Molecular comparisons of the β_2 -microglobulin-binding site in Class I major-histocompatibility-complex α -chains and proteins of related sequences

V. Andrew TYSOE-CALNON,* Jane E. GRUNDY† and Stephen J. PERKINS*‡

*Department of Biochemistry and Chemistry and †Department of Virology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K.

β_2 -Microglobulin (β_2m) binds non-covalently to the α_1 , α_2 and α_3 domains of the α -chain of Class I major-histocompatibility-complex (MHC) molecules. On the basis of the crystal structures of human leucocyte antigens HLA-A2.1 and HLA-Aw68.1, we have used molecular-graphics analyses to define 44 contact points between 19 α -chain residues and 18 β_2m residues. In 88 other α -chain sequences from the HLA-A, HLA-B, HLA-C, HLA-D, HLA-E, HLA-F and HLA-G locus products in man and the H-2, Qa and Tla loci in mouse, 37 contact sites were conserved to 90 % or more, and in β_2m sequences from seven other species 40 % of contact sites were totally conserved. Four distinct regions form the contact points between the α -chain and β_2m , one on each of the α_1 and α_2 domains and two on the α_3 domain. We have further studied the α -chain sequences of three non-MHC molecules, human CD1 and rat Fc receptor (FcRn), known to bind to β_2m , and a third molecule, the putative product of the H301 (UL18) gene of human cytomegalovirus (CMV). CMV has been shown to bind β_2m , and it has been postulated that the H301-gene product, which has sequence similarity to Class I HLA, is the protein responsible. These sequences exhibited much lower residue conservation with the MHC-linked group, although the α_3 domain was the most highly conserved, and gaps and insertions were required for optimal alignments with the 90 α -chain sequences. Of the 44 β_2m - α -chain contacts defined for Class I HLA, 24 α -chain contact sites were conserved in CD1, 25 in FcRn and 17 in the H301-gene product. For CD1 and FcRn, the majority of the conserved β_2m contacts were found in the α_2 domain and the major contact region in the α_3 domain. Together with the use of secondary-structure predictions, it was concluded that the binding of β_2m in CD1 and FcRn was MHC-like at the α_3 domain, and probably also at the α_2 domain for FcRn, but non-MHC-like for the α_1 domain of both molecules and the α_2 domain of CD1. In the H301-gene product sequence, only the β_2m contacts with the main region of the α_3 domain were noticeably conserved. Structural evidence for β_2m binding in H301 in this α_3 contact region was, however, much weakened by a sequence insertion, the presence of proline residues, and marked deviations in the secondary-structure predictions. If the H301-gene product is responsible for the binding of β_2m to cytomegalovirus, this occurs in a non-MHC-like manner.

INTRODUCTION

Class I major-histocompatibility-complex (MHC) membrane proteins are heterodimeric structures expressed on the surface of most cells. These present small endogenously processed peptides in a binding groove for recognition by the CD8-positive T-lymphocyte (Long & Jacobson, 1989). They consist of an α -chain (M_r 45000), which contains three extracellular domains α_1 , α_2 and α_3 , in association with β_2 -microglobulin (β_2m) (M_r 11700) (Bjorkman *et al.*, 1987a; Kourilsky & Claverie, 1989). In man Class I molecules are encoded by three major loci HLA-A, HLA-B and HLA-C, and in the mouse the MHC Class I loci are H-2-K, H-2-D and H-2-L. The mouse also possesses Qa and Tla loci, which encode MHC Class I-like molecules similar to H-2 (Steinmetz & Hood, 1983). The HLA-E, HLA-F and HLA-G loci are believed to be the human counterparts of Qa and Tla (Geraghty *et al.*, 1987, 1990; Koller *et al.*, 1988; Shimizu *et al.*, 1988). A large number of human and mouse α -chains have been sequenced and the results show that the α -chains are highly polymorphic, with most of the variation occurring within the α_1 and α_2 domains. Protein crystallography on HLA-A2.1 and HLA-Aw68.1 (Bjorkman *et al.*, 1987a,b; Garrett *et al.*, 1989) showed that the peptide-binding groove lies in the α_1 and α_2 domains between two long α -helices positioned upon an eight-

stranded β -sheet. The structures of the α_3 domain and β_2m resemble that of the immunoglobulin constant domain (Becker & Reeke, 1985).

In HLA, β_2m forms non-covalent contacts with all three α -chain domains (Bjorkman *et al.*, 1987a). Work with human Daudi cells shows that β_2m binding is required for the surface expression of the α -chain (Owen *et al.*, 1980; Sege *et al.*, 1981; de Preval & Mach, 1983). It is believed that β_2m binding also affects the conformation of the α -chain (Lancet *et al.*, 1979; Maziarz *et al.*, 1988). In addition, β_2m does not appear to bind unless the correct peptide has first bound to the α -chain in order to facilitate the formation of a β_2m -binding site on it (Townsend *et al.*, 1989, 1990; Junggren *et al.*, 1990).

Three other proteins exhibit sequence similarity to Class I molecules. CD1 antigens are expressed on the surface of immature cortical thymocytes, and consist of a glycosylated heavy chain (M_r 43000–49000) non-covalently associated with β_2m (Martin *et al.*, 1986, 1987; Balk *et al.*, 1989). Only the α_3 domain of CD1 is reported to show sequence similarity to Class I HLA (Calabi *et al.*, 1989). Five CD1 genes have been identified in man. The Fc receptor in the intestinal epithelial cells of neonatal rats (FcRn) mediates the uptake of IgG from milk, and contains two components of M_r 45000–53000 and 14000, the latter being β_2m (Simister & Mostov, 1989). FcRn is also found in mice, and

Abbreviations used: MHC, major histocompatibility complex; HLA, human leucocyte antigen (human MHC Class I molecules); H-2, mouse MHC Class I molecules; α -chain, heavy chain of Class I antigens, containing the α_1 , α_2 and α_3 domains; β_2m , β_2 -microglobulin; FcRn, rat Fc receptor; CMV, cytomegalovirus; H301, Class I-like gene found in CMV.

‡ To whom correspondence and requests for reprints should be addressed.

1991

studies with β_2 m-deficient mice showed that the intestinal cells do not bind IgG, indicating that association with β_2 m is required for functional expression (Zijlstra *et al.*, 1990). The DNA sequence of cytomegalovirus (CMV) showed that it encodes an H301-gene product with approx. 20 % similarity to Class I antigens, although no protein has yet been identified on the virus or in CMV-infected cells. Since β_2 m is known to bind to CMV (Grundt *et al.*, 1987a), it has been proposed that the H301-gene product is the β_2 m-binding protein of CMV (Beck & Barrell, 1988). When the H301 gene is expressed in BHK cells with the use of a vaccinia construct, it binds β_2 m (Browne *et al.*, 1990).

Previous studies have reported sequence alignments of various Class I molecules (Klein & Figueroa, 1986; Brown *et al.*, 1988; Parham *et al.*, 1988; Watts *et al.*, 1989). Others have analysed the sequence variations within the α_1 and α_2 domains, in particular those residues that make contact with the peptide and the T-cell receptor (Ajitkumar *et al.*, 1988; Maziarz *et al.*, 1988; Bouillot *et al.*, 1989; Grey *et al.*, 1989; Maryanski *et al.*, 1989; Mattson *et al.*, 1989). Limited secondary-structure predictions have been reported (Vega *et al.*, 1984; Novotny & Auffray, 1984). On the basis of the unrefined X-ray data of Bjorkman *et al.* (1987a) at 0.35 nm resolution, limited analyses of the α -chain- β_2 m contacts have been reported (Ennis *et al.*, 1988; Tysoe-Calnon *et al.*, 1990). Now, using the recently refined crystallographic data for HLA-A2.1 and HLA-Aw68.1 at 0.26 nm resolution (Garrett *et al.*, 1989), sequence comparisons based on the alignment of 90 α -chain and eight β_2 m sequences, and secondary-structure predictions, we have analysed the β_2 m- α -chain interactions for Class I HLA and H-2 molecules and their close analogues. This then permitted an assessment of whether or not the α -chain sequences for CD1, FcRn and H301-gene product were compatible with similar sites of binding of β_2 m.

MATERIALS AND METHODS

Sequence data

Ninety α -chain sequences in Class I and Class I-like molecules were used.

Human molecules were HLA-A1, HLA-B8, HLA-B14, HLA-B18, HLA-Bw41, HLA-Bw42, HLA-B44.2, HLA-Bw60, HLA-

Bw65, HLA-Cw2.2, HLA-A2.1, HLA-A2.2F, HLA-A2.2Y, HLA-A2.2, HLA-Aw68.2, HLA-A2.4a, HLA-A2.4b, HLA-A3.1, HLA-A3.2, HLA-A11, HLA-Aw24, HLA-A32, HLA-Aw68.1, HLA-Aw69, HLA-B7.1, HLA-B7.2, HLA-B13, HLA-Bw47, HLA-B27.1, HLA-B27.2, HLA-B27.3, HLA-B27.4, HLA-B27f, HLA-B40, HLA-B44.1, HLA-Bw58, HLA-Cw1, HLA-Cw2.1, HLA-Cw3 (Parham *et al.*, 1988), HLA-A11E, HLA-A11K (Mayer *et al.*, 1988), HLA-C4 (Davidson *et al.*, 1985), HLA-328 (Srivastava *et al.*, 1985), HLA-E, HLA-F and HLA-G (Koller *et al.*, 1988; Shimizu *et al.*, 1988; Geraghty *et al.*, 1990; Ellis *et al.*, 1990).

Mouse molecules were H-2D^b, H-2D^k, H-2D^d, H-2D^{dm1}, H-2D^q, H-2L^q, H-2A216, H-2K^b, H-2K^{bm1}, H-2K^{bm5}, H-2K^{bm6}, H-2K^{bm10}, H-2K^{bm11}, H-2K^{bm23}, H-2K^{bm3}, H-2K^{bm4}, H-2K^{bm8}, H-2K^q, H-2K^d, H-2K^{d.24}, H-2K^k, H-2K^{km1}, H-2K^{km2}, H-2K^q, H-2K^{w28}, H-2L^d, Q1^k, Q2^k, Q4^b, Q5^k, Q10^k, Q10^q, Q4^k, Q7^b, Q8^b, Q7^d, Q10^b, Tlaa-1, Tlaa-2, Tlaa-3, MT3b, MT13c (Watts *et al.*, 1989), H-2K^{k.d1} (Arnold *et al.*, 1984) and H-2K^{w29.7} (Morita *et al.*, 1985).

Seven further similar sequences correspond to non-MHC Class I-like proteins: human CD1a, CD1b, CD1c, CD1e (R2) and CD1d (R3) (Martin *et al.*, 1986; Balk *et al.*, 1989); FcRn of neonatal rat (Simister & Mostov, 1989); H301-gene product from human CMV (Beck & Barrell, 1988). β_2 m sequences were used as follows: rat (Mauxion & Kress, 1987), bovine (Groves & Greenberg, 1982), human (Suggs *et al.*, 1981), rabbit (Gates *et al.*, 1979), guinea pig (Wolfe & Cebra, 1980), mouse (Gates *et al.*, 1981; Parnes & Seidman, 1982) and chicken and turkey (Welinder *et al.*, 1991).

Sequence alignments

The α_1 , α_2 and α_3 domains in the set of 43 HLA and three HLA-E,F,G sequences (human) and that of the 28 H-2, 11 Qa and five Tla sequences (mouse) were aligned manually in two species-dependent groups. These alignments are available from the authors on request. The CD1, FcRn and H301-gene product sequences were aligned manually against an HLA consensus sequence derived from this (Figs. 1a-1c). The six β_2 m sequences were aligned as a group (Fig. 1d). The alignments maximized the similarity of conserved and conservatively replaced residues (Taylor, 1986), and minimized the occurrence of gaps in order to keep the secondary structure coherent and to facilitate averaging of the structure predictions (Perkins *et al.*, 1988). A conservative replacement was defined by residue families according to Venn diagrams (Taylor, 1986): tiny, glycine, alanine and serine; aliphatics, isoleucine, valine, leucine and methionine; aromatics, phenylalanine, histidine, tyrosine and tryptophan; positive, lysine, arginine and histidine; negative, aspartic acid and glutamic acid; hydroxy, serine and threonine; amides, glutamine and asparagine. The consensus HLA sequence (Fig. 1) is based on the most abundant residue at that position in the alignment of all 43 HLA sequences.

Contact sites

The contact sites between β_2 m and the α -chain were defined on the basis of a co-ordinate scan that listed all possible atomic contacts within 0.4 nm in the structures for HLA-A2.1 and HLA-Aw68.1 (Brookhaven Database codes 3HLA and 2HLA; Garrett *et al.*, 1989). Only those contacts between two atoms that formed a likely non-covalent interaction are listed in Table 1. Hydrogen bonds were defined by non-hydrogen distances of 0.27-0.32 nm, and hydrophobic and ionic interactions were defined by non-hydrogen distances of 0.32-0.4 nm. A Silicon Graphics Personal Iris 4D2OG Workstation was used with INSIGHT software (Biosym Technologies, San Diego, CA, U.S.A.) for molecular displays.

Fig. 1. Alignments of the α_1 , α_2 and α_3 domains in the CD1, FcRn and H301-gene product sequences with a HLA consensus sequence and of β_2 m sequences from six species

In the 90 MHC-linked sequences, the α_1 , α_2 and α_3 sequences are 90, 92 and 92 residues in length; the β_2 m sequences are 99 residues in length. (a) In the α_1 domain, totally conserved contact residues in CD1, FcRn and H301-gene product are marked with a double asterisk above the sequences. Contact residues that are 100 % conservatively replaced are marked with a single asterisk. Contact residue or residue family conservation greater than 75 % is marked with a plus sign. Residues denoted by a dash are identical with those in the HLA consensus sequence. Beneath the HLA consensus sequence, residue conservation in HLA is indicated as above by an asterisk or plus sign, and the positions of contact residues with β_2 m are denoted by c. Contact residues are stippled. The positions of the four β -strands and two α -helices are marked by b and a respectively, as defined in the crystal structures of HLA-Aw68.1 (top) and HLA-A2.1 (bottom). Residues are numbered with a vertical line at every tenth residue. Below the sequences, double points indicate residues that are totally conserved with the HLA consensus sequence, and single points those that are 100 % conservatively replaced. (b) and (c) The corresponding data for the α_2 and α_3 domains are presented in that order, to follow (a). (d) For the β_2 m sequences, residues denoted by a dash are identical with those in the human β_2 m sequences. The numbers above the sequences denote the α_1 , α_2 or α_3 domain with which the residue is in contact. Other details are as specified in (a).

Secondary-structure predictions

Averaged secondary-structure predictions were performed on each of these eight groups of sequences by the Robson method (Garner *et al.*, 1978; Perkins *et al.*, 1988), using the programs PREDRB, PREDRB3 and PREDRB4 in order to align the sequences and analyse their averaged predictions (in units of centinats). The α_1 and α_2 domains were analysed using an unbiased prediction since α -helices and β -sheets were both observed in the crystal structures. The α_3 and β_2 m domains were analysed using a β -biased prediction to follow the crystallographic results. The observed α -helices and β -strands are represented by a and b in Fig. 1, to follow the analysis by Garrett *et al.* (1989) of the crystal structures using the DSSP procedure of Kabsch & Sander (1983). The conventional fourth β -strand of the immunoglobulin fold of the α_3 and β_2 m domains contains an extra residue that disrupted the hydrogen-bonding pattern. This is denoted as b5 and b6 in Fig. 1. The other six conventional β -

strands are designated b1 to b3 and b7 to b9; however, b9 was not specified for the α_3 domain. In each of the α_3 and β_2 m domains, there is an extra β -strand between the third and fourth strands, denoted as b4 in Fig. 1.

RESULTS

Sequence alignments

The 90 MHC-linked α -chain sequences were aligned in groups corresponding to the 43 HLA, 28 H-2, 11 Qa, three HLA-E,F,G and five Tla sequences. The α_1 , α_2 and α_3 domains were of lengths 90, 92 and 92 residues respectively (Fig. 1). Most notably, alignments were readily achieved without gaps or insertions, apart from single three-residue insertions at residues 275–277 in the mouse H-2D^d, H-2D^{dm1}, Q7^b and Q7^d sequences. The eight β_2 m sequences could also be readily aligned without gaps or insertions, except for a gap at residue 47 in bovine β_2 m (Fig. 1d).

Table 1. Percentage conservation of contact sites in different α -chain and β_2 m sequences compared with HLA-A2.1 and HLA-Aw68.1

Residues are numbered to follow Fig. 1. The percentages correspond to percentage total conservation of the residue with that of HLA-A2.1 and HLA-Aw68.1 or of human β_2 m. The contacts are defined from examination of the crystal co-ordinate sets 2HLA (HLA-Aw68.1) and 3HLA (HLA-A2.1) in the Brookhaven Database (see the Materials and methods section). Abbreviations: sc, side chain; mc, main chain; H-bond, hydrogen bond. The superscripts ^a and ^b indicate contacts found only in HLA-Aw68.1 and HLA2.1 respectively.

β_2 m contact		β_2 m residue conservation (%)	α -Chain contact	Type of interaction	α -Chain residue conservation (% of total)								H301-gene product (1)
					HLA (43)	H-2 (28)	Qa (11)	HLA-EFG (3)	Tla (5)	CDI (5)	FcRn (1)		
α_1 domain													
Asp-53 CO ₂ ⁻	sc	100	Gln-32 NH ₂	sc	H-bond	73	19	70	100	80	100	100	0
^a Asp-53 CO ₂ ⁻	sc	100	Arg-35 NH	sc	H-bond	100	100	100	100	40	0	0	0
Asp-53 CO ₂ ⁻	sc	100	Arg-35 NH ₂ ⁺	sc	Ionic	100	100	100	100	40	0	0	0
^b Asp-53 CO ₂ ⁻	sc	100	Arg-48 NH	sc	H-bond	100	93	90	100	60	0	0	0
^a Asp-53 CO ₂ ⁻	sc	100	Arg-48 NH ₂ ⁺	sc	Ionic	100	93	90	100	60	0	0	0
^a Ser-55 OH	sc	100	Tyr-27 OH	sc	H-bond	100	100	100	100	100	0	0	0
Phe-56 CH	sc	100	Phe-8 CH	sc	Hydrophobic	100	100	100	100	100	20	0	0
(3 CH-CH contacts in both HLA-A2.1 and HLA-Aw68.1)													
Tyr-63 OH	sc	71	Tyr-27 OH	sc	H-bond	100	100	100	100	100	0	0	0
α_2 domain													
Ile-1 N-terminus	mc	100	Asp-119 C=O	mc	H-bond	100	96	73	100	100	0	0	100
^b Arg-3 NH ₂	sc	29	Gly-120 C=O	mc	H-bond	100	100	100	100	100	100	100	100
His-31 NH	sc	100	Gln-96 C=O	sc	H-bond	98	100	100	100	100	100	100	0
Trp-60 C=O	mc	100	Gln-96 NH ₂	sc	H-bond	98	100	100	100	100	100	100	0
Trp-60 NH	sc	100	Asp-122 CO ₂ ⁻	sc	H-bond	98	100	100	100	100	100	0	0
Phe-62	sc	100	Gln-96 NH ₂	sc	H-bond	98	100	100	100	100	100	100	0
(2 ring contacts in HLA-A2.1 and 4 contacts in HLA-Aw68.1)													
α_3 domain													
Lys-6 NH ₃ ⁺	sc	43	Glu-232 CO ₂ ⁻	sc	Ionic	100	100	91	100	100	0	0	0
Gln-8 NH ₂	sc	100	Glu-232 C=O	mc	H-bond	100	100	91	100	100	0	0	0
Gln-8 C=O	sc	100	Arg-234 NH ₂	sc	H-bond	100	100	100	100	100	0	0	0
Tyr-10 OH	sc	100	Pro-235 C=O	mc	H-bond	100	100	100	100	100	100	100	100
Ser-11 C=O	mc	100	Gln-242 NH ₂	sc	H-bond	98	100	100	100	100	0	0	0
^b Arg-12 NH ₂	sc	100	Ala-236 C=O	mc	H-bond	100	100	91	100	100	0	0	0
^a Arg-12 NH	sc	100	Gly-237 C=O	mc	H-bond	100	100	91	100	100	0	100	0
Arg-12 NH ₂	sc	100	Gly-237 C=O	mc	H-bond	100	100	91	100	100	0	100	0
^a Arg-12 NH	sc	100	Asp-238 CO ₂ ⁻	sc	H-bond	100	100	82	100	100	80	100	100
Arg-12 C=O	mc	100	Gln-242 NH ₂	sc	H-bond	98	100	100	100	100	0	0	0
Asn-24 NH ₂	sc	100	Ala-236 C=O	mc	H-bond	100	100	91	100	100	0	0	0
Tyr-26 CH	sc	100	Pro-235 CH	sc	Hydrophobic	100	100	100	100	100	100	100	100
(8 CH-CH contacts in HLA-A2.1, 3 contacts in HLA-Aw68.1)													
^b Asp-98 CO ₂ ⁻	sc	57	His-192 NH ₂ ⁺	sc	Ionic	100	100	100	100	100	0	0	0
Asp-98 C=O	mc	57	Arg-202 NH	sc	H-bond	100	100	91	100	100	0	0	0
Asp-98 C=O	mc	57	Arg-202 NH ₂	sc	H-bond	100	100	91	100	100	0	0	0
Met-99 C-terminus	mc	29	Arg-234 NH ₂ ⁺	sc	Ionic	100	100	100	100	100	0	0	0
(2 contacts in HLA-A2.1 and HLA-Aw68.1)													
Met-99 C-terminus	mc	29	Trp-244 NH	sc	H-bond	100	100	100	100	100	0	100	0

In contrast, in the non-MHC-linked sequences, the alignments with the HLA consensus sequence in Figs. 1(a)–1(c) required a total of six gaps and five insertions in the five CD1 sequences, six gaps and five insertions in the FcRn sequence, and six gaps and nine insertions in the H301-gene product sequence.

Conservation of residues

α_1 domain. The 43 HLA sequences demonstrated 54% total residue conservation and 10% conservative replacement (Fig. 1a). The greatest variation between the sequences lies within the a2 α -helix, and between the b3 and b4 β -strands, but not within the β -strands. Comparable results were found for the HLA-E,F,G, H-2, Qa and Tla sequences. Much smaller conservations were, however, observed for the CD1, FcRn and H301 sequences. There was only 1% total conservation and 13% conservative replacement of residues between CD1 and the HLA consensus sequence (see Martin *et al.*, 1986; Calabi *et al.*, 1989). The FcRn sequence demonstrated 26% total conservation and 18% conservative replacement with the HLA consensus, and the H301-gene product sequence was 30% totally conserved and 11% conservatively replaced. The areas of variability were spread evenly over the domain in the CD1, FcRn and H301-gene product sequences.

α_2 domain. In the 43 HLA sequences, 58% of the residues were totally conserved and 10% were conservatively replaced (Fig. 1b). Similar percentages were found for the HLA-E,F,G, H-2, Qa and Tla sequences. More areas of variability were found in the α_2 domain than in the α_1 domain, principally in the four α -helices a1 to a4, and also in the b1 β -strand, and between the b1 and b2 β -strands. Residues between the b2 β -strand and the a1 α -helix constituted the largest conserved region in all three groups. Again much smaller conservation was found in the CD1, FcRn and H301-gene product sequences when compared with the HLA consensus sequence, between 11% and 26% of residues being totally conserved and between 15% and 18% conservatively replaced.

α_3 domain. Residue conservation was greater in the α_3 domain than in the α_1 and α_2 domains, and there were virtually no regions of particular sequence variability. The 43 HLA sequences showed that 78% of residues were totally conserved and 7% were conservatively replaced (Fig. 1c), and similar percentages were found for the HLA-E,F,G, H-2, Qa and Tla sequences. Despite this improved conservation, a smaller amount of conservation was again seen in the CD1, FcRn and H301-gene product sequences. These showed 23–36% total conservation and 16–18% conservative replacement.

β_2 m. Within this group, the sequences showed 35% total conservation and 16% conservative replacement (Fig. 1d). Generally most of the sequence variation was found between the β -strands.

Definition of β_2 m contact sites with the α -chain

Our analysis of the Brookhaven HLA-A2.1 and HLA-Aw68.1 co-ordinates at 0.26 nm resolution (Garrett *et al.*, 1989) in Table 1 shows that 19 residues on the α -chain make 44 contacts with 18 residues on human β_2 m. Nine instances of different contacts found between the two co-ordinate sets were considered not to be significant (Table 1), since these reflected minor conformational re-arrangements. Fig. 2 shows the global location of the β -strands and the residues involved in forming the contact sites between the α_1 and α_2 domains of HLA-A2.1 and β_2 m (Fig. 2a), and likewise between the α_3 domain and β_2 m (Fig. 2b).

The principal interactions between the α -chain and β_2 m are summarized in Figs. 3(a)–3(e). Between the α_1 domain and β_2 m, several side chains pointed into a central pocket, where Asp-53 (β_2 m) formed two ionic and two hydrogen bonds with Gln-32, Arg-35 and Arg-48 on the α_1 domain (Fig. 3a). Also, Ser-55 and Tyr-63 (β_2 m) made well-defined hydrogen bonds with the hydroxy group of Tyr-27 (α_1) in HLA-Aw68.1 (Fig. 3b), although in HLA-A2.1 only the hydrogen bond involving Tyr-63 was observed. Between β_2 m and the α_2 domain in a buried contact

Fig. 2. General views of the β_2 m- α -chain contacts

Only the fragments of secondary structures involved in the contact regions are shown. α -Carbon positions of 19 contact residues in the α -chain are marked by large $\bullet$ or $\circ$ symbols, and the 18 contact residues in β_2 m are marked by large $\blacktriangle$ or $\triangle$ symbols. All other α -carbon atoms are marked by small $\bullet$ symbols. The α -chain is highlighted relative to β_2 m. (a) The b1–b4 strands of the α_1 and α_2 domains with five and four contact residues respectively are shown above two loops in β_2 m (residues 30–32 and 52–64) and the N-terminus of β_2 m (residues 1–3), on which are marked the nine β_2 m contact residues ($\bullet$ and $\blacktriangle$, α_1 -domain contacts; $\circ$ and $\triangle$, α_2 -domain contacts). (b) The b1–b2–b7–b5/b6 β -sheet of the α_3 domain with ten contact residues is shown above the b1–b2–b7–b5/b6 β -sheet and the C-terminus of β_2 m with nine contact residues ($\bullet$ and $\blacktriangle$, contacts between the four-stranded β -sheets of α_3 domain and β_2 m; $\circ$ and $\triangle$, contacts between the four-stranded β -sheets of α_3 domain and the C-terminus of β_2 m).

Fig. 3. Selected molecular views of the principal contacts between the α -chain and β_2m (Table 1)

All the non-hydrogen atoms are shown. Ionic or hydrogen bonds or hydrophobic interactions are denoted by broken lines. Except in (d), the distance between the non-hydrogen atoms (N, C or O) is reported in nm. (a) Interaction of Asp-53 (β_2m) with Gln-32, Arg-35 and Arg-48 (α_1 domain). (b) Interaction of Ser-55, Phe-56 and Tyr-63 (β_2m) with Phe-8 and Tyr-27 (α_1 domain). (c) Interaction of His-31, Trp-60 and Phe-62 (β_2m) with Gln-96 and Asp-122 (α_2 domain). (d) Interaction of Lys-6, Gln-8, Tyr-10, Ser-11, Arg-12, Asn-24 and Tyr-26 (β_2m) with Glu-232, Arg-234, Pro-235, Ala-236, Gly-237, Asp-238 and Gln-242 (α_1 domain), as shown in Fig. 2(b). (e) Interaction of Asp-98 and Met-99 (β_2m) with His-192, Arg-202, Arg-234 and Trp-244 (α_3 domain).

region, well-defined hydrogen bonds connecting His-31, Trp-60 and Phe-62 on β_2 m and Gln-96 on the α_2 domain formed strategic contacts in both HLA-Aw68.1 and HLA-A2.1 (Fig. 3c).

The 25 contacts formed between the α_3 domain and β_2 m could be subdivided into two major patches (Fig. 2b), an upper one situated close to the α_1 and α_2 domains (Fig. 3d), and a lower one at the other end of the α_3 domain involving the C-terminus of β_2 m (Fig. 3e). The upper patch was constructed from a large number of closely spaced interactive contacts between these domains. Six of these contacts involved the α -chain main-chain carbonyl groups of Gln-232, Pro-235, Ala-236 and Gly-237, which formed hydrogen bonds with the side chains of Gln-8, Tyr-10, Arg-12 and Asn-24 in β_2 m. In HLA-Aw68.1 there were three hydrophobic interactions between the aromatic rings of Tyr-26 and Pro-235 compared with eight in HLA-A2.1. The lower patch involved the two C-terminal residues Asp-98 and Met-99 of β_2 m and His-192, Arg-202, Arg-234 and Trp-244 on the α_3 domain.

There were fewer contacts between β_2 m and each of the α_1 and α_2 domains than with the α_3 domain. The α_1 and α_2 domains formed two ionic, three hydrophobic and 14 hydrogen-bond interactions with β_2 m, compared with four ionic, eight hydrophobic and 13 hydrogen-bond interactions between the α_3 domain and β_2 m (Table 1).

Contacts in MHC-linked molecules. Since no gaps or insertions were present in these sequence alignments, we have looked to see whether the α -chain- β_2 m contact residues in HLA-Aw68.1 and HLA-A2.1 are conserved in the same positions in all 90 α -chain and eight β_2 m sequences (Table 1).

In the α_1 domain, five residues made ten contacts with four residues in human β_2 m (Table 1 and Fig. 2a). Of these ten contacts, nine were totally conserved among the other 41 HLA sequences, seven contacts were totally conserved in each of the H-2 and Qa groups, all ten were conserved in the HLA-E,F,G group, but only five were conserved in the Tla group. In the Tla group, the lowest conservation (40%) corresponded to the contact with Arg-35 (Fig. 3a). The least conserved contact was that between Asp-53 on β_2 m and Gln-32 on the α -chain (Fig. 3a), especially for the H-2 sequences.

In the α_2 domain, four residues made nine contacts with five residues on human β_2 m (Table 1 and Fig. 2a). All nine of these contacts were conserved to at least 98% in the HLA sequences, and to at least 96% in the H-2 sequences. Eight contacts were totally conserved in the Qa sequences, and all nine were totally conserved in the HLA-E,F,G and Tla sequences. Asp-119 was the least conserved contact residue in the α_2 domain, and was conservatively replaced with glutamic acid in several H-2 and Qa sequences.

In the α_3 domain, ten residues made 25 contacts with nine residues on human β_2 m (Table 1 and Fig. 2b). All 25 contact points were conserved to more than 98% for the 43 HLA sequences, all 25 were totally conserved in the H-2, HLA-E,F,G and Tla sequences, and 17 contacts were totally conserved in the Qa sequences (Table 1). The ionic bonds formed between Asp-98 and Asp-99 (C-terminus of β_2 m), and His-192 and Arg-234 (α_3 domain) were highly conserved.

In human β_2 m, 18 residues made 44 contacts with 19 residues on the α -chain (Table 1). Comparisons with the other seven β_2 m sequences showed that 35 of the 44 contacts were totally conserved. The two residues that demonstrated least conservation were Arg-3 (29%) and Met-99 (29%) (Fig. 3e).

Contacts in non-MHC-linked molecules. There was far less conservation at the 44 contact sites defined for HLA-A2.1 and HLA-Aw68.1 in the CD1, FcRn and H301-gene product sequences, with the exception of the α_2 domains of CD1 and FcRn (Table 1). Of these sites 24 were totally conserved in CD1,

25 in FcRn and 17 in H301-gene product (including contacts involving main-chain carbonyl groups only).

In the α_1 domain, only one of the ten contact sites was totally conserved in the CD1 and FcRn sequences, and in H301-gene product none of the ten sites was conserved. In the five CD1 and the FcRn sequences Tyr-27 was conservatively replaced with Trp-27 (Fig. 3b). In the FcRn sequence another conservative replacement of Phe-8 with His-8 was found (Figs. 1a and 3b).

In the α_2 domain, eight of the nine contact sites were totally conserved in CD1, and seven in FcRn. The non-conserved contact site in these sequences was the main-chain carbonyl contact at Asp-119. The aspartic acid residue at position 122 was conservatively replaced with glutamic acid in the FcRn sequence. In H301-gene product only two of the nine α_2 -domain contact sites were conserved, at positions 119 and 120; however, both these contacts involve main-chain carbonyl groups.

In the α_3 domain, nine of the 25 contact sites were totally conserved in CD1, 13 in FcRn and ten in H301-gene product. If contacts involving main-chain carbonyl groups were counted as conserved, these numbers were increased to 14 (CD1), 16 (FcRn) and 15 (H301-gene product). Of the conserved contact sites, only one in the FcRn sequence involved the patch at the lower end of the α_3 domain (Fig. 3e). In each group eight of the 25 contact sites were accounted for by the hydrophobic interactions involving Tyr-26 (β_2 m) and Pro-235 (α_3 domain). The two residues that demonstrated highest conservation in these three groups were Asp-238 and Pro-235 (Fig. 3d).

In summary, most of the conserved contact sites in CD1 and FcRn lay in the α_2 domain and the top patch of the α_3 domain. In H301-gene product, only those contact sites involving the top patch of the α_3 domain were conserved to any degree. All three groups showed virtually no conservation of contact sites in the α_1 domain and the bottom patch of the α_3 domain (Fig. 3e).

Table 2. Mean percentage correlation of secondary-structure predictions compared with those observed in the crystal structures of HLA-A2.1 and HLA-Aw68.1

	α_1 domain*	α_2 domain*	α_3 domain†
MHC-like sequences			
HLA	43	64	71
H-2	42	63	78
Qa	50	66	77
HLA-E,F,G	47	57	76
Tla	47	38	75
Non-MHC-like sequences			
CD1	41	34	73
FcRn	48	49	73
CMV H301-gene product	52	17	55
			β_2 m domain†
β_2 m			72

* Unbiased prediction result based on PREDRB4 output.

† β -Sheet-biased prediction result based on PREDRB4 output.

Fig. 4. Robson secondary-structure prediction for α -chains of Class I HLA and the CMV H301-gene product sequence

(a) The average of 43 predictions for the HLA sequences is shown in centinats. The α_1 and α_2 domains were analysed using unbiased predictions, and the α_3 domain was analysed using β -sheet-biased predictions. Above the predictions, the crystallographically observed α -helices and β -strands of HLA are denoted by bars and lettering to follow Fig. 1 (○, α -helix; ●, β -strand). The broken lines correspond to coil and turn predictions. (b) The prediction for the CMV H301-gene product sequence is shown, as in (a).

Secondary-structure predictions

The Robson secondary-structure predictions for each group of sequences were compared with the observed secondary structures of the HLA-A2.1 and HLA-Aw68.1 molecules. The results were expressed as the mean percentage of residues where the predicted α -helix or β -strand structure matched the observed structure (Table 2). Similar analyses were performed for the CD1, FcRn and H301-gene product sequences in order to assess whether these related proteins would be predicted to fold in a similar or different manner to the HLA structures in the regions of the β_2 m- α -chain contacts, even though the α_1 and α_2 domains of CD1 show less than 20% sequence conservation with HLA, and these sequences generally exhibit smaller residue conservation and gaps and insertions found in these sequences when compared with the MHC-linked sequences.

α_1 domain. The averaged prediction for the α_1 domain in 43 HLA sequences agrees with the observed crystal structure for 43% of residues (Fig. 4a and Table 2). This low percentage was mainly due to mismatching of the long α -helix a2 with the predicted structure. Examination of the 43 individual predictions showed that from approximately residue 60 onwards the predictions exhibited much variation, and corresponded to the poor sequence conservation in that region (Fig. 1a). However the three observed β -strands b1–b3 at the beginning of the α_1 domain correlated well with the predictions (64% of residues matching). The percentage correlation of the averaged predictions for the H-2, Qa, HLA-E,F,G and Tla sequences were slightly higher than that for HLA (Table 2).

In the non-MHC-linked group, the prediction for the α_1 domain of H301-gene product (Fig. 4b) matched with 52% of the secondary structure when compared to the HLA crystal structure. The β -strands b1–b3 and most of the long α -helix a2 correlated well with the prediction. The FcRn and CD1 predic-

tions (not shown) showed a good correlation between the predicted structure and the observed long α -helix a2, but little of the observed β -sheet was predicted. The overall percentages of these three predictions were similar to that of the MHC-linked group (Table 2).

α_2 domain. The middle section of Fig. 4(a) shows that in the α_2 domain of the 43 HLA sequences several sections of the observed β -strands b1–b3 correlated well with the predictions, as did the four observed α -helices a1–a4. The overall correlation of the predicted with the observed structures was 64% (Table 2). The predictions for the H-2, Qa and HLA-E,F,G sequences were similar to those of the HLA with well-matched predictions of α -helices, but with minor mismatches in the prediction of β -strands. The corresponding percentage of the Tla prediction was surprisingly low.

The CD1, FcRn and H301-gene product predictions matched the observed structures for less than 50% of residues (Table 2). The prediction for FcRn suggested extra α -helix in the first half of the domain, and that for CD1 predicted extra β -sheet in the second half. The correlation between the H301-gene product prediction and the HLA crystal structure was poor, with matching at only 17% of residues (i.e. this is worse than random). For example, the H301-gene product prediction (Fig. 4b) showed only parts of the observed β -strands b1, b2 and b4, and little of the four α -helices a1–a4 in the second half of the domain.

α_3 domain and β_2 m. The third section of Fig. 4(a) shows that in the α_3 domain all six observed β -strands (b1, b2, b3, b5/b6, b7 and b8) were predicted for the HLA sequences with a match of 71% of residues using a β bias. Similar results were seen for other groups of sequences and also for β_2 m (Table 2), with good matches of 72–78%. The exception was the H301-gene product sequence, in which the observed β -strands b5/b6 and b7 were not predicted (Fig. 4b), and the match was decreased to 55%.

DISCUSSION

Analyses of the refined HLA crystal structures (Garrett *et al.*, 1989), sequence alignments and structure predictions permitted the main features of β_2 m binding to α -chains to be identified. Thus, for the α_1 domain, the contact with Asp-53 on β_2 m (Fig. 3a) was seen to be important. There were no instances in any of the 90 MHC-linked α -chain sequences where the three residues in contact with Asp-53 (β_2 m) (Table 1) were all absent. This was true even for the least-conserved contact sites, which occurred in the α_1 domain of the H-2 and Tla sequences with Asp-53 (β_2 m). For the ten contact sites in the α_2 domain, only that between Ile-1 (β_2 m) and Asp-119 (α_2 domain) was noticeably less conserved in the H-2 and Qa groups (Table 1). However, since this site involves the main-chain carbonyl group of Asp-119 (α_2 domain), an aspartic acid side chain is not required, and this means that all ten α_2 domain contact sites were maintained in the MHC-linked group. Four of these contacts were unusual, in that they appeared to involve a hydrogen-bond interaction of delocalized π electrons on the aromatic ring of Phe-62 with the amino group of Gln-96 (Fig. 3c; Tüchsen & Woodward, 1987).

The principal contact region was formed between the upper patch of the α_3 domain and β_2 m, which contained 19 of the total of 44 contact sites (Figs. 2b and 3d). All these sites lay on the fourth and fifth α_3 β -strands b5/b6 and b7 and the region between these. The lower patch formed six contacts with the C-terminus of β_2 m, and these may act as a stabilizer for the binding of β_2 m (Fig. 3e). Of the 25 contacts between the α_3 domain and β_2 m, 15 were conserved to more than 98 %, and a further eight to more than 90 % among the MHC-linked groups (Table 1). Generally, even though the Qa group exhibited slightly less conservation of contact residues, many of the non-conserved residues were involved in making contacts with main-chain carbonyl groups, so there was no stringent requirement for side-chain conservation.

Of the 44 contact sites on β_2 m (Table 1), 35 sites were totally conserved in all eight species. Of the nine non-conserved contacts, five involved main-chain groups, so the side chains were not crucial for contact formation (Table 1). In summary, the binding of β_2 m to the α -chain of Class I molecules involves 44 contacts, 40 of which were totally conserved in eight β_2 m sequences, and 37 of which were conserved to 90 % or more in 90 α -chain sequences.

In contrast, poor conservation of the contact sites was observed in the non-MHC-linked groups. In CD1, 55 % of the β_2 m contact sites were totally conserved, 57 % were conserved in FcRn, but only 39 % in H301-gene product (including main-chain contacts). The lack of contact site conservation was most notable in the α_1 domain and the bottom patch of the α_3 domain for all three non-MHC-linked groups, and in the upper patch of the α_3 domain of H301-gene product (Table 1). However, the total conservation of all residues in the α_1 and α_2 domains of CD1 compared with the HLA consensus sequence was less than 20 %, and for FcRn and H301-gene product was less than 30 % in each case.

Even though the predictions for the α_1 and α_2 domains exhibit moderate correlations in the range 42–66 % for the MHC-linked group of sequences, the β -strand predictions correlate well with the β_2 m contact sites. The low percentage prediction of the α -helices with the observed structure corresponds to the greatest sequence variation in the alignments; however, these helices are not directly involved with β_2 m binding. This may reflect a role for the antigenic peptide in stabilizing the conformation of the α_1 and α_2 domains, especially the α -helices. Peptide binding may facilitate the formation of contacts between the α_1 domain and β_2 m (Townsend *et al.*, 1989, 1990; Ljunggren *et al.*, 1990). The

predictions for the α_3 domain in the MHC-linked groups and β_2 m correlates well, with 71–78 % of residues matched (Table 2). The β -strands observed in the crystal structure were clearly seen (Fig. 4a). These high matches for the α_3 domain and β_2 m complement the observation that 23 out of 25 contact sites formed between these two domains were 90–100 % conserved in the 90 MHC-linked sequences. This is consistent with the view that these interactions are important for β_2 m binding. All the structure predictions for the Class I sequences were thus in good agreement with crystallographic data in the contact regions between the α -chain and β_2 m.

The smaller sequence conservation and the gaps and insertions in the FcRn and H301-gene product sequences and the α_3 domain of CD1 led to diminished correlations between the structure predictions and the HLA crystallographic data (Table 2). The smaller degrees of sequence similarity for these molecules may affect their secondary and tertiary structures. The FcRn prediction demonstrated more of the a2 α -helix, but less of the β -strands b1–b4, whereas the H301-gene product prediction demonstrated a number of β -strands in the first half of the domain, and two α -helices in the second half, giving its prediction a higher correlation with the observed crystal structures than the averaged prediction of the 43 HLA sequences. The predictive percentages for the α_2 domain were lower than those for the MHC-linked group, with a tendency towards predicting β -structure for the whole of the domain for H301-gene product. For the CD1 sequences, the α_1 domain prediction was similar to that for FcRn, and the α_2 domain also had a tendency to predict more β -structure, similarly to the H301-gene product prediction (Martin *et al.*, 1987; Calabi *et al.*, 1989). The predictions for the α_3 domains were similar to those of the MHC-linked groups, except for the H301-gene product sequence, which was matched at 20 % fewer residues. Thus, even though the predictions indicated that folding could occur in an MHC-like manner for the α_3 domain in each of these groups, the predictions for the other domains tended to suggest that the rest of the α -chain was folding differently.

Biochemical evidence shows that β_2 m binds to CD1 and FcRn (Martin *et al.*, 1986, 1987; Simister & Mostov, 1989). This suggests for FcRn that, if this molecule has an HLA-like structure, the two conserved contact areas in the α_2 domain and the top of the α_3 domain may be sufficient to permit β_2 m binding. The FcRn sequence alignment supports this conclusion, as no gaps or insertions occurred in the β -strand regions involved in β_2 m binding. The structure prediction for the α_2 and α_3 domains of FcRn showed that the predictions for the β -strand regions containing the conserved contact sites (b1, b2 and b3 in the α_2 domain and b5/b6 and b7 in the α_3 domain) correlated well with the crystal structure, even though other regions of the α -chain may appear to fold differently. The structure prediction for CD1 shows a propensity for β -sheet structure in the α_2 domain, and the prediction for the α_3 domain demonstrates 73 % correlation with the observed crystallographic structures. Given the known binding of β_2 m to CD1 and FcRn, the binding-site sequence similarity is of great interest, but the predictions on these folded structures lack strength. In summary, the top of the α_3 domain of CD1 and FcRn and the α_2 domain of FcRn would be predicted to bind β_2 m in an MHC-like manner, but the α_1 domain of both molecules and the α_2 domain of CD1 may fold differently and therefore bind to β_2 m in a non-MHC-like manner. Biochemical evidence exists to support this conclusion for CD1, since it has been shown that CD1 forms an intermolecular complex with HLA on human thymocytes (Amiot *et al.*, 1988). If CD1 is binding to β_2 m already bound to the α -chain of HLA, it presumably cannot bind to β_2 m using the same contact sites.

In vivo, CMV exists as β_2 m-coated particles (McKeating *et al.*,

1987), and the virus has been shown to bind β_2m *in vitro* (Grundy *et al.*, 1987a). It has been postulated that, because of sequence similarity to Class I HLA molecules, the H301-gene product is a likely candidate for the β_2m -binding protein in CMV (Beck & Barrell, 1988; Wiley, 1988). Indeed, recent data suggest that when the H301 gene is expressed in BHK cells with the use of a vaccinia virus vector, the resultant protein could bind β_2m (Browne *et al.*, 1990). However, as yet, there is no biochemical evidence to show that the H301-gene product is expressed in infected cells or on the virus particle.

The sequence alignment for the α_1 domain of H301-gene product shows a gap and an insertion in the region that has been defined as β_2m -binding for the two crystallized HLA molecules. Although the secondary-structure predictions for this domain were the best matched for all the eight groups of sequences against the observed structure, the presence of this gap and insertion and the observation that none of the contact sites on this domain are conserved on H301-gene product make it unlikely that this domain could bind β_2m in an MHC-like manner. In the α_2 domain, the H301-gene product sequence alignment of Fig. 1(b) shows that the two conserved contact sites were found at the centre of a region of 42 residues without gaps or insertions. However, the prediction for the α_2 domain suggested that the β -strand b3 of HLA (close to this contact site) becomes an α -helix in H301-gene product (Fig. 4b). In the α_1 domain and the top patch of the α_3 domain, the poorly conserved contacts (Table 1 and Figs. 3a, 3b and 3d) do not appear sufficient for β_2m binding to the α -chain of H301-gene product in an MHC-like manner. In the alignment with the HLA consensus, there is a two-residue insertion at the centre of what would otherwise be a conserved α_3 -domain patch for β_2m binding (Fig. 1c). This insertion was also identified in the alignment proposed by Beck & Barrell (1988), and by an automated alignment using the ALIGN software (Protein Identification Resource, National Biomedical Research Foundation, Washington, DC, U.S.A.). This insertion was coincident with the failure to predict the observed β -strand b5/b6 (and possibly b7 also) that was achieved with all the other α -chain predictions (Fig. 4b and Table 2). In H301-gene product this region contains four proline residues in close proximity, which would not favour regular secondary structures. These poor correlations with the HLA crystal structure do not indicate that β_2m could bind to the H301-gene product in an MHC-like manner. If, as recent data suggest, β_2m can bind to the H301-gene product (Browne *et al.*, 1990), it is concluded that this binding is non-MHC-like.

It has been proposed that CMV uses Class I molecules as a receptor for cell attachment and entry (Grundy *et al.*, 1987b). The model proposed is that the β_2m -binding protein on CMV binds to Class I HLA on the cell surface by displacing β_2m from the α -chain, and binding to the α -chain via viral-bound β_2m (Grundy *et al.*, 1987b,c). Such a model implies that CMV and HLA bind to β_2m at different sites, and there are some experimental data to support this conclusion. The monoclonal antibody C21 does not recognize β_2m in association with the Class I α -chain, but does recognize free β_2m , suggesting that the C21 epitope on β_2m is masked when bound to the α -chain (Liabeuf *et al.*, 1981). However, C21 does recognize β_2m bound to CMV (Grundy *et al.*, 1987c), and thus the C21 epitope is not masked when β_2m binds to CMV. This strongly suggests that the binding site for β_2m on CMV is different to that on the α -chain of Class I. It is interesting to speculate that, if H301-gene product is the β_2m -binding protein of CMV, it may form an intermolecular complex with HLA in a manner similar to that of the complex formed between CD1 and HLA (Amiot *et al.*, 1988). In this case, the binding of H301-gene product to β_2m would be non-MHC-like. Further biochemical studies are required to elucidate the precise nature of β_2m binding to CMV.

Note added in proof (received 8 May 1991)

Examination of two recently reported avian β_2m sequences from chicken and turkey (Welinder *et al.*, 1991) shows that the β_2m - α -chain contact sites are overwhelmingly conserved in non-mammalian β_2m , even though the percentage of totally conserved residues in β_2m has decreased from 50% to 35%. This re-analysis has been incorporated into the text.

We thank the Science and Engineering Research Council for a studentship to V.A.T.-C., and the Wellcome Trust and the Medical Research Council for the provision of a Silicon Graphics Workstation. We thank A. S. Nealis for computational assistance.

REFERENCES

- Ajitkumar, P., Geier, S. S., Kesari, K. V., Borriello, F., Nakagawa, M., Bluestone, J. A., Saper, M. A., Wiley, D. C. & Nathenson, S. G. (1988) *Cell* **54**, 47–56
- Amiot, M., Dastot, H., Degos, L., Dausset, J., Bernard, A. & Bousmell, L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4451–4454
- Arnold, B., Burgert, H. G., Archibald, A. L. & Kvist, S. (1984) *Nucleic Acids Res.* **12**, 9473–9487
- Balk, S. P., Bleicher, P. A. & Terhorst, C. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 252–256
- Beck, S. & Barrell, B. G. (1988) *Nature (London)* **331**, 269–272
- Becker, J. W. & Reeke, N. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4225–4229
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987a) *Nature (London)* **329**, 506–512
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987b) *Nature (London)* **329**, 512–518
- Bouillot, M., Choppin, J., Cornille, F., Martinon, F., Papo, T., Gomard, E., Fournie-Zaluski, M.-C. & Levy, J. P. (1989) *Nature (London)* **339**, 473–475
- Brown, J. H., Jardetsky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J. & Wiley, D. C. (1988) *Nature (London)* **332**, 845–850
- Browne, H., Smith, G., Beck, S. & Minson, T. (1990) *Nature (London)* **347**, 770–772
- Calabi, F., Jarvis, J. M., Martin, L. H. & Milstein, C. (1989) *Eur. J. Immunol.* **19**, 285–292
- Davidson, W. F., Kress, M., Khoun, G. & Jay, G. (1985) *J. Biol. Chem.* **260**, 13414–13423
- de Preval, C. & Mach, B. (1983) *Immunogenetics* **17**, 133–140
- Ellis, S. A., Palmer, M. S. & McMichael, A. J. (1990) *J. Immunol.* **144**, 731–735
- Ennis, P. D., Jackson, A. P. & Parham, P. (1988) *J. Immunol.* **141**, 642–651
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
- Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. (1989) *Nature (London)* **342**, 692–696
- Gates, F. T., III, Coligan, J. E. & Kindt, T. J. (1979) *Biochemistry* **18**, 2267–2272
- Gates, F. T., III, Coligan, J. E. & Kindt, T. J. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 554–558
- Geraghty, D. E., Koller, B. H. & Orr, H. T. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9145–9149
- Geraghty, D. E., Wei, X., Orr, H. T. & Koller, B. H. (1990) *J. Exp. Med.* **171**, 1–18
- Grey, H. M., Buus, S., Colon, S., Miles, C. & Sette, A. (1989) *Philos. Trans. R. Soc. London B* **323**, 545–552
- Groves, M. L. & Greenberg, R. (1982) *J. Biol. Chem.* **257**, 2619–2626
- Grundy, J. E., McKeating, J. A. & Griffiths, P. D. (1987a) *J. Gen. Virol.* **68**, 777–784
- Grundy, J. E., McKeating, J. A., Ward, P. J., Sanderson, A. R. & Griffiths, P. D. (1987b) *J. Gen. Virol.* **68**, 793–803
- Grundy, J. E., McKeating, J. A. & Griffiths, P. D. (1987c) *MHC + X: Complex Formation and Antibody Induction (Proc. Workshop-Symposium Meeting, Amsterdam)* (Ivanyi, P., ed.), pp. 171–177, Springer-Verlag, Berlin
- Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- Klein, J. & Figueroa, F. (1986) *Immunol. Today* **7**, 41–44
- Koller, B. H., Geraghty, D. E., Shimizu, Y., DeMars, R. & Orr, H. T. (1988) *J. Immunol.* **141**, 897–904
- Kourilsky, P. & Claverie, J.-M. (1989) *Adv. Immunol.* **45**, 107–193

- Lancet, D., Parham, P. & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3844–3848
- Liabeuf, A., De Kaouel, C., Le, B., Kourilsky, F. M., Malissen, B., Manuel, Y. & Sanderson, A. R. (1981) *J. Immunol.* **127**, 1542–1548
- Ljunggren, H.-G., Stam, N. J., Öhlén, C., Neeffjes, J. J., Höglund, P., Heemels, M.-T., Bastin, J., Schumacher, T. N. M., Townsend, A., Kärre, K. & Ploegh, H. L. (1990) *Nature (London)* **346**, 476–480
- Long, E. O. & Jacobson, S. (1989) *Immunol. Today* **10**, 45–48
- Martin, L. H., Calabi, F. & Milstein, C. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9154–9158
- Martin, L. H., Calabi, F., Lefebvre, F.-A., Bilsland, C. A. G. & Milstein, C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9189–9193
- Maryanski, J. L., Abastado, J. P., MacDonald, H. R. & Kourilsky, P. (1989) *J. Immunol.* **193**, 193–196
- Mattson, D. H., Shimojo, N., Cowan, E. P., Baskin, J. J., Turner, R. V., Shvetsky, B. D., Coligan, J. E., Maloy, W. L. & Biddison, W. E. (1989) *J. Immunol.* **143**, 1101–1107
- Mauxion, F. & Kress, M. (1987) *Nucleic Acids Res.* **15**, 7638
- Mayer, W. E., Jonker, M., Klein, S., Ivanyi, P., van Seventer, G. & Klein, J. (1988) *EMBO J.* **7**, 2765–2774
- Maziarz, R. T., Burakoff, S. J. & Bluestone, J. A. (1988) *J. Immunol.* **140**, 4372–4377
- McKeating, J. A., Griffiths, P. D. & Grundy, J. E. (1987) *J. Gen. Virol.* **68**, 785–792
- Morita, T., Delarbre, C., Kress, M., Kourilsky, P. & Gachelin, G. (1985) *Immunogenetics* **21**, 367–383
- Novotny, J. & Auffray, C. (1984) *Nucleic Acids Res.* **12**, 243–255
- Owen, M. J., Kissonerghis, A.-M. & Lodish, H. F. (1980) *J. Biol. Chem.* **255**, 9678–9684
- Parham, P., Lomen, C. E., Lawlor, D. A., Ways, P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M. & Ennis, P. D. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4005–4009
- Parnes, J. R. & Seidman, J. G. (1982) *Cell* **29**, 661–669
- Perkins, S. J., Haris, P. I., Sim, R. B. & Chapman, D. (1988) *Biochemistry* **27**, 4004–4012
- Sege, K., Rask, L. & Peterson, P. A. (1981) *Biochemistry* **20**, 4523–4530
- Shimizu, Y., Geraghty, D. E., Koller, B. H., Orr, H. T. & DeMars, R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 227–231
- Simister, N. E. & Mostov, K. E. (1989) *Nature (London)* **337**, 184–187
- Srivastava, R., Duceaman, B. W., Biro, P. A., Sood, A. K. & Weissman, S. M. (1985) *Immunol. Rev.* **84**, 93–121
- Steinmetz, M. & Hood, L. E. (1983) *Science* **222**, 727–733
- Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6613–6617
- Taylor, W. R. (1986) *J. Mol. Biol.* **188**, 233–258
- Townsend, A., Öhlén, C., Bastin, J., Ljunggren, H.-G., Foster, L. & Kärre, K. (1989) *Nature (London)* **340**, 443–448
- Townsend, A., Elliott, T., Cerundolo, V., Foster, L., Barber, B. & Tse, A. (1990) *Cell* **62**, 285–295
- Tüchsen, E. & Woodward, C. (1987) *Biochemistry* **26**, 8073–8078
- Tysoe-Calnon, V. A., Grundy, J. E., Nealis, A. S. & Perkins, S. J. (1990) *Biochem. Soc. Trans.* **18**, 930–931
- Vega, M. A., Bragado, R., Ezquerro, A. & Lopez de Castro, J. A. (1984) *Biochemistry* **23**, 823–831
- Watts, S., Wheeler, C., Morse, R. & Goodenow, R. S. (1989) *Immunogenetics* **30**, 390–392
- Welinder, K. G., Jespersen, H. M., Walther-Rasmussen, J. & Skjødt, K. (1991) *Mol. Immunol.* **28**, 177–182
- Wiley, D. C. (1988) *Nature (London)* **331**, 209–210
- Wolfe, P. B. & Cebra, J. J. (1980) *Mol. Immunol.* **17**, 1493–1505
- Zijlstra, M., Bix, M., Simister, N. E., Loring, J. M., Raulet, D. H. & Jaenisch, R. (1990) *Nature (London)* **344**, 742–746

Received 30 August 1990/18 December 1990; accepted 23 January 1991