Article

Cardiolipin Interactions with Proteins

Joan Planas-Iglesias,[1] Himal Dwarakanath,[1] Dariush Mohammadyani,[2,3] Naveena Yanamala,[3] Valerian E. Kagan,[3] and Judith Klein-Seetharaman[1,2,*]

[1]Division of Metabolic and Vascular Health, Medical School, University of Warwick, Coventry, United Kingdom; [2]Department of Bioengineering and [3]Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT Cardiolipins (CL) represent unique phospholipids of bacteria and eukaryotic mitochondria with four acyl chains and two phosphate groups that have been implicated in numerous functions from energy metabolism to apoptosis. Many proteins are known to interact with CL, and several cocrystal structures of protein-CL complexes exist. In this work, we describe the collection of the first systematic and, to the best of our knowledge, the comprehensive gold standard data set of all known CL-binding proteins. There are 62 proteins in this data set, 21 of which have nonredundant crystal structures with bound CL molecules available. Using binding patch analysis of amino acid frequencies, secondary structures and loop supersecondary structures considering phosphate and acyl chain binding regions together and separately, we gained a detailed understanding of the general structural and dynamic features involved in CL binding to proteins. Exhaustive docking of CL to all known structures of proteins experimentally shown to interact with CL demonstrated the validity of the docking approach, and provides a rich source of information for experimentalists who may wish to validate predictions.

INTRODUCTION

Cardiolipin (CL) is an unusual anionic phospholipid with two negatively charged phosphate groups in its polar head and four hydrophobic acyl chains (1,2). CL is the signature phospholipid of the inner bacterial membrane and the inner mitochondrial membranes of eukaryotic cells where its presence is important for the proper structural arrangements and functioning of a number of proteins (3,4). In addition to its structural role for the proteins of the electron transport chain, it also plays roles in different stages of apoptosis (5), adenylate translocase ion transport (6), and translocation by NDPK-D (7). Lately, extramitochondrial functions of CL in regulating the process of mitophagy (8–11) and orchestrating the NLRP3 inflammasome and maturation of IL-1β (12,13) have been documented. Not surprisingly, CL has been implicated in a number of different diseases, such as Barth syndrome, diabetes, heart disease, radiation damage, Parkinson's disease, and others. Some of these roles are membrane related, i.e., aberrant CL concentrations and distributions were shown to affect the structure, dynamics, and permeability of the mitochondrial membranes and CL is known to appear on the surface of the cell and also on the plasma membrane during apoptosis. On the other hand, interactions with proteins are crucial for understanding CL's functions and dysfunctions. Numerous proteins have been reported to interact with CL (reviewed in (14,15)), and in cases where cocrystal structures are known, individual CL binding patches (BPs) have been identified. However, when there is no structure, the location and mode of binding can be difficult to establish. For example, for cytochrome *c* at least three distinct BPs have been proposed (16–21). Here, we aimed to retrieve a comprehensive list of all known CL-protein pairs to date. The properties of the BPs were investigated in detail using molecular modeling approaches for the available cocrystal structures.

Molecular modeling has been used extensively to predict interactions between small molecules and proteins, essential for example in the drug discovery process (22). Identifying and studying the BPs of specific ligands to a group of proteins helps understanding of the molecular mechanisms governing such interactions (23). In turn, this knowledge may help in the design of proteins with specific functions (24,25), and in development of inhibitors and drugs targeting the identified binding region (26). A wide variety of approaches have been developed and employed to address the BP identification problem in silico (recently reviewed in (27)). Most detailed methods, such as the 3D QSAR relationship, require abundant previous structural knowledge on the binding regions of the ligand to its target. These methods have been successfully used, e.g., for G-protein coupled receptors (28). However, when computational resources and structural information are limited, patch analysis of the biophysical properties of the binding sites become particularly informative (29). Here, we analyzed the trends of amino acid distribution and secondary structure usage in CL-BPs, to differentiate between the regions that bind the polar headgroup and the acyl chains of the CL molecules, respectively.

Submitted March 19, 2015, and accepted for publication July 13, 2015.

*Correspondence: j.klein-seetharaman@warwick.ac.uk

Joan Planas-Iglesias and Himal Dwarakanath contributed equally to this work.

Editor: Ozlem Keskin.

0006-3495/15/09/1282/13

http://dx.doi.org/10.1016/j.bpj.2015.07.034

Additionally, taking advantage of the recently discovered overrepresentation of certain supersecondary structural motifs in protein-protein interaction interfaces (30), we searched for such motifs in CL-BPs. Finally, where experimental structural data was not available for analysis, computational predictions by docking methods were performed to provide valuable insight into the binding modes of the ligand to its target (31,32). We therefore also conducted a comprehensive analysis of the potential validity of such predictions in the case of CL-binding proteins.

MATERIALS AND METHODS

Preparing a database of known CL-binding proteins

A list of proteins experimentally demonstrated to interact with CL (the gold standard), was compiled using four different strategies: 1) extensive literature search, 2) retrieval from the UniProtKB database (33), 3) retrieval from the Protein Data Bank (PDB) (34), and 4) search for the term cardiolipin binding in the Gene Ontology browser—Quick GO (35), as follows. In addition, where available, the PDB entry was obtained from the Cross-reference section of the protein description in the UniProtKB database.

Extensive literature search

Starting with previous reviews of CL-binding proteins (14,15) we extended the search to publications in Pubmed using cardiolipin binding as the search term. Different attributes of the CL-binding proteins detailed in the identified publications (such as their name, gene name, organism name, EC number, PDB entries) were then used to search UniProtKB. The UniProt entry with the best match of the protein attributes was included in the gold standard.

Retrieval from the UniProtKB database

The keyword cardiolipin was used to search the UniProtKB database, resulting in further identification of proteins that interact with CL. Such interactions were later verified by literature review.

Retrieval from the PDB

Additional CL-protein interactions were identified by searching for structures containing CL in the PDB files (34). The chain IDs in the remarks were used to retrieve the UniProt entry of the protein.

Gene ontology cardiolipin binding term search

The GO:1901612 ID, which corresponds to the gene ontology term cardiolipin binding, was used to search the Quick GO browser. The search results revealed additional CL-binding proteins, which were recorded along with the UniProtKB references provided in the Quick GO search results. Literature search was used to verify the interactions of these proteins with CL.

Predicting 3D structures of proteins in the gold standard without known structures

For proteins without annotated PDB entries, 3D models were obtained using the online protein structure homology-modeling server—SWISS-MODEL (36,37). Models having QMEAN Z-scores >-2.0 were selected and included for obtaining structure-based features of the protein. For models having a QMEAN4 score <-2.0 and Torsion Z-score >-2.0, the Optimize plugin (38) for the PyMol software (10,39), was used to optimize the models. The optimized models were then uploaded to the QMEAN server (40,41) for reevaluation. The models with a QMEAN Z-score >-2.0 were included for obtaining structure-based features of the protein.

Homology assessment

To avoid overrepresentation of close homologs in the analysis of BPs thus ensuring no redundant information about CL-binders is included in the gold standard, we removed proteins sharing >90% sequence identity using the CD-HIT Suite (42).

Definition of BPs

CL-BPs were defined as the set of amino acids in the interacting protein chain containing at least one atom within a distance of 5 Å to any CL atom, for each protein chain/CL molecule pair. This definition was narrowed further to define three subsets of BPs to consider the specific nature of negative electric charges in the CL polar head and hydrophobic interactions involving the acyl chains. We thus defined phosphate binding patches (PBPs) as the set of amino acids in the interacting protein chain containing at least one atom within a range distance of 5 Å to any phosphorous atom in the CL molecule. Analogously, we defined hydroxyl binding patches (OBPs) as the set of amino acids in the interacting protein chain containing at least one atom within a range distance of 5 Å to the oxygen atom in the central glycerol moiety in CL molecule. Hence, a PBP or a OBP will always be a subset of the corresponding general BP given by the first definition in this subsection. Acyl binding patches (ABPs) were defined as the asymmetric difference set operation between the BP and the union of PBP and OBP for each CL-protein chain pair (i.e., an ABP is defined by the subset of residues in the BP neither belonging to the corresponding PBP nor to the corresponding OBP). All these definitions were applied to each crystallographic structure obtained from the PDB that contained 3D coordinates for both the CL molecule and its binding protein.

Sequence- and structure-based features

For proteins with known 3D structure, their amino acid composition (primary structure) was obtained directly from the corresponding PDB file. The assignment of secondary structure in those proteins, and the solvent accessibility was obtained by using DSSP 2.0.4 (43). We categorized the different annotations of DSSP as follows: H, G, and I as Helix; E and B as strand; any other annotation as Coiled coil. Whenever DSSP was unable to produce an assignment for a particular amino acid, this amino acid was excluded from the secondary structure analysis. Finally, the assignment of supersecondary structures or loops on such proteins was obtained from Markov Clustering classification terms listed in the ArchDB database (44). Relative accumulated frequencies and propensities of amino acids and secondary structures, and relative accumulated frequencies of loops were computed and represented using R (45) (see Metrics section below). Such computation was applied to five different data sets: 1) the complete protein chain, 2) BPs, 3) PBPs, 4) OBPs, and 5) ABPs. For this purpose, a loop was considered to be included in the relevant protein region for a given analysis scope if at least one amino acid of the loop belonged to such a region.

Metrics

The relative frequency of an element in a population is given by

$$F(i) = N(i)/T, \quad (1)$$

where $N(i)$ represents the total number of observations of such element and T the total number of elements in the population. In this work, we refer to

the relative accumulated frequencies of a certain amino acid as the total number of observed instances of such amino acid in a given region (i.e., whole chain or any type of binding patch) over the total number of amino acids (any one) in the same region. For secondary structures, we consider the total number of residues within a certain structural type in a given region over the total number of residues in such structural type on the same region. For loops, the relative accumulated frequency would be the number of instances of a particular loop type in a certain region over the total number of loops in that region.

Propensity of a certain amino acid or secondary structure element is defined as in (29):

$$P(i) = \frac{F(i)}{N(i)_{surf}/T_{surf}} = \frac{F(i)}{K(i)}, \quad (2)$$

where $N(i)_{surf}$ is the total count of exposed instances of such amino acid or of residues in such secondary structure element, and T_{surf} is the total number of exposed amino acids. We considered a residue exposed if its relative solvent accessibility (46) was higher than 25% (47). Within a defined set of protein crystals, every amino acid or secondary structural element will have a characteristic $N(i)_{surf}/T_{surf}$ ratio, $K(i)$. Compared to frequencies, the computation of propensities takes into account the natural inclination of a certain element (amino acid or secondary structural element) to be exposed in the protein surface. Hence, propensities will highlight, among the studied population, elements which are abundant in relation to the natural tendency of such element to be exposed. To study BPs, this measure can pinpoint key elements for the interaction (48).

Statistical significance of differences between the different measures analyzed was assessed using the Chi-squared test as implemented in R (45). The null model considered the compared elements have equal $F(i)$ (see Eq. 1) and equal $K(i)$ (see Eq. 2). P-values from these Chi-squared tests are given throughout this work. For a given comparison, a p-value $<= 0.05$ denotes statistically significant differences, whereas a p-value >0.05 denotes that the differences observed between the two analyzed elements could also be observed by chance. To stress this breakpoint, we illustrate p-values in a double colored scale, linear values from dark red (1) to light pink (0.05) to denote nonsignificant p-values and a logarithmic scale ranging from light blue (0.05) to dark blue (0) for significant p-values.

Docking

Proteins were classified into two groups for structural analysis, 1) CL cocrystal structures, i.e., proteins having a PDB entry with a CL structure present in the PDB file; 2) proteins having a PDB entry, but no CL structures included; 3) proteins only having a Swiss model entry. Docking was essential for proteins belonging to the second and third category, but it was also performed for the proteins in the first category to test the compatibility of the docking results with the CL-binding sites given in the PDB files. AutoDock Tools (49) was used to obtain PDBQT files of the receptors (proteins) and ligand (CL), and for determining the dimensions of the grid box for each protein chain, which acts as the search space for docking the ligand. The polar hydrogens were added to the protein chain molecule followed by computing Gasteiger charges for the molecule. The result was used to compute the corresponding PDBQT file. The CL species selected as ligand in all cases irrespective of the CL species bound in the cocrystal structures was 1,1′,2,2′-tetralinoleoyl-CL (TLCL), as it is an oxidizable species of CL and the most abundant CL species found in mammalian mitochondria (50). AutoDock Vina (51) was employed to run molecular docking simulations. Default parameters were used and the docking specific parameters are provided in File S5. Results were analyzed using PyMol (39) and all pymol sessions are provided in Files S3 and S5 for single chain and multiple chain dockings, respectively. The top nine different docking poses of CL were manually classified into binding sites, i.e., those modes of the docked CL that bound the receptor at the same location were grouped into a common binding site. For proteins having CL costructures in the PDB file, overlap of each predicted binding site with each crystallographic CL site was recorded. An overlap was identified whenever the phosphate heads of the docked CL were in proximity to those of the CL costructure and both the molecules shared similar orientations of the acyl chains. Proteins with no overlapping predicted site to the crystallographic one were redocked and this time the top 50 different docking poses were considered for manual classification into binding sites and posterior analysis.

Protein chains were docked individually, i.e., if other chains were present in the pdb file, they were removed for docking. Therefore, in a few isolated cases, we found some of the CL docking poses to be partially overlapping with these other chains. These instances were recorded as clashes. We also retrieved the amino acid residues of each receptor protein at 5 Å distance from each CL docking pose.

RESULTS

Identification of CL-binding proteins

We identified a total of 62 CL-binding proteins by different procedures described in the Materials and Methods section, and these proteins are listed in Table 1. Detailed information about these gold standard proteins is provided in Table S1. 26 of them were identified from extensive literature search, 12 were retrieved from the UniprotKB database, 22 were obtained from the PDB, and 2 from GO term search. Crystallographic data containing CL molecules and their protein binding partners were available for 22 of these proteins, represented in 34 different PDB protein chains. For another 21, crystallographic data was available but not in the presence of the CL binding molecule (30 PDB protein chains). 3D structure was assigned to six proteins using homology modeling. We were not able to assign 3D structure to the 13 remaining proteins. To avoid overrepresentation of near identical structures, we pruned our analysis set of proteins sharing >90% sequence identity with others using CD-hit (42). See Table S1 for details on which protein chains were removed. This procedure resulted in a set containing 21 proteins with CL cocrystals observed in 15 different PDB records.

Identification and analysis of CL-BPs

As some of the protein chains had more than one CL molecule bound, a total of 27 CL-BPs were identified in 21 protein cocrystal structures with CL. Within BPs, we differentiated between regions with preference for binding of the polar headgroup (PBPs and OBPs) of CL molecule and regions with preference for binding the acyl chains of CL (ABPs) (see Materials and Methods section). 15 of the 21 proteins were also mapped to supersecondary structures in the ArchDB database, referred hereinafter as loops (44). Loops were mapped to 14 BPs, 12 ABPs, 5 OBPs, and 9 PBPs.

TABLE 1 Proteins in gold standard

Protein Name	Mb	CL	PDB_Chain	Ref.
Alpha-MGlcDAG synthase	Y	NA	1Z2T_A	(52)
Phospholipase A2	N	NA	1POC_A	(53)
Cyt. *c* oxidase sub. 1	Y	NA	2OCC_A	(54)
Cyt. c oxidase sub. 7A1, mit.	Y	NA	2OCC_J	(54)
Cyt. *c* oxidase sub. 7C, mit.	Y	NA	2OCC_L	(54)
ADP/ATP translocase 1	**Y**	**CDL**	**1OKC_A**	(55)
Cyt. *b*-*c*1 complex sub. 1, mit.	**Y**	**CDL**	**1SQP_A**	(56)
Cyt. *c*1, heme protein, mit.	**Y**	**CDL**	1SQP_D, **2A06_D**, Q	(56), (57)
Cytochrome *b*	**Y**	**CDL**	**2A06_C**, P	(57)
Adrenodoxin, mit.	Y	NA	2BT6_A, B	(58,59)
ATP synthase sub. beta, mit.	Y	NA	2CK3_ D, E, F	(60)
Cyt. *b*-*c*1 complex subunit 8	**Y**	**CDL**	2YBB_G, **1SQP_G**	(61), (56)
Cholesterol desmolase	Y	NA	3MZS_ A,B,C,D	(58,62)
Protein S100-B	N	NA	3RM1_A	(63)
Cyt. *c* oxidase sub. 6A2, mit.	**Y**	**CDL**	3**WG7_G**, T	(64)
Cyt. *c* oxidase sub. 3	**Y**	**CDL**	**3WG7_C**, P	(64)
FDH-N sub. gamma	**Y**	**CDL**	**1KQF_C**	(65)
FDH-N sub. beta	**Y**	**CDL**	**1KQG_B**	(65)
Cyt. *c*	Y	NA	3O1Y_ A, B, C	(66)
Chr. replication initiator prot. DnaA	Y	NA	1J1V_A	(67)
Cyt. *b*-556	**Y**	**CDN**	**2ACZ_C**	(68)
DNA topoisomerase 1	N	NA	1MW9_X	(69)
Cyt. *b*	**Y**	**CDL**	**3TGU_C**, P	(70)
Mit. ubiquinol-cyt. *c* reductase ubiquinone-binding prot. qp-c	**Y**	**CDL**	**3TGU_G**	(70)
Mit. cyt. *c*1, heme protein	**Y**	**CDL**	**3TGU_Q**	(70)
Prohibitin	Y	NA	1LU7_A	(71)
Beta-2-glycoprotein 1	N	NA	3OP8_A, B	(72,73)
ABC transporter 10 prot.	**Y**	**CDL**	**3ZDQ_A**	(74)
Acyl-CoA thioesterase THEM5	Y	NA	4AE7_A	(75)
Caspase-8	Y	NA	4JJ7_A	(76)
Apoptosis regulator BAX	Y	NA	4BD2_A	(77)
Uncharacterized protein	**Y**	**CDL**	**4A2N_B**	(78)
BH3-interacting domain death agonist	Y	NA	2VOI_B	(79,80)
Antigen-presenting glycoprot. CD1d1	**Y**	**CD4**	**3MA7_A**, C	(81)
Neutral ceramidase	N	NA	2zws_A	(82)
Carnitine/acylcarnitine translocase	Y	NA	2BMN_A	(83)
Reaction center prot. M chain	**Y**	**CDL**	**4LWY_M**	(84)
Reaction center prot. M chain	Y	CDL	4N7L_M	(85)
ADP,ATP carrier prot. 2	**Y**	**CDL**	**4C9G_A**	(86)
ADP,ATP carrier prot. 3	**Y**	**CDL**	**4C9J_A**, B	(86)
Cyt. *b*	**Y**	CDL, CN3, **CN5**	1KB9_C, 3CX5_N, **3CX5_C**	(87), (88), (88)
Cyt. *c*1, heme prot. mit.	**Y**	CDL, **CN3**	1P84_D, **3CX5_D**	(89), (88)
Sapecin	N	NA	1L4V_A	(90)
Malate dehydrogenase, mit.	Y	NA	model	(59)
Thiosulfate sulfurtransferase	Y	NA	model	(91)
Tafazzin	Y	NA	model	(76)
Serine/threonine-prot. kinase N1	Y	NA	model	(92)
Creatine kinase S-type, mit.	Y	NA	model	(93)
Mit. glycerol-3-phosphate dehydrogenase	Y	NA	model	(94)
NADH dehydrogenase sub. 1	Y	NA	NA	(95)
Phosphate carrier prot mit.	Y	NA	NA	(96)
Prot. UPS1, mit.	Y	NA	NA	(97)
Prot. UPS2, mit.	Y	NA	NA	(97)
Lysocardiolipin acyltransferase 1	Y	NA	NA	(98)
Phospholipid scramblase 3	Y	NA	NA	(99,100)
Stomatin-like prot. 2, mit.	Y	NA	NA	(101)
MICOS complex subunit MIC27	Y	NA	NA	(102)
Phospholipid-transporting ATPase IC	Y	NA	NA	(103)
Glycerol-3-phosphate acyltransf. 1, mit.	Y	NA	NA	(104)
Carbamoyl-phosphate sinthase I	Y	NA	NA	(105)
Neoverrucotoxin subunit alpha	N	NA	NA	(106)
AMP deaminase 1 isoform M	N	NA	NA	(107)

(*Continued on next page*)

Table 1. Continued

The column Protein name shows the UniProtKB recommended name of the protein. Some records use alternate and short names as provided in UniProtKB for convenience. Recurring words are abbreviated: mit. for mitochondrial, cyt. for cytochrome, sub. for subunit, chr. for chromosomal, prot. for protein, and transf. for transferase. The column Mb indicates if protein is a membrane or nonmembrane protein (Y and N, respectively). CL shows the structure ID(s) of the cardiolipin costructure(s) given in the PDB file of the protein. PDB_chain indicates the chain of the protein's PDB structure known to interact with cardiolipin. Records of proteins included in the nonredundant cocrystals subset are shown in bold face. Note that the reaction center protein M chain in PDB 4NL7 was excluded from this subset because it is 100% identical to reaction center protein M chain in PDB 4LWY, although both proteins are from different organisms.

We have used the set of CL-binding protein cocrystal structures to analyze the characteristics of CL-BPs in proteins in terms of their amino acid composition (Figs. 1 and S1), secondary structure (Figs. 2 and S2), and supersecondary structure/loops (Fig. 3). Finally, we compared the values of these three features in the CL-BPs, ABPs, OBPs, and PBPs to the values obtained from the complete CL-interacting proteins.

Electrostatic-driven amino acid bias trends in CL-binding regions

The amino acid composition shown in Fig. 1 indicates that there is an overrepresentation of positively charged residues in PBPs and OBPs, primarily Arg and Lys residues but also His, when compared to ABPs. This suggests an important role of electrostatic interactions in attracting CL molecules to their protein binders, as has been proposed for specific proteins previously (108–110). The fact that negatively charged residues (Asp and Glu) are underrepresented in BPs further supports this hypothesis. Strikingly, although Asp is only underrepresented in PBPs, demonstrating its incompatibility to interact with the negative polar head of CL, Glu is completely disallowed in the whole CL-binding region (Figs. 1 and S1), suggesting there exist further restrictions for the usage of this amino acid in CL-BPs.

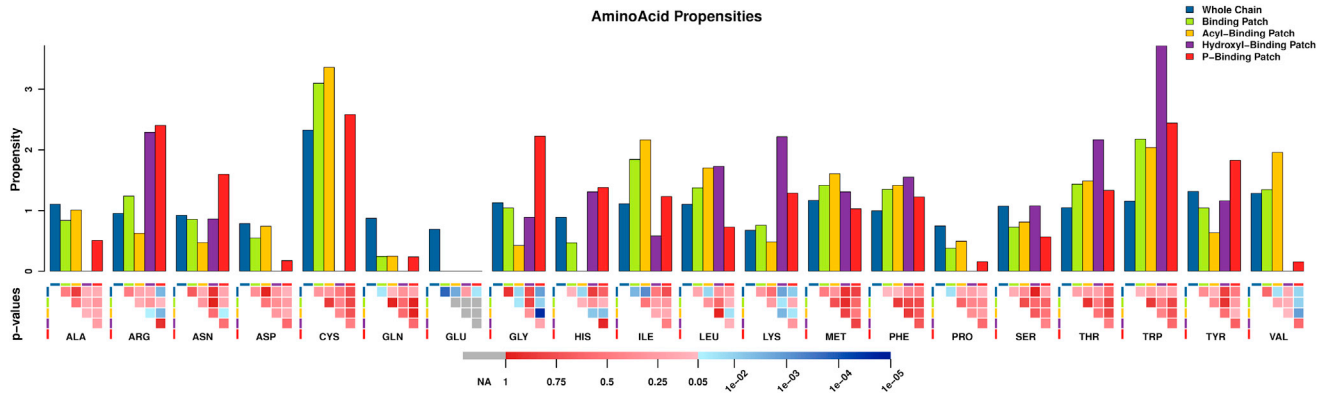

FIGURE 1 Propensities of observed amino acids in all protein chains analyzed (*blue*), BPs (*green*), ABPs (*dark yellow*), OBPs (*purple*), and PBPs (*red*). The *p*-values for the statistical significance of differences between propensities of each amino acid in different BPs are shown in the lower panel. NA (*gray*) indicates that two empty sets were compared and hence no statistics can be computed. In red and using a linear scale, nonsignificant *p*-values ranging from 1 (*darker red*, indicates that greater differences could always be observed in a random set) to 0.05 (*lighter red*) are shown. In blue and using a logarithmic scale, significant *p*-values are shown, from 0.05 (*light blue*, differences observed could be reproduced by random only in 5 over 100 cases) to *p*-values near 0 (*dark blue*, the probability to observe such differences by chance is the lowest).

Hydrophobic interactions with the acyl chains

As expected, hydrophobic amino acids dominate the ABPs (Figs. 1 and S1). In particular, Leu, Ile, and Val are most frequent. The accumulated frequencies of Leu, Ile, and Val alone represent 24% of the amino acid content of the studied proteins, and sum up to 15% in the PBPs. In comparison, these hydrophobic amino acids are overrepresented in BPs and ABPs, representing 31% and 40% of their amino acid content, respectively. Although these three amino acids are the most strikingly overrepresented ones, Phe, Thr, Met, and Trp are also more frequent in BPs when compared to all chains. Grouping amino acids by hydrophobicity alone is of course a simplification, and this is most clearly visible when looking at Tyr. Although a hydrophobic amino acid, it is rarely found in ABPs, either due to the hydrophilic OH group or its role in regulation, i.e., the potential to be phosphorylated, which would prevent CL binding. Significance values of frequencies and propensities comparisons between different amino acids are provided in File S1.

Flexibility trends in CL-binding regions

A remarkable bias was observed in the distribution of amino acids related to protein flexibility (Fig. 1): Gly, which imparts flexibility is highly overrepresented and Pro, which generally introduces rigidities in the 3D structure of proteins is highly underrepresented in PBSs. This observation suggests a requirement for higher flexibility in the region binding the polar headgroups of CL. The opposite is observed for the ABPs. Here, Gly is underrepresented, although Pro is also but less strongly underrepresented when compared to all chains. This indicates that structured regions favor the binding of the acyl chains.

Secondary structure analysis

The analysis of secondary structure compositions is shown in Figs. 2 and S2. Overall, helices are most frequent, but this is expected as the majority of known CL-binding proteins with structures are integral membrane proteins of the helical bundle type (Table 1). We therefore concentrated on the differences in the preference for secondary structures in BPs, PBPs, and ABPs as compared to the full chains. Although ABPs are mainly composed of helical secondary structures and only contain ~10% unorganized regions, the percentage of residues with coiled-coil conformation remarkably rises to almost 40% in PBPs. Although this value is similar to that of the whole amino acid chain, the important fact is its increased percentage of coiled-coil conformation in BPs and especially PBPs when compared to ABPs (Fig. S2). In fact, propensities analysis reveal a reduced preference of ABPs for coiled-coil conformation (~0.25) in comparison the PBPs, which display a normal trend for this secondary structure element (propensity ~1) (Fig. 2). Although the intrinsic rigidity of Pro induces the disruption of regular secondary structures (111), the increased coiled-coil organization in PBPs cannot be attributed to this effect because the abundance of Pro is clearly diminished in such regions (Fig. 1). Hence, the increased preference for coiled-coil conformation in PBPs depicts the larger requirements for flexibility in the binding of the polar CL heads, further supporting the flexibility trends observed in the amino acid compositions (see above). In line with this conclusion is also the complete absence of strand secondary structure conformations in PBPs. In contrast to the PBPs, the ABPs prefer helices, and there is a clear trend even beyond the overall abundance of helices in our gold standard data set. Strand is also an acceptable structural element, although the percentage is no higher than observed in all chains. In line with this observation,

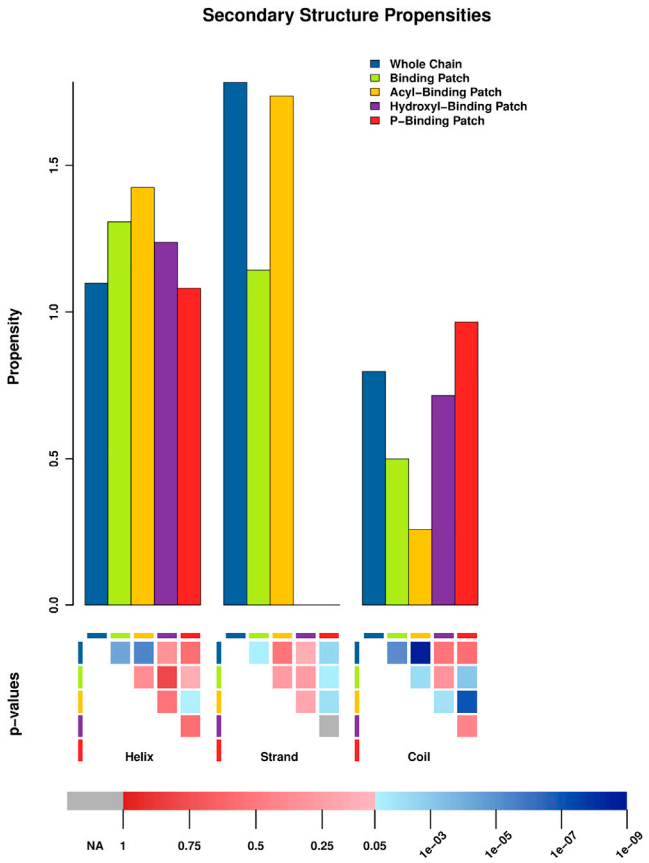

FIGURE 2 Relative propensity of observed secondary structures in all protein chains (*blue*), BPs (*green*), ABPs (*dark yellow*), OBPs (*purple*), and PBPs (*red*). The *p*-values for the statistical significance of differences between propensities of each amino acid in different BPs are shown in the lower panel, using the same color legend described in Fig 1.

coiled coil is strongly disfavored in ABPs. These observations also support the conclusions from amino acid preferences, that the acyl chains favor rigid structured elements to anchor on.

Supersecondary structure motifs in CL-binding regions

Recent studies show that loops, defined as supersecondary structural motifs formed by a coiled region flanked by two correlative regular secondary motifs (112), to be useful features in the prediction of protein-protein interactions (30). A database of such supersecondary structural motifs has been built (44). With the aim of identifying putative specific structural footprints of CL-binding sites, we extended our analysis of CL-binding regions to the loops in this database. 94 different types of loops were observed in regions of CL-binding proteins, 23 of which were located in different types of BPs (Tables S2 and S3). 20 of these 23 were represented in ABPs and only 12 of them were in PBPs. Moreover, three of the loop types are exclusive to PBPs, 11 are exclusive to ABPs, and 9 loop types are shared among PBPs and ABPs. The frequencies of these loops in the CL-binding proteins are shown in Fig. 3 *A*, and cartoon images of these loop structures are depicted in Fig. 3 *B*. Most frequent are loops GH_2.19.1 and HH_1.4.16. Most loops were observed more frequently in PBPs than in ABPs, with the exception of HH_1.4.16. PBP-exclusive loops are: 1) HG.0.1.1 for which two instances found in PBPs (1OKC_A residues 26–40 and 4N7L_M residues 82–91), and 2) EH.2.25.1 and EH.4M.6.1, which represent the same supersecondary structural motif and was observed once in PBPs (1SQP_A residues 317–348). All these loops represent hydrophobic residue-rich helices in their original structures and contain positively charged residues near the flexible region of the loop. Only the N-terminal β-strand of the EH motif in 1SQP_A is not in close proximity to the polar head of CL.

Of the nine loop types shared among PBPs and ABPs (Fig. 3 *A*), GH.2.19.1 shows the most remarkable preference for PBPs. Four instances of this loop are present in three of the proteins included in our analysis, and all of them are also present in both PBPs and ABPs. This fact suggests a possible relationship between this supersecondary structure and the CL-binding function. Due to the lesser number of loops mapped in BPs than in proteins (23 and 94, respectively), and in PBPs than in ABPs (12 and 20, respectively), the relative frequency of this loop is progressively higher in BPs, ABPs, and PBPs. Remarkably, all structures clustered in this loop type contain a Gly residue in the second position of its flexible region (44), thus making it an extremely conserved residue. This Gly is preceded by a positively charged residue (Arg or Lys) in CL-binding proteins. A second positively charged residue is located near this position, allowing for the stabilization of the double negatively charged CL polar head. Fig. 4 illustrates these details in the example of the ATP/ADP carrier protein (PDB ID 1OKC), which has two different CL-binding sites. Strikingly, in both of the binding sites a GH.2.19.1 loop is present, showing that this particular protein adopted the same structural solution for binding two CL molecules. Thus, the presence of a GH.2.19.1 loop with a positively charged residue preceding a crucial flexibility conferring Gly in the phosphate binding region and a helix providing an anchor for acyl chain binding, may be considered a footprint for CL binding.

Validity of predicted CL-binding sites

Approximately one-third of the proteins (22 of 62, 21 after removing sequence redundancy) in the gold standard data set have known CL bound cocrystal structures, and a further 21 have structures without CL (Table 1). This represents a potentially rich source of structural features for analysis and prediction of CL-binding sites. CL-protein docking predictions have been used successfully in the structural analysis of CL interactions previously (9,113–116) and

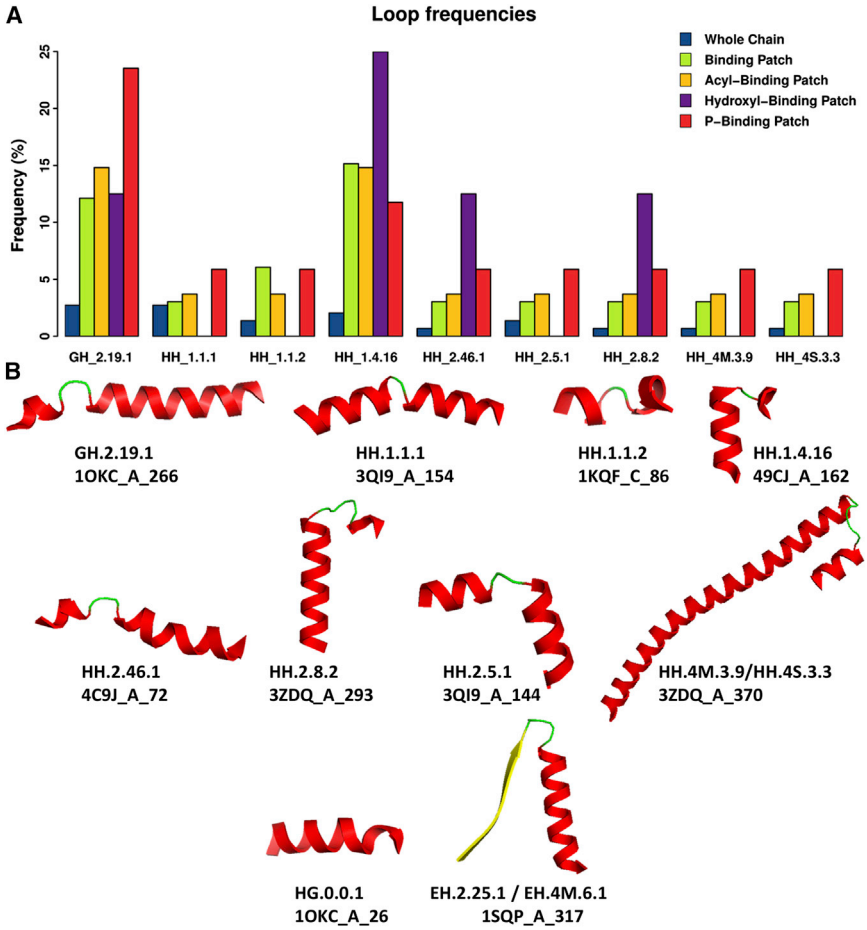

FIGURE 3 (*A*) Relative frequencies of observed supersecondary motifs (*loops*) in all protein chains (*blue*), BPs (*green*), ABPs (*dark yellow*), OBPs (*purple*), and PBPs (*red*). (*B*) Cartoon images of the loops present in PBPs. Different secondary structure elements are shown in color (*helices in red*, *strands in yellow*, and *coiled-coil regions in green*). Each motif is identified in the first text line below by the code of its corresponding loop type(s). The PDB file from where each loop structure shown was obtained, protein chain and starting residue of the supersecondary structure are provided in the second line below each cartoon.

have been useful in shedding light on the mechanisms governing such interactions. We therefore docked TLCL, the most abundant CL species in mitochondria, to all structures after removing all bound ligands in them, and analyzed the results (see Materials and Methods section). Of the top-ranked poses predicted by Autodock Vina, binding sites were manually clustered together and highlighted in File S2. In most cases, we predicted more than one binding site. However, checking for overlap with the known CL-binding sites in the 21 cocrystal structures (Table 2) indicated that the prediction of binding sites by docking is very effective. For 18 of the 21 analyzed proteins, the predicted CL docking poses overlapped with the crystallographic CL molecule annotated in the proteins' structure. Fig. 5 illustrates this overlap for the ATP/ADP carrier protein in its monomeric and dimeric forms. Clashes (docking poses where CL molecule overlap with known protein-protein interaction interfaces, see Materials and Methods) were observed in 10 of the 18 binding sites, but seven of them had at least one overlapping docking pose that was free of clashes. The other 8 of 18 were completely free of clashes.

Only TLCL was docked in this study, although there were five different crystallographic identities of CL species (CDL, CN3, CN5, CDN, and CD4). The fact that docking was still able to identify the known CL-binding site despite the different CL species used in docking provides important evidence that docking with AutoDock Vina (51) is a highly reliable method for identification of general CL-binding sites in proteins. All of the poses obtained in the docking are available in the Supporting Material as PyMol (39) sessions (File S3), not only for the CL-cocrystal structures but also structures lacking CL. This provides a rich source of predictions that experimentalists can use to verify CL-binding sites.

Comparison of studied trends in crystallographic CL-protein complexes to docking and negative data sets

Taking advantage of all docked poses, we defined four new sets of CL-binding protein costructures. The first one contains all docking results in Table 2. The second and third ones respectively contain the particular docking poses overlapping (positive docking control) and nonoverlapping (negative control) with the crystallographic structure as defined in Table 2 and Supporting Material File S2. The fourth data set contains the results of docking TLCL to 30

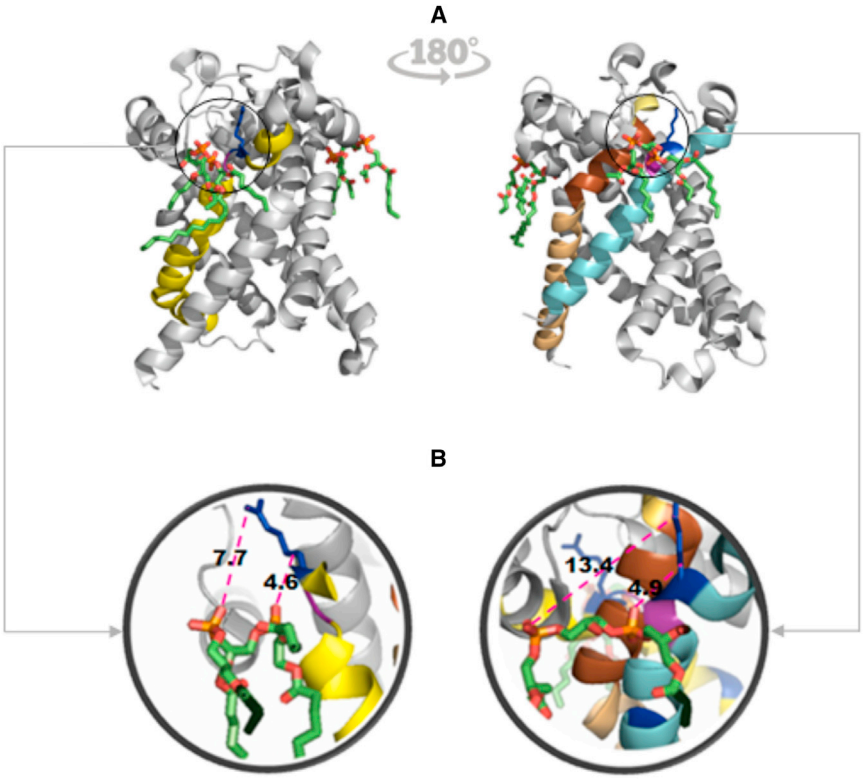

FIGURE 4 Loops of ADP/ATP translocase 1 (PDB 1OKC chain A) interacting with two cardiolipin molecules. (*A*) One CL molecule (*left*) interacts with loop GH_2.19.1 (1OKC_A_66; residues 66–99; colored in *yellow*).The second CL molecule (*right*) interacts with a different instance of loop GH_2.19.1 (1OKC_A_266; residues 266–291; colored in *cyan*), showing that both binding sites adopt similar structural conformation. The second molecule also interacts with loop HH_1.1.2 (1OKC_A_4; residues 4–37; colored in *light orange*) and loop HG_0.1.1 (1OKC_A_26; residues 26–40; colored in *pale yellow*). Overlapping region between HH_1.1.2 and HG_0.1.1 (residues 26–37) is shown in *brown*). Positively charged amino acids Arg and Lys in loop GH_2.19.1 are shown in blue, and the conserved Gly in the loop motif is colored in purple. (*B*) Enlarged images of areas circled in (*A*) showing interaction between positively charged residues and the phosphate groups of CL in the two different binding sites, where CL binding is accomplished by different geometries. The geometry of binding is described by measured distances between two proximal oxygen atoms in the different CL phosphate groups to carbon beta and distal nitrogen in relevant Arg (4.6 Å and 7.7 Å, respectively, *left*) or Lys (4.9 Å and 13.4 Å, respectively, *right*).

nonredundant proteins for which no cocrystal was available. Results of this comparison are provided in Figs. S3–S10 (amino acid and secondary structure elements propensity values), and in File S4 (statistical significance for differences between trends in crystallographic complexes and trends in the different docking data sets defined previously) in the Supporting Material. Remarkably, docking results by Autodock Vina of overlapping binding sites (docking positives) reproduce the electrostatic and hydrophobic features described for crystallographic complexes, but not the flexibility trends. For instance, docking positives as well as the overall docking results show an enrichment of positively charged residues in PBPs and OBPs, whereas in ABPs hydrophobic residues are predominant. If secondary structural elements are considered, the ABPs preference for strand conformation (compared to PBPs and OBPs) is then confirmed. However, not even the controlled positives from docking experiments are able to reproduce the unique trend for Gly in PBPs or the preference for coiled-coil conformation in PBPs and OBPs seen in the experimentally observed crystal structure binding sites.

DISCUSSION

In this work we made three contributions to the field of CL binding to proteins. 1) We collected a gold standard data set of all known CL binding proteins. This is the only systematic and to the best of our knowledge comprehensive list available to date. 2) Our BP analysis of the compiled set of cocrystal structures not only confirmed for some CL binding proteins the experimentally validated expectation that hydrophobicity and charge play a role in binding, but we also revealed a new, to our knowledge, and unexpected feature important for binding: flexibility. Thus, a typical binding patch is composed of positively charged or nonpolar residues in PBPs and ABPs, respectively, with a remarkable overrepresentation of Arg and Lys in PBPs and of Leu, Ile, and Val in ABPs. In contrast, negatively charged residues (Asp and Glu) are either underrepresented or disallowed in PBPs. Furthermore, the region binding the polar headgroups of the CL molecule shows a differential trend toward increased flexibility, with a higher content of Gly and coiled-coil secondary structure, and a lower content of Pro. The only and remarkable case in which crystallographic data did not comply with these trends was that of the hydrophobic residue Tyr, overrepresented in PBPs. Tyr is a widely used residue in signaling due to its ability of being phosphorylated. Such phosphorylation may prevent CL binding and could therefore be disfavored here. In addition, Tyr usually participates in hydrogen bonding and it may have other roles besides binding. For example, it may participate in structural arrangements essential for catalytic functions, similar to Tyr-67 in cytochrome *c* (117). Finally, in the investigation of loop supersecondary structural motifs (44) in BPs, loop GH.2.19.1., characteristic for the mitochondrial ATP/ADP carrier, appeared particularly prominently. It is

TABLE 2 Overlap analysis for PDBs with CL costructure(s)

UniProt	PDB	C.	CL ID	O.	Cl.	#DP	#OP	#CP	#NCP	#BS	#OS	#NCS	Homo-log Struct-ure	BE (kcal/mol)	RMSD
P0AEK7	1KQF	C	CDL	N	NA	50	NA	NA	NA	5	NA	NA	NA	NA	NA
P0AAJ3	1KQG	B	CDL	Y	Y	9	9	8	1	1	1	1	NA	−4.3	10.9
P02722	1OKC	A	CDL	Y	N	9	3	0	3	4	2	2	NA	−4.0	11.6
P31800	1SQP	A	CDL	N	NA	50	NA	NA	NA	5	NA	NA	NA	NA	NA
P13271	1SQP	G	CDL	Y	N	9	1	0	1	3	1	1	NA	−3.0	14.3
P00157	2A06	C	CDL	Y	Y	50	5	2	3	5	1	1	2A06_P	−5.5	13.3
P00125	2A06	D	CDL	Y	Y	9	4	4	0	3	1	0	NA	−3.5	12.8
P69054	2ACZ	C	CDN	Y	N	50	1	0	1	8	1	1	NA	−5.2	11.8
P00163	3CX5	C	CN5	Y	Y	50	4	1	3	8	1	1	NA	−6.8	17.5
P07143	3CX5	D	CN3	Y	Y	9	9	9	0	1	1	0	NA	−4.5	15.9
P11609	3MA7	A	CD4	Y	N	9	7	0	7	2	1	1	NA	−7.0	12.6
P18946	3TGU	C	CDL	Y	Y	50	10	1	9	8	1	1	3TGU_P	−4.8	12.7
D0VX32	3TGU	G	CDL	Y	Y	50	5	4	1	5	1	1	NA	−2.9	14.7
D0VX26	3TGU	Q	CDL	Y	Y	50	1	1	0	10	1	0	NA	−5.0	12.0
P00415	3WG7	C	CDL	Y	Y	50	4	1	3	6	1	1	NA	−5.8	12.5
P07471	3WG7	G	CDL	N	NA	50	NA	NA	NA	6	NA	NA	NA	NA	NA
Q9NRK6	3ZDQ	A	CDL	Y	N	9	3	0	3	2	1	1	NA	−5.8	13.5
Q8TMG0	4A2N	B	CDL	Y	N	9	2	0	2	3	1	1	NA	−4.6	13.4
P18239	4C9G	A	CDL	Y	N	50	3	0	3	8	1	1	NA	−3.5	12.4
P18238	4C9J	A	CDL	Y	N	50	2	0	2	8	2	2	NA	−3.4	14.5
P0C0Y9	4LWY	M	CDL	Y	Y	50	4	1	3	8	1	1	NA	−5.8	13.6

The column UniProt shows the UniProt ID of the protein. PDB indicates the PDB structure ID. C. lists the chain of the PDB structure, which corresponds to the UniProt ID. CL ID indicates the structure ID of the CL molecule that binds with the protein. O. shows whether the crystallographic CL molecule overlaps with any of the docked CLs or not (Y and N, respectively); Cl. indicates whether any clash has occurred between the overlapping docking poses and other chains of the PDB structure or not (Y and N, respectively); #DP gives the number of docking poses of the docked CL; #OP indicates the number of docking poses of the docked CL overlapping with the CL crystallographic costructure; #CP shows the number of overlapping modes that clash with other chains of the PDB structure; #NCP Number of overlapping modes which do not clash with any chain of the PDB structure; #BS shows the number of different manually clustered binding sites of the docked CL; #OS indicates the number of binding sites including at least one pose overlapping with the CL crystallographic costructure; #NCS shows the number of binding sites including at least one pose which do not clash with any chain of the PDB structure; Homolog structure indicates an alternate structure file for the same protein (UniProt ID) in which an overlap between crystallographic CL and any docked one was observed in cases where such overlap did not occur for the PDB and chain shown in columns 2 and 3. Crystal structures referred to in columns 2 and 3 were the ones included in the nonredundant cocrystals subset described in main text. BE kcal/mol shows the lowest binding energy of the docked CL's overlapping docking pose. RMSD shows the root mean-square deviation between the lowest energy overlapping docked pose and the crystallographic pose of CL. NA here refers to no overlap observed. Proteins in which no overlap between the crystallographic CL molecule and any docked one was observed show NA in Cl., #DP, #OP, #CP, #NCP, #BS, #OS, #NCS, and BE kcal/mol columns.

tempting to speculate that this loop containing a positively charged residue preceding the crucial flexibility conferring Gly in the PBP region, and a helix providing an anchor for acyl chain binding, may be considered a footprint for CL binding. 3) We carried out an exhaustive docking of TLCL to all known structures of proteins experimentally shown to interact with CL. This not only demonstrates the validity of the docking approach, but also provided new, to our knowledge, insights for designing further experiments.

CL binding has initially been thought of as being dominated by electrostatic forces (17,19,118), and our analysis supports this notion. However, numerous studies have proposed additional hydrophobic forces to contribute to binding (19,21,66), a hypothesis also supported by our findings. A popular model combining these two features is a two-step process where electrostatic attraction initiates the interaction of CL with protein, and hydrophobic contacts further consolidate the binding to be of high affinity (20,66). Recently, the role of electrostatic attraction has been challenged (86) because the crystal structures of the yeast mitochondrial ATP/ADP carrier (PDBs 4C9G, 4C9J) show that the lateral chains of the positively charged residues preceding the central conserved Gly are oriented in a plane perpendicular to that defined by the phosphate groups in the CL molecule, and pointing in the direction opposite to the phosphates. This led to the speculation that instead of direct electrostatic interaction, CL interacts with the positively charged end of the helix dipoles, which correspond to the C-terminal secondary structure in loops GH.2.19.1 shown in Fig. 4. Further insight can be gained from analysis of the cattle mitochondrial ATP/ADP carrier (PDB 1OKC) (55), which contains two different CL-BPs. In one of them (Fig. 4, *right*), the lateral chain of the preceding Lys is oriented analogous to the yeast mitochondrial ATP/ADP carrier, whereas the other binding site shows the lateral chain of the Arg preceding Gly to be parallel to the plane defined by the phosphate groups in the CL molecule (Fig. 4, *left*). Our finding that flexibility appears to be an important feature for binding of the phosphate groups, would suggest that a wide spectrum of geometries of amino acid side chains can be involved in CL binding to the same

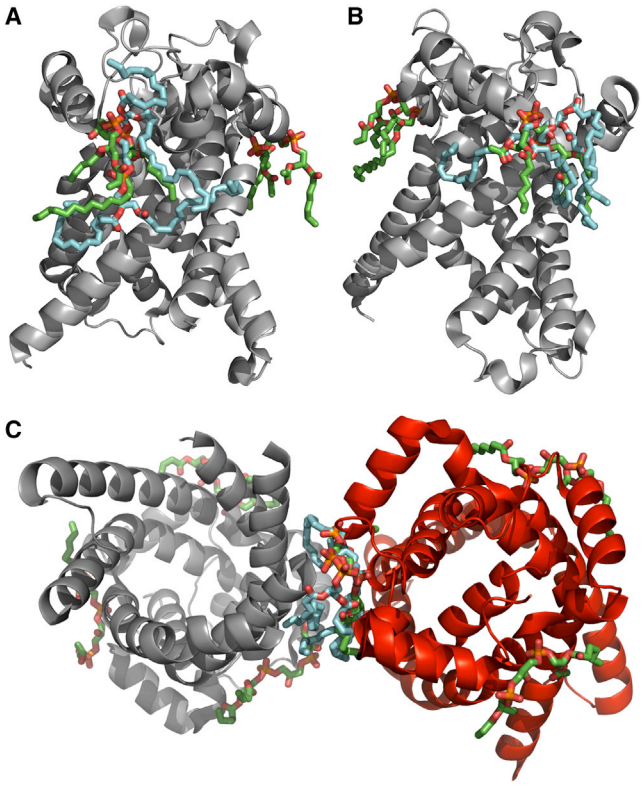

FIGURE 5 Overlap of docked CL molecule with CL costructures of 1OKC and 4C9J. The docked CL is colored in cyan and the CL costructures are colored in green. (*A*) Overlap with CDL 800 molecule of 1OKC; (*B*) Overlap with CDL 801 molecule of 1OKC; (*C*) Overlap with CDL 802 molecule of 4C9J chain A (*red*).

supersecondary motif. The high flexibility in these areas would make it likely that the Lys side chains are not fixed in space (either parallel or perpendicular) but are likely to be dynamic and able to adopt multiple conformations. This view is compatible with both models, direct electrostatic interactions of the phosphate groups with positively charged amino acids or with the electric dipole created by the helix. It is remarkable that loop GH.2.19.1, which we propose here as a footprint of CL binding, is compatible with all features, the previously proposed positive charge and/or the electric helix dipole and the hydrophobic anchor. In addition, we propose a direct role of flexibility in facilitating binding.

An important issue to address in the future is the role of multiple chains in binding CL. The majority of CL cocrystal structures contain multiple chains (14 of 18 pdb files) and even though the HETATM field in the pdb code identifies each CL molecule with a specific chain, there are contributions of neighboring chains to the binding pockets. This can yield further insight into binding but can also help improve the docking. Docking of the large protein complexes becomes highly challenging and multiple grid boxes are required to cover the docking space. For illustration purposes, we selected three pdb files containing multiple chains (File S6) and compared docking of the single chains with docking of all chains (File S7). The comparison indicates that highly complementary information can be obtained by the two approaches, requiring further in-depth investigation for the contribution of interfaces to CL-binding pockets in the future.

Although we have observed clear trends in BP properties in CL-binding proteins, we have not yet explored specificity. For example, other anionic phospholipids or even other doubly negatively charged small molecules such as GDP or ADP may show similar general structural and dynamics features in their respective binding pockets. Hence, it is premature to regard the trends described here as specific to CL binding. Conversely, we cannot ascertain if all CL binding modes will be the same. In particular, there may be different modes depending on the role of the CL interaction with proteins. For example, in respirasomes, CL acts as a staple between the subunits. In contrast, in cytochrome *c*, CL appears to remain membrane bound while interacting with the protein. In some proteins, CL binding has no demonstrated effect on function. Thus, different functional roles of CL binding may be accomplished by yet unidentified modes of binding and these may be specific to a given role.

At least for some types of CL-binding proteins—those represented by the current data set—our studies open the door for developing a generic algorithm for qualitatively predicting protein-CL interactions and for stratifying computed binding sites from docking. By comparing a predicted binding site for the presence of specific amino acid types, secondary structures and supersecondary structures, it might be possible to accurately select the actual binding pose of CLs from several high-ranking docking poses. Furthermore, the availability of a gold standard data set can be used in combination with suitable features to develop a general classification model that can be applied at the system-wide level to predict if a given protein may interact with CL or not. Currently, work is underway to develop such a classifier.

SUPPORTING MATERIAL

Ten figures, three tables, and seven supporting files are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00769-9.

AUTHOR CONTRIBUTIONS

J.K.S. initiated and oversaw the study. J.P.I. and J.K.S. designed the research and interpreted the results. J.P.I., H.D., and D.M. generated and analyzed the data presented. N.K. and V.E.K. contributed critical discussion. J.P.I., H.D., and J.K.S. wrote the article and all authors participated in editing. All authors read and accepted this manuscript.

ACKNOWLEDGMENTS

The authors thank the Human Frontier Science Program and National Institutes of Health (grant NIH U19AIO68021) for funding this study.

REFERENCES

1. Daum, G. 1985. Lipids of mitochondria. *Biochim. Biophys. Acta.* 822:1–42.
2. Schlame, M., S. Brody, and K. Y. Hostetler. 1993. Mitochondrial cardiolipin in diverse eukaryotes. Comparison of biosynthetic reactions and molecular acyl species. *Eur. J. Biochem.* 212:727–735.
3. Belikova, N. A., Y. Y. Tyurina, …, V. E. Kagan. 2009. Heterolytic reduction of fatty acid hydroperoxides by cytochrome *c*/cardiolipin complexes: antioxidant function in mitochondria. *J. Am. Chem. Soc.* 131:11288–11289.
4. Xu, Y., R. I. Kelley, …, M. Schlame. 2003. Remodeling of cardiolipin by phospholipid transacylation. *J. Biol. Chem.* 278:51380–51385.
5. Houtkooper, R. H., and F. M. Vaz. 2008. Cardiolipin, the heart of mitochondrial metabolism. *Cell. Mol. Life Sci.* 65:2493–2506.
6. Rück, A., M. Dolder, …, D. Brdiczka. 1998. Reconstituted adenine nucleotide translocase forms a channel for small molecules comparable to the mitochondrial permeability transition pore. *FEBS Lett.* 426:97–101.
7. Francois-Moutal, L., O. Marcillat, and T. Granjon. 2014. Structural comparison of highly similar nucleoside-diphosphate kinases: molecular explanation of distinct membrane-binding behavior. *Biochimie.* 105:110–118.
8. Chu, C. T., H. Bayır, and V. E. Kagan. 2014. LC3 binds externalized cardiolipin on injured mitochondria to signal mitophagy in neurons: implications for Parkinson disease. *Autophagy.* 10:376–378.
9. Chu, C. T., J. Ji, …, V. E. Kagan. 2013. Cardiolipin externalization to the outer mitochondrial membrane acts as an elimination signal for mitophagy in neuronal cells. *Nat. Cell Biol.* 15:1197–1205.
10. Kagan, V. E., C. T. Chu, …, H. Bayir. 2014. Cardiolipin asymmetry, oxidation and signaling. *Chem. Phys. Lipids.* 179:64–69.
11. Li, X. X., B. Tsoi, Y. F. Li, H. Kurihara, and R. R. He. 2015. Cardiolipin and its different properties in mitophagy and apoptosis. *J. Histochem. Cytochem.* 63:301–311.
12. Iyer, S. S., Q. He, …, F. S. Sutterwala. 2013. Mitochondrial cardiolipin is required for Nlrp3 inflammasome activation. *Immunity.* 39: 311–323.
13. O'Neill, L. A. 2013. Cardiolipin and the Nlrp3 inflammasome. *Cell Metab.* 18:610–612.
14. Schlame, M., and M. Ren. 2009. The role of cardiolipin in the structural organization of mitochondrial membranes. *Biochim. Biophys. Acta.* 1788:2080–2083.
15. Schlame, M., D. Rua, and M. L. Greenberg. 2000. The biosynthesis and functional role of cardiolipin. *Prog. Lipid Res.* 39:257–288.
16. Kawai, C., F. M. Prado, …, I. L. Nantes. 2005. pH-Dependent interaction of cytochrome *c* with mitochondrial mimetic membranes: the role of an array of positively charged amino acids. *J. Biol. Chem.* 280:34709–34717.
17. Kostrzewa, A., T. Páli, …, D. Marsh. 2000. Membrane location of spin-labeled cytochrome *c* determined by paramagnetic relaxation agents. *Biochemistry.* 39:6066–6074.
18. Rytömaa, M., and P. K. Kinnunen. 1994. Evidence for two distinct acidic phospholipid-binding sites in cytochrome *c*. *J. Biol. Chem.* 269:1770–1774.
19. Rytömaa, M., and P. K. Kinnunen. 1995. Reversibility of the binding of cytochrome *c* to liposomes. Implications for lipid-protein interactions. *J. Biol. Chem.* 270:3197–3202.
20. Sinibaldi, F., B. D. Howes, …, R. Santucci. 2010. Extended cardiolipin anchorage to cytochrome *c*: a model for protein-mitochondrial membrane binding. *J. Biol. Inor. Chem.* 15:689–700.
21. Tuominen, E. K., C. J. Wallace, and P. K. Kinnunen. 2002. Phospholipid-cytochrome *c* interaction: evidence for the extended lipid anchorage. *J. Biol. Chem.* 277:8822–8826.
22. Jacob, L., and J. P. Vert. 2008. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics.* 24:2149–2156.
23. Sael, L., and D. Kihara. 2012. Detecting local ligand-binding site similarity in nonhomologous proteins by surface patch comparison. *Proteins.* 80:1177–1195.
24. Looger, L. L., M. A. Dwyer, …, H. W. Hellinga. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature.* 423:185–190.
25. Bonet, J., J. Segura, …, N. Fernandez-Fuentes. 2014. Frag'r'Us: knowledge-based sampling of protein backbone conformations for de novo structure-based protein design. *Bioinformatics.* 30:1935–1936.
26. Joseph-McCarthy, D., A. J. Campbell, …, D. Moustakas. 2014. Fragment-based lead discovery and design. *J. Chem. Inf. Model.* 54:693–704.
27. Konc, J., and D. Janežič. 2014. Binding site comparison for function prediction and pharmaceutical discovery. *Curr. Opin. Struct. Biol.* 25:34–39.
28. Rodríguez, D., and H. Gutiérrez-de-Terán. 2013. Computational approaches for ligand discovery and design in class-A G protein-coupled receptors. *Curr. Pharm. Des.* 19:2216–2236.
29. Jones, S., and J. M. Thornton. 1997. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.* 272:133–143.
30. Planas-Iglesias, J., J. Bonet, …, B. Oliva. 2013. Understanding protein-protein interactions using local structural features. *J. Mol. Biol.* 425:1210–1224.
31. Sousa, S. F., A. J. Ribeiro, …, M. J. Ramos. 2013. Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Curr. Med. Chem.* 20:2296–2314.
32. Brewerton, S. C. 2008. The use of protein-ligand interaction fingerprints in docking. *Curr. Opin. Drug Discov. Devel.* 11:356–364.
33. Consortium, T. U. 2014. Activities at the Universal Protein Resource (UniProt). *Nucl. Acids Res.* 42:D191–D198.
34. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
35. Binns, D., E. Dimmer, …, R. Apweiler. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics.* 25:3045–3046.
36. Biasini, M., S. Bienert, …, T. Schwede. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42:W252–W258.
37. Bordoli, L., F. Kiefer, …, T. Schwede. 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4:1–13.
38. O'Boyle, N. M., M. Banck, …, G. R. Hutchison. 2011. Open Babel: An open chemical toolbox. *J. Cheminform.* 3:33.
39. The PyMOL Molecular Graphics System, Version 1.5.0.4. Schrödinger, LLC, New York, NY.
40. Benkert, P., M. Biasini, and T. Schwede. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 27:343–350.
41. Benkert, P., M. Künzli, and T. Schwede. 2009. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37:W510–W514.
42. Huang, Y., B. Niu, …, W. Li. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 26:680–682.
43. Joosten, R. P., T. A. te Beek, …, G. Vriend. 2011. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* 39:D411–D419.
44. Bonet, J., J. Planas-Iglesias, …, B. Oliva. 2014. ArchDB 2014: structural classification of loops in proteins. *Nucleic Acids Res.* 42:D315–D319.
45. R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

46. Adamczak, R., A. Porollo, and J. Meller. 2004. Accurate prediction of solvent accessibility using neural networks-based regression. *Proteins.* 56:753–767.

47. Adamczak, R., A. Porollo, and J. Meller. 2005. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins.* 59:467–475.

48. Keskin, O., C. J. Tsai, H. Wolfson, and R. Nussinov. 2004. A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications. *Protein Sci.* 13:1043–1055.

49. Morris, G. M., R. Huey, …, A. J. Olson. 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30:2785–2791.

50. Tyurina, Y. Y., V. Kini, …, V. E. Kagan. 2006. Mechanisms of cardiolipin oxidation by cytochrome *c*: relevance to pro- and antiapoptotic functions of etoposide. *Mol. Pharmacol.* 70:706–717.

51. Trott, O., and A. J. Olson. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31:455–461.

52. Berg, S., M. Edman, …, A. Wieslander. 2001. Sequence properties of the 1,2-diacylglycerol 3-glucosyltransferase from *Acholeplasma laidlawii* membranes. Recognition of a large group of lipid glycosyltransferases in eubacteria and archaea. *J. Biol. Chem.* 276:22056–22063.

53. Hovius, R., J. Thijssen, …, B. de Kruijff. 1993. Phospholipid asymmetry of the outer membrane of rat liver mitochondria. Evidence for the presence of cardiolipin on the outside of the outer membrane. *FEBS Lett.* 330:71–76.

54. Arnarez, C., S. J. Marrink, and X. Periole. 2013. Identification of cardiolipin binding sites on cytochrome *c* oxidase at the entrance of proton channels. *Sci. Rep.* 3:1263–1270.

55. Pebay-Peyroula, E., C. Dahout-Gonzalez, …, G. Brandolin. 2003. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature.* 426:39–44.

56. Esser, L., B. Quinn, …, D. Xia. 2004. Crystallographic studies of quinol oxidation site inhibitors: a modified classification of inhibitors for the cytochrome *bc*(1) complex. *J. Mol. Biol.* 341:281–302.

57. Huang, L. S., D. Cobessi, …, E. A. Berry. 2005. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc1 complex: a new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. *J. Mol. Biol.* 351:573–597.

58. Pember, S. O., G. L. Powell, and J. D. Lambeth. 1983. Cytochrome P-450scc-phospholipid interactions. Evidence for a cardiolipin binding site and thermodynamics of enzyme interactions with cardiolipin, cholesterol, and adrenodoxin. *J. Biol. Chem.* 258:3198–3206.

59. Ou, W. J., A. Ito, …, T. Omura. 1988. Specific binding of mitochondrial protein precursors to liposomes containing cardiolipin. *J. Biochem.* 103:589–595.

60. Eble, K. S., W. B. Coleman, …, C. C. Cunningham. 1990. Tightly associated cardiolipin in the bovine heart mitochondrial ATP synthase as analyzed by 31P nuclear magnetic resonance spectroscopy. *J. Biol. Chem.* 265:19434–19440.

61. Althoff, T., D. J. Mills, …, W. Kühlbrandt. 2011. Arrangement of electron transport chain components in bovine mitochondrial supercomplex I1III2IV1. *EMBO J.* 30:4652–4664.

62. Lambeth, J. D. 1981. Cytochrome P-450scc. Cardiolipin as an effector of activity of a mitochondrial cytochrome P-450. *J. Biol. Chem.* 256:4757–4762.

63. Zolese, G., A. Tangorra, …, R. Donato. 1988. Interaction of S-100b protein with cardiolipin vesicles as monitored by electron spin resonance, pyrene fluorescence and circular dichroism. *Cell Calcium.* 9:149–157.

64. Hirata, K., K. Shinzawa-Itoh, …, H. Ago. 2014. Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. *Nat. Methods.* 11:734–736.

65. Jormakka, M., S. Törnroth, …, S. Iwata. 2002. Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science.* 295:1863–1868.

66. Kalanxhi, E., and C. J. Wallace. 2007. Cytochrome *c* impaled: investigation of the extended lipid anchorage of a soluble protein to mitochondrial membrane models. *Biochem. J.* 407:179–187.

67. Yung, B. Y., and A. Kornberg. 1988. Membrane attachment activates dnaA protein, the initiation protein of chromosome replication in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 85:7202–7205.

68. Horsefield, R., V. Yankovskaya, …, S. Iwata. 2006. Structural and computational analysis of the quinone-binding site of complex II (succinate-ubiquinone oxidoreductase): a mechanism of electron transfer and proton conduction during ubiquinone reduction. *J. Biol. Chem.* 281:7309–7316.

69. Mizushima, T., S. Natori, and K. Sekimizu. 1992. Inhibition of *Escherichia coli* DNA topoisomerase I activity by phospholipids. *Biochem. J.* 285:503–506.

70. Hao, G. F., F. Wang, …, G. F. Yang. 2012. Computational discovery of picomolar Q(o) site inhibitors of cytochrome *bc*1 complex. *J. Am. Chem. Soc.* 134:11168–11176.

71. Thuaud, F., N. Ribeiro, …, L. Désaubry. 2013. Prohibitin ligands in cell death and survival: mode of action and therapeutic potential. *Chem. Biol.* 20:316–331.

72. Kertesz, Z., B. B. Yu, …, R. B. Sim. 1995. Characterization of binding of human beta 2-glycoprotein I to cardiolipin. *Biochem. J.* 310: 315–321.

73. Matsuura, E., Y. Igarashi, …, T. Koike. 1992. Heterogeneity of anticardiolipin antibodies defined by the anticardiolipin cofactor. *J. Immunol.* 148:3885–3891.

74. Shintre, C. A., A. C. Pike, …, E. P. Carpenter. 2013. Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc. Natl. Acad. Sci. USA.* 110:9710–9715.

75. Zhuravleva, E., H. Gut, …, B. A. Hemmings. 2012. Acyl coenzyme A thioesterase Them5/Acot15 is involved in cardiolipin remodeling and fatty liver development. *Mol. Cell. Biol.* 32:2685–2697.

76. Gonzalvez, F., Z. T. Schug, …, E. Gottlieb. 2008. Cardiolipin provides an essential activating platform for caspase-8 on mitochondria. *J. Cell Biol.* 183:681–696.

77. Sani, M. A., E. J. Dufourc, and G. Gröbner. 2009. How does the Bax-alpha1 targeting sequence interact with mitochondrial membranes? The role of cardiolipin. *Biochim. Biophys. Acta.* 1788: 623–631.

78. Yang, J., K. Kulkarni, …, D. Barford. 2011. Mechanism of isoprenylcysteine carboxyl methylation from the crystal structure of the integral membrane methyltransferase ICMT. *Mol. Cell.* 44:997–1004.

79. Wang, Y., and N. Tjandra. 2013. Structural insights of tBid, the caspase-8-activated Bid, and its BH3 domain. *J. Biol. Chem.* 288:35840–35851.

80. Gonzalvez, F., F. Pariselli, …, P. X. Petit. 2005. tBid interaction with cardiolipin primarily orchestrates mitochondrial dysfunctions and subsequently activates Bax and Bak. *Cell Death Differ.* 12:614–626.

81. Dieudé, M., H. Striegl, …, J. Rauch. 2011. Cardiolipin binds to CD1d and stimulates CD1d-restricted $\gamma\delta$ T cells in the normal murine repertoire. *J. Immunol.* 186:4771–4781.

82. Kita, K., N. Sueyoshi, …, M. Ito. 2002. Activation of bacterial ceramidase by anionic glycerophospholipids: possible involvement in ceramide hydrolysis on atopic skin by *Pseudomonas ceramidase*. *Biochem. J.* 362:619–626.

83. Indiveri, C., A. Tonazzi, …, F. Palmieri. 1991. Kinetic characterization of the reconstituted carnitine carrier from rat liver mitochondria. *Biochim. Biophys. Acta.* 1065:231–238.

84. Kljashtorny, V. G., T. Y. Fufina, …, A. G. Gabdulkhakov. 2014. Molecular dynamic studies of reaction centers mutants from *Rhodobacter sphaeroides* and his mutant form L(M196)H+H(M202)L. *Crystallogr. Rep.* 59:536–541.

85. Saer, R. G., J. Pan, …, J. T. Beatty. 2014. Structural and kinetic properties of *Rhodobacter sphaeroides* photosynthetic reaction centers containing exclusively Zn-coordinated bacteriochlorophyll as bacteriochlorin cofactors. *Biochim. Biophys. Acta.* 1837: 366–374.

86. Ruprecht, J. J., A. M. Hellawell, …, E. R. Kunji. 2014. Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proc. Natl. Acad. Sci. USA.* 111:E426–E434.

87. Lange, C., J. H. Nett, …, C. Hunte. 2001. Specific roles of protein-phospholipid interactions in the yeast cytochrome *bc*1 complex structure. *EMBO J.* 20:6591–6600.

88. Solmaz, S. R., and C. Hunte. 2008. Structure of complex III with bound cytochrome *c* in reduced state and definition of a minimal core interface for electron transfer. *J. Biol. Chem.* 283:17542–17549.

89. Palsdottir, H., C. G. Lojero, …, C. Hunte. 2003. Structure of the yeast cytochrome *bc*1 complex with a hydroxyquinone anion Qo site inhibitor bound. *J. Biol. Chem.* 278:31303–31311.

90. Komano, H., K. Homma, and S. Natori. 1991. Involvement of sapecin in embryonic cell proliferation of *Sarcophaga peregrina* (flesh fly). *FEBS Lett.* 289:167–170.

91. Zardeneta, G., and P. M. Horowitz. 1993. Physical characterization of a reactivatable liposome-bound rhodanese folding intermediate. *Biochemistry.* 32:13941–13948.

92. Morrice, N. A., B. Gabrielli, …, R. E. Wettenhall. 1994. A cardiolipin-activated protein kinase from rat liver structurally distinct from the protein kinases C. *J. Biol. Chem.* 269:20040–20046.

93. Müller, M., R. Moser, …, E. Carafoli. 1985. Cardiolipin is the membrane receptor for mitochondrial creatine phosphokinase. *J. Biol. Chem.* 260:3839–3843.

94. Beleznai, Z., and V. Jancsik. 1989. Role of cardiolipin in the functioning of mitochondrial L-glycerol-3-phosphate dehydrogenase. *Biochem. Biophys. Res. Commun.* 159:132–139.

95. Fry, M., and D. E. Green. 1981. Cardiolipin requirement for electron transfer in complex I and III of the mitochondrial respiratory chain. *J. Biol. Chem.* 256:1874–1880.

96. Cheneval, D., M. Müller, and E. Carafoli. 1983. The mitochondrial phosphate carrier reconstituted in liposomes is inhibited by doxorubicin. *FEBS Lett.* 159:123–126.

97. Osman, C., M. Haag, …, T. Langer. 2009. The genetic interactome of prohibitins: coordinated control of cardiolipin and phosphatidylethanolamine by conserved regulators in mitochondria. *J. Cell Biol.* 184:583–596.

98. Zhao, Y., Y. Q. Chen, …, G. Cao. 2009. The microsomal cardiolipin remodeling enzyme acyl-CoA lysocardiolipin acyltransferase is an acyltransferase of multiple anionic lysophospholipids. *J. Lipid Res.* 50:945–956.

99. Inuzuka, T., A. Inokawa, …, M. Maki. 2013. ALG-2-interacting Tubby-like protein superfamily member PLSCR3 is secreted by an exosomal pathway and taken up by recipient cultured cells. *Biosci. Rep.* 33:e00026.

100. Van, Q., J. Liu, …, G. M. Hatch. 2007. Phospholipid scramblase-3 regulates cardiolipin de novo biosynthesis and its resynthesis in growing HeLa cells. *Biochem. J.* 401:103–109.

101. Christie, D. A., C. D. Lemke, …, J. Madrenas. 2011. Stomatin-like protein 2 binds cardiolipin and regulates mitochondrial biogenesis and function. *Mol. Cell. Biol.* 31:3845–3856.

102. Weber, T. A., S. Koob, …, A. S. Reichert. 2013. APOOL is a cardiolipin-binding constituent of the Mitofilin/MINOS protein complex determining cristae morphology in mammalian mitochondria. *PLoS One.* 8:e63683.

103. Ray, N. B., L. Durairaj, …, R. K. Mallampalli. 2010. Dynamic regulation of cardiolipin by the lipid pump Atp8b1 determines the severity of lung injury in experimental pneumonia. *Nat. Med.* 16:1120–1127.

104. Vancura, A., and D. Haldar. 1994. Purification and characterization of glycerophosphate acyltransferase from rat liver mitochondria. *J. Biol. Chem.* 269:27209–27215.

105. Brandt, M. A., and S. G. Powers-Lee. 1991. The interaction of cardiolipin with rat liver carbamoyl phosphate synthetase I. *Arch. Biochem. Biophys.* 290:14–20.

106. Ueda, A., M. Suzuki, …, K. Shiomi. 2006. Purification, properties and cDNA cloning of neoverrucotoxin (neoVTX), a hemolytic lethal factor from the stonefish *Synanceia verrucosa* venom. *Biochim. Biophys. Acta.* 1760:1713–1722.

107. Purzycka-Preis, J., and M. M. Zydowo. 1987. Regulatory effect of pig heart phospholipids on heart muscle AMP-deaminase. *Int. J. Biochem.* 19:565–568.

108. Milon, L., P. Meyer, …, M. L. Lacombe. 2000. The human *nm23*-H4 gene product is a mitochondrial nucleoside diphosphate kinase. *J. Biol. Chem.* 275:14264–14272.

109. Schlattner, U., M. Tokarska-Schlattner, …, R. M. Epand. 2009. Mitochondrial kinases and their molecular interaction with cardiolipin. *Biochim. Biophys. Acta.* 1788:2032–2047.

110. Vacheron, M. J., E. Clottes, …, C. Vial. 1997. Mitochondrial creatine kinase interaction with phospholipid vesicles. *Arch. Biochem. Biophys.* 344:316–324.

111. MacArthur, M. W., and J. M. Thornton. 1991. Influence of proline residues on protein conformation. *J. Mol. Biol.* 218:397–412.

112. Oliva, B., P. A. Bates, …, M. J. Sternberg. 1997. An automated classification of the structure of protein loops. *J. Mol. Biol.* 266:814–830.

113. Kagan, V. E., Y. Y. Tyurina, …, H. Bayir. 2015. Cardiolipin signaling mechanisms: collapse of asymmetry and oxidation. *Antioxid. Redox Signal.* 22:1667–1680.

114. Tyurina, Y. Y., S. M. Poloyac, …, V. E. Kagan. 2014. A mitochondrial pathway for biosynthesis of lipid mediators. *Nat. Chem.* 6:542–552.

115. Kapralov, A. A., N. Yanamala, …, V. E. Kagan. 2011. Topography of tyrosine residues and their involvement in peroxidation of polyunsaturated cardiolipin in cytochrome *c*/cardiolipin peroxidase complexes. *Biochim. Biophys. Acta.* 1808:2147–2155.

116. Atkinson, J., A. A. Kapralov, …, V. E. Kagan. 2011. A mitochondria-targeted inhibitor of cytochrome *c* peroxidase mitigates radiation-induced death. *Nat. Commun.* 2:497.

117. Kagan, V. E., H. A. Bayir, …, G. Borisenko. 2009. Cytochrome *c*/cardiolipin relations in mitochondria: a kiss of death. *Free Radic. Biol. Med.* 46:1439–1453.

118. Beyer, K., and M. Klingenberg. 1985. ADP/ATP carrier protein from beef heart mitochondria has high amounts of tightly bound cardiolipin, as revealed by 31P nuclear magnetic resonance. *Biochemistry.* 24:3821–3826.