Supporting Information

Electron Transfer Interactome of Cytochrome c

Alexander N. Volkov$^{1,2}$ and Nico A. J. van Nuland$^{1,2}$

2. Department of Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium.

1. Supplementary Figures S2
2. Supplementary Table S9
3. Supplementary Text S10
4. Supplementary Videos S15
Supplementary Figures

Figure S1. Distributions of Cc CMs around CcP in the ternary complexes colored by the heme-W191 (A) or heme-heme (B) distances and the corresponding ET rates (see the scale bar). Protein orientation is the same as in Figure 3A in the main text. Cc in the crystallographic orientation is shown as yellow surface. See Videos S3 and S4 for an expanded view of (A) and (B), respectively.

Figure S2. Distribution of Cc CMs around the subunit II of CCO in binary (A) and ternary (B) complexes colored by the heme-CuA distances and the corresponding ET rates (see the scale bar). Protein orientation is the same as in Figure 4C in the main text. (B) Cc bound to the high-affinity site of CCO is shown as a molecular surface.
Figure S3. Cc-Cbc1. (A) Crystal structure of the complex [S1]. The intramitochondrial region modeled in this work is shown as a blue molecular surface. Cc is colored yellow, heme groups of Cc and Cc1 are in red, the other Cbc1 redox cofactors are in orange. Horizontal lines indicate approximate location of the mitochondrial membrane. The antibody fragments used for protein crystallization are removed for clarity. (B) Close-up of the crystallographic Cc-Cbc1 orientation. (C) The molecular surface of the modeled Cbc1 region colored by the electrostatic potential (see the scale bar). Protein orientation is the same as in (B). (D) Distribution of Cc CMs around Cbc1 colored by the heme-heme distances and the corresponding ET rates (see the scale bar). Protein orientation is the same as in (B). (E) Distribution of the intermolecular heme-heme distances (left) and ET rates (right). In the right panel, the red and black traces indicate the ET rates calculated with $\lambda = 0.7$ and 1 ev, respectively. The solid vertical lines indicate the intermolecular heme-heme distance (left panel) and the corresponding ET rates (right panel; thick – calculated with $\lambda = 0.7$, thin – with $\lambda = 1$ ev) in the crystallographic orientation. The dashed line denotes the fastest experimentally measured intermolecular ET rate in the Cc-Cbc1 system [S2].
Figure S4. Intermolecular ET in the Cc-Fcb₂ complex with the simulated interdomain motion. (A) Simulated ensemble of Fcb₂ domain-domain orientations with residues 89-99 as a mobile linker. The green mesh is a reweighted atomic probability density map [S3], plotted at a threshold of 20% maximum, for the overall distribution of the Cb₃ domains among 100 generated structures. Ten representative, low-energy solutions are shown as blue ribbons, with heme groups in red sticks. Crystallographic monomer is in cartoon. Protein orientation is the same as in Figure 5B in the main text. (B) Distributions of the intermolecular heme-heme distances (top) and ET rates (bottom) for the crystallographic (red) and the simulated domain-domain orientations (black). The latter was calculated for 10 solutions shown in (A), with the black line denoting the average.
Figure S5. Ce-SOX. (A) Crystal structure of chicken SOX [S4]. The monomer modeled in this work is shown as a molecular surface. Cb₃ heme domain is colored yellow; heme groups and Mo atoms are in red. (B) The molecular surface of the SOX monomer colored by the electrostatic potential (see the scale bar). Protein orientation is the same as in (A). (C) Distribution of Ce CMs around the SOX heme b₃ domain in the crystallographic orientation colored by the heme-heme distances and the corresponding ET rates (see the scale bar). Protein orientation is the same as in (A). (D) The green mesh is a reweighted atomic probability density map [S3], plotted at a threshold of 10 % maximum, for the overall distribution of the SOX Cb₃ domains among 100 generated structures. Fifteen representative, low-energy solutions are shown as blue ribbons, with heme groups in red sticks. Crystallographic monomer is in cartoon. (E) Distribution of the intermolecular heme-heme distances (left) and ET rates (right) for the crystallographic (red) and the simulated domain-domain orientations (black), the latter calculated for 15 solutions shown in (D), with the black line denoting the average.
Figure S6. Functional epitope of Cc in the complex with Cb₅. (A) Cc and Cb₅ in the hypothetical model of Salemme [S5] are in yellow and blue, respectively, with heme groups shown as sticks and iron atoms as red spheres. (B) Distribution of Cb₅ CMs around Cc colored by the heme-heme distances (see the scale bar). Protein orientation is the same as in (A). See Video S7 for an expanded view.

Figure S7. Functional epitope of Cc in the complex with CcP. (A) Crystal structure of the complex [S6]. Cc and CcP are in yellow and blue, respectively, with heme groups shown as sticks and iron atoms as red spheres. (B) Distribution of CcP CMs around Cc colored by the heme-W191 distances (see the scale bar). Protein orientation is the same as in (A). See Video S8 for an expanded view.
Figure S8. Cc-Erv1. (A) Homology model of ScErv1 based on the crystal structure of AtErv1 [S7]. Redox-active groups are shown as red sticks. The isoalloxazine ring of FAD and the C130-C133 disulfide bridge are indicated by the labels. (B) Distribution of Cc CMs around ScErv1 colored by the flavin-heme (left) or disulfide-heme (right) distances and the corresponding ET rates (see the scale bar). Protein orientation is the same as in (A). (C) Distribution of the intermolecular distances (top) and ET rates (bottom) for flavin-heme (red traces) and disulfide-heme (black traces). The filled and open symbols in the bottom panel refer to the $k_{ET}$ rates calculated with $\lambda = 0.7$ and 1 eV, respectively.
Figure S9. The workflow in the computational protocol used for the Cc-CcP complex. All scripts, input files, and some of the output data are provided in the Dataset S1.
## Supplementary Table

### Table S1. Molecular systems modeled in this study.

<table>
<thead>
<tr>
<th>Stationary protein&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Moving protein</th>
<th>Atoms&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Structures&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>yeast CcP (2PCC) [S6]</td>
<td>yeast Cc (2PCC) [S6]</td>
<td>6,475</td>
<td>94,913,856</td>
</tr>
<tr>
<td>bovine Cb&lt;sub&gt;2&lt;/sub&gt; (1CYO) [S22]</td>
<td>yeast Cc (1YCC) [S23]</td>
<td>3,255</td>
<td>59,797,440</td>
</tr>
<tr>
<td>yeast Cbc1 (3CX5) [S1] chains D (62-261); E (91-215); F; I (46-58); O (62-261); P (91-215); Q; T (46-58)</td>
<td>yeast Cc (1YCC) [S23]</td>
<td>14,629</td>
<td>35,435,232</td>
</tr>
<tr>
<td>bovine CCO (1V54) [S24] chains A (37-55, 115-144, 209-232, 287-302, 361-374, 428-451); B (2-20, 85-228); C (31-43, 99-127, 184-196, 255-261); D (99-147); G (41-84); H; I (39-73); J (53-54); K (33-54); L (40-47); M (33-43)</td>
<td>horse Cc (1HRC) [S25]</td>
<td>11,442</td>
<td>40,481,856</td>
</tr>
<tr>
<td>yeast Fcb&lt;sub&gt;2&lt;/sub&gt; (1KBI) [S26] chains A and B. Missing heme domain in chain B modeled based on that in chain A. Missing residues in the 300-308 loop modeled.</td>
<td>yeast Cc (1YCC) [S23]</td>
<td>18,064</td>
<td>38,677,824-54,820,800</td>
</tr>
<tr>
<td>chicken SOX (1SOX) [S4] chain A. Missing atoms in residues 85, 86, 89, and 94 modeled.</td>
<td>horse Cc (1HRC) [S25]</td>
<td>8,939</td>
<td>34,992,000-69,470,784</td>
</tr>
<tr>
<td>residues 84-183 of the yeast Erv1 homology-modeled on the structure of AtErv1 (2HJ3) [S7]</td>
<td>yeast Cc (1YCC) [S23]</td>
<td>3,482</td>
<td>85,769,280</td>
</tr>
</tbody>
</table>

<sup>a</sup> Residue selections for parts of modeled protein chains are given in parentheses; <sup>b</sup> total number of atoms in the system; <sup>c</sup> number of structures sampled per run.
Supplementary Text

Text S1. Description of the computational protocol.

The computational protocol was implemented as a pipeline of several Xplor-NIH scripts, where output of one script is the input for the next. A typical workflow is illustrated on the example of the Cc-CcP complex in Figure S9. See below for a brief description of the scripts.

**gen.cc.inp, gen.ccp.inp**

These scripts generate Xplor parameter and structure files for Cc and CcP and add missing hydrogen atoms according to the standard Xplor topology and parameter sets (topallhdg.pro, top11 pep, and parallhdg.pro). For the heme groups and their ligands we used the topology and parameter files available in our group; however, those can also be obtained from the HIC-Up database (http://xray.bmc.uu.se/hicup/). The Cc and CcP coordinates were taken from the PDB entry for the Cc-CcP complex (2PCC). Coordinates for two protein chains and cofactors were placed in separate files and used as input.

**orient.inp**

This script combines Cc and CcP into a single molecular structure; places pseudoatoms (atom name G, residue name GB) at the proteins’ CMs; orients the protein complex in the reference frame defined in Figure 1A (main text) so that CcP and Cc CMs appear at the coordinate origin and on the positive z axis, respectively; and outputs the reference coordinate file (ref.cc-ccp.pdb). Further, the script finds the protein-protein orientation with the minimal Cc-CcP heme-heme separation at the original, fixed CM positions (i.e. the structure with the frontal Cc orientation, Figure 1B in the main text), which is used as an input for the subsequent conformational search.

**search.py**

The script performs a complete conformational search of the Cc-CcP binding geometries and outputs the coordinates of the Cc CMs, which define the coverage of the space sampled. As a rule, the entire search was split among 10 runs, using 10 CPUs each. The current run number and the total number of CPUs used to carry out the search (given by the variables `run_number` and `nproc_tot`, respectively) define the number of ($\theta$, $\varphi$) increments processed in each job. For each ($\theta$, $\varphi$) increment, the script outputs the coordinates of the Cc CMs with the smallest Cc-CcP heme-heme (runX_cmY_hemZ.pdb) and heme-W191 (runX_cmY_trpZ.pdb) distances as the b-factors. X, Y, and Z are the respective run, job, and output structure numbers. Translational ($\theta$, $\varphi$) and rotational ($\chi$, $\psi$, $\xi$) coordinates for each Cc CM are given in the captions of the output files. First 10 output files are given for bench-marking purposes.

**cm_hem.pml, cm_trp.pml**

Pymol scripts for data visualization. See Figure 3C,E in the main text for details.
Text S2. Details of the ET rate calculations.

The ET rates were calculated from Equation 8 in the main text using the $\Delta G^o$ and $\lambda$ values given below.

**Cc-CeP.** For the ET from yeast Cc to the CcP W191$^{+}$ radical, the driving force matches the reorganization energy ($-\Delta G^o \approx \lambda$), so that the ET is described by the maximum, activationless rate constant $k_{ET} \approx k_{ET,max}$ [S8]. For the ET to the CcP heme oxyferryl, $\lambda = 0.7$ eV [S9] and $\Delta G^o = -797$ mV can be calculated from the reduction potentials ($\Delta E$) of yeast Cc, $\Delta E = 290$ mV [S10], and the CcP oxyferryl, $\Delta E = 1087$ mV [S11].

**Cc-Cbs.** For the heme-heme ET between bovine Cb$_5$ and yeast Cc, $\lambda = 0.7$ eV [S12] and $\Delta G^o = -270$ mV is calculated from $\Delta E = 20$ mV for Cb$_5$ [S13] and $\Delta E = 290$ mV for Cc [S10].

**Cc-Cbc1.** The reduction potentials of yeast Cc and Cc$_1$ hemes are nearly identical (i.e. $\Delta G^o \approx 0$) [S2], while the reorganization energy is estimated to be in the range $\lambda = 0.7 - 1$ eV [S2].

**Cc-CCO.** With $\Delta E = 245$ mV for the CuA site [S14] and $\Delta E = 220$ mV for the horse Cc bound to CCO [S15], $\Delta G^o = -25$ mV. The reorganization energy for the intermolecular ET step, $\lambda$, can be calculated from the reorganization energies of Cc, $\lambda_1 \approx 0.7$ eV [S16], and the CuA site, $\lambda_2 \approx 0.4$ eV [S17], as $\lambda = (\lambda_1 + \lambda_2)/2 = 0.55$ eV.

**Cc-Fcb2.** With $\Delta E = 6$ mV for the heme b$_5$ of Fcb$_2$ [S18] and $\Delta E = 290$ mV for the yeast Cc [S10], $\Delta G^o = -284$ mV; $\lambda = 0.7$ eV as reported for the Cc-Cb$_5$ couple [S12].

**Cc-SOX.** With $\Delta E = 20$ mV for the heme group in Cb$_5$ [S13] and $\Delta E = 260$ mV for the horse Cc [S19], $\Delta G^o = -240$ mV; $\lambda = 0.7$ eV [S12].

**Cc-Erv1.** Given the values of $\Delta E = 290$ mV for the yeast Cc [S10], $\Delta E = 215$ mV for the Erv1 flavin [S20], and $\Delta E = 150$ mV for the C130-C133 disulfide in Erv1 [S20], the activation energy for the ET from the Erv1 flavin or C130-C133 disulfide is $\Delta G^o = -75$ mV or -140 mV, respectively. A range of reorganization energies ($\lambda = 0.7 - 1$ eV) was used in the calculations.
Text S3. Cc-Erv1 complex.

Erv1 is a flavin-dependent sulfhydryl oxidase that aids the import of cysteine-rich proteins in the mitochondrial intermembrane space [S20]. It accepts electrons from Mia40 – an enzyme catalyzing formation of disulfide bonds – and passes them to Cc, acting as the electron sink. Erv1 contains three redox-active centers: two disulfides, C30-C33 (distal) and C130-C133 (proximal), and a flavin adenine nucleotide (FAD). The distal cysteine pair, believed to mediate Erv1 homodimerization, was shown to be the electron entry site in ALR, a flavin-dependent sulfhydryl oxidase in humans [S21]. In ALR, the electrons are further relayed to the proximal disulfide and then the FAD group, the electron donor to Cc [S21]. A similar electron flow is expected in Erv1; however, direct ET from the proximal disulfide to Cc was also suggested [S20]. In order to investigate the ET properties of the two redox sites, we modeled the interaction of Cc with the *Saccharomyces cerevisiae* Erv1 (ScErv1).

As the ScErv1 structure is not known, we built a structural model based on the X-ray coordinates of a homologous protein from *Arabidopsis thaliana* (AtErv1). Given that the first 80 residues, constituting a flexible N-terminal segment involved in protein dimerization and interaction with Mia40, are not observed in the AtErv1 structure, the ScErv1 model was built for residues 84-183 containing the proximal disulfide and the FAD group (Figure S8A). The ET-competent Cc-Erv1 orientations are located in well-defined patches on the front face of the protein (Figure S8B), with a number of protein-protein geometries featuring short ET distances (Figure S8C). Thus, from the geometrical point of view, it appears that both the proximal disulfide and the FAD group of ScErv1 can sustain fast ET to Cc.
Text S4. Supplementary references.


Supplementary Videos

Video S1. Distribution of Cc CMs around CcP colored by the heme-W191 distances and the corresponding ET rates. For details, see the legend to Figure 3 in the main text.

Video S2. Distribution of Cc CMs around CcP colored by the heme-heme distances and the corresponding ET rates. For details, see the legend to Figure 3 in the main text.

Video S3. Distribution of Cc CMs around CcP in ternary complexes colored by the heme-W191 distances and the corresponding ET rates. See the legend to Figure S1 for details.

Video S4. Distribution of Cc CMs around CcP in ternary complexes colored by the heme-heme distances and the corresponding ET rates. See the legend to Figure S1 for details.

Video S5. Distribution of Cc CMs around Cb5 in binary complexes colored by the heme-heme distances and the corresponding ET rates. For details, see the legend to Figure 6 in the main text.

Video S6. Distribution of Cc CMs around Cb5 in ternary complexes colored by the heme-heme distances and the corresponding ET rates. For details, see the legend to Figure 6 in the main text.

Video S7. Functional epitope of Cc in the complex with Cb5. Distribution of Cb5 CMs around Cc colored by the heme-heme distances. See the legend to Figure S6 for details.

Video S8. Functional epitope of Cc in the complex with CcP. Distribution of CcP CMs around Cc colored by the heme-W191 distances. See the legend to Figure S7 for details.