Multimeric assembly and biochemical characterization of the Trax/Translin endonuclease complex

Yuan Tian[1,2,7], Dhirendra K. Simanshu[1,7], Manuel Ascano Jr[3,7], Ruben Diaz-Avalos[4], Ah Young Park[5], Stefan A. Juranek[3], William J. Rice[4], Qian Yin[6], Carol V. Robinson[5], Thomas Tuschl[3,8] & Dinshaw J. Patel[1]

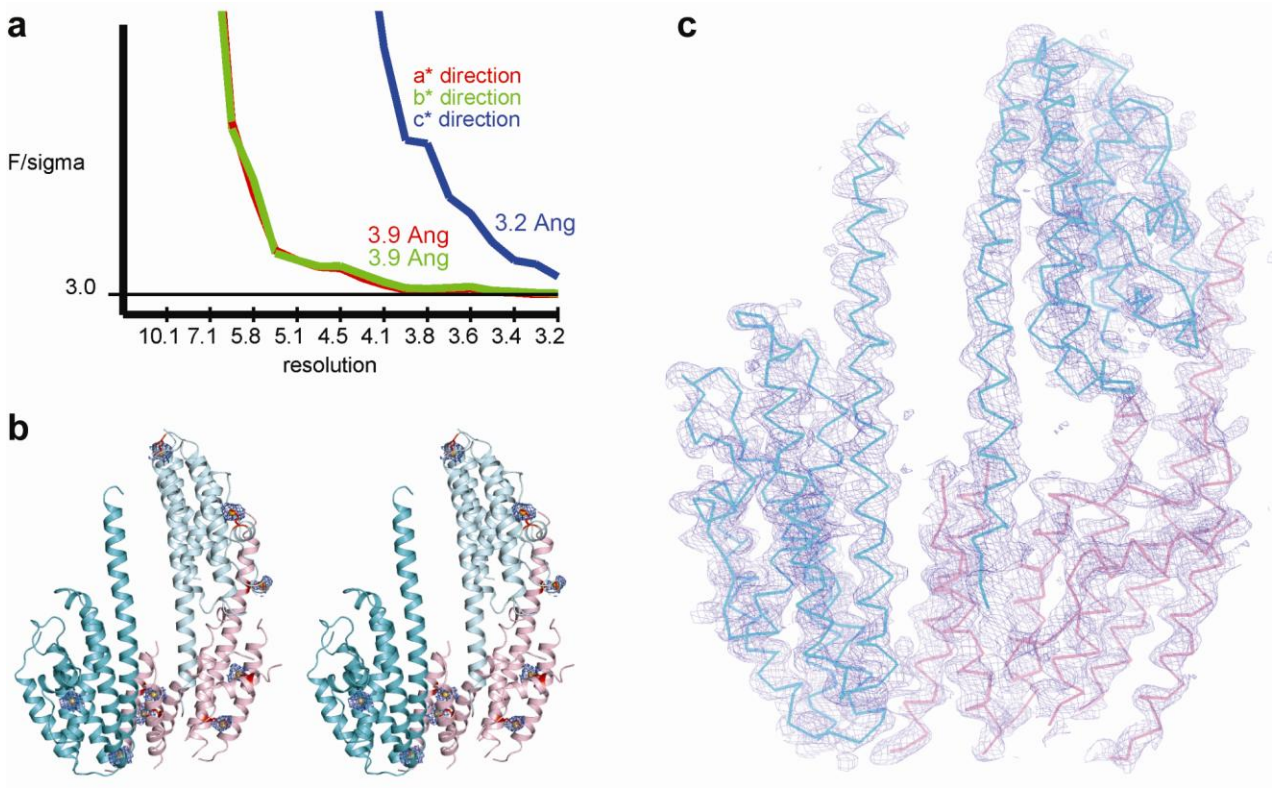
[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065 USA. [2]Graduate Program in Neuroscience, Weill Medical College of Cornell University, New York, NY, 10065 USA. [3]The Rockefeller University and [8]HHMI Laboratory of RNA Molecular Biology, New York, NY, 10065. [4]New York Structural Biology Center, New York, NY 10027 USA. [5]Department of Chemistry, University of Oxford, Oxford, OX1 3QZ, UK. [6]Department of Biochemistry, Weill Medical College of Cornell University, New York, NY, 10065 USA

[7]These authors contributed equally to this work.

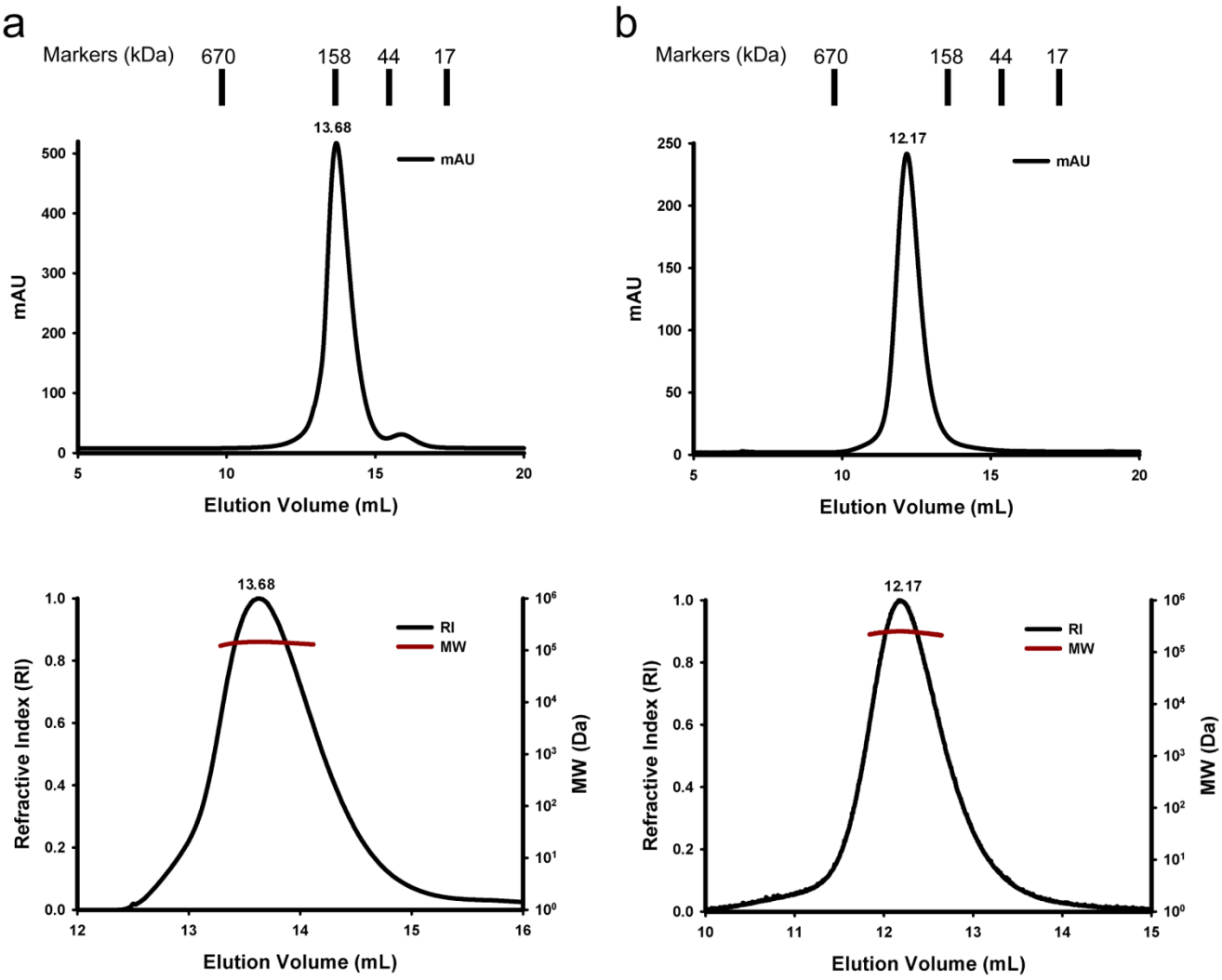
Correspondence should be addressed to D.J.P. (pateld@mskcc.org) and T.T (ttuschl@mail.rockefeller.edu)

Supplementary materials

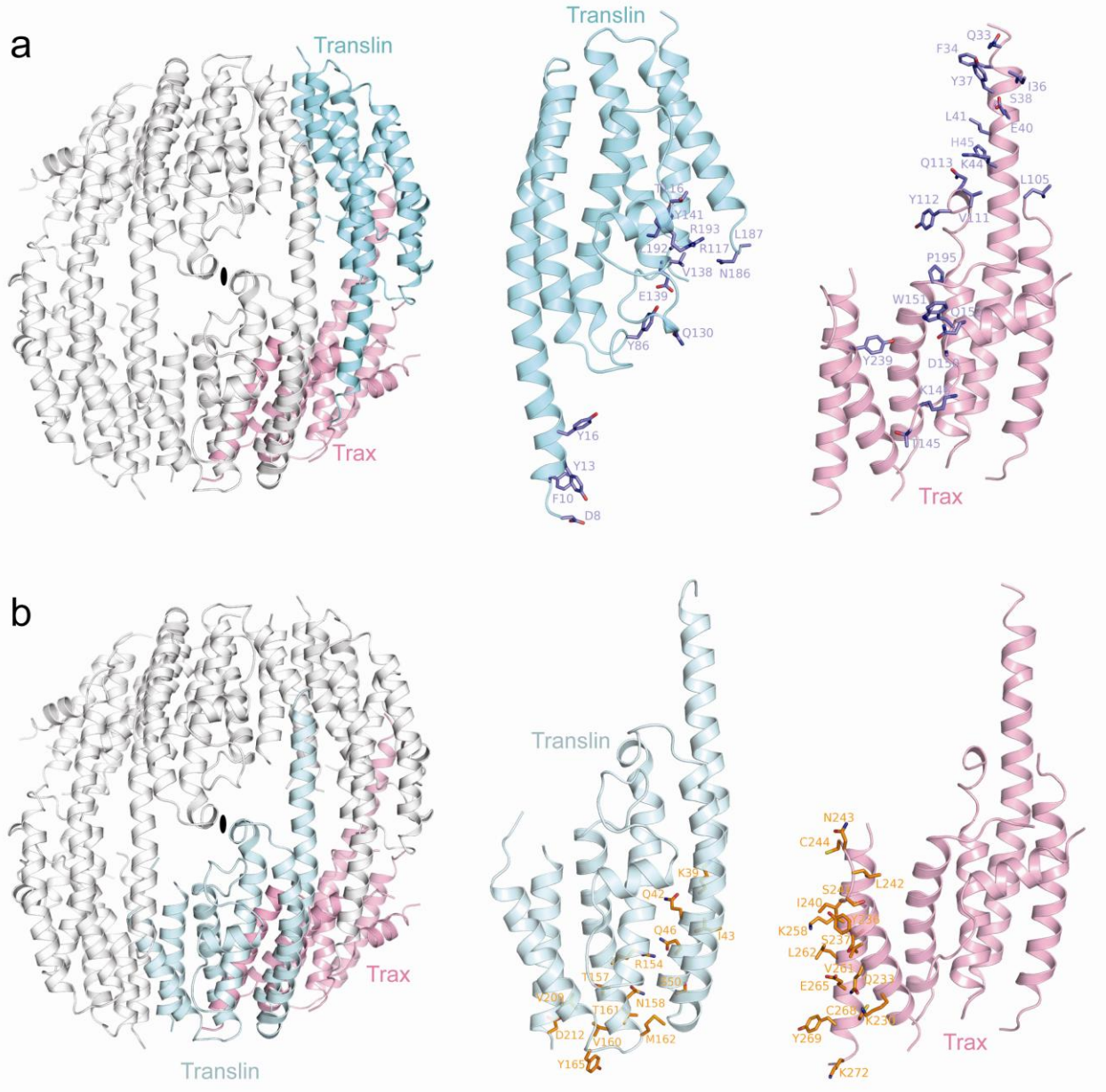
List of supplementary materials:

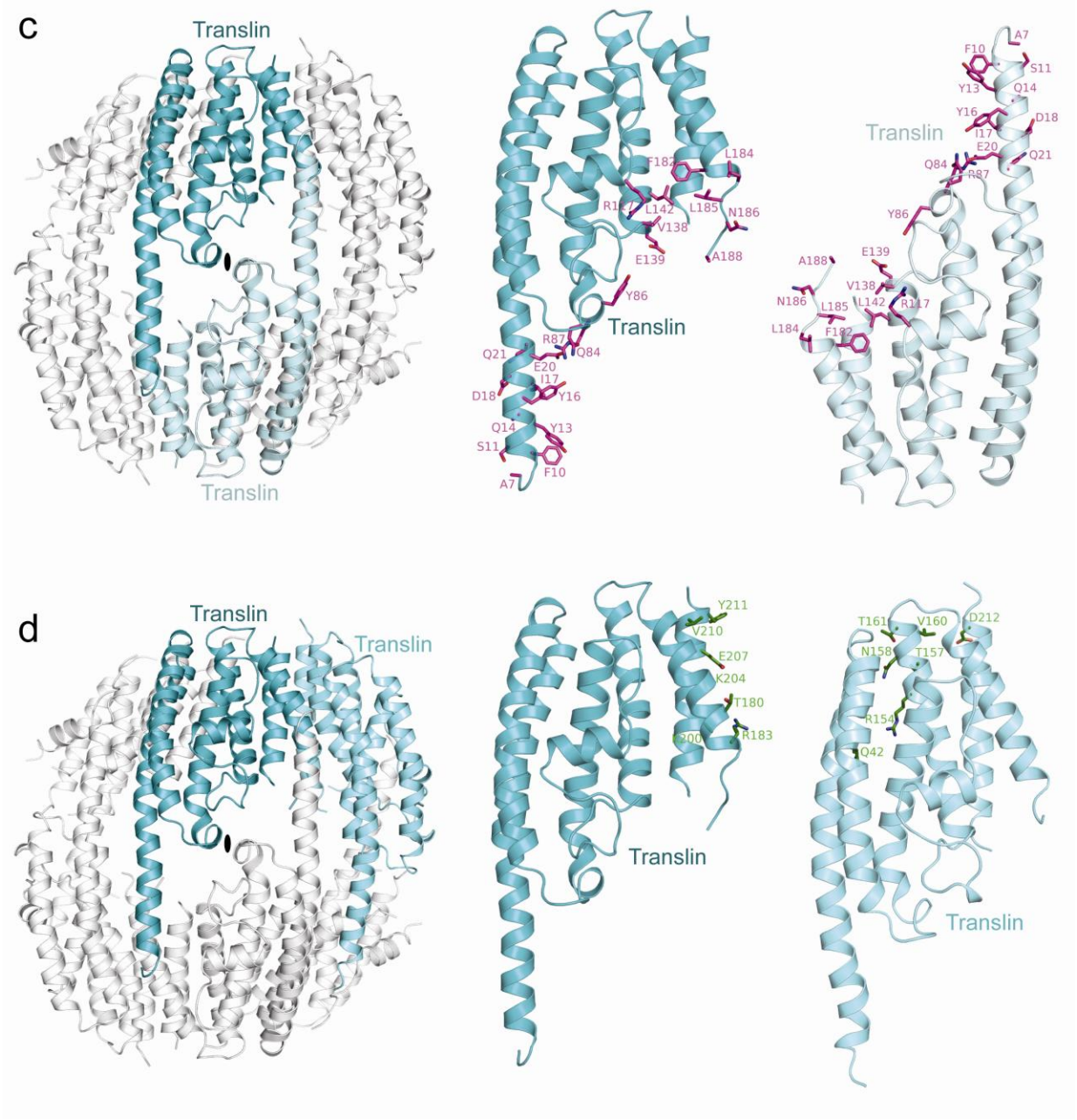


Supplementary Fig. 1. Anisotropic diffraction analysis, positions of selenomethionine residues and initial experimental map for truncated C3PO**. (a)** Data collected from truncated C3PO crystal was corrected for anisotropic diffraction using NIH anisotropic diffraction server, which suggested the resolution limit cut-off for the three principle axes of the ellipsoid. F/sigma is plotted for the three principle directions of the crystal (red, green, blue). The resolutions at which F/sigma drops below 3.0 for each of the three principle axes are selected by default as the resolution limits for ellipsoidal truncation. **(b)** A stereo view identifying positions of seleno-methionine (SeMet) residues in the asymmetric unit consisting of two molecules of Translin (two shades of cyan) and one molecule of Trax (pink). Protein shown in ribbon representation, while side chain atoms of seleno-methionine residues shown in stick representation along with anomalous difference map (blue mesh) contoured at 2.5σ for nine Se atoms (shown as balls colored in orange). **(c)** Initial experimental solvent-flattened map (contoured at 1σ) obtained after phasing and density-modification is shown along with final refined model of truncated C3PO. Two molecules of Translin (colored in cyan) and one molecule of Trax (colored in pink) present in the asymmetric unit are shown in wire representation.

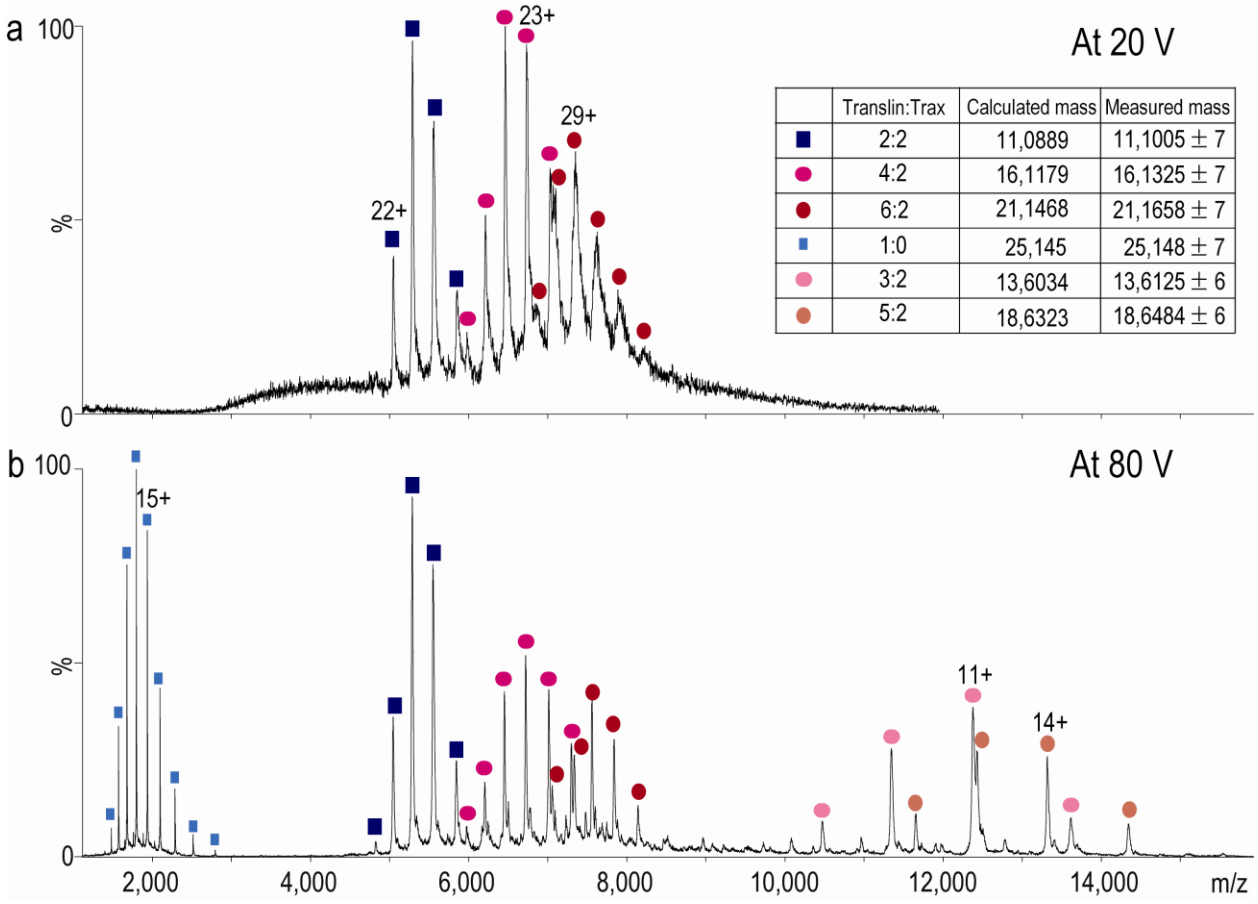


Supplementary Fig. 2. Gel-filtration and multi-angle light scattering (MALS) profile of C3PO. Gel-filtration (top panel) and MALS (bottom panel) profile of **(a)** truncated C3PO and **(b)** full-length C3PO. Locations of molecular weight standards in gel-filtration profile are marked. In MALS profile, the trace in black corresponds to light scattering signals at 90°. The trace in dark red shows the variation in molecular mass determination as a function of peak position or elution volume.

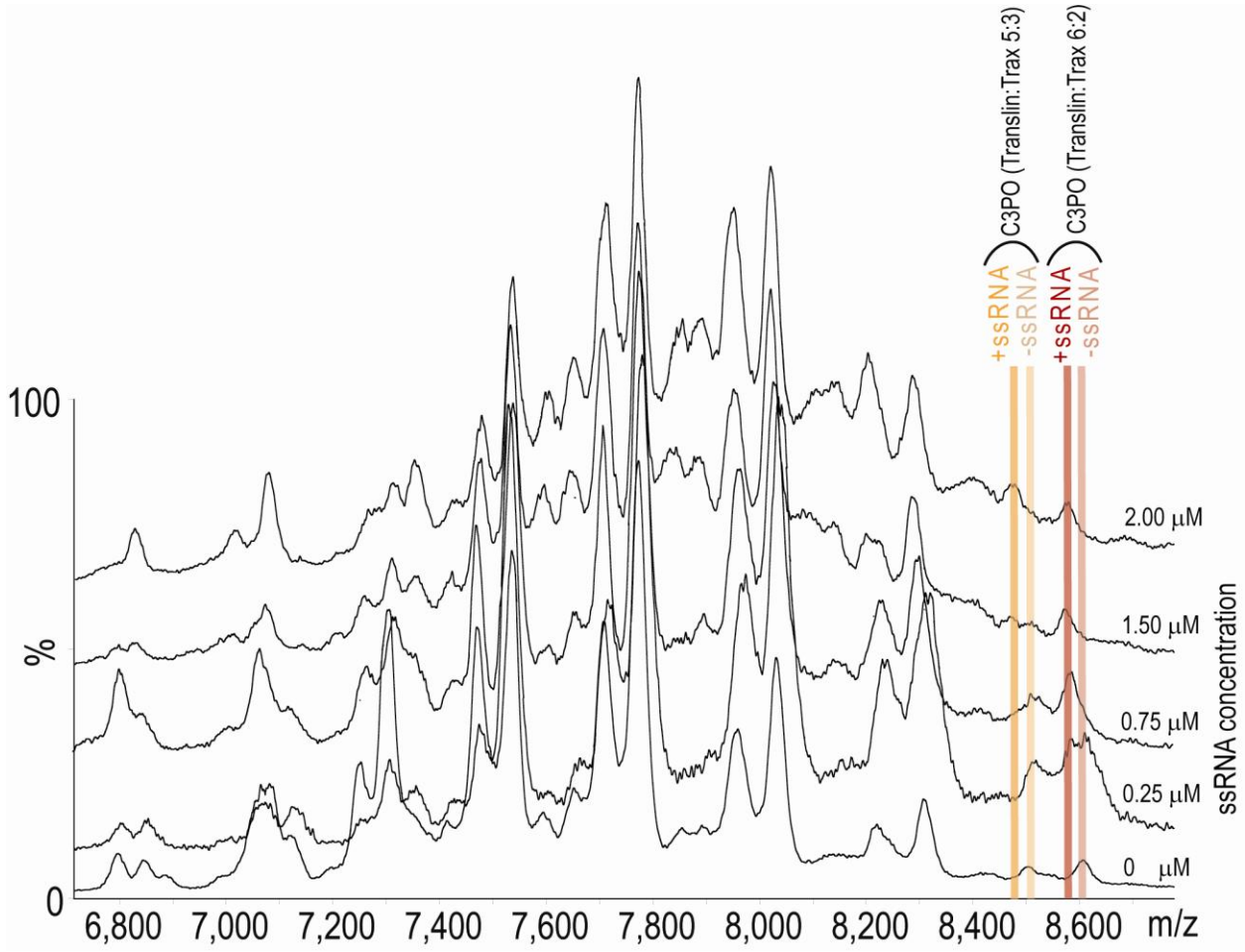

a
Translin
Trax
Translin
N16
Y141
R193
L187
I192
R117
V138
N186
E139
Q130
Y86
Y16
Y13
F10
D8
Q33
F34
Y37
I36
S38
L41
E40
H45
Q113
K44
L105
Y112
V111
P195
W151
Y239
D145
K14
I145
Trax
b
Trax
Translin
Translin
K39
Q42
Q46
I43
R154
T157
E50
V209
T161
N158
D212
V160
M162
Y165
N243
C244
L242
S241
I240
K258
S237
L262
V261
D233
E265
C268
K231
Y269
K272
Trax



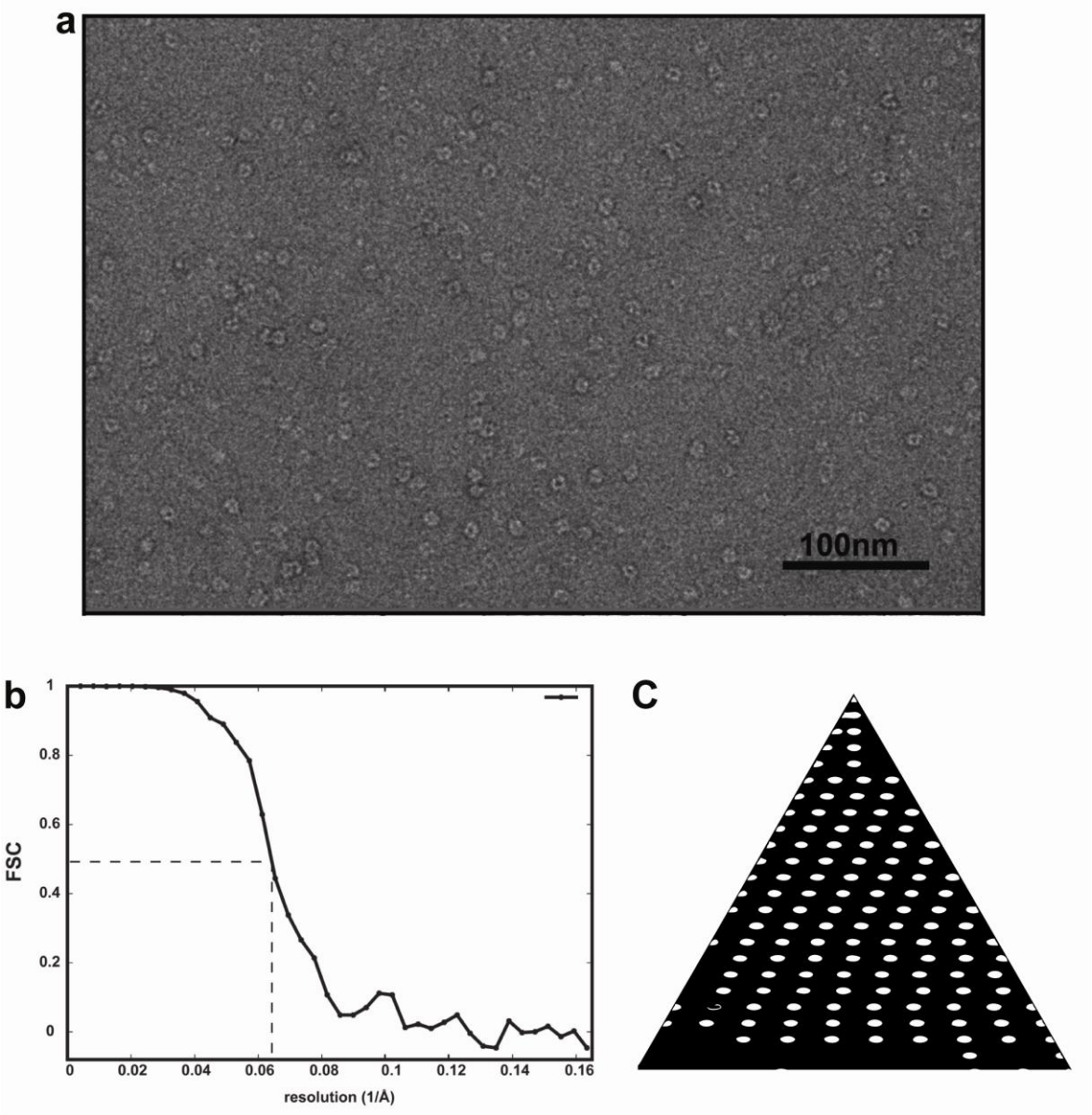
Supplementary Fig. 3. Interaction interface of Translin and Trax in truncated C3PO**.** A Translin-Trax heterodimer along with a Translin molecule from the middle Translin-Translin homodimer present in the asymmetric unit is related to rest of the truncated C3PO complex via 2-fold crystallographic symmetry. Amino acid residues involved in inter-subunit interaction at two-different interfaces of Trax with heterodimeric Translin and homodimeric Translin are shown in panel **(a)** and **(b)**, respectively. Amino acid residues present at the subunit interaction interface in two Translin molecules forming a homodimer and between two adjacent Translin molecules with one each from a homodimeric Translin and heterodimeric Trax-Translin are shown in panel **(c)** and **(d)**, respectively. Side-chain atoms of residues present at the subunit interface (cut-off < 4 Å) are shown in stick representation and are colored differently.



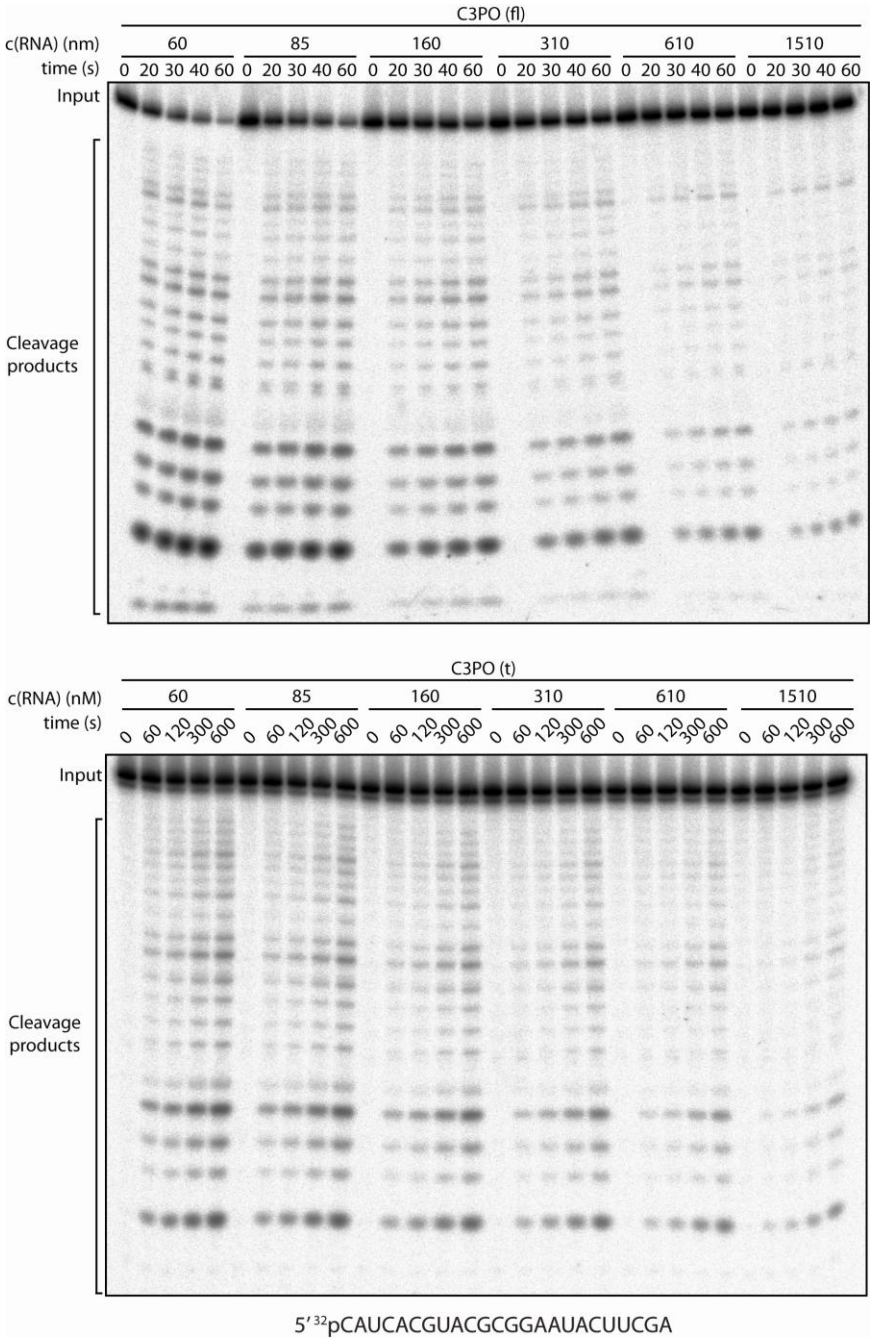
Supplementary Fig. 4. Mass spectra of the truncated C3PO. **(a)** Mass spectrum of the truncated C3PO complex at the concentration of 6 μM, obtained at low collision energy (20 V). Three well-resolved charge series are identified and assigned to the truncated Translin: Trax complexes with stoichiometries of 2:2, 4:2 and 6:2. **(b)** At 80 V the monomer of Translin is dissociated from the 4:2 and 6:2 (Translin:Trax) complexes and the stripped complexes with 3:2 and 5:2 stoichiometries are observed at high m/z (10,000-16,000).



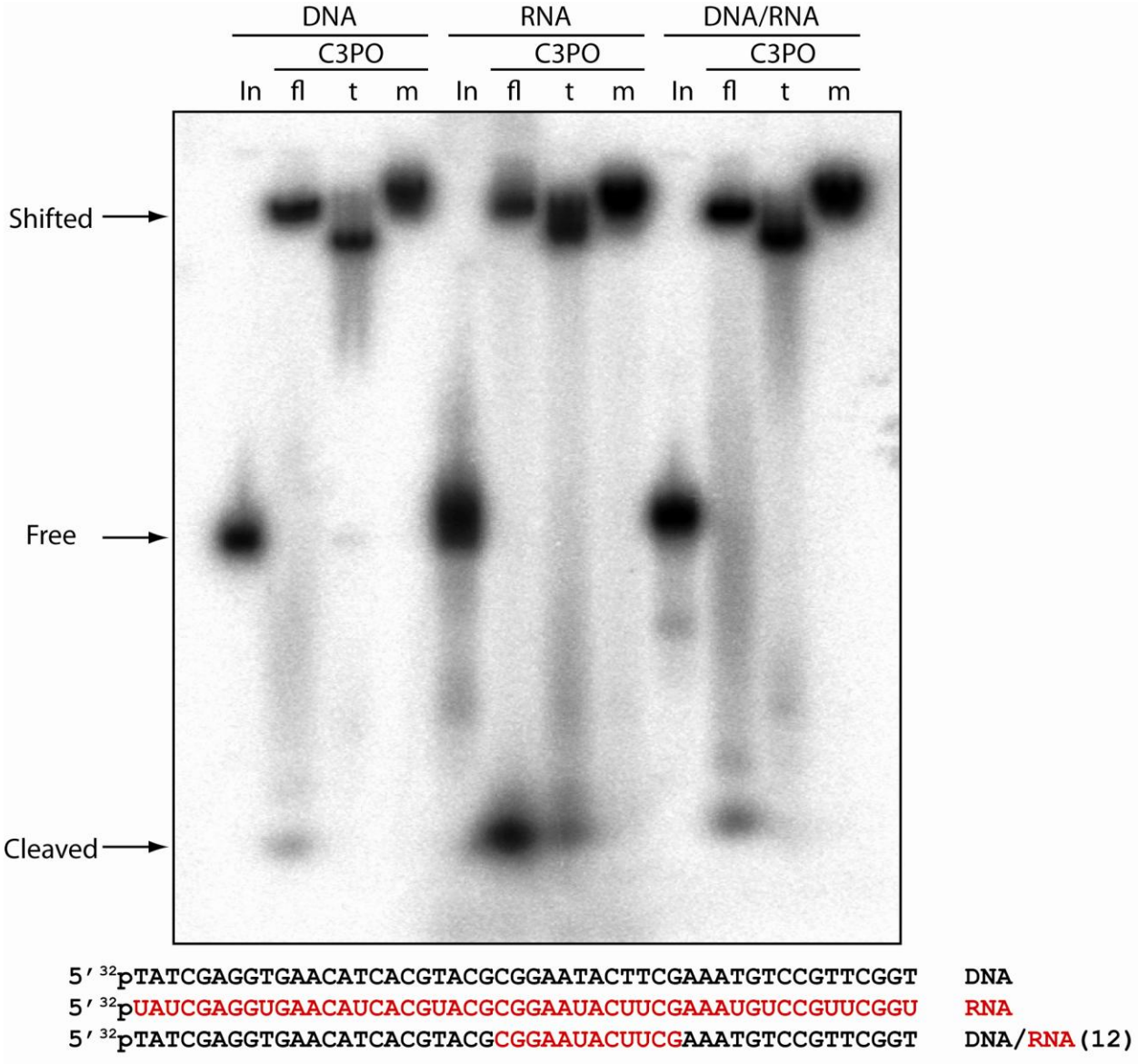
Supplementary Fig. 5. Titration of ssRN to the full-length C3PO**.** Changes in mass spectra of the full-length C3PO were observed in the presence of ssRNA at varied concentrations (noted on each spectrum). Due to a small mass increase and broadness of peaks, the charge series for the ssRNA bound C3PO complexes were resolved only at high m/z. Brown and orange lines represent the C3PO having Translin and Trax in 6:2 and 5:3 ratios, respectively. Light and dark lines within each colour represent each C3PO complex free and bound to ssRNA, respectively.



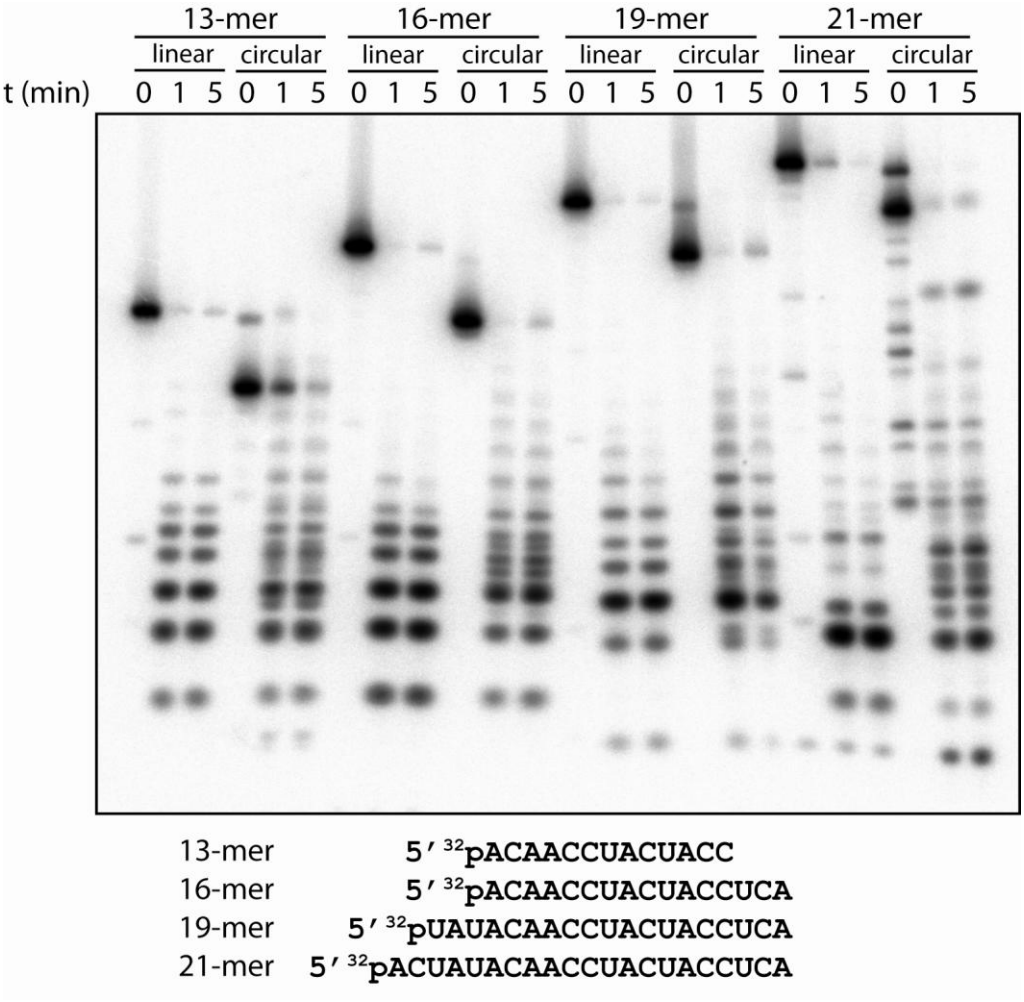
Supplementary Fig. 6. Negative staining of full-length C3PO. **(a)**Typical electron micrograph of the full-length C3PO over continuous carbon and stained with 1% uranyl acetate. **(b)** Resolution estimation for 3D image reconstructions using Fourier shell correlation (FSC) curve. FSC curve showing the correlation of two reconstructions computed with the two different halves of the dataset. The intersection at the 0.5 correlation mark (dashed line) is taken as the resolution limit in the data, and in this case corresponds to ~15 Å. **(c)** A plot showing the distribution of particles among the different classes used for EM reconstruction. Classes are plotted in a triangle to illustrate the Euler angles used for generating projections. The top of the triangle represents a view along the Z-axis, while the base represents a fraction of the circle in the XY plane. D4 symmetry was used for generating projections. Brightness of each spot indicates the number of particles; the lighter the spot, the greater the number of particles in that particular class. The number of particles per class varied between 12 and 91, with a mean of 43.



Supplementary Fig. 7. Kinetic determination of C3PO nucleolytic activity. A 24-nt radiolabeled oligoribonucleotide was used to measure the catalytic activity of full-length (fl) and truncated (t) C3PO. Shown here are representative results used for the C3PO kinetic measurements. Full-length C3PO reactions were performed using 10 nM protein from 0 to 60 s; truncated C3PO measurements were done using 50 nM protein from 0 to 600 s. The observed maximum velocity of full-length C3PO was derived by determining the rate of RNA cleavage at 1510 nM; under this condition, observed V = 7.38 nM/s and k_{obs} = 0.74 s^{-1}. The maximum velocity observed truncated C3PO was measured using 1510 nM RNA substrate and 50 nM C3PO. For truncated C3PO the measured velocity was 0.22 nM/s (k_{obs} = 0.004 s^{-1}).



Supplementary Fig. 8. C3PO complexes bind DNA, RNA, and DNA/RNA chimeric oligonucleotides independent of their catalytic activity. C3PO complexes were incubated with radiolabeled DNA, RNA and a chimeric DNA oligonucleotide containing 12 internal RNA residues (in red). The incubations were performed in the same C3PO reaction buffer as before; however, the resulting reactions were then separated on a 6% native polyacrylamide gel. Abbreviations: In, input RNA.



Supplementary Fig. 9. Cleavage of circular RNA substrates of varying length by C3PO. Oligoribonucleotides of lengths between 13 and 21 nt were radiolabeled, an aliquot of which was further treated with T4 RNA ligase 1 to generate circularized substrates. 100 nM of linear or circularized RNA was then incubated with 100 nM of C3PO for the indicated times. Reactions were stopped and separated on an 18% denaturing polyacrylamide gel.

Supplementary methods

Protein purification

The cell extract was prepared using a French press in buffer A (20 mM Tris-HCl pH 8.0, 1 M NaCl, 30 mM Imidazole). After centrifugation, the supernatant was loaded onto a Ni-affinity column, which was then extensively washed with buffer A and eluted with buffer B (1 M Imidazole pH 8.0, 500 mM NaCl). The eluant protein fractions were pooled and dialyzed in buffer C (20 mM Tris-HCl pH 8.0, 50 mM KCl) overnight at 4 ºC, along with protease Ulp1 for removal of the fusion tag. This protocol was followed by a second round of purification using a Ni-affinity column to remove the fusion tag, His-tag containing protease and any undigested fusion protein. The flow-through containing the C3PO complex was further purified by size-exclusion column chromatography using HiLoad 26/60 Superdex-200 prep grade column (GE healthcare) in buffer C. Peak fractions obtained from gel filtration were pooled and concentrated using centrifugal concentrators. Protein purity was assessed by SDS-PAGE. For phasing, the L-seleno-methionine (SeMet) derivatized protein was produced using the feedback-inhibition of methionine synthesis pathway in minimum media (M9). Purification of seleno-methionine containing protein was performed essentially in the same way as for the native protein.

Structure determination and refinement

The initial phases were improved by solvent flattening and histogram matching with SOLOMON[37] as implemented in SHARP[31]. In this map, 485 residues in 24 helical fragments were assigned using the ARP/wARP module[38], which recognizes helical structural motifs in electron density maps. The structure was built through several rounds of manual model fitting in COOT[39] and refinement using PHENIX.REFINE[40] coupled with *non-crystallographic symmetry (NCS) averaging, and* secondary structure and hydrogen-bond restraints. The model was further improved by refinement using rigid body, simulated annealing, positional and TLS refinement at different stages[41-43].

Five Se atoms present in Trax were located on different α-helices and thus were very useful during amino acid assignments. The previously solved crystal structure of *Drosophila* Translin[13] helped in building the two Translin molecules present in the asymmetric unit. A large solvent channel (solvent content 77%) and non-uniform crystal packing interactions in different directions explain the poor and anisotropic diffraction from the C3PO crystals (**Supplementary Fig 1a**). The majority of the model has an interpretable electron density map, with the exception of a stretch of residues present at the N- and C-termini and loops present between the helical

segments in the Trax molecule for which no clear electron density was observed. These segments were omitted in the final model. We were unable to build side-chains for 43 aa out of a total of 598 aa residues present in the asymmetric unit of C3PO. The geometry of the final model was checked using PROCHECK[44]. The mostly allowed and additionally allowed regions of the Ramachandran plot contained 90.2 % and 8 %, respectively, of non-glycine and non-proline residues. Figures were prepared using the program PyMol[45] and ESPript[46].

Mass spectrometry (MS)

Typically, 2 µl of the aqueous protein samples were electrosprayed from a gold-coated borosilicate capillary prepared in-house[47]. The following instrument parameters were used: capillary voltage 1.7 kV, cone voltage 100 V, extractor cone 5V, collision energy 20-120 V, MCP 2900 V, indicated pressure: 3.5×10^{-3} mbar source pirani, 1.3×10^{-4} mbar analyser, 1.3×10^{-6} mbar ToF. The simulated mass spectrum was generated automatically using an algorithm[29] implemented in Python, which extends on earlier approaches for simulating mass spectra[48,49].

EM image processing

Individual images of the particles (~8,000) were selected using the program *boxer* from the EMAN suite[50] and JWEB[51] for the tilt pairs. First, the particles (388) from the untilted images were classified using SPIDER, and from the most populated classes, 3D reconstructions were computed using the particles from the tilted images[52], producing thus a 3D reconstruction using the random conical tilt method. The reconstructions from the most populated classes were used as the initial models for refinement (16 cycles) of the structure, using all the selected particles to produce a final average. Both EMAN and SPIDER were used to refine the data, producing very similar final averages. The CTF of the images was calculated using *ctfit* and the phases flipped accordingly. The first 12 cycles of refinement were done without imposing symmetry, and it was only for the final four cycles that the apparent symmetry (D4) was imposed, giving at the end the volume shown in **Fig. 3b**. The volume produced without the imposition of the symmetry has a very apparent D4 symmetry.

EM labeling/tagging experiments

We attempted to label the Trax molecule so as to pinpoint their positions within the full-length C3PO complex. First, we attempted using Ni-NTA gold particles to target His-tag epitopes inserted either at the N-terminus, the C-terminus or in the loop region present in center of Trax. To minimize the non-specific binding of gold, bovine serum albumin was added to the C3PO

solution. In the case of labels targeted to the C-terminus of Trax, we selected ~14,000 particles for refinement using an initial model of the 3-D reconstruction of the undecorated C3PO. Although the gold particles were visible in the micrographs, there was no density arising from them in the average, indicating that the binding to the His-tag did not occur with full occupancy and/or in a specific fashion. Similar results were obtained with other gold labels. We then attempted to use a construct of Trax having a MBP (Maltose Binding protein) tag at C-terminus within the full-length C3PO context to visualize the extra mass of the MBP tag. This effort was also unsuccessful, mainly due to the fact that it was difficult to get very well-defined particles of the complex, although 20,000 particles were selected for averaging in this case.

Multi-angle light scattering (MALS)

Molar mass values of purified full-length C3PO and truncated C3PO were determined by gel filtration and MALS. The protein was injected on to a Superdex 200 HR 10/300 gel filtration column equilibrated in a buffer containing 20 mM Tris-HCl pH8.0 and 50 mM KCl. The chromatography system was coupled to a three-angle light scattering detector (mini-DAWN EOS) and refractive index detector (Optilab DSP, Wyatt Technology). Data were collected every 0.5 second at a flow rate of 0.2 ml/minute. Data analysis was carried out using the program ASTRA, yielding the molar mass and mass distribution of the sample.

5' end radiolabeling of oligonucleotides

Typically, 100 pmoles of oligonucleotide was incubated with 5 pmoles of γ-^{32}P ATP and 10 units of T4 PNK in 1 x T4 PNK buffer (70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT) in a 20 µl reaction for 15 minutes at 37ºC; followed by addition of 1000 pmoles of non-radiolabeled ATP for another 5 minutes. Circular RNA was similarly radiolabeled but then isolated by phenol-chloroform and ethanol precipitation and self-ligated by incubating it with T4 Rnl1 (0.2 µg/µl) in 50 µl of buffer containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM □-mercaptoethanol, 0.2 mM ATP, 0.1 mg/ml acetylated BSA, 15% DMSO for 1 h at 37ºC. All reactions were stopped by addition of 1 volume of stop buffer (8 M urea, 10 mM EDTA, and bromophenol blue).

Kinetic determination of the catalytic activities of C3PO

Nuclease reactions were carried out at various protein concentrations using full-length C3PO and 100 nM RNA substrate to determine optimal initial velocity conditions at a reasonable time scale (data not shown); 10 nM was chosen accordingly. Master mixes, with a volume sufficient

for all tested time points, were prepared containing 60, 85, 160, 310, 610, and 1510 nM RNA (where 10 nM RNA of each was radiolabelled) in C3PO reaction buffer. 10 nM (final) full-length C3PO was pre-aliquoted into six reaction tubes, to which each master mix was immediately added by multi-channel pipette. 10 µl aliquots from each RNA concentration reaction were removed and immediately added to an equal volume of stop buffer for a total of four time points (20, 30, 40, and 60 sec). A 0 sec time point for each RNA concentration was prepared in which the original reaction and volume was prepared but without C3PO protein, substituting reaction buffer. Truncated C3PO kinetic experiments were similarly performed except that 50 nM protein was used and the time points were extended to 0, 60, 120, 300, and 600 sec. 2 µl of each reaction were separated by denaturing polyacrylamide gel electrophoresis and processed as before. Full-length C3PO reactions were done in quadruplicate and truncated C3PO done in duplicate.

Removal of 3' and 2',3'-cyclic phosphates from partially alkaline-hydrolysed oligonucleotides

Where indicated, 3' and 2',3'-cyclic phosphates were removed from alkaline hydrolyzed oligonucleotides by first phenol-chloroform and ethanol precipitating them, then incubating in 50 µl containing 60 units of T4 PNK in 1 x PNK buffer for 1 h at 37ºC. 10 µg proteinase K was added and incubated for 20 min at 37ºC. Samples were phenol-chloroform extracted twice then ethanol precipitated and resuspended in H_2O.

Oxidation and β-elimination reactions

C3PO-digested oligonucleotides, recovered by phenol-chloroform and ethanol precipitation, were resuspended in 17.5 µl of borax buffer, pH 8.6 (4.38 mM $Na_2B_4O_7{\cdot}10H_2O$, 50 mM H_3BO_3). 2.5 µl of $NaIO_4$ (28.6 mM final) was added and the reactants incubated for 10 min in the dark at 24ºC. Glycerol (5% final) was added to quench the reaction and incubated for 10 min in the dark at 24ºC. The samples were concentrated in a SpeedVac (Eppendorf) down to 5 µl. 50 µl of borax buffer, pH 9.5 (33.75 mM $Na_2B_4O_7{\cdot}10H_2O$, 50 mM H_3BO_3, pH adjusted by NaOH) was added and the samples were incubated for 90 min at 45ºC.

Supplementary References

37. Abrahams, J.P. & Leslie, A.G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**, 30-42 (1996).

38. Morris, R.J. et al. Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat* **11**, 56-9 (2004).

39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

40. Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. CCP4 Newsl. Vol. 42 (2005).

41. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

42. Adams, P.D. et al. Recent developments in the PHENIX software for automated crystallographic structure determination. *J Synchrotron Radiat* **11**, 53-5 (2004).

43. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).

44. Laskowski, R.A., Macarthur, M.W., Moss, D.S. & Thornton, J.M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).

45. DeLano, W.L. & Lam, J.W. PyMOL: A communications tool for computational models. *Abstracts of Papers of the American Chemical Society* **230**, U1371-U1372 (2005).

46. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320-3 (2003).

47. Hernandez, H. & Robinson, C.V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat Protoc* **2**, 715-26 (2007).

48. Sobott, F., Benesch, J.L., Vierling, E. & Robinson, C.V. Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J Biol Chem* **277**, 38921-9 (2002).

49. van Breukelen, B., Barendregt, A., Heck, A.J. & van den Heuvel, R.H. Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Commun Mass Spectrom* **20**, 2490-6 (2006).

50. Ludtke, S.J., Baldwin, P.R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97 (1999).

51. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-9 (1996).

52. Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc* **146**, 113-36 (1987).