Supplemental Figures

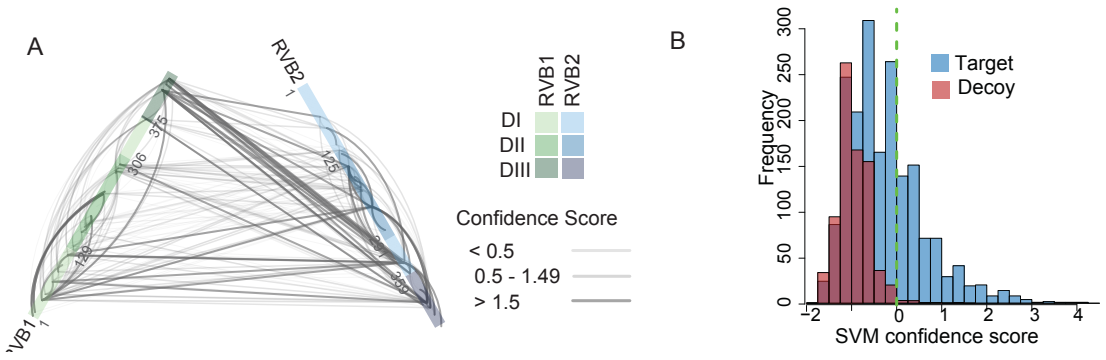

Figure S1 (related to Figure 1): XL-MS of Rvb1/Rvb2 and Rvb1/Rvb2/Ino80INS
(A) XL-MS data for Rvb1/Rvb2 alone. Rvb1 (green) and Rvb2 (blue) are shown with different shades representing DI, DII, and DIII (see legend, right). The confidence of the MS identification (SVM score) is represented by the thickness and transparency of the lines (see Supplemental Experimental Procedures).
(B) Decoy analysis of cross-links. The distribution of cross-linked residue pairs in which both cross-linked amino acids matched the target protein database are plotted in blue with respect to the SVM confidence score used for classification. Crosslinked residue pairs in which either residue matched the decoy (scrambled) sequence database are in red. An SVM score of 0.0 (marked by green dotted line) was taken as the acceptance criteria resulting in a residue-pair level false discovery rate of 1.7%.

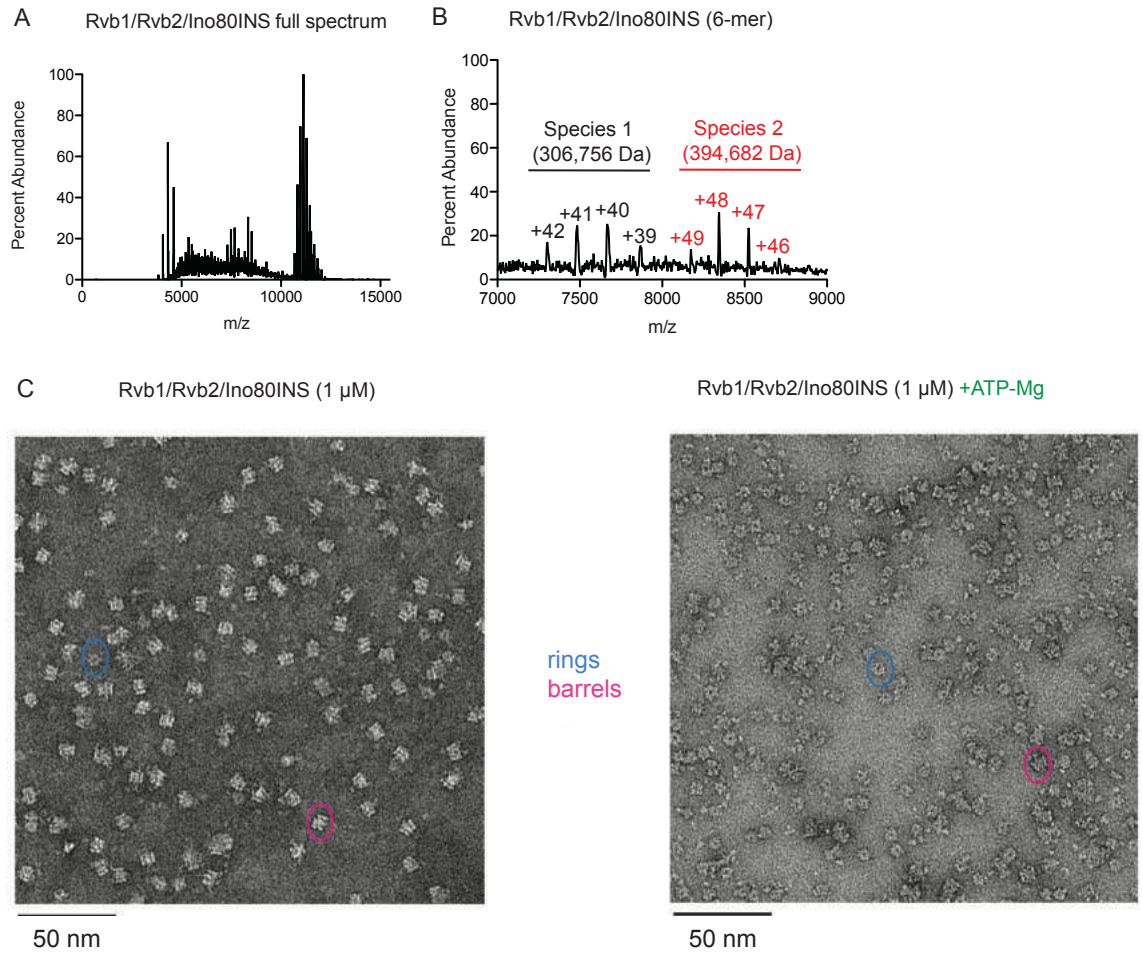

Figure S2 (related to Figure 2, Table S2): Native-MS and negative stain EM of the Rvb1/Rvb2/Ino80INS complex.

(A) Full raw spectrum of Rvb1/Rvb2/Ino80INS by native-MS.

(B) Charge states for two species of the Rvb1/Rvb2/Ino80INS sample that contains a Rvb1/Rvb2 hexamer. Species 1 (black) has a mass that is most consistent with a Rvb1/Rvb2 heterohexamer while Species 2 (red) has a mass that is most consistent with a Rvb1/Rvb2 heterohexamer bound to one copy of Ino80INS.

(C) Left, negative stain image of 2 µM Rvb1/Rvb2/Ino80INS complexes. Right, the same sample of Rvb1/Rvb2/Ino80INS with 20 µM ATP-Mg added, and after a 20-minute incubation at room temperature. For each image, barrel-like structures are highlighted by pink circles and ring-like structures are highlighted by blue circles. Scale bar below each image is equal to 50 nm.

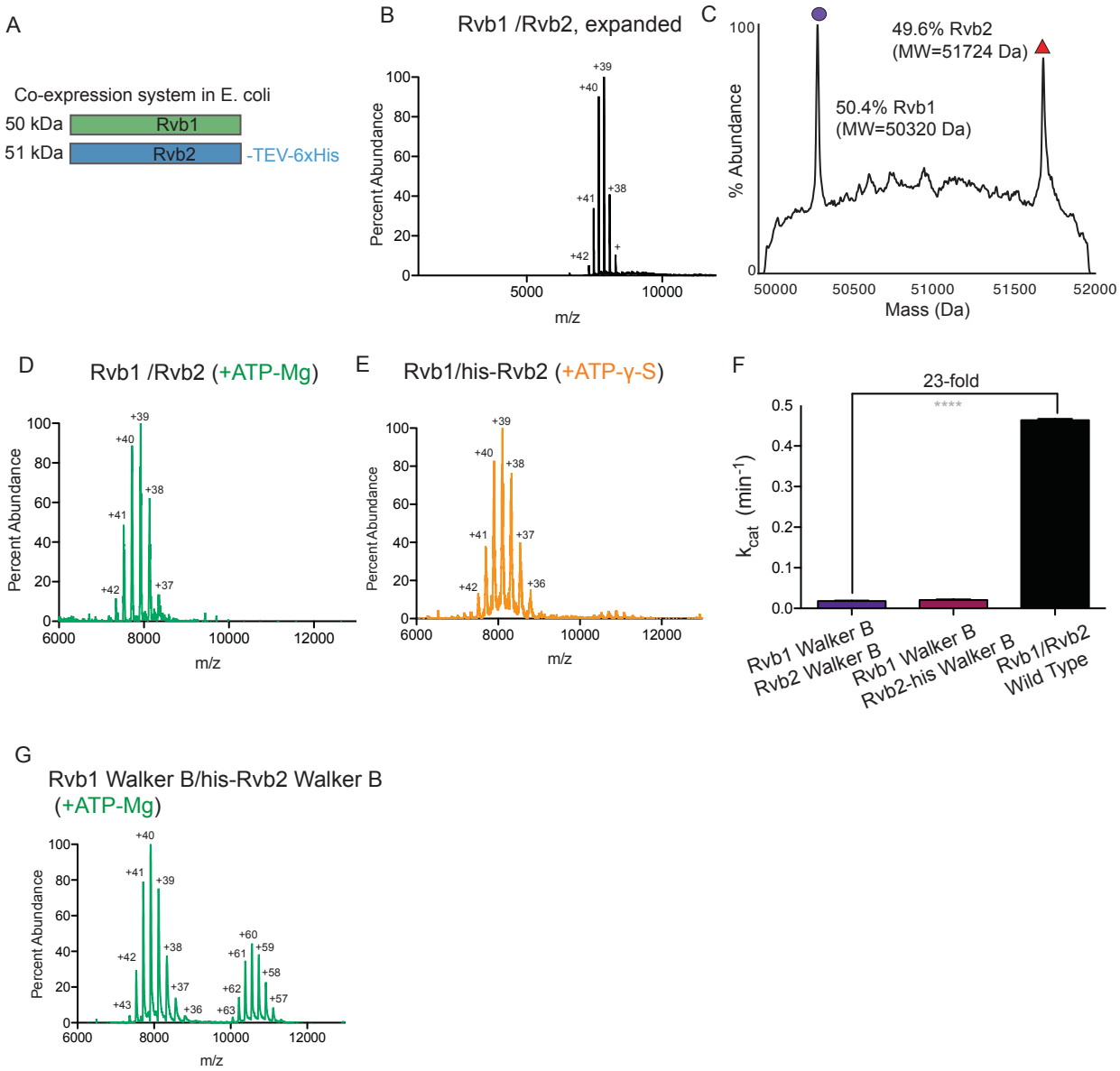

Figure S3 (related to Figure 3, Table S3, and Table S4): Native-MS of Rvb1/Rvb2 complexes.
(A) Co-expression system used for experiments with Rvb1/Rvb2 alone. The his-tag on Rvb2 is removed by tev cleavage unless otherwise noted. See also experimental procedures.
(B) Expanded view of Rvb1/Rvb2 spectrum.
(C) Deconvoluted ESI-mass spectrum of intact Rvb1/Rvb2 complexes. 0.25 µM of Rvb1/Rvb2 was denatured by addition of 50% acetonitrile and 10% acetic acid. The percent abundance of each peak was measured by taking the area under each peak.
(D) Raw spectrum of 1 µM Rvb1/Rvb2 with 20 µM ATP-Mg added.
(E) Raw spectrum of 1 µM Rvb1/his-Rvb2 with 20 µM ATP-γ-S added.
(F) Maximal rate constants (k_{cat}) for ATP hydrolysis activity by Walker B mutants of Rvb1/Rvb2. k_{cat} values for Rvb1 D311N/Rvb2 D296N (purple bar) and Rvb1 D311N/his-Rvb2 D296N (red bar) are plotted next to k_{cat} value of wild-type Rvb1/Rvb2. Experiments were performed using 10 µM Rvb monomer and saturating ATP. Bars represent the mean ± SEM of 3 replicates. **** = $p < 0.0001$.
(G) Raw spectrum of 1 µM Rvb1 Walker B/his-Rvb2 Walker B with 20 µM ATP-Mg added.

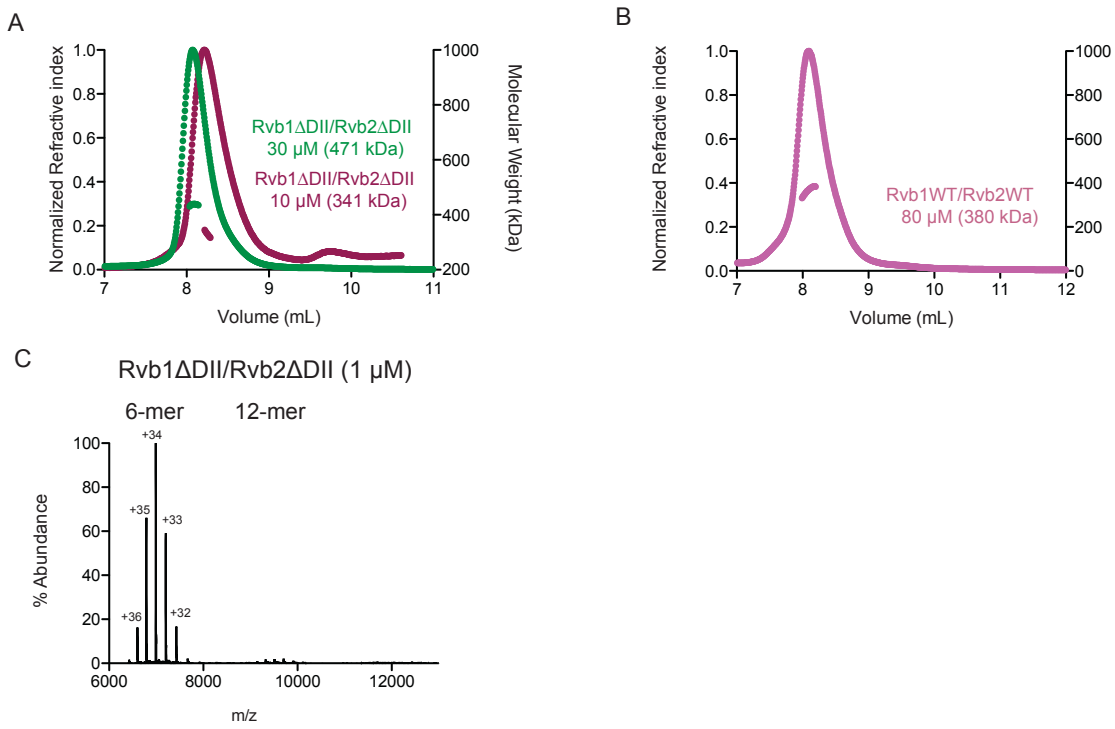

Figure S4 (related to Figure 4 and Table S5): SEC-MALS of Rvb1/Rvb2 complexes.
(A) SEC-MALS of Rvb1 Δ DII/Rvb2 Δ DII complexes at 10 μM (purple) and 30 μM (green). The molecular weights shown are an average of the calculated molecular weight values across the main peak.
(B) SEC-MALS of wild-type Rvb1/Rvb2 at 80 μM, along with calculated molecular weight.
(C) Native-MS raw spectrum of 1 μM Rvb1 Δ DII/Rvb2 Δ DII.

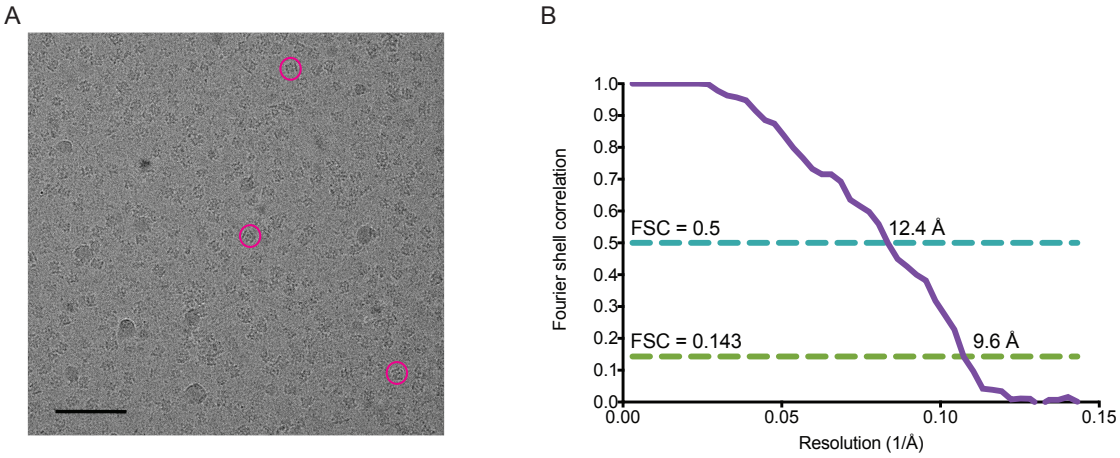

Figure S5: Cryo-EM recontruction of the Rvb1/Rvb2/Ino80INS complex (related to Figure 5).
(A) Representative cryo-EM micrograph showing mono-dispersed and relatively homogeneous Rvb1/Rvb2Ino80INS complex particles. Scale bar is equal to 50 nm.
(B) Fourier shell correlation (FSC) curve generated after refining two separate half-maps using Relion 3D auto-refine.

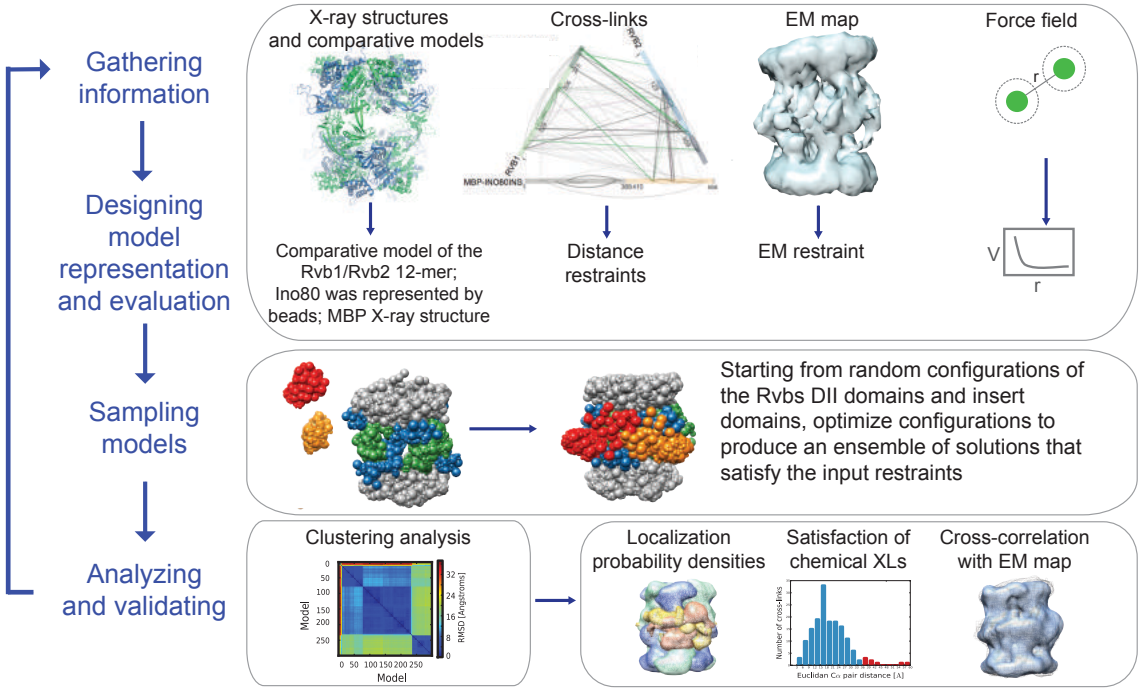

Figure S6: The 4-stage scheme for integrative structure modeling of the Rvb1/Rvb2/Ino80INS complex (related to Figure 6, Figure S7, Table S6).
Our integrative approach proceeds through four stages: (1) gathering of data, (2) representation of subunits and translation of the data into spatial restraints, (3) configurational sampling to produce an ensemble of models that satisfies the restraints, and (4) analysis of the ensemble.

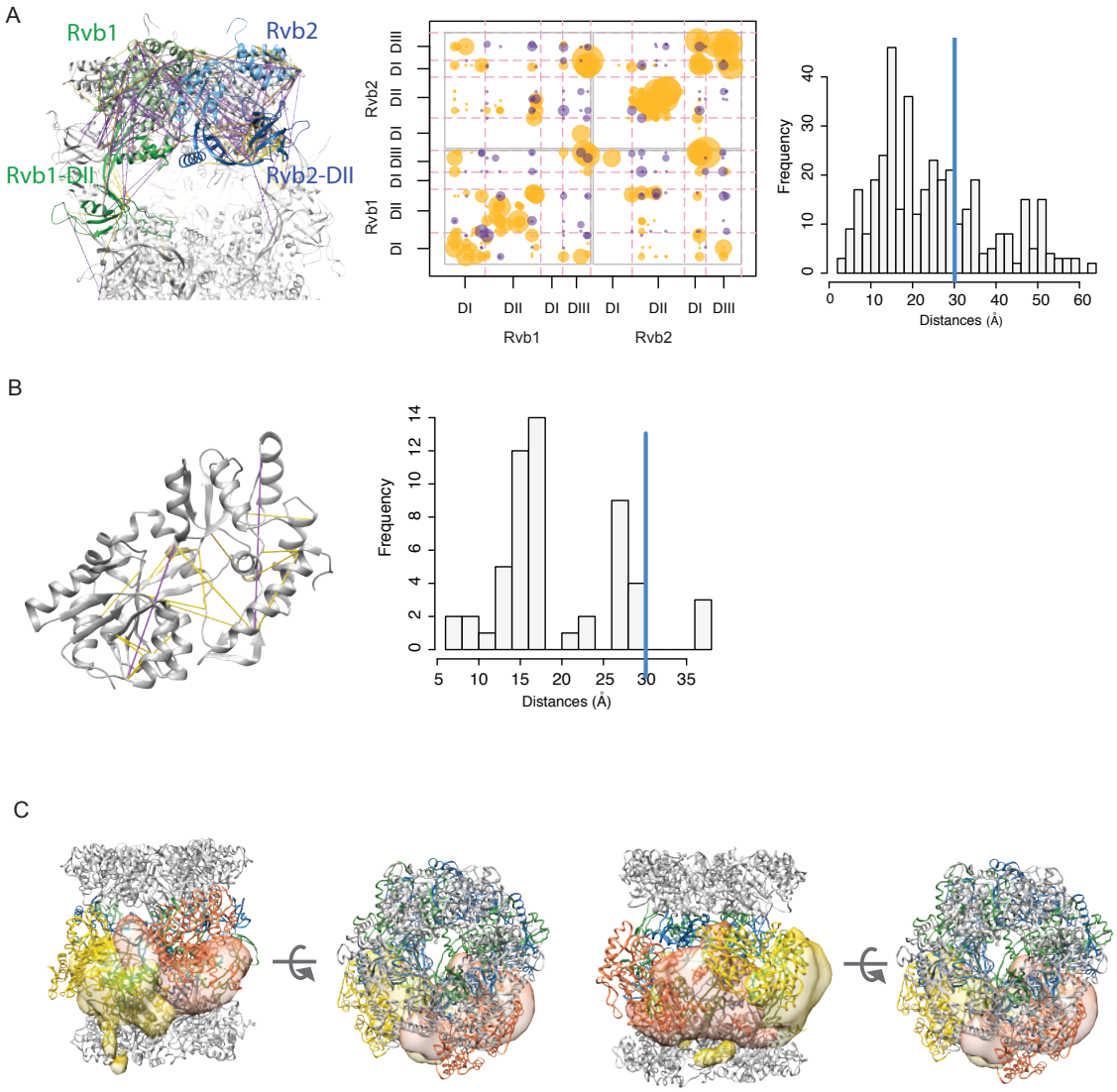

Figure S7: Analysis of integrative structure modeling of Rvb1/Rvb2/Ino80INS complex (related to Figure 6, Figure S6, Table S6).

(A) Analysis of the Rvb-Rvb crosslinked residue pairs in the Rvb1/Rvb2/Ino80INS dataset.
Left: 132 satisfied XLs (58%; yellow) and 94 violated XL's (42%; purple) are mapped onto a model of the Rvb1/Rvb2 dodecamer, using a 30 Å cutoff. For clarity, only XLs for the first subunit of Rvb1 (green) and Rvb2 (blue) are shown. Middle: the yellow and purple circles correspond to the XLs that are satisfied or violated in the X-ray structure, respectively. The size of each circle is proportional to the spectral count of the cross-link. Right: histogram of the distances between crosslinked residues in the X-ray structure. The vertical blue line represents the 30 Å cutoff used to define satisfied and violated XLs.
(B) Analysis of the cross-linked residues in MBP. Left: Mapping MBP-MBP crosslinks onto the X-ray structure of shows that two of the MBP XLs are violated (in purple). This observation indicates that these cross-links are inter-molecular. Right: Histogram of the distances between cross-linked residues in the X-ray structure. The vertical blue line represents the 30 Å cutoff.
(C) Structure of the Rvb1/Rvb2/Ino80INS complex in Cluster 2 (left) and Cluster 3 (right). The Rvb1/Rvb2 dodecamer is represented by the centroid structure of the cluster. The top and bottom rings are colored grey, while the Rvb1 and Rvb2 domains are colored green and blue, respectively. The Ino80INS domains are shown as localization densities (orange and yellow). The MBPs are shown in ribbon representation (orange and yellow).

Supplemental Tables

Table S1 (related to Figure 1 and Figure S1): Summary of XL-MS data for Rvb1/Rvb2 and Rvb1/Rvb2/Ino80INS

Rvb1/Rvb2 crosslinks observed				
	Rvb1	Rvb2	Ino80	Total [a]
Rvb1	111			
Rvb2	156	106		
Ino80	na	na	na	
Total	267	106	0	373 (2300)
Rvb1/Rvb2/Ino80INS crosslinks observed				
	Rvb1	Rvb2	Ino80	Total [a]
Rvb1	69			
Rvb2	85	73		
Ino80	42	38	42	
Total	196	111	42	349 (1420)

[a] Numbers of unique cross-linked residue pairs. Numbers in parentheses are spectral counts.

Table S2 (related to Figure 2, Figure S2): Analysis of the Rvb1/Rvb2/Ino80INS complex by native mass spectrometry.

Rvb1/Rvb2/Ino80INS Hexameric Species							
Hexamer Species 1				Hexamer Species 2			
Mass observed (Da)	Standard devation	Predicted stoichiometry	% Error	Mass observed (Da)	Standard deviation	Predicted stoichiometry	% Error
306,756	73	3 Rvb1 3 Rvb2	0.7%	394,682	8232	3 Rvb1 3 Rvb2 1 MBP-Ino80INS	3.0%
Rvb1/Rvb2/Ino80INS Dodecameric Spcies							
Dodecameric Species 1				Dodecameric Species 2			
Mass observed (Da)	Standard devation	Predicted stoichiometry	% Error	Mass observed (Da)	Standard deviation	Predicted stoichiometry	% Error
790,087	387	6 Rvb1 6 Rvb2 2 MBP-Ino80INS	3.0%	863,256	917	6 Rvb1 6 Rvb2 3 MBP-Ino80INS	2.4%
Rvb1/Rvb2/Ino80INS + ATP-Mg							
Hexameric Species				Dodecameric Species			
Mass observed (Da)	Standard devation	Predicted stoichiometry	% Error	Mass observed (Da)	Standard deviation	Predicted stoichiometry	% Error
308813	47	3 Rvb1 3 Rvb2 5 nucleotides	0.9%	n.a.	n.a.	n.a.	n.a.

Table S3 (related to Figure 3, Figure S3): Predicted molecular weights for different stoichiometries of Rvb1/Rvb2 hexamers

Rvb1:Rvb2	Predicted MW (Da)	% Error (based on observed mass)
3:3	306930	0.19%
4:2	305526	0.46%
2:4	308334	0.46%
1:5	309738	0.91%
5:1	304122	0.91%

Table S4 (related to Figure 4 and Table S3): Molecular mass distribution of Rvb1/Rvb2 complexes by native mass spectrometry

[a] Standard deviation based on n=2

Sample	Hexamer					Dodecamer				
	% of total area	Expected mass (Da)	Observed mass (Da)	Standard deviation (Da) [a]	# of NT bound [b]	% of total area	Expected mass (Da)	Observed mass (Da)	Standard deviation (Da) [a]	# of NT bound [b]
Rvb1/Rvb2	100	306,930	306,352	40	0	0	n.a.	n.a.	n.a.	n.a.
Rvb1/Rvb2 +ATP-Mg	100	306,930	308,693	250	4-5	0	n.a.	n.a.	n.a.	n.a.
Rvb1/his-Rvb2	0	313,692	n.a.	n.a.	n.a.	100	627,384	627,407	102	1-2
Rvb1/his-Rvb2 + ATP-Mg	14	313,692	315,729	71	4-5	86	627,384	631,079	424	8-9
Rvb1/his-Rvb2 + ADP-Mg	100	313,692	315,745	210	5-6	0	627,384	n.a.	n.a.	n.a.
Rvb1/his-Rvb2 + AMP-PNP-Mg	100	313,692	315,452	175	4-5	0	627,384	n.a.	n.a.	n.a.
Rvb1/his-Rvb2 + ATP-γ-S-Mg	100	313,692	315,796	220	2-3	0	627,384	n.a.	n.a.	n.a.
Rvb1 Walker B/ his-Rvb2 Walker B	83	313,686	314,564	83	1-2	17	627,372	629,314	129	5-6
Rvb1 Walker B/ his-Rvb2 Walker B+ ATP-Mg	80	313,686	316,884	494	6-7	20	627,372	633,685	494	13-14

[b] Number of nucleotides bound is estimated by dividing the difference between observed and expected mass by the mass of ATP Mg (531 Da), ADP-Mg (451 Da), or AMP-PNP-Mg (528 Da).

Table S5 (related to Figure 4 and Figure S4): Molecular weight predictions from MALS experiments

Sample	Observed molecular weight (kDa)	Expected molecular weight for hexamer (kDa)[a]	Expected molecular weight for dodecamer (kDa)[a]
Rvb1/Rvb2	345 kDa	307 kDa	614 kDa
Rvb1/his-Rvb2	593 kDa	314 kDa	627 kDa
Rvb1ΔDII, Rvb2ΔDII	341 kDa	237 kDa	474 kDa

[a] Expected molecular weights were calculated for a 3:3 stoichiometry of Rvb1:Rvb2 for the hexamer and 6:6 stoichiometry for Rvb1:Rvb2 for the dodecamer.

Table S6 (related to Figure 6, S6, S7): Summary of integrative structure determination of the Rvb1/Rvb2/Ino80INS complex

Modeling programs	• Python Modeling Interface (PMI), version b47d68e • Integrative Modeling Platform (IMP, version 2.5) • MODELLER 9.13
Homology detection and structure prediction	• HHPred, PSIPRED, DISOPRED
Structured components	• Rvb1/Rvb2 dodecamer comparative models based on PDB 4WVY • MBP X-ray structure (PDB 4EFB) • Ino80 helices residues 1137-1144, 1148-1156, 1193-1201, 1127-1237 and 1251-1261.
Unstructured components	• Rvb1 residues: 1-51, 143-165, 211-238 • Rvb2 residues: 142-155, 205-229, 249-266, 444-354 • Ino80INS, expect helices defined as structured parts.
Stoichiometry	6:6:2(=Rvb1:Rvb2:MBP-Ino80INS)
Spatial restraints	Sequence connectivity, protein excluded volume, chemical cross-links and 3D electron microscopy *See Supplemental Experimental Procedures.*
Sampling method	Replica Exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm. 32 replicas were used in 20 independent runs, at the temperature range of 1.0 - 2.5.
Monte Carlo moves	- Random translation and rotation of rigid bodies (up to 1.5 Å and 0.03 radians, respectively). - Random translation of individual beads in the flexible segments (up to 2 Å).
Clustering analysis	3 clusters of 196, 180 and 124 structures. See figures 6 and S7.
Sampling exhaustiveness	p-value = 0.15
Precision of the clusters	• Cluster 1: 25 Å (all components); 15.9 Å (Rvbs) 39.3 Å (Ino80ins) • Cluster 2: 26.7 Å (all components); 14.8 Å (Rvbs) 40.2 Å (Ino80ins) • Cluster 3: 12.1 Å (all components); 6.2 Å (Rvbs) 14.3 Å (Ino80ins)
Chemical cross-links satisfied in the clusters	• Cluster 1: 97% • Cluster 2: 94% • Cluster 3: 91% • All clusters: 98%
3D EM map cross-correlation coefficients (ccc)	• Cluster 1: ccc = 0.94 • Cluster 2: ccc = 0.96 • Cluster 3: ccc = 0.87
Visualization and plotting	UCSF Chimera 1.10, XlinkAnalyzer, R

Supplemental Experimental Procedures

Protein expression and purification

Rvb1 and Rvb2 were co-expressed on a pet28a-DUET plasmid, a kind gift from the Hopfner laboratory at the Ludwig Maximilian University in Munich. The Ino80 insertion (Ino80INS, residues 1022-1294 of the yeast Ino80 ATPase) or the Swr1 insertion (Swr1INS, residues 1003-1237 of the yeast Swr1 ATPase) were cloned into the pMAL vector (NEB) containing a C-terminal His_{x8} tag, a gift from the Lim lab at UCSF. For each prep, the appropriate constructs were freshly transformed into Rosetta (DE3) *E. coli* (Novagen). 1 L cultures of transformed cells were grown to ~0.3-0.4 at 37°C in 2X LB media with 1X NaCl, then switched to 18°C and induced at OD_{600} = 0.6-0.7 with 0.3 mM IPTG for approxminately 18 hours. Cells were then harvested, resuspended in lysis buffer (25 mM HEPES pH 7.5, 300 mM KCl, 7.5 mM imidazole, 10% glycerol v/v, 2 μg/mL aprotinin, 1 μg/mL peptstatin A, 3 μg/mL leupeptin,1 mM PMSF) and lysed by high pressure with an Emulsiflex-C3 homogenizer (Avestin). Lysates were cleared by spinning at 30,000 x g for 30 minutes, and cleared lysates were incubated with TALON cobalt resin (Clontech) (1.5 mL slurry per 1 L culture) for 1 hour. Resin was batch-washed in 50 mL conicals for one minute before spinning down (1,000 x g) and removing the supernatant. For Rvb1/Rvb2 alone preps, resin was incubated with lysis buffer containing 100 μM ATP-Mg for 30 minutes, and then batch washed with lysis buffer for 30 minutes. For Rvb1/Rvb2/ID preps, the ATP wash step was skipped. Resin was transferred to a 20-mL disposable column (BioRad) and washed with 50 mLs of lysis buffer. Protein of interest was eluted in elution buffer (25 mM HEPES pH 8, 300 mM KCl, 500 mM imidazole). For Rvb1/Rvb2 untagged preps, TEV protease (homemade) was added to a final concentration of 0.075-0.15 mg/mL, and the mixture was dialyzed overnight into size exclusion (SEC) buffer, which contains 25 mM Hepes pH 7.5, 10% glycerol v/v, 300 mM KCl, and 2 mM dithiothreitol (DTT). For Rvb1/his-Rvb2 or Rvb1/Rvb2/ID preps, TEV protease was not added and the dialysis step was skipped. Proteins were then purified by size exclusion chromatography on a HiLoad Superdex 200 26/60 column (320 mL bed volume; GE Lifesciences) in SEC buffer. Fractions were analyzed on an SDS-PAGE gel, pooled, concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Native-MS

Native mass spectrometry was carried out using the Exactive Plus EMR instrument (Rose et al., 2012) (Thermo Scientific, San Jose, CA) that was externally calibrated using a 5 mg/mL CsI solution prepared in water. Prior to analysis, the protein samples were buffer-exchanged into 150 mM ammonium acetate, pH 7.5 using MicroBiospin-6 columns (Bio-Rad, Hercules, CA) that had been pre-equilibrated in the same buffer. Protein samples were introduced into the mass spectrometer using offline Au/Pd-coated borosilicate emitters (NanoES Spray Capillaries, Medium, ES380, Thermo Scientific) at a flow rate of 10 – 40 nL/min. Spectra were acquired over the range *m/z* 500 – 20,000 in positive ion mode, were averaged, and then exported for deconvolution and subsequent generation of the zero-charge mass values using PeakSeeker (Lu et al., 2015). Samples were analyzed with the following experimental parameters: spray voltage (0.8 – 1.5 kV), injection flatapole = 5; inter flatapole lens = 5; bent flatapole = 5; transfer multipole = 6.1; C-trap entrance lens = 8.9, source DC offset (25 V), fragmentation energies (CE = 20 – 50 and CID = 50 – 100), injection times (200 usec), trapping gas pressure (7.5), resolution (17,500 arbitrary units), capillary temperature (250 °C), S-len RF levels (200 V), microscans (10), and AGC ($1e^6$). Nucleotide – Mg complexes were prepared fresh by incubating equimolar ratio of nucleotide and $MgCl_2$ for 30 minutes at 4 °C and then freshly diluting to the final working concentration. Protein (~ 0.5 to 1 μM) was incubated with excess nucleotide (20 μM) immediately prior to analysis.

XL-MS

Cross-linking analysis was generally performed as in "Procedure B" of Robinson et al., 2015 (Robinson et al., 2015), with minor modifications. 80 μM of Rvb1/Rvb2, 46 μM of Rvb1/Rvb2/MBP-ino80 insertion, and 44 μM of Rvb1/Rvb2/Swr1IN insertion were buffer exchanged into 25 mM Hepes, pH 7.5, 150 mM KCl and crosslinked with 2 mM Bis[sulfosuccinimidyl]suberate (BS3,Thermoscientific Pierce) on ice for 30 minutes. Samples were quenched with 50 mM Tris, pH 8.5. Crosslinked proteins were pelleted by centrifuging at 21,000 x g. The pellet was washed once with cold acetone, and then resuspended in 50 mM ammonium bicarbonate containing 8M Urea and 10 mM

Tris(2-carboxyethyl)phosphine (TCEP). After incubation at 56 degrees for 20 minutes, the sample was cooled, iodoacetamide was added to a final concentration of 20 mM to alkylate –SH groups. Sample was then trypsinized overnight at 37°C. Following digestion, peptides were desalted using a Macrotrap C_{18} cartridge (Michrom Bioresources). Cross-linked products were then enriched by size-exclusion chromatography (Superdex Peptide, GE Healthcare Life Sciences). Fractions eluting between 0.9 and 1.4 ml were dried, resuspended in 0.1% formic acid and then analyzed with a Q-Exactive Plus mass spectrometer (Thermo Scientific) coupled with a nanoelectrospray ion source (Easy-Spray, Thermo) and NanoAcquity UPLC system (Waters). Enriched fractions were separated on a 15 cm x 75 µm ID PepMap C_{18} column (Thermo) using a 90-minute gradient from 3-27% solvent B (A: 0.1% formic acid in water, B: 0.1% formic acid in acetonitrile). Precursor MS scans were measured in the Orbitrap scanning from 350-1500 m/z (mass resolution: 70,000). The ten most intense triply charged or higher precursors were isolated in the quadrupole (isolation window: 4 m/z), dissociated by HCD (normalized collision energy: 25), and the product ion spectra were measured in the Orbitrap (mass resolution: 17,500). A dynamic exclusion window of 15 sec was applied and the automatic gain control targets were set to 3e6 (precursor scan) and 5e4 (product scan).

XL-MS data analysis

Peaklists were generated using Proteome Discoverer 1.4 (Thermo) and searched for cross-linked peptides with an in-house version of Protein Prospector 5.14.0 (Trnka et al., 2014). 85 of the most intense peaks from each product ion spectrum were searched against a database containing 11 target protein sequences (yeast Rvb1, Rvb2, Ino80 and Swr1 insertion constructs as well as several minor contaminating proteins identified in a prior, unbiased search of the data) concatenated with 110 randomized decoy protein sequences using mass tolerances of 9 ppm (precursor scan) and 20 ppm (product scan). Carbamidomethylation of cysteine was considered as a fixed modification. N-terminal methionine loss with and without acetylation, peptide N-terminal glutamine conversion to pyroglutamate, oxidation of methionine, and "dead-end" modification of lysine and the protein N-terminus by Tris quenched BS3 were considered as variable modifications in addition to cross-linking by BS3. Up to 3 variable modifications per peptide were considered. Trypsin specificity with 3 missed cleavages was used to generate theoretical peptides.

Cross-link spectral matches (CSM) were initially discarded if the following Protein Prospector parameters fell outside the threshold values: peptide score below 20, peptide/protein, and score difference below 0. A linear support vector machine (SVM) model was constructed based on three parameters ("score", "score difference", and "peptide 2 rank") to further classify CSMs between decoy and target classes (Robinson et al., 2015; Trnka et al., 2014). 5,208 spectra were classified with an SVM confidence score above the acceptance threshold of 0 at a false discovery rate (FDR) of 0.6%. The FDR at the residue pair level was 1.7%. CSMs where either peptide was less than 4-amino acids long were discarded. When spectra could be interpreted by multiple cross-linked peptide-pairs, CSMs were considered to be unambiguous if the second best match had a confidence score 0.3 lower than the top match. The remaining spectral redundancy came exclusively from ambiguous site localizations, where there was sufficient evidence to identify both peptides, but the position of the modified amino acid remained ambiguous. In these cases, the CSM was flagged as ambiguous and all possible sites were reported.

Homology models of yeast Rvb1Rvb2 dodecamers were constructed using modeller 9.12 using the crystal structure of the Rvb1 Rvb2 complex from *C. thermophilum* (pdb:4WVY) as the template. The sequences are over 70% identical.

Cross-linking results and annotated HCD spectra may be viewed online using Protein Prospector's MS-Viewer program:
http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer
Search key: ***uvtzc60yzg***

Integrative structure modeling of the Rvb1/Rvb2/Ino80INS complex

We computed a structural model of the yeast Rvb1/Rvb2/Ino80INS complex using an integrative approach based on data from X-ray crystallography, cross-linking mass spectrometry, and cryo-EM. The complex consists of 2 copies of the Ino80INS-MBP construct and 2 Rvb1/Rvb2 rings, each ring consisting of 3 copies of Rvb1 and Rvb2. Our integrative structure determination proceeds through four stages (Figure S6A) (Alber et al., 2007a; Russel et al.,

2012): (1) gathering of data, (2) representation of subunits and translation of the data into spatial restraints, (3) configurational sampling and scoring to produce an ensemble of structures that satisfies the restraints, and (4) analysis and validation of the ensemble structures. The modeling protocol (i.e., stages 2, 3, and 4) was scripted using the Python Modeling Interface (PMI), a library for modeling macromolecular complexes based on our open-source Integrative Modeling Platform (IMP) package, version 2.5 (http://integrativemodeling.org) (Russel et al., 2012). Further details of the integrative modeling procedures are provided in previous publications (S. J. Kim et al., 2014; LoPiccolo et al., 2015; Luo et al., 2015; Robinson et al., 2015; Webb et al., 2011). Files containing the input data, scripts, and output structures are available online (http://salilab.org/Rvbs).

1. Gathering of data: A comparative model of the yeast Rvb1/Rvb2 double-ring hexamer was computed based on the atomic X-ray structure of the double-heterohexameric Rvb1/Rvb2 rings from *Chaetomium thermophilum* (PDB access code 4WVY) (Lakomek et al., 2015), using MODELLER 9.15 (Sali and Blundell, 1993). The atomic X-ray structure of the MBP domain was extracted from the PDB (4FED (M. Kim, 2013)). The structure of the yeast Ino80INS is not known and no putative homologs of known structure were detected, using HHpred (Söding et al., 2005) and ModWeb (Pieper et al., 2014). Secondary structure and disordered regions of the Ino80INS domain were predicted by PSIPRED (McGuffin et al., 2000), JPRED (Drozdetskiy et al., 2015), and DISOPRED (Ward et al., 2004).
We used the 350 unique BS3 cross-links of the Rvb1/Rvb2/Ino80INS complex (Figures 1 and S1, and Table S1) and the consensus cryo-EM model of the Rvb1/Rvb2/Ino80INS determined in this study (Figure 5).

2. Representation of subunits and translation of the data into spatial restraints: To maximize computational efficiency while avoiding using too coarse a representation, we represented the Rvb1/Rvb2/Ino80INS complex in a multi-scale fashion. In particular, the domains of the Rvb1/Rvb2/Ino80INS complex subunits were coarse-grained using beads of varying sizes representing either a rigid body or a flexible string, based on the available crystallographic structures and comparative models, as follows. First, the comparative models and crystallographic structures of the Rvb1s, Rvb2s and MBPs components were coarse-grained into three representations at different resolutions. Two of the representations corresponded to beads obtained at different levels of coarse-graining of the atomic structure. At the first scale (i.e., fine scale), each bead corresponded to an individual residue, and was centered at the position of the Cα atom. At the second scale (i.e., coarse scale), each bead represented 10-residue segments and was positioned at the center of mass of all atoms of the corresponding segment. At the third scale, the system was represented by a 3D density map corresponding to the Gaussian mixture model (GMM) (Kawabata, 2008). The atomic structures were converted into a GMM by first sampling points from the density, and then fitting the sample using the expectation-maximization algorithm implemented in scikit-learn (Pedregosa et al., 2011). For each component, we set the number of Gaussians to be approximately the number of residues in the component divided by 50. We excluded loop regions Rvb1$^{1\text{-}51}$, Rvb1$^{143\text{-}165}$, Rvb1$^{211\text{-}238}$, Rvb2$^{444\text{-}454}$, Rvb2$^{142\text{-}155}$, Rvb2$^{205\text{-}229}$, and Rvb2$^{249\text{-}266}$, which were represented as flexible strings of beads, as described below. Second, the predicted helical regions of the Ino80INS components (Ino80$^{1137\text{-}1144}$, Ino80$^{1148\text{-}1156}$, Ino80$^{1193\text{-}1201}$, Ino80$^{1227\text{-}1237}$, and Ino80$^{1251\text{-}1261}$) were modeled as ideal helices and represented as described for crystallographic structures and comparative models; for each helix, we used two Gaussians for the GMM representation. Finally, the remaining regions without a crystallographic structure, comparative model, or predicted secondary structure were represented by a flexible string of beads corresponding to up to 10 residues each. In addition, each of these beads was represented by a spherical Gaussian (the centers of a bead and its Gaussian were identical). The radius of the bead and the variance of the Gaussians were set to describe the average molecular volume and the molecular electron density of the polypeptide segments, respectively.

Rigid body definition: We defined the Rvb1 domains as follows: Rvb1$^{1\text{-}128}$, Rvb1$^{129\text{-}305}$, and Rvb1$^{306\text{-}463}$ correspond to domains DI, DII, and DIII, respectively. Similarly, we defined the Rvb2 domains as follow: Rvb2$^{1\text{-}124}$, Rvb2$^{125\text{-}290}$, and Rvb2$^{291\text{-}471}$ correspond to domains DI, DII, and DIII, respectively. The DI/DIII domain of Rvb1/2 copies 1-3 were defined as a single rigid body (top ring; Rvb1.1$^{1\text{-}128}$-Rvb1.1$^{306\text{-}463}$- Rvb2.1$^{1\text{-}124}$-Rvb2.1$^{125\text{-}290}$- Rvb1.2$^{1\text{-}128}$-Rvb1.2$^{306\text{-}463}$- Rvb2.2$^{1\text{-}124}$-Rvb2.2$^{125\text{-}290}$ -Rvb1.3$^{1\text{-}128}$-Rvb1.3$^{306\text{-}463}$- Rvb2.3$^{1\text{-}124}$-Rvb2.3$^{125\text{-}290}$). Similarly, The DI/DIII domain of Rvb1/2 copies 4 through 6 were defined as a single rigid body (bottom ring; Rvb1.4$^{1\text{-}128}$-Rvb1.4$^{306\text{-}463}$-Rvb2.4$^{1\text{-}124}$-Rvb2.4$^{125\text{-}290}$- Rvb1.5$^{1\text{-}128}$-Rvb1.5$^{306\text{-}463}$- Rvb2.5$^{1\text{-}124}$-Rvb2.5$^{125\text{-}290}$ -Rvb1.6$^{1\text{-}128}$-Rvb1.6$^{306\text{-}463}$- Rvb2.6$^{1\text{-}124}$-Rvb2.6$^{125\text{-}290}$). Each of the 12 Rvb1/2 DII domains was defined as a single independent rigid body. Consequently, the Rvb1/Rvb2 complex was described by 14 rigid bodies. Each copy of MBP as well of each of the ideal helices of Ino80INS domains were represented as a single rigid body. In each rigid body, the beads have their relative distances constrained during configurational sampling, whereas in a flexible string the beads are restrained by the

sequence connectivity, as described in our previous studies (Algret et al., 2014; LoPiccolo et al., 2015; Shi et al., 2014).

Scoring function: With this representation in hand, we next encoded the spatial restraints into a Bayesian scoring function (Rieping et al., 2005; Shi et al., 2014) based on the information gathered in Stage 1, as follows.
First, the excluded volume restraints were applied to each bead in 10-residue (or the closest) bead representations, using the statistical relationship between the volume and the number of residues that it covered (Alber et al., 2007a; LoPiccolo et al., 2015; Shen and Sali, 2006).
Second, we applied the sequence connectivity restraint, using a harmonic upper bound on the distance between consecutive beads in a subunit, with a threshold distance equal to four times the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density (Alber et al., 2007a; LoPiccolo et al., 2015; Shi et al., 2014).
Third, the collected BS3 cross-links were used to construct the Bayesian scoring function that restrained the distances spanned by the cross-linked residues (35 A) (Erzberger et al., 2014; LoPiccolo et al., 2015; Shi et al., 2015), taking into account the ambiguity due to multiple copies of identical subunits (Algret et al., 2014; LoPiccolo et al., 2015); the ambiguous cross-link restraint considers all possible pairwise distances in multiple copies of identical subunits, scoring only the shortest distance. The cross-link restraint was applied to the fine scale representation for the X-ray structures and comparative models as well as to flexible beads.
Fourth, the EM 3D restraint corresponded to the cross-correlation coefficient between the GMM representation of each component and the GMM representation of the consensus EM map (Kawabata, 2008; Robinson et al., 2015). The EM map was approximated by a 680 Gaussian GMM computed using the expectation-maximization as implemented in scikit-learn (Pedregosa et al., 2011); 680 Gaussians appeared to be sufficient to reproduce the significant features of the map.

3. Configurational sampling and scoring: Structural models of the Rvb1/Rvb2/Ino80INS complex were computed using Replica Exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm (Rieping et al., 2005; Shi et al., 2015). The Monte Carlo moves included random translation and rotation of rigid bodies (up to 1.5 Å and 0.03 radians, respectively) and random translation of individual beads in the flexible segments (up to 2 Å). 32 replicas were used for each run, with temperatures ranging between 1.0 and 2.5. A structure model was saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every rigid body and flexible bead once. The sampling produced a total of ~500,000 conformations from the 20 independent runs. 500 top-scoring structures from Step 3 were subjected to the subsequent analysis in Stage 4.

4. Analysis and validation of the ensemble:

Input information and output structures were analyzed to estimate structure precision and accuracy, detect inconsistencies among input data and missing information, and to suggest more informative future experiments. Assessment begins with structural clustering of the modeled structures produced by sampling, followed by assessment of the thoroughness of structural sampling, estimating structure precision based on variability in the ensemble of good-scoring structures and quantification of the structure fit to the input information. These validations are based on the nascent wwPDB effort on archival, validation, and dissemination of integrative structure models (Sali et al., 2015). We now discuss each one of these points in turn.

Clustering: A prerequisite for analysis of the ensemble of structures generated by satisfying the input data is clustering of these structures (Alber et al., 2007a; Shi et al., 2014; Zeng-Elmore et al., 2014). We used Cα root-mean-square deviation (RMSD) quality-threshold clustering (Shi et al., 2014). For the Rvb1/Rvb2/Ino80INS complex, the 500 best scoring models grouped into 3 clusters. The average RMSD between the major (196 structures) and minor clusters (180 and 124 structures) was 86.4 Å and 77.3 Å respectively. Despite the large RMSD between clusters, the localization of all components is effectively identical between the clusters, differing mainly in the orientation of the Ino80INS-MBP domain (Figure S7C). Most importantly, our functional interpretation of the structure is the same regardless of which cluster we analyze.

Sampling convergence: For stochastic methods, thoroughness of sampling can be assessed by showing that two independent runs or set of runs (e.g., using random starting configurations or different random number generator seeds) do not result in significantly different solutions (Alber et al., 2007a; Algret et al., 2014; Shi et al., 2014). Given two or more sets of structures from independent runs, we first cluster structures from all sets together,

followed by assessing whether or not the runs contribute evenly to the population of each cluster, using the p-value from the χ-square contingency test for homogeneity of proportions (McDonaldUniversity of Delaware, 2009). For the Rvb1/Rvb2/Ino80INS complex, the significant p-value of 0.15 indicated that our Monte Carlo algorithm sampled all top-scoring solutions at the resolution better than the precision of the dominant cluster. The caveat is that passing this sampling test is not absolute evidence of thorough sampling; a positive outcome of the test may be misleading if, for example, the landscape contains only a narrow, and thus difficult to find, pathway to the pronounced minimum corresponding to the correct structure.

Estimating structure precision based in variability in the ensemble of good-scoring structures: The ensemble of the top-scoring structures was analyzed in terms of the precision of its structural features (Alber et al., 2007a; 2007b). The spread around the maximum describes how precisely the feature is determined from the input information. The precision of component position is quantified as the average root-mean-square fluctuation (RMSF) across all pairs of in the cluster, after least-squares superposition onto the centroid structure structures (Shi et al., 2014) (Figure S5C), likely provided the lower bound on accuracy.For Cluster 1, the precision of the complex is 25 Å. The precision of the Rvbs dodecamer and the Ino80INS-MBP domain is 15.9 and 39.9 Å, respectively. This precision is sufficiently high to pinpoint the locations of the Ino80INS and DII domains (Figures 1A and S5C, and Table 1), demonstrating the quality of the data including the cross-links and 3D-EM map. For Clusters 2 and 3, the precision of the complex is 25 Å (14.8 Å for the Rvbs and 40.2 Å for the Ino80INS-MBP domains) and 12.2 Å (6.2 Å for the Rvbs and 14.3 Å for the Ino80INS-MBP domains), respectively. The localization probability density maps of every Rvb1/Rvb2/Ino80INS subunit as well as the whole complex were computed from the dominant cluster of the 189 solutions (Figures 6B) and the other two clusters (Figure S7C).

Fit to input data: The ensemble of solutions was assessed in terms of how well they satisfied information from which they were computed, including the cross-links, the 3D-EM, the excluded volume and sequence connectivity restraints.

First, 97%, 95%, and 91% of all crosslinks are satisfied by at least one structure in Clusters 1, 2, and 3, respectively. 98% of all crosslinks are satisfied by at least one structure in the three clusters; a cross-link restraint was satisfied by a cluster if the Cα-Cα distance between the cross-linked residues (considering restraint ambiguity) was less than 30 Å in any of the structures in the cluster.

Second, the solutions fit the 3D-EM consensus map with an average cross-correlation of 0.94.

Third, 99% of the top 500 solutions satisfied the excluded volume and sequence connectivity restraints under a combined score threshold of 50.

Supplemental References

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Sali, A., Rout, M.P., 2007b. The molecular architecture of the nuclear pore complex. Nature 450, 695–701. doi:10.1038/nature06405

Algret, R., Fernandez-Martinez, J., Shi, Y., Kim, S.J., Pellarin, R., Cimermančič, P., Cochet, E., Sali, A., Chait, B.T., Rout, M.P., Dokudovskaya, S., 2014. Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. Mol. Cell Proteomics 13, 2855–2870. doi:10.1074/mcp.M114.039388

Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J., 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43, W389–94. doi:10.1093/nar/gkv332

Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimermančič, P., Boehringer, D., Sali, A., Aebersold, R., Ban, N., 2014. Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. Cell 158, 1123–1135. doi:10.1016/j.cell.2014.07.044

Kawabata, T., 2008. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. Biophys. J. 95, 4643–4658. doi:10.1529/biophysj.108.137125

Kim, S.J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T.M., Shi, Y., Markina-Inarrairaegui, A., Bonanno, J.B., Sauder, J.M., Burley, S.K., Chait, B.T., Almo, S.C., Rout, M.P., Sali, A., 2014. Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. Mol. Cell Proteomics 13, 2911–2926. doi:10.1074/mcp.M114.040915

LoPiccolo, J., Kim, S.J., Shi, Y., Wu, B., Wu, H., Chait, B.T., Singer, R.H., Sali, A., Brenowitz, M., Bresnick, A.R., Backer, J.M., 2015. Assembly and Molecular Architecture of the Phosphoinositide 3-Kinase p85α Homodimer. J. Biol. Chem. 290, 30390–30405. doi:10.1074/jbc.M115.689604

Luo, J., Cimermančič, P., Viswanath, S., Ebmeier, C.C., Kim, B., Dehecq, M., Raman, V., Greenberg, C.H., Pellarin, R., Sali, A., Taatjes, D.J., Hahn, S., Ranish, J., 2015. Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. Mol. Cell 59, 794–806. doi:10.1016/j.molcel.2015.07.016

McDonald, J.H., University of Delaware, 2009. Handbook of Biological Statistics.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405. doi:10.1093/bioinformatics/16.4.404

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pieper, U., Webb, B.M., Dong, G.Q., Schneidman-Duhovny, D., Fan, H., Kim, S.J., Khuri, N., Spill, Y.G., Weinkam, P., Hammel, M., Tainer, J.A., Nilges, M., Sali, A., 2014. ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. 42, D336–46. doi:10.1093/nar/gkt1144

Rieping, W., Habeck, M., Nilges, M., 2005. Inferential structure determination. Sci Signal 309, 303–306. doi:10.1126/science.1110428

Sali, A., Berman, H.M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S.K., Markley, J., Nakamura, H., Adams, P., Bonvin, A.M.J.J., Chiu, W., Peraro, M.D., Di Maio, F., Ferrin, T.E., Grünewald, K., Gutmanas, A., Henderson, R., Hummer, G., Iwasaki, K., Johnson, G., Lawson, C.L., Meiler, J., Marti-Renom, M.A., Montelione, G.T., Nilges, M., Nussinov, R., Patwardhan, A., Rappsilber, J., Read, R.J., Saibil, H., Schröder, G.F., Schwieters, C.D., Seidel, C.A.M., Svergun, D., Topf, M., Ulrich, E.L., Velankar, S., Westbrook, J.D., 2015. Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. Structure 23, 1156–1167. doi:10.1016/j.str.2015.05.013

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815. doi:10.1006/jmbi.1993.1626

Shen, M.-Y., Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. Protein Sci. 15, 2507–2524. doi:10.1110/ps.062416606

Shi, Y., Pellarin, R., Fridy, P.C., Fernandez-Martinez, J., Thompson, M.K., Li, Y., Wang, Q.J., Sali, A., Rout, M.P., Chait, B.T., 2015. A strategy for dissecting the architectures of native macromolecular assemblies. Nat. Methods 12, 1135–1138. doi:10.1038/nmeth.3617

Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–8. doi:10.1093/nar/gki408

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., Jones, D.T., 2004. The DISOPRED server for the prediction of protein disorder. Bioinformatics 20, 2138–2139. doi:10.1093/bioinformatics/bth195

Webb, B., Lasker, K., Schneidman-Duhovny, D., Tjioe, E., Phillips, J., Kim, S.J., Velázquez-Muriel, J., Russel, D., Sali, A., 2011. Modeling of proteins and their assemblies with the integrative modeling platform. Methods Mol. Biol. 781, 377–397. doi:10.1007/978-1-61779-276-2_19

Zeng-Elmore, X., Gao, X.-Z., Pellarin, R., Schneidman-Duhovny, D., Zhang, X.-J., Kozacka, K.A., Tang, Y., Sali, A., Chalkley, R.J., Cote, R.H., Chu, F., 2014. Molecular architecture of photoreceptor phosphodiesterase elucidated by chemical cross-linking and integrative modeling. J. Mol. Biol. 426, 3713–3728. doi:10.1016/j.jmb.2014.07.033