

De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes

Kalli Kappel¹, Shiheng Liu^{2,3}, Kevin P. Larsen^{1,4}, Georgios Skiniotis^{4,5}, Elisabetta Viani Puglisi⁴, Joseph D. Puglisi⁴, Z. Hong Zhou^{2,3}, Rui Zhao⁶ and Rhiju Das^{1,7,8*}

Increasingly, cryo-electron microscopy (cryo-EM) is used to determine the structures of RNA–protein assemblies, but nearly all maps determined with this method have biologically important regions where the local resolution does not permit RNA coordinate tracing. To address these omissions, we present de novo ribonucleoprotein modeling in real space through assembly of fragments together with experimental density in Rosetta (DRRAFTER). We show that DRRAFTER recovers near-native models for a diverse benchmark set of RNA–protein complexes including the spliceosome, mitochondrial ribosome, and CRISPR–Cas9–sgRNA complexes; rigorous blind tests include yeast U1 snRNP and spliceosomal P complex maps. Additionally, to aid in model interpretation, we present a method for reliable in situ estimation of DRRAFTER model accuracy. Finally, we apply DRRAFTER to recently determined maps of telomerase, the HIV-1 reverse transcriptase initiation complex, and the packaged MS2 genome, demonstrating the acceleration of accurate model building in challenging cases.

Recent advances in cryo-EM have led to new structural insights into many biologically important ribonucleoprotein (RNP) assemblies, including the spliceosome, ribosome, telomerase, and CRISPR complexes^{1–4}. For the increasing number of these maps with regions of high-resolution density (<4.0 Å), it is possible to manually trace atomic coordinates to obtain full-atom models⁵. However, most high-resolution maps still contain regions of lower resolution in which manual coordinate tracing is not feasible^{6,7}. For these regions, as well as for the sizable number of maps determined at lower resolution, atomic coordinates are often obtained by fitting of known structures of smaller subcomponents into the density⁸. This procedure presents a particular challenge for RNA–protein assemblies, as it is typically difficult to experimentally determine the coordinates of RNA subcomponents in isolation. For this reason, RNA coordinates are frequently omitted from models of RNP complexes^{9–12}, which highlights the critical need for computational methods that can accurately build RNA coordinates de novo into density maps of RNP assemblies.

The majority of existing computational methods focus on protein model building and refinement^{13–16}. These methods, many of which are based on well-established structure prediction algorithms, are able to build proteins de novo into both high- and lower-resolution maps, but at best can handle the presence of predetermined RNA structures¹⁷. In principle, RNA structure prediction algorithms¹⁸ could be similarly adapted for modeling of RNA coordinates de novo into cryo-EM maps of RNPs, but these methods have not yet been expanded to model RNA–protein complexes. Tools capable of modeling RNA into density maps are therefore limited to automated coordinate tracing within high-resolution maps¹⁹ and

refinement of reasonable initial structures. Developed primarily for high-resolution crystallographic density maps, refinement tools such as ERRASER, PHENIX, RCrane, and RNABC can be used to improve the quality of RNA structures^{20–23}. Molecular dynamics flexible fitting refines reasonable starting structures, which are often previously determined structures of alternative conformational states, into density maps ranging from low- to high-resolution and has been successfully applied to large RNP assemblies such as the ribosome to generate accurate atomic models of different functional states²⁴. However, to our knowledge there are currently no tools that are capable of building RNA structures de novo into low-resolution density maps.

Here, we have developed a computational framework for de novo ribonucleoprotein modeling in real space through assembly of fragments together with experimental density in Rosetta (DRRAFTER). DRRAFTER automatically builds missing RNA coordinates into cryo-EM maps of RNPs through fragment-based folding and docking. Structures are assessed by low-resolution and full-atom Rosetta score functions, which evaluate both the energy of the conformations and agreement with the density map. We benchmarked DRRAFTER on pairs of high-resolution (≤ 3.7 Å) and lower-resolution density maps for ten small RNA–protein complexes, the mitochondrial ribosome (mitoribosome), spliceosomal U4/U6.U5 tri-snRNP, and CRISPR–Cas9–sgRNA complexes, and performed additional blind tests on maps of the yeast U1 snRNP and spliceosomal P complex. These tests show that the accuracy of DRRAFTER models is comparable to that of models built by individual fitting of subcomponent crystal structures, and that DRRAFTER model accuracy can be reliably estimated in silico. Additionally, application

¹Biophysics Program, Stanford University, Stanford, CA, USA. ²Electron Imaging Center for Nanomachines, California NanoSystems Institute, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ³Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA, USA. ⁴Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ⁵Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ⁶Department of Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO, USA. ⁷Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. ⁸Department of Physics, Stanford University, Stanford, CA, USA. *e-mail: rhiju@stanford.edu

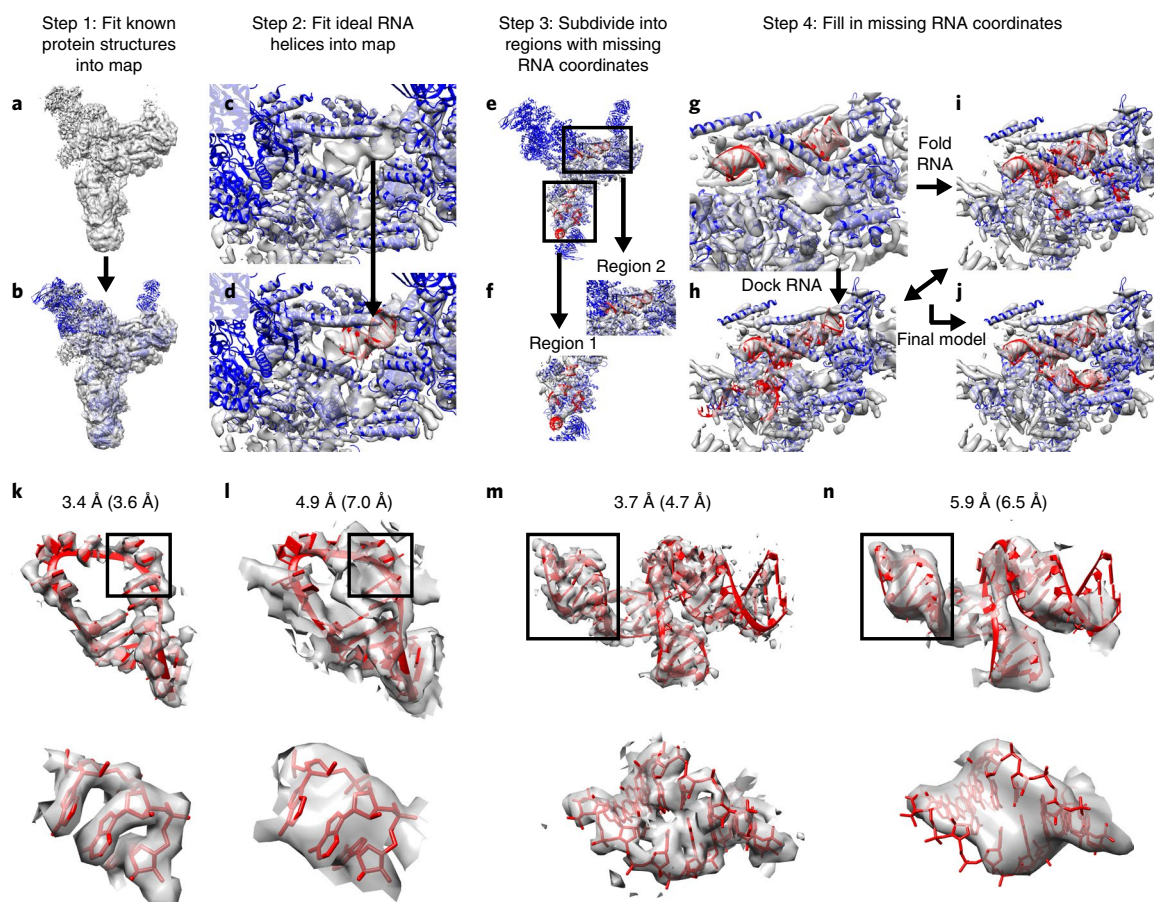


Fig. 1 | The DRRAFTER framework. **a–j**, Overview of the DRRAFTER pipeline. **a**, The starting cryo-EM density map (here the 5.9 Å spliceosomal tri-snRNP map⁹, gray). **b**, Individual protein structures (blue) are first fit into the density (here using Chimera). **c,d**, Ideal RNA helices are then fit into the density map (red). **e**, Subregions around the RNA helices where RNA coordinates are missing are visually identified. **f**, For each subregion, surrounding proteins and RNA helices are extracted from the larger model. **g–j**, Each of these sub-structures is input into the DRRAFTER protocol in Rosetta (**g**), during which RNA coordinates are filled in through a Monte Carlo simulation involving (**h**) docking moves to optimize rigid body orientations within the density map and (**i**) RNA fragment insertions to fold the RNA (RNA coordinates colored red). Models are scored initially with a low-resolution RNA–protein energy function, which accounts for RNA–RNA and RNA–protein interactions, and finally by an all-atom potential, each supplemented with a score term that rewards agreement with the density map to produce (**j**) final models that fit into the density map. **k–n**, Examples of high- and lower-resolution cryo-EM density maps. The high-resolution mitribosome loop 1 coordinates (red) in (**k**) the 3.4 Å (3.6 Å local resolution) density map³⁸ and (**l**) the 4.9 Å (7.0 Å local resolution) density maps (gray)¹⁰. The high-resolution spliceosomal tri-snRNP U5 three-way junction coordinates (red) in the (**m**) 3.7 Å (4.7 Å local resolution)³⁵ and (**n**) the 5.9 Å (6.5 Å local resolution) density maps (gray)⁹. Bottom panels show zoomed-in views of the regions boxed in the top panels. Surrounding proteins and RNA are not shown for clarity.

of our method to the recently determined 8.9 and 8.0 Å resolution telomerase and HIV-1 reverse transcriptase initiation complex (RTIC) maps recovered models that agree within error with previously published manually built models while requiring significantly reduced human effort, demonstrating that DRRAFTER can be used to accelerate and reduce bias in model building for lower-resolution maps of RNPs. Finally, we used DRRAFTER to build a full-atom model of 1,508 resolved nucleotides of the packaged MS2 genome, which until now had not been possible.

Results

DRRAFTER overview. An overview of the DRRAFTER framework is shown in Fig. 1a–j. Briefly, known structures of protein components as well as RNA helices should first be individually fit into a density map (Fig. 1a–d). This step is manual but rapid. For map subregions with missing RNA coordinates (Fig. 1e,f), full-atom models based on a user-supplied RNA secondary structure are automatically constructed within Rosetta through fragment-based RNA folding and docking (Fig. 1g–j). During this stage, models are scored initially

with the Rosetta low-resolution RNA–protein potential and finally with a full-atom energy function. Both energy functions account for RNA–RNA and RNA–protein interactions and are also supplemented with a score term that monitors agreement with the density map. The ten best-scoring models are then refined with the PHENIX-ERRASER pipeline to produce the final structures²⁰.

Benchmarking DRRAFTER performance. To evaluate the accuracy of the method, we benchmarked DRRAFTER on RNA–protein systems with pairs of density maps at high (≤ 3.7 Å) and lower resolution (4.5–7 Å overall; 5.0–9.8 Å local resolution). Examples of the high- and lower-resolution density maps are shown in Fig. 1k–n. In the highest-resolution maps (3.6 Å local resolution; Fig. 1k), individual RNA bases, base pairs, and phosphates can easily be identified. At intermediate resolutions (4–5 Å; Fig. 1m), these features are more difficult to visually identify. In lower-resolution maps (~6–12 Å; Fig. 1l,n), RNA helices can be seen clearly, but the base-pairing register is ambiguous and non-helical regions are difficult to discern.

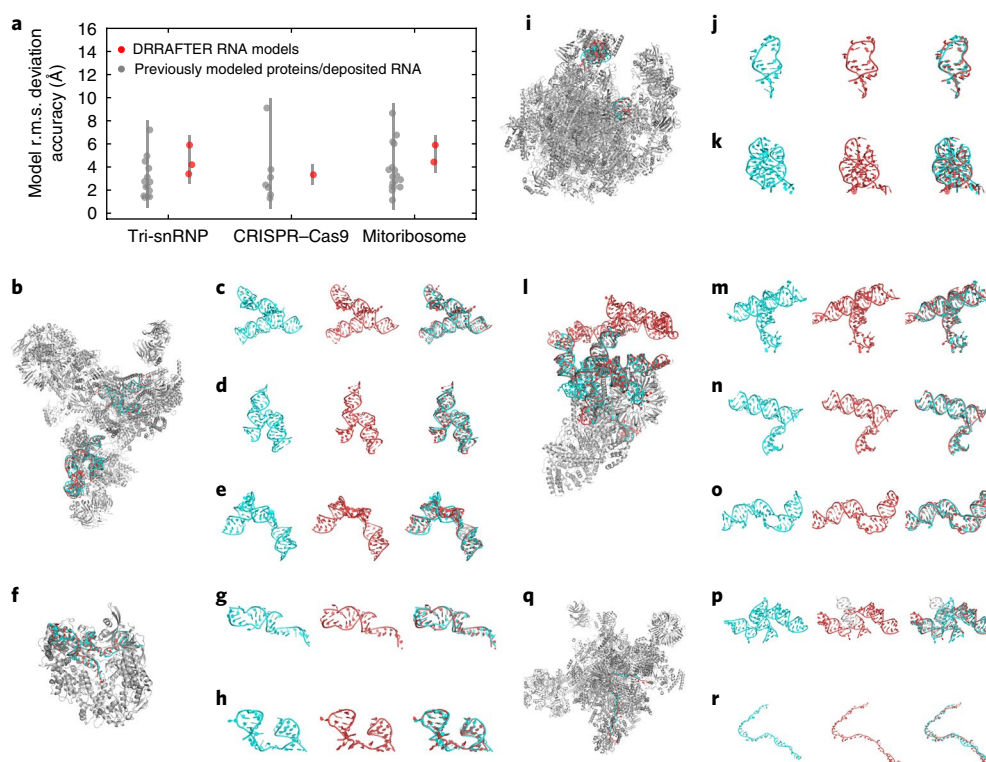


Fig. 2 | DRRAFTER recovers near-native models over a diverse benchmark set and two blind test cases. **a**, R.m.s. deviation values of DRRAFTER models (red; each region modeled is plotted as a separate point) and previously modeled low-resolution protein and RNA coordinates (gray; each protein or region of RNA is plotted as a separate point) compared with later determined high-resolution coordinates. **b, f, i, l, q**, DRRAFTER models built into low-resolution maps (RNA colored red) overlaid with high-resolution coordinates (RNA colored cyan; protein colored silver; PDB IDs listed in parentheses) for **(b)** the spliceosomal tri-snRNP (5GAN), **(f)** CRISPR-Cas9-sgRNA complex (4ZT0), **(i)** mitoribosome (4V19 and 4V1A), **(l)** yeast U1 snRNP (5UZ5), and **(q)** yeast spliceosomal P complex (6BK8). **c-e, g, h, j, k, m-p, r**, High-resolution RNA coordinates (left, cyan), RNA coordinates from DRRAFTER models built into low-resolution maps (middle, red), and high-resolution coordinates and DRRAFTER models overlaid (right, high-resolution coordinates colored cyan, DRRAFTER models colored red) for the spliceosomal tri-snRNP, **(c)** U4/U6 three-way junction, **(d)** U5 three-way junction, **(e)** U5 internal loop II; CRISPR-Cas9-sgRNA complex, **(g)** sgRNA residues 11-30 and 57-68, **(h)** sgRNA residues 69-99, mitoribosome, **(j)** loop 1, **(k)** loop 2, yeast U1 snRNP (blind), **(m)** core four-way junction, **(n)** yeast three-way junction, **(o)** SL2-2, **(p)** yeast-specific four-way junction (DRRAFTER model of SL3-2, SL3-3, and SL3-5 colored red; DRRAFTER model of SL3-4 colored white in order to show one of the unusual strong departures from the high-resolution structure (see text)), yeast spliceosomal P complex, and **(r)** ligated exon.

The benchmark set included ten small RNA-protein crystal structures for which we simulated density maps at both 5.0 and 7.0 Å resolution (Supplementary Fig. 1)²⁵⁻³⁴ and three large RNP machines with published experimental density maps containing regions where RNA coordinates had not previously been modeled: the *Saccharomyces cerevisiae* spliceosomal tri-snRNP³⁵, the *Streptococcus pyogenes* CRISPR-Cas9-sgRNA complex^{36,37}, and the *Sus scrofa* mitoribosome^{10,38} (Fig. 2). These systems represent a diverse range of RNA and RNA-protein structures including complex RNA junctions and interactions between proteins and both single-stranded and highly structured RNAs.

To first establish the baseline target accuracy, we compared coordinates from the three lower-resolution experimental maps for the protein regions (for all three systems) and RNA regions (for the mitoribosome only) that were modeled into those maps to the later determined high-resolution coordinates. The root-mean-square deviation (r.m.s. deviation) ranged from 1.3 to 9.1 Å (Fig. 2a; see Methods). We then used DRRAFTER to build models of the ten small RNA-protein systems using the 5 and 7 Å simulated density maps, as well as six regions of the three large RNP machines using the lower-resolution experimental maps (local resolutions varied from 5.0 to 9.8 Å). Qualitatively, the DRRAFTER models closely recapitulate the overall folds of the high-resolution coordinates in all cases (Fig. 2b-k, Supplementary Fig. 1, and Supplementary Fig. 2a-g).

The r.m.s. deviation accuracy of DRRAFTER models ranges from 0.7 to 6.2 Å (best of ten models, median of ten models was similar; see Table 1 and Supplementary Table 1), which is within our targeted baseline accuracy range (Fig. 2a). Additionally, the real-space correlation coefficients of the RNA models are comparable to the correlation of the high-resolution coordinates to the lower-resolution map (Supplementary Table 2 and Supplementary Fig. 3).

To test the applicability of DRRAFTER to higher-resolution density maps, we also used DRRAFTER to build models into the high-resolution experimental density maps of each benchmark RNP or, for the ten small crystal structures, simulated maps at 3 Å resolution (Supplementary Fig. 1). While the reported resolutions for the experimental maps were all better than 3.7 Å, the local resolution varied from 2.9 to 5.7 Å (Supplementary Table 3). Compared to the published manually generated coordinates, the r.m.s. deviation values of the DRRAFTER models ranged from 0.3 to 3.9 Å (Supplementary Table 3), with the worst r.m.s. deviation for the spliceosomal tri-snRNP U5 internal loop II (3.9 Å), which also had the lowest resolution density (5.7 Å). These results suggest that while the DRRAFTER framework is primarily intended for cases where manual coordinate tracing is not feasible, it can be used to automatically build coordinates into high-resolution maps, though in some cases final manual adjustments may be necessary and careful visual inspection is always recommended.

Table 1 | Summary of r.m.s. deviation accuracies

Systems	Number of test cases	Reported map resolution range (Å)	Mean reported map resolution (Å)	Local map resolution range (Å)	Mean local map resolution (Å)	Mean of the best r.m.s. deviations of the ten top-scoring models (Å)	Mean convergence estimate (Å)
Small RNPs, lower-resolution simulated maps ¹	20	5.0–7.0	6.0	5.0–7.0	6.0	2.4	3.5
Small RNPs, higher-resolution simulated maps ¹	10	3.0	3.0	3.0	3.0	1.4	2.5
Large RNPs, lower-resolution experimental maps ²	19	4.5–10.5	8.4	5.0–12.4	9.5	4.9	7.5
Large RNPs, higher-resolution experimental maps ³	6	2.9–3.7	3.5	2.9–5.7	4.1	2.4	3.4
Blind tests, experimental maps ⁴	6	5.4–6.0	5.9	6.6–7.3	6.7	3.6	5.9

¹Small RNPs: *E. coli* L25-5S rRNA (1DFU), Sex-lethal RRM (1B7F), Ribotoxin restrictocin—SRL analog (1JBS), SmpB-tmRNA complex (1P6V), HutP antitermination complex (1WPU), mRNA-binding domain of SelB elongation factor (1WSU), NusA transcriptional regulator (2ASB), methyltransferase RumA in complex with rRNA (2BH2), PP7 coat protein and viral RNA (2QUX), Puf4 bound to 3' UTR of target transcript (3BX2). Complete data are provided in Supplementary Table 1 and Supplementary Table 3. ²Large RNPs built into lower-resolution experimental maps: U4/U6.U5 tri-snRNP U4/U6 3WJ, U5 3WJ, U5 IL II; mitochondrial ribosome loop 1, loop 2; CRISPR-Cas9; MS2 packaged genome S1 + S2, S3, S4, S5 + S6, S7, S8, S9-1, S9-2, S10, S12, S15 + S16; yeast U1 snRNP yeast-specific 4WJ, SL2-2. Complete data are provided in Supplementary Table 1. ³Large RNPs built into higher-resolution experimental maps: U4/U6.U5 tri-snRNP U4/U6 3WJ, U5 3WJ, U5 IL II; mitochondrial ribosome loop 1, loop 2; CRISPR-Cas9. Complete data are provided in Supplementary Table 3. ⁴Blind tests: yeast U1 snRNP core 4WJ, core 4WJ only, yeast 3WJ, yeast-specific 4WJ, SL2-2; yeast P complex ligated exon. Complete data are provided in Supplementary Table 1.

As an additional test, we compared the accuracy of DRRAFTER models to the accuracy of models manually built into lower-resolution maps. For most of the test cases in our benchmark set, RNA coordinates were not previously built into the lower-resolution maps. However, we were able to perform this test on the mitochondrial ribosome, for which manually built RNA coordinates were deposited for the lower-resolution (4.9 Å) map for several regions (where coordinates were not taken from the homologous *E. coli* ribosome structure). The accuracies of the DRRAFTER and deposited manually built models, determined by comparison to the higher-resolution coordinates (from the 3.4 Å map), were comparable (Supplementary Fig. 4). This result suggests that DRRAFTER is a comparable alternative to manual modeling, when it is possible, into lower-resolution maps.

Blind tests of DRRAFTER performance. As a rigorous challenge, K.K. and R.D. performed blind tests of the DRRAFTER pipeline on early stage 6.0 and 5.4 Å resolution maps of the yeast U1 snRNP and spliceosomal P complex, respectively, prior to the publication of higher-resolution maps with resolutions of 3.6 and 3.3 Å, respectively (kept hidden by S.L., H.Z., and R.Z.)^{39,40}. The yeast U1 snRNP modeling was carried out over a period of 3 d, during which we built DRRAFTER models of five subregions covering the majority of the 568-nucleotide U1 snRNA. A previously published structure of the core human U1 snRNP helped identify the location of the core four-way junction in the map, but because the human structure did not fit well in the density map and the yeast snRNA is significantly larger than the human U1 snRNA (568 versus 164 nucleotides), nearly the entire RNA was modeled de novo (Fig. 2l). Blind DRRAFTER models of the core four-way junction (LR1/LR2, SL1, SL2-1, SL3-1) (Fig. 2m and Supplementary Fig. 2h) and yeast-specific three-way junction regions (SL3-1, SL3-2, SL3-6) (Fig. 2n and Supplementary Fig. 2i) achieved r.m.s. deviation values of 3.1 and 2.4 Å, respectively, with residues within the four-way junction (residues 16–17, 45–46, 167–168, and 544–545) reaching 1.6 Å r.m.s. deviation accuracy. The best model of SL2-2 achieved r.m.s. deviation accuracy of 4.0 Å (Fig. 2o and Supplementary Fig. 2j), although we noted that models of this region suffered from a lack

of compute time (~450 models generated versus the target of 3,000 models). When later revisited with additional computational expenditure (~3,000 models generated), the r.m.s. deviation dropped to 2.5 Å. The best model of the yeast-specific four-way junction over SL3-2, SL3-3, and SL3-5 achieved r.m.s. deviation accuracy of 4.3 Å (Fig. 2p and Supplementary Fig. 2k). SL3-4 was excluded from the final r.m.s. deviation calculation because we were unable to build a model that fit into the density, as determined by visual inspection. After unblinding the high-resolution coordinates, we learned that the proposed secondary structure for this region, which was enforced during the DRRAFTER modeling, was incorrect. When this region was subsequently revisited with the corrected secondary structure, we were able to build models with SL3-4 in the density, and the r.m.s. deviation accuracy over the entire yeast-specific four-way junction improved slightly to 4.2 Å. Finally, we could not assess the accuracy of models that we built for the peripheral SL3-7 domain because coordinates were not built into the final high-resolution map, which only showed diffuse density for that region. We provide a complete all-atom model for the yeast U1 snRNP, including these peripheral regions, in Supplementary Data 1.

When modeling the yeast spliceosomal P complex, we discovered that the majority of the density could be modeled well by the previously published structure of the C* complex, the state immediately prior to P complex formation in the catalytic cycle of the spliceosome^{41,42}. We therefore focused our attention on the structure of the ligated exon, which is not yet present in the C* complex. This long single-stranded RNA region proved challenging to model as indicated by two measures. First, the density in this region was at 7.3 Å resolution, considerably poorer than the overall 5.4 Å resolution of the map. Second, our final pool of DRRAFTER models exhibited substantial structural heterogeneity (Supplementary Fig. 2l). Indeed, while our models cluster around the high-resolution coordinates, the r.m.s. deviation accuracy of our best model was 6.2 Å, poorer than for the majority of the test cases in our benchmark set (Fig. 2q,r).

Estimating DRRAFTER model accuracy. Inspired by the challenge of these blind tests, we sought to develop a method to estimate the accuracy of DRRAFTER models in silico. This would allow model

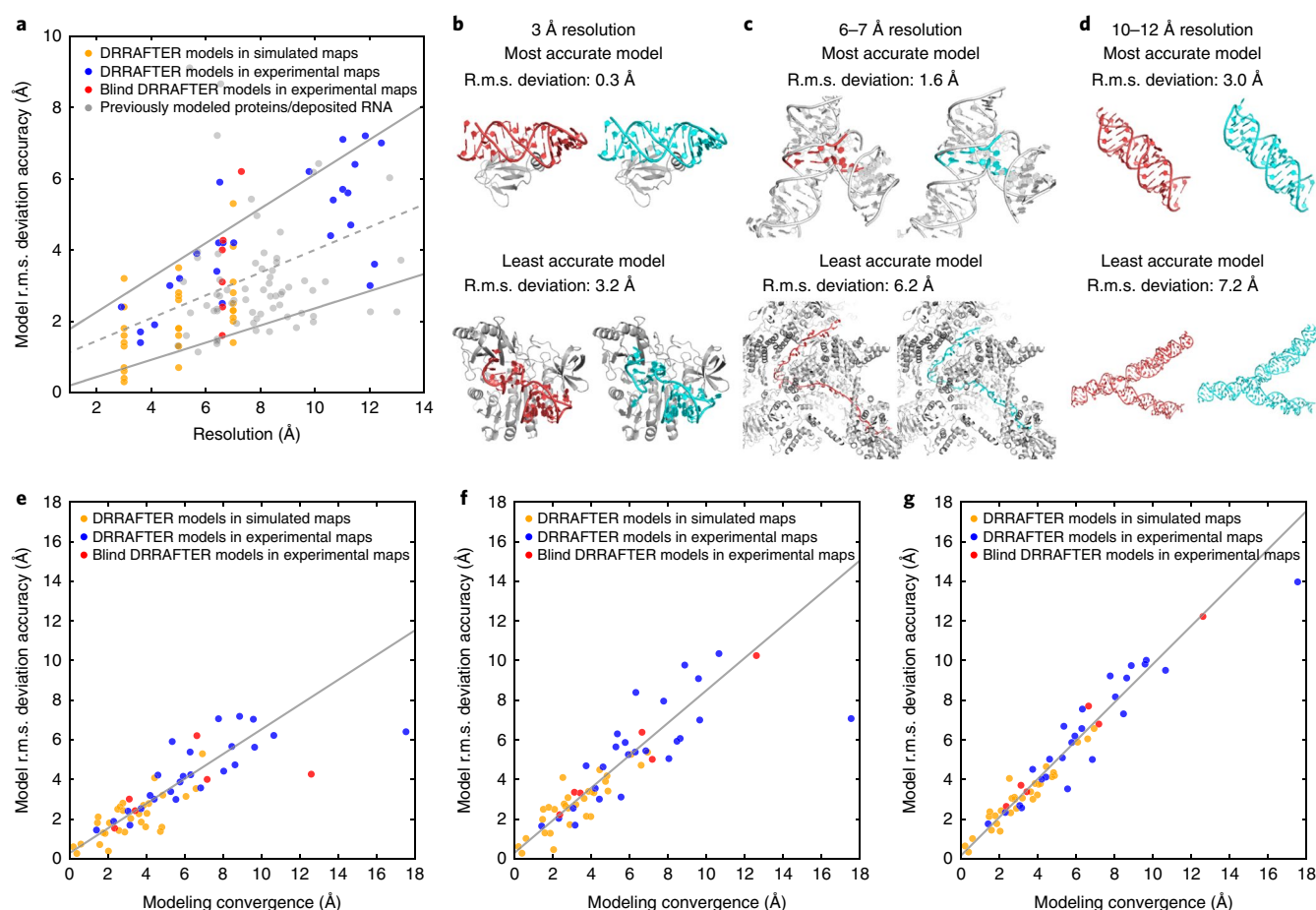


Fig. 3 | Estimating DRRAFTER model accuracy. **a**, R.m.s. deviation accuracy versus local map resolution (Supplementary Table 1 and Supplementary Table 3) for DRRAFTER models built into high- and low-resolution simulated (gold, $N=30$) and experimental maps (blue, $N=25$), blind DRRAFTER models built into low-resolution experimental maps (red, $N=6$), and previously modeled low-resolution protein and RNA coordinates (gray, $N=67$). The best-fit line (dashed gray) is given by $y=0.32x+0.81$ (total number of systems = 128). The best-fit upper and lower bound lines (solid gray) are given by $y=0.48x+1.29$ and $y=0.24x-0.04$, respectively (see Methods). **b-d**, Examples of the most accurate (top) and least accurate (bottom) DRRAFTER models for maps at **(b)** 3 Å, **(c)** 6–7 Å, and **(d)** 10–12 Å. For each panel, DRRAFTER models are shown on the left with the RNA colored red and the protein colored gray, and the high-resolution coordinates are shown on the right with the RNA colored cyan and the protein colored gray. **b**, Top, *E. coli* L25-5S rRNA. Bottom, methyltransferase RumA in complex with rRNA. **c**, Top, yeast U1 snRNP core four-way junction (surrounding RNA residues colored gray). Bottom, yeast spliceosomal P complex ligated exon. **d**, Top, MS2 packaged genome region S9-2. Bottom, region S7. **e-g**, R.m.s. deviation accuracy versus DRRAFTER modeling convergence for **(e)** the most accurate of the ten top-scoring DRRAFTER models (points for DRRAFTER models built into simulated density maps colored gold ($N=30$); points for DRRAFTER models built into experimental density maps colored blue ($N=25$); points for blind DRRAFTER models colored red ($N=6$); total number of systems = 61), **(f)** the centroid of the ten top-scoring DRRAFTER models (colors as in **e**), and **(g)** the mean r.m.s. deviation to native across the ten top-scoring DRRAFTER models (colors as in **e**). The best-fit lines (solid gray; excluding the two points with convergence >12 Å: MS2 S15 + S16 and blind yeast U1 snRNP yeast-specific four-way junction) are given by **(e)** $y=0.62x+0.28$, **(f)** $y=0.82x+0.30$, and **(g)** $y=0.97x+0.17$.

quality to be quantitatively determined in realistic modeling scenarios. We identified two metrics that are predictive of final model accuracy. First, the local resolution places approximate bounds on the final modeling accuracy (Fig. 3a), though there is still considerable variation in model accuracy across different test cases for maps with similar resolution (Fig. 3b–d). Regions of highly structured RNA tend to be predicted more accurately with DRRAFTER, while regions of long single-stranded RNA are often more challenging to model accurately (Fig. 3b–d). The correlation between resolution and model accuracy is significant (two-tailed $P=4\times 10^{-8}$ for Pearson's correlation coefficient, $N=128$, Supplementary Table 1) but weak ($R^2=0.21$), suggesting that there are additional factors that determine model accuracy. Second, we assessed the convergence of DRRAFTER models by calculating the average pairwise r.m.s. deviation over the ten best-scoring models (Supplementary

Tables 1 and 3 and Supplementary Fig. 2). This convergence estimate is correlated with the accuracy of the best of the top ten models (Fig. 3e and Supplementary Table 1, $R^2=0.67$, two-tailed $P=6\times 10^{-16}$, $N=61$; excluding models with convergence >12 Å: $R^2=0.78$, two-tailed $P=3\times 10^{-20}$, $N=59$), the centroid of the top ten models (Fig. 3f and Supplementary Table 1, $R^2=0.72$, two-tailed $P=4\times 10^{-18}$, $N=61$; excluding models with convergence >12 Å: $R^2=0.82$, two-tailed $P=2\times 10^{-22}$, $N=59$), and the mean accuracy of the top ten models (Fig. 3g and Supplementary Table 1, $R^2=0.93$, two-tailed $P=4\times 10^{-36}$, $N=61$; excluding models with convergence >12 Å: $R^2=0.92$, two-tailed $P=1\times 10^{-33}$, $N=59$). Based on these results, we suggest that prior to modeling, the local map resolution be used to place bounds on the expected modeling accuracy, and after modeling is completed, the convergence of the DRRAFTER models be used to reliably estimate modeling accuracy.

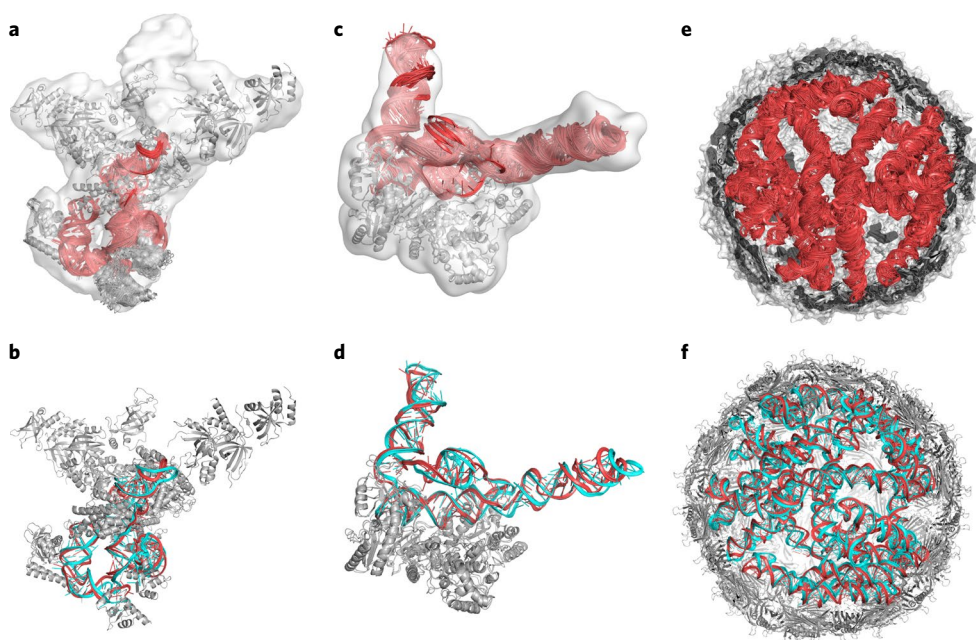


Fig. 4 | DRRAFTER can accelerate manual model building into low-resolution density maps. a,c,e, Overlay of ten best-scoring DRRAFTER models (RNA colored red, protein colored gray, density map colored transparent light gray) for **(a)** telomerase, **(c)** HIV-1 RTIC, and **(e)** the packaged MS2 genome (built into the 10.5 Å resolution map). Regions with more variability between models are estimated to be less accurate. **b,d**, Overlay of DRRAFTER models with previously built manual models for **(b)** telomerase (DRRAFTER model colored red for RNA and gray for protein, previously built manual model colored cyan for RNA and gray for protein⁴³) and **(d)** HIV-1 RTIC (coloring as in **b**⁴⁴). A single DRRAFTER model (centroid) of the ten best scoring is shown for clarity. **f**, Overlay of DRRAFTER models built into independently determined 3.6 Å (RNA colored cyan, protein colored gray, includes coordinates from PDB ID 5TC1⁴⁶) and 10.5 Å (RNA colored red, protein colored gray) resolution maps of the packaged MS2 genome.

Application to challenging targets. For RNP targets of exceptional biological value, researchers have committed extraordinary efforts to manually piece together RNA models within low-resolution maps of RNPs. In the few cases where this manual model building is actually feasible, it is extremely time-consuming and subject to considerable bias. We therefore wanted to test whether DRRAFTER could be used to accelerate model building and reduce human bias in these cases. We applied DRRAFTER to the recently determined 8.9 Å map of *Tetrahymena* telomerase and the 8.0 Å map of the HIV-1 RTIC, where models of the RNA had previously been built manually^{43,44}. The DRRAFTER models agree well with the published models with mean r.m.s. deviation values over the top ten models of 5.7 Å for HIV-1 RTIC and 7.6 Å for telomerase (6.6 Å excluding the poorly converged single stranded RNA residues 52–68) (Fig. 4a–d). Building these models with DRRAFTER required only a few hours of human effort, versus the days to weeks that are usually required for manual model building. Additionally, by using DRRAFTER to build these models we were able to calculate their expected accuracy. Using the convergence of the DRRAFTER models, we estimate that the best of the ten DRRAFTER models have r.m.s. deviation accuracies to the ‘true’ coordinates of 3.5 Å for telomerase (convergence = 5.2 Å), and 4.2 Å (convergence = 6.3 Å) for the HIV-1 RTIC RNA. After this modeling was performed, a higher-resolution (4.8 Å) structure of *Tetrahymena* telomerase with telomeric DNA became available⁴⁵. Comparison with DRRAFTER models confirmed that the accuracy of the de novo modeled regions was close to the predicted value and that region by region, the accuracies of the DRRAFTER models are similar to the accuracies of the previously published manually built telomerase model, again confirming that DRRAFTER provides a comparable alternative to time-consuming manual model building (Supplementary Table 4).

Finally, we applied DRRAFTER to the recently determined 3.6 Å map of the packaged MS2 genome⁴⁶. Despite the high resolution

overall, the local resolution in the region of the packaged RNA was not high enough for a full-atom model to be built, with the exception of several protein-bound RNA hairpins. With DRRAFTER, we were able to build a model of 1,508 nucleotides (Fig. 4e,f and Supplementary Data 2) with estimated accuracies of 2.4–6.0 Å (convergence = 3.8–9.7 Å). As a final test of DRRAFTER accuracy, we additionally applied the framework to the previously published 10.5 Å map of the packaged MS2 genome and compared the resulting models to those based on the 3.6 Å map⁴⁷. The r.m.s. deviation values are between 3.0 and 7.2 Å; qualitatively, the models agree very well, and many of the differences in the models reflect underlying differences in the 3.6 and 10.5 Å maps (Supplementary Fig. 5).

Discussion

For systems representing all major classes of RNPs with maps of a wide range of resolutions, DRRAFTER was able to successfully build near-native coordinates in regions where manual coordinate tracing was difficult or intractable. Over a benchmark set of both simulated and experimental maps, DRRAFTER models consistently recovered native RNA folds. Separate blind tests of the method demonstrate that the DRRAFTER framework can be successfully applied in realistic modeling settings. Additionally, even in cases where manual modeling into low-resolution maps may be feasible, it is slow, painstaking, and can suffer from errors; DRRAFTER can be used to accelerate and reduce bias from the process. DRRAFTER has the added advantage over manual modeling of providing a way to estimate model accuracy, which should aid in interpretation of final models. Overall, we expect that DRRAFTER will be widely useful for building RNA coordinates into cryo-EM maps.

The tests presented here suggest three main areas for future improvement of the DRRAFTER pipeline. First, DRRAFTER relies on having accurate RNA secondary structure information. In some cases, the current DRRAFTER pipeline may be able to distinguish

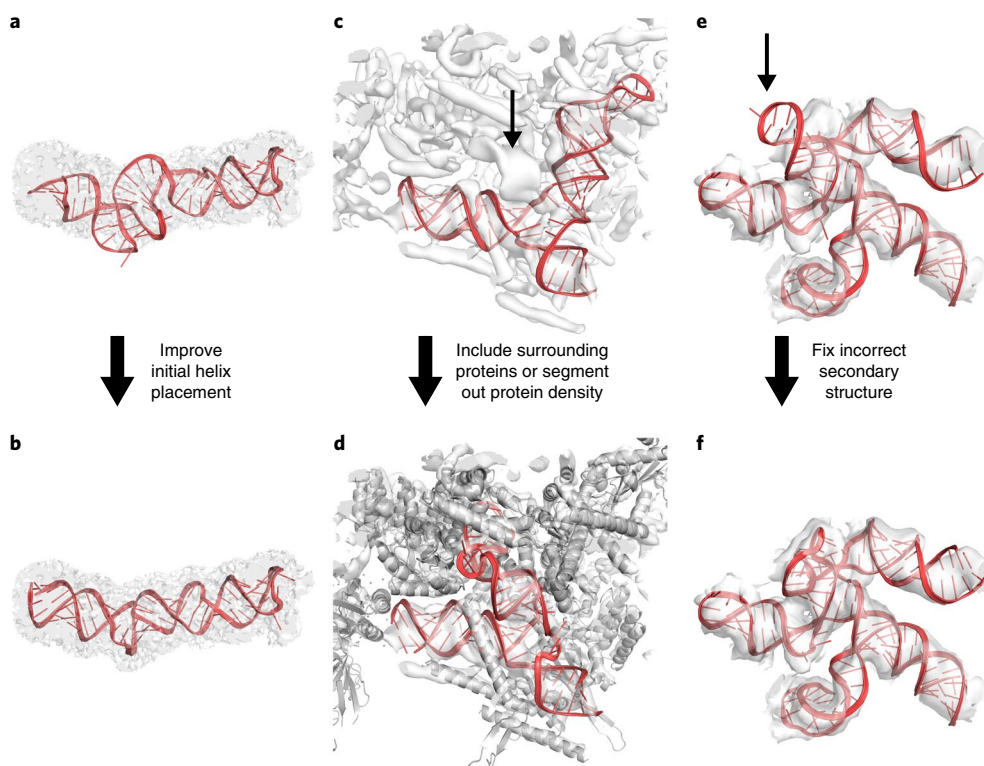


Fig. 5 | Typical mistakes that may occur during DRRAFTER modeling and possible solutions. **a**, Poor initial helix placement can lead to distorted final models, as shown here for residues 153–227 of the packaged MS2 genome (RNA colored red, density map colored transparent gray). **b**, This can be fixed either by improving the initial helix placement or by skipping the helix placement step and letting DRRAFTER determine the helix placement de novo. Here, this improved model was built by omitting the initial helix placements. **c**, When proteins are not included during DRRAFTER modeling, RNA models may be built into protein density as shown here for the spliceosomal tri-snRNP U4/U6 three-way junction (RNA colored red). The actual density for the RNA is indicated with the black arrow. **d**, This can be fixed either by including the surrounding proteins during the DRRAFTER modeling, as shown here (proteins colored gray), or by segmenting the protein density out of the map before modeling. **e**, Visual inspection can identify models that do not fit well in the density map, as shown here for SL3-4 of the yeast spliceosomal U1 snRNP (black arrow). This can be caused by inadequate sampling, in which case building more models and/or increasing the number of cycles used to build each model should solve this problem. Alternatively, some of the modeling assumptions, such as the RNA secondary structure, or fixed positions of surrounding RNA or protein residues may be incorrect. **f**, In this case, the secondary structure assumed as part of the initial modeling was incorrect. When the secondary structure was corrected, we were able to build DRRAFTER models that fit in the density map.

between different secondary structure possibilities; for the U1 snRNP yeast-specific four-way junction test case, models with the incorrect secondary structure were unable to fit into the density, while later models with the corrected secondary structure fit well. However, this strategy is unlikely to be feasible in cases where large sections of an RNA secondary structure are unknown and/or the number of possible secondary structures is large. We expect that combining cryo-EM data and the DRRAFTER pipeline with nuclear magnetic resonance or biochemical techniques that probe RNA secondary structure will be critical to solving accurate structures for many RNPs⁴⁸.

Second, improvement to the final accuracy of DRRAFTER models will require advances in structure refinement tools. Existing refinement methods such as the PHENIX-ERRASER pipeline used here work best with high-resolution density maps and near atomic accuracy starting models. DRRAFTER model refinement will benefit from new tools that can handle more substantial structural changes and focus on refinement into lower-resolution maps.

Third, DRRAFTER does not remodel protein backbones or build missing protein coordinates. DRRAFTER may therefore build RNA coordinates into nearby unfilled protein density. This challenge can often be overcome by segmenting out density that is visually recognizable as belonging to a protein prior to DRRAFTER modeling. However, in some cases it is difficult to distinguish between density belonging to proteins and RNA. It may also be more challenging

to sample the correct protein-bound RNA conformation when the protein partner is not present. Ultimately, integrating DRRAFTER with existing protein structure modeling tools will be necessary to complete the pipeline for RNP model building.

Lastly, DRRAFTER automates RNA model building and error estimation, but final visual inspection should still play an important role in the modeling process. We present a graphical overview of typical mistakes that may occur when applying DRRAFTER and possible fixes (Fig. 5). We recommend visually inspecting at least the top ten DRRAFTER models; a similar process has been powerful for our ERRASER tool^{49,50}. Particularly when the modeling error is predicted to be high, visual examination can identify regions for which modeling assumptions, such as the secondary structure or initial placements of proteins and RNA helices, may be incorrect.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41592-018-0172-2>.

Received: 16 March 2018; Accepted: 31 July 2018;
Published online: 30 October 2018

References

- Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* **24**, 791–799 (2017).
- Feigon, J., Chan, H. & Jiang, J. S. Integrative structural biology of *Tetrahymena* telomerase – insights into catalytic mechanism and interaction at telomeres. *FEBS J.* **283**, 2044–2050 (2016).
- Jiang, F. G. & Doudna, J. A. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111 (2015).
- von Loeffelholz, O. et al. Focused classification and refinement in high-resolution cryo-EM structural analysis of ribosome complexes. *Curr. Opin. Struct. Biol.* **46**, 140–148 (2017).
- Zhou, Z. H. Atomic resolution cryo electron microscopy of macromolecular complexes. *Adv. Protein Chem. Struct. Biol.* **82**, 1–35 (2011).
- Leschziner, A. E. & Nogales, E. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 43–62 (2007).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EMEM density maps. *Nat. Methods* **11**, 63–65 (2014).
- Frank, J. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 303–319 (2002).
- Nguyen, T. H. D. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
- Greber, B. J. et al. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**, 515–519 (2014).
- Chaker-Margot, M. et al. Architecture of the yeast small subunit processome. *Science* **355**, eaal1880 (2017).
- Li, X. J. et al. Structure of ribosomal silencing factor bound to *Mycobacterium tuberculosis* ribosome. *Structure* **23**, 1858–1865 (2015).
- DiMaio, F. & Chiu, W. Tools for model building and optimization into near-atomic resolution electron cryo-microscopy density maps. *Methods Enzymol.* **579**, 255–276 (2016).
- Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153 (2015).
- Frenz, B. et al. RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat. Methods* **14**, 797–800 (2017).
- Kim, D. N. and K. Y. Sanbonmatsu, Tools for the cryo-EM gold rush: going from the cryo-EM map to the atomistic model. *Biosci Rep.* **37**, BSR20170072 (2017).
- Wang, R. Y. R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, e17219 (2016).
- Dawson, W. K. & Bujnicki, J. M. Computational modeling of RNA 3D structures and interactions. *Curr. Opin. Struct. Biol.* **37**, 22–28 (2016).
- Cowtan, K. Automated nucleic acid chain tracing in real time. *IUCr* **1**, 387–392 (2014).
- Chou, F. C. et al. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Keating, K. S. & Pyle, A. M. Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc. Natl Acad. Sci. USA* **107**, 8177–8182 (2010).
- Wang, X. Y. et al. RNABC: forward kinematics to reduce all-atom steric clashes in RNA backbone. *J. Math. Biol.* **56**, 253–278 (2008).
- Trabuco, L. G. et al. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
- Lu, M. & Steitz, T. A. Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-angstrom resolution. *Proc. Natl Acad. Sci. USA* **97**, 2023–2028 (2000).
- Handa, N. et al. Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature* **398**, 579–585 (1999).
- Yang, X. J. et al. Crystal structures of restriction-inhibitor complexes with implications for RNA recognition and base flipping. *Nat. Struct. Mol. Biol.* **8**, 968–973 (2001).
- Gutmann, S. et al. Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB. *Nature* **424**, 699–703 (2003).
- Kumarevel, T., Mizuno, H. & Kumar, P. K. Structural basis of HutP-mediated anti-termination and roles of the Mg²⁺ ion and L-histidine ligand. *Nature* **434**, 183–191 (2005).
- Yoshizawa, S. et al. Structural basis for mRNA recognition by elongation factor SelB. *Nat. Struct. Mol. Biol.* **12**, 198–203 (2005).
- Beuth, B. et al. Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. *EMBO J.* **24**, 3576–3587 (2005).
- Lee, T. T., Agarwalla, S. & Stroud, R. M. A unique RNA fold in the RumA-RNA-Cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120**, 599–611 (2005).
- Chao, J. A. et al. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
- Miller, M. T., Higgin, J. J. & Hall, T. M. T. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat. Struct. Mol. Biol.* **15**, 397–402 (2008).
- Nguyen, T. H. D. et al. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 angstrom resolution. *Nature* **530**, 298–302 (2016).
- Jiang, F. G. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
- Jiang, F. G. et al. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
- Greber, B. J. et al. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283–286 (2014).
- Li, X. N. et al. CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat. Commun.* **8**, 1035 (2017).
- Liu, S. et al. Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**, 1278–1283 (2017).
- Yan, C. Y. et al. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155 (2017).
- Fica, S. M. et al. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380 (2017).
- Jiang, J. S. et al. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).
- Larsen, K. P. et al. Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature* **557**, 118–122 (2018).
- Jiang, J. et al. Structure of Telomerase with telomeric DNA. *Cell* **173**, 1179–1190 (2018).
- Dai, X. H. et al. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116 (2017).
- Koning, R. I. et al. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.* **7**, 12524 (2016).
- Cheng, C. Y. et al. RNA structure inference through chemical mapping after accidental or intentional mutations. *Proc. Natl Acad. Sci. USA* **114**, 9876–9881 (2017).
- Chou, F. C. et al. RNA structure refinement using the ERRASER-Phenix pipeline. *Methods Mol. Biol.* **1320**, 269–282 (2016).
- Kapral, G. J. et al. New tools provide a second look at HDV ribozyme structure, dynamics and cleavage. *Nucleic Acids Res.* **42**, 12833–12846 (2014).

Acknowledgements

We thank members of the Das lab for useful discussions and members of the Rosetta community for discussions and code sharing. We thank J. Feigon and her lab for sharing their coordinates for the 8.9 Å *Tetrahymena* telomerase cryo-EM structure. Calculations were performed on the Stanford Sherlock cluster and Stanford BioX cluster, supported by NIH Shared Instrumentation Grant 1S10RR02664701. This work was supported by a Gabilan Stanford Graduate Fellowship (K.K.), the National Science Foundation (GRFP to K.K.), and the National Institutes of Health through awards T32 GM008294 (K.P.L. and K.K.), NIGMS R35 GM122579 (R.D.), R21 CA121487 (R.D.), R01 GM121487 (R.D. and P. Bradley), R01 GM114178 (R.Z.), R01 GM126157 (R.Z.), and R01 GM071940 (Z.H.Z.).

Author contributions

K.K. and R.D. designed the computational approach. K.K. implemented the method and performed the tests and analysis. S.L., Z.H.Z., and R.Z. provided the U1 snRNP and P complex blind test cases. K.P.L., G.S., E.V.P., and J.D.P. provided the HIV-1 RTIC test case and provided initial feedback on the method. K.K. and R.D. wrote the manuscript with input from S.L., K.P.L., G.S., E.V.P., J.D.P., Z.H.Z., and R.Z.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41592-018-0172-2>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

Methods

The DRRAFTER pipeline. For each system, all available structures of individual proteins were collected from the Protein Data Bank (PDB) and then fit into the cryo-EM density map in Chimera using the “Fit in Map” function³¹. Ideal A-form RNA helices were built with the Rosetta tool, `rna_helix.py`, and then fit into the maps in Chimera³¹. Following conventional protocols^{9–12}, these steps were performed manually, but completed rapidly (minutes per structure). Regions with missing RNA coordinates were identified and subdivided by visual inspection. The surrounding RNA helices and proteins were extracted from the overall model of the RNP and used as the input to the Rosetta DRRAFTER run.

The Rosetta stage consists of a modified version of the FARFAR method, run through the Rosetta `rna_denovo` application^{52,53}. The method was updated so that both proteins and density maps can be included. There are two stages to this protocol. First, a low-resolution Monte Carlo stage, which includes standard RNA fragment insertion moves to fold the RNA, now allows docking moves that optimize the placement of RNA helices and proteins. Docking moves for RNA helices include rotations and translations about the helical axis, in addition to the standard random rigid body perturbations. During this stage, the proteins are treated as rigid bodies. Each conformation is scored with the low-resolution RNA–protein potential in Rosetta⁵⁴, augmented by the `elec_dens_fast` score term, which scores the agreement between the map and model⁵⁵.

After the low-resolution stage, the structure goes through full-atom refinement. First, the structure is subjected to energy minimization in which the RNA, as well as the protein side chains within a 20.0 Å distance of any RNA atom, are allowed to move. Then, the structure is further refined through single residue fragment insertions, side chain packing, and small rigid body perturbations. The structure is then subjected to a second round of energy minimization. Scoring during these phases is performed with the full-atom Rosetta energy function, which includes terms that describe hydrogen bonding, electrostatics, torsional energy, van der Waals interactions, and solvation, and is also supplemented with the density score term `elec_dens_fast`^{55,56}. This score function is available within Rosetta as `rna_hires_with_protein.wts`. The top ten models are output from the run, with the centroid model highlighted, to be visually inspected and to allow final manual selection.

The DRRAFTER code is freely available to academic users as part of the Rosetta software package in Rosetta 3.10 and in weekly releases after March 14, 2018 (<https://www.rosettacommons.org>) and is automatically compiled along with ERRASER, which is already in routine use for RNA and RNP cryo-EM.

An example Rosetta command line is as follows:

```
DRRAFTER.py -fasta fasta.txt -secstruct secstruct.txt
-start_struct my_starting_structure.pdb -map_file my_cryoEM_map.mrc -map_reso 7.0 -residues_to_model A:20-30
-job_name my_drrafter_run
```

where `fasta.txt` is a FASTA file listing the full sequence of the complex, `secstruct.txt` is a file containing the secondary structure in dot bracket notation (with dots for protein residues), `-residues_to_model` (here given a value of A:20–30) specifies the residues that should be built in the DRRAFTER run, `my_starting_structure.pdb` is the PDB file containing all fit protein structures and RNA helices, `-map_file` specifies the density map, `-map_reso` specifies the resolution of the map, and `-job_name` specifies a name for the run (which controls the names of the output files). Documentation and a demo are available at <https://www.rosettacommons.org>.

We calculated modeling convergence by taking the average of the pairwise r.m.s. deviation values over the RNA region being modeled for the ten best-scoring DRRAFTER models. An example command line to calculate convergence and corresponding error estimates is as follows:

```
DRRAFTER.py -estimate_error -final_structures model_1.pdb
model_2.pdb model_3.pdb model_4.pdb model_5.pdb
model_6.pdb model_7.pdb model_8.pdb model_9.pdb
model_10.pdb
```

Approximately 3,000 DRRAFTER models were generated in all cases, and the ten top-scoring models were then subjected to the PHENIX-ERRASER pipeline²⁰. For the PHENIX runs, secondary structure restraints were automatically generated with `phenix.secondary_structure_restraints` and applied during refinement with `phenix.real_space_refine`. Additionally, coordinate restraints were applied for all residues in RNA helices. During the ERRASER runs, the first base pair of each RNA helix was kept fixed, as were any residues contacting a protein surface, or near enough that ERRASER introduced protein–RNA clashes if the residue was not kept fixed.

Model analysis. The r.m.s. deviation values (reported in Supplementary Table 1) were calculated over RNA heavy atoms after initial alignment over protein heavy atoms. These calculations were carried out in Rosetta and Pymol. The r.m.s. deviation values for previously modeled coordinates in the spliceosomal tri-snRNP were calculated for protein structures that had been fit into the lower-resolution (5.9 Å) density map in Chimera following the description in the methods section

of the original paper⁹ versus the high-resolution coordinates of the corresponding proteins in PDB 5GAN³⁵. Homologous protein structures that were docked into the lower-resolution map were omitted from this calculation. For the mitoribosome, r.m.s. deviation values were calculated between the coordinates deposited with the lower-resolution (4.9 Å) map (PDB 4CE4) and the high-resolution (3.4 Å) map (PDB 4V1A and 4V19) for proteins present in both, as well as for RNA regions that could not have been modeled by simple threading of the *E. coli* ribosome structure. For the Cas9–sgRNA complex, the protein coordinates were taken from the crystal structure of CRISPR–Cas9 in complex with sgRNA and double-stranded DNA (PDB 5F9R) and broken up into domains, and each of these was individually fit into the cryo-EM density map³⁶; r.m.s. deviation values between these regions and the high-resolution crystal structure (PDB 4ZT0) were calculated.

Local map resolution was calculated with Resmap⁷, then loaded into Chimera along with the corresponding high-resolution coordinates. The “Values at Atom Positions” tool in Chimera was used to find the local resolution at the positions of each of the atoms in the high-resolution structure. The values at the positions of all of the RNA atoms for the region being modeled were averaged (with a Python script) to give the local resolution for that region.

Best-fit lines describing the upper and lower bounds of DRRAFTER model accuracy versus local resolution (Fig. 3a) were calculated using the minimum r.m.s. deviation values (lower bound) or 90th-percentile r.m.s. deviation values (upper bound) in each 1 Å bin ranging from 2.5 to 12.5 Å local resolution.

Real-space correlation coefficients were calculated for RNA coordinates being modeled only (surrounding proteins were not included to facilitate comparison between high- and low-resolution coordinates) using the PHENIX tool `phenix.get_cc_mtz_pdb` with `fix_xyz = True` and `scale = True`. The “Map correlation in region of model” was reported.

Figures were generated with Pymol and UCSF Chimera. The versions of all software used in this study are listed in the Nature Research Reporting Summary.

Statistics. Pearson’s correlation coefficients were calculated for local resolution (determined as described above) versus model accuracy for a total of 128 models, of which 30 were DRRAFTER models built into simulated maps, 25 were DRRAFTER models built into experimental maps, 6 were blind DRRAFTER models built into experimental maps, and 67 were previously modeled low-resolution protein and RNA coordinates. Pearson’s correlation coefficients were also calculated for the mean, median, and best model accuracy out of the ten top-scoring DRRAFTER models versus modeling convergence (calculated as described above) for 61 systems of which 30 were DRRAFTER models built into simulated maps, 25 were DRRAFTER models built into experimental maps, and 6 were blind DRRAFTER models built into experimental maps. Two-tailed *P* values are reported for all correlation coefficients.

Simulated benchmark. Ten systems were chosen from the nonredundant set of RNA–protein complexes with corresponding unbound protein structures available, described in ref. ⁵⁷. The specific systems were selected manually to represent a diversity of types of RNA–protein interactions (unbound protein structures listed in parentheses): 1DFU (1B75), 1B7F (3SXL), 1JBS (1AQZ), 1P6V (1K8H), 1WPU (1WPV), 1WSU (1LVA), 2ASB (1K0R), 2BH2 (1UWV), 2QUX (2QUD), and 3BX2 (3BWT). For each of these systems, density maps were simulated at 3.0, 5.0, and 7.0 Å resolution with the `pdb2vol` tool in the Situs package⁵⁸. Unbound protein structures (listed above) were fit into the simulated density maps using Chimera’s Fit in Map tool. Ideal RNA helices for helical segments of RNA were generated with `rna_helix.py` in Rosetta and then fit into the maps using Chimera’s Fit in Map tool. For systems that contained only single-stranded RNA, an ideal A-form nucleotide was fit approximately into the map—throughout the later DRRAFTER simulation, it was allowed to change its conformation and orientation within the map. The remaining RNA residues were also built with the DRRAFTER protocol in Rosetta. The full protein structures were included in the simulations, and were allowed to dock as rigid bodies within the density map. The ideal RNA helices were also subjected to docking within the map to optimize their final placement.

Spliceosomal tri-snRNP modeling. All proteins listed in Extended Data Table 1 of the original paper⁹ were fit into the full tri-snRNP density map (EMD-2966), as well as the structure of the C-terminal fragment of PRP3, which had since been solved (PDB 4YHU)⁵⁹. Ideal RNA helices were fit into the map for all helical parts of the three regions modeled: the U5 snRNA three-way junction (residues 35–53, 62–91, and 103–119), the U5 snRNA internal loop II (residues 4–40, 114–144), and the U4/U6 snRNA three-way junction consisting of U4 snRNA residues 1–64 and U6 snRNA residues 55–80. All RNA helices were allowed to move as rigid bodies throughout the DRRAFTER runs. Proteins were kept fixed. In each case, the density map was approximately segmented around the region of interest with the Segment Map tool in Chimera (Segger v1.9.4). R.m.s. deviation values were calculated relative to the coordinates from the 3.7 Å map, PDB 5GAN³⁵.

For DRRAFTER models built into the 3.7 Å map (EMD-8012), the protein structures were taken from the corresponding PDB entry, 5GAN. Ideal RNA helices were fit into the map and DRRAFTER runs were performed as described above.

Mitoribosome modeling. DRRAFTER models were built extending from the coordinates deposited with the 4.9 Å map (EMD-2490), PDB 4CE4, for two regions for which RNA coordinates were missing¹⁰. “Loop 1” consisted of RNA residues 401–407, and “loop 2” consisted of RNA residues 495–547. Connected RNA residues were included in the simulations. For loop 2, an initial model of residues 502–522 and 529–544 was built by taking H43 and H44 from the *E. coli* ribosome structure (PDB 4YBB) and threading in the mitoribosome sequence⁶⁰. This model was fit approximately into the density map with the Fit in Map function in Chimera and then included as a rigid body, allowed to rotate and translate, in the DRRAFTER run. Models were similarly built into the 3.4 Å map (EMD-2787)³⁸, but surrounding protein and RNA coordinates were taken from PDB structures 4V19 and 4V1A (deposited with the 3.4 Å map). We additionally built DRRAFTER models for 17 regions where manually built models had been deposited for the 4.9 Å map: residues 96–99, 220–223, 226–228, 271–274, 591–595, 612–617, 709–710, 720–728, 742–748, 772–774, 803–814, 886–889, 1,124–1,128, 1,185–1,188, 1,237–1,240, 1,488–1,492, and 1,543–1,551.

CRISPR–Cas9–sgRNA modeling. Protein coordinates were taken from the crystal structure of CRISPR–Cas9 in complex with sgRNA and double-stranded DNA (PDB 5F9R)³⁶. The protein was split up into seven domains (Arg, CTD, HNH, Helical-I, Helical-II, Helical-III, and RuvC) and each was fit individually into the 4.5 Å cryo-EM map (EMD-3276)³⁶. The protein domains were kept fixed throughout the DRRAFTER run. Ideal A-form RNA helices were fit into the map for all helical sections of the sgRNA. Models were built for sgRNA residues 11–99, but r.m.s. deviation values were computed only over residues with coordinates in the 2.9 Å crystal structure, PDB 4ZT0 (residues 11–30 and 57–99)³⁷. Models were similarly built into the 2.9 Å crystallographic density map (4ZT0), but with the protein coordinates taken from 4ZT0.

Blind yeast U1 snRNP modeling. Modeling was performed with a 6.0 Å resolution map of the yeast U1 snRNP from an earlier stage of processing than the later published 3.6 Å map³⁹. The core four-way junction region of the map was identified by fitting the structures of the human U1 snRNP (3CW1 and 3PGW) into the map^{61,62}. Structures of the seven yeast Sm proteins (B, D1, D2, D3, E, F, and G) were taken from PDB 5GMK and fit into the map with the Fit in Map tool in Chimera⁶³. Homology models of PRP39 and PRP42 were generated with Modeller and fit into the map⁶⁴. A homology model of the U1-70K RRM was fit into the map and later allowed to move as a rigid body. We assumed that the U1 snRNA would adopt the secondary structure proposed in the literature⁶⁵. Ideal RNA helices were fit into the map for all helical regions of the RNA. DRRAFTER models were built for five regions of the RNA: the core four-way junction (residues 11–60, 154–178, and 534–559), the SL3-1/SL3-2/SL3-6 yeast-specific three-way junction (residues 172–185, 304–325, and 526–539), the yeast-specific four-way junction (residues 181–202 and 236–308), and SL3-7 (residues 310–531).

Blind yeast spliceosomal P complex modeling. Models of the P complex ligated exon were built into a 5.4 Å resolution map from an earlier stage of processing than the later published 3.3 Å map⁴⁰. Previously determined structures of the yeast spliceosomal C* complex were fit into the map (PDB 5MQ0 and 5WSG), which allowed identification of the density for the ligated exon^{41,42}. Coordinates for PRP22 were taken from the C* complex (5MQ0) and fit into the density map individually. The coordinates of the RNA bound to PRP22 were modeled by taking the structure of PRP43 in complex with RNA (PDB 5I8Q) and aligning it to PRP22, then taking the resulting RNA coordinates from the complex⁴⁶. These RNA coordinates were kept fixed relative to PRP22 in all DRRAFTER runs. DRRAFTER runs were set up with varying numbers of nucleotides spanning the exon–exon junction and the active site in PRP22, ranging from 10 to 20 nucleotides. Models were selected from the runs with the fewest number of nucleotides spanning the exon–exon junction and the PRP22 active site in which there were no breaks in the RNA chain (13 and 14 nucleotides).

HIV-1 RTIC modeling. Approximate initial locations for all helical segments of the HIV-1 RNA and bound tRNA were determined by fitting ideal A-form helices into an 8.0 Å map of the HIV-1 RTIC⁴⁴. The alternative tRNA secondary structure was assumed as in the previously published manual modeling. Protein coordinates were taken from the previously published model. Final refinement was carried out only with PHENIX, as was carried out for the previously published model. The 15 best-scoring models were visually inspected and the top ten without large distortions in the primer binding site helix were selected as the final set of ten best-scoring models.

Tetrahymena telomerase modeling. All proteins described in the original paper⁴³ were fit into the 8.9 Å map of *Tetrahymena* telomerase (EMD-6443). Additionally, the RNA pseudoknot (5KMZ)⁴³, RNA residues 155–159 bound to the N-terminal domain of the human La protein (2VOP)⁶⁷, the structure of the RNA TBE bound to the TRBD (5C9H)⁶⁸, and the RNA stem IV loop (2M21)⁶⁹, RNA stem IV (4ERD)⁷⁰, “half” an ideal A-form helix for the template RNA, and ideal A-form helices for the remaining helical regions of the RNA were fit into the map with Fit

in Map in Chimera and then each allowed to move individually in the subsequent DRRAFTER runs. The full RNA was modeled as a single region.

MS2 packaged genome modeling. The packaged MS2 genome was modeled based on the 3.6 Å map (EMD-8397) using the published proposed secondary structure⁴⁶. Because the RNA density in this map is noisy, a 1.5 Å Gaussian filter was applied to the map in Chimera prior to RNA modeling (similarly, RNA density in the original paper⁴⁶ was examined after low-pass filtering to 6 Å resolution). Models were built for ten regions: S1 + S2 (residues 29–227, 341–369); S3 (residues 372–583); S4 (residues 888–943); S5 + S6 (residues 963–1,119); S7 (residues 1,132–1,283); S8 (residues 1,714–1,806); S9-1 (residues 1,837–1,896); S9-2 (residues 1,900–1,940); S10 (residues 1,960–2,122); S12 (residues 1,810–1,826, 2,202–2,340); S15 + S16 (residues 2,346–2,353, 2,757–2,661, 3,088–3,111, 3,249–3,382). The published coordinates for the protein capsid and bound RNA hairpins were kept fixed (5TC1)⁴⁶. One ideal RNA helix for each region was fit into the map; the initial coordinates of the remaining helices were not provided for the DRRAFTER run (and were therefore determined by the initial random perturbations to the RNA structure). For comparison, models were similarly built into the 10.5 Å map (EMD-3403)³⁷, without the high-resolution coordinates of the RNA hairpins. Because the 3.6 and 10.5 Å maps differed significantly in regions S9-1 and S9-2 (Supplementary Fig. 5), r.m.s. deviation values for these regions were calculated after alignment over all RNA heavy atoms. For all other regions, r.m.s. deviation values were calculated over RNA heavy atoms after alignment over all protein residues.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The DRRAFTER code is freely available to academic users as part of the Rosetta software package in weekly releases starting with 2018.12 at <https://www.rosettacommons.org/> and is also available in the Rosetta 3.10 release. Instructions for setting up Rosetta and running the DRRAFTER software are available at https://www.rosettacommons.org/docs/latest/application_documentation/rna/drrafter. A demo is available at <https://www.rosettacommons.org/demos/latest/public/drrafter/README>. All necessary files for the demo are included with Rosetta in the folder ROSETTA_HOME/demos/public/drrafter/, where ROSETTA_HOME is the path to your Rosetta directory.

Data availability

The accession codes used in this study are as follows: *E. coli* L25-5S rRNA (PDB 1DFU and 1B75), sex-lethal RRM (PDB 1B7F and 3SXL), ribotoxin restrictocin sarcin-ricin loop analog (PDB 1JBS and 1AQZ), the SmpB–tmRNA complex (PDB 1P6V and 1K8H), the HutP antitermination complex (PDB 1WPU and 1WPV), the mRNA-binding domain of the SelB elongation factor (PDB 1WSU and 1LVA), the NusA transcriptional regulator (PDB 2ASB and 1K0R), the methyltransferase RumA in complex with rRNA (PDB 2BH2 and 1UWV), the PP7 coat protein and viral RNA (PDB 2QUX and 2QUD), Puf4 bound to the 3' UTR of the target transcript (PDB 3BX2 and 3BWT), tri-snRNP (EMD-2966 and EMD-8012), PDB 4YHU and 5GAN, in addition to all PDB codes listed in Extended Data Table 1 of ref. ⁹), the mitochondrial ribosome (EMD-2490 and EMD-2787; PDB 4CE4, 4V19, and 4V1A), CRISPR–Cas9–sgRNA complex (EMD-3276; PDB 5F9R and 4ZT0), U1 snRNP (EMD-8622; PDB 3CW1, 3PGW, 5GMK and 5UZ5), the spliceosomal P complex (PDB 5MQ0, 5WSG, 5I8Q and 6BK8), HIV-1 RTIC (described in ref. ⁴⁴), and *Tetrahymena* telomerase (EMD-6443; PDB 5KMZ, 2VOP, 5C9H, 2M21 and 4ERD), MS2 packaged genome (EMD-8397 and EMD-3403; PDB 5TC1). The DRRAFTER models of the U1 snRNP (from the 3.6 Å map) and the packaged MS2 genome (from the 3.6 Å map) are available in Supplementary Data 1 and 2. DRRAFTER models for all other systems are available at <https://purl.stanford.edu/jj049gk5411>.

References

- Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **7**, 291–294 (2010).
- Kappel, K. & Das, R. Sampling native-like structures of RNA–protein complexes through Rosetta folding and docking. *bioRxiv*, Preprint at <https://www.biorxiv.org/content/early/2018/06/05/339374> (2018).
- DiMaio, F. et al. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
- Alford, R. F., et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
- Perez-Cano, L. & Fernandez-Recio, J. Optimal protein–RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* **78**, 25–35 (2010).

58. Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).
59. Liu, S. et al. A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *eLife* **4**, e07320 (2015).
60. Noeske, J. et al. High-resolution structure of the *Escherichia coli* ribosome. *Nat. Struct. Mol. Biol.* **22**, 336–341 (2015).
61. Weber, G. et al. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184 (2010).
62. Krummel, D. A. P. et al. Crystal structure of a ten-subunit human spliceosomal U1 snRNP at 5.5 angstrom resolution. *Biophys. J.* **100**, 198–198 (2011).
63. Wan, R. X. et al. Structure of a yeast catalytic step I spliceosome at 3.4 angstrom resolution. *Science* **353**, 895–904 (2016).
64. Eswar, N. et al. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **15**, 5.6.1–5.6.30 (2006).
65. Kretzner, L., Krol, A. & Rosbash, M. *Saccharomyces cerevisiae* U1 small nuclear-RNA secondary structure contains both universal and yeast-specific domains. *Proc. Natl Acad. Sci. USA* **87**, 851–855 (1990).
66. He, Y. Z. et al. Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA* **23**, 1110–1124 (2017).
67. Kotik-Kogan, O. et al. Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. *Structure* **16**, 852–862 (2008).
68. Jansson, L. I. et al. Structural basis of template-boundary definition in *Tetrahymena* telomerase. *Nat. Struct. Mol. Biol.* **22**, 883–888 (2015).
69. Richards, R. J. et al. Structural study of elements of *Tetrahymena* telomerase RNA stem-loop IV domain important for function. *RNA* **12**, 1475–1485 (2006).
70. Singh, M. et al. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell* **47**, 16–26 (2012).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Code described in this study is available as part of the Rosetta software package in releases after March 14, 2018, excluding Rosetta 3.9, at www.rosettacommons.org. Documentation and a demo are also available at https://www.rosettacommons.org/docs/latest/application_documentation/rna/drrafter and <https://www.rosettacommons.org/demos/latest/public/drrafter/README>.

Data analysis

Rosetta (versions after March 2018, excluding Rosetta 3.9), Chimera version 1.11.2, Segger version 1.9.4, PyMOL version 1.8.4.2, python v2.7, Situs version 2.7.2, PHENIX version 1.12, Modeller version 9.19, ResMap version 1.1.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The accession codes used in this study are as follows: E. coli L25-5S rRNA (PDB 1DFU and 1B75), sex-lethal RRM (PDB 1B7F and 3SXL), ribotoxin restrictocin sarcin-ricin loop analog (PDB 1JBS and 1AQZ), SmpB-tmRNA complex (PDB 1P6V and 1K8H), HutP antitermination complex (PDB 1WPU and 1WPV), mRNA binding domain of SelB elongation factor (PDB 1WSU and 1LVA), NusA transcriptional regulator (PDB 2ASB and 1KOR), methyltransferase RumA in complex with rRNA (PDB 2BH2 and 1UWV), PP7 coat protein and viral RNA (PDB 2QUX and 2QUD), Puf4 bound to 3' UTR of target transcript (PDB 3BX2 and 3BWT), tri-snRNP (EMD 2966 and 8012; PDB 4YHU and 5GAN, in addition to all PDB codes listed in Extended Data Table 1 of [9]), mitochondrial ribosome (EMD 2490 and 2787; PDB 4CE4, 4V19, and 4V1A), CRISPR-Cas9-sgRNA complex (EMD 3276; PDB 5F9R, 4ZT0), U1 snRNP (EMD 8622; PDB 3CW1, 3PGW, 5GMK, 5UZ5), spliceosomal P complex (PDB 5MQO, 5WSG, 5I8Q, 6BK8), HIV-1 RTIC (described in [44]), and Tetrahymena telomerase (EMD 6443; PDB 5KMZ, 2VOP, 5C9H, 2M21, 4ERD), MS2 packaged genome (EMD 8397 and 3403; PDB 5TC1). The DRRAFTER models of the U1 snRNP (from the 3.6 Å map) and the packaged MS2 genome (from the 3.6 Å map) are available in the Supplementary Information.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. The sample size (number of systems modeled with DRRAFTER) was a result of our effort to model systems representing all major classes of RNPs with maps with a wide range of resolutions.
Data exclusions	No data was excluded.
Replication	Experimental replication was not attempted.
Randomization	Not applicable. All experiments in this study were performed in silico.
Blinding	Blinding to group allocation is not applicable to this study as all experiments were performed in silico.

Materials & experimental systems

Policy information about [availability of materials](#)





n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Research animals
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Method-specific reporting

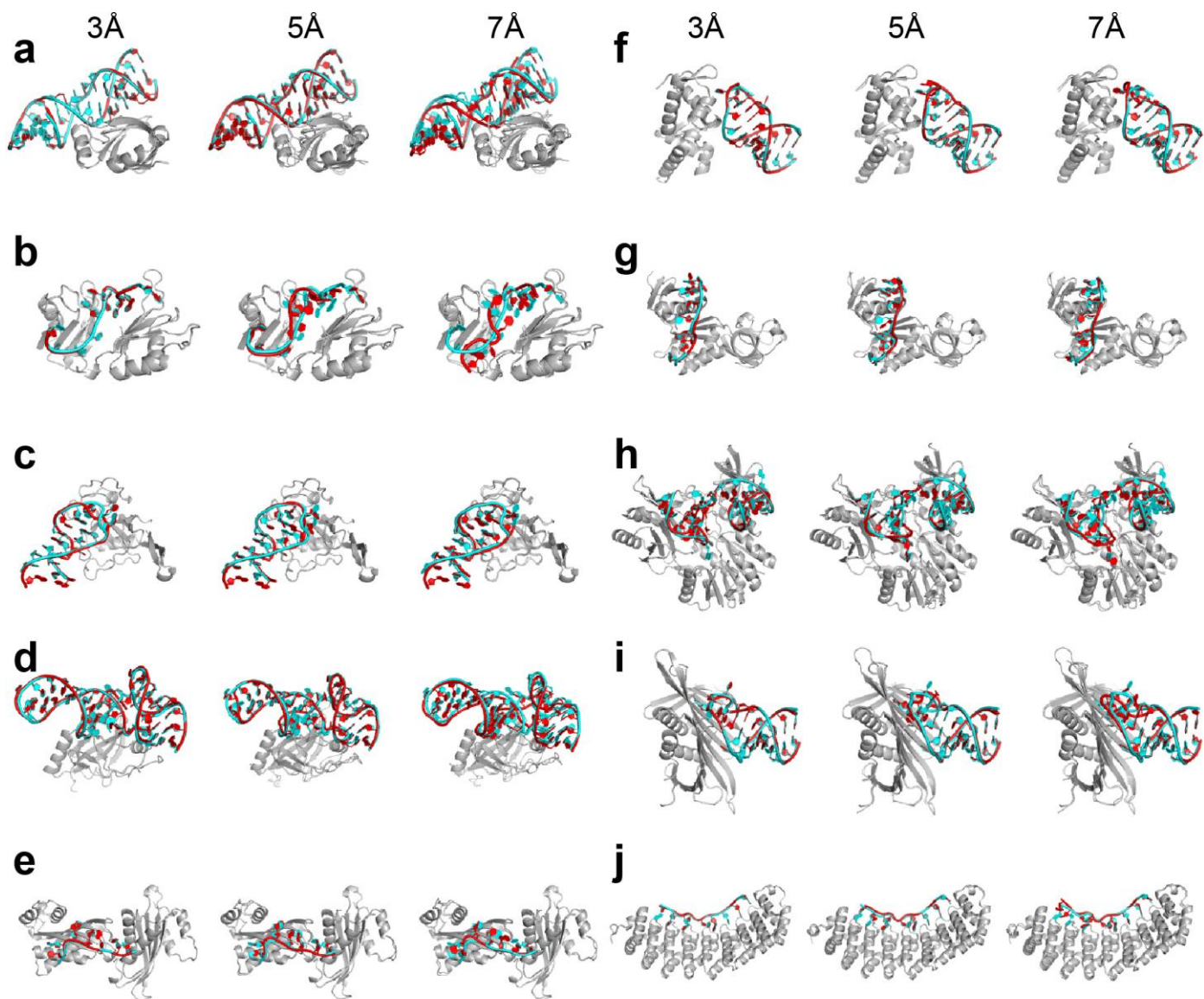
n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Magnetic resonance imaging

In the format provided by the authors and unedited.

De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes

Kalli Kappel ¹, Shiheng Liu^{2,3}, Kevin P. Larsen ^{1,4}, Georgios Skiniotis ^{4,5}, Elisabetta Viani Puglisi⁴, Joseph D. Puglisi ⁴, Z. Hong Zhou^{2,3}, Rui Zhao⁶ and Rhiju Das ^{1,7,8*}

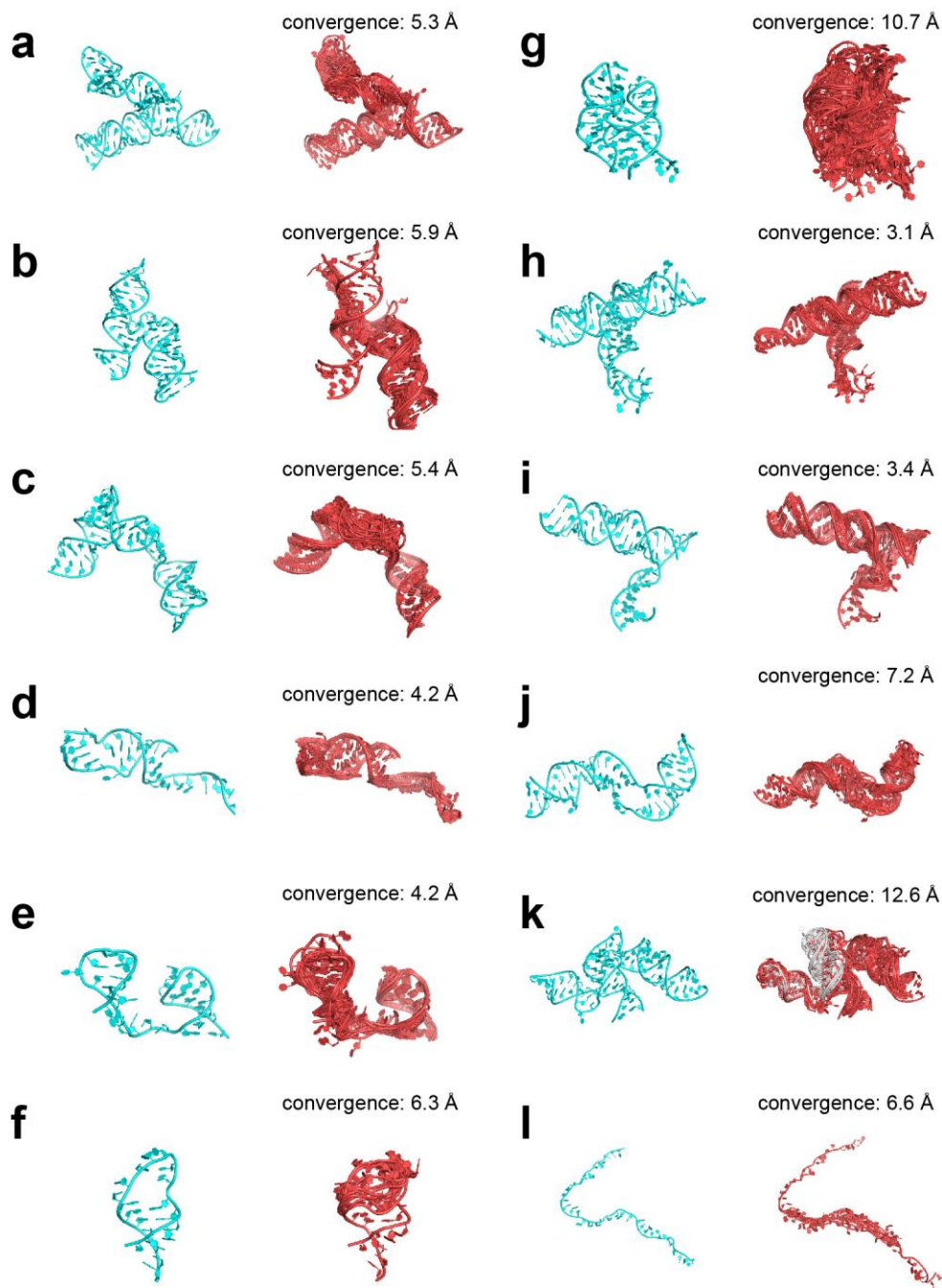
¹Biophysics Program, Stanford University, Stanford, CA, USA. ²Electron Imaging Center for Nanomachines, California NanoSystems Institute, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ³Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA, USA. ⁴Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ⁵Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ⁶Department of Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO, USA. ⁷Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. ⁸Department of Physics, Stanford University, Stanford, CA, USA. *e-mail: rhiju@stanford.edu



Supplementary Figure 1

DRRAPTER models for ten small RNA-protein systems.

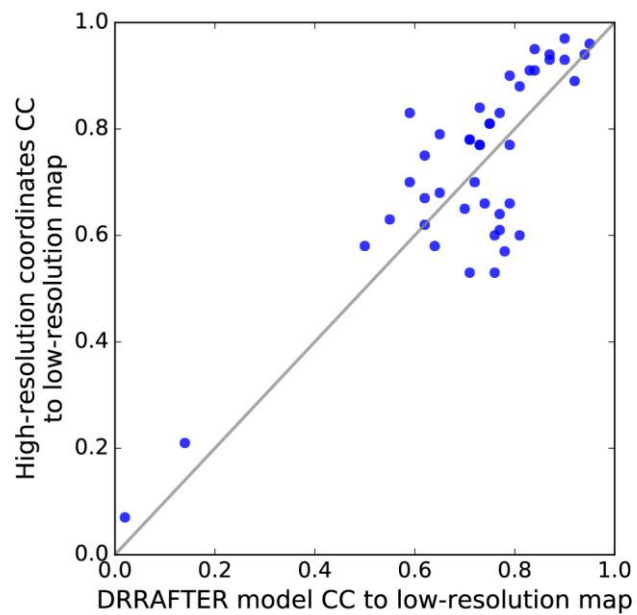
a–j, The best RMSD DRRAPTER models of the top ten scoring (RNA colored red, protein colored gray) overlaid with the deposited crystallographic coordinates (RNA colored cyan, protein colored gray) shown for 3-Å (left), 5-Å (middle), and 7-Å (right) simulated density maps for *E. coli* L25-5S rRNA (1dfu) (**a**), sex-lethal RRM (1b7f) (**b**), ribotoxin restrictocin sarcin-ricin loop analog (1jbs) (**c**), SmpB-tmRNA complex (1p6v) (**d**), HutP antitermination complex (1wpu) (**e**), the mRNA-binding domain of the SelB elongation factor (1wsu) (**f**), the NusA transcriptional regulator (2asb) (**g**), the methyltransferase RumA in complex with rRNA (2bh2) (**h**), the PP7 coat protein and viral RNA (2qux) (**i**), and Puf4 bound to the 3' UTR of target transcript (3bx2) (**j**).



Supplementary Figure 2

Convergence of DRRAFTER models.

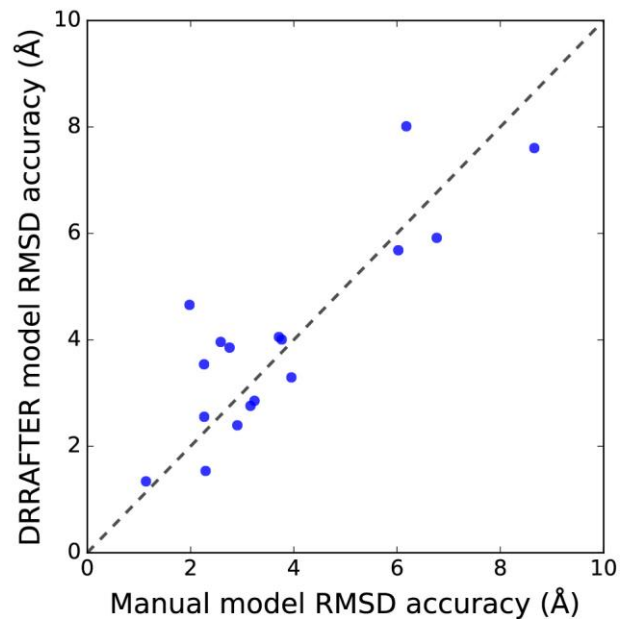
a–l, High-resolution RNA coordinates (cyan, left) and the top ten scoring DRRAFTER models (red, right) for the spliceosomal tri-snRNP U4/U6 three-way junction (**a**), U5 three-way junction (**b**), and U5 internal loop II (**c**); the CRISPR–Cas9–sgRNA complex sgRNA residues 11–30 and 57–68 (**d**) and sgRNA residues 69–99 (**e**); mitoribosome loop 1 (**f**) and loop 2 (**g**); yeast U1 snRNP (blind) core four-way junction (**h**), yeast three-way junction (**i**), SL2-2 (**j**), and yeast four-way junction (**k**) (DRRAFTER models of SL3-2, SL3-3, and SL3-5 colored red; DRRAFTER models of SL3-4 colored white); and yeast spliceosomal P complex (blind) ligated exon (**l**).



Supplementary Figure 3

Assessing the agreement between DRRAFTER models and the lower-resolution density maps versus the agreement of high-resolution coordinates to the same lower-resolution density maps.

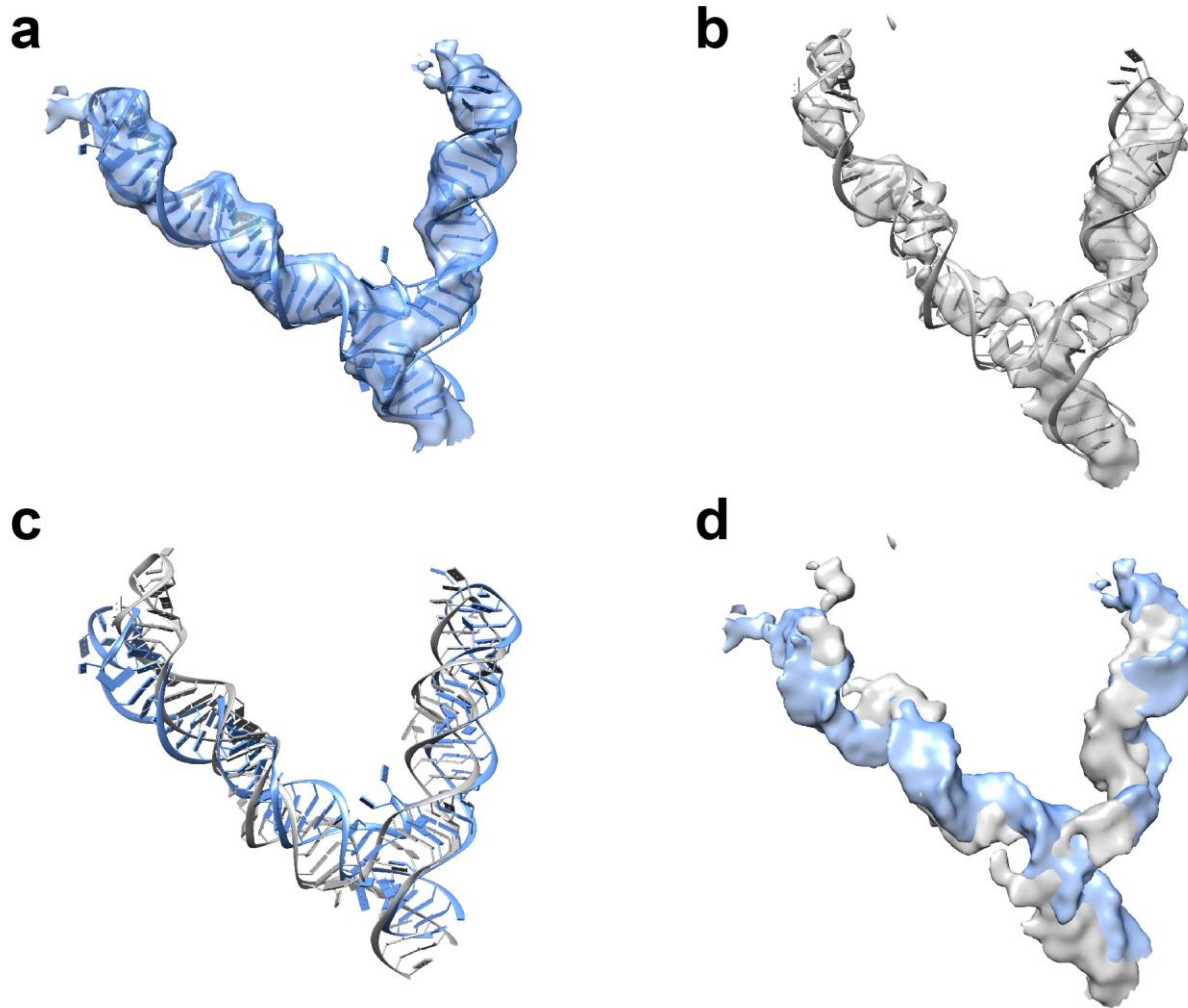
Real-space correlation coefficients (CC) for the best RMSD DRRAFTER models out of the top ten scoring for all systems described in Supplementary Table 1 to lower-resolution density maps are plotted against CC values for the corresponding high-resolution coordinates to the same lower-resolution density maps (see Methods for details).



Supplementary Figure 4

Comparing the accuracy of DRRAFTER models to manual models built into the lower-resolution mitoribosome map.

DRRAFTER models were built for all regions in the mitoribosome for which manual models were previously deposited. The accuracy of the manual (previously deposited) and DRRAFTER models was determined by comparing to the higher-resolution mitoribosome coordinates.



Supplementary Figure 5

Differences between the 3.6-Å and 10.5-Å maps of the packaged MS2 genome.

a, The model of region S9 (S9-1 and S9-2) (blue) built into the 3.6-Å map (blue, transparent). **b**, The model of region S9 (gray) built into the 10.5-Å map (gray, transparent). Each of the models fits well in the map in which it was built. There are, however, differences between the two models. **c**, Models built into the 3.6-Å and 10.5-Å maps overlaid. These differences are largely due to the underlying differences in the maps. **d**, The 3.6-Å (blue) and 10.5 Å (gray) maps overlaid.

Supplementary Table 1. Local resolution and RMSD accuracy for all systems

System		Reported, Local Map Resolution (Å)	Best, Centroid, Mean (standard deviation) RMSD of top 10 scoring models (Å)	Convergence Estimate (Å)
<i>E. coli</i> L25-5S rRNA (1dfu) ¹		5.0, 5.0	0.7, 1.0, 1.0 (2.1)	0.6
		7.0, 7.0	2.5, 4.1, 4.0 (2.4)	2.5
Sex-lethal RRM (1b7f) ¹		5.0, 5.0	3.5, 4.7, 6.0 (1.2)	6.6
		7.0, 7.0	5.3, 5.4, 6.6 (1.5)	6.9
Ribotoxin restrictocin – SRL analog (1jbs) ¹		5.0, 5.0	1.8, 2.5, 2.4 (0.3)	2.1
		7.0, 7.0	2.0, 2.7, 3.1 (0.4)	3.2
SmpB-tmRNA complex (1p6v) ¹		5.0, 5.0	2.8, 3.3, 3.8 (0.9)	4.1
		7.0, 7.0	3.1, 5.3, 5.9 (1.7)	6.1
HutP antitermination complex (1wpu) ¹		5.0, 5.0	1.3, 1.3, 1.8 (0.4)	1.9
		7.0, 7.0	2.3, 3.4, 3.8 (0.9)	4.0
mRNA binding domain of SelB elongation factor (1wsu) ¹		5.0, 5.0	1.8, 2.0, 2.1 (0.2)	1.5
		7.0, 7.0	2.1, 2.5, 2.4 (0.2)	1.5
NusA transcriptional regulator (2asb) ¹		5.0, 5.0	1.6, 2.1, 3.2 (0.9)	4.0
		7.0, 7.0	1.4, 3.9, 4.1 (1.2)	4.7
Methyltransferase RumA in complex with rRNA (2bh2) ¹		5.0, 5.0	2.7, 3.4, 3.8 (0.8)	3.9
		7.0, 7.0	4.1, 4.5, 4.6 (0.5)	4.4
PP7 coat protein and viral RNA (2qux) ¹		5.0, 5.0	2.6, 2.7, 3.0 (0.3)	2.6
		7.0, 7.0	2.8, 3.1, 3.1 (0.3)	2.8
Puf4 bound to 3' UTR of target transcript (3bx2) ¹		5.0, 5.0	1.3, 1.7, 2.5 (1.0)	2.9
		7.0, 7.0	2.3, 3.0, 3.4 (0.8)	3.6
U4/U6.U5 tri-snRNP	U4/U6 3WJ	5.9, 6.4	3.4, 5.6, 5.1 (0.9)	5.3
	U5 3WJ	5.9, 6.5	4.2, 5.2, 6.2 (2.0)	5.9
	U5 IL II	5.9, 6.5	5.9, 6.3, 6.7 (0.7)	5.4
Mitochondrial ribosome	Loop 1	4.9, 7.0	4.2, 8.4, 7.6 (2.0)	6.3
	Loop 2	4.9, 9.8	6.2, 10.4, 9.5 (2.1)	10.7
CRISPR-Cas9	Full RNA	4.5, 5.0	3.2, 3.6, 4.0 (0.7)	4.2
MS2 packaged genome	S1+S2	10.5, 11.0	5.7, 5.9, 7.3 (2.2)	8.5
	S3	10.5, 12.4	7.0, 9.1, 9.8 (1.5)	9.6
	S4	10.5, 10.6	4.4, 5.1, 8.2 (3.4)	8.0
	S5+S6	10.5, 11.2	5.6, 7.0, 10.0 (3.0)	9.6
	S7	10.5, 11.8	7.2, 9.8, 9.7 (1.5)	8.9
	S8	10.5, 10.7	5.4, 5.4, 6.6 (0.8)	6.3
	S9-1 ²	10.5, 12.2	3.6, 5.4, 5.0 (1.2)	6.8
	S9-2 ²	10.5, 12.0	3.0, 3.1, 3.5 (0.4)	5.6
	S10	10.5, 11.0	7.1, 8.0, 9.2 (1.1)	7.8
	S12	10.5, 11.3	4.7, 6.1, 9.1 (2.9)	8.6
S15+S16	10.5, 11.5	6.4, 7.1, 14.0 (9.9)	17.5	
Yeast U1 snRNP	Core 4WJ ³	6.0, 6.6	3.1, 3.4, 3.7 (0.5)	3.1
	Core 4WJ only ³	6.0, 6.6	1.6, 2.2, 2.6 (0.8)	2.4
	Yeast 3WJ ³	6.0, 6.6	2.4, 3.3, 3.4 (0.5)	3.4

	Yeast-specific 4WJ ³	6.0, 6.6	4.3, 10.3, 12.2 (5.9)	12.6
	SL2-2 ³	6.0, 6.6	4.0, 5.0, 6.8 (3.6)	7.2
	Yeast-specific 4WJ	6.0, 6.6	4.2, 4.6, 5.0 (0.4)	4.6
	SL2-2	6.0, 6.6	2.5, 4.7, 4.5 (0.8)	3.7
Yeast P complex ³	Ligated exon ³	5.4, 7.3	6.2, 6.4, 7.7 (1.4)	6.6
Overall (average)			3.6, 4.8, 5.5	5.5

¹Simulated density maps.

²To account for differences in the high- and low-resolution maps (Figure S5), RMSDs were calculated after alignment over RNA residues.

³Blind predictions.

Supplementary Table 2. Real-space correlation coefficients (CC) for the best RMSD DRRAFTER models out of the top ten scoring and for the corresponding high-resolution RNA coordinates within the lower-resolution map

System		Reported, Local Map Resolution (Å)	Model CC (RNA only)	High-res structure CC (RNA only)
<i>E. Coli</i> L25-5S rRNA (1dfu) [*]		5.0, 5.0	0.95	0.96
		7.0, 7.0	0.90	0.93
Sex-lethal RRM (1b7f) [*]		5.0, 5.0	0.59	0.83
		7.0, 7.0	0.59	0.70
Ribotoxin restrictocin – SRL analog (1jbs) [*]		5.0, 5.0	0.87	0.93
		7.0, 7.0	0.92	0.89
SmpB-tmRNA complex (1p6v) [*]		5.0, 5.0	0.84	0.95
		7.0, 7.0	0.83	0.91
HutP antitermination complex (1wpu) [*]		5.0, 5.0	0.75	0.81
		7.0, 7.0	0.70	0.65
mRNA binding domain of SelB elongation factor (1wsu) [*]		5.0, 5.0	0.90	0.97
		7.0, 7.0	0.94	0.94
NusA transcriptional regulator (2asb) [*]		5.0, 5.0	0.81	0.88
		7.0, 7.0	0.72	0.70
Methyltransferase RumA in complex with rRNA (2bh2) [*]		5.0, 5.0	0.79	0.90
		7.0, 7.0	0.73	0.84
PP7 coat protein and viral RNA (2qux) [*]		5.0, 5.0	0.87	0.94
		7.0, 7.0	0.84	0.91
Puf4 bound to 3' UTR of target transcript (3bx2) [*]		5.0, 5.0	0.77	0.83
		7.0, 7.0	0.65	0.68
U4/U6.U5 tri-snRNP	U4/U6 3WJ	5.9, 6.4	0.75	0.81
	U5 3WJ	5.9, 6.5	0.79	0.77
	U5 IL II	5.9, 6.5	0.74	0.66
Mitochondrial ribosome	Loop 1	4.9, 7.0	0.14	0.21
	Loop 2	4.9, 9.8	0.02	0.07
CRISPR-Cas9	Full RNA	4.5, 5.0	0.62	0.62
MS2 packaged genome	S1+S2	10.5, 11.0	0.78	0.57
	S3	10.5, 12.4	0.76	0.53
	S4	10.5, 10.6	0.55	0.63
	S5+S6	10.5, 11.2	0.76	0.60
	S7	10.5, 11.8	0.77	0.61
	S8	10.5, 10.7	0.62	0.67
	S9-align1	10.5, 12.2	0.71	0.53
	S9-align2	10.5, 12.0	0.81	0.60
	S10	10.5, 11.0	0.77	0.64
	S12	10.5, 11.3	0.79	0.66
	S15+S16	10.5, 11.5	0.64	0.58
Yeast U1 snRNP	Core 4WJ [†]	6.0, 6.6	0.73	0.77
	Core 4WJ only [†]	6.0, 6.6	0.50	0.58
	Yeast 3WJ [†]	6.0, 6.6	0.73	0.77
	Yeast-specific	6.0, 6.6	0.62	0.75

	4WJ [†]			
	SL2-2 [†]	6.0, 6.6	0.71	0.78
	Yeast-specific 4WJ	6.0, 6.6	0.65	0.79
	SL2-2	6.0, 6.6	0.71	0.78
Yeast P complex [†]	Ligated exon [†]	5.4, 7.3	0.44	0.55

*Simulated density maps.

[†]Blind predictions.

Supplementary Table 3. Local resolution and RMSD accuracy for DRRAFTER models based on high-resolution maps

System		Reported, Local Map Resolution (Å)	Best, Centroid, Mean (standard deviation) RMSD of top 10 scoring models (Å)	Convergence Estimate (Å)
<i>E. Coli</i> L25-5S rRNA (1dfu) [*]		3.0, 3.0	0.3, 0.3, 0.3 (0.4)	0.4
Sex-lethal RRM (1b7f) [*]		3.0, 3.0	1.6, 4.1, 4.3 (1.1)	4.8
Ribotoxin restrictocin – SRL analog (1jbs) [*]		3.0, 3.0	1.3, 2.6, 2.2 (0.5)	1.8
SmpB-tmRNA complex (1p6v) [*]		3.0, 3.0	2.4, 2.6, 3.0 (0.4)	2.8
HutP antitermination complex (1wpu) [*]		3.0, 3.0	0.6, 0.6, 0.6 (0.2)	0.2
mRNA binding domain of SelB elongation factor (1wsu) [*]		3.0, 3.0	0.7, 1.3, 1.4 (0.4)	1.6
NusA transcriptional regulator (2asb) [*]		3.0, 3.0	1.8, 2.1, 3.0 (0.6)	3.7
Methyltransferase RumA in complex with rRNA (2bh2) [*]		3.0, 3.0	3.2, 3.4, 4.2 (1.0)	4.9
PP7 coat protein and viral RNA (2qux) [*]		3.0, 3.0	1.4, 2.8, 2.3 (0.7)	2.6
Puf4 bound to 3' UTR of target transcript (3bx2) [*]		3.0, 3.0	0.4, 0.5, 1.4 (0.9)	2.0
U4/U6.U5 tri-snRNP	U4/U6 3WJ	3.7, 4.1	1.9, 2.0, 2.3 (0.3)	2.3
	U5 3WJ	3.7, 4.7	3.0, 3.0, 4.1 (1.4)	4.4
	U5 IL II	3.7, 5.7	3.9, 5.9, 5.9 (0.7)	5.8
Mitochondrial ribosome	Loop 1	3.4, 3.6	1.7, 1.7, 2.6 (0.7)	3.2
	Loop 2	3.4, 3.6	1.4, 1.6, 1.8 (0.1)	1.4
CRISPR-Cas9	Full RNA	2.9, 2.9 [†]	2.4, 2.5, 2.7 (0.2)	3.1
Overall (average)			1.8, 2.3, 2.6	2.8

^{*} Simulated density maps.

[†] Crystallographic density map, local resolution was not calculated.

Supplementary Table 4. Predicted and actual accuracy (compared to the 4.8 Å coordinates [1]) of *Tetrahymena* telomerase DRRAFTER models.

Residues		Convergence (Å)	Predicted mean RMSD (Å)	Actual mean RMSD (Å) (compared to 4.8 Å coordinates)	Manually built model RMSD (Å)
Regions treated as rigid bodies	14-19, 37-40 (template boundary)	1.2	1.3	3.0	5.8
	69-99 (pseudoknot)	2.0	2.1	3.4	7.4
	156-159	3.0	3.1	13.6	7.3
	112-116, 150-154 (part of SL-4)	3.3	3.4	3.9	4.5
	128-142 (part of SL-4)	1.6	1.8	4.8	5.8
	119-125, 144-148 (part of SL-4)	1.5	1.6	4.7	2.8
	4-8, 103-107 (S1)	4.2	4.3	6.6	3.0
	21-35 (SL-2)	1.3	1.4	6.1	4.8
	43-51 (template)	4.6	4.7	12.3	12.5
Modeled <i>de novo</i>	Remaining residues not listed above	9.0	8.9	11.0	8.8
Overall (all residues)		5.2	5.2	7.8	7.2

References

1. Jiang, J., et al., Structure of Telomerase with Telomeric DNA. *Cell*, 2018. 173(5): p. 1179-1190 e13.