ELSEVIER

Biochimica et Biophysica Acta 1316 (1996) 85–101

BBA Biochimica et Biophysica Acta

Natural ligand motifs of closely related HLA-DR4 molecules predict features of rheumatoid arthritis associated peptides

Thomas Friede [a], Volker Gnau [b], Günther Jung [b], Wieland Keilholz [a], Stefan Stevanović [a], Hans-Georg Rammensee [a,*]

[a] *Department of Tumor Virus Immunology (0620), German Cancer Research Center, Im Neuenheimer Feld 242, D-69120 Heidelberg, Germany*
[b] *The Institute of Organic Chemistry, University of Tübingen, Auf der Morgenstelle 18, D-72076 Tübingen, Germany*

Received 28 July 1995; accepted 30 January 1996

Abstract

Rheumatoid arthritis (RA), one of the most common autoimmune disorders, is believed to be mediated via T lymphocytes and genetic studies have shown that it is strongly associated with HLA-DR4. The DR4 subtypes DR4Dw4, DR4Dw14 and DR4Dw15 represent increased risk factors for RA, whereas DR4Dw10 is not associated with the disorder. Our study determines and compares the natural ligand motifs of these MHC class II molecules and identifies 60 natural ligands. At relative position 4 (P4), only the RA-associated DR4 molecules allow, or even prefer, negatively charged amino acids, but do not allow those which are positively charged (Arg, Lys). In the case of DR4Dw10 the preference for these amino acids is reversed. The results predict features of the putative RA-inducing peptide(s). A remarkable specificity, almost exclusively for negative charges (Asp, Glu), is found at P9 of the DR4Dw15 motif. This specificity can be ascribed to amino acid β57 of the DRβ chain, and gives an important insight into the β57-association of another autoimmune disease, insulin-dependent diabetes mellitus type I.

Keywords: Rheumatoid arthritis; Autoimmunity; Peptide; Peptide motif; T cell epitope prediction; HLA-DR; MHC class II

1. Introduction

Many autoimmune diseases are associated with particular HLA class II alleles [1]. Rheumatoid arthritis, for example, occurs with relative high frequency in individuals expressing HLA-DR4. Among the DR4 alleles, DR4Dw4, Dw14 and Dw15 are associated with this disorder to a variable degree, whereas DR4Dw10 is not, only weakly, or only in restricted populations [2–5]. Sequence differences among these DR4 alleles affect only few amino acids [6,7], most of which are located around the peptide binding groove. Thus, the development of rheumatoid arthritis might be dependent on $CD4^+$ T cells recognizing an epitope, or epitopes, presented by the disease-associated DR4 molecules but not by the non-associated ones. Such epitopes could be derived from normal self proteins, such as collagen or other joint-associated proteins, or from micro-organisms present locally or systemically [1,8,9]. In any case, such epitopes should be central for disease development and — once identified — for prevention, diagnosis and therapy of autoimmune disease.

One way to gain information on such potentially disease-associated T-cell epitopes is to study carefully the peptide specificity of the disease-associated HLA molecules. The analysis of the differences in peptide fine specificity of disease-associated and non-associated allelic forms of HLA molecules should be particularly informative. We chose to analyse HLA-DR4Dw4, DR4Dw14 and DR4Dw15 as molecules highly associated with rheumatoid arthritis, and DR4Dw10 as a closely related molecule which is not, or only weakly associated.

The analysis of natural ligands can be a convenient and successful method of obtaining information on the peptide specificity of MHC class II molecules [10]. Pool sequencing of natural ligand mixtures of DR molecules indicated allele specific motifs, which can be confirmed and refined by alignment of individual ligands. The advantage of this approach is that it is possible to detect not only the influences of MHC peptide binding specificity but also the effects of antigen processing on the natural ligands. On the

* Corresponding author. Present address: Institute of Cell Biology, Department of Immunology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany. Fax: +49 7071 295653.

PII S0925-4439(96)00010-5

other hand, obtaining motif information solely by attempting to align individual ligands without pool sequences has proved to be difficult [11–15]. The more common approach of relying purely on peptide binding assays for motif determination (reviewed in Ref. [16]) has a disadvantage because it is biased towards peptide binding only, and does not take into account the effects of antigen processing. Furthermore, the peptide binding approach only provides partial information since the number of peptide variations required to obtain complete information is usually too high to be practical, with the exception of the ingenious phage peptide library approach [16–18].

In addition to pool sequencing and analysis of individual natural DR4 ligands, we took advantage of the recently published three-dimensional structure of multipeptide/DR1 and monopeptide/DR1 complexes [19,20]. X-ray analysis of such molecules indicated 5 pockets, each of which has a certain specificity for peptide ligand side chains. The specificity of the more restricted DR1 pockets (at relative positions 1, 4, 6 and 9) had been deduced both from peptide binding studies [18] as well as by ligand sequencing analysis [10], and can be well explained by the characteristics of the corresponding pocket with regard to size, charge and hydrophobicity. Comparing the nature of the putatively corresponding pockets in the four DR4 molecules under study with our sequencing results gave a surprisingly clear and detailed insight into the peptide presentation requirements of these DR4 molecules, making it possible to predict characteristics of putative RA-epitopes. While this manuscript was in preparation, Hammer et al. [21] reported a similar conclusion based on a very different technique, the peptide library approach.

In addition to those RA-related motif features, we report a peculiar specificity at P9 of the DR4Dw15 motif that is of importance for the functional consequences of the β57-polymorphism and its association with IDDM.

The motifs reported in the present paper are already included in recent reviews [22,23].

2. Materials and methods

2.1. Cell lines

The following homozygous EBV-transformed human B-cell lines were used to isolate the different HLA-DR4 molecules: LRM (A2; B44; DRB1*0401; DQA1*03; DQB1*0301; DP4), Al 10 (A26; B38; DRB1*0402; DQA1*03; DQB1*0302; DP4), MFF (A11,31; B60; Bw6; Cw3; DRB1*0404), WTA (DRB1*0405; DQA1*03; DQB1*0401). PCR typing was performed by J. Mytilineos. Serological and cellular typing results are from the HLA defined collection, European Collection of Animal Cell Cultures (ECACC), Salisbury; and from G. Pawelec. Due to its haplotype family, HLA-DR4 is tightly linked to HLA-DR53 (DRB4*0101). Since it is known that DRB1-genes are transcribed five to ten times higher than the respective DRB4-genes [24], the peptide contribution of DR53 molecules, expected to co-precipitate with DR4 molecules in our experiments, should be small or negligible.

2.2. Antibodies

Immunoprecipitations were performed with L243 antibodies [25] specific for HLA-DR. Hybridomas were grown in roller culture. Supernatant was precipitated with ammonium sulfate and the antibodies were further purified on a Protein A-Sepharose column (Pharmacia, Freiburg, Germany).

2.3. Cell culture

For bulk culture, cells were grown at 37°C in 2 l glass roller bottles (Schott Duran, Frankfurt/Main, Germany) with up to 1 l Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum (FCS), 5% newborn calf serum (NCS), glutamine and 2-mercaptoethanol. Cells were split every three to four days. For harvest, (typically 24–28 l) cell suspensions were pelleted with a continuous flow rotor (Beckman, Munich, Germany). Cell pellets were either used directly for peptide extraction or stored at −70°C.

2.4. MHC precipitation and peptide extraction

This was done as described [26]. Briefly, pellets (20–40 ml) were lysed in 1% (v/v) NP-40, 1mM PMSF, 0.02% (w/v) pepstatin, 0.02% (w/v) leupeptin and 0.02% (w/v) aprotinin (Boehringer Mannheim, Germany). The lysate was ultracentrifuged for 65 min at 86,000 g. The supernatant was passed at least twice over Sepharose beads (Pharmacia, Freiburg, Germany) coupled with glycine and then over Sepharose beads coupled with L243 antibodies. Peptides were acid-released from the loaded beads by treatment with 0. 1% (v/v) trifluoroacetic acid (TFA). Glycine beads were treated the same way, using extracted material as a control for molecules sticking nonspecifically to beads (see Fig. 1). In order to remove antibody molecules and HLA-DR α- and β-chains from the peptides, the solution was passed through ultrafiltration membranes (Centricon 10, Amicon, Witten) of 10 kDa exclusion size. The filtrate was separated on a reversed-phase HPLC column (μRPC C2/C18; 2.1 × 100 mm; Pharmacia), using a SMART system (Pharmacia). Eluent A was 0.1% (v/v) TFA in water, Eluent B was 0.081% (v/v) TFA in 80% acetonitrile/water. All eluents used were of HPLC grade (Riedel de Haen, Frankfurt, Germany). The gradient used was the same as previously described [10]. Peptide fractions were pooled, whereby dominant single peaks were omitted. Pools and dominant single peaks were sequenced separately.

2.5. Peptide analysis

Pools and single peaks were sequenced by Edman degradation on a pulsed-liquid protein sequencer 476A with on-line PTH-amino acid analysis (Applied Biosystems, Weiterstadt, Germany) as described [27]. Cys was not modified, and therefore not detectable.

2.6. Interpretation of pool sequencing

In the raw data tables, amino acids are ordered according to their hydropathy [28], which facilitates perception of certain physico-chemical groups clustering in cycles of Edman degradation. Two arbitrary levels of significance were used for evaluation of data. An increase in the amount of a given PTH amino acid at a given cycle of more than 50% as compared to the previous cycle, is considered highly significant as indicated by **bold** typing. An increase of less than 50% is considered to reflect a significant signal and is indicated by *italic* typing the respective numbers. In some instances, a decrease smaller than the lag effect usually found for a given amino acid is taken as an indication that some of the peptides in the pool have this amino acid. Such values are also *italicised.*

3. Results

3.1. Analysis of DR4Dw4 eluted peptides

Large batches of the EBV-transformed B-cell line LRM (DRB1*0401) were grown in liquid cell culture. DR molecules were immunoprecipitated using L243 antibodies. A potential complication here is the co-expression of HLA-DR4 linked HLA-DR53 genes. However, since DR53 expression is usually much lower than that of DR4, peptide contribution of DR53 in our experiments should be small or negligible. (See also Section 2). DR-associated peptides were extracted by TFA-treatment and separated by HPLC (Fig. 1A). Material eluting in large distinct peaks was collected into individual fractions and sequenced by Edman degradation. The remaining fractions (as indicated in

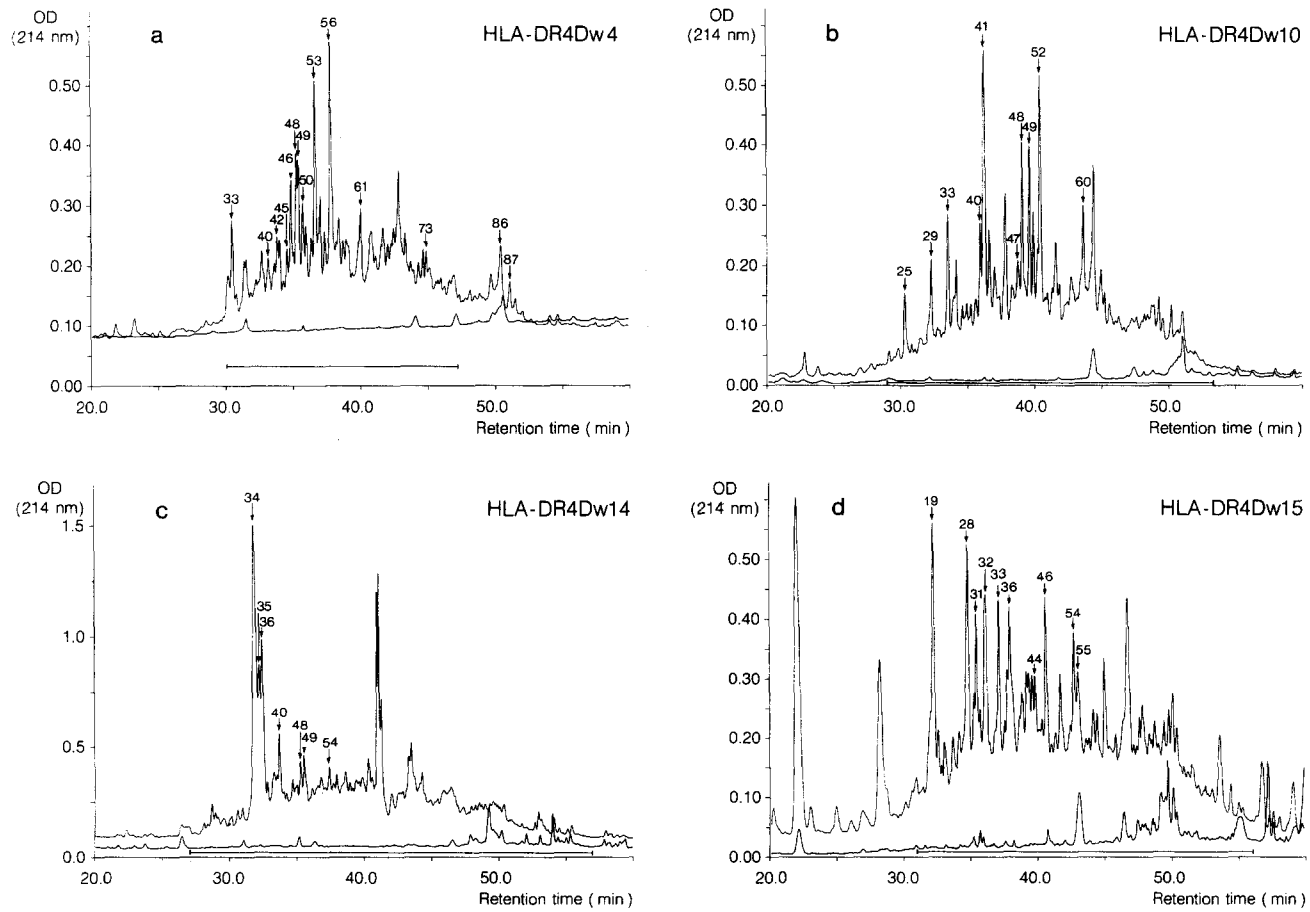

Fig. 1. HPLC separation of DR4 ligands. Peptides dissociated from immunoprecipitated DR4Dw4 (a), DR4Dw10 (b), DR4Dw14 (c) and DR4Dw15 (d) molecules were separated by reversed phase HPLC. The arrows indicate peaks collected and sequenced individually. Fraction numbers of these peaks are also indicated in Tables 2, 4, 6, and 8. The horizontal bars indicate the fractions pooled for pool sequencing. Numbered peaks, however, were omitted from these pools. The lower curve shows the extinction of the glycine control at 214 nm.

Table 1
HLA-DR4Dw4 pool sequencing

Cycle of Edman degradation	Amino acid residues (in pmol) [a]																		
	I	V	L	F	W	M	A	G	T	S	Y	P	H	Q	N	D	S	K	R
1	139.4	290.4	217.6	162.6	17.3	39.7	166.3	–	–	120.0	114.7	29.7	31.3	27.5	27.4	125.4	173.3	96.3	35.6
2	64.2	160.5	133.9	29.5	7.6	18.7	69.9	–	–	–	32.8	**155.8**	–	–	7.2	44.4	–	94.2	–
3	63.7	111.7	*136.9*	**68.2**	*8.4*	**28.4**	**121.1**	–	–	70.6	**80.6**	121.7	35.6	98.7	**44.8**	**122.7**	*221.7*	**151.3**	69.0
4	58.7	96.1	107.0	**205.0**	**14.9**	**30.8**	103.8	–	–	55.6	**143.2**	96.7	30.8	70.8	**56.1**	111.6	191.3	129.2	*88.2*
5	*75.1*	**140.4**	97.5	160.6	**18.6**	**33.8**	102.5	–	–	37.7	**151.3**	66.6	23.7	**144.3**	49.2	111.2	169.6	72.4	**101.1**
6	**98.4**	120.8	94.1	*187.1*	**19.7**	30.5	*118.1*	–	–	*44.9*	101.0	59.8	21.7	65.0	46.4	83.7	127.8	**110.3**	**117.4**
7	87.0	**168.7**	*101.9*	107.4	14.4	20.7	**147.7**	–	–	33.3	55.7	49.1	15.1	52.8	45.6	*96.2*	*179.2*	**123.9**	100.6
8	51.7	92.9	85.7	62.6	10.3	20.1	**162.3**	–	–	33.1	35.0	*50.1*	11.9	*75.2*	41.0	*96.2*	**191.3**	80.6	**154.2**
9	*52.9*	*96.1*	72.1	**117.2**	8.2	*25.9*	139.5	396.6	61.3	*43.8*	26.0	*51.0*	*13.8l*	*50.5*	**61.2**	58.5	166.9	56.7	98.7
10	37.3	78.0	*76.5*	41.0	6.3	19.7	116.7	220.3	*60.1*	*46.7*	23.1	*50.7*	**17.7**	38.3	**75.8**	51.8	*231.3*	42.5	66.7
11	34.6	55.2	69.0	23.6	3.7	*23.7*	*170.3*	121.4	40.2	*54.6*	*24.4*	*57.5*	**21.0**	*51.3*	46.6	46.7	126.9	42.3	58.5
12	*42.7*	41.4	*70.1*	17.4	2.7	16.9	160.7	90.4	28.8	38.3	20.3	40.6	14.8	**63.5**	27.8	43.9	*134.2*	37.7	*72.3*
13	33.5	33.5	53.8	13.7	1.9	*21.9*	*172.8*	78.0	22.0	33.3	17.4	29.6	11.1	**76.2**	27.1	35.8	67.6	*50.0*	59.6
14	30.8	29.5	42.8	*18.6*	2.2	12.4	139.6	57.0	*25.1*	25.9	16.5	27.3	9.9	58.4	25.5	35.5	49.7	*52.1*	51.7
15	21.7	23.1	29.8	*18.2*	1.6	8.6	65.0	44.7	23.9	*35.1*	16.3	*29.1*	9.0	43.0	18.3	28.0	43.3	36.8	51.1
16	13.9	16.7	20.7	12.6	1.9	**13.3**	37.6	34.6	16.8	19.1	12.0	26.6	7.4	*59.3*	14.2	27.3	34.6	26.4	41.4
17	9.3	14.6	15.8	7.8	1.2	9.8	27.6	27.8	11.8	11.1	7.3	21.4	6.0	32.1	10.1	22.6	23.6	17.4	**63.8**
18	6.6	11.2	12.1	5.7	1.0	**27.1**	20.7	25.4	9.2	9.4	7.1	16.9	5.5	18.5	7.9	15.9	20.3	13.2	42.6

[a] Amino acids are ordered according to their hydropathy [28]. Bold numbers indicate increases of more than 50% as compared to the pmol amounts of the same amino acid at the previous position. Italicized numbers indicate smaller increases or decreases lower than expected from tailing effects. A dash indicates technical failure to detect a particular amino acid.

Table 2
HLA-DR4Dw4 motif and ligands

	Relative position										Source	Fraction
	1	2	3	4	5	6	7	8	9			
Anchor or preferred residues [a]	**F, Y, W,** I, L, V, M			F, W, I, L, V, A, D, E **no R, K**		N, S T,Q H,R	polar, charged, aliphatic		polar aliphatic K			
	1	2	3	**4**	5	**6**	**7**	8	**9**			
VDDTQ	**F**	V	R	**F**	D	**S**	**D**	A	**A**	SQRMEP...	HLA-A2 (28–?)	53, 56
DTQ	**F**	V	R	**F**	D	**S**	**D**	A	**A**	SQR	HLA-A2 (30–44)	40
	F	V	R	**F**	D	**S**	**D**	A	**A**	SQR	HLA-A2 (33–44)	33
	F	V	R	**F**	D	**S**	**D**	A	**A**	SQRMEP	HLA-A2 (33A7)	42, 49
VDDTQ	**F**	V	R	**F**	D	**S**	**D**	A	**A**	SQRM	HLA-A2 (28–45)	46, 55
	F	V	R	**F**	D	**S**	**D**	A	**A**	SQRM	HLA-A2 (33–45)	45
VDDTQ	**F**	V	R	**F**	D	**S**	**D**	A	**A**	SPRGEP...	HLA-C (28–?)	50, 54
DGKD	**Y**	I	A	**L**	N	E	**D**	L	**S**	S	HLA-B44 (143–156)	
LSS	**W**	T	A	**A**	D	**T**	**A**	A	**Q**	ITQ	HLA-B44 (154–168)	48
LSS	**W**	T	A	**A**	D	**T**	**A**	A	**Q**	IT	HLA-B44 (154–167)	49
IY	**F**	R	N	**Q**	K	G	**S**	H	**S**	GLQPTGFL	HLA-DR4β (252–270)	61
DVA	**F**	V	K	**D**	Q	**T**	**V**	I	**Q**	NTD	bovine transferrin (68–82)	40
YDHN	**F**	V	K	**A**	I	**N**	**A**	I	**Q**	KSW	cathepsin C (170–185)	73, 55
KHKV	**Y**	A	C	**E**	V	**T**	**H**	Q	**G**	...	Igκ chain C region (80–?)	
HKV	**Y**	A	C	**E**	V	**T**	**H**	Q	**G**	L...	Igκ chain C region (81–?)	
DGP	**F**	R	I	**I**	T	V	**P**	A	**A**	LDY	unknown	86, 87
TGN	**Y**	R	I	**E**	S	V	**L**	S	**S**		Sphingolipid activator protein 3 (165–176)	
GERA	**M**	T	K	**D**	N	**N**	**L**	L	**G**	...	HSC 70 (445–?) [b]	
XXX	**Y**	E	X	**A**	L	**S**	**L**	P	**S**	K...	no match	
GSLF	**V**	Y	N	**I**	T	**T**	**N**	K	**Y**	KAFLKQ	VLA-4 (229–247)	(Chicz et al., 1993)
SPEDF	**V**	Y	Q	**F**	K	G	**M**	C	**Y**	F	HLA-DQβ3 3.2 chain (24–38)	(Chicz et al., 1993)
AAPYEKEVP	**L**	S	A	**L**	T	**N**	**I**	L	**S**	AQL	PAI-I (261–281)	(Chicz et al., 1993)
GVYF	**Y**	L	Q	**W**	G	**R**	**S**	T	**L**	VSVS	Ig heavy chain (121–?)	(Chicz et al., 1993)
AEALERM	**F**	L	S	**F**	P	**T**	**T**	K	**T**		bovine hemoglobin (26–41)	(Chicz et al., 1993)

[a] Deferred from pool sequence, individual ligands, and pocket structure.
[b] HSC 70, heat-shock cognate protein 70; PAI-1 plasminogen activator inhibitor 1; VLA-4, a cell surface heterodimer in the integrin superfamily of adhesin receptors.
[c] Particles without fraction number are from an experiment other than that shown in Fig. 1.

Table 3
HLA-DR4Dw10 pool sequencing

Cycle of Edman degradation	Amino acid residues (in pmol) [a]																	
	I	V	L	F	M	A	G	T	S	Y	P	H	Q	N	E	D	K	R
1	143.2	102.8	148.2	111.9	22.0	83.0	160.5	23.0	28.5	52.1	17.1	6.3	14.9	10.7	36.8	39.7	22.0	11.2
2	76.8	*106.1*	74.5	54.0	14.0	**143.1**	91.8	*29.2*	*33.2*	45.0	**136.7**	*9.0*	**33.0**	**26.8**	*47.2*	**57.2**	**53.6**	**25.0**
3	*96.8*	70.8	*85.4*	*59.9*	*15.8*	65.1	**134.6**	22.0	31.3	43.6	75.0	**15.8**	**62.4**	**39.3**	**69.6**	**99.0**	**89.9**	**60.1**
4	**148.1**	67.2	**164.0**	41.4	15.5	55.4	84.2	16.2	22.5	26.3	62.7	14.8	48.3	**47.3**	54.1	83.7	79.9	**71.4**
5	123.1	*75.0*	114.2	*42.3*	14.9	*67.6*	66.6	**24.7**	22.2	23.8	41.8	12.4	*50.7*	39.1	44.0	70.0	74.1	**99.4**
6	89.7	73.0	106.4	26.4	14.7	58.7	62.7	18.5	20.1	21.1	35.2	**19.7**	46.3	34.7	37.5	53.6	**114.0**	93.3
7	75.4	52.2	*113.2*	*29.6*	**25.9**	53.0	53.5	17.7	*22.8*	*21.6*	*35.7*	**30.2**	44.2	27.5	36.6	43.7	87.0	93.1
8	43.9	*52.8*	88.8	*30.4*	19.9	51.1	53.3	13.4	17.0	16.0	*43.2*	**33.2**	33.3	*28.9*	35.1	39.5	68.5	*114.1*
9	28.4	38.8	70.1	16.8	12.5	44.8	*59.2*	**24.6**	*33.2*	12.3	34.7	24.0	28.7	**45.4**	32.4	*51.3*	44.2	92.7
10	26.1	31.1	46.2	*22.5*	8.3	43.0	52.4	21.5	27.0	*17.0*	21.2	**45.0**	**48.1**	38.4	26.3	40.9	42.1	86.0
11	19.2	29.1	*51.0*	16.7	*8.7*	*52.5*	44.5	19.6	19.9	16.4	*24.2*	35.6	33.7	*45.2*	24.5	32.5	*48.0*	*87.0*
12	18.4	26.7	41.5	10.4	6.6	51.6	*47.8*	13.8	18.3	15.6	19.0	24.3	**74.1**	37.8	*25.4*	24.1	38.4	59.3
13	14.2	24.4	25.7	**25.1**	6.4	42.8	35.7	9.9	15.6	*19.4*	*20.2*	16.4	54.3	29.2	22.0	20.5	32.5	40.7
14	11.0	21.9	20.1	23.9	3.7	*50.3*	26.3	*12.6*	15.4	15.7	*20.3*	10.6	39.0	21.6	*19.0*	19.0	30.2	*44.7*
15	8.7	15.5	14.4	17.4	3.5	38.6	25.0	11.0	10.8	**30.3**	16.9	7.3	25.4	16.0	14.7	16.9	22.6	37.4
16	6.7	11.6	10.9	16.5	2.8	27.8	20.3	5.9	7.8	24.0	11.6	5.8	16.6	11.9	11.2	*25.1*	19.5	27.6
17	4.6	8.2	10.0	10.6	2.4	18.2	*23.6*	4.2	5.2	18.1	11.1	4.4	11.9	7.5	8.8	22.6	13.1	20.0
18	3.9	6.3	8.3	6.0	2.3	11.6	19.2	4.0	4.7	11.0	9.0	3.7	8.5	6.1	6.8	18.8	*16.2*	14.7
19	2.8	5.2	6.4	3.6	1.7	8.1	15.5	2.5	3.8	6.3	6.7	2.8	6.8	4.5	5.5	13.8	14.0	10.6
20	2.4	4.2	4.9	2.5	1.4	6.5	10.8	1.6	3.2	4.3	5.6	2.0	5.3	3.3	4.2	12.1	12.2	8.5

[a] Organization of data as in Table 1.

Table 4
HLA-DR4Dw10 motif and ligands

	Relative position										Source	Fraction
	1	2	3	**4**	5	**6**	**7**	8	**9**			
Anchor or preferred residues	**V, I L, M**			Y, F, W I, L, M R, N, H **no D, E**		N, Q S, T, K	R, K H, N Q, P; rare D, E		polar, aliphatic, H			
	1	2	3	**4**	5	**6**	**7**	8	**9**			
GPDGR	**L**	L	R	G	H	**N**	**Q**	F	**A**	YDGKDY...	HLA-B38 (128–?)	40
GPDGR	**L**	L	R	G	H	**N**	**Q**	F	**A**	YDGKD	HLA-B38 (128–146)	
GPDGR	**L**	L	R	G	H	**N**	**Q**	F	**A**	YDGK	HLA-B38 (128–145)	
GPDGR	**L**	L	R	G	H	**N**	**Q**	F	**A**	YDG	HLA-B38 (128–144)	33
GR	**L**	L	R	G	H	**N**	**Q**	F	**A**	YDGK	HLA-B38 (131–145)	29
I	**I**	K	G	V	R	**K**	**S**	N	**A**	AERRG	HLA-DR α (238–252)	25
	I	Y	F	**R**	N	**Q**	**K**	G	**H**	SGLQPTGFLS	DR4β (248–266)	52
	I	Y	F	**R**	N	**Q**	**K**	G	**H**	SGLQP	DR4β (248–261)	
			F	**R**	N	**Q**	**K**	O	**H**	SGLQP	DR4β (250–261)	
F	**I**	Y	F	**R**	N	**Q**	**K**	G	**H**	SGLQPTGFLS	DR4β (249–266)	60
LPKPPKPVSK	**M**	R	M	A	T	P	L	L	**Q**		invariant chain (97–?)	
FDQK	**I**	V	E	**W**	D	**S**	**R**	K	**S**	KYFE	BLAST-1 (62–78)	48
DQK	**I**	V	E	**W**	D	**S**	**R**	K	Y	F	BLAST-1 (63–77)	
IKI	**I**	S	K	**I**	E	**N**	**H**	E	**G**	VRR	pyruvate kinase (264–278)	
IKI	**I**	S	K	**I**	E	**N**	**H**	E	**G**	VR	pyruvate kinase (264–277)	
FGR	**I**	G	R	**L**	V	**T**	**R**	A	**A**	FNSG	GAPDH (11–25)	49
FGR	**I**	G	R	**L**	V	**T**	**R**	A	**A**	FN	GAPDH (11–23)	
GFGR	**I**	G	R	**L**	V	**T**	**R**	A	**A**	FNSG	GAPDH (10–25)	
CNE	**I**	I	N	**W**	L	D	**K**	N	**Q**		HSC 70 (574–585)	
QPD	**L**	R	Y	**L**	F	L	**N**	G	**N**		Leucine-rich α2-glycoprotein (200–211)	41

[a] Deferred from pool sequence, individual ligands, and pocket structure. GAPDH, glycerine aldehyde 3-phosphate dehydrogenase; BLAST-1, B-lymphocyte activation marker; HSC 70, heat-shock cognate protein 70.

Table 5
HLA-DR4Dw14 pool sequencing

Cycle of Edman degradation	Amino acid residues (in pmol) [a]																	
	I	V	L	F	M	A	G	T	S	Y	P	H	Q	N	E	D	K	R
1	673.2	487.2	644.3	238.4	127.6	401.4	626.4 [b]	77.2	91.8 [b]	216.5	37.8	14.0	41.1	42.0	123.4	87.8	92.2	25.4
2	350.9	413.8	319.4	77.6	73.5	*485.6*	323.2	61.7	*104.9* [b]	125.7	**457.0**	**26.1** [b]	**111.7**	**71.5**	**182.7**	**184.3**	**193.7**	**40.7**
3	*442.8*	371.0	297.8	*85.7*	48.8	290.5	*412.7*	42.5	75.7 [b]	88.5	203.2	**48.3** [b]	**111.5**	**120.9**	144.1	**184.9**	185.5	**45.7**
4	**780.6**	344.1	*395.4*	73.2	**102.9** [b]	258.9	251.2	27.2	75.6 [b]	73.0	129.8	0.7	**180.8**	80.6	*151.7*	143.3	122.4	27.1
5	**858.8**	**500.8**	**637.6**	*93.1*	242.4 [b]	245.5	162.2	*30.5*	39.1	69.8	101.5	0.8	109.8	*84.2*	103.8	*125.6*	*145.0*	**62.7** [b]
6	650.2	368.5	415.6	**168.3**	117.8	*330.1*	*184.4*	*41.5*	*46.4*	**118.5** [b]	76.8	–	90.4	*98.6*	83.6	109.1	**241.5**	**80.9** [b]
7	*676.7*	288.2	354.3	**186.9** [b]	*138.5*	*331.9*	128.8	26.3	42.5	**157.5** [b]	**134.0**	–	81.7	*198.5*	77.6	94.5	98.7	58.7
8	563.9	244.5	334.2	**332.6** [b]	*143.3*	268.2	*135.3*	**39.3**	*43.3*	67.7	113.9	1.2	61.1	98.0	68.2	*101.6*	54.6	52.1
9	253.7	241.2	231.0	147.4	105.5	254.6	*142.2*	**84.4** [b]	**64.8**	43.4	95.6	–	*73.2*	157.4	*94.3*	*130.9*	49.3	42.9
10	225.1	188.4	178.0	63.3	63.3	*295.8* [b]	117.6	**101.6** [b]	46.9	*58.5*	65.4	–	67.5	206.0	65.1	127.6	35.8	16.2
11	183.6	178.8	**306.9**	61.9	**103.1** [b]	**435.8** [b]	*144.5*	43.6	*39.9*	37.4	*65.9*	0.2	*68.8*	128.4	*72.4*	108.5	*42.1*	**34.9**
12	155.8	167.5	285.8	49.5	**108.4** [b]	343.0	*171.6*	23.6	*49.8* [b]	*51.7*	51.3	1.2	*80.4*	80.2	62.9	84.7	39.7	**39.8**
13	122.1	160.6	150.2	*50.1*	44.8	273.3	*147.9*	*24.3*	*55.5* [b]	42.8	36.9	0.8	75.7	49.5	51.7	75.6	*44.4*	**43.1** [b]
14	106.5	146.9	108.4	*52.3*	32.6	225.8	*169.7*	*29.7*	34.8	39.4	*41.6* [b]	–	*91.5*	47.8	*51.9*	73.7	*51.8*	**67.4** [b]
15	75.0	117.9	65.3	33.5	20.1	140.7	111.8	*30.7*	23.0	31.8	41.4 [b]	–	40.9	*48.9*	26.6	57.1	32.1	33.9
16	59.1	91.4	52.0	27.9	14.6	93.8	*130.5* [b]	26.0	17.9	24.4	25.7	0.9	38.7	36.8	*32.2*	46.2	20.9	18.5
17	40.9	60.0	44.1	21.1	10.3	65.1	103.7	20.3	13.3	16.8	17.3	–	24.9	20.4	20.0	34.3	14.4	*23.1* [b]
18	31.1	42.6	33.7	15.4	8.0	50.2	97.0	15.8	11.0	10.8	12.0	0.5	11.4	10.0	16.4	21.9	10.2	13.4
19	27.1	35.8	26.7	12.5	7.0	41.1	69.4	15.4	10.0	8.3	10.9	–	8.1	8.1	*18.6* [b]	20.3	9.5	11.1
20	22.7	31.3	20.9	10.1	5.3	30.1	48.9	11.8	7.2	7.2	9.7	–	6.4	5.7	13.2	15.6	7.6	10.8
21	4.1	3.0	2.3	1.0	0.5	2.8	6.0	1.3	0.7	0.5	0.9	0.5	0.1	0.5	*14.3*	1.5	1.2	3.2

[a] Organization of data as in Table 1.
[b] Signal can be allocated to a dominant peptide: GSHSMRYF... or SHSMRYF....

Fig. 1A) were pooled and sequenced as such (Table 1). The pool sequencing data appeared similar, in principle, to that of other class II ligands [10]. Thus, a Pro signal was prominent at cycle 2 (probably as a result of enzymatic activity during processing) [10] and a number of hydrophobic amino acids were found to be clustered over about three cycles. Such clusters reflect the fixed anchor positions used by ligands to interact with the peptide-accommodating groove of class II molecules. The occurrence in clusters rather than at fixed positions in the pool sequencing data is due to the variable distance between the peptide N-terminus and the first anchor in the different ligands. For example, cycles 3, 4 and 5 show strong signals for a number of hydrophobic or aromatic residues, especially Tyr, Phe and Met (Table 1). Another cluster is seen at positions 6, 7 and 8, this time dominated by Val and Ala, but lacking Tyr. The first cluster centered at cycle 4 represents the first anchor of ligands (set at relative position 1), and the second cluster centered at cycle 7, the second one.

Such analysis led to a basic DR4Dw4-motif with hydrophobic anchors at relative position 1 (corresponding to cycle 4) and relative position 4 (reflecting cycle 7) (Table 2). The other details of the motif shown in Table 2 are derived from a careful comparison of pool sequencing results, individual ligands, and DR1 crystal structure (see discussion). A number of individual ligands, corresponding to the larger peaks in Fig. 1A, could be sequenced directly (Table 2). Some of the most prominent DR4Dw4 ligands are from the HLA-A2 molecule co-expressed by LRM cells. The high abundance of A2-derived ligands is well illustrated by the fact that the corresponding sequences can be followed even in the pool sequencing data. For example, the partial sequence QFVRFDSDA can be easily followed from cycle 5 through 13 (Table 1), standing out as unclustered signals. All individual DR4Dw4 ligands sequenced, including several reported by Chicz et al. [14], are shown in alignment to the motif. Details of alignment, in addition to the hydrophobic anchors at relative positions P1 and P4, such as the occurrence of exceptional hydrophilic residues at P4 and the occupancy of P6, P7 and P9 are covered in Section 4.

3.2. DR4Dw10

DR-associated peptides were isolated from large batches of A1-10 cells, being homozygous for DR4Dw10 (DRB1*0402), and separated by HPLC (Fig. 1B). Material eluting in large, distinct peaks was collected and sequenced separately. All other fractions eluting between 28.0–58.5 min elution time were sequenced as a pool (Table 3). As with all other class II ligand pools we have sequenced so far, a sharp Pro signal at cycle 2 is evident [10,29]. This signal is so well reproducible, irrespective of the class II molecule considered, that it is used as a marker for the quality of the respective peptide isolations. A clustering of hydrophobic amino acids, especially Leu and Ile, is visible for cycles 3, 4 and 5. Significant clustering of hydrophilic residues at cycles 2 and 3 is also evident, as previously found for other class II ligand pool sequences [10]. Less pronounced is a cluster of hydrophobic residues around cycle 7, consisting of Met, Leu and Phe. These data lead to a basic DR4Dw10 ligand motif with a hydrophobic anchor at a relative position P1, reflecting the cluster at cycles 3, 4 and 5. A second anchor is then at P4, with preference for hydrophobic residues (Table 4). Twenty-one individual ligands, derived from 10 different proteins, were sequenced. The most abundant ligands are from HLA-B38, which is co-expressed in A1-10 cells. One peptide is from invariant chain. About half of the peptides can be well aligned to the basic motif with hydrophobic residues at P1 and P4, while the other half has an aliphatic residue to be aligned with P1 but no corresponding hydrophobic residue at P4. This indicates that P4 must allow other than hydrophobic residues as well. At first glance, the further positions within the ligands do not present themselves as clear anchors. However, a careful analysis of aligned ligands, pool sequencing results, and pocket structure reveals degenerate but still distinct specificities at P6, P7 and P9 (see discussion).

Table 6
HLA-DR4Dw14 motif and ligands

	Relative position									Source	Fraction
	1	2	3	**4**	5	**6**	**7**	8	**9**		
Anchor or preferred residues	**V,I L,M**			F,Y,W I,L V,M,A D,E **no R,K**		N,T S,Q R	polar, charged, aliphatic		polar, aliphatic, K		
	1	2	3	**4**	5	**6**	**7**	8	**9**		
GSHS	**M**	R	Y	**F**	H	**T**	**A**	M	**S** RPGRGE	HLA-B60 (1–?)	34, 35, 36, 48
SHS	**M**	R	Y	**F**	H	**T**	**A**	M	**S** RPGRGE	HLA-B60 (2–?)	34, 35, 36, 48
YDNS	**L**	K	I	**I**	S	**N**	**A**	S	C TTN	GAPDH (139–154)	40

[a] Deferred from pool sequence, individual ligands, and pocket structure. GAPDH, glycerine aldehyde 3-phosphate dehydrogenase.

Table 7
HLA-DR4Dw15 pool sequencing

Cycle of Edman degradation	Amino acid residues (in pmol) [a]																		
	I	V	L	F	W	M	A	G	T	S	Y	P	H	Q	N	E	D	K	R
1	85.5	106.1	106.5	66.1	4.3	13.4	110.6	91.0	35.0	34.0	104.5	13.1	1.9	12.4	12.5	55.8	46.0	40.1	12.1
2	41.2	85.3	*114.7*	35.4	*5.1*	11.6	*141.9*	73.0	29.4	*36.9*	57.3	**275.4**	**14.4**	**33.8**	**31.5**	**97.5**	**81.3**	*48.8*	*30.8*
3	*46.4*	73.2	109.7	**66.8**	4.5	**30.6**	104.1	72.8	*31.7*	29.3	**103.5**	184.7	**17.8**	**47.5**	**38.3**	**96.6**	**100.4**	*52.5*	*39.5*
4	*54.8*	71.5	98.9	**105.7**	*6.2*	23.8	63.6	62.9	*39.2*	28.2	**173.0**	99.2	**18.6**	**68.0**	28.6	85.4	79.8	50.8	*42.7*
5	**90.1**	*83.0*	*106.3*	**106.1**	5.1	21.5	61.5	43.1	*43.0*	19.0	**187.5**	70.1	**20.0**	**78.3**	*29.6*	66.3	68.9	*58.0*	*45.6*
6	**97.4**	*117.4*	*106.8*	93.8	*5.8*	**38.5**	*86.2*	32.0	36.0	*27.7*	120.0	56.7	18.4	69.3	*29.9*	*71.3*	59.7	55.2	*53.6*
7	**95.0**	**159.7**	*114.5*	89.5	5.0	**40.8**	**119.2**	29.0	30.2	23.8	88.7	53.5	15.5	44.4	*33.8*	46.3	50.8	42.0	*57.4*
8	77.4	**199.0**	84.4	61.8	2.5	26.9	81.7	*30.2*	29.0	23.5	77.1	*61.7*	10.3	30.2	32.6	28.0	49.6	36.7	*58.8*
9	57.2	164.6	64.9	43.8	**4.5**	*29.3*	71.0	*35.6*	*37.7*	*28.1*	53.9	*64.8*	**19.4**	*39.9*	**57.3**	28.6	*83.9*	28.3	*63.4*
10	38.5	133.6	60.0	33.4	4.2	17.3	*76.4*	*39.7*	35.4	*29.5*	35.8	*62.6*	12.2	32.8	**72.5**	22.5	67.3	26.0	49.3
11	32.7	*133.9*	51.9	21.6	3.8	*22.0*	*86.5*	37.5	29.5	22.5	27.2	*64.4*	9.5	*40.0*	51.4	**32.9**	*73.3*	*27.9*	48.3
12	28.0	91.0	41.1	14.7	2.4	13.6	58.3	27.8	19.8	16.8	18.7	59.9	8.3	*39.8*	36.9	**48.6**	**95.3**	25.1	*49.2*
13	23.0	64.9	30.6	12.5	*2.7*	10.5	42.5	*29.4*	17.7	*17.7*	18.8	49.8	*9.1*	36.8	24.5	45.8	**120.5**	*26.7*	*56.3*
14	16.8	40.7	20.7	8.6	1.0	6.3	34.9	23.1	14.8	15.9	16.9	33.2	7.2	34.1	*28.6*	34.3	92.7	24.3	38.6
15	14.5	30.2	16.9	**13.6**	**3.5**	4.7	23.4	n.d.	n.d.	3.6	13.2	23.5	3.2	13.0	12.8	16.9	40.9	17.9	n.d.
16	9.1	20.3	10.2	6.9	1.8	3.4	20.9	21.6	11.1	10.1	13.1	16.9	*3.4*	*14.2*	*23.1*	*19.8*	*50.7*	*18.2*	18.7
17	7.3	14.4	7.8	5.2	0.7	2.7	14.8	19.5	n.d.	9.7	9.5	13.2	1.8	9.7	16.4	13.7	40.9	13.2	16.1
18	6.2	11.2	5.8	5.2	*1.5*	2.3	10.0	16.0	7.3	8.1	6.9	9.7	1.3	6.4	13.7	8.3	29.2	9.6	12.0
19	4.2	8.9	4.4	3.9	0.7	1.3	7.3	13.5	4.9	6.4	4.7	6.7	0.9	4.5	10.6	5.9	21.2	6.6	7.5
20	2.9	7.0	3.3	2.8	0.6	1.0	6.9	9.9	3.7	5.5	3.4	5.1	0.8	3.8	8.5	4.4	15.1	5.0	7.0

[a] Organization of data as in Table 1.

3.3. DR4Dw14

DR ligands were isolated from DR4Dw14 homozygous MFF cells (DRB1*0404) (Fig. 1C). Pool sequencing again gave a very strong stand-alone signal for Pro at cycle 2 (Table 5). The sequences of a pair of individual ligands, corresponding to the N-terminus of HLA-B60 (GSHSMRYF...) and its N-terminal truncation (SHSMRYF...) can be followed through the pool sequence. This results in doublets of the respective residues at consecutive cycles. See, for example, Arg signals at cycles 5 and 6, Tyr at 6 and 7 and Phe at 7 and 8. In addition to these dominant sequences, several significant clusters can be observed. The most significant one is at cycles 4 and 5, both of which show very strong signals for the aliphatic residues Ile, Leu and Val. A second cluster of hydrophobic residues, of lesser intensity, is at cycles 7 and 8 for Met, Ile and Ala. In addition, the Phe signals at 7 and 8 are stronger than would be expected, if derived only from the two dominant B60 peptides. Thus, Phe is probably part of this cluster. Cycles 9 and 10 indicate a cluster for the hydrophilic residue Asn and potentially Thr (belonging to the two dominant B60 peptides). Further details on the considerations leading to the motif shown in Table 6 are in Section 4.

3.4. DR4Dw15

Source of peptides was the cell line WTA (DRB1*0405). Pool sequencing of HPLC-purified peptides (Fig. 1D) again gave a strong and sharp signal for Pro at cycle 2, and a cluster of hydrophilic residues at cycles 2 and 3 (Table 7). A very significant cluster of the aromatic residues Tyr and Phe is evident in cycles 3, 4 and 5, accompanied by weak signals for the aliphatic residues Met and Ile. A second cluster of aliphatic residues, occurring at cycles 6, 7 and 8, is dominated by Val, Ile and Ala but lacks the aromatic residues Tyr and Phe. Other clusters are at cycles 9 and 10 for Asn and other hydrophilic residues, and very significantly, at 12 and 13 for Asp, accompanied by other hydrophilic residues.

A basic motif indicated by this pattern of clusters is as follows (Table 8). P1 (represented by the cluster at cycles 3, 4 and 5) is an anchor dominated by the aromatic residues Tyr and Phe, but which also allows aliphatic residues. P4 (corresponding to the cluster at cycles 6, 7 and 8) is an anchor with preference for aliphatic residues. P6 shows a preference for Asn and other hydrophilic residues, whereas P9 is an anchor dominated by the negatively charged Asp. Seventeen individual peptides, derived from 9 different proteins, were sequenced (Table 8). All

Table 8
HLA-DR4Dw15 motif and ligands

	Relative position									Source		Fraction
	1	2	3	**4**	5	**6**	**7**	8	**9**			
anchor or preferred residues	**F, Y,** W V, I L, M			V, I, L, M D, E		N, S T, Q K, D	polar, charged, aliphatic		**D, E,** Q			
	1	2	3	**4**	5	**6**	**7**	8	**9**			
YPTQRAR	**Y**	Q	W	**V**	R	C	**N**	P	**D**	SNS	PGSG (1–19)	19
QRAR	**Y**	Q	W	**V**	R	C	**N**	P	**D**	SNS	PGSG (4–19)	19
RAR	**Y**	Q	W	**V**	R	C	**N**	P	**D**	SNS	PGSG (5–19)	19
KPPQ	**Y**	I	A	**V**	H	V	**V**	P	**D**	Q	MIF (32–45)	28, 44
FRE	**F**	K	L	S	K	V	W	R	**D**	QH	Transferrin receptor (173–186)	
FRE	**F**	K	L	S	K	V	W	R	**D**	Q	Transferrin receptor (173–185)	
RE	**F**	K	L	S	K	V	W	R	**D**	QH	Transferrin receptor (174–186)	28
RE	**F**	K	L	S	K	V	W	R	**D**	Q	Transferrin receptor (174–185)	31
VEPDH	**Y**	V	V	**V**	O	A	**Q**	R	**D**	A	Transferrin receptor (397–411)	33
EPDH	**Y**	V	V	**V**	O	A	**Q**	R	**D**	A	Transferrin receptor (398–411)	32
THY	**Y**	A	V	A	V	V	**K**	K	**D**	TDFK	similar to transferrin	
KELK	**I**	D	I	**I**	P	**N**	**P**	Q	**E**	R	Hsp 90-beta (68–81)	33
YLL	**Y**	Y	T	**E**	F	**T**	**P**	T	**E**	KD	β_2-microglobulin (83–96)	54
LL	**Y**	Y	T	**E**	F	**T**	**P**	T	**E**	KDEY	β_2-microglobulin (84–98)	
CAIHAKR	**V**	T	I	**M**	P	**K**	**D**	I	**Q**	LA...	Histone H3 (110–?)	36
APNT	**F**	K	T	**L**	D	**S**	W	R	**D**		ras-related protein RAB-7 (rat) (86–98)	36
VADK	**I**	Q	L	**I**	N	**N**	**M**	L	**D**		Phosphoglycerate kinase (216–228)	55, 46

[a] Deferred from pool sequence, individual ligands, and pocket structure. PGSG, secretory granule proteoglycan core protein; MIF, macrophage migration inhibitory factor. Three peptide sequences (KVHGSLARAGKVRGQTPKVA..., SSHKTFRIKRFLAKKQKQNR..., GKFMKPGKVVLVLAGRYS...) of ribosomal proteins were not aligned because their ligand nature appears to be questionable for several reasons: (i) they represent the N-terminal parts of the respective protein, (ii) the C-terminal residues could not be characterised due to their unusual length of more than 20 residues, (iii) they have been found frequently in ligand extractions of different class II proteins.

Table 9
Relations between DR pocket composition, preference for ligand residues, and association with rheumatoid arthritis

DR molecule	Pocket for relative position					RA association
	P1	P4	P6	P7	P9	
DRB1 * 0101 (DR1)						intermediate
polymorphic residues	*86 Gly*	*13 Phe, 70 Gln, 71 Arg*	*11 Leu, 13 Phe*	*28 Gln, 67 Leu, 71 Arg*	*9 Trp, 57 Asp*	
ligand specificity	**Y, V, L, F, I**, A, M, W	**L, A**, I, V, M, N, Q	**A, G**, S, T, P		L, A, I, V, N, F, Y	
DRB1 * 0401 (DR4Dw4)						strong
polymorphic residues	*86 Gly*	*13 His, 70 Gln, 71 Lys*	*11 Val, 13 His*	*28 Asp, 67 Leu, 71 Lys*	*9 Glu, 57 Asp*	
ligand specificity	**F, Y, W**, I, L, V, M	F, W, I, L, V, A, D, E; **no R, K**	N, S, T, Q, H, R	polar, charged, aliphatic	polar, aliphatic, K	
DRB1 * 0402 (DR4Dw10)						absent
polymorphic residues	*86 Val*	*13 His, 70 Asp, 71 Glu*	*11 Val, 13 His*	*28 Asp, 67 Ile, 71 Glu*	*9 Glu, 57 Asp*	
ligand specificity	**V, I, L, M**	Y, F, W, I, L, M, H, R, N; **no D, E**	N, Q, S, T, K	R, K, H, N, Q, P; rare D, E	polar, aliphatic, H	
DRB1 * 0404 (DR4Dw14)						strong
polymorphic residues	*86 Val*	*13 His, 70 Gln, 71 Arg*	*11 Val, 13 His*	*28 Asp, 67 Leu, 71 Arg*	*9 Glu, 57 Asp*	
ligand specificity	**V, I, L, M**	F, Y, W, I, L, V, M, A, D, E: **no R, K**	N, S, T, Q, K, D	polar, charged, aliphatic	polar, aliphatic, K	
DRB1 * 0405 (DR4Dw15)						intermediate
polymorphic residues	*86 Gly*	*13 His, 70 Gln, 71 Arg*	*11 Val, 13 His*	*28 Asp, 67 Leu, 71 Arg*	*9 Glu, 57 Ser*	
ligand specificity	**F, Y**, W, V, I, L, M	V, I, L, M, F, D, E	N, S, T, Q, K, D	polar, charged, aliphatic	**D, E**, Q	

can be aligned very well to the basic motif, especially with respect to P1, P4 and P9. Details of the motif indicated in Table 8 are discussed below.

4. Discussion

We have analyzed natural peptide ligands of four closely related DR4 molecules which differ in their association with rheumatoid arthritis. Clustering of amino acid signals in pool sequences makes it possible to interpret basic allele-specific motifs consisting of at least two anchors at relative position P1 and P4 of peptide ligands. Individual peptides could be well aligned to these motifs.

The structural information of a monopeptidic DR1 crystal indicated 5 pockets in the DR1 groove, accommodating the side chains of the residues at relative positions P1, P4, P6, P7 and P9 of the peptide [19,20]. The closely related DR4 molecules are expected to have a similar pocket arrangement but with variance of individual pocket specificities. The pool sequences can be interpreted in the light of the predicted pocket structures (Table 9). P1 of ligands is reflected by a cluster of hydrophobic residues stretching around sequencing cycles 3 to 5, as known from previous analysis of DR1 ligands [10]. The residues contributing to the P1 pocket in the DR1/influenza peptide crystal are identical in DR4Dw4 and DR4Dw15. Indeed, occupancy of P1 in the respective ligand mixtures is very similar to the P1 specificity of DR1 ligands: hydrophobic residues with a preference for aromatic amino acids (Table 9). For DR4Dw4, this P1 specificity has also been detected by peptide binding assays [17,30–32]. In contrast, DR4Dw10 and DR4Dw14, differing from DR1 (β86Gly) at only one position in the P1 pocket (β86Val) show a different P1 specificity: hydrophobic residues with preference for aliphatic amino acids. This difference in pocket specificity imposed might be explained by a reduction of the P1 pocket volume by the larger β86Val side chain as opposed to the β86Gly hydrogen atom in DR1 [19,20,29,33,34].

The pocket for the P4 side chain (pocket No. 2 [20]) is, in the case of DR1, more degenerate in its specificity as compared to the P1 pocket. P4 of DR1 ligands can be occupied not only by aliphatic residues but also by polar Asn or Gln [10,18–20]. DRβ chain residues polymorphic among DR1 and the four DR4 molecules considered here and contributing to the P4 pockets are β13, β70 and β71. Polymorphism of this pocket is indeed reflected in the P4 specificity of these 5 molecules, although all of them show a preference for hydrophobic residues at P4. The P4 pocket has one more positive charge (β13His) in DR4Dw4, Dw14 and Dw15 as compared to the P4 pocket of DR1 (β13Phe). Thus, one should expect negative charges to be allowed at P4 of ligands, and this should be reflected at cycles 7 and 8 of the pool sequences. Indeed, for DR4Dw4 (Table 1) the absolute amounts of Asp and Glu were generally higher than those of Arg and Lys, Asp and Glu signals increased at both cycles 7 and 8, and the significant Arg signal at cycle 8 is isolated (not clustered), indicating that it belongs to an individual dominant peptide. Indeed, this Arg signal can be allocated to represent P3 (absolute position 8) of some of the dominant DR4Dw4 ligands derived from HLA-A2. Even with the isolated Arg signal included, the ratio of cumulated Asp and Glu values versus Arg and Lys in cycles 7 and 8 of the pool sequence is greater than 1 (1.23) and His signals are weak and decreasing. This observation gains significance if the calculation is repeated for DR4Dw10. Here, the P4 pocket has two additional negative charges (β70Asp and β71Glu) as compared to Dw4, and thus, positive but not negative charges are expected at P4 of ligands. Inspection of Asp/Glu vs. Arg/Lys signals at cycles 7 and 8 of pool sequence (Table 3) indeed indicates much higher signals of the latter, with a ratio of cumulative Asp/Glu signals vs. Arg/Lys of 0.43. In addition, His signals are very significant and increasing. Thus, the P4 pocket of DR4Dw10 appears to allow positively but disfavours negatively charged residues at P4 of ligands. In contrast, the corresponding pocket of DR4Dw4 allows negative charges but should not allow positive ones at P4 of ligands. Indeed, of the 24 Dw4 ligands aligned in Table 2, five have Asp or Glu at P4, and none has a positive charge. Likewise, peptide binding experiments indicated that positive charges at P4 of peptides inhibit binding to DR4Dw4 [31]. The results of two recent papers by Hammer et al. [21,35] based on synthetic peptide libraries are in accordance with our results with respect to the P4 specificity. Hammer et al. found Lys and Arg at P4 to be favoured in binding to DRB1*0402 as compared to DRB1*0401, whereas DRB1*0404 was similar in this respect to DRB1*0401, which is in perfect agreement with our findings. For Asp and Glu at the same position, Hammer et al. found that the peptides bound weaker both to DRB1*0402 (200- to 500-fold) and DRB1*0404 (20- to 50-fold), as compared to DRB1*0401. Our results fit well with those findings; the relative preference of hydrophobic residues at P4 of DRB1*0404 over DRB1*0401 associated peptides might be interpreted as a consequence of the Lys/Arg substitution at position 71 from DRB1*0401 to DRB1*0404. For DR4Dw14 and Dw15, the charge distribution in the P4 pocket is similar to that of DR4Dw4. Indeed, the Asp/Glu versus Arg/Lys ratio at cycles 7 and 8 of the Dw14 pool sequence is similar as for Dw4 (1.29), so that the P4 specificities of DR4Dw4 and DR4Dw14 are very similar (Table 9). For DR4Dw15, however, the levels of Asp/Glu and Arg/Lys at cycles 7 and 8 are not much different. Two of the 17 individual ligands aligned in Table 8 carry Glu at P4, so that negatively charged residues appear to be tolerated at P4.

Two important points emerge from this consideration of P4 specificities of the different DR4 molecules: i) without the X-ray crystallography data on pocket structure, the detailed evaluation of DR4 ligand pool sequences would

not have been possible, and ii) the P4 specificity appears to be one major feature discriminating the different DR4 molecules, confirming P4 as an allele-specific anchor [16].

The P6 pocket of DR1 (pocket No. 3 [20]) prefers small residues [10,16,18,20]. All 4 DR4 molecules considered here differ from DR1 in two β-chain residues contributing to the P6 pocket (β11Val and β13His). Indeed, all 4 DR4 molecules have a similar specificity here which is different from P6 of the DR1 molecule. P6 of these DR4 molecules accommodate polar residues, especially Asn, Gln, Ser, Thr as well as to a limited extent charged ones. The latter amino acids have not been reported previously. Hammer et al. [17,30] found that positive amino acids are disfavoured at P6. In addition to this, Sette et al. [31] found negatively charged amino acids to have a negative effect on binding. Both groups solely considered positive or negative effects of certain substitutions in synthetic or recombinant peptides on binding to DR4Dw4 molecules without looking at constraints of processing. The latter constraint might heavily influence the natural peptide repertoire presented by class II molecules. The 66 individual DR4 ligands aligned in Tables 2 and 4, 6 and 8 indeed confirm the rather degenerate nature of P6 in that, in addition to polar and charged residues, aliphatic residues are also found, which this time is in agreement with the results by Sette et al. [31].

The P7 pocket (pocket No. 4 [20]) is even more degenerate, at least for DR1, where no specificity could be detected [10,16,18] as well as for DR4Dw4, Dw14 and Dw15. Since DR4Dw10, however, has replaced the basic β71 (Arg or Lys) residue of DR1 and the other DR4 molecules by a negatively charged β71Glu, cycles 10 and 11 of the pool sequence as well as P7 of the 21 individual Dw10 ligands aligned in Table 4 were inspected. The expected preference for positively charged residues is indicated by the cumulative ratio of 0.47 for Asp/Glu values versus Arg/Lys at cycles 10 and 11 of for Dw10 (as opposed to 2.41 for Dw4, 2.54 for Dw14 and 1.29 for Dw15), and P7 of the aligned ligands was positively charged in 12 peptides (belonging to 5 nested sets) out of the 21 Dw10 ligands, whereas a negative charge was found only once. Thus, P7 is rather nonspecific in DR1, DR4Dw4, DR4Dw14 or DR4Dw15 allowing polar, charged and aliphatic residues, whereas P7 of DR4Dw10 has a preference for positively charged residues and disfavours negative charges.

The pocket for P9 side chains (Pocket No. 5 [20]) of DR1 has a preference for aliphatic residues while allowing aromatic and some polar residues as well [10,16,18,20]. The P9 pockets of DR4Dw4, Dw10 and Dw14 differ from that of DR1 at one residue, β9Glu instead of β9Trp. This change is obviously responsible for the lack of a significant clustering of hydrophobic residues at P9, that is cycles 12 through 14 of pool sequences. Instead, signals for polar, aliphatic and positively charged residues are found here, as well as at P9 of aligned ligands. Thus, P9 of DR4Dw4, Dw10 and Dw14 is a rather degenerate anchor, more characterized by residues disfavoured - negatively charged residues and aromatic ones - rather than preferred. P9 would not have been assigned as an anchor of DR4Dw4, Dw10 and Dw14 based on ligand sequencing alone, if the structure analysis of DR1 had not predicted such an anchor.

In contrast, P9 of DR4Dw15 is different. The P9 pocket of this molecule differs from those of the other three DR4 molecules (β57Asp) in one residue, β57Ser. Thus, the P9 pocket of DR4Dw15 has one negative charge less as compared to the P9 pockets of DR1, DR4Dw4, Dw10 and Dw14. The cluster of amino acids at cycles 12 to 14 reflects this difference very clearly: it shows a striking preference for the negatively charged residues Asp and Glu (Table 7). Indeed, almost all of the DR4Dw15 ligands aligned in Table 8 have Asp or Glu at P9. Thus, P9 appears to be an anchor for negatively charged residues or, to a lesser extent, polar residues, as indicated by the pool sequence cluster and P9 of individual ligands. This implies that the P9 pocket of DR4Dw15 must carry a positive countercharge; this is probably provided by α76Arg. Since α76Arg is conserved among the DR molecules considered, it is assumed that the positive charge at α76Arg is neutralized by β57Asp in DR1, DR4Dw4, Dw10 and Dw14 molecules.

Comparing these detailed peptide motifs, it becomes apparent that each of the 5 pockets can contribute to allele-specific peptide binding. P1 is the least variable one, limited to the Gly/Val polymorphism at β86 of all DRβ chains and determining the preference of either aromatic or aliphatic residues at P1 of ligands. In contrast, specificity of the P4 pocket, appears to be highly variable, as indicated by our present data on DR4 subtypes as well as in previous data on DR1, DR3 and other molecules [10,14,16–18,20,30,36]. P6 is rather similar among the DR4 subtypes studied here, which is well explained by identical P6 pockets, whereas P6 pocket composition as well as specificity is different in DR1. Specificity of the P7 pocket is rather degenerate; only for DR4Dw10 is it more pronounced, in that positive or polar residues at P7 of ligands are preferred. P9 specificity is degenerate in DR4Dw4, Dw10 and Dw14 but not in DR4Dw15 and DR1, and has either a pronounced preference for negatively charged and polar residues (DR4Dw15) or a preference for hydrophobic residues (DR1).

As indicated above, the combined information gathered from pool sequencing, individual ligand sequencing, and crystallographic analysis is a powerful tool for determining allele-specific peptide motifs of class II molecules. The polymorphism of DRβ residues contributing to pockets in one particular DR1/peptide complex can well explain the polymorphic specificities of several other DR molecules, including the DR4 subtypes studied here. However, since the contribution of individual DRβ positions to pockets might be different in different DR molecules, interpretation

of pocket structures in other DR molecules should be done with caution [20]. Indeed, it appears that DR molecules with the same composition of pocket residues, as defined for the DR1/influenza peptide complex, may have different pocket specificities. For example, P4-pocket residues of DR4Dw14 and Dw15 are identical whereas the pocket specificity is different: P4 of Dw14 ligands shows a preference for aromatic rather than aliphatic residues and does not tolerate positively charged residues, whereas Dw15 prefers aliphatic residues, and shows much less discrimination of charged residues. Thus, pocket specificity cannot be predicted to the point by the DRβ sequence alone, since DRβ residues other than those used in the DR1/influenza peptide complex might contribute to pocket structure in different DR molecules.

The central question of this paper is the relation between peptide specificity and disease association of DR4 molecules. Although there is some disagreement among authors concerning the levels of association between rheumatoid arthritis (RA) and DR expression reported for different populations, there is still a certain consensus. DR4Dw4 and Dw14 are strongly associated with RA as is DR4Dw15, but, in this case, only in the Japanese population, DR1 intermediately, and DR4Dw10 is not at all or weakly associated [1–5,37]. Thus, it should be of interest to inspect the peptide motifs of the different DR4 molecules for features correlating to the levels of disease association. Most striking in this respect are the different preferences for charged residues at P4 of ligands. Both strongly RA-associated Dw4 and Dw14 do not tolerate positive charges at P4, whereas they have a moderate preference for negatively charged residues. The opposite is found for the least RA-associated DR4 molecule, Dw10; here P4 does not tolerate Asp or Glu, but has a preference for positively charged residues. For DR4Dw15 as well as for DR1, the charge discriminations at P4 are much less pronounced. Another difference in charge preference is at P7; whereas the strongly or intermediately RA-associated DR1, DR4Dw4, Dw14 and Dw15 molecules are rather degenerate in P7 specificity, the non-associated DR4Dw10 shows a preference for positive charges at P7 of ligands. On the other hand, the specificity of P1 and P6 does not seem to show any correlation with RA association, at least within the DR4 groups considered here. The β86 Val/Gly polymorphism, influencing P1 specificity, does not seem to be important here; it has been indicated however, that this difference is correlated with RA association in another context [2]. When the most strongly RA-associated DR4 molecules, DR4Dw4 in the Caucasoid population and DR4Dw15 in the Japanese population are compared, it strikes the eye that they differ in P9. DR4Dw15 shows a prominent specificity for negatively charged amino acids, whereas DR4Dw4 shows a preference for small hydrophobic and polar amino acids, respectively. This might be indicative of an epitope difference dependent on the regional occurrence of the possible disease-inducing agents analogous to the difference in protective DR alleles with respect to different malaria-causing *Plasmodium* strains [38]. As a consequence of these different P9 specificities, a negatively charged amino acid should be included at P9 if a common RA-inducing peptide motif in the Japanese population is to be postulated.

Table 10
Candidates for RA inducing peptides?

a. Suspected features of RA-inducing peptides			**L** **I** **V** **M**	X	X	**D** **E**	X	**T** **S** **N** **Q**	D E	X	X
	Known B cell or T cell antigen										
b. Fitting stretches within antigens possibly involved in RA	Mycobacterial 19kD lipoprotein antigen precursor [a] 88–96		**V**	L	T	**D**	G	**N**	P	P	E
	Mycobacterial HSP 65 [b,c]	83–91	**V**	A	G	**D**	G	**T**	T	T	A
		173–181 *	**I**	T	V	**E**	E	**S**	N	T	F
		183–191	**L**	Q	L	**E**	L	**T**	E	G	M
		323–331	**V**	T	K	**D**	E	**T**	T	I	V
	Calreticulin/Ro/SS-A [d]	97–105	**V**	K	H	**E**	Q	**N**	I	D	C
		176–184	**V**	R	P	**D**	N	**T**	Y	E	V
		184–192	**V**	K	I	**D**	N	**S**	Q	V	E
	Collagen α1(II)precursor [e]	1312–1320	**L**	S	T	**E**	O	**S**	Q	N	I
		1352–1360	**I**	R	A	**E**	G	**N**	S	R	F

* Reported as T cell epitope by Danieli et al. [41].
[a] Tan et al. [9].
[b] Danieli et al. [41].
[c] Quayle et al. [42]; Routsias et al. [43].
[e] Worthington et al. [44].

Interestingly, DR4Dw15 lacks an aspartate at position 57 of its β-chain, as does HLA-DQ8 which is strongly associated with Insulin-dependent Diabetes (Type I) and its murine homolog H-2A^{g7} of the NOD mouse. Peptides bound to the latter two MHC class II molecules have been studied [39,40]; although a preference for a negative charge at P9 has been noted, such an extraordinary specificity as for P9 in DR4Dw15 has not been found.

Under the assumption that RA is caused by a T-cell response against a peptide presented by the RA-associated DR molecules but not by the non-associated ones, one can now start to think about the features such a hypothetical peptide should have. Or, in other words, how should a peptide look that is presented well by DR4Dw4, Dw14 and DR4Dw15, moderately by DR1, and not at all by DR4Dw10 molecules? Based on the above considerations and the data summarized in Table 9, the hypothetical RA-inducing peptide should look as follows: P1 is hydrophobic (aliphatic or aromatic), P4 is Asp or Glu. P6 is Ser or Thr (common for P6 residues of DR1 and DR4 ligands). P7 is not specified but more likely is occupied by an Asp or Glu rather than another residue. P9 is not specified for the Caucasoid population; for the Japanese, it would be Asp or Glu. Thus, a sequence like LXXEXSEXX (LXXEXSEXE for Japan) would be the core of such a speculative RA inducing peptide, which, as a natural DR ligand, would include 3 or 4 flanking residues at either side of the core sequence. The likely source of such a peptide would be a self protein expressed in the joint, such as collagen, or a bacterial antigen crossreactive to a self protein of the joint. For some of the antigens that have been reported to be possibly involved in RA [9,41–44], we performed an epitope search based on the above hypothesis (Table 10). The result indicated a number of fitting stretches within these proteins, including part of a T cell epitope of mycobacterial HSP 65 reported by Danieli et al. [41]. We are aware that our exercise here is highly speculative but we think it provides a useful basis for the design of future experiments, the aim of which is to identify peptides inducing RA as well as other autoimmune diseases, and to base new strategies for prevention and therapy on this information.

Acknowledgements

We thank A. Vogt for introducing T.F. into techniques and for unpublished data, S. Faath for technical assistance, and B. Stiller and A. Jordan for preparing the manuscript. We further thank Z. Nagy for helpful discussions, J. Mytilineos for HLA typing, and G. Pawelec and B. Takacs for cell lines. This work was supported by Hoffmann-La Roche, Nutley (N.J.), by a grant from the Bundesminister fur Forschung und Technologie, and by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 120 and the Leibnizprogramm).

References

[1] Lechler, R., ed. (1994), HLA and Disease, Academic Press, London.
[2] Nelson, J.L. et al. (1991) in HLA 1991, (K. Tsuji, M. Aizawa and T. Sasazuki, eds.), Vol. 1, pp. 772–774, Oxford Science Publications, Oxford.
[3] Nelson, J.L., Mickelson, E., Masewicz, S., Barrington, R., Dugowson, C., Koepsell, T. and Hansen, J.A. (1991) Tissue Antigens 38, 145–151.
[4] Hillarby, M.C., Hopkins, J., and Grennan, D.M. (1991) Tissue Antigens 37, 39–41.
[5] Wordsworth, B.P. and Bell, J.I. (1992) Springer Semin. Immunopathol. 14, 59–78.
[6] Becking, A., Pluschke, G., Krawinkel, U., Melchers, I., Peter, H.H. and Lang, B. (1993) Eur. J. Immunogenet. 20, 83–89.
[7] Marsh, S.G.E. and Bodmer, J.G. (1993) Eur. J. Immunogenet. 20, 47–79.
[8] Rook, G. and McCulloch, J. (1992) Arthritis Rheum. 35, 1409–1412.
[9] Tan, P.L., Farmiloe, S., Young, J., Watson, J.D. and Skinner, M.A. (1992) Arthritis Rheum. 35, 1419–1426.
[10] Falk, K., Rötzschke, O., Stevanović, S., Jung, G. and Rammensee, H.-G. (1994) Immunogenetics 39, 230–242.
[11] Rudensky, A.Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. and Janeway, C.A. (1991) Nature 353, 622–627.
[12] Rudensky, A.Y., Preston-Hurlburt, P., Al-Ramadi, B.K., Rothbard, J. and Janeway, C.A. (1992) Nature 359, 429–431.
[13] Hunt, D.F., Michel, H., Dickinson, T.A., Shabanowitz, J., Cox, A.L., Sakaguchi, K., Appella, E., Grey, H.M. and Sette, A. (1992) Science 256, 1817–1820.
[14] Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A.A., Lane, W.S. and Strominger, J.L. (1993) J. Exp. Med. 178, 27–47.
[15] Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Vignali, D.A.A., and Strominger, J.L. (1992) Nature 358, 764–768.
[16] Sinigaglia, F. and Hammer, J. (1994) Curr. Opin. Immunol. 6, 52–56.
[17] Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B. and Sinigaglia, F. (1993) Cell 74, 197–203.
[18] Hammer, J., Takacs, B. and Sinigaglia, F. (1992) J. Exp. Med. 176, 1007–1013.
[19] Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1993) Nature 364, 33–39.
[20] Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1994) Nature 368, 215–221.
[21] Hammer, J., Gallazi, F., Bono, E., Karr, R.W., Guenot, J., Valsasnini, P., Nagy, Z.A. and Sinigaglia, F. (1995) J. Exp. Med. 181, 1847–1855.
[22] Rammensee, H.-G., Friede, T. and Stevanović, S. (1995) Immunogenetics 41, 178–228.
[23] Rammensee, H.-G. (1995) Curr. Opin. Immunol. 7, 85–96.
[24] Stunz, L.L., Karr, R.W. and Anderson, R.A. (1989) J. Immunol. 143, 3081–3086.
[25] Lampson, L.A. and Levy, R. (1980) J. Immunol. 125, 293–299.
[26] Falk, K., Rötzschke, O., Stevanović, S., Jung, G. and Rammensee, H.-G. (1991) Nature 351, 290–296.
[27] Stevanović, S. and Jung, G. (1993) Analyt. Biochem. 212, 212–220.
[28] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[29] Rötzschke, O. and Falk, K. (1994) Curr. Opin. Immunol. 6, 45–51.
[30] Hammer, J., Belunis, C., Bolin, D., Papadopoulos, J., Walsky, R., Higelin, J., Danho, W., Sinigaglia, F. and Nagy, Z.A. (1994) Proc. Natl. Acad. Sci. USA 91, 4456–4460.
[31] Sette, A., Sidney, J., Oseroff, C., Del Guercio, M.F., Southwood, S., Arrhenius, T., Powell, M.F., Colon, S.M., Gaeta, F.C.A. and Grey, H.M. (1993) J. Immunol. 151, 3163–3170.
[32] Hill, C.M., Liu, A., Marshall, K.W., Mayer, J., Jorgensen, B., Yuan, B., Cubbon, R.M., Nichols, E.A., Wicker, L.S. and Rothbard, J.B. (1994) J. Immunol. 152, 2890–2898.
[33] Stern, L.J. and Wiley, D.C. (1994) Structure 2, 245–251.

[34] Demotz, S., Barbey, C., Corradin, G., Amoroso, A. and Lanzavecchia, A. (1993) Eur. J. Immunol. 23, 425–432.
[35] Hammer, J., Bono, E., Gallazi, F., Belunis, C., Nagy, Z.A. and Sinigaglia, F. (1994) J. Exp. Med. 180, 2353–2358.
[36] Malcherek, G., Falk, K., Rötzschke, O., Rammensee, H.-G., Stevanović, S., Gnau, V., Jung, G. and Melms, A. (1993) Int. Immunol. 5, 1229–1237.
[37] de Vries, N., Ronningen, K.S., Tilanus, M.G., Bouwens-Rombouts, A., Segal, R., Egeland, T., Thorsby, E., van de Putte, L.B. and Brautbar, C. (1993) Tissue Antigens 41, 26–30.
[38] Powis, S.H. and Geraghty, D.E. (1995) Immunol. Today 16, 466–468.
[39] Reich, E.-P., von Grafenstein, H., Barlow, A., Swenson, K.E., Williams, K. and Janeway, C.A. (1994) J. Immunol. 152, 2279–2288.
[40] Chicz, R.M., Lane, W.S., Robinson, R.A., Trucco, M., Strominger, J.L. and Gorga, J.C. (1994) Int. Immunol. 6, 1639–1649.
[41] Danieli, M.G., Markovits, D., Gabrielli, A., Corvetta, A., Giorgi, P.L., van der Zee, R., Van Embden, J.D., Danieli, G. and Cohen, I.R. (1992) Clin. Immunol. Immunopathol. 64, 121–128.
[42] Quayle, A.J., Wilson, K.B., Li, S.G., Kjeldsen-Kragh, J., Oftung, F., Shinnick, T., Sioud M., Forre, O., Capra, J.D. and Natvig, J.B. (1992) Eur. J. Immunol. 22, 1315–1322.
[43] Routsias, J.G., Tzioufas, A.G., Sakarellos-Daitsiotis, M., Sakarellos, C. and Moutsopoulos, H.M. (1993) Clin. Exp. Immunol. 91, 437–441.
[44] Worthington, J., Turner, S., Brass, A. and Morgan, K. (1993) Br. J. Rheumatol. 32, 658–662.