Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme

Tien-Hao Chen[1,2], Akiko Tanimoto[1], Nikoloz Shkriabai[3], Mamuka Kvaratskhelia[3], Vicki Wysocki[1,*] and Venkat Gopalan[1,2,*]

[1]Department of Chemistry & Biochemistry, The Ohio State University, Columbus, OH 43210, USA, [2]Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA and [3]College of Pharmacy, The Ohio State University, Columbus, OH 43210, USA

ABSTRACT

Among all enzymes in nature, RNase P is unique in that it can use either an RNA- or a protein-based active site for its function: catalyzing cleavage of the 5′-leader from precursor tRNAs (pre-tRNAs). The well-studied catalytic RNase P RNA uses a specificity module to recognize the pre-tRNA and a catalytic module to perform cleavage. Similarly, the recently discovered proteinaceous RNase P (PRORP) possesses two domains – pentatricopeptide repeat (PPR) and metallonuclease (NYN) – that are present in some other RNA processing factors. Here, we combined chemical modification of lysines and multiple-reaction monitoring mass spectrometry to identify putative substrate-contacting residues in *Arabidopsis thaliana* PRORP1 (*At*PRORP1), and subsequently validated these candidate sites by site-directed mutagenesis. Using biochemical studies to characterize the wild-type (WT) and mutant derivatives, we found that *At*PRORP1 exploits specific lysines strategically positioned at the tips of it's V-shaped arms, in the first PPR motif and in the NYN domain proximal to the catalytic center, to bind and cleave pre-tRNA. Our results confirm that the protein- and RNA-based forms of RNase P have distinct modules for substrate recognition and cleavage, an unanticipated parallel in their mode of action.

INTRODUCTION

The 5′ maturation of precursor tRNAs (pre-tRNAs) is catalyzed by RNase P, an essential endonuclease (1). RNase P in the three domains of life displays variation in its subunit composition and even in the make-up of its catalytic center (2–7). Broadly, RNase P can be classified into two architectures: (i) a ribonucleoprotein (RNP) form, where a catalytic RNA is associated with one, up to 5 and 10 protein cofactors in Bacteria, Archaea and Eukarya (nucleus), respectively; and (ii) an RNA-free, protein-alone form found in nucleus/organelles of four out of the five eukaryal super-groups (4,8). Because the proteinaceous RNase P (PRORP) variant shares no sequence similarity with the protein subunits of the RNP forms, the two RNase P types reflect independent evolutionary paths. The basis for this diversification remains unclear, as is the coexistence of these two forms in the same cell but in different compartments (4,7–10).

Unlike the RNP variant first described a few decades ago (1,11), PRORP was discovered recently due to the painstaking purification and characterization of human mitochondrial RNase P (7), which showed that three proteins (MRPP1, MRPP2 and MRPP3) are essential for this activity. However, MRPP3 homologs in *Arabidopsis thaliana* were soon after shown to be active as single (recombinant) polypeptides and named PRORPs (9). Subsequent reports confirmed activities of recombinant PRORPs from the alga *Ostreococcus tauri* (10), the protozoan *Trypanosoma brucei* (12) and the moss *Physcomitrella patens* (13).

Since the discovery of PRORP (7), there has been interest in examining whether the two structurally distinct forms of RNase P are functionally comparable. Despite different active-site scaffolds, both RNase P variants exploit a Mg^{2+}-bound water as the nucleophile to hydrolyze the scissile phosphodiester bond in pre-tRNAs (14–16). Moreover, genetic complementation assays revealed that PRORP rescues mutations that eliminate the activity of the two-subunit RNase P RNP in *Escherichia coli* and the 10-subunit RNP in *Saccharomyces cerevisiae* (9,12,17). Results from these genetic studies are remarkable since *in vitro* PRORP exhibits a k_{cat}/K_m up to three-orders of magnitude lower than those of bacterial or eukaryotic RNase P RNP variants (6,14,16).

*To whom correspondence should be addressed. Tel: +1 614 292 1332; Fax: +1 614 292 6773; Email: gopalan.5@osu.edu
Correspondence may also be addressed to Vicki Wysocki. Tel: +1 614 292 8687; Fax: +1 614 292 1685; Email: wysocki.11@osu.edu

While it is difficult to correlate catalytic and fitness landscapes, understanding the mechanisms of substrate recognition and cleavage strategies of the two forms of RNase P is essential to determine if each has distinctive traits that prove useful under select conditions. Also, such efforts may yield general insights into the transition from an RNA to a protein world.

The *A. thaliana* (*At*) nuclear genome encodes three PRORPs. While *At*PRORP1 is targeted to mitochondria and chloroplasts, *At*PRORP2 and *At*PRORP3 are localized to the nucleus and appear to be functionally redundant (9,18). *At*PRORP1 cannot be knocked out (9). Its knock-down by RNAi in stable transgenic plants revealed that, although protein synthesis was adversely affected in both organelles, only photosynthesis was defective and mitochondrial respiration was unaffected (19). Strikingly, the effects on processing of individual organellar tRNAs in these knock-down lines were variable (19); similar findings were reported from knock-down of *AtPRORP1* by transient virus-induced gene silencing and from knock-out of organellar PRORPs in *P. patens* (13,18). Also, recombinant *At*PRORP1, *At*PRORP2 and *At*PRORP3 exhibit some substrate selectivity as evidenced by a 13-fold variation in catalytic efficiencies for processing of select nuclear and organellar pre-tRNAs (20). Together, these results suggest that organellar PRORPs might preferentially process certain pre-tRNAs. To understand this substrate bias, we investigated here the structural basis for substrate recognition by *At*PRORP1.

The high-resolution crystal structure of *At*PRORP1 revealed a sequential arrangement of N-terminal pentatricopeptide repeat (PPR), central zinc-finger and C-terminal Nedd4-BP1 YacP nuclease (NYN) domains (15). The central domain orients and stabilizes the PPR and NYN domains, which constitute the arms of a V-shape structure (15). PPR proteins are prevalent in plant and other eukaryotic proteomes, and play critical roles in regulating RNA stability and editing, transcription and translation (21,22). PPR domains are composed of different numbers of PPR motifs ($\sim$35 aa), each of which adopt a helix-turn-helix fold and enable modular RNA recognition by using two or three amino acids for readout of a specific nucleotide (23–28). The NYN domain, which is found in all three domains of life, is an ancient nuclease motif that utilizes Mg^{2+} for phosphodiester cleavage (29). Based on the domain architecture of *At*PRORP1, a division of labor between the PPR and NYN domains with respect to substrate recognition and cleavage, respectively, was expected (4,15). Mutagenesis data provide preliminary support for this idea. First, deleting the first four of the five tandemly-arranged PPR motifs in *At*PRORP1 resulted in a 34-fold pre-tRNA binding defect and decreased cleavage rate by >2000-fold (15). Also, mutating the conserved active-site aspartates residues in the NYN domain eliminated *At*PRORP1 activity (9,15). Second, two observations led to the idea that *At*PRORP1 might recognize the 'C$_{56}$ G/A$_{57}$ A$_{58}$' sequence in the T-loop of pre-tRNAs (30): (i) footprinting assays on a UV-crosslinked *At*PRORP1–pre-tRNA complex revealed that nucleotides in D/T-loops were protected from nuclease digestion (31); (ii) the nucleotide code predicted based on the key residues in the highly conserved PPR motifs 2, 3 and 4. Indeed, sin-

gle time-point assays showed that changing the identity of C_{56} or A_{57} or A_{58} led to a variable (up to 6-fold) decrease in activity (30), although adverse effects from alterations to T-loop structure were not ruled out.

There are now two models that show docking of the pre-tRNA on *At*PRORP1, with the PPR motifs recognizing the D/T-loop and the NYN domain positioned near the cleavage site (30,31), but there are few details in terms of which residues in *At*PRORP1 contact the pre-tRNA. To this end, we employed chemical footprinting based on lysine modification coupled with multiple-reaction monitoring mass spectrometry (MRM MS) to identify the pre-tRNA^{Cys} contacting sites in *At*PRORP1. Moreover, to validate that the identified sites were *bona fide*, we used site-directed mutagenesis followed by pre-tRNA^{Cys} binding and cleavage assays. Our results show that multiple lysines in the PPR and NYN (but not the central) domains in *At*PRORP1 contact the pre-tRNA, and support the idea that despite their independent evolution both RNase P variants utilize two distinct modules to accomplish substrate recognition and cleavage.

MATERIAL AND METHODS

Molecular cloning, overexpression and purification of *At*PRORP1

See Supplementary Information for details.

Labeling of pre-tRNA^{Cys} for binding and activity assays

The DNA templates for *in vitro* transcription were amplified using PCR with Phusion High-Fidelity DNA polymerase (New England Biolabs). The template for PCR was pBT7-ptCys-5523 (see Supplementary Information) (32). To generate pre-tRNA^{Cys} with a 5-nt 5′-leader and 0-nt 3′-trailer (i.e. terminating at the discriminator position; Supplementary Figure S1), the DNA templates were generated using 5′-***taatacgactcactata***ggtttggctaggtaacataatg-3′ (***bold italicized*** font denotes the T7 RNA polymerase promoter) and 5′-aggccaaggacggagtc-3′. Alternatively, changing only the reverse primer to 5′-gttatttctggtcatgaagc-3′ permitted generation of a template to transcribe pre-tRNA^{Cys} with a 5-nt 5′-leader and 23-nt 3′-trailer (Supplementary Figure S1). After *in vitro* transcription by T7 RNA polymerase, the reactions were subjected to DNase I treatment, phenol-chloroform extraction, extensive dialysis against water and precipitation with ethanol. The concentration of pre-tRNA^{Cys} with or without the 23-nt 3′-trailer was then determined using Abs_{260} values and the extinction coefficients of 1.012 or 0.827 $\mu M\,cm^{-1}$, respectively, which we experimentally determined by hydrolysis of these RNAs with sodium hydroxide.

For use in binding assays, pre-tRNA^{Cys} was 3′-labeled with fluorescein. The RNA was first subjected to periodate oxidation and then reacted with fluorescein-5-thiosemicarbazide (33,34). For use in cleavage assays, pre-tRNA^{Cys} was radiolabeled at its 5′ end. The *in vitro* transcripts were de-phosphorylated with an alkaline phosphatase, and 5′-labeled with [γ-^{32}P]-ATP and Optikinase (Affymetrix). The labeled RNA was then gel purified using denaturing PAGE [8% (w/v) polyacrylamide, 7 M urea].

The choice of substrates for the footprinting, binding and cleavage assays was empirically determined based on results from our preliminary studies. First, for our initial chemical footprinting coupled to matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) MS studies, we used a pre-tRNA^{Cys} containing a 55-nt 5′-leader and a 23-nt 3′-trailer because this substrate was previously used in *At*PRORP1 binding and cleavage assays (15). During our subsequent ESI MS/MS experiments, we found that the results obtained with the 55-nt 5′-leader and 23-nt 3′-trailer variant were quite similar to those obtained with substrates in which the leader and trailer lengths were decreased to 5 nt and 0 nt, respectively (Supplementary Table S1). Because trimming of the substrate was expected to minimize possible non-specific interactions, all the MS data were obtained using pre-tRNA^{Cys} with a 5-nt 5′-leader and 0-nt 3′-trailer. This substrate was also used for the cleavage assays. Second, for the fluorescence polarization binding assays, we typically introduce a fluorophore at the 3′-terminus. Because a bulky fluorophore at the 3′-end of a pre-tRNA with a 0-nt 3′-trailer was expected to interfere with binding to *At*PRORP1, we decided to use a variant with a 5-nt 5′-leader and 23-nt 3′-trailer. Pre-tRNA^{Cys} with a 23-nt 3′-trailer is cleaved only 2-fold slower than the 0-nt counterpart (data not shown), indicating that these two substrates are recognized and cleaved comparably by *At*PRORP1.

Biotinylation of *At*PRORP1 and *At*PRORP1–pre-tRNA^{Cys} complex

For these experiments, 21 μM pre-tRNA^{Cys} in binding buffer [20 mM HEPES (pH 7.2), 100 mM NH$_4$OAc, 4 mM DTT] was refolded by incubating at 95°C for 3 min, 37°C first for 10 min and then for 30 min in the presence of 1 mM Ca(OAc)$_2$, and finally slow cooling to 25°C. Subsequently, 8.3 μM *At*PRORP1 was added to promote complex formation for 15 min at 25°C. The resulting *At*PRORP1–pre-tRNA^{Cys} complex was then incubated with 1 mM N-hydroxysuccinimidobiotin (NHSB) in a final volume of 20 μl at 25°C for 30 min, and the reaction quenched by addition of 200 mM Tris-HCl (pH 7.5). Control reactions with *At*PRORP1 alone excluded pre-tRNA^{Cys}.

In-gel pepsin digestion

Subsequent to biotinylation, 30 μl of 0.3 μg/μl of *At*PRORP1 was gel purified using SDS-PAGE [12% (w/v) polyacrylamide]. The gel was stained with Bio-safe Coomassie (Bio-Rad) for 45 min, and destained in water for 1 h with shaking. Visualized protein bands were excised and cut into ∼1 × 1 mm slices, and destained twice (or more times if the color remained) using 50% (v/v) methanol in 25 mM NH$_4$HCO$_3$. The gel slices were then dehydrated in 50% (v/v) acetonitrile in 25 mM NH$_4$HCO$_3$ and then in 100% (v/v) acetonitrile, before being completely dried in a Speed-Vac. The sample was reduced in 25 mM DTT at 56°C for 20 min, and alkylated with 55 mM iodoacetamide in the dark at 25°C for 20 min. After reduction and alkylation, the gel pieces were rinsed twice with water, and then dehydrated as described above, and dried using a SpeedVac. The in-gel digestion was initiated by adding 100 μl of 0.01 μg/μl pepsin (Promega), which had been dissolved in 50 mM NH$_4$HCO$_3$ (pH 2, adjusted with formic acid), and incubating at 37°C for 8 h with shaking. The reaction was centrifuged at 12 000 g for 5 min, and the supernatant was collected and kept on ice. The pellet containing the gel pieces was immersed in 60% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid, sonicated using a water bath sonicator (VWR) for 30 min and centrifuged at 12 000 g for 5 min. The supernatant from this second centrifugation was combined with the first, completely dried in SpeedVac, and stored at −80°C until analysis.

For Glu-C proteolysis of *At*PRORP1, the post-biotinylation processing steps until dehydration were the same as those with pepsin. Subsequently, digestion with 100 μl of 0.01 μg/μl Glu-C (Promega) was performed in 50 mM NH$_4$HCO$_3$ (pH 7.5) at 37°C overnight with shaking. The reaction was quenched using 0.5% (v/v) formic acid and centrifuged at 12 000 g for 5 min. The supernatant from the in-gel digest with Glu-C was processed similarly as pepsin (see above).

Liquid chromatography-coupled electrospray ionization (LC-ESI) tandem mass spectrometry

For an initial inventory of pre-tRNA^{Cys} contacting sites in *At*PRORP1 (Supplementary Table S1), we employed a bottom-up proteomics approach (35,36). The peptides obtained from in-gel digestion with pepsin were analyzed on an Orbitrap Elite (Thermo Fisher Scientific) mass spectrometer, which was directly coupled to a nanoACQUITY UPLC (Waters). Approximately 0.5 μg of peptides [dissolved in 0.1% (v/v) formic acid] was injected onto a Waters NanoACQUITY UPLC HSS T3 C18 column (75 μm × 150 mm, 1.8 μm particle size). The peptides were separated using mobile phase A [0.1% (v/v) formic acid] and B [acetonitrile, 0.1% (v/v) formic acid] over a 40-min gradient (5 to 35% B) at flow rate of 0.5 μl/min. The eluate was ionized in a Nanospray Flex ion source (Thermo Fisher Scientific) at 1.7 kV, and directly introduced into the mass spectrometer with the capillary heated at 275°C. MS was performed at high resolving power (240 000 FWHM) using the Orbitrap, and MS/MS was performed in the front-end Velos Pro linear ion trap. The top 10 most intense peaks in the full MS scan (400-1600 *m/z*) were sent to collision-induced dissociation for fragmentation with an isolation width of 2 *m/z*. Dynamic exclusion was used with a repeat count of 1 and the exclusion duration was set to 15 s. The peptide sequences were identified from LC-ESI MS/MS data using SEQUEST HT in Proteome Discoverer 1.4 (Thermo Fisher Scientific) at a 1% false discovery rate. This software was also used to quantitate the precursor ion area of each identified peptide.

Multiple-reaction monitoring (MRM)

MRM was performed using a Waters Xevo TQ-S triple-quadrupole mass spectrometer, which was directly coupled to a Waters nanoACQUITY UPLC. Approximately 0.1 μg [dissolved in 0.1% (v/v) formic acid] of the digested sample was injected, and separated over a 10-min gradient (2 to 35%B). Electrospray source conditions were applied with 2.7 kV capillary, 50 V cone,

20 V source offset, 110 to 150°C source temperature, and 2 to 7 bar Nebulizer gas flow. The collision energy for each targeted peptide was predicted by Skyline software (https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view) (37). This software was also used for visualization and quantitation of the transitions of the targeted peptides.

Quantitating levels of protection

Substrate-contacting sites were first determined by a semi-quantitative method that used LC-ESI MS/MS data. The peak area of a peptide without lysine ($_{145}$AEAATESSPNPGLSRG$_{160}$ derived from *At*PRORP1) was used as a reference for normalization of peak areas of biotinylated peptides. The normalized value for each lysine from PRORP1 either alone or when it was bound to pre-tRNA^{Cys} was compared, and the degree of protection was calculated for each peptide as:

$$\text{Level of protection} = \frac{At\text{PRORP1 alone}}{At\text{PRORP1–pre-tRNA complex}}$$

In instances where the lysine residue(s) was present in multiple peptides, the protection level calculated for each of these peptides was averaged to obtain a mean value.

For the MRM experiments, more than three transitions were selected for each peptide, with the exception of instances where only two predominant transitions were detected due to the proximity of two biotinylated lysines in a small peptide. Peak areas of all transitions from each biotinylated peptide were summed, and normalized against that of a similar summation performed for the reference peptide ($_{145}$AEAATESSPNPGLSRG$_{160}$).

Fluorescence polarization (FP)-based binding assays

Indicated amounts of *At*PRORP1 were mixed with 20 nM 3′-FTSC labeled pre-tRNA^{Cys} in 20 mM HEPES pH 7.2, 10 mM Ca(OAc)$_2$, 100 mM NH$_4$OAc, 4 mM DTT and 5% (v/v) glycerol. After incubation in a 384-well plate (Corning Costar black round bottom) for at least 10 min at 22–25°C, fluorescence polarization values were collected using M1000 PRO (Tecan) with G factor set to 1.2. We determined ΔP at each *At*PRORP1 concentration tested by subtracting the polarization values (P) observed in the presence of different *At*PRORP1 concentrations from those observed with the substrate alone, and used KaleidaGraph (Synergy) to obtain the dissociation constants by fitting to $\Delta P = \frac{\Delta P_{max} \times [At\text{PRORP1}]}{K_D + [At\text{PRORP1}]}$. While representative data from a single trial with wild type *At*PRORP1 and each mutant tested are shown (Supplementary Figure S5), the curve-fit errors for the K_D values were ≤23% in three or more trials.

Cleavage assays

For determining RNase P activity, 5′-[^{32}P]-labeled pre-tRNA^{Cys} (<40 nM) was refolded in 20 mM HEPES (pH 7.2), 10 mM Mg(OAc)$_2$, 100 mM NH$_4$OAc, 4 mM DTT, 5% (v/v) glycerol. The reaction at 37°C was initiated by adding 2 μM *At*PRORP1, which has been pre-incubated in the same buffer as the substrate for 10 min. At defined time points, 5-μl aliquots (from a total assay volume of 50 μl) were withdrawn and quenched using 10 μl stop solution [7 M urea, 20% (v/v) phenol, 0.2% (w/v) SDS, 10 mM EDTA, 0.05% (w/v) bromophenol blue and 0.05% (w/v) xylene cyanol]. Reaction contents were separated using denaturing PAGE [8% (w/v) polyacrylamide, 7 M urea], and visualized using the Typhoon Phosphorimager (GE Healthcare). We employed ImageQuant (GE Healthcare) to quantitate the amount of substrates and products, and used KaleidaGraph (Synergy) to calculate the rate (k_{obs}) by fitting to $P_t = P_{max} \times (1 - e^{-k_{obs}t})$, where P_t is the product formed at time t. While representative data from a single trial with wild type *At*PRORP1 and each mutant tested are shown (Supplementary Figure S6), the curve-fit errors for the k_{obs} values were ≤17% in three or more trials.

RESULTS

Probing surface topology of lysines to map AtPRORP1–pre-tRNA^{Cys} contacts

Due to the availability of a crystal structure for *At*PRORP1 and kinetic studies describing its cleavage of *Arabidopsis* mitochondrial pre-tRNA^{Cys} (14,15), we chose this enzyme-substrate pair for our experimental studies to map RNA–protein contacts. We employed a high-resolution MS-based protein footprinting approach (38–42) in which an amino acid-modifying agent is used to distinguish in an RNA/DNA-binding protein those positions with decreased accessibility when the nucleic acid ligand is present. We focused on NHSB-mediated biotinylation of lysines because (i) lysine-phosphate is one of the most common amino acid-nucleotide contacts in RNA–protein complexes (43–45), and (ii) there are 43 lysines distributed across the three structural domains of *At*PRORP1 (496 aa total in the construct we used). Briefly, this method entails treating the free *At*PRORP1 and the *At*PRORP1–pre-tRNA complex (assembled in the presence of Ca^{2+} to prevent cleavage) with NHSB before separation by SDS-PAGE. The polypeptides are then subjected to in-gel proteolytic digestion and MS analysis to identify the positions that are protected by the pre-tRNA (Figure 1).

The choice of Ca^{2+} in the above experiments was guided by previous findings. First, *At*PRORP1 shows no detectable pre-tRNA cleavage activity in the presence of Ca^{2+}, even though fluorescence polarization assays show tight substrate binding (15); we have confirmed these findings under our assay conditions (see below). Second, a recent hybrid quantum mechanics/molecular mechanics free energy calculation showed that differences in ligand-Ca^{2+} bond distances and charge transfer, and not active-site architectural changes, underlie the inability of Ca^{2+} to support RNase H catalysis (46). This reason is likely to apply to other phosphoryl transfer enzymes including *At*PRORP1.

Improving the coverage of modified lysines in AtPRORP1

After establishing the dissociation constant of *At*PRORP1 for binding pre-tRNA^{Cys} [158 ± 33 nM in 1 mM Ca(OAc)$_2$] using FP assays, we formed the enzyme-substrate complex at μM concentrations using a 2.5-fold excess of pre-tRNA^{Cys} over *At*PRORP1 to minimize the presence of un-
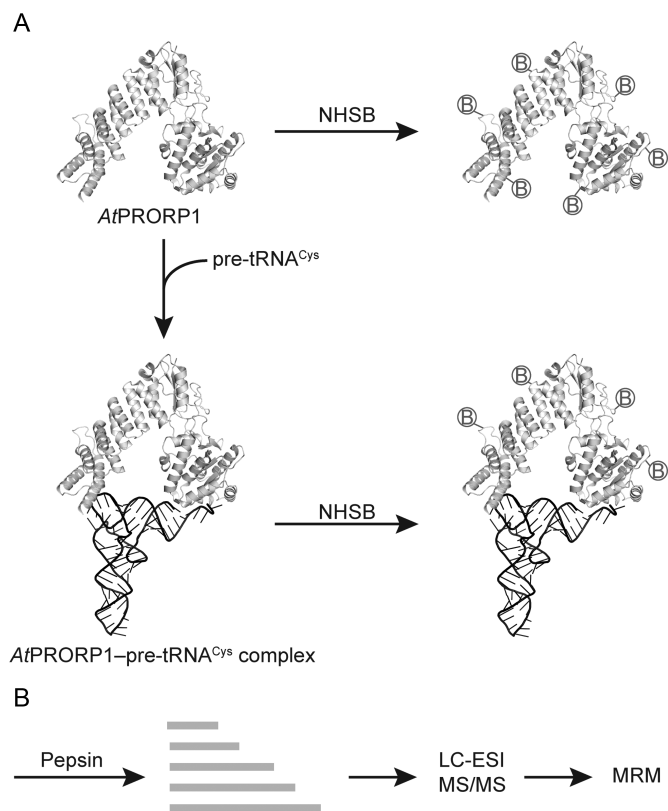
A

*At*PRORP1

NHSB

pre-tRNA^Cys

NHSB

*At*PRORP1–pre-tRNA^Cys complex

B

Pepsin

LC-ESI
MS/MS

MRM

Figure 1. Scheme for identification of substrate-contacting lysines in *At*PRORP1 by N-hydroxysuccinimido-biotin (NHSB) modification and mass spectrometry. (**A**) Incubation with NHSB is expected to biotinylate (circled B) all solvent-exposed lysines in free *At*PRORP1 but not those interacting with pre-tRNA^Cys in the substrate-bound form of *At*PRORP1. (**B**) Subsequent to NHSB modification, *At*PRORP1 was digested with pepsin, and the resulting peptides were subjected to LC-ESI MS/MS to identify those with substrate-contacting lysines. Multiple reaction monitoring (MRM) was then used to accurately quantitate the extent of modification of each lysine in these candidate peptides. Ribbon diagrams were generated using PDB files 4G23 (PRORP1) and 4TNA (tRNA).

bound protein. After modification with NHSB (30 min at 22°C) and purification by SDS-PAGE, we followed established procedures (39) using trypsin digestion followed by MALDI-ToF MS to identify the substrate-contacting sites. We found only seven lysines to be modified by NHSB, of which only one (K439) showed strong protection upon binding to pre-tRNA^Cys (Supplementary Figure S2).

To increase the coverage of biotinylated lysines in *At*PRORP1, we reasoned that a different protease and or an ionization method that could be coupled with LC might prove beneficial. The clustering of lysines and arginines in the primary sequence of *At*PRORP1 and the differential modification of lysines by biotinylation suggested that alternatives to trypsin would be desirable. Therefore, we sought to use either endoproteinase-GluC or pepsin. With GluC digestion followed by LC ESI-MS/MS, we increased coverage to 28 biotinylated lysines, but K439, which we identified in our MALDI-ToF study above, was not one of them (see Supplementary Table S1 for additional comments on coverage of unmodified versus biotinylated lysines). When we switched to pepsin, a protease that prefers to cleave

C-terminal to hydrophobic residues, we could identify 39 lysines: 9, 14, 7 and 9 in the unstructured regions, PPR, central and NYN domains, respectively (Supplementary Table S1), and this list includes K439. Because we could achieve ∼90% coverage (39/43 lysines) of *At*PRORP1 with pepsin, we chose to complete this study with pepsin.

From pepsin-generated peptides, we could reliably identify the substrate-contacting residues by comparing the peak intensity of individual peptides containing biotinylated lysines; a peptide without lysines was used for normalization (Supplementary Table S1). As the raw intensities of peptides/fragment ions are comparable, any difference is likely to stem from the difference in solvent accessibility of individual lysines in the presence and absence of the substrate. Comparison of the normalized value for lysine-modified peptides derived from *At*PRORP1, free versus substrate-bound form, allowed us to calculate the protection levels at each lysine (1, no change; >1 or <1, protected from or susceptible to biotinylation upon binding to pre-tRNA^Cys).

Use of MRM-MS to quantitate the protection level of individual lysines

Because data-dependent LC-ESI MS/MS is designed to identify multiple peptides from a large sample pool, not all peptides of interest might be detected and quantitated with accuracy. Moreover, in peptides with two lysines (e.g. $_{100}$LKQKLDM$_{106}$), low numbers of daughter ions preclude pinpointing which lysine was differentially modified by NHSB. Therefore, we used the semi-quantitative ESI MS/MS to obtain a primary inventory of peptides that deserved further consideration by MRM, which accurately determines the protection levels of *individual* lysines (Figure 2A). By exploiting quadrupoles in tandem, MRM isolates a specific protonated peptide, fragments it and transmits only specific product ions for analysis. Thus, MRM, by using multiple precursor-product transitions, quantitates low-abundance peptides with high accuracy (47).

From LC-ESI MS/MS (Supplementary Table S1), lysine-containing peptides with protection levels >1.4 were selected for accurate quantitation by MRM experiments. Using MRM results and a slightly more stringent protection level cut-off >2, we identified 11 lysines as potential pre-tRNA^Cys–contacting sites (Figure 2B; K86, K88, K89 and K90 in the N-terminal unstructured region; K101, K103, K109 and K110 in the PPR domain; K439, K451 and K460 in the NYN domain). The highest protection levels were observed for K101 and K109 (∼27 and 13, respectively; Figure 2B). We did not detect protection of lysines in the central domain. Importantly, our results show that MRM can be used to parse the protection levels of proximal lysines even in a small peptide (Figure 2A).

Substrate-contacting sites are located in PPR and NYN domains of AtPRORP1

An individual lysine in *At*PRORP1 might show decreased biotinylation in the presence of pre-tRNA^Cys because it either makes a direct contact with the substrate or it is present in a region that undergoes conformational changes.
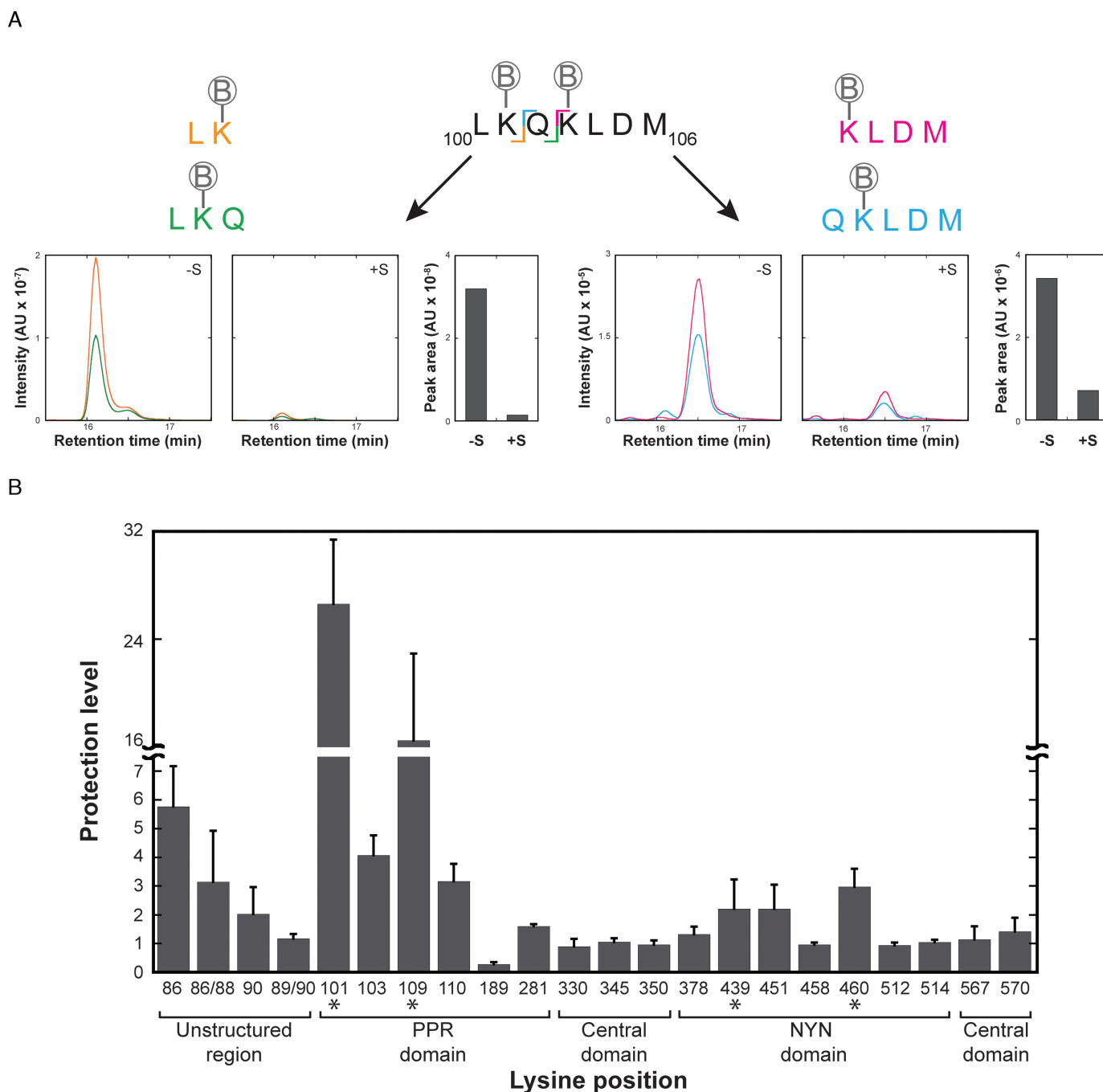
Figure 2. Determination of substrate-contacting lysines in *At*PRORP1. (**A**) Quantitation of two lysines in the fragmented $_{100}$LKQKLDM$_{106}$ peptide by MRM. The peak areas corresponding to individual transitions for biotinylated K101 and K103 were summed and compared in the presence and absence of pre-tRNA^{Cys} to calculate the protection levels; encircled B: biotinylation. For the sake of illustration, a peptide with two biotinylated lysines is depicted; however, MRM analyses were always performed using a peptide with a single biotinylated lysine. (**B**) MS data were used to calculate protection levels of 20 lysines. Values plotted are the mean and standard deviation from three independent MRM experiments. Residues marked with asterisks were characterized further. AU, arbitrary units; S, substrate.

To distinguish between these possibilities, we generated single and double K→A substitutions at four positions that showed high protection levels (Figure 2B), and compared the pre-tRNA^{Cys} cleavage and binding ability of these mutant derivatives relative to the wild-type (WT). If a lysine identified by MS is not a *bona fide* contact residue, its mu-

tation to alanine is not expected to affect either binding or cleavage.

*At*PRORP1 mutant derivatives (K101A, K109A, K439A, K460A, K101A/109A, K101A/K439A and K109A/K439A) were overexpressed in *E. coli* and purified using the same procedure as for the WT. Some control experiments were performed before characterization of

these mutants. First, we confirmed that there was no *E. coli* RNase P present in the these recombinant preparations using reverse transcription PCR, which can identify even 100 fmol of the RNA subunit of *E. coli* RNase P (data not shown). Second, although substitution of K→A (either singly or in pairs) in α-helices was expected to cause minimal structural perturbation, we used circular dichroism (CD) spectroscopy to assess if these mutations cause measurable structural alterations. The CD spectra for the WT *At*PRORP1 and the mutant derivatives are similar (Supplementary Figure S3) and consistent with that expected for a predominantly α-helical protein (15). Our spectra also mirror those reported for *At*PRORP1 and *At*PRORP2 using synchrotron radiation CD (31). Last, before we determined the rate of cleavage, we validated that the mutant derivatives cleaved pre-tRNA^{Cys} at the expected site and exhibited the same fidelity as the WT (Supplementary Figure S4).

Using FP binding assays (Figure 3A), we determined the dissociation constants (K_D values) of *At*PRORP1 WT and mutants for binding to pre-tRNA^{Cys} (5-nt 5′-leader and 23-nt 3′-trailer). For the WT, we obtained a K_D of 250 ± 34 nM; a K_D of 700 nM for pre-tRNA^{Cys} with a 53-nt 5′-leader and a 24-nt 3′-trailer has been reported (15). Compared to *At*PRORP1 WT, all the single substitution mutants have binding defects as evident from the ~1.6- to 3.4-fold increase in their K_D values (Figure 3B and Supplementary Figure S5; Table S2), confirming their substrate-binding role as identified by MS. Among these mutants, *At*PRORP1 K101A has the highest K_D value consistent with its large protection level (Figure 2B). Among the double mutants, *At*PRORP1 K101A/K109A exhibits a 6-fold increase in K_D relative to the WT, and appears to be cumulative of the effects observed with K101A (3.4-fold) and K109A (1.6-fold) (Supplementary Table S2). The K_D determined for *At*PRORP1 K101A/K439A is comparable to that for K101A, while that of K109A/K439A is slightly higher than either K109A or K439A alone.

When we performed the binding assays using pre-tRNA^{Cys} with a 5-nt 3′-trailer (instead of a 23-nt trailer), *At*PRORP1 displayed a K_D of 77 ± 19 nM, indicating that trimming the trailer from 23 to 5 nt decreased the K_D by 3-fold. This decrease might result from the artefact of placing the fluorescein close to the enzyme's active site, a possibility made likely by the observation that shortening the trailer of pre-tRNA^{Gly} from 40 to 0 nt did not affect $K_{M(STO)}$ of *At*PRORP3 (48). Nevertheless, even with this substrate, the K101A or K439A mutants displayed weaker binding relative to the WT (K_D for K101A increased 3.8-fold and K439A increased by 1.7-fold). These results indicate that the trend we observed with the longer-trailer pre-tRNA^{Cys} is maintained (not shown), and that *At*PRORP1 does not contact the long trailer.

K439 in the NYN domain plays a key role in pre-tRNA cleavage

We used time-course assays to study the cleavage of pre-tRNA^{Cys} by *At*PRORP1 or its mutant derivatives under single-turnover conditions (2 μM *At*PRORP1 and <40 nM pre-tRNA) (Figure 4 and Supplementary Figure S6). While
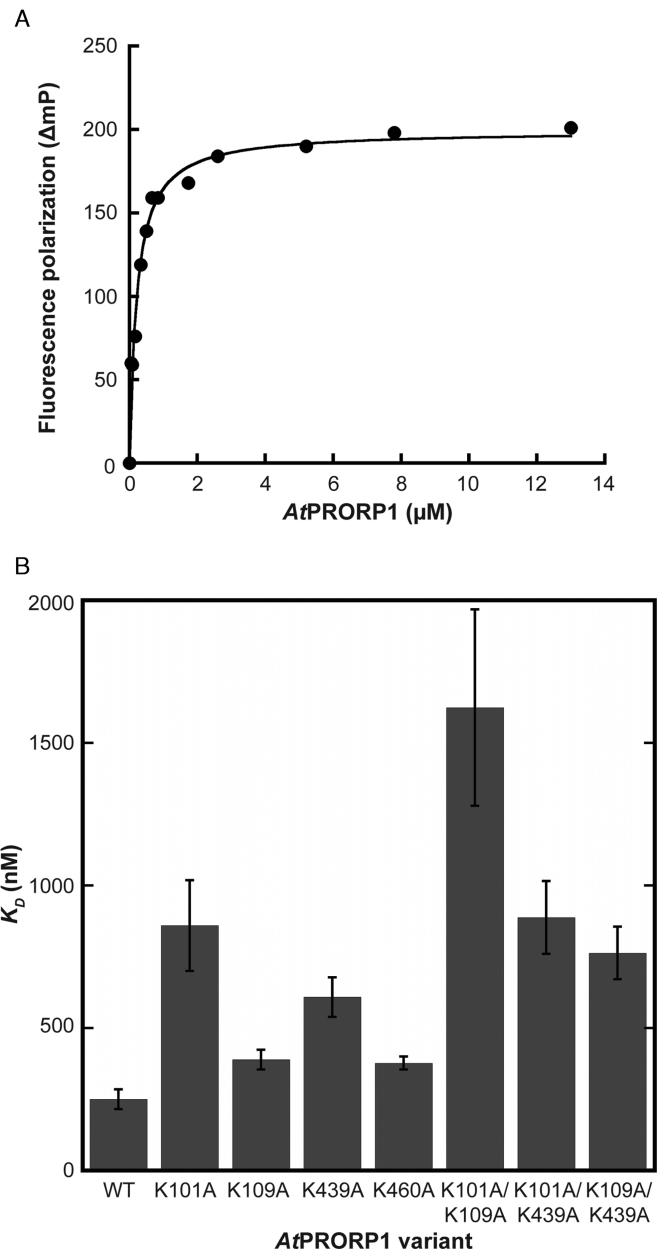
Figure 3. Fluorescence polarization (FP) binding assays. (**A**) Representative binding curves of *At*PRORP1 WT for 3′-FTSC-labeled pre-tRNA^{Cys} (5-nt 5′-leader and 23-nt 3′-trailer). (**B**) Dissociation constants (K_D) of *At*PRORP1 WT and mutants for binding to pre-tRNA^{Cys}. Each K_D value plotted represents the mean and standard deviation calculated from at least three independent experiments.

the observed rates (k_{obs}) of *At*PRORP1 K101A, K109A and K460A are comparable to WT, *At*PRORP1 K439A displays a 3-fold decrease (Figure 4C; Supplementary Table S3). A similar decrease is evident in the k_{obs} of the two *At*PRORP1 double mutants that contain K439A. Also, while there was no difference in the k_{obs} of *At*PRORP1 WT, K101A or K109A, the double mutant K101A/K109A shows an unanticipated modest *increase* in rate (2.1-fold; Supplementary Table S3).
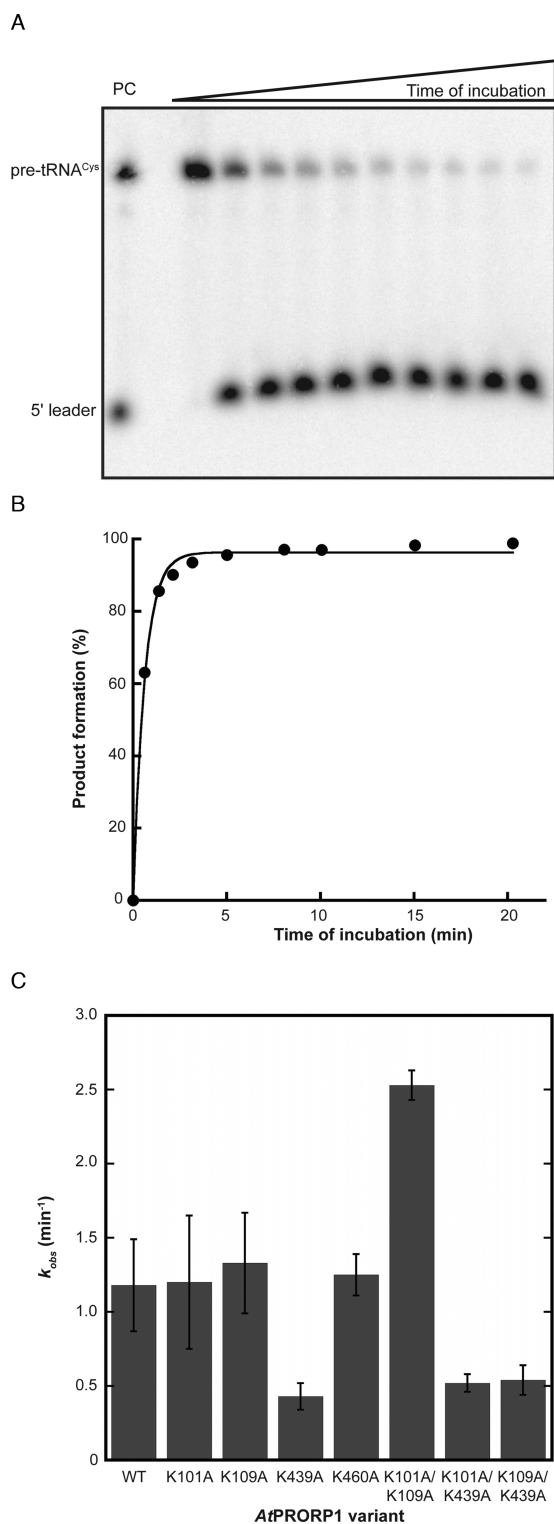
A

B

C

Figure 4. Cleavage of pre-tRNA^{Cys} by *At*PRORP1 WT and mutants. (**A**) Representative time course and (**B**) curve fitting of *At*PRORP1 WT-mediated cleavage of 5′-[^{32}P]-labeled pre-tRNA^{Cys} (5-nt 5′-leader and 0-nt 3′-trailer). PC, positive control, cleavage by *E. coli* RNase P. (**C**) Observed rates of *At*PRORP1 WT and mutants for pre-tRNA^{Cys} cleavage. Rates plotted are the mean and standard deviation calculated from at least three independent experiments. Pre-tRNA cleavage was assessed using an 8% (w/v) polyacrylamide/7 M urea gel.

DISCUSSION

Mapping RNA–protein interactions using MRM mass spectrometry-aided footprinting

NHSB modification and mass spectrometry have been used successfully to map nucleic acid–protein interactions (39,40,42), e.g. between HIV1 integrase and its DNA substrate (41) and between Cbf5 (a pseudouridine synthase) and the H/ACA guide RNA (38). Here, we made some important changes to this method while mapping the *At*PRORP1–pre-tRNA^{Cys} interface. We have demonstrated the utility of MRM MS for accurate quantitation, and documented the influence of protease (pepsin instead of trypsin) and ionization method (ESI *in lieu* of MALDI) on overall sequence coverage. Moreover, given the preponderance of lysines in RNA-binding proteins, the application of MRM MS to parse the specific contributions of proximal lysines, which are typically inaccessible, should promote wider use of this footprinting method for mapping contact sites in RNP complexes. Efficient modification of other residues [e.g. arginines (42,49)] under conditions that are compatible with stable RNP complexes will add appeal.

Substrate contacting sites in *At*PRORP1: implications for catalysis

MRM MS helped us quantitate the decreased accessibility of 11 lysines in *At*PRORP1 upon binding to pre-tRNA^{Cys} (Figure 2): four in the N-terminal unstructured region, four in the PPR domain, none in the central domain and three in the NYN domain (Figure 5). This finding suggests that lysines in the PPR and NYN domains are used to engage in electrostatic interactions with the substrate. Indeed, our sequence analysis of 60 *At*PRORP1 homologs (Figure 6) confirms that many of these lysines, although not universally conserved, are often substituted by arginines. The protection levels of the lysines identified in the central domain are ∼1 (Figure 2 and Supplementary Table S1). Although we cannot completely rule it out, the central domain does not appear to be directly involved in substrate binding, reaffirming its role as a structural element that enables functional coupling of PPR and NYN domains (15). The protection levels of the lysines in the PPR domain are higher than those in the NYN domain, indicating importance of the PPR domain for substrate recognition.

Solvent accessibility considerations preclude a direct correlation between protection levels and contribution to substrate binding. Changes in biotinylation observed in the presence of pre-tRNA^{Cys} could also arise from conformational changes. To examine if such structural changes (and not *bona fide* contact sites) led to protection levels >1, we employed mutagenesis studies (see below). However, there are instances where a protection level <1, reflecting increased NHSB reactivity, was observed. For example, K189 (in PPR3), whose side-chain projects to the other side of the substrate-binding concave cleft (Figure 5), becomes more solvent accessible in the presence of pre-tRNA^{Cys} and exhibits a protection level of 0.3 (Supplementary Table S1, Figure 2B).

*At*PRORP1 has 5.5 PPR motifs in its PPR domain. Each PPR motif (∼35 aa) adopts a helix-turn-helix fold, with the
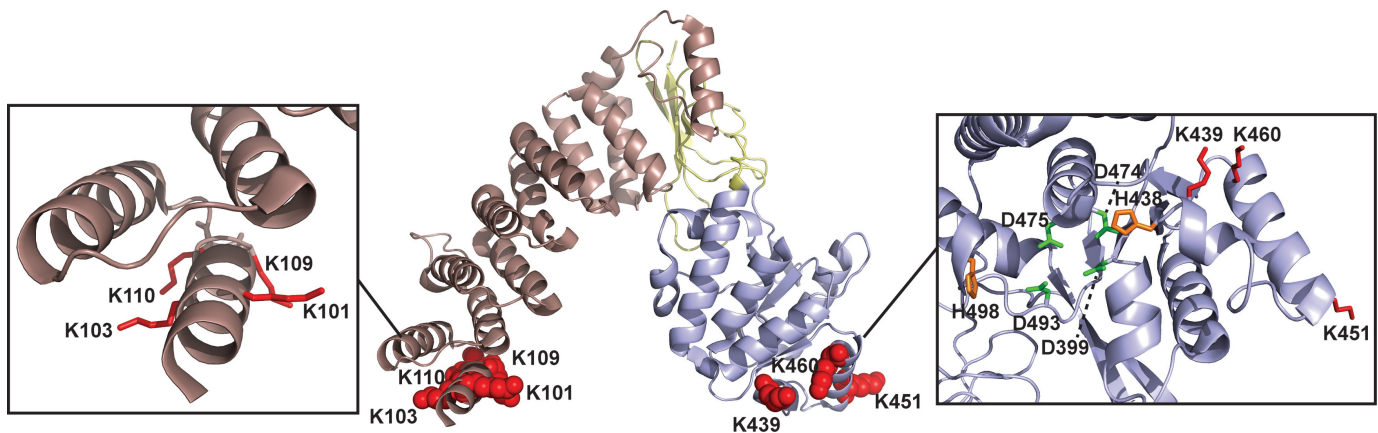
Figure 5. Map of substrate-contacting residues in *At*PRORP1 structure. Three structural domains are depicted in three different colors: pentatricopeptide repeat (PPR), central and Nedd4-BP1 YacP nuclease (NYN) domains in brown, yellow and light blue, respectively. Lysines with protection levels >2, as determined by MRM (Figure 2B), are shown as red spheres (or as sticks in enlarged inset boxes). The right panel also depicts in green the four aspartates that coordinate the Mg^{2+} for cleaving the scissile phosphodiester bond, and in orange the two histidines proposed to be important for maintaining the local fold.
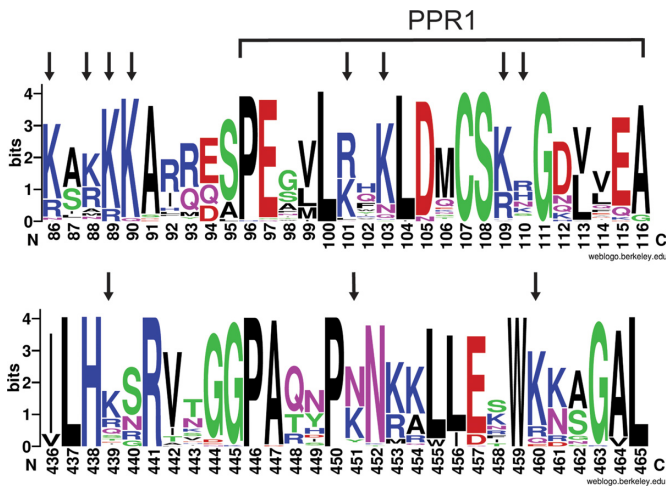

Figure 6. Location of substrate-contacting lysines in a sequence alignment of 60 *At*PRORP1 homologs. These sequences were the top-scoring hits in NCBI BlastP when *At*PRORP1 was used as the query. Black arrows indicate the substrate-contacting lysines at positions 86, 88, 89, 90, 101, 103, 109, 110, 439, 451 and 460 that were identified in our experiments. Top panel: Logo map of the N-terminal unstructured region (positions 86 to 95) and part of PPR1 (positions 96 to 115). Bottom panel: Logo map of the NYN domain (positions 436 to 455). Logo maps were generated using the web-based Berkeley WebLogo application (http://weblogo.berkeley.edu/logo.cgi).

two helices in anti-parallel orientation. Sequence analyses uncovered an RNA recognition code in which bases in the target RNA are specifically recognized by amino acids at positions 4 and 34 in the PPR motif (e.g. T4 and N34 help recognize an adenine) (23–27); this code has been validated by binding studies that tested reciprocal recoding (23) and structural studies (25,26,28). Examining the sequence conservation and structure of *At*PRORP1 led to the idea that PPRs 2 to 4 might hold the key for pre-tRNA recognition, and indeed mutagenesis of the 4th position in PPR2 and PPR3 did result in a modest 1.6- to 3-fold decrease in activity (30). Additionally, a recent study reported that mutating

positions 4 and 34 in PPR3 of *At*PRORP3 led to a 100-fold decrease in k_{obs} for pre-tRNA cleavage (48). However, due to the absence of lysines in PPR2 and PPR3 of *At*PRORP1, the NHSB approach could not test the substrate-binding roles of PPR2 and PPR3. The four substrate-contacting lysines that we identified are in PPR1 − K101 and K103 in the middle of the first helix and K109 and K110 in the loop that connects the first and second helix (Figure 5). Our binding assays showed that K101A, K109A and K101A/K109A exhibit a 3.2-, 1.5- and 6-fold higher K_D value compared to the WT and establishes an important role for PPR1 for substrate binding. Some previous findings support this inference. Deletion of PPR1 and part of PPR2 in *O. tauri* PRORP eliminated its activity (10). A single-endpoint assay showed that deletion of the entire PPR1 in *At*PRORP1 led to a 20-fold decrease in its activity (30); although structural alterations in this mutant were not ruled out, it was postulated that PPR1 might serve as a capping/stabilizing helix for PPR 2–4. Another explanation might be that PPR1 anchors the substrate for cleavage.

As the 4th residue of PPR1 in the inner concave surface of the V-shaped *At*PRORP1, K101 offers an interesting extension to the PPR recognition code. While sequence conservation was used to support the claim that PPR2 and PPR3 might be critical, our results suggest that PPR1 uses a positively charged residue (K or R) at the 4th position to help bind the substrate through an electrostatic interaction. Such a model is consistent with inferences from high-resolution structures of PPR domains with their RNA ligands that show some code-independent recognition and thematic variations for some of the motifs (25,26,28). It remains to be seen how K101 might network with the 4th residue in PPR2 (N136) and PPR3 (T180), given that these three residues all project into the substrate-binding concave cleft (Figure 5). Although N136 and T180 were proposed to contact C56 and A57, respectively, in the T-loop of pre-tRNAs (30), the recognition of A57 by PPR3 in *At*PRORP3 could not be established using reciprocal recoding experiments with either pre-tRNA or minimal substrates (48).

In the NYN domain, we reasoned that K439, K451 and K460 might contact pre-tRNA^{Cys} given their NHSB-modification protection levels >2; the ε-nitrogen atoms of K439 and K460 are only ∼5 Å apart (PDB: 4G23). Indeed, K439A and K460A showed a ∼2-fold binding defect compared to WT, and K439A also has a ∼3-fold lower cleavage rate (Figures 3 and 4). In another study, *At*PRORP1 H438A was found to have a 2-fold lower k_{obs} without affecting substrate binding (15). It is possible that there might be other factors including different assay conditions which might underlie the substrate-binding defect of K439A (this study) but not H438A (15). Taken together, H438, K439 and K460, which are proximal to the highly conserved aspartates essential for coordinating catalytically-important Mg^{2+} (Figure 5, D399, D474, D475 and D493), through electrostatic interactions facilitate contacts with the 5′-leader and position the cleavage site proximal to key Mg^{2+} ions. K439 in the NYN domain is important for pre-tRNA cleavage.

Substrate recognition by the two classes of RNase P

Our results and those from other recent reports (20,48) allow important conclusions to be drawn in terms of similarities and differences between the two RNase P variants. Foremost, the mode of functioning and overall architecture of the two variants might largely reflect how two different catalytic scaffolds coevolved with a highly conserved element of the tRNA structure, namely the 'elbow' region (the tertiary contact between the D- and T-loop) (20,48,50,51) (Mao, Chen, Gopalan and Krisebom, in preparation). It has been pointed out that a defining attribute of the tRNA structure is the elbow, which has been exploited as a recognition determinant by different cellular RNAs (rRNA, RNase P, riboswtiches) and proteins that interact with the tRNA (50,51).

Second, the recognition of this elbow in the pre-tRNA must be coupled to cleavage of the pre-tRNA between −1 (last position of the 5′ leader) and +1 (first position of the mature tRNA) (48). The crystal structure of the RNase P ribozyme reveals a bimodular structure comprising a specificity (S) domain and a catalytic (C) domain, with the S domain using inter-digitated T loops to recognize the elbow (52). Biochemical studies show how a productive interaction between the pre-tRNA and the S domain might lead to an induced-fit that positions the chemical groups/Mg^{2+} in the C domain to cleave the substrate (53,54). Similarly, in the case of PRORP, the PPR domain might be responsible for recognizing the elbow (or its structural variant) and signaling the NYN domain to cleave. Evidence for such functional crosstalk between the PPR and NYN domains is indicated by our result that weakening PPR interactions with pre-tRNA^{Cys} (as in the case of K101A/K109A) enhances the rate of cleavage by ∼2-fold (Figure 4; Supplementary Table S3). Interestingly, mutations in the bacterial RNase P RNA's S domain affect substrate binding, cleavage-site selection and either increase or decrease the rate of cleavage in a substrate identity-dependent manner (53). The increased rates associated with weaker ground state binding observed with both forms of RNase P (41) might reflect a trade-off between selectivity and catalytic efficiency (55).

Last, our results demonstrate the importance of electrostatic interactions for substrate recognition and cleavage by PRORP. While this inference is not unexpected given that such substrate steering is commonly used to decrease futile encounters in protein-based catalysts (56,57), the RNase P RNA alone cannot exploit this strategy as it cleaves another polyanion (i.e. the pre-tRNA). Instead, the ribozyme relies on non-covalent interactions, notably Watson-Crick base-pairing with the substrate; for electrostatic interactions, it utilizes the protein cofactor (4,52,54,58). These differences have implications for their respective mechanisms of action (i.e. whether product release is rate limiting (14)), and illustrate how RNA- and protein-based scaffolds perform the same reaction with substrate recognition and catalysis strategies that exploit their respective building blocks.

The ability of PRORPs to replace *in vivo* the two-subunit RNase P RNP in *E. coli* or the 10-subunit RNP in *S. cerevisiae* (9,12,17) has inspired the notion that the different forms of RNase P might indicate constructive neutral evolution (or a form of 'runaway bureaucracy' (17,59,60)) and that functional gains might not have been the drivers of evolutionary complexity of RNase P. Although additional studies are needed, parallels in substrate recognition by different RNase P scaffolds provide support for this premise.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Altman,S. (2007) A view of RNase P. *Mol. Biosyst.*, **3**, 604–607.
2. Esakova,O. and Krasilnikov,A.S. (2010) Of proteins and RNA: the RNase P/MRP family. *RNA*, **16**, 1725–1747.
3. Evans,D., Marquez,S.M. and Pace,N.R. (2006) RNase P: interface of the RNA and protein worlds. *Trends Biochem. Sci.*, **31**, 333–341.
4. Lai,L.B., Vioque,A., Kirsebom,L.A. and Gopalan,V. (2010) Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett.*, **584**, 287–296.

5. Rossmanith,W. (2012) Of P and Z: mitochondrial tRNA processing enzymes. *Biochim. Biophys. Acta*, **1819**, 1017–1026.

6. Howard,M.J., Liu,X., Lim,W.H., Klemm,B.P., Fierke,C.A., Koutmos,M. and Engelke,D.R. (2013) RNase P enzymes: divergent scaffolds for a conserved biological reaction. *RNA Biol.*, **10**, 909–914.

7. Holzmann,J., Frank,P., Loffler,E., Bennett,K.L., Gerner,C. and Rossmanith,W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*, **135**, 462–474.

8. Lechner,M., Rossmanith,W., Hartmann,R.K., Tholken,C., Gutmann,B., Giege,P. and Gobert,A. (2015) Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol. Biol. Evol.*, **32**, 3186–3193.

9. Gobert,A., Gutmann,B., Taschner,A., Gossringer,M., Holzmann,J., Hartmann,R.K., Rossmanith,W. and Giege,P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.*, **17**, 740–744.

10. Lai,L.B., Bernal-Bayard,P., Mohannath,G., Lai,S.M., Gopalan,V. and Vioque,A. (2011) A functional RNase P protein subunit of bacterial origin in some eukaryotes. *Mol. Genet. Genomics*, **286**, 359–369.

11. McClain,W.H., Lai,L.B. and Gopalan,V. (2010) Trials, travails and triumphs: an account of RNA catalysis in RNase P. *J. Mol. Biol.*, **397**, 627–646.

12. Taschner,A., Weber,C., Buzet,A., Hartmann,R.K., Hartig,A. and Rossmanith,W. (2012) Nuclear RNase P of Trypanosoma brucei: a single protein in place of the multicomponent RNA-protein complex. *Cell Rep.*, **2**, 19–25.

13. Sugita,C., Komura,Y., Tanaka,K., Kometani,K., Satoh,H. and Sugita,M. (2014) Molecular characterization of three PRORP proteins in the moss Physcomitrella patens: nuclear PRORP protein is not essential for moss viability. *PLoS One*, **9**, e108962.

14. Howard,M.J., Klemm,B.P. and Fierke,C.A. (2015) Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.*, **290**, 13454–13464.

15. Howard,M.J., Lim,W.H., Fierke,C.A. and Koutmos,M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16149–16154.

16. Pavlova,L.V., Gossringer,M., Weber,C., Buzet,A., Rossmanith,W. and Hartmann,R.K. (2012) tRNA processing by protein-only versus RNA-based RNase P: kinetic analysis reveals mechanistic differences. *Chembiochem*, **13**, 2270–2276.

17. Weber,C., Hartig,A., Hartmann,R.K. and Rossmanith,W. (2014) Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet.*, **10**, e1004506.

18. Gutmann,B., Gobert,A. and Giege,P. (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.*, **26**, 1022–1027.

19. Zhou,W., Karcher,D., Fischer,A., Maximova,E., Walther,D. and Bock,R. (2015) Multiple RNA processing defects and impaired chloroplast function in plants deficient in the organellar protein-only RNase P enzyme. *PLoS One*, **10**, e0120533.

20. Howard,M.J., Karasik,A., Klemm,B.P., Mei,C., Shanmuganathan,A., Fierke,C.A. and Koutmos,M. (2016) Differential substrate recognition by isozymes of plant protein-only ribonuclease P. *RNA*, **22**, 782–792.

21. Barkan,A. and Small,I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442.

22. Filipovska,A. and Rackham,O. (2013) Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. *RNA Biol.*, **10**, 1426–1432.

23. Barkan,A., Rojas,M., Fujii,S., Yap,A., Chong,Y.S., Bond,C.S. and Small,I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.*, **8**, e1002910.

24. Yagi,Y., Hayashi,S., Kobayashi,K., Hirayama,T. and Nakamura,T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*, **8**, e57286.

25. Yin,P., Li,Q., Yan,C., Liu,Y., Liu,J., Yu,F., Wang,Z., Long,J., He,J., Wang,H.W. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168–171.

26. Coquille,S., Filipovska,A., Chia,T., Rajappa,L., Lingford,J.P., Razif,M.F., Thore,S. and Rackham,O. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.*, **5**, 5729.

27. Takenaka,M., Zehrmann,A., Brennicke,A. and Graichen,K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*, **8**, e65343.

28. Ke,J., Chen,R.Z., Ban,T., Zhou,X.E., Gu,X., Tan,M.H., Chen,C., Kang,Y., Brunzelle,J.S., Zhu,J.K. *et al.* (2013) Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.*, **20**, 1377–1382.

29. Anantharaman,V. and Aravind,L. (2006) The NYN domains: novel predicted RNAses with a PIN domain-like fold. *RNA Biol.*, **3**, 18–27.

30. Imai,T., Nakamura,T., Maeda,T., Nakayama,K., Gao,X., Nakashima,T., Kakuta,Y. and Kimura,M. (2014) Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TpsiC loop in precursor tRNAs. *Biochem. Biophys. Res. Commun.*, **450**, 1541–1546.

31. Gobert,A., Pinker,F., Fuchsbauer,O., Gutmann,B., Boutin,R., Roblin,P., Sauter,C. and Giege,P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat. Commun.*, **4**, 1353.

32. Tsai,H.Y., Lai,L.B. and Gopalan,V. (2002) A Modified pBluescript-based vector for facile cloning and transcription of RNAs. *Anal. Biochem.*, **303**, 214–217.

33. Pagano,J.M., Clingman,C.C. and Ryder,S.P. (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA*, **17**, 14–20.

34. Qiu,C., Liu,W.-Y. and Xu,Y.-Z. (2015) Fluorescence labeling of short RNA by oxidation at the 3′-end. *Methods Mol. Biol.*, **1297**, 113–120.

35. Link,A.J., Eng,J., Schieltz,D.M., Carmack,E., Mize,G.J., Morris,D.R., Garvik,B.M. and Yates,J.R. 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.*, **17**, 676–682.

36. Yates,J.R. 3rd (2004) Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 297–316.

37. MacLean,B., Tomazela,D.M., Shulman,N., Chambers,M., Finney,G.L., Frewen,B., Kern,R., Tabb,D.L., Liebler,D.C. and MacCoss,M.J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, **26**, 966–968.

38. Baker,D.L., Seyfried,N.T., Li,H., Orlando,R., Terns,R.M. and Terns,M.P. (2008) Determination of protein-RNA interaction sites in the Cbf5-H/ACA guide RNA complex by mass spectrometric protein footprinting. *Biochemistry*, **47**, 1500–1510.

39. Kvaratskhelia,M. and Grice,S.F. (2008) Structural analysis of protein-RNA interactions with mass spectrometry. *Methods Mol. Biol.*, **488**, 213–219.

40. McKee,C.J., Kessl,J.J., Norris,J.O., Shkriabai,N. and Kvaratskhelia,M. (2009) Mass spectrometry-based footprinting of protein-protein interactions. *Methods*, **47**, 304–307.

41. Zhao,Z., McKee,C.J., Kessl,J.J., Santos,W.L., Daigle,J.E., Engelman,A., Verdine,G. and Kvaratskhelia,M. (2008) Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J. Biol. Chem.*, **283**, 5632–5641.

42. Mendoza,V.L. and Vachet,R.W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.*, **28**, 785–815.

43. Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.

44. Ellis,J.J., Broom,M. and Jones,S. (2007) Protein-RNA interactions: structural analysis and functional classes. *Proteins*, **66**, 903–911.

45. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.

46. Rosta,E., Yang,W. and Hummer,G. (2014) Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J. Am. Chem. Soc.*, **136**, 3137–3144.

47. Lange,V., Picotti,P., Domon,B. and Aebersold,R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.*, **4**, 222.

48. Brillante,N., Gossringer,M., Lindenhofer,D., Toth,U., Rossmanith,W. and Hartmann,R.K. (2016) Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.*, **44**, 2323–2336.

49. Akinsiku,O.T., Yu,E.T. and Fabris,D. (2005) Mass spectrometric investigation of protein alkylation by the RNA footprinting probe kethoxal. *J. Mass Spectrom.*, **40**, 1372–1381.

50. Zhang,J. and Ferre-D'Amare,A.R. (2016) The tRNA Elbow in Structure, Recognition and Evolution. *Life (Basel)*, **6**, 3.

51. Chan,C.W., Chetnani,B. and Mondragon,A. (2013) Structure and function of the T-loop structural motif in noncoding RNAs. *Wiley Interdiscip. Rev. RNA*, **4**, 507–522.

52. Reiter,N.J., Osterman,A., Torres-Larios,A., Swinger,K.K., Pan,T. and Mondragon,A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784–789.

53. Brannvall,M., Kikovska,E., Wu,S. and Kirsebom,L.A. (2007) Evidence for induced fit in bacterial RNase P RNA-mediated cleavage. *J. Mol. Biol.*, **372**, 1149–1164.

54. Kirsebom,L.A. and Trobro,S. (2009) RNase P RNA-mediated cleavage. *IUBMB Life*, **61**, 189–200.

55. Tawfik,D.S. (2014) Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.*, **21**, 73–80.

56. Bar-Even,A., Milo,R., Noor,E. and Tawfik,D.S. (2015) The moderately efficient enzyme: futile encounters and enzyme floppiness. *Biochemistry*, **54**, 4969–4977.

57. Harel,M., Spaar,A. and Schreiber,G. (2009) Fruitful and futile encounters along the association reaction between proteins. *Biophys. J.*, **96**, 4237–4248.

58. Kirsebom,L.A. and Svard,S.G. (1994) Base pairing between Escherichia coli RNase P RNA and its substrate. *EMBO J.*, **13**, 4870–4876.

59. Gray,M.W., Lukes,J., Archibald,J.M., Keeling,P.J. and Doolittle,W.F. (2010) Cell biology. Irremediable complexity? *Science*, **330**, 920–921.

60. Stoltzfus,A. (2012) Constructive neutral evolution: exploring evolutionary theory's curious disconnect. *Biol. Direct*, **7**, 35.