ARTICLES
https://doi.org/10.1038/s41594-019-0206-1

nature structural & molecular biology

An anti-CRISPR protein disables type V Cas12a by acetylation

Liyong Dong[1,4], Xiaoyu Guan[1,4], Ningning Li[2,4], Fan Zhang[1], Yuwei Zhu[1], Kuan Ren[1], Ling Yu[1], Fengxia Zhou[1], Zhifu Han[3], Ning Gao[2] and Zhiwei Huang[1*]

Phages use anti-CRISPR proteins to deactivate the CRISPR–Cas system. The mechanisms for the inhibition of type I and type II systems by anti-CRISPRs have been elucidated. However, it has remained unknown how the type V CRISPR–Cas12a (Cpf1) system is inhibited by anti-CRISPRs. Here we identify the anti-CRISPR protein AcrVA5 and report the mechanisms by which it inhibits CRISPR–Cas12a. Our structural and biochemical data show that AcrVA5 functions as an acetyltransferase to modify *Moraxella bovoculi* (Mb) Cas12a at Lys635, a residue that is required for recognition of the protospacer-adjacent motif. The AcrVA5-mediated modification of MbCas12a results in complete loss of double-stranded DNA (dsDNA)-cleavage activity. In contrast, the Lys635Arg mutation renders MbCas12a completely insensitive to inhibition by AcrVA5. A cryo-EM structure of the AcrVA5-acetylated MbCas12a reveals that Lys635 acetylation provides sufficient steric hindrance to prevent dsDNA substrates from binding to the Cas protein. Our study reveals an unprecedented mechanism of CRISPR–Cas inhibition and suggests an evolutionary arms race between phages and bacteria.

CRISPR–Cas adaptive immune systems provide bacteria and archaea with a nucleic acid sequence–specific defense mechanism against phages or plasmid invaders[1–4]. CRISPR RNA (crRNA)-guided single Cas proteins (class II) or multi-component proteins (class I) recognize and degrade target DNA or RNA sequences that are complementary to the guide sequence of crRNA. The CRISPR–Cas systems are divided into six types (I–VI). Notably, the type II CRISPR–Cas9 (refs. [5–7]) and type V CRISPR–Cas12a[8–10] systems have been harnessed as the widely used tools for genome editing and various biotechnological applications. RNA-guided Cas effector protein Cas9 or Cas12a recognizes a specific sequence segment of a protospacer-adjacent motif (PAM) located in the target double-stranded DNA (dsDNA) by the PAM-interacting domain. Following separation of two strands of target dsDNA by the PAM-interacting domain[11,12], the target strand forms a heteroduplex with guide RNA bearing the complementary sequences, resulting in cleavage of both strands of the target dsDNA by the nuclease domains of Cas9 or Cas12a.

In response to the CRISPR–Cas immune systems, phages have evolved protein inhibitors of CRISPR–Cas that are called anti-CRISPRs. Anti-CRISPR proteins enable phages to shut down and escape the CRISPR–Cas defense systems. A number of distinct families of anti-CRISPR proteins that target type I and type II CRISPR–Cas systems have been discovered[13–18]. Biochemical and structural studies have shown that these anti-CRISPR proteins inactivate the CRISPR–Cas systems by blocking target DNA binding[17] or endonuclease activity of Cas proteins[19]. Anti-CRISPR proteins targeting type V CRISPR–Cas systems were also reported in two recent studies[20,21]. Direct interaction with a Cas protein is a common strategy used by anti-CRISPR proteins for inhibition of the CRISPR–Cas systems. It remains unknown whether and how anti-CRISPRs can use other strategies such as an enzymatic activity to serve this purpose.

Results

Discovery of anti-CRISPR candidates for V-A CRISPR–Cas systems. Koonin et al.[22] previously showed that nearly all spacers target mobile genetic elements, including free elements and prophages integrated into host genomes. To counter such targeting, especially lethal self-targeting[23], phages, prophages and other mobile genetic elements (Supplementary Fig. 1a) have evolved anti-CRISPR strategies that inhibit diverse CRISPR–Cas systems. To identify inhibitors of the type V-A CRISPR-Cas system (anti–CRISPR type V-A proteins (AcrVAs)), we analyzed all available completely sequenced bacterial genomes from the NCBI RefSeq database using the CRISPRminer pipeline[24] (available at http://www.microbiome-bigdata.com/CRISPRminer/). In total, we found 24 genomes bearing the V-A subtype CRISPR–Cas systems, with approximately 40% of them belonging to the *Moraxella* genus. Interestingly, all of the identified genomes from the *Moraxella bovoculi* (Mb) species carried a V-A CRISPR (Cas12a) system. We therefore performed a comprehensive analysis of the spacers from the Mb V-A CRISPR arrays, and found that 31 (~25%) out of the 122 unique spacers potentially target the prophage regions (Supplementary Fig. 1b). This result suggests an antagonistic relationship between V-A CRISPR and prophages. Notably, four prophage regions were the hot spots of spacer attack but tolerated multiple self-targeting spacers (Supplementary Note 1), suggesting that these prophage regions are likely to contain anti-CRISPRs.

To map the anti-CRISPRs of the type V-A CRISPR-Cas12a system, we first screened 50 gene products encoding fewer than 300 amino acids from the second prophage region (prophage 2) in Mb 22581, which encodes a Cas12a system and tolerates most self-targeting spacers (Supplementary Fig. 1b). We then tested their MbCas12a-inhibiting activity using in vitro and in vivo assays. We successfully purified 49 proteins of the 50 anti-CRISPR candidates. The data from the cleavage inhibition assay showed that two of the purified proteins, candidates 49 and 50, strongly inhibited the DNA cleavage activity of MbCas12a (Supplementary Fig. 1c).

[1]HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin, China. [2]State Key Laboratory of Membrane Biology, Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing, China. [3]School of Life Sciences, Tsinghua University, Beijing, China. [4]These authors contributed equally: Liyong Dong, Xiaoyu Guan, Ningning Li. *e-mail: huangzhiwei@hit.edu.cn

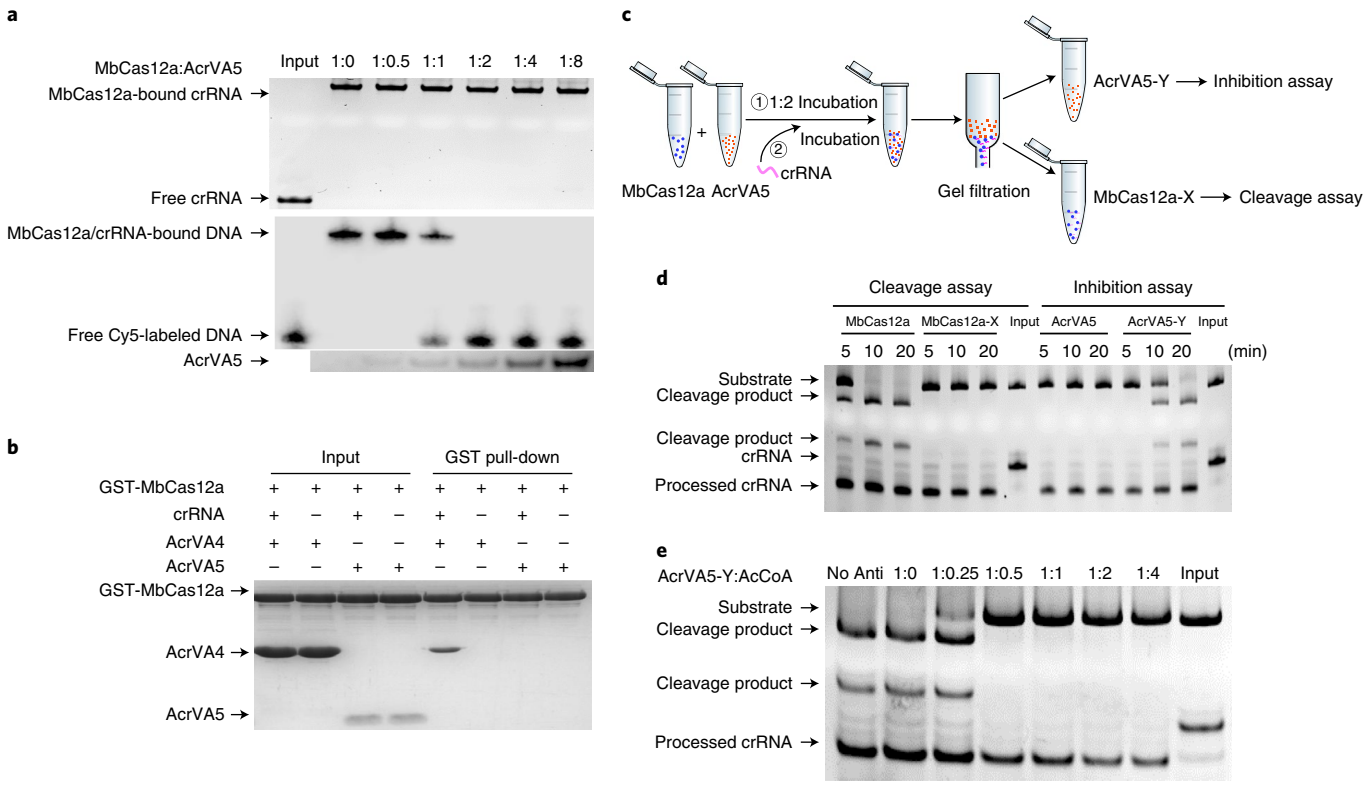

Fig. 1 | AcrVA5 inhibits MbCas12a activity through acetyltransferase activity. a, EMSA used to test the crRNA- and dsDNA-binding activity of MbCas12a in the presence of AcrVA5. Data shown are representative of three independent experiments. **b**, GST pull-down assays used to test interactions between AcrVA4 or AcrVA5 and the GST-tagged MbCas12a in the presence or absence of crRNA. Data shown are representative of three replicates. **c**, Diagram of the assay scheme for the separation of MbCas12a-X and AcrVA5-Y. **d**, Time course of cleavage and inhibition assays using purified MbCas12a-X and AcrVA5-Y, respectively. Representative time-points are shown at the top of each lane. Data shown are representative of three independent experiments. **e**, Inhibition assay using AcrVA5-Y protein in the presence of increasing concentrations of acetyl-CoA. The molar ratios of AcrVA5-Y:acetyl-CoA are shown at the top of each lane. The molar ratio of AcrVA5: Mbcas12a is 2:1. Data shown are representative of three independent experiments. Uncropped gel images for **a**,**b**,**d** and **e** are shown in Supplementary Dataset 1.

To test whether these two anti-CRISPR candidates are active in cells, we performed a plasmid-targeting assay to assess their inhibition activity in bacteria. The plasmids of candidates 49 and 50 displayed approximately 100- to 1,000-fold enhancements in transforming protospacer sequence-containing plasmids compared to other candidate plasmids or empty plasmids (Supplementary Fig. 1d), consistent with the data from the cleavage assays. These data collectively establish candidates 49 and 50 as anti-CRISPRs of MbCas12a. The same anti-CRISPR proteins were independently reported in a recent study[20]. Interestingly, Marino et al[21] discovered two different anti-proteins from another *Mb* strain through a different method. Based on the nomenclature of these studies, we refer to candidates 49 and 50 as AcrVA4 and AcrVA5, respectively.

Inhibitory effects of the cleavage of MbCas12a DNA by AcrVA4 and AcrVA5. Our in vitro inhibition assays showed that the proteins of AcrVA4 and AcrVA5 potently inhibited dsDNA cleavage by MbCas12a (Supplementary Fig. 2a). Interestingly, when MbCas12a and AcrVA4 were incubated for a few minutes and then the crRNA of MbCas12a was added into the reaction mixture, the inhibitory effect of AcrVA4 was notably promoted (Supplementary Fig. 2a). Unexpectedly, MbCas12a pre-incubated with the AcrVA4 protein was also greatly compromised in its activity of processing pre-crRNA. By contrast, a similar effect on activity was not detected for the AcrVA5 protein (Supplementary Fig. 2a). To explore the mechanisms of AcrVA4- and AcrVA5-mediated inhibition of MbCas12a, we tested the dsDNA binding activity of MbCas12a in the presence of either of these two anti-CRISPRs using EMSA (electrophoretic mobility shift assay). The results from the assays showed that proteins of both AcrVA4 and AcrVA5 substantially reduced the dsDNA-binding activity of MbCas12a (Fig. 1a and Supplementary Fig. 2b). By contrast, these two proteins had no detectable effect on the interaction of MbCas12a with crRNA (Fig. 1a and Supplementary Fig. 2b). Pull-down assays using glutathione *S*-transferase (GST)-fused MbCas12a showed that AcrVA4 interacted with the crRNA-bound but not free form of MbCas12a (Fig. 1b), reminiscent of the anti-CRISPR AcrIIA4, which targets only crRNA-bound SpCas9 (refs. [17,25]). Surprisingly, however, neither the free MbCas12a nor the crRNA-bound state was found to interact with AcrVA5 in similar assays (Fig. 1b). The detailed mechanism of AcrVA4-mediated inhibition of MbCas12a will be addressed in a different study. The current manuscript focuses on the inhibition mechanism of AcrVA5.

AcrVA5 impairs DNA cleavage activity by acetylation of MbCas12a. One explanation for the potent MbCas12a-inhibiting but not MbCas12a-binding activity of AcrVA5 may be that the Cas enzyme had been covalently modified by the anti-CRISPR protein, resulting in permanent inactivation of the Cas enzyme. Consistent with this hypothesis, careful sequence analysis revealed that the anti-CRISPR protein contains a segment of 29-Lys-Arg-Gln-Gly-Ile-Gly-34, which shares striking homology with the conserved (Arg/Gln)-X-X-Gly-X-(Gly/Ala) motif found in GCN5-related NAT (GNAT) superfamily acetyltransferases[26]. To test whether MbCas12a can be permanently inactivated by AcrVA5, we first incubated the two proteins with a molar ratio of approximately 1:2

Table 1 | Crystal data collection and refinement statistics

	Se_AcrVA5	Native_AcrVA5 (6IUF)
Data collection		
Space group	P4132	P4132
Cell dimensions		
a, *b*, *c* (Å)	143.91, 143.91, 143.91	143.42, 143.42, 143.42
α, *β*, *γ* (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.9792	0.9792
Resolution (Å)	50.00–2.75 (2.85–2.75)[a]	50.00–2.05 (2.09–2.05)[a]
R_{merge}	0.106 (0.893)	0.052 (0.733)
$I/\sigma(I)$	29.3 (4.2)	52.1 (2.2)
Completeness (%)	100.0 (100.0)	99.9 (99.1)
Redundancy	20.4 (19.8)	11.7 (6.0)
Refinement		
Resolution (Å)		50.00–2.05
No. of reflections		31,986
R_{work}/R_{free}		0.152/0.186
No. of atoms	,	
Protein		1,532
Ligand/ion		108
Water		130
B factors		
Protein		32.6
Ligand/ion		32.3
Water		44.4
R.m.s.d.		
Bond lengths (Å)		0.014
Bond angles (°)		1.683

[a]Values in parentheses are for the highest-resolution shell.

Table 2 | Cryo-EM data collection, refinement and validation statistics

	AcrVA5-acetylated MbCas12a (EMDB 9742) (PDB ID 6IV6)
Data collection and processing	
Magnification	165,000
Voltage (kV)	300
Electron exposure ($e^-/Å^2$)	81/dose weighting
Defocus range (μm)	1.5–3
Pixel size (Å)	0.83
Symmetry imposed	C1
Initial particle images (no.)	578 K (after 2D classification)
Final particle images (no.)	93 K
FSC threshold	0.143
Map resolution range (Å)	200.0–3.6
Refinement	
Initial model used (PDB code)	5ID6
Model resolution (Å) FSC threshold	3.6 (0.143)
Model resolution range (Å)	200.0–3.6
Map sharpening *B* factor ($Å^2$)	−162
Model composition	
Nonhydrogen atoms	10,158
Protein residues	1,209
Nucleotide	26
B factors ($Å^2$)	
Protein	52.6
Ligand	33.28
R.m.s.d.	
Bond lengths (Å)	0.005
Bond angles (°)	1.016
Validation	
MolProbity score	1.69
Clashscore	4.91
Poor rotamers (%)	0.61
Ramachandran plot	
Favored (%)	93.26
Allowed (%)	6.66
Disallowed (%)	0.08

for 20 min at room temperature and then crRNA was added to the mixture (Fig. 1c). After incubation for a further 10 min, the mixture was subjected to gel filtration. In support of the pull-down data, the MbCas12a and AcrVA5 proteins were well separated in the assay (Supplementary Fig. 3a), confirming that the two proteins did not interact with each other. The MbCas12a protein recovered from the assay (hereinafter referred to as 'MbCas12a-X') displayed little activity cleaving dsDNA substrate (Fig. 1d). Unexpectedly, the inhibitory effect of the AcrVA5 protein recovered (hereinafter referred to as 'AcrVA5-Y') on MbCas12a-catalysed dsDNA cleavage was also significantly impaired (Fig. 1d). Members of the GCN5-related NAT superfamily of acetyltransferases employ acetyl-CoA as a cofactor for their enzymatic activity. Thus, one possibility for the unexpected observation is that AcrVA5 freshly purified from bacteria contained acetyl-CoA, and that the acetyl-CoA-bound form of AcrVA5 was converted into a CoASH-bound form after reaction with MbCas12a owing to consumption of the acetyl-CoA, making the anti-CRISPR enzymatically inactive. If this is the case, addition of acetyl-CoA is predicted to restore the inhibition activity AcrVA5-Y. Indeed, in the presence of 8 μM of acetyl-CoA, AcrVA5-Y displayed similar activity to the freshly purified AcrVA5 protein in inhibiting MbCas12a-mediated dsDNA cleavage (Fig. 1e). Taken together, these biochemical data suggest that AcrVA5 functions as an acetyltransferase to inhibit MbCas12a.

Crystal structure of AcrVA5 and its acetylation mechanism. To further verify whether AcrVA5 functions as an acetyltransferase, we determined the crystal structure of the anti-CRISPR protein at 2.05 Å (Table 1). Structural alignment using the DALI server[27] revealed that the amino-terminal acetyltransferase NatD[28] (PDB ID 4U9W) is the closest structural homologue of AcrVA5, with a root mean squared deviation (r.m.s.d.) of 1.6 Å over 86 aligned Cα atoms (Fig. 2a), despite their low sequence homology (7% identity). This observation provides structural evidence that AcrVA5 is an acetyltransferase. In the structure, AcrVA5, which consists of a pair of crossing α-helices bundle packing against a five-stranded anti-parallel β-sheet, forms a homodimer (Supplementary Fig. 3b), as observed in other GNAT family acetyltransferases. Compared to NatD, AcrVA5 is more compact because of its smaller size (92 amino acids). As seen in other GNAT family members, acetyl-CoA is bound in the structure of AcrVA5 (Fig. 2b) and adopts a

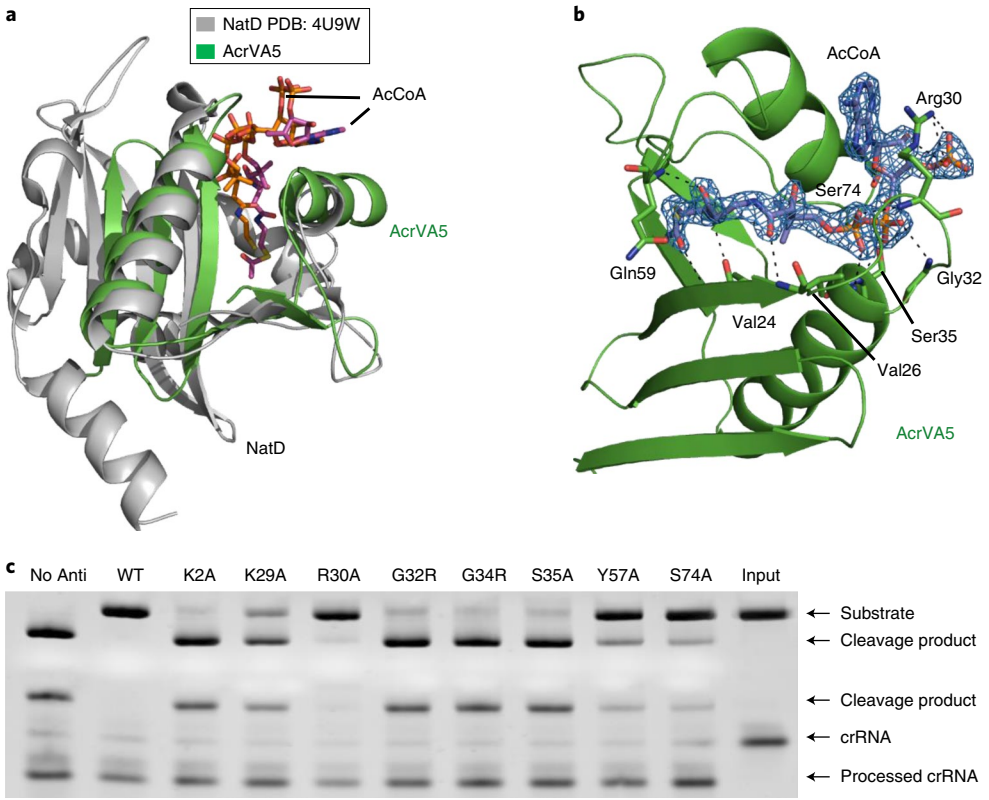

Fig. 2 | AcrVA5 adopts a typical acetyltransferase structure. a, Superimposition of the AcrVA5 (green) and the N-terminal acetyltransferase NatD (PDB ID 4U9W) (grey). The two acetyl-CoA (AcCoA) molecules are shown as stick diagrams. **b**, Detailed interactions of AcrVA5 with acetyl-CoA. Hydrogen bonds are shown as dashed lines. The electron density map of acetyl-CoA is shown in blue mesh. The 2*Fo*–*Fc* electron density map is contoured at 2.0 σ. **c**, Cleavage inhibition assays to verify structural determinants for specific acetyl-CoA recognition by AcrVA5 using wild-type (WT) and mutant AcrVA5 proteins. Data shown are representative of three independent experiments. Uncropped gel images for **c** are shown in Supplementary Dataset 1.

bent conformation with a sharp turn at the pantothenate moiety[26]. This structural observation further supports our biochemical data (Fig. 1e). The acetyl head and the pantetheine group of acetyl-CoA are buried inside the classic 'V'-shaped channel formed by two β-strands (Fig. 2b), forming hydrogen bonding interactions with the main chains of residues Val24, Val26 and Gln59. The diphosphate moiety of acetyl-CoA is bound predomiantly by the main chains of Arg30, Gly32, Gly34 and Ser35 from the atypical 'P-loop' 29-Lys-Arg-Gln-Gly-Ile-Gly-34 and the following α-helix (Fig. 2b). The K2A, K29A, G32R, G34R, S35A, Y57A and S74A mutations predicted to impair acetyl-CoA recognition abolished or greatly reduced the inhibition ability of AcrVA5 (Fig. 2c), providing further support to the theory that acetyl-CoA has a role in the inhibition of MbCas12a by AcrVA5.

AcrVA5 inhibits MbCas12a through acetylation of the critical PAM-recognition residue Lys635. The biochemical and structural data presented above establish AcrVA5 as an acetyltransferase to inhibit MbCas12a. We next sought to elucidate the structural mechanism of MbCas12a inhibition by AcrVA5. To this end, we first mapped the acetylation site (or sites) of MbCas12a by AcrVA5. We therefore co-expressed the two proteins in *Escherichia coli* and purified the MbCas12a protein for mass spectrometry analysis. The mass spectrometry data showed that several lysine residues of MbCas12a had been acetylated in the cells (Supplementary Dataset 2). This result indicates that AcrVA5 acetylation of MbCas12a is independent of crRNA. Notably, one of the mapped acetylated sites is Lys635 of MbCas12a (Fig. 3a). The crystal structure of *Lachnospiraceae bacterium* Cas12a (LbCas12a) has been solved. In the structure, Lys595, which is equivalent to Lys635 of MbCas12a, specifically interacts with N3 of dA(3*) and O2 of dT(2) from the 5′-TTTA-3′ PAM sequence[29]. The specific interaction has an essential role in unwinding dsDNA substrate and in the nuclease activity of LbCas12a. Modelling suggested that acetylation of the equivalent residue of MbCas12a can generate steric hindrance to block dsDNA binding. To test this idea, we first purified acetylated MbCas12a protein by AcrVA5 using the method described above and then determined a cryo-EM structure of the acetylated MbCas12a at a resolution of 3.6 Å (Fig. 3b,c, Supplementary Fig. 4 and Table 2). As expected, the cryo-EM structure strikingly resembles the crystal structure (Fig. 3d) of LbCas12a. The quality of the density map is sufficient to discern the acetylated side chain of residue Lys635 (Fig. 3e). The acetylation of Lys635 not only results in the loss of hydrogen bonding interactions with dT2 and dA(3*) but may also generate steric hindrance with neighbouring PAM DNA (Fig. 3f), thus preventing MbCas12a from binding dsDNA, as demonstrated by our biochemical data (Fig. 1b). Taken together, our biochemical and structural data show that AcrVA5 inhibits MbCas12a through acetylation of the critical PAM-interacting residue Lys635.

AcrVA5 acetylates lysine- but not arginine-dependent PAM-recognition residues of Cas12a. It is of interest to note that Lys635 of MbCas12a is not absolutely conserved among MbCas12a proteins from different Mb strains (Fig. 4a). MbCas12a proteins from some strains have an arginine residue at this position. This replacement may not abrogate the catalytic activity of an MbCas12a protein, because lysine and arginine residues are comparable in size and charge. Indeed, one such Cas12a protein, Cas12a from Mb strain 58069 (Mb2Cas12a), was fully active in cleaving dsDNA. By contrast, such replacement would prevent Mb2Cas12a from being acetylated

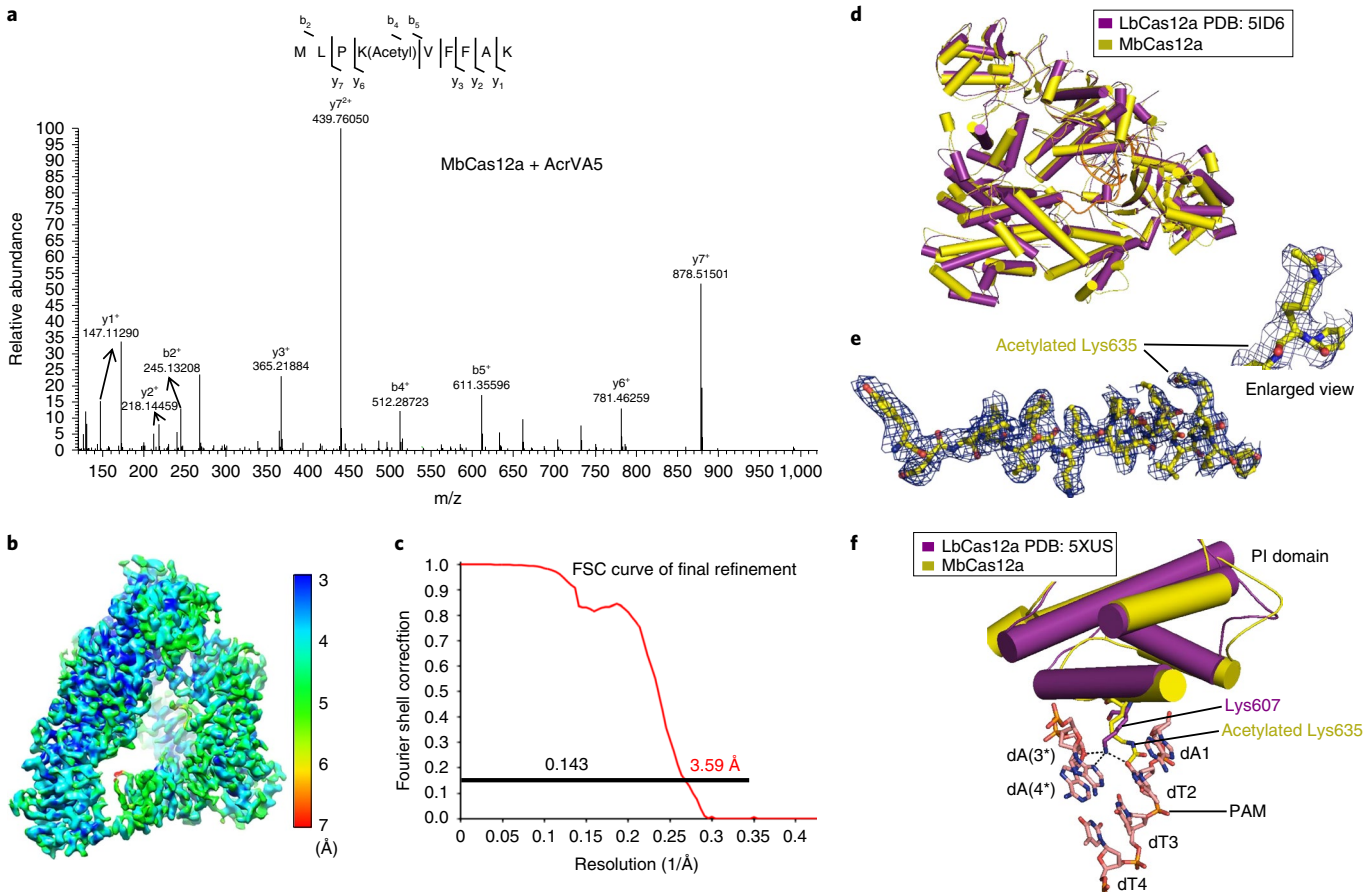

Fig. 3 | Lys635 acetylation by AcrVA5 sterically blocks MbCas12a from PAM recognition. a, The mass spectrum of MbCas12a co-expressed with AcrVA5 in *E. coli*. The mass spectrum charged ion showed that Lys635 is acetylated in the peptide MLPKVFFAK. The b and y ions indicate the fragmentations containing the N terminus and C terminus of the peptide, respectively. **b**, Local resolution heat map of the final density map (MbCas12a in the compact state) of the cryo-EM structure. **c**, FSC curve of the final refined density map of MbCas12a in the compact state. **d**, The overall structural superimposition of the Lys635-acetylated MbCas12a and LbCas12a (PDB ID 5ID6). **e**, Density of the MbCas12a peptide containing the acetylated residue Lys635 in the final reconstruction. **f**, Superimposition of the PAM-interacting domains of the Lys635-acetylated MbCas12a and the PAM-bound LbCas12a (PDB ID 5XUS). PAM bases and PAM-interacting residue Lys607 of LbCas12a are shown as stick diagrams. Detailed hydrogen bond interactions of PAM with Lys607 are shown as dashed lines.

by AcrVA5 at this position, thus making Mb2Cas12a irresponsive to AcrVA5 inhibition. Consistent with this hypothesis, AcrVA5 had no effect on the dsDNA-cleaving activity of Mb2Cas12a (Fig. 4b). As expected, an Mb2Cas12a mutant with Arg625 substituted by lysine displayed dsDNA-cleaving activity similar to that of wild-type protein. Remarkably, however, the activity of this Arg625Lys variant was blocked completely by pre-incubation of the variant protein with AcrVA5 for 20 min (Fig. 4b), similar to the effect observed with MbCas12a. Similar results were also obtained with the Cas12a protein from *L. bacterium COE1* (Lb2Cas12a), which contains the arginine residue at the equivalent position of Arg625 in Mb2Cas12a (Supplementary Fig. 5a). These results indicate that Arg625 is a structural determinant dictating Mb2Cas12a inhibition by AcrVA5. To test whether the equivalent residue of MbCas12a has a similar role in determining AcrVA5 inhibition, we generated an MbCas12a mutant with Lys635 substituted by arginine. The mutant protein cleaved dsDNA substrate, although with slightly lower activity than wild-type MbCas12a (Fig. 4c). However, this mutation rendered MbCas12a completely insensitive to AcrVA5 in cleaving dsDNA substrates (Fig. 4c). Fully consistent with the results observed with Cas12a proteins from Mb strains, AcrVA5 strongly inhibited dsDNA cleavage for wild-type Cas12a from the *L. bacterium ND2006* strain (LbCas12a), which uses Lys595 for specific PAM recognition. Bycontrast, AcrVA5 had no effect on dsDNA cleavage by the Lys595Arg mutant of LbCas12a (Supplementary Fig. 5b). Although Lys635 of MbCas12a is conserved in *Acidaminococcus sp.* Cas12a (AsCas12a), AsCas12a has been shown to be insensitive to AcrVA5 inhibition[20]. This may result from the variable residues surrounding Lys607 of AsCas12a (Fig. 4a).

Discussion

Of the four Mb strains examined (22581, 28389, 33362 and 58069), AcrVA5 occurs in the strains in which the Cas12a proteins use the lysine residue for PAM recognition (the first three strains here) (Fig. 4d). Interestingly, the 58069 strain, which lacks that anti-CRISPR protein, encodes Cas12a that uses arginine for PAM interaction at the same position. These results suggest an evolutionary arms race between phages and bacteria.

In summary, we have identified an anti-CRISPR of the type V CRISPR–Cas12a system and revealed its inhibition mechanism, which has not been previously characterized. Our study also demonstrates that phages have evolved diverse mechanisms to inhibit the CRISPR–Cas adaptive immune systems[18,25,30–33]. Covalent modifications of Cas proteins by anti-CRISPR can permanently inactivate their enzymatic activity and thus probably act as a more efficient strategy with which to inhibit CRISPR–Cas-mediated

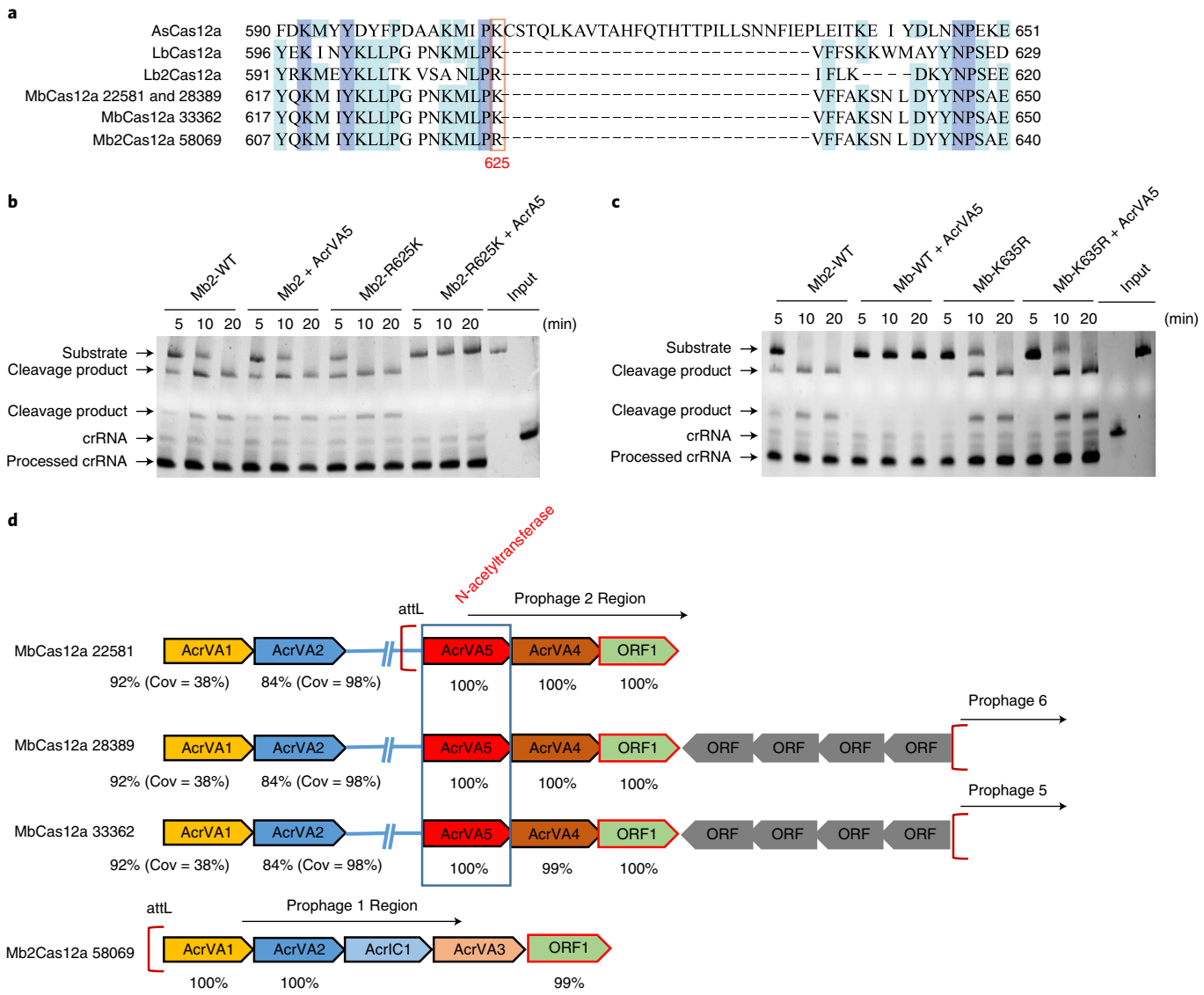

Fig. 4 | Mutation of the critical PAM-recognition lysine residue to arginine results in escape of MbCas12a inhibition by AcrVA5. a, Sequence alignment of AsCas12a, LbCas12a (from strains ND2006 and COE1, respectively) and MbCas12a (from strains 22581, 33362 and 58069) around the PAM-recognition region. Red outline highlights the critical PAM-recognition site 625 in the 58069 MbCas12a. **b**,Time course of cleavage inhibition assays used to test the inhibition activities of AcrVA5 against WT or the Arg625Lys mutant of Mb2Cas12a. Mb2Cas12a from strain 58069 denotes a close MbCas12a ortholog (94% identity). **c**, Time course of cleavage inhibition assays were used to test the inhibition activities of AcrVA5 against WT and Arg635Lys mutant MbCas12a. Uncropped gel images for **b** and **c** are shown in Supplementary Dataset 1. **d**, AcrVA orthologs in four *M. bovoculi* strains (22581, 28389, 33362 and 58069). AcrVA5 is absent in the 58069 genome. Cov, coverage (percent of the query sequence that overlaps the subject sequence); ORF, open reading frame; attL, attachment site of the integrated phage genome.

immunity. Although acetylation of the Lys635 residue of MbCas12a by AcrVA5 blocks target DNA binding, the functions of acetylation on other sites of MbCas12a and potentially other proteins are worthy of study in the future. We also used AcrVA5-Y, Cas12a crRNA and acetyl-CoA at a 0.1:1:100 ratio in the Cas12a inhibition assay and found that 90% of the DNA cleavage activity was inhibited (not shown), indicating that the inhibition activity of AcrVA5 can be promoted substantially when the amount of acetyl-CoA is increased. Thus, our biochemical data suggest that the amount of AcrVA5 and acetyl-CoA could be two layers of determinant for CRISPR-Cas12a deactivation in bacteria. Moreover, it will also be interesting to explore the possibility of whether phages also have anti-CRISPR proteins to target crRNA binding. Although many Cas proteins with high *in vitro* activity have been identified, only a few were shown to be efficient for genome editing[34]. The precise reasons for this remain unclear. But given the existence of many types of enzymes in human or other eukaryotic cells, it is conceivable that a Cas protein introduced into the cells might be covalently modified by them, thus effectively inactivating the potential toxic activity associated with the Cas protein. Future studies that address this possibility will not just contribute to understanding of anti-CRISPR-mediated inhibition of CRISPR–Cas systems but also,and perhaps more importantly, to identifying and engineering more efficient CRISPR-Cas systems for genome editing.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41594-019-0206-1.

Received: 6 December 2018; Accepted: 21 February 2019;
Published online: 1 April 2019

References

1. Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
2. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
3. Westra, E. R. et al. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46**, 311–339 (2012).
4. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237–266 (2013).
5. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
6. Platt, R. J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
7. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
8. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
9. Kleinstiver, B. P. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
10. Tang, X. et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3**, 17103 (2017).
11. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
12. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
13. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
14. He, F. et al. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat Microbiol* **3**, 461–469 (2018).
15. Pawluk, A. et al. Naturally occurring off-switches for CRISPR-Cas9. *Cell* **167**, 1829–1838 e9 (2016).
16. Rauch, B. J. et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **168**, 150–158 e10 (2017).
17. Dong et al. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**, 436–439 (2017).
18. Stanley, S. Y. & Maxwell, K. L. Phage-encoded anti-CRISPR defenses. *Annu. Rev. Genet.* **52**, 445–464 (2018).
19. Gunaratne, R. et al. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. *J. Arthroplasty* **32**, 3854–3860 (2017).
20. Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M. & Doudna, J. A. Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* **362**, 236–239 (2018).
21. Marino, N. D. et al. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science* **362**, 240–242 (2018).
22. Shmakov, S. A. et al. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio* **8**, e01397-17 (2017).
23. Heussler, G. E. & O'Toole, G. A. Friendly fire: biological functions and consequences of chromosomal targeting by CRISPR-cas systems. *J. Bacteriol.* **198**, 1481–1486 (2016).
24. Zhang, F. et al. CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun. Biol.* **1**, 180 (2018).
25. Yang, H. & Patel, D. J. Inhibition mechanism of an anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. *Mol. Cell* **67**, 117–127 e5 (2017).
26. Salah Ud-Din, A. I., Tikhomirova, A. & Roujeinikova, A. Structure and functional diversity of GCN5-related N-acetyltransferases (GNAT). *Int. J. Mol. Sci.* **17,** E1018 (2016).
27. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
28. Magin, R. S., Liszczak, G. P. & Marmorstein, R. The molecular basis for histone H4- and H2A-specific amino-terminal acetylation by NatD. *Structure* **23**, 332–341 (2015).
29. Yamano, T. et al. Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Mol. Cell* **67**, 633–645 e3 (2017).
30. Chowdhury, S. et al. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47–57 e11 (2017).
31. Harrington, L. B. et al. A broad-spectrum inhibitor of CRISPR-Cas9. *Cell* **170**, 1224–1233.e15 (2017).
32. Maxwell, K. L. et al. The solution structure of an anti-CRISPR protein. *Nat .Commun.* **7**, 13134 (2016).
33. Bondy-Denomy, J. et al. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139 (2015).
34. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).

Acknowledgements

We thank J. He at the Shanghai Synchrotron Radiation Facility (SSRF) for help with data collection. We thank the Core Facilities at the School of Life Sciences, Peking University, for assistance with negative-staining EM and the Electron Microscopy Laboratory and the cryo-EM platform of Peking University for help with cryo-EM data collection. The computation was supported by the High-performance Computing Platform of Peking University. We thank X. Meng and H. Deng at the Center of Biomedical Analysis, Tsinghua University, for protein MS analysis. We thank J. Chai for critical reading of the manuscript. This research was funded by the National Natural Science Foundation of China grant no. 31825008 and 31422014 to Z.H.; 31700655 to N.L. N.L. is supported by Young Elite Scientists Sponsorship Program by CAST.

Author contributions

F.Z. and F.Z. conducted bioinformatics analysis. L.D. and X.G. expressed, purified and characterized proteins, and crystallized AcrVA5. Y.Z. and N.L. carried out crystallographic and cryo-EM studies. R.K. and L.Y. performed in vivo experiments. L.D. and X.G. performed biochemical assays. L.D., X.G., Z.H., N.G. and Z.H. wrote the paper. All authors contributed to the manuscript preparation. Z.H. designed the experiments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0206-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Z.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Strains and plasmids. *E. coli* Trans-T1 was used as the host for plasmid construction. Different proteins used in this study were expressed in *E. coli* C43 (DE3) cells. All of the plasmids used in this study are listed in Supplementary Table 1. The cDNAs of full-length MbCas12a, Mb2Cas12a, LbCas12a, Lb2Cas12a and 49 Acr candidates were synthesized and sub-cloned into the bacterial expression vector pGEX-6P-1 (GE Healthcare, with an N-terminal GST tag). The plasmid pBBR1-MbCas12a that was used for C43 (DE3) gene editing via the CRISPR–MbCas12a system was derived from pBBR1-Cas9 (ref. [35]) and constructed by Gibson assembly. The MbCas12a gene was amplified from the plasmid pMbCas12a and driven by the L-arabinose inducible pBAD promoter, while the corresponding crRNA was transcribed from a constitutive promoter listed in Supplementary Table 2. Targeted and non-targeted sequences for pCmr[36] were constructed by ligating annealed primer pairs into the BbsI sites of pBBR1-MbCas12a and pBBR1-Cas9.

Inhibition of MbCas12a cleavage in *E. coli*. The pBBR1-MbCas12a and pCmr plasmids (1 μg of each) were co-transformed individually with the 49 Acr candidates into *E. coli* C43 (DE3) to test the inhibitory effects of these Acr candidates on MbCas12a-mediated dsDNA cleavage. Electro-competent cells were prepared and transformed with MicroPulser (BIO-RAD) following the manufacturer's protocol. The transformation was recovered at 30 °C for 2 h. The cells were spread on LB agar plates with kanamycin (50 μg ml^{-1}), ampicillin (100 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}), and incubated at 30 °C for 24 h. A single colony was picked up in 5 ml LB medium with kanamycin (50 μg ml^{-1}) and ampicillin (100 μg ml^{-1}). After overnight culture at 30 °C, 0.3 mM isopropyl-β-D-thiogalactoside (IPTG) was added to induce the expression of candidate proteins at 25 °C. After 4 h of induction, 4 mg ml^{-1} L-arabinose was added to the culture to initiate the editing process for 8 h at 25 °C. Bacterial suspension was collected for measurement of absorbance (as optical density (OD)) at 600 nm. Survival (colony-forming units (CFU)) was enumerated for *E. coli* C43 (DE3) by serial dilution and selection on solid LB medium with chloramphenicol (34 μg ml^{-1}) after adjusting the OD value to 0.6.

Protein expression and purification. The proteins were expressed in *E. coli* C43 (DE3) cells. Expression of the recombinant protein was induced by 0.3 mM IPTG at 20 °C. After overnight induction, the cells were collected by centrifugation, Cas12a proteins were resuspended in buffer A (25 mM Tris-HCl, pH 8.0, 1 M NaCl, 3 mM dithiothreitol (DTT)) supplemented with 1 mM protease-inhibitor PMSF (phenyl methane sulphonyl fluoride, Sigma). The cells were subjected to lysis by sonication and cell debris was removed by centrifugation at 23,708 *g* for 40 min at 4 °C. The lysate was first purified using glutathione sepharose 4B (GS4B) beads (GE Healthcare). The beads were washed and the bound proteins were cleaved by precision protease in buffer B (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 3 mM DTT) overnight at 4 °C to remove the GST tag. The cleaved Cas12a proteins were eluted from GS4B resin, and further fractionated using a heparin sepharose column and ion exchange chromatography via fast protein liquid chromatography (FPLC) (AKTA Pure, GE Healthcare). AcrVA4 and AcrVA5 proteins were resuspended in buffer B, and purified as described above. Further fractionation was carried out with ion exchange chromatography.

Crystallization, data collection, structure determination and refinement. Crystals of AcrVA5 were generated by mixing the protein with an equal amount of well solution (2 μl) by the hanging-drop vapour-diffusion method. Crystals grew to their maximum size in 8 days in the solution containing 0.1 M NaAC, pH 5.1, 2 M $(NH_4)_2SO_4$ at 20 °C. Before data collection, the crystals were transferred into the cryo-protectant buffer containing the crystallization buffer plus 25% (w/v) glycerol, and flash-frozen in liquid nitrogen.

Diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) at beam line BL17U1 using a DECTRIS EIGER 16 M detector at 100 K. All frames were collected at a wavelength of 0.9792 Å. The crystals belonged to space group $P4_132$ with two subunits per asymmetric unit. The data were processed using HKL2000 (ref. [37]). Initial phases were obtained with a selenomethionine (SeMet)-crystal diffracting to 2.75 Å by the Se-SAD method using AutoSol[38]. The phases were then extended to the 2.05 Å dataset collected from a native crystal. The electron density calculated to 2.05 Å was sufficient for model building with the program COOT[39]. The built model was refined using PHENIX[40]. The Ramachandran plot showed a 98.93% favoured region and 1.07% allowed region. The structure figures were prepared using PYMOL.

In vitro transcription and purification of crRNA. The crRNAs were transcribed in vitro using T7 polymerase and purified using corresponding concentration denaturing polyacrylamide gel electrophoresis (PAGE). Transcription template (dsDNA) for crRNA was generated by PCR. Buffer containing 0.1 M HEPES-K pH 7.9, 12 mM $MgCl_2$, 30 mM DTT, 2 mM spermidine, 2 mM of each type of NTP, 80 μg ml^{-1} home-made T7 polymerase and 500 nM transcription template was used for transcription reactions. The reactions were conducted at 37 °C for 2–6 h and stopped by freezing at −80 °C for 1 h. Pyrophosphate precipitated with Mg^{2+} at 4 °C, and DNA templates precipitated with spermidine. After the removal of the precipitate, RNAs were precipitated using ethanol. The RNA-containing pellets were then resuspended and further purified by gel electrophoresis on a denaturing (8 M urea) polyacrylamide gel. RNA bands were excised from the gel and recovered with the Elutrap System followed by ethanol precipitation. RNAs were resuspended in diethyl pyrocarbonate H_2O and stored at −80 °C.

In vitro cleavage assay. In vitro dsDNA cleavage reactions were performed in a 20 μl buffer system containing 2 μg MbCas12a, 0.3 μg crRNA and 0.3 μg dsDNA. Target DNA sequence containing a protospacer target sequence and a 5′-TTC-3′ PAM motif was cloned into the pUC18 vector. To test AcrVA-mediated inhibition of dsDNA cleavage by MbCas12a, the purified MbCas12a protein was first incubated with AcrVA4 or AcrVA5 protein in cleavage buffer (50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 100 mM NaCl, 5 mM DTT) at room temperature for 10 min. crRNA was then added to the mixture and incubated at room temperature for 5 min. The molar ratio of MbCas12a:AcrVA used for the assay ranged from 1:0 to 1:8. dsDNA (300 ng) was added to the reaction mixture. Cleavage reactions were conducted at 37 °C for 20 min, or as long as required, in cleavage buffer (50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 100 mM NaCl, 5 mM DTT). The reactions were stopped by adding 2×TBE-urea gel loading buffer and quenched 95 °C for 2 min. Cleavage products were separated on TBE-urea 8% PAGE and visualized by EB staining.

GST pull-down assay. Purified GST-MbCas12a protein was incubated with purified Acr proteins and crRNA with a molar ratio of 1:8:2 at 4 °C for 15 min. GS4B resin (40 μl) was added to each reaction system and incubated at 4 °C for 10 min. After washing three times with buffer B (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 3 mM DTT), the reactions were monitored using sodium dodecyl sulfate (SDS)-PAGE and visualized by Coomassie blue staining. The experiment was repeated three times.

Gel filtration assay. MbCas12a and LbCas12a proteins were purified as described above. MbCas12a or LbCas12a, crRNA and AcrVA5, with a molar ratio of 1:2:2, were incubated at 4 °C for 1 h in buffer B supplemented with 2 mM $MgCl_2$. The samples were applied to a size-exclusion chromatography column (Superdex-200 increase 10/300 GL, GE Healthcare) equilibrated with buffer C (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 3 mM DTT). The assays were performed with a flow rate of 0.5 ml min^{-1} and an injection volume of 1 ml for each run. Samples taken from relevant fractions were applied to SDS-PAGE and visualized by Coomassie blue staining.

Electrophoretic mobility shift assay. EMSAs were carried out using the catalytically inactive MbCas12a protein (D874A/E968A). Target DNA strand with 5′-end labelling was purchased from GENEWIZ. Target and non-target DNA strands were hybridized with a molar ratio of 1.5:1 after the probe labelling reaction. Reactions were performed in a 20 μl buffer system containing 120 ng Cas12a protein, 40 ng crRNA and 5 ng dsDNA, with the molar ratio of dCas12a:AcrVA ranging from 1:0 to 1:8. All binding reactions were conducted at room temperature for 30 min in the buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 3 mM DTT and 2 mM $MgCl_2$. Products of the reaction were separated using 6% native polyacrylamide gels and visualized by fluorescence imaging.

Sample preparation and mass spectrometry. Gel bands of proteins were excised for in-gel digestion, and proteins were identified by mass spectrometry. In brief, proteins were disulfide reduced with 25 mM DTT and alkylated with 55 mM iodoacetamide. In-gel digestion was performed using sequencing grade modified trypsin in 50 mM ammonium bicarbonate at 37 °C overnight. The peptides from in-gel digestions were extracted twice with 1% trifluoroacetic acid in 50% acetonitrile aqueous solution for 30 min. The peptide-containing extracts were then centrifuged in a SpeedVac to concentrate the samples.

For liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis, peptides were separated by a 60 min gradient elution at a flow rate 0.300 μl min^{-1} with the EASY-nLC 1000 system (Thermo Scientific), which was directly interfaced with the Thermo Orbitrap Fusion mass spectrometer. The analytical column was a homemade fused silica capillary column (75 μm ID, 150 mm length) packed with C-18 resin (300 A, 5 μm). Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid. The Orbitrap Fusion mass spectrometer was operated in the data-dependent acquisition mode using Xcalibur3.0 software and there was a single full-scan mass spectrum in the Orbitrap (350–1,550 m/z, 120,000 resolution) followed by 3 s data-dependent MS/MS scans in an ion-routing multipole at 30% normalized collision energy (HCD). The MS/MS spectra from each LC-MS/MS run were searched against the selected database using the Proteome Discovery searching algorithm (version 1.4). All MS/MS spectra corresponding to acetylated peptides were examined manually.

Electron microscopy. Negative-staining EM was used to examine the quality of MbCas12a-crRNA samples from multiple biochemical preparations. Aliquots of samples (4-μl) were added to the glow-discharged copper grids, washed and stained with 2% uranyl acetate. The grids with stained samples were checked using a Tecnai T20 electron microscope (FEI) operated at an acceleration voltage of 120 kV. The concentration of samples for cryo-grid preparation were adjusted based on the results of negative-staining EM (~20 times higher). Aliquots of samples (4-μl) were applied to the glow-discharged holy-carbon gold grids (Quantifoil, R1.2/1.3,

400 mesh), blotted and quickly frozen using a Vitrobot Mark IV (Thermo Fisher Scientific), under conditions of constant temperature (4 °C) and humidity (100%). The cryo-grids were screened using a Talos Arctica elcetron microscope (Thermo Fisher Scientific) operated at an acceleration voltage of 200 kV. Selected grids were recovered and transferred into a Titan Krios microscope (Thermo Fisher Scientific) operated at an acceleration voltage of 300 kV for data collection. The images were collected automatically using SerialEM[41] in a movie mode with a nominal magnification of x165,000 and with defocus ranging from 1.5 µm to 3.0 µm. The movies (32 frames each) were recorded using a K2 summit camera equipped with a GIF Quantum energy filter (Gatan) in a super-resolution mode with a dose rate of 10.2 $e^{-}/s/Å^{2}$, and a total exposure time of 8 s. The calibrated pixel size at object scale (super-resolution) was 0.414 Å.

Image processing. Raw movie stacks (4,192) were acquired. The movie stacks were pre-processed with gain-reference correction, drift correction, electron-dose weighting and twofold binning using MotionCor2 (ref. [42]), generating summed images with or without dose weighting. The parameters for contrast transfer function (CTF) of each micrograph were evaluated based on summed images without dose weighting using Gctf[43]. The summed images and CTF power spectra were further screened using SPIDER[44] to discard low-quality micrographs. Approximately 2,000 particles of MbCas12a-crRNA complex were manually picked and processed with two-dimensional (2D) classification using RELION3.0 (ref. [45]), and several 2D class averages with a higher signal-to-noise ratio were selected as templates for further particle auto-picking on all micrographs using RELION3.0. In total, 1,084 K particles were auto-picked from images without dose weighting and subjected to 2D classification to exclude bad and noise particles. Particles with 2D averages reasonably resolved (578) were selected for further processing (Supplementary Fig. 4b). To improve the performance of three-dimensional (3D) classification, the selected particles were first processed with one round of 3D refinement using RELION3.0 with an initial model calculated using CisTEM[46], re-centered on the basis of the refined *x*- and *y*- coordinates, and re-extracted from dose-weighted images. To separate different conformational states of the complex as fully as possible, different 3D classification parameters were tested. Three main conformational states were identified by the first round of 3D classification using RELION3.0, MbCas12a in a compact state (~20% of the particles), MbCas12a in a compact state but with an unstable PAM-interacting domain (~20% of the particles) and MbCas12a in an extended state (~60% of particles). The first group of particles displayed more high-resolution features and was refined to a resolution of 3.72 Å with a global mask imposed. To further improve the resolution of the PAM-interacting domain, which harbours the modified Lys635 residue, the first group of particles was subjected to one round of supervised 3D classification with two references (presence or absence of PAM-interacting domain) applied and the parameter '—skip_alignment' added. Supervised classification resulted in exclusion of a further ~17% of the particles (Supplementary Fig. 4c). A final set of 93 K particles were refined to 3.65 Å, and further improved to 3.59 Å with per-particle local defocus applied using CTF refinement in RELION3.0. The local resolution of the PAM-interacting domain was significantly improved after the supervised 3D classification. The final density map was sharpened by a B-factor auto-evaluated using RELION3.0. All resolution estimations were determined by gold standard FSC at 0.143, with the mask effect corrected. The local resolution heat map of the final density map was calculated using ResMap[47] and displayed using UCSF Chimera[48].

Model building and structure refinement of acetylated MbCas12a. Model building was carried out based on the 3.6 Å reconstruction map. The atomic coordinate of Cas12a from *L. bacterium* (PDB ID 5ID6; LbCas12a) was fitted manually to the density map by CHIMERA[48] to generate a starting model, followed by manual rebuilding using COOT[39]. The crRNA was built de novo using COOT. The model was refined using the phenix.real_space_refine[40] application with secondary structure and geometry restraints. The final models were evaluated by MolProbity[49] and Ramachandran plot[50]. Statistics of the map reconstruction and model refinement are presented in Table 2.

Bioinformatics approach for discovering AcrVA candidates. All available completely sequenced bacterial genomes downloaded from the NCBI RefSeq database were analyzed using our CRISPRminer pipeline (available at http://www.microbiome-bigdata.com/CRISPRminer/) to detect V-A CRISPR–Cas systems. The prediction procedure underlying CRISPRminer is implemented using python scripts and a series of public tools. In brief, the CRISPR arrays were identified using either CRISPRFinder[51] or PILER-CR[52]. Then the Cas gene clusters and their (sub)type associated with the CRISPR arrays were identified using MacSyFinder[51] or HmmScan, respectively. Next, all complete genomes bearing V-A CRISPR–Cas systems were searched for prophage regions using the online webserver provided by PHASTER[53]. A spacer-prophage targeting network was built based on a comprehensive targeting analysis between the spacers from V-A CRISPR arrays and the prophage regions on these genomes using blastn (E-value < 0.01 and with less than three mismatches).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The atomic coordinates and structure factors of AcrVA5 have been deposited in the Protein Data Bank under accession code 6IUF. The atomic coordinates of acetylated MbCas12a have been deposited in the Protein Data Bank under accession code 6IV6. The corresponding maps have been deposited in the Electron Microscopy Data Bank under accession code EMD-9742. The data sets generated and analyzed are available from the corresponding authors on request.

References

35. Xiong, B. et al. Genome editing of Ralstonia eutropha using an electroporation-based CRISPR-Cas9 technique. *Biotechnol. Biofuels* **11**, 172 (2018).
36. Feng, X., Zhao, D., Zhang, X., Ding, X. & Bi, C. CRISPR/Cas9 assisted multiplex genome editing technique in *Escherichia coli*. *Biotechnol. J.* **13**, e1700604 (2018).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. McCoy, A. J. et al. Phaser crystallographic software. *J .Appl. Crystallogr.* **40**, 658–674 (2007).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
41. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
42. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
43. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
44. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**, 1941–1974 (2008).
45. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7,** e42166 (2018).
46. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7,** e35383 (2018).
47. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
48. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
50. Hovmoller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallogr. D* **58**, 768–776 (2002).
51. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
52. Edgar, R. C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18 (2007).
53. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).

nature research

Corresponding author(s): Zhiwei Huang

Last updated by author(s): Feb 14, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Omega Lum C v2.1; HKL2000; XDS;ImageQuant LAS 4000;Chemi XR5;
Data analysis	COOT;PYMOL; Microsoft Excel 2016 was typically used for all presented statistical analyses

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6IUF（AcrVA5),6IV6(MbCas12a：EMDB-9742).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined based on prior published data from similar experiments .
Data exclusions	No data was excluded from the study.
Replication	Every experiment reported was done at least three times with consistent results.
Randomization	We report on in vitro experiments where randomization is not required, However, all experiments were done in a 'randomized' fashion in the sense that random culture dishes containing the same cell population.
Blinding	We report on in vitro experiments where blinding is not required.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*

Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*

Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*