Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3543–3547, June 1983
Biochemistry

Structure and nucleotide sequence of the heavy chain gene of HLA-DR

(B-cell histocompatibility gene structure/precursor protein sequence/immunoglobulin-like domain)

HRIDAY K. DAS, SIMON K. LAWRANCE, AND SHERMAN M. WEISSMAN

Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Communicated by Stanley N. Cohen, January 24, 1983

ABSTRACT **We have used a 175-nucleotide-long primer extension product corresponding to the 5′ end of HLA-DR α-chain mRNA to isolate a genomic clone from a human DNA library. The entire *HLA-DRα* gene is contained in two contiguous *Eco*RI fragments spanning about 7.5 kilobases (kb); most of the sequence has been determined. The 5′ end of the gene is contained in a 4.4-kb fragment, and the coding segments and the 3′ untranslated region are contained in a 3.1-kb fragment. The gene is split into five exons. The 5′ untranslated region, the leader peptide, and the first two NH_2-terminal amino acids are fused into the first exon. Exons 2 and 3 represent two extracellular coding domains of mature p34. The transmembrane domain, cytoplasmic domain, and part of the 3′ untranslated region are merged into a fourth exon. The rest of the 3′ untranslated region is in exon 5. The predicted amino acid sequence of mature p34, as deduced from its gene structure, has 229 residues and reveals a single potential disulfide loop (between cysteine residues 107 and 163) as well as a 22-amino acid residue membrane integrated segment (residues 193–214). Fifteen amino acids (residues 215–229) reside on the cytoplasmic side of the plasma membrane. There is considerable amino acid sequence homology between the second external domains of p34 and p29, as well as the immunoglobulin-like third domain of HLA-B7, and β_2-microglobulin and the homologous constant region domains of the light and heavy chains of immunoglobulins.**

Human chromosome 6 and murine chromosome 17 contain a cluster of genes, known as the major histocompatibility complex, that encode a number of cell surface glycoproteins involved in various immune interactions (1–3). The HLA-A, -B, and -C and H2-K, -D, and -L glycoproteins are designated class I antigens, and HLA-DR and I(a) are class II antigens. Class I antigens are expressed on almost all types of cells and consist of a heavy chain (M_r, 45,000) associated noncovalently with β_2-microglobulin (β_2m), a protein of M_r 11,400. In contrast, the class II antigens have limited tissue distribution and are predominantly expressed on B lymphocytes and macrophages. Class II antigens consist of a heavy α chain (M_r, 34,000; p34) bound noncovalently to a light β chain (M_r, 29,000; p29). The class II antigens are polymorphic; the p34 chain has limited polymorphism, whereas the p29 chains vary among individuals (4–7).

Structural analysis of murine and human class I antigens and their genes shows that they are divided into a series of discrete domains—a hydrophobic leader, three extracellular or external domains, a transmembrane domain, and a cytoplasmic domain (8–13). Similar to immunoglobulin domains, both the second and third external domains of class I antigens have a centrally placed disulfide loop.

Recently, two groups of investigators (14, 15) reported the amino acid sequence of p29 for class II antigens. One of these groups isolated a cDNA clone for a p29-like polypeptide (16). Several investigators (17–19) have isolated cDNA clones of *HLA-DRα*. To date, only a limited amino acid sequence from the NH_2 terminus of p34 is known (20).

Earlier work in this laboratory used the primer extension technique for cloning class I and II genes (19, 21). In the case of class II genes, a 20-nucleotide-long primer was synthesized and the 5′ γ-^{32}P-labeled 20-mer was hybridized to poly(A)$^+$ mRNA from B-cell line LG-2 and extended in the presence of four deoxynucleoside triphosphates and reverse transcriptase to give a 175-nucleotide-long product encoding the 5′ end of *HLA-DRα* mRNA (19, 22), which was used to isolate a genomic clone (22). The genomic DNA was subcloned into plasmid pBR328. Two clones with inserts of 3.1 and 4.1 kilobases (kb) hybridized with the p34 probe. The 5′ end of the gene was contained in a 2-kb *Bgl* II–*Bgl* II fragment of the 4.4-kb *Eco*RI insert, whereas all the coding sequence and the 3′ untranslated region of the gene were present in the 3.1-kb *Eco*RI fragment. The entire gene was contained within these two contiguous *Eco*RI fragments spanning about 7.5 kb.

We report here the nucleotide sequence of the 7.5-kb fragment of a human genomic clone of *HLA-DRα*. This sequence includes five exons and predicts the complete amino acid sequence of the mature precursor p34 polypeptide. Analysis of the structure of the *HLA-DRα* gene and the deduced amino acid sequence of mature p34 revealed that, like the class I genes, exons 2 and 3 of *HLA-DRα* may correlate with the two structurally defined external protein domains of p34. Unlike the class I antigens, the transmembrane and the cytoplasmic domain of p34 are fused into exon 4. Considerable sequence homology exists between the second external domain of mature p34 and p29, as well as the third external domain of HLA-B7, β_2m, and the corresponding homologous constant region domains of immunoglobulin light and heavy chains. These results suggest that an immunoglobulin-like domain is also preserved in the p34 polypeptide.

MATERIALS AND METHODS

Materials. Restriction enzymes and T4 DNA ligase were purchased from New England BioLabs and Bethesda Research Laboratories. The dNTPs were obtained from Collaborative Research (Waltham, MA), *Escherichia coli* DNA polymerase large fragment was from P-L Biochemicals, T4 polynucleotide kinase was made from a T4-infected *E. coli* strain. [γ-^{32}P]ATP (2,000–3,000 Ci/mmol; 1 Ci = 3.7×10^{10} Bq) and [α-^{32}P]dNTP (7,900 Ci/mmol) were purchased from Amersham.

Subcloning of *Eco*RI- and *Hin*dIII-Digested Genomic Fragments. Subcloning of 3.1-kb and 4.4-kb *Eco*RI fragments and

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.

Abbreviations: kb, kilobase(s); bp, base pair(s); β_2m, β_2-microglobulin.

a 2.3-kb *Hin*dIII fragment in pBR328 and pBR322 was done as described (22).

Restriction Map of *HLA-DRα*. A restriction map for *HLA-DRα* was constructed by the procedure of Smith and Birnstiel (23) as described (22).

DNA Sequence Analysis. To determine the sequence of the 3.1-kb *Eco*RI fragment, the plasmid was digested with *Eco*RI and *Hin*dIII, and 2-kb and 1.1-kb *Eco*RI–*Hin*dIII insert fragments were purified from a 40-cm-long 4% acrylamide gel. Similarly, a 2-kb *Bgl* II–*Bgl* II fragment of the 4.4-kb *Eco*RI insert was also purified. These DNA fragments were digested with different restriction enzymes and labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase or at the 3′ end with the appropriate [α-^{32}P]dNTP and *E. coli* DNA polymerase I large fragment or [^{32}P]cordycepin and terminal transferase.

The rest of the 4.4-kb *Eco*RI fragment and a portion of the 2.3-kb *Hin*dIII fragment were analyzed by cutting the whole plasmid with the appropriate enzymes and labeling the 3′ end of the DNA fragments with [α-^{32}P]dNTP and *E. coli* DNA polymerase large fragments. The DNA fragments were subjected to sequence analysis by the method of Maxam and Gilbert (24).

RESULTS AND DISCUSSION

We isolated a genomic clone from a partial *Eco*RI-digested human B-cell (JY) DNA library in phage λ Charon 4A with the 175-mer probe as described (22); we also isolated an additional six partial homologous genes. This original clone (*HLA-DRα*) was subcloned into the *Eco*RI sites of pBR328, and two subclones containing 3.1-kb and 4.4-kb inserts were found to hybridize with the 175-mer.

Restriction Map of *HLA-DRα*. Fig. 1 shows the restriction map of clone *HLA-DRα*. We determined the restriction maps of the 3.1-kb and 4.4-kb *Eco*RI fragments separately. Because only the 2-kb-long *Bgl* II–*Bgl* II fragment of the 4.4-kb *Eco*RI insert hybridized with the 175-nucleotide-long probe (data not shown), the 5′ portion of the gene must be contained within this fragment. Exons and introns were assigned only on the basis of DNA sequence data. In order to construct a restriction map of *HLA-DRα*, we also subcloned the genomic insert into the *Hin*dIII site of pBR322. Only a single clone, containing a 2.3-kb *Hin*dIII insert, hybridized with the 2-kb *Eco*RI–*Hin*dIII DNA fragment of the 3.1-kb *Eco*RI insert. The 2.3-kb *Hin*dIII fragment contained a single *Eco*RI restriction site approximately 250 base pairs (bp) from its 5′ *Hin*dIII end. The nucleotide sequence of the 5′ *Hin*dIII end across the *Eco*RI site was identical with that of the 3′ *Hin*dIII–*Eco*RI DNA fragment of the 4.4-kb *Eco*RI clone and also overlapped the 5′ *Eco*RI end of the 3.1-kb *Eco*RI clone. This result showed that *HLA-DRα* gene is contained in two contiguous *Eco*RI fragments spanning about 7.5 kb.

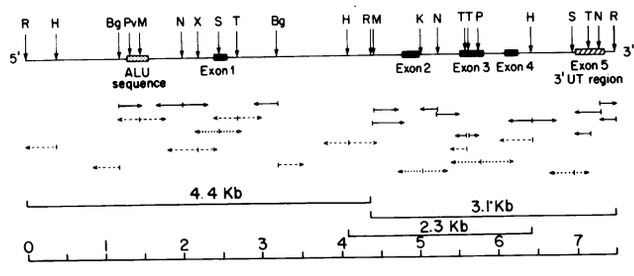

FIG. 1. Restriction map of the *HLA-DRα* gene and positions of individual sequences obtained. The 4.4-kb and 3.1-kb *Eco*RI fragments and 2.3-kb *Hin*dIII fragment were isolated for DNA sequence analysis. The location of exons are indicated by black rectangles. The hatched box denotes the 3′ untranslated region. The *Alu* sequences are also shown. E, *Eco*RI; H, *Hin*dIII; Bg, *Bgl* II; Pv, *Pvu* II; M, *Msp* I; N, *Nco* I; X, *Xba* I; S, *Sst* I; T, *Taq* I; K, *Kpn* I; P, *Pst* I. Solid arrows, length of the sequence obtained in the 5′ → 3′ direction with [γ-^{32}P]ATP-labeled fragments; dashed arrows, length of sequence obtained in the 3′ → 5′ direction obtained by 3′-end-labeling with [α-^{32}P]dNTP; dotted arrows, length of sequence obtained in the 3′ → 5′ direction by labeling the 3′ end with cordycepin.

Nucleotide Sequence of *HLA-DRα*. The nucleotide sequence of *HLA-DRα* was determined by the method of Maxam and Gilbert (24). The DNA sequence starting from the 5′ *Bgl* II end is given in Fig. 2. The nucleotide sequence was determined on both strands at many parts of the insert. The sequence was confirmed by analysis twice in the same direction. Comparison of the genomic sequence with that of the 175-nucleotide-long extension product (19) and with the cDNA sequence of *HLA-DRα* (18) clearly indicates that *HLA-DRα* is split into five exons. Exon 1, assigned on the basis of the nucleotide sequence of the 175-mer extension product, contains the 5′ untranslated region and a leader peptide containing 25 amino acids, of which 18 are hydrophobic. In contrast to class I genes (10–13), the first two NH_2-terminal amino acids are contained in exon 1. In the case of class I genes, exon 1 was identified by characterizing the donor–acceptor sites in the intron–exon junction because the sequence of the 5′ end of the mRNA was not known. In our case, with the primer extension method, the 5′ sequence of *HLA-DRα* was obtained and hence the sequence of the 5′ hydrophobic signal peptide is confirmed.

Exons 2–5 were assigned on the basis of the presence of typical donor–acceptor sites and confirmed by cDNA sequence data (18). Exon 2 contains amino acid residues 3–84 and may represent the first external domain of the p34 polypeptide. Amino acid residues 85–178 are contained in exon 3 which comprises the second external domain of p34. Exon 4 contains amino acid residues 179–229 and a stop codon at amino acid position 230. The predicted amino acid sequence data in this region indicate the presence of a stretch of hydrophobic amino acids (residues 193–214) followed by 15 amino acids (residues 215–229) of which most are hydrophilic. In the case of class I antigens and also the human p29 polypeptide, it has been found that a stretch of hydrophobic amino acids at the COOH terminus represents the membrane-integrated region. This is followed by a sequence of charged amino acids representing the cytoplasmic domain of the proteins (14, 25). Thus, in the case of p34, the hydrophobic amino acid residues 193–214 represent the membrane-integrated region of p34 and residues 219–229 are exposed inside the cytoplasm. By analogy with the structure of class I genes, amino acid residues 179–214 and residues 215–229 represent the transmembrane and cytoplasmic domains of the p34 polypeptide, respectively (25).

The structure of class I genes revealed that the transmembrane domain is present in a separate exon and the cytoplasmic domain is split into several exons (10–12). The sequence of the class I genes in this region raises the possibility that alternative splicing events may produce alternative forms of class I mRNA, resulting in proteins containing different cytoplasmic domain. Because both the transmembrane and cytoplasmic domains of p34 are fused into a single exon, there is less possibility of alternative splicing events in this region of *HLA-DRα*. Thus, only one form of p34 polypeptide is implicated by the gene structure. Exon 4 also contains 11 nucleotide residues from the 3′ untranslated region of the gene. The rest of the 3′ untranslated region is contained in exon 5. In the case of the *HLA-B7* gene, the 3′ untranslated region is also present in a separate exon, whereas the 3′ untranslated region of a mouse transplantation antigen gene is fused with the last cytoplasmic domain in a single exon (10).

Correction. In the article "Structure and nucleotide sequence of the heavy chain gene of HLA-DR," by Hriday K. Das, Simon K. Lawrance, and Sherman M. Weissman, which appeared in number 12, June 1983, of *Proc. Natl. Acad. Sci. USA* (**80**, 3543–3547), a whole line of sequences and some symbols were omitted during the revision of Fig. 2 on page 3545. Failure to fuse sequences properly at two restriction sites also led to duplications of a few residues. These and a few other nucleotides have been corrected.

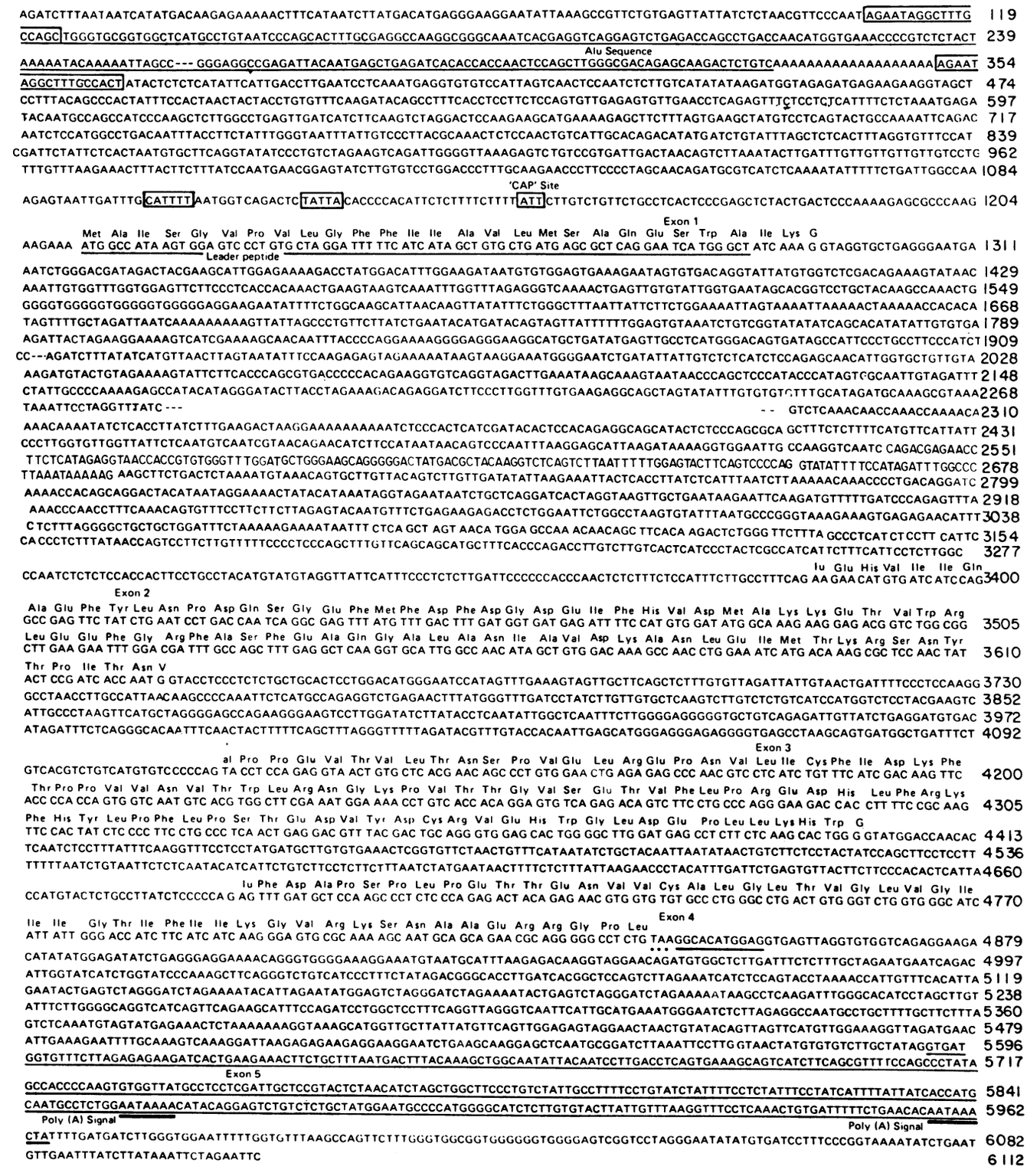

FIG. 2. DNA sequence of the 6.3-kb *Bgl* II–*Eco*RI DNA fragment of *HLA-DRα* gene. The amino acid sequences encoded by exons of *HLA-DRα* are shown above the DNA sequences. The 3′ untranslated region is underlined. The region of *Alu* sequences, the CAT box, the TATA box (promoter), and the mRNA transcription initiation site (CAP site) are marked. ***, Stop codon; polyadenylylation signals are also indicated. ---, Stretch of DNA for which the sequence has not been determined. Numbers on the right have not been adjusted for the estimated lengths of segments of DNA represented by dashes.

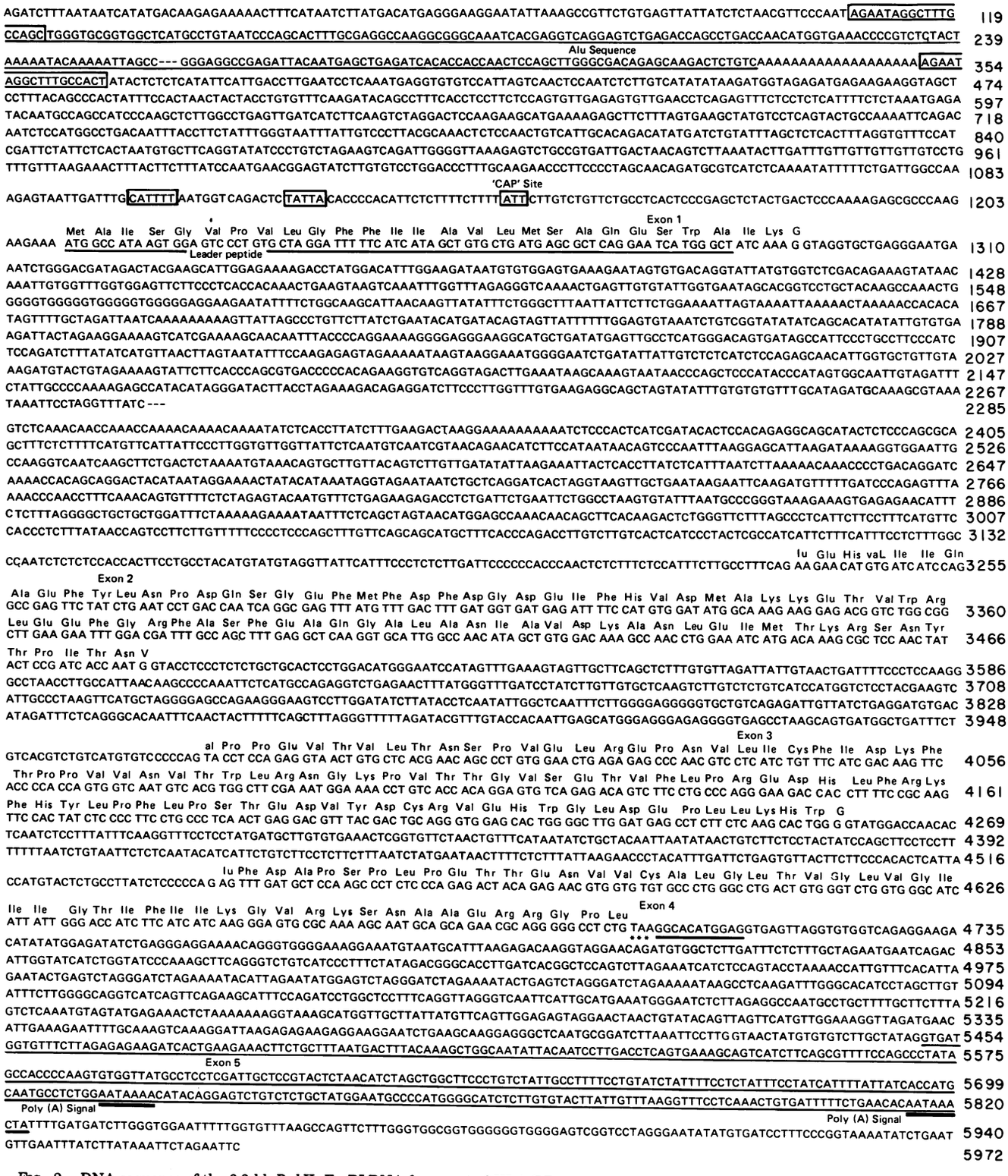

FIG. 2. DNA sequence of the 6.3-kb *Bgl* II–*Eco*RI DNA fragment of *HLA-DRα* gene. The amino acid sequences encoded by exons of *HLA-DRα* are shown above the DNA sequences. The 3′ untranslated region is underlined. The region of *Alu* sequences, the CAT box, the TATAA box (promoter), and the mRNA transcription initiation site (CAP site) are marked. ***, stop codon; polyadenylylation signals are also indicated.

Correlation of the structural domains of p34 polypeptide with the different exons of *HLA-DRα* is presented in Fig. 3. Nucleotide sequence data show the presence of two poly(A) signal sites characterized by the sequence A-A-T-A-A-A. The second poly(A) signal site exists about 97 nucleotide downstream from the first one. Because the two poly(A) signal sites also exist in the cDNA of *HLA-DRα* mRNA and at the same distance apart, only the second poly(A) signal site seems to be used for polyadenylylation of the *HLA-DRα*.

Comparison of the nucleotide sequence of the 5′ end of the *HLA*-DRα gene with the sequence of the 175-nucleotide extension product (19, 22) shows that about 60 nucleotides upstream from the initiation codon there exists an ATT sequence that most probably is the transcription initiation site or cap site.

All eukaryotic genes contain an A+T-rich region about 30 nucleotides upstream from the cap site which is known as the

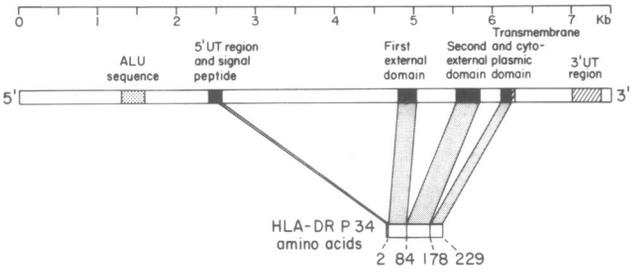

FIG. 3. Organization of exons and introns in *HLA-DRα* gene. Exons 1–4 are shown as black boxes; exon 5, which represents the 3′ untranslated region, is shown as the hatched box. Exon 1 has been assigned on the basis of the DNA sequences of the 175-mer. Exons 2–5 are assigned on the basis of the presence of donor and acceptor sites in the sequence, confirmed by the published cDNA sequence (18) and by a personal communication from H. Erlich. Protein domains are postulated on the basis of protein comparison.

TATA box (26). At approximately 70 nucleotides upstream from the transcription initiation site, another region, known as the CAT box, is found which is required for efficient transcription of the gene (27). Like other eukaryotic genes, *HLA-DRα* has a short TATTA sequence approximately 30 nucleotides upstream from the potential cap site and probably represents the promoter of the gene. But, unlike other genes, there exists a C-A-T-T-T-T sequence approximately 48 nucleotides upstream of the cap site, which may represent the potential CAT box. It remains to be seen whether this sequence is required for efficient transcription of the gene.

Approximately 1 kb upstream from the transcription initiation site, an *Alu* sequence (28) is found flanked by direct terminal repeats.

Predicted Amino Acid Sequence of the Mature Precursor p34 Chain. The complete amino acid sequence of the mature precursor p34 chain is shown in Fig. 4. The amino acid sequence has been deduced from the nucleotide sequence of exons 1–4. The first 25 amino acids, most of which are hydrophobic, represent the signal peptide of p34. The deduced amino acid sequence predicts that the mature protein is composed of 229 amino acid residues. From the protein comparison (25), it is likely that approximately the first 84 amino acids represent the first external domain and amino acid residues 85–178 comprise the second external domain. In the second external domain there are two cysteine residues, at positions 107 and 163, which form a disulfide loop encompassing 55 amino acids. There are two sites, Asn^{78}-Tyr^{79}-Thr^{80} and Asn^{118}-Val^{119}-Thr^{120}, which represent the region of glycosylation of the membrane protein. Thus, carbohydrate moieties are most likely linked to these asparagines, one in each of the two extracellular domains.

Amino acid residues 179–214 represent the transmembrane domain. The 22 amino acids (residues 193–214) that we presume are integrated into the lipid bilayer of the plasma membrane are preceded by 14 amino acids (residues 179–192) of which most are negatively charged and followed by a heptapeptide containing basic amino acids (Lys^{215}, Arg^{218}, and Lys^{219}), which could represent the region of the p34 polypeptide integrated with the phosphate residues of phospholipids to the plasma membrane. There is another cysteine residue at position 195 in the membrane-integrated segment which does not appear to take part in disulfide bond formation. There is little homology between the amino acid sequence in the membrane-integrated regions of p34, p29, and class I transplantation antigens. Unlike class I antigens, only a short segment—10 amino acids (residues 220–229)—marks the COOH-terminal region of p34. As expected, there is little homology between and among class I and class II antigens in the cytoplasmic domain. A single serine residue at position 220 represents the potential site for phosphorylation as in class I antigens (25) and p29 (14, 15).

Amino Acid Sequence Homology Among p34, p29, HLA-B7, β_2m, and Immunoglobulin Constant Region. In order to find possible amino acid sequence homologies, we compared the p34 sequence with that of p29, human class I antigen HLA-B7, β_2m, and various regions of immunoglobulin. We found considerable amino acid sequence homology between the second external domain of p34 and p29, as well as the third external domain of HLA-B7, the constant region of human κ chains, the three domains of γ chains, and β_2m. Fig. 5 compares amino acid sequences of p34 (residues 152–182) and p29, HLA-B7, constant region domains of human κ and γ chains, and β_2m.

Although the sequence homology is scattered throughout the second external domain of p34 (about 25–30% amino acid positions are identical), it is especially noticeable around cysteine residue 163. Between amino acid residues 159 and 170 of p34, eight amino acids are identical with p29 and HLA-B7 and seven amino acids are identical with β_2m, human immunoglobulin κ light chain constant region domain, and three domains of γ chains. The data suggest that all three classes—immunoglobulin, class I, and class II—appear to have a similar domain in the protein. The disulfide loop in the second external domain of p34, p29, in the third external domain of HLA-B7, and β_2m encompasses exactly 55 amino acids. The amino acid sequence in the third external domain of HLA-B7 and the second ex-

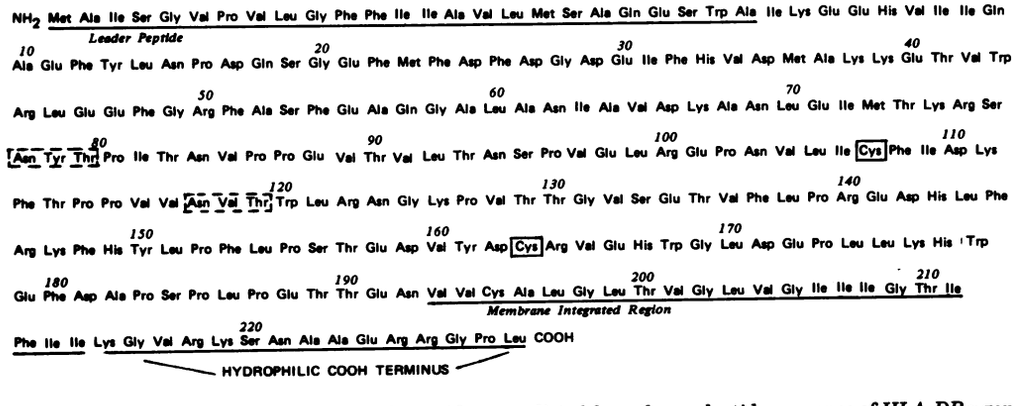

FIG. 4. Amino acid sequence of precursor mature p34 polypeptide as predicted from the nucleotide sequence of *HLA-DRα* gene. The leader peptide, membrane-integrated region, and hydrophilic COOH-terminal region are underlined. Amino acids 1–84 represent the first external domain, 85–178 are the second external domain, 179–214 are the transmembrane domain, and 215–229 are the cytoplasmic domain. Lys^{215} to Ser^{220} represents the amino acid sequence required to anchor the lipid bilayer on the cytoplasmic side. The single potential disulfide loop is between Cys^{107} and Cys^{163}, shown in boxes. The two potential sites for glycosylation at Asn^{78} and Asn^{118} are shown in boxes.

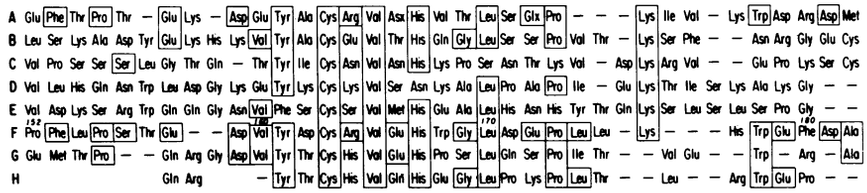

FIG. 5. Comparison of amino acid sequences of human p34 and of p29, HLA-B7, human β_2m, constant region of human κ chain, and three domains of human γ chains. Sequences of β_2m, κ chain, and γ chains are taken from Peterson *et al.* (29); those of p29 and HLA-B7 are from Larhammar *et al.* (14) and Ploegh *et al.* (25), respectively. Identical residues are enclosed in boxes. Numbering is for p34 polypeptide. A, β_2m (residues 70–100); B, EU-CL (residues 181–214); C, EU-CH^1 (residues 188–220); D, EU-CH^2 (residues 308–341); E, EU-CH^3 (residues 412–446); F, HLA-DR p34 (residues 152–182); G, HLA-DR p29 (residues 162–190); H, HLA-B7 (residues 255–276).

ternal domain of p29 have about 50% homology with p34. The homology is prominent around Cys^{107} and Cys^{163}. The homology with p29 around the region of Cys^{107} is strong. Between residues 103 and 107, three amino acid positions are identical and there are two conservative substitutions. Between residues 107 and 163, 19 amino acid positions are identical and 3 conservative substitutions are found between p34 and p29. There is little amino acid sequence homology in the first extracellular domains of p34 and p29 or the first and second extracellular domains of HLA-B7.

After this manuscript was prepared, the complete nucleotide sequence of p34 cDNA and the complete amino acid sequence of p34 as deduced from the cDNA of p34 were published (30). The nucleotide sequence of the cDNA matched completely with that of the exons of *HLA-DRα*, except that the codon in use for the cDNA at amino acid position 82 was ATT instead of ATC. Also, on the cDNA sequence, amino acid position 178 was a termination codon, TAG, instead of TGG encoding tryptophan. During the review of this manuscript, a partial nucleotide sequence of *HLA-DRα* was published (31). Comparison of our sequence with that of Korman *et al.* shows 19 differences, all of which are single-base substitutions in the introns.

This investigation was supported by National Cancer Institute Grant CA-30938 to S.M.W.; H.K.D. was supported by a postdoctoral fellowship from the Cancer Research Institute, Inc., New York.

1. McDevitt, H. O. (1976) in *The Role of the Products of the Histocompatibility Gene Complex in Immune Responses*, eds. Katz, D. H. & Benacerraf, B. (Academic, New York), pp. 257–275.
2. Ferrone, S., Allison, J. P. & Pellegrino, M. A. (1978) *Contemp. Top. Mol. Immunol.* **7**, 239–281.
3. Amos, D. B. & Kostyu, D. D. (1980) *Adv. Hum. Genet.* **10**, 137–208.
4. McMillan, M., Cecka, J. M., Murphy, D. B., McDevitt, H. O. & Hood, L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5135–5139.
5. Silver, J., Walker, L. E., Reisfeld, R. A., Pellegrino, M. A. & Ferrone, S. (1979) *Mol. Immunol.* **16**, 37–42.
6. Springer, T. A., Kauf, J. F., Terhorst, C. & Strominger, J. L. (1977) *Nature (London)* **268**, 213–218.
7. Allison, J. P., Walker, L. E., Russell, W. A., Pellegrino, M. H., Ferrone, S., Reisfeld, R. A., Frelinger, J. A. & Silver, J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3953–3956.
8. Colligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. & Nathanson, S. G. (1981) *Nature (London)* **291**, 35–39.
9. Lopez de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T., Mann, D. L. & Strominger, J. L. (1979) *Biochemistry* **18**, 5704–5711.
10. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, A. W. & Hood, L. (1981) *Cell* **25**, 683–692.
11. Moore, K. W., Sher, B. T., Sun, H.-Y., Eakle, K. A. & Hood, L. E. (1982) *Science* **215**, 679–682.
12. Biro, P. A., Pan, J., Sood, A. K., Reddy, V. B. & Weissman, S. M. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, in press.
13. Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1994–1998.
14. Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. & Peterson, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3687–3691.
15. Kratzin, H., Yang, C.-Y., Gotz, H., Pauly, E., Kolbel, S., Egert, G., Thinnes, F. P., Warnet, P., Altevogt, P. & Hilschmann, N. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1665–1669.
16. Wiman, K., Larhammar, D., Claesson, K., Gustafsson, K., Schenning, L., Bill, P., Bohme, J., Denero, M., Dobberstein, B., Hammerling, U., Kvist, S., Servenius, B., Sundelin, J., Peterson, P. A. & Rask, L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1703–1707.
17. Lee, J. S., Trowsdale, J. & Bodmer, W. F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 545–549.
18. Korman, A. J., Knudsen, P. J., Kauffman, J. F.& Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1844–1848.
19. Stetler, D., Das, H., Nunberg, J. H., Saiki, R., Sheng-Dong, R., Mullis, K. B., Weissman, S. M. & Erlich, H. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5966–5970.
20. Klein, J. (1979) *Science* **203**, 516–521.
21. Sood, A. K., Pereira, D. & Weissman, S. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 616–620.
22. Das, H. K., Biro, P. A., Cohen, S. N., Erlich, H. A., von Gabain, A., Lawrence, S. K., Lemaux, P. G., McDevitt, H. O., Peterlin, B. M., Schulz, M.-F., Sood, A. K. & Weissman, S. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1531–1535.
23. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
25. Ploegh, H. L., Orr, H. T.& Strominger, J. L. (1981) *Cell* **24**, 287–299.
26. Goldberg, P. (1979) Dissertation (Stanford Univ., Stanford, CA).
27. Benoist, E., O'Hare, K., Bretnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8**, 127–140.
28. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **216**, 1065–1070.
29. Peterson, P. A., Cunningham, B. A., Berggård, I. & Edelman, G. M. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1697–1701.
30. 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.
31. Korman, A. J., Aufray, C., Schamboeck, A. & Strominger, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6013–6017.