Three-dimensional structure of a peptide extending from one end of a class I MHC binding site

Edward J. Collins*, David N. Garboczi*† & Don C. Wiley†‡

* Department of Molecular and Cellular Biology,
† Howard Hughes Medical Institute, Harvard University,
7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
‡ To whom correspondence should be addressed

Class I major histocompatibility complex (MHC) molecules present peptides to $CD8^+$ T cells for immunological surveillance (reviewed in ref. 1). The structures of complexes of class I MHC molecules with octamer, nonamer and decamer peptides determined until now[2–8] show a common binding mode, with both peptide termini bound in conserved pockets at the ends of the peptide binding site. Length variations were accommodated by the peptide bulging[5,6] or zig-zagging[4] in the middle. Here we describe the structure of a decamer peptide which binds with the carboxy-terminal residue positioned outside the peptide binding site. Several protein side chains have rearranged to allow the peptide to exit. The structure suggests that even longer peptides could bind. The energetic effect of the altered mode of binding has been assessed by measuring the stability of the complex to thermal denaturation. Peptides bound in this novel manner are stable at physiological temperature, raising questions about their role in T-cell recognition and their production by proteolytic processing.

TABLE 1 Melting temperatures (T_m) of complexes containing peptides related to MLLSVPLLLG

Peptide	Source	T_m(°C)	ΔT_m(°C)
MLLSVPLLL	Calreticulin	65.8	—
MLLSVPLLLG	Calreticulin	51.7	−14.1
MLLSVPLLLGR	Experimental	48.0	−17.8

Each HLA-A2.1/peptide complex was purified and melting temperature assayed as described previously[14]. Briefly, the complexes (4–12 μM) were subjected to increasing temperature while the change in circular dichroic signal at 218 nm was monitored every 1.0 °C in an Aviv 62DS spectropolarimeter. The T_m was derived from the least-squares fit of the plot of molar ellipticity versus temperature using a two-state unfolding algorithm[14]. Data shown above are the average of two independent experiments with 2–3 melting curves per experiment. The error associated with the fitted parameter T_m is less than 0.5 °C. ΔT_m is the change in the melting temperature from that of the nonamer MLLSVPLLL.

Following the observation that long peptides without appropriately positioned anchor residues bind class I molecules *in vivo*[9-13] and the suggestion that a decamer from the signal sequence of calreticulin might not bind in the carboxylate end of the binding site of HLA-A2[10], we determined the structure of that decamer (MLLSVPLLLG) complexed with HLA-A2 by X-ray diffraction (Fig. 1, methods). Initial electron density maps were interpretable for the entire length of the peptide. The C-terminal glycine projects up out of the peptide binding cleft (Figs 1 and 2*a*). The first nine residues of the decamer bind to HLA-A2.1 almost identically to conventional nonamers; the locations of the amino terminus and the P2 anchor residue overlap those from a collection of viral peptides bound to HLA-A2 (Fig. 2*b*), and the P9 side chain and carbonyl oxygen overlap the P9 anchors and one oxygen of the carboxylate groups of conventionally bound nonamers (Fig. 2*c*). The pattern of hydrogen bonds made at the C terminus is different, however. In the conventional mode of binding (Fig. 2*d*), four hydrogen bonds are made to the peptide carboxylate oxygens, three from conserved MHC residues (Thr 143, Lys 146, Tyr 84) and one from a non-conserved residue (Thr 80 through H_2O). Only one of those bonds (Thr 143) is preserved in the calreticulin decamer complex, although one new hydrogen bond is also formed between Lys 146 and the exiting P10 carboxylate group of the calreticulin decamer (Fig. 2*e*).

Two of the four class I residues that participate in hydrogen bonding to conventionally bound peptides have relocated in the calreticulin peptide/HLA-A2 complex. Tyr 84 has rotated away from the peptide P9 carbonyl removing the potential for a hydrogen bond (Fig. 2*e*). Lys 146 has raised out of the site to form a salt bridge with the P10 carboxylate. These movements create an opening through which the peptide C terminus can extend out into solvent (compare Fig. 3*a* and *b*).

We have assessed the energetic consequences of this novel binding mode by comparing the thermal stability of the calreticulin decamer complex to that of a complex with a peptide with the same first 9 amino acids (MLLSVPLLL), containing anchor residues at conventional positions (P2 Leu, P9 Leu) (Table 1). The complex with the nonamer peptide thermally denatures at 65.8 °C, a temperature comparable to the naturally occurring antigenic peptide of the influenza virus M_1 protein (residues 58–66) complexed to HLA-A2[14]. The decamer, MLLSVPLLLG, denatures at 14 °C lower, 51.7 °C. This 14 °C difference is about half that observed (−23 °C) when all four C-terminal hydrogen bonds (Fig. 2*d*) are eliminated by replacing the carboxylate group with a methyl group on the influenza virus M_1 peptide[14]. Because the crystal structure of the calreticulin decamer suggests that longer peptides might extend even farther out of the binding site, we tested a synthetic 11-residue oligonucleotide with an arginine (a non-anchor residue for HLA-A2) added to the C terminus of the decamer. That complex denatured at 48.0 °C, an additional loss of only 3.7 °C compared to the decamer, arguing that it also extends out of the site.

Peptides bound to class I molecules *in vivo* are derived from cellular proteins by one of two mechanisms. Most appear to arise from cytoplasmic proteolysis followed by transport into the endoplasmic reticulum (ER) (reviewed in ref. 1), but some derive from signal sequences that enter the ER lumen as part of a secreted protein[9,15]. The decamer (MLLSVPLLLG) studied here is the N-terminal 10 amino acids of the 17 amino-acid signal sequence of calreticulin. It is found bound to HLA-A2 even in cells lacking peptide transporter functions[9]. The nonamer, MLLSVPLLL, binds with substantially higher affinity than the decamer[10] (Table 1), yet is not detected bound to HLA-A2 *in vivo*[9]. These observations suggest that the nonamer is not produced *in vivo*.

It has been suggested that peptides may be trimmed[16], by for example a carboxy peptidase, while bound to class I molecules. The X-ray structure (Fig. 3*b*) shows that the peptide carboxylate oxygens, even in this peptide which extends out of the site, are bound too closely to the MHC protein to fit simultaneously into the active site of a protease. Instead, trimming of successive

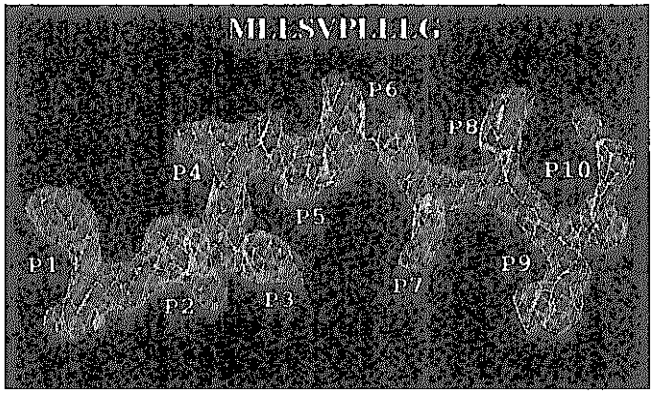

FIG. 1 'Omit' $||2F_o|-|F_c||$ electron density for the peptide MLLSVPLLLG. The electron density is calculated from phases derived from iterative cycles of non-crystallographic real-space averaging of the two molecules in the asymmetric unit using a model without peptide. The path of the peptide is clearly visible and the glycine at P10 is located out of the peptide binding site.

METHODS. Soluble HLA-A2 complexed with the peptide MLLSVPLLLG was produced as described previously[20] and co-crystallized. X-ray data were collected to 1.9 Å resolution on supercooled crystals (−165 °C) as 0.15° oscillation frames with a Xentronics detector on an Elliot GX-13 rotating anode generator with a 100 μm focal cup and focusing mirrors[21]. The space group was determined by examination of the differences in intensities of potential pairs of reflections across putative mirror planes within the data[22] and confirmed by molecular replacement and by examining the electron density maps. The space group is P1 with a unit cell of $a=50.41$ Å, $b=62.91$ Å, $c=74.81$ Å, $\alpha=81.99°$, $\beta=76.43°$, $\gamma=78.06°$. The R_{merge} for data is 5.4% from 30.0–2.07 Å, and 20.5% in the resolution shell from 2.16–2.07 Å. The molecular replacement solution was found by a rigid body fit of the reverse transcriptase peptide/HLA-A2.1 model[5] with X-PLOR[4]. The initial correlation coefficient before the rigid body fit was 64.8% from 10.0 to 2.5 Å. Cycles of real-space averaging with phase extension (G. J. Kleywegt and T. A. Jones, manuscript in preparation), manual model building using the program O[23], and automated refinement using X-PLOR[24] were done to refine the model to 2.0 Å resolution. Although the automated refinement target was the residual based on data in the subset R_{work}, the independent parameter R_{free} (ref. 25) had to drop for the cycle to be considered successful. The final R_{free} is 29.6%, the R_{work} is 23.4% from 6.0–2.0 Å with all structure factors over 2 sigma. Each final monomer model has 251 bound waters. The geometric deviations from ideality are: r.m.s. bonds = 0.018 Å; and r.m.s. angles = 3.35 degrees. $R_{free}=(\Sigma_h|F_o-F_c|)/(\Sigma_hF_o)$, $\forall h\in\{\text{free set}\}$; $R_{work}=(\Sigma_h|F_o-F_c|)/(\Sigma_hF_o)$, $\forall h\in\{\text{working set}\}$. $R_{merge}=(\Sigma_{hkl}|I-\langle I\rangle|)/(\Sigma_{hkl}\langle I\rangle)$, $\forall hkl\in\{\text{independent Miller indices}\}$.

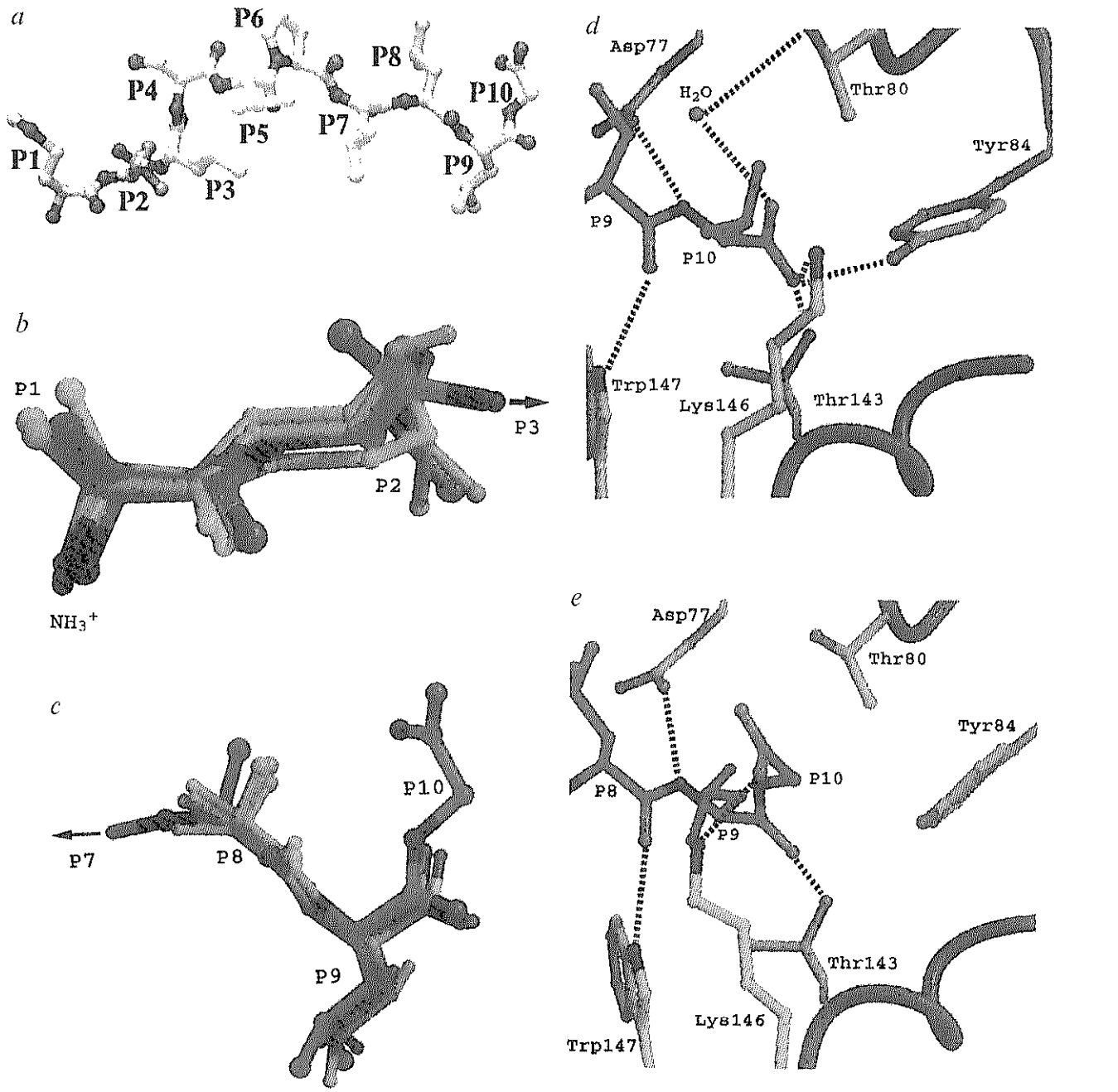

FIG. 2 The decamer, MLLSVPLLLG, binds to HLA-A2.1 in the conventional manner until residue P9. *a*, The path of the entire peptide. *b*, A superposition of the first two residues from the four nonamers previously determined[4] in complex with HLA-A2.1 (grey bonds) and the first two residues from MLLSVPLLLG (blue bonds). The side chains of the P1 residues have been removed for clarity (N-terminal nitrogen dark blue; P1 carbonyl oxygen red). *c*, A superposition of the last two residues of four nonamers[4] (grey bonds) in complex with HLA-A2.1 and the final three residues of MLLSVPLLLG complexed with HLA-A2.1 (blue bonds). The P8 side chains have been removed for clarity (carbonyl oxygens red; amide nitrogens dark blue). *d*, An example of conventional hydrogen bonds, residues P9 and P10 of hepatitis B nucleocapsid peptide decamer (FLPSDFFPSV) bound to HLA-A2.1[4]. *e*, Hydrogen bonds for residues P8, P9 and P10 of calreticulin peptide (MLLSVPLLLG). *b*–*e*, were produced with the program HYDRASTER which was written by S. Watowich and L. Gross and based in part on programs by W. Anderson (RASTER3D) and R. Hubbard (HYDRA). The four nonamers in *b* and *c* and sources are: ILKEPVHGV (HIV-1 reverse transcriptase); LLFGYPVYV (HTLV-1 Tax); GILGFVFTL (influenza virus M peptide); TLTSCNTSV (HIV-1 gp120)[22].

C-terminal peptide residues might occur by a cycle of dissociation of a few C-terminal residues, proteolytic cleavage of the exposed terminal residue, and rebinding. By such a 'nibbling' mechanism a nonamer might never be produced if the decamer bound tightly enough to be transported to the cell surface with the MHC molecule.

The extraction of long peptides from class I molecules *in vivo*[9,13] and the thermal stability well above 37 °C measured here, suggest that the unconventional mode of peptide binding observed crystallographically may be found on cell surfaces. It has been argued from both peptide pool sequences[13] and thermodynamic data[14], that peptides could extend from either end of the binding site. The peptide extensions out of the binding site must be geometrically different from those found on class II MHC molecules (compare Fig. 3*a* here with Fig. 4 in ref. 18) because of differences in the shape of the binding sites[17,19]. It will

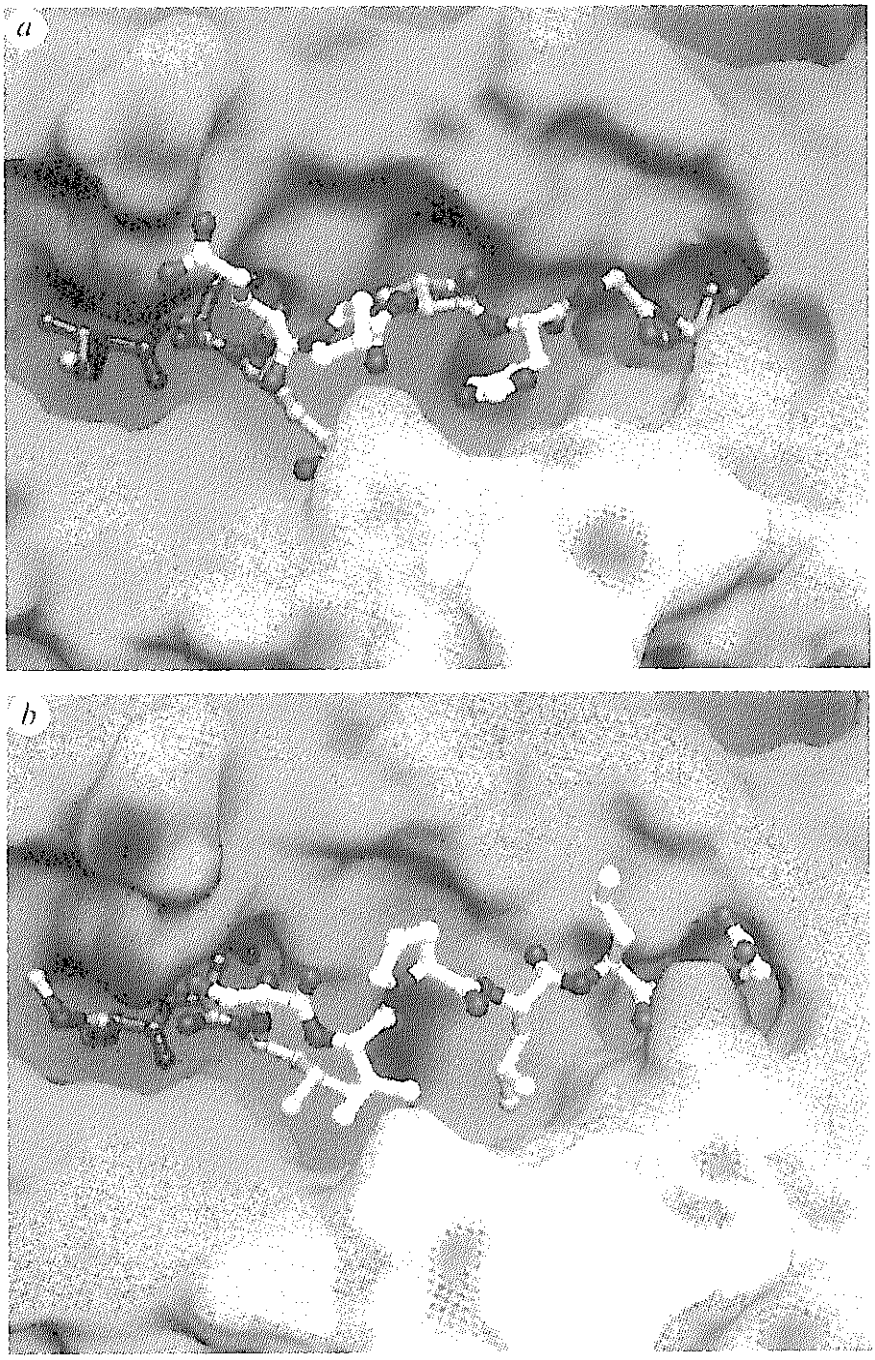

FIG. 3 The molecular surface of HLA-A2.1 complexed with a 'conventionally bound' peptide (*a*) and MLLSVPLLLG (*b*). *a*, HLA-A2.1 with ILKEPVHGV. The C terminus of the peptide is on the right-hand side. Lys 146 of HLA-A2.1 (visible as a positive charge, blue) forms a 'cover' under which the C terminus of the peptide is bound[4]. *b*, HLA-A2.1 with MLLSVPLLLG (calreticulin). The movement of Tyr 84 and Lys 146 (visible as a positive charge, blue) forms a groove through which the peptide C terminus exits. The electrostatic potential colouring of Figs 2*a*, 3*a* and *b* were produced using the program GRASP (A. Nicholls, personal communication).

be interesting to see if the novel surfaces presented by peptides extending out of the ends of class I MHC binding sites are recognized by $\alpha\beta$ or $\gamma\delta$ T cells. □

Received 22 June; accepted 19 August 1994.

1. Germain, R. N. & Margulies, D. H. *A. Rev. Immun.* **11,** 403–450 (1993).
2. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353,** 321–325 (1991).
3. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Cell* **70,** 1035–1048 (1992).
4. Madden, D. R., Garboczi, D. N. & Wiley, D. C. *Cell* **75,** 693–708 (1993).
5. Guo, H.-C. *et al. Nature* **360,** 364–366 (1992).
6. 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).
7. Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. *Science* **257,** 919–927 (1992).
8. Young, A. C. M., Zhang, W., Sacchettini, J. C. & Nathenson, S. G. *Cell* **76,** 39–50 (1994).
9. Henderson, R. A. *et al. Science* **255,** 1264–1266 (1992).
10. Chen, Y. *et al. J. Immun.* **152,** 2874–2881 (1994).
11. Huczko, E. L. *et al. J. Immun.* **151,** 2572–2587 (1993).
12. Olsen, A. C. *et al. Eur. J. Immun.* **24,** 385–392 (1994).
13. Urban, R. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **91,** 1534–1538 (1994).
14. Bouvier, M. & Wiley, D. C. *Science* **265,** 398–402 (1994).
15. Wei, M. & Cresswell, P. *Nature* **356,** 443–446 (1992).
16. Rotzschke, O. *et al. Nature* **348,** 252–254 (1990).
17. Brown, J. H. *et al. Nature* **364,** 33–39 (1993).
18. Stern, L. J. *et al. Nature* **368,** 215–221 (1994).
19. Stern, L. J. & Wiley, D. C. *Structure* **2,** 245–251 (1994).
20. Garboczi, D. N., Hung, D. T. & Wiley, D. C. *Proc. natn. Acad. Sci. U.S.A.* **89,** 3429–3433 (1992).
21. Harrison, S. C. *J. Appl. Crystallogr.* **1,** 84–90 (1968).
22. Garboczi, D. N., Madden, D. R. & Wiley, D. C. *J. molec. Biol.* **239,** 581–587 (1994).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* **A47,** 110–119 (1991).
24. Brünger, A. *X-PLOR (version 3.0) Manual* (Yale University, New Haven, Conn, 1992).
25. Brünger, A. T. *Nature* **355,** 472–475 (1992).

ACKNOWLEDGEMENTS. We thank V. H. Engelhard and Y. Chen for sharing data before publication, J. Strominger and colleagues for continued scientific collaborations on MHC molecules and L. Stern and T. Jardetzky for critical reading. E.J.C. was supported by the Irvington Institute for Medical Research and D.N.G. by a NIH National Research Service Award. D.C.W. is an Investigator of the Howard Hughes Medical Institute. The coordinates have been deposited with the Protein Data Bank, Brookhaven, New York.