Crystal Structure of Mouse CD1: An MHC-Like Fold with a Large Hydrophobic Binding Groove

Z.-H. Zeng,* A. R. Castaño,† B. W. Segelke, E. A. Stura, P. A. Peterson, I. A. Wilson‡

CD1 represents a third lineage of antigen-presenting molecules that are distantly related to major histocompatibility complex (MHC) molecules in the immune system. The crystal structure of mouse CD1d1, corresponding to human CD1d, at 2.8 Å resolution shows that CD1 adopts an MHC fold that is more closely related to that of MHC class I than to that of MHC class II. The binding groove, although significantly narrower, is substantially larger because of increased depth and it has only two major pockets that are almost completely hydrophobic. The extreme hydrophobicity and shape of the binding site are consistent with observations that human CD1b and CD1c can present mycobacterial cell wall antigens, such as mycolic acid and lipoarabinomannans. However, mouse CD1d1 can present very hydrophobic peptides, but must do so in a very different way from MHC class Ia and class II molecules.

CD1 is a family of nonpolymorphic cell surface glycoproteins encoded outside the major histocompatibility complex (MHC), but with a distant relation to MHC class I and class II molecules (*1*). CD1 appears to be more closely related to MHC class Ia and Ib proteins by sequence homology, domain organization (α1, α2, α3, and β2m) (*1*), and association with β2 microglobulin (β2m) than to class II molecules. Classical MHC Ia molecules present short peptides (8 to 10 residues) (*2*) from endogenously synthesized proteins to $CD8^+$ cytotoxic T lymphocytes (CTLs), whereas class II molecules present series of longer nested peptides (13 to 25 residues in length) (*3*) from exogenous origin to $CD4^+$ T cells. A number of genes in and outside the MHC also encode nonpolymorphic class I–like or class Ib proteins that share sequence similarity with class Ia molecules, as well as association with the non-MHC encoded β2 microglobulin. Some of these specialized molecules, such as H-2M3, present a distinct set of peptide (*4*, *5*) or nonpeptide antigens (*6*).

The CD1 family can be divided into two groups by sequence homology. Group I, consists of CD1a, -b, and -c isotypes, and group II contains CD1d; CD1e is proposed to be an intermediate isotype (*7*). Only the group II CD1d isotypes are preserved in human (hCD1), mouse (mCD1), rabbit, and rat. Sequence similarity is substantially higher for the same isotypes from different species than for different isotypes within the same species (*1*, *8*) suggesting that each group of CD1 molecules could have a different function.

CD1 isotypes are differentially expressed with restricted tissue distribution, including immature cortical thymocytes (*1*), dendritic cells, and a subpopulation of B lymphocytes, and can interact with both γδ and αβ T cells (*9*–*11*). Different antigens may be presented by the different isotypes. Nonpeptide, processed forms of mycobacterial cell wall constituents, such as mycolic acid lipids and lipoarabinomannan lipoglycans (LAMs), can be presented by hCD1b (*12*–*14*) and hCD1c (*15*). Mouse CD1d1 (mCD1d1) has been shown to be capable of binding and presenting long peptides (12 to 22 residues) that contain a specific hydrophobic core motif (*11*). Most CD1s appear to be localized to endosomal MHC class II compartments (MIICs) in which MHC class II molecules are thought to be loaded with exogenous antigen (*16*). In the CD1b-presentation pathway, the mannose receptor is responsible for the uptake and delivery of LAMs and their transport to late endosomes and MIICs (*14*), in a process facilitated by the acid pH of the endosome (*17*). The role of accessory molecules in CD1-T cell interactions is unclear; mCD1d1 is recognized by $CD8^+$ (*11*) and $CD4^+$ (*18*) T cells, and hCD1b can interact with $CD4^-CD8^-$ T cells (*13*). A subset of $NK1^+$ T cells with an invariant V_α T cell receptor (TCR) also appears to be specific for human and mouse CD1 (*19*, *20*). Also unresolved is whether CD1 isotypes of group I present only lipids or glycolipids whereas group II CD1d presents hydrophobic peptides. Like H-2M3, which binds *N*-formylated peptides (*4*), CD1 may have evolved to present antigens that are predominantly found in microbial pathogens.

Many questions have arisen about the possible structure of CD1. Specifically, how does CD1 interact with ligand, with CD8 or CD4, and with β2m (*21*)? Some secondary structure predictions have forecasted that CD1 folds differently from MHC class I and class II molecules (*7*), suggesting that CD1

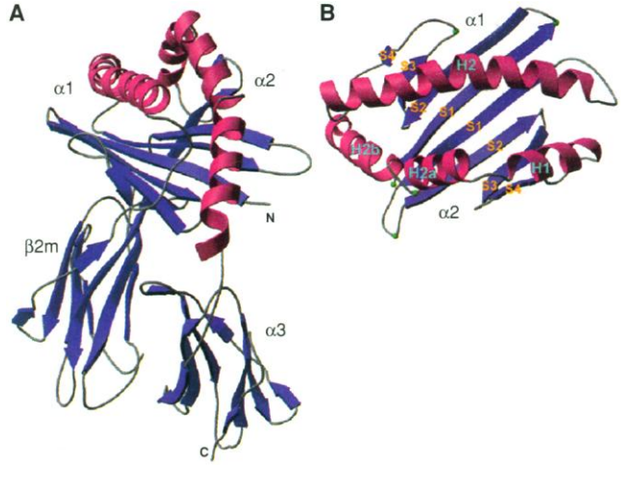

Fig. 1. The mouse CD1d1 structure. (**A**) The ribbon diagram of mouse CD1d1 is shown with each domain labeled, with β strands in purple and α helices in red. CD1 folds very much like MHC class I molecules with similar domain arrangements and secondary structure elements, although there are two notable differences in secondary structure; CD1 does not have the single-turn α1 H1 helix of class I molecules and CD1 has a different arrangement in the bulge of the β2m S4 strand compared to H-2K^b (*5*). (**B**) Top view of the ligand binding α1 and α2 domains. Each of the secondary structure elements are labeled with H referring to α helix and S referring to β strand. Green spheres mark the Cα positions for the Asn residues of possible N-linked glycoslylation sites. Generated with MOLSCRIPT (*48*) and RASTER3D (*49*).

Z.-H. Zeng, B. W. Segelke, E. A. Stura, and I. A. Wilson are in the Department of Molecular Biology and the Skaggs Institute for Chemical Biology at the Scripps Research Institute, La Jolla, CA 92037, USA. A. R. Castaño and P. A. Peterson are at the R. W. Johnson Pharmaceutical Research Institute at the Scripps Research Institute, La Jolla, CA 92037, USA.

*Present address: Department of Protein Engineering, Institute of Biophysics, Chinese Academy of Sciences, Chao Yang District, Beijing 100101, China.

†Present address: Centro Nacional de Biologia Fundamental, Instituto de Salud Carlos III, Ctra Pozuelo Km2, 28220 Majadahonda (Madrid), Spain.

‡To whom correspondence should be addressed. E-mail: wilson@scripps.edu

may have a different mode of antigen presentation and T cell interaction. Actually CD1 has slightly greater amino acid sequence similarity to the neonatal Fc receptor (FcRn), which participates in the immunity of the newborn via uptake of maternal immunoglobulin G (IgG) from breast milk into the blood stream of the suckling infant. Although FcRn adopts an MHC-like fold, it does not have a functional binding groove or support antigen presentation (*22*). To gain insight into the functions of CD1, we have determined the crystal structure of mCD1d1 molecule at 2.8 Å resolution in the absence of added antigen.

Mouse CD1d1 structure and the hydrophobic groove. The extracellular region of mCD1d1 associated with β2m was obtained as described (*11*), and crystallized. The crystal structure was determined by molecular replacement, with FcRn (*22*) and HLA-B27 (*5*) as search models (Table 1 and Fig. 1).

Despite the low sequence identity, and contrary to some predictions of secondary structure (*7*), the overall structure of mCD1d1 has a strikingly similar secondary, tertiary, and quaternary structure (Figs. 1 and 2) to those of MHC class I and II molecules and related proteins (*5*, *22*, *23*). The α chain folds into the standard three domains (α1, α2, α3) and is closely associated with β2m. The α1 and α2 domains form a binding groove superdomain composed of an eight-stranded antiparallel β-pleated sheet with two long antiparallel α-helical structures that sit atop and traverse the β-sheet platform (Fig. 1B). The α3 and β2m domains have c-type Ig folds (*24*). All the secondary structural units of MHC class I molecules (*5*) are present in mCD1d1, except for the first α1 helix (H1), which adopts an irregular, extended conformation (Fig. 1).

The α1 domain of CD1 is structurally more similar (*25*) to MHC class I molecules and FcRn than to class II, whereas the α2 domain is more similar (*25*) to FcRn than to class I and is quite distinct from class II. For the entire α1-α2 superdomain, CD1 is as similar to FcRn as it is to class I and is most distinct from class II molecules (*25*) (Fig. 2). The Ig domains of CD1 are equally similar to all class I molecules, including FcRn, and more distant from class II molecules (*25*).

However, some important differences are evident in the quaternary assembly. Relative to β2m, the CD1 α1α2 domains have a slight rotation and translation compared to the arrangement in class I molecules (*5*, *26*) or to the equivalent α2α1-β1 relation in class II molecules (*23*). The total buried surface area between β2m and the heavy chain is slightly less than for class I molecules (*27*), but it is more hydrophobic (*28*), which may account for the CD1-β2m complex being more stable to thermal denaturation (*29*). The most notable difference in quaternary structure occurs in α3-β2m association. The pseudo-dyadic rotation necessary to superimpose α3 onto β2m in CD1 is more similar to class II than to class I molecules; the large translation component is common to all MHCs (*26*). In this particular assembly, the outermost α3 S1 strand now comes closer to β2m and results in a slightly different pattern of H bonding and van der Waals interaction. The β bulge that breaks the S4 strand in most β2m structures is also found in mCD1d1 β2m, unlike its counterpart in H-2K^b (*5*).

The CTLs specific for CD1 can be CD8$^+$ (*11*, *30*), CD4$^+$ (*18*), or double negative (*10*, *19*). The highly acidic loop, α3 residues 220 to 229, from class I implicated in CD8 binding (*31*) and its equivalent CD4 binding to class II (*32*), retains some structure similarity in CD1 and even in FcRn, which does not bind accessory molecules (*33*). However, a single residue deletion occurs in the putative binding loop (*31*, *32*) of CD1 compared with other MHC class I and II molecules.

The relative position of the helical segments and outermost strands differs in CD1 and causes a significant alteration in the dimensions of the cleft formed between α1 and α2 (Fig. 2B). The width of the cleft is dictated by the distance between the α2 H1, H2a, and H2b helical segments and the α1 H2 helix (Figs. 1 and 2). Class I and class II molecules have the widest cleft (18 to 20 Å at the middle), whereas the binding groove of CD1 is narrower and of nearly constant width (~14 Å); FcRn has the narrowest dimensions (10 to 13 Å) between the opposing α helices (Fig. 2A). Proline-162 at the junction between the α2 H2a and H2b helices of FcRn has been proposed (*22*) to be responsible for the concerted movement of the α2 H2a and H1 helices (and the α2 S3 and S4 β strands) toward

Table 1. Crystallization and structure determination of mCD1d1. The best crystals were obtained from 5 μl of mCD1d1 at 6 mg/ml (*53*) with 2.5 μl of precipitant [2.0 M $(NH_4)_2SO_4$, 0.15 M sodium citrate, 2 mM $ZnSO_4$, pH 6.2]. Crystals were enlarged by the addition of 3 μl of protein-precipitant mixture pre-equilibriated at a 3% lower concentration of ammonium sulfate. The crystals are monoclinic $P2_1$, with two molecules per asymmetric unit, 44 percent solvent content, V_m of 2.19 Å^3/dalton (*54*) and cell dimensions a = 103.4 Å, b = 76.2 Å, c = 59.3 Å, and β = 102.3°. Data were collected on a Siemens multiwire detector mounted on an Elliott GX-18 rotating anode. Data were processed with XENGEN (*55*). Initial phase estimates were calculated by molecular replacement (MR) with AMORE from the CCP4 suite of programs (*56*). Initial searches were made using each of MHC molecules available at the time in the Protein Data Bank (*5*, *23*, *57*) and FcRn (*22*). The clearest solutions and initial phase estimates were obtained with FcRn (*22*) or HLA-B27 (*5*) as search models. Normalized structure factors (E) from 20.0 to 4.0 Å were used in the rotation and translation searches. The cut-off radius for Patterson vectors was set at 30 Å. Correct placement of molecules in the asymmetric unit was aided by interpretation of the native Patterson map and the self-rotation function (*58*). After initial placement, the position of the first molecule was refined with X-PLOR Patterson correlation refinement (*59*). Refinement and map calculations were done using X-PLOR (*59*) while model rebuilding was done using TURBO (*60*). The model was rebuilt into 10 percent omit maps. Noncrystallographic symmetry (NCS) restraints were applied to all main chain atoms and side chain atoms not involved in crystal contacts or intermolecular contacts within the asymmetric unit (2258 atoms, total). The final model was refined to a crystallographic R value of 19% at 2.67 Å. The post-refinement R_{free} with partial NCS applied is 33%.

	Collected	Unique	Redundancy	R_{sym}*
Reflections (N)	57,464	22,974	2.5	9.5%
Resolution range (Å)	20.0–10.0	10.0–2.9	2.9–2.67	
Completeness (%)	99.5	86.0	53.0	
Average I/σ	30.2	8.9	1.5	
Resolution (Å)	10.0–2.67	6.0–2.67		
R value (%)	19.0	18.5		
Post refinement R_{free}† (%)	33.0			
Average B† (Å^2)	22.8			
	Bonds	Angles	Dihedrals	Impropers
Rmsd	0.013 Å	1.95°	27°	1.7°
	Favored	Allowed	Generous	Unfavored
Ramachandran plot (%)‡	84.2	15.4	0.5	0.0
	C_α	Main chain	All atoms	
Number of atoms	372	1487	3013	
Rmsd 2 molecules§ (Å)	0.29	0.38	0.86	

*Weighted $R_{sym} = 100 * [\Sigma_j\Sigma_i (\langle I_{ij} - \langle I \rangle_j)/\sigma_{ij})^2/\Sigma_j\Sigma_i(I_{ij}/\sigma_{ij})^2]^{1/2}$ where $\langle I \rangle_j$ is the mean of the I_{ij} observations and σ_{ij} is the standard deviation of I_{ij}. †The post-refinement R_{free} was calculated by subjecting the final coordinates to one cycle each of simulated annealing, positional refinement, and group B value refinement with 90 percent of the data. As a separate check, an R value of 34% was calculated against an independently collected data set (~20% complete to 2.7 Å). Three loops, 87 to 95, 107 to 109, and 196 to 201, were excluded from the calculation, as their average B values were greater than 50 Å^2. ‡The Ramanchandran plot was generated in PROCHECK (*55*) and the rmsd's from ideality were calculated with X-PLOR (*57*). §Comparison between two molecules of the asymmetric unit in the crystal.

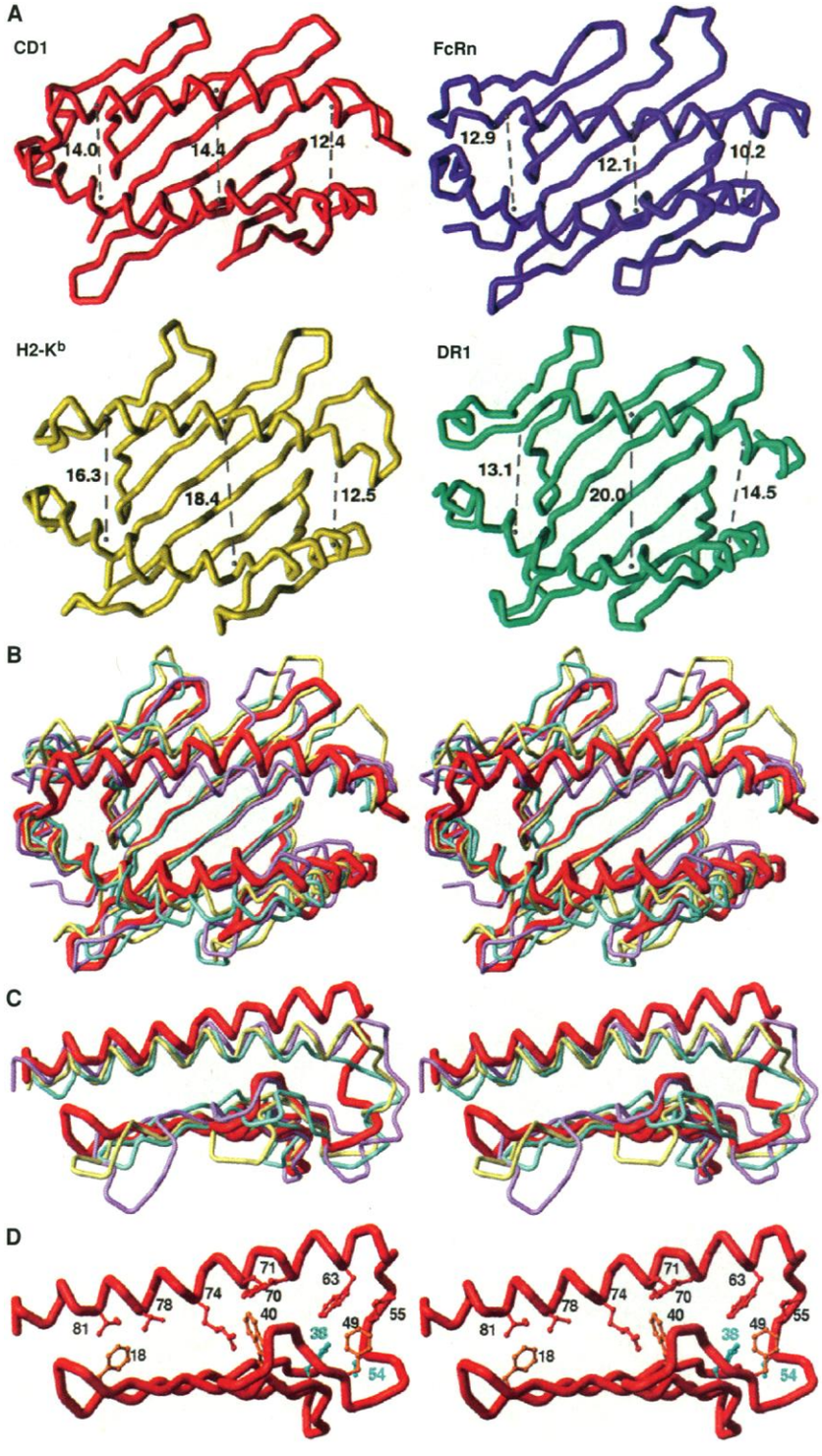

Fig. 2. Comparison of mouse CD1d1, MHC class I and II, and FcRn binding domains. (**A**) The α1α2 domains are viewed from above, looking into the ligand binding groove. The α1α2 domains are shown for mCD1d1, FcRn, and H-2K^b and the α1 and β1 domains for HLA-DR1. The interdomain distances from the center of the opposing helices are shown at three representative points. CD1 has a much narrower groove when compared with class I or class II MHC molecules; the α1-α2 inter-helical distance in FcRn is even less. (**B**) The α1α2 domains of mCD1d1, H-2K^b, and FcRn, are shown in stereoview superimposed with the α1 and β1 domains of HLA-DR1. The β-sheet platform superimposes closely, with the exception of a few loops and the external strands, while there are a variety of arrangements for the helical segments. The long α1 H2 helix bends such that its NH_2-terminal end is closer to α2 helices as in FcRn. The α2 H2a segment of FcRn and CD1 is notably closer to the α1 H2 helix than in class I and class II molecules. The α2 H1 helix in CD1 is in a similar position to class I and class II but differs from FcRn. The Cα trace of CD1 is shown in thick red tubes, FcRn in purple, H-2K^b in yellow, and HLA-DR1 in light green. (**C**) Viewed from the α1 side, the α1 domains of the four MHC-like molecules in (A) are shown superimposed on their β strands. The α1 H2 helix of CD1 is notably higher than in the other molecules with the segment of structure preceding the H2 helix in an extended rather than helical conformation. (**D**) CD1 conserved residues 18, 40, and 49 (orange) interact with the residues from the α1 H2 helix (shown in red). Also shown are Ser^{54} and His^{38} (cyan), which make main chain–side chain contacts that may stabilize the extended loop preceding the H2 helix. Some extra electron density near the CD1 group II conserved NH_2-linked glycosylation site at Asn^{42}, which approaches the α1 H2 helix, may also contribute to stabilization of the H2 helix by interaction with group II conserved Arg^{74}. The transformations to superimpose homologous Cα's were calculated with SUPPOS (*50*), and the figure was generated with MIDAS (*51*).

α1, effectively collapsing the groove at one end. CD1 has a structurally homologous proline (Pro^{169}) and a corresponding shift in the α2 H2a helix, S3 and S4 strands (Fig. 3). However, unlike FcRn, the α2 H1 helix in CD1 is not shifted substantially toward α1 (Fig. 3). Thus, a proline in this position in non-classical MHC class Ib molecules appears to affect the relative location of the α2 H2a helix, but positioning of the α2 H1 helix can be uncoupled from the effects of the proline substitution. In fact, the α2 H1 helix position in FcRn is more likely a consequence of stabilization by several hydrogen bonds and nine pairs of van der Waals contacts between α1 and α2 that are not present in CD1 (Fig. 3).

The CD1 α1 helix is raised 4 to 6 Å higher above the β sheet than in other MHC molecules or FcRn (Fig. 2C) and results in a deeper groove. Three highly conserved side chains from the β-sheet platform appear to contribute to the elevation of the α1 H2 helix; Phe^{49}, Trp^{40}, and Phe^{18} contact residues at the NH_2-terminal, central, and COOH-terminal portions of α1 H2, respectively. Other significant differences are the rearrangement of the external α1 β strands (S3 and S4) and deformation of the short α1 H1 helix into an irregular, extended structure that is stabilized by several main chain–side chain H bonds (Fig. 2D).

The altered dispositions of the α1 and α2 helical segments are also accompanied by a concomitant adjustment of the corresponding S3 and S4 strands toward the nearest helix. As a consequence, the whole β-sheet platform has substantially greater curvature than in class I and class II molecules (Fig. 2C). The S4 strand of α2 becomes detached from the sheet, as does the small α2 S4 strand in FcRn (*22*) and also

the α1 S4 strand of class II molecules (*23*). Despite the movement of the two S4 strands toward the α helices, fewer overall interactions occur between the α helices and the β sheet mainly because of the increased α1 H2 helix elevation.

MHC class I and class II molecules accommodate bound peptide antigens in a groove formed between the α helices and the β-sheet platform (*5*). Polymorphic residues in the groove determine the size and stereochemistry of different pockets so that individual peptide motifs can be conferred by each MHC allele (*23*, *34*). The number and disposition of these pockets is not fixed, or even well defined, in that polymorphic substitutions can markedly change the surface topography of the groove (*5*) (Fig. 4). The prototypic groove of class I molecules consists of six pockets (labeled A to F) (*5*) and is closed at both ends (Fig. 4). Class II molecules (*23*) have seven main pockets (labeled 1 to 9) (*35*) in a groove that is open ended (Fig. 4), while FcRn (*22*) has no functional groove (Fig. 4).

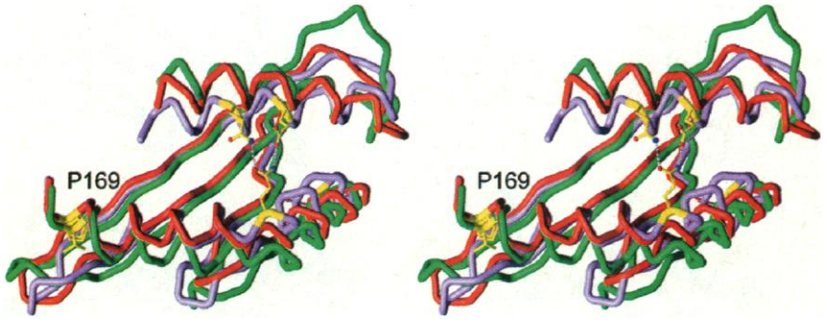

Fig. 3. The proline kink in the H2 helical region of the α2 domain. A portion of the α1 and α2 backbones of CD1 (red), FcRn (purple), and H-2K^b (green) are shown superimposed in stereo. Residues that are involved in closing the groove are shown in yellow. A proline at the beginning of the H2b helix is at the initiating position of a kink in the H2a helix that effectively narrows the center and COOH-end of the groove in FcRn. Specific hydrogen bonds from residues in FcRn close the COOH-end of the groove by forming specific interactions from the COOH-end of the α1 helix to the α2 H1 helix. These residues differ in CD1 and do not interact.

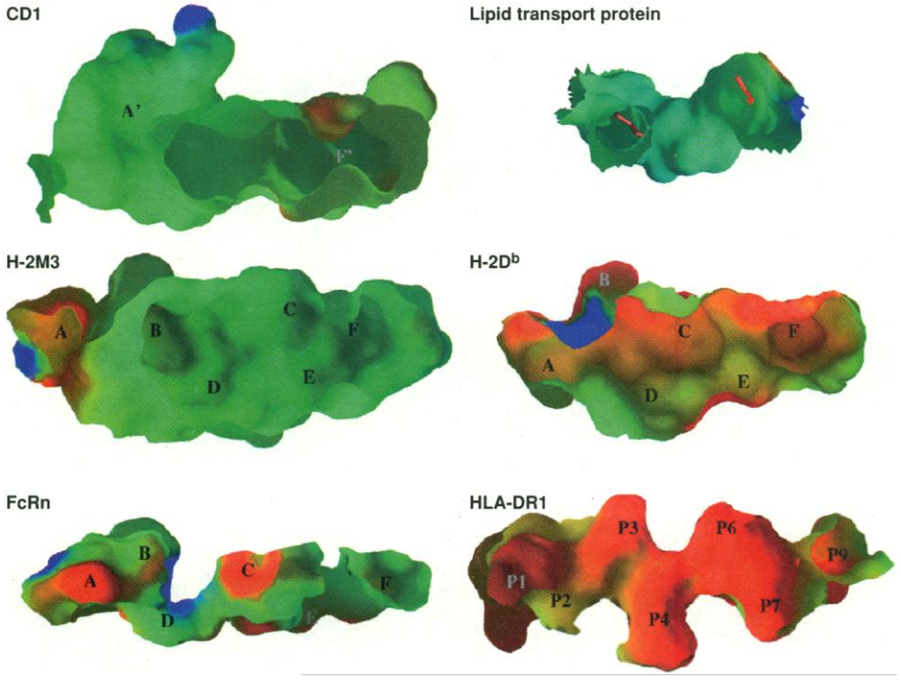

Fig. 4. Comparison of molecular surfaces of the binding grooves. The cut away molecular surface (*37*) of six MHC or MHC-like molecules are shown with electrostatic potentials mapped to their surface. The ligand binding pocket of Maize nonspecific lipid binding protein with palmitic acid bound is also shown for comparison. The groove molecular surface of CD1, like that of H-2M3 and nonspecific lipid binding protein, is mostly neutral and hydrophobic compared to the grooves of the other MHC molecules or to FcRn. CD1 has by far the largest and deepest groove of all molecules represented (*37*) (pocket A′ is approximately the same size (in surface area) as the entire binding site of nonspecific lipid binding protein). Many of the specificity pockets present in class I molecules have merged to form only two very large pockets in CD1. In FcRn (*22*), there is essentially no groove. The peptide specificity pockets are labeled A to F for class I (*5*) and P1-P9 (*35*) for class II (*23*) molecules. The FcRn is labeled to show the corresponding location of pockets in class I and is not intended to indicate binding pockets. Electrostatics potentials were calculated in GRASP (*52*) with full formal charge. Positive potential (≥15 mV) is colored blue, neutral potential (0 mV) is colored green, and negative potential (≤ −10 mV) is colored red.

Despite structural similarity in α1 and α2 with FcRn, CD1 does have a substantial ligand binding groove. However, the putative ligand binding site of CD1 is quite distinct from other MHC-like molecules (Fig. 4). The proximity of the α2 H2a helix to α1 narrows the groove in the center, while the raised elevation of the α1 helix deepens it (Fig. 2). The modified arrangement of secondary structural elements, coupled with the nature of the side chains that line the groove, substantially increases the dimensions of the groove but reduces the number of pockets (Fig. 4). The substitution of several conserved aromatic residues in class I molecules by smaller residues in CD1 helps to create a single large pocket, defined here as A′, that subsumes pockets A, B, C, and D of the class I MHC groove (Fig. 4). Pocket A′ has a cylindrical shape, 7 Å in height and about 12 Å in width, with a depression in its center due to proximity of the side chains of α1 Phe^{70} (H2) at the top and α1 Cys^{12} (S1) at the bottom. At the opposite end of the groove, pockets E and F combine into another, narrower pocket, F′, which is considerably smaller than pocket A′. The A′ pocket has maximum width (*36*) of ~15 Å and is 18 to 20 Å long, whereas the F′ pocket is ~10 Å wide and 10 to 12 Å long. The combined length of the A′ and F′ pockets is ~30 Å, which is slightly longer than the standard 25 Å long groove in MHC class I molecules (*5*). The groove is closed at either end and is only accessible through a narrowed entrance, 18 Å long and 6 to 7 Å wide, extending from the center of the groove to the center of the F′ pocket (Fig. 4). Above the A′ pocket, several residues are in van der Waals contact and form a wide bridge that partially covers the groove (Figs. 4 and 5). H-2K^b, H-2D^b, and HLA-B27 (*5*) each have a small bridge in a similar position that is formed either by interdomain H bonds or a salt bridge [from residues α1 66 (K^b and D^b) or 62 (B27) to α2 163]. Although extensive intradomain interactions form either side of the bridge of

Table 2. Comparison of binding grooves of CD1, MHC class I and II, and FcRn. The binding groove was defined as a set of atoms meeting each of three criteria (*37*): They must have ≥ 0.1 Å^2 solvent accessibility to a 1.4 Å probe, ≤ 1 Å^2 solvent accessibility to a 5 Å probe, and contribute atoms to the contiguous scribed molecular surface between or below the α1 and α2 helical axes and above the β-sheet platform. This definition differs subtly from previous definitions used to define the grooves of FcRn (*22*) and H-2K^b (*5*) in that the groove is considered as a set of atoms rather than a set of residues and only surface that contributes to the contiguous groove is part of the surface area. Row 2 shows the number of residues contributing to the groove molecular surface; row 3 gives the number of carbon atoms that these residues contribute to this surface; row 4 the number of polar atoms; and row 5 the number of charged atoms. Row 6 shows the surface area of the groove (*37*) and row 7 shows the MHC surface area buried (*27*) on ligand binding where applicable. The number of residues, hydrophobicity, and surface area are significantly larger in CD1, whereas the number of polar and charged atoms are fewer.

	CD1d1*	CD1d1†	H-2K^b	H-2D^b	HLA-A2	HLA-B27	H-2M3	FcRn	HLA-DR1
Residues	46	46	29	30	25	33	39	22	34
Carbon atoms	102	108	50	57	46	62	75	29	53
Polar atoms	9	10	11	16	9	11	14	9	13
Charged atoms	1	1	8	5	6	10	4	5	5
Surface area (Å^2)	1310	1390	763	763	747	891	898	509	725
Buried surface (Å^2)‡	N.A.	N.A.	1182	1220	1165	1062	1122	N.A.	1220

*Molecule 1 of the asymmetric unit. †Molecule 2 of the asymmetric unit. ‡The buried surface area (*27*) upon peptide binding. The buried surface area is greater than the groove surface area because it contains an appreciable area that would be excluded from the groove surface by the solvent accessibility criteria used to define the groove surface (*37*). The buried surface can be approximated by adding a 2.8 Å wide skirt to the groove surface. The surface area of this approximation of buried surface for H-2K^b is 1259 Å^2, whereas the surface area for a similar surface for CD1d1 is 1756 Å^2.

CD1, only one interdomain van der Waals contact is actually made across the bridge.

In MHC class I, a cluster of tyrosines in pocket A (Tyr^7, Tyr^{59}, and Tyr^{171}) close the groove and form a conserved hydrogen bonding network with the NH_2-terminus of the bound peptide (*5*). In mouse CD1d1, the A′ pocket wall is made up of entirely hydrophobic residues (Phe^{10}, Trp^{63}, and Phe^{171}) eliminating the possibility of a hydrogen bonding network that might anchor the NH_2-terminus of a bound peptide ligand (Fig. 5). The binding groove of H-2M3 (*5*) is also incapable of anchoring the NH_2-terminus of peptides since the A pocket is nearly closed and separated from the rest of the groove (Fig. 4). Instead, H-2M3 binds shorter, mostly hydrophobic, NH_2-terminal formylated peptides in pockets B through F (*4*, *5*). At the F′ end, the residues that interact with the COOH-terminus of class I peptides and close this end of the groove (Thr^{80}, Tyr^{84}, Thr^{143}, Lys^{146}, and Trp^{147}) are also substituted by mostly hydrophobic residues (Asp^{83}, Met^{87}, Pro^{146}, Val^{149}, and Leu^{150}, respectively) in CD1 (Fig. 5). Although the groove is closed at this end, several of the residues contributing to the COOH-terminal wall are poorly ordered in the structure or have high B values, perhaps indicative of some mobility that might allow for longer ligands to protrude from this end.

The CD1 binding groove has the largest surface area of any MHC antigen presenting molecule whose structure has been determined so far (*37*) (Table 2). Of the 29 residues that form the binding groove in class I H-2K^b (*5*), all of the structurally homologous residues in CD1 are either solvent accessible or part of the groove (with the exception of Leu^{175} (171 in class I), which is completely buried. Despite its size, the CD1 groove has only 10 residues with a capacity to hydrogen bond to bound ligand (Table 2), which is less than the number of polar residues used to bind ligand by any MHC except H-2M3. The only polar atoms of the groove molecular surface are concentrated at the bottom of the A′ pocket (Gln^{14}, Ser^{28}, and Arg^{74}) and at the entrance to the groove (Tyr^{73}, Ser^{76}, Asp^{80}, and Thr^{156}) (Fig. 5). Of the 16 side chains used for hydrogen bond formation with the main chain of bound peptides in class I (*5*), eight residues in homologous positions are polar in CD1, with six of them located in α1. Similarly, for the corresponding nine main chain interacting residues in class II (*23*), five are polar in CD1.

Ligand binding. This limited number of potential hydrogen bond donors and acceptors in CD1 could be sufficient to engage a peptide ligand, as seen in H-2M3 (*5*), where only five residues form hydrogen bonds (*38*) with bound peptide. However, the distribution of these polar atoms (mostly in the α1 domain) is not compatible with peptide binding in a manner analogous to other MHC molecules. The possibility of forming an extensive hydrogen bonding network, concentrated either at both ends of the peptide, as in class I (*5*, *34*), or on both sides of the longitudinal axis of the groove, as in class II (*23*), is not apparent in CD1. More probably, extensive hydrophobic interactions in deeply buried pockets, similar to some of the peptide interactions with H-2M3 (*5*), would be the chief determinant for ligand binding. Nevertheless, a peptide binding motif defined for mCD1d1 from a phage display library and peptide binding analysis (*11*) is highly hydrophobic (F/WxxL/I/MxxW) and would have complementarity to the chemical properties of the putative binding groove. However, it is not apparent from the structure how the remainder of the long peptide antigen might interact in the groove. Small conformational changes, such as rearrangement of side chains forming the bridge over the A′ pocket or of the COOH-terminal end of the groove, could increase accessibility to the groove. Small, but significant, conformational changes that alter the shape of the groove do occur in peptide–class I complexes, but these are mostly confined to the α2 H1 and H2a helices at the COOH-terminus

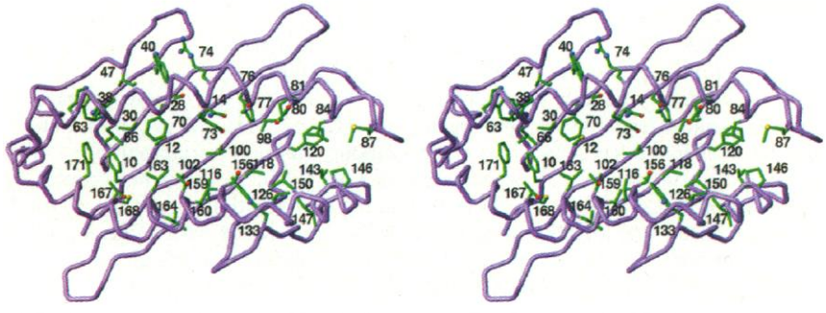

Fig. 5. Residues contributing to the CD1 binding groove. The Cα trace of mCD1d1 α1α2 domains and labeled side chains of residues making up the groove molecular surface (*37*) shown looking into the ligand binding groove. Most of the residues contributing to the groove are hydrophobic. The Cα trace is in purple tubes, carbon side chain atoms in green, nitrogen atoms in blue, oxygen atoms in red, and sulfur atoms in yellow. This figure was generated with MIDAS (*51*).

of the groove (*5*, *34*). It is not known whether any larger conformational changes occur in the transition from empty to ligand-filled forms of MHC molecules (*39*, *40*) because no structures of empty MHCs are available. Some conformational changes in CD1 upon peptide (*29*) and lipid (*17*) binding have been proposed from circular dichroism studies.

Binding of lipid or glycolipid antigens is compatible with the hydrophobic character of the CD1 groove. However, CD1-lipid interactions are likely to be different from the binding of fatty acids by fatty acid binding proteins, which bind lipid through electrostatic interactions between basic residues at the bottom of the binding cavity and the carboxylate group of the fatty acid (*41*). The putative ligand binding groove of mCD1d1 does not have amphipathic character; the groove is mostly of neutral charge with no exposed basic groups on its floor to interact with a buried carboxylate group. Rather CD1 may interact with lipid analogously to nonspecific lipid transport proteins (*42*) (Fig. 4) which bury the hydrophobic tail of their ligand in an electrostatically neutral pocket that is lined with hydrophobic residues. The carbohydrate moieties of the lipoglycans would then most likely be on the outside of the binding groove as there are no specific clusters of hydrophilic and aromatic residues in the groove that are commonly seen in the binding sites of carbohydrate binding proteins (*43*). Some discontinuous electron density is present in both the A′ and F′ pockets of the ligand binding groove for this mCD1d1 structure, for which no specific ligand was added in the protein purification and crystallization. However, strong, continuous, and unbranched density deep in the ligand binding groove has been observed in a second mCD1d1 structure (*44*) and is consistent with the fragmented density observed in the first structure. Interpretation of this density and analysis of the composition of the putative ligand requires further study.

The CD1 family and T cell specificity. CD1 isotypes are essentially nonpolymorphic, which suggests that they may function for presentation of specialized sets of antigens. In accordance with this view, only a few allelic variants of mCD1d1 have been detected in several mice subspecies, with very limited substitutions (three to six changes and only one in the groove) (*45*). The mouse CD1d2 (mCD1d2), has 17 possible substitutions, 6 of which are in the groove, but functional mCD1d2 protein has yet to be detected.

Although detailed predictions are premature, the overall folding of other CD1 molecules is likely to be very similar to mCD1d1. Few insertions or deletions are seen and residues participating in interdomain contacts and interactions between secondary structural elements are significantly more conserved than the rest of the protein (Fig. 5). The CD1-specific features, such as the elevation of the α1 H2 helix, are probably maintained because the sequences of the 50 to 55 turn and β sheet that hold up the helix are mostly conserved. An additional disulfide bond between Cys^{133} and Cys^{147} in hCD1b is predicted from the mCD1d1 structure (Fig. 5).

Conservation of general features of the groove among different isotypes is more difficult to predict. Pocket A′ is probably enclosed in all CD1 molecules, as residues that form the bridge over the A′ pocket are mostly conserved. The bridge may be larger in group I CD1 molecules due to a possible Glu^{62}-Arg^{170}, group I–specific salt bridge. In CD1a, the entrance to the groove may be narrowed by a possible Arg^{73}-Glu^{156} salt bridge and an Arg^{76}-Asn^{153} H bond. Significant differences may be found at the COOH-terminus of the groove, where the residues that partially close pocket F′ are not highly conserved, suggesting that the positions of the α2 H1 and H2a helical segments may vary. As this region corresponds to a less ordered part of the groove structure, its shape and ligand interaction may vary considerably.

While sequence identity among known CD1 molecules is equally distributed between the binding groove and the rest of the α1α2 domain, conservation by chemical homology is higher for groove residues than for the rest of the domains generally. Most of this chemical conservation is present in both group I and group II molecules. However, four residues (76, 84, 118, and 156) in the F′ pocket appear only to be conserved in CD1 group II whereas six other residues have group-specific size and properties (86, 142, and 147 in the F′ pocket and 12, 30, and 74 in the A′ pocket). It is likely then that CD1 molecules from both groups can similarly accommodate predominately hydrophobic ligands, although a few binding site residues may confer isotype specificity to ligand binding.

Residues on the exposed surface of the α1 H2, α2 H1, and α2 H2a helices are among the most variable in the collection of CD1 sequences and may confer T cell specificity for different isotypes. mCD1d1 selects and reacts with a specific NK1.1 T cell population bearing an invariant TCR Vα14,Jα281 chain paired with a limited number of Vβ chains (*19*). A similar population of T cells, presumed to be the first source of IL-4 (*46*), has been found in humans (*47*). It is not known whether ligand binding to CD1 is required for the selection or stimulation of such T cells (or both), but it is possible that CD1d-specific residues may be involved in the selection of this restricted TCR repertoire. In this respect, a combination of CD1d-specific residues, not shared with any other CD1 isotypes, is found at the NH_2-terminus of the α1 H2 helix, pointing up and away from the binding groove.

REFERENCES AND NOTES

1. A. J. McMichael *et al.*, *Eur. J. Immunol.* **9**, 205 (1979); F. Calabi and C. Milstein, *Nature* **323**, 540 (1986); L. H. Martin, F. Calabi, C. Milstein, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9154 (1986).
2. H. G. Rammensee, K. Falk, O. Rotzschke, *Annu. Rev. Immunol.* **11**, 213 (1993).
3. R. N. Germain and D. H. Margulies, *ibid.*, p. 403.
4. C. R. Wang, B. E. Loveland, K. F. Lindahl, *Cell* **66**, 335 (1991); P. H. Gulden *et al.*, *Immunity* **5**, 73 (1996); L. L. Lenz, B. Dere, M. J. Bevan, *ibid.*, p. 63.
5. H-2K^b: D. H. Fremont, M. Matsumura, E. A. Stura, P. A. Peterson, I. A. Wilson, *Science* **257**, 919 (1992); W. Zhang, A. C. Young, M. Imarai, S. G. Nathenson, J. C. Sacchettini, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8403 (1992); D. H. Fremont, E. A. Stura, M. Matsumura, P. A. Peterson, I. A. Wilson, *ibid.* **92**, 2479 (1995). H-2M3: C. R. Wang *et al.*, *Cell* **82**, 655 (1995). H-2D^b: A. C. Young, W. Zhang, J. C. Sacchettini, S. G. Nathenson, *ibid.* **76**, 39 (1994). HLA-A2: P. J. Bjorkman *et al.*, *Nature* **329**, 506 (1987); P. J. Bjorkman *et al.*, *ibid.*, p. 512; D. R. Madden, D. N. Garboczi, D. C. Wiley, *Cell* **75**, 693 (1993); E. J. Collins, D. N. Garboczi, D. C. Wiley, *Nature* **371**, 626 (1994). HLA-B27: D. R. Madden, J. C. Gorga, J. L. Strominger, D. C. Wiley, *ibid.* **353**, 321 (1991). HLA-Aw68: T. P. Garrett, M. A. Saper, P. J. Bjorkman, J. L. Strominger, D. C. Wiley, *ibid.* **342**, 692 (1989); M. L. Silver, H. C. Guo, J. L. Strominger, D. C. Wiley, *ibid.* **360**, 367 (1992); E. J. Collins, D. N. Garboczi, M. N. Karpusas, D. C. Wiley, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1218 (1995). HLA-B53: K. J. Smith *et al.*, *Immunity* **4**, 215 (1996). HLA-B35: K. J. Smith *et al.*, *ibid.*, p. 203.
6. C. Nataraj *et al.*, *Int. Immunol.* **8**, 367 (1996).
7. F. Calabi, J. M. Jarvis, L. Martin, C. Milstein, *Eur. J. Immunol.* **19**, 285 (1989).
8. S. A. Porcelli, *Adv. Immunol.* **59**, 1 (1995).
9. S. P. Balk *et al.*, *Science* **253**, 1411 (1991).
10. S. Porcelli *et al.*, *Nature* **341**, 447 (1989).
11. A. R. Castaño *et al.*, *Science* **269**, 223 (1995).
12. E. M. Beckman *et al.*, *Nature* **372**, 691 (1994); A. Bendelac, *Science* **269**, 185 (1995).
13. P. A. Sieling *et al.*, *Science.*, p. 227.
14. T. I. Prigozy *et al.*, *Immunity* **6**, 187 (1997).
15. E. M. Beckman *et al.*, *J. Immunol.* **157**, 2795 (1996).
16. M. Sugita *et al.*, *Science* **273**, 349 (1996).
17. W. A. Ernst and R. L. Modlin, personal communication.
18. S. Cardell *et al.*, *J. Exp. Med.* **182**, 993 (1995).
19. A. Bendelac *et al.*, *Science* **268**, 863 (1995).
20. P. Dellabona *et al.*, *J. Exp. Med.* **177**, 1763 (1993).
21. S. P. Balk *et al.*, *Science* **265**, 259 (1994).
22. W. P. Burmeister, L. N. Gastinel, N. E. Simister, M. L. Blum, P. J. Bjorkman, *Nature* **372**, 336 (1994); W. P. Burmeister, A. H. Huber, P. J. Bjorkman, *ibid.*, p. 379.
23. HLA-DR1: J. H. Brown *et al.*, *ibid.* **364**, 33 (1993); L. J. Stern *et al.*, *ibid.* **3658**, 215 (1994); T. S. Jardetzky *et al.*, *ibid.*, p. 711. HLA-DR3: P. Ghosh, M. Amaya, E. Mellins, D. C. Wiley, *ibid.* **378**, 457 (1995). I-E^k: D. H. Fremont, W. A. Hendrickson, P. Marrack, J. Kappler, *Science* **272**, 1001 (1996).
24. P. Bork, L. Holm, C. Sander, *J. Mol. Biol.* **242**, 309 (1994).
25. The average rmsd for α1 between CD1 and MHC class I molecules (*5*) is ~1.9 Å for 70 homologous Cα positions, 1.8 Å between CD1 and FcRn (*22*) for 62 Cα positions, and ~4.0 Å between CD1 and class II (*23*) for 62 Cα positions. The average rmsd is ~0.9 Å for α2 between CD1 and FcRn for 62 Cα

positions, ~1.5 Å between CD1 and class I molecules for 72 Cα positions, and ~3.5 Å between CD1 and class II for 83 Cα positions. For the combined α1α2 domain (α1-β1 for HLA-DR1), the rmsd is ~2.6 Å between CD1 and FcRn or class Ia (*5*) molecules for 174 Cα positions, 3.4 Å between CD1 and class Ib H-2M3 (*5*) for 174 Cα positions, and 4 Å between CD1 and class II HLA-DR1 (*23*) for 145 Cα positions. The rmsd is ~2.0 Å for 90 Cα positions in α3 between CD1 and class I molecules or FcRn and 2.7 Å for 90 Cα positions between CD1 in the α3 domain and HLA-DR1 in the homologous β2 domain. The rmsd is 0.66 Å between CD1 and H-2M3 and 1.04 Å between H-2K^b for 99 homologous Cα positions of the β2m domain. The rmsd values were calculated with SUPPOS (*50*) and a structure-based sequence alignment.

26. After aligning all homologous molecules on their β2m domains, the transformation necessary to align the α1α2 domains with CD1 is a 0.4° to 1° rotation and 3.9 to 11.4 Å translation for class I molecules (*5*), a 1.2° rotation and 8.6 Å translation for class II HLA-DR1 (*23*), and 1.0° rotation and 5.3 Å translation for FcRn (*22*). Aligning β2m on α3 requires a 170° rotation and 11.6 Å translation for CD1, 146° to 160° rotation and 12.9 to 13.6 Å translation for class I (*5*), 176° rotation and 13.5 Å translation for class II HLA-DR1 (*23*), and 159° rotation and 12.0 Å translation for FcRn (*22*).
27. The buried surface was determined by a distance calculation performed in GRASP (*52*); all surface vertices on one surface within 3.4 Å of any vertex of another surface is considered part of the buried surface. The total buried surface area between heavy and light chain in CD1 is about 2700 Å^2 and for class I molecules is ~2860 Å^2.
28. The proportion of hydrophilicity and hydrophobicity was calculated by mapping a simple atom-specific atom type hydrophobicity or hydrophilicity parameter on to the molecular surface in GRASP (*52*).
29. A. R. Castaño and P. A. Peterson, unpublished data.
30. M. Teitell *et al.*, *J. Immunol.* **158**, 2143 (1997).
31. R. D. Salter *et al.*, *Nature* **345**, 41 (1990).
32. R. König, L. Y. Huang, R. N. Germain, *ibid.* **356**, 796 (1992).
33. The rmsd for residues 223 to 233 is 1.0 to 1.3 Å between CD1 and class I (*5*), 1.8 Å between CD1 and FcRn (*22*), and 1.8 to 2.1 Å between CD1 and HLA-DR1 (*23*). Most MHC and MHC-like molecules have relatively high *B* values in this region of the molecule, making interpretation difficult. H-2K^b (*5*) has an average B value for all atoms of 26 Å^2 and a *B* value of 68 Å^2 in the CD8 binding loop. HLA-DR1 (*23*) has an average *B* value for all atoms of 34 Å^2 and a *B* value of 44 Å^2 in the CD4 binding loop. CD1 has an average *B* value of 22 Å^2 for all atoms and a *B* value of 26 Å^2 for the homologous loop. FcRn (*23*) has an average *B* value of 20 Å^2 for all protein atoms and a *B* value of 43 Å^2 for the homologous loop. The values for the rmsd's were calculated with SUPPOS (*50*).
34. M. A. Saper, P. J. Bjorkman, D. C. Wiley, *J. Mol. Biol.* **219**, 277 (1991); M. Matsumura, D. H. Fremont, P. A. Peterson, I. A. Wilson, *Science* **257**, 927 (1992).
35. Pockets numbers correspond to the position of the residue in the bound peptide ligand. Residues at positions 5 and 8 point up and out of the groove. Hence, there are no pockets 5 and 8.
36. The reported lengths and widths are measured distances between atom centers. The total length of the groove is a distance between atoms contributing to the groove molecular surface, one from the wall of the A′ pocket and one from the wall of the F′ pocket, at approximately equal elevation from the β-sheet platform. The widths of the pockets are measured between two atoms, one in the α1 domain and one in the α2 domain, that contribute to the molecular surface of the pockets and have nearly equal elevation above the β-sheet platform. The lengths of the individual pockets are distances between atoms in the terminal wall of the pocket and atoms of Leu^{100}. The distances reported are the maximum distances meeting these criteria.
37. The groove molecular surface was defined as the collection of vertices (*52*) from the scribed molecular surface (generated by a 1.4 Å surface probe) between the α1 and α2 domains, which are more than 2.5 Å away from atoms accessible to a 6 Å probe. Groove residues were considered to be any residues that contribute atoms to the groove molecular surface. Atoms were identified as contributing to the groove molecular surface if they were within 0.05 Å of the scribed surface and had more than 0.1 Å^2 accessibility to a 1.4 Å probe. Portions of the groove molecular surface within the van der Waals radius of atoms accessible (>1 Å^2) to a 5 Å probe were deleted from the surface. Surface areas were calculated from scribed molecular surfaces in GRASP (*52*).
38. Van der Waals contacts are assumed between any pair of atoms with their centers closer than 4 Å, and hydrogen bonds were assigned by HBPLUS (*61*).
39. M. R. Jackson, E. S. Song, Y. Yang, P. A. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 12117 (1992).
40. H. G. Ljunggren *et al.*, *Nature* **346**, 476 (1990).
41. J. M. LaLonde, M. A. Levenson, J. J. Roe, D. A. Bernlohr, L. J. Banaszak, *J. Biol. Chem.* **269**, 25339 (1994).
42. D. H. Shin, J. Y. Lee, K. Y. Hwang, K. K. Kim, S. W. Suh, *Structure* **3**, 189 (1995).
43. N. K. Vyas, M. N. Vyas, F. A. Quiocho, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1792 (1983); N. K. Vyas, M. N. Vyas, F. A. Quicho, *J. Biol. Chem.* **266**, 5226 (1991).
44. B. W. Segelke, A. R. Castaño, P. A. Peterson, I. A. Wilson, unpublished data. The second mCD1d1 structure determination was conducted in the presence of a hydrophobic peptide, p99 (*11*).
45. C. R. Wang, personal communication.
46. T. Yoshimoto and W. E. Paul, *J. Exp. Med.* **179**, 1285 (1994).
47. P. Dellabona, E. Padovan, G. Casorati, M. Brockhaus, A. Lanzavecchia, *ibid.* **180**, 1171 (1994).
48. D. J. Bacon and W. F. Anderson, *J. Mol. Graph.* **6**, 219 (1988).
49. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991); E. A. Merrit and M. E. P. Murphy, *Acta Crystallogr. D* **50**, 869 (1994).
50. F. M. D. Villieux, *DEMON program suite Manual* (Grenoble, France, 1993).
51. UCSF MGL, *UCSF MidasPlus User's Manual* (School of Pharmacy, University of California at San Francisco, San Francisco, 1995).
52. B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).
53. Soluble mCD1d1 and β2m were produced in *Drosophila melanogaster* cells (*39*). Oligonucleotides and the polymerase chain reaction were used to engineer mCD1d1 cDNA constructs (*9*) encoding the complete CD1 extracellular domain (280 amino acids) attached to a COOH-terminal $(His)_6$ tag. This construct was cloned into a pRMHa-3 vector and cotransfected with a mouse β2m-pRMHa-3 construct into *D. melanogaster* SC2 cells (*11*). Stable transfectants of CD1 and β2m were established, and secreted CD1-β2m complexes were isolated at room temperature by affinity chromatography on a Ni-NTA column (Qiagen) and purified by fast-performance liquid chromatography (FPLC) Mono Q (Pharmacia) and Mono S (Pharmacia) columns, or gel filtration, to a single band on an SDS gel and concentrated to 5 to 8 mg/ml.
54. B. W. Matthews, *J. Mol. Biol.* **33**, 491 (1968).
55. A. J. Howard *et al.*, *J. Appl. Crystallogr.* **20**, 383 (1987).
56. Collaborative Computational Project, Number 4, *Acta Crystallogr. D* **50**, 760 (1994).
57. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
58. A 10σ peak in the native Patterson map at (0.05, 0.5, 0.5) corresponds to 4σ peak in the kappa = 180° section of the self-rotation function, indicative of a non-crystallographic pseudo twofold axis passing close to (0.0, 0.0, 0.25) and parallel to the b axis.
59. A. T. Brünger, *X-PLOR*, version 3.1 (Yale Univ. Press, New Haven, CT, ed. 3.1, 1992).
60. T. A. Jones, in *Computational Chemistry* D. Sayre, Ed. (Clarendon Press, Oxford, 1982), pp. 303–317.
61. I. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994).
62. We thank Dr. P. J. Bjorkman for the prereleased FcRn coordinates and helpful discussions and Dr. T. Yeates for helpful discussions on molecular replacement. Supported in part by grants from the National Institutes of Health, CA-58896 (I.A.W.) and the Ministerio de Educacion y Ciencia of Spain (A.R.C.). This is publication MB-10266 from the Scripps Research Institute. The coordinates for mCD1d1 have been deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, NY) with accession code 1CD1.

19 September 1996; accepted 6 May 1997