Structure-Sensitive RNA Footprinting of Yeast Nuclear Ribonuclease P†

Anthony J. Tranguch,‡ David W. Kindelberger,§ Christopher E. Rohlman,‖ Jae-Yong Lee,⊥ and David R. Engelke*

Department of Biological Chemistry, The University of Michigan, Ann Arbor, Michigan 48109-0606

Received October 6, 1993; Revised Manuscript Received November 23, 1993®

ABSTRACT: Several enzymatic and chemical reagents were used to probe the secondary structure of *Saccharomyces cerevisiae* nuclear RNase P RNA in the presence and absence of its protein components. Double-stranded regions were detected with RNase V1 and single-stranded regions with RNase ONE (*Escherichia coli* RNase I). Nucleotides not paired at Watson–Crick positions were monitored with dimethyl sulfate, kethoxal, and 1-cyclohexyl-3-[2-(N-methylmorpholinio)ethyl]carbodiimide *p*-toluenesulfonate. The results supported most aspects of the previously proposed, phylogenetically-derived RNA secondary structure, although minor refinements allowed incorporation of both the biochemical and phylogenetic data. Digestion of the RNase P protein(s) with proteinase K gave enhanced reactivities to structure probes at selected positions, indicating regions of the RNA made inaccessible by the presence of the protein subunit(s). The regions of RNA protected in the yeast nuclear holoenzyme were considerably more extensive than that seen in the *Escherichia coli* holoenzyme, consistent with the observation that the protein moiety generally comprises a larger percentage of the RNase P holoenzyme in eukaryotes than in eubacteria.

RNase P is an essential ribonucleoprotein enzyme that cleaves precursor tRNA (pre-tRNA)[1] molecules to produce mature 5′-termini (Altman, 1989). Unique among the known ribozymes, RNase P recognizes the conserved, higher-order structure of its substrates and so can bind and cleave numerous different pre-tRNAs of unrelated sequence. To understand the nature of this interaction, it is essential to examine the structure of the RNase P holoenzyme.

Although RNase P activity has been found in eubacteria, archaebacteria, and eukaryotes, the composition and features of these enzymes appear to be different (Darr et al., 1992). In eubacteria, where RNase P has been studied most extensively, the holoenzyme is composed of a single RNA subunit (~400 nt) and a single protein subunit (~14 kDa). The eubacterial RNA component alone was shown to be capable of substrate binding and catalysis *in vitro* (Guerrier-Takada et al., 1983). Subsequently, the RNA from *Escherichia coli* was probed with a variety of structure-specific enzymatic (Guerrier-Takada & Altman, 1984) and chemical reagents (Shiraishi & Shimura, 1988; Knap et al., 1990), and a conserved secondary structure for eubacterial RNase P RNA was derived by phylogenetic sequence comparisons (James et al., 1988). Despite the demonstration of RNA (alone)-mediated catalysis and the fact that RNA comprised ~90% (by density) of the holoenzyme (Akaboshi et al., 1980; Gardiner & Pace, 1980), the eubacterial protein subunit was nevertheless essential for RNase P activity *in vivo*, purportedly to assist with electrostatic shielding between the phosphate backbones (Reich et al., 1988). Comparison of the eubacterial RNase P proteins revealed conservation of basic and aromatic residues at a few positions (Brown & Pace, 1992);[2] however, no structure for any of the proteins has been proposed. Furthermore, the structure of the eubacterial holoenzyme has yet to be elucidated, although footprinting analysis has suggested some regions of interaction between the RNA and protein moieties (Vioque et al., 1988).[2]

In contrast to the situation for the eubacterial RNase P, the RNA components from the archaebacterial and eukaryotic enzymes have not yet exhibited RNA (alone)-mediated catalysis. In fact, it is currently unclear whether RNase P from chloroplasts contains an RNA subunit at all (Wang et al., 1988). Although the archaebacterial RNAs can theoretically conform to the phylogenetically-derived eubacterial structure, it remains uncertain why these RNAs do not catalyze the RNase P reaction *in vitro* (Nieuwlandt et al., 1991). Difficulties in deriving an unequivocal structure for the mitochondrial RNA subunits involve their extremely high (>85%) A+U sequence content (Morales et al., 1989). The structures of the nuclear RNase P RNAs have recently been shown to contain some regions of structural homology to the eubacterial RNAs (Forster & Altman, 1990; Tranguch & Engelke, 1993). Whereas the protein subunit comprises only a small percentage of the eubacterial RNase P, protein accounts for ~50% (by density) of the eukaryotic holoenzymes (Kline et al., 1981; Lawrence et al., 1987; Morales et al., 1989; Doria et al., 1991; Jayanthi & Van Tuyle, 1992) and at least one of the archaebacterial holoenzymes (Darr et al.,

† This work was supported by National Institutes of Health Grant NIH GM 34869 to D.R.E. Oligonucleotide synthesis was partially subsidized by National Institutes of Health Grant P30CA46592 to the University of Michigan Cancer Center. Computer assistance was facilitated by National Institutes of Health Grant M01 RR00042 to the University of Michigan Clinical Research Center.

* To whom correspondence should be addressed. Tel: 313-763-0641; Fax: 313-763-4581.

‡ Fellow in the Medical Scientist Training Program at the University of Michigan and supported by a Young Scientist M.D./Ph.D. Scholarship provided by the Life and Health Insurance Medical Research Fund.

§ Supported by NIH National Research Service Award 5-T32-GM07544-15 from the National Institute of General Medical Sciences.

‖ Present address: Department of Chemistry, Pomona College, Claremont, CA 91711-6333.

⊥ Present address: Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI 48109.

® Abstract published in *Advance ACS Abstracts,* January 15, 1994.

[1] Abbreviations: pre-tRNA, precursor tRNA; *RPR1*, nuclear RNase P RNA gene from *Saccharomyces cerevisiae*; YNE, yeast nuclear extract; bp, base pair; nt, nucleotide; RNase V1, ribonuclease from cobra *Naja naja oxiana* venom; DMS, dimethyl sulfate; KE, kethoxal, β-ethoxy-α-ketobutyraldehyde; CMCT, 1-cyclohexyl-3-[2-(*N*-methylmorpholinioethyl]carbodiimide *p*-toluenesulfonate; HEPES, *N*-(2-hydroxyethyl)-piperazine-*N*′-2-ethanesulfonic acid; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; $Tween_{20}$, polyoxyethylene(20)sorbitan monolaurate; SDS, sodium dodecyl sulfate.

[2] S. J. Talbot and S. Altman, personal communication.

0006-2960/94/0433-1778$04.50/0

1990). To date, the only noneubacterial RNase P protein that has been isolated is a single, 105-kDa polypeptide from *Saccharomyces cerevisiae* mitochondrial RNase P (Morales et al., 1992; Dang & Martin, 1993); this protein is considerably larger than the eubacterial subunit (14 kDa). It is possible that the noneubacterial RNase P protein subunits may have assumed some or all of this enzyme's substrate binding and catalytic properties. However, it has been demonstrated biochemically (Doersen et al., 1985; Krupp et al., 1986; Lee & Engelke, 1989) and genetically (Miller & Martin, 1983; Cherayil et al., 1987; Lee et al., 1991) that these RNA components are functionally indispensible. This further underscores the need to study the structure of the RNase P RNA in the context of the holoenzyme.

Our laboratory had previously identified the RNA component of *S. cerevisiae* nuclear RNase P (Lee & Engelke, 1989), characterized its gene (*RPR1*), and demonstrated through defective alleles that it was an essential component of nuclear RNase P (Lee et al., 1991). We subsequently characterized the RNase P RNAs and their genes from six *Saccharomyces* species and derived secondary structure models for the RNAs via comparative sequence analysis (Tranguch & Engelke, 1993). A common core of primary and secondary structure emerged for the *Saccharomyces* RNase P RNAs, and the *Schizosaccharomyces* homologs conformed in large part to this consensus structure. Comparison of the yeast core to the eubacterial consensus structure and to the vertebrate homologs revealed a number of similarities, suggesting that RNase P RNA from diverse sources may share a core of structurally conserved elements.

In the current study, we tested the validity of the phylogenetically-derived RNA structure of *S. cerevisiae* nuclear RNase P and assessed which regions of the RNA might interact with its protein subunit(s). For these purposes, enzymatic and chemical reagents were employed to probe the structure of the RNA in the presence and absence of the RNase P protein component(s). RNase V1 was used to cleave double-stranded and stacked regions and RNase ONE (RNase I from *E. coli*) to cleave single-stranded regions. Chemicals that were used to modify unpaired nucleotide bases included: dimethyl sulfate (DMS) at the N1 position of adenines and N3 of cytosines; kethoxal (KE) at N1 and 2-NH_2 of guanines; and 1-cyclohexyl-3-[2-(*N*-methylmorpholinio)ethyl]carbodiimide *p*-toluenesulfonate (CMCT) at N3 of uracils and, rarely, N1 of guanines. The results strongly supported most aspects of the phylogenetic structure, although minor structural refinements were made to accommodate both the biochemical and phylogenetic data. Regions of the RNA potentially involved in protein interactions were deduced from the observation of enhanced reactivities to structure probes following deproteinization of the RNase P holoenzyme. An RNA footprint was proposed for the *S. cerevisiae* nuclear RNase P and compared to the footprint from the *E. coli* holoenzyme (Vioque et al., 1988).[2]

EXPERIMENTAL PROCEDURES

Buffers. Buffer A: 20 mM Na-HEPES, pH 7.9, 10% glycerol, 8 mM $MgCl_2$, 1 mM DTT, and 1% $Tween_{20}$. Buffer B: 10 mM Tris, pH 7.8, 8 mM $MgCl_2$, and 100 mM NaCl. Buffer C: 10 mM Tris, pH 7.8, and 0.1 mM EDTA. Buffer D: 50 mM Tris, 10 mM EDTA, 0.5% SDS, and 100 mM sodium acetate, pH 4.8. Buffer E: 300 mM sodium acetate, pH 4.8, 2.5 mM EDTA, and 1 μg/μL glycogen. Buffer F: 50 mM K-HEPES, pH 7.8, 8 mM $MgCl_2$, and 100 mM NaCl. Buffer G: buffer F with K-borate, pH 8.0, replacing K-HEPES. Buffer H: 50 mM K-HEPES, pH 7.8, 100 mM NaCl, and 1.25 μg/μL glycogen. Buffer I: buffer H with K-borate, pH 7.0, replacing K-HEPES. Buffer J: 50 mM K-HEPES, 2.5 mM EDTA, 0.25% SDS, and 300 mM sodium acetate, pH 4.8. Buffer K: buffer J with K-borate replacing K-HEPES. Buffer L: 34 mM Tris, pH 8.3, 40 mM KCl, 6 mM $MgCl_2$, and 2 mM DTT.

Preparation of RNase P. As a source of nuclear RNase P holoenzyme, a high-salt extract of nuclei from 12 L of *S. cerevisiae* strain PP1002 (Piper & Stråby, 1989; gift of Kerstin Stråby) was prepared as described previously (Evans & Engelke, 1990), with minor modifications. The $MgCl_2$ concentration was maintained at 8 mM in all buffers throughout the preparation, and the final pellet was resuspended in 25 mL of buffer A, aliquoted, and stored at −80 °C prior to use. Typically, 5 μL of this yeast nuclear extract (YNE) was used per 50-μL final reaction volume. The concentrations of nucleic acids and proteins were determined to be approximately 32.5 and 100 μg/5 μL of YNE, respectively, as calculated from absorbances at 260 and 280 nm. In samples subjected to deproteinization, YNE was preincubated with proteinase K (500 μg/mL final concentration) at 25 °C for 30 min prior to enzymatic or chemical probing. This level of protein digestion was found to cause the disappearance of all visible Coomassie blue-stained bands (>19 kDa) on 10% denaturing polyacrylamide gels (data not shown).

Structure-Specific Enzymatic Probing. Digestions with RNase V1 (Pharmacia) or RNase ONE (RNase I from *E. coli*; Promega) were performed at 25 °C for 20 min in buffer B. Nucleases were diluted with buffer C to produce 0.090, 0.225, and 0.450 units/μL RNase V1 or 1.50, 3.75, and 7.50 units/μL RNase ONE. Dilutions were made 15 min prior to addition and kept on ice until use. For each enzyme, a control was treated in parallel, with omission of the nuclease probe. To begin the digestions, 1 μL of reagent was added per 50 μL of reaction and mixed gently by pipetting. Reactions were terminated by the addition of an equal volume of buffer D. Following two extractions with phenol equilibrated in 0.1 M Tris, pH 8.0, and one extraction with chloroform, the nucleic acids were precipitated in buffer E and 70% ethanol. Pellets were washed, dried, resuspended in water, and stored at −80 °C until subjected to primer extension analysis. Typically, one 50-μL probing reaction (5 μL of YNE) provided enough RNA for a single primer extension reaction. Probing reactions were performed on proportionally larger scale for extensions with multiple primers.

Chemical Probing. Chemical modification reactions were similar to those described previously (Stern et al., 1988). Modifications were performed at 25 °C for 20 min in buffer F for DMS (Kodak) or KE (Upjohn), or in buffer G for CMCT (Aldrich). Dilutions of the reagents were made 15 min prior to addition (5 min for DMS) as follows: DMS was diluted with 100% ethanol to give final reaction concentrations of 8.4, 21.1, and 42.2 mM (1 μL of diluted reagent/50 μL of reaction); KE with water to give 5.7, 22.8, and 57.0 mM (1 μL/50 μL); CMCT with buffer G to give 25, 100, and 250 mM (25 μL/50 μL). (It was necessary to briefly incubate the 500 mM CMCT stock solution at 37 °C prior to addition to prevent precipitation.) For each chemical reagent, a control was treated in parallel, with omission of the chemical probe. Immediately following addition of the chemical reagent, samples were mixed gently by pipetting. Reactions were terminated by the addition of 200 μL of stop buffer (buffer H for DMS and CMCT,

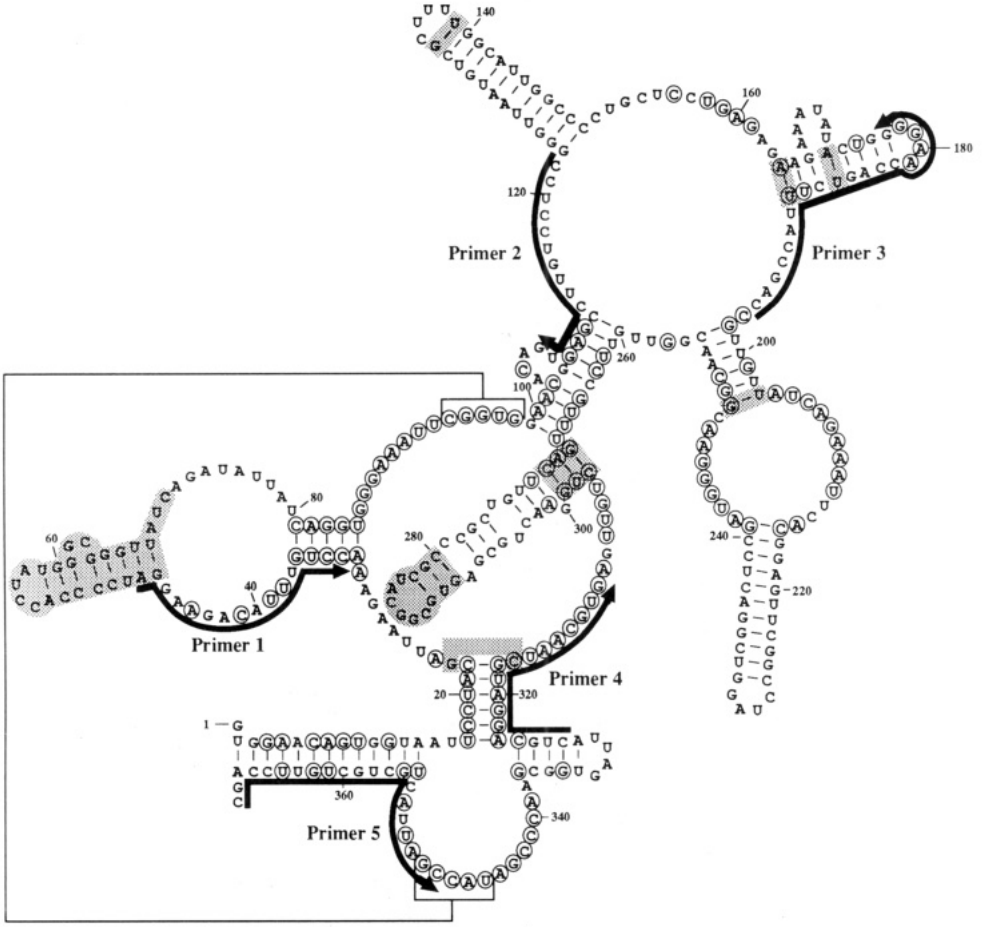

FIGURE 1: Previous secondary structure model for *S. cerevisiae* nuclear RNase P RNA based on comparative sequence analysis (Tranguch & Engelke, 1993). Nucleotides found to be invariant among the *Saccharomyces* nuclear RNAs are circled. Nucleotide positions are numbered every 20 nt. Deoxyoligonucleotide primers used for primer extension analysis in conjunction with structure-specific probes are depicted as curved arrows next to their complementary sites. Shaded regions of the model were refined subsequent to direct structure probing experiments (cf. Figure 3).

buffer I for KE) and 1125 μL of 100% ethanol, followed immediately by precipitation on dry ice. The pellets were washed, dried, and resuspended in 200 μL of buffer J (buffer K for KE samples). Following two extractions with water-saturated phenol and one extraction with chloroform, the nucleic acids were precipitated in 70% ethanol. Pellets were washed, dried, resuspended in water (25 mM K-borate, pH 7.0, for KE samples), and stored at −80 °C until subjected to primer extension analysis. Similar to the enzymatic probing, the chemical modifications were scaled up proportionally, performed in batch, and aliquoted prior to primer extensions.

Primers. Deoxyoligonucleotide primers specifically hybridizing to the mature RNA of *RPR1* were synthesized by the DNA Synthesis Facility/University of Michigan Biomedical Research Core. The sequence of the primers and the complementary positions in the RNA were as follows:

primer 1: 5′-d(TCCTTCTGTAAACAGG)-3′ (33–48)

primer 2: 5′-d(GGAGGACAAGGCTCCAC)-3′ (106–122)

primer 3: 5′-d(GTCGGTAAAGACTGGTTCCC)-3′ (177–196)

primer 4: 5′-d(GACGTCCTACGATTGCAC)-3′ (310–327)

primer 5: 5′-d(GCTGGAACAGCAGCAGTAATCG)-3′ (348–369)

Primers were 5′-end-labeled with [γ-^{32}P]ATP (New England Nuclear) and T4 polynucleotide kinase (Gibco/BRL) and purified by electrophoretic separation in 15% denaturing polyacrylamide gels (Lee et al., 1991).

Primer Extension Analysis. Primers were individually annealed (45 °C, 30 min, buffer L) to RNA isolated from the enzymatic and chemical probing experiments and extended (45 °C, 25 min) with avian myeloblastosis virus reverse transcriptase (Life Sciences). The extensions were terminated by precipitating in buffer E (minus glycogen for the chemically-modified samples) and 70% ethanol; otherwise, the reactions were performed as described previously (Lee et al., 1991). The labeled extension products were resolved by electrophoresis in 6% polyacrylamide/8 M urea/1× TBE gels (Sambrook et al., 1989). Gels were then dried and exposed to Hyperfilm-MP (Amersham Corp.). Positions of the primer extension stops ("hits") were determined by direct comparison with dideoxynucleotide chain-termination sequencing (Tabor & Richardson, 1987) of an *RPR1* genomic clone using the same labeled primers.

RESULTS AND DISCUSSION

Probing Strategy. The RNA from *S. cerevisiae* nuclear RNase P (*RPR1* RNA), both in the holoenzyme and after

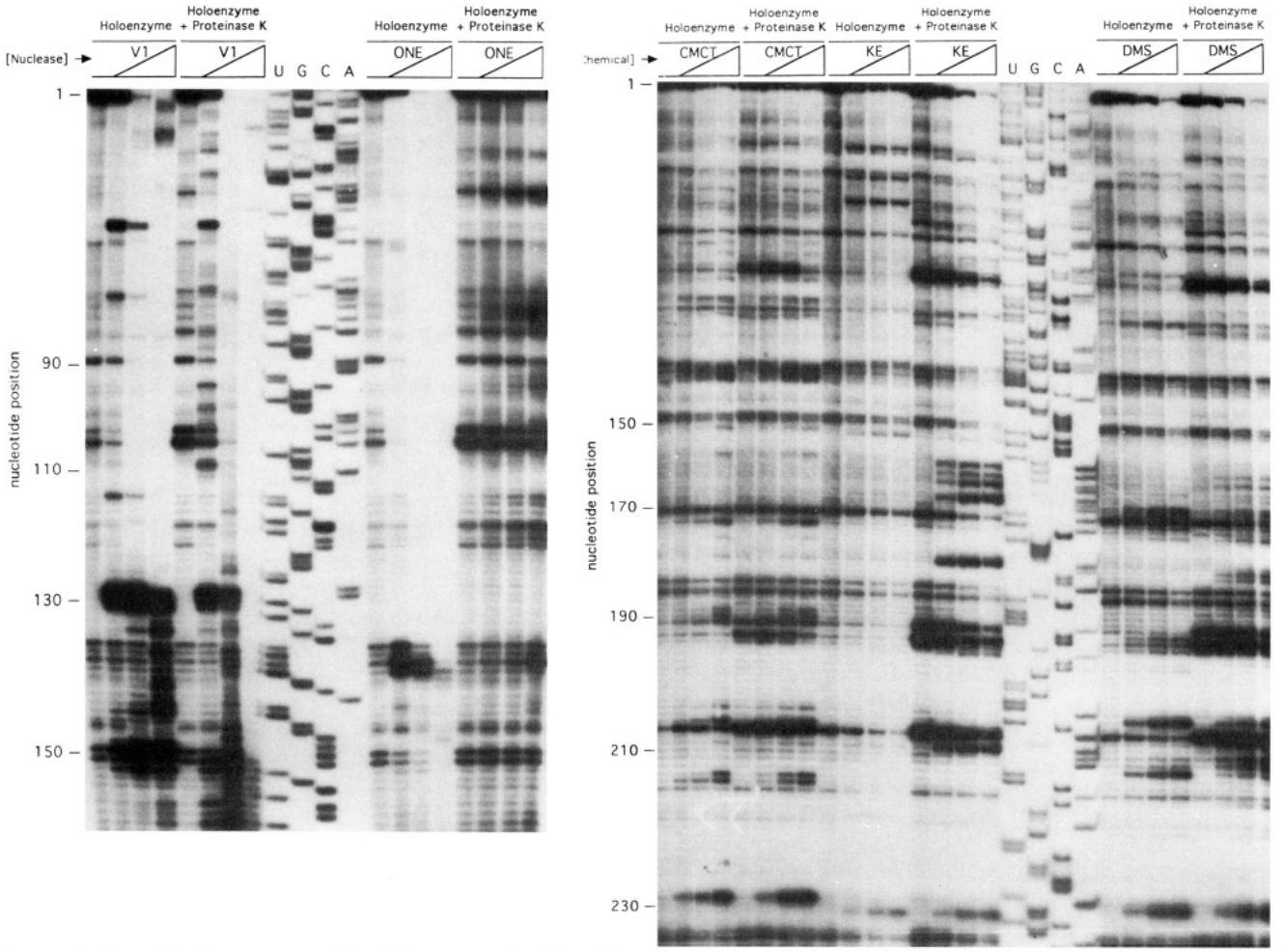

FIGURE 2: Examples of structure-probing/primer extension data. Direct structural probing of RNase P RNA in the presence ("Holoenzyme") or absence ("Holoenzyme + Proteinase K") of its protein subunit(s). Yeast nuclear extract (−/+ preincubation with proteinase K) was treated with a limiting amount of a structure-specific nuclease [left panel; V1 = RNase V1 (stacked- and double strand-specific), ONE = RNase I from *E. coli* (single strand-specific)] or chemical reagent [right panel; DMS (A- and C-specific; A > C), KE (G-specific), CMCT (U- and G-specific; U > G)]. Each treatment (indicated above the lanes) was performed with increasing amounts of the nuclease or chemical (see Experimental Procedures for details; left lanes represent mock incubations). Cleavage or modification sites were detected using primer extension analysis with one of five primers specific to the RNase P RNA (see Figure 1). For reference, the same primers were used in DNA sequencing reactions (lanes U, G, C, and A). Primer extension products were fractionated on 6% polyacrylamide/8 M urea gels. Typical autoradiograms are shown. (Left panel) Nuclease digestions; extensions with primer 3. (Right panel) Chemical modifications; extensions with primer 4. Numbers to the left of each panel represent nucleotide positions in the RNase P RNA.

deproteinization, was probed with structure-specific enzymatic and chemical reagents. Unlike its counterparts from eubacterial RNase P, *RPR1* RNA has not yet been shown to have catalytic activity in the absence of protein. Therefore, to be interpreted in terms of a functionally relevant structure, probing studies on *RPR1* RNA must be performed in the context of the holoenzyme. Conversely, potential sites of RNA–protein interaction may be identified by comparing accessibility to probes before and after deproteinization of the RNase P holoenzyme. To acquire both types of information, the following strategy was employed. Yeast nuclear extract was used as a source of RNase P holoenzyme in order to limit breakdown of the ribonucleoprotein complex, as is often seen upon further purification (Lee & Engelke, 1989). The holoenzyme or proteinase K-treated RNase P was subjected to limited digestion or modification with one of the following structure-specific probes: RNase V1, specific for double-stranded or stacked single-stranded regions (Lockard & Kumar, 1981; Lowman & Draper, 1986); RNase ONE, single-stranded regions (Meador et al., 1990); DMS, N1 position of adenines and N3 of cytosines (A > C); KE, N1 and 2-NH_2 of guanosines; and CMCT, N3 of uracils and N1 of guanosines (U > G). The RNA was then isolated, and sites of nuclease cleavage or chemical modification (i.e., "hits") were assessed by primer extension analysis (Stern et al., 1988).

Five primers were used in this study to span the entire length of the RNase P RNA. Positions to which these primers hybridized are shown in Figure 1 relative to the secondary structure model, previously proposed on the basis of phylogenetic analysis of yeast nuclear RNase P RNAs (Tranguch & Engelke, 1993). Refinements to this structure, based on results from the current study, are shown as shaded areas in this figure and are incorporated into the structure models displayed in subsequent figures. Note that the structure of the extreme 3′-end of the RNA (the final 26 nucleotides) could not be directly assessed due to the inherent limitations of the primer extension technique (Stern et al., 1988).

Representative data from the primer extensions of RNase P RNA from both holoenzyme and proteinase K-treated samples probed with structure-specific reagents are shown in Figure 2. Titrations of the nucleases (Figure 2, left panel) and chemical reagents (right panel) were routinely performed to determine the amount of each reagent required to produce less than one hit per RNA molecule, thereby reducing the possibility of induced structural rearrangements. In practice, lanes used to assign hits were chosen only if greater than 50%

of the RNA species appeared unmodified (Christiansen & Garrett, 1988). (Note that, for deproteinized samples, higher concentrations of RNase ONE were required to overcome residual proteinase K activity.) Titrating each probe also allowed observation of positions that were cleaved or modified in a concentration-dependent manner. Artifactual bands, presumably arising from nicks, posttranscriptional modifications, or strong structural features in the template RNA, were distinguished from sites attacked by structure probes by their occurrence in transcripts using unmodified control RNA, which had otherwise been subjected to identical treatment. Twenty positions in the RNase P RNA from the holoenzyme samples and eight additional positions in the proteinase K-treated samples (identified by small filled circles next to nucleotide positions in Figures 3 and 4, respectively) could not be evaluated due to these background primer extension stops. Hits were assigned by direct comparison with DNA sequencing lanes run in parallel. Enzymatic cleavages were mapped to the phosphodiester linkage 5′ of the primer extension stop, whereas chemical modifications were mapped to the nucleotide one position up (5′) in the corresponding dideoxy sequencing lane from the stop itself.

RNase P RNA Structure in the Holoenzyme. A total of 136 enzymatic and chemical hits were observed when the RNA was probed in the context of the RNase P holoenzyme. These hits are mapped onto a refined secondary structure model of the *RPR1* RNA, as shown in Figure 3. Minor refinements were made to the previous model (cf. Figure 1; Tranguch & Engelke, 1993) to allow incorporation of both biochemical and phylogenetic data. In the current model, only 9 of the 136 hits are potentially inconsistent with the structure as drawn. In the discussion that follows, structural refinements, inconsistent hits, and the general accessibility of the holoenzyme's RNA to probes will be analyzed.

Subsequent to probing, the stems found at nucleotide positions 47–71 and at 280–291 were reorganized to account for the observed biochemical data. Prior to rearrangement, the majority of hits falling within region 47–71 contradicted the previous structure. Following rearrangement of this region, all hits except for one within 47–71 are now consistent with the new structure (Figure 3). A weak DMS hit was observed at A48, suggesting that this nucleotide may not be involved in base-pairing (at least at its N1 position) and the proximal 3 base pairs (bp) of this stem may not form. However, an RNase V1 hit on the 3′-side of A72 suggests the potential presence of a nearby double-stranded region (see discussion below). Furthermore, inspection of the homologous region in *Saccharomyces* and *Schizosaccharomyces* RNase P RNAs revealed that the rearranged stem, including the proximal 3 bp portion, could be supported phylogenetically; i.e., similar rearrangements could be made in the other *Saccharomyces* RNAs. In the new structure, the stem–internal loop–hairpin stem domain found within positions 32–85 displays 10 nucleotides (nt) in the 5′-side of the internal loop (formerly 10–11 nt in the yeast RNAs) and 7–10 nt in the 3′-side (formerly 10–13 nt). Since maintenance of the sequence and structural homology depends upon formation of the proximal 3 base pairs in question, the phylogenetic evidence overrules the apparently contradictory DMS hit at A72 and argues for preservation of the structure as drawn in Figure 3.

A less dramatic but biochemically- and phylogenetically-consistent structural rearrangement was made to the 280–291 portion of stem 272–299. Although only one hit (CMCT at U290) was inconsistent with the former structure, it spurred a rearrangement in which the alignment of this stem between the RNAs from *S. cerevisiae* and *Saccharomyces carlsbergensis* became clear. The hairpin loop now contains 4 nt (UACG in *S. cerevisiae*, CACG in *S. carlsbergensis*), and the 1 bp mismatch displays a 2 nt insertion (291–292) on the 3′-side in *S. cerevisiae*; otherwise, the sequences and structures within positions 277–294 are identical in both of these yeast RNAs.

Other changes made to the structural model consisted primarily of breaking single base pairs throughout the molecule on the basis of biochemical data, phylogenetic arguments, or both. The base pair between G134 and U139 (134/139 bp) was broken due to moderate chemical hits on both members of the pair and to lack of phylogenetic support at this position. Although a weak chemical hit was seen on only one member of bp 203/250, this pair was broken because phylogenetic covariations have not yet been observed at this position. Two base pairs in the bulged stem 164–189 were broken. Base pair 173/186 was opened due to DMS modification of A173 and RNase ONE cleavage on the 5′-side of U186, whereas the 164/189 bp was opened to provide better phylogenetic alignment of this region in the *Saccharomyces* RNAs. This provides a consistent 3 nt spacing between the conserved UGAG sequence (158–161) and this stem, and a constant 7 nt in the 3′-side of the stem. Three contiguous base pairs at positions 269–271/301–303 were broken due to lack of covariation support at these positions and in order to provide a constant 3 nt spacing between neighboring structural elements (i.e., bulged stem 99–112/260–268 and stem 272–300) in the *Saccharomyces* RNase P RNAs and their eubacterial counterparts (Brown et al., 1993). Finally, a base pair was formed between G23 and C317. Although a U/C mismatch at this position must be postulated for *Saccharomyces kluyveri*, there is covariation support when the *Schizosaccharomyces* RNAs are taken into account. The 1 bp extension in yeast RNAs makes this stem (17–23/317–323) consistent with eubacterial RNAs in terms of stem length (7 bp) and spacing to a universally conserved AU sequence (315–316).

One refinement that was possible but not made to the structural model was a 3 bp extension (90–92/350–352) of the long-distance interaction 94–98/345–349 (shown enclosed by brackets in Figure 3). If this extension was made (with a bulged U at position 93), this entire structural element would seem to be completely homologous to its counterparts in the eubacterial RNase P RNAs. However, although the biochemical data do not argue for or against the extension, phylogenetic evidence within the yeast RNAs argues against these additional base pairs, as described elsewhere (Tranguch & Engelke, 1993).

The few hits that are inconsistent with this model deserve further consideration. Weak chemical hits were seen at U2, A328, and U333, calling into question whether bp 2/367 or bp 328/333 actually form (note that A367 could not be assessed). Since covariations are seen at these positions in the *Saccharomyces* RNAs, base-pairing is thought to be maintained. Therefore, chemical attack at these A/U pairs might be explained in terms of possible instability of weak base pairs (i.e., A/U or G/U) at the ends of stems. The inconsistent DMS hit at A48 has already been addressed. The most troubling hit in the entire molecule is a moderate DMS modification of A100 within the bulged stem 99–112/260–268. There is no covariation support for the bp at this position in the *Saccharomyces* RNAs, and multiple folding arrangements are possible in this central region of the molecule. However, since the biochemical data do not favor any of the

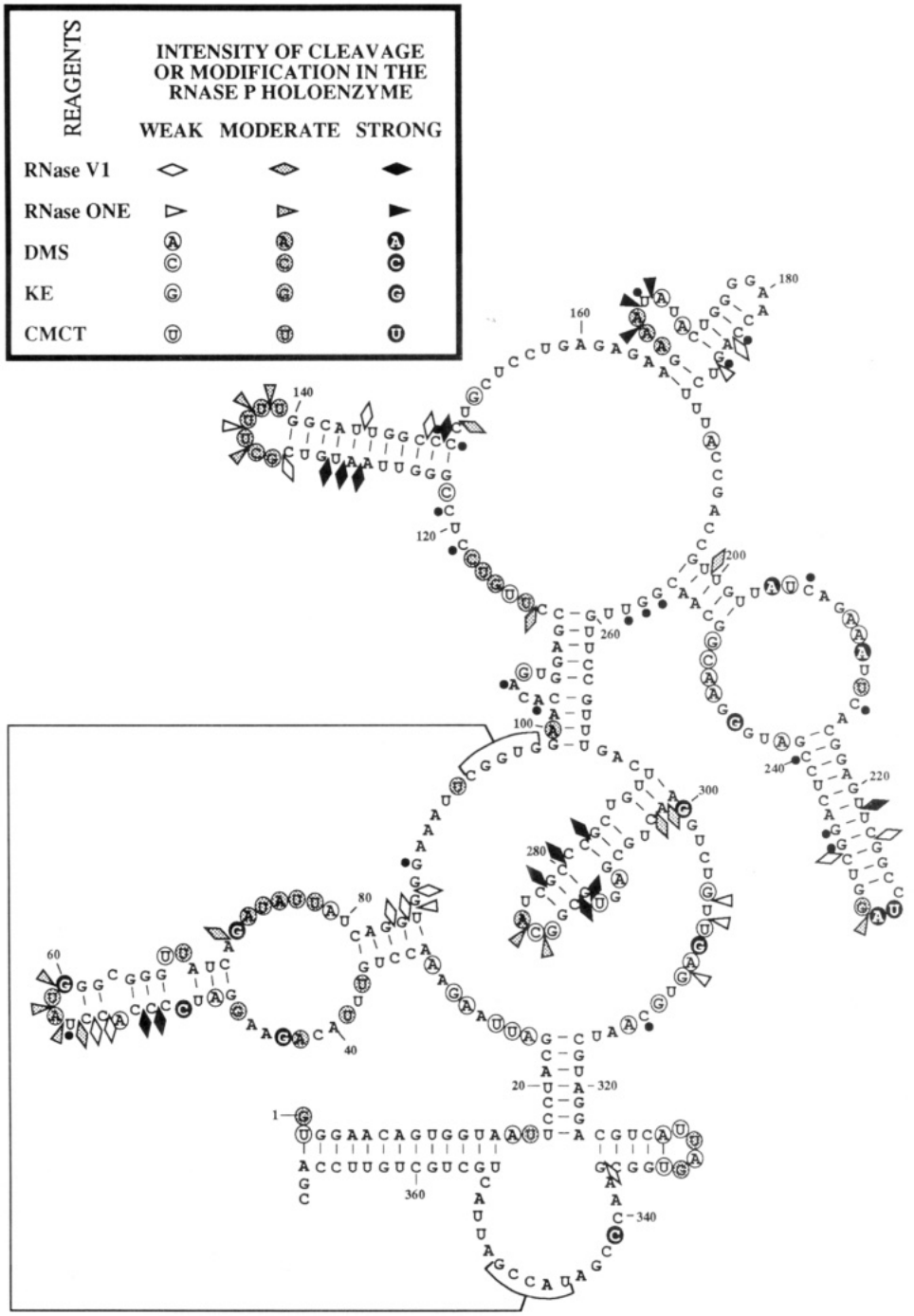

FIGURE 3: Structure-specific enzymatic and chemical probing of the RNase P holoenzyme. The RNase P holoenzyme was exposed to a variety of reagents that preferably target double-stranded (RNase V1) or single-stranded (RNase ONE, DMS, KE, and CMCT) regions of the RNA. "Hits" were mapped by primer extension analysis. Premature primer extension stops seen in control samples are denoted by a small filled circle *next* to the nucleotide; all other symbols are defined in the key. Indicated reactivities (weak, moderate, or strong) were from visual estimates of band intensities. The secondary structure model has been adjusted (cf. Figure 1) to be consistent with both the holoenzyme structure probing data and the phylogenetic comparisons.

phylogenetically-consistent alternate forms we have identified, this structure is being left as it is, with reservations, for the original rationale (Tranguch & Engelke, 1993).

Finally, four RNase V1 cleavages that do not fall immediately within stems merit explanation. Each of these hits (i.e., positions 72, 86, 113, and 151) are located two positions away from the 3′-side of a stem structure. It should be noted that RNase V1 cleaves not only within double-stranded regions but often one to two positions away from the base-paired region itself (Lowman & Draper, 1986). RNase V1 also reportedly recognizes noncanonical base pairs and stacked single-stranded regions (Lockard & Kumar, 1981). Thus, structural interpretation of RNase V1 hits is often not entirely straightforward, and these particular hits might or might not be consistent with the current structural model.

Accessible regions of the holoenzyme's RNA can be distinguished from inaccessible regions (i.e., nucleotides either buried in the folded RNA's interior or blocked by protein components) by analyzing the presence and intensity of structure-specific hits. In general, the most accessible areas of the RNA appear to be the distal portions of long hairpin stems. For example, moderate and strong hits are observed with both nuclease and chemicals in the hairpin loop and stem region of the stems encompassing nucleotides 47–71, 123–150, 216–241, and 272–299. These structural elements are thought to project away from the holoenzyme's core into

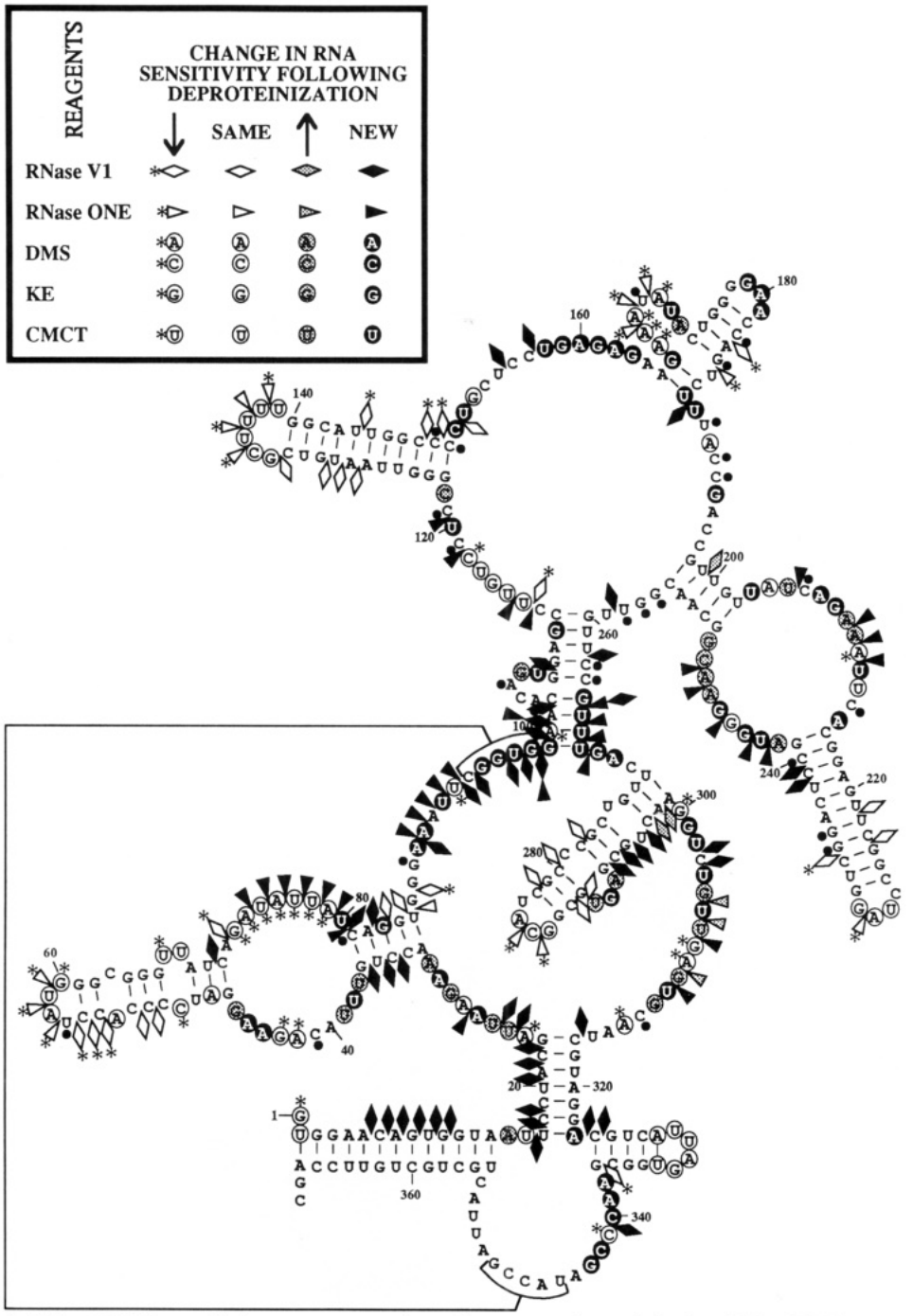

FIGURE 4: Changes in RNase P RNA sensitivity to structure probes subsequent to deproteinization. RNase P holoenzyme was preincubated with proteinase K prior to exposure to structure-specific reagents. "Hits" were identified by primer extension analysis, and the reactivities were compared with those found in the holoenzyme (Figure 3). Premature primer extension stops seen in control samples prevented assignments at those positions and are denoted by a small filled circle *next* to the nucleotide; all other symbols are defined in the key. Note that the shading of symbols (grey or black) refers to an increase in sensitivity to probes when the deproteinized RNase P RNA is compared to the holoenyzme, rather than to absolute strength of the hit. Enhanced reactivity is interpreted as exposure of protein-protected regions or rearrangement in the RNA structure following digestion of the RNase P protein subunit(s). Asterisks next to open symbols denote decreased reactivities, which could be interpreted as structural rearrangements.

solution, on the basis of their susceptibility to structure probes. Two of these stems, 47–71 and 216–241, are each part of larger structural motifs known as (proximal) stem–internal loop–hairpin stem elements, found within nucleotides 32–85 and 198–255, respectively. Whereas the hairpin stem portion of these elements is susceptible to both enzymatic and chemical reagents, the internal loops display only chemical sensitivities, and the proximal stems show only one or two mild RNase V1 hits. This suggests that these portions are positioned closer to other structures within the holoenzyme. Thus, large probes such as the nucleases are less likely to cleave within these regions due to steric hindrance, yet the physically smaller chemical reagents can still easily access these sites. A similar situation appears to exist for the small hairpin stem 324–337.

Although the bulged portion of hairpin stem 165–188 displays susceptibility to both enzymatic and chemical probes, its hairpin loop does not exhibit even weak modification by chemical reagents. This is unique in the *RPR1* RNA since this is the only hairpin loop that appears to be inaccessible in the holoenzyme. This region is therefore suspected to be involved in either RNA tertiary interactions or protein contacts. Another interesting observation is that stems connecting distant

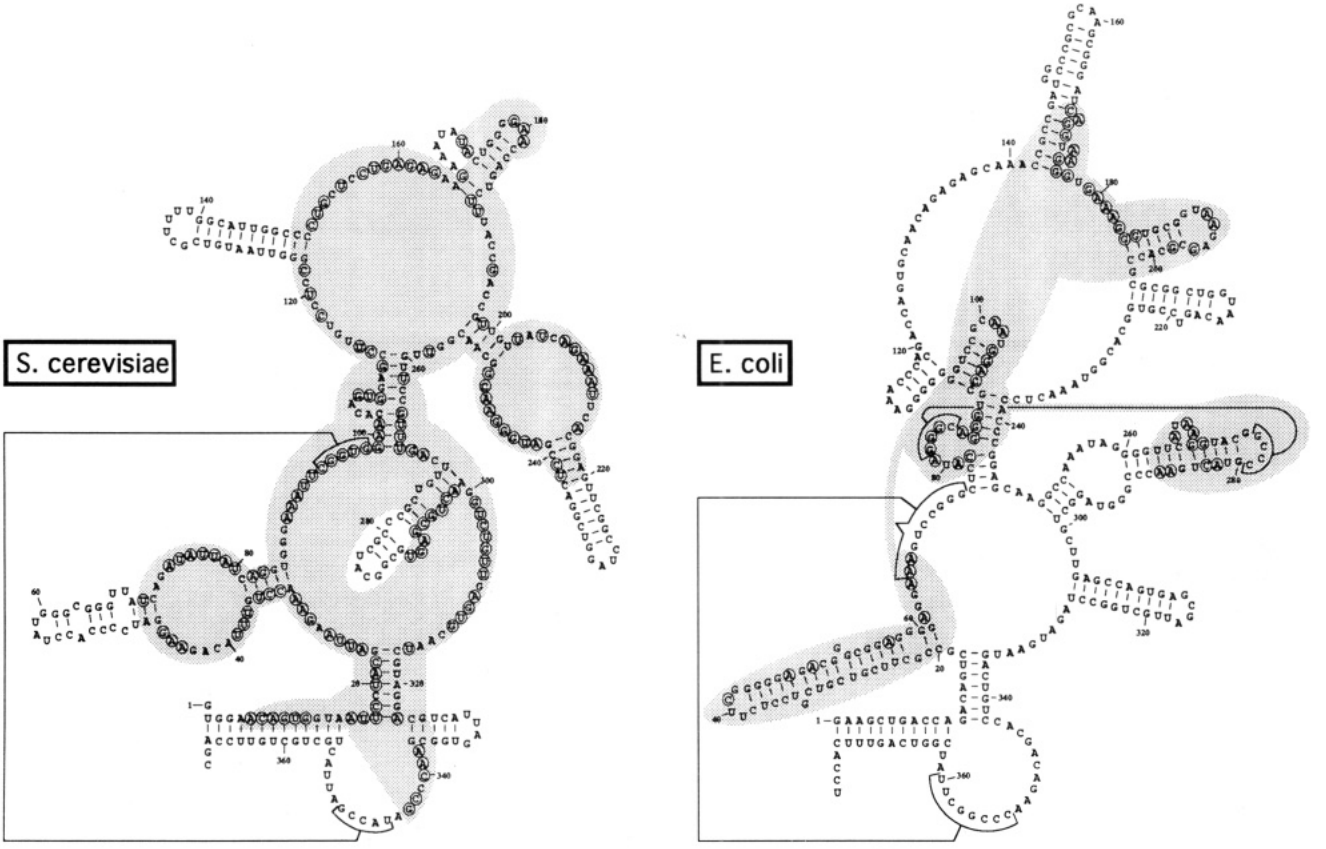

FIGURE 5: Comparison of the protein-protected regions in the RNase P RNAs from *S. cerevisiae* (nuclear) and from *E. coli*. In the *RPR1* RNA structure (*S. cerevisiae*; left), positions displaying increased or new sensitivities to RNase V1, RNase ONE, DMS, KE, or CMCT following deproteinization are indicated by circled nucleotides (for the purpose of visual clarity, nuclease hits are mapped to the nucleotide 5′ of the cleaved phosphodiester bond). In the M1 RNA structure (*E. coli*; right; Brown et al., 1993), circled nucleotides indicate positions protected from RNase T1 cleavage (Vioque et al., 1988) or from DMS modification (see footnote 2; data used by permission of Talbot and Altman) in the presence of the C5 protein. In both structures, shaded areas indicate clusters of nucleotides that are protected in the presence of the protein subunits. Note that the final 26 nucleotides at the 3′-end of *RPR1* RNA could not be directly assessed.

segments of the RNA's primary sequence (i.e., stems 2–14/355–367, 17–23/317–323, 94–98/345–349, and 99–112/260–268) show no cleavage by RNase V1 in the holoenzyme. (Note that the 3′-sides of two of these stems, 345–349 and 355–367, could not be directly assessed by primer extension because they are too close to the 3′-end of the RNA.)

Of the remaining loop regions located internally, only three segments exhibited stretches that were accessible to structure probes. Nucleotides 114–118 displayed moderate sensitivity to chemical reagents. Weak chemical modifications were detected within positions 24–31. The third region, 305–314, was hit by both enzymatic and chemical probes, suggesting that this loop may be closer to the holoenzyme's surface. This is interesting when one takes into account that a portion of this segment overlaps with one of the three areas that contain patches of sequence found to be invariant among all RNase P RNAs (i.e., regions 309–316, 87–98, and 339–349). Furthermore, the G at position 308 was strongly hit by not only KE but also CMCT. This is unusual since CMCT modification of Gs was generally found to be weak and infrequent compared to that of Us in the *RPR1* RNA. However, the single most intense CMCT hit in this entire RNA was found at G308, suggesting that the local chemical environment is altered in such a way as to make the N1 position of G308 a much better nucleophile toward CMCT modification. This position is of particular interest because it borders a strongly conserved segment from 309–316 that is suspected of being part of the RNase P active site (Burgin & Pace, 1990). A functional role for this position would be compatible with the fact that the holoenzyme is sensitive to inactivation by low concentrations of CMCT (Lee, 1991) and the observation that this exceptional G sensitivity vanishes when the holoenzyme is stripped of protein.

Changes in Accessibility to Structure Probes following Protein Digestion. As stated earlier, only the holoenzyme form of RNase P from *S. cerevisiae* exhibits catalytic activity. Unlike the case for the eubacterial RNAs, the *RPR1* RNA subunit by itself has not yet been shown to bind substrate or catalyze the RNase P reaction. It has not been determined whether the protein subunit(s) merely assist in properly folding the RNA into an active conformation, or whether the protein subunit(s) have assumed, at least in part, binding and/or catalytic functions. RNA structure probing in the absence of protein is therefore not particularly relevant to the "active" conformation, but it can be useful in indicating which areas of the RNA are being protected by the presence of protein. Interpretation of the resulting RNA "footprint" is not straightforward, however, since areas of direct RNA–protein interaction cannot be distinguished from structural rearrangements induced by protein digestion that expose sites previously buried in the folded RNA's interior. Nevertheless, by employing this approach, candidate regions of RNA–protein interaction can potentially be identified, and the influence of protein on the RNA's structure can be examined.

Yeast nuclear extract containing the RNase P holoenzyme was subjected to protein digestion with proteinase K under nondenaturing aqueous conditions and subsequently probed with structure-specific reagents. A total of 248 enzymatic and chemical hits were observed when the deproteinized *RPR1* RNA was probed—almost twice as many hits as seen in the

holoenzyme. The presence and intensities of these hits were compared to those observed in the holoenzyme's RNA, and the changes in RNA sensitivity to structure probes following deproteinization are mapped in Figure 4 onto the refined secondary structure model of *RPR1* RNA. The changes in RNA sensitivity were divided into four classes: "decreased", "same", "increased", and "new". "Decreased" hits were defined as those hits that appeared less intense in deproteinized samples as compared to the holoenzyme; this category included the 30 hits observed only in the holoenzyme. These hits might be due either to structural rearrangements following protein digestion or simply to a concomitant drop in attack at previous sites when more sites became available. "Same" hits displayed approximately equal intensities in holoenzyme and proteinase K-treated samples. "Increased" hits appeared more intense following deproteinization, whereas "new" hits were seen only in the proteinase K-treated samples. "Increased" and "new" might be interpreted either as exposure of protein-protected regions or as structural rearrangements subsequent to protein digestion. It is interesting to note that all regions that are inaccessible in the holoenzyme display some degree of susceptibility to structure probes following deproteinization. We cannot rule out the possibility that proteinase K digestion leaves peptides still bound to the RNA, thus masking the full extent to which the RNA becomes solvent accessible upon protein removal. However, more stringent peptide removal increases the chances of RNA structure rearrangement and was avoided with the understanding that even the dramatic increase in RNA accessibility might not represent the full footprint.

Given the large number of decreased, increased, and new hits seen in Figure 4, the RNA is likely to undergo some degree of rearrangement following protein digestion, but it is difficult to judge how extensive this rearrangement might be. Only 28 of the 142 new hits are inconsistent with the refined structure model. Of these inconsistent hits, 14 are found in two regions, 94–102 and 265–268, which are presumed to be involved in long-distance interaction 94–98/345–349 and stem 99–102/265–268. The occurrence of both single-stranded and double-stranded hits, sometimes at the same position, within these regions suggests that these two structural elements are unstable when the RNase P protein contingent is digested. The remaining inconsistent hits are spread out over the RNA molecule. Some of these hits might have trivial explanations (e.g., RNase V1 hits at positions 24 and 25 might be due to stem 17–23/317–323 "zipping up"; chemical modifications at G166 and U188 might be due to the instability of the 2 bp stem 165–166/187–188 in the absence of protein). Explanations for other hits are less obvious (e.g., RNase V1 hits at positions 155, 157, 302, 303, and 340; KE modification of G83).

Although the existence of increased and new hits cannot be used to distinguish direct protein contacts from induced structural rearrangements, these types of hits do indicate which regions of the RNA are protected, directly or indirectly, by the presence of the protein. Figure 5 summarizes the protein-protected regions in *RPR1* RNA and compares them to the C5 protein-protected regions in M1 RNA, the RNase P subunits from *E. coli* (Vioque et al., 1988).[2] These patterns of protein protection might not be directly comparable for three reasons: (1) except for DMS, different structure probes were used to footprint the two RNAs; (2) some of the structural elements in one RNA might not have a structural homolog in the other RNA, and vice versa; and (3) M1 RNA retains catalytic activity in the absence of its protein subunit (Guerrier-Takada et al., 1983), whereas *RPR1* RNA does not. Nevertheless, the regions of RNA protected in the *S. cerevisiae* nuclear RNase P were considerably more extensive than that observed in the *E. coli* holoenzyme. This is not surprising since the protein moiety generally comprises a larger percentage of the the RNase P holoenzyme in eukaryotes than in eubacteria. The refined secondary structure model and the footprint of protein-protected regions provided in this work will help to guide mutational analysis of RNase P holoenzyme assembly and function.

ACKNOWLEDGMENT

We are grateful to Simon J. Talbot and Sidney Altman for communication of results prior to publication, The Upjohn Co. for providing us with kethoxal, and Don G. Gilbert for allowing us to use a prerelease version of LoopDloop, the computer drawing program used to produce the RNA structural models. We thank members of the Engelke and Thiele labs for helpful discussions during the course of these studies. We acknowledge support services provided by the University of Michigan, including the DNA Synthesis Facility/University of Michigan Biomedical Research Core and the University of Michigan Clinical Research Center.

REFERENCES

Akaboshi, E., Guerrier-Takada, C., & Altman, S. (1980) *Biochem. Biophys. Res. Commun. 96*, 831–837.

Altman, S. (1989) *Adv. Enzymol. Relat. Areas Mol. Biol. 62*, 1–36.

Brown, J. W., & Pace, N. R. (1992) *Nucleic Acids Res. 20*, 1451–1456.

Brown, J. W., Haas, E. S., & Pace, N. R. (1993) *Nucleic Acids Res. 21*, 671–679.

Burgin, A. B., & Pace, N. R. (1990) *EMBO J. 9*, 4111–4118.

Cherayil, B., Krupp, G., Schuchert, P., Char, S., & Söll, D. (1987) *Gene 60*, 157–161.

Christiansen, J., & Garrett, R. (1988) *Methods Enzymol. 164*, 456–468.

Dang, Y. L., & Martin, N. C. (1993) *J. Biol. Chem. 268*, 19791–19796.

Darr, S. C., Pace, B., & Pace, N. R. (1990) *J. Biol. Chem. 265*, 12927–12932.

Darr, S. C., Brown, J. W., & Pace, N. R. (1992) *Trends Biochem. Sci. 17*, 178–182.

Doersen, C.-J., Guerrier-Takada, C., Altman, S., & Attardi, G. (1985) *J. Biol. Chem. 260*, 5942–5949.

Doria, M., Carrara, G., Calandra, P., & Tocchini-Valentini G. P. (1991) *Nucleic Acids Res. 19*, 2315–2320.

Evans, C. F., & Engelke, D. R. (1990) *Methods Enzymol. 181*, 439–450.

Forster, A. C., & Altman, S. (1990) *Cell 62*, 407–409.

Gardiner, K., & Pace, N. R. (1980) *J. Biol. Chem. 255*, 7507–7509.

Guerrier-Takada, C., & Altman, S. (1984) *Biochemistry 23*, 6327–6634.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983) *Cell 35*, 849–857.

James, B. D., Olsen, G. J., Liu, J., & Pace, N. R. (1988) *Cell 52*, 19–26.

Jayanthi, G. P., & Van Tuyle, G. C. (1992) *Arch. Biochem. Biophys. 296*, 264–270.

Kline, L., Nishikawa, S., & Söll, D. (1981) *J. Biol. Chem. 256*, 5058–5063.

Knap, A. K., Wesolowski, D., & Altman, S. (1990) *Biochimie 72*, 779–790.

Krupp, G., Cherayil, B., Frendewey, D., Nishikawa, S., & Söll, D. (1986) *EMBO J. 5*, 1697–1703.

Lawrence, N., Wesolowski, D., Gold, H., Bartkiewicz, M., Guerrier-Takada, C., McClain, W. H., & Altman, S. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 233–238.
Lee, J.-Y. (1991) Ph.D. Thesis, University of Michigan, Ann Arbor, MI.
Lee, J.-Y., & Engelke, D. R. (1989) *Mol. Cell. Biol. 9*, 2536–2543.
Lee, J.-Y., Rohlman, C. E., Molony, L. A., & Engelke, D. R. (1991) *Mol. Cell. Biol. 11*, 721–730.
Lockard, R. E., & Kumar, A. (1981) *Nucleic Acids Res. 9*, 5125–5140.
Lowman, H. B., & Draper, D. E. (1986) *J. Biol. Chem. 261*, 5396–5403.
Meador, J., III, Cannon, B., Cannistraro, V. J., & Kennell, D. (1990) *Eur. J. Biochem. 187*, 549–553.
Miller, D. L., & Martin, N. C. (1983) *Cell 34*, 911–917.
Morales, M. J., Wise, C. A., Hollingsworth, M. J., & Martin, N. C. (1989) *Nucleic Acids Res. 17*, 6865–6881.
Morales, M. J., Dang, Y. L., Lou, Y. C., Sulo, P., & Martin, N. C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9875–9879.
Nieuwlandt, D. T., Haas, E. S., & Daniels, C. J. (1991) *J. Biol. Chem. 266*, 5689–5695.
Piper, P. W., & Stråby, K. B. (1989) *FEBS Lett. 250*, 311–316.
Reich, C., Olsen, G. J., Pace, B., & Pace, N. R. (1988) *Science 239*, 178–181.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Shiraishi, H., & Shimura, Y. (1988) *EMBO J. 7*, 3817–3821.
Stern, S., Moazed, D., & Noller, H. F. (1988) *Methods Enzymol. 164*, 481–489.
Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.
Tranguch, A. J., & Engelke, D. R. (1993) *J. Biol. Chem. 268*, 14045–14055.
Vioque, A., Arnez, J., & Altman, S. (1988) *J. Mol. Biol. 202*, 835–848.
Wang, M. J., Davis, N. W., & Gegenheimer, P. (1988) *EMBO J. 7*, 1567–1574.