Comprehensive deletion scan of anti-CRISPR AcrIIA4 reveals essential and dispensable domains for Cas9 inhibition

Authors

Annette B Iturralde[1,2], Cory A Weller[1,3], Meru J Sadhu[1]

[1] Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

[2] Present address: Biomedical Sciences Graduate Program, University of Virginia, Charlottesville, Virginia, USA

[3] Present address: Center for Alzheimer's and Related Dementias, National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA

Email: meru.sadhu@nih.gov

Running title

Systematic deletion scan of AcrIIA4

Abstract

Delineating a protein's essential and dispensable domains provides critical insight into how it carries out its function. Here, we developed a high-throughput method to synthesize and test the functionality of all possible in-frame and continuous deletions in a gene of interest, enabling rapid and unbiased determination of protein domain importance. Our approach generates precise deletions using a CRISPR library framework that is free from constraints of gRNA target site availability and efficacy. We applied our method to AcrIIA4, a phage-encoded anti-CRISPR protein that robustly inhibits SpCas9. Extensive structural characterization has shown that AcrIIA4 physically occupies the DNA-binding interfaces of several SpCas9 domains; nonetheless, the importance of each AcrIIA4 interaction for SpCas9 inhibition is unknown. We used our approach to determine the essential and dispensable regions of AcrIIA4. Surprisingly, not all contacts with SpCas9 were required, and in particular, we found that the AcrIIA4 loop that inserts into SpCas9's RuvC catalytic domain can be deleted. Our results show that AcrIIA4 inhibits SpCas9 primarily by blocking PAM binding, and that its interaction with the SpCas9 catalytic domain is inessential.

Keywords

AcrIIA4 / anti-CRISPR / deletion / MAVE / SpCas9

41 Introduction

42

43 CRISPR systems consist of RNA-guided endonucleases that cleave nucleic acid
44 sequences specified by the variable spacer sequence of a bound guide RNA (gRNA)
45 (1). CRISPR systems are widespread in bacteria and archaea, where they can provide
46 adaptive immunity against viruses, plasmids, and other mobile genetic elements
47 (MGEs) through gRNAs that recognize foreign DNA or RNA and target their cleavage.
48 The adaptability of CRISPR endonucleases to cleave specific DNA sequences or
49 otherwise recruit proteins to specific genomic sites is used in myriad ways in molecular
50 biology research, predominantly featuring the type IIA Cas9 protein from *Streptococcus*
51 *pyogenes*, referred to as SpCas9. gRNA-bound SpCas9 cleaves double-stranded DNA
52 if its sequence matches the gRNA's 20-nucleotide (nt) spacer sequence followed by the
53 protospacer motif (PAM) sequence of NGG.

54 To counter the native role of CRISPR in host defense, MGEs encode a variety of
55 anti-CRISPR proteins, or Acrs (2). Acrs have diverse strategies to block CRISPR
56 cleavage, including through disruption of gRNA function, DNA binding, or DNA
57 cleavage. Acrs have been repurposed technologically to regulate CRISPR, including to
58 restrict CRISPR activity to a particular tissue type or to reduce cytotoxicity associated
59 with CRISPR (3, 4). One of the most well-characterized Acrs is AcrIIA4, a potent
60 inhibitor of SpCas9 identified from a *Listeria monocytogenes* prophage (5). AcrIIA4 is a
61 compact 87-amino acid protein composed of a central three-stranded β-sheet flanked
62 by three α-helices (6–8), as well as a 3_{10} helix in the β1-β2 loop that deforms upon
63 SpCas9 binding (9). AcrIIA4 occupies the DNA-binding interfaces of several domains of

64 SpCas9 (Fig. 1A). AcrIIA4 binds the SpCas9 C-terminal domain (CTD), which binds the
65 PAM sequence of target DNA, using its β3 strand, α3 helix, and β2-β3, β3-α2, and α2-
66 α3 loops. Residues in AcrIIA4's α1-β1 and β2-β3 loops interact with the SpCas9
67 topoisomerase-homology (TOPO) domain, which aids in PAM binding. Finally, the β1-
68 β2 loop of AcrIIA4 occupies the active site of SpCas9's RuvC domain, which cleaves
69 the nontarget DNA strand. However, the individual importance of these interactions to
70 SpCas9 inhibition has not been determined.

71 To precisely determine which domains of AcrIIA4 are essential for anti-CRISPR
72 function, we generated nearly all possible in-frame deletions of all sizes in AcrIIA4 and
73 tested their impact on SpCas9 inhibition. Functional domains of proteins have been
74 mapped for decades by determining the effects of deleting segments of the protein (10).
75 Augmenting such approaches to assay all possible in-frame deletions enables full and
76 precise delineation of a protein's essential and dispensable domains in an unbiased
77 fashion (11), but creating all possible sizes of deletions across a gene scales
78 quadratically with gene size and quickly amounts to thousands of deletions,
79 necessitating a high-throughput approach. CRISPR technology can be used for high-
80 throughput genetic interrogation, by synthesizing a library of DNA molecules encoding
81 gRNAs and/or repair templates and using it to generate genetic variants in a pool of
82 cells that contains CRISPR machinery (12–14). Here, we developed a high-throughput
83 pooled CRISPR approach to comprehensively determine the effects of deletions in a
84 protein and applied it to AcrIIA4. We find that the AcrIIA4 regions responsible for
85 blocking the PAM-interacting domains of SpCas9 are essential, whereas the segment
86 involved in blocking the RuvC catalytic domain is dispensable.

87

88 Results

89

90 We developed an approach for comprehensive deletion mutagenesis that uses
91 CRISPR in the budding yeast *Saccharomyces cerevisiae* (Fig 1B). We used
92 synonymous variants to generate two alleles of *acrIIA4* with maximum DNA sequence
93 divergence while reflecting yeast codon usage, resulting in 61% DNA identity. We
94 inserted the two *acrIIA4* alleles in tandem into the yeast genome at the *HO* locus,
95 placing them after a galactose-inducible promoter and placing a synthetic Cas9 target
96 site and the *URA3* gene from *Kluyveromyces lactis* between them. To generate an
97 *acrIIA4* allele carrying a deletion between any pair of codons n_1 and n_2, we direct
98 CRISPR cleavage at the synthetic target site between the *acrIIA4* copies and provide a
99 repair template with homology to the first acrIIA4 copy up to codon n_1, followed by
100 homology to the second copy starting from codon n_2. Thus, the use of this repair
101 template during homology-directed repair produces a single fusion *acrIIA4* gene that
102 has the region from codon n_1 to codon n_2 deleted. The fusion event also deletes the
103 CRISPR target site and *URA3*, precluding further cutting and allowing for the selection
104 of cells with successful deletions through their ability to grow on media containing 5-
105 fluoroorotic acid (5-FOA), which is toxic to *S. cerevisiae* cells expressing *URA3* (15).
106 There are 3,741 possible unique in-frame *acrIIA4* deletions. To generate a
107 comprehensive library of deletions we used large-scale array-based synthesis to
108 synthesize a pool of all possible deletion-directing repair templates, composed of 100 nt
109 matching the first *acrIIA4* copy followed by 100 nt matching the second. We cloned the

110 repair template oligos into three replicate 2-micron plasmid libraries for high-copy
111 maintenance in yeast cells. Massively parallel sequencing of the repair templates from
112 two of the resulting plasmid libraries revealed 3,573 (95%) repair templates were
113 represented in at least one of the replicate plasmid pools. The plasmid pools also
114 encoded a gRNA targeting the novel sequence inserted between the *acrIIA4* genes,
115 which was designed to have at least six mismatches with any other SpCas9 target site
116 in the *S. cerevisiae* genome. Note that as the gRNA target site is constant, this
117 approach is not constrained by the density of gRNA target sites in the *acrIIA4* gene, nor
118 is there any aspect of variability in gRNA efficacy.

119 We transformed the repair template library into the *acrIIA4-URA3-acrIIA4* yeast
120 strain. The strain also carried two copies of SpCas9, one under a *MAL62* promoter and
121 the other under a *GAL*L promoter, allowing SpCas9 to be activated by either maltose or
122 galactose (16–18). Both promoters are repressed by glucose. Transformed cells were
123 selected on glucose plates and then re-plated on maltose to induce the maltose-
124 responsive SpCas9 to generate deletion-carrying *acrIIA4* alleles. Resulting colonies
125 were then grown on 5-FOA plates to select cells that had deleted *URA3* and thus were
126 carrying the designed *acrIIA4* deletion alleles.

127 We sequenced the *acrIIA4* alleles present in the 5-FOA-resistant population by
128 massively parallel sequencing. We recovered 3,305 distinct programmed deletions
129 across the three replicates, 88% of the total possible deletions (Fig. 2). Deletions were
130 recovered across the entire range of sizes.

131 Next, we determined the ability of the *acrIIA4* deletion alleles to inhibit SpCas9
132 activity. We generated a plasmid expressing a gRNA that robustly targets the essential

133 gene *ARB1* (14); usage of this gRNA for CRISPR cutting is lethal. To identify *acrIIA4*
134 deletion alleles that maintained anti-CRISPR function, we transformed the *ARB1*-
135 targeting gRNA plasmid into our pool of cells with programmed *acrIIA4* deletions and
136 induced SpCas9 and *acrIIA4* expression on galactose plates. Approximately 30% as
137 many colonies survived CRISPR induction compared to glucose plates without CRISPR
138 expression. We used massively parallel sequencing of the *acrIIA4* genomic locus to
139 identify the *acrIIA4* alleles present in survivors. Read counts were normalized against
140 the count of each allele from the uninduced glucose condition (Fig. 3A).
141 We observed three distinct clusters of AcrIIA4 deletions that maintain function, in the
142 β1-β2 loop, the β3-α2 loop, and the end of the α3 helix (Fig. 3A, Supplementary Fig. 1).
143 Given the contacts between the β1-β2 loop and the catalytic site of the RuvC domain,
144 the loop's tolerance of deletions was notable. The loop is eight amino acids long; we
145 found that AcrIIA4 retained anti-CRISPR function even when six of these amino acids
146 were deleted (Fig. 3B). As this comprises the major RuvC-interacting surface, we
147 propose that RuvC binding is dispensable for AcrIIA4's anti-CRISPR function.
148 AcrIIA4 makes contacts with the SpCas9 CTD through residues in the β3 strand, α3
149 helix, and β2-β3, β3-α2, and α2-α3 loops. In CRISPR targeting, arginines Arg1333 and
150 Arg1335 of the SpCas9 CTD engage the two guanines in the NGG PAM sequence.
151 Residues 67-70 in the α2-α3 loop and α3 helix of AcrIIA4, which includes three
152 negatively charged residues, interact with the positively charged arginines (6).
153 Interestingly, this segment of AcrIIA4 tolerated single amino-acid deletions, but not any
154 larger deletions (Fig. 3C). We propose that this interaction is required for SpCas9
155 inhibition and can tolerate the removal of a single negative charge from AcrIIA4, but no

156 more. Supporting the importance of this interaction, we see that Asn39 in the β2-β3
157 loop, which also contacts Arg1335, was not dispensable. In contrast, the β3-α2 loop,
158 which contacts a distal component of the SpCas9 CTD, was a hotspot for tolerated
159 deletions and could be shortened by up to four amino acids (Fig. 3D). This suggests
160 that the critical CTD contacts are precisely the ones at the PAM-interacting interface.
161 AcrIIA4 makes relatively fewer contacts to SpCas9's TOPO domain, which is also
162 involved in PAM binding. TOPO residues that interact with the PAM are contacted by
163 the AcrIIA4 residues Asp14 and Asn36 (6), both of which were partially dispensable
164 (Fig. 3E). We noted that Asn36 is adjacent to Asn35 in the β2-β3 loop, and while each
165 asparagine was individually partially dispensable, simultaneous deletion of both was not
166 tolerated. Thus, binding of TOPO by AcrIIA4 appears to be important for AcrIIA4
167 function, with functional redundancy buffering the loss of individual contacts.
168 We observed a large number of tolerated deletions towards the end of the C-
169 terminal helix α3 (Fig. 3E), with up to six terminal residues being dispensable. This
170 segment does not make any contacts with SpCas9 and is largely separate from the rest
171 of AcrIIA4. Consistent with the C-terminus being mutable, a C-terminal GFP fusion does
172 not disrupt anti-CRISPR activity (19). In contrast, deletions of seven or more terminal
173 residues were not tolerated. These deletions remove Ile80, which contributes to the
174 hydrophobic core of AcrIIA4 along with Leu6, Ile10, and Phe55. Intriguingly, we
175 observed that three six-residue deletions encompassing Ile80 retained partial function in
176 all three replicates: Δ78-83, Δ79-84, and Δ80-85 (Supplementary Fig. 2). These
177 deletions move Leu86 to position 80, which we propose partially reconstitutes Ile80's
178 role in AcrIIA4 stability.

179 We were curious whether the dispensable segments we identified in AcrIIA4 were
180 ever absent in naturally occurring AcrIIA4 homologs. We identified 124 homologs of
181 AcrIIA4 through a BLASTp search of the NCBI sequence database (20). Aligning these
182 to AcrIIA4 revealed that 48 had gaps relative to AcrIIA4, falling into six unique patterns
183 (Fig. 4A). Strikingly, most observed alignment gaps were found in regions we identified
184 as dispensable. These included gaps in the β1-β2 and β3-α2 loops and truncations of
185 the C-terminus. The largest gap was seen in several homologs that had a β1-β2 loop
186 five amino acids shorter than in AcrIIA4, consistent with our observation that the AcrIIA4
187 β1-β2 loop could be shortened by up to six amino acids while maintaining anti-CRISPR
188 function.
189 One alignment gap was present in a non-dispensable segment in the AcrIIA4
190 homolog from *Listeria innocua* (NCBI ID WP_185342258.1, hereafter referred to as
191 $AcrIIA4_{Lin}$), which appeared to have a one-amino acid shortening of the indispensable
192 C-terminal end of the β2 strand, in addition to a five-amino acid shortening of the
193 dispensable β1-β2 loop. We generated an Alphafold model of the $AcrIIA4_{Lin}$ protein
194 structure (Fig. 4B) and found that the β2 strand in the model was five amino acids long,
195 the same length as in AcrIIA4, while the modeled β1-β2 loop in $AcrIIA4_{Lin}$ was a full six
196 amino acids shorter than in AcrIIA4, consistent with our observed deletion tolerances.
197
198
199 Discussion
200

201 AcrIIA4 is a robust inhibitor of SpCas9. It physically interacts with SpCas9 through
202 an extensive interface that involves several distinct SpCas9 domains, including domains
203 involved in PAM recognition (CTD and TOPO) and DNA cleavage (RuvC), though the
204 importance of each individual interaction to SpCas9 inhibition was unknown. To
205 elucidate the mechanism of AcrIIA4 inhibition of SpCas9, we comprehensively deleted
206 segments of AcrIIA4 and assayed their function. We found that the interaction with the
207 RuvC domain, mediated by the β1-β2 loop of AcrIIA4, is dispensable, as this eight-
208 amino acid loop could be shortened by up to six amino acids without loss of anti-
209 CRISPR activity. Our finding that AcrIIA4 does not need to occupy the RuvC active site
210 to inhibit SpCas9 is consistent with the fact that AcrIIA4 robustly inhibits dCas9 from
211 binding DNA, which necessarily precedes DNA cleavage (8). Still, it remained possible
212 that AcrIIA4 binding to the RuvC domain contributed to the binding energy required to
213 block DNA binding. Based on our findings, binding to Cas9's PAM-interacting domains
214 is sufficient.
215 Though the physical interactions between AcrIIA4 and SpCas9 are largely mediated
216 by the loops between secondary structural elements, we observed that none of the α
217 helices and β strands of AcrIIA4 were fully dispensable. This likely reflects
218 intramolecular interactions necessary for AcrIIA4 stability, matching the observation that
219 evolutionarily observed indels rarely occur in α helices and β strands (21). Consistent
220 with this hypothesis, we observed that partial deletions of helices α1 and α3 were
221 tolerated. These helices are at the very start and end of the protein, and the tolerated
222 deletions were not near the AcrIIA4 core, so deletion of these segments would not be

223 predicted to disrupt AcrIIA4's structure. Deletion scans of other proteins have similarly
224 found higher tolerance for deletions at the N- and C-termini of domains (22)..
225 Interestingly, we found several segments that could not be deleted despite being
226 composed of smaller dispensable segments, as has been observed previously in other
227 proteins (23). This could occur through functional redundancy, as we propose for amino
228 acids 67-70, which has several negatively charged residues and can tolerate single
229 amino acid deletions, but no further. Alternatively, a segment could have no direct role
230 in protein function, but could link functional components that need to be a minimum
231 distance apart. For instance, while multiple deletions of six amino acids were tolerated
232 in the eight-amino acid β1-β2 loop, deletions of seven amino acids were not;
233 presumably such deletions disrupt the relative positioning of strands β1 and β2. The
234 ubiquity of non-additive deletion effects highlights the utility of comprehensively testing
235 deletions, which can reveal functional redundancy and dispensable segments that could
236 be missed with a low-throughput approach.
237
238 Comprehensive scanning of deletion effects has many applications in addition to
239 mapping proteins' essential and dispensable segments. It can also be used in genetic
240 engineering to produce improved or novel protein function, as indels have been shown
241 to have larger effects on protein function than missense variants (24). Similarly,
242 deletions are an important source of evolutionary novelty (25, 26), so deletion scanning
243 is useful to understand evolutionary trajectories. Deletion scanning can also be used in
244 clinical genetics to characterize potential disease-causing deletion alleles.
245 Understanding the effects of deletion variants is essential for interpreting the risk of

246 genetic disease from genome sequences: after single-nucleotide variants, deletions are
247 the most common type of variant on ClinVar (27), and each individual person is
248 predicted to harbor approximately 200 in-frame insertions or deletions relative to the
249 human reference genome (28).

250 Our strategy for deletion scanning involved chimerizing two copies of *acrIIA4*
251 following CRISPR cutting between them. A previous study used CRISPR to generate
252 cuts within a single copy of yeast *SGS1*, which were repaired using repair templates
253 that directed 20-amino acid deletions across the gene body (29). By cutting between
254 two copies of our target gene, we did not require PAM sequences close to the regions
255 being deleted, allowing us to comprehensively characterize deletions of small to large
256 size. In addition, our strategy uses a constant gRNA, which avoids issues arising from
257 variable gRNA targeting efficiency. Finally, by concurrently deleting a *URA3* gene
258 between the target gene copies, we could select for cells carrying deletion alleles.

259 Deletion scans have also been carried out using non-CRISPR approaches. One
260 strategy is to use oligonucleotide synthesis technology to directly generate deletion
261 alleles, rather than to generate repair templates that direct deletions following CRISPR
262 cutting. As oligonucleotides synthesized at large scale are currently limited to lengths of
263 300 nt or fewer, to date this approach has been applied to short genes, segments of
264 genes, or small deletions within larger genes (11, 22, 23, 26). Another strategy is to
265 introduce restriction sites into the gene of interest, cleave, and then self-anneal to
266 produce truncated gene products. For deletion scanning methods that use this
267 framework (24, 30), a challenge is that novel sequences are inserted at the deletion
268 junction for most or all of the deletion alleles.

269 An inherent challenge to the concept of deletion scanning is that the number of
270 potential deletions scales quadratically with the size of the protein, which could
271 theoretically be impractical for a large gene even when using a high-throughput
272 approach. We note that oligonucleotide library synthesis currently can produce up to
273 244,000 sequences in a single order. With our approach, that would allow
274 comprehensive deletion scans of proteins up to 700 residues, encompassing most
275 proteins. More importantly, our approach allows specific deletions to be programmed,
276 such as deletions under a certain size or deletions within predicted domains or
277 secondary structure elements. Screening select deletions would lead to a linear scaling
278 with protein length and thus prevent the number of tested deletions from becoming
279 unmanageable.
280 For applications other than mapping essential domains, such as determining the
281 evolutionary or pathogenic potential of genetic variants, it can be interesting to test
282 insertions as well as deletions (11, 26). Our method could be easily adapted to
283 additionally generate internal duplicated segments, which are the most common form of
284 naturally occurring insertion mutations (31), by having repair templates in which the start
285 point in the second gene copy comes before the end point of the first copy.
286 Homologous proteins often have divergent functions, and chimeras between
287 them can map the causative sequence differences. For instance, the yeast GTPases
288 Sec4 and Ypt1 have different functions in the secretory pathway that were mapped to
289 differences in loops 2 and 7 (32, 33). Chimeras have also been used to study orthologs
290 between species, such as the human and macaque homologs of the innate immune
291 protein TRIM5α, whose differences in HIV restriction were mapped to a loop in the

292 SPRY domain (34–37). Our method could be readily adapted to generate chimeras
293 between two different genes, rather than to make deletions. Generating and
294 characterizing chimeras in high throughput would enable rapid fine-resolution
295 determination of specific amino acids causing functional differences between homologs.
296

297

298 References

299 1. M. Jinek, *et al.*, Structures of Cas9 Endonucleases Reveal *RNA*-Mediated
300 Conformational Activation. *Science* **343**, 1247997 (2014).
301 2. A. R. Davidson, et al., Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems.
302 *Annu. Rev. Biochem.* **89**, 1–24 (2020).
303 3. J. Lee, et al., Tissue-restricted genome editing in vivo specified by microRNA-
304 repressible anti-CRISPR proteins. RNA 1421–1431 (2019).
305 4. C. Li, et al., HDAd5/35++ Adenovirus Vector Expressing Anti-CRISPR Peptides
306 Decreases CRISPR/Cas9 Toxicity in Human Hematopoietic Stem *Cell*s. *Mol. Ther. -*
307 *Methods Clin. Dev.* **9**, 390–401 (2018).
308 5. B. J. Rauch, et al., Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell **168**,
309 150-158.e10 (2017).
310 6. D. Dong, et al., Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR
311 protein. *Nature* **546**, 436–439 (2017).
312 7. H. Yang, D. J. Patel, Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4
313 Targeting SpyCas9. *Mol. Cell* **67**, 117-127.e5 (2017).

314 8. J. Shin, et al., Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci. Adv.* **3**, e1701620
315 (2017).
316 9. I. Kim, et al., Solution structure and dynamics of anti-CRISPR AcrIIA4, the Cas9
317 inhibitor. *Sci. Rep.* **8**, 3883 (2018).
318 10. S. Munro, H. R. B. Pelham, Use of peptide tagging to detect proteins expressed
319 from cloned genes: deletion mapping functional domains of Drosophila hsp70. *The*
320 *EMBO Journal* **3**, 3087–3093 (1984).
321 11. M. Seuma, B. Lehner, B. Bolognesi, An atlas of amyloid aggregation: the impact of
322 substitutions, insertions, deletions and truncations on amyloid beta fibril nucleation. *Nat.*
323 *Commun.* **13**, 7084 (2022).
324 **12**. O. Shalem, et al., Genome-Scale CRISPR-Cas9 Knockout Screening in Human
325 Cells. Science 343, 84–87 (2014).
326 13. G. M. Findlay, et al., Accurate classification of BRCA1 variants with saturation
327 genome editing. Nature **562**, 217–222 (2018).
328 14. M. J. Sadhu, et al., Highly parallel genome variant engineering with CRISPR–Cas9.
329 *Nat. Genet.* **50**, 510–514 (2018).
330 15. J. D. Boeke, F. L. Croute, G. R. Fink, A positive selection for mutants lacking
331 orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance.
332 *Mol. Gen. Genet. MGG* **197**, 345–346 (1984).
333 16. A. W. Gibson, et al., Constitutive Mutations of the Saccharomyces cerevisiae MAL-
334 Activator Genes MAL23, MAL43, MAL63, and mal64. *Genetics* **146**, 1287–1298 (1997).
335 17. R. L. Finley, H. Zhang, J. Zhong, C. A. Stanyon, Regulated expression of proteins
336 in yeast using the MAL61–62 promoter and a mating scheme to increase dynamic

range. Gene **285**, 49–57 (2002).

18. C. A. Weller, et al., Highly complete long-read genomes reveal pangenomic variation underlying yeast phenotypic diversity. *Genome Res* gr.277515.122 (2023). https://doi.org/10.**11**01/gr.277515.122.

19. E. M. Basgall, et al., Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology*+ **164**, 464–474 (2018).

20. E. W. Sayers, et al., Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **52**, D33–D43 (2023).

21. S. Pascarella, P. Argos, Analysis of insertions/deletions in protein structures. *J. Mol. Biol.* **224**, 461–471 (1992).

22. M. Topolska, A. Beltran, B. Lehner, Deep indel mutagenesis reveals the impact of insertions and deletions on protein stability and function. *bioRxiv* 2023.10.06.561180 (2023). https://doi.org/10.1101/2023.10.06.561180.

23. C. B. Macdonald, et al., DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *Genome Biol.* **24**, 36 (2023).

24. S. Emond, et al., Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. Nat. Commun. 11, 3469 (2020).

25. K. R. McCarthy, et al., Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science **371**, 1139–1142 (2021).

26. J. Tenthorey, et al., Indels allow antiviral proteins to evolve functional novelty inaccessible by missense mutations. (2024). https://doi.org/10.1101/2024.05.07.592993.

360 27. M. J. Landrum, et al., ClinVar: improving access to variant interpretations and
361 supporting evidence. *Nucleic acids Res.* **46**, D1062–D1067 (2017).
362 28. R. M. Durbin, et al., A map of human genome variation from population-scale
363 sequencing. Nature **467**, 1061–1073 (2010).
364 29. X. Guo, et al., High-throughput creation and functional profiling of DNA sequence
365 variant libraries using CRISPR–Cas9 in yeast. *Nat Biotechnol* **36**, 540–546 (2018).
366 30. A. Shams, et al., Comprehensive deletion landscape of CRISPR-Cas9 identifies
367 minimal RNA-guided DNA-binding modules. Nat. Commun. 12, 5664 (2021).
368 31. S. B. Montgomery, et al., The origin, evolution, and functional impact of short
369 insertion–deletion variants identified in 179 human genomes. *Genome Res.* **23**, 749–
370 761 (2013).
371 32. B. Dunn, T. Stearns, D. Botstein, Specificity domains distinguish the Ras-related
372 GTPases Ypt1 and Sec4. Nature **362**, 563–565 (1993).
373 33. P. Brennwald, P. Novick, Interactions of three domains distinguishing the Ras-
374 related GTP-binding proteins Ypt1 and Sec4. Nature 362, 560–563 (1993).
375 34. M. Stremlau, M. Perron, S. Welikala, J. Sodroski, Species-Specific Variation in the
376 B30.2(SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency
377 Virus Restriction. *J. Virol.* **79**, 3139–3145 (2005).
378 35. M. W. Yap, S. Nisole, J. P. Stoye, A Single Amino Acid Change in the SPRY
379 Domain of Human Trim5α Leads to HIV-1 Restriction. *Curr. Biol.* **15**, 73–78 (2005).
380 36. S. L. Sawyer, L. I. Wu, M. Emerman, H. S. Malik, Positive selection of primate
381 TRIM5α identifies a critical species-specific retroviral restriction domain. *Proc National*
382 *Acad Sci* **102**, 2832–2837 (2005).

383 37. D. Perez-Caballero, T. Hatziioannou, A. Yang, S. Cowan, P. D. Bieniasz, Human
384 Tripartite Motif 5α Domains Responsible for Retrovirus Restriction Activity and
385 Specificity. J. Virol. 79, 8969–8978 (2005).
386 38. D. M. Becker, V. Lundblad, Introduction of DNA into yeast cells. *Curr. Protoc. Mol.*
387 *Biol.* **Chapter 13**, Unit13.7 (2001).
388 39. D. Mumberg, R. Muller, M. Funk, Regulatable promoters of Saccharomyces
389 cerevisiae: comparison of transcriptional activity and their use for heterologous
390 expression. Nucleic Acids Res 22, 5767–5768 (1994).
391 40. K. A. Curran, et al., Short Synthetic Terminators for Improved Heterologous Gene
392 Expression in Yeast. *Acs Synth Biol* **4**, 824–832 (2015).
393 41. M. Valencia, et al., NEJ1 controls non-homologous end joining in Saccharomyces
394 cerevisiae. Nature **414**, 666–9 (2001).
395 42. J. Jumper, et al., Highly accurate protein structure prediction with AlphaFold.
396 Nature **596**, 583–589 (2021).
397 43. E. C. Meng, et al., UCSF ChimeraX : Tools for structure building and analysis.
398 *Protein Sci.* **32**, e4792 (2023).

399

400

401 **Methods**

402 **Generation of base strains and plasmids**

403 All strains, plasmids, primers, and oligonucleotides are included in Supplementary
404 tables 1-5.

405 All plasmids were transformed into NEB 5-alpha Competent *Escherichia coli* cells (New
406 England Biolabs) and purified using Qiaprep Spin Miniprep Kits or Plasmid Plus Maxi
407 Kits for plasmid libraries (QIAGEN). All PCR products were purified using QIAquick
408 PCR Purification Kits or MinElute PCR Purification Kits (QIAGEN). Plasmid and
409 genomic sequences were verified through Sanger sequencing (Eurofins Genomics,
410 Louisville, KY, USA). All yeast transformations were done according to standard lithium
411 acetate transformation procedures (38).

412 A plasmid carrying the two *acrIIA4* alleles was synthesized by Genewiz (MSp513). The
413 sequence of the second *acrIIA4* allele matched the sequence used by Basgall, et al.
414 (19). The sequence of the first *acrIIA4* allele was generated using custom python scripts
415 in which each codon was selected to share as few bases as possible with the
416 corresponding codon of the second *acrIIA4* allele. When multiple codons met that
417 criterion, the codon was randomly selected from the available codons with weights
418 matching codon usage in the *S. cerevisiae* genome. The first *acrIIA4* allele was
419 preceded by the *GAL*L promoter (39) and the second was followed by the T_{synth8}
420 terminator (40). The *acrIIA4* genes were flanked by 120-bp homology arms to the *HO*
421 locus. This Acr construct was amplified using PfuUltraII fusion HS DNA Polymerase
422 (Agilent Technologies) with the primers ANx87 and ANx88 and was transformed into

423 strain MS262 (MSy15), which had *NEJ1* deleted to eliminate non-homologous end-
424 joining as a repair pathways for DSBs (41) in favor of homology-directed repair
425 (Supplementary Table 1). To increase transformation efficiency, during the
426 transformation we directed a double-stranded DNA break at the *HO* locus by co-
427 transforming the Acr construct with a plasmid constitutively expressing SpCas9 from the
428 *TEF* promoter, acquired from Addgene (plasmid #: 162606; designated MSp181 in
429 Supplementary Table 2), and a PCR construct encoding a gRNA targeting the yeast *HO*
430 locus. The gRNA construct was generated by amplifying the targeting sequence from
431 pMS58 (MSp20) with primers ANx84 and ANx85 (Supplementary Table 3). Positive
432 transformants of the *acrIIA4* construct into the *HO* locus were selected by growth on
433 plates lacking uracil and leucine (CSM -Ura -Leu; Sunrise Science Products) and
434 confirmed by Sanger sequencing. A sequence-verified clone, in which MSp181 had
435 been spontaneously lost, was designated MSy261.

436 MSy261 was used to perform the first deletion scan replicate experiment, as described
437 in the section “Creation of acrIIA4 deletion alleles”. However, plasmids in the repair-
438 template plasmid library (whose generation is described in the section “Design of repair
439 templates”) that contained the sequence of codons 26-35 of the first copy of *acrIIA4*
440 were strongly depleted from the plasmid library for unclear reasons, presumably related
441 to that sequence posing a challenge either for oligonucleotide synthesis or library
442 cloning. We thus generated a second strain with these codons re-coded, MSy262,
443 which was used for the second and third deletion scan replicate experiments
444 (Supplementary Fig. 3). This was done by amplifying two segments of the *acrIIA4*
445 construct on MSy261, upstream and downstream of codons 26-35, using Herculase II

446 Fusion DNA Polymerase (Agilent) and introducing the alternate sequence by including it
447 in the primers used for amplification (ANx80-83). To make MSy262, the two amplicons
448 were transformed into MSy14, which carries a galactose-inducible SpCas9 plasmid,
449 with the *HO*-targeting gRNA described above, and immediately plated onto CSM -Leu -
450 Ura galactose plates.

451 Our strategy of introducing deletions into *acrIIA4* involved cutting with SpCas9 between
452 the *acrIIA4* copies. For that step, we generated a maltose-inducible SpCas9 plasmid to
453 induce SpCas9 expression while maintaining repression of the galactose-inducible
454 *acrIIA4*, to prevent the anti-CRISPR activity of acrIIA4 from inhibiting the generation of
455 deletions. We inserted a DNA construct containing the *MAL63* transcription factor gene,
456 *MAL61* maltose transporter gene, and *MAL62* promoter sequence, ordered from
457 Genewiz (MSp514), upstream of the SpCas9 gene on MSp14. We named the resulting
458 plasmid MSp500. The sequence of the *MAL* genes was taken from the genome
459 sequence of YJM145 (MSY28) (18). Amplification of the *MAL* locus sequence was done
460 with KAPA Hi-Fi Hotstart DNA Polymerase (and primers ANx16.5 and 17.5), and
461 inserted into MSp14 via Gibson assembly (New England Biolabs, Ipswich, MA, USA)
462 following a digestion of the plasmid with SpeI-HF (New England Biolabs).

463 MSy261 was transformed with a plasmid carrying a galactose-inducible SpCas9, p415-
464 GalL-Cas9-CYC1t (Addgene plasmid #:107922; MSp6) and transformants were
465 selected on CSM -Ura -Leu glucose plates. One colony was picked and transformed
466 with MSp500, with transformants selected on CSM -Ura -Leu + clonNAT (100ug/mL)
467 glucose plates (MSy263). MSy262, already carrying a galactose-inducible SpCas9, was

468 also transformed with MSp500 and plated on CSM -Ura -Leu + clonNAT (100ug/mL)
469 glucose plates to produce MSy264.

470

471 **Design of repair templates**

472 Custom python scripts were used to generate repair template sequences for directing all
473 possible in-frame deletions within *acrIIA4*. In brief, to generate the repair template for
474 the deletion from the *m*th to the *n*th codon, we appended the 100 nt up to the *m*th codon
475 of the first *acrIIA4* allele to the 100 nt after the *n*th codon of the second *acrIIA4* allele.
476 For deletions close to the start of *acrIIA4*, their repair templates included sequence from
477 the *GAL*L promoter preceding the first copy, and for deletions close to the end, their
478 repair templates included the *acrIIA4* terminator and the sequence of the *HO* locus
479 flanking the insertion site. We included the start codon and stop codon as deletable
480 codons. We also designed repair templates that generated a clean chimera of the two
481 *acrIIA4* alleles into a single *acrIIA4* without deletion, for 85 repair templates. Once all
482 sequences were created, 15-nt primer sites were added to the beginning and end of the
483 strings, making all sequences 230 characters long. Sequences were then ordered as an
484 oligonucleotide pool through Agilent (Santa Clara, California, USA). This process was
485 performed once to generate the repair templates used for replicate 1, and then again
486 with the modified sequence of the first *acrIIA4* allele to generate the repair templates
487 used for replicates 2 and 3 (Supplementary tables 4 and 5). For the purpose of this
488 experiment, we proceeded with data analysis of 3,741 sequences, excluding deletions
489 of the start or stop codons, as well as repair templates that generated clean chimeras.

490 MSp218 was used as the backbone of the three replicate repair template plasmid pools,
491 and was generated from pMEL16 (MSp50), a *HIS3* SpCas9 gRNA plasmid acquired
492 from Jean-Marc Daran (Addgene plasmid #:107922). For cloning in the repair
493 templates, we engineered EcoRI and BamHI sites downstream of the gRNA gene. In
494 addition to carrying the repair templates, we designed MSp218 to encode a gRNA
495 sequence targeting a sequence inserted between the two *acrIIA4* alleles,
496 ACGTATCGGTACATGCGCAT. This sequence was picked to have equal
497 representation of each of the four bases, as well as to have at least six mismatches to
498 any other SpCas9 target sites in the yeast genome.

499 We amplified the repair template oligonucleotide pools using primers that additionally
500 added 12 nt of random DNA on each end to serve as barcodes as well as homology
501 arms for Gibson assembly (ANx23 and 24); PCR was performed with PfuUltra II Fusion
502 High-fidelity DNA Polymerase for 15 cycles. MSp218 was double digested using EcoRI-
503 HF and BamHI-HF (New England Biolabs) on the cloning site, and the oligo amplicons
504 were inserted via Gibson assembly. 50uL of Gibson assembly were transformed into
505 1000uL of 5-alpha competent *E. coli* cells for each replicate (New England Biolabs).
506 Replicate 1 produced 100,000 colonies, Replicate 2 produced 100,000 colonies, and
507 Replicate 3 produced 80,000 colonies. Colonies were washed off plates and plasmid
508 libraries were isolated with the Qiagen Plasmid Plus Maxi kit.

509

510 **Creation of *acrIIA4* deletion alleles**

511 The repair template plasmid pools were transformed en masse (2.05 µg of Replicate 1;
512 3.5 µg of Replicates 2 and 3) into approximately 10^9 cells of MSy263 or MSy264 as

513 appropriate, and plated on CSM -Ura -Leu -His +clonNAT glucose plates. Replicate 1
514 produced 260,000 colonies, Replicate 2 produced 300,000 colonies, and Replicate 3
515 produced 250,000 colonies. Cells were then harvested by washing the plate with
516 glucose -H,L,U +clonNAT liquid media, and a 100-mL culture of these cells was then
517 started at an optical density (OD_{600}) of 0.002 using the same media. The culture was left
518 to grow overnight at 30ºC until it reached an OD_{600} of 1. Approximately 2×10^9 cells
519 were then plated over multiple CSM Maltose -His -Leu +cloNAT plates to induce
520 SpCas9 from MSp500, resulting in 77,000 colonies for Replicate 1, 66,000 colonies for
521 Replicate 2, and 80,000 colonies for Replicate 3. Cells were harvested by washing
522 plates with YPD, and approximately 4×10^7 cells were replated onto 5-fluoroorotic acid
523 (5-FOA) -Leu -His plates. On 5-FOA plates, Replicate 1 produced 1.6×10^6 colonies,
524 Replicate 2 produced 1×10^6 colonies, and Replicate 3 produced 1×10^6 colonies.
525

526 **Characterization of *acrIIA4* deletion allele function**

527 We generated a plasmid carrying a gRNA that directs CRISPR cleavage of the essential
528 gene *ARB1* by inserting an *ARB1*-targeting gRNA sequence into pLK88 (MSp11), a
529 *URA3* high-copy plasmid from Leonid Kruglyak. An oligonucleotide containing the gRNA
530 sequence was ordered from Integrated DNA Technologies (ANx86) and inserted by
531 Gibson assembly into MSp11 following a double digestion with SphI-HF and BstEII-HF
532 (New England Biolabs). This plasmid was named MSp501.

533 Yeast cells with *acrIIA4* deletion alleles were washed off 5-FOA plates using CSM -Leu
534 -His glucose liquid media. A 50-mL culture was started at OD_{600} of 0.01 using the same
535 liquid media, and it was left to grow overnight at 30ºC. Once grown overnight, the

536 culture was used to start a new culture in YPD starting at OD_{600} of 0.125, allowed to
537 double twice, and then transformed with MSp501. In Replicate 1, 1.5 µg of MSp501 was
538 transformed into approximately 10^9 cells; in Replicates 2 and 3, 2 µg were transformed
539 into approximately 10^9 cells. Transformations were plated on CSM -Leu -His -Ura
540 glucose plates, producing 300,000 colonies in Replicate 1, 600,000 colonies in
541 Replicate 2, and 800,000 colonies in Replicate 3. Plates were then scraped once more
542 and 1×10^6 cells were plated on CSM -Leu -His -Ura galactose plates to simultaneously
543 induce the *acrIIA4* deletion alleles and SpCas9 to target *ARB1*. An equal number of
544 cells were plated on CSM -Leu -His -Ura glucose plates as controls. On glucose,
545 Replicate 1 produced 1×10^6 colonies, Replicate 2 produced 900,000 colonies, and
546 Replicate 3 produced 1×10^6 colonies. On galactose, Replicate 1 produced 300,000
547 colonies, Replicate 2 produced 280,000 colonies, and Replicate 3 produced 370,000
548 colonies.
549

550 **Data collection and sequencing**

551 Yeast genomic DNA was extracted from $4x10^7$ cells (per library) harvested off of the 5-
552 FOA plates, galactose plates, and final glucose plates using the DNeasy Blood & Tissue
553 Kit (QIAGEN). We then amplified the chimeric *acrIIA4* locus from the genomic DNA,
554 using primers containing overhangs for attaching Illumina adapters. The forward primers
555 for the genomic DNA annealed upstream of the first *acrIIA4* allele and the reverse
556 primers annealed downstream of the second allele, such that they would amplify all
557 chimeric *acrIIA4* alleles. Additionally, yeast plasmid DNA was extracted from $5x10^7$ cells
558 harvested off of galactose and final glucose plates of replicates 2 and 3 using Qiaprep

559 Spin Miniprep Kit (QIAGEN). We amplified the 200-bp repair template region from the
560 initial *E. coli* plasmid pools of replicates 1 and 2 and the yeast plasmid extractions from
561 replicates 2 and 3. Genomic primers (ANx25-34) and plasmid primers (ANx35-44) were
562 staggered to maintain sequence diversity on the Illumina flow cell. All PCRs were
563 performed using KAPA HiFi Hotstart ReadyMix DNA Polymerase for 20 cycles. A
564 second PCR was performed to attach the Nextera adapter index primers using KAPA
565 HiFi Hotstart ReadyMix DNA Polymerase for 10 cycles, after which samples were
566 prepared for sequencing according to Illumina Miseq sequencing protocols.

567 For all three replicates, we sequenced the *acrIIA4* alleles from genomic DNA of cells
568 isolated from glucose or galactose following introduction of the *ARB1*-targeting gRNA,
569 and from cells isolated after the 5-FOA selection. For Replicates 1 and 2, we
570 additionally sequenced the repair templates from plasmid DNA following cloning of the
571 repair template libraries, while for Replicates 2 and 3 we sequenced the repair
572 templates from yeast plasmid DNA following selection on galactose and glucose. Each
573 of these pools were sequenced as unique Illumina libraries using Illumina Miseq 500
574 cycle V2 kits. In total, four Miseq runs were performed: once for sequencing the
575 genomic reads and *E. coli* plasmid pool of replicate 1, once for sequencing the *E. coli*
576 plasmid pool of replicate 2, and once for sequencing the genomic and plasmid reads of
577 replicates 2 and 3.

578

579 **Data analysis**

580 PEAR software (Version 0.9.8) was used to process and align forward and reverse
581 Miseq reads. Custom python and bash scripts were used to match Miseq sequencing
582 reads of the genomic and plasmid DNA amplicons to the possible *acrIIA4* deletion
583 alleles and generate read counts of perfectly matched reads to each deletion in each
584 library. For the individual glucose and galactose genomic libraries, counts were first
585 normalized by dividing the number of reads corresponding to a specific deletion by the
586 total amount of reads in that library. We then divided all galactose values by their
587 corresponding glucose value to calculate the enrichment of *acrIIA4* deletions in cells
588 performing a lethal CRISPR edit. To merge the data from the three replicates for display
589 in Fig. 3, we took the median of the galactose/glucose read count values for a given
590 deletion across the replicates, in order to minimize the effects of escaper mutants that
591 avoided SpCas9 lethality through means other than anti-CRISPR activity. A deletion
592 was only counted as observed in a given replicate if it had at least five reads in glucose.
593 If a deletion was not observed in one replicate, the lower of the remaining two values
594 was taken. If there were no counts recorded for a given deletion in all three replicates, it
595 was represented by a blank pixel on the heat map. For comparison, we also took the
596 mean of the galactose/glucose read count values across replicates, and found that the
597 dispensability of each analyzed domain was unaffected (Supplementary Fig. 4). We
598 also found that repair templates were enriched following CRISPR induction that
599 generated AcrIIA4 deletions we found to maintain anti-CRISPR function (Supplementary
600 Fig. 5). Custom R scripts were used to generate heatmaps using R package ggplot2
601 (Version 4.2.2).

602 To determine the statistical significance of the domains that appeared to be dispensable
603 (β1-β2 loop, β3-α2 loop, and the end of the α3 helix), we calculated the mean of the
604 functional scores in the smallest triangular area encompassing the entirety of each
605 domain, for each replicate. We then acquired the mean of an equal number of randomly
606 selected data points as were contained in each triangular area and repeated this 1000
607 times (Supplementary Fig. 1) for comparison by the Monte Carlo method.

608 To find naturally occurring AcrIIA4 homologs with deletions relative to the canonical
609 AcrIIA4, we performed a BLASTp search on the NCBI BLASTp server using standard
610 parameters. We used custom R scripts to identify homologs with deletions, and
611 manually verified that the deletions did not result from incomplete sequencing.

612 All statistical measurements and visualizations were done through RStudio (Version
613 4.2.2) and ggplot2 (Version 3.4.3). Crystal structure visualizations were generated with
614 the PyMOL molecular visualization system, Version 2.5.0 and Version 3.0.0 Open-
615 Source, Schrödinger, LLC. Alphafold was used to generate the models of AcrIIA4
616 homologs (42) through ChimeraX (43).

617

618 **Data Availability**

619 - Python and R scripts:

620 https://github.com/annetteig/AcrIIA4-Functional-Domains

621

622 **Acknowledgements**

We thank Bryan Thurtle-Schmidt, Elena Zehr, Simone Giovanetti, and all other members of the Sadhu lab for helpful discussion. We thank Jean-Marc Daran, Lars Steinmetz, and Leonid Kruglyak for strains and plasmids. This work was supported by the Intramural Research Program of the National Human Genome Research Institute, NIH (1ZIAHG200401).

Author contributions

A.B.I., C.A.W., and M.J.S. designed research; A.B.I. performed formal analysis; A.B.I. performed investigation; A.B.I., C.A.W., and M.J.S. developed methodology; C.A.W. and M.J.S. performed supervision; A.B.I. and M.J.S. prepared visualizations; and A.B.I. and M.J.S. wrote the paper. All authors read and revised the paper.

Disclosure and competing interest statement

The authors declare no competing interests.

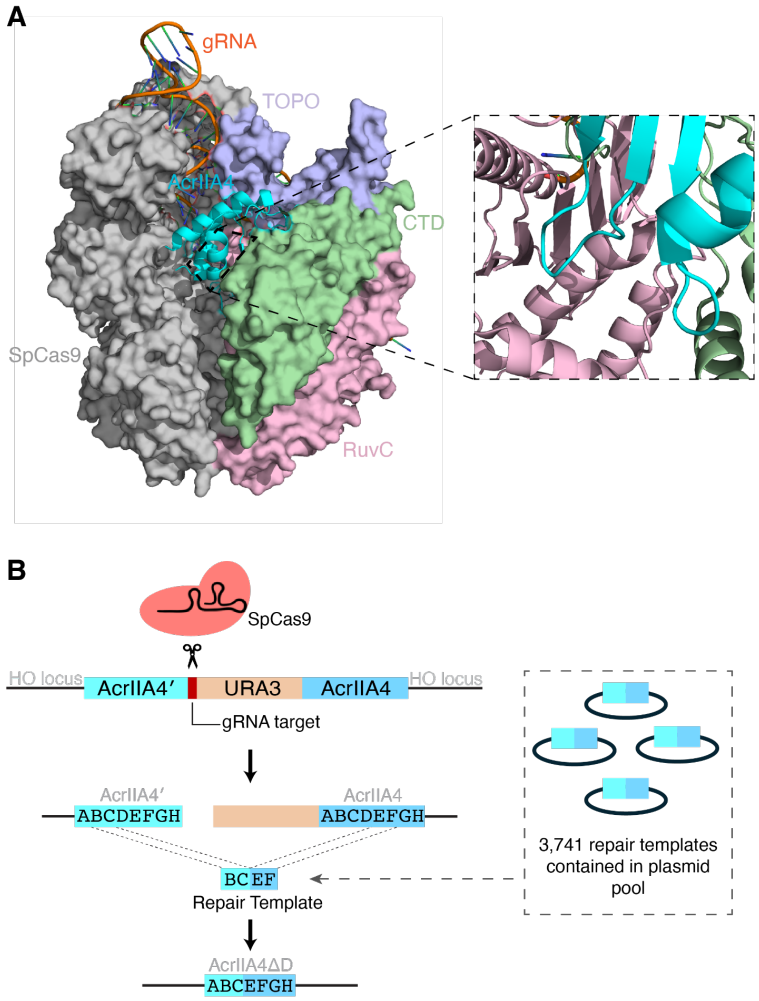


Figure 1. Comprehensive generation of in-frame deletions in AcrIIA4, an inhibitor of SpCas9. (A) Structure of AcrIIA4 bound to SpCas9 (PDB 5VW1) (7). AcrIIA4 (cyan) binds SpCas9's CTD (green), TOPO (lavender), and RuvC (pink) domains. The inset shows the AcrIIA4-RuvC binding interface, which is largely obscured in the larger view by CTD. (B) Strategy for generating all possible in-frame deletions in AcrIIA4 in yeast. CRISPR cleavage using a constant gRNA is followed by repair directed by one of 3,741 deletion-generating repair templates.

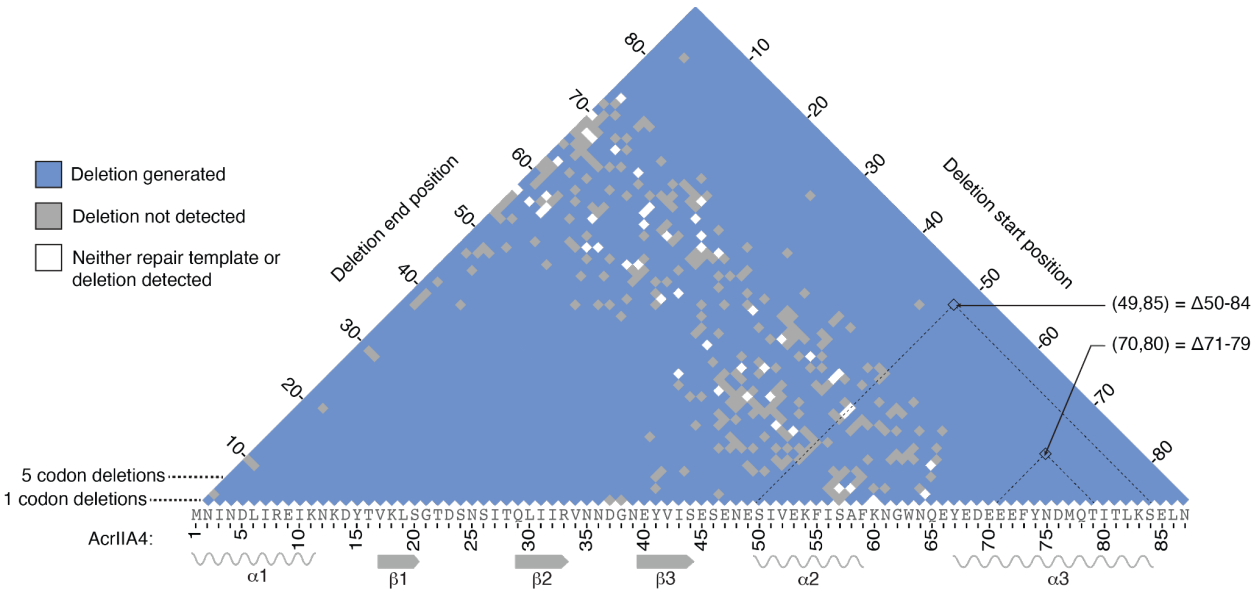


Figure 2. High-throughput sequencing of AcrIIA4 alleles following deletion introduction identified 87% of the possible in-frame deletions, spanning the range of possible deletion sizes.

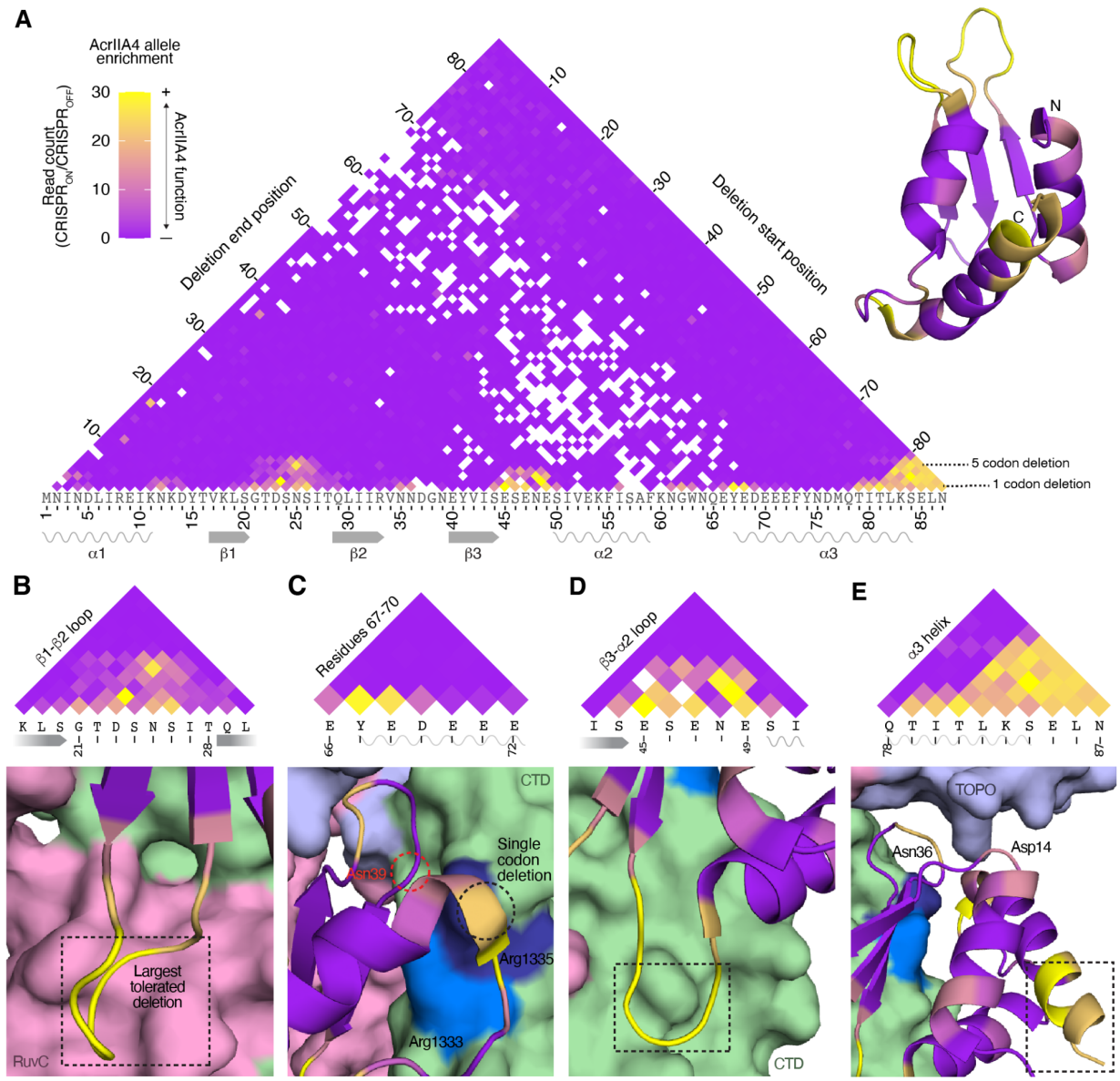
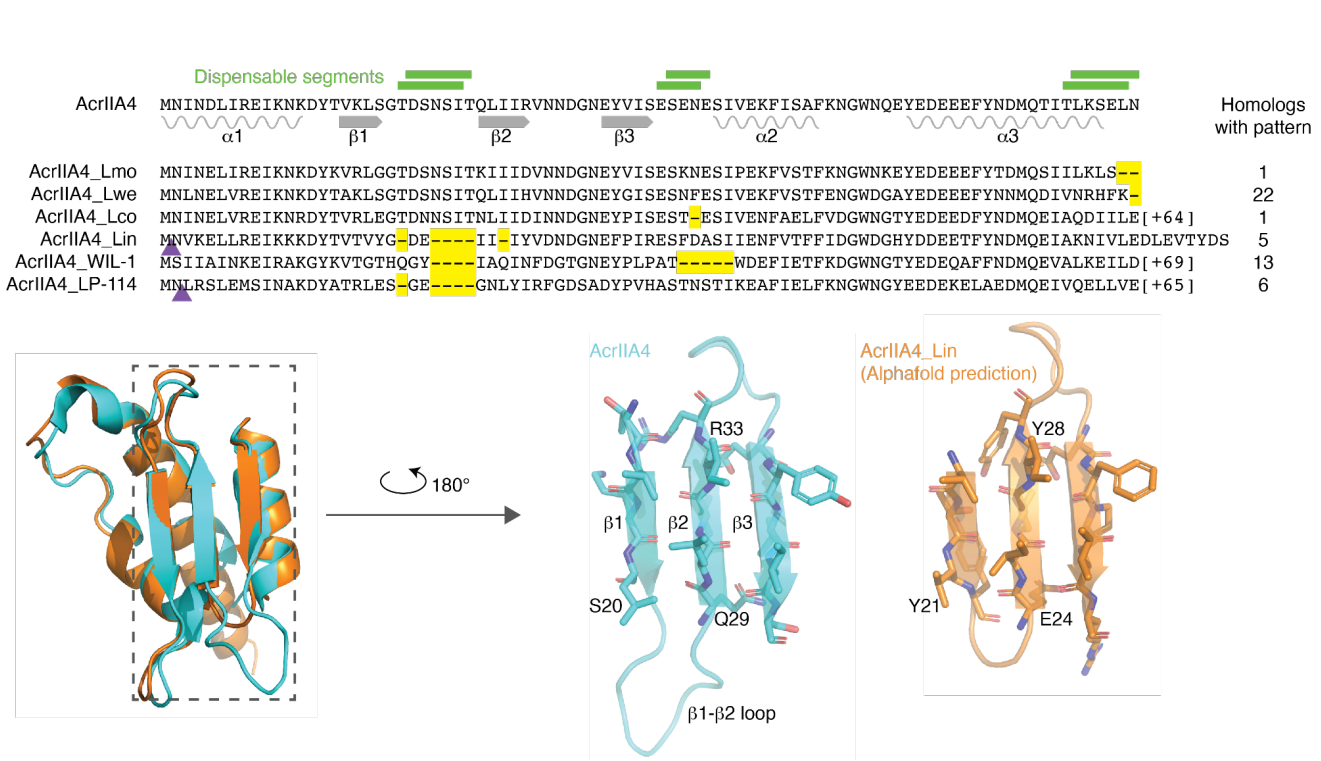


Figure 3. Functional characterization of AcrIIA4 deletion alleles. (A) Functionality of AcrIIA4 deletion alleles, determined by their read counts in CRISPR$_{ON}$ conditions normalized against CRISPR$_{OFF}$ read counts. Deletions are colored by their median value across three replicate experiments to minimize the effects of outlier escaper mutants; alleles only present in two replicates are represented by the lower of the two values. Residues in the AcrIIA4 crystal structure (right) are colored by the maximum

658 functionality of deletions encompassing that residue (six-residue deletions or shorter).
659 (B-E) Dashed boxes demarcate the largest tolerated AcrIIA4 deletion in the view. (B)
660 Large dispensable regions within the RuvC-interacting β1-β2 loop. (C) AcrIIA4 residues
661 67-70, which interact with Arg1333 and Arg1335, are tolerant of single amino acid
662 deletions. Asn39, which interacts with Arg1335, is not dispensable. (D) Large
663 dispensable regions within the CTD-interacting β3-α2 loop. (E) Asp14 and Asn36,
664 TOPO-interacting residues, are partially dispensable. The end of the C-terminal helix α3
665 is dispensable; note that the final two amino acids are not resolved in the crystal
666 structure but were also dispensable.



667

668 **Figure 4.** Deletions in naturally observed homologs of AcrIIA4 occur in experimentally
669 determined dispensable segments. (A) Alignment of six AcrIIA4 homologs representing
670 the identified deletion patterns. Yellow dashes signify deletions relative to AcrIIA4, while
671 purple triangles demarcate the positions of insertions. The positions of select segments
672 identified as dispensable in Fig. 3 are illustrated with green lines above the AcrIIA4
673 sequence. From top to bottom, the AcrIIA4 homolog NCBI IDs and source are
674 EAD0751510.1 isolated from *L. monocytogenes*; MBC1462049.1 from *L. welshimeri*,
675 WP_185587529.1 from *L. cossartiae*, WP_185342258.1 from *L. innocua*,
676 YP_010843666.1 from phage WIL-1, and YP_009045111.1 from phage LP-114. (B) The
677 Alphafold-predicted structure of AcrIIA4_Lin has a full-length β2 strand.

678