

Nucleotide sequence of the human parathyroid hormone gene

(Z-DNA/ π VX plasmid/intervening sequence/domain evolution)

THOMAS J. VASICEK*^{†‡}, BARBARA E. MCDEVITT*^{†§}, MASON W. FREEMAN*[†], BARBARA J. FENNICK*[†],
GEOFFREY N. HENDY*^{†¶}, JOHN T. POTTS, JR.[†], ALEXANDER RICH^{||}, AND HENRY M. KRONENBERG*[†]

*Endocrine Genetics Unit and [†]Department of Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts 02114; and
^{||}Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contributed by Alexander Rich, December 29, 1982

ABSTRACT From a λ phage gene library we have isolated phage containing the gene encoding human preproparathyroid hormone. The phage were isolated by using both the plaque-hybridization technique and the *in vivo* recombination–selection technique. The human preproparathyroid hormone gene contains two intervening sequences that separate the gene into a 5' non-coding domain, a "prepro" sequence domain, and a domain containing the parathyroid hormone sequence and the 3' noncoding region. The gene is approximately 4,200 base pairs long. Restriction endonuclease analysis of human leukocyte DNA shows that the haploid human genome contains one copy of the preproparathyroid hormone gene. A 14-base-pair sequence of alternating purines and pyrimidines that has the potential of adopting the Z-DNA conformation lies 134 base pairs upstream from the presumed site of initiation of transcription.

Parathyroid hormone (PTH) is the principal homeostatic regulator of blood calcium, and, in turn, the secretion of PTH is closely regulated by the blood level of calcium. While the regulation of secretion of PTH has been studied extensively, the regulation of the synthesis of PTH has received comparatively little attention. We (1) and others (2, 3) have cloned cDNA encoding bovine PTH. More recently, we have cloned cDNA encoding human PTH as well (4). Here we describe the isolation and DNA sequence analysis of genomic DNA encoding human PTH. The human PTH gene contains two intervening sequences that separate the mRNA sequence into three functional domains. DNA blotting experiments show that the haploid human genome contains only one PTH gene.

METHODS

Screening a λ Phage Library. A human gene library in phage Charon 4A, derived from fetal liver, was provided by T. Maniatis (5). The library was first screened by the procedure of Benton and Davis (6), using plasmid pPTHm113 (4) containing human PTH cDNA as a nick-translated (7) probe. The library was subsequently screened by the method of Seed (8). Plasmid DNA was isolated from *Escherichia coli* strain W3110r⁻m⁺(p3)(π VX) by an alkaline miniprep procedure (9), the DNA was electrophoresed through a 0.7% agarose gel, and the π VX plasmid was isolated from the gel. By using phage T4 DNA ligase, the *Bgl* II/*Xba* I fragment from pPTHm122 (4) was inserted into the corresponding sites on the π VX plasmid. The resultant plasmid, π VX-PTH, containing the PTH cDNA linked to the selectable marker *supF*, was introduced into *E. coli* strain W3110r⁻m⁺(p3). One million library phage were then amplified on one plate of the resultant strain W3110r⁻m⁺(p3)(π VX-PTH). Then 5×10^8 of the resultant phage were grown on one plate of strain W3110r⁻m⁺Su⁻, a strain containing no amber suppressor tRNA gene. Growth of Charon 4A phage on *E. coli* W3110r⁻m⁺Su⁻

selected for phage that had incorporated the *supF* gene, encoding an amber suppressor tRNA, from the π VX-PTH plasmid. All work with organisms containing recombinant DNA was performed in a P1 physical containment facility according to the then current guidelines of the National Institutes of Health.

Subcloning PTH Gene Fragments. Fragments of DNA generated by restriction enzyme digestion of λ hPTH1 and λ hPTH2 were subcloned in plasmid pBR322, after blot-hybridization analysis (10) was used to determine which fragments contained portions of the PTH gene. Fragments resulting from cleavage with one restriction endonuclease were ligated to DNA from plasmid pBR322 that had been cleaved with the same enzyme and treated with calf intestinal alkaline phosphatase to prevent intramolecular religation. Fragments resulting from cleavage with two restriction endonucleases were ligated with the appropriate fragment of DNA from pBR322, which had also been cleaved with those enzymes. DNA fragments were isolated from agarose gels by using glass powder to bind the DNA (11) and were ligated by using T4 DNA ligase.

DNA Sequence Analysis. All DNA sequences were determined by the chemical method of Maxam and Gilbert (12). DNA was end-labeled either with polynucleotide kinase or with the large fragment of DNA polymerase I.

Blot-Hybridization of Uncloned Human DNA. DNA was isolated from blood leukocytes from 50 ml of blood from a normal volunteer (13). DNA was digested to completion with 10-fold excess enzyme, according to suppliers' protocols. Completion of the digestions was assayed by the addition of small amounts of an appropriate plasmid to an aliquot of the reaction and electrophoretic analysis of the products of digestion of the aliquot. Then 10 μ g of digested DNA was electrophoresed through a 1.0% agarose gel and transferred to nitrocellulose. Hybridization was performed according to a dextran sulfate protocol (14), using purified PTH cDNA fragments as probes. Nick-translation was performed with all four α -³²P-labeled deoxyribonucleoside triphosphates; the resultant DNA contained 10⁸ cpm/ μ g. The intensities of autoradiographic bands were measured with a Joyce-Loebl microdensitometer. Calculations of the number of copies of the PTH gene in the human genome used 3.3×10^9 base pairs (bp) as the size of the haploid human genome (15), 44,000 bp as the size of λ hPTH1, and 43,000 bp as the size of λ hPTH2.

RESULTS

Isolation of λ Phage Containing the Human PTH Gene. We first screened a human fetal liver genomic DNA library constructed in phage Charon 4A by Lawn *et al.* (5). We used the plaque-hybridization method, with nick-translated plas-

Abbreviations: PTH, parathyroid hormone; bp, base pair(s).

[†] Present address: Dept. of Biological Chemistry, Harvard Medical School, Boston, MA 02115.

[§] Present address: New Jersey Medical School, Newark, NJ 07103.

[¶] Present address: Dept. of Medicine, The Middlesex Hospital, London, U.K.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

mids pPTHm113 and pPTHm122 (4) as probes. Only one plaque, λ hPTH1, was positive. A partial physical map of λ hPTH1 was constructed by performing blot-hybridization analysis of λ hPTH1 DNA cleaved with a variety of restriction endonucleases and probed with specific DNA fragments from pPTHm122. This analysis demonstrated that λ hPTH1 contained the PTH gene's coding region and 3' noncoding region but that the 5' noncoding region was missing from the phage.

We chose to rescreen the λ phage library by using the recombination-selection method of Seed (8). We isolated from pPTHm122 the 288-bp *Bgl* II/*Xba* I fragment, which contains the PTH coding sequence and part of the mRNA's 3' noncoding region. This fragment was inserted into the plasmid π VX, using the *Bgl* II and *Xba* I sites in the "polylinker" region of that plasmid. Following Seed's protocol, we passed 10^6 phage from the gene library through *E. coli* strain W3110r⁻m⁺(p3) that had been transfected with π VX-PTH. Then 5×10^8 of the resultant phage were plated on *E. coli* W3110r⁻m⁺ S⁻ cells, which do not contain an amber suppressor tRNA gene and therefore do not allow the production of Charon 4A phage. Sixty plaques survived that selection. Six of the 60 plaques were screened by restriction enzyme digestion analysis and found to be identical. One plaque, λ hPTH2, from among these 6 was chosen for further analysis. Digestion of λ hPTH2 DNA with restriction endonucleases and subsequent blot-hybridization experiments using fragments of pPTHm122 as probes demonstrated that λ hPTH2 contained the entire human PTH gene with the π VX-PTH plasmid inserted, as expected, into the region between the *Bgl* II and *Xba* I sites used in constructing π VX-PTH (Fig. 1).

Examination of the Human PTH Gene in Uncloned DNA.

In order to verify that the two λ phage that we had isolated contained the normal human PTH gene free of artifactual rearrangements introduced during cloning, we compared the sizes of DNA fragments generated by restriction endonuclease digestion of human leukocyte genomic DNA with the sizes of corresponding fragments from λ hPTH1 and λ hPTH2. After agar-

ose gel electrophoresis and transfer to nitrocellulose, DNA fragments were hybridized with probes specific for either the 5' end or the 3' end of the human PTH gene. We used the 2,150-bp *Hind*III fragment which includes the first exon and part of IVS1 from pPTHg105 (see Fig. 1) as the 5'-specific probe. An 800-bp *Hpa* II fragment of pPTHm122 (4) was used as a 3'-specific probe. This fragment contains 100 bp of pBR322 as well as 547 bp of DNA corresponding to the second exon and part of the third exon. It also contains 49 bp at the end of the first exon; this short sequence does not form detectable hybrids under the conditions of hybridization used here. Fig. 2 demonstrates the comigration of the resultant radioactive fragments from the λ phage and leukocyte DNA. Further, because no unexpected bands appeared in the lanes containing leukocyte DNA, the results suggest that the human PTH gene is represented just once in the haploid human genome. To confirm this suggestion, we measured the intensities of the PTH-specific bands on the autoradiograms and compared them with the intensities of the bands generated from known amounts of DNA from λ hPTH1 and λ hPTH2. With the 5'-specific PTH probe, this analysis yielded 1.0 PTH gene per haploid genome; with the 3'-specific PTH probe, the result was 0.6 PTH gene per haploid genome.

DNA Sequence Analysis of the Human PTH Gene. Restriction endonuclease analysis and gene blotting experiments suggested that two intervening DNA sequences interrupted the PTH gene, one approximately 3,400 bp in length, and the other one about 100 bp in length. We determined the entire sequence of the gene, excluding the sequence of the internal portion of the first intervening sequence, and, in addition, determined the sequence of several hundred base pairs of DNA flanking the PTH gene. Fig. 1 illustrates the restriction enzyme map of fragments of λ hPTH1 and λ hPTH2 subcloned in pBR322 to facilitate sequence analysis, and also indicates the sequencing strategy. Fig. 3 illustrates the DNA sequence.

DISCUSSION

DNA Sequence Data. Fig. 1 shows that most of the sequence depicted in Fig. 3 was determined by analyzing both strands of

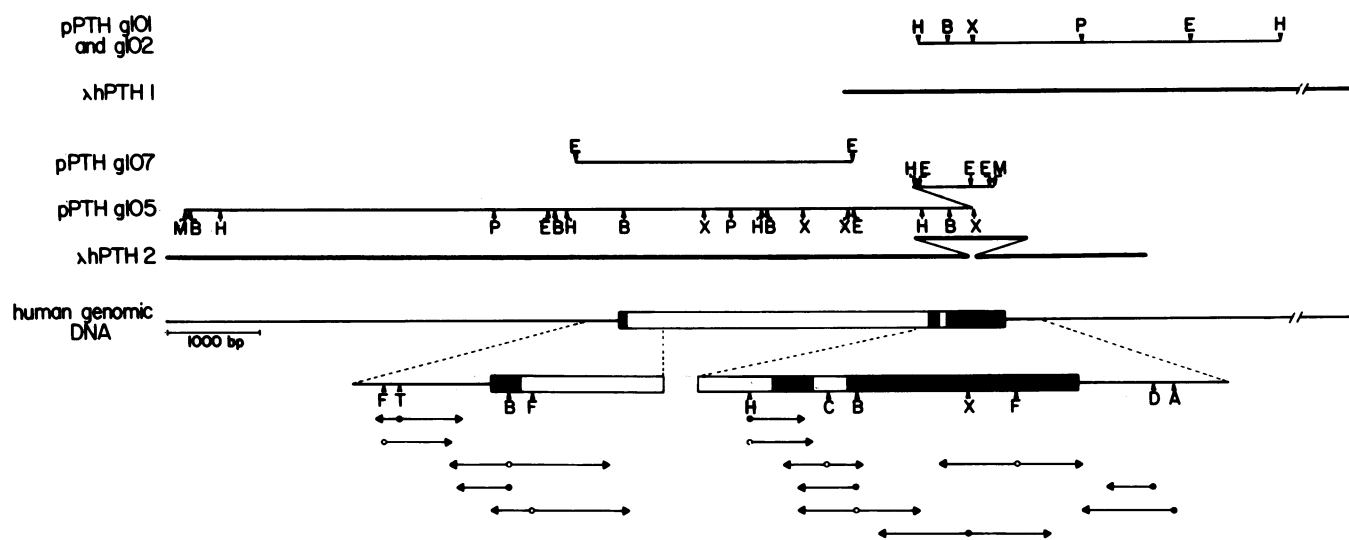


FIG. 1. Physical map of the human PTH gene. Horizontal lines indicate the length of human DNA inserted into the indicated λ phage and plasmid subclones. The raised portion of λ hPTH2 indicates the π VX-PTH insert. pPTHg101 and pPTHg102 were constructed by inserting the indicated fragment of λ hPTH1 into *Hind*III-cleaved pBR322. They differ only by their orientation in pBR322. pPTHg105 and pPTHg107 were constructed by inserting the indicated fragments of λ hPTH2 into *Bam*HI-cleaved and *Eco*RI-cleaved pBR322. Restriction enzyme maps of the human portions of pPTHg101 and pPTHg105 are indicated. The human PTH gene is indicated on the line "human genomic DNA" by the thick line; black areas are exons, white areas are introns. More detailed and magnified copies of portions of the PTH gene are given below the human genomic DNA line. Arrows at the bottom of the figure indicate the DNA sequence analysis strategy. Open circles represent 5' ends of fragments labeled with polynucleotide kinase. Closed circles represent 3' ends of fragments labeled with the large fragment of DNA polymerase. Arrows show how far the DNA sequence could be reliably read without reference to other data. The broken line at the end of λ hPTH1 indicates that 9,000 bp have been deleted from the figure for ease of presentation. A, *Ava* I; B, *Bgl* II; D, *Dde* I; E, *Eco*RI; F, *Hinf*I; H, *Hind*III; M, *Bam*HI; P, *Hpa* I; X, *Xba* I.

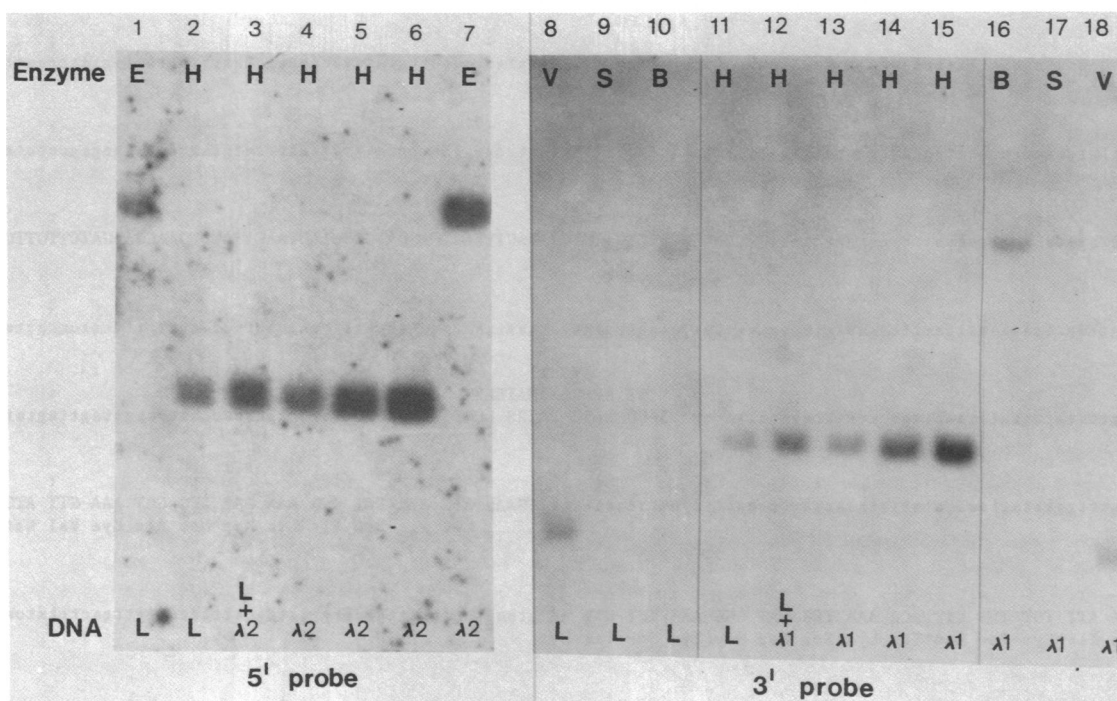


FIG. 2. Blot-hybridization analysis of human leukocyte DNA and cloned DNA. High molecular weight leukocyte DNA (L) and DNA from λ hPTH1 (λ 1) and λ hPTH2 (λ 2) were digested with a series of restriction endonucleases (E, *EcoRI*; H, *HindIII*; V, *Pvu II*; S, *Sst I*; B, *Bgl II*) and applied to wells as indicated in the figure. Wells 1–7 were probed with the 5'-specific probe; wells 8–18 were probed with the 3'-specific probe. Ten micrograms of leukocyte DNA or 300 μ g of phage DNA was applied to each well except for the following wells: wells 3 and 12, 10 μ g of leukocyte DNA plus 150 μ g of phage DNA; wells 4 and 13, 150 μ g of phage DNA; wells 6 and 15, 450 μ g of phage DNA.

plasmid DNA and that the ordering of fragments generated by cleavage with restriction endonucleases was confirmed by sequencing across cleavage sites. The sequence of the human PTH mRNA predicted by analysis of the cloned genomic DNA confirms without exception the sequence determined from analysis of five cDNA clones (4). We are thus quite confident of the accuracy of the AUG sequence found in the 5' noncoding region of the mRNA sequence, for example. Further, the genomic DNA sequence confirms the sequence of the region encoding human preproPTH. This sequence, derived from a presumably normal fetal liver, allows us to deduce the human PTH sequence from an individual with no parathyroid gland disease. [Both the previously determined protein sequences (16) and cDNA sequence (4) were derived from parathyroid tumor materials.] As expected, we can conclude that patients with hyperparathyroidism make a structurally normal parathyroid hormone.

We have not determined the start point of PTH gene transcription either by using an *in vitro* transcription system or by analyzing the far 5' terminus of the human PTH mRNA. Consequently, the assignment of the transcription start site at nucleotide $-3,566$ in Fig. 3 is only an approximation based on three considerations: (i) The bovine PTH mRNA, whose 5' noncoding region closely resembles the 5' noncoding region of the human mRNA, begins at three tightly clustered A residues (corresponding to nucleotides $-3,568$, $-3,566$, and $-3,561$ of the human gene) as determined from reverse transcription of mRNA by Weaver *et al.* (3). (ii) Nucleotide $-3,566$ is 29 nucleotides from the sequence T-A-T-A-T-A, commonly found 27–31 nucleotides from transcription start sites (17). (iii) Transcription usually starts with an A (17), thus making nucleotide $-3,568$, a G residue in the human sequence, a less likely start site.

Organization of Intervening Sequences in the Human PTH Gene. By comparing the sequence of human PTH cDNA (4) and the sequence of human PTH genomic DNA, we can deduce the location of two intervening DNA sequences interrupting the

PTH gene. The first is approximately 3,400 bp long and interrupts the 5' noncoding region of the mRNA five nucleotides before the start of the coding region. The second intervening sequence is 103 bp long and comes between the second and third nucleotide encoding lysine-29 of the prepro-PTH molecule. Because of redundant sequences at the splice junctions, the exact splice start and stop sites cannot be assigned unambiguously. However, if we require that the splice donor and acceptor sequences follow the consensus sequence pattern of other donor and acceptor sequences (17), then the junctions can be assigned without ambiguity. In both introns the sequences of the donor and acceptor junction closely follow the consensus sequence predicted from previous analyses.

The first intervening sequence comes close to the end of the 5' noncoding region of the mRNA and therefore follows the pattern noted by Gilbert (18), that intervening sequences often separate mRNAs into functional domains. The exact location of the intervening sequence is intriguing, because it comes precisely after the potential AUG codon in the 5' noncoding region of the mRNA. After removal of the intervening sequence, the AUG is then followed by a UGA termination codon in the mRNA. The location of the AUG just before the start of the intervening sequence raises the possibility that alternative splicing patterns might allow the AUG to direct the synthesis of a second protein encoded by another nucleotide sequence. The first intervening sequence contains an in-frame termination codon only 15 codons after the AUG; therefore, an unspliced RNA could not use the AUG effectively. The possibility remains that the splice donor sequence could ligate to more than one acceptor sequence, however. Of course, the juxtaposition of the unusual 5' noncoding AUG and the intervening sequence may be coincidental. The bovine PTH mRNA has the sequence GUG instead of AUG in the 5' noncoding sequence (1, 3), suggesting that the 5' noncoding AUG is not of vital significance. On the other hand, it is striking that two of the other rare cases of mRNAs with 5'

PTH. Comparison of the "pro" sequences of the pro-PTH molecules from several species reveals that none of the first four amino acids of the "pro" sequences is conserved; in contrast, the last two residues are always Lys-Arg (21). The human PTH gene's second intervening sequence, then, separates the "pre" sequence and the variable portion of the "pro" sequence from the basic residues Lys-Arg and the PTH sequence. According to Gilbert's hypothesis, through evolutionary time, the PTH sequence carries with it the two basic residues required for cleavage of the hormone sequence from any peptide to which splicing events might fuse it.

Number of PTH Genes. Comparison of the sizes of DNA fragments generated by restriction enzyme digestion of uncloned human DNA and DNA from λ hPTH1 and λ hPTH2 suggests that major rearrangements did not occur during the cloning. Two kinds of data suggest that there is only one PTH gene in the human haploid genome. First, the blots of enzyme digests of uncloned DNA contain only bands predicted by the maps of λ hPTH1 and λ hPTH2. These blots contain fragments generated by several restriction enzymes and probes covering both the far 5' and 3' ends of the gene. These results strongly suggest that the human genome contains only one PTH gene, but they cannot eliminate the possibility that the DNA sequences flanking multiple human PTH genes are tightly conserved in the genome, resulting in identical physical maps of the multiple PTH genes. This possibility is eliminated by the second, quantitative, experiment, in which we measured in uncloned DNA the amount of DNA recognized by PTH-specific probes. By comparing the intensity of autoradiographic bands in DNA blots of uncloned genomic DNA and known amounts of cloned DNA, we could conclude that the genomic bands had the intensities predicted if the PTH gene is a unique gene. In a control well, we mixed together cloned and uncloned DNA and found that the resultant band after blotting had the intensity predicted by simple addition of the intensities associated with each of the DNAs. This result shows that the several micrograms of unrelated DNA in the wells containing uncloned DNA did not interfere with the electrophoresis, blot-transfer, or hybridization of the PTH DNA. Because these blots were hybridized under stringent conditions, we cannot eliminate the possibility that the human genome contains PTH-like genes so different from the PTH gene that they could not be recognized by PTH gene probes.

Potential Z-DNA Structure. The existence of a 14-bp segment of alternating purines and pyrimidines (-3,699 to -3,686) 134 bp upstream from the presumed site of initiation raises many interesting questions regarding the possible role of left-handed Z-DNA in the regulation of human PTH transcription. Negative supercoiling is the major driving force for stabilizing segments of left-handed Z-DNA. Nordheim *et al.* (22) have recently shown that antibody to Z-DNA binds to a 14-bp sequence of alternating purines and pyrimidines with one base pair out of alternation, when the sequence in the plasmid pBR322 is negatively supercoiled in the physiological range. The alternating sequence found near human PTH is 14 bp in length; however, it is part of an 18-bp segment of alternating purines and pyrimidines with one base pair out of alternation. It is thus longer than the segment that has been observed to form Z-DNA in the plasmid pBR322. The sequence near the PTH gene is largely of the type $(dC-dA)_n(dG-dT)_n$, and DNA segments of this type have been found to form Z-DNA at levels of negative supercoiling that are within the physiological range (23). Hamada *et al.* (24) have shown that longer sequences of the type $(dC-dA)_n(dG-dT)_n$ are widely dispersed through the eukaryotic genome in a number of species. In their assay they determined that there were over 50,000 copies of this type, 50 bp or greater in length, in the human genome.

Recently a number of proteins have been found in the nuclei of *Drosophila* cells that have the ability to bind specifically to left-handed Z-DNA but not to right-handed B-DNA (25). These proteins were found to have the property of stabilizing the Z-DNA conformation, and in particular were shown to bind to negatively supercoiled plasmids containing the sequence $(dC-dA)_n(dG-dT)_n$. It is reasonable to believe that similar Z-DNA binding proteins exist in human cells. These proteins may be able to stabilize the Z-DNA conformation in the nucleotide sequence 134 bp from the presumed site of transcription initiation of the PTH gene. At the present time we do not know what the effects would be of having Z-DNA binding protein attached to a segment of Z-DNA at this site. If we postulate that the Z-DNA binding proteins bind tightly to this site, and if this site is in the promoter region, the proteins could block the attachment of RNA polymerase and decrease the level of transcription. Alternatively, if such proteins were readily released from this Z-DNA binding site and the local region of Z-DNA converted to one of B-DNA, the release would enhance the negative superhelical density in that region, which, in turn, might promote the level of RNA polymerase binding. Further experimentation is necessary to show whether this region on the 5' side of the PTH gene plays any role in the regulation of gene expression.

We thank Mary Douvadjian for secretarial help. This work was supported by National Institutes of Health Grants AM 21529 and AM 11794.

- Kronenberg, H. M., McDevitt, B. E., Majzoub, J. A., Nathans, J., Sharp, P. A., Potts, J. T., Jr., & Rich, A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4981-4985.
- Weaver, C. A., Gordon, D. F. & Kemper, B. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4073-4077.
- Weaver, C. A., Gordon, D. F. & Kemper, B. (1982) *Mol. Cell. Endocrinol.* **28**, 411-424.
- Hendy, G. N., Kronenberg, H. M., Potts, J. T., Jr., & Rich, A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7365-7369.
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1981) *Cell* **15**, 1157-1174.
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180-182.
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237-251.
- Seed, B. (1983) *Nucleic Acids Res.*, in press.
- Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523.
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517.
- Vogelstein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 615-619.
- Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
- Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32-38.
- Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683-3687.
- Lewin, B. (1980) *Gene Expression* (Wiley, New York), 2nd Ed., Vol. 2, p. 962.
- Keutmann, H. T., Sauer, M. M., Hendy, G. N., O'Riordan, J. L. H. & Potts, J. T., Jr. (1978) *Biochemistry* **17**, 5723-5729.
- Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349-383.
- Gilbert, W. (1978) *Nature (London)* **271**, 501.
- Young, R. A., Hagenbuchle, O. & Schibler, U. (1981) *Cell* **23**, 451-458.
- Potts, J. T., Jr., Kronenberg, H. M. & Rosenblatt, M. (1982) *Adv. Protein Chem.* **35**, 323-396.
- Cohn, D. V. & MacGregor, R. R. (1981) *Endocrinol. Rev.* **2**, 1-6.
- Nordheim, A., Lafer, E. M., Peck, L. J., Wang, J. C., Stollar, B. D. & Rich, A. (1982) *Cell* **31**, 309-318.
- Nordheim, A. & Rich, A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1821-1825.
- Hamada, H., Petrino, M. G. & Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6465-6469.
- Nordheim, A., Tesser, P., Azorin, F., Kwon, Y. H., Möller, A. & Rich, A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7729-7733.