Supporting Information

Barnwal et al. 10.1073/pnas.1214102110

SI Experimental Procedures

Cloning, Overexpression, and Purification. All proteins were cloned with N-terminal His_6-GB1 and GST tags unless specifically mentioned. His_6-tagged constructs were used for NMR spectroscopy, whereas GST-tagged proteins were used for pull-down experiments. All constructs were expressed in Rosetta cells grown in LB or minimal M9 media containing kanamycin or chloramphenicol as needed. Cells were grown at 37 °C followed by a temperature drop to either 30 °C or 18 °C before IPTG induction. Induced cells were further grown for 4–14 h followed by pelleting and storage at −80 °C. His_6-tagged proteins were purified using nickel-affinity chromatography. The His_6-tag was removed by cleavage with tobacco etch virus (TEV) protease, followed by size-exclusion chromatography. GST-tagged proteins were lysed and kept in 250 μL stocks at −80 °C before their use in pull-down experiments.

Sixteen-nucleotide-long RNA oligonucleotides were purchased from IDT (www.idtdna.com). The lyophilized samples were dissolved in water, ethanol precipitated and purified by polyacrylamide gel electrophoresis (PAGE). The RNA was dialyzed against 100 mM NaCl/20 mM Tris (pH 8.0) buffer overnight, followed by lyophilization and storage at −20 °C until use. The lyophilized powder was dissolved in H_2O or 99.9% ^2H_2O depending on the NMR experiments.

Site-Directed Mutagenesis. Thirteen single mutants and 7 multiple mutants of Hrp1 were generated using QuikChange Site-directed mutagenesis kit (Agilent) (Table S1). The primers used for mutagenesis are given below. All recombinant plasmids were confirmed by DNA sequencing.

FORWARD PRIMER (L166Q) - g ttc att ggt ggt CAG aat tgg gac act ac

REVERSE PRIMER (L166Q) - GTA GTG TCC CAA TTC TGA CCA CCA ATG AAC

FORWARD (L175S) - ct acg gaa gat aat TCT cgc gaa tat ttt g

REVERSE (L175S) - CAA AAT ATT CGC GAG AAT TAT CTT CCG TAG

FORWARD (V185D) - gt aag tat ggt acc GAC act gat ttg aaa atc

REVERSE (V185D) - GAT TTT CAA ATC AGT GTC GGT ACC ATA CTT AC

FORWARD (L205S) - ct aga ggg ttc ggt ttc TCT tct ttt gaa aaa cct tc

REVERSE (L205S) - GAA GGT TTT TCA AAA GAA GAG AAA CCG AAC CCT CTA G

FORWARD (V217E)- gt gtt gat gaa gtg GAA aag aca caa c

REVERSE (V217E) - GTT GTG TCT TTT CCA CTT CAT CAA CAC

FORWARD (I222S) - gtg gta aag aca caa cat TCT ctc gat ggt aaa g

REVERSE (I222S) - CTT TAC CAT CGA GAC TAT GTT GTG TCT TTA CCA C

FORWARD (I228N) - ctc gat ggt aaa gtt AAT gac cca aaa aga gc

REVERSE (I228N) - AGC TCT TTT TGG GTC ATT AAC TTT ACC ATC GAG

FORWARD (I234N) - cca aaa aga gct AAT cca aga gac gag c

REVERSE (I234N) - GCT CGT CTC TTG GAT TAG CTC TTT TTG G

FORWARD (V247D) - c ggt aaa atc ttt GAT ggt ggt att ggt c

REVERSE (V247D) - GAC CAA TAC CAC CAT CAA AGA TTT TAC CG

FORWARD (I269S) - ag tgg ggt acg TCT atc gat gcg caa c

REVERSE (I269S) - GTT GCG CAT CGA TAG ACG TAC CCC ACT

FORWARD (L274Q) - c gat gcg caa CAG atg tta gat aag g

REVERSE (L274Q) - CCT TAT CTA ACA TCT GTT GCG CAT CG

FORWARD (V300D) - c gcc gtt gac aga GAT tgt cag aat aaa ttc

REVERSE (V300D) - GAA TTT ATT CTG ACA ATC TCT GTC AAC GGC G

FORWARD (I313N) - ttc aaa gat cgc aag AAT gaa atc aag aga gct gag cc

REVERSE (I313N) - GGC TCA GCT CTC TTG ATT TCA TTC TTG CGA TCT TTG AA

Pull Down Experiments. GST pull-down experiments were executed using various GST-Hrp1 constructs and the His_6–Rna14–Rna15 complex. A 150-μL lysate of Hrp1 (RBD or full length) was treated with a 305-μL lysate of Rna14–Rna15 complex either in the presence or absence of RNA in the binding buffer (230 mM NaCl/50 mM Tris, pH 8.0/3% Glycerol/5 mM β-mercaptoethanol, BMe). The binding reaction was added to 100 μL of GST resin (Qiagen) and mixed at 4 °C for 4–5 h before resins were washed six to eight times with binding buffer. The proteins were eluted with buffer containing 10 mM reduced Glutathione and loaded onto 12% SDS/PAGE gels, transferred to Immune-Blot polyvinylidene fluoride membranes (Bio-Rad). These membranes were either probed with mouse monoclonal IgG_1 to GST (Santa Cruz Biotechnology) or with mouse monoclonal IgG_1 to His_6 (Santa Cruz Biotechnology) followed by goat anti-mouse IgG as a secondary antibody (Santa Cruz Biotechnology). The membrane was soaked and dried with LumiGLO peroxidase chemiluminescent substrate kit (www.kpl.com) for 3–4 min before detecting the protein using mini-medical automatic film processors (AFP Imaging).

NMR Experiments. All NMR experiments were recorded on Bruker Avance 500 or 600 MHz spectrometers, equipped with triple resonance cryoprobes. Sample temperature was kept at 298 K and buffer containing 100 mM NaCl, 20mM Tris (pH 8.0) and 2 mM DTT was used, unless otherwise stated. The ^1H-^{13}C methyl-TROSY experiments (1–3) for Hrp1 mutants were performed on Bruker Avance 600 spectrometer at 298 K. Chemical shifts were referenced against 2,2-dimetyl-2-silapentane-5-sulfonic acid (DSS).

Haddock Modeling of CF I. We used the results of NMR chemical shift perturbation experiments, site-directed mutagenesis and

pull-down experiments to generate a structural model of the CF I complex including Hrp1, the pre-mRNA, and a dimer of Rna14-Rna15. We used the online multidock server of HADDOCK to model this complex of nearly 300 kDa molecular weight. The docking procedure was split into two steps. In step 1, we modeled all protein components using I-TASSER based on different PDB templates (http://zhanglab.ccmb.med.umich.edu/I-TASSER). The protein Rna14 was modeled based on templates 2OND, 2OOE, and 2L9B, whereas Rna15 was modeled using 2X1A, 2KM8, 2J8P, 1P1T, and 2L9B. The Hrp1 model was made using 2CJK and 2KM8 whereas the pre-mRNA was taken from the ternary complexes of the Hrp1 RRM, and Rna15 RRM with RNA (2CJK and 2KM8).

In step 2, all components were docked in multiple steps using our experimental results and the information available from previous work (4–10). In substep 1, we reduced the structural complexity by modeling only the Rna14–Rna15 complex based on the information available. Rna14 is reported to interact with Rna15 through its C-terminal region, amino acids 640–677, whereas Rna15 associates with Rna14 through residues 148–228. Furthermore, Rna14 forms a homo-dimer through residues 320–480 of one monomer to residues 516–585 of the other monomer. No passive residues were defined between Rna14 proteins but ambiguous interaction restraints (AIRs) were generated using the HADDOCK server. These restraints were combined in the multidocking routine of HADDOCK to generate a model of the Rna14–Rna15 dimer where residues 639–670 of Rna14 and residues 149–228 of Rna15 were allowed to move during the annealing. Apart from these AIRs, we selected the option of center of mass restraints with a value of 1.0 in docking the complex. This option enforces the contact between different protein molecules. During the rigid body docking (it0), a total of 1500 structures were calculated with 7 trials for rigid body minimization. A total of 280 structures were selected for semiflexible refinement with explicit water. RMSD for clustering was kept to its default value of 7.5 Å with a cluster size of 4.

In substep 2, the Rna14–Rna15 dimer was docked with Hrp1 and RNA using the multidock approach. We used both active and passive residues from Hrp1 to RNA and vice versa. Similarly, AIRs within Rna15 and the RNA were defined based on previous studies. AIRs from Hrp1 to Rna14 were defined based on the experimental results presented here, whereas residues 1–320 of Rna14 were defined as passive residues to Hrp1 and were allowed to move freely during the annealing protocol. A similar docking protocol was followed as in substep 1, except for the number of calculated structures in rigid body docking (it0). Here, a total of 300 structures were selected for semiflexible refinement with other default parameters. No further structure minimization was done for this complex of nearly 300 kDa molecular weight.

The CF I model can be better viewed using Fig. S7. In Fig. S7*B*, chemical shift perturbation results are mapped onto the Hrp1 protein, whereas Fig. S7 *C–E* depict the schematic representation of the CF I complex.

Yeast Strains, Media, and Plasmids. The following *Saccharomyces cerevisiae* strains used in this study were described: PSY818 (MATα hrp1::HIS3 ura3 ade2 ade8 his3 leu2 lys1 trp1||plus pRS316–HRP1), Y198 (*MATa, ura3-1, leu2-3, his3-11, trp1-1, ade2-1, hrp1::HIS3, YCp-hrp1-5*), and Y199 (*MATa, ura3-1, leu2-3, his3-11, trp1-1, ade2-1, hrp1::HIS3, YCp-hrp1-6*). Yeast strains SL61 and SL63 were created by transforming PSY818 with pMHY50 (*CEN, LEU2, Myc-his6-HRP1*) or pMH50-hrp1-I228N (*CEN, LEU2, Myc-his6- hrp1-I228N*), respectively. Yeast were grown and maintained on YPD (1% yeast extract, 2% peptone, 2% glucose) or on selective media as needed. Yeast strains were transformed using the lithium acetate method. Loss of URA3 plasmids from yeast cells was accomplished by plating on solid medium containing 5-FOA. These procedures are described in Burke et al. (11).

1. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445(7128):618–622.
2. Barnwal RP, Atreya HS, Chary KV (2008) Chemical shift based editing of CH3 groups in fractionally 13C-labelled proteins using GFT (3, 2)D CT-HCCH-COSY: Stereospecific assignments of CH3 groups of Val and Leu residues. *J Biomol NMR* 42(2):149–154.
3. Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* 418(6894):207–211.
4. Legrand P, Pinaud N, Minvielle-Sébastia L, Fribourg S (2007) The structure of the CstF-77 homodimer provides insights into CstF assembly. *Nucleic Acids Res* 35(13): 4515–4522.
5. Qu X, et al. (2007) The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3′-end processing. *J Biol Chem* 282(3): 2101–2115.
6. Carneiro T, et al. (2008) Inactivation of cleavage factor I components Rna14p and Rna15p induces sequestration of small nucleolar ribonucleoproteins at discrete sites in the nucleus. *Mol Biol Cell* 19(4):1499–1508.
7. Leeper TC, Qu X, Lu C, Moore C, Varani G (2010) Novel protein-protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. *J Mol Biol* 401(3):334–349.
8. Pancevac C, Goldstone DC, Ramos A, Taylor IA (2010) Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3′-end processing factors. *Nucleic Acids Res* 38(9):3119–3132.
9. Mackereth CD (2011) Chemical shift assignments of a minimal Rna14p/Rna15p heterodimer from the yeast cleavage factor IA complex. *Biomol NMR Assign* 5(1):93–95.
10. Moreno-Morcillo M, Minvielle-Sébastia L, Fribourg S, Mackereth CD (2011) Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex. *Structure* 19(4):534–545.
11. Burke D, Ambrose D, Strathern JN (2005) *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).

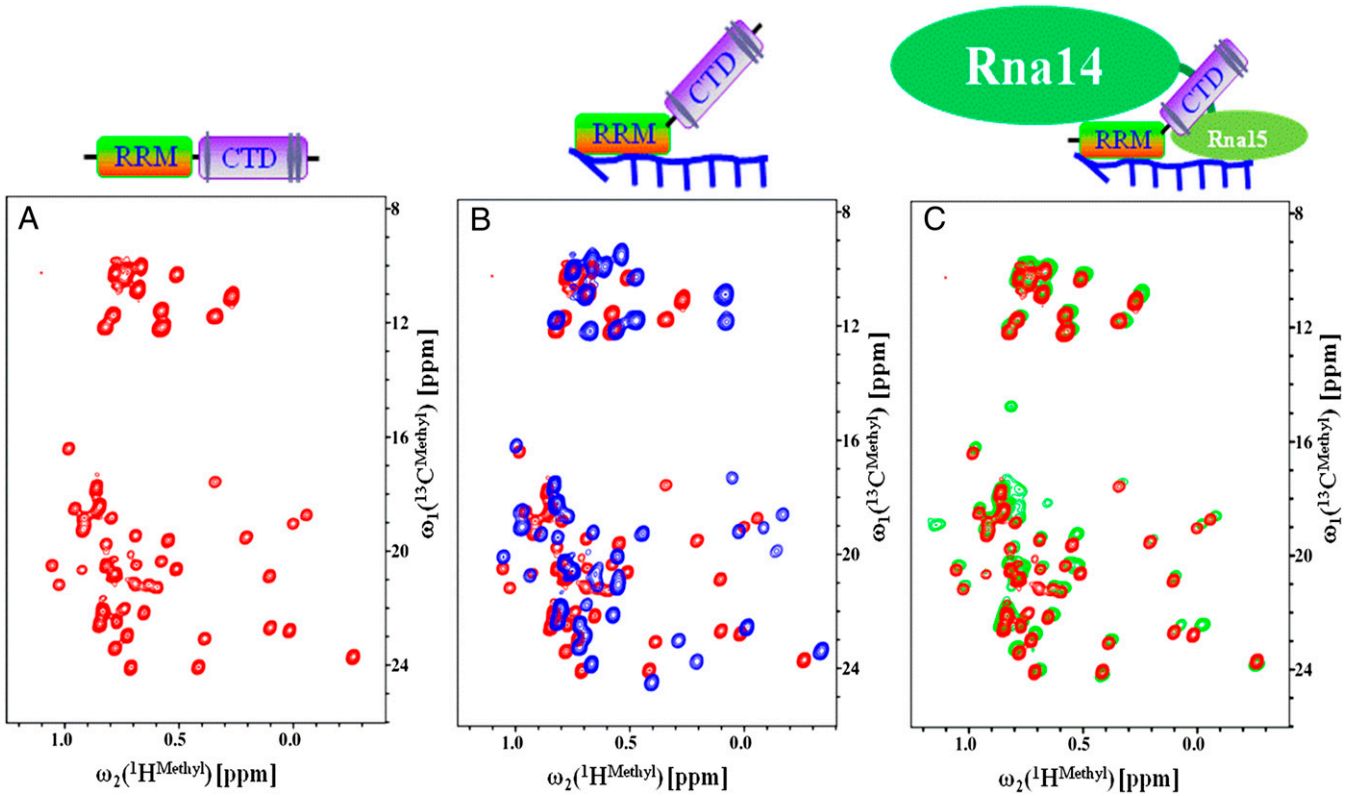

Fig. S1. [^{1}H-^{13}C]-HMQC spectra of labeled Hrp1 (RRMCTD construct). (*A*) ^{1}H-^{13}C HMQC spectra of {Ile (δ1)-[$^{13}CH_3$], Leu ($^{13}CH_3$, $^{12}CD_3$), Val ($^{13}CH_3$, $^{12}CD_3$)} U- [^{15}N, ^{12}C, ^{2}H] labeled sample of free Hrp1 protein, in red color. (*B* and *C*) Perturbations (*B*) are observed in the HMQC upon addition of the 14 nucleotides RNA followed by the addition of the ∼220-kDa complex of Rna14–Rna15 as a heterotetramer (*C*). In *B*, the red color represents the free RRMCTD protein, whereas the blue color represents peaks perturbed upon addition of RNA. In *C*, red and green colors represent free protein and RRMCTD with RNA and Rna14–Rna15, respectively. All figures were generated using CARA (http://cara.nmr.ch).

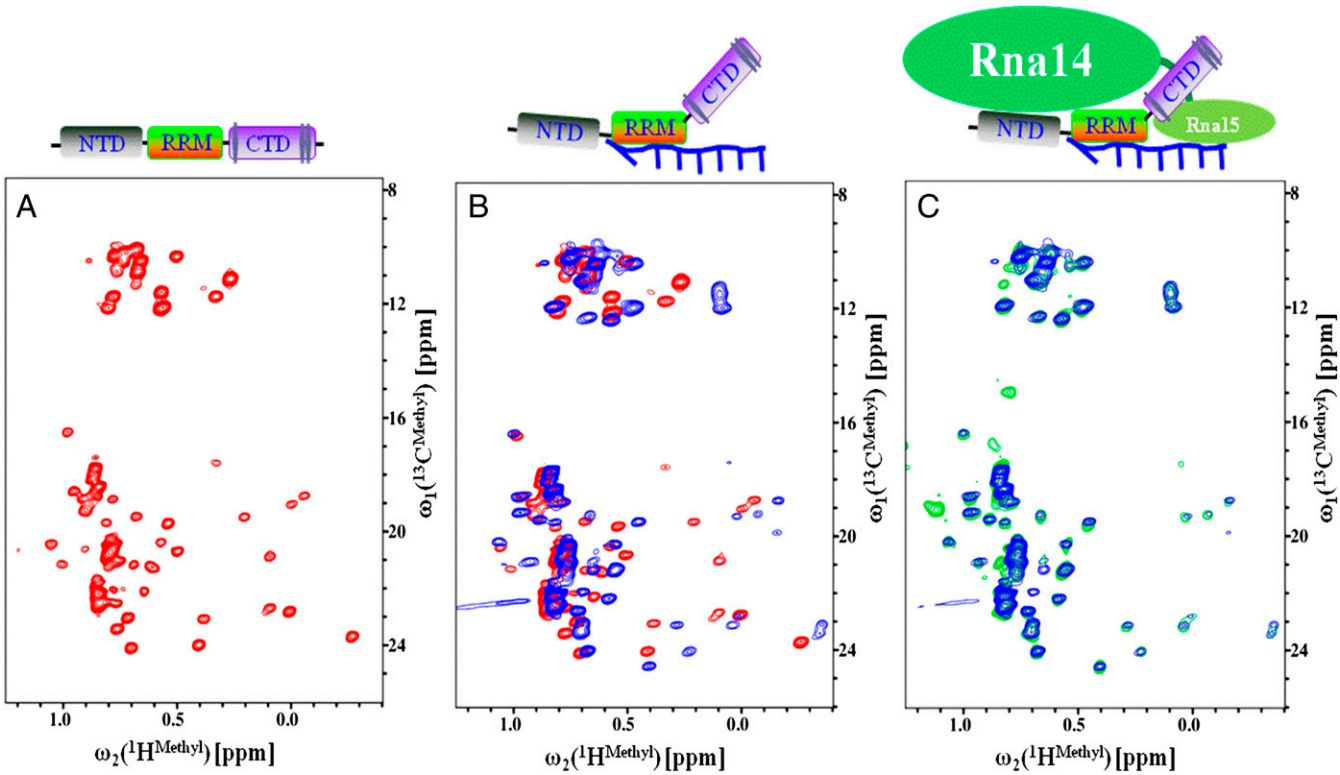

Fig. S2. Methyl [^{1}H-^{13}C]-TROSY spectra of full-length Hrp1. (*A*) ^{1}H-^{13}C HMQC spectra of {Ile (δ1)-[$^{13}CH_3$], Leu ($^{13}CH_3$, $^{12}CD_3$), Val ($^{13}CH_3$, $^{12}CD_3$)} U-[^{15}N, ^{12}C, ^{2}H] labeled sample of the free protein in red. (*B* and *C*) Perturbations (*B*) are observed in the HMQC spectrum upon addition of the 14-nt RNA, followed by the addition of the ~220-kDa complex of Rna14–Rna15 (*C*). In *B*, the red color represents the free protein, whereas the blue color represents Hrp1 bound to RNA. In *C*, blue and green colors represent Hrp1 bound to RNA and Hrp1 with RNA and Rna14–Rna15, respectively. All figures were generated using CARA (http://cara.nmr.ch).

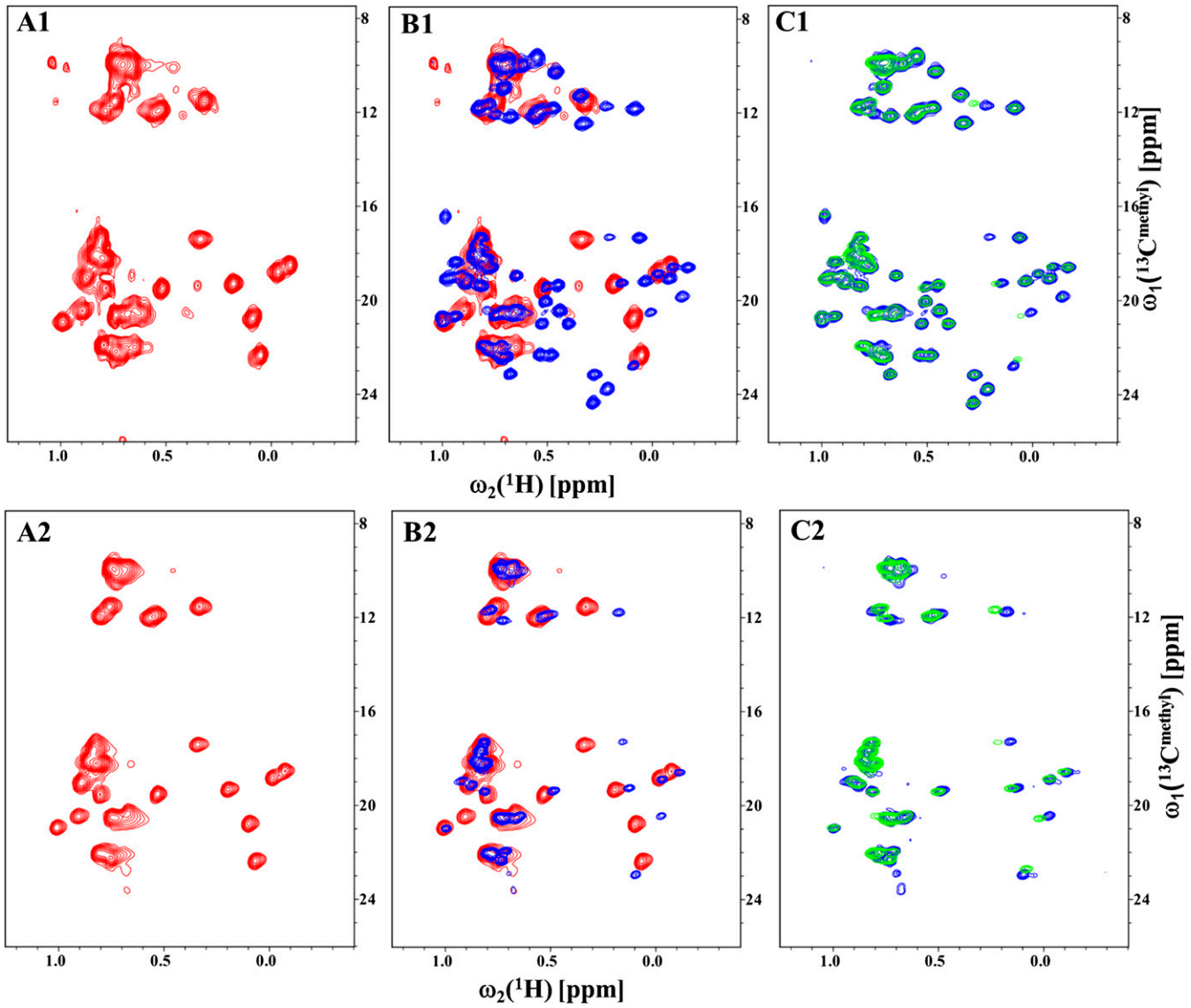

Fig. S3. [^{1}H-^{13}C]-methyl chemical shift perturbation experiments for Hrp1 mutant L205S and I228N. Here, *A1*, *B1* and *C1* are used for mutant I205S, whereas *A2*, *B2*, and *C2* are used for mutant I228N. (*A1*) The protein L205S in its free form. (*B1*) Superposition of the free protein (L205S, red) with RNA-bound protein (blue). (*C1*) Overlay of the protein–RNA complex (blue) compared with the same complex after addition of the Rna14–Rna15 complex (green). Chemical shift changes in the protein upon addition of RNA are clearly visible, but there are no further changes upon addition of Rna14–Rna15. Thus, the single mutation L250S essentially abolishes the interaction between Hrp1 and Rna14–Rna15. (*A2*) The protein I228N in its free form. (*B2*) Superposition of the free protein (I228N, red) with RNA-bound protein (blue). (*C2*) Overlay of the protein–RNA complex (blue) compared with the same complex after addition of the Rna14–Rna15 complex (green). All images were generated using Topspin 3.0 and CARA.

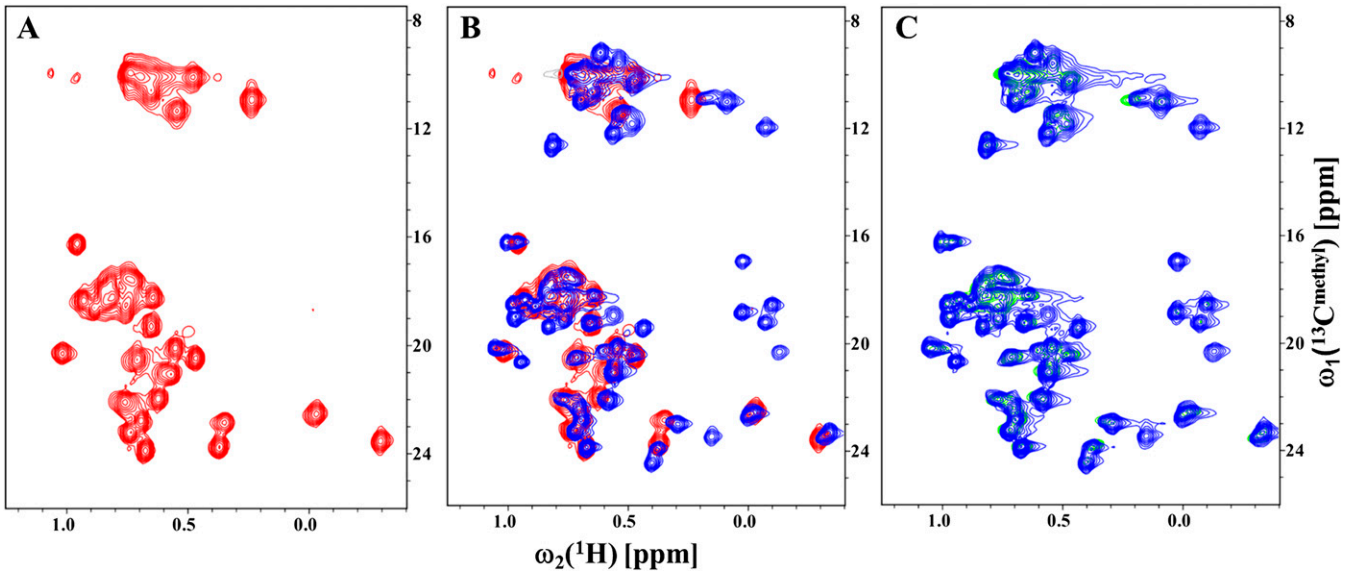

Fig. S4. [^{1}H-^{13}C]-HMQC experiments for Hrp1 mutant I313N. (*A*) The protein in its free form. (*B*) Superposition of the free protein (red) with RNA-bound protein (blue). (*C*) Overlay of the protein–RNA complex (blue) compared with the same complex after addition of the Rna14–Rna15 complex (green). All images were generated using Topspin 3.0 and CARA.

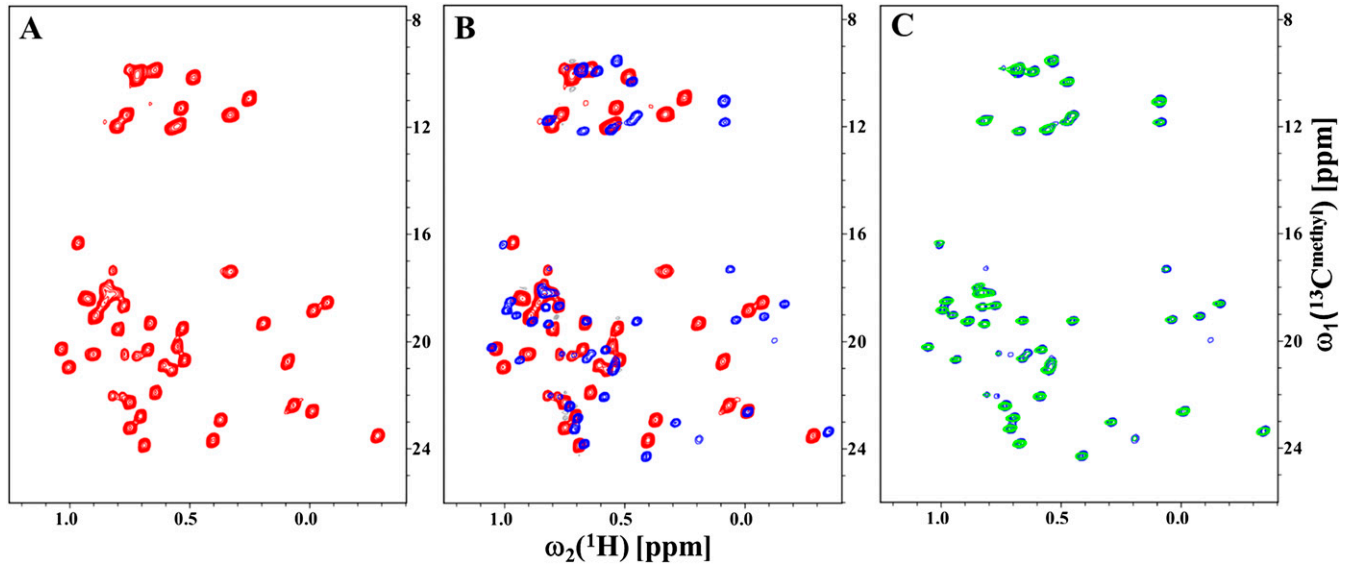

Fig. S5. [^{1}H-^{13}C]-Methyl TROSY experiments for Hrp1 mutant I222S. (*A*) The protein in its free form. (*B*) Superposition of the free protein (red) with RNA--bound protein (blue). (*C*) Overlay of the protein–RNA complex (blue) compared with the same complex after addition of the Rna14–Rna15 complex (green). All images were generated using Topspin 3.0 and CARA.

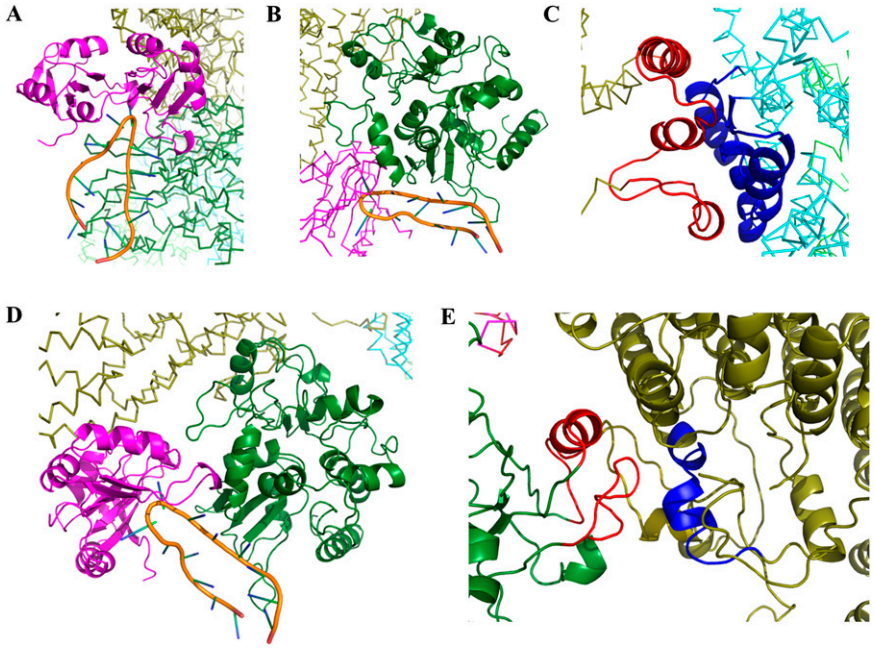

Fig. S6. Protein–protein interactions mapped on the model of the CF I complex. (*A*) The interaction of Hrp1 with RNA; Hrp1 is colored in magenta, whereas RNA is shown as a single-stranded ribbon. (*B*) The interaction of Rna15 with RNA; the protein is shown in dark green color, whereas the RNA is shown as an orange ribbon. (*C*) The Rna14-Rna14 dimer interface is shown in red (residues Asn536-Glu583 of one monomer) and blue (residues Ser365-Asp395 of the other Rna14 monomer). (*D*) Highlights of Hrp1, RNA, and Rna15 in the complete model of the CF I complex. The orientation of the RNA and Rna15 has been changed by an angle of 75° from the NMR structure 2KM8. (*E*) Interface between Rna14 and Rna15; interacting residues from Rna14 are shown in blue, whereas interacting residues from Rna15 are shown in red.

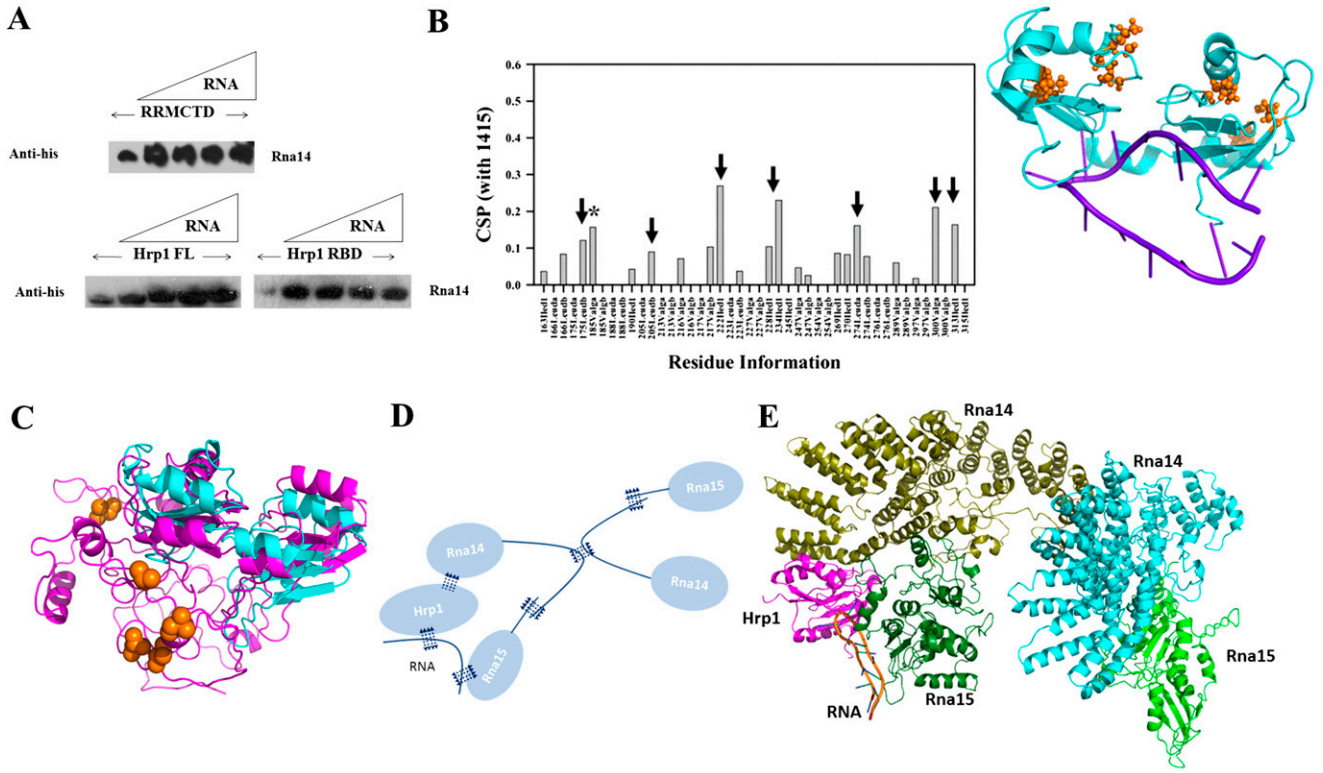

Fig. S7. (*A*) GST Pull-down of Hrp1 (RBDCTD), Hrp1 (full length; FL), and Hrp1 (RBD) with His-tagged Rna14–Rna15. The first lane in each gel is Hrp1 with Rna14–Rna15 in absence of RNA, whereas following lanes are with increasing amount of RNA concentrations. (*B*) Chemical shift perturbation for Hrp1 upon addition of Rna14–Rna15. In the graph, methyl signals corresponding to L175δ2, L205δ2, I228δ1, L274δ1, V300γ2, and I313δ1 show the larger perturbation (shown with black arrow) along with V185γ1 (indicated with asterisk) on binding to Rna14–Rna15. The side chain of V185 is not directly involved in the binding, as seen from our model and projected toward RNA. In the model, the highly perturbed methyl side chains are highlighted as orange colored spherical dots; structure 2KM8 is used to show the perturbations. (*C*) Overlay of Hrp1–RBD and RBDCTD. Here, RBD is colored in cyan, whereas RBDCTD is shown in magenta. The extra methyl containing residues of CTD are shown with orange sphere. (*D* and *E*) Schematic diagram of the CF I complex. (*D*) All protein factors including Hrp1, Rna14, and Rna15 are shown as simple cartoons; the RNA is shown as curved line. The interactions between the partners are shown by arrows. In the case of Hrp1, only the Hrp1->Rna14 interface is defined. Here, all protein factors except Hrp1 and RNA are modeled using I-TASSER. (*E*) Complete model of CF I generated using HADDOCK.

Table S1. Mutagenesis information used under this study

Type of mutation	Residue modified
Single mutation (His_6- and GST- versions)	L166Q, L175S, V185D, L205S, V217E, I222S, I228N, I234N, V247D, I269S, L274Q, V300D, I313N
Multiple mutation (His_6- only)	(a) L166Q+L205S+I313N, (b) L175S+I222S+I228N, (c) V185D+L234N +L274Q, (d) V300D+I269S+V217E+V247D, combination of (a)+(b), combination of (a)+(c), combination of (a)+(d)

Table S2. Residues used to generate the TBL file

Interface partner	Residue	Residue number involved for the interface partner
Rna14 to Rna15	Active	640,643,644,653,655
	Passive	639,642,645,654,656,657,658,659,663,668
Rna15 to Rna14	Active	149,151,152,199,200,228
	Passive	150,153,154,161,165,182,189,196,203,204,227
Rna15 to RNA	Active	27,29,31,56,57,58,59,60,61,62,63,66,69,70,71,98,99,100,102
	Passive	28,30,64,67,68,103
Rna to Rna15	Active	11,12,13,14,15,16
	Passive	9,10
Rna to Hrp1	Active	3,4,5,6,7,8
	Passive	9,10
Hrp1 to Rna	Active	162,164,165,168,191,201,202,204,244,246,248,273,275,279,284,286,288
	Passive	160,161,163,167,169,189,192,199,227,229,231,232,235,236,245,249,271, 278,281,285,287,317,320
Hrp1 to Rna14	Active	175,205,228,274,300,313
	Passive	166,185, 217,222,234,247,269
Rna14 to Hrp1	Active	None
	Passive	1–316
Rna14 (1) to Rna14 (2)	Active	320–480
	Passive	Not defined
Rna14 (2) to Rna14 (1)	Active	516–585
	Passive	Not defined