H-2D^d Exploits a Four Residue Peptide Binding Motif

By Maripat Corr, Lisa F. Boyd, Eduardo A. Padlan,* and David H. Margulies

*From the Molecular Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, and the *Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892*

Summary

We have characterized the amino acid sequences of over 20 endogenous peptides bound by a soluble analog of H-2D^d, H-2D^d_s. Synthetic analogs corresponding to self, viral, tumor, or motif peptides were then tested for their ability to bind to H-2D^d by serologic epitope induction assays using both purified soluble protein and cell surface H-2D^d. The dominant primary sequence motif included glycine at position 2, proline at position 3, and a hydrophobic COOH terminus: leucine, isoleucine, or phenylalanine at position 9 or 10. Ancillary support for high affinity binding was contributed by a positively charged residue at position 5. Three-dimensional computer models of H-2D^d_s/peptide complexes, based on the crystallographic structure of the human HLA-B27/peptide complex, showed that the basic residue at position 5 was in position to form a salt bridge with aspartic acid at position 156, a polymorphic residue of the H-2D^d heavy (H) chain. Analysis of 28 such models, including 17 based on nonamer self-peptides, revealed considerable variation in the structure of the major histocompatibility complex (MHC) surrounding peptide residue 1, depending on the size and charge of the side chain. Interactions between the side chains of peptide residues 5 and 7, and 6 and 8 commonly occurred. Those peptide positions with limited sequence variability and least solvent accessibility may satisfy structural requirements for high affinity binding of the peptide to the MHC class I H chain, whereas the highly variable positions of the peptide (such as positions 4, 6, and 8) may contribute more to the T cell epitopes.

The class I MHC encoded antigens are expressed at the cell surface as ternary structures in which a 46-kD H chain is noncovalently bound to both a small peptide fragment 8–10 amino acids long and a 12-kD L chain, β_2-microglobulin (β_2m)[1] (1, 2). The heterodimeric H/L chain complexes bind to proteolytic fragments from the cell's endogenous pool or to those resulting from infection with an intracellular pathogen, or cellular dysregulation. At the cell membrane this ternary complex can then interact with an appropriate TCR to initiate activation. The crystal structures of the human MHC class I antigens HLA-A2 (3–5), HLA-Aw68 (6), HLA-B27 (7, 8), and the murine H-2K^b (9–11) reveal a molecular anatomy that reflects the function of these molecules. The H chain associates with the peptide in a groove between the two helices of the α_1 and α_2 domains supported by a floor of β-pleated sheet. Six pockets (termed A through F) open onto the main peptide binding groove which can accommodate peptide termini and amino acid side chains having appropriate size, polarity, and charge (5, 6).

The hallmark of MHC class I molecules is their polymorphism, which dictates not only their ability to interact with a variety of TCR, but their capacity to bind to a variety of peptides with different primary sequences. The definition of sequence motifs characteristic of certain MHC molecules has been approached in several ways. Preferred amino acids at particular positions were determined by automated Edman degradation of a pool of peptides that was isolated by acid elution and reverse phase (RP) HPLC separation (2, 12, 13). Subsequently, individual or limited pools of peptides have been identified by the sequencing of single HPLC fractions using classical chemical, radiochemical, or mass spectrometry techniques (14–17).

A full understanding of the influence of MHC polymorphism on both peptide and T cell interactions awaits the accumulation of not only a large data set of primary sequences of MHC molecules, but also a full library of bound peptides and structures of MHC/peptide complexes. Motifs of peptides restricted to additional MHC molecules should be elucidated, as other molecules may use different polymorphic regions in varying the repertoire of selected peptides. Likewise, more structures of peptide/MHC complexes need to

[1] *Abbreviations used in this paper:* β_2m; β_2-microglobulin; rms, root mean square; RU, resonance unit.

be determined to extend our comprehension of the amino acid sequence and the structural basis of antigen selection by MHC class I molecules. One approach is to evaluate a number of homogenous MHC/peptide structures crystallographically. Another complementary and pragmatic approach is to identify a large number of cognate peptides for many different class I molecules and to utilize modern homology modeling methods to describe the likely structures of the MHC/peptide complexes. As part of our continuing study of MHC/peptide interactions, we present the characterization of the H-2D^d peptide motif, the binding of a large set of peptides, and the construction and analysis of three-dimensional models derived from them. The H-2D^d peptide motif is unique, based on four amino acid residues, and reflects several MHC amino acid sequence polymorphisms.

Materials and Methods

Protein and Synthetic Peptides. The soluble analog of H-2D^d, H-2D^d_s, is a chimeric molecule expressed as the α_1, α_2, and α_3 domains of H-2D^d with the 27 residues of the COOH terminus of Q10^b (18, 19). Protein was immunoaffinity purified from transfectant L cell culture supernatants on a column of Sepharose 4B coupled to the mAbs 34-5-8S, 34-4-20S, 34-1-2S, and 34-2-12S (20), and concentrated by ultracentrifugation (Centricon 30; Amicon, Beverly, MA). Synthetic peptides (provided by J. Coligan, Biological Resources Branch [BRB], National Institute of Allergy and Infectious Diseases, Frederick, MD) were synthesized using standard Fmoc chemistry on peptide synthesizers and repurified by RP-HPLC. Peptide purity was >90%. The peptides were dissolved in dH_2O, except for those representing fraction 46 which were insoluble in water and were reconstituted in 20% DMSO.

Peptide Elution and Sequence Analysis. 500 μg purified H-2D^d_s were brought to 0.1% TFA at ambient temperature and separated by RP-HPLC using a Vydac C18 column (Separations Group, Hesperia, CA) (4.6 × 250 mm, 5-μM particle size). The sample was loaded in 0.1% TFA in dH_2O (buffer A) for 15 min and then eluted with a 0–60% linear gradient of 0.1% TFA in acetonitrile (buffer B) over 60 min. After the gradient, the remaining material was eluted for 10 min with 80% buffer B. The flow rate was 1 ml/min and the absorbance was monitored at both 215 and 280 nm. 1-ml fractions were collected, dried by vacuum centrifugation, and recovered fractions 30–53 were sequenced by automated Edman degradation on a sequenator (model 470A; Applied Biosystems, Inc., Foster City, CA) standardized to a repetitive yield of 94%. Residues were identified by amino acid analysis of a PTH analyzer (model 120A; Applied Biosystems, Inc.). All sequencing and amino acid analysis was performed by the BRB. Homologous sequences were identified by searching the National Biomedical Research Foundation (NBRF) data bank using MacVector v.3.5 (IBI, New Haven, CT) and the nonredundant data bank at the National Library of Medicine (Bethesda, MD) which includes: SWISS-PROT, PIR, and GenPept using BLAST (21).

Serological Detection Using Surface Plasmon Resonance. The binding of the α_2 domain–specific mAb 34-5-8S (22) to various H-2D^d_s/peptide complexes was assessed by surface plasmon resonance using the BIAcore™ (Pharmacia, Piscataway, NJ). The carboxyl groups of the carboxymethylated dextran matrix of the gold surface Sensor chip CM5™ (Pharmacia Biosensor AB) were chemically activated during a 6-min continuous flow (5 μl/min) exposure to a solution of 50 mM *N*-hydroxysuccinimide and 200 mM *N*-ethyl-*N*′-(3-dimethylaminopropyl) carbodiimide hydrochloride, and then covalently linked to the free amines of 34-5-8S (0.02 mg/ml in 10 mM NaAcetate, pH 4.5) for 6 min. The remaining reactive groups were then blocked with 1 M ethanolamine hydrochloride, pH 8.5. To enrich for H-2D^d_s molecules lacking preassociated peptides, protein purified as described above was passed over a 34-5-8S coupled Sepharose 4B column and the flowthrough was concentrated by Centricon 30 ultracentrifugation. 50 μl of 10 μg/ml 34-5-8S-depleted H-2D^d_s protein was added to 50 μl of varying amounts of peptide in 10 mM Hepes (pH 7.5), 150 mM NaCl, 3.4 mM EDTA, 0.05% Tween 20 (HBS) or HBS alone at room temperature to achieve equilibrium (at least 15 h). 30 μl aliquots were analyzed for direct binding to the immmobilized 34-5-8S on the matrix surface by plasmon resonance. The flow rate was constant at 5 μl/min and the Ab surface was regenerated with a 1-min wash of 5 mM phosphoric acid. Control experiments indicated that high concentrations (>200 μM) of free peptide failed to bind the Ab-coupled surface. 1,000 resonance units (RU) correspond to a surface concentration of ~1 ng/mm^2.

Serological Detection Using Immunofluorescence and Flow Cytometry. LKD8 cells express low cell surface levels of H-2D^d at 37°C but can be stabilized in their expression of class I molecules in the presence of exogenous peptide (23, 24). These cells were propagated in DMEM supplemented with 10% FCS, 2 mM glutamine, nonessential amino acids, 50 μg/ml gentamicin, and 5 × 10^5 M 2-ME. The cells were washed with DMEM 10% FCS after trypsinization and a second time with medium lacking serum. Cells (10^6 per well in 24-well plate [Costar Corp., Cambridge, MA]) were incubated overnight at 37°C in the presence of graded amounts of synthetic peptide and 2 μg/ml human β_2m (Calbiochem, San Diego, CA) in NCTC-109 (GIBCO BRL, Gaithersburg, MD) containing 0.5% BSA. The cells were stained with 34-5-8S (2.7 μg/ml in DMEM) followed by FITC-conjugated goat F(ab′)$_2$ anti–mouse Ig (Cappel Laboratories, Organon Teknika, Durham, NC) on ice. Propidium iodide was added, and viable cells were analyzed on a FACScan® (Becton Dickinson & Co., Mountain View, CA).

Analysis of Sequence Variability. Peptide sequences were analyzed for sequence variability according to the algorithm of Wu and Kabat (25) and for structural variability according to Padlan (26) using the amino acid dissimilarity values of Grantham (27), as described previously (17).

Molecular Modeling. Three-dimensional coordinates for the HLA-B27/peptide complex based the 2.1 Å crystal were kindly provided by D. Madden and D. Wiley (Harvard University, Cambridge, MA) (8). The amino acid structures corresponding to the H-2D^d sequence for residues 1 to 276, and for murine β_2m (28, 29) were substituted at the positions of the corresponding residues in the HLA-B27 structure using the computer graphics program QUANTA 3.3 (Molecular Simulations, Inc., Hudson, MA) implemented on a work station (INDIGO; Silicon Graphics, Mountain View, CA). The alpha carbons, C_α, of the substituted amino acids were superimposed on those of the corresponding residues of HLA-B27 and the side chains were maximally overlapped. HLA-B27 and H-2D^d have 74% sequence identity over the 276 amino acid residues for which three-dimensional structure is available. There was no need to introduce gaps in either sequence to maximize the homology alignment. Similarly, antigenic peptide residues were substituted onto the corresponding positions of the HLA-B27 bound peptide. Prolines were substituted without difficulty in the *trans* configuration in all cases. After the substitution, the amino acid side chains were regularized to eliminate geometric incompatibilities, and the resulting coordinates were subjected first to 100 steps of minimization using the Adopted Basis Newton Raphson method

in QUANTA (30). Using X-PLOR 3.0 (31) these structures were subjected to simulated annealing (32, 33) by heating to 1,000°K and slow cooling to 300°K. The final cooled structure was subjected to 120 steps of minimization by the method of Powell (34) using X-PLOR. The only constraints invoked in the simulated annealing were the disulfide bonds joining residues 101 to 164 and 203 to 259 of the H chain and residues 25 to 80 of the β_2m L chain. Final coordinates for each structure were compared for root mean square (rms) deviations, molecular surface exposure, and atomic contacts. Connolly surfaces (35) were calculated for a water probe of 1.7 Å radius. The validity of the simulated annealing procedure in building these models is emphasized by the demonstration that a model of the HLA-A2 molecule built by the same procedure compared favorably to the crystal structure of HLA-A2 with rms deviation of all 276 C_α atoms of <1.5 Å. Atomic coordinates for the models described here will be made available on request.

Results

Identification of the Sequence Motif for H-2D^d Bound Peptides. Our laboratory has previously investigated the stimulation of both alloreactive and peptide-specific T cells using purified preparations of a secreted soluble analog of H-2D^d (19, 36, 37). As a step in the detailed characterization of the molecular structures recognized by H-2D^d–restricted T cells, we analyzed the peptides that copurify through immunoaffinity isolation with H-2D^d_s. Approximately half a milligram of soluble molecules, purified from transfected L cell supernatants as described in Materials and Methods, was denatured with TFA and separated by RP-HPLC (Fig. 1). Fractions containing small pools of peptides were collected and sequenced by automated Edman degradation (Table 1). The sequences were typically 9 or 10 amino acids in length, consistent with the findings of several other studies evaluating MHC class I/bound peptides from detergent extracts of cells (2, 16, 38–41), or from engineered soluble H-2L^d (17).

Although two fractions (30 and 31) yielded no sequence,

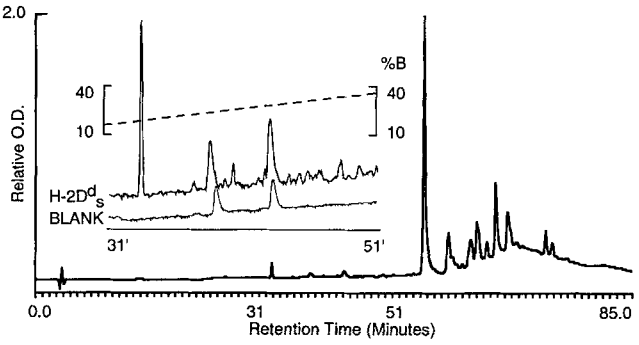


Figure 1. RP-HPLC profile of H-2D^d_s molecules. H-2D^d_s molecules were immunoaffinity purified, denatured with 0.1% TFA, and loaded onto a C18 RP-HPLC column as described in Materials and Methods. Peptides were then eluted with a 0–60% linear acetonitrile gradient between 15 and 75 min. The region of peptide elution is magnified in the inset (*bottom*, blank injection; *top*, H-2D^d_s/peptide separation). Absorbance shown is at 215 nm.

and the repetitive yields from fraction 32 suggested that its sequence was incomplete, reliable data were obtained from fractions 33–53. On inspection, nearly all peptide sequences contained glycine at position 2, proline at position 3, and a hydrophobic residue (leucine, isoleucine, or phenylalanine) at the COOH terminus. An additional conserved residue was frequently identified at position 5 where positively charged arginine or lysine predominated. (In fraction 49, the basic residue was identified at position 6, after glycine). This H-2D^d motif (glycine at position 2, proline at position 3, positively charged residue at position 5, and a hydrophobic COOH-terminal residue) was also present in two previously described H-2D^d restricted peptides: the HIV-IIIB envelope glycoprotein p18-I-10 (42) and the tum$^-$ P35B peptide (43, 44) (Table 1).

To identify possible sources for the copurified peptides, several databases were scanned for homologous sequences, as described in Materials and Methods. The best protein matches and fractions of homology are indicated in Table 1. Although no perfect matches to known proteins were made, the search detected proteins derived from a wide variety of cellular compartments. Homology to serum-derived proteins was not observed, consistent with the view that the copurified peptides derive from the cell that synthesizes the MHC molecule.

Synthetic analogs were prepared based on the sequences of those peptides that were eluted from H-2D^d_s (Table 2). In cases in which the amino acid assignment could not be made with certainty, an alanine residue was incorporated at those positions, except for alternate COOH-terminal residues in which case a leucine or isoleucine was preferred. For some of the sequenced peptides there were two possible terminal residues at positions 9 and 10, either of which was consistent with the motif. In those cases both the decamer and the nonamer truncation were made. In our peptide binding studies, we also used the originally described H-2D^d–restricted quindecamer peptide from the HIV-IIIB gp120 glycoprotein RIQRGPGRAFVTIGK (p18) (45), and the optimal decamer RGPGRAFVTI (p18-I-10), as well as the undecamer tum$^-$ H-2D^d–restricted peptide NGPPHSNNFGY (tum$^-$ P35B) (43), and a nonamer truncation that would conform to the elucidated motif NGPPHSNNF (tum$^-$ P35B-F9). To evaluate the effect of the anchor residues, four motif peptides were analyzed: AGPAAAAAL (pD239), AGPARAAAL (pD2359), GPGGGGGGL (pL29G), and GPPPPPPL (pL28P). The length of the different spacing elements in the motif peptides was based in part on the previous findings of Maryanski et al. (46), who demonstrated that a proline spacer one residue shorter than expected was the most effective in generating a motif peptide for binding to H-2K^d.

Each of the synthetic peptides was analyzed by RP-HPLC as well, to confirm that it was representative of the sequenced peptides. Of the synthetic peptides, pD36, pD38A, pD39, pD40, pD41, pD44, pD45, pD47, pD48A and B, and pD53B had the same retention times as those fractions from which their sequences were obtained. The synthetic self-peptides, which had alanine inserted for uncharacterized residues, did not have the same retention time as the original sequenced

Table 1. *Sequences of Eluted Self-peptides*

Fraction	Sequence 1	2	3	4	5	6	7	8	9	10	Homology
30	No sequence										
31	No sequence										
32	s	**E**	Q	D	L	N	f				Elongation factor rat (6/7)
33	s	X	H	K	**E**	Q	**P**	**A**	**T**		Transforming protein spi-1 human (7/8)
34	s	X	**P**	**K**	**T**	D	X	Q	T	L	Insulin receptor precursor murine (7/8)
35	K	**G**	P	I	T	V	Q	I			Fiber protein adenovirus2 (6/8)
36	V	**G**	**P**	**Q**	**K**	N	**E**	**N**	**L**		Transferrin receptor rat (6/9)
37	s	**G**	**P**	R	K	X	I	X	L		mRNA CD40 mouse (6/7)
38	a	**G**	P	D	R	T	E	K	X	L	Fibronectin precursor human (6/9)
39	k	**G**	P	D	K	G	N	E	F		Metalloproteinase 2 inhibitor mouse (7/9)
40	i	**G**	**P**	E	R	G	H	N	L		Hypoxanthine phosphoribosyltransferase human (7/9)
41	D	**G**	P	V	R	E	H	N	L		Urease *Canavalia ensiformis* (7/9)
42	K	**G**	**P**	E	R	X	**N**	**G**	L		Pyruvate kinase rat (6/9)
43	s	**G**	P	E	R	G	E	K	L		Proliferating cell nucleolar antigen P40 human (7/9)
44	D	**G**	**P**	V	**R**	G	I	S	I		Ribosomal protein S17 rat (8/9)
45	N	G	P	Q	R	I	Y	N	L		Alpha 1 macroglobulin rat (6/9)
46	s	**G**	P	v	a	L	V	N	F	I	Superoxide dismutase horse (6/10)
47	I	**G**	P	N	R	a	f	**N**	**F**		HIV 1 envelope protein (6/9)
48	S	G	P	E	R	I	L	S	X	Y	Heterogeneous nuclear ribonucleoprotein complex K (8/9)
49	V	**G**	**P**	S	G	K	y	F	I	**L**	T complex associated testes peptide mouse (7/10)
50	F	G	P	y	k	L	N	R	**L**		Feline leukemia virus envelope polyprotein (8/9)
51	F	**G**	**P**	I	K	F	n	V	L	T	Ran *Canis familiaris* (7/10)
52	A	G	P	d	R	f	I	X	X	**M**	Amidotransferase II *Streptomyces griseus* (6/8)
53	F	**G**	**P**	y	R	F	Y	**V**	**L**	**T**	LasA protein *Pseudomonas aeruginosa* (7/10)
Known H-2D^d–restricted antigens											
	N	G	P	P	H	S	N	N	F	G	Y tum-P35B
	R	G	P	G	R	A	F	V	T	I	p18-I-10

The material in recovered HPLC fractions 30–53 was sequenced by automated Edman degradation. The typeface classifications for amino acids reported include: boldface for a unique residue, plain uppercase for a quantitatively dominant residue, and plain lowercase for a residue with the least assurance due to low quantity or heterogeneity at that position. Positions for which an assignment could not be confidently made are designated with an X. Homologous proteins were identified by searching multiple data banks as described in Materials and Methods and are followed by the fraction of exact amino acid matches with the identified residues in the sequenced fraction. Accession numbers for best matches were: fraction 32, A25440 (62); fraction 33, S10892 (63); fraction 34, J05149 (64); fraction 35, A03845 (65); fraction 36, A34549 (66); fraction 37, M83312 (67); fraction 38, X02761 (68); fraction 39, M82858 (69); and fraction 40, M26434 (70); fraction 41, M65260 (71); fraction 42, B26186 (72); fraction 43, X15610 (73); fraction 44, A24028 (74); fraction 45, M77183 (75); fraction 46, A00515 (76); fraction 47, M61648 (77); fraction 48, S74678 (78); fraction 49, M28821 (79); fraction 50, A24300 (80); fraction 51, Z11922 (81); fraction 52, S18620 (82); and fraction 53, M20982 (83).

fraction, suggesting that other amino acids were at these positions, possibly cysteine residues which cannot be identified by this sequencing technique.

Synthetic Peptides Induce Reactivity with an Epitope-specific mAb As we had characterized sequences which putatively represented peptides from denatured H-2D^d_s molecules, we wished to confirm their ability to bind to H-2D^d_s. In this study we used an indirect assay to evaluate this binding: the addition of exogenous synthetic peptide to H-2D^d_s to induce the formation of the 34-5-8S epitope, which is specific for the H-2D^d α_2 domain and conformationally dependent on the presence of bound peptide (23, 47). A population of H-2D^d_s molecules that was relatively free from preassociated peptides (prepared as described in Materials and Methods) was incubated with titrated amounts of synthetic peptide in solution and then assayed for the ability to bind to immobilized 34-5-8S Ab on a solid phase by real time surface plasmon resonance. RU are directly proportional to the amount of

Table 2. *Sequences of Synthetic Self-peptides*

Name	Sequence 1	2	3	4	5	6	7	8	9	10
pD34	S	G	P	K	T	D	A	Q	T	L
pD35	K	G	P	I	T	V	Q	I		
pD36	V	G	P	Q	K	N	E	N	L	
pD37	S	G	P	R	K	A	I	A	L	
pD38A	A	G	P	D	R	T	E	K	L	
pD38B	A	G	P	D	R	T	E	K	A	L
pD39	K	G	P	D	K	G	N	E	F	
pD40	I	G	P	E	R	G	H	N	L	
pD41	D	G	P	V	R	E	H	N	L	
pD42	K	G	P	E	R	A	N	G	L	
pD43	S	G	P	E	R	G	E	K	L	
pD44	D	G	P	V	R	G	I	S	I	
pD45	N	G	P	Q	R	I	Y	N	L	
pD46A	S	G	P	V	A	L	V	N	F	I
pD46B	S	G	P	V	A	L	V	N	F	
pD47	I	G	P	N	R	A	F	N	F	
pD48A	S	G	P	E	R	I	L	S	I	
pD48B	S	G	P	E	R	I	L	S	L	
pD48C	S	G	P	E	R	I	L	S	A	Y
pD49A	V	G	P	S	G	K	Y	F	I	
pD49B	V	G	P	S	G	K	Y	F	I	L
pD50	F	G	P	Y	K	L	N	R	L	
pD51	F	G	P	I	K	F	N	V	L	
pD53A	F	G	P	Y	R	F	Y	V	L	
pD53B	F	G	P	Y	R	F	Y	V	L	T
pD2359	A	G	P	A	R	A	A	A	L	
pD239	A	G	P	A	A	A	A	A	L	
pD29G		G	P	G	G	G	G	G	L	
pL28P		G	P	P	P	P	P	P	L	
p18	R I Q	R	G	P	G	R	A	F	V	T I G I
p18-I-10		R	G	P	G	R	A	F	V	T I
tum⁻ 35B		N	G	P	P	H	S	N	N	F G Y
tum⁻ 35B-F9		N	G	P	P	H	S	N	N	F

A panel of synthetic peptides was patterned after the sequences identified for the acid-eluted H-2D^d_s self-peptides with alanine residues inserted at positions where the sequences was not clearly defined. Each synthetic peptide is designated by the HPLC fraction from which the sequnce was originally identified. Peptides pD239, pD2359, pL29G, and pL28P, each of which contain elements of the H-2D^d peptide motif, separated by repetive single residue spacers, were also synthesized.

mass bound, and this assay provides results that are similar to those of a conventional ELISA. These results are plotted in Fig. 2, *A–H*. All nonamer peptides tested, the octamer pD35 (Fig. 2 *G*), several decamers (p18-I-10, all panels; pD38B, Fig. 2 *C*; pD49B, Fig. 2 *D*; pD48C, Fig. 2 *E*; pD34, Fig. 2 *G*; pD53B, and pD46A, Fig. 2 *F*), and the tum⁻ P35B undecamer (Fig. 2 *A*) effectively induced the 34-5-8S epitope.

Two motif nonamer peptides were also examined, pD239 (containing motif residues at positions, 2, 3, and 9) and pD2359 (containing arginine additionally at position 5) (Fig. 2 *B*). The motif peptide with arginine at position 5 (pD2359) induced the 34-5-8S epitope more effectively than the motif peptide with alanine at that position (pD239), suggesting that the positively charged residue at position 5 contributed to the interaction with the H chain. The only endogenous peptide sequence with an alanine at position 5 was fraction 46. Two peptide analogs representing this sequence were made, a nonamer and a decamer, both of which required an organic solvent for solubilization. The requirement for greater apparent concentrations of these peptides to induce the epitope (Fig. 2 *F*) most likely represented the relative insolubility of these peptides in aqueous solution, masking their true affinity for H-2D^d_s. Both peptides which had threonine at position 5, pD34 and pD35, efficiently induced the 34-5-8S epitope.

Different peptides were made representing other fractions that had multiple possible overlapping COOH-terminal residues: fractions 38, 48, 49, and 53. In each case the nonameric peptides had higher apparent binding affinities for the H chain than the related decamers, but both peptides from each set were able to bind. Thus, all the synthetic peptides corresponding to the endogenous peptide sequences bound H-2D^d_s, as measured by this epitope induction assay. In the case of the previously described H-2D^d–restricted peptide tum⁻ P35B (Fig. 2 *A*), the undecamer peptide weakly induced the epitope with a half maximal point about two orders of magnitude less than the nonamer truncation that terminated in phenylalanine. Similarly, the P18 quindecamer did not induce the epitope, but the nested decamer accomplished this readily (Fig. 2 *C*). Thus, all the synthetic peptides corresponding to the endogenous peptide sequences bound H-2D^d_s, as measured by this epitope induction assay.

To demonstrate that the binding of these endogenous peptide analogs to H-2D^d_s was not unique to the soluble protein system, we evaluated their ability to bind to membrane-associated molecules. We used an H-2D^d–transfected embryonic cell line, LKD8, which expresses low levels of H-2D^d at 37°C (23). Surface expression of H-2D^d on this cell line can be stabilized by the addition of exogenous cognate peptides in the presence of β_2m, as the cells have an inherent defect in efficiently providing peptides to nascent class I proteins before their transit to the cell membrane (24). The cells were incubated with synthetic peptides (see Materials and Methods) and then assayed for the presence of 34-5-8S epitope by flow cytometry (see Fig. 3).

Fig. 3 *A* depicts representative examples of fluorescence histograms with increasing peptide concentrations in the foreground. In rank order, the most effective peptide in inducing the 34-5-8S epitope was P18-I-10 followed by the nonamer truncated tum⁻ P35B-F9, pD2359, pD39, pD239, and tum⁻ P35B. The results of the analysis of all the peptides are plotted in Fig. 3 *B*. As with the epitope induction assay employing purified H-2D^d_s protein described above, the poor

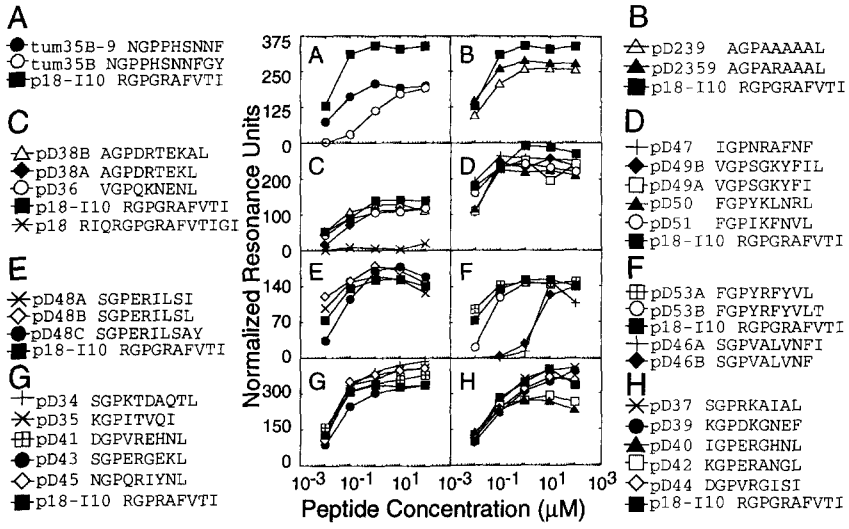


Figure 2. Peptides generate the α_2 domain conformationally sensitive 34-5-8S epitope on H-2D^d molecules. Varying concentrations of synthetic peptides as indicated were added in solution to purified H-2D^d. The subsequent reactivity to the immobilized Ab, 34-5-8S, was assayed using real time plasmon surface resonance as described in Materials and Methods. Normalized RU represent the difference in RU in the presence and absence of added peptide. Under these conditions, free p18-I-10 or pD2359 at concentrations as high as 222 μM gave a signal of <7 RU.

epitope stabilization by the peptides pD46A and B may reflect their relative aqueous insolubility and not their true ability to associate with the class I chain. In all other cases in which nonamers were compared to decamers of the same fraction, the nonameric peptide was more effective at preserving the epitope than the decameric peptide (pD38, pD48, pD49, and pD53). Nevertheless, the p18-I-10 decamer was the most effective peptide studied.

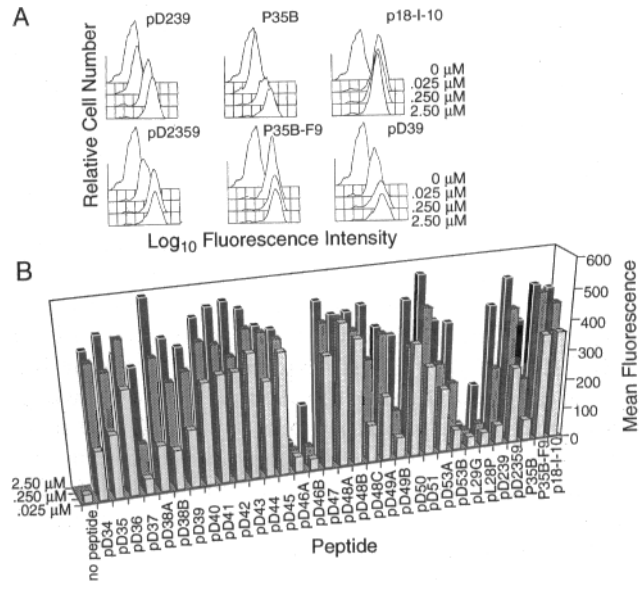


Figure 3. Peptides stabilize the α_2 domain conformationally sensitive 34-5-8S epitope on H-2D^d–expressing embryonic cells. LKD8 cells were incubated in serum-free medium in the presence of titrated amounts of peptide (2.5, 0.25, and 0.025 μM) and human β_2m overnight and then analyzed for reactivity to 34-5-8S by indirect immunofluorescence. (*A*) The induction of the 34-5-8S epitope with pD239 and pD2359, p18, p18-I-10, tum$^-$ P35B, and tum$^-$ P35B-9 with histogram plots. (*B*) Summary plot of mean fluorescence by peptide vs peptide concentration for all peptides tested.

The synthetic peptides induced the expression of H-2D^d to varying degrees with the exception of the polyglycine-substituted motif peptide (pL29G), which did not stabilize the epitope at all (Fig. 3 *B*). Of the motif peptides, the alanine-substituted spacing residues were apparently better than the proline or glycine residues at supporting binding by the anchor residues (compare p239, pD2359, pL29G, and pL28P). The most effective motif peptide at stabilizing the 34-5-8S epitope was pD2359 with an arginine at position 5, suggesting that although the positive charge is not essential for peptide binding, it promotes a stronger interaction with the H chain. Qualitatively, the cell surface epitope induction assay was in agreement with the assay using soluble, purified H-2D^d.

Sequence and Structural Variability. Although inspection of the sequences of the endogenous peptides suggested that the gly-pro sequence at residues 2 and 3 and the hydrophobic residue at the COOH terminus were conserved, we also elected to evaluate the sequences obtained by objective sequence and structural variability criteria. The eluted peptides that were demonstrated to be capable of binding to H-2D^d were grouped into decamers and nonamers. Omitting unassigned residues, these sequences were then analyzed for sequence and structural variability (Fig. 4) (25, 26). Since the Wu/Kabat (25) sequence variability provided an indication of how many different amino acids were observed at a given position relative to the frequency of the most common amino acid at that position, this calculation treated both conservative and radical substitutions equivalently. The structural variability algorithm imposed a bias dependent on the degree of chemical dissimilarity of the observed substitutions. The sequence and structural variabilities correlated well for the nonameric sequences: pD36, pD37, pD39, pD40, pD41, pD42, pD43, pD44, pD45, pD47, pD50, and pD51. Notably, positions 2, 3, 5, and 9 were most restricted in the observed variability, and the most varied residue was at position 1 (Fig. 4, *A* and *C*). Fewer decamer sequences were available for this analysis: pD38, pD46, pD48, pD49, pD53, and p18-I-10. Both the sequence and structural variabilities were limited at positions

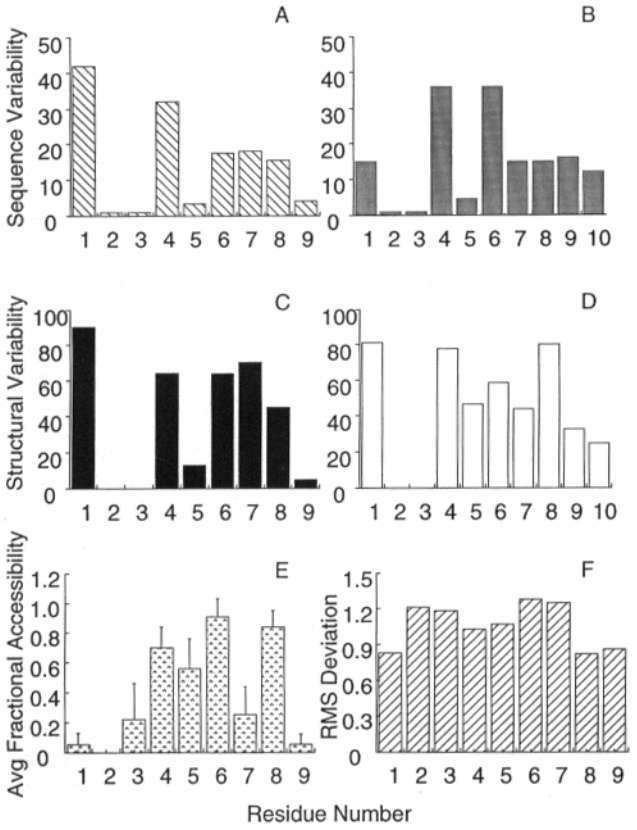


Figure 4. Sequence and structural variability of 12 nonamer and 6 decamer H-2D^d–binding peptides. Nonamer sequences of peptides pD36, pD37, pD39, pD40, pD41, pD42, pD43, pD44, pD45, pD47, pD50, and pD51 were compared according to the methods of Wu and Kabat (25) (*A*), and Padlan (26) (*C*), as were decamer sequences of peptides pD38, pD46, pD48, pD49, pD53, and p18-I-10 (*B* and *D*, respectively). The fractional accessibility of each amino acid residue, defined as the amount of surface area that is accessible to a solvent probe in the three-dimensional structure divided by the surface area that would be exposed if the residue (*X*) were in an isolated Gly-X-Gly tripeptide with the same backbone configuration as the corresponding tripeptide in the structure, was computed according to Padlan (61) (*E*). Deviation of the C_α position from the average was calculated (*F*).

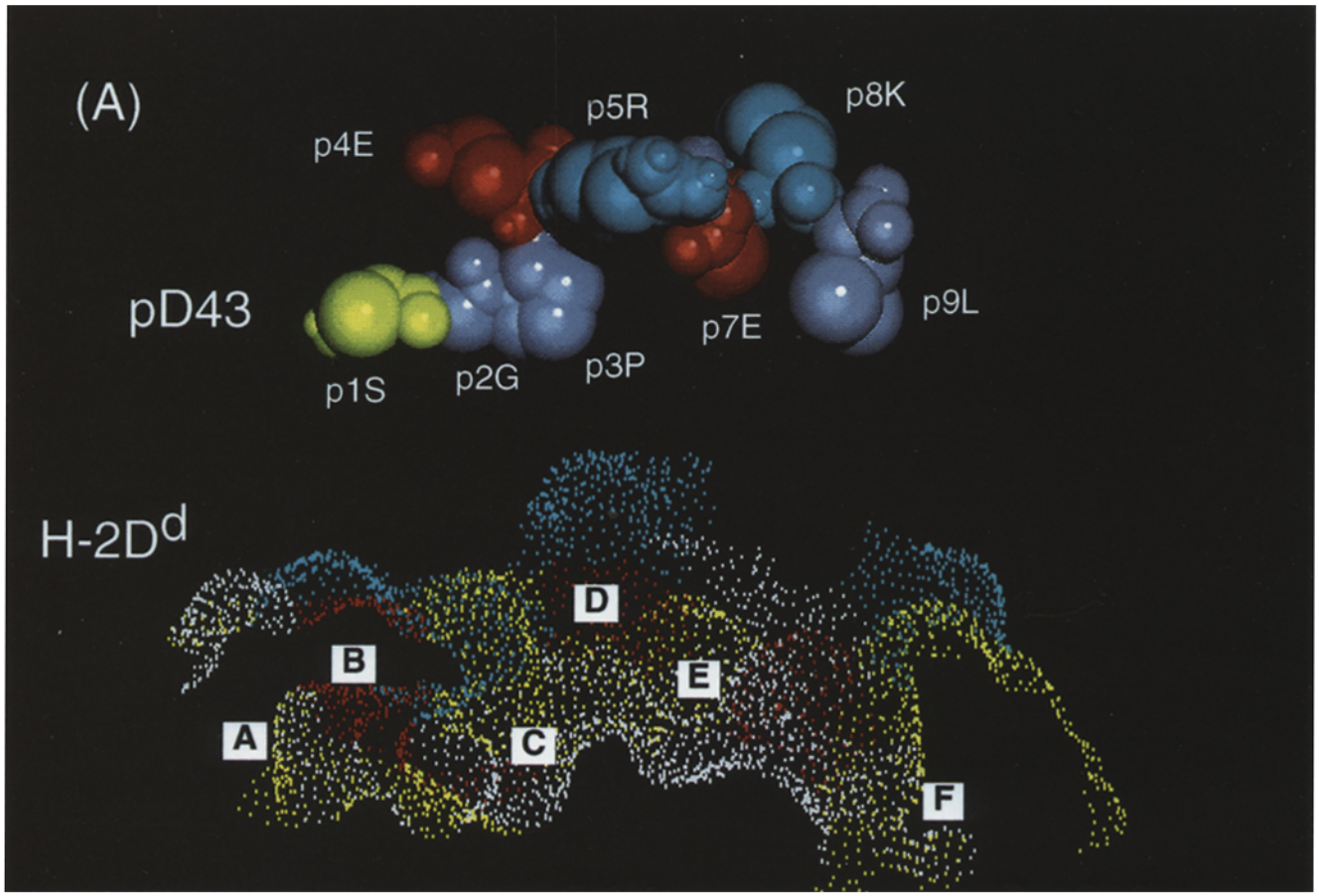
2, 3, and 10 (Fig. 4, *B* and *D*). There appeared to be less biochemical restriction at position 5 than in the nonameric sequences, which may represent flexibility in the middle of bound peptides, allowing the binding cleft to accommodate longer peptides.

Construction and Analysis of Molecular Models of H-2D^d/Peptide Complexes. To examine the structural basis of the interaction of the H-2D^d molecule with self, viral, and tumor-specific peptides, we built three-dimensional models of the H-2D^d/peptide complexes based on the x-ray crystallographic structure of the human MHC class I molecule HLA-B27/peptide complex (7). The atomic coordinates of this structure provided a template for modeling nonamer peptides with motifs suggestive of interactions in the B and F pockets of the binding groove, since HLA-B27 revealed anchors at peptide positions 2 and 9. Molecular models were constructed for 17 nonamer self-peptides, the truncated peptide tum$^-$ P35B-F9, and two motif peptides pD239 (AGPAAAAAL) and pD2359 (AGPARAAAL).

After the modeling routine described in Materials and Methods, which omits water, the resulting structures were examined for contacts between the bound peptide and the H chain, as well as the surface areas of both the peptide and MHC that were buried on binding (Table 3). Individual self-peptides had buried surface areas from 534 to 733 Å^2 with a mean of 655 Å^2, and the corresponding H-2D^d surface buried was from 740 to 1024 Å^2 with a mean of 917 Å^2. Peptides made extensive contacts with the class I molecule: an average of about 162 van der Waals contacts with a range of 134 to 208, an average of 21 hydrogen bonds (range 15 to 29), and from 1 to 7 ion pairs (salt bridges). Notably, pD239, the minimal motif peptide, had the lowest values for buried surface, and the fewest van der Waals contacts, hydrogen bonds, and salt bridges. The α-amino group of the first peptide residue and the carboxylate group of the COOH-terminal residue were frequently involved in salt bridges. There was no simple correlation between either the amount of surface area buried or the number or kinds of contacts and the apparent strength of binding to H-2D^d as assessed by the epitope induction assays employed.

Side chains at each of the nine peptide positions in the 17 nonamer H-2D^d–binding self-peptides were compared for their rms deviation from the average position and for the average fractional accessibility to solvent (Fig. 4, *E* and *F*). Positions 1, 2, and 9 had very little if any accessibility to solvent and amino acids at positions 4, 6, and 8 had the greatest (Fig. 4 *E*). Rms deviation from average position was smallest for positions 1, 8, and 9, and greatest for positions 2, 3, 6, and 7. Side chain/side chain interactions of nonadjacent amino acids were quantitated for the 17 models of the H-2D^d bound nonamer self-peptides. Between positions 5 and 7, 67 interactions were counted, and between residues 6 and 8, 11 interactions were counted. In addition, residues 7 and 9 had four interactions, and residue 5 interacted with residue 3 or 4 once.

Similar to published crystallographic structures, the peptide lies in an extended conformation with the NH_2 and COOH termini anchored at the left and right sides of the peptide binding cleft (Fig. 5, *A* and *B*). Our modeling suggests that the molecular architecture surrounding the NH_2 terminus of the peptide is determined in large part by the chemical character of the amino acid side chain at position 1. It is proposed, for example, that in the A pocket the NH_2-terminal amine of pD43 (SGPERGEKL) hydrogen bonds to Y7 and Y59 of H-2D^d and the hydroxyl of the serine side chain hydrogen bonds to Y171 (Fig. 6 *A*). In contrast, as shown in Fig. 6 *B*, the amine of the acidic aspartate of peptide pD41 (DGPVREHNL) would be in position to hydrogen bond to E63 and Y171 and the carboxylic acid side chain forms salt bridges with R62 and R66. Alternatively, the amine of pD42 (KGPERANGL) would hydrogen bond to Y7, Y45, and Y59, but cannot directly interact with Y171 (Fig. 6 *C*). The NH_2-terminal lysine forms a salt bridge with the side chain of E163 and hydrogen bonds to the backbone carbonyl oxygen of E163 as well (data not shown). The


(A)
p4E
p5R
p8K
pD43
p7E
p9L
p1S
p2G
p3P
H-2D^d
D
B
E
A
C
F

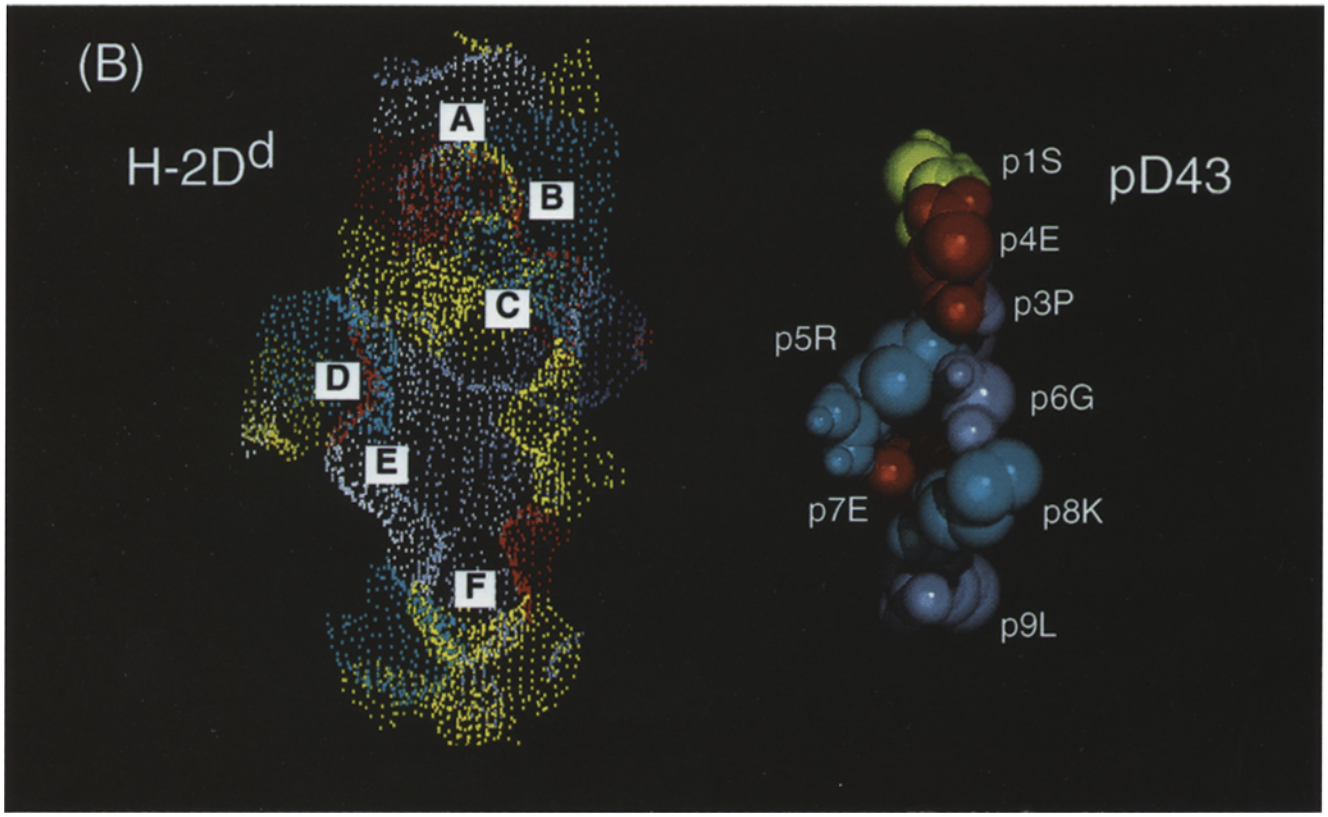

(B)
H-2D^d
A
B
C
D
E
F
p1S
pD43
p4E
p3P
p5R
p6G
p7E
p8K
p9L

Table 3. *H-2D^d/Peptide Interactions*

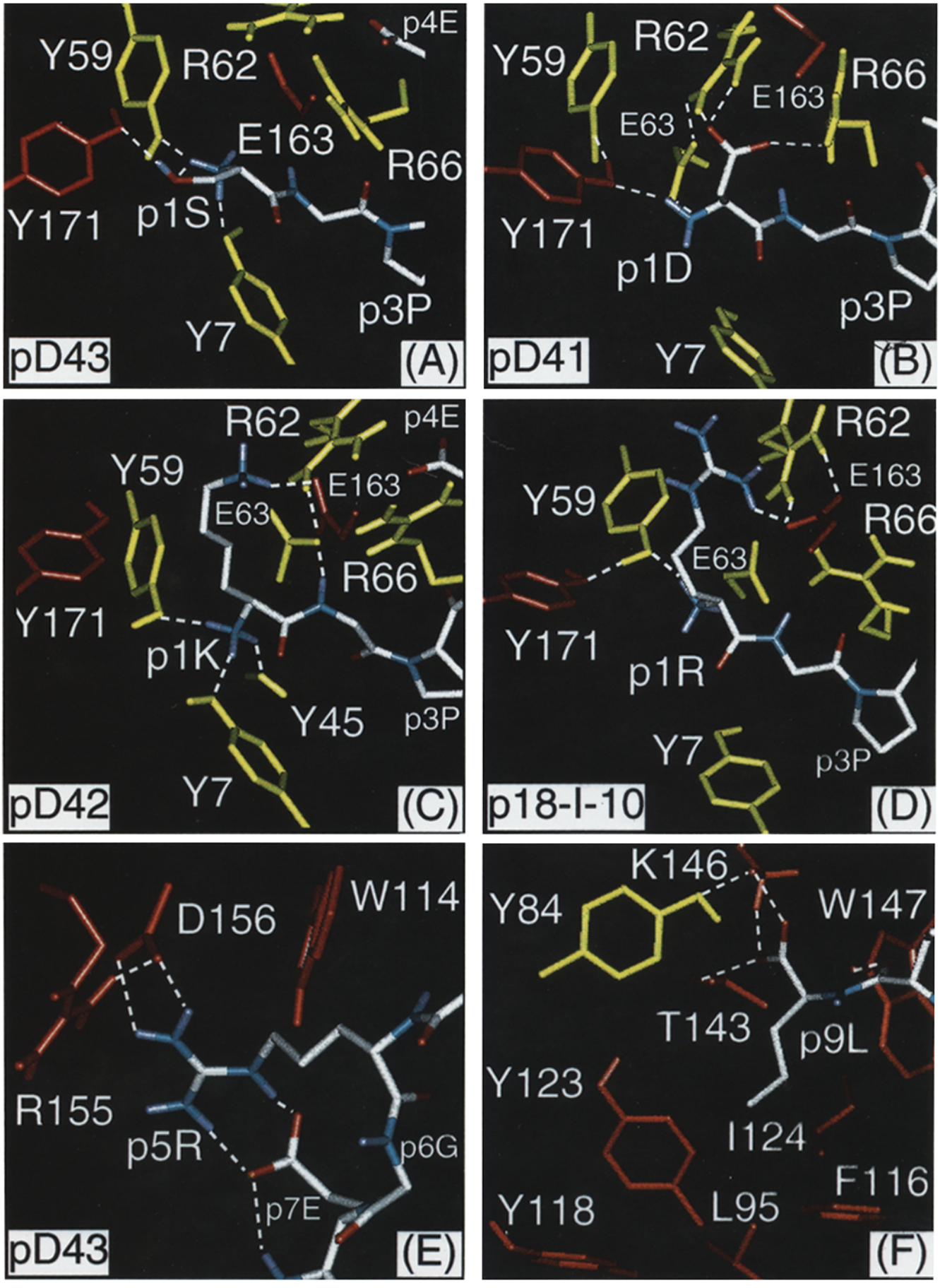
Ligand	Sequence	Peptide surface buried (Å^2)	H-2D^d Surface buried (Å^2)	Number of van der Waals contacts	Number of Hydrogen bonds	Number of Ion pairs
pD36	VGPQKNENL	662*	939*	160‡	23§	3‖
pD37	SGPRKAIAL	630	868	152	21	3
pD38	AGPDRTEKL	632	944	134	23	5
pD39	KGPDKGNEF	654	955	163	24	7
pD40	IGPERGHNL	683	1024	177	25	6
pD41	DGPVREHNL	665	940	145	21	6
pD42	KGPERANGL	638	924	152	26	6
pD43	SGPERGEKL	664	937	185	29	5
pD44	DGPVRGISI	625	872	129	23	4
pD45	NGPQRIYNL	678	812	199	27	3
pD46	SGPVALVNF	534	740	122	17	1
pD47	IGPNRAFNF	733	977	194	19	3
pD48	SGPERILSI	603	858	150	23	5
pD49	VGPSGKYFI	646	824	144	17	1
pD50	FGPYKLNRL	673	957	169	17	2
pD51	FGPIKFNVL	688	1030	176	14	3
pD53	FGPYRFYVL	730	992	208	15	3
pD239	AGPAAAAAL	547	693	120	11	1
pD2359	AGPARAAAL	615	886	163	22	3
tum$^-$35B-F9	NGPPHSNNF	656	1003	145	18	3

* Molecular surface lost on peptide binding was calculated for the indicated peptide/MHC models using the Connolly algorithm (35).
‡ Two atoms are said to be in van der Waals contact if they are within the sum of their van der Waal's radii plus 0.5 Å of each other.
§ Two atoms are said to be hydrogen bonded if they are within 2.9 ± 0.5 Å of each other.
‖ Two oppositely charged atoms are said to form an ion pair if they are within 2.85 ± 0.5 Å of each other.

larger side chain of arginine at position 1 of the viral peptide, p18-I-10, is accessible to solvent and also interacts with E163 (Fig. 6 *D*). In this case the NH_2-terminal amine hydrogen bonds with Y59. Comparison of Fig. 6, *C* and *D* indicates the kind of adjustments that residues Y7, Y171, R62, and R66 can make even when the conserved substitution of lysine to arginine at position 1 of the peptide occurs.

To illustrate other interactions observed between the bound peptide and MHC, as well as between peptide side chains, we will describe the model of the pD43/H-2D^d complex in some detail. We observe that conserved glycine and proline, representing residues 2 and 3, are constrained beneath the canopy formed by residues R62, R66, and E163 which form a salt bridge between the α_1 and α_2 helices (Fig. 6 *A*). These two peptide residues interact primarily with R66. Residue 4, a glutamic acid, is directed toward solvent and interacts with R66 from above through 21 atomic contacts. Residue 5, in this case the basic arginine, is oriented toward the D pocket (5, 6) under the α_2 helix, and is also inclined parallel to the peptide backbone toward the COOH terminus through interactions not only with aspartate 156, but also with the glutamate at position 7 of the peptide (Fig. 6 *E*). Peptide residues 6 and 8 point away from the cleft and are close enough to hydrogen bond with each other. The COOH terminus is anchored in the F pocket through hydrogen bonding of the penultimate carbonyl with the invariant W147; hydrophobic interactions of the COOH-terminal leucine side chain with residues A81 (data not shown), L95, F116, Y118,

Figure 5. Molecular model of the H-2D^d/pD43 complex. A van der Waals space filling model of the pD43 peptide is shown in lateral (*A*) and coronal (*B*) views, displaced from the corresponding calculated molecular surface representation of the H-2D^d residues that line the binding pocket. Colors represent polarity of the amino acid side chains: red, acidic; blue, basic; yellow, polar, noncharged; and light blue, nonpolar. The approximate locations of the pockets analogous to those described by Saper (5) and Garrett (6) are indicated by letters A through F.


Y59
R62
p4E
E163
R66
Y171
p1S
p3P
Y7
pD43
(A)
Y59
R62
E163
R66
E63
Y171
p1D
p3P
pD41
Y7
(B)
R62
p4E
Y59
E163
E63
R66
Y171
p1K
Y45
p3P
pD42
Y7
(C)
R62
E163
Y59
R66
E63
Y171
p1R
Y7
p3P
p18-I-10
(D)
W114
D156
R155
p5R
p6G
p7E
pD43
(E)
K146
Y84
W147
T143
p9L
Y123
I124
F116
Y118
L95
(F)

Y123, and I124; a hydrogen bond between the carboxylate terminus and T143; and a salt bridge formed with K146 (Fig. 6 *F*).

In addition to the set of nonamer peptide/H-2D^d models, we built models of one octamer (pD35, KGPITVQI) and six decamer peptide/H-2D^d complexes. The resulting model of the pD35/H-2D^d complex maintained the side chain of the COOH-terminal residue in the F pocket, and had the side chain of the threonine at position 5 oriented for hydrogen bonding to D156. To preserve the architecture revealed by the X-Gly-Pro-X-(Arg, Lys)-X-X-X-(Leu, Ile) motif, five alternate models were constructed for each decamer peptide by introducing the additional residue after position 5, 6, 7, 8, or 9. After the minimization/slow cooling procedure described in Materials and Methods, the decamer models with insertions after residue 5, 6, or 7 resulted in structures with similar backbone configurations and side chain orientations, whereas addition after position 8 or 9 resulted considerable distortion of the modelled structures. Also, these structures frequently showed side chain/side chain interactions of residues 5 with 7, and 6 with 8 as did the nonamer structures. A direct comparsion of the peptide structures of the octamer pD35, nonamer pD43, and decamer p18-I-10, is displayed in Fig. 7. The backbone configuration dictated by the proline at position 3, as well as the interaction of peptide residue 5 with D156, the anchoring of the hydrophobic COOH terminus and the side chain/side chain interactions permits these peptides of different length to be bound by the same MHC molecule.

Discussion

In the past several years, beginning with the seminal observations of Falk et al. (48) and van Bleek and Nathenson (14), it has become apparent that both endogenous and viral peptides associated with MHC class I molecules can be identified and sequenced. Such studies provided the basis for a structural understanding of the interaction of particular MHC class I molecules with particular peptides. Here we have described a large number of peptides that were copurified with a soluble analog of H-2D^d, studied their binding by induction of a serologic epitope, and analyzed their structural interaction with the class I H chain using computer-generated molecular models based on the crystallographic coordinates of the HLA-B27/peptide complex. This analysis not only permitted the determination of the binding motif for H-2D^d–restricted peptides, but also an evaluation of the peptide/MHC and intrapeptide interactions that contributed to a stable ternary complex.

The peptides that copurified with the H-2D^d soluble analog likely represented the more tightly bound subset of peptides, having been selected by the protein production and immunopurification procedures. Those peptides identified were predominately nonamers and decamers, however one octamer was confidently identified. The length of these peptides was consistent with the findings of others who have characterized MHC class I–associated peptides. However, unlike the previously described peptide motifs involving two residues, the amino acid composition of the H-2D^d peptides was restricted at four positions: glycine at 2, proline at 3, basic residue at 5, and hydrophobic residue at the COOH terminus. The motif was confirmed with the epitope induction due to the peptides: pD239 (AGPAAAAAL) and pD2359 (AGPARAAAL). The suggestion that pD2359 binds better than pD239 has been confirmed by a solution phase competition assay (data not shown). In addition, the motif is consistent with that of two known viral and tumor peptides, p18-I-10 and tum35B. The identification of this motif supported the suggestion that a nonamer derivative of the undecamer tum35B, tum35B-F9 would bind H-2D^d, and raises the possibility that this nonamer might be antigenic as well.

The structural constraints implied by the four motif residues were examined with the aid of 17 independent molecular models of H-2D^d_s/self-peptide/mβ_2m complexes and two similar models with the alanine-substituted motif peptides (pD239 and pD2359). We have described four different patterns of interactions in the A pocket dependent on the biochemical character of the side chain at position 1 of the peptide. The differences in surface exposure of the side chain of the first peptide residue and the complex interaction patterns with the residues that line the A pocket dependent on the side chain suggest that mAbs that bind in the region may be able to discriminate different MHC/peptide combinations. Indeed, this behavior has been observed for several anti-H-2D^d (49) and anti-H-2K^b (50, 51) mAbs.

Whereas the A pocket exhibits a degree of plasticity to accommodate side chains of different size and charge, the B pocket of H-2D^d has strict requirements for acceptable amino acids. The restriction to glycine-proline appears to be more a configurational than a spatial constraint, evidenced by the relatively large rms deviation at those positions. Indeed glycine is restricted in its conformational freedom when followed by proline. The ϕ/ψ plot for glycine in glycine-proline from nonhomologous structures determined to 2.5 Å resolution or better shows a predominant clustering around $\psi = 180°$, suggesting that glycine-proline nearly always adopts an extended conformation (52). Other class I motifs that employ proline include H-2L^d (17), HLA-B35, and HLA-B53 (40), which favor this residue at position 2. In the case of H-2L^d, modeling analysis suggested that the B pocket was shallow and markedly spatially constrained.

The side chain of the fifth peptide residue is oriented toward the D pocket, enabling salt bridging (through a basic group) or hydrogen bonding (through threonine e.g., pD34)

Figure 6. Models of H-2D^d/peptide complexes. Three-dimensional molecular models of complexes of pD43 (*A*, *E*, and *F*); pD41 (*B*); pD42 (*C*); and p18-I-10 (*D*) are shown. The region of the A pocket is shown in the same view for the four different models in *A–D*. The region of the D pocket of pD43 is in (*E*) and that of the F pocket in (*F*). Amino acid residues of H-2D^d from the α_1 helix are in yellow and those from the α_2 helix are in red. The peptide residues are shown as N, blue; C, white; O, red; and polar H, light blue. Hydrogen bonds are shown as dotted lines. For clarity, backbone atoms of H-2D^d residues are omitted, and only a selection of the residues is portrayed.

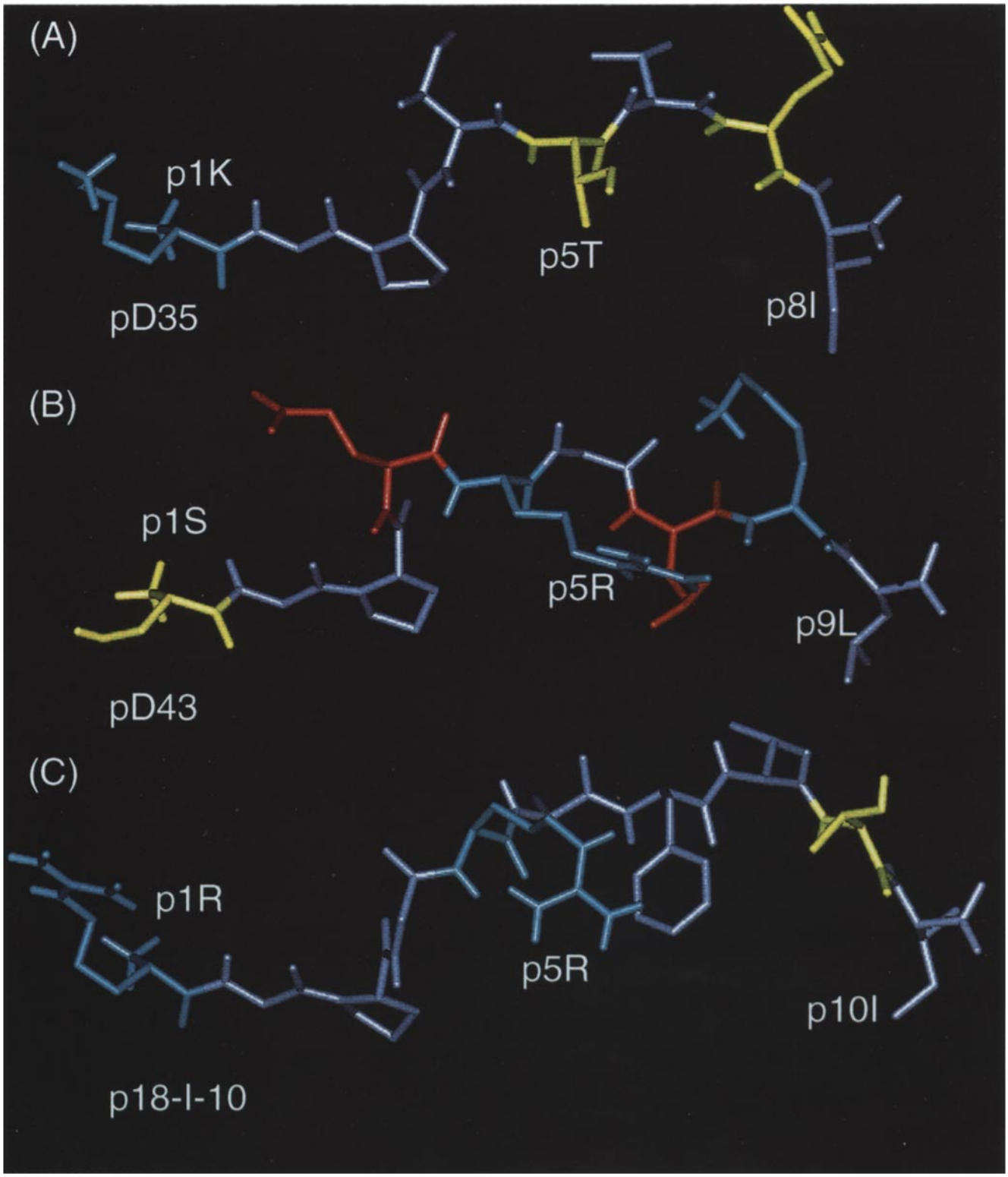


Figure 7. Molecular models of three H-2D^d bound peptides. The structures of models of an octamer pD35 (*A*), a nonamer pD43 (*B*), and a decamer p18-I-10 (*C*) are shown as licorice models, with the NH_2-terminal residues to the left and the COOH-terminal residues to the right. Coloring is by polarity as indicated in the legend to Fig. 5.

to the polymorphic aspartic acid 156. The contribution of the positively charged residue to the interaction of the peptide with the H chain is highlighted by the doubling of the number of hydrogen bonds, the increase in van der Waals contacts from 120 to 163 and in ion pairs in the models of the two alanine-substituted motif peptides pD239 and pD2359, which differ solely by the arginine residue at position 5. The peptide-dependent epitope induction data support the view that the middle positively charged residue enhances the interaction of the peptide with the H chain and can act as a fourth anchoring position.

The D156 polymorphism is shared with H-2D^d by H-2K^k and H-2D^k (53). The peptide binding motif of H-2K^k inferred from inspection of previously described viral and malarial peptides is most likely glutamic acid at position 2 and isoleucine at position 9 (54, 55). There is no indication that these peptides have a conserved central positively charged residue. Unlike H-2K^k and H-2D^k, however, H-2D^d also has polymorphisms at W97 and W114 which form a ridge in the floor of the groove. The side chain/side chain interactions noted in MHC-bound H-2D^d–restricted peptides, namely the interaction of 5 with 7, and 6 with 8, serve to kink the otherwise fully extended peptide backbone into a bend permitting the peptide to accommodate this ridge. H-2K^k and H-2D^k both have arginine and glutamic acid at positions 97 and 114, respectively, which are capable not only of interacting with each other, but also of interacting with bound peptide. The W97/W114 pair in H-2D^d may contribute to a lower local dielectric constant than that expected for the R97/E114 pair (in H-2K^k and H-2D^k), enhancing the electrostatic interaction between the basic group at peptide position 5 and D156. Thus, the summed effects of D156, W97, and W114 are likely to contribute to the advantage of a basic residue at position 5 for H-2D^d peptides.

The positioning of the peptide by its interaction with residue 156 of the H chain impacts on the conformational or combinatorial interface necessary to effect signaling through the TCR. This behavior is suggested by the failure of the H-2K^b mutant bm1, which differs at positions 152 (E to A), 155 (R to Y), and 156 (L to Y), to present antigen to an appropriately restricted effector T cell despite evidence that it associates with the appropriate peptide (56, 57). The analysis of the role of residue 156 in peptide presentation has been refined by single-site mutational studies in HLA-A2 (58, 59). These studies showed that mutations at 156 could either enhance or abrogate functional peptide presentation depending on the peptide.

In contrast to H-2K^b, the C pocket is not prominent in the calculated surface representation of the binding cleft for H-2D^d, and appears to play little if any role in the interactions with peptide since there is no preference for hydrophobic residues in the center of the H-2D^d/bound peptides. The E pocket in H-2D^d is shallow as well, and its contribution to peptide binding appears small. The deep hydrophobic F pocket, like that of a number of other MHC class I molecules examined (HLA-A2, H-2K^b, and H-L^d), seems to play a major role in anchoring the COOH terminus of H-2D^d bound peptides. In other MHC molecules, HLA-B27 (8) and HLA-A3 (39), the F pocket appears capable of accommodating peptides with positively charged COOH-terminal residues as well as hydrophobic ones.

Thus, the H-2D^d bound peptides have several characteristics that contribute to stable interactions with the MHC molecule: defined length (8–10 residues), the above mentioned motif, and the propensity for side chain/side chain interactions. Each of these qualities may contribute not only to specific interactions with H-2D^d itself, but may reflect entropic advantages. Short peptides clearly have less entropy than longer ones, prolines severely limit entropy as compared to all other amino acids by limiting the available torsion angles in the backbone, and side chain/side chain interactions must also contribute to a peptide with less potential disorder. Further binding and structural analysis of more MHC-restricted peptides and their substituted analogs can allow critical evaluation of this possibility.

Molecular models made for decamer/H-2D^d_s complexes, designed to preserve the constraints of the motif residues, illustrate the flexibility of the peptide backbone between anchoring residues, particularly residues 6, 7, 8, and 9. This bend is consistent with crystallographic data on decamer HLA-Aw68 complexes (60) The crystal structures of H-2K^b complexed with a nonamer or an octamer peptide (9, 11) describe a similar kink in the peptide backbone between the A and C pocket residues. In our analysis of nonamer peptides, residues that are the most solvent accessible (residues 4, 6, and 8) are also those with the most variability in amino acid selection. These residues are likely to be epitopic (i.e., available for TCR recognition), and the restricted accessibility of peptide residue 7 may reflect the high frequency of interactions between residues 5 and 7. For the p18-I-10 decamer, because of a presumably larger kinked backbone between residues 5 and 10, all the intervening amino acids are accessible to solvent, consistent with the identification of residues 6, 7, and 8 as T cell epitopic positions in the context of H-2D^d (42).

Thus, the general rules that govern the interaction of MHC molecules with peptides continue to hold. Antigenic peptides are a preferred length, so that both NH_2- and COOH-terminal groups can be bound, and they have a particular subset of residues defined by hydrophobicity, size, polarity, or charge to bind at anchor positions. These binding requirements then allow the other, more solvent-exposed peptide residues to vary. Our studies of H-2D^d bound peptides illustrate several unique aspects of this developing theme: the exploitation of four rather than one or two peptide residues for peptide binding; the plasticity of the molecular architecture in accommodating the NH_2-terminal peptide residue, and the identification of a large number of intrapeptide interactions, in this case likely to permit the peptide backbone to accommodate a ridge in the floor of the MHC molecule. As we develop more precise understanding of the MHC/peptide interaction and its structural causes and conformational consequences, we may progress to a further characterization of a deeper problem—how TCR interact with the MHC/peptide complex.

We thank J. Coligan and the BRB for sequencing and synthesizing peptides, S. Starnes for secretarial assistance, C. Hoes for technical support, and D. Preuss and A. Graeff for computer support. We thank W. Biddison, S. Khilko, S. Kozlowski, R. Ribaudo, S. Sadegh-Nasseri, and M. Zauderer for their advice and comments on the manuscript.

Address correspondence to Dr. David H. Margulies, Molecular Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, NIH, Building 10, Room 11N311, Bethesda, MD 20892.

Received for publication 24 May 1993 and in revised form 18 August 1993.

References

1. Bjorkman, P., and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem.* 59:253.
2. Falk, K., O. Rötzschke, S. Stevanovic, G. Jung, and H. Rammensee. 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature (Lond.).* 351:290.
3. Bjorkman, P., M. Saper, B. Samraoui, W. Bennett, J. Strominger, and D. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature (Lond.).* 329:506.
4. Bjorkman, P.J., M.A. Saper, B. Samraoui, W.S. Bennett, J.L. Strominger, and D.C. Wiley. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature (Lond.).* 329:512.
5. Saper, M., P. Bjorkman, and D. Wiley. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* 219:277.
6. Garrett, T., M. Saper, P. Bjorkman, J. Strominger, and D. Wiley. 1989. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature (Lond.).* 342:692.
7. Madden, D., J. Gorga, J. Strominger, and D. Wiley. 1991. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature (Lond.).* 353:321.
8. Madden, D., J. Gorga, J. Strominger, and D. Wiley. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell.* 70:1035.
9. Fremont, D., M. Matsumura, E. Stura, P. Peterson, and I. Wilson. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science (Wash. DC).* 257:919.
10. Matsumura, M., D. Fremont, P. Peterson, and I. Wilson. 1992. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science (Wash. DC).* 257:927.
11. Zhang, W., A. Young, M. Imarai, S. Nathenson, and J. Sacchettini. 1992. Crystal strucure of the major histocompatibility complex class I H-2K^b molecule containing a single viral peptide: implications for peptide binding and T-cell receptor recognition. *Proc. Natl. Acad. Sci. USA.* 89:8403.
12. Rötzschke, O., K. Falk, K. Deres, H. Schild, M. Norda, J. Metzger, G. Jung, and H. Rammensee. 1990. Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature (Lond.).* 348:252.
13. Falk, K., O. Rötzschke, K. Deres, J. Metzger, G. Jung, and H. Rammensee. 1991. Identification of naturally processed viral nonapeptides allows their quantification in infected cells and suggests an allele-specific T cell epitope forecast. *J. Exp. Med.* 174:425.
14. Van Bleek, G., and S. Nathenson. 1990. Isolation of an endogenously processed immunodominant viral peptide from the class I H-2K^b molecule. *Nature (Lond.).* 348:213.
15. Jardetzky, T., W. Lane, R. Robinson, D. Madden, and D. Wiley. 1991. Identification of self peptides bound to purified HLA-B27. *Nature (Lond.).* 353:326.
16. Hunt, D., R. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. Cox, E. Appella, and V. Engelhard. 1992. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science (Wash. DC).* 255:1261.
17. Corr, M., L.F. Boyd, S.R. Frankel, S. Kozlowski, E.A. Padlan, and D.H. Margulies. 1992. Endogenous peptides of a soluble major histocompatibility complex class-I molecule, H-2L^d-sequence motif, quantitative binding, and molecular modeling of the complex. *J. Exp. Med.* 176:1681.
18. Margulies, D.H., A. Ramsey, L.F. Boyd, and J. McCluskey. 1986. Genetic engineering of an H-2D^d/Q10^b chimeric histocompatibility antigen: purification of soluble protein from transformant cell supernatants. *Proc. Natl. Acad. Sci. USA.* 83:5252.
19. Kozlowski, S., T. Takeshita, W.H. Boehncke, H. Takahashi, L.F. Boyd, R.N. Germain, J.A. Berzofsky, and D.H. Margulies. 1991. Excess beta 2 microglobulin promoting functional peptide association with purified soluble class I MHC molecules. *Nature (Lond.).* 349:74.
20. Ozato, K., N.M. Mayer, and D.H. Sachs. 1982. Monoclonal antibodies to mouse major histocompatibility complex antigens. *Transplantation (Baltimore).* 34:113.
21. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403.
22. Abastado, J., C. Jaulin, M. Schutze, P. Langlade-Demoyen, F. Plata, K. Ozato, and P. Kourilsky. 1987. Fine mapping of epitopes by intradomain K^d/D^d recombinants. *J. Exp. Med.* 166:327.
23. Otten, G., E. Bikoff, R. Ribaudo, S. Kozlowski, D.H. Margulies, and R.N. Germain. 1992. Peptide and beta 2-microglobulin regulation of cell surface MHC class I conformation and expression. *J. Immunol.* 148:3723.
24. Bikoff, E., L. Jaffe, R.K. Ribaudo, G. Otten, R.N. Germain, and E. Robertson. 1991. MHC class I surface expression in embryo-derived cell lines inducible with peptide or interferon. *Nature (Lond.).* 354:235.
25. Wu, T., and E. Kabat. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132:211.
26. Padlan, E. 1977. Structural implications of sequence variability

in immunoglobulins. *Proc. Natl. Acad. Sci. USA.* 74:2551.
27. Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. *Science (Wash. DC).* 185:862.
28. Gates, F., J. Coligan, and T. Kindt. 1981. Complete amino sequence of murine beta 2-microglobulin: structural evidence for strain-related polymorphism. *Proc. Natl. Acad. Sci. USA.* 78:554.
29. Parnes, J., and J. Seidman. 1982. Structure of wild-type and mutant mouse beta 2-microglobulin genes. *Cell.* 29:661.
30. Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and K. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry.* 4:187.
31. Brünger, A. X-PLOR Manual, Version 2.1. 1990. Yale University Press, New Haven. 382 pp.
32. Kirkpatrick, S., C. Gelatt, Jr., and M. Vecchi. 1983. Optimization by simulated annealing. *Science (Wash. DC).* 220:671.
33. Wilson, S., and W. Cui. 1990. Applications of simulated annealing to peptides. *Biopolymers.* 29:225.
34. Powell, M.J.D. 1977. Restart procedures for the conjugate gradient method. *Mathematical Programming.* 12:241.
35. Connolly, M. 1983. Analytical molecular surface calculation. *Journal of Applied Crystallography.* 16:548.
36. McCluskey, J., L. Boyd, P. Highet, J. Inman, and D. Margulies. 1988. T cell activation by purified, soluble, class I MHC molecules. Requirement for polyvalency. *J. Immunol.* 141:1451.
37. Kozlowski, S., M. Corr, T. Takeshita, L.F. Boyd, C.D. Pendleton, R.N. Germain, J. Berzofsky, and D.H. Margulies. 1992. Serum angiotensin-1 converting enzyme activity processes a human immunodeficiency virus 1 gp160 peptide for presentation by major histocompatibility complex class I molecules. *J. Exp. Med.* 175:1417.
38. Wei, M., and P. Cresswell. 1992. HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature (Lond.).* 356:443.
39. Dibrino, M., K.C. Parker, J. Shiloach, M. Knierman, J. Lukszo, R.V. Turner, W.E. Biddison, and J.E. Coligan. 1993. Endogenous peptides bound to HLA-A3 possess a specific combination of anchor residues that permit identification of potential antigenic peptides. *Proc. Natl. Acad. Sci. USA.* 90:1508.
40. Hill, A.V.S., J. Elvin, A.C. Willis, M. Aidoo, C.E.M. Allsopp, F.M. Gotch, X.M. Gao, M. Takiguchi, B.M. Greenwood, A.R.M. Townsend, et al. 1992. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature (Lond.).* 360:434.
41. Sutton, J., S. Rowlandjones, W. Rosenberg, D. Nixon, F. Gotch, X.M. Gao, N. Murray, A. Spoonas, P. Driscoll, M. Smith, et al. 1993. A sequence pattern for peptides presented to cytotoxic T-lymphocytes by HLA-B8 revealed by analysis of epitopes and eluted peptides. *Eur. J. Immunol.* 23:447.
42. Shirai, M., C. Pendleton, and J.A. Berzofsky. 1992. Broad recognition of cytotoxic T cell epitopes from the HIV-1 envelope protein with multiple class I histocompatibility molecules. *J. Immunol.* 148:1657.
43. Szikora, J.P., A. VanPel, V. Brichard, M. Andre, N. VanBaren, P. Henry, E. DePlaen, and T. Boon. 1990. Structure of the gene of tum− transplantation antigen P35B: presence of a point mutation in the antigenic allele. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:1041.
44. Szikora, J.P., A. Vanpel, and T. Boon. 1993. Tum− mutation P35B generates the MHC-binding site of a new antigenic peptide. *Immunogenetics.* 37:135.
45. Takahashi, H., J. Cohen, A. Hosmalin, K. Cease, R. Houghten, J. Cornette, C. DeLisi, B. Moss, R.N. Germain, and J.A. Berzofsky. 1988. An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex molecule-restricted murine cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA.* 85:3105.
46. Maryanski, J., A. Verdini, P. Weber, F. Salemme, and G. Corradin. 1990. Competitor analogs for defined T cell antigens: peptides incorporating a putative binding profile and polyproline or polyglycine spacers. *Cell.* 60:63.
47. Ribaudo, R.K., and D.H. Margulies. 1992. Independent and synergistic effects of disulfide bond formation, beta 2-microglobulin, and peptides on class I MHC folding and assembly in an in vitro translation system. *J. Immunol.* 149:2935.
48. Falk, K., O. Rözschke, and H. Rammensee. 1990. Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature (Lond.).* 348:248.
49. Abastado, J., S. Darche, H. Jouin, C. Delarbre, G. Gachelin, and P. Kourilsky. 1989. A monoclonal antibody recognizes a subset of the H-2D^d mouse major class I antigens. *Res. Immunol.* 140:581.
50. Čatipović, B., J. Dal Porto, M. Mage, T.E. Johansen, and J.P. *Schneck. 1992. Major histocompatibility complex conformational epitopes are peptide specific. J. Exp. Med.* 176:1611.
51. Bluestone, J.A., S. Jameson, S. Miller, and R. Dick II. 1992. Peptide-induced conformational changes in class I heavy chains alter major histocompatibility complex recognition. *J. Exp. Med.* 176:1757.
52. MacArthur, M., and J. Thornton. 1991. Influence of proline residues on protein conformation. *J. Mol. Biol.* 218:397.
53. Watts, S., C. Wheeler, R. Morse, and R. Goodenow. 1989. Amino acid comparison of the class I antigens of mouse major histocompatibility complex. *Immunogenetics.* 30:390.
54. Gould, K.G., H. Scotney, and G.G. Brownlee. 1991. Characterization of two distinct major histocompatibility complex class I K^k-restricted T-cell epitopes within the influenza A/PR/8/34 virus hemagglutinin. *J. Virol.* 65:5401.
55. Cossins, J., K.G. Gould, M. Smith, P. Driscoll, and G.G. Brownlee. 1993. Precise prediction of a K^k-restricted cytotoxic T cell epitope in the NS1 protein of influenza virus using an MHC allele-specific motif. *Virology.* 193:289.
56. Nathenson, S., J. Geliebter, G. Pfaffenbach, and R. Zeff. 1986. Murine major histocompatibility complex class-I mutants: molecular analysis and structure-function implications. *Annu. Rev. Immunol.* 4:471.
57. Falk, K., O. Rötzschke, and H.-G. Rammensee. 1992. A self peptide naturally presented by both H-2K^b and H-2K^{bm1} molecules demonstrates MHC restriction of self tolerance at the molecular level. *Int. Immunol.* 4:321.
58. Utz, U., S. Koenig, J. Coligan, and W. Biddison. 1992. Presentation of three different viral peptides, HTLV-1 Tax, HCMV gB, and influenza virus M1, is determined by common structural features of the HLA-A2.1 molecule. *J. Immunol.* 149:214.
59. McMichael, A., F. Gotch, J. Santos-Aguado, and J. Strominger. 1988. Effect of mutations and variations of HLA-A2 on recognition of a virus peptide epitope by cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA.* 85:9194.
60. Guo, H.C., T.S. Jardetzky, T.P.J. Garrett, W.S. Lane, J.L. Strominger, and D.C. Wiley. 1992. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature (Lond.).* 360:364.
61. Padlan, E. 1990. On the nature of antibody combining sites: unusual structural features that may confer on these sites an

enhanced capacity for binding ligands. *Proteins.* 7:112.
62. Oleinikov, A.V., G.G. Jokhadze, and Y.B. Alakhov. 1989. Primary structure of rat liver elongation factor 2 deduced from the cDNA sequence. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 248:131.
63. Ray, D., S. Culine, A. Tavitan, and F. Moreau Gachelin. 1990. The human homologue of the putative proto-oncogene spi-1: characterization and expression in tumors. *Oncogene.* 5:663.
64. Flores-Riveros, J.R., E. Sibley, T. Kastelic, and M.D. Lane. 1989. Substrate phosphorylation catalyzed by the insulin receptor tyrosine kinase. Kinetic correlation to autophosphorylation of specific sites in the beta subunit. *J. Biol. Chem.* 264:21557.
65. Herisse, J., and F. Galibert. 1981. Nucleotide sequences of the EcoRI E fragment of adenovirus 2 genome. *Nucleic Acids Res.* 9:1229.
66. Roberts, K.P., and M.D. Griswold. 1990. Characterization of rat transferrin receptor mRNA in testes in Sertoli cells in culture. *Mol. Cell. Endocrinol.* 14:531.
67. Torres, R.M., and E.A. Clark. 1992. Differential increase of an alternatively polyadenylated mRNA species of murine CD40 upon B cell activation. *J. Immunol.* 148:620.
68. Kornblihtt, A.R., K. Umezawa, K. Vibe-Pedersen, and F.E. Baralle. 1985. Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1755.
69. Shimizu, S., K. Malik, H. Sejima, J. Kishi, T. Hayakawa, and O. Koiwai. 1992. Cloning and sequencing of the cDNA encoding a mouse tissue inhibitor of metalloproteinase-2. *Gene.* 114:291.
70. Argos, P., M. Hane, J.M. Wilson, and W.N. Kelly. 1983. Molecular basis of hypoxanthine guanine phosphribosyltransferase deficiency in a patient with the Lesch Nyhan syndrome. *J. Clin. Invest.* 71:1331.
71. Riddles, P.W., V. Whan, R.L. Blakely, and B. Zerner. 1992. Cloning and sequencing of a jackbean urease encoding cDNA. *Gene.* 108:265.
72. Naguchi, T., H. Inoue, and T. Tanaka. 1986. The m1 and m2 type isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing. *J. Biol. Chem.* 261: 13807.
73. Reddy, A.B., A. Chatterjee, L.I. Rothblum, A. Black, and H. Busch. 1989. Isolation and characterization of complementary DNA to proliferating cell nucleolar antigen P40. *Cancer Res.* 49:1763.
74. Nakanishi, O., M. Oyanagi, Y. Kuwano, T. Tanaka, T. Nakayawa, H. Mitsui, Y.I. Nabeshima, and K. Ogatata. 1985. Molecular cloning and nucleotide sequences of cDNAs for rat liver ribosomal proteins S17 and L30. *Gene.* 35:289.
75. Eggertsen, G., G. Hudson, B. Shiels, D. Reed, K. Lonberg-Holm, and G.H. Fey. 1991. Sequence of rat alpha 1 macroglobulin, a broad range proteinase inhibitor from the alpha-macroglobulin complement family. *Mol. Biol. & Med.* 8:287.
76. Lerch, K., and D. Ammer. 1981. Amino acid sequence of copper zinc oxide dismutase from horse liver. *J. Biol. Chem.* 256:11545.
77. La Rosa, G.J., J.P. Davide, K. Weihold, J.A Waterbury, A.T. Profy, J.A. Lewis, A.J. Langois, G.R. Dressman, R.N. Boswell, P. Shadduck, et al. 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science (Wash. DC).* 249:932.
78. Matunis, M.J., W.M. Michael, and G. Dreyfus. 1992. Characterization and primary structure of poly(C) binding heterogenous nuclear ribonuclear protein complex K protein. *Mol. Cell. Biol.* 12:164.
79. Sarvetnick, N., J.-Y. Tsai, H. Fox, S.H. Pilder, and L. Silver. 1990. A mouse chromosome 17 gene encodes a testes specific transcript with unusual properties. *Immunogenetics.* 31:283.
80. Stewart, M.A., M. Warnok, A. Wheeler, N. Wilkie, J.I. Mullins, D.E. Onions, and J.C. Neil. 1986. Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. *J. Virol.* 58:825.
81. Dupree, P., V.M. Olkkonnen, and P. Chavrier. 1992. Sequence of a canine cDNA clone encoding a Ran/TC4 GTP-binding protein. *Gene.* 120:325.
82. Pissowotzki, K., K. Mansouri, and W. Piepersberg. 1991. Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes str ELMB2N. *Mol. & Gen. Genet.* 231:113.
83. Schad, P.A., and B.H. Iglewski. 1988. Nucleotide sequence and expression in *E. coli* of the *Pseudomonas aeruinosa* Las A gene. *J. Bacteriol.* 170:2784.