

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

The molecular determinants behind genome editing activity: insights into TnpB

Permalink

<https://escholarship.org/uc/item/94z269m7>

ISBN

9798189280146

Author

Zhou, Zehan

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

The molecular determinants behind genome editing activity: insights into TnpB

By

Zehan Zhou

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer A. Doudna, Chair

Professor Michael P. Rapé

Professor Xavier Darzacq

Professor Kimberley D. Seed

Spring 2026

Abstract

The molecular determinants behind genome editing activity: insights into TnpB
by

Zehan Zhou

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Jennifer A. Doudna, Chair

TnpB is a compact RNA-guided endonuclease and evolutionary ancestor of CRISPR-Cas12 that offers a promising platform for genome engineering. However, the genome-editing activity of TnpBs remains limited and its underlying determinants are poorly understood. Here, we used biochemical and single-molecule assays to examine the DNA-unwinding mechanism of *Yongiibacter multivorans* TnpB (Ymu1 TnpB). DNA unwinding proceeds through a discrete, long-lived partially unwound intermediate state before reaching a fully unwound open state. The open state forms inefficiently and collapses readily in the absence of negative supercoiling. An optimized variant, Ymu1-WFR, stabilizes formation of both the intermediate and open states, resulting in enhanced DNA cleavage *in vitro* and increased genome editing *in vivo*. These findings identify the physical basis for the observed minimal activities of natural TnpBs, revealing how stabilizing specific unwinding states enables efficient DNA targeting.

Table of Contents

Acknowledgements	iii
Chapter 1: Introduction	1
Chapter 2 Rationally engineered TnpB proteins can be purposed for plant genome engineering	
2.1 Introduction	3
2.2 Results	
2.2.1 Ymu1 TnpB is an enzyme that can be utilized for plant genome applications.....	3
2.2.2 Ymu1 TnpB protein variants accelerate DNA cleavage in plants.....	6
2.2.3 In vitro cleavage recapitulates the efficiency observations made in plant protoplasts.....	8
2.2.4 Mutagenesis of Ymu1 TnpB influences its off target editing efficiency.....	9
2.2.5 Differential enrichment of rational engineering-derived mutations across the TnpB family tree.....	11
2.3 Discussion	12
2.4 Limitations of the study	13
2.5. Supplementary Figures	14
2.6 Methods	17
2.7 Supplementary Tables	26
2.8 Declaration of interests	49
Chapter 3: Engineering TnpB proteins enables them to bypass intrinsic barriers in DNA target engagement	
3.1 Introduction	50
3.2 Results	
3.2.1 AuRBT revealed step-specific enhancement of DNA unwinding by Ymu1 mutations.....	51
3.2.2 Negative supercoiling stabilizes unwound states in WT Ymu1.....	53
3.2.3 Stepwise DNA substrate unwinding limits WT Ymu1 cleavage.....	55
3.2.4 Structural analysis of Ymu1 TnpB mutations.....	56
3.3 Discussion	59
3.4 Limitation of the study	60
3.5 Supplementary Figures	61
3.6 Methods	63
3.7 Supplementary Tables	69

3.8 Declaration of interests.....	83
References.....	83

Acknowledgements

It takes a village to raise a child and nowhere does this feel truer than in science. I am truly grateful and humbled by the amazing scientists and friends that I have come to know in my four years at Berkeley MCB- their support has been truly invaluable in my learning experience here, and I truly could not have done this PhD without all their insight, their guidance, and their kindness. I deeply appreciate how much I have learned not only about science, but also about myself in this process, and this is thanks to the amazing people who I have the honour and privilege of surrounding myself with.

First and foremost, I thank my advisor Jennifer Doudna for taking a chance on me and for investing me with her faith and confidence. I originally came to Berkeley as a committed chemical biology student and I had no idea about how CRISPR worked- much less the biochemical and structural techniques and way of thinking that forms the bread and butter of the lab. In my meetings with Jennifer, I have come to see that she embodies the very essence of the growth mindset: she is not afraid to prompt you to re-examine your ideas, to challenge your thinking, and to guide you through the processes of scientific thinking, writing, and presentation. Hearing her tell me about her own journey in science, from how she took time to plan out her morning experiments to her detailing grant applications and scientific narrative construction, I have come to deeply appreciate and admire her approach to science and to life. She is generous with her time and her support, and she is always there for whenever you need her- I remembered that when my experiments weren't going well, she took the initiative to schedule weekly meetings with me, and in all of my group meetings, she stepped up to advocate for my learning experience in challenging me to think about science. Jennifer is the embodiment of the Socratic method of pedagogy, and she serves as a role model for me in the type of scientist I wish to become.

My experience in this project would not be complete without the selflessness and generosity of Honglue Shi. I chose to embark on this project with him since I was interested in seeing how understanding the basic principles behind genome engineering and understanding how RNA behaves in a genome editing context. Honglue not only took the pains to coach me through each of the biochemical techniques, but supported me when the project hit roadblocks, guided me through chatting with collaborators, mentored me in the writing process and in the figure making process, and served as a source of great inspiration and comfort throughout. I am deeply indebted to his mentorship and his care, and I cannot wait to see where life takes him.

I wanted to thank my amazing collaborators at the Bryant lab, Irenka and Zev Bryant himself, and Trevor Weiss and Steve Jacobsen at the Jacobsen lab. Irenka was such an incredibly talented collaborator- her tireless work on the RBT assays proved to be instrumental to our narrative, and working with someone as dedicated but also as happy and joyful as her was such an amazing experience. Zev was an incredible source of wisdom and guidance, both in terms of single molecular work and in terms of scientific thinking and investigation and learning that we went to high school a 10-minute drive apart really was a true treat. The Bryant lab's rigour and dedication to this project was unparalleled, and I could not have collaborated with a better group. Steve Jacobsen was a deeply inspirational person to work with- his scientific acuity and his insight were a boon to our project, and his wicked sense of humour added to the joy of collaborating with him. Lone Jackson, ride on! Trevor's selflessness with data and with his plant biology work

formed the basis of our publication and our science, and he was an amazing source of insight and support throughout all the ups and downs of the publication process. I got to witness a master class in science through Trevor and Steve, as well as talking to Maris and Jasmine in their visits here, and I could not have been more grateful for that. It was amazing talking to Brittney and Rachel about their DMS insights in the formulation of ideas as well, and Dave Savage's group provided a great basis of engineering design understanding. As well, I would like to shout out Petr Skopintsev, whose structural insights and whose generosity with his AI-generated TnpB structure enabled us to think about mechanistic hypotheses in terms of TnpB residues.

I want to also extend a huge thank you to Kaihong, Keana, and Jenny. Kaihong, you made the lab a second home to me. From setting me up with the appropriate reagents and equipment to being a source of inspiration and advice, you made the lab seem like family, where I could always go to get support and a listening ear. I am so grateful for Enbo's infectious joy and enthusiasm in the lab, and experiencing the Doudna lab with the both of you was such an immense honour. Keana, thank you so much for all your help in managing my lab expenses, helping me with my paperwork, but also for being such an amazing friend and supporter in my whole process of navigating science. People like you make the world a brighter and better place. Jenny, thank you so much for helping me with reagents and ordering, as well as being a source of fun conversation. I also wanted to shout out Elaine while she worked in the lab- she was such a friendly presence, and chatting to her about cats and Chicago and scuba diving really added so much joy to my time at the Doudna lab.

Before I get to the great members of the Doudna lab, I'd also like to shout out some of my cohort friends. **BESSICA!!!!** (actually Jessica Nguy)- thank you for being such an incredible friend, from when we visited Berkeley as new recruits to the rotation process, second year quals studying and beyond. Thank you for being my temple buddy, for your perfect punctuation while texting, and for making such delicious food- I have always loved your wine braised ribs. Thank you for also allowing me to drop by and talk about stupid things and to be hyperactive. Thank you, Anna Trzcinski, for being my workout buddy after Peter, for going hiking with me, and for being such a fun friend to drive around with. We really should go freshwater swimming together. I'm sorry I couldn't uphold my end of the bargain in going to live in Switzerland after the PhD, but I really hope that you uphold yours and go live in Japan. Thank you to Cynthia Terrace and Simon Chau for all the fun we had in getting dim sum, rotating sushi, Korean barbecue, New Years celebrations and the like. Thank you to my other cohort friends- Aditya Udupa, Gavin Dury, Christy George (special shoutout to Christy for being a great source of advice in thinking about the physician trajectory), Nate Dempsey, Anna Christenson, Erin Newringesen, Jan Cajulao, Sophia Swartz, and so many more. I really love my cohort.

I also want to thank the great faculty at Berkeley for making this PhD such an incredible experience. I want to begin by thanking the amazing members of my thesis committee for their advice and their guidance throughout the process. Thank you, Micha Rape, for your kindness throughout my academic journey- from your amazing advice in my quals to chatting about how you started off as a professor at Berkeley to your help and guidance at the MTx faculty meetings, I could not be more grateful for your help and your guidance throughout. Thank you, Xavier Darzacq, for being an amazing rotation mentor- I loved my time in the Tjian-Darzacq lab and I learned so much. Thank you, Kim

Seed, for your insights into prokaryotic biology during my thesis committee meetings. Thank you, Dan Nomura, for an amazing rotation in your group, and for being such a wonderful professor to GSI with. And I want to thank Lauren Orr and Halime Yilmaz for a wonderful first rotation, and Jack Ferrie and Jonathan Karr for an outstanding third one. And thank you Roberto Zoncu for your kindness at MTx faculty meetings, for our wonderful experience teaching together, and for being a gracious lab host when I came here. We have some of the best faculty in the world at Berkeley MCB, and what an amazing privilege it is to study here.

Now back to the amazing lab members- thank you so much, David Colognori, for being my rotation mentor and for being that initial friendly presence in the Doudna lab to help me navigate the group and science. It was so great being able to sit next to you when I started, and you will always serve as a source of inspiration for me. Thank you so much, Wayne Ngo, for your confidence and your warm support of me in thinking through projects and in understanding the scientific process. It was so great to get to know you better at the Keystone, and perhaps my greatest regret in my PhD was not having gotten to chat with you sooner. Thank you, Matt Kan, for your advice in thinking about combining clinical practice and science- you are who I aspire to be, and I am so thankful to be sitting next to you and your wonderful team. Anna Vardapetyan, Alanna Yang, Aysel Ogul, and Sophie Wang, it was so cool to sit right by and shoot the breeze with you all, and to lollygag. Thank you, TER-RY Zhang- it's great chatting with you about structural biology, Jollibee, and having general outings and fun with you and Kenny. Thank you, MA-RE-NA Trinidad- you are one of the most brilliant and graceful people I know, and I always feel so happy whenever you come by to the bench and work on your projects. Thank you for your generosity in helping me with RNAseq analysis and for going out with me to fun vegetarian restaurants. I promise you, one day, I'll go back to being vegetarian and we can really celebrate. Thank you, Owen and Kevin, for being the dynamic duo- I literally learned about TnpB through Owen, and I am so thankful for how he is ready to drop everything at a moment's notice to help you out. Kevin, you were one of the first faces I met at the lab, and working with you on the RNA-targeting Cas9 was such a joy. I will miss your sardonic sense of humour and your incomparable work ethic. We are all so honoured to be in your presence. Thank you, Luis, for being such a fun, cool, and chill dude (we're going to Puerto Rico!) and Santi for your podcast next to my bench. Thank you, Ron Boger, for being a true friend and being down to talk about whatever! Thank you, Kai Chen- sitting next to you was the greatest honour of my scientific life, and you have helped me so much in thinking about what it means to become a faculty member and how to best communicate my science rigourously. You have such an amazing way of seeing the bigger questions at hand, and your passion for your craft is infectious. Thank you, Alex Ehrenberg, for your insights into evolutionary biology and for inspiring me to have faith in myself and continue research, even when I didn't necessarily have it. I really owe you a big one. Thank you, Bel Esain, for always believing in me and for being such a friendly presence in the lab. Thank you, Erin Doherty (Dr. D!!!), for your patience and your guidance in science and in mentorship- it is always such a source of joy to see you at lab, and I cannot thank you more for your reassurances and your passion as I go through the scientific process. Thank you, Kamakshi, for your immense contributions to our project and your even bigger contributions to my lab life- from going on temple visits and restaurant treks to parties, you are such an amazing friend and a source of joy. Sir

Evan DeTurk, thank you for your classy charm and big heart, and for all the fun we all had. I am so excited for you! Thank you Deirdre Killebrew for being such an amazing presence in the U-19 grant: from teaching me about how you communicated as a dolphin trainer and seeing how that can be applied to being a good scientist to working with me on troubleshooting ddPCR questions, to letting me stay in your house and chill with Calypso (we got tons of great exercise together, and you inspired me to think about getting a dog!), you have truly made a second scientific home in this groundbreaking therapeutic collaboration. Thank you, Rania, for your technical guidance and for our fun conversations on Italian food and travel. Thank you, Dr. Kenny Loi- you are absolutely cracked at science, and you've always been a source of happiness and joy in my time in the lab. I can always rely on you for fun hugs and great thinking, and you've made the lab such a place of joy and comfort to me.

If there is anybody in lab that I would like to highlight most, it must be Peter Yoon. I first got to know Peter better when I started rotating in the lab and he told me that he was committed to stealing reagents from anybody who went to Harvard (I went to Harvard and felt really nervous). When I learned that he kept turtles, I had never felt a greater sense of joy and camaraderie: my love of animals is literally why I went to study biology, and it is rare to see someone with the same hobbies as you. You got me hooked onto turtles- I started off turtle-sitting the hatchlings, and you gave me this beautiful breeding pair which I am so grateful for. I love these turtles, and I really wish that you could give fish a shot. Oh well. I'm always down if you want to build Grace a pea puffer tank or a killifish setup. But seriously, thank you for being the best friend that I could literally ask for. Seriously. From doing the grand aquarium tour after I got placed into the Doudna lab to being my workout buddy when I started off to teaching me about CRISPR-Cas ancestry to traveling to Florida with me to see the Daytona Reptile show and going to Oahu together, you are truly an amazing scientist and an even better friend. I will always miss you dropping by to drink tea and eat snacks, and Dr. Loi also popping over to chill and to chew the fat together. I personally believe that you are the most innovative scientist in our lab, and your character is unparalleled. I am so happy that you have been in the lab for the whole time that I was there, and I am so excited to be graduating with you. At this point, I consider you family and am so honoured to have known you in my time at the lab.

Speaking of family, I would be remiss if I did not thank my incredible brother, Ray, who is an immense source of my pride and joy. You are so incredibly talented and committed to science, and I cannot wait to see who you will become. Speaking to you is always an immense source of joy for me and thank you for your kindness and support in my journey. Thank you, Mom and Dad, for being the most amazing parents ever. I could always count on you whenever I needed advice, guidance, or just general life conversations. You have always been there for me, and you are the purest embodiment of what I know to be success.

Namo Buddhaya. Namu Dharmaya. Namu Sanghaya.

Chapter 1: Introduction

Genome editing has revolutionized the field of biology ever since its inception. Starting with the first isolation description of restriction enzymes in the 1970s, which cleave DNA in a sequence specific manner, harnessing proteins to modify the genome have facilitated the manipulation of genes, including gene deletion and the insertion of new sequences (Loenen et. al. 2014). Adopting I-SceI endonuclease for the creation of double stranded breaks in mouse chromosomes showcased that genome editing can also be applied to eukaryotic cells, opening the window to therapeutic and agricultural applications of the technology (Rouet, Smith, and Jasin 1994). These double stranded breaks can be both purposed for gene deletion or gene insertion, as the break can stimulate homologous recombination at the target site (Jasin 1996). The target specificity conferred by these restriction enzymes comes from the protein structure recognizing specific DNA sequences and subsequent developments in genome editing, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), sought to build out the programmability of these tools by tiling different DNA binding domains together to target specific sequences of DNA (Wood et. al. 2011, Pacesa et. al. 2024). While these proteins are highly active and very sequence specific, the laborious process required for reprogramming these proteins to target a new sequence, which entails protein engineering, have impeded their wider applicability (Pacesa et. al. 2024).

In 2007, prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)- Cas (CRISPR-associated) systems were discovered to act as adaptive immunity defence against foreign DNA sequences (Barrangou et. al. 2007, Makarova et. al. 2017). These systems consist of a CRISPR array which contains repetitive DNA sequences (repeats) separated by unique spacers which encode for the genomic sequence to target and either multiple or a single effector protein (Makarova et. al. 2017). The classification of CRISPR-CAS systems depends on the components of the effector portion: Class 1 CRISPR-CAS systems have multiple components, and Class 2 CRISPR-CAS systems have a single component effector protein that contains nuclease domains which perform cleavage on the target substrate (Koonin, Makarova and Zhang et. al. 2017). The majority of CRISPR-Cas proteins used for genome editing purposes are Class 2 CRISPR-Cas proteins: within this family are included the type II systems, with Cas 9 as the effector protein, the type V systems, with Cas12 as the effector protein, and the type VI systems, with Cas13 as the effector protein. Cas9 and Cas12 proteins target DNA, while Cas13 targets RNA (Koonin and Makarova 2019). Cas9 cleaves DNA through cleaving the target strand with its HNH nuclease domain and the nontarget strand with the RuvC nuclease domain (Nishimasu et. al. 2014), while Cas12 cleaves both strands using a RuvC mediated mechanism (Zetsche et. al. 2015, Dong et. al. 2016, Cofsky et. al. 2020). Due to their programmable nature and their single-component architecture, Cas9 and Cas12 proteins are commonly used in genome editing applications (Wang and Doudna 2023).

Delivery of Cas9 and Cas12 nucleases are often mediated by delivery vehicles such as adeno-associated virus (AAV) in mammalian cells (Wang et. al. 2024) and tobacco rattle virus (TRV) in plants (Weiss et. al. 2025). However, a significant issue is the packaging limit of these delivery vehicles: while AAV has a packaging limit of 4.7-5.0 kB (Kosaka et. al. 2025) and TRV has a packaging limit of ~1.3 kB (Liu and Page 2008), the commonly used SpCas9 is ~4.1kB (Nishimasu et. al. 2014) and LbCas12a is ~3.7 kB

(Dong et. al. 2016). Given the sheer size of the canonical CRISPR-Cas nucleases, there has been a concerted effort to discover smaller CRISPR-Cas proteins to overcome the delivery problem. The discovery of small, hypercompact CRISPR-Cas12 orthologues represented an initially promising strategy to find genome editors that could be effectively delivered (Pausch et. al. 2020, Al-Shayeb et. al. 2022, Liu et. al. 2019, Harrington et. al. 2018). However, these small effector proteins do not have high activity in the cellular environment, making the application of these editors difficult.

A long-standing hypothesis in the CRISPR-Cas field was that Cas9 and Cas12 have transposon-derived origins, as they show the highest sequence similarity to the TnpB proteins from the IS605 family of transposons (Koonin, Makarova, and Zhang 2017). The TnpB genes are extremely abundant across the tree of life and are associated with insertion sequences: while the mobilization of these insertion sequences is modulated by the TnpA transposase, the function of TnpB had been unclear given that this protein was not required for transposition (Pasternak et. al. 2013). This interest in understanding the similarities between TnpB and CRISPR-Cas inspired a biochemical investigation of TnpB, where it was uncovered to function similarly to CRISPR-Cas, as a programmable RNA-guided endonuclease (Karvelis et. al. 2021). However, TnpB proteins are a third the size compared to CRISPR-Cas (Sasnauskas et. al. 2023, Nakagawa et. al. 2023). TnpB genome editing has been noted to work in both prokaryotes and eukaryotes, being applied to effect genome editing in plants and animals alike (Li et. al. 2024, Zhang et. al. 2024, Xiang et. al. 2024). Despite the small size of the TnpB proteins, they exhibit greater cleavage activity and specificity compared to many small CRISPR-Cas systems, making them a target of interest for their genome editing applications (Xiang et. al. 2024).

While TnpB proteins exhibit greater cleavage activity compared to the small CRISPR-Cas proteins, the activity is still significantly lower compared to the canonical CRISPR-Cas proteins often in application. Furthermore, while the CRISPR-Cas proteins have been subjected to in-depth biochemical and biophysical analyses on their mechanism, the field has hitherto viewed the TnpB cleavage mechanism as rather similar to CRISPR-Cas, given the sheer similarities in the domain architecture (Dong et. al. 2016, Sasnauskas et. al. 2023, Nakagawa et. al. 2023). However, TnpB proteins have evolved under very different evolutionary pressures: to enable transposon retention after mobilization, TnpB has been shown to have homing endonuclease activity (Meers et. al. 2023). Upon the excision of the insertion sequence mediated by the TnpA transposase, the T-rich transposon-adjacent motif (TAM) at the 5' end of the insertion sequence is brought in proximity to the spacer sequence at the 3' end of the insertion sequence, creating the necessary target sequence for TnpB-mediated double strand break generation and subsequent homology-directed repair to enable transposon retention (Meers et. al. 2023). As this mechanism of transposon-mediated retention requires that the nuclease cleave the host's genomic DNA, TnpB proteins may be subject to different selection pressure, as compared to CRISPR-Cas proteins which target foreign DNA for destruction. In this thesis, I will first detail how my research enabled the development of a rational engineering strategy to produce TnpB proteins that have improved cleavage activity and then describe my mechanistic investigation in how TnpB binding and unwinding of DNA and how modification of distinct residues facilitates this process. I ultimately hope that this work provides a blueprint for utilizing engineering to understand

the basic mechanism of these ancestral proteins and to pave the way for future engineering efforts, rooted in a deep mechanistic understanding of the protein biology.

Chapter 2: Rationally engineered TnpB proteins can be purposed for plant genome engineering

2.1 Introduction

TnpB proteins are bacterial RNA-guided homing endonucleases encoded by IS605 and IS607 transposable elements and proposed to be the evolutionary ancestors of CRISPR-Cas12 enzymes (Karvelis et. al. 2021, Sasnauskas et. al. 2023, Koonin, Makarova, and Zhang 2017, Meers et. al. 2023). Recent cryo-EM structures revealed that, like Cas12, TnpB adopts a bi-lobed architecture with conserved wedge (WED), recognition (REC) and RuvC domains, along with a gating lid subdomain (Sasnauskas et. al. 2023, Nakagawa et. al. 2023). TnpB activity depends on a non-coding RNA, called the reRNA that contains a scaffold for protein binding and a ~16-nucleotide (16-nt) reprogrammable guide region (Karvelis et. al. 2021). Guided by the reRNA, TnpB recognizes complementary DNA sequences adjacent to a transposon-adjacent motif (TAM), enabling RNA-DNA base pairing to form an *R*-loop structure prior to cleavage (Karvelis et. al. 2021, Sasnauskas et. al. 2023, Nakagawa et. al. 2023)

The compact size of TnpBs makes them attractive for genome editing in systems where delivery constraints require smaller editing enzymes (Xiang et. al. 2024, Li et. al. 2024), particularly for plant editing applications for which enzyme delivery using size-constrained RNA viruses is desirable (Weiss et. al. 2025). However, proof-of-principle studies in plants have thus far reported only modest TnpB-catalyzed editing efficiencies, including in *Arabidopsis thaliana* and other plant systems (Weiss et. al. 2025, Thornton et. al. 2026, Zhang et. al. 2024, Karmakar et. al. 2024, Lv et. al. 2024). Overcoming this limitation in TnpB genome engineering requires engineering of this protein: previous efforts have focused on directed evolution and deep mutational scanning, which were used to identify multiple activity-enhancing mutations. However, these approaches require intensive screening and often are not necessarily scalable to other TnpB proteins. To approach a method of rationalizing TnpB engineering, we approached TnpB engineering at the protein-DNA interaction interface. Both approaches yielded a higher efficiency genome editing tool which can be purposed for plant delivery purposes.

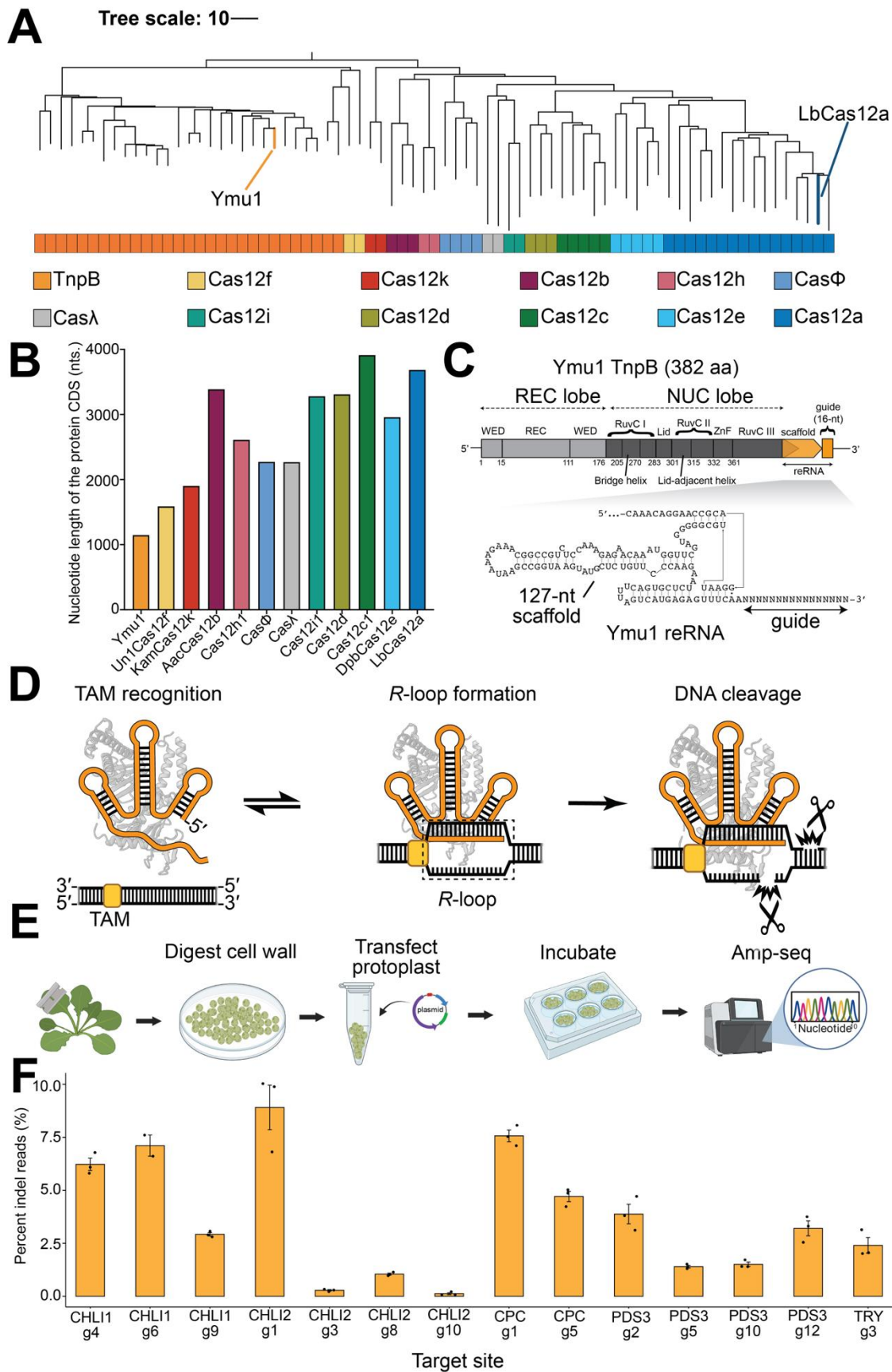
2.2 Results

2.2.1 Ymu1 TnpB is an enzyme that can be utilised for plant genome applications

TnpB enzymes are found in the IS605 and IS607 transposon families and share a deep evolutionary relationship with type V CRISPR-Cas12 systems, with Cas12f proteins representing the closest related clade (Fig. 2.2.1.A). Ymu1 TnpB stands out for its unusually compact size (382 aa), less than one third that of LbCas12a (Fig. 2.2.1B). Among the TnpB orthologues tested, Ymu1 TnpB expressed robustly in *E. coli* and formed stable, homogeneous ribonucleoprotein (RNP) complexes with its reRNA (Fig. 2.5.1A-C), comprising a 127-nt scaffold and a 16-nt guide sequence (Fig. 2.2.1C and Fig. 1.5.1D). Purified Ymu1 RNPs displayed reproducible cleavage of double-stranded DNA (dsDNA) substrates *in vitro* using a previously validated target sequence (Target 1:

5'-TCTTCTGGATTGTTGT-3') from biochemical and single-molecule studies of Cas12 enzymes (Cofsky et. al. 2020, Aris et. al. 2025) (Fig. 2.5.1E-F), enabling mechanistic *in vitro* studies of the reaction pathway as well as the impact that rationalized engineering would have on it (Fig. 2.2.1D).

Previous studies have demonstrated genome-editing activities of Ymu1 TnpB across bacterial, mammalian, and plant systems (Xiang et. al. 2024, Weiss et. al. 2025). To establish a quantitative cellular baseline for mechanistic analysis, we evaluated Ymu1 TnpB activity across multiple endogenous genomic targets in *Arabidopsis* protoplasts using a previously established single-transcript expression system (TnpB-reRNA-HDV) (Weiss et. al. 2025) (Fig. 2.2.1E; Methods). Across all fourteen tested loci, editing efficiencies remained below 10% (Fig. 2.2.1F), consistent with prior reports (Weiss et. al. 2025). Together, these results establish Ymu1 TnpB as a compact and biochemically tractable RNA-guided nuclease with constrained genome-editing efficiency in plants.

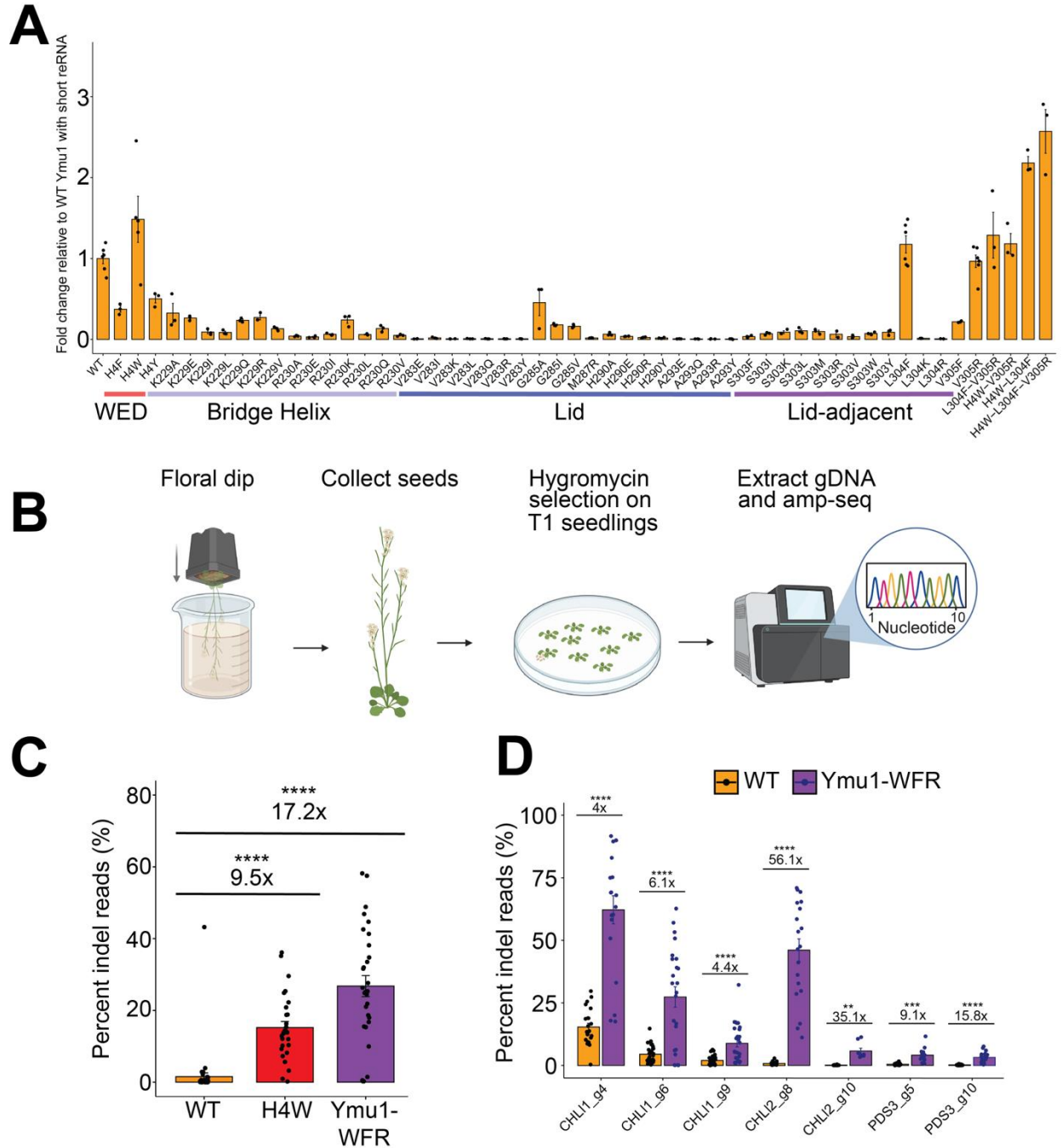


2.2.1. A reconstituted Ymu1 TnpB system for mechanistic and genome-editing studies. (A) Phylogenetic placement of Ymu1 TnpB. Dendrogram of selected TnpB orthologues (N = 29) and Cas12-family nucleases (N = 46) showing their evolutionary relationships. Ymu1 TnpB and LbCas12a are highlighted. (B) Size of Ymu1 TnpB relative to other Cas12 enzymes. Bar plot comparing coding-sequence (CDS) length (nucleotides) of Ymu1 TnpB with representative Cas12 subtypes. (C) Genomic architecture and RNA-protein organization of Ymu1 TnpB. Domain map showing recognition lobe (WED and REC domains) and catalytic lobe (RuvC, lid, ZnF domains), with the bridge helix positioned within RuvC I and the lid-adjacent helix positioned within RuvC II. The reRNA locus lies immediately downstream of the TnpB gene; the guide region is 16-nt. The secondary structure model of the 127-nt reRNA is shown below. (D) Cartoon of the TnpB reaction pathway. TAM binding, R-loop formation, and DNA cleavage are illustrated schematically, with TAM and R-loop positions indicated. (E) Schematic of the *Arabidopsis* protoplast editing assay. Workflow consists of leaf digestion, protoplast isolation, plasmid transfection, incubation, genomic DNA extraction, amplicon library preparation, and next-generation sequencing. (F) Genome-editing activity of Ymu1 TnpB across endogenous *Arabidopsis* loci (Table 2.7.S1). All plant experiments were conducted using a 16-nt guide length (1.5.1.G). Target sites are labeled by gene name and guide number (e.g., *CHL11* g4), where the guide number designates the programmable 16-nt region of the reRNA directed to that site. Percentage of insertion-deletion (indel) reads (y-axis) obtained from amplicon sequencing at multiple genomic sites (x-axis). Bars represent mean $\pm$ SEM (n = 3 biological replicates).

2.2.2. Ymu1 TnpB protein variants accelerate DNA cleavage in plants

The modest editing efficiency of Ymu1 TnpB prompted us to explore whether amino-acid substitutions could improve its performance. Candidate residues were selected based on their predicted proximity to the DNA target in an AF3 structural model of the Ymu1 ternary complex scaffolded from the established Dra2 TnpB structure. Our residue selection was guided in part by positions identified as mutationally sensitive in DMS of the related Dra2 TnpB (Thornton et al., 2026) (Methods). We identified three structural regions to be of great interest in strengthening potential protein-DNA interactions: (i) the WED domain, where H4 is positioned adjacent to the first RNA-DNA base pair; (ii) the bridge helix (K229 and R230), which may couple to distal R-loop formation; and (iii) a cluster of residues spanning the lid and lid-adjacent helix (N282-V305), a region predicted to influence conformational dynamics during R-loop progression and that includes mutationally sensitive positions in Dra2 TnpB (Methods). To screen amino-acid substitutions, we adopted a short reRNA scaffold engineered through SHAPE-guided truncation and stabilization of the 127-nt reRNA (Methods) that preserves the core secondary structure, as confirmed by SHAPE-MaP analysis (Fig. 1.5.2A-B). Both reRNA scaffolds show comparable editing efficiency in protoplast assays (Fig. 1.5.2C). We examined a total of 54 variants (both single and combinatorial substitutions) in *Arabidopsis* protoplasts targeting the *PDS3* g2 site (also termed Target 2: 5'-AAGGCAAATTCGCCGC-3') (Fig. 2.2.2A). Two mutations, a WED-domain mutation H4W and a lid-adjacent-helix mutation L304F, increased editing activity relative to wild-type (WT) Ymu1 (Fig. 2.2.2A). The nearby mutation V305R, which lies in the same lid-adjacent helix, performed similarly to WT in this assay and was retained for further analysis. Analogous substitutions at all three

positions have precedent in prior mutagenesis of Dra2 TnpB: hydrophobic aromatic substitutions are favoured at the position homologous to H4, the native Phe of Dra2 TnpB is preferred over Leu at the position homologous to L304, and mutation to Arg is favoured at the position homologous to V305 (Thornton et. al. 2026). Combining these substitutions yielded a triple variant (H4W-L304F-V305R, hereafter Ymu1-WFR) that consistently outperformed WT and the individual single mutants in the protoplast genome-editing assay (Fig. 2.2.2A).

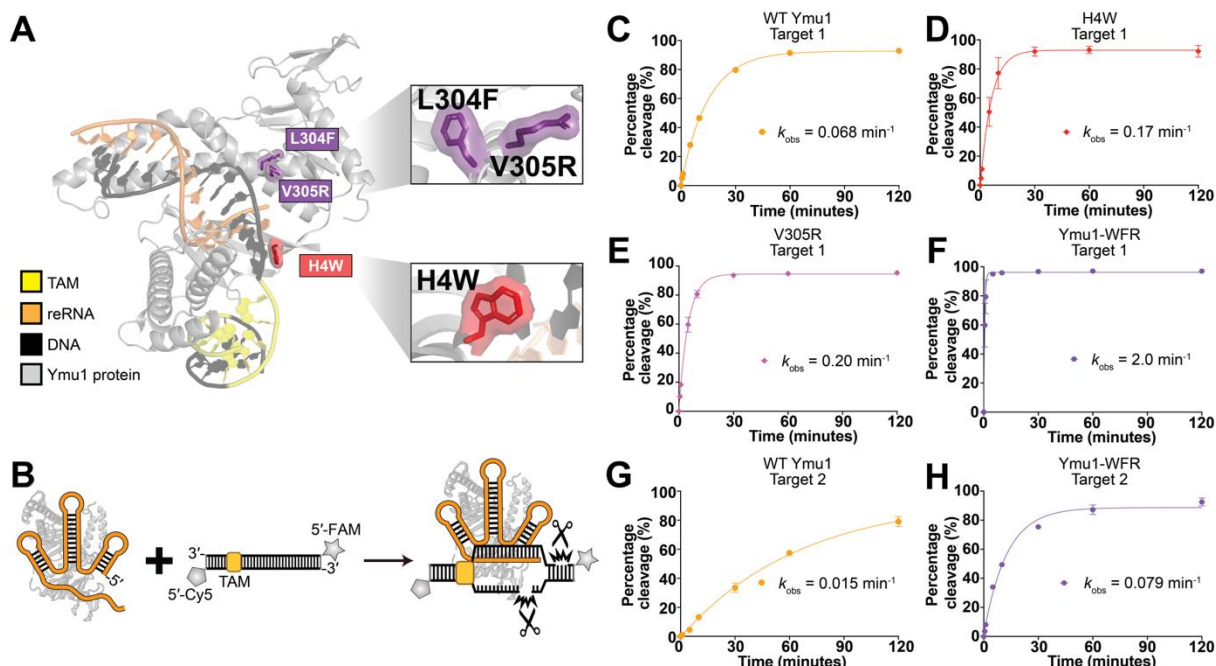


2.2.2. Targeted mutagenesis of Ymu1 TnpB identifies structural determinants that enhance genome-editing activity. (A) Mutational screen of Ymu1 TnpB in *Arabidopsis* protoplasts. Bar plot showing fold change in indel percentage relative to WT Ymu1 TnpB

with the short reRNA (y-axis) for amino-acid substitutions (x-axis) at the endogenous *PDS3* g2 site (Target 2: 5'-AAGGCAAATTGCGCCGC-3'). Structural domains are indicated beneath residue labels. Combinations of selected activating mutations are shown on the right side of the plot, including the triple variant H4W-L304F-V305R ("Ymu1-WFR"). Bars represent mean $\pm$ SEM ($n = 3$ or 6 biological replicates). **(B)** Workflow for T1 whole-plant genome-editing assays. Schematic showing *Agrobacterium* floral dip, T1 seed harvest, hygromycin selection, genomic DNA isolation, and amplicon sequencing. **(C)** Editing efficiencies of WT Ymu1, H4W, and Ymu1-WFR with the 127-nt reRNA in T1 *Arabidopsis* plants at the *PDS3* g2 site. Bar plot shows percentage of indel reads (mean $\pm$ SEM). Fold change and p-values relative to WT are indicated above bars. WT Ymu1 editing data were previously published only for this locus (Weiss et. al. 2025) and are replotted here for direct comparison with mutant variants generated in this study; all experiments were performed using identical methods. **(D)** Editing efficiencies of WT Ymu1 and Ymu1-WFR with the 127-nt reRNA across multiple endogenous genomic loci in T1 plants. Bar plot shows percentage of indel reads (mean $\pm$ SEM). Fold change and p-values relative to WT are indicated above each site. All WT data in **(D)** were generated in this study.

2.2.3. In vitro cleavage recapitulates the efficiency observations made in plant protoplasts

To understand how the enhanced variants achieve higher efficiency, we compared their biochemical cleavage kinetics *in vitro* (Fig. 2.2.3.B; Methods). We measured the cleavage activities of the three single mutants (H4W, L304F, and V305R) compared to WT Ymu1 on a 60-bp dsDNA substrate corresponding to Target 1 (Fig. 2.2.3C-F). Compared to WT Ymu1 (Fig. 2.2.3C and Fig. 2.5.3F), both H4W and V305R mutants cleaved DNA faster (~3-fold) (Fig. 3D-E and Fig. 2.5.3G-H), while the L304F mutant did not display significantly improved kinetics (Fig. 2.5.3E, 2.5.3I). Ymu1-WFR exhibited a pronounced ~28-fold increase in cleavage rate relative to WT (Fig. 2.2.3F and Fig. 2.5.3J). To test whether this enhancement persists in the presence of excess non-target DNA, we performed cleavage assays with a 30-fold mass excess of salmon sperm DNA. Under these competitive conditions, Ymu1-WFR remained 58-fold faster than WT Ymu1 (Fig. 2.5.3K-N). To test whether this improvement was target-specific, we performed the same cleavage assays using RNPs targeting Target 2, corresponding to the *Arabidopsis* *PDS3* g2 site used in plants. Under the same conditions, Ymu1-WFR cleaved Target 2 > 5-fold faster than WT Ymu1 (Fig. 2.2.3G-H and Fig. 2.5.3O-P), mirroring the trend in editing efficiencies observed in plant assays (Fig. 2.2.2A, 2.2.2C and Fig. 2.5.2D). Together, these results suggest that the superior genome-editing performance of Ymu1-WFR arises from intrinsic acceleration of its overall DNA cleavage kinetics, motivating the mechanistic analyses that follow.



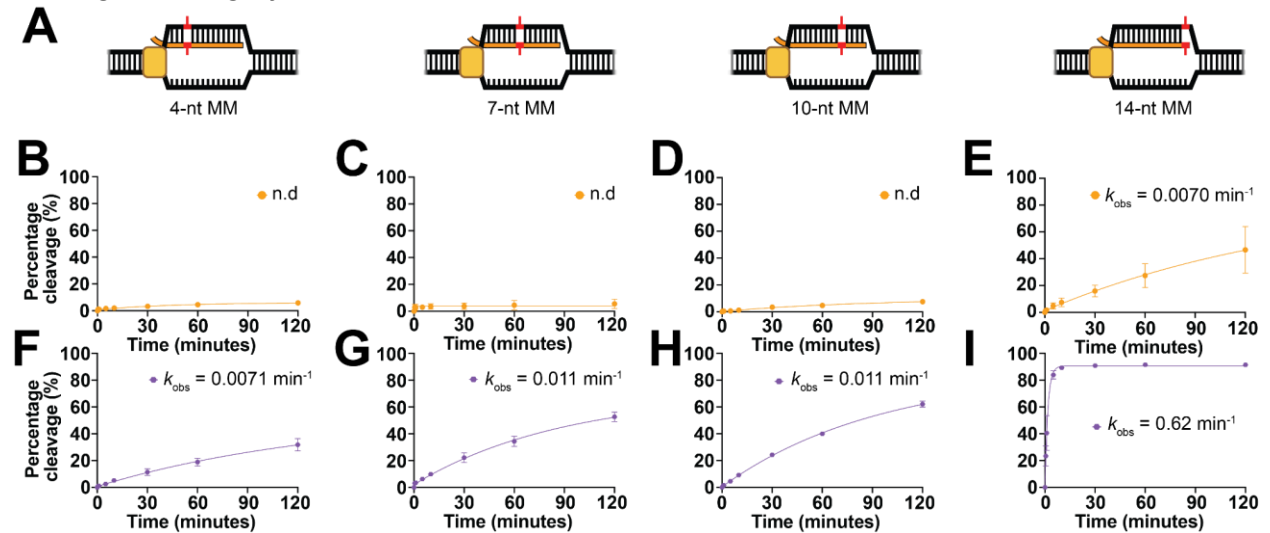
2.2.3. Ymu1 TnpB variants accelerate DNA cleavage *in vitro*. (A) AlphaFold3 (AF3) model of the Ymu1-WFR ternary complex. The three substitutions (H4W, L304F, V305R) are highlighted, with zoomed views showing their local environments within the WED and RuvC lobes. (B) Schematic of the DNA cleavage assay using a labeled double-stranded DNA substrate. The non-target strand (NTS) is 5'-labeled with Cy5 and the target strand (TS) is 5'-labeled with FAM. (C-H) NTS cleavage profiles for Ymu1 TnpB variants under enzyme-excess conditions (100 nM RNP, 10 nM dsDNA) (Methods). The homogeneity of all the RNP samples were analyzed in Fig. 2.5.3A-D. Percentage of cleaved substrate (y-axis) is plotted over time (x-axis). Each point represents mean $\pm$ SD ($n = 3$ independent reactions); k_{obs} values reflect the mean from three independent mono-exponential fits on each reaction time course. Shown are (C) WT Ymu1 (Target 1), (D) H4W (Target 1), (E) V305R (Target 1), (F) Ymu1-WFR (Target 1), (G) WT Ymu1 (Target 2), (H) Ymu1-WFR (Target 2). L304F (Target 1) NTS, cleavage in the presence of salmon sperm competitor DNA, and all TS cleavage profiles are provided in Fig. S3E-P. The single-turnover behavior was verified under substrate-excess conditions (100 nM dsDNA, 20 nM RNP) (Fig. S3Q-R) (Methods). All k_{obs} values (mean $\pm$ SD) are summarized in Table 2.7.S2.

2.2.4 Mutagenesis of Ymu1 TnpB influences its off target editing efficiency

Having established that the WFR mutations enhance on-target cleavage by stabilizing DNA unwinding intermediates, we next asked whether this stabilization also increases tolerance to guide-target mismatches. To evaluate mismatch discrimination, we performed cleavage assays using dsDNA substrates containing single nucleotide RNA-DNA mismatches at positions 4, 7, 10, and 14 within the guide-target heteroduplex for both WT Ymu1 and Ymu1-WFR (Fig. 2.2.4B-I). Both variants displayed strong position-dependent mismatch sensitivity. Mismatches at positions 4, 7, and 10 severely inhibited cleavage for both variants: WT cleavage remained below 10% at 120 min, precluding reliable kinetic fitting, while WFR reached 32-62% cleavage but at substantially reduced rates relative to the matched substrate ($k_{obs} = 0.007\text{-}0.011 \text{ min}^{-1}$, representing a $>100\text{-}$

fold reduction) (Fig. 2.2.4B-D, 2.2.4F-H, Table 2.7.S3). In contrast, a mismatch at position 14 had a more modest effect, reducing WT cleavage ~10-fold and WFR cleavage only ~3-fold relative to their respective matched substrates ($k_{\text{obs}} = 0.006$ and 0.575 min^{-1} , respectively; Fig. 2.2.4.E, 2.2.4.I Table 2.7.S3). These results indicate that while both WT and WFR maintain strong discrimination against mismatches in the TAM-proximal and mid-guide regions, WFR shows increased tolerance at the TAM-distal position, consistent with stabilization of the fully unwound state partially compensating for distal mismatch penalties.

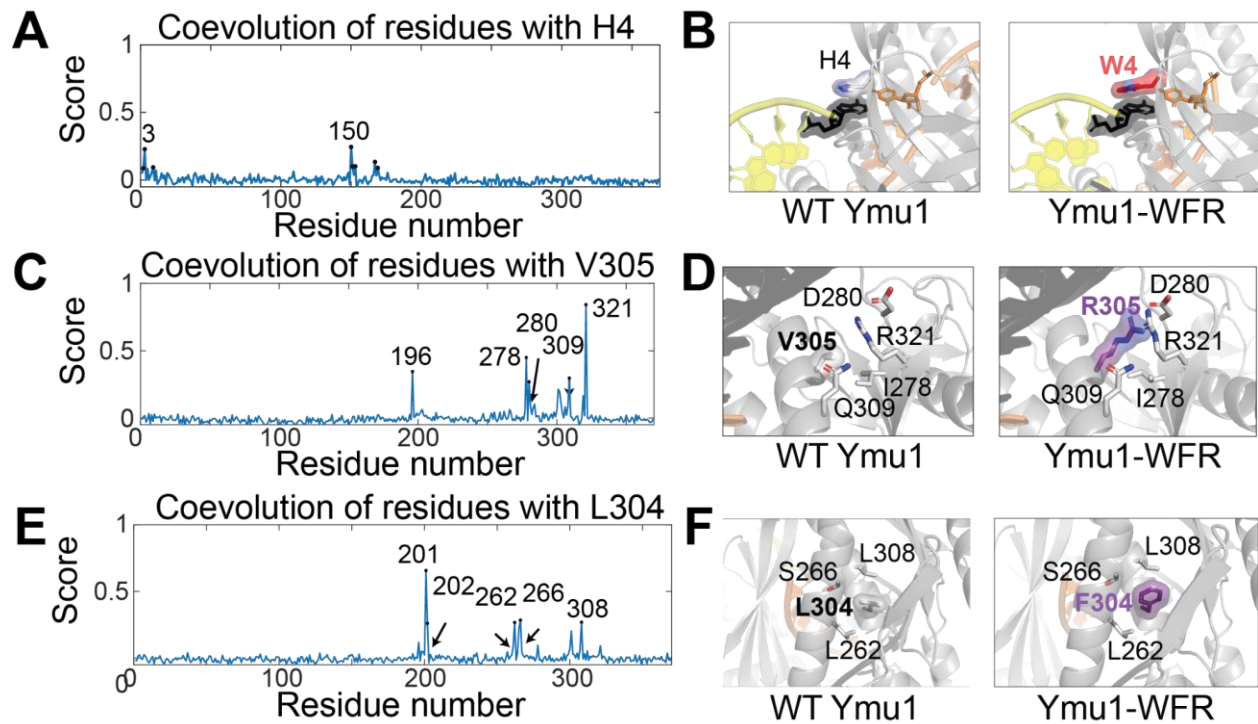
The observation that Ymu1-WFR cleaves mismatched substrates at elevated rates compared to WT raised the question of whether the enhanced activity of Ymu1-WFR also leads to elevated off-target editing *in vivo*. To address this, we performed whole-genome (2,406-5,605x coverage) sequencing of three transgene-free *Arabidopsis* progeny lines derived from a parent plant in which Ymu1-WFR achieved 70.3% and 65.1% somatic editing at the *CHL11* g4 and *PDS3* g12 sites, respectively (Weiss et al., 2026). All three progeny harbored edits at both loci. Cas-OFFinder (Bae et al., 2014) identified 82 and 48 candidate off-target sites for *CHL11* g4 and *PDS3* g12 (Methods), respectively, and no off-target SNPs or indels were detected at any predicted sites (Tables 2.7.S4-5). After filtering against control samples, only 44 unique variants were identified genome-wide, predominantly SNPs rather than the indels characteristic of TnpB cleavage, and none overlapped with predicted off-target sites (Table 2.7.S6). Under the conditions tested, no off-target editing by Ymu1-WFR was detected.



2.2.4. Mismatch tolerance of the WT Ymu1 and Ymu1-WFR occurs in a position-specific manner. (A) Schematic of dsDNA substrates containing single nucleotide RNA-DNA mismatches at positions 4, 7, 10, and 14 within the guide-target heteroduplex. NTS cleavage profiles for (B-E) WT Ymu1 on substrates with mismatches at (B) position 4, (C) position 7, (D) position 10, and (E) position 14 and for (F-I) Ymu1-WFR on substrates with mismatches at (F) position 4, (G) position 7, (H) position 10, and (I) position 14. Percentage of cleaved substrate (y-axis) is plotted over time (x-axis). Each point represents mean $\pm$ SD ($n = 3$ independent reactions); k_{obs} values reflect the mean from three independent mono-exponential fits on each reaction time course ($n=3$). n.d., not determined due to low cleavage levels ($<10\%$) that precluded reliable fitting. k_{obs} values (mean $\pm$ SD) for both NTS and TS cleavage are provided in Table 2.7.S3.

2.2.5. Differential enrichment of rational engineering-derived mutations across the TnpB family tree

To understand how the mutations in our Ymu1 TnpB engineering reflect across nature, we combined sequence co-evolution analysis (Ovchinnikov et. al. 2014) of ~4,500 non-redundant TnpB homologs, AlphaFold3 (AF3) modeling of the Ymu1-reRNA-DNA ternary complex. While sequence co-evolution analysis revealed little evidence of coupling between H4 and other amino acids in TnpB (Fig. 2.2.5A), AF3 modeling places H4 adjacent to the first RNA-DNA base pair of the heteroduplex (Fig. 2.2.5.B). Substitution of His with Trp may introduce a larger aromatic surface to enhance π -stacking at the heteroduplex terminus, stabilizing initial heteroduplex formation. Co-conservation analysis identifies several evolutionarily coupled residues surrounding V305 (I278, D280, Q309, and R321) (Fig. 2.2.5C), consistent with the AF3 model in which V305R is immediately adjacent to this network (Fig. 2.2.5D). More specifically, the AF3-predicted arginine 305 residue in the lid-adjacent helix is oriented to form an interaction with D280 on the counterpart RuvC β -strand, with I278, Q309, and R321 forming additional contacts, suggesting a molecular mechanism for lid domain stabilization in the ternary conformation. Co-evolution analysis of L304 likewise identifies a hydrophobic cluster (L262, S266, L308), consistent with AF3 modeling, which places L304 in a tightly packed hydrophobic pocket formed by these residues from the lid-adjacent helix and the bridge helix (Fig. 2.2.5E-F). Substitution with a bulkier aromatic side chain (L304F) could provide improved packing within this pocket. Consistent with this interpretation, deep mutational scanning of Dra2 TnpB indicates that the native Phe is preferred over Leu at the position homologous to Ymu1 L304 (Thornton et. al. 2026). Together, these structural and covariation analyses support a model in which L304F and V305R strengthen interactions within the lid-proximal region of the RuvC domain, increasing stabilization to protein binding to DNA.



2.2.5. The distinct residues in Ymu1-WFR are enriched at different levels across diverse TnpB orthologues. (A) Co-evolution profile for residue 4 with other protein residues across ~4,500 non-redundant TnpB homologs. (B) AF3 ternary-complex model of Ymu1 TnpB highlighting residue 4 in the WED domain. The H4W substitution is positioned adjacent to the first RNA-DNA base pair of the heteroduplex. (C) Co-evolution profile for residue 305. (D) AF3 ternary-complex model highlighting residue 305 within the lid-adjacent helix, proximal to a network of residues in the catalytic RuvC domain. (E) Co-evolution profile for residue 304. (F) AF3 ternary-complex model highlighting residue 304, located within a hydrophobic interface between the lid-adjacent and bridge helices.

2.3. Discussion

Consistent with prior deep mutational scanning of Dra2 TnpB, which uncovered numerous activity-enhancing substitutions (Thornton et. al. 2026), our results suggest that naturally occurring TnpB variants are not optimized for maximal DNA-cutting activity. Our results establish that targeted mutagenesis of the wedge and lid-adjacent regions of the TnpB lead to increased genome editing in TnpB proteins *in vitro*, in plant protoplasts, and in whole plants. Stacking the H4W, L304F, and V305R mutations of Ymu1 TnpB prompt a synergistic improvement in cleavage activity, but this improvement is also correlated with increased mismatch tolerance in the TAM-distal areas of the sequence *in vitro*, but not in the whole plant context. Our work showcases how rationalized targeting of different domains on the TnpB protein is a viable alternative to the extensive screening previously established for TnpB proteins (Thornton et. al. 2026) and a method to efficiently engineer different TnpB proteins from across the family tree to become highly active, hypercompact plant genome editors.

Structural hypotheses informed by sequence covariation, AF3 modeling, and cryo-EM structures of related TnpB orthologues (Sasnauskas et. al. 2023, Nakagawa et. al. 2023) suggest how the H4W may stabilize early RNA-DNA engagement and how

L304F/V305R may affect protein conformational features associated with full R-loop formation. The same principle may explain why mutations at homologous positions enhance Dra2 TnpB activity (Thornton et al., 2026), suggesting that similar energetic bottlenecks exist across TnpB orthologues, even though the optimal substitution at each position can be orthologue specific. Consistent with this framework, these activating positions have been independently shown to enhance editing across multiple TnpB orthologues in both human cells and plants (Thornton et al., 2026; Nagalakshmi et al., 2026; Weiss et al., 2026). Notably, the enhanced unwinding stability of Ymu1-WFR results in elevated mismatch tolerance *in vitro*, yet whole-genome sequencing of edited *Arabidopsis* plants revealed no detectable off-target editing, suggesting that the stringent TAM requirement and/or cellular context may provide additional layers of specificity beyond R-loop discrimination, which warrants further investigation.

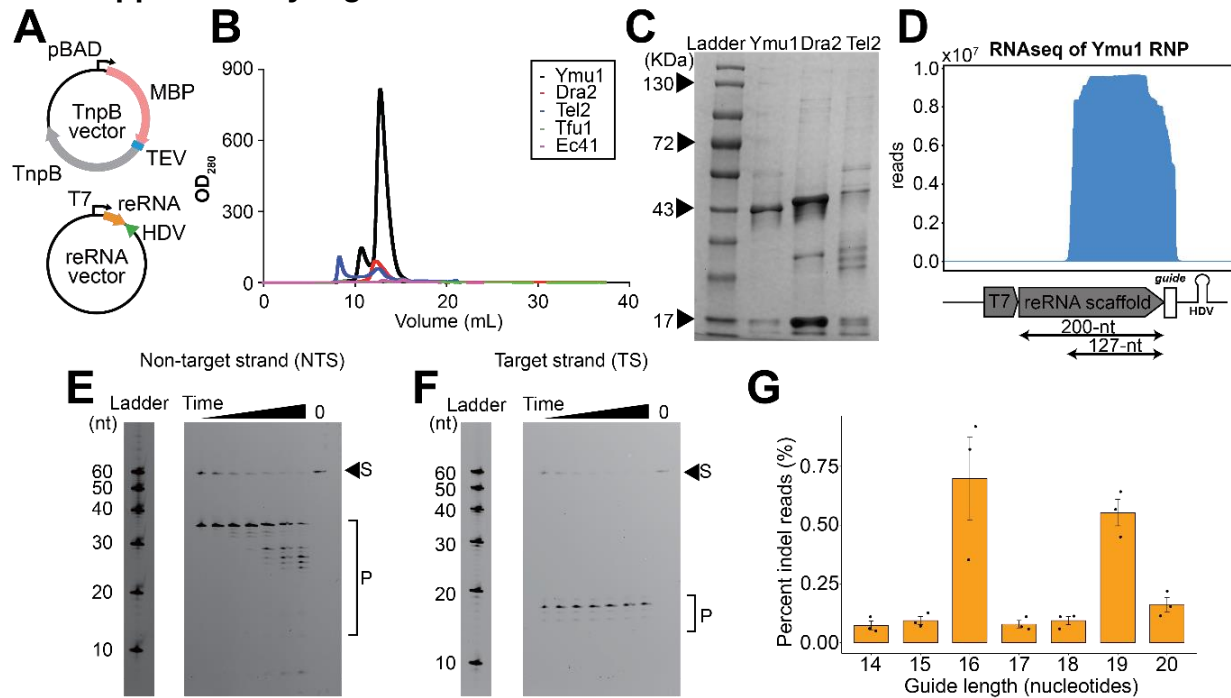
Unlike CRISPR-Cas proteins, TnpB functions endogenously as homing endonucleases which enable transposon retention after excision by making double-strand breaks at the transposon excision site to induce homology-directed repair (Meers et al. 2023). As excessive amounts of double stranded breaks prove toxic to a cell, TnpB proteins are under selection pressure to have tightly controlled, specific cleavage activity. The de-enrichment of the initial H4W mutant across the TnpB family tree is tightly correlated with the suggested function of this residue. As the H4W mutation stabilizes the initialization of the DNA hybridization process of TnpB after TAM recognition, an overly promiscuous initiation step may promote overly high levels of toxic, off-target cleavage activity and would thus be selected against. The L304F and V305R mutations, in contrast, stabilize the RNA-DNA hybrid once the initial target screening steps have been overcome and are selected for across the TnpB family tree. Engineering TnpB proteins to overcome this initial hurdle and to have increased stability in DNA binding prompts a more CRISPR-Cas like cleavage profile, where cleavage level is high but so too is the risk of potential off-target cleavage.

Taken together, by connecting the rationalized engineering of TnpB with modeling, we can identify steps required to reduce the barriers of TnpB to cleave DNA, providing a blueprint to engineering other hypercompact genome editors and a means to make TnpB more CRISPR-Cas-like, expanding its range of applications.

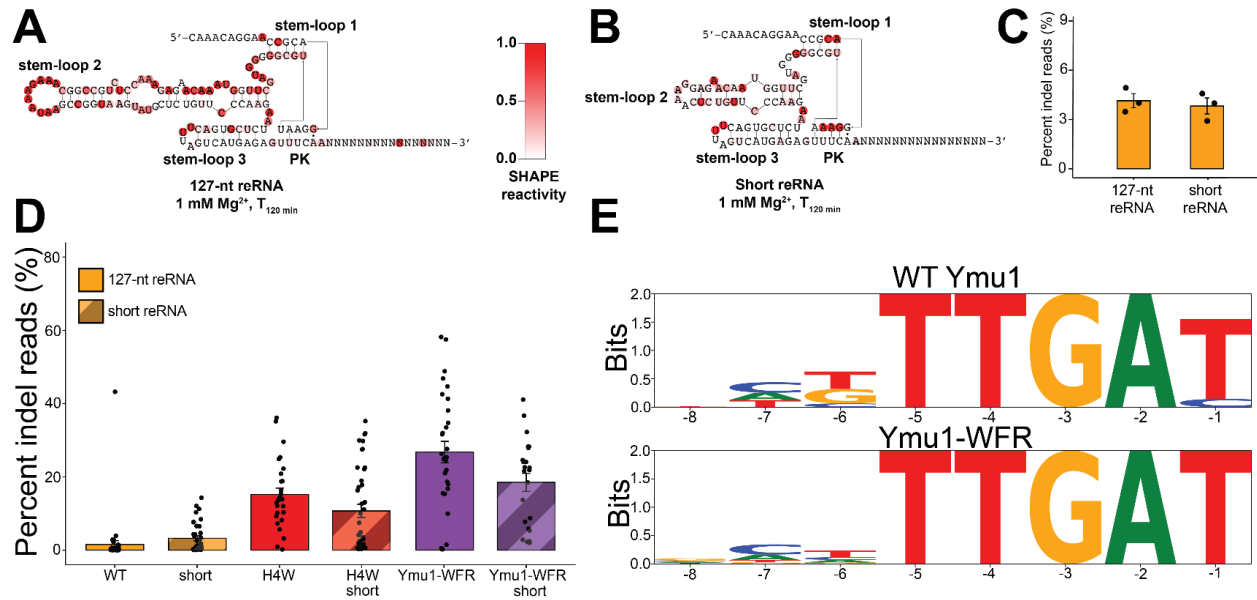
2.4. Limitations of the Study

This study provides mechanistic insights into how DNA unwinding dynamics govern TnpB activity; however, several limitations should be considered. First, our biochemical cleavage assays were performed on short oligonucleotide or plasmid substrates in purified systems, which do not fully recapitulate the complexity of target search in cellular environments, including chromatinized genomic DNA. Second, the genome editing and off-target experiments in this study were conducted exclusively in *Arabidopsis*, and whether the enhanced activity and specificity profile of Ymu1-WFR extend to mammalian or other cellular contexts has not been assessed. Finally, the relationship between the *in vitro* mismatch tolerance observed for Ymu1-WFR and the absence of detectable off-target editing *in vivo* likely involves additional layers of selectivity not captured by biochemical assays and warrants further investigation.

2.5. Supplementary Figures

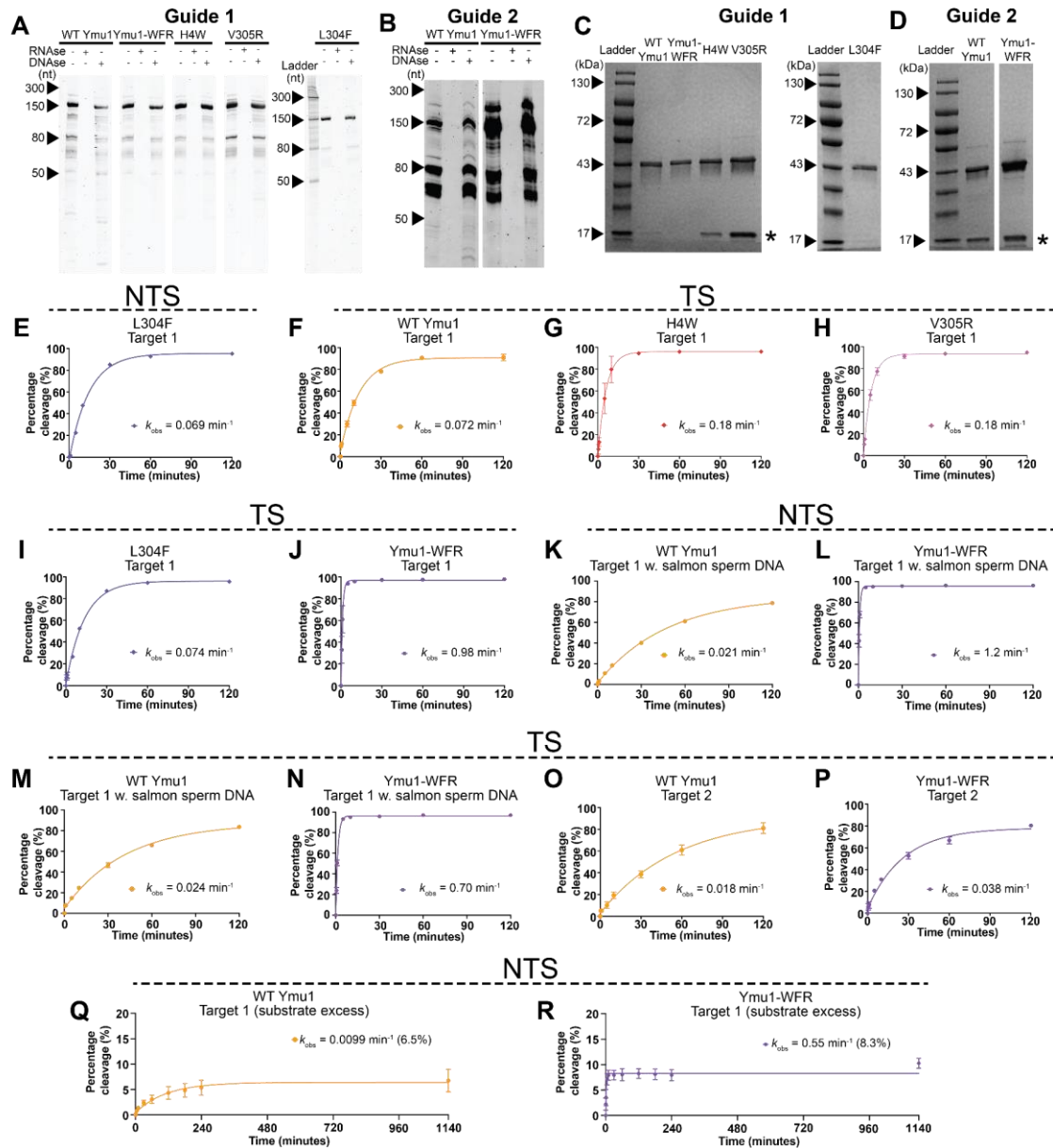


2.5.1. Additional characterization of Ymu1 TnpB and guide configuration. (A) Schematic of the bacterial expression system used for TnpB ribonucleoprotein (RNP) production. TnpB protein is expressed from an arabinose-inducible promoter, while the reRNA is expressed from a separate plasmid under a T7 promoter and processed by a downstream HDV ribozyme. (B) Size-exclusion chromatography traces of five TnpB orthologues (Ymu1, Dra2, Tel2, Tfu1, Ec41). Absorbance at 280 nm (y-axis) is plotted against elution volume (x-axis, mL). (C) SDS-PAGE analysis of purified TnpB orthologues corresponding to panel A, confirming protein homogeneity following affinity and size-exclusion purification. (D) RNA-seq analysis of Ymu1 reRNA expressed in *E. coli* from a T7 promoter using a 200-nt reRNA scaffold. (E-F) Representative denaturing gels of Target 1 cleavage by reconstituted Ymu1 RNP, showing (E) non-target strand (NTS) and (F) target strand (TS) cleavage. The uncleaved substrate (S) and cleavage product (P) are indicated. (G) Guide-length screen in *Arabidopsis* protoplasts using a single-transcript TnpB-reRNA-HDV design with the WT TnpB and 127-nt reRNA. The percentage of indel reads (y-axis) at the endogenous PDS3 g2 target site using guide length ranging from 14-20 nts (x-axis). Bars represent mean $\pm$ SEM ($n = 3$ biological replicates).



2.5.2. Structural and functional analysis of Ymu1 reRNA and scaffold variants (related to Fig. 2).

(A-B) SHAPE-MaP reactivity profile for the **(A)** 127-nt reRNA and **(B)** short reRNA scaffold. Reactivities are mapped onto the predicted secondary structure, with major stem-loop elements and pseudoknot (PK) indicated. **(C)** Protoplast editing efficiencies comparing the 127-nt and short reRNA scaffolds using WT Ymu1. Percentage of indel reads (y-axis) obtained from amplicon sequencing the *Arabidopsis* PDS3 g2 site (x-axis). Bars represent mean $\pm$ SEM (n = 3 biological replicates). **(D)** Comparison of T1 whole-plant editing efficiencies targeting *Arabidopsis* PDS3 g2 using full-length versus short reRNA scaffolds. Bar plot shows percentage of indel reads (mean $\pm$ SEM). **(E)** Bacterial TAM-discovery assay for WT Ymu1 and Ymu1-WFR. Sequence logos summarize nucleotide enrichment adjacent to the target site.



2.5.3. Additional cleavage assays and biochemical quality controls (related to Fig. 3) (A-B) Denaturing PAGE analysis of the reRNA homogeneity for RNP samples used in cleavage assays targeting (A) Target 1 and (B) Target 2. Gels are shown under untreated, RNase-treated, and DNase-treated conditions. (C-D) SDS-PAGE analysis of the TnpB protein homogeneity RNP samples used in cleavage assays targeting (C) Target 1 and (D) Target 2. The asterisk (*) indicates a ~17 kDa minor band identified by mass spectrometry as an N-terminal REC lobe fragment. The percentage of full-length (~44 kDa) protein estimated by densitometry is: (C) WT 84.0%, H4W 73.2%, V305R 61.0%, L304F 82.3%, Ymu1-WFR 82.8% (Target 1); (D) WT 64.4%, Ymu1-WFR 69.8% (Target 2). (E) NTS cleavage profiles for L304F (Target 1). (F-L) TS cleavage profiles for Ymu1 TnpB variants. Percentage of cleaved substrate (y-axis) is plotted over time (x-axis). Each point represents mean $\pm$ SD ($n = 3$ independent reactions); k_{obs} values reflect the mean from three independent mono-exponential fits on each reaction time course ($n=3$). (F) WT

Ymu1 (Target 1). **(G)** H4W (Target 1). **(H)** V305R (Target 1). **(I)** L304F (Target 1). **(J)** Ymu1-WFR (Target 1). **(K-N)** NTS and TS cleavage profiles in the presence of 30-fold mass excess salmon sperm competitor DNA (Methods). **(K)** WT Ymu1 NTS (Target 1). **(L)** Ymu1-WFR NTS (Target 1). **(M)** WT Ymu1 TS (Target 1). **(N)** Ymu1-WFR TS (Target 1). **(O-P)** TS cleavage profiles for **(O)** WT Ymu1 (Target 2) and **(P)** Ymu1-WFR (Target 2). **(Q-R)** Substrate-excess cleavage assays (100 nM dsDNA, 20 nM RNP) for **(Q)** WT Ymu1 and **(R)** Ymu1-WFR on Target 1. Note that low cleavage levels (<10%) limit the precision of k_{obs} estimates under these conditions. k_{obs} values (mean $\pm$ SD) are also provided in Table 2.7.S2.

2.6. Methods

Plasmid construction

All plasmids used in this study were constructed using a combination of PCR amplification, Gibson assembly, and Golden Gate assembly. Primers were synthesized by Integrated DNA Technologies (IDT). PCR reactions were performed with PrimeSTAR GXL DNA Polymerase (Takara Bio, R050A) according to the manufacturer's instructions. Synthetic DNA fragments (gBlocks) were ordered from Twist Bioscience. For Golden Gate cloning, we used BbsI-HF (New England Biolabs, R3539), BsmBI (New England Biolabs, R0739), and PaqCI (AarI; New England Biolabs, R0745) in combination with T4 DNA Ligase (New England Biolabs, M0202) or used the NEBridge Golden Gate Assembly Kit BsmBI-v2 (New England Biolabs, E1602). Gibson assemblies were performed with the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, E2621). All constructs were transformed into either DH5 α (Fisher Scientific, 18-265-017), NEB 5-alpha (New England Biolabs, C2987H), Mach1TM T1^R (ThermoFisher Scientific, C862003) or NEB 10-beta (New England Biolabs, C3019) competent *E. coli* cells and verified by Plasmidsaurus whole-plasmid sequencing to confirm complete construct integrity. A complete list of key plasmids, primer sequences, guide sequences, and vector maps is provided in Tables 2.7.S1 and 2.7.S7-10.

Bacterial TnpB protein expression plasmids

To express TnpB proteins in bacteria, we built pBAD-10 $\times$ His-MBP-TEV-TnpB expression vectors. Coding sequences (CDSs) for TnpB homologs (Ymu1, Dra2, Tel2, Tfu1, and Ec41) were obtained either from a previously described Ymu1 plasmid pMK061 (Weiss et. Al. 2025). or as gBlocks that included both the CDS and the downstream noncoding region encoding the reRNA (Xiang et. al. 2024). For bacterial expression, all TnpB CDSs were cloned into pHS354, a BbsI-based dropout vector containing an araBAD promoter, an N-terminal 10 $\times$ His-MBP-TEV tag, and an mKate stuffer cassette flanked by BbsI sites. Each CDS was PCR-amplified with primers introducing BbsI-compatible overhangs and inserted into pHS354 by BbsI digestion and Golden Gate ligation, replacing the mKate stuffer with the TnpB open reading frame and yielding 10 $\times$ His-MBP-TEV-TnpB constructs (Table 2.7.S9). Point mutations (H4W, L304F, V305R), the catalytic inactivation mutation (dYmu1/E279A), and combinatorial variants (Ymu1-WFR, dH4W, dWFR) were generated using primer-encoded substitutions. Mutant constructs were assembled by GXL PCR followed by two-fragment Gibson assembly.

Bacterial reRNA expression plasmids

To express reRNAs in bacteria, we built T7-reRNA-HDV expression vectors. We first constructed pHS357, a BsmBI-based dropout vector containing a T7 promoter, a BsmBI stuffer cassette, and a 3' HDV ribozyme. The reRNA scaffolds and their corresponding 16-nt guide sequences were PCR-amplified with primers introducing BsmBI overhangs and inserted into pHS357 by replacing the stuffer cassette via BsmBI Golden Gate assembly (Table 2.7.S10). For Ymu1, we tested two reRNA architectures: a 200-nt scaffold corresponding to the native genomic noncoding region (pHS610) and a shorter 127-nt processed scaffold identified by bacterial small-RNA sequencing (pHS607). Notably, expression of the 200-nt scaffold in bacteria yields a processed ~127-nt reRNA species (Fig. 2.5.1D), consistent with endogenous RNA processing. Direct expression of the 127-nt scaffold resulted in higher yield of RNP in bacteria and was used for all DNA cleavage assays and equilibrium AuRBT measurements. The 200-nt scaffold construct was used to reconstitute the dYmu1 RNP employed in torque-driven AuRBT experiments. For all reRNA-expression constructs, we tested guide sequences corresponding to both Target 1 (5'-TCTTCTGGATTGTTGT-3') (Cofsky et. al. 2020, Aris et. al. 2025) and Target 2 (5'-AAGGCAAATTCGCCGC-3', targeting *PDS3* g2 site) (Weiss et. al. 2025).

Bacterial TAM-profiling assay plasmids

Single-transcript TnpB-reRNA constructs for bacterial TAM profiling were expressed from a tetracycline-inducible TetR/pTet promoter, with TnpB and reRNA encoded on the same transcript and an HDV ribozyme positioned at the 3' end to define the reRNA terminus, as described previously (Weiss et. al. 2025). The WT Ymu1 single-transcript construct (pHS516) was used as reported. To generate the Ymu1-WFR variant (pHS751), the three substitutions (H4W, L304F, and V305R) were introduced by PCR using mutagenic primers, and the single-transcript cassette was reassembled by Gibson assembly into the same expression scaffold.

Plant genome-editing plasmids

Plant genome-editing constructs for *Arabidopsis thaliana* consisted of a single transcriptional unit expressing TnpB, the reRNA, and a 3' HDV ribozyme (TnpB-reRNA-HDV), driven by the UBQ10 promoter and terminated by the RbcS-E9t terminator.

To construct a subset of Ymu1 protein variant plant editors targeting *PDS3* g2 locus, we used pMK525, a TnpB-reRNA ccdB dropout cloning vector for PaeI golden gate cloning to insert gene fragments (from Twist Bioscience) containing the Ymu1 TnpB, reRNA, and *PDS3* g2 sequence. Additional plant-editing vectors targeting other endogenous genomic loci were created by cloning guide sequences into Ymu1 guide ccdB dropout vectors pMK025, pTW2503, or pTW2543. Guide sequences were introduced by phosphorylating and annealing complementary IDT oligos and inserting the resulting duplex into pMK025, pTW2503, or pTW2543 using PaeI Golden Gate assembly. All the endogenous target sequences are listed in Table 2.7.S1. All assembled plasmids were transformed into NEB 10-beta cells and validated by whole-plasmid sequencing. The benchmark WT Ymu1 protoplast experiments (Fig. 2.2.1F) and all whole-plant data in the main figures (Fig. 2.2.2C-D) were generated using the 127-nt reRNA scaffold. For the *PDS3* g2 locus, WT Ymu1, H4W, and Ymu1 WFR were additionally tested in parallel using the short reRNA

scaffold to assess compatibility (Fig. 2.5.2D). All mutagenesis screening experiments in protoplast were performed using the short reRNA scaffold (Fig. 2.2.2A).

For high-throughput screening of protein mutations in *Arabidopsis* protoplasts targeting *PDS3* g2, we constructed a single-transcript vector (pKV86) containing a short reRNA scaffold derived from pMK061. To generate pKV86, we amplified two backbone fragments (12 kb and 800 bp) from pMK061 and assembled them with a 200-bp gBlock encoding the short reRNA plus guide sequence in a three-part Gibson reaction. To generate TnpB variant constructs for protoplast screening, we used pHS550 as the backbone, which is a *PaqCI* dropout plasmid containing an mRFP1 stuffer under the UBQ10 promoter, followed by an HDV ribozyme and the RbcS-E9t terminator. Point mutations were encoded directly on *PaqCI*-compatible primers. The TnpB-short-reRNA-guide cassette (as in pKV86) was split into two non-overlapping PCR fragments, with the junction positioned such that each desired mutation resided on one fragment. These fragments were amplified separately and assembled with the pHS550 backbone using a three-part *PaqCI* Golden Gate reaction, yielding complete single-transcript TnpB-reRNA-HDV constructs. Combinatorial variants were generated similarly, except that the cassette was divided into three PCR fragments to accommodate multiple mutated regions, which were assembled with the pHS550 backbone in a four-part *PaqCI* Golden Gate reaction.

Plant genome-editing experiments

Protoplast experiments were performed as described in a previous study (Weiss et. al. 2025). WT *Arabidopsis* Columbia ecotype (Col-0) seeds were grown on Jiffy pucks under 12-h/12-h light/dark photoperiod with low-light condition at 20 °C for about 4 weeks. Mesophyll protoplast isolation was then performed as described previously (Yoo et. al. 2007). Isolated protoplast cells were transfected with 20 µg of plasmid and incubated in 6-well plates at 26 °C for 48 h with a 37 °C heat-shock treatment for 2 h at 16 h post transfection. At 48 h post transfection, protoplasts were collected for genomic DNA extraction.

For transgenic plant editing experiments, the *Arabidopsis* Columbia ecotype (Col-0) was used. Transgenic plants were created using the standard floral dip protocol with the Agl0 *Agrobacterium* strain (Zhang et. al. 2006). After the floral dip procedure, seeds were harvested and plated on ½ MS plates containing 40 µg ml⁻¹ hygromycin B, kept at 4 °C in the dark for 3 days to stratify, and then grown under a 16-h/8-h light/dark cycle at 23 °C. After 12 days, transgenic seedlings (except for transgenic seedlings targeting *PDS3* g2) were collected for genomic DNA extraction. The transgenic seedlings expressing Ymu1 targeting *PDS3* g2 were transplanted from the hygromycin B selection plates to soil and grown in a greenhouse (23 °C) for three weeks. After three weeks, leaf tissue was sampled for genomic DNA extraction by collecting and pooling three random leaves on an individual plant. Protoplast and transgenic plant tissue samples underwent genomic DNA extraction using Qiagen DNeasy plant mini kit (Qiagen, 69106). Amplicon sequencing was performed using the Illumina NovaSeqX platform. Analysis of amplicon sequencing was performed using the CrispRvariants R package (v.1.14.0) as previously described (Weiss et. al. 2025). Primers used for amplicon sequencing can be found in Table 2.7.S1.

Whole-genome off-target analysis

Three transgene-free *Arabidopsis* progeny lines harboring heritable edits generated by Ymu1-WFR at the *CHLI1* g4 and *PDS3* g12 sites were obtained from a previous study (Weiss et al., 2026). Two wild-type Col-0 replicates sequenced in a prior study (Weiss et al., 2025) were used as controls. Genomic DNA (50 ng) was sheared to approximately 300 bp using a Covaris S2 sonicator, and libraries were prepared using the Kapa Hyper DNA Kit (Roche, KK8502) following the manufacturer's instructions. Sequencing achieved an average depth of approximately 3,737x (2,406x to 5,605x) coverage, with greater than 99% of the genome covered by mapped reads.

Whole-genome off-target analysis was performed following a previously established pipeline (Weiss et al., 2025) with minor modifications. Briefly, reads were mapped to the TAIR10 reference genome with BWA-MEM (v0.7.17) (Li et al., 2013), and duplicates were marked using GATK (v4.2.0.0) MarkDuplicatesSpark (McKenna et al., 2010). SNPs and InDels were jointly called by GATK HaplotypeCaller and Strelka2 (v2.9.2) (Kim et al., 2018), retaining only variants identified by both callers (intersection via BEDTools v2.26.0) (Quinlan and Hall et al., 2010). Standard GATK hard filters were applied (SNPs: QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 4.0; InDels: QD < 2.0, FS > 200.0, SOR > 10.0). Variants present in the wild-type control samples were subtracted, and remaining candidates with depth below 30x were excluded. All retained variants were manually inspected in BAM alignments to remove calling artifacts. Potential off-target sites were predicted using Cas-OFFinder (v3.0.0) (Bae et al., 2014), allowing up to four mismatches across the TAM and guide RNA sequences (Table 2.7.S4-5).

TnpB-reRNA co-purification

Briefly, *Escherichia coli* BL21-AI cells (ThermoFisher Scientific, C607003) were co-transformed with the 10×His-MBP-TnpB expression plasmid and the corresponding T7-reRNA-HDV plasmid. Starter cultures were grown in 2XYT medium supplemented with ampicillin (0.1 mg/mL), chloramphenicol (0.034 mg/mL), and tetracycline (0.01 mg/mL) for ~18 h at 37°C. A 5-mL aliquot of the starter culture was then used to inoculate 4.5 L of 2XYT expression medium containing the same antibiotics. Cultures were grown at 37°C until they reached an OD600 of 0.6-0.8, at which point TnpB expression was induced with 0.2% (w/v) L-arabinose. Induction proceeded overnight at 16°C for ~16 h.

Cells were harvested by centrifugation and resuspended in Lysis Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 25 mM imidazole; 5% (v/v) glycerol; 1 mM TCEP) supplemented with cOMplete EDTA-free protease inhibitor (Roche). Cells were lysed by sonication and clarified by ultracentrifugation. The supernatant was applied to Ni-NTA resin (QIAGEN), washed five times with Wash Buffer (20 mM Tris-HCl, pH 8.0; 500 mM NaCl; 25 mM imidazole; 5% (v/v) glycerol; 1 mM TCEP) and eluted using 300 mM imidazole in the same buffer. Eluted protein was dialyzed overnight at 4°C in Dialysis Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 25 mM imidazole; 5% glycerol; 1 mM TCEP) in the presence of TEV protease.

Following TEV cleavage, the digest was passed over a HisTrap column (Cytiva) to separate the cleaved TnpB protein from the His-MBP tag. The flow-through containing tag-free TnpB was concentrated and further purified using a HiTrap Heparin HP affinity column (Cytiva), eluting the protein with Ion-Exchange Buffer (20 mM Tris-HCl, pH 8.0; X mM NaCl; 10% (v/v) glycerol; 1 mM TCEP), where a NaCl gradient from 250 mM to 1 M.

The fraction was then concentrated and finally purified with size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in Gel Filtration Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 10% (v/v) glycerol; 1 mM TCEP). Purified TnpB was concentrated, aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C. RNP concentrations were quantified by measuring A260 on a NanoDrop spectrophotometer using extinction coefficients calculated for each reRNA.

The homogeneity of proteins within purified RNPs was assessed by SDS-PAGE. Protein samples were mixed with 5× SDS Loading Dye (10% SDS; 30% glycerol; 75 mM EDTA; 250 mM Tris-HCl, pH 6.8; supplemented with bromophenol blue) at a 4:1 ratio and boiled at 95°C for 2 min. Samples were loaded onto Mini-PROTEAN TGX Precast Gels (Bio-Rad) and electrophoresed in a 1× SDS running buffer at 150 V for 45 min. Gels were stained using InstantBlue Coomassie Protein Stain (Abcam, ab119211) for ≥ 30 min, rinsed with MilliQ water, and imaged on a ChemiDoc MP system under the Coomassie Blue Gel 590/110 White Trans setting.

To assess the homogeneity of reRNA within purified RNPs, 3.5 µL of 2 µM RNP was prepared for three conditions: untreated, RNase-treated, and DNase-treated. RNase digestion was performed with DNase- and protease-free RNase A (ThermoFisher Scientific, EN0531), and DNase digestion was performed using RNase-free DNase I (New England Biolabs, M0303S) with 0.5 µL 10× DNase I Reaction Buffer (New England Biolabs, B0303S). Gel Filtration Buffer was added to each sample to a final volume of 5 µL. Reactions were incubated at 37°C for 10 min and quenched by adding an equal volume of colourless denaturing loading dye (94% formamide, 30 mM EDTA, 400 µg/mL heparin). Samples were resolved on a 15% denaturing urea-PAGE gel in 0.5× TBE for 1 h at 25 W. The gel was stained with SYBR Gold (ThermoFisher Scientific, S11494) for 30 min and imaged using a ChemiDoc MP imaging system (Bio-Rad) under the SYBR Gold 590/110 Blue Trans setting.

Oligonucleotide substrates for cleavage assays

All DNA oligonucleotides were synthesized by IDT and HPLC-purified. For the double-stranded DNA (dsDNA) substrates used in cleavage assays, the target strand (TS) was 5'-labeled with FAM and the non-target strand (NTS) was 5'-labeled with Cy5, unless otherwise specified.

To generate dsDNA substrates, equimolar amounts (typically 40 µM each) of complementary single-stranded oligonucleotides were mixed in DNA Annealing Buffer (10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA), heated to 95 °C for 5 min, and then slowly cooled to 35°C over 45 min on a thermocycler. Annealed products were quantified by measuring A260 on a NanoDrop spectrophotometer (ThermoFisher Scientific), and concentrations were calculated using manufacturer-provided extinction coefficients.

The annealing of dsDNA was confirmed by electrophoresis on an 8% native polyacrylamide gel run at 150 V at 4°C for 4 h. Working stocks were diluted to 200 nM and stored at -20 °C. The same procedure was used to prepare canonical and mismatched DNA substrates. Sequences of all DNA oligonucleotides used in this study are provided in Table 2.7.S11.

Oligonucleotide cleavage assay

For DNA cleavage assays, TnpB-reRNA RNP complexes were prepared by incubating 100 nM RNP in Cleavage Buffer (1× condition: 10 mM Tris-HCl, pH 7.5; 100 mM NaCl; 2 mM MgCl₂; 1 mM EDTA; 1 mM DTT) at 26 °C for 15 min. Cleavage reactions were initiated by adding the pre-assembled RNP (76 µL) to annealed dsDNA substrate (4 µL) to yield a final concentration of 10 nM dsDNA and 100 nM TnpB RNP in 1× Cleavage Buffer. For competitive cleavage assays, salmon sperm DNA (Fisher Scientific, AM9680) was added to the reaction at a 30-fold mass excess relative to the dsDNA substrate prior to initiation (final concentration: 12 ng/µL salmon sperm DNA to 10 nM (0.4 ng/µL) dsDNA substrate). For substrate-excess assays, the concentrations were adjusted to 100 nM dsDNA and 20 nM TnpB RNP.

Reactions were quenched at defined time points (0.5, 1, 5, 10, 30, 60, and 120 min) by mixing an equal volume of 2× Quench Buffer (94% formamide; 30 mM EDTA; 400 µg/mL heparin). Samples were resolved on 15% denaturing urea-PAGE gels in 0.5× TBE at 40 W for 1 h. Gels were imaged using a Amersham Typhoon scanner (Cytiva), detecting FAM at 488 nm with a Cy2 emission filter (525BP20) or detecting Cy5 at 635 nm with a Cy5 emission filter (670BP30).

Band intensities were quantified using Bio-Rad ImageLab 6.1. Cleavage extent was calculated as the fraction of cleaved product relative to total lane signal. For all datasets, cleavage kinetics were fitted in GraphPad Prism using a mono-exponential model:

$$Y = A \times (1 - e^{-k_{obs} \times t})$$

where Y is the fraction cleaved, A is the reaction amplitude, k_{obs} is the apparent rate constant, and t is time (min). The k_{obs} values are reported as mean ± SD from independent mono-exponential fits to each reaction time course (n=3).

Small RNA sequencing

Purified reRNA was isolated directly from assembled TnpB RNP complexes using the RNA Clean & Concentrator-5 Kit (Zymo Research) according to the manufacturer's instructions. Approximately 20 ng of purified RNA was used for end-repair treatment with QuickCIP (New England Biolabs, M0525S) and T4 PNK (New England Biolabs, M0201S). RNA samples were incubated in a 20 µL reaction containing T4 PNK buffer (New England Biolabs, B0201S) and 1 µL QuickCIP at 37 °C for 30 min, followed by heat-inactivation at 85 °C for 10 min. Without additional cleanup, 1 µL T4 PNK was added directly to the same reaction and incubated at 37 °C for 30 min; ATP was then added to a final concentration of 5 mM and the mixture incubated for an additional 30 min at 37 °C. End-repaired RNA was purified again using the RNA Clean & Concentrator-5 Kit.

RNA-seq libraries were prepared using the Colibri™ Stranded RNA Library Prep Kit for Illumina Systems (Invitrogen, A38994024) following the manufacturer's protocol. Purified RNA (10 µL) was combined with 10 µL of 2× Adapter Mix (BLUE), and adapter hybridization was performed in a thermocycler using a cooling ramp of -0.5 °C s⁻¹ with incubation at 65 °C for 10 min and 20 °C for 5 min. A ligation master mix (26.25 µL 2× Ligation Buffer (YELLOW) + 5.25 µL 10× Ligation Enzyme Mix) was added to the hybridization reaction for a total volume of 50 µL, followed by incubation at 20 °C for 15 min with the thermocycler lid off. Reverse transcription was performed by adding 50 µL of RT master mix (42 µL 2.5× RT Buffer (RED) + 10.5 µL 10× SuperScript™ IV Enzyme Mix) to the 50 µL ligation reaction and incubating at 50 °C for 10 min, followed by 85 °C

for 5 min (lid at 90 °C). cDNA was purified using Dynabeads Cleanup Beads with two sequential bead cleanups and eluted in 20 µL Elution Buffer. Libraries were amplified using KAPA HiFi DNA Polymerase (Roche, KK2101) with an annealing temperature of 65 °C to minimize primer-dimer formation. Amplified libraries were size-selected by gel extraction on E-Gel™ EX 4% Agarose Gels (Invitrogen, G401004), quantified using the KAPA Library Quantification Kit (Roche, KK4873), and sequenced at the Innovative Genomics Institute Next-Generation Sequencing Core using an Illumina NextSeq 1000/2000 P2 v3 kit.

Paired-end reads were trimmed and merged with fastp (v0.23.2) (Chen et. al. 2018). Merged reads with lengths between 100 and 150 nucleotides were extracted using fastq-filter (v0.3.0) (<https://github.com/LUMC/fastq-filter>), then mapped to TnpB loci with BWA (v0.7.17)⁴⁶. The resulting alignments were sorted with SAMTools (v1.17) (Li et. al. 2009) and converted to BigWig format with deepTools2 (v3.5.1)⁴⁸ for visualization in Integrative Genomics Viewer (IGV) (Li et. al. 2009).

SHAPE-MaP RNA Structure Probing

SHAPE-MaP was performed as described previously (Smola et. al. 2015) with minor modifications. For each reaction, 1 µg purified RNP was mixed with 12 µL nuclease-free water and 6 µL 3.3× folding buffer (333 mM HEPES pH 8.0, 333 mM NaCl). Eight microliters of this mixture were combined with 1 µL 10× MgCl₂ (10 mM) and incubated at 37 °C. RNA was modified by adding 1 µL freshly prepared 100 mM 1M7 in DMSO or 1 µL DMSO, then brought to 100 µL and purified using the Qiagen RNeasy Micro Kit (QIAGEN, 74007), eluting in 14 µL water. Ten microliters of modified or control RNA were mixed with 0.8 µL random primer 9 (200 ng/µL) and 0.2 µL transcript-specific primer (2 µM), heated to 65 °C for 5 min, cooled on ice, and combined with 8 µL 2.5× MaP buffer (5× Pre-MaP buffer mixed 1:1 with 30 mM MnCl₂). After incubation at 25 °C for 2 min, 1 µL SuperScript II (ThermoFisher Scientific, 18-064-071) was added and reverse transcription was performed at 25 °C for 10 min, 42 °C for 3 h, and 70 °C for 15 min. cDNA was diluted to 68 µL and purified with MicroSpin G-25 columns (Cytiva, 27-5325-01). Double-stranded cDNA was generated with the NEBNext Second Strand Synthesis Module (New England Biolabs, E6111S) in an 80 µL reaction (68 µL cDNA, 8 µL buffer, 4 µL enzyme mix), incubated at 16 °C for 2.5 h, purified with a Zymo DNA Clean & Concentrator-5 kit, and eluted in 10 µL water. DNA was quantified by Qubit dsDNA HS (ThermoFisher Scientific, Q32851). Libraries were prepared with the Illumina Nextera XT kit. Tagmentation used 10 µL Tagment DNA buffer, 5 µL dsDNA (0.2 ng/µL), and 5 µL Amplicon Tagment Mix, incubated at 55 °C for 5 min and quenched with 5 µL NT buffer. Indexed PCR contained 25 µL tagmented DNA, 15 µL NPM, and 5 µL each index primer; cycling was 72 °C 3 min; 95 °C 30 s; 14 cycles of 95 °C 10 s, 55 °C 30 s, 72 °C 30 s; and 72 °C 5 min. Products were purified using AMPure XP beads (1.8×), washed twice with 80% ethanol, and eluted in 15 µL water. Libraries were checked on an Agilent TapeStation and molarity calculated with the Illumina Molarity Calculator and sequenced at the Innovative Genomics Institute Next-Generation Sequencing Core using an Illumina NextSeq 1000/2000 P1 v3 kit. Illumina reads were processed with ShapeMapper2 using: `shapemapper --name <run_name> --target <target.fasta> --modified --folder <modified_fastqs_dir> --untreated --folder <untreated_fastqs_dir>` SHAPE reactivities were visualized in VARNA on the Ymu1 reRNA 2D structure.

Bacteria TAM assays

TAM depletion assays were performed in *E. coli* NEB 10-beta cells harboring a plasmid library in which all 8-nt TAM sequences are positioned immediately upstream of a single target site, following a modified version of the protocol described in a previous study (Yoon et. al. 2023). For each Ymu1 variant (WT, pHS718; Ymu1-WFR, pHS751), 100 ng of the corresponding TnpB-expression plasmid was electroporated into 50 μ L NEB 10-beta cells already containing the TAM library. Electroporation was performed at 1.7 kV using a time constant of 4.0-4.7 ms. Cells were recovered in NEB 10-beta/Stable Outgrowth Medium (New England Biolabs, B9035S) at 37 °C for 1 h. Following recovery, 50 μ L and 100 μ L aliquots were plated onto Nunc™ Square BioAssay Dishes (Thermo Scientific, 240845) containing LB agar supplemented with chloramphenicol, carbenicillin, and 2 nM anhydrotetracycline (Sigma-Aldrich, 37919-100MG-R) to induce TnpB expression. Plates were spread with sterile glass beads for 5-10 min and incubated at room temperature for 2 days to allow TAM depletion. Cells were harvested by adding 10 mL LB medium containing chloramphenicol and carbenicillin, resuspending the colonies, and recovering at 37 °C for 1 h.

For each library, 1 mL and 3 mL aliquots of the recovered culture were subjected to plasmid extraction using the QIAprep Spin Miniprep Kit (QIAGEN) and eluted in 50 μ L nuclease-free water. The TAM region was amplified using Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, M0530) and prepared for deep sequencing. Libraries were sequenced on a NovaSeq X Plus 1.5B flow cell (Illumina) with 150 bp paired-end reads at the UCLA Broad Stem Cell Research Center High-Throughput Sequencing Core. Raw FASTQ files were quality-trimmed using Trim Galore v0.6.10 (Babraham Bioinformatics) with a minimum read length of 75 bp and a quality threshold of Q20. Paired-end reads were merged using PEAR v0.9.6 (Zhang et. al. 2014). Downstream analysis was performed in Python 3.8 using Biopython v1.85 for FASTQ parsing, pandas v1.4.0 and NumPy v1.22.1 for TAM extraction and quantification, and custom scripts to calculate log₂ fold-changes relative to both the naïve library and a non-targeting negative control. Motif logos were generated from position-weight matrices derived from the 20 most depleted TAM sequences using Logomaker v0.8.7 and matplotlib v3.5.1.

AF3 structural modeling

A SHAPE-informed AlphaFold3 (AF3) model of the Ymu1 TnpB-reRNA-DNA complex was generated using the public AlphaFold Server (<https://alphafoldserver.com>) (Abramson et. al. 2024). AF3 was run in multi-chain mode using the exact protein, RNA, and DNA sequences listed in Table 2.7.S12. Specifically, the protein chain comprised full-length Ymu1 TnpB (382 aa). The RNA chain contained the 127-nt reRNA scaffold with a 16-nt guide base-paired to Target 1, including the TAM. To preserve experimentally supported RNA secondary features, we introduced nucleotide substitutions during AF3 runs that enforce known elements such as the pseudoknot (guided by SHAPE reactivity). The DNA duplex was represented by a 22-nt target strand spanning the TAM and protospacer and an 8-nt non-target strand containing only the TAM and immediately downstream bases, which prevents unintended TS-NTS annealing while preserving the local R-loop interface. Our AF3 modeling of Ymu1 TnpB consistently converges on a conformation in which the lid domain occludes the RuvC active site and positions the catalytic residues far from the scissile phosphates, matching the ternary state captured

experimentally in Dra2 TnpB (ternary conformation 1) (Sasnauskas et. al. 2023). Because the detailed architecture of the distal reRNA scaffold remains uncertain and is not central to our mechanistic conclusions, we do not interpret or discuss its fine structural features. Accordingly, in all figures we display only the protein and DNA components of the AF3 model (omitting the reRNA scaffold) for clarity.

Protein mutagenesis

Residues for mutagenesis were selected based on their nucleic acid-proximal positions in the AF3 model and their predicted roles in target DNA recognition, and positions identified as mutationally sensitive in DMS of Dra2 TnpB (Thornton et al. 2026). At the N-terminus, H4 in the WED domain sits adjacent to the first DNA-RNA base pair and was substituted with hydrophobic aromatic residues (W, F, Y) to probe potential contributions to initiating strand separation. Aromatic substitutions at this position are enriched in Dra2 TnpB DMS (Thornton et al. 2026), and similar engineering has been applied to Cas12f (Hino et al., 2023). Two bridge-helix residues, K229 and R230, were each substituted with a panel of charge-altering and hydrophobic residues to evaluate potential coupling to distal *R*-loop formation. Finally, we introduced substitutions across a cluster of residues within the lid/lid-adjacent segment (V283, G285, M287, H290, A293, S303, L304, V305). The lid subdomain undergoes conformational rearrangement upon *R*-loop formation in Cas12a and related compact Cas12s (Stella et al., 2018; Pausch et al., 2021; Guan et al., 2025), and this region includes positions homologous to mutationally sensitive residues in Dra2 TnpB, including E302 (Ymu1 S303) and I304 (Ymu1 V305) (Thornton et al. 2026). At S303, nine substitutions were tested; at L304, three (F, K, R); at V305, two (F, R). Combinatorial variants were generated by pairing the activating single mutations (L304F-V305R, H4W-V305R, H4W-L304F, and H4W-L304F-V305R). In total, 54 single and combinatorial variants were screened (Fig. 2.2.2A).

Bioinformatic analysis

For generation of the tree used in Fig. 2.2.1A, a selection of TnpB and Cas12 genes spanning diverse types were selected from a prior study (Xiang et. al. 2024) and CasPEDIA database (Adler et. al. 2023) and folded using ColabFold2 (Mirdita et. al. 2022) on default settings. The dendrogram was generated using DaliLite.v5 (Holm et. al. 2019) with the --matrix flag.

For additional phylogenetic analyses of the three substitutions (Fig. 2.2.5), TnpB homologous sequences were first identified using PSI-BLAST via the MPI Bioinformatics Toolkit (<https://toolkit.tuebingen.mpg.de>), searching against the NR70 databases. PSI-BLAST was run for 8 iterations using the BLOSUM45 scoring matrix, with E-value cutoffs of 1×10^{-3} for both reporting and inclusion in subsequent iterations, and a maximum of 10,000 target hits. The resulting multiple sequence alignments were used directly for downstream analysis. Each MSA was filtered to remove sequences containing gaps at any of the analyzed positions. Residue frequencies at each column were computed by tallying amino acid symbols and dividing by the total number of non-gap residues (i.e., the number of sequences in the filtered alignment). To identify residues that co-evolve and are likely to mediate interactions across structural interfaces in TnpB, we trained a Potts model on the TnpB multiple sequence alignment (MSA). Specifically, we used the TnpB MSA derived from the NR70 dataset (as described above) and removed all

alignment columns containing gaps in the Ymu1 TnpB reference sequence. We then analyzed this MSA with the GREMLIN Colab notebook⁵⁸ and obtained average product-corrected (APC) coupling matrices. In these matrices, we examined positions 4, 304 and 305 corresponding to Ymu1 residues H4, L304 and V305, and selected the top five positions with APC scores > 5 standard deviations above the mean. H4 was weakly coupled to Q3, L150, indicating low significance of these potential contacts for the overall structure across TnpB homologs. Two clusters of strongly coupled residues: (1) D201, N202, L262, S266, L308 for L304, and (2) D196, I278, D280, Q309, R321 for V305, were mapped onto WT Ymu1 and Ymu1-WFR AF3 structures (Fig. 2.2.5 B, D, and F)..

2.7. Supplemental Tables

Table S1. Target site sequences and amplicon-sequencing primers used in plant genome-editing experiments (related to Fig. 2.2.1 and Methods)

Gene ID	Target site	Target site sequence (5'-3')	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
AT4G14210	AtPDS3_g2	aaggcaaattcgccgc	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTgaagcagttgtgagtt aagttggaga	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Tttgtcttaagcgcttga gaagtgg
AT4G18480	AtCHLI1_g4	CTGTTACCTGAGATTA	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTGTGGTGTTATG ATTATGGGAGATAG AG	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TTCGCAATAAC AGGAACTTGCT C
AT4G18480	AtCHLI1_g6	GAAGTTAATCTCTTGG	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTAAGCCTTTGAG CCTGGTTTG	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TCGGGTGAGAA ATCGAAATCCC
AT4G18480	AtCHLI1_g9	CGGTTTGGTATGCATG	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTGCGAGGTTTAT CTTGATCGGTTC	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TTCACGGAAAT CCTTTGGGTTA CTA
AT5G45930	AtCHLI2_g1	CTCTGTAGCGACATTC	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTtgcagAGAACT TTCTGGAAGAATCCA	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Ttcattccaaaggttca agcttttaatc

AT5G45930	AtCHLI2_g3	TTACCAGTTCCTC TAT	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTGGACAAGATG AGATGAAGCTATGC CTT	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TCTCGGGTCTG AGTTATACGGA TCA
AT5G45930	AtCHLI2_g8	GAAGTTAATCTTTT GG	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTAGGCGTTTGA GCCTGGACTACTA	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TAGCAGGATGA GAAATCGATAT CCCTTC
AT5G45930	AtCHLI2_g10	TCAGCTGCATCTG GTT	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTTAGCTAAAGCT AATAGAGGGATTCTT TATGTTGAT	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TGTGGTCTAAG CTCTCCTTCTTC AGGA
AT2G46410	AtCPC_g1	cggtttgagtctgatt	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTCCAAGGCTTCT TGTTCCGAAG	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TCCAACGAGTT TATACATCCGA GAA
AT2G46410	AtCPC_g5	tacagcatgtttgtat	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTtgattcttagcaaaac atattcttaatttatgtca	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Tccgataaaaaccgc ataaagtttgt
AT4G14210	AtPDS3_g5	taccatcctaaagta	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTgacgtcaggaagaa catggtcatttg	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Tagaacatttcagcgc taatgctacaa
AT4G14210	AtPDS3_g10	taactgtactacctc	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTtgttgcagcttatgaca cacaataag	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Tgaatctttatctactg gcaatcata
AT4G14210	AtPDS3_g12	gcgttgagcatataa	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTgaaccgacccgaga agagatttg	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC Ttggaatacacacattt gtacaacca

AT5G53200	AtTRY_g3	AGCAGGAAGAGTT CCT	ACACTCTTTCCCTAC ACGACGCTCTTCCG ATCTtttattatgaaaatata aatgctaatacttgggat	GTGACTGGAGT TCAGACGTGTG CTCTTCCGATC TTGGCGTCGTT TATCAGCAAAG
-----------	----------	----------------------	--	---

Table S2. Rate constants (k_{obs}) for DNA cleavage assay (n=3) (related to Fig. 2.2.3, and Fig. 2.5.3)

Protein	Target	Strand	Mean (min^{-1})	SD (min^{-1})
WT Ymu1	1	NTS	0.0676	0.0033
		NTS (with salmon sperm)	0.0212	0.0017
		NTS (substrate excess)	(0.00994)†	0.0040
		TS	0.0717	0.0090
		TS (with salmon sperm)	0.0238	0.0018
		TS (substrate excess)	(0.00541)†	0.0025
	2	NTS	0.0152	0.0019
		TS	0.0179	0.0019
H4W	1	NTS	0.168	0.039
		TS	0.176	0.055
V305R	1	NTS	0.198	0.025
		TS	0.179	0.023
Ymu1-WFR	1	NTS	2.00	0.74
		NTS (with salmon sperm)	1.23	0.18
		NTS (substrate excess)	(0.555)†	0.20
		TS	0.980	0.33

		TS (with salmon sperm)	0.705	0.058
		TS (substrate excess)	(0.323)†	0.11
	2	NTS	0.0793	0.0038
		TS	0.0380	0.0032
L304F	1	NTS	0.0692	0.0043
		TS	0.0737	0.0023

†Fitted values are shown for completeness but are not interpreted quantitatively due to total cleavage <10% at the final time point, which renders k_{obs} unreliable for kinetic comparison.

Table S3. Rate constants (k_{obs}) for DNA-RNA mismatch cleavage assay (n=3) (related to Fig. 2.2.4)

Strand	DNA target	WT Ymu1 mean (min^{-1})	WT Ymu1 SD (min^{-1})	WFR mean (min^{-1})	WFR SD (min^{-1})
NTS	4 nt mismatch	(0.0671)†	(0.0849)†	0.00710	0.000944
	7 nt mismatch	(1.26)†	(1.22)†	0.0111	0.00274
	10 nt mismatch	(0.0119)†	(0.00712)†	0.0106	0.000752
	14 nt mismatch	0.00696	0.00320	0.621	0.221
TS	4 nt mismatch	(2.129)†	(0.681)†	0.0192	0.00987
	7 nt mismatch	(0.503)†	(0.848)†	0.0158	0.00188
	10 nt mismatch	(0.285)†	(0.468)†	0.0182	0.00307
	14 nt mismatch	0.00739	0.00674	0.335	0.0589

†Fitted values are shown for completeness but are not interpreted quantitatively due to total cleavage <10% at the final time point, which renders k_{obs} unreliable for kinetic comparison.

Table S4. Cas-OFFinder predicted off-target sites for *CHLI1* g4 (related to Fig. 2.2.4)

Genomic DNA (5'-3')	Chromosome	Position	Strand	Mismatches	Classification	Variants Detected
TTGATtTGTTACCTGAG ATag	chr1	385713	+	3	Off-target	No SNP/InDel
aTGATCTGTTACCTGAtc TTt	chr1	456752	-	4	Off-target	No SNP/InDel
aTGATgaGTTACCTGAG ATTg	chr1	1091615	+	4	Off-target	No SNP/InDel
TTGATCTcTTACCTccGA TTt	chr1	5232246	+	4	Off-target	No SNP/InDel
TTGtTCTGcTtCCaGAGA TTA	chr1	5486474	-	4	Off-target	No SNP/InDel
TTGATtTGTTACCTaAtAT Tg	chr1	6483939	-	4	Off-target	No SNP/InDel
TTtATCTaTTtCCTGAGAA TA	chr1	8044117	-	4	Off-target	No SNP/InDel
TTcATCTGTTcCtTGtGAT TA	chr1	8655476	+	4	Off-target	No SNP/InDel
TTGtTCTGTTACCTGAtc TTt	chr1	8701665	-	4	Off-target	No SNP/InDel
TTGAaCTGTTtCCTGtGA aTA	chr1	9221199	-	4	Off-target	No SNP/InDel
TTGcTgTGTTACCTGAG AaTt	chr1	1067279 1	-	4	Off-target	No SNP/InDel
TTcATCTcgTACCTGAtAT TA	chr1	1105313 5	+	4	Off-target	No SNP/InDel
TTGAaaaGTTAgCTGAG ATTA	chr1	1271607 5	+	4	Off-target	No SNP/InDel
TactTgTGTTACCTGAGA TTA	chr1	1348221 7	-	4	Off-target	No SNP/InDel
aTcATCTGTTtCCTGAGA aTA	chr1	1360070 5	+	4	Off-target	No SNP/InDel
TgGcTCTGTTACaTGAA TTA	chr1	1369190 6	+	4	Off-target	No SNP/InDel

gTGATCTtTTcCCTGAaA TTA	chr1	1483263 2	+	4	Off-target	No SNP/InDel
gTGATCTtTTcCCTGAaA TTA	chr1	1491338 2	-	4	Off-target	No SNP/InDel
gTGATgTGTTcCCTGAaA TTA	chr1	1602902 3	+	4	Off-target	No SNP/InDel
TTGATaTaTTtCCTGAGA TgA	chr1	1764856 4	+	4	Off-target	No SNP/InDel
TTGAgCTaTTAgaTGAGA TTA	chr1	1849922 2	-	4	Off-target	No SNP/InDel
TTGATCTGTTtCaGAGA TTt	chr1	1994876 5	+	4	Off-target	No SNP/InDel
TTGATCTGTTACCTaAct TaA	chr1	2151907 2	+	4	Off-target	No SNP/InDel
TTGATgTGcTtCCTGAcA TTA	chr1	2232501 1	-	4	Off-target	No SNP/InDel
TcGATCTGTgACCTcAG ATTt	chr1	2425020 7	+	4	Off-target	No SNP/InDel
TTGATCTcTTAgCTcAGA TTc	chr1	2639011 0	-	4	Off-target	No SNP/InDel
TTGtTCTGTTtCCTcAaAT TA	chr1	2971510 1	-	4	Off-target	No SNP/InDel
TTGATCTGaTgCCTGAG AcTc	chr1	2990742 3	+	4	Off-target	No SNP/InDel
TTtATCTaTTtCCTGAGAc TA	chr2	4774056	-	4	Off-target	No SNP/InDel
TTGAgCTGTTtCCTGtGA TaA	chr2	6423541	+	4	Off-target	No SNP/InDel
TTGAagTGTTACgTGAG AcTA	chr2	6984586	-	4	Off-target	No SNP/InDel
TTGATaTgaTACaaGAG ATTA	chr2	8181489	-	4	Off-target	No SNP/InDel
TTGAaCTGTTcCCTtAGA TaA	chr2	8826433	-	4	Off-target	No SNP/InDel

TTGATCTGTTAttaaAGATTA	chr2	9993416	+	4	Off-target	No SNP/InDel
TTGATgTGTgACCTGAtATTA	chr2	11593142	-	3	Off-target	No SNP/InDel
TTGATCTaTTACCTaAGATcA	chr2	12290104	-	3	Off-target	No SNP/InDel
TTGATCTcTTAaTGAGgTTA	chr2	12900822	+	4	Off-target	No SNP/InDel
TgGATCTGaaAtCTGAGATTa	chr2	13747304	-	4	Off-target	No SNP/InDel
TTtcTCTGTTAtCTGAGAAATA	chr2	14024696	-	4	Off-target	No SNP/InDel
TTGATaTcTTACTTGAGATTt	chr2	15133368	+	4	Off-target	No SNP/InDel
aTGAACTGTTACCTGAGATTc	chr2	15659366	+	3	Off-target	No SNP/InDel
TTGATCaGTTACaTttGATTA	chr2	18248786	-	4	Off-target	No SNP/InDel
TTGtTaTGTTAgCTGAGATTg	chr2	18317151	+	4	Off-target	No SNP/InDel
TTGAaCTGTTtCaTGAGcTTA	chr2	18905832	-	4	Off-target	No SNP/InDel
TTaATCTaTaACCTtAGATTa	chr3	218050	-	4	Off-target	No SNP/InDel
TTGATCgGTTAggTGAGATTg	chr3	2589316	-	4	Off-target	No SNP/InDel
TTGATaTGTTtCaTGAGATgA	chr3	4058168	-	4	Off-target	No SNP/InDel
TTGATtTGTTgtCTGAGATTt	chr3	6726855	+	4	Off-target	No SNP/InDel
gTcATCTGTTtCCTGAGaTA	chr3	11854134	+	4	Off-target	No SNP/InDel
TTGATCTGTTgatTGAGATTt	chr3	17319931	+	4	Off-target	No SNP/InDel

TTtTCTGTTAgCTGAGA TgA	chr3	1835044 7	-	4	Off-target	No SNP/InDel
TTGAgCTGTTgtCTGAGA TTc	chr3	2035248 2	+	4	Off-target	No SNP/InDel
TTGATCTcTTtCCTtAGAT Tt	chr3	2082028 0	-	4	Off-target	No SNP/InDel
TTtATgTGTTACtGAaAT TA	chr4	4089641	-	4	Off-target	No SNP/InDel
TTGATtTGgTACCTGAGA cTg	chr4	5859717	+	4	Off-target	No SNP/InDel
TTGATtTGgTACCTGAGA cTg	chr4	5877024	+	4	Off-target	No SNP/InDel
TTGATtTGgTACCTGAGA cTg	chr4	5903286	+	4	Off-target	No SNP/InDel
TTtATCTGTTACCTGAaA aaA	chr4	8342444	-	4	Off-target	No SNP/InDel
TTaATCTGTTAaCTaAaA TTA	chr4	9193485	+	4	Off-target	No SNP/InDel
TTGATCTGTTACCTGAG ATTA	chr4	1020275 3	-	0	On-target	Edited (InDel)
cgGATCaGTTACCTGAG cTTA	chr4	1112667 9	+	4	Off-target	No SNP/InDel
TTGgTtTGTTACCTGAGA agA	chr4	1291727 8	-	4	Off-target	No SNP/InDel
TTGATgTGTTAgCTGAaA TgA	chr4	1660318 8	-	4	Off-target	No SNP/InDel
TTGATCTtTTACCaGAcA TTc	chr5	1158214	-	4	Off-target	No SNP/InDel
TTGATCTtTTACCaGAGg cTA	chr5	3746803	-	4	Off-target	No SNP/InDel
cTGgTCTGTcACCTGAG ATaA	chr5	5295761	+	4	Off-target	No SNP/InDel
TTGATCTGaTAtCTcAGA TTc	chr5	6998326	+	4	Off-target	No SNP/InDel

aTGATCTGTcACCTGtaA TTA	chr5	9245291	-	4	Off-target	No SNP/InDel
TgGtTCTGTTtCCTGAGA TTt	chr5	9636891	+	4	Off-target	No SNP/InDel
TTtAcCTGTTcCCTGAGtT TA	chr5	9794815	+	4	Off-target	No SNP/InDel
aTcATCTGTTAgCTtAGA TTA	chr5	1062926 8	-	4	Off-target	No SNP/InDel
TaGAaCTGTTACCTcAG ATTt	chr5	1163226 5	+	4	Off-target	No SNP/InDel
TTcATCTaTTtCCTGAGA aTA	chr5	1243696 5	-	4	Off-target	No SNP/InDel
TTGATaTGTTtCCTcAGA aTA	chr5	1399295 1	-	4	Off-target	No SNP/InDel
TTGATCTGTaAttTGAaAT TA	chr5	1601530 3	+	4	Off-target	No SNP/InDel
TTGATtTGcTtCCTGAGA TcA	chr5	1862868 7	+	4	Off-target	No SNP/InDel
aTGtTCTGgTACCTGAGA TTg	chr5	1972823 8	+	4	Off-target	No SNP/InDel
TaGATgTGTTACCTGAa ATcA	chr5	1984484 7	-	4	Off-target	No SNP/InDel
TTGATCTtggACCaGAGA TTA	chr5	2051441 6	-	4	Off-target	No SNP/InDel
TTGcTCTGaTACCTGAG AggA	chr5	2096942 0	-	4	Off-target	No SNP/InDel
TTGATCTGTTACtTGAag aTA	chr5	2343559 1	-	4	Off-target	No SNP/InDel
TTGATCaGTTACtTgTGA TgA	chr5	2485260 8	-	4	Off-target	No SNP/InDel
TTGATaTGTTcCaaGAGA TTA	chr5	2688851 4	+	4	Off-target	No SNP/InDel

Potential off-target and on-target sites predicted by Cas-OFFinder (v3.0.0) for the *CHL1* g4 site, allowing up to 4 mismatches across the TAM (**bolded** here) and guide sequence (5'-**TTGATCTGTTACCTGAGATTA**-3'). **Genomic DNA**: matched genomic sequence with mismatches shown in lowercase. **Chromosome**: Arabidopsis thaliana TAIR10 chromosome. **Position**: genomic coordinate. **Strand**: (+) forward, (-) reverse.

Mismatches: number of mismatches between gRNA and genomic DNA. **Classification:** on-target or off-target. **Variants Detected:** SNPs or InDels identified at this site across all three sequenced progeny lines after filtering against wild-type controls.

Table S5. Cas-OFFinder predicted off-target sites for *PDS3* g12 (related to Fig. 2.2.4)

Genomic DNA (5'-3')	Chromosome	Position	Strand	Mismatches	Classification	Variants Detected
TTGATGCGTTtGAGaATcTgA	chr1	603266	+	4	Off-target	No SNP/InDel
TTGATaCtTTGGAGCtTATcA	chr1	1112526	-	4	Off-target	No SNP/InDel
TTGATGatTTGGtcCATATAA	chr1	2663557	+	4	Off-target	No SNP/InDel
TTGATGtGTTtGAGatTATAA	chr1	4238242	-	4	Off-target	No SNP/InDel
TTGATGCGTTGGtGCATAatc	chr1	5305662	+	4	Off-target	No SNP/InDel
TTGgTGtGgTGaAGCATAA	chr1	5880417	-	4	Off-target	No SNP/InDel
TTGtTtCGTaGGAGCATAATat	chr1	9814485	-	4	Off-target	No SNP/InDel
TTGATGCaTTatAtCATATAA	chr1	10322285	-	4	Off-target	No SNP/InDel
TTGATGatTTGGAtgATATAA	chr1	11292702	-	4	Off-target	No SNP/InDel
TTGgTGcTTGGtGCATAATtA	chr1	13817767	-	4	Off-target	No SNP/InDel
TgGATatGTTGGAGgATATAA	chr1	17843545	+	4	Off-target	No SNP/InDel
TTGAgGCGTTGGAGaATATAA	chr1	18348024	+	2	Off-target	No SNP/InDel
TTGATGCtTTtGcGCATAaAA	chr1	19411643	+	4	Off-target	No SNP/InDel
TTGATGgGTTGGtGgATAgAA	chr1	19953639	+	4	Off-target	No SNP/InDel
TTGATGatgTGGAcCATATAA	chr1	23704848	-	4	Off-target	No SNP/InDel

TTGATGCGTTGGAtagTc TAA	chr1	2428816 1	+	4	Off-target	No SNP/InDel
TTGtTGcTTGGAGCATc TgA	chr1	2557151 7	-	4	Off-target	No SNP/InDel
TTGtTatGTTGGAtCATAT AA	chr1	2855945 5	-	4	Off-target	No SNP/InDel
TaaATGCGTTGcAGCATc TAA	chr1	2938906 7	+	4	Off-target	No SNP/InDel
TTGgTGcTTGGtGCATA TtA	chr2	3221482	+	4	Off-target	No SNP/InDel
TTGATGtGTTGtAcCATAa AA	chr2	8176569	+	4	Off-target	No SNP/InDel
TTGATGCGTTGaAGCtT AaAA	chr3	9345906	+	3	Off-target	No SNP/InDel
aTGATGCcTTtGAGCAgA TAA	chr3	1271979 1	-	4	Off-target	No SNP/InDel
TTGATGCGTTGGAGCAT cTct	chr3	1518126 4	+	3	Off-target	No SNP/InDel
TTGATGCGTTtGAGaATg aAA	chr3	1536938 3	-	4	Off-target	No SNP/InDel
TTGAaGCGaTGGAGaAg ATAA	chr3	1802166 3	-	4	Off-target	No SNP/InDel
TTGATGCGTTtGtGCATA Tgt	chr3	2102081 7	-	4	Off-target	No SNP/InDel
TTGAgGCGTTGGtGCAC ATgA	chr3	2199732 3	-	4	Off-target	No SNP/InDel
TTaAgGCGTTGGAGCAg ATcA	chr4	833322	-	4	Off-target	No SNP/InDel
TcGATcCGTTtGAGCATA TAt	chr4	2315047	+	4	Off-target	No SNP/InDel
TTGATGcTTGGgGaATA cAA	chr4	4218388	+	4	Off-target	No SNP/InDel
TTGATGCGTTGGAGCAT ATAA	chr4	8195559	-	0	On-target	Edited (InDel)

TTGATGCGTTGaAGaAc AaAA	chr4	1101914 8	-	4	Off-target	No SNP/InDel
TTaATaCGTTGcAGCATA TAg	chr4	1122813 5	-	4	Off-target	No SNP/InDel
cTGATGCGaTGGAGCAT AcAg	chr4	1334682 6	+	4	Off-target	No SNP/InDel
TTcATGCGTTGcActATAT AA	chr4	1463170 3	-	4	Off-target	No SNP/InDel
TTGAgGCtTTtGAGCATA TAt	chr5	124947	+	4	Off-target	No SNP/InDel
TTGATGCGTatGAatATA TAA	chr5	4677310	-	4	Off-target	No SNP/InDel
TcGAaTCGTTGGAAtCATA TAA	chr5	5784463	+	4	Off-target	No SNP/InDel
TTtATtCGTatGAGCATAT AA	chr5	6069715	-	4	Off-target	No SNP/InDel
TTGATcCGTTGGAGgAa AaAA	chr5	6389952	-	4	Off-target	No SNP/InDel
gTGATGCGTTGGAGtATc TtA	chr5	8922973	-	4	Off-target	No SNP/InDel
TTGATGCGgTGAaATA TAt	chr5	1590810 3	-	4	Off-target	No SNP/InDel
TTGATGCcTTGaAGCATc TAc	chr5	1659042 5	-	4	Off-target	No SNP/InDel
aTacTGCGTTtGAGCATA TAA	chr5	1861416 3	+	4	Off-target	No SNP/InDel
TTGATGatTTGGAAtgATAT AA	chr5	1885182 9	+	4	Off-target	No SNP/InDel
TTGATtgaTTGGAaCATA TAA	chr5	2264483 8	+	4	Off-target	No SNP/InDel
TTGgTGCGTgGaAGCAT AaAA	chr5	2338726 9	+	4	Off-target	No SNP/InDel
TTGAaGCcTTGGAGaAa ATAA	chr5	2639008 4	-	4	Off-target	No SNP/InDel

As described for Table 2.7.S4, for the PDS3 g12 site (5'-TTGATGCGTTGGAGCATATAA-3') .

Table S6. Filtered variants identified in whole-genome sequencing of Ymu1-WFR edited progeny (related to Fig. 2.2.4)

Chromosome	Start	End	Line 126	Line 153	Line 176	Gene
chr1	929479	929579	het SNP	ND	ND	
chr1	5170759	5170859	homo SNP	homo SNP	homo SNP	
chr1	9550381	9550481	homo SNP	homo SNP	homo SNP	
chr1	10641247	10641347	ND	ND	het SNP	
chr1	11146968	11147068	homo SNP	homo SNP	homo SNP	
chr1	12874513	12874613	ND	ND	het InDel	
chr1	18122011	18122111	het SNP	het SNP	het SNP	
chr1	18437377	18437477	ND	ND	het SNP	
chr1	24972594	24972694	het SNP	ND	ND	
chr2	2069663	2069763	het SNP	ND	ND	
chr2	4196468	4196568	homo SNP	homo SNP	homo SNP	
chr2	5067286	5067386	homo SNP	homo SNP	homo SNP	
chr2	6261428	6261528	het SNP	homo SNP	homo SNP	
chr2	11117881	11117981	homo SNP	homo SNP	homo SNP	
chr2	14876309	14876409	ND	het SNP	homo SNP	
chr3	5279155	5279255	ND	ND	het SNP	
chr3	5818127	5818227	ND	ND	homo SNP	
chr3	11547524	11547624	homo SNP	homo SNP	homo SNP	

chr3	12152993	12153093	homo SNP	homo SNP	homo SNP	
chr3	14199881	14199981	het SNP	het SNP	het SNP	
chr3	14449540	14449640	homo SNP	homo SNP	homo SNP	
chr3	15396806	15396906	homo SNP	homo SNP	homo SNP	
chr3	20015222	20015322	ND	ND	het SNP	
chr3	23186204	23186304	homo SNP	ND	ND	
chr4	944800	944900	homo SNP	homo SNP	homo SNP	
chr4	3268734	3268834	het SNP	het SNP	het SNP	
chr4	4025667	4025767	het SNP	het SNP	het SNP	
chr4	6179688	6179789	het SNP	het SNP	het SNP	
chr4	7629083	7629183	ND	het SNP	ND	
chr4	8178664	8178764	het SNP	ND	ND	
chr4	8195000	8196007	homo InDel	het InDel	homo InDel	<i>CHLI1</i> (on-target)
chr4	9457406	9457506	ND	het SNP	homo SNP	
chr4	10202091	10203301	homo InDel	homo inDel	homo InDel	<i>PDS3</i> (on-target)
chr4	10939399	10939499	homo SNP	ND	ND	
chr4	15195953	15196053	ND	het SNP	ND	
chr4	16978520	16978620	ND	het SNP	ND	
chr5	4786957	4787057	ND	het SNP	het SNP	
chr5	10002510	10002610	homo SNP	homo SNP	homo SNP	
chr5	11784970	11785070	homo SNP	homo SNP	homo SNP	

chr5	13657156	13657256	ND	het SNP	homo SNP	
chr5	15163048	15163153	homo SNP	homo SNP	homo SNP	
chr5	19033499	19033599	ND	ND	homo SNP	
chr5	19492943	19493043	het SNP	ND	ND	
chr5	21434792	21434892	ND	homo SNP	het SNP	
chr5	22109092	22109192	ND	ND	het SNP	
chr5	24819665	24819765	het SNP	ND	ND	

Variants remaining after filtering against wild-type control samples, excluding sites with sequencing depth below 30x and manually inspected to remove calling artifacts. **Chromosome, Start, End:** genomic coordinates of the variant. **Line 126, Line 153, Line 176:** genotype call for each progeny line (homo SNP: homozygous SNP; het SNP: heterozygous SNP; het InDel: heterozygous insertion-deletion; homo InDel: homozygous deletion; ND: not detected). **Gene:** annotated gene if the variant overlaps a known locus. On-target edits at *CHL1* and *PDS3* are highlighted.

Table S7. Plasmid vectors used in this study (related to Methods)

Internal ID	System	Purpose
pZZ09	<i>E.coli</i>	Bacterial expression of Ymu1 protein
pHS355	<i>E.coli</i>	Bacterial expression of Dra2 protein
pZZ049	<i>E.coli</i>	Bacterial expression of Tel2 protein
pZZ034	<i>E.coli</i>	Bacterial expression of Tfu1 protein
pZZ053	<i>E.coli</i>	Bacterial expression of Ec41 protein
pZZ069	<i>E.coli</i>	Bacterial expression of V305R Ymu1 protein
pZZ080	<i>E.coli</i>	Bacterial expression of H4W Ymu1 protein
pHS749	<i>E.coli</i>	Bacterial expression of Ymu1-WFR protein
pZZ089	<i>E.coli</i>	Bacterial expression of L304F Ymu1 protein
pHS607	<i>E.coli</i>	Bacterial expression of Ymu1 200-nt reRNA for target 1
pHS610	<i>E.coli</i>	Bacterial expression of Ymu1 127-nt reRNA for target 1
pHS612	<i>E.coli</i>	Bacterial expression of Ymu1 short reRNA for target 1
pZZ062	<i>E.coli</i>	Bacterial expression of Ymu1 short reRNA for target 2
pZZ064	<i>E.coli</i>	Bacterial expression of Ymu1 127-nt reRNA for target 2

pZZ013	<i>E.coli</i>	Bacterial expression of Dra2 200-nt reRNA for target 1
pZZ052	<i>E.coli</i>	Bacterial expression of Tel2 200-nt reRNA for target 1
pZZ035	<i>E.coli</i>	Bacterial expression of Tfu1 200-nt reRNA for target 1
pZZ054	<i>E.coli</i>	Bacterial expression for Ec41 200-nt reRNA for target 1
pHS516	<i>E.coli</i>	Bacterial TAM assay vector encoding WT Ymu1 and WT reRNA expressed as a single transcript, with a 16-nt guide flanked by HDV
pHS751	<i>E.coli</i>	Bacterial TAM assay vector encoding Ymu1-WFR and WT reRNA expressed as a single transcript, with a 16-nt guide flanked by HDV
pTW2036	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 14 nt spacer length targeting AtPDS3_g2
pTW2035	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 15 nt spacer length targeting AtPDS3_g2
pMK061	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 16 nt spacer length targeting AtPDS3_g2
pTW2034	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 17 nt spacer length targeting AtPDS3_g2
pTW2033	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 18 nt spacer length targeting AtPDS3_g2
pTW2032	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 19 nt spacer length targeting AtPDS3_g2
pTW2031	<i>Arabidopsis</i>	WT Ymu1 and WT reRNA with 20 nt spacer length targeting AtPDS3_g2
pTW2314	<i>Arabidopsis</i>	G285A Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2315	<i>Arabidopsis</i>	G285I Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2317	<i>Arabidopsis</i>	G285V Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2321	<i>Arabidopsis</i>	H4W-L304F Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2322	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2397	<i>Arabidopsis</i>	H4W-V305R Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2335	<i>Arabidopsis</i>	L304F-V305R Ymu1 variant with short reRNA targeting AtPDS3_g2
pTW2145	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHL11_g4
pTW2149	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHL11_g6
pTW2125	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHL11_g9

pTW2126	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHLI2_g1
pTW2128	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHLI2_g3
pTW2147	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHLI2_g8
pTW2134	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCHLI2_g10
pTW2198	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCPC_g1
pTW2202	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtCPC_g5
pMK064	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtPDS3_g5
pMK068	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtPDS3_g10
pMK070	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtPDS3_g12
pTW2196	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA targeting AtTRY_g3
pTW2541	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtPDS3_g2
pTW2532	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtCHLI1_g4
pTW2533	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtCHLI1_g6
pTW2534	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtCHLI1_g9
pTW2536	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtCHLI2_g8
pTW2537	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtCHLI2_g10
pTW2542	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtPDS3_g5
pTW2543	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA targeting AtPDS3_g10
pTW2471	<i>Arabidopsis</i>	H4W Ymu1 with WT reRNA targeting AtPDS3_g2
pTW2503	<i>Arabidopsis</i>	H4W-L304F-V305R Ymu1 with WT reRNA ccdB gRNA cloning vector
pTW2453	<i>Arabidopsis</i>	H4W Ymu1 with WT reRNA ccdB gRNA cloning vector
pMK525	<i>Arabidopsis</i>	ccdB TnpB cloning vector
pMK025	<i>Arabidopsis</i>	WT Ymu1 with WT reRNA ccdB gRNA cloning vector
pKV100	<i>Arabidopsis</i>	H4Y Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV101	<i>Arabidopsis</i>	H4F Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV102	<i>Arabidopsis</i>	H4W Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV106	<i>Arabidopsis</i>	K229A Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV105	<i>Arabidopsis</i>	K229E Ymu1 variant with short reRNA targeting AtPDS3_g2

pKV111	<i>Arabidopsis</i>	K229I Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV112	<i>Arabidopsis</i>	K229L Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV104	<i>Arabidopsis</i>	K229Q Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV103	<i>Arabidopsis</i>	K229R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV113	<i>Arabidopsis</i>	K229V Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV110	<i>Arabidopsis</i>	R230A Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV109	<i>Arabidopsis</i>	R230E Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV114	<i>Arabidopsis</i>	R230I Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV107	<i>Arabidopsis</i>	R230K Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV115	<i>Arabidopsis</i>	R230L Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV108	<i>Arabidopsis</i>	R230Q Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV116	<i>Arabidopsis</i>	R230V Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV136	<i>Arabidopsis</i>	V283E Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV138	<i>Arabidopsis</i>	V283I Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV141	<i>Arabidopsis</i>	V283K Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV140	<i>Arabidopsis</i>	V283L Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV137	<i>Arabidopsis</i>	V283Q Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV134	<i>Arabidopsis</i>	V283R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV135	<i>Arabidopsis</i>	V283Y Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV139	<i>Arabidopsis</i>	M287R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV128	<i>Arabidopsis</i>	H290A Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV129	<i>Arabidopsis</i>	H290E Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV126	<i>Arabidopsis</i>	H290R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV127	<i>Arabidopsis</i>	H290Y Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV132	<i>Arabidopsis</i>	A293E Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV133	<i>Arabidopsis</i>	A293Q Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV130	<i>Arabidopsis</i>	A293R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV131	<i>Arabidopsis</i>	A293Y Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV120	<i>Arabidopsis</i>	S303F Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV118	<i>Arabidopsis</i>	S303I Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV124	<i>Arabidopsis</i>	S303K Ymu1 variant with short reRNA targeting AtPDS3_g2

pKV117	<i>Arabidopsis</i>	S303L Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV125	<i>Arabidopsis</i>	S303M Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV123	<i>Arabidopsis</i>	S303R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV119	<i>Arabidopsis</i>	S303V Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV122	<i>Arabidopsis</i>	S303W Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV121	<i>Arabidopsis</i>	S303Y Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV160	<i>Arabidopsis</i>	L304F Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV158	<i>Arabidopsis</i>	L304K Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV161	<i>Arabidopsis</i>	V305R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV162	<i>Arabidopsis</i>	V305F Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV161	<i>Arabidopsis</i>	V305R Ymu1 variant with short reRNA targeting AtPDS3_g2
pKV87	<i>Arabidopsis</i>	WT Ymu1 with short reRNA targeting AtPDS3_g2
pHS550	<i>Arabidopsis</i>	mRFP drop out vector with backbones for protoplast assay

Plasmids pHS516, pMK025, pMK061, pMK64, pMK68, pMK70 are from a previous study (Weiss et. al. 2025).

Table S8. Primers used for Gibson assembly-based site-directed mutagenesis of Ymu1 TnpB for bacterial expression (related to Methods)

Oligonucleotide	Internal ID	Sequence (5'-3')
Reverse primer for amplifying Ymu1 backbone plasmid	oZZ080R	ttgacggcttgacggagtagcatagggttgacg
Forward primer for amplifying Ymu1 backbone plasmid	oZZ081F	gcagggattctgcaaaccctatgctactccgtc
Forward primer to induce the E279A mutation in Ymu1	oZZ086F	AAACCACGATATCATCTGTATCgcgGACCTTAAC GTAAAGGGCATGATGC

Reverse primer to induce the E279A mutation in Ymu1	oZZ087R	GCCCTTAACGTTAAGGTCcgcGATACAGATGATA TCGTGGTTTTTGACTATCTCTGTAC
Reverse primer to induce the H4W mutation in Ymu1	oZZ227R	GTATTCATAGGCTTTccaCTGCAGCATTGCATTG GATTGG
Forward primer to induce the H4W mutation in Ymu1	oZZ228F	AATGCAATGCTGCAGtggAAAGCCTATGAATACC GTATCTATCCAGATAAGAAG
Forward primer to induce the both L304F/V305R mutations in Ymu1	oZZ233F	TGGACGAGCttccgaTCGAAACTGCAGTACAAGG CTTC
Reverse primer to induce the both L304F/V305R mutations in Ymu1	oZZ234R	CAGTTTCGAtcggaaGCTCGTCCATGATACATCAG AGATGC
Reverse primer to induce the L304F mutation in Ymu1	oZZ235R	CTGCAGTTTCGATACgaaGCTCGTCCATGATACA TCAGAGATGC
Forward primer to induce the L304F mutation in Ymu1	oZZ236F	GTATCATGGACGAGCttcGTATCGAAACTGCAGT ACAAGGCTTC
Forward primer to induce the V305R mutation in Ymu1	oZZ217F	TCATGGACGAGCCTTagaTCGAAACTGCAGTAC AAGGCTTC

Reverse primer to induce the V305R mutation in Ymu1	oZZ220R	GTACTGCAGTTTCGAtctAAGGCTCGTCCATGAT ACATCAGAG
---	---------	---

Table S9. Amino acid sequence of the Ymu1 TnpB bacterial expression construct (related to Methods)

ID	Protein sequence
WT Ymu1	MKSSHHHHHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDT GIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPD KAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGV NAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSN IDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQ MSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIE ENLYFQSNAMLQHKAYEYRIYPDKKQETLIAKTIGSSRYVYNHFLELW EYEETGKGLTYACSKLLTKLKRDPETVWLCEVDKFSLQNSLRNLSDAF SRFFKGQNEHPQFKSKSPRQSYTTQYTNNNIAVSGNCLKLPKLGLVKF ADSREMKGRILNATVRRKSSGKFFVSILCEEEICELPKTDSSVGIDLGIDF AVMSDGSRHDNNHFTRQMEERLRREQRKLARRALAAEKRGISLSEARN YQKQRRKVARLYEKVANQRKEYLNKLSTEIVKNHDIICIEDLNVKGM MRNHKLAKSISDVSWTSLVSKLQYKASWYGKEVIRISRWFPSSQICSECGHK DRKKPLHVREWTCVPVCHAHHDRDVNAARNILAEGLRIRALTPGS

10 His-tags; MBP; TEV site; TnpB protein

Table S10. DNA sequence of the 127-nt reRNA expression cassette (related to Methods)

ID	DNA sequence (5'-3')
127-nt reRNA	TAATACGACTCACTATAGGCAAACAGGAACCGCAGGAATTGCGGGGGT AGCTTGGTAAACAAGAGAAACCTCTGCCGGCAAAGAAATAAGCCGG TAAGTATGCTCTGTTCCCAAGAATCTCGTGACTTTAGTCATGAGAGTT TCAATCTTCTGGATTGTTGTGGCCGGCATGGTCCCAGCCTCCTCGCTGGCG CCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC

T7 promoter; reRNA scaffold; guide sequence (for Target 1); HDV ribozyme

Table S11. Sequences of all the oligonucleotide substrates used in DNA cleavage assays (related to Methods)

Internal ID	Description	Sequence (5'-3')	Figures
dZZ01	Ymu1 DNA substrate 0-16 bp corresponding to target 1; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATTCTTCTGGATTG TTGTAAGCAGCATTTGAGC AAAAATCT	Fig. 2.2.3C, 2.2.3.F, 2.5.F, 2.5.K, 2.5.M-N, 2.5.Q-R
dZZ02	Ymu1 DNA substrate 0-16 bp corresponding to target 1; TS; FAM	/56-FAMN/AGATTTTTTGCTCAAA TGCTGCTTACAACAATCCA GAAGAATCAATCATTATGA CCTCCGG	Fig. 2.2.3C, 2.2.3.F, 2.5.F, 2.5.K, 2.5.M-N, 2.5.Q-R
dZZ013	Ymu1 DNA substrate 0-16 bp corresponding to target 2; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATAAGGCAAATTCG CCGCAAGCAGCATTTGAG CAAAAATCT	Fig. 2.2.3G-H, 2.5.3O-P
dZZ014	Ymu1 DNA substrate 0-16 bp corresponding to target 2; TS; FAM	/56-FAMN/AGATTTTTTGCTCAAA TGCTGCTTGCGGCGAATTT GCCTTATCAATCATTATGA CCTCCGG	Fig. 2.2.3G-H, 2.5.3O-P
dZZ017	Ymu1 DNA substrate with the 4th base mismatched from the target 1 guide; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATTCTACTGGATTG TTGTAAGCAGCATTTGAGC AAAAATCT	Fig. 2.2.4B, 2.2.4F
dZZ018	Ymu1 DNA substrate with the 4th base mismatched from the target 1 guide; TS; Cy2	/56-FAMN/AGATTTTTTGCTCAAA TGCTGCTTACAACAATCCA GTAGAATCAATCATTATGA CCTCCGG	Fig. 2.2.4B, 2.2.4F

dZZ025	Ymu1 DNA substrate with the 7th base mismatched from the target 1 guide; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATTCTTCTCGATTG TTGTAAGCAGCATTGAGC AAAAATCT	Fig. 2.2.4C, 2.2.4G
dZZ026	Ymu1 DNA substrate with the 7th base mismatched from the target 1 guide; TS; Cy2	/5Cy5/CCGGAGGTCATAAT GATTGATTCTTCTCGATTG TTGTAAGCAGCATTGAGC AAAAATCT	Fig. 2.2.4C, 2.2.4G
dZZ033	Ymu1 DNA substrate with the 10th base mismatched from the target 1 guide; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATTCTTCTGGAATG TTGTAAGCAGCATTGAGC AAAAATCT	Fig. 2.2.4D, 2.2.4H
dZZ034	Ymu1 DNA substrate with the 10th base mismatched from the target 1 guide; TS; Cy2	/56- FAMN/AGATTTTTGCTCAAA TGCTGCTTACAACATTCCA GAAGAATCAATCATTATGA CCTCCGG	Fig. 2.2.4D, 2.2.4H
dZZ029	Ymu1 DNA substrate with the 14th base mismatched from the target 1 guide; NTS; Cy5	/5Cy5/CCGGAGGTCATAAT GATTGATTCTTCTGGATTG TAGTAAGCAGCATTGAGC AAAAATCT	Fig. 2.2.4E, 2.2.4I

dZZ030	Ymu1 DNA substrate with the 14th base mismatched from the target 1 guide; TS; Cy2	/56- FAMN/AGATTTTTTGCTCAAA TGCTGCTTACTACAATCCA GAAGAATCAATCATTATGA CCTCCGG	Fig. 2.2.4E, 2.2.4I
--------	---	--	------------------------

Table S12. AF3 query sequences for the WT Ymu1 ternary complex (related to Methods)

Biomolecule	Sequences
Protein	MLQHKAYEYRIYPDKKQETLIAKTIGSSRYVYNHFLELWNKEYEE TGKGLTYACSKLLTKLKRDPETVWLCEVDKFSLQNSLRNLSDA FSRFFKGQNEHPQFKSKSPRQSYTTQYTNNNIAVSGNCLKLPK LGLVKFADSREMKGRILNATVRRKSSGKFFVSILCEEEICELPKTD SSVGIDLGIIDFAVMSDGSRHDNNHFTRQMEERLRREQRKLARR ALAAEKRGISLSEARNYQKQRRKVARLYEKVANQRKEYLNKLST EIVKNHDIICIEDLNVKGMMRNHKLAKSISDVSWTSLVSKLQYKAS WYGKEVIRISRWFPSSQICSECGHKDRKKPLHVREWTCPVCHA HHRDVAARNILAEGLRIRALTPGS
RNA (5'-3')	CCGCACCCCCUGCGGGGGUAGCUUGGUAAACAAGAGAAACC UCUGCCGGCAAAGAAUAAGCCGGUAAGUAUGCUCUGUUC CAAGAAUCUCGUGACUUUAGUCAUGAGAGGGGGGAUCUUCU GGAUUGUU
DNA (5'-3')	AACAATCCAGAAGAATCAATCA
DNA (5'-3')	TGATTGAT

2.8 Declaration of interests

Trevor Weiss, Maris Kamalu, Honglue Shi, Jennifer A. Doudna, and Steven E. Jacobsen have filed a patent covering aspects of this work. The Regents of the University of California have patents issued and pending for CRISPR technologies on which the authors are inventors. Jennifer A. Doudna is a cofounder of Azalea Therapeutics, Caribou Biosciences, Editas Medicine, Evercrisp, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. Jennifer A. Doudna is a scientific advisory board member at Evercrisp, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, The Column Group, and Inari. Jennifer A. Doudna is Chief Science Advisor to Sixth Street; is a Director at Johnson & Johnson, Altos, and Tempus; and has a

research project sponsored by Apple Tree Partners. Steven E. Jacobsen. is a cofounder and consultant for Inari Agriculture and a consultant for Terrana Biosciences, Invaio Sciences, Sail Biomedicines, Six Street and Zymo Research. David F. Savage is a cofounder and member of the scientific advisory board at Scribe Therapeutics.

Chapter 3: Engineering TnpB proteins enables them to bypass intrinsic barriers in DNA target engagement

3.1 Introduction

CRISPR-Cas 12 proteins engage with DNA through initial searching of the DNA target, then recognition of the PAM sequence and hybridization of its RNA guide, and subsequent cleavage mediated by the RuvC domain (Singh et. al. 2018, Jeon et. al. 2018). Given the similar domain architecture of TnpB, the overarching hypothesis of TnpB cleavage is that it occurs in the same manner as CRISPR-Cas (Karvelis et. al. 2021, Sasnauskas et. al. 2023, Nakagawa et. al. 2023). However, profiling of TnpB cleavage in bacterial and mammalian cells showcases a different cleavage profile compared to the CRISPR-Cas proteins (Xiang et. al. 2024). This differential in cleavage activity has motivated engineering efforts to improve TnpB activity (Marquart et. al. 2024, Skopintsev et. al. 2025, Thornton et. al. 2026). Bulk biochemical assays showed that WT Ymu1 cleaved DNA at a markedly slower rate than enhanced variants. Such kinetic limitations sometimes arise during DNA unwinding and *R*-loop formation in CRISPR-Cas enzymes (Aris et. al. 2025, Gong et. al. 2018, Eggers et. al. 2024, Shi et. al. 2025). Although biophysical models of *R*-loop progression and intermediate states have been established for Cas enzymes (Aris et. al. 2025, Shi et. al. 2025, Ivanov et. al. 2020), analogous information remains unexplored for TnpB. The mechanistic basis behind the low activity of natural TnpB proteins is poorly understood, as is how engineering the variants can enable TnpB proteins to overcome barriers in cleavage.

Here, we address this gap by identifying molecular determinants that limit TnpB efficiency. We selected Ymu1 TnpB as a representative protein, given its cleavage activity in diverse cellular environments (Xiang et. al. 2024) and its biochemical tractability. Combining bulk biochemical assays and single-molecule gold rotor-bead tracking (AuRBT), we found that *Yungiiibacter multivorans* TnpB (Ymu1 TnpB) unwinds DNA through a stepwise pathway comprising two distinct transitions: an initial closed-to-intermediate transition and a subsequent intermediate-to-open completion, forming a fully unwound state. Mutational analysis revealed that these transitions are modulated by the WED domain and an α -helix flanking the lid subdomain (hereafter, lid-adjacent helix): a WED-domain mutation promoted early transition, whereas mutations in the lid-adjacent helix stabilize more extensive unwinding and favour progression to the fully unwound state. Together, these perturbations differentially stabilize specific unwound conformations, enabling enhanced DNA cleavage *in vitro* and editing efficiency in plants. These findings establish a stepwise DNA unwinding model for the conformational control in TnpB and provide a mechanistic framework for rational engineering hypercompact, high-performance RNA-guided nucleases.

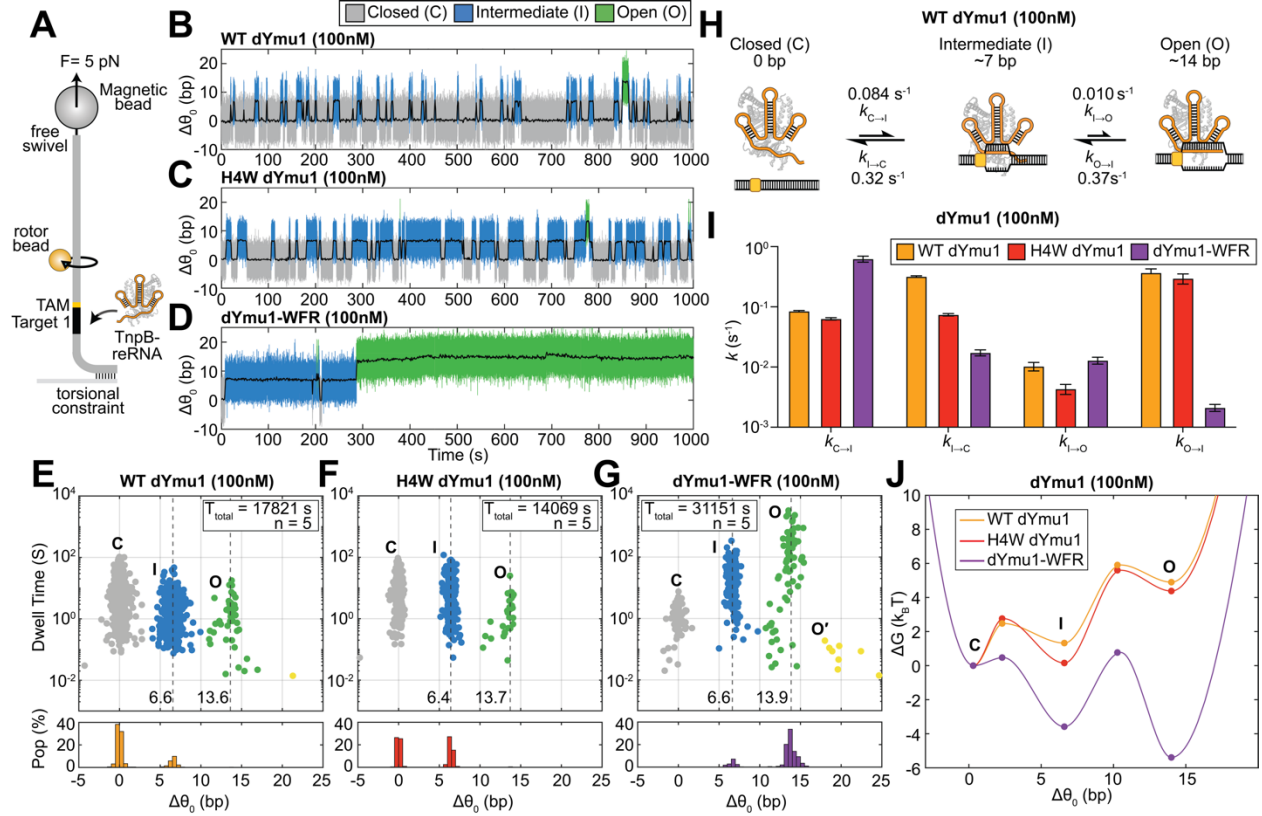
3.2 Results

3.2.1. AuRBT revealed step-specific enhancement of DNA unwinding by Ymu1 mutations

To directly resolve DNA unwinding dynamics and identify potential kinetic bottlenecks in Ymu1 TnpB, we employed gold rotor-bead tracking (AuRBT), a single-molecule torque spectroscopy method which directly reports DNA unwinding in real time with base-pair resolution (Lebel et. al. 2014). In this assay, a dsDNA tether containing a single TAM adjacent to Target 1 is coupled to a gold nanoparticle, whose rotation reports changes in DNA twist ($\Delta\theta_0$) as the duplex unwinds and rewinds (Fig. 3.2.1A and Fig. 3.5.1A; Methods). To prevent cleavage and enable repeated sampling of unwinding transitions, all measurements were performed using catalytically dead WT Ymu1 (WT dYmu1, E279A) and the corresponding H4W dYmu1 and dYmu1-WFR variants (Fig. 3.5.1B-D). We examined unwinding dynamics in a torsionally relaxed DNA tether in the presence of 100 nM RNP (Fig. 3.2.1A) and observed equilibrium fluctuations in twist resulting from reversible RNP binding and *R*-loop formation. To resolve unwinding intermediates and quantify transition dynamics, we applied automated change-point detection (Wiggins 2015, Lamont and Wiggins 2016) to continuous trajectories, enabling identification of unwinding states as well as their dwell times and interconversion rates. All three proteins tested produced transitions between three well-separated states: a closed duplex state (C), a partially unwound ~ 7 bp intermediate state (I), and a ~ 14 bp fully unwound open state (O) (Fig. 3.2.1B-G). The extent of the O state is consistent with the RNA-DNA heteroduplex lengths (~ 12 - 16 bp) required for efficient TnpB cleavage by a related orthologue (Dra2 TnpB) (Sasnauskas et. al. 2023, Nakagawa et. al. 2023). The ~ 7 bp I state aligns with prior observations (Sasnauskas et. al. 2023) that Dra2 TnpB can stably bind targets containing TAM-distal mismatches beyond position ~ 8 , indicating a binding-competent but cleavage-incompetent *R*-loop state, reminiscent of the “seed” intermediates observed in other CRISPR-Cas enzymes (Aris et. al. 2025, Shi et. al. 2025, Ivanov et. al. 2020, Szczelkun et. al. 2014, Madariaga-Marcos et. al. 2025). Thus, Ymu1 unwinding involves a C $\rightarrow$ I initiation step that can be followed by an I $\rightarrow$ O completion step. Introducing mismatches within the first 1-4 nt of the RNA-DNA heteroduplex abolished the observed unwinding states (3.5.1E-H), as expected for selective on-target formation of the I state.

While displaying the same three-state behavior, the Ymu1 variants have markedly different state occupancies and transition rates. At 100 nM RNP, WT dYmu1 resided predominantly in the C state, with frequent reversions from the I state and only rare, short-lived excursions to the O state (Fig. 3.2.2B, E, H, and I). The H4W substitution, positioned near the first RNA-DNA base pair, stabilized the intermediate state (Fig. 3.2.2C, F) by suppressing collapse from I to C (Fig. 3.2.2I and Table 3.7.S1). In contrast, the triple substitution dYmu1-WFR reshaped the unwinding kinetics at both C $\leftrightarrow$ I and I $\leftrightarrow$ O stages (Fig. 3.2.1D, G, and I). dYmu1-WFR strongly suppressed collapse back to C ($k_{I\rightarrow C}$) (Fig. 3.2.1I and Table 3.7.S1), which increases the probability of progression to the O state. In addition, dYmu1-WFR dramatically reduced reversal from the O state ($k_{O\rightarrow I}$), biasing the equilibrium population toward full *R*-loop formation (Fig. 3.2.1I and Table 3.7.S1). Together, these kinetic effects enable efficient and persistent formation of the fully unwound state under single-molecule conditions (Fig. 3.2.1J). Collectively, these results

show that H4W dYmu1 primarily enhances early unwinding by stabilizing the I state, whereas the combined dYmu1-WFR substitutions promote both initiation and propagation of the *R*-loop, providing a kinetic basis for the enhanced biochemical and cellular activities of Ymu1-WFR.



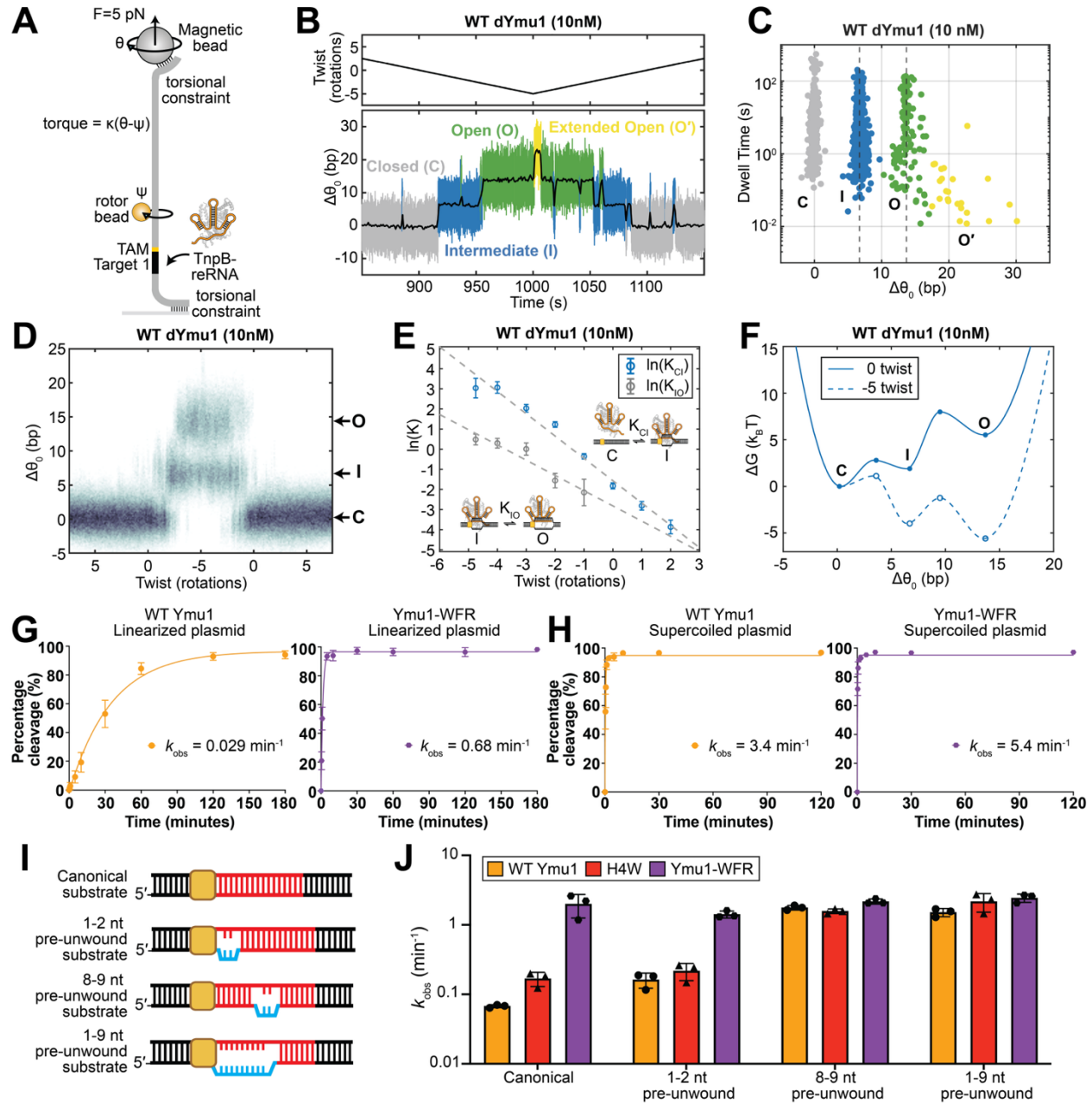
3.2.1. Mutations in distinct domains of Ymu1 TnpB enhance sequential steps of DNA unwinding. (A) Schematic AuRBT experiment for measuring *R*-loop dynamics at bp resolution. A single, torsionally relaxed ~5-kb DNA molecule containing a single TAM (5'-TTGAT-3') flanking the Target 1 is stretched between a coverslip and a magnetic bead under 5 pN tension. An 80-nm gold rotor bead (orange) is attached above the target site. In this geometry, changes in the mean rotor-bead angle report directly on changes in equilibrium twist ($\Delta\theta_0$, expressed in units of base-pairs unwound by assuming unwinding of B-form DNA). dYmu1 TnpB RNP (Fig. 3.5.1B-C), showing no detectable dsDNA cleavage in the AuRBT buffer (Fig. 3.5.1D), was introduced at 100 nM. (B-D) Representative trajectories showing $\Delta\theta_0$ (bp) over time (sec) for (B) WT Ymu1, (C) H4W, (D) Ymu1-WFR. Distinct states are scored using automated change-point detection (see Methods), including closed (C, gray), intermediate (I, blue), and open (O, green). Low-pass-filtered traces (1 Hz) are shown in black. (E-G) (Top) Scatter plots of state dwell times (sec) versus $\Delta\theta_0$ (bp) for all merged states (see Methods) under equilibrium conditions. (Bottom) Histogram of occupation probability, representing the lifetime-weighted population at each corresponding $\Delta\theta_0$ with a bin size of 0.5 bp. Distinct unwinding states include closed (C, gray), intermediate (I, blue), and open (O, green). Some short-lived events are also seen with extended levels of opening (O', yellow). The total collection time (T_{total}) and the number of DNA tethers analyzed (n) are reported in the figure legend. (E) WT Ymu1. (F) H4W. (G) Ymu1-WFR. The unmerged scatter plots

are shown in Fig. 3.5.1I-K. (H) Kinetic models summarizing transition rates between the closed, intermediate, and open states of WT dYmu1 at 100 nM RNP. (I) Bar graph of RBT transition kinetics of WT dYmu1, H4W dYmu1, and dYmu1-WFR at 100 nM RNP (Table 3.7.S1). Errors were calculated assuming Poisson statistics of state transitions. (J) Free-energy landscapes for DNA unwinding, overlaid for WT (orange), H4W (red), and Ymu1-WFR (purple). AuRBT trace statistics, including the number of tethers analyzed, total tracking time, and number of detected transitions, are summarized in Table 3.7.S2.

3.2.2. Negative supercoiling stabilizes unwound states in WT Ymu1

Our equilibrium measurements indicate that WT dYmu1 rarely stabilizes a fully unwound *R*-loop on torsionally relaxed DNA, raising the question of how this enzyme achieves productive activity in its native bacterial context. In bacterial genomes, DNA is pervasively negatively supercoiled (Sinden et. al. 1980, Wang 1996, Wang 2002, Finzi and Olson 2016). Because negative supercoiling is known to modulate strand separation and *R*-loop formation for diverse RNA-guided nucleases (Aris et. al. 2025, Ivanov et. al. 2020, Szczelkun et. al. 2014, Rutkauskas et. al. 2022, Newton et. al. 2023), we tested whether DNA supercoiling alone could reshape the DNA unwinding landscape of WT Ymu1. We performed non-equilibrium AuRBT experiments in which the DNA tether was repeatedly ramped between positively and negatively supercoiled conditions (Fig. 3.2.2A; Methods) (Aris et. al. 2025, Ivanov et. al. 2020), driving *R*-loop formation and collapse. As in prior work, changes in measured torque (Fig. 3.5.2C) were converted to changes in equilibrium twist ($\Delta\theta_0$) (Fig. 3.2.2B and Fig. 3.5.2D), and transition rates between scored *R*-loop states (Fig. 3.2.2C-D and Fig. 3.5.2E) were analyzed as a function of imposed twist (Fig. 3.5.2F-G) and used to determine supercoiling-dependent equilibrium constants (Fig. 3.2.2E) and construct free energy landscapes (Fig. 3.2.2F).

The locations of metastable *R*-loop states observed during twist-ramping experiments closely matched those determined in Fig. 3.2.1H, and the O state remained poorly accessible under zero imposed twist (Fig. 3.2.2B-F). As negative supercoiling was progressively applied, both the C $\leftrightarrow$ I initiation step and the I $\leftrightarrow$ O propagation step shifted toward unwound states (Fig. 3.2.2B-D and Fig. 3.5.2C-E). At approximately -5 Twist (~1% negative supercoiling), both the I and O states are substantially stabilized (Fig. 3.2.2D-F, Fig. 3.5.2F-G, and Table 3.7.S3), resembling the persistent unwound states in Ymu1-WFR. Under these conditions, we occasionally observed short-lived excursions beyond the ~14 bp O state (Fig. 3.2.2B-C and Fig. 3.5.2D-E), consistent with suggested transient downstream duplex opening required for the cleavage mechanism of TnpB (Nakagawa et. al. 2023). As expected, negative supercoiling can partially compensate for intrinsic unwinding limitations in WT dYmu1 by stabilizing unwound conformations. Together with the variant comparisons, these measurements illustrate how DNA mechanics and protein mutations act on the same energetic landscape for *R*-loop formation.



3.2.2. Negatively supercoiled or pre-unwound DNA enhances Ymu1 activity. (A) Schematic of the non-equilibrium, torque-driven AuRBT assay. Here the DNA molecule is torsionally constrained. Twist is imposed by rotating the magnets (θ), and torque is inferred from the difference in magnet position (θ) and rotor-bead angle (ψ) using $\tau = \kappa(\theta - \psi)$. A separately prepared batch of dYmu1 TnpB RNP (Fig. 3.5.2A-B) was used at 10 nM for the torque-driven measurements. (B) Representative torque-driven unwinding trajectory of WT dYmu1. (Top) Imposed twist (rotations) and (Bottom) $\Delta\theta_0$ (in units of base-pairs unwound; see Methods) over time (sec). Distinct states are scored using the automated change-point detection algorithm Steppi (Wiggins 2015, Lamon and Wiggins 2016), including closed (C, gray), intermediate (I, blue), open (O, green), and extended open (O', yellow) states. Low-pass-filtered traces (1 Hz) are shown in black. This trace is a portion of the trajectory shown in Fig. 3.5.2D. (C) Scatter plot of dwell times (sec) versus

$\Delta\theta_0$ (bp) after merging (Methods) for torque-driven measurements of WT dYmu1. Corresponding unmerged dwells are shown in Fig. 3.5.2E. Distinct unwinding states include closed (C, gray), intermediate (I, blue), open (O, green), and an extended open (O', yellow). (D) Heat map showing the distribution of $\Delta\theta_0$ (bp) as a function of imposed twist (rotations). (E) Plot of $\ln(K_{C \rightarrow I})$ and $\ln(K_{I \rightarrow O})$ versus imposed twist (rotations), with linear fits. Data points were calculated from the ratio of the transition rate constants, which are plotted in Fig. 3.5.2F-G and listed in Table 3.7.S3. Errors were calculated assuming Poisson statistics of state transitions. Fit parameters are summarized in Table 3.7.S4. AuRBT trace statistics are summarized in Table 3.7.S5. (F) Free-energy landscapes for DNA unwinding. Solid line: 0 rotations; dashed line: -5 rotations. Landscapes were constructed using linear fits to $\ln(K)$ and $\ln(k)$ data (Table 3.7.S4). (G-H) Cleavage of linearized (G) and negatively supercoiled (H) 2.2 kb plasmid substrates by WT Ymu1 and Ymu1-WFR. Percentage of cleaved substrate (y-axis) is plotted over time (x-axis). Each point represents mean $\pm$ SD ($n = 3$ independent reactions); k_{obs} values reflect the mean from three independent mono-exponential fits on each reaction time course. See also Table 3.7.S6. (I) Schematics of partially pre-unwound DNA substrates used for kinetic perturbation. (J) Bar graph of apparent cleavage rates (k_{obs}) for WT Ymu1, H4W, and Ymu1-WFR across substrates with different initial unwinding (mean $\pm$ SD) (Tables 3.7.S7). Shown here are NTS cleavage. TS cleavage profiles are provided in Fig. 3.5.2H.

To test whether the supercoiling dependence observed in single-molecule measurements extends to the overall cleavage activity of enzymes, we compared WT Ymu1 and Ymu1-WFR cleavage on 2.2 kb negatively supercoiled versus linearized plasmid DNA substrates. On the linearized substrate, Ymu1-WFR cleaved 23-fold faster than WT Ymu1 ($k_{obs} = 0.68$ and 0.029 min^{-1} , respectively; Fig. 3.2.2G), consistent with the magnitude of enhancement observed on short oligonucleotide substrates (Fig. 2.2.3C and 3F). In contrast, on the negatively supercoiled plasmid, Ymu1-WFR and WT Ymu1 cleaved at comparable rates, differing by only 1.6-fold ($k_{obs} = 5.4$ and 3.4 min^{-1} , respectively; Fig. 3.2.2H). Thus, WT Ymu1 benefits substantially more from negative supercoiling (~ 117 -fold) than Ymu1-WFR (~ 8 -fold), largely eliminating the rate difference between the two variants on supercoiled DNA. This is consistent with the single-molecule observation that both negative supercoiling (Fig. 3.2.2F) and WFR mutations (Fig. 3.2.1J) are relieving the unwinding limitations that constrain WT activity on torsionally relaxed DNA.

3.2.3. Stepwise DNA substrate unwinding limits WT Ymu1 cleavage

To further challenge our model for Ymu1 function and investigate the impact of *R*-loop state stabilization in the context of the active endonuclease, we conducted cleavage assays in the presence of site-specific DNA-DNA mismatches designed to favour *R*-loop formation by reducing the energetic cost of duplex unwinding. Unlike supercoiling, which globally favours unwinding, site-specific mismatches specifically stabilize *R*-loops containing the targeted nucleotide positions. We used substrates containing localized duplex pre-unwinding to preferentially assist either $C \leftrightarrow I$ (1-2 nt near the TAM), $I \leftrightarrow O$ (8-9 nt), or both steps (1-9 nt) of *R*-loop formation (Fig. 3.2.2I).

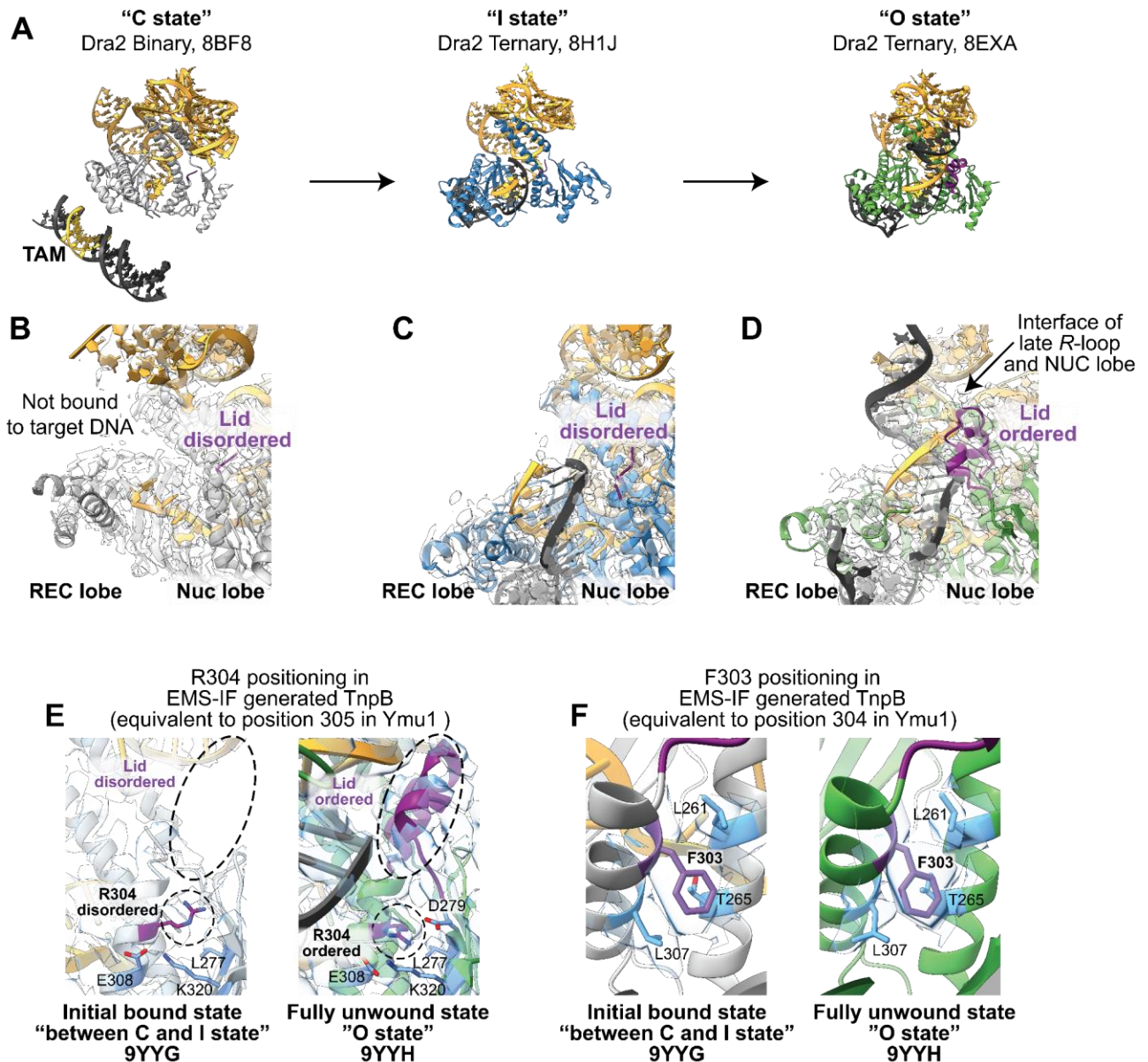
For WT Ymu1, all three pre-unwound substrates produced improvements in cleavage rates, as expected, with substantially larger effects for substrates that stabilized

the fully unwound O state (Fig. 3.2.2J and Fig. 3.5.2H). For H4W, the 1-2 nt pre-unwound substrate yielded only a small additional increase (Fig. 3.2.2J and Fig. 3.5.2H), which may be explained by this substitution's stabilization of the I state, such that its formation is not strongly limiting under the conditions of the assay. By contrast, 8-9 nt or 1-9 nt pre-unwound substrates produced large enhancements (Fig. 3.2.2J and Fig. 3.5.2H), as expected since with H4W, the transition from the I to the O state remains highly inefficient, and can be favoured by reducing the energetic cost of the I to O transition. Ymu1-WFR displayed uniformly high cleavage activity across all substrates (Fig. 3.2.2J and Fig. 3.5.2H). This suggests that once the $C \leftrightarrow I$ and $I \leftrightarrow O$ steps have both been strongly enhanced in WFR, stabilization of unwinding transitions is ineffective in accelerating overall cleavage under these conditions. Together, these measurements show that H4W selectively stabilizes the I state, whereas WFR produces major stabilization of both the I and O states, providing a mechanistic explanation for the dramatically enhanced genome-editing efficiency of Ymu1-WFR.

3.2.4. Structural analysis of Ymu1 TnpB mutations

To understand how mutations in Ymu1-WFR reshape the DNA unwinding landscape, we used AlphaFold3 (AF3) modeling of the Ymu1-reRNA-DNA ternary complex and compared them with available cryo-EM structures of related TnpB homologues (Konnin et. al. 2017, Meers et. al. 2023, Nakagawa et. al. 2024) (see Methods). Together, these analyses suggest two classes of effects: stabilization of early RNA-DNA engagement and modulation of protein conformational transitions that favour later, more extensively unwound states.

We propose a structural framework relating cryo-EM conformations of a related orthologue (Dra2) to the unwinding intermediates resolved by AuRBT. The Dra2 protein-RNA binary complex (Sasnauskas et. al. 2023) is most relevant to the closed (C) state (Fig. 3.2.3A-B). The Dra2 ternary structure reported by Nakagawa et al. 2023 exhibits cryo-EM density consistent with limited ordering of the TAM-distal RNA-DNA heteroduplex, and the strongly resolved proximal heteroduplex principally engages the REC lobe (Fig. 3.2.3A, C). We propose that the partially unwound intermediate (I) state observed in our experiments, in which only a short RNA-DNA heteroduplex is formed, may adopt a protein-RNA-DNA architecture resembling this structure. By contrast, the ternary structure reported by Sasnauskas et al. 2023, which resolves ~15 bp of ordered heteroduplex, may correspond to the fully unwound open (O) state (Fig. 3.2.3A, D). In this putative O-state architecture, the extended RNA-DNA heteroduplex engages both the REC and NUC lobes, including the bridge helix and lid domain (Fig. 3.2.3D). Across these structures, the lid domain appears disordered in the C state and becomes progressively ordered. We propose the lid may have transient interactions with the partially formed RNA-DNA heteroduplex in the I state and then becomes well-ordered and engages the heteroduplex in the fully unwound O state, suggesting coupling between lid ordering and late-stage R-loop formation as observed in related Cas12 enzymes (Stella et. al. 2018, Pausch et. al. 2021, Guan et. al. 2025). AF3 modeling of the Ymu1 ternary complex converges on a similar fully unwound O-state architecture (Methods), providing a structural reference for interpreting mutational effects discussed shortly.

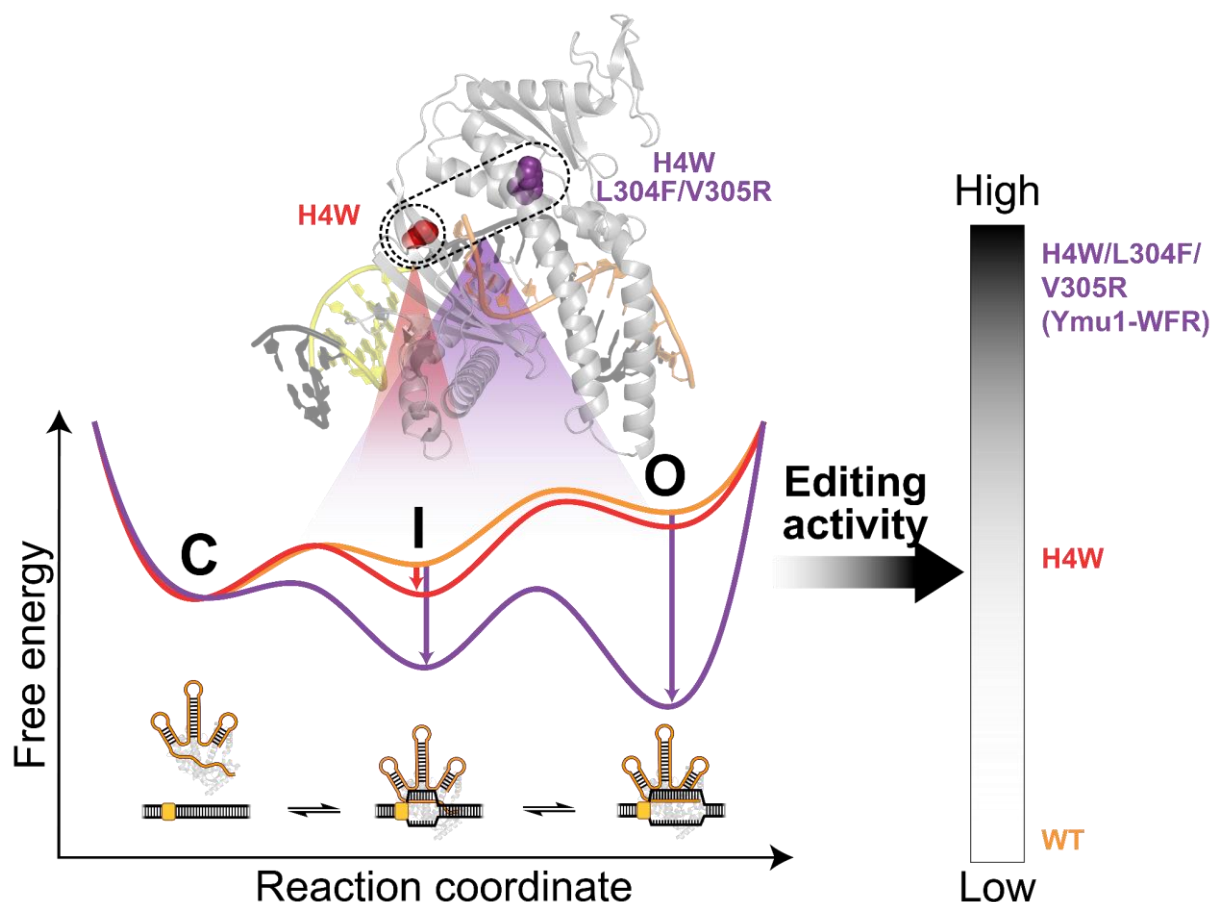


3.2.3. Analyses supporting mutational effects at positions 4, 305, and 304 (related to Fig. 6).

(A) Comparison of cryo-EM structures of Dra2 TnpB in distinct conformational states. The closed (“C”) state is defined by intact duplex DNA and corresponds to a DNA-free or binary-like protein complex conformation in solution. The Dra2 binary complex lacking target DNA (Sasnauskas et. al. 2023) (PDB: 8BF8) is therefore shown as a reference for the protein conformation associated with the C state. The ternary state reported by Nakagawa et al. 2023 (PDB: 8H1J) is proposed to correspond to the intermediate (“I”) state, and the ternary state reported by Sasnauskas et al. 2023 (PDB: 8EXA) is proposed to correspond to the fully-formed open (“O”) state. **(B-D)** Experimentally resolved heteroduplex and lid domain structures across conformational states of Dra2 TnpB. **(B)** In 8BF8 (“C state”), the lid domain is disordered and DNA is not unwound. **(C)** In 8H1J (“I state”), the lid remains mostly unstructured, possibly allowing transient interactions with a partially formed DNA heteroduplex, which mostly binds the REC lobe. 6-8 bp of

heteroduplex are most strongly resolved in the density; 8 bp are shown in the figure, fewer than the 12 bp included in the 8H1J model (Nakagawa et. al. 2023). **(D)** In 8EXA (“O state”), the lid domain becomes ordered and facilitates binding of TnpB with the RNA-DNA heteroduplex along its whole length, consistent with the fully unwound O state. The heteroduplex is stabilized by both the REC and the NUC lobes. **(E-F)** Local structural environments of residues in cryo-EM structures of an ESM inverse-folding (ESM-IF) generated TnpB homologous to the Ymu1 WFR substitutions (Skopintsev et. al. 2025). **(E)** Residue R304 in the ESM-IF-generated TnpB (homologous to Ymu1 V305R substitution) does not form cryo-EM density-supported contacts in the initial bound state (PDB: 9YYG), whereas it is associated with a strong density in the fully unwound state (PDB: 9YYH). **(F)** Residue F303 in the ESM-IF-generated TnpB (homologous to Ymu1 substitution L304F) is structured in both early and late *R*-loop states. For each panel, the left view shows an initial bound state preceding the *R*-loop formation (PDB: 9YYG), and the right view shows the fully unwound state (PDB: 9YYH). Density and model views illustrate state-dependent ordering of the lid region and associated residues. Cryo-EM maps in **(B-D)** were contoured at 7.0, 0.5, 8.0, respectively, whereas **(E-F)** were contoured at 0.1 (dataset-dependent map units).

Recent cryo-EM structures of an artificial TnpB designed using evolutionary information and ESM inverse folding (ESM-IF), in which F303 and R304 occupy positions homologous to Ymu1 L304 and V305 in a similar residues context, provide additional structural explanation to these residues as potential stabilizers of protein occupancy on DNA (Skopintsev et. al. 2025). Across reported conformations, both the lid domain and the Ymu1’s V305-homologous residue R304 are unresolved in the initial TAM-bound structure (which may correspond to a state intermediate between C and I) but become ordered in the fully unwound state (O) (Fig. 3.2.3E). Based on this experimental observation, we suggest that Ymu1’s V305R similarly supports formation of the RuvC and lid domain conformations in the late unwound states (Fig. 3.2.3E). In contrast, F303 is structured in both early and late states (Fig. 3.2.3F), and may contribute to overall RuvC fold stabilization across the unwinding pathway. The position of the H4W residue nearing the TAM sequence, as well as its function in facilitating the initial opening of DNA (Fig. 3.2.2I-J) makes it likely that this residue is involved in the initial C-to-I transition and thus, we hypothesize that the H4W residues and the L304F/V305R residues act in a stepwise manner in improving TnpB cleavage of the DNA target (Fig. 3.2.4).



3.2.4. Mechanistic model for how Ymu1 TnpB mutations reshape the unwinding energy landscape to enhance genome editing. The schematic model illustrates how the three mutations in Ymu1-WFR modulate DNA-unwinding along a three-step pathway that includes closed (C), intermediate (I) and open (O) states. H4W stabilizes the I state, whereas L304F and V305R, positioned in the lid-adjacent helix, further stabilize both the I and O states. Together, these effects yield progressively higher genome-editing efficiencies from WT→H4W→WFR. Our model suggests that tuning local energetic checkpoints for R-loop formation provides a generalizable strategy for rational engineering of compact RNA-guided nucleases with enhanced activity.

3.3. Discussion

RNA-guided nucleases recognize DNA through the energetics of base-pair exchange, yet the quantitative architecture of this process has remained poorly defined for compact enzymes such as TnpB. Our work reveals that TnpB traverses an unwinding pathway that includes formation of a partially unwound DNA intermediate state followed by a fully unwound open state. Under relaxed DNA conditions, both states are prone to collapse, and reversal is favoured over progression from the intermediate to the open state. Consequently, forward progress into the fully unwound state is intrinsically limited, and TnpB-catalyzed DNA cleavage occurs only during rare, transient visits to this configuration. Thus, the stability and dynamics of conformational states upstream of strand scission constrains TnpB activity. These conformational dynamics are sensitive to protein mutagenesis and DNA supercoiling and may be further modulated by genomic

context and target sequence. While stepwise *R*-loop formation has been described for Cas12 enzymes (Aris et al., 2025) that share structural homology with TnpB, the quantitative measurements reported here reveal a substantially more reversible unwinding landscape for the wild-type enzyme. The reversibility of DNA substrate unwinding combined with strict TAM dependence naturally bias TnpB toward reversible target sampling rather than favouring the catalytically competent *R*-loop conformation. Such properties are consistent with longstanding hypotheses that transposon-associated nucleases need to balance mobility with host genome stability (Meers et. al. 2023, Thornton et. al. 2025). When deployed in eukaryotic genomes for genome editing, the local DNA topology may further modulate this balance: bulk nuclear DNA is torsionally relaxed (Corless and Gilbert et al., 2016), but localized negative supercoiling arising from transcription could partially stabilize the O state for WT TnpB at some loci, which could contribute to the target-site-dependent variation in editing efficiency.

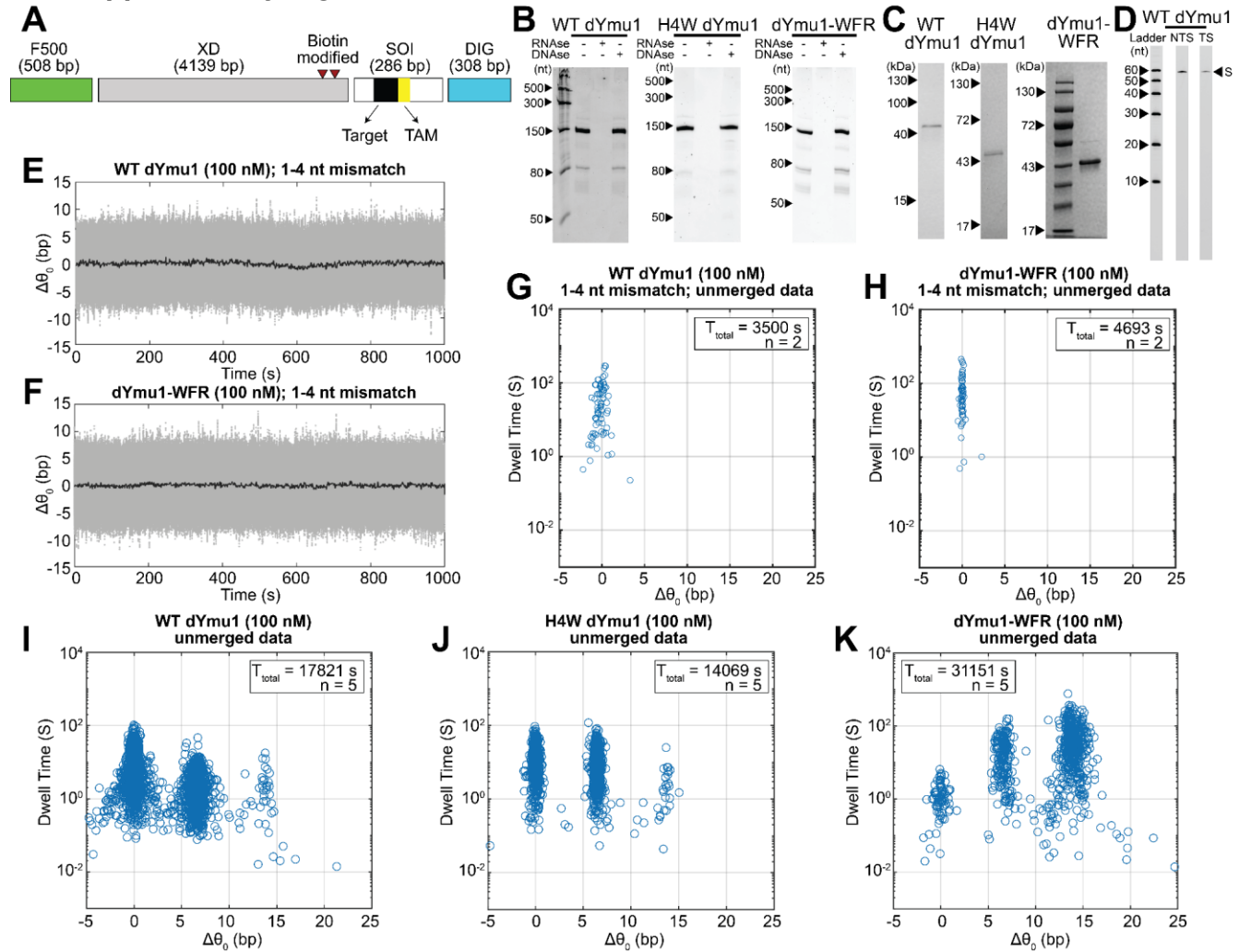
By defining the quantitative contributions of each DNA unwinding step, our work also illuminates a modular route for enzyme improvement. The three activating mutations identified here act at distinct energetic states: H4W facilitates *R*-loop initiation, whereas the addition of L304F and V305R stabilizes more extensive DNA unwinding. This separation of effects demonstrates that compact nucleases can be modulated through a small number of targeted point mutations rather than extensive redesign of the protein scaffold. The same principle may explain why mutations at homologous positions enhance Dra2 TnpB activity (Thornton et al. 2026), suggesting that similar energetic bottlenecks exist across TnpB orthologues, even though the optimal substitution at each position can be orthologue specific. Consistent with this framework, these activating positions have been independently shown to enhance editing across multiple TnpB orthologues in both human cells and plants (Thornton et al., 2026; Nagalakshmi et al., 2026; Weiss et al., 2026).

Taken together, these findings position TnpB as a mechanistic reference point for understanding how RNA-guided systems diversified from compact ancestors. The reversible, stepwise unwinding pathway we describe provides a plausible framework by which CRISPR-associated nucleases could have evolved faster DNA cleavage kinetics through stabilization of specific unwound intermediates (Aris et. al. 2025, Shi et. al. 2025, Ivanov et. al. 2020). In addition, stabilization of the fully unwound state may also be critical for emerging applications of TnpB that rely on durable DNA binding without cleavage, such as transcriptional modulation³¹ and base editing (Luo et. al. 2025). More broadly, comparative unwinding-landscape measurements across TnpB orthologues and related RNA-guided enzymes including IscB and Fanzors, where activity-enhancing variants are already emerging (Yan et. al. 2024, Xue et. al. 2024, Wei et. al. 2026, Zhao et. al. 2025), may enable further rational engineering of compact RNA-guided nucleases.

3.4. Limitations of the study

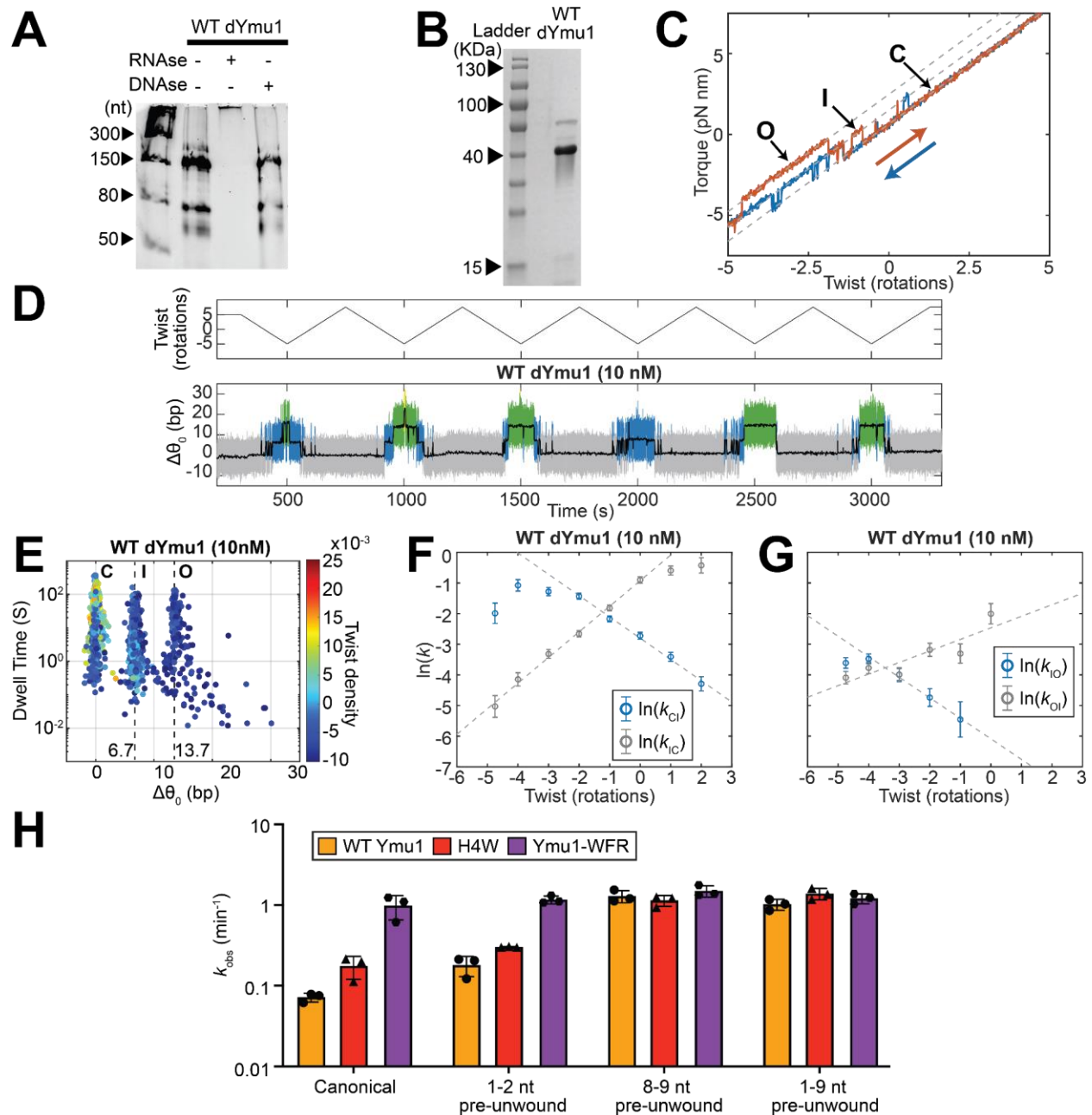
The mechanistic model developed here is based on single-molecule measurements of a single TnpB orthologue (Ymu1) using one target DNA sequence. Extending these measurements to additional target sequences, guide-target mismatches, supercoiling-dependent conditions, as well as to other Ymu1 variants and TnpB orthologues, would further test and generalize the stepwise unwinding framework.

3.5. Supplementary Figures



3.5.1. Controls and validation for equilibrium AuRBT experiments. **(A)** Schematic of DNA tether used in AuRBT. Tethers were assembled by ligating four PCR-generated fragments (Table 3.7.S12). The F500 and DIG segments contained fluorescein- and digoxigenin-modified dUTPs for bead and coverslip attachment, respectively. The transducer segment (XD) carried two biotin modifications for rotor-bead linkage. The sequence-of-interest (SOI) included a 5'-TTGAT-3' TAM followed by the Target 1 sequence for dYmu1 TnpB binding (Table 3.7.S13). The full tether sequence is provided in Table 3.7.S14. **(B)** Denaturing PAGE analysis of the reRNA homogeneity for WT dYmu1, H4W dYmu1, and dYmu1-WFR TnpB samples used in torsionally-relaxed equilibrium AuRBT experiments. Gels are shown under untreated, RNase-treated, and DNase-treated conditions. **(C)** SDS-PAGE analysis of the TnpB protein homogeneity for WT dYmu1, H4W dYmu1, and dYmu1-WFR samples used in torsionally-relaxed equilibrium AuRBT experiments. **(D)** Representative denaturing gel of Target 1 oligonucleotide dsDNA cleavage by reconstituted WT dYmu1 (E279A) RNP in AuRBT buffer conditions showing no detectable cleavage after 120 min. **(E-F)** Representative trajectories of $\Delta\theta_0$ (bp) over time (sec) using tethers containing DNA-RNA mismatches introduced into nt 1-4 of the target DNA sequence. **(E)** WT dYmu1 and **(F)** dYmu1-WFR.

Low-pass-filtered traces (1 Hz) are shown in black. **(G-H)** Scatter plots of dwell times (sec) versus $\Delta\theta_0$ (bp) for all unmerged Steppi-assigned dwells for experiments with 1-4 nt DNA-RNA mismatches. **(G)** WT dYmu1 and **(H)** dYmu1-WFR. The total collection time (T_{total}) and the number of DNA tethers analyzed (n) are reported in the figure legend. *R*-loop formation events are not observed, consistent with a requirement for initial seed matching. **(I-K)** Scatter plots of state dwell times (sec) versus $\Delta\theta_0$ (bp) for all unmerged Steppi-assigned states of **(I)** WT dYmu1, **(J)** H4W dYmu1, and **(K)** dYmu1-WFR. The total collection time (T_{total}) and the number of DNA tethers analyzed (n) are reported in the figure legend. Complete AuRBT trace statistics, including the numbers of detected transitions, are summarized in Table 3.7.S2.



3.5.2. Additional analyses for DNA unwinding, supercoiling, and mismatch cleavage experiments (A) Denaturing PAGE analysis of the reRNA homogeneity for the WT dYmu1 sample used in torque-driven AuRBT experiments. RNA gels are shown under untreated, RNase-treated, and DNase-treated conditions. (B) SDS-PAGE of purified dYmu1 TnpB protein used in the torque-driven AuRBT experiments. The WT dYmu1 RNP sample in 3.2.2 and Fig. 3.5.2 was prepared independently from that used in equilibrium AuRBT assays. (C) Torque vs. twist for a twist ramping cycle in the presence of WT dYmu1. Blue (unwinding) and orange (rewinding) arrows indicate the direction of ramping. (D) Representative twist ramping trace (a portion of this trace is shown in 3.2.2B). (Top) Imposed twist (rotations). (Bottom) $\Delta\theta_0$ (expressed in units of base-pairs unwound assuming changes arise from local DNA unwinding of B-DNA). Transitions between distinct states are scored using automated change-point detection followed by merging to prevent overscoring (Wiggins 2015, Lamon and Wiggins 2016). States are colour-coded as closed (C, gray), intermediate (I, blue), open (O, green), and extended open (O', yellow). Low-pass-filtered traces (1 Hz) are shown in black. (E) Scatter plot of state dwell times (sec) (y-axis) versus $\Delta\theta_0$ (bp) (x-axis), coloured by twist density at state onset. Twist density was calculated as imposed twist (in turns) divided by the relaxed linking number of the DNA tether ($Lk_0 = N/10.5$ where N is the length of the tether in bp). (F-G) Plots of $\ln(k)$ versus imposed twist (rot), for forward and reverse transition rates k_{ij} in s^{-1} of (F) $C \leftrightarrow I$ and (G) $I \leftrightarrow O$ transitions for WT dYmu1. Data points are listed in Table 3.7.S3 and linear fit parameters are summarized in Table 3.7.S4. AuRBT trace statistics, including the number of tethers analyzed, total tracking time, and number of detected transitions, are summarized in Table 3.7.S5. (H) Pre-unwound DNA cleavage assays showing apparent cleavage rates (k_{obs}) as a function of initial duplex opening of WT, H4W, and Ymu1-WFR. Shown are TS cleavage. NTS cleavage profiles are provided in Fig. 3.2.2J.

3.6. Methods

Rotor bead tracking

Reagents

AuRBT studies were conducted with catalytically inactive Ymu1 TnpB RNP (dYmu1, dH4W, and dYmu1-WFB) with E279A mutation. DNA tethers for torque-driven and torsionally relaxed assays were assembled by ligating restriction enzyme digested PCR products as previously described (Lebel et. al. 2014, Ivanov et. al. 2018, Ivanov et. al. 2020). Detailed information on the tether construction, including building blocks, can be found in Fig. 3.5.1A and Tables 3.7.S12-S14. Magnetic beads (ThermoFisher, Dynabeads MyOne Carboxylic Acid) were crosslinked with an antibody to Fluorescein/Oregon Green (ThermoFisher, A889) via EDC (ThermoFisher, 77149), as previously described (Gore et. al. 2006). These beads were prepared and stored at 4°C for up to two months.

Chambers

Flow chambers were constructed as previously described (Shi et. al. 2025), by sandwiching custom laser-cut Nescofilm channels between a hole-punched vinyl coverslip and glass coverslip spin coated with 0.1% nitrocellulose in isopentyl acetate (Ladd Research, 10800). On the day before experiments, streptavidin-coated gold

nanospheres (20 μ L original suspended volume per channel) with a nominal diameter of 80 nm (Cytodiagnostics, AC-80-04-15) were washed twice in Au wash buffer (50 mM Tris-HCl, pH 8.0; 0.05% TWEEN-20), centrifuging at 2000g for 5 minutes between washes. These were finally resuspended to their original suspended volume in blocking buffer (40 mM Tris-HCl, pH 8.0; 0.5 M NaCl; 0.2% TWEEN-20; 0.01% sodium azide; 5 mg/mL BSA). DNA tethers ($\sim$ 2 pM) and magnetic beads (1.5 μ L original suspended volume per channel) were added to the gold nanospheres in the blocking buffer and incubated overnight at 4°C on a rotator. Channels were also incubated overnight at 4°C with 12 μ g/mL anti-digoxigenin in the PBS buffer. On the day of experiments, channels were (1) incubated with blocking buffer supplemented with 0.25% w/v casein for 1 h, (2) incubated with DNA tethers and both magnetic and rotor beads for 1 h, and (3) washed with approximately 15 channel volumes of C12T buffer (20 mM Tris-HCl, pH 8.0; 150 mM KCl; 5 mM $MgCl_2$; 1 mM TCEP) supplemented with 0.2% Tween-20 (Sigma, P9416) and 0.2 mg/ml BSA (Invitrogen, 15561-020). All flow steps were controllably carried out with a syringe pump.

Microscopy

Experiments were conducted using a custom-built AuRBT microscope as previously described (Lebel et. al. 2014). Briefly, this comprised a modified Nikon Eclipse Ti-S inverted microscope, with evanescent excitation provided by an intensity-stabilized 845-nm laser diode (Lumics LU0845M200) directly coupled to a polarization-maintaining fiber. A half-wave plate was used to achieve s-polarization at the sample interface. The return beam was collected on a position-sensitive detector to provide a signal for focus stabilization. Scattered light was collected with a Nikon Apo TIRF objective (60 $\times$ /1.49 numerical aperture, oil) and imaged through an optical path splitter (Cairn, Optosplit III) onto a high-speed CMOS camera (Mikrotron, EoSens CL). Magnetic tweezers were implemented using a pair of 0.25 inch $\times$ 0.25 inch $\times$ 0.5 inch rectangular neodymium magnets (K&J Magnetics B448) mounted on a stage equipped with rotary and vertical servomotors (Physik Instrumente C-150.P and DPhysik Instrumente M-126.PD1). Samples were mounted on a three-axis nanopositioning stage (Mad City Labs PDQ-series).

Data Collection

Rotor bead tracking was done at 5 kHz, with DNA tethers subject to 5 pN of tension. To ensure proper attachments to the glass coverslip and magnetic bead, all DNA molecules were recorded for at least 5 minutes prior to data collection. Attachment type of the DNA molecule to the magnetic bead (free swivel or torsionally constrained) was further determined by rotating the magnets. For torque-driven assays, DNA twist was ramped between (+7.5, -5) by rotating the magnets at 3 RPM. Torque was calculated from the angular deflection of a transducer segment of torsional stiffness 0.26 pN nm/rad (Bryant et. al. 2003, Ivanov et. al. 2020). The zero-torque angular position of the magnets was calculated as the position of maximum extension of the DNA tether at low force (Oberstrass et. al. 2012, Oberstrass et. al. 2013) and this zero was subtracted to obtain the reported twist values. A software delay between recorded and actual magnet angle was left uncorrected because of negligible effects on the results for the slow magnet rotation speed used in this study.

Prior to use in experiments, RNP was incubated at 100 nM in RBT buffer (1× condition: 10 mM Tris-HCl, pH 7.5; 10 mM MgCl₂; 1 mM EDTA; 100 mM NaCl; 1 mM DTT) for 15 minutes at 37°C. Before data collection, we flowed in approximately 3 channel volumes of RNP in the RBT buffer supplemented with 0.2% Tween-20 (Sigma, P9416) and 0.2 mg/mL BSA (Invitrogen, 15561-020). RBT buffer was prepared from a 10× frozen stock for each experiment.

Data Processing

Torsionally Relaxed Assays

Rotor bead tracking and analysis was performed as previously described (Shi et. al. 2025). Unlike that work, DNA tethers (Fig. 3.5.1A and Tables 3.7.S20-22) had multiple modifications at each end for attachment to the surface and magnetic bead, but only tethers with free-swivel attachments to the magnetic bead (verified by cycling the magnets) were used for data collection. Cumulative rotor bead angles were directly converted to units of base pairs unwound assuming B-DNA helicity of 10.5 bp/turn.

Transitions between states were scored using the Steppi change-point analysis tool (Wiggins et. al. 2015), modeling the data as originating from an Ornstein-Uhlenbeck process. Global stiffness and coupling parameters were fixed by analyzing a portion of the trace before the introduction of any RNP. The only free model parameters were thus the mean angle of the rotor bead and the change point time. Adjacent dwells occupying the same state of *R*-loop formation (C, I, O) were subsequently merged as in prior work (Ivanov et. al. 2020, Aris et. al. 2025, Shi. et. al. 2025). Boundaries between *R*-loop states were calculated using the arithmetic mean between lifetime-weighted average unwinding values of adjacent states. Traces were zeroed so that the predominant closed state corresponded to $\Delta\theta_0=0$ bp.

Transition rates between states were calculated by dividing the number of transitions from state *i* to state *j* by the total time spent in state *i*. Transitions were counted starting after the first transition event, typically C → I. Cartoons of free energy landscapes (Fig. 3.2.1J) were drawn with reference to the kinetic data shown in Table 3.7.S7. Apparent equilibrium constants were computed using

$$K_{ij} = \frac{k_{ij}}{k_{ji}}$$

where k_{ij} represents the transition rate from state *i* to state *j*. Estimated free energy differences were then computed using

$$\Delta G_{ij} = -k_B T \ln(K_{ij})$$

where k_B represents the Boltzmann constant and *T* is temperature. Well positions for C, I, and O states were set to the lifetime-weighted average $\Delta\theta_0$ in each merged state cluster, and barrier heights relative to wells were represented as

$$\Delta G^\ddagger_{ij} = -k_B T \ln(k_{ij})$$

Torque-Driven Assays

Rotor bead tracking and analysis was performed as previously described (Ivanov et. al. 2020, Aris et. al. 2025). We expressed torque changes due to TnpB *R*-loop formation in

units of base pairs unwound, assuming any change in equilibrium twist ($\Delta\theta_0$) of the tether is the result of bubble formation on B-DNA (10.5 bp per turn)

$$\Delta\theta_0(\text{bp}) = 10.5 * \frac{\tau - \tau_B}{2\pi\kappa}$$

where κ is the torsional stiffness of the full DNA tether obtained from the slope of the torque-twist curve and $\tau_B = \kappa\theta$ is the expected torque for unperturbed B-DNA.

R-loop states were identified using change point analysis as above. Twist-dependent transition rates, apparent equilibrium constants, and free energies were calculated (Ivanov et. al. 2020) using an approach adapted from an earlier procedure for analyzing force-dependent protein unfolding (Zhang et. al. 2013). Briefly, the number of transitions between states *i* and *j* in each twist bin was counted and normalized to obtain each rate k_{ij} . Twist bins with fewer than 3 sampled transitions were excluded from analysis. Standard errors of transition rates were calculated by assuming Poisson statistics for transition events. Apparent equilibrium constants were then calculated as

$$K_{ij} = \frac{k_{ij}}{k_{ji}}$$

As described previously (Ivanov et. al. 2020, Aris et. al. 2025), transitions between *R*-loop intermediates were modeled as transitions between states with changes in equilibrium twist $\Delta\theta_{ij}$ and differences in free energy ΔG_{ij} on a DNA polymer with torsional stiffness κ . Model parameters are derived from a linear fit of $\ln(K_{ij})$ as a function of imposed twist θ :

$$\ln(K_{ij}) = \beta\kappa\Delta\theta_{ij}\theta - \beta\Delta G_{ij}(0)$$

where $\beta = \frac{1}{k_B T}$. Estimated free energy differences were obtained as $\Delta G_{ij} = -k_B T \ln(K_{ij})$.

Cartoon energy landscapes were drawn with reference to the above calculations, with barrier heights and well depths derived from linear fits to the transition rates and equilibrium constants, respectively, and the remainder of the energy landscapes drawn using arbitrary interpolant curves. Barrier locations were estimated by fitting a selected approximately linear portion for each $\ln(k_{ij})$ vs twist plot, assuming:

$$\ln(k_{ij}(\theta)) = \beta\kappa\Delta\theta_{ij}^\ddagger\theta + \ln(k_{ij}(0))$$

The $\Delta\theta_{ij}^\ddagger$ fit values were then scaled for display, so that the fractional positions of the displayed transition states correspond to $\Delta\theta_{ij}^\ddagger / \Delta\theta_{ij}$, where $\Delta\theta_{ij}$ is obtained from the equilibrium fits described above. Scaling was applied because fit values of $\Delta\theta_{ij}$ did not reliably correspond to the physical differences in the $\Delta\theta_0$ values extracted from lifetime-weighted averaging, which were used to plot well positions.

Bacterial TnpB protein expression plasmids

To express TnpB proteins in bacteria, we built pBAD-10×His-MBP-TEV-TnpB expression vectors. The coding sequence (CDS) for Ymu1 TnpB was obtained from a previously

described Ymu1 plasmid pMK061 (Weiss et. al. 2025). For bacterial expression, the Ymu1 CDS were cloned into pHS354, a BbsI-based dropout vector containing an araBAD promoter, an N-terminal 10×His-MBP-TEV tag, and an mKate stuffer cassette flanked by BbsI sites. Each CDS was PCR-amplified with primers introducing BbsI-compatible overhangs and inserted into pHS354 by BbsI digestion and Golden Gate ligation, replacing the mKate stuffer with the TnpB open reading frame and yielding 10×His-MBP-TEV-TnpB constructs (Table 3.7.S16). The catalytic inactivation mutation (dYmu1/E279A), and combinatorial variants dH4W, dWFR) were generated using primer-encoded substitutions. Mutant constructs were assembled by GXL PCR followed by two-fragment Gibson assembly.

Bacterial reRNA expression plasmids

To express reRNAs in bacteria, we built T7-reRNA-HDV expression vectors. We first constructed pHS357, a BsmBI-based dropout vector containing a T7 promoter, a BsmBI stuffer cassette, and a 3' HDV ribozyme. The reRNA scaffolds and their corresponding 16-nt guide sequences were PCR-amplified with primers introducing BsmBI overhangs and inserted into pHS357 by replacing the stuffer cassette via BsmBI Golden Gate assembly (Table 3.7.S17). We chose a 127-nt processed scaffold identified by bacterial small-RNA sequencing (pHS607). Direct expression of the 127-nt scaffold resulted in higher yield of RNP in bacteria and was used for all DNA cleavage assays and equilibrium AuRBT measurements. The 200-nt scaffold construct was used to reconstitute the dYmu1 RNP employed in torque-driven AuRBT experiments. For all reRNA-expression constructs, we tested guide sequences corresponding to both Target 1 (5'-TCTTCTGGATTGTTGT-3') (Cofsky et. al. 2020, Aris et. al. 2025) and Target 2 (5'-AAGGCAAATTCGCCGC-3', targeting *PDS3* g2 site) (Weiss et. al. 2025).

Plasmid substrates for cleavage assays

To generate the plasmid substrate used in supercoiled versus linearized cleavage assays, a preannealed oligonucleotide duplex encoding the Target 1 sequence and TAM with 4-bp overhangs compatible with BsmBI-generated ends was cloned into the pGGaselect Destination Vector (New England Biolabs, N0309) using the NEBridge Golden Gate Assembly Kit (BsmBI-v2) (New England Biolabs, E1602). The resulting plasmid (pHS697, ~2.2 kb) was verified by whole-plasmid sequencing. The full plasmid sequence is provided in Table 3.7.S10.

TnpB-reRNA co-purification

Briefly, *Escherichia coli* BL21-AI cells (ThermoFisher Scientific, C607003) were co-transformed with the 10×His-MBP-TnpB expression plasmid and the corresponding T7-reRNA-HDV plasmid. Starter cultures were grown in 2XYT medium supplemented with ampicillin (0.1 mg/mL), chloramphenicol (0.034 mg/mL), and tetracycline (0.01 mg/mL) for ~18 h at 37°C. A 5-mL aliquot of the starter culture was then used to inoculate 4.5 L of 2XYT expression medium containing the same antibiotics. Cultures were grown at 37°C until they reached an OD600 of 0.6-0.8, at which point TnpB expression was induced with 0.2% (w/v) L-arabinose. Induction proceeded overnight at 16°C for ~16 h.

Cells were harvested by centrifugation and resuspended in Lysis Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 25 mM imidazole; 5% (v/v) glycerol; 1 mM TCEP)

supplemented with cOmplete EDTA-free protease inhibitor (Roche). Cells were lysed by sonication and clarified by ultracentrifugation. The supernatant was applied to Ni-NTA resin (QIAGEN), washed five times with Wash Buffer (20 mM Tris-HCl, pH 8.0; 500 mM NaCl; 25 mM imidazole; 5% (v/v) glycerol; 1 mM TCEP) and eluted using 300 mM imidazole in the same buffer. Eluted protein was dialyzed overnight at 4°C in Dialysis Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 25 mM imidazole; 5% glycerol; 1 mM TCEP) in the presence of TEV protease.

Following TEV cleavage, the digest was passed over a HisTrap column (Cytiva) to separate the cleaved TnpB protein from the His-MBP tag. The flow-through containing tag-free TnpB was concentrated and further purified using a HiTrap Heparin HP affinity column (Cytiva), eluting the protein with Ion-Exchange Buffer (20 mM Tris-HCl, pH 8.0; X mM NaCl; 10% (v/v) glycerol; 1 mM TCEP), where a NaCl gradient from 250 mM to 1 M. The fraction was then concentrated and finally purified with size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in Gel Filtration Buffer (20 mM Tris-HCl, pH 8.0; 250 mM NaCl; 10% (v/v) glycerol; 1 mM TCEP). Purified TnpB was concentrated, aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C. RNP concentrations were quantified by measuring A₂₆₀ on a NanoDrop spectrophotometer using extinction coefficients calculated for each reRNA.

The homogeneity of proteins within purified RNPs was assessed by SDS-PAGE. Protein samples were mixed with 5× SDS Loading Dye (10% SDS; 30% glycerol; 75 mM EDTA; 250 mM Tris-HCl, pH 6.8; supplemented with bromophenol blue) at a 4:1 ratio and boiled at 95°C for 2 min. Samples were loaded onto Mini-PROTEAN TGX Precast Gels (Bio-Rad) and electrophoresed in a 1× SDS running buffer at 150 V for 45 min. Gels were stained using InstantBlue Coomassie Protein Stain (Abcam, ab119211) for ≥ 30 min, rinsed with MilliQ water, and imaged on a ChemiDoc MP system under the Coomassie Blue Gel 590/110 White Trans setting.

To assess the homogeneity of reRNA within purified RNPs, 3.5 µL of 2 µM RNP was prepared for three conditions: untreated, RNase-treated, and DNase-treated. RNase digestion was performed with DNase- and protease-free RNase A (ThermoFisher Scientific, EN0531), and DNase digestion was performed using RNase-free DNase I (New England Biolabs, M0303S) with 0.5 µL 10× DNase I Reaction Buffer (New England Biolabs, B0303S). Gel Filtration Buffer was added to each sample to a final volume of 5 µL. Reactions were incubated at 37°C for 10 min and quenched by adding an equal volume of colourless denaturing loading dye (94% formamide, 30 mM EDTA, 400 µg/mL heparin). Samples were resolved on a 15% denaturing urea-PAGE gel in 0.5× TBE for 1 h at 25 W. The gel was stained with SYBR Gold (ThermoFisher Scientific, S11494) for 30 min and imaged using a ChemiDoc MP imaging system (Bio-Rad) under the SYBR Gold 590/110 Blue Trans setting.

Oligonucleotide substrates for cleavage assays

All DNA oligonucleotides were synthesized by IDT and HPLC-purified. For the double-stranded DNA (dsDNA) substrates used in cleavage assays, the target strand (TS) was 5'-labeled with FAM and the non-target strand (NTS) was 5'-labeled with Cy5, unless otherwise specified.

To generate dsDNA substrates, equimolar amounts (typically 40 µM each) of complementary single-stranded oligonucleotides were mixed in DNA Annealing Buffer

(10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA), heated to 95 °C for 5 min, and then slowly cooled to 35 °C over 45 min on a thermocycler. Annealed products were quantified by measuring A260 on a NanoDrop spectrophotometer (ThermoFisher Scientific), and concentrations were calculated using manufacturer-provided extinction coefficients.

The annealing of dsDNA was confirmed by electrophoresis on an 8% native polyacrylamide gel run at 150 V at 4°C for 4 h. Working stocks were diluted to 200 nM and stored at -20 °C. The same procedure was used to prepare canonical and mismatched DNA substrates. Sequences of all DNA oligonucleotides used in this study are provided in Table 3.7.S11.

Plasmid cleavage assay

For plasmid cleavage assays, negatively supercoiled pHS697 (~2.2 kb) was purified from DH5α cells (Fisher Scientific, 18-265-017) using the QIAprep Spin Miniprep Kit (QIAGEN, 27106). Supercoiling was confirmed by agarose gel electrophoresis. To prepare the linearized substrate, supercoiled pHS697 was digested with BsaAI using CutSmart buffer at 37°C (New England Biolabs, R0531) and purified using the QIAquick Gel Extraction Kit (QIAGEN, 28706). TnpB-reRNA RNP complexes were prepared by incubating 100 nM RNP in Cleavage Buffer (10 mM Tris-HCl, pH 7.5; 100 mM NaCl; 2 mM MgCl₂; 1 mM EDTA; 1 mM DTT) at 26 °C for 15 min. Cleavage reactions were initiated by adding pre-assembled RNP (66 μL) to the plasmid substrate (16.5 μL) to yield final concentrations of 100 nM TnpB RNP and 3 nM plasmid in 1× Cleavage Buffer. Reactions were quenched at defined time points (linearized: 0.5, 1, 5, 10, 30, 60, 120, 180 min; supercoiled: 0.25, 0.5, 1, 2, 5, 10, 30, 120 min) by adding the samples to an equivalent volume of Quenching Buffer (0.2% SDS, 50 mM EDTA). 1/5 the volume of 6x Orange DNA Loading Dye (Thermo Scientific) were added to the samples for gel visualization and products were resolved on 0.6% agarose gels in 1× TAE supplemented with 0.001% SYBR-SAFE DNA Gel Stain (Invitrogen) and imaged using a ChemiDoc MP Imaging System (Bio-Rad). Band intensities were quantified using ImageLab 6.1 (Bio-Rad) under the Blue Tray, SYBR Safe Gel (590/100) setting under Auto Optimal exposure and cleavage kinetics were fitted as described above.

3.7. Supplemental Tables

Table S1. Transition rate constants from equilibrium AuRBT assays with fully matched target sequences (related to Fig. 3.2.1 and Fig. 3.5.1)

RNP	Conc	$k_{c \rightarrow l}$ (s ⁻¹)	$k_{c \rightarrow l}$ error (s ⁻¹)	$k_{l \rightarrow c}$ (s ⁻¹)	$k_{l \rightarrow c}$ error (s ⁻¹)	$k_{l \rightarrow o}$ (s ⁻¹)	$k_{l \rightarrow o}$ error (s ⁻¹)	$k_{o \rightarrow l}$ (s ⁻¹)	$k_{o \rightarrow l}$ error (s ⁻¹)
WT dYmu1	100 nM	0.084 199	0.0024 522	0.316 62	0.0092 328	0.0102 31	0.00165 97	0.365 36	0.05927
H4W dYmu1	100 nM	0.063 184	0.0029 022	0.073 818	0.0033 763	0.0043 241	0.00081 717	0.295 39	0.05684 7
dYmu1 -WFR	100 nM	0.623 08	0.0710 06	0.017 29	0.0019 577	0.0128 56	0.00168 81	0.002 112	0.00028 222

Table S2. AuRBT trace statistics for equilibrium AuRBT assays (related to Fig. 3.2.1, Fig. 3.5.1, and Methods)

Condition	Tethers	Unique Chambers	Time Tracked (s)	C → I	I → C	I → O	O → I	C → O	O → C
WT dYmu1 (100nM)	5	3	17821	1179	1176	38	38	2	3
H4W dYmu1 (100nM)	5	5	14069	474	478	28	27	1	2
dYmu1-WFR (100nM)	5	4	31151	77	78	58	56	2	1
WT dYmu1 (100nM) 1-4 nt mismatch	2	1	3500	0	0	0	0	0	0
dYmu1-WFR (100 nM) 1-4 nt mismatch	2	1	4693	0	0	0	0	0	0

Column “i → j” gives number of transitions observed from state i to state j.

Table S3. Transition rate constants of WT dYmu1 for non-equilibrium, torsion-driven AuRBT assays with fully matched target sequences (related to 3.2.2 and Fig. 3.5.2)

Twist	$k_{C \rightarrow I}$ [s^{-1}]	$k_{C \rightarrow I}$ error [s^{-1}]	$k_{I \rightarrow C}$ [s^{-1}]	$k_{I \rightarrow C}$ error [s^{-1}]	$k_{I \rightarrow O}$ [s^{-1}]	$k_{I \rightarrow O}$ error [s^{-1}]	$k_{O \rightarrow I}$ [s^{-1}]	$k_{O \rightarrow I}$ error [s^{-1}]
-4.75	0.1371	0.0457	0.0066	0.0023	0.0270	0.0047	0.0167	0.0034
-4	0.3410	0.0633	0.0159	0.0035	0.0310	0.0048	0.0230	0.0042
-3	0.2784	0.0366	0.0363	0.0052	0.0185	0.0037	0.0184	0.0042
-2	0.2387	0.0231	0.0699	0.0071	0.0087	0.0025	0.0415	0.0088
-1	0.1139	0.0112	0.1624	0.0152	0.0043	0.0025	0.0367	0.0116
0	0.0657	0.0073	0.4068	0.0413	NA	NA	0.1347	0.0449

1	0.0331	0.0049	0.5525	0.0815	NA	NA	NA	NA
2	0.0137	0.0031	0.6544	0.1636	NA	NA	NA	NA
3	0.0021	0.0012	0.9507	0.4753	NA	NA	NA	NA
4	NA	NA	NA	NA	NA	NA	NA	NA

Table S4. Parameters from linear fits of $\ln(K_{ij})$ and $\ln(k_{ij})$ vs imposed twist (related to Fig. 3.2.2 and Fig. 3.5.2)

Transition	$\Delta\theta_{ij}$ (bp)	$\Delta G_{ij}(0)$ (k _B T)	$\Delta G_{ij}(-5)$ (k _B T)	$\Delta\theta^{\ddagger}_{ij}$ (bp)	$\ln(k_{ij}(0))$	$\Delta\theta^{\ddagger}_{ji}$ (bp)	$\ln(k_{ji}(0))$
C $\leftrightarrow$ I	5.3	1.6	-3.9	3.2	-2.8	-3.9	-0.9
I $\leftrightarrow$ O	3.6	2.8	-1.2	3.1	-6.1	-1.7	-2.5

$\Delta\theta_{ij}$, reported in number of base pairs unwound, gives the predicted difference in equilibrium twist between states i and j based on the slope of the $\ln(K_{ij})$ vs twist plot. $\Delta G_{ij}(0)$ and $\Delta G_{ij}(-5)$ give the predicted free energy difference between the states at 0 and -5 twist, respectively. Linear fits calculated from equally-weighted $\ln(K_{ij})$ data points. $\Delta\theta^{\ddagger}_{ij}$, reported in the number of base pairs unwound, gives the predicted location of the transition state for the i $\rightarrow$ j transition in relation to the predicted equilibrium twist for state i. $\ln(k_{ij}(0))$ gives the natural log of the predicted transition rate (in s⁻¹) from state i to state j at 0 twist. Linear fits calculated from equally-weighted data points selected from a linear portion of $\ln(k_{ij})$ data.

Table S5. AuRBT trace statistics for non-equilibrium torque-driven assays (related to 3.2.2, Fig. 3.5.2, and Methods)

Condition	Tethers	Unique Chambers	Cycles
WT dYmu1 (10nM)	13	13	68

Twist interval	C $\rightarrow$ I	I $\rightarrow$ C	I $\rightarrow$ O	O $\rightarrow$ I
[-5, -4.5]	9	8	33	24
[-4.5, -3.5]	29	21	41	30
[-3.5, -2.5]	58	49	25	19
[-2.5, -1.5]	107	96	12	22

[-1.5, -0.5]	104	114	3	10
[-0.5, 0.5]	82	97	1	9
[0.5, 1.5]	45	46	0	0
[1.5, 2.5]	19	16	0	0
[2.5, 3.5]	3	4	0	0
[3.5, 4.5]	2	2	0	0
[4.5, 5.5]	0	0	0	0
[5.5, 6.5]	0	0	0	0
[6.5, 7.5]	0	0	0	0

Column "i → j" gives number of transitions observed from state i to state j.

Table S6. Rate constants (k_{obs}) for plasmid cleavage assay (n=3) (related to 3.2.2)

Protein	Type	Mean (min^{-1})	SD (min^{-1})
WT Ymu1	Supercoiled	3.41	1.35
	Linearized	0.0288	0.00671
Ymu1-WFR	Supercoiled	5.40	0.934
	Linearized	0.683	0.137

Table S7. Rate constants (k_{obs}) for pre-unwound DNA cleavage assay (n=3) (related to 3.2.2 and Fig. 3.5.2)

Strand	DNA target	WT Ymu1	WT Ymu1	H4W mean (min^{-1})	H4W SD (min^{-1})	Ymu1-WFR	Ymu1-WFR SD (min^{-1})
--------	------------	---------	---------	--------------------------------	------------------------------	----------	-----------------------------------

		mean (min ⁻¹)	SD (min ⁻¹)			mean (min ⁻¹)	
NTS	Canonical substrate	0.0676	0.0033	0.168	0.039	2.00	0.74
	1-2 nt pre-unwound substrate	0.162	0.040	0.217	0.060	1.42	0.16
	8-9 nt pre-unwound substrate	1.77	0.13	1.57	0.12	2.17	0.18
	1-9 nt pre-unwound substrate	1.51	0.20	2.17	0.65	2.44	0.33
TS	Canonical substrate	0.0717	0.0090	0.176	0.055	0.980	0.33
	1-2 nt pre-unwound substrate	0.180	0.050	0.301	0.004	1.16	0.13
	8-9 nt pre-unwound substrate	1.29	0.22	1.14	0.18	1.49	0.24
	1-9 nt pre-unwound substrate	1.02	0.16	1.38	0.22	1.21	0.17

Table S8. Plasmid vectors used in this study (related to Methods)

Internal ID	System	Purpose
pZZ09	<i>E.coli</i>	Bacterial expression of Ymu1 protein
pZZ031	<i>E.coli</i>	Bacterial expression of dYmu1 protein
pZZ069	<i>E.coli</i>	Bacterial expression of V305R Ymu1 protein
pZZ080	<i>E.coli</i>	Bacterial expression of H4W Ymu1 protein
pZZ087	<i>E.coli</i>	Bacterial expression of H4W dYmu1 protein
pHS749	<i>E.coli</i>	Bacterial expression of Ymu1-WFR protein
pHS750	<i>E.coli</i>	Bacterial expression of dYmu1-WFR protein
pZZ089	<i>E.coli</i>	Bacterial expression of L304F Ymu1 protein
pHS607	<i>E.coli</i>	Bacterial expression of Ymu1 200-nt reRNA for target 1

pHS610	<i>E.coli</i>	Bacterial expression of Ymu1 127-nt reRNA for target 1
pHS612	<i>E.coli</i>	Bacterial expression of Ymu1 short reRNA for target 1
pZZ062	<i>E.coli</i>	Bacterial expression of Ymu1 short reRNA for target 2
pZZ064	<i>E.coli</i>	Bacterial expression of Ymu1 127-nt reRNA for target 2
pZZ013	<i>E.coli</i>	Bacterial expression of Dra2 200-nt reRNA for target 1
pZZ052	<i>E.coli</i>	Bacterial expression of Tel2 200-nt reRNA for target 1
pZZ035	<i>E.coli</i>	Bacterial expression of Tfu1 200-nt reRNA for target 1
pZZ054	<i>E.coli</i>	Bacterial expression for Ec41 200-nt reRNA for target 1
pHS516	<i>E.coli</i>	Bacterial TAM assay vector encoding WT Ymu1 and WT reRNA expressed as a single transcript, with a 16-nt guide flanked by HDV
pHS751	<i>E.coli</i>	Bacterial TAM assay vector encoding Ymu1-WFR and WT reRNA expressed as a single transcript, with a 16-nt guide flanked by HDV
pHS697	<i>E.coli</i>	Substrate plasmid for cleavage assay encoding TAM and Target 1

Table S9. Primers used for Gibson assembly-based site-directed mutagenesis of Ymu1 TnpB for bacterial expression (related to Methods)

Oligonucleotide	Internal ID	Sequence (5'-3')
Reverse primer for amplifying Ymu1 backbone plasmid	oZZ080R	ttgacggcttgacggagtagcataggggttg cag
Forward primer for amplifying Ymu1 backbone plasmid	oZZ081F	gcagggattctgcaaaccctatgctactcc gtc
Forward primer to induce the E279A mutation in Ymu1	oZZ086F	AAACCACGATATCATCTGTAT CgcgGACCTTAACGTTAAGGG CATGATGC
Reverse primer to induce the E279A mutation in Ymu1	oZZ087R	GCCCTTAACGTTAAGGTCgcg GATACAGATGATATCGTGGTT TTTGACTATCTCTGTAC

Reverse primer to induce the H4W mutation in Ymu1	oZZ227R	GTATTCATAGGCTTTccaCTGC AGCATTGCATTGGATTGG
Forward primer to induce the H4W mutation in Ymu1	oZZ228F	AATGCAATGCTGCAGtggAAAG CCTATGAATACCGTATCTATC CAGATAAGAAG
Forward primer to induce the both L304F/V305R mutations in Ymu1	oZZ233F	TGGACGAGCttccgaTCGAAACT GCAGTACAAGGCTTC
Reverse primer to induce the both L304F/V305R mutations in Ymu1	oZZ234R	CAGTTTCGAtcggaaGCTCGTC CATGATACATCAGAGATGC
Reverse primer to induce the L304F mutation in Ymu1	oZZ235R	CTGCAGTTTCGATACgaaGCT CGTCCATGATACATCAGAGAT GC
Forward primer to induce the L304F mutation in Ymu1	oZZ236F	GTATCATGGACGAGCttcGTAT CGAAACTGCAGTACAAGGCTT C
Forward primer to induce the V305R mutation in Ymu1	oZZ217F	TCATGGACGAGCCTTagaTCG AAACTGCAGTACAAGGCTTC
Reverse primer to induce the V305R mutation in Ymu1	oZZ220R	GTA CTGCAGTTTCGAtctAAGG CTCGTCCATGATACATCAGAG

Table S10. DNA sequence of plasmid cleavage substrate (related to Methods)

pHS697

CGAAAAATCAATAATCAGACAACAAGATGTGCGAACTCGATATTTTACACGACTC
TCTTTACCAATTCTGCCCCGAATTACACTTAAAACGACTCAACAGCTTAACGTTGG
CTTGCCACGCATTACTTGACTGTAAACTCTCACTCTTACCGAACTTGGCCGTAA
CCTGCCAACCAAAGCGAGAACAAAACATAACATCAAACGAATCGACCGATTGTTA
GGTAATCGTCACCTGCAGGAAGGTTTAAACGCATTTAGGTGACACTATAGAAGTG
TGTATCGCTCGAGGGATCCGAATTCGAAGTCTTGGTACGGAGCGAGACCGGAG
GTCATAATGATTGATT**CTTCTGGATTGTTGT**AAGCAGCATTTGAGCAAAAATCTGT
TGCCCATGGTCTCACCATTCTGTAGACTTCTTAATTAAGACGTCAGAATTCTCGA
GGCGGCCGCATGTGAGTCTCCCTATAGTGAGTCGTATTAATTTTCGCGGGCGGAA
CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGTAGCA
CCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGC
CACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCAT
CACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTGCGCTT
GCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGG
CCACGTTTAAATCAAACCTGGTGAACTCACCCAGGGATTGGCTGAGACAAAAAA
CATATTCTCAATAAACCCCTTTAGGGAAATAGGCCAGGTTTTTCACCGTAACACGCC
ACATCTTGCGAATATATGTGTAGAACTGCCGGAAATCGTCGTGGTATTCACTCC
AGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAAACAAGGGTGAA
CACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATG
AGCATTATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAACTTGTGCTT
ATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTAT
AGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTG
GGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAG
CTCCTGAAAATCTCGATAACTCAAAAAATACGCCCCGGTAGTGATCTTATTTTATTA
TGGTGAAAGTTGGAACCTCTTACGTGCCGATCAAAGTCTCATTTTCGCCAAAAGT
TGTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCC
GTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG
CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGA
GCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAAT
ACTGTTCTTCTAGTGATAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAC
CGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG
ATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC
AGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCT
TCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG
GAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCT
GTCGGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAATGCGGCCTTTTTACGGTTCCTGGCC
TTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGA
TAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGC

bolded region represents the 16-nt Target 1 sequence (the TS sequence);
italicized region represents the TAM (the complementary sequence of 5'-TTGAT-3');

Table S11. Sequences of all the oligonucleotide substrates used in DNA cleavage assays (related to Methods)

Internal ID	Description	Sequence (5'-3')	Figures
dZZ01	Ymu1 DNA substrate 0-16 bp corresponding to target 1; NTS; Cy5	/5Cy5/CCGGAGGTCA TAATGATTGATTCTTC TGGATTGTTGTAAGC AGCATTGAGCAAAA ATCT	Fig. 3.2.2J, 3.5.2H
dZZ01C	Ymu1 DNA substrate 2-16 bp corresponding to target 1 (substrate pre-unwound at bases 1 and 2) ; NTS; Cy5	/5Cy5/CCGGAGGTCA TAATGATTGATAGTT CTGGATTGTTGTAAG CAGCATTGAGCAAA AATCT	Fig. 3.2.2J, 3.5.2H
dZZ01F	Ymu1 DNA substrate 0-7 and 10-16 bp corresponding to target 1 (substrate pre-unwound at bases 8 and 9); NTS; Cy5	/5Cy5/CCGGAGGTCA TAATGATTGATAGAA GACCTTTGTTGTAAG CAGCATTGAGCAAA AATCT	Fig. 3.2.2J, 3.5.2H
dZZ01G	Ymu1 DNA substrate 10-16 bp corresponding to target 1 (substrate pre-unwound from bases 1 to 9); NTS; Cy5	/5Cy5/CCGGAGGTCA TAATGATTGATTCTTC TGCCTTGTTGTAAGC AGCATTGAGCAAAA ATCT	Fig. 3.2.2J, 3.5.2H
dZZ02	Ymu1 DNA substrate 0-16 bp corresponding to target 1; TS; FAM	/56- FAMN/AGATTTTTGCT CAAATGCTGCTTACA ACAATCCAGAAGAAT CAATCATTATGACCT CCGG	Fig. 3.2.2J, 3.5.2H

Table S12. Primer sequences used to generate DNA tether fragments in AuRBT (related to Fig. 3.5.1 and Methods).

	Primer Sequence (5'-3')	Length	Template	Digest
F500	gatcgaagacacttag ACAACCCACA AGTATAGAGGCTCCTATG GACGCGGATATAATGACATTT CCTAAC	520	pFO-SE2	BbsI
XD	gaaggggtctcatgac TC ACTATAGG GCGAATTGGAGCTCCACCGC G gaaggggtctcactaa CACTAAAGGG AACAAAAGCTGGTAC	4165	pFO-SE2	BsaI
SOI	gatcgaagacacctgc TG ACGCATC AGTCAGTACTACTGACG gatcgaagacacgtca GTGGTAGCT GCTGTGCGAGTG	290	TnpB Gblock	BbsI
SOI 1-4MM	gatcgaagacacctgc TG ACGCATC AGTCAGTACTACTGACG gatcgaagacacgtca GTGGTAGCT GCTGTGCGAGTG	290	TnpB 1-4MM Gblock	BbsI
SPMDIG	GCGCAGCACGCAGATACACTC gatcgaagacacgcag CATACGTCT GCGTCGCTG	320	C12Gblock	BbsI

Biotin-modified nucleotides are **bolded**.

Table S13. Target sequences in the SOI (related to Fig. 3.5.1 and Methods).

DNA Target	Sequence (5'-3')
Target 1	<i>TTGATTCTTCTGGATTGTTGT</i>
1-4 mismatch	<i>TTGATAGAACTGGATTGTTGT</i>

TAM sequence *italicized*.

Table S14. The full DNA tether sequence with Target 1 (related to Fig. 3.5.1 and Methods)

<u>GACGCGGATATAATGACATTTCTAACTTTTGGGCAAAAATTCGCTATCATATGC</u> <u>GAGAACCGTTTGCGGAGTTTCTCGGGACACTAGTTCTTGTCATTTTTGGTGTTGG</u> <u>TGGTAATCTTCAAGCAACTGTAACAAAAGGTAGTGGTGGTTCCTATGAATCCCTA</u> <u>TCATTTGCATGGGGGTTTCGGTTGTATGCTTGGTGTTTACGTCGCAGGCGGTATTA</u> <u>GTGGTGGTCATATTAACCCTGCTGTTACGATTTCAATGGCAATTTTTCGAAAATTC</u> <u>CCCTGGAAAAGGTGCCCGTATATATTGTTGCTCAGATTATCGGTGCATATTTTG</u>
--

GAGGAGCTATGGCTTATGGTTATTTTTGGAGCTCTATCACAGAATTTGAGGGAGG
TCCGCACATAAGAACAACGGCGACCGGTGCGTGTTTGTTTACTGATCCAAAGTCT
TACGTCACGTGGAGAAATGCCTTCTTTGACGAATTCATAGGAGCCTCTATACTTG
TGGGTTGTCTAACACTAAAGGGAACAAAAGCTGGTACCGGGCCCCCCTCGAGC
GGTACCCCACTTACCCACCCCGGAAATTTGAGTTATAAACGTTGTTTGAGCTTTA
CCTAGTCTTGGTCGATCAAAAGTTCTGGTACCTTTTCACCATGTCTCCCCCCTTAT
TCATATAAAAAGAAGCGTATAATCGCACAGTATAACGCTCCTCTGATATATGATCT
AGACCCAAGTAATGAGTTACGAATCTGGGAGGTCATCCTCCTCTTCCGAGAGTA
CACGGCCACCAACGCTAAAAGAAGAACCTAATGGTAAAATAGCTTGGGAAGAAA
GTGTCAAAAATCTAGGGAAAATAACGAAAATGACAGCACTCTCTTGAGGCGAAA
GCTAGGTGAGACTCGAAAAGCAATTGAACTGGAGGATCATCGAGAAATAAACTT
TCTGCTTTGACACCCTTGAAAAAGTGTTGACGAGAGGAAGGATTCGGTACAA
CCACAGGTCCCTTCCATGGGTTTTACTTATTCTTTGCCTAATTTGAAGACTTTAAA
CAGTTTTTTCAGATGCTGAGCAAGCACGTATAATGCAAGATTATCTATCCAGGGGG
GTAAATCAAGGCAACAGTAATAATTATGTAGACCCACTATATCGGCAATTAAATCC
AACTATGGGTAGTAGCAGGAACAGGCCTGTTTGGAGTTTAAATCAGCCGTTACC
GCATGTATTGGATCGAGGCTTGGCAGCAAAGATGATACAAAAGAATATGGATGC
AAGGTCCCGCGCATCATCGAGACGAGGGTCGACCGATATTTCAAGGGGGGGTT
CTACTACGTCAGTGAAAGACTGGAAAAGGCTCCTTAGAGGTGCAGCACCGGGTA
AAAAGCTTGGTGACATCGAAGCTCAAACGCAACGCGATAATACTGTTGGTGACG
ATGTGAAACCTACTAAGTTAGAGCCTGAAAACCCACAAAAGCCCTCTAACACGCA
TATTGAGAATGTTTCACGTAAGAAAAAGCGTACTTCGCATAATGTCAATTTTTCAT
TAGGCGATGAAAGCTACGCATCCTCCATAGCCGATGCAGAATCCAGAAAATTAAA
GAACATGCAAACCCTCGATGGTTCTACTCCGGTTTATACGAAGCTTCCTGAAGAA
CTTATTGAAGAGGAAAATAAAAGTACGAGTGCATTAGATGGTAATGAAATTGGTG
CCTCAGAAGATGAAGACGCGGATATAATGACATTTCTAACTTTTGGGCAAAAAT
TCGCTATCATATGCGAGAACCGTTTGCGGAGTTTCTCGGGACACTAGTTCTTGTC
ATTTTTGGTGTTGGTGTAATCTTCAAGCAACTGTAACAAAAGGTAGTGGTGGTT
CCTATGAATCCCTATCATTTGCATGGGGGTTTCGGTTGTATGCTTGGTGTTTACGT
CGCAGGCGGTATTAGTGGTGGTCATATTAACCCTGCTGTTACGATTTCAATGGCA
ATTTTTCGAAAATTCCCCTGGAAAAAGGTGCCCGTATATATTGTTGCTCAGATTAT
CGGTGCATATTTTGGAGGAGCTATGGCTTATGGTTATTTTTGGAGCTCTATCACA
GAATTTGAGGGAGGTCCGCACATAAGAACAACGGCGACCGGTGCGTGTTTGTTT
ACTGATCCAAAGTCTTACGTCACGTGGAGAAATGCCTTCTTTGACGAATTCATAG
GAGCCTCTATACTTGTGGGTTGTTTGATGGCGCTATTGGATGATAGTAATGCTCC
ACCTGGCAATGGTATGACCGCATTAAATTATTGGATTCTTAGTCGCTGCAATTGGT
ATGGCCCTTGGATATCAAACAAGTTTCACAATCAATCCTGCAAGAGATCTCGGTC
CTCGCATATTTGCTTCCATGATTGGCTATGGTCCACATGCTTTTCATCTCACACAT
TGGTGGTGGACATGGGGAGCCTGGGGTGGTCCAATTGCCGGCGGTATTGCTGG
AGCACTCATATATGACATTTTCATTTTTACTGGATGCGAATCCCCAGTCAACTACC
CAGACAACGGTTATATTGAGAATAGGGTAGGCTGCAGCCCGGGGGATCCACTAG
TTCTAGAGCGGCCGCGGTACCCCACTTACCCACCCCGGAAATTTGAGTTATAA
ACGTTGTTTGAGCTTTACCTAGTCTTGGTCGATCAAAAGTTCTGGTACCTTTTCAC
CATGTCTCCCCCTTATTCATATAAAAAGAAGCGTATAATCGCACAGTATAACGCT
CCTCTGATATATGATCTAGACCCAAGTAATGAGTTACGAATCTGGGAGGTCATCC
TCCTCTTCCGAGAGTACACGGCCACCAACGCTAAAAGAAGAACCTAATGGTAAAA

TAGCTTGGGAAGAAAGTGTCAAAAAATCTAGGGAAAATAACGAAAATGACAGCAC
 TCTCTTGAGGCGAAAGCTAGGTGAGACTCGAAAAGCAATTGAAACTGGAGGATC
 ATCGAGAAATAAACTTTCTGCTTTGACACCCTTGAAAAAAGTGGTTGACGAGAGG
 AAGGATTCGGTACAACCACAGGTCCCTTCCATGGGTTTTACTTATTCTTTGCCTAA
 TTTGAAGACTTTAAACAGTTTTTTCAGATGCTGAGCAAGCACGTATAATGCAAGATT
 ATCTATCCAGGGGGGTAAATCAAGGCAACAGTAATAATTATGTAGACCCACTATA
 TCGGCAATTAAATCCAACCTATGGGTAGTAGCAGGAACAGGCCTGTTTGGAGTTTA
 AATCAGCCGTTACCGCATGTATTGGATCGAGGCTTGGCAGCAAAGATGATACAA
 AAGAATATGGATGCAAGGTCCCGCGCATCATCGAGACGAGGGTTCGACCGATATT
 TCAAGGGGGGGTTCTACTACGTCAGTGAAAGACTGGAAAAGGCTCCTTAGAGGT
 GCAGCACCGGGTAAAAAGCTTGGTGACATCGAAGCTCAAACGCAACGCGATAAT
 ACTGTTGGTGCAGATGTGAAACCTACTAAGTTAGAGCCTGAAAACCCACAAAAGC
 CCTCTAACACGCATATTGAGAATGTTTCACGTAAGAAAAAGCGTACTTCGCATAA
 TGTCATTTTTTTCATTAGGCGATGAAAGCTACGCATCCTCCATAGCCGATGCAGAA
 TCCAGAAAATTAAGAACATGCAAACCCTCGATGGTTCTACTCCGGTTTATACGA
 AGCTTCCTGAAGAACTTATTGAAGAGGAAAATAAAAGTACGAGTGCATTAGATGG
 TAATGAAATTGGTGCCTCAGAAGATGAAGACGCGGATATAATGACATTTCTAAC
 TTTTGGGCAAAAATTCGCTATCATATGCGAGAACCGTTTTCGGAGTTTCTCGGGA
 CACTAGTTCTTGTCAATTTTTTGGTGGTGGTAAATCTTCAAGCAACTGTAACAAAA
 GGTAAGTGGTGGTTCCTATGAATCCCTATCATTTGCATGGGGGTTTCGGTTGTATGC
 TTGGTGGTTTACGTCGCAGGCGGTATTAGTGGTGGTCATATTAACCCTGCTGTTAC
 GATTTCAATGGCAATTTTTTCGAAAATTCCCCTGGAAAAAGGTGCCCGTATATATTG
 TTGCTCAGATTATCGGTGCATATTTTGGAGGAGCTATGGCTTATGGTTATTTTTTG
 AGCTCTATCACAGAATTTGAGGGAGGTCCGCACATAAGAACAACGGCGACCGGT
 GCGTGGTTTGTACTGATCCAAAGTCTTACGTCACGTGGAGAAATGCCTTCTTTG
 ACGAATTCATAGGAGCCTCTATACTTGTGGGTTGTTTGATGGCGCTATTGGATGA
 TAGTAATGCTCCACCTGGCAATGGTATGACCGCATTAAATTATTGGATTCTTAGTC
 GCTGCAATTGGTATGGCCCTTGGATATCAAACAAGTTTCACAATCAATCCTGCAA
 GAGATCTCGGTCCTCGCATATTTGCTTCCATGATTGGCTATGGTCCACATGCTTT
 TCATCTCACACATTGGTGGTGGACATGGGGAGCCTGGGGTGGTCCAATTGCCGG
 CGGTATTGCTGGAGCACTCATATATGACATTTTCAATTTTACTGGATGCGAATCCC
 CAGTCAACTACCCAGACAACGGTTATATTGAGAATAGGGTAGGGGCCGCCACCG
 CGGTGGAGCTCCAATTCGCCCTATAGTGAGTCAGTGGTAGCTGCTGTGCGAGTG
 TATCTGCGTGCTGCGCTAGTGAGATCACGTGCTCATGATACGTGATCTCGTCGC
 GTAGTGCAGCGAGATCGCGTATGTACTGCTT**ACAACAATCCAGAAGAATCAAGC**
 AGCGATATCTCGTGTGCAGACAGCTCGCGACGATGCGCTACGTCATCTAGTAGT
 GCGTACTGACTCTCAGTATACGACATATCTAGACGCTGCTGACTGTATATCTCTA
 TAGCTCAGGCATGCTAGTCTAGCGTCAGTAGTACTGACTGATGCGTCAG**CAGCA**
TACGTCTGCGTCGCTGTGGTGTATACGTGCGACATACTCAGAGTCGCGCTCTCT
ACGGTGTCACTCGAGCTGTAGATGCTCAGTGCGATACAGTACGCTCCATGCTG
CTCATCAGCTGCTCAGCTCTGTATCTGCTCTGACATGACGTGACTCAGCGCGAG
CGTATGATCGAGATACTGATCGTACACTCACTGGTCAGTGTGTAGCGTGACTGC
GTGCAGCTCAGTCAGAGCATCACTCCTCGCCATCATGTCCGAGACTAGTGGTG
GTAGCTGCTGTGCGAGTGTATCTGCGTGCTGCGC

A represents adenine nucleotides paired with biotin-modified thymines, where the streptavidin-coated gold rotor bead can be attached;

bolded region represents the 16-nt Target 1 sequence (the TS sequence);
italicized region represents the TAM (the complementary sequence of 5'-TTGAT-3');
bolded and underlined region represents the segment with multiple incorporated dUTP-digoxigenin (Roche);
italicized and underlined region represents the segment with multiple incorporated fluorescein-dUTPs (Roche).

Table S15. AF3 query sequences for the WT Ymu1 ternary complex (related to Methods)

Biomolecule	Sequences
Protein	MLQHKAYEYRIYPDKKQETLIAKTIGSSRYVYNHF LELWNKEYEETGKGLTYACSKLLTKLRDPETV WLCEVDKFSLQNSLRNLSDAFSRFFKGQNEHPQ FKSKKSPRQSYTTQYTNNNIAVSGNCLKLPKLGL VKFADSREMKGRI LNATVRRKSSGKFFVSILCEE EICELPKTDSSVGIDLGII DFAVMSDGSRHDNNHF TRQMEERLRREQRKLARRALAAEKRGISLSEAR NYQKQRRKVARLYEKVANQRKEYLNKLSTEIVKN HDIICIEDLNVKGM MRNHKLAKSISDVSWTSLVSK LQYKASWYGKEVIRISRWFPSSQICSECGHKDRK KPLHVREWTCPVCHAHHDRDVNAARNILA EGLRI RALTPGS
RNA (5'-3')	CCGCACCCCCUGCGGGGGUAGCUUGGUA AAC AAGAGAAACCUCUGCCGGCAAAGAAUAAGCC GGUAAGUAUGCUCUGU UCCCAAGAAUCUCGU GACUUUAGUCAUGAGAGGGGGGAUCUUCUGG AUUGUU
DNA (5'-3')	AACAATCCAGAAGAATCAATCA
DNA (5'-3')	TGATTGAT

Table S16. Amino acid sequence of the Ymu1 TnpB bacterial expression construct (related to Methods)

ID	Protein sequence
WT Ymu1	MKSSHHHHHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGGKKFEKDT GIKVTVEHPDKLEEKFPQVAATGDGPDII FWAHDRFGGYAQSGLLAEITPD KAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKD LLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVD

	NAGAKAGLTFLVDLIKHKHMNADTDYSIAEAAFNKGETAMTINGPWAWSN IDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPO MSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNNNNNNNNNLGIE ENLYFQSNAMLQHKAYEYRIYPDKKQETLIAKTIGSSRYVYNHFLELWNK EYEETGKGLTYACSKLLTKLKRDPETVWLCEVDKFSLQNSLRNLSDAF SRFFKGQNEHPQFKSKKSPRQSYTTQYTNNNIAVSGNCLKLPKLGLVKF ADSREMKGRLNATVRRKSSGKFFVSILCEEEICELPKTDSSVGIDLGIIIDF AVMSDGSRHDNNHFTRQMEERLRREQRKLARRALAAEKRGISLSEARN YQKQRRKVARLYEKVANQRKEYLNKLSTEIVKNHDIICIEDLNVKGMNRN HKLAKSISDVSWTSLVSKLQYKASWYGKEVIRISRWFPSSQICSECGHK DRKKPLHVREWTCPVCHAHHRDVAARNILAEGLRIRALTPGS
--	--

10 His-tags; MBP; TEV site; TnpB protein

Table S17. DNA sequence of the 127-nt reRNA expression cassette (related to Methods)

ID	DNA sequence (5'-3')
127-nt reRNA	TAATACGACTCACTATAGG CAAACAGGAACCGCAGGAATTGCGGGGGT AGCTTGGTAAACAAGAGAAACCTCTGCCGGCAAAGAAATAAGCCGG TAAGTATGCTCTGTTCCCAAGAATCTCGTGACTTTAGTCATGAGAGTT TCAATCTTCTGGATTGTTGT <u>GGCCGGCATGGTCCCAGCCTCCTCGCTGGCG</u> CCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC

T7 promoter; **reRNA scaffold**; **guide sequence (for Target 1)**; HDV ribozyme

Declaration of interests

Trevor Weiss, Maris Kamalu, Honglue Shi, Jennifer A. Doudna, and Steven E. Jacobsen have filed a patent covering aspects of this work. The Regents of the University of California have patents issued and pending for CRISPR technologies on which the authors are inventors. Jennifer A. Doudna is a cofounder of Azalea Therapeutics, Caribou Biosciences, Editas Medicine, Evercrisp, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. Jennifer A. Doudna is a scientific advisory board member at Evercrisp, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, The Column Group, and Inari. Jennifer A. Doudna is Chief Science Advisor to Sixth Street; is a Director at Johnson & Johnson, Altos, and Tempus; and has a research project sponsored by Apple Tree Partners. Steven E. Jacobsen is a cofounder and consultant for Inari Agriculture and a consultant for Terrana Biosciences, Invaio Sciences, Sail Biomedicines, Six Street and Zymo Research.

References

1. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493-500 (2024).
2. Adler, B.A. et al. CasPEDIA Database: a functional classification system for class 2 CRISPR-Cas enzymes. *Nucleic Acids Research* 52, D590-D596 (2023).

3. Al-Shayeb B, Skopintsev P, Soczek KM, Stahl EC, Li Z, Groover E, Smock D, Eggers AR, Pausch P, Cress BF, Huang CJ, Staskawicz B, Savage DF, Jacobsen SE, Banfield JF, Doudna JA. Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell*. 185(24):4574-4586.e16 (2022).
4. Aris, K.D.P. et al. Dynamic basis of supercoiling-dependent DNA interrogation by Cas12a via R-loop intermediates. *Nat Commun* 16, 2939 (2025).
5. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 315(5819):1709-12 (2007).
6. Bryant, Z. et al. Structural transitions and elasticity from torque measurements on DNA. *Nature* 424, 338-341 (2003).
7. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890 (2018).
8. Cofsky, J.C. et al. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *eLife* 9, e55143 (2020).
9. Dong, D., Ren, K., Qiu, X. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
10. Eggers, A.R. et al. Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell* 187, 3249-3261 e14 (2024).
11. Finzi, L. & Olson, W.K. The emerging role of DNA supercoiling as a dynamic player in genomic structure and function. *Biophysical Reviews* 8, 1-3 (2016).
12. Gong, S., Yu, H.H., Johnson, K.A. & Taylor, D.W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Reports* 22, 359-371 (2018).
13. Gorbenko, F. et al. Directed evolution of compact RNA-guided nucleases for homology-directed repair in mammalian cells. *bioRxiv*, 2025.10.27.684765 (2025).
14. Gore, J. et al. Mechanochemical analysis of DNA gyrase using rotor bead tracking. *Nature* 439, 100-104 (2006).
15. Guan, K. et al. Comparative characterization of Cas12f orthologs reveals mechanistic features underlying enhanced genome editing efficiency. *bioRxiv*, 2025.08.14.670346 (2025).
16. Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, Cofsky JC, Kyrpides NC, Banfield JF, Doudna JA. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*. 362(6416):839-842 (2018).
17. Hino, T. et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* 186, 4920-4935 e23 (2023).
18. Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics* 35, 5326-5327 (2019).
19. Ivanov, I.E. et al. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc Natl Acad Sci U S A* 117, 5853-5860 (2020).
20. Ivanov, I.E. et al. Multimodal Measurements of Single-Molecule Dynamics Using FluoRBT. *Biophys J* 114, 278-282 (2018).
21. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases *Trends Genet*. 1996; 12:224-228

22. Karmakar, S. et al. A miniature alternative to Cas9 and Cas12: Transposon-associated TnpB mediates targeted genome editing in plants. *Plant Biotechnology Journal* 22, 2950-2953 (2024).
23. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 599, 692-696 (2021).
24. Koonin, E.V., Makarova, K.S. Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biol. Evol.* 9 (10), 2812-2825 (2017)
25. Koonin EV, Makarova KS. Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci.* 374(1772):20180087 (2019)
26. Koonin, E.V., Makarova, K.S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37, 67-78 (2017).
27. Kosaka M, Sajiki AF, Fujita K, Yamada K, Kawano K, Hirazawa K, Matsuno T, Takada M, Shimozaawa N, Inoue KI, Yuki K, Nishiguchi KM. Evaluation of the loading capacity and patterns of packaged DNA in AAV genomes of different sizes using long-read sequencing. *Mol Ther Methods Clin Dev.* 2025 Apr 19;33(2):101474
28. LaMont, C.H. & Wiggins, P.A. The Development of an Information Criterion for Change-Point Analysis. *Neural Computation* 28, 594-612 (2016).
29. Lebel, P., Basu, A., Oberstrass, F.C., Tretter, E.M. & Bryant, Z. Gold rotor bead tracking for high-speed measurements of DNA twist, torque and extension. *Nat Methods* 11, 456-62 (2014).
30. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* (2013).
31. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9 (2009).
32. Li, Z. et al. Engineering a transposon-associated TnpB- ω RNA system for efficient gene editing and phenotypic correction of a tyrosinaemia mouse model. *Nature Communications* 15, 831 (2024).
33. Liu JJ, Orlova N, Oakes BL, Ma E, Spinner HB, Baney KLM, Chuck J, Tan D, Knott GJ, Harrington LB, Al-Shayeb B, Wagner A, Brötzmann J, Staahl BT, Taylor KL, Desmarais J, Nogales E, Doudna JA. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature.* (7743):218-223. (2019).
34. Liu, E., Page, J.E. Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. *Plant Methods* 4, 5 (2008)
35. Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* 42(1), 3-19 (2014)
36. Luo, J. et al. STAGE: A compact and versatile TnpB-based genome editing toolkit for *Streptomyces*. *Proc Natl Acad Sci U S A* 122, e2509146122 (2025).
37. Lv, Z. et al. Targeted mutagenesis in *Arabidopsis* and medicinal plants using transposon-associated TnpB. *Journal of Integrative Plant Biology* 66, 2083-2086 (2024).
38. Madariaga-Marcos, J. et al. Structural and mechanistic insights into the sequential dsDNA cleavage by SpCas12f1. *Nucleic Acids Research* 53 (2025).

39. Meers, C. et al. Transposon-encoded nucleases use guide RNAs to promote their selfish spread. *Nature* 622, 863-871 (2023).
40. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nature Methods* 19, 679-682 (2022).
41. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*.156(5):935-49 (2014)
42. Nakagawa, R. et al. Cryo-EM structure of the transposon-associated TnpB enzyme. *Nature* 616, 390-397 (2023).
43. Newton, M.D. et al. Negative DNA supercoiling induces genome-wide Cas9 off-target activity. *Mol Cell* 83, 3533-3545 e5 (2023).
44. Oberstrass, F.C., Fernandes, L.E. & Bryant, Z. Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA. *Proc Natl Acad Sci U S A* 109, 6106-11 (2012).
45. Oberstrass, F.C., Fernandes, L.E., Lebel, P. & Bryant, Z. Torque Spectroscopy of DNA: Base-Pair Stability, Boundary Effects, Backbending, and Breathing Dynamics. *Physical Review Letters* 110, 178103 (2013).
46. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* 3, e02030 (2014).
47. Pacesa, M., Pelea, O., Jinek, M. Past, present, and future of CRISPR genome editing technologies *Cell*, 187, 1076-1100 (2024).
48. Pasternak C, et al. 2013. ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. *Mol Microbiol.* 88(2):443–455.
49. Pausch P, Al-Shayeb B, Bisom-Rapp E, Tsuchida CA, Li Z, Cress BF, Knott GJ, Jacobsen SE, Banfield JF, Doudna JA. CRISPR-Cas Φ from huge phages is a hypercompact genome editor. *Science*. 369(6501):333-337 (2020).
50. Pausch, P. et al. DNA interference states of the hypercompact CRISPR–Cas Φ effector. *Nature Structural & Molecular Biology* 28, 652-661 (2021).
51. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42, W187-91 (2014).
52. Robinson, J.T. et al. Integrative genomics viewer. *Nat Biotechnol* 29, 24-6 (2011).
53. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol.* 14(12):8096-106 (1994).
54. Rutkauskas, M. et al. A quantitative model for the dynamics of target recognition and off-target rejection by the CRISPR-Cas Cascade complex. *Nat Commun* 13, 7460 (2022).
55. Sasnauskas, G. et al. TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* 616, 384-389 (2023).
56. Shi, H. et al. Rapid two-step target capture ensures efficient CRISPR-Cas9-guided genome editing. *Mol Cell* 85, 1730-1742 e9 (2025).
57. Sinden, R.R., Carlson, J.O. & Pettijohn, D.E. Torsional tension in the DNA double helix measured with trimethylpsoralen in living *E. coli* cells: Analogous measurements in insect and human cells. *Cell* 21, 773-783 (1980).

58. Skopintsev, P. et al. Structure and evolution-guided design of minimal RNA-guided nucleases. *bioRxiv*, 2025.12.08.692503 (2025).
59. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* 10, 1643-69 (2015).
60. Stella, S. et al. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* 175, 1856-1871 e21 (2018).
61. Szczelkun, M.D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* 111, 9798-803 (2014).
62. Thornton, B.W., Weissman, R.F., Rodriguez, J.E. et al. Engineered TnpB genome editors for plants and human cells identified by ribonucleoprotein mutational scanning. *Nat Biotechnol* (2026).
63. Wang, J.C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Reviews Molecular Cell Biology* 3, 430-440 (2002).
64. Wang, J.C. DNA topoisomerases. *Annu Rev Biochem* 65, 635-92 (1996).
65. Wang JY, Doudna JA. CRISPR technology: A decade of genome editing is only the beginning. *Science*. 2023 Jan 20;379(6629):eadd8643.
66. Wang Y, Jiang H, Li M, Xu Z, Xu H, Chen Y, Chen K, Zheng W, Lin W, Liu Z, Lin Z, Zhang M. Delivery of CRISPR/Cas9 system by AAV as vectors for gene therapy. *Gene*, 927:148733 (2024).
67. Wei, Y. et al. Engineering eukaryotic transposon-encoded Fanzor2 system for genome editing in mammals. *Nat Chem Biol* 22, 48-57 (2026).
68. Weiss, T. et al. Viral delivery of an RNA-guided genome editor for transgene-free germline editing in Arabidopsis. *Nature Plants* (2025).
69. Wiegand, T. et al. TnpB homologues exapted from transposons are RNA-guided transcription factors. *Nature* 631, 439-448 (2024).
70. Wiggins, P.A. An information-based approach to change-point analysis with applications to biophysics and cell biology. *Biophys J* 109, 346-54 (2015).
71. Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Meyer BJ. Targeted genome editing across species using ZFNs and TALENs. *Science*. 2011 Jul 15;333(6040):307
72. Xiang, G. et al. Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nature Biotechnology* 42, 745-757 (2024).
73. Xue, N. et al. Engineering IscB to develop highly efficient miniature editing tools in mammalian cells and embryos. *Mol Cell* 84, 3128-3140 e4 (2024).
74. Yan, H. et al. Assessing and engineering the IscB-omegaRNA system for programmed genome editing. *Nat Chem Biol* 20, 1617-1628 (2024).
75. Yoo, S.D., Cho, Y.H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2, 1565-72 (2007).
76. Yoon, P.H. et al. Eukaryotic RNA-guided endonucleases evolved from a unique clade of bacterial enzymes. *Nucleic Acids Research* 51, 12414-12427 (2023).

77. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I.M., Makarova, K.S., Esslezbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V. & Zhang, F., Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 2015; 163, 759-771
78. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614-20 (2014).
79. Zhang, R. et al. IsDge10 is a hypercompact TnpB nuclease that confers efficient genome editing in rice. *Plant Communications* 5 (2024).
80. Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W. & Chua, N.H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* 1, 641-6 (2006).
81. Zhang, Y. & Dudko, O.K. A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc Natl Acad Sci U S A* 110, 16432-7 (2013).
82. Zhao, G. et al. Engineered hypercompact Fanzor-omegaRNA system with enhanced genome editing activity. *Mol Cell* 85, 4138-4151 e4 (2025).