The EMBO Journal vol.3 no.2 pp.447–452, 1984

Both α and β chains of HLA-DC class II histocompatibility antigens display extensive polymorphism in their amino-terminal domains

Lena Schenning, Dan Larhammar*, Per Bill, Klas Wiman[2], Ann-Kristin Jonsson, Lars Rask[1] and Per A. Peterson

Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, and [1]Swedish University for Agricultural Sciences, Box 562, S-751 22 Uppsala, Sweden

[2]Present address: Memorial Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10021, USA
*To whom reprint requests should be sent
Communicated by P.A. Peterson

At least three class II antigens, all composed of an α and a β subunit, are encoded in the human major histocompatibility complex, i.e., DR, DC and SB. Two cDNA clones, encoding a DCα and a DCβ chain, respectively, were isolated from a cDNA library of the lymphoblastoid cell line Raji (DR3,w6). The two polypeptides predicted from the nucleotide sequences of these clones are each composed of a signal peptide, two extracellular domains, a hydrophobic transmembrane region and a short cytoplasmic tail. Comparison of the DCα sequence with two previously published partial sequences shows that the majority of the differences is located in the amino-terminal domain. The differences are not randomly distributed; a cluster of replacements is present in the central portion of the amino-terminal domain. Likewise, the allelic polymorphism of the DCβ chains occurs preferentially in the amino-terminal domain, where three minor clusters of replacements can be discerned. The non-random distribution of the variability of DCα and β chains may be due to phenotypic selection against replacement substitutions in the second domains of the polypeptides.

Key words: cDNA/class II antigens/HLA/MHC/polymorphism

Introduction

The class II antigens of the major histocompatibility complex (MHC) are expressed on B lymphocytes, macrophages and activated T lymphocytes (Thorsby, 1977; Evans *et al.*, 1978) and participate in interactions between different cell types of the immune system (Benacerraf, 1981). They are dimers composed of an α chain of ~34 000 daltons non-covalently associated with a β chain of ~28 000 daltons (Klareskog *et al.*, 1977). Three distinct class II antigens have been identified in man, denoted DR, DC and SB (Klareskog *et al.* 1977; Tosi *et al.*, 1978; Shaw *et al.*, 1980). Two additional class II heterodimers have recently been described; BR and FA (Tanigaki and Tosi, 1982; Watson *et al.*, 1983).

cDNA clones sequenced in our and other laboratories have provided the complete primary structures of DRα and β chains (Larhammar *et al.*, 1982a; Lee *et al.*, 1982; Long *et al.*, 1983; Gustafsson *et al.*, in preparation). Both subunits are composed of two extracellular domains of similar size, a membrane-spanning hydrophobic segment, and a short cytoplasmic tail. The DC subunits are similarly organized (Larhammar *et al.*, 1982b; Auffray *et al.*, 1982). The extensive homologies between the different class II antigens suggest that they have evolved from a common ancestor through processes involving gene duplications.

A remarkable feature of the class II histocompatibility antigens is their extensive polymorphism. The exact role of the polymorphism is as yet unknown, as is the mechanism of action of the class II antigens. The polymorphism of the DR antigens is almost exclusively limited to the β chain (Charron and McDevitt, 1979; Silver and Ferrone, 1979; Gustafsson *et al.*, in preparation). Less is known about the polymorphism of the DC antigens. In contrast to the DRα chain, evidence from genomic hybridizations and partial DCα sequences (Auffray *et al.*, 1982, 1983; Götz *et al.*, 1983) suggest that the DCα chain is polymorphic, as is the DCβ chain (Larhammar *et al.*, 1982b, 1983b; Böhme *et al.*, 1983; Wake *et al.*, 1982). This dual polymorphism of the DC chains may be a reflection of a function distinct from that of the DR chains.

Apart from their obvious importance in a number of immunological functions, the class II antigens of the MHC serve as an excellent model system for the analysis of genetic polymorphism at the DNA level. Thus, by determining a large number of sequences, it may eventually be possible to gain insight into the mechanisms generating the polymorphism. To this end, we describe here the complete sequences of two cDNA clones encoding HLA-DC antigen chains.

Results

Isolation and characterization of DC cDNA clones

Two cDNA libraries were made from Raji cell mRNA (Wiman *et al.*, 1982). The library enriched for DRα mRNA by size fractionation was screened with a fragment from a DCα cDNA clone (Auffray *et al.*, 1982) as probe. Out of ~10 000 clones, 38 reacted with the probe. A clone denoted pII-α-5, containing a long insert, was chosen for further characterization. Its nucleotide sequence was determined as shown in Figure 1. The nucleotide sequence and the deduced amino acid sequence are shown in Figure 2.

The library enriched for β-chain mRNA was screened with a fragment from the DCβ cDNA clone pII-β-1 (Larhammar *et al.*, 1982b) as probe. Out of 20 000 clones, 90 reacted with the probe. A clone denoted pII-β-2 was chosen for further characterization since its restriction map showed similarities to that of pII-β-1, yet displayed unique features. The clone pII-β-2 was sequenced as described in Figure 1. The nucleotide sequence and the translated amino acid sequence are shown in Figure 4.

DCα nucleotide sequence and predicted amino acid sequence

Ten nucleotides precede the AUG initiation codon in the DCα clone (see Figure 2). The first 23 amino acids constitute the signal peptide. Like most other signal sequences (von Heijne, 1983), the DCα signal sequence has a core of hydrophobic amino acids, and residues with small side chains at positions −3 (Cys) and −1 (Gly). The processed DCα chain

consists of 231 amino acids, 193 of which constitute the extracellular part. The DCα chain predicted from pII-α-5 is one amino acid shorter than that predicted from pDCH1 (Auffray *et al.*, 1982) and that described by Götz *et al.* (1983) due to a deletion at position 56 (Figure 5). A stretch of 23 mainly hydrophobic residues comprises the trans-membrane region, and 15 residues reside on the cytoplasmic side of the plasma membrane. Three cysteines are present in the extracellular part. Those at positions 109 and 165 are conserved in DCα and DRα chains (Auffray *et al.*, 1982; Larhammar *et al.*, 1982a; Götz *et al.*, 1983) as well as in their murine homologues (Benoist *et al.*, 1983; Hyldig-Nielsen *et al.*, 1983; Mathis *et al.*, 1983), and form an intradomain disulfide bond. The cysteine at position 47 of the pII-α-5 DCα chain is not found in the other DCα chains. It is located in what appears to be a highly polymorphic region (see Figure 5). A cysteine is also present at position 11 in the DCα sequence reported by Götz *et al.* (1983). Like the DRα chain and the murine α chains, the DCα chains have a cysteine in the trans-membrane region (position 197). Since this residue is conserved in all α chains known, it may have a specific function like serving as the attachment site for fatty acid. The attachment sites for N-linked carbohydrate at positions 80 and 120 are also conserved in all human and murine α chains known to date.

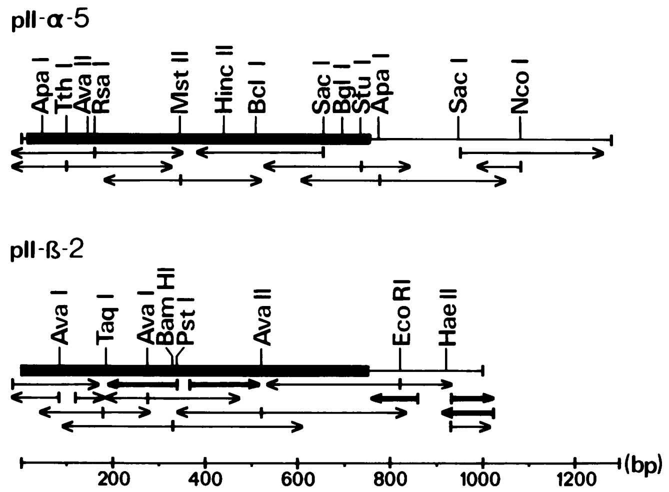

Fig. 1. Restriction maps of the HLA-DC cDNA clones pII-α-5 and pII-β-2. Coding parts of the inserts are indicated by filled boxes. Arrows show sequences determined by the chemical degradation procedure after labelling at 5′ ends (fine arrows) or 3′ ends (bold arrows).

The 3′-untranslated region of pII-α-5 encompasses ~500 bp, whereas that of pDCH1 consists of only 130 bp. Comparision with a DCα gene (A.-K. Jonsson, unpublished data) shows that part of the difference between the two cDNA clones can be accounted for by differential splicing of the intron separating the two exons encoding the 3′-untranslated region (Figure 3). The mRNA transcribed from the gene encoding pII-α-5 has retained ~210 bp of the intron preceding the last exon of the gene encoding pDCH1 (see Figure 2). Hybridization with a DCα restriction fragment to mRNA isolated from Raji cells, the heterozygous cell line that served as the source of the cDNA libraries, reveals two distinct mRNA species (data not shown). However, no cDNA clone representing the shorter DCα mRNA species has been found in the Raji library. A reasonable explanation for this negative result is the fact that the cDNA library used was constructed from mRNA enriched by size fractionation for DRα message, which is several hundred nucleotides longer than the shorter DCα mRNA species. Whether the intron segment retained in pII-α-5 is a unique feature of one of the Raji alleles or will occur also in other allelic products remains to be seen.

The polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976) found in pDCH1 corresponds to AACAAA in pII-α-5, which is consistent with the presence of ~160 additional nucleotides at the 3′ end of pII-α-5, as compared with pDCH1. The aberrant polyadenylation signal of pII-α-5 apparently does not prevent transport of the mRNA from the nucleus to the cytoplasm, probably because correct poly(A)

```
           -23                                                                                  -1  +1
           Met Ile Leu Asn Lys Ala Leu Met Leu Gly Ala Leu Ala Leu Thr Thr Val Met Ser Pro Cys Gly Gly Glu Asp Ile Val        4
TGGGAAGAGG ATG ATC CTA AAC AAA GCT CTG ATG CTG GGG GCC CTT GCC CTG ACC ACC GTG ATG AGC CCC TGT GGA GGT GAA GAC ATT GTG       91

Ala Asp His Val Ala Ser Tyr Gly Val Asn Leu Tyr Gln Ser Tyr Gly Pro Ser Gly Gln Tyr Thr His Glu Phe Asp Gly Asp Glu Gln       34
GCT GAC CAC GTC GCC TCT TAT GGT GTA AAC TTG TAC CAG TCT TAC GGT CCC TCT GGC CAG TAC ACC CAT GAA TTT GAT GGA GAT GAG CAG      181

Phe Tyr Val Asp Leu Gly Arg Lys Glu Thr Val Trp Cys Leu Pro Val Leu Arg Gln Phe Arg Phe Asp Pro Gln Phe Ala Leu Thr Asn       64
TTC TAC GTG GAC CTG GGG AGG AAG GAG ACT GTC TGG TGT TTG CCT GTT CTC AGA CAA TTT AGA TTT GAC CCG CAA TTT GCA CTG ACA AAC      271
                                                                                       ↓
Ile Ala Val Leu Lys His Asn Leu Asn Ser Leu Ile Lys Arg Ser [Asn Ser Thr] Ala Ala Thr Asn Glu Val Pro Glu Val Thr Val Phe       94
ATC GCT GTC CTA AAA CAT AAC TTG AAC AGT CTG ATT AAA CGC TCC [AAC TCT ACC] GCT GCT ACC AAT GAG GTT CCT GAG GTC ACA GTG TTT      361

Ser Lys Ser Pro Val Thr Leu Gly Gln Pro Asn Ile Leu Ile [Cys] Leu Val Asp Asn Ile Phe Pro Pro Val Val [Asn Ile Thr] Trp Leu      124
TCC AAG TCT CCC GTG ACA CTG GGT CAG CCC AAC ATC CTC ATC [TGT] CTT GTG GAC AAC ATC TTT CCT CCT GTG GTC [AAC ATC ACA] TGG CTG      451

Ser Asn Gly His Ser Val Thr Glu Gly Val Ser Glu Thr Ser Phe Leu Ser Lys Ser Asp His Ser Phe Phe Lys Ile Ser Tyr Leu Thr      154
AGC AAT GGG CAC TCA GTC ACA GAA GGT GTT TCT GAG ACC AGC TTC CTC TCC AAG AGT GAT CAT TCC TTC TTC AAG ATC AGT TAC CTC ACC      541
                                                                                                             ↓
Leu Leu Pro Ser Ala Glu Glu Ser Tyr Asp [Cys] Lys Val Glu His Trp Gly Leu Asp Lys Pro Leu Leu Lys His Trp Glu Pro Glu Ile      184
CTC CTC CCT TCT GCT GAG GAG AGT TAT GAC [TGC] AAG GTG GAG CAC TGG GGC CTG GAC AAG CCT CTT CTG AAA CAC TGG GAG CCT GAG ATT      631

Pro Ala Pro Met Ser Glu Leu Thr Glu Thr Val Val Cys Ala Leu Gly Leu Ser Val Gly Leu Val Gly Ile Val Val Gly Thr Val Phe      214
CCA GCC CCT ATG TCA GAG CTC ACA GAG ACT GTG GTC TGC GCC CTG GGA TTG TCT GTG GGC CTC GTG GGC ATT GTG GTG GGC ACT GTC TTC      721
                                                                                           ↓
Ile Ile Arg Gly Leu Arg Ser Val Gly Ala Ser Arg His Gln Gly Pro Leu ***                                                      231
ATC ATC CGA GGC CTG CGT TCA GTT GGT GCT TCC AGA CAC CAA GGG CCC TTG TGA ATCCCATCCTGGAATGGAAGGTGCATCGCCATCTACAGGAGCAGAAGA      823

GTGGACTTGCTACATGACCTAGCATTATTTTCTGGCCCCATTTATCATATCCCTTTTCTCCTCCAAATGTTTCTCCTCTCACCTCTTCTGTGGGACTTAAATTGCTATATCTGCTCAGAG     943
                                                   (↓)
CTCACAAATGCCTTTGAATTATTTCCCTGACTTCCTGATTTTTTTCTTCTTAAGTGTTACCTACTAAGAGTTGCCTGGAGTAAGCCACCCAGCTACCTAATTCCTCAGTAACCTCCATCT    1063

ATAATCTCCATGGAAGC[AACAAA]TTCCCTTTATGAGATATATGTCAAATTTTTCCATCTTTCATCXAGGGCTGACTGAAACCGTGGCTAAGAATTGGGAGACTCTCTTGTTTCAAGCCAA    1183

TTTAACATCATTTACCAGATCATTTGTCATGTCCAGTAACACAGAAGCAACCAACTACAGTATAGCCTGATAACATGA                                               1261
```

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the DCα cDNA clone pII-α-5. Cysteines involved in the intramolecular disulfide bond and the attachment sites for N-linked carbohydrates are within boxes, as is the aberrant polyadenylation signal. Arrows mark exon boundaries. The splice junction utilized in the gene encoding pDCH1 is indicated by the arrow within parentheses.

signals are present further downstream in the DCα gene (see Figure 3). Like most cDNA clones isolated from the two Raji libraries, pII-α-5 may be truncated at the 3′ end.

DCβ nucleotide sequence and predicted amino acid sequence

The insert of pII-β-2 starts at a position corresponding to amino acid −21 in the signal peptide (see Figure 4). (The signal sequence of the only DCβ chain whose sequence is complete is 32 amino acids long; Larhammar *et al.*, 1983b.) The processed DCβ chain encoded by pII-β-2 displays the characteristic features of other human and murine class II β chains (Larhammar *et al.*, 1982b, 1983a, 1983b; Kratzin *et al.*, 1981; Long *et al.*, 1983; Choi *et al.*, 1983; Malissen *et al.*, 1983; Saito *et al.*, 1983; Gustafsson *et al.*, in preparation). Four cysteines are present in the extracellular part, forming two intradomain disulfide loops. A single site is available for attachment of N-linked carbohydrate at position 19. The hydrophobic trans-membrane region encompasses 21 amino acids. The intra-cytoplasmic segment is shorter than those of DRβ and murine β chains due to a splice junction mutation causing a separate cytoplasmic 'exon' encoding eight amino acids to be non-expressed (Hyldig-Nielsen *et al.*, 1984). The 3′-untranslated region of pII-β-2 ends prematurely 90 bp upstream of the polyadenylation signals (cf. pII-β-1; Larhammar *et al.*, 1982b).

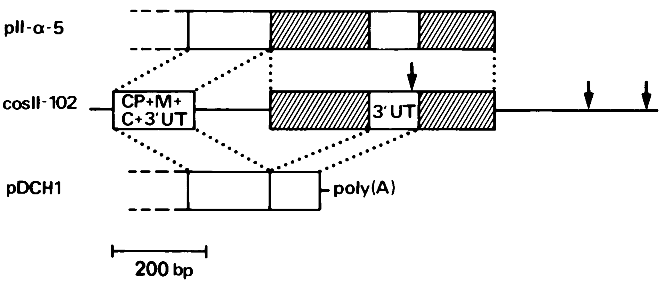

Fig. 3. Schematic representation of the differential splicing occuring in the 3′ part of DCα genes. The additional nucleotide stretches found in pII-α-5, as compared with pDCH1 (Auffray *et al.*, 1982), are indicated by the hatched boxes. Arrows mark polyadenylation signals in the DCα gene encoded by cosII-102 (A.-K. Jonsson, unpublished data). Abbreviations used: CP, connecting peptide; M, membrane-spanning segment; C, cytoplasmic stretch; 3′UT, 3′-untranslated region.

Discussion

Southern blot hybridizations suggest that the human genome contains two DCα-related and two DCβ-related genes (Böhme *et al.*, 1983; Auffray *et al.*, 1983; Wake *et al.*, 1982). One DCα-related and one DCβ-related gene appear virtually non-polymorphic as judged from the positions of several restriction enzyme sites. In contrast, alleles of the other DCα and DCβ gene show extensive restriction site polymorphism. It is not known whether both DCα-related and both DCβ-related genes are expressed. The three DCβ clones discussed here (Figure 6) display extensive homology to each other in the 3′-untranslated region. Since the invariant DCβ-related gene does not cross-hybridize to the 3′-untranslated region of these clones (G. Andersson, unpublished data), we conclude that the three DCβ clones correspond to allelic products. In contrast, we cannot formally prove that the DCα sequences (Figure 5) are derived from true alleles. However, pII-α-5 and pDCH1 display >90% homology in their 3′-untranslated regions, why we provisionally consider them as allelic products.

The polymorphism of the DR antigens is known to reside primarily in the β chains (Charron and McDevitt, 1979; Silver and Ferrone, 1979). Indeed, in the DRα chain, only a single variant position has been found although DRα sequences from six different sources have been described (Larhammar *et al.*, 1982a; Korman *et al.*, 1982a, 1982b; Lee *et al.*, 1982; Yang *et al.*, 1982; Das *et al.*, 1983). In contrast, the α chains of the DC antigens display extensive polymorphism (Figure 5). Although the two previously published DCα sequences are incomplete, corresponding to 85% and 53% of the polypeptide chains, respectively (Auffray *et al.*, 1982; Götz *et al.*, 1983), alignment of these sequences with the DCα sequence reported here conclusively shows that most of the polymorphism is expresed in the amino-terminal domain (residues 1−86, see Figure 5). Among the residues available for comparison, the amino-terminal domain of the DCα chain encoded by pII-α-5 displays 10 and 16 differences, respectively, to the correspon-

```
 -21                                                                          -1  +1
 Asp Leu Arg Val Ala Thr Val Thr Leu Met Leu Ala Ile Leu Ser Ser Ser Leu Ala Glu Gly Arg Asp Ser Pro Glu Asp Phe Val      8
A GAC CTT CGG GTA GCA ACT GTC ACC TTG ATG CTG GCG ATC CTG AGC TCC TCA CTG GCT GAG GGC AGA GAC TCT CCC GAG GAT TTC GTG     88

Tyr Gln Phe Lys Gly Leu Cys Tyr Phe Thr Asn Gly Thr Glu Arg Val Arg Gly Val Thr Arg His Ile Tyr Asn Arg Glu Glu Tyr Val   38
TAC CAG TTT AAG GGC CTG TGC TAC TTC ACC AAC GGG ACG GAG CGC GTG CGG GGT GTG ACC AGA CAC ATC TAT AAC CGA GAG GAG TAC GTG  178

Arg Phe Asp Ser Asp Val Gly Val Tyr Arg Ala Val Thr Pro Gln Gly Arg Pro Val Ala Glu Tyr Trp Asn Ser Gln Lys Glu Val Leu   68
CGC TTC GAC AGC GAC GTG GGG GTG TAC CGG GCA GTG ACG CCG CAG GGG CGG CCT GTT GCC GAG TAC TGG AAC AGC CAG AAG GAA GTC CTA  268

Glu Gly Ala Arg Ala Ser Val Asp Arg Val Cys Arg His Asn Tyr Glu Val Ala Tyr Arg Gly Ile Leu Gln Arg Arg Val Glu Pro Thr   98
GAG GGG GCC CGG GCG TCG GTG GAC AGG GTG TGC AGA CAC AAC TAC GAG GTG GCG TAC CGC GGG ATC CTG CAG AGG AGA GTG GAG CCC ACA  358

Val Thr Ile Ser Pro Ser Arg Thr Glu Ala Leu Asn His His Asn Leu Leu Ile Cys Ser Val Thr Asp Phe Tyr Pro Ser Gln Ile Lys  128
GTG ACC ATC TCC CCA TCC AGG ACA GAG GCC CTC AAC CAC CAC AAC CTG CTG ATC TGC TCG GTG ACA GAT TTC TAT CCA AGC CAG ATC AAA  448

Val Arg Trp Phe Arg Asn Asp Gln Glu Glu Thr Ala Gly Val Val Ser Thr Pro Leu Ile Arg Asn Gly Asp Trp Thr Phe Gln Ile Leu  158
GTC CGG TGG TTT CGG AAT GAT CAG GAG GAG ACA GCC GGC GTT GTG TCC ACC CCC CTC ATT AGG AAC GGT GAC TGG ACC TTC CAG ATC CTG  538

Val Met Leu Glu Met Thr Pro Gln Arg Gly Asp Val Tyr Thr Cys His Val Glu His Pro Ser Leu Gln Ser Pro Ile Thr Val Glu Trp  188
GTG ATG CTG GAA ATG ACT CCC CAG CGT GGA GAT GTC TAC ACC TGC CAC GTG GAG CAC CCC AGC CTC CAG AGC CCC ATC ACC GTG GAG TGG  628

Arg Ala Gln Ser Glu Ser Ala Gln Ser Lys Met Leu Ser Gly Val Gly Gly Phe Val Leu Gly Leu Ile Phe Leu Gly Leu Gly Leu Ile  218
CGG GCT CAG TCT GAA TCT GCC CAG AGC AAG ATG CTG AGT GGC GTT GGA GGC TTC GTG CTG GGG CTG ATC TTC CTT GGG CTT GGC CTT ATC  718

Ile Arg Gln Arg Ser Arg Lys Gly Leu Leu His ***                                                                          229
ATC CGT CAA AGG AGT CGG AAA GGG CTT CTG CAC TGA CTCCTGAGACTGTTTTAACTAAGACTGGTTATCACTCTTCTGTGATGCCTGCTTGTCCCTGCCCAGAATTCC  826

CAGCTGCCTGTGTCAGCTTGTCCCCCTGAGATCAAAGTCCTACAGTGGCTGTCACGCAACCACCAGGTCATCTCCTTTCATCCCCACCCCAAGGCGCTGGCTGTGACTCTGCTTCCTGCA  946

CTGACCCAGAGCCACTGCCTGTACATGGCCAGCTGCGTCTACTCAGG                                                                          993
```

Fig. 4. Nucleotide sequence and predicted amino acid sequence of the DCβ cDNA clone pII-β-2. Cysteines and the attachment site for N-linked carbohydrate are within boxes. Arrows mark exon boundaries inferred from a DCβ gene (Larhammar *et al.*, 1983b).

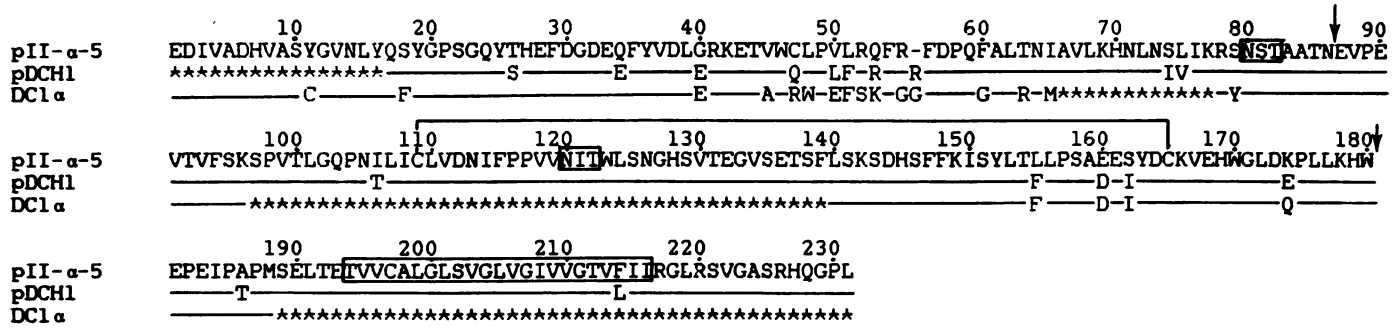

Fig. 5. Comparison of DCα amino acid sequences. The pII-α-5 sequence is derived from a cDNA clone of the DR3,w6 cell line Raji, and pDCHl is from a cDNA clone of a DR4,w6 cell line (Auffray *et al.*, 1982). DC1α is a protein sequence from a DR2,2 cell line (Götz *et al.*, 1983). Stars denote amino acid residues not available for comparison. Arrows mark exon boundaries. Sites for addition of N-linked carbohydrate and the membrane-spanning segment are within boxes.

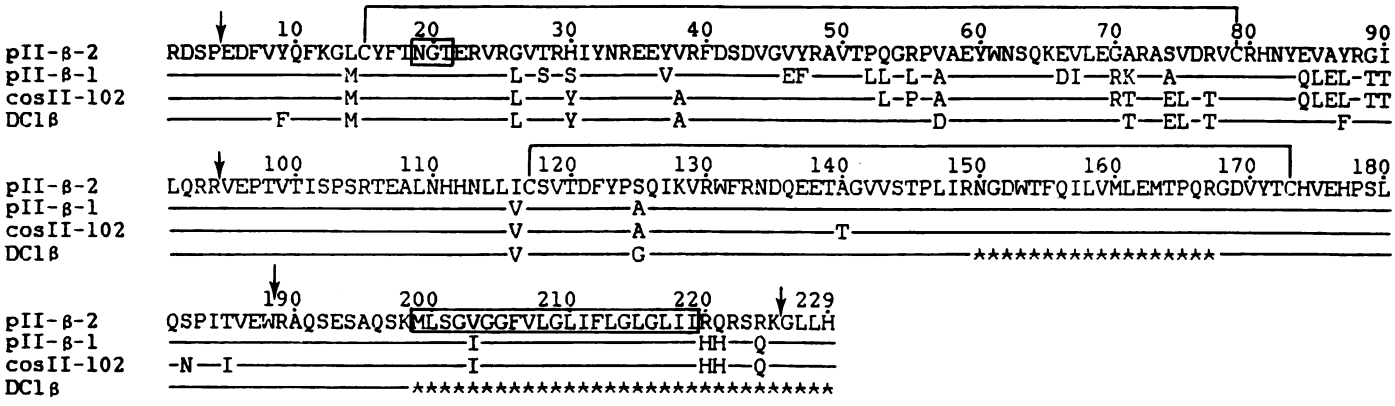

Fig. 6. Comparison of DCβ amino acid sequences. The pII-β-2 and pII-β-1 sequences are derived from cDNA clones of the DR3,w6 cell line Raji (Larhammar *et al.*, 1982b) and cosII-102 from a gene of a DR4,4 individual (Larhammar *et al.*, 1983b). DC1β is a protein sequence of a DR2,2 cell line (Götz *et al.*, 1983). Stars denote amino acid residues not available for comparison. Arrows mark exon boundaries. The site for addition of N-linked carbohydrate and the membrane-spanning segment are within boxes.

ding parts of pDCH1 and DC1α. The second domain (residues 87–180), on the other hand, shows only five and four differences, respectively. A similarly uneven distribution of the allelic polymorphism has also been found in the α chain of the murine structural homologue of DCα, i.e., I-Aα (Benoist *et al.*, 1983). The amino acid replacements are scattered over the entire amino-terminal domain in both DCα and I-Aα. However, a cluster of differences can be discerned in both polypeptides around amino acid 50. At only four of the positions available for comparison do all three DCα chains display unique amino acids (Figure 5; positions 47, 50, 53 and 174). Three of these positions occur in the cluster of variable residues. Although many of the replacements involve amino acids with similar chemical properties, some non-conservative exchanges also occur, e.g., at the positions where all three DCα chains are unique. Interestingly, at all four of these positions, the DRα sequence displays one of the variant amino acids.

The second DCα domain, which is immunoglobulin-like and comparatively constant, shows higher homology to the invariant DRα chain (~65%) than does the amino-terminal domain (~45% homology). The trans-membrane regions of the two types of α chains are more homologous (almost 80%) than the extracellular portions. Likewise, the β chains of DC and DR show less homology in the amino-terminal domains (~60%; residues 1–94), than in the second domains (~70%; residues 95–188). Again, the trans-membrane regions are well conserved, displaying ~80% homology, most likely due to a conservative pressure acting to maintain hydrophobicity. Also, these regions are well conserved within each allelic series. Possibly, the second domains and trans-membrane domains are crucial in the interaction between α and β chains of the different dimers, and have therefore been allowed to diverge less. Conversely, the differences between the various types of α and β chains, respectively, may have to be large enough to prevent formation of hybrid antigens, e.g., DCαDRβ.

Within the DCβ allelic series, the polymorphism in the amino-terminal domain is even more pronounced than in the DCα chains (see Figure 6). As in the DCα chains, amino acid replacements are found along the entire DCβ amino-terminal domain. However, three minor clusters of relatively more variability can be found (positions 52–57, 70–77, and 84–90). At no position do all four DCβ chains have unique amino acids. However, three different amino acid residues occur at seven positions, six of which are located in the amino-terminal domain. As in the DCα chains, several amino acid replacements involve residues with markedly different properties. At six of the seven positions where three different amino acids occur, at least one of the amino acid residues is also found in DRβ chains at the corresponding position (Gustafsson *et al.*, in preparation). This, together with the evidence from the DCα comparison, may indicate that although the conservative selection pressure appears to be low in the amino-terminal domains (see below), at a certain polymorphic position only a limited repertoire of amino acids is

compatible with the overall structure of the polypeptide.

The mechanism for the generation of the genetic polymorphism of the MHC antigens is far from understood. The preferential location of amino acid replacements in the amino-terminal domains of the class II antigen chains may suggest that substitutions are introduced by a process different from that generating substitutions in the second domains. However, Gustafsson *et al.* (in preparation) have demonstrated that the numbers of silent substitutions are virtually identical in the amino-terminal and the second domains of both DR and DCβ chains. This observation indicates that the exons of the two domains accumulate mutations at a common basic mutation rate. The numbers of replacement substitutions in the first domains were found to be in accordance with this mutation rate, suggesting that the conservative selection pressure on amino acid replacements is low in these parts of the polypeptide chains. In contrast, the numbers of replacement substitutions in the second domains are significantly lower than expected from the numbers of silent substitutions, implying that severe constraints act on replacement substitutions in this region. Unfortunately, a statistical evaluation of the DCα polymorphism is not possible since only one DCα nucleotide sequence has been published previously. Moreover, that sequence is incomplete. In fact, comparison of the pII-α-5 and pDCH1 sequences indicates a higher basic mutation rate in the first DCα domain than in the second domain. Additional sequence information is required to allow conclusions to be drawn.

The non-random distribution of the amino acid replacements in the first and second domains of DCα and β chains may suggest that events related to gene conversion (Efstratiadis *et al.*, 1980; Weiss *et al.*, 1983) may contribute to the polymorphism. In keeping with this notion is the observation that the sequences of cosII-102 and DC1β are identical at positions 71–77, yet differ extensively at positions 84–90 (see Figure 6). Here, instead, cosII-102 is identical to pII-β-1 (also at the nucleotide level). Thus, in the two regions mentioned, cosII-102 displays sequences suggesting that this DCβ gene is a hybrid between DC1β and pII-β-1. Such cross-wise similarities have also been observed between different class I histocompatibility antigens (Weiss *et al.*, 1983). Whether the polymorphism has arisen due to point mutations or gene conversion events (or both), fixation seems to be due to selection at the phenotypic level. As in antibody chains (Baltimore, 1981), certain regions of the α and β chains seem to be more amenable to amino acid replacements than other regions. Whether this propensity affects both structural and functional aspects of the class II antigens remains to be seen.

Since both DC subunits are polymorphic, the DC antigen may display more extensive polymorphism than DR, whose α subunit is virtually invariant. However, the polymorphic DCα and DCβ genes are located <15 kb apart in the MHC (Bo Servenius, unpublished data), so few recombination events would be expected to occur between the two genes. The DCα gene may therefore co-evolve with a linked DCβ gene to an extent which disallows heterodimer formation between the DCα chain and DCβ chains of non-linked alleles (trans-gene complementation). Thus, the number of different functional DC antigens may not necessarily be the product of the numbers of α and β alleles. However, the conservation of the second domains and membrane regions of the α and β chains may ascertain heterodimer formation between products of non-linked alleles.

Materials and methods

Construction and screening of cDNA libraries

Isolation of mRNA from the lymphoblastoid cell line Raji (DR3,w6) and cDNA synthesis after size fractionation of mRNA have been described (Wiman *et al.*, 1982). The cDNA library enriched for DRα clones was screened by colony hybridization (Maniatis *et al.*, 1982) with the nick-translated (Rigby *et al.*, 1977) 583-bp *Rsa*I-*Stu*I fragment containing most of the coding part of the DCα cDNA clone pDCH1 (Auffray *et al.*, 1982). The cDNA library enriched for β-chain clones was screened (Gergen *et al.*, 1979) with the 627-bp *Ava*I fragment covering almost the entire coding part of the DCβ cDNA clone pII-β-1 (previously denoted pDR-β-1; Larhammar *et al.*, 1982b).

Nucleotide sequence determination

Nucleotide sequences were determined by the chemical degradation procedure (Maxam and Gilbert, 1980). Two sequencing errors have been corrected in the DCβ sequence of pII-β-1 (previously denoted pDR-β-1; Larhammar *et al.*, 1982b), both involving *Bst*NI sites, whose second Cs are methylated in the *Escherichia coli* strain used and therefore behave aberrantly in the chemical degradation reactions. Position 499 should be a C and position 853 should be a C. Neither of these corrections influences the deduced amino acid sequence of pII-β-1.

Acknowledgements

We are grateful to Ms. Anita Moren for skilful technical assistance and Ms. Elisabetta Rossi for help with computer analyses. We thank Drs. C. Auffray and J.L. Strominger for the DCα probe. This work was supported by grants from the Swedish Cancer Society, Marcus Borgström's Fund, and King Gustav V's 80-years' Fund.

References

Auffray,C., Korman,A.J., Roux-Dosseto,M., Bono,R. and Strominger,J.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6337-6341.

Auffray,C., Ben-Nun,A., Roux-Dosseto,M., Germain,R.N., Seidman,J.G. and Strominger,J.L. (1983) *EMBO J.*, **2**, 121-124.

Baltimore,D. (1981) *Cell*, **26**, 295-296.

Benacerraf,B. (1981) *Science (Wash.)*, **212**, 1229-1238.

Benoist,C.O., Mathis,D.J., Kanter,M.R., Williams,V.E.,II and McDevitt, H.O. (1983) *Cell*, **34**, 169-177.

Böhme,J., Owerbach,D., Denaro,M., Lernmark,A., Peterson,P.A. and Rask.L. (1983) *Nature*, **301**, 82-84.

Charron,D.J. and McDevitt,H.O. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 6567-6571.

Choi,E., McIntyre,K., Germain,R.N. and Seidman,J.G. (1983) *Science (Wash.)*, **221**, 283-286.

Das,H.K., Lawrence,S.K. and Weissman,S.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3543-3547.

Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Weissman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C. and Proudfoot, N.J. (1980) *Cell*, **21**, 653-668.

Evans,R.L., Faldetta,T.J., Humphreys,R.E., Pratt,D.M., Yunis,E.J. and Schlossman,S.F. (1978) *J. Exp. Med.*, **148**, 1440-1445.

Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.

Götz,H., Kratzin,H., Thinnes,F.P., Yang,C.-y., Kruse,T., Pauly,E., Kölbel, S., Egert,G., Wernet,P. and Hilschmann,N. (1983) *Hoppe-Seyler's Z. Physiol. Chem.*, **364**, 749-755.

Hyldig-Nielsen,J.J., Schenning,L., Hammerling,U., Widmark,E., Heldin,E., Lind,P., Servenius,B., Lund,T., Flavell,R.A., Lee,J.S., Trowsdale,J., Schreier,P., Sablitzky,F., Larhammar,D., Peterson,P.A. and Rask.L. (1983) *Nucleic Acids Res.*, **11**, 5055-5071.

Hyldig-Nielsen,J.J., Larhammar,D., Servenius,B., Andersson,G., Rask,L. and Peterson,P.A. (1984) *Nature*, in press.

Klareskog,L., Sandberg-Trägårdh,L., Rask,L., Lindblom,J.B., Curman,B. and Peterson,P.A. (1977) *Nature*, **265**, 248-251.

Korman,A.J., Auffray,C., Schamboeck,A. and Strominger,J.L. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 6013-6017.

Korman,A.J., Knudsen,P.J., Kaufman,J.F. and Strominger,J.L. (1982b) *Proc. Natl. Acad. Sci. USA*, **79**, 1844-1848.

Kratzin,H., Yang,C.-y., Götz,H., Pauly,E., Kölbel,S., Egert,G., Thinnes, F.P., Wernet,P., Altevogt,P. and Hilschmann,N. (1981) *Hoppe-Seyler's Z. Physiol. Chem.*, **362**, 1665-1669.

Larhammar,D., Gustafsson,K., Claesson,L., Bill,P., Wiman,K., Schenning, L., Sundelin,J., Widmark,E., Peterson,P.A. and Rask,L. (1982a) *Cell*, **30**, 153-161.

Larhammar,D., Schenning,L., Gustafsson,K., Wiman,K., Claesson,L., Rask,L. and Peterson,P.A. (1982b) *Proc. Natl. Acad. Sci. USA,* **79**, 3687-3691.
Larhammar,D., Hammerling,U., Denaro,M., Lund,T., Flavell,R.A., Rask. L. and Peterson,P.A. (1983a) *Cell,* **34**, 179-188.
Larhammar,D., Hyldig-Nielsen,J.J., Servenius,B., Andersson,G., Rask.L. and Peterson,P.A. (1983b) *Proc. Natl. Acad. Sci. USA,* **80**, 7313-7317.
Lee,J.S., Trowsdale,J., Travers,P.J., Carey,J., Grosveld,F., Jenkins,J. and Bodmer,W.F. (1982) *Nature,* **299**, 750-752.
Long,E.O., Wake,C.T., Gorski,J. and Mach,B. (1983) *EMBO J.,* **2**, 389-394.
Malissen,M., Hunkapiller,T. and Hood,L. (1983) *Science (Wash.),* **221**, 750-754.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning,* published by Cold Spring Harbor Laboratory Press, NY, pp. 387-389.
Mathis,D.J., Benoist,C.O., Wiliams,V.E.,II, Kanter,M.R. and McDevitt, H.O. (1983) *Cell,* **32**, 745-754.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
Rigby,P.W.J., Dieckmann,M., Rhoades,C. and Berg,P. (1977) *J. Mol. Biol.,* **113**, 237-251.
Saito,H., Maki,R.A., Clayton,L. and Tonegawa,S. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 5520-5524.
Shaw,S., Johnson,A.H. and Shearer,G.M. (1980) *J. Exp. Med.,* **152**, 565-580.
Silver,J. and Ferrone,S. (1979) *Nature,* **279**, 436-437.
Tanigaki,N. and Tosi,R. (1982) *Immunol. Rev.,* **66**, 5-37.
Thorsby,E. (1977) *Transplant. Proc.,* **9**, 393-400.
Tosi,R., Tanigaki,N., Centis,D., Ferrara,G.B. and Pressman,D. (1978) *J. Exp. Med.,* **148**, 1592-1611.
von Heijne,G. (1983) *Eur. J. Biochem.,* **133**, 17-21.
Wake,C.T., Long,E.O. and Mach,B. (1982) *Nature,* **300**, 372-374.
Watson,A.J., DeMars,R., Trowbridge,I.S. and Bach,F.H. (1983) *Nature,* **304**, 358-361.
Weiss,E., Golden,L., Zakut,R., Mellor,A., Fahrner,K., Kvist,S. and Flavell, R.A. (1983) *EMBO J.,* **2**, 453-462.
Wiman,K., Larhammar,D., Claesson,L., Gustafsson,K., Schenning,L., Bill, P., Böhme,J., Denaro,M., Dobberstein,B., Hammerling,U., Kvist,S., Servenius,B., Sundelin,J., Peterson,P.A. and Rask.L. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 1703-1707.
Yang,C.-y., Kratzin,H.J., Götz,H., Thinnes,F.P., Kruse,T., Egert,G., Pauly,E., Kölbel,S., Wernet,P. and Hilschmann,N. (1982) *Hopper-Seyler's Z. Physiol. Chem.,* **363**, 671-676.

Received on 14 November 1983