Biochemistry

Large-Scale Photolithographic Synthesis of Chimeric DNA/RNA Hairpin Microarrays To Explore Sequence Specificity Landscapes of RNase HII Cleavage

Jory Lietard,*,[†] Masad J. Damha,[‡] and Mark M. Somoza*,[†]

[†]Institute of Inorganic Chemistry, Faculty of Chemistry, University of Vienna, Althanstraße 14 (UZA II), 1090 Vienna, Austria
[‡]Department of Chemistry, McGill University, 801 Rue Sherbrooke Ouest, Montreal, QC H3A 0B8, Canada

S *Supporting Information*

ABSTRACT: Ribonuclease HII (RNase HII) is an essential endoribonuclease that binds to double-stranded DNA with RNA nucleotide incorporations and cleaves 5′ of the ribonucleotide at RNA−DNA junctions. Thought to be present in all domains of life, RNase HII protects genomic integrity by initiating excision repair pathways that protect the encoded information from rapid degradation. There is sparse evidence that the enzyme cleaves some substrates better than others, but a large-scale study is missing. Such large-scale studies can be carried out on microarrays, and we employ chemical photolithography to synthesize very large combinatorial libraries of fluorescently labeled DNA/RNA chimeric sequences that self-anneal to form hairpin structures that are substrates for *Escherichia coli* RNase HII. The relative activity is determined by the loss of fluorescence upon cleavage. Each substrate includes a double-stranded 5 bp variable region with one to five consecutive ribonucleotide substitutions. We also examined the effect of all possible single and double mismatches, for a total of >9500 unique structures. Differences in cleavage efficiency indicate some level of substrate preference, and we identified the 5′-dC/rC-rA-dX-3′ motif in well-cleaved substrates. The results significantly extend known patterns of RNase HII sequence specificity and serve as a template using large-scale photolithographic synthesis to comprehensively map landscapes of substrate specificity of nucleic acid-processing enzymes.
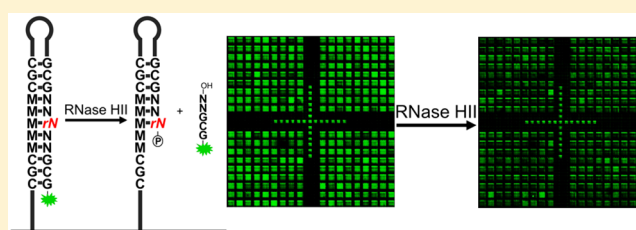
The presence of RNA in genomic DNA was long thought to be limited to the short, transient RNA primers serving as an initiating site of DNA polymerization during DNA replication, but there is evidence that RNA nucleotides may also be misincorporated by DNA polymerases in this process.[1,2] Misincorporation of RNA during DNA replication appears to be a common, widespread phenomenon occurring in bacteria as well as in eukaryotic organisms and is likely caused, or at least influenced, by a strong excess of cellular ribonucleoside monophosphate (rNMP) over deoxyribonucleoside monophosphate,[3,4] and recent work suggests that misincorporation of RNA nucleotides is actually a highly frequent error, with reports counting anywhere from 2000 rNMP insertions per replication cycle in prokaryotes to 1 million in higher eukaryotes.[1,5,6] The introduction of a 2′-OH group within DNA creates genomic instability as the DNA is now more susceptible to degradation, but the ribonucleotide excision repair (RER) mechanism very efficiently removes unwelcome rNMPs,[3,7−9] thereby safeguarding the integrity of genomic DNA and, at the same time, revealing why rNMP misincorporation had remained largely unnoticed.[10] In RER, the RNase H type 2 enzyme (labeled RNase H2 in eukaryotes and HII in prokaryotes) first cleaves the phosphodiester bond 5′ to the RNA insert and the cleaved 5′ strand is then extended with DNA polymerase δ, leading to displacement of the RNA-containing strand. The displaced region carrying the RNA

misincorporation, or flap, is cleaved by the enzyme Fen1, and the two DNA-only strands are then joined together by ligation.[11−13] Failure to remove ribonucleotides from the newly synthesized DNA leaves DNA open to single- and double-strand breaks, aberrant recombination, mutagenesis and slows DNA replication further.[5,14−16] The absence of RNase H type 2 or a dysfunctional version of it has severe consequences. The absence of RNase H2 in mice is associated with embryonic lethality,[6] while mutations in the RNase H2-coding gene in humans cause Aicardi-Goutières syndrome (AGS), a neurological disorder gravely affecting the brain.[17] In bacteria, however, mutation of the rnhB gene encoding the enzyme seems to be better tolerated by the organism.[5,18,19] While the RNase H type 2 enzymes both cleave single RNA inserts within double-stranded DNA 5′ to the 5′-RNA−DNA-3′ junction, they also recognize and cleave longer stretches of RNA, such as the RNA−DNA fragments assembled during DNA replication, Okazaki fragments, leaving behind the copied DNA strand with a single RNA nucleotide at the 5′ end.[11,20,21] This enzymatic activity was coined junction ribonuclease.[22,23] Importantly, RNase HII enzymes poorly

process pure RNA:DNA hybrids,[11,24] which are substrates of RNase H type 1/I, as they indeed require the presence of at least one paired DNA nucleotide 3′ to the last RNA base. The enzyme is a monomeric structure in prokaryotes[25] and trimeric in eukaryotic RNase H2.[26] At the catalytic center of RNase HII, the 2′-hydroxyl group contacts the side chains of three highly conserved amino acids, with coordination of a Mg^{2+} ion to the bridging oxygen atoms of the phosphodiester bonds 5′ and 3′ to the RNA insert.[27] The RNA−DNA junction is sensed with a tyrosine side chain stacking with the deoxyribose sugar 3′ to the RNA nucleotide. This feature of the binding interaction alone suffices to understand why RNA-only strands are not substrates of bacterial RNase HII.

It might seem counterintuitive to envisage the existence of a sequence preference in RNase H enzymes, given that their apparent function is to correct for any type of RNA error, yet a certain amount of data puts forth the idea that RNase H types 1 and 2 can process some substrate sequences better than others. For instance, some sequence preference in RNase H type 1-mediated cleavage was previously mentioned[20] and recently investigated and uncovered.[28−30] Also, RNase H2 cannot process abasic and oxidized incorporations, indicating that they are quite sensitive to details of nucleobase structure and not just the presence of a 2′-OH.[31] In RNase H2, it was originally found that susceptibility to RNase H2-mediated degradation in double-stranded DNA modified with a single RNA monomer followed the order rA > rU > rC > rG.[21] In the context of *in situ* synthesis of a high-density RNA microarray,[32] we also started to address the question of sequence specificity in the enzymatic cleavage by *Escherichia coli* RNase HII.[33] We found that for single RNA inserts, RNase HII better processes substrates containing rC as the RNA modification as well as, interestingly, those carrying a dC 5′ to the RNA insert. We now wish to expand on these findings and conduct a deeper analysis of the *E. coli* RNase HII sequence specificity by including all possible stretches of ribonucleotides two to five nucleotides in length, as well as all possible single and double mismatches in the vicinity of the cleavage site.

■ MATERIALS AND METHODS

Nucleic Acid Photolithography. Our current protocol for microarray synthesis by photolithography is the sum of recent technical improvements over the standard of manufacture.[33−39] Combined photolithography and *in situ* DNA and RNA synthesis using phosphoramidite chemistry can be described as follows. In the paired array system, two microscope slides (Schott Nexterion Glass D) are used for a single synthesis. One of the two slides is first drilled at two locations with a 0.9 mm diamond bit with a CNC router (Stepcraft), rinsed with deionized water in an ultrasonic bath for 30 min, and then dried. Slides, drilled and nondrilled, are then silanized with *N*-(3-triethoxysilylpropyl)-4-hydroxybutyramide (10 g, Gelest SIT8189.5) by being submerged in a 500 mL solution of a 95:5 EtOH/H_2O mixture with 1% AcOH for 4 h at room temperature. The slides are then rinsed with 2 × 500 mL of a 95:5 EtOH/H_2O mixture with 1% AcOH for 20 min, cured overnight in a vacuum oven preheated at 120 °C, and then stored in a desiccator at room temperature until further use. A drilled slide and a nondrilled slide are then assembled in a synthesis cell, separated by a 50 μm thick PTFE gasket, which is then attached to an Expedite 8909 DNA Synthesizer (PerSeptive Biosystems). The DNA synthesizer controls the delivery of all reagents and solvents to the

synthesis cell and follows standard synthesis protocols. The cell is fixed at the focal plane of incoming 365 nm ultraviolet (UV) light, which will trigger the removal of the photosensitive nitrophenylpropoxycarbonyl (NPPOC) protecting group at the 5′ end of the growing oligonucleotide strand. UV light is generated by a high-power UV light-emitting diode (Nichia NVSU333A), is spatially homogenized, and then reaches a Digital Micromirror Device (DMD) consisting of 1024 × 768 individually addressable mirrors 14 μm in size (Texas Instruments). The DMD is electronically controlled by a computer that uses the generated masks to command the proper tilting of micromirrors in the DMD. ON mirrors, corresponding to white pixels in the masks, will reflect the incoming UV light onto the synthesis area of the glass slides in the cell. OFF mirrors, corresponding to black pixels in the masks, will reflect UV light away from the glass slides. During UV deprotection, the slides are immersed in a 1% (w/w) solution of imidazole in DMSO (Biosolve), and the exposure proceeds for 70 s at a radiant power of ∼85 mW/cm^2, yielding a radiant energy density of 6 J/cm^2.

Besides the basic exposure solvents, other solvents and reagents are standards of automated DNA synthesis: activator (0.25 M 4,5-dicyanoimidazole in acetonitrile, Biosolve), dry ACN (<30 ppm of H_2O), and oxidizer (20 mM I_2 in a pyridine/THF/H_2O mixture, Sigma-Aldrich). The coupling step lasts 15 s for DNA phosphoramidites [protected with *tert*-butylphenoxyacetyl protecting group (tac) for dA, *i*PrPac for dG, and isobutyryl for dC, FlexGen], 2 min for rU, and 5 min for rA, rC, and rG phosphoramidites. RNA phosphoramidites are protected at the 5′ end with NPPOC, at the 2′-OH with an acetal levulinyl ester (ALE), and at the nucleobase with levulinyl for rC and rA and with dimethylformamidine (dmf) for rG. RNA 2′-*O*-ALE phosphoramidites were prepared by ChemGenes according to published procedures.[32] DNA and RNA phosphoramidites are diluted to 30 mM in ACN prior to microarray synthesis. After coupling, a capping step is introduced whereby 5′-dimethoxytrityl (DMTr) dT phosphoramidite (30 mM in ACN, Sigma-Aldrich) is allowed to couple for 60 s. Because microarray photolithography does not require the use of an acidic solution to deblock the 5′ end of the oligonucleotide before the next coupling event, coupling with DMTr-dT can essentially be regarded as capping of the oligonucleotide strands that failed to couple with the previous NPPOC DNA or RNA amidite. A short (3 s) oxidation step is then performed before proceeding with UV illumination and the beginning of the next cycle.

Before synthesis of the hairpin sequences, a T_{20} linker is first synthesized on the entire synthesis area of the glass slides. After synthesis of the hairpin sequences, the interstitial space between features is passivated by first removing NPPOC groups and then coupling with DMTr-dT phosphoramidite. The last synthesis cycle is the terminal labeling of the hairpins with Cy3 phosphoramidite (Link Technologies). Cy3 amidite is freshly diluted into dry ACN as a 50 mM solution and then coupled to 5′-OH oligonucleotide termini for 2 × 300 s.

Chemical Deprotection. After synthesis, the nucleobase, 2′-OH, and phosphate protecting groups must be removed from the ribo- and deoxyribonucleotides. First, the cyanoethyl group on the phosphates is cleaved in a 2:3 solution of anhydrous triethylamine in acetonitrile (90 min at room temperature in a 50 mL Falcon tube with gentle agitation). After being rinsed twice in acetonitrile (20 mL in a Falcon tube), the arrays are dried in a centrifuge and then transferred
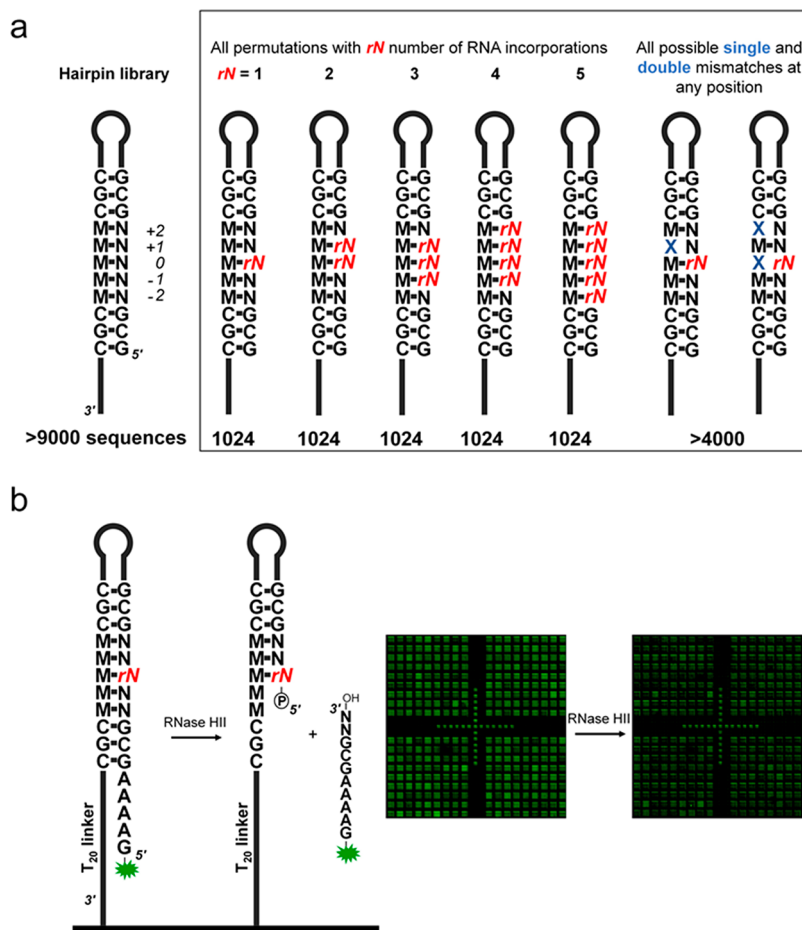
Figure 1. (a) Library hairpin design. The loop is a DNA TCCT tetranucleotide. Single and double mismatches have been introduced on two sequence templates: 5′-GCrCC and 5′-AArUAA. (b) Schematic representation of the outcome of enzymatic cleavage of Cy3-labeled hairpins (left). Small scan excerpt (≈0.5% of the total synthesis area) of the hairpin library before and after cleavage with RNase HII (right).

into a 0.5 M solution of hydrazine hydrate (1.2 mL) in a 3:2 pyridine/acetic acid mixture (50 mL in a Falcon tube for 2 h at room temperature) to remove the protecting groups on the 2′-OH and RNA. After another washing and drying step (as above), a final deprotection step in a 1:1 solution of ethylenediamine in ethanol for 1 h at room temperature fully removes protecting groups on the DNA nucleobases. The resulting deprotected arrays were washed twice with nuclease-free water, dried, and stored in a desiccator until further use.

RNase HII Assays and Data Analysis. After the deprotection procedure, the hairpins folded and the slides were incubated with a buffered solution of *E. coli* recombinant RNase HII (5 units, New England Biolabs M0288S) at 37 °C [10 mM KCl, 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, and 0.1% Triton X-100 (pH 8.8)], following the manufacturer's instructions. After 1 h, the arrays were washed in water and scanned. The cleavage efficiency is calculated from the ratio of the Cy3 fluorescence intensity after or before RNase HII and, relative to the fluorescence intensity of the uncleavable, DNA-only hairpin. The fluorescence intensities are corrected for background fluorescence. The cleavage efficiency is obtained by performing the following calculations:

$$A = \frac{I_{\text{hairpin before}} - I_{\text{background before}}}{I_{\text{DNA-hairpin before}} - I_{\text{background before}}}$$

$$B = \frac{I_{\text{hairpin after}} - I_{\text{background after}}}{I_{\text{DNA-hairpin after}} - I_{\text{background after}}}$$

$$\text{cleavage efficiency} = 1 - B/A$$

The recorded cleavage efficiency is an average from five independent measurements (±standard deviation). The 20 best cleaved hairpin sequences in each series (top 2% of 1024 combinations) were used for motif searching, which was rendered as a sequence logo using Weblogo 3.6 (http://weblogo.threeplusone.com). The decrease in cleavage efficiency for sequences containing mismatches was calculated relative to the cleavage efficiency of the corresponding full-match sequence. For example, for the mismatched hairpin sequence GAAAAGCGA**A**r*U*A**A**GCGTCCTCGC**TTAGT**-CGC (mismatch base pair underlined), its cleavage efficiency was normalized to that of GAAAAGCGA**A**r*U*A**A**GCGTCC-TCGC**TTATT**CGC, the ratio yielding the decrease in cleavage efficiency. Heat maps for single and double mismatches were generated using a Pivot Table and Conditional Formatting in Microsoft Excel. Sequence logos[40] were generated by WebLogo (weblogo.berkeley.edu), and then the resulting image was manually edited to label the RNA nucleotides.
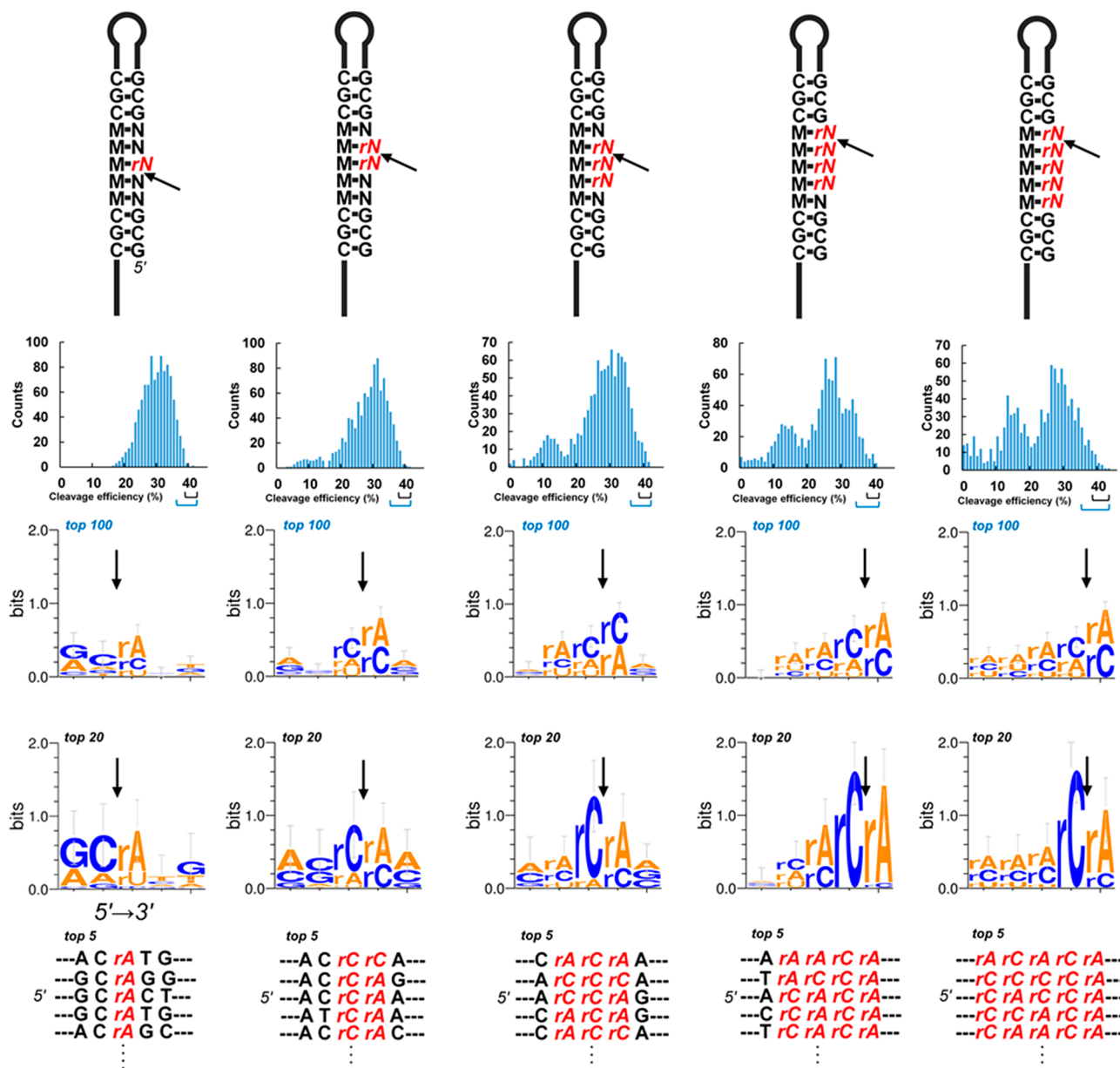
Figure 2. In each of the 1×, 2×, 3×, 4×, and 5× RNA-modified series, the number of hairpins per percent cleavage efficiency and the sequence logos from the 10% and 2% (top 100 and 20, respectively) most-cleaved hairpins (the corresponding region in the counts/percent cleavage is shown with a small bracket under the *x* axis). The large arrows point at the only cleavage site for hairpins with single RNA inserts and at the most likely cleavage site in hairpins containing consecutive RNA incorporations. The top five most-cleaved hairpins sequences for each RNA-modified series are then listed below the sequence logos.

■ RESULTS AND DISCUSSION

To comprehensively explore the activity landscape of RNase HII, we designed a library of DNA/RNA chimeric hairpins as substrates of this endoribonuclease (Figure 1). Each hairpin is composed of an 11 bp stem and a four-nucleotide loop of the TCCT sequence. The stem consists of two invariable 3 bp CGC:GCG "clamps", to stabilize the hairpin structure under the temperature and salt conditions of the RNase HII assay,[41] and a variable 5 bp middle section [nucleotides "M" and "N" (Figure 1a)]. A Cy3 label terminates the hairpin construct, along with a single-stranded GAAAA tag that serves to increase the intensity of Cy3 fluorescence, as well as to make it insensitive to sequence-specific fluorescence originating in the variable region.[42] The 5′ segment of the stem hosts the ribonucleotide inserts, and cleavage 5′ to the RNA leads to loss

of a short, Cy3-labeled segment, which can be converted into enzymatic cleavage efficiency. In terms of library elements, we set out to prepare hairpins from all possible permutations in the 5 bp variable region carrying either one, two, three, four, or five consecutive RNA bases (5 × 1024), as well as all possible single and double mismatches in two specific templates: 5′-GC*r*CCC and 5′-AA*r*UAA. The former was previously found to be a good substrate for RNase HII, while the latter displayed intermediate cleavage efficiency.[33] With mismatched sequences totaling >4000, the chimeric hairpin library contains >9000 unique elements that were synthesized in parallel, with multiple replicates, on a single glass substrate using maskless nucleic acid photolithography and 5′-photoprotected DNA and RNA phosphoramidites.[32,34,35] After deprotection and folding, the hairpins were incubated with *E. coli* RNase HII,
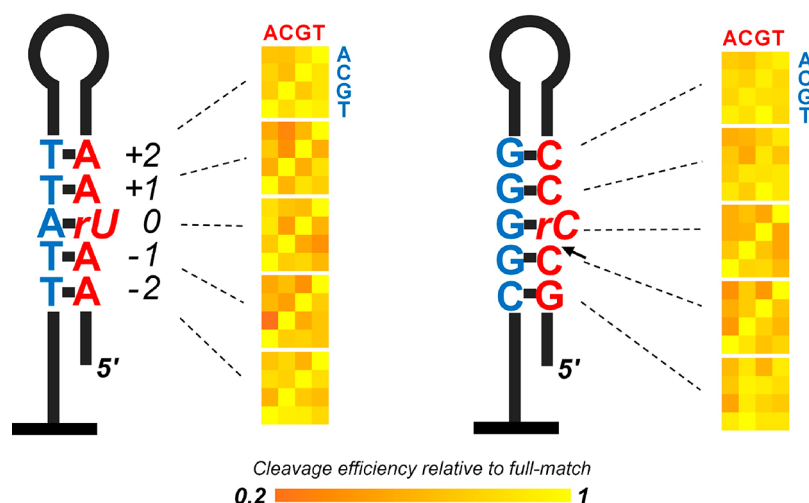
Figure 3. Effect of single mismatches on the cleavage efficiency of hairpins containing a single RNA nucleotide. The cleavage efficiency of mismatched constructs is calculated relative to that of the full-match construct and depicted in the form of heat maps. The middle heat map (for position 0) was obtained from RNA·DNA mismatches, instead of DNA·DNA mismatches at the four other positions. The arrow marks the cleavage site.

and the array was then subsequently washed and scanned. Fluorescent scanning and subsequent data extraction clearly show differences in the loss of fluorescence, ranging from 0% to ≈45% loss relative to the pure DNA hairpin (Figure 1b and Figure S1). We attribute the residual fluorescence to synthetic errors, which are likely caused by incomplete photodeprotection (95−96% per cycle). With 22 nucleotides in the hairpin stem, correct hairpin sequences thus amount to 32−40% of all oligonucleotides in each feature (0.95^{22}−0.96^{22}), which correlates well with the recorded cleavage efficiencies. Incomplete photodeprotection leads to deletion errors that affect all sequences with equal probability. The deletions result in oligonucleotides that cannot form duplexes and therefore do not participate in the pool of potential substrates.

For hairpins containing one to five RNA inserts, the subset of the top 10% most cleaved sequences is equally populated with hairpins containing one, two, or three RNA nucleotides but less well represented with sequences counting four or five consecutive RNA nucleotides (Table S1). Conversely, the subset of low cleavage rates is overrepresented with hairpins modified with four or five RNA bases, the large majority of which are rG-rich sequences (Table S2). Indeed, in the bottom 500 least cleaved hairpins, almost all possible sequences containing three, four, or five rG inserts (contiguous or not) are found: one $(rG)_5$ substrate, 19 of 21 possible instances of four rG nucleotides, and 154 of 176 sequences presenting three rG units. This effect was not observed for dG-rich hairpins, which hints at the conjoined role of guanine and the 2′-OH group in leading to a low cleavage efficiency. It may, alternatively, be due to misfolding, quartet formation, or fluorescence artifacts. We then looked at the subset of poor and better substrates in each of the 1×, 2×, 3×, 4×, and 5× RNA-modified series. Sequence motifs for the top 100 most-cleaved sequences reveal the existence of specific ribo- and deoxyribonucleotides preferentially found around the cleavage site (Figure 2), and these preferences appear to be gradually stronger when selecting the 20 and then five most-cleaved sequences. In single RNA-modified hairpins, the ribonucleotide base most commonly found in the 20 better-cleaved constructs (of 1024) appears to be rA, closely followed by rU

(Figure 2), in agreement with earlier work,[21] but in contrast to the omnipresence of rC in the shorter hairpins studied previously.[33] For DNA bases flanking the RNA modification, we noted clear differences in sequence preference between the regions upstream and downstream of the RNA. There seems to be no preference for a specific DNA base at the position 3′ to the RNA, which may be surprising because it is the location for DNA sensing by RNase HII. Upstream of the RNA insert, however, and especially at the position immediately 5′ to the RNA, there emerges a stronger consensus with dC being the preferred base for better-cleaved sequences, while the −2 position, further upstream, prefers purine nucleobases. In poorly cleaved hairpins, the picture is reversed with the region downstream of the RNA showing a clearer consensus than the upstream region. Indeed, directly 3′ to the RNA modification, we find dT as the most common nucleobase, and dC at position +2, yet 5′ to the RNA there is no indication of base preference (Figure S2). As was previously described,[21,33] the presence of rG corresponds to a low cleavage efficiency.

A very particular sequence motif takes shape as the number of consecutive RNA incorporations increases (Figure 2), specifically, with the overwhelming presence of cytidine and adenosine at the sites of RNA insertion in highly cleaved hairpins. Indeed, the 5′-dC-rA-3′ duo identified in the single RNA series carries over to the double RNA series, with 5′-rC-rA-3′ being prevalent in the 20 most-cleaved hairpins. The DNA base 5′ to the RNA−RNA section has a less distinct signature, suggesting it is a weaker factor for a high cleavage rate. Then, starting from three and up to five consecutive RNA nucleotides, the 5′-rC-rA-dX-3′ motif is almost always found around the cleavage site of the better-cleaved substrates. All additional RNA nucleotides 5′ to the rC-rA pair show reduced base selectivity, further decreasing with an increasing distance from the 5′-RNA−DNA-3′ junction. In summary, the sequence motifs presented here underline the importance of the 5′-rX$_1$/dX$_1$-rX$_2$-dX$_3$-3′ trinucleotide in RNase HII-mediated cleavage and how the nature of the rX$_1$/dX$_1$-rX$_2$ nucleobases influences its efficiency. Previous work on multiple, consecutive incorporations of rA in double-stranded DNA (dsDNA) showed that *E. coli* RNase HII predominantly
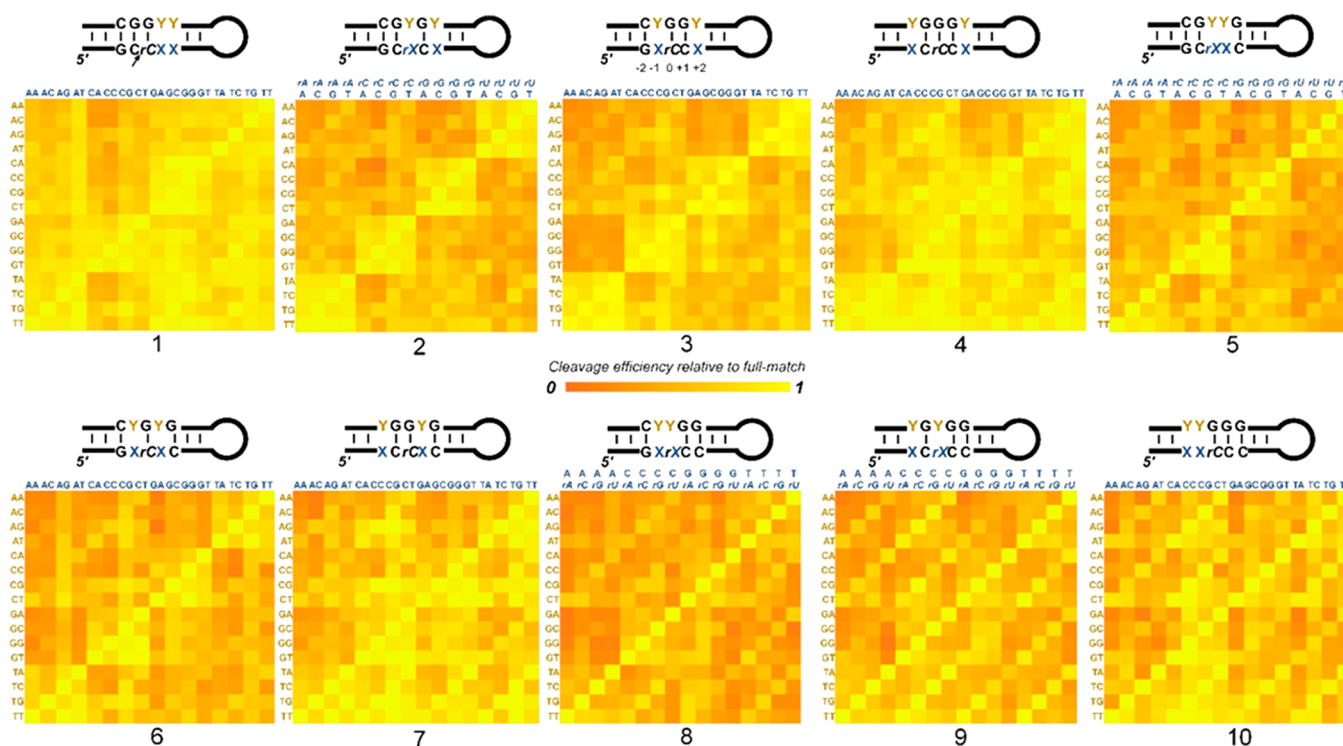
Figure 4. Cleavage efficiencies of hairpins with the 5′-GCrCCC motif where two nucleobases have been replaced with a mismatching base. The cleavage efficiency is calculated relative to that of the full match and depicted in the form of heat maps, for each of the 10 possible locations for two mismatches. In blue are the DNA or RNA bases from the 5′ segment of the stem [to be read 5′ → 3′ in the two-dimensional (2D) heat maps], and in gold the DNA bases from the 3′ segment of the stem (to be read 3′ → 5′ in the 2D heat maps).

cleaves 5′ to the 5′-rA-DNA-3′ junction and to a much lesser extent at the other rA-rA intersections,[43] and the strong sequence consensus found here at the 5′-rX$_1$-rX$_2$-dX$_3$-3′ region supports this observation. The r(CAA) motif has recently been detected in efficiently cleaved substrates of RNase H type 1.[28] Given the topological similarities between the catalytic subunits of RNases H,[27] the preference for rC/dC-rA for a high rate of cleavage may not be entirely surprising, even though it remains to be explained. Very clear cleavage motifs in substrates containing multiple, continuous RNA nucleotides may also indicate some level of sequence preference in the processing of Okazaki fragments. Of interest is also the absence of rU and rG in the better RNase HII substrates and their presence in poorly cleaved sequences, further suggesting that the sequence preference does not hinge upon either A·T/U or C·G base pair or pyrimidine/purine recognition but rather upon interaction with the actual nucleobase. In addition, the fact that nucleobase identity directly around the cleavage site (rC/dC and rA) becomes more evident as one moves toward higher cleavage efficiency hints at the possibility of a combined role of the neighboring C and A bases.

To summarize, the cleavage assay performed on a series of hairpins containing one to five consecutive RNA bases has shown that the preferentially cleaved phosphodiester bond is between dC and rA, or between rC and rA in the case of two or more consecutive RNA inserts. On the other hand, the presence of rG is met with a significantly lower cleavage efficiency. Finally, sequence specificity seems to be localized at the site of RNA incorporation as well as directly 5′ to it, yet the position 3′ to the RNA displays no particular preference for any DNA base.

We next looked at the effect of mismatches on RNase HII-mediated cleavage efficiency. To do so, we selected two templates, 5′-GCrCCC and 5′-AArUAA, and first introduced single mismatches anywhere in the template or the complementary region. The results are shown in Figure 3. We first observed that mismatches seem to have a noticeable, yet perhaps not catastrophic, effect on cleavage efficiency. At positions −2 and +2, further from the recognition and cleavage site, mismatches mildly affect cleavage efficiency (between 75% and 85% of a full match's cleavage rate), with position −2 being slightly more sensitive than position +2 in both cases. As the location of mismatch insertion draws closer to the cleavage site itself, mispairing decreases the cleavage efficiency, with the strongest effect recorded for dA·dG (template·complement) at position −1 (cleavage efficiency down 80% compared to that of the AArUAA full match). Reciprocally, dG·dA also leads to poor cleavage. In fact, at position −1, mismatches involving dA seem globally less well tolerated than other nucleobases. Mismatches at the RNA·DNA level appear to be less detrimental to cleavage efficiency than those at position −1. Still, mismatches involving rU hinder enzymatic cleavage more than rA, rC, or rG, with the classical rU·dG wobble base pair strongly affecting cleavage (decreased by 60%) when the reciprocal rG·dT pair decreased it by only 20%. The +1 position displays yet another mismatch profile, with dC·dC or dC·dA mismatches being the least cleaved hairpins (40% decrease in the case of GCrCCC and 70% in the case of AArUAA). Overall, the analysis of single mismatches (60 different sequences per series) allows us to tentatively surmise a positional effect of a given mismatch pair on the enzymatic hydrolysis by RNase HII.

We then introduced a second mismatched pair within the two templates designated above and had the overall effect on cleavage efficiency mapped in Figure 4. Within the 2560 different values reported in the 10 heat maps are full matches as well as all single mismatches. A closer look at the single mismatches in this particular context shows that the presence of a single dA·dG mismatch profoundly decreases the cleavage efficiency and seems to only be somewhat tolerated when found at position +1. In fact, dA·dG and dG·dA mismatches make up half of the 10% least-cleaved hairpins containing a single mismatch (150 sequences).

The dual mismatch constructs have a distinct pattern of cleavage efficiency. For instance, mismatches at both positions −2 and +2 (heat map 4) are, as expected, the least disrupting to enzymatic cleavage, save for patches of purine·purine clashes, only averaging a 30% decrease in total cleavage compared to a corresponding full match. In addition, two mismatches at both positions +1 and +2 (heat map 1) do not dramatically decrease the cleavage efficiency unless it involves dC at position +1, which was already noted in Figure 3. On the other hand, double mismatches at positions −1 and 0 (heat map 8) more strongly affect cleavage, averaging a 65% decrease compared to a matched sequence. In the AArUAA template, the presence of two mismatches is generally met with a much lower cleavage efficiency and a weaker dependence on the position of the mismatches within the variable region (Figure S3). Taken together, these results additionally highlight the relatively robust cleavage activity of E. coli RNase HII in the presence of mispaired ribonucleotides, which had been observed before.[44]

Finally, we monitored the enzymatic degradation of all 9506 RNA-containing hairpins over time. The well-cleaved substrates identified previously were found to be largely hydrolyzed already after 5 min with RNase HII (Figure S4). In other words, the sequence motifs presented in Figure 2, and in particular the preferred 5′-r/dC-rA-3′ dinucleotide at the cleavage site, already appear at the earliest time point of the assay. Poorly cleaved substrates on the other hand, such as 5′-AArGTC, display much slower hydrolysis rates, which signals the existence of large differences in the catalytic efficiency and turnover number of RNase HII between hairpin sequences.

CONCLUSION

In conclusion, we have successfully prepared a complex library of DNA−RNA hairpins spanning the entire sequence permutation set of a 5 bp long variable region in the stem with one to five consecutive RNA incorporations, as well as a large series of single and double mismatches. The resulting >80000 sequences, replicates included, were synthesized in parallel and in situ using nucleic acid photolithography, which can now handle DNA and RNA phosphoramidite chemistries. Multiple RNase HII assays, from a commercial source of the bacterial enzyme, performed under biologically relevant conditions on the DNA−RNA chimeric microchips have uncovered a substrate preference localized around the cleavage site. At the 5′-RNA−DNA-3′ junction that RNase HII senses to detect the presence of RNA in dsDNA, the enzyme prefers rA as the ribonucleotide but shows no preference for the 3′ DNA base. However, RNase HII prefers rC or dC 5′ to the RNA. The incorporation of mismatches in the hairpin library revealed how the position of the mismatch affects the cleavage efficiency, with position −1 (5′ to the RNA) being more sensitive to mispairing than position 0 or +1 (3′ to the RNA).

This study contributes to the understanding of the underlying mechanisms of the maintenance of genome integrity, but it also suggests that nucleobase identity around and at the site of RNA incorporation plays a role in the efficiency and rate of cleavage mediated by RNase HII. Solution-phase data will be helpful not only to validate our observations but also to identify the reasons for the apparent existence of nucleotide preference in the cleavage. Similarly, whether a stem−loop structure influences the enzymatic processing and whether a standard double-stranded format leads to the same preference for C and A bases is currently unknown. The data gathered and presented herein and, in particular, the identification of better-cleaved substrates may in addition become a useful biotechnological tool for the design of nucleic acid sequences that can be programmatically cleaved by addition of the appropriate complementary sequence and enzyme, for instance, in RNase H2-dependent polymerase chain reaction or in the development of nucleic acid-based logic circuits.[45,46] The origin of differing cleavage efficiencies remains elusive, but DNA and RNA microarrays are expected to be suitable platforms to provide clues about the binding and recognition profile of RNA-cleaving enzymes.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.bio-chem.9b00806.

> Cleavage efficiency of all experimental sequences in spreadsheet format (XLSX)

> Detailed library design, microarray scans, lists of good and poor substrates, additional heat maps, time-dependent enzymatic assays, and sequence motifs of poor enzyme substrate (PDF)

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: jory.lietard@univie.ac.at.
*E-mail: mark.somoza@univie.ac.at.

ORCID Ⓘ
Jory Lietard: 0000-0003-4523-6001
Masad J. Damha: 0000-0002-4458-1623
Mark M. Somoza: 0000-0002-8039-1341

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) McElhinny, S. A. N., Watts, B. E., Kumar, D., Watt, D. L., Lundstrom, E. B., Burgers, P. M. J., Johansson, E., Chabes, A., and Kunkel, T. A. (2010) Abundant ribonucleotide incorporation into

DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. U. S. A.* 107, 4949−4954.

(2) Sassa, A., Yasui, M., and Honma, M. (2019) Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. *Genes Environ.* 41, 3.

(3) Williams, J. S., Lujan, S. A., and Kunkel, T. A. (2016) Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* 17, 350−363.

(4) Traut, T. W. (1994) Physiological Concentrations of Purines and Pyrimidines. *Mol. Cell. Biochem.* 140, 1−22.

(5) Yao, N. Y., Schroeder, J. W., Yurieva, O., Simmons, L. A., and O'Donnell, M. E. (2013) Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12942−12947.

(6) Reijns, M. A., Rabe, B., Rigby, R. E., Mill, P., Astell, K. R., Lettice, L. A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., Devenney, P. S., Sexton, D., Grimes, G., Holt, I. J., Hill, R. E., Taylor, M. S., Lawson, K. A., Dorin, J. R., and Jackson, A. P. (2012) Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* 149, 1008−1022.

(7) Klein, H. L. (2017) Genome instabilities arising from ribonucleotides in DNA. *DNA Repair* 56, 26−32.

(8) McElhinny, S. A. N., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundstrom, E. B., Johansson, E., Chabes, A., and Kunkel, T. A. (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* 6, 774−781.

(9) Cerritelli, S. M., and Crouch, R. J. (2016) The Balancing Act of Ribonucleotides in DNA. *Trends Biochem. Sci.* 41, 434−445.

(10) Jinks-Robertson, S., and Klein, H. L. (2015) Ribonucleotides in DNA: hidden in plain sight. *Nat. Struct. Mol. Biol.* 22, 176−178.

(11) Rydberg, B., and Game, J. (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16654−16659.

(12) Vaisman, A., and Woodgate, R. (2015) Redundancy in ribonucleotide excision repair: Competition, compensation, and cooperation. *DNA Repair* 29, 74−82.

(13) Wallace, B. D., and Williams, R. S. (2014) Ribonucleotide triggered DNA damage and RNA-DNA damage responses. *RNA Biol.* 11, 1340−1346.

(14) Caldecott, K. W. (2014) Ribose-An Internal Threat to DNA. *Science* 343, 260−261.

(15) Chon, H., Sparks, J. L., Nowotny, M., Rychlik, M., Burgers, P. M., Cerritelli, S. M., and Crouch, R. J. (2013) RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* 41, 3130−3143.

(16) Oivanen, M., Kuusela, S., and Lonnberg, H. (1998) Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by Bronsted acids and bases. *Chem. Rev.* 98, 961−990.

(17) Crow, Y. J., Leitch, A., Hayward, B. E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., Baumann, C., Baxter, P., Bertini, E., Chandler, K. E., Chitayat, D., Cau, D., Dery, C., Fazzi, E., Goizet, C., King, M. D., Klepper, J., Lacombe, D., Lanzi, G., Lyall, H., Martinez-Frias, M. L., Mathieu, M., McKeown, C., Monier, A., Oade, Y., Quarrell, O. W., Rittey, C. D., Rogers, R. C., Sanchis, A., Stephenson, J. B. P., Tacke, U., Till, M., Tolmie, J. L., Tomlin, P., Voit, T., Weschke, B., Woods, C. G., Lebon, P., Bonthron, D. T., Ponting, C. P., and Jackson, A. P. (2006) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat. Genet.* 38, 910−916.

(18) Minias, A. E., Brzostek, A. M., Minias, P., and Dziadek, J. (2015) The Deletion of rnhB in Mycobacterium smegmatis Does Not Affect the Level of RNase HII Substrates or Influence Genome Stability. *PLoS One* 10, No. e0115521.

(19) Kouzminova, E. A., Kadyrov, F. F., and Kuzminov, A. (2017) RNase HII Saves rnhA Mutant Escherichia coli from R-Loop-Associated Chromosomal Fragmentation. *J. Mol. Biol.* 429, 2873−2894.

(20) Cerritelli, S. M., and Crouch, R. J. (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 276, 1494−1505.

(21) Eder, P. S., Walder, R. Y., and Walder, J. A. (1993) Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* 75, 123−126.

(22) Murante, R. S., Henricksen, L. A., and Bambara, R. A. (1998) Junction ribonuclease: An activity in Okazaki fragment processing. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2244−2249.

(23) Ohtani, N., Tomita, M., and Itaya, M. (2008) Junction ribonuclease: a ribonuclease HII orthologue from Thermus thermophilus HB8 prefers the RNA-DNA junction to the RNA/DNA heteroduplex. *Biochem. J.* 412, 517−526.

(24) Randall, J. R., Hirst, W. G., and Simmons, L. A. (2018) Substrate Specificity for Bacterial RNases HII and HIII Is Influenced by Metal Availability. *J. Bacteriol.* 200, e00401−00417.

(25) Tadokoro, T., and Kanaya, S. (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.* 276, 1482−1493.

(26) Shaban, N. M., Harvey, S., Perrino, F. W., and Hollis, T. (2010) The Structure of the Mammalian RNase H2 Complex Provides Insight into RNA·DNA Hybrid Processing to Prevent Immune Dysfunction. *J. Biol. Chem.* 285, 3617−3624.

(27) Rychlik, M. P., Chon, H., Cerritelli, S. M., Klimek, P., Crouch, R. J., and Nowotny, M. (2010) Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. *Mol. Cell* 40, 658−670.

(28) Kielpinski, L. J., Hagedorn, P. H., Lindow, M., and Vinther, J. (2017) RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic Acids Res.* 45, 12932−12944.

(29) Schultz, S. J., Zhang, M. H., and Champoux, J. J. (2004) Recognition of internal cleavage sites by retroviral RNases H. *J. Mol. Biol.* 344, 635−652.

(30) Schultz, S. J., Zhang, M. H., and Champoux, J. J. (2010) Multiple Nucleotide Preferences Determine Cleavage-Site Recognition by the HIV-1 and M-MuLV RNases H. *J. Mol. Biol.* 397, 161−178.

(31) Malfatti, M. C., Balachander, S., Antoniali, G., Koh, K. D., Saint-Pierre, C., Gasparutto, D., Chon, H., Crouch, R. J., Storici, F., and Tell, G. (2017) Abasic and oxidized ribonucleotides embedded in DNA are processed by human APE1 and not by RNase H2. *Nucleic Acids Res.* 45, 11193−11212.

(32) Lackey, J. G., Mitra, D., Somoza, M. M., Cerrina, F., and Damha, M. J. (2009) Acetal Levulinyl Ester (ALE) Groups for 2′-Hydroxyl Protection of Ribonucleosides in the Synthesis of Oligoribonucleotides on Glass and Microarrays. *J. Am. Chem. Soc.* 131, 8496−8502.

(33) Lietard, J., Ameur, D., Damha, M., and Somoza, M. M. (2018) High-density RNA microarrays synthesized in situ by photolithography. *Angew. Chem., Int. Ed.* 57, 15257−15261.

(34) Hölz, K., Schaudy, E., Lietard, J., and Somoza, M. M. (2019) Multi-level patterning nucleic acid photolithography. *Nat. Commun.* 10, 3805.

(35) Agbavwe, C., Kim, C., Hong, D., Heinrich, K., Wang, T., and Somoza, M. M. (2011) Efficiency, Error and Yield in Light-Directed Maskless Synthesis of DNA Microarrays. *J. Nanobiotechnol.* 9, 57.

(36) Sack, M., Kretschy, N., Rohm, B., Somoza, V., and Somoza, M. M. (2013) Simultaneous Light-Directed Synthesis of Mirror-Image Microarrays in a Photochemical Reaction Cell with Flare Suppression. *Anal. Chem.* 85, 8513−8517.

(37) Kretschy, N., Holik, A. K., Somoza, V., Stengele, K. P., and Somoza, M. M. (2015) Next-Generation o-Nitrobenzyl Photolabile Groups for Light-Directed Chemistry and Microarray Synthesis. *Angew. Chem., Int. Ed.* 54, 8555−8559.

(38) Sack, M., Holz, K., Holik, A. K., Kretschy, N., Somoza, V., Stengele, K. P., and Somoza, M. M. (2016) Express photolithographic DNA microarray synthesis with optimized chemistry and high-efficiency photolabile groups. *J. Nanobiotechnol.* 14, 14.

(39) Holz, K., Lietard, J., and Somoza, M. M. (2017) High-Power 365 nm UV LED Mercury Arc Lamp Replacement for Photochemistry and Chemical Photolithography. *ACS Sustainable Chem. Eng.* 5, 828−834.

(40) Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res. 18*, 6097−6100.

(41) Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N., and Ansari, A. Z. (2006) Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. U. S. A. 103*, 867−872.

(42) Kretschy, N., Sack, M., and Somoza, M. M. (2016) Sequence-Dependent Fluorescence of Cy3-and Cy5-Labeled Double-Stranded DNA. *Bioconjugate Chem. 27*, 840−848.

(43) Haruki, M., Tsunaka, Y., Morikawa, M., and Kanaya, S. (2002) Cleavage of a DNA-RNA-DNA/DNA chimeric substrate containing a single ribonucleotide at the DNA-RNA junction with prokaryotic RNases HII. *FEBS Lett. 531*, 204−208.

(44) Shen, Y., Koh, K. D., Weiss, B., and Storici, F. (2012) Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H. *Nat. Struct. Mol. Biol. 19*, 98−105.

(45) Dobosy, J. R., Rose, S. D., Beltz, K. R., Rupp, S. M., Powers, K. M., Behlke, M. A., and Walder, J. A. (2011) RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnol. 11*, 80.

(46) Wang, F., Lu, C. H., and Willner, I. (2014) From Cascaded Catalytic Nucleic Acids to Enzyme-DNA Nanostructures: Controlling Reactivity, Sensing, Logic Operations, and Assembly of Complex Structures. *Chem. Rev. 114*, 2881−2941.