Identification of self peptides bound to purified HLA-B27

T. S. Jardetzky*, W. S. Lane†, R. A. Robinson†, D. R. Madden*‡ & D. C. Wiley*§

* Department of Biochemistry and Molecular Biology, § Howard Hughes Medical Institute, ‡ Committee on Higher Degrees in Biophysics, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
† Harvard Microchemistry Facility, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

A pool of endogenous peptides bound to the human class I MHC molecule, HLA-B27, has been isolated. Microsequence analysis of the pool and of 11 HPLC-purified peptides provides information on the binding specificity of the HLA-B27 molecule. The peptides all seem to be nonamers, seven of which match to protein sequences in a database search. These self peptides derive from abundant cytosolic or nuclear proteins, such as histone, ribosomal proteins, and members of the 90K heat-shock protein family.

CLASS I major histocompatibility complex (MHC) molecules bind short peptides derived from intracellular proteins, forming complexes which are transported to the cell surface and accessible to immune system surveillance. In an infected cell, viral peptides complexed with class I MHC molecules are recognized by specific cytotoxic T-lymphocytes (CTL), and the infected cell is lysed[1-3]. During their development in the thymus, CTL go through a complex process of positive and negative selection[4,5], generating a population of T cells poised to react to non-self peptides bound to self class I MHC molecules, while eliminating many self-reactive T cells. T-cell interaction with MHC molecules on thymic cells is important for this repertoire selection.

In an uninfected cell, class I MHC molecules are probably bound to peptides derived from self proteins. The peptide-binding sites of three human class I MHC molecules contain significant extra electron density, potentially because of bound endogenous peptides, as observed by X-ray crystallographic methods[6-8,45]. Class I heavy- and light-chain assembly and stable cell-surface expression depends on the binding of short peptides[9-15]. Empty class I MHC molecules are much less stable than complexes formed with antigenic viral peptides, and short peptides can induce a conformational change in the class I heavy chain[16]. Finally, low relative molecular mass (M_r) material eluted from purified class I molecules could be interpreted as a complex mixture of potential self peptides[17,18], with peptide motifs which differed between MHC alleles (ref. 18 and G. Van Bleek and S. Nathenson, in preparation).

Here we present the sequences of 11 self peptides isolated from a pool of peptides bound to the human class I MHC molecule, HLA-B27. Seven of the peptides can be matched to proteins found in sequence databases and derive from intracellular proteins. Alignment of the self peptides with HLA-B27-restricted viral peptides reveals a sequence motif, which can be complementary to the antigen-binding site seen in the HLA-B27 crystal structure[8].

Isolation of peptides bound to HLA-B27

Peptides were acid-eluted from purified papain-solubilized HLA-B27 (ref. 19) and recovered after ultrafiltration (Fig. 1), followed by amino-acid analysis. Amino-acid amounts 15–30-fold above background were found after acid treatment of HLA-B27. Assuming the length of bound peptides to be 8–9 amino acids[17,18,20], the amount of peptidic material in the acid-eluted pool corresponds to a 50% (molar) yield, although HLA-B27 fragments also contribute to this estimate. An additional fraction, which remains to be fully characterized, was collected after reduction of the denatured HLA-B27. The combined yield from both pools corresponds to about 65% of the peptidic material expected.

The acid-eluted pool was analysed by automated Edman degradation[21]. The pool sequence did not show any dominant peptide sequence, but in cycle 2 there was 20 times more arginine than all other amino acids. Thus, arginine is probably an anchor residue[18], important for peptide binding to HLA-B27.

The acid-eluted pool was fractionated by reversed-phase HPLC (Fig. 2). Single peaks from these separations were collected and microsequenced by automated Edman degradation[21]. Some of these peaks contained more than one peptide, allowing both a primary and a secondary sequence to be determined. The confidence with which a given amino-acid assignment was

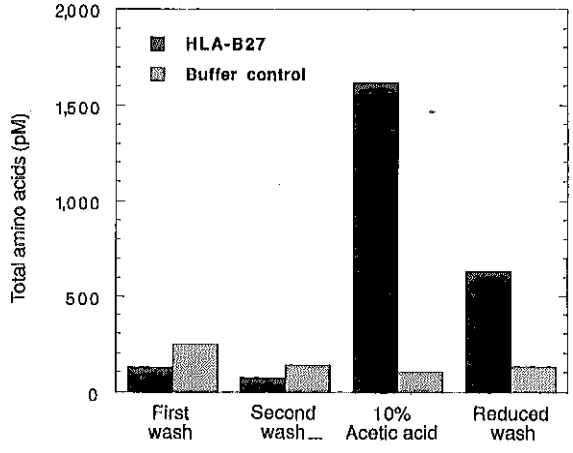

FIG. 1 Isolation of peptides bound to purified HLA-B27. HLA-B27 (B*2705) was papain-cleaved and purified as described[18] from IG-2 lymphoblastoid cells. Concentrated HLA-B27 was further purified by HPLC gel filtration using a Waters SW-300 gel filtration column run in 25 mM MES pH 6.5, 150 mM NaCl, 0.1% NaN_3. About 300 μg HLA-B27 was spin-concentrated using an Amicon Centricon-10 (M_r cutoff of 10K) ultrafiltration device, to a final volume of 60 μl. Ammonium acetate (1 ml, 50 mM, pH 7.5) was added and concentrated. The flow-through provided the first wash. This procedure was repeated (second wash). Acetic acid (1 ml, 10%) was added, the sample was heated at 100 °C for 5 min and concentrated. The flow-through provided the acid-eluted low M_r pool. An equivalent volume (60 μl) of 10% ammonium hydroxide was added to the retentate followed by 1.2 μl β-mercaptoethanol. At this point a precipitate was observed. The sample was incubated for 1 h at 37 °C. Acetic acid (1 ml, 10%) was added, at which point the precipitate dissolved. The sample was spin-concentrated and the flow-through collected. The four pools were concentrated to 100 μl in a Savant Speed-Vac and lyophilized. The pools were then dissolved in 100 μl water and 5% was analysed for amino-acid content on an ABI 420A/130A derivatizer/HPLC after hydrolysis *in vacuo* with 6M HCl for 24 h.

TABLE 1 Peptide sequences

Peak	Yield (pmol)	Peptide sequence													Homologous protein
ELUTED PEPTIDES		1	2	3	4	5	6	7	8	9					
2-27	16	**R**	**R**	**Y**	**Q**	**K**	**S**	**T**	**E**	**L**					Human histone H3, H3.3
	4	**R**	**R**	**I**	**K**	E	**I**	**V**	**K**	K					Human Hsp89α
2-14	4	**R**	**R**	**V**	**K**	**E**	V	V	**K**	k					Human Hsp89β
2-83	4	**R**	**R**	**W**	**L**	**P**	**A**	G	d	a					Human elongation factor 2
2-25a	3	**R**	**R**	**S**	**K**	**E**	**I**	**T**	**V**	**R**					Human ATP-dependent RNA helicase
2-42	16	**G**	**R**	**I**	**D**	**K**	**P**	**I**	**L**	**K**					Ribosomal protein (yeast/slime mould)
2-62b	3	F*	**R**	Y	**N**	**G**	**L**	i	**H**	r					Rat 60S ribosomal protein L28
1-14	1.5	**K**	**R**	**F**	**E**	**G**	**L**	**T**	**Q**	**R**					—
2-45	6	R	**R**	**F**	**T**	**R**	**P**	**E**	**H**	—					—
2-46	7	**R**	**R**	**I**	**S**	**G**	**V**	**D**	**R**	**Y**					—
2-62a	14	**A**	**R**	**L**	**F**	**G**	**I**	**R**	**A**	**K**					—
2-25b	3	[**P**	**P**	**K**	**T**	**H**	**V**	**T**	**H**	**H**	**P**	...]			HLA-B27 (182-)
2-93	11	[**G**	**S**	**H**	**S**	**M**	**R**	**Y**	**F**	**H**	**T**	**S**	**V**	...]	HLA-B27 N terminus
VIRAL PEPTIDES															
		S†	**R**	**Y**	**W**	**A**	**I**	**R**	**T**	**R**					Influenza A nucleoprotein 383-391
		K	**R**	**W**	**I**	**I**	**L**	**G**	**L**	**N**	**K**	**I**	**V**		HIV GAG p24 protein 265-276
		G	**R**	**A**	**F**	**V**	**T**	**I**	**G**	**K**					HIV gp120 314-322

Isolated peptides were HPLC purified as described in Fig. 2 and microsequenced using an ABI 477A protein sequencer as in ref. 21. Yields are based on initial sequence levels. Bold, full confidence sequence data; not bold, high confidence sequence data; lower case, most significant PTH-amino acid, but less than high confidence; underlined, where peaks contain two peptides, PTH-amino-acid level was consistent with the presence of the amino acid in both sequences; brackets, peptides derived from HLA-B27, not bound in the site.

* Ser and Lys in cycle 1 and Phe in cycle 3 also consistent with sequencing data.

† Ser not present in synthetic peptides used in ref. 27.

made is divided into four different categories (Table 1). All peptide sequences were found in the 1-16 picomole range, where 30 picomole represents an occupancy of about 1% of the HLA-B27 molecules. Each of the sequences was used to search protein and nucleic-acid databases (refs 22, 23 and Sequence Analysis Software Package, Genetics Computer Group, Madison).

Peptide homology

Five of the isolated peptides match to human proteins with intracellular locations and two peptides were derived from HLA-B27 (Table 1). The primary sequence data from peak 1-8 and peak 2-27 match the human H3 and H3.3 histone proteins, whereas the secondary sequence matches a member of the human 90K heat-shock protein family, Hsp89α. A second peptide, found in peak 2-14, matches exactly to a closely related heat-shock protein, Hsp89β (ref. 24), and the two heat-shock peptides derive from the same region of their associated protein sequences. A peptide which matches the human elongation factor 2 protein sequence was found in peak 2-83.

In peak 2-25, two peptide sequences were observed in a 1:1 molar ratio. After identification of an acid cleavable[25] fragment of HLA-B27 (peak 2-25b, Table 1), the remaining peptide sequence in peak 2-25 was matched to the sequence of a human ATP-dependent RNA helicase (peak 2-25a, Table 1).

A representative large peak, eluting later in the acetonitrile gradient (Fig. 2*b*, peak 2-93), was sequenced and found to be a long N-terminal peptide from HLA-B27. The two HLA-B27 peptides (peak 2-25b and peak 2-93) are probably peptides generated by papain cleavage and acid treatment and not specifically bound in the antigen-binding site.

Two peptides (peaks 2-42 and 2-62) have significant sequence similarity to ribosomal proteins, whose human homologues are not in the databases. The peptide from peak 2-42 is similar to three proteins with one amino acid mismatched in each alignment: the slime mould (*Dictyostelium discoideum*) ribosomal protein L2, the yeast (*Schizosaccharomyces pombe*) ribosomal protein K37, and the yeast (*S. pombe*) ribosomal protein rpKD4. The single mismatches occurred in the same sequence position and are the conservative substitutions methionine or leucine, instead of the experimentally determined isoleucine at position 7 (Table 1). The sequencing data suggest that the human protein will have isoleucine at this position.

In the secondary sequence data from peak 2-62, a second potential ribosomal protein peptide was found. An exact match is found with the rat ribosomal protein L28 (peak 2-62b, Table 1). This peptide sequence may be conserved in the rat and human proteins.

Four of the sequenced peptides did not match any database proteins with statistical significance. These peptides were found in peaks 1-13, 2-45, 2-46 and 2-62a and all gave high confidence sequence data. Three of the four peptides provided nine cycles of clear sequence data. The PTH amino-acid yields for peak 2-45 tapered off in the later cycles and were too low to conclude that the peptide is only eight amino acids long.

The HLA-B27 peptide-binding motif

The sequencing data from the peptide pool and the HPLC-purified peptides (Table 1) share sequence patterns, the most striking being the arginine at position 2. The interpretation that the sequence patterns reflect the specificity of the HLA-B27 peptide-binding site is supported by the observation that three viral peptides which bind to HLA-B27 (refs 26, 27 and K. Parker, *et al.*, in preparation) align well with the self-peptide sequences (Table 1). In addition, three of the self peptides (KRFEGLTQR, RRYQKSTEL, GRIDKPILK; single-letter amino-acid code) have been synthesized and stimulate the assembly of denatured HLA-B27 (ref. 28 and E. Collins, unpublished data). A tally of all the observed residues at each position of the peptides (P1–P9) is presented in Fig. 3. The number of times an amino acid is found at each peptide position is graphically represented using the structural data for HLA-B27 (ref. 8).

The most restricted peptide position is P2, with arginine found in all of the peptide sequences, followed by P1 and P9, which show a preference for positively charged amino acids. Further constraints are found at positions P3 (hydrophobic) and P6 (nonpolar or small polar residues) in the peptides. The remaining positions, P4, P5, P7 and P8 are fairly unrestricted, with both

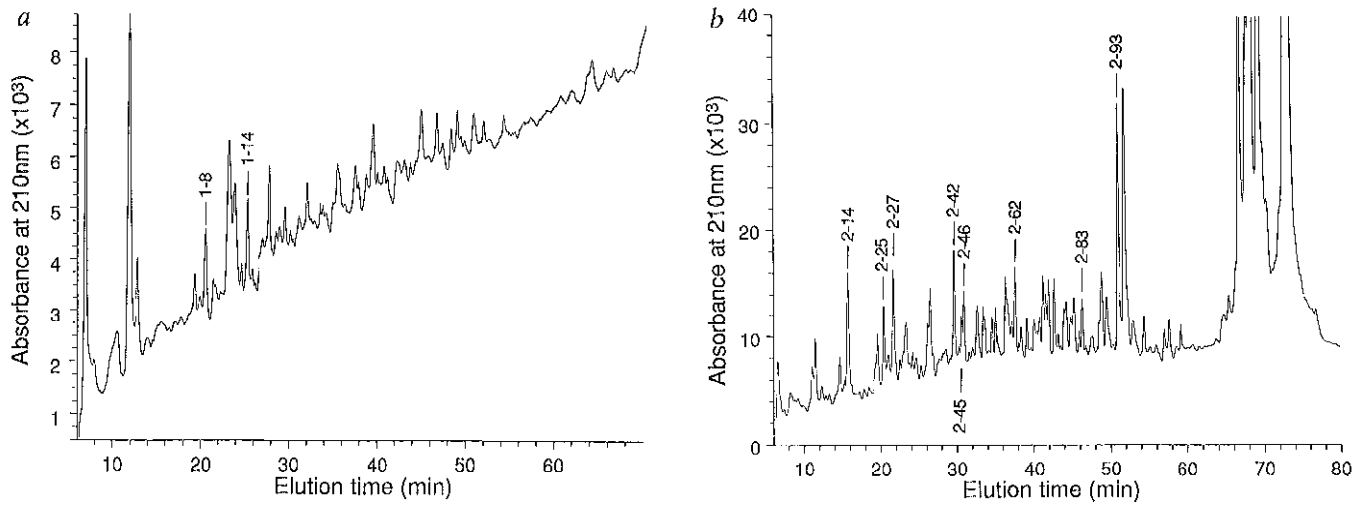

FIG. 2 Narrowbore reversed-phase HPLC separation of peptides. *a*, Five per cent of the HLA-B27 acid-eluted pool was loaded onto a 2.1 mm × 150 mm Vydac C-18 reversed-phase column. *b*, Fifty per cent of the HLA-B27 acid-eluted pool was loaded onto a 2.1 mm × 150 mm Vydac C-8 reversed-phase column. The samples were chromatographed on a Hewlett-Packard 1090 HPLC/1040 diode array detector, using a modification of a previously described gradient[21,44]. Peaks are labelled with the prefix 1 or 2 to identify the column run and fraction number. Fourteen peaks were chosen for sequencing based on peak symmetry, resolution, ultraviolet absorbance and spectra, with 10 providing sequence data.

negatively and positively charged side chains as well as polar and nonpolar residues represented.

Conclusions

Individual peptide sequences have been determined from a pool of peptides associated with purified HLA-B27 molecules. A heterogeneous population of peptides is eluted on acid denaturation of HLA-B27 in yields consistent with stoichiometric MHC-peptide complexes. Sequence analysis of this peptide mixture reveals a predominance of arginine at cycle 2, remarkably consistent with the expected specificity constraints of the HLA-B27 peptide-binding site[8]. Narrow-bore HPLC separation allowed the isolation and microsequence analysis of 11 individual self peptides. These self peptides share a sequence motif, which can be used to align three viral peptides known to bind HLA-B27. In addition, three selected self peptides stimulate the *in vitro* assembly of HLA-B27. These data support the conclusion that the peptide sequences reported here are derived from HLA-B27 peptide complexes found in the lymphoblastoid cell line LG-2 grown in cell culture.

The compilation and alignment of 14 individual peptide sequences provides an HLA-B27 peptide-binding motif, which can be interpreted using the X-ray crystallographic structure of HLA-B27 (ref. 8). Class I MHC molecules must bind a large number of diverse peptides to guarantee T-cell mediated immunity to infection. Assuming each peptide position is completely independent, the limited motif data in Fig. 3 represent a potential for binding ($6 \times 8 \times 11 \times 8 \times 7 \times 7 \times 11 \times 6$), or 13 million, unique peptide sequences.

Seven of the self peptides are similar to intracellular proteins, with either cytosolic (ribosomal proteins, elongation factor 2 and heat-shock proteins, ref. 29) or nuclear (histone, helicase, ref. 30) locations. This intracellular localization is consistent with the observation that class I MHC molecules bind antigenic peptides from non-self proteins degraded in the cytoplasm[31-33], and that cells with a defective peptide transport mechanism

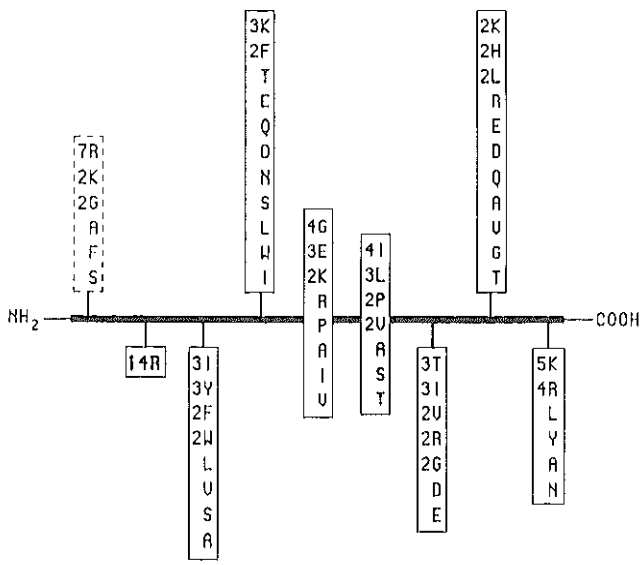

FIG. 3 The HLA-B27 peptide binding motif. The peptide sequences shown in Table 1 were used to compile a list of the amino acids found at positions P1–P9 for nonamer peptides bound to HLA-B27. Residues which occur once at a position are listed in single-letter amino-acid code, whereas multiple occurences are indicated by number. P1–P9 are represented from left to right, with the far left equivalent to the N terminus of the peptide. The placement of the boxed peptide positions reflects information based on the analysis of the HLA-B27 structure[8]. Boxed positions which point down (P2, P3, P7, P9) are those for which observed extra electron density points down into HLA-B27 pockets. Boxed positions which point up (P4 and P8) are those which point directly out of the HLA-B27 peptide binding site. Boxed positions which are centred (P5 and P6) are those for which side chains lie across the peptide-binding site. The dashed box for peptide position P1 reflects the fact that no electron density was observed for this peptide side chain. But model building suggests that the longer side chains of arginine and lysine at P1 can point up from the binding site.

express low amounts of class I molecules on their surface[9,34–38]. No obvious protease cleavage sites were found in common in the self-protein sequences, which could generate the observed self peptides.

Most of the identified self peptides derive from abundantly expressed proteins. Histones and ribosomal proteins are present in millions of copies in the cell[39,40]. Elongation factor 2 is similar to Ef-Tu, one of the most abundant proteins in *Escherichia coli*[40], and the heat-shock proteins Hsp89α and Hsp89β are abundantly and constitutively expressed in cells cultured in serum[24,29]. Highly expressed proteins might be expected to provide more peptide for class I MHC molecules than poorly expressed proteins and thereby be better represented in the self-peptide pool. But the number of peptide complexes formed and presented at the cell surface will probably depend on other factors as well.

All the processes which alter the origins and the amounts of self peptides bound to class I MHC molecules at the cell surface could have effects on T-cell development, the allorecognition of non-self MHC (ref. 46), the induction of autoimmune disease, and the immune surveillance of infected or cancerous cells. Protein degradation rates affect viral peptide presentation to CTL (ref. 41) and the specific expression of a normal protein may affect tumour rejection[42]. The mouse homologue of the human Hsp89α protein has also been identified as a tumour transplantation antigen[43].

Note added in proof: All of the large peaks to the right of peak 2-93 in Fig. 2*b* contain peptides longer than 12 amino acids, with sequences starting at the N-terminus of the HLA-B27 heavy chain or β_2-microglobulin, or the acid-cleaned band of the heavy chain. □

Received 23 July; accepted 23 August 1991.

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248,** 701–702 (1974).
2. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42,** 457–467 (1985).
3. Townsend, A. R. M. & Bodmer, H. *A. Rev. Immun.* **7,** 601–624 (1989).
4. Strominger, J. L. *Science* **244,** 943–950 (1989).
5. Von Boehmer, H. *A. Rev. Immun.* **8,** 531–556 (1990).
6. Bjorkman, P. J. *et al. Nature* **329,** 506–512 (1987).
7. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *Nature* **342,** 692–696 (1989).
8. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353,** 321–325 (1991).
9. Townsend, A. R. M. *et al. Nature* **340,** 443–448 (1989).
10. Townsend, A. R. M. *et al. Cell* **62,** 285–295 (1990).
11. Schumacher, T. N. M. *et al. Cell* **62,** 563–567 (1990).
12. Lie, W. R. *et al. Nature* **344,** 439–441 (1990).
13. Cerundolo, V. *et al. Nature* **345,** 449–452 (1990).
14. Ljungren, H. G. & Karre, K. *Nature* **346,** 476–480 (1990).
15. Kvist, S. & Hamann, U. *Nature* **348,** 446–448 (1990).
16. Elliott, T., Cerundolo, V., Elvin, J. & Townsend, A. *Nature* **351,** 402–406 (1991).
17. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348,** 213–216 (1990).
18. Falk, K., Rötzschke, O., Stevanović, S., Jung, G. & Rammensee, H.-G. *Nature* **351,** 290–296 (1991).
19. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. *J. biol. Chem.* **252,** 7555–7567 (1977).
20. Rötzschke, O. *et al. Nature* **348,** 252–254 (1990).
21. Lane, W. S., Galat, A., Harding, M. W. & Schreiber, S. L. *J. Prot. Chem.* **10,** 151–160 (1991).
22. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85,** 2444–2448 (1988).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215,** 403–410 (1990).
24. Hickey, E., Brandon, S. E., Smale, G., Lloyd, D. & Weber, L. A. *Molec. cell. Biol.* **9,** 2615–2626 (1989).
25. Terhorst, C., Robb, R., Jones, G. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **74,** 4002–4006 (1977).
26. Nixon, D. F. *et al. Nature* **336,** 484–487 (1988).
27. Huet, S. *et al. Int. Immun.* **2,** 311–316 (1990).
28. Silver, M. L., Parker, K. C. & Wiley, D. C. *Nature* **350,** 619–622 (1991).
29. Lai, B.-T., Chin, N. W., Stanek, A. E., Keh, W. & Lanks, K. W. *Molec. cell. Biol.* **4,** 2802–2810 (1984).
30. Iggo, R. D. *et al. Molec. cell. Biol.* **11,** 1326–1333 (1991).
31. Morrison, L. A., Lukacher, A. E., Braciale, V. L., Fan, D. P. & Braciale, T. J. *J. exp. Med.* **163,** 903–921 (1986).
32. Moore, M. W., Carbone, F. R. & Bevan, M. J. *Cell* **54,** 777–785 (1988).
33. Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324,** 575–577 (1986).
34. Spies, T. *et al. Nature* **348,** 744–747 (1990).
35. Monaco, J. J., Cho, S. & Attaya, M. *Science* **250,** 1723–1726 (1990).
36. Trowsdale, J. *et al. Nature* **348,** 741–744 (1990).
37. Deverson, E. V. *et al. Nature* **348,** 738–741 (1990).
38. Spies, T. & DeMars, R. *Nature* **351,** 323–324 (1991).
39. Alberts, B. *et al.* in *Molecular Biology of the Cell* 2nd edn 496 (Garland, New York, 1989).
40. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. in *Molecular Biology of the Gene* 4th edn 466, 652, 677, 725 (Benjamin/Cummings, Menlo Park, 1987).
41. Townsend, A. R. M. *et al. J. exp. Med.* **168,** 1211–1224 (1988).
42. Van den Eynde, B., Lethé, B., Van Pel, A., De Plaen, E. & Boon, T. *J. exp. Med.* **173,** 1373–1384 (1991).
43. Ullrich, S. J., Robinson, E. A., Law, L. W., Willingham, M. & Appella, E. *Proc. natn. Acad. Sci. U.S.A.* **83,** 3121–3125 (1986).
44. Stone, K. L. *et al.* in *Techniques in Protein Chemistry* (ed. Hugli, T.) 377–391 (Academic, San Diego, 1989).
45. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. *J. molec. Biol.* **219,** 277–319 (1991).
46. Matzinger, P. & Bevan, M. J. *Cell Immunol.* **29,** 1–5 (1977).

ACKNOWLEDGEMENTS. We thank J. Gorga and J. Strominger for their initial and continuing collaboration on the HLA-B27 crystal structure and A. Haykov for technical assistance, L. Stern for helpful discussions, E. Collins and K. Parker for sharing unpublished data and M. Frayser and K. Svenson for growing LG-2 cells. The work was supported by the Howard Hughes Medical Institute. T.S.J. is a postdoctoral fellow of the Cancer Research Institute.