Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 7313–7317, December 1983
Immunology

Exon–intron organization and complete nucleotide sequence of a human major histocompatibility antigen DCβ gene

(major histocompatibility complex/class II antigen)

DAN LARHAMMAR, JENS JØRGEN HYLDIG-NIELSEN, BO SERVENIUS, GÖRAN ANDERSSON, LARS RASK, AND PER A. PETERSON

Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, Box 562, S-751 22 Uppsala, Sweden

Communicated by Jean Dausset, August 15, 1983

ABSTRACT We have determined the complete nucleotide sequence of a human class II histocompatibility antigen DCβ gene. The gene spans more than 7 kilobases and contains five exons corresponding to the different domains of the DCβ polypeptide. The exon–intron organization is thus analogous to that of class II antigen α-chain genes, class I antigen heavy chain genes, and the constant parts of immunoglobulin genes, emphasizing further the evolutionary relationship among these molecules. The mature polypeptide deduced from the DCβ gene shows 93% and 88% homology, respectively, to sequences derived from two DCβ cDNA clones of other haplotypes. The allelic polymorphism of DCβ chains resides predominantly in the first extracellular domain, whereas the rest of the polypeptide is virtually constant. The exons of the DCβ gene display high homology to the corresponding exons of a murine I-Aβ gene. Also, the introns show significant homology. The DCβ chains lack eight amino acids in the cytoplasmic tail, as compared to DR and I-A β chains. This is probably due to a nonfunctional splice junction of DCβ genes, causing a separate cytoplasmic exon to be nonexpressed.

The D region of the human major histocompatibility complex (MHC) harbors genes involved in the regulation of the immune response (1). The known products of this region, the class II antigens, are polymorphic glycoproteins composed of two dissimilar subunits: an α chain of approximately 34,000 daltons and a β chain of approximately 28,000 daltons (2). The polymorphism is primarily carried by the β chains (3). Class II antigens are expressed on the surface of cells involved in various aspects of an immune response such as B lymphocytes, subpopulations of T lymphocytes, and macrophages (4).

Three different class II heterodimers, denoted DR (2), DC (5), and SB (6), have been identified by immunochemical techniques and primed lymphocyte typing. Also a fourth antigen called BR has been proposed (7). Southern blot analyses with fragments derived from β-chain cDNA clones provide evidence that the class II family contains more than three β-chain genes (8). The exact function of the different class II antigens in cell interactions is as yet unknown.

Through cloning of cDNAs corresponding to class II antigen α and β chains, we have recently shown that the two chains display structural homology to each other as well as to class I antigens and immunoglobulins (9–11), suggesting that these immunologically important proteins have evolved from a common ancestor by gene duplications.

As a step towards the understanding of the evolution and mechanisms for generation and maintenance of the polymorphism of class II antigens, we present here the nucleotide sequence of a human class II antigen β-chain gene.

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.

MATERIALS AND METHODS

Isolation of Genomic Clone. A genomic library was constructed from DNA donated by an HLA-homozygous human individual typed to be Dw4/DR4 (unpublished data) in the cosmid vector pHEP (12). The library was screened with a 627-base-pair (bp) *Ava* I fragment containing almost the entire coding part of the DCβ cDNA clone pII-β-1 [previously denoted pDR-β-1 (9)]. Screening of the cosmid library and growth and analysis of cosmid clones were performed as described (13), as were Southern blot analyses (14). DNA probes were labeled by nick-translation (15).

DNA Sequence Determination. Nucleotide sequences were determined with the chemical degradation procedure (16) and a modification of the dideoxy chain termination method (17) using exonuclease III and synthetic oligonucleotide primers (unpublished results).

RESULTS AND DISCUSSION

Isolation of a Cosmid Containing a DCβ Gene. Genomic DNA from a panel of DR-homozygous human individuals has been analyzed in Southern blotting experiments using a fragment from a DCβ cDNA clone as probe (8). All blots with DNA digested with different enzymes show one strong constant band and one or two strong polymorphic bands. The β-chain gene of cosmid clone cosII-102 accounts for the strong polymorphic DCβ bands of the genomic donor DNA. Thus, this clone was chosen for further characterization.

Exon–Intron Organization of the DCβ Gene. Two overlapping fragments containing the DCβ gene [a 7.7-kilobase (kb) *Eco*RI fragment and an 11-kb *Bam*HI fragment] were subcloned in pUC9 to facilitate sequence determination (see Fig. 1). A restriction map of the DCβ gene and the sequencing strategy are shown in Fig. 1. Exons were localized by comparison with the DCβ cDNA clone pII-β-2 (unpublished results). The nucleotide sequence and the translated amino acid sequence are shown in Fig. 2. The DCβ gene encompasses more than 7 kb and contains five exons correlating with the different domains of the DCβ polypeptide (see Fig. 3). The first exon corresponds to the 5′ untranslated region, the signal sequence, and four amino acids of the first domain. The remainder of the first domain and the second domain are encoded by exons 2 and 3, respectively. Exon 4 encodes the connecting peptide, the membrane-spanning segment, and six amino acids of the cytoplasmic tail. The last four amino acids of the cytoplasmic tail are encoded by a separate exon also containing the 3′ untranslated region. All splice junctions conform to the G-T—A-G rule (18). Overall, the exon–intron organization is analogous to that of genes for class II antigen α chains (refs. 19 and 20; unpub-

Abbreviation: bp, base pair(s).

FIG. 1. Restriction map of the DCβ gene of cosmid clone cosII-102. Exons are shown as filled boxes. The hatched box shows the nonfunctional cytoplasmic exon, whereas the open box corresponds to the cosmid vector. Subcloned fragments are indicated by bars below the map. Sequencing strategy is shown by bold arrows for the dideoxy chain termination method with synthetic primers and fine arrows for the chemical degradation procedure. At sites indicated by an asterisk, no overlapping sequence was obtained. Scale shows distances in kb.

lished data) and class I antigens (21, 22), as well as to the constant parts of immunoglobulin genes (23). This finding, together with the previously described structural homologies between the class II, class I, and immunoglobulin polypeptides (9–11), suggests that these molecules, all of which are of central importance in immune responses, have evolved from a com-

```
GGATCCCCACTTAATTTGCCCTACTGAAAGAATCCCAAGTATAAAAACAACCAGTTTTTAATCAATATTACAAAGATGTTTACTGTTGAATCGCATTTTTCTTTGGCTTCTTAAAATCCC  120
TTAGGCATTCAATCTTCAGCTCTTCTATAATTGAGAGGAAGTTTTCACCTCAAATGTTCATCCAGTGCAATTTGAAGACGTCACAGTGCCAGGGACTGGATTGAGAACCTTCACAAAAAA  240
AAATGTCTGCCTAGAGACAGATTAGGTCCTTCAGCTCCAGTGCTGATTGGTTCCTTTCCAAAGGACCATCCAATCCTGCCACGCAGGGAAACATCCACAGGTTTTTATTCTTTCTGCCAG  360

                                                           Met Ser Trp Lys Lys Ala Leu Arg Ile Pro Gly Gly Leu Arg     -19
GTACATCAGATCCATCAGGTCCAAGCTGTGTTGACTACCACTACTTTTCCCTTCGTCTCAATT ATG TCT TGG AAG AAG GCT TTG CGG ATC CCT GGA GGC CTT CGG  465
                                                                    -1  +1
Val Ala Thr Val Thr Leu Met Leu Ala Met Leu Ser Thr Pro Val Ala Glu Gly Arg Asp Ser Pro G                                  4
GTA GCA ACT GTG ACC TTG ATG CTG GCG ATG CTG AGC ACC CCG GTG GCT GAG GGC AGA GAC TCT CCC G GTAAGTGCAGGGCCACTGCTCTCCAGAGCC  562
GCCACTCTGGGAACAGGCTCTCCTTGGGCTGGGGTAGGGGGATGGTGATCTCCATGATCTCGGACACAATCTTTCATCAACATTTCCTCTCTTTGGGGAAAGAGAACGATGTTGCATTCC  682
CATTTATCTTTTAGTGATGAGGTGAGCACAGTCGGATCCCCATCCTACAGGCTTAAGCCTGGAATATAAGGAGAGAGGAAAGAGGAGACAAAGTGTACATTTACTACCAGTGACAGGACA  802
AAGTGAGCATGGGGTTATTTTTGAAGATACGAATTTCTCCAGAGACACAGCAGGATTTGTCATTTAGGCGTGCCCCAAGACTTTGCCTGACTAAATATTATGGGATCCTGCATTGGGAAA  922
TGTAAGGCAGCAATGGTGTCTGTAGTCTCCGTATTTGAGGAAAAGTTGTCTGTATTCCTGACTGACTGGAGCGTTTGTGGAGGCAAAATCTTGGTACTGAAGGAAGCTGACTGGATGACC  1042
ACAGACAGGGAGTCTTCAGGTTTCACTGATTTATGGGCAAATGGTGACTTCAGTGGGATTCAGAGACCCGAGTTGGTGGACTGAATTTAGCAGAAAGGAGGATGTAAAGAAGGAAATAA  1162
CATATATTGTGAAACCACTCATTTCAGACACAGCACAATACTTTACATAAATTCTCTCTCACTCCTTCTAACATCCTGTGTGTAGATATCATGATTTTCTTTTACACAATTATACTTGTG  1282
ATATGGATATTCTGTTACATACCTGCCCTAGCTGGTGACTGCCACAGCTTAATTGGAATCTAGTTTATCAAATTCAAAAGCTTGTGCTCTTTCCATGAATAAATGTTTCTTTCTAGGACT  1402
CGGAGGTGTAGGTCCTTTCTAACATAAAAGTGAGTGAACCTCACATGGCATTCGGAAGGGTAAATCCAGGCATGGAAGGGAGGTATTTTACCGAGGGACCAAGAGAATACGCATATCAG  1522
AACGAGGACAGGCTTAATTTCTGGACCCGTCTCATCATTCCCTTGAACTCACAGGTTTATGTGGATAATTTTATCTCTGAGGTTTCCAGGAGCTCAATGGAAAATGGGATTTCATGCGAG  1642
AGCGCCCTGATTCCCTCTAAGTGCAGAGGTCTATGTAAAATCAGCCCGACTGCCTCTTCCCTCGGTTCACAGGCTCCGGCAGGGACAGGGCTTTCCGCCCTTTCCTGCCTGCAGGAAGGC  1762
CGATTCCCGAAGCCCCCAGAGAGGGCGGGCAGGGCTGGGCAGAGCCGCCGGGCGGATCACAAGTCTGGAGCGCCAGGCACGGGCGGCGGGAACTGGAGGTCGCGCGGGCGGTTCCACAG  1882

                                                                                        lu Asp Phe Val Tyr Gln Phe Lys     12
CTCCGGGCCGGGTCAGGGCGGCGGCTGCGGGGGCGGCCGGGCTGGGGCCGGGCCGGGGCCTGACTGACCGGCCGGTGATTCCCCGCAG AG GAT TTC GTG TAC CAG TTT AAG  1993

Gly Met [Cys] Tyr Phe Thr [Asn Gly Thr] Glu Arg Val Arg Leu Val Thr Arg Tyr Ile Tyr Asn Arg Glu Glu Tyr Ala Arg Phe Asp Ser     42
GGC ATG [TGC] TAC TTC ACC [AAC GGG ACG] GAG CGC GTG CGT CTT GTG ACC AGA TAC ATC TAT AAC CGA GAG GAG TAC GCA CGC TTC GAC AGC   2083

Asp Val Gly Val Tyr Arg Ala Val Thr Pro Leu Gly Pro Pro Ala Ala Glu Tyr Trp Asn Ser Gln Lys Glu Val Leu Glu Arg Thr Arg     72
GAC GTG GGG GTG TAT CGG GCG GTG ACG CCG CTG GGG CCG CCT GCC GCC GAG TAC TGG AAC AGC CAG AAG GAA GTC CTG GAG AGG ACC CGG   2173

Ala Glu Leu Asp Thr Val [Cys] Arg His Asn Tyr Gln Leu Glu Leu Arg Thr Thr Leu Gln Arg Arg V                                94
GCG GAG TTG GAC ACG GTG [TGC] AGA CAC AAC TAC CAG TTG GAG CTC CGC ACG ACC TTG CAG CGG CGA G GTGAGCGTCGTCGCCCGTCTGCGAGGCCCA  2270
TCCTTGGCAGGGGCCCAGAGTCTCTGCCGCAGGAGGGGCGAAGGGGGCGCGGCCTCTGGAACCTGAGCCTTGTTCGTTCCACCCCGGCTGACAGGAGGAGGCGGTGGGGGTGCTGGGGGC  2390
TGGTGCATCGGAGGGGCAGGGACCTAGGGCAGAGCAGGGGGACAGGCAGAGTTGGTCAAGCTGCCTAGTTTCGCCCCATCCTCCCCGTCCGTCGGCCTCGCCCTCTGCTCTGCACGTTCT  2510
TGCCTCGTGCCTTATGCGTTTGCCTCCTCGTGCCTTACCTTTACTAAGCAGTTCTCTCTGCCCCCAATTTCCGCCCTCTTCCCCTGCCCGCCCGCCCGGCTAGCACTGCCGCACCCGGCA  2630
AGGTCCACCTACACAGCTCATGCAGTGGGAAGCTTCAGACTTGGCCTGATGGAGTTAGGGCTGCCCCACAACTGCGCGCAGGGCATCCAGCAATTACAGTTGTCAAATAAGATATTTTGA  2750
CTTTTGACTTCAAATTATTATTCATCGTAATTCTGTTTTCTTAAATGGCTCTCATTCATGGCGGAGCTCTTTGAGATGAGAGTGTTTTAATCATTGCATGCTTGGTATCTGACACATTGA  2870
CTGGCATGTGGTATGAGCTCAATGATCTTCTGTTAAATTAATGAATAAATGTGCTCAGCTGCCCAATCCACTTAGGCTCAAGGGAAAGCAGAGGATAAATAGAGCCTTAAAAATGGACTTT  2990
TATTAATTATTTTCTGTCATTTTGCTTAATTCTTTAAAGTAAACTCTTATTGACTTGGATCTTAATAGAGTTTGTGAATACAAAGTCTGAGGAAAAAAGTGTTTGCTAAAAATAAAAACA  3110
ACACTTGAATGATGTTTGTAAGGCAGTTTTAATTTCTTAGAAAAGCTGAACAAATGGCACAATGCAAAGAGCAGAAGTTTTGGAATAAATAGATTGAAGCAATTAAATTATTGGATAAAA  3230
ATAGTTTCAGGTTGCTTTTGGCTTAGGTTCTCCCCTCCCCCCATCACTATCCACTTCAGGAATAAACATTCTGAAAGTCAATTTTACCCATTTAGTGAGCACTTATTTCTAGACAGTTAC  3350
CTTATCAAATACCGTCTATGTTACGTCATTTAATCTCACAGTTGCCTGTGCATTAGAGATTAGCATCACCACTTTATATATCCTAATATTAGTACATGACAAACACTTTAAGTAATCAGC  3470
CCACAAGTACTCACCAAGACCTTAAGCCTCCCAAAGTACACAATATTCTTTATGTTCTTCACTACACATCTATAGAGTCAAAGGGACATAAAGCCTTGTTAAAGCCAGTTTTGACTAGAA  3590
GCAGCAATGAGTCTCTTCCTGTTTGATCTCCATGTTAATGGGACAAAATGATACTTTCAAGGCATTGAAAATTTATGGATTAATCAATCCCTAGGTCTGACTCCAGTGTTACCTATGCAG  3710
ATTCACAAAACTTTTAGTTTACTTTAAACTCCCTTGCCTTCTTTTGACTCACATCGTAGTGCCAGCAAGTACTTACATTTTTGCTTATTTTGGGTCTATTCCATAAAATTTATTTTATCA  3830
TCTTTCTCATAAGTTTGTGCCCTCTATTTTACTCCCAGTTCTGTGTAAGATTGAACAAATCTTATAAGGCCACATAGCTGACTGTGATTTCAGGTGGACTCCAXGAAGGAGAACCAAAGA  3950
AAAGTTCAAGTCCAAGCACAAACCGTGATTCCTTCCGGATGATGGCTCAAGAGTGCCTTTTAATTGGGGTGCAACCTGCTGACCTCAGCAAATCTCAGCTATATTTATATGTTCACATTA  4070
CAGGCTCATTAACCTAGGCTGATCTCTGCAAGGATCTCAGAATATTTTCTACAGATAACATACATGATAATATCTGATTTCAGGACAAAAAAGTAATTCTCAATAGCAAGGGAATGGAGT  4190
AGGGTAGACAGCTAGTAATTAAACTCACTTGTATGTTAAAAATAAATTAAGGAAAAAAAGAAAATGAGAGAACATATTACTAAATAAAGAAAGCATACATTAAATATTTACTATAGTTTC  4310
```

(*Fig. 2 continues on the next page.*)

```
ACACTAAGAGAATAAAGGAAATGCAATAAAGTGGCCTGAAAGATAAAGGATGAGACGTGTAAAGAGACAGGGAAAGATGTGTCATTTTTTTACTATGAGCAGCAGTCTGAGAAGATAAAG   4430
GAATCGAGTTATGGGCAAACATGATGTTTGATCAGTGTTATTTGTTTTGAAGGCCTGCCTACTTTTGTTTCAAATATTACAGACTTTTGAAATCACATTCTTTTTGTTTTTTGCTGTCTG   4550
TTACTAGATCGCACATTCTGTAAAGGCAGGGACCATGGTATGTTGTTTATCTTTGGATTCTCAGTGATTGTCATATTTATATTTGTTGAATGAATCTTAATCCAAGACTTGGACTCCAGG   4670
                                                                                               al  Glu Pro Thr Val      99
TATCTTTCCACTCTGGTTCCAAGGAGGGACCATTCCCAATGGTAGACGTGCTGTGTGGTCTCACATCTCACTCCTGTCTTTTCCTGTTTACTGCCCTCAG TG  GAG CCC ACA GTG   4784
Thr Ile Ser Pro Ser Arg Thr Glu Ala Leu Asn His His Asn Leu Leu Val [Cys] Ser Val Thr Asp Phe Tyr Pro Ala Gln Ile Lys Val   129
ACC ATC TCC CCA TCC AGG ACA GAG GCC CTC AAC CAC CAC AAC CTG CTG GTC [TGC] TCA GTG ACA GAT TTC TAT CCA GCC CAG ATC AAA GTC   4874
Arg Trp Phe Arg Asn Asp Gln Glu Glu Thr Thr Gly Val Val Ser Thr Pro Leu Ile Arg Asn Gly Asp Trp Thr Phe Gln Ile Leu Val   159
CGG TGG TTT CGG AAT GAC CAG GAG GAG ACA ACT GGC GTT GTG TCC ACC CCC CTT ATT AGG AAC GGT GAC TGG ACC TTC CAG ATC CTG GTG   4964
Met Leu Glu Met Thr Pro Gln Arg Gly Asp Val Tyr Thr [Cys] His Val Glu His Pro Ser Leu Gln Asn Pro Ile Ile Val Glu Trp A     188
ATG CTG GAA ATG ACT CCC CAG CGT GGA GAC GTC TAC ACC [TGC] CAC GTG GAG CAC CCC AGC CTC CAG AAC CCC ATC ATC GTG GAG TGG C GT   5054
AAGGGGATATTGAGTTTCTGTTACTATGGGCCCCACAAGACAAAGGGCAGAGCTCCTTCTGACCCATTCCTTCCCATCTCTTATCCCTGATGTCACTACTGAGCTGGGGATCACAGGAGA   5174
CTAGAGCACCTCTTGCTCCATGGCAAGTGCATCAGAAGAATCCTGATCTCATCACCTTTCCAGATGCTAGGGAAATTATTCTACGTACTGTTTCTCCAGATCCCAGTCCTGATAGCTCGG   5294
AGGGACTTATTATTAGGGCTGGTGACTGGGATCTTAGGGTTTAAGGTATGGATGAGTTCCTGAGGAGTGGAGATCTGCTTCCCCGCTCTCTCACCTACTCACTATACCGAAGGACCTATT   5414
GGCTGGCTTTCCCCTCCCTTAGGGGTGGTCTGAATGGAGGACXAGGTTCCTTTGACACTTTCACCTCCTGCATCTCAGACTGGACTTCAGCTCCTCAGCAGGGATGCTATGGGGTGTAGG   5534
                                  rg  Ala Gln Ser Glu Ser Ala Gln Ser Lys Met Leu Ser Gly Ile Gly Gly Phe Val Leu Gly Leu   210
GACAAACGCTGACACTCAGGCTCTGCTTCTTAG GG  GCT CAG TCT GAA TCT GCC CAG AGC AAG ATG CTG AGT GGC ATT GGA GGC TTC GTG CTG GGG CTG   5632
Ile Phe Leu Gly Leu Gly Leu Ile Ile His His Arg Ser Gln Lys G                                                             225
ATC TTC CTC GGG CTG GGC CTT ATT ATC CAT CAC AGG AGT CAG AAA G GTGAGGAACCCCAAGGGAAAAGGGGAAGATGGCTGTGACCCAGACCCTCTGTTCAGG    5736
GAGGTCCTGTCTCTAGATGTGGCTCTTTCCTCCTGACCCTGAGAGGAAGAAAACTGAGCTGGAGGTGGGAGGAGACAGGACAAGATTGGAGGAGGCATTGGAATCTGATTTTACTAGTTG   5856
AAAGGTAGCCCTGTCACACAGGTGACTGATAGAGCTTATTCCAGGATATACTTACCATTCATCATCTCATTGGCTCCTTTCCAAAAGCTTCCTCCATTAAGAGGGTCAGAGCCTTGGCCT   5976
CCTTGCCTTCTAGTGACAATTTTCTTTGTTTAGGGGATTTTAAATTAGGGTACTTAAGGCCTTGAAGAACATGAGTGGTAAGAGAATATAACTCTAATTAAGTCACATGTGTCATTTTC   6096
CTTTGGGGTGAAAGAGTGGCTGTTTGTGTAATGAGACCTTTCTCTGCATAACTTCCTTTTGTAAGACCTCAAGGGCCTCCACCAGCAGGTGATATTTCAGCCATGAGCCAGTGTGGGGGG   6216
GCACAGGTGTAAGAGGCAAGAGCATGAGCTGAATGCACCTGACCACAATGGTCTCTGTTCATGGTATATTTGCTGCTATGAGGATCAAGACTTAGGGTCGAAGTTTGCCAGTTTCTAGGA   6336
ATCTCCAGAGGTTGTTCCCCAGAACCAAGCCTTAACTTTGGTGGTATCTTCTTGTGAAATGTGAAGCCAGAACCACAGCTTAAATGTTAGACAAGAGGATGATGCCCACTTTGTGCCACA   6456
TGTTGGTGGCTACTGCCTGTAGGCATTTTCCAGTGACTGAAAGAGGCTGCTAGTGGTAGGGATGAGGTATCATCCAATTTTCTAAAAAGATTGAACCCGTTCATATTCCCCAGAAGAGTAA   6576
CAGCTGTTCCGCCACTTCCCACATATCTGCATCAAGCTGAAGTTCTGTGTCCTCACGAGCTGATTTCACCTTTGCACAGATCTTGCGGGAGGTGACAATAATACATTCTGGACCTCAGCT   6696
                                                                                           ly  Leu Leu His ***       229
TTCTCTGTCTGAAGCTGCAGGGGGCCCCTGAGGGGTGGGGGAGATTGCAGGCCCACCAGCGTACCCTGTGCTGATCATCCCTCTTCTCTCTTCTTCAG GG  CTC CTG CAC TGA   6808
CTCCTGAGACTATTTTAACTGGGATTGGTTATCACTTTTCTGTAACGCCTGCTTGTCCCTGCCCAGAATTCCCAGCTGCCTGTGTCAGCCTGTCCCCCTGAGATCAGAGTCCTACAGTGG   6928
CTGTCACGCAGCCACCAGGTCATCTCCTTTCATCCCCACCTCGAGGCTGATGGCTGTGACCCTGCTTCCTGCACTTACCCAGAGCCTCTGCCTGTGCACGGCCAGCTGCGTCTACTGAGG   7048
CCCCAAGGGGTTTCTGTTTCTATTCTCTCCTCAGACTGCTCAAGAGAAGCACATGAAAACCATTACCTGACTTTAGAGCTTTTTTACATA[ATTAAA]CATGATCCTGAGTTATCTGTATTC   7168
TCAACTTCCTTAATTGAGCAGAGGCAGGAAATCACTGCAGAATGAAGGAACATACCTGAGGTGACCCAGCCAACCTGTGCCCAGAAGGAGGGTTGTACCTGAAA   7272
```

FIG. 2. Nucleotide sequence and translated amino acid sequence of the DCβ gene. Underlined in the 5′ part of the gene are the putative "CAT," "TATA," and "cap" elements, with an arrow marking the putative cap site. Cysteine residues and the glycosylation site are within boxes. The nonfunctional cytoplasmic exon is underlined, as is the 3′ untranslated region. The polyadenylylation signal is within a box.

mon ancestor by processes involving gene duplications.

Polymorphism of DCβ Chains Resides in the First Extracellular Domain. The β chain encoded by cosII-102 has a signal peptide of 32 amino acids displaying the characteristic features of archetypal signal sequences. It contains a core of hydrophobic amino acid residues, and the amino acid preceding the mature protein has no side chain (24). The mature DCβ chain contains 229 amino acid residues (9). Four cysteines, most probably forming two disulfide loops, are present as in all class II β-chain sequences known to date. Likewise, the single glycosylation site at asparagine-19 is common to all known β chains. A hydrophobic membrane-spanning segment of 21 residues is followed by a short cytoplasmic tail of 10 residues.

Comparisons between the DCβ-chain amino acid sequences deduced from cosII-102 and two cDNA clones of other haplotypes (9, 11) reveal a conspicuously uneven distribution of amino acid replacements along the polypeptides (see Fig. 3). The first extracellular domain encoded by cosII-102 displays 84% and 80% homology to the pII-β-1 and pII-β-2 DCβ chains, respectively, whereas the second domain shows 97% and 95%

```
            -30       -20       -10        -1 | 1        10        20        30        40        50
cosII-102   MSWKKALRIPGGLRVATVTLMLAMLSTPVAEG | RDSPEDFVYQFKGMCYFT[NGT]ERVRLVTRYIYNREEYARFDSDVGVYRAVTPLGPP
pII-β-2                 D-----------I--SSL--- | --------------L-----------G--H------V--------------Q-R-
pII-β-1                                     - | ---------------------------S-S------VV--------EF----L--L-

            60        70        80        90        100       110       120       130       140
cosII-102   AAEYWNSQKEVLERTRAELDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVST
pII-β-2     V---------GA--SV-R------EVAY-GI---------------------I--------S---------------A----
pII-β-1     ----------DI---K--AV-R------------------------------------------------------A----

            150       160       170       180       190       200        210       220
cosII-102   PLIRNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSK | MLSGIGGFVLGLIFLGLGLII | HHRSQKGLLH
pII-β-2     ---------------------------------------S--T---------- | ---V----------------- | RQ--R-----
pII-β-1     ---------------------------------------S--T---------- | --------------------- | ----------
```

FIG. 3. Comparison of amino acid sequences deduced from the DCβ gene and two DCβ cDNA clones, pII-β-1 (9) and pII-β-2 (11). The standard one-letter code is used; a horizontal line indicates identity with the amino acid in the sequence above it. Arrows show exon boundaries. The boundary between the signal peptide and the mature protein and the boundaries of the membrane-spanning segment are marked with vertical bars. The glycosylation site is within the box.

homology, respectively. Thus, within one subset of class II β chains, the first extracellular domain carries most of the allelic polymorphisms, whereas the second domain appears virtually constant (25). In the nucleotide sequences encoding the first domains of the three DCβ chains, mutations in the first and second positions of codons are at least as prevalent as third-base mutations. This pattern of nucleotide substitutions is also found in the variable domain exons of immunoglobulins (26–28), particularly in the codons of the hypervariable regions (29). Models involving either positive selection favoring substitutions leading to amino acid replacements (26) or suppression of silent mutations (28) have been proposed. Whereas the class I antigen polymorphism may be due partly to gene conversion events (30, 31), the DCβ pattern of nucleotide substitutions is consistent with multiple independent mutational events, as has been suggested for the immunoglobulins (27). In the nucleotide stretches encoding the DCβ second domain, the pattern of substitutions is as expected for a conserved region—i.e., silent third-base substitutions in the codons are more frequent than first- and second-base substitutions. At only 4 of 34 variant positions do all three DCβ chains have unique amino acids. This low number may be due to structural constraints, allowing only a few different amino acids at a certain position.

Exon 4, encoding the connecting peptide, the membrane-spanning segment, and the first six amino acids of the cytoplasmic tail, is highly conserved. Its translated amino acid sequence is identical to that of the cDNA clone pII-β-1 and differs from that deduced from pII-β-2 at only four positions.

The 3′ untranslated region of the DCβ gene in cosII-102 shows 98% and 92% homology to pII-β-1 and pII-β-2, respectively. The polyadenylylation signal A-T-T-A-A-A (ref. 9; see also ref. 32 and references therein) is present 330 bp downstream of the termination codon. Although the sequence determination was extended more than 100 bp 3′ of the region, no additional polyadenylylation signal was found.

In all four exons available for comparison, the DCβ gene shows greater homology to pII-β-1 than to pII-β-2. The mature DCβ chain deduced from the gene displays an overall homology of 93% and 88%, respectively, to the amino acid sequences deduced from pII-β-1 and pII-β-2. The DCβ chain encoded by cosII-102 thus appears to be more closely related to that deduced from pII-β-1 than to that of pII-β-2.

DCβ Genes Lack a Cytoplasmic Exon. The cytoplasmic tail of DCβ chains is eight amino acids shorter than that of DRβ chains (11). The nucleotide sequence of the murine I-Aβ gene (33), the structural homologue of DCβ, has provided evidence that a separate exon of 24 nucleotides accounts for this difference. A nucleotide stretch of 24 bp is indeed present in the DCβ gene, displaying clear homology to the corresponding segments of the DRβ cDNA clones and the Aβ gene. The absence of this segment in the two DCβ cDNA clones whose sequences have been determined to date is probably due to an altered 5′ splice junction of this exon. Instead of the preferred sequence A-G$\downarrow$G (18), A-A$\downarrow$G is found. Whether the difference in cytoplasmic tails between DCβ and DRβ chains has any functional significance remains to be investigated. Although this exon is not expressed in the DCβ genes, it remains well conserved at the genomic level (21 of the 24 nucleotides are identical with the Aβ cytoplasmic exon). However, the entire intron between exons 4 and 5 of the DCβ gene is highly homologous to the corresponding stretch of the Aβ gene, indicating that there could be a general conservative pressure acting on the nucleotides of this part of the DCβ gene. Also, the third DCβ intron shows clear homology to the corresponding Aβ intron. Significant homology can also be detected in parts of intron 1 and intron 2 by using a computer alignment program (34).

Promoter Region. The transcriptionally important DNA sequences G-G-$^{C}_{T}$-C-A-A-T-C-T (CAT) and TATA are expected at about 70 bp and 30 bp upstream from the cap site of transcribed genes, respectively (18). Because no full-length DCβ cDNA clone is available that can define the cap site, assignment of these elements can only be tentative. A computer comparison of the 5′ part of the DCβ gene with the corresponding region of the Aβ gene aligns a TATA-like sequence and a possible cap sequence with the proposed Aβ TATA and cap sequences, respectively (33). These putative TATA and cap elements, as well as a CAT-like sequence, are underlined in Fig. 2. Provided the proposed cap site is correct, the DCβ mRNA will have a 5′ untranslated region of 329 nucleotides, unless an intron occurs in this region. The homology between the DCβ gene and the Aβ gene in the region extending from the putative cap sites to the initiation codons is about 60%, counting insertions as mismatches.

The nucleotide stretches flanking the first domain exon are extremely rich in G+C, a feature shared with the nucleotide stretches flanking the first domain exon of the Aβ gene (33) as well as the first and second domain exons of class I heavy chain genes in man (21) and mouse (22), but not of the DRα (19), I-Eα (ref. 20; unpublished data), and β_2-microglobulin (35) genes. The significance of the G+C-rich stretches is unclear. However, it is interesting to note that they flank exons encoding polymorphic domains.

Concluding Remarks. The DCβ-chain gene reported here corresponds to the highly polymorphic DCβ gene detected in Southern blotting analyses of genomic DNA. The gene displays all the characteristics of a functional gene. The polymorphic region of DCβ chains is predominantly located in the first extracellular domain. The exon–intron organization of the DCβ gene is similar to that of class II antigen α-chain genes, class I antigen heavy chain genes, and the constant parts of immunoglobulin genes; the similarity supports previous evidence for an evolutionary relationship among these molecules.

We thank Ms. E. Rossi for help with computer analyses, Ms. P. Ågren for technical assistance, and Dr. P. Lind and Mr. P. Wenkler for preparation of oligonucleotide primers. This work was supported by grants from the Swedish Cancer Society, King Gustav V's 80-Years Fund, and Marcus Borgström's Fund.

1. Benacerraf, B. (1981) *Science* **212**, 1229–1238.
2. Klareskog, L., Sandberg-Trägårdh, L., Rask, L., Lindblom, J. B., Curman, B. & Peterson, P. A. (1977) *Nature (London)* **265**, 248–251.
3. Charron, D. J. & McDevitt, H. O. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6567–6571.
4. Hämmerling, G. J. (1976) *Transplant. Rev.* **40**, 64–82.
5. Tosi, R., Tanigaki, N., Centis, D., Ferrara, G. B. & Pressman, D. (1978) *J. Exp. Med.* **148**, 1592–1611.
6. Shaw, S., Johnson, A. H. & Shearer, G. M. (1980) *J. Exp. Med.* **152**, 565–580.
7. Tanigaki, N. & Tosi, R. (1982) *Immunol. Rev.* **66**, 5–37.
8. Böhme, J., Owerbach, D., Denaro, M., Lernmark, Å., Peterson, P. A. & Rask, L. (1983) *Nature (London)* **301**, 82–84.
9. Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. & Peterson, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3687–3691.
10. Larhammar, D., Gustafsson, K., Claesson, L., Bill, P., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P. A. & Rask, L. (1982) *Cell* **30**, 153–161.
11. Larhammar, D., Andersson, G., Andersson, M., Bill, P., Böhme, J., Claesson, L., Denaro, M., Emmoth, E., Gustafsson, K., Hammerling, U., Heldin, E., Hyldig-Nielsen, J. J., Lind, P., Schenning, L., Servenius, B., Widmark, E., Rask, L. & Peterson, P. A. (1983) *Hum. Immunol.*, in press.
12. Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 6715–6732.

13. Grosveld, F. G., Dahl, H.-H. M., de Boer, E. & Flavell, R. A. (1981) *Gene* **13**, 227–237.
14. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
15. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
19. Das, H. K., Lawrence, S. K. & Weissman, S. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3543–3547.
20. Mathis, D. J., Benoist, C. O., Williams, V. E., II, Kanter, M. R. & McDevitt, H. O. (1983) *Cell* **32**, 745–754.
21. Malissen, M., Malissen, B. & Jordan, B. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 893–897.
22. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. & Hood, L. (1981) *Cell* **25**, 683–692.
23. Gough, N. (1981) *Trends Biochem. Sci.* **6**, 203–205.
24. Kreil, G. (1981) *Annu. Rev. Biochem.* **50**, 317–348.
25. Kaufman, J. F. & Strominger, J. L. (1982) *Nature (London)* **297**, 694–697.
26. Baltimore, D. (1981) *Cell* **24**, 592–594.
27. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
28. Selsing, E., Miller, J., Wilson, R. & Storb, U. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4681–4685.
29. Rechavi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J. B., Zakut, R. & Givol, D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4405–4409.
30. Boss, J. M., Gillam, S., Zitomer, R. S. & Smith, M. (1981) *J. Biol. Chem.* **256**, 12958–12961.
31. Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 242–246.
32. Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. & Flavell, R. A. (1983) *EMBO J.* **2**, 453–462.
33. Larhammar, D., Hammerling, U., Denaro, M., Lund, T., Flavell, R. A., Rask, L. & Peterson, P. A. (1983) *Cell* **34**, 179–188.
34. Orcutt, B. C., Dayhoff, M. O. & Barker, W. C. (1982) *ALIGN: Alignment Score Program*, NBR Report 820501-08710 (National Biomedical Research Foundation, Washington, DC).
35. Parnes, J. R. & Seidman, J. G. (1982) *Cell* **29**, 661–669.