Pillars Article: Structure of the human class I histocompatibility antigen, HLA-A2.

P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger and D. C. Wiley

J Immunol 2005; 174:6-12; http://www.jimmunol.org/content/174/1/6.citation
Structure of the human class I histocompatibility antigen, HLA-A2

P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger & D. C. Wiley

HLA (human leukocyte antigen) molecules are polymorphic membrane glycoproteins found on the surface of nearly all cells. Multiple genetic loci within the major histocompatibility complex (MHC) encode these proteins, and one individual simultaneously expresses several polymorphic forms from a large pool of alleles in the population. HLA molecules (also known as class I histocompatibility antigens) are the targets of antibodies and cytotoxic T lymphocytes (CTL) during rejection of foreign transplants. They are also recognized by T cells together with viral antigens on infected cell surfaces, a phenomenon known as MHC restricted recognition. In contrast to antibodies, that can bind to free virus or soluble antigens, T-cell receptors only recognize foreign antigens that are associated with a particular HLA molecule. Because one individual expresses a limited set of different HLA molecules, a central question has been how these few HLA molecules can interact with so many foreign antigens. Limitations in the ability of a particular HLA molecule to associate with all antigens may explain the linkage of histocompatibility antigens to variations in susceptibility to human diseases, and the immune system's responsiveness to particular antigens.

Recent work has shown that virus-specific CTL will lyse an uninfected target cell of appropriate class I specificity to which peptide fragments of a viral protein have been added. T-helper cells had previously been shown to recognize fragments of protein antigens in association with class II histocompatibility
molecules. Short synthetic peptides (10–20 residues long) have also been shown to bind to purified class II proteins at what appears to be a single binding site. By analogy, it is likely that the homologous class I molecules bind antigenic peptides, and that the HLA-peptide complex is recognized by T-cell receptors on CTL.

Class I molecules (HLA-A, B, C in humans, H–2K, D, L in mice) are composed of two polypeptide chains: a heavy chain (relative molecular mass 44,000; M, 44K) which spans the membrane bilayer, and the non-covariently attached light chain, β2-microglobulin (β2-m, 12K). The extracellular portion of the heavy chain is divided into three domains, α1, α2, and α3 (ref. 15), each ~90 amino acids long and encoded on separate exons. The α1 domain and β2m are relatively conserved and show amino-acid sequence homology to immunoglobulin constant or variable domains, but have been reported to show weak sequence homology to each other. We have crystallized a soluble fragment of HLA-A2 (ref. 22), composed of α1, α2, and β2m, after removing the transmembrane anchor of the heavy chain by papain digestion. The structure determination of HLA-A2 was undertaken to gain an understanding of where the polymorphic residues are located on the structure, how and where antibodies and T-cell receptors recognize the molecule, and how the molecule interacts with foreign antigens. Here we report a description of HLA-A2 from a 3.5 Å X-ray crystallographic structure determination. The membrane-proximal α1 and β2m domains have tertiary structures resembling antibody domains, but are paired by a novel interaction not previously seen in immunoglobulin structures. The α2 and α3 domains are nearly identical to each other in structure, and are not similar to immunoglobulin constant or variable domains. The domains α1 and α2 form a platform composed of a single β-pleated sheet topped by α-helices with a long groove between the helices. Electron density which is not a part of the HLA molecule is observed in this site between the helices, presumably representing an unknown bound antigen. The location of polymorphic residues, antibody binding sites, and mutations affecting recognition by CTL are analysed in the accompanying article, which defines the binding site for processed foreign antigens.

### Structure determination

Soluble HLA-A2 was purified after papain digestion of plasma membranes from the homozygous human lymphoblastoid cell line JY. Papain cleaves the HLA heavy chain at residue 271, thirteen residues from the transmembrane region, yielding a molecule composed of α1, α2, α3, and β2m. When purified, 3.4-mg protein is obtained from 200 litres of human tissue culture cells.

The crystallization and a proposed packing model for the two crystal forms of HLA-A2 and HLA-A28 have been described previously. HL-A2 crystallizes in two space groups: monoclinic P21 and orthorhombic P212121. Both crystal forms grow as very thin (20 μm) plates ~0.5 mm square. Occasional 100 μm thick orthorhombic crystals have been grown. Data from the 20 μm thick monoclinic crystals were collected using synchrotron X-ray sources. Part of the data from the orthorhombic crystals were collected using a Xeonite area detector. Data collection and phasing statistics are presented in Table 1. A monoclonic electron density map was calculated to 3.5 Å resolution using isomorphous differences and anomalous scat-
Fig. 2 Schematic representation of the structure of HLA-A2. The β-strands are shown as thick arrows in the amino to carboxy direction, α-helices are represented as helical ribbons. Connecting loops are depicted as thin lines. Disulphide bonds are indicated as two connected spheres. a (Facing page), schematic representation of the four domains of HLA-A2. The molecule is shown with the membrane proximal immunoglobulin-like domains (α1, β2m) at the bottom, and the polymorphic α2 and α3 domains at the top. The indicated C-terminus of α1 is the papain cleavage site. In the membrane-bound molecule, another 13 amino acids extend past the cleavage site toward the membrane, which we assume to be horizontal at the bottom of this picture. This orientation places the helices of the α1 and α2 domains, and the antigen recognition site located between them, on the top surface of the molecule. The domains α1 and α2 form a platform with a single eight-stranded β-pleated sheet (seen edge on), covered by α-helices. b, Schematic representation of the top surface of HLA-A2. The α1 and α2 domains are shown as viewed from the top of the molecule, showing the surface that is presumably contacted by a T-cell receptor (90° from Fig. 2e). The N terminus of α1 is indicated. Each domain consists of four antiparallel β-strands followed by a long helical region, and the domains pair to form a single eight-stranded β-sheet topped by α-helices. The disulphide bond connects residue 164 in the α2 long helix to residue 101 in the first β-strand of α2. The residue analogous to Cys101 in α1 is Ser11; it is 8 Å from Ser71 (Cα to Cα) on the α1 long helix. c, α-carbon backbone of the α1 and α2 domains. The domains are shown in the same orientation as in b. Every fifth residue and the oligosaccharide attachment site (86) are labelled.

Fig. 3 (Right) Schematic representation of the secondary structure of HLA-A2. The four β-strands and helical regions in α1 occupy equivalent locations as the β-strands and α-helices in α2, and the two domains are folded into homologous tertiary structures (Fig. 4). In α3 and β2m, the seven β-strands fold into structures resembling immunoglobulin constant domains. The positions of cysteine residues involved in disulphide bonds, and the location of the glycosylation site at residue 86 are indicated.

Fig. 4 (Facing page) Comparison of the structures of the α1 and α2 domains. a, The α-carbon backbone of α1, N and C termini are labelled. The α-carbon backbone of α1, N and C termini are labelled.

Fig. 5 (Facing page) Location of the pseudo-symmetry axes in the HLA-A2 molecule. Domains α1 (red) and α2 (yellow) are related by an approximate dyad axis of symmetry (orange arrow). (The relationship is a rotation of ~177° followed by a 0.7 Å translation.) Domains α3 (blue) and β2m (purple) are related by a 146° rotation (blue arrow) followed by a 13-Å translation. The two axes do not intersect, and deviate from co-linearity by 25° in projection.

Fig. 6 (Facing page) Van der Waals surface representation (1) of the top of the HLA-A2 molecule (a) showing the deep groove identified as the antigen recognition site (2) and the electron density (b) found in this site in crystals of HLA-A2. The surface was generated with the program GRIDCTR written by Mark Handschumacher and Fred Richards. The molecule is shown from the top with the Ca backbone (pink) oriented as in Fig. 2b,c. a, Van der Waals surface of the α1 and α2 domains (blue) showing a deep groove running between the helical regions in the α1 and α2 domains. This groove has been identified as the recognition site for processed antigens (see text and ref. 26 for details). b, The extra electron density (red contours at 0.8σ) found in both crystal forms of HLA-A2 is shown superimposed on the Van der Waals surface of α1 and α2 (blue). The extra density represents a bound molecule or mixture of molecules that co-purified and co-crystallized with HLA-A2. Some regions of the extra density may be wide enough to accommodate a peptide in an α-helical conformation, but in the absence of the knowledge of its composition, or whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution of the electron density maps (3.5 Å).
tering from four heavy atom derivatives and modified by iterative solvent flattening. Due to the poor quality of the heavy atom derivatives, the map was not fully interpretable, although the molecular outline and the two immunoglobulin-like domains of α₁ and α₂ were evident. Subsequent electron density maps derived from many cycles of model building, phase combination and CORELS refinement permitted the complete structures of α₁ and β₂m to be determined, and ~80% of the main chain of α₁ and α₂ to be fit as segments of polyalanine. To complete tracing the polypeptide chain in α₁ and α₂, a single isomorphous replacement electron density map was calculated from 3.5 Å data from orthorhombic crystals. The relationship between the two crystal forms was established by a six-dimensional real-space search using the monoclinic partial model, and confirmed by the location of a common heavy atom derivative. The monoclinic and orthorhombic maps were averaged inside their molecular envelopes using the Bricogne algorithms. Solvent region densities were replaced by their average value, and new phases were calculated by Fourier transformation of the maps. Ten cycles of averaging in both crystal forms produced significant phase changes and R-factor improvements (Table 1), allowing the complete polypeptide chain to be traced unambiguously with the aid of the amino acid sequence (details will be published elsewhere; Saper et al., manuscript in preparation). Despite the limited resolution (3.5 Å), 80% of the side chains are fitted in well-defined density (Fig. 1). In the future, least squares refinement using 2.7 Å native data in either space group will allow a more precise positioning of side chains.

Description of the structure

HLA-A2 consists of two pairs of structurally similar domains: α₁ has the same tertiary fold as α₂, while β₂m has the same tertiary fold as β₂m. As oriented in Fig. 2a, the α₁ and β₂m domains are closest to the bottom of the figure. These domains are proximal to the cell membrane, which would probably be horizontal in this orientation (see Fig. 2 legend). Domains α₁ and α₂ form the top of the molecule with their helical sides facing away from the cell. A view of the 'top' of HLA-A2 (Fig. 26) shows the surface that is presumably recognized by T cells. The cross-section of the top of the molecule is about 50 Å × 40 Å, and the molecule is about 70 Å in length.

Structures of α₁ and β₂m

The α₁ and β₂m domains are both β-sandwich structures composed of two antiparallel β-sheet sheets, one with four β-strands and one with three β-strands, connected by a disulfide bond. This tertiary structure has been described for constant regions of immunoglobulin molecules (for review, see refs 36, 37) and was expected from the significant sequence homology between α₁, β₂m and constant regions. The structure of human β₂m as bound to the HLA-A2 heavy chain also appears similar to the structure of the free monomer of bovine β₂m. A least-squares comparison between the α-carbon coordinates of the β-strand region of α₁ with β₂m indicates an average r.m.s. positional difference of 1.4 Å. A similar comparison of α₁ and β₂m to an immunoglobulin constant region (Cμs from human Fc) yields an r.m.s. difference of 1.6 Å in each case. Similar comparisons among antibody constant domains give r.m.s. differences of 0.6-1.1 Å.

Structures of α₁ and α₂

The α₁ and α₂ domains each consist of an antiparallel β-sheet sheet spanned by a long α-helical region that is C-terminal to the four β-strands in the sheet (Figs 3, 4a, b). In the helical region of α₁, a short α-helix (50-55) precedes a longer curved α-helix (58-84). These helices are almost at right angles (~110°). In α₂, a short helix (138-148) precedes a longer helix (151-173) at an angle of ~130°. The long helix in α₁ is knitted at residue 162. An additional short helix in α₁ (177-181) connects α₁ to α₂. A disulfide bond in α₂ connects residue 101 in the N-terminal β-strand to the long helix at residue 164. In Fig. 4d, α₁ is superimposed on α₂ to show their structural similarity (overall r.m.s. difference 3.1 Å). Published secondary structure predictions can be compared with Fig. 3.

Interdomain contacts

The α₁ and β₂m domains contact each other in an interaction not found between pairs of constant domains in the known antibody structures (for example, Cμs and Cγ1 in an Fab, the C₁s dimer in Fc). Antibody constant domains are related by a nearly exact dyad (180°) symmetry axis, with their four-stranded β-sheets forming the contact interface. In HLA-A₂, although the immunoglobulin-like domains contact each other with their four-stranded sheets (Table 2), they are related by a 146° rotation followed by a 13 Å translation (Fig. 5).

The structurally similar α₁ and α₂ domains (Fig. 4a, b) are paired in the HLA molecule by an approximate dyad axis of symmetry, such that the four β-strands from each domain form a single antiparallel β-sheet with eight strands (Fig. 4c). This β-sheet is topped by the helical regions from each domain, which are separated by about 18 Å (centre to centre). The interdomain contact (Fig. 4c) includes hydrogen bonding between the N-terminal β-strands of each domain at the centre of the eight-stranded sheet, and contacts between the C-terminal ends of the helical regions and β-sheet residues in the other domain (Fig. 4c, Table 2). The presence of approximate dyad symmetry has previously been observed between homologous domains in a single polypeptide chain, and is termed 'intramolecular dimerization'. The particular dimeric interaction seen between α₁ and α₂ involves the creation of a single β-sheet from two domains, has also been observed in a number of intermolecular interactions (concanavalin A2, aspartate carbamoyltransferase B chains33, alcohol dehydrogenase35), indicating that the α₁-α₂ domain interface could be preserved in an intermolecular dimer such as that in class II heterodimeric antigens (see below).

Pseudo-symmetry axes, domain interfaces

The screw axis relating the two immunoglobulin-like domains of HLA-A₂ (146° rotation and 13 Å translation) and the approximate dyad axis relating α₁ and α₂ are neither co-linear nor intersecting. In projection, the two axes deviate by 25° from co-linearity (Fig. 5). The presence of a local dyad axis had been anticipated by earlier crystallographic work, but the relationship between the immunoglobulin-like domains in HLA was unexpected. As this arrangement of domains is found in both crystal forms of HLA-A₂, it is not likely to result from crystal-packing forces. Table 2 lists the residues in each domain that are within 4 Å of another domain. The residues in contact across the α₁-β₂m interface are in the four-stranded β-sheets, but the contact pairs are not symmetrical as would be expected if the domains were related by a dyad. The β₂m domain also interacts with the central β-strands and loops of the α₁-α₂ β-sheet. The interaction of β₂m with all three domains of the heavy chain of HLA was suggested by immunological and circular dichroism data, and results in a molecule that is shaped quite differently from an Fab28, 37.

Carbohydrate

Both human and mouse class I histocompatibility antigens have an N-linked, complex oligosaccharide attached at Asn 86 (ref. 46). This residue is in a loop connecting α₁ to α₂. Carbohydrate was visible in the electron density map extending away from the protein structure, but analysis of its structure awaits high resolution refinement. Several studies have demonstrated that glycosylation at Asn 86 is not necessary for the association of β₂m with the HLA heavy chain or cell-surface expression. In murine class I antigens there is a second glycosylation site at Asn 176, and H-2L4, H-2D9 and H-2K3 molecules have a third at Asn 256 (ref. 49). Residue 176 is located in α₂ at a
position structurally homologous to Asn 66 in α1. The presence of carbohydrate at position 176 would increase the approximate symmetric nature of the α1–α2 ‘intramolecular dimer’. Residue 256 is located on the loop between strands 5 and 6 in α1 (Fig. 3) and is accessible to solvent in HLA-A2.

**Antigen recognition site**

A deep groove ~25 Å long and 10 Å wide runs between the two long α-helices of the α1 and α2 domains (Fig. 6a). The sides of the groove are formed by side chains from these helices, and the bottom is formed by side chains from the central β-stands of the α1–α2 β-sheet. The groove is located on the top surface of the molecule (as oriented in Fig. 2a), and is therefore a likely candidate for the binding site for the foreign antigen that is recognized together with HLA by a T-cell receptor. The dimensions of the site are consistent with the expectation that class I molecules bind a processed antigen, probably a peptide. The site is lined with both polar and non-polar side chains. Some of these residues have been identified to be critical for T-cell recognition, as described more fully in an accompanying article.

In this site, we observe a large continuous region of electron density that is not accounted for by the polypeptide chain of the HLA molecule (Fig. 6b). This extra density is found in electron density maps calculated in both space groups. The density level is comparable to the density level of the protein, suggesting that most or all of the crystalline HLA molecules have a molecule(s) bound in the proposed antigen binding site. It seems likely that the extra density is the image of a peptide or mixture of peptides that are purified and co-crystallized with HLA-A2. Potential sources of such peptides are: (1) protein fragments produced during papain cleavage; (2) peptides from the fetal calf serum used in the growth of JY cells; (3) processed antigenic peptides from Epstein Barr virus used in the transformation of JY cells20; and/or (4) endogenous ‘self’ peptides created during normal recycling and degradation of cellular proteins. In the absence of knowledge of the composition of the bound molecule(s) and whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution (3.5 Å) of the electron density maps.

**Other histocompatibility antigens**

The structure of HLA-A2 can be used as a starting point for modelling other class I and class II histocompatibility antigens. Using the extensive sequence homology between murine and human class I molecules, it is possible to locate residues critical for murine CTL and serological epitopes on the HLA-A2 structure (see accompanying article)

Class II molecules have a domain structure similar to class I, but the four domains of the class II chains are arranged on two polypeptide chains of roughly equal size that span the membrane bilayer (reviewed in ref. 51). The membrane-proximal domains of each chain are immunoglobulin-like with homology to α1 and β1m. The N-terminal domains of the class II chains presumably contain the binding site for antigenic peptides, and one of these domains has weak sequence homology to class I α1 and α2 (refs. 52). Two features of the HLA-A2 molecule suggest that class II molecules could adopt a similar structure: (1) The ‘intramolecular dimerization’ between the α1 and α2 domains resulting in an edge-to-edge β-sheet contact is similar to intramolecular dimer contacts observed in multisubunit molecules. It is therefore possible that class II N-terminal domains can form an intermolecular dimer with a structure similar to the α1 and α2 region of HLA-A2, thereby creating a binding site for antigens similar to that in HLA-A2. Residues in the first half of the membrane-distal domains of class II molecules have recently been shown to be critical for chain association5, a result consistent with the proximity of the N-terminal β-strands of the class I and class II domains. Also shown in Fig. 4c (2) In the HLA-A2 structure, the N-terminus of β2m is 11 Å from the C-terminus of the α1 domain. By replacing the loop that connects α1 and α2 (residues 85–93) with an extended chain connecting the end of α1 to the N-terminus of β2m, a four-domain structure can be constructed with two domains per polypeptide chain that is similar to HLA-A2 (J. Brown et al., manuscript in preparation).

These observations indicate that the HLA-A2 structure may be a good first-order model for interpreting the locations of polymorphic and functional residues on many class I and class II histocompatibility antigens. We thank Anastasia Haykov for excellent technical assistance; Hans Bartunik and Klaus Bartels, Keith Moffat and Wilfried Schindl and Paul Phizackerley for help with synchrotron data collection, Mike Silver and Tom Garrett for help with the figures and our colleagues in the structural molecular biology group. We also thank Drs Dean Mann (NIH), D. Michael Strong and James Woody (U.S. Naval Research Unit) and Don Giard (MIT Cell Culture Facility) for provision of cells which made this work possible. B.S. thanks the Algerian Ministère de l’Enseignement et de la Recherche Scientifique for a postdoctoral leave. P.J.B. held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

*Note added in proof*: Alpha-carbon coordinates are being deposited in the Brookhaven Data Bank.