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

Reprinted from Nature, Vol. 329, No. 6139, pp. 506-512, 8 October 1987

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*§

* Department of Biochemistry and Molecular Biology, Harvard University, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
† Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA
‡ Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA (P.J.B.) and Max-Planck-Institute für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FDR (W.S.B.)
§ To whom correspondence should be addressed

The class I histocompatibility antigen from human cell membranes has two structural motifs: the membrane-proximal end of the glycoprotein contains two domains with immunoglobulin-folds that are paired in a novel manner, and the region distal from the membrane is a platform of eight antiparallel β-strands topped by α-helices. A large groove between the α-helices provides a binding site for processed foreign antigens. An unknown 'antigen' is found in this site in crystals of purified HLA-A2.

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[1,2]. They are also recognized by T cells together with viral antigens on infected cell surfaces, a phenomenon known as MHC restricted recognition[3]. 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[4], and the immune system's responsiveness to particular antigens[5].

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[6]. T-helper cells had previously been shown to recognize fragments of protein antigens in association with class II histocompatibility

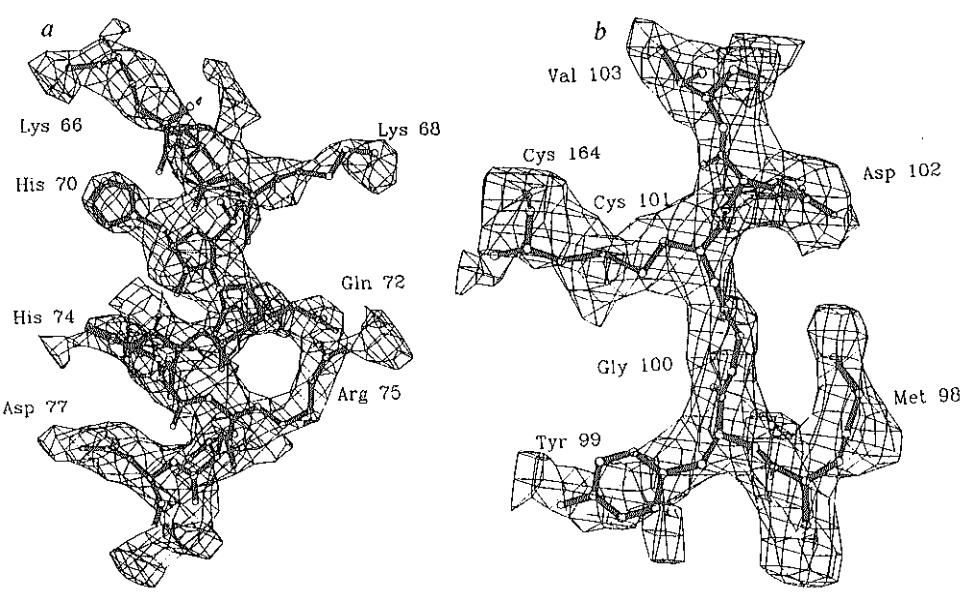

Fig. 1 Representative portions of the HLA-A2 model superimposed on the 3.5 Å averaged electron density map. Electron density is shown for residues 66–77 in the α_1 long helix (*a*) and for residues 98–103 in β-strand 1 of α_2 (*b*). Contours are drawn at 0.9 σ. Rotation and translation parameters were determined that relate the monoclinic and orthorhombic maps, and these maps were iteratively averaged[34]. The input $P2_12_12_1$ map was calculated with phases from a single derivative, and the input $P2_1$ map was calculated with phases from four derivatives combined with phases from a partial model of the structure (model did not include side chains in α_1 and α_2; see text for details). As a control against bias from the partial model included in the $P2_1$ map, the real space averaging was repeated starting with the model-free, uninterpretable monoclinic and orthorhombic maps, resulting in a similar and interpretable averaged map. It is possible that some side chains located on the surface of the molecule or near crystal contacts have different conformations in the two crystal forms. The final positioning of side chains awaits extension to higher resolution and refinement. (Current *R* factor between $P2_1$ calculated and observed amplitudes is 26.2% with good geometry after 24 cycles of 6–3.2 Å refinement using the program CORELS[33].)

Table 1 Data collection, phasing and map averaging statistics

Mean figure of merit: Monoclinic – 0.60 for 5,759 reflections
(3.5 Å resolution) Orthorhombic – 0.40 for 4,549 reflections

Derivative	d_{lim} (Å)	Number of films (source)‖	Number of measurements	Number of reflections, (% complete)	R†	r.m.s. f_h/E‡ (15–3.5 Å)
$P2_1$ ($a = 60.4$ Å, $b = 80.4$ Å, $c = 56.5$ Å, $\beta = 120.4°$)						
Native	2.7	120 (DESY/CHESS)	80,891	13,513 (96%)	0.101	—
K_2PtCl_4	2.7	44 (DESY)	42,299	12,808 (89%)	0.095	1.78
HgI_2	2.7	40 (DESY)	37,101	12,696 (81%)	0.105	0.92
K_2OsCl_6	2.7	35 (DESY)	30,888	12,110 (78%)	0.091	0.52
PCMP*	2.8	54 (SSRL/CHESS)	34,295	10,247 (83%)	0.114	0.97
$P2_12_12_1$ ($a = 60.2$ Å, $b = 80.4$ Å, $c = 112.2$ Å)						
Native	2.8	24 (CHESS/XEN)	30,590	10,638 (70%)	0.087	—
K_2PtCl_4	3.5	– (XEN)	15,106	5,028 (70%)	0.094	1.45

Map averaging statistics at 3.5 Å resolution

	R§		r.m.s. phase difference¶ (degrees)	
	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Cycle 1	0.342	0.343	47	38
Cycle 5	0.247	0.255	68	78
Cycle 10	0.166	0.185	73	83

Film data were collected on 5° oscillation photographs ($P2_1$) or 2–3° oscillation photographs ($P2_12_12_1$) using synchrotron X-ray sources, and processed and scaled as described[28]. The Xentronics area detector data were recorded as 5′ oscillation frames, then merged into batches of 20 frames and scaled to the native orthorhombic film data. The monoclinic derivatives used in phasing were chosen after screening over 100 heavy atom compounds, and recording complete film data sets from 13 different heavy atom derivatives, 9 of which did not prove useful in phasing. The most useful monoclinic derivative (K_2PtCl_4) was the only dataset with an overall phasing power (r.m.s. f_H/E) greater than 1.0 to 3.5 Å resolution. The other derivative datasets had phasing powers greater than 1.0 only at lower resolution (~6.0 Å). The K_2PtCl_4 orthorhombic data were truncated at 3.5 Å resolution due to poor quality at higher resolution, and these data determine the resolution limit of the averaged electron density maps (Fig. 1*a*,*b*). Statistics are presented for 10 cycles of real-space averaging of the monoclinic and orthorhombic electron density maps.

* *p*-Chloromercuriphenol; † $R_I = \sum|I - \langle I\rangle|/\sum\langle I\rangle$; ‡ f_h, heavy atom structure factor; E is the residual lack of closure; ‖ DESY, Deutches Elektronen Synchrotron; CHESS, Cornell High Energy Synchrotron Source; SSRL, Stanford Synchrotron Radiation Laboratory; XEN, Xentronics Area Detector.

§ R, crystallographic R-factor between observed structure factors and calculated structure factors from the averaged map.

¶ r.m.s. phase difference between calculated phases and phases used in the starting map.

molecules[7]. Short synthetic peptides (10–20 residues long) have also been shown to bind to purified class II proteins[8,9] at what appears to be a single binding site[10]. 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[11].

Class I molecules (HLA-A, B, C in humans, H-2K, D, L in mice) are composed of two polypeptide chains[12]; a heavy chain (relative molecular mass 44,000; M_r 44K) which spans the membrane bilayer, and the non-covalently attached light chain, β_2-microglobulin (β_2m—12K)[13,14]. The extracellular portion of the heavy chain is divided into three domains, α_1, α_2 and α_3 (ref. 15), each ~90 amino acids long and encoded on separate exons[16]. The α_3 domain and β_2m are relatively conserved and show amino-acid sequence homology to immunoglobulin constant domains[15,17–20]. The polymorphic α_1 and α_2 domains show no significant sequence homology to immunoglobulin constant or variable domains, but have been reported to show weak sequence homology to each other[21]. We have crystallized a soluble fragment of HLA-A2 (ref. 22), composed of α_1, α_2, α_3 and β_2m, after removing the transmembrane anchor of the heavy chain by papain digestion[23,24]. (HLA-A2, formerly named Mac, was the first polymorphic human leukocyte antigen identified[25].)

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 α_3 and β_2m domains have tertiary structures resembling antibody domains, but are paired by a novel interaction not previously seen in immunoglobulin structures. The α_1 and α_2 domains are nearly identical to each other in structure, and are not similar to immunoglobulin constant or variable domains. The domains α_1 and α_2 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[26].

Structure determination

Soluble HLA-A2 was purified after papain digestion of plasma membranes from the homozygous human lymphoblastoid cell line JY[23,24]. Papain cleaves the HLA heavy chain at residue 271, thirteen residues from the transmembrane region[27], yielding a molecule composed of α_1, α_2, α_3 and β_2m. 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[22,28]. HLA-A2 crystallizes in two space groups: monoclinic $P2_1$ and orthorhombic $P2_12_12_1$. 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 Xentronics area detector[29,30]. Data collection and phasing statistics are presented in Table 1.

A monoclinic electron density map was calculated to 3.5 Å resolution using isomorphous differences and anomalous scat-

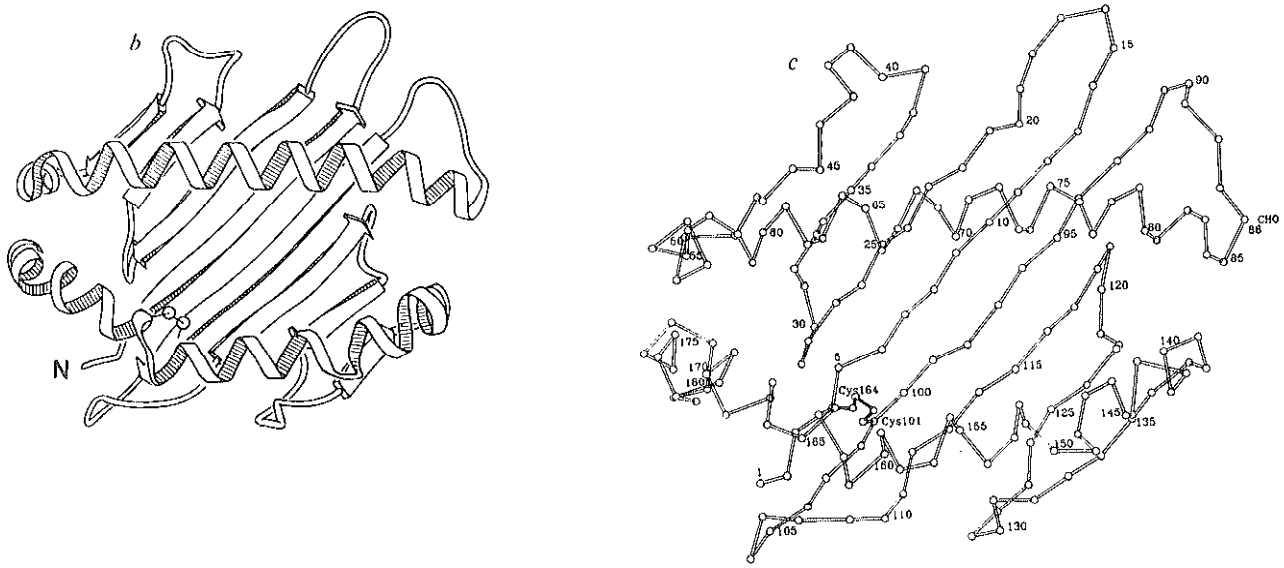

Fig. 2 Schematic representation of the structure of HLA-A2. The β-strands are shown as thick arrows in the amino to carboxy direction, α-helices are represented as helical ribbons. Connecting loops are depicted as thin lines. Disulphide bonds are indicated as two connected spheres. *a* (Facing page), schematic representation of the four domains of HLA-A2. The molecule is shown with the membrane proximal immunoglobulin-like domains (α_3, β_2m) at the bottom, and the polymorphic α_1 and α_2 domains at the top. The indicated C-terminus of α_3 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 α_1 and α_2 domains, and the antigen recognition site located between them, on the top surface of the molecule. The domains α_1 and α_2 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 α_1 and α_2 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. 2*a*). The N terminus of α_1 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 α_2 long helix to residue 101 in the first β-strand of α_2. The residue analogous to Cys 101 in α_1 is Ser 11; it is 8 Å from Ser 71 (C_α to C_α) on the α_2 long helix. *c*, α-carbon backbone of the α_1 and α_2 domains. The domains are shown in the same orientation as in *b*. Every fifth residue and the oligosaccharide attachment site (86) are labelled.

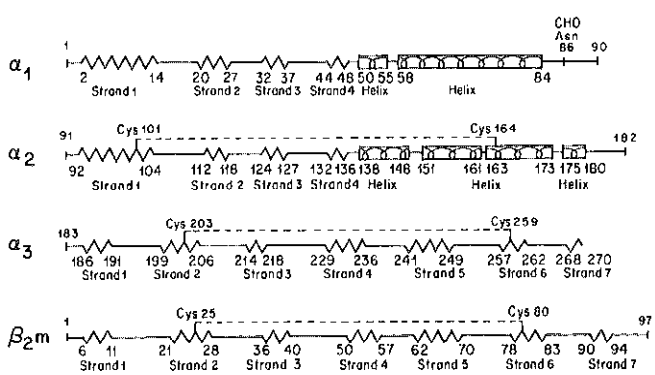

Fig. 3 (Right) Schematic representation of the secondary structure of HLA-A2. The four β-strands and helical regions in α_1 occupy equivalent locations as the β-strands and α-helices in α_2, and the two domains are folded into homologous tertiary structures (Fig. 4). In α_3 and β_2m, the seven β-strands 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 α_1 and α_2 domains. *a*, The α-carbon backbone of α_1. N and C termini are labelled. *b*, The α-carbon backbone of α_2. N and C termini are labelled. *c*, The α_1 (blue) and α_2 (red) domains as they are paired in the HLA-A2 structure. The N-terminal β-strand in each domain is depicted with main-chain atoms to emphasize the hydrogen bonds (yellow lines) between these strands that form a single, eight-stranded β-pleated sheet from the four-stranded sheets in each domain. The rest of each domain is shown as the C_α backbone. Yellow stars highlight the C_α atom of any residue with a main or side chain atom within 4 Å of an atom located in the partner domain (see Table 2). The N terminus of α_1 is labelled. *d*, The α-carbon backbones of α_1 (blue) and α_2 (red) superimposed to show structural similarity; α_1 and α_2 were aligned after a least-squares fit of the α-carbon coordinates of the β-strand and α-helical residues in each domain. The average r.m.s. positional difference is 3.1 Å.

Fig. 5 (Facing page) Location of the pseudo-symmetry axes in the HLA-A2 molecule. Domains α_1 (red) and α_2 (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 α_3 (blue) and β_2m (purple) are related by a 146° rotation (blue arrow) followed by a 13-Å translation. The two axes do not intersect, and deviate from colinearity by 25° (in projection).

Fig. 6 (Facing page) Van der Waals surface representation[54] of the top of the HLA-A2 molecule (*a*) showing the deep groove identified as the antigen recognition site[26] 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 Cα backbone (pink) oriented as in Fig. 2*b*,*c*. *a*, Van der Waals surface of the α_1 and α_2 domains (blue) showing a deep groove running between the helical regions in the α_1 and α_2 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 α_1 and α_2 (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).

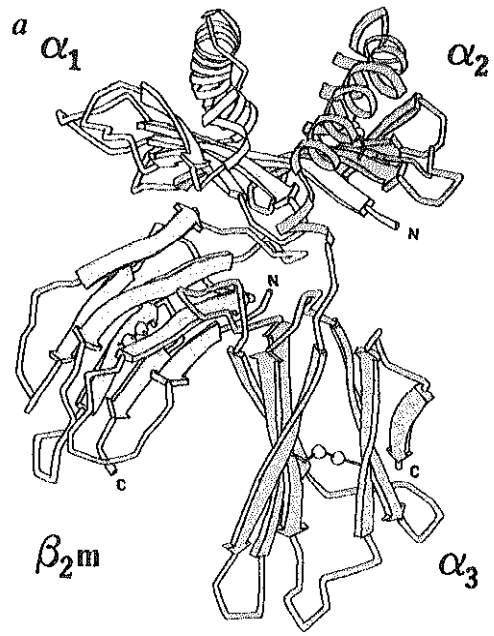

Fig. 2*a*

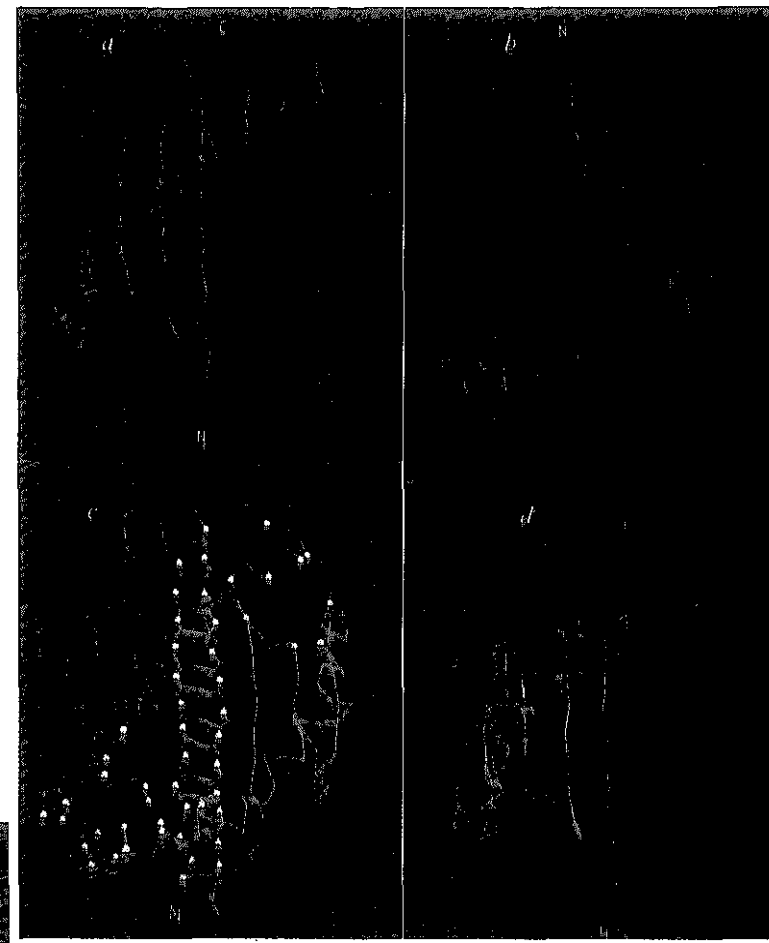

Fig. 4

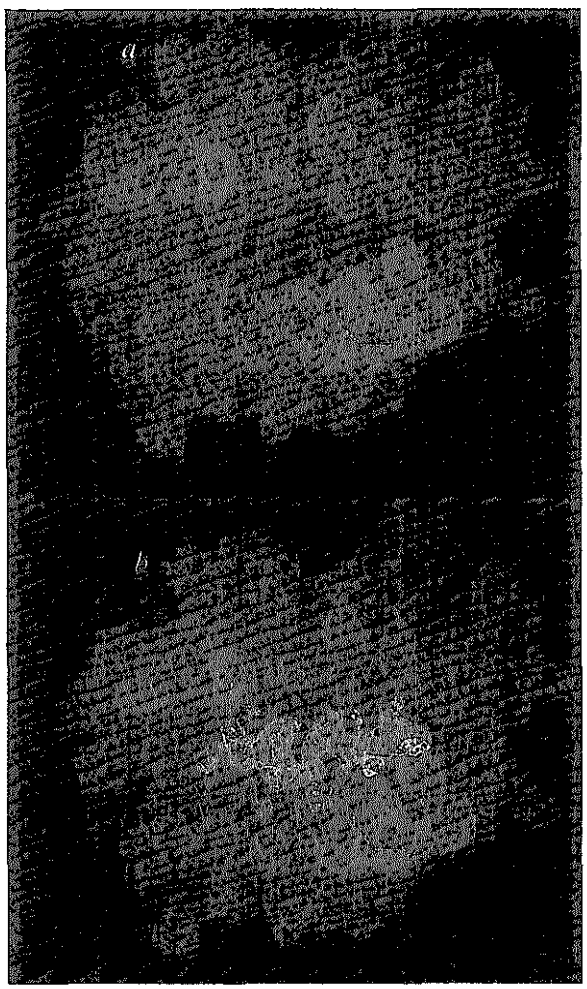

Fig. 6

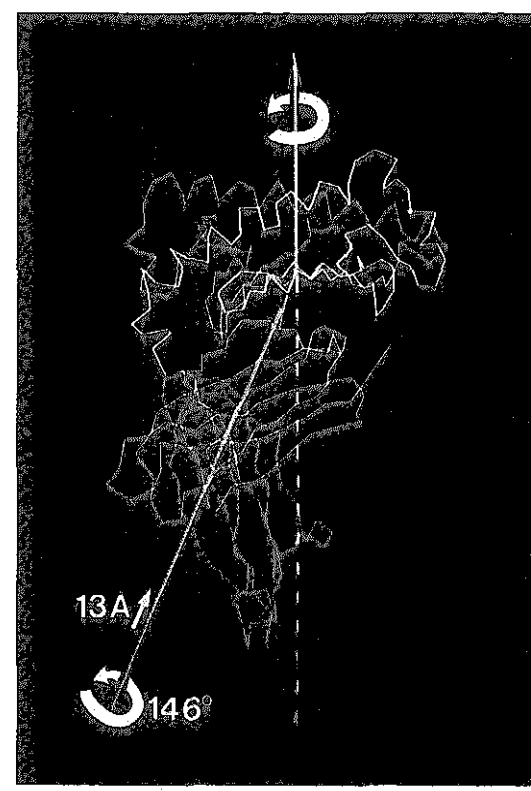

Fig. 5

tering from four heavy atom derivatives and modified by iterative solvent flattening[31]. 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 α_3 and β_2m were evident. Subsequent electron density maps derived from many cycles of model building, phase combination[32] and CORELS refinement[33] permitted the complete structures of α_3 and β_2m to be determined, and ~80% of the main chain of α_1 and α_2 to be fit as segments of polyalanine. To complete tracing the polypeptide chain in α_1 and α_2, 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[34]. 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[35] (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: α_1 has the same tertiary fold as α_2, while α_3 has the same tertiary fold as β_2m. As oriented in Fig. 2*a*, the α_3 and β_2m 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 α_1 and α_2 form the top of the molecule with their helical sides facing away from the cell. A view of the 'top' of HLA-A2 (Fig. 2*b*) 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 α_3 and β_2m

The α_3 and β_2m domains are both β-sandwich structures composed of two antiparallel β-pleated sheets, one with four β-strands and one with three β-strands, connected by a disulphide 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 α_3, β_2m and constant regions[15,17–20]. The structure of human β_2m as bound to the HLA-A2 heavy chain also appears similar to the structure of the free monomer of bovine β_2m[38]. A least-squares comparison between the α-carbon coordinates of the β-strand regions of α_3 with β_2m indicates an average r.m.s. positional difference of 1.4 Å. A similar comparison of α_3 and β_2m to an immunoglobulin constant region (C_{H3} 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[38].

Structures of α_1 and α_2

The α_1 and α_2 domains each consist of an antiparallel β-pleated sheet spanned by a long α-helical region that is C-terminal to the four β-strands in the sheet (Figs 3, 4*a*,*b*). In the helical region of α_1, a short α-helix (50–55) precedes a longer curved α-helix (58–84). These helices are almost at right angles (~110°). In α_2, a short helix (138–148) precedes a longer helix (151–173) at an angle of ~130°. The long helix in α_2 is kinked at residue 162. An additional short helix in α_2 (177–181) connects α_2 to α_3. A disulphide bond in α_2 connects residue 101 in the N-terminal β-strand to the long helix at residue 164. In Fig. 4*d*, α_1 is superimposed on α_2 to show their structural similarity (overall r.m.s. difference 3.1 Å). Published secondary structure predictions[39,40] can be compared with Fig. 3.

Interdomain contacts

The α_3 and β_2m domains contact each other in an interaction not found between pairs of constant domains in the known antibody structures (for example, C_{H1} and C_L in an Fab, the C_{H3} 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[36,37]. 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 α_1 and α_2 domains (Fig. 4*a*,*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. 4*c*). This β-sheet is topped by the helical regions from each domain, which are separated by about 18 Å (centre to centre). The interface between α_1 and α_2 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. 4*c*, 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'[41]. The particular dimeric interaction seen between α_1 and α_2, involving the creation of a single β-sheet from two domains, has also been observed in a number of intermolecular dimers (concanavalin A[42], aspartate carbamoyltransferase R chains[43], alcohol dehydrogenase[55]), indicating that the α_1-α_2 domain interface could be preserved in an intermolecular dimer such as that in class II histocompatibility 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 α_1 and α_2 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[22,28], 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 α_3-β_2m 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 β_2m domain also interacts with the central β-strands and loops of the α_1-α_2 β-sheet. The interaction of β_2m with all three domains of the heavy chain of HLA was suggested by immunological[44] and circular dichroism data[45], and results in a molecule that is shaped quite differently from an Fab[36,37].

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 α_1 to α_2. Carbohydrate is visible in the HLA 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 β_2m with the HLA heavy chain or cell-surface expression[47,48]. In murine class I antigens there is a second glycosylation site at Asn 176, and H-2L^d, H-2D^b and H-2K^d molecules have a third at Asn 256 (ref. 49). Residue 176 is located in α_2 at a

Table 2 Inter-domain contacts

α_1		β_2	α_3
1	104	8	231, 244
2	104	10	234, 242
3	103, 168, 172, 180	11	242
4	101, 102, 164, 168	12	238
5	99, 100, 101, 164, 168, 171	13	188, 206
6	98, 99	22	237
7	98, 99	26	233, 234
8	97	65	237
9	95, 96, 97, 99	97	204
10	95, 96		
11	94, 95	β_2	α_1
12	93	33	12
13	92	53	25, 33, 35
28	171, 179	55	27
29	179	56	8, 9, 10
32	171	62	10
51	174, 175	63	6, 27
54	174		
55	170		
59	167, 171	β_2	α_2
63	171	1	119, 120
70	99	31	94, 96, 119
74	95, 97	56	96, 97
78	95	60	96, 115, 116, 117, 121
81	118, 123	62	96
84	123, 139, 142, 143		
85	118		
87	118	α_3	α_1
		209	29, 30
		211	30
		234	27, 33
		α_3	α_2
		181	176, 177, 179
		182	180
		209	179

All residues (main chain or side chain) of one domain within 4 Å of the second domain are listed. This table is included as a guide for interpreting mutagenesis and exon shuffling experiments that exchange domains between different histocompatibility molecules. The contact list may require minor updating after crystallographic refinement. The residue numbers can be assigned to their secondary structural location on the HLA-A2 structure using Fig. 3. The α_1–α_2 contacts are shown on the structure of these domains in Fig. 4*c*.

position structurally homologous to Asn 86 in α_1. The presence of carbohydrate at position 176 would increase the approximate symmetric nature of the α_1–α_2 'intramolecular dimer'. Residue 256 is located on the loop between strands 5 and 6 in a_3 (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 α_1 and α_2 domains (Fig. 6*a*). 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-strands of the α_1–α_2 β-sheet. The groove is located on the top surface of the molecule (as oriented in Fig. 2*a*), 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[26].

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. 6*b*). 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 co-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 cells[50]; and/or (4) endogenous 'self' peptides created during normal recycling and degradation of cellular proteins. In the absence of the 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)[26].

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 α_3 and β_2m. 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 α_1 and α_2 (ref. 52). Two features of the HLA-A2 molecule suggest that class II molecules could adopt a similar structure: (1) The 'intramolecular dimerization' between the α_1 and α_2 domains resulting in an edge-to-edge β-sheet contact is similar to intermolecular 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 α_1 and α_2 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 association[53], a result consistent with the proximity of the N-terminal β-strands of the class I α_1 and α_2 domains shown in Fig. 4*c*. (2) In the HLA-A2 structure, the N-terminus of β_2m is 11 Å from the C-terminus of the α_1 domain. By replacing the loop that connects α_1 and α_2 (residues 85–93) with an extended chain connecting the end of α_1 to the N-terminus of β_2m, 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 Schildkamp 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 Ministeré de l'Enseignment 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.

Received 17 August; accepted 26 August, 1987.

1. Klein, J. in *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin, 1975).
2. Bach, F. H. & van Rood, J. J. *New Engl. J. Med.* **295,** 806–813 (1976).
3. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248,** 701–702 (1974).
4. Zinkernagel, R. M. *A. Rev. Microbiol.* **33,** 201–213 (1979).
5. McDevitt, H. O. & Tyan, M. L. *J. exp. Med.* **128,** 1–11 (1968).
6. Townsend, A. R. M. *et al. Cell* **44,** 959–968 (1986).
7. Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* **158,** 303–316 (1983).
8. Babbit, B., Allen, G., Matsueda, E., Haber, E. & Unanue, E. *Nature* **317,** 359–361 (1985).
9. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83,** 3968–2971 (1986).
10. Guillet, J. G., Laim-Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* **324,** 260–263 (1986).
11. Forman, J. *Adv. Immun.* (in the press).
12. Cresswell, P., Turner, M. J. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **70,** 1503–1507 (1973).
13. Grey, A. M. *et al. J. exp. Med.* **138,** 1608–1612 (1973).
14. Peterson, P. A., Rask, L. & Lindblom, J. B. *Proc. natn. Acad. Sci. U.S.A.* **71,** 35–39 (1974).
15. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. *Proc. natn. Acad. Sci. U.S.A.* **76,** 5839–5842 (1979).
16. Malissen, M., Malissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* **79,** 893–897 (1982).
17. Tragardh, L., Rask, L., Wiman, K., Fohlman, J. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **76,** 5839–5842 (1979).
18. Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **69,** 1697–1701 (1972).
19. Smithies, O. & Poulik, M. D. *Science* **175,** 187–189 (1972).
20. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. *Nature* **282,** 266–270 (1979).
21. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18,** 5711–5719 (1979).
22. Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *J. molec. Biol.* **186,** 205–210 (1985).
23. Turner, M. J. *et al. J. biol. Chem.* **250,** 4512–4519 (1975).
24. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. *J. biol. Chem.* **252,** 7555–7567 (1977).
25. Dausset, J. *Acta haematol.* **20,** 156 (1958).
26. Bjorkman, P. J. *et al. Nature* **329,** 512–518 (1987).
27. Springer, T. A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73,** 2481–2485 (1976).
28. Bjorkman, P. J. thesis, Harvard University (1984).
29. Durbin, R. *et al. Science* **232,** 1127–1132 (1986).
30. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Cryst.* **20,** 235–242 (1987).
31. Wang, B. C. *Meth. Enzym.* **115,** 90–111 (1985).
32. Rice, D. W. *Acta crystallogr.* **A37,** 491–500 (1981).
33. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **33,** 800–804 (1977).
34. Bricogne, G. *Acta crystallogr.* **A32,** 832–847 (1976).
35. Koller, B. H. & Orr, H. T. *J. Immun.* **134,** 2727–2733 (1985).
36. Marquart, M. & Deisenhofer, J. *Immun. Today* **3,** 160 (1982).
37. Amzel, L. M. & Poljak, R. J. *A. Rev. Btochem.* **48,** 961–997 (1979).
38. Becker, J. W. & Reeke, G. N. *Proc. natn. Acad. Sci. U.S.A.* **82,** 4225–4229 (1985).
39. Vega, M. A., Bragado, R., Ezquerra, A. & Lopez de Castro, J. A. *Biochemistry* **23,** 823–831 (1984).
40. Novotny, J. & Auffray, C. *Nucleic Acid Res.* **12,** 243–255 (1984).
41. McLachlan, A. D. in *Protein Folding* (ed. Jaenicke, R.) 79–99 (Elsevier, North Holland, 1980).
42. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250,** 1525–1545 (1975).
43. Krause, K. L., Volz, K. W. & Lipscomb, W. N. *J. molec. Biol.* **193,** 527–553 (1987).
44. Krangel, M. S., Biddison, W. E. & Strominger, J. L. *J. Immun.* **130,** 1856–1862 (1979).
45. Lancet, D., Parham, P. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3844–3848 (1979).
46. Nathenson, S. G. & Cullen, S. E. *Biochim. biophys. Acta* **344,** 1–25 (1974).
47. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *J. Immun.* **126,** 270–275 (1981).
48. Santos-Aguado, J., Biro, P. A., Fuhrmann, U., Strominger, J. L. & Barbosa, J. A. *Molec. cell. Biol.* **7,** 982–990 (1987).
49. Kimball, E. S. & Coligan, J. E. *Contemp. Top. molec. Immun.* **9,** 1–63 (1983).
50. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73,** 910–944 (1976).
51. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36,** 1–13 (1984).
52. Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79,** 3687–3691 (1982).
53. Sant, A. J., Braunstein, N. S. & Germain, R. N. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
54. Lee, B. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).
55. Eklund, H. *et al. J. Molec. Biol.* **102,** 27–59 (1976).

Printed in Great Britain by Turnergraphic Limited, Basingstoke, Hampshire