Crystal Structure of a Full-Length Human Tetraspanin Reveals a Cholesterol-Binding Pocket

Graphical Abstract

Highlights

- The structure of full-length CD81 reveals a cone-like architecture
- A large intramembrane cavity exists within the transmembrane region
- Cholesterol binds CD81 within the intramembrane cavity
- MD simulations suggest that CD81 can adopt open and closed conformations

Authors

Brandon Zimmerman, Brendan Kelly, Brian J. McMillan, Tom C.M. Seegar, Ron O. Dror, Andrew C. Kruse, Stephen C. Blacklow

Correspondence

andrew_kruse@hms.harvard.edu (A.C.K.), stephen_blacklow@hms.harvard.edu (S.C.B.)

In Brief

The tetraspanin CD81 contains a large intramembrane cavity occupied by cholesterol, indicating the potential for functional modulation by small molecules.
Crystal Structure of a Full-Length Human Tetraspanin Reveals a Cholesterol-Binding Pocket

Brandon Zimmerman,1,2 Brendan Kelly,3 Brian J. McMillan,1,2 Tom C.M. Seegar,1,2 Ron O. Dror,3 Andrew C. Kruse,1,* and Stephen C. Blacklow1,2,4,5,*

1Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
2Department of Cancer Biology, Dana Farber Cancer Institute, Boston, MA 02215, USA
3Departments of Computer Science and of Molecular and Cellular Physiology and Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA
4Department of Pathology, Brigham and Women’s Hospital, Boston, MA 02115, USA
5Lead Contact
*Correspondence: andrew_kruse@hms.harvard.edu (A.C.K.), stephen_blacklow@hms.harvard.edu (S.C.B.)
http://dx.doi.org/10.1016/j.cell.2016.09.056

SUMMARY

Tetraspanins comprise a diverse family of four-pass transmembrane proteins that play critical roles in the immune, reproductive, genitourinary, and auditory systems. Despite their pervasive roles in human physiology, little is known about the structure of tetraspanins or the molecular mechanisms underlying their various functions. Here, we report the crystal structure of human CD81, a full-length tetraspanin. The transmembrane segments of CD81 pack as two largely separated pairs of helices, capped by the large extracellular loop (EC2) at the outer membrane leaflet. The two pairs of helices converge at the inner leaflet to create an intramembrane pocket with additional electron density corresponding to a bound cholesterol molecule within the cavity. Molecular dynamics simulations identify an additional conformation in which EC2 separates substantially from the transmembrane domain. Cholesterol binding appears to modulate CD81 activity in cells, suggesting a potential mechanism for regulation of tetraspanin function.

INTRODUCTION

Tetraspanins comprise a large family of four-pass transmembrane proteins arising evolutionarily in protists (Huang et al., 2005). In early unicellular eukaryotes, tetraspanins are believed to play a role in the dynamic regulation of membrane morphology, an activity thought to be subsequently co-opted for cell-cell interactions. As a result, there is an evolutionary link between the emergence of tetraspanins and the development of multicellularity (Huang et al., 2005). The development of specialized cell-cell interactions and new cell types likely selected for the duplication and differentiation of tetraspanins that evolved to effect specific functions. Thus, primitive fungi have a single tetraspanin, while the size of the tetraspanin family has grown to ten representatives in the sea urchin Strongylocentrotus purpuratus, 17 in the tunicate Ciona intestinalis, and 33 in Homo sapiens (Garcia-Espana et al., 2008).

Despite their common ancestry, the vertebrate tetraspanins have acquired a variety of discrete and important biological functions, which highlight their critical, though underappreciated, role in mammalian physiology (Hemler, 2005). Studies of tetraspanin knockouts in mice and other organisms have identified essential functions for tetraspanins in the immune, reproductive, genitourinary, and auditory systems. For instance, mice lacking CD151/tspan24 have abnormalities in hemostasis and lymphocytes that respond abnormally to mitogenic stimulation (Le Naour et al., 2000; Wright et al., 2004). Patients with frameshift mutations in CD151 present with a similar bleeding disorder, along with hereditary nephritis, deafness, and epidermolysis bullosa (Karamatic Crew et al., 2004). Additionally, CD63/tspan30 knockout mice exhibit altered water homeostasis with increased urinary flow and water intake (Schroeder et al., 2009), and CD9/tspan29 null mice are sterile due to failure of sperm-egg fusion (Le Naour et al., 2000).

The biological effects of many tetraspanins appear to be attributable to their activities as modulators of central signal transduction pathways. There is accumulating evidence that the six tetraspanins in the C8 subclass regulate Notch signaling by promoting ADAM10 trafficking and enzymatic maturation in both flies and mammals (Dornier et al., 2012; Haining et al., 2012; Jouannet et al., 2016; Noy et al., 2016). A number of other tetraspanins appear to exert their effects by influencing integrin signaling, either indirectly, by modulating responses of integrins to their ligands (van Spriel et al., 2012; Wee et al., 2015), or by directly and stably associating with particular integrin heterodimers (Yauch et al., 1998). Moreover, knockout of CD37 in mice predisposes them to development of B cell lymphoma, a phenotype attributed to unrestrained interleukin 6 (IL-6) signaling in the absence of CD37 (de Winde et al., 2016).

CD81/TAPA-1 (target of antiproliferative antibody 1)/tspan28 was first identified as the target of an antibody discovered in a screen for the inhibition of proliferation in a lymphoma cell line (Oren et al., 1990). CD81, together with ME491/CD63 and CD37, became the founding members of the tetraspanin (TM4SF) protein family in mammals (Hemler, 2005; Huang et al., 2005). CD81 is among the most well characterized tetraspanins because of its essential role in B cell biology.
(Cherukuri et al., 2004b; Mattila et al., 2013), forming a multi-protein complex with CD19, CD21/CR2, and CD225 to regulate B cell receptor function. The importance of CD81 in B cell function is further highlighted by the recent report of a common variable immunodeficiency patient who has a germline mutation that produces an alternatively spliced, truncated form of CD81 (van Zelm et al., 2010), which appears to sequester CD19 intracellularly (Vences-Catalan et al., 2015). CD81 is also a host factor that interacts with the hepatitis C virus (HCV) E2 envelope protein and is required for efficient HCV entry (Bradbury et al., 1992; Pileri et al., 1998).

Tetraspanins are the largest family of transmembrane proteins in mammals for which detailed structural information about the intramembrane domain remains unavailable. This lack of information, together with a limited understanding about the structural relationship between the intramembrane domain and the extracellular region of the protein, has greatly hindered efforts to understand the molecular mechanisms of action of these proteins. The tetraspanins are predicted to contain a number of shared structural features, including intracellular N and C termini, small (EC1) and large (EC2) extracellular loops, four transmembrane regions (TMs), and a short intracellular loop between TM2 and TM3 (Yanez-Mo et al., 2009). All tetraspanins also possess intracellular cysteine residues, which are typically palmitoylated (Yanez-Mo et al., 2009). These cysteines and their palmitoylation are required for efficient interaction with certain associated proteins and the formation of tetraspanin microdomains known as the tetraspanin web (Hopf et al., 2012), which is thought to be generated by tetraspanin-tetraspanin interactions as well as by heterologous interactions with other associated proteins in “signaling hubs” (Levy and Shoham, 2005). Although structures have been reported for the extracellular EC2 regions of Schistosoma mansoni TSP-2 and human CD81, they provide limited insight into other family members because of the poor conservation of the EC2 region, and they offer no information about the intramembrane portion of the protein, which is the most highly conserved part of the molecule (Stipp et al., 2003).

We report here the structure of human CD81 (PDB: 5TCX) as a representative example of a full-length tetraspanin. The protein contains a bound cholesterol molecule in a large intramembrane pocket between two largely independent pairs of transmembrane helices capped by the EC2 domain, and specific cholesterol binding to this site is observed in vitro. The EC2 domain appears to more readily adopt an “open” conformation in molecular dynamics simulations when cholesterol is not present in its binding site, consistent with the modulation of tetraspanin function by mutations that compromise cholesterol binding. Together, the structural, computational, and biochemical studies suggest a model for tetraspanin function, as well as a route to modulating tetraspanin activity as a therapeutic strategy in a variety of different diseases.

**RESULTS**

**Crystal Structure of CD81**

To define the overall architecture of an intact tetraspanin, elucidate interdomain relationships between the EC2 and the intramembrane region, and gain insight into tetraspanin-ligand interactions and the tetraspanin web, we purified full-length CD81 from insect cells (Figures S1A and S1B), crystallized the protein using the lipidic cubic phase method, and determined its structure. As is common for integral membrane protein crystals, the diffraction pattern was anisotropic, with strong scattering along two reciprocal space axes and weaker scattering along the third. Prior to refinement, we performed ellipsoidal truncation with resolution limits of 5.5 Å along the a* axis, 2.95 Å along the b* axis, and 2.95 Å along the c* axis (Table S1). To obtain phase information, we utilized a fragment-based iterative molecular replacement approach (Kruse et al., 2013). Briefly, we first located EC2 of CD81 by molecular replacement using a previously reported structure of the CD81 EC2 region (PDB: 1IV5, chain A) as a search model (Kitadokoro et al., 2002) (Figure S1C). After placing this domain, we used four more rounds of iterative molecular replacement using a polyalanine alpha helix search model to locate the four alpha helices comprising the transmembrane domain. This model was then used as a starting point for model building and refinement (Figure S2).

The overall structure of CD81 resembles a waffle cone in which the EC2 domain covers an intramembrane cavity bounded by the four transmembrane helices (Figure 1A). No electron density is visible for the small extracellular loop (EC1), suggesting this region is disordered. The overall fold of the four transmembrane helices does not resemble that of any other integral membrane protein of known structure. The transmembrane region consists of two largely separated pairs of antiparallel helices: one pair comprises TM1/TM2 and the other TM3/TM4. The two pairs of helices only converge close to the cytoplasmic side of the membrane through contacts between TM2 and TM3. The central cavity bounded by the four transmembrane helices and the bottom face of EC2 encloses a total volume of 3,300 Å³ (Figure 1B).

**Evolutionary Conservation of Tetraspanin Structure**

To confirm the relative positions of the transmembrane helices and assess our assignments of the residues in this region, we performed an evolutionary coupling analysis, which compares homologous sequences to determine amino acid residues that are correlated through evolution (Marks et al., 2012). The correlation map for CD81 indicates that strong evolutionary couplings occur between helices one and two, between helices three and four, and within the EC2 domain, as predicted by the X-ray structure. We extended this analysis to additional tetraspanin proteins dispersed throughout phylogeny to examine whether this unusual fold is evolutionarily conserved, or a specific feature of CD81 (Figure 2). CD81 (Figure 2A), hypothetical protein FGSG_08695 from the fungus Fusarium graminearum PH-1 (Figure 2B), tetraspan 3A from the fruit fly Drosophila melanogaster (Figure 2C), and AX4 tetraspanin family protein from the eukaryotic slime mold Dictostelium discoideum (Figure 2D) all exhibit the same pattern of evolutionary coupling, providing support for the inference that the fold seen in the CD81 structure is conserved among all tetraspanins.

A more detailed analysis of sequence conservation patterns for CD81 homologs across species (Figures 3A and 3A) and for tetraspanin paralogs within humans (Figures 3B and 3B) highlights the tight evolutionary constraints within the transmembrane...
region and the greater variability of the second extracellular loop by comparison. Strikingly, highly conserved residues within EC2 are located on extracellular helices one and two at their points of contact with the junction between EC1 and the first two transmembrane segments, accounting for the “closed” conformation of the EC2 cap. Contacts between the EC2 and the transmembrane region include hydrophobic interactions between L35 of the EC2 cap. Contacts between the EC2 and the transmembrane segments, accounting for the “closed” conformation of the EC2 cap. We observe unexpected additional Fc-Fc electrostatic interactions within this pocket. On the basis of the shape of the electron density within this pocket, we tentatively identified the additional density as a bound cholesterol molecule (Figures 4 and S4, related to Figure 4). Within 4 Å of the bound cholesterol are a number of hydrophobic and aromatic residues, including F21 of TM1; I64, V68, V71, M72, and V75 of TM2; F94, L98, and L101 of TM3; and V212 and M216 of TM4 (Figure 4A). N18 and E219, two polar residues belonging to TM1 and TM4, respectively, form hydrogen bonds to the cholesterol hydroxyl group. Analysis of the CD81 sequences from all 37 available homologs reveals that these 13 amino acids are nearly 100% conserved (Figure S4). Comparison of the CD81 sequence with its 32 human paralogs shows N18 is conserved in 27 of the 33 human tetraspanins. However, E219 is only present in CD81 and tspan10, with the majority of mammalian tetraspanins having a polar glutamate or glutamine residue (82%) on the preceding turn of the helix and a glycine residue (64%) at this position instead.

**Analysis of Intramembrane Binding Pocket**

The most striking feature of the CD81 structure is the large, hydrophobic pocket bounded by the four transmembrane helices and the EC2 cap. We observe unexpected additional Fc-Fc electrostatic interactions within this pocket. On the basis of the shape of the electron density within this pocket, we tentatively identified the additional density as a bound cholesterol molecule (Figures 4 and S4, related to Figure 4). Within 4 Å of the bound cholesterol are a number of hydrophobic and aromatic residues, including F21 of TM1; I64, V68, V71, M72, and V75 of TM2; F94, L98, and L101 of TM3; and V212 and M216 of TM4 (Figure 4A). N18 and E219, two polar residues belonging to TM1 and TM4, respectively, form hydrogen bonds to the cholesterol hydroxyl group. Analysis of the CD81 sequences from all 37 available homologs reveals that these 13 amino acids are nearly 100% conserved (Figure S4). Comparison of the CD81 sequence with its 32 human paralogs shows N18 is conserved in 27 of the 33 human tetraspanins. However, E219 is only present in CD81 and tspan10, with the majority of mammalian tetraspanins having a polar glutamate or glutamine residue (82%) on the preceding turn of the helix and a glycine residue (64%) at this position instead.

**CD81 Binds Specifically to Cholesterol**

To determine directly whether CD81 specifically binds cholesterol, we assessed the cholesterol-binding ability of wild-type CD81 immunopurified from HEK293T cells using a radioactive binding assay and compared it with the Beta-lactam binding receptor BlaR, a negative control four-pass transmembrane protein prepared similarly (Figure 4B). Wild-type CD81 immunoprecipitates recover ~15-fold more cholesterol than immunoprecipitates from untransfected control cells or from cells expressing BlaR. To determine whether cholesterol binding by CD81 relies on specific interactions seen in the structure of CD81, we mutated residue E219 to either alanine or glutamine, as it is a critical polar contact with the cholesterol hydroxyl in our structure (the hydrogen bond between E219 and cholesterol is formed 90% of the time in the simulations described in the following section). The E219A and E219Q point mutants of CD81 recover ~50% less cholesterol than wild-type (Figure 4C), whereas a G26F/G30F double mutant on the external face of helix one does not detectably affect cholesterol recovery (Figure S4). Finally, to address whether CD81 and the transmembrane pocket are capable of binding other lipids, we explored the binding of radioactive estradiol and palmitate using the same radioligand-binding assay. The recovery of both estradiol and palmitate is unaffected by the E219Q mutation (though it does appear that the amount of bound estradiol is reduced somewhat upon introduction of the E219Q mutation, suggesting that it might have very weak affinity for the cholesterol-binding pocket, consistent with the fact that it is a cholesterol derivative).

**Molecular Dynamics Simulations Identify an Open Conformation of EC2**

We performed molecular dynamics simulations of CD81 in a hydrated lipid bilayer, both with and without cholesterol bound in the intramembrane pocket. In three of nine simulations of the apoprotein (i.e., with cholesterol removed), EC2 transitioned to
an open conformation in which it disengaged from TM1 and TM2 (Figure 5A). EC2 remained in this open conformation for the remainder of these three simulations. After opening, EC2 is flexible and dynamic relative to the TM domain (Figure 5C); however, the same fully open conformation is observed in all three simulations that display opening.

The opening motion involves a substantial straightening of TM3 and TM4. During EC2 opening, a salt bridge between D196 on EC2 and K201 on TM4 that stabilizes the closed state breaks, leading to extension and straightening of TM4. A newly formed salt bridge between K116 and D117 stabilizes the extended form of TM3 observed in the open state (Figure 5B).

Interestingly, cholesterol-bound simulations consistently maintained a closed conformation, similar to the crystal structure, in which EC2 remained in contact with TM1 and TM2 (Figure 5A). In two of the nine simulations that we initiated with cholesterol bound, however, cholesterol dissociated from the binding pocket into the membrane, exiting through the gap between TM1 and TM4. In one of these simulations, EC2 transitioned to the fully open conformation after cholesterol dissociated. EC2 remained in that conformation for the rest of the simulation. The fact that we only observed opening when cholesterol was absent from the binding pocket suggests that the presence of cholesterol may stabilize the closed conformation and that the absence of cholesterol may favor opening (p = 0.03; see STAR Methods).

TM1 and TM2 undergo substantial motion in simulations both in the presence and in the absence of cholesterol (Figure S5). The intracellular end of TM1 remains in close contact with TM4 in
simulations of the apoprotein, stabilized by a hydrogen bond between the side chains of N18 and E219. When cholesterol is bound, it competes for interaction with E219, often leading the intracellular end of TM1 to separate from TM4.

**Cholesterol Binding Regulates CD81 Function**

Numerous reports have highlighted the importance of CD81 in CD19 export to the cell surface, and a CD81 truncation in a human patient results in a combined variable immunodeficiency phenotype and intracellular retention of CD19 in cell-based assays (van Zelm et al., 2010; Vences-Catalán et al., 2015). Here, we used a flow cytometry assay to measure the amount of CD19 at the surface of transfected cells and the ability of CD81 to increase the amount of CD19 detected at the cell surface. In the absence of added CD81, 293T cells transfected with FLAG-tagged CD19 show a minimal increase in surface staining compared to untransfected cells. Upon co-transfection of wild-type CD81, we see a 10-fold increase in surface staining compared to CD19 with control vector (Figure 6A). Moreover, when cholesterol binding is compromised by either the E219A or E219Q mutation, the amount of CD19 surface staining increases further by an additional 50% when comparable amounts of CD81 protein are expressed and present at the cell surface, whereas no effect is seen when the G26F/G30F mutant is tested (Figure 6B).

**DISCUSSION**

We have solved the crystal structure of a full-length human tetraspanin and defined an unknown intramembrane binding pocket for tetraspanins. The discovery of the pocket with cholesterol bound was then used to design experiments for interrogation of cholesterol binding and its potential role in tetraspanin function.

Our structure of CD81 revealed a monomeric form of a tetraspanin, which contrasts with the dimeric structure of the isolated EC2 fragment. The putative dimerization interface seen in the structure of the isolated EC2 fragment is located on its bottom face in the full-length protein within 3.5 Å of the TM1/TM2 bundle, indicating that the dimer is likely a non-native interaction driven by lattice packing effects in the absence of the transmembrane regions of the protein.

Early modeling work proposed that the four TM helices would form a tightly associated four-helix bundle (Seigneuret, 2006), yet our structure reveals the transmembrane region to fold as two largely separated pairs of antiparallel helices, a conclusion supported by evolutionary coupling analysis (Figure 2). The overall similarity seen among diverse family members argues strongly that all tetraspanin proteins possess this transmembrane fold. The intramembrane pocket seen in the structure between TM1/4 and TM2/3 is likely to be accessible only through lateral diffusion within the membrane plane, because entrance from the extracellular space is precluded by the presence of EC2 above the pocket.

Much attention has been previously given to the concept of a tetraspanin web, where tetraspanins form homooligomers or heterooligomers with other tetraspanins to form higher-order complexes and protein-rich microdomains in the cell membrane (Charrin et al., 2003a; Horváth et al., 1998; Levy and Shoham, 2005; Rubinstein et al., 2013). Recently, the existence of the tetraspanin web has come under more scrutiny, as early experiments done with detergents did not effectively disrupt secondary interactions (Dornier et al., 2012; Zuidscherwoude et al., 2015). In our CD81 crystals, individual monomers in the lattice pack such that adjacent subunits in the membrane plane lie in antiparallel orientations (Figure S6), an arrangement that is necessarily non-physiological. Though the absence of lateral homotypic packing interactions among adjacent subunits does not in itself exclude the possibility of higher-order CD81 complexes or the...
potential for assembly of a tetraspanin web, there is also evidence among the uroplakin tetraspanins that tetraspanin monomers do not interact with one another. Uroplakins Ia and Ib, in complex with their accessory proteins uroplakin II and IIIa, form hexameric lattices. Despite the close proximity of the six tetraspanin molecules in the lattice, the lattice is entirely bridged by interactions between the non-tetraspanin partners, uroplakin II and IIIa (Min et al., 2006). A recent study using super-resolution microscopy also shows that tetraspanins, including CD81, lie in closer proximity to their non-tetraspanin binding partners than to other tetraspanins in the membrane, again suggesting that tetraspanins need not be constitutively oligomeric in the cellular milieu.

Figure 4. CD81 Binds Cholesterol within Its Intramembrane Cavity

(A) CD81-cholesterol interactions. CD81 residues within 4 Å of the bound cholesterol molecule are rendered as sticks and labeled in the zoomed in view (right). An Fo-Fc omit map of electron density contoured at 2.0 σ is shown for the bound cholesterol. N18 and E219 coordinate the cholesterol hydroxyl group. Views of the pocket in surface representation are shown in open-book form projecting onto the TM1/2 bundle (left) and the TM3/4 bundle (right). CD81 residues at the ligand interface are colored orange.

(B) CD81 wild-type, CD81 mutant proteins (E219A, E219Q), and BlaR were immunopurified from HEK293T cells. Proteins captured on FLAG beads were used for radioactive cholesterol binding experiments. WCL, whole-cell lysate; Sol, solubilized protein; IP, immunoprecipitation.

(C) Cholesterol binding by immunopurified proteins. 1,2-3H-cholesterol was incubated with immunopurified FLAG-CD81 or FLAG-BlaR from 293T cells and bound cholesterol was measured. The figure represents three independent experiments performed in duplicate. Statistical analysis was performed using ANOVA, and a Bonferroni post hoc test was performed comparing all columns. **p < 0.01; ***p < 0.001.

(D) Specificity of lipid binding to CD81. Immunopurified FLAG-CD81 from 293T cells prepared as in (C) was incubated with 1,2-3H-cholesterol (C), 2,4,6,7-3H(N)-estradiol (E), and 9,10-3H(N)-palmitic acid (P). Bound 3H-lipid was measured in a scintillation counter. The figure represents three independent experiments performed in duplicate. See also Figure S4.
comes from a study showing that cholesterol in the host cell highlighting the central role of cholesterol to tetraspanin biology (rin et al., 2003b). Perhaps the most compelling evidence proximity of cholesterol and tetraspanins within the membrane (Charrin et al., 2004). Early work demonstrated a physical link between cholesterol and CD81 as well as CD9 and CD82 using photoactivatable cholesterol crosslinking, demonstrating the close proximity of cholesterol and tetraspanins within the membrane (Charrin et al., 2003b). Perhaps the most compelling evidence highlighting the central role of cholesterol to tetraspanin biology comes from a study showing that cholesterol in the host cell membrane was required for both maintenance of a CD81 monoclonal antibody epitope and for CD81-dependent infection by Plasmodium falciparum sporozoites, but not by sporozoites that were CD81 independent (Silvie et al., 2006). Our results reveal that CD81 binds cholesterol and that E219 is an important residue in the binding pocket, offering a potential mechanism for how tetraspanins might detect cholesterol or other membrane lipids (whereas CD81 and tspan10 both possess a polar residue in this position, most other tetraspanins have a polar residue one helical turn earlier, with other cholesterol-interacting residues highly conserved throughout evolution).

Our molecular dynamics simulations suggest that CD81 may exist in both closed and open conformations and that the equilibrium between closed and open conformations may be shifted toward the closed state when cholesterol is bound. Mutations that interfere with cholesterol binding by CD81 result in enhanced delivery of its binding partner (CD19) to the cell surface. On the basis of these observations, we speculate that the closed state, favored when cholesterol is bound, disfavors partner binding and, hence, protein export (Figure 6C).

See also Figure S5.

Figure 5. Molecular Dynamics Simulations Reveal an Open Conformation in which EC2 Separates from the Transmembrane Domain

(A) Closed (blue) and open (red) states of CD81, as observed in cholesterol-bound and apoprotein simulations, respectively. The open conformation is characterized by substantial domain separation and straightening of TM helices 3 and 4.

(B) The salt bridge from EC2 to TM4 (D196 – K201) stabilizes the closed conformation and breaks during opening, while a different salt bridge (K116 – D117) formed upon opening helps stabilize the open conformation.

(C) Interdomain distance (measured between alpha carbons of F58 and F126) as a function of time for an apoprotein simulation in which the domains separate and a cholesterol-bound simulation in which they do not. Thin traces show values every 1 ns, and thick traces are smoothed.
by loading and unloading cargo in response to variation in lipid composition among different membrane compartments. Accumulating evidence that tetraspanins such as CD81 have growth-promoting activity in certain human cancers has led to their emergence as potential therapeutic targets (Hemler, 2014). Indeed, monoclonal antibodies directed at the tetraspanin26/CD37 have moved into phase 1 or phase 2 clinical trials for the treatment of chronic lymphocytic leukemia and non-Hodgkin lymphoma (Deckert et al., 2013; Zhao et al., 2007). The unexpected discovery of the intramembrane ligand-binding pocket represents a potential targetable site for modulating tetraspanin function using small molecules and provides a route forward for elucidation of the molecular mechanism of action of this enigmatic class of important but poorly understood proteins.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR RESOURCE AND REAGENT SHARING

Figure 6. Cholesterol Binding Regulates CD81-Mediated Export of CD19
(A and B) HEK293T cells were transfected with cDNA encoding the indicated proteins, and cell-surface amounts of CD19 (A) and CD81 (B) were assessed by flow cytometry. Histograms shown represent four independent experiments done in triplicate. Statistical analysis was performed using ANOVA, and a Bonferroni post hoc test was performed comparing all columns. ***p < 0.01.
(C) Proposed model for modulation of cargo binding in response to cholesterol. CD81 favors a closed conformation when cholesterol is bound (left) and more readily accesses an open conformation, which allows more efficient export of its cargo CD19 (modeled in cartoon form), when not bound to cholesterol (right).
● EXPERIMENTAL MODEL AND SUBJECT DETAILS
● METHOD DETAILS
  o Expression and Purification
  o Crystallography and Data Collection
  o Phasing and Structure Refinement
  o Evolutionary Coupling Analysis
  o Sequence Conservation Analysis
  o Radioactive Lipid Binding Assay
  o Molecular Dynamics Simulation: System Setup
  o MD Simulation Protocol
  o MD Simulation Analysis
  o Flow Cytometry Assay
● QUANTIFICATION AND STATISTICAL ANALYSIS
● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, six figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.09.056.

AUTHOR CONTRIBUTIONS
The overall project was designed and developed by B.Z., T.C.M.S., A.C.K., and S.C.B. Molecular cloning, protein purification, and crystallization experiments were conducted by B.Z. Data collection was performed by B.Z. and A.C.K. Data processing, phase calculation, and structure refinement were carried out by B.Z., B.J.M., and A.C.K. B.K. and R.O.D. performed and analyzed molecular dynamics simulations. The manuscript was written by B.Z. and S.C.B. with input from B.K., R.O.D., and A.C.K.

ACKNOWLEDGMENTS
We would like to thank beamline staff at APS GM/CA beamline 23ID-B for their superb technical assistance and support and N. Latorraca for helpful suggestions regarding simulation analysis. Financial support for this work was provided by NIH grants NCI 5 RO1 CA092433 (S.C.B.) and TIPS 0D021345 (A.C.K.) and by a Terman Faculty Fellowship (R.O.D.). B.Z. was supported by a CIHR post-doctoral fellowship, and T.C.M.S. was supported by training grant 2T32 HL007627.

Received: May 2, 2016
Revised: July 12, 2016
Accepted: September 29, 2016
Published: October 27, 2016

REFERENCES

Cell 167, 1041–1051, November 3, 2016 1049
Genomics
Chen, S., et al. (2005). The phylogenetic analysis of tetraspanins projects the
Huang, S., Yuan, S., Dong, M., Su, J., Yu, C., Shen, Y., Xie, X., Yu, Y., Yu, X.,
Chem.
Huang, J., and MacKerell, A.D., Jr. (2013). CHARMM36 all-atom additive pro-
Jouannet, S., Saint-Pol, J., Fernandez, L., Nguyen, V., Charrin, S., Boucheix, C.,
differentially regulate the cleavage of ADAM10 substrates, Notch activation
Brady, R.L., Daniels, G., and Anstee, D.J. (2004). CD161, the first member of
the tetraspanin (TM4) superfamily detected on erythrocytes, is essential for
the correct assembly of human basement membranes in kidney and skin.
Blood 104, 2217–2223.
Kitatokoro, K., Ponassi, M., Galli, G., Petracca, R., Falugi, F., Grandi, G., and
Bolognesi, M. (2002). Subunit association and conformational flexibility in
the correct assembly of human basement membranes in kidney and skin.
Biocl 104, 2177–2223.
Klaidokoro, K., Ponassi, M., Galli, G., Petracca, R., Falugi, F., Grandi, G., and
Bolognesi, M. (2002). Subunit association and conformational flexibility in
the head subdomain of human CD81 large extracellular loop. Biol. Chem.
383, 1447–1452.
Klauda, J.B., Venable, R.M., Freites, J.A., O’Connor, J.W., Tobias, D.J., Mon-
dragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., Jr., and Pastor, R.W.
(2010). Update of the CHARMM all-atom additive force field for lipids: valida-
Kreise, A.C., Manglik, A., Koblika, B.K., and Weiss, W.I. (2013). Applications of
Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pepko, T., and
compromise—A mixed precision model for GPU accelerated molecular dy-
MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D.,
Mattila, P.K., Feke, C., Depol, D., Terebor, B., Montaner, B., Ottopby, K.L.,
and tetraspanin networks organize receptor nanoclusters to regulate B cell
spanin functions as revealed by the cryo-EM structure of uropaklin complexes
Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano,
M., and Siltz, P. (2013). Collaboration gets the most out of software. eLife 2,
e01456.


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexa Fluor 488 anti-human CD19 Antibody</td>
<td>BioLegend</td>
<td>302219; RRID: AB_389313</td>
</tr>
<tr>
<td>APC anti-human CD81 (TAPA-1) Antibody</td>
<td>BioLegend</td>
<td>349510; RRID: AB_2564021</td>
</tr>
<tr>
<td>ANTI-FLAG M2 Affinity Gel</td>
<td>Sigma Aldrich</td>
<td>A2220; RRID: AB_10063035</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2-3H]-cholesterol</td>
<td>Perkin Elmer</td>
<td>NET139250UC</td>
</tr>
<tr>
<td>2,4,6,7-3H(N)-estradiol</td>
<td>Perkin Elmer</td>
<td>NET013250UC</td>
</tr>
<tr>
<td>9,10-3H(N)-palmitic acid</td>
<td>Perkin Elmer</td>
<td>NET043001MC</td>
</tr>
<tr>
<td>Benzonase recombinant nuclease</td>
<td>Sigma Aldrich</td>
<td>E1014</td>
</tr>
<tr>
<td>Iodoacetamide</td>
<td>Sigma Aldrich</td>
<td>I1149</td>
</tr>
<tr>
<td>In-Fusion HD Cloning Plus</td>
<td>Clontech</td>
<td>638911</td>
</tr>
<tr>
<td>ESF 921 Insect Cell Culture Medium, Protein-Free</td>
<td>Expression systems</td>
<td>96-001</td>
</tr>
<tr>
<td>BestBac 2.0, v-cath/chia Deleted Linearized Baculovirus DNA</td>
<td>Expression systems</td>
<td>91-002</td>
</tr>
<tr>
<td>Fugene HD</td>
<td>Promega</td>
<td>E2311</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>n-Dodecyl-β-D-Maltoside (Anagrade)</td>
<td>Anatrace</td>
<td>D310</td>
</tr>
<tr>
<td>5-Choesten-3β-ol Hemisuccinate</td>
<td>Steraloids</td>
<td>C6823-000</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Invitrogen</td>
<td>11668019</td>
</tr>
</tbody>
</table>

### Deposited Data

<table>
<thead>
<tr>
<th>Deposited Data</th>
<th>Source</th>
<th>PDB:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD81 Structure</td>
<td>This paper</td>
<td>5TCX</td>
</tr>
</tbody>
</table>

### Experimental Models: Cell Lines

<table>
<thead>
<tr>
<th>Experimental Models: Cell Lines</th>
<th>Source</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>293T (ATCC CRL-3216)</td>
<td>ATCC</td>
<td>CRL-3216</td>
</tr>
<tr>
<td>Spodoptera frugiperda ovarian tissue (SF9)</td>
<td>Expression systems</td>
<td>94-001S</td>
</tr>
</tbody>
</table>

### Recombinant DNA

<table>
<thead>
<tr>
<th>Recombinant DNA</th>
<th>Source</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>pVL1392</td>
<td>Expression systems</td>
<td>91-030</td>
</tr>
<tr>
<td>pCDNA3.1 (+) hygro</td>
<td>Invitrogen</td>
<td>V87020</td>
</tr>
<tr>
<td>gBlocks</td>
<td>Integrated DNA Technologies</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Software and Algorithms

<table>
<thead>
<tr>
<th>Software and Algorithms</th>
<th>Source</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amber14</td>
<td>Case et al., 2014</td>
<td><a href="http://ambermd.org/">http://ambermd.org/</a></td>
</tr>
<tr>
<td>GraphPad Prism 7.0</td>
<td>N/A</td>
<td><a href="http://www.graphpad.com/scientific-software/prism/">http://www.graphpad.com/scientific-software/prism/</a></td>
</tr>
<tr>
<td>EVFold conservation server</td>
<td>Hopf et al., 2012</td>
<td><a href="http://evfold.org/evfold-web/evfold.do">http://evfold.org/evfold-web/evfold.do</a></td>
</tr>
<tr>
<td>BD Accuri C6 software</td>
<td>BD Accuri C6</td>
<td><a href="https://www.bdbiosciences.com/instruments/accuri/features/software.jsp">https://www.bdbiosciences.com/instruments/accuri/features/software.jsp</a></td>
</tr>
<tr>
<td>UCLA-DOE LAB — Diffraction Anisotropy Server</td>
<td>Strong et al., 2006</td>
<td><a href="https://services.mbi.ucla.edu/anisoscale/">https://services.mbi.ucla.edu/anisoscale/</a></td>
</tr>
<tr>
<td>SBGrid Consortium</td>
<td>Morin et al., 2013</td>
<td><a href="https://sbgrid.org/software/">https://sbgrid.org/software/</a></td>
</tr>
<tr>
<td>XDS</td>
<td>Kabsch, 2010</td>
<td><a href="https://sbgrid.org/software/">https://sbgrid.org/software/</a></td>
</tr>
<tr>
<td>Phenix (1.10_2155)</td>
<td>Afonine et al., 2012</td>
<td><a href="https://sbgrid.org/software/">https://sbgrid.org/software/</a></td>
</tr>
<tr>
<td>Coot</td>
<td>Emsley and Cowtan, 2004</td>
<td><a href="https://sbgrid.org/software/">https://sbgrid.org/software/</a></td>
</tr>
<tr>
<td>Pymol</td>
<td>Schrödinger Team, 2010</td>
<td><a href="https://sbgrid.org/software/">https://sbgrid.org/software/</a></td>
</tr>
</tbody>
</table>
CONTACT FOR RESOURCE AND REAGENT SHARING

For additional information about reagents and resources, contact the Lead Contact, Stephen Blacklow, at Stephen_blacklow@hms.harvard.edu.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

CD81 protein for crystallographic studies was expressed in Spodoptera frugiperda ovarian tissue (SF9) cells grown in grown in ESF 921 Insect Cell Culture Medium, Protein-Free media at 27°C. Protein was harvested 42 hr after baculovirus infection and isolated from the SF9 membrane fraction.

Radioactivity experiments were performed with protein isolated from HEK293T cells. HEK293T cells were grown in DMEM media supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. HEK293T cells were also used in the flow cytometry assays.

METHOD DETAILS

Expression and Purification

The sequence encoding full-length human CD81 was assembled using synthetic gene blocks (gBlocks, Integrated DNA Technologies) and inserted into the baculovirus transfer vector pVL1392 with an amino-terminal FLAG epitope tag followed by a 3C protease cleavage site. Four intracellular cysteine residues at positions 6, 9, 227 and 228 were mutated to serine to prevent disulfide cross-linking during purification. This protein was expressed in SF9 insect cells using the BestBac system (Expression Systems) according to the manufacturer’s instructions with Fugene HD as the transfection reagent. Infection was performed when cells reached a density of 4 x 10^6 cells/ml, and flasks were then shaken at 27°C for 42 hr prior to harvest. Cells were harvested by centrifugation and frozen at −80°C until purification. After frozen cell paste was thawed, cells were lysed by osmotic shock in 20 mM HEPES (pH 7.4), 2 mM magnesium chloride, 2 mg/ml iodoacetamide (Sigma Aldrich) and 1:100,000 (v:v) benzonase nuclease (Sigma Aldrich). Lysed cells were centrifuged at 18,000 RPM in a Sorvall RC 5C Plus centrifuge with an SS-34 rotor for 15 min. CD81 was then extracted from the pellet using a glass dounce tissue grinder to homogenize lysed cells in a solubilization buffer consisting of 250 mM NaCl, 20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 1% (w/v) n-Dodecyl-β-D-Maltoside (DDM - Anagrade; Anatrace), 0.1% (w/v) cholesterol hemisuccinate (CHS; Steraloids) and 2 mg/ml iodoacetamide. Samples were stirred for 2 hr at 4°C, then centrifuged at 20,000 RPM for 20 min. Supernatant containing solubilized protein was filtered through a glass microfiber prefiter and loaded by gravity flow onto 3 mL of M2 anti-FLAG antibody affinity resin (Sigma Aldrich). The resin was then washed with 50 mL of buffer containing 100 mM NaCl, 20 mM HEPES (pH 7.4), 1% glycerol, 0.1% DDM, and 0.01% CHS. Bound protein was eluted in the same buffer supplemented with 0.2 mg/mL FLAG peptide. 3C protease was added (1:100 w:w) and incubated with CD81 at 4°C overnight. CD81 was further purified by size exclusion chromatography (SEC) on a Sephadex S200 column (GE Healthcare) in buffer containing 0.1% DDM, 0.01% CHS, 100 mM NaCl, and 20 mM HEPES (pH 7.4). CD81 was biochemically pure but consistently eluted as two broad peaks during SEC. The SEC-purified protein was concentrated to 30 - 40 mg/mL and flash frozen with liquid nitrogen in aliquots of 8 µL. Samples were stored at −80°C until use for crystallography. Purity and monodispersity of crystallographic samples was evaluated by SDS-PAGE and analytical SEC, respectively (Figure S1).

Crystallography and Data Collection

Purified CD81 was reconstituted into lipidic cubic phase by mixing with a pre-made 10:1 (w:w) mix of monoolein (Hampton Research) with cholesterol (Sigma Aldrich) at a ratio of 1.5:1.0 lipid:protein by mass, using the coupled syringe reconstitution method (Caffrey and Cherezov, 2009). All samples were mixed at least 100 times. The resulting phase was dispensed in 50 mL of buffer containing 100 mM NaCl, 20 mM HEPES (pH 7.4), 1% glycerol, 0.1% DDM, and 0.01% CHS. Bound protein was eluted in the same buffer supplemented with 0.2 mg/mL FLAG peptide. 3C protease was added (1:100 w:w) and incubated with CD81 at 4°C overnight. CD81 was further purified by size exclusion chromatography (SEC) on a Sephadex S200 column (GE Healthcare) in buffer containing 0.1% DDM, 0.01% CHS, 100 mM NaCl, and 20 mM HEPES (pH 7.4). CD81 was biochemically pure but consistently eluted as two broad peaks during SEC. The SEC-purified protein was concentrated to 30 - 40 mg/mL and flash frozen with liquid nitrogen in aliquots of 8 µL. Samples were stored at −80°C until use for crystallography. Purity and monodispersity of crystallographic samples was evaluated by SDS-PAGE and analytical SEC, respectively (Figure S1).

Data collection was performed at Advanced Photon Source GM/CA beamline 23ID-B. An initial grid raster with 80 x 30 µm beam dimensions was performed, followed by a sub-raster using a 20 µm beam to locate crystals in the loop. Additional rastering was performed using a 10 µm beam diameter to optimally position crystals for data collection. Data collection used a 10 µm beam and diffraction images were collected in 1 degree oscillations at a wavelength of 1.033 Å. The final dataset for CD81 was compiled by merging data from five crystals using XDS (Kabsch, 2010). Data are summarized in Table S1, using the format of Harrison and colleagues (Corbett et al., 2010).

Visual inspection of diffraction frames revealed strong anisotropy, with weaker diffraction in the reciprocal a* axis and stronger diffraction in the other two directions. We thus performed ellipsoidal truncation on the merged dataset using the UCLA anisotropy diffraction server (Strong et al., 2006). Resolution limits along reciprocal space axes were chosen based on an
F/sigF > 2 criterion, giving resolution limits of 5.50 Å, 2.95 Å and 2.95 Å along the reciprocal space a*, b* and c* axes, respectively. Anisotropic B scaling was also applied using the server to compensate for differences in intensity along each reciprocal space axis.

Phasing and Structure Refinement
To obtain phases, molecular replacement was performed in Phenix with Phaser using chain A from a CD81 ectodomain X-ray structure, PDB ID 1IV5 (Kitadokoro et al., 2002) (Figure S1C). A subsequent molecular replacement search was conducted using a polyalanine helix as a search model, with the ectodomain included as a fixed partial model. This procedure, which located one transmembrane helix, was performed three more times, in order to locate all four transmembrane helices. For refinement, reciprocal space optimization of XYZ coordinates and individual atomic B-factor parameters was performed with standard Phenix restraints in phenix.refine (Afonine et al., 2012); optimization for X-ray/stereochemistry weight and X-ray/ADP weight was also performed. Extensive additional model building was carried out manually in Coot (Emsley and Cowtan, 2004) and refined following each round of model building. Near the end of refinement, custom geometric restraints were placed on the two disulfide bonds present in the EC2. As a control for register assignment, the structure was built and register assigned using two independent approaches. First, the model was manually built and register assigned by inspection of electron density in combination with consideration of evolutionary conservation of residues interacting within the transmembrane domains (Marks et al., 2012). In parallel, sequence register was independently assigned automatically with phenix.autobuild, which resulted in the same register assignment. Representative composite OMIT map density is shown in Figure S2. Cholesterol was manually placed into an Fo-Fc difference map. Following refinement, structure quality was assessed using MolProbity (Chen et al., 2010), and figures were prepared in PyMOL (Schrödinger Team, 2010). All crystallographic data processing, refinement, and analysis software was compiled and supported by the SBGrid Consortium (Morin et al., 2013).

Evolutionary Coupling Analysis
For evolutionary coupling analysis, the EVcouplings option on http://evfold.org was used (Marks et al., 2012). The sequence of the indicated tetraspanins was used as a template and default search settings were used, except that alpha helical TM domain was set to “Yes.” The top 90 evolutionary coupling pairs were chosen for display.

Sequence Conservation Analysis
For conservation analysis, multiple sequence alignment aligning human CD81 to its most similar 150 homologs was analyzed using the Consurf server (Ashkenazy et al., 2010; Landau et al., 2005). The multiple sequence alignment was generated using a protein sequence BLAST on the NCBI database. For conservation analysis among the 33 human tetraspanins, the sequences were aligned using Clustal Omega and then analyzed using the Consurf server (Sievers et al., 2011).

Radioactive Lipid Binding Assay

[1,2-3H]cholesterol (45 Ci/mmol), 2,4,6,7-3H(N)-estradiol (94 Ci/mmol), and 9,10-3H(N)-palmitic acid (30 Ci/mmol) were purchased from PerkinElmer Life Sciences. FLAG-tagged CD81 was subcloned into the mammalian expression vector pcDNA3.1 using NheI and BamHI restriction sites and mutations were made using gBlocks from IDT. FLAG-Blar was constructed using PCR, and subcloned into the same restriction vector pcDNA3.1 using the same restriction sites. HEK293T cells were transfected with empty vector, FLAG-tagged CD81 (wild-type, E219A, E219Q), or FLAG-tagged Blar. Cells were lysed 36 hr later with 20 mM HEPES (pH 7.4), 2 mM MgCl2, 1 μl benzozane and 2 mg/ml iodoacetamide. Lysed cells were harvested by centrifugation at 15,000 RPM for 10 min at 4°C. Pellets were resuspended in 1% DDM, 0.1% CHS, 20 mM HEPES (pH 7.4) and 250 mM NaCl and solubilized for 2 hr. The samples were clarified by centrifugation at 15,000 RPM for 10 min at 4°C. The soluble fraction was then incubated with 15 μl of anti-M2 FLAG beads per sample for 1 hr. Beads were recovered and washed three times with 0.1% DDM, 0.01% CHS, 20 mM HEPES (pH 7.4) and 100 mM NaCl. FLAG-CD81 conjugated beads were then incubated with 1 μCi of 3H-lipid for 1 hr. Beads were washed three times used a spin filter column and remaining 3H-lipid was counted using a Beckman Coulter LS6500 scintillation counter.

Molecular Dynamics Simulation: System Setup
Simulations of CD81 were based on the cholesterol-bound crystal structure described in this manuscript. The receptor was simulated in two distinct conditions: (1) the cholesterol-bound crystal structure; (2) the same structure with cholesterol removed.

Prime (Schrödinger) was used to model in missing side chains and the missing residues E86, S87 and 41-54 of EC1. The crystalized protein construct with cysteine mutations at S6, S9, S227 and S228 was employed during simulation. S-(2-amino-2-oxoethyl)cysteines 80 and 89 were returned to natural cysteines. Hydrogen atoms were added, and protein chain termini were capped with the neutral groups acetyl and methylamide. Titratable residues were left in their dominant protonation state at pH 7.0. An alternative side-chain orientation of K11 was used to prevent early formation of potentially artifactual interactions with E219 that was observed in initial simulations. For apoprotein simulation conditions, cholesterol was removed.

The prepared protein structures were aligned on the transmembrane helices to the z-axis, and internal waters added with Dowser (Hermans et al., 2003; Zhang and Hermans, 1996) (in addition to internal waters resolved in the crystal structure). The structures were then inserted into a pre-equilibrated palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer, and solvated with 0.15 M NaCl in explicitly
represented water, then neutralized by removing chloride ions. Final system dimensions were approximately 80 × 70 × 96 Å³, including about 124 lipids, 29 sodium ions, 25 chloride ions, and 10600 water molecules.

**MD Simulation Protocol**

We used the CHARMM36 parameter set for protein molecules, lipid molecules, cholesterol, and salt ions, and the CHARMM TIP3P model for water; protein parameters incorporated CMAP terms (Best et al., 2012a, 2012b; Huang and MacKerell, 2013; Klauda et al., 2010; MacKerell et al., 1998).

Simulations were performed on GPUs using the CUDA version of PMEMD (Particle Mesh Ewald Molecular Dynamics) in Amber14 (Case, 2014; Le Grand et al., 2013; Salomon-Ferrer et al., 2013).

Prepared systems were minimized, then equilibrated as follows: The system was heated using the Langevin thermostat from 0 to 100 K in the NVT ensemble over 12.5 ps with harmonic restraints of 10.0 kcal·mol⁻¹ Å⁻² on the non-hydrogen atoms of lipid, protein, and ligand, and initial velocities sampled from the Boltzmann distribution. The system was then heated to 310 K over 125 ps in the NPT ensemble with semi-isotropic pressure coupling and a pressure of one bar. Further equilibration was performed at 310 K with harmonic restraints on the protein and ligand starting at 5.0 kcal·mol⁻¹ Å⁻² and reduced by 1.0 kcal·mol⁻¹ Å⁻² in a stepwise fashion every 2 ns, for a total of 10 ns of additional restrained equilibration.

We performed nine simulations of cholesterol-bound CD81, and another nine simulations of unliganded CD81, for a total of 18 simulations. These simulations were conducted in the NPT ensemble at 310 K and 1 bar, using a Langevin thermostat and Monte Carlo barostat. In each of these simulations, we performed 5 ns of unrestrained equilibration followed by a production run of 1.0–2.5 μs. Simulations initiated with cholesterol bound totaled 13.7 μs, and simulations initiated with cholesterol removed totaled 14.8 μs.

Simulations used periodic boundary conditions, and a time step of 2.5 fs. Bond lengths to hydrogen atoms were constrained using SHAKE. Non-bonded interactions were cut off at 9.0 Å, and long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method with an Ewald coefficient β of approximately 0.31 Å and B-spline interpolation of order 4. The FFT grid size was chosen such that the width of a grid cell was approximately 1 Å.

**MD Simulation Analysis**

Trajectory snapshots were saved every 100 ps during production simulations. Trajectory analysis was performed using VMD (Humphrey et al., 1996) and CPPTRAJ (Roe and Cheatham, 2013), and visualization was performed using VMD and PyMol (Schrödinger Team, 2010). Trajectories were aligned to the crystal structure, along TM3 and TM4.

To determine how frequently the hydrogen bond between the hydroxyl group of cholesterol and the carboxylate side-chain of E219 was formed in simulation, we calculated the percentage of time the distance between the cholesterol hydroxyl hydrogen and either carboxylate oxygen of E219 was 3.3 Å or less during simulations.

To calculate a p-value associated with the alternative hypothesis that the absence of cholesterol in the binding pocket favors opening, we proceed as follows. We define the p-value as the probability that we would have observed data that favors the alternative hypothesis at least as strongly as the data we observed under the null hypothesis that opening is equally likely when simulating CD81 with or without cholesterol in the binding pocket. Given that we observed no opening events when cholesterol was bound and four opening events when it was not, we calculate the probability, under the null hypothesis, of observing no opening events when cholesterol was bound and four or more opening events when it was not. Because we do not have an a priori estimate of the probability of opening in a given simulation, we maximize the calculated p-value over all possible opening probabilities—in other words, we report the largest of a family of possible p-values.

More specifically, we assume that, under the null hypothesis, the number of opening events observed in t μs of simulation of closed-state CD81 is given by a Poisson distribution with mean q, where q is a proportionality or rate constant. The value of q is unknown but assumed to be the same in simulations with or without cholesterol bound. We calculate, as a function of q, the probability of observing no opening events when cholesterol is bound (in 11.04 μs of simulation of closed, cholesterol-bound CD81) and four or more opening events when cholesterol is not bound (in 12.55 μs of simulation of closed CD81 with no cholesterol bound, including the relevant portion of simulations in which cholesterol dissociated). Our calculated p value, 0.027, is the maximum value of this probability over all possible values of q (p = 0.03; see STAR Methods).

**Flow Cytometry Assay**

HEK293T cells were transfected with empty vector, Flag-tagged CD19 or ProtC-tagged CD81 (wild-type, E219A, E219Q). Cells were washed with PBS 36 hr post-transfection and resuspended with 20mM HEPES (pH 7.4), 150mM NaCl, and 0.1% BSA. Cells were incubated for 15min with Alexa 488-anti-CD19 (Molecular probes) and APC-anti-CD81 (BioLegend) per manufacturers recommendation. Cells were washed and plated into a 96 well U-bottom dish and sorted on a BD Accuri C6 flow cytometer. Experiments were done in triplicate or quadruplicate at least 4 independent times.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Bar graphs display mean ± SD. p values were calculated by one-way ANOVA followed by post hoc Bonferroni tests where applicable using GraphPad Prism. All radioactivity experiments were performed three independent times with duplicate measurements.
(Figure 4C). Flow cytometry experiments were performed four independent times with triplicate measurements (Figures 6A and 6B). Statistical analyses relevant to the structural model are included in Table S1, related to Figure 1.

DATA AND SOFTWARE AVAILABILITY

The structure reported in the paper is deposited in the PDB under code: 5TCX. GraphPad Prism 7.0 is available from an institutional license at Harvard Medical School. The EVfold server is a freely available online tool created by the Marks lab at Harvard Medical School and the Sander lab at Memorial Sloan Kettering Cancer Center. All software used in structure determination (XDS, Phenix, Pymol and Coot) are accessible via the Structural Biology Grid (SBGrid) at Harvard Medical School.
Figure S1. Assessment of CD81 Biochemical Quality and Phase Calculation, Related to Figure 1

(A) SDS-PAGE analysis of purity. Left lane shows molecular weight standard, right lane is FLAG-purified CD81.

(B) Analytical size exclusion profile of purified CD81 in DDM detergent buffer on a Superdex 200 column.

(C) Overlay of CD81 ECL structures. Ribbon diagrams of EC2 of CD81 from this structure superimposed on five other EC2 structures. Red: EC2 from this structure, blue: 1G8Q, chain A, green: 3X0E, chain A, cyan: 3X0E, chain B, orange: 1IV5, chain A and magenta: 1IV5, chain B.
Figure S2. Assessment of Helix Register Assignment, Related to Figures 1 and 2

(A–D) Composite omit 2F_o-F_c electron density map contoured at 1.0 σ for TM1 (A), TM2 (B), TM3 (C) and TM4 (D).
Figure S3. Sequence Conservation among CD81 Homologs and Paralogs and ECL2 and TM1/TM2 Homolog Interface, Related to Figure 3
(A and B) Surface representation of CD81 homologs (A) and mammalian paralogs (B). Analysis was performed using Consurf, colored in a sliding scale from teal (poorly conserved) to crimson (highly conserved). Yellow indicates residues with insufficient data for assignment.
(C) Interface between TM1/TM2 and EC2. Residues at the interface between the TM1/2 bundle and EC2 are shown as sticks colored by conservation.
Figure S4. Electron Density, Cholesterol Binding, and Conservation of Pocket Residues, Related to Figures 4 and 6

(A) Composite omit map (2F_o-F_c) of protein density contoured at 2σ (blue) and difference map (F_o-F_c) showing difference electron density (green) modeled as cholesterol (yellow sticks). The protein model is in magenta.

(B) Immunopurified proteins. Wild-type and mutant (E219Q, G26F/G30F) CD81 proteins were immunopurified from HEK293T cells. Proteins captured on FLAG beads were used for radioactive cholesterol binding experiments. WCL: whole cell lysate; Sol: solubilized protein; IP: immunoprecipitated protein.

(C) Cholesterol binding. 1,2-3H-cholesterol was incubated with immunopurified FLAG-CD81 proteins and bound cholesterol was measured.

(D) Sequence conservation in the cholesterol binding pocket. Sequence conservation was determined using ConSurf as described in the STAR Methods. Residues within 4 Å of the cholesterol binding site are denoted by a red asterisk. Percentage conservation is denoted in the table at the right.
In both cholesterol-bound and apoprotein simulations, the transmembrane helices undergo substantial motion. Snapshots from cholesterol-bound and apoprotein simulations are shown in blue and red, respectively, while the crystal structure is shown in gray for comparison. In the apoprotein simulations, a hydrogen bond forms between N18 and E219 that keeps TM1 and TM4 in close proximity. In cholesterol-bound simulations, as in the crystal structure, cholesterol forms a hydrogen bond with E219, disrupting interaction between N18 and E219 and often leading to separation of TM1 from TM4.
Figure S6. Lattice Contacts, Related to Figure 1
(A–C) Lattice packing of CD81 viewed parallel to the membrane plane (A). (B) and (C) show a view normal to the membrane and a second parallel view, respectively.