Supporting Information

Garcia-Alles et al. 10.1073/pnas.1105627108

SI Materials and Methods

Reagents and Cell Lines. Recombinant HLA-A2 in complex with the melan A peptide (ELAGIGILTV) was purchased from the Recombinant Protein Platform (Institut National de la Santé et de la Recherche Médicale). Bovine brain sulfatides (cerebroside sulfates, sulfatides) was purchased from Sigma and the tri-palmitoylated lipopeptide Pam_3CSK_4 from InvivoGen. All other commercial lipids were from Avanti Polar Lipids. *Mycobacterium tuberculosis* diacylated sulfoglycolipids were purified according to published protocols (1, 2), and the synthetic sulfoglycolipid (SGL) analogs SGL1 and SGL12 were prepared as previously described (3). To prepare lipid vesicles, single lipids or mixtures of lipids in chloroform/MeOH were dried under a N_2 flow in glass tubes, and the residue was rehydrated with periodic agitation for 1 h at room temperature in the water volume required to obtain 0.5- to 5-mM stock solutions, which were then submitted to 10 freeze-thaw cycles and 10 min of bath sonication.

Recombinant soluble CD1e (including the *N*-terminal propeptide) and CD1b were expressed in, respectively, S2 *Drosophila* cells and mouse J558 cells and purified as described previously (4, 5). Recombinant soluble CD1a was expressed exactly like rsCD1b and purified by affinity chromatography using an agarose-conjugated antibody recognizing the BirA-peptide (Avidity). Saposins A–C were prepared in *Escherichia coli* and purified according to a published protocol (6).

Lipid Binding to CD1 Molecules. Briefly, 6- to 10-μg aliquots of rsCD1e-2, scCD1e, or rsCD1b at 5–20 μM final concentration were shaken in glass vials at 800 rpm and 30 °C for the time indicated (typically, 30–120 min) in the presence of liposomes containing the specified single lipid or mixture of lipids (100–200 μM) in 50 mM sodium acetate (pH 5.0) containing 50 mM NaCl, 1 mM DTT, and 1 mM EDTA. Sodium acetate was replaced by Hepes for incubations at neutral pH. Experiments with rsCD1b were also carried out in the presence of CD1e or protein controls at 20 μM final concentration (or as noted in the legends for Fig. 3 and Fig. S6). After the indicated time, the samples were cooled on ice and 2-μL aliquots were applied to an isoelectric focusing (IEF) gel. Electrophoresis was performed in a PhastGel system (Amersham Biosciences) for 600 AVh (accumulated volt hours). Proteins in the gels were detected by staining with Coomassie R 350. The gels were scanned on a GS-800 calibrated densitometer (Bio-Rad), and protein bands were quantified using the Quantity One 4.6.3 software program (Bio-Rad) after subtraction of the background.

Native Nano Electrospray Ionization (ESI) Mass Spectrometry. Before MS experiments, a sample of fully deglycosylated scCD1e was buffer-exchanged extensively against 50 mM aqueous ammonium acetate (pH 6.7) using 5-kDa centrifuge filters (Millipore). Other details concerning the instrumentation and procedure are exactly as described in ref. 5.

Structure Determination, Analysis, and Presentation. A crystal was immersed for 1 min in 0.4 μL of crystallization solution supplemented with 20 mM Na acetate and 20% ethyleneglycol before being flash-frozen in liquid nitrogen. Diffracted intensities were collected from this single crystal at the European Synchrotron Radiation Facility (Grenoble, France) on the beamline ID14 EH4, tuned to 0.9295 Å, using an ADSC CCD detector. The diffraction images were processed using MOSFLM (7), scaled with SCALA (8), and further processed with the CCP4 program package (9). Data processing statistics are summarized in Table S1. The crystals belonged to the monoclinic space group *C2* (a = 206.29 Å, b = 45.93 Å, c = 65.63 Å, β = 91.48°) and contained a single β2-microglobulin (β2m)-CD1e molecule per asymmetric unit.

The structure was solved by the molecular replacement method using the program PHASER (10), feeding independently the two polypeptide chains from a CD1e structure built on a homology model (4). A single unambiguous solution was obtained with a *Z*-score of 15.6 for data between 51.6 and 3.5 Å resolution. This solution was refined with REFMAC (11) using data between 51.6 and 2.90 Å resolution. The F_o–F_c electron density maps allowed us to model a Fuc(α1–6)-GlcNAc group onto the heavy chain Asn16. Translation libration screw-motion (TLS) parameters were refined for each polypeptidic chain. The final model includes 375 residues, 2 sugar moieties, and 1 glycerol. Statistics of the final refined structure are given in Table S1. Ramachandran analysis showed 245 residues in preferred regions (77%), 61 in allowed regions (19%), 7 residues in generously allowed regions (2%), and 5 outliers (1.6%). The protein stereochemistry was validated using PROCHECK (12).

Molecular surface areas and volumes were calculated using the CASTp server (13). A solvent probe of 1.7 Å was used to map the internal cavities and pockets of the proteins. Fig. 1 and Figs. S3 and S4 were generated with The PyMOL Molecular Graphics System (2002) (DeLano Scientific) and Hollow (14).

1. Gilleron M, et al. (2001) Acylation state of the phosphatidylinositol mannosides from Mycobacterium bovis bacillus Calmette Guerin and ability to induce granuloma and recruit natural killer T cells. *J Biol Chem* 276:34896–34904.
2. Gilleron M, et al. (2004) Diacylated sulfoglycolipids are novel mycobacterial antigens stimulating CD1-restricted T cells during infection with Mycobacterium tuberculosis. *J Exp Med* 199:649–659.
3. Guiard J, et al. (2008) Synthesis of diacylated trehalose sulfates: Candidates for a tuberculosis vaccine. *Angew Chem Int Ed Engl* 47:9734–9738.
4. de la Salle H, et al. (2005) Assistance of microbial glycolipid antigen processing by CD1e. *Science* 310:1321–1324.
5. Garcia-Alles LF, et al. (2006) Endogenous phosphatidylcholine and a long spacer ligand stabilize the lipid-binding groove of CD1b. *EMBO J* 25:3684–3692.
6. Qi X, Grabowski GA (2001) Differential membrane interactions of saposins A and C: Implications for the functional specificity. *J Biol Chem* 276:27010–27017.
7. Leslie AGW (1987) Computational aspects of protein crystal data analysis. *Proceedings of the Daresbury Study Weekend*, eds Helliwell JR, Machin PA, Papiz MZ (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK), pp 39–50.
8. Evans PR (1993) Data collection and processing. *Proceedings of the CCP4 Study Weekend* (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK), pp 114–122.
9. Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
10. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458–464.
11. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
13. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci* 7:1884–1897.
14. Ho BK, Gruswitz F (2008) HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* 8:49.

Fig. S1. Structure of di- and triacylated lipids. *Bis*-(monooleoylglycero)phosphate (*S,S*) (BMP), sulfatides (SLF), sphingomyelin (SM), synthetic analogs of diacylated sulfoglycolipids from *M. tuberculosis* (SGL12 and SGL1), phosphatidylinositol (PI), *sn*-(3-oleoyl-2-hydroxy)-glycerol-1-phospho-*sn*-3′-(1′,2′-dioleoyl)-glycerol (*S,R*) (HemiBMP), *N*-palmitoyl-S-[2,3-bis(palmitoyloxy)-(2*RS*)-propyl]-[*R*]-cysteinyl-[*S*]-seryl-[*S*]-lysyl-[*S*]-lysyl-[*S*]-lysyl-[*S*]-lysine (Pam_3CSK_4), phosphatidylcholine (PC), phosphatidylserine (PS).

Fig. S2. CD1e binds di- and triacylated lipids. IEF gels of rsCD1e-2 (first row) expressed in *Drosophila* S2 cells and scCD1e (second row) produced in human M10 cells after incubation at pH 7.5 in the absence (first lane) or presence of a 10-fold molar excess of the indicated lipids (structures shown in Fig. S1). Ac_2SGL refers to diacylated sulfoglycolipids from *M. tuberculosis*.

Fig. S3. Structure of CD1e. (*A*) Global fold of scCD1e shown as a ribbon diagram with the heavy chain colored in green and β2m in blue. The Fuc(α1–6)-Glc*N*Ac glycan at Asn16 and glycerol moiety modeled between the α3 heavy chain domain and the β2m are shown as sticks with carbon, oxygen, and nitrogen atoms colored in yellow, red, and blue, respectively. The emplacement of the nonmodeled loop comprising residues Q120–I122 is indicated by the arrow. (*B*) Top stereoview of the A′ pocket of the scCD1e groove. The electron density of protein backbone atoms (green mesh) is from the final 2*FoFc* map contoured at 2σ. The same map was contoured at 1σ and colored in blue for the side-chain atoms of the residues lining the groove. Residue numbers are indicated. Side-chain carbon atoms are colored according to their localization at the A′ pocket (green), F′ pocket (gray), or the interface between A′ and F′ pockets (light blue). Oxygen, nitrogen, and sulfur atoms are in red, blue, and yellow, respectively. Side chains of residues that could not be modeled due to insufficient electron density are included in the picture, e.g., Phe54. (*C*) Top stereoview of the F′ pocket of the scCD1e groove. Other details are as in *B*.

Fig. S4. Comparison of the groove architecture of CD1e and other CD1 isoforms. Top (*Left* column) and side (*Center* column) views of α1–α2 domains, with the transparent molecular surface of grooves rendered in light green (scCD1e), gray (hCD1b, PDB ID 2H26), light blue (hCD1d, 1ZT4), yellow (hCD1a, 1ONQ), and light magenta (hCD1c, 3OV6). The profile of the backbone atoms is colored in yellow with the exception of the α-helical portions, which are in red. Phosphatidylcholine (PC) associated with CD1b, α-galactosyl-ceramide bound to CD1d, sulfatides bound to CD1a, and MPM bound to CD1c are shown as sticks with carbon, oxygen, nitrogen, and phosphorus/sulfur atoms colored in yellow, red, blue, and orange, respectively. Unknown ligand (UL) and C_{12} ligands of CD1b and CD1c, respectively, are colored in orange. The side-chain atoms of residues differing among CD1 isotypes and causing notable changes in groove architecture are shown as green sticks, with the corresponding Cα depicted as balls. (*Right* column) Comparison of the main portals of grooves of human CD1 proteins. Molecular surfaces are presented colored as in the accompanying panels for each molecule, with the exception of portal-lining residues that appear in orange for all CD1. Side-chain atoms from these residues are shown as sticks, with carbon atoms in black, oxygen atoms in red, and nitrogen atoms in blue. The maximal aperture of the portal in the direction that runs parallel to the helices is indicated for comparison. In the case of human CD1c (*Bottom Right*), the side chain of Glu80 splits the cleft in two. Two distances are therefore indicated to suggest the possibility that such a side chain could move away.

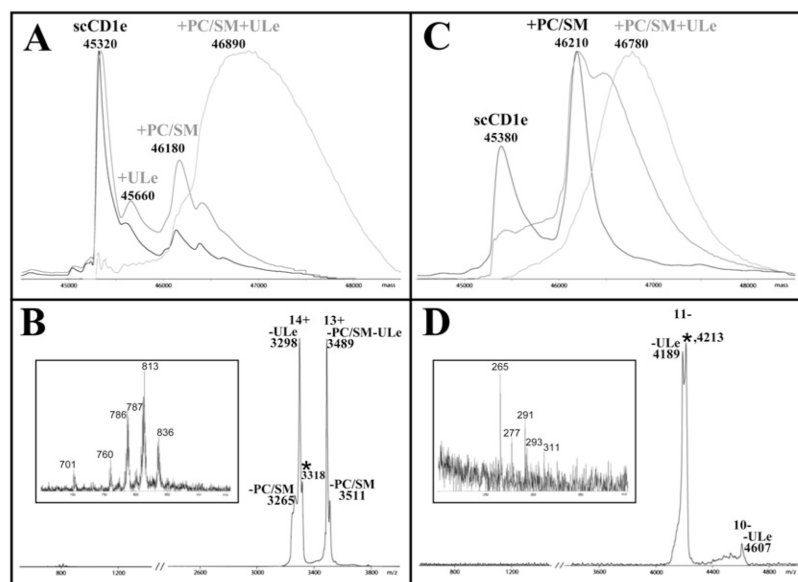

Fig. S5. Endogenous ligands associated with scCD1e. (*A*) Convoluted "neutralized" mass spectrum obtained by averaging over the charged states detected in positive-ion mode native ESI-MS of fully deglycosylated scCD1e. Masses originating from apo-scCD1e, scCD1e-ULe, scCD1e-phosphatidylcholine (PC)/sphingomyelin (SM), and scCD1e-PC/SM-ULe are labeled. The first spectrum was recorded using high desolvation energy (black line) and the two other spectra at progressively decreasing energies (dark gray and light gray lines, respectively). Decreasing the desolvation energy prevents ligand dissociation but causes a broadening of the signals and displacement of the peaks toward higher-molecular-weight values due to the heterogeneous association of water and other small molecules. (*B*) Tandem MS after selection of 14^+ precursor ions at 3,318 *m/z* corresponding to a presumed scCD1e-PC/SM-ULe species (asterisk). Products appearing inside the 3,200- to 3,600-*m/z* interval originate from the dissociation of either PC or SM (−PC/SM); the loss of 280 Da, which might correspond to a spacer molecule (−ULe); or the simultaneous loss of PC/SM and ULe. Signals corresponding to PC (*m/z* 760.6, 786.6, 812.6, 836.6) and SM (701.7, 787.7, 813.7), but not to ULe, were detected in the low-mass range of this spectrum. (*Inset*) The enlarged 650- to 950-*m/z* region. (*C*) Convoluted "neutralized" mass spectrum obtained by averaging over the charged states detected in negative-ion mode in native ESI-MS of fully deglycosylated scCD1e. Other details are as for *A*. (*D*) Tandem MS after selecting 11^- precursor ions at 4,213 *m/z*, which are attributed to scCD1e-PC/SM-ULe species (asterisk). The observed dissociation products correspond to loss of neutral or negatively charged species of 260–270 Da. (*Inset*) An enlarged 200- to 400-*m/z* region of a second spectrum recorded at higher collision energy. Low-intensity peaks are detected, which could correspond to fatty acids: *m/z* 265 ($C_{17:2}$), 277 ($C_{18:3}$), 291 ($C_{19:3}$), 293 ($C_{19:2}$), and 311 ($C_{20:0}$).

Fig. S6. (*A*) Lipid binding to CD1e is not affected by pH. ScCD1e was incubated with 10-fold molar excess of *bis*-(monoacylglycero)phosphate (BMP) at pH 7.5 (□) or 5.0 (●) for the time indicated before IEF separation and quantification by densitometry scanning. (*B*) CD1–lipid complexes are stable in the absence of external lipids. IEF gel of purified CD1–lipid complexes after incubation at pH 5.0 for the time and at the temperature indicated. (*C*) CD1e removes SGL12 bound to CD1b. Purified rsCD1b-SGL12 was incubated with and without (negative control) indicated proteins under the conditions of Fig. 3*A*. In experiments with β2m, saposin A or C, and BSA, a 15-, 3-, 10- and 2-fold molar excess was added, respectively. Percentages of loaded complex were quantified and normalized against negative controls (i.e., 100% rsCD1b-SGL12) to remove the contribution of spontaneous dissociation. (*D*) Selectivity of lipid editing by CD1e. Purified rsCD1b-SGL12 (□) or rsCD1b-SGL1 (■) complexes (7 μM final concentration) were incubated with the indicated amounts of rsCD1e-2 under the conditions of Fig. 3*A* before separation by IEF and quantification. The percentage of loaded complex remaining after incubation is plotted. (*E*) CD1e transfers lipids from vesicles to CD1b. IEF analysis of the products after incubation of rsCD1b (20 μM) with the specified lipids (150 μM) for 30 min at pH 5.0 and 37 °C in the presence or absence of rsCD1e-2 (14 μM). (*F*) Substoichiometric amounts of CD1e promote formation of CD1b–BMP. Mixtures containing rsCD1b (16 μM), the amount of rsCD1e-2 indicated on the *x* axis, and BMP (160 μM) were incubated at pH 7.5 (□) or 5.0 (●) for 1 h, 37 °C, before IEF separation and quantification. Data are representative of at least three independent identical or close experiments.

Table S1. Data collection and refinement statistics

Data collection	
Space group	*C*2
Cell dimensions	
a, *b*, *c* (Å)	206.29, 45.93, 65.63
α, β, γ (°)	90.0, 91.48, 90.0
Resolution (Å)	103.11–2.90 (3.06–2.90)
R_{sym}	0.063 (0.372)
I/σ	7.0 (1.7)
Completeness (%)	94.7 (96.7)
Redundancy	2.0 (2.0)
Refinement	
Resolution (Å)	51.55–2.90
No. reflections	12,460
R_{work}/R_{free}	0.2408/0.2955
No. atoms	2898
Protein	2868
Glycan	24
Glycerol	6
Average B-factors	
From Wilson statistics	79.9
Protein	87.7
Glycan	83.7
Glycerol	66.4
rmsd	
Bond lengths (Å)	0.017
Bond angles (°)	1.832

Values in parentheses are for the highest-resolution shell.