The amino acid sequence and gene organization of the heavy chain of the HLA-DR antigen: Homology to immunoglobulins

(major histocompatibility complex/Ig-like domain)

ALAN J. KORMAN, CHARLES AUFRAY, ANTON SCHAMBOECK, AND JACK L. STROMINGER

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

Contributed by Jack L. Strominger, July 6, 1982

ABSTRACT The amino acid sequence of the heavy chain of HLA-DR antigens has been elucidated from the analysis of a genomic clone coding for this protein. A 3.2-kilobase *EcoRI* fragment includes four exons containing 227 amino acids out of 229 in the mature HLA-DR heavy chain. One exon ($\alpha 2$) encodes a domain of 94 amino acids with strong sequence homology both to Ig constant region domains and to Ig-like domains in HLA-B7, β_2 -microglobulin, and the HLA-DR light chain. These results support a structure for the HLA-DR antigen heterodimer consisting of four extracellular domains, two of which are Ig-like [one in the heavy chain ($\alpha 2$) and one in the light chain ($\beta 2$)]. The third is the amino-terminal polymorphic domain in the light chain ($\beta 1$), and the fourth is an invariant domain in the heavy chain ($\alpha 1$).

The HLA-DR antigens, also called DR antigens, are two-chain membrane glycoproteins consisting of a 34,000-dalton heavy chain in noncovalent association with a 29,000-dalton light chain(s) (reviewed in ref. 1). The heavy chain is invariant between individuals whereas the light chain(s) is polymorphic. Differences in these antigens are responsible for proliferation in the mixed lymphocyte reaction. The DR antigens are transplantation antigens that can also be serologically defined. In the mouse, the I-E antigens are structurally and functionally homologous to the HLA-DR antigens and have been genetically mapped to the I region of the 17th chromosome. These Ia antigens, which include the related I-A locus product, are the products of the immune response genes. These genetic loci were first defined by the capacity of different strains of mice to produce an antibody response to a defined protein antigen. It was later shown that the I-region antigens are expressed on the surface of antigen presenting cells (macrophages) and together with "antigen" stimulate I-region-restricted T helper cells to proliferate. However, the site and mechanism of action of the I region effect remains to be fully clarified (2, 3).

Until recently, there has been little information regarding the primary structure of the DR and Ia antigens. Current interest in the molecular biology of the major histocompatibility complex (MHC) has resulted in approaches to this information by using recombinant DNA methods. The primary structure of the HLA-DR heavy chain, as elucidated from the analysis of the gene that encodes it, is presented here. The sequence reveals that the DR heavy-chain gene has an exon encoding a sequence that is highly homologous to Ig and to other Ig-homologous domains of the transplantation antigens—the $\alpha 3$ region of HLA-B7, β_2 -microglobulin (β_2m), and the $\beta 2$ region of DR light chains. These data together with recent sequence data and structural work (4–6) on the DR light chain suggest a structure for the DR and Ia antigens that is similar to the structure for

the class I transplantation antigens—HLA-A, -B, -C and H2-K, -D, -L (7, 8).

MATERIALS AND METHODS

Isolation of Genomic Clones. Genomic clones for the HLA-DR heavy chain were isolated from the *Hae* III/*Alu* I partial library of human fetal liver DNA in Charon 4A [kindly supplied by T. Maniatis (9)] by hybridization with labeled *Pst* I/*Bgl* II fragment from pDRH-2 (10). DNA from positive phage was isolated and digested with *EcoRI* and a 3.2-kilobase (kb) band hybridizing with pDRH-2 was identified by hybridization after transfer to nitrocellulose paper. The DNA from one of these phages (λ DRH6A) was digested with *EcoRI* and ligated to phosphatase-treated *EcoRI*-cut pUR222 (11). This mixture was used to transform competent F[–]Z[–]M15recA cells and plated on 5-chloro-4-bromo-3-indolyl- β -D-galactoside indicator plates. White colonies were screened and a clone containing the 3.2-kb *EcoRI* fragment was grown [pUR5(3.2RI)].

cDNA clones for the heavy chain of HLA-DR were isolated from a library prepared from size-fractionated mRNA from the JY (DR4, w6) cell line by hybridization with the 3.2-kb *EcoRI* fragment of λ DRH6A (12). All sequences were obtained by using the method of Maxam and Gilbert, substituting 60% formic acid for A+G-specific cleavage reaction (13).

S1 Nuclease Mapping. pUR5(3.2RI) was digested with *Bal* I and the 3' ends were labeled with ³²P-labeled nucleotide triphosphates by T4 DNA polymerase (14). The DNA was hybridized with 20 μ g of total JY mRNA and digested with S1 nuclease as described by Berk and Sharp (15).

Computer Analysis. The comparison of the HLA-DR heavy chain sequence to other proteins was performed by W. Barker of the National Biomedical Research Foundation (Washington, DC) using the ALIGN program (16) (see Table 1).

RESULTS AND DISCUSSION

Isolation of Genomic Clones. The isolation of a cDNA clone (pDRH-2) coding for 31 amino acids of the DR heavy chain and the entire 3' untranslated region of the mRNA after immunopurification of polysomes by a monoclonal antibody against the isolated DR heavy chain has been reported (10). A human genomic library of DNA fragments in Charon 4A (9) was screened by hybridization with pDRH-2. Four independently hybridizing clones were isolated, each of which contained a 3.2-kb *EcoRI* fragment that hybridizes with pDRH-2. In addition, ³²P-labeled cDNA complementary to purified HLA-DR heavy

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); MHC, major histocompatibility complex; β_2m , β_2 -microglobulin; C region, Ig constant region; DR, DR-H, and DR-L, HLA-DR antigen(s) and its heavy and light chain(s).

chain mRNA hybridized with the 3.2-kb *EcoRI* fragment. In hybridizations to *EcoRI*-digested genomic DNA, Lee *et al.* (17) reported a single band of 3.4 kb hybridizing with their nearly full-length clone for the HLA-DR heavy chain. These data suggest that the DR heavy chain gene is a single-copy sequence.

The 3.2-kb *EcoRI* fragment from one of the λ genomic clones (ADRH6A) was subcloned into the *EcoRI* site of pUR222 (11). The organization and nucleotide sequence of the coding portions of the DR heavy chain gene and some of the introns contained in this subclone are presented in Fig. 1. The coding portions of the gene were determined by comparison with the nucleic acid sequences of cDNA clones for the DR heavy chain, comparison with the amino-terminal protein sequence, and S1 nuclease mapping of RNA-DNA hybrids. This DNA segment encodes four exons of the DR heavy chain: (i) $\alpha 1$, consisting of amino acids 3–84; (ii) $\alpha 2$, consisting of amino acids 85–178, whose sequence bears a marked resemblance to Ig constant

regions (C regions); (iii) an exon encoding amino acids 179–229, consisting of a connecting peptide, the hydrophobic transmembrane segment, and the hydrophilic intracytoplasmic carboxyl terminus, followed by a stop codon [this exon also contains 11 base pairs (bp) of the 3' untranslated region]; and (iv) an exon containing the remaining part of the 3' untranslated region. An exon(s) containing the 5' untranslated region and the signal sequence including amino acids 1 and 2 (Ile-Lys) of the mature DR heavy chain is not present in this DNA segment. The existence of a signal sequence was shown by cell-free translation of the HLA-DR heavy chain mRNA (18). Primer extension of a 230-bp *Sau3A/Sau3A* fragment (beginning at amino acid 7) indicates that ≈ 150 bp of DR-H mRNA should be encoded 5' to the 3.2-kb *EcoRI* fragment.

The $\alpha 1$ Exon. Approximately 400 bp from the 5' end of the 3.2-kb *EcoRI* fragment begins an open reading frame corresponding to the amino-terminal protein sequence of the mature

FIG. 1. The organization and partial nucleotide sequence of the gene for the HLA-DR heavy chain. (Upper) The intron/exon organization of the HLA-DR heavy chain gene is shown along with a restriction enzyme map of the 3.2-kb *EcoRI* subclone and in comparison with two cDNA clones for the DR heavy chain (see text). CP, connecting peptide; TM, transmembrane region; CY, intracytoplasmic region; 3'UT, 3' untranslated region. Φ , *Sau3A*; Δ , *Taq I*; $\downarrow$, *Hinf I*. (Lower) The partial nucleotide sequence of the 3.2-kb subclone is presented with nucleotides numbered as approximate distance from the 5' *EcoRI* site. The 3' untranslated region and the glycosylation sites are underlined. The numbers above the deduced protein sequence and the determined NH₂-terminal sequence (top line) refer to the amino acid number of the mature DR heavy chain.

FIG. 2. Determination of the 3' intron-exon border of the $\alpha 1$ domain by S1 nuclease mapping. ^{32}P -Labeled *Bal* I fragments totaling 5,000 cpm were hybridized to total JY mRNA, digested with S1 nuclease, and run on a 5% polyacrylamide sequence gel. Lanes: M, DNA size markers in bp; 1, control: input probe; 2, control: treatment in the absence of mRNA; 3, treatment in the presence of JY mRNA. A protected species at 73–76 bp is detected.

DR heavy chain (19). The exon begins with bases 2 and 3 of a Glu codon at position 3 and an open reading frame extends ≈ 500 bp downstream. Two potential splice sites are present within this reading frame. The location of the 3' end of this exon was determined by S1 nuclease mapping (Fig. 2). DNA labeled at the unique *Bal* I site and hybridized with total JY mRNA protects from S1 nuclease digestion a segment of DNA $\approx 75 \pm 2$ nucleotides long. This localizes the 3' splice junction at codon 84 including the first nucleotide of the next codon.

The $\alpha 1$ exon contains 82 amino acids of the DR heavy chain. It contains no cysteines and contains all the methionines in the molecule. By comparison to domains of HLA-B7 (7) and the DR light chain (5), it appears to be ≈ 10 amino acids smaller. The sequence has been compared to Ig variable and C regions and to HLA-B7, $\beta_2\text{m}$, HLA-DR light chain, and Thy-1 by using the ALIGN program (Table 1). It has no homology to any of these proteins.

The $\alpha 2$ Exon and Homology to Ig. The nucleotide sequence around the single *Pst* I site revealed that this site is situated in an open reading frame that contains two cysteine codons (the second of which is the T-G-C of the *Pst* I site). A disulfide loop was anticipated from the mobility difference of the DR heavy chain in the presence or absence of reducing agents (19). The

exon boundaries were established in several ways. A cDNA library of size-selected JY mRNA was screened by using the 1,400-bp *Eco*RI/*Pst* I and 1,800-bp *Pst* I/*Eco*RI fragments as probes. A number of strongly hybridizing clones have been identified. One clone, pDRH-3, is identical to the sequence of the genomic clone in the $\alpha 2$ region and codes for amino acids 121–193. pDRH-3 therefore identifies the 3' end of this exon (50 bp from the *Pst* I site) at amino acid 178. Towards the 5' end of this open reading frame (≈ 220 bp upstream of the *Pst* I site), a series of A-G dinucleotides is followed by a T+C-rich region typical of acceptor splice junctions. In a recent compilation of splice junction sequences (27), the average distance between the A-G of the splice site and the nearest A-G upstream is 41 bp. The minimal distance observed is 15 bp. Examination of the sequence of the genomic clone in this region places the splice site after the 5'-most A-G and leaves two nucleotides of a codon to splice with a single nucleotide from $\alpha 1$. The next A-G dinucleotide is found 36 bp in 5' direction. In addition, a *Staphylococcus aureus* V8 protease peptide from the heavy chain (amino acids 72–88) has been isolated that spans the $\alpha 1$ – $\alpha 2$ junction, together with many other peptides that confirm the sequence (P. Knudsen, personal communication).

The $\alpha 2$ exon encodes a domain of 94 amino acids. The deduced protein sequence has all the general features that give it an unmistakable homology to C regions and to similar domains in HLA-B7 ($\alpha 3$), $\beta_2\text{m}$, and the HLA-DR light chain [DR-L ($\beta 2$)]. These homologies in amino acid sequence are believed to represent similar patterns of tertiary folding (28). Different C regions have as much homology to each other by amino acid sequence as they do to the $\alpha 3$ domain of HLA-B7, $\beta_2\text{m}$, and DR-L ($\beta 2$) and, as shown here, the $\alpha 2$ domain of DR-H. Despite differences in primary sequence, C regions appear quite similar in structure when examined by x-ray crystallography (29).

The $\alpha 2$ domain of the HLA-DR heavy chain contains two cysteines that form a disulfide loop of 55 amino acids and has a Trp at amino acid 121. These residues are invariant in all C regions and are found in similar positions in HLA-B7 ($\alpha 3$) and DR-L ($\beta 2$). (In $\beta_2\text{m}$, Trp is replaced by Leu.) The greatest amino acid homology among the Ig-homologous domains of the HLA antigens and C regions is found around the two cysteines (Fig. 3). These homologies correspond to the alternating pattern of hydrophobic amino acids whose side chains fill the internal spaces between two faces of a C region. These faces are composed of four β -strands that form the x-face (fx1–4) and three β -strands (fy1–3) that form the y-face on the opposing side. For example, the sequence in fy2 consisting of Tyr(or Phe)-X-Cys-X-Val-X-His is highly conserved in C regions and is present in all the sequences examined. When the residues in a β -strand are not identical, they are often replaced by similar hydrophobic residues. It is in the bends or non- β pleated segments, b1–6,

Table 1. Homology of DR $\alpha 1$ and $\alpha 2$ to Ig C regions and Ig-homologous domains by using the ALIGN program

	V_H	$C_{\mu H1}$	$C_{\mu H2}$	$C_{\mu H3}$	$C_{\mu H4}$	$C_{\mu H5}$			DR-H	HLA-B7				DR-L	DR-L	
	(1–119)	(1–103)	(104–215)	(216–321)	(322–433)	(218–324)	C_α	C_λ	$\alpha 2$	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\beta_2\text{m}$	$\beta 1$	$\beta 2$	Thy-1
									(85–178)	(1–90)	(91–181)	(182–271)	(1–100)	(1–94)	(95–187)	(1–111)
DR-H																
$\alpha 1$ (1–84)	0.49	0.54	–0.67	–0.23	–0.23	0.63	–0.10	–1.65	–1.05	1.37	–0.04	–0.13	0.49	0.72	–0.14	0.90
DR-H																
$\alpha 2$ (85–178)	3.57	3.80	7.01	4.29	7.30	5.51	7.16	8.92	—	–0.43	–0.24	7.63	10.01	2.22	11.46	2.10

Statistical relatedness (in A scores) of HLA-DR heavy chain $\alpha 1$ and $\alpha 2$ domains to Ig domains [V_H (20), C_μ (21), C_γ (22), C_κ (23), C_λ (24)] and HLA-A, -B (7), $\beta_2\text{m}$ (25), DR light chain (5), and Thy-1 (26) by using the ALIGN program (16). This program calculates a value for the maximal alignment between any two sequences. The scoring is based on a mutation data matrix calculated from a rate of accepted point mutations (PAMs), in this case, 250 plus a bias parameter of 6. A penalty of 6 for each gap is also included. A score of 6 or greater is considered significant when the proteins are thought to be unrelated. A score of 3 or greater may be considered significant if the proteins are believed to be related.

DR-H	VPP[EV]-TVLTNSPVLELREP[NV]I[Q]FIDK[ET]PPVNV[IT]MLRW-[G]KPVTTG						
DC1	EVPP[EV]-TVVFSKSPVTLGEPNTL[IL]CLNDNTFPPVNV[IT]MLRW-[G]HVSVEG						
DR-L	VPP[EV]-TVVYPSKQPTLQHHNL[LV]CSMSG[FY]PGS[ET]VRFVFLW-[G]EKEKAG						
HLA-B7	ADPPK-[THVTHHPISDHEAT-[LR]CMALG[FY]PAE[IT]LTMOR-[G]EDDTQ						
β_2m	RTPKI-[GVYSRHPAENGKSNFLN]CYMSG[FY]HPSD[IE]VDLLK-[G]ER-IE						
C_H3	REPP[EV]TLPPSREEMTKNQVS[LT]CLMKG[FY]PSD[IE]V[ES]D[EG]-[G]-PE						
	fx1	b1	fx2	b2	fy1	b3	
DR-H	VSETVFLPRE-[D]H[F]RKFHYLPFL[ST]-[E]D-VYD[CR]VE[H]G[LD]-[E]PLLKHW						
DC1	VSETVSFLSKS-[D]HS[F]FKISYLTFL[PSA]-[E]D-IYD[CK]VE[H]G[LD]-[E]PLLKHW						
DR-L	GVSTGLIQD-[D]MT[F]QTLVNLLETVP[RS]G-[E]-VYTC[Q]VE[H]PSVT-[S]PLTVEW						
HLA-B7	DTLVETRPAG[RT]FEKWA[V]V-PS-[G]-EQR[RY]TC[Q]VE[H]G[LP]-[K]PLTLRW						
β_2m	KVENSDFSFSK[DN]SYFYLLYYTEFT[PT]-[E]KDEY[AC]RVN[H]VTL[LS]-[Q]KIVK[M]						
C_H3	NYKTTTPVLDS[DG]S[F]LYSKLTVDKSRW[EG]NV[SC]V[VM]EAL[H]NHYT[Q]KSL						
	fx3	b4	fx4	b5	fy2	b6	fy3

FIG. 3. Comparison of the amino acid sequence of the $\alpha 2$ domain of HLA-DR heavy chain to Ig-like domains in DC1 heavy chain (12), HLA-B7 (7), HLA-DR light chain (5), β_2m (25), and Ig C_H3 (22). The alignments are made to illustrate those features that are held in common with the structure of a C region determined by x-ray crystallography. The underlined segments refer to stretches of β -pleated sheet, whereas underlined residues refer to those amino acids that reside in the internal portion of the folded Ig domain (28).

where amino acid sequence homology is poorest and where deletions and insertions are frequently found in C regions. In addition, the fx3 strand is the least conserved among C regions and this segment also appears dissimilar between HLA-DR-H, HLA-DR-L, HLA-B7, and β_2m . The fx3 strand forms the fewest interdomain contacts between two C regions (28). The close relatedness of these sequences has also been unequivocally demonstrated statistically by using the ALIGN program (Table 1).

Included in the homology comparison (Fig. 3) is part of the sequence of the human DC1 heavy chain. By using the DR-H genomic clone as a hybridization probe, a cDNA clone encoding the sequence of the heavy chain of this second HLA-DR locus has been identified (12). In addition, the DC1 heavy chain has been isolated and its amino terminus has been sequenced (R. Bono, personal communication). This chain is the human equivalent of the murine I-A α chain. The sequence of the $\alpha 2$ domain of the DC1 heavy chain (amino acids 88–181) has a 65% amino acid sequence homology to the DR heavy chain in this domain, with considerable blocks of identity. Those features that are indicative of strong homology to Ig have been conserved. These include (i) the invariant cysteines and tryptophan and (ii) strong homology around the cysteines.

The relatedness of the Ig-homologous domains in the class I and class II transplantation antigens has also been analyzed at the nucleic acid level. The nucleotide sequences of two blocks around the first and second cysteine of the Ig-homologous domains of HLA-DR-L (6), HLA-A, -B (30), and β_2m (25) were compared to the DR-H sequence (first Cys, amino acids 103–121; second Cys, amino acids 158–178). No gaps are needed. There is considerable nucleic acid homology between many of the sequences—i.e., 58% between DR-H and DR-L in the first Cys block; 56% between DR-H and DR-L in the second Cys block; 56% between DR-H and HLA-B7 in the second Cys block. In addition, many positions are conserved among all four sequences (16 of 69 in the first block; 19 of 63 in the second block). Moreover, there is considerable conservation of third base-coding assignments in these segments, frequently even when the amino acid is not conserved (11 of 23 third base identities between DR-L and DR-H in the first Cys block; 11 of 21 between DR-L and DR-H in the second Cys block; 12 of 21 between HLA-A, -B and DR-H in the second Cys block). These observations are consistent with divergent evolution from a common ancestral gene for these conserved domains of MHC antigens and Ig.

The Connecting Peptide–Transmembrane–Intracytoplasmic Exon. A single exon encodes amino acids leading from $\alpha 2$ to the membrane (the connecting peptide), the transmembrane region, and intracytoplasmic domain, in addition to a stop codon and 11 bp of the 3' untranslated region. The 3' and 5' boundaries of this exon were determined by comparison to the cDNA clones, pDRH-2, and pDRH-3, respectively. The exon begins at amino acid 179 with a stretch of 13 amino acids rich in proline and glutamic acid. High concentrations of papain cleave the heavy chain in native HLA-DR at a site in the extracellular portion of the molecule (31). This cleavage site is presumably within this sequence.

Following this connecting peptide is a sequence of 23 hydrophobic amino acids (amino acids 192–214) that traverse the membrane including a free cysteine residue. B lymphoblastoid cells metabolically labeled with [3H]palmitate incorporate label into the DR heavy chain at a site within the transmembrane region. This cysteine residue is a potential site of this posttranslational modification (J. Kaufman and M. Krangel, personal communication). A comparison of this transmembrane segment to similar regions in HLA-B7 and H-2K, -D reveals no homology other than the conservation of hydrophobicity. However, the DC1 heavy chain (12) and the DR heavy chain have an unusually high degree of homology in this region (74%), suggesting a requirement for some conserved quaternary interaction with other proteins in the membrane.

Following the transmembrane region is the COOH-terminal intracytoplasmic segment of 15 amino acids (amino acids 215–229). It contains numerous positively charged residues and a potential site of phosphorylation at the single serine residue. The primary papain cleavage site is after this serine (31). Another cDNA clone (pDRH-5) from the JY library contains a single base substitution (T $\rightarrow$ G) converting a Leu [which is present in the genomic sequence and also in pDRH-2 (made from LB mRNA)] to a Val at position 217 (see Fig. 1). This change preserves the character of the amino acid at this position.

The 3' Untranslated Region. Approximately 750 bp from the previous exon is a single exon that encodes the majority (384 bp) of the 3' untranslated region. A sequence of 4 bp (G-T-G-A) that includes a second inframe stop codon is present at the beginning of this exon as well as at the end of the previous exon. There is only one potential splice that produces the sequence found in the cDNA clones, and this places this second stop codon within this exon. For most genes analyzed the final stop codon and 3' untranslated region are present in the same exon. This is not the case for the DR heavy chain gene. However, because the first stop codon is in the same reading frame as the termination codon in this exon, a single base change may have created a truncated protein that resulted in this division of the 3' untranslated region.

There are a number of nucleotide differences between the genomic DNA and cDNA clone pDRH-2 in the 3' untranslated region. One of these differences changes a *Sau*3A [12 bp from the beginning of a poly(A) tail] to a *Bgl* II site found in pDRH-2. This would result in a readily detectable restriction fragment-length polymorphism. There are also two polyadenylation signals (A-A-T-A-A) in this exon separated by ≈ 100 bp.

Glycosylation Sites. The mature heavy chain has two carbohydrate side chains—one susceptible to cleavage by endoglycosidase D and the other to endoglycosidase H (D. A. Shackelford, personal communication). The deduced amino acid sequence contains two carbohydrate addition sites (Asn-X-Thr/Ser) at amino acid 78 in $\alpha 1$ and at amino acid 118 in $\alpha 2$, 11 residues from the first cysteine. If one accepts a similar folding pattern for DR-H $\alpha 2$ as for Ig C domains, then this carbohydrate moiety would be present on a β -strand (fy1) on the three-strand

FIG. 4. The structure of an HLA-DR antigen (taken from ref. 1). The molecule consists of two chains that traverse the lipid bilayer and have short hydrophilic intracytoplasmic COOH-termini. The extracellular portion consists of four domains as shown. $\Leftarrow$ and $\rightarrow$ refer to endoglycosidase D- and endoglycosidase H-susceptible N-linked carbohydrate moieties—two in the heavy chain and one in the light chain. The respective positions of the two types of carbohydrates in the heavy chain are not known. P, C, and T refer to positions of proteolysis by papain, chymotrypsin, and trypsin, respectively, in the native molecule.

face of the domain (situated between two valines presumably pointing inward). This suggests that this face is exposed towards the solvent and that the four-strand face is interacting with the light chain.

The Structure of a DR Antigen. A general model for the structure of a DR antigen or class II antigen is presented in Fig. 4. Recent structural work and amino acid sequence for DR-L chain show that its extracellular portion is composed of two domains of roughly 93 amino acids. Each domain contains a similar sized disulfide loop. The NH_2 -terminal domain is polymorphic, whereas the COOH-terminal domain (as discussed earlier) is homologous to Ig and DR-H as well as HLA-B7 ($\alpha 3$) and $\beta 2m$. A DR antigen would consist of four extracellular domains. Two of these domains are homologous to C regions [DR-H ($\alpha 2$) and DR-L ($\beta 2$)] and it is presumed that these two domains interact in a manner analogous to C regions in Fc fragments and to C_H and C_L in Fab. The disposition of the carbohydrate side chain on the DR-H $\alpha 2$ domain and the location of the protease-sensitive sites in the DR-L chain (ref. 4; J. Kaufman, personal communication) are consistent with this mode of interaction. Upon this conserved base would lie two additional domains—a polymorphic domain of the light chain ($\beta 1$) and a fourth conserved domain in the heavy chain ($\alpha 1$). A similar structure for the HLA-DR antigen has been deduced by Larhammar *et al.* from the analysis of cDNA clones for the heavy and light chains (6, 32).

The class I transplantation antigens, HLA-A, -B and H-2K, -D, unlike the class II antigens, HLA-DR and Ia, are present on all cells. They are recognized in conjunction with viral antigens (and as modified self-antigens) by MHC-restricted T killer cells. These antigens are also composed of four extracellular domains: two Ig-homologous domains in HLA-B7 ($\alpha 3$) and $\beta 2m$ and two polymorphic domains ($\alpha 1$ and $\alpha 2$) in the heavy chain. Preliminary crystallographic evidence (unpublished results) suggests that the HLA-A antigen has a 2-fold axis of symmetry and its crystals have noteworthy similarities to some crystals of Fc and Fab fragments. Studies of the effect of chemical modification of $\beta 2m$ on its exchange into the HLA-A, -B antigen (unpublished results) suggest that the four-strand face of $\beta 2m$ and presumably $\alpha 3$ interact in the manner of C regions. These features of sequence homology and domain structure between the class I and class II antigens suggest a similarity in their general overall structure. It is probable that they function in a similar mechanistic fashion and are perhaps recognized in a similar fashion by receptors for MHC molecules. The question of the

significance of these domains in structural (perhaps as rudimentary structural units), in functional (such as the effector functions served by C regions), or in evolutionary terms remains.

We thank J. Lillie for enthusiasm and technical assistance, J. Kaufman for an idea, C. Lehn, M. Krangel, P. Bjorkman, D. Goldberg, W. Herr, A. Kamb, and members of the laboratories of J.L.S. and T. Maniatis for help and discussion. This work was supported by research grants from National Institutes of Health (AM 30241) and the Kroc Foundation.

- Shackelford, D. A., Kaufman, J., Korman, A. J. & Strominger, J. L. (1982) *Immunol. Rev.* **66**, 129–183.
- Benacerraf, B. (1981) *Science* **212**, 1229–1238.
- Nagy, Z. A., Baxevanis, C. N., Ishii, N. & Klein, J. (1981) *Immunol. Rev.* **60**, 59–83.
- Kaufman, J. & Strominger, J. L. (1982) *Nature (London)* **297**, 694–697.
- Kratzin, H., Yang, C. Y., Gotz, H., Pauley, E., Kolbel, S., Egert, G., Thinnies, F. P., Wernet, P., Alterogt, P. & Hilschmann, N. (1981) *Hoppe-Seyler's Z. Physiol. Biochem.* **362**, 1665–1669.
- Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. & Peterson, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3687–3691.
- Orr, H. T., Lopez de Castro, J., Lancet, D. & Strominger, J. L. (1979) *Biochemistry* **18**, 5711–5719.
- Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. & Nathenson, S. G. (1981) *Nature (London)* **291**, 35–39.
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
- Korman, A. J., Knudsen, P. J., Kaufman, J. F. & Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1844–1848.
- Rutter, U., Koenen, M., Otto, K. & Muller-Hill, B. (1981) *Nucleic Acids Res.* **9**, 4087–4097.
- Auffray, C., Korman, A. J., Roux-Dosseto, M., Bono, R. & Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, in press.
- Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- O'Farrell, P. N., Kutter, E. & Nakanishi, M. (1980) *Mol. Gen. Genet.* **179**, 421–435.
- Berk, A. & Sharp, P. (1977) *Cell* **12**, 721–732.
- Dayhoff, M. O. (1975) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3.
- Lee, J. S., Trowsdale, J. & Bodmer, W. F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 545–549.
- Korman, A. J., Ploegh, H. L., Kaufman, J. F., Owen, M. J. & Strominger, J. L. (1980) *J. Exp. Med.* **152**, 65s–82s.
- Springer, T. A., Kaufman, J. F., Terhorst, C. & Strominger, J. L. (1977) *Nature (London)* **268**, 213–216.
- Liu, V. Y. S., Low, T. L. K., Infante, A. & Putnam, F. W. (1976) *Science* **193**, 1017–1020.
- Putnam, F. W., Florent, G., Paul, C., Shinoda, T. & Shimizu, A. (1973) *Science* **182**, 287–291.
- Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Putishauser, U. & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 78–85.
- Langer, B., Steinmetz-Kayne, M. & Hilschmann, N. (1968) *Z. Physiol. Chem.* **349**, 945–951.
- Titani, K., Shinoda, T. & Putnam, F. W. (1969) *J. Biol. Chem.* **244**, 3550–3560.
- Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
- Williams, A. F. & Gagnon, J. (1982) *Science* **214**, 696–703.
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
- Beale, D. & Feinstein, A. (1976) *Q. Rev. Biophys.* **9**, 135–180.
- Anzel, L. M. & Poljak, R. (1979) *Annu. Rev. Biochem.* **48**, 961–997.
- Malissen, M., Malissen, B. & Jordan, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 893–897.
- Kaufman, J. F. & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6304–6308.
- 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.