

Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle

Hwai-Chen Guo*, Theodore S. Jardetzky*,
Thomas P. J. Garrett†, William S. Lane‡,
Jack L. Strominger* & Don C. Wiley†

* Department of Biochemistry and Molecular Biology and
† Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue,
Cambridge, Massachusetts 02138, USA
‡ Harvard Microchemistry Facility, 16 Divinity Avenue, Cambridge,
Massachusetts 02138, USA

WE report here the determination and refinement to 1.9 Å resolution by X-ray cryo-crystallography the structure of HLA-Aw68. The averaged image from the collection of bound, endogenous peptides clearly shows the atomic structure at the first three and last two amino acids in the peptides but no connected electron density in between. This suggests that bound peptides, held at both ends, take alternative pathways and could be of different lengths by bulging out in the middle. Peptides eluted from HLA-Aw68 include peptides of 9, 10 and 11 amino acids, a direct indication of the length heterogeneity of tightly bound peptides. Peptide sequencing shows relatively conserved 'anchor' residues at position 2 and the carboxy-terminal residue. Conserved binding sites for the peptide N and C termini at the ends of the class I major histocompatibility complex binding groove are apparently dominant in producing the long half-lives of peptide binding and the peptide-dependent stabilization of the class I molecule's structure.

Class I histocompatibility glycoproteins bind peptides from the cytoplasm of cells and 'present' them at the cell surface for surveillance by cytotoxic T lymphocytes (CTL)¹. Foreign peptides, such as those from viral infections, complexed with major

histocompatibility complex (MHC) molecules present a novel antigenic surface to CTL, and are therefore recognized, stimulating CTL to kill the invaded cell. Peptide binding must be extremely 'tight', with a long half-life, in order to resist dissociation at the cell surface where the free peptide concentration is essentially near zero. Studies of peptides eluted from class I MHC molecules show that nonamers (that is, peptides that are nine amino acids long) are preferred²⁻⁷ and that peptides from each MHC allele have a sequence 'motif' with certain positions ('anchors') conserved or nearly so⁵ (reviewed in ref. 8).

The structure of HLA-Aw68, previously determined at 2.6 Å resolution⁹, has been determined and refined to 1.9 Å resolution from X-ray diffraction data collected from a single crystal frozen at -165 °C (see Fig. 1 legend). At 1.9 Å resolution, the electron density of the bound peptides is very readily interpreted at both ends of the HLA binding groove (Fig. 1a) as an atomic model of the N-terminal and C-terminal parts of bound peptides. Models of Glu-Val-Ala and Ala-Arg have been built into the first three (P1, P2, P3) and last two (PC-1, PC carboxy-terminal) positions of the peptide, respectively (based on eluted sequences). The real-space correlation coefficients for the main chain, a measure of how well the peptide atomic model fits the ($F_o - F_c$) difference electron density¹⁰, are reasonably good for P1 (0.9), P2 (0.89) and PC (0.92), and comparable to that for the whole protein model (0.93) fit to the 2 $F_o - F_c$ electron density map. By contrast, the middle part of the peptide density is sparse and unconnected (Fig. 1b, c), presumably representing an average of diverse conformations and/or lengths. The peptide model at the P3 and PC-1 positions, which might be positioned differently for various conformations or lengths in the peptide's centre, have correspondingly lower real-space correlation coefficients, 0.75 for P3 and 0.83 for PC-1.

The collection of peptides bound to HLA-Aw68 makes the same atomic contacts from the peptide's main chain and termini to a set of conserved residues at both ends of the binding cleft (Fig. 2), as observed from a single influenza virus peptide to HLA-Aw68¹¹ and from a collection of endogenous, nonameric

TABLE 1. Self-peptide sequences from HLA-Aw68

Cycle	Peptide sequence											M_{calc}	M_{obs}	Homologous protein
	1	2	3	4	5	6	7	8	9	10	11			
E	V	A	P	P	E	Y	H	R				1,097.2	1,097 ± 1	—
A	V	A	A	V	A	A	R	R				884.0	884 ± 1	—
D	V	F	R	D	P	A	L	K				1,060.2	1,060 ± 1	Ribosomal 60S homologue
E	V	A	P	P	E	Y	H	R	K			1,225.4	1,225 ± 0.5	—
E	V	I	L	I	D	P	F	H	K			1,210.4	1,213 ± 4	—
T	V	F	D	A	K	R	L	I	G	R		1,275.5	ND	Human Hsp70 protein B/Hsc70
K	T	G	G	P	I	Y	K	R						Influenza Np 91-99

Peptides were isolated, HPLC-purified and then microsequenced using an ABI model 477A protein sequencer as described in ref. 7. M_{calc} , calculated mass of peptides in daltons; M_{obs} , observed mass determined using either a Finnigan TSQ700 or LaserMat mass spectrometer; ND, not determined; Np peptide from refs 11 and 18. P2 and the C-terminal residues (boxed) are discussed in the text. The self peptides sequenced have a minimum occupancy relative to HLA-Aw68 from 0.12 to 0.4% based on initial amino-acid yields from sequencing and the starting amount of HLA-Aw68 (20 nanomol). (The residues at P1 and P3 are also consistent with the HLA-Aw68 structure as described here and in the accompanying paper¹¹.) Asp and Glu at P1 could contact Arg 62 and non-polar residues at P3; Phe, Ile and Ala face a non-polar pocket similar to that described for HLA-B27 at P3 (ref. 12).

FIG. 1 Endogenous peptides bound to HLA-Aw68. *a*, Side view of the electron density within 1.8 Å of the first three and last two residues of the collection of endogenous peptides bound to HLA-Aw68. The 1.8 Å cutoff only truncated density at the peptide's centre, as shown below. *b*, Side view as in *a*, but including unconnected density near the peptide's middle. *c*, Top view, including unconnected density near the peptide's middle. (The α_1 α -helix runs from left to right. An atomic model of Glu-Val-Ala is included in the left side (N-terminal) density and Ala-Arg in the right side (C-terminal) density.) Figure generated with FRODO²⁶. Maps are calculated with coefficients ($F_o - F_c$) in the resolution range 12–1.9 Å and contoured at 2.3σ , where F_o is the observed structure factor amplitudes and F_c are amplitudes calculated from the atomic model (described below). To avoid bias, no atomic model for bound peptides was included in the refinement of the protein, or to generate the electron density maps shown in this report.

METHODS. 1.9 Å X-ray diffraction data were collected from one crystal frozen in a film of buffer²⁷ containing 20% glycerol at -165°C on a Xentronics Detector²⁸ and processed with BUDDHA²⁹, $R_{\text{merge}} = 7.3\%$. The structure was determined by molecular replacement using the 2.6 Å resolution room temperature coordinates⁹ and refined with XPLOR³⁰ (details to be published elsewhere). The current atomic model of 2,998 non-hydrogen protein atoms and 186 water molecules (no peptide included) has an R -factor = 21.8% ($F > 3\sigma$ 6–1.9 Å), $R_{\text{free}} =$ (ref. 31) 27.7% ($F > 0, 6$ –1.9 Å), r.m.s. $\Delta_{\text{bonds}} = 0.017$ Å; r.m.s. $\Delta_{\text{angles}} = 3.2^\circ$.

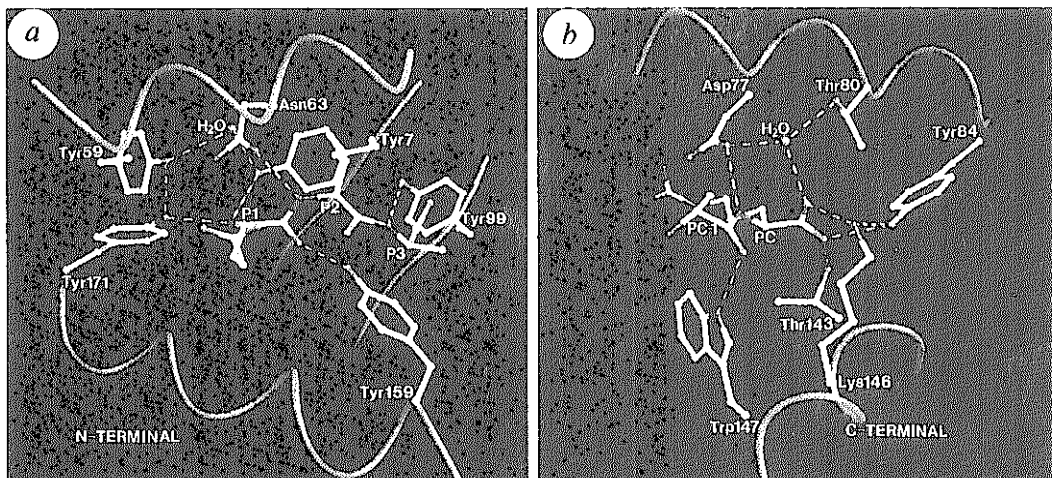
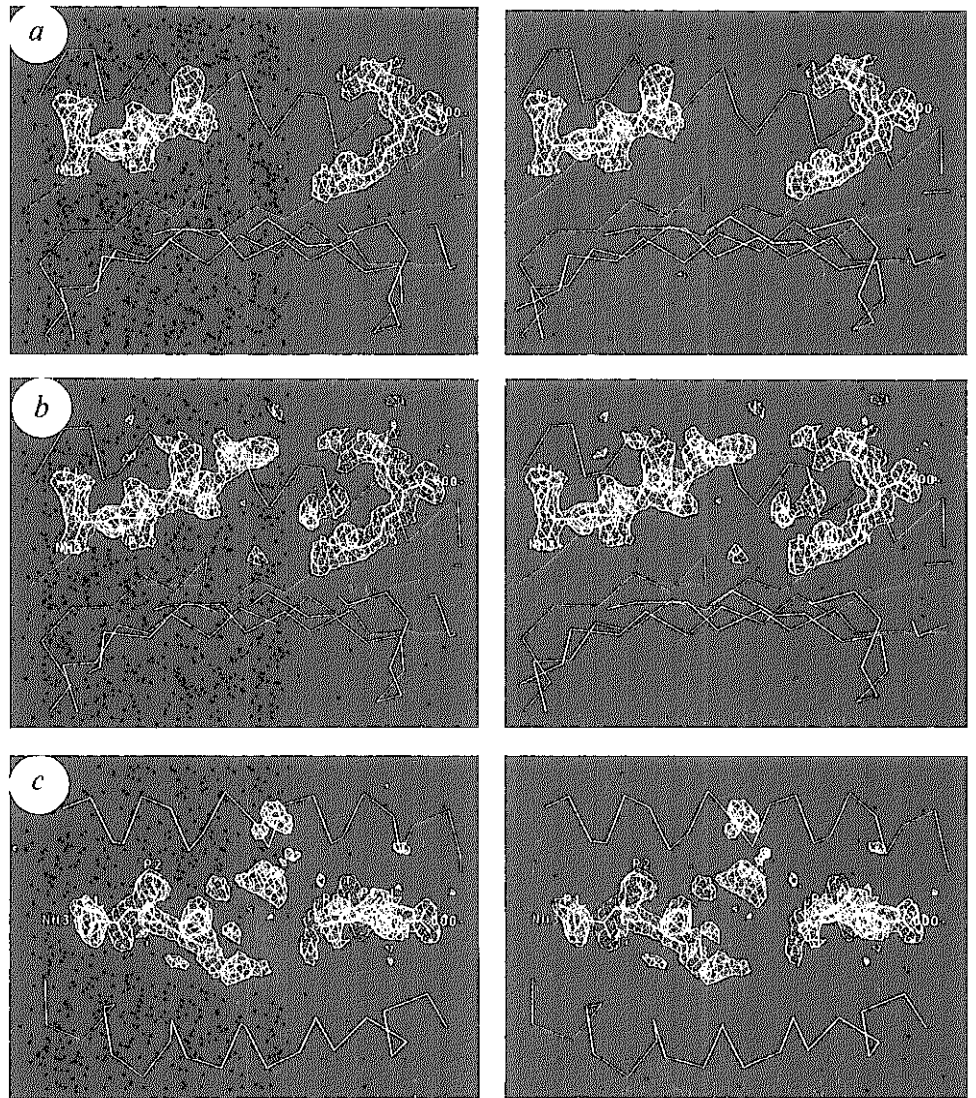


FIG. 2 A detailed view of the peptide model and hydrogen-bonding network at the N (*a*) and C (*b*) termini of the collection of endogenous peptides bound to HLA-Aw68. *a*, Potential hydrogen bonds (green dotted lines to conserved residues, purple to others) (distances less than 3.2 Å) between the peptide N terminus and HLA-Aw68. *b*, Potential hydrogen bonds between the peptide's C terminus and HLA-Aw68. HLA-Aw68 α -carbon trace in red; bonds coloured by atom type: yellow, peptide carbon; white, HLA-carbon; blue, nitrogen; red, oxygen. Figure generated with HYDRA^{TER}, modified by S. Watowich and L. Gross from programs written by D. Bacon and W. Anderson (RASTER3D) and R. Hubbard (HYDRA). Figure view chosen for comparison with Fig. 5 of ref. 13.

peptides to HLA-B27^{7,12,13}. In all three cases, the peptides' termini link together conserved residues on the sides (Tyr 59, Tyr 159, Tyr 171, Tyr 84, Thr 143, Lys 146, Trp 147) and bottom (Tyr 7, Tyr 123) of the peptide binding groove in a network of hydrogen-bonds (green in Fig. 2). This concentration of interactions at the peptides' ends not only holds the peptides in the groove, but seem to explain the peptide-dependent stabilization of the HLA structure¹⁴. Similar contacts have also been observed for two peptides bound to murine H-2K^b (refs 15, 16).

The absence of connected density corresponding to the middle of the collection of peptides bound to HLA-Aw68 (P4 to PC-2) might be interpreted as an indication that short, nonameric peptides adopt different conformations as they extend from one end to the other. The suggestion that a single peptide might adopt different conformations when bound to MHC molecules¹⁷ is not supported by the crystallographic structures of a single influenza virus peptide (Np 91-99) bound to HLA-Aw68 (ref. 11) and two other peptides bound to murine H-2K^b (refs 15, 16), which show a unique conformation for those bound octamers and nonamers. In addition, the electron density corresponding to a collection of endogenous peptide nonamers⁷ bound to crystalline HLA-B27 showed one pathway of clear density at high resolution (2.1 Å), readily interpretable as a common conformation for the collection of bound nonameric peptides¹³.

These observations suggest that the electron density at the centre of the collection of peptides bound to HLA-Aw68 might be missing because the bound peptides are of different lengths and consequently bulge out of the binding site to different degrees. We acid-eluted, fractionated and sequenced endogenous peptides^{4,7} bound to HLA-Aw68. Six self peptides were sequenced with lengths ranging from 9 to 11 amino acids (Table 1). These peptides represent a random, although small sampling of the self peptides bound to HLA-Aw68, based on high-performance liquid chromatography peak height and separation, indicating that length heterogeneity is abundant in this pool. Furthermore, there is a small amino-acid 'anchor' at P2 (Val or Thr) and a positively charged amino acid 'anchor' (Arg or Lys) at the C-terminal position (boxed in Table 1) in six peptides sequenced so far, and in the influenza virus peptide, Np 91-99, that is restricted to HLA-Aw68 (refs 11, 18). This argues that

all of these peptides are bound at both ends by their charged termini and the P2 and PC anchors. Models of Val and Arg fit the electron density at P2 and PC with side-chain correlation coefficients of 0.81 and 0.84, comparable to the 0.89 for the HLA-Aw68 side chains. Peptides from 8 to 13 amino acids long have been reported in tight complexes with class I molecules and those sequences indicate that the extra length 'inserts' between anchor residues as found here (refs 2, 19-23, and A. Bertoletti *et al.*, personal communication).

It is not clear what limits the size of 'bulges' *in vivo* (for example, transporters, proteasomes, proteolysis) or *in vitro*, although length variability might be expected to be smaller for cases where the peptide residues contacting the MHC molecule are near the peptide centre (like H-2K^b, P5, and HLA-B27, P7). A large enough insertion might even prevent a T-cell receptor from approaching the MHC molecule's surface. In HLA-peptide complexes with nonameric peptides, more than one half of the peptide side chains are exposed, primarily in the middle of the sequence, to recognition by T cells¹¹⁻¹³. Where longer peptides can bulge out of the binding site, as in HLA-Aw68, each 'bulged' position can probably accommodate many different amino acids, substantially increasing the diversity of sequences that can be presented.

Crystallographic and sequence data argue that interactions at the ends of peptides dominate in binding to HLA-Aw68. The conservation of the HLA sites for binding peptide termini suggests that the same substantial number of interactions for binding a peptide and for stabilizing the MHC molecule's structure (Fig. 2) will be found in all tight peptide/class I molecule complexes. 'Anchor' peptide residues bind in polymorphic pockets that differ among alleles^{9,11-13,24,25} and so might be expected to supply variable amounts of stabilization both to the peptide-HLA interaction and to the HLA molecule's stability. Polymorphic pockets therefore presumably determine which peptide sequences can bind^{9,24} in such a way that the peptides' ends will fit into both of the conserved peptide terminal binding sites. This suggests that short peptide termini joined by a long non-peptide linker could be designed to bind to HLA, achieving both long half-life, 'tight' binding and HLA structure stabilization. □

Received 4 August; accepted 2 October 1992.

1. Brodsky, F. M. & Guagliardi, L. E. *A. Rev. Immun.* **9**, 707-744 (1991).
2. Rötzschke, O. *et al. Nature* **348**, 252-254 (1990).
3. van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213-216 (1990).
4. van Bleek, G. M. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11032-11036 (1991).
5. Falk, K., Rötzschke, O., Stevanović, S., Jung, G. & Rammensee, H.-G. *Nature* **351**, 290-296 (1991).
6. Hunt, D. F. *et al. Science* **255**, 1261-1263 (1992).
7. Jardotzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. *Nature* **353**, 326-329 (1991).
8. Rötzschke, O. & Falk, K. *Immun. Today* **12**, 447-455 (1991).
9. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *Nature* **342**, 692-696 (1989).
10. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110-119 (1991).
11. Silver, M. L., Guo, H.-C., Strominger, J. L. & Wiley, D. C. *Nature* **360**, 367-369 (1992).
12. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353**, 321-325 (1991).
13. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Cell* **70**, 1035-1048 (1992).
14. Townsend, A. *et al. Nature* **340**, 443-448 (1989).
15. Zhang, W., Young, A. C. M., Imarai, M., Nathenson, S. G. & Sacchettini, J. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8403-8407 (1992).
16. Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. *Science* **267**, 919-927 (1992).
17. Peccoud, J., Dellabona, P., Allen, P., Benoist, C. & Mathis, D. *EMBO J.* **9**, 4215-4223 (1990).

18. Cerundolo, V., Tse, A. G. D., Salter, R. D., Parham, P. & Townsend, A. *Proc. R. Soc.* **244**, 169-177 (1991).
19. Robbins, P. A. *et al. J. Immun.* **143**, 4098-4103 (1989).
20. Silver, M. L., Parker, K. C. & Wiley, D. C. *Nature* **350**, 619-622 (1991).
21. Wei, M. L. & Cresswell, P. *Nature* **356**, 443-446 (1992).
22. Henderson, R. A. *et al. Science* **255**, 1264-1266 (1992).
23. Deres, K. *et al. Eur. J. Immun.* **22**, 1603-1608 (1992).
24. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. *J. molec. Biol.* **219**, 277-319 (1991).
25. Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
26. Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
27. Tong, T.-Y. *J. appl. Crystallogr.* **23**, 387-391 (1990).
28. Durbin, R. M. *et al. Science* **232**, 1127-1132 (1986).
29. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Crystallogr.* **20**, 235-242 (1987).
30. Brünger, A. T. *X-PLOR Manual* (Version 2.1) (Yale University, New Haven, 1990).
31. Brünger, A. T. *Nature* **355**, 472-475 (1992).

ACKNOWLEDGEMENTS. We acknowledge the earlier contributions made by M. Saper and P. Bjorkman to the determination of the structure of HLA-Aw68 to 2.6 Å resolution⁹. We thank J. Gorga for supervising the protein purifications, A. Haykov and K. Svenson for excellent technical assistance and R. Robinson for peptide separation H.-C.G. and T.S.J. are Cancer Research Institute fellows. J.L.S. acknowledges support from the NIH. D.C.W. is supported and T.P.J.G. was supported by the Howard Hughes Medical Institute.

