LETTER TO THE EDITOR

Cryo-EM structure of the ssDNA-activated SPARTA complex

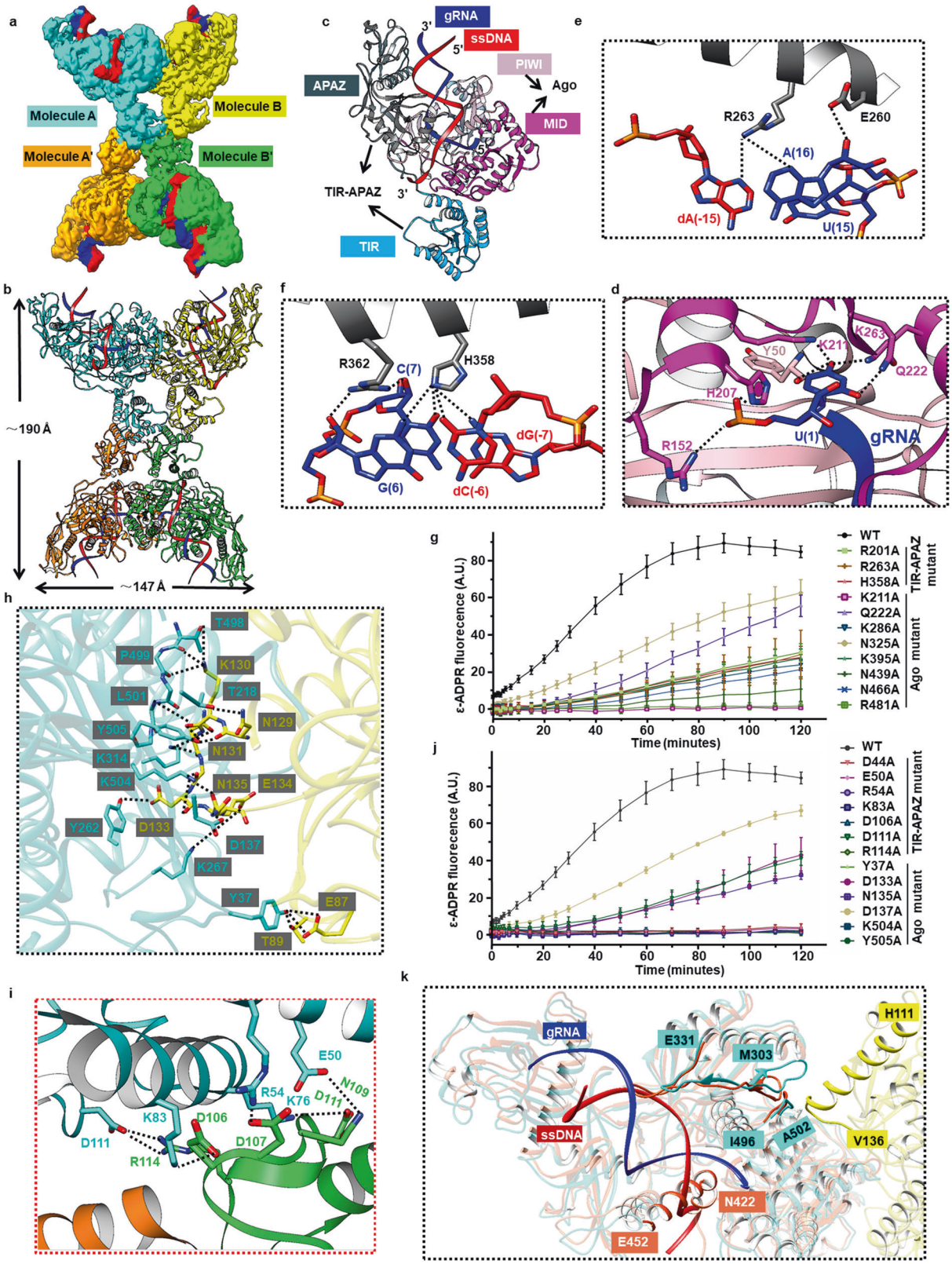
Cell Research (2023) 33:731–734; https://doi.org/10.1038/s41422-023-00850-y

Dear Editor,

Argonaute (Ago) proteins represent a highly conserved gene family, and are found in ~10% bacteria, ~30% archaea and almost all eukaryotes.[1] The eukaryotic Ago (eAgo) proteins are the core element of RNA-induced silencing complex, which is a key player in RNA interference pathways and regulates gene expression at transcriptional, posttranscriptional and translational levels.[2] The prokaryotic Ago (pAgo) proteins function in host defense against foreign genetic elements and can be divided into three categories: long pAgos, short pAgos, and PIWI-RE proteins.[3] A recent study has shown that short pAgo physically interacts with TIR-APAZ protein to constitute a catalytically inactive SPARTA (short prokaryotic Argonaute and the associated TIR-APAZ) immune system.[4] Small RNA-mediated detection of exogenous single-stranded DNA (ssDNA) triggers the activation of the SPARTA complex, which results in cellular NAD^+ depletion and consequentially cell death.[4] However, the molecular mechanisms underlying guide RNA (gRNA)-mediated invading DNA recognition and subsequent activation of the SPARTA complex remain opaque. In this study, we determined the cryo-EM structure of the SPARTA complex bound to gRNA and target ssDNA at 2.84 Å resolution (Supplementary information, Figs. S1, S2 and Table S1). The structure of SPARTA–gRNA–ssDNA complex is composed of four SPARTA complexes (Ago/TIR-APAZ heterodimer) and four gRNA–ssDNA heteroduplexes with a 4:4 stoichiometry, measuring 147 Å × 190 Å in size (Fig. 1a, b). Tetrameric assembly of the SPARTA complex adopts a butterfly-shaped architecture (A, A', B and B') (Fig. 1a, b). Four TIR domains locate in the center to form the body, and two dimeric Ago–APAZ heterodimers stand on each side to shape the wings with an included angle of ~90° (Supplementary information, Fig. S3b–d). Each two adjacent Ago–APAZ heterodimers bind respective gRNA–ssDNA heteroduplex at opposite sides (Fig. 1a, b). The short Ago protein consists of the MID and PIWI domains only, and shares 21% and 17% sequence identities with the MID-PIWI domains of *Cb*Ago (from *Clostridium butyricum*) and *Tt*Ago (from *Thermus thermophilus*), respectively (Fig. 1c; Supplementary information, Fig. S4a). Structural superpositions of the monomic Ago–APAZ–gRNA–ssDNA with *Cb*Ago–gDNA–tgDNA (PDB: 6QZK) and *Tt*Ago–siDNA–tgDNA (PDB: 4NCB) reveal that the architecture and conformation of the MID-PIWI domains between these three structures are highly similar (root mean square deviation (RMSD) values of 3.4 Å over 269 residues and 3.2 Å over 175 residues) (Supplementary information, Fig. S4b). Considering that the APAZ domain does not exhibit obvious sequence homology to the proteins with available structures in the PDB database, the Dali server was used to search for proteins whose structures are similar to that of the APAZ domain. The results revealed that the APAZ domain was structurally homologous to the tandem N and L1 domains of *Tt*Ago (PDB: 4N41), with a Dali Z-score of 10.9 and an RMSD of 3.9 Å for 180 Cα atoms. The TIR domain of TIR-APAZ protein displays a sandwich fold, with five α helices packing against the central five-stranded parallel β sheet (Supplementary information, Fig. S3d). Structural superimposition indicated that the TIR domain in Molecule B (B') rotates ~180° compared to that in Molecule A (A'), which breaks the symmetry of the tetramer (Supplementary information, Fig. S3e). In contrast, the architectures of Molecules A and B are almost identical to those of Molecules A' and B', respectively.

The well-resolved electron density enabled us to build a 5′-phosphorylated 19-nt gRNA and a 24-nt target ssDNA, located at the central basic channel formed by Ago and the APAZ domain of TIR-APAZ. The remaining 2 nt at the C-terminus of the gRNA and 7 nt at the N-terminus of the target ssDNA are disordered (Supplementary information, Fig. S3a). The gRNA is fully complementary to the target ssDNA spanning positions 3–19 to form a heteroduplex (A(3):dT(−3)–A(19):dT(−19)), leaving the 5′-end U(1) of gRNA inserted into the MID domain and the 3′-end dC(−2)–dA(5) of target ssDNA extending towards the solvent (Supplementary information, Fig. S5). The 5′-phosphorylated U(1) makes extensive hydrogen bonds with residues Arg^{152}, His^{207}, Lys^{211}, Gln^{222}, Lys^{263} and the C-terminal carboxyl group of Ago (Fig. 1d). These results are consistent with the observation that the SPARTA complex is preferentially activated by the binding of 5′-phosphorylated RNA-guided target ssDNA, and that gRNA 3′-end binding is not necessary for SPARTA activation.[4] The MID (residues Arg^{72}, Arg^{225}, Thr^{228}, Arg^{243}, Lys^{247}), PIWI (residues Lys^{286}, Asn^{325}, Gln^{327}, Lys^{362}, Arg^{364}, Lys^{390}, Lys^{395}, Asn^{439}, Asn^{466}, Asn^{468}, Arg^{481}, Lys^{485}) and APAZ (residues Arg^{201}, Arg^{209}, Lys^{211}, Arg^{257}, Asn^{270}, Ser^{288}, Lys^{313}, Lys^{315}, Lys^{328}, His^{340}, Ser^{359}, Arg^{361}, Lys^{366}) domains make extensive hydrogen bonds with the backbone phosphates of the gRNA–ssDNA heteroduplex (Supplementary information, Fig. S5). An important discovery is that two pairs of spatially neighboring residues (His^{358}:Arg^{362} and Glu^{260}:Arg^{263}) on the APAZ domain anchor into the minor groove of the gRNA–ssDNA heteroduplex at each end, and form hydrogen bonding interactions with the ribose of G(6)–C(7), and A(16) of gRNA, respectively (Fig. 1e, f). Furthermore, these residues contribute hydrogen bonds with the nucleobases of G(6):dC(−6), C(7):dG(−7), U(15):dA(−15) and A(16):dT(−16) base pairs (Fig. 1e, f). Our result supports the observation that mismatches between the gRNA and target ssDNA at guide positions 5–10 and 14–15 attenuated the activity of SPARTA complex.[4] We speculated that the unique structural feature of APAZ domain could facilitate the RNA-guided base pairing with the target ssDNA and monitor possible mismatches, compensating for the loss of PAZ domain in the SPARTA complex. As the SPARTA complex could be activated upon guide-mediated target binding, we investigated whether the residues involved in the recognition of the gRNA–ssDNA heteroduplex play a critical role in the activation of SPARTA complex. We constructed eleven mutants, including Ago^{K211A}, Ago^{Q222A}, Ago^{K286A}, Ago^{N325A}, Ago^{K395A}, Ago^{N439A}, Ago^{N466A}, Ago^{R481A}, TIR-$APAZ^{R201A}$, TIR-$APAZ^{R263A}$, and TIR-$APAZ^{H358A}$. Consistent with the structural observation, these mutations resulted in modest or severe reductions in the NADase activity of SPARTA complex (Fig. 1g), suggesting that guide-mediated target binding may be the trigger for the activation of SPARTA complex.

Received: 28 April 2023 Accepted: 30 June 2023
Published online: 25 July 2023



Both the MID domain of Ago and the TIR domain of TIR-APAZ participate in the tetramerization of the SPARTA–gRNA–ssDNA complex. The MID domains of two neighboring Agos pack against each other with a back-to-back arrangement, positioning the PIWI domains in opposite directions. This creates two symmetric binding interfaces, mediated by extensive hydrogen bonds and salt bridges. For simplicity, we only discuss one of them as an example. The structure shows that residue Tyr[37] (MID-A) forms three hydrogen bonds with the side chains of residues Glu[87] and Thr[89] (MID-B) (Fig. 1h). Residues Asp[137], Thr[218], Tyr[262] and Lys[314]

Fig. 1 Cryo-EM structure of the SPARTA–gRNA–ssDNA complex. a, **b** Cryo-EM density map and cartoon illustration of the side view of the tetrameric SPARTA–gRNA–ssDNA complex. Four asymmetric molecules are shown in cyan, yellow, orange, and green, respectively. **c** Presentation of the monomeric SPARTA–gRNA–ssDNA complex. The MID, PIWI, TIR, APAZ domains, gRNA and target ssDNA are shown in magenta, pink, light-blue, black, dark-blue and red, respectively. **d** Interaction details of the 5′-U(1) recognition by the SPARTA complex. The residues involved in the interaction are labeled and shown as sticks. **e**, **f** Interaction details of the SPARTA complex with the nucleobases and riboses of the gRNA–ssDNA heteroduplex. The residues involved in the interactions are labeled and shown as sticks. **g** Effect of the residues involved in the recognition of the gRNA–ssDNA heteroduplex on the activity of the SPARTA complex. Error bars represent SD; $n = 3$. **h**, **i** Enlarged view of the tetrameric interface of the SPARTA–gRNA–ssDNA complex. The residues involved in the interactions are labeled and shown as sticks, and hydrogen bonds are indicated by dashed lines. **j** Effect of the residues involved in the tetrameric assembly on the activity of the SPARTA complex. Error bars represent SD; $n = 3$. **k** Structural comparison of Molecule A in the tetrameric SPARTA–gRNA–ssDNA complex with the Apo-SPARTA complex predicted by AlphaFold. The SPARTA–gRNA–ssDNA complex is colored the same as in **b**, and the Apo-SPARTA complex is colored orange red.

(MID-A) contribute another five hydrogen bonds with residues Asn^{135}, Asn^{129}, Asp^{133} and Asn^{131} (MID-B) (Fig. 1h). Residue Lys^{267} (MID-A) makes a salt bridge with Glu^{134} (MID-B) (Fig. 1h). In addition, residues 498–505 at the C-terminus of Ago (MID-A) form nine hydrogen bonds with the adjacent main-chain carbonyl of residues 129–135 (MID-B) to further stabilize the interface (Fig. 1h). The tetrameric TIR domains display an architecture of two parallel strands, similar to that observed in the human TLR adaptor proteins, such as myeloid differentiation primary-response gene 88 (MyD88) and MyD88-adaptor-like protein (MAL).[5] Each strand is constituted of two head-to-tail-arranged TIR domains, held together through the "BE" interface. The "BE" interface includes the area around the BB loop (βB–αB) of one subunit and the βD, βE and αE-enclosed EE surface of the neighboring subunit (Supplementary information, Fig. S3d). The two strands are linked by the "BCD" interface mediated by extensive hydrogen bonds and salt bridges, involving the residues (Glu^{50}, Arg^{54}, Lys^{76}, Lys^{83} and Asp^{111}) on the BC surface (αB–αC) of one subunit and the residues (Asp^{106}, Asp^{107}, Asn^{109}, Asp^{111}, and Arg^{114}) on the CD surface (αD and the CD loop) of another subunit (Fig. 1i). To investigate whether the residues mediating the formation of the tetrameric SPARTA–gRNA–ssDNA complex are required for NADase activity, we designed six Ago mutants (Ago^{Y37A}, Ago^{D133A}, Ago^{N135A}, Ago^{D137A}, Ago^{K504A}, Ago^{Y505A}) and seven TIR-APAZ mutants (TIR-$APAZ^{D44A}$, TIR-$APAZ^{E50A}$, TIR-$APAZ^{R54A}$, TIR-$APAZ^{K83A}$, TIR-$APAZ^{D106A}$, TIR-$APAZ^{D111A}$, TIR-$APAZ^{R114A}$). As expected, these mutations obviously reduced the NADase activity of SPARTA complex, confirming the significance of tetramerization of SPARTA complex for the activation of NADase activity (Fig. 1j).

Size exclusion chromatography analyses of Apo-SPARTA, SPARTA–gRNA and SPARTA–gRNA–ssDNA revealed that only the SPARTA–gRNA–ssDNA formed a tetrameric complex with a 4:4:4:4 stoichiometry (TIR-APAZ:Ago:guide:target) (Supplementary information, Fig. S6a). To investigate the molecular mechanism underlying target ssDNA binding-triggered activation of SPARTA complex,[4] we predicted the structure of Apo-SPARTA complex using AlphaFold.[6] Structural superposition of the Apo-SPARTA complex with monomeric SPARTA–gRNA–ssDNA (Molecule A) reveals that the architectures of the Ago and TIR-APAZ proteins between these two structures are highly similar (RMSD of 1.2 Å over 810 residues) (Fig. 1k). The gRNA-mediated binding of target ssDNA causes conformational changes of the loops (sensor loops) spanning residues 303–331 and 496–502, a region located at the interactive interface of two Agos (Fig. 1k). As shown in Fig. 1h, several residues at this region contribute to the tetramerization of SPARTA complex. The size exclusion chromatography results showed that Ago mutant ($Ago^{N131A/E132A/K314A}$) is monomeric, confirming that the sensor loops are vital for the tetramerization of SPARTA complex (Supplementary information, Fig. S6b). These results support that the sensor loops might serve as a checkpoint to monitor the fidelity of base pairing between gRNA and target ssDNA. Additionally, the gRNA–ssDNA heteroduplex also makes a steric clash with the C-terminal α-helix (residues 422–452) of TIR-APAZ protein, which may restrict the self-association of TIR domains. Thus, we conclude that the allosteric effect of sensor loops drives the dimerization of two neighboring Agos, which, together with the self-association of four TIR domains, fulfills the tetramerization of SPARTA complex.

The tetramerization of SPARTA–gRNA–ssDNA complex brings four adjacent TIR domains together, which creates two composite catalytic sites, each formed by an asymmetric TIR homodimer. Structural superimposition of the TIR domains between Molecule A (A') and Molecule B (B') indicates that the BB loops of active TIR subunits undergo a drastic conformational change, which switches the catalytic pockets from closed to open (Supplementary information, Fig. S7a). TIR domain-mediated NAD^+ depletion is a common immune strategy found in both prokaryotic and eukaryotic immune systems. Moreover, TIR domain-containing proteins can catalyze NAD^+ for the production of signaling molecules (ADPR, cADPR or the distinct cADPR isomers) to activate the downstream immunity.[7] Self-association to form an oligomer is a general mechanism of the activation of TIR domains, which has been found in plant nucleotide-binding leucine-rich repeat receptors (NLRs)[8,9] and human sterile alpha and TIR motif-containing protein 1 (SARM1).[10] Recently, the cryo-EM structure of the TIR domain of *Ab*Tir (from *Acinetobacter baumannii*, $AbTir^{TIR}$) bound to NAD^+ mimic has been resolved, which helps understand the process of TIR-mediated NAD^+ hydrolysis to produce nicotinamide (NAM) and ADPR or the distinct cADPR isomers (2′cADPR or 3′cADPR).[7] The TIR domain of TIR-APAZ shares 24% sequence identity with the $AbTir^{TIR}$. Sequence analysis of these two TIR domains showed that the residues involved in the recognition of NAD^+ mimic are conserved despite some differences (Supplementary information, Fig. S7b). Interestingly, the tetrameric $AbTir^{TIR}$ also displays a scaffold assembly, similar to the TIR domains of TIR-APAZ, MyD88 and MAL. Structural superimposition of the tetrameric TIR domains with $AbTir^{TIR}$ complexed with NAD^+ mimic (PDB ID: 7UXU) indicated that residues Thr^{11}, Trp^{33}, Asp^{35}, Leu^{39}, Arg^{71}, Glu^{77}, His^{141} of TIR-B together with residues Tyr^{105}, Asn^{116}, Lys^{121} of TIR-A form a pocket for NAD^+ binding, and that residue Glu^{77} is very likely an executor for the catalytic reaction (Supplementary information, Fig. S7c). The residue Trp^{204} of $AbTir^{TIR}$ vital for ADPR cyclization is mutated to Glycine in TIR-APAZ (Supplementary information, Fig. S7c). This is consistent with the finding that activated SPARTA complex converts NAD^+ into ADPR and NAM inducing NAD^+ depletion-triggered cell death.[4]

Taken together, our results indicate that the TIR-APAZ protein compensates for the lack of N-PAZ domain in short pAgo, cooperating with the MID-PIWI domain of short Ago to function as a sensor for exogenous ssDNA. gRNA-mediated target ssDNA binding triggers the conformational changes of the SPARTA complex for tetramerization, activating the NADase activity of TIR domain.

Minghui Guo[1,4], Yuwei Zhu[1,4], Zhiying Lin[1,4], Dehui Yang[1], Anqi Zhang[1], Changyou Guo[1] and Zhiwei Huang[1,2,3]✉

[1]HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China. [2]Westlake Center for Genome Editing, Westlake Laboratory of Life Sciences and Biomedicine, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. [3]New Cornerstone Science Laboratory, Shenzhen, Guangdong, China. [4]These authors contributed equally: Minghui Guo, Yuwei Zhu, Zhiying Lin. ✉email: huangzhiwei@hit.edu.cn

DATA AVAILABILITY

The atomic coordinate of the SPARTA–gRNA–ssDNA complex has been deposited to the Protein Data Bank under the accession number 8J7S. The corresponding map has been deposited in the Electron Microscopy Data Bank under the accession number EMD-36047.

REFERENCES

1. Hutvagner, G. & Simard, M. J. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32 (2008).
2. Iwakawa, H. O. & Tomari, Y. *Mol. Cell* **82**, 30–43 (2022).
3. Swarts, D. C. et al. *Nature* **507**, 258–261 (2014).
4. Koopal, B. et al. *Cell* **185**, 1471–1486.e19 (2022).
5. O'Neill, L. A. & Bowie, A. G. *Nat. Rev. Immunol.* **7**, 353–364 (2007).
6. Jumper, J. et al. *Nature* **596**, 583–589 (2021).
7. Manik, M. K. et al. *Science* **377**, eadc8969 (2022).
8. Ma, S. et al. *Science* **370**, eabe3069 (2020).
9. Martin, R. et al. *Science* **370**, eabd9993 (2020).
10. Gerdts, J., Brace, E. J., Sasaki, Y., DiAntonio, A. & Milbrandt, J. *Science* **348**, 453–457 (2015).

ACKNOWLEDGEMENTS

We thank the EM and NMR platform at School of Life Science and Technology of Harbin Institute of Technology and the cryo-EM facility, the high-performance computing center core facility of Westlake University for providing supports. This research was funded by the National Natural Science Foundation of China (31825008 and U21A20276), the Tencent Foundation through the XPLORER PRIZE, the New Cornerstone Science Foundation, and Heilongjiang Touyan Team (HITTY-20190034) to Z.H., Postdoctoral Scientific Research Fund of Heilongjiang Province (AUGA4120002723) to Y.Z., and the Fellowship of China National Postdoctoral Program for Innovative Talents (AUGA4150000720) to M.G.

AUTHOR CONTRIBUTIONS

Z.H. conceived and supervised the whole project. M.G. and Z.L. prepared the protein samples for negative-staining and cryo-EM analysis. M.G., Z.L., A.Z. and C.G. did negative-staining, prepared the cryo-EM grids and collected the cryo-EM data. M.G., Z.L. and D.Y. did functional experiments. Y.Z. and Z.H. performed cryo-EM data processing and model building. Y.Z. wrote the manuscript with contributions from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41422-023-00850-y.

Correspondence and requests for materials should be addressed to Zhiwei Huang.

Reprints and permission information is available at http://www.nature.com/reprints