Supporting Information

Quan et al. 10.1073/pnas.1409076111

SI Materials and Methods

Protein Expression and Purification. The vesicular stomatitis virus matrix (M) protein (residues 44–229) was cloned into a pGEX 4T-1 vector (GE Healthcare) modified to contain a tobacco etch virus (TEV) protease cleavable N-terminal GST tag. Ala-Val-Leu-Ala (residues 121–124) of the M protein was replaced with Gly-Gly-Ser-Gly by QuikChange mutagenesis (Stratagene) to improve solubility and avoid aggregation. The mutated M protein (residues 44–229 containing the AVLA to GGSG mutation) is denoted by $*M_{44–229}$. Human mRNA export factor 1 (Rae1; residues 31–368; denoted by $Rae1_{31–368}$) and human nucleoporin 98 (Nup98) GLEBS motif (residues 157–213; denoted by $Nup98_{157–213}$) were cloned into a pFastBac Dual vector (Invitrogen) modified to contain a TEV-cleavable N-terminal His tag (1). The expression of His_6-Nup98 and Rae1 was under control of polyhedrin and p10 promoters, respectively.

GST-tagged $*M_{44–229}$ was expressed in *Escherichia coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene). Protein expression was induced at an OD_{600} of 1.0 with 0.5 mM isopropyl-β-D-thiogalactoside at 20 °C for 16 h in LB containing 100 μg/L ampicillin and 34 μg/L chloramphenicol. Cells were harvested and lysed with a cell disruptor (Avestin) in a wash buffer [50 mM Hepes, pH 7.0, 300 mM NaCl, 2 mM β-mercaptoethanol (β-ME)] supplemented with a protease inhibitor mix containing 1 mM PMSF, 5 mg/L aprotinin, 1 mg/L pepstatin, and 1 mg/L leupeptin. The GST-$*M_{44–229}$ protein was first purified using a glutathione Sepharose 4 fast-flow column (GE Healthcare). The column was washed extensively using the wash buffer, and GST-$*M_{44–229}$ was eluted using the wash buffer supplemented with 10 mM glutathione. After GST tag removal by TEV, the protein sample was applied to a HiTrap SP column (GE Healthcare) and separated using a 200–400 mM NaCl gradient in a buffer containing constant 10 mM Hepes (pH 7.0). $*M_{44–229}$ was further purified by size-exclusion chromatography using a Superdex 200 column (GE Healthcare) equilibrated with a storage buffer [10 mM Hepes, pH 7.0, 300 mM NaCl, 0.5 mM Tris[2-carboxyethyl] phosphine (TCEP)].

For coexpression of $Rae1_{31–368}$ and $Nup98_{157–213}$, Hi5 cells (Invitrogen) were infected with a recombinant baculovirus and grown in suspension for 2 d. The cells were harvested and lysed by sonication in a lysis buffer containing 50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole, 2 mM β-ME, 1 mM PMSF, 5 mg/L aprotinin, 1 mg/L pepstatin, and 1 mg/L leupeptin. The protein complex was purified by Ni^{2+}-affinity chromatography followed by TEV digestion to remove the N-terminal His_6 tag on $Nup98_{157–213}$. The presence of both $Nup98_{157–213}$ and $Rae1_{31–368}$ was confirmed by SDS/PAGE. The protein-containing fractions were applied to a HiTrap Q column (GE Healthcare) in a buffer containing 10 mM Tris (pH 8.0), 100 mM NaCl, and 2 mM β-ME; under this condition, Rae1•Nup98 does not bind to the column. $Rae1_{31–368}$•$Nup98_{157–213}$ complex-containing fractions were pooled, concentrated, and further purified over a Superdex 200 column (GE Healthcare) equilibrated with storage buffer. Equimolar amounts of $Rae1_{31–368}$•$Nup98_{157–213}$ heterodimer and $*M_{44–229}$ protein were mixed and subjected to size-exclusion chromatography using a Superdex 200 column in storage buffer. Fractions containing the $*M_{44–229}$•$Rae1_{31–368}$•$Nup98_{157–213}$ complex (referred to as M•Rae1•Nup98) were then concentrated to 5 mg/mL and flash-frozen in liquid nitrogen for storage at −80 °C. A complex of full-length Rae1 ($Rae1_{FL}$; residues 1–368) and $Nup98_{157–213}$ was prepared as previously reported (1) and used for light-scattering analyses and EMSAs.

Crystallization and Structure Determination. Crystals of M•Rae1•Nup98 were obtained at 20 °C by vapor diffusion in sitting drops using 1 μL complex (at 5 mg/mL) and 1 μL reservoir solution consisting of 0.1 M Hepes (pH 7.5), 11% (wt/vol) PEG 10000, and 10% (vol/vol) 2-methyl-2,4-pentanediol. The crystals grew to their maximum size of 200 × 200 × 50 μm within 3 d. The crystals were transferred in three steps of increasing glycerol concentration to a cryoprotectant solution containing 0.1 M Hepes (pH 7.5), 11% (wt/vol) PEG 10000, 10% (vol/vol) 2-methyl-2,4-pentanediol, and 20% (vol/vol) glycerol. X-ray diffraction data were collected at the 24ID-C Beamline at the Advanced Photon Source, Argonne National Laboratory. X-ray intensities were processed using the HKL2000 denzo/scalepack package (2), and subsequent calculations were carried out using the CCP4 program package (3). Initial electron density map was obtained by molecular replacement using the coordinates of Rae1•Nup98 (Protein Data Bank ID code 3MMY) and the vesicular stomatitis virus Indiana M protein (Protein Data Bank ID code 1LG7) as search models in Phaser (4). The model of M•Rae1•Nup98 was built using Coot (5) and refined using Refmac (6). Details of the data collection and refinement statistics are in Table S1.

Multiangle Light Scattering. The purified $Rae1_{FL}$•$Nup98_{157–213}$ complex, the $*M_{44–229}$ protein, and the preincubated mixture of all three proteins were injected separately at 2 mg/mL onto a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in a buffer containing 10 mM Tris (pH 8.0), 150 mM NaCl, and 0.5 mM TCEP. The chromotography system was connected to a multiangle light-scattering instrument consisting of an 18-angle light-scattering detector (DAWN HELEOS) and a refractive index detector (Optilab-rEX; Wyatt Technology). Data analysis was carried out with the ASTRA program (Wyatt Technology).

EMSA. A 10-mer poly(U) ssRNA was synthesized with Alexa-488 at the 5′ end (Integrated DNA Technologies). A solution of ssRNA was first incubated with a solution of $Rae1_{FL}$•$Nup98_{157–213}$ in a buffer containing 10 mM Tris (pH 8.0), 150 mM NaCl, 0.5 mM TCEP, and 5% (vol/vol) glycerol for 5 min at room temperature. The $*M_{44–229}$ protein was added to the mixture at different concentrations (0–2.5 μM) and incubated for an additional 10 min at room temperature. The final concentrations of ssRNA, $Rae1_{FL}$•$Nup98_{157–213}$, and $*M_{44–229}$ protein in a 10-μL mixture were 0.2, 2.0, and 0–2.5 μM, respectively, as indicated in Fig. 3; 10-μL samples were loaded onto a prerun 6% native PAGE gel that was prepared and run in a 45 mM Tris (pH 8.5; titrated with boric acid) buffer. After electrophoresis, the ssRNA was visualized using the fluorescence signal from Alexa-488. To determine the region of $*M_{44–229}$ that is involved in displacing ssRNA from $Rae1_{FL}$•$Nup98_{157–213}$, a 13-mer finger peptide (residues 49–61) and an 11-mer thumb peptide (residues 213–223) were synthesized and individually tested by EMSA as described above at the final concentration indicated in Fig. 3. To test for different nucleic acids, we used five different types: 21-mer random sequence RNA and DNA libraries, 21-mer gene recruitment sequence I dsRNA and dsDNA (7), and yeast tRNA (Roche). Samples were prepared as described above at the final concentration indicated in Fig. 4 and visualized by SYBR-GOLD staining (Invitrogen).

Interface Calculation, Illustration, and Figures. Both the interface residues and the buried surface area between the proteins were calculated using the PDBePISA server (www.ebi.ac.uk/msd-srv/prot_int/).

Figures were generated using PyMOL (www.pymol.org). The electrostatic potential maps were calculated using APBS (8). Sequence alignments were generated using ClustalX (9) and colored using Aline (10).

1. Ren Y, Seo H-S, Blobel G, Hoelz A (2010) Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. *Proc Natl Acad Sci USA* 107(23):10406–10411.
2. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
3. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
4. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4): 658–674.
5. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
6. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
7. Seo H-S, Blus BJ, Jankovic NZ, Blobel G (2013) Structure and nucleic acid binding activity of the nucleoporin Nup157. *Proc Natl Acad Sci USA* 110(41):16450–16455.
8. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.
9. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947–2948.
10. Bond CS, Schüttelkopf AW (2009) ALINE: A WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* 65(Pt 5):510–512.

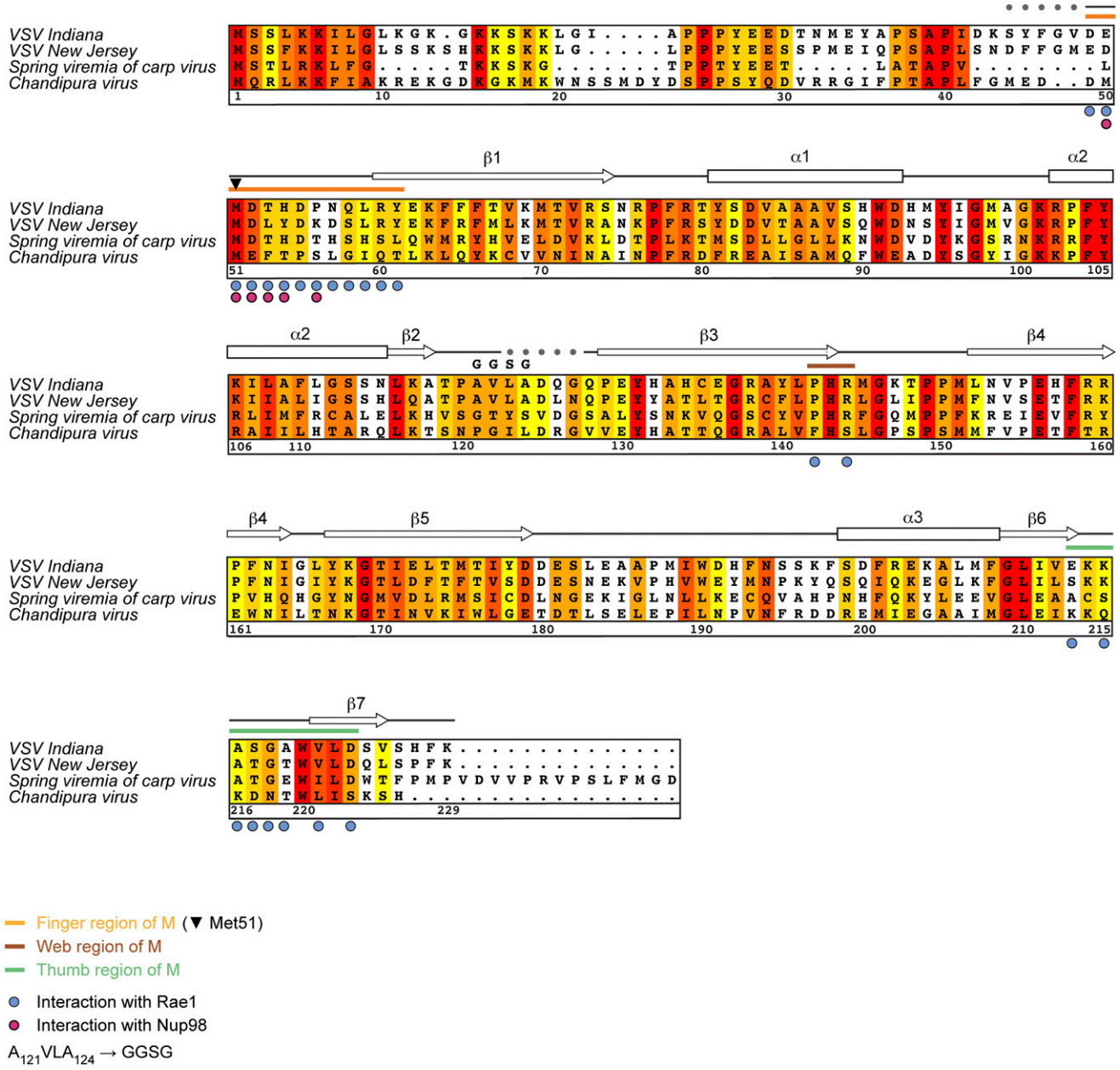


Fig. S1. Multispecies sequence alignment of M protein. Representative sequences of M homologs are aligned with respect to vesicular stomatitis virus (VSV) Indiana M. Overall sequence conservation at each position is shaded in a color gradient from yellow (40% similarity) to red (100% identity). The numbering of residues shown below the alignment is that of the VSV Indiana M protein. Secondary structure elements of the VSV Indiana M protein are indicated by arrows (β-strands), rectangles (α-helices), and gray lines (connecting loops). Dotted lines mark missing residues from the final structure because of disorder. Blue and magenta dots below the alignment indicate residues involved in Rae1 and Nup98 binding, respectively. Three prominent features of the finger, web, and thumb regions of M are marked by lines above the residues and colored in orange, brown, and green, respectively (Figs. 1 and 2). Methionine at residue 51 (Met51) is indicated by the upside-down triangle.

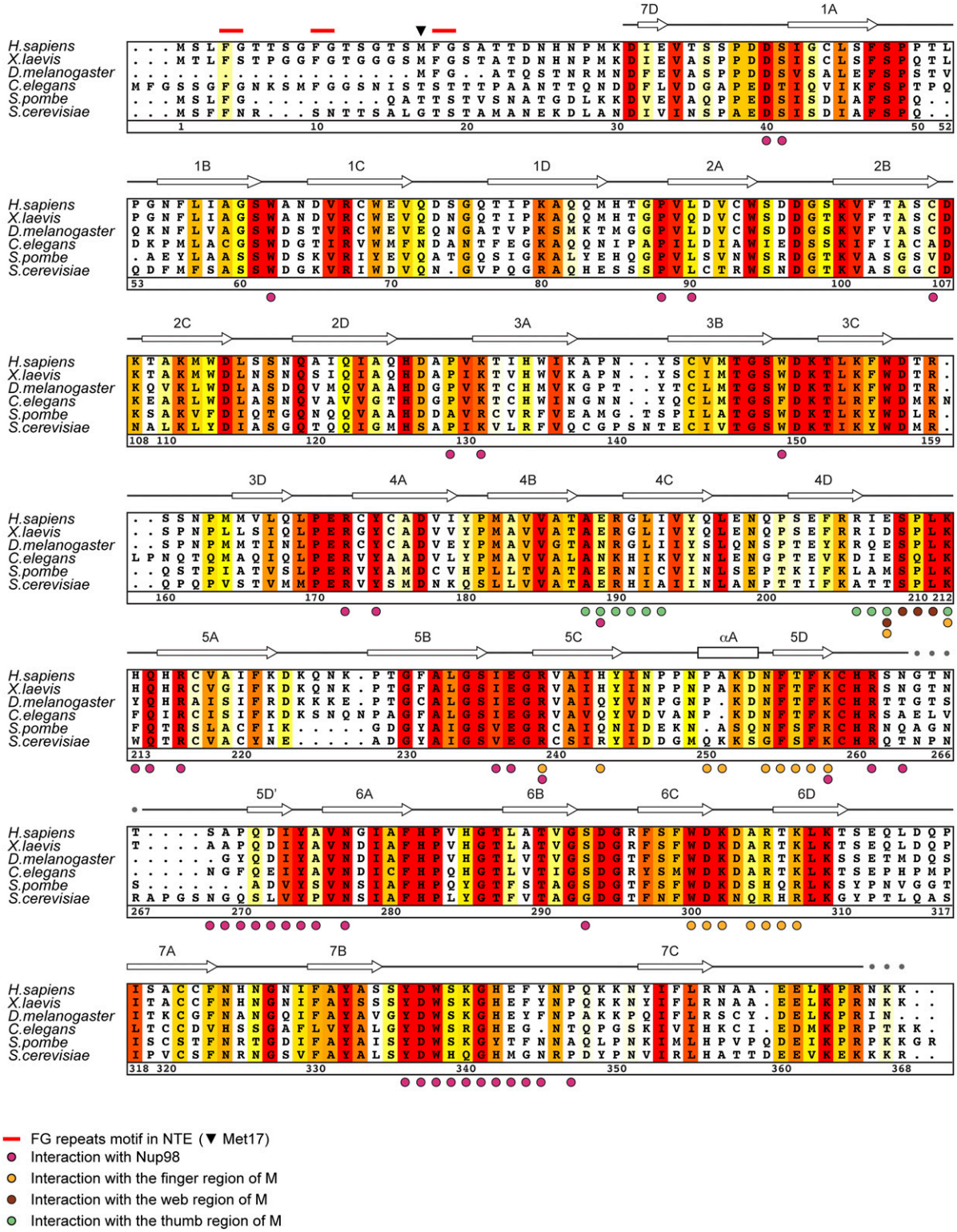
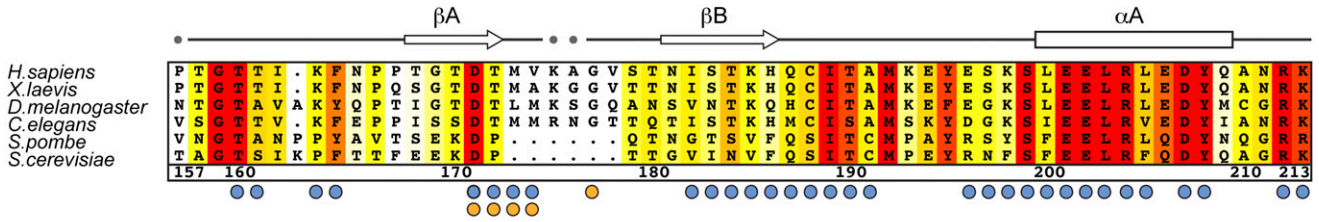


Fig. S2. Multispecies sequence alignment of Rae1 homologs. Representative sequences of Rae1 homologs are aligned with respect to human Rae1. Overall sequence conservation at each position is shaded in a color gradient from yellow (60% similarity) to red (100% identity). The numbering of residues shown below the alignment is that of human Rae1. Secondary structure elements are indicated by arrows (β-strands), rectangles (α-helices), and gray lines (connecting loops). Dotted lines mark missing residues from the final structure because of disorder. Magenta and yellow dots below the alignment mark residues involved in the interactions with Nup98 and M protein, respectively. Phenylalanine-Glycine (FG) repeats in the N-terminal extension (NTE) of Rae1 are marked by a red line above the residues, and Methionine at residue 17 (Met17) is indicated by the upside-down triangle.



Interaction with Rae1

Interaction with the finger region of M

Fig. S3. Multispecies sequence alignment of the Nup98 GLEBS motif. Representative sequences of Nup98 homologs are aligned with respect to human Nup98. Overall sequence conservation at each position is shaded in a color gradient from yellow (40% similarity) to red (100% identity). The numbering of residues shown below the alignment is that of human Nup98. Secondary structure elements are indicated by arrows (β-strands), rectangles (α-helices), and gray lines (connecting loops). Dotted lines mark missing residues from the final structure because of disorder. Blue and yellow dots below the alignment indicate residues involved in the interaction with Rae1 and M protein, respectively.

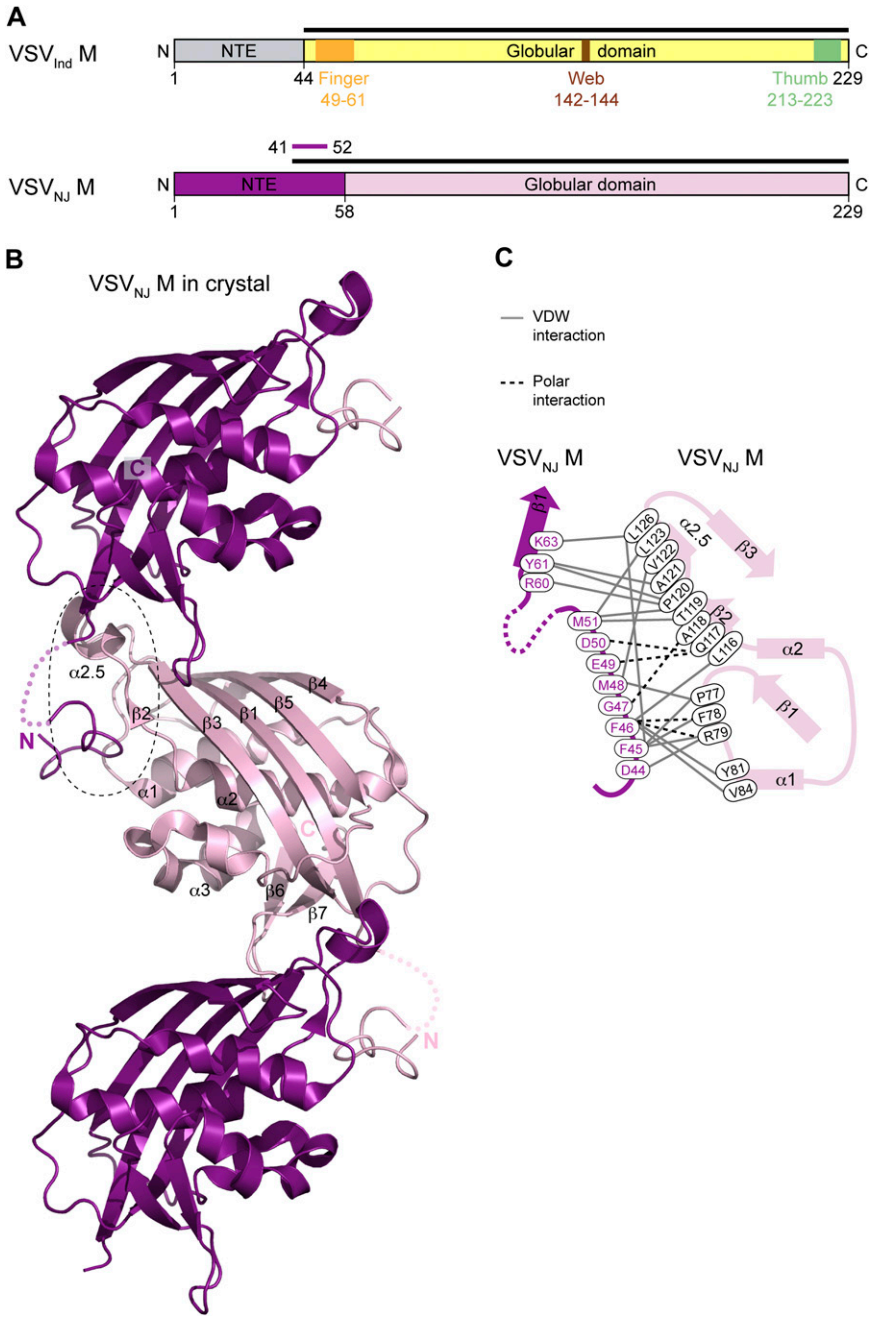


Fig. S4. Vesicular stomatitis virus New Jersey (VSV$_{NJ}$) M protein forms noncovalent linear polymers *in crystallo*. (*A*) Schematic representation of the domain architecture. The boundaries for each domain are color-coded and marked with residue numbers. For vesicular stomatitis virus Indiana (VSV$_{Ind}$) M protein, the N-terminal extension segment (NTE; gray) and globular domain (yellow) are indicated. In the globular domain of VSV$_{Ind}$ M protein, positions of the three prominent features interacting with Rae1•Nup98 are indicated with residue numbers: finger (orange), web (dark brown), and thumb (green). For VSV$_{NJ}$ M protein, NTE (magenta) and globular domain (pink) are indicated. The magenta bar above the domain architecture represents the segment (residues 41–52) that forms intermolecular contacts. Black bars above the domain architecture correspond to the crystallized fragments. (*B*) The interactions between the NTE and the globular domain of VSV$_{NJ}$ M protein that give rise to the noncovalent linear polymers observed in the crystals are shown (Protein Data Bank ID code 2W2R). Three VSV$_{NJ}$ M proteins are colored in alternating magenta and pink. The dotted lines indicate missing residues between the NTE and the globular domain. (*C*) Residues involved at the intermolecular interface between the NTE and the globular domain of VSV$_{NJ}$ M protein are shown. Black dotted lines indicate polar interactions including salt bridges and hydrogen bonds. Gray lines indicate van der Waals (VDW) interactions.

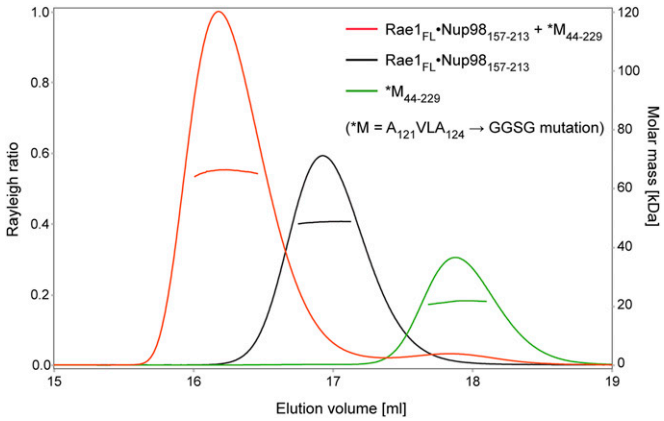


Fig. S5. Assembly of the M•Rae1•Nup98 complex analyzed by size-exclusion chromatography coupled to multiangle light scattering. The Rayleigh ratio and the measured molar masses of $*M_{44–229}$, $Rae1_{FL}$•$Nup98_{157–213}$, and the mixture of both are plotted against the elution volumes from a Superdex 200 10/300 GL gel filtration column. The measured molar mass of the $*M_{44–229}$•$Rae1_{FL}$•$Nup98_{157–213}$ complex (65.4 ± 0.7 kDa) is in agreement with the calculated molecular mass of a complex with 1:1:1 stoichiometry (69.7 kDa).

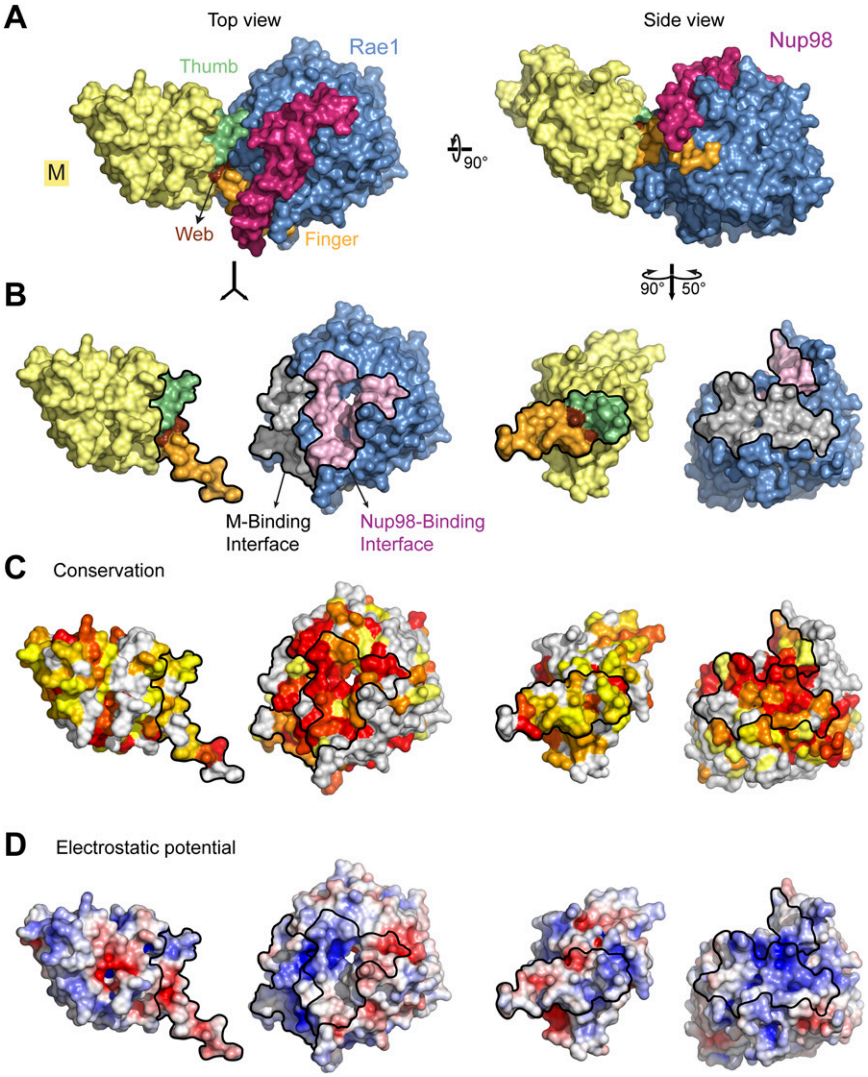


Fig. S6. Surface properties of the interface within M•Rae1•Nup98. (*A*) Space-filling model of the trimeric complex is shown in the same orientation and colored as in Fig. 1*D*; *Left* is a top view, and *Right* is a side view. (*B*) Interfaces of M, Rae1, and Nup98 are exposed on the space-filling models of M and Rae1. A top view of M and Rae1 is shown separately in *Left*. M and Rae1 are rotated with respect to each other for a better view of the interfaces and presented in *Right*. Residues of Rae1 at the interfaces with M and Nup98 are colored in light gray and magenta, respectively, and contoured using a black line. Residues of the M protein involved in forming three prominent features of the structure (finger, web, and thumb) that interact with Rae1 are colored in orange, dark brown, and green, respectively, and contoured using a black line. (*C*) Conserved residues of M and Rae1 (Figs. S1 and S2) are mapped onto the surface and shaded in a color gradient from light yellow to red (40–100% sequence conservation for M and 60–100% sequence conservation for Rae1). (*D*) Electrostatic potential surface of M and Rae1 is depicted in a gradient from red (–5 k_BT/e) to blue (+5 k_BT/e). The orientations of all surface representations in *B*–*D* are identical in each column.

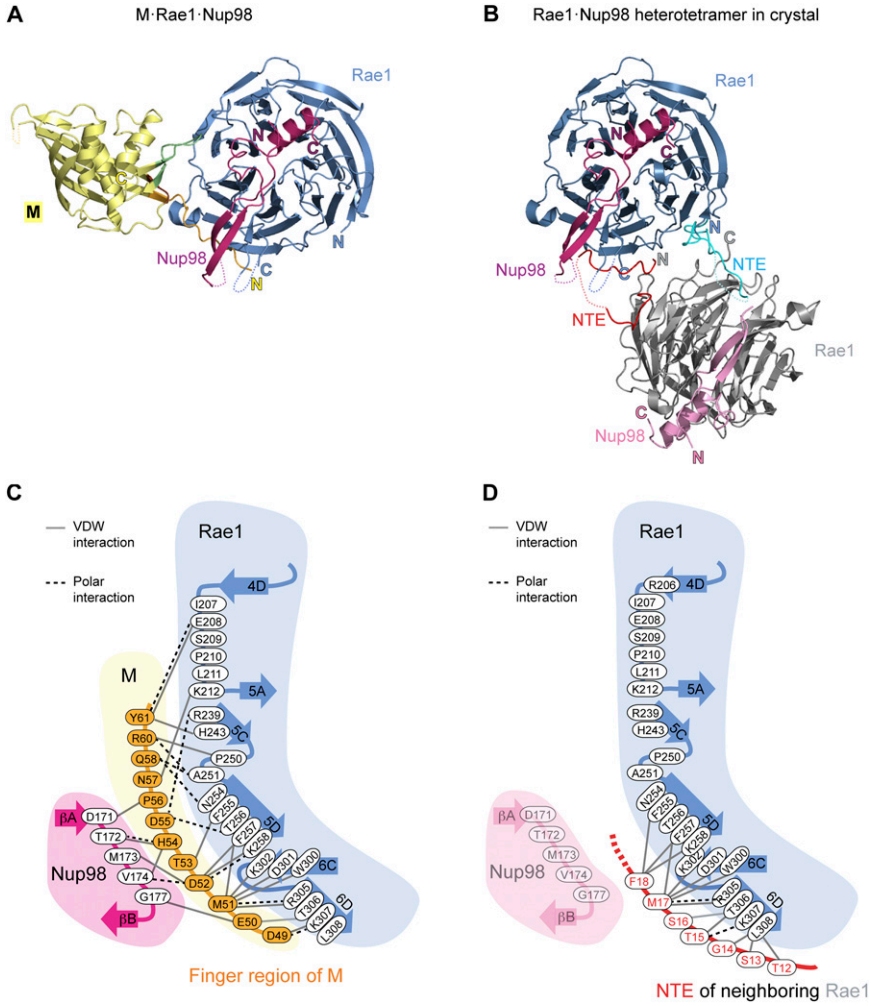


Fig. S7. Structural comparison between M•Rae1•Nup98 and Rae1•Nup98 heterotetramer. (*A*) The structure of $*M_{44–229}$•$Rae1_{31–368}$•$Nup98_{157–213}$ is colored as in Fig. 1. (*B*) Two heterodimers of $Rae1_{FL}$•$Nup98_{157–213}$ are colored in gray and pink for one heterodimer and blue and magenta for the other heterodimer. The N-terminal extension (NTE) segments of Rae1s are colored in light blue and red. All residues involved at the intermolecular interface (*C*) between the finger region of M and Rae1•Nup98 and (*D*) between NTE of Rae1 and neighboring Rae1•Nup98 are shown. Black dotted lines indicate polar interactions, including salt bridges and hydrogen bonds. Gray lines indicate van der Waals (VDW) interactions.

Table S1. Data collection and refinement statistics

	*$M_{44–229}$•$Rae1_{31–368}$•$Nup98_{157–213}$
Data collection	
Synchrotron	Advanced Photon Source
Beamline	24ID-C
Space group	$P42_12$
Cell dimensions	
a, *b*, *c* (Å)	*a* = 141.4, *b* = 141.4, *c* = 78.5
α, β, γ (°)	α = 90.0, β = 90.0, γ = 90.0
Wavelength (Å)	0.97950
Resolution (Å)	30–3.15 (3.20–3.15)
R_{sym} (%)	9.3 (62.7)
$<I/\sigma I>$	15.5 (2.3)
Completeness (%)	97.2 (97.3)
Redundancy	5.0 (4.6)
Refinement	
Resolution (Å)	30–3.15
No. reflections	
Total	12,717
Test set	666
R_{work}/R_{free} (%)	21.3/28.9
No. atoms	4,471
B factor	46.9
rmsds	
Bond lengths (Å)	0.008
Bond angles (°)	1.07
Ramachandran statistics (%)	
With favored	93.4
Within allowed	100.0
Outliers	0

Highest-resolution shell is shown in parentheses.