



Cite this: DOI: 10.1039/d5nh00811e

Received 14th December 2025,  
Accepted 12th January 2026

DOI: 10.1039/d5nh00811e

rsc.li/nanoscale-horizons

## RNA nanostructures based on three-letter coding with non-canonical base pairs

Jianqiu Zhao,<sup>ab</sup> Yan Qin,<sup>ab</sup> Qiancheng Xiong,<sup>id</sup><sup>c</sup> Fang Fang<sup>id</sup><sup>\*d</sup> and Bryan Wei<sup>id</sup><sup>\*ab</sup>

Synthetic RNA nanostructures are typically composed of four nucleotides (A, U, G, and C) following a canonical base pairing rule (A-U/G-C). G-U wobble pairs are commonly employed in many RNA nanostructures, but other non-canonical base pairing remains underexplored. In this work, we design RNA nanostructures with only three nucleotides instead of four. Besides Watson–Crick G-C base pairs, we incorporate A-C non-canonical base pairs into this three-letter coding scheme and allow selective nanostructure assembly from mixed DNA templates. With the new paradigm, we produce a variety of RNA nanostructures, further expanding the possibilities of rational molecular design.

## Introduction

Over the past two decades, RNA nanotechnology<sup>1</sup> has been developed in parallel with DNA nanotechnology,<sup>2</sup> with a shared rule of canonical Watson–Crick base pairing (A-T/U and G-C). From the pioneering designs of self-assembled tectoRNA modules<sup>3,4</sup> to the more recent co-transcriptional folding methodologies, RNA nanostructures whose complexity rivals that of their counterpart DNA nanostructures have been produced.<sup>5,6</sup> In the conventional design of synthetic RNA nanostructures, duplex segments of A-form RNA conformation are typically composed of continuous canonical Watson–Crick base pairs. Conversely, a substantial proportion of RNA base interactions in natural RNA systems involve highly versatile non-canonical base pairs, such as G-U, G-A, A-A, G-G, C-C, U-U, U-C, and A-C.<sup>7,8</sup> Similarly, analogous non-canonical base pairing interactions

## New concepts

This is the first time three-letter coding has been applied in synthetic nucleic acid constructs. The total omission of a certain nucleotide species for an entire RNA construct enables crucial differentiation from its counterparts with common four-letter coding. Such differentiation can be applied in RNA information encryption and other synthetic biology applications.

are also applicable in synthetic DNA constructs.<sup>9,10</sup> Among these non-canonical interactions, wobble A-C base pairing (*cis* Watson–Crick/Watson–Crick) is one of the most common pairings in the RNA double helix,<sup>7</sup> which exhibits a single hydrogen bond, while its protonated variant (wobble A<sup>+</sup>·C) stabilizes two hydrogen bonds with the extra one formed by the N1 nitrogen atom of A (Fig. 1).<sup>11,12</sup> Naturally, A-C base pairs play an important role in the stabilization of the anticodon stem of tRNA<sup>13</sup> and are also a basic structural element of many ribozymes.<sup>14–16</sup>

In the sequence design of many RNA nanostructures, G-U wobble pairs are employed to mitigate secondary structures of the corresponding DNA templates,<sup>5</sup> but exploration of rational design based on other non-canonical base pairing remains limited. Herein, we present a three-letter coding scheme (A, C, and G) for RNA nanostructures with the integration of non-canonical A-C base pairs, which are structurally similar to G-U wobble pairs.<sup>11</sup> Using this scheme, we establish successful folding across diverse RNA nanostructures. Wobble pairs may reduce

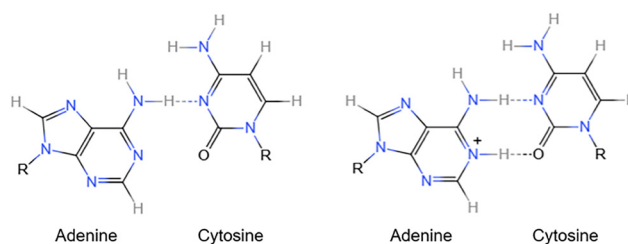


Fig. 1 Chemical structures of wobble A-C and A<sup>+</sup>·C base pairs at physiological pH.

<sup>a</sup> School of Life Sciences, Tsinghua University, Beijing, 100084, China.  
E-mail: bw@tsinghua.edu.cn

<sup>b</sup> Center for Synthetic and Systems Biology, Tsinghua University, Beijing, 100084, China

<sup>c</sup> Bioprocessing Technology Institute, Agency for Science, Technology and Research (A\*STAR), Singapore 138668, Singapore

<sup>d</sup> Reproductive and Genetic Center, The First Affiliated Hospital of USTC, Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China. E-mail: ffang24@ustc.edu.cn

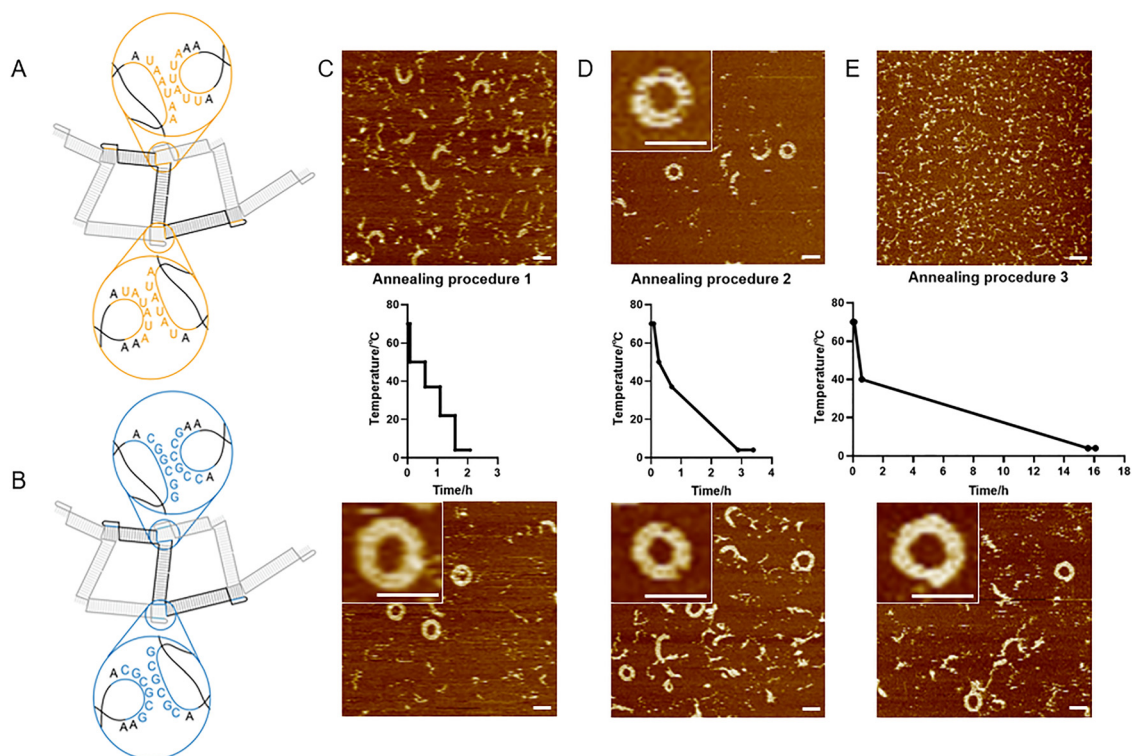


duplex stability and thereby limit folding and self-assembly robustness. At the same time, the introduction of A-C base pairs enables the complete omission of a specific species of nucleotide (e.g. U) throughout an entire construct. Such an intentional nucleotide species omission establishes a critical distinction from conventional four-letter coded constructs. As a demonstration, we selectively assemble a three-letter-coded target construct from mixed templates containing only three out of four canonical NTPs. Our implementation offers opportunities for nucleic acid information encryption<sup>17–19</sup> and the creation of recoded genomes/transcripts incorporating unnatural bases.<sup>20–22</sup>

## Results

To investigate the appropriate nucleotide composition for kissing loops (KLs), we derived a Z-shaped motif with asymmetric arm lengths from an earlier report to form RNA constructs of closed-ring configuration.<sup>23</sup> Specifically, we adopted typical branched KLs between a bulged helix and a hairpin loop *via* Watson-Crick base pairing over a 6-nucleotide region. The asymmetric arm lengths were set to 14 and 25 base pairs. The folding of our RNA nanostructures relies on interactions of two hierarchies. A stem-loop intermediate initially forms, with stems

paired through complementary intramolecular segments. Then, the pre-designed KLs connect *via* intermolecular or ring-closing interactions, directing the self-assembly into the desired geometry. KL interactions must satisfy both binding strength and orthogonality requirements. With that in mind, we systematically assessed KLs composed exclusively of A-U (Fig. 2A and Fig. S1) or G-C (Fig. 2B and Fig. S3) base pairs with three different annealing procedures (middle panels of Fig. 2C–E): (1) step cooling, (2) short ramp cooling, and (3) long ramp cooling. Our atomic force microscopy (AFM) results showed that KLs exclusively with G-C base pairs formed complete structures following any of the three annealing procedures (bottom panels of Fig. 2C–E and Fig. S3), while KLs exclusively with A-U base pairs gave rise to closed rings in procedure 2 (top panel of Fig. 2D and Fig. S1) with a limited yield and extremely rare ring appearance in procedure 1 (Fig. S2). From the results, KLs exclusively with G-C base pairs demonstrated more reliable self-assembly capability than their counterparts exclusively with A-U base pairs. The limited stability of A-U base pairs suggested that three-letter coding of A, U and G with even weaker G-U wobble pairs was not preferable. As unpaired As are structurally required in KL sequences to maintain the coaxial alignment,<sup>24</sup> we adopted three-letter coding of A, C and G for our RNA nanostructures.



**Fig. 2** Comparison of the assembly of RNA rings whose kissing loops (KLs) involve A-U base pairs and G-C base pairs. (A) Strand diagram of the Z-shaped motif with asymmetric arm lengths with KLs connected exclusively with A-U base pairs. KL interactions are depicted in orange. (B) Strand diagram of the Z-shaped motif with asymmetric arm lengths, with KLs connected exclusively with G-C base pairs. KL interactions are depicted in blue. (C) Atomic force microscopy (AFM) results following annealing procedure 1, step cooling from 70 °C to 4 °C. (D) AFM results following annealing procedure 2, short ramp cooling from 70 °C to 4 °C (fast cooling from 70 °C to 37 °C and slower cooling from 37 °C to 4 °C). (E) AFM results following annealing procedure 3, long ramp cooling from 70 °C to 4 °C (fast cooling from 70 °C to 40 °C and overnight cooling from 40 °C to 4 °C). For (C)–(E), top: AFM images with KLs with A-U base pairs; middle: annealing procedure; bottom: AFM images with KLs with G-C base pairs. Insets show magnified views. Scale bars: 50 nm.



Next, we redesigned the sequence of the Z-shaped motif exclusively with G-C pairs, including stems and KLs. However, such a sequence arrangement only led to short fragments according to AFM imaging results (Fig. 3A and Fig. S4). Then, we incorporated varying numbers of A-C base pairs in the sequence. When A-C base pairs were introduced at a percentage of  $\sim 5\%$  or  $\sim 10\%$ , polymers with closed-ring configuration successfully self-assembled (Fig. 3B, C and Fig. S5, S6). A design incorporating  $\sim 5\%$  A-C base pairs exhibited improved self-assembly performance, with a more uniform diameter distribution (Fig. S7). This trend was consistent with the Mfold predictions of reduced undesired secondary structures.<sup>25</sup> Therefore, a composition of  $\sim 5\%$  A-C base pairs was adopted in the rest of this study. Besides A-C base pairs being positioned at stem regions, we also tested constructs with A-C base pairs positioned at loop regions (Fig. S8). Similarly, the desired ring structures were produced with a comparable yield. Under physiological pH, A-C and A<sup>+</sup>-C wobble pairs coexist (Fig. 1).<sup>26,27</sup> The structural details of A-C base pairs have been previously characterized by NMR spectroscopy<sup>12</sup> and X-ray crystallography.<sup>26</sup>

We then extended this three-letter coding strategy to additional RNA nanostructures. Our next RNA construct was a ribbon assembled from a Z-shaped motif with symmetric arm lengths (14 base pairs).<sup>23</sup> Similar to the original design with conventional four-letter coding, three-letter coded RNA ribbons and rings of different sizes were observed under AFM (Fig. 4A and Fig. S9). Unlike the four-letter coded counterparts, the three-letter ribbons did not extend to micrometer lengths. Similarly, we designed an RNA square based on  $90^\circ$  kink motifs<sup>28,29</sup> and an RNA triangle based on open three-way junction motifs.<sup>29,30</sup> The 6-nt loop sequences of branched KLs<sup>23</sup> were modified from GGAGGC to GGCGGC and from GCGAGC to GCGGCG, and those of  $180^\circ$  KLs<sup>29</sup> from GGAGGC to GGCGGC. Other structural motifs, including  $90^\circ$  kinks<sup>28,29</sup> (AACUA to AACCA) and tetraloops<sup>31</sup> (GAAA), were adapted as U-free variants from published designs. AFM

imaging confirmed the proper folding into expected geometries (Fig. 4B, C and Fig. S10, S11).

With total omission of Us for the entire RNA constructs, we then sought to selectively produce three-letter coded constructs from mixed three- and four-letter DNA templates in a co-transcription system. For this system, we designed structure Z (Z-shaped motif with asymmetric arm lengths with  $\sim 5\%$  A-C base pairs) with three-letter coding and structure S with conventional four-letter coding. Using mixed templates of Z and S, we compared transcription outputs from different pools of input nucleoside triphosphate. In transcription reactions containing only VTPs (ATP, CTP and GTP), only the three-letter-coded structure Z was transcribed and yielded properly self-assembled products (Fig. 5A and Fig. S12 A, C). In contrast, reactions with all four NTPs (ATP, CTP, GTP and UTP) generated both structure Z and structure S (Fig. 5B; Fig. S12B and C). Thus, inclusion or exclusion of UTP functions as a critical switch enabling or disabling the production of specific constructs in our co-transcription system.

## Discussion

In this study, we designed and constructed RNA nanostructures composed of three nucleotides. Our results indicate that G-C base pairs in KLs led to better self-assembly performance. Therefore, we adopted A, C and G in our three-letter coding strategy. Integration of A-C non-canonical base pairs was carefully optimized to ensure proper structural formation, resulting in various RNA nanostructures with three-letter coding. Taking advantage of the omission of U nucleotides for the entire constructs, we manage to selectively produce target constructs from mixed templates with three- and four-letter coding. Our investigation of the inclusion of additional non-canonical base pairing has further expanded the design freedom of RNA nanostructures.

Compared to RNA nanostructures based on conventional four-letter coding, our three-letter designs showed slightly

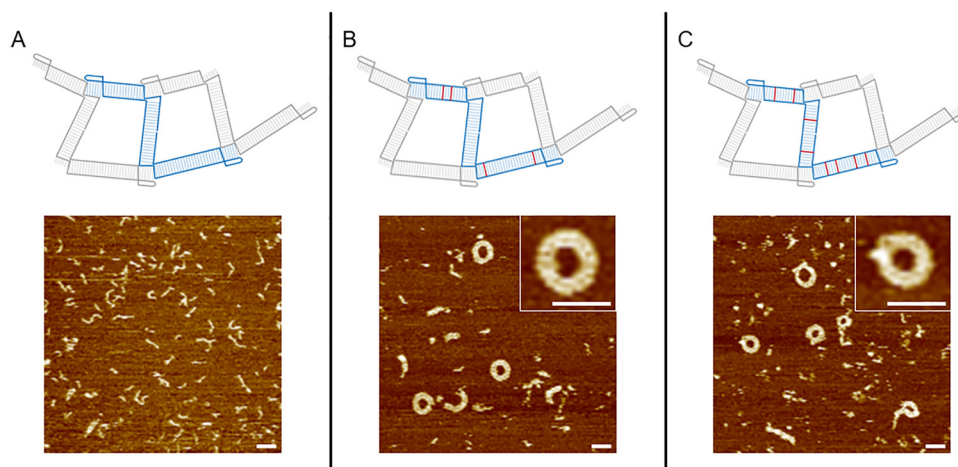


Fig. 3 Adjustment of A-C base pair composition. (A) Z-shaped motif with asymmetric arm lengths with no A-C base pair inserted. (B) Z-shaped motif with asymmetric arm lengths with  $\sim 5\%$  A-C base pairs. (C) Z-shaped motif with asymmetric arm lengths with  $\sim 10\%$  A-C base pairs. Top: Strand diagrams (A-C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images (insets show magnified views). Scale bars: 50 nm.



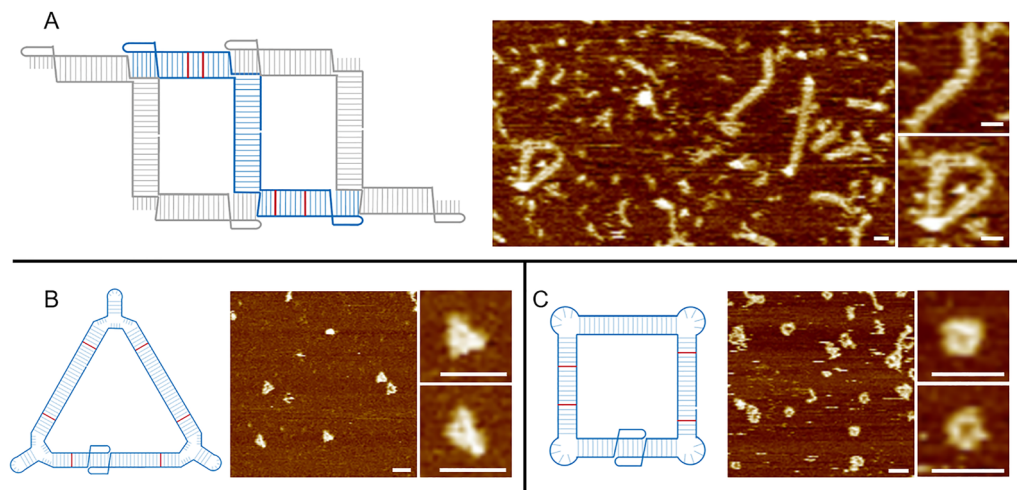


Fig. 4 Diverse three-letter-coded RNA nanostructures. (A) Ribbon. (B) Triangle. (C) Square. Left: Strand diagrams (four A-C base pairs depicted in red); right: AFM images (insets show magnified views). Scale bars: 25 nm.

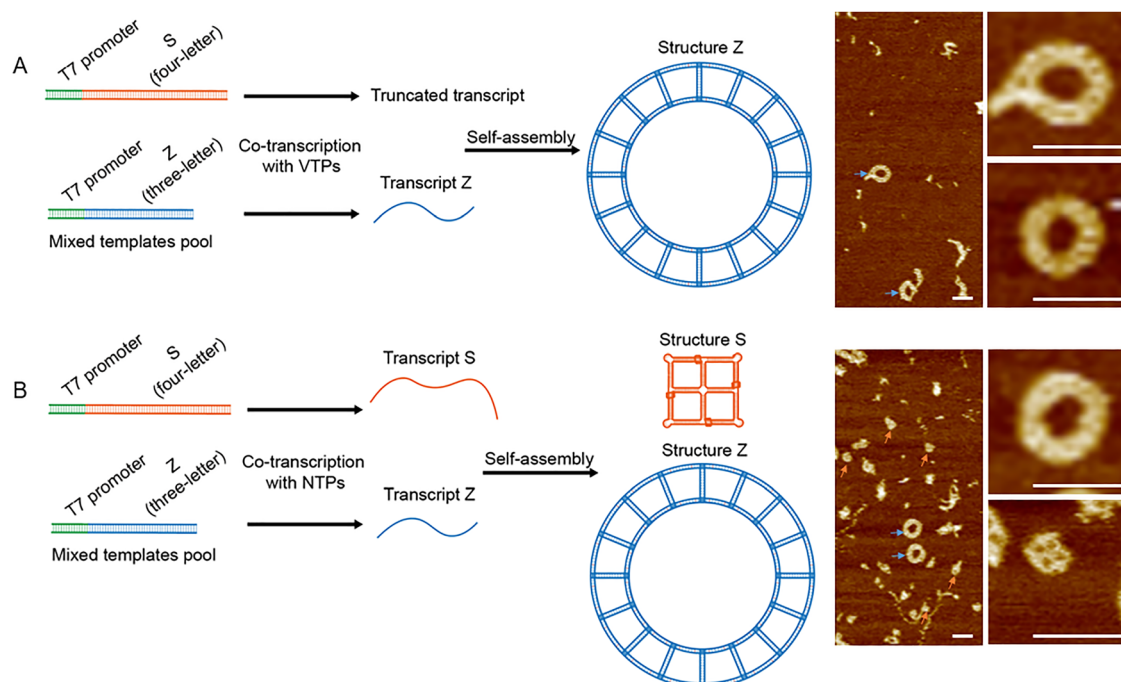


Fig. 5 Co-transcription system with mixed templates of Z (three-letter coding) and S (four-letter coding). (A) Co-transcription with VTPs (ATP, CTP and GTP). Left: Diagrams of co-transcription to generate structure Z with VTPs; right: AFM images. (B) Co-transcription with NTPs (ATP, CTP, GTP and UTP). Left: Diagrams of co-transcription to generate structures S and Z with NTPs; right: AFM images (blue and orange arrows point to structures Z and S, respectively). Scale bars: 50 nm.

reduced self-assembly performance (Fig. S13). Moreover, it is necessary to optimize synthesis and amplification strategies for GC-rich DNA templates. Further sequence optimization could improve the production efficiency of RNA nanostructures with three-letter coding. Beyond A-C base pairing and three-letter coding of A, C and G, more non-canonical base pairing based on natural bases (unmodified or modified) and synthetic bases can further expand the RNA design toolbox. The intentional omission of a certain nucleotide species from entire RNA

constructs not only differentiates three-letter designs from conventional four-letter constructs but also establishes an additional design dimension. The excluded nucleotide species can be reserved for dedicated functionality, enabling orthogonal control over RNA production and downstream self-assembly through nucleotide availability rather than sequence redesign. As shown in our selective target structure production from mixed template pools, this design principle provides a foundation for implementations of RNA information encryption. Similar to synthetic



biology exploration, in which certain codons are entirely left out of the translation system and reassigned as novel synthetic codons alongside corresponding novel amino acids,<sup>20–22</sup> the omission and reservation of a certain nucleotide in RNA constructs offers programmability to introduce new functionalities.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information is about methods, experimental details, and sequences. See DOI: <https://doi.org/10.1039/d5nh00811e>.

## Acknowledgements

This work is supported by: National Key Research and Development Program of China (grant No. 2021YFF1200200), and grants from Tsinghua University Initiative Scientific Research Program (to B. W.).

## References

- 1 E. Poppleton, N. Urbanek, T. Chakraborty, A. Griffo, L. Monari and K. Göpflich, *RNA Biol.*, 2023, **20**, 510–524.
- 2 N. C. Seeman and H. F. Sleiman, *Nat. Rev. Mater.*, 2018, **3**, 17068.
- 3 L. Jaeger, E. Westhof and N. B. Leontis, *Nucleic Acids Res.*, 2001, **29**, 455–463.
- 4 A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam, E. Oroudjev, H. G. Hansma and L. Jaeger, *Science*, 2004, **306**, 2068–2072.
- 5 C. Geary, P. W. K. Rothmund and E. S. Andersen, *Science*, 2014, **345**, 799–804.
- 6 C. Geary, G. Grossi, E. K. McRae, P. W. K. Rothmund and E. S. Andersen, *Nat. Chem.*, 2021, **13**, 549–558.
- 7 A. Roy, S. Panigrahi, M. Bhattacharyya and D. Bhattacharyya, *J. Phys. Chem. B*, 2008, **112**, 3786–3796.
- 8 S. Lemieux and F. Major, *Nucleic Acids Res.*, 2002, **30**, 4250–4263.
- 9 N. Wu, L. Niu, J. Liu, Y. Luo, P. Hao, X. Sun, B. Wang, B. Liu, F. Chen, C. Fan and Y. Zhao, *Angew. Chem., Int. Ed.*, 2025, **18**, e202513760.
- 10 E. Chen, M. Trajkovski, H. K. Lee, S. Nyovanie, K. N. Martin, W. L. Dean, M. Tahiliani, J. Plavec and L. A. Yatsunyk, *Nucleic Acids Res.*, 2024, **52**, 3390–3405.
- 11 L. Yang, Z. Zhong, C. Tong, H. Jia, Y. Liu and G. Chen, *J. Am. Chem. Soc.*, 2018, **140**, 8172–8184.
- 12 J. D. Puglisi, J. R. Wyatt and I. Tinoco Jr., *Biochemistry*, 1990, **29**, 4215–4226.
- 13 P. C. Durant and D. R. Davis, *J. Mol. Biol.*, 1999, **285**, 115–131.
- 14 A. L. Cerrone-Szakal, D. M. Chadalavada, B. L. Golden and P. C. Bevilacqua, *RNA*, 2008, **14**, 1746–1760.
- 15 N. B. Suslov, S. DasGupta, H. Huang, J. R. Fuller, D. M. J. Lilley, P. A. Rice and J. A. Piccirilli, *Nat. Chem. Biol.*, 2015, **11**, 840–846.
- 16 P. Dagenais, N. Girard, E. Bonneau and P. Legault, *Rev. RNA*, 2017, **8**, e1421.
- 17 Y. Zhang, F. Wang, J. Chao, M. Xie, H. Liu, M. Pan, E. Kopperger, X. Liu, Q. Li, J. Shi, L. Wang, J. Hu, L. Wang, F. C. Simmel and C. Fan, *Nat. Commun.*, 2019, **10**, 5469.
- 18 C. T. Clelland, V. Risca and C. Bancroft, *Nature*, 1999, **399**, 533–534.
- 19 S. Zhou, Y. Wei, Y. Zhang, H. H. Iu and H. Zhang, *Integration*, 2025, **101**, 102336.
- 20 M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs, *Science*, 2013, **342**, 357–360.
- 21 J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. Dunkelmann, V. Beranek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott and J. W. Chin, *Nature*, 2019, **569**, 514–518.
- 22 M. W. Grome, M. T. A. Nguyen, D. W. Moonan, K. Mohler, K. Gurara, S. Wang, C. Hemez, B. J. Stenton, Y. Cao, F. Radford, M. Kornaj, J. Patel, M. Prome, S. Rogulina, D. Sozanski, J. Tordoff, J. Rinehart and F. J. Isaacs, *Nature*, 2025, **639**, 512–521.
- 23 D. Liu, C. W. Geary, G. Chen, Y. Shao, M. Li, C. Mao, E. S. Andersen, J. A. Piccirilli, P. W. K. Rothmund and Y. Weizmann, *Nat. Chem.*, 2020, **12**, 249–259.
- 24 E. Ennifar, P. Walter, B. Ehresmann, C. Ehresmann and P. Dumas, *Nat. Struct. Biol.*, 2001, **8**, 1064–1068.
- 25 M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.
- 26 B. Pan, S. N. Mitra and M. Sundaralingam, *J. Mol. Biol.*, 1998, **283**, 977–984.
- 27 H. T. Allawi and J. SantaLucia Jr., *Biochemistry*, 1998, **37**, 9435–9444.
- 28 S. M. Dibrov, H. Johnston-Cox, Y. Weng and T. Hermann, *Angew. Chem., Int. Ed.*, 2007, **46**, 226–229.
- 29 M. Li, M. Zheng, S. Wu, C. Tian, D. Liu, Y. Weizmann, W. Jiang, G. Wang and C. Mao, *Nat. Commun.*, 2018, **9**, 2196.
- 30 A. Lescoute and E. Westhof, *RNA*, 2006, **12**, 83–93.
- 31 H. A. Heus and A. Pardi, *Science*, 1991, **253**, 191–194.

