as corresponding to sites in the cDNA. The presence or absence of intervening sequences was determined by mapping the distances between these sites. Introns were also detected by the presence of restriction sites within exon-containing fragments which were not found in the cDNA.

Exon A occurs within the 2.2 kbp EcoRI fragment (subclone p2d) in λPNP2. A BamHI site 700 bp from the 5' end of this fragment corresponds to the BamHI site at position 51 in the cDNA. This BamHI site in the genomic clone is flanked by two NcoI sites which are about 200 bp apart. The 3' NcoI site corresponds to the NcoI site which contains the translational initiation codon in the cDNA. Exon A hybridizes with probe fragment 1 (Figure 3).

Exon B hybridizes with probe 2 and is contained in a 250 bp EcoRI fragment (subclone p2f).

Exon C, which also hybridizes with probe 2, is found close to exon D, which hybridizes with probes 2 and 3 and weakly with probe 4, in the 3.3 kbp Eco RI fragment (subclone p2b) in λPNP2. A KpnI site in exon C and an EcoRV site in exon D correspond to sites in the cDNA. These sites are separated by about 350 bp in the genomic clone but only 113 bp in the cDNA.

Exon E, which hybridizes with probes 4, 5 and 6, is contained in a 1.1 kbp BamHI-PstI fragment within the 3.2 kbp EcoRI fragment (subclone p4b) in λPNP4. There appear to be no introns in the region of the gene corresponding to cDNA probe 5, since the PvuII sites which form the boundaries of this fragment are separated by about 400 bp in both the genomic and cDNA clones. It was not determined whether the sequences in λPNP4 corresponding to probes 4 and 6 are immediately adjacent to this PvuII fragment or if they are separated by short intervening sequences.

DISCUSSION

This paper describes the nucleotide sequence of a cDNA clone which contains the entire coding region of the human purine nucleoside phosphorylase message. Active PNP protein has been expressed from this cDNA clone in PNP-deficient *E. coli* cells in a genetic complementation assay (7). In addition, a mammalian expression vector construction containing this cDNA expresses PNP enzymatic activity in both mouse and hamster cell lines (19).

Although the cDNA clone pPNP1 contains the entire coding region, it does not contain a complement of the entire messenger RNA. As previously reported, the PNP mRNA is approximately 1700 bases long (7), while the sequence of the cDNA clone described herein is approximately 300 nucleotides smaller. In addition, this cDNA clone does not terminate in a polyadenylate tract, although oligo-d(T) chromatography enriches for the PNP mRNA (7). Of note is the existence of a consensus poly(A) addition signal (18) within the cDNA sequence but at a site that is 200 nucleotides from the most 3' nucleotide of the molecule. This suggests that the authentic poly(A) addition signal is still farther 3' than the sequences that have been cloned in pPNP1. Although there are reports of messenger RNAs with multiple poly(A) addition signals and multiple processing sites (20,21), there appears at least in HeLa cells to be only a single functional polyadenylation site for PNP mRNA, since the length of the mRNA from these cells is homogeneous.

The cloned cDNA for PNP has permitted the isolation and characteri-

zation of human PNP genomic clones. Exon-containing sequences span a region of approximately 9 kbp in this gene with a minimum of four intervening sequences varying in length from about 200 bp to 3 kbp. λPNP1 contains at least 10 kbp 5' to the first exon and thus almost certainly contains the promoter region for human PNP. This gene is expressed ubiquitously in humans although there are some tissue-specific differences in levels of expression (22). Thus the PNP promoter region may be useful as a regulatory signal allowing expression of downstream genes in many or all tissues. This promoter is probably an ideal candidate for use in constructing vectors for possible gene replacement therapy in cases of PNP deficiency. Studies are now proceeding in our laboratory to express human PNP in mouse bone marrow cells as a model for gene replacement therapy in humans afflicted with the immunodeficiency disease which results from the absence of PNP activity in T-lymphocytes.

These genomic probes will be useful in understanding at the molecular level the pathogenesis of PNP deficiency disease. A nonfunctional PNP molecule in one patient seems to have an internal insertion of several amino acid residues (23,24). The availability of the wild type genomic clones will permit comparative studies of the genetic defect in this PNP-deficient patient and in others (25-29). In addition, the amino acid sequence of PNP derived from the cDNA sequence will aid those laboratories studying the crystallographic structure of the protein (30) and allow comparative studies of the functional domains predicted from the amino acid sequences of other nucleoside and nucleotide metabolizing enzymes or binding proteins (31).

ACKNOWLEDGMENTS

We thank Jane Gitschier for valuable advice. This work was supported in part by USPHS grant number AI-17435 from the National Institutes of Health. DWM was an Investigator, Howard Hughes Medical Institute.

* Current address: Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080.
** Current address: Department of Biology, University of Utah, Salt Lake City, Utah 84112.

REFERENCES

1. Giblett, E.R., Ammann, A.J., Wara, D.W., Sandman, R. and Diamond, L.K. (1975) Lancet *i*, 1010-1013.
2. Martin, D.W., Jr. and Gelfand, E.W. (1981) Ann. Rev. Biochem. 50, 845-877.
3. Cohen, A., Doyle, D., Martin, D.W., Jr. and Ammann, A.J. (1976) New Engl. J. Med. 295, 1449-1454.

4. Ullman, B., Gudas, L.J., Clift, S.M. and Martin, D.W., Jr. (1979) Proc. Natl. Acad. Sci. USA 76, 1074-1078.
5. Agarwal, R.P. and Parks, R.E., Jr. (1969) J. Biol. Chem. 244, 644-647.
6. Zannis, V., Doyle, D. and Martin, D.W., Jr. (1978) J. Biol. Chem. 253, 504-510.
7. Goddard, J.M., Caput, D., Williams, S.R. and Martin, D.W., Jr. (1983) Proc. Natl. Acad. Sci. USA 80, 4281-4285.
8. Parks, R.E., Jr. and Agarwal, R.P. (1972) in The Enzymes, Boyer, P.D., Ed., 3rd ed. Vol. 7, pp. 483-514, Academic Press, New York.
9. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
11. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) Cell 15, 1157-1174.
12. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
13. Southern, E.M. (1975) J. Molec. Biol. 98, 503-517.
14. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Natl. Acad. Sci. USA 72, 1184-1188.
15. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.
16. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
17. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
18. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
19. McIvor, R.S. and Martin, D.W., Jr. (1984) manuscript in preparation.
20. Moore, C.L. and Sharp, P.A. (1984) Cell 36, 581-591.
21. Chin, D.J., Gil, G., Russell, D.W., Liscum, L., Luskey, K.L., Basu, S.K., Okayama, H., Berg, P., Goldstein, J.L. and Brown, M.S. (1984) Nature 308, 613-617.
22. Carson, D.A., Kaye, J. and Seegmiller, J.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5677-5681.
23. Gudas, L.J., Zannis, V.I., Clift, S.M., Ammann, A.J., Staal, G.E.J. and Martin, D.W., Jr. (1978) J. Biol. Chem. 253, 8916-8924.
24. McRoberts, J.A. and Martin, D.W., Jr. (1980) J. Biol. Chem. 255, 5605-5615.
25. Stoop, J.W., Eijsvoogel, V.P., Zegers, B.J.M., Blok-Schut, B., van Bekkum, D.W. and Ballieux, R.E. (1976) Clin. Immunol. Immunopathol. 6, 289-298.
26. Biggar, W.D., Giblett, E.R., Ozere, R.L. and Grover, B.D. (1978) J. Pediatr. 92, 354-357.
27. Carapella-De Luca, E., Aiuti, F., Lucarelli, P., Bruni, L., Baroni, C.D., Imperato, C., Roos, D. and Astaldi, A. (1978) J. Pediatr. 93, 1000-1003.
28. Rich, K., Arnold, W., Fox, I. and Palella, T. (1978) Pediatr. Res. 12, 485.
29. Siegenbeek van Heukelom, L.H., Staal, G.E.J., Stoop, J.W. and Zegers, B.J.M. (1976) Clinica Chimica Acta 72, 117-124.
30. Cook, W.J., Ealick, S.E., Bugg, C.E., Stoeckler, J.D. and Parks, R.E., Jr. (1981) J. Biol. Chem. 256, 4079-4080.
31. Argos, P., Hanei, M., Wilson, J.M. and Kelley, W.N. (1983) J. Biol. Chem. 258, 6450-6457.