

Sequence analysis and structure–function correlations of murine *q*, *k*, *u*, *s*, and *f* haplotype I-A_β cDNA clones

(major histocompatibility complex/protein structure/sequence homology/immune response genes)

PILA ESTESS*†, ANN B. BEGOVICH‡, MAY KOO*, PATRICIA P. JONES‡, AND HUGH O. MCDEVITT*§

Departments of *Medical Microbiology and †Medicine, Stanford University School of Medicine and ‡Department of Biological Sciences, Stanford University, Stanford, CA 94305

Contributed by Hugh O. McDevitt, January 6, 1986

ABSTRACT I-A_β-chain cDNA clones from mice of the *q*, *k*, *u*, *s*, and *f* haplotypes have been isolated and sequenced. Nucleotide sequence comparisons among these five A_β chains show considerable allelic variation in the region encoding the first external (β1) domain of the mature A_β protein. The β1 domain variability is clustered into three discrete regions, two of which divide the A_β chains into subgroups, suggesting an evolutionary history for the separation of alleles in inbred strains of mice. The amino acid sequences of these five chains are compared to each other and to previously published I-A_β chains. Correlations are made between the primary structural differences and the serologic and immune response characteristics mapping to the I-A subregion.

The intimate relationship between I-region associated (Ia) antigens and immune responsiveness to a multitude of natural and synthetic antigens has been known for some time (1, 2). Recent studies using either anti-Ia antisera or monoclonal antibodies to block immune responses strongly suggest that Ia antigens are the products of immune response (*Ir*) genes (3–5), but how a limited number of Ia molecules differentially regulate responses to a myriad of antigens is yet to be resolved. It seems likely that their role in the presentation of antigen to immunocompetent cells is of fundamental importance and that primary structural alterations in the α and β chains of the class II heterodimer result in a failure to present antigen in a configuration appropriate for recognition. These alterations must also account for the multiple epitopes recognized by both anti-Ia antibodies and alloreactive T cells.

In an effort to address the nature of primary structural differences among class II molecules and how these differences affect immune responses, we have undertaken the cloning and sequencing of cDNAs encoding I-A_β chains from several mouse haplotypes. This should facilitate the eventual assignment of particular serologic epitopes and restriction elements to specific amino acids in these chains. Such analysis is a first step in designing experiments in which immune responsiveness can be manipulated via structural alterations in Ia molecules and then, perhaps, understood.

MATERIALS AND METHODS

Construction and Screening of cDNA Libraries. cDNA libraries from B10.A (*Ia^k*), B10.PL (*Ia^d*), A.TH (*Ia^s*), B10.G (*Ia^q*), and B10.M (*Ia^f*) mice were provided by D. Mathis and C. Benoist and have been described (6, 7). Screening for I-A_β clones was done with a human DQ_β cDNA clone (8). Other I-A_β cDNA clones (*q*, *u*, *s*, and *f*) were detected using either the full-length I-A_β^k clone or a subcloned 245-base-pair (bp) fragment containing 14 bp of 5' untranslated material and

nucleotides encoding the leader peptide and the first 50 amino acids of the mature I-A_β^k protein.

Characterization and Sequencing of β-Chain cDNA Clones. Clones hybridizing to the human DQ_β or the murine A_β^k cDNA probes were tested for the presence of restriction enzyme sites consistent with A_β but not E_β coding sequences (9). The longest such clones were subcloned into the *Eco*RI site of pBR322 and from there into the *Eco*RI site of M13 mp8 for sequencing. Sequencing was performed by the dideoxy-sequencing method of Biggin *et al.* (10). All clones were sequenced using the M13 universal primer (UP) from the *Eco*RI ends, as shown in Fig. 1. In addition, five 18-bp synthetic oligonucleotide primers homologous to conserved regions of the known A_β and DQ_β sequences (8, 11–14) were constructed and used to obtain sequence information on portions of the clones inaccessible from the *Eco*RI ends (Fig. 1, B–H). Circular double-stranded M13 clones containing A_β were also cut with *Bst*EII (in I-A_β) plus *Bam*HI (in M13) to delete part of the clone. These were religated to generate M13 templates containing a ≈230-bp 5' or a 630-bp 3' fragment, depending on the original orientation of the A_β clone in M13. These subclones were sequenced using the M13 universal primer from the *Bst*EII site toward the *Eco*RI ends. A similar approach was also used with a *Bal*I site in the *u* haplotype clone, religated to the *Sma*I site in M13.

RESULTS AND DISCUSSION

I-A_β cDNA Nucleotide Sequences. The nucleotide sequences for the coding regions of the five (*q*, *k*, *u*, *s*, and *f*) A_β cDNA clones isolated are shown in Fig. 2. The sequence of the A_β^k molecule from nucleotide 96 through the 3' untranslated region has been published (11). Nucleotide identity in the β1 domain among these five haplotype cDNAs ranges from 87% to 98%. In contrast, when comparisons are made between the leader peptide, β2, transmembrane, or intracytoplasmic coding regions of these molecules, homology is 95% or greater. While nucleotide differences between molecules are present throughout, there is a definite clustering of substitutions in the region encoding the first external domain (β1) of the mature protein. As with class II α and β chains of both the mouse and human, the majority of the β1 domain substitutions are productive (8, 9, 11–17).

Of particular interest is the region of nucleotides 262–282, which encode amino acids 61–67. Nine bases present in the *q* haplotype β chain are absent in the *k*, *u*, *s*, and *f* haplotype sequences. In these four chains, these nine bases, CCG-GAGATC, are replaced by three others, TAC, producing A_β genes that are 6 bp shorter than that of the *q* haplotype. As indicated in Fig. 3, this alteration, previously detected in the A_β^k gene, was not found in the A_β^d or A_β^f molecules (11–13).

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: Ia, I-region associated; bp, base pair(s).

†Present address: Becton Dickinson Immunocytometry Systems, Mountain View, CA 94039.

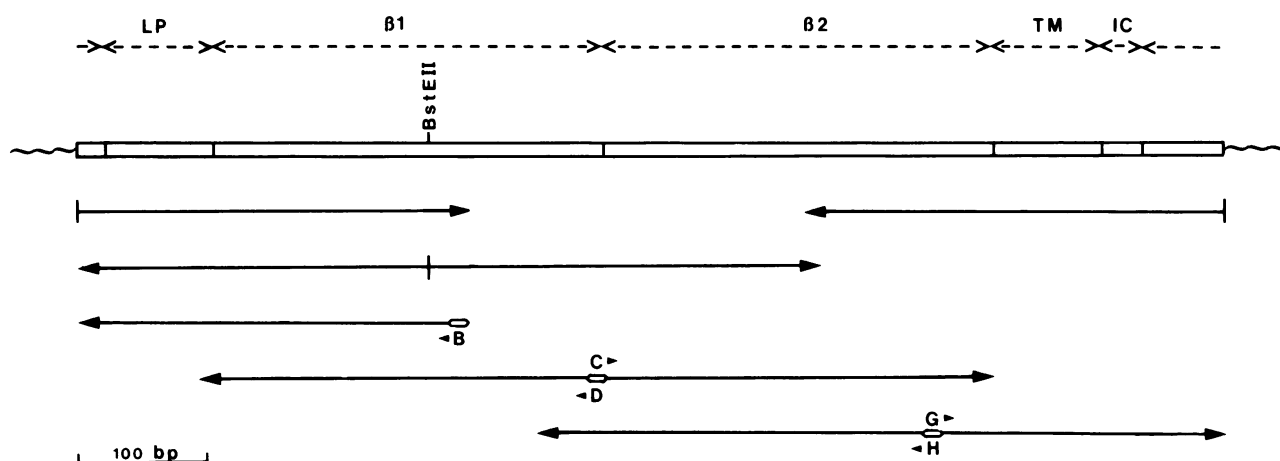


FIG. 1. Sequencing strategy for I- A_{β} cDNA clones. The open box represents full-length cDNA clones. Solid arrows indicate the direction of sequencing using either universal primer (I) or 18-bp synthetic oligonucleotide primers (\Leftarrow , B-H). The *BstEII* site shown is that used for subcloning. The structural domains as indicated above the dashed arrows. LP, leader peptide; $\beta 1$, first external domain; $\beta 2$, second external domain; TM, transmembrane region; IC, intracytoplasmic tail.

This marked structural alteration effectively divides the A_{β} chains into two distinct subgroups, consistent with what is already known about their antibody reactivity and immune response (*Ir* gene) patterns (Table 1). Since there is identity within a group in this region, including the sharing of a silent base change at nucleotide 267 (asparagine-62) in the *k/u/s/f* group, the molecular event that gave rise to the two variations of A_{β} chains probably occurred prior to the further separation of these alleles in contemporary laboratory mice. The *d/b/q* group is more like the E_{β} and human class II β -chain sequences in this region (8, 9, 14-19) and probably represents the progenitor sequence from which the other arose. Alternatively, a gene conversion-like event in this region resulting in a 6-bp longer gene could have generated the *d*-type sequence from the *k*. E_{β}^d and E_{β}^q genes are identical to the A_{β}^d sequence in this region, and either could have acted as donor (9, 16). This altered region overlaps the site of the putative A^{bml2}_{β} gene conversion event (18, 19), suggesting that it may be an area readily subject to such occurrences.

A second notable region is that from nucleotides 115-123 (encoding amino acids 12-14) in the $\beta 1$ domain (Fig. 2 and 3). Here the *k* and *u* molecules share an identical sequence but are completely distinct from the other three β chains published here and from the previously published *d* and *b* β chains (11-13). The latter five chains share a different nearly identical sequence. Again, one has to consider either an event occurring prior to the separation of the *k* from the *u* allele, or the possibility of identical evolutionary events in the two haplotypes. The *k/u* 9-residue sequence has also been found in the I- A_{β} chain of at least one strain of inbred wild mice (M. Golubić, R. Schöpfer, F. Figueroa, and J. Klein, personal communication). This sequence is not present in any other class II molecules sequenced thus far and, in data bank searches, does not appear in any mouse genomic context that bears homology to I- A_{β} . Thus, the donor of this sequence, if it is in fact the result of a gene conversion-type event, remains unknown.

I- A_{β} Protein Sequences. The predicted $\beta 1$ -domain amino acid sequences encoded by the five I- A_{β} cDNAs examined here are shown in Fig. 4 along with the published *d*, *b*, and *k* sequences (11-13). Cysteine residues probably involved in intradomain disulfide bonding are present at positions 15 and 79. A single canonical asparagine-linked carbohydrate attachment site is present at positions 19-21 in all seven chains. Among the A_{β} chains, the *d*, *b*, and *q* haplotypes show the highest degree of first domain similarity, consistent with their serological and, to some extent, immune response characteristics (Table 1). A_{β}^d appears to be the most distinct but it is most homologous to A_{β}^b and A_{β}^q , again consistent with

serological relatedness. Several regions of marked variability are present. The three most prominent are at positions 9-14, 61-67, and 85-89. An additional focus of variability is present at positions 26 and 28. The existence of three or four regions of hypervariability has also been seen in the first external domain of E_{β} chains (10, 15, 16), and two, perhaps three, in the I-A $\alpha 1$ domain (7). The two nucleotide regions that divide the A_{β} chains into subgroups are also apparent in the protein sequence. The alternative sequences in both areas probably have marked structural effects and are strong candidates for conferring some of the phenotypic characteristics of these molecules (see below).

The second domains, transmembrane and intracytoplasmic regions of A_{β} chains are very highly conserved, as are the 27 amino acids of the leader peptides ($\geq 95\%$; Fig. 2). While the functional role of the $\beta 1$ domain is probably intimately connected with its freedom to vary, the presumed functions of these other domains (processing, heterodimer formation, membrane anchoring, and perhaps intracytoplasmic signaling) appear to be associated with limited structural heterogeneity. With respect to the latter, while the $\beta 1$ and $\beta 2$ domains of A_{β} and E_{β} share 58% and 65% of residues, respectively, the leader peptide show only 31% homology (refs. 9, 12, 13, and 17; this paper). Perhaps this reflects a role for this segment in the restricted association between A and E region α and β chains on the surfaces of normal cells (20-22). The pronounced differences in the leader sequences may restrict such interactions at a very early stage in Ia biosynthesis. Alternatively, since recent experiments suggest that the first half of the $\beta 1$ domain restricts I-A α - β pairing (23), this region may place limitations on A/E associations as well. The intracytoplasmic regions of I- A_{β} chains, unlike E_{β} , contain no phosphorylatable serines. This may reflect as yet undetermined functional differences between the two types of molecules.

A_{β} First Domains Are Equally Homologous to Those of DQ_{β} and DR_{β} . Based on amino-terminal amino acid sequence, I-E molecules were considered to be the murine DR homologues, while I-A molecules corresponded more closely to DQ (24). Comparison of the seven complete A_{β} protein sequences with three DQ_{β} sequences (17) indicates that 72% of comparable amino acid positions are identical in 9 of the 10 chains. When A_{β} protein sequences are compared to four DR_{β} sequences (17), only 59% of positions are identical in 10 of the 11 chains. However, this greater similarity between A_{β} and DQ_{β} is due to very high second-domain relatedness, the first domain of A_{β} being almost equally homologous to those of DQ_{β} and DR_{β} (54% vs. 51%). In the remainder of the molecule, A_{β} vs.

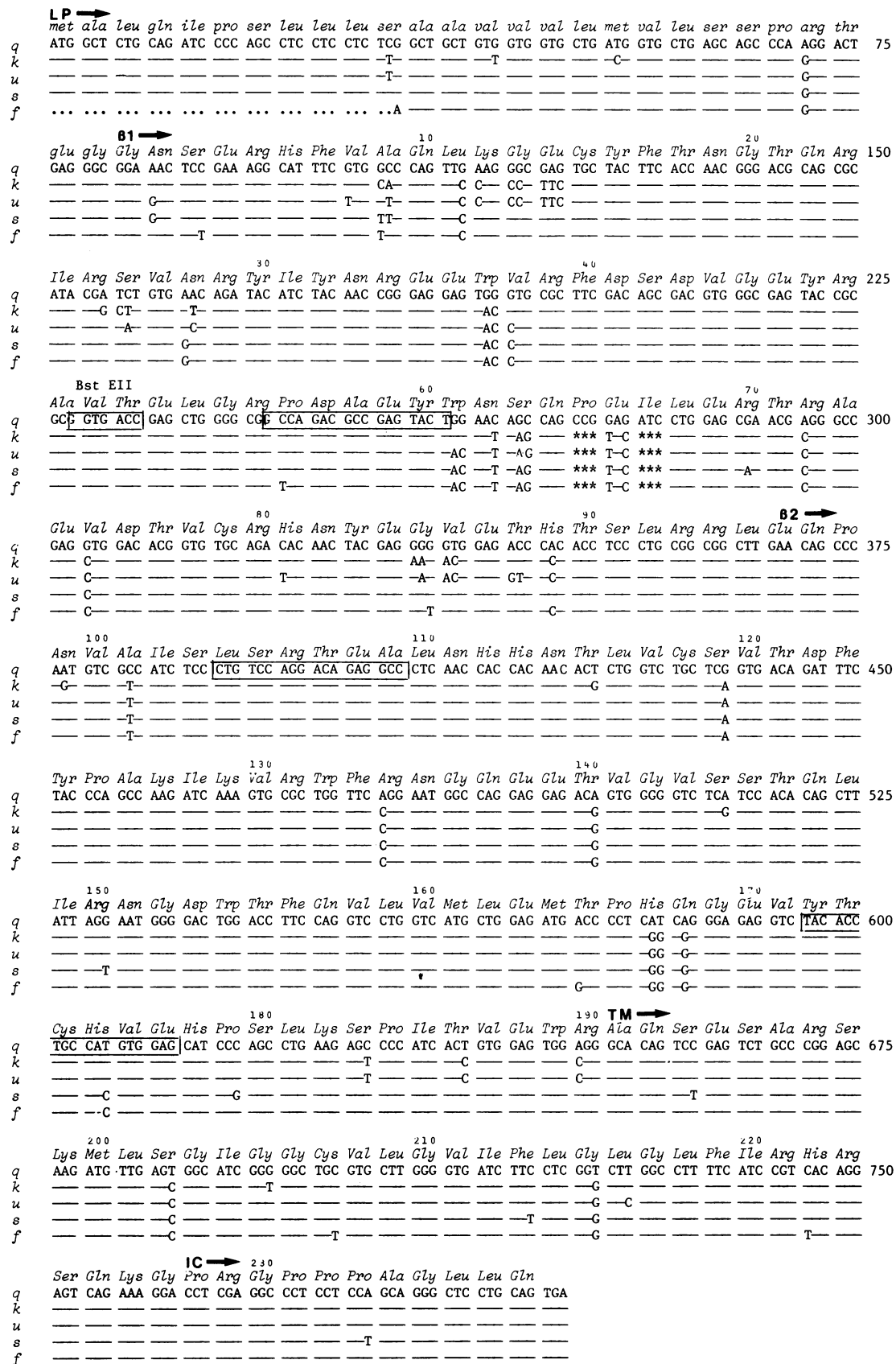


FIG. 2. Nucleotide sequence of I-A β cDNA coding regions. Dashed lines indicate identity with the *q* haplotype sequence; only differences are noted. The structural domains are indicated as in Fig. 1. Asterisks indicate gaps relative to the *q* haplotype sequence. Sequences corresponding to the 18-bp oligonucleotide primers and the *Bst*EII restriction enzyme site used in sequencing are boxed. The amino acid sequence is that predicted from the *q* haplotype DNA sequence. Numbers above the line refer to amino acid position; those to the far right refer to nucleic acid position. Dots indicate regions not sequenced.

DQ β = 84% and A β vs. DR β = 65%. Such calculations suggest that, in addition to the known allelic variability, the evolu-

tionary history of the first domain of DQ β may differ substantially from that of the remainder of the molecule.

	12	13	14		62	63	64	65	66	67
	<i>Lys</i>	<i>Gly</i>	<i>Glu</i>		<i>Asn</i>	<i>Ser</i>	<i>Gln</i>	<i>Pro</i>	<i>Glu</i>	<i>Ile</i>
<i>q</i>	AAG	GGC	GAG	AAC	AGC	CAG	CCG	GAG	ATC
<i>d</i>	---	---	---	---	---	---	---	---	---
<i>i</i>	-T-	---	---	---	---	---	---	---	---
<i>s</i>	---	---	---	-T	-AG	---	***	T-C	***
<i>f</i>	---	---	---	-T	-AG	---	***	T-C	***
<i>u</i>	C-	CC-	TTC	-T	-AG	---	***	T-C	***
<i>k</i>	C-	CC-	TTC	-T	-AG	---	***	T-C	***
	<i>Gln</i>	<i>Pro</i>	<i>Phe</i>		<i>Asn</i>	<i>Lys</i>	<i>Gln</i>	*	<i>Tyr</i>	*

FIG. 3. Nucleotide sequence encoding amino acids 12-14 and 62-67 of A_β chains. The amino acid sequence for the *q* and *k* chains are indicated above and below the nucleotide sequences, respectively. *d*, *b*, and *k* sequences are from refs. 11-13.

Localization of Antibody Reactivity and Immune Response to Particular Amino Acid Residues. A main purpose for determining primary structure of class II molecules is to attempt to localize regions responsible for *Ir* gene function, major histocompatibility complex restriction, and alloreactivity, and thus to understand how Ia antigens function in the immune system. Table 1 summarizes some of the immune response characteristics and serologic specificities that map to the A region or to the A_β gene. Indicated in the table are residues in I-A α and β chains that correlate with particular reactivity patterns and are likely candidates for affecting immune responsiveness, T-cell allorecognition, or serologic epitopes.

Ia.1, Ia.17, and the determinant recognized by the MK-S4 monoclonal antibody all reside on the I-A_β chains (refs. 25-28; P.P.J., unpublished data). The Ia.1 specificity is shared by the *k*, *u*, and *f* haplotypes (29). Only one amino acid residue, proline-89 in the third region of clustered variability, is consistent with this reactivity pattern. Ia.17 is present on the *k*, *u*, *s*, and *f* A_β chains (29). This is the same clustering of haplotypes seen for the region encoding amino acids 65-67, and this area probably determines this serologic specificity. Allen *et al.* (30) have shown that alteration of amino acid 69 in an I-A_β chain from a glutamic acid to a lysine removes the Ia.17 specificity, supporting the idea that this region is critical to the determinant. The MK-S4 determinant is expressed on the *u*, *s*, and *f* haplotype molecules (ref. 25; P.P.J., unpublished data; R.B. Fritz, personal communication). The tyrosine at position 61 is unique to these chains and may comprise the serological epitope or affect tertiary structure in such a way as to generate the epitope. Position

Table 1. Ia specificity and antigen responder status of seven mouse haplotypes

Haplo-type	Ia specificity			Antigen				MBP1-37*
	Ia.1	Ia.17	MK-S4	HGAL	TGAL	PheGAL		
<i>q</i>	-	-	-	NR	NR	R	-	
<i>k</i>	+	+	-	R	NR	R	+	
<i>u</i>	+	+	+	NR	NR	NT	+	
<i>s</i>	-	+	+	NR	NR	NR	-	
<i>f</i>	+	+	+	NR	NR	R	-	
<i>d</i>	-	-	-	NR	R	R	-	
<i>b</i>	-	-	-	NR	R	R	-	
β-chain residues	89	65-67	38, 61	9, 28, 85, 99	40, 86	70	12-14, 86	
α-chain residues†	NA	NA	NA	57, 75	None	72, 73	11, 53, 56	

NR, nonresponder; R, responder; NT, not tested; NA, not applicable.

*Amino acid residues 1-37 of myelin basic protein; +, disease; -, no disease.

†Data from ref. 7.

38, also unique in these three chains, has a more conservative interchange (Val → Leu).

The genetic control of responsiveness to poly(Tyr,Glu)-poly(DLAla)-poly(Lys), poly(His,Glu)-poly(DLAla)-poly(Lys) and poly(Phe,Glu)-poly(DLAla)-poly(Lys) has been studied extensively and maps to the I-A subregion (31-36). The response characteristics for each haplotype whose I-A α and β chains have been sequenced are shown in Table 1. Mice of the *d* and *b* haplotype share responsiveness to poly(Tyr, Glu)-poly(DLAla)-poly(Lys) (33). Only two β-chain amino acid positions, 40 and 86, are consistent with this pattern. The tyrosine/phenylalanine interchange at position 40 may not dramatically alter protein structure, while the proline at position 86 in A_β^d and A_β^b could result in a markedly different tertiary conformation. No α-chain position conforms to the *d/b* positive, *q/k/u/s/f* negative pattern (7).

Of the seven haplotypes listed, only mice that are I-A^k respond to the antigen poly(His,Glu)-poly(DLAla)-poly(Lys) (33). Comparing the seven A_β sequences, several positions are potentially involved: residues 9, 28, and 85 in the β1 domain, and residue 99 in the β2 domain. Each of the alterations in the nonresponder haplotype chains is probably significant enough to alter tertiary conformation. However, the lysine at position 85 in A_β^k introduces a net charge difference from all of the other β chains and may be the most critical residue. In addition, either position 57 or 75 in the I-A α chain may be involved (7).

Five of the seven haplotypes listed in Table 1 respond to poly(Phe,Glu)-poly(DLAla)-poly(Lys); *s* haplotype mice do not respond (33) and *u* haplotype mice have not been tested. There is only one position where all five responder haplotype β chains are the same but different from *s*: position 70. The glutamine/arginine interchange results in a charge difference and may be sufficient to alter responsiveness to this antigen. Although not tested, this would predict that *u* haplotype animals will also respond to poly(Phe,Glu)-poly(DLAla)-poly(Lys), since there is no position where *u* and *s* β chains differ from those of the other five haplotypes. In the α chains, two positions, 72 and 73, could also be involved (7).

A fourth antigen that has recently been of interest to workers in this laboratory is the amino-terminal 37-residue peptide of myelin basic protein. This peptide has been shown to be encephalitogenic in *k* and *u* haplotype mice, causing experimental allergic encephalomyelitis (EAE), a disease very similar to multiple sclerosis (37, 38). The susceptibility of mice to this disease is governed by genes in the I region (39, 40), and I-A restricted MBP1-37-specific T-cell clones have recently been described that also produce disease when administered *in vivo* (41, 42). Residues 12-14 in the *k* and *u* chains are markedly different from those of the other five chains. Analyses done according to Hopp and Woods (43) predict that the net effect of the three-amino acid change from Lys-Gly-Glu (A_{βq}, -*d*, -*s*, -*f*) to Gln-Pro-Phe (A_{βk}, -*u*) would be to create a region of markedly increased hydrophobicity. Such a change could affect overall conformation of the β1 domain. This distinct three-residue stretch may be involved in the response to this as well as other antigens and in generating *k/u*-specific serologic epitopes such as Ia.31 (29). Position 86 in the β chain and positions 11, 53, and 56 in the α chains (7) may also be involved in susceptibility to peptide-induced EAE.

The primary structures of seven I-A α and β chains have now been completed by the sequencing of their cDNAs (refs. 7 and 11-13; this paper). These sequence data allow the comparison of molecules known to differ both antigenically and functionally and the identification of amino acid residues potentially responsible for these differences. In conjunction with analogous studies on the molecular nature of antigens, it should eventually be possible to construct models for the structure of an Ia-plus-antigen complex. The involvement of

