SUPPLEMENTAL INFORMATION

Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex

Yi Shi[1]*, Javier Fernandez-Martinez[2]*, Elina Tjioe[3]*, Riccardo Pellarin[3]*, Seung Joong Kim[3]*, Rosemary Williams[2], Dina Schneidman-Duhovny[3], Andrej Sali[3], Michael P. Rout[2], and Brian T. Chait[1]

[1]*Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA*

[2]*Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA*

[3]*Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall, 1700 4th Street, Suite 503B, University of California, San Francisco, San Francisco, CA 94158, USA*

**These authors contributed equally to this work as co-first authors.*

Corresponding authors

Brian T. Chait, Box 170, Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA.

tel: +1 212 327 8849; fax: +1 212 327 7547; chait@rockefeller.edu

Michael. P. Rout, Box 213, Laboratory of Cellular and Structural Biology, Rockefeller

University, 1230 York Avenue, New York, NY 10021, USA.

tel: +1 212 327 8135; fax: +1 212 327 7193; rout@rockefeller.edu

Andrej Sali, UCSF MC 2552, Byers Hall at Mission Bay, Suite 503B, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA

tel: +1 415 514 4227; fax: +1 415 514 4231; sali@salilab.org

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Purification of the endogenous Nup84 complex

To purify the native *S. cerevisiae* Nup84 complex, we used a strain in which the Nup84-encoding gene was genomically tagged with PrA preceded by the human rhinovirus 3C protease (ppx) target sequence (GLEVLFQGPS). Harvested cells, grown in YPD at 30°C, were frozen in liquid nitrogen and cryogenically lysed in a Retsch PM 100 planetary ball mill (http://lab.rockefeller.edu/rout/protocols). A total of 10-20 g of frozen cell powder were resuspended in 9 volumes of resuspension buffer (20 mM Hepes/KOH pH 7.4, 500 mM NaCl, 2 mM $MgCl_2$, 0.1% (w/v) CHAPS, 1 mM DTT, 1/500 (v/v) Protease Inhibitor Cocktail (Sigma)). Cell lysate was clarified by centrifugation at 20,000 g for 10 minutes. IgG Ab conjugated magnetic beads (Invitrogen) at a concentration of 50 µL slurry/g of frozen powder were added to the clarified cell lysate and incubated for 30 minutes at 4°C. Beads were washed three times with 1 ml of resuspension buffer without protease inhibitors. Beads were resuspended in 10 µL of buffer and 3 µg of ppx protease were added. Beads were incubated at 4°C for 45 minutes with constant agitation. A magnet was used to separate the beads from the supernatant, 100 µL of digestion buffer were used to wash the remaining beads, and the total volume of sample was then centrifuged at 20,000 g for 10 minutes. The sample was then processed for cross-linking as described in the Experimental Procedures section.

Bayesian scoring function

The Bayesian approach (1) estimates the probability of a model, given information available about the system, including both prior knowledge and newly acquired experimental data. The model $M \equiv (X, \{\alpha_i\})$ includes the structure coordinates X and additional parameters $\{\alpha_i\}$. Using Bayes' theorem, the posterior probability $p(M \mid D, I)$, given data D and prior knowledge I, is $p(M \mid D, I) \propto p(D \mid M, I) p(M, I)$, where the likelihood function $p(D \mid M, I)$ is the probability of observing data D, given I and M, and the prior is the probability of model M, given I. To define the likelihood function, one needs a forward model that predicts the data point (i.e., the presence of a cross-link between two given residues) given any model M, and a noise model that specifies the distribution of the deviation between the observed and predicted data points. The Bayesian scoring function is the negative logarithm of $p(D \mid M, I) p(M \mid I)$, which ranks the models identically to the posterior probability.

Briefly, the forward model f_n is computed as the probability of randomly picking two points $\tilde{\mathrm{r}}_i$ and $\tilde{\mathrm{r}}_j$ within the spheres centered on the C_α atoms of the cross-linked residues, with coordinates r_i and r_j, with unknown radii σ_i and σ_j, such that the distance between them $\tilde{r}_{ij}$ is lower than the maximum cross-linker length l_{XL}; the radii σ_i and σ_j are proxies for the uncertainty of forming a cross-link, given structural model X. To reduce the number of parameters in the model, we utilized a single uncertainty parameter σ for all residues. We imposed l_{XL} = 21 Å and 16 Å for the DSS and EDC cross-linkers, respectively.

The likelihood function for a cross-link d_n is $p(d_n \mid X, I) = \psi \cdot (1 - f_n(X)) + f_n(X) \cdot (1 - \psi)$, where ψ is the uncertainty of observing a cross-link that is approximately equal to the fraction of false-positive cross-links. We set ψ to 5%. The joint likelihood function $p(D \mid M, I)$ for a dataset $D = \{d_n\}$ of N_{XL} independently observed cross-links is the product of likelihood functions for each data point. When multiple residue pairs $(n, n+1, ...)$ are assigned to an individual cross-link observation, the ambiguity is taken account for by the following compound likelihood $p(d_{n,n+1,...} \mid X, I) = 1 - \prod_n (1 - p(d_n \mid X, I))$.

The model prior $p(M \mid I)$ is defined as a product of the priors $p(X)$ and $p(\sigma)$ on the structural coordinates X and uncertainty σ, respectively. The prior $p(X)$ includes the excluded volume restraints, the sequence connectivity restraints, the EM2D restraint, and a weak restraint whose score depends linearly on the distance between cross-linked residues, with a slope of 0.01 Å^{-1}. $p(\sigma)$ is a uniform distribution over the interval [0,100].

A detailed description of the representation, scoring function, and sampling is described elsewhere (2, 3).

EM 2D Restraint and Scoring Function

The EM 2D restraint fits a given model to an EM class average (4) and computes a score that quantifies the match (5). The computation proceeds in three stages: (i) generation of alternative model projections, (ii) alignment of the class average and each model projection, and (iii) calculation of the fitting score for each projection.

Projection generation. The highest resolution bead representation of the evaluated model is used (Figure S3). The electron density of each bead is represented by a 3D Gaussian function projected onto a 2D grid. The pixel size of the resulting projection image is equal to the pixel size of the class average (5.9Å). The intensity of each grid point is:

$$\rho_{i,j} = \frac{m}{\sqrt{2\pi}\sigma} \sum_{k} \exp[-\frac{(\Delta r)^2}{2\sigma}]$$

where *m* is the mass of the bead, $\Delta r$is the distance between the bead center and the *(i, j, k)* point in 3D space, and σ is 0.425 times the resolution of the EM class average (30 Å here) (6). For sufficient coverage, 400 projections are generated by uniform sampling of the unit sphere (7). In the subsequent evaluations of the scoring function, to save time only projections that are within 20 degrees on the unit sphere from the previously selected best projection are considered; this periodic update is possible because the model changes slowly during sampling. All projections are calculated from scratch every 100 model sampling iterations.

Alignment of the class average and model projection. The projections and the class

average are segmented to identify the pixels originating from the modeled system, using thresholding and connected component labeling (8). Next, the principal axes of the segmented areas are computed. Each projection is aligned efficiently to the class average using only the principal axes of the segmented areas.

Calculation of the fitting score. For each pair of the aligned class average and a projection, the score is calculated as the cross-correlation coefficient between the two images. The final restraint score is the negative logarithm of the highest cross-correlation coefficient. The relative weight of the EM 2D restraint in the total score of a model was set to 500 based on an enumeration of the weight, requiring reasonable satisfaction of all types of restraints.

SUPPLEMENTAL TABLES

Supplemental Table S1. DSS cross-linking MS data (attached separately)

Note that only the identification with the best score is shown.

Supplemental Table S2. EDC cross-linking MS data (attached separately)

Note that only the identification with the best score is shown. We were unable to unambiguously locate the conjugation sites for ~20% of the EDC cross-links, because many result from peptides containing consecutive and/or adjacent carboxylic acids. These ambiguities were accounted for in our modeling calculations, resulting in 123 possible EDC cross-links used for further analysis.

A	59 DSS cross-links					47 EDC cross-links				
	Protein 1	Residue 1	Protein 2	Residue 2	Median Distance [Å]	Protein 1	Residue 1	Protein 2	Residue 2	Median Distance [Å]
Inter-molecular cross-links	Sec13	245	Nup85	2	77.3	Nup133	253	Nup84	340	201.6
	Sec13	183	Seh1	198	71.8	Sec13	2	Nup120	990	75.3
	Sec13	252	Nup85	734	69.9	Nup145c	141	Seh1	284	72.7
	Nup145c	39	Nup85	163	64.4	Nup145c	39	Nup85	164	64.2
	Nup145c	43	Nup85	163	63.7	Nup85	705	Seh1	144	59.2
	Nup85	60	Nup145c	43	62.8	Sec13	2	Seh1	279	55.2
	Nup85	718	Seh1	198	59.7	Nup145c	62	Nup85	2	52.9
	Nup133	966	Nup84	625	56.5	Nup145c	63	Nup85	2	52.9
	Nup145c	1	Nup85	163	55.0	Sec13	2	Nup145c	283	52.7
	Nup145c	39	Nup85	2	53.2	Seh1	40	Nup145c	44	41.3
	Seh1	207	Nup85	2	49.4	Nup85	536	Nup145c	70	40.9
	Seh1	212	Nup85	2	45.9	Nup145c	39	Nup85	715	40.8
	Sec13	2	Seh1	198	45.8	Seh1	79	Nup85	2	40.4
	Nup145c	59	Nup85	680	39.7	Nup145c	39	Nup85	717	36.5
	Nup85	680	Nup145c	43	39.7	Sec13	2	Seh1	141	32.0
	Sec13	83	Nup145c	649	39.5	Nup145c	43	Seh1	138	30.8
	Nup145c	39	Nup85	680	38.9	Nup145c	39	Seh1	138	29.6
	Seh1	236	Nup85	2	38.3	Nup145c	39	Nup85	587	24.8
	Seh1	100	Nup85	2	34.9	Nup145c	39	Seh1	141	24.8
	Seh1	40	Nup85	564	34.6	Nup145c	43	Seh1	141	24.7
	Sec13	2	Nup145c	141	34.4	Nup84	154	Nup145c	335	24.0
	Nup145c	59	Nup85	564	27.3	Nup84	154	Nup145c	338	20.1
	Nup145c	1	Nup85	564	25.4	Nup133	936	Nup84	713	20.0
	Seh1	1	Nup85	163	24.7	Sec13	282	Nup145c	438	19.4
	Sec13	244	Nup145c	437	23.8	Sec13	2	Nup145c	22	19.3
	Nup145c	183	Sec13	282	23.7	Nup85	585	Nup145c	42	18.9
	Nup145c	694	Nup85	729	23.5	Sec13	2	Nup145c	42	18.3
	Nup85	438	Seh1	244	22.5	Sec13	2	Nup145c	60	18.3
	Nup120	972	Nup85	729	22.5	Nup84	712	Nup133	935	18.0
	Nup120	972	Nup85	734	22.3	Seh1	40	Nup85	2	17.6
	Nup120	943	Nup145c	672	22.0	Nup84	143	Nup145c	335	17.1
	Seh1	34	Nup85	2	21.9	Nup145c	43	Nup85	587	17.0
	Nup120	972	Nup85	733	20.6	Sec13	2	Nup145c	72	16.8
	Nup120	972	Nup145c	694	20.6	Sec13	2	Nup145c	61	16.8
	Nup120	972	Nup145c	681	20.1	Sec13	282	Nup145c	179	16.6
	Nup120	943	Nup145c	681	19.6	Sec13	283	Nup145c	437	15.5
	Sec13	2	Nup145c	39	19.3	Sec13	2	Nup145c	164	14.4
	Sec13	31	Nup145c	141	19.0	Seh1	2	Nup85	34	14.1
	Seh1	236	Nup85	64	18.9	Seh1	1	Nup85	33	13.9
	Nup133	936	Nup84	712	18.8	Seh1	1	Nup85	32	13.9
	Sec13	2	Nup145c	59	18.3	Sec13	10	Nup145c	164	13.4
	Sec13	2	Nup145c	43	18.3	Nup85	17	Seh1	1	12.2
	Sec13	282	Nup145c	437	18.2	Seh1	1	Nup85	4	12.2
	Seh1	40	Nup85	2	17.6	Nup85	2	Seh1	1	12.2
	Sec13	288	Nup145c	437	16.3	Sec13	286	Nup145c	437	11.7
	Sec13	2	Nup145c	163	15.8	Seh1	2	Nup85	4	10.8
	Sec13	282	Nup145c	174	15.7	Seh1	2	Nup85	5	10.8
	Sec13	65	Nup145c	509	14.9					
	Seh1	1	Nup85	39	13.9					
	Seh1	1	Nup85	30	13.9					
	Seh1	1	Nup85	89	13.1					
	Sec13	44	Nup145c	168	13.0					
	Seh1	1	Nup85	2	12.2					
	Nup145c	353	Nup84	154	12.1					
	Sec13	2	Nup145c	168	12.0					
	Seh1	2	Nup85	1	10.8					
	Sec13	277	Nup145c	174	10.7					
	Sec13	34	Nup145c	168	10.3					
	Sec13	2	Nup145c	174	9.5					

Supplemental Table S3. (A) The median distances of the inter-molecular cross-links calculated using the 6,520 solutions.

Violated cross-links are highlighted in orange.

B

Intra-molecular cross-links

104 DSS cross-links					76 EDC cross-links				
Protein 1	Residue 1	Protein 2	Residue 2	Median Distance [Å]	Protein 1	Residue 1	Protein 2	Residue 2	Median Distance [Å]
Nup85	293	Nup85	2	62.2	Nup133	936	Nup133	392	118.8
Nup85	231	Nup85	2	54.5	Nup145c	593	Nup145c	201	48.5
Nup133	506	Nup133	59	38.6	Nup133	946	Nup133	1097	30.6
Nup133	1072	Nup133	936	30.6	Nup145c	250	Nup145c	213	25.5
Nup85	85	Nup85	2	26.6	Nup84	544	Nup84	565	23.8
Sec13	2	Sec13	31	24.8	Nup133	936	Nup133	952	23.2
Nup133	59	Nup133	5	24.2	Nup133	1072	Nup133	1032	22.7
Nup84	573	Nup84	544	23.8	Nup145c	39	Nup145c	61	22.1
Nup84	592	Nup84	560	23.4	Sec13	2	Sec13	47	20.0
Nup145c	183	Nup145c	174	23.2	Nup84	143	Nup84	164	19.2
Nup84	544	Nup84	670	23.0	Nup120	400	Nup120	441	18.9
Nup133	191	Nup133	239	22.8	Seh1	207	Seh1	279	18.8
Nup133	1034	Nup133	946	22.8	Seh1	207	Seh1	283	18.8
Nup133	913	Nup133	946	22.7	Nup85	30	Nup85	164	18.7
Nup133	544	Nup133	488	22.2	Nup133	913	Nup133	1097	18.4
Nup85	718	Nup85	734	22.2	Nup145c	219	Nup145c	187	18.2
Nup85	718	Nup85	733	22.1	Nup133	946	Nup133	935	17.8
Nup145c	593	Nup145c	694	21.8	Nup145c	91	Nup145c	70	17.3
Nup133	532	Nup133	488	21.7	Nup120	865	Nup120	894	17.2
Nup145c	1	Nup145c	43	21.6	Nup120	865	Nup120	898	17.2
Nup120	17	Nup120	735	21.3	Nup145c	642	Nup145c	686	16.7
Nup145c	668	Nup145c	593	21.2	Nup145c	91	Nup145c	60	16.5
Nup133	946	Nup133	912	20.8	Nup145c	91	Nup145c	44	16.5
Nup145c	183	Nup145c	197	20.5	Nup120	38	Nup120	429	16.4
Nup145c	1	Nup145c	39	20.2	Nup133	913	Nup133	989	16.4
Seh1	207	Seh1	198	20.2	Nup120	735	Nup120	403	16.2
Nup85	60	Nup85	39	20.1	Seh1	37	Seh1	89	16.2
Nup85	39	Nup85	60	20.1	Nup85	87	Nup85	59	16.2
Nup133	168	Nup133	191	19.8	Nup145c	39	Nup145c	60	16.2
Nup85	60	Nup85	2	19.8	Nup145c	672	Nup145c	700	16.2
Nup84	23	Nup84	14	19.8	Nup133	1142	Nup133	1068	16.0
Nup133	1072	Nup133	1035	19.5	Nup145c	681	Nup145c	700	16.0
Nup145c	683	Nup145c	672	19.2	Nup145c	681	Nup145c	704	16.0
Seh1	40	Seh1	34	19.1	Nup133	912	Nup133	1097	15.8
Nup145c	642	Nup145c	694	18.8	Nup120	400	Nup120	440	15.8
Nup145c	694	Nup145c	649	18.8	Nup133	187	Nup133	127	15.8
Nup120	400	Nup120	735	18.7	Seh1	201	Seh1	279	15.5
Nup133	1072	Nup133	1034	18.7	Seh1	201	Seh1	283	15.5
Nup85	30	Nup85	163	18.5	Nup120	206	Nup120	219	14.7
Nup145c	67	Nup145c	43	18.5	Nup133	486	Nup133	442	14.5
Nup84	602	Nup84	544	17.7	Nup133	484	Nup133	442	14.5
Nup120	735	Nup120	412	17.6	Nup133	103	Nup133	131	14.2
Sec13	2	Sec13	34	17.4	Nup133	103	Nup133	130	14.2
Nup84	23	Nup84	460	17.2	Nup133	394	Nup133	454	14.2
Nup85	680	Nup85	734	17.1	Nup133	1142	Nup133	1069	14.2
Nup85	89	Nup85	2	17.0	Nup133	168	Nup133	131	14.0
Nup85	247	Nup85	237	16.8	Nup120	735	Nup120	405	14.0
Nup145c	642	Nup145c	681	16.7	Nup145c	309	Nup145c	293	13.9
Nup145c	642	Nup145c	683	16.7	Sec13	277	Sec13	294	13.9
Nup145c	284	Nup145c	357	16.7	Nup133	506	Nup133	562	13.6
Nup85	39	Nup85	2	16.6	Nup133	171	Nup133	127	13.5
Nup85	30	Nup85	1	16.6	Seh1	198	Seh1	279	13.4
Nup85	30	Nup85	2	16.6	Nup84	544	Nup84	598	12.9
Seh1	248	Seh1	207	16.4	Sec13	2	Sec13	37	12.7
Nup145c	48	Nup145c	39	16.2	Nup145c	219	Nup145c	231	12.2
Nup145c	59	Nup145c	39	16.2	Nup133	1144	Nup133	1068	12.0
Nup145c	694	Nup145c	672	16.2	Nup145c	193	Nup145c	223	12.0
Nup84	20	Nup84	459	16.1	Nup120	400	Nup120	523	11.6
Nup145c	681	Nup145c	711	16.0	Nup145c	183	Nup145c	231	11.6
Nup145c	694	Nup145c	681	16.0	Sec13	186	Sec13	239	11.6
Nup145c	707	Nup145c	681	16.0	Nup84	143	Nup84	175	11.5
Sec13	277	Sec13	249	15.9	Sec13	22	Sec13	41	11.4
Nup84	669	Nup84	712	15.9	Nup85	85	Nup85	63	11.3
Nup133	171	Nup133	191	15.6	Nup145c	183	Nup145c	226	10.7
Sec13	252	Sec13	245	15.5	Sec13	288	Sec13	250	10.5
Nup120	400	Nup120	531	15.5	Nup133	480	Nup133	63	10.2
Nup133	278	Nup133	239	14.9	Seh1	37	Seh1	92	10.1
Nup84	573	Nup84	560	14.7	Nup145c	193	Nup145c	220	9.9
Nup120	206	Nup120	217	14.7	Nup84	619	Nup84	599	9.9
Nup133	488	Nup133	442	14.5	Nup84	592	Nup84	626	9.7
Nup133	115	Nup133	59	14.3	Nup120	17	Nup120	421	9.6
Nup145c	227	Nup145c	183	14.2	Nup84	143	Nup84	173	7.9
Nup120	972	Nup120	963	14.2	Nup133	171	Nup133	195	7.2
Nup145c	204	Nup145c	532	13.9	Sec13	22	Sec13	39	6.4
Nup84	66	Nup84	382	13.4	Sec13	2	Sec13	41	5.3
Nup85	102	Nup85	163	13.4	Nup133	171	Nup133	197	5.1
Seh1	1	Seh1	40	13.4					
Nup84	460	Nup84	20	12.9					
Sec13	44	Sec13	2	12.8					
Seh1	182	Seh1	198	12.5					
Sec13	244	Sec13	288	12.3					
Nup84	114	Nup84	107	12.2					
Nup120	17	Nup120	412	12.2					
Nup120	400	Nup120	530	11.8					
Sec13	288	Sec13	245	11.4					
Nup84	592	Nup84	625	11.3					
Nup85	85	Nup85	64	11.3					
Nup85	247	Nup85	345	11.0					
Nup85	536	Nup85	564	10.5					
Nup145c	360	Nup145c	353	10.5					
Nup84	669	Nup84	676	10.2					
Nup145c	163	Nup145c	174	10.1					
Nup120	384	Nup120	362	9.4					
Nup85	231	Nup85	238	8.8					
Nup85	729	Nup85	734	8.5					
Sec13	288	Sec13	249	8.3					
Seh1	236	Seh1	212	7.9					
Seh1	82	Seh1	100	6.9					
Nup145c	236	Nup145c	219	6.6					
Seh1	108	Seh1	79	6.2					
Nup85	729	Nup85	733	6.1					
Nup133	973	Nup133	976	5.2					
Nup133	966	Nup133	963	4.7					
Sec13	186	Sec13	172	4.4					

Supplemental Table S3. (B) The median distances of the intra-molecular cross-links calculated using the 6,520 solutions

Violated cross-links are highlighted in orange.

Number of constrained interfaces	Cluster Index	Number of structures	Precision (dRMSD) [Å]
3	1	1,257	15.4
	2	1,010	12.7
0	1	577	14.8
	2	404	10.8

Supplemental Table S4. Precisions calculated on the hub region

The precisions of each dominant cluster of solutions, obtained with and without the 3 crystallographic dimer constraints, were calculated considering only the five subunits in the hub region.

Number of constrained interfaces	Cluster Index	Number of structures	Precision (dRMSD) [Å]
3	1	5,064	17.3
	2	1,475	17.3
0	1	2,885	17.3
	2	1,401	19.2

Supplemental Table S5. Overall precisions of the solutions

The precisions of each dominant cluster of solutions, obtained with and without the 3 crystallographic dimer constraints, were calculated considering all subunits of the Nup84 complex.

Nup84 complex components	Estimated Volume (from sequence)	Volume Threshold (x 2.5)	Residue number begins at	Residue number ends at
Nup84	110,189	275,473	1	726
Nup85	102,719	256,798	1	744
Nup120-NTD	98,861	247,153	1	712
Nup120-CTD	45,771	114,428	713	1037
Nup120 (whole)	144,632	361,580	1	1037
Nup133	161,305	403,263	1	1157
Nup145c-NTD	16,617	41,543	1	125
Nup145c-middle	59,693	149,233	126	553
Nup145c-CTD	21,819	54,548	554	712
Nup145c (whole)	98,129	245,323	1	712
Seh1	47,336	118,340	1	349
Sec13	39,980	99,950	1	297
Entire Nup84 complex	704,290	1,760,725		

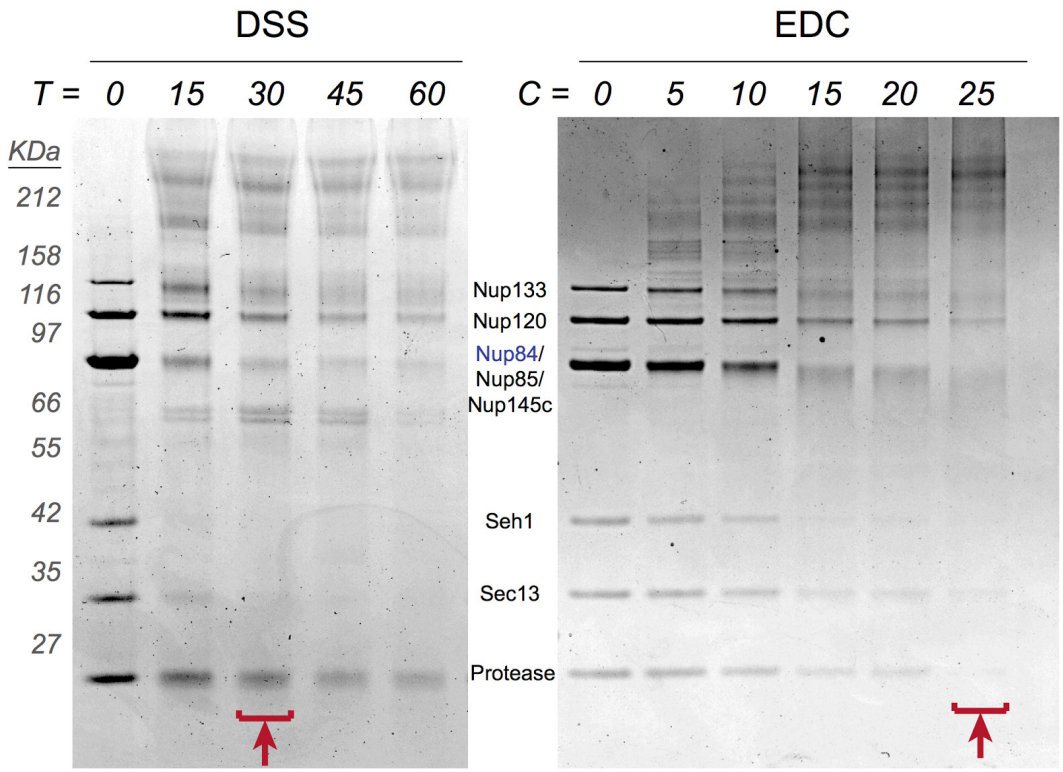
Supplemental Table S6. Volume thresholds used for the localization density maps of the Nup84 complex components

The volume of each Nup84 complex component was estimated from sequence in the units of $Å^3$ (http://www.basic.northwestern.edu/biotools/proteincalc.html). Each volume was then multiplied by 2.5 to obtain the volume threshold for contouring the localization density map.

Rank	log(e)	log(I)	Coverage %	unique peptide	total peptide	Mr	Accession	Description
1	-466.3	7.34	34	49	102	133.2	YKR082W	NUP133
2	-430.4	7.6	24	41	97	145.6	YGL092W	NUP145
3	-343.1	7.56	36	37	92	83.6	YDL116W	NUP84
4	-305.3	7.27	29	35	71	120.4	YKL057C	NUP120
5	-239.8	7.5	73	24	62	33	YLR208W	SEC13
6	-223.9	6.79	32	25	58	84.8	YJR042W	NUP85
7	-223.7	7.51	52	24	62	39.1	YGL100W	SEH1
8	-121.7	5.92	24	14	21	69.2	sp|ALBU_BOVIN|	Serum albumins
9	-41.8	6.2	20	6	10	35.8	YJR009C	TDH2
10	-38.4	5.55	6.6	4	7	69.4	YLL024C	SSA2
11	-34.2	6.34	20	3	14	25.4	sp|TRY1_BOVIN|	Cationic trypsin
12	-22.5	5.11	8.7	4	5	44.7	YCR012W	PGK1
13	-19	4.71	5.2	3	3	66.6	YDL229W	SSB1
14	-15.2	4.89	4.6	3	3	59.5	sp|K1C10_HUMAN|	K1C10_HUMAN (keratin)
15	-11.9	5.28	3.9	2	3	50	YBR118W	TEF2
16	-11.6	5.06	18	2	4	14.5	YIL148W	RPL40A
17	-9.9	4.75	7.5	2	2	28.8	YML063W	RPS1B
18	-9.7	4.46	6.4	2	2	33.7	YLR340W	RPP0
19	-9.4	4.45	2.8	2	3	65.8	sp|K2C1_HUMAN|	K2C1_HUMAN (keratin)
20	-9	4.83	5	2	2	34.2	YLR018C	POM34
21	-8.7	5.06	15	2	2	14.5	YCR031C	RPS14A
22	-6.3	4.69	10	2	2	18.2	YBR191W	RPL21A
23	-6	4.39	3.8	1	1	29.4	YHR203C	RPS4B
24	-4.5	4.12	1.7	1	1	74.1	YML031W	NDC1
25	-4.1	5.02	4.2	1	2	26.5	YNL178W	RPS3
26	-4	4.73	1.6	1	1	61.5	YLR044C	PDC1
27	-3.6	4.88	5.2	1	1	25.7	YLR407W	YLR407W
28	-3.3	4.95	2	1	1	37.6	YBR145W	ADH5
29	-3.3	4.48	2.5	1	2	39.6	YKL060C	FBA1
30	-3.1	4.53	2.5	1	1	34.8	YMR116C	ASC1
31	-3	4.12	1.8	1	2	72	YLL004W	ORC3
32	-2.8	4.44	2.7	1	1	32.9	YDR233C	RTN1
33	-2.7	3.88	4.3	1	1	27.4	YFR031C-A	RPL2A
34	-2.6	4.13	6.6	1	1	17	YDR064W	RPS13
35	-2.3	3.9	0.5	1	1	303.3	YLR454W	FMP27
36	-2.2	5.18	2.1	1	2	46.9	YHR174W	ENO2
37	-2	4.85	1.3	1	4	60.5	YHR207C	SET5
38	-2	4.37	5.8	1	1	17.7	YBR048W	RPS11B
39	-1.9	4.55	4.8	1	1	27.9	YLR048W	RPS0B
40	-1.9	3.98	0.6	1	2	93	YLL034C	RIX7
41	-1.8	4.38	4.4	1	1	25	YJR123W	RPS5
42	-1.8	4.37	4.2	1	1	46.4	YLR225C	YLR225C
43	-1.6	4.07	3.8	1	1	38.1	YOL151W	GRE2
44	-1.6	4.64	3.3	1	1	57.2	YMR131C	RRB1
45	-1.6	4.95	1.9	1	1	82	YMR237W	BCH1
46	-1.5	4.45	1.5	1	1	62	YMR199W	CLN1
47	-1.4	4.42	6.9	1	1	22.7	YMR264W	CUE1
48	-1.4	4.89	1.3	1	1	77.5	YLR260W	LCB5
49	-1.4	3.72	0.6	1	1	130.8	YGL006W	PMC1
50	-1.4	4.15	8.5	1	1	8.9	YHR021C	RPS27B
51	-1.3	4.72	1.9	1	1	91.9	YDL126C	CDC48
52	-1.3	3.72	4.8	1	1	29.9	YNL312W	RFA2
53	-1.3	4.8	2.7	1	1	40.2	YBR076W	ECM8
54	-1.3	4.4	1.8	1	1	54.5	YAL038W	CDC19
55	-1.2	4.12	4.8	1	1	17.8	YEL054C	RPL12A
56	-1.2	5.08	2.6	1	1	56.1	YGL202W	ARO8
57	-1.2	4.31	3.2	1	1	52.6	YMR153W	NUP53
58	-1.2	3.41	4.1	1	1	16.7	YKL066W	YKL066W
59	-1.1	5.13	3.6	1	1	15.8	YER050C	RSM18
60	-1.1	4.84	2.8	1	1	42.3	Q0075	AI5_BETA
61	-1.1	5.84	1.5	1	2	140.8	YPL137C	GIP3
62	-1.1	4.59	6.5	1	1	32.1	YDR287W	YDR287W
63	-1.1	4.25	0.4	1	1	190.4	YER190W	YRF1-2
64	-1	3.53	1.1	1	1	114.9	YBL037W	APL3

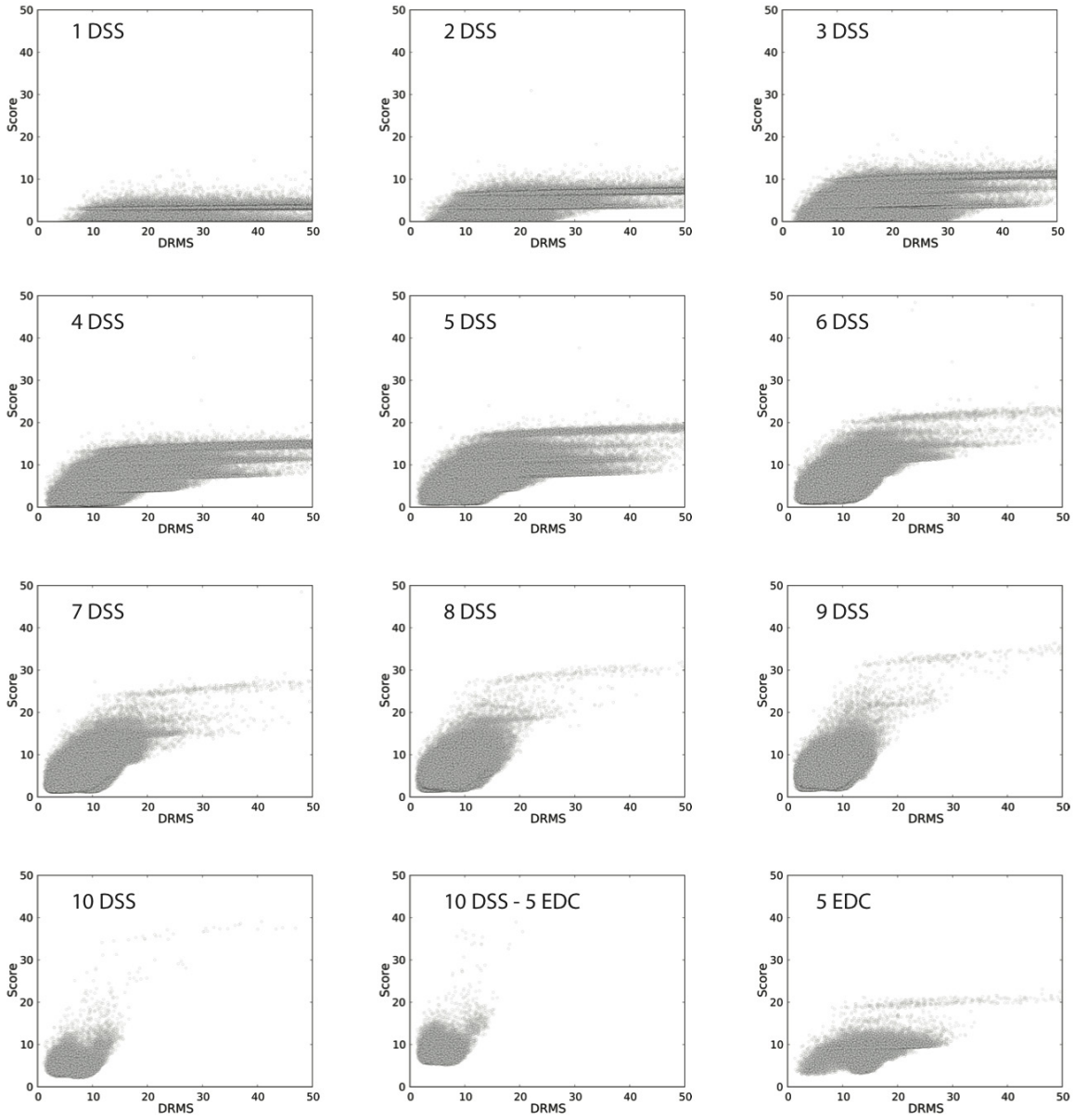
Supplemental Table S7. Protein partners of the Nup84 complex identified by X!Tandem. The FPR of the identification was estimated to be 0.79%.

SUPPLEMENTAL FIGURES



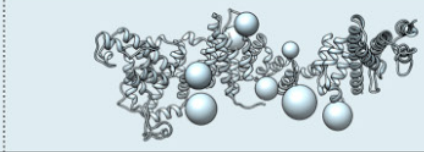
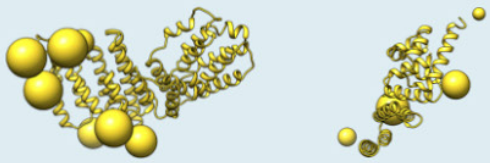
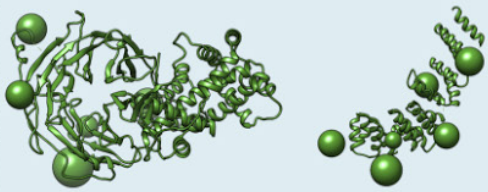
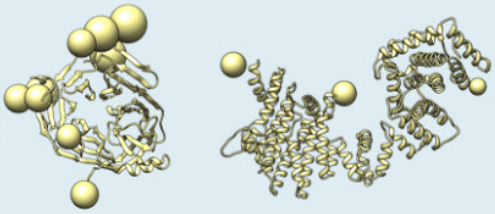
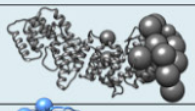
Supplemental Figure S1. Nup84 complex pullout and optimized cross-linking titrations by either DSS or EDC cross-linker

The Nup84 complex was immuno-purified, natively eluted, and titrated to find the optimal cross-linking conditions that show shifting of the complex bands without causing over-crosslinking. On the left, SDS-PAGE gel showing cross-linking titration using 1 mM DSS and different time points (T, in minutes). On the right side, SDS-PAGE gel showing cross-linking titration using increasing concentrations (C, in mM) of EDC. Approximately 0.5 µg of complex per point were used. The time or concentration point selected for further analysis is identified with a red line and arrow. The identity of the proteins is shown in the middle (tagged protein in blue) and the molecular size reference on the left.



Supplemental Figure S2. Correlation between the number of cross-links and the accuracy of dimer models

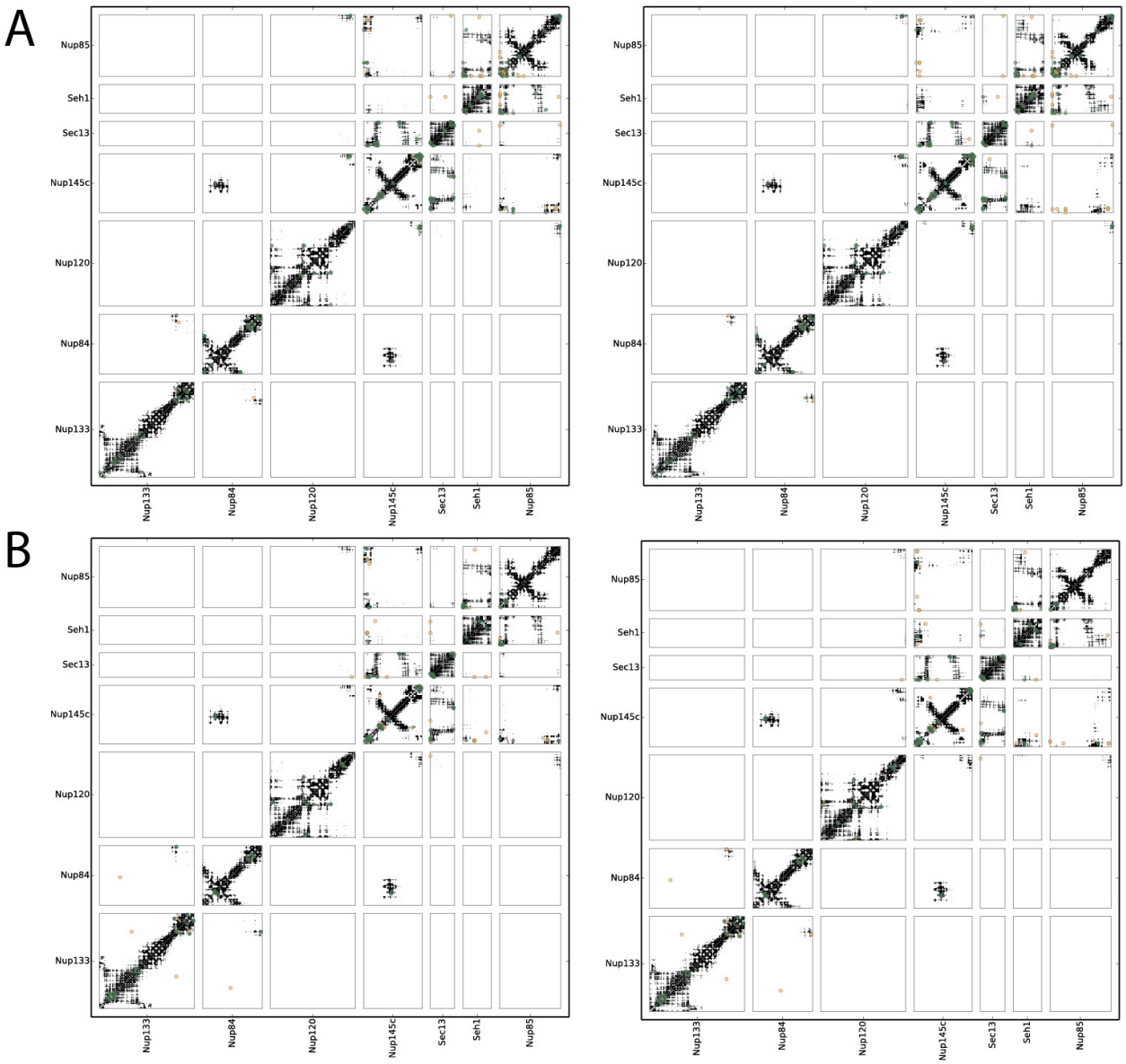
Scores (*i.e.*, the sum of excluded volume and cross-link restraint scores) are plotted as a function of the C_α dRMSD of the Nup145c-Sec13 dimer models with respect to the crystallographic interface (PDB code 3IKO (9)).

7 Components		10 Rigid domains	81 Segments				
			39 structured segments with templates		42 disordered and/or unknown segments without a template		
			begins at	ends at	begins at	ends at	64 flexible beads
Nup 84		**Nup84 (1-726)** 1-6: flexible string of a bead (DISOPRED) 7-436: X-ray 3JRO_C, 3IKO_C (100% seq id) 429-488: Model 3F3F_G (10% seq id, HHpred) 489-505: flexible string of beads (DISOPRED) 506-726: Model 3CQC_A (18% seq id, HHpred)	7 27 96 136 372 506 575	20 80 126 364 484 562 726	1 21 81 127 365 485 563	6 26 95 135 371 505 574	1 1 1 1 1 2 1
Nup 85		**Nup85 NTD (1-529)** 1-43: flexible string of beads (DISOPRED) 44-555: X-ray 3F3F_D, 3EWE_D (100% seq id)	44 135 461	122 427 529	1 123 428	43 134 460	3 1 2
		Linker (flexible string of a bead)			530	532	1
		Nup85 CTD (533-744) 532-655: Model 2QX5_B (18% seq id, HHpred) 532-743 Model 4LCT_A (14% seq id, HHpred) 744: flexible string of a bead (PSIPRED)	533 620 680	602 671 743	603 672 744	619 679 744	1 1 1
Nup120		**Nup120 NTD (1-712)** 1-712: X-ray 3F7F_A, 3HXR_A (100% seq id)	1 53 221 311 440	29 212 305 429 712	30 213 306 430	52 220 310 439	2 1 1 1
		Linker (flexible string of a bead)			713	726	1
		Nup120 CTD (727-1037) 727-1037: Model 4FHN_B, 4FHN_D (14% seq id, HHpred)	727 805 903 921 1023	781 892 910 1010 1037	782 893 911 1011	804 902 920 1022	2 1 1 1
Nup 133		**Nup133 NTD (1-480)** 1-55: flexible string of beads (DISOPRED) 56-480: Model VpNup133 (in press) (46% seq id, Muscle)	56 86 133 162 193 206 258	78 125 144 184 200 249 480	1 79 126 145 185 201 250	55 85 132 161 192 205 257	3 1 1 1 1 1 1
		Linker (flexible string of a bead)			481	489	1
		Nup133 CTD (490-1157) 490-945: Model 3I4R (15% seq id, HHpred) 946-1157: X-ray 3KFO_A (100% seq id)	490 772	763 1155	764 1156	771 1157	1 1
Nup 145c		**Nup145c (1-712)** 1:125: flexible string of beads (DISOPRED) 126-553: X-ray 3IKO_B, 3JRO_A, 3BG1_B, 3BG0_B (100% seq id) 554-712: flexible string of beads (PSIPRED)	126 151 182	144 175 553	1 145 176 554	125 150 181 712	7 1 1 8
Seh1		**Seh1 (1-349)** 1-346: X-ray 3F3F (100% seq id) 347-349: flexible string of a bead (DISOPRED)	1 288	248 346	249 347	287 349	2 1
Sec 13		**Sec13 (1-297)** 1: flexible string of a bead (DISOPRED) 2-296: X-ray 2PM7_D (100% seq id) 297: flexible string of a bead (DISOPRED)	2 166	158 296	1 159 297	1 165 297	1 1 1

Supplemental Figure S3. Representation of the Nup84 complex proteins for integrative modeling

The domains of the Nup84 complex subunits were represented by beads of varying sizes, arranged into either a rigid-body (column 4) or a flexible string (column 5), based on the available crystallographic structures and comparative models. The linkers between rigid bodies are highlighted in red, in column 5. The atomic structures for some

of the yeast Nup84 complex components and their close homologs have been previously determined by X-ray crystallography (9-19). For predicted non-disordered domains of the remaining sequences, comparative models were built with MODELLER 9.13 (20) based on the closest known structure detected by HHPred (21, 22) and the literature. Secondary structure and disordered regions were predicted by PSIPRED (23, 24) and DISOPRED (25), respectively.

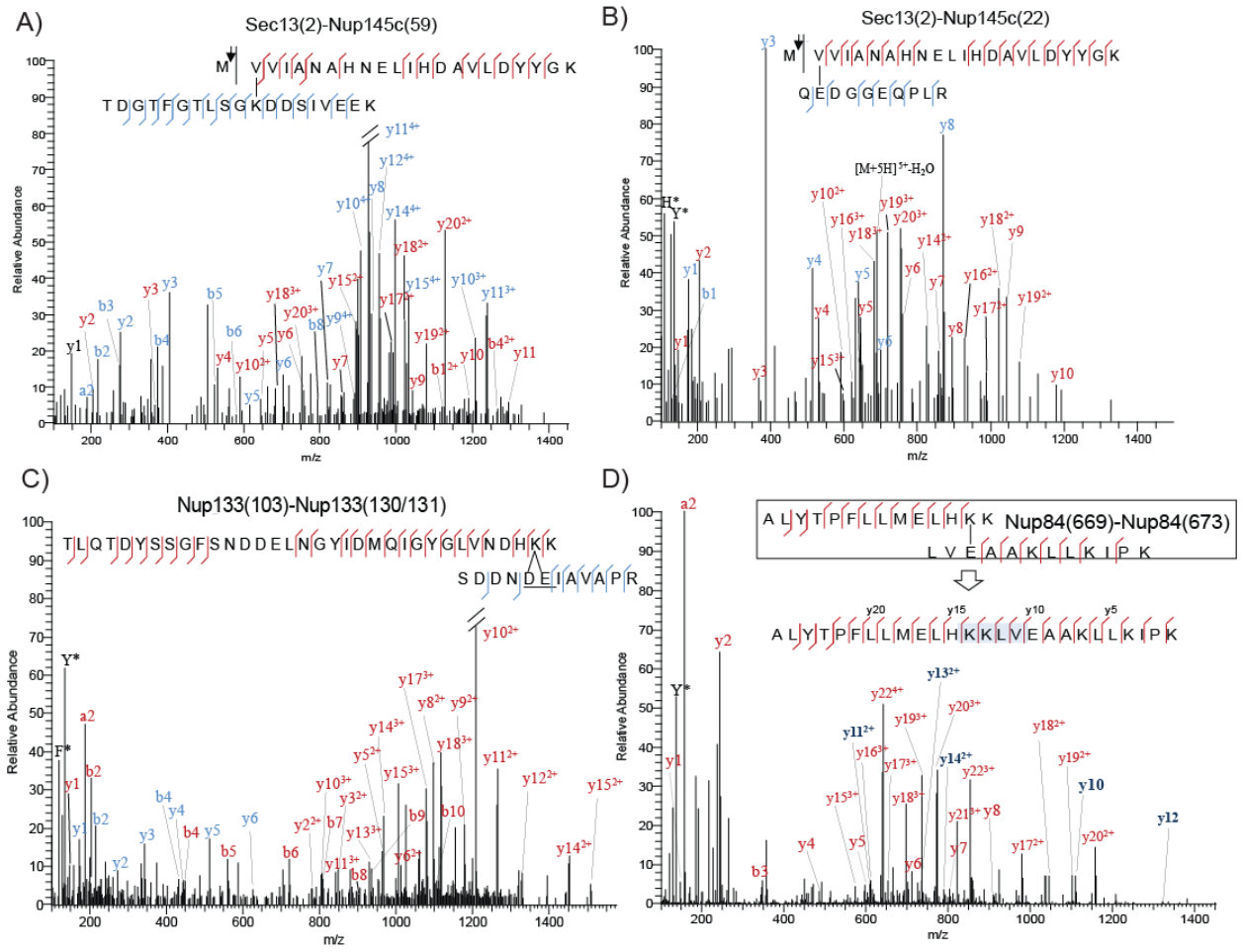


Supplemental Figure S4. The cross-link validation and normalized contact frequency maps

(A) Validation of the 163 DSS cross-links (green circle=satisfied, orange circle=violated), and the normalized contact frequency (scaled in black dots) are plotted together for each of the two clusters (left and right columns, respectively). The three crystallographic interfaces were constrained for generating the two clusters.

(B) Validation of the 123 EDC cross-links (green circle=satisfied, orange

circle=violated), and the normalized contact frequency (scaled in black dots) are plotted together for each of the two clusters (left and right columns, respectively). The three crystallographic interfaces were constrained for generating the two clusters.



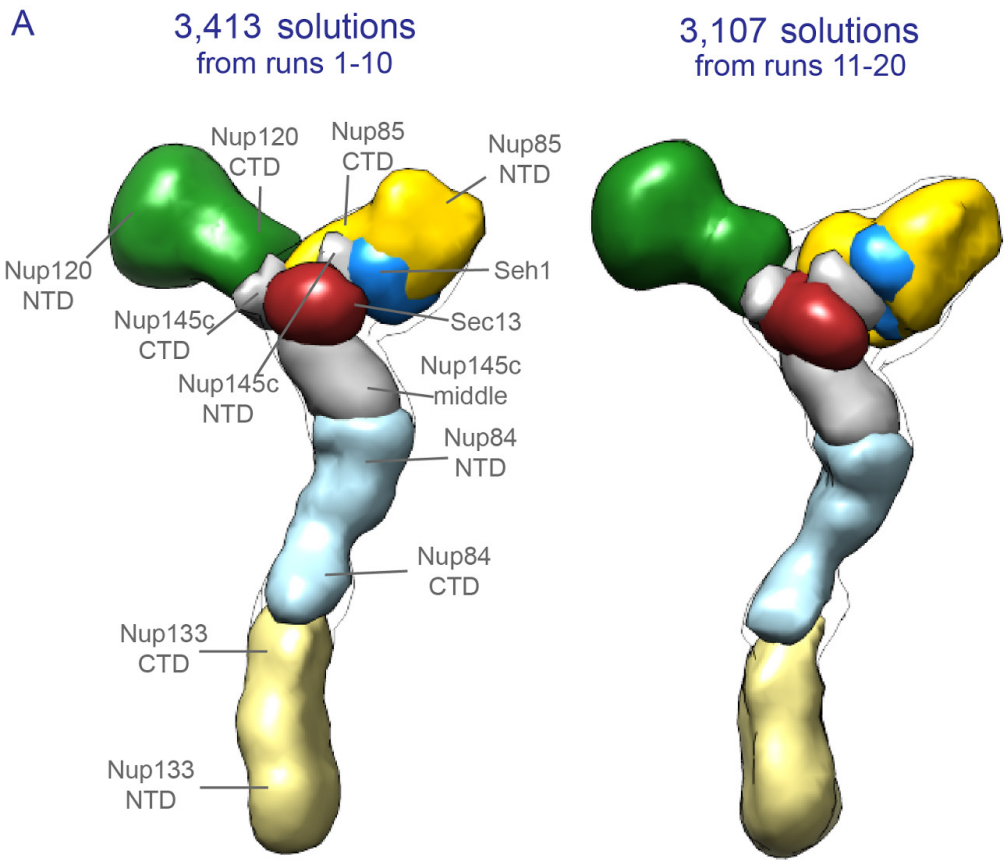
Supplemental Figure S5. The MS/MS spectra of cross-linked peptides

(A) A representative HCD MS/MS spectrum of DSS cross-link (m/z=899.052, z=5) between the second residue of Sec13 (after cleavage of N-terminal methionine) and lysine 59 of Nup145c. b and y ions are labeled accordingly.

(B) A representative HCD MS/MS spectrum of EDC cross-link (m/z=693.547, z=5) between the 2nd residue of Sec13 and glutamic acid 22 of Nup145c. H* and Y* indicate immonium ions of histidine and tyrosine, respectively.

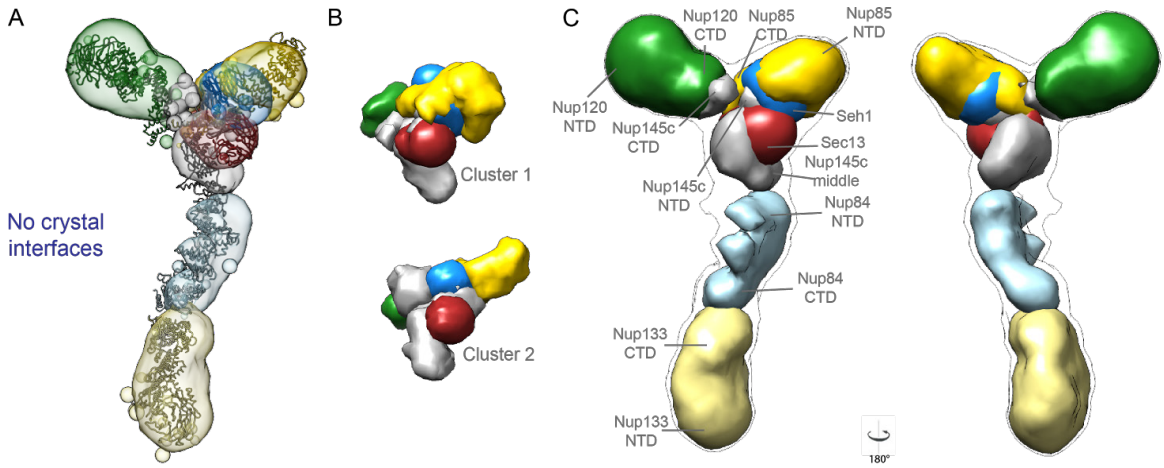
(C) A representative ambiguous EDC cross-link spectrum (m/z=1025.273, z=5) in which either aspartic acid 130 or glutamic acid 131 is potentially cross-linked.

(D) A representative MS/MS spectrum (m/z=602.573, z=5) of a false positive EDC cross-link of Nup84(669) - Nup84(673) identified by the software. The peptide was misassigned as an intra-molecular EDC cross-link and could be distinguished by the y ions (y10-y14).



Supplemental Figure S6. Thoroughness of sampling good-scoring solutions

(A) The thoroughness of configurational sampling was assessed by comparing a subset of 3,413 solutions from runs 1-10 to another subset of 3,107 solutions from independent runs 11-20. Each subset of solutions was converted into a density map of any volume element being occupied by a given protein (the 'localization density map') using VMD (26), contoured at the threshold that results in 2.5 times its volume estimated from sequence (Supplemental Table S6). Importantly, the two localization density maps were similar to each other, demonstrating that the sampling is likely to have sampled well all solutions that satisfy the input restraints.



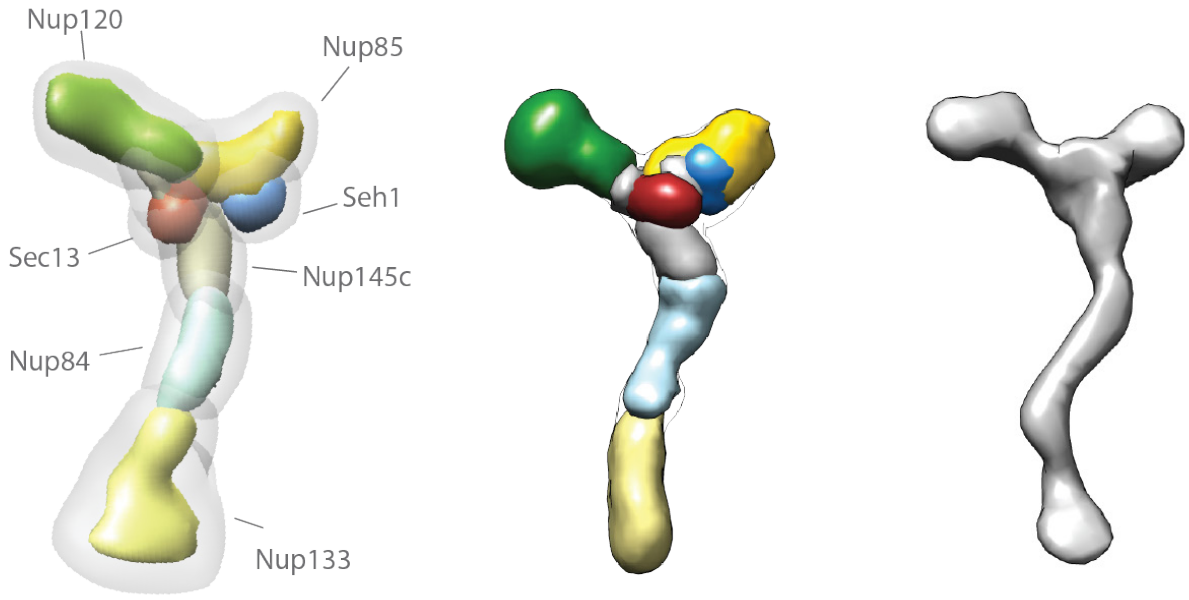
Supplemental Figure S7. An ensemble of solutions without constraining the 3 crystallographic interfaces

Modeling calculations were repeated by omitting the three crystallographic interface constraints (14, 18, 27). The localization density maps of the Nup84 subunits (solid contour surfaces) and the entire complex (transparent surfaces) were then computed using the resulting 4,286 solutions, and contoured at the threshold of approximately 3 times its volume estimated from sequence. The similarity of this localization probability map to that obtained with the 3 crystallographic interface constraints validates the *in vivo* relevance of the crystallographic interfaces.

(A) A representative single Nup84 complex structure (colored ribbon) is shown along with the localization density maps of the individual subunits.

(B) The localization density maps of the two clusters (comprising 577 and 404 solutions, respectively) computed on the hub region (Nup120-CTD, Nup85, Nup145c, Sec13, and Seh1) are shown.

(C) Front and back views of the localization density maps of the Nup84 subunits and the entire complex.



Supplemental Figure S8. **Assessment of the CX-MS integrative Nup84 complex structure through comparison with previously determined structures using other sources of data**

(LEFT) For comparison, a localization density map in our previous study is shown, calculated primarily using the domain mapping data and the EM class average (27).

(MIDDLE) The molecular architecture presented in this study is shown, calculated primarily using the CX-MS and the EM class average.

(RIGHT) A negative-stain EM 3D map is shown for the same complex, generated from single particle EM reconstruction (28).

For Supplemental Figures S9-S11, the annotated spectra are separately attached.

Supplemental Figure S9. 163 Unique MS/MS HCD spectra of DSS cross-links reported in the current study (prepared by spectrum viewer pLabel 2.4, attached separately)

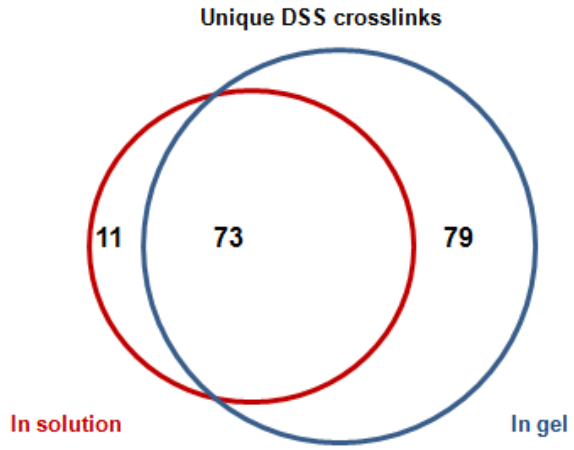
The MS2 tolerance is < 20 ppm.

Supplemental Figure S10. 104 Unique MS/MS HCD spectra of EDC cross-links reported in this study (prepared by spectrum viewer pLabel 2.4, attached separately).

The MS2 tolerance is < 20 ppm. We were unable to unambiguously locate the conjugation sites for ~20% of the EDC cross-links, because many result from peptides containing consecutive and/or adjacent carboxylic acids. These ambiguities were accounted for in our modeling calculations, resulting in 123 possible EDC cross-links used for further analysis.

Supplemental Figure S11. A few examples of potential false positive identifications at expected 5% FDR after use of additional filters/ manual verification of the spectra (prepared by spectrum viewer pLabel 2.4, attached separately).

The MS2 tolerance is < 20 ppm.



Supplemental Figure S12. A comparison of in-gel and in solution digestion of the unique DSS cross-linked peptides identified from the Nup84 cross-linked samples.

SUPPLEMENTAL REFERENCES

1. Rieping, W., Habeck, M., and Nilges, M. (2005) Inferential structure determination. *Science* 309, 303-306

2. Erzberger, J., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C., Cimermancic, P., Boehringer, D., Sali, A., Aebersold, R., and Ban, N. (2014) Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell*, in press

3. Pellarin, R., Bonomi, M., Spill, Y., Nilges, M., DeGrado, W., and Sali, A. (in preparation) Modeling multiple structural states of macromolecules by cysteine cross-linking.

4. Schneidman-Duhovny, D., Rossi, A., Avila-Sakar, A., Kim, S. J., Velazquez-Muriel, J., Strop, P., Liang, H., Krukenberg, K. A., Liao, M., Kim, H. M., Sobhanifar, S., Dotsch, V., Rajpal, A., Pons, J., Agard, D. A., Cheng, Y., and Sali, A. (2012) A method for integrative structure determination of protein-protein complexes. *Bioinformatics (Oxford, England)* 28, 3282-3289

5. Velazquez-Muriel, J., Lasker, K., Russel, D., Phillips, J., Webb, B. M., Schneidman-Duhovny, D., and Sali, A. (2012) Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proceedings of the National Academy of Sciences of the United States of America* 109, 18821-18826

6. Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008) Protein structure fitting and refinement guided by cryo-EM density. *Structure* 16, 295-307

7. Saff, E. B., and Kuijlaars, A. B. J. (1997) Distributing many points on a sphere. *Mathematical Intelligencer* 19, 5-11

8. Ballard, D. H., and Brown, C. M. (1982) *Computer vision*, Prentice-Hall

9. Nagy, V., Hsia, K. C., Debler, E. W., Kampmann, M., Davenport, A. M., Blobel, G., and Hoelz, A. (2009) Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proceedings of the National Academy of Sciences of the United States of America* 106, 17693-17698

10. Berke, I. C., Boehmer, T., Blobel, G., and Schwartz, T. U. (2004) Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J Cell Biol* 167, 591-597

11. Whittle, J. R., and Schwartz, T. U. (2009) Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *The Journal of biological chemistry*

12. Boehmer, T., Jeudy, S., Berke, I. C., and Schwartz, T. U. (2008) Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Molecular cell* 30, 721-731

13. Sampathkumar, P., Gheyi, T., Miller, S. A., Bain, K. T., Dickey, M., Bonanno, J. B., Kim, S. J., Phillips, J., Pieper, U., Fernandez-Martinez, J., Franke, J. D., Martel, A., Tsuruta, H., Atwell, S., Thompson, D. A., Emtage, J. S., Wasserman, S. R., Rout, M. P., Sali, A., Sauder, J. M., and Burley, S. K. (2011) Structure of the C-terminal domain of Saccharomyces cerevisiae Nup133, a component of the nuclear pore complex. *Proteins* 79, 1672-1677

14. Brohawn, S. G., and Schwartz, T. U. (2009) Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat Struct Mol Biol* 16, 1173-1177

15. Seo, H. S., Ma, Y., Debler, E. W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A.

(2009) Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proceedings of the National Academy of Sciences of the United States of America* 106, 14281-14286

16. Leksa, N. C., Brohawn, S. G., and Schwartz, T. U. (2009) The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* 17, 1082-1091

17. Debler, E. W., Ma, Y., Seo, H. S., Hsia, K. C., Noriega, T. R., Blobel, G., and Hoelz, A. (2008) A fence-like coat for the nuclear pore membrane. *Molecular cell* 32, 815-826

18. Brohawn, S. G., Leksa, N. C., Spear, E. D., Rajashankar, K. R., and Schwartz, T. U. (2008) Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* 322, 1369-1373

19. Fath, S., Mancias, J. D., Bi, X., and Goldberg, J. (2007) Structure and organization of coat proteins in the COPII cage. *Cell* 129, 1325-1336

20. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* 234, 779-815

21. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics (Oxford, England)* 21, 951-960

22. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* 33, W244-248

23. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* 292, 195-202

24. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* 41, W349-357

25. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics (Oxford, England)* 20, 2138-2139

26. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14, 33-38, 27-38

27. Fernandez-Martinez, J., Phillips, J., Sekedat, M. D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J. D., Williams, R., Stokes, D. L., Chait, B. T., Sali, A., and Rout, M. P. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. *J Cell Biol* 196, 419-434

28. Kampmann, M., and Blobel, G. (2009) Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nat Struct Mol Biol* 16, 782-788