

# Primary structural evidence that the *H-2D<sup>q</sup>* region encodes at least three distinct gene products: *D<sup>q</sup>*, *L<sup>q</sup>*, and *R<sup>q</sup>*

(histocompatibility antigens/radiosequence analysis/tryptic peptide map analysis/*H-2* gene evolution)

ERIK P. LILLEHOJ\*, TED H. HANSEN†, DAVID H. SACHS‡, AND JOHN E. COLIGAN\*§

\*Laboratory of Immunogenetics, Building 5, Room B1-04, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20205; †Department of Genetics, Washington University School of Medicine, St. Louis, MO 63310; and ‡Transplantation Biology Section, Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205

Communicated by Donald C. Shreffler, January 11, 1984

**ABSTRACT** Class I gene products *D*, *L*, and *R* encoded by the *D* region of the *H-2<sup>q</sup>* haplotype were isolated by sequential immunoprecipitation from Nonidet P-40 extracts of B10.AKM murine spleen cells. Primary structural comparisons of these molecules were made by tryptic peptide map analyses and radiochemical amino acid sequence determinations. Partial NH<sub>2</sub>-terminal radiosequence analyses of the peptides generated by CNBr cleavage of these three molecules defined the *R<sup>q</sup>* molecule as a gene product that is distinct from the *D<sup>q</sup>* and *L<sup>q</sup>* molecules. Comparative tryptic peptide map analyses of [<sup>3</sup>H]arginine-, [<sup>3</sup>H]lysine-, or [<sup>3</sup>H]mannose-labeled molecules were performed to determine if the *D<sup>q</sup>* and *L<sup>q</sup>* molecules could be distinguished. The two molecules were identical in their [<sup>3</sup>H]lysine-labeled peptides (17 in common) and [<sup>3</sup>H]mannose-labeled glycopeptides but could be differentiated by their [<sup>3</sup>H]arginine-labeled peptides (18 of 21 in common). These results suggest an amino acid sequence homology of 99% or greater between the *D<sup>q</sup>* and *L<sup>q</sup>* molecules. These data therefore provide structural evidence for three *D* region gene products and suggest that their antigenic differences are not due to post-translational glycosylation differences. Thus, the *D* region of the *H-2<sup>q</sup>* haplotype may provide an example of a relatively recent class I gene duplication, indicating that these genes are in a dynamic state of evolution through expansion or contraction of the gene pool.

Class I genes of the major histocompatibility complex are expressed as polymorphic integral membrane glycoproteins of approximately 45,000 *M<sub>r</sub>* in noncovalent association with β<sub>2</sub>-microglobulin (1). The murine class I transplantation antigens have been genetically mapped to the *H-2K* and *H-2D* regions of chromosome 17 (2, 3). Although originally each *H-2K* and *H-2D* region was believed to encode a single gene product (4-6), it has now been well documented that certain haplotypes express multiple, antigenically distinct molecules determined by their *K* and *D* regions (see ref. 1 for review). For example, the *D* region of the *H-2<sup>q</sup>* haplotype has been reported to encode at least three antigenically distinct molecules, designated *D<sup>q</sup>*, *L<sup>q</sup>*, and *R<sup>q</sup>* (7), and the *D* region of the *H-2<sup>d</sup>* haplotype possibly encodes as many as five antigenically distinct molecules: *D<sup>d</sup>*, *L<sup>d</sup>*, *R<sup>d</sup>*, *M<sup>d</sup>*, and *L2<sup>d</sup>* (7, 8).

The molecular basis for the existence of these multiple *D* region-encoded molecules is unclear except for the *D<sup>d</sup>* and *L<sup>d</sup>* molecules, which have been shown to have distinct primary structures (9) and to be encoded by separate genes (10). The fact that cosmid cloning studies have defined only three genes in the *D* region of the *H-2<sup>d</sup>* haplotype, one of which is thought to be a pseudogene (11), raises the question as to whether the additional molecules thought to be encoded within the *D* region of certain strains of mice are the

products of distinct genes or whether they originate by other means such as post-translational modifications of a single gene product. Studies of the *D* region of *H-2<sup>b</sup>* haplotype mice have provided evidence for only a single gene (12) and gene product (13), indicating that the number of *D* region genes may be variable, although other investigators have used serological methods to describe multiple molecules mapping to the *D<sup>b</sup>* region (14).

In addition to possible variability in the number of *D* region gene products, several *D* region molecules show much higher levels of homology than has previously been observed for other class I molecules. Sequence studies (15) have shown that the *D<sup>b</sup>* and *L<sup>d</sup>* molecules are the two most closely related *H-2* class I molecules so far studied, having 94% homology. Tryptic peptide maps have barely been able to reveal differences in the *D<sup>q</sup>*, *L<sup>q</sup>*, and *L<sup>d</sup>* molecules (16), and no serological differences are evident between the *L<sup>q</sup>* and *L<sup>d</sup>* molecules (17). These studies are further complicated by the fact that the *L<sup>q</sup>* preparations undoubtedly also contained the *R<sup>q</sup>* molecule (7).

Studies of *D* region-encoded molecules therefore raise the following questions: (i) How many *D* region genes are expressed? (ii) How are different *D* region gene products of the same haplotype related to each other? (iii) How are the *D* region genes of different haplotypes related? (iv) Is the number of different class I genes in each *D* region fixed or in a state of evolutionary flux due to genetic mechanisms of duplication and deletion? To begin to answer these questions, we have compared the primary structures of several *D* region-encoded molecules.

## MATERIALS AND METHODS

**Alloantisera and Monoclonal Antibodies.** The reagents used and their source and serological specificity are listed in Table 1. Alloantiserum anti-H-2.30 was produced as described (18). Monoclonal antibodies 30-5-7 and 28-14-8 were produced from hybridoma cell lines grown in ascites and have been previously characterized (19).

**Preparation of Radiolabeled Antigens.** Murine B10.AKM (*K<sup>k</sup>*, *D<sup>q</sup>*) spleen cells were cultured with <sup>3</sup>H-labeled amino acids (histidine, isoleucine, leucine, tyrosine, arginine, or lysine; Amersham) as described (7). Biosynthetic incorporation of D-[2-<sup>3</sup>H(N)]mannose (New England Nuclear) was performed as described (20). Labeled cells were lysed with 0.5% Nonidet P-40 and the glycoprotein fraction was obtained by chromatography on a column of lentil lectin coupled to Sepharose 4B (7). Class I *H-2* molecules were isolated in the order *D<sup>q</sup>*, *L<sup>q</sup>*, and *R<sup>q</sup>* by sequential immunoprecipitation from the same glycoprotein fraction (see *Results*).

**Preparation of Peptides by Papain and CNBr Cleavage.** The immunoprecipitated molecules were exposed to limited papain digestion (Sigma) in 0.01 M Tris-HCl/0.15 M NaCl, pH

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.

§To whom reprint requests should be addressed.

Table 1. Reagents used and their specificity on antigens of B10.AKM mice [ $K^k$  ( $D^q$ ,  $L^q$ ,  $R^q$ )]

Reagent*	Serologic specificity	Molecule(s) detected
(LP.R.III $\times$ B10.A) anti-B10.AKM <sup>†</sup> (anti-H-2.30)	H-2.30	$D^q$
BALB/c-H-2 <sup>dm2</sup> anti-BALB/c <sup>‡</sup> (monoclonal 30-5-7)	H-2.65	$D^q$ , $L^q$
C3H anti-C3H.SW <sup>‡</sup> (monoclonal 28-14-8)	H-2.64	$D^q$ , $L^q$ , $R^q$

\*To isolate the  $D^q$ ,  $L^q$ , and  $R^q$  molecules by immunoprecipitation, the reagents were employed in the order given from top to bottom (see Fig. 1).

<sup>†</sup>Previously characterized by Hansen *et al.* (18).

<sup>‡</sup>Previously characterized by Ozato *et al.* (19).

7.4. The papain cleavage fragments were separated from  $\beta_2$ -microglobulin by gel filtration on Sephadex G-75 (2.5  $\times$  140 cm) in 1 M formic acid/0.01% 2-mercaptoethanol and lyophilized. CNBr digestion was performed as described (9), and the resulting peptides were separated by gel filtration on Sephacryl S-200 (2.0  $\times$  195 cm) in 6 M guanidine-HCl/0.5% acetic acid. The disulfide-linked peptides isolated from this digestion were reduced and carboxamidomethylated as previously described (9), separated by gel permeation chromatography as above, and desalted.

**Automated Amino Acid Sequence Analysis.** Extensive details of the radiochemical amino acid sequencing procedures have been published (see ref. 21 for review).

**Tryptic Peptide Map Analysis.** Tryptic peptides of [<sup>3</sup>H]arginine-, [<sup>3</sup>H]lysine-, or [<sup>3</sup>H]mannose-labeled  $D^q$ ,  $L^q$ , and  $R^q$  molecules were prepared using a modification of the procedure described by Brown *et al.* (22). Briefly, reduced and alkylated 45,000  $M_r$  class I molecules were purified by gel filtration on Sephacryl S-300 in 0.05 M Tris-HCl/0.5% NaDodSO<sub>4</sub>/0.01% 2-mercaptoethanol, pH 7.5, dialyzed against 0.001% NaDodSO<sub>4</sub>, and lyophilized. The isolated class I molecules were dissolved in 2 M formic acid, precipitated

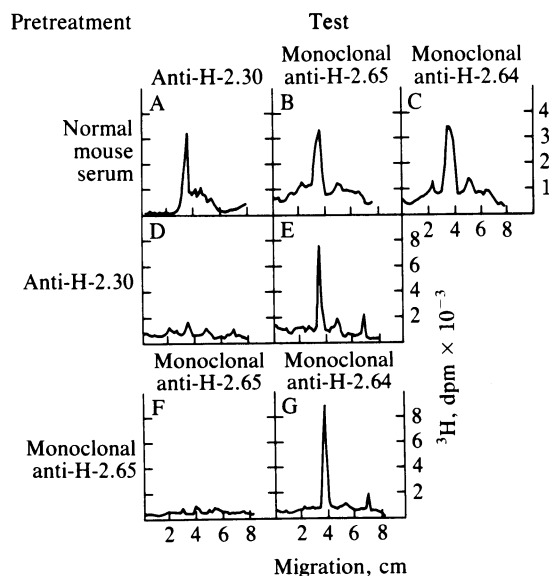


FIG. 1. Sequential immunoprecipitation of the  $D^q$ ,  $L^q$ , and  $R^q$  molecules. Radiolabeled B10.AKM glycoproteins were pretreated with normal mouse serum, anti- $D^q$  alloantiserum (anti-H-2.30), or monoclonal antibody 30-5-7 (anti-H-2.65) as shown at the left. Supernatants were then tested with anti-H-2.30, anti-H-2.65, or monoclonal antibody 28-14-8 (anti-H-2.64) as shown at the top. The secondary immunoprecipitates were analyzed by NaDodSO<sub>4</sub>/polyacrylamide gel electrophoresis.

with acetone, and resuspended in 0.05 M NH<sub>4</sub>HCO<sub>3</sub>, pH 7.4, and tosylphenylalanyl chloromethyl ketone-treated trypsin (Millipore) in 0.001 M HCl was added at a weight ratio of 10:1 (carrier protein:enzyme). The tryptic peptides were resolved by reversed-phase HPLC on a Supelcosil C<sub>18</sub> column (150 mm  $\times$  4.6 mm inside diameter, Supelco, Bellefonte, PA) at a flow rate of 1 ml/min with a Spectra Physics (Santa Clara, CA) SP8000B liquid chromatograph. After initial isocratic elution for 5 min in 0.1% CF<sub>3</sub>COOH, linear gradient elution to 0.1% CF<sub>3</sub>COOH/30% (vol/vol) CH<sub>3</sub>CN was run for 95 min and followed by gradient elution to 0.1% CF<sub>3</sub>COOH/100% CH<sub>3</sub>CN in 10 min. Fractions (0.5 ml) were analyzed by liquid scintillation counting.

## RESULTS

**Sequential Immunoprecipitation.** The  $D^q$ ,  $L^q$ , and  $R^q$  molecules were isolated from radiolabeled cell lysates by sequential immunoprecipitation and analyzed by NaDodSO<sub>4</sub>/polyacrylamide gel electrophoresis (Fig. 1). Since the  $D^q$  and  $L^q$  molecules possess both the H-2.64 and H-2.65 serological specificities and the  $R^q$  molecule the H-2.64 specificity, it was necessary to verify complete clearance of the  $D^q$  molecule prior to isolation of the  $L^q$  antigen and likewise complete removal of the  $L^q$  molecule before precipitation of the  $R^q$  glycoprotein. When B10.AKM glycoprotein fractions were treated with anti-H-2.30 alloantiserum ( $D^q$  specific), all molecules reactive with this reagent were removed (Fig. 1D) while monoclonal antibody 30-5-7 (anti-H-2.65) still precipitated a 45,000  $M_r$  product, the  $L^q$  molecule (Fig. 1E). After isolation of the  $L^q$  molecule, the absence of reaction of the

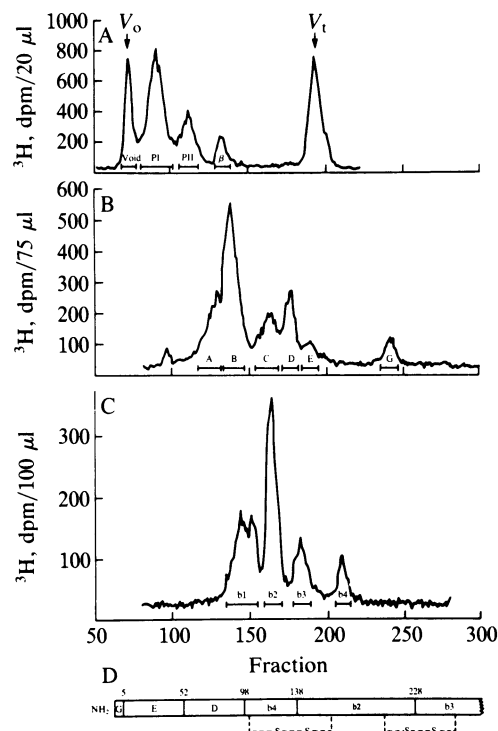


FIG. 2. (A) Chromatography of peptides from papain digestion of the  $D^q$  molecule labeled with [<sup>3</sup>H]leucine, [<sup>3</sup>H]tyrosine, and [<sup>3</sup>H]arginine on Sephadex G-75. The fraction size was 2.5 ml and the flow rate was 13 ml/hr.  $\beta_2$ -microglobulin. (B) Chromatography of peptides from CNBr cleavage of the PI fragment of the  $D^q$  molecule labeled with [<sup>3</sup>H]leucine, [<sup>3</sup>H]tyrosine, and [<sup>3</sup>H]arginine on Sephacryl S-200. The fraction size was 2.0 ml and the flow rate was 12 ml/hr. (C) Rechromatography of pool B after reduction and carboxamidomethylation. (D) Alignment of the peptides produced by CNBr digestion. Numbers are amino acid residues.

remaining antigen supernatant with monoclonal antibody 30-5-7 indicated complete removal of the L<sup>q</sup> antigen (Fig. 1F) while still leaving a 45,000 M<sub>r</sub> product reactive with monoclonal antibody 28-14-8 (anti-H-2.64), the R<sup>q</sup> molecule (Fig. 1G). These results establish that each precipitated molecule was quantitatively free of the previously isolated molecule and also confirm that the D<sup>q</sup> region encodes at least three antigenically distinct class I molecules. (These results do not, however, preclude the existence of multiple gene products in any of the immunoprecipitated fractions.)

**Isolation of Fragments After Papain Cleavage.** The immunoprecipitated D<sup>q</sup>, L<sup>q</sup>, and R<sup>q</sup> glycoproteins were initially subjected to papain cleavage and the digestion products were resolved by gel filtration chromatography. Fig. 2A illustrates a representative elution profile of the D<sup>q</sup> molecule. With all three molecules, a major peptide (PI, ≈37,000 M<sub>r</sub>) was separated from undigested material eluting in the void volume (V<sub>0</sub>), a smaller papain fragment (PII, ≈28,000 M<sub>r</sub>), β<sub>2</sub>-microglobulin (β<sub>2</sub>m), and small digestion product peptides eluting in the total volume (V<sub>1</sub>). The major fragment consists of approximately the NH<sub>2</sub>-terminal 280 residues of the intact molecule as a result of papain digestion between the third extracellular domain and the hydrophobic transmembrane region (9). The PII fragment is generated by papain cleavage between the second and third extracellular domains and corresponds approximately to the NH<sub>2</sub>-terminal 180 residues (9).

**Isolation and Alignment of Peptides Produced by CNBr Cleavage.** Fig. 2B illustrates a representative gel filtration elution profile after CNBr cleavage of the major fragment from papain digestion of the D<sup>q</sup> molecule. Six pools were resolved, one of which (pool B) could be separated into four additional pools after reduction and alkylation (Fig. 2C). Sequence analyses of the peptides present in these pools and comparison of these sequences with the L<sup>d</sup> molecule indicated that pools D, E, G, b2, b3, and b4 contained pure peptides that had the following order: G (residues 1–5), E (6–52), D

(53–98), b4 (99–138), b2 (139–228), and b3 (229–papain cleavage site). This alignment is shown in Fig. 2D. Pools A and b1 contained undigested and partial cleavage product peptides (9). Pool C contained a peptide encompassing positions 6 to 98 that results from incomplete CNBr digestion at Met-52. Analyses of the L<sup>q</sup> and R<sup>q</sup> molecules yielded identical peptides.

**Amino Acid Sequence Determination.** Fig. 3 shows the amino acid sequence determined for each of the D<sup>q</sup>, L<sup>q</sup>, and R<sup>q</sup> peptides and their alignment to the L<sup>d</sup> amino acid sequence. All three H-2<sup>q</sup> haplotype molecules differ in sequence from the L<sup>d</sup> molecule. Of the positions analyzed, the D<sup>q</sup> and L<sup>q</sup> molecules differ from L<sup>d</sup> at positions 81, 105, 121, and 155. The D<sup>q</sup> molecule also differs from L<sup>d</sup> at position 83. The R<sup>q</sup> molecule differs from L<sup>d</sup> at positions 95 and 155. For the H-2<sup>q</sup> haplotype molecules, R<sup>q</sup> can be distinguished from both the D<sup>q</sup> and L<sup>q</sup> molecules at three positions (81, 105, and 121) of the 34 and 33 comparable positions, respectively. In contrast, of 55 positions comparable, the D<sup>q</sup> and L<sup>q</sup> molecules are indistinguishable.

**Tryptic Peptide Map Analyses.** Because the D<sup>q</sup> and L<sup>q</sup> molecules appeared identical by partial radiochemical amino acid sequence analyses, we performed tryptic peptide map comparisons to discern possible primary structural differences between these two molecules. Fig. 4A shows a comparison of the [<sup>3</sup>H]arginine-labeled tryptic peptides of the D<sup>q</sup> and L<sup>q</sup> molecules resolved by reverse-phase HPLC. Two peptides unique to the D<sup>q</sup> molecule and one peptide unique to the L<sup>q</sup> molecule are indicated by the arrows. These differences between the [<sup>3</sup>H]arginine-labeled peptides of the D<sup>q</sup> and L<sup>q</sup> molecules are reproducible since they have been observed repeatedly with two different antigen preparations and several different tryptic digests of each preparation. In contrast, comparative tryptic mapping of the [<sup>3</sup>H]lysine-labeled D<sup>q</sup> and L<sup>q</sup> molecules demonstrated 100% coincidence of peptides (Fig. 4B). HPLC analysis of [<sup>3</sup>H]mannose-labeled tryptic peptides revealed that the variations evident in

L <sup>d</sup>		5	10	15	20	25	30	35	40	45																																								
	G	P	H	S	M	R	Y	F	E	T	A	V	S	R	P	G	L	G	E	P	R	Y	I	S	V	G	Y	V	D	N	K	E	F	V	R	F	D	S	D	A	E	N	P	R	Y					
D <sup>q</sup>			H	(M)	R	Y	F					V	R	L		R	Y	I	V	Y	V				K	F	V	R	F																					
L <sup>q</sup>			H	(M)	R	Y	F					V	R	L		R	Y	I	V	Y	V					F	V	F																					(Y)	
R <sup>q</sup>			H	(M)	R	Y							R	L		R	Y		<sup>I</sup> <sub>V</sub> *	Y								(R)																						
L <sup>d</sup>				50	55	60	65	70	75	80	85	90																																						
	E	P	Q	A	P	W	M	E	Q	E	G	P	E	Y	W	E	R	I	T	Q	I	A	K	G	Q	E	Q	W	F	R	V	N	L	R	T	L	L	G	Y	N	Q	S	A	G						
D <sup>q</sup>				(M)		(P)	Y	R	I	I	K																F	R	V	L	R	X	L	R	Y	Y														
L <sup>q</sup>				(M)		(P)	Y	R	I	I	(K)																F	R	L	R	X	L	Y	Y																
R <sup>q</sup>				(M)			Y	R	I	I																																								
L <sup>d</sup>					95	100	105	110	115	120	125	130	135																																					
	G	T	H	T	L	Q	W	M	Y	G	C	D	V	G	S	D	G	R	L	L	R	G	Y	E	Q	F	A	Y	D	G	R	D	Y	I	A	L	N	E	D	L	K	T	W	T	A					
D <sup>q</sup>					(M)Y		L	L	Y		Y	L	I	L	L																																			
L <sup>q</sup>					(M)Y		L	L	(Y)			L	I	L	(L)																																			
R <sup>q</sup>					H	I	(M)		X	L	L																	X	(I)	(L)																				
L <sup>d</sup>					140	145	150	155	160	165	170	175	180																																					
	A	D	M	A	A	Q	I	T	R	K	W	E	Q	A	G	A	E	Y	Y	R	A	Y	L	E	G	E	C	V	E	W	L	H	R	Y	L	K	N	G	N	A	T	L	L							
D <sup>q</sup>					(M)	I	R	R			H	Y	Y	L																																				
L <sup>q</sup>					(M)	I	R	R			H	Y	Y	L																																				
R <sup>q</sup>					(M)		(R)	(R)			X	†	(Y)	(Y)	L																																			
L <sup>d</sup>					225	230	235	240	245	250	255																																							
	G	E	E	L	T	Q	D	M	E	L	V	E	T	R	P	A	G	D	G	T	F	Q	K	W	A	S	V	V	V	P	L	G	K	E	Q															
D <sup>q</sup>					(M)	L																																												
L <sup>q</sup>					(M)	L																																												
R <sup>q</sup>					(M)																																													

FIG. 3. Partial amino acid sequence of D<sup>q</sup>, L<sup>q</sup>, and R<sup>q</sup> compared to the L<sup>d</sup> sequence. The L<sup>d</sup> sequence is from Evans *et al.* (23) and Moore *et al.* (24). The standard one-letter symbols for amino acids are used (25). X is not His, Ile, Leu, Tyr, Arg, or Met. Methionine residues in parentheses were assigned according to homology of the CNBr peptides to the L<sup>d</sup> molecule. Other residues in parentheses indicate tentative assignments based upon the radioactivity in the butyl chloride extract and homology to the L<sup>d</sup> molecule in the absence of HPLC data.

\*Both Ile and Tyr were detected at position 23 of the R<sup>q</sup> molecule during separate sequence analyses. This observation was repeated using a different antigen preparation containing both Ile and Tyr.

†X refers to the absence of Leu or Tyr at position 155 of the R<sup>q</sup> molecule.

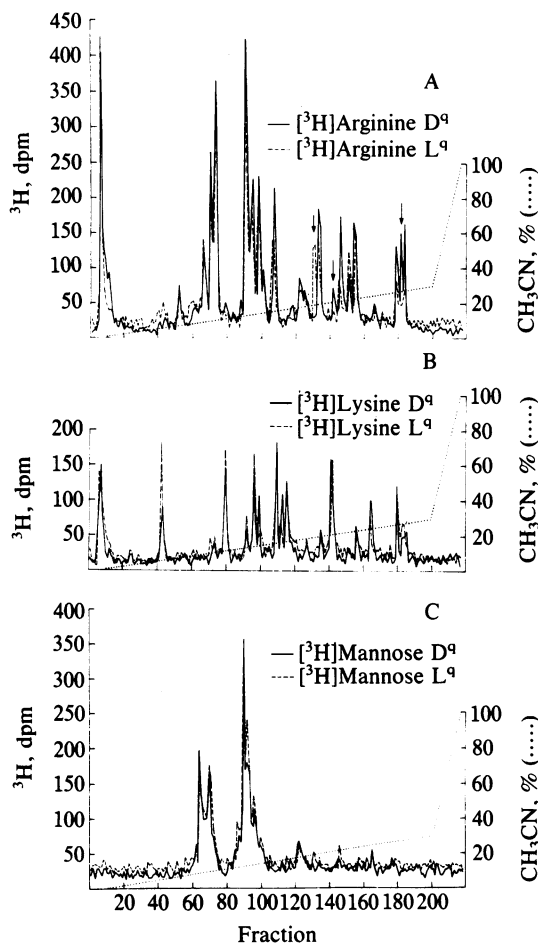


FIG. 4. Comparative tryptic peptide map analyses of the  $D^a$  (—) and  $L^a$  (---) molecules by reverse-phase HPLC. Both molecules were digested simultaneously and mapped consecutively. [ $^{14}\text{C}$ ]Arginine-labeled tryptic peptides of  $\beta_2$ -microglobulin served as an internal standard to verify alignment of the separate maps. The arrows indicate the peptides demonstrating differences between the two molecules. (A) Comparison of the [ $^3\text{H}$ ]arginine-labeled peptides. (B) Comparison of the [ $^3\text{H}$ ]lysine-labeled peptides. (C) Comparison of the [ $^3\text{H}$ ]mannose-labeled peptides.

the [ $^3\text{H}$ ]arginine-labeled peptides were not due to differences in N-linked glycosylation (Fig. 4C).

Comparative tryptic peptide mapping of the  $R^a$  molecule revealed multiple [ $^3\text{H}$ ]arginine and [ $^3\text{H}$ ]lysine-labeled peptide differences with both the  $D^a$  and  $L^a$  molecules (data not shown). Similar analyses were performed with the  $L^d$ ,  $D^b$ , and  $K^b$  molecules. A summary of the tryptic peptide homologies among these molecules is given in Table 2. These results indicate that the  $D^a$ ,  $L^a$ , and  $R^a$  molecules are closely related but unique gene products. They confirm their primary structure relationships seen by amino acid sequence analyses—i.e., these three molecules are highly homologous.

**Amino Acid Sequence Homology of the  $D^a$ ,  $L^a$ , and  $R^a$  Molecules with Other Murine Class I Antigens.** Table 3 compares the primary sequences of the three  $D^a$ -end molecules to each other and to those of other murine class I antigens. Although extensive primary structure determinations have been reported for these other class I molecules, only those positions assigned residues in the  $D^a$ ,  $L^a$ , and  $R^a$  molecules have been compared. As shown, the  $D^a$ ,  $L^a$ ,  $R^a$ ,  $L^d$ , and  $D^b$  molecules are generally more homologous to each other than to other murine class I molecules. Also, with few exceptions, this homology is greater than that found between other class I molecules.

Table 2. Tryptic peptide homology among the  $D^a$ ,  $L^a$ ,  $R^a$ ,  $L^d$ , and  $D^b$  molecules

Molecules compared	Peptides compared	No. shared peptides/ total no. peptides	Total comparison	% peptide homology
$D^a$ vs. $L^a$	{Lys Arg}	{15/15 18/21}	33/36	92
$D^a$ vs. $R^a$	{Lys Arg}	{7/20 17/26}	24/46	52
$L^a$ vs. $R^a$	{Lys Arg}	{7/20 15/27}	22/47	47
$D^a$ vs. $L^d$	{Lys Arg}	{8/16 15/26}	23/42	55
$R^a$ vs. $L^d$	{Lys Arg}	{5/16 16/29}	21/45	47
$D^a$ vs. $D^b$	Arg	16/28	16/28	57
$R^a$ vs. $D^b$	Arg	14/28	14/28	50
$D^a$ vs. $K^b$	Arg	10/29	10/29	34
$R^a$ vs. $K^b$	Arg	7/32	7/32	22

## DISCUSSION

A distinguishing feature of the genes located within the  $H-2K$  and  $-D$  regions is their high level of genetic polymorphism. This polymorphism is reflected in two ways: first, there are multiple loci, which apparently have arisen by gene duplication, that encode the major transplantation antigens and, second, multiple alleles exist for each locus, perhaps as many as 100 in some cases (30). While the physiological function(s) of class I antigens have not yet been definitively established, it would seem likely that such extensive polymorphism must have some evolutionary importance. Work by many investigators indicates that foreign molecules, such as viral antigens expressed on the host cell surface, are recognized by cytotoxic T lymphocytes in the context of the host's own class I antigens (see ref. 31 for review). This finding has been taken as a clue to the physiological function of these molecules and to the role of polymorphism. Accordingly, the recognition of viral antigens in the context of host class I antigens might serve to focus effector immune cells on the source of the virus, the infected cell, rather than on free viral particles, as a more effective means of combating infection. Polymorphism would serve to expand the repertoire of associative recognition—e.g., certain allelic products might be suitable to present a particular antigen rather than other products of that same or another similar locus.

The exact number of expressed H-2 class I molecules is currently a point of controversy, especially for the  $D$  region (see Introduction). Although serological evidence for as many as five antigenically distinct  $D$  region molecules has been described ( $D$  region of the  $H-2^d$  haplotype), gene expression data have provided evidence for only two functioning genes (11). In this report, we provide primary structural evidence for the existence of three distinct  $D$  region molecules encoded by the  $H-2^a$  haplotype. Our finding that the differences are apparently not due to carbohydrate variation implies that the  $H-2^a$  haplotype has at least three functioning  $D$  region genes.

While most H-2 class I molecules are about 80–90% homologous (Table 3; ref. 1), the  $D^a$ ,  $L^a$ , and  $R^a$  molecules are estimated to be 90–99% homologous at the protein level (Tables 2 and 3). Previously, it had been noted that the  $D^a$  and  $L^a$  molecules share an exceptionally high antigenic relatedness (19). The structural homology that our results indicate to exist between the  $D^a$  and  $L^a$  molecules, approximately 99% at the amino acid sequence level, suggests that these molecules may be the products of relatively recently duplicated genes. If this is so, it may mean that the number of H-2

Table 3. Homologies between class I molecules of the murine major histocompatibility complex

	D <sup>a</sup>	L <sup>a</sup>	R <sup>a</sup>	L <sup>d</sup>	D <sup>d</sup>	K <sup>d</sup>	D <sup>b</sup>	K <sup>b</sup>	K <sup>k</sup>
	% identical residues								
D <sup>a</sup>		100*	91	92	77	76	86	73	85
L <sup>a</sup>	55		91	93	76	78	88	76	90
R <sup>a</sup>	34	33		94	74	73	90	76	80
L <sup>d</sup>	62	57	35		78	75	90	78	79
D <sup>d</sup>	61	59	41	73		77	77	82	75
K <sup>d</sup>	63	59	37	73	73		74	75	81
D <sup>b</sup>	64	59	41	73	73	73		79	75
K <sup>b</sup>	63	58	38	73	73	73	73		77
K <sup>k</sup>	46	39	30	52	52	52	52	52	
	Number of residues compared								

Numbers above and right of the diagonal are the percent of residues that are identical. Numbers below and to the left are the numbers of residues compared. Only those positions corresponding to residue assignments in the D<sup>a</sup>, L<sup>a</sup>, and R<sup>a</sup> molecules are compared. L<sup>d</sup> is from Evans *et al.* (23) and Moore *et al.* (24); D<sup>d</sup> is from Nairn *et al.* (26); K<sup>d</sup> is from Kvist *et al.* (27); D<sup>b</sup> is from Maloy and Coligan (15); K<sup>b</sup> is from Coligan *et al.* (28); and K<sup>k</sup> is from Lillehoj and Coligan (29).

\*The D<sup>a</sup> and L<sup>a</sup> molecules are indistinguishable in the 55 residues compared. Tryptic peptide map comparisons of these two molecules suggest that they are >99% homologous at the primary structural level (see Fig. 4).

class I genes is in a dynamic state in that they undergo relatively frequent expansion or contraction. The D region molecules of the H-2<sup>a</sup> haplotype would then represent an example of a recent expansion of H-2 genes, and the *dml* and *dm2* mutants (32) could be examples of relatively recent contractions of expressed H-2 genes.

Protein structural data (9) indicate that the L<sup>d</sup> molecule is no more like D<sup>d</sup> than any other currently described H-2 molecule, suggesting that the L<sup>d</sup> gene is probably not the result of a recent duplication of the D<sup>d</sup> gene. In contrast, our results extend previous observations demonstrating a close relationship of L<sup>d</sup> to the D<sup>a</sup>, L<sup>a</sup>, and R<sup>a</sup> molecules (18, 19). This similarity suggests that H-2<sup>d</sup> could have acquired the L<sup>d</sup> gene by recombination with H-2<sup>a</sup> or another closely related haplotype. Recent data indicating that individual viral antigens can be recognized only in the context of particular class I molecules—e.g., vesicular stomatitis virus can be recognized in the context of L<sup>d</sup> but not the K<sup>d</sup>, D<sup>d</sup>, or R<sup>d</sup> molecules (33)—suggest that it may be desirable for a species to expand the number of class I transplantation genes expressed. Once expanded, polymorphism would be generated by genetic mechanisms such as point mutation and gene conversion. The D region of the H-2<sup>a</sup> haplotype may represent a relatively early stage of this process.

The authors are grateful to Frank Turner and Michael Raum for technical assistance and Ginny Shaw for editorial assistance. In addition, we thank Drs. Thomas Kindt, Keiko Ozato, Saswati Chatterjee-Hasrouni, Elliot Cowan, W. Lee Maloy, and Marie Rose van Schravendijk for valuable suggestions and criticism. E.P.L. was supported by a fellowship from the American Cancer Society. T.H.H. was supported in part by National Institutes of Health Grant AI19687.

- Kimball, E. S. & Coligan, J. E. (1983) *Contemp. Top. Mol. Immunol.* **9**, 1–63.
- Klein, J. & Shreffler, D. C. (1971) *Transplant. Rev.* **6**, 3–29.
- Snell, G. D., Cherry, M. & Démant, P. (1971) *Transplant. Proc.* **3**, 183–186.
- Lengerova, A., Peknicova, J. & Pokorna, Z. (1974) *J. Immunogenet.* **1**, 239–248.
- Hauptfeld, V. & Klein, J. (1975) *J. Exp. Med.* **142**, 288–298.
- Cullen, S. E., Schwartz, B. D., Nathenson, S. G. & Cherry, M. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1394–1397.
- Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. & Sachs, D. H. (1981) *J. Immunol.* **126**, 1713–1716.
- Démant, P., Iványi, D., Oudshoorn-Snoek, M., Calafat, J. & Roos, M. H. (1981) *Immunol. Rev.* **60**, 5–22.
- Coligan, J. E., Kindt, T. J., Nairn, R., Nathenson, S. G., Sachs, D. H. & Hansen, T. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1134–1138.
- Margulies, D. H., Evans, G. A., Ozato, K., Camerini-Otero, R. D., Tanaka, K., Appella, E. & Seidman, J. G. (1983) *J. Immunol.* **130**, 463–470.
- Goodenow, R. S., McMillan, M., Nicolson, J., Sher, B. T., Eakle, K., Davidson, N. & Hood, L. (1982) *Nature (London)* **300**, 231–237.
- Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28**, 489–498.
- Melino, M. R., Nichols, E. A., Strausser, H. R. & Hansen, T. H. (1982) *J. Immunol.* **129**, 222–226.
- Iványi, D. & Démant, P. (1983) *J. Immunol.* **131**, 1080–1084.
- Maloy, W. L. & Coligan, J. E. (1982) *Immunogenetics* **16**, 11–22.
- Rose, S. M., Hansen, T. H. & Cullen, S. E. (1980) *J. Immunol.* **125**, 2044–2050.
- Levy, R. B. & Hansen, T. H. (1980) *Immunogenetics* **10**, 7–17.
- Hansen, T. H., Iványi, P., Levy, R. B. & Sachs, D. H. (1979) *Transplantation* **28**, 339–342.
- Ozato, K., Hansen, T. H. & Sachs, D. H. (1979) *J. Immunol.* **125**, 2473–2477.
- Cowan, E. P., Cummings, R. D., Schwartz, B. D. & Cullen, S. E. (1982) *J. Biol. Chem.* **257**, 11241–11248.
- Coligan, J. E. & Kindt, T. J. (1981) *J. Immunol. Methods* **47**, 1–11.
- Brown, J. L., Kato, K., Silver, J. & Nathenson, S. G. (1974) *Biochemistry* **13**, 3174–3178.
- Evans, G. A., Margulies, D. H., Camerini-Otero, D., Ozato, K. & Seidman, J. G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1994–1998.
- Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. A. & Hood, L. (1982) *Science* **215**, 679–682.
- IUPAC-IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem.* **243**, 3557–3559.
- Nairn, R., Nathenson, S. G. & Coligan, J. E. (1981) *Biochemistry* **20**, 4739–4745.
- Kvist, S., Roberts, L. & Dobberstein, B. (1983) *EMBO J.* **2**, 245–254.
- Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. & Nathenson, S. G. (1981) *Nature (London)* **291**, 35–39.
- Lillehoj, E. P. & Coligan, J. E. (1984) *Mol. Immunol.*, in press.
- Goetze, D., Nadeau, J., Wakeland, E. K., Berry, R. J., Bonhomme, F., Egorov, I. K., Hjorth, J. P., Hoogstraal, H., Vives, J., Winking, H. & Klein, J. (1980) *J. Immunol.* **124**, 2675–2681.
- Zinknagel, R. M. & Doherty, P. C. (1979) *Adv. Immunol.* **27**, 51–177.
- Nairn, R., Yamaga, K. & Nathenson, S. G. (1980) *Annu. Rev. Genet.* **14**, 241–277.
- Ciavarrá, R. & Forman, J. (1982) *J. Exp. Med.* **156**, 778–790.