www.nature.com/cr
www.cell-research.com

Cell Research

LETTER TO THE EDITOR

Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1

Cell Research (2020) 30:272–275; https://doi.org/10.1038/s41422-019-0233-9

Dear Editor,

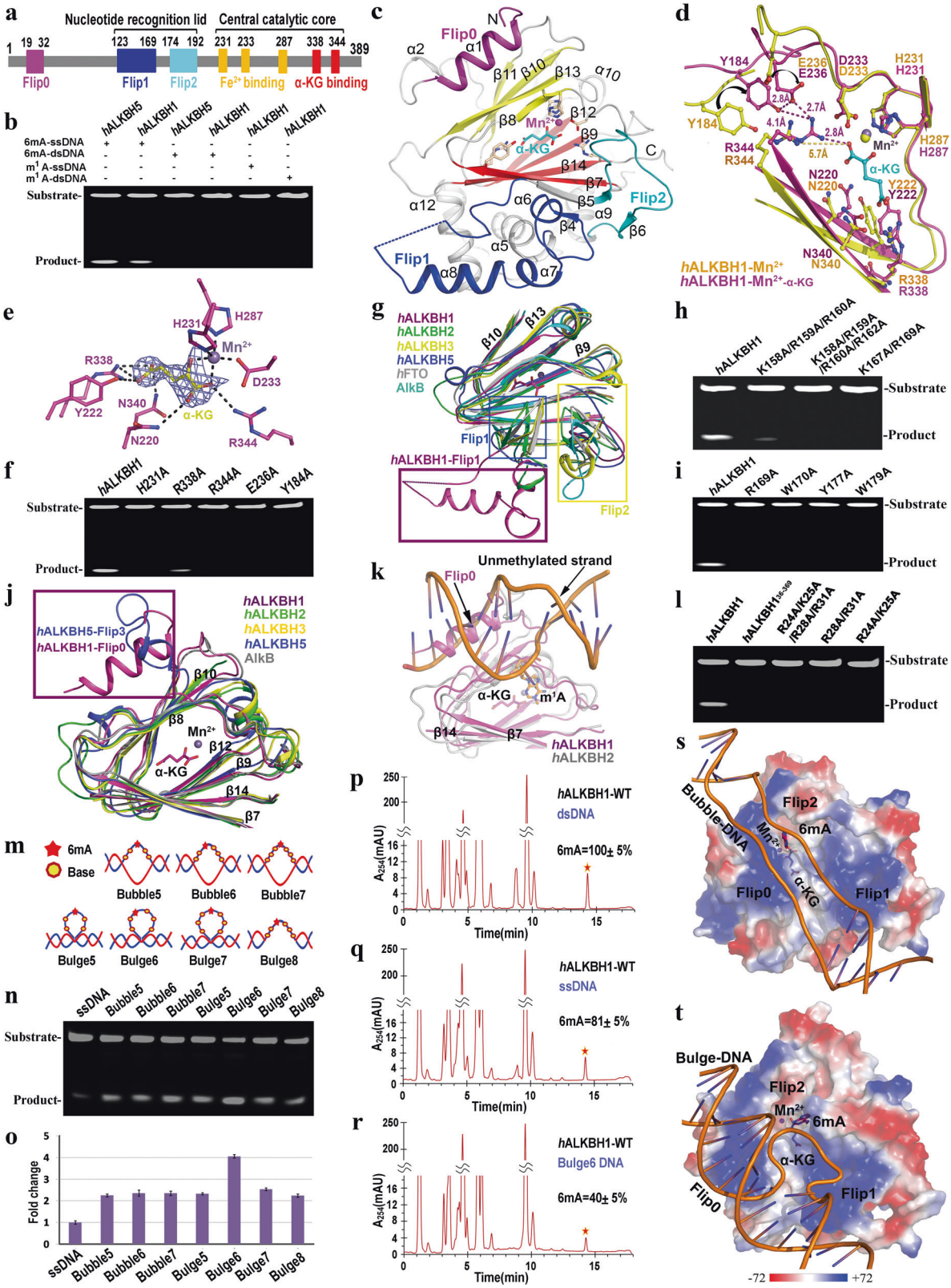
DNA N^6-methyladenine (6mA) modification is common in prokaryotes[1] and eukaryotes,[2] involving in gene regulation, transposon, stem cell differentiation, and human tumors. At present, it has been confirmed that 6mA is ubiquitous in the human genome, and [G/C]AGG[C/T] is the most prominent motif for 6mA modification.[3] Human ALKBH1 (*h*ALKBH1), one of the nine human homologs of the AlkB family, is an Fe(II) and α-ketoglutarate (α-KG)-dependent dioxygenase and highly conserved in mammals. AlkB family proteins can repair damaged DNA/RNA or other lesions. *h*ALKBH1 exhibits demethylation activity toward 6mA, and its abnormal expression has been found in many human cancers and developmental defects, such as tissue malformation and gender imbalance.[3–5] DNA methyltransferase N6AMT1 and demethylase *h*ALKBH1 mediate the methylation and demethylation of DNA 6mA in the human genome, respectively. The abnormal distribution of 6mA has been found in many cancers.[3,4] Interestingly, *h*ALKBH1 is also reported to have demethylation activity toward other four kinds of substrates, such as histone H2A,[6] m^3C on DNA and RNA,[7] m^5C or m^1A on tRNA.[8–10] Therefore, the molecular function of *h*ALKBH1 is still controversial and the functional mechanism is unclear. In addition, the sequence identity between *h*ALKBH1 and other solved structures is less than 19%, which hinders the understanding of action mechanism and the potential drug applications.

Here we determined two crystal structures of binary and ternary complex of *h*ALKBH1$_{19–369}$, *h*ALKBH1$_{19–369}$-Mn^{2+} and *h*ALKBH1$_{19–369}$-Mn^{2+}-α-KG at resolutions of 1.97 and 2.8 Å, respectively (Supplementary information, Fig. S1 and Table S1). *h*ALKBH1$_{19–369}$ (referred to as *h*ALKBH1 hereafter) contained the enzymatically active center and retained full demethylation activity toward 6mA ssDNA, but not toward 6mA on dsDNA or m^1A on ssDNA, the same as *h*ALKBH1 WT (Fig. 1a, b). The overall structures of *h*ALKBH1-Mn^{2+}-α-KG and *h*ALKBH1-Mn^{2+} are similar, with root mean square deviation of 1.4 Å. The *h*ALKBH1 contains a unique N-terminal Flip0, the nucleotide recognition lid (NRL, containing Flip1 and Flip2), and the central catalytic core. A highly conservative double-stranded β-helix (DSBH) fold in the catalytic core is a characteristic of the α-KG-dependent dioxygenase superfamily (Fig. 1c; Supplementary information, Fig. S2). Four antiparallel β-strands β7, β9, β12, and β14 form the major sheet, whereas β8, β10, β11 and β13 form the minor sheet. *h*ALKBH1 has a conservative HxD…H metal ion-binding sequence and an R…R α-KG-binding sequence. Apart from Mn^{2+}, α-KG is further stabilized by three hydrogen bonds formed by the side chains of Asn220, Asn340 and Tyr222 as well as three salt bridges formed by Arg338 and Arg344 (Fig. 1d, e). Isothermal titration calorimetry analysis of the *h*ALKBH1-α-KG interaction revealed a dissociation constant (K_d) of 4.5 μM and a 1:1 stoichiometry (Supplementary information, Fig. S3b). The total surface area of *h*ALKBH1 decreased obviously from 14,445 to 13,893 $Å^2$ during the addition of α-KG, indicating that *h*ALKBH1 whole structure shrank with the association of α-KG, which caused many conformational changes including those of active sites, β5, β8, Flip1 and Flip2 (Supplementary information, Movie S1, Fig.S3c, d). Moreover, of the eight catalytic residues, Arg344 and Tyr222 showed significant conformational changes, interacting with α-KG through moving 2.5 Å and 2.3 Å toward the active center, respectively. Besides, Tyr184 and Glu236 of Flip2 exhibited the most obvious conformational changes and moved to Arg344, forming a stabilization triangle by hydrogen bonds (Fig. 1d). The stability triangle of Tyr184-Arg344-Glu236 near the α-KG-binding site is essential for 6mA ssDNA demethylation activity. Any single mutation of the triangle abolished the demethylation activity (Fig. 1e, f), but did not reduce the DNA-binding affinity (Supplementary information, S3e). Interestingly, these large conformational changes induced by α-KG were not observed in other AlkB members,[11–14] revealing a novel function of the essential triangle as a scaffold for the catalytic activity of *h*ALKBH1.

Compared with other members of the AlkB family, the NRL of *h*ALKBH1 has several unique structural features (Supplementary information, S4). The *h*ALKBH1 Flip1 region is unique and long, leaving a larger binding space over the active site pocket (Fig. 1g). K158A/R159A/R160A/R162A and K167A/R169A mutants completely lost the 6mA demethylation activity, and the activity of K158A/R159A/R160A mutant was significantly compromised (Fig.1h). Moreover, the Flip2 contains a pair of antiparallel β-sheets and a long loop with high B factors (Supplementary information, Fig. S5). Structure-based sequence analyses and mutagenesis experiments confirmed that key residues in the NRL region, such as Arg169, Trp170, Tyr177, and Trp179, are potential determinants of *h*ALKBH1 for 6mA recognition and demethylation (Fig. 1i; Supplementary information, Fig. S6). The distinctive composition and conformation of Flip1 and Flip2 are likely to confer substrate selectivity on *h*ALKBH1.

Intriguingly, a significant structural feature of *h*ALKBH1 at the N-terminus is the Flip0 (residues 19–32). It is highly conserved in ALKBH1 among various mammalian species, but not in *Os*ALKBH1[15] or other AlkB family members (Fig. 1j; Supplementary information, Fig. S7). When superimposed with the *h*ALKBH2-dsDNA complex structure, the Flip0 region of *h*ALKBH1 is well accommodated by the modified strand, and the modified nucleotide enters the catalytic pocket. However, Flip0 seriously collides with the unmethylated strand (Fig. 1k). We hypothesized that Flip0 impedes the access of paired dsDNA to the active site and this may be the structural basis for the selectivity of *h*ALKBH1 toward single-stranded substrates. EMSA data revealed that dsDNA and ssDNA can bind to *h*ALKBH1, and EMSA and chromatographic analyses showed a direct binding between Flip0 and ssDNA/dsDNA (Supplementary information, Figs. S8 and S9). However, *h*ALKBH1 has no demethylation activity on 6mA of dsDNA. Based on the above results, we believe that Flip0 is disadvantageous for 6mA dsDNA to enter the *h*ALKBH1 catalytic pocket, and that dsDNA and *h*ALKBH1 are a non-productive combination in the presence of Flip0, rather than a productive way. Flip0 binds tightly adjacent to the minor sheet of the DSBH fold, and most of the interactions are hydrophobic with all the

Received: 31 May 2019 Accepted: 2 September 2019
Published online: 12 February 2020



involved hydrophobic residues conserved in ALKBH1 proteins of different species except for *Os*ALKBH1 (Supplementary information, Fig. S7a). R24A/K25A, R28A/R31A, R24A/K25A/R28A/R31A mutants and Flip0-deleted mutant *h*ALKBH1$_{36–369}$ completely lost demethylation activity toward 6mA ssDNA (Fig. 1l). Therefore, the unique Flip0 is essential for *h*ALKBH1 activity and discriminating single-stranded from paired double-stranded substrates. This structural characteristic and function of *h*ALKBH1 Flip0 are greatly different from those of Flip3 in *h*ALKBH5 (Fig. 1j),[11] whose structure can be disrupted by reduction of the disulfide bond. So *h*ALKBH1 uses a novel substrate recognition mechanism that is distinct from those of known AlkB demethylases.

Fig. 1 Structural and biochemical studies of *h*ALKBH1. **a** Primary structure of *h*ALKBH1. **b** Detection of *h*ALKBH1 demethylation activity by electrophoresis. The demethylated *h*ALKBH1 product at 1 h was enzymatically digested using nuclease *Dpn*II. The results showed that *h*ALKBH1 exhibited demethylation activity toward ssDNA containing 6mA, but not 6mA dsDNA, m^1A ssDNA or m^1A dsDNA. **c** A cartoon representation of *h*ALKBH1-Mn^{2+}-α-KG structure. Mn^{2+}, purple ball. The residues without visible electron density, dashed lines. Flip0, purple. Flip1, blue. Flip2, cyan. **d** Structural comparison of the active centers of *h*ALKBH1-Mn^{2+}-α-KG (purple) and *h*ALKBH1-Mn^{2+} (yellow). Interactions between Arg-344, Tyr-184, and Glu-236 are shown in dashed lines. **e** The interaction network around Mn^{2+} and α-KG. The $2mF_o$-DF_c electron density map (light blue), 1.5 σ. **f** Demethylation activity of the wide-type *h*ALKBH1 and its mutants toward 6mA ssDNA at 1 h. **g** Structural comparison of *h*ALKBH1 (purple) with other AlkB proteins. AlkB (cyan; PDB 2FDH), *h*ALKBH2 (green; 3BTZ), *h*ALKBH3 (yellow; 2IUW), *h*ALKBH5 (tv-blue; 4NRM), and *h*FTO (gray; 4QKN). *h*ALKBH1-Flip1, Flip1 of other AlkB proteins and Flip2 are highlighted in purple, blue and yellow boxes, respectively. **h, i** Mutations of the key residues in the lid region greatly impair *h*ALKBH1 demethylation activity toward 6mA ssDNA at 1 h. **j** Structural comparison of *h*ALKBH1 (purple) with AlkB, *h*ALKBH2, *h*ALKBH3, *h*ALKBH5. The *h*ALKBH1-Flip0 and *h*ALKBH5-Flip3 are highlighted within the purple frame. **k** Structural alignment of *h*ALKBH1 with *h*ALKBH2-dsDNA complex (dsDNA in orange; 3BTZ). The unmethylated strand of dsDNA would sterically clash with Flip0 of *h*ALKBH1. m^1A and α-KG (purple) are shown in sticks. **l** Mutations in the Flip0 region greatly impair *h*ALKBH1 demethylation activity at 1 h. **m** Schematic diagram of bubble- and bulge-structured DNA. **n**, **o** Detection of *h*ALKBH1 demethylation activity at 30 min by electrophoretic and statistical analysis. *h*ALKBH1 had the strongest demethylation activity toward Bugle6. Error bars, SD of three replicates. **p–r** HPLC detection of 6mA demethylation of dsDNA, ssDNA or Bulge6 DNA respectively by *h*ALKBH1. The relative content of 6mA is the same when using T, C or G as internal standard, respectively. **s**, **t** Proposed model of *h*ALKBH1 binding to Bubble6 (**s**) or Bulge6 DNA (**t**). DNA substrates, orange. Surface colored as a gradient ranging from red (negative) to blue (positive). Surface potential (±72 kBT/e) of *h*ALKBH1. The 6mA nucleobase inserted into the catalytic pocket is indicated

Next, we examined the surface charge distribution of *h*ALKBH1. The region of positive charge extends along with Flip1, Flip0, and Flip2 near the active center of the DSBH domain, and forms a large substrate-binding groove. Combining the substrate-binding ability and catalytic activity of *h*ALKBH1 mutants with mutations in Flip0, Flip1 and Flip2 regions, a possible binding model of *h*ALKBH1 with ssDNA is proposed: one end of single-stranded nucleic acid interacts with Flip1, and the other end extends along the direction of the groove formed by Flip2 and Flip0 (Supplementary information, Fig. S10d).

Compared with other AlkB family proteins (Supplementary information, Fig. S10b), the distribution of positive charges of *h*ALKBH1 is different and the area is very large. The positive surface of *h*ALKBH1 significantly exceeds that of ssDNA binding in the *h*ALKBH1-ssDNA model. In addition, the demethylation activity of *h*ALKBH1 on ssDNA is weak, probably because ssDNA is not the most suitable substrate. Based on structure, we introduced various substrates, such as hemi-methylated DNA bubble and bulge with different numbers of mismatched base pairs in the middle of double-stranded DNA (Fig. 1m; Supplementary information, Fig. S11). Because the HPLC method was not sensitive enough, we developed a high-throughput methylation-sensitive restriction digest assay to detect the demethylation activity of *h*ALKBH1 toward DNA bubble/bulge. Among different DNA bubbles, *h*ALKBH1 displayed the highest demethylation activity when the number of mismatched base pairs was 5–7. The demethylation activity of *h*ALKBH1 on Bubble6 DNA was twice that toward ssDNA (Supplementary information, Fig. S12). Intriguingly, *h*ALKBH1 had strongest activity on Bulge6 DNA (Fig. 1m–o). Similar results were obtained by using the method of HPLC (Fig. 1p–r; Supplementary information, Fig. S13). Furthermore, we tested the demethylation activity toward DNA bulge with 6mA at different mismatch position, and found that the activity was the highest at the third mismatched base pair. Notably, the demethylation activity of *h*ALKBH1 to Bulge6 DNA was 3–4 folds relative to ssDNA (Fig. 1m–o; Supplementary information, Figs. S13 and S14), and the variation of *h*ALKBH1 demethylation activity toward DNA bulge/bubble, ssDNA or dsDNA was not due to different binding affinities between substrates and *h*ALKBH1 (Supplementary information, Fig. S15). Therefore, structure-based substrate screening reveals that DNA bubble/bulge is more suitable for *h*ALKBH1. The 6mA demethylation activity of *h*ALKBH5 on dsDNA, ssDNA, DNA bubble and bulge was also assayed. *h*ALKBH5 is a member of the AlkB family, and is also a Fe(II)/2-OG-dependent dioxygenase. Results showed that when the substrate was Bubble6 or Bulge6, the 6mA demethylation activity of *h*ALKBH5 was 41% or 52.5% of that toward ssDNA, respectively (Supplementary information, Fig. S16). Therefore, DNA bubble and bulge are specific substrates for *h*ALKBH1.

Based on the binding and activity assays (Supplementary information, Fig. S17), we presented hypothetical structural models of *h*ALKBH1-Bubble6 DNA and *h*ALKBH1-Bulge6 DNA (Fig. 1s, t). The length of the substrate-binding groove is ~40 Å, suitable for binding 5–7-nt bubble-structured DNA, consistent with the experimental results. The opening of the bubble structure promotes the binding and makes the 6mA insert deeper into the active center (Fig. 1s). In the *h*ALKBH1-Bulge6 DNA binding model, the double-stranded DNA extends along the groove formed by Flip0, Flip1, and Flip2. The Bulge6 DNA binds to the Flip1 and Flip2 near the active pocket (Fig. 1t). The conformation allows a DNA circle with diameter less than 12 Å, consistent with the experimental result that *h*ALKBH1 displayed the highest activity toward 6-nt bulge-structured DNA. In this binding model, 6mA at the third mismatched base pairs is more suitable to penetrate into the active center. Compared with ssDNA, partly unpaired dsDNA is more biologically relevant in genomic DNA, such as in mismatch repair. Our structure-based substrate screening reveals that DNA bulge and bubble, rather than ssDNA, are preferential substrates of *h*ALKBH1 and have more physiological significance. Further investigations are required to elucidate the detailed molecular mechanism.

In summary, our research revealed several unique structural features of *h*ALKBH1 and found its novel native substrates. Our findings can further be used to study the regulatory mechanism of 6mA modification in different basic biological processes and in the field of DNA epigenetics, to guide future drug research.

The atomic coordinates and structure factors of the *h*ALKBH1-Mn^{2+}-α-KG and *h*ALKBH1-Mn^{2+} complex have been deposited in the Protein Data Bank under the accession code 6IE2 and 6IE3.

ACKNOWLEDGEMENTS

We thank Prof. Dong-Cai Liang and Wenqing Xu for their advice and support throughout the project. We also thank Prof. Gongyi Zhang and Shaodong Dai for critical reading. We thank the staff of the beamlines BL19U1 and BL17U at Shanghai Synchrotron Radiation Facility for assistance during data collection. We also thank Yuanyuan Chen and Zhenwei Yang for help with ITC and Octet analyses at Protein science research platform of Institute of Biophysics. This work was supported by the National Natural Science Foundation of China Grants 31570794, 31872713, 31629002 and 31371310; the Chinese Academy of Sciences Pilot Strategic Science and Technology Projects B Grants XDB08010300; Extramural Scientists of State Key Laboratory of Agrobiotechnology Grant [2019SKLAB6-14].

AUTHOR CONTRIBUTIONS

X-X.Y. and Z.C. conceived the project. L-F.T. and Y-P.L. carried out protein purification, related binding, and enzymatic analysis. L-F.T., Y-P.L., Q.T. and L.C carried out crystallization and X-ray crystal work. L-F.T. and Y-P.L performed structural refinement. L.C., W.S. and W.W. contributed to molecular cloning and sample preparation. L-F.T., Y-P.L., Z.C. and X-X.Y. analyzed data and wrote the paper. All authors participated in manuscript revision and analysis of biochemical data.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at https://doi.org/10.1038/s41422-019-0233-9.

Competing interests: The authors declare no competing interests.

Li-Fei Tian[1,2], Yan-Ping Liu[2], Lianqi Chen[2,3], Qun Tang[2], Wei Wu[1], Wei Sun[2], Zhongzhou Chen[1] and Xiao-Xue Yan[2]

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, 100193 Beijing, China; [2]National Laboratory of Biomacromolecules, Chinese Academy of Sciences (CAS) Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 100101 Beijing, China and [3]College of Life Sciences, University of Chinese Academy of Sciences, 100049 Beijing, China

These authors contributed equally: Li-Fei Tian, Yan-Ping Liu

Correspondence: Zhongzhou Chen (chenzhongzhou@cau.edu.cn) or Xiao-Xue Yan (snow@ibp.ac.cn)

REFERENCES

1. Wion, D. & Casadesus, J. *Nat. Rev. Microbiol.* **4**, 183–192 (2006).
2. Luo, G. Z. et al. *Nat. Rev. Mol. Cell Biol.* **16**, 705–710 (2015).
3. Xiao, C. L. et al. *Mol. Cell* **71**, 306–318 (2018).
4. Xie, Q. et al. *Cell* **175**, 1228–1244 (2018).
5. Wu, T. P. et al. *Nature* **532**, 329–333 (2016).
6. Ougland, R. et al. *Stem Cells* **30**, 2672–2682 (2012).
7. Westbye, M. P. et al. *J. Biol. Chem.* **283**, 25046–25056 (2008).
8. Haag, S. et al. *EMBO J.* **35**, 2104–2119 (2016).
9. Kawarada, L. et al. *Nucleic Acids Res.* **45**, 7401–7415 (2017).
10. Liu, F. et al. *Cell* **167**, 816–828 (2016).
11. Feng, C. et al. *J. Biol. Chem.* **289**, 11571–11583 (2014).
12. Yang, C. G. et al. *Nature* **452**, 961–965 (2008).
13. Zhu, C. X. & Yi, C. Q. *Angew. Chem. Int. Ed.* **53**, 3659–3662 (2014).
14. Zhang, X. et al. *Proc. Natl Acad. Sci. USA.* **116**, 2919–2924 (2019).
15. Zhou, C. et al. *Nat. Plants* **4**, 554–563 (2018).