

## SUPPLEMENTARY INFORMATION

### **Bayesian weighing of electron cryo-microscopy data for integrative structural modeling**

Massimiliano Bonomi<sup>1,\*</sup>, Samuel Hanot<sup>2</sup>, Charles H. Greenberg<sup>3</sup>, Andrej Sali<sup>3</sup>, Michael Nilges<sup>2</sup>,  
Michele Vendruscolo<sup>1</sup>, Riccardo Pellarin<sup>2,\*‡</sup>

*<sup>1</sup>Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom*

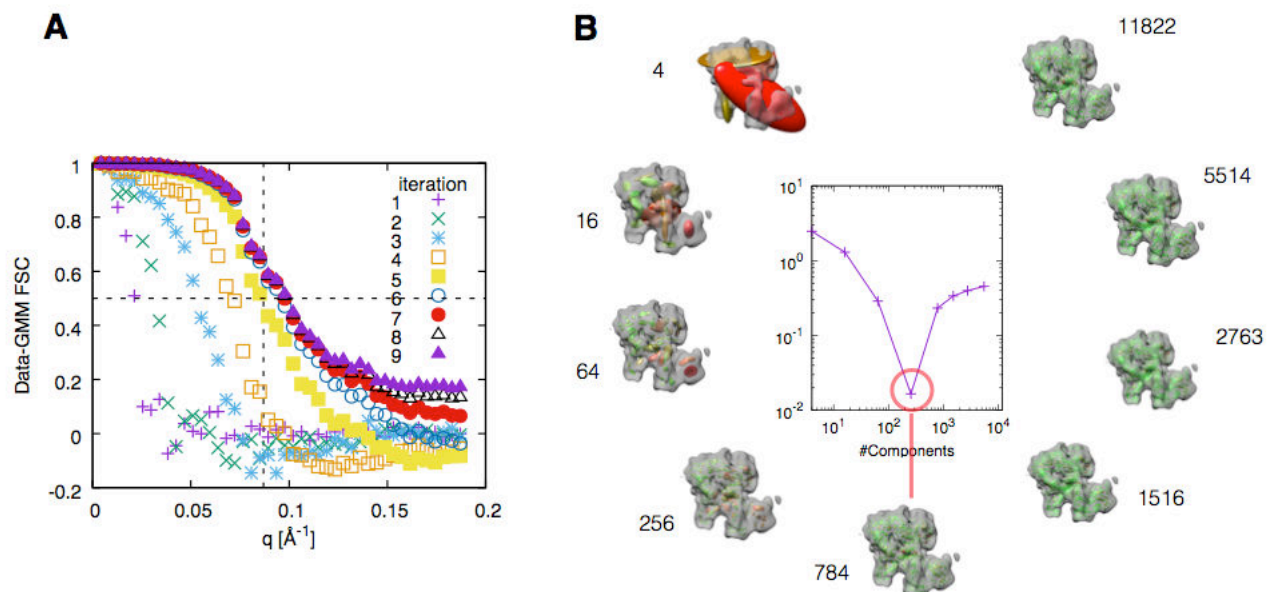
*<sup>2</sup>Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, 75015 Paris, France*

*<sup>3</sup>Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Sciences,  
and California Institute for Quantitative Biomedical Sciences, University of California, San Francisco,  
California 94158, United States*

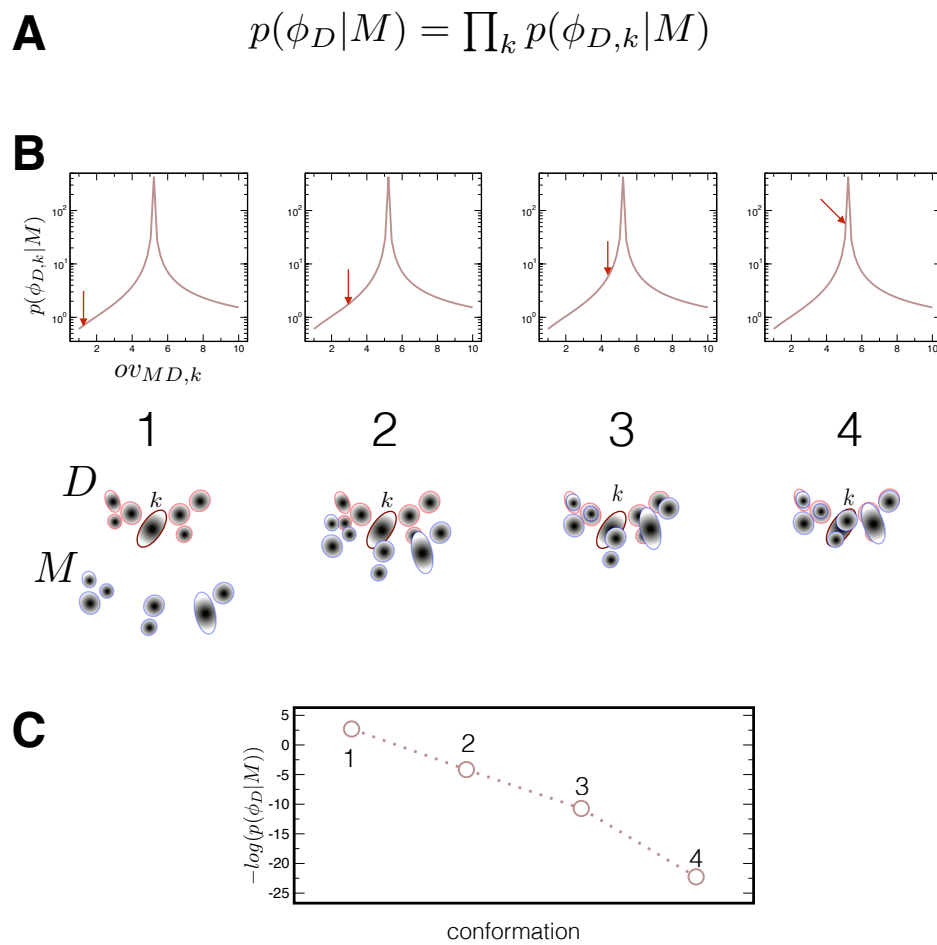
\*Correspondence to: mb2006@cam.ac.uk; riccardo.pellarin@pasteur.fr

‡Lead contact

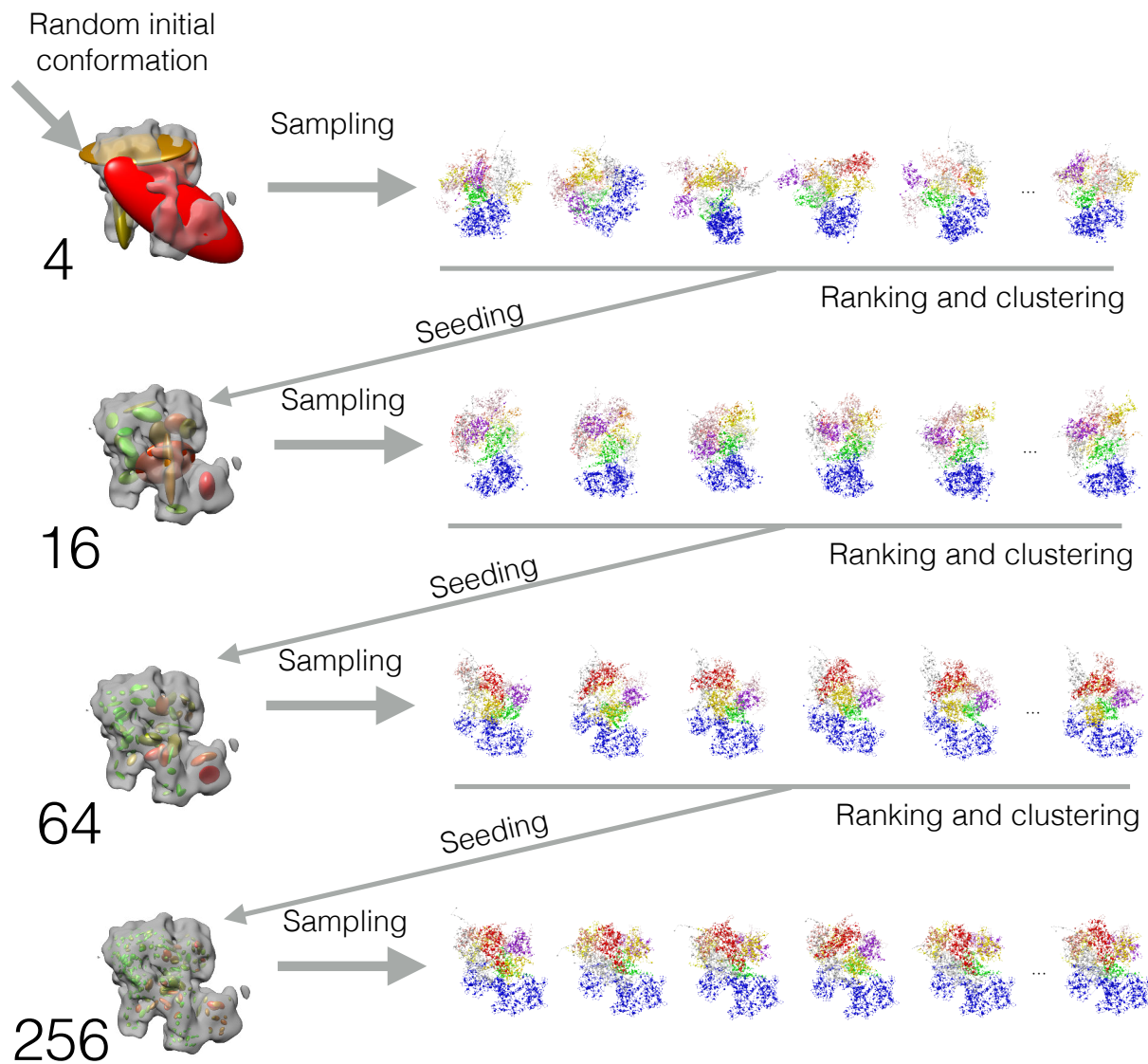
## Supplementary Figures



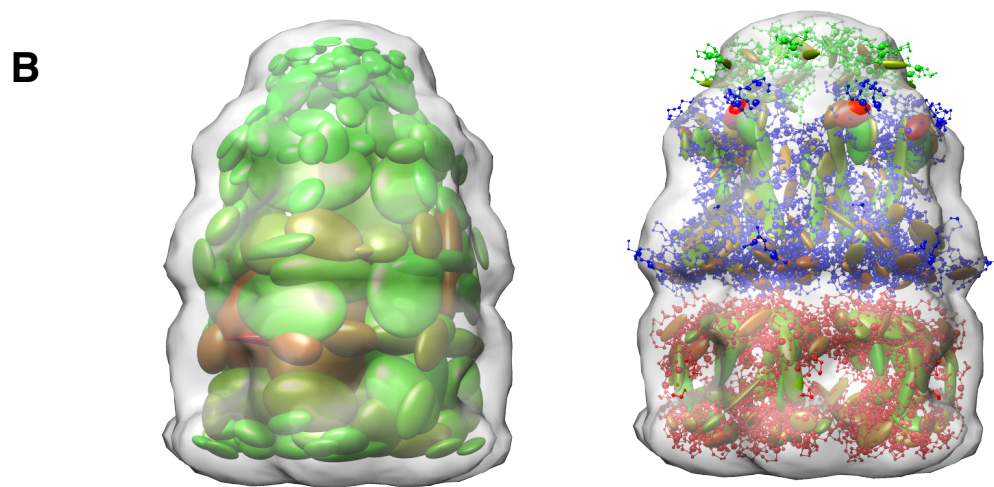
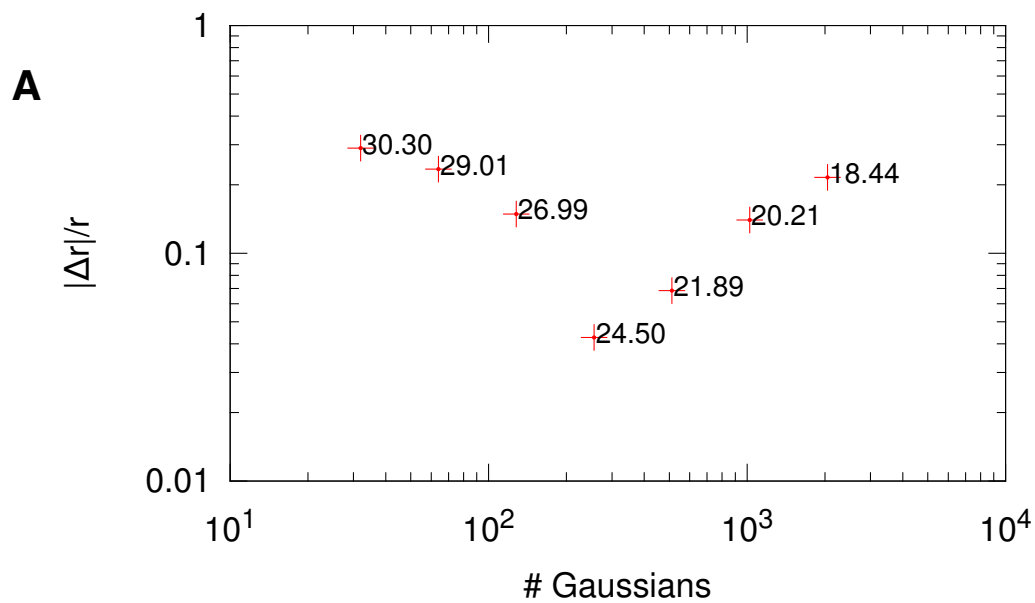
**Figure S1: Selection of the optimal data-GMM.** (A) Fourier Shell Correlation of each GMM in the series with the original map. The horizontal dotted line represents the 0.5 threshold, and the vertical dotted line marks the inverse of the resolution of the original map (11.5  $\text{\AA}$ ). (B) Ellipsoid representation of the GMM series, together with the absolute relative deviation  $|\Delta r|/r$  between the data-GMM resolution and the density map resolution. The optimal GMM, defined as the GMM which minimizes  $|\Delta r|/r$ , is highlighted.



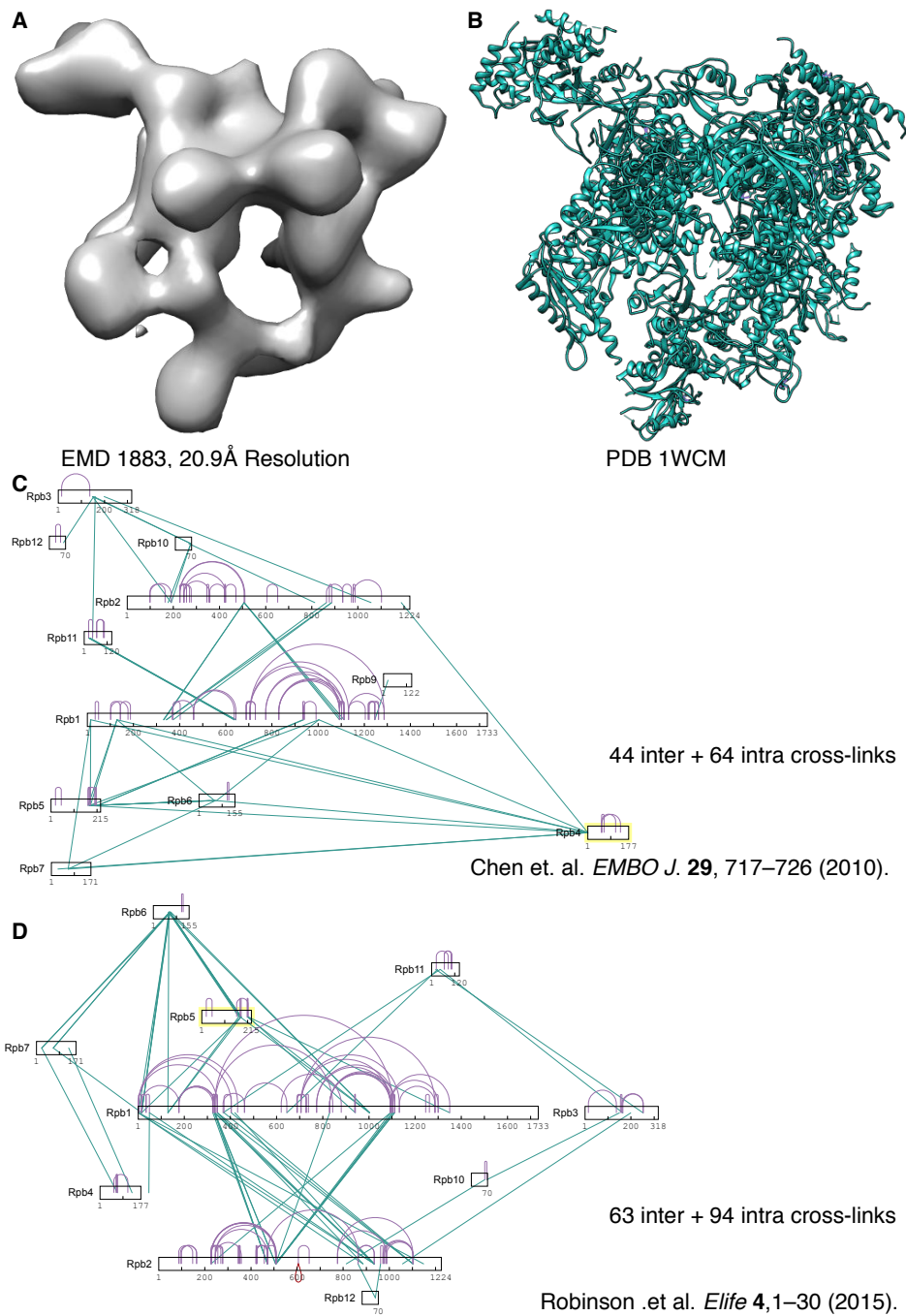
**Figure S2: Likelihood and score.** (A) Likelihood expressed as the product of individual data-component terms. (B) Representation of the evolution of the likelihood as a function of the overlap between the data- and model-GMMs. (C) Scores associated with each conformation.



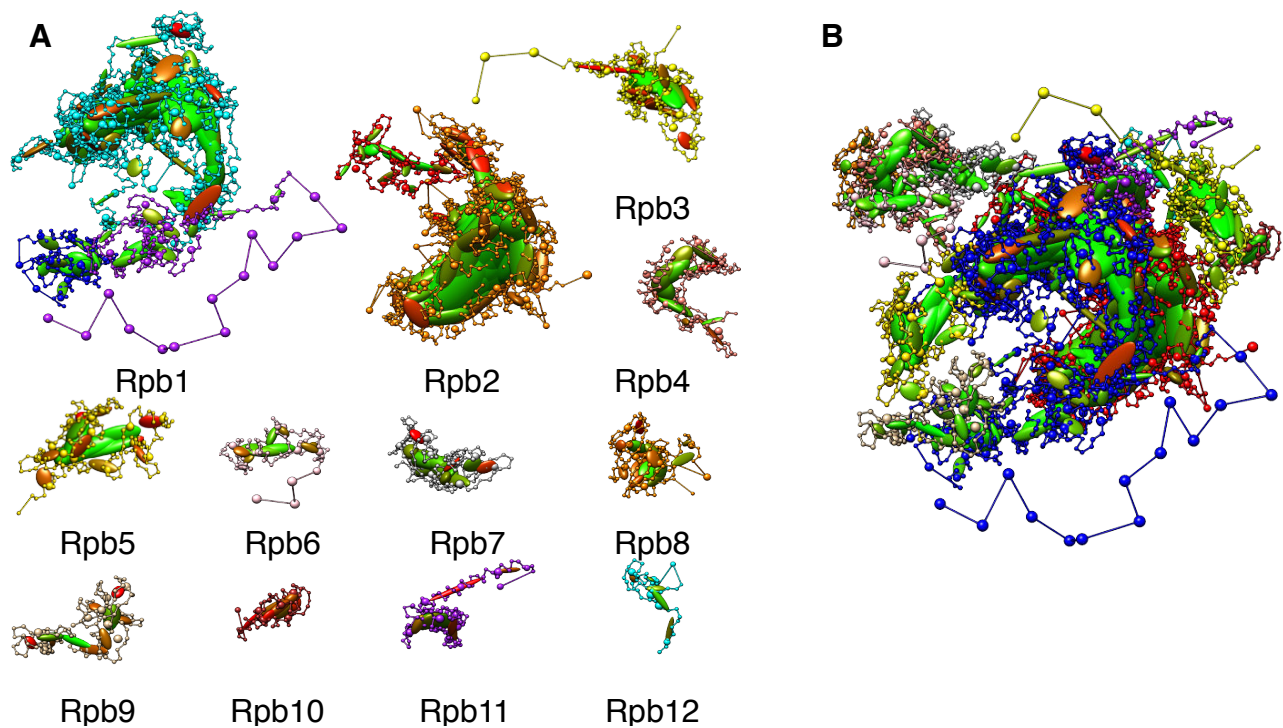
**Figure S3: Schematic representation of the iterative sampling protocol.** For the four iterations of the exosome modeling, the data-GMM is represented as a set of ellipsoids along with resulting models that were selected as seed for the next iteration. A subset of models as structurally diverse as possible is selected from the 100 best scoring models produced at every iteration and used as starting configurations for the following iteration.



**Figure S4: Modeling of the GroEL/ES complex.** (A) Absolute relative deviation  $|\Delta r|/r$  between the data-GMM resolution and the density map resolution as a function of the number of Gaussians of the data-GMM; (B) The model-GMM is represented as a set of ellipsoids (left) and representation of the best scoring model (right). Each subunit is represented as a string of connected beads, where the diameter of the beads defines the coarse-graining resolution.

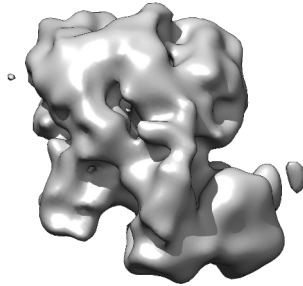


**Figure S5: Modeling of the RNA polymerase II.** (A) The cryo-EM map (EMDB code 1883) used in the modeling. (B) The prior structure (PDB code 1WCM) used to initialize the coordinates of the beads. (C, D) Visual representation of the two XL-MS datasets used. The sequence of each subunit is represented as a rectangle, with numbers indicating the residue indexes in the sequence. Cross-links between two residues are represented as a line connecting the two residues. The intra-molecular and inter-molecular cross-links are represented as curved purple lines and straight green lines, respectively.



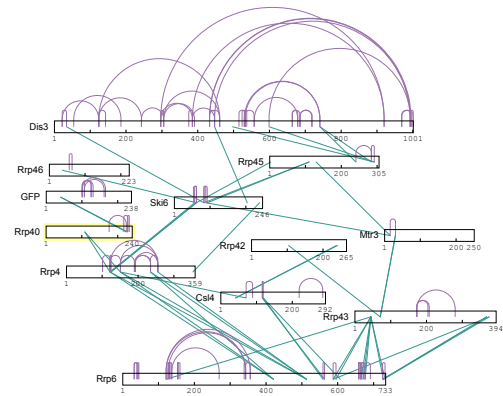
**Figure S6: Model representation of the RNA polymerase II.** (A) Representation of each subunit, with different colors for each individual rigid body. (B) Representation of the best scoring model. Each subunit is represented as a string of connected beads, where the diameter of the beads defines the coarse-graining resolution. The model-GMM is represented as a set of ellipsoids.

**A** EMDB 3367 ~12Å resolution



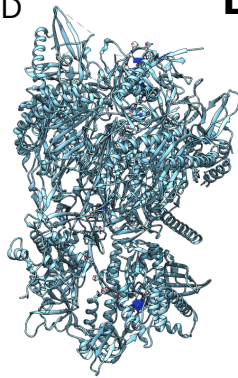
Liu JJ, et al. *cell research* 26 822-837 (2016)

**B** 26 inter + 72 intra cross-links



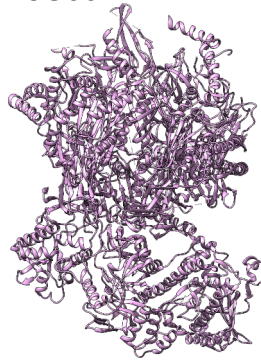
Shi, Y. et al. *Nature Methods* 12, 1135–1138 (2015).

**C** PDB 4IFD



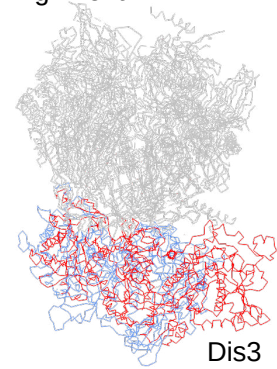
Makino, D.L., Conti, E. (2013) *Nature* 495: 70-75

**D** PDB 5G06



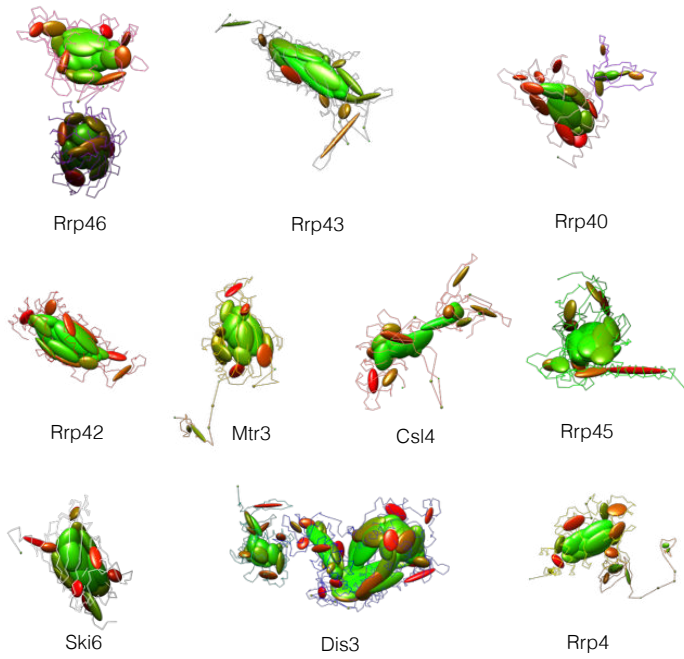
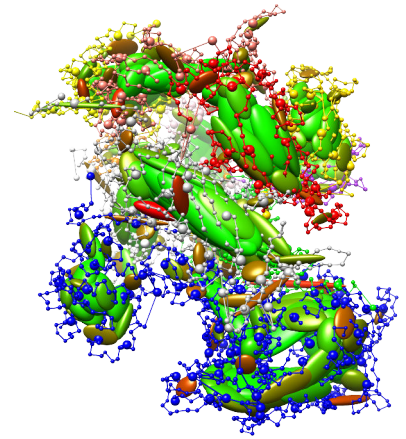
Liu JJ, et al. *cell research* 26 822-837 (2016)

**E** Alignment



**Figure S7: Modeling of the exosome.** (A) The cryo-EM map (EMDB code 3367) and (B) a visualization of the XL-MS dataset used in the modeling. (C) Prior structure (PDB code 4IFD), (D) reference structure (PDB code 5G06), and (E) an alignment of the two, in which the conformational change between RNA-bound (blue) and RNA-free (red) structures is highlighted.



**A****B**

**Figure S8: Model representation of the exosome.** (A) Representation of each subunit, with different colors for each individual rigid body. (B) Representation of the best scoring model. Each subunit is represented as a string of connected beads, where the diameter of the beads defines the coarse-graining resolution. The model-GMM is represented as a set of ellipsoids.

	2UZX:A 289		2UZX:B 727	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.5	0.5, 3.3	0.3	0.5, 0.6
C1	0.8	0.9, 6.7	0.5	0.6, 0.8

	3R5D:A 347		3R5D:B 347		3R5D:C 347	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.7	0.9, 2.0	0.9	1.6, 0.4	0.6	0.8, 2.8
C1	0.8	1.2, 1.5	0.9	1.4, 2.4	0.7	0.6, 3.8

	1CS4:A 225		1CS4:B 212		1CS4:C 394	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	1.3	2.0, 1.6	1.1	0.7, 1.3	0.4	0.2, 0.3
C1	1.3	1.0, 2.9	0.9	0.8, 0.3	0.9	0.7, 0.4

	2WVY:A 737		2WVY:B 737		2WVY:C 737	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.4	0.5, 1.1	0.3	0.3, 0.0	0.5	0.4, 0.4
C1	0.6	0.9, 1.0	0.3	0.2, 0.4	0.4	0.4, 0.4

	2DQJ:H 114		2DQJ:L 107		2DQJ:Y 129	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.7	1.2, 0.2	1.4	0.8, 6.3	1.2	1.1, 2.0
C1	0.9	1.2, 1.2	2.0	2.3, 6.5	1.3	1.5, 1.7
C2	13.6	1.0, 179.0	3.6	3.0, 14.3	1.1	1.0, 1.7
C3	1.6	1.0, 5.8	12.6	3.7, 169.7	1.4	1.1, 2.2

	1VCB:C 160		1VCB:A 118		1VCB:B 112	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.7	0.4, 5.3	0.6	0.9, 1.5	0.8	1.1, 1.0
C1	0.6	0.6, 4.5	0.7	0.9, 1.8	1.1	1.7, 0.8

	2GC7:C 105		2GC7:D 147		2GC7:A 386		2GC7:B 131	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	1.1	0.9, 0.8	0.5	0.6, 0.5	0.7	0.8, 1.2	0.6	0.5, 1.0
C1	1.6	1.5, 0.9	0.6	0.2, 0.8	1.1	0.9, 2.9	0.5	0.8, 0.3

	2BO9:A 308		2BO9:C 308		2BO9:B 222		2BO9:D 222	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.8	0.9, 0.3	0.2	0.2, 0.2	0.8	1.0, 2.5	0.7	0.8, 1.3
C1	0.9	0.8, 0.5	0.7	1.1, 0.1	1	0.6, 4.0	0.7	0.9, 1.9

	2BBK:H 355		2BBK:J 355		2BBK:L 125		2BBK:M 125	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.9	1.2, 0.7	1.3	1.9, 1.0	1.4	1.8, 0.7	1.4	2.1, 1.4
C1	1.2	1.2, 0.8	1.2	1.7, 0.9	1.1	1.5, 0.6	2.2	3.2, 1.9

	1GPQ:A 135		1GPQ:B 135		1GPQ:C 129		1GPQ:D 129	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.9	1.4, 1.0	1.0	1.7, 0.4	1.0	0.4, 1.4	1.1	1.0, 2.0
C1	0.9	1.0, 1.4	1.2	2.0, 0.5	1.2	1.0, 1.8	1.3	1.6, 2.4

	3V6D:A 556		3V6D:B 428		3V6D:P 21		3V6D:T 27	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.7	1.0, 1.6	0.6	1.0, 0.7	2.9	1.9, 32.6	2.2	1.7, 8.8
C1	1.0	1.2, 3.9	0.7	0.6, 2.4	4.0	2.0, 70.4	3.4	1.7, 23.8
C2	22.2	38.5, 1.1	16.1	27.9, 1.6	19.5	32.4, 90.0	19.7	32.7, 51.2

	3SFD:A 622		3SFD:B 252		3SFD:C 140		3SFD:D 103	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.3	0.2, 0.1	1.0	1.3, 1.9	1.6	1.9, 1.3	1.1	1.7, 4.1
C1	0.3	0.2, 0.1	1.0	1.3, 1.9	1.6	1.8, 1.1	1.1	1.7, 4.1

	3PDU:A 287		3PDU:B 287		3PDU:C 287		3PDU:D 287	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.5	0.7, 0.5	0.9	1.1, 1.2	0.7	1.2, 0.4	0.7	1.1, 0.4
C1	0.6	0.7, 0.7	1.2	1.6, 2.0	0.9	1.5, 0.3	0.8	1.3, 0.5

	3NVQ:A 590		3NVQ:E 590		3NVQ:B 476		3NVQ:F 476	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	0.5	0.7, 0.2	0.5	0.7, 1.3	0.3	0.2, 0.2	0.7	0.8, 0.4
C1	0.6	0.8, 0.4	0.8	0.7, 1.8	0.2	0.2, 0.2	0.5	0.7, 0.2

	2Y7H:A 464		2Y7H:B 529		2Y7H:C 529		2Y7H:D 20		2Y7H:E 20	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	1.0	1.5, 0.7	1.1	1.4, 1.9	0.8	1.0, 1.4	2.0	3.1, 9.7	2.0	3.0, 10.6
C1	1.5	2.6, 1.0	1.3	2.0, 1.3	0.7	1.1, 1.2	3.1	3.4, 17.5	3.1	2.7, 14.2

	1SUV:A 639		1SUV:B 639		1SUV:C 329		1SUV:D 329		1SUV:E 345		1SUV:F 345	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	2.9	4.9, 0.6	2.8	4.8, 0.1	3.3	5.6, 0.6	3.4	5.8, 0.5	2.9	5.0, 0.2	3.0	5.2, 0.1
C1	2.9	4.9, 0.6	3	5.2, 0.1	3.3	5.6, 0.6	3.4	5.8, 0.5	2.9	5.0, 0.2	3.0	5.2, 0.1

	1Z5S:A 156		1Z5S:B 82		1Z5S:C 172		1Z5S:D 83	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	1.0	1.6, 1.0	12.0	1.2, 171.0	0.7	0.9, 0.7	1.4	2.3, 1.4
C1	1.3	2.0, 1.7	12.3	2.2, 178.9	1.0	1.4, 0.9	1.3	1.9, 4.1

	3LU0:A 329		3LU0:B 329		3LU0:C 1342		3LU0:D 1407		3LU0:E 91	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	10.1	10.2, 35.5	9.6	7.5, 30.5	4.8	3.6, 1.1	2.6	3.1, 0.3	5.5	7.4, 20.7
C1	10.3	10.9, 33.7	9.8	7.3, 24.1	4.2	3.2, 0.8	2.8	3.1, 0.4	11.3	13.5, 34.6
C2	9.9	11.0, 30.3	27.3	21.2, 50.4	5.5	3.2, 1.6	2.3	3.5, 0.3	43.5	72.6, 151.7

	1MDA:A 103		1MDA:B 103		1MDA:H 368		1MDA:J 368		1MDA:L 121		1MDA:M 121	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	10.3	2.0, 178.4	10	3.1, 172.0	1.4	2.2, 0.4	2.1	2.7, 2.5	2.4	0.8, 2.3	1.7	2.3, 2.8
C1	10.4	3.6, 176.1	1.8	2.4, 1.2	1.9	2.8, 1.4	2.1	3.4, 0.3	2.0	3.4, 3.0	9.4	4.0, 174.0

	3PUV:A 381		3PUV:B 381		3PUV:E 378		3PUV:F 514		3PUV:G 296	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	15.7	17.3, 14.6	24.7	16.9, 134.1	0.6	0.7, 1.9	1.3	1.7, 1.2	1.3	2.1, 0.4
C1	16.1	17.6, 17.7	25.1	18.1, 136.7	1.1	0.9, 2.7	1.7	2.0, 0.7	1.5	2.2, 1.4

	1TYQ:A 418		1TYQ:B 394		1TYQ:C 372		1TYQ:D 300		1TYQ:E 178		1TYQ:F 168		1TYQ:G 151	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	1.0	0.7, 3.8	3.1	2.6, 4.9	1.2	0.6, 2.3	21.0	7.7, 144.5	14.7	1.6, 175.7	12.1	14.1, 154.7	2.4	2.9, 5.0
C1	0.9	1.0, 0.9	3.6	3.3, 5.4	1.1	1.2, 1.2	21.3	7.9, 144.8	14.9	1.5, 178.4	13.1	16.0, 154.5	2.6	3.4, 8.1
C2	28.3	36.0, 147.4	3.3	2.7, 5.3	1.4	1.0, 2.7	32.2	49.0, 140.7	0.6	0.6, 1.0	10.2	12.4, 30.6	3.3	4.7, 9.7
C3	28.5	36.5, 148.6	1.6	2.3, 3.6	0.8	1.0, 1.0	31.9	48.9, 141.0	1.0	1.5, 0.4	10.1	12.8, 30.5	3.0	2.9, 10.8

**Table S1: Detailed results of the benchmark.** Rmsd (in Ångstrom) and Placement scores (PS, in Ångstrom and angular degrees) of every subunit for the best-rmsd model and the best scoring model of each structural cluster. By construction, cluster C1 contains the absolute best scoring model, whose properties are reported in **Table 2**. Each column is titled with the PDB code, chain id, and number of residues. For instance, 2B09:A 308 refers to chain A of PDB id 2B09, which is composed of 308 residues.

rank	rank	rmsd [Å]	p(10)	CC	PS [Å,°]
best rmsd	99	8.6	0.98	0.85	4.4, 18.2
C1	0	9	0.98	0.85	5.4, 18.1
C2	1	30.9	0.53	0.82	5.2, 103.3
C3	84	25	0.55	0.82	5.3, 94

	groEL-trans (524)		groEL-cis (524)		groES (97)	
	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]	rmsd [Å]	PS [Å,°]
best rmsd	4.4	2.9, 0.3	9	5.7, 11.8	18.3	4.8, 148.9
C1	6.2	5.6, 2.9	9.1	4.6, 12.4	17.4	7.9, 130.7
C2	43.2	2.9, 177.5	11.1	6.4, 19.7	21.8	10.6, 153.3
C3	33.4	3.1, 178.6	13.3	5.4, 19.7	20.4	16.3, 38.3

**Table S2: Results for the GroEL/GroES modeling.** (Top) Rmsd (in Ångstrom) Placement scores (PS, in Ångstrom and angular degrees) p(10) and cross-correlation (CC) for the best-rmsd structure, and the best scoring structures of clusters C1, C2 and C3. (Bottom) rmsd and PS contributions for the three subunits.

	Reference	rmsd [Å]				PS [Å,°]			
		Total	GroEL-cis	GroEL-trans	GroES	Total	GroEL-cis	GroEL-trans	GroES
This Work		9.0	9.1	6.2	17.4	5.4, 18.1	4.6, 12.4	5.6, 2.9	7.9, 130.7
gmfit	Kawabata 2008	14.7							
Attract-EM <sup>1</sup>	de Vries 2012	11.1	5.7	12.7	19.7				
IQP <sup>2</sup>	Zhang 2010	(best) 8.6				5.7,17.7			
MultiFit <sup>3</sup>	Lasker 2009					11.0,84			
ISD	Habeck 2017		3.5 (both GroEL rings)		2.0-20.0 <sup>4</sup>				
y-TEMPy <sup>5</sup>	Pandurangan 2015			11.7				8.9,21.0	
PowerFit	Zundert 2015		3.8-7.3	2.9-5.5	4.2-4.6				
gEMfitter	Hoang 2013		4.0-4.5	2.5-2.8	5.3-6.1				
COLORES <sup>6</sup>	Chacon 2002	Failed							
ADP_EM	Garzon 2006		<1.3 (both GroEL rings)		Failed				
Segger	Pintilie 2012		5.07	3.06	6.03				

**Table S3: Comparison of GroEL/ES modeling with other software.** Rmsd (in Ångstrom) and Placement scores (PS, in Ångstrom and angular degrees) for the whole complex (Total) or its subunits.

<sup>1</sup> Before refinement.

<sup>2</sup> Fit into a manual segmentation of GroEL-trans ring. Only best-rmsd is reported.

<sup>3</sup> Fine grained search.

<sup>4</sup> GroES was found to have two clusters with similar scores.

<sup>5</sup> Only GroEL-trans fitted.

<sup>6</sup> In the gEMfitter article it is reported that COLORES was not able to model the heptameric double-ring by fitting GroEL-cis and -trans separately.