Volume 11 Number 24 1983 Nucleic Acids Research

Organization of the transcriptional unit of a human class II histocompatibility antigen: HLA-DR heavy chain

Anton Schamboeck, Alan J.Korman, Alexander Kamb and Jack L.Strominger

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Received 10 August 1983; Revised and Accepted 7 November 1983

ABSTRACT

A total of 5724 base pairs of a recombinant phage DNA containing a human HLA-DR heavy chain gene including flanking regions has been analyzed. The regions corresponding to all the exons have been identified. The sites of initiation of transcription and polyadenylation have been determined. A large intron of 2399 base pairs separates the first exon containing the 5' untranslated region and the signal peptide from the second exon containing the N-terminal peptide domain.

INTRODUCTION

The major histocompatiblity complex (MHC) in man is located on chromosome 6 and is composed of 3 classes of antigens.

Class I is a highly polymorphic gene family defined by three genetic loci, HLA-A, HLA-B and HLA-C (1). On the nucleic acid level transcripts of class I gene products will crosshybridize to each other as well as to species homologous to the murine TL and Qa antigens (2). Class I antigens are membrane glycoproteins noncovalently associated with β2-microglobulin (the gene for which is located on chromosome 15). They are commonly expressed in all tissue cells, and it has been shown that they are recognized by cytotoxic T cells along with viral antigens or as alloreactive targets (3,4).

Class II antigens (Ir or immune response genes) are encoded by a genetic region called HLA-D. They represent a gene family of more limited polymorphism expressed on the cell surface of lymphoblastoid cells and in tissue involved in immune processes (5,6). They are two chain membrane glycoproteins composed of a 34000 dalton heavy chain (α) in noncovalent association with a polymorphic 29000 dalton light chain (β). Both chains pierce the membrane (1). It has been shown that HLA-D region products together with antigen are presented on the cell surface of macrophages and are recognized by syngeneic T cells resulting in T cell proliferation (7,8,9,10). So far experimental evidence has been presented for the existence of 3 class II related antigens

© IRL Press Limited, Oxford, England.

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

called DR, DC and SB (11,12,13,14,15). Class III antigens are peptides of the complement group (16).

We have focussed on the complete structural analysis of the HLA-DR heavy chain gene in an attempt to broaden our understanding about regulation and action of this gene product and for comparison with other α chain genes in man and in other species.

METHODS

DNA cloning and sequence analysis: A human genomic DNA library (fetal liver, HLA type unknown) in phage lambda Charon 4A (17) was screened with the probe pDRH-2 (18) as described previously (19). DNA fragments of the recombinant phage were subcloned into the plasmid vector pUC9. Restriction sites (Fig. 1) were labeled at their 5' ends using polynucleotide kinase (PL Biochemicals) and ^{32}P-γATP (10 mCi/nMol, a gift of Rick Myers), or at their 3' ends using the Klenow fragment of Polymerase I and ^{32}P-αdNTP's and subjected to the base specific degradation reactions according to Maxam and Gilbert (20). The cleavage products were analyzed on 80 x 30 x 0.03 cm 8% polyacrylamide urea gels.

Extraction of RNA and S_1 nuclease mapping: Total RNA was extracted by the method of Auffray and Rougeon (21) from the lymphoblastoid B cell line JY (DR 4,w6). 20 µg of total RNA together with 5000 cpm of an end-labeled DNA fragment in a total volume of 20 µl containing 0.5 M NaCl, 40 mM PIPES, 1 mM EDTA and 8 % formamide (Mallinckrodt) was denatured for 5 minutes at 75°C and annealed for 12 hours at 42°C. The sample was diluted 10 fold at 37°C into buffer containing 0.2 M NaCl, 4 mM $ZnCl_2$, 30 mM NaAc pH 4.5 and 40 U of S_1 nuclease (PL Biochemicals) and incubated for 30 minutes at 37°C. After phenol/chloroform extraction and isopropanol precipitation, samples were applied to 5% polyacrylamide urea gels as described (22).

RESULTS

Analysis of genomic clone lambda DRH-6A

A human genomic library in lambda Charon 4A (17) was screened with the cDNA probe pDRH-2 (18), representing the carboxy-terminal 31 amino acids and the entire 3' untranslated region of the HLA-DR heavy chain, as described previously (19). A phage, lambda DRH-6A, was isolated containing a 3.2 kb EcoRI fragment, which hybridized to pDRH-2 and has been characterized in detail (19). It represented all the coding portions of the DR heavy chain polypeptide except for the signal sequence including amino acids one and two

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

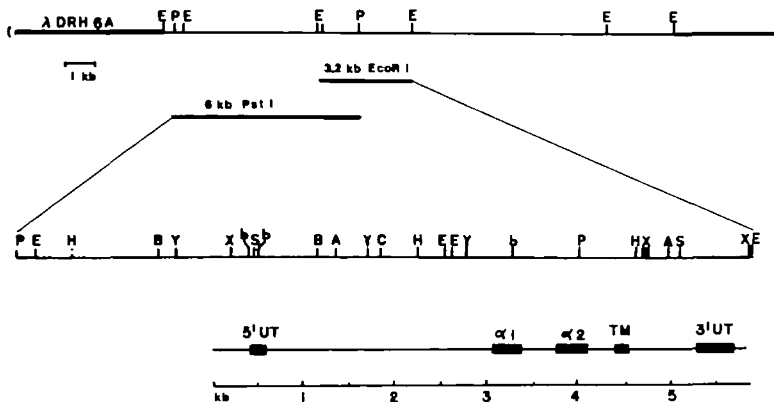


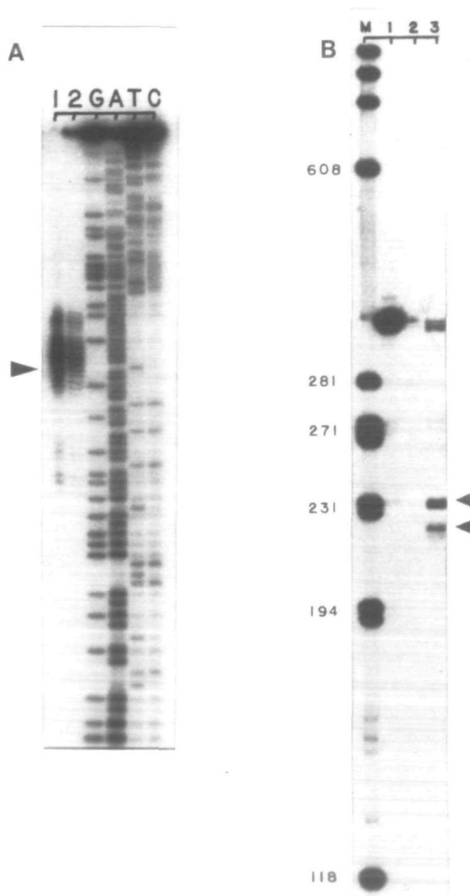
Fig. 1: Restriction map of the genomic clone lambda DRH-6A and the overlapping subclones 3.2 kb EcoRI and 6 kb PstI. The symbols for restriction enzymes are: A AccI, B BglII, b BalI, C ClaI, E EcoRI, H HindIII, P PstI, S SstI, X XbaI, Y XmaI. Intron-exon organization and the region sujected to nucleotide sequence analysis is shown below the map. Solid blocks represent exons.

of the mature DR heavy chain and the 5' untranslated region. The missing part was identified as follows: a 225 bp Sau3A primer fragment (position 3010 to 3235, see Fig. 3 below) was purified by polyacrylamide gel electrophoresis, denatured and hybridized to total JY RNA. In a primer extension reaction (23) a product which was 150-160 bp larger was found (results not shown). The same extension product was used as a probe to identify a 6 kb PstI fragment which was subcloned into the pUC9 plasmid vector and analyzed by restriction mapping (Fig. 1) and sequence analysis (see Fig. 3 below). It appeared from these studies and from the S_1 nuclease mapping experiment described below that the 5' untranslated region and the signal peptide are encoded in one exon of 148 nucleotides separated by an intron of 2399 nucleotides from the next downstream exon. The complete transcriptional unit of the HLA-DR heavy chain is therefore split into 5 exons (Fig. 1 and Fig. 3): 5' untranslated region, signal sequence and amino acids one and two of the α1 domain in exon 1; α1 domain in exon 2; α2 domain in exon 3; connecting peptide, transmembrane, intracytoplasmic region and 11 bp of the 3' untranslated region in exon 4 and the remainder of the 3' untranslated region in exon 5. The overall intron-exon organization has been confirmed by the structure of the corresponding cDNA clones (23,24,25,26).

The initiation site of transcription

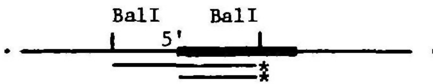
The 5' end of HLA-DR mRNA was mapped approximately to position 449 by S_1

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024



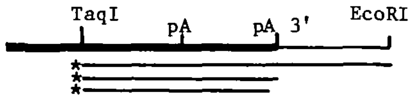
Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

Fig.2:
A: Determination of the 5' initiation site of transcription of the HLA-DR α chain gene by S_1 nuclease mapping. A 135 bp BalI fragment (position 380 - 515 Fig 3) was 5' end labeled, strand separated and annealed to 30 μg of total JY RNA (lane 1) or to 3 μg of polyA selected JY mRNA (lane2) as described in Methods. The same fragment was subjected to the base specific degradation reactions according to Maxam & Gilbert and run side by side on a 5% polyacrylamide urea gel (lanes GATC). The labeled probe and protected fragments are shown in the small schemes below.



B: Determination of the 3' end by S_1 nuclease mapping. The probe was a 3' end labeled 360 bp TaqI - EcoRI fragment (position 5365 - 5724 Fig. 3). M, marker fragments in basepairs; lane 1: input DNA probe, lane 2: annealing of probe in the absence of RNA, lane 3: annealing of probe in the presence of RNA shows

two protected fragments (indicated by arrows) differing by 10 - 15 bp corresponding to the second polyA addition site at position 5570 in Fig. 3. Labeled probe and protected fragments are shown below.



nuclease mapping (Fig. 2A). A 5' end labeled BalI fragment of 135 bp (corresponding to position 380 to 515 in Fig. 3) was hybridized to total JY RNA and treated with S_1 nuclease. The protected fragment could be mapped to the corresponding DNA sequence as shown in Fig. 2A. The Cap position 449 in Fig. 3 may indicate the most upstream possible transcript; the major start positions are to be expected in the region 5 to 10 bp downstream (arrow in Fig. 2A). The first AUG codon is located at position 513 resulting in a 60 to 65 bp 5' untranslated sequence followed by 25 amino acids of the signal peptide. This sequence was identical to the partial sequences of the 5' untranslated region of an HLA-DR cDNA clone (27).

The termination site of transcription

The cDNA clone pDRH-2 (18) apparently contained a complete 3' untranslated region running into a stretch of polyA. According to this transcript the termination signal would have to be placed at position 5570 AATAAA in Fig. 3. However two potential polyA addition sites are located in the 3' untranslated region corresponding to position 5472 and 5570 respectively. An S_1 nuclease mapping experiment was performed in order to show if one of the two termination signals is used preferentially. A 360 bp TaqI-EcoRI fragment (position 5365-5724) was 3' end labeled, denatured and hybridized to total JY RNA as described in the methods section. After S_1 nuclease treatment the only detectable products of 220 and 235 bp mapped close to position 5596 (Fig. 2B) indicating that the termination signal 5570 was used exclusively in JY cells. No shorter transcripts corresponding to the first termination signal at position 5472 could be detected. However the detectable signal corresponding to the second polyadenylation site was split into two bands differing by 10-15 nucleotides, indicating two transcripts, the first one mapping to approximately position 5585 and the second to position 5596. In this region a BglII restriction site polymorphism has been reported (23). The finding of two different transcripts in the 3' end may indicate the existence of two slightly different alleles of the HLA-DRα chain gene in the cell line JY (DR4,w6). The same two bands were found with RNA from the cell line LB (DRw6,w6 but not homo-

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

zygous throughout the MHC).

Structural elements in the DNA sequence

The complete sequence (Fig. 3) was determined from the two overlapping subclones shown in Fig. 1 by the Maxam and Gilbert sequencing protocol (20). With the initiation site at position 449 and termination at position 5596 a precursor transcript of approximately 5150 bp + 150 polyA would be predicted. The DNA sequence has been searched for elements of symmetry, repeats and consensus sequences by Homology Matrix computer programs (28). Some of the more obvious ones are listed below: position 318-325 and 331-338 imperfect tandem repeat may correspond to the minus 100 initiation signal described by McKnight *et al*. (29); position 381 TTGGCCAA corresponding to the minus 65 to 75 initiation region CCAAT (30); position 441 TATTA Hogness-box candidate; position 423-436 a 7 bp inverted repeat; position 451-463 a 7 bp inverted repeat located in the 5' Cap region which may in part be responsible for the broad signal observed in the S_1 nuclease mapping in Fig. 2A; position 849 the element TGGGGG repeated 4 times; position 993 homopolymer 8 A; position 1965 homopolymer 10 A; position 4746-4826 shows a 23 bp element repeated 3 times in intron 4; position 5421-5438 a 9 bp direct repeat; position 5501-5516 7 bp palindrom; position 5472 and 5570 AATAAA termination and polyadenylation signals; two additional AATAAA signals were observed in the first large intron position 1581 and 2300.

Extensive regions of the human HLA-DRα chain gene and its murine homolog IEα (31,32) have been compared. The coding regions are 82% homologous and intron sequences in the range of 55% to 65% except for a 90 bp element located in the first large intron (corresponding to position 2050 to 2140 in Fig. 3) showing 75% homology. It will be of interest to search other class II antigen related sequences for homology to this element.

An interesting observation was made in the promotor region. A segment of 95 bp (position 37 to 130 in Fig. 4) of the human and murine promotor regions are nearly identical except for a 31 bp insertion in the human sequence as compared to the murine, changing the promotor sequence drastically (Fig. 4). The putative TATA-box and the 7 bp inverted repeat position 423-436 (Fig. 3) is contained within this 31 bp insertion. This segment of the human gene substitutes for the murine TATA-box region and results in a shift converting the murine TATA-box region into the minus 55 region in the human gene, possibly resulting in a more heterogeneous Cap site.

By comparison of the promotor regions of the murine light chain gene IEβ with the analogous sequences of DRα and IEα Saito *et al*. (33) have observed

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

```
   1  AGTACTGCCA AATTCGAGAC AATCTCCATG ACCTGACAAT TTACCTTCTA TTTGGGTAAT
  61  TTATTGTCCC TTACGCAAAC TCTCCAACTG TCATTGCACA GACATATGAT CTGTATTTAG
 121  CTCTCACTTT AGGTGTTTCC ATTGATTCTA TTCTCACTAA TGTGCTTCAG GTATATCCCT
 181  GTCTAGAAGT CAGATTGGGG TTAAAGAGTC TGTCCGTGAT TGACTAACAG TCTTAAATAC
 241  TTGATTTGTT GTTGTTGTTG TCCTGTTTGT TTAAGAACTT TACTTCTTTA TCCAATGAAC
 301  GGAGTATCTT GTGTCCTGGA CCCTTTGCAA GAACCCTTCC CCTAGCAACA GATGCGTCAT
 361  CTCAAAATAT TTTTCTGATT GGCCAAAGAG TAATTGATTT GCATTTTAAT GGTCAGACTC
                                  Cap
 421  TATTACACCC CACATTCTCT TTTCTTTTAT TCTTGTCTGT TCTGCCTCAC TCCCGAGCTC 5'UT
                                    MetAlaIl eSerGlyVal ProValLeuG
 481  TACTGACTCC CAAAAGAGCG CCCAAGAAGA AAATGGCCAT AAGTGGAGTC CCTGTGCTAG SIG
      lyPhePheIl eIleAlaVal LeuMetSerA laGlnGluSe rTrpAlaIle LysG
 541  GATTTTTCAT CATAGCTGTG CTGATGAGCG CTCAGGAATC ATGGGCTATC AAAGGTAGGT
 601  GCTGAGGGAA TGAAATCTGG GACGATAGAC TACGAAGCAT TGGAGAAAAG ACCTATGGAC
 661  ATTTGGAAGA TAATGTGTGG AGTGAAAGAA TAGTGTGACA GGTATTATGT GGTCTCGACA
 721  GAAAGTATAA CAAATTGTGG TTTGGTGGAG TTCTTCCCTC ACCACAAACT GAAGTAAGTC
 781  AAATTTGGTT TAGAGGGTCA AAACTGAGTT GTGTATTGAT GAATAGCACG GTCCTGCTAC
 841  AAGCCAAACT GGGGGTGGGG GTGGGGGTGG GGGAGGAAGA ATATTTTCTG GCAAGCATTA
 901  ACAAGTTATA TTTCTGGGCT TTAATTATTC TTTCTGGAAA ATTAGTAAAA TTAAAAACTA
 961  AAAACCACAC ATAGTTTTGT TAGAATTAAA TGAAAAAAAA AGTTATTAGC CCTGTTCTTA
1021  TCTGAATACA TGATACAGTA GTTATTTTTT GGAGTGTAAA TCCTGTCGGT ATATATTGAG
1081  CACATATATT GTGTTGAAGA TTACTAGAAG GAAAAGTCAT CAAAAAGCAA CAATTTACCC
1141  CAGGAAAAGG GGAGGGAAGG CATGCTGATA TGAGTTGCCT CATGGGACAG TGATAGCCAT
1201  TCCCTGCCTT CCCATCTCCA TGGTACAGCA GATCTTATAT CATGTTAACT TAGTAATATT
1261  TCCAAGAGAG TAGAAAAATA AGTAAGGAAA TGGGGAATCT GATATTATTG TCTCTCATCT
1321  CCAGAGCAAC ATTGGTGCTG TTGTAAAGAT GTACTGTAGA AAAGTATTCT TCACCCAGCG
1381  TGACCCCCAC AGAAGGTGTC AGGTAGACTT GGAATAAGCA AAGTAATAAC CCAGCTCCCA
1441  TACCCATAGT GGCAATTGTA GATTTCTATT GCCCCAAAAG AGCCATACAT AGGGATACTT
1501  ACCTAGAAAG ACAGAGGATC TTCCCTTGGT TTGTGAAGAG GCAGCTAGTA TATTTGTGTG
1561  TGTTTGCATA GATGGAAGTA AATAAATTCC TAGGTTTATC AATACACAGT CAAACATTTA
1621  AAATCTCTCA TCTTGGCTGG GCACGGTGGC TCACGCTAAT CCCAGCACTT TGGGAGGCCG
1681  AGGCGGGCGG ATCACGAGGT CAAGAGATCG AGACCGTCCT GGGCAACATG GTGAAACCCC
1741  GTCTCTACTA AAATACAAAA AATTAGCTGG GTATGGTGGC ACACGCCTGT AGTCCCAGCT
1801  ACTCGGGAGG CTGAGGCAGG AGGATTGCTT GGCCCGGGAG GCGGAGGTTG CAGTGAGCTG
1861  AGATGGTGCC ACTGCACTCC AGCCTGGCGA TAGAGCAAGA CTCCGTCTCA AACAACCAAA
1921  CCAAAACAAA ACAAAATATC TCACCTTATC TTTGAAGACT AAGGAAAAAA AAAATCTCCC
1981  ACTCATCGAT ACACTCCACA GAGGCAGCAT ACTCTCCCAG TGTAGCTTTC TCTTTTCATG
2041  TTCATTATTC CCTTGGTGTT GGTTATTCTC AATGTCAATC GTAACAGAAC ATCTTCCATA
2101  ATAACAGTCC CAATTTAAGG AGCATTAAGA TAAAAGGTGG AATTGCCAAG GTCAATCCAG
2161  ACGAGAACCT TCTCATAGAG GTAACCACCG TGTGGGTTTG GATGCTGGGA AGCAGGGGGA
2221  CTATGACGCT ACAAGGTCTC AGTCTTAATT TTTGGAGTAC TTCAGTCCCC AGGTATATTT
2281  TCCATAGATT TGGCCCTTAA ATAAAAAGAA GCTTCTGACT CTAAAATGTA AACAGTGCTT
2341  GTTACAGTCT TGTTGATATA TTAAGAAATT ACTCACCTTA TCTCATTTAA TCTTAAAAAC
2401  AAACCCCTGA CAGGATCAAA ACCACAGCAG GACTACATAA TAGGAAAACT ATACATAAAT
2461  AGGTAGAATA ATCTGCTCAG GATCACTAGG TAAGTTGCTG AATAAGAATT CAAGATGAAA
2521  AAGATCCCAG AGTTTAAAAC CCAACCTTTC AAACAGTGTT TCCTTCTTCT TAGAGTACAA
2581  TGTTCTGAGA AAGAGATCCT CTGGAATTCT GGCCTAAGTG TATTTAATGC CCGGGTAAAG
2641  AAAGTGAGAG AACATTTCTC TTTAGGGGCT GCTGCTGGAT TTCTAAAAAG AAAATAATTT
2701  CTCAGCTAGT AACATGGAGC CAAACAACAG CTTCACAAGA CTCTGGGTTC TTTAGCCCTC
2761  ATCTCCTTCA ATCCACCCTC TTATAACCAG TCCTTCTTGT TTTTCCCCTC CCAGCTTTGT
2821  TCAGCAGCAT GCCCTTCACC CAGACCTTGT CTTGTCACTC ATCCCTACTC GCCATCATTC
2881  TTTCATTCCT CTTGGCCCAA TCTCTCTCCA CCACTTCCTG CCTACACGTA TGTAGGTCAC
                                                                  lu
2941  CATTCCCCTC TCTTGATTCC CCCCGCCCAA CTCTCTTTCT CCATTTCTTG CCTTTCAGAA
      GluHisValI leIleGlnAl aGluPheTyr LeuAsnProA spGlnSerGl yGluPheMet
3001  GAACATGTGA TCATCCAGGC CGAGTTCTAT CTGAATCCTG ACCAATCAGG CGAGTTTATG α1
      PheAspPheA spGlyAspGl uIlePheHis ValAspMetA laLysLysGl uThrValTrp
3061  TTTGACTTTG ATGGTGATGA GATTTTCCAT GTGGATATGG CAAAGAAGGA GACGGTCTGG
```

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

```
      ArgLeuGluG luPheGlyAr gPheAlaSer PheGluAlaG lnGlyAlaLe uAlaAsnIle
3121  CGGCTTGAAG AATTTGGACG ATTTGCCAGC TTTGAGGCTC AAGGTGCATT GGCCAACATA   α1
      AlaValAspL ysAlaAsnLe uGluIleMet ThrLysArgS erAsnTyrTh rProIleThr
3181  GCTGTGGACA AAGCCAACCT GGAAATCATG ACAAAGCGCT CCAACTATAC TCCGATCACC
      AsnV
3241  AATGGTACCT CCCTCTCTGC TGCACTCCTG GACACGGGAA TCCATAGTTT GAAAGTAGTT
3301  GCTTCAGCTC TTTGTGTTAG ATTATTGTAA CTGATTTTCC CTCCAAGGGC CTAACCTTGC
3361  CATTAACAAG CCCCAAATTC TCATGCCAGA GGTCTGAGAA CTTTATGGGT TTGATCCTAT
3421  CTTGTTGTGC TCAAGTCTTG TCTCTGTCAT CCATGGTCTC CTACGAAGTC ATTGCCCTAA
3481  GTTCATGCTA GGGGAGCCAG AAGGGAAGTC CTTGGATATC TTATACCTCA ATATTGGCTC
3541  AATCTCTTGG GGAGGGGGTG CTGTCAGAGA CTGTTATCTG AGGATGTGAC ATAGACTTCT
3601  CAGGGCACAA TTTCAACTAC TTTTTCAGCT TTAGGGTTTT TAGATACGTT TGTACCACAA
3661  TTGAGCATGG GAGGGAGAGG GGTGAGCCTA AGCAGTGACG GCTGATTCTG TCACGTCTGT
                 alProP roGluValTh rValLeuThr AsnSerProV alGluLeuAr
3721  CATGTGTCCC CCAGTACCTC CAGAGGTAAC TGTGCTCACG AACAGCCCTG TGGAACTGAG
      gGluProAsn ValLeuIleC ysPheIleAs pLysPheThr ProProValV alAsnValTh
3781  AGAGCCCAAC GTCCTCATCT GTTTCATCGA CAAGTTCACC CCACCAGTGG TCAATGTCAC   α2
      rTrpLeuArg AsnGlyLysP roValThrTh rGlyValSer GluThrValP heLeuProAr
3841  GTGGCTTCGA AATGGAAAAC CTGTCACCAC AGGAGTGTCA GAGACAGTCT TCCTGCCCAG
      gGluAspHis LeuPheArgL ysPheHisTy rLeuProPhe LeuProSerT hrGluAspVa
3901  GGAAGACCAC CTTTTCCGCA AGTTCCACTA TCTCCCCTTC CTGCCCTCAA CTGAGGACGT
      lTyrAspCys ArgValGluH isTrpGlyLe uAspGluPro LeuLeuLysH isTrpG
3961  TTACGACTGC AGGGTGGAGC ACTGGGGCTT GGATGAGCCT CTTCTCAAGC ACTGGGGTAT
4021  GGACCAACAC TCAATCTCCT TTATTTCAAG GTTTCCTCCT ATGATGCTTG TGTGAAACTC
4081  GGTGTTCTAA CTGTTTCATA ATATCTGCTA CAATTAATAT AACTGTCTTC TCCTACTATC
4141  CAGCTTCCTC CTTTTTTTAA TCTGTAATTC TCTCAATACA TCATTCTGTC TTCCTCTTCT
4201  TTAATCTATG AATAACTTTT CTCTTTCTAA AGAACCCTAC ATTTGATTCT GAGTGTTACT
                                                    luPhe AspAlaProS
4261  TCTTCCCACA CTCATTACCA TGTACTCTGC CTTATCTCCC CCCAGAGTTT GATGCTCCAA   CP
      erProLeuPr oGluThrThr GluAsnValV alCysAlaLe uGlyLeuThr ValGlyLeuV
4321  GCCCTCTCCC AGAGACTACA GAGAACGTGG TGTGTGCCTT GGGCCTGACT GTGGGTCTGG   TM
      alGlyIleIl eIleGlyThr IlePheIleI leLysGlyLe uArgLysSer AsnAlaAlaG
4381  TGGGCATCAT TATTGGGACC ATCTTCATCA TCAAGGGATT GCGCAAAAGC AATGCAGCAG   CY
      luArgArgGl yProLeu***
4441  AACGCAGGGG GCCTCTGTAA GGCACATGGA GGTGAGTTAG GTGTGGTCAG AGGAAGACAT
4501  ATATGGAGAT ATCTGAGGGA GGAAAATCAG GGTGGGGAAA GGAAATGTAA TGCATTTAAG
4561  AGACAAGGTA GGAACAGATG TGGCTCTTGA TTTCTCTTTG CTAGAATGAA TCAGACATTG
4621  GTATCATCTG GTACCCCAAA GCTTCAGGGT CTGTCATCCC TTTCTATAGA CGGGCACCTT
4681  GATCACGGCT CCAGTCTTAG AAATCATCTC CAGTACCTAA AACCATTGTT TCACATTAGA
4741  ATACTGAGTC TAGGGATCTA GAAAATACAT TAGAATATGG AGTCTAGGGA TCTAGAAAAT
4801  ACTGAGTCTA GGGATCTAGA AAAATAAGCC TCAAGATTTG GGCACATCCT AGCTTGTATT
4861  TCCTGGGGCA GGTCATCAGT TCAGAAGCAT TTCCAGATCC TGGCTCCTTT CAGGTTAGGG
4921  TCAATTCGTT GCATGAAATG GGAATCTCTT AGAGGCCAAT GCCTGCTTTT GCTTCTTTAG
4981  TCTCAAATGT AGTATGAGAA ACTCTAAAAA AGGGTAAGAC ATGGTTGCTT ATTATGTTCA
5041  GTTGGAGAGT AGGAACTAAC TGTATACAGT TAGTTCATGT TGGAAAGGTT AGATGAACAT
5101  TGAAAGAATT TTGCAAAGTC AAAGGATTAA GAGAGAAGAG GAAGGAATCT GAAGCAAGGA
5161  GCTCAAAACG GATCTTAAAT TCCTTGGTAA CTATGTGTGT CTTGCTATAG GTGATGGTGT
5221  TTCTTAGAGA GAAGATCACT GAAGAAACTT CTGCTTTAAT GACTTTACAA AGCTGGCAAT
5281  ATTACAATCC TTGACCTCAG TGAAAGCAGT CATCTTCAGC GTTTTCCAGC CCTATAGCCA
5341  CCCCAAGTGT GGTTATGCCT CCTCGATTGC TCCGTACTCT AACATCTAGC TGGCTTCCCT   3'UT
5401  GTCTATTGCC TTTTCCTGTA TCTATTTCCT CTATTTCCTA TCATTTTATT ATCACCATGC
5461  AATGCCTCTG GAATAAAACA TACAGGAGTC TGTCTCTGCT ATGGCCCATG GGGCATCTCT
5521  TGTGTACTTA TTGTTTAAGG TTTCCTCAAA CTGTGATTTT TCTGAACACA ATAAACTATT
5581  TGATGATCTT GGGTGGAATT TTTGGTGTTT AAGCCAGTTC TTTGGGTGGC GGTGGGGGGT
5641  GGGGAGTCGG TCCTGGGGAA TATATGTGAT CCTTTCCCGG TAAAATATCT GAATGTTGAA
5701  TTTATCTTAT AAATTCTAGA ATTC
```


Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

two conserved elements not otherwise discernable between the nearly identical human and murine 5' flanking regions (Fig. 4 and ref. 31). These two elements (indicated by stippled boxes in Fig. 4) are separated by 20 nucleotides and are located 50-100 bp upstream of the Cap site. In addition the two corresponding elements are strikingly conserved in the murine IAβ gene (34), in a human HLA-DCβ gene (J. Boss personal communication) and in a human HLA-DCα gene (J. Lillie personal communication). Recent work indicates that γ-interferon can induce class II antigen expression in peripheral blood monocytes (35). The role of those conserved elements mentioned above need to be defined and delineated with respect to developmental versus inducible gene expression.

DISCUSSION

It has been shown for class I antigens (36,37) that intron-exon organization follows the general principle of exon blocks containing the coding information for the functional peptide domains. A similar arrangement has been observed in the class II heavy chain gene described in this paper. An unusually large intron (2399 bp) searates the first from the second exon, as compared to class I antigens (36). This structural arrangement with a large intron separating the 5' exon is comparable to the homologous murine IEα chain gene (31,32,38) and also to the 5' end structure of the β2-microglobulin gene (39). A further difference in intron-exon organization between class I and II heavy chains is observed in the 3' terminal portion. The class II gene contains transmembrane and intracytoplasmic region in one exon block of 165 bp separated from the 3' untranslated portion of 390 bp by an intron of 790 bp. The corresponding region in class I heavy chains is split into several smaller blocks (36,37). The difference in structural arrangement of transmembrane and intracytoplasmic portion between class I and II antigens may imply a difference in function. By 'Southern blotting' using DNA from various cell lines or individuals and probing with an HLA-DRα chain probe, a constant

Fig. 3: Nucleotide sequence and peptide domains of an HLA-DRα (heavy) chain gene. The sequence was deduced from the two overlapping subclones 3.2 kb EcoRI and 6 kb PstI of lambda DRH-6A indicated in Fig. 1 as described in Methods. Exon 1: 5' untranslated region, signal peptide (25 amino acids) and amino acids 1 and 2 of the α1 domain, Exon 2: α1 peptide domain, Exon 3: α2 peptide domain, Exon 4: connecting peptide (13 amino acids), transmembrane region (23 amino acids), intracytoplasmic region (15 amino acids) and 11 bp of the 3' untranslated region, Exon 5: 3' untranslated region. Glycosylation sites correspond to Asn position 3223 and Asn 3833. A disulfide loop forms between Cys at position 3800 and Cys 3968 respectively.

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

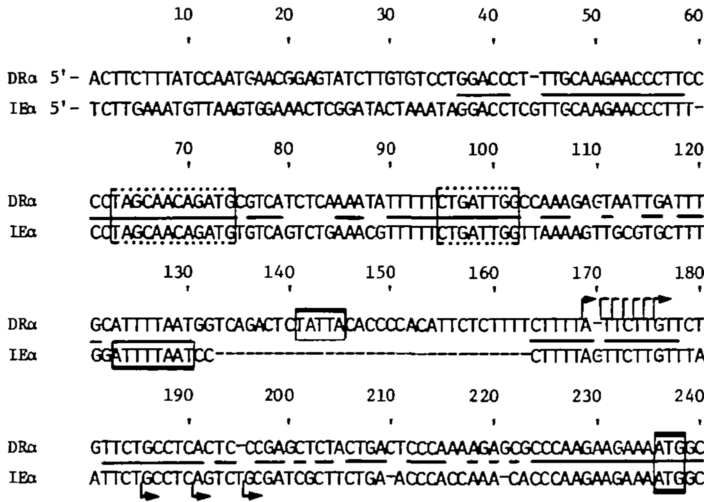


Fig. 4: Aligned promotor regions of the human HLA-DRα and the murine IEα chain genes. Major mRNA starts are indicated by arrows. AUG codons and putative TATA-boxes are boxed. Conserved regions are underlined. The stippled boxes indicate conserved elements in the initiation region of class II antigens described by Saito et al. (33). The murine sequence was from ref. (32).

restriction pattern has been observed with the exception of one BglII site polymorphism (23), corresponding to position 5615 in Fig. 3. The HLA-DRα chain is not responsible for the HLA-DR phenotypes defined by serology.

The human genome also encodes the HLA-DR related DC and SB α chain genes. The coding DNA sequence of the DC-1 heavy chain shows 57% overall homology to HLA-DR, and SB has a similar degree of homology (14, and C. Auffray in press). In contrast to HLA-DR the DC heavy chain shows population polymorphism (15).

It is unlikely that the DR related α chain genes interfere in S_1 nuclease mapping using end labeled probes as was performed in this study, because the position carrying the label shows little sequence homology.

It has been shown that two size classes of β2-microglobulin mRNA differing by 250 bp are transcribed from one gene. The two mRNA species map to different polyA addition sites in the DNA sequence (40). This result is in contrast to our finding for the HLA-DRα chain gene where apparently only one of the two potential polyA addition sites is recognized in JY cells (Fig. 2B).

It has been pointed out earlier by several authors (1,19,41) that the peptide domain located closest to the membrane is most conserved (approximately 40%) among the class I and II histocompatibility antigens. The region shows

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

striking homology to β2-microglobulin and to the constant region of immunoglobulins, pointing towards a common ancestor for these classes of genes.

Another analysis of the human HLA-DRα chain gene has been published recently (42). The coding sequences are in complete agreement except for two positions 4359 T→C silent, and 4419 T→G changing $Leu_{217} \rightarrow Val_{217}$, see Fig. 3 and ref. (19,24). They may reflect allelic variation as they have been found in corresponding cDNA clones (19,24). Intron and flanking sequences differ at 48 positions. Some of the descrepancies in our sequence may be the result of sequencing errors (estimated error rate less than 1%) or allelic variation. Point mutations may also be considered as the DNA is propagated in cell lines and bacterial vectors under no selective pressure. The following gaps not indicated by dashes exist in their sequence (42) as compared to ours corresponding to positions in Fig. 2 of their presentation: 9 bp gap position 1910 and 151 bp gap position 2538. The statement made in their work that the entire HLA-DRα chain gene is contained in two contiguous EcoRI fragments of 4.4 kb and 3.1 kb respectively is not in agreement with our mapping and sequencing data and inconsistent with their own presented DNA sequence showing two EcoRI sites at positions 2735 and 2832 in their Fig. 2. Our analysis shows that a 98 bp EcoRI fragment maps between the fragments mentioned above.

In addition a gene expression study carried out by C. Rabourdin-Combe and B. Mach (43) using the lambda phage DNA (λDRH-6A) described in this paper confirmed that the gene is functional as would be expected from our analysis.

ACKNOWLEDGMENTS

We are grateful to Dario Grossberger, David Arnot and members of the laboratory of Tom Maniatis for expert technical advice and gifts of restriction enzymes, to Magali Roux-Dosseto and Dietmar Kappes for gifts of mRNA and to Alice Furamoto-Dawson for typing the manuscript. Steve Dunn helped us in preparing large scale phage DNA. This work was supported by grants from the Swiss National Foundation (to A.S.), from the National Institute of Health (AI-30241) and from the Kroc Foundation.

REFERENCES

1. Strominger J.L. et al. (1981) in the Role of the Major Histocompatibility Complex in Immunobiology, (Dorf M.E. Ed.) Garland STPM Press, New York p. 115-172
2. Orr H. and DeMars R. (1982) Nature 302, 534-536
3. Zinkernagel R. and Doherty P. (1974) Nature 251, 547-548
4. Zinkernagel R. and Doherty P. (1979) Adv. Immunol. 27, 51-177
5. Nixon D., Pan-Yun-Ting J. and Frelinger J. (1982) Immunology

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

Today 3, 339-342
6. Pober J. and Gimbrone Jr. M. (1982) Proc. Natl. Acad. Sci. USA 79, 6641-6645
7. Benacerraf B. (1981) Science 212, 1229-1238
8. Lamb J. and Feldman M. (1982) Nature 33, 456-458
9. Lamb J., Eckels D., Lake P., Johnson A., Hartzman R. and Woody J. (1982) J. Immunol. 128, 233-238
10. Fischer A., Beverley P. and Feldman M. (1981) Nature 294, 166-168
11. Tosi R., Tanigaki M., Centis N., Ferrara G. and Pressman D. (1978) J. Exp. Med. 148, 1592-1611
12. Cook R., Vitteta E., Uhr J. and Capra J. (1979) Mol. Immunol. 16, 29-35
13. DeKretser T., Crumpton M., Bodmer J. and Bodmer W. (1982) Eur. J. Immunol. 12, 600-606
14. Auffray C., Korman A., Roux-Dosseto M., Bono R. and Strominger J. (1982) Proc. Natl. Acad. Sci. USA 79, 6337-6341
15. Auffray C., Ben-Nun A., Roux-Dosseto M., Germain R., Seidman J. and Strominger J. (1983) EMBO J. 2, 121-124
16. Shreffler A. (1982) Histocompatibility Antigens Structure and Function ed. Parham P. and Strominger J.,Chapman and Hall p. 189-219
17. Lawn R., Fritsch E., Parker R., Blake G. and Maniatis M. (1978) Cell 15, 1157-1174
18. Korman A., Knudsen P., Kaufman J. and Strominger J. (1982) Proc. Natl. Acad. Sci. USA 79, 1844-1848
19. Korman A., Auffray C., Schamboeck A. and Strominger J. (1982) Proc. Natl. Acad. Sci. USA 79, 6013-6017
20. Maxam A. and Gilbert W. (1980) Methods in Enzymol. 65, 499-560
21. Auffray C.and Rougeon F. (1980) Eur. J. Biochem. 107, 303-314
22. Maniatis R., Fritsch E. and Sambrook J. (1982) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory
23. Stetler D., Das H., Nunberg J., Saiki R., Sheng-Dong R., Mullis K., Weissman S. and Erlich H., (1982) Proc. Natl. Acad. Sci. USA 79, 5966-5970
24. Lee J., Trowsdale J. and Bodmer W. (1982) Proc. Natl. Acad. Sci. USA 79, 545-549
25. Larhammar D., Gustafsson K., Claesson L., Bill P., Wiman K., Schenning L., Sundelin J., Widmark E., Peterson P. and Rask L. (1982) Cell 30, 153-161
26. Long E., Wake C., Strubin M., Gross N., Accolla R., Carrel S. and Mach B. (1982) Proc. Natl. Acad. USA 79, 7465-7469
27. Wake C., Long E., Strubin M., Gross N., Acolla R., Carrel S. and Mach B. (1982) Proc. Natl. Acad. Sci. USA 79, 6979-6983
28. Pustell J. and Kafatos P. (1982) Nucleic Acid Res. 10, 4765-4782
29. McKnight S., Gavis E., Kingsburg R. and Axel R. (1981) Cell 25, 386-398
30. Breathnach R. and Chambon P. (1981) Ann. Rev. Biochem. 50, 349-383
31. Mathis D., Benoist C., Williams W., Kanter M. and McDevitt H. (1983) Cell 32, 745-754
32. Hyldig-Nielsen J., Schenning L., Hammerling U., Widmark E., Heldin E., Lind P., Servenius B., Lund T., Flavell R., Lee J., Trowsdale J., Schreier P., Zablitzky F., Larhammar D., Peterson P. and Rask L. (1983) Nucleic Acids Res. 11, 5055-5071
33. Saito H., Maki R., Clayton L. and Tonegawa S. (1983) Proc. Natl. Acad. Sci. USA 80, 5520-5524
34. Larhammar D., Hammerling U., Denaro M., Lund T., Flavell R., Rask L. and Peterson P. (1983) Cell 34, 179-188
35. Basham T. and Merigan T. (1983) J. Immunol. 130, 1492-1494
36. Malissen M., Malissen B. and Jordan B. (1982) Proc. Natl. Acad. Sci.

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024

USA 79, 893-897
37. Evans G., Margulies D., Camerini-Otero R., Ozato D. and Seidman J. (1982) Proc. Natl. Acad. USA 79, 1994-1998
38. McNicholas J., Steinmetz., Hunkapiller T., Jones P. and Hood L. (1982) Science 218, 1229-1232
39. Parnes J. and Seidman J. (1982) Cell 29, 661-669
40. Parnes J., Robinson R. and Seidman J. (1983) Nature 302, 449-452
41. Kaufman J. and Strominger L. (1982) Nature 297, 694-696
42. Das H., Lawrance S. and Weissman S. (1983) Proc. Natl. Acad. Sci. USA 80, 3543-3547
43. Rabourdin-Combe C. and Mach B. (1983) Nature 303, 670-674

Downloaded from https://academic.oup.com/nar/article/11/24/8663/2379311 by guest on 25 April 2024