Volume 17 Number 8 1989 Nucleic Acids Research

Nucleotide sequence of the human α-galactosidase A gene

Ruth Kornreich, Robert J.Desnick and David F.Bishop

Division of Medical and Molecular Genetics, Mount Sinai School of Medicine, 100th Street and Fifth Avenue,
New York, NY 10029, USA

Submitted March 9, 1989 EMBL accession no. X14448

Human α-galactosidase A (α-D-galactoside galactohydrolase; EC 3.2.1.22; α-Gal A) is a lysosomal hydrolase that cleaves terminal α-galactosyl residues from glycosphingolipids and glycopeptides and is the enzyme deficient in Fabry disease (1). The gene encoding this enzyme has been assigned to the region Xq21.33→Xq22 (1). Previously, we reported the sequence of the full-length cDNA, and the genomic organization and 5' flanking sequence of this 12 kb housekeeping gene (2,3). Here, we present the complete nucleotide sequence of the gene including 1179 bp of 5' flanking and 1169 bp of 3' flanking sequence. The 12,436 bp sequence derived from α-Gal A genomic clones (3) was determined on both strands in their entirety by the method of Sanger et al. (4). The seven exons are shown in bold and a CT-rich region in intron 1 is underlined.

```
   1 CCCTTCTGTA GGGGCAGAGA GGTTCTACTT CATTACTGCG TCTCCTGGGA AGGCCATCAG GACTGCTGGC TAAAGTGGGA ACCAGGACTC TTTGTGAGTT
 101 AAGAATTTGT GTATTTATAT GTGTGTTATA CACATTTTTT AAAAAACTGT AACGACATCA GGTTGAGCAG TCGTCTCCGG GTGGTGAATT ATGTGTATTT
 201 TTAAATTTTA TACTATATTG TTATTTTTCA AATGTTCGAA ATTGAATATG TAGATTGTTG TTATCAGCAG AAAAATAAAC ATTATTCAAA TACTCTATTC
 301 AGTAAAGTAA TTTATTGGGC GCCTTTGTCA AGCACGCATT TGCCTAGATG TGACTCTACA GATAAAATTC ACTTGGGGCC TCCCCTTACA GACAATCAGG
 401 CAGTGGAGAC TGAGTGCCTG AATGGATAGA CCAGCACTCA GACCACTATT TTCAGTATCT GTTTTTCTTA ACTCAGGGCC GTGGTTTTCA AACGTTTTTC
 501 GCCTTACGGT CACCCTTAGG GTCCCCCGAG ACCGGCCCAG ACAGACAGAT ATACAAAAAC ACATACACAG TCATGAGCGT CCACCATTTC CCCACCAGGC
 601 GCAGCACAGG CGGCTTCCCG GCACTGAGAT GGGGGGGAGG AGGGAGAGAG CGCGAGGGGG GAGGGGAAAG CAGAGAACGA AAGAGGCGGA GGCGGCCCCC
 701 GAACCCCGCT CTGGTCTTCA TCATCACCAC CCCTGGGTCC CCAGTTCCCA CCCACACACC AACCTCTAAC GATACCGGGT AATTTTCCTC CTTCTTCCCT
 801 CAAACGGCTA TAGCGAGACG GTAGACGACG ACCAGAACTA CTTCTGCTCA CGTAAGCGAG TAATCACGTG AGCGCCTACG TCATGTGAGA TCTCGGTCAC
 901 GTGAGCAACT CTCGGCTTAA ACTCGGGATC ACTAAGGTGC CGCACTTCCT TCTGGTATGG AAATAGGGCG GGTCAATATC AAGAAAGGAA GAGGGTGATT
1001 GGTTAGCGGA ACGTCTTACG TGACTGATTA TTGGTCTACC TCTGGGGATA ACCGTCCCAG TTGCCAGAGA AACAATAACG TCATTATTTA ATAAGTCATC
1101 GGTGATTGGT CCGCCCCTGA GGTTAATCTT AAAAGCCCAG GTTACCCGCG GAAATTTATG CTGTCCGGTC ACCGTGACAA TGCAGCTGAG GAACCCAGAA
1201 CTACATCTGG GCTGCGCGCT TGCGCTTCGC TTCCTGGCCC TCGTTTCCTG GGACATCCCT GGGGCTAGAG CACTGGACAA TGGATTGGCA AGGACGCCTA
1301 CCATGGGCTG GCTGCACTGG GAGCGCTTCA TGTGCAACCT TGACTGCCAG GAAGAGCCAG ATTCCTGCAT CAGGTATCAG ATATTGGGTA CTCCCTTCCC
1401 TTTGCTTTTC CATGTGTTTG GGTGTGTTTG GGGAACTGGA GAGTCTCAAC GGGAACAGTT GAGCCCGAGG GAGAGCTCCC CCACCCGACT CTGCTGCTGC
1501 TTTTTTATCC CCAGCAAACT GTCCCGAATC AGGACTAGCC CTAAACTTTC TCTGTGTGAC CTTTCCTGGG ATGGGAGTCC GGCCAGCGGC CCCTGTTTCT
1601 TTCTCTCTCT CTCTCTCTCT CGTTCTCCTT CTCTTTCTCT TTCTCTTCTT TCCTCTCTCT TTCTCTCTCT CCCTGCCCGG TTCTCTTTTT TCACTGCTCC
1701 TTGCAGAGCA GGGCCACCCC ATAGGCAGTG TGCCCAAAGT AGCCCTGCCC GGTTCTATTC AGACCCTTCT TGTGAACTTC TGCTCTTCCT CTGCCGGGTG
1801 CTAACCGTTA GAACATCTAG GGTGGGTAGG AGGAATGGGG AACTAAGATT CGTGCCATTT TTTCTCCTTT TGGGGTCGTG GATTTCTCGG CAGTATCTCG
1901 AGGGAGTTAG AGAGACCATA AGGTCGCTGA GATCTCTCCC ACCTCGCCCA TGAGCGTGGC ATCAGGCTGG AAGGTTGACA TGGAGGAACT TTATACATTT
2001 ACACCTTTGC GTGAGGGTTG AGGCTGGATT AGATAGGTAT TGAACATATC TGACCCTCAC AATCCTTATC TGTAAATTGG GATTACAACC TTTTAATTTC
2101 AGGGAGCTGA CAAAAAAAAT CTGAAAAATA GTTCTTATCT CACACAGGTG AGTTTTCAAG GAGATAACCT ATTTAAAGTA CATAGCACAG CGCTTGACCA
2201 TTCAACTGCG CTTACAGAGC AAATGTTCAA TGGGAAAATG AATGTAAATC TACAAATCTG AATGAATATG TGTATTTTTC TGGAGAGAGG ATATTTACCT
2301 TTCTTCAAAT TCTCAAAGGG CTCTGTGATT TAAAAAAGGT TAGGAATCAC TGATAGATGT TGGTAAAAGG TGGCAGTCAC AGTACATTTC TGTGTCCATA
2401 AGTTATTCCT ATGAATATCT TTATAGATAA AGTCAGGATG TTGGTCAGAC ATCACAGAAG AAATTGGCCT TGTAAGTTTC ATGTGACCCT GTGGTACAGT
2501 ATGTGTGGCA ATTTTGCCCA TCACGGATTT TTTTTTATTG GTATTTGCAT CTGATTATAA AACTAATGCA TGATCATTGC AAAAAATGTA GATAAAGAAG
2601 AGCAAAATGA AAATAAAGAT TTCCCCCCAC CGTTCCACCA CCCAGAAATA ATCATGGTTT AAATGTTAAT ATACAACCTT ACAATTGTTT TCTATATAAA
2701 TGAAAACATA GATTTCTTTA TTTCATTATT TTCCATAAAA AATGGATCAT GTTTATGTCA TGTTTGGCTA ATGGCAAGAC CCTGGCACCC AGTCTGGGCT
2801 CAAATTCTGC CTCATTGTTA CTTAGCCCTG TGACATTGGG TAAATTACAC TTTTTTTTTT TTTTTTTTTT TGAGACGGGG TCTCGCTCTG TCGCCCAGGC
2901 TGGAGTGCAG TGGCACGATC TCGGCTCACT GCAAGTCCGC CTCCTGGGTT CACGCCATTC TTCTGCCTCA GCCTCCCGAG TAGCTGGGAC TACAGGCGCC
3001 TGCCACCACG CCTGGCTCTT TTTTTTTTTT TTTTTTTTTT TAGTACAGAC GGGGTTTCAC CATGTTAGCC AGGGTGGTCT CAATCTCCTG ACCTCGTGAT
3101 TCGCCCGCCT CAGCCTCCCA AAGTGCTGGT GTGAGCCACC GTGCCCAGCC TTACTTTTTT TTTTGAGAGG GGGTCTCACT CTGTCACCCA GGTTGGAGTG
3201 CAGTGGCGCG ATCTCTGCTC AGTGCAAACT CCACCTCCCG GGTTTAAGCA GTTCTCCTGT CGTAGTCTCC TGAGTAGCTG GGATTACAGG CACACCACCA
3301 CGGCCAGCTA ATTTTTGTAT TTTCAGTAGA GACGGGTTTC ACCATGTTGC CCAAGCTGGT CTCGAACTCC TGGCCTCAAG TGATCTGCCC GCCTTGGCCT
3401 CCCAGAGTGC TGGGATTACA GGTGTGAGCC ACCGCACCCG GCCTCTTTTT TCTTTTTTAG TCTATCATAC CTTGCAAATA CAGTGGTTCT TCCTATGTGT
3501 TGGTTTTGAT ATTTATGTAA TCAAACACAT CAGTTTTTCC TTTCTGATTT CTGACTTTGG GGTCATGCTG AGAAAGTCCT TTCCTACCTG AAGATAATAC
3601 AGTATATACG TTTCTTACTA GTATTTTTGT GGATTTTTAA AATATTTAAA TCTTTAGTCC ATCTGAACTT GTTCTTCTAT CAGAAATGCC ACATTTAATA
3701 AATAATAAGT CCCATGGTAT CAGATGGCTG GAAGGACCTC TTTCGAAACT TTGTTTAATT CCATTAATCT GTGTATTCTT ATTCTAATGC TAATAGTTCC
3801 ACACTAGCTT CCTTTATCTT TTTTTTCTTT TTTTTTTTTT TTTTGAGCTG GAGTTTCGCT CTTGTTGCCC AGGCTGGAGT ACAATGTCAC GATCTCGGTT
3901 CACCGCAACC TCCGCCTCCC AGGTTCAAGC AATTCTCCTG CCTCATCCTC GCGAGTAGCT GGAATTACAG GCATGCGCCA CCACGCCTAG CTATTTTGTA
4001 TTTTTAGTAG AGATGGGGTT TCTCCATGTT GGTCAGGCTG GTCTCAAACT CCCAGCCTCA GGTGATCTGC CTGCCTCGGC CTCCCAAAAT GCTGTTATTA
4101 CAGGCGTGAG CCACCACGCC CAGCCTTCAT CTTTTAATGA ATGTACATGT ATGTAATCTT TTAGGTGAAC TTTTTGTAAT GTTGTGCCAA GTTCCTTAAA
4201 AAGCCCTTTT GGAAGCTGGG CAGGTGGCCA CGCCTGTAAT CCCAGCATTT TGGGAGTCTG AGGCAGGTGG ATCACTTGAG GCCAGGAGTT CAAGACTAGC
4301 CTAGCCAAAA TGCAAAACCC TGTCTCTACT AAAGATACAA AAATTAGCCG GATGCGATGG CACATGCCTG TAATCTCAGC TACTCGGGAG GCTGAGGTAG
4401 AAGAATCGCT TGAACCGGGG AGGCAGAGGT TGCAGTGAGC AAGATGGCGC CACTGCACTC CAGCCTGGGT GACAGAGGGA GACTCCATCT CAAAAAAAAA
4501 AAAAAAAAAA AAGATAAAAA GGAAACCTAA GTACTCTTGG GCTTTGTTAA GGATTTTGTT AAATATACAA AGGATTGCAG GGAAAATTAA CTTATTTTTA
4601 ATATTGAGTA TGCTTATCCA AGAGCAAAAT AATATTTCTC CATTTATTCA AATCATTTAG GAGCATCATA GTTTTAACAT ATGGGCCTTG CACGTATCTT
4701 AAATTTATCT CTAGGCATTT TAGGTTGTTC AGTTGTTCTT GTGAATGGGA TCTTTTTCTC CAAATAGGAT TATTGTTGAT ATCTGTTGAT TATGTTAACT
4801 TTGTAGTTTC TGACTTTACT GAACTGTCTT CTTAGATCTA ATACTCTTTT CAATTTCATC ATATATTTCT CATTCCTATT TTGTTTGGGG TTTTTAGGGC
4901 GGGAATATTA ACGGGATAAG AGAGACAAAA GAAAATCTGG AAAAACAATT CATTTTACCT TACATTGCTT GTGATTACTA CCACACTATT ACTGGGTTGG
5001 AAAAAATTGT GAAATCCCAA GGTGCCTAAT AAATGGGAGG TACCTAAGTG TTCATTTAAT GAATTGTAAT GATTATTGGA ATTTCTCTTT CAGTGAGAAG
5101 CTCTTCATGG AGATGGCAGA GCTCATGGTC TCAGAAGGCT GGAAGGATGC AGGTTATGAG TACCTCTGCA TTGATGACTG TTGGATGGCT CCCCAAAGAG
5201 ATTCAGAAGG CAGACTTCAG GCAGACCCTC AGCGCTTTCC TCATGGGATT CGCCAGCTAG CTAATTATGT GAGTTTATAG ATAATGTTCT TGTTCATTCA
5301 GAGGACTGTA AGCACTTCTG TACAGAAGCT TGTTTAGAAA CAGCCCTCAT GGCCGGGCGT GGTGGCTCAC GCTGTAATCC CAACACTTTG GGAGGCCGAG
5401 GCGGGTGGAT CACCTGAGGT CAAGAGTTCA AGACCAGCCT GGCCAACATG GTGAAACCCC AACTCTATTA AAAGTACAAA AAATTAGCTG GGCATGGTGG
5501 TGAACGCCTG TAACCCCAGC TACTTGGGAG GCTGAGGCAG GAGAATCGCT TGAACCCAGG AGGTGGAAGT TTCAGTGAGC TGAGATCACG CCATTGCACT
5601 CTAGCCTGGG CAACAAAAGA GAAACTCCAT CTCAAAAAAA AAAACAAGGA AAAAAAGAAA CAGCCCTCAT GACACTTAGA AAGTAGAATA GCTGGCTGTT
5701 ATCTGAACAT TGAATTGTAA GGCTTATCAG GTGGACTTTG CATTCCATCA GCAGACAATT TTTTTTTTTT TTTTTTTTTG AGATGGAGTC TCATTCTGTC
5801 TCCCAGGCTG GAGGGCAGTG GTGCGATCTC GGCTCACTGC AAGCTCCACC TCCTGGGTTC ATGCCATTCT CCTGCCTCAG CCTCCCAAGT AGCTGGGACC
5901 ACAGGCACCC GCCACCATGC CCAGTTAATT TTTTGTATTT TTAGTAGAGA CGGGGTTTCA CCATGTTAGC CAAGATGGTC TCGATCTCCT GACCTCGTGA
6001 TCCGCCCACC TCGGCCTCCC AAAGTGCTGG GATTACAGGC ATGAGCCACC GCGCCTAGCC TACAAATGTT TTGTAATAGC TCTTGAGGCC CATCTTGGAG
```

© IRL Press

```
 6101 TTCTCCTTTT GCTAAAACCA CTGAACTCTC TAGGAGGAAA AAGGAACTTG GTTCTTGACA TATGTGTGCA TGTATTTCCA TATAACCTTT AGGAAGCTAT
 6201 TGCAATGGTA CTATAAACTA GAATTTTAGA AGATAGAAGG AAAATATTCT GGAGATCATT GAAGAGAAAT GGAGTCCAAC ACTAGTTAAA GATGATGAAG
 6301 ACAGATTTTT TTTTTTGACG GAGTCTCGCT CTGTCGCCCA GGCTGGAGTG CAGTGGCACA ATCTCAGCTC ACTGCAACCC TCCACCTCTT GGGTTCAAGT
 6401 GATTCTCCTG CCTCAGCCTC CCAAGTAGCT GGGACTACAG GCGCACACCA CCACGCCCGG CTAATTTTTG TATTTTTAGT AGAGACAAGG TTTCACCATA
 6501 TTCGCCAGGC TGGTCTCGAA CTCCTGACCT TGTAATCCGC CCACCTTGGC CTCCCAAAGT GCTGGGATTA CAGGCATGAG CCACCACGCC CGGCCGATGA
 6601 AGACAGATTT TATTCAGTAC TACCACAGTA GAGGAAAGAG CCAAGTTCAA TTCCAAATAC AACAAAGACA GGTGGAGATT TATAGCCAAT GAGCAGATTG
 6701 AGGGGGTCAG TGGATGGAAT ATTTAAGAAG ACATCAAGGG TAGGGAGCTT CTTGCTAAAG CTTCATGTAC TTAAACAAGA AGGGTGGGGG ATGAGGGAAA
 6801 TTGATCAGAT ATCAATGGTG GCAGTATTGA CTTAGCAGGA TTCTTGCTAA GAGGTCTTGC TAGGACAGAC ATAGGAAGCC AAGGTGGAGG TCTAGTCGAA
 6901 AAGAAGGCTC ATCAGAGAAG TCTAACTAAA GTTTGGTCAA GAAGAGTCTT TGTCAAGGTA AATCTATCAT TTCCCTCAAA AGGTAATTTT CAGGATCCCA
 7001 TCAGGAAGAT TAGCATGGCT GCTAGCTTTC TCCTCAGTTC TGGGCTATAG CTCACATGCC TAGTTTGAAC TAGCTCAGCA GAACTGGGGG ATTTATTCTT
 7101 TGTCTTCCAA CAAACTCATC TGGATGATTT TGGGGGTTTG TGGGGAAAAG CCCCCAATAC CTGGTGAAGT AACCTTGTCT CTTCCCCCAG CCTGGAATGG
 7201 TTCTCTCTTT CTGCTACCTC ACGATTGTGC TTCTACAATG GTGACTCTTT TCCTCCCTCT CATTTCAGGT TCACAGCAAA GGACTGAAGC TAGGGATTTA
 7301 TGCAGATGTT GGAAATAAAA CCTGCGCAGG CTTCCCTGGG AGTTTTGGAT ACTACGACAT TGATGCCCAG ACCTTTGCTG ACTGGGGAGT AGATCTGCTA
 7401 AAATTTGATG GTTGTTACTG TGACAGTTTG GAAAATTTGG CAGATGGTAA TGTTTCATTC CAGAGATTTA GCCACAAAGG AAAGAACTTT GAGGCCATGG
 7501 TAGCTGAGCC AAAGAACCAA TCTTCAGAAT TTTAAATACC CTGTCACAAT ACTGGAAATA ATTATTCTCC ATGTGCCAGA GCTCCCATCT CTTCTCTTTC
 7601 AGTTCATTAA TTAATTAATT AATTCATGTA AAATCCATGC ATACCTAACC ATAGCTAATA TTGTGCACTT ATAATTCAAG AGGGCTCTAA GAGTTAATTA
 7701 GTAATTGTAA CTCTCTATAA CATCATTTAG GGGAGTCCAG GTTGTCAATC GGTCACAGAG AAAGAAGCAT CTTCATTCCT GCCTTTCCTC AATATACACA
 7801 CCATCTCTGC ACTACTTCCT CAGAACAATC CCAGCAGTCT GGGAGGTACT TTACACAATT TAAGCACAGA GCAACTGCCT GTCCCTGCTG CTAGTTTAAA
 7901 CATGAACCTT CCAGGTAGCC TCTTCTTAAA ATATACAGCC CCAGCTGGGC ATGATGGCTC ATGCCTGTAA TCCTAGCACT TTGGGAGGCT GAGGCGGGTG
 8001 GATTACTTGA GGTCAGGAGT TCGAGACCAC CCTGGCCAAC ATGGTGAAAC CCCATCTCTA GTAAAAATAC AAAAATTAGC TGACTTTGGT GGCACATGCC
 8101 TGTAATCCCA GCTACTTGGG AAGCTGAGAC AGAAGAGTCA CTTGAACCTG GGAAACAGAG GTTGCAGTGA GCCAAGATCG CACCACTGCA CTCCACCCTG
 8201 GATGACAGAC TGAACCCCAT CTCAAAAAAT TAAAATAAAA TAAAATAAAA TAACTATATA TATAGCCCCA GCTGGAAATT CATTTCTTTC CCTTATTTTA
 8301 CCCATTGTTT TCTCATACAG GTTATAAGCA CATGTCCTTG GCCCTGAATA GGACTGGCAG AAGCATTGTG TACTCCTGTG AGTGGCCTCT TTATATGTGG
 8401 CCCTTTCAAA AGGTGAGATA GTGAGCCCAG AATCCAATAG AACTGTACTG ATAGATAGAA CTTGACAACA AAGGAAACCA AGGTCTCCTT CAAAGTCCAA
 8501 CGTTACTTAC TATCATCCTA CCATCTCTCC CAGGTTCCAA CCACTTCTCA CCATCCCCAC TGCTGTAATT ATAGCCTAAG CTACCATCAC CTGGAAAGTC
 8601 ATCCTTGTGT CTTCCCCTTT ATTTCACCAT TCATGTCCTG TCTATCAACA GTCCTTCCAC CAGTATCTCT AAAATATCTC CTGAATCAGC CCACTTCCTT
 8701 CCATCTTCAC TACATGCACC CTGGCCTTCC AAGCTACTAT CGGCTCTCAA CCAGACTGCT GGGACCACCT GATCTCTCTG CTTCCACTCT GTCTCAACCC
 8801 CCATCTATTT TCCAAGCAGC ACTAGAGTTA TCATATTAAA ATGTAAATAT CAGTTTTTTT TTTAAAGAAA AAAACCCTGA GACTTAACAG AGTTATAAAA
 8901 AATATAAATG TCATCATCAG TTCCCTGCTT AAAACCCTTA ACTCGCTTCC AATTGCACTT GGAATGAAAC CAAACTGCAC TGATCCAGCC CTTGCCTGCC
 9001 TCCCCAAAGT CCAAGGGGTC ATGGCTCTTT CCCTGGCTAC ACTGGTTTTC TTTCTGTCCC TCAACACTGC AAGCCTATTG CTGCCCCAGG GCCTTTACAC
 9101 TTGCTTTTTT TCTGCCTAGA ACAGTTCTTC CCCAAAGATT TTTAAAGGGC CGGGCTCCTT AACATTGAAG TCGCAGACCA AACGCCACAT ATGCAGACAG
 9201 TTCTTCTCTA ACTACTTTAA AATAGCCCTC TGTCCATTCA TTCTTCATCA CATTAACCTG TTTAATTTTC TTCTCAGAGC TCCACACTAT TTGGAAGTAT
 9301 TTGTTGACTT GTTACCATGT CTCCCCACTA GAGTGTAAGT TTCATGAGGG CAGGGACCTT GTCTGACTTT GACTGTATCT CTCGCATATG GTTAAGTGTT
 9401 AAATAGTTAT TTATGGAATG AATCCCTATT ATTCCCTCAT TATCTCTGCA AAATAGTCTT TTTTCTCAAC ATCTTAAACC TGATATCCCA CCTGCCTATC
 9501 TACAAACTTT TTTTTTGCGA CAGAGTCTCA CTGTCACCCA GGCTAGAGTG CAGTGGCGCC ATCTCGGCTC ACTGCAACCT CCGCCTCCCG GGTTTAAGCG
 9601 ATTCTCTTGC CTCAGCCTCC CAGTAGCTGG GATTATAGGC GTGCGCTACC ACATCTGGCT AATTTTTGTA TTTTTAGTAG AGATGGTTTC ACCATGTTGG
 9701 CCAGGCTTGT CTCGAACTCC TGACCTCAGA TGATCCACCT GCCTCGGCCT CCCAAAGTGC TGGGATTACA GGCATGAGCC ACCGTGCCCA GCCTCTACAA
 9801 ACTTTTTATT CCATTAACAA ACTATATGCT GGGATTTAAG TTTTCTTAAT ACTTGATGGA GTCCTATGTA ATTTTCGAGC TTTTAATTTT ACTAAGACCA
 9901 TTTTAGTTCT GATTATAGAA GTAAATTAAC TTTAAGGGAT TTCAAGTTAT ATGGCCTACT TCTGAAGCAA ACTTCTTACA GTGAAAATTC ATTATAAGGG
10001 TTTAGACCTC CTTATGGAGA CGTTCAATCT GTAAACTCAA GAGAAGGCTA CAAGTGCCTC CTTTAAACTG TTTTCATCTC ACAAGGATGT TAGTAGAAAG
10101 TAAACAGAAG AGTCATATCT GTTTTCACAG CCCAATTATA CAGAAATCCG ACAGTACTGC AATCACTGGC GAAATTTTGC TGACATTGAT GATTCCTGGA
10201 AAAGTATAAA GAGTATCTTG GACTGGACAT CTTTTAACCA GGAGAGAATT GTTGATGTTG CTGGACCAGG GGGTTGGAAT GACCCAGATA TGGTAAAAAC
10301 TTGAGCCCTC CTTGTTCAAG ACCCTGCGGT AGGCTTGTTT CCTATTTTGA CATTCAAGGT AAATACAGGT AAAGTTCCTG GGAGGAGGCT TTATGTGAGA
10401 GTACTTAGAG CAGGATGCTG TGGAAAGTGG TTTCTCCATA TGGGTCATCT AGGTAACTTT AAGAATGTTT CCTCCTCTCT TGTTTGAATT ATTTCATTCT
10501 TTTTCTCAGT TAGTGATTGG CAACTTTGGC CTCAGCTGGA ATCAGCAAGT AACTCAGATG GCCCTCTGGG CTATCATGGC TGCTCCTTTA TTCATGTCTA
10601 ATGACCTCCG ACACATCAGC CCTCAAGCCA AAGCTCTCCT TCAGGATAAG GACGTAATTG CCATCAATCA GGACCCCTTG GGCAAGCAAG GGTACCAGCT
10701 TAGACAGGTA AATAAGAGTA TATATTTTAA GATGGCTTTA TATACCCAAT ACCAACTTTG TCTTGGGCCT AAATCTATTT TTTTCCCTTG CTCTTGATGT
10801 TACTATCAGT AATAAAGCTT CTTGCTAGAA ACATTACTTT ATTTCCAAAA TAATGCTACA GGATCATTTT AATTTTTCCT ACAAGTGCTT GATAGTTCTG
10901 ACATTAAGAA TGAATGCCAA ACTAACAGGG CCACTTATCA CTAGTTGCTA AGCAACCACA CTTTCTTGGT TTTTCAGGGA GACAACTTTG AAGTGTGGGA
11001 ACGACCTCTC TCAGGCTTAG CCTGGGCTGT AGCTATGATA AACCGGCAGG AGATTGGTGG ACCTCGCTCT TATACCATCG CAGTTGCTTC CCTGGGTAAA
11101 GGAGTGGCCT GTAATCCTGC CTGCTTCATC ACACAGCTCC TCCCTGTGAA AAGGAAGCTA GGGTTCTATG AATGGACTTC AAGGTTAAGA AGTCACATAA
11201 ATCCCACAGG CACTGTTTTG CTTCAGCTAG AAAATACAAT GCAGATGTCA TTAAAAGACT TACTTTAAAA TGTTTATTTT ATTGCCAACT ACTACTTCCT
11301 GTCCACCTTT TTCTCCATTC ACTTTAAAAG CTCAAGGCTA GGTGGCTCAT GCCTGTAATC CCAGCACTTT GGGAGGCTGA GGCGGGCAGA TCACCTGAGG
11401 TCGGGACTTT GAGACCCGCC TGGACAACAT GGTGAAACCC CATTTCTAAT AAAAATATAA AAATTAGCCA GGTGTGGTGG CGCACCTGTG GTCCCAGCTA
11501 CTCTGGGGGC TGAGGCATGA GAATCGCTTG AACCCGGGAG TGGAGGTTGC ATTGAGCTGA GATCATGCCA CCTCACTCCA GCCTGGGCAA CAAAGATTCC
11601 ATCTCAAAAA AAAAAAAAAA GCCAGGCACA GTGGCTCATG CCTGGAATCC CAGCACTTTT GGAAGCTGAG GCAGGCAGAT CACTTGAGGT TAGGATTTCA
11701 AGACCAGCCT GGCTAACATA GTAAAGCCCT GTCTCTACTA AAAATACAAA AATTAGCCAG GTATGGTGGC GAGCTTCTGT AGCCCCAGCT ACTCAGGAGA
11801 CTGAGGCAGG AGAATCACTT GAACCCGGGA AGTGGGGGGG TGCAGTGACC CAAGATCACG CCACTGCATT CCAGCCTGGG CAACAGAGCA AGACTCCATC
11901 TCAAAAAAAA AAGTTCTATT TCCTTGAATA AAATTTTCCG AAGTTTAAAC TTTAGGAATA AAACTATTAA ACCCGTATTT ACTCATCCAG ATACCCACCC
12001 CCCTTGTTGA GATTCTCTCC CAATTATCAA AATGTGTAGC ATATTTAACT ACCAAGAGCT AAACATCATT AAGACTGAAA TGTATTAAGA AGGATGTATA
12101 GGCCAGGCAC GGTGTCTCAC GCCTGTAATC CCAACACTTT GGGAGGCCAA GTCGGGCGGA TCACGAGGTC AGGAGATGGA GACCATCCTG GCCAACATGG
12201 TGAAACCCCC TCTCTACTAA AAATACAAAA ATTAGCCAGG CAGGTGGCAG GCACCTGTAA TCCCAGCTAC TCCAGAGGCT GAGGCAGGAC AATCACTTGA
12301 ACCTGGGAGG CAGAGGCTGC AGTGAGCTGA GGTTGTACCA ATTGCACTCC AGCCTAGGTA ACGAGCAACA CTCCATCTCA AAAAAAGAAA AAAAAAAAGA
12401 TGTATAATTT GGAACTGTTA AGAGGCATTT TAAAGA
```

Acknowledgements

We are indebted to C.M. Eng, T. Fitzmaurice, Y.A. Ioannou and R.W. Gotlib for their suggestions and expert assistance and to H.J. Giordano for skillful technical help. This work was supported in part by a grant (5 R01 DK34045) from the National Institutes of Health and a grant (1-578) from the March of Dimes Birth Defects Foundation. R.K. was the recipient of a predoctoral fellowship from the National Institutes of Health (5 T32 HDO7105).

References

1. Desnick, R.J. and Bishop, D.F. in Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (eds.) The Metabolic Basis of Inherited Disease, 6th edn., McGraw Hill, New York, in press.
2. Bishop, D.F., Calhoun, D.H., Bernstein, H.S., Hantzopoulos, P., Quinn, M. and Desnick, R.J. (1986) Proc. Natl. Acad. Sci. USA, **83**, 4859-4863.
3. Bishop, D.F., Kornreich, R. and Desnick, R.J. (1988) Proc. Natl. Acad. Sci. USA, **85**, 3903-3907.
4. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, **74**, 5463-5467.

© IRL Press