Supplementary Materials for

Structure of the STRA6 receptor for retinol uptake

Yunting Chen†, Oliver B. Clarke†, Jonathan Kim, Sean Stowe, Youn-Kyung Kim, Zahra Assur, Michael Cavalier, Raquel Godoy-Ruiz, Desiree C. von Alpen, Chiara Manzini, William S. Blaner, Joachim Frank, Loredana Quadro, David J. Weber, Lawrence Shapiro, Wayne A. Hendrickson and Filippo Mancia

correspondence to: fm123@cumc.columbia.edu

This PDF file includes:

Figs. S1 to S14
Tables S1 to S3
Captions for Movies S1 to S3

Other Supplementary Materials for this manuscript includes the following:

Movies S1 to S3

Table S1: Data collection and validation statistics for STRA6-CaM cryo-EM structure.

Dataset	***Danio rerio* STRA6-CaM**
Pixel size (Å)	1.255
Number of grids	2
Number of micrographs (total)	2613
Number of micrographs (final)	1657
Number of picked particles	317416
Final number of particles	56615
Box size (pixels)	200
Symmetry	C2
Resolution, overall	3.9
Est. accuracy rotations (degrees)	2.49
Est. accuracy translations (pixels)	0.72
Applied B-factor (sharpening)	-113.68
Molprobity clashscore	5.23
Ramachandran favored (%)	94.03
Ramachandran allowed (%)	5.28
Ramachandran outliers (%)	0.69
Rotamer outliers (%)	0.16

Table S2: Data collection and refinement statistics for CaM-CaMBP2 complex crystal structure. Values in the highest resolution shell (1.801-1.739) are given in parentheses.

Wavelength	0.97919
Resolution range	25.37 - 1.739 (1.801 - 1.739)
Space group	C 1 2 1
Unit cell	156.5, 37.2, 35.7; 90, 94.4, 90
Total reflections	67579 (3076)
Unique reflections	21134 (2084)
Multiplicity	3.2 (2.6)
Completeness (%)	98.8 (97.8)
Mean I/sigma(I)	6.7 (1.2)
Wilson B-factor	18.29
R-merge	0.095 (0.658)
R-meas	0.131 (0.882)
CC1/2	0.986 (0.609)
Reflections used in refinement	21098 (2083)
Reflections used for R-free	1026 (115)
R-work	0.2279 (0.3448)
R-free	0.2672 (0.4032)
Number of non-hydrogen atoms:	1504
macromolecules	1364
ligands	10
Protein residues	172
RMS(bonds)	0.007
RMS(angles)	0.88
Ramachandran favored (%)	100
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Molprobity clashscore	2.97
Average B-factor:	32.35
macromolecules	32.06
ligands	30.25
solvent	35.52
Number of TLS groups	5

Table S3: STRA6 MWS missense mutations

Mutation	zSTRA equivalent position	Structural element	Reference
P90L	P77	TM1TM2L	Pasutto, F. *et al.,* Am. J. Human Gen. 2007
P293L	P272	JM1TM6L	Pasutto, F. *et al.,* Am. J. Human Gen. 2007
T321P	T300	TM6	Pasutto, F. *et al.,* Am. J. Human Gen. 2007
T644M	T613	CaMBP2	Pasutto, F. *et al.,* Am. J. Human Gen. 2007
R655C	R626	αCT	Pasutto, F. *et al.,* Am. J. Human Gen. 2007
G217E	A194	TM5	White, T. *et al.,* Mol. Vis. 2008
Q438R	Q417	TM8	Chassaing, N. *et al.,* Human Mut. 2009
R638P	R609	CaMBP2	Chassaing, N. *et al.,* Human Mut. 2009
D560H/R655H	D539/R626	JML/ αCT	Segel, R. *et al.,* Am. J. Med. Gen. 2009
G304K	F283	TM6	Casey, J. *et al.,* Human Mut. 2011
R408Q	R387	TM7TM8L	Slavotinek, A.M. *et al.,* Clin. Genet. 2015
G339S	G318	TM6TM7L (lid helix)	Kawaguchi, R., *et al.,* J. Biol. Chem. 2008

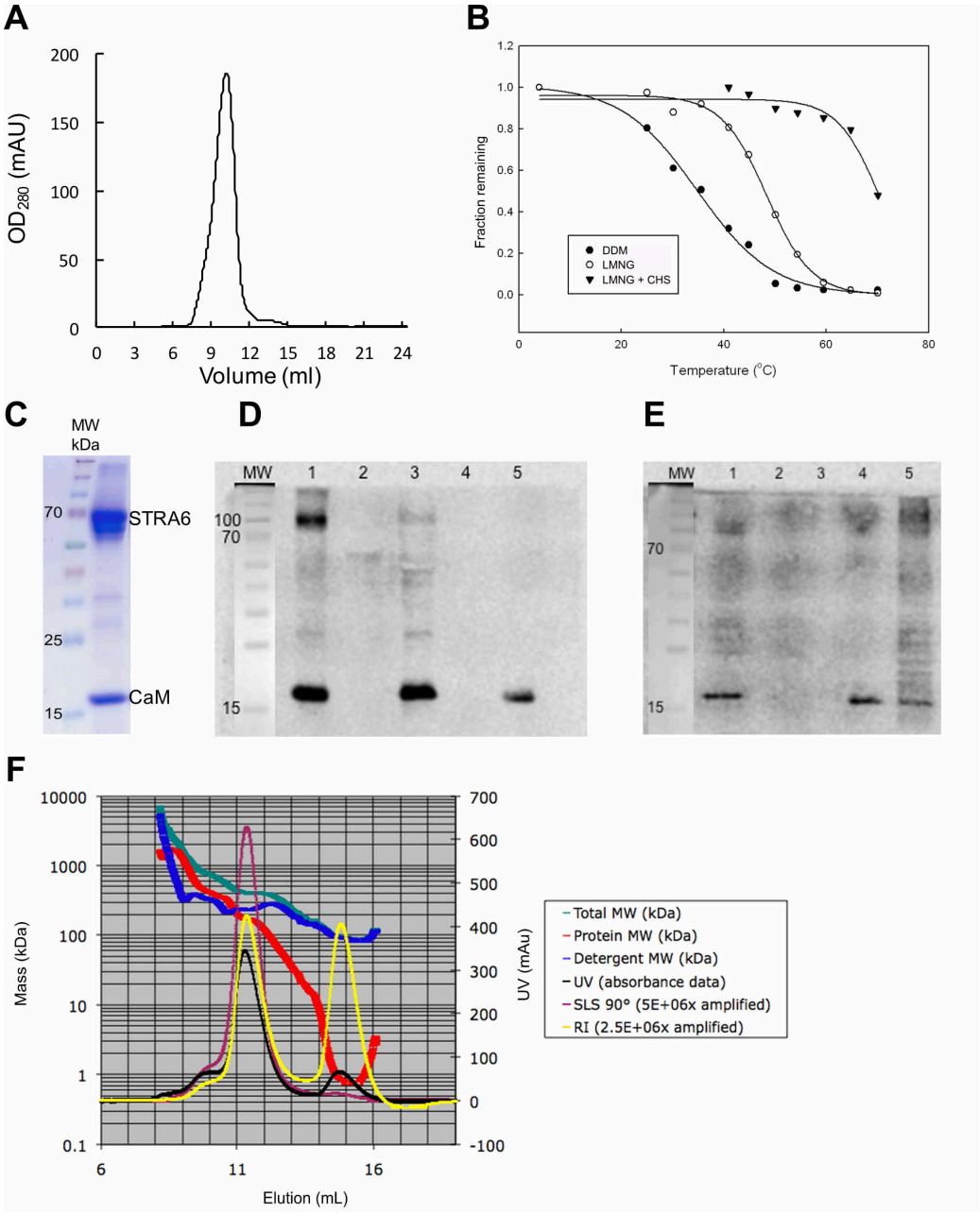


Fig. S1. Purification and biochemical characterization of zebrafish STRA6. **(A)** Elution profile from size exclusion chromatography of STRA6 purified in buffer containing LMNG and CHS. **(B)** Denaturation curves of STRA6 purified in βDDM (black circles), LMNG (MNG, white circles) and LMNG/CHS (MNG+CHS, black triangles). The fraction of folded protein remaining is plotted against a control sample maintained at 4 °C. **(C)** Purified STRA6 analyzed on SDS-PAGE denaturing gel, stained by Coomassie blue. **(D)** Immunoblot of STRA6-CaM complex purified from transiently transfected HEK cells. Probed with anti-CaM antibody. Lanes as follows: (1): zebrafish STRA6-CaM purified from insect cells and reconstituted in amphipol; (2): Pulldown from HEK cells transfected with control plasmid pFM1.2; (3): Pulldown from HEK cells transfected with expression plasmid for zebrafish STRA6; (4): Empty; (5):

Purified recombinant human CaM (5ng). **(E)** Immunoblot of STRA6-CaM complex immunoprecipitated using an anti-STRA6 antibody from native zebrafish tissue. Probed with anti-CaM antibody. Lanes as follows: (1): zebrafish STRA6 immunoprecipitated using antibodies against the recombinant protein purified from insect celle; (2): Immunoprecipitation using control antibody; (3): Empty; (4) Purified recombinant human CaM; (5): Supernatant from immunoprecipitation (5µg total protein). **(F)** Oligomerization state of STRA6 in a solution containing βDDM, determined by multi angle light scattering coupled to refractive index and UV measurements. The estimated molecular weight of the protein component is ~200 kDa (left peak), consistent with STRA6 being a dimer in solution associated with two molecules of CaM. The peak to the right corresponds to excess βDDM in micelles of ~100 kDa.

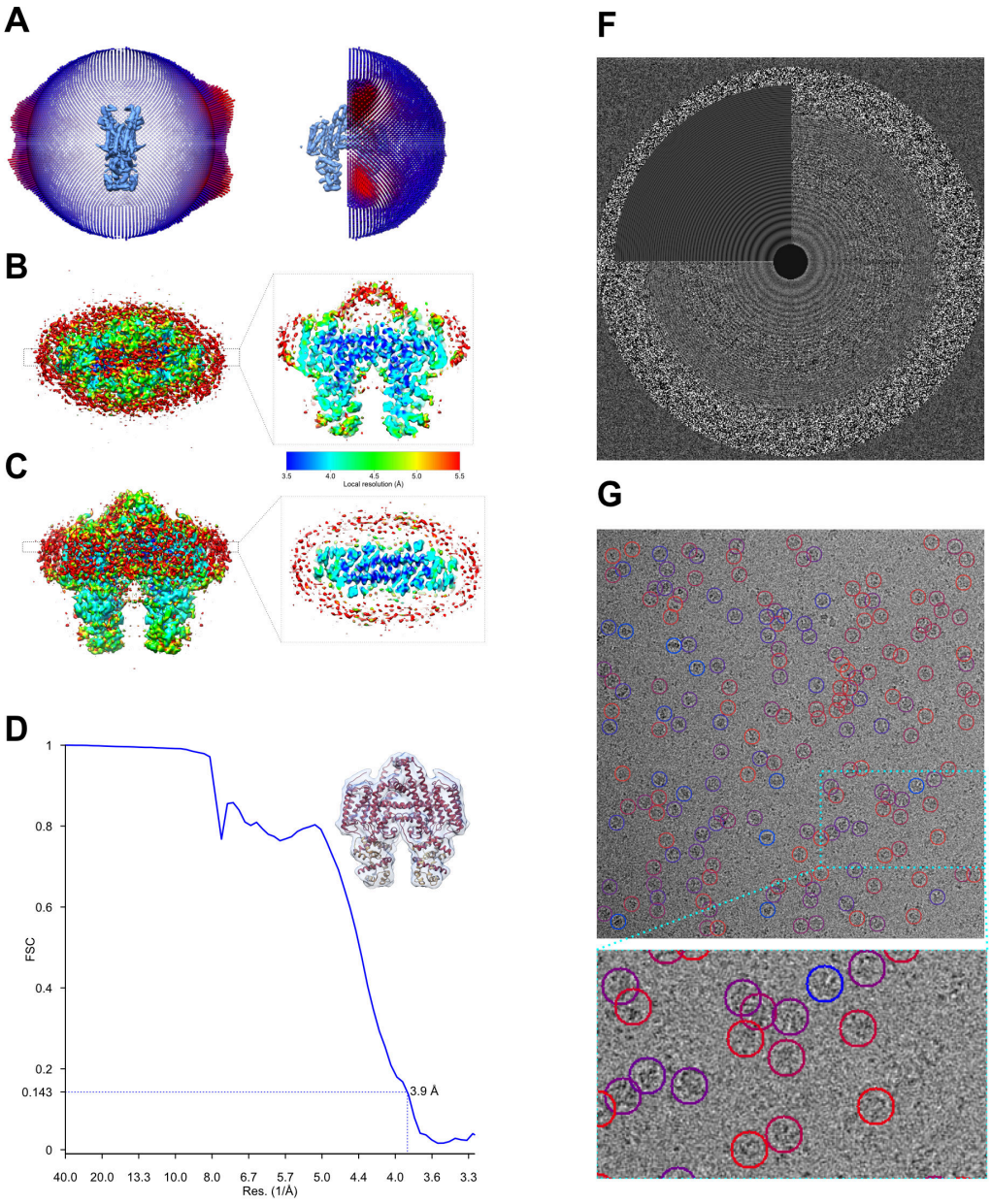

A
B
C
3.5
4.0
4.5
5.0
5.5
Local resolution (Å)
D
FSC
1
0.8
0.6
0.4
0.2
0.143
0
3.9 Å
40.0
20.0
13.3
10.0
8.0
6.7
5.7
5.0
4.4
4.0
3.6
3.3
Res. (1/Å)
E
F
G

Fig. S2. Cryo-EM analysis of *Danio rerio* STRA6. (A) Orientation distribution of particles in the final C2-symmetric reconstruction. The reconstruction is positioned in the center of a sphere as reference, and a histogram displaying the frequency of each unique orientation is represented on the surface of the sphere, with the length of each cylindrical bar proportional to the population size of particles in the final reconstruction with that angular orientation. **(B)** Final reconstruction colored by local resolution, calculated using ResMap (*68*), viewed from the extracellular region. A central vertical slab through the TM region is represented at right. **(C)** As B, viewed parallel to the membrane, with a slab taken through the TM region displayed in the inset. **(D)** Fourier Shell Correlation plot of the final refined reconstruction. The resolution at which the curve drops below 0.143 is indicated, and a thumbnail of the mask used for FSC calculation is inset, overlayed on the atomic model. **(E)** Four representative 2D class averages (upper panels) with maximum intensity projections of the final reconstruction in matching orientations (lower panels) **(F)** Power spectrum of a representative stack, calculated as an average of the power spectra of the frames in the stack with a window size of 3 frames. The contrast transfer function, fit using CTFFIND4 to a resolution of 3.0 Å, is displayed in the upper left quadrant. **(G)** Representative micrograph with particles picked using *relion_autopick* (prior to 2D and 3D classification) circled and colored by the autopick figure of merit. A magnified view of an indicated section of the micrograph is represented below the main panel.

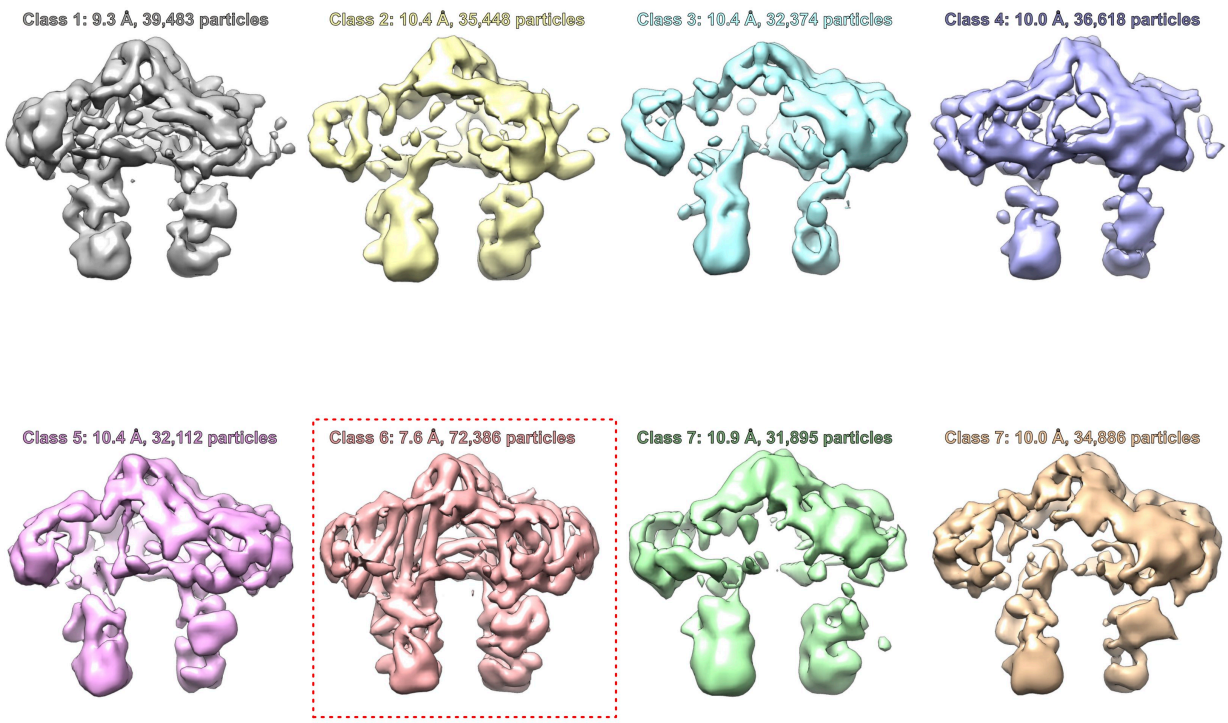


Fig. S3. Analysis of heterogeneity by 3D classification. The eight classes obtained from 3D classification with Relion 1.3 are viewed parallel to the membrane with the resolution where the SSNR of the class drops below 2.0, as well as the number of particles in each class, labeled above each panel. Class 6, which was chosen for high-resolution refinement, is enclosed within a dashed red box.

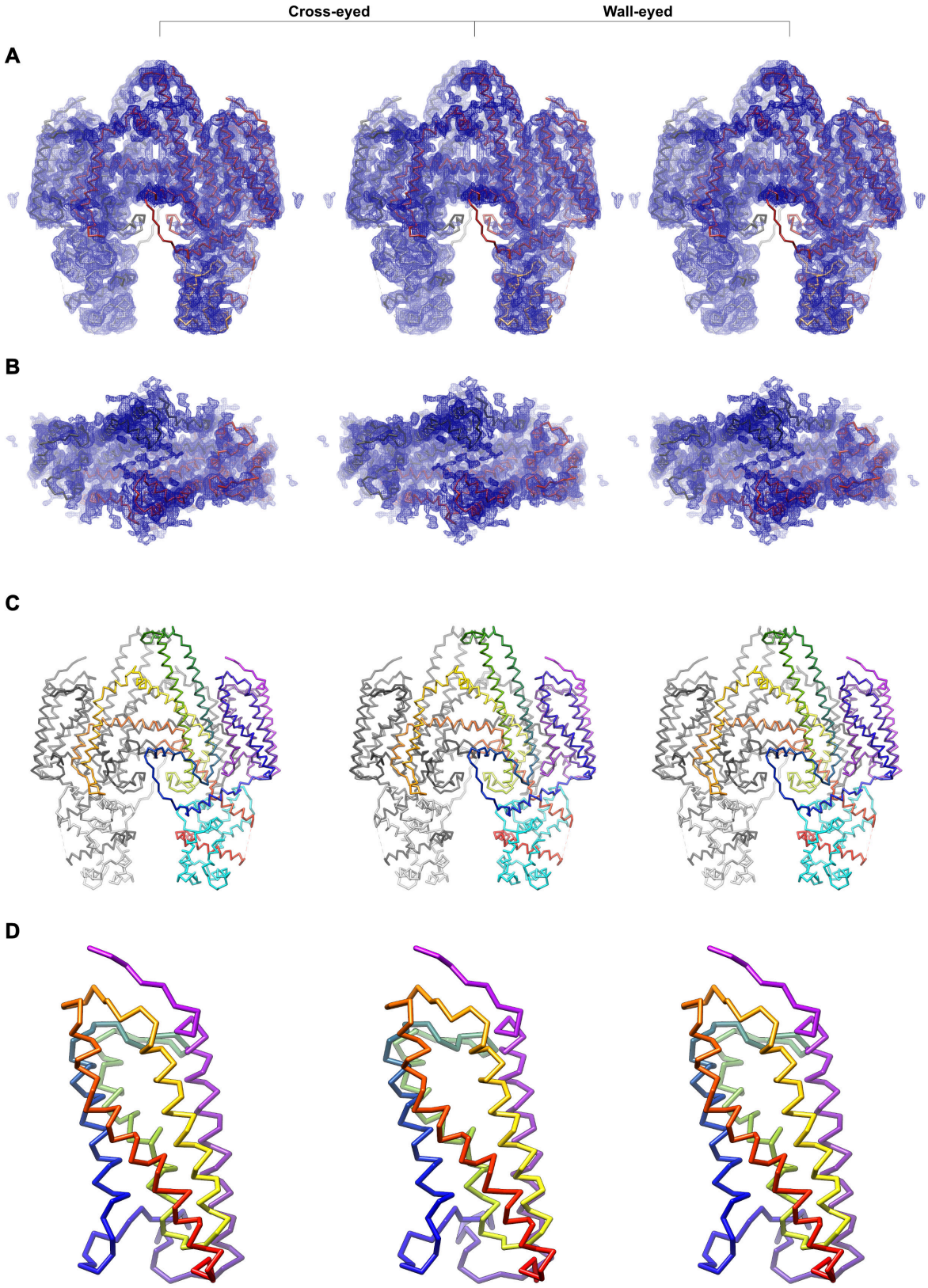


Fig. S4. STRA6 fold and global quality of the cryo-EM density map. All panels are represented in stereo, as indicated. All density maps are contoured at 6

X R.M.S. **(A** & **B)** Two views of the density maps are shown, together with the Cα trace of the entire structure, in which one STRA6 protomer is displayed in dark red, and one CaM in yellow. **(C)** STRA6-CaM is depicted in licorice representation. One protomer of STRA6 is represented in spectral coloring, from violet N-terminus to red C-terminus. One protomer of CaM is colored cyan. **(D)** The TM1-TM5 bundle (residues 31-209) is depicted in licorice representation, in spectral coloring from violet (31) to red (209). The viewpoint is parallel to the membrane and to the IM helix axis. The N-terminal segment of STRA6, preceding TM1 is adjacent to the TM4-5 loop. The relative position of these two segments is consistent with the presence of a predicted disulfide bond between C31 and C171 (*15*). A distinctive feature of the NTD is that TM3 has a two-proline kink in the middle of the membrane, separating it into two segments (TM3a and TM3b) with an interhelical angle of 17°

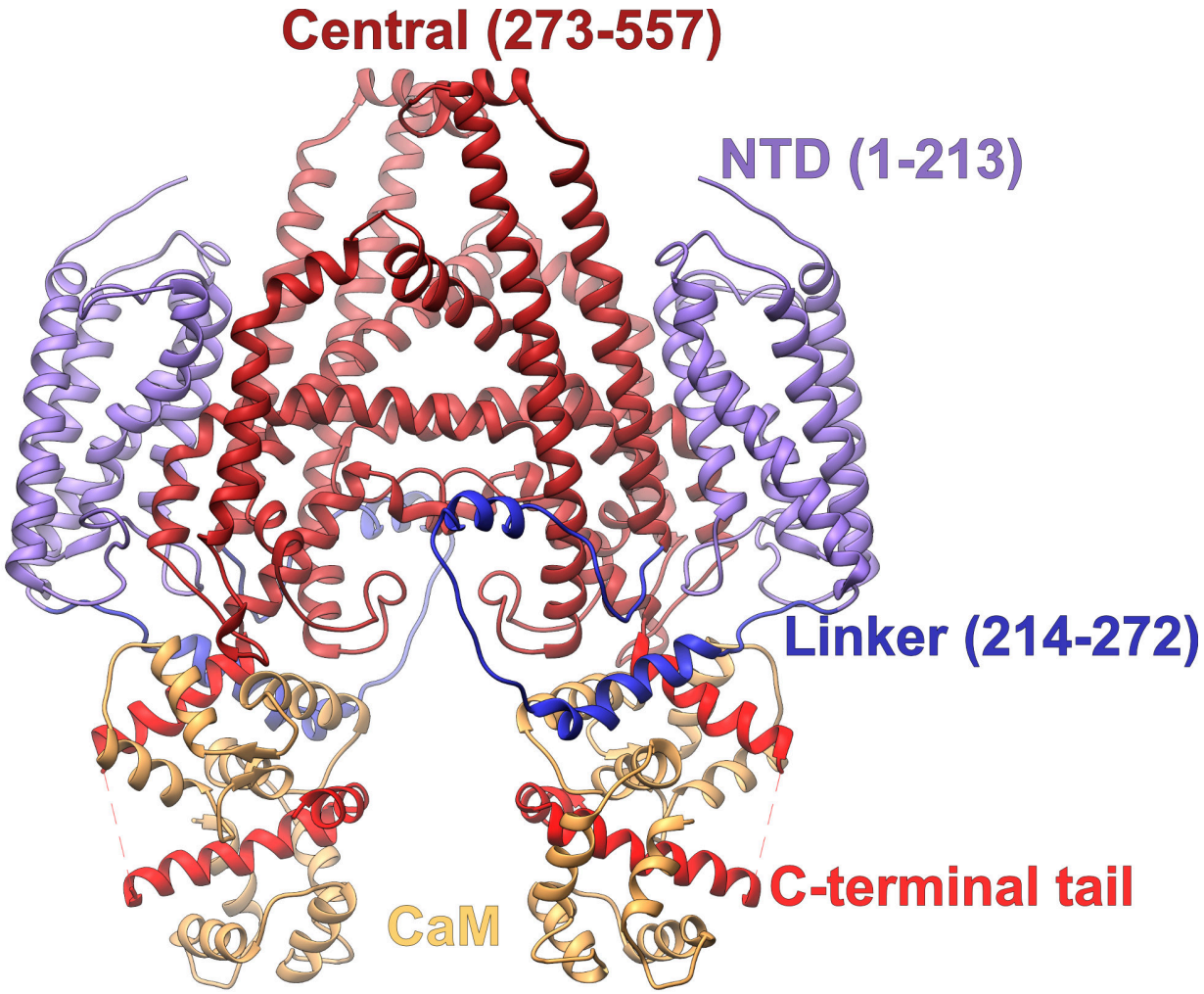


Fig. S5 Domain representation of the STRA6-CaM complex. Ribbon representation of the STRA6-CaM complex colored by domain, and labeled. Residue ranges are provided in parenthesis.

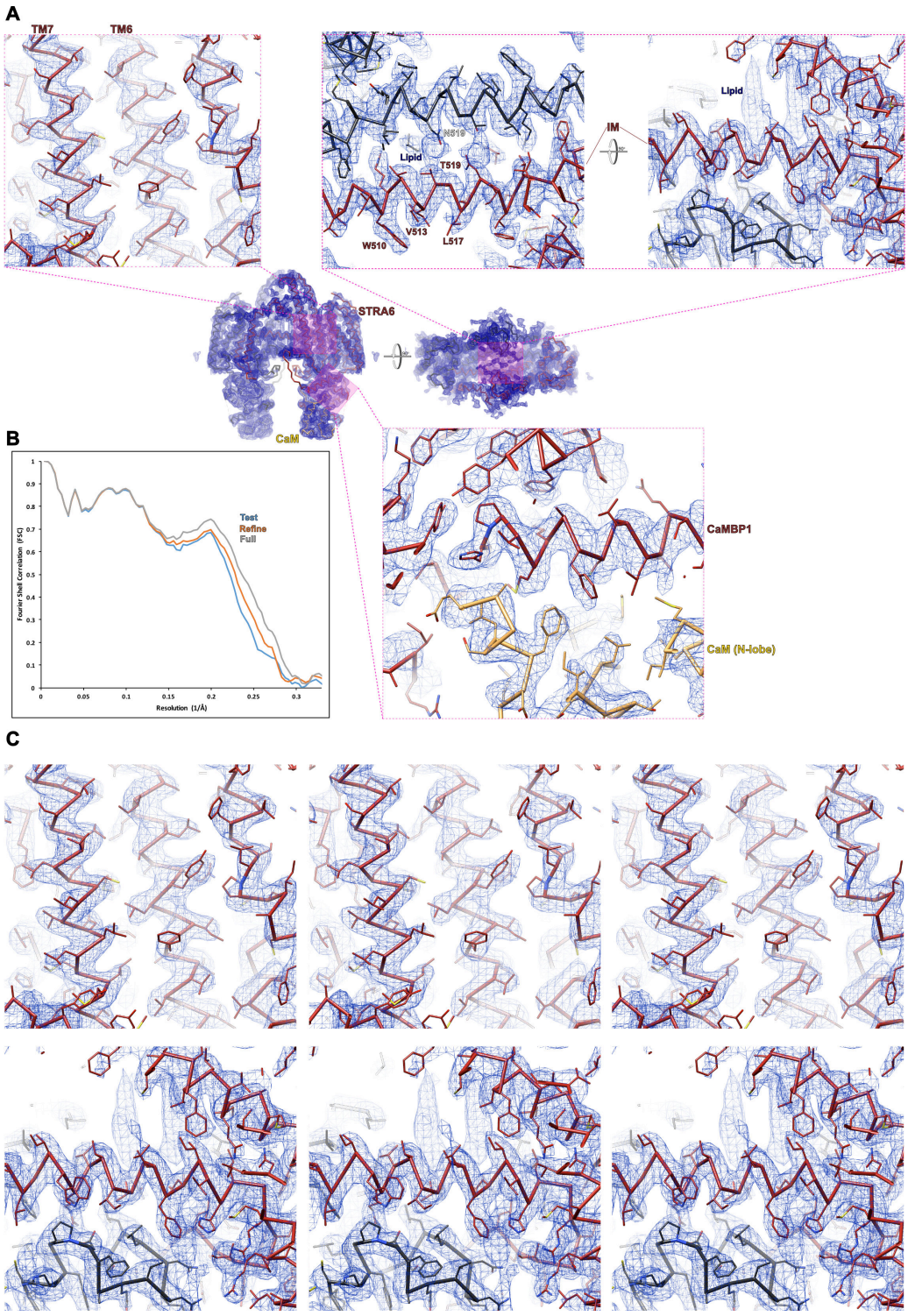
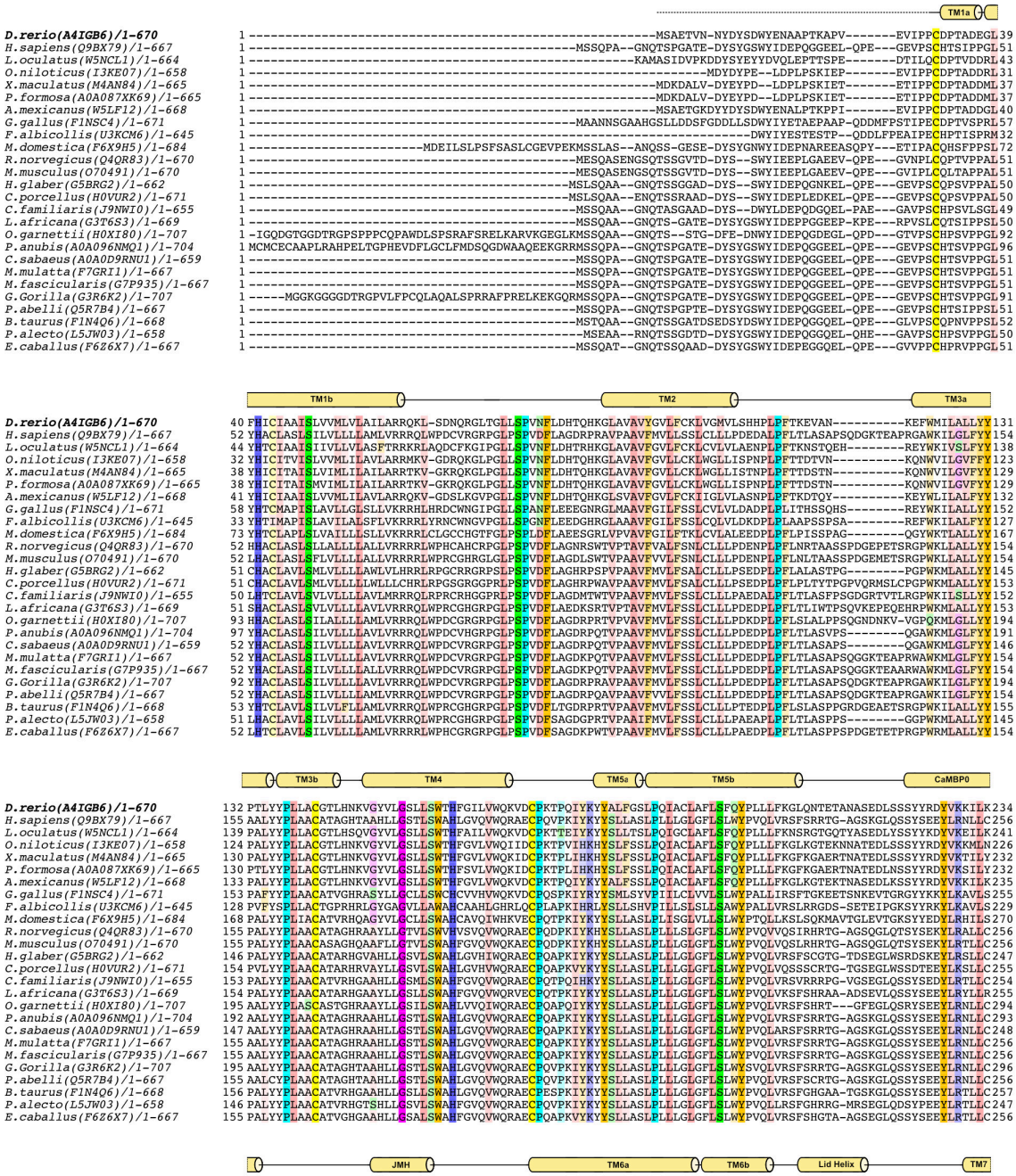


Fig. S6. Quality of the cryo-EM density map. (A) Two views of the STRA6-CaM complex are represented in the center and three insets are arranged around it, showing density quality in specific regions of the structure, with the structure represented as a Cα trace with side chains in stick representation. **(B)** Model-map FSC curves calculated for the full map used for initial model building (gray); the test half-map withheld from the cross-validation refinement run (orange) and the

half-map used in the cross-validation refinement (blue). For details of model-map FSC calculations see Materials & Methods. Stereo views of two of the insets from (**A**) are shown in panel **(C)**. Density maps are contoured at 6 X R.M.S.


TM1a
TM1b
TM2
TM3a
TM3b
TM4
TM5a
TM5b
CaMBP0
JMH
TM6a
TM6b
Lid Helix
TM7

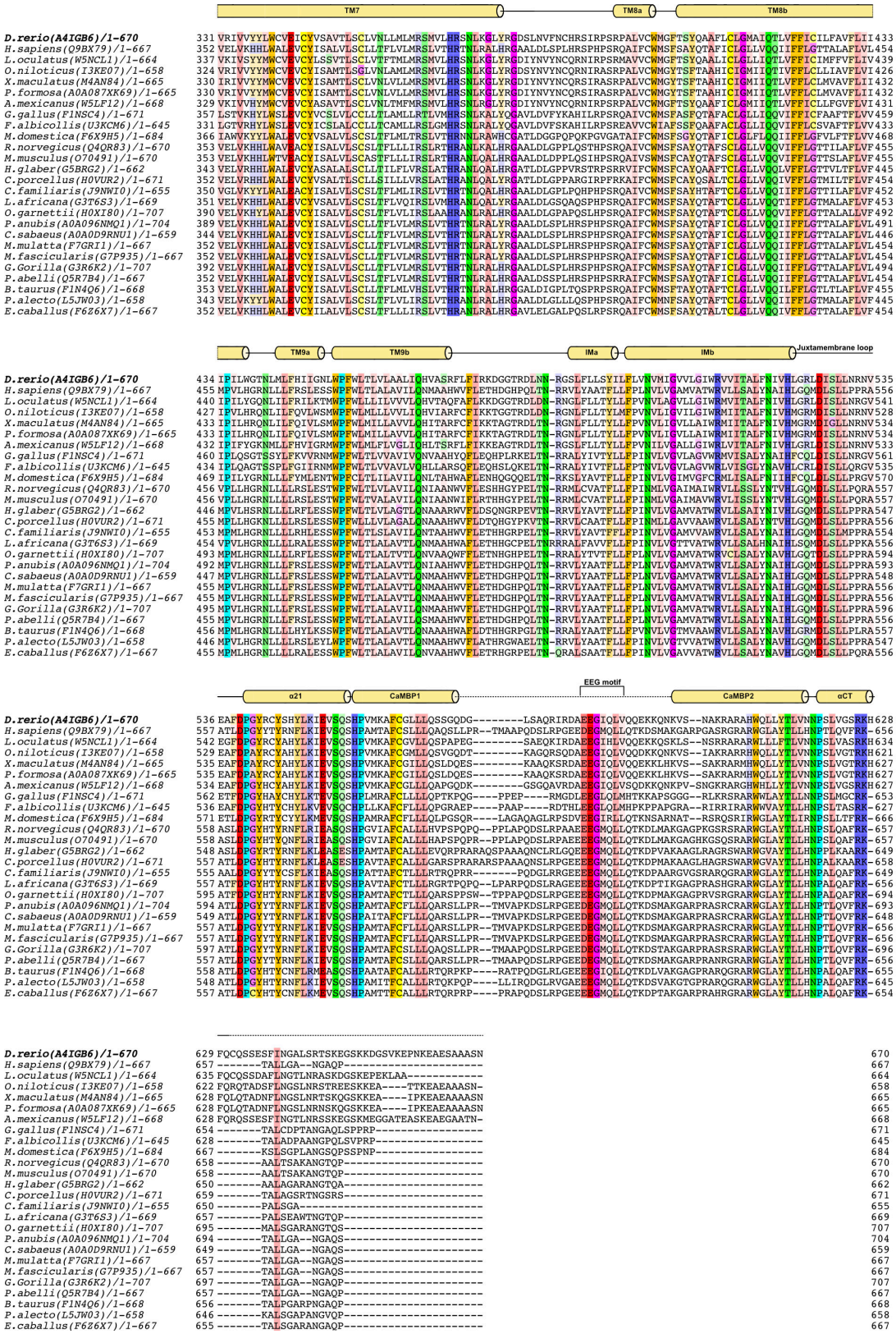


Fig. S7. STRA6 sequence alignment. Sequence alignment (generated using MUSCLE (*64*)) comprising 26 STRA6 sequences, where the intensity of the color is proportional to the degree of conservation, colored and visualized using Jalview

(*65*). Residues are colored by their physicochemical properties using a modified version of the Zappo scheme: aliphatic residues (I, L, V, A & M) are in salmon, aromatics (F, W & Y) in orange, positively charged (K, R & H) in blue, negatively charged (D, E) in red, polars (S, T, N, Q) in green, G in magenta, P in cyan, C in yellow. Secondary structure elements are labeled and displayed above the alignment, in yellow. Solid and dotted indicate regions of the protein, which are visible and disordered in the structure, respectively.

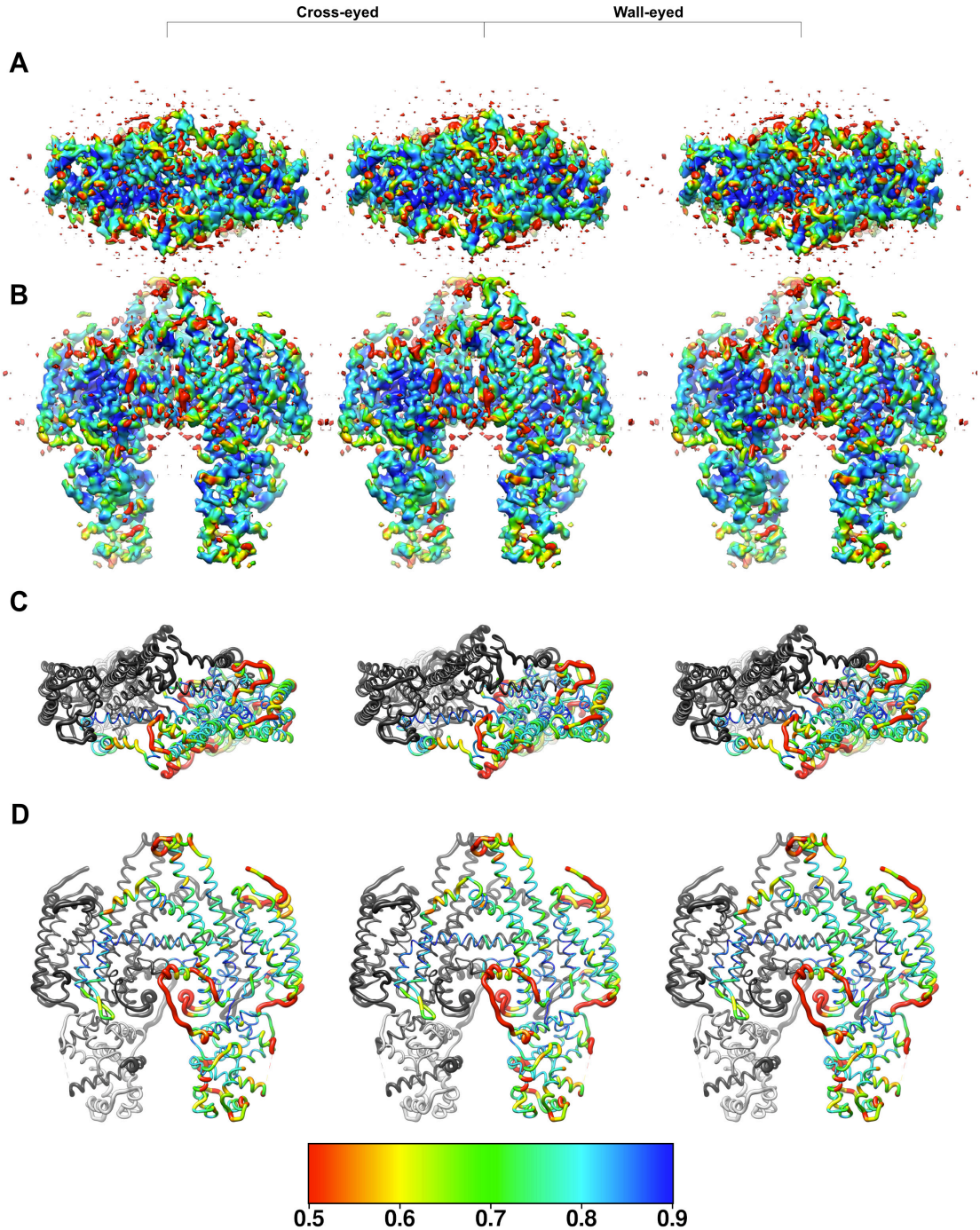


Fig. S8. Correlation between the STRA6 model and the density map. All panels are represented in stereo as indicated. The same color key is used for all panels and is reproduced at the bottom of the page. (**A** and **B)** Two views of the experimental density map, contoured at 6XR.M.S and colored by local correlation with a synthetic density map generated using *phenix.model_map* (version 1.10.1-dev-2427) from a model with B-factors estimated using *phenix.real_space_refine*. The synthetic map had the same sharpening B-factor applied as was used in post-processing of the cryo-EM map. Local correlation was calculated in a 5 voxel sliding window using the "*vop localcorrelation*" command in UCSF Chimera. **(C**

and **D)** Two views of the STRA6-CaM atomic model, with one protomer colored by the local correlation of each residue, averaged over the atomic positions. The radius of the backbone trace is inversely proportional to the correlation, such that a thicker trace indicates lower correlation with the experimental map.

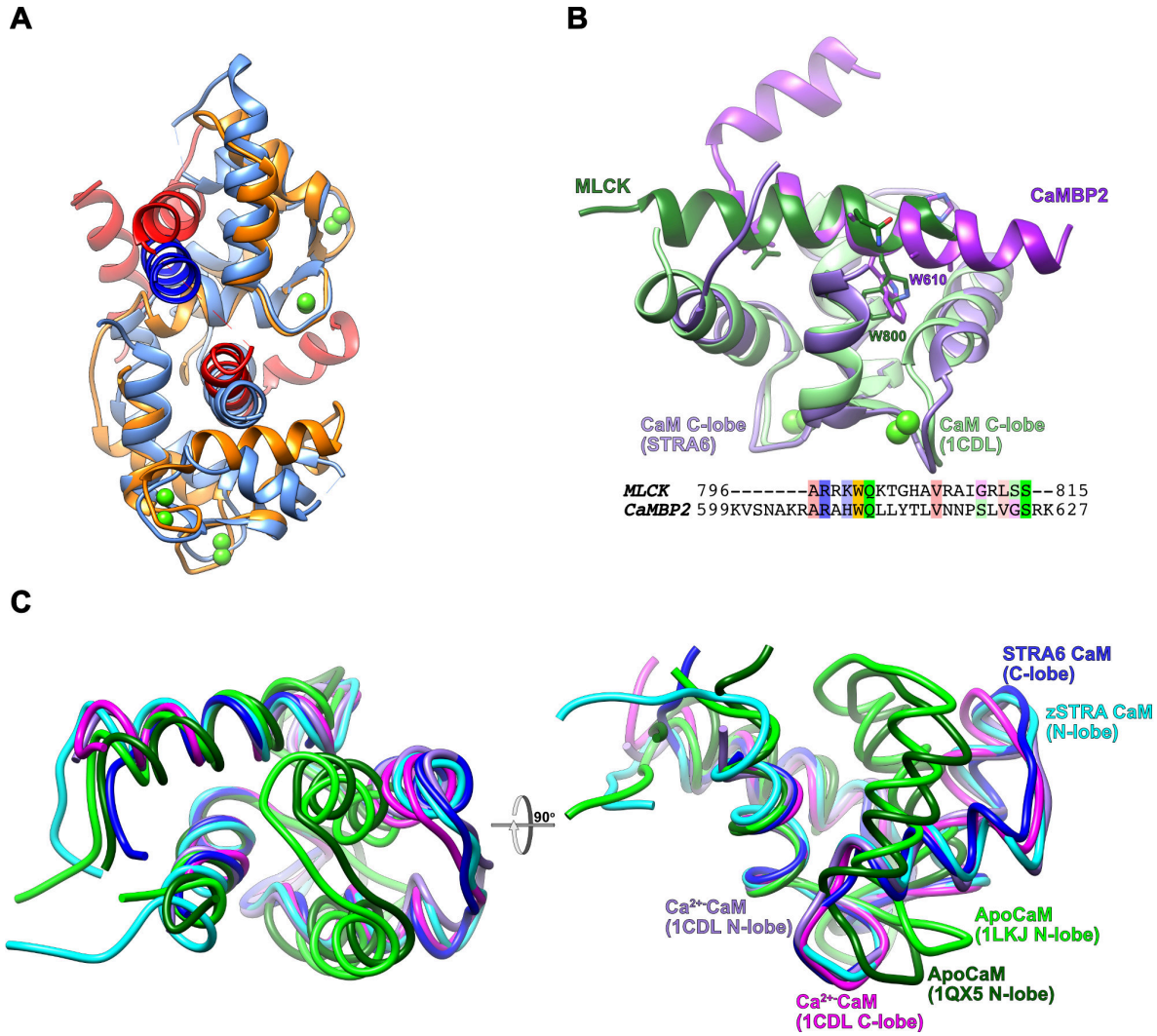


Fig. S9. Comparison of STRA6 bound CaM to other EF hand proteins and prior CaM structures. (A) The Cα trace of the structure of CaM (orange) bound to the three STRA6 peptides (red) is shown superimposed on the nearest match in the PDB (light/dark blue; PDB accession code: 3L19). Ca^{2+}-ions are represented as green spheres. **(B)** Superposition of the C-lobe of CaM bound to STRA6 CaMBP2, in light and dark purple respectively, with the complex of the MLCK target peptide and the C-lobe of Ca^{2+}-CaM (PDB accession code: 1CDL), in light and dark green respectively. Structurally homologous residues are shown in stick representation. A structure-based sequence alignment of the MLCK target peptide with CaMBP2 is shown below the structural superposition. **(C)** A superposition of the N- and C-lobes of CaM from the STRA6-CaM complex, displayed in cyan and blue respectively, with the N- and C- lobes from a Ca^{2+}-CaM structure (PDB accession code: 1CDL) in light purple and magenta respectively, and with two independent apo-CaM structures, the N-lobe of PDB accession code 1LKJ and the N-lobe of PDB accession code 1QX5, in light and dark green respectively.

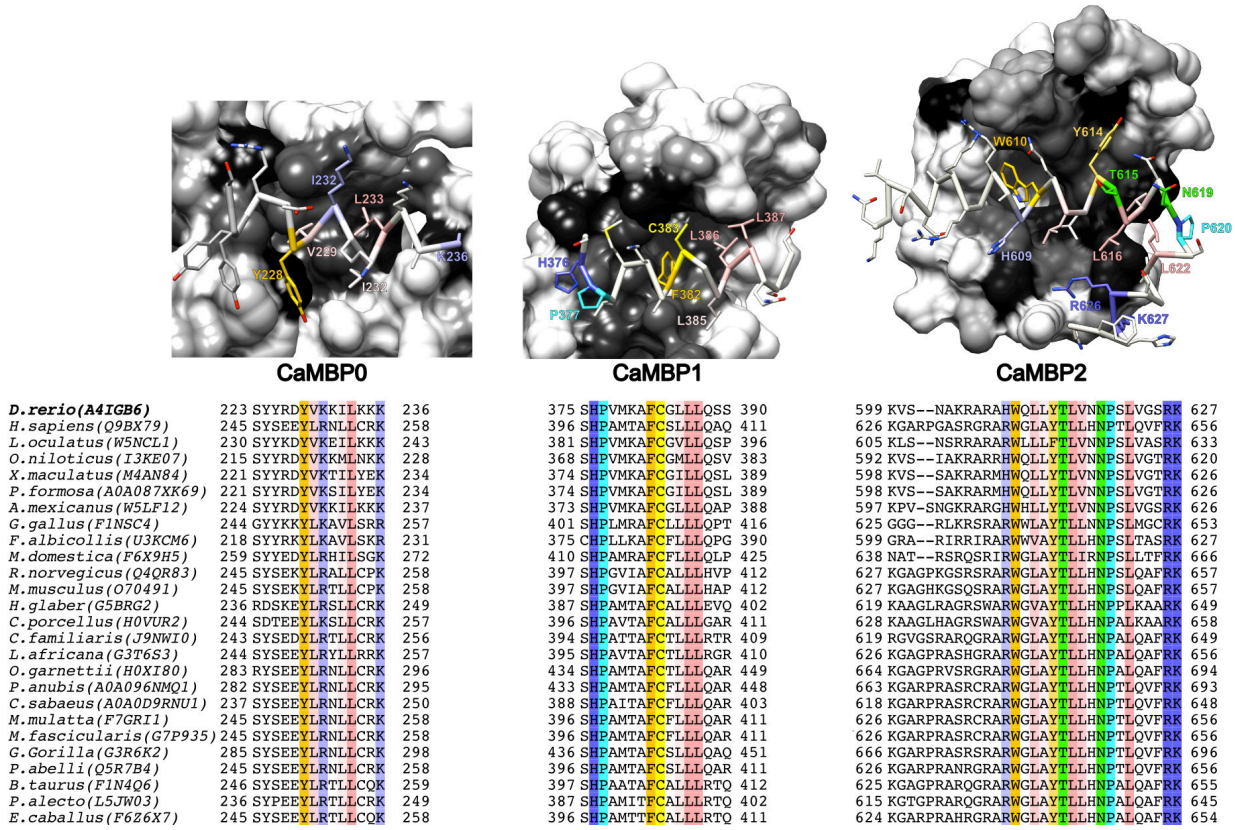


Fig. S10. Molecular details of the STRA6-CaM interaction. Separate views of the three CaM-binding peptides of STRA6 represented as Cα traces with side chains shown in stick, overlaid over a hydrophobic surface colored according to the Kyte-Doolittle hydropathy scale (*71*) from white (-4.5, most polar) to black (4.5, most hydrophobic). For CaMBP0, CaMBP1 is also represented as part of the hydrophobic surface, as it forms part of the CaMBP0-binding cleft.

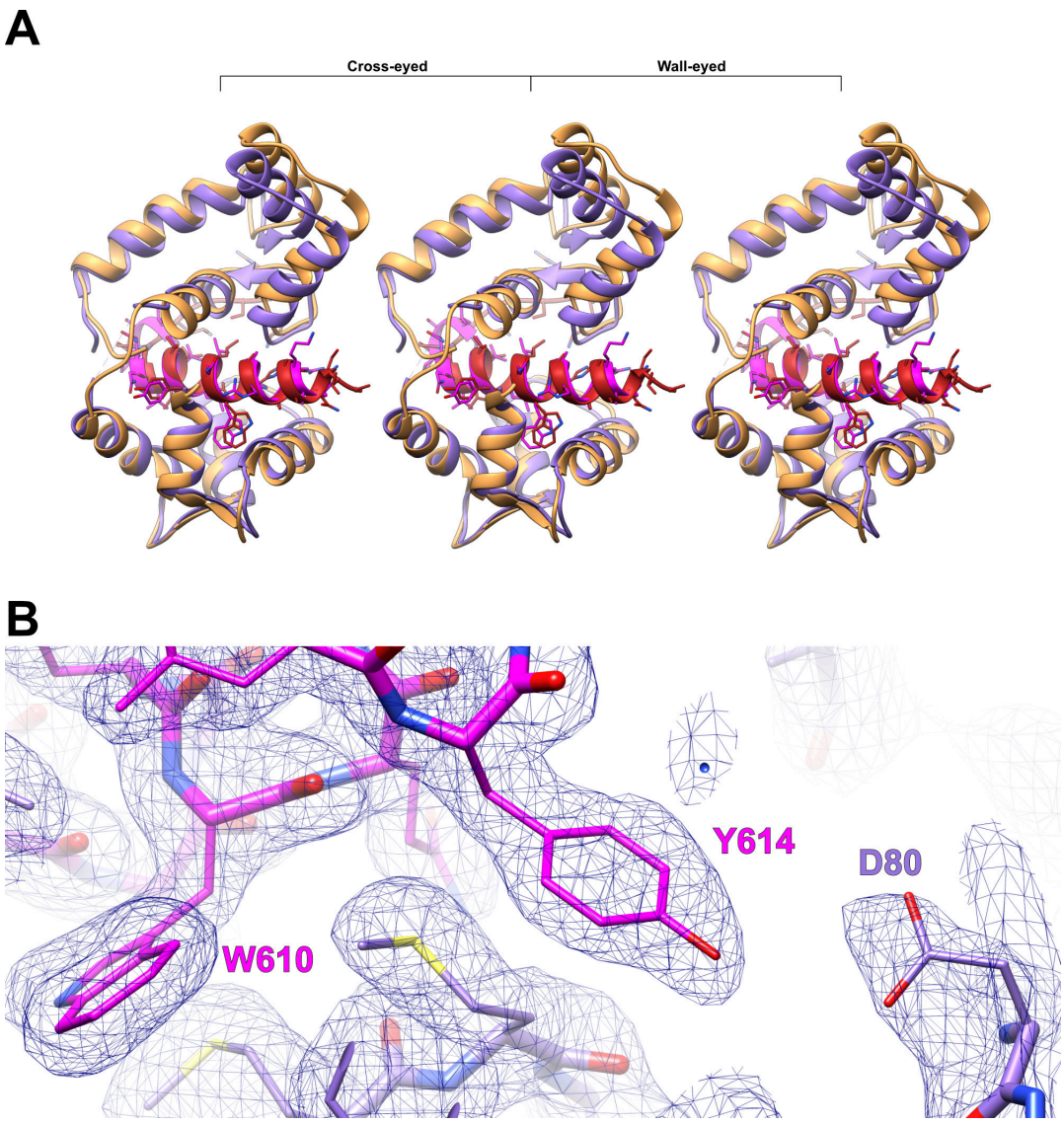


Fig. S11. Crystal Structure of CaM-CaMBP2. (A) Stereo representation of an overlay of the CaM-CaMBP2 crystal structure (in purple and magenta, respectively) and the corresponding region of the STRA6-CaM cryo-EM structure (in dark red, and gold, respectively). **(B)** Representative region of the 2Fo-Fc electron density map of the CaM-CaMBP2 crystal structure contoured at 1 x R.M.S.

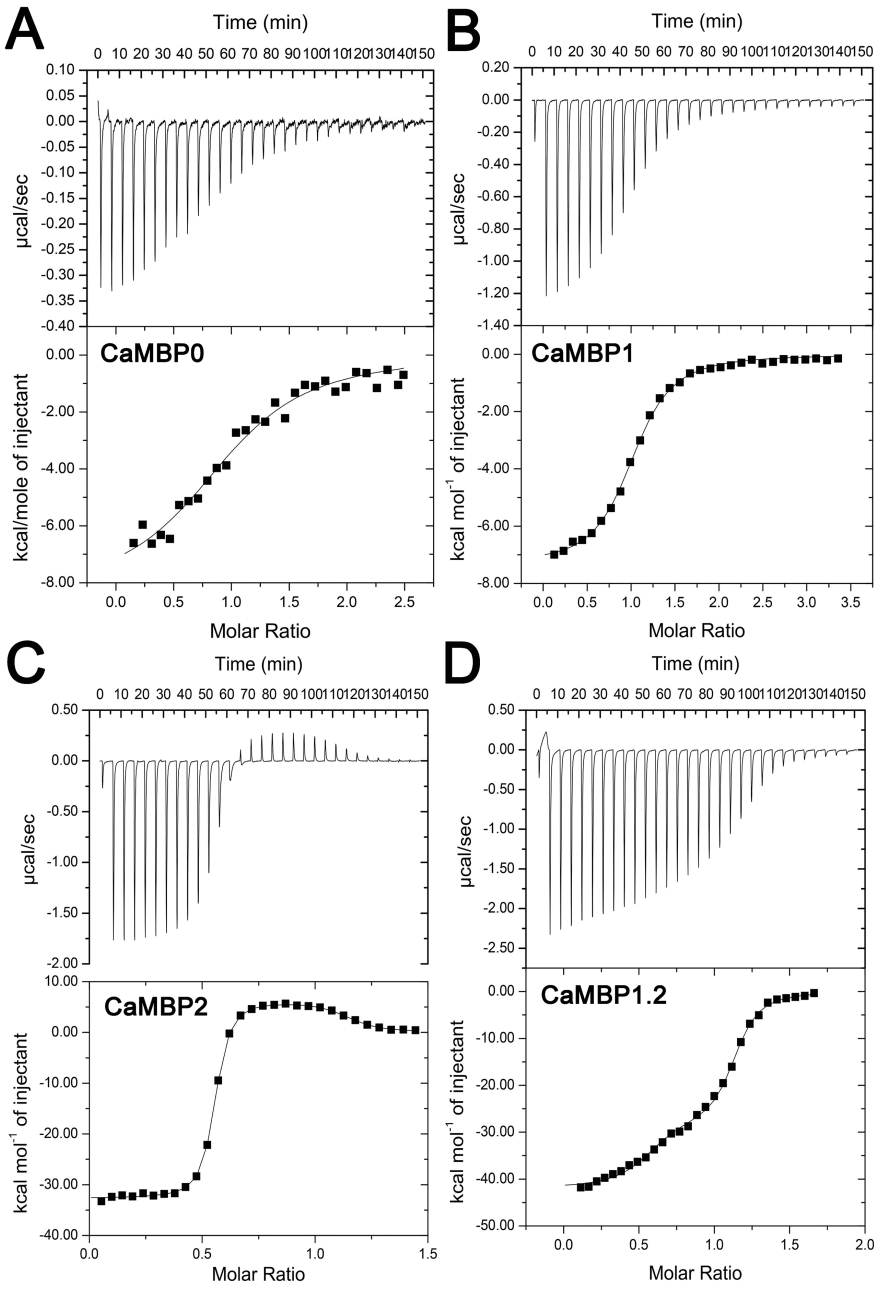


Fig. S12. ITC thermograms of STRA6 peptides – CaM interactions. Representative ITC thermograms for the reversed titrations used to observe CaM interactions with STRA6 peptides: CaM interaction with **(A)** CaMBP0, **(B)** CaMBP1, **(C)** CaMBP2, and **(D)** CaMBP1.2 under calcium-loaded conditions. All experiments were performed in 20 mM HEPES, pH 7.4, 50 mM NaCl, 10 mM $CaCl_2$, 5 mM $MgCl_2$, and 0.5 mM TCEP at 37°C. CaMBP1 titrations into CaM were not performed, because the peptide aggregated at concentrations greater 0.1 mM in aqueous solutions due to its hydrophobicity, preventing its use as a titrant.

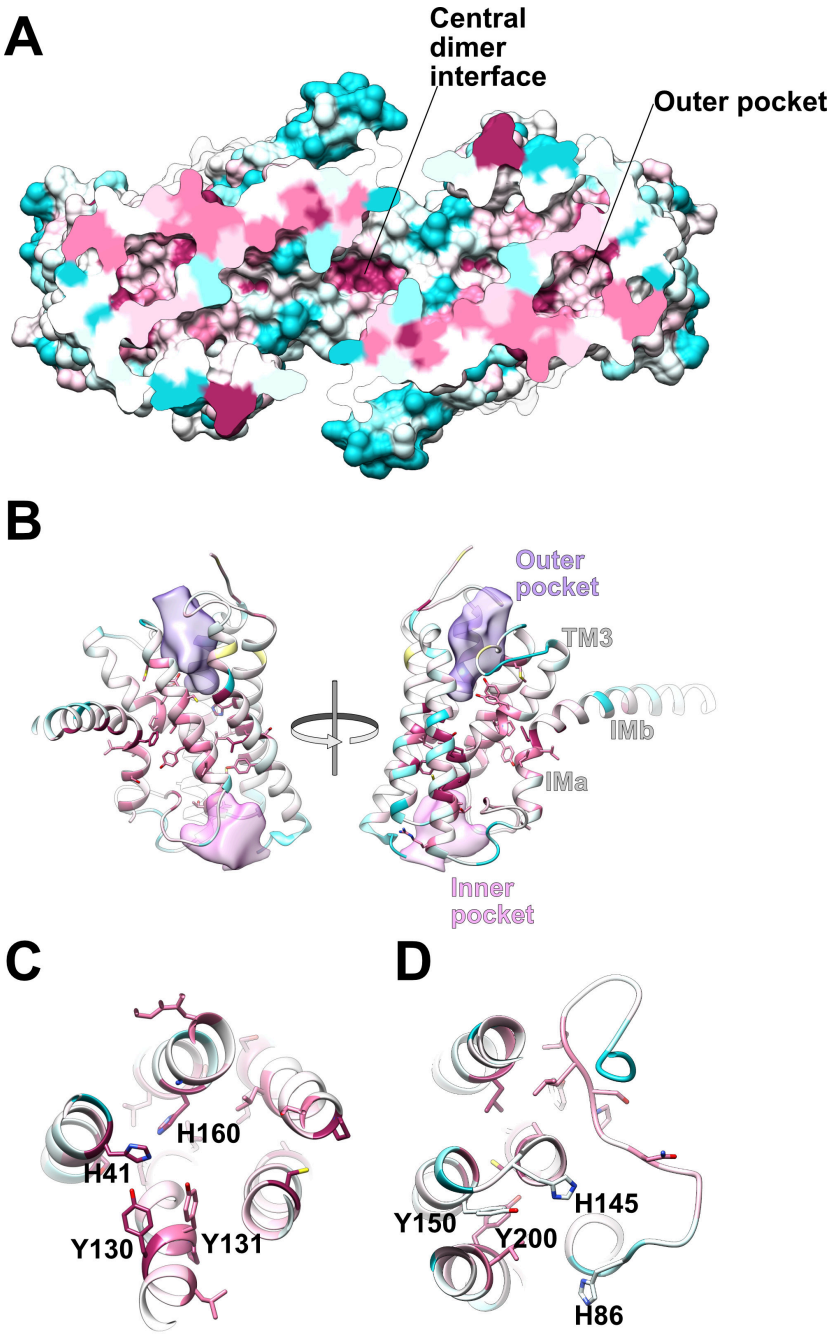


Fig. S13. Conservation of STRA6 within the outer cleft and pockets. (A) Slab in the membrane plane of the STRA6 conservation surface using the Consurf server, viewed from the extracellular region. The STRA6 molecular surface is colored from cyan (least conserved) to maroon (most conserved). **(B)** Ribbon representation of the NTD, IMa and IMb colored by conservation as in (A), with conserved residues represented in stick, and the bounding surfaces of each pocket shown in purple and pink transparent surfaces for the outer and inner pockets, respectively. **(C)** Molecular details of conserved interactions within the outer pocket. **(D)** Molecular details of interactions within the outer pocket.

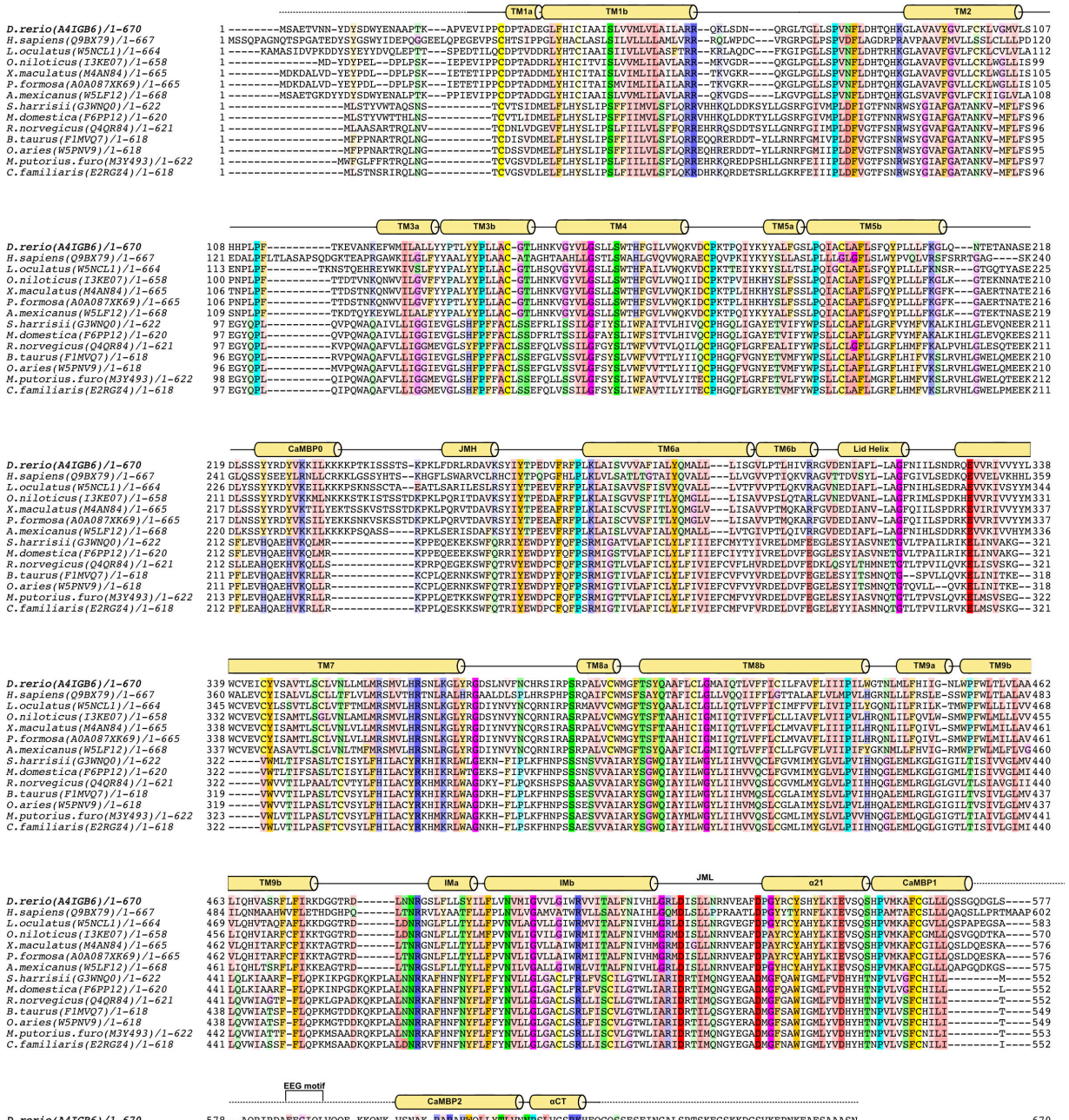


Fig. S14. Sequence alignment between STRA6 and RBPR2. Sequence alignment (generated using MUSCLE (*64*)) comprising seven STRA6 sequences and seven putative RBPR2 sequences, where the intensity of the color is proportional to the degree of conservation, colored and visualized using Jalview (*65*). Residues are colored by their physicochemical properties using a modified version of the Zappo scheme: aliphatic residues (I, L, V, A & M) are in salmon, aromatics (F, W & Y) in orange, positively charged (K, R & H) in blue,

negatively charged (D, E) in red, polars (S, T, N, Q) in green, G in magenta, P in cyan, C in yellow. Secondary structure elements are labeled and displayed above the alignment, in yellow. Solid and dotted indicate regions of the protein, which are visible and disordered in the structure, respectively.

Movie S1. Density quality of STRA6. The STRA6-CaM complex is depicted in ribbon representation with one STRA6 protomer in spectral coloring (violet N-terminus to red C-terminus). The associated protomer of CaM is colored cyan. The density map is represented as blue mesh and contoured at 7 X R.M.S.

Movie S2. STRA6-CaM interactions. CaM is represented as a molecular surface colored by hydrophobicity according to the Kyte-Doolittle scale *(71)* from -4.5 (most polar) in light blue to 4.5 (most hydrophobic) in orange. The N-lobe of CaM is initially at the top of the screen. CaMBP0, 1 and 2 are depicted in ribbon representation with side chains shown in stick. The CaMBPs are colored by conservation from cyan (least conserved) to maroon (most conserved), based on the alignment presented in **Fig. 2** and **Fig. S7**.