Direct interactions of CEACAM1 and CD36 with LPS and each other

Weidong Hu[1], Patty Wong[2], Supriyo Battacharya[3], Zhuo Li[4], Lin Li[1], Eric Aniogo[1], Jitender Jitender[1], Teresa Hong[1], Zhifang Zhang[5], Paul Yazaki[1], Maciej Kujawski[1]and John E. Shively[1]

[1]Department of Immunology and Theranostics, Arthur Riggs Diabetes and Metabolism Research Institute of City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA.

[2]Department of Radiation Oncology, City of Hope National Medical Center, 1500 East Duarte Road, Duarte CA 91010, USA.

[3]Department of Integrative Genomics and Bioinformatics, Beckman Research Institute of City of Hope, 1500 East Duarte Road, Duarte CA 91010, USA.

[4]Electron Microscopy Core Facility, Beckman Research Institute of City of Hope, 1500 East Duarte Road, Duarte CA 91010, USA.

[5]Department of Surgery, City of Hope National Medical Center. 1500 East Duarte Road, Duarte, CA 91010, USA.

Address Correspondence to:

Dr. John E. Shively; jshively@coh.org

KEY WORDS: CEACAM1, LPS, CD36

ABSTRACT

Introduction: Liposaccharide (LPS), a ubiquitous bacterial micellar component of ingested food, is neutralized by a variety of mechanisms that help to establish a threshold, which when exceeded results in an inflammatory TLR4 mediated response. Notably, both CEACAM1 and CD36 affect downstream signaling of TLR4 to LPS. Furthermore, CEACAM1 associates with CD36 in hepatocytes, regulating lipid storage and bile acid (BA) secretion that includes reverse transport of LPS to the intestine. **Methods**: Direct binding of LPS-Ra micelles to soluble CEACAM1 or soluble CD36 was analyzed by surface plasmon surface resonance (SPR), size exclusion chromatography (SEC) and transmission electron microscopy (TEM). Direct binding of CEACAM1 to CD36 was analyzed by SPR and proximity ligation assays. Molecular models were generated by Alpha Fold and Molecular Dynamics. **LPS Binding to sCEACAM1 and sCD36**: SPR binding constants of K_D= 1.04 x 10^{-10} M and K_D= 3.38 x 10^{-10} M were obtained for LPS-Ra micelle binding to sCEACAM1 and sCD36, respectively. On SEC, the molecular sizes of LPS-Ra micelles bound to sCEACAM1 and sCD36 were approximately 500 and 800 kDa, respectively. In addition, LPS binding to both was reduced by BAs sodium cholate and sodium deoxycholate. Alpha Fold predicted a binding site of LPS-Ra to CD36, while Molecular Dynamic studies of an N-domain mutant of CEACAM1, that breaks a conserved salt bridge, revealed the presence of an open form that is predicted to bind LPS. s**CEACAM1 to sCD36 Binding:** A K_D of 5.28 x 10^{-8} M was obtained for sCEACAM1 binding to immobilized sCD36 by SPR. Antibody-based-proximity ligation demonstrated the association of full length CEACAM1 and CD36 on the surface of hepatic cells and when co-expressed on HEK cells, while biotin-based proximity ligation demonstrated association of the cytoplasmic domain of CEACAM1 to the cytoplasmic domain of a CD36-BioID2 fusion protein co-expressed in HEK cells. Alpha Fold predicted both head-to-head (trans) and side-to-side (cis) binding of the N-domain of CEACAM1 to CD36, from which a membrane model of their cis-interaction could account for the proximity ligation results. **Conclusion**: Both CEACAM1 and CD36 share a common LPS micelle binding function, as well as binding to each other, and together, may regulate uptake and excretion of micellar LPS.

Word count: 359

INTRODUCTION

Human CEACAM1 is a single pass transmembrane glycoprotein with 3-4 extra-cellular Ig-like domains and two different cytoplasmic signaling domains formed by alternative splicing (1). It occurs as inactive cis-dimers on epithelial cells and immune cells, and upon activation by inside-out signaling, forms trans-dimers between cells that modulate a variety of co-receptor signaling pathways, including the TCR of T-cells (1) and the BCR of B-cells (2,3). Dimer formation is dependent on a Gly-xxx-xxx-xxx-Gly motif in its transmembrane domain (4), leading to subsequent homophilic binding of the GFCC'C'' faces of its distal N-domain (5,6). Several groups have expressed the unglycosylated N-domain of CEACAM1 in *E. Coli* that exhibits a tight formation of the homodimer (7,8). This is in contrast to the expression of full length, glycosylated, soluble CEACAM1 isoforms with 3 or 4 extracellular domains (sCEACAM1-3D and -4D) that were shown to be elongated monomers by ultracentrifugation (9). Solution phase NMR studies at first predicted a monomer for partially de-glycosylated N-domain (10) that was later amended to a dimeric structure (11). These conflicting studies suggest that the N-domain of CEACAM1 may not only interact with itself but also with addition ligands that depend on the presence of its glycosylation, and/or its appendant Ig-like domains.

Clues to the existence of additional ligand specificity for CEACAM1 are found in studies on its effect on LPS signaling. LPS is routinely shed from growing and/or dying gram negative bacteria as a lipid micelle (12). LPS, along with fatty acids, is taken up by CD36 in the small intestine (13) where CEACAM1 is also expressed (14,15). Like all intestinally absorbed lipids that enter the lacteal lymphatics, micellular LPS is incorporated into circulating lipid particles including chylomicrons, VLDL, LDL and HDL where it does not usually represent a threat to the organism (16,17). On the other hand, systemic LPS may lead to life threatening sepsis and is detected by binding proteins such as LPS binding protein (LBP) that binds to a LPS monomer. LBP bound LPS is sequentially transferred first to CD14 and then to a TLR4-MD2 complex resulting in a potent inflammatory response (18). Since LBP is an acute phase protein produced by the liver (19), it is likely that other proteins are capable of binding either monomeric or micellar LPS. Candidates include, among others, soluble CD14 (sCD14), soluble CD36 (sCD36), and as we propose now, soluble CEACAM1 (sCEACAM1), all of which are found in normal serum (20-22). Since the liver is a major organ of LPS clearance, it is not surprising that it can excrete LPS into the bile (23), returning it to the small intestine for disposal, similar to the reverse transport of cholesterol (24,25). Candidate LPS receptors found in bile canaliculi are CD36 (26) and CEACAM1 (27). In fact, a soluble form of CEACAM1, found in bile and

described as biliary glycoprotein 1 (BGP1) (28,29), suggests that it plays a dual role in bile, excreting both excess cholesterol and LPS. Among other proteins, BGP1 is known to bind cholesterol (30). Amino acid sequence analysis revealed that BGP1 is a truncated form of CEACAM1 that includes its IgV-like N-domain followed by several IgC2-like domains (31). It is likely produced from CEACAM1 by proteolysis since it was detected by us in Crispr/Cas9 deleted CEACAM1 HepG2 cells that were transfected with cytoplasmic domain variants of CEACAM1 (32).

Comprehensive studies showed that CEACAM1 is highly expressed in neutrophils and urinary epithelia and that the Opa proteins of *Neisseria gonorrhea* bind CEACAM1 allowing the bacteria to invade and spread infection partly due to the inhibitory signaling motif of CEACAM1 (33,34). Other studies identified that many pathogenic bacteria express CEACAM1 binding adhesins that allowed them to circumvent the inflammatory response in the airway, urogenital tract, stomach, and intestine, including invasion into the submucosa (35-37). We previously showed that CEACAM1 is phosphorylated by Src upon LPS binding to TLR4 in neutrophils, downregulating the Syk activation of IL-1β by the NLRP3 inflammasome (38). In addition, IP injection of LPS into wild type vs CEACAM1 KO mice revealed that a CEACAM1 modulated the IL-6 mediated inflammatory response was more severe in the KO mice (39). Binding of LPS to the N-domain of CEACAM1 expressed in *E. coli* was shown by Servin and co-workers, along with TLR4 mediated Erk1/2 and p38 MAPK signaling in CEACAM1 transfected MDCK cells (40). Overall, these studies provided evidence that CEACAM1 downregulates LPS signaling, acting as a threshold of activation and signaling, depending on its expression levels in various cells. However, the proposal that CEACAM1 can directly bind LPS has not been verified.

Lipid dysbiosis in the liver is strongly associated with both CEACAM1 and CD36, apparently in opposite directions, since abrogation of *Ceacam1* leads to fatty liver disease (41), while abrogation of *Cd36* can correct hepatosteatosis (42). When CEACAM1 was deleted in HepG2 hepatocytes by Crispr/Cas9, these cells upregulated *CD36*, and the two proteins co-immunoprecipitated with Src (32), a critical tyrosine kinase involved with both CEACAM1 (43,44) and CD36 (45,46) signaling. CD36 not only modulates TLR4 signaling (47), but is also responsible for transport of fatty acids into intestinal enterocytes and across the endothelium (46). Subclinical LPS leakage and inflammation are found in the intestines of CD36 KO mice (48), demonstrating the important role of LPS uptake by CD36 in the intestine (13). Genetic alleles resulting in low or absent levels of CD36 show that similar intestinal effects may occur in man (49). In fact, there may be a fundamental connection between type 2 diabetes in obesity

and LPS since high serum levels of LPS found in obesity are immediately reversed by gastric sleeve surgery (50,51), a procedure with profound effects on food/lipid uptake in the duodenum.

Given these considerations, we tested the hypotheses that CEACAM1 and CD36 both directly bind LPS, and also, directly interact with each other.

METHODS

Materials

Antibodies used in these studies are described in the sections below. Lipopolysaccharides (LPS) rough strains from *E. coli* EH 100 (Ra mutant) were purchased from Sigma. Sodium dodecyl sulfate 0.5% (w/v), HBS-N buffer and HBS-EP buffer were purchased from Cytiva. Sodium cholate hydrate and sodium deoxycholate monohydrate were purchased from Thermofisher Scientific. Human recombinant ectodomain CD36 (Gly 30 – Asn 439) was purchased from AcroBiosystems.

Expression of CEACAM1 proteins

Plasmid construction and membrane expression of CEACAM1-4L has been previously described (52,53). The soluble ectodomain of human sCEACAM1-4D (4 domains) and soluble N-domain of CEACAM1 containing C-terminal His6 tags with its secretion signal sequence were cloned into the mammalian expression vector pIBS54. The constructs were transiently expressed in Expi293 cells using the ExpiFectamine™ 293 Transfection Kit (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's protocol. Cells were maintained in Expi293 expression medium at 37 °C in a humidified incubator with 8% CO_2 and agitation at 125 rpm. The culture supernatant was collected seven days after transfection.

The harvested medium was clarified by centrifugation and treated overnight at 4 °C with AG1-X8 resin (Bio-Rad, Hercules, CA) to remove residual nucleic acids. The treated supernatants were diluted with wash buffer to get 10mM imidazole concentration then applied to a 5 ml Ni-NTA Superflow column (Qiagen, Germantown, MD) pre-equilibrated with buffer containing 50 mM sodium phosphate, 300 mM sodium chloride, and 10 mM imidazole at pH 7.4. The column was washed with the same buffer containing 20 mM imidazole, and bound protein was eluted with 250 mM imidazole. The eluate was buffer-exchanged into 0.05 M MES, 0.01 M potassium phosphate buffer, pH 6.5, using 10 kDa molecular weight cutoff centrifugal filters. The IMAC-purified proteins were further polished using ceramic hydroxyapatite (CHT) Type I (Bio-Rad, 20 µm, 7 mL) chromatography to remove aggregates. The column was equilibrated with 0.05 M MES, 0.01 M potassium phosphate buffer at pH 6.5. After loading, the column was washed with 0.05 M MES, 0.05 M potassium phosphate buffer, and the proteins were eluted using a linear gradient up to 0.2 M potassium phosphate. Protein elution was monitored by absorbance at 280 nm. Monomeric fractions were confirmed by size-exclusion chromatography (Superdex 200 Increase 10/300 GL, Cytiva) and exchanged into PBS for subsequent experiments

Surface Plasmon Resonance (SPR)

The experiments were carried out on a Biacore™ X100 (GE Healthcare) at 25°C. Human recombinant human CD36 (CD6-H5221 from Acro Biosystems) was diluted into 10 mM sodium acetate (pH 4.5) to final concentration of 12.5 µg/mL. The CD36 was immobilized to two CM5 chips with resonance units (RU) of ~430 and ~2500. The immobilized CD36 with 430 RUs was used to study the interaction with Ceacam1, and the one with 2500 RUs was used to study the interaction with LPS and/or bile acids. The expression of human recombinant sCEACAM1-4D was described above. A concentration of 50 µg/mL of sCEACAM1-4D was prepared in 10 mM sodium acetate (pH 4.5) for immobilization onto a CM5 with RU of ~2300. Concentration of N-domain of Ceacam1 of 50 ug/ml in 10 mM sodium acetate (pH 4.5) is used for immobilization onto a CM5 chip with RU of ~1050. The running buffer HBS-EP was used for CEACAM1 interaction with immobilized sCD36, and the chip was regenerated with 0.1 N HCl. The running buffer HBS-N was used for the interaction between LPS or bile acids with immobilized sCEACAM1-4D or sCD36, and the chip was regenerated with 0.25% or 0.5% SDS water solution. Typical contact and dissociation times for binding studies were set at 180 and 700 seconds, respectively. The contact and stabilization times for regeneration were set to 30 and 700 seconds, respectively. Data were analyzed using Biacore™ X100 software and excel.

Size exclusion chromatography and calculation of Stoke's radius

Samples were run on a Superdex 200 Increase 10/300 GL) Cytiva, Marlborough, MA) in pH 7 PBS at a flow rate of 0.5 mL/min with UV absorption at 215 and 280 nm. The column was calibrated with standards of known Stokes radius (BioRad gel calibration standards) and calculations performed according to La Verde et al. (54)

Transmission electron microscopy

A 5 µl drop of protein complexes isolated from SEC was applied to a freshly plasma-cleaned 200-mesh formvar/carbon coated grid (Ted Pella Prod No. 01801) for 1 minute, blotted to a thin film using filter paper, floated on drops of Milli-Q H_2O for 20 seconds (3x), blotted to a thin film again using filter paper, and immediately stained with 1% (w/v) uranyl acetate for 1 minute. The stain was then blotted away with filter paper and the grid subsequently air-dried. Transmission electron microscopy was performed using an FEI Tecnai 12 electron microscope operating at 120 kV equipped with a Gatan OneView CMOS camera. Images were collected at nominal magnification of 52,000x (0.21 nm/pixel) using Gatan Digital Micrograph. Images were acquired at a nominal underfocus of -1.5 um.

Particle size analysis was carried out using automatic particle counting function in Image J (55). For sCEACAM1-4D (Fig 6), individual particles in the 52,000x images were selected using automated picking protocols, followed by several rounds of reference-free 2D alignment and classification algorithms using XMIPP (56) to sort them into self-similar groups.

Proximity ligation assays

Proximity ligation assay (Duolink PLA, Millipore Sigma) for CD36 and CEACAM1 was performed accordingly to the manufacturer's protocol. Briefly, cells were grown in Nunc Lab-Tek II Chamber Slide System (Thermo Fisher), fixed and permeabilized, blocked and incubated with primary antibodies: huCD36 (D8L9T) Rabbit mAb (Cell Signaling Technology, Danver, MA) and huCEACAM1 (MAB22441) Mouse mAb (R&D Systems, Mineapolis, MN) overnight in 4ºC. Next day slides were incubated with anti-Mouse and anti-Rabbit PLA probes, ligated and the signal was amplified by polymerase. Slides were mounted with In Situ Mounting Medium with DAPI and analyzed using a LSM900 confocal microscope (Carl Zeiss Microscopy, White Plains, NY).

Biotin proximity ligation assay. Transient expression of CD36-BioID2 and CEACAM-1 or both was performed using the Expi293F mammalian expression system (Thermofisher, Waltham, MA, USA). Briefly, 75 x 10^6 viable cells were transfected with 25 µg of CD36-BioID2 and/or CEACAM1-4L plasmids in a 125mL vented, plain bottom flask (PBV125). For co-expression, equal volumes of both CD36-BioID2 and CEACAM-1 plasmids (12.5 µg each) were also transfected, and a flask with no DNA was set as a negative control. On day 4 following transfection, the cells were collected and stained with either anti-CD36 (FAB19551G) or anti-CEACAM-1 (R&D systems, FAB2244P) antibody to assess the level of expression using flow cytometry.

Proximity labelling was performed by adding biotin (Thermo Fisher, B20656) to a final concentration of 50 µM to each culture groups and incubated overnight at 37ºC. Next, 20 x 10^6 cells were transferred to a 15 ml conical tube and were washed twice with 15 ml ambient PBS. Following the wash, the cells were lysed with 2mL of lysis buffer with 20ul protease inhibitors added. The lysed cells were transferred to a 2 mL microcentrifuge tube and incubated for 30 min on ice. After the incubation, the lysate was centrifuged at 15,000 xg for 5 min at 4ºC and the clarified supernatant transferred to a new 2 ml tube. The labelled protein was isolated by incubating the lysate with 100 µL of streptavidin beads (Thermo Fisher, 65801D) and incubated on a rotator for 4 hrs at 4ºC. Thereafter, the beads were washed three times to remove unbound protein and resuspended with 400ul of SDS-PAGE sample buffer for Western blot analysis.

Western blot analysis was done with a western detection protocol using the Quick Western Kit (Li-Cor Biosciences, Lincoln, NE). Briefly, 20 uL of the lysate resuspended in SDS-PAGE sample buffer were boiled for 5 min at 95°C and run at 150 V for 60 minutes on a denaturing non-reducing NuPAGE 4 – 12% Bis-Tris SDS-PAGE gel (Thermo Fisher) and later transferred to PVDF membrane (Bio-Rad, Hercules, CA, USA). The membrane was blocked for 1 hour with Intercept blocking buffer at room temperature and then incubated with Quick Western Detection Solution. This detection solution contains the primary antibody of choice diluted according to the manufacturer, IRDye 680RD detection reagent (1:1000 dilution) in Intercept blocking buffer solution, 0.2% Tween20, and 0.02% SDS. The antibodies used as primary antibodies are: CD36 (Human CD36/SR-B3; Cat no: AF1955) diluted to 1:1000 or CEACAM-1 (Human CEACAM-1/CD66a; cat no: MAB22441) diluted to 1:2000. Both antibodies were purchased from R&D Systems. The membrane blot was incubated overnight at 4°C on a platform shaker. After the incubation, the detection solution was carefully decanted, and the membrane was rinsed 3 times with 15mL of 1X PBS-T (1X PBS + 0.1% Tween 20) and scanned with the Odyssey Imaging System using the 700 nm channel.

AlphaFold models of CD36-CEACAM1 interaction and LPS bound CD36 and CEACAM1 dimer:

We used AlphaFold v3.0 installed on a GPU cluster and obtained 50 models for each system. CD36 and CEACAM1 sequences were obtained from Uniprot. The chemical formula for LPS-Lipid-A was supplied to AlphaFold in the form of a SMILES string.

Molecular Dynamics simulations of open form of CEACAM1 N domain:

The starting structure for the CEACAM1 N domain was derived using AlphaFold v3.0 by submitting the protein sequence to the AlphaFold Server (https://alphafoldserver.com) (57). The input for MD simulation was prepared using the tleap module of AmberTools (58). The protein structure was solvated in a rectangular water box of dimensions 94Å x 81Å x 81Å and Na+/Cl- ions were added to neutralize the total charge and achieve an effective salt concentration of 0.15M. The system was parameterized using the FF19SB forcefield and the TIP4PEW water model (59,60). The system was minimized for 2000 cycles of which, first 1000 were via steepest descent and the rest using the conjugate-gradient method. Next, the system was gradually heated from 0 to 310K over 30 ns at constant volume, during which, the protein heavy atoms were restrained by applying a harmonic force with force-constant 500 kcal/mol/Å. Following this, the system was simulated for a further 50 ns in the NPT ensemble (310K, 1 atm) during which,

the harmonic restraints were gradually reduced to zero. A final equilibration run was performed for 50 ns without any restraints. The production runs consisted of two independent simulations lasting for 1.8-1.9 μs starting with randomly generated velocities. Temperature and pressure were maintained using the Langevin thermostat with collision frequency gamma ln set to 1.0 The MD simulations were performed using GPU accelerated AMBER v22. Hydrogen mass repartitioning was applied to enable a timestep of 4 femtosecond (61).

Molecular dynamics simulations of D82A mutant of glycosylated CEACAM1-N domain:

In the AlphaFold derived structure of human CEACAM1-N domain, D82 was mutated to alanine by direct deletion of the aspartate side chain and changing the residue name to ALA in the pdb file. The structure of the glycosylated CEACAM1-N domain was derived using the Glycoprotein Builder web interface at GLYCAM-Web (https://glycam.org/gp/)(62), where the positions N77 and N81, the sites nearest to D82, were glycosylated as depicted in Supplementary Fig. S10. The MD system was prepared using the Glycan Modeler module of CHARMM-GUI (63). The system was solvated in a water box of dimensions 92 Å X 92 Å X 92 Å and Na+ and Cl- ions were added to neutralize the net charge and achieve an effective salt concentration of 0.15M. The system was parameterized using the CHARMM36m forcefield (64) and simulated using the GPU accelerated AMBER v22. Equilibration started with 5000 cycles of minimization of which, the first 2500 were via steepest descent and the rest using the conjugate-gradient method. Following this, the system was heated to 310K over 125 ps and then equilibrated for 10 ns in the NVT ensemble with a timestep of 1 femtosecond. During the minimization and heating phases, the protein heavy atom positions were restrained by applying harmonic forces with force constant 1 kcal/mol. Further, the glycan dihedrals were restrained using harmonic forces with a force constant of 1 kcal/mol. The production run continued for 2.5 us in the NPT ensemble (310K, 1 atm). Temperature and pressure were maintained using the Langevin thermostat with collision frequency gamma ln set to 1.0 Hydrogen mass repartitioning was applied to enable a timestep of 4 femtosecond (61).

Modeling of LPS-Ra bound to CEACAM1-N domain:

For modeling bound LPS, we selected the MD conformation of the D82A-CEACAM1-N domain showing the largest distance between N81 and N61, ensuring an open conformation with adequate space to dock LPS (Fig. 11A-B). Due to the large size and high number of rotatable

bonds, the entire LPS molecule could not be docked to CEACAM1 using conventional docking. We therefore adopted a piecemeal approach, where the di-glucosamine segment of lipid-A was first docked to the CEACAM1 cavity using GLIDE SP (65). The ligand conformations were generated outside the protein cavity using MacroModel, followed by the docking of the ligand structural ensemble to CEACAM1. During docking, the vdW radii of both protein and ligand atoms were scaled by 0.5. The best pose with the lowest docking score was selected for the next phase. in which the alkyl chains were manually added using the Molecule Builder interface of Maestro, followed by local minimization of the modeled atoms, keeping the protein rigid (Fig. 11C). Finally, with lipid-A in place, the core glycan moiety was docked to CEACAM-1 and the pose with the lowest docking score was selected. The lipid-A and core glycan were then covalently connected using the molecule builder interface of Maestro. The final structure was subjected to thorough minimization using the PrimeX module of Maestro.

Location of primary data:

Primary data is available from the senior author by request.

RESULTS

Direct binding of LPS-Ra to immobilized sCEACAM1-4D by SPR. For this study we used LPS-Ra that contains all of the core glycan but not the O-specific glycan of LPS. LPS-Ra is a micellar, water soluble LPS that is a well-studied LPS molecule, including an x-ray structure of its monomeric form in the TLR4-MD2-LPS complex (66). To approximate a natural state for LPS, we made no attempt to disrupt LPS micelles in this study. Since human CEACAM1 is expressed with an IgV-like N-domain followed by 2-3 extracellular IgC domains due to alternative mRNA splicing (1), we chose the largest splice variant with 4 extracellular domains (sCEACAM1-4D) as representative for this study. Expression and purification of sCEACAM1-4D is shown in **Supplemental Fig S1**. It should be noted that we had previously expressed both sCEACAM1-3D and sCEACAM1-4D variants and showed they were both *monomers* when analyzed by ultracentrifugation (9). These results are in contrast to numerous studies on expressed soluble CEACAM1 N-domain showing it is a dimer, but in agreement with the finding that membrane bound CEACAM1 dimerization requires their transmembrane domains for dimerization (4). When sCEACAM1-4D was immobilized on a CM5 chip and binding studies with LPS-Ra performed over the range of 0-4.0 μM, a K_D of 1.04 x 10^{-10} M was obtained (**Fig 1A**). The dissociation of LPS-Ra from immobilized CEACAM1-4D was very slow, suggesting that CEACAM1-4D was binding a high molecular weight, micellar form of LPS-Ra with multiple copies of its subunits resulting in a high avidity interaction.

Since CEACAM1-4D has 4 domains, its interaction with LPS could be with any one or a combination of its domains. Given that the N-domain of CEACAM1 has been shown to convey its functional activity, namely homotypic cell adhesion (1), and CEACAM1-4D was shown to be a monomer by ultracentrifugation studies (9), we attempted to block LPS binding to the N-domain with an N-domain specific antibody, T84.1 (67). When immobilized sCEACAM1-4D was first subjected to binding of T84.1 followed by LPS (0.5 μM), the RUs bound were similar to binding of LPS (0.5 μM) alone (**Supplemental Fig S2A**), demonstrating that this antibody did not block LPS binding. Since this experiment did not definitely rule out the N-domain as an LPS binding domain, we also expressed sCEACAM1-N (**Supplemental Fig S2B**) and found LPS still bound with K_D of 3.56 x 10^{-9} M (**Supplemental Fig S2C**). We conclude that (a) at a minimum, LPS-Ra binds to the N-domain, and (b) LPS-Ra binds to a site other than the well-studied GFCC'C'' binding site on the N-domain (6).

Binding of bile acids to sCEACAM1-4D by SPR

Since CEACAM1 expressed in the liver and intestine is exposed to bile acids (BAs), as well as LPS, we first performed binding studies of CEACAM1-4D with increasing concentrations of two BAs, sodium cholate, a primary BA, and sodium deoxycholate**,** a secondary BA**,** to immobilized sCEACAM1-4D **(Fig 1B-C).** Sodium cholate (0.2- 3.2 mM) in the absence of LPS showed low binding to CEACAM1-4D compared to sodium deoxycholate (0.05- 0.80 mM). In both cases the bound BAs exhibited slow rates of dissociation; however, no reliable rate constants were obtained for these two BAs binding to CEACAM1-4D.

Direct binding of LPS-Ra to immobilized sCD36 by SPR

Since our previous studies showed that CEACAM1 and CD36 co-immunoprecipitated from hepatocytes (32) and CD36 was known to affect LPS uptake in the intestine (68), we asked if we could demonstrate direct binding of LPS to sCD36, similar to sCEACAM1-4D. When soluble sCD36, obtained from a commercial source, previously validated by SPR binding studies to oxLDL (69), was immobilized on a CM5 chip and binding studies with LPS-Ra performed over the range of 0-4.0 μM, a K_D of 3.38×10^{-10} M was obtained (**Fig 1D**). Although the kinetic association constant k_a for sCD36 was similar to that for sCEACAM1-4D, the dissociation constant k_d for both sCD36 and sCEACAM1-4D were too small to be determined precisely, suggesting both were binding LPS Ra micelles.

Binding of bile acids bind to sCD36 by SPR

Similar to CEACAM1-4D, CD36 is highly expressed in the intestine (68) and exposed to BAs during digestion. When BA binding studies were performed with immobilized sCD36, we found higher binding with deoxycholate than with cholate **(Fig 1E-F).** Thus, BAs bind to both sCEACAM1-4D and sCD36 but with much less mass (RUs) than LPS, suggesting that the BAs bind as monomers while the LPS binds as micelles. The micellar nature of the LPS-Ra complexes with sCEACAM1-4D and sCD36 was later confirmed by SEC and TEM studies.

Effect of BAs on the binding of LPS-Ra to sCD36 or sCEACAM1-4D by SPR

Having established that sCEACAM1-4D binds both LPS-Ra and BAs, we wanted to determine the effect of BAs on its binding to LPS-Ra, a situation expected to occur during intestinal digestion. The experimental setup included runs with no LPS-Ra (**Fig 2A**), runs with a fixed concentration of LPS-Ra (2 μM) with increasing concentrations of deoxycholate (**Fig 2B**) and corrected sensograms with subtraction of RUs in the absence of LPS (**Fig 2C**). The results indicate that the RU values from LPS-Ra-deoxycholate mixtures are much less than the sum of RU from individual LPS and deoxycholate alone. This could be due to both LPS-Ra and deoxycholate binding to the same site on CEACAM1-4D, but since molecular size of deoxycholate is smaller, the binding of deoxycholate instead of LPS-Ra may reduce the RU signal significantly. Another possible reason is that deoxycholate may reduce the micelle size of LPS-Ra leading to a reduced RU signal. Using equivalent concentrations of cholate, the study was repeated (**Fig 2D-F**), indicating that the effect of cholate on LPS binding was less than that of deoxycholate. Since deoxycholate is known to reduce the size of LPS micelles (70), changes in the total binding (measured in Ru) of LPS-Ra micelles to CEACAM1-4D in the presence of BAs is a reasonable explanation.

The effect of BAs on LPS binding to immobilized sCD36 (**Fig 3**) gave similar results to CEACAM1-4D (**Fig 2**). Since it appears that BAs compete for the same binding site on sCD36, or CEACAM1-4D we replotted the data as normalized RU (± correction for the binding of BA only) vs the molar ratio of BA/LPS-Ra. In the case of sCD36, a smooth curve was obtained for cholate, indicating that cholate reduced the LPS-Ra binding RUs mainly by one mechanism, that is reduction of the micelle size of LPS-Ra. For the curve of deoxycholate there is a discontinuity around a molar ratio of 50, suggesting that in addition to the reduced the size of LPS-Ra micelles, the deoxycholate also competes to bind the same site on sCD36 as its concentration increases. (**Supplementary Fig S3A**). Since deoxycholate is more efficient than cholate in reducing the size of LPS micelles (70) these results are expected. Similar results were obtained for the effect of BAs on sCEACAM1-4D with cholate giving essential linear results and deoxycholate giving steeper, higher order curves (**Supplementary Fig S3B**). In both cases, deoxycholate was more effective in suppressing LPS-Ra binding to both proteins.

Direct binding of LPS-Ra to sCD36 or sCEACAM1-4D by SEC

To determine if the binding of LPS-Ra could occur in solution phase, sCD36 (35 µg/0.1 mL; 5 µM) or sCEACAM1-4D (35 µg/0.1 mL; 2.9 µM), were incubated with increasing amounts of LPS-Ra (0-150 µg/0.1 mL) for 60 min at 37°C, then the entire volume (0.1 mL) analyzed by SEC. It should be noted that LPS-Ra is UV transparent at 214nm and thus only changes in the retention times of sCD36 or sCECAM1-4D are observed. The results show that both sCD36 and sCEACAM1-4D generate higher molecular complexes that reach a maximum at about 100 µg of LPS-Ra (**Fig 4**). Since the injected complex undergoes a dynamic dilution of about 10-fold during chromatography, it is possible that a certain amount of complex dissociation occurs preventing conversion to 100% complex formation. In any case, the molecular size of the CD36-LPS-Ra complex was estimated by reference to standards of known Stokes radii as 7.9 nm (ca 500 kDa for a globular particle) and 8.9 nm (800 kDa) for the CEACAM1-4D-LPS-Ra complex (**Supplemental Fig S4**). Whereas sCD36 can be approximated as a globular structure, that is not the case for sCEACAM1-4D that was shown by us to be a prolate ellipsoid with an axial ratio of 8 and a Stokes radius of 5.5 nm by ultracentrifugation (9). In good agreement, the calculated Stokes radius of CEACAM1-4D from **Supplemental Fig S4** is also 5.5 nm.

Since both CEACAM1 and CD36 have been shown to modulate TLR4 signaling (38,71), and TLR4 is a well-studied receptor for monomeric LPS-Ra (66), we tested the possibility that either sCD36 or sCEACAM1-4D could directly interact with sTLR4. At least by coincubations of each with sTLR4 and analysis by SEC, no interactions were observed in the µM range (data not shown). In addition, sTLR4 did not interact with LPS-Ra under the same conditions as above (data not shown), a result consistent with the requirement that TLR4 must bind a monomeric form of LPS in the presence of the accessory protein MD2 (66). Since no attempt was made to generate monomeric LPS-Ra by approaches such as sonication (66) or incubation with LPS-binding protein (72), and given the large size of the LPS-Ra complexes generated for both sCD36 and sCEACAM1-4D, we conclude that both are incorporated into apparently well-defined LPS-Ra micelles. Therefore, the particle sizes of the complexes isolated by SEC were further examined by transmission electron microscopy (TEM)

Particle sizes of LPS-Ra complexes with sCD36 or sCEACAM1-4D by TEM

When the LPS-Ra complex of sCD36 was collected by SEC (see above) and examined by TEM, a range of ellipsoid particle sizes were observed: with a median minor axis of 8.4 nm and major axis of 12.3 nm (**Fig 5**). The calculated radius (6.1 nm) of the median particle is less than the Stokes radius obtained by SEC (7.9 nm), while the maximum radius of 8.6 nm is larger than obtained by SEC. In spite of this discrepancy, the gaussian distribution of particle sizes seen in TEM agrees well with the symmetrical peak seen on SEC. Similar results were observed for the LPS-Ra-CEACAM1-4D complex: with a median minor axis of 7.4 nm and major axis of 12.0 nm (**Fig 6**). While the median radius calculated from TEM (6.0 nm) does not agree with that obtained from SEC (8.9 nm), the maximum radius (8.9 nm) is in good agreement.

Given the predicted self-association of CEACAM1 with itself (43), we analyzed CEACAM1-4D by TEM. Six particle shapes were identified and their numbers out of 499 particle shapes analyzed (one set shown in **Fig 7)**. Although we expected to see only 4 domain particles (eg, id 1 and 5, **Fig 7**), many appear to have only 3 domains (eg, id 2 and 6), and some with a kink between two of the domains (eg, id 4). The measured length of an extended 4 domain particle was 17.5 nm. This corresponds to a predicted radius of 8.75 nm, roughly 1.6 times the Stokes Radius calculated by SEC (5.5 nm) or by ultracentrifugation (9). We conclude that CEACAM1-4D adopts a more compact structure in solution vs more elongated particles as seen by TEM. Importantly, the predicted length of CEACAM1-4D was 16.2 nm (or a radius of 8.1 nm) by Alpha Fold (**Supplemental Fig S5**) and that calculation does not include 2.7 nm of an unstructured domain between the A2 and TM domains. Since the unstructured sequence may not exist as an extended structure as predicted by Alpha Fold, it could easily account for an additional 1.4 nm, leading to good agreement with the longest TEM measurement. In addition, the occasional kinked structures observed by TEM agree well with the Apha Fold predicted structure that shows the A2 domain at a 120-degree angle to the other 3 domains. It is notable that we saw no evidence of head-to-head dimers in the TEM, arguing that the dimerization of the isolated N-domain that forms dimers in solution (6) may not be representative of a glycosylated CEACAM1 with its additional C2-Ig-like domains.

Since LPS, whether as the highly glycosylated smooth form or the less glycosylated rough form (LPS-Ra), self-assembles into large irregular sized micelles on their own (73), it appears that eukaryotes developed LPS binding molecules that reduce the LPS micelle size making it more manageable for recognition. It appears that this is the case for both CEACAM1 and CD36, for

which the addition of their soluble forms leads to more defined micelle sizes. Other examples include BAs (70), LPS binding protein and CD14 (17), and more recently, caspase-4 (74). In the case of Cas-4, the LPS-Ra micelle size is gradually reduced to a more uniform size by increasing ratios of Cas-4 to LPS-Ra on SEC, similar to that shown in **Fig 4** for sCEACAM-1-4D and sCD36. Thus, we can add CEACAM1 and CD36 to a growing list of LPS micelle modifiers.

Molecular modeling of LPS-Ra monomer binding to sCD36

Since the newer versions of Alpha Fold (alphafold server.com) allow docking of small molecules to structures present in the PDB database, we asked if monomeric LPS-Ra could dock to sCD36 or sCEACAM1-4D. Although our LPS-CD36 or LPS-CEACAM1 binding studies were performed with micellar LPS, we considered docking monomeric LPS-Ra as a reasonable first step. In the case of sCD36 we found a structure in which two of the acyl chains of LPS-Ra inserted into a hydrophobic cavity of CD36 (**Fig 8**) that may normally accommodate an endogenous fatty acid such as the palmitic acid (among others) found in the original x-ray structure (75). Given that LPS-Ra has two beta 1,6-linked glucosamine residues, one with two acyl chains and the other with four (76), it appears that the size of the hydrophobic cavity of sCD36 can only accommodate the half of the lipid A portion of the molecule with two acyl chains and that the other half with four acyl chains are forced into the surface/aqueous environment. Since this arrangement is energetically unfavorable, we speculate that the four-acyl chains portion of the lipid A could be more favorably accommodated in a retained LPS micelle. Returning to the monomeric LPS structure shown in **Fig 8**, we noticed that two tyrosine residues, Tyr-192 and Tyr-203 on CD36 loops may interact with the two phosphates on the glucosamine dimer of LPS plus an additional phosphate on one of the sugars found in the LPS-Ra structure (**Supplemental Fig S6**).

As a first attempt to making a model of the LPS-Ra-CD36 micelle, we first considered the well described micelle of LPS-Ra with deoxycholate of an approximate molecular size 20 kDa (70). Assuming an approximate molecular weight of an LPS-Ra monomer as 5 kDa, the micelle model shown in **Fig 9A**, includes 4 LPS-Ra units (two shown in the xy plane, two not shown along the ±z- axis). Prior to addition of deoxycholate, the LPS-Ra size was much larger and rather undefined (70). By expanding model A without deoxycholate to include more LPS-Ra units the micelle can grow in size, but may expose acyl chains at its ends leading to more complex particle sizes. However, if proteins such as CD36 can bind exposed hydrophobic acyl

chains, then the end problem could be solved, especially if only one half of the dimeric structure of LPS-Ra was exposed at the ends of the micelle. We propose that a simple rotation about the glycosidic bond of the hexosamine dimer would expose just two acyl chains, the situation we observed in the Apha Fold structure of the CD36-LPS-Ra complex shown in **Fig 8**. An abbreviated model of this situation for just LPS-Ra (CD36 not shown) is shown in **Fig 9B**. We believe this model explains the rather uniform size of the CD36-LPS-Ra complex observed in both the SEC (**Fig 4**) and TEM (**Fig 5**) studies. Since our SPR studies show BAs do not interfere with LPS-Ra binding to CD36, we speculate that the final size of the CD36-LPS-Ra complex would be even less with added BAs, exactly the case observed in our SPR studies. Further studies are required to validate this model.

Molecular modeling of LPS-Ra monomer binding to sCEACAM1-N

Although an association of the unglycosylated N-domain of CEACAM1 (expressed in *E. coli*) with LPS (the smooth form), *as well as E. coli*, was shown by the Servin group in 2011 (40), their observation received little attention perhaps due to the well-known property of the CEACAM1 N-domain to form dimers and participate in homophilic cell-cell interactions (7) and its binding to *E. coli* Dr adhesins (77). While the binding motif of several LPS binding proteins such as LBP (72) and MD2 (66) are known, the molecular details of LPS binding to the CEACAM1 N-domain presents a challenge. We approached this problem by asking if Alpha Fold could dock the lipid A moiety of LPS-Ra to sCEACASM1-4D (**Supplemental Fig S7**), but we were not satisfied with the results that exposed so many acyl chains to the aqueous environment or the protein surface. On reflection, we considered that a unique, functionally unexplained, feature of the N-domain, namely the lack of a structure stabilizing disulfide bond in its IgV-like N-domain, may accommodate the acyl chains of lipid A if the hydrophobic faces of its beta sheet sandwich could adopt an open structure. An example of a beta sheet protein that does exactly that is MD2 (78), the critical LPS binding protein in the TLR4 complex (66).

As a first step in exploring this idea, and starting from the predicted x-ray structure of the N-domain (8) we performed a molecular dynamics study of the N-domain out to 2 μsec that revealed an ensemble of open structures (**Fig 10A-E**). However, none of these opened the beta-sandwich wide enough to accommodate the acyl chains of LPS due to the presence of the Asp-82 to Arg-95 salt bridge at the distal region of the N-domain (79). This observation led us to examine residues surrounding the salt bridge to see if there was an internal mechanism that

would facilitate breaking the salt bridge along with stabilization of an alternate open structure. In this regard, we noticed two key features that might impact the stability of the salt bridge, namely the presence of two nearby glycosylation sites (Asn-81 and Asn-77), one of which was adjacent to the Asp-82 of the salt bridge and present on a short alpha helix that might be expected to unwind if the glycans of LPS-Ra interacted with the glycans of the N-domain. As a first step, we introduced a D82A mutation to break the salt bridge and reran the MD simulation program to see if the beta-sheet sandwich would open as predicted. Using N81, the first glycosylation site, and N61 as distance markers, it can be seen by MD that the N-domain undergoes a stable conformation change in the μsec time frame (**Fig 11 A-B**). Using the open structure as a template, we next tried to dock the entire LPS-Ra structure, but it was too large for the program parameters. As a next approach we broke LPS-Ra into pieces, first docking lipid A without its acyl chains, then sequentially adding its flexible acyl chains and inner core glycan, we obtained a structure in which LPS-Ra had a good fit with the N-domain (**Fig 11C).** Note that this model places a portion of the lipid A moiety within the now open beta sandwich (**Fig 11D**). Details of the ribbon structure of the N-domain show key residues K35 and R43 making salt bridges with the two phosphates of docked LPS-Ra (**Fig 12A-B**). All contact residues with the lipid A moiety minus its acyl chains are shown in **Fig 12C**. Similar to CD36, we predict that the N-domain only interacts with a portion of the LPS-Ra micelle. Future studies will be directed at evaluating our model by site directed mutagenesis.

Interaction of sCEACAM1 with sCD36 by SPR

CD36 and CEACAM1 are co-expressed in several cell types that respond to LPS, including hepatocytes (27,42) and macrophages (80,81). In the case of intestinal enterocytes, the primary source of nutritional fatty acid uptake, both CEACAM1 and CD36 are expressed (13,82) and CD36 has been shown to uptake LPS along with fatty acids (13). The possibility that the two receptors could associate with each other was established by us in HepG2 hepatocytes by co-immunoprecipitation with specific antibodies (32). Indeed, SPR studies showed direct binding of immobilized sCD36 with sCEACAM1 with a K_D of 5.28 x 10^{-8} M. (**Fig 13**). However, when the two soluble proteins were incubated together and run on SEC, there was no evidence of complex formation. This is not a surprising result given that CEACAM1 dimer formation requires it to be in the membrane with a prior interaction between its TM domains (4). Therefore,

we sought further evidence that CEACAM1 and CD36 could directly interact by other approaches.

Proximity ligation of CEACAM1-4L with CD36

Two types of proximity ligation assays (PLA) were performed to determine if these two receptors were in tight association on cell membranes. In the first approach, specific antibodies to CEACAM1 and CD36 were ligated with complementary oligonucleotides that served as a template for amplification with fluorescent labeled probes on HepG2 hepatocytes that expressed both CEACAM1-4L (CEACAM1 with four extracellular domains and a 72-amino acid long cytoplasmic domain) and CD36 or CD36 only as a control. Analysis of the cells by confocal microscopy after PLA showed a strong cell surface signal for the CEACAM1-CD36 double positive sample vs the negative control lacking CEACAM1 (**Fig 14A-B**). In addition, HEK cells double transfected with plasmids for full length CEACAM1-4L and CD36 showed strong signals in the PLA assay vs untransfected HEK controls (**Fig 14C-D**). Flow analysis of the HEK cells for none, single, and double transfections are shown below.

In a second PLA assay involving biotinylation of adjacent cytoplasmic domains by BioID2 (83), HEK cells were transfected with plasmids expressing a C-terminal CD36-BioID2 fusion protein only, or CD36-BioID2 fusion protein (**Supplemental Fig S8**) plus CEACAM1-4L (52) followed by flow analysis with CEACAM1 and CD36 specific antibodies (**Fig 15A-C**). Since the flow analysis revealed a good co-expression of the double transfectants, we proceeded with the proximity labeling assay. The cell double transfectants were incubated with biotin, lysed, the biotinylated proteins isolated on streptavidin beads, and the bound and eluted proteins were immunoblotted with anti-CD36 or anti-CEACAM1 antibodies. The results show that the CD36-BioID2 transfectants are biotinylated in the singly transfected cells, and both CEACAM1 and CD36-BioID2 co-transfectants are biotinylated in the double transfected cells (**Fig 15D-E**).

CD36 has two transmembrane domains, each with short cytoplasmic tails of which the C-terminal cytoplasmic tail has a lysine residue that can be biotinylated by the fusion protein BioID2, thus explaining why cells transfected with CD36-BioID2 only were self-biotinylated. The multiple bands for CD36-BioID2 may be due to various glycoforms of CD36 and/or proteolysis of the fusion protein, but in any case, the presence of a single lysine in the C-terminal cytoplasmic domain of CD36 is a likely target of biotinylation by the BioID2 fusion partner. In addition, the

72-amino acid long cytoplasmic domain of CEACAM1-4L contains 4 lysine residues, two of which are at its distal end, making it an excellent target for biotinylation by CD36-BioID2.

The fact that two different PLAs, one directed at the extracellular side of the plasma membrane, and one directed at the cytoplasmic side of the plasma membrane, show CD36 and CEACAM1 are expressed in close proximity to each other, suggests that they can cooperate in cell signaling. This may explain their similar roles in regulating LPS-TLR4 signaling and their ability to bind LPS-Ra micelles.

Alpha Fold model of the CD36-CEACAM1 association

Since one of the features of Alpha Fold allows the docking of two previously published protein structures from the PDB data base, we were able to obtain an ensemble of structures for the N-domain of CEACAM1 docked to sCD36. One of the predicted structures replaced the original binding site of malarial protein PfEMP1 of *Plasmodium falciparum* (75) with the N-domain of CEACAM1 (**Fig 16A**). A prominent feature of this structure was the docking of critical residue F153 from sCD36 onto the hydrophobic GFCC'C'' face of the N-domain of CEACAM1 (**Fig 16B**), its preferred site of interaction (75). Additional interaction pairs in the N-domain of CEACAM1 to CD36 are shown in **Fig 16C**. However, since this structure would be incompatible with a cis interaction of the N-domain of CEACAM1 with CD36 as predicted from the PLA assays, we selected an alternative Alpha Fold predicted structure for consideration (**Fig 17A**). In this model, the N-domain of CEACAM1 interacts with the side of sCD36 and is in a better position to form a cis-interaction. Key residues between CD36 and the N-domain of CEACAM1 are shown in **Fig 17B**, along with a pair of residues on flexible loops that are likely to interact. Taking this model a step further, we used Alpha Fold to generate models of both full-length proteins including their transmembrane domains inserted into a generic plasma membrane (**Fig 18**). The model predicts that the cis-version of the CD36-CEACAM1 interaction places their transmembrane domains in proximity. In fact, if the TM domain of CEACAM1 closely associates with the N-terminal TM domain of CD36, its C-terminal TM domain could accommodate the fusion partner BioID2, enabling the biotin proximity labeling study (**Fig 15**). Since the formation of cis-dimers of CEACAM1 depend on the G-XXX-G motif in their TM domains (4), a sequence motif also found in the TM domains of CD36 (including an A-XXX-A motif), it is reasonable to speculate that CEACAM1 and CD36 self-associate via these shared motifs in their TM domains (**Supplemental Fig S9**). Although we selected the cis model of interaction to explain our PLA

data, we cannot rule out the possibility that trans interactions are possible when two cells interact. Future studies will examine this possibility.

DISCUSSION

Since its first description as a cell-cell adhesion molecule (CCAM-105) in rat hepatocytes by Obrink and co-workers (14,84) and later cloning studies that identified it as ratCEACAM1 (85), the over-riding function of CEACAM1 was ascribed to its homotypic cell-cell adhesive properties. Later, this function was naturally extended to the related proteins CEACAM5 and CEACAM6 (older names CEA and NCA or non-specific cross reacting antigen, respectively) that were shown to be related to CEACAM1 by cross-reactive antibodies that often made it difficult to distinguish the tumor associated antigen CEA from its normal tissue relatives (86). With the advent of cDNA cloning, the size of the gene family grew to over 20 genes in which their chief distinguishing feature was their structurally conserved IgV-like N-domains followed by a variable number of Ig-C2-like domains (87,88). Comprehensive structure-function studies showed that the N-domain conferred the cell-cell adhesive properties of the CEACAM gene family members but failed to account for the lack of the structure stabilizing disulfide bond found between the B and E beta-strands present in the majority of IgV-like family proteins (79). Although early studies by us showed bacteria binding to CEA family members (89-91), subsequent studies (35,92-106), focused attention on bacterial adhesins binding CEACAMs for epithelial invasion, rather than CEACAMs binding bacterial products for host defense.

A hint of a possible additional function for CEACAM1, besides cell-cell or interaction with bacterial adhesins, came early in the field when Hammarstrom and coworkers (28) identified a CEA-cross-reactive protein in bile that they named Biliary Glycoprotein 1 (BGP1). Upon protein sequencing (31) and molecular cloning studies (107), BGP1 was shown to have sequence identity to CEACAM1, including an IgV-like N-domain and IgC2-like domains, but lacking TM and cytoplasmic domains. Later, BGP1 was shown to bind cholesterol in bile (30), suggesting the possibility that CEACAM1 had lipid binding properties. The idea that CEACAM1 could bind other lipids such as LPS was first put forward by Servin and co-workers (40) who showed that CEACAM1 transfected epithelial cells bound LPS and that bacterial expressed CEACAM1 N-domain bound LPS in a SPR assay. In spite of this proposed novel function of CEACAM1, this finding received little attention, perhaps due to the prevailing dogma that CEACAM1 was just a

cell-cell adhesion protein and that bacterial expressed adhesins were responsible for bacterial binding to CEACAM1 family members (77).

It became clear to us that if the LPS binding function of CEACAM1, and by extension, the entire gene family, were true, then more comprehensive data was required. It is our hope that the studies presented here, although limited in scope to CEACAM1, will open the field to a new, more physiologically meaningful function for the CEACAM family, namely binding LPS, and by extension, other bacterial lipids. We propose that complex organisms that developed alongside exposure to environmental LPS, regarded LPS as both a friend and foe: friend as a ubiquitous source of energy due to its high content of both carbohydrates and fatty acids, and foe as a shed marker of live bacterial growth within the organism, at the least threatening infection, and at the worst, conferring lethality. In general, since slower growing complex organisms compete poorly with rapidly dividing bacteria, they used shed bacterial membranes, in this example LPS, as a marker of potential danger. Thus, they developed a detection system that could titrate LPS levels and warn the host when its levels exceeded a safe threshold. But as long as these levels were below the danger point, this detection system mainly functioned to neutralize LPS and/or utilize it as a source of energy. Later, it became obvious that conditions that increase intestinal permeability, including obesity, were a chronic source of inflammation caused by bacterial products such as LPS (108).

We propose that the expansion of the CEACAM1 gene family across numerous species and tissue expression (88) is evidence of its LPS binding function since LPS can be found on almost all epithelial tissues where CEACAM1 is expressed, especially since its immune function is induced by inflammatory cytokines such as IFNγ (109,110). This function may also explain why many strains of pathogenic bacteria have developed adhesin proteins that bind CEACAM1 (77,111), allowing them to subvert their host LPS detection/protective functions, and at the same time, invade the host epithelial barrier. This may be evidence of the “red queen” hypothesis (112). Since the presence of systemic LPS is associated with life-threatening sepsis (113), it is not surprising that T- and B-cells, as well as granulocytes and macrophages, express CEACAM1 and become activated upon binding LPS (38). The apparent mystery that engagement of CEACAM1 by LPS results in an inhibitory response via its ITIMs (38) is not that strange when considering that CEACAM1 helps set the *threshold* of an immune response, unlike TLR4 whose highly orchestrated engagement generates a robust immune response. By highly orchestrated we mean the sequential action of LPS binding proteins such as LBP and

CD14 are required to produce a monomeric species of LPS (18) rather than LPS micelles that can bind CEACAM1.

The expression of CEACAM1 in the intestine where LPS is continuously exposed from dying gram negative bacteria is of special significance. Lipid micelles, including LPS, are solubilized upon exposure to bile acids (114), and as shown in this study, can be taken up as discrete particles by CEACAM1 even in the absence of bile acids. Although the fate of the LPS micelles bound to CEACAM1 on enterocytes was not addressed in this study, we can speculate on at least two fates. In fate one, LPS micelles are internalized in cooperation with CD36, a receptor already implicated in the intestinal uptake of LPS (13). Once internalized, the LPS micelles may be incorporated into newly forming chylomicrons that are secreted in the lacteal lymphatics. It is well known that LPS can be detected in chylomicrons, VLDL, LDL and HDL, where it is shown to be non-pathogenic (17). Once these lipid particles reach the liver, they are repackaged as VLDL and LDL and the LPS excreted in bile (23). If the LPS reaches a high enough level, hepatocytes might produce the enzyme acyloxyacylhydrolase that removes acyl chains from the lipid A moiety, resulting in a further detoxification mechanism (115). In any case, bile acids play a major role in the management of LPS in both the intestine and liver. In fate two, the LPS from membrane bound CEACAM1 may be exchanged with soluble BGP1 present in the bile. We can speculate that this soluble form of CEACAM1 plays an important role in binding LPS in both the bile where LPS is present (116,117) and in the intestine, resulting in the combined biliary-intestinal excretion of LPS. Systemic circulation of BGP1 may also play a role in LPS clearance and is worth further functional studies.

Since both cholesterol and LPS are subject to reverse transport in bile, they may compete for each other. If levels of one rise, the other's removal may be impaired leading to higher levels of both. Thus, high levels of cholesterol would lead to high levels of LPS, and vice versa. Either situation is undesirable and their interdependence may not be appreciated in promoting better health. We speculate that treatment of high cholesterol levels with CEACAM1, especially a soluble form, may be an option for patients with poor statin control. It is also possible that increased intestinal permeability that leads to chronic low levels of LPS may compete with cholesterol removal and require the use of statins.

The fact that the structurally distinct CEACAM1 and CD36 share in binding LPS and can interact with each other is reminiscent of the role of multiple LPS binding proteins in the transfer of LPS monomers to TLR4. This common theme suggests that LPS per se is not dangerous until it

exceeds a threshold, otherwise the host would be in a constant state of inflammation. Indeed, abrogation of either the *Ceacam1* or *Cd36* gene in mice leads to smoldering inflammation that increases with age and lipid consumption. In the case of liver specific *Ceacam1* LKO mice (118), just as in global *Ceacam1* KO mice (119), its absence leads to obesity, insulin resistance and kidney and vascular pathologies *even on a normal diet over time*. While the functional absence of CD36 in certain human populations does not cause life threatening pathologies (120), there are no examples of loss of the human *CEACAM1* gene, suggesting that its function is essential to life. Since mice have two alleles of *Ceacam1*, as well as a second gene (*Ceacam2*), the loss of one is compensated by the expression of the other (121). It is also worth mentioning that LPS species from many commensal bacteria are unable to deliver an inflammatory response via TLR4 signaling, suggesting that the engagement of different species of LPS with CEACAM1 and CD36 may be worth studying. Furthermore, the expression of both CEACAM1 (122) and CD36 (123) on the surface of platelets is functionally important, linking LPS to capillary clotting during infection (124,125).

Although we focused our studies on the binding of LPS-Ra to CEACAM1 and CD36, we hypothesize that their binding functions are not limited to LPS from gram negative bacteria, but should also extend to gram positive lipids such as teichoic acid and TLR1/2 binding lipopeptides such as Pam3CSK4. In this respect, CEACAM1 is known to associate with TLR2 (126), supporting our hypothesis. Therefore, further lipid binding studies with CEACAM1 are required, as well as their complexes with bile acids. In addition, we speculate that each of the bile acids may have a preferential binding to each of the bacterial lipids, partly explaining their species specificity in different organisms reflecting not only their diets but also their diet's contamination with different bacteria. In this respect, we observed a stronger binding of deoxycholate/LPS-Ra micelles than of cholate/LPS-Ra micelles to both sCEACAM1 and sCD36. Although deoxycholate is a bacterially produced secondary bile acid, its concentration in bile is nearly as high as cholate itself (127).

Finally, it is worth pointing out that chronic over consumption of nutritional lipids that can lead to obesity and type 2 diabetes (T2D) is associated with high levels of circulating LPS (128,129). Interventions such as gastric sleeve reduction of stomach volume has been shown to not only reverse T2D in mice (130) and in man (131), but also can return serum LPS levels to normal (50). This apparent causal relationship among LPS, T2D and gastric sleeve surgery suggests that management of LPS uptake/excretion in the intestine and liver may be an excellent target for the prevention and management of T2D. A model of the uptake and recycling/excretion of

LPS in the small intestine (**Fig 19**) builds on the known uptake of LPS by CD36 (48) and our postulated role of membrane bound CEACAM1 and the secreted form of CEACAM1 (BGP1) binding LPS. The model predicts that CD36 is mainly responsible for LPS uptake, while CEACAM1/BGP1 is responsible for LPS recycling/excretion. Based on this model, further studies on the role of both CD36 and CEACAM1 in lipid management are warranted. Furthermore, we suggest that recombinant forms of CEACAM1, especially its N-domain, as a natural substance or probiotic could be used to treat diseases associated with intestinal leakage of bacterial lipids, as exemplified by LPS.

REFERENCES

1. Gray-Owen, S. D., and Blumberg, R. S. (2006) CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol* **6**, 433-446
2. Lobo, E. O., Zhang, Z., and Shively, J. E. (2009) Pivotal advance: CEACAM1 is a negative coreceptor for the B cell receptor and promotes CD19-mediated adhesion of B cells in a PI3K-dependent manner. *J Leukoc Biol* **86**, 205-218
3. Greicius, G., Severinson, E., Beauchemin, N., Obrink, B., and Singer, B. B. (2003) CEACAM1 is a potent regulator of B cell receptor complex-induced activation. *J Leukoc Biol* **74**, 126-134
4. Patel, P. C., Lee, H. S., Ming, A. Y., Rath, A., Deber, C. M., Yip, C. M., Rocheleau, J. V., and Gray-Owen, S. D. (2013) Inside-out signaling promotes dynamic changes in the carcinoembryonic antigen-related cellular adhesion molecule 1 (CEACAM1) oligomeric state to control its cell adhesion properties. *J Biol Chem* **288**, 29654-29669
5. Kim, W. M., Huang, Y. H., Gandhi, A., and Blumberg, R. S. (2019) CEACAM1 structure and function in immunity and its therapeutic implications. *Semin Immunol* **42**, 101296
6. Gandhi, A. K., Huang, Y. H., Sun, Z. J., Kim, W. M., Kondo, Y., Hanley, T., Beauchemin, N., and Blumberg, R. S. (2024) Structural aspects of CEACAM1 interactions. *Eur J Clin Invest* **54 Suppl 2**, e14357
7. Gandhi, A. K., Sun, Z. J., Huang, Y. H., Kim, W. M., Yang, C., Petsko, G. A., Beauchemin, N., and Blumberg, R. S. (2022) Structural analysis of human CEACAM1 oligomerization. *Commun Biol* **5**, 1042
8. Gandhi, A. K., Sun, Z. J., Kim, W. M., Huang, Y. H., Kondo, Y., Bonsor, D. A., Sundberg, E. J., Wagner, G., Kuchroo, V. K., Petsko, G. A., and Blumberg, R. S. (2021) Structural basis of the dynamic human CEACAM1 monomer-dimer equilibrium. *Commun Biol* **4**, 360

9. Schumann, D., Huang, J., Clarke, P. E., Kirshner, J., Tsai, S. W., Schumaker, V. N., and Shively, J. E. (2004) Characterization of recombinant soluble carcinoembryonic antigen cell adhesion molecule 1. *Biochem Biophys Res Commun* **318**, 227-233
10. Zhuo, Y., Yang, J. Y., Moremen, K. W., and Prestegard, J. H. (2016) Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1). *J Biol Chem* **291**, 20085-20095
11. Zhuo, Y., Yang, J. Y., Moremen, K. W., and Prestegard, J. H. (2020) Correction: Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). *J Biol Chem* **295**, 3748
12. Kumar, P., Schroder, E. A., Rajaram, M. V. S., Harris, E. N., and Ganesan, L. P. (2024) The Battle of LPS Clearance in Host Defense vs. Inflammatory Signaling. *Cells* **13**
13. Cifarelli, V., and Abumrad, N. A. (2018) Intestinal CD36 and Other Key Proteins of Lipid Utilization: Role in Absorption and Gut Homeostasis. *Compr Physiol* **8**, 493-507
14. Hansson, M., Blikstad, I., and Obrink, B. (1989) Cell-surface location and molecular properties of cell-CAM 105 in intestinal epithelial cells. *Exp Cell Res* **181**, 63-74
15. Schirbel, A., Rebert, N., Sadler, T., West, G., Rieder, F., Wagener, C., Horst, A., Sturm, A., de la Motte, C., and Fiocchi, C. (2019) Mutual Regulation of TLR/NLR and CEACAM1 in the Intestinal Microvasculature: Implications for IBD Pathogenesis and Therapy. *Inflamm Bowel Dis* **25**, 294-305
16. Vors, C., Pineau, G., Drai, J., Meugnier, E., Pesenti, S., Laville, M., Laugerette, F., Malpuech-Brugere, C., Vidal, H., and Michalski, M. C. (2015) Postprandial Endotoxemia Linked With Chylomicrons and Lipopolysaccharides Handling in Obese Versus Lean Men: A Lipid Dose-Effect Trial. *J Clin Endocrinol Metab* **100**, 3427-3435
17. Levels, J. H., Marquart, J. A., Abraham, P. R., van den Ende, A. E., Molhuizen, H. O., van Deventer, S. J., and Meijers, J. C. (2005) Lipopolysaccharide is transferred from high-density to low-density lipoproteins by lipopolysaccharide-binding protein and phospholipid transfer protein. *Infect Immun* **73**, 2321-2326
18. Park, B. S., and Lee, J. O. (2013) Recognition of lipopolysaccharide pattern by TLR4 complexes. *Exp Mol Med* **45**, e66
19. Schumann, R. R., Kirschning, C. J., Unbehaun, A., Aberle, H. P., Knope, H. P., Lamping, N., Ulevitch, R. J., and Herrmann, F. (1996) The lipopolysaccharide-binding protein is a secretory class 1 acute-phase protein whose gene is transcriptionally activated by APRF/STAT/3 and other cytokine-inducible nuclear proteins. *Mol Cell Biol* **16**, 3490-3503
20. Landmann, R., Reber, A. M., Sansano, S., and Zimmerli, W. (1996) Function of soluble CD14 in serum from patients with septic shock. *J Infect Dis* **173**, 661-668
21. Garcia-Monzon, C., Lo Iacono, O., Crespo, J., Romero-Gomez, M., Garcia-Samaniego, J., Fernandez-Bermejo, M., Dominguez-Diez, A., Rodriguez de Cia, J., Saez, A., Porrero, J. L., Vargas-Castrillon, J., Chavez-Jimenez, E., Soto-Fernandez, S., Diaz, A., Gallego-Duran, R., Madejon, A., and Miquilena-Colina, M. E. (2014) Increased soluble CD36 is linked to advanced steatosis in nonalcoholic fatty liver disease. *Eur J Clin Invest* **44**, 65-73
22. Draberova, L., Cerna, H., Brodska, H., Boubelik, M., Watt, S. M., Stanners, C. P., and Draber, P. (2000) Soluble isoforms of CEACAM1 containing the A2 domain: increased serum levels in patients with obstructive jaundice and differences in 3-fucosyl-N-acetyl-lactosamine moiety. *Immunology* **101**, 279-287
23. Osnes, T., Kierulf, P., Skar, V., Skar, A. G., Ovstebo, R., and Osnes, M. (1992) Quantification of lipopolysaccharides in human bile with or without gram-negative bacteria. *Scand J Gastroenterol* **27**, 453-458

24. Pownall, H. J., Rosales, C., Gillard, B. K., and Gotto, A. M., Jr. (2021) High-density lipoproteins, reverse cholesterol transport and atherogenesis. *Nat Rev Cardiol* **18**, 712-723
25. Azzam, K. M., and Fessler, M. B. (2012) Crosstalk between reverse cholesterol transport and innate immunity. *Trends Endocrinol Metab* **23**, 169-178
26. Shibao, C. A., Celedonio, J. E., Tamboli, R., Sidani, R., Love-Gregory, L., Pietka, T., Xiong, Y., Wei, Y., Abumrad, N. N., Abumrad, N. A., and Flynn, C. R. (2018) CD36 Modulates Fasting and Preabsorptive Hormone and Bile Acid Levels. *J Clin Endocrinol Metab* **103**, 1856-1866
27. Horst, A. K., Najjar, S. M., Wagener, C., and Tiegs, G. (2018) CEACAM1 in Liver Injury, Metabolic and Immune Regulation. *Int J Mol Sci* **19**
28. Svenberg, T., Hammarstrom, S., and Hedin, A. (1979) Purification and properties of biliary glycoprotein I (BGP I). Immunochemical relationship to carcinoembryonic antigen. *Mol Immunol* **16**, 245-252
29. Svenberg, T., Wahren, B., and Hammarstrom, S. (1979) Elevated serum levels of a biliary glycoprotein (BGP I) in patients with liver or biliary tract disease. *Clin Exp Immunol* **36**, 317-325
30. Jirsa, M., Muchova, L., Draberova, L., Draber, P., Smid, F., Kuroki, M., Marecek, Z., and Groen, A. K. (2001) Carcinoembryonic antigen-related cell adhesion molecule 1 is the 85-kilodalton pronase-resistant biliary glycoprotein in the cholesterol crystallization promoting low density protein-lipid complex. *Hepatology* **34**, 1075-1082
31. Hinoda, Y., Neumaier, M., Hefta, S. A., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L. J., Shively, J. E., and Paxton, R. J. (1988) Molecular cloning of a cDNA coding biliary glycoprotein I: primary structure of a glycoprotein immunologically crossreactive with carcinoembryonic antigen. *Proc Natl Acad Sci U S A* **85**, 6959-6963
32. Chean, J., Chen, C. J., Gugiu, G., Wong, P., Cha, S., Li, H., Nguyen, T., Bhatticharya, S., and Shively, J. E. (2021) Human CEACAM1-LF regulates lipid storage in HepG2 cells via fatty acid transporter CD36. *J Biol Chem* **297**, 101311
33. Boulton, I. C., and Gray-Owen, S. D. (2002) Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. *Nat Immunol* **3**, 229-236
34. Lee, H. S., Ostrowski, M. A., and Gray-Owen, S. D. (2008) CEACAM1 dynamics during neisseria gonorrhoeae suppression of CD4+ T lymphocyte activation. *Journal of immunology* **180**, 6827-6835
35. Voges, M., Bachmann, V., Kammerer, R., Gophna, U., and Hauck, C. R. (2010) CEACAM1 recognition by bacterial pathogens is species-specific. *BMC Microbiol* **10**, 117
36. van Sorge, N. M., Bonsor, D. A., Deng, L., Lindahl, E., Schmitt, V., Lyndin, M., Schmidt, A., Nilsson, O. R., Brizuela, J., Boero, E., Sundberg, E. J., van Strijp, J. A. G., Doran, K. S., Singer, B. B., Lindahl, G., and McCarthy, A. J. (2021) Bacterial protein domains with a novel Ig-like fold target human CEACAM receptors. *EMBO J* **40**, e106103
37. Saberi, S., Shans, N., Ghaffari, S., Esmaeili, M., and Mohammadi, M. (2024) The role of CEACAMs versus integrins in Helicobacter pylori CagA translocation: a systematic review. *Microbes Infect* **26**, 105246
38. Lu, R., Pan, H., and Shively, J. E. (2012) CEACAM1 negatively regulates IL-1beta production in LPS activated neutrophils by recruiting SHP-1 to a SYK-TLR4-CEACAM1 complex. *PLoS Pathog* **8**, e1002597
39. Zhang, Z., La Placa, D., Nguyen, T., Kujawski, M., Le, K., Li, L., and Shively, J. E. (2019) CEACAM1 regulates the IL-6 mediated fever response to LPS through the RP105 receptor in murine monocytes. *BMC Immunol* **20**, 7
40. Lievin-Le Moal, V., Beau, I., Rougeaux, C., Kansau, I., Fabrega, S., Brice, C., Korotkova, N., Moseley, S. L., and Servin, A. L. (2011) Apical expression of human full-length

hCEACAM1-4L protein renders the Madin Darby Canine Kidney cells responsive to lipopolysaccharide leading to TLR4-dependent Erk1/2 and p38 MAPK signalling. *Cell Microbiol* **13**, 764-785
41. Heinrich, G., Ghadieh, H. E., Ghanem, S. S., Muturi, H. T., Rezaei, K., Al-Share, Q. Y., Bowman, T. A., Zhang, D., Garofalo, R. S., Yin, L., and Najjar, S. M. (2017) Loss of Hepatic CEACAM1: A Unifying Mechanism Linking Insulin Resistance to Obesity and Non-Alcoholic Fatty Liver Disease. *Front Endocrinol (Lausanne)* **8**, 8
42. Wilson, C. G., Tran, J. L., Erion, D. M., Vera, N. B., Febbraio, M., and Weiss, E. J. (2016) Hepatocyte-Specific Disruption of CD36 Attenuates Fatty Liver and Improves Insulin Sensitivity in HFD-Fed Mice. *Endocrinology* **157**, 570-585
43. Muller, M. M., Klaile, E., Vorontsova, O., Singer, B. B., and Obrink, B. (2009) Homophilic adhesion and CEACAM1-S regulate dimerization of CEACAM1-L and recruitment of SHP-2 and c-Src. *J Cell Biol* **187**, 569-581
44. Hu, W., Bagramyan, K., Bhatticharya, S., Hong, T., Tapia, A., Wong, P., Kalkum, M., and Shively, J. E. (2021) Phosphorylation of human CEACAM1-LF by PKA and GSK3beta promotes its interaction with beta-catenin. *J Biol Chem* **297**, 101305
45. Hao, J. W., Wang, J., Guo, H., Zhao, Y. Y., Sun, H. H., Li, Y. F., Lai, X. Y., Zhao, N., Wang, X., Xie, C., Hong, L., Huang, X., Wang, H. R., Li, C. B., Liang, B., Chen, S., and Zhao, T. J. (2020) CD36 facilitates fatty acid uptake by dynamic palmitoylation-regulated endocytosis. *Nat Commun* **11**, 4765
46. Peche, V. S., Pietka, T. A., Jacome-Sosa, M., Samovski, D., Palacios, H., Chatterjee-Basu, G., Dudley, A. C., Beatty, W., Meyer, G. A., Goldberg, I. J., and Abumrad, N. A. (2023) Endothelial cell CD36 regulates membrane ceramide formation, exosome fatty acid transfer and circulating fatty acid levels. *Nat Commun* **14**, 4029
47. Chavez-Sanchez, L., Garza-Reyes, M. G., Espinosa-Luna, J. E., Chavez-Rueda, K., Legorreta-Haquet, M. V., and Blanco-Favela, F. (2014) The role of TLR2, TLR4 and CD36 in macrophage activation and foam cell formation in response to oxLDL in humans. *Hum Immunol* **75**, 322-329
48. Cifarelli, V., Ivanov, S., Xie, Y., Son, N. H., Saunders, B. T., Pietka, T. A., Shew, T. M., Yoshino, J., Sundaresan, S., Davidson, N. O., Goldberg, I. J., Gelman, A. E., Zinselmeyer, B. H., Randolph, G. J., and Abumrad, N. A. (2017) CD36 deficiency impairs the small intestinal barrier and induces subclinical inflammation in mice. *Cell Mol Gastroenterol Hepatol* **3**, 82-98
49. Shibao, C. A., Celedonio, J. E., Ramirez, C. E., Love-Gregory, L., Arnold, A. C., Choi, L., Okamoto, L. E., Gamboa, A., Biaggioni, I., Abumrad, N. N., and Abumrad, N. A. (2016) A Common CD36 Variant Influences Endothelial Function and Response to Treatment with Phosphodiesterase 5 Inhibition. *J Clin Endocrinol Metab* **101**, 2751-2758
50. Clemente-Postigo, M., Roca-Rodriguez Mdel, M., Camargo, A., Ocana-Wilhelmi, L., Cardona, F., and Tinahones, F. J. (2015) Lipopolysaccharide and lipopolysaccharide-binding protein levels and their relationship to early metabolic improvement after bariatric surgery. *Surg Obes Relat Dis* **11**, 933-939
51. Troseid, M., Nestvold, T. K., Rudi, K., Thoresen, H., Nielsen, E. W., and Lappegard, K. T. (2013) Plasma lipopolysaccharide is closely associated with glycemic control and abdominal obesity: evidence from bariatric surgery. *Diabetes Care* **36**, 3627-3632
52. Jin, L., Li, Y., Chen, C. J., Sherman, M. A., Le, K., and Shively, J. E. (2008) Direct interaction of tumor suppressor CEACAM1 with beta catenin: identification of key residues in the long cytoplasmic domain. *Exp Biol Med (Maywood)* **233**, 849-859
53. Li, C., Chen, C. J., and Shively, J. E. (2009) Mutational analysis of the cytoplasmic domain of CEACAM1-4L in humanized mammary glands reveals key residues involved in lumen formation: stimulation by Thr-457 and inhibition by Ser-461. *Exp Cell Res* **315**, 1225-1233

54. La Verde, V., Dominici, P., and Astegno, A. (2017) Determination of Hydrodynamic Radius of Proteins by Size Exclusion Chromatography. *Bio Protoc* **7**, e2230
55. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675
56. Sorzano, C. O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H., Carazo, J. M., and Pascual-Montano, A. (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**, 194-204
57. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500
58. Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., Forouzesh, N., Giese, T. J., Götz, A. W., Gohlke, H., Izadi, S., Kasavajhala, K., Kaymak, M. C., King, E., Kurtzman, T., Lee, T.-S., Li, P., Liu, J., Luchko, T., Luo, R., Manathunga, M., Machado, M. R., Nguyen, H. M., O'Hearn, K. A., Onufriev, A. V., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Risheh, A., Schott-Verdugo, S., Shajan, A., Swails, J., Wang, J., Wei, H., Wu, X., Wu, Y., Zhang, S., Zhao, S., Zhu, Q., Cheatham, T. E., III, Roe, D. R., Roitberg, A., Simmerling, C., York, D. M., Nagan, M. C., and Merz, K. M., Jr. (2023) AmberTools. *Journal of Chemical Information and Modeling* **63**, 6183-6191
59. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020) ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **16**, 528-552
60. Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics* **120**, 9665-9678
61. Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **11**, 1864-1874
62. Grant, O. C., Wentworth, D., Holmes, S. G., Kandel, R., Sehnal, D., Wang, X., Xiao, Y., Sheppard, P., Grelsson, T., Coulter, A., Miller, G., Foley, B. L., and Woods, R. J. (2025) Generating 3D Models of Carbohydrates with GLYCAM-Web. *bioRxiv*
63. Park, S. J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., Joung, I., Joo, K., Lee, J., and Im, W. (2019) CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320-331
64. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71-73
65. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749

66. Park, B. S., Song, D. H., Kim, H. M., Choi, B. S., Lee, H., and Lee, J. O. (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191-1195
67. Nap, M., Hammarstrom, M. L., Bormer, O., Hammarstrom, S., Wagener, C., Handt, S., Schreyer, M., Mach, J. P., Buchegger, F., von Kleist, S., and et al. (1992) Specificity and affinity of monoclonal antibodies against carcinoembryonic antigen. *Cancer Res* **52**, 2329-2339
68. Cao, D., Luo, J., Chen, D., Xu, H., Shi, H., Jing, X., and Zang, W. (2016) CD36 regulates lipopolysaccharide-induced signaling pathways and mediates the internalization of Escherichia coli in cooperation with TLR4 in goat mammary gland epithelial cells. *Sci Rep* **6**, 23132
69. Jay, A. G., Chen, A. N., Paz, M. A., Hung, J. P., and Hamilton, J. A. (2015) CD36 binds oxidized low density lipoprotein (LDL) in a mechanism dependent upon fatty acid binding. *J Biol Chem* **290**, 4590-4603
70. Shands, J. W., Jr., and Chun, P. W. (1980) The dispersion of gram-negative lipopolysaccharide by deoxycholate. Subunit molecular weight. *J Biol Chem* **255**, 1221-1226
71. Stewart, C. R., Stuart, L. M., Wilkinson, K., van Gils, J. M., Deng, J., Halle, A., Rayner, K. J., Boyer, L., Zhong, R., Frazier, W. A., Lacy-Hulbert, A., El Khoury, J., Golenbock, D. T., and Moore, K. J. (2010) CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat Immunol* **11**, 155-161
72. Eckert, J. K., Kim, Y. J., Kim, J. I., Gurtler, K., Oh, D. Y., Sur, S., Lundvall, L., Hamann, L., van der Ploeg, A., Pickkers, P., Giamarellos-Bourboulis, E., Kubarenko, A. V., Weber, A. N., Kabesch, M., Kumpf, O., An, H. J., Lee, J. O., and Schumann, R. R. (2013) The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity. *Immunity* **39**, 647-660
73. Noronha, M. A., D'Angelo, N. A., Pérez-Sánchez, G., Severino, P., Foglio, M. A., Greaves, T. L., Pereira, J. F. B., and Lopes, A. M. (2023) Self-assembling micelles of lipopolysaccharides (LPS) for loading hydrophobic (bio)molecules. *J Mol Liq* **372**
74. An, J., Kim, S. H., Hwang, D., Lee, K. E., Kim, M. J., Yang, E. G., Kim, S. Y., and Chung, H. S. (2019) Caspase-4 disaggregates lipopolysaccharide micelles via LPS-CARD interaction. *Sci Rep* **9**, 826
75. Hsieh, F. L., Turner, L., Bolla, J. R., Robinson, C. V., Lavstsen, T., and Higgins, M. K. (2016) The structural basis for CD36 binding by the malaria parasite. *Nat Commun* **7**, 12837
76. Gorman, A., and Golovanov, A. P. (2022) Lipopolysaccharide Structure and the Phenomenon of Low Endotoxin Recovery. *Eur J Pharm Biopharm* **180**, 289-307
77. Korotkova, N., Yang, Y., Le Trong, I., Cota, E., Demeler, B., Marchant, J., Thomas, W. E., Stenkamp, R. E., Moseley, S. L., and Matthews, S. (2008) Binding of Dr adhesins of Escherichia coli to carcinoembryonic antigen triggers receptor dissociation. *Mol Microbiol* **67**, 420-434
78. Ohto, U., Fukase, K., Miyake, K., and Satow, Y. (2007) Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **316**, 1632-1634
79. Watt, S. M., Teixeira, A. M., Zhou, G. Q., Doyonnas, R., Zhang, Y., Grunert, F., Blumberg, R. S., Kuroki, M., Skubitz, K. M., and Bates, P. A. (2001) Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site. *Blood* **98**, 1469-1479
80. Silverstein, R. L., and Febbraio, M. (2009) CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Sci Signal* **2**, re3
81. Samineni, S., Zhang, Z., and Shively, J. E. (2013) Carcinoembryonic antigen-related cell adhesion molecule 1 negatively regulates granulocyte colony-stimulating factor

production by breast tumor-associated macrophages that mediate tumor angiogenesis. *Int J Cancer* **133**, 394-407
82. Terahara, K., Yoshida, M., Taguchi, F., Igarashi, O., Nochi, T., Gotoh, Y., Yamamoto, T., Tsunetsugu-Yokota, Y., Beauchemin, N., and Kiyono, H. (2009) Expression of newly identified secretory CEACAM1(a) isoforms in the intestinal epithelium. *Biochem Biophys Res Commun* **383**, 340-346
83. Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., and Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **27**, 1188-1196
84. Edlund, M., Gaardsvoll, H., Bock, E., and Obrink, B. (1993) Different isoforms and stock-specific variants of the cell adhesion molecule C-CAM (cell-CAM 105) in rat liver. *Eur J Biochem* **213**, 1109-1116
85. Budt, M., Michely, B., Muller, M. M., Reutter, W., and Lucka, L. (2002) Secreted CEACAM1 splice variants in rat cell lines and in vivo in rat serum. *Biochem Biophys Res Commun* **292**, 749-755
86. Engvall, E., Shively, J. E., and Wrann, M. (1978) Isolation and characterization of the normal crossreacting antigen: homology of its NH2-terminal amino acid sequence with that of carcinoembryonic antigen. *Proc Natl Acad Sci U S A* **75**, 1670-1674
87. Hammarstrom, S. (1999) The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* **9**, 67-81
88. Kammerer, R., and Zimmermann, W. (2010) Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biol* **8**, 12
89. Leusch, H. G., Drzeniek, Z., Markos-Pusztai, Z., and Wagener, C. (1991) Binding of Escherichia coli and Salmonella strains to members of the carcinoembryonic antigen family: differential binding inhibition by aromatic alpha-glycosides of mannose. *Infect Immun* **59**, 2051-2057
90. Sauter, S. L., Rutherfurd, S. M., Wagener, C., Shively, J. E., and Hefta, S. A. (1993) Identification of the specific oligosaccharide sites recognized by type 1 fimbriae from Escherichia coli on nonspecific cross-reacting antigen, a CD66 cluster granulocyte glycoprotein. *J Biol Chem* **268**, 15510-15516
91. Sauter, S. L., Rutherfurd, S. M., Wagener, C., Shively, J. E., and Hefta, S. A. (1991) Binding of nonspecific cross-reacting antigen, a granulocyte membrane glycoprotein, to Escherichia coli expressing type 1 fimbriae. *Infect Immun* **59**, 2485-2493
92. Virji, M., Watt, S. M., Barker, S., Makepeace, K., and Doyonnas, R. (1996) The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of Neisseria meningitidis and Neisseria gonorrhoeae. *Mol Microbiol* **22**, 929-939
93. Virji, M., Evans, D., Hadfield, A., Grunert, F., Teixeira, A. M., and Watt, S. M. (1999) Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules. *Mol Microbiol* **34**, 538-551
94. Virji, M., Evans, D., Griffith, J., Hill, D., Serino, L., Hadfield, A., and Watt, S. M. (2000) Carcinoembryonic antigens are targeted by diverse strains of typable and non-typable Haemophilus influenzae. *Mol Microbiol* **36**, 784-795
95. Bos, M. P., Kuroki, M., Krop-Watorek, A., Hogan, D., and Belland, R. J. (1998) CD66 receptor specificity exhibited by neisserial Opa variants is controlled by protein determinants in CD66 N-domains. *Proc Natl Acad Sci U S A* **95**, 9584-9589
96. Hill, D. J., and Virji, M. (2003) A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1. *Mol Microbiol* **48**, 117-129

97. Bos, M. P., Hogan, D., and Belland, R. J. (1999) Homologue scanning mutagenesis reveals CD66 receptor residues required for neisserial Opa protein binding. *J Exp Med* **190**, 331-340
98. Berger, C. N., Billker, O., Meyer, T. F., Servin, A. L., and Kansau, I. (2004) Differential recognition of members of the carcinoembryonic antigen family by Afa/Dr adhesins of diffusely adhering Escherichia coli (Afa/Dr DAEC). *Mol Microbiol* **52**, 963-983
99. Javaheri, A., Kruse, T., Moonens, K., Mejias-Luque, R., Debraekeleer, A., Asche, C. I., Tegtmeyer, N., Kalali, B., Bach, N. C., Sieber, S. A., Hill, D. J., Koniger, V., Hauck, C. R., Moskalenko, R., Haas, R., Busch, D. H., Klaile, E., Slevogt, H., Schmidt, A., Backert, S., Remaut, H., Singer, B. B., and Gerhard, M. (2016) Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nat Microbiol* **2**, 16189
100. Kuespert, K., Roth, A., and Hauck, C. R. (2011) Neisseria meningitidis has two independent modes of recognizing its human receptor CEACAM1. *PLoS One* **6**, e14609
101. Behrens, I. K., Busch, B., Ishikawa-Ankerhold, H., Palamides, P., Shively, J. E., Stanners, C., Chan, C., Leung, N., Gray-Owen, S., and Haas, R. (2020) The HopQ-CEACAM Interaction Controls CagA Translocation, Phosphorylation, and Phagocytosis of Helicobacter pylori in Neutrophils. *mBio* **11**
102. Koniger, V., Holsten, L., Harrison, U., Busch, B., Loell, E., Zhao, Q., Bonsor, D. A., Roth, A., Kengmo-Tchoupa, A., Smith, S. I., Mueller, S., Sundberg, E. J., Zimmermann, W., Fischer, W., Hauck, C. R., and Haas, R. (2016) Helicobacter pylori exploits human CEACAMs via HopQ for adherence and translocation of CagA. *Nat Microbiol* **2**, 16188
103. Bonsor, D. A., Zhao, Q., Schmidinger, B., Weiss, E., Wang, J., Deredge, D., Beadenkopf, R., Dow, B., Fischer, W., Beckett, D., Wintrode, P. L., Haas, R., and Sundberg, E. J. (2018) The Helicobacter pylori adhesin protein HopQ exploits the dimer interface of human CEACAMs to facilitate translocation of the oncoprotein CagA. *EMBO J* **37**
104. Moonens, K., Hamway, Y., Neddermann, M., Reschke, M., Tegtmeyer, N., Kruse, T., Kammerer, R., Mejias-Luque, R., Singer, B. B., Backert, S., Gerhard, M., and Remaut, H. (2018) Helicobacter pylori adhesin HopQ disrupts trans dimerization in human CEACAMs. *EMBO J* **37**
105. Schmitter, T., Agerer, F., Peterson, L., Munzner, P., and Hauck, C. R. (2004) Granulocyte CEACAM3 is a phagocytic receptor of the innate immune system that mediates recognition and elimination of human-specific pathogens. *J Exp Med* **199**, 35-46
106. Catton, E. A., Bonsor, D. A., Herrera, C., Stalhammar-Carlemalm, M., Lyndin, M., Turner, C. E., Soden, J., van Strijp, J. A. G., Singer, B. B., van Sorge, N. M., Lindahl, G., and McCarthy, A. J. (2023) Human CEACAM1 is targeted by a Streptococcus pyogenes adhesin implicated in puerperal sepsis pathogenesis. *Nat Commun* **14**, 2275
107. Barnett, T. R., Drake, L., and Pickle, W., 2nd. (1993) Human biliary glycoprotein gene: characterization of a family of novel alternatively spliced RNAs and their expressed proteins. *Mol Cell Biol* **13**, 1273-1282
108. Escalante, J., Artaiz, O., Diwakarla, S., and McQuade, R. M. (2025) Leaky gut in systemic inflammation: exploring the link between gastrointestinal disorders and age-related diseases. *Geroscience* **47**, 1-22
109. Chen, C. J., Lin, T. T., and Shively, J. E. (1996) Role of interferon regulatory factor-1 in the induction of biliary glycoprotein (cell CAM-1) by interferon-gamma. *J Biol Chem* **271**, 28181-28188
110. Gencheva, M., Chen, C. J., Nguyen, T., and Shively, J. E. (2010) Regulation of CEACAM1 transcription in human breast epithelial cells. *BMC Mol Biol* **11**, 79

111. Tchoupa, A. K., Schuhmacher, T., and Hauck, C. R. (2014) Signaling by epithelial members of the CEACAM family - mucosal docking sites for pathogenic bacteria. *Cell Commun Signal* **12**, 27
112. Dawkins, R. (1996) *The blind watchmaker : why the evidence of evolution reveals a universe without design*, Norton, New York
113. Gabarin, R. S., Li, M., Zimmel, P. A., Marshall, J. C., Li, Y., and Zhang, H. (2021) Intracellular and Extracellular Lipopolysaccharide Signaling in Sepsis: Avenues for Novel Therapeutic Strategies. *J Innate Immun* **13**, 323-332
114. Westergaard, H., and Dietschy, J. M. (1976) The mechanism whereby bile acid micelles increase the rate of fatty acid and cholesterol uptake into the intestinal mucosal cell. *J Clin Invest* **58**, 97-108
115. Wang, Z., Ojogun, N., Liu, Y., Gan, L., Xiao, Z., Feng, J., Jiang, W., Chen, Y., Zou, B., Yu, C., Li, C., Ashuo, A., Li, X., Fu, M., Wu, J., Chu, Y., Munford, R. S., and Lu, M. (2025) A host enzyme reduces metabolic dysfunction-associated steatotic liver disease (MASLD) by inactivating intestinal lipopolysaccharide. *Elife* **13**
116. Yu, J., Meng, Z., Liu, X., Zheng, Y., Xue, D., Hao, C., and Wang, L. (2024) Lipopolysaccharide in Bile Promotes the Neutrophil Extracellular Traps-Induced Gallstone Formation by Activating the Gallbladder Immune Barrier. *Immunotargets Ther* **13**, 789-803
117. Chen, L., Ma, S., Cao, A., and Zhao, R. (2024) Bile acids promote lipopolysaccharide clearance via the hepato-biliary pathway in broiler chickens. *Ecotoxicol Environ Saf* **282**, 116767
118. Zaidi, S., Asalla, S., Muturi, H. T., Russo, L., Abdolahipour, R., Belew, G. D., Iglesias, M. B., Feraudo, M., Leon, L., Kuo, E., Liu, X., Kumarasamy, S., Ghadieh, H. E., Gatto-Weis, C., Zarrinpar, A., Duarte, S., and Najjar, S. M. (2024) Loss of CEACAM1 in hepatocytes causes hepatic fibrosis. *Eur J Clin Invest* **54**, e14177
119. Russo, L., Ghadieh, H. E., Ghanem, S. S., Al-Share, Q. Y., Smiley, Z. N., Gatto-Weis, C., Esakov, E. L., McInerney, M. F., Heinrich, G., Tong, X., Yin, L., and Najjar, S. M. (2016) Role for hepatic CEACAM1 in regulating fatty acid metabolism along the adipocyte-hepatocyte axis. *J Lipid Res* **57**, 2163-2175
120. Love-Gregory, L., Kraja, A. T., Allum, F., Aslibekyan, S., Hedman, A. K., Duan, Y., Borecki, I. B., Arnett, D. K., McCarthy, M. I., Deloukas, P., Ordovas, J. M., Hopkins, P. N., Grundberg, E., and Abumrad, N. A. (2016) Higher chylomicron remnants and LDL particle numbers associate with CD36 SNPs and DNA methylation sites that reduce CD36. *J Lipid Res* **57**, 2176-2184
121. Ghanem, S. S., Heinrich, G., Lester, S. G., Pfeiffer, V., Bhattacharya, S., Patel, P. R., DeAngelis, A. M., Dai, T., Ramakrishnan, S. K., Smiley, Z. N., Jung, D. Y., Lee, Y., Kitamura, T., Ergun, S., Kulkarni, R. N., Kim, J. K., Giovannucci, D. R., and Najjar, S. M. (2016) Increased Glucose-induced Secretion of Glucagon-like Peptide-1 in Mice Lacking the Carcinoembryonic Antigen-related Cell Adhesion Molecule 2 (CEACAM2). *J Biol Chem* **291**, 980-988
122. Wong, C., Liu, Y., Yip, J., Chand, R., Wee, J. L., Oates, L., Nieswandt, B., Reheman, A., Ni, H., Beauchemin, N., and Jackson, D. E. (2009) CEACAM1 negatively regulates platelet-collagen interactions and thrombus growth in vitro and in vivo. *Blood* **113**, 1818-1828
123. Ghosh, A., Murugesan, G., Chen, K., Zhang, L., Wang, Q., Febbraio, M., Anselmo, R. M., Marchant, K., Barnard, J., and Silverstein, R. L. (2011) Platelet CD36 surface expression levels affect functional responses to oxidized LDL and are associated with inheritance of specific genetic polymorphisms. *Blood* **117**, 6355-6366
124. Martyanov, A. A., Maiorov, A. S., Filkova, A. A., Ryabykh, A. A., Svidelskaya, G. S., Artemenko, E. O., Gambaryan, S. P., Panteleev, M. A., and Sveshnikova, A. N. (2020)

Effects of bacterial lipopolysaccharides on platelet function: inhibition of weak platelet activation. *Sci Rep* **10**, 12296
125. Armstrong, M. T., Rickles, F. R., and Armstrong, P. B. (2013) Capture of lipopolysaccharide (endotoxin) by the blood clot: a comparative study. *PLoS One* **8**, e80192
126. Slevogt, H., Zabel, S., Opitz, B., Hocke, A., Eitel, J., N'Guessan P, D., Lucka, L., Riesbeck, K., Zimmermann, W., Zweigner, J., Temmesfeld-Wollbrueck, B., Suttorp, N., and Singer, B. B. (2008) CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nat Immunol* **9**, 1270-1278
127. Hussaini, S. H., Pereira, S. P., Murphy, G. M., and Dowling, R. H. (1995) Deoxycholic acid influences cholesterol solubilization and microcrystal nucleation time in gallbladder bile. *Hepatology* **22**, 1735-1744
128. Wang, K., Lai, W., Min, T., Wei, J., Bai, Y., Cao, H., Guo, J., and Su, Z. (2024) The Effect of Enteric-Derived Lipopolysaccharides on Obesity. *Int J Mol Sci* **25**
129. Hersoug, L. G., Moller, P., and Loft, S. (2016) Gut microbiota-derived lipopolysaccharide uptake and trafficking to adipose tissue: implications for inflammation and obesity. *Obes Rev* **17**, 297-312
130. Du, J., Tian, J., Ding, L., Trac, C., Xia, B., Sun, S., Schones, D. E., and Huang, W. (2017) Vertical sleeve gastrectomy reverses diet-induced gene-regulatory changes impacting lipid metabolism. *Sci Rep* **7**, 5274
131. Jin, Z. L., and Liu, W. (2021) Progress in treatment of type 2 diabetes by bariatric surgery. *World J Diabetes* **12**, 1187-1199

FIGURES AND FIGURE LEGENDS

Fig 1. Surface plasmon resonance analysis of the binding of *E. coli* LPS-Ra or BAs to sCD36 and sCEACAM1-4D. **A.** sCEACAM1-4D (50 µg/ml) was immobilized on a CM5 chip and LPS-Ra binding performed over the range 0-4 µM. **B-C**. Binding of different concentrations of sodium cholate (**B**) or sodium deoxycholate (**C**) to immobilized sCEACAM1-4D was studied. **D**. sCD36 (50 µg/ml) was immobilized to a BIACore CM5 chip and binding studies with LPS-Ra (0 to 4 µM) was performed. **E-F**. Increasing concentrations of sodium cholate (**E**) or sodium deoxycholate (**F**) were run through the immobilized sCD36. (Information given in the methods) The kinetic constants for LPS binding to sCEACAM1-4D are $K_a = 5.85 \times 10^3$ 1/Ms, $K_d = 6.08 \times 10^{-7}$ 1/s, and $K_D = 1.04 \times 10^{-10}$ M.. The kinetic constants for LPS binding to sCD36 are $K_a = 3.69 \times 10^3$ 1/Ms, $K_d = 1.25 \times 10^{-6}$ 1/s, and $K_D = 3.38 \times 10^{-10}$ M..

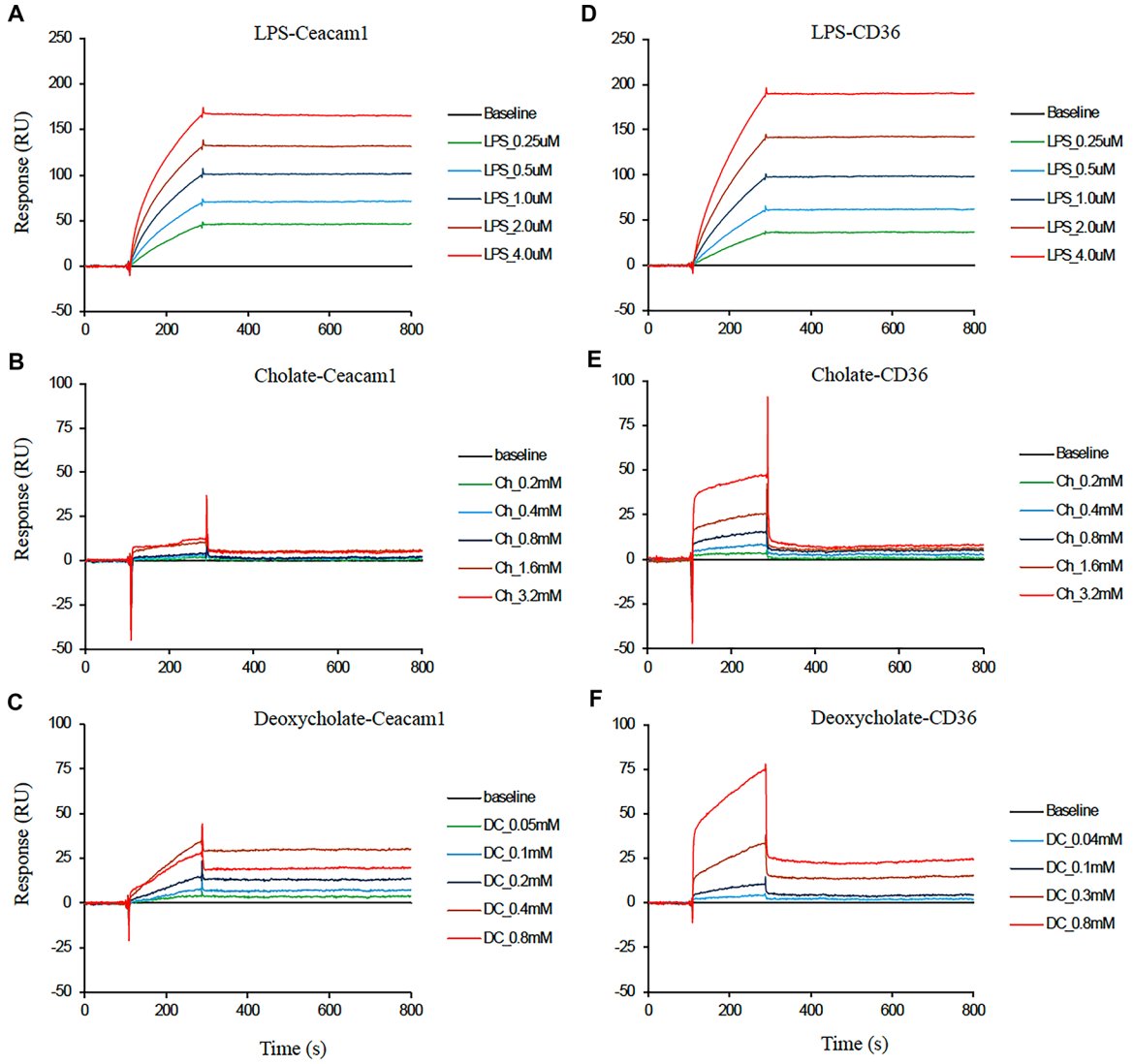


Fig 2. Surface plasmon resonance analysis of the binding of *E. coli* LPS-Ra to sCEACAM1-4D in the presence of sodium deoxycholate or sodium cholate. **A.** sCEACAM1-4D (50 µg/ml) immobilized to a CM5 chip was treated with increasing concentrations (0-384 µM) of sodium deoxycholate (DC) in the absence of LPS. LPS (2 µM) only is shown as a reference. **B**. Repeat of **A** with constant amount of LPS (2 µM) with added DC. **C**. Sensograms corrected for deoxycholate only binding (**B** minus **A**). **D-F**. Same set of experiments with sodium cholate (Ch). Chips were regenerated between runs with 0.5% SDS in water.

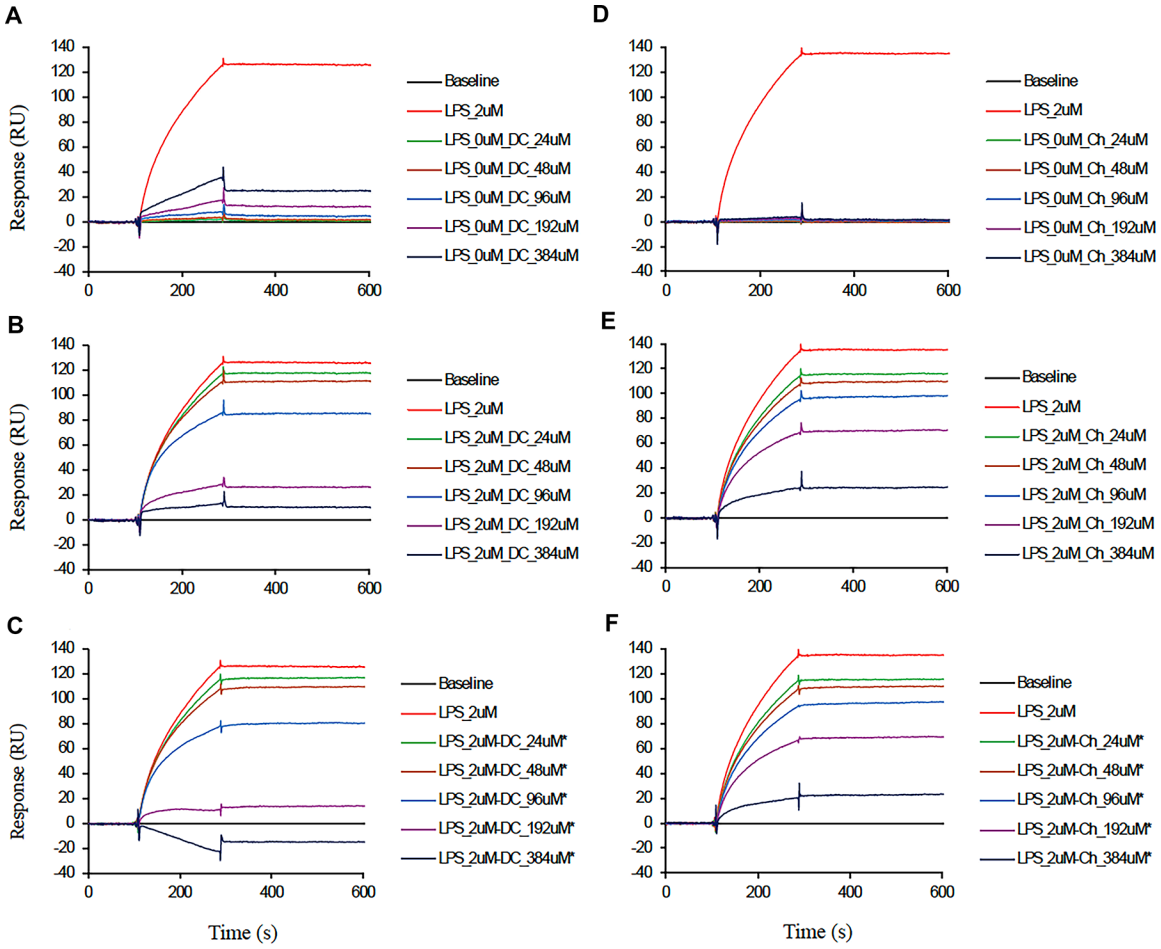


Fig 3. Surface plasmon resonance analysis of the binding of *E. coli* LPS-Ra to sCD36 in the presence of sodium deoxycholate or sodium cholate. **A.** sCD36 (12.5 μg/ml) immobilized to a CM5 chip was treated with increasing concentrations (0-384 μM) of sodium deoxycholate (DC) in the absence of LPS. LPS (2 μM) only is shown as a reference. **B**. Repeat of **A** with constant amount of LPS (2 μM) with added DC. **C**. Sensograms corrected for deoxycholate only binding (**B** minus **A**). **D-F**. Same set of experiments with sodium cholate (Ch). Chips were regenerated between runs with 0.25% SDS in water.

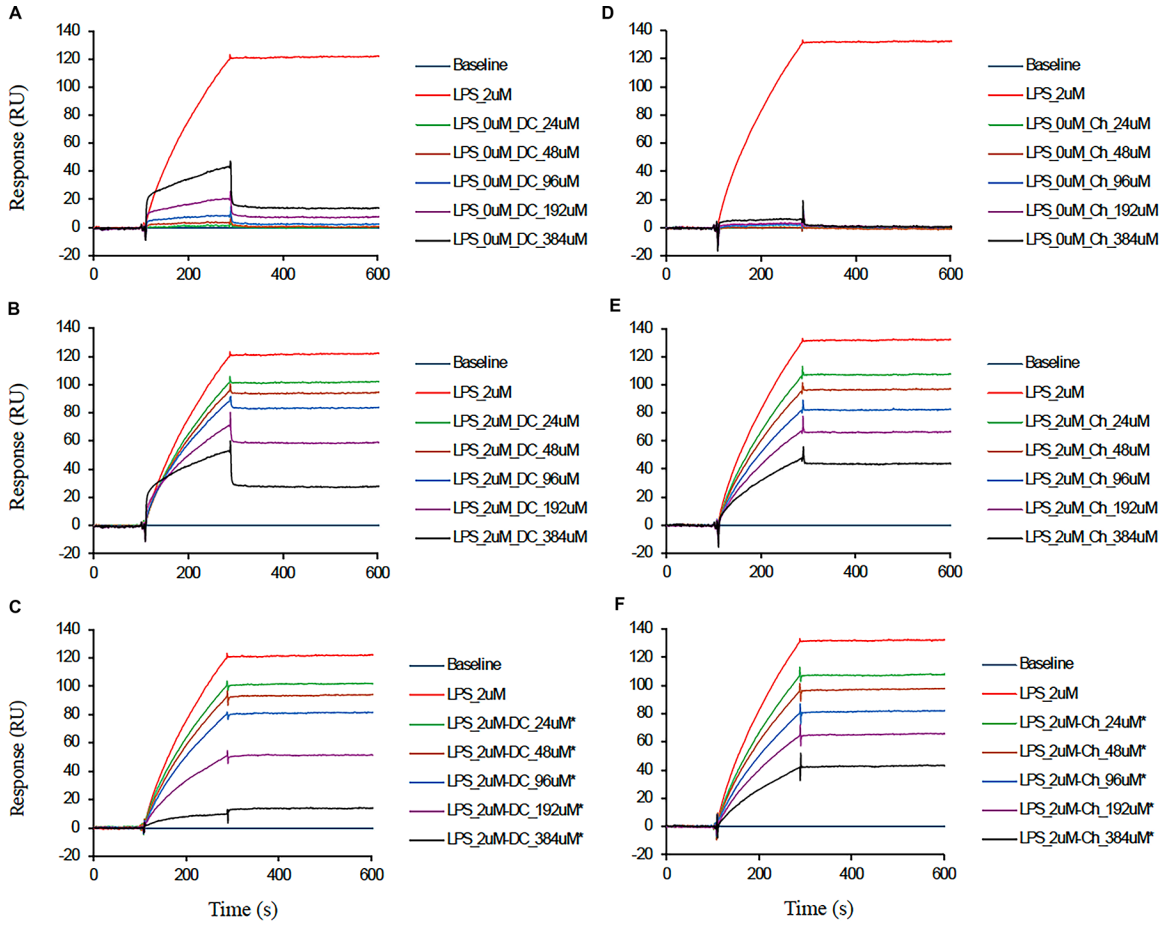


Fig 4. Size exclusion chromatography analysis of the binding of *E. coli* LPS-Ra to sCD36 and sCEACAM1-4D. A-D. sCD36 (35 µg, 0.5 nmol) was incubated with increasing amounts of LPS-Ra (0-150 µg, 0-30 nmoles assuming a monomeric molecular weight of 5 kDa) for 1 hr at RT in a volume of 0.1 mL and the entire amount injected onto a Superdex 200i column in PBS at a flow rate of 0.5 mL/min monitored at 215 and 280 nm. **E-F**. sCEACAM1-4D (35 µg, 0.29 nmoles) was incubated with increasing amounts of LPS-Ra and analyzed as described in **A-D**. Note: LPS-Ra is UV transparent at 215 nm in these studies.

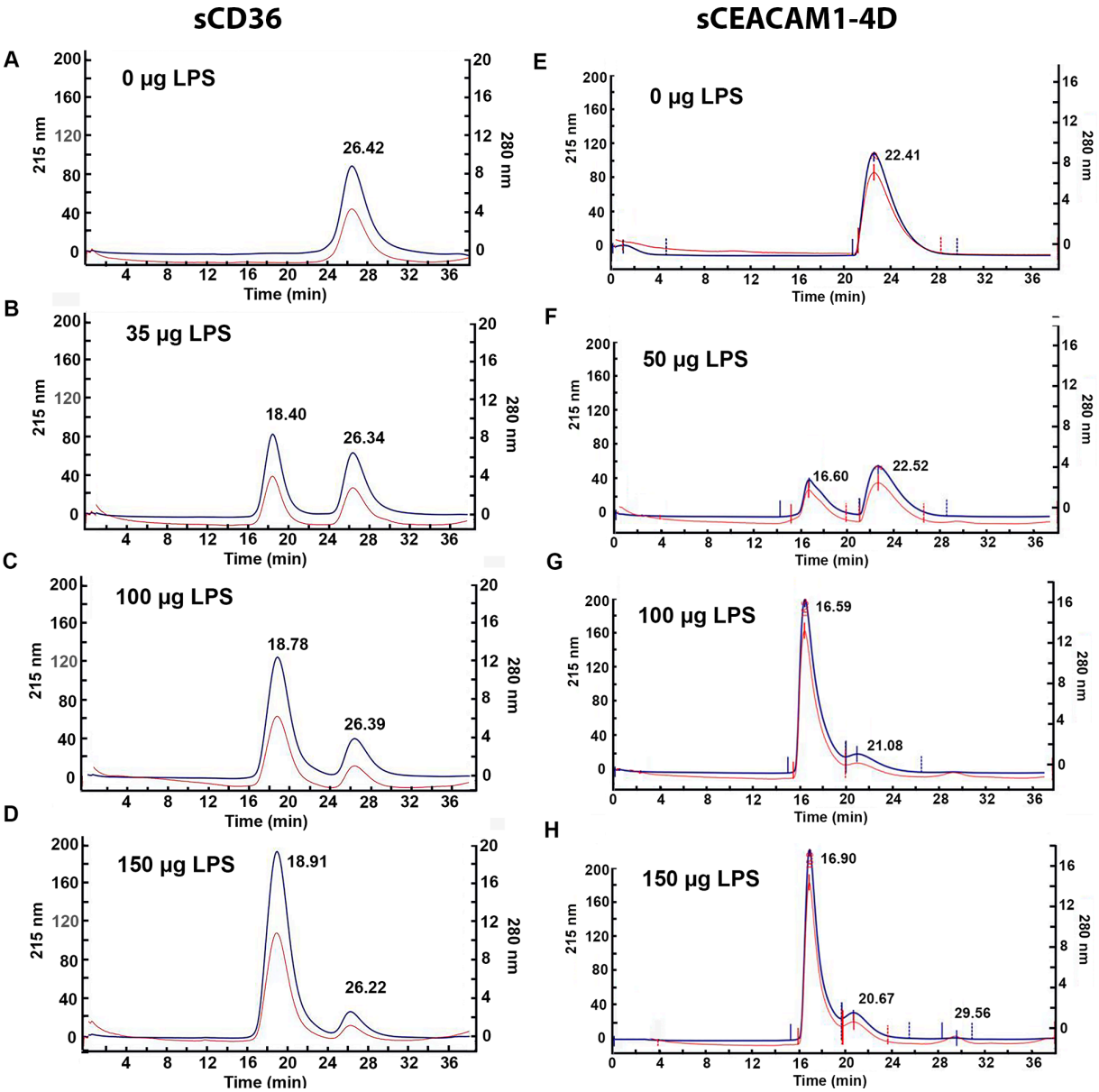
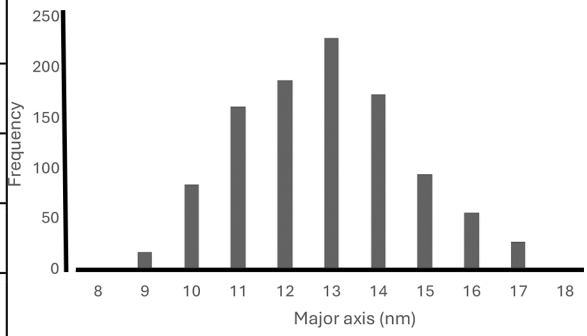


Fig 5. Transmission electron microscopy analysis of *E. coli* LPS-Ra complexes with sCD36. The sCD36-LPS-Ra complex isolated from SEC (see **Fig 4**) was analyzed by TEM. Particle size analysis shows measurement and distribution of ellipsoid complexes**.**

	Major axis (nm)	**Minor axis (nm)**
Mean	12.34	8.57
Std. Dev.	1.78	1.22
Median	12.34	8.40
Min	8.22	5.43
Max	17.15	13.07



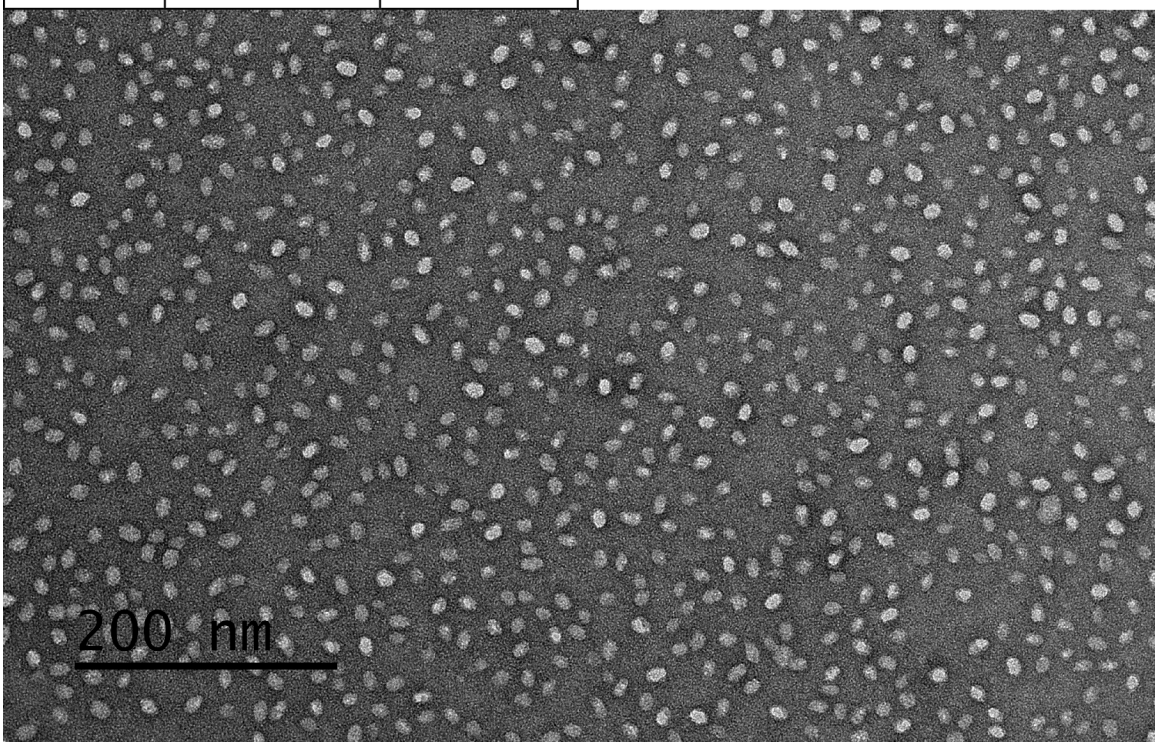
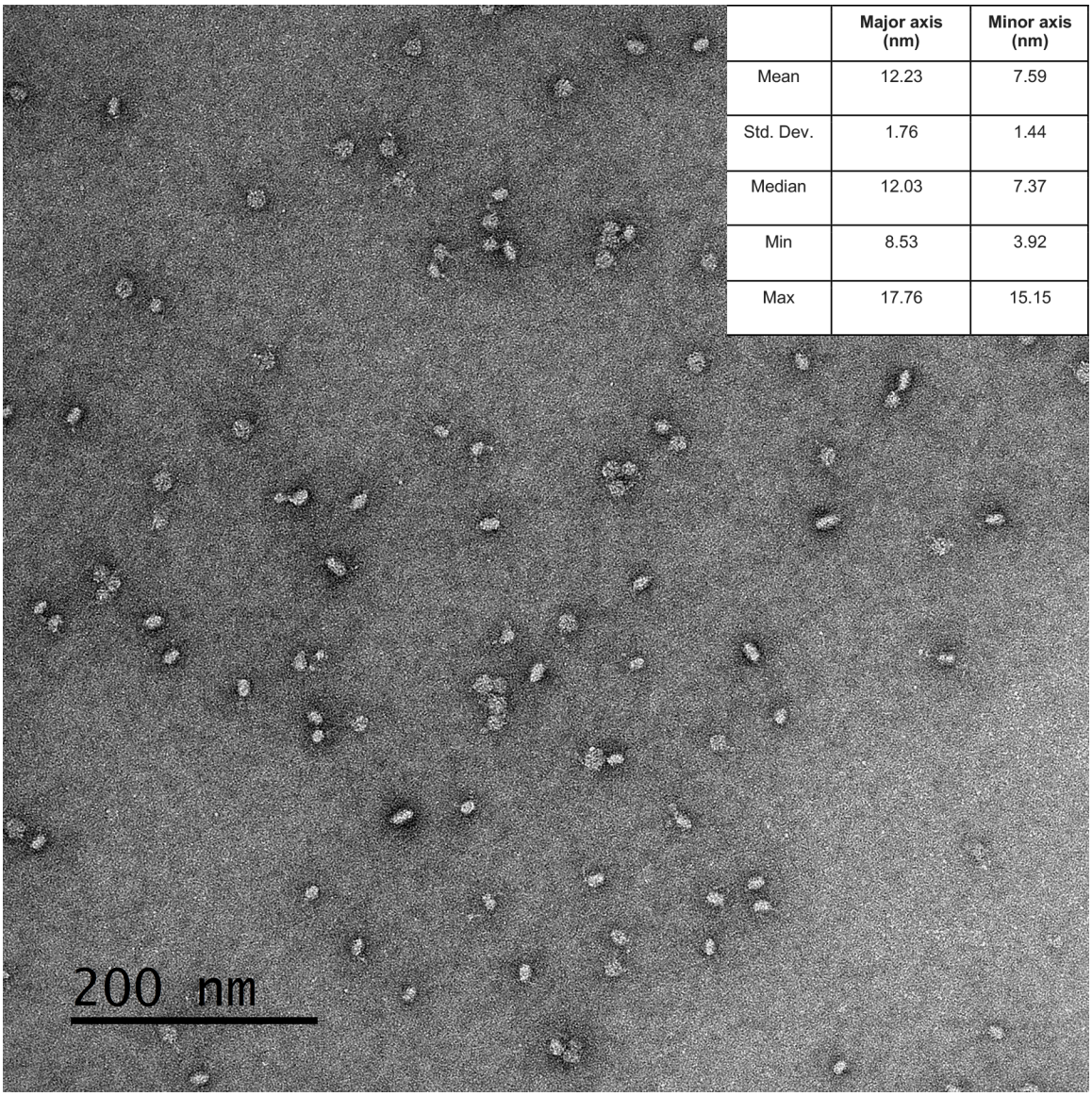


Fig 6. Transmission electron microscopy analysis of *E. coli* LPS-Ra complexes with sCEACAM1-4D. The sCEACAM1-4D-LPS-Ra complex isolated from SEC (see **Fig 4**) was analyzed by TEM. Example of one field with scale bar and particle size analysis**.** A total of 10 fields were analyzed for particle size, n = 559.



	Major axis (nm)	Minor axis (nm)
Mean	12.23	7.59
Std. Dev.	1.76	1.44
Median	12.03	7.37
Min	8.53	3.92
Max	17.76	15.15

Fig 7. Transmission electron microscopy analysis of sCEACAM1-4D. Field with scale bar**.** Six particle shapes were quantitated as shown on the right.

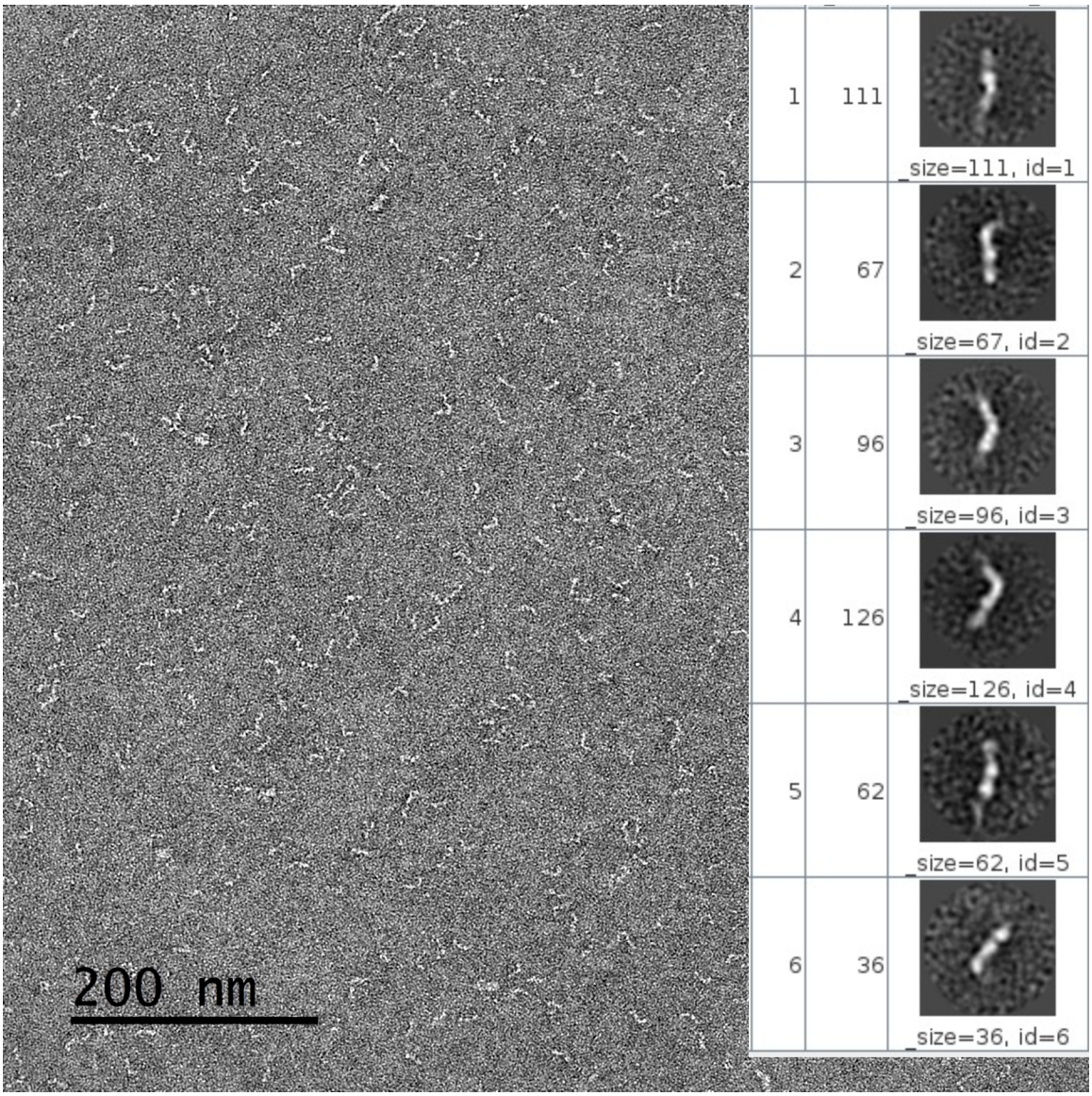


Fig 8. Alpha Fold model of CD36-LPS-Ra complex. **A**. Ribbon model of CD36 with stick model of LPS-Ra docked to CD36 (green: full length CD36, Alpha Fold model; magenta: sCD36 x-ray structure). **B**. Detail shows insertion of two acyl chains (Ac1 on the 2-amino group of GlcN and acyl 2 (Ac2) on the 3-OH of Ac1) into cavities of CD36. The interior free fatty acid (green) bound to CD36 was retained in the structure.

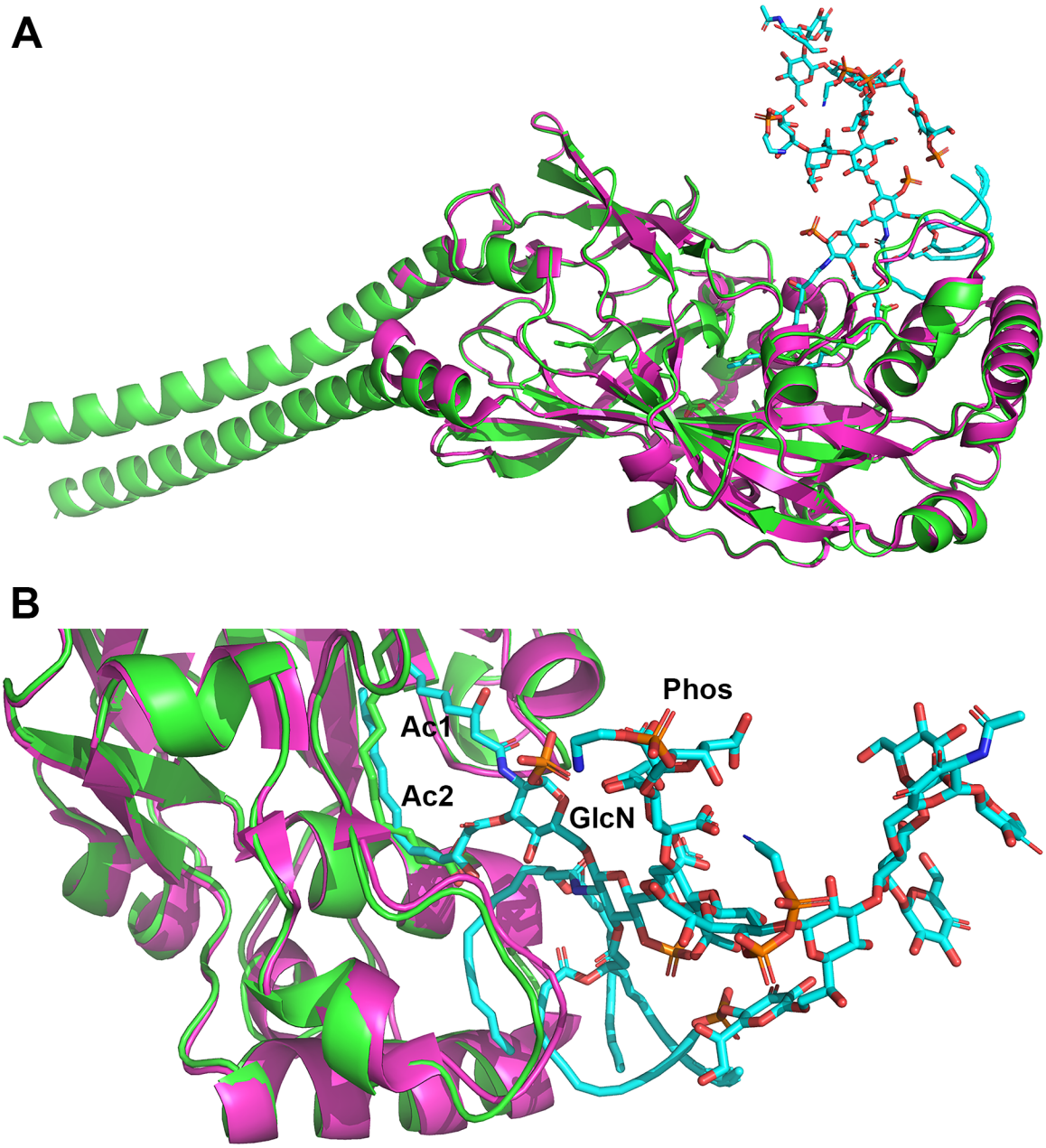


Fig 9. Model of LPS-Ra micelle interactions with bile acids or CD36. A. Micelle model with LPS-Ra with exposed hydrophobic acyl chains capped with deoxycholic acid (red). For clarity, two additional molecules of LPS-Ra in ±z-axis directions not shown). **B.** Abbreviated model with LPS Ra micelle showing rotation about glycosidic bond between hexosamine units exposing two acyl chains (magenta) that may insert into the hydrophobic core of CD36 (CD36 not shown). Divalent metal ions bridging phosphates shown in light blue.

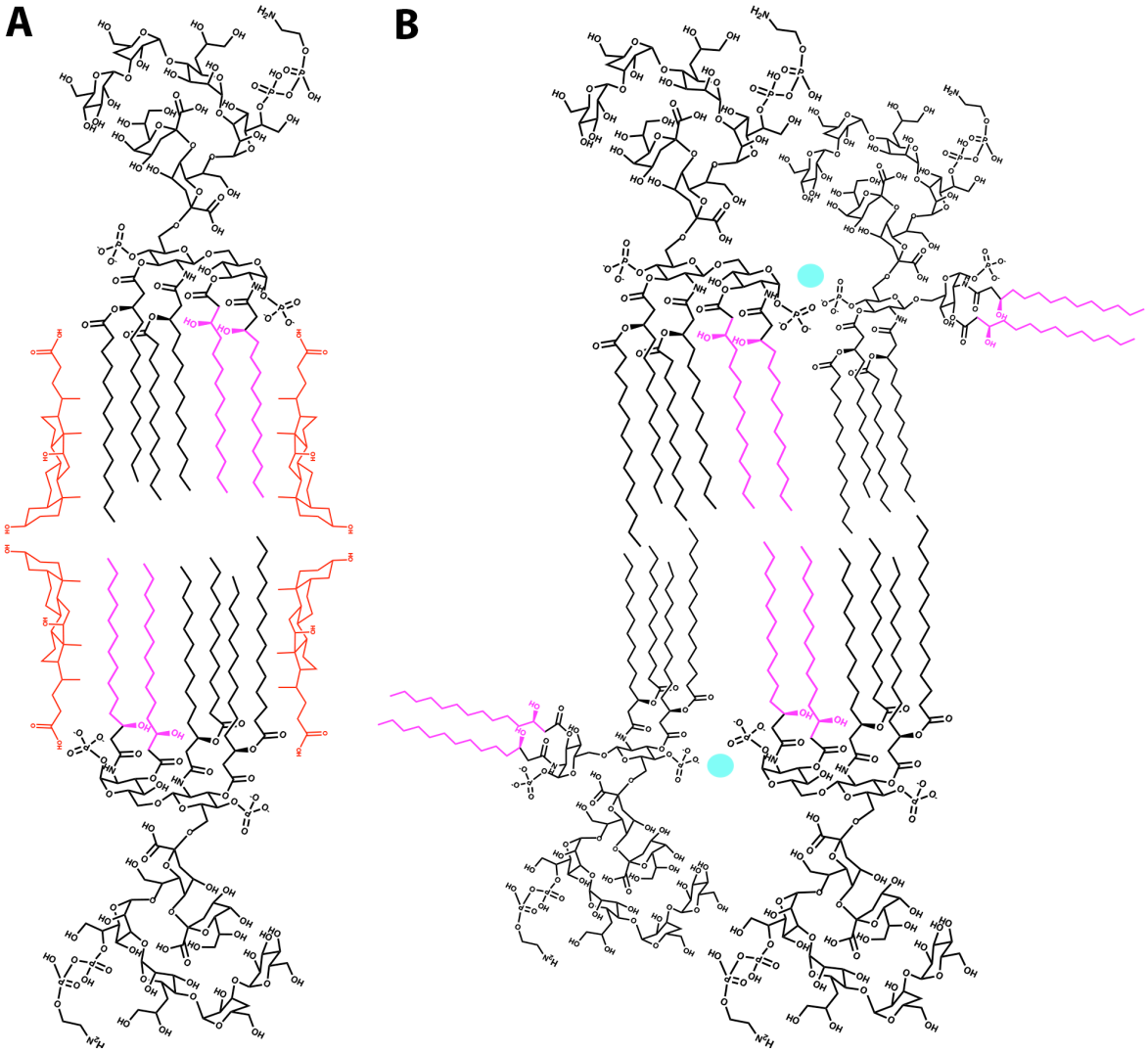


Fig 10. Molecular dynamics model of open form of CEACAM1-4D. A. Frequency distribution of the distance between position marker amino acids Q53 and P69 in CEACAM1 N-domain over 2 μsec time frame. Corresponding distance in the crystal structure (pdb ID: 4WHD) is highlighted by vertical red dotted line. **B-C.** Closed (**B**) and open (**C**) poses showing position marker amino acids Q53 and P69 in CEACAM1 N-domain. Note the loss of secondary structures C'' in going from closed to open pose. **D**. Superimposed open and closed poses with highlighted secondary structures that change the most. **E**. Side view of open pose with highlighted secondary structure showing open gap between the beta-sheet sandwich and salt bridge (R64-D82) that keeps the structure closed at the bottom of the beta-sandwich. Note residue N81 that is glycosylated and lies adjacent to D82 on a helical loop.

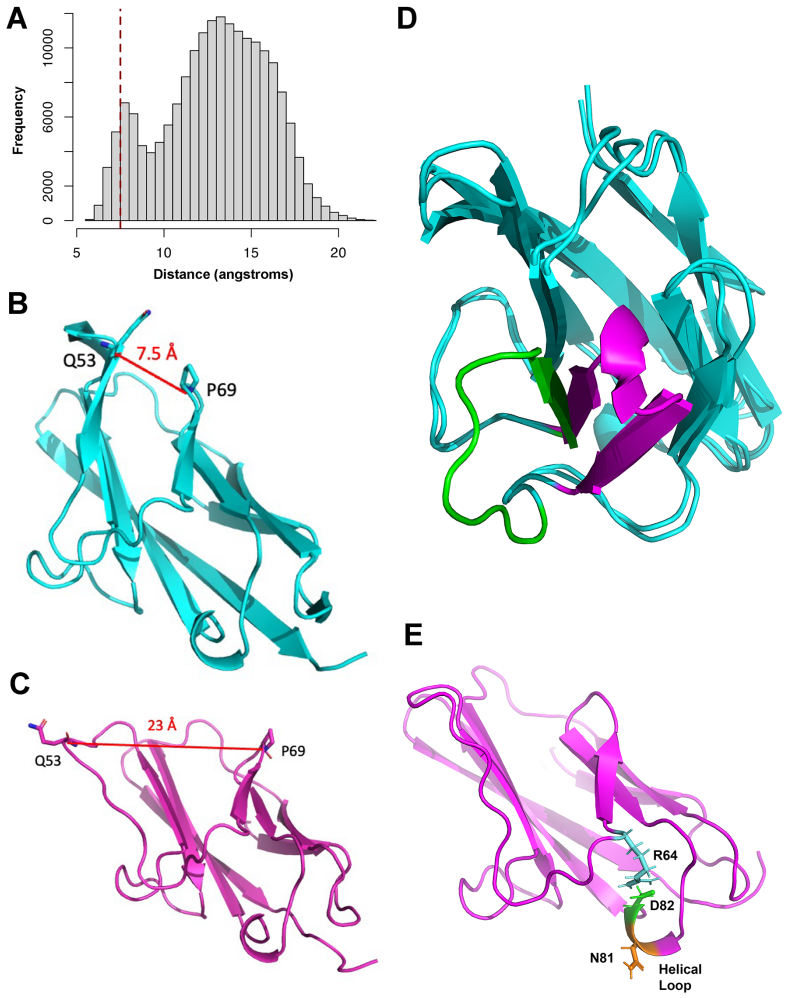


Fig 11. Molecular dynamics model of D82A mutant of CEACAM1-N modeled with LPS-Ra. A. Ribbon model of D82A mutant of CEACAM1-N showing distance markers residue N81 with attached glycan and N61. **B.** MD simulation of distance between two marker residues over 2 x 1 μsec simulation. **C.** Using the open form (greatest distance between markers) from **B**, LPS-Ra was sequentially (left to right) docked into a surface model of CEACAM1-N. **D.** Ribbon model with beta-strands shown (A to G) of CEACAM1-N with docked minimum structure of LPS-Ra.

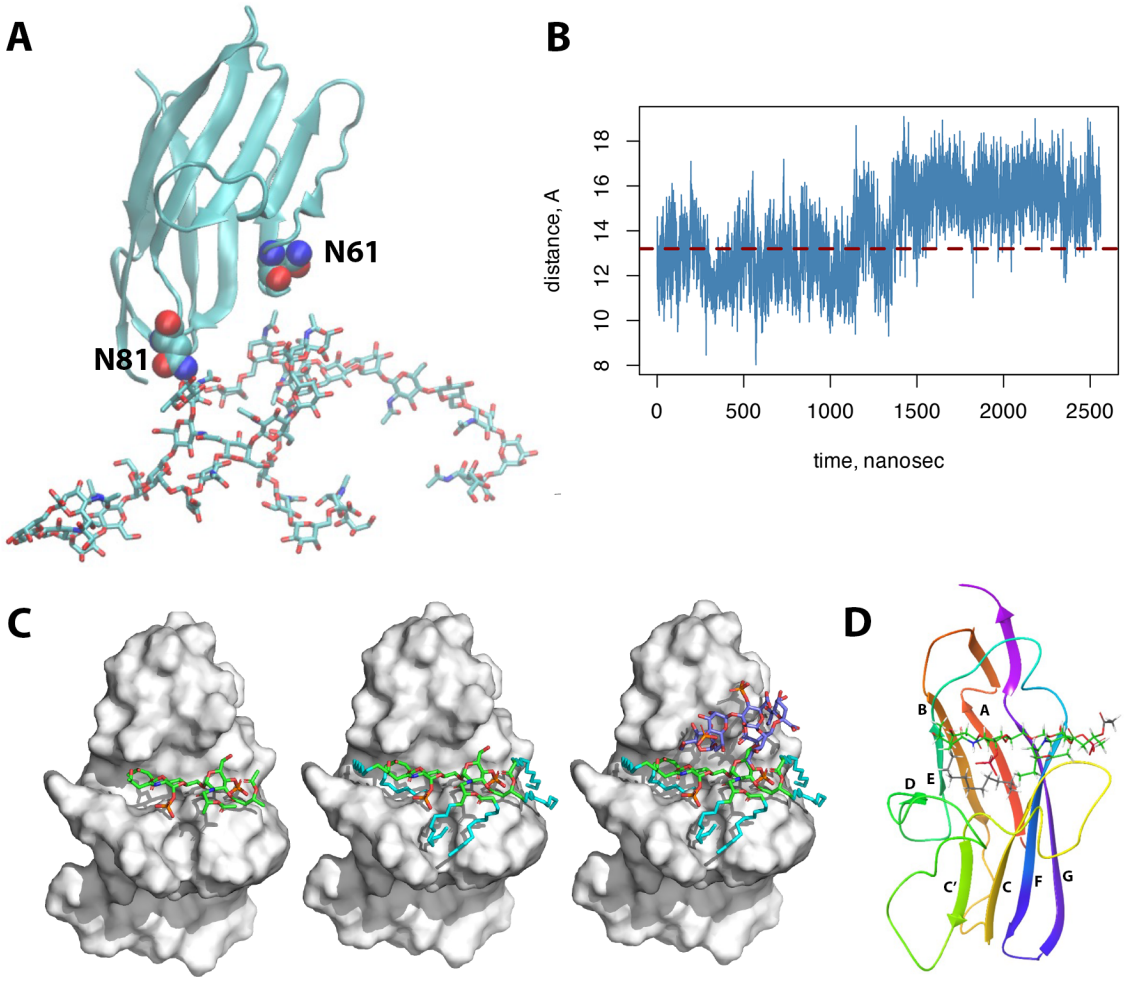


Fig 12. Open model of CEACAM1-N with docked LPS-Ra and critical residues. A. Ribbon model of open form of CEACAM1-N with docked LPS-Ra (core glycan in dark blue, di-glucosamine in green, acyl chains in cyan). **B.** Detail showing interaction of the two phosphates in LPS-Ra with residues K35 and R43 in CEACAM1-N. **C.** Contact residue cartoon for CEACAM1-N and minimum structure of LPS-Ra.

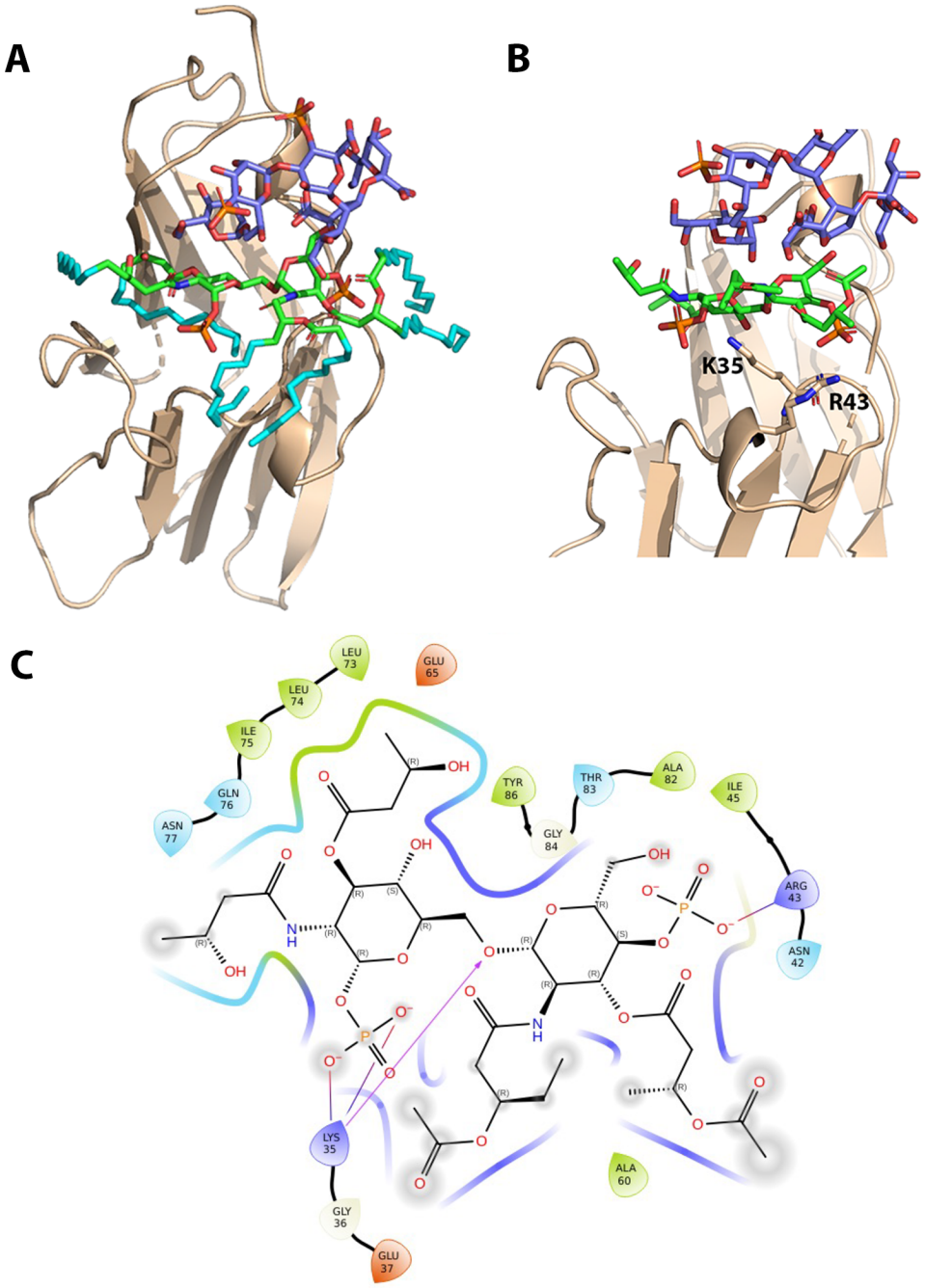


Fig 13. Surface plasmon resonance analysis of the binding of sCEACAM1-3D to sCD36 . sCD36 (12.5 µg/ml) was immobilized to a BIACore CM5 chip and binding studies with CEACAM1 (0-2.0 µM) performed. Chips were regenerated between runs with 0.1 N HCl solution. The kinetic constants are k_a= 3.25 x 10^4 $M^{-1}s^{-1}$ and k_d= 1.72 x 10^{-3} s^{-1} with K_D = 5.28 x 10^{-8} M.

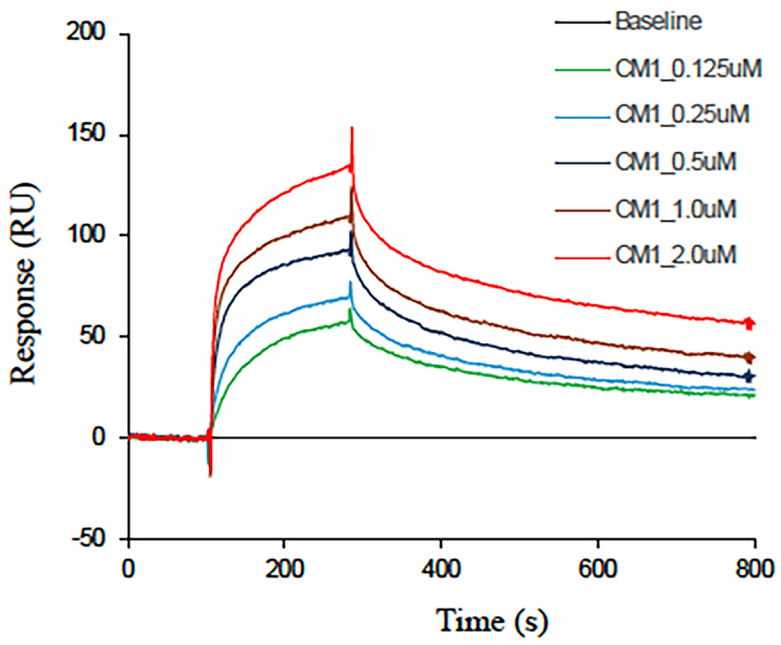


Fig 14. Cell surface proximity ligation assays for CD36 and CEACAM1-4L. A-B. HepG2 hepatocytes expressing only CD36 (**A**) or both CD36 and CEACAM1-4L (**B**) were cross-linked with CD36 and CEACAM1 specific antibodies conjugated with complementary oligonucleotides and amplified with fluorescent labeled primers. Dots were counted over several fields for both conditions. **C-D**. HEK cells expressing only CEACAM1-4L (**C**) or both CEACAM1-4L and CD36-BioID2 (**D**) were treated as in **A-D**. Dots were counted over several fields for each condition.

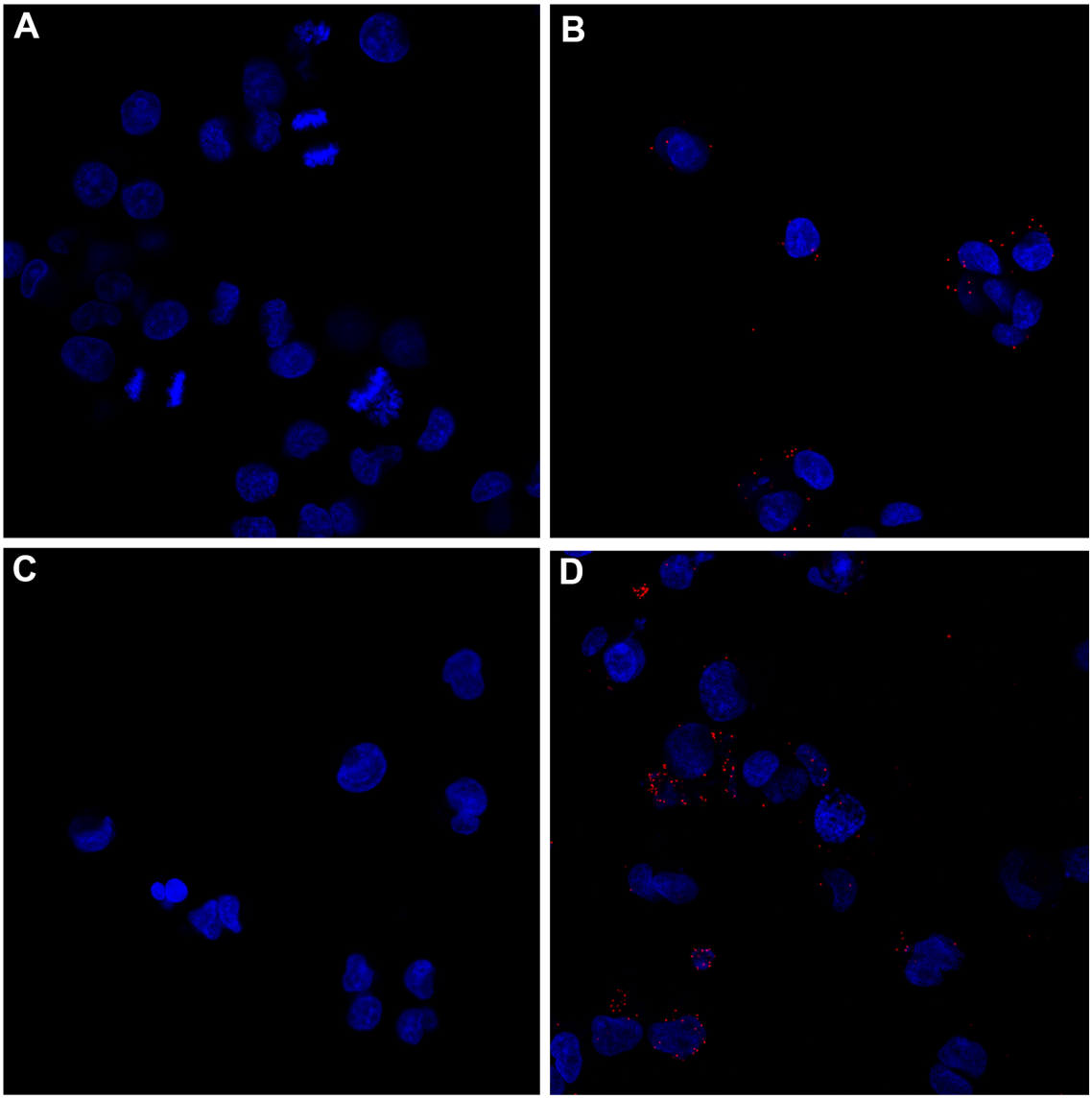


Fig 15. Cytoplasmic domain biotin proximity ligation assays for CD36-BioID2 and CEACAM1-4L. A-C. Flow analysis of CEACAM1-4L and CD36-BioID2 expression in transfected HEK cells. (**A**) HEK cells untransfected controls. (**B**) HEK cells transfected with CD36-BioID2 only. (**C** HEK cells transfected with both CD36-BioID2 and CEACAM1-4L. **D-E**. Immunoblots of HEK cells expressing only CD36-BioID2 (lane 1) or both CD36-BioID2 and CEACAM1-4L (lane 2), immunostained with antibodies to CD36 (**D**) or CEACAM1-4L (**E**). Transfected cells were grown for 4 days, incubated with 50 μM biotin overnight, washed twice with PBS, lysed with 1% NP40, incubated with streptavidin beads for 4 hrs, eluted with SDS running buffer, run on SDS gels and immunoblotted with either anti-CD36 or anti-CEACAM1 antibodies.

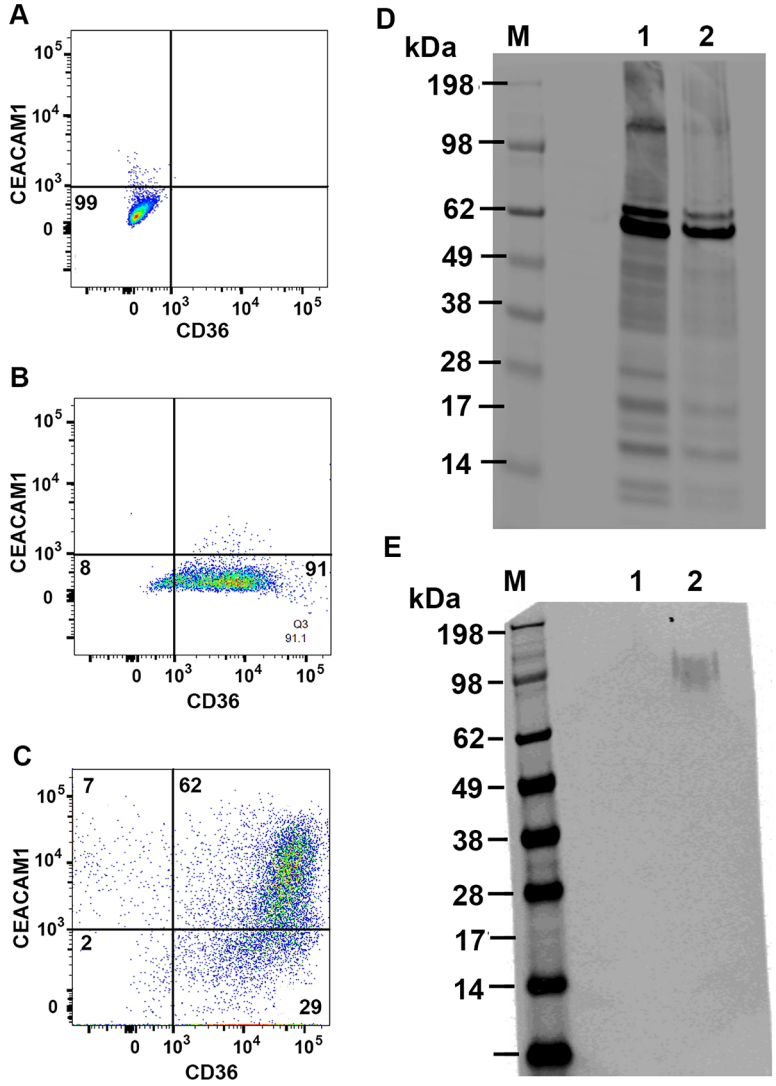


Fig 16. Alpha Fold model of CEACAM1-4D trans interaction with CD36. A. Trans interaction of CEACAM1-4D with CD36. **B**. View 1 of key amino acid interaction pairs of CEACAM1 N-domain with CD36 in trans model. **C.** View 2 of key amino acid interaction pairs of CEACAM1 N-domain with CD36 in trans model.

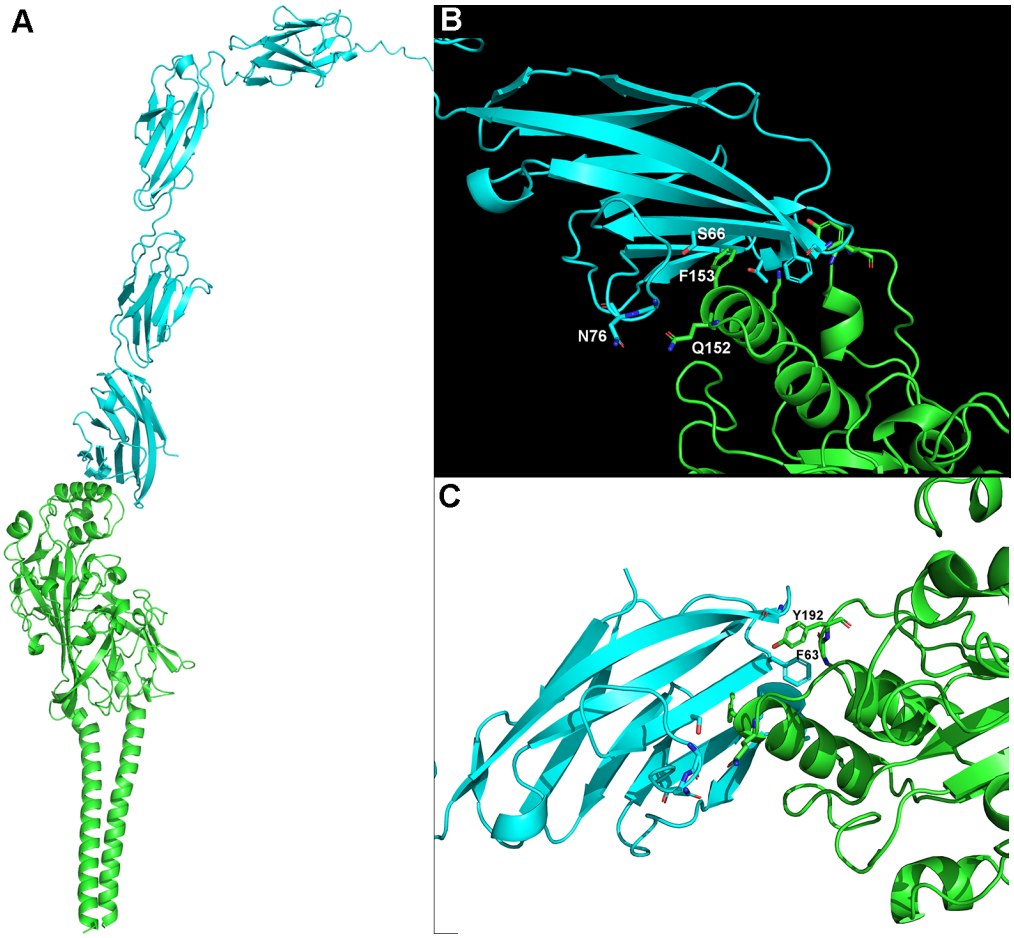


Fig 17. Alpha Fold model of CEACAM1-4D cis interaction with CD36. A. Cis interaction of CEACAM1-4D (magenta) with CD36 (cyan). **B**. Contact amino acids between CEACAM1 N-domain and CD36 in the cis interaction model. The middle residues (N97-E335) are 5-7 Å apart but are both on flexible loops.

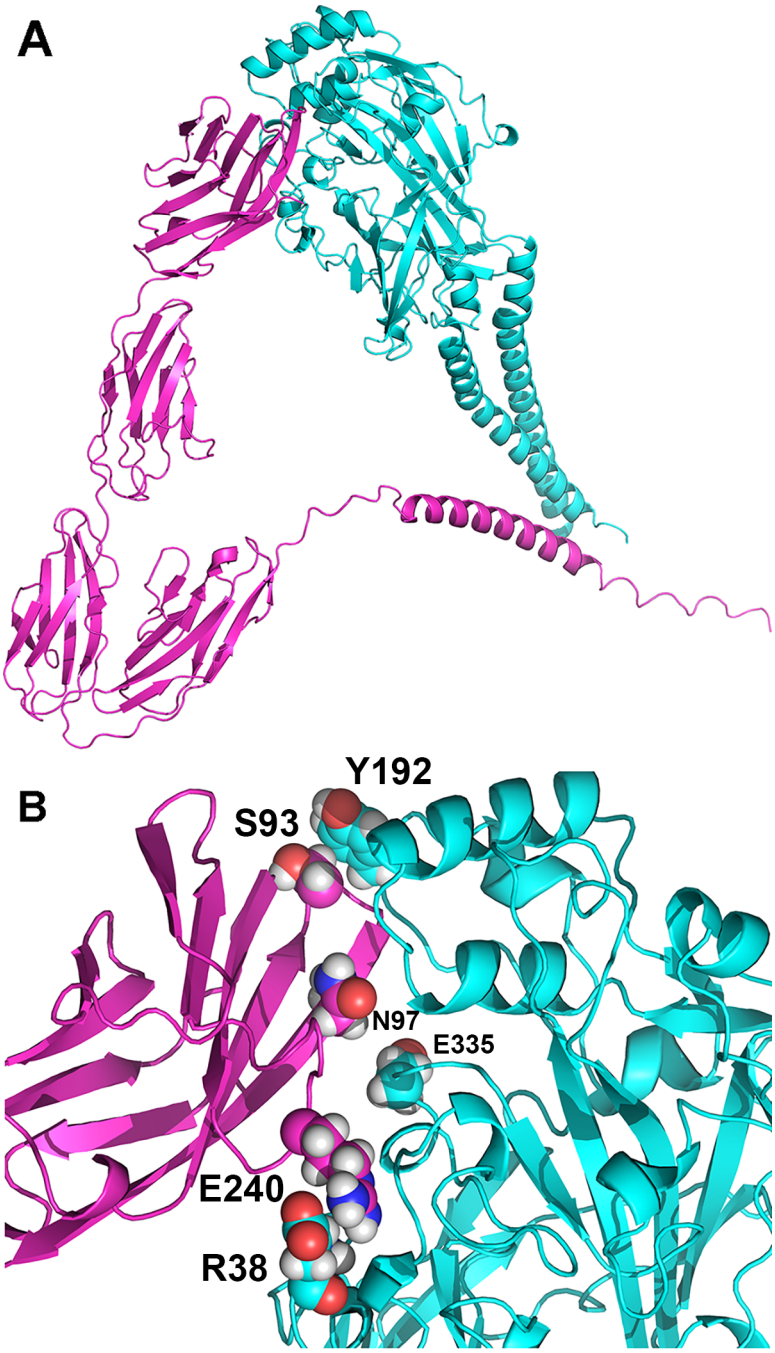


Fig 18. Alpha Fold membrane model of CEACAM1-4L interaction with CD36. A. Side view of alpha fold structure taken from Fig 15 and embedded in a generalized plasma membrane. **B.** Top-down view.

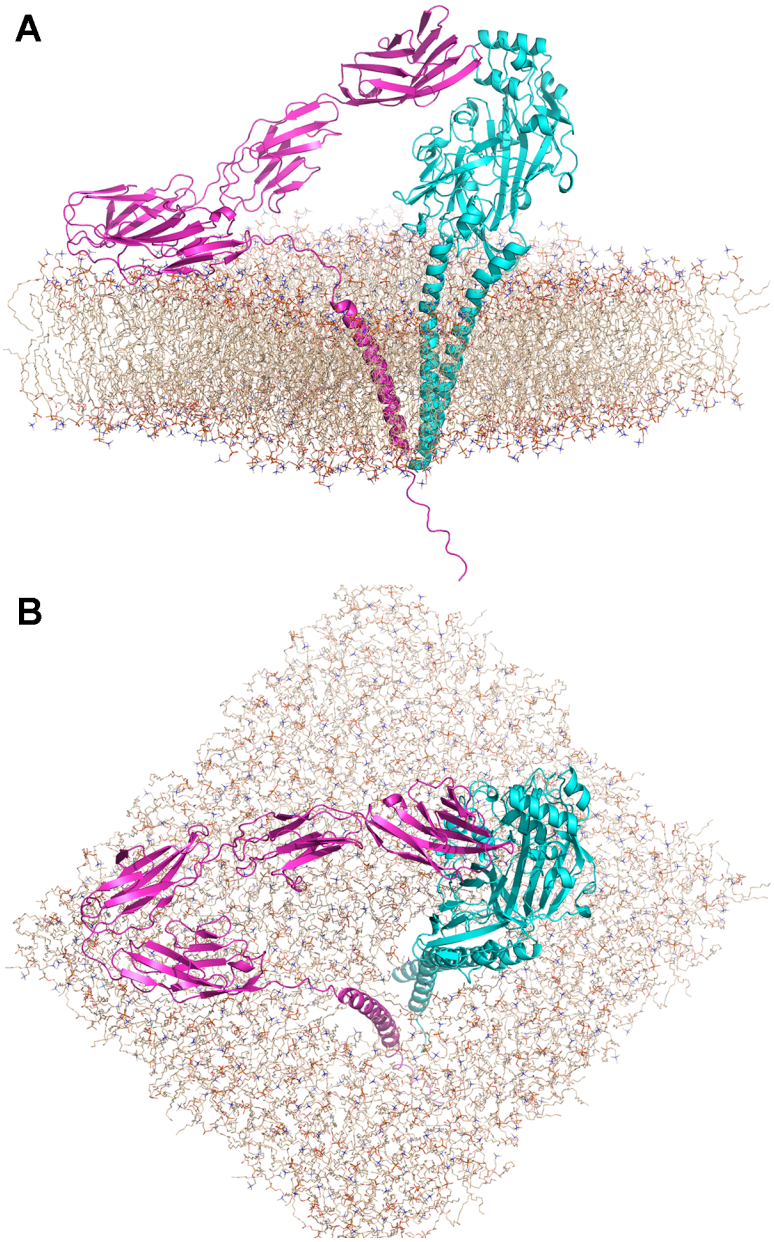
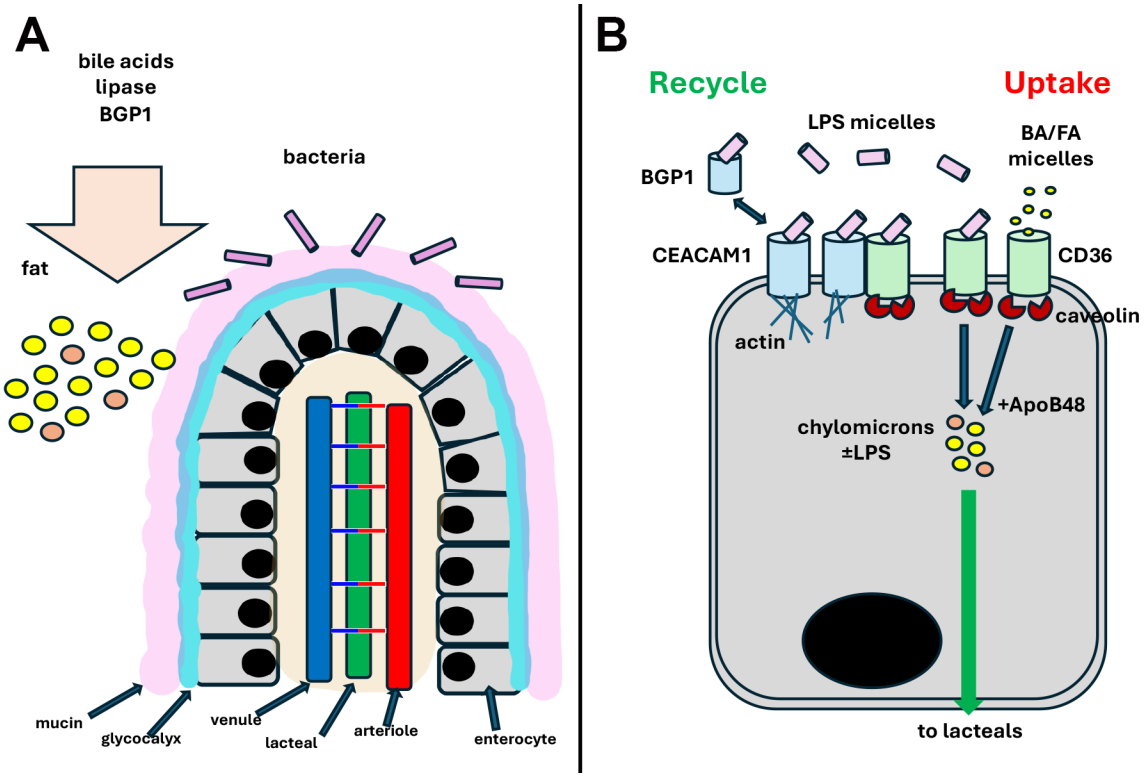



Fig 19. Enterocyte uptake-recycle model of LPS micelles. A. Intestinal villus showing adherence of bacteria to mucous layer during ingestion of fat, bile and pancreatic lipase. **B.** CD36 and CEACAM1 receptors on individual enterocyte and their interactions with LPS micelles.