Nucleic Acids Research, Vol. 19, No. 22 6263–6268

Genomic organization and complete sequence of the human gene encoding the β-subunit of the cGMP phosphodiesterase and its localisation to 4p16.3

Bernhard Weber, Olaf Riess, Gordon Hutchinson, Colin Collins, Biaoyang Lin, David Kowbel, Susan Andrew, Keith Schappert[1] and Michael R.Hayden*
Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 2B5 and [1]Department of Genetics, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

Received July 31, 1991; Revised and Accepted October 22, 1991 EMBL accession nos X62692–X62695 (incl.)

ABSTRACT

As part of the search for the Huntington disease (HD) gene we have cloned and sequenced 34 kb of genomic DNA containing the full-length gene for the β-subunit of the human cGMP phosphodiesterase (β-cGMP PDE). This gene is localized to 4p16.3 about 700 kb proximal to the 4p telomere and represents the most telomeric gene characterized on 4p to date. We show that this gene is comprised of 22 exons spanning approximately 43 kb of genomic DNA. We also provide 400 bp immediately 5′ to the putative initiator methionine and 700 bp of 3′ flanking sequences. Northern blot analysis of several human tissues revealed a highly abundant 3.5 kb transcript and a minor signal of 4.5 kb in retinal tissue. Alignment of the deduced amino acid sequence to the previously identified β-subunits of the cGMP PDEs of mouse and cow demonstrates highly significant similarities and, therefore, confirms the identity of the cloned gene. A defect in the β-subunit of the cGMP PDE gene has been shown recently to be the cause for the retinal degeneration in the rd mouse. The cloning of the human homolog and the knowledge of its genomic organization with exon/intron boundaries will allow rapid assessment of the role of this gene in the causation of human retinopathies.

INTRODUCTION

Approaches to cloning genes for specific diseases with unknown biochemical defects have concentrated on the identification of minimal candidate regions likely to contain the defective gene. With this strategy, the gene is cloned based on its chromosomal location rather than its functional properties. Alternatively, the candidate gene approach bases the identification of a disease gene on a theoretical hypothesis as to the pathogenesis of the disorder. In the future, we are increasingly likely to see the cloning of particular disease genes by a combination of the candidate gene approach in conjunction with the knowledge of the location of the mutant gene to a particular chromosomal region.

We have previously reported on the cloning and physical characterization of a candidate region for the Huntington disease (HD) gene within chromosomal region 4p16.3 (1). A total of 15 CpG-rich islands were characterized and sequences surrounding these islands were tested for cross species conservation. This strategy allowed the rapid identification of putative coding sequences within the cloned region. As part of the search for the HD gene, we now report the cloning and characterization of a gene from this candidate region. GeneBank alignment established significant similarities to previously cloned genes from mouse and cow and identifies the isolated gene as the β-subunit of the cyclic GMP phosphodiesterase (cGMP PDE).

In retinal rod cells, cGMP PDE is involved in the phototransduction pathway and consists of two large catalytic subunits, α and β, and two small inhibitory γ polypeptides (2,3). The genes encoding the α-subunit of cGMP PDE of human (4), cow (3,5) and mouse (6), the β-subunit of cow (7) and mouse (8) and the γ-subunit of human (9), cow (10) and mouse (11) have previously been cloned and assigned to human or mouse chromosomes. The α-subunit was mapped to 5q31.2–5q34 in humans (4) and to chromosome 18 in mouse (12), and the β-subunit was localized in mouse to chromosome 5 (8, 13) and the γ-subunit to mouse chromosome 11 (14).

In rod cells the signalling cascade begins with the absorption of a photon by rhodopsin which is localized in the rod outer segment membranes. The photon excitement triggers a conformational change in the receptor protein which binds a membrane-associated G protein, transducin. This catalyzes a GDP-GTP exchange reaction on the α-subunit of transducin and the dissociation from the β/γ subunit complex. The GTP bound α-subunit of transducin then activates a photoreceptor-specific phosphodiesterase (PDE) complex which hydrolyzes 3′, 5′-cyclic GMP to 5′ GMP. As cGMP acts allosterically on the sodium channels of the plasma membrane a reduction of intracellular concentrations of cGMP leads to the closure of the channels,

* To whom correspondence should be addressed

consequently to hyperpolarization of the plasma membrane and, thereby, visual signaling (15,16).

Mutations in any one of the many genes involved in the complex biochemistry of the eye could theoretically impair vision. Until recently the biochemical defects underlying retinal diseases were mostly unknown. However, the application of DNA technology to these problems has led to new insights. A few forms of human hereditary retinal degeneration such as gyrate atrophy of the choroid and retina (17), choroideraemia (18), and autosomal dominant retinitis pigmentosa (19) have been linked to mutations in specific genes and it has recently been demonstrated that the gene encoding the β-subunit of the rod cGMP PDE is responsible for the autosomal recessive retinal degeneration in the rd mouse (8). The isolation of the human homolog to the mouse β-subunit of the cGMP PDE will now allow rapid determination as to whether any human retinal disease is caused by a defect in this gene. Moreover, the knowledge of the exon/intron sequences of this gene will rapidly facilitate the assessment of possible mutations without the need for obtaining retinal RNA from affected persons.

MATERIALS AND METHODS

Isolation of overlapping cosmid clones and sequence analysis

Overlapping cosmids c16Dp, c17QB1, and c18E were isolated as part of a cosmid walking project with the goal of cloning a candidate region containing the HD gene. A detailed description of the cosmid libraries, construction, screening procedures, determination of overlapping fragments and restriction mapping has been described elsewhere (1). Sequence analysis of purified plasmid subclones (vector: Bluescript KS II(+),Stratagene) was done using the dideoxy nucleotide chain termination method (20) and the Sequenase DNA sequencing kit (US Biochemicals).

Alu element-mediated PCR cloning

Alu primer A1S (TCATGTCGACGCGAGACTCCATCTC-AAA) was used in an Alu element-mediated PCR reaction as described previously (21) in order to synthesize and subclone small fragments from cosmid c16Dp.

Identification of intron/exon junctions and sequence data analysis

Tentative splice junctions were first determined by alignment of the genomic sequence with the bovine β-subunit cDNA sequence in the GenBank database (acc.no JO 5553). Splice consensus sequences were then scored by calculating the 'discrimination energy' (22, 23). Nucleotide sequence data from the bovine and mouse (acc.no X55968) β-subunit and from the bovine (acc.no. M 26043) and human (acc.no. M26061) α-subunit cGMP PDEs have been extracted from GenBank Version 67. The translated peptide sequences were then aligned using Hein's program (24).

Localisation of the β-subunit cGMP PDE locus by pulsed field gel electrophoresis analysis

The preparation of high molecular weight DNA and the separation of large restriction fragments using the transverse alternating field electrophoresis system (Beckman Inc.) were performed as described previously (1).

RNA preparation, Northern blot analysis

RNAs from various tissues were isolated using a single step method by homogenization in guanidinium isothiocyanate as described (25). Poly(A)+ RNA was purified using oligo(dT) push columns (Stratagene). RNA samples were size fractionated on a 2.2M formaldehyde/1% agarose gel and transferred to Hybond N membranes. Hybridizations were done at 65°C in 0.5M sodium phosphate buffer pH 7.2/7% sodium dodecyl sulfate (26)

RESULTS

Cloning and structure of the β-subunit of the cGMP-PDE gene

We recently reported the isolation of 460 kb of contiguously overlapping genomic DNA from a candidate region for the HD gene (1). In order to rapidly define exonic sequences within the contig DNA we choose an approach which included the characterization of CpG-rich islands and the subsequent testing of adjacent sequences for cross-species conservation.

Probe K1, a single copy 1.1 kb subclone of cosmid c16Dp (fig 1) which was generated by using Alu element-mediated PCR showed cross species conservation in several mammalian species (data not shown). Clone K1 was sequenced and searched in the GenBank database for sequence similarities with previously described genes. Computer alignment revealed significant similarities between 138 nucleotides (nt) of subclone K1 and the β-subunits of the cyclic GMP phosphodiesterase from cow (124/138 nt identical) and mouse (123/138 nt identical).

In order to determine the complete genomic locus of this gene, we initiated the sequencing of genomic fragments between CpG-rich islands I and II (fig 1). As the K1 fragment is flanked by

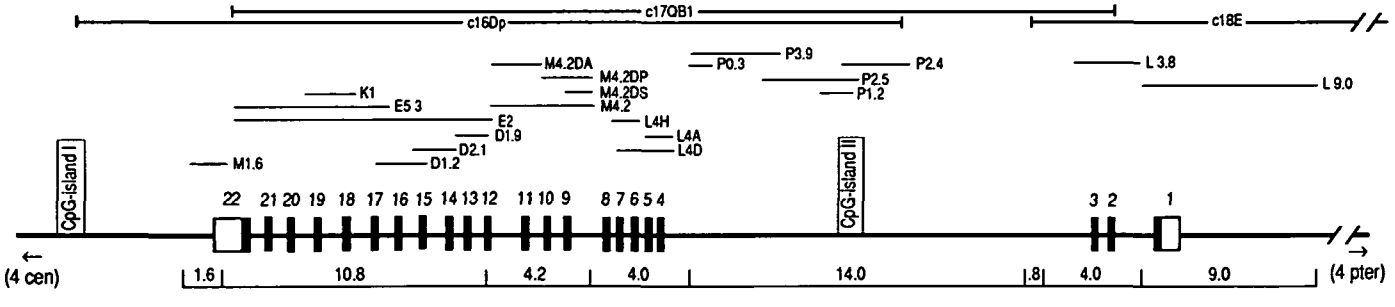

Figure 1. Physical map and genomic organization of exons 1 to 22 of the human β-subunit of the cGMP PDE. The upper panel shows three overlapping cosmids c16Dp, c17QB1, and c18E together encompassing the β-subunit cGMP PDE locus. The horizontal bars below depict subclones used in the sequencing analysis or in Southern and Northern blot experiments. The CpG-rich islands I and II have been previously defined (1). Solid boxes represent exons 1 to 22. 5' and 3' flanking regions are shown as open boxes. EcoRI fragments and corresponding sizes in kb are shown below. The 5'-3' orientation of the β-cGMP PDE locus within 4p is indicated.

these two islands within a distance of approximately 30 kb of genomic DNA (fig 1) and the mouse β-subunit of the cGMP PDE detects a transcript of approximately 3.1 kb (8) we reasoned that the complete gene could well be contained within the 30 kb of genomic DNA.

EcoR1 fragments from cosmids c16 Dp and c17 QB1 were further digested with different enzymes in order to obtain overlapping fragments in a convenient size range for sequence analysis (fig 1). Alignment of the obtained sequences with the cDNAs of the bovine (7) and mouse (8) β-subunits of the cGMP PDEs revealed that 19 exons including the 3′ flanking region were contained within the 30 kb of genomic DNA (fig 1). However, nucleotides 1 to 712 as numbered from the initiation codon (ATG, A=1) of the bovine cDNA were not contained within the sequenced genomic fragments.

In order to identify the missing 5′ exons we synthesized primer pde1 corresponding to nt 675–700 of the bovine β-subunit cGMP PDE cDNA and hybridized this oligomer to the EcoRI digested fragments from overlapping cosmids c16Dp, c17QB1, and c18E. Pde1 hybridized specifically to a 4 kb EcoRI fragment in cosmid c18E at least 16 kb distal to the last identified exon and approximately 8 kb distal to CpG-rich island II (fig 1). We sequenced this exon (90 bp) and its intronic boundaries and confirmed significant nucleotide similarities (over 82%) with the bovine and mouse homologs. Subsequently, primers pde 3 (nt 567–588) and pde 6r (nt 425–446) identified the two most 5′ exons located in the 4 kb and the adjacent 9 kb EcoRI fragments, respectively (fig1). The complete coding sequence including 400 bp of 5′ and 700 bp of 3′ flanking sequences are given in fig 2. An inserted Alu repeat within the 3′ flanking region and

```
   1 GAGCCCAGGTCTGGGCTCCCAGGAGGCCTGGGGCCACCTGTGGGACCAACCCAAACCCTCAGGCAGCTCCACCAGCAGAAAGCGTCATGCTGTGTGTGGGATAACCTCCTCAGGGGATTAGTTCTCAGGGCCTGGTCCGCTAGAAGCCAA
 151 ATGCCTGTGCCCAGCCACCCGTGACGCCTGCAACAGGCAGGAGATCCCCCAACAGTTACTCCCAGCCTTCATTCCACAGGGTCTGGTTTTCCTGGAAGGTGGGAAGTCCCAGGGTCTGAGGAGAGGGAGCGCAGGCCCCCATTTGTAGGA
 301 GTGAGTCAGCTGACCCGCCCCCGGGGTTCCTAATCTCACTAAGAAACTTGCTATGACAGGTTCCTGGAGTCCATGCGTGCCTGGAGCAGCAGCGTCTCCAGGGACAGGCAGCCACCATGAGCCTCAGTGAGGAGCAGGCCCGGAGCTTTC
   1                                                                                                                       M  S  L  S  E  E  Q  A  R  S  F

 451 TGGACCAGAACCCCGATTTTGCCCGCCAGTACTTTGGGAAGAAACTGAGCCCTGAGAATGTTGGCCGCGGCTGCGAGGACGGGTGCCCGCCGGACTGCGACAGCCTCCGGGACCTCTGCCAGGTGGAGGAGAGCACGGCGCTGCTGGAGC
  12 L  D  Q  N  P  D  F  A  R  Q  Y  F  G  K  K  L  S  P  E  N  V  G  R  G  C  E  D  G  C  P  P  D  C  D  S  L  R  D  L  C  Q  V  E  E  S  T  A  L  L  E

 601 TGGTGCAGGATATGCAGGAGAGCATCAACATGGAGCGCGTGGTCTTCAAGGTCCTGCGGCGCCTCTGCACCCTCCTGCAGGCCGACCGCTGCAGCCTCTTCATGTACCGCCAGCGCAACGGCGTGGCCGAGCTGGCCACCAGGCTTTTCA
  62 L  V  Q  D  M  Q  E  S  I  N  M  E  R  V  V  F  K  V  L  R  R  L  C  T  L  L  Q  A  D  R  C  S  L  F  M  Y  R  Q  R  N  G  V  A  E  L  A  T  R  L  F

 751 GCGTGCAGCCGGACAGCGTCCTGGAGGACTGCCTGGTGCCCCCCGACTCCGAGATCGTCTTCCCACTGGACATCGGGGTCGTGGGCCACGTGGCTCAGACCAAAAAGATGGTGAACGTCGAGGACGTGGCCGAG | TGCCCTCACTTCA
 112 S  V  Q  P  D  S  V  L  E  D  C  L  V  P  P  D  S  E  I  V  F  P  L  D  I  G  V  V  G  H  V  A  Q  T  K  K  M  V  N  V  E  D  V  A  E | C  P  H  F
                                                                                                                                          1 exon 2
 901 GCTCATTTGCTGACGAGCTCACTGACTACAAGACAAAGAATATGCTGGCCACACCCATCATGAATGGCAAAGACGTCGTGGCGGTGATCATGGCAGTGAACAAGCTCAACGGCCCATTCTTCACCAGCGAAGACGAAGAT | GTGTTCT
 161 S  S  F  A  D  E  L  T  D  Y  K  T  K  N  M  L  A  T  P  I  M  N  G  K  D  V  V  A  V  I  M  A  V  N  K  L  N  G  P  F  F  T  S  E  D  E  D | V  F
                                                                                                                                                2 exon 3
1051 TGAAGTACCTGAATTTTGCCACGTTGTACCTGAAGATCTATCACCTGAGCTACCTCCACAACTGCGAGACGCGCCGCGGCCAG | GTGCTGCTGTGGTCGGCCAACAAGGTGTTTGAGGAGCTGACGGACATCGAGAGGCAGTTCCACA
 210 L  K  Y  L  N  F  A  T  L  Y  L  K  I  Y  H  L  S  Y  L  H  N  C  E  T  R  R  G  Q | V  L  L  W  S  A  N  K  V  F  E  E  L  T  D  I  E  R  Q  F  H
                                                                                      3 exon 4
1201 AGGCCTTCTACACGGTGCGGGCCTACCTCAACTGCGAGCGGTACTCCGTGGGCCTCCTGGACATGACCAAGGAGAAG | GAATTTTTTGACGTGTGGTCTGTGCTGATGGGAGAGTCCCAGCCGTACTCGGGCCCACGCACGCCTGATG
 259 K  A  F  Y  T  V  R  A  Y  L  N  C  E  R  Y  S  V  G  L  L  D  M  T  K  E  K | E  F  F  D  V  W  S  V  L  M  G  E  S  Q  P  Y  S  G  P  R  T  P  D
                                                                                4 exon 5
1351 GCCGG | GAAATTGTCTTCTACCAAGTGATCGACTACCTCCTCCACGGCAAGGAGGAGATCAAGGTCATTCC | CACACCCTCAGCCGATCACTGGGCCCTGGCCAGCGGCCTTCCAAGCTACGTGGCAGAAAGCGGCTTT | ATTT
 308 G  R | E  I  V  F  Y  Q  V  I  D  Y  L  L  H  G  K  E  E  I  K  V  I  P |  T  P  S  A  D  H  W  A  L  A  S  G  L  P  S  Y  V  A  E  S  G  F | I
      5 exon 6                                                        6 exon 7                                                          7 exon 8
1501 GTAACATCATGAATCGTTCCGCTGACGAAATGTTCAAATTTCAG | GAAGGGGCCCTGGACGACTCCGGGTGGCTCATCAAGAATGTGCTGTCCATGCCCATCGTCAACAAGAAGGAGGAGATTGTGGGAGTCGCCACATTTTACAACA
 355 C  N  I  M  N  R  S  A  D  E  M  F  K  F  Q | E  G  A  L  D  D  S  G  W  L  I  K  N  V  L  S  M  P  I  V  N  K  K  E  E  I  V  G  V  A  T  F  Y  N
                                              8 exon 9
1651 GGAAAGACGGGAAGCCCTTTGACGAACAGGACGAGGTTCTCATGGAG | TCCCTGACACAGTTCCTGGGCTGGTCAGTGATGAACACCGACACCTACGACAAGATGAACAAGCTGGAGAACCGCAAGGACATCGCACAGGACATGGTCC
 404 R  K  D  G  K  P  F  D  E  Q  D  E  V  L  M  E | S  L  T  Q  F  L  G  W  S  V  M  N  T  D  T  Y  D  K  M  N  K  L  E  N  R  K  D  I  A  Q  D  M  V
                                                  9 exon 10
1801 TTTACCACGTGAAGTGCGACAGGGACGAGATCCAGCTCATCCTG | CCAACCAGAGCGCGCCTGGGGAAGGAGCCTGCTGACTGCGATGAGGACGAGCTGGGCGAAATCCTG | AAGGAGGAGCTGCCAGGGCCCACCACATTTGACA
 453 L  Y  H  V  K  C  D  R  D  E  I  Q  L  I  L | P  T  R  A  R  L  G  K  E  P  A  D  C  D  E  D  E  L  G  E  I  L | K  E  E  L  P  G  P  T  T  F  D
                                              10 exon 11                                               11 exon 12
1951 TCTACGAATTCCACTTCTCTGACCTGGAGTGCACCGAACTGGACCTGGTCAAATGTGGCATCCAGATGTACTACGAGCTGGGCGTGGTCCGAAAGTTCCAGATCCCCCAGGAG | GTCCTGGTGCGGTTCCTGTTCTCCATCAGCAAAG
 501 I  Y  E  F  H  F  S  D  L  E  C  T  E  L  D  L  V  K  C  G  I  Q  M  Y  Y  E  L  G  V  V  R  K  F  Q  I  P  Q  E | V  L  V  R  F  L  F  S  I  S  K
                                                                                                        12 exon 13
2101 GGTACCGTAGAATCACCTACCACAACTGGCGCCACGGCTTCAACGTGGCCCAGACGATGTTCACGCTGCTCATG | ACCGGCAAACTGAAGAGCTACTACACGGACCTGGAGGCCTTCGCCATGGTGACAGCCGGCCTGTGCCATGACA
 550 G  Y  R  R  I  T  Y  H  N  W  R  H  G  F  N  V  A  Q  T  M  F  T  L  L  M | T  G  K  L  K  S  Y  Y  T  D  L  E  A  F  A  M  V  T  A  G  L  C  H  D
                                                                   13 exon 14
2251 TCGACCACCGCGGCACCAACAACCTGTACCAGATGAAG | TCCCAGAACCCCTTGGCTAAGCTCCACGGCTCCTCGATTTTGGAGCGGCACCACCTGGAGTTTGGGAAGTTCCTGCTCTCGGAGGAG | ACCCTGAACATCTACCAGA
 599 I  D  H  R  G  T  N  N  L  Y  Q  M  K | S  Q  N  P  L  A  K  L  H  G  S  S  I  L  E  R  H  H  L  E  F  G  K  F  L  L  S  E  E | T  L  N  I  Y  Q
                                     14 exon 15                                                                            15 exon 16
2401 ACCTGAACCGGCGGCAGCACGAGCACGTGATCCACCTGATGGACATCGCCATCATCGCCACGGACCTGGCCCTGTACTTCAA | GAAGAGAGCGATGTTTCAGAAGATCGTGGATGAGTCCAAGAACTACCAGGACAAGAAGAGCTGGG
 647 N  L  N  R  R  Q  H  E  H  V  I  H  L  M  D  I  A  I  I  A  T  D  L  A  L  Y  F  K |  K  R  A  M  F  Q  K  I  V  D  E  S  K  N  Y  Q  D  K  K  S  W
                                                                        16 exon 17
2551 TGGAGATCCTGTCCCTGGAGACGACCCGGAAGGAGATCGTCAT | GGCCATGATGATGACAGCCTGCGACCTGTCTGCCATCACCAAGCCCTGGGAAGTCCAGAGCAAG | GTCGCACTTCTCGTGGCTGCTGAGTTCTGGGAGCAAG
 696 V  E  I  L  S  L  E  T  T  R  K  E  I  V  M |  A  M  M  M  T  A  C  D  L  S  A  I  T  K  P  W  E  V  Q  S  K | V  A  L  L  V  A  A  E  F  W  E  Q
                                         17 exon 18                                      18 exon 19
2701 GTGACTTGGAAAGGACAGTCTTGGATCAGCAGCCCATT | CCTATGATGGACCGGAACAAGGCGGCCGAGCTCCCCAAGCTGCAAGTGGGCTTCATCGACTTCGTGTGCACATTCGTGTACAAG | GAGTTCTCTCGTTTCCACGAAG
 744 G  D  L  E  R  T  V  L  D  Q  Q  P  I | P  M  M  D  R  N  K  A  A  E  L  P  K  L  Q  V  G  F  I  D  F  V  C  T  F  V  Y  K | E  F  S  R  F  H  E
                                     19 exon 20                                                                                   20 exon 21
2851 AGATCCTGCCCATGTTCGACCGACTGCAGAACAATAGGAAAGAGTGGAAGGCGCTGGCTGATGAGTATGAGGCCAAAGTGAAGGCTCTGGAGGAGAAGGAGGAGGAGGAGAGGGTGGCAGCCAAGAAAG | TAGGCACAGAAATTTGCA
 792 E  I  L  P  M  F  D  R  L  Q  N  N  R  K  E  W  K  A  L  A  D  E  Y  E  A  K  V  K  A  L  E  E  K  E  E  E  E  R  V  A  A  K  K |V  G  T  E  I  C
                                                                                                                                      21 exon 22
3001 ATGGCGGCCCAGCACCCAAGTCTTCAACCTGCTGTATCCTGTGAGCACTGTCCCATGGGACCCTATGGCTCCCTCAATCTTCACCCACTAGGATTTGGGTTCTGCCTGTGGCTATTTGCTACAAGAGGTTAGGAAGCCCAAGAAAATGAC
 841 N  G  G  P  A  P  K  S  S  T  C  C  I  L  *

3151 TGAAGATCATTCTGGATATTTTAATTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCCGTGGCACGATCTCAGCTCACTGCAACCTCCACCTCCCAGTTCAAGCGATTCTCGTGCCTCAGCCTCC
3301 TGAGTAGCTGGGACTACAGGCGCCCACCACCACATGCTAATTTTTGTATTTTCAGTACAGATGGGGTTTCACCATATTGGGCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCACCCGCCTCAGCTTCCTGAAGTGCTGGGATTAC
3451 AGGCATGAGCCACCACGTTTAGCCTGTTTTTATAAACTGAAGCCAACTGTGAATAAACTGTAGCCTACATTACTCATCCATTTTTGGATAGTTACCACTGGGAGACCTTTGAAAAGGGTCCATGAACTCTGAAATCACTGAGAACATTTG
3601 CAGCCACACATGTACATATGTGTACACAGGTAGACAGATGGACACAGGCCGTTTCTCATCCAGTTTAGGAAAACACACATGCTCAGGAATTCAGAATAAAATAAACAGAAAACTACTTGCCCATTATCTGTGGTAAAATGATAGCGGAAG
```

Figure 2. Nucleotide sequences of exons 1 to 22 of the β-subunit of the cGMP PDE including 400 bp of 5′ and 700 bp of 3′ flanking region. Exon/intron boundaries are marked by a vertical bar. The putative stop codon TGA is indicated by an asterisk. The nucleotide sequence is numbered from the transcription start codon ATG based on the comparison to the bovine homolog position (translation initiation codon ATG with A=1) (7). The deduced amino acid sequence is given in a one letter code below. An inserted Alu repeat at the 3′ flanking region and two possible polyadenylation signals are underlined.

approximately 120 bp downstream the putative stop codon is underlined (fig 2)

Table 1 summarizes the data for the exon/intron organization of exons 1 to 22 of the β-subunit of the human cGMP PDE. Alignment of the genomic sequence to the bovine β-subunit of the cGMP PDE in the GenBank database (acc.no JO 5553) allowed a first tentative assignment of donor and acceptor splice junctions. The discrimination energy score of donor and acceptor splice recognition sequences (22, 23) was then calculated defining the most likely splice sites. This score is included in the table to indicate the strength of consensus of our assigned splice junctions with previously published sequences. This method scores a 'perfect' consensus as zero, with worsening agreement indicated by an increasing positive score. The maximum (worst) score for acceptors is 42.5 and for donors is 30.1. A survey of 764 human acceptor/donor sites found that the average acceptor score is 5.1 (standard deviation 2.4) and the average donor score is 3.16 (standard deviation 1.81) (23). Note that the donor splice junction for exon 20 contains a 'GC' rather than the 'GT' that is almost universal in donor splice recognition sequences. Despite this anomaly, the sequence has a score of 5.68, well within the range expected for true splice junctions, and suggesting that the remainder of the consensus area somehow compensates for the lack of a 'GT'.

Conservation of the human, mouse and bovine β-subunits of the cGMP PDE and comparison to the human and bovine α-subunits

Alignment of the deduced peptides of the β-subunit of the human cGMP phosphodiesterase (fig 2) to the mouse (8) and bovine (7) homologs and in addition to the human and bovine α-subunits (4) reveals significant sequence similarities (data not shown). The primary amino acid sequence from the human β-subunit is 92% and 91% identical to the mouse and bovine β-subunits, respectively, and 71% identical to the human α-subunit of the cGMP PDE gene. The bovine and mouse β-subunits have an amino acid residue similarity of 93%. This significant cross species conservation of the deduced amino acid sequences demonstrates the identity of the cloned gene as the β-subunit of the cGMP PDE. These data also indicate a close phylogenetic relationship of the α- and the β-subunits of the cGMP PDEs.

Table 1: Exon/intron organization for the human beta subunit of the cGMP PDE gene

Acceptor Sequence	Score[a]	Exon No.	Size (bp)	Donor Sequence	Score[a]	Intron (bp)
		1	469	GAG GTGGGT	2.83	>1600
TCCTCTCTTGCGGCAG T	5.92	2	151	GAT GTGAGT	3.18	>800
GCAATTCCTGTTTCAG G	6.10	3	90	CAG GTACCA	5.78	>16000
GTGCCCCTCCCTCCAG G	3.19	4	141	AAG GTGAGG	1.10	80
ACCTCTTCTCTGCCAG G	3.67	5	75	CGG GTGAGT	1.41	650
GTCTTCTGCTTCTCAG G	3.24	6	65	TCC GTAAGT	5.76	1080
TGCTCTCTGCCCACAG C	4.57	7	67	TTT GTGAGT	5.12	230
TCTGATGCTTTTTCAG A	6.34	8	48	CAG GTATCT	5.14	600
GACACCGCTCCCGCAG G	6.90	9	150	GAG GTAAGC	1.79	300
CTGCCATCCCCTCCAG T	5.67	10	144	CTG GTGCGG	4.27	1300
CTGTGTCTCTGTGTAG C	7.83	11	66	CTG GTAAGA	2.54	1400
ACAATCCCTCCCACAG A	6.08	12	147	GAG GTGGGA	3.89	>1700
CTCTCGGCTCCCCCAG G	3.83	13	108	ATG GTACGT	3.61	250
GGCCCTGTCTCTACAG A	5.25	14	111	AAG GTAGGC	3.04	500
ACCACCTGTGTAACAG T	10.12	15	87	GAG GTTGGT	5.67	580
GGCATAACCTCCGCAG A	9.25	16	101	CAA GTGCGC	5.00	250
TGTTCCCTGGGTTCAG G	6.89	17	108	CAT GTGAGC	3.25	650
CGGTTTGTGTCTGCAG G	5.22	18	64	AAG GTTAGA	4.20	300
CTTCCTGTGCCTCCAG G	3.41	19	75	ATT GTGAGT	4.02	1200
CTGTTCCTCCCACCAG C	5.26	20	84	AAG GCGAGT	5.68	1240
TCTCGACTCCCCTCAG G	4.74	21	151	AAG GTCTGG	6.69	>1400
CTGTCTTCTCTTGCAG T	3.49	22	62			

a) The acceptor and donor splice junctions have been scored as the discrimination energy (22, 23). A perfect consensus match would have a score of zero, while the worst score for acceptors would be 42.5 and for donors would be 30.1. All splice junctions score within the 95% confidence interval for known splices, except the acceptor of exon 15, which falls slightly outside of this range. The donor splice junction of exon 20 contains a 'GC' rather than the usual 'GT'. In a recent survey in GenBank Version 58, 17 such nonconforming sites have been described (46).

Northern blot analysis

We analyzed the expression pattern of the β-subunit gene of the cGMP PDE in various adult human tissues including frontal cortex, adrenal, lung, esophagus, skeletal muscle, kidney, liver, and retina (fig 3a). Northern blot filters with total RNA of these tissues were hybridized with probe K1, containing exons 18 and 19 (fig 1, fig 3a). The retinal RNA shows a major hybridization signal corresponding to a 3.5 kb transcript and a minor signal at 4.5 kb. All other tissues show no hybridization signal. In addition we hybridized the K1 probe to poly (A)+ RNA isolated from frontal cortex, skeletal muscle, adrenal, and lung total RNA. A normal exposure of the autorad does not reveal a hybridization signal in one of these tissues (fig 3b). However, an extreme overexposure detects a weak signal only in frontal cortex poly (A)+ RNA approximately 500−800 bp smaller than the major retinal transcript (data not shown).

Physical localisation of locus D4S228 containing the β-cGMP PDE gene

We have previously reported the localisation and orientation of a 460 kb cosmid contig close to the 4p telomere (1). The gene for the β-subunit of the human cGMP PDE maps to the most distal region of this contig. Fig 4 summarizes the physical mapping data obtained from PFGE analysis with three probes which allows the construction of a physical long range map of approximately 720 kb and which extends to the 4p telomere. Probes P1.2 (D4S228), a 1.5 kb single copy probe from the region proximal to CpG-rich island II (fig 1), 281 (D4S133) (27)

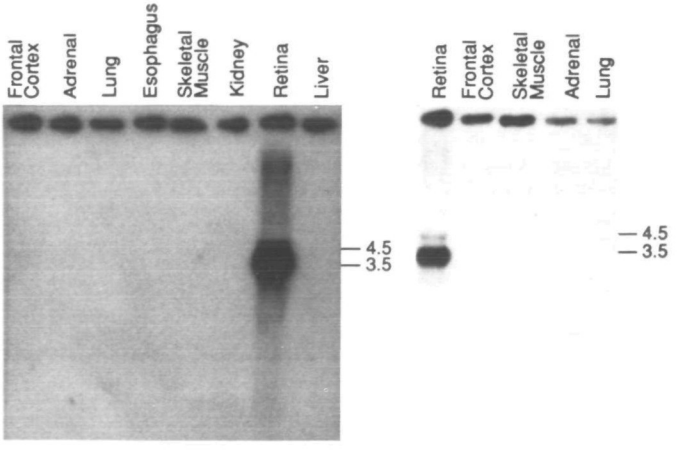

Figure 3. Northern blot analysis of 20 μg each of total (A) and 2 μg each of poly(A)+ (B) RNA from several human tissues. The retinal lane in (B) contains 5 μg of total RNA. The sizes were estimated by comparing to a commercially available RNA ladder (BRL).

and D5 (D4S90) (28) hybridize to the same 800 kb NotI fragment (fig 4). The three loci also share a common 900 kb partial MluI fragment. A partially methylated MluI site, however, separates loci D4S133 and D4S228 from D4S90. D4S133 is physically closely linked to D4S228. Both loci detect a common 90 kb NruI fragment. A SacII site which is part of CpG-rich island I (fig 1) separates the two loci and places D4S133 proximal to this CpG island.

In summary, our PFGE data demonstrate physical linkage between D4S228 and the flanking loci D4S133 and D4S90. Since it is well established that D4S90 is located within 350 kb from the 4p telomere (29, 30) we conclude that the gene for the β-subunit of the cGMP PDE is located approximately 650 to 700 kb proximal to the 4p telomere (arrow in fig 4) and represents to date the most telomeric gene characterized on 4p.

DISCUSSION

As part of the search for the Huntington disease gene we have cloned and sequenced the complete genomic locus containing the gene encoding the β-subunit of the human cGMP PDE. We identified 22 exons including 5′ and 3′ flanking regions. In addition, we determined the orientation and the precise localisation of the gene approximately 700 kb from the 4p telomere in 4p16.3. The gene is predominantly expressed in retina with a transcript size of about 3.5 kb. In addition, we also detected expression with very low abundancy and an altered sized transcript in brain.

The deduced peptide sequence of the identified gene has been aligned to known peptide sequences in GeneBank and identifies this gene unambiguously as the human homolog of the previously described β-subunits of the mouse (8) and bovine (7) cGMP PDE gene. Moreover, analysis of the C-terminus of the peptide between position 555 and 792 (Met = 1 of the β-subunit of the bovine cGMP PDE) reveals that over 85% of the amino acid residues are identical between the human α- and β-subunit of the cGMP PDEs. This supports earlier findings of a strong conservation in this region in all members of the phosphodiesterase family (5, 31) which has led to the recognition of this domain as having significant functional importance as the catalytic core for cGMP hydrolysis (32, 33).

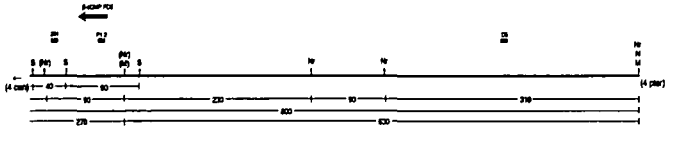

Figure 4. Long range restriction map of loci D4S133 (probe 281), D4S228 (probe P1.2), and D4S90 (probe D5). Probe P1.2 is isolated from an internal location of the β-subunit of the cGMP PDE gene (fig 1) and, therefore, localizing the gene approximately 700 kb proximal to the 4p telomere. The direction of transcription of the gene relative to the 4p telomere is indicated by the arrow. Restriction enzymes: NotI (N), MluI (M), NruI (Nr), and SacII (S). Brackets denote partially methylated enzyme recognition sites. Numbers are given in kb.

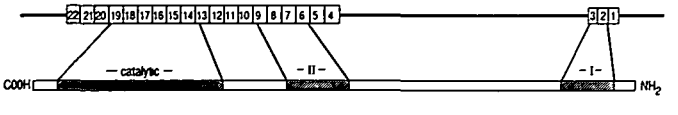

Figure 5. Correlation between genomic organization of exons 1 to 22 and the three proposed domains (catalytic core, noncatalytic binding domain I and noncatalytic binding domain II) (29, 30) in the primary amino acid sequence of the photoreceptor cGMP PDEs. The proposed domain I is encoded by the three most 5′ exons which are physically separated from the other 19 exons by at least 17 kb of intronic sequences.

To our knowledge, this is the first report providing a thorough analysis of the genomic organization of a cGMP PDE gene. Thus far only cDNAs for the human α- and γ- and the mouse and bovine α-, β-, and γ-subunits of the cGMP PDEs have been described (4, 5, 7, 33, 34). We have now sequenced the complete genomic locus encoding the proposed catalytic core, the putative noncatalytic binding domain II, and the noncatalytic cGMP-binding domain I (33, 34) (Fig 2, fig 5).

The analysis of the genomic organization of the gene reveals some interesting features. The putative catalytic core and the noncatalytic cGMP binding domain II are encoded by 19 exons which are tightly clustered within 17 kb of DNA, whereas the three most 5′ exons are physically separated by a large intron of at least 17 kb in size (fig 1, fig 5). In addition, this 17 kb intron contains a strong CpG-rich island which is defined by one SacII, one EagI, and two BssHII sites within 1kb of DNA and which reveals a G+C content of over 80% (1). CpG-rich islands have been frequently found to mark the 5′ ends of genes (35).

Interestingly, the three physically separated 5′ exons encode the putative noncatalytic cGMP binding domain I (fig 5) which can be regarded as a tandem repeat of the noncatalytic cGMP binding domain II (33, 34). Tissues other than the retina might not include the three most 5′ exons by using an alternate promotor. This alternate promotor could be represented by sequences around CpG-rich island II (fig 1) and would be expected to generate a modified transcript at least 700 bp smaller than the retina homolog. It is interesting to note that the altered size of the rare brain transcript could be explained by assuming a differential transcriptional start site in brain tissue. Alternatively, differential splicing of the primary transcript in brain could also account for a smaller transcript. Isolation of full length retina and brain cDNAs corresponding to this locus are currently under way and are required to resolve this question (Collins et al, in preparation).

Recently, evidence has been provided that the degenerative process in the retinal degeneration (rd) mouse is caused by a defect in the β-subunit of the rod cGMP PDE (8). The rd mouse is considered an animal model for autosomal recessive retinitis pigmentosa (RP) as homozygous mice have been shown to display hereditary progressive degeneration of retinal photoreceptors (36, 37). Retinal degeneration in these mice is preceded by elevated levels of cGMP in the retina as a result of deficient cGMP PDE activity (38, 39).

In human, RP is clinically and genetically heterogenous and is a major cause of blindness with a prevalence ranging from 1/3000 to 1/7000 (40, 41). In the caucasian population, RP is inherited most frequently (84%) as an autosomal recessive trait (42). However, even within this category of autosomal recessive RP, a large number of clinically distinct retinopathies can be defined including Usher syndrome, Leber's amaurosis and Stargardt's disease. The cloning of the human homolog for the mouse β-subunit will now allow assessment as to whether any form of human retinal degeneration corresponds to the previously described mouse model. Moreover, the knowledge of the genomic sequences of the exon/intron boundaries will facilitate

rapid analysis of affected patients for DNA sequence alterations without the need for retinal RNA.

The rd locus has been mapped in the mouse between the genes for α-feto protein (Afp) and β-glucuronidase (Gus) (43). This region is well outside the conserved linkage group between mouse chromosome 5 and human chromosome 4p which has been established between markers D4S62h and Kit (44, 45). If the location of the β-subunit of the cGMP PDE can be definitely mapped between Afp and Gus in the mouse, we would have to assume more than one syntenic region for distal human 4p. A chromosomal rearrangement in the mouse which interrupts a syntenic group and places more distal markers from human 4p to a more proximal position in the mouse would have to be postulated.

The β-subunit of the cGMP PDE maps to a previously described candidate region for the Huntington Disease (HD) gene. This gene, however, is predominantly expressed in human retina and the absence of any obvious retinal changes in patients with HD would appear to make it less likely that a defect in this gene causes HD. However, alternate splicing or expression of a defective gene with different tissue vulnerabilities could theoretically result in tissue specific defects. Final confirmation that this gene is not involved in HD awaits detailed sequence analysis of affected HD patients and their comparison with normal controls.

ACKNOWLEDGEMENT

The technical advice of Dr. Allen Delaney is greatly appreciated. We thank Dr. Sandra Youngman for probe D5 and Dr. Richard M. Myers for probe 281. We wish to thank C. Munroe for preparation of the figures. This research is supported by grants from MRC-Canada and the Canadian Genetic Diseases Network. M.R. Hayden is an established investigator of the Children's Hospital of B.C. Colin Collins is a predoctoral scholar of the Huntington Society of Canada and O. Riess is an MRC visiting scientist.

REFERENCES

1. Weber B, Collins C, Kowbel D, Riess O, Hayden MR (1991). Genomics, in press.
2. Baehr W, Devlin MJ, Applebury ML (1979) J Biol Chem **254**: 11669–11677
3. Deterre P, Bigay J, Forquet F, Robert M, Chabre M (1988). Proc Natl Acad Sci USA **85**: 2424–2428
4. Pittler SJ, Baehr W, Wasmuth JJ, McConnell DG, Champagne MS, Van Tuinen P, Ledbetter D, Davis RL (1990). Genomics **6**: 272–283
5. Ovchinnikov YA, Gubanov VV, Khramtsov NV, Ischenko KA, Zagranichny VE, Muradov KG, Shuraeva TM, Lipkin VM (1987). FEBS Lett **223**: 169–173
6. Pittler SJ, Baehr W (1991) In: Farber DB, Chader GJ (eds), The molecular biology of the retina. Wiley-Liss, New york, pp 33–66
7. Lipkin VM, Khramtsov NV, Vasilevskaya IA, Atabekova NV, Muradov KG,Gubanov VV, Li T, Johnston JP, Volpp KJ, Applebury ML (1990). J Biol Chem **265**: 12955–12959
8. Bowes C, Li T, Danciger M, Baxter LC, Applebury ML, Farber DB (1990). Nature **347**:677–680
9. Piriev NI, Purishko VA, Khramtsov NV, Lipkin VM (1990). Doklady Bioch Proc Acad Sci (USSR) **315**: 275–277
10. Ovchinnikov YA, Lipkin VM, Kumarev VP, Gubanov VV, Khramtsov NV, Akhmedov NB, Zagranichny VE, Muradov KG (1986). FEBS Lett **204**: 288–292
11. Tuteja N, Farber DB (1988). FEBS Lett **232**: 182–186
12. Dancinger M, Kozak CA, Li T, Applebury ML, Farber DB (1990a). Expl Eye Res **351**: 185–189
13. Dancinger M, Bowes C, Kozak CA, LaVail MM, Farber DB (1990b). Invest Ophthalmol Vis Sci **31**: 1427–1432
14. Dancinger M, Tuteja N, Kozak CA, Farber DB (1989). Expl Eye Res **48**: 303–308
15. Stryer L (1986). Ann Rev Neurosci **9**: 87–119
16. Chabre M, Deterre P (1989). Eur J Biochem **179**: 255–266
17. Mitchell GA, Brody LC, Looney JE, Steel G, Suchanek M, Dowling C, DerKaloustian V, Kaiser-Kupfer M, Valle D (1988). J Clin Invest **81**: 630–633
18. Cremers FPM, Van De Pol DJR, Van Kerkhoff LPM, Wieringa B, Ropers H-H (1990). Nature **347**: 674–677
19. Dryja TP, McGee TL, Hahn LB, Cowley GS, Olsson JE, Reichel E, Sandberg MA, Berson EL (1990). N Engl J Med **323**: 1302–1307
20. Sanger F, Nicklen S, Coulson AR (1977). Proc Natl Acad Sci USA **74**: 5463–5467
21. Brooks-Wilson AR, Goodfellow PN, Povey S, Nevanlinna HA, DeJong PJ, Goodfellow PJ (1990). Genomics **7**: 614–620
22. Berg OG, van Hippel PH (1988) J Mol Biol **200**: 709–723
23. Penotti FE (1991) J Theor Biol **150**: 385–420
24. Hein J (1990). Methods Enzym **183**: 626–645
25. Comczinski P, Sacchi N (1987). Anal Biochem **162**: 156–159
26. Church GM, Gilbert W (1984). Proc Natl Acad Sci USA **81**: 1991–1995
27. Pritchard CA, Casher D, Uglum E, Cox DR, Myers RM (1989). Genomics **4**: 408–418
28. Bucan M, Zimmer M, Whaley WL, Poustka A, Youngman S, Allitto BA, Ormondroyd E, Smith B, Pohl TM, MacDonald M, Bates GP, Richards J, Volinia S, Gilliam TC, Sedlacek Z, Collins FS, Wasmuth JJ, Shaw DJ, Gusella JF, Frischauf AM, Lehrach H (1990). Genomics **6**:1–15
29. Doggett NA, Cheng JF, Smith CL, Cantor CR (1989). Proc Natl Acad Sci USA **86**: 10011–10014
30. Bates GP, MacDonald ME, Baxendale S, Sedlacek Z, Youngman S, Romano D, Whaley WL, Allitto BA, Poustka A, Gusella JF, Lehrach H (1990). Am J Hum Genet **46**: 762–775
31. Stroop SD, Charbonneau H, Beavo JA (1989). J Biol Chem **264**, 13718–13725
32. Charbonneau H, Beier N, Walsh KA, Beavo JA (1986). Proc Natl Acad Sci USA **83**, 9308–9312
33. Charbonneau H, Prusti RK, LeTrong H, Sonnenburg WK, Mullaney PJ, Walsh KA, Beavo JA (1990). Proc Natl Acad Sci USA **87**, 288–292
34. Li T, Volpp K, Applebury ML (1990). Proc Natl Acad Sci USA **87**: 293–297
35. Lindsay S, Bird AP (1987). Nature **327**: 336–338
36. Noell WK (1958). AMA Arch Ophthalmol **60**: 702–733
37. Karli P (1963). Pr Gr Ophthalm **14**: 51–89
38. Farber DB, Lolley RN (1974). Science **186**: 449–451
39. Farber DB, Lolley RN (1976). J Cyclic Nucleotide Res **2**: 139–148
40. Bundey S, Crews SJ (1984). II Prevalence J Med Genet **21**: 417–420
41. Bunker CH, Berson EL, Bromley WC, Hayes RP, Roderick TH (1984). Amer J Ophthalmol **97**: 357–365
42. Boughman JA, Conneally PM, Nance WE (1980). Am J Hum Genet **32**: 223–235
43. Davisson MT, Roderick TH, Hillyard AL, Doolittle DP(1989). Mouse Newsletter **84**: 15–23
44. Yang-Feng TL, Ullrich A, Francke U (1987). Cytogenet Cell Genet **46**: 723
45. Cheng SV, Martin GR, Nadeau JH, Haines JL, Bucan M, Kozak CA, MacDonald ME, Lockyer JL, Ledley FD, Woo SLC, Lehrach H, Gilliam TC, Gusella JF (1989). Genomics **4**: 419–426
46. Senapathy P, Shapiro MB, Harris NL (1990). Methods in Enzymology **183**: 252–278