
Molecular cloning and complete primary sequence of human erythrocyte porphobilinogen deaminase

N.Raich, P.H.Romeo*, A.Dubart, D.Beaupain, M.Cohen-Solal and M.Goossens

Unité INSERM U.91, Hôpital Henri Mondor, 94010 Creteil, Cedex, France

Received 16 June 1986; Accepted 30 June 1986

ABSTRACT

We have cloned and sequenced a cDNA clone coding for human erythrocyte porphobilinogen deaminase. It encompasses the translated region, part of the 5' and the 3' untranslated regions. The deduced 344 amino acid sequence is consistent with the molecular weight and the partial amino-acid sequence of the NH₂ terminal of the purified erythrocyte enzyme. Southern analysis of human genomic DNA shows that its gene is present as a single copy in the human genome and Northern analysis demonstrates the presence of a single size species of mRNA in erythroid and non-erythroid tissues and in several cultured cell lines. Quantitative determinations indicate that the amount of PBG-D mRNA is modulated both by the erythroid nature of the tissue and by cell proliferation, probably at the transcriptional level.

INTRODUCTION

Porphobilinogen deaminase (PBG-D, EC 4-3-1.8) is the third enzyme of the biosynthetic pathway leading to the production of heme. It catalyses the head to tail condensation of four molecules of the monopyrrole porphobilinogen, to form the linear tetrapyrrole, hydroxymethylbilane which is then converted by uroporphyrinogen III synthase to uroporphyrinogen III (1). In humans, deficiency in its activity is responsible for a dominant hereditary disease: Acute Intermittent Porphyria (A.I.P.) (2). Studies using antibodies specifically directed against human erythrocyte PBG-D show that inactive cross reacting immunological material is present in about 20% of patients with this disorder but absent from the remainder (3). Therefore Acute Intermittent Porphyria appears to be, at the molecular level, a heterogeneous disorder.

Although widely distributed in tissues, the enzymes of the heme biosynthetic pathway are particularly active in the liver where heme synthesis is regulated mainly by the activity of δ -aminolevulinic acid synthase (4) and in bone marrow, where the activity of all the enzymes of the pathway progressively augment during erythroid differentiation. Thus the control of

heme biosynthesis in erythroid cells and in hepatocytes appears to be different and a recent report (5) suggests that Porphobilinogen Deaminase may be the primary enzyme controlling heme synthesis during erythropoietin induced erythroid differentiation. Furthermore, Grandchamp et al. (6) have shown that in Mouse Erythroleukemia Cells (M.E.L.), capable of undergoing *in vitro* erythroid differentiation, induction by DMSO is followed by an increase in the copy number of Porphobilinogen Deaminase mRNA. This process precedes accumulation of α and β globin mRNAs, indicating a concerted and time dependent regulation.

This paper reports the cloning and analysis of a cDNA sequence complementary to human erythrocyte PBG-D mRNA. It contains a reading frame of 1029 bp which encodes for 344 amino-acids, a 5' non coding region of 81 bp and a complete 3' non coding region of 266 bp excluding the poly(A) tail. Using this cDNA as a hybridization probe, we confirm that PBG-D is encoded by a single copy gene per haploid genome and show by Northern blot analysis that there is only one size species of mRNA in erythroid and non-erythroid tissues and in cultured cell lines. Furthermore we demonstrate that the concentration of PBG-D mRNA is modulated in tissues undergoing erythroid differentiation in a manner that is different to that in immortalized cell lines where the PBG-D gene seems to have the same rate of transcription irrespective of the cell phenotype.

MATERIALS AND METHODS

General procedures

A previous communication (7) described the methods used for *in vitro* translation in a messenger-dependent rabbit reticulocyte cell-free system, PBG-D immunoprecipitation from *in vitro* translation products, and SDS-polyacrylamide gel electrophoresis.

Cell culture

The different cell lines studied were a gift from Dr. U. Testa. Cells were grown in suspension culture in a modified RPMI 1640 medium containing 10% heat inactivated newborn fetal calf serum plus 2% human serum, in a humidified atmosphere with 5% CO₂. All cell lines were grown exponentially to a density of 10⁵ to 10⁶ cells per ml.

Plasmid isolation, insert purification and nick translation

Plasmids containing rat or human PBG-D cDNA inserts were prepared from chloramphenicol treated bacteria by thermal denaturation and sedimentation of chromosomal DNA (8), followed by treatment with RNAase and proteinase K and extraction with buffered phenol. The recombinant plasmids were further

purified by the acid phenol method (9). The cDNAs were excised from the recombinant plasmid, isolated by agarose gel electrophoresis, electroeluted and nick-translated as described (10).

Preparation and fractionation of poly(A⁺) RNA

Total RNA was extracted by the LiCl method (11) from the spleen, removed for therapeutic reasons, of a child with β -thalassemia major. Total RNA from the cell lines was prepared by extraction with the proteinase K-SDS procedure (12). Poly(A)⁺ RNA was then selected by chromatography on oligo dT-cellulose (type T3, Collaborative Research, Waltham, MA) (13). Messenger RNA obtained from the spleen was fractionated by preparative gel electrophoresis (14) and size enriched fractions containing the PBG-D sequences were pooled and ethanol precipitated.

cDNA cloning and analysis

Synthesis of double stranded cDNA complementary to enriched poly (A⁺) RNA was accomplished as described by Wickens et al. (15). Double stranded cDNAs longer than 800 bp were purified by polyacrylamide gel electrophoresis followed by electroelution. They were then inserted in the PstI site of pBR 322 using the homopolymeric tailing and hybridization method (16). The resulting hybrid molecules were used to transform E. coli strain MC1061 which was rendered competent for uptake of plasmids (17). Recombinant clones were stored frozen on nitrocellulose filters after high density plating (18).

Clones were screened with ³²P labelled cDNAs complementary to rat mRNA PBG-D using the colony hybridization method of Grunstein and Hogness (19) as modified by Thayer (20). After hybridization and washing, positive colonies were visualized by exposing filters to Kodak AR5 Xray film at -80°C.

For further characterization, plasmids from positive colonies were isolated from 5 ml overnight cultures by the boiling method, digested with PstI and electrophoresed in 1% agarose gel. The clone PBG-DH4 which contains the largest insert was further characterized and used as labelled nucleic acid hybridization probe.

Preparation and blot analysis of DNA

Human leucocyte DNA was isolated (21) and digested with different restriction enzymes (Boehringer Mannheim). Following digestion, the DNA was ethanol precipitated, resuspended and run on a 1% agarose gel. The gel was blotted onto a nitrocellulose filter and hybridized with the radioactive labelled probe at 42° C in 50% formamide for 24 hours. The filter was

washed extensively and autoradiographed at -80°C for 48 hours with intensifying screens.

RNA analysis

RNA denaturation, fractionation by electrophoresis in formaldehyde-agarose gels and transfer to nitrocellulose filter were done as described (22). Hybridization and washing were performed according to Thomas (23) and autoradiograms were scanned with a densitometer.

In vitro transcription and hybridization

Nuclei were extracted from cell lines as described (24). Isolated nuclei were then resuspended in 20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSF, 50 % glycerol and stored at -70°C . Nuclear transcription, isolation of ^{32}P -labelled RNA and hybridization to filter-bound DNA were done as described (25). To measure the relative rate of transcription, the following cloned DNAs were immobilized on "gene screen plus" (New Engl. Nuclear) as probes: i) PBG-DH₄ cDNA which represent the DNA complementary to PBG-D mRNA. This cDNA recognizes a unique gene by Southern hybridization analysis, ii) pBR322, iii) plasmid-containing sequences coding for 28S ribosomal RNA. Non specific binding to vector DNA was less than 50 cpm.

DNA sequence analysis

The chemical modification method of Maxam and Gilbert (26) was used for all the DNA sequence determinations. After digestion with the appropriate restriction endonuclease, the plasmid was labelled either at the 5' end with T₄ polynucleotide kinase and ($\gamma^{32}\text{P}$) ATP, or at the 3' end with DNA polymerase I (Klenow fragment) and ($\alpha^{32}\text{P}$)dATP or terminal transferase and ($\alpha^{32}\text{P}$)ddATP. After digestion with a second restriction enzyme or strand separation, single end-labelled fragments were isolated from polyacrylamide gels for DNA sequence analysis.

RESULTS AND DISCUSSION

Identification of PBG-D cDNA clones

As a source of RNA we selected the spleen, removed for therapeutic purpose, from a patient with β thalassemia major. This organ contained erythropoietic islands and PBG-D mRNA represented about 0.05% of the total mRNA. The mRNA, enriched 10 fold in PBG-D sequence by size fractionation, was used to construct a cDNA library. Screening of this library with the specific rat PBG-D cDNA obtained previously (7) allowed us to isolate several human PBG-D clones. Plasmid DNA was then prepared and the size of the

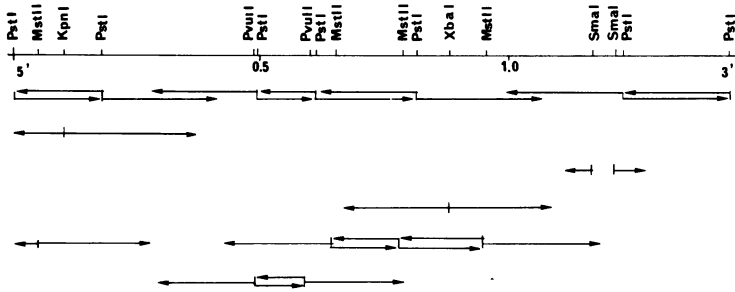


Fig. 1 : Strategy for sequencing human porphobilinogen deaminase cDNA PBG-DH4. The top line is a restriction map for the entire cDNA sequence including the two Pst I cloning sites. The map numbering proceeds in a 5' to 3' direction. The arrows below the map indicate the extent and direction of sequences determined. Each fragment was sequenced at least two times.

inserts examined by agarose gel electrophoresis. The largest PBG-D cDNA, PBG-DH 4, contains 1,450 bp. Since the size of PBG-D mRNA is 1,600 bp (see below), this clone seemed likely to correspond to most, if not all, the coding sequence and was used for further analysis.

Nucleotide sequence of human erythrocyte porphobilinogen deaminase cDNA

The restriction endonuclease map of the human cDNA insert and the strategy for its sequencing are shown in Fig. 1. The nucleotide sequence analysis revealed that PBG-DH 4 contains a coding sequence of 1,038 bp followed by a sequence corresponding to the mRNA 3' terminal poly(A) tail. The PBG-DH 4 recombinant contains a single large open reading frame that starts with an ATG codon at position 82 and extends to position 1116 (Fig. 2). The 5' non-coding region is 81 bases long whereas there are 267 bases of 3' non-coding region (excluding the poly(A) tail) containing the consensus poly(A) addition signal, AATAAA, starting at position 1161 and preceding the poly(A) tail by 14 bases.

Two findings indicate that the ATG codon located at position 82 of the cDNA is the initiator methionine codon. First, two stop codons are found in phase upstream from this putative initiator codon at position 26 and 49. Second, Fig. 2 shows that the sequence of the 19 amino acid residues adjacent to the NH₂ terminal residue of PBG-D purified from human erythrocytes (27) (unpublished data from Dr. G.H. Elder) is identical to the sequence deduced from the nucleotide sequence downstream from the methionine codon at position 82.

The ATG translation start codon lies within the sequence

methionine is included, the reading frame encodes 344 amino acids and accounts for a protein having a molecular weight of 37,627 daltons in good agreement with that estimated by SDS-polyacrylamide gel electrophoresis (37,000 - 44,500) (27, 30). All these results confirm that the deduced amino acid sequence represents the actual sequence of human erythrocyte porphobilinogen deaminase. Its deduced amino acid composition is shown in Table I and disagrees with the one proposed by Anderson and Desnick (30) who found no methionine or tryptophan, only one tyrosine and three histidines (Table I). Although this discrepancy can be easily explained for methionine and tryptophane, the low number of histidines and tyrosines found might be related to the presence of traces of sodium azide eluting with the protein during the G100 gel filtration (31).

It has been suggested that human PBG-D has one catalytic site and two substrate-binding sites (3). Our data were analyzed by computer programs but did not show any internal duplication in either the nucleotide or the amino acid sequences, indicating that the enzyme may in fact have only one binding site.

Knowledge of the complete sequence of the coding portion of human erythrocyte PBG-D cDNA will be useful for further characterization of the molecular defects responsible for AIP. As there is genetic heterogeneity in this disease [some patients with subnormal enzyme activity display normal immunoreactivity with anti porphobilinogen deaminase antibodies whereas others display a corresponding decrease in immunoreactivity (32)], the establishment of the nucleotide sequence of the normal gene and of the amino acid sequence of the normal protein should make it possible to define structural mutations which underly different variants found in patients with AIP.

Genomic Southern analysis of human porphobilinogen deaminase gene

Knowing that PBG-DH 4 contains the whole coding sequence of PBG-D as well as part of the 5' and the 3' untranslated region, we used it as a probe to explore the distribution of porphobilinogen deaminase coding sequences within the human genome. A single band was observed with the PBG-D probe in the Eco-RI and Hind III digests of human genomic DNA (Fig. 3). Since the PBG-D cDNA used as a hybridization probe contains no internal cleavage sites for Hind III and EcoRI, each of the single hybridizing restriction fragments obtained with these enzymes is likely to contain a single copy of the gene coding for human porphobilinogen deaminase. Following digestion of human DNA with Bam HI or Pst I, for which there are

TABLE I
AMINO ACID COMPOSITION OF HUMAN PORPHOBILINOGEN DEAMINASE

AMINO ACID	DEDUCED FROM PBG-DH ⁴ SEQUENCE	RESIDUES/37000 g PROTEIN (30)
Cys	4	4
Lys	19	17
His	13	3
Arg	21	10
Asn	10	Asn + Asp = 36
Asp	19 } 29	
Thr	20	18
Ser	18	21
Glu	20	Gln + Glu = 40
Gln	19 } 39	
Pro	16	12
Gly	27	30
Ala	30	23
Val	25	17
Met	5	0
Ile	20	15
Leu	43	39
Tyr	3	1
Phe	9	8
Trp	2	0

respectively one and several internal sites in PBG-D cDNA, two and several bands were expected and obtained (Fig. 3). These results establish the existence of only one gene coding for PBG-D and exclude the possibility that two distinct structural genes are tandemly repeated in the same chromosomal region [i.e. one coding for the erythrocytic PBG-D and one for a non-erythroid isoenzyme (33)].

Northern blot analysis of PBG-D mRNA

In order to determine the size of PBG-D mRNA, Northern blot analysis of mRNA obtained from various erythroid or non erythroid tissues and cultured cell lines was performed. Only one size species of mRNA was found in all cases suggesting that these mRNAs sequences are not qualitatively different in erythroid and non erythroid cells. The possibility of several species of mRNA which differ by less than 50 nucleotides cannot, however, be ruled out by this kind of experiment. The size of PBG-D mRNA was estimated to be 1600 bases, based upon its electrophoretic mobility relative to known standards (Fig. 4). Since PBG-DH⁴ contains 1380 bases excluding the poly(A) tail, this confirms that this cDNA covers the entire translated region.

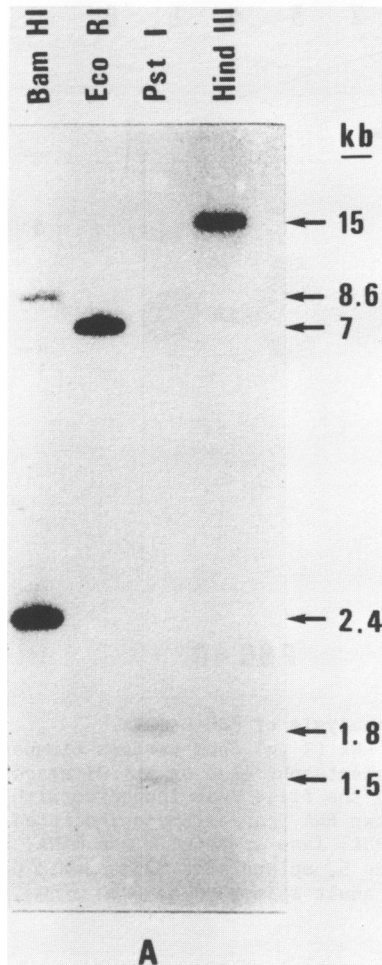


Fig. 3 : Southern analysis of human genomic restriction fragments containing sequences coding for PBG-D.

High molecular weight DNA was extracted from human leucocytes, cleaved with the restriction endonucleases indicated, electrophoresed on a 1% agarose gel and transferred to a nitrocellulose filter. The filter was hybridized with ^{32}P -labelled PBG-D probe. Bacteriophage λ DNA Hind III digest was used as DNA size markers.

The analysis of the relative intensity of the bands obtained after autoradiography showed that in the tissues tested, the concentration of PBG-D mRNA present is highest in those that have some erythroid activity (fetal liver, spleen with erythropoietic cells islands). This is in good agreement with previous determinations of enzymatic activities and suggests

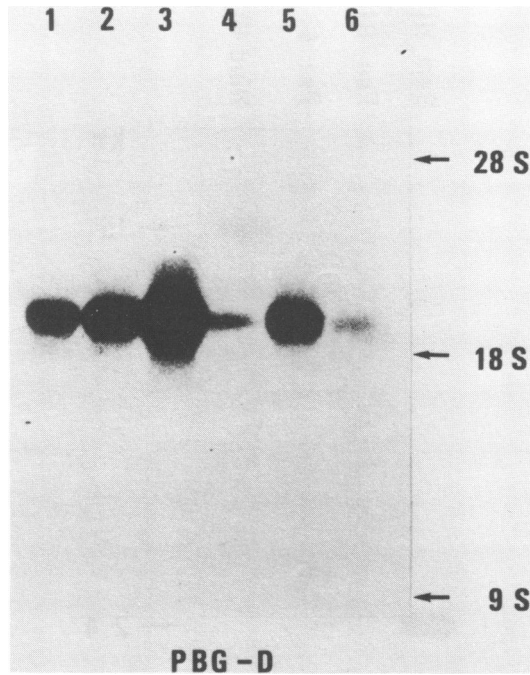


Fig. 4 : Northern blot analysis of PBG-D mRNA

Human poly(A⁺) RNA (5 µg) from various tissues and cell lines was denatured and electrophoresed on a 1.2% agarose formaldehyde gel. After blotting, the filter was incubated with PBG-D probe. The position of known RNA fragments are indicated. Lane 1, HEL mRNA; Lane 2, K562 mRNA; lane 3, fetal liver mRNA; lane 4, normal adult liver mRNA; lane 5, spleen of a thalassemic patient mRNA and lane 6, normal adult spleen mRNA.

that the tissue specific control of the expression of PBG-D gene occurs at a pretranslational level. Evaluation of the transcriptional activity of this gene in different tissues will indicate whether this regulation is at a transcriptional and/or at a post-transcriptional level (see below).

Among the cell lines tested (Fig. 5), the concentration of PBG-D mRNA was similar in HEL cells, which have some erythroid properties (34), and in cell lines such as HL 60 or Hela which have no erythroid properties (35, 36). Furthermore, K562 cells (37) had only twice as much PBG-D mRNA as HEL and HL 60 cells. The simplest explanation of these results is that the expression of the PBG-D gene in those cell lines is affected both by the stage of differentiation and by the fact that they are transformed cells. Thus the different pattern of PBG-D expression in normal cells and their

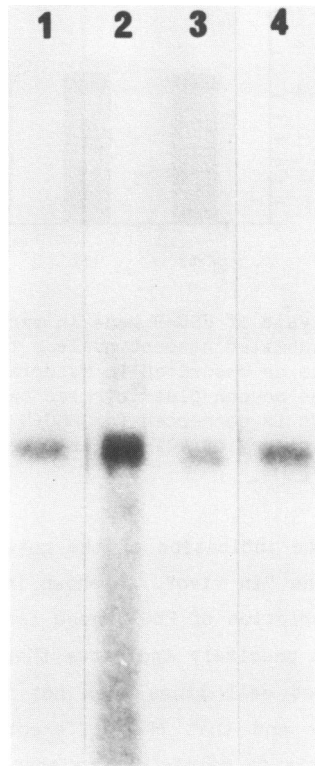


Fig. 5 : Northern blot analysis of PBG-D mRNA in various cell lines
Total RNA (12 μ g) from various human cell lines was denatured and electrophoresed on a 1.2% agarose formaldehyde gel. After blotting, the filter was incubated with PBG-D probe.
Lane 1, HL60 RNA; lane 2, K562 RNA; lane 3, Hela RNA;
lane 4, HEL RNA;

transformed counterparts indicates that great caution must be exercised in using erythroleukemic cell lines as model systems for studying normal erythroid differentiation. A recent report (38) has also shown that carbonic anhydrase is aberrantly and constitutively expressed in human erythroleukemia cells HEL, and this might also be the case for other proteins.

PBG-D gene transcription in erythroid and non erythroid cell lines

To determine why the same quantities of PBG-D mRNA are present in HEL and HL 60 cell lines, we measured the rate of transcription of this gene in isolated nuclei. Since reinitiation does not occur when isolated nuclei are transcribed "in vitro", the amount of nucleotide incorporation into RNA in

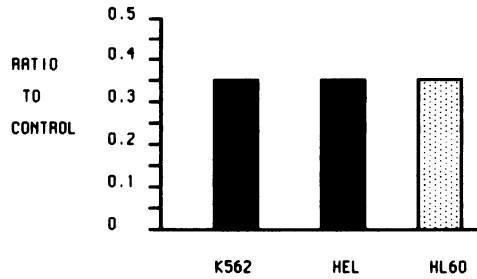


Fig. 6 : Transcription analysis of PBG-D gene in various cell lines.

3.10⁷ cpm of ³²P labelled nascent nuclear transcripts obtained from each cell line as described in Material and Methods were hybridized to "gene screen plus" circles carrying 5 µg of PBG-D cDNA. Incorporation is corrected for PBG-D gene length and for a hybridization efficiency of 40% and is given as a ratio to the 28S ribosomal control.

such an assay is an accurate indication of the rate of transcription initiation in a particular gene "in vivo". As shown in Fig. 6, we found that the relative rate of transcription of PBG-D gene is the same in K 562, HEL and HL 60 cell lines. This precisely indicates that the type of differentiation of those transformed cell lines does not influence the level of transcription of this gene and that the different amount of PBG-D mRNA between K 562 and HL 60 cells is mainly due to post-transcriptional mechanisms.

Our data clearly indicate that this housekeeping gene can be regulated both by tissue differentiation and neoplastic cell line transformation. These results, together with the recent report of a variant of A.I.P. in which an inherited PBG-D defect is present in liver but not in erythrocytes (33) support the hypothesis that a subtle transcriptional control mechanism exists which can lead to the production of the same amount of hybridizing RNA from two different promoters: one which controls expression during erythroid differentiation and another, ubiquitous, which is modulated during cell transformation. Fine mapping of the mRNA expressed in different tissues and studies on the gene structure are needed to clarify the promoter organization and to elucidate some of the mechanisms involved in this differential expression.

ACKNOWLEDGEMENTS

We thank Pr. J. Rosa for continuous support and encouragement, Y. Blouquit, B. Grandchamp, A. Munnich for helpful discussions, M. Titeux for

technical assistance in cell cultures, C. Valentin for technical assistance in nucleotide sequencing and A.M. Dulac and M. Tassier for assistance in the preparation of the manuscript. We thank Pr. G.H. Elder, Department of Medical Biochemistry, University of Wales College of Medicine, Cardiff, UK, for providing the amino acid sequence data. Amino acid sequencing was performed by the SERC Protein Sequencing Unit, Department of Biochemistry, University of Leeds, UK (Dr. J.B.C. Findlay).

*To whom correspondence should be addressed

REFERENCES

1. Battersby, A.R., Fookes, C.J.R., Matcham, G.W.J. and McDonald, E. (1980) *Nature*, **285**, 17-21.
2. Meyer, U.A., Strand, L.J., Doss, M., Rees, A.C. and Marver, H.S. (1972), *N. Engl. J. Med.*, **286**, 1277-1282.
3. Anderson, P.M., Reddy, R., Anderson, K.E. and Desnick, R.S. (1981), *J. Clin. Invest.*, **68**, 1-12.
4. Granick, S. and Sassa, S. (1971) in *Metabolic Regulation* (Vogel, H.S. ed.) vol. 5, pp 77-141, Academic Press, New York.
5. Beru, N. and Goldwasser, E. (1985) *J. Biol. Chem.*, **260**, 9251-9257.
6. Grandchamp, B., Beaumont, C., de Verneuil, H. and Nordmann, Y. (1985) *J. Biol. Chem.*, **260**, 9630-9635.
7. Grandchamp, B., Romeo, P.H., Dubart, A., Raich, N., Rosa, J., Nordmann, Y. and Goossens, M. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 5036-5040.
8. Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.*, **114**, 193-198.
9. Zasloff, A., Ginder, G.D., Felsenfeld, G. (1978) *Nucleic Acids Res.*, **5**, 1139-1151.
10. Durnam, D.M. and Palmiter, R.D. (1979) *Anal. Biochem.*, **131**, 385-393.
11. Itoh, N., Nose, K. and Okamoto, H. (1979) *Eur. J. Biochem.*, **97**, 1-9.
12. Ares, M.J. and Howell, S.H. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5577-5581.
13. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408-1412.
14. Brûlet, P. and Roskam, W.G. (1983) *Gene*, **22**, 59-68.
15. Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978) *J. Biol. Chem.* **253**, 2483-2495.
16. Michelson, A.M. and Orkin, S.H. (1982) *J. Biol. Chem.* **257**, 14773-14782.
17. Hanahan, D. (1983) *J. Mol. Biol.*, **166**, 557-580.
18. Hanahan, D. and Meselson, M. (1980) *Gene*, **10**, 63-67.
19. Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 3961-3966.
20. Thayer, R.E. (1979) *Anal. Biochem.* **98**, 60-63.
21. Goossens, M. and Kan, Y.W. (1980) *Meth. Enzymol.*, **76**, 805-807.
22. Hoffmann, L.M., Fritsch, M.K., and Gorski, J. (1981) *J. Biol. Chem.*, **256**, 2597-2600.
23. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
24. Weber, J., Jelinek, W., Darnell, J.E.Jr (1977) *Cell*, **10**, 611-616.
25. Vaulont, S., Munnich, A., Marie, J., Reach, G., Pichard, A.L., Simon, A.P., Besmond, C., Barbry, P. and Kahn, A. (1984) *Biochem. Biophys. Res. Commun.*, **125**, 135-141.

Nucleic Acids Research

26. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzymol.*, **65**, 499-560.
27. Brown, R.C., Elder, G.H., Urquihart, A.J. (1985) *Biochem. Soc. Trans.* **13**, 1227-1228.
28. Kozak, M. (1984) *Nucl. Acids Res.* **12**, 857-872.
29. Kozak, M. (1984) *Nature*, **308**, 241-246.
30. Anderson, P.M. and Desnick, R.J. (1980) *J. Biol. Chem.*, **255**, 1993-1999.
31. Walker, J.M., Hastings, J.R.B. and Johns, E.W. (1980) *J. Chrom.*, **189**, 105-107.
32. Desnick, R.J., Ostasiewicz, L.T., Tishler, P.A. and Mustajoki, P. J. *Clin. Invest.* (1985) **76**, 865-874.
33. Mustajoki, P. (1981) *Ann. Intern. Med.*, **95**, 162-166.
34. Martin, P. and Payannopoulou, T. (1982) *Science*, **216**, 1233-1235.
35. Collins, S.S., Gallo, R.C. and Gallagher, R.E. (1977) *Nature*, **270**, 347-349.
36. Gey, G.O., Coffmann, W.D., and Kubicek, M.T. (1952) *Cancer Res.*, **12**, 264-270.
37. Anderson, L.C., Jokinen, M. and Gahmberg, C.G. (1979) *Nature*, **278**, 364-365.
38. Frankel, S.R., Walloch, J., Hirata, R.K., Bondurant, M.C., Villanueva, R. and Weil, S.C. (1985) *Proc. Natl. Acad. Sci., USA*, **82**, 5175-5179.