Supplementary Information for

Identification of cardiolipin binding sites on cytochrome *c* oxidase at the entrance of proton channels.

C. Arnarez[a], S.J. Marrink[a], X. Periole[a,*]

[a]Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[1]To whom correspondence should be addressed. E-mail: x.periole@rug.nl

Content:

1- **Supplementary Figure**

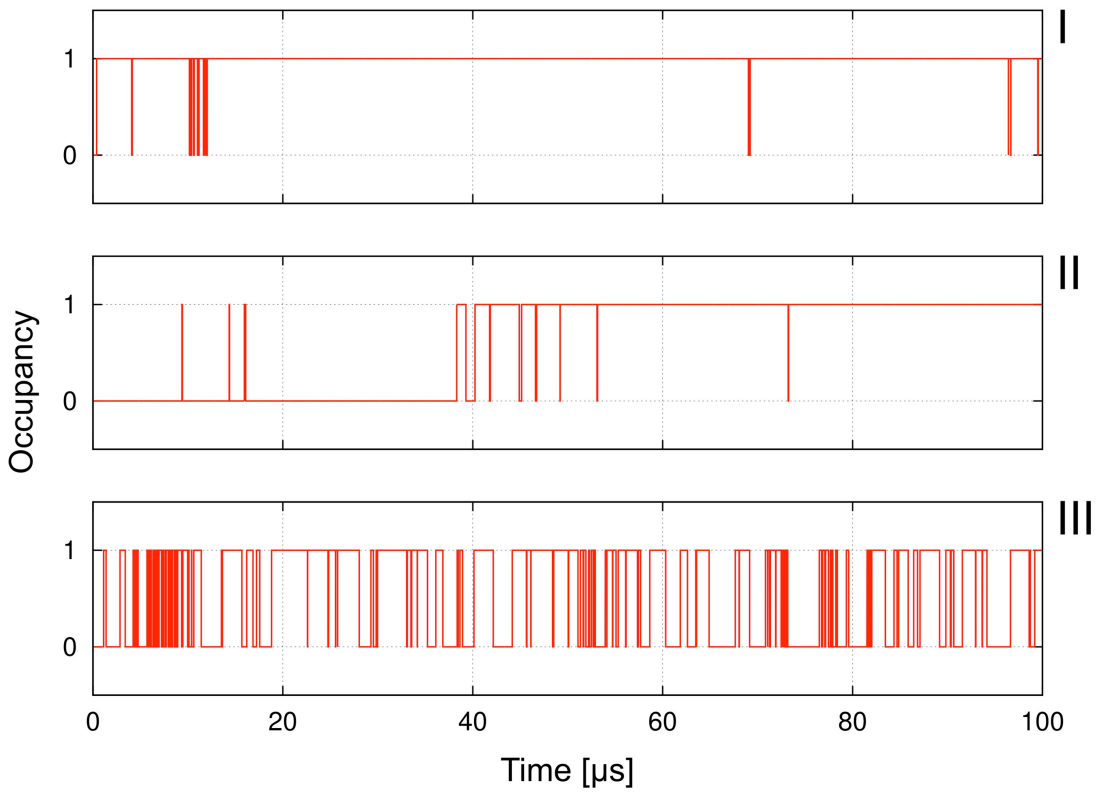

Figure S1 | Occupancy of site I to III (Top to bottom). Illustration of the hierarchy of CL's binding site strength and dynamics. Sites I and II and strong binder and are only sporadically disrupted. In Site I it is actually the same CL that is bound the all time of simulation. In site II it is a unique CL that binds at t=40 µs and remains for the rest of the simulation, but three other unsuccessfully attempted to bind earlier in the simulation (t~10-20 µs). Both behaviors contrast with the very dynamic one observed in site III and offering enough statistics to compute the auto-correlation analysis described in Supplementary Methods. This analysis was not possible in the cases of site I and site II in which cases the statistics is obviously not sufficient. Instead we qualitatively estimated the lifetime to 50 and 30 µs. Much longer simulations would be required to get more precise values.

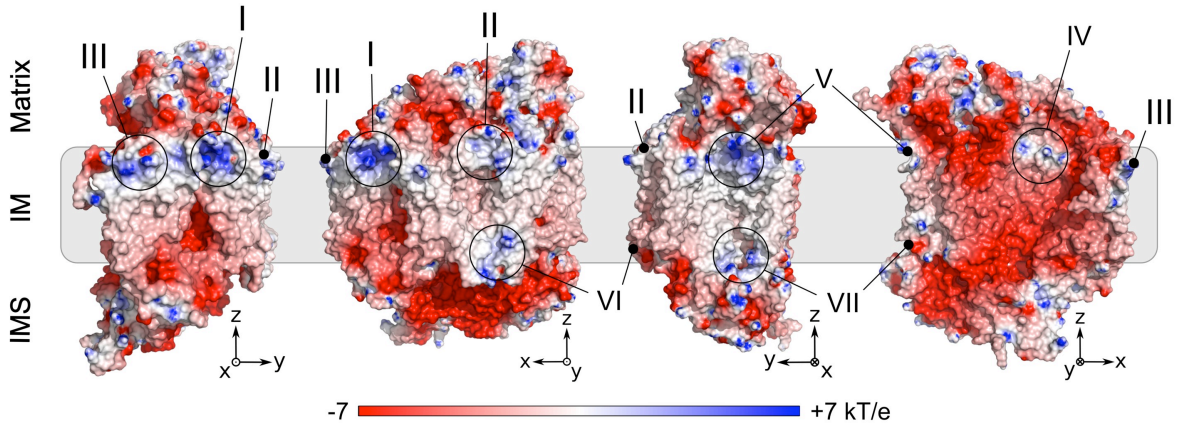

Figure S2 | Electrostatic map of the CIV. The electrostatic potential was computed using the DelPhi package[2]. Four orientations are shown with the position of the sites defined in this study (see main manuscript) also indicated. Positive regions are reported in blue, negative in red. Note the strong correlation between regions of the protein surface with a positive electrostatic potential and the location of the CL's binding sites (see also CL's densities shown in Figure 2 of the main manuscript). The approximate location of the inner membrane (IM) is indicated by a grey shade. The IM defines the separation between the inter membrane space (IMS) and the matrix compartments.

2- Legend for the supplementary animation. The animation reports on the dynamics of binding/unbinding of CLs on CIV. The system is viewed from the matrix side and the trace of the protein backbone is shown in gray sticks. The red and orange wireframes indicate the CL densities (averaged over a 100 μs CGMD simulation; see Fig. 2 in the main manuscript) located on the upper leaflet (matrix side) and the lower leaflet (inter membrane space, IMS) of the bilayer, respectively. Simulation time is indicated in μs (us) on the lower-left corner of the frame. CLs are shown as colored sticks (one color per CL) with the linker bead shown as a large sphere during the first 12 μs; The remaining of the time only the linker bead is shown to ease the

visualization. The positions of the CLs are taken as the average over 0.2 μs windows to smooth out the high frequency motion. The POPC lipid molecules and the aqueous phase are also omitted from the representation. See Fig. 1 in the main manuscript to have a more complete representation of the system.

3- Supplementary Table

Table S1 | Table of correspondence between the letter (top row and used in the main manuscript) and roman number subunit names[1] (bottom row).

A	B	C	D	E	F	G	H	I	J	K	L	M
I	II	III	IV	Va	Vb	VIa	VIb	Vic	VIIa	VIIb	VIIc	VIII

4- Cardiolipin binding sites versus lipids found in crystal structures

Phosphatidylethanolamines (PE) lipids in the crystal occupy the CL binding sites III and IV predicted by our simulations (PE1 and PE3 in main-Fig. 4). This is surprising since in contrast to CLs PE headgroups are uncharged and CL's tail is double the size of PE one. Other stabilizing factors must be involved. In the case of site III, it is extremely polar (main-Fig. 2D), which seems not ideal for binding a PE lipid headgroup. However, in the crystal structure PE3 shares the site III (as defined in the simulations) with a cholic acid molecule[3], which interacts with G:R26 (a CL ligand in site III) and is located at the interface between the two CIV monomers. The location of site III close to the dimer interface is likely to lead to different accessibilities of the binding site in the monomeric form of CIV used in the simulation compared to the dimeric form found in the crystal. One can speculate that the binding of a CL is not possible to the dimeric form of CIV, and is replaced by a PE/cholate. We cannot, however, rule out the possibility of preferential binding of PE molecule at this location in the monomer since PEs are

not present in our simulation. The CL occupancy of these sites is relatively low, as well as their binding lifetime (main-Table 1).

Triglyceride (TGL) molecules occupy CL binding sites II, V, and VII molecule in the crystal structure (TG2, TG1, and TG3 in main-Fig. 4, respectively). TGLs and CLs have bulky tails and might therefore mimic each other. It is quite possible that on one hand TGLs replace CLs weakly bound to CIV and lost during the purification process or on the other hand that the lack of TGLs in our simulations leads CLs to replace them. Either way, this behavior would be consistent with the observations that the aliphatic tails are in fact most important for CL binding to complex IV[4,5], that detergent molecules can at least partially occupy other lipid binding sites[6] and finally that the hydrophobic tails of lipids are often better resolved due to the crystallization process, which disrupts the ionic interactions between lipid head groups and polar residues[6,7]. It is at this point difficult to tell what really happens because TGLs are not present in our systems and we cannot therefore exclude that TGLs bound to CIV in sites V and VII stronger than CLs do. However, in the case of site II, a CL is most likely bound (see above and reference[8]) and thus suggests that TGLs might replace CL in the crystal. Moreover there is to our best knowledge no report of a function associated with TGL molecules and linked to the activity of CIV or the respiratory chain. It is also only recently[3] that TGL have been detected in the mitochondrial membrane[9]. The double occupancy of sites V and VII, which are on the same face of the protein but on opposite leaflets, illustrates the size of the cavity of the protein; It is large enough to actually accommodate both sites, so four CLs. They have a relative high occupancy but a low residence time at the exception of site VIIa.

The sites occupied by phosphatidylglycerols (PGs) in the crystal structure are not occupied by CLs in the simulations, despite that PG headgroups carry an anionic charge (-1 *e*) similar to CLs

(-2 *e*). PG1 and PG2 (main-Fig. 4) are deeply buried into the complex and thus difficult to access by CLs from the bulk membrane. In the cases of PG3 and PG4 (main-Fig. 4) it is apparent from our simulations that CLs do not have notable affinity for these sites. PG4 is located in a groove of the protein in which the bulky tail of CL cannot enter. PG3 is at the interface of the CIV dimer and it is not clear why a CL does not bind in the simulation.

5- Extended Methods

Molecular models. The model of CIV was built based on its experimentally determined structure from bovine heart mitochondria. The *Protein Data Bank* (PDB) entries 1occ and 2occ[2,10] were used to built a complete model of CIV. Structures of bovine heart mitochondria CIV available in the PDB contain 13 subunits labeled A to M. Although CIV is found in a dimeric form in the crystals, it was simulated as a monomer, its functional form. Details on the complex and its subunits are given in main-Fig. 1A.

In the structure of CIV four CLs co-crystallize with the protein in its dimeric form (main-Fig. 1A). The consideration of the symmetry of the dimer arrangement in the crystals results in three CL binding sites per monomer (main-Fig. 1A). One CL binding site, CL2, is located on a membrane-exposed protein surface with its headgroup facing the matrix side of the inner-membrane. The two other sites, CL1a/b, are located at the interface of the dimer of the complex with their head groups facing the IMS. In contrast to earlier studies[1,3,8] we differentiate CL1*a* and CL1*b* because although the symmetry of the dimer makes them equivalent they do form two dissimilar binding sites on each monomer. We simulate CIV in a monomeric form and since they are exposed to the membrane bulk the CL positions described above were not included in the initial conformation of the system.

System set-up. The atomic protein structure described above was converted to a CG model and embedded into a pre-equilibrated POPC/CL membrane patch with a 20:1 PC/CL ratio. The system was energy minimized and simulated for 10 ns with positional restraints on the backbone beads of the protein to relax the solvent and side-chains before starting production runs. The system shown in main-Fig. 1B contains the protein (1780 residues; 4117 beads), a POPC bilayer (966 lipids; 12,558 beads) including CLs (48 CLs; 1296 beads) and the aqueous phase (51,549 water beads and 103 sodium ions). In total, the system contains slightly less than 70,000 CG beads and runs at a speed of ~1.3 μs/day using 120 CPUs with a clock rate of 2.6 GHz.

Simulation details. All CG simulations were performed using the GROMACS simulation package version 4.0. The systems were described with the MARTINI CG force field for biomolecules (version 2.0[3,11]) and its extension to proteins (version 2.1[4,5,12]) together with the ElNeDyn approach[6,13]. Conventional simulation setups associated with the use of the MARTINI force field were used. That includes an integration time step of 20 fs for production runs (10 fs during equilibration) and non-bonded interactions cutoff at a distance r_{cut} = 1.2 nm. The Lennard-Jones potential is shifted from r_{shift} = 0.9 nm to r_{cut}. The electrostatic potential is shifted from r_{shift} = 0.0 nm to r_{cut} using a relative dielectric screening constant of 15. The protein, membrane bilayer (POPC and CL) and aqueous phase (water and Na^+ ions) were coupled independently to an external temperature bath at 300 K using a Berendsen thermostat[6,7,14] with a relaxation time of 0.5 ps. The pressure was weakly coupled (Berendsen barostat[8,14]) to an external bath at 1 bar using a relaxation time of 1.2 ps and a semi-isotropic pressure scheme. CG sodium ions were used to neutralize the systems. CL topology (main-Fig. 1C) and parameters were taken from the

work of Dahlberg *et al.*[3,15]. The ElNeDyn approach, which defines an elastic network between the backbone beads to control the conformation of a protein, was used on each subunit of the complexes separately; *i.e.* springs are not present between the subunits. The integrity of the complex is only dependent on non-bonded interactions. Extend of the network was 0.9 nm and the force constant of the springs was set to 500 kJ mol^{-1} nm^{-2} [9,13]. Trajectory snapshots were saved every 500 ps for analysis purpose unless otherwise specified. The complex and its subunits were both numerically and structurally stable over the time scale of the simulations. Three 20 μs simulations starting with randomized initial velocities were initially performed and analyzed to judge the convergence of a few observables. The results presented are extracted from a 100 μs simulation, with the first μs of the simulations excluded, considered as an equilibration period. The root-mean-square deviation of the protein complex (with respect to the experimental structures) typically reached values of 0.3-0.4 nm indicative of its structural stability.

Binding site definition. The densities of the CLs were used to localize their potential binding sites at the protein surface and were computed by means of the "occupancy" option of the VolMap tool of VMD[16] using a 0.2 nm resolution grid. An occupancy level at least 5 times higher than the level corresponding to the bulk region was used to identify the binding sites. To characterize the CL binding sites in greater detail we used the following protocol. For each frame of the 100 μs simulation of the complex we defined a list of residues simultaneously in contact with a unique CL. A contact was counted when a residue was within 0.7 nm of the CL head group (phosphatidyl and glycerol CG beads). The time series of the lists obtained were then sorted using a clustering algorithm[17]: the lists with at least 66% of common residues were clustered together and the most populated cluster (*i.e.* the lists composing it) was removed from

the pool of lists; the procedure was repeated on the remaining lists until the pool was empty. The residues present in at least 80% of the lists composing a cluster were selected as the representative set of residues of the cluster. Note that several other protocols were tested for the definition of the binding sites. The final definition of the sites was only slightly sensitive to the protocol used.

Binding site averaged occupancy. The occupancy of a site *j* by a CL *L* at a time t was determined as:

$$\xi_L^j(t) = \begin{cases} 1, & d_{jL}(t) \leq \delta_L \\ 0, & d_{jL}(t) > \delta_L \end{cases}$$

where $d_{jL}(t)$ is the distance between the center of mass of the CL head group and the center of mass of the binding site *j*, δ_L is a cutoff value set to 0.8 nm (accounts for the relative large size of a CL head group). The instantaneous values of the distance were smoothed out over windows of 50 ns. The time series of the site occupancies were additionally smoothed out by averaging over 2.5 ns time windows, which resulted in an effective averaging of ~52.5 ns. These precautions were necessary to discard the fast leave and return of CLs from the site considered. The error on the occupation values was estimated by splitting the trajectory in 10 slices of 10 μs.

Binding site residence time. The residence time (also referred to as the survival or lifetime[18,19]), θ_L^j, of a CL L in a binding site *j* was obtained from the normalized time-correlation, $\sigma_L^j(t)$, which can be expressed as:

$$\sigma_L^j(t) = \frac{1}{T-t}\sum_{\upsilon=0}^{T-t} \rho_L^j(\upsilon, \upsilon + t)$$

where T is the total time of simulation and $\rho_L^j(\upsilon, \upsilon + t)$ a function taking the value of 1 if the CL L has occupied the site j for a time t after being in contact with the site at time υ, and 0 otherwise. ρ_L^j was obtained from the time series of the site's occupancy by a CL as described above. In practice the CLs were not differentiated since they occupy a site one at the time and as before a 52.5 ns window averaging was used to discard the fast leave and return movements of CLs. The residence time of CLs in a specific site j, θ^j, was obtained by fitting the time-correlation function relative to this site using: $\sigma^j(t) \approx \exp\left(-\frac{t}{\theta^j}\right)$. Note that the values of the CL's residence time were quite sensitive to the exact protocol used. Notably the smoothing of the raw data to remove the quick binding/unbinding events was found important to report on a realistic behavior of the CL.

Potentials of mean force (PMFs). PMFs were used to quantify the binding affinities of specific lipid molecules with respect to CL binding sites on CIV. PMFs were calculated using the umbrella sampling technique. The system used was composed of a CIV embedded in a patch of ~360 POPC and one molecule of the lipid of interest. We tested the binding of CL, TGL and POPC molecules. For each case we ran simulations with a set of umbrella potentials that covered from the lipid-bound situation to free in the bulk membrane (the profile reaches a plateau value). The umbrella potentials were applied to the distance between the center-of-mass of the lipid head group and of the binding site using a harmonic potential with a 1000 kJ mol^{-1} nm^{-2} force constant. This choice resulted in 20 to 30 umbrella simulations spaced by 0.1 nm. We reduced the positional and rotational motions of the protein by applying a weak position restrain on three backbone beads of the protein. These position restraints were not applied on the z coordinate and used a 200 kJ mol^{-1} nm^{-2} force constant (0.42 kcal mol^{-1} $Å^{-2}$). We applied the distance restraint

between the centers-of-masses along the direction normal to the protein surface only; therefore an additional weak position restraint (100 kJ mol^{-1} nm^{-2} force constant) was applied on one bead of the lipid head group (GL5 for CL, GLY for TGL and PO4 for POPC) to restrict its exploration of the direction parallel to the protein surface (perpendicular to the normal of the protein surface). The direction normal to the membrane plane was let free. Each umbrella simulation was run for 250 to 700 ns; the first 50 ns were excluded from the analysis. The weighted histogram analysis method (WHAM)[20] was used to combine and unbias the simulations, and to produce the free energy profiles. The Bayesian bootstrapping method was used to estimate the error on the calculation.

Supplementary References

1. Kadenbach, B., Jarausch, J., Hartmann, R. & Merle, P. Separation of mammalian cytochrome c oxidase into 13 polypeptides by a sodium dodecyl sulfate-gel electrophoretic procedure. *Anal Biochem* **129**, 517–521 (1983).
2. Li, L. *et al.* DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys* **5**, 9 (2012).
3. Shinzawa-Itoh, K. *et al.* Structures and physiological roles of 13 integral lipids of bovine heart cytochrome *c* oxidase. *EMBO J.* **26**, 1713–1725 (2007).
4. Robinson, N. C., Zborowski, J. & Talbert, L. H. Cardiolipin-Depleted Bovine Heart Cytochrome-C-Oxidase - Binding Stoichiometry and Affinity for Cardiolipin Derivatives. *Biochemistry* **29**, 8962–8969 (1990).
5. Robinson, N. C. Functional Binding of Cardiolipin to Cytochrome *c* oxidase. *J. Bioenerg. Biomembr.* **25**, 153–163 (1993).
6. Qin, L., Hiser, C., Mulichak, A., Garavito, R. M. & Ferguson-Miller, S. Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16117–16122 (2006).
7. Palsdottir, H. & Hunte, C. Lipids in membrane protein structures. *Biochim. Biophys. Acta* **1666**, 2–18 (2004).
8. Sedlák, E., Panda, M., Dale, M. P., Weintraub, S. T. & Robinson, N. C. Photolabeling of cardiolipin binding subunits within bovine heart cytochrome c oxidase. *Biochemistry* **45**, 746–754 (2006).
9. Daum, G. Lipids of mitochondria. *Biochim. Biophys. Acta* **822**, 1–42 (1985).
10. Tsukihara, T. *et al.* The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 A. *Science* **272**, 1136–1144 (1996).
11. Marrink, S.-J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The

MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
12. Monticelli, L. *et al.* The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).
13. Periole, X., Cavalli, M., Marrink, S.-J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).
14. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
15. Dahlberg, M. Polymorphic phase behavior of cardiolipin derivatives studied by coarse-grained molecular dynamics. *J. Phys. Chem. B* **111**, 7194–7200 (2007).
16. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).
17. Daura, X. *et al.* Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Edit.* **38**, 236–240 (1999).
18. Garcia, A. E. & Stiller, L. Computation of the mean residence time of water in the hydration shells of biomolecules. *J. Comput. Chem.* **14**, 1396–1406 (1993).
19. Periole, X., Rampioni, A., Vendruscolo, M. & Mark, A. E. Factors that affect the degree of twist in β-sheet structures: a molecular dynamics simulation study of a cross-β filament of the GNNQQNY peptide. *J. Phys. Chem. B* **113**, 1728–1737 (2009).
20. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).