Engineering cardiolipin binding to an artificial membrane protein

Authors
Mia L. Abramsson,[1] Robin A. Corey,[2]* Jan Škerle,[3,4] Louise J. Persson,[5] Olivia Andén,[6] Abraham O. Oluwole,[7,8] Rebecca J. Howard,[6,] Erik Lindahl,[6,9] Carol V. Robinson,[7,8] Kvido Strisovsky,[4] Erik G. Marklund,[5] David Drew,[3] Phillip J. Stansfeld[10]* & Michael Landreh[1,11]*

Affiliations
[1] Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 171 65 Solna, Sweden
[2] School of Physiology, Pharmacology & Neuroscience, University of Bristol, BS8 1TD Bristol, UK
[3] Department of Biochemistry and Biophysics, Stockholm University, 104 05 Stockholm, Sweden
[4] Institute of Organic Chemistry and Biochemistry, Academy of Science of the Czech Republic, 160 00 Prague, Czech Republic
[5] Department of Chemistry – BMC, Uppsala University, 751 23 Uppsala, Sweden
[6] Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, 171 65 Solna, Sweden
[7] Department of Chemistry, University of Oxford, OX1 3QZ Oxford, UK
[8] Kavli Institute for Nanoscience Discovery, University of Oxford, OX1 3QU Oxford, UK
[9] Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 171 65 Solna, Sweden
[10] School of Life Sciences & Chemistry, University of Warwick, CV4 7AL Coventry, UK
[11] Department for Cell and Molecular Biology, Uppsala University, 751 24 Uppsala, Sweden

* Correspondence: robin.corey@bristol.ac.uk, phillip.stansfeld@warwick.ac.uk, or michael.landreh@icm.uu.se

Abstract
Integral membrane proteins carry out essential functions in the cell, and their activities are often modulated by specific protein-lipid interactions in the membrane. Here, we elucidate the intricate role of cardiolipin (CDL), a regulatory lipid, as a stabilizer of membrane proteins and their complexes. Using the *in silico*-designed model protein TMHC4_R (ROCKET) as a scaffold, we employ a combination of molecular dynamics simulations and native mass spectrometry to explore the protein features that facilitate preferential lipid interactions and mediate stabilization. We reveal that the spatial arrangement of positively charged residues as well as local conformational flexibility are pivotal in distinguishing stabilizing from non-stabilizing CDL interactions. Extending our insights to naturally occurring proteins, we identify a stabilizing CDL site within the *E. coli* rhomboid intramembrane protease GlpG and uncover its regulatory influence on enzyme substrate preference. This work establishes a framework for engineering functional lipid interactions, paving the way for the design of proteins with membrane-specific properties or functions.

Introduction

Biological membranes, which are vital for cellular life, provide a specific and highly adaptable lipid environment for membrane proteins that govern numerous cellular functions (1). The exact roles that the different membrane lipids play in the regulation of membrane proteins often go unacknowledged, as their highly dynamic interactions challenge conventional analytical methods. Despite these obstacles, evidence has consistently highlighted the crucial role of lipids (2), for example as allosteric regulators (3), facilitating protein oligomerization (4), or locally affecting the properties of the membrane (5). The simplest form of lipid-mediated regulation is the stabilization of specific protein conformations (6), resulting in the observation of individual lipid molecules in high-resolution structures (7). These "structural" lipids often display increased residence times at their binding sites which distinguish them from non-regulatory, "annular" lipids (8).

Cardiolipin (CDL) is a prime example of a lipid with regulatory activity for both bacterial and mitochondrial membrane proteins (9). Due to its unique structure, comprised of two phosphate groups which both potentially carry a negative charge, and four acyl chains, CDL mediates the assembly of membrane protein oligomers, for example in the respiratory chain supercomplexes (10). The double phosphate groups can create strongly attractive electrostatic interactions with basic side chains, which makes CDL an idea model lipid to understand interactions, but it also exhibits more specific patterns. Of note, both the head groups and all four acyl chains are thought to be important components of supercomplex stabilization (11). Similarly, CDL plays an essential role in the dimerization of the Na^+/H^+ antiporter NhaA, which increases the exchanger activity to protects the bacteria from osmotic stress (12, 13). Therefore, sites displaying preferential CDL binding may indicate lipid-activated regulatory mechanisms. To address this possibility, we have previously used coarse-grained molecular dynamics (CG-MD) simulations to map CDL binding sites on *E. coli* inner membrane proteins with published structures, identifying specific amino acids and binding site geometries that mediate preferential interactions with CDL (14). Although such CDL "fingerprints" are found in a wide range of proteins with different activities, they stop short of clarifying the functional role of lipids at these sites, with predictions of their functionality remaining largely speculative. Addressing this knowledge gap requires monitoring both the molecular interactions as well as the structure or stability of membrane protein complexes. For instance, thermal-shift assays provide data on lipid binding and associated changes in protein (15). Moreover, native mass spectrometry (nMS) has gained traction for membrane protein analysis, revealing the influence of lipids on oligomerization (16), binding affinities (17), and conformational stability (18). Monitoring mass shifts captures individual lipid interactions across multiple protein populations, while gas-phase dissociation provides insight into lipid stabilization. nMS thus captures key features of regulatory lipid interactions, and is especially powerful when coupled with MD which provides insight at the atomistic level (19).

Being able to connect individual lipid binding events to the stability of a protein complex is a crucial step towards predicting functionally important CDL interactions. We reasoned that a combined MD and nMS strategy may reveal basic requirements for CDL-mediated stabilization. However, the sequence and structures of membrane proteins are evolutionarily entrenched with the lipid composition of their surrounding membrane. To reduce the system to first principles, we turned to TransMembrane Helical Core Tetramer_Rocket-shaped (TMHC4_R, hereafter referred to as ROCKET), an artificial membrane protein tetramer whose sequence was derived from Rosetta Monte Carlo calculations (20). ROCKET includes a generic lipid-water interface composed of a

ring of aromatic residues and a ring of positively charged residues on the cytoplasmic side. Into the ROCKET scaffold, we designed several CDL binding sites based on our observations from *E. coli* proteins and tested their effect on tetramer stability using nMS. We find that local dynamics and the spatial distribution of charged residues distinguish stabilizing from non-stabilizing sites. Screening our database of *E. coli* CDL binding sites (https://osf.io/gftqa/) for similar features, we uncover a highly stabilizing CDL interaction in the membrane protease GlpG, which regulates the substrate preference of the enzyme. Together, this demonstrates the potential to design functional CDL sites on artificial proteins that can recognize membrane compositions.

Results

Design of a CDL binding site in ROCKET

As first step, we characterized the inherent lipid binding properties of ROCKET (Fig 1a) through CG-MD simulations of the protein in a mimetic *E. coli* membrane. The membrane composition was modeled with a distribution of POPE, POPG, and CDL in a ratio of 67:23:10, and the system was simulated for 5 x 10 µs while monitoring the lipid interactions. We observed abundant lipid interactions, with CDL displaying markedly more localized binding than POPE or POPG (Fig S1). The N-terminal region on the first transmembrane helix bound CDL with average occupancy of 71 % and average residence time of 35 ns (at R9). The site, which we termed Site 1, consists of three basic residues (R9, K10, and R13) and an aromatic residue (W12), which corresponds to a consensus CDL binding motif (Fig 1b) (*14*). We also observed a second site, involving W12 and R66 on helix 2 of the neighboring subunit. This site, termed Site 2, exhibited significantly lower occupancy of 56 % and an average residence time of 35 ns (at R66). Both sites represent distinct lipid binding modes: Site 1 is a high-occupancy site away from the protein core with extensive head-group interactions, and Site 2 is a lower-occupancy site with extensive acyl chain contacts close to the protein core. We therefore decided to use these two sites, which arose from purely statistical distribution of charged and aromatic residues, as basis for engineering a stabilizing CDL site.

To separate the two CDL binding modes, we generated ROCKET mutants *in silico* and performed CG-MD with two subunits each of ROCKET and ROCKET mutants. We found that substituting the charged residues of Site 1 with alanine (R9A/K10A/R13A, Fig 1c) redirected preferential CDL binding to Site 2 (Fig 1d, e). In this mutant, which we termed ROCKET^{AAXWA}, the Site 2 had an occupancy of 53 % and an increased average residence time of 47 ns (R66), whereas Site 1 had a reduced occupancy of 45 % and a residence time of 45 ns (R9A). This occupancy difference was quantified by CG simulations and showed significant reduction of total CDL binding between ROCKET and ROCKET^{AAXWA} (Fig 1d, S1).

Next, we evaluated the potential for lipid-mediated stabilization at both sites using gas-phase atomistic MD simulations, which allows for a direct comparison with nMS. We applied a pulling force between two adjacent subunits of ROCKET and ROCKET^{AAXWA} tetramers with and without bound CDL and determined the force required to separate the protein chains (Fig 1f). We found that separating the adjacent helices from the neighbouring subunits of ROCKET^{AAXWA} by 1.1 nm, the point at which non-covalent interactions between the transmembrane helices are disrupted, required more force when CDL was present (Fig 1g, Fig S2). ROCKET, on the other hand, displayed a lower significant difference in force with or without CDL. Snapshots from the simulations reveal that the lipid forms multiple contacts with both subunits adjacent to Site 2 in

ROCKET^{AAXWA}, which likely gives rise to the stabilizing effect (Fig 1h). We conclude that channeling the CDL molecules to inter-helix sites may be a prerequisite for lipid-mediated stabilization.

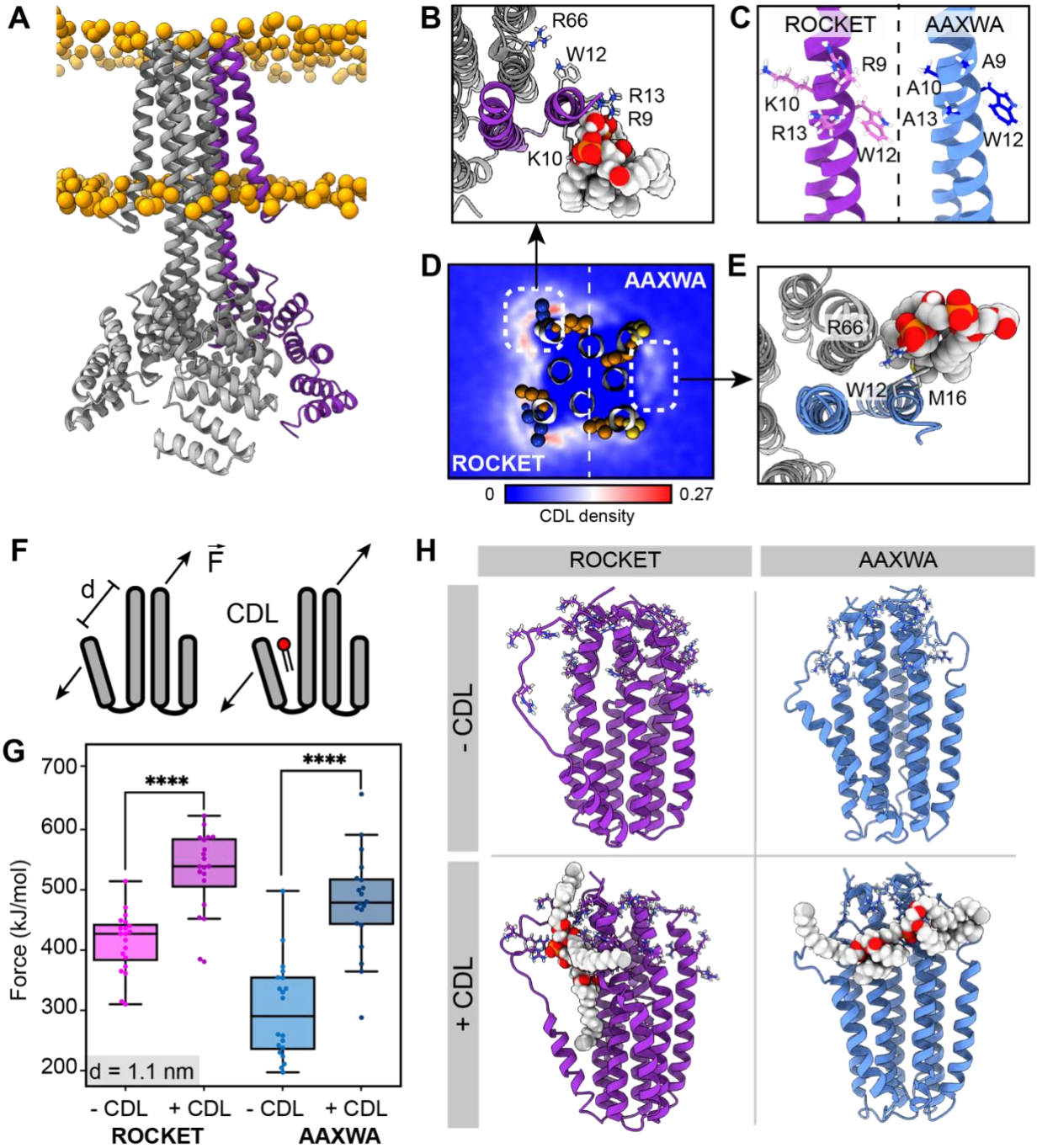

Fig 1. ROCKET contains two CDL binding sites with different structural implications.
(A) Structure of ROCKET (PDB ID 6B85) in the membrane, with one protein subunit highlighted in purple and phosphate headgroups fo the membrane shown as orange spheres. The structure was obtained from MemProtMD.

(B) Top view of CDL binding to Site 1 in equilibrated CD-MD simulations. CDL is shown as spacefill, and residues R9, K10, W12, and R13 as sticks. The CDL-binding subunit is highlighted in purple.
(C) Design of the ROCKET^{AAXWA} variant. Site 1 on helix 1 of ROCKET is shown on the left (purple), and ROCKET^{AAXWA} with the mutations R9A/K10A/R13A on the right (blue).
(D) CG-MD-derived CDL densities around a heterotetramer composed of two ROCKET subunits (left) and two ROCKET^{AAXWA} subunits (right). Site 1 on ROCKET and Site 2 on ROCKET^{AAXWA} are highlighted by dashed boxes. R9, K10, W12, and R13 are shown as spheres.
(E) Top view of CDL binding to Site 2 in ROCKET^{AAXWA}. Interacting residues W12, M16, and R66 on the neighboring subunit (grey) are shown as sticks.
(F) Setup of gas-phase MD simulations for dissociation of ROCKET and ROCKET^{AAXWA} with and without lipids.
(G) Plots of the integral of the force required to separate helix 1 and 2 (d = 1.1 nm), for ROCKET (purple) (p=3.85*10-7) and ROCKET^{AAXWA} (blue) (p=2.93*10^{-8}) with and without bound CDL show a more pronounced increase in stability for CDL-bound ROCKET^{AAXWA} compared to ROCKET (two-tailed t-test with n=20).
(H) Snapshots from gas-phase MD simulations show broad interactions of CDL across the subunits of lipid-bound ROCKET^{AAXWA} (blue) and more localized interactions with fewer intermolecular contacts for ROCKET (purple).

<u>Inter-helix CDL binding stabilizes ROCKET^{AAXWA} in the gas-phase</u>
Having derived two ROCKET variants with distinct CDL binding modes from MD simulations, we turned to cryogenic electron microscopy (cryo-EM) and nMS to investigate their lipid interactions experimentally. We first analyzed ROCKET and ROCKET^{AAXWA} in the presence of CDL by cryo-EM. The resulting density maps, refined to a resolution of 3.8 and 3.9 Å respectively, show essentially identical architectures that agree with the previously solved crystal structure of ROCKET (*20*), confirming that the mutations do not disrupt the native structure (Fig S3, Table S1). Although a definitive atomic-level molecular model was not possible at this resolution, we also observed in both maps a diffuse non-protein density which partially overlaps the head-group of CDL predicted in Site 2 (Fig S3). To determine lipid binding preferences, we therefore reconstituted the proteins into liposomes composed of polar *E.coli* polar lipid extracts (Fig 2a). By releasing the proteins from the liposomes inside the mass spectrometer and monitoring the intensity peaks corresponding to apo- and lipid-bound protein, we can compare lipid preferences of both variants (Fig 2a). We find that tetrameric ROCKET retains up to three CDL molecules, which can be identified by their characteristic 1.4 kDa mass shift, as well as a significant amount of phospholipids between 700 and 800 Da (Fig 2b). The data thus show a preference for CDL, which constitutes only 10% of the liposome. Interestingly, nMS of ROCKET^{AAXWA} revealed a similarly specific retention of up to three CDL molecules, although the intensity of the lipid adducts was reduced by approximately 50% (Fig 2c). The mass spectra show that the preference for CDL is preserved in the ROCKET^{AAXWA} variant, while the occupancy is reduced, indicating either lower affinity in solution or lower stability of the protein-lipid complex in the gas-phase.

To validate these findings, we analyzed lipid binding to detergent-solubilized ROCKET. Having optimized detergent conditions for maintaining the intact ROCKET tetramer (Fig S4), we performed a competition assay where we mixed both variants in C8E4 detergent containing a limiting amount of CDL (50 µM) and monitored binding with nMS. As expected, both variants

bound a similar number of CDL molecules per tetramer, however, ROCKET displayed significantly more intense lipid adducts than ROCKET^{AAXWA} (Fig S4). The nMS data are therefore consistent with CDL binding preferentially to Site 1 in ROCKET and preferentially to Site 2 in the ROCKET^{AAXWA} variant, in good agreement with the MD simulations. Next, we explored how CDL binding to either site affects the stability of ROCKET, using the oligomeric state in nMS as a measure. For this purpose, we employed gas-phase dissociation of the intact protein complexes. Briefly, collisions with gas molecules in the ion trap of the mass spectrometer cause thermal unfolding of a single subunit in the complex, which is then ejected as a highly charged, unfolded monomer (Fig 2d) (*21*). By comparing the peak intensities of the monomers that are ejected simultaneously from two protein oligomers, we can obtain information about their relative stabilities. Therefore, by adding CDL to an equimolar mixture of ROCKET and ROCKET^{AAXWA}, dissociating the resulting complexes, and monitoring changes in monomer signal intensities, we can determine whether lipid binding to Site 1 or Site 2 affects tetramer stability (Fig 2e). Importantly, by comparing changes in peak intensities with and without CDL while keeping all other conditions constant, we can avoid interference from changes in gas-phase fragmentation or ionization efficiency. nMS of ROCKET and ROCKET^{AAXWA} shows the release of highly charged monomers which can be distinguished based on their masses (Fig 2f). We then added CDL to the protein solution and repeated the measurement using identical conditions. We observed a reduction in the peak intensities of ROCKET^{AAXWA} monomers compared to ROCKET (Fig 2g). We do not observe a change in the charge state distributions for tetramers or monomers, or notable fragmentation. Therefore, the change in monomer ratio suggests that CDL stabilizes the ROCKET^{AAXWA} tetramer to a greater extent than the ROCKET tetramer. Considering the predictions from CG- and gas-phase MD, we conclude that the increase in stability is due to the preferential binding of CDL to the inter-helix Site 2 in the AAWXA variant. CDL binding to the distal Site 1, as preferred in ROCKET, involves fewer intermolecular contacts, and does therefore not exhibit a similarly stabilizing effect.

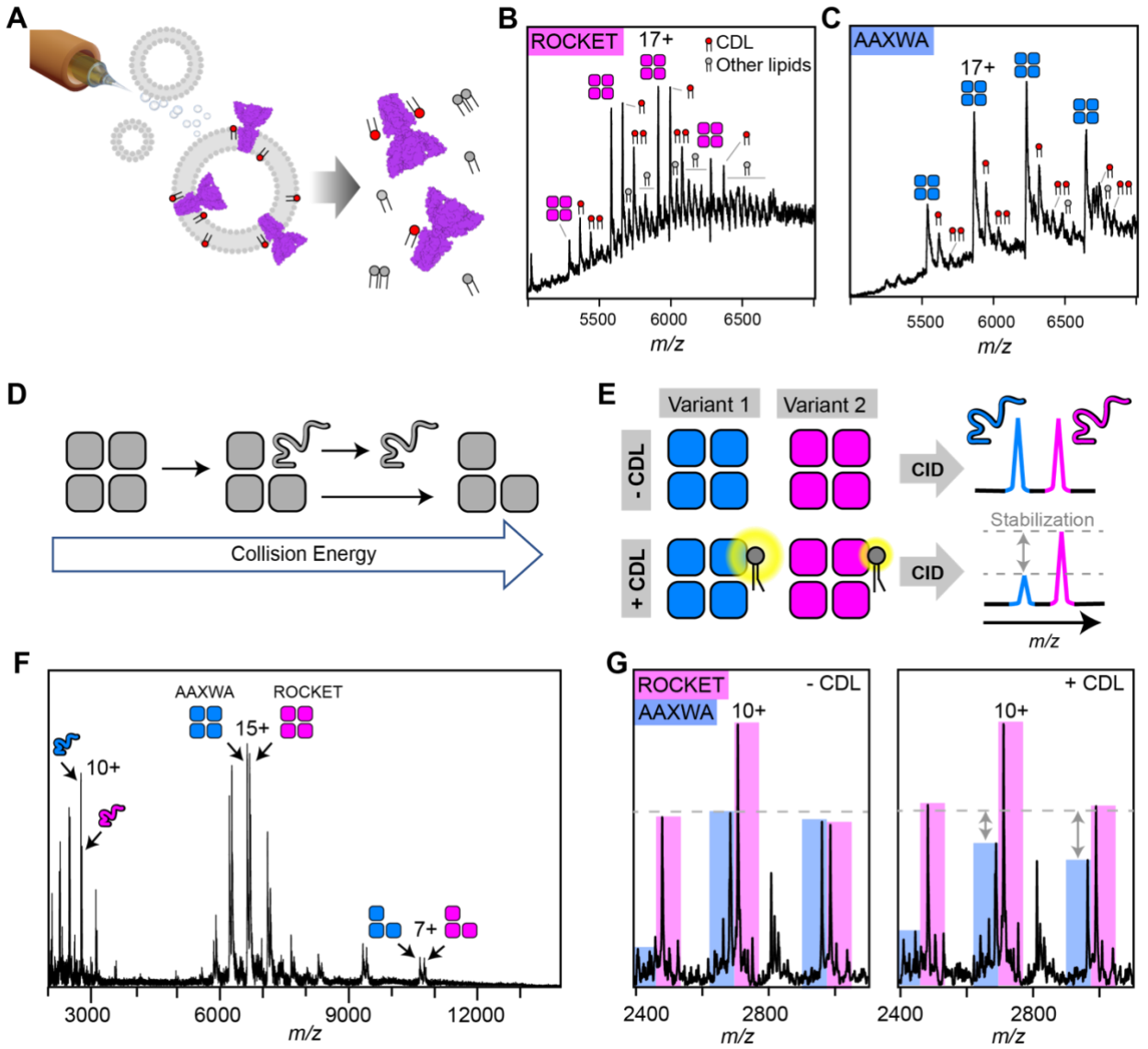

Fig 2. nMS analysis of lipid binding and lipid mediated stabilization of ROCKET and ROCKET^{AAXWA}.

(A) Schematic depiction of electrospray ionization (ESI) process for proteo-liposomes, leading to the ejection of protein-lipid complexes into the gas-phase.

(B) A representative mass spectrum of ROCKET released from proteoliposomes shows tetramers with 1-3 bound CDL molecules, as judged by the characteristic mass shift of 1.4 kDa, as well as additional lipids with molecular weights between 700 and 800 Da.

(C) Release of ROCKET^{AAXWA} from proteoliposomes shows retention of 1-3 CDL molecules per tetramer. The reduced lipid adduct intensity compared to ROCKET indicates reduced lipid binding and/or complex stability.

(D) Schematic illustrating the process of gas-phase subunit unfolding and ejection from ROCKET tetramers at increasing collision energies.

(E) A nMS assay to assess CDL-mediated stabilization of ROCKET and ROCKET^{AAXWA}. Simultaneous dissociation of ROCKET and ROCKET^{AAXWA} leads to the ejection of unfolded monomers, which can be quantified by nMS (top row). Addition of CDL to the same mixture results

in lipid binding to tetramers. If CDL binding stabilizes one tetrameric variant more than the other, the amount of ejected monomers will be reduced accordingly (bottom row).
(F) Representative mass spectrum of ROCKET and ROCKET^{AAXWA} at a collision voltage of 220 V. Intact tetramers are seen in the middle, ejected monomers and stripped trimers are seen in the low and high *m/z* regions, respectively.
(G) Zoom of the low *m/z* region of a mixture of 25 µM each of ROCKET and ROCKET^{AAXWA} with a collision voltage of 200 V before (left) and after (right) the addition of 50 µM CDL. The three main charge states for both variants can be distinguished based on their mass difference. Addition of CDL reduces the intensity of the ROCKET^{AAXWA} monomer peaks compared to ROCKET (dashed line).

Multiple structural features impact CDL-mediated stabilization

The finding that inter-helix CDL binding stabilizes a tetrameric membrane protein in the gas-phase recapitulates a key feature of both prokaryotic and eukaryotic membrane proteins (*16*, *22*). Unlike naturally evolved proteins, however, the extraordinary stability and mutation tolerance of the ROCKET scaffold enables us to dissect further the requirements for CDL-mediated stabilization. We therefore applied the above MS strategy to quantitatively assess lipid-mediated stabilization in our model system between two protein variants using ROCKET^{AAXWA} as an internal reference. We can determine relative stability changes upon CDL addition for different ROCKET variants by plotting the ratio of mutation (ROCKET^{MUT}) to total protein monomers (ROCKET^{MUT} + ROCKET^{AAXWA}) with and without CDL. If ROCKET^{AAXWA} is stabilized more than the variant of interest, the ratio increases with CDL addition (Fig 3a). As expected, the AAXWA mutation significantly increased the stabilizing effect of CDL, as determined from four independent repeats (Fig 3b and 3c). With this assay, we then explored whether introducing different structural features into Site 1 could turn it into a stabilizing CDL binding comparable to Site 2. As a first hypothesis, we reasoned that a destabilization of the core of ROCKET might increase the effect of CDL. We introduced a destabilizing mutation (A61P) in helix 2, right below the headgroup region, theorizing that the proline-induced kink would destabilize the ROCKET tetramer. AlphaFold2 predictions (*23*, *24*) indicated that the ROCKET^{A61P} mutation does not affect the tetrameric state (*25*), which was confirmed by nMS. To our surprise, ROCKET^{A61P} exhibited significantly less CDL stabilization than ROCKET^{AAXWA}, and was comparable to ROCKET (Fig 3d, Fig S5). This finding indicates that the introduction of a proline in the protein core does not sufficiently destabilize the protein to be counteracted by lipid binding.

As second hypothesis, we reasoned that additional lipid binding sites may increase the effect of CDL binding. We therefore mutated residue D7 to alanine and S8 to arginine. The D7A/S8R variant retains the high-affinity Site 1 on helix 1, but includes an additional site composed of R8 on helix 1 and R66 and E68 on helix 2 which does not overlap with Site 1 (Fig 3d). Quantification of the monomer release with and without CDL suggests a shift towards increased stability with CDL, albeit not as pronounced as for ROCKET^{AAXWA} (Fig 3d, Fig S5). This data leads us to speculate that Site 1 may still bind the bulk of the available CDL molecules, or that the salt bridge S8-E68 may contribute increased stability in a CDL-independent manner.

Having explored directed lipid binding (AAXWA), increased lipid binding (D7A/S8R) and core destabilization (A61P), we reasoned that the flexibility of the CDL binding site may affect stabilization. This feature is challenging to implement in the ROCKET scaffold, since it is designed

around a tightly folded hydrogen bond network with a melting temperature of > 90°C (*20*). We therefore decided to untether helix 1 from the core of the protein by mutating R66, which forms a salt bridge with D7, to alanine. The AlphaFold2 model shows helix 1 being tilted away form the core, creating a large hydrophobic gap in the acyl chain region and turning Site 1 towards the neighboring subunit (Fig 3f, Fig S5). Interestingly, ROCKET^{R66A} showed lower signal intensities in native mass spectra than all other variants, which may indicate overall lower stability in detergent. Quantification of the monomer release with and without CDL revealed no significant difference compared to ROCKET^{AAXWA}, which we attribute to CDL stabilizing both proteins to a similar extent (Fig 3f). Stabilization of ROCKET^{R66A} may occur in several ways, for example via acyl chain contacts between helix 1 and 2, or by moving CDL bound top Site 1 closer to the next subunit.

From the designed ROCKET variants, we can conclude that binding involving different helices, as in ROCKET^{AAXWA}, and connecting flexible regions, as in ROCKET^{R66A}, gives rise to the most pronounced CDL stabilization of our system (Fig 3g). Core destabilization, as well as introduction of additional headgroup contacts, had less of an impact, although the specific properties of the engineered protein scaffold may mitigate potential effects to some extent.

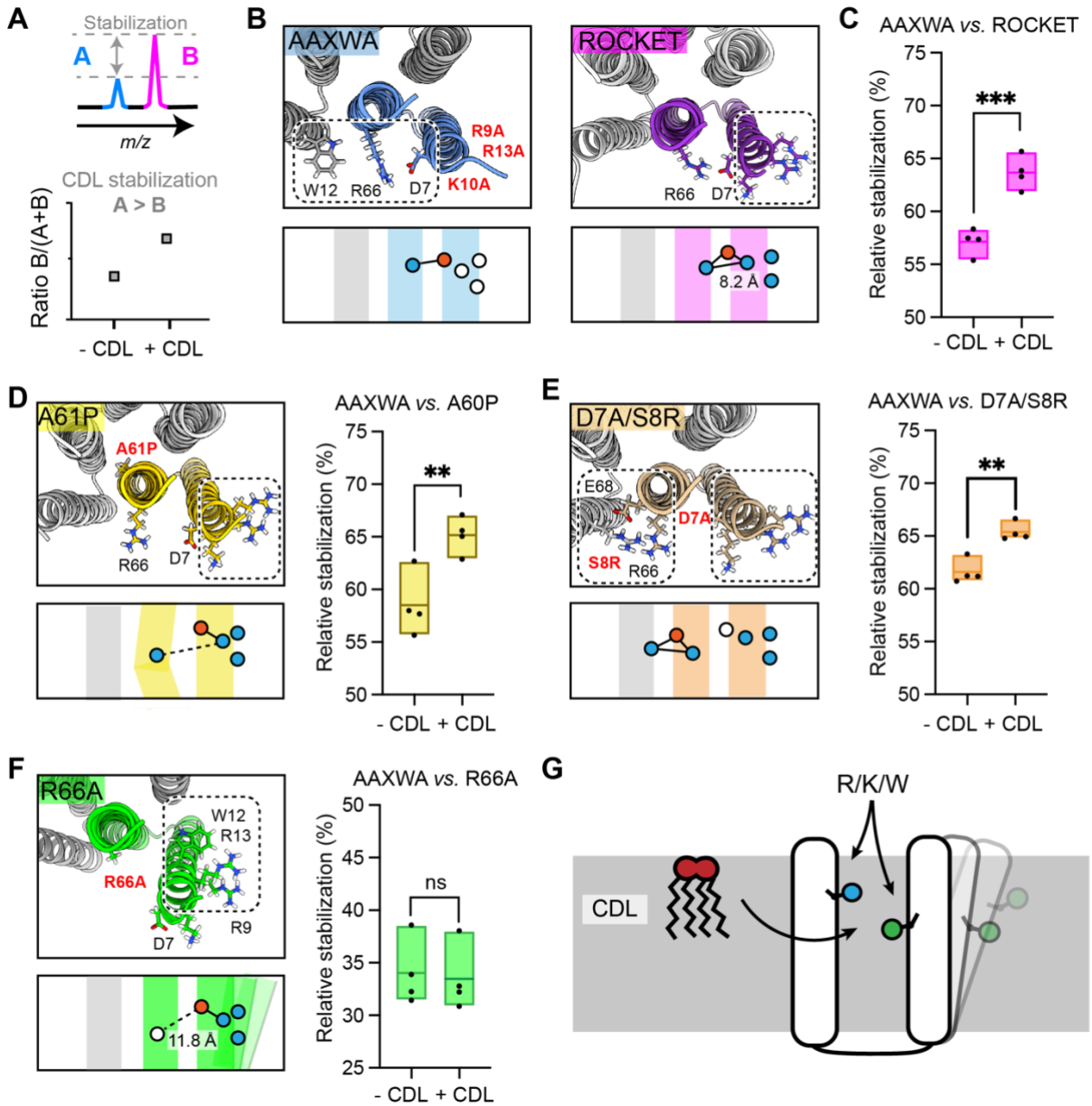

Fig 3. Assessment of lipid-mediated stabilization effects on ROCKET variant.
(A) Principle for pairwise analysis of protein stabilization by CDL. Peaks representing monomers released of protein A and B display intensity changes upon lipid addition. Plotting the peak intensities as a ratio of B to total protein (A + B) shows an increase upon CDL addition, if A is stabilized more than B.
(B) Residues involved in CDL headgroup binding to ROCKET and ROCKET^{AAXWA} were derived from CG-MD simulations (Fig 1) and are shown based on AlphaFold2 models as top view, with the area occupied by CDL as a dashed rectangle, and as side view schematic. A single subunit is colored (purple for ROCKET and blue for ROCKET^{AAXWA}).
(C) Plotting the peak intensity ratios of ROCKET to (ROCKET^{AAXWA} + ROCKET) in the presence and absence of CDL shows a decrease in ROCKET^{AAXWA} monomers when CDL is added ($p=0.0007$, two-tailed t-tests with $n=4$).

(D) The CDL binding site and location of the A61P mutation (yellow) mapped on the AlphaFold2 model of ROCKET^{A61P} and shown as a schematic as a side view below. Intensity ratios show significantly more pronounced stabilization of ROCKET^{AAXWA} than ROCKET^{A61P} (p=0.0084, two-tailed t-tests with n=4).
(E) Introduction of a second CDL binding site in the D7A/S8R variant (orange) mapped on the AlphaFold2 model of ROCKET$^{D7A/S8R}$ and shown as a schematic as a side view below. The shift in intensity ratios show that ROCKET^{AAXWA} is still stabilized to a greater extent (p=0.0019, two-tailed t-tests with n=4), but with a smaller margin than ROCKET or ROCKET^{A61P}.
(F) The R66A mutation, designed to disconnect helix 1 from the tetrameric protein core, results in an outward rotation of the CDL binding site, as shown in the AlphaFold2 model (green) and the side view schematic. Intensity ratios show no change upon CDL addition, suggesting that ROCKET^{R66A} is stabilized to a similar extent as ROCKET^{AAXWA} (p=0.8113, two-tailed t-tests with n=4).
(G) Conceptual diagram depicting structural features that promote CDL-mediated stabilization. Distributing the residues that interact with the lipid headgroup, usually basic and aromatic residues, between two helices, as well as Involvement of flexible protein segments, indicated by an outward movement of the right helix, also enhances stabilization by CDL.

Identification of a stabilizing CDL binding site in the *E.coli* rhomboid intramembrane protease GlpG

As outlined above, the features that cause CDL-mediated stabilization of the ROCKET scaffold were designed based on observations from our large-scale CG-MD investigation of CDL interactions with of *E. coli* membrane proteins. We therefore asked whether the same features could indicate stabilizing, and by extension, functionally relevant CDL interactions in naturally occurring proteins. To test this hypothesis, we evaluated our database of monomeric *E. coli* membrane proteins CDL sites (https://osf.io/gftqa/) for the sequence distribution of basic residues to find binding sites that span multiple helices. We reasoned that if two or more basic residues that interact with the same CDL molecule are located further apart in the sequence than approximately 30 positions, they have a high likelihood of being on separate helices, whereas sites spanning less than 10 residues are confined to a single helix or loop. Plotting the maximum sequence distance on a log scale reveals a bimodal distribution, with 75 CDL sites spanning less than 30 positions, and 180 sites spanning more than 30 positions (Fig 4a, Fig S5). Site 1 in ROCKET is in the first group, with four positions between R9 and R13. This approach does not consider interfacial CDL molecules in homo-oligomers, which may bind *via* single residues on different subunits. We therefore limited the dataset to monomeric proteins with CDL sites spanning >30 residues which we manually inspected to find CDL sites linking potentially flexible regions.

The rhomboid intramembrane protease GlpG from *E.coli* contains a CDL site between R92 on helix 1 and K167 on helix 4, as well as one aromatic residue on each (W98 and Y160). Importantly, R92 borders on a disordered linker region that connects helix 1 to an N-terminal cytoplasmic domain (Fig 4b). These features suggest that CDL binding may have structural implications. We therefore measured whether any lipids imparted thermal stabilization in a GFP-based thermal shift assay (*15*). Measuring the fraction of GFP-tagged soluble protein after heating to 63 °C in the presence of different lipids, we found that phosphatidyl ethanolamine (PE), phosphatidyl choline (PC), as well as none *E. coli* lipid sphingomyelin (SM) had no stabilizing effect, whereas phosphatidyl glycerol (PG) and CDL resulted in near-complete protection form heat-induced precipitation (Fig 4c).

When performing CG-MD of GlpG in the model *E. coli* bilayer, we observed that CDL bound exclusively to the predicted site between helix 1 and 4, with an occupancy of 96.2% with an average residence time of 170 ns, but no preferential POPG binding was observed anywhere on the protein (Fig 4d, Fig S6). The data suggest that the stabilizing effect of POPG in the thermal shift assay is due to binding to this site in the absence of CDL, whereas CDL outcompetes POPG in the mixed lipid environment of the membrane. To test whether GlpG binds CDL in the membrane, we used nMS analysis. We have previously noted a 1.4 kDa adduct on purified GlpG (*26*). Collisional activation revealed preferential loss of the adduct as a single species, confirming that it is a single lipid molecule, *i.e.* CDL (Fig 4e). These findings are in good agreement with the presence of a double-lipid density in this location in the crystal structure of GlpG (Fig S6) (*27*). The apparent high specificity for CDL prompted us to investigate the effect of CDL and PG on GlpG protease activity. For this purpose, we used a cleavage assay with fluorescently labeled substrates that represent either a soluble "extramembrane" substrate (KSp63) or transmembrane helix (KSp96) peptides containing the same GlpG cleavage site (*28*, *29*). We found that PG and CDL both have a positive effect on the cleavage of the soluble substrate KSp63 compared to detergent-only, with both lipids increasing the cleavage rates by approximately 40% (Fig 4f). PG mildly reduced the cleavage rate of the transmembrane substrate KSp96 to approximately 90% compared to detergent-only. Strikingly, CDL caused drastic inhibition of transmembrane substrate cleavage, with only 20% activity remaining (Fig 4f). To find out how the lipid binding site, which is located on the cytoplasmic side, can affect substrate access to the active site, which faces the periplasm, we performed CG-MD simulations of the open and closed conformations of GlpG (*30*). We found that both conformations bound CDL in the same place, which agrees with the observation that CDL affects the flexibility of the linker connecting the N-terminal domain (*30*). Interestingly, the acyl chains of the tightly bound lipid extend into the gap between helices 2 and 5 (Fig S6), the lateral gate through which transmembrane substrates access the active site of GlpG (*30*). Loop 5, whose movement is critical for catalytic activity, remains unaffected, providing a rationale for the selective inhibition of transmembrane substrate cleavage by CDL. Although the biological role of CDL-regulated substrate preferences of GlpG remains to be clarified, our results demonstrate the identification of a functionally relevant CDL binding site in an *E. coli* membrane protein based on the insights from artificial protein design.

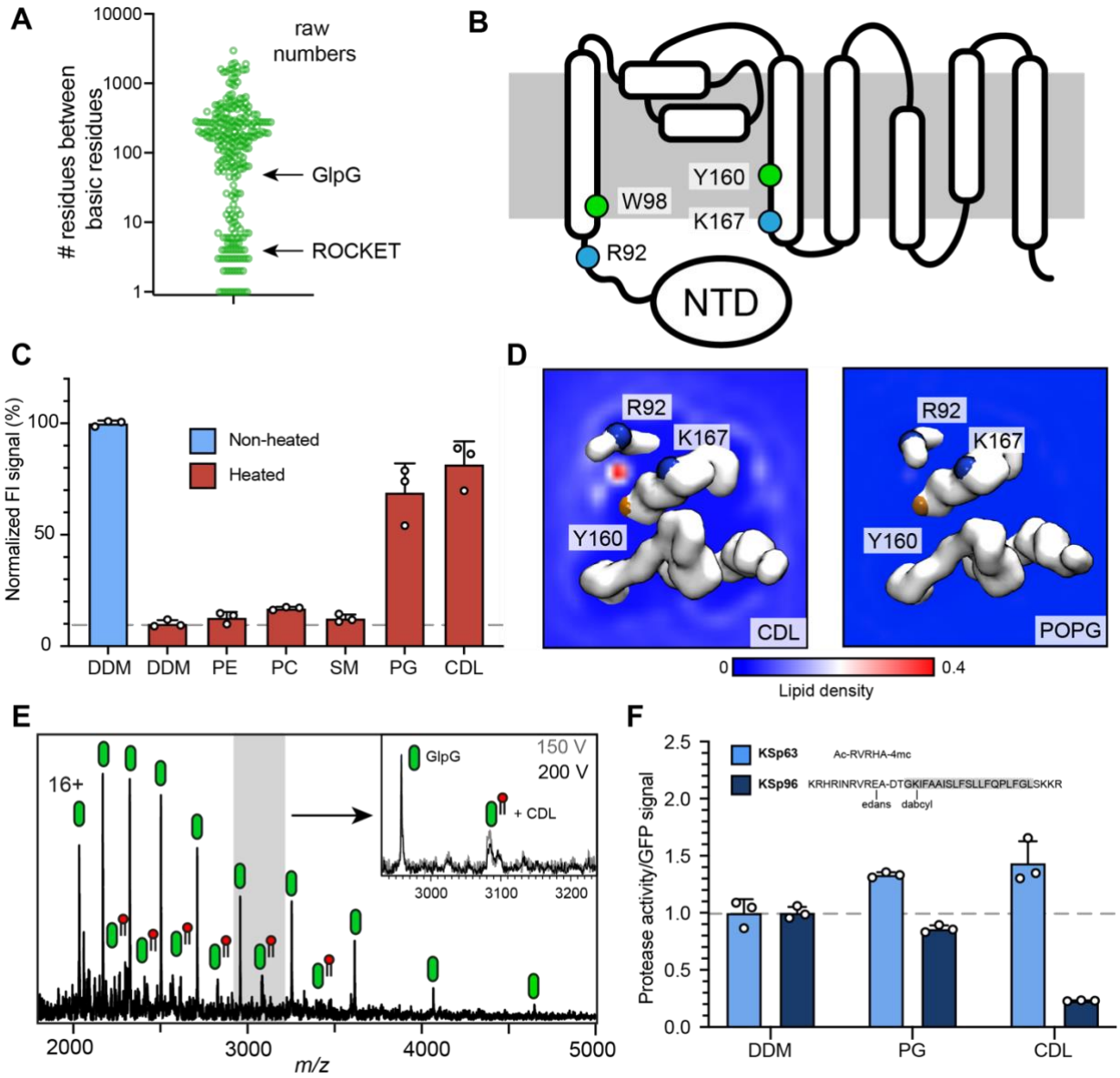

Fig 4. **Identification of a functional CDL binding site in the GlpG membrane protease.**
(A) Distances between basic residues within the same CDL binding site and their occurrence in *E. coli* membrane proteins. GlpG contains a CDL binding site with two basic residues separated by 75 positions. Site 1 in ROCKET (four positions) is indicated for reference.
(B) Topology diagram of GlpG indicating the locations of basic (R; arginine, L; lysine) and aromatic (W; tryptophan, Y; tyrosine) residues that interact with CDL.
(C) GFP-thermal shift assay of GlpG in detergent or in the presence of PE, PC, SM, PG, or CDL. The average fluorescence intensity (FI) indicates the fraction of soluble protein before (unheated) or after heating to 63 °C and removing precipitated protein. Data are normalized against the non-heated control in detergent. All measurements were performed in triplicates (n = 3).
(D) CG-MD-derived Lipid density plots for CDL (left) and POPG (right) around the transmembrane region of GlpG (PDB ID 2IC8) viewed from the cytoplasmic side. CDL, but not POPG, exhibits preferential binding to the R92-K167 site.

(E) nMS spectra of GlpG released from detergent micelles show a 1.4 kDa adduct which can be removed through collisional activation of the protein.
(F) GlpG proteolytic cleavage rates (specific activity, by normalizing to the GFP signal) for soluble (extramembrane) substrate (KSp63) or transmembrane substrate (KSp96, transmembrane helix shaded) in the presence of PG or CDL. Both lipids increase the cleavage rate for the soluble substrate by approx. 40%. For the transmembrane substrate, addition of PG causes a moderate reduction in cleavage activity, whereas CDL causes near-complete inhibition of cleavage.

Discussion

In this study, we have systematically investigated the requirements for CDL binding and stabilization of membrane proteins, using the scaffold protein ROCKET as a model system. Our findings reveal that CDL binding sites have few sequence requirements, which is evident from the fact that ROCKET contains two interaction sites solely from statistical distribution of aromatic and charged amino acids at the membrane interface (Fig 1). Both sites exhibit characteristic features, including electrostatic interactions with the negatively charged lipid headgroups and flanking aromatic residues that align the acyl chains with the hydrophobic transmembrane domain. By channeling CDL to either one of the two sites, we find they differ significantly in their ability to exert a stabilizing effect. Therefore, although CDL binding sites share structural features, their ability to contribute to lipid-mediated stabilization appears to be multifaceted.

By designing different CDL binding sites in ROCKET, we demonstrate that commonly observed structural features (the number of headgroup-interaction residues, local flexibility, and the involvement of multiple transmembrane helices) impact lipid-mediated stabilization to different extent. Building upon the insights garnered from ROCKET, we extended our investigation to *E. coli* proteins and identified GlpG as a case study for CDL's impact on function. Previous studies have shown that GlpG conformation is impacted by the surrounding lipids and membrane geometry (*30*, *31*), and CDL has been suggested to affect proteolytic activity (*31*). Our observation that CDL acts as an allosteric activator for the cleavage of soluble substrates while exerting an inhibitory effect on the processing of transmembrane substrates establishes CDL as a regulator of GlpG activity. These findings suggest that the features that mediate lipid stabilization in ROCKET are hallmarks of functionally important lipid binding sites in native membrane proteins.

Although we can identify relatively intuitive features of CDL interaction sites, we find that the connection between lipid binding and stabilization is not clear-cut. For example, one destabilizing mutation increases lipid-mediated stabilization whereas another does not (compare ROCKET^{R66A} and ROCKET^{A61P}). Furthermore, being able to bind more lipids do not translate to forming more stable complexes (compare ROCKET and ROCKET^{AAXWA}). The reasons are likely two-fold: Firstly, our approach has methodological limitations, as gas-phase stability is not easily correlated with condensed phase stability. In case of CDL, increasing the number of molecular contacts likely translates to stabilizing effects in both phases. However, more subtle contributions may be lost, making more detailed investigations challenging. Here, we have focused here on headgroup contacts, which can be readily predicted with CG-MD and analyzed with nMS. Secondly, our approach highlights that protein-lipid interactions are complex. The ROCKET scaffold is, *per* design, extremely stable and has no function. Thus, it does not capture the dynamic nature of most membrane proteins, which, in turn, is often related to lipid-mediated regulation (*12*). The key finding of our study is that emulating the architecture of a binding site is only the first step to uncovering the principles of lipid regulation, and that the effect of lipids on the local dynamics are

critical to understanding how they shape membrane protein function. Integrated design approaches based not only on static structures but on dynamic models are therefore the key to designing membrane proteins with membrane-specific functions.

In summary, we present a protein design-driven approach to decipher mechanisms by which CDL regulates membrane protein stability. Our findings illuminate critical protein features that are challenging to predict or design *de novo*. The methodology developed here, integrating MD simulations, nMS, and machine learning, offers a promising pathway to uncover novel CDL binding sites. The findings not only contribute to our understanding of lipid-protein interactions but highlight potential avenues for the design of membrane proteins with tailored stability and function, potentially informing therapeutic strategies targeting membrane protein dysfunctions.

Materials and Methods

Coarse-grained molecular dynamics simulations of ROCKET variants.

CG systems were built using PDB 6B85 (TMHC4_R) or 2IC8/2NRF (GlpG). Protein atoms were converted to the CG Martini 3 force field (*32*) using the martinize method (*33*). Additional bonds of 500 kJ mol^{-1} nm^{-2} were applied between all protein backbone beads within 1 nm. Where used, mutations were made in PyMOL (*34*). Proteins were built into membranes composed of 10% CDL, 23% POPG, and 67% POPE for ROCKET, or 10% CDL, 10% POPG, and 80% POPE or just 10% POPG and 90% POPE for GlpG. Membranes were built using the insane protocol (*35*). All systems were solvated with Martini waters and Na^+ and Cl^- ions to a neutral charge and 150 mM. Systems were minimized using the steepest descents method, followed by 1 ns equilibration with 5 fs time steps, then by 100 ns equilibration with 20 fs time steps, before 5 × 10 µs production simulations using 20 fs time steps, all in the NPT ensemble at 323 K with the V-rescale thermostat (τ_t = 1.0 ps) and semi-isotropic Parrinello-Rahman pressure coupling at 1 bar (τ_p = 12.0 ps). The reaction-field method was used to model long-range electrostatic interactions. Bond lengths were constrained to the equilibrium values using the LINCS algorithm. Density analyses was performed using the VolMap tool of VMD, with the default settings (*36*). Lipid binding sites and lipid-residue interactions were determined using the PyLipID package, which provides both occupancy and residence time data (*37*). Simulations were run in Gromacs 2022 (*38*, *39*).

Gas-phase molecular dynamics simulations of ROCKET variants.

In order to assess how lipid binding affects the stability of ROCKET and the $ROCKET^{AAWXA}$ variant in the gas-phase, molecular dynamics simulations were performed, in which chains were pulled apart and the force was measured. The simulations were done on truncated models comprised of the membrane-spanning domain that included residues S2-V79, with a neutral C-terminus (COOH) added to residue 79. In aiming to reflect the 16+ charge state, which was observed in mass spectrometry experiments, residues D38 and E78 were protonated, while all other titratable residues were given their pKa-based protonation state at pH 7. The addition of eight protons to the truncated model is intended to replicate the distribution of exposed acidic sites on the entire protein, which are likely to become protonated in the experiments. Simulations were performed with the GROMACS MD package (*40*), version 2023.3, and the July 2022 version of the CHARMM36 force field (*41*). The models were simulated with and without cardiolipin bridging the helices of neighbouring subunits. The proteins were placed in cubic boxes with sides of 999.9 nm, and the cutoff radii for Coulomb and van der Waals interaction were set to 333.3 nm.

This set-up allows the use of the Verlet buffer scheme (*42*) for neighbour searching and GPU acceleration while avoiding artefacts from the periodicity (*43*).
Virtual sites (*44*) were used for hydrogens, for which cardiolipin was generated with MkVsites (*45*). The four models were equilibrated with a steepest descent energy minimization until convergence, followed by temperature coupling over 10 ps, using a 0.5-fs time step, and the Berendsen (*46*) thermostat set to 300 K. For the production simulations, the thermostat was changed to velocity rescale (*47*), retaining the temperature of 300 K. All bonds were constrained with LINCS(*48*), using an order of 4 and one iteration, which allowed for a 5-fs time step to be used. Helices were pulled apart with the center-of-mass pull code. The reference groups included, respectively, the Ca atoms of residues 6-12 and 63-69 on the two chains with which cardiolipin was interacting. The groups were pulled with an umbrella potential and a harmonic force constant of 1000 kJ/mol/nm^2, using the distance in the x, y, z directions as geometry, and a pull rate of 0.1125 nm/ns. Twenty replicate pulling simulations of 20 ns each were performed using a random seed for the velocities.

Protein engineering of ROCKET mutants.

All protein mutagenesis was performed with the Q5 site-directed mutagenesis kit from NEB (E0554). Primers were designed with the NEBaseChanger tool (https://nebasechanger.neb.com/) and synthesized (Eurofins Genomics) with the sequences listed (Table S2). The original plasmid used as template for the PCR reactions was based on wildtype ROCKET plasmid (Table S3) and the reactions utilized the Q5 Hot Start High-Fidelity DNA Polymerases with annealing temperature optimized for each individual reaction (Table S2). PCR products were prepared using kinase, ligase, and DpnI (KDL) mixture, according to the manufacturer protocol, transformed into NEB 5-alfa cells, plated on Luria-Bertani (LB) agar plates containing kanamycin (50 µg/mL) and incubated overnight at 37°C. Individual colonies were selected for growth in 5 mL LB cultures at 37°C for 12-16 hours. 1-5 mL of cell pellet was harvested for plasmid extraction with the Monarch Plasmid DNA Miniprep Kit (New England Biolabs) (T1010). Mutagenesis was confirmed by sequencing.

Cloning and expression of ROCKET variants.

Synthetic genes with N-terminal histidine tag (either 6-His or 10-His for ROCKET^{AAXWA}) were synthesized by Genscript Inc or derived from mutagenesis in a pet26b (+) expression vector (Table S3). These plasmids were transformed into *E. coli* BL21 (DE3) (New England BioLabs). Selection was carried out on LB agar plates containing kanamycin (50 µg/mL) and incubated overnight at 37°C. Pre-cultures were grown under the same condition overnight and used to inoculate 400 mL LB media at a 1:100 dilution for protein expression. The expression cultures were incubated at 37°C until OD_{600} reached 0.8-1.0, at which protein expression was induced with 0.2 mM isopropylthio-β-galactoside (IPTG). Post-induction, the cultures were incubated at 18°C overnight for protein expression. Cells were harvested by centrifugation at 8 000 x g for 20 minutes.

Purification of ROCKET variants.

Cell pellets were homogenized in resuspension buffer (25 mM TRIS pH 8.0, 150 mM NaCl) and subjected to probe sonication. n-Dodecyl-beta-maltoside (DDM) was added to the lysate to a final concentration of 1 % (w/v) and the mixture was incubated overnight at 4°C with shaking. Following solubilization, the solution was centrifuged at 10 000 x g for 10 minutes the supernatant was then

filtered through a 0.2 µm syringe filter. The cleared supernatant was applied to a Ni^{2+} Sepharose High Performance column HisTrap (Cytiva) pre-equilibrated with wash buffer (25 mM TRIS pH 8.0, 150 mM NaCl, 30 mM imidazole, and 0.1 % DDM). The column was subsequently washed with the same buffer to remove unbound material.
Protein was eluted with an elution buffer (25 mM TRIS pH 8.0, 150 mM NaCl, 300 mM imidazole and 0.1 % (w/v) DDM). Elution fractions were collected and analyzed for protein content by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated using Amicon Ultra-15 centrifugal filter units with a 100-kDa molecular weight cutoff (Merck Millipore) Concurrently, the buffer was changed to remove imidazole, resulting in a final storage buffer (25 mM TRIS pH 8.0, 150 mM NaCl and 0.1 % (w/v) DDM) for downstream applications.

Cryo-EM sample preparation and data acquisition.
5 mg/mL ROCKET and $ROCKET^{AAXWA}$ with 100 µM CDL were frozen on Quantifoil 1.2/1.3 Au 300 mesh grids (Quantifoil Micro Tools). 3 µL of sample was applied to each grid which was then blotted for 3 s and plunge-frozen into liquid ethane using FEI Virtobot Mark IV (Thermo Fisher Scientific). Micrographs were collected on Krios G3i electron microscope (Thermo Fisher Scientific) operated at 300 kV equipped with Gatan BioQuantum K3 image filter and a Ceta-D detector. Movies were collected at a nominal 165,000 x magnification, resulting in a pixel size of 0.5076 Å. A total dose of 60 $e^-/Å^2$ was used to collect 29 frames over 1 s. The target defocus range was set between -0.6 to -1.8 µm, in steps of 0.2 µm.

Image processing & model building.
Data processing was performed using the RELION 4.0.1 pipeline (*49*). Motion correction was performed using Relion's own implementation (*50*) and CTF estimation was done with CtfFind4.1 (*51*). For the WT dataset, 2D references from 2D classification of manually picked particles were used for initial autopicking, and the picked particles were used to generate a low-resolution 3D reconstruction that was used and a 3D reference for the final round of autopicking. Due to high similarity in 2D classes from manually picked particles for the WT and MUT5 dataset, the same low resolution 3D reconstruction from the WT processing pipeline was used directly as a 3D reference for autopicking in the MUT5 data set. For both datasets, the auto picked particles were used for an ab initio reconstruction. The particles were then further refined, and the data cleaned using several rounds of 3D classification and 3D auto-refinement, followed by CTF parameters refinement and particle polishing before a final 3D auto-refinement and post-processing.

Preparation of proteoliposomes for native mass spectrometry
E. coli polar lipid extract, with a composition of 67.0% PE, 23.2% PG, 9,8% CDL (Avanti Polar Lipids) was dissolved in a 1:1 mixture of chloroform and methanol. The solvent was then removed under vacuum using a SpeedVac concentrator (Savant SPD1010) until completely dry. The dried lipid film was resuspended in buffer (25 mM TRIS pH 8.0, 150 mM NaCl) and vortexed vigorously to ensure homogeneity. Large unilamellar vesicles (LUVs) were prepared by extrusion though a pair of 0.4 µm polycarbonate membrane. The target protein-to-lipid ration was 1:100 (protein:lipid by weight) and the mixture was incubated at 37°C for 30 minutes to allow for protein incorporation into the lipid vesicles. Post-incubation, the samples were dialyzed against 500 mM ammonium acetate, pH 8.0, overnight to facilitate buffer exchange and removal of remaining detergent. The particle size distribution of LUVs and proteoliposomes were monitored by dynamic light scatter

(DLS) using a Viscotek model 802 DLS instrument with an internal laser (825-832 nm). Data processing was performed with OmniSIZE2.

Native mass spectrometry of proteoliposomes.

Proteoliposome samples were subjected to ESI nMS were performed on the Q Exactive Ultra-high range (UHMR) mass spectrometer (Thermo Fisher Scientific). The MS capillaries were custom-pulled and coated in-house (*52*). We set the capillary voltage to 1.5 kV and maintained the source temperature at 270°C. In-source trapping voltage was applied at 300 V to enhance ion desolvation, and the higher-energy collision dissociation (HCD) voltage was set to 200 V to facilitate ion transmission. The ultra-high vacuum pressure within the MS was measured at 6.01 x 10^{-10} mbar. Data was analyzed using Xcalibur 2.2 (Thermo Fisher).

Protein preparation for native mass spectrometry.

Immediately prior to MS analysis, the purified protein was subjected to size exclusion chromatography (SEC) using a Superdex 200 Increase 10/300 GL column (Cytiva). The detergent exchange process was conducted with native compatible buffer (200 mM ammonium acetate pH 8.0, 0.5% C8E4). The fractions containing the tetrameric state of ROCKET protein was collected for direct analysis with nMS.

Native mass spectrometry.

ESI-MS spectra were recorded on a Waters Synapt G1 wave ion mobility mass spectrometer, modified for high-mass analysis (MS Vision), and equipped with an offline nanospray source. The ESI-MS parameters were set as follows: capillary voltage at 1.5 kV, cone voltage at 100 V, source pressure maintained at 8 mbar, and source temperature regulated at 30°C. To optimize the detection of protein-detergent complexes, the trap voltage was varied from 90-240 V. For assessments of lipid binding ROCKET and ROCKET^{AAXWA} variants were prepared in samples with and without 50 µM 16:0 cardiolipin (Avanti Polar Lipid). The nMS setting were the same as above, with a trap voltage of 170 V. For each condition, four protein-lipid mixtures were prepared and measured separately. The mass spectra were analyzed using MassLynx software version 4.1 (Waters), and the intensities of apo- and lipid-bound tetramers were quantified with mMass V3.9.0.

To measure lipid-mediated stabilization, 15 µM ROCKET or ROCKET^{MUT} were mixed with 15 µM ROCKET^{AAXWA} at an equimolar ratio to achieve comparable intensities for tetrameric species at a collision voltage of 170 V. The mixtures were the supplemented with 16:0 cardiolipin (Avanti Polar Lipid) to a final lipid concentration of 25 µM. nMS setting were consistent with previous experiments and the collision voltage set to 200-220 V to allow optimal detection of unfolded monomers. Four spectra were recorded for each condition (n=4). The mass spectra were analyzed using MassLynx software version 4.1 (Waters), and the intensities of the monomers were quantified with mMass V3.9.0.

Screening of lipid impact on of GlpG-GFP fusion protein stability.

GlpG-GFP fusion protein was expressed and purified as described previously (*15*). The purified fusion protein was diluted in buffer containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 % (w/v) β-OG and 1% (w/v) DDM to a final concentration of fusion protein of 1 µM, and individual lipids were added to this preparation to a final concentration of 0.3 mg/mL. The used lipids were from Avanti Polar Lipids: 18:1 PE (cat no. 850725P), 18:1 PG (cat no. 840475P), 18:1 PC (cat no.

850375P), 18:1 CDL (cat no. 710335P) and brain SM (cat no. 860062P), and DDM was used as a negative control. Samples were incubated at 63°C for 10 min (the negative control at 4°C) followed by centrifugation at 20 000xg at 4°C for 45 minutes. Fluorescence of the supernatant was measured with TECAN Infinite M1000 spectrophotometer with excitation at 488 nm and emission at 512 nm.

Coarse-grained molecular dynamics simulations of GlpG.

Post-CG simulation snapshots were converted to all-atom description using the CG2AT approach (*53*) Atoms were described using the CHARMM36m force field (*54*) with TIP3P water. Systems were energy minimized using the steepest descents method, and subsequently equilibrated with positional restraints on heavy atoms for 100 ps in the NPT ensemble at 303 K with the V-rescale thermostat (τ_t = 1.0 ps) and semi-isotropic Parrinello-Rahman pressure coupling at 1 bar (τ_p = 5.0 ps), with a compressibility of 4.5×10^{-5} bar^{-1}. The Particle-Mesh-Ewald (PME) method was used to model long-range electrostatic interactions. van der Waals (VDW) interactions were cut off at 1.2 nm. Bond lengths were constrained to the equilibrium values using the LINCS algorithm. Production simulations were run in triplicate with 2 fs time steps for *ca*. 530 ns unless specified. Simulations were run in Gromacs 2022 (*38*, *39*).

Determination of protease activity of GlpG.

To determine GlpG activity, the GlpG-GFP supernatant collected after centrifugation was diluted 1:10 into 50 mM phosphate buffer pH 7.4, 150 mM NaCl, 0.05 % (w/v) PEG 8000, 20 % (v/v) glycerol, and 0.05 % (w/v) DDM. Lyophilized substrate were dissolved in the same buffer, with further addition of 5% (v/v) DMSO in case of the soluble substrate KSp63 (*28*), and preincubated at 37°C. The concentration of substrates in these master mixes were 400 µM for the 'soluble' (extramembrane) substrate KSp63 and 50 µM for the transmembrane substrate KSp96 (*29*). Cleavage reactions were initiated by mixing the enzyme solution and substrate master mix in 1:1 ratio, yielding final concentrations of 200 µM KSp63 and 25 µM KSp96. Protease activity was measured at 37°C by reading fluorescence intensity continuously every 30 seconds in TECAN Infinite M1000 for 40 min or until the upper detection limit was reached. Excitation was set to 355 nm and emission to 450 nm for the soluble substrate KSp63, or to 335 nm and 493 nm, respectively, for the transmembrane substrate KSp96.

Acknowledgments

The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at the PDC Center for High Performance Computing, KTH Royal Institute of Technology, partially funded by the Swedish Research Council through grant agreement (2022-06725). Additional computational resource came from ARCHER and JADE UK National Supercomputing Services, provided by HECBioSim, the UK High End Computing Consortium for Biomolecular Simulation (https://www.hecbiosim.ac.uk), which is supported by the EPSRC (EP/L000253/1). EM data was collected at the Karolinska Institutet 3D-EM facility (https://ki.se/cmb/3d-em). M.L.A is supported by a VR Research Environment grant (2019-02433). M.L. is supported by a KI faculty-funded Career Position, a Cancerfonden Project grant (22-2023 Pj), a VR Starting Grant (2019-01961), and a Consolidator Grant from the Swedish Society for Medical Research (SSMF). K.S. and J.S. have been supported by the Operational Programme European Regional Development Fund (no. CZ.02.1.01/0.0/0.0/16_019/0000729).

K.S. also acknowledges support from the InterCOST programme of the Ministry of Education, Youth and Sports of the Czech Republic (project no. LUC23180).C.V.R. and A.O.O. are supported by the Medical Research Council (MRC) programme grant (MR/V028839/1). D.D. acknowledges support from Göran Gustafssons Foundation. O.A. was supported by a doctoral fellowship from Sven och Lilly Lawskis stiftelse. R.J.H. and E.L. by grants from the Swedish Research Council (2019-02433, 2021-05806) and Swedish e-Science Research Center. L.J.P and E.G.M are supported by a project grant from the Swedish Research Council (2020- 04825). P.J.S. acknowledges Wellcome (208361/Z/17/Z), MRC (MR/S009213/1), BBSRC (BB/P01948X/1, BB/R002517/1, BB/S003339/1 and BB/Y003187/1), and the Howard Dalton Centre for funding. P.J.S. acknowledges Sulis at HPC Midlands+, which was funded by the EPSRC on grant EP/T022108/1, and the University of Warwick Scientific Computing Research Technology Platform for computational access.

Author contributions

Conceptualization and design: M.L.A., R.A.C., D.D., P.J.S., M.L.
Acquisition of data: M.L.A., R.A.C., J.S., L.J.P., O.A., A.O.O.
Analysis of data: M.L.A., R.A.C., J.S., L.J.P., O.A., A.O.O.
Supervision: R.J.H., E.L., C.V.R., K.S., E.G.M., D.D., P.J.S., M.L.
Writing—original draft: M.L.A., M.L.
Writing—review & editing: M.L.A., R.A.C., J.P.S., L.P., O.A., A.O.O., R.J.H., E.L., C.V.R., K.S., E.G.M., D.D., P.J.S., M.L.

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

Cryo-EM density maps of ROCKET and ROCKET^{AAXWA} in detergent micelles have been deposited in the Electron Microscopy Data Bank under accession number EMD-50106 and EMD-50107 respectively. All data are available in the main text or the supplementary materials.

References

1. I. Levental, E. Lyman, Regulation of membrane protein structure and function by their lipid nano-environment. *Nat. Rev. Mol. Cell Biol.*, 1–16 (2022).

2. V. Corradi, B. I. Sejdiu, H. Mesa-Galloso, H. Abdizadeh, S. Yu. Noskov, S. J. Marrink, D. P. Tieleman, Emerging Diversity in Lipid–Protein Interactions. *Chem. Rev.* **119**, 5775–5848 (2019).

3. Z. Cournia, A. Chatzigoulas, Allostery in membrane proteins. *Curr. Opin. Struct. Biol.* **62**, 197–204 (2020).

4. K. Renard, B. Byrne, Insights into the Role of Membrane Lipids in the Structure, Function and Regulation of Integral Membrane Proteins. *Int. J. Mol. Sci.* **22**, 9026 (2021).

5. J. C. Bozelli, S. S. Aulakh, R. M. Epand, Membrane shape as determinant of protein properties. *Biophys. Chem.* **273**, 106587 (2021).

6. C. Hunte, Specific protein-lipid interactions in membrane proteins. *Biochem. Soc. Trans.* **33**, 938–942 (2005).

7. H. Palsdottir, C. Hunte, Lipids in membrane protein structures. *Biochim. Biophys. Acta BBA - Biomembr.* **1666**, 2–18 (2004).

8. M. Landreh, M. T. Marty, J. Gault, C. V. Robinson, A sliding selectivity scale for lipid binding to membrane proteins. *Curr. Opin. Struct. Biol.* **39**, 54–60 (2016).

9. A. Musatov, E. Sedlák, Role of cardiolipin in stability of integral membrane proteins. *Biochimie* **142**, 102–111 (2017).

10. K. Pfeiffer, V. Gohil, R. A. Stuart, C. Hunte, U. Brandt, M. L. Greenberg, H. Schägger, Cardiolipin stabilizes respiratory chain supercomplexes. *J. Biol. Chem.* **278**, 52873–52880 (2003).

11. R. A. Corey, N. Harrison, P. J. Stansfeld, M. S. P. Sansom, A. L. Duncan, Cardiolipin, and not monolysocardiolipin, preferentially binds to the interface of complexes III and IV. *Chem. Sci.* **13**, 13489–13498 (2022).

12. M. Landreh, E. G. Marklund, P. Uzdavinys, M. T. Degiacomi, M. Coincon, J. Gault, K. Gupta, I. Liko, J. L. P. Benesch, D. Drew, C. V. Robinson, Integrating mass spectrometry with MD simulations reveals the role of lipids in Na+/H+ antiporters. *Nat. Commun.* **8**, 13993 (2017).

13. A. Rimon, R. Mondal, A. Friedler, E. Padan, Cardiolipin is an Optimal Phospholipid for the Assembly, Stability, and Proper Functionality of the Dimeric Form of NhaA Na+/H+ Antiporter. *Sci. Rep.* **9**, 17662 (2019).

14. R. A. Corey, W. Song, A. L. Duncan, T. B. Ansell, M. S. P. Sansom, P. J. Stansfeld, Identification and assessment of cardiolipin interactions with E. coli inner membrane proteins. *Sci. Adv.* **7**, eabh2217.

15. E. Nji, Y. Chatzikyriakidou, M. Landreh, D. Drew, An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins. *Nat. Commun.* **9**, 4253 (2018).

16. K. Gupta, J. A. C. Donlan, J. T. S. Hopper, P. Uzdavinys, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld, C. V. Robinson, The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421–424 (2017).

17. S. Schrecke, Y. Zhu, J. W. McCabe, M. Bartz, C. Packianathan, M. Zhao, M. Zhou, D. Russell, A. Laganowsky, Selective regulation of human TRAAK channels by biologically active phospholipids. *Nat. Chem. Biol.* **17**, 89–95 (2021).

18. A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin, C. V. Robinson, Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–175 (2014).

19. J. R. Bolla, R. A. Corey, C. Sahin, J. Gault, A. Hummer, J. T. S. Hopper, D. P. Lane, D. Drew, T. M. Allison, P. J. Stansfeld, C. V. Robinson, M. Landreh, A Mass-Spectrometry-Based Approach to Distinguish Annular and Specific Lipid Binding to Membrane Proteins. *Angew. Chem.* **132**, 3551–3556 (2020).

20. P. Lu, D. Min, F. DiMaio, K. Y. Wei, M. D. Vahey, S. E. Boyken, Z. Chen, J. A. Fallas, G. Ueda, W. Sheffler, V. K. Mulligan, W. Xu, J. U. Bowie, D. Baker, Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046 (2018).

21. S.-J. Hyung, C. V. Robinson, B. T. Ruotolo, Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chem. Biol.* **16**, 382–390 (2009).

22. E. Pyle, A. C. Kalli, S. Amillis, Z. Hall, A. M. Lau, A. C. Hanyaloglu, G. Diallinas, B. Byrne, A. Politis, Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chem. Biol.* **25**, 840-848.e4 (2018).

23. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

24. M. Akdel, D. E. V. Pires, E. P. Pardo, J. Jänes, A. O. Zalevsky, B. Mészáros, P. Bryant, L. L. Good, R. A. Laskowski, G. Pozzati, A. Shenoy, W. Zhu, P. Kundrotas, V. R. Serra, C. H. M. Rodrigues, A. S. Dunham, D. Burke, N. Borkakoti, S. Velankar, A. Frost, J. Basquin, K. Lindorff-Larsen, A. Bateman, A. V. Kajava, A. Valencia, S. Ovchinnikov, J. Durairaj, D. B. Ascher, J. M. Thornton, N. E. Davey, A. Stein, A. Elofsson, T. I. Croll, P. Beltrao, A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* **29**, 1056–1067 (2022).

25. J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, T. Tlusty, AlphaFold2 Can Predict Single-Mutation Effects. *Phys. Rev. Lett.* **131**, 218401 (2023).

26. H.-Y. Yen, M. L. Abramsson, M. T. Agasid, D. Lama, J. Gault, I. Liko, M. Kaldmäe, M. Saluri, A. A. Qureshi, A. Suades, D. Drew, M. T. Degiacomi, E. G. Marklund, T. M. Allison, C. V. Robinson, M. Landreh, Electrospray ionization of native membrane proteins proceeds via a charge equilibration step. *RSC Adv.* **12**, 9671–9680 (2022).

27. K. R. Vinothkumar, Structure of rhomboid protease in a lipid environment. *J. Mol. Biol.* **407**, 232–247 (2011).

28. E. Poláchová, K. Bach, E. Heuten, S. Stanchev, A. Tichá, P. Lampe, P. Majer, T. Langer, M. K. Lemberg, K. Stříšovský, Chemical Blockage of the Mitochondrial Rhomboid Protease PARL by Novel Ketoamide Inhibitors Reveals Its Role in PINK1/Parkin-Dependent Mitophagy. *J. Med. Chem.* **66**, 251–265 (2023).

29. A. Tichá, S. Stanchev, J. Škerle, J. Began, M. Ingr, K. Švehlová, L. Polovinkin, M. Růžička, L. Bednárová, R. Hadravová, E. Poláchová, P. Rampírová, J. Březinová, V. Kašička, P. Majer, K. Strisovsky, Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases. *J. Biol. Chem.* **292**, 2703–2713 (2017).

30. Y. Wang, Y. Ha, Open-cap conformation of intramembrane protease GlpG. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2098–2102 (2007).

31. S. Urban, M. S. Wolfe, Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl. Acad. Sci.* **102**, 1883–1888 (2005).

32. P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries, S. J. Marrink, Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).

33. P. C. Kroon, F. Grunewald, J. Barnoud, M. van Tilburg, P. C. T. Souza, T. A. Wassenaar, S. J. Marrink, Martinize2 and Vermouth: Unified Framework for Topology Generation. *eLife* **12** (2023).

34. Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 2.0 (2015).

35. T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, S. J. Marrink, Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

36. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).

37. W. Song, R. A. Corey, T. B. Ansell, C. K. Cassidy, M. R. Horrell, A. L. Duncan, P. J. Stansfeld, M. S. P. Sansom, PyLipID: A Python Package for Analysis of Protein–Lipid Interactions from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **18**, 1188–1201 (2022).

38. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

39. M. Abraham, A. Alekseenko, V. Basov, C. Bergh, E. Briand, A. Brown, M. Doijade, G. Fiorin, S. Fleischmann, S. Gorelov, G. Gouaillardet, A. Grey, M. E. Irrgang, F. Jalalypour, J. Jordan, C. Kutzner, J. A. Lemkul, M. Lundborg, P. Merz, V. Miletic, D. Morozov, J. Nabet, S. Pall, A. Pasquadibisceglie, M. Pellegrino, H. Santuz, R. Schulz, T. Shugaeva, A. Shvetsov, A. Villa, S. Wingbermuehle, B. Hess, E. Lindahl, GROMACS 2024-beta Manual. doi: 10.5281/zenodo.10090355 (2023).

40. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

41. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. Jr. MacKerell, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).

42. S. Páll, B. Hess, A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **184**, 2641–2650 (2013).

43. L. Konermann, Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. *J. Phys. Chem. B* **121**, 8102–8112 (2017).

44. H. Feroz, H. Kwon, J. Peng, H. Oh, B. Ferlez, C. S. Baker, J. H. Golbeck, G. C. Bazan, A. L. Zydney, M. Kumar, Improving extraction and post-purification concentration of membrane proteins. *Analyst* **143**, 1378–1386 (2018).

45. P. Larsson, R. C. Kneiszl, E. G. Marklund, MkVsites: A tool for creating GROMACS virtual sites parameters to increase performance in all-atom molecular dynamics simulations. *J. Comput. Chem.* **41**, 1564–1569 (2020).

46. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

47. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

48. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

49. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021).

50. J. Zivanov, T. Nakane, S. H. W. Scheres, A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

51. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

52. H. Hernández, C. V. Robinson, Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**, 715–726 (2007).

53. O. N. Vickery, P. J. Stansfeld, CG2AT2: an Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **17**, 6472–6482 (2021).

54. J. Huang, A. D. MacKerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).