PNAS
www.pnas.org

Supplementary Information for

A FERREDOXIN-LINKED FLAVOENZYME DEFINES A FAMILY OF PYRIDINE NUCLEOTIDE-INDEPENDENT THIOREDOXIN REDUCTASES

Rubén M. Buey, David Fernández-Justel, José M. de Pereda, José Luis Revuelta, Peter Schürmann, Bob B. Buchanan and Monica Balsera

Bob B. Buchanan
Email: view@berkeley.edu

Monica Balsera
Email: monica.balsera@csic.es

This PDF file includes:

Supplementary text
Figs. S1 to S9
Tables S1 to S3
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1812781115

Supplementary Materials and Methods

Gene cloning, expression and protein purification. Cysteine to serine mutants in CaTrx2 (CaTrx2C32S) and CaFFTR2 (CaFFTR2C131S) were prepared according to (1). The selected Cys belong to the redox active CxxC motif in CaTrx2 and CaFFTR2 ($C^{29}xxC^{32}$ and $C^{131}xxC^{134}$, respectively); the mutation to Ser precludes the disruption of the intermolecular disulfide formed between the catalytic Cys of the proteins (Cys^{32} and Cys^{134}, respectively) (2). All constructs were corroborated by DNA sequencing. The expression plasmids were transformed into *E. coli* Rosetta (DE3) cells and grown at 37 ºC in LB media. Gene expression was induced with 0.1 mM IPTG at OD 0.6. The cultures were incubated at 20 ºC overnight. Cells were harvested by centrifugation and the pellet suspended in 20 mM Tris-HCl pH 7.6, 300 mM NaCl and 10% glycerol. Cells were disrupted by sonication and insoluble material was separated by centrifugation. The supernatant, containing the protein of interest, was loaded on a His-Trap HP column (GE Healthcare), and after a washing step with 60 mM imidazol, protein was eluted by 500 mM imidazole. After incubation with TEV protease in buffer 20 mM Tris-HCl pH 8/500 mM NaCl/0.5 mM DTT, the solution was again loaded on a Ni-column, to remove the uncleaved protein and the TEV protease (His-tagged) itself, and the collected fractions, containing the protein of interest, were passed through a Sephacryl S-300 column (GE Healthcare) in buffer 20 mM Tris-HCl pH 8, 150 mM NaCl. FFTR proteins were incubated with an excess of FAD for 2 h before loading on the gel filtration column. The assay mixture for the enzymatic assay contained 2 mM NADPH, 1 μM spinach ferredoxin-NADP reductase (FNR; Sigma), 20 μM spinach Fdx (Sigma), 2 μM CaFFTR2 and 0.5 μM CaTrx2, in 50 mM Tris-HCl pH 7.5, 100 mM NaCl. Components were pipetted under a stream of dry nitrogen into a 96-well microtiter plate. The reaction was followed spectrophotometrically at 340 nm.

Crystallization, X-ray diffraction data collection and refinement. Crystals were transferred to Paratone-N (Hampton Research) for cryoprotection before flash freezing in liquid nitrogen. Collected X-ray diffraction data were integrated with XDS and scaled using XSCALE (3, 4). Diffraction data for the CaFFTR2-Trx2 complex were severely anisotropic and were processed and truncated anisotropically by using the software STARANISO (5) as implemented in the autoPROC pipeline (6). The STARANISO protocol produced a best resolution limit of 2.89 Å and a worst-resolution limit of 3.88 Å. Initial models were obtained for CaFFTR2 and CaTrx2 by molecular replacement using PHASER (7) with *Gloeobacter violaceus* DTR (PDB code 5j60) and an *Escherichia coli* Trx mutant (PDB code 3dyr), respectively, as search models (8, 9). All structures were iteratively refined by alternating visual inspection of the electron density maps, manual modeling with Coot (10) and automatic refinement using the Phenix crystallographic software suite

(11). Rigid body, gradient-driven positional, simulated annealing, and restrained individual isotropic B-factor and Translation–Liberation–Screw rotation (TLS) (12) were used for structure refinement. Data collection and refinement statistics are summarized in SI Appendix, Table S2. The refined structures include most of the polypeptide chains, with the exception of short, disordered loop regions in CaTrx2 in complex with CaFFTR2. All protein structures were rendered using Pymol (13)**.** The electrostatic potential surfaces were computed using the APBS (14) plug-in for Pymol. The atomic coordinates and structure factors were deposited to the Protein Data Bank (PDB codes are shown in SI Appendix, Table S1 and Table S2). Interface residues were analyzed using the PISA method (15).

Small-angle X-ray scattering. Size exclusion chromatography-SAXS (SEC-SAXS) data were collected for CaFFTR2 and the complex CaFFTR2-CaTrx2 using a Superdex S200 (GE Healthcare) column. The SEC-SAXS data were analyzed using the software Chromixs, and Crysol was used to compare the crystal structures with the experimental scattering profiles (16).

Sequence analysis of selected FFTR-related protein sequences. Blastp program (17) was used to retrieve homolog protein sequences from National Center for Biotechnology Information database. Protein multiple sequence alignment was performed with ClustalX (18) on a set of manually selected sequences.

Homology modeling and protein docking. The *Clostridium acetobutylicum* Fdx structure was predicted using the SWISS-MODEL homology modeling server (19), with the known structure of *Clostridium pasteurianum* Fdx (PDB code 1clx) as the template. The rigid-body docking step was performed using the FRODOCK program (20). FRODOCK is a fast spherical harmonics-based protein-protein interaction tool that searches for favorable binding positions to generate a number of bound conformations ranked by higher accuracy solutions. We took as input the structural models of CaFFTR2 or CaFFTR2-CaTrx2 and CaFdx. The DynDom program (21) was used to determine dynamic domain and hinge regions.

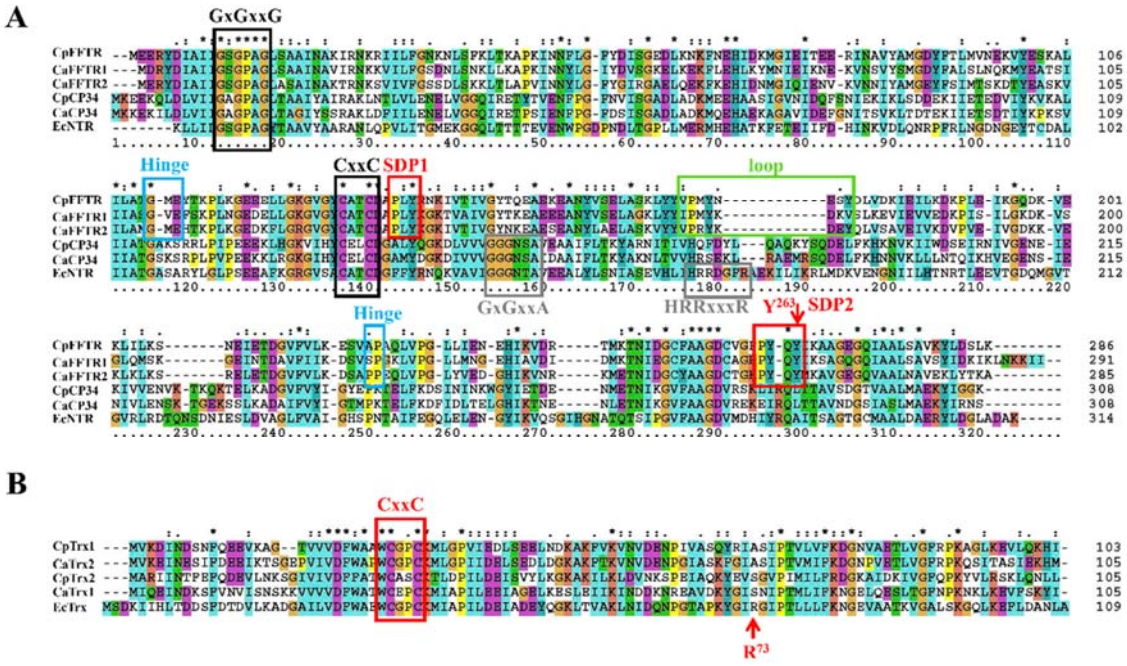

Fig. S1. Amino acid sequence comparison of *C. acetobutylicum* NTR-related proteins and Trx. (A) Protein multiple sequence alignment of *C. pasteurianum* FFTR (CpFFTR) and two forms of *C. acetobutylicum* FFTR (CaFFTR1 and CaFFTR2); CP34 from *C. pasteurianum* (CpCP34) and *C. acetobutylicum* (CaCP34); *E. coli* NTR (EcNTR). Black boxes include the conserved motif for FAD binding (GxGxxG) and the redox active Cys (CxxC). Grey boxes mark the GxGxxA motif for pyridine nucleotide binding in NTR and CP34, and HRRxxxR motif for NADPH-binding in NTR. Note that CP34 functions with NADH. Hinge regions and specificity-determinant positions (SDPs) are highlighted by blue and red boxes, respectively. The conserved Tyr residue (Y^{263}) stacked at the *si*-side of FAD is indicated with a red arrow. A loop region in FFTR sequences is marked with a green box; (B) Protein multiple sequence alignment of Trxs of *C. pasteurianum* (CpTrx1), *C. acetobutylicum* (CaTrx1) and *E. coli* (EcTrx) with the redox active WCGPC motif; *C. pasteurianum* and *C. acetobutylicum* Trx-related forms are included (CpTrx2 and CaTrx2, respectively). The redox-active motif (CxxC) of the proteins and the amino acid Arg73 of *E. coli* Trx are indicated by a red box and an arrow, respectively.

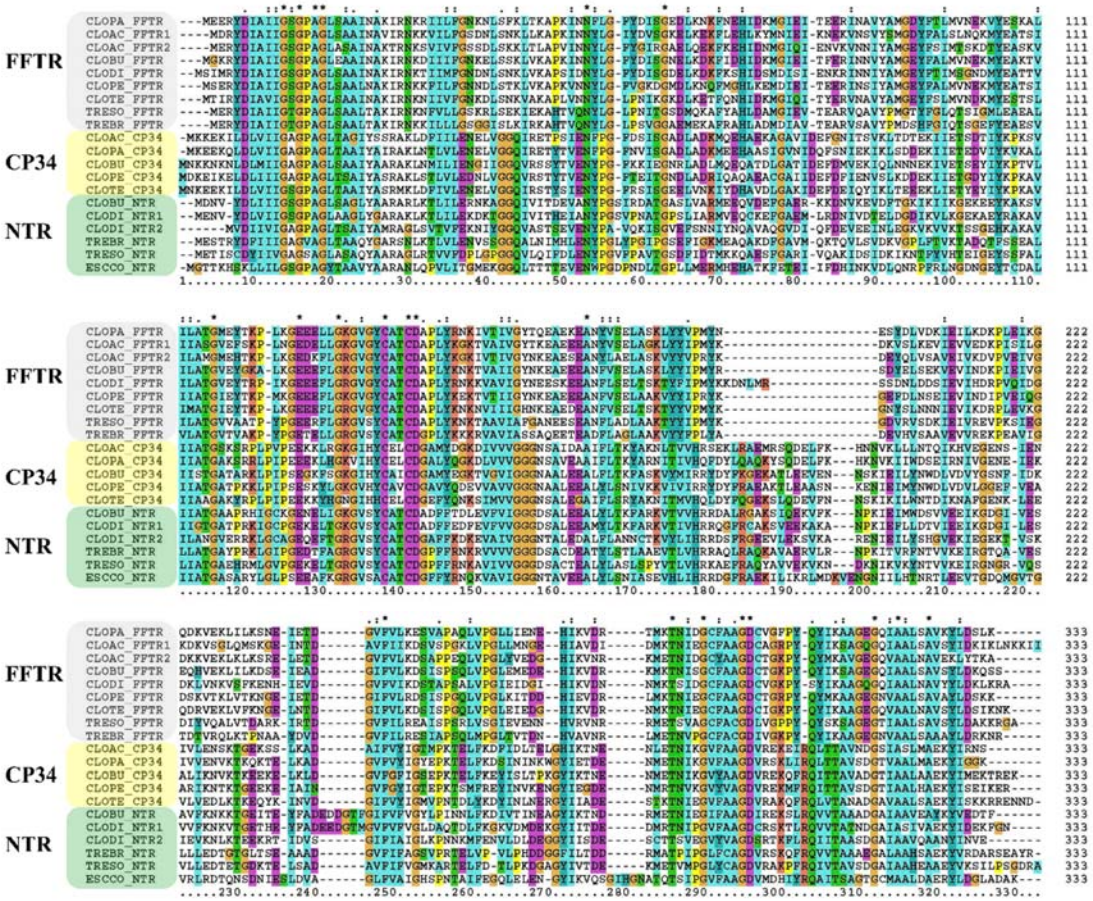

Fig. S2. Multiple protein sequence alignment of selected sequences of FFTR (grey box), CP34 (yellow box) and NTR (green box) families. In the alignment the sequences are indicated with an abbreviation of the name of the organism: CLOPA: *Clostridium pasteurianum*; CLOAC: *Clostridium acetobutylicum*; CLODI, *C. difficile*; CLOBO: *C. botulinum*; CLOTE: *C. tetani*; CLOPE: *C. perfringens*; TRESO: *Treponema socranskii*; TREBR: *Treponema brennaborense*; ESCCO: *Escherichia coli*.

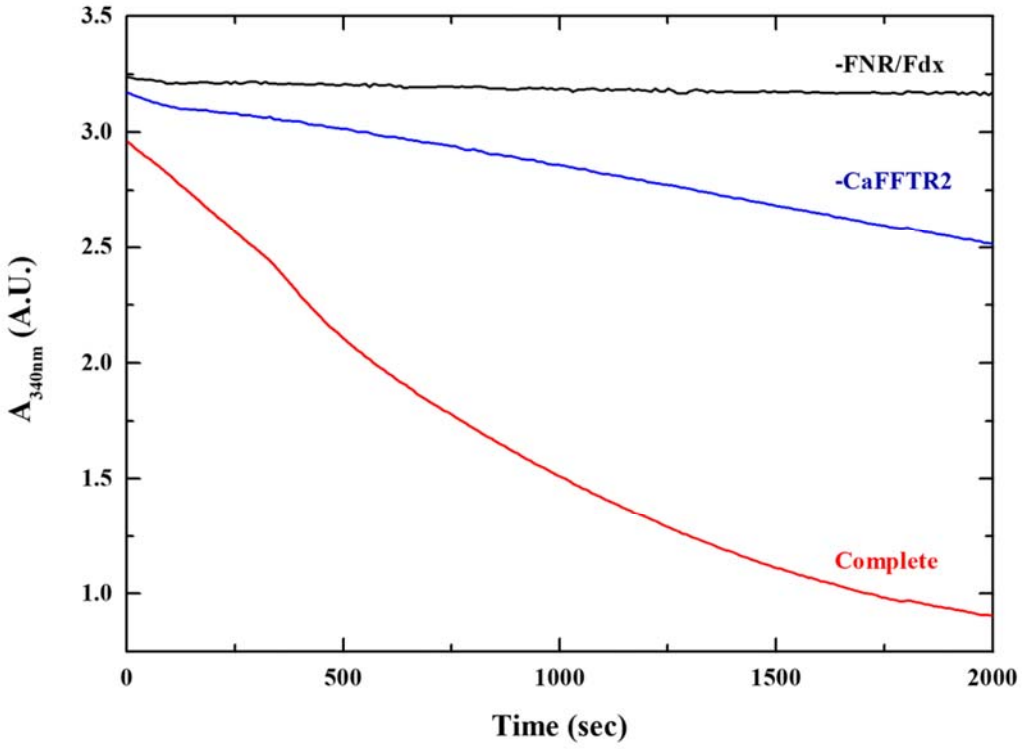

Fig. S3. Spectrometric assay of CaFFTR2 activity in an *in vitro*-reconstituted assay system. The complete mixture contained NADPH, FNR+Fdx, CaFFTR2 and CaTrx2. When indicated, one of the components was omitted in the reaction. The sample mixtures were prepared under nitrogen, and measurements were performed aerobically. As a consequence, part of the reducing equivalents is bypassed to oxygen as described in (22). The highest activity is detected when all components are present in the assay mixture (red line).

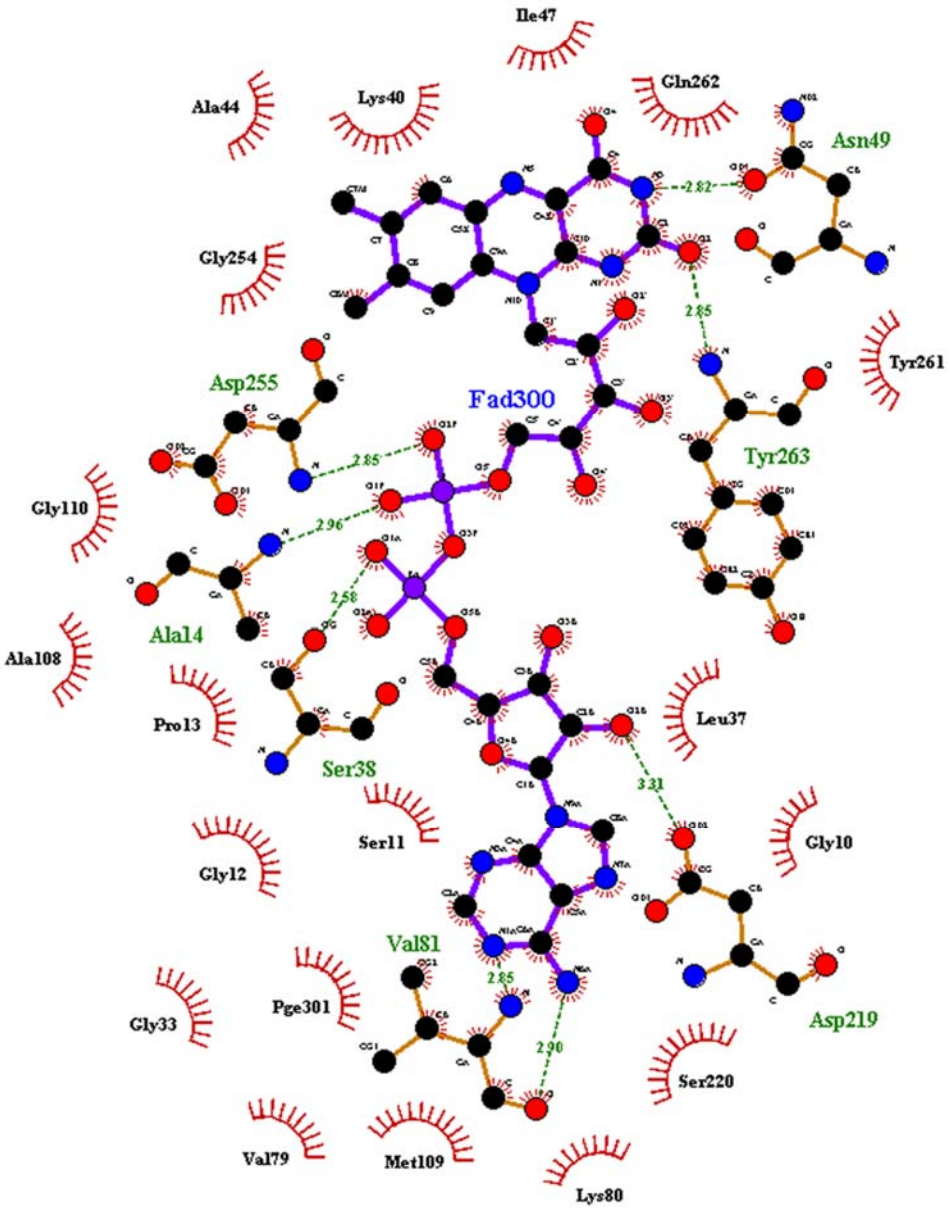

Fig. S4. Residues involved in the interaction with FAD. Hydrophobic interactions are represented by spoked arcs and hydrogen bonds are indicated by dashed green lines.

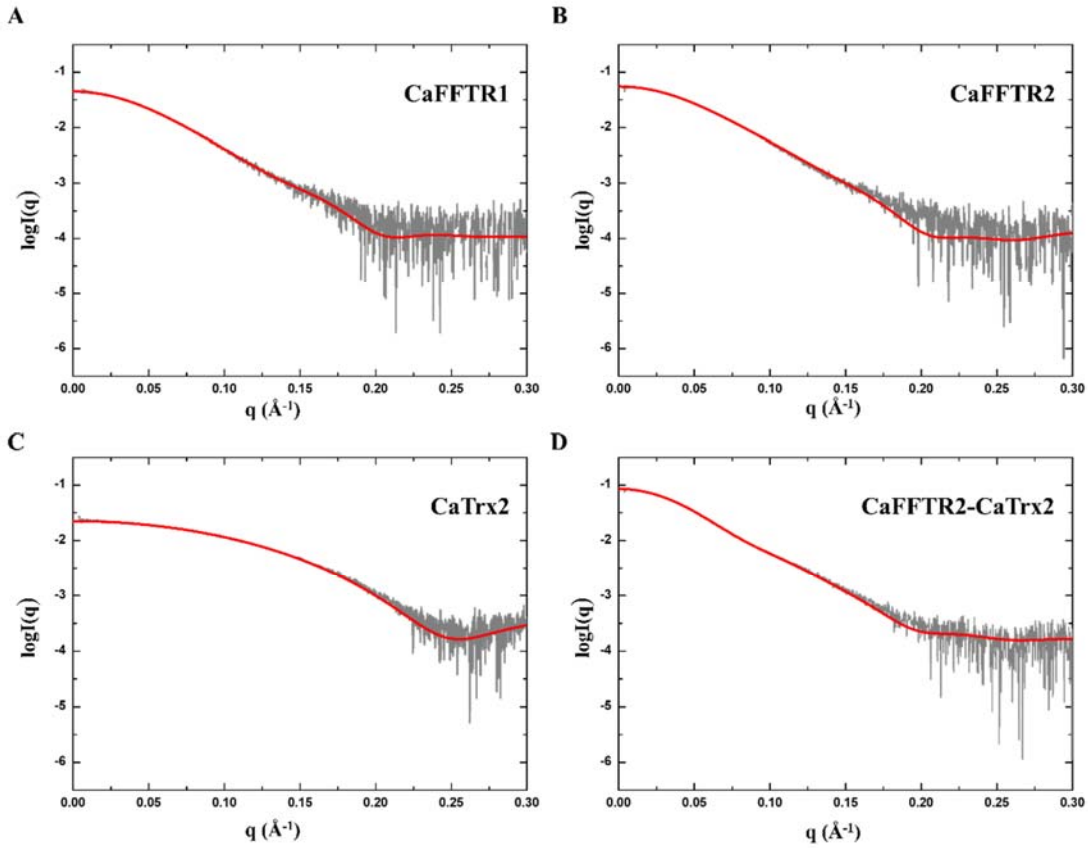

Fig. S5. Fitting of the experimental SAXS data (grey lines) of (A) CaFFTR1, (B) CaFFTR2, (C) CaTrx2, and (C) CaFFTR2-CaTrx2 proteins on the theoretical scattering curves (red lines), calculated by Crysol (16) from the crystallographic structures. A good agreement between the experimental and theoretical SAXS profiles can be observed for all the studied proteins, especially in the low and middle range of the scattering vector values (up to 0.15 $Å^{-1}$), which contains most of the low-resolution information on the protein shape and conformational transitions in solution (23).

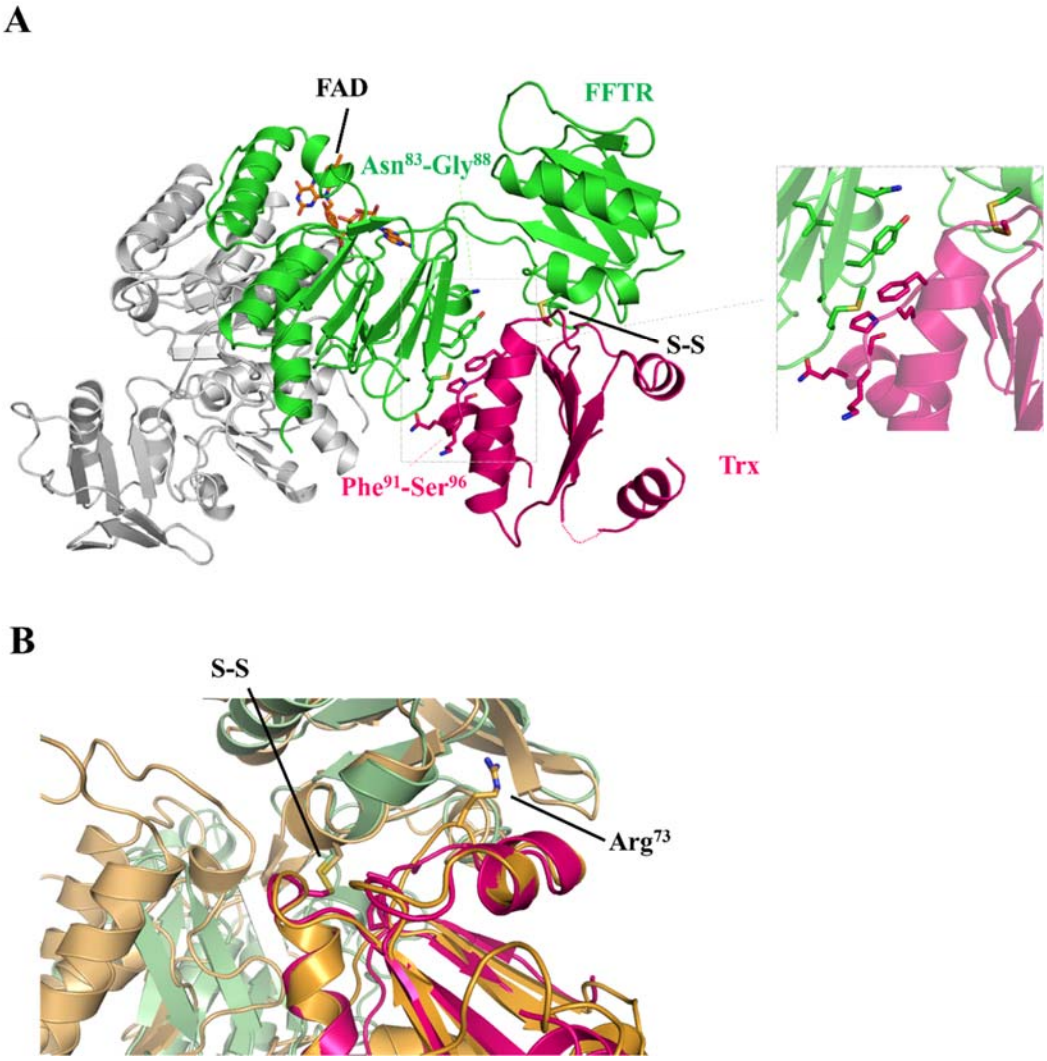

Fig. S6. FFTR-Trx complex formation. (A) Trx interaction with the FAD domain of FFTR. FFTR and Trx structures are represented in ribbon diagrams in green and magenta colors, respectively. Residues at the interface of the FAD domain of FFTR and Trx are shown in sticks; on the right, the interaction interface is zoomed. The second monomer of FFTR is colored in grey; (B) Zoom-in of the interaction region between CaFFTR2 and CaTrx2, and EcNTR and EcTrx (PDB code 1f6m) in the protein complexes. For the structural alignment, the coordinates of the Trx molecules (shown in magenta for CaTrx2 and orange for EcTrx) were fixed. FFTR and NTR are shown as ribbon diagrams in light green and light orange colors, respectively. Cys of the CxxC motifs and the Arg73 of *E. coli* Trx that is determinant for EcNTR-EcTrx interaction are shown in sticks representation.

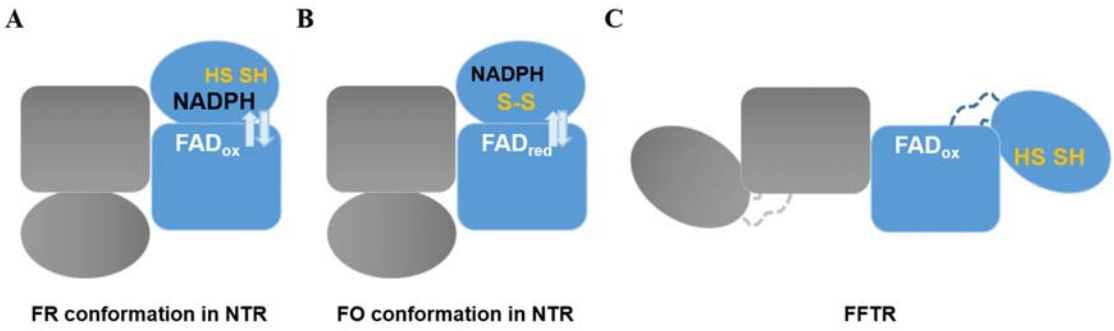

Fig. S7. Schematic representations of NTR and FFTR homodimers. (A) Flavin-reducing (FR) conformation of NTR, where FAD is in contact with NADPH; (B) Flavin-oxidizing (FO) conformation of NTR, where FAD is in contact with the redox active S-S; (C) FFTR conformation obtained in this work, where FAD is exposed to the solvent. Cofactors and redox active groups of only one of the monomers are shown. The two arrows in (A) and (B) indicate the two β-strands between domains in NTRs. Dashed lines in (C) represent the two extended loops connecting domains in FFTR.

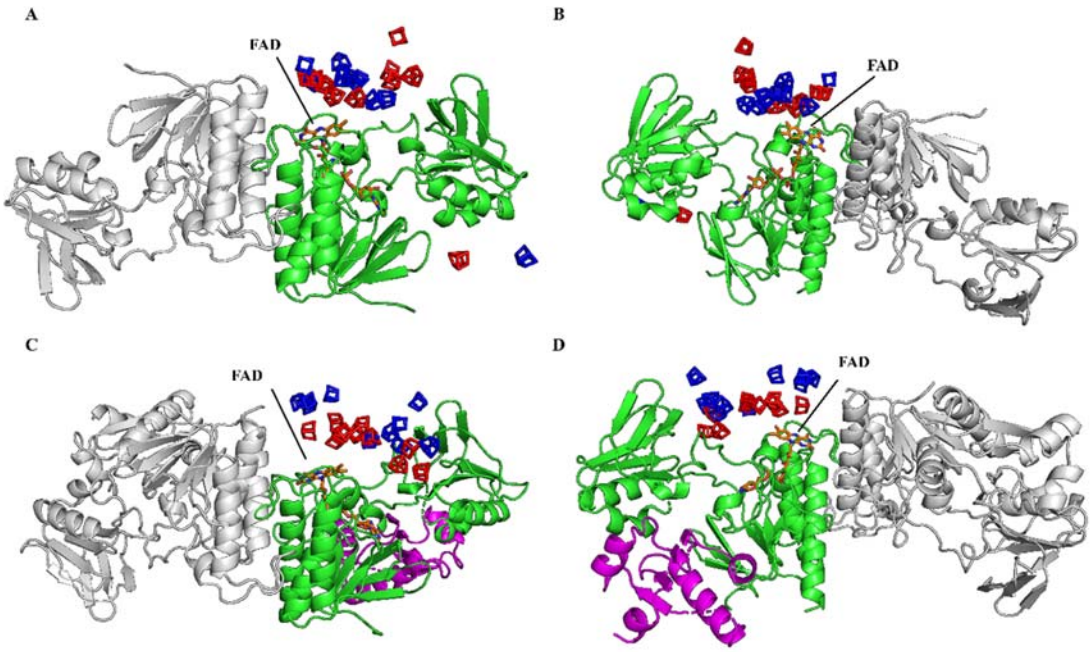

Fig. S8. Top 10 predictions of the interaction between CaFFTR2 and CaFdx selected by the molecular docking tool FRODOCK (24). (A and B) CaFFTR2 and CaFdx interaction; (C and D) CaFFTR2-CaTrx2 complex and CaFdx interaction. For the sake of simplicity, only the 2[4Fe-4S] clusters of the Fdx have been included, in red and blue colors, respectively.

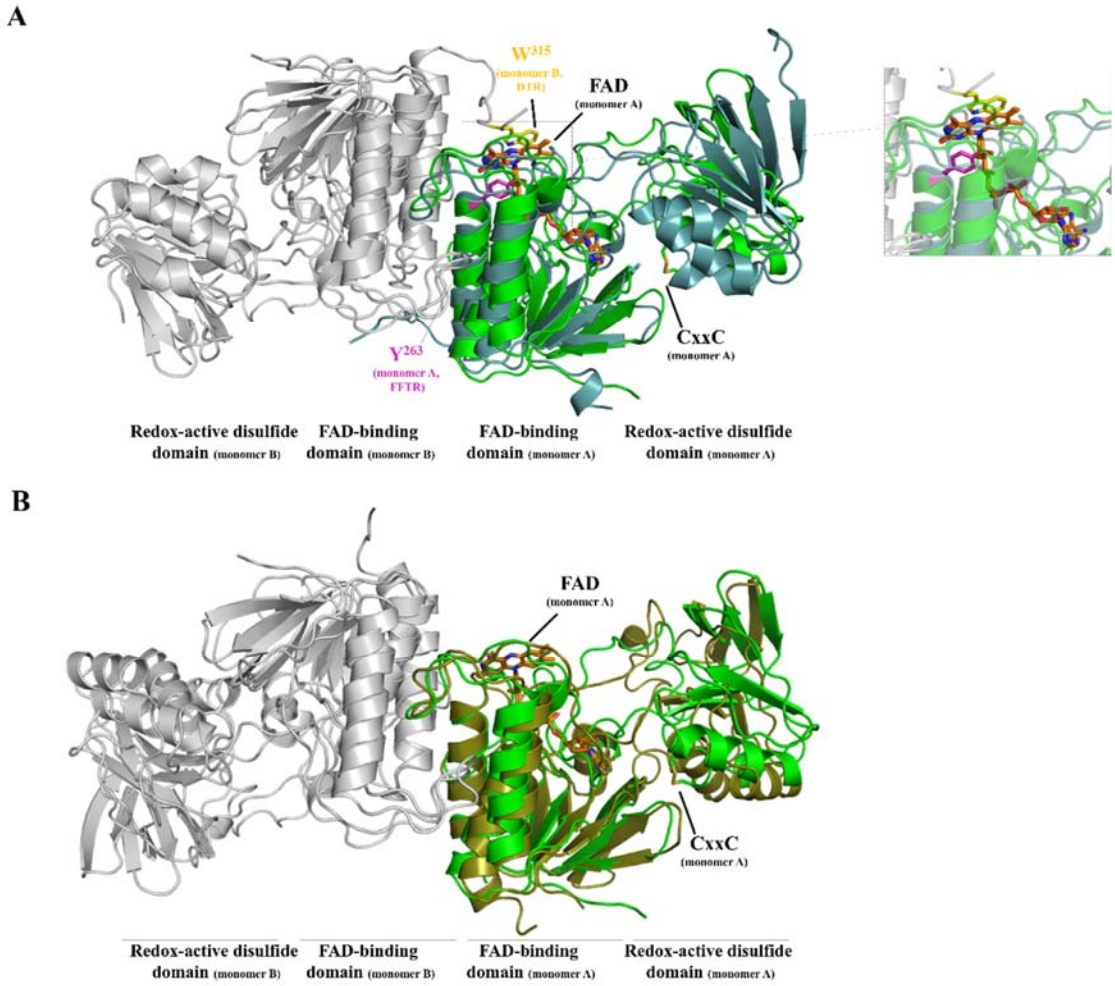

Fig. S9. Structural alignment of CaFFTR2 (green) and (A) *Gloeobacter violaceus* DTR (teal; PDB code 5J60), and (B) *Desulfovibrio vulgaris* TRi (olive; PDB code 5nii) homodimers. For the alignment the coordinates of the FAD cofactor (carbons in orange) were fixed. The aromatic amino acids W^{315} (yellow) and Y^{263} (violet), which stacks over FAD in FFTR and DTR, respectively, are displayed in sticks. A zoom-in view of the interaction is presented at a right panel in (A).

Table S1. Experimental crystallization conditions.

Protein	PDB code	Crystallization conditions
CaFFTR1	6GNC	0.2 M calcium chloride, 0.1 m HEPES pH 7.5, 28% (v/v) PEG-400
CaTrx2	6GN9	0.2 M calcium acetate, 0.1 M sodium cacodylate pH 6.5, 40% (v/v) PEG-600
CaFFTR2	6GNA	0.2 M sodium chloride, 0.1 M sodium acetate pH 4.5, 40% (v/v) PEG-300
CaFFTR2	6GNB	0.2 M sodium thiocyanate, 0.1 M HEPES pH 7.0, 15% (v/v) pentaerythritol propoxylate (5/4 PO/OH)
CaFFTR2-Trx2	6GND	0.02 M DL-glutamic acid monohydrate, 0.02 M DL-alanine, 0.02 M glycine, 0.02 M DL-lysine monohydrochloride, 0.02 M DL-serine, 0.1M imidazole/MES monohydrate pH 6.5, 20 % (v/v) ethylene glycol, 10% (w/v) PEG-8000

Table S2. Data collection and refinement statistics.

	CaTrx2	CaFFTR2	CaFFTR2	CaFFTR1	CaFFTR2-Trx2
Diffraction data statistics					
Wavelength (Å)	1.00001	0.97926	0.97923	0.97918	0.97918
Resolution range (Å)	37.89 - 1.75 (1.81 - 1.75)	40.42 - 1.29 (1.34 - 1.29)	42.93 - 1.90 (1.96 - 1.90	53.34 - 1.64 (1.70 - 1.64)	87.18 - 2.89 (2.99 - 2.89)
Space group	P 2 2_1 2_1	C 2	C 2	C 2 2 2	C 2
Unit cell (Å/º)	30.03 42.39 84.52 90 90 90	83.22 40.94 81.09 90 103.76 90	111.73 41.11 129.16 90 94.37 90	80.81 175.78 67.11 90 90 90	115.23 174.35 114.30 90 119.80 90
Total reflections	141913 (14005)	433369 (41289)	309769 (31269)	781222 (77493)	140615 (8897)
Unique reflections	11442 (1111)	64697 (6246)	46948 (4604)	58960 (5784)	27743 (1371)
Multiplicity	12.4 (12.6)	6.7 (6.6)	6.6 (6.8)	13.3 (13.4)	5.1 (6.5)
Completeness (%)	99.82 (99.19)	97.58 (94.93)	99.59 (98.56)	99.87 (99.52)	92.70 (75.00)¶
Mean I/sigma(I)	15.95 (1.02)	16.88 (1.04)	10.71 (1.17)	34.62 (2.95)	14.10 (1.50)
R-pim	0.0267 (0.691)	0.0179 (0.750)	0.0452 (0.613)	0.010 (0.273)	0.033 (0.489)
CC1/2	0.999 (0.398)	0.999 (0.578)	0.996 (0.504)	1 (0.943)	0.999 (0.609)
Refinement statistics					
Resolution range (Å)	37.89 - 1.75	40.42 - 1.295	42.93 - 1.895	53.34 - 1.639	87.18 - 2.889
R-work	0.21 (0.34)	0.19 (0.57)	0.20 (0.35)	0.22 (0.32)	0.24 (0.57)
R-free	0.24 (0.32)	0.21 (0.54)	0.22 (0.34)	0.23 (0.35)	0.28 (0.53)
Non-H atoms	813	2568	4780	2425	10678
macromolecules	755	2181	4299	2106	10390
ligands	16	102	153	67	220
solvent	42	285	328	252	68

Protein residues	102	284	567	287	1479
RMS					
bonds	0.005	0.008	0.016	0.011	0.013
angles	1.29	1.08	1.08	1.55	1.31
Ramachandran (%)					
favored	98.98	96.79	97.16	96.84	96.33
allowed	1.02	2.50	2.66	3.16	3.26
outliers	0.00	0.71	0.18	0.00	0.42
Rotamer outliers (%)	2.60	0.44	1.14	0.00	0.66
Average B-factor	41.74	33.79	42.10	50.83	84.77
macromolecules	41.07	32.50	42.25	50.44	85.60
ligands	67.75	34.27	36.16	46.78	58.40
solvent	43.80	43.47	42.99	55.22	44.08
PDB code	6GN9	6GNA	6GNB	6GNC	6GND

Statistics for the highest-resolution shell are shown in parentheses. 5% of reflections were used for calculation of R_{free}.
¶ Values shown in the table for FFTR2-Trx2 correspond to STARANISO ellipsoidal completeness. Spherical completeness values are 63.30 (12.60).

Table S3. Clostridia strains naturally producing butanol (25).

Strains	**FFTR**	**NTR**
Clostridium acetobutylicum	+	-
Clostridium aurantibutyricum	+	-
Clostridium beijerinckii	+	+
Clostridium butyricum	+	-
Clostridium cadavaris	+	+
Clostridium carboxidiverans	+	+
Clostridium chauvoei	+	-
Clostridium felsineum	+	-
Clostridium pasteurianum	+	-
Clostridium puniceum	+	+
Clostridium roseum	+	-
Clostridium sachharobutylicum	+	-
Clostridium sachharoperbutylacetonium	+	-
Clostridium septicum	+	-
Clostridium sporogenes	+	+
Clostridium tetanomorphum	+	-

References

1. Li J, Li C, Xiao W, Yuan D, Wan G, Ma L (2008) Site-directed mutagenesis by combination of homologous recombination and DpnI digestion of the plasmid template in *Escherichia coli*. *Anal Biochem* 373(2):389-391.
2. Lennon BW, Williams CH, Jr. , Ludwig ML (2000) Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase. *Science* 289(5482):1190-1194.
3. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Cryst D* 66(2):133-144.
4. Kabsch W (2010) XDS. *Acta Cryst D* 66(2):125-132.
5. Tickle IJ, Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. (2018) STARANISO (Global Phasing Ltd., Cambridge, United Kingdom).
6. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T, Bricogne G (2011) Data processing and analysis with the *autoPROC* toolbox. *Acta Cryst D* 67(4):293-302.
7. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* 40(4):658-674.
8. Buey RM, Galindo-Trigo S, López-Maury L, Velázquez-Campoy A, Revuelta JL, Florencio FJ, de Pereda JM, Schürmann P, Buchanan BB, Balsera M (2017) A new member of the thioredoxin reductase family from early oxygenic photosynthetic organisms. *Mol Plant* 10(1):212-215.
9. Ren G, Stephan D, Xu Z, Zheng Y, Tang D, Harrison RS, Kurz M, Jarrott R, Shouldice SR, Hiniker A, Martin JL, Heras B, Bardwell JCA (2009) Properties of the thioredoxin fold superfamily are modulated by a single amino acid residue. *J Biol Chem* 284(15):10150-10159.
10. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Cryst D* 66(4):486-501.
11. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* 66:213-221.
12. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* 57(Pt 1):122-133.
13. The PyMOL Molecular Graphics System (2015) (Schrödinger, LLC), Version 1.8.
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037-10041.
15. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774-797.
16. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Cryst* 50(4):1212-1225.
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
18. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24):4876-4882.

19. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31(13):3381-3385.
20. Ramírez-Aportela E, López-Blanco JR, Chacón P (2016) FRODOCK 2.0: fast protein–protein docking server. *Bioinformatics* 32(15):2386-2388.
21. Hayward S, Berendsen HJC (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins: Structure, Function, and Bioinformatics* 30(2):144-154.
23. Yonekura-Sakakibara K*, et al.* (2000) Analysis of reductant supply systems for ferredoxin-dependent sulfite reductase in photosynthetic and nonphotosynthetic organs of maize. *Plant Physiol.* 122(3):887-894.
24. Putnam CD, Hammel M, Hura GL, & Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40(3):191-285.
25. Ramírez-Aportela E, López-Blanco JR, & Chacón P (2016) FRODOCK 2.0: fast protein–protein docking server. *Bioinformatics* 32(15):2386-2388.
26. Schiel-Bengelsdorf B, Montoya J, Linder S, & Dürre P (2013) Butanol fermentation. *Environ. Technol.* 34(13-14):1691-1710.