Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3296–3300, May 1989
Immunology

Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis

(binding motif/class II molecules/vaccines/autoimmunity)

ALESSANDRO SETTE*, SØREN BUUS†, ETTORE APPELLA‡, JOHN A. SMITH§, ROBERT CHESNUT*, CRAIG MILES*, SONIA M. COLON*, AND HOWARD M. GREY*¶

*Cytel, 11099 North Torrey Pines Road, La Jolla, CA 92037; †Institute for Experimental Immunology, 71, Nørre Alle, DK-2100 Copenhagen Ø, Denmark; ‡Laboratory of Cell Biology, National Cancer Institute, Bethesda, MD 20892; and §Departments of Pathology and Molecular Biology, Massachusetts General Hospital, and Department of Pathology, Harvard Medical School, Boston, MA 02114

Communicated by Clement L. Markert, January 23, 1989

ABSTRACT **We have previously experimentally analyzed the structural requirements for interaction between peptide antigens and mouse major histocompatibility complex (MHC) molecules of the *d* haplotype. We describe here two procedures devised to predict specifically the capacity of peptide molecules to interact with these MHC class II molecules (IA^d and IE^d). The accuracy of these procedures has been tested on a large panel of synthetic peptides of eukaryotic, prokaryotic, and viral origin, and also on a set of overlapping peptides encompassing the entire staphylococcal nuclease molecule. For both sets of peptides, IA^d and IE^d binding was successfully predicted in ≈75% of the cases. This suggests that definition of such sequence "motifs" could be of general use in predicting potentially immunogenic peptide regions within proteins.**

There is at present compelling evidence in support of the concept that T cells recognize a complex formed between major histocompatibility complex (MHC) molecules and "processed" antigen. The structural requirements for this interaction have been the object of intense investigation. In a previous study (1), the effect of a large number of single amino acid substitutions of an ovalbumin-derived peptide that binds strongly to IA^d [Ova(323–339)] on its capacity to bind purified IA^d molecules was examined. Significant changes were only detected in a six-amino acid core region [Ova(327–332)] that had also been strongly implicated as crucial for binding by analysis of a set of N- and C-terminal truncated analogs. In a subsequent study (2), it was demonstrated that unrelated peptides that are good IA^d binders contained regions ("motifs") that were structurally similar to each other and to Ova(327–332). More recently, it was shown (A.S., L. Adorini, S.B., E.A., and H.M.G., unpublished data) that a different amino acid sequence pattern characterizes the interaction between peptide antigens and the other class II MHC molecule, IE^d. The motif recognized by IE^d molecules appears to be independent of the motif recognized by IA^d and is defined by the presence of positively charged amino acids (refs. 3 and 4; A.S., L. Adorini, S.B., E.A., and H.M.G., unpublished data).

In the present study, this information was used to derive a procedure that identifies motifs. Using a large panel of peptides, the presence of either an IA^d or IE^d motif as determined by this procedure was compared to the experimentally determined capacity of a given peptide to bind IA^d and IE^d. The peptide data base contained 62 peptides and, in addition, a set of peptides encompassing the entire staphylococcal nuclease molecule. In these sets of peptides, IA^d and IE^d binding was successfully predicted in ≈75% of the cases, suggesting that similar procedures should be useful in selecting peptides capable of interacting with other MHC specificities, identify possible autoantigens in self molecules or specific blockers of autoimmune diseases, and because of the previously demonstrated strong correlation between MHC binding and immunogenicity (5) be useful in selecting potentially immunogenic peptide regions within proteins.

MATERIALS AND METHODS

Peptide Synthesis and Purification. Peptides derived from various viral, bacterial, and eukaryotic proteins were synthesized as described (1, 2, 4). Peptides were HPLC isolated and their purity was determined by amino acid composition analysis. Peptides used in this study were routinely ≥95% pure.

Ia Purification and Binding Assay. IA^d and IE^d molecules were purified by affinity chromatography as described (2, 4–6), and the capacity of each peptide to bind either IA^d or IE^d molecules was assessed in a competition assay by a gel filtration procedure (6). To allow comparison of data obtained in different experiments, a relative binding was calculated for each peptide by dividing the 50% inhibition doses of the positive control for inhibition [unlabeled Ova(323–337) or λ repressor(12–26) for IA^d and IE^d, respectively] by the 50% inhibition doses of each peptide. The 50% inhibition doses for Ova(323–337) and λ repressor were routinely in the 1–10 μM range. Based on its binding capacity relative to that of the positive control, each peptide was then scored as a good binder (relative binding, >0.1), intermediate binder (relative binding, 0.1–0.01), or nonbinder (relative binding, <0.01).

Sequence Similarity Analysis. Peptide sequences were searched for the presence of the IA^d and IE^d motifs as defined by the procedures described below. Peptide sequences were also searched for predicted T-cell sites according to the procedures described by Rothbard and Taylor (7) and Berzofsky and associates (8, 9). In the case of the Berzofsky procedure, the corresponding algorithm was run on a Univax, using an averaging block length of 7 (9). Statistical significances were evaluated by a simple χ^2 test (10).

RESULTS

We previously identified an IA^d binding motif on the basis of the structural similarities between unrelated IA^d binding peptides using the six-amino acid core sequence that had been shown to be crucial for IA^d binding of Ova(323–339) [i.e., Ova(327–332)] as a "master sequence." The present study defines more precisely a motif as a sequence of six amino acids that contains at each position either the original

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; Ova, chicken egg ovalbumin.

¶To whom reprint requests should be addressed.

amino acid from the Ova(327–332) sequence (Val-His-Ala-Ala-His-Ala), a structurally similar amino acid, or one that has been found experimentally to be tolerated when substituted for an original amino acid (Table 1).

To obtain a numerical figure to be used in the analysis, we arbitrarily assigned a value of 3 to substitutions with little or no effect on IA^d binding (change in 50% inhibition doses up to a factor of 2), 2 to substitutions with some effect (50% inhibition dose reduction by a factor of 2–10), and 1 to substitutions with drastic effects (50% inhibition dose reduction by a factor of >10). Whenever the information on the effect of a given substitution was not available, we assigned it the number obtained with an analog that had a similar amino acid at that position, or if this information was not available, we arbitrarily extrapolated its most likely effect on the basis of the Dayhoff tables (11). For example, in position 1 the effect of a Val → Glu substitution was extrapolated from the effect of the closely related substitution Val → Asp, and therefore Glu was assigned a "substitution value" of 1; the same value was assigned to the Val → Trp substitution, based on Dayhoff tables, which show that a Val → Trp substitution is rather infrequently observed within a series of phylogenetically related proteins, and so on. The complete set of such "substitution values" is shown in Table 2. Values that were determined experimentally are underlined. For each particular six-residue sequence, the six substitution values corresponding to each position were then multiplied by each other to obtain a "motif number." A sequence yielding a motif number >400 was then operationally defined as an IA^d binding motif. This corresponds to defining a valid motif as one that contains either a peptide of all acceptable substitutions ($3^6 = 729$) or a single substitution with moderate deleterious effect ($3^5 \times 2 = 486$). Finally, whenever a peptide or protein sequence was analyzed, the motif numbers for all the possible six-residue sequences were calculated, and the sequence was scored as positive if at least one hexamer sequence with a motif value >400 was found. We have derived a simple microcomputer program that automatically performs these operations.

The approach that was used to define and detect an IE^d binding motif was simpler, reflecting the observation that the common feature shared by most of the good IE^d binders is the presence of multiple basic amino acids (refs. 3 and 4; A.S., L. Adorini, S.B., E.A., and H.M.G., unpublished data). The crucial IE^d interacting regions, as defined by truncation analysis in six unrelated peptides that are good IE^d binders, contain a basic residue (Arg, His, or Lys) at position 1, 2, or 3, and two more basic residues at positions 4 and 6. Position 5 is noncharged and tends to be hydrophobic (Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Ser, Thr, Val, Tyr are considered permissible). Thus, the IE^d motif is defined by a "basic-basic-noncharged-basic" motif (Table 3). Since this motif is relatively simple and "all-or-none," a computerized procedure was not necessary for its identification, and sequences were screened for the presence of the motif by simple visual inspection.

Using these procedures, we analyzed the set of peptides shown in Table 4 for the presence or absence of IA^d and IE^d motifs. These peptides, derived from various viral, prokaryotic, and eukaryotic proteins, have all been tested (except where noted) for binding to IA^d and IE^d molecules (refs. 1–6; A.S., L. Adorini, S.B., E.A., and H.M.G., unpublished data; our unpublished observations). It is apparent from Table 4 that a very strong correlation exists between the

Table 1. IA^d motifs of unrelated peptides with good IA^d binding capacity

Peptide	Residue number					
	1	2	3	4	5	6
Ova(323–339)	V	H	A	A	H	A
Ha(130–142)	V	T	A	A	C	S
Ha(187–206)	V	G	T	Y	V	S
Myo(63–78)	V	T	V	L	T	A
Myo(108–118)	I	H	V	L	H	S
$A^d\alpha$(6–20)	I	T	V	Y	Q	S
Nase(101–120)	V	R	Q	G	L	A
Nase(1–20)	A	T	L	I	K	A
Ova(317–327)	L	K	I	S	Q	A
p. cyt(88–104)	L	K	Q	A	T	A
CS(388–393)	A	K	M	E	K	A
λ rep(12–26)	A	R	R	L	K	A
HEL(74–86)	C	S	A	L	L	S
Ova(313–322)	L	S	G	I	S	S

One-letter codes for amino acids are shown. Ha, influenza hemagglutinin; Myo, sperm whale myoglobin; Nase, staphylococcal nuclease; p. cyt, pigeon cytochrome *c*; CS, malaria circumsporozoite; λ rep, λ repressor cI protein; HEL, hen egg white lysozyme.

Table 2. Substitution values for each of the six positions of the IA^d motif corresponding to the Ova(327–332) sequence (VHAAHA)

Amino acid	Position (residue)					
	1 (V)	2 (H)	3 (A)	4 (A)	5 (H)	6 (A)
A	3	2	3	3	3	3
C	2	2	2	2	3	2
D	1	2	2	2	2	1
E	1	2	2	2	2	1
F	1	1	2	2	2	1
G	1	2	3	2	1	2
H	1	3	2	2	3	1
I	3	2	3	3	3	2
K	1	3	1	2	3	1
L	3	2	3	3	3	2
M	2	1	3	3	2	2
N	1	2	2	2	1	1
P	2	1	2	3	2	2
Q	1	2	3	2	3	1
R	1	3	2	3	3	1
S	2	3	3	3	3	3
T	2	3	3	3	3	3
V	3	2	3	3	3	2
W	1	1	2	2	2	1
Y	1	2	3	3	3	1

Values represent the experimentally observed (underlined) or predicted (not underlined) effect on IA^d binding of substituting each of the 20 amino acids (single-letter code) in the six-amino acid core region of Ova(327–332) ranked as described in the text. Predicted values are obtained as described in the text. Experimental values are from ref. 2 and unpublished observations.

Table 3. IE^d motifs in unrelated peptides with good IE^d binding capacity

Peptide	Residue number					
	1	2	3	4	5	6
Myo(108–118)	H	V	L	H	S	R
Nase(111–130)	H	E	Q	H	L	R
HSV(8–23)		R	F	R	G	K
HEL(105–120)		R	N	R	C	K
HIV p 17(17–32)		K	Y	K	L	K
Dynorphin(1–13)		R	I	R	P	K
HIV p 17(9–24)		K	I	R	L	R
λ rep(12–26)			R	R	L	K
CS(325–341)			K	K	I	K

Underlined residues designate the IE^d motif (basic residue, basic residue, noncharged residue, basic residue) described in the text. Amino acids are designated by the single-letter code. HSV, herpes simplex virus gd protein. All other abbreviations are as in Table 1.

Table 4. Presence of IA^d and IE^d binding motifs in peptides tested for Ia^d binding

Peptide origin	Peptide	Peptide sequence	IA^d binding	IA^d motif	IE^d binding	IE^d motif
Chicken ovalbumin	Ova(111–122)	YPILPEYLQCVK	–	–	–	–
	Ova(312–322)	LSGISSAESLK	++	+	–	–
	Ova(307–317)	SSSANLSGISS	–	–	–	–
	Ova(323–336)	ISQAVHAAHAEINEAGR	++	+	–	–
	Ova(370–381)	HIATNAVLFFGR	–	–	–	–
	Ova(317–327)	SAESLKISQAV	++	+	ND	ND
Hen egg white lysozyme	HEL(34–65)	FESNFNTQATNR	ND	ND	–	–
	HEL(46–61)	NTDGSTDYGILQINSR	+	–	+	–
	HEL(74–86)	NLCNIPCSALLSS	++	+	–	–
	HEL(81–96)	SALLSSDITASVNCAK	+	+	–	–
	HEL(98–113)	IVSDGDGMNAWVAWRN	ND	ND	–	–
	HEL(105–120)	MNAWVAWRNRCKGTDV	–	–	++	+
	HEL(107–129)	AWVAWRNRCKGTDVQAWIRGCRL	+	+	ND	ND
	HEL(116–129)	KGTDVQAWIRGCRL	ND	ND	++	–
Cytochromes	m. cyt(88–103)	ANERADLIAYLKQATK	–	–	+	–
	p. cyt(88–104)	KAERADLIAYLKQATAK	++	+	–	–
	h. cyt(11–27)	VQKCAQCHTVEKGGKHK	+	–	ND	ND
Sperm whale myoglobin	Myo(63–78)	KHGVTVLTALGAILKK	++	+	+	–
	Myo(103–113)	YLEFISEAIIH	–	–	–	–
	Myo(108–118)	SEAIIHVLHSR	++	+	++	+
	Myo(115–125)	LHSRHPGDFGA	–	+	–	–
	Myo(132–153)	NKALELFRKDIAAKYKELGYQG	–	–	++	–
MHC molecules	IA^k_α(6–20)	ADHVGSYGITVYQSP	++	+	–	–
	IA^k_α(50–64)	PEFAQLRRFEPQGGL	–	–	–	–
	IA^d_β(40–58)	YDSDVGEYRAVTELGRPDA	–	–	–	–
	IA^d_β(64–80)	QPEILERTRAEVDTACR	–	–	–	–
	IA^d_α(60–76)	PQGGLQNIAAEKHNLGI	+	–	–	–
Malaria proteins	$(NANP)_3$	NANPNANPNANP	–	–	–	–
	Nonamer repeat	DGQPAGDRADGQPAGDRA	–	–	ND	ND
	CS(325–341)	EPSDKHIEQYLKKIKNS	–	–	++	+
	CS(382–394)	KIAKMEKASSVFNVVNS	++	+	–	–
Influenza proteins	Ha(111–120)	FERFEIFPKE	–	–	+	–
	Ha(121–131)	FPKESSWPNHN	–	–	–	–
	Ha(127–137)	WPNHNTNGVTA	–	–	ND	ND
	Ha(130–142)	HNTNGVTAACSHE	++	+	–	–
	Ha(136–146)	TAASSHEGKS	–	–	ND	ND
	Ha(187–206)	RTLYQNVGTYVSVGTSTLNK	++	+	ND	ND
	NP(147–158) Arg^-	TYQRTRALVTG	+	+	–	–
	NP(366–379)	ASNENMDAMESSTL	–	–	–	–
Herpes simplex virus	HSVgd(8–23)	SLKMADPNRFRGKDLP	–	–	++	+
	HSVgd(186–210)	IPPSACLSPQAYQQG	++	+	–	–
	HSVgd(245–260)	APYTSTLLPPELSETP	++	–	–	–
	HSVgd(204–219)	PQAYQQGVTDSIGML	++	–	–	–
λ repressor	λrep(12–26)	LEDARRLKAIYEKKK	++	+	++	+
HIV p_s	p13(97–112)	QKQEPIDKELYPLTSL	++	–	–	–
	p13(105–123)	ELYPLTSLRSLFGNDPSSQ	++	+	–	–
	p17(1–16)	GARASVLSGGELDKWE	–	+	–	–
	p17(9–24)	GGELDKWEKIRLRPGG	–	–	++	+
	p17(17–32)	KIRLRPGGKKKYKLKH	–	–	++	+
	p17(33–48)	IVWASRELERFAVNPG	–	–	+	–
	p17(57–73)	RQILGQLQPSLQTGSE	+	–	–	–
	p17(107–122)	EEQNKSKKKAQQAAA	–	+	–	–
	p17(115–135)	KAQQAAAAGTGNSSQVSQNY	+	+	–	–
	p25(33–48)	SPEVIPMFSALSEGAT	++	–	–	–
	p25(97–112)	REPRGSDIAGTTSTLQ	–	+	–	–
	p25(217–233)	ACQGVGGPGHKARVLA	–	–	–	+
Miscellaneous	Angiotensin	DRVYHIPFHLLVYS	+	–	–	–
	Dynorphin(1–13)	YGGFLRRIRPKLK	–	–	++	+
	$(YT)_4$VHAAH	YTYTYTYTVHAAH	+	–	–	–
	JF39	AHRVQLGPRSLQVLLIP	++	–	–	–
	$(YT)_6$	YTYTYTYTYTYT	–	–	ND	ND
	Bradykinin potentiator B	EGLPPRPKIPP	–	–	–	–

Binding values are from refs. 1–6; A.S., L. Adorini, S.B., E.A., and H.M.G., unpublished data; and our unpublished observations. Relative binding figures are derived as described in *Materials and Methods*. Peptides were ranked for binding capacity as follows: ++, good binders (relative binding, >0.1); +, intermediate binders (relative binding, 0.01–0.1); –, negative binders (relative binding, <0.01); ND, not determined. Peptides were also scored as positive or negative for the presence of the IA^d and IE^d binding motifs using the procedures described in the text. m. cyt, moth cytochrome *c*; h. cyt, horse cytochrome *c*; HIV p_s, human immunodeficiency virus gag proteins. All other abbreviations are as in Tables 1 and 3.

binding properties of this set of peptides for IA^d and IE^d and the presence of the appropriate motif. In the case of the IA^d motif, the motif was present in 14 of 19 good binders (74%), 4 of 10 intermediate binders (40%), and 4 of 30 nonbinders (13%). In the case of IE^d, the motif was present in 8 of 10 good binders (80%), 0 of 5 intermediate binders, and 1 of 39 negative binders (3%).

Most of the peptides analyzed herein were synthesized and studied for MHC binding because they had been previously characterized as being T-cell epitopes. Consequently, we were interested in expanding the motif analysis to a more random set of peptides. For this purpose, a set of overlapping peptides encompassing the entire staphylococcal nuclease protein was examined (Table 5). In this set of peptides, 2 of 2 good IA^d binders and 3 of 4 intermediate binders (75%) contained the IA^d motif, while 0 of 8 nonbinders contained the IA^d binding motif. When the same set of peptides was analyzed for IE^d binding and presence of IE^d motifs, the 1 of 1 good binder and 2 of 4 intermediate binders (50%) contained the IE^d binding motif, and 2 of the 9 nonbinders (22%) contained the IE^d binding motif.

Thus, the data for this unbiased set of peptides were consistent with those obtained with the set of T-cell epitopic peptides and supported the strong correlation between the Ia binding capacity of a peptide and the presence of the associated amino acid sequence motif. As shown in Table 6, in which the data in Tables 4 and 5 are combined, the fraction of good IA^d and IE^d binders containing the motifs was in the 75–80% range, and in the 20–50% range for intermediate binders, whereas only ≈10% of the nonbinders were found to contain the IA^d and IE^d motifs.

These data were also compiled in a different form, which considered only peptides that were positive for a motif and analyzed these for binding to the relevant MHC molecule (Table 6). This tabulation, although disregarding most of the peptides, is of particular interest because it demonstrates the value of using these motifs to predict MHC binding regions and immunogenic peptides from a protein sequence. When this analysis was performed, the overall predictive success rates were 85% and 79% for IA^d and IE^d, respectively.

Table 6. Distribution of IA^d and IE^d motifs in peptide sets of different binding capacities and effectiveness of motifs in predicting Ia^d binding

Peptide set	Presence of the motif for IA^d		Presence of the motif for IE^d	
Distribution of motifs				
Good binders	16/21	(76%)	9/11	(82%)
Intermediate binders	7/19	(50%)	2/9	(22%)
Nonbinders	4/38	(11%)	3/48	(6%)
P value (χ^2)	<0.001		<0.001	
Fraction of motif-positive peptides capable of binding*				
Original peptide collection (based on data from Table 4)	18/22	(82%)	8/9	(89%)
Nase (based on data from Table 5)	5/5	(100%)	3/5	(60%)
Overall	23/27	(85%)	11/14	(79%)

Good, intermediate, and nonbinders are as defined in the Table 4 legend. Nase, staphylococcal nuclease.
*Good (relative binding, >0.1) and intermediate (relative binding, 0.01–0.1) binding peptides were scored as positive in this compilation.

DISCUSSION

In the present report it is shown that most peptides capable of binding a particular Ia^d isotype (A^d or E^d) share structural motifs detectable at the level of primary amino acid sequence (Tables 1 and 3) and that most peptides that do not bind lack these structural features. The definition of these motifs is quite broad, and at any given residue position at least three and usually more amino acids are compatible with good Ia binding. These data are in keeping with the concept that an Ia–antigen interaction must demonstrate broad specificity for a few Ia specificities to be capable of binding a large number of peptide antigens. Despite this broad specificity, such motifs can be used quite successfully to predict peptides that are capable of interacting with a given Ia^d molecule. Such predictions could be of considerable interest in an attempt to select synthetic vaccine candidates, to identify possible autoantigens in self molecules, and to select strong Ia binders that might function as specific inhibitors of MHC-associated autoimmune diseases.

With the data base of peptides used in this study, it is of interest to compare the predictability of the above motifs with the predictions based on two other published procedures by Berzofsky and associates (8, 9) and by Rothbard and Taylor (7). These procedures are based on the detection of amphipathic α-helical segments (8) and on sequence pattern recognition (7), respectively, and are aimed at predicting T-cell epitopes, disregarding specificity at the level of isotype or allele of restriction element, or even of the animal species analyzed.

When these algorithms were examined for their ability to predict peptide binders to a single Ia^d isotype (A vs. E), there was no statistically significant association between binding and the presence of the structural feature defined by the algorithm. A significant difference between binding and nonbinding peptides could only be detected if the data were

Table 5. Presence of IA^d and IE^d binding motifs in a nested set of overlapping peptides encompassing the entire staphylococcal nuclease (Nase) molecule

Peptide	Peptide sequence	IA^d binding	IA^d motif	IE^d binding	IE^d motif
Nase(1–20)	ATSTKKLHKEPATLIKAIDG	++	+	−	+
Nase(11–30)	PATLIKAIDGDTVKLMYKGQ	+	+	−	−
Nase(21–40)	DTVKLMYKGQPMTFRLLLVD	−	−	−	−
Nase(31–50)	PMTFRLLLVDTPETKHPKKG	−	−	+	+
Nase(41–60)	TPETKHPKKGVEKYGPEASA	−	−	−	+
Nase(51–70)	VEKYGPEASAFTKKMVENAK	−	−	+	−
Nase(61–80)	FTKKMVENAKKIEVEFDKGQ	+	−	−	−
Nase(71–90)	KIEVEFDKGQRTDKYGRGLA	−	−	−	−
Nase(81–100)	RTDKYGRGLAYIYADGKMVN	+	+	−	−
Nase(91–110)	YIYADGKMVNEALVRQGLAK	+	+	−	−
Nase(101–120)	EALVRQGLAKVAYVYKPNNT	++	+	+	−
Nase(112–130)	AYVYKPNNTHEQHLRKSEA	−	−	++	+
Nase(121–140)	HEQHLRKSEAQAKKEKLNIW	−	−	+	+
Nase(131–149)	QAKKEKLNIWSEDNADSGQ	−	−	−	−

accumulated to determine the predictability of a peptide being a binder to Ia^d regardless of isotype specificity. With this analysis, the amphipathic α-helix motif was present in 51% of the Ia^d binders and in 27% of the nonbinders ($P < 0.05$). Similarly, Rothbard and Taylor's motif was present in 59% of the Ia^d binders but also in 35% of the nonbinders ($P < 0.05$).

The failure of these algorithms to predict specific MHC–peptide interactions is perhaps not surprising, considering the fact that these two algorithms were derived by considering T-cell immunogenicity and not MHC binding, and although immunogenicity is dependent on MHC binding, immunogenicity of a peptide clearly has structural features besides those that are required for MHC binding (5). It would thus appear that, although these algorithms could be of some use in predicting T-cell epitopes, specific predictive procedures described herein are likely to be more useful, since in addition to predicting epitopes, they are capable of identifying possible autoantigens or potential MHC-specific blockers of autoimmune diseases.

When the peptides used in this study were analyzed to determine whether there was a positive correlation between IA^d and IE^d binding capacities, none could be detected (data not shown). This is consistent with the different structural requirements of IA^d and IE^d in their interaction with peptide molecules and underlines the importance of predictive procedures specific for each individual restriction element. Moreover, MHC recognition of distinct patterns in peptide molecules is also consistent with the notion that most (if not all) of each protein molecule can be immunogenic provided that enough different MHC specificities are examined (12). It could be argued that this concept also makes biological sense, in that if a binding motif common to all MHC molecules were to exist, bacteria and viruses could escape from all immune surveillance simply by evolving to remove the binding motif from their protein sequences.

In summary, the data presented herein suggest that it is possible to predict successfully $\approx$75% of the strong Ia^d binding peptides, while for $\approx$20% of the good binders the presence of an Ia^d binding motif could not be detected. This could be due to an inability to recognize the Ia^d binding motif in some of the peptides, or, as already suggested (2), some peptides might bind to Ia^d in different conformations, which might in turn require distinct structural motifs. Further experiments will be required to analyze these possibilities, to ascertain the efficiency of these procedures in predicting the Ia binding characteristics of peptide antigens, and to determine whether similar procedures could also be applied to other MHC molecules.

The expert secretarial assistance of Joyce Joseph is gratefully acknowledged. We wish to thank F. Sinigaglia, V. Nussenzweig, J. Freed, and G. Corradin for the generous gift of some of the peptides whose Ia^d binding capacity is included in this study. This work was supported in part by U.S. Public Health Service Grants A109758 (H.M.G.), AI18634 (H.M.G.), AI25280 (R.C.), CA49394 (R.C.), and Hoechst Aktiengesellschaft, Frankfurt/Main, F.R.G. (J.A.S.).

1. Sette, A., Buus, S., Colon, S., Smith, J. A., Miles, C. & Grey, H. M. (1987) *Nature (London)* **328,** 395–399.
2. Sette, A., Buus, S., Colon, S. M., Miles, C. & Grey, H. M. (1988) *J. Immunol.* **141,** 45–48.
3. Adorini, L., Sette, A., Buus, S., Grey, H. M., Darsley, M., Lehmann, P., Doria, G., Nagy, Z. & Appella, E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5181–5185.
4. Sette, A., Buus, S., Colon, S., Miles, C. & Grey, H. M. (1989) *J. Immunol.* **142,** 35–40.
5. Buus, S., Sette, A., Colon, S., Miles, C. & Grey, H. (1987) *Science* **235,** 865–870.
6. Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. (1986) *Cell* **47,** 1071–1077.
7. Rothbard, J. & Taylor, W. R. (1988) *EMBO J.* **7,** 93–101.
8. DeLisi, C. & Berzofsky, J. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7048–7053.
9. Margalit, H., Spouge, J. L., Cornette, J. L., Cease, K. B., DeLisi, C. & Berzofsky, J. A. (1987) *J. Immunol.* **138,** 2213–2229.
10. Bailey, N. T. (1981) *Statistical Methods in Biology* (Wiley, New York), 2nd Ed.
11. Dayhoff, M. O. & Eck, R. V. (1968) *Atlas of Protein Sequences and Structure* (Natl. Biomed. Res. Found., Washington, DC).
12. Gammon, G., Shastri, N., Cogswell, J., Wilbur, S., Sadegh Naneri, S., Kaych, V., Miller, A. & Sercarz, E. (1988) *Immunol. Rev.* **98,** 53–73.