**EXAMPLES OF AMINO ACID SEQUENCES OF 2 PROTEINS**

**1. Amino acid sequence of a protein involved in the production of collagen. In the genetic disease *osteogenesis imperfecta,*** **this protein is defective.**

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTGPPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEPGEPGQTGPAGARGPAGPPGKAGEDGHPGKPGRPGERGVVGPQGARGFPGTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPNGSAGPPGFPGAPGPKGEIGAVGNAGPTGPAGPRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPPGAAGTTGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPVGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPAGPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAVGTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRDGARGAHGAVGAPGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGGKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTGEVGAVGPPGFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGLPGSRGERGLPGVAGAVGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNIGPVGAAGAPGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGDFYRADQPRSAPSLRPKDYEVDATLKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMEAIKVYCDFPTGETCIRAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADHEFFVDIGPVCFK

**TOTAL NUMBER OF AMINO ACIDS: 1366**

**2. Amino acid sequence of the enzyme *phenylalanine hydroxylase*.**

**A defect in this protein causes the genetic disease *PKU***.

MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKKKDTVPWFPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNYRHGQPIPRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGFRLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGAPDEYIEKLATIYWFTVEFGLCKQGDSIKAYGAGLLSSFGELQYCLSEKPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQRIEVLDNTQQLKILADSINSEIGILCSALQKIK

**TOTAL NUMBER OF AMINO ACIDS: 452**

**DNA NUCLEOTIDE (or "base") SEQUENCES OF GENES CODING FOR THE TWO PROTEINS**

**1. Nucleotide (or base) sequence of COL1A1, a gene that codes for a protein involved in production of collagen. A mutation in this gene causes the disease *osteogenesis imperfecta*.**

atgctatcgtttgtagacaccaggacgcttcttctcctagccgtgaccctttgcctcgct gtgtcccata gtgtttccaa acttggaaag ggcgggggag ggcgggagga tgcggagggcggaggtatgc agacaacgag tcagagtttc cccttgaaag cctcaaaagt gtccacgtcctcaaaaagaa tggaaccaat ttaagaagcc agccccgtgg ccacgtccct tcccccattcgggccctcct ctgcgccccc gcaggctcct cccagctgtg gctgcccggg cccccagccccagccctccc attggtggag gcccttttgg aggcacccta gggccaggga aacttttgccgtataaatag ggcagatccg ggatttgtta ttttagcacc acggcagcag gaggtttcggctaagttgga ggtactggcc acgactgcat gcccgcgccc gccatgtacctccgccggtgacccagg gctctgcgac acaaggagtc gcatgtctaa gtgctagatgctcagctttgtggatacg cggactttgt tgctgcttgc agtaacctta tgcctagcaa catgccaatctttacaagag gaaactgtaa gaaagggccc agccggagat agaggaccac gtggagaaaggggtccacca ggccccccag gcagagatgg tgaagatggt cccacaggcc ctcctggtccacctggtcct cctggccccc ctggtctcgg tgggaacttt gctgctcagt atgatggaaaaggagttgga cttggccctg gaccaatggg cttaatggga cctagaggcc cacctggtgcagctggagcc ccaggccctc aaggtttcca aggacctgct ggtgagcctg gtgaacctggtcaaactggt cctgcaggtg ctcgtggtcc agctggccct cctggcaagg ctggtgaagatggtcaccct ggaaaacccg gacgacctgg tgagagagga gttgttggac cacagggtgctcgtggtttc cctggaactc ctggacttcc tggcttcaaa ggcattaggg gacacaatggtctggatgga ttgaagggac agcccggtgc tcctggtgtg aagggtgctggtgccccggaactccag gtcaaacagg agcccgtggt cttcctggtg agagaggacgtgttggtgcc cctggtccag ctggtgcccg tggaagtgat ggaagtgtgg gtcccgtaggtcctgctggt cctaatgggt ctgctggccc tccaggtttc ccaggtgccc ctggtcccaagggtgaaatt ggagctgttg gtaacgctgg tcctactgga cccgccggtc agtgggtctt ccaggcctct ccggccccgt tggacctcct ggtaatcctg gagcaaacggccttactggt gccaagggtg ctgctggcct tcccggcgtt gctggggctc ccggcctccctggaccccgc ggtattcctg gccctcctgg tgctgccggt actactggtg ccagaggacttgttggtgag cctggtccag ctggctccaa aggagagagc ggtaacaagg gtgagcccggctccgctggt ccccaaggtc ctcctggtcc cagtggtgaa gaaggaaaga gaggccctaatggggaagct ggatctgccg gccctccagg acctcctggg ctgagaggta gtcctggttctcgtggtctt cctggagctg atggcagagc tggcgtcatg ggccctcctg gtagtcgtggtgcaagtggc cctgctggag tccgaggacc taatggagat gctggtcgcc ctggggagcctggtctcatg ggacccagag gtcttcctgg ttcccctgga aatatcggcc ccgctggaaaagaaggtcct gtcggcctcc ctggcatcga cggcaggcct ggcccaattg gccccgttggagcaagagga gagcctggca acattggatt ccctggaccc aaaggcccca ctggtgaccctggcaaaaac ggtgataaag gtcatgctgg tcttgctggt gctcggggtg ctccaggtcctgatggaaac aatggtgctc agggacctcc tggaccacag ggtgttcaag gtggaatgaacagggt cccgctggtc ctccaggctt ccagggtctg cctggcccct caggtcccgctggtgaagtt ggcaaaccag gagaaagggg tctccatggt gagtttggtc tccctggtcctgctggtcca agaggggaac gcggtccccc aggtgagagt ggtgctgccg gtcctactggtcctattgga agccgaggtc cttctggacc cccagggcct gatggaaaca agggtgaacc tggtgtggtt ggtgctgtgg gcactgctgg tccatctggt cctagtggac tcccaggagagaggggtgct gctggcatac ctggaggcaa gggagaaaag ggtgaacctg gtctcagaggtgaaattggt aaccctggca gagatggtgc tcgtggtgct catggtgctg taggtgcccctggtcctgct ggagccacag gtgaccgggg cgaagctggg gctgctggtc ctgctggtcctgctggtcct cggggaagcc ctggtgaacg tggcgaggtc ggtcctgctg gccccaacggatttgctggt ccggctggtg ctgctggtca accgggtgct aaaggagaaa gaggaggcaaagggcctaag ggtgaaaacg gtgttgttgg tcccacaggc cccgttggag ctgctggcccagctggtcca aatggtcccc ccggtcctgc tggaagtcgt ggtgatggag gcccccctggtatgactggt ttccctggtg ctgctggacg gactggtccc ccaggaccct ctggtatttctggccctcct ggtccccctg gtcctgctgg gaaagaaggg cttcgtggtc ctcgtggtgaccaaggtcca gttggccgaa ctggagaagt aggtgcagtt ggtccccctg gcttcgctggtgagaagggt ccctctggag aggctggtac tgctggacct cctggcactc caggtcctcagggtcttctt ggtgctcctg gtattctggg tctccctggc tcgagaggtg aacgtggtctacctggtgtt gctggtgctg tgggtgaacc tggtcctctt ggcattgccg gccctcctggggcccgtggt cctcctggtg ctgtgggtag tcctggagtc aacggtgctc ctggtgaagctggtcgtgat ggcaaccctg ggaacgatgg tcccccaggt cgcgatggtc aacccggacacaagggagag cgcggttacc ctggcaatat tggtcccgtt ggtgctgcag gtgcacctggtcctcatggc cccgtgggtc ctgctggcaa acatggaaac cgtggtgaaa ctggtccttctggtcctgtt ggtcctgctg gtgctgttgg cccaagaggt cctagtggcc cacaaggcattcgtggcgat aagggagagc ccggtgaaaa ggggcccaga ggtcttcctg gcttcaagggacacaatgga ttgcaaggtc tgcctggtat cgctggtcac catggtgatc aaggtgctcctggctccgtg ggtcctgctg gtcctagggg ccctgctggt ccttctggcc ctgctggaaaagatggtcgc actggacatc ctggtacggt tggacctgct ggcattcgag gccctcagggtcaccaaggc cctgctggcc cccctggtcc ccctggccct cctggacctc caggtgtaagcggtggtggt tatgactttg gttacgatgg agacttctac agggctgacc agcctcgctcagcaccttct ctcagaccca aggactatga agttgatgct actctgaagt ctctcaacaaccagattgag acccttctta ctcctgaagg ctctagaaag aacccagctgcacatgccgtgacttgaga ctcagccacc cagagtggag cagcggttac tactggaccccaaccaaggatgcact atggaagcca tcaaagtata ctgtgatttc cctaccggcaaacct

**TOTAL NUMBER OF BASES: 4098**

**2. Nucleotide (or base) sequence of gene for protein *phenylalanine hydroxylase*. A mutation in this gene causes *PKU*.**

atgtccactgcggtcctg gaaaacccag gcttgggcag gaaactctct gactttggacaggaaacaagctatattgaa gacaactgca atcaaaatgg tgccatatca ctgatcttcactcaaagaagaagttggt gcattggcca aagtattgcg cttatttgag gagaatgatg taaacctgacccacattgaa tctagacctt ctcgtttaaa gaaagatgag tatgaatttt tcacccatttggataaacgt agcctgcctg ctctgacaaa catcatcaagatcttgaggc atgacattggtgccactgtc catgagcttt cacgagataa gaagaaagac acagtgccggttcccaagaaccattcaa gagctggaca gatttgccaa tcagattctc agctatggcggaactggatgctgaccac cctggtttta aagatcctgt gtaccgtgca agacggaagc agtttgctgacattgcctac aactaccgcc atgggcagcc catccctcga gtggaataca tggaggaagaaagaaaaca tggggcacag tgttcaagac tctgaagtcc ttgtataaacccatgcttgctatgagtac aatcacattt ttccacttct tgaaaagtac tgtggcttcc atgaagataacattccccag ctggaagacg tttctcaatt cctgcagact tgcactggtt tccgcctccgacctgtggct ggcctgcttt cctctcggga tttcttgggt ggcctggcct tccgagtcttccactgcaca cagtacatca gacatggatc caagcccatg tataccccaacctgacatctgccatgag ctgttgggac atgtgccctt gttttcagat cgcccagttttcccaggaaatt ggccttgcct ctctgggtgc acctgatgaa tacattgaaa agctcgccacaatttactgg tttactgtgg agtttgggct ctgcaaacaa ggagactcca taaaggcatatggtgctggg ctcctgtcat cctttggtga attacagtac tgcttatcag agaagccaaagcttctcccc ctggagctgg agaagacagc catccaaaat tacactgtca cggagttccagcccctgtat tacgtggcag agagttttaa tgatgccaag gagaaagtaggaactttgctgccacaata cctcggccct tctcagttcg ctacgaccca tacacccaaa ggattgaggcttggacaat acccagcagc ttaagatttt ggctgattcc attaacagtg aaattggaatcctttgcagt gccctccaga aaataaagta aagccatgga cagaatgt

**TOTAL NUMBER OF BASES: 1356**