Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs

Sander Granneman, Grzegorz Kudla, Elisabeth Petfalski, and David Tollervey

Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland

Edited by Christine Guthrie, University of California, San Francisco, CA, and approved April 17, 2009 (received for review February 25, 2009)

The U3 small nucleolar ribonucleoprotein (snoRNP) plays an essential role in ribosome biogenesis but, like many RNA–protein complexes, its architecture is poorly understood. To address this problem, binding sites for the snoRNP proteins Nop1, Nop56, Nop58, and Rrp9 were mapped by UV cross-linking and analysis of cDNAs. Cross-linked protein–RNA complexes were purified under highly-denaturing conditions, ensuring that only direct interactions were detected. Recovered RNA fragments were amplified after linker ligation and cDNA synthesis. Cross-linking was successfully performed either in vitro on purified complexes or in vivo in living cells. Cross-linking sites were precisely mapped either by Sanger sequencing of multiple cloned fragments or direct, high-throughput Solexa sequencing. Analysis of RNAs associated with the snoRNP proteins revealed remarkably high signal-to-noise ratios and identified specific binding sites for each of these proteins on the U3 RNA. The results were consistent with previous data, demonstrating the reliability of the method, but also provided insights into the architecture of the U3 snoRNP. The snoRNP proteins were also cross-linked to pre-rRNA fragments, with preferential association at known sites of box C/D snoRNA function. This finding demonstrates that the snoRNP proteins directly contact the pre-rRNA substrate, suggesting roles in snoRNA recruitment. The techniques reported here should be widely applicable to analyses of RNA–protein interactions.

Author contributions: S.G. and D.T. designed research; S.G. and E.P. performed research; S.G. and G.K. contributed new reagents/analytic tools; S.G., G.K., and D.T. analyzed data; and S.G., G.K., and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. E-mail: d.tollervey@ed.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901997106/DCSupplemental.
snoRNA copurified with Rrp9-HTP, significantly above background levels (Fig. 2B, nickel, eluates).

Quantification of several Rrp9 cross-linking and analysis of cDNAs (CRAC) experiments revealed that 6.4% (H11001/H110021.7%) of Rrp9-HTP and 0.2% (H11001/H110020.063%) of the U3 snoRNA was recovered in the nickel eluates, indicating that 3.1% of U3 snoRNA was UV-cross-linked to Rrp9.

Fig. 1. The CRAC technique. (A) Purification of protein–RNA complexes. Cells were UV-irradiated in Petri dishes on ice. Extracts were incubated with IgG beads and tagged proteins were released by TEV protease cleavage. Cross-linked complexes were purified via nickel affinity purification under denaturing conditions. Purified proteins were detected by Western analysis and cross-linked RNAs were detected by Northern analysis. (B) Schematic representation of a protein fused to either the HTP tag (Upper) or the TAP tag (Lower). Prot A: Staphylococcus aureus Protein A IgG binding domain. (C) Identification of RNA binding sites. Partially Rnase-digested RNPs were incubated with nickel beads to immobilize His-tagged proteins (blue ovals) and covalently attached RNAs (red lines). Cross-linked RNAs were 3’ dephosphorylated, ligated to the adenylated linker (blue line), radioactively-labeled with polynucleotide kinase, and then ligated to the 5’ linker (green line). After release by imidazole treatment, radioactive RNPs were resolved on Bis-Tris NuPAGE gels and transferred to nitrocellulose. Bands corresponding to the predicted Mr of the target protein were excised and digested with proteinase K, and recovered RNAs were amplified by RT/PCR. The PCR products were gel-purified and sequenced. ddC: dideoxy-cytidine. InvddT: inverted dideoxythymidine. The asterisk indicates the UV cross-linking site.

Fig. 2. Mapping Rrp9 cross-linking sites. (A) Rrp9-HTP is specifically recovered on nickel beads. Extracts from cells expressing HTP or TAP-tagged Rrp9 were purified as in Fig. 1A. Five percent of the TEV eluates (5% Input) and the nickel eluates (Nickel eluates) were resolved on 4–12% Bis-Tris NuPAGE gels and detected by Western analysis. (B) Rrp9-HTP is cross-linked to U3. RNA extracted from 2% of the TEV eluates (2% Input) and nickel eluates (Nickel eluates) was analyzed by Northern analysis. (C) Rrp9 UV cross-links to the U3 snoRNA in vivo. Rrp9-HTP (lanes 1 and 3–6) or Rrp9-TAP (lane 2) were purified as shown in Fig. 1C. UV cross-linking was performed in vitro (lanes 1 and 2) or in vivo (lanes 3–6) and cross-linked U3 was detected by Northern analysis. The UV dose (J/cm2) is indicated above each lane. (D) Protein is cross-linked to radiolabeled RNA. CRAC was performed with strains expressing Rrp9-HTP with (lane 2) or without (lane 1) RNase treatment. (E) Contaminant proteins are not associated with RNA. CRAC was performed with strains expressing Rrp9-HTP with (lane 2) or without (lane 1) UV cross-linking. The asterisk indicates frequently-detected contaminants. Dashed red boxes indicate regions from which cross-linked RNA was extracted. (F) Multiple sequence alignment for the major Rrp9 binding sites. The black box indicates where deletions were frequently identified. The dashed red box indicates 2 primer extension stops detected after cross-linking (Fig. S1). (G) Histogram displaying locations of Rrp9-associated RNA fragments mapped to the U3 snoRNA (x axis). Percentage (y axis) is the number of reads mapped to that nucleotide divided by the total of U3 reads.
Cross-linking was also performed in vivo by UV-irradiating intact yeast cells in suspension in a Petri dish on ice. A time course of UV exposure revealed that ~4-fold more irradiation of yeast cells was required to replicate the in vitro cross-linking efficiency (Fig. 2C). These results demonstrate that cross-linking of Rrp9 to the U3 snoRNA can readily be detected in vitro and in vivo.

RNA binding sites were mapped by cloning and sequencing (Fig. 1C). To reduce the size of cross-linked RNAs, TEV eluates were partially digested with RNase A + T1. The bound proteins should largely protect their RNA binding sites, yielding small fragments containing the cross-linking sites. Guanidine hydrochloride was subsequently added to 6 M to inactivate the RNAses and disrupt the RNP particles. His₆-tagged proteins and covalently-attached RNA fragments were immobilized on nickel resin and extensively washed to remove the guanidine. Cross-linked RNA sequences were dephosphorylated with alkaline phosphatase to remove terminal 2' and 3' phosphates resulting from RNase cleavage and ligated on-bead to the 5' linker. RNAs were 5'-phosphorylated by T4 polynucleotide kinase in the presence of [γ-3²P]ATP, followed by ligation of the 5' linker. Both linker ligation reactions were performed on the nickel beads (Fig. 1C), which reduced the need for RNA gel purification steps, decreased recovery of linker multimers, and virtually eliminated cloning of contaminating bacterial rRNA, which is generally present in commercial preparations of recombinant proteins.

Proteins, together with attached radiolabeled RNA fragments flanked by linkers, were eluted from the nickel beads, trichloroacetic acid-precipitated, resolved on Bis-Tris NuPAGE gels, transferred to nitrocellulose membranes, and visualized by autoradiography. Analyses of strains expressing HTP-tagged Rrp9 revealed a radioactive band in the gel that migrated near the expected molecular mass of Rrp9 after either in vitro or in vivo cross-linking (Fig. 2D, lane 2). Without prior RNase digestion we observed a smear in the top half of the gel (Fig. 2D, compare lane 2 with lane 1), indicating that the radioactive bands represent Rrp9 cross-linked to RNA. Radiolabeled ~55-kDa bands were detected in many samples, including nontagged and noncross-linked negative controls (asterisk in Fig. 2E, lane 1 and lane 2), and appear to be nonspecific.

The linkers add ~10 kDa to the mass of the cross-linked protein–RNA complex, so we excised radiolabeled bands from membranes that migrated with and above the free primer (Fig. 2E, lane 3). Membrane slices were incubated with proteinase K to release cross-linked RNAs, which were then amplified by RT-PCR using linker-specific primers. To minimize recovery of primer dimers, PCR products >60 bp were gel-purified. The 3' linkers used were 5' adenylated, 3' blocked (dideoxyctydine) DNA oligonucleotides, which can be efficiently ligated by T4 RNA ligase in the absence of ATP (10). Under these conditions, only the adenylated DNA oligonucleotide can be ligated to RNA, greatly reducing the background. We initially used the published RLS RNA oligonucleotide (7) as the 5' linker. However, we often observed concatamerization of the RLS linker in cloned fragments, reducing the number of relevant clones. We therefore designed a DNA–RNA hybrid 5' linker that contains an inverted dideoxyctydime at the 5' end that completely blocked 5' linker concatamerization.

For Sanger sequence analyses, gel-purified fragments were cloned in the pCR4TOPO vector and transformed into E. coli and individual clones were sequenced. The histogram in Fig. 2G shows the distribution of cloned U3 fragments cross-linked to Rrp9-HTP along the U3a gene. Among the sequenced clones, 67% mapped to U3 (n = 68). The rest were apparently random rRNA fragments. An Rrp9 binding site at the interface between helix 2 and helix 4 of U3 (residues 193–206) was found in >70% of U3 clones (Figs. 2G and 3). A second Rrp9 binding site was identified in helix 4 of the U3 snoRNA. We frequently found small deletions in the center of these sequences (Fig. 2F), presumably reflecting errors in reverse transcription at the nucleotide that is the site of protein–RNA cross-linking (6, 7). These results are in good agreement with previous predictions of an Rrp9 binding site near the box B/C motif (8, 9, 11).

Primers were also performed on RNA extracted from nickel eluates of the Rrp9-HTP strain, as an alternative method to map UV cross-linking sites on U3 (4) (Fig. S1 and SI Text). Although the intensities of the signals were always low, 2 primer extension stops were reproducibly enriched in Rrp9 nickel eluates (U198-G199), which were located near the center of the region where deletions were frequently found in cloned U3 sequences (Fig. 2F). Thus, the primer extension analysis confirmed the major Rrp9 binding site identified in the CRAC analysis.

We next performed CRAC on each of the common box C/D snoRNP proteins except Smn13, which could not be HTP-tagged, likely because the tag interfered with the normal function of the protein. The efficiency of purification of HTP-tagged Nop1, Nop56, and Nop58 was similar to that of Rrp9-HTP (Fig. 4A, lanes 1–8), and all were cross-linked to RNA in vivo (Fig. 4A, lanes 9–12). Sanger sequencing of ~50 clones obtained from each of several independent CRAC experiments confirmed cross-linking of the common box C/D snoRNP proteins to box C/D snoRNAs in vivo and in vitro, ~5% of which were U3 hits.

Cross-Linking and Deep Sequencing: Common Box C/D snoRNP Proteins Cross-Link to Specific Sites in the U3 snoRNA. To increase the number of U3 sequences available for analysis, we modified the
CRAC protocol by using linkers that are compatible with Illumina Solexa deep sequencing. In vivo CRAC results are summarized in Table S1. Analyses of the deep sequencing data revealed frequent deletions and substitutions at specific locations, which likely represent the cross-linked nucleotides. To map the reads we used the Novoalign program (www.novocraft.com), which performs gapped alignments and reports mutations in single end reads. We aligned the reads against the entire yeast genome and a yeast noncoding RNA database (see SI Text).

Between 4.6 million and 8.6 million sequence reads were obtained for each library, of which 1.5 million to 5.5 million could be mapped to the yeast genomic sequence (Table S1). Between 74% and 90% of mapped sequences recovered with Nop1, Nop56, and Nop58 were derived from box C/D snoRNAs, whereas only ~1% corresponded to box H/ACA snoRNA sequences (Table S1). Box C/D snoRNAs represented only 0.5% of sequences recovered by using a nontagged control strain (Table S1), demonstrating the sensitivity and specificity of the CRAC method. The U3 snoRNA results are discussed below. Analyses of other box C/D snoRNAs will be presented elsewhere.

Nop1, Nop56, and Nop58 were primarily cross-linked to the 3′ end of the U3 snoRNA, near the conserved box D motif (Fig. 4B). However, approximately a quarter of the U3 sequences cross-linked to Nop58 were mapped to the 5′ end of U3 in the 5′ hinge and a small fraction also included the C′ box (Fig. 4B). The average length of the mapped sequenced fragments (not including linkers) was between 22 and 34 nt (Table S1), which is expected to increase the apparent overlap between individual peaks. To better localize the binding sites of the box C/D snoRNPs in the 3′ end of U3 we analyzed only reads between 15 and 18 nt in length, which is long enough to identify unique genomic sites (Fig. 4C). The resulting graph revealed 2 major peaks for Nop1, one in helix 3 and another near the 3′ end of the RNA. Nop56 primarily cross-linked to helix 3, whereas Nop58 almost exclusively bound at box D in U3.

Sanger sequencing of cloned CRAC products had indicated that mutations and deletions were indicative of the actual site of cross-linking. We therefore mapped the distribution of U3 deletions and substitutions in the deep sequencing data. Strikingly, 48% of the U3 sequences cross-linked to Nop58 contained substitutions of A322 and/or G323 (Fig. 4D, and 18% of the sequences contained deletions of A322 and/or G323 (Fig. 4E). In contrast, mutations at these positions were rarely found for Nop1 or Nop56 (Fig. 4D and E), demonstrating that these mutations were specifically connected to Nop58. The 5′ domain of U3 RNA was also cross-linked to Nop58, and 90% of these sequences contained substitutions at C39. We conclude that Nop58 directly binds the U3 snoRNA at G323 and U324 in stem II adjacent to box D and at C39 in the 5′ domain of U3 (see Fig. 3).
The U3 snoRNA is encoded by 2 genes, the products of which differ at a few positions (12). One site of difference is located in a bulge in helix 3 around nucleotide \( H_300 \) (Table S2), because 10% of the U3 sequences mapped to the 3' end had deletions here. The Nop56 and Nop58 CRAC data rarely included deletions in this region (Fig. 4E), suggesting this is a specific binding site for Nop1. Substitutions and deletions were identified at the same positions by Sanger sequencing of cDNA clones from multiple independent experiments (Table S2), demonstrating that they are not a consequence of Illumina Solexa sequencing errors.

Collectively, these data indicate that Nop1, Nop56, and Nop58 bind at specific positions in the 3' region of the U3 snoRNA and that Nop58 also contacts the 5' domain of the U3 snoRNA (see Fig. 3).

The cross-linking of Nop58 to the 5' domain of U3, which is involved in pre-rRNA base-pairing interactions, prompted us to analyze cross-linking of the common box C/D snoRNP proteins to the pre-rRNA (Fig. 5). Relative to a control dataset (Fig. 5D), the snoRNP proteins recovered more hits in the 5' external transcribed spacer (ETS) region of the pre-rRNA. In each case there was a peak around the known U3 binding site at +470 (Fig. 5, blue line ~500) (13). A substantial number of hits were also detected at the U14 base-pairing site near the 5' end of 18S (Fig. 5, blue line ~750), an interaction that is essential for 18S rRNA synthesis (14). The sites of box C/D snoRNP protein cross-linking to the rRNA sequence was compared with the distribution of known sites of snoRNA-directed 2'-O-methylation (Figs. S2 and S3 and Table S3). As expected, the methyl-transferase Nop1 most significantly bound the pre-rRNA close to rRNA methylation sites. Approximately 60% of Nop1-cross-linked reads in 18S rRNA and 65% of reads in 25S rRNA were located within 20 nt of a methylation site, whereas \( \approx 32\% \) would be expected if the reads were randomly distributed over the rRNA (Table S3). Nop58 reads were also significantly enriched close to methylation sites in both 18S and 25S (Table S3). In contrast, Nop56 significantly cross-linked close to methylation sites in 18S but not in the 25S rRNA. We conclude that common box C/D snoRNP proteins not only interact with the snoRNAs but also directly contact the RNA substrates.

Discussion

Mapping Protein–RNA Binding Sites. The studies reported here show that CRAC can be used in vitro and in vivo to pinpoint protein–RNA interaction sites in the U3 snoRNA and pre-rRNA. In general, it may be assumed that the in vivo cross-linking will more faithfully reflect the genuine protein–RNA interactions, “in vivo veritas.” However, the available data on RNP composition, with which the cross-linking data might be integrated, was largely obtained on complexes analyzed in vitro. Sanger sequencing of individual clones and deep sequencing each have their advantages. For many RNPs a small number of sequences will be enough to clearly identify the binding sites, especially when it is evident that the protein of interest primarily cross-links to 1 site on the RNA. However, the common box C/D snoRNP proteins bind to 47 snoRNAs and, as shown here, to the pre-rRNA, so greater depth of coverage was required to increase the confidence that all significant binding sites had been identified. The deep sequencing data were challenging to analyze and required the development of software tools to handle the large datasets. We are in the process of setting up a publicly-accessible, Galaxy-based web server to provide the tools for analyses of CRAC datasets.

In both Sanger sequencing and deep sequencing analyses, we observed sites at which nucleotide substitutions and deletions were repeatedly identified, generally with 1 specific protein. CLIP analyses of the mouse RNA binding protein Nova also yielded RNA fragments containing nucleotide substitutions in the Nova YCAY

**Fig. 5.** snoRNP proteins directly bind the pre-rRNA. (A–C) Binding sites for Nop1, Nop56, and Nop58 across the 5' region of the pre-rRNA. (D) CRAC results for the untagged control strain. Red lines indicate sites of snoRNA-directed 2'-O-methylation. Blue lines indicate the site U3 base-pairing in the 5' ETS and the U14 base-pairing in the 18S rRNA, which are required for pre-rRNA processing. Hits on the complete 35S pre-rRNA region are presented in Figs. S2 and S3.
Locations of Core Protein Binding Sites in the U3 snoRNA. Two Rp9 binding sites were mapped in U3, adjacent to the box B/C motif in helices 2 and 4. This location is consistent with previous studies (8, 9, 11), and the major Rp9 cross-linking site was confirmed by primer extension. The archaeal dual guide box C/D snoRNA architecture is symmetric with 2 copies each of the orthologues of Nop1, Nop56/58, and Smu13 (15). In U3, 2 binding sites had been identified for Smu13 (8, 16), suggesting that these might associate with 2 copies of Nop1 and single copies of Nop56 and Nop58. However, the stoichiometry of the snoRNP proteins and their exact binding sites were unclear. The cross-linking data indicate that Nop1 has at least 2 binding sites in the 3′ domain of U3; within helix 3 and close to box D. Consistent with binding near the 3′ end, mutations in Nop1 can alter the site of 3′ end formation of box C/D snoRNAs (17). Specific binding sites for Nop56 and Nop58 in the 3′ domain of U3 were clearly distinguishable. Nop56 mainly cross-linked to helix 3, whereas Nop58 primarily bound the 5′ end of U3 close to the highly-conserved box C/D stem II. Unexpectedly, Nop58 also cross-linked to the 5′ hinge region of the U3, which base-pairs with the 5′ ETS at position 470 on the pre-rRNA (18).

The mechanisms by which the numerous snoRNA find their specific binding sites within the very large and complex preribosomes remain unclear. Analyses of cross-linking between the snoRNP proteins and the pre-rRNA showed significant enrichment for sequences close to sites of snoRNA-directed RNA methylation, consistent with their association with the methylation-guide box C/D snoRNAs. Relative to the nontagged cross-linking analysis, Nop1, Nop56, and Nop58 each showed clearly increased association with the 5′ ETS region of the pre-rRNA, which is not methylated but is bound by U3. In each case, there was a substantial signal in the region around the U3-binding site at +470. Cross-links were also found over the 18S RNA region (+83–95) that base-pairs with domain A of U14, an interaction essential for pre-rRNA processing (19). These results demonstrate that Nop1, Nop56, and Nop58 each directly contact the RNA substrate at snoRNA binding sites, suggesting roles in promoting snoRNA–rRNA association and/or snoRNA-dependent changes in preribosome structure.

Materials and Methods

Strains and Media. Growth, handling, and transformation of yeast involved standard techniques. All strains were constructed in the background of BY4741. Yeast strains used are listed in Table S4.

CRAC Method and Bioinformatics Analyses. The technique is described in Fig. 1. A more detailed CRAC protocol and description of the Linux/Unix (Bash, Awk, and Perl) scripts used for sequence analyses is provided in SI Text. Oligonucleotide sequences are listed in Table S5.

Acknowledgments. We thank members of D.T.’s laboratory for critical reading of the manuscript and helpful discussions; Alastair Kerr and Shaun Webb for bioinformatics support; members of the Edinburgh Gene Pool Sequencing Facility for DNA sequencing, and the Swann Building kitchen staff for media preparation. This work was supported by the Welcome Trust (D.T.), European Molecular Biology Organization long-term fellowships (to S.G. and G.K.), a Marie Curie Intra-European fellowship (to G.K.), and Biotechnology and Biological Sciences Research Council Grant BB/D019621/1.