










A resurrected ancestor of Cas12a expands target access and substrate recognition for nucleic acid editing and detection

Received: 11 April 2024

Accepted: 2 October 2024

Published online: 31 October 2024

 Check for updates

Ylenia Jabalera ^{1,8}, Igor Tascón ^{2,3,8}, Sara Samperio ¹,
Jorge P. López-Alonso ^{3,4}, Monika Gonzalez-Lopez¹, Ana M. Aransay ^{1,5},
Guillermo Abascal-Palacios ^{2,3}, Chase L. Beisel ^{6,7},
Iban Ubarretxena-Belandia ^{2,3} ✉ & Raul Perez-Jimenez ^{1,2} ✉

The properties of Cas12a nucleases constrict the range of accessible targets and their applications. In this study, we applied ancestral sequence reconstruction (ASR) to a set of Cas12a orthologs from hydrobacteria to reconstruct a common ancestor, ReChb, characterized by near-PAMless targeting and the recognition of diverse nucleic acid activators and collateral substrates. ReChb shares 53% sequence identity with the closest Cas12a ortholog but no longer requires a T-rich PAM and can achieve genome editing in human cells at sites inaccessible to the natural FnCas12a or the engineered and PAM-flexible enAsCas12a. Furthermore, ReChb can be triggered not only by double-stranded DNA but also by single-stranded RNA and DNA targets, leading to non-specific collateral cleavage of all three nucleic acid substrates with similar efficiencies. Finally, tertiary and quaternary structures of ReChb obtained by cryogenic electron microscopy reveal the molecular details underlying its expanded biophysical activities. Overall, ReChb expands the application space of Cas12a nucleases and underscores the potential of ASR for enhancing CRISPR technologies.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their CRISPR-associated (Cas) nucleases confer bacteria and archaea with adaptive immunity against foreign mobile genetic elements^{1–4}. These CRISPR–Cas systems rely on the ability of the nucleases to use a CRISPR RNA (crRNA) as a guide, with the additional requirement of a non-self sequence (that is, protospacer-adjacent motif (PAM) in DNA, protospacer-flanking sequence (PFS) in RNA) flanking the complementary nucleic acid target. Target recognition then enacts a myriad of immune activities, from cutting target nucleic acids to inducing cellular dormancy or cell death through widespread RNA

or DNA degradation. The adeptness of Cas nucleases at RNA-guided targeting of nucleic acids has catapulted their repurposing for genome editing and diagnostics among many applications⁵. In particular, Cas12a nucleases offer a promising alternative to Cas9 due to relying on T-rich PAMs, generating staggered double-strand breaks in their DNA target and using a short crRNA with no need for a transactivating-crRNA (tracrRNA). Also, Cas12a nucleases can process their own crRNA lending to multiplexing and exert fewer off-targets^{6,7}. Finally, activated Cas12a collaterally cleave single-stranded DNA (ssDNA), which has been used for signal amplification as part of CRISPR-based molecular diagnostics⁸.

¹Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Derio, Spain. ²Ikerbasque Foundation for Science, Bilbao, Spain. ³Instituto Biofisika (UPV/EHU, CSIC), University of the Basque Country, Leioa, Spain. ⁴Basque Resource for Electron Microscopy, Leioa, Spain. ⁵CIBERehd, ISCIII, Madrid, Spain. ⁶Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany. ⁷Medical Faculty, University of Würzburg, Würzburg, Germany. ⁸These authors contributed equally: Ylenia Jabalera, Igor Tascón.

✉ e-mail: ivan.ubarrechena@ehu.es; raulpic@cicbiogune.es

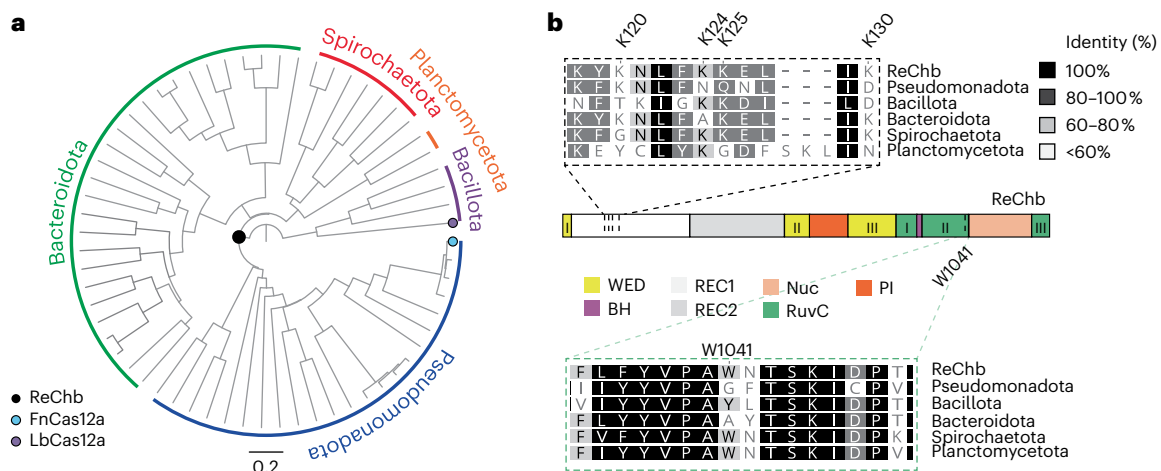


Fig. 1 | ASR of Cas12a yields ReChb. **a**, Phylogenetic tree of sequences used for ASR of ReChb. **b**, The domain architecture of ReChb. Both the WED and RuvC domains are formed by three discontinuous segments of the protein sequence. Top, aligned sequences associated with ReChb and representative Cas12a

nucleases from REC1. Bottom, aligned sequences associated with ReChb and representative Cas12a nucleases from RuvC-II domain. Highlighted sequences indicate conserved positions mutated in ReChb.

Despite its potential, Cas12a is also less efficient than other Cas nucleases. To bridge this gap, engineered versions relying on a few mutations to natural nucleases, such as Cas12a from *Acidaminococcus* sp. (AsCas12a), *Lachnospiraceae bacterium* (LbCas2a) and *Francisella novicida* (FnCas12a), have broadened PAM preferences^{9–12}. Other mutations have been reported to improve editing efficiencies¹³, and efforts have been made to expand substrate recognition, such as the recognition of RNA substrates by the PAM-distal end of the crRNA¹⁴. Nevertheless, even these engineered forms of Cas12a still exhibit properties that, at most, partially deviate from their natural counterparts, displaying similar functionalities.

To overcome these limitations, we adopted an approach based on ancestral sequence reconstruction (ASR). Using ASR, we previously resurrected active Cas9 nucleases that no longer exist in nature¹⁵. These proteins, derived from type II-A Cas9 orthologs from Clostridia and Bacilli classes, retraced the evolution of Cas9 up until modern *Streptococcus pyogenes* Cas9 and displayed PAM-flexible target recognition, gRNA and substrate promiscuity and even collateral cleavage. Although the ancient Cas9 orthologs lost the ability to cleave double-stranded DNA (dsDNA) and exhibited limited genome editing activity as dual nickases, the orthologs suggested that ancestral reconstruction could unlock expanded properties of functions of other Cas nucleases. In the present study, we applied ASR to engineer a fully functional, resurrected Cas nuclease derived from orthologs of Cas12a from hydrobacteria phyla^{8,16}. The resulting resurrected Cas12a ancestor, which we call ReChb, yielded PAM-flexible editing in human cells across three cell lines, seven genes and 16 target sites that outperformed natural and engineered Cas12a nucleases enAsCas12a (ref. 9) and could both target and collaterally cleave RNA, ssDNA and dsDNA. ReChb thus offers a unique and versatile tool for a multitude of CRISPR-based biotechnological applications, from gene therapy to molecular diagnostics.

Results

Applying ASR of Cas12a nucleases to generate ReChb

To reconstruct the sequence of ReChb using ASR, we first collected 63 homologs of Cas12a sequences from species belonging to the phyla Pseudomonadota, Planctomycetota, Spirochaetota and Bacteroidota and belonging to the clade hydrobacteria and Bacillota (as outgroup) (Fig. 1a and Extended Data Fig. 1). A phylogeny using Bayesian inference was constructed following previously described methods¹⁵. The internal node representing a common ancestor of hydrobacteria (excluding the outgroup) was selected for resurrection, resulting in an amino

acid sequence with 53% identity shared with the *Francisella novicida* Cas12a (FnCas12a)¹⁶ and a posterior probability average of 0.92. This ancestor is estimated to have diverged around 3,000 million years ago (Mya) (confidence interval (CI): 2978.5–3030.4 Mya (ref. 17)). The alignment of the reconstructed ReChb against representative sequences from the other phyla shows substantial differences in the REC1 domain (Fig. 1b) with a K-rich region unique to ReChb. In addition, of note is the presence of a tryptophan (W1041) in a highly conserved region of the RuvC-II domain (Fig. 1b), which is conserved in uncharacterized Cas12a sequences from Planctomycetota and Spirochaetota but is absent in the other orthologs (Supplementary Fig. 1). The reconstructed sequence was synthesized and expressed in *Escherichia coli* with higher expression yields than FnCas12a (Supplementary Fig. 2).

ReChb efficiently edits human cells

FnCas12a, the query sequence for ReChb, has been shown to display variable genome editing activity in human cells^{6,16,18}, whereas the ancestral Cas9 nucleases did not exhibit any measurable editing activity when targeting a single site. To evaluate the editing activity of ReChb, we quantified the formation of insertions or deletions (indels) via the generation of dsDNA breaks in HEK293 cells using crRNA guides from FnCas12a targeting sites flanked by a canonical 5'-TTTV-3' PAM in the genes *DNMT1* (5'-TTTC-3'), *AAVSI* (5'-TTTG-3') and *EMX1* (5'-TTTG-3'). After transiently transfecting ReChb and crRNA expression constructs, we employed a T7 endonuclease mismatch assay to quantify indel formation. Compared to FnCas12a, which exhibited detectable indel formation (1%) only at *AAVSI*, ReChb displayed robust indel formation (15–35%) against the target in each of the three genes (Extended Data Fig. 2a,b). Considering that only a limited number of type II and type V orthologs have been successfully employed for genome editing in mammalian cells, these results highlight the potential of ReChb as a gene editing tool.

ReChb exhibits robust PAM-flexible editing in human cells

Given the PAM flexibility of the ancestral version of SpCas9 (ref. 15), we next explored the PAM requirements for ReChb. We applied an in vitro cleavage assay in which purified ReChb loaded with an in vitro transcribed crRNA complex was incubated with a linear DNA target flanked by seven random nucleotides. The PAM-containing cut DNA fragment was then gel purified and submitted for amplicon sequencing (Fig. 2a). The resulting normalized read counts revealed a modest preference for C/T at the -2 and -3 positions (Fig. 2b,c). However, all sequences

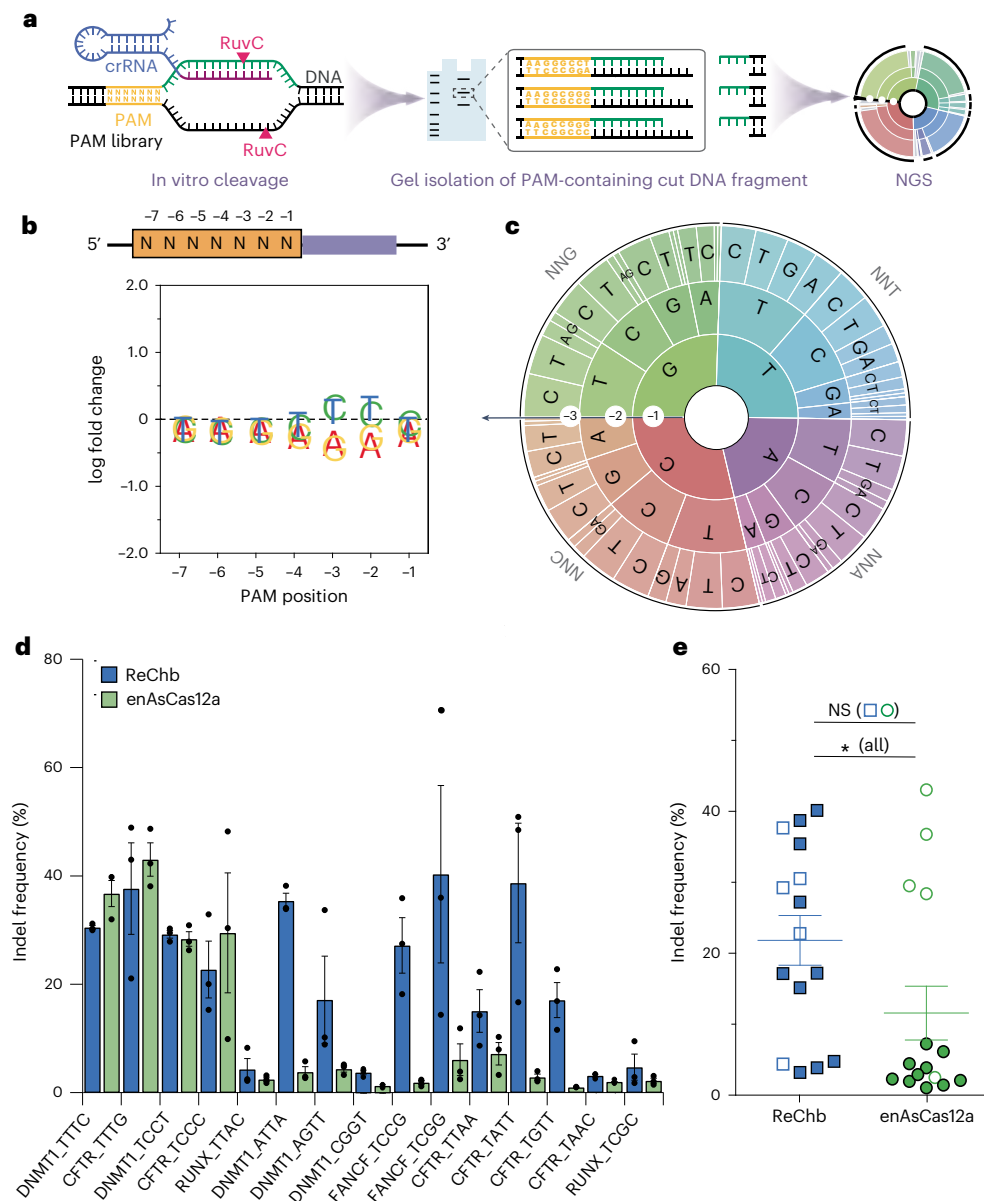


Fig. 2 | ReChb exhibits broad PAM recognition activity in human cells.

a, Scheme of in vitro determination of PAM preference using a substrate library encoding random 7-nt PAMs with next-generation sequencing (NGS). **b**, **c**, 7-nt WebLogo (**b**) and 3-nt PAM wheels (Krona plots) (**c**) recognized by ReChb. **d**, On-target activities of ReChb and enAsCas12a across sites encoding canonical and non-canonical PAM sequences, quantified using TIDE¹⁹. Error bars represent the mean \pm s.e.m. ($n = 3$ independent experiments). **e**, Assessment of the

editing activities of ReChb and enAsCas12a on-target sites harboring canonical and non-canonical PAM sequences. Error bars represent the mean \pm s.e.m. ($n = 15$ target sites). Not significant (NS) differences: ^{NS} $P > 0.05$ (^{NS} $P = 0.1670$); * $P < 0.05$ ($*P = 0.0185$) (two-tailed Welch's t -tests). NS differences for those target sites with PAMs recognized by enAsCas12a (that is, 5'-TTTC-3', 5'-TTTG-3', 5'-TCCT-3', 5'-TCCC-3' and 5'-TTAC-3') are indicated as empty symbols.

were identified in the gel-extracted DNA (Source Data 1), suggesting that the resurrected nuclease has PAM biases but no obvious requirements. This PAM-flexible activity was verified in vitro using both the non-T-rich PAM sequences 5'-ATTG-3', 5'-ATTT-3', 5'-ACCC-3', 5'-ATCA-3' and 5'-ATGG-3' on a supercoiled plasmid as well as 5'-TAAA-3', 5'-TCCC-3' and 5'-TGGG-3' on a FAM-labeled dsDNA substrate (Supplementary Fig. 3). Given the parallels to the PAM flexibility of the ancestral form of SpCas9 (ref. 15), flexible PAM recognition could be a general feature of resurrected Cas nucleases.

After demonstrating the ability of ReChb to target non-canonical sites in vitro, we evaluated its targeting flexibility in human cells in comparison to the engineered Cas12a variant enAsCas12a (ref. 9). enAsCas12a nuclease was selected because it has been associated with the broadest PAM recognition of Cas12a nucleases to date, with recognition

of canonical and non-canonical PAMs⁹. To provide a more comprehensive analysis, we targeted 15 sites in four human genes (*DNMT1*, *CFTR*, *FANCF* and *RUNX*) and analyzed the frequency of indels by Sanger-based indel detection methodology, which provides precise results for indels with a frequency above 1–2% (ref. 19). These experiments yielded robust editing activities (ranges 30–37%) at two target sites with a canonical PAM (5'-TTTC-3' and 5'-TTTG-3') as well as two target sites with T/C-rich (5'-TCCT-3' and 5'-TCCC-3') PAMs (ranges 22–29%) but also at sites with PAMs not recognized efficiently by enAsCas12a, such as 5'-TCGG-3' (40%), 5'-TATT-3' (38%), 5'-AGTT-3' (17%) or 5'-TGTT-3' (17%) (Fig. 2d). At four sites with a 5'-TTAC-3', 5'-TCGC-3', 5'-CCGT-3' or 5'-TAAC-3' PAM, both nucleases yielded poor indel frequencies (<5%), due either to limited PAM recognition or to other factors, such as poor single guide RNA (sgRNA) performance or site accessibility. Across the other sites, ReChb

efficiently edited those targeted by enAsCas12a and, more importantly, demonstrated robust genome editing at sites not recognized by enAsCas12a (Fig. 2e). These data show that ReChb enables robust editing of sites with non-canonical PAMs, with a tested PAM preference toward 5'-NYYN-3', 5'-NRYN-3' and 5'-NYRN-3' in eukaryotic cells, extending beyond those recognized by enAsCas12a. Additionally, ReChb could drive indel formation when transfected as a ribonucleoprotein (RNP) complex in two other cell lines (HeLa and Hs27 fibroblast cells) (Supplementary Fig. 4). Across the four target sites in four genes (*DNMT1*, *SOD1*, *FANCF* and *CFTR*) representing both canonical and non-canonical PAMs, ReChb showed robust editing between 10% and 60%. Finally, off-target analysis of five on-target sites revealed that the indel frequencies at the predicted/potential off-target sites²⁰ were below 1% (Source Data 2), much lower than those at corresponding on-target sites. Thus, ReChb exhibits PAM-flexible editing in human cells that outperforms enAsCas12a in a head-to-head comparison.

ReChb accepts dsDNA, ssDNA and RNA as well as altered crRNAs

Given the broadened PAM recognition of ReChb as one biochemical feature of this nuclease, we explored whether it can expand beyond dsDNA targets and ssDNA collateral substrates traditionally recognized by Cas12a nucleases. Beginning with target recognition, we found that ReChb cleaved complementary dsDNA (Fig. 3a), ssDNA (Fig. 3b) and ssRNA (Fig. 3c) targets containing the reverse complement of the signature T-rich PAM sequence of canonical Cas12a substrates under *in vitro* conditions¹⁶. Although the dsDNA target encoded within a supercoiled plasmid was cleaved into a linearized form, the ssDNA and RNA targets were completely degraded. In contrast, FnCas12a completely cleaved the dsDNA and ssDNA targets and marginally cleaved the RNA target under the same conditions (Extended Data Fig. 3), in line with previous work⁸. We next turned to collateral substrate cleavage, now with the knowledge that ReChb could be activated by dsDNA, ssDNA or ssRNA. For all activating substrates, ReChb degraded non-target dsDNA (Fig. 3d), ssDNA (Fig. 3e) and ssRNA (Fig. 3f). FnCas12a degraded non-target ssDNA but only in the presence of ssDNA and dsDNA targets (Extended Data Fig. 3). Interestingly, the closely related Cas12a2 (that shares 26% of sequence identity with ReChb) nucleases recognize RNA but not ssDNA targets and collaterally cleave ssRNA, ssDNA and dsDNA^{21,22}, suggesting that the ancestrally resurrected nuclease captures the targeting properties of both nucleases despite Cas12a and Cas12a2 respectively specializing in DNA and ssRNA target recognition and ASR drawing only from the Cas12a branch. Critically, the flexible targeting properties exhibited by ReChb extend beyond those of any individual characterized Cas nuclease found in modern organisms^{8,22–24}.

Based on the observed collateral cleavage activity of ReChb, we assessed ReChb as a tool for molecular diagnostics. Cas12a has been extensively used for the detection of specific dsDNA and ssDNA sequences, with cleavage of ssDNA beacons for signal production²⁵. However, RNA remains a poor substrate for Cas12a, requiring workarounds such as supplying a partial dsDNA target¹⁴. To explore the

potential of ReChb, we incubated a ReChb–crRNA RNP complex with a dsDNA, ssDNA or ssRNA target as well as short ssRNA (Fig. 3h) and ssDNA (Fig. 3i) beacons fusing a fluorophore and quencher (Fig. 3g). All combinations of targets and molecular beacons resulted in increased fluorescence compared to no target. In contrast, an FnCas12a–crRNA RNP yielded increased fluorescence only when combining a dsDNA or ssDNA target with the ssDNA molecular beacon. Therefore, ReChb could expand existing capabilities for CRISPR-based molecular diagnostics with a single nuclease capable of accepting different types of nucleic acid biomarkers and molecular beacons.

Traditional CRISPR diagnostics systems based on collateral cleavage by Cas12a nucleases are restricted by the presence of PAM sequence before the target sequence, compromising the universal recognition of any target. Based on PAM-flexible recognition of ReChb, we envisioned that, by combining the collateral cleavage of ReChb next to its PAM-flexible recognition, we could overcome the main limitation of this technology. To test this hypothesis, we first compared the collateral cleavage of ReChb with those exposed by wild-type LbCas12a (ref. 8), a widely used Cas12a nuclease for molecular diagnostics²⁵, after the recognition of a dsDNA target with a T-rich PAM (Fig. 4a). The activity exhibited by ReChb (0.0045 AU s⁻¹) was lower than that of LbCas12a (0.0051 AU s⁻¹) under equivalent conditions (Fig. 4b,d) at 37 °C. Reducing the temperature from 37 °C to 25 °C (room temperature) (Supplementary Fig. 5) considerably decreased the activity of both nucleases (Fig. 4b). However, at room temperature, ReChb retained its ability to recognize target substrates at low concentrations (2 pM; Fig. 4c) compared to LbCas12a, which showed a higher detection limit (10 nM; Fig. 4c). This is in line with previously reported results, where higher temperatures yielded optimal cleavage activity for Cas12a nucleases^{26,27}. Thus, ReChb exhibits catalytic properties similar to LbCas12a, a leading nuclease for CRISPR-based diagnostics, while also offering PAM flexibility and measurable activity at room temperature well suited for point-of-care molecular diagnostics.

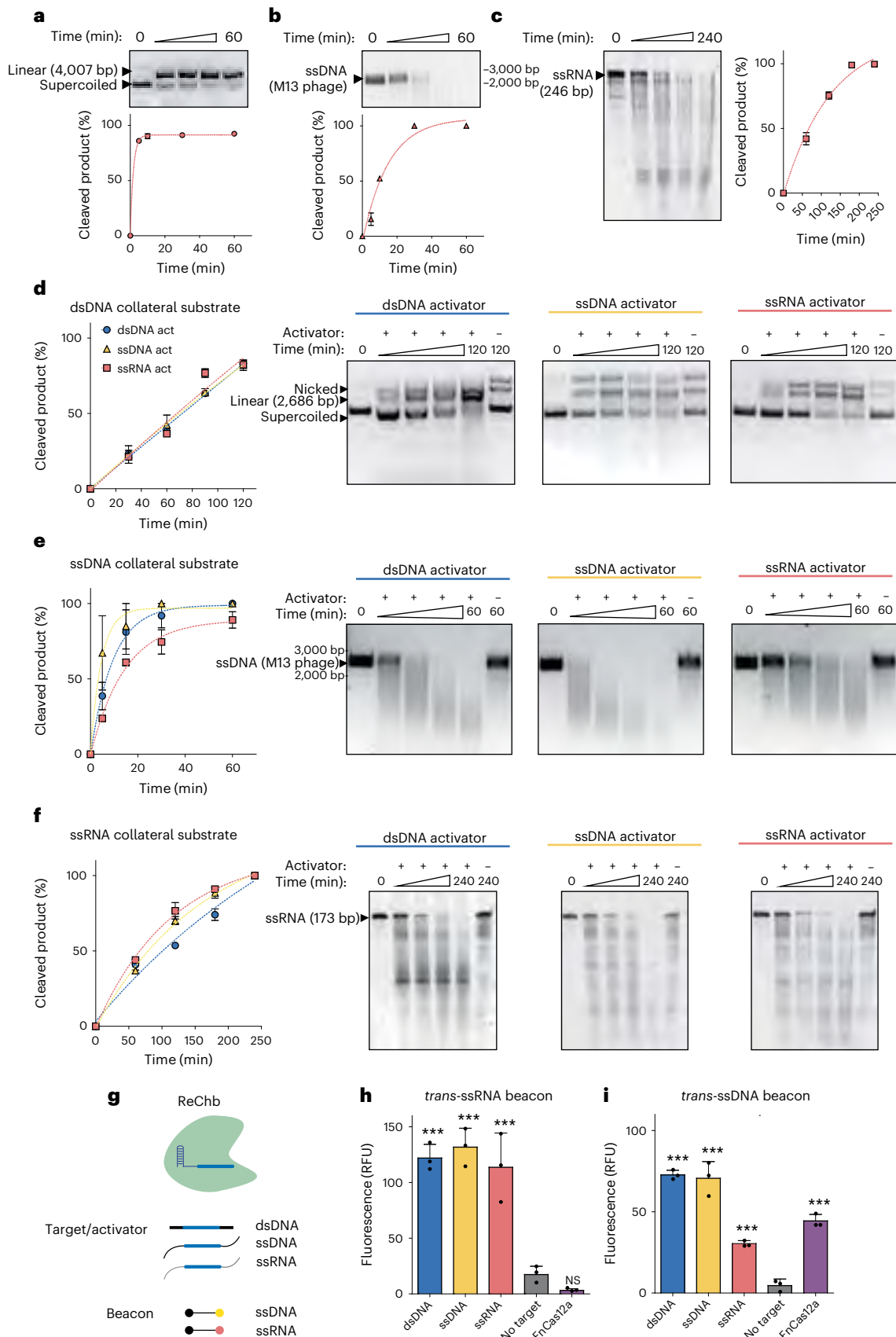
Given the ability of ReChb to recognize non-canonical PAMs with similar activities to that of LbCas12a, we applied ReChb in an application requiring dsDNA recognition and PAM flexibility: detection of a short (21-bp) variable region of a single amplicon of 16S rDNA²⁸ (Fig. 4e). By applying this approach, it is possible to amplify the target region using universal primers that anneal to the conserved region of 16S rDNA and, thus, avoid difficulties associated with multiplexed pre-amplification. To evaluate this approach, we designed four different crRNAs that hybridize to four variable regions (from hypervariable region V1) of bacterial 16S rDNA with a non-canonical PAM (GTCCG). Importantly, this region does not possess a T-rich stretch and, thus, cannot be recognized by traditional Cas12a-based diagnostics. Specifically, we selected four bacterial pathogens, *Acinetobacter baumannii*, *E. coli*, *Klebsiella pneumoniae* and *Staphylococcus aureus*, which are relevant pathogens associated with sepsis in newborns²⁸. Then, we evaluated the collateral cleavage activity for each 16S dsDNA fragment and each crRNA. We found that ReChb triggered fluorescence release only when there was a match between the dsDNA 16S fragment and the

Fig. 3 | Characterization ReChb activity *in vitro*. **a**, *In vitro* double-strand break activity of ReChb against a 4,007-bp supercoiled plasmid DNA as a function of time. Exponential fit of the data is expressed as percent of cleaved product. Assays were repeated three times independently with similar results. **b**, *In vitro* activity of ReChb against M13 phage-derived ssDNA (M13 phage) as a function of time. Exponential fit of the data is expressed as percent of cleaved product. Assays were repeated three times independently with similar results. **c**, *In vitro* activity of ReChb against ssRNA as a function of time. Exponential fit of the data is expressed as percent of cleaved product. ReChb exhibits collateral activity against non-target nucleic acids *in vitro*. Assays were repeated three times independently with similar results. **d–f**, Timecourse of collateral activity against non-target dsDNA (**d**), ssDNA (**e**) and ssRNA (**f**) by a ReChb–crRNA complex activated with target dsDNA, ssDNA and ssRNA substrates. Assays were repeated

three times independently with similar results. See Extended Data Fig. 3 for the same characterization using FnCas12a as nuclease. **g**, Nucleic acid detection assays. Quantification of maximum fluorescence signal generated after incubating ReChb–crRNA activator with a custom ssRNA beacon (**h**) or ssDNA beacon (**i**) for 2 h and 30 min at 37 °C, respectively. For FnCas12a sample, nuclease was target activated with a dsDNA substrate. The negative control was prepared using nuclease-free water instead of the nucleic acid target activator. All data are shown as mean ± s.d. ($n = 3$ independent experiments). ^{NS} $P > 0.05$; ^{***} $P < 0.001$ (two-tailed Welch's *t*-tests). *P* values for ssRNA beacon samples versus negative control: dsDNA (^{***} $P = 0.0002$); ssDNA (^{***} $P = 0.0001$); ssRNA (^{***} $P = 0.0003$); and FnCas12a (^{NS} $P = 0.6704$). *P* values for ssDNA beacon samples versus negative control: dsDNA (^{***} $P = 0.0001$); ssDNA (^{***} $P = 0.0001$); ssRNA (^{***} $P = 0.0004$); and FnCas12a (^{***} $P = 0.0001$). NS, not significant; RFU, relative fluorescence units.

crRNA (Fig. 4f). As expected, LbCas12a was not able to recognize any of these DNA targets (Supplementary Fig. 6). We noticed crRNA from *K. pneumoniae* also triggered fluorescence release when paired with the other three pathogens, likely due to gRNA sequence similarity with

the seed region, as previously reported²⁹. In fact, the mismatch profile analysis of ReChb showed that the protein did not tolerate single or double mismatches at positions 1–20 nucleotides (nt), only triggering a reduced fluorescence signal to single mismatches at positions



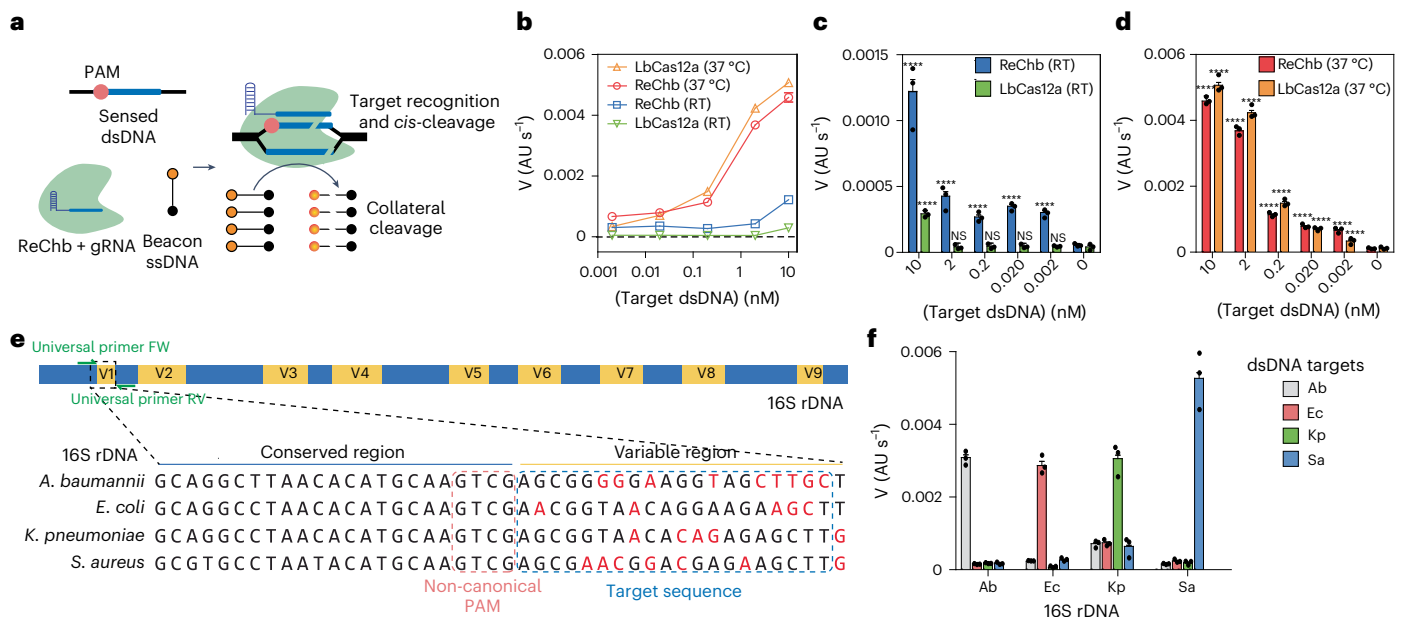


Fig. 4 | ReChb as molecular diagnostic tool for PAM-free dsDNA target recognition. **a**, Schematic of in vitro collateral cleavage assay for dsDNA recognition by ReChb. The ReChb–gRNA complex recognizes and cleaves the dsDNA target, which triggers non-specific collateral cleavage on ssDNA. **b**, Sensitivity of ReChb and LbCas12a using a dsDNA with canonical PAM (5'-TTTC-3') as target substrate and ssDNA beacon as collateral substrate. Velocities were obtained by linear regression analysis of the linear regions of the progress curves. Dots represent the mean ($n = 3$ independent experiments). **c**, Velocities for dsDNA target recognition and posterior collateral cleavage of ssDNA beacon by ReChb and LbCas12a at room temperature. Error bars represent the mean \pm s.d. ($n = 3$ independent experiments). ^{NS} $P > 0.05$; ^{****} $P < 0.0001$ (two-tailed Welch's t -tests). P values for ReChb velocities versus (Target dsDNA) = 0 nM: 10 nM (^{****} $P < 0.0001$); 2 nM (^{****} $P < 0.0001$); 0.2 nM (^{****} $P < 0.0001$); 0.02 nM (^{****} $P < 0.0001$); and 0.002 nM (^{****} $P < 0.0001$). P values for LbCas12a velocities versus (Target dsDNA) = 0 nM: 10 nM (^{****} $P < 0.0001$); 2 nM (^{****} $P < 0.0001$); 0.2 nM (^{****} $P < 0.0001$); 0.02 nM (^{****} $P < 0.0001$); and 0.002 nM (^{****} $P < 0.0001$). **e**, 16S dsDNA from four different pathogens including both conserved and variable regions used for dsDNA target detection with non-canonical PAM sequence (5'-GTTC-3'). FW, forward; RV, reverse. **f**, Specific detection of pathogens based on 16S rDNA with ReChb and each specific crRNA using universal primers for specific amplification of variable region (yellow, V1) using conserved regions of 16S rDNA (blue). Ab, *Acinetobacter baumannii*; Ec, *Escherichia coli*; Kp, *Klebsiella pneumoniae*; Sa, *Staphylococcus aureus*. Error bars represent the mean \pm s.d. ($n = 3$ independent experiments). NS, not significant; RT, room temperature.

collateral cleavage of ssDNA beacon by ReChb and LbCas12a at 37 °C. Error bars represent the mean \pm s.d. ($n = 3$ independent experiments). ^{NS} $P > 0.05$; ^{****} $P < 0.0001$ (two-tailed Welch's t -tests). P values for ReChb velocities versus (Target dsDNA) = 0 nM: 10 nM (^{****} $P < 0.0001$); 2 nM (^{****} $P < 0.0001$); 0.2 nM (^{****} $P < 0.0001$); 0.02 nM (^{****} $P < 0.0001$); and 0.002 nM (^{****} $P < 0.0001$). P values for LbCas12a velocities versus (Target dsDNA) = 0 nM: 10 nM (^{****} $P < 0.0001$); 2 nM (^{****} $P < 0.0001$); 0.2 nM (^{****} $P < 0.0001$); 0.02 nM (^{****} $P < 0.0001$); and 0.002 nM (^{****} $P < 0.0001$). **e**, 16S dsDNA from four different pathogens including both conserved and variable regions used for dsDNA target detection with non-canonical PAM sequence (5'-GTTC-3'). FW, forward; RV, reverse. **f**, Specific detection of pathogens based on 16S rDNA with ReChb and each specific crRNA using universal primers for specific amplification of variable region (yellow, V1) using conserved regions of 16S rDNA (blue). Ab, *Acinetobacter baumannii*; Ec, *Escherichia coli*; Kp, *Klebsiella pneumoniae*; Sa, *Staphylococcus aureus*. Error bars represent the mean \pm s.d. ($n = 3$ independent experiments). NS, not significant; RT, room temperature.

7–20 nt or double mismatches at the end of the PAM-distal region (Supplementary Fig. 7). However, this signal was considerably reduced compared to that displayed by ReChb with a target dsDNA substrate. These results demonstrate the specificity and flexibility of ReChb for molecular diagnostics.

Given ReChb's flexibility in target and collateral substrate recognition, we reasoned that ReChb could also exhibit flexibility in crRNA recognition. To determine the crRNA scaffold recognition determinants of ReChb, in vitro dsDNA cleavage assays were performed by incubating the enzyme with crRNAs from various CRISPR types and Cas12a species (Extended Data Fig. 4a–c). ReChb linearized plasmid DNA using targeting crRNAs from not only different Cas12a nucleases but also from Cas9 and Cas12j. The use of crRNAs with distinct repeats from different types and subtypes is not associated with modern CRISPR–Cas systems and, at most, was observed with the ancestral Cas9 (ref. 15). ReChb could also process its own precursor (pre-)crRNA comprising multiple crRNAs in the same transcript (Extended Data Fig. 4d). The pattern of cleavage products (65 nt) against a single customized pre-crRNA transcript⁷ with four repeats matched with the expected length obtained from cutting the 5' end of a direct repeat hairpins¹⁶. Thus, ReChb can process its own crRNA guides similar to type V effector nucleases^{30,31}, and it can share crRNAs with different types of Cas nucleases. This capability of ReChb should be relevant for multiplexed genome editing applications that employ customized CRISPR arrays for multi-site targeting⁷.

Structure of ReChb ternary complex

To shed light on the mechanism behind the flexibility of PAM and substrate recognition exerted by ReChb, we solved the cryogenic electron microscopy (cryo-EM) structure of the nuclease effector bound to a crRNA and different substrates. The 3.1-Å resolution cryo-EM structure (Protein Data Bank (PDB): 8QWE) captures a ternary complex of ReChb bound to crRNA and target dsDNA containing a T-rich PAM (Fig. 5, Supplementary Figs. 8 and 9 and Extended Data Table 1). Overall, ReChb displays a bilobed architecture resembling the characteristic oval 'sea conch' shape described for Cas12a (ref. 32). The recognition (REC) lobe comprises both REC1 and REC2 domains, and the nuclease (NUC) lobe comprises the PAM-interacting (PI), wedge (WED), RuvC, Nuc and bridge helix (BH) domains (Fig. 5a). The crRNA–target DNA heteroduplex (R-loop; Fig. 5b and Extended Data Fig. 5) threads through the positively charged central channel between the two lobes. The target strand (TS) hybridizes with the crRNA, and the dissociated non-target strand (NTS) traces a straight path through the groove of the nuclease site. There is unambiguous cryo-EM density downstream of the cleaved end of the TS, which we attribute to a post-cleavage product (Fig. 5c,d). As in Cas12a, the nuclease active site of ReChb lies in a pocket at the interface between Nuc and RuvC. The RuvC domain carries the three (D878, E967 and D1216) highly conserved active site residues (Extended Data Fig. 6). In contrast to other Cas12a nucleases^{23,33,34}, the nuclease site of ReChb embraces the dissociated NTS through stacking interactions with W1041 and F971 residues.

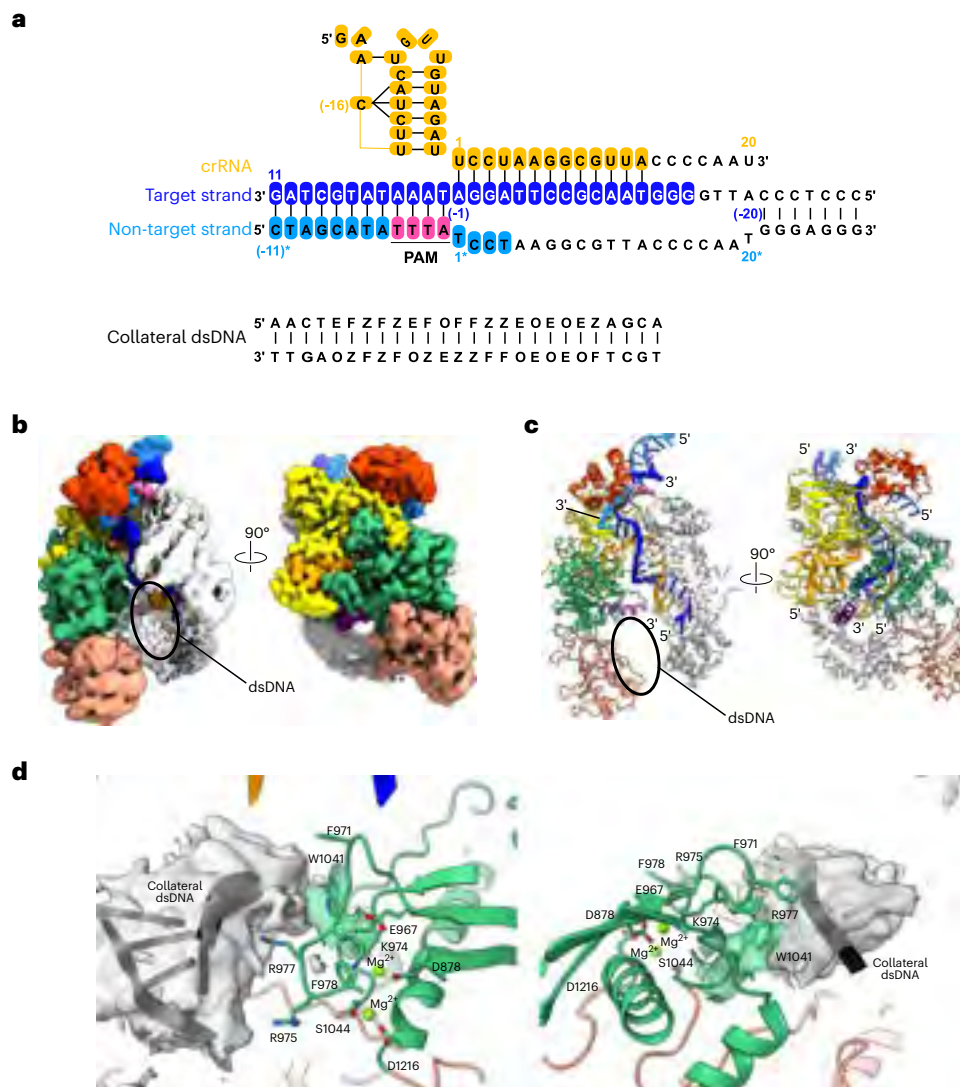


Fig. 6 | Quaternary complex among ReChb, crRNA, target dsDNA and collateral dsDNA. **a**, Sequences of the crRNA guide, target dsDNA and collateral dsDNA used to form the quaternary complex. Nucleotides with colored background are visible in the cryo-EM map, whereas uncolored nucleotides are disordered. The central region of the non-hydrolysable collateral dsDNA includes the following phosphothiates: F, A-phosphorothioate; O, C-phosphorothioate; E, G-phosphorothioate; Z, T-phosphorothioate. **b**, Unsharpened cryo-EM map obtained of the quaternary complex colored by nucleic acid and protein

domains. Unassigned density attributed to part of the collateral dsDNA is highlighted. **c**, Cryo-EM structure of the quaternary complex colored and oriented as in **b**. An oval marks the putative location of the collateral dsDNA. **d**, Close-up views of front (left) and back (right) of the ReChb active site showing the cryo-EM density attributed to the collateral dsDNA fitted with an idealized 4-nt dsDNA fragment. Continuous to the collateral dsDNA density is the density for Trp 1041 shown in transparent green. Selected active site and lid-loop residues are shown as sticks.

Structure of ReChb quaternary complex

The 3.0-Å resolution cryo-EM structure (PDB: 8QWF) (Fig. 6, Supplementary Figs. 8 and 9 and Extended Data Table 1) of a quaternary complex of ReChb bound to crRNA, target dsDNA containing a T-rich PAM and a non-specific collateral dsDNA with both strands containing non-hydrolysable phosphorothioate modifications (Fig. 6a) provides a framework to understand its collateral nuclease activity. As shown for Cas12a nucleases²³, after cleavage, the TS remains hybridized to the crRNA. However, the cryo-EM density for most of the dissociated NTS is absent, indicating its displacement from the ternary complex pathway (Fig. 6b,c), maintaining the ReChb–crRNA complex in the catalytically activated conformation. Specifically, the structure captures RuvC in a pre-hydrolysis state, during which the non-hydrolysable collateral substrate, visible only at low-density thresholds, is attempting to enter its active site (Fig. 6d). This weak density for the collateral substrate is consistent with a constant motion of the substrates, aligning with biochemical data showing dsDNA degradation occurring through

multiple-turnover DNA nicking (Fig. 3d). Therefore, the conformational arrangements of the complex and the PAM-distal target DNA displacement increase accessibility to the RuvC active site, enabling rapid substrate capture and cleavage in *trans* as measured for ReChb (Fig. 3). In fact, with the notable exception of Cas12a2, other natural Cas12 nucleases have difficulties collateral cleaving dsDNA substrates at short incubation times^{21,23,38}. This limitation has been attributed to the RuvC active site not being able to accommodate duplex DNA properly. In addition, the Nuc domain might also act as a physical barrier, limiting cleavage in *trans*²¹. Analogously to Cas12a2 (ref. 21), ReChb can nick, linearize and degrade supercoiled non-specific plasmid DNA. Using FnCas12a as a reference, structures of FnCas12a post-cleavage state (PDB: 5MGA)³² reveal that the presence of the crRNA–TS complex hinders the path to the catalytic site, creating a narrow pathway to reach the RuvC active site, which is incompatible with rigid dsDNA substrates (Extended Data Fig. 8). This contrasts with the accessible ReChb RuvC active site in the post-cleavage state, offering a structural basis for the

wide range of collateral substrates degraded by ReChb (Extended Data Fig. 8). In fact, this mechanism resembles that of Cas12a2, where the absence of Nuc domain and the presence of an exposed RuvC active site afford high substrate accessibility, enabling the cleavage a wide range of substrates in *trans*²¹. Overall, the structural determination together with the biochemical assays highlight the versatility of ReChb for substrate recognition as well as nucleic acid editing.

Discussion

Protein design techniques using computational methods offer an opportunity to improve and even design catalysts with properties not found in natural enzymes³⁹. The increasing number of protein sequences in databases and the advent of methods for sequence alterations and design, including deep learning methods and language models^{40–42}, have the potential to expand the universe of biomolecules and revolutionize the fields of biotechnology and synthetic biology. ASR is rapidly gaining prominence in this context, as it not only provides important evolutionary information but also is able to generate protein sequences not found today^{43–47}. ASR is arguably the only technique that can handle non-natural sequences with many residues (>1,000) and with substantial identity alterations relative to natural proteins.

We previously established that ASR can be used to uncover important aspects of the evolution of CRISPR–Cas systems¹⁵. Here we demonstrate that it is possible to obtain a nuclease with a set of properties that have not, to our knowledge, been found yet in any existing natural Cas nuclease, surpassing the functionality and versatility of any known Cas nuclease. As shown previously¹⁵, ancestral proteins display a functional promiscuity that is present in ReChb perhaps as its most prominent feature. ReChb can recognize any nucleic acid form as a target and is able to carry out specific or non-specific *cis/trans* cleavage. This can be accomplished without the need for the recognition of a strict PAM and with a variety of crRNA guide sequences. ReChb can process its own crRNA and is proficient at genome editing in human cells at a wide range of sites. The promiscuity of ReChb toward nucleic acids could lead to cell toxicity when it is applied as a genome editing tool, although, as observed with other Cas12a nucleases⁴⁸, no signs of cell toxicity were detected during cell experiments in this study. This suggests that these promiscuous collateral activities are likely only observed *in vitro* and do not negatively impact the protein's on-target activity *in vivo*. Although ReChb is capable of collateral cleavage of ssRNA, our hypothesis is that, unlike Cas13, ReChb binds to its DNA target, limiting its ability to freely diffuse and affect surrounding RNA. The cryo-EM structures provide a framework to understand the differences between synthetic ReChb and natural Cas nucleases and provide information for future fine-tuning of the tool.

All of these features have been achieved using ASR, unlike other protein design techniques based on rational design where only one feature is improved at a time—for example, in the effort made to achieve relaxed recognition of PAM using protein engineering techniques, such as that performed for enAsCas12a. In terms of evolution, the different promiscuous features found in the ancestors of Cas9 (ref. 15) and Cas12 reveal that the exact connections between them are unknown and hidden, and surely interconnected mechanisms are yet to be discovered.

ReChb is versatile, as it can be used for both genome editing and diagnostics based on nucleic acid detection. In recent years, assays based on the collateral activity of Cas12a and Cas13 have been developed as diagnostic tools²⁵. Cas12a and Cas13 can be efficiently activated only by DNA and RNA, and, thus, depending on the sequence to be detected, one enzyme or the other normally must be used. ReChb is not limited in this way, as it can be activated by any nucleic acid and is able to identify any type of genetic target. In addition, we demonstrate that the PAM-flexible recognition ability of ReChb allows overcoming one of the main limitations of Cas12-based diagnostics—recognition of dsDNA targets without the PAM restriction.

To our knowledge, ReChb is the only reported nuclease with such expanded activities (Supplementary Table 1). A nuclease with these features might well exist in nature, but searching the natural enzyme space could prove arduous, costly and time-consuming. Similarly, machine-learning-based computational methods are still in their infancy and are not yet capable of designing enzymes with complex and controlled functions, such as large-scale conformational changes. The recently developed deep learning methods (OpenCRISPR-1 (ref. 49)), although promising, have not yet demonstrated the ability to design proteins with new functionalities. These limitations highlight the ability of ASR to generate complex synthetic enzymes with multiple and improved properties and open new avenues for its combination with deep learning and language methods.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-024-02461-3>.

References

1. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
2. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
3. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
4. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19 (2010).
5. Wang, J. Y. & Doudna, J. A. CRISPR technology: a decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
6. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
7. Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
8. Chen, J. S. et al. CRISPR–Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
9. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
10. Toth, E. et al. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic Acids Res.* **46**, 10272–10285 (2018).
11. Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).
12. Tran, M. H. et al. A more efficient CRISPR–Cas12a variant derived from *Lachnospiraceae bacterium* MA2020. *Mol. Ther. Nucleic Acids* **24**, 40–53 (2021).
13. Ma, E. et al. Improved genome editing by an engineered CRISPR–Cas12a. *Nucleic Acids Res.* **50**, 12689–12701 (2022).
14. Rananaware, S. R. et al. Programmable RNA detection with CRISPR–Cas12a. *Nat. Commun.* **14**, 5409 (2023).
15. Alonso-Lerma, B. et al. Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins. *Nat. Microbiol.* **8**, 77–90 (2023).
16. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).

17. Hedges, S. B. & Kumar, S. *The Timetree of Life* (Oxford Univ. Press, 2009).
18. Tu, M. et al. A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res.* **45**, 11295–11304 (2017).
19. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
20. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
21. Bravo, J. P. K. et al. RNA targeting unleashes indiscriminate nuclease activity of CRISPR–Cas12a2. *Nature* **613**, 582–587 (2023).
22. Dmytrenko, O. et al. Cas12a2 elicits abortive infection through RNA-triggered destruction of dsDNA. *Nature* **613**, 588–594 (2023).
23. Swarts, D. C. & Jinek, M. Mechanistic insights into the cis- and trans-acting DNase activities of Cas12a. *Mol. Cell* **73**, 589–600.e4 (2019).
24. Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
25. Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F. & Collins, J. J. CRISPR-based diagnostics. *Nat. Biomed. Eng.* **5**, 643–656 (2021).
26. Huyke, D. A. et al. Enzyme kinetics and detector sensitivity determine limits of detection of amplification-free CRISPR–Cas12 and CRISPR–Cas13 diagnostics. *Anal. Chem.* **94**, 9826–9834 (2022).
27. Nalefski, E. A. et al. Kinetic analysis of Cas12a and Cas13a RNA-guided nucleases for development of improved CRISPR-based diagnostics. *iScience* **24**, 102996 (2021).
28. Srinivasan, R. et al. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS ONE* **10**, e0117617 (2015).
29. Chunlei, J. et al. TracrRNA reprogramming enables direct PAM-independent detection of RNA with diverse DNA-targeting Cas12 nucleases. *Nat. Commun.* **15**, 5909 (2024).
30. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
31. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR–Cas12a. *Mol. Cell* **66**, 221–233.e4 (2017).
32. Stella, S., Alcon, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).
33. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
34. Stella, S. et al. Conformational activation promotes CRISPR–Cas12a catalysis and resetting of the endonuclease activity. *Cell* **175**, 1856–1871.e21 (2018).
35. Nishimasu, H. et al. Structural Basis for the Altered PAM Recognition by Engineered CRISPR–Cpf1. *Mol. Cell* **67**, 139–147.e2 (2017).
36. Strohkendl, I. et al. Cas12a domain flexibility guides R-loop formation and forces RuvC resetting. *Mol. Cell* **84**, 2717–2731.e6 (2024).
37. Worle, E., Newman, A., D'Silva, J., Burgio, G. & Grohmann, D. Allosteric activation of CRISPR–Cas12a requires the concerted movement of the bridge helix and helix 1 of the RuvC II domain. *Nucleic Acids Res.* **50**, 10153–10168 (2022).
38. Murugan, K., Seetharam, A. S., Severin, A. J. & Sashital, D. G. CRISPR–Cas12a has widespread off-target and dsDNA-nicking effects. *J. Biol. Chem.* **295**, 5538–5553 (2020).
39. Hanreich, S., Bonandi, E. & Drienovska, I. Design of artificial enzymes: insights into protein scaffolds. *ChemBioChem* **24**, e202200566 (2023).
40. Yeh, A. H. et al. De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
41. Lovelock, S. L. et al. The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).
42. Madani, A. et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099–1106 (2023).
43. Manteca, A. et al. Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins. *Nat. Struct. Mol. Biol.* **24**, 652–657 (2017).
44. Perez-Jimenez, R. et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* **18**, 592–596 (2011).
45. Zakas, P. M. et al. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* **35**, 35–37 (2017).
46. Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β -lactamases. *J. Am. Chem. Soc.* **135**, 2899–2902 (2013).
47. Risso, V. A. et al. De novo active sites for resurrected Precambrian enzymes. *Nat. Commun.* **8**, 16113 (2017).
48. Marino, N. D., Pinilla-Redondo, R. & Bondy-Denomy, J. CRISPR–Cas12a targeting of ssDNA plays no detectable role in immunity. *Nucleic Acids Res.* **50**, 6414–6422 (2022).
49. Ruffolo, J. A. et al. Design of highly functional genome editors by modeling the universe of CRISPR–Cas sequences. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.04.22.590591> (2024).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Methods

Engineering of ReChb sequence

ReChb was engineered through computational methodologies, specifically ASR¹⁵. We use the BLAST tool with custom parameters and criteria—that is, a maximum of 1,000 hits and minimum identity of 35% to ensure the selection of Cas12a sequences (individually inspected) and BLOSUM62 scoring matrix. E-values were virtually zero for all sequences. Sixty-three sequences were selected following similar proportions of sequences in each phylum as in the database (Extended Data Fig. 1). Sequences belong to five bacterial phyla: Pseudomonadota, Planctomycetota, Spirochaetota, Bacteroidota and Bacillota, as outgroup. Alignment of sequences was performed using MUSCLE software on the MEGA platform and manually edited to eliminate gaps, poorly aligned sites and divergent regions. We inferred the best evolutionary model using MEGA, resulting in the JTT with gamma distribution model (eight categories), Yule model for speciation and length chain of 100 million generations, sampling every 1,000 generations. Phylogeny was carried out using BEAST version 2.6.6 package software (<https://beast.community/>) including the BEAGLE library for parallel processing and based on Bayesian inference using MCMC. Divergence times were estimated using the RelTime method⁵⁰ implemented in MEGA with discrete eight-category gamma distribution for evolutionary rates. We set calibration times using information from the TToL^{17,51} in three major clades with 95% CI. Finally, ASR was performed by maximum likelihood using PAML version 4.9 (<https://evomics.org/resources/software/molecular-evolution-software/paml/>) with a gamma distribution of eight categories for variable replacement rates across sites. Posterior probabilities were calculated for all amino acids, and the residue with highest posterior probability was chosen for each site. The reconstructed sequence displays average posterior probability of 0.92 and amino acid sequence identities of 52% with respect to FnCas12a and 26% with respect to SuCas12a2.

In vitro characterization

Expression and purification of ReChb. The ReChb gene was synthesized and codon optimized for *E. coli* cell expression. ReChb was cloned in pET-28a(+) expression vector and transformed in *E. coli* BL21 (DE3) (Life Technologies) for protein expression. Cells were incubated in LB medium at 30 °C at 160 r.p.m. until optical density at 600 nm (OD₆₀₀) reached 0.6, and IPTG (400 μM) was added for protein induction overnight at 18 °C. Cells were pelleted by centrifugation at 5,000g. Pellets were resuspended in extraction buffer (Tris 50 mM, NaCl 500 mM, pH 8, imidazole 10 mM) supplemented with EDTA-free protease inhibitor (Thermo Fisher Scientific). Then, the pellet was sonicated for three cycles for 10 min at 30% amplitude. Cell debris was separated by ultracentrifugation at 33,000g for 1 h. For purification, the supernatants were mixed with Ni-NTA agarose beads (Thermo Fisher Scientific) and incubated for 1 h. Then, the beads were washed with 50 column volumes with 40 mM imidazole. After that, imidazole concentration was reduced to 10 mM, and an aliquot of protease 3C was added. After 1 h of incubation, protein was eluted in elution buffer (Tris-HCl 50 mM, NaCl 500 mM, pH 8, imidazole 10 mM). The protein was further purified by size exclusion chromatography using a Superose 6 10/300 GL column (GE Healthcare) and eluted in 50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM MgCl₂. For protein purification verification, SDS-PAGE was used with 8% gels. The protein concentration was calculated by measuring the absorbance at 280 nm in a NanoDrop 2000C.

crRNA synthesis

crRNA (Supplementary Table 2) was synthesized using a HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs (NEB)). The DNA sequences include the T7 promoter at the 5' end and the sequence from crRNA with the target sequence at the 3' end. ssDNA oligos were hybridized, and the reaction was incubated overnight. Then, the crRNA was purified following the protocol of the Monarch RNA Purification Columns Kit.

Nucleic acid cleavage assays

For analysis of targeted cleavage (Supplementary Table 3) on supercoiled DNA, 30-μl reactions of 170 nM ReChb-crRNA and 100 nM dsDNA target in 1× NEB 3.1 buffer (50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl₂, 100 μg ml⁻¹ BSA) were incubated at 37 °C for varying incubation times. In the case of ssDNA substrate, 30-μl reactions of 80 nM ReChb-crRNA and 30 nM ssDNA target in 1× NEB 3.1 buffer were incubated at 37 °C for varying incubation times. Reactions were stopped by adding 6× loading dye (NEB) with EDTA and running a 1–2% agarose gel. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad). Cleavage was quantified by ImageJ. Finally, for ssRNA targeted cleavage, 30-μl reactions of 250 nM ReChb-crRNA and 120 nM ssRNA target in 1× NEB 3.1 buffer were incubated at 37 °C for varying times. Reactions were stopped by adding 2× loading dye (NEB) with EDTA and running 10% TBE polyacrylamide gels. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad). Cleavage was quantified by ImageJ.

For analysis of non-targeted (collateral or *trans*) cleavage (Supplementary Tables 4 and 5), 30-μl reactions of 70 nM ReChb-crRNA, 70 nM target substrate (dsDNA, ssDNA or ssRNA) and 120 nM collateral substrate (dsDNA, ssDNA or ssRNA) in 1× NEB 3.1 were incubated at 37 °C for varying incubation times. Reactions were stopped by adding 2× loading dye (NEB) with EDTA or 6× loading dye (NEB) with EDTA and running 10% TBE polyacrylamide gels or a 1–2% agarose gel, respectively. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad). Cleavage was quantified by ImageJ.

PAM library construction

A DNA library (Supplementary Tables 6 and 7) comprising seven random nucleotides was created and subsequently cloned into the plasmid pUC18 by GenScript. This random library was transformed in XL1-Blue *E. coli* and amplified several times to achieve maximal variability in the PAM sequences. Subsequently, an 855-bp polymerase chain reaction (PCR) fragment was generated using the primers detailed in Supplementary Table 7 from the DNA library containing the seven random nucleotides.

PAM determination

PAM determination assay was performed incubating of 170 nM ReChb-crRNA (targeting 23 nt downstream the seven random nucleotides) and 100 nM PCR fragment from the DNA library in 1× NEB 3.1 buffer (50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl₂, 100 μg ml⁻¹ BSA) for 1 h at 37 °C. Reaction was stopped by adding 6× loading dye (NEB) with EDTA and running a 2% agarose gel. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad). The longer fragment, which contained the seven random nucleotides, was purified from the agarose gel with GeneJet Gel Extraction Kit (Thermo Fisher Scientific) and sequenced by Ion Torrent, and the obtained reads were mapped in the reference sequence. Each PAM sequence was quantified, and its frequency was calculated from the total PAM previously extracted. From the frequency of each PAM, PAM wheel was generating, following previously published methodology⁵².

For in vitro cleavage of different PAM sequences, different DNA fragments carrying each PAM (Supplementary Table 8) were cloned into Zero-Blunt TOPO plasmid. The cleavage assay was performed under the conditions described previously. For FAM-labeled dsDNA substrates (Supplementary Table 3), the results were analyzed using a 15% urea-PAGE gel and visualized for fluorescein fluorescence.

Pre-crRNA processing

For pre-crRNA processing, pre-crRNA arrays (Supplementary Tables 9 and 10) were synthesized using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). T7 transcription was performed for 16 h, and then RNA was purified using the Monarch RNA Cleanup Kit (NEB). In vitro cleavage was performed with the purified recombinant protein. ReChb (330 mM)

and in vitro transcribed pre-crRNA arrays (100 nM) were incubated at 37 °C 1× NEB 3.1 buffer for 1 h. Reactions were stopped by adding 2× loading dye (NEB) with EDTA and running 10% TBE polyacrylamide gels. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad).

Nucleic acid detection by ReChb

For nucleic acid detection using fluorescent reporter probes, a 70 nM concentration of the ReChb–crRNA complex was combined with RNase or DNase Alert (250 nM; Integrated DNA Technologies (IDT)) and a 70 nM concentration of the target substrate (dsDNA, ssDNA or ssRNA) for 2 h and 30 min at 37 °C, respectively. The *trans*-cleavage assay was performed in 100-μl reactions, and the resulting endpoint fluorescence was measured using a multi-mode microplate reader (BioTek Instruments), with excitation/emission wavelengths of 490/520 nm for RNase Alert and 500/560 nm for DNase Alert. A negative control was prepared using nuclease-free water instead of the nucleic acid target.

For ReChb comparison with LbCas12a, the wild-type LbCas12a was purified using pMBP-LbCas12a expression plasmid (pMBP-LbCas12a was a gift from Jennifer Doudna (Addgene, plasmid 113431; <http://n2t.net/addgene:113431>; RRID: Addgene_113431)). Protein purification was conducted as previously described⁸, and the protein stocks were diluted into storage buffer (10 mM Tris-HCl, 300 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 500 μg ml⁻¹ BSA, 50% glycerol, pH 7.5) at 1 μM. For in vitro collateral cleavage assay, purified proteins (ReChb and LbCas12a, 100 nM), gRNA (100 nM), target dsDNA (10, 2, 0.2, 0.02 and 0.002 nM, respectively) and ssDNA reporter (DNase alert, 1,000 nM; IDT) were added into 1× NEB r2.1 (Supplementary Table 11). The *trans*-cleavage assay was performed in 5-μl reactions by measuring the fluorescence on a multi-mode microplate reader (BioTek Instruments) using a 96-well V-bottom plate (Thermo Fisher Scientific) with excitation/emission wavelengths 500/560 nm. Timecourses were run for 2 h with an interval of 1 min between reads. Observed rates were obtained by regression analysis of the linear regions of the progress curves. A negative control was prepared using nuclease-free water instead of the nucleic acid target.

For 16S rDNA detection of multiple pathogens related to sepsis, purified proteins (ReChb and LbCas12a, 100 nM), gRNA (100 nM), target dsDNA fragments (10 nm) and ssDNA reporter (DNase alert, 1,000 nM; IDT) were added into 1× NEB r2.1 (Supplementary Table 11).

For one single mismatch and two mismatch profiles of ReChb, purified ReChb (100 nM), gRNA (100 nM), target dsDNA fragments with one or two mismatches across the target site (10 nm; Supplementary Table 12) and ssDNA reporter (DNase alert, 1,000 nM; IDT) were added into 1× NEB r2.1.

In-cell characterization

For endogenous gene editing in human cells, cells were maintained in DMEM (Gibco) supplemented with 10% heat-inactivated FBS (Gibco) in an incubator at 37 °C and 5% CO₂. For plasmid transfection experiments, HEK293T cells were seeded onto 24-well plates (Thermo Fisher Scientific) 24 h before transfection. Cells were transfected using jetPRIME (Polyplus) at 70–80% confluency following the manufacturer's recommended protocol. For each well of a 24-well plate, a total of 500 ng of plasmid DNA (humanized version of ReChb cloned into pcDNA3.1, Addgene plasmid 69976, for FnCas12a (pY004 (pcDNA3.1-hFnCpf1) was a gift from Feng Zhang (Addgene, plasmid 69976; <http://n2t.net/addgene:69976>; RRID: Addgene_69976)) and Addgene plasmid 107941, for enAsCas12a (pCAG-enAsCas12a(E174R/S542R/K548R)-NLS(nuc)-3×HA (AAS848) was a gift from Keith Joung and Benjamin Kleinstiver (Addgene, plasmid 107941; <http://n2t.net/addgene:107941>; RRID: Addgene_107941))) and 300 ng of PCR amplicons comprising a U6 promoter driving expression of the different crRNA scaffolds and target sites were used (Supplementary Tables 13 and 14).

HEK293T cells were transfected with DNA, as described above. Cells were incubated at 37 °C for 48 h after transfection before genomic

DNA extraction. Genomic DNA was extracted using the GenElute mammalian genomic DNA miniprep kit (Merck) following the manufacturer's protocol. The genomic DNA was PCR amplified, and, after validating amplification on a 1% agarose gel, the amplicon was prepared for Sanger sequencing with the primer closest to the expected cleavage site. The chromatograms obtained from each sequencing reaction were analyzed using TIDE analysis¹⁹. As an alternative, editing activity of ReChb was also evaluated using EnGen Mutation Detection Kit (NEB). In brief, the genomic region flanking the target site for each gene was PCR amplified (Supplementary Table 13) using Q5 Hot Start High Fidelity, and PCR products were subjected to a re-annealing process to enable heteroduplex formation. After re-annealing, products were treated with T7 Endonuclease I and analyzed on 2% agarose gels. Gels were dyed with SYBR Gold (Thermo Fisher Scientific) and imaged with ChemiDoc XRS+ System (Bio-Rad). Cleavage was quantified by ImageJ. Quantification was based on relative band intensities. Indel percentage was determined by the following formula:

$$\%Indels = 100x \left(1 - \sqrt{1 - \frac{(b+c)}{(a+b+c)}} \right),$$

where *a* is the integrated intensity of the undigested PCR product band, and *b* and *c* are the integrated intensities of each cleavage product band.

For RNP experiments in HeLa and Hs27 fibroblast cells, RNPs were complexed by mixing 70 pmol of ReChb and 140 pmol of gRNA (Supplementary Table 14) in 50 μl of Opti-MEM at room temperature for 15 min. Then, RNPs were mixed with 4 μl of CRISPRMAX and 2.5 μl of Cas9 Plus reagent (Invitrogen) and then carefully dropped into existing cell culture media for transfection. Cells were incubated at 37 °C for 48 h after transfection before genomic DNA extraction. Genomic DNA was extracted using the GenElute mammalian genomic DNA miniprep kit (Merck) following the manufacturer's protocol. The genomic DNA was PCR amplified, and, after validating amplification on a 1–2% agarose gel, the amplicon was prepared for Sanger sequencing with the primer closest to the expected cleavage site. The chromatograms obtained from each sequencing reaction were analyzed using TIDE analysis¹⁹.

For off-target analysis, predicted off-target sites in the genome using Cas-OFFinder²⁰ were PCR amplified with the primers listed in Source Data 2, and, after validating amplification on a 1–2% agarose gel, the amplicon was prepared for Sanger sequencing with the primer closest to the expected cleavage site. The chromatograms obtained from each sequencing reaction were analyzed using TIDE analysis¹⁹.

Cryo-EM structural determination

Cryo-EM data collection. For all specimens (see Supplementary Note 2 for details on specimen preparation), immediately after adding 0.05% CHAPS, a 4-μl aliquot of the mixture was adsorbed onto a glow-discharged QUANTIFOIL R 1.2/1.3 300 mesh grid (Quantifoil) and vitrified in liquid ethane with a Leica EM GP2 cryoplunger (Leica) using front-side blotting for 2 s at 95% humidity. The vitrified complex specimens were imaged in-house using a 300-kV Krios G4 (Thermo Fisher Scientific) equipped with a BioContinuum/K3 camera (Gatan) operating in counting mode at a calibrated 0.8238 Å per pixel. Employing a 1–1.6-μm underfocus range, we recorded three movies per hole with a total accumulated dose of 50 e⁻/Å² over 50 frames. Movies were recorded automatically using EPU 2 (Thermo Fisher Scientific) with aberration-free image shift (AFIS) and fringe-free imaging (FFI).

Cryo-EM data processing. Initial processing of complex specimen data followed the same general processing strategy (Supplementary Figs. 8 and 10). Initial frame alignment and contrast transfer function (CTF) estimation were performed using cryoSPARC live⁵³, after which processing steps were carried out in cryoSPARC 4.0 (ref. 53) (see Supplementary Figs. 8–10 and Supplementary Note 2 for details on the cryo-EM data processing and three-dimensional reconstruction).

Model building. Initial models were obtained using automated model building with ModelAngelo⁵⁴ and completed by manual intervention in Coot⁵⁵ (see Supplementary Note 2 for details). To improve the fit and to optimize stereochemistry of the atomic model, real-space refinements with secondary structure and geometry restraints were performed using Phenix⁵⁶.

Statistics and reproducibility. Results are representative of least two independent experiments. All replication attempts showed similar results. Analyses were done using GraphPad Prism (GraphPad Software) and Igor Pro 6.32A.

Reporting summary

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

Data availability

Next-generation sequencing data from PAM determination were deposited at the National Center for Biotechnology Information's Sequence Read Archive under BioSample accessions SAMN38227368 (ref. 57) and SAMN38227369 (ref. 58) (BioProject: PRJNA1168665). The EM maps of apo ReChb (accession numbers EMD-18691 and EMD-18692) and its ternary and quaternary complexes (EMD-18693 and EMD-18694) have been deposited in the Electron Microscopy Data Bank (<http://www.ebi.ac.uk/pdbe/emdb/>). The atomic coordinates of ReChb and its ternary and quaternary complexes have been deposited in the Protein Data Bank (<https://www.rcsb.org/>) under PDB ID codes 8QWD (ref. 59), 8QWE (ref. 60) and 8QWF (ref. 61), respectively. Other data supporting the findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

References

50. Tamura, K. et al. Estimating divergence times in large molecular phylogenies. *Proc. Natl Acad. Sci. USA* **109**, 19333–19338 (2012).
51. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timeliness, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
52. Leenay, R. T. et al. Identifying and visualizing functional PAM diversity across CRISPR–Cas systems. *Mol. Cell* **62**, 137–147 (2016).
53. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
54. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
57. MIGS cultured bacterial/archaeal sample from *Escherichia coli*. <https://www.ncbi.nlm.nih.gov/biosample/38227368> (2023).
58. MIGS cultured bacterial/archaeal sample from *Escherichia coli*. <https://www.ncbi.nlm.nih.gov/biosample/38227369> (2023).
59. Lopez-Alonso, J. P., Ubarretxena-Belandia, I. & Tascon, I. Apo ReChb Cas. <https://doi.org/10.2210/pdb8qwd/pdb> (2024).
60. Lopez-Alonso, J. P., Ubarretxena-Belandia, I. & Tascon, I. Ternary complex of ReChb Cas – crRNA – target dsDNA. <https://doi.org/10.2210/pdb8qwe/pdb> (2024).
61. Lopez-Alonso, J. P., Ubarretxena-Belandia, I. & Tascon, I. Quaternary complex of ReChb Cas – crRNA – target dsDNA – collateral dsDNA. <https://doi.org/10.2210/pdb8qwf/pdb> (2024).

Acknowledgements

This work has been supported by grants PID2019-109087RB-I00 and PID2022-141347OB-I00 to R.P.-J.; PID2019-104423GB-I00 and PID2022-143177NB-I00 to I.U.-B.; and PID2021-126263OA-I00 to

G.A.-P., funded by the Spanish Ministry of Science and Innovation (MICIU/AEI/10.13039/501100011033). This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement number 964764 to R.P.-J. and from a European Research Council Consolidator Award (grant agreement number 865973) to C.L.B. The content presented in this paper represents the views of the authors, and the European Commission has no liability with respect to the content. Y.J. acknowledges financial support from Juan de la Cierva program grant FJC2021-047689-I, funded by the Spanish Ministry of Science and Innovation (MICIU/AEI/10.13039/501100011033) and by European Union NextGenerationEU/PRTR. R.P.-J. and Y.J. acknowledge financial support from the Spanish Foundation for the Promotion of Research of Amyotrophic Lateral Sclerosis (FUNDELA). J.P.L.-A. acknowledges financial support from a PTA contract granted by the MCIN/AEI. CIC bioGUNE support was provided from the Department of Industry, Tourism and Trade of the Government of the Autonomous Community of the Basque Country (Elkartek Research Programs 2020–2023), the Innovation Technology Department of the Bizkaia County and MINECO for the Severo Ochoa Center of Excellence Accreditation (CEX2021-001136-S). CIBERehd is funded by the Instituto de Salud Carlos III. High-resolution cryo-EM data collection was performed at the Basque Resource for Electron Microscopy, supported primarily by the Department of Education and the Innovation Fund of the Basque Government, the Fundación Biofísica Bizkaia and the Spanish Ministry of Science and Innovation, through the Plan de Recuperación, Transformación y Resiliencia (PRTR), funded by the NextGenerationEU (PRTR-C17.11) program.

Author contributions

R.P.-J. conceived the project. R.P.-J., Y.J., I.T., I.U.-B. and C.L.B. designed research and planned experiments. Y.J. and R.P.-J. performed the phylogenetic analysis and ancestral sequence reconstruction. Y.J. and S.S. cloned and expressed proteins and performed in vitro experiments and functional validation of ReChb in mammalian and bacterial cells. Y.J., M.G.-L. and A.M.A. performed the sequencing experiments and bioinformatic analysis. I.T., J.P.L.-A., G.A.-P. and I.U.-B. designed, performed, analyzed and represented structural data. All authors participated in discussions and provided ideas for the work. Y.J., I.T., C.L.B., I.U.-B. and R.P.-J. wrote the paper, with additional input from all authors.

Competing interests

R.P.-J. and Y.J. are co-inventors on a patent application filed by CIC bioGUNE. C.L.B. is a co-founder and officer of Leopard Biosciences, a co-founder and member of the scientific advisory board for Locus Biosciences and a member of the scientific advisory board for Benson Hill. All other authors declare no competing interests.

Additional information

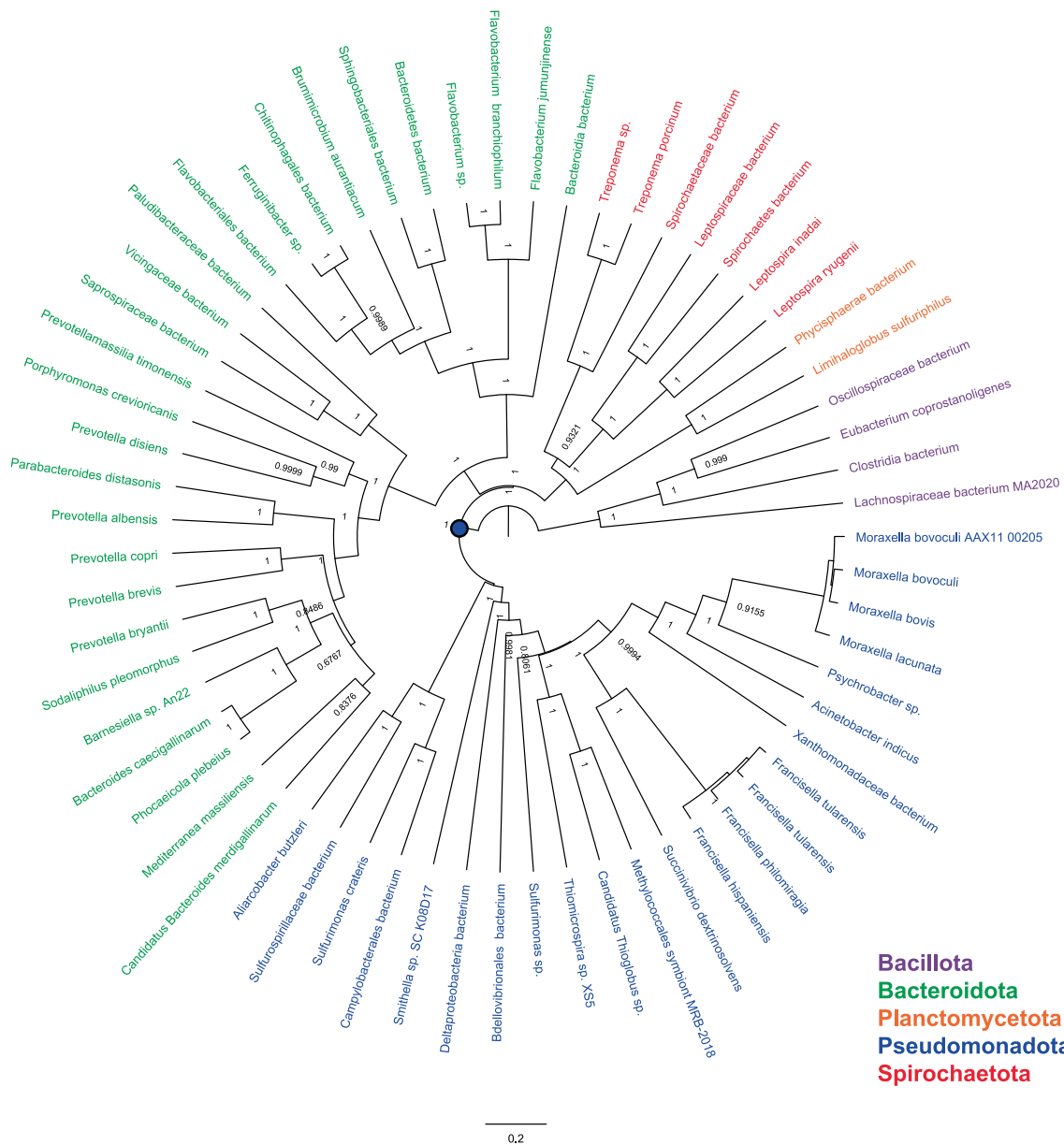
Extended data is available for this paper at <https://doi.org/10.1038/s41587-024-02461-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-024-02461-3>.

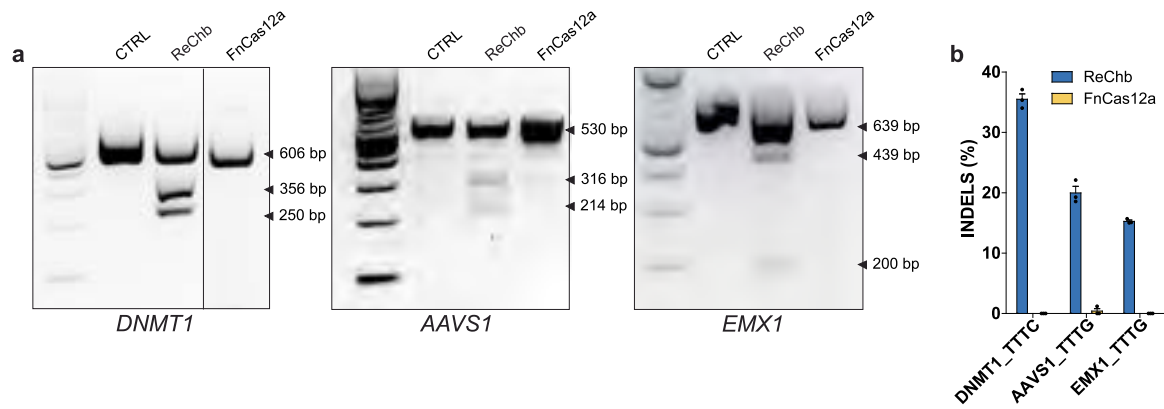
Correspondence and requests for materials should be addressed to Iban Ubarretxena-Belandia or Raul Perez-Jimenez.

Peer review information *Nature Biotechnology* thanks Eric Gaucher and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

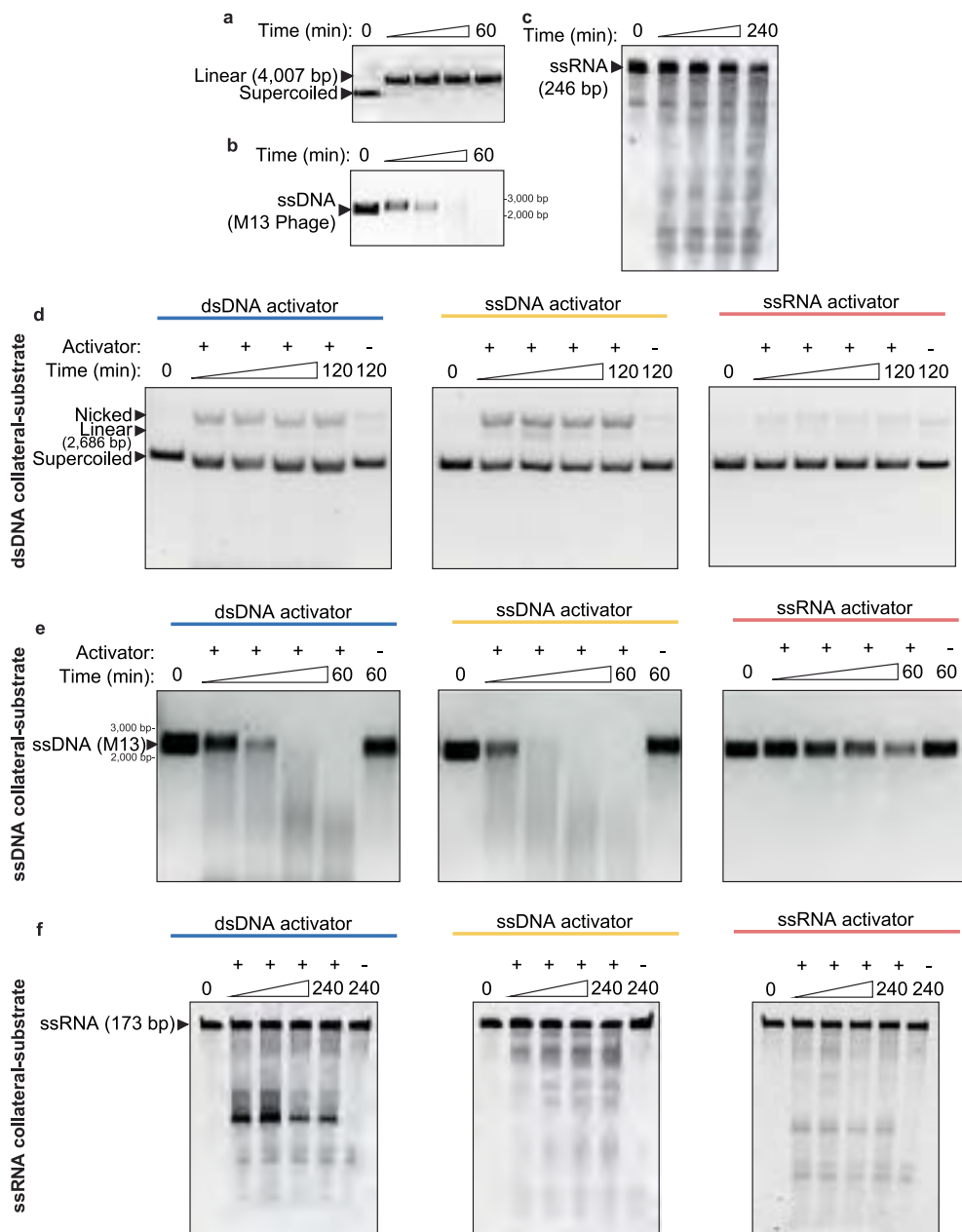
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Phylogenetic tree of sequences used for ASR of ReChb. Posterior probabilities for each node are indicated. A blue circle marks the node selected for ancestral sequence reconstruction.

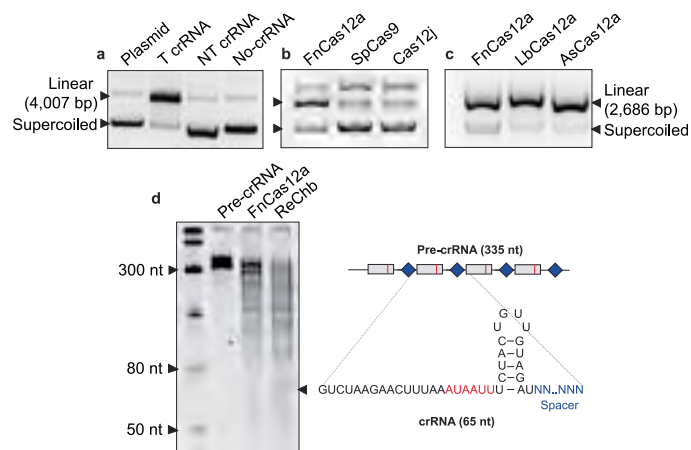


Extended Data Fig. 2 | Editing activity of ReChb in human cells. (a) Editing activity in HEK293T cells. T7 endonuclease mismatch assay for genes *DNMT1*, *AAVS1* and *EMX1* mediated by ReChb and FnCas12a. (b) Quantification of indel efficiency achieved by ReChb at the target sites. Error bars represent the mean \pm SD, (n = 3, independent experiments).



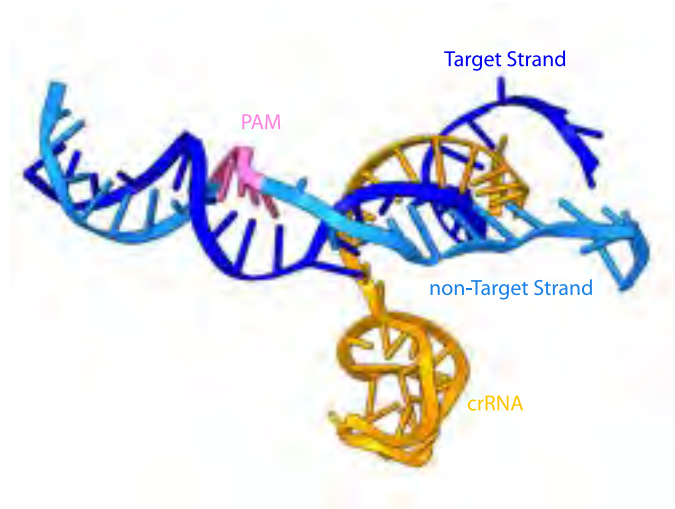
Extended Data Fig. 3 | Activities of FnCas12a ortholog. Cleavage activity of FnCas12a against target (a) dsDNA, (b) ssDNA and (c) ssRNA. Assays were repeated three times independently with similar results. Collateral cleavage

activity against non-target (d) dsDNA, (e) ssDNA and (f) ssRNA by an FnCas12a/crRNA complex activated by target nucleic-acid substrates. Assays were repeated three times independently with similar results.

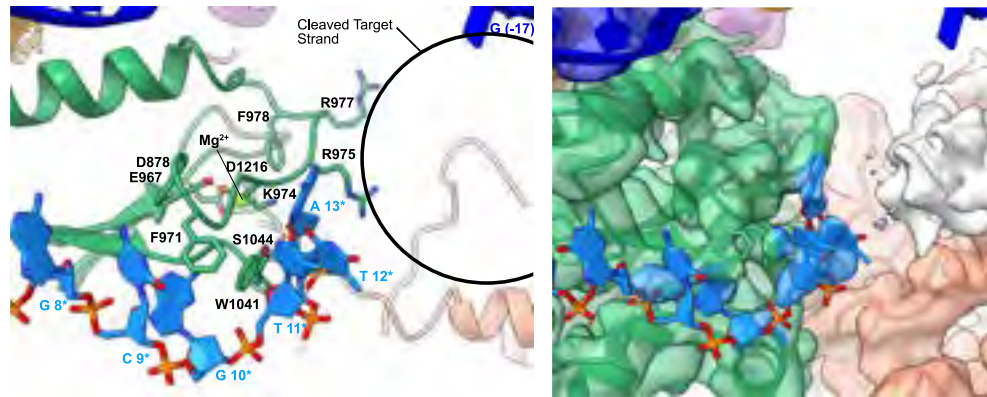


Extended Data Fig. 4 | Flexible use of distinct crRNAs by ReChb while retaining pre-crRNA processing. (a) Target-dependent *in vitro* activity of ReChb using target crRNA (T), non-target crRNA (NT) and no-crRNA. Assays were repeated three times independently with similar results. *In vitro* cleavage

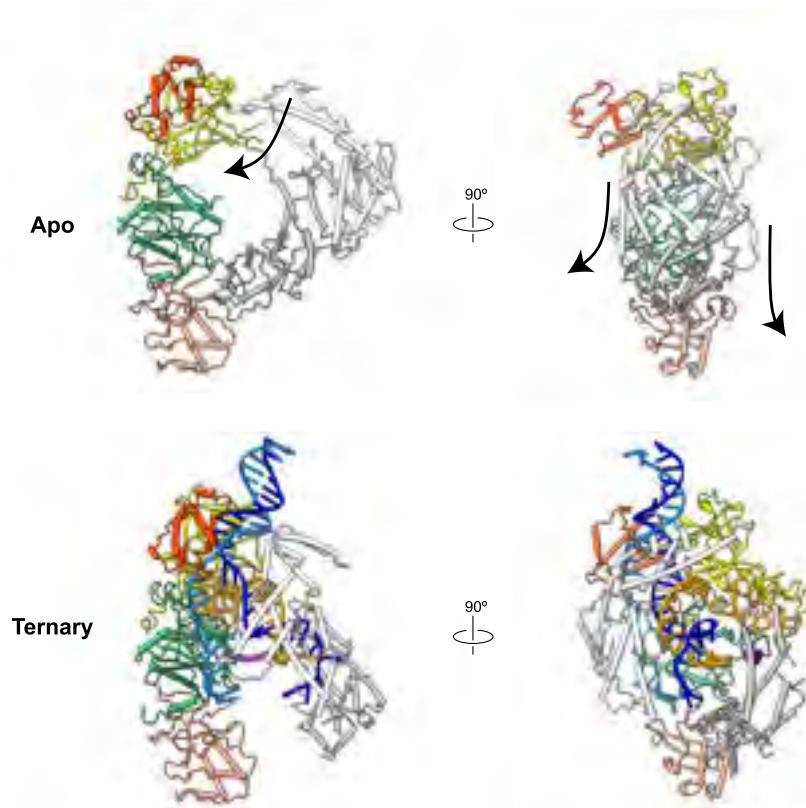
assay against supercoiled dsDNA as a function of crRNAs from different CRISPR (b) types and (c) species. Assays were repeated three times independently with similar results. (d) *In vitro* processing of FnCas12a pre-crRNA (four repeats, 335 nt) transcript with purified FnCas12a and ReChb.



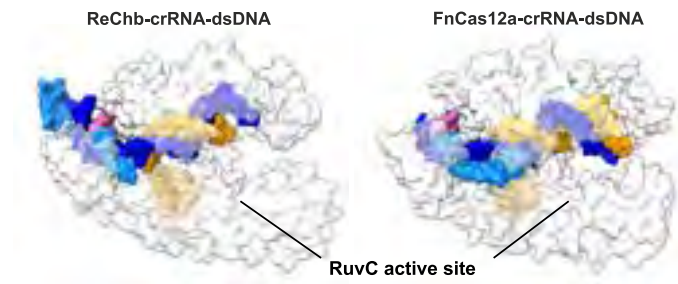
Extended Data Fig. 5 | Details of crRNA-DNA interactions. Structures of the crRNA and target dsDNA in the ternary complex.



Extended Data Fig. 6 | Close-up of the ternary complex active site. On the left, the active site and lid-loop residues are shown as sticks and the non-target DNA strand nucleotides as filled sticks. A circle marks the putative location of the cleaved target strand DNA. On the right, the fitting of the active site region on the cryo-EM density map.



Extended Data Fig. 7 | Cryo-EM structure of apo ReChb. Comparison of apo (top) and ternary complex (bottom) structures of ReChb coloured by nucleic acid and protein domain. Arrows depict the direction of movement of the REC1 and REC2 domains upon RNA/DNA binding.



Extended Data Fig. 8 | Post-cleavage conformation of ReChb (left) and FnCas12a (right; PDB: [5MGA](#)). Nucleotides are coloured as in Fig. 5b.

Extended Data Table 1 | Cryo-EM structure determination

	Apo ReChb (EMD-18691) (PDB 8QWD)	Ternary complex ReChb -crRNA- target dsDNA (EMD-18693) (PDB 8QWE)	Quaternary complex ReChb -crRNA- target dsDNA- collateral dsDNA (EMD-18694) (PDB 8QWF)
Data collection and processing			
Magnification	105000	105000	105000
Voltage (kV)	300	300	300
Electron exposure (e-/Å ²)	50.979	50.979	50.084
Defocus range (µm)	1-1.6	1-1.6	1-1.6
Pixel size (Å)	0.8238	0.8238	0.8238
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	3324256	5463714	7589433
Final particle images (no.)	254316	137913	300509
Map resolution (Å)	3.33	3.14	3.08
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	2.9-6.5	2.9-6.5	2.9-6.5
Refinement			
Initial model used (PDB code)	-	-	-
Model resolution (Å)	3.3	3.1	3.0
FSC threshold	0.143	0.143	0.143
Model resolution range (Å)	2.9-6.5	2.9-6.5	2.9-6.5
Map sharpening <i>B</i> factor (Å ²)	-137.6	-103.2	-124.0
Model composition			
Non-hydrogen atoms	9204	11713	11499
Protein residues	1103	1185	1186
Ligands	2	5	6
<i>B</i> factors (Å ²)			
Protein	127.16	101.79	31.38
Nucleotide	-	93.27	35.04
Ligand	69.59	28.99	21.31
R.m.s. deviations			
Bond lengths (Å)	0.002	0.005	0.002
Bond angles (°)	0.507	0.645	0.499
Validation			
MolProbity score	1.79	1.63	1.69
Clashscore	8.29	6.94	7.59
Poor rotamers (%)	0.20	0.00	0.00
Ramachandran plot			
Favored (%)	95.02	96.33	95.99
Allowed (%)	4.98	3.67	4.01
Disallowed (%)	0.00	0.00	0.00

Cryo-EM data collection, refinement and validation statistics.

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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

Data analysis

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

NGS Sequencing data from PAM determination were deposited at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under BioSample accessions SAMN38227368, SAMN38227369. The EM maps of apo ReChb, accession numbers EMD-18691 and EMD-18692, and its ternary and quaternary complexes, EMD-18693 and EMD-18694 have been deposited in the Electron Microscopy Data Bank (<http://www.ebi.ac.uk/pdbe/emdb/>). The atomic

coordinates of ReChb, and its ternary and quaternary complexes, have been deposited in the Protein Data Bank (www.pdb.org) under PDB ID codes 8QWD, 8QWE and 8QWF, respectively. Other data supporting the findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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	No statistical methods were used to predetermine sample size. At least three independent replicates were performed for in vitro and in cells experiments using plasmids, and two independent replicates for in cell experiment using RNP. Sample sizes were determined based on our previous experiences and what were described in similar experiments in publised papers.
Data exclusions	No data was excluded. For structure determination, junk particles were excluded by image classification in order to identify a subset of particles that led to the highest resolution 3D-reconstruction.
Replication	Information on replication of the experiments is reported in the figure legends. In vitro and in vivo results are representative of least 2 independent experiments. All replication attempts showed similar results.
Randomization	Randomization was not applied in line wih standar experimental procedures in molecular biology.
Blinding	Blinding was not applied in line with standard experimental procedures in molecular biology.

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HEK293T (ATCC CRL-1573); HeLa (ATCC CRM-CCL-2); Hs27 fibroblast (ATCC CRL-1634)
Authentication	None of the cell lines used were authenticated
Mycoplasma contamination	Cells were confirmed to be free of mycoplasma by analyzing supernatant media monthly using MycoAlert™ PLUS (Lonza).
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>