The murine Ia α chains, E $_{\alpha}$ and A $_{\alpha}$, show a surprising degree of sequence homology

(Ia antigens/major histocompatibility complex/immune-response genes/recombinant DNA/DNA sequence analysis)

CHRISTOPHE O. BENOIST, DIANE J. MATHIS, MADGE R. KANTER, VIRGINIA E. WILLIAMS II, AND
HUGH O. MCDEVITT

Department of Medical Microbiology, Stanford University Medical School, Stanford, California 94305

Contributed by Hugh O. McDevitt, October 29, 1982

ABSTRACT The I region of the murine major histocompatibility complex codes for a group of glycoproteins, the Ia antigens, thought to be involved in the control of immune responsiveness. Each Ia antigen complex contains a "heavy chain," a "light chain," and the "invariant chain." We describe here the isolation and characterization of genomic and cDNA clones for one of the heavy chains, A $_{\alpha}$. The complete nucleotide sequence of the cDNA clone is presented, and the predicted amino acid sequence is compared with that of another α chain, E $_{\alpha}$. About 50% of the amino acids are identical, a finding somewhat unexpected on the basis of preliminary protein sequence data.

The mouse major histocompatibility complex (MHC) codes for a variety of cell-surface antigens thought to be important in cell-cell recognition, especially in an immunological context (1, 2). Among these cell-surface molecules are the Ia antigens, glycoproteins that map to the I region of the MHC and that appear to control the ability of an animal to mount an effective response to certain foreign antigens (1, 2). Two different Ia antigen complexes have been described thus far: the A complex is composed of two subunits, A $_{\alpha}$ (M_r , 34,000) and A $_{\beta}$ (M_r , 29,000), which map to the I-A subregion; the E complex consists of E $_{\alpha}$ (M_r , 34,000), which maps to the I-E subregion, and E $_{\beta}$ (M_r , 29,000), which maps to I-A. The invariant chain, Ii, is found in both A and E complexes, but its structural gene does not map to the MHC (1, 2). As of yet, structural information about the Ia antigens remains scant; in particular, it has proven very difficult to obtain protein sequence data. Therefore, we have exploited the techniques of molecular genetics in the hope of determining more rapidly the structural organization of these immunologically important cell-surface molecules.

We describe here the isolation and identification of a genomic and cDNA clone for the murine A $_{\alpha}$ gene. The complete nucleotide sequence of the cDNA clone is presented, and the predicted amino acid sequence of the A $_{\alpha}$ protein is compared with that of the E $_{\alpha}$ gene product (unpublished data). This analysis reveals a surprising similarity between the two gene products. (The A $_{\alpha}$ and E $_{\alpha}$ genes were cloned from "a" haplotype mice. Because mice of the a haplotype are actually "k" haplotype in the I region, we refer to the genes as A $_{\alpha}^k$ and E $_{\alpha}^k$ throughout the paper.)

MATERIALS AND METHODS

Mice. B10.AKM mice were purchased from The Jackson Laboratories. C. S. David of the Mayo Clinic, Rochester, MN, provided the B10.MBR mice. All other mouse strains were produced in our breeding colony at Stanford.

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.

Genomic Cloning. An A/J thymus DNA genomic library was produced in Charon 4 λ phage by described methods (3). Briefly, high molecular weight A/J thymus DNA was partially digested with *EcoRI*, the digest was size-fractionated on a sucrose gradient, a 20-kilobase (kb) fraction was isolated and ligated to the purified *EcoRI*-cut arms of Charon 4, and the ligation mix was packaged and used to infect *Escherichia coli* C600. One million clones were screened with a ³²P-labeled cDNA clone for the human p34 gene (4). Replicate filters were probed at 44, 39, 30, and 23°C in 50% formamide/0.75 M NaCl/0.075 M Na citrate, pH 7, and washed at 23°C in 0.30 M NaCl/0.030 M Na citrate, pH 7 (5). Candidate clones were plaque-purified and analyzed by restriction enzyme mapping and Southern blotting (6).

In order to identify a low-copy-number probe for the putative A $_{\alpha}$ clone, λ A/Jp34.170, we digested the clone with various restriction enzymes, electrophoresed the digests, transferred them to nitrocellulose, and probed the filter with nick-translated total A/J liver DNA. A 2.85-kb *Pst* I fragment did not hybridize with this probe but did hybridize with the p34 cDNA clone. We subcloned the fragment into the *Pst* I site of pBR322 and called the resulting clone pPAA1 (see Fig. 1 a and b).

cDNA Cloning. Poly(A)⁺ RNA from spleens of B10.A mice was used as a template for the synthesis of double-stranded (ds) cDNA with reverse transcriptase and DNA polymerase I. The hairpin loop was cleaved with S1 nuclease, and the ds cDNA was ligated to a 50-fold molar excess of *EcoRI* linkers. After digestion with *EcoRI*, the largest ds cDNA molecules were isolated on a Bio-Gel A-5m column and ligated into the *EcoRI* site of the λ phage λ GT10 (T. Huynh and R. Davis, personal communication). After *in vitro* packaging, the library was amplified and screened for possible A $_{\alpha}$ clones with nick-translated pPAA1. Several clones were isolated, and the cDNA insert was subcloned into pBR322. The longest of these, pAAC6, was chosen for further study.

Isolation of Genomic DNA and Southern Blotting. Successful isolation of liver DNA required that the donor mouse be deprived of food for 12–18 hr just prior to sacrifice. Subsequently, the liver was removed, cleaned of the gall bladder, rinsed in phosphate-buffered saline, and frozen in liquid nitrogen. The frozen liver was crushed to a fine powder with a mortar and pestle, and DNA was isolated by described methods (7) entailing NaDodSO₄/proteinase K digestion and phenol/chloroform extraction.

For Southern blots, 5–10 μ g of liver DNA was digested with the appropriate restriction enzyme, the digest was electrophoresed on an agarose gel and transferred to nitrocellulose paper,

Abbreviations: MHC, major histocompatibility complex; kb, kilobase(s); ds, double-stranded.

FIG. 1. Restriction enzyme maps of the A_α clones. All maps are in the same orientation with respect to the 5' to 3' direction of transcription. (a) Map of the genomic DNA clone $\lambda A/Jp34.170$. (b) Map of the insert in the plasmid pPAA1, a subclone of $\lambda A/Jp34.170$ (dashed lines). The insert (a *Pst* I fragment) contains no repetitive elements and all of the sequences hybridizing to the human *p34* cDNA clone (see text). (c) Map of the cDNA insert in the plasmid pAAC6. The protein-coding sequences are delineated by a thick line, the 3' untranslated region by a thinner line. Broken lines join the *BstEII* and *EcoRI* sites found at an identical distance from each other in pPAA1 and pAAC6 (120 bp). Arrows under the map summarize the strategy used for the sequence determination of pAAC6. Solid arrows, Sanger and Coulson sequence assay; broken arrows, Maxam and Gilbert sequence assay.

and the filter was hybridized with an A_α probe (6).

DNA Sequence Determination. Detailed descriptions of the methods of sequence analysis have been published (8–10). In short, our strategy was to purify the cDNA insert of pAAC6, self-ligate the inserted fragment, shear the resulting concatamers to a size of about 200–1,000 base pairs, clone these random fragments into M13mp8, screen the resulting clones with the original cDNA insert, and determine the sequence of selected clones by the dideoxy technique.

Occasionally, this strategy resulted in “holes” in the sequence. In such cases, we filled in the gaps by isolating specific restriction enzyme fragments and determining their sequence by the technique of Maxam and Gilbert (10). The sequence of essentially all of the cDNA insert was determined either on both strands or by both techniques on the same strand (see Fig. 1d).

RESULTS

Isolation of a Putative A_α^k Genomic Clone. An A/J thymus genomic library was probed at various stringencies with a 32 P-labeled cDNA clone for the human *p34* gene. Two classes of positive clones were obtained. One type of recombinant produced strong signals even at the highest stringency of hybridization (44°C). This class, exemplified by $\lambda A/Jp34.13$, had two *EcoRI* fragments, 8.3 and 2.9 kb in length, which hybridized to the human probe. $\lambda A/Jp34.13$ contained a gene eventually identified as E_α^k (unpublished data). The other type of recombinant gave relatively weak signals at the highest-stringency

Table 1. MHC haplotypes of the parental and recombinant mouse strains

Strains	Haplotypes of MHC regions						
	K	I					D
		A	B	J	E	C	
B10	b	b	b	b	b	b	b
B10.A(1R)	k	k	k	k	k	d	b
B10.A(3R)	b	b	b	b	k	d	d
B10.A(5R)	b	b	b	k	k	d	d
B10.A(4R)	k	k	b	b	b	b	b
B10.A	k	k	k	k	k	d	d
B10.AKM	k	k	k	k	k	k	q
B10.MBR	b	k	k	k	k	k	q

Each of the strains bears a different MHC haplotype on the B10 background. The haplotypes (11) of the K and D regions and of the subregions of I are listed for each strain.

FIG. 2. Mapping of pPAA1 sequences to the I-A subregion. Genomic DNA was isolated from livers of the mouse strains indicated. After digestion with *Pst* I, the DNAs were analyzed on Southern blots by using nick-translated pPAA1 as a probe. 1R, 3R, 5R, and 4R, strains B10.A(1R), B10.A(3R), B10.A(5R), and B10.A(4R), respectively.

Pst I 50 100
 5'...AATTCCCGGGAGGTCTCTGCAGCAGAGCTCTGATTCTGGGGTCTCGCCCTGACCACCATGCTCAGCCTCTGCGGAGGTGAAGACGACATTGAGGCCG
 150 200
 ACCACGTAGGCTCCTATGGTATAACTGTATATCAGTCTCCTGGAGACATTGGCCAGTACACATTGAATTTGATGGTGATGAGTGTCTATGTGGACTT
 250 300
 GGATAAGAAGGAGACTGTCTGGATGCTTCTGAGTTTGCTCAACTGAGAAGATTGAGCCCCAAGGTGGACTGCAAAACATAGCTACAGGAAAAACAAC
 350 400
 TTGGAAATCTTGACTAAGAGGTCAAATTCACCCAGCTACCAATGAGGCTCCTCAAGCGACTGTGTTCCCAAGTCTCTGTGCTGCTGGGTGAGCCCA
Sau3A I 450 500
 ACACCTTATCTGCTTTGTGGACAACATCTTCCCTCCTGTGATCAACATCACATGGCTCAGAAATAGCAAGTCAGTCACAGACGGCGTTTATGAAACCGA
Hinc II 550 600
 CTTCTTCGTCAACCGTGAATTCCTTCCACAAGCTGTCTTATCTCACCTTCATCCCTTCTGACGATGACATTTATGACTGCAAGGTGGAGCACTGGGGC
 650 700
 CTGGAGGAGCCGTTCTGAAACACTGGGAACCTGAGATTCCAGCCCCATGTGAGAGCTGACAGAGACTGTGGTGTGTGCCCTGGGGTGTCTGTGGGCC
Sau3A I 750 800
 TTGTGGGCATCGTGGTGGGCACCATCTTCATCATTCAAGGCTGCGATCAGGTGGCACCTCCAGACACCCAGGGCCTTTATGAGTACACCCCTGGAAAGG
Bst E II 850 900
 AAGGCGTGTGCCCTCTTCATGGAAGAAGTGGTGTGCTGGTGACCTGGCACAGTGTGTTTTCTGGACCAATTTATGGTGTCTTTTCTTCTTCAAGT
 950
 ACCCCAACTTGCTTTTCCCTTGGCCCTGAGGCTGTCCCTCTCACAGCTCACACCCCTTGAATTC...3'

FIG. 3. Nucleotide sequence of the cDNA insert in pAAC6. The sense strand of the sequence is shown. The *EcoRI* linker present at the 5' end is indicated by dots. The sequence coding for the leader peptide is delineated by a dashed line and continued with solid lines that underline the sequences coding for the mature protein up to the termination codon (boxed). Some of the restriction enzyme sites presented in Fig. 1c are bracketed.

screening, although the intensity of the signal increased somewhat at lower stringencies (37 and 30°C). In this class, which included λ /Jp34.170, sequences homologous to human *p34* were contained in a 10.9-kb *EcoRI* fragment (Fig. 1a).

As an initial step in identifying λ /Jp34.170, we sought to map its sequences within the *I* region of the MHC, using restriction enzyme fragment polymorphism amongst recombinant mouse strains. We first identified and then subcloned into pBR322 a 2.85-kb *Pst I* fragment that did not contain repetitive sequences (see Fig. 1a and b). DNA from this clone, pPAA1, was nick-translated and used to probe restriction enzyme-digested DNA from several recombinant mouse strains. Preliminary experiments (data not shown) indicated that the enzyme *Pst I* revealed a restriction enzyme polymorphism between mice of *b* and *k* haplotypes. That is, pPAA1 hybridized to a 2.85-kb band in DNA digests from mice of *k* haplotype, but to a band of only 2.7 kb in digests from mice of *b* haplotype.

This polymorphism could be exploited to map the pPAA1 sequences to the *I-A* subregion. Table 1 shows the *H-2* haplotypes of the relevant recombinant strains, whereas Fig. 2 shows the restriction enzyme fragment pattern revealed by probing DNA from these recombinant strains with pPAA1. B10.A and

B10 are parental strains for the recombinants B10.A(1R), B10.A(3R), B10.A(5R), and B10.A(4R). Strain B10.A(1R) mice exhibit the *k*-haplotype restriction enzyme fragment pattern (i.e., a band of 2.85 kb) and, thus, the pPAA1 sequences reside to the left of *H-2D*. B10.A(3R) and B10.A(5R) mice have the 2.7-kb band characteristic of the *b* haplotype, placing the pPAA1 sequences to the left of *I-J*. The *k*-haplotype pattern shown by B10.A(4R) mice further localizes the pPAA1 sequences to the left of *I-B*. B10.AKM and B10 mice are parental strains for the recombinant strain B10.MBR. Because B10.MBR mice have the *k*-haplotype 2.85-kb band, the pPAA1 sequences must map to the right of *H-2K*. Therefore, these sequences appear to be localized within the *I-A* subregion, and our clone could code for A_α , A_β , or E_β or for some unknown entity. We hypothesized that it coded for A_α on the basis of its homology, revealed by hybridization, to the human *p34* gene.

Isolation of a Putative A_α^k cDNA Clone. To facilitate the structural analysis of our putative A_α^k genomic clone, we cloned the corresponding cDNA sequences. A B10.A spleen cDNA bank was screened with the pPAA1 probe, and a putative A_α^k cDNA clone was identified and subcloned (pAAC6). A restriction enzyme map for this clone is presented in Fig. 1c. Clone

A_α^k Protein Sequence NH₂. * - * - * -Ile- * -Ala- * - * - * - * -Tyr- * - *
 pAAC 6 Sequence NH₂.Glu-Asp-Asp-Ile-Glu-Ala-Asp-His-Val-Gly-Ser-Tyr-Gly-Ile-
 * -Val-Tyr- * - * - * - * - * - * -Tyr-.....
 Thr-Val-Tyr-Gln-Ser-Pro-Gly-Asp-Ile-Gly-Gln-Tyr-.....

FIG. 4. Comparison of the known amino acid residues in the A_α^k protein (11) (top line; stars indicate unknown residues) with the sequence deduced from the nucleotide sequence of pAAC6 (bottom line).

pAAC6 was mapped to the *I-A* subregion by the same procedure used for pPAAl (data not shown).

Identification of pAAC6 as A_α^k . In order to unambiguously identify our clones as bearing the A_α^k gene, we sequenced the cDNA insert of clone pAAC6 in its entirety. The strategy for sequence analysis is shown in Fig. 1c, and the nucleotide sequence of the pAAC6 insert is shown in Fig. 3. We searched the predicted amino acid sequence for a stretch corresponding to a region of the A_α chain that has been subjected to very limited protein sequence analysis (12). The nucleotide sequence of clone pAAC6 does predict a stretch of amino acids identical to that found in the A_α protein (Fig. 4). On this basis, and supported by our mapping data, we conclude that pAAC6 codes for the A_α chain. We infer also that the genomic clone $\lambda A/Jp34.170$ includes the A_α^k gene.

Although the initiation codon is not present in pAAC6, we believe, by analogy with the E_α gene (unpublished data), that most of the sequences coding for the leader peptide of A_α are represented. pAAC6 also lacks the extreme 3' end of the messenger RNA because an *EcoRI* site is present in the 3' untranslated region. (The ds cDNA was cloned by ligation to *EcoRI*

linkers, followed by cleavage with *EcoRI* without prior treatment with *EcoRI* methylase.)

Comparison of Amino Acid Sequences of A_α^k and E_α^k Chains. We have recently isolated and determined the sequence of genomic and cDNA clones for the E_α^k gene (unpublished data). These data allowed a delineation of the intron-exon organization of the E_α gene and, thus, a somewhat clearer understanding of the organization of protein domains. In Fig. 5, we compare the predicted amino acid sequence for the A_α^k gene product with that for the E_α^k gene product.

About 50% of the amino acid residues are identical. In general, there are not large clusters of homology but rather invariant residues interspersed with stretches of nonhomology. The most dissimilar regions are the cytoplasmic tails (amino acid residues 215–229 in E_α) and the leader peptides. Also quite divergent are the amino termini of the mature proteins (residues 1–23) and the short stretches (residues 179–192) that link the second external domains to the transmembrane regions. Residues 24–84 of the first domains are better conserved, as are the second domains (55% and 60% identity, respectively). The greatest similarity is in the transmembrane regions (residues 192–214 in E_α) where 18 of 23 amino acids are identical.

Aα	 Pro Gly Gly Leu Cys Ser Arg Ala Leu Ile Leu Gly Val Leu Ala Leu Thr Thr Met Leu Ser Leu Cys Gly Gly	] LP
Eα	NH ₂ -Met Ala Thr Ile Gly Ala Leu Val Leu Arg Phe Phe Phe Ile Ala Val Leu Met Ser Ser Gln Lys Ser Trp Ala	
	Glu Asp Asp Ile Glu Ala Asp His Val Gly Ser Tyr Gly Ile Thr Val Tyr Gln Ser Pro Gly Asp Ile Gly Gln Tyr	] D1
	Ile Lys Glu Glu His Thr Ile Ile Gln Ala Glu Phe Tyr Leu Leu Pro Asp Lys Arg Gly Glu Phe	
	1 10	
	Thr Phe Glu Phe Asp Gly Asp Glu Leu Phe Tyr Val Asp Leu Asp Lys Lys Glu Thr Val Trp Met Leu Pro Glu Phe	
	Met Phe Asp Phe Asp Gly Asp Glu Ile Phe His Val Asp Ile Glu Lys Ser Glu Thr Ile Trp Arg Leu Glu Glu Phe	
	23 36	] D2
	Ala Gln Leu Arg Arg Phe Glu Pro Gln Gly Gly Leu Gln Asn Ile Ala Thr Gly Lys His Asn Leu Glu Ile Leu Thr	
	Ala Lys Phe Ala Ser Phe Glu Ala Gln Gly Ala Leu Ala Asn Ile Ala Val Asp Lys Ala Asn Leu Asp Val Met Lys	
	49 62	
	Lys Arg Ser Asn Ser Thr Pro Ala Thr Asn	] TC
	Glu Arg Ser Asn Asn Thr Pro Asp Ala Asn	
	75 84	
	Glu Ala Pro Gln Ala Thr Val Phe Pro Lys Ser Pro Val Leu Leu Gly Gln Pro Asn Thr Leu Ile Cys Phe Val Asp	
	Val Ala Pro Glu Val Thr Val Leu Ser Arg Ser Pro Val Asn Leu Gly Glu Pro Asn Ile Leu Ile Cys Phe Ile Asp	
	85 98	] D2
	Asn Ile Phe Pro Pro Val Ile Asn Ile Thr Trp Leu Arg Asn Ser Lys Ser Val Thr Asp Gly Val Tyr Glu Thr Ser	
	Lys Phe Ser Pro Pro Val Val Asn Val Thr Trp Leu Arg Asn Gly Arg Pro Val Thr Glu Gly Val Ser Glu Thr Val	
	111 124	
	Phe Phe Val Asn Arg Asp Tyr Ser Phe His Lys Leu Ser Tyr Leu Thr Phe Ile Pro Ser Asp Asp Asp Ile Tyr Asp	] TC
	Phe Leu Pro Arg Asp Asp His Leu Phe Arg Lys Phe His Tyr Leu Thr Phe Leu Pro Ser Thr Asp Asp Phe Tyr Asp	
	137 150	
	* Cys Lys Val Glu His Trp Gly Leu Glu Glu Pro Val Leu Lys His Trp	
	Cys Glu Val Asp His Trp Gly Leu Glu Glu Pro Leu Arg Lys His Trp	
	163 176	] TC
	Glu Pro Glu Ile Pro Ala Pro Met Ser Glu Leu Thr Glu Thr Val Val Cys Ala Leu Gly Leu Ser Val Gly Leu Val	
	Glu Phe Glu Glu Lys Thr Leu Leu Pro Glu Thr Lys Glu Asn Val Val Cys Ala Leu Gly Leu Phe Val Gly Leu Val	
	179 192	
	Gly Ile Val Val Gly Thr Ile Phe Ile Ile Gln Gly Leu Arg Ser Gly Gly Thr Ser Arg His Pro Gly Pro Leu-COOH	] TC
	Gly Ile Val Val Gly Ile Ile Leu Ile Met Lys Gly Ile Lys Lys Arg Asn Val Glu Arg Arg Gln Gly Ala Leu-COOH	
	205 218 230	

FIG. 5. Comparison of the A_α and E_α protein sequences. The A_α protein sequence (top line throughout) was deduced from the nucleotide sequence (Fig. 3), whereas the E_α sequence (bottom lines) was obtained similarly from cDNA and genomic sequences (unpublished data). The sequences have been divided into four blocks; the leader peptide (LP), the first external domain (D1), the second external domain (D2), and the transmembrane and cytoplasmic regions (TC). Separation into these blocks is based on the known intron-exon structure of the E_α gene (unpublished data), except for the division between LP and D1 based on the known NH₂-terminal protein sequences (11). The numbering follows the E_α sequence. Identical residues are shown by a line between the sequences. The glycosylation sites are boxed, and the two cysteine residues of the second domain are starred.

Certain E_α amino acid residues are of identifiable structural importance. For example, residues 78–80 and 118–120 probably serve as sites for carbohydrate attachment, a process requiring the sequence Asn-X-Ser/Thr (13). Likewise, the cysteine at position 107 probably forms an intrachain disulfide bridge with the cysteine at position 163 if, as suspected (14), this region of the E_α chain forms an immunoglobulin-like fold. Not surprisingly, both the carbohydrate attachment sites and the disulfide bridge sites are conserved in the A_α chain. Also conserved are the proline at position 114, tryptophan at position 121, valine at position 165, and histidine at position 167. Larhammer *et al.* (14) have pointed out that identical residues in the human p34 protein sequence align with corresponding invariant residues in immunoglobulin constant domains and, thus, may indicate that the second external protein domain is folded like an immunoglobulin domain.

DISCUSSION

We have described the isolation and characterization of genomic and cDNA clones for the murine A_α gene. This represents the first reported isolation of an A_α gene. Our initial identification of the clones was based on homology with E_α, as revealed by hybridization, and on chromosome mapping through restriction enzyme polymorphism amongst recombinant mouse strains. Comparison of amino acid sequence data derived from the nucleotide sequence of the cDNA clone with limited published protein sequence data confirmed our identification of the clones as corresponding to the A_α gene.

It seems reasonable to hypothesize that A_α and E_α chains have analogous functions because each participates as the "heavy chain" in an Ia complex expressed on the cell surface (1, 2). Yet, limited protein sequence data, encompassing amino acid residues 1–27, revealed no homology between the two M_r 34,000 proteins (12). Furthermore, in strains bearing haplotypes *b* and *s*, which express E_β but not E_α chains, there is no detectable association between the E_β and A_α chains (15). It also has proven difficult until recently (16) to isolate antibodies that are crossreactive with the A and E complexes. Thus, our finding that the A_α and E_α chains share extensive homology comes as somewhat of a surprise and serves to caution against predicting degrees of protein homology on the basis of such limited data. With this in mind, we have reexamined the question of which mouse α chain is really the homologue of the human p34 gene product by comparing our A_α and E_α protein sequences with the human p34 protein sequence recently described by Larhammer *et al.* (14) and by S. M. Weissman and colleagues (personal communication). The E_α protein was identical to p34 in 76% of its residues, while A_α shared only 57%. Hence, E_α appears to be the homologue of p34 protein, confirming a conclusion based on limited protein sequence information (17).

Because we already knew the intron-exon organization of the E_α gene (unpublished data), we were able to analyze E_α and A_α homologies in the context of domain organization. The domain of greatest sequence conservation was the transmembrane region, where 78% of the amino acid residues were identical. Such pronounced homology could signify that the α chains interact in this region, either with the β chains, or with Ii, a nonpolymorphic protein thought to interact with A and E complexes within the membrane (1, 2). Aside from the first 23 amino acids, quite divergent, an intermediate degree of sequence conservation was characteristic of the extracellular domains. The remainder of the first extracellular domains exhibited 55% identity, while the second showed 60%. Perhaps the greater overall sequence constraint imposed on the second extracellular do-

main is related to their similarity in structure to immunoglobulin constant regions (14). Not surprisingly, the leader peptides exhibited no conservation of sequence. More unexpected, however, was the lack of sequence homology within the cytoplasmic tails, especially in light of the fact that the corresponding region is conserved amongst different class I molecules (18–20). Perhaps the lack of homology in the intracellular domains of the A_α and E_α chains signifies that these proteins interact with different molecules in the cytoplasm. Especially interesting in this respect is that E_α lacks a serine in this stretch of sequence, whereas A_α has two serine residues that might serve as phosphorylation sites. Alternatively, the sequence divergence within the cytoplasmic tails could merely reflect the fact that there is no functional constraint on this particular amino acid sequence. However, this explanation is rendered unlikely by the fact that the amino acid sequence of the cytoplasmic tails is about 60% conserved between the E_α and human p34 proteins (unpublished data).

In short, the A_α and E_α chains are quite similar proteins. It remains to determine what advantage a cell derives from having a dual Ia antigen system, as well as to assess what functional role (if any) the differences in structure between A_α and E_α play.

We acknowledge, gratefully, the gift of a p34 cDNA clone by Dr. Henry Erlich, Cetus Corporation, and Dr. Sherman Weissman, Yale University. We thank the Stanford SUMEX staff and the Molgen group for the use of the computer and programs. D.M. and C.B. are recipients of fellowships from The Leukemia Society of America. The laboratory is supported by Grants AI 07757 and AI 18367 to H.O.M.

- McDevitt, H. O. (1981) *J. Immunogenet.* 8, 287–295.
- Klein, J., Juretic, A., Baxevanis, C. N. & Nagy, Z. A. (1981) *Nature (London)* 291, 455–460.
- Maniatis, T., Hardison, R. C., Lacy, E., Lamer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* 15, 687–701.
- Stetler, D., Das, H., Nunberg, J., Saiki, R., Shen-Dong, R., Mullis, K. B., Weissman, S. M. & Erlich, H. A. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5966–5970.
- Benton, W. D. & Davis, R. W. (1977) *Science* 196, 180–182.
- Southern, E. (1975) *J. Mol. Biol.* 98, 503–517.
- Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* 36, 32–38.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* 143, 161–178.
- Fuhrman, S. A., Deininger, P. L., La Porte, P., Friedman, T. & Guiduscheck, P. (1981) *Nucleic Acids Res.* 9, 6349–6456.
- Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499–580.
- Altman, P. & Katz, D., eds. (1979) *Inbred and Genetically Defined Strains of Laboratory Animals Part I*, (Fed. Am. Soc. Exp. Biol., Bethesda, MD), p. 126.
- Cecka, J. M., McMillan, M., Murphy, D. B., McDevitt, H. O. & Hood, L. (1979) *Eur. J. Immunol.* 9, 955–963.
- Wagh, P. V. & Bahl, O. P. (1981) *CRC Crit. Rev. Biochem.* 10, 307–377.
- Larhammer, 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.
- Jones, P., Murphy, D. & McDevitt, H. (1981) *Immunogenetics* 12, 321–337.
- Bhattacharya, A., Dorf, M. & Springer, T. A. (1981) *J. Immunol.* 127, 2488–2495.
- McMillan, M., Cecka, J. M., Murphy, D. B., McDevitt, H. O. & Hood, L. (1978) *Immunogenetics* 6, 137–147.
- Breggeregere, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, G., Wiman, K., Larhammer, D., Peterson, P. A., Gachelin, G., Kourilsky, P. & Dobberstein, B. (1981) *Nature (London)* 292, 78–81.
- Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. A. & Hood, L. (1982) *Science* 215, 679–682.
- Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. & Hood, L. (1981) *Cell* 25, 683–692.