The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation

D. R. Madden*, J. C. Gorga†*, J. L. Strominger† & D. C. Wiley†*

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

X-ray crystallography reveals electron density in the antigen-binding site of HLA-B27 that is an interpretable image of nonameric peptides in a largely extended conformation. Clear density exists for the main chain and several side chains and is consistent with the sequence of 11 nonameric self-peptides eluted from HLA-B27 (see accompanying article). Pockets in the antigen-binding cleft bind four side chains and the amino and carboxyl termini of the peptide.

**Cytotoxic** events in the cellular immune response begin with specific cell-to-cell recognition. The hypervariable immunoglobulin-like T-cell receptor on the surface of cytotoxic T cells (CTL) binds to the complex between a short peptide processed from a protein antigen and a target cell's class I major histocompatibility glycoprotein (reviewed in ref. 2). The range of peptides that can be presented for immune surveillance is enhanced by the coexpression of several class I histocompatibility antigens on the cell, and, more importantly, by the ability of each histocompatibility molecule to bind many peptides of diverse sequences. In the population the major histocompatibility complex (MHC) antigens are highly polymorphic, having been selected to increase the diversity of peptides that can be presented to CTL (refs 3, 4).

The three-dimensional structures of the human class I histocompatibility antigens HLA-A2 and HLA-Aw68 revealed a peptide-binding groove lined with polymorphic residues5,6,7. Electron density bound in the groove, although not definitively interpretable, seemed to represent short peptides bound to the crysalline HLA (refs 3, 3, 6), and a polyalanine mer was modelled into the density associated with HLA-A2 (ref. 7; M. A. Saper, unpublished data). The inability to dissociate the endogenous mixture of peptides and readily prepare molecules with one peptide bound8 has made it difficult to determine either how one peptide binds to a given site or how a given site can accommodate any one of a large number of peptides with different sequences. Indeed, peptide binding seems to be associated with the folding and assembly of class I MHC molecules9,10 and only strategies involving reconstitution or de novo biosynthesis of class I molecules in the presence of single peptides will yield stoichiometric complexes11,15,38-40.

Here we examine the unusually clear electron density in the binding site of HLA-B27. It indicates that the amino-acid sequence of bound nonameric peptides would be restricted at four positions (P2, P3, P7, P8, where P6 is the 9th position in the peptide) to a B27-specific sequence motif, because these peptide side chains are bound in 'pockets' in the bottom and sides of the binding groove. To a varying degree, the pockets can accommodate several different peptide side chains, consistent with evidence from peptide sequence1. A T cell receptor could interact directly with the four or five peptide side chains (P1, 4, 5, 6, 8) that appear accessible and could perhaps indirectly sense amino-acid substitutions at the more buried positions through shifts in the peptide main chain. The main chain of the first two and last two peptide amino acids and the peptide's charged termini seem to bind to conserved polar atoms at both ends of the binding groove.

**Structural determination**

Because HLA-B2705 crystallized with two molecules in the crystallographic asymmetric unit16, an atomic model could be determined on the basis of the coordinates of HLA-A2 but free of bias from that structure. Iterative real-space phase averaging yielded high-quality electron density maps from starting models with one domain omitted at a time, and the model for that domain was built only into clear density (see Table 1 caption). The refined model was used to calculate new maps, which were again improved by phase averaging. Until the final stage, the model was extended only into clear electron density. Throughout refinement, both averaged and unaveraged maps were examined, and non-crystallographic symmetry constraints or restraints were applied. The final structure has a crystallographic R factor of 18.5% from 6.0-3.0 Å, and has reasonable geometric parameters (see Table 1). The structure of HLA-B27, the histocompatibility antigen associated with the disease ankylosing spondylitis (reviewed in ref. 17), is very similar to that of HLA-A2 and HLA-Aw68 (refs 5-7), as is to be expected from the 90% sequence identity. Details of the structure will be described after high-resolution refinement (D.R.M. et al., in preparation).

Electron density suggestive of a short peptide was observed in the peptide-binding groove in the earliest maps, as well as in the final refined maps (Fig. 1a, b). Presumably, it is the averaged image of a mixture of different bound peptides1. To analyse the protein-peptide interaction, we built several peptide models and subjected them to refinement. But to avoid introducing bias into the electron density maps the peptide models were not included in the HLA-B27 refinement or in any of the electron density maps discussed here. The geometry of the current peptide model is standard; a data base search using the diagonal facility of FRADO (refs 18, 19) reveals a number of overlapping tetra- and pentapeptides with similar geometries.

**Peptide conformation**

The current model for a peptide in the cleft is a nonamer in a largely extended conformation, with side chains projecting in a roughly alternating pattern, as is found in extended B strands. There is a kink at amino acids three and four, such that the side chain of position 3 (P3) points forward along the direction of the peptide, fitting under the main chain of P5. The side chain of P4 points up (Fig. 1c). P1 and P2 are close to the floor of the cleft, with the P2 side chain in the '4 pocket' previously described15 (Fig. 2). There is no side chain density for P1, but a long side chain could reach solvent. The segment from P3 to
FIG. 1 The peptide model and extra electron density in the peptide binding cleft of HLA-B27. a: Top view of the cleft, with the α1 α-helix at the top of the figure. Red: Co trace of the HLA-B27 binding groove helices, with Trp 147 side chain in yellow. Yellow: peptide model blue. 2Fo-Fc real-space phase-averaged electron density within an envelope describing the cleft. contour equals 1σ. b: Side view of the cleft, looking from the α2 α-helix (absent) towards the α1 α-helix. c: Peptide model, viewed as in b.

METHODS. As an initial model of the bound peptide, a polyalanine octapeptide was built into the observed extra density with its N terminus at the left-hand (Y171) end of the cleft, and regularized using FRODO. Hypothetical side chains were added at several positions, as suggested by the electron density and the binding environment. Another model with the reverse sequence was built into the density with its N terminus at the right-hand (Y164) end of the cleft. The two models were subjected to a simulated annealing slow-cooling protocol from 4,000 K to 300 K, and only the one with its N terminus at the left-hand end (Tyr 171) of the cleft achieved an acceptable peptide conformation. This peptide also fits the density somewhat better, with its penultimate carbonyl oxygen in a prominent electron density bulge and in a position to hydrogen-bond to the pyrrole nitrogen of the conserved Trp 147. An additional residue was added at the N terminus, yielding a nonamer. This basic structure was modified to accommodate biochemical sequence information as it became available. The model peptide (sequence ARYAALSTEL) is based on the histone peptide sequence eluted from HLA-B27 (ref. 1) with alanine at positions 1, 4, 5.

FIG. 2 Side chain specificity. The charged and polar side chains of HLA-B27 in the '45 pocket', together with model of peptide side chain P2 (arginine, green) and electron density (blue, contoured as in Fig. 1).
TABLE I. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Fraction complete</th>
<th>Number of independent reflections</th>
<th>Fraction I &gt; 2σ</th>
<th>Rmerge</th>
<th>Number of multiply-measured reflections</th>
<th>Rcryst</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0-4.10</td>
<td>0.97</td>
<td>1,852</td>
<td>0.95</td>
<td>0.085</td>
<td>1,710</td>
<td>0.393</td>
</tr>
<tr>
<td>6.40-4.78</td>
<td>0.97</td>
<td>2,504</td>
<td>0.95</td>
<td>0.090</td>
<td>2,325</td>
<td>0.204</td>
</tr>
<tr>
<td>4.78-4.08</td>
<td>0.96</td>
<td>4,084</td>
<td>0.94</td>
<td>0.103</td>
<td>2,372</td>
<td>0.160</td>
</tr>
<tr>
<td>4.06-3.69</td>
<td>0.88</td>
<td>3,262</td>
<td>0.82</td>
<td>0.119</td>
<td>2,727</td>
<td>0.175</td>
</tr>
<tr>
<td>3.59-3.26</td>
<td>0.95</td>
<td>2,504</td>
<td>0.66</td>
<td>0.078</td>
<td>1.3</td>
<td>0.210</td>
</tr>
<tr>
<td>3.26-3.00</td>
<td>0.49</td>
<td>2,215</td>
<td>0.50</td>
<td>0.032</td>
<td>7</td>
<td>0.238</td>
</tr>
<tr>
<td>10-3 Å</td>
<td>0.75</td>
<td>15,211</td>
<td>0.61</td>
<td>0.096</td>
<td>8,154</td>
<td>0.218</td>
</tr>
<tr>
<td>9-3 Å</td>
<td>0.73</td>
<td>1,474-45</td>
<td>0.79</td>
<td>0.105</td>
<td>6,359</td>
<td>0.183</td>
</tr>
</tbody>
</table>

R.m.s. deviations in geometry: bond length: 0.017 Å; bond angle, 3.7°; dihedral angle, 25.9°. Data were collected from 5 crystals on a Xrnetronics area detector, processed using DUDU programs and scaled with the CCP4 package (P. Evans, personal communication). Initial phases were determined by molecular replacement (refs 33–35) and P. Fitzgerald, personal communication). The molecules are related by an approximately twofold rotational symmetry axis (179° rotation, 3 Å translation), within 1° of the crystallographic a axis, as expected from pseudo-1 mm precision photographs and the existence of a nearly isomorphous P3 crystalline. To derive a model independent of the HLA-AS structure, the four P. 50 amino-acid domains α1, α2, α3, and α4, as well as the glycosylation sites, were individually omitted from the model used to calculate the phases for density maps, which were each iteratively real-space phased-averaged to convergence. Each map yielded clear electron density for the omitted domain, and an unbiased partial model (69% complete) of the whole structure was constructed one domain at a time. Model was extended only into clear density, and its geometry regularized by FRODO and X-PLOR. The process was repeated until it yielded no new density. Subsequent refinement included least-squares positional and temperature factor refinement, as well as positional refinement by simulated annealing slow-cooling. Two molecular envelopes and non-crystallographic symmetry relationships were determined. In the final stage, the resolution range for refinement was 6.5-3 Å, and strict non-crystallographic symmetry restraints were relaxed. Several further rounds of map averaging, building and refinement yielded the final structure. Several charged or polar surface sidechains were without significant electron density. No water molecules have been included in the model, Rmerge = Σ(Il/⟨Il⟩) / Σ(Io), Rcryst = Σ[|Fo| - |Fc|] / Σ|Fo|. l, measured diffraction intensity; o, standard deviation on I. Fo and Fe observed and calculated structure factor amplitudes.

P9 is twisted so that P5 and P6 point roughly back and forth across the width of the binding site, with P5 towards the α2-α helix and P6 towards the α1-α helix, whereas side chains P7, P8, and P9 alternatively point approximately into, out of, and into the binding site (Fig. 1a, b).

Overall, P2, P3, P7 and P9 face inward to contact the HLA molecule, and the terminal carboxyl and amino groups fit into the ends of the grooves (Fig. 1b, c); P4 and P8 point up towards the solvent; P5 and P6 could be simultaneously in contact with the HLA molecule and accessible to direct contact by a T-cell receptor. The side chain density is relatively short for both P5 and P6. In terms of the previously described pocket, containing polymorphic residues in the antigen-binding site, peptide side chain P2 fits into pocket B (near 45), P3 into pocket D (near 156), P7 into pocket E (near 152), and P9 into pocket C (near 74 and 116). The N terminus and P1 are in pocket A, and the C terminus and part of P9 in pocket F. Although there is no electron density to suggest that longer peptides may be bound, model building demonstrates that it is possible to extend the peptide at either end. Longer peptides would presumably result in the loss of some hydrogen bonds to the charged terminal groups, as well as a salt bridge at the C terminus (see below).

Peptide specificity

The size and amino-acid composition of the sites binding P2, P3, P7, and to some extent P7, suggest restrictions on the side chains that could be accommodated at these positions in a peptide's sequence. The most obvious of these restrictions is at position P2, where clear electron density reaches deep into the largely hydrophobic pocket that ends near a negatively charged Givu 45, and Cys 67 (Fig. 2). This site would be complementary to a long positively charged peptide side chain. In fact, Jardetzky et al. find arginine at P2 in all eleven self-peptides eluted from HLA-B27, a result consistent with the sequences of three viral peptides known to bind to B27 (refs 20, 21; K. C. Parker et al., manuscript submitted).

Peptide side chain P3 binds in a predominantly non-polar site near Leu 156 which contains two aromatic side chains (Tyr 99 and Tyr 159), so that a hydrophobic or aromatic side chain would be expected on the peptide. A polar atom (from His 114) is accessible at the side of the pocket so that a tyrosine could also be accommodated. Most of the HLA-B27 specific peptide sequences examined have hydrophobic or aromatic side chains at P3. Residues as large as Trp and as small as Ser are found. The current model accommodates even these extremes, although it is possible that the binding of water molecules and small adjustments of the peptide main chain will occur in some cases.

P7 and P9 fit into two pockets, which in HLA-B27 form a large cavity between Asp 74 and Trp 147 which has a charged and polar surface. The electron density for the side chain at P7 is extensive, but bifurcated, pointing towards His 114 and Asp 97, and therefore difficult to interpret. This suggests that different side chains at P7 may be accommodated by contacting different residues in HLA-B27. Seven different side chains are found at P7 in the 14 HLA-B27 restricted peptide sets at P7 (ref. 1). Determining exactly how the binding site can accommodate, for example, either an arginine or a glutamic acid at P7 or whether for some peptide sequences P7 might rotate to point out of the site, will require higher resolution and possibly structures with single peptides bound. At position P9, the side chain could reach deep into the site to contact polar and negatively charged residues Asp 116 or Asp 77, or if shorter, to contact non-polar atoms of nearby Thr 143, Trp 147, Leu 81 or Tyr 123. The electron density at P9 does not fill the pocket, so shorter polar or even hydrophobic side chains may also be present. This is consistent with the six residues found at P9 so far (K, R, L, Y, N, A) (ref. 1).

Side chain P1 seems likely to influence the position of the peptide main chain atoms. A small side chain at this position permits considerable rotation around the residue 1 psi angle. A large side chain at position P1 would probably extend towards the solvent, restricting the main chain conformation and resulting in the loss of some hydrogen bonds (see later). Both small and large side chains are found at P1 in the available B27-restricted peptide sequences.

The side chains at positions 4, 5, 6, and 8 are all oriented towards the solvent or else horizontally within the cleft, such that the side chains are potentially exposed to solvent. The sequences at P4, 5 and 8 show the expected wide range of side chains. P6, although exposed to solvent, shows only short electron density pointing between alanine 68 and the aliphatic chain of Lys 70, both on the α1-α helix, which may restrict the permissible sequences. Consistent with this, P6 is restricted to small
polar and non-polar residues in the known 
pptide sequences. Further sequences 
may, however, reveal more diversity if the 
side chain can bend to point out of the site.

From model building experiments in which different side 
chains are placed at the protein-binding positions of the peptide, it 
is possible to accommodate the observed diversity of sequence 
with at most minor modifications of the overall main chain 
conformation.

**Peptide termini**

Two clusters of amino acids conserved in essentially all class I 
histocompatibility antigen sequences, one at each end of the 
peptide-binding groove, are positioned to contact polar 
atoms of the main chain of the first two and last two peptide 
argin amino acids and the charged termini, features common to all 
peptides. The importance of conserved residues for recognizing 
common aspects of peptides, such as their main chain and 
termini, was anticipated and a role for tyrosines 7 and 171 
and threonine 143 was demonstrated by site-directed 
mutagenesis (F. Latron et al., manuscript in preparation).

At the carboxyl terminus of the peptide, the positively charged 
ε-amino group of lysine 146 and the hydroxyl groups of tyrosine 
84 and threonine 143 are in positions to donate hydrogen bonds 

to the peptide carboxylate (Fig. 3b). Residues 146 and 84 are 
conserved, and residue 143 is threonine or serine so that the 
polar hydroxyl group is conserved. The carboxyl oxygen of the 
peptide backbone amino acid P8 appears positioned to hydrogen 

to the pyrrole nitrogen of the conserved tryptophan 147 (Figs 1a 
and 3b).

As already discussed, the interactions with the amino terminus are 
likely to depend on the P1 side chain. Hydrogen bonds to 
conserved tyrosines 59 (from the P1 carboxyl oxygen) and 171 
(from the amino-terminal nitrogen) are possibly independent 
of the nature of the P1 side chain (Fig. 3a). Given a small side 
chain, additional hydrogen bonds may be made to conserved 

![Diagram](image)

**Fig. 3 Binding environment of the peptide termini. a, N-terminal and b, 
C-terminal two residues of the peptide model (stick figure). Residues of 
HLA-B27 currently within hydrogen-bonding distance of polar atoms in the 
model (solid forward arrow) and those potentially interacting with peptide 
through water molecules or conformational shifts not ruled out by the density 
are labelled. Boxed, residue on the floor of the cleft; dashed arrow, hydrogen 
bond possible if P1 side chain is small.**

**Discussion**

From X-ray crystallographic evidence, the peptides bound to 
HLA-B27 seem to be nonaners whose conformation is that of 
ex tended β strand, with a kink at positions 3 and 4. The quality 
of the electron density from a probably heterogeneous mixture of 
peptides suggests that peptides of diverse sequence are 
accommodated within a relatively narrow range of backbone 
conformations. All the geometric features of this conformation 
are standard, and have been observed within a database of 
protein structures determined at high resolution by X-ray crystal-

lography. The N-to-C orientation is consistent with that 
suggested for HLA-A2 (F. Latron et al., manuscript submitted) 
and that suggested for a peptide bound to the class II histo-
compatibility antigen HLA-DR7 (ref. 22) on the basis of experi-
ments in which compensating changes in the histocompatibility 

tagon and peptide sequences were discovered. Models largely 

composed of extended structure have been proposed for pep-
tides bound to H2-Kb (ref. 23) and to class I (refs 22, 24-25) 
based on binding experiments and model building.

The peptide model is consistent with the pattern of sequence 
restriction observed in naturally processed self-peptides eluted 
from HLA-B27 (ref. 1). The side chains at positions 2, 3, 7 and
9 are in close contact with the protein, in binding pockets whose geometry and chemistry determines an HLA-B27-specific sequence motif: arginine at 2, bulky non-polar or tyrosine and some small residues at 3, a mixture at 7, and at 8 Arg or Lys preferred but not non-polar accepted. Sequence motifs have been suggested from binding data for class I molecules and from sequencing of the pool of bound peptide eluted from several class I alleles (ref. 26; G. M. Van Bluck and S. C. Nathenson, manuscript submitted). The particular stringency of binding at position P2 (14/14 Arg) coincides with the depth and narrowness of the '45 pocket'. The importance of a single amino acid is reminiscent of peptide binding to HLA-DR1, where a polyalanine peptide containing one tyrosine binds as well as the viral peptide from which it was derived. Other pockets are more accommodating, binding any one of a small set of peptide side chains. P3, P7 and P9 bind in comparatively broad pockets, consistent with the large variation in residues found at these positions. P7 and P9 bind in a particularly wide cavity lined with polar and charged residues, and the density at P7 suggests that side chains may fit into more than one location within this pocket. In general, accommodation may also require the involvement of water molecules and in some cases movement of the peptide main chain which may be visible to the T-cell receptor. At P6 the electron density is too poor to definitively explain the apparent sequence restriction at that position (see above). About one half of the bound peptide's side chains (P4, 5, 6, 8 and possibly 1) appear accessible enough to be contacted directly by the T-cell receptor and it is likely that the others, especially P3 and P7, could be recognized indirectly if they perturb the peptide backbone's location.

The polar atoms from conserved amino acids at the two ends of the binding site are here observed to contact the main chain polar atoms of the first two and the last two amino acids and the charged amino and carbonyl termini. The pockets at the ends of the class I MHC antigen-binding site therefore seem to be designed to recognize the ends of short peptides. This is consistent with the findings that peptides eluted from class I molecules are octamers and nonamers, and that the natural short peptide eluted from H2-D b forms a much more stable complex than peptides even one amino acid longer.

The mode of peptide binding seen here may have features general to other class I molecules, as the electron density in the binding sites of both HLA-A2 (refs 3, 5, 7) and HLA-Aw68 (ref. 6) is consistent with that described here (see Fig. 14 of ref. 7). Both showed peptide density deep in both ends of the binding groove and much more exposed at a bend or kink near the middle of the site. Anchor residues, positions of conserved sequence in peptides that bind a given MHC molecule, have been deduced from the sequencing of pools of peptides eluted from HLA-A2, H2-K b , H2-K a , and -D a , and are found at positions P2 and P3 and at the C-terminal position, like the structure presented here. An anchor at P5 is observed in peptides eluted from H2-D b and H2-K c (ref. 26; G. M. Van Bluck and S. C. Nathenson, manuscript submitted). In HLA-B27, P5 is relatively unconstrained by its contact with the HLA molecule. In other alleles, however, amino-acid differences in that region or a slightly altered conformation of the bound peptide could easily constrain the sequence at P5.

Our observations offer a general answer to the question of how one HLA molecule can form tight complexes with any one of a large number of peptides with diverse sequences. First, the MHC molecule determines the conformation of the bound peptide, and accommodates its side chains in binding-sites pockets. Second, the peptide is only specifically recognized at four of its nine side chains, with further interactions involving main chain atoms at the two ends of the peptide and the charged termini. Third, some of the HLA pockets that recognize side chains can accommodate some diversity in peptide sequence. These strategies presumably evolved to increase the probability that at least one peptide expressed by an invading genome will bind a class I histocompatibility antigen and be presented to the immune system. For class II histocompatibility antigens, where the stimulus may be only a single protein antigen, peptide binding may be expected to involve more conserved contacts, fewer side-chain specificity pockets or pockets with an even greater capacity for accommodating diversity.

Received 22 July; accepted 14 August 1991.


Acknowledgements: We thank D. Grant (Mit Cell Culture Facility) for H. A. Svecson and M. Fray for large-scale culturing of L.2 cells, and P. Kirsanuso and A. A. Hay for technical assistance. This work was supported by the NSF (D.R.M.), the NIH (U.S.S. and D.R.M.), the Medical Foundation/Clinton A. King Trust Fund (C.G.), and the Howard Hughes Medical Institute (C.G.).