Structural basis for Cas9 off-target activity

Martin Pacesa[1], Chun-Han Lin[2,3], Antoine Cléry[4], Aakash Saha[5], Pablo R. Arantes[5], Katja Bargsten[1,6], Matthew J. Irby[2], Frédéric H.T. Allain[4], Giulia Palermo[5], Peter Cameron[2,7], Paul D. Donohoue[2], Martin Jinek[1,8]

Address:

[1]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[2]Caribou Biosciences, 2929 Seventh Street #105, Berkeley, CA 94710, United States

[3]Present address: LinkedIn, Sunnyvale, CA 94085, United States

[4]Institute of Biochemistry, ETH Zurich, Hönggerbergring 64, CH-8093 Zurich, Switzerland

[5]Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States

[6]Present address: leadXpro AG, Villigen, Switzerland

[7]Present address: Spotlight Therapeutics, Hayward, CA 94545, United States

[8]Lead contact

Corresponding author: Martin Jinek (jinek@bioc.uzh.ch)

SUMMARY

The target DNA specificity of the CRISPR-associated genome editor nuclease Cas9 is determined by complementarity to a 20-nucleotide segment in its guide RNA. However, Cas9 can bind and cleave partially complementary off-target sequences, which raises safety concerns for its use in clinical applications. Here we report crystallographic structures of Cas9 bound to *bona fide* off-target substrates, revealing that off-target binding is enabled by a range of non-canonical base pairing interactions and preservation of base stacking within the guide–off-target heteroduplex. Off-target sites containing single-nucleotide deletions relative to the guide RNA are accommodated by base skipping or multiple non-canonical base pairs rather than RNA bulge formation. Additionally, PAM-distal mismatches result in duplex unpairing and induce a conformational change of the Cas9 REC lobe that perturbs its conformational activation. Together, these insights provide a structural rationale for the off-target activity of Cas9 and contribute to the improved rational design of guide RNAs and off-target prediction algorithms.

42 ## INTRODUCTION

43 Cas9, the effector nuclease of prokaryotic Type II CRISPR adaptive immune systems
44 (Makarova et al., 2020), cleaves double-stranded DNA substrates complementary to a guide
45 CRISPR RNA (crRNA) (Jinek et al., 2012). By changing the sequence of the guide RNA, the
46 target DNA specificity of the CRISPR-Cas9 system is readily programmable (Jinek et al.,
47 2012), a feature that has been widely exploited for genome engineering applications (Anzalone
48 et al., 2020). Cas9 functions in conjunction with a trans-activating crRNA (tracrRNA), which
49 is required both for crRNA loading and subsequent DNA binding and cleavage (Deltcheva et
50 al., 2011; Jinek et al., 2012). Target DNA binding and cleavage is further dependent on the
51 presence of a protospacer-adjacent motif (PAM) flanking the target sequence (Anders et al.,
52 2014; Jinek et al., 2012). Due to its high activity and 5'-NGG-3' PAM specificity,
53 *Streptococcus pyogenes* Cas9 (SpCas9) remains the most widely used CRISPR-Cas nuclease
54 for gene editing applications. However, despite a high intrinsic accuracy in generating targeted
55 DNA breaks, SpCas9 can nevertheless cleave genomic DNA sequences with imperfect
56 complementarity to the guide RNA, resulting in off-target editing (Cameron et al., 2017; Hsu
57 et al., 2013; Pattanayak et al., 2013; Tsai et al., 2015). The off-target activity of SpCas9, as well
58 as other Cas9 enzymes, thus presents a safety concern for their therapeutic applications.

59 Off-target sites typically contain one or several nucleobase mismatches relative to the
60 guide RNA (Cameron et al., 2017; Tsai et al., 2017; Tsai et al., 2015). Recent studies have
61 established that the type of mismatch, its positioning, and the total number of mismatches are
62 important determinants of off-target DNA binding and cleavage (Boyle et al., 2017; Boyle et
63 al., 2021; Doench et al., 2016; Jones et al., 2021; Zhang et al., 2020). PAM-proximal
64 mismatches within the seed region of the guide RNA (gRNA)-target strand (TS) DNA
65 heteroduplex typically have a dramatic impact on substrate DNA binding and R-loop formation
66 (Boyle et al., 2021; Ivanov et al., 2020; Singh et al., 2016; Zhang et al., 2020). In contrast,

67 PAM-distal mismatches are compatible with stable binding; however, their presence often
68 results in the formation of a catalytically incompetent complex (Boyle et al., 2021; Dagdas et
69 al., 2017; Ivanov et al., 2020; Jones et al., 2021; Sternberg et al., 2015; Yang et al., 2018; Zhang
70 et al., 2020). In addition, Cas9 has been shown to cleave off-target substrates containing
71 insertions or deletions relative to the guide RNA sequence, which have been proposed to be
72 recognized through the formation of nucleotide "bulges" in the gRNA-TS DNA heteroduplex
73 (Boyle et al., 2021; Cameron et al., 2017; Doench et al., 2016; Jones et al., 2021; Lin et al.,
74 2014; Tsai et al., 2015).

75 Numerous computational tools have been developed to predict possible genomic off-
76 target sites based on sequence similarity (Bae et al., 2014; Stemmer et al., 2015). However, the
77 majority of actual off-target cleavage events remain unpredicted (Cameron et al., 2017; Tsai et
78 al., 2015). Furthermore, although Cas9 is able to bind genomic sites harbouring as few as five
79 complementary nucleotides, only a relatively small number of off-target sites are actually
80 cleaved and result in detectable off-target editing in cells (Kuscu et al., 2014; O'Geen et al.,
81 2015; Wu et al., 2014). Several structures of target-bound Cas9 complexes have been
82 determined to date (Anders et al., 2014; Jiang et al., 2016; Nishimasu et al., 2014; Zhu et al.,
83 2019) that have shed light on the mechanism of on-target binding and cleavage. However, the
84 same processes for off-target sites remain poorly understood.

85 To elucidate the mechanism of mismatch tolerance of Cas9, we determined crystal
86 structures of a comprehensive set of *bona fide* off-target–bound complexes. These structures
87 reveal that the formation of non-canonical base pairs and preservation of heteroduplex shape
88 underpin the off-target tolerance of Cas9. We also observe that multiple consecutive
89 mismatches can be accommodated by base skipping of a guide RNA nucleotide, as opposed to
90 nucleotide bulging. Finally, the structure of an off-target complex containing three PAM-distal
91 mismatches exhibits REC2/3 domain rearrangements, which likely perturbs conformational

activation of Cas9 and thus modulates cleavage efficiency. Taken together, our structural data reveal the diversity of mechanisms enabling off-target recognition and lay the foundation for improved off-target prediction and engineering optimized CRISPR-Cas9 complex designs for gene editing.

96 # RESULTS

97 ## *In vitro* profiling reveals diversity of Cas9 off-targets

98 Multiple studies have investigated the off-target activity of Cas9, suggesting context-dependent
99 tolerance of nucleobase mismatches between the guide RNA and off-target DNA sequences
100 (Boyle et al., 2021; Cameron et al., 2017; Lazzarotto et al., 2020; Tsai et al., 2015; Zhang et al.,
101 2020). To investigate the effect of mismatches on Cas9 binding and cleavage, we performed
102 the SITE-Seq® assay (Cameron et al., 2017) to define the off-target landscapes of 12 well-
103 studied guide RNAs to select suitable off-targets for further evaluation (**Table S1, Table S2**)
104 (**Figure S1A, Table 1**). The SITE-Seq assay analysis revealed a total of 3,848 detectable off-
105 target sites at the highest Cas9 ribonucleoprotein (RNP) concentration, with a median of 5
106 mismatches per off-target site (**Figure S1C**). The detected mismatches covered all possible
107 base mismatch combinations and were distributed throughout the length of the gRNA-TS DNA
108 heteroduplex (**Figure S1D-E**).

109 To probe the thermodynamics of on- and off-target substrate DNA binding and the
110 kinetics of DNA cleavage, we focused on a subset of four guide RNAs (*AAVS1*, *FANCF*,
111 *PTPRC-tgt2,* and *TRAC*) and a total of 15 *bona fide* off-target sites detectable *in vivo* (Cameron
112 et al., 2017; Donohoue et al., 2021; Tsai et al., 2017; Tsai et al., 2015) (**Figure 1A**) that covered
113 all 12 possible base mispair types. To benchmark the editing efficiency of the selected off-target
114 substrates *in vivo*, we transfected human primary T cells with recombinant Cas9 RNPs at
115 multiple concentrations and quantified indel frequencies. We observed efficient cellular editing
116 at the on-target site for each guide RNA at all RNP concentrations, and detected editing at 7 of
117 the 15 off-target sites in our set, including the deletion-containing FANCF off-target #3 (**Figure**
118 **S3, Table S3**).

119 In vitro nuclease activity assays revealed that all selected off-target sequences were
120 cleaved slower than the corresponding on-target substrates, with 20–500-fold reductions in the

121 calculated rate constants (**Table 1, Figure S2A**). To distinguish whether the cleavage defects
122 were due to slow R-loop formation or perturbations in downstream steps, we also quantified
123 cleavage kinetics using partially single-stranded PAMmer DNA substrates (Anders et al., 2014;
124 O'Connell et al., 2014). These experiments revealed that the slower cleavage kinetics of most
125 off-target substrates was due to perturbed R-loop formation (**Table 1, Figure S2B**). However,
126 for some off-targets, notably *AAVS1* off-targets #2 and #5, *FANCF* off-targets #3, #4, and #6,
127 and *PTPRC-tgt2* off-target #1, the rate of PAMmer substrate cleavage was more than 100-fold
128 slower as compared to their respective on-target sequences (**Table 1**, **Figure S2B**), indicating
129 perturbations in the conformational activation checkpoint downstream of guide-target
130 hybridization or inhibition of cleavage by direct steric hindrance of the Cas9 HNH domain
131 (Chen et al., 2017; Dagdas et al., 2017).

132 Complementary quantification of substrate DNA binding using DNA nanolever
133 (switchSENSE) methodology revealed perturbations in the binding affinities of most off-targets
134 (**Table 1, Data S1**). Notably, the reductions in binding affinities were almost entirely due to
135 increased dissociation rates (k_{off}), while on-binding rates (k_{on}) were largely unperturbed (**Table**
136 **1**), indicating that most of the tested off-targets promote DNA dissociation, likely due to R-loop
137 collapse. However, there was little correlation between the observed reductions in cleavage
138 rates and binding constants (**Figure S2C**), confirming that the molecular basis for off-target
139 discrimination by Cas9 is not based on substrate binding alone, in agreement with prior studies
140 (Boyle et al., 2021; Chen et al., 2017; Dagdas et al., 2017; Jones et al., 2021; Yang et al., 2018;
141 Zhang et al., 2020). The dissociation rate (k_{off}) correlated significantly only with the number of
142 mismatches located in the seed region (R^2=0.46, p=0.001) (**Figure S2C**), suggesting that seed
143 mismatches promote R-loop collapse and non-target strand (NTS) rehybridization (Boyle et al.,
144 2017; Gong et al., 2018; Singh et al., 2016; Sternberg et al., 2014).

145 Finally, we compared the measured cleavage rate constants (k_{obs}) with predicted data
146 from a leading biophysical model of Cas9 off-target cleavage that accounts for mismatch
147 number and position, but utilises position-independent weights for mismatch type (Jones et al.,
148 2021). Despite good overall correlation between the model and our data (R^2=0.46, p=0.004)
149 (**Figure S2C**), there were several prominent outliers (*AAVS1* off-target #2 and off-target #3,
150 and *FANCF* off-target #4), suggesting that accurate modelling of off-target interactions requires
151 accounting for position-specific effects of individual mismatch types (**Figure S2D**).

152 **Crystallographic analysis of off-target interactions**

153 To obtain insights into the structural basis of off-target recognition and mismatch tolerance, we
154 employed a previously described approach (Anders et al., 2014) to co-crystallize Cas9 with
155 sgRNA guides and partially duplexed off-target DNA substrates (**Figure 1C**). Focusing on our
156 set of *AAVS1*, *FANCF*, *PTPRC-tgt2* and *TRAC* off-targets (**Figure 1A**), covering all 12 possible
157 mismatch types, we determined a total of 15 off-target complex structures at resolutions of
158 2.25–3.30 Å (**Figure 1C**, **Table S4)**. Overall, the off-target complex structures have very
159 similar conformations, with the Cas9 polypeptide backbone superimposing with a mean root
160 mean square deviation of 0.41Å over 1330 Cα atoms (as referenced to the *FANCF* on-target
161 complex structure, excluding *FANCF* off-target #4, as discussed below). Of note, the *AAVS1*
162 on-target complex structure reveals substantial repositioning of the REC2 domain, where it
163 undergoes a 12° rotation (relative to *FANCF* and *TRAC* on-target complexes) (**Figure S4A,**
164 **Video S1**), with concomitant shortening of the α-helix comprising residues 301-305 and
165 restructuring of the loop comprising residues 175-179 (**Figure S4B**), enabled by the absence of
166 crystal contacts involving the REC2 and REC3 domains.

167 However, the structures display considerable local variation of the gRNA-TS DNA
168 heteroduplex conformation (**Table S5)**. Base base pairing and base stacking are mostly
169 preserved throughout the heteroduplexes (**Table S6**), with the exception of positions 1–3 within

170 the PAM-distal end of the guide-TS duplex, where the presence of mismatches results in duplex
171 unpairing (**Figure S5**). Interestingly, such unpairing results in the base stacking of the terminal
172 residue of the sgRNA on top of the unpaired DNA base, as observed in a recent cryo-EM
173 structure of a PAM-distal mismatch off-target complex (Bravo et al., 2022). Despite the
174 observed conformational variation, the off-target structures preserve almost all intermolecular
175 contacts between the Cas9 protein and the bound nucleic acids (**Figure S6**), further
176 underscoring the structural plasticity of Cas9 in accommodating mismatch-induced
177 heteroduplex distortions.

178 **Non-canonical base-pairing interactions facilitate off-target recognition**

179 A substantial fraction of the base pair mismatches observed in the determined off-target
180 complex structures (34 out of 49) is accommodated by non-canonical base pairing interactions
181 that preserve at least one hydrogen bond between the mispaired bases. The most common off-
182 target mismatches, both in our data set (**Table S1, Table S2**) and as reported by other studies
183 (Boyle et al., 2017; Doench et al., 2016; Hsu et al., 2013; Jones et al., 2021; Pattanayak et al.,
184 2013; Zhang et al., 2020), are rG-dT (**Figure S7**) and rU-dG (**Figure 2A, Figure S8**), which
185 have the potential to form wobble base pairs (Kimsey et al., 2015). Indeed, all rG-dT
186 mismatches in the determined structures are accommodated by wobble base pairing. Observed
187 at duplex positions 4, 13 and 15, the dT base undergoes a ~1 Å shear displacement into the
188 major groove of the gRNA–TS DNA heteroduplex to form the wobble base pair (**Figure S7A-**
189 **E**), whereas at duplex position 2, wobble base pairing is enabled by a minor groove
190 displacement of the rG base (**Figure S7F-G**). In contrast, rU-dG base pairs in the determined
191 structures exhibit considerable structural variation. At duplex position 10, the rU base is able
192 to undergo the major groove displacement required for wobble base-pairing (**Figure 2A, Figure**
193 **S8A**). In contrast, at duplex position 5, the backbone of the RNA strand makes extensive
194 contacts with Cas9 (**Figure S8B-D, Figure S6**). As a result, the rU-dG base pairs are instead

195 accommodated by compensatory shifts of the dG base to maintain hydrogen-bonding
196 interactions (**Figure S8B-D**). At duplex position 9 in the *TRAC* off-target #1 complex, the rU-
197 dG mismatch is accommodated by wobble base-pairing enabled by a minor groove
198 displacement of the dG base (**Figure S8E**). At the same duplex position in the *AAVS1* off-target
199 #1 and #2 complexes, however, this mismatch occurs next to rC-dA and rC-dT mismatches,
200 respectively, and adopts the sterically prohibited Watson-Crick geometry (**Figure S8F-G**),
201 implying a tautomeric shift or base deprotonation to accommodate this otherwise unfavorable
202 base pairing mode (**Figure S8H**). Collectively, these observations suggest that the ability of
203 rU-dG (and likely rG-dT) mismatches to form wobble base pairing interactions is determined
204 not only by backbone constraints at the specific position within the gRNA-TS DNA
205 heteroduplex (**Figure S6)**, but also by local sequence context and/or the presence of
206 neighboring mismatches (**Figure S8C-G**).

207 rA-dC or rC-dA mismatches can also form wobble-like base pairs when the adenine
208 base is protonated at the N1 position (Garg and Heinemann, 2018; Wang et al., 2011). In the
209 rA-dC mispairs found at duplex position 4, the dC base undergoes a wobble displacement
210 compatible with the formation of two hydrogen bonds with the adenine base, indicative of
211 adenine protonation (**Figure 2B**, **Figure S9A**). At other duplex positions in our data set, the
212 rA-dC or rC-dA mispairs are instead accommodated by slight displacements of the adenine
213 base within the base pair plane resulting in the formation of a single hydrogen bond in each
214 case (**Figure S9B-D**).

215 Accommodating purine-purine mismatches by Watson-Crick-like interactions would
216 normally require severe distortion of the guide–off-target duplex to increase its width by more
217 than 2 Å (Leontis et al., 2002). At positions where the duplex width is constrained by Cas9
218 interactions **(Figure S6**), rG-dA and rA-dG mispairs are accommodated by *anti*-to-*syn*
219 isomerization of the adenine base to form two hydrogen-bonding interactions via its Hoogsteen

220 base edge. This is observed at duplex position 11 in the *AAVS1* off-target #4 complex (rG-dA
221 mispair) (**Figure 2C**) and at position 7 in the *AAVS1* off-target #2 complex (rA-dG mispair)
222 (**Figure S9E**). Similarly, the rG-dG mispair at duplex position 13 in the *FANCF* off-target #5
223 complex is accommodated by Hoogsteen base-pairing as a result by *anti*-to-*syn* isomerization
224 of the guide RNA base (**Figure 2D**). Overall, the observed Hoogsteen base pairing interactions
225 are near-isosteric with Watson-Crick base pairs and maintain duplex width without excessive
226 backbone distortion (**Table S6**).

227 **Duplex backbone rearrangements accommodate otherwise non-productive mismatches**

228 Whereas wobble (G-U/T or A-C) and Hoogsteen (A-G or G-G) base pairs are generally
229 compatible with the canonical A-form geometry of an RNA-DNA duplex, other nucleotide
230 mismatches only form non-isosteric base pairs that require considerable distortion of the
231 (deoxy)ribose-phosphate backbone. The formation of pyrimidine-pyrimidine base pairs is
232 expected to occur by a substantial reduction in duplex width (Leontis et al., 2002). This is
233 observed at the rU-dC mismatch at duplex position 9 in the *FANCF* off-target #1 complex
234 (**Figure 3A**). Here, the guide RNA backbone is able to shift towards the target DNA strand,
235 resulting in a reduction of the C1'–C1' distance to 8.65 Å as compared to 10.0 Å in the *FANCF*
236 on-target complex. This facilitates the formation of two hydrogen bonding interactions within
237 the rU-dC base pair, which is further enabled by a substantial increase in base propeller twist
238 (**Figure 3A**). In contrast, rC-dT mismatches remain unpaired at duplex positions 6 and 7 in
239 (**Figure S10A-B**), or form only a single hydrogen bond at position 8 (**Figure S10C**), likely due
240 to backbone steric constraints at these positions imposed by Cas9 interactions (**Figure S6**). Of
241 note, the *FANCF* off-target #7 rC-dT mismatch is bridged by hydrogen bonding interactions
242 with the side chain of Arg895 inserted into the minor groove of the heteroduplex (**Figure**
243 **S10B**); however, the interaction is not essential for the tolerance of rC-dT mismatches at this
244 position (**Figure S11, Table S7**). Backbone steric constraints also likely influence the

245 formation of rU-dT base pairs (**Figure S6**). At duplex position 7, the mismatch remains
246 unpaired (**Figure S10D**), whereas productive pairing is seen at duplex positions 8 and 9,
247 facilitated by distortions of the guide RNA and TS backbone, respectively (**Figure 3B, Figure**
248 **S10E-F**).

249 rC-dC mismatches only form productive hydrogen-bonding interactions if bridged by a
250 water molecule or when one of the cytosine bases is protonated (Leontis et al., 2002). Only the
251 former is observed in the determined structures, at duplex position 5 in the *AAVS1* off-target #2
252 complex (**Figure S10G**). In contrast, at duplex position 8 and 15, the bases remain unpaired
253 while maintaining intra-strand base stacking interactions (**Figure 3C, Figure S10H**). Similarly,
254 rA-dA mismatches are unable to form productive hydrogen-bonding interactions within the
255 constraints of an A-form duplex (Leontis et al., 2002). Accordingly, the rA-dA mismatch at
256 duplex position 5 is accommodated by extrusion of the dA nucleobase out of the base stack into
257 the major groove of the duplex (**Figure 3D**). As the duplex width is constrained at this position
258 by Cas9 (**Figure S6**), the base extrusion is enabled by local distortion of the TS backbone
259 (**Figure S12A**). Analysis of our SITE-Seq assay data set revealed that off-target rA-dA
260 mismatches occur at all positions within the gRNA–TS DNA heteroduplex (**Figure S12B**), in
261 agreement with previous studies (Boyle et al., 2017; Boyle et al., 2021; Doench et al., 2016;
262 Jones et al., 2021; Zhang et al., 2020). This suggests that rA-dA mismatches do not encounter
263 steric barriers within Cas9 that would disfavour their presence, which is consistent with the
264 absence of specific contacts with Cas9 along the length of the major groove of the heteroduplex
265 (**Figure S6**).

266 To gain further insights into the mechanism of rA-dA mismatch accommodation and
267 extrapolate our structural observations to other heteroduplex positions, we performed molecular
268 dynamics (MD) simulations by introducing single rA-dA mismatches at all 20 positions along
269 the heteroduplex in the context of the catalytically active state of Cas9 (PDB: 6O0Y) and

270 compared the resulting trajectories to the corresponding on-target system. rA-dA mismatch
271 simulations at positions 5, 18, and 19 showed very good agreement with the determined crystal
272 structures (*PTPRC-tgt2 off-target #1*, *FANCF off-target #6*, *AAVS1 off-target #2)*, predicting
273 both dA base extrusion and phosphate backbone distortions (**Figure S12C-D**). In-depth
274 analysis of the rA-dA conformational dynamics was performed by computing the geometrical
275 descriptors defining the base pair complementarity along the heteroduplex (**Figure S12D,**
276 **Figure S13**) (Lavery et al., 2009). These reveal that at all positions, the rA-dA mismatch is
277 primarily accommodated by base extrusion from the duplex, with positions 3, 5 ,15, 16, 18, 19
278 and 20 undergoing dA base extrusion, and positions 10 and 17 rA base extrusion. Notably, at
279 duplex positions 2, 4, 6, 8, 11 and 12, the broad or bimodal distributions of base pair shear and
280 opening parameters are indicative of considerable conformational dynamics, allowing the rA-
281 dA mismatch to be accommodated by either rA or dA base extrusion (**Figure S13A**).

282 **PAM-proximal mismatches are accommodated by TS distortion due to seed sequence**
283 **rigidity**

284 The seed sequence of the guide RNA (nucleotides 11-20) makes extensive interactions with
285 Cas9, both in the absence and presence of bound DNA (Anders et al., 2014; Jiang et al., 2015;
286 Nishimasu et al., 2014; Zhu et al., 2019). Structural pre-ordering of the seed sequence by Cas9
287 facilitates target DNA binding and contributes to the specificity of on-target DNA recognition
288 (Jiang et al., 2015; O'Geen et al., 2015; Wu et al., 2014). Conversely, binding of off-target
289 DNAs containing PAM-proximal mismatches is inhibited and results in accelerated off-target
290 dissociation (Boyle et al., 2017; Boyle et al., 2021; Ivanov et al., 2020; Jones et al., 2021; Singh
291 et al., 2016; Zhang et al., 2020). Nevertheless, Cas9 does tolerate most base mismatch types
292 within the seed region of the guide RNA, leading to detectable off-target DNA cleavage (Boyle
293 et al., 2021; Doench et al., 2016; Jones et al., 2021; Zhang et al., 2020). In particular, the first
294 two PAM-proximal positions display a markedly higher tolerance for mismatches than the rest

295 of the seed region (Cofsky et al., 2022; Doench et al., 2016; Hsu et al., 2013; Mekler et al.,
296 2017; Zeng et al., 2018); this is supported by our SITE-Seq assay data as the frequency of
297 mismatches at PAM-proximal positions 18-20 is 1.4-2.0x higher than at seed positions 14-17
298 (**Figure S1D-E**).

299 Unlike the seed region of the guide RNA, the complementary PAM-proximal TS
300 nucleotides are not directly contacted by Cas9 in the pre-cleavage state and are thus under fewer
301 steric constraints (**Figure S6**), with the exception of duplex position 20 in which the
302 phosphodiester group of the TS nucleotide makes extensive interactions with the phosphate
303 lock loop of Cas9 (Anders et al., 2014) (**Figure S14A**). In agreement with this, four of our off-
304 target complex structures reveal that PAM-proximal base mismatches are accommodated solely
305 by structural distortions of the TS backbone (**Figure 4**), while the conformation of the guide
306 RNA backbone and base stacking within the seed region remain unperturbed (**Figure S14B-**
307 **C)**. The presence of an rA-dA mismatch in the PAM-proximal position 18 results in the
308 extrusion of the TS nucleobase into the major groove (**Figure 4A**), likely due to steric
309 constraints on duplex width at this position. In contrast, the rA-dA mismatch at duplex position
310 19 is instead accommodated by a marked distortion in the TS backbone that results in increased
311 duplex width, which preserves base stacking within the duplex in the absence of productive
312 pairing between the adenine bases (**Figure 4B**, **Figure S14B**). Similarly, the rA-dG mismatch
313 at position 19 is accommodated by a ~2 Å displacement of the TS backbone, increasing duplex
314 width. This not only preserves base stacking but also facilitates rA-dG base paring by two
315 hydrogen bonding interactions via their Watson-Crick edges (**Figure 4C**, **Figure S14C**). This
316 off-target complex also contains a rU-dG mismatch at duplex position 20 which does not
317 undergo wobble base pairing as the rU20 nucleotide is extensively contacted by Cas9 (**Figure**
318 **S6**) and unable to shift towards the major groove and is instead accommodated by a slight shift
319 in the dG nucleotide (**Figure 4C**). Finally, the rU-dT base mismatch at duplex position 20 in

320 the *AAVS1* off-target #4 complex remains unpaired and the dT base lacks ordered electron
321 density (**Figure 4D**). This is likely a result of the dT nucleotide maintaining contact with the
322 phosphate-lock loop of Cas9 (**Figure S14A**), which prevents a reduction in the duplex width
323 and precludes productive base pairing. Overall, these observations indicate that off-target
324 DNAs containing mismatches to the seed sequence of the guide RNA can be accommodated by
325 Cas9 due to limited interactions with the TS DNA in the seed-binding region

326 Cas9 recognizes off-targets with single-nucleotide deletions by base skipping or via 327 multiple mismatches

328 A substantial fraction of off-target sites recovered in our SITE-Seq assay analysis (46.4%, when
329 not considering the possibility of nucleotide insertions or deletions) contained six or more
330 mismatched bases to the guide RNA (**Figure S1C**, **Table S1, Table S2**). Such off-target
331 sequences have previously been proposed to be accommodated by bulging out or skipping of
332 nucleotides (Boyle et al., 2021; Cameron et al., 2017; Doench et al., 2016; Jones et al., 2021;
333 Lin et al., 2014; Tsai et al., 2015), which would result in a shift of the nucleotide register to re-
334 establish correct base pairing downstream of the initially encountered mismatch. The *PTPRC-*
335 *tgt2* off-target #1 and *FANCF* off-target #3 sites are predicted to contain single nucleotide
336 deletions at duplex positions 15 and 17, respectively (**Figure 1A**). The structures of the
337 corresponding off-target complexes reveal that the single nucleotide deletions in these off-target
338 substrates are not accommodated by bulging out the unpaired guide RNA nucleotide. Instead,
339 the conformations of the guide RNAs remain largely unperturbed and the off-target TS DNAs
340 "skip over" the unpaired gRNA bases to resume productive base-pairing downstream (**Figure**
341 **5A-B**). The seed regions of the gRNAs make extensive interactions with the bridge helix and
342 the REC1 domain, whereas the TS DNA phosphate backbones are displaced by almost 3 Å
343 (**Figure S17A-B**). The base pair skips are accommodated by considerable buckling and tilting
344 of the base pairs immediately downstream of the skip site (**Table S6**). An additional

345 consequence of the base pairing register shift is the formation of non-canonical base pairs
346 between the off-target DNA and the extra 5'-terminal guanine nucleotides present in the guide
347 RNA as a consequence of *in vitro* transcription by T7 RNA polymerase (**Figure S17C-D**). This
348 potentially explains the impact of the 5'-guanines on both R-loop stability and *in vitro* cleavage
349 activity (Kulcsar et al., 2020; Mullally et al., 2020; Okafor et al., 2019).

350 Originally, our SITE-Seq assay analysis classified the *AAVS1* off-target #2 as a single-
351 nucleotide deletion at duplex position 9 (**Figure 5C**). Unexpectedly, the structure of the *AAVS1*
352 off-target #2 complex instead reveals that the off-target substrate is bound in the unshifted
353 register, resulting in the formation of five base mismatches in the PAM-distal half of the guide
354 RNA–TS duplex (**Figure 5D**), including a partially paired rC-dC mismatch at position 5, an
355 rA-dG Hoogsteen pair at position 7, a partially paired rC-dT mismatch at position 8, and a
356 tautomeric rU-dG pair at position 9. The backbone conformations of the gRNA and the off-
357 target TS exhibit minimal distortions and are nearly identical with the corresponding on-target
358 heteroduplex (**Figure 5D**). This implies that some mismatch combinations might
359 synergistically result in guide RNA and TS backbone geometries that mimic the on-target
360 conformation and result in off-target tolerance. To test this hypothesis, we reverted the rC-dT
361 mismatch at position 8 to the on-target rC-dG pair, thereby reducing the total amount of off-
362 target mismatches from 6 to 5 (**Figure 5C**). The resulting off-target substrate (*AAVS1* off-target
363 #2-rev) exhibited substantially reduced cleavage rates in both dsDNA and PAMmer formats,
364 as well as a significantly increased dissociation rate (**Figure 5E, Table S8**). We replicated these
365 results for two other *AAVS1* off-target substrates with different mismatch patterns that were
366 initially predicted to contain PAM-distal deletions (**Figure S18, Table S8**). Together, these
367 results suggest that some *bona fide* off-target substrates containing multiple consecutive
368 mismatches, the reversal of one mismatch may affect the structural integrity of the guide RNA–

369 TS DNA heteroduplex and interfere with DNA binding and/or conformational activation of
370 Cas9, despite a reduction in the total number of mismatches.

371 To further investigate the accommodation of PAM-distal deletions in Cas9 off-target
372 sites, we selected an additional efficiently cleaved off-target substrate (*CD34* off-target #9,
373 **Figure S19A,B**), containing a dT substitution at position 17 (resulting in a rU-dT mismatch
374 with the *CD34* sgRNA) and a predicted single-nucleotide deletion at position 6. Crystal
375 structure of the resulting complex revealed that the rU-dT mismatch involves the *syn* conformer
376 of the rU17 nucleotide (**Figure S19C**). The deletion at position 6 is accommodated by base
377 skipping, leaving the base of rA6 unpaired within the guide-TS heteroduplex stack (**Figure**
378 **S19E**) and the register shift results enables the 5'-terminal guanine nucleotide of the gRNA
379 (introduced during in vitro transcription) to base pair with 3'-terminal TS nucleotide (**Figure**
380 **S19F**). The backbone distortions in the TS are accompanied by compensatory rearrangements
381 of the REC2 and REC3 domains of Cas9 (**Figure S19G**), which preserve all contacts with the
382 heteroduplex (**Figure S6**).

383 Collectively, these results indicate that deletion-containing off-target complexes are
384 accommodated either by RNA base skipping, as opposed to RNA nucleotide bulging, or by the
385 formation of multiple base mismatches, with the precise mechanism dependent in part on the
386 position of the deletion.

387 **PAM-distal mismatches perturb the Cas9 conformational checkpoint**

388 *FANCF* off-target #4, which contains three PAM-distal mismatches at positions 1-3 and a G-U
389 mismatch in position 10 (**Figure 1A**), is reproducibly the top ranking off-target site for the
390 *FANCF* guide RNA, as detected by SITE-Seq assay analysis at the lowest Cas9 RNP
391 concentrations (**Table S1, Table S2**). The off-target substrate exhibits slow cleavage kinetics
392 *in vitro* with both dsDNA and PAMmer substrates (**Figure 2B, Figure S2A-B**), indicating a
393 perturbation of the conformational activation checkpoint of Cas9. The structure of the *FANCF*

394 off-target #4 complex reveals that the RNA–DNA heteroduplex is unpaired at positions 1-3 as
395 a result of the PAM-distal mismatches, with nucleotides 1-2 of the guide RNA and 19-20 of the
396 TS disordered (**Figure 6A**). Furthermore, Cas9 undergoes structural rearrangements of its REC
397 lobe and the HNH domain (**Figure 6B, Video S2**), resulting in a root mean square displacement
398 of the REC2 and REC3 domains of 3.7 Å (1,315 Cα atoms) relative to the *FANCF* on-target
399 complex structure. The REC3 domain undergoes a 19-degree rotation (**Figure 6B**), facilitated
400 by extending the helix comprising residues 703-712 through restructuring of loop residues 713-
401 716 (**Figure S20A)**, to accommodate the altered guide RNA conformation. The REC2 domain
402 rotates 32° away from the REC3 domain (**Figure 6B**). This is accompanied by restructuring of
403 the hinge loop residues 174-180 and disordering of loops 258-264, 284-285, and 307-309.
404 Concomitantly, the HNH domain rotates 11° away from the heteroduplex, as compared to the
405 *FANCF* on-target structure, to accommodate distortion of the TS DNA (**Figure 6B**).

406 The unpaired 5' end of the sgRNA is located at the interface between the REC3 and the
407 RuvC domain and maintains interactions with heteroduplex-sensing residues Lys510, Tyr515,
408 and Arg661 of the REC3 domain (**Figure S20B).** In contrast to the corresponding on-target
409 complex structure, the unpaired 3' end of the off-target TS breaks away from the REC3 lobe
410 and instead points towards the REC2 domain, forming unique interactions with Arg895,
411 Asn899, Arg905, Arg919 and His930 in the HNH domain (**Figure S20C**). These interactions
412 (**Figure S20D)** could be responsible for the observed repositioning of the REC lobe and HNH
413 domain, and they may impede the formation of a cleavage-competent complex.

414 The conformation of the *FANCF* off-target #4 complex is distinct from the
415 conformations observed in cryo-EM reconstructions of the pre- and post-cleavage states of the
416 Cas9 complex (Zhu et al., 2019) (**Figure S21A-B**). Instead, the off-target complex structure
417 most closely resembles that of a high-fidelity variant xCas9 3.7 containing amino acid
418 substitutions that disrupt interactions with the TS DNA (Guo et al., 2019). Although the xCas9

419 3.7 complex adopts a slightly different REC lobe conformation (**Figure S21C**), the PAM-distal
420 duplex also undergoes unpairing at positions 1–3 and displays a comparable degree of structural
421 disorder (**Figure S21D**). These structural observations thus suggest that the presence of
422 multiple mismatches in the PAM-distal region of a guide RNA–off-target DNA duplex leads to
423 conformational perturbations in the DNA-bound complex that resemble the structural
424 consequences of specificity-enhancing mutations in high-fidelity Cas9 variants.
425
426

427 ## DISCUSSION

428 The off-target activity of Cas9 has been extensively documented in prior genome editing,
429 biochemical and biophysical studies (Boyle et al., 2017; Boyle et al., 2021; Doench et al., 2016;
430 Jones et al., 2021; Lazzarotto et al., 2020; Zhang et al., 2020). Although numerous methods
431 have been devised for computational prediction of genomic off-target sites and their
432 experimental validation, these have reported highly variable mismatch tolerance profiles
433 depending on the screening method and the target sequence. Thus, a comprehensive
434 understanding of this phenomenon is still lacking, particularly as to whether off-target tolerance
435 has an underlying structural basis. In this study, we used the SITE-Seq assay to examine the
436 off-target landscape of 12 well-studied guide RNAs, observing a broad variation of cleavage
437 activities associated with individual off-target substrates. To shed light on the molecular
438 mechanisms underpinning off-target activity, we determined atomic structures of a
439 representative set of 16 off-target complexes, thus providing fundamental insights into the
440 structural aspects of off-target recognition.

441 ### Role of non-canonical base pairing in off-target recognition

442 The principal, and largely unexpected, finding of our structural analysis is that the majority of
443 nucleotide mismatches in *bona fide* off-target substrates are accommodated by non-canonical
444 base pairing interactions. These range from simple rG-dT/rU-dG wobble or rA-dG/rG-dA/rG-
445 dG Hoogsteen base pairing interactions, to pyrimidine-pyrimidine pairs that rely on
446 (deoxy)ribose-phosphate backbone distortions that reduce duplex width. With the notable
447 exception of rA-dA mismatches, which are accommodated by extrahelical base extrusion, the
448 structural rearrangements associated with base mismatch accommodation largely preserve base
449 stacking, which is the primary determinant of nucleic acid duplex stability (Yakovchuk et al.,
450 2006). For some off-target sequences, our structures are suggestive of base protonation or
451 tautomerization facilitating hydrogen bonding interactions in otherwise non-permissive

452 mismatches, such as rA-dC. These rare base pair forms have been previously observed in both
453 RNA and DNA duplexes and are thought to be important contributors to DNA replication and
454 translation errors (Kimsey et al., 2015; Kimsey et al., 2018). Future studies employing
455 complementary structural methods, such as nuclear magnetic resonance, will help confirm the
456 occurrence of non-canonical base states in off-target complexes.

457 The mismatch tolerance of Cas9 can be explained primarily by two factors. Firstly, Cas9
458 does not directly contact the major- or minor-groove edges of the guide RNA–TS DNA
459 heteroduplex base pairs at any of the duplex positions and thus lacks a steric mechanism to
460 enforce Watson-Crick base pairing (**Figure S6**). This is further underscored by Cas9's tolerance
461 of base modifications in target DNA, including cytosine 5-hydroxymethylation and, at least at
462 some duplex positions, glucosyl-5-hydroxymethylation (Vlot et al., 2018). In this respect, Cas9
463 differs from other molecular systems, notably the ribosome and replicative DNA polymerases,
464 which enhance the specificity of base-pairing by direct readout of base-pair shape and steric
465 rejection of mismatches (Kunkel and Bebenek, 2000; Rodnina and Wintermeyer, 2001; Timsit,
466 1999).

467 Secondly, Cas9 is a multidomain protein that displays considerable conformational
468 plasticity and is therefore able to accommodate local distortions in the guide–TS duplex
469 geometry by compensatory rearrangements of the REC2, REC3 and HNH domains (Chen et
470 al., 2017; Donohoue et al., 2021). Indeed, in most off-target structures reported in this study,
471 almost all atomic contacts between Cas9 and the guide–TS heteroduplex are preserved (**Figure**
472 **S6**). Thus, Cas9 only detects guide-target mismatches by indirect readout of the guide RNA–
473 TS DNA heteroduplex width, except at the PAM-distal end of the heteroduplex where base
474 mismatches result in duplex unpairing, as discussed below. Our observations are consistent with
475 recent molecular dynamics simulation studies showing that internally positioned mismatches
476 within the guide RNA–TS DNA heteroduplex are readily incorporated within the heteroduplex

477 and have only minor effects on Cas9 interactions (Mitchell et al., 2020). The lack of a steric
478 base-pair enforcement mechanism and the resulting off-target promiscuity likely reflects the
479 biological function of Cas9 in CRISPR immunity, where mismatch tolerance enables targeting
480 of closely related viruses and hindering immune evasion by mutations or covalent base
481 modifications (Deveau et al., 2008; Semenova et al., 2011; van Houte et al., 2016; Yaung et al.,
482 2014). On the other hand, such conformational plasticity also enables the incorporation of
483 various chemical modifications in the guide RNA that are compatible with the A-form geometry
484 of the heteroduplex (Cromwell et al., 2018; Donohoue et al., 2021; Hendel et al., 2015; O'Reilly
485 et al., 2019; Yin et al., 2018; Yin et al., 2017). These could potentially disfavour the formation
486 of certain non-canonical base pairs or reduce the backbone flexibility of the guide RNA, thereby
487 enhancing off-target discrimination. However, the effect of specific guide RNA modifications
488 on particular types of mismatches is yet to be closely examined mechanistically (Donohoue et
489 al., 2021).

490 **Structural rigidity of the guide RNA seed region and implications for off-target**
491 **recognition**

492 The seed sequence of the Cas9 guide RNA (nucleotides 11-20) is the primary determinant of
493 target DNA binding, a consequence of its structural pre-ordering in an A-like conformation by
494 extensive interactions with Cas9 (Anders et al., 2014; Jiang et al., 2015; Nishimasu et al., 2014;
495 Zhu et al., 2019). Our data indicate that structural rigidity of the guide RNA seed sequence also
496 affects off-target recognition, as base mismatches in the seed region of the heteroduplex can
497 only be accommodated by conformational distortions of TS DNA, which is subject to only a
498 few steric constraints, notably at position 20 due to interactions with the phosphate lock loop
499 (Anders et al., 2014). This increases the energetic penalty of base mispairing in the seed region
500 of the heteroduplex, and thus contributes to mismatch sensitivity of Cas9 within the seed region.
501 Although structural distortions of TS DNA facilitate binding of off-target substrates containing

502 seed mismatches, they may nevertheless inhibit off-target cleavage by steric hindrance of the
503 HNH domain, thereby further contributing to off-target discrimination. The contrasting
504 structural plasticities of the guide RNA and TS DNA strands are manifested in the differential
505 activities of Cas9 against off-targets containing rU-dG and rG-dT mismatches within the seed
506 region (Boyle et al., 2021; Doench et al., 2016; Hsu et al., 2013; Jones et al., 2021; Zhang et
507 al., 2020). Whereas rG-dT mismatches can be readily accommodated by wobble base pairing,
508 seed sequence rigidity is expected to hinder rU-dG wobble base pairing. Combined with a lower
509 energetic penalty associated with rG-dT mismatch binding (binding an off-target with an rG-
510 dT mismatch requires unpairing a dT-dA base pair in the off-target DNA, while rU-dG off-
511 target recognition requires dC-dG unpairing), these effects thus help Cas9 discriminate against
512 rU-dG mismatches in the seed region.

513 **Recognition of off-targets containing insertions and deletions**

514 *Bona fide* off-target sites containing insertions or deletions have been detected in a number of
515 studies (Boyle et al., 2021; Cameron et al., 2017; Doench et al., 2016; Jones et al., 2021; Tsai
516 et al., 2015). Nucleotide "bulging" has been proposed as a mechanism to recognize such off-
517 targets, which would otherwise result in large numbers of consecutive base mismatches.
518 However, as Cas9 encloses the guide RNA–TS DNA heteroduplex in a central channel and
519 makes extensive interactions along the entire length of the guide RNA strand, the formation of
520 RNA bulges is precluded by steric clashes, pointing to a different mechanism.

521 Indeed, our structural analysis provides no evidence for RNA bulging. Instead, the
522 structures of *PTPRC-tgt2* off-target #1 and *FANCF* off-target #3 complexes reveal that off-
523 target sequences predicted to contain single-nucleotide deletions in the seed region of the
524 heteroduplex are recognized by base skipping, resulting in an unpaired guide RNA base within
525 the duplex stack. This is enabled by the lack of extensive contacts of Cas9 with the TS, while
526 rigid coordination of the guide RNA in the seed region disfavours extrahelical guide RNA

527 bulging. However, base skipping in the seed region of the heteroduplex likely incurs a large
528 energetic penalty. As the seed region of the TS DNA is devoid of Cas9 contacts in the pre-
529 cleavage state (Zhu et al., 2019), off-targets containing single-nucleotide insertions in the seed
530 region of the heteroduplex are likely to be recognized by DNA nucleotide bulging, likewise
531 incurring a large energetic penalty as unwinding an off-target DNA sequence containing an
532 insertion requires breaking an extra base pair. Additionally, TS DNA distortion might inhibit
533 cleavage by steric hindrance of the HNH domain. These observations thus explain why Cas9
534 appears to tolerate mismatches better than insertions or deletions and why deletions and
535 insertions within the seed region are particularly deleterious (Boyle et al., 2021; Cameron et al.,
536 2017; Doench et al., 2016; Jones et al., 2021).

537 In contrast, off-target sequences containing insertions or deletions in the PAM-distal
538 region of the heteroduplex (positions 1-10) can be recognized either by base skipping, as in the
539 case of *CD34* off-target #9, or bound in the unshifted register, with multiple base mismatches
540 accommodated by non-canonical base pairing interactions, as seen for *AAVS1* off-target #2,
541 which was previously predicted to contain an RNA bulge or skip (Cameron et al., 2017;
542 Lazzarotto et al., 2020). Which of the two mechanisms is used likely depends on the off-target
543 sequence and the position of the deletion, which in turn dictate the number of mismatches
544 between the guide and the off-target sequence in the unshifted register. Off-target
545 accommodation by multiple non-canonical base pairs likely relies on their synergistic effect to
546 mimic the on-target heteroduplex geometry, which enables unperturbed binding, as supported
547 by our observations that mismatch reversal in several off-target substrates reduces their rates of
548 cleavage. Although our structural analysis does not examine off-target substrates containing
549 insertions, we posit that PAM-distal insertions are recognized as multiple mismatches due to
550 steric hindrance of extrahelical bulging. Thus, our structural findings suggest that a substantial

551 fraction of off-target sites predicted to contain insertions or deletions may be bound via multiple
552 mismatches instead (Boyle et al., 2021; Doench et al., 2016; Jones et al., 2021).

553 **PAM-distal base pairing and the conformational checkpoint of Cas9**

554 Upon substrate DNA hybridization and R-loop formation, Cas9 undergoes conformational
555 activation of its nuclease domains (Bravo et al., 2022; Pacesa et al., 2022; Zhu et al., 2019).
556 The Cas9 REC3 domain plays a key role in the process, as it senses the integrity of the PAM-
557 distal region of the guide RNA–TS DNA heteroduplex and allosterically regulates the REC2
558 and HNH domains, providing a conformational checkpoint that traps Cas9 in a
559 conformationally inactive state in the absence of PAM-distal hybridization (Chen et al., 2017;
560 Dagdas et al., 2017; Palermo et al., 2018; Zhu et al., 2019). Our structural data confirm that
561 mismatches at the PAM-distal end of the heteroduplex (positions 1-3) result in heteroduplex
562 unpairing, incomplete R-loop formation and structural repositioning of the REC3 domain
563 (**Figure 6**), indicating a perturbation of the Cas9 conformational checkpoint. We envision that
564 the observed conformational state mimics the structural effect of 5'-truncated guide RNAs,
565 which have been shown to improve targeting specificity (Fu et al., 2014). Furthermore,
566 similarities with the structure of a high-fidelity Cas9 variant (Guo et al., 2019) suggest a shared
567 underlying mechanism for increased specificity.

568 **Implications for off-target prediction**

569 Our structural data reveal that Cas9 plays a limited steric role in off-target discrimination insofar
570 as only sensing the integrity and general shape of the guide–target heteroduplex. Off-target
571 activity is thus largely determined by the kinetics and energetics of R-loop formation, *i.e.,* off-
572 target DNA strand separation and guide RNA–TS DNA hybridization, and the Cas9
573 conformational activation checkpoint. We observe on multiple occasions that a given mismatch
574 adopts different conformational arrangements depending on its position along the guide RNA–
575 TS DNA heteroduplex, as further supported by MD simulations of rA-dA mismatches. This

576 poses a challenge for *ab initio* modelling of off-target activity, as biophysical models of off-
577 target binding and cleavage are bound to be of limited accuracy unless they incorporate
578 position-dependent energetic penalties for each base mismatch type and for deletions, as well
579 as position- and base-specific penalties for insertions (Boyle et al., 2021; Jones et al., 2021;
580 Zhang et al., 2020). As illustrated by our study, MD simulations can complement experimental
581 data to provide structural information on specific mismatches in remaining positions within the
582 heteroduplex. Thus, ongoing structural and computational studies, combined with machine
583 learning approaches, will assist in generating complete models for off-target prediction.

584 Furthermore, as certain off-target sequences that are incompatible with dsDNA cleavage
585 can undergo NTS nicking (Fu et al., 2019; Jones et al., 2021; Murugan et al., 2020; Zeng et al.,
586 2018), future bioinformatic models need to be able to predict off-target nicking activity as well.
587 Furthermore, accurate modelling of off-target interactions remains difficult due to context-
588 dependent effects, as documented in previous studies showing that the binding and cleavage
589 defects of consecutive mismatches deviate from additivity (Boyle et al., 2021; Cameron et al.,
590 2017; Lazzarotto et al., 2020; Zhang et al., 2020). Indeed, our structural data rationalize this by
591 showing that the conformation of a given base mismatch is highly sensitive to the presence of
592 neighbouring mismatches. As seen in the case of *AAVS1* off-target #2 complex, multiple
593 mismatched bases can synergistically combine to preserve an on-target-like heteroduplex
594 conformation that passes the REC3 conformational checkpoint, supporting nearly on-target
595 efficiencies of cleavage (Zhang et al., 2020). This is in line with recent cryo-EM structural
596 studies suggesting that indirect readout of heteroduplex conformation is coupled to nuclease
597 activation, while mismatches disrupt this coupling (Bravo et al., 2022; Pacesa et al., 2022).
598 Critically, reversion of one of the mismatches in this off-target substrate impairs cleavage
599 activity. Similar effects have been described for other DNA binding proteins such as
600 transcription factors, where mismatches modulate transcription factor binding activity by

601 affecting the conformation of the DNA duplex (Afek et al., 2020). In an analogy with Cas9,
602 these proteins check for correct binding sites through indirect sequence readout by sampling
603 for the correct duplex shape rather than base sequence (Abe et al., 2015; Kitayner et al., 2010;
604 Rohs et al., 2009a; Rohs et al., 2009b).

605 In conclusion, structural insights presented in this study establish an initial framework
606 for understanding the molecular basis for the off-target activity of Cas9. In conjunction with
607 ongoing computational studies, these findings will help achieve improved energetic
608 parametrization of off-target mismatches and deletions/insertions, thus contributing to the
609 development of more accurate off-target prediction algorithms and more specific guide RNA
610 designs. In doing so, these studies will contribute towards increasing the precision of CRISPR-
611 Cas9 genome editing and the safety of its therapeutic applications.

612 Limitations of the study

613 The structural dataset presented in this study is necessarily restricted in scope. Although all
614 mismatch types are covered, a much larger collection of off-target complex structures would be
615 required to cover all mismatches at all positions to achieve a complete structural overview of
616 the mismatch tolerance underpinning intrinsic off-target activity of Cas9. Although MD
617 simulations can help fill in the gaps in the experimental data, our findings suggest that mismatch
618 accommodation appears to be sequence context-dependent, thus limiting their predictive power.
619 Although we provide insights into off-target substrates containing single-nucleotide deletions,
620 it is presently difficult to predict which mechanism occurs at a given heteroduplex position.
621 Moreover, we currently lack structural data for off-target substrates containing insertions,
622 which will be addressed in future studies. Finally, not all off-targets (~50%) targets examined
623 in this study are efficiently cleaved in cells, despite detectable cleavage in vitro, implying that
624 other factors, including chromatin structure and DNA repair, affect Cas9 off-targeting in the

625 context of eukaryotic genomes. Thus, further work combining structural and computational
626 approaches will be needed to accurately predict the off-target activity of Cas9.

627 ## Author contributions

628 M.P., P.C., P.D.D., and M.J. conceived the study. M.P. purified wild-type Cas9, performed *in*
629 *vitro* cleavage assays, crystallized ternary Cas9 complexes, solved the structures, and
630 performed structural analysis along with M.J.; A.C. performed switchSENSE binding
631 measurements, under the supervision of F.H.T.A; A.S. and P.R.A. performed molecular
632 dynamics simulations, under the supervision of G.P.; M.J.I. performed the SITE-Seq assay; C-
633 H.L. wrote the computational off-target classification model and P.D.D. and P.C. analysed the
634 output; K.B. purified dCas9, transcribed sgRNAs, and prepared DNA substrates for in vitro
635 cleavage assays; M.P., P.C., P.D.D., and M.J. wrote the manuscript.

636

637 ## Conflict of interest statement

638 P.D.D. and M.J.I are current employees of Caribou Biosciences, Inc., and C-H.L. and P.C. are
639 former employees of Caribou Biosciences, Inc. M.J. is a co-founder of Caribou Biosciences,
640 Inc. M.J., C-H.L., M.J.I., P.C. and P.D.D. are named inventors on patents and patent
641 applications related to CRISPR-Cas technologies.

642

643 ## Acknowledgements

644 We thank members of the Jinek laboratory and Caribou Biosciences, Inc. for discussion and
645 critical reading of the manuscript. We thank Christelle Chanez and Michael Schmitz for their
646 assistance with preparing reagents. We thank Kenny B. Jungfer for assistance with
647 crystallographic data collection. We thank John Hawkins and Ilya Finkelstein for providing
648 access to the NucleaSEQ prediction pipeline. We thank Vincent Olieric, Meitian Wang, and
649 Takashi Tomizaki (Swiss Light Source, Paul Scherrer Institute) for assistance with
650 crystallographic data collection and Nena Matscheko from Dynamic Biosensors for support
651 with switchSENSE experiments. We thank the NCCR RNA and Disease for providing

infrastructural support. This work was supported by the Swiss National Science Foundation Grant 31003A_182567 (to M.J.). M.J. is an International Research Scholar of the Howard Hughes Medical Institute and Vallee Scholar of the Bert L & N Kuggie Vallee Foundation. G.P. would like to thank the National Science Foundation under Grant Number CHE-1905374 and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM141329 for supporting the research reported in this publication.

658 ## FIGURES AND LEGENDS

659 **Figure 1. Biochemical and structural analysis of Cas9 off-targets.**

660 (**A**) Guide RNA and (off-)target DNA sequences selected for biochemical and structural
661 analysis. Matching bases in off-targets are denoted by a dot; nucleotide mismatches and
662 deletions (–) are highlighted. (**B**) Top: Schematic representation of the guide RNA (orange),
663 TS (blue), and NTS (black) sequences used for crystallisation. The PAM sequence in the DNA
664 is highlighted in yellow. Bottom: Structure of the Cas9 *FANCF* on-target complex. Individual
665 Cas9 domains are coloured according to the legend; substrate DNA target strand (TS) is
666 coloured blue, non-target strand (NTS) black, and the guide RNA orange.

667 **Figure 2. Cas9 off-target binding is enabled by non-canonical base pairing.**

668 Close-up views of (**A**) rU-dG wobble base pair at duplex position 10 in *FANCF* off-target #2
669 complex, (**B**) rA-dC wobble base pair at position 4 in *FANCF* off-target #2 complex, (**C**) rG-
670 dA Hoogsteen base pair at duplex position 11 in *AAVS1* off-target #4 complex and (**D**) rG-dG
671 Hoogsteen base pair at duplex position 13 in *FANCF* off-target #5 complex. Hydrogen bonding
672 interactions are indicated with dashed lines. Numbers indicate interatomic distances in Å.
673 Corresponding on-target Watson-Crick base pairs are shown in white. Dashed arrows indicate
674 *anti-syn* isomerization of the dA and rG bases to enable Hoogsteen-edge base pairing. A bound
675 monovalent ion, modelled as K^+, is depicted as a purple sphere.

676 **Figure 3. Duplex backbone distortions facilitate formation of non-canonical base pairs.**

677 (**A**) Close-up view of the rU-dC base pair at duplex position 9 in *FANCF* off-target #1 complex,
678 facilitated by lateral displacement of the guide RNA backbone. (**B**) Zoomed-in view of the rU-
679 dT base pair at position 9 in *FANCF* off-target #5 complex. (**C**) Zoomed-in view of the rC-dC
680 mismatch at duplex position 8 in *AAVS1* off-target #3 complex. The distances between the
681 cytosine bases indicate lack of hydrogen bonding. (**D**) Zoomed-in view of the rA-dA mismatch
682 at duplex position 5 in *PTPRC-tgt2* off-target #1 complex. Hydrogen bonding interactions are

683 indicated with dashed lines. Corresponding on-target Watson-Crick base pairs are shown in
684 white (for *PTPRC-tgt2*, the *FANCF* on-target structure was used and bases were mutated *in*
685 *silico*). Numbers indicate interatomic distances in Å. Bound water molecule is depicted as red
686 sphere.

687 **Figure 4. TS distortion facilitates mismatch accommodation in the seed region of the**
688 **guide–off-target heteroduplex.**

689 (**A**) Close-up view of the rA-dA mismatch at position 18 in *FANCF* off-target #6 complex,
690 showing major groove extrusion of the dA base. (**B**) Close-up view of the rA-dA mismatch at
691 position 19 in *AAVS1* off-target #2 complex, showing retention of the dA base in the duplex
692 stack. (**C**) Close-up rU-dT mismatch at the PAM-proximal position 20 in *AAVS1* off-target #4
693 complex. Residual electron density indicates the presence of an ion or solvent molecule.
694 Refined $2mF_o-DF_c$ electron density map of the heteroduplex, contoured at 1.5σ, is rendered as
695 a grey mesh. Structurally disordered thymine nucleobase for which no unambiguous density is
696 present is in grey. (**D**) Zoomed-in view of the rA-dG base pair at position 19 and the unpaired
697 rU-dG mismatch at position 20 in *AAVS1* off-target #5 complex. Arrows indicate
698 conformational changes in the TS backbone relative to the on-target complex.

699 **Figure 5. Off-targets with single-nucleotide deletions are accommodated by base skipping**
700 **or multiple consecutive mismatches**.

701 (**A**) Zoomed-in view of the base skip at duplex position 15 in the *PTPRC-tgt2* off-target #1
702 complex. (**B**) Zoomed-in view of the base skip at duplex position 17 in the *FANCF* off-target
703 #3 complex. (**C**) Schematic depiction of alternative base pairing interactions in the *AAVS1* off-
704 target #2 complex. *AAVS1* off-target #2-rev substrate was designed based on the *AAVS1* off-
705 target #2, with the reversal of a single mismatch in the consecutive region back to the
706 corresponding canonical base pair. (**D**) Structural overlay of the *AAVS1* off-target #2 (coloured)

707 and *AAVS1* on-target (white) heteroduplexes. (**E**) Cleavage DNA kinetics of AAVS1 on-target,
708 off-target #2 and off-target #2-rev substrates.

709 **Figure 6. *FANCF* off-target #4 exhibits conformational changes in the REC2/3 and HNH**
710 **domains due to PAM-distal duplex unpairing.**

711 (**A**) Close-up view of the unpairing of mismatched bases at the PAM-distal end of the *FANCF*
712 off-target #4 heteroduplex. The last two nucleotides on each strand could not be modelled due
713 to structural disorder. (**B**) Overlay of the *FANCF* off-target #4 and *FANCF* on-target complex
714 structures. The *FANCF* off-target #4 complex is coloured according to the domain legend in
715 **Figure 1A**, *FANCF* on-target complex is shown in white. The REC1, RuvC, and PAM-
716 interaction domains have been omitted for clarity, as no structural differences are observed.

717

718

719 ## SUPPLEMENTAL FIGURES AND TABLES

720 **Figure S1. Off-target profiling of selected genomic sites using SITE-Seq.**

721 (**A**) Selected genomic targets and the corresponding guide RNA sequences selected for the
722 SITE-Seq assay off-target profiling. Heatmap indicates frequency of nucleotide identity across
723 each position for the selected targets. (**B**) SITE-seq assay analysis for RNPs assembled with
724 indicated crRNAs. The numbers of detected off-target sites are shown as a function of RNP
725 concentration. Checked boxes indicate recovery of the on-target site. n=3 replicates per sample.
726 (**C**) Number of off-target sites recovered by the SITE-Seq assay are shown as a function of the
727 number of mismatches between the guide RNA and the off-target sequence. (**D**) Frequency of
728 nucleotide mismatches at each guide RNA–off-target DNA heteroduplex position for all off-
729 target sites identified in (B). (**E**) Number of total identified mismatches per heteroduplex
730 position.

731

732 **Figure S2. *in vitro* cleavage of selected Cas9 off-target substrates.**

733 (**A**) *In vitro* clevage kinetics of fully double stranded on- and off-target synthetic DNA
734 substrates labelled on the target-strand for each guide RNA used in the study. Black triangles
735 in the substrate schematic (top) indicate position of cleavage sites. Each data point represents a
736 mean of four independent replicates. Error bars represent standard deviation for each time point.
737 (**B**) *In vitro* clevage kinetics of partially single stranded (PAMmer) on- and off-target substrates.
738 (**C**) Heatmap representation of mutual correlations between measured kinetic and
739 thermodymamic parameters including cleavage (k_{obs}), substrate DNA binding (k_{on}), substrate
740 dissociation (k_{off}) rate constants, equilibrium dissociation constant (K_d) with numbers of
741 nucleotide mismatches in the off- target sites (total and within seed), the GC content of the
742 spacer (%GC) and cleavage rate predicted using the NucleaSeq algorithm (NucleaSeq k_{obs}).
743 The values were calculated across all off-targets for both dsDNA (lower left half, in blue), and

744 partially single stranded (PAMmer) substrates (upper right half, in red). ns, no significant
745 correlation. (**D**) Correlation between measured and NucleaSeq-predicted k_{obs} rate constants.
746 Off-target sites with significant deviations are highlighted in yellow.
747
748 **Figure S3. On and off-target cellular editing activity.**
749 **(A)** Editing activity of crystalized on- and off-targets in primary human T cells. Matching bases
750 in off-targets are denoted by a dot, whereas base pair mismatches and deletions (–) are
751 highlighted. Asterisk indicated single-nucleotide polymorphism in PBMC donor sequence
752 different from hg38 genomic reference. Editing rates are from highest Cas9 concentration
753 shown in (B), n=3 replicates per sample, limit of detection = 0.1%. **(B)** Dose titration of Cas9
754 in human T cells for selected on and off-target sites. Off-target numbering correspond to sample
755 shown in **(A)**, and/or reference number ("ref") as indicated in **Table S1**. Cas9 concentration
756 shown below x-axis (1:3 molar ratio of Cas9 to guides), N=3 replicates per sample.
757
758 **Figure S4. Alternative REC2 conformation in AAVS1 on-target.**
759 (**A**) Overlay of REC2 domain conformations in the *AAVS1* (pink), *FANCF* (purple) and *TRAC*
760 (light blue) on-target complexes (**B**) Close-up view of helix REC2 helix spanning Cas9 residues
761 174-180. Linear and angular displacements of the helix in the *AAVS1* on-target complex relative
762 to the *FANCF* and *TRAC* on-target complexes are indicated.
763
764 **Figure S5. PAM-distal mismatches result in unpairing and disordering of guide RNA**
765 **nucleobase in position 1.**
766 Close-up views of the PAM-distal end of the guide RNA-TS heteroduplex in (**A**) *AAVS1* off-
767 target #2, (**B**) *AAVS1* off-target #4, (**C**) *FANCF* off-target #5 and (**D**) *TRAC* off-target #1
768 complexes. Arrowheads indicate nucleotides with disordered bases. Refined $2mF_o-DF_c$

769 electron density maps of the heteroduplexes are rendered as a grey mesh and contoured at 1.2σ
770 for (A) and 1.0σ for (B)-(D).
771
772 **Figure S6. Cas9-nucleic acid interactions in on-target complexes.**
773 Schematic diagram depicting Cas9 residues interacting with the guide RNA-target DNA
774 heteroduplex. Dotted lines represent hydrogen bonding interactions; dashed lines represent salt
775 bridges; solid lines represent stacking/hydrophobic interactions. Target strand is coloured in
776 blue, guide RNA in orange. Phosphates are represented by circles, ribose moieties by
777 pentagons, and nucleobases by rectangles.
778
779 **Figure S7. Wobble base pairing of rG-dT mismatches.**
780 Close-up views of rG-dT mismatches at (**A**) heteroduplex position 15 in the *TRAC* off-target
781 #1 complex, (**B**) position 13 in *FANCF* off-target #7 complex, (**C**) position 13 in *FANCF* off-
782 target #2 complex, (**D**) position 4 in *AAVS1* off-target #5 complex, (**E**) position 4 in *AAVS1* off-
783 target #1 complex, (**F**) position 2 in *TRAC* off-target #1 complex and (**G**) position 2 in *TRAC*
784 off-target #2 complex. Corresponding on-target Watson-Crick base pairs are shown in white.
785 Monovalent ions, modeled as K^+, are depicted as purple spheres. In (A)-(E), the dT base is
786 displaced into the major groove and forms a canonical wobble base pair with the rG base. In
787 (F)-(G), the the rG base instead shifts towards the minor groove to facilitate wobble pairing.
788
789 **Figure S8. rU-dG wobble base pairs adopt duplex position-dependent conformations.**
790 Close-up views of of rU-dG mismatches at (**A**) heteroduplex position 10 in *FANCF* off-target
791 #4 complex, (**B**) position 5 in *FANCF* off-target #1 complex, (**C**) position 5 in *FANCF* off-
792 target #3 complex, **(D**) position 5 in *FANCF* off-target #6 complex, (**E**) *TRAC* off-target #1
793 complex, and (**F**) *AAVS1* off-target #2, and (**G**) *AAVS1* off-target #1 complex. Corresponding

794 on-target Watson-Crick base pairs are shown in white. Bound potassium ions are depicted as
795 purple spheres. In (A), rU-dG wobble base pairing is achieved by minor groove displacement
796 of the guanine base. In (B)-(D), the rU-dG mismatches adopt atypical conformations. In (E),
797 the guanine base is shifted into the minor groove to form a wobble base pair, whereas at the
798 identical heteroduplex position in (F) and (G), the rU-dG base pairs do not engage in wobble
799 pairing, instead adopting alternative tautomeric forms. **(H)** Schematic depicting hydrogen
800 bonding interactions between rU and dG bases in (F) and (G).
801
802 **Figure S9. Additional non-canonical base pairs within Cas9 off-target complexes.**
803 (**A**) Close-up view of rA-dC wobble base pairing at position 4 in *FANCF* off-target #3 complex.
804 The base pair geometry is consistent with base protonation or tautomerism to enable productive
805 hydrogen bonding between the bases. (**B**) Close-up view of rC-dA mismatch at position 11 of
806 *FANCF* off-target #7 complex, facilitated by base tilting at positions 11 and 12. (**C**) Close-up
807 view of partially paired rC-dA mismatch at position 8 in *AAVS1* off-target #1 complex. (**D**)
808 Close-up view of rA-dC mismatch at position 10 in *AAVS1* off-target #5 complex. (**E**) Close-
809 up view of Hoogsteen-edge rA-dG base pair at position 7 in *AAVS1* off-target #2 complex.
810 Corresponding on-target Watson-Crick base pairs are shown in white.
811
812 **Figure S10. Preservation of base stacking in pyrimidine-pyrimidine off-target**
813 **mismatches.**
814 (**A**) Close-up view of rC-dT mismatch at position 6 in *FANCF* off-target #6 complex. (**B**) Close-
815 up view of rC-dT mismatch at position 7 in *FANCF* off-target #7 complex, bridged by Arg895.
816 The arginine sidechain in the corresponding on-target complex is shown in white. (**C**) Close-
817 up view of rC-dT base pairing at position 8 of *AAVS1* off-target #2. (**D**) Close-up view of rU-
818 dT mismatch at position 7 in *TRAC* off-target #2 complex. (**E**) Close-up view of rU-dT pairing

819 at position 9 in *PTPRC*-tgt2 off-target #1 complex, facilitated by base propeller twisting. (**F**)
820 Close-up view of rU-dT pairing at position 8 in *TRAC* off-target #2 complex, enabled by
821 backbone shift of the RNA strand. (**G**) Close-up view of partially paired rC-dC mismatch at
822 position 5 in *AAVS1* off-target #2 complex, bridged by a water molecule. (**H**) Close-up view of
823 rC-dC mismatch at position 15 in *AAVS1* off-target #3 complex. Corresponding on-target
824 Watson-Crick base pairs are shown in white (for *PTPRC-tgt2*, the *FANCF* on-target structure
825 was used and bases were mutated *in silico*).
826
827 **Figure S11. Cas9 R895A mutation of Cas9 has no significant impact on FANCF off-target**
828 **#7 cleavage or binding.**
829 (**A**) Kinetic analysis of *FANCF* on- and off-target substrate DNA cleavage by wild-type and
830 R895A Cas9 proteins.
831
832 **Figure S12. Tolerance of adenine-adenine mismatches within the heteroduplex.**
833 (**A**) Close-up view of rA-dA mismatch at position 18 in *FANCF* off-target #6 complex, overlaid
834 with the *FANCF* on-target structure (white). (**B**) Number of rA-dA off-target mismatches per
835 heteroduplex position recovered in the SITE-Seq assay for all analysed genomic targets.
836 Percentages indicate frequency of rA-dA mismatches recovered in the particular position as a
837 fraction of total number of rA-dA mismatches. (**C**) Close-up view of MD-simulated rA-dA
838 mismatches (colored), overlaid with crystallographic ground truth (white) at positions 5
839 (PTPRC-tgt2 off-target #1), 18 (FANCF off-target #6), and 19 (AAVS1 off-target #2). (**D**)
840 Probability density plots of the most representative geometrical base pair parameters (viz., shear
841 in blue, opening in violet, stretch in orange and buckle in red), describing the mode of rA-dA
842 mismatch tolerance at different positions (1 – 20) of the heteroduplex, as computed via MD
843 simulations for the systems with (shaded) and without mismatches (unshaded). Vertical line in

844 positions 5, 18 and 19 indicates base pair parameter value of the respective X-ray structure. For
845 each system, data are reported considering an aggregate sampling of ~1.5 μs.
846
847 **Figure S13. Base pair parameter probability density plots of MD simulated rA-dA**
848 **mismatches.** (**A**) Shear, (**B**) stretch, (**C**) buckle, (**D**) opening, (**E**) stagger, and (**F**) propeller-
849 twist base pair parameter for the rA-dA mismatch at different heteroduplex, as computed from
850 molecular dynamics simulations for the systems with (shaded) and without mismatches
851 (unshaded). For each system, data are reported considering an aggregate sampling of ~1.5 μs.
852
853 **Figure S14. Lack of protein contacts with the target DNA strand in the seed region allows**
854 **for large phosphate backbone distortions.**
855 (**A**) Schematic overview of Cas9 interactions within the PAM-proximal seed region of the guide
856 RNA-TS DNA heteroduplex. (**B**) Close-up view of the seed region in *AAVS1* off-target #2
857 complex, overlaid with the AAVS1 on-target heteroduplex (white), showing structural
858 distortion of the TS due to rA-dA mismatch at seed position 19. (**C**) Close-up view of the seed
859 region in AAVS1 off-target #5 complex, overlaid with the AAVS1 on-target heteroduplex
860 (white), showing structural distortion due to rA-dG and rU-dG mismatches at positions 19 and
861 20, respectively. Red lock icon indicates position of the phosphate lock residue in (B) and (C).
862
863 **Figure S15. Schematic representation of mismatch, insertion, and deletion classification**
864 **algorithm for the SITE-Seq assay analysis of off-target sites.**
865 Schematic represents classification algorithm of off-target sites with putative insertions and
866 deletions.
867

868 **Figure S16. Positional mapping of nucleotide insertions/deletions during of SITE-Seq**
869 **assay off-target profiling.**
870 (**A**) Number of recovered off-target sites per genomic target as a function of RNP concentration
871 classified as containing either only mismatches, single-nucleotide deletions, or single-
872 nucleotide insertions. (**B**) Frequency of positional mismatch occurrence per genomic target for
873 mismatched off-targets. (**C**) Frequency of nucleotide mismatches within the heteroduplex for
874 all off-target sites (n=3445 sites for both (B) and (C)). (**D**) Frequency of single-nucleotide
875 deletions for all off-target sites. (n=277 sites). (**E**) Frequency of single-nucleotide insertions for
876 all off-target sites. (n=116 sites).
877
878 **Figure S17. Recognition of off-target sites containing deletions in the seed region.**
879 (**A**) Close-up view of base skipping within the seed region of the guide RNA-off-target DNA
880 heteroduplex in *FANCF* off-target #3 complex, overlaid with the on-target heteroduplex
881 (white). (**B**) Close-up view of base skipping within the seed region of the guide RNA-off-target
882 DNA heteroduplex in *PTPRC-tgt2* off-target #1 complex, overlaid with *FANCF* on-target
883 heteroduplex (white). (**C**) Close-up view of non-canonical base pairs at the 5'-terminus of the
884 guide RNA in *FANCF* off-target #3 complex involving guanosine nucleotides introduced
885 during in vitro transcription of the guide RNA. (**D**) Close-up view of non-canonical base pairs
886 at the 5'-terminus of the guide RNA in *PTPRC-tgt2* off-target #1 complex involving guanosine
887 nucleotides introduced during in vitro transcription of the guide RNA.
888
889 **Figure S18. Reversal of a single mismatch in the PAM-distal region reduces the cleavage**
890 **efficiency of off-targets containing multiple consecutive mismatches**
891 (**A**) Schematic depiction of possible mismatch tolerance mechanisms through PAM-distal
892 deletions (denoted -A) or the formation of multiple mismatches (denoted -B) in *AAVS1* off-

893 targets. Off-targets denoted with "-rev" indicate the reversal of a single mismatch in the
894 consecutive mismatch region back to the corresponding canonical base pair. *In vitro* cleavage
895 kinetics of fully double stranded (**B**) partially single stranded (PAMmer) (**C**) on- and off-target
896 DNA substrates for off-targets with putative multiple consecutive mismatches in the PAM-
897 distal region. Black triangles in the substrate schematic (top) indicate position of cleavage sites.
898 Each data point represents a mean of four independent replicates. Error bars represent standard
899 deviation for each time point.
900
901 **Figure S19. PAM-distal RNA base skip in *CD34* off-target #9 induces REC lobe**
902 **rearrangement.**
903 (**A**) Schematic depiction of a PAM-distal RNA-skip in the sequence of *CD34* off-target #9. *In*
904 *vitro* cleavage kinetics of fully double stranded (**B**) partially single stranded (PAMmer) (**C**)
905 *CD34* on-target and off-target #9 DNA substrates. Black triangles in the substrate schematic
906 (top) indicate position of cleavage sites. Each data point represents a mean of four independent
907 replicates. Error bars represent standard deviation for each time point. (**D**) Close-up view of
908 rU-dT mismatch at position 17 in *CD34* off-target #9 complex, where the rU17 nucleotide
909 undergoes *anti-syn* isomerisation. Corresponding on-target Watson-Crick base pairs are shown
910 in white (the *TRAC* on-target structure was used and bases were mutated *in silico*). (**E**) Zoomed-
911 in view of the PAM-distal base skip at duplex position 6 in the *CD34* off-target #9 complex.
912 (**F**) Close-up view of rG-dC Watson-Crick base pair at the 5'-terminus of the guide RNA in
913 *CD34* off-target #9 complex, involving guanosine nucleotides introduced during in vitro
914 transcription of the guide RNA. (**G**) Structural overlay of the *CD34* off-target #9 complex with
915 the *TRAC* on-target complex. The REC1, RuvC, and PAM interaction domains have been
916 omitted for clarity in all panels. The *CD34* off-target #9 complex domains are colored according
917 to **Figure 1A**. The on-target complex is colored white.

918

919 **Figure S20. Altered heteroduplex interactions in *FANCF* off-target #4 complex.**

920 (**A**) Overlay of REC3 domain helix 703-712 in *FANCF* off-target #4 complex (wheat) with
921 *FANCF* on-target complex (orange). (**B**) Close-up view of REC3 domain interactions with the
922 guide RNA strand in *FANCF* off-target #4 complex. (**C**) Close-up view of TS DNA interactions
923 established by HNH domain in *FANCF* off-target #4 complex. (**D**) Schematic diagram
924 depicting Cas9 residues interacting with the guide RNA-off-target DNA heteroduplex in
925 *FANCF* off-target #4 complex. Dotted lines represent hydrogen bonding interactions, dashed
926 lines represent salt bridges, solid lines represent stacking/hydrophobic interactions. Target
927 strand is coloured blue, guide RNA orange. Phosphates are represented by circles, ribose
928 moieties by pentagons, and nucleobases by rectangles.

929

930 **Figure S21. Conformational rearrangements of REC2/3 AND HNH domains in *FANCF***
931 **off-target #4 complex.**

932 (**A**) Structural overlay of the *FANCF* off-target #4 complex with cryo-EM structure of a pre-
933 catalytic (State I) Cas9 complex (PDB: 6O0Z). (**B**) Structural overlay of the *FANCF* off-target
934 #4 complex with the cryo-EM structure of a post-catalytic (State II) Cas9 complex (PDB:
935 6O0Y). (**C**) Structural overlay of the *FANCF* off-target #4 complex with the crystallographic
936 structure of the high-fidelity xCas9 3.7 variant (PDB: 6AEG). The REC1, RuvC, and PAM
937 interaction domains have been omitted for clarity in all panels, as no significant structural
938 changes were observed in these domains. The *FANCF* off-target #4 complex domains are
939 colored according to **Figure 1A**. The overlaid structures are coloured white. (**D**) Overlay of the
940 PAM-distal heteroduplex region in *FANCF* off-target #4 and xCas9 3.7 on-target complexes.
941 Target strand is coloured in blue, guide RNA is coloured orange.

942

943 **Table 1. Biochemical analysis of the binding and cleavage efficiency of selected off-target**
944 **sites.**

945 Kinetic and thermodynamic parameters of off-target substrates. The cleavage rate constants
946 (k_{obs}) were derived from single-exponential function fitting of measured cleavage rates of the
947 target-strand labelled DNA substrates. The binding and dissociation rate constants (k_{on} and k_{off})
948 and the equilibrium dissociation constant (K_d) were determined using a DNA nanolever binding
949 (switchSENSE) assay. Intervals indicate standard error of mean.

950

Gene	Target	24 h cleavage (%)	k_{obs} (min^{-1})	k_{on} ($M^{-1} . s^{-1}$)	k_{off} (s^{-1})	K_d (pM)	24 h cleavage (%) PAMmer	k_{obs} (min^{-1}) PAMmer
AAVS1	on-target	92.2	1.624 ± 0.159	$3.95 \pm 0.26 \times 10^{6}$	$5.74 \pm 0.73 \times 10^{-5}$	14.5 ± 2.1	95.0	0.562 ± 0.073
AAVS1	off-target #1	92.7	0.071 ± 0.005	$4.73 \pm 0.71 \times 10^{6}$	$6.29 \pm 0.29 \times 10^{5}$	13.3 ± 2.1	87.5	0.239 ± 0.031
AAVS1	off-target #2	95.1	0.034 ± 0.002	$8.75 \pm 0.31 \times 10^{6}$	$3.38 \pm 0.08 \times 10^{-3}$	386 ± 16	85.5	0.004 ± 0.0001
AAVS1	off-target #3	96.7	0.051 ± 0.002	$3.30 \pm 0.11 \times 10^{6}$	$2.51 \pm 0.15 \times 10^{-3}$	761 ± 51	82.7	0.081 ± 0.010
AAVS1	off-target #4	94.4	0.004 ± 0.0002	$1.09 \pm 0.03 \times 10^{7}$	$3.28 \pm 0.06 \times 10^{-3}$	301 ± 9	89.6	0.065 ± 0.007
AAVS1	off-target #5	18.9	0.00013 ± 0.0003	ND	ND	ND	70.1	0.007 ± 0.0001
FANCF	on-target	97.5	0.238 ± 0.013	$3.45 \pm 0.19 \times 10^{6}$	$7.46 \pm 0.97 \times 10^{-5}$	21.6 ± 3.1	98.3	0.565 ± 0.054
FANCF	off-target #1	35.1	0.001 ± 0.0001	$3.97 \pm 0.06 \times 10^{6}$	$2.09 \pm 0.06 \times 10^{-3}$	528 ± 17	97.4	0.069 ± 0.006
FANCF	off-target #2	62.4	0.001 ± 0.0002	$1.42 \pm 0.03 \times 10^{6}$	$2.45 \pm 0.06 \times 10^{-3}$	1730 ± 60	92.9	0.233 ± 0.018
FANCF	off-target #3	0.0	0	$1.22 \pm 0.05 \times 10^{7}$	$2.37 \pm 0.11 \times 10^{-3}$	193 ± 13	4.2	0.00003 ± 0.0003

FANCF	off-target #4	53.0	0.0005 ± 0.0002	$3.35 \pm 0.05 \times 10^{6}$	$1.91 \pm 0.07 \times 10^{-3}$	571 ± 24	38.9	0.001 ± 0.0001
FANCF	off-target #5	80.4	0.001 ± 0.0001	$1.27 \pm 0.05 \times 10^{6}$	$2.55 \pm 0.10 \times 10^{-3}$	2010 ± 110	92.9	0.058 ± 0.005
FANCF	off-target #6	8.2	0.00006 ± 0.0003	$1.50 \pm 0.09 \times 10^{6}$	$2.03 \pm 0.15 \times 10^{-3}$	1350 ± 130	66.6	0.001 ± 0.00005
FANCF	off-target #7	5.2	0.004 ± 0.0003	$2.95 \pm 0.20 \times 10^{6}$	$3.21 \pm 0.23 \times 10^{-3}$	1090 ± 110	94.5	0.013 ± 0.001
PTPRC-tgt2	on-target	96.8	0.459 ± 0.022	$6.08 \pm 0.39 \times 10^{6}$	$2.19 \pm 0.21 \times 10^{-4}$	36.0± 4.2	95.5	0.074 ± 0.015
PTPRC-tgt2	off-target #1	0.0	0	$1.22 \pm 0.07 \times 10^{7}$	$2.39 \pm 0.12 \times 10^{-3}$	196 ± 15	91.4	0.001 ± 0.0003
TRAC	on-target	97.7	0.381 ± 0.015	$1.02 \pm 0.21 \times 10^{6}$	$3.23 \pm 0.53 \times 10^{-4}$	31.8 ± 8.5	93.5	0.181 ± 0.013
TRAC	off-target #1	95.8	0.019 ± 0.001	$1.37 \pm 0.45 \times 10^{6}$	$1.77 \pm 0.1 \times 10^{-4}$	130 ± 44	90.7	0.081 ± 0.006
TRAC	off-target #2	65.0	0.001 ± 0.00006	$9.43 \pm 1.12 \times 10^{6}$	$3.27 \pm 0.27 \times 10^{-4}$	34.6 ± 5.0	88.4	0.026 ± 0.002

Table S1. SITE-Seq assay results for Cas9 off-target profiling of 12 selected genomic sites. Related to Figure S1

Columns indicate the recovered off-target sequence; motif location; number of substitutions in recovered target sequence compared to the on-target (substitutions); strand designation of PAM; the lowest recovery concentration of each target; and whether the off-target is predicted to contain inserts or deletions based. Off-target sites recovered at lower concentrations were also recovered at higher concentrations (*e.g.,* all 4 nM sites were also recovered at 16 nM, 64 nM, and 256 nM).

Table S2. List of recovered off-target sequences aligned to the corresponding on-target sequence. Related to Figures S1 and S19

963 Off-target alignments classified by genomic target and by the presence of insertions, deletions
964 or purely mismatched targets. Indexes correspond to off-target sequence numbering in Table
965 S1.
966
967 **Table S3. Cellular editing data, Related to Figure S3**
968
969 **Table S4. Crystallographic data collection and refinement statistics of Cas9 on-target and**
970 **off-target complexes, Related to STAR methods**
971
972 **Table S5. Summary of non-canonical base pairing and nucleobase distortions observed in**
973 **structural data, Related to Figure 1**
974
975 **Table S6. 3DNA 2.0 analysis of the helical parameters and sugar puckering of**
976 **characterised on-target and off-target duplexes, Related to Figure 1**
977 **Table S7. Effect of Cas9 R895A on off-target cleavage kinetics, Related to Figure S11**
978 Kinetic and thermodynamic parameters of *FANCF* on- and off-target substrate DNA cleavage
979 by wild-type and R895A Cas9. Cleavage rate constants (k_{obs}) were derived from single-
980 exponential function fitting. Substrate binding and dissociation rate constants (k_{on} and k_{off}) and
981 the equilibrium dissociation constant (K_d) were determined using a DNA nanolever
982 (switchSENSE) binding assay. Intervals indicate standard error of mean.
983
984 **Table S8. Cleavage kinetics of off-targets with predicted PAM-distal deletions, Related to**
985 **Figures S18 and S19**
986 Kinetic and thermodynamic parameters determined for *AAVS1* and *CD34* off-targets with
987 predicted PAM-distal deletions. The apparent cleavage rate constants (k_{obs}) were derived from

988 a single-exponential function fitting of measured cleavage. Substrate binding (k_{on}) and
989 dissociation (k_{off}) constants were determined using SwitchSENSE assay. Intervals indicate
990 standard error of mean. NM indicates that values were not measured.
991
992 **Table S9. List of oligonucleotides used in this study, Related to STAR methods**
993
994 **Video S1. Morph of REC2 domain between AAVS1 on-target and FANCF on-target**
995 **structures, Related to Figure S4**
996
997 **Video S2. Morph between FANCF off-target #4 and FANCF on-target showing**
998 **conformational changes in the REC2, REC3, and HNH domains, Related to Figure S20**
999
1000 **Data S1. SwitchSENSE association and dissociation curves, Related to Table 1**
1001
1002

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Martin Jinek (jinek@bioc.uzh.ch)

Material availability

This study did not generate new unique reagents.

Data and code availability

- Coordinates have been deposited to the Protein Data Bank under the PDB accession numbers: 7QQO, 7QR7, 7QQP, 7QQQ, 7QQR, 7QQS, 7QQT, 7QQU, 7QQV, 7QQW, 7QQX, 7QR5, 7QQZ, 7QR8, 7QR0, 7QR1 and 7ZO1. Raw sequencing files for the SITE-Seq assay have been deposited to BioProject. Accession numbers are listed in the key resource table. All deposited data is publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

The *E. coli* strains Rosetta 2 (DE3) was used for recombinant protein expression for structural studies and biochemical *in vitro* experiments. T cells used for cellular editing experiments were isolated from human primary blood mononuclear cells were purchased from STEMCELL Technologies.

1026 # METHODS

1027 ## DNA oligonucleotides and substrates

1028 Sequences of DNA oligonucleotides used in this study are summarised in **Table S9**.
1029 Crystallisation substrates were synthesised by Sigma Aldrich without further purification,
1030 sgRNA transcription templates and ATTO-532-labelled cleavage substrates were synthesised
1031 by Integrated DNA Technologies, Inc., with PAGE and HPLC purification, respectively.
1032 Partially double stranded crystallisation substrates were prepared by mixing complementary
1033 oligonucleotides in a 1:1 molar ratio (as determined by 260 nm absorption), heating to 95 °C
1034 for 5 minutes and slow cooling to room-temperature. Cleavage substrates were prepared
1035 similarly, except that a 2-fold molar excess of the non-target strand was used.

1036 ## Cas9 protein expression and purification

1037 *Streptococcus pyogenes* Cas9 wild type protein and the nuclease dead mutant (D10A, H840A)
1038 were both recombinantly expressed for 16 hours at 18 °C in *Escherichia coli* Rosetta 2 (DE3)
1039 (Novagen) N-terminally fused to a hexahistidine affinity tag, the maltose binding protein
1040 (MBP) polypeptide, and the tobacco etch virus (TEV) protease cleavage site. Cells were
1041 resuspended and lysed in 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 5 mM imidazole, and
1042 supplemented with added protease inhibitors. Clarified lysate was loaded on a 10 ml Ni-NTA
1043 Superflow column (QIAGEN), washed with 7 column volumes of 20 mM HEPES-KOH pH
1044 7.5, 500 mM KCl, 5 mM imidazole, and eluted with 10 column volumes of 20 mM HEPES-
1045 KOH pH 7.5, 250 mM KCl, 200 mM imidazole. Salt concentration is adjusted and protein is
1046 loaded on a 10 ml HiTrap Heparin HP column (GE Healthcare) equilibrated in 20 mM HEPES-
1047 KOH pH 7.5, 250 mM KCl, 1 mM DTT. The column is washed with 5 column volumes of 20
1048 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT, and Cas9 is eluted with 17 column
1049 volumes of 20 mM HEPES-KOH pH 7.5, 1.5 M KCl, 1 mM DTT, in a 0-32% gradient (peak
1050 elution around 500 mM KCl). His_6-MBP tag was removed by TEV protease cleavage overnight

1051 with gentle shaking. The untagged Cas9 was concentrated and applied to a Superdex 200 16/600
1052 (GE Healthcare) and eluted with 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 1 mM DTT.
1053 Purified protein was concentrated to 10 mg/ml, flash frozen in liquid nitrogen and store
1054 at -80 °C. DTT was omitted in the size-exclusion step of the purification when protein was used
1055 for switchSENSE measurements.

1056 sgRNA transcription and purification

1057 sgRNAs are transcribed from a double stranded PCR product template amplified from a plasmid
1058 in a 5 ml transcription reaction (30 mM Tris-HCl pH 8.1, 25 mM $MgCl_2$, 2 mM spermidine,
1059 0.01% Triton X-100, 5 mM CTP, 5 mM ATP, 5 mM GTP, 5 mM UTP, 10 mM DTT, 1 µM
1060 DNA transcription template, 0.5 units inorganic pyrophosphatase (Thermo Fischer), 250 µg
1061 homemade T7 RNA polymerase. The reaction is incubated at 37 °C for 5 hours, and then treated
1062 for 30 minutes with 15 units of RQ1 DNAse (Promega). The transcribed sgRNAs are
1063 subsequently PAGE purified on an 8% denaturing (7 M urea) polyacrylamide gel, and lastly
1064 ethanol precipitated and resuspended in DEPC treated water.

1065 Crystallisation of Cas9 ternary complexes and structure determination

1066 To assemble the Cas9 on-/off-target ternary complexes, the Cas9 protein is first mixed with the
1067 sgRNA in a 1:1.5 molar ratio and incubated at room temperature for 10 minutes. Next, the
1068 binary complex is diluted to 2 mg/ml with 20 mM HEPES-KOH 7.5, 250 mM KCl, 1 mM DTT,
1069 2 mM $MgCl_2$ buffer, pre-annealed 100 µM DNA substrate is added in a 1:1.8 molar ratio and
1070 the complex is incubated another 10 minutes at room temperature. For crystallisation, 1 µl of
1071 the ternary complex (1-2 mg/ml) is mixed with 1 µl of the reservoir solution (0.1 M Tris-acetate
1072 pH 8.5, 0.3-0.5 M KSCN, 17-19% PEG3350) and crystals are grown at 20 °C using the hanging
1073 drop vapour diffusion method. In some cases, microseeding was be used to improve crystal
1074 morphology. Crystals are typically harvested after 2-3 weeks, cryoprotected in 0.1 M Tris-
1075 acetate pH 8.5, 0.4 M KSCN, 30% PEG3350, 15% ethylene glycol, 1 mM $MgCl_2$, and flash-

1076 cooled in liquid nitrogen. Diffraction data was obtained at beamlines PXI and PXIII of the
1077 Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) and were processed using
1078 the XDS package (Kabsch, 2010). Structures were solved by molecular replacement through
1079 the Phaser module of the Phenix package (Adams et al., 2010) using the PDB ID: 5FQ5 model
1080 omitting the RNA-DNA target duplex from the search. Model adjustment and duplex building
1081 was completed using COOT software (Emsley et al., 2010). Atomic model refinement was
1082 performed using Phenix.refine (Adams et al., 2010). Protein-nucleic acid interactions were
1083 analysed using the PISA web server (Krissinel and Henrick, 2007). Characterisation of the
1084 guide-protospacer duplex was performed using the 3DNA 2.0 web server (Li et al., 2019).
1085 Structural figures were generated using PyMOL and ChimeraX (Pettersen et al., 2021).

1086 *In vitro* nuclease activity assays

1087 Cleavage reactions were performed at 37 °C in reaction buffer, containing 20 mM HEPES pH
1088 7.5, 250 mM KCl, 5 mM $MgCl_2$ and 1 mM DTT. First, Cas9 protein was pre-incubated with
1089 sgRNA in 1:1.25 ratio for 10 minutes at room temperature. The protein-RNA complex was
1090 rapidly mixed with the target strand-ATTO-532-labelled dsDNA, to yield final concentrations
1091 of 1.67 µM protein and 66.67 nM substrate in a 7.5 µl reaction. Time points were harvested at
1092 1, 2.5, 5, 15, 45, 90, 150 minutes, and 24 hours. Cleavage was stopped by addition of 2 µl of
1093 250 mM EDTA, 0.5% SDS and 20 µg of Proteinase K. Formamide was added to the reactions
1094 with final concentration of 50%, samples were incubated at 95 °C for 10 minutes, and resolved
1095 on a 15% denaturing PAGE gel containing 7M urea and imaged using a Typhoon FLA 9500
1096 gel imager. Depicted error bars correspond to the standard deviation from four independent
1097 cleavage reactions. Rate constants (k_{obs}) were extracted from single exponential fits: [Product]
1098 $= A*(1-\exp(-k_{obs}*t))$

1099 switchSENSE analysis

1100 The target strands (TS) containing a 3' flanking sequence complementary to the ssDNA
1101 covalently bound to the chip electrode, and the non-target strands (NTS) (**Table S9**) were
1102 resuspended in a buffer containing 10 mM Tris-HCl pH 7.4, 40 mM NaCl, and 0.05% Tween
1103 20. The matching TS:NTS duplex is pre-annealed and hybridised to the chip anode. The Cas9
1104 protein was mixed with the sgRNAs at a 1:2 protein:RNA molar ratio, and the complex was
1105 incubated for 30 min at 37 °C in association buffer containing 20 mM HEPES-KOH pH 7.5,
1106 150 mM KCl, 2 mM $MgCl_2$, 0.01% Tween 20. All switchSENSE experiments were performed
1107 on a DRX analyser using CAS-48-1-R1-S chips (Dynamic Biosensors GmbH, Martinsried,
1108 Germany). Kinetics experiments were performed at 25 °C in association buffer, with an
1109 association time of 5 min, dissociation time of 20 min, and a flow rate of 50 µl/min.

1110 SITE-Seq assay

1111 SITE-Seq assay reaction conditions were performed as described previously (Cameron et al.,
1112 2017). Briefly, high molecular weight genomic DNA (gDNA) was purified from human
1113 primary T cells using the Blood & Cell Culture DNA Maxi Kit (Qiagen) according to the
1114 manufacturer's instructions. RNPs comprising the guides were biochemically assembled for
1115 gDNA digestion. Specifically, equal molar amounts of crRNA and tracrRNA were mixed and
1116 heated to 95 °C for 2 min then allowed to cool at room temperature for ~5 min. Three-fold
1117 molar excess of the guides were incubated with *Streptococcus pyogenes* Cas9 (SpCas9) in
1118 cleavage reaction buffer (20 mM HEPES pH 7.4, 150 mM KCl, 10 mM MgCl2, 5% glycerol)
1119 at 37 °C for 10 min. In a 96-well plate format, 10 µg of gDNA was treated with 0.2 pmol (4
1120 nM), 0.8 pmol (16 nM), 3.2 pmol (64 nM), and 12.8 pmol (256 nM) of each RNP in 50 µL total
1121 volume in cleavage reaction buffer. Each cleavage reaction was performed in triplicate.
1122 Negative control reactions were assembled in parallel and did not include RNP. gDNA was
1123 treated with RNPs for 4 hours at 37 °C. SITE-Seq assay library preparation and sequencing

1124 was performed as described previously and the final library was loaded onto the Illumina
1125 NextSeq platform (Illumina, San Diego, CA), and ~1-3 M reads were obtained for each sample.

1126 SITE-Seq assay analysis and selection for cellular validation

1127 SITE-Seq assay recovered off-targets were filtered for sites that had read-pileups proximal to
1128 the expected cut site, a PAM comprising at least one guanine base, fewer than 12 mismatches
1129 (reasoning that sites with 12 or more mismatches are likely spurious peaks not resulting from
1130 Cas9-induced double-strand breaks), and all sites with 11 mismatches were visually inspected
1131 and included in analysis if a putative deletion or insertion would result in a reduction of at least
1132 4 mismatches relative to the spacer sequence.

1133 *In silico* mismatch, deletion, and insertion prediction algorithm

1134 Predictive classification of SITE-Seq assay recovered off-target sites as pure mismatches,
1135 deletions, or insertions was executed using a scoring algorithm which consisted of the following
1136 sequential steps (**Figure S18**):

1137 (i) For each off-target, a gap library was generated where a single nucleotide gap was
1138 introduced between each nucleotide in the off-target sequence.

1139 (ii) The off-target gap library was then aligned to the spacer sequence and each
1140 alignment was scored based on the number of matched bases between the spacer
1141 and gapped off-target pair. If the gapped off-target with the highest alignment score
1142 improved alignment by at least 4 nucleotides relative to the non-gapped spacer–off-
1143 target alignment, the off-target sequence was marked as a single-nucleotide deletion
1144 and removed from subsequent analysis.

1145 (iii) The remaining pool of off-targets were then aligned to a spacer gapped library where
1146 a single nucleotide gap was introduced at each positing in the spacer.

1147 (iv) The spacer gap library was then aligned to each off-target sequence and each
1148 alignment was scored based on the number of matched bases between the off-target

1149 and the gapped spacer pair. If the gapped spacer with the highest alignment score
1150 improved alignment by at least 4 nucleotides relative to the non-gapped spacer–off-
1151 target alignment, the off-target sequence was marked as a single-nucleotide insertion
1152 and removed from subsequent analysis.
1153 (v) The remaining off-target for which the spacer–off-target alignment was not
1154 improved by single-nucleotide deletions or insertions were annotated as a
1155 mismatched off-target.
1156 *Generation of nucleoprotein complexes (RNP) for electroporation*
1157 Synthetic crRNA and tracrRNA oligonucleotides were ordered from Integrated DNA
1158 Technologies (Coralville, IA) as AltR^{TM} reagents. Ribonucleoprotein complexes (RNPs) were
1159 formulated by incubating 480 pmols of each crRNA and tracrRNA and heated to 95 °C for 2
1160 min then allowed to cool at room temperature for ~5 min. Annealed guides were then mixed
1161 with 160 pmols of Cas9 protein (1:3 molar ratio of Cas9 to guides) in RNP assembly buffer (20
1162 mM HEPES pH 7.4, 150 mM KCl, 10 mM $MgCl_2$, 5% glycerol) and incubated at 37 °C for 10
1163 min. RNP were then serial diluted two-fold by mixing with equal volume of RNP assembly
1164 buffer across seven concentrations down to 2.5 pmol of Cas9 and 7.5 pmols of crRNA-
1165 tracrRNA. RNPs were kept at 4 °C or on ice until transfection.
1166 *T cell handling and transfections*
1167 T cells were isolated from peripheral blood mononuclear cells (PBMCs) using the RoboSep-S
1168 cell isolation platform (STEMCELL Technologies) with EasySep Human T cell Isolation Kit
1169 (STEMCELL Technologies). Cells were then activated for 3 days in "Complete media"
1170 (Immunocult-XF T cell expansion media (STEMCELL Technologies), CTS Immune Cell
1171 serum replacement (Fisher Scientific), 1x antibiotic-antimycotic (Fisher Scientific), and
1172 recombinant human Interleukin-2 (rhIL-2, 100 units/mL)) in the presence of anti-CD3/CD28
1173 Dynabeads (Fisher Scientific) at a bead:cell ratio of 1:1. On the third day, beads were removed

1174 and cells expanded in Complete media for 24 hours. Cells were then harvested, washed once
1175 with phosphate-buffered saline (PBS) and resuspended at 10^7 cells/mL in P4 Primary Cell
1176 Nucleofector Solution (Lonza). 20 µL of resuspended cells (corresponding to ~ 200,000 cells)
1177 were then combined with 2.5 µL of RNP and electroporated using the Lonza 4D 96-well
1178 Nucleofector electroporation system using pulse code CA137. Transfected cells were then
1179 recovered in 180 µL/sample of Complete media. After 48 hours, the electroporated T cells were
1180 pelleted and gDNA was isolated by adding 50 µL/well QuickExtract DNA extraction solution
1181 (Epicentre), followed by incubation at 37 °C for 10 minutes, 65 °C for 6 minutes, and 95 °C for
1182 3 minutes. The isolated gDNA was diluted with 50 µL sterile water to achieve ~ 2,000 genome
1183 equivalents/µL and samples were stored at -20 °C until NGS analysis.

1184 *Targeted amplicon sequencing*

1185 Cas9 on- and off-target sites were amplified in a two-step PCR reaction. In brief, 3.75 µl
1186 (corresponding to ~7,500 cells) of lysate was used as a template for PCR amplification with Q5
1187 Hot-Start High Fidelity DNA Polymerase (NEB) and unique primer pairs containing an internal
1188 locus-specific region and an outer Illumina-compatible adapter sequence. A second PCR
1189 reaction targeting the outer-adapter sequence was performed to append unique indices to each
1190 amplicon. Sites were sequenced on a MiSeq with 2 × 151 paired-end reads and v2 chemistry,
1191 or NextSeq with 2 × 151 paired- end reads and v2 or v2.5 chemistry (Illumina), and aligned to
1192 the hg38 genomic assembly. For each site, indels were tallied if they occurred within 3
1193 nucleotides of the putative Cas9 cut site, and editing efficiencies were calculated by subtracting
1194 the percentage of indels in untransfected cells from the percentage of indels in RNP-transfected
1195 cells. Depth of coverage was ~5,000–20,000 reads per amplicon, and all samples with <500
1196 reads aligning to the predicted amplicon were discarded. Lower limit of detection (*e.g.,* ≥0.1%)
1197 was determined by titration of NIST genomic standards and assessment of expected versus
1198 measured values (Data not shown).

1199 Molecular Dynamics (MD) simulations

1200 MD simulations were performed using a well-established protocol, previously employed in
1201 several computational studies of CRISPR-Cas9 (Mitchell et al., 2020; Ricci et al., 2019). The
1202 Amber ff14SB force field (Maier et al., 2015) was adopted, including the ff99bsc1 corrections
1203 for DNA (Ivani et al., 2016) and the ff99bsc0+χOL3 corrections for RNA (Banas et al., 2010;
1204 Zgarbova et al., 2011). Explicit water molecules were described using the TIP3P model
1205 (Jorgensen et al., 1983), while the Li & Merz model was used for Mg2+ ions (Li and Merz,
1206 2014). An integration time step of 2 fs was applied. All bond lengths involving hydrogen atoms
1207 were constrained using the SHAKE algorithm. Temperature control (300 K) was performed via
1208 Langevin dynamics (Turq et al., 1977), with a collision frequency $\gamma = 1$. Pressure control was
1209 accomplished by coupling the system to a Berendsen barostat (Berendsen et al., 1984), at a
1210 reference pressure of 1 atm and with a relaxation time of 2 ps. The systems were subjected to
1211 energy minimization to relax water molecules and counter ions, keeping the protein, the RNA,
1212 DNA and Mg2+ ions fixed with harmonic position restraints of 300 kcal/mol · Å2. Then, the
1213 systems were heated up from 0 to 100 K in the canonical ensemble (NVT), by running two
1214 simulations of 5 ps each, imposing position restraints of 100 kcal/mol · Å2 on the above-
1215 mentioned elements of the system. The temperature was further increased up to 200 K in ~100
1216 ps of MD in the isothermal-isobaric ensemble (NPT), reducing the restraint to 25 kcal/mol ·
1217 Å2. Subsequently, all restraints were released, and the temperature of the systems was raised
1218 up to 300 K in a single NPT simulation of 500 ps. After ~ 1.1 ns of equilibration, ~10 ns of
1219 NPT runs were carried out allowing the density of the systems to stabilize around 1.01 g cm-3.
1220 Finally, production runs were carried out in the NVT ensemble, collecting ~500 ns in three
1221 replicates for each of the systems. These simulations have been performed using the GPU-
1222 empowered version of AMBER 20 (D.A. Case, 2021; Salomon-Ferrer et al., 2013). Analysis
1223 of the RNA:DNA conformational dynamics has been done using the CURVES+ code (Lavery

1224 et al., 2009). Molecular simulations were based on three X-ray structures of CRISPR-Cas9:
1225 TRAC on-target (PDB: 7OX8), AAVS1 on-target #2 and FANCF on-target, displaying 20 base
1226 pair long RNA:TS DNA hybrid and a cleaved NTS. To study the accommodation of rA-dA
1227 mismatches at various positions along the heteroduplex, we systematically mutated the base
1228 pairs to rA-dA mismatches (TRAC on-target for positions 1, 11, 15, and 20; FANCF on-target
1229 for positions 3-5, 9, and 18; AAVS1 on-target for positions 2, 6-8, 10, 12-14, 16, 17, and 19),
1230 using the LEaP tool in the AMBER 20 code (Lavery et al., 2009). The obtained model systems
1231 were embedded in explicit waters, and counterions were added to neutralize the total charge at
1232 physiological conditions, leading to periodic simulation cells of ~145*115*150 Å3 and a total
1233 of ~220,000 atoms for each system.

References

Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H.J., Rohs, R., and Mann, R.S. (2015). Deconvolving the recognition of DNA shape from sequence. Cell *161*, 307-318.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Afek, A., Shi, H., Rangadurai, A., Sahay, H., Senitzki, A., Xhani, S., Fang, M., Salinas, R., Mielko, Z., Pufall, M.A.*, et al.* (2020). DNA mismatches reveal conformational penalties in protein-DNA recognition. Nature.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol *38*, 824-844.

Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473-1475.

Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham, T.E., 3rd, Sponer, J., and Otyepka, M. (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. J Chem Theory Comput *6*, 3836-3849.

Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.v., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics *81*, 3684-3690.

Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. Proc Natl Acad Sci U S A *114*, 5461-5466.

Boyle, E.A., Becker, W.R., Bai, H.B., Chen, J.S., Doudna, J.A., and Greenleaf, W.J. (2021). Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. Sci Adv *7*.

Bravo, J.P.K., Liu, M.S., Hibshman, G.N., Dangerfield, T.L., Jung, K., McCool, R.S., Johnson, K.A., and Taylor, D.W. (2022). Structural basis for mismatch surveillance by CRISPR-Cas9. Nature *603*, 343-347.

Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., Gradia, S., Vidal, B., Garner, E., Slorach, E.M.*, et al.* (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nat Methods *14*, 600-606.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407-410.

Cofsky, J.C., Soczek, K.M., Knott, G.J., Nogales, E., and Doudna, J.A. (2022). CRISPR-Cas9 bends and twists DNA to read its sequence. Nat Struct Mol Biol *29*, 395-402.

Cromwell, C.R., Sung, K., Park, J., Krysler, A.R., Jovel, J., Kim, S.K., and Hubbard, B.P. (2018). Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. Nat Commun *9*, 1448.

D.A. Case, H.M.A., K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman (2021). Amber 2021. University of California, San Francisco.

Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., and Yildiz, A. (2017). A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv *3*, eaao0027.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol *190*, 1390-1400.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.

Donohoue, P.D., Pacesa, M., Lau, E., Vidal, B., Irby, M.J., Nyer, D.B., Rotstein, T., Banh, L., Toh, M.S., Gibson, J.*, et al.* (2021). Conformational control of Cas9 by CRISPR hybrid RNA-DNA guides mitigates off-target activity in T cells. Mol Cell *81*, 3637-3649 e3635.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Fu, B.X.H., Smith, J.D., Fuchs, R.T., Mabuchi, M., Curcuru, J., Robb, G.B., and Fire, A.Z. (2019). Target-dependent nickase activities of the CRISPR-Cas nucleases Cpf1 and Cas9. Nat Microbiol *4*, 888-897.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol *32*, 279-284.

Garg, A., and Heinemann, U. (2018). A novel form of RNA double helix based on G.U and C.A(+) wobble base pairing. RNA *24*, 209-218.

Gong, S., Yu, H.H., Johnson, K.A., and Taylor, D.W. (2018). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep *22*, 359-371.

Guo, M., Ren, K., Zhu, Y., Tang, Z., Wang, Y., Zhang, B., and Huang, Z. (2019). Structural insights into a high fidelity variant of SpCas9. Cell Res *29*, 183-192.

Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B.*, et al.* (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat Biotechnol *33*, 985-989.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Ivani, I., Dans, P.D., Noy, A., Perez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goni, R., Balaceanu, A.*, et al.* (2016). Parmbsc1: a refined force field for DNA simulations. Nat Methods *13*, 55-58.

Ivanov, I.E., Wright, A.V., Cofsky, J.C., Aris, K.D.P., Doudna, J.A., and Bryant, Z. (2020). Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. Proc Natl Acad Sci U S A *117*, 5853-5860.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477-1481.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jones, S.K., Jr., Hawkins, J.A., Johnson, N.V., Jung, C., Hu, K., Rybarski, J.R., Chen, J.S., Doudna, J.A., Press, W.H., and Finkelstein, I.J. (2021). Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. Nat Biotechnol *39*, 84-93.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics *79*, 926-935.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Kimsey, I.J., Petzold, K., Sathyamoorthy, B., Stein, Z.W., and Al-Hashimi, H.M. (2015). Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. Nature *519*, 315-320.

Kimsey, I.J., Szymanski, E.S., Zahurancik, W.J., Shakya, A., Xue, Y., Chu, C.C., Sathyamoorthy, B., Suo, Z., and Al-Hashimi, H.M. (2018). Dynamic basis for dG*dT misincorporation via tautomerization and ionization. Nature *554*, 195-201.
Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., and Shakked, Z. (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. Nat Struct Mol Biol *17*, 423-429.
Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J Mol Biol *372*, 774-797.
Kulcsar, P.I., Talas, A., Toth, E., Nyeste, A., Ligeti, Z., Welker, Z., and Welker, E. (2020). Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5'G-extended sgRNAs. Nat Commun *11*, 1223.
Kunkel, T.A., and Bebenek, K. (2000). DNA replication fidelity. Annu Rev Biochem *69*, 497-529.
Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol *32*, 677-683.
Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D., and Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res *37*, 5917-5929.
Lazzarotto, C.R., Malinin, N.L., Li, Y., Zhang, R., Yang, Y., Lee, G., Cowley, E., He, Y., Lan, X., Jividen, K.*, et al.* (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. Nat Biotechnol.
Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res *30*, 3497-3531.
Li, P., and Merz, K.M. (2014). Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. Journal of Chemical Theory and Computation *10*, 289-297.
Li, S., Olson, W.K., and Lu, X.J. (2019). Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. Nucleic Acids Res *47*, W26-W34.
Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res *42*, 7473-7485.
Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput *11*, 3696-3713.
Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P.*, et al.* (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol *18*, 67-83.
Mekler, V., Minakhin, L., and Severinov, K. (2017). Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. Proc Natl Acad Sci U S A *114*, 5443-5448.
Mitchell, B.P., Hsu, R.V., Medrano, M.A., Zewde, N.T., Narkhede, Y.B., and Palermo, G. (2020). Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9. Frontiers in Molecular Biosciences *7*.
Mullally, G., van Aelst, K., Naqvi, M.M., Diffin, F.M., Karvelis, T., Gasiunas, G., Siksnys, V., and Szczelkun, M.D. (2020). 5' modifications to CRISPR-Cas9 gRNA can change the dynamics and size of R-loops and inhibit DNA cleavage. Nucleic Acids Res.
Murugan, K., Seetharam, A.S., Severin, A.J., and Sashital, D.G. (2020). CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. J Biol Chem.
Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.
O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature *516*, 263-266.
O'Geen, H., Henry, I.M., Bhakta, M.S., Meckler, J.F., and Segal, D.J. (2015). A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res *43*, 3389-3404.
O'Reilly, D., Kartje, Z.J., Ageely, E.A., Malek-Adamian, E., Habibian, M., Schofield, A., Barkau, C.L., Rohilla, K.J., DeRossett, L.B., Weigle, A.T.*, et al.* (2019). Extensive CRISPR RNA modification reveals

chemical compatibility and structure-activity relationships for Cas9 biochemical activity. Nucleic Acids Res *47*, 546-558.
Okafor, I.C., Singh, D., Wang, Y., Jung, M., Wang, H., Mallon, J., Bailey, S., Lee, J.K., and Ha, T. (2019). Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. Nucleic Acids Res *47*, 11880-11888.
Pacesa, M., and Jinek, M. (2021). Mechanism of R-loop formation and conformational activation of Cas9. bioRxiv, 2021.2009.2016.460614.
Pacesa, M., Loeff, L., Querques, I., Muckenfuss, L.M., Sawicka, M., and Jinek, M. (2022). Mechanism of R-loop formation and conformational activation of Cas9. bioRxiv, 2021.2009.2016.460614.
Palermo, G., Chen, J.S., Ricci, C.G., Rivalta, I., Jinek, M., Batista, V.S., Doudna, J.A., and McCammon, J.A. (2018). Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. Q Rev Biophys *51*.
Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol *31*, 839-843.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci *30*, 70-82.
Ricci, C.G., Chen, J.S., Miao, Y., Jinek, M., Doudna, J.A., McCammon, J.A., and Palermo, G. (2019). Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. ACS Cent Sci *5*, 651-662.
Rodnina, M.V., and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. Annu Rev Biochem *70*, 415-435.
Rohs, R., West, S.M., Liu, P., and Honig, B. (2009a). Nuance in the double-helix and its role in protein-DNA recognition. Curr Opin Struct Biol *19*, 171-177.
Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009b). The role of DNA shape in protein-DNA recognition. Nature *461*, 1248-1253.
Salomon-Ferrer, R., Gotz, A.W., Poole, D., Le Grand, S., and Walker, R.C. (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput *9*, 3878-3888.
Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A *108*, 10098-10103.
Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat Commun *7*, 12778.
Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633.
Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.
Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.
Timsit, Y. (1999). DNA structure and polymerase fidelity. J Mol Biol *293*, 835-853.
Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods *14*, 607-614.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., *et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.
Turq, P., Lantelme, F., and Friedman, H.L. (1977). Brownian dynamics: Its application to ionic solutions. The Journal of Chemical Physics *66*, 3039-3044.
van Houte, S., Ekroth, A.K., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., *et al.* (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385-388.

1446 Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C.,
1447 Jinek, M., van der Oost, J.*, et al.* (2018). Bacteriophage DNA glucosylation impairs target DNA binding
1448 by type I and II but not by type V CRISPR-Cas effector complexes. Nucleic Acids Res *46*, 873-885.
1449 Wang, W., Hellinga, H.W., and Beese, L.S. (2011). Structural evidence for the rare tautomer hypothesis
1450 of spontaneous mutagenesis. Proc Natl Acad Sci U S A *108*, 17644-17648.
1451 Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E.,
1452 Konermann, S., Chen, S.*, et al.* (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in
1453 mammalian cells. Nat Biotechnol *32*, 670-676.
1454 Yakovchuk, P., Protozanova, E., and Frank-Kamenetskii, M.D. (2006). Base-stacking and base-pairing
1455 contributions into thermal stability of the DNA double helix. Nucleic Acids Res *34*, 564-574.
1456 Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., and Chen, C. (2018). The Conformational Dynamics of
1457 Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. Cell Rep *22*, 372-382.
1458 Yaung, S.J., Esvelt, K.M., and Church, G.M. (2014). CRISPR/Cas9-mediated phage resistance is not
1459 impeded by the DNA modifications of phage T4. PLoS One *9*, e98811.
1460 Yin, H., Song, C.Q., Suresh, S., Kwan, S.Y., Wu, Q., Walsh, S., Ding, J., Bogorad, R.L., Zhu, L.J.,
1461 Wolfe, S.A.*, et al.* (2018). Partial DNA-guided Cas9 enables genome editing with reduced off-target
1462 activity. Nat Chem Biol *14*, 311-316.
1463 Yin, H., Song, C.Q., Suresh, S., Wu, Q., Walsh, S., Rhym, L.H., Mintzer, E., Bolukbasi, M.F., Zhu, L.J.,
1464 Kauffman, K.*, et al.* (2017). Structure-guided chemical modification of guide RNA enables potent non-
1465 viral in vivo genome editing. Nat Biotechnol *35*, 1179-1187.
1466 Zeng, Y., Cui, Y., Zhang, Y., Zhang, Y., Liang, M., Chen, H., Lan, J., Song, G., and Lou, J. (2018). The
1467 initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. Nucleic Acids Res *46*, 350-
1468 361.
1469 Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E., 3rd, and Jurecka, P.
1470 (2011). Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum
1471 Chemical Calculations of Glycosidic Torsion Profiles. J Chem Theory Comput *7*, 2886-2902.
1472 Zhang, L., Rube, H.T., Vakulskas, C.A., Behlke, M.A., Bussemaker, H.J., and Pufall, M.A. (2020).
1473 Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. Nucleic
1474 Acids Research.
1475 Zhu, X., Clarke, R., Puppala, A.K., Chittori, S., Merk, A., Merrill, B.J., Simonovic, M., and
1476 Subramaniam, S. (2019). Cryo-EM structures reveal coordinated domain motions that govern DNA
1477 cleavage by Cas9. Nat Struct Mol Biol *26*, 679-685.

1478

Figure 1

A

Gene	sequence	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	P	A	M
AAVS1	guide RNA	G	G	G	G	G	C	C	A	C	U	A	G	G	G	A	C	A	G	G	A	U			
AAVS1	on-target		C	C	C	C	G	G	T	G	A	T	C	C	C	T	G	T	C	C	T	A	A	C	C
AAVS1	off-target1		·	·	·	T	·	·	·	A	G	·	·	·	·	·	·	·	·	·	·	·	C	C	C
AAVS1	off-target2		A	·	·	·	C	·	G	T	G	·	·	·	·	·	·	·	·	·	A	·	A	C	C
AAVS1	off-target3		·	·	·	·	·	·	·	C	·	·	·	·	·	·	C	·	·	·	·	·	T	C	C
AAVS1	off-target4		A	·	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	T	A	C	C
AAVS1	off-target5		·	·	·	T	·	·	·	·	·	C	·	·	·	·	·	·	·	·	G	G	A	C	C
FANCF	guide RNA	G	G	G	A	A	U	C	C	C	U	U	C	U	G	C	A	G	C	A	C	C			
FANCF	on-target		C	C	T	T	A	G	G	G	A	A	G	A	C	G	T	C	G	T	G	G	A	C	C
FANCF	off-target1		·	·	·	·	G	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	T	C	C
FANCF	off-target2		·	·	·	C	·	·	·	·	·	G	·	·	T	·	·	·	·	·	·	·	T	C	C
FANCF	off-target3	G	·	·	·	C	G	·	·	·	·	·	·	·	·	·	·	·	—	·	·	·	G	C	C
FANCF	off-target4		T	G	G	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	A	C	C
FANCF	off-target5		A	·	·	·	·	·	·	·	T	·	·	·	G	·	·	·	·	·	·	·	A	C	C
FANCF	off-target6		·	·	·	·	G	T	·	·	·	·	·	·	·	·	·	·	·	A	·	·	T	C	C
FANCF	off-target7		·	·	·	·	·	·	T	·	·	·	A	·	T	·	·	·	·	·	·	·	T	C	C
PTPRC-tgt2	guide RNA	G	G	C	A	A	A	A	C	U	C	A	A	C	C	C	U	A	C	C	C	C			
PTPRC-tgt2	on-target		C	G	T	T	T	T	G	A	G	T	T	G	G	G	A	T	G	G	G	G	A	C	C
PTPRC-tgt2	off-target1	T	·	·	·	·	A	·	·	T	·	·	·	·	·	·	—	·	·	·	·	·	A	C	C
TRAC	guide RNA	G	A	G	A	G	U	C	U	C	U	C	A	G	C	U	G	G	U	A	C	A			
TRAC	on-target		T	C	T	C	A	G	A	G	A	G	T	C	G	A	C	C	A	T	G	T	G	C	C
TRAC	off-target1		C	T	·	·	·	·	·	·	G	·	·	·	·	·	T	·	·	·	·	·	A	C	C
TRAC	off-target2		·	T	·	·	·	·	T	·	T	·	·	·	·	·	·	·	·	·	·	·	A	C	C

B

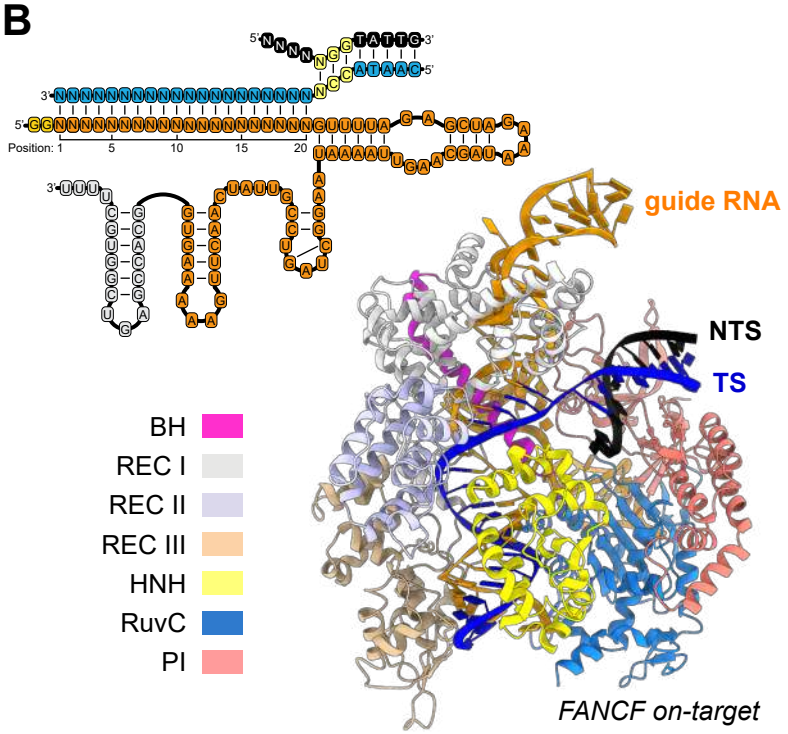


Figure 2

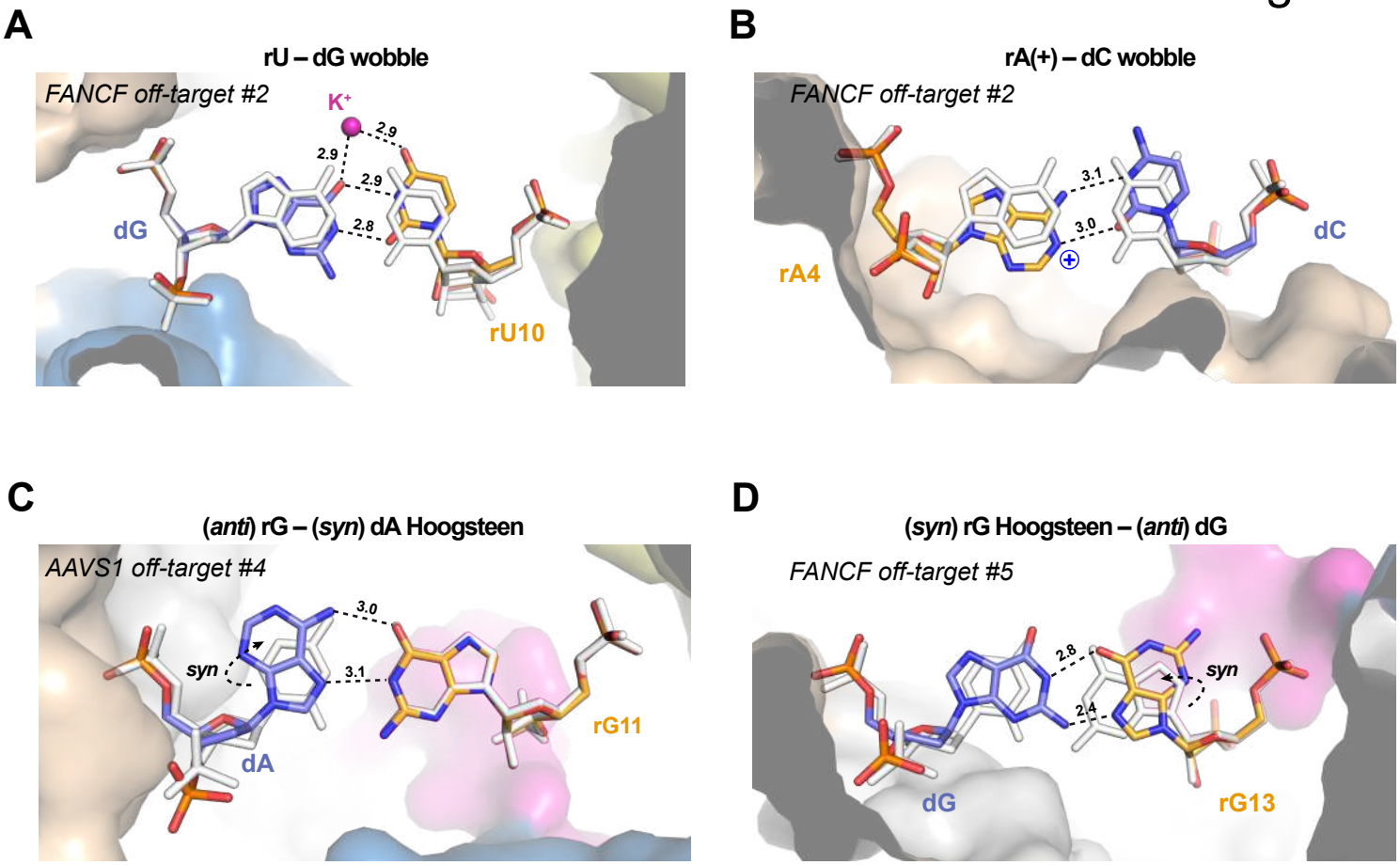

A
rU – dG wobble
FANCF off-target #2
K+
2.9
2.9
2.9
2.8
dG
rU10
B
rA(+) – dC wobble
FANCF off-target #2
3.1
3.0
rA4
dC
C
(anti) rG – (syn) dA Hoogsteen
AAVS1 off-target #4
3.0
3.1
syn
dA
rG11
D
(syn) rG Hoogsteen – (anti) dG
FANCF off-target #5
2.8
2.4
syn
dG
rG13

Figure 3

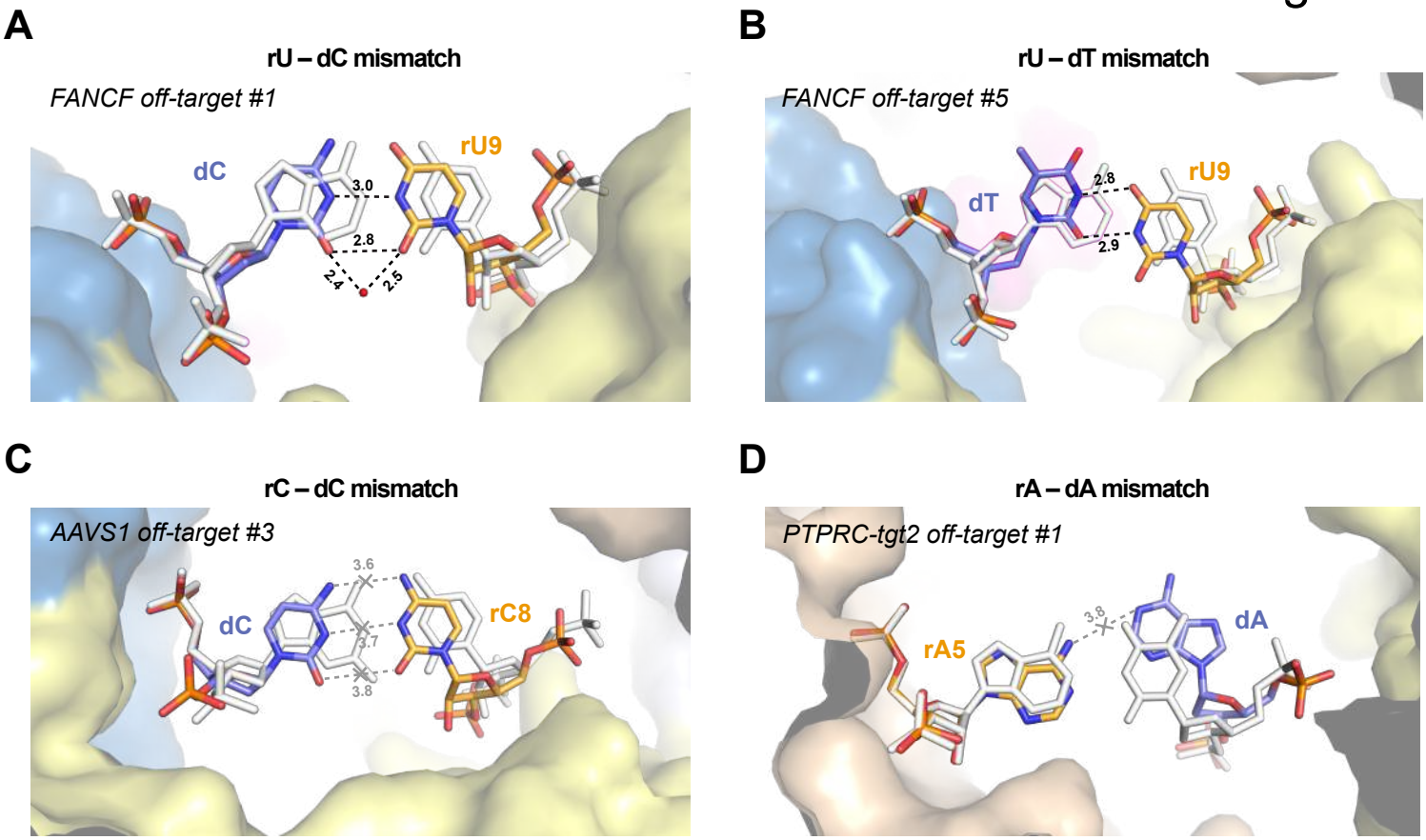

A
rU – dC mismatch
FANCF off-target #1
dC
rU9
3.0
2.8
2.4
2.5
B
rU – dT mismatch
FANCF off-target #5
dT
rU9
2.8
2.9
C
rC – dC mismatch
AAVS1 off-target #3
dC
rC8
3.6
3.7
3.8
D
rA – dA mismatch
PTPRC-tgt2 off-target #1
rA5
dA
3.8

Figure 4

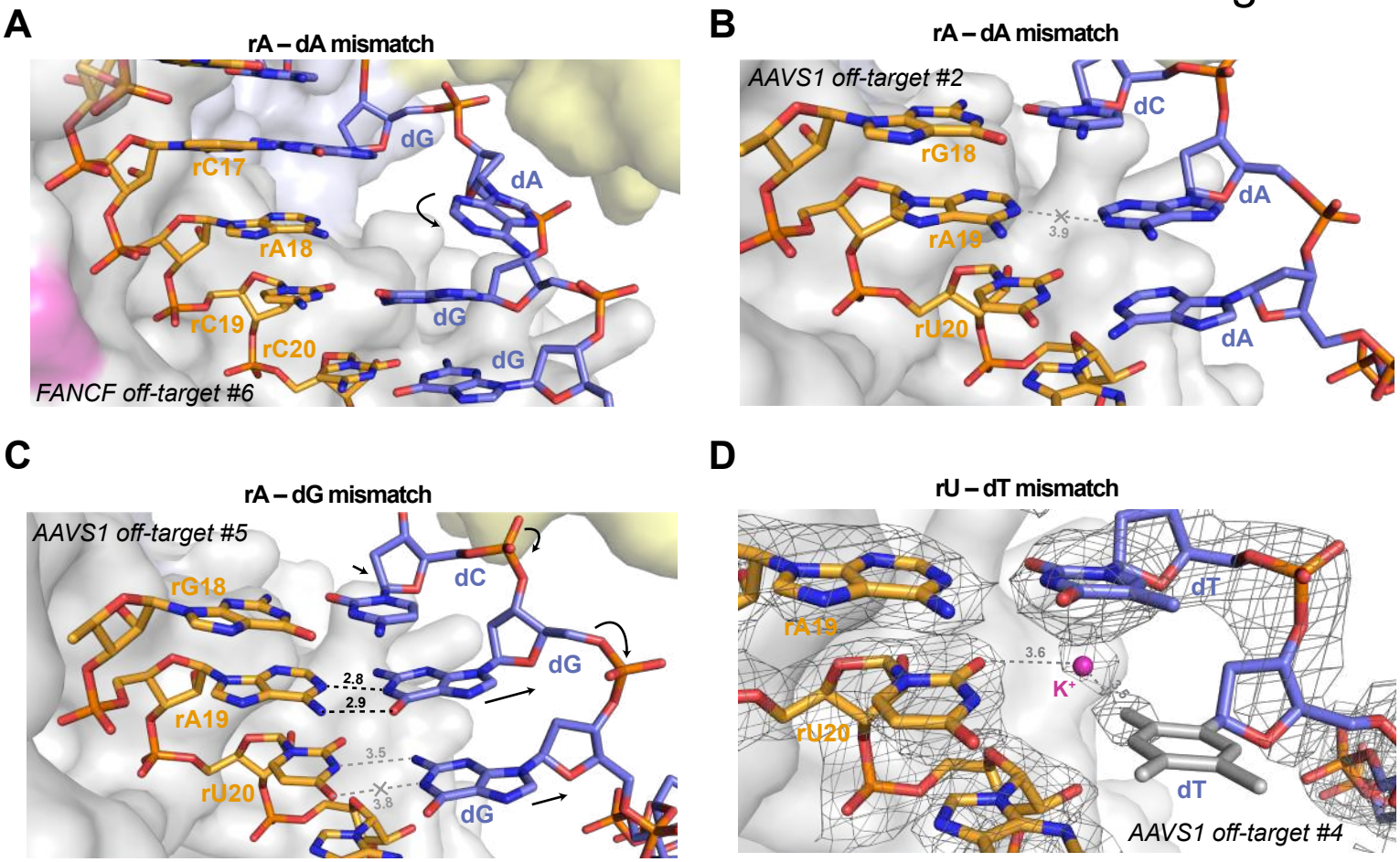
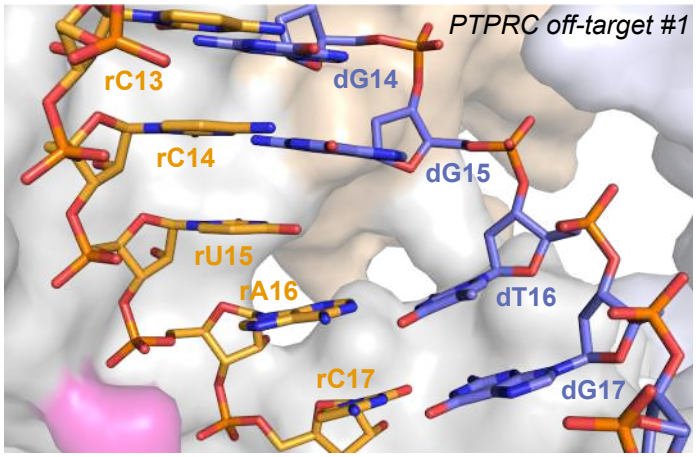
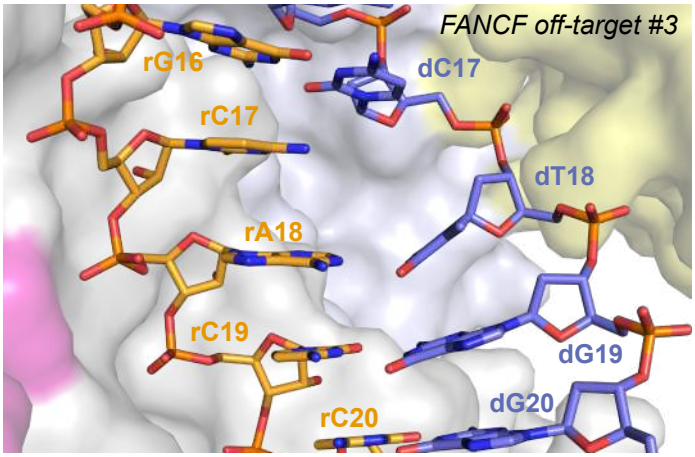

A
rA – dA mismatch
dG
rC17
dA
rA18
rC19
dG
rC20
dG
FANCF off-target #6
B
rA – dA mismatch
AAVS1 off-target #2
dC
rG18
dA
rA19
3.9
rU20
dA
C
rA – dG mismatch
AAVS1 off-target #5
rG18
dC
dG
2.8
2.9
rA19
3.5
rU20
3.8
dG
D
rU – dT mismatch
dT
rA19
3.6
K+
rU20
dT
AAVS1 off-target #4

Figure 5

A



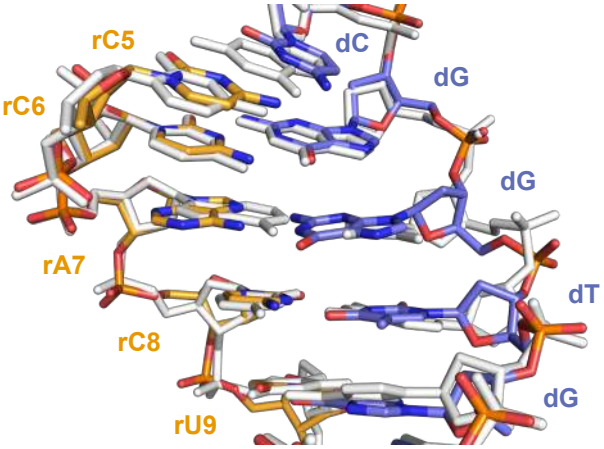
B



C

Gene	sequence	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	P	A	M
AAVS1	guide RNA	G	G	G	G	G	C	C	A	C	U	A	G	G	G	A	C	A	G	G	A	U			
AAVS1	on-target		C	C	C	C	G	G	T	G	A	T	C	C	C	T	G	T	C	C	T	A	A	C	C
AAVS1	off-target2-A		A	·	·	·	C	·	G	T	G	·	·	·	·	·	·	·	·	·	A	·	A	C	C
AAVS1	off-target2-B	A	·	·	·	·	·	·	·	·	—	·	·	·	·	·	·	·	·	·	A	·	A	C	C
AAVS1	off-target2-rev		A	·	·	·	C	·	G	·	G	·	·	·	·	·	·	·	·	·	A	·	A	C	C

D



E

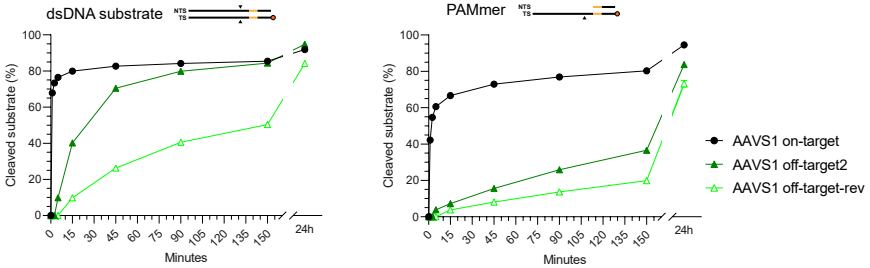
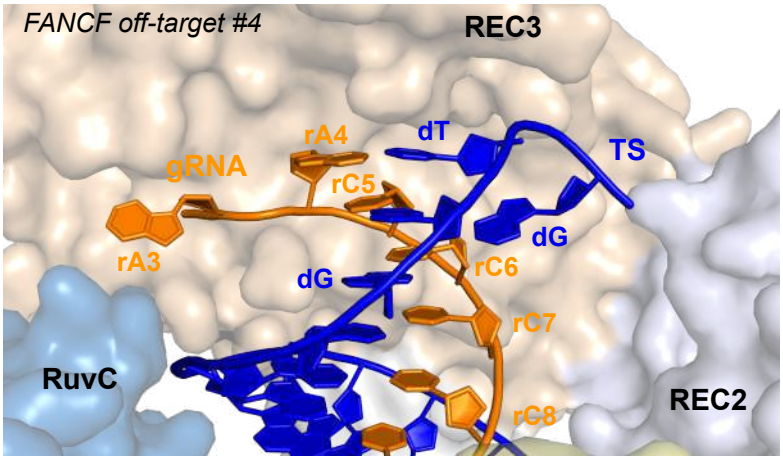


Figure 6

A


FANCF off-target #4
REC3
rA4
dT
TS
gRNA
rC5
dG
rA3
rC6
dG
rC7
RuvC
REC2
rC8

B

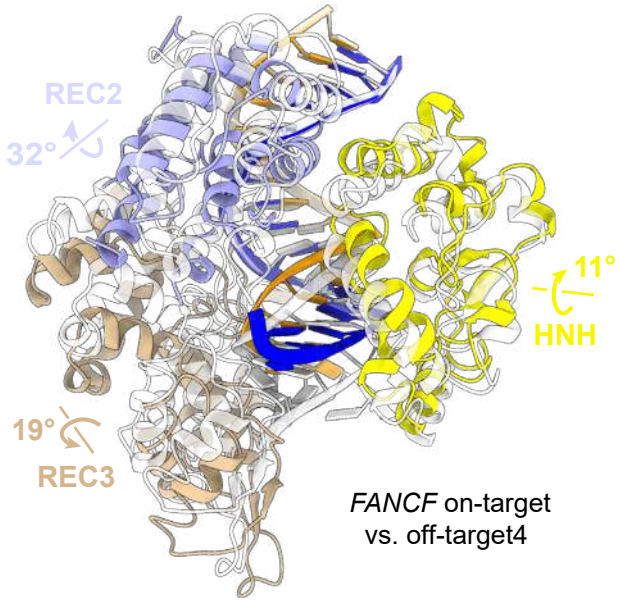

REC2
32°
11°
HNH
19°
REC3
FANCF on-target
vs. off-target4

Figure S1

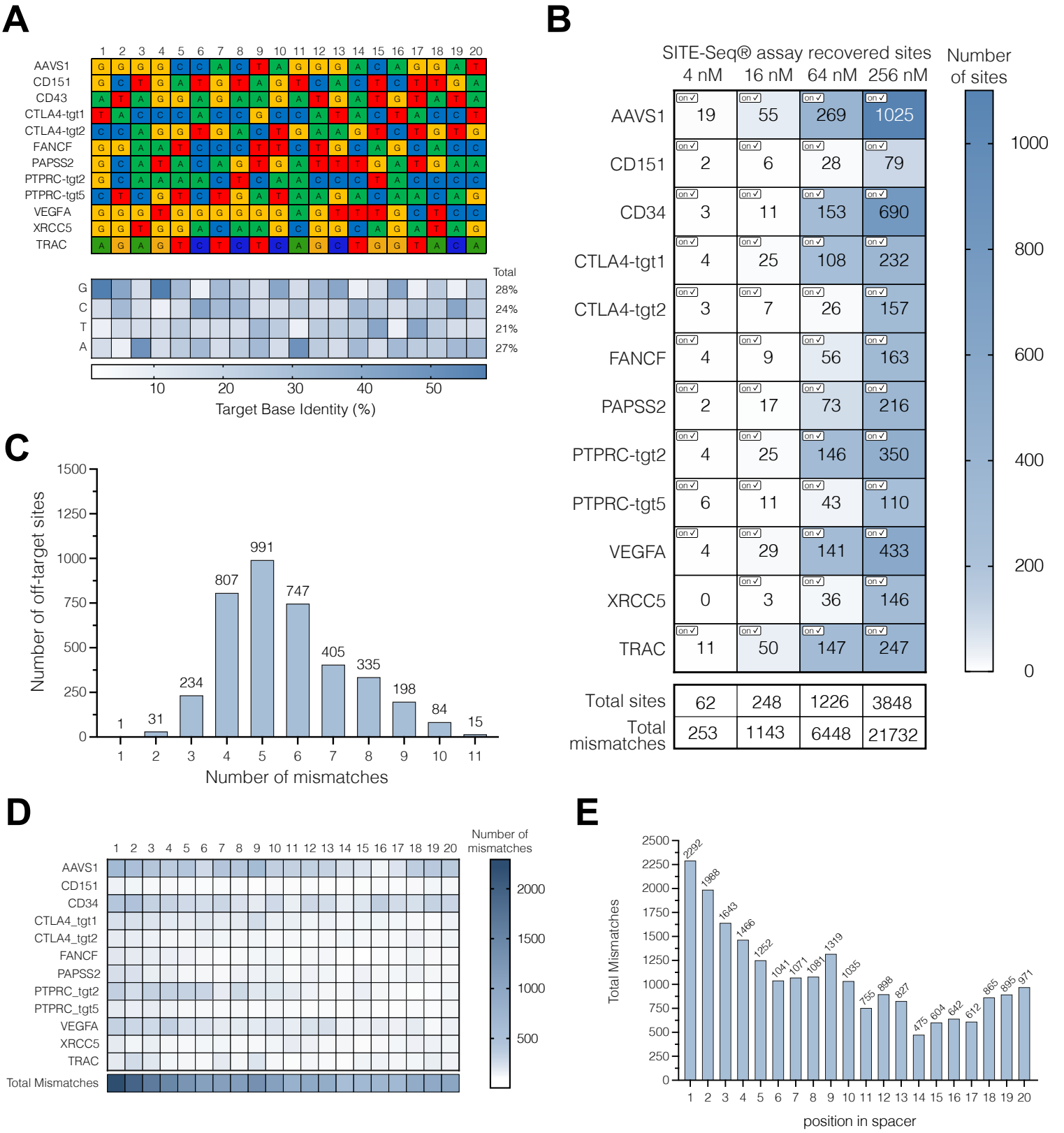

A
AAVS1
CD151
CD43
CTLA4-tgt1
CTLA4-tgt2
FANCF
PAPSS2
PTPRC-tgt2
PTPRC-tgt5
VEGFA
XRCC5
TRAC
Total
28%
24%
21%
27%
Target Base Identity (%)
B
SITE-Seq® assay recovered sites
4 nM 16 nM 64 nM 256 nM
Number of sites
AAVS1 19 55 269 1025
CD151 2 6 28 79
CD34 3 11 153 690
CTLA4-tgt1 4 25 108 232
CTLA4-tgt2 3 7 26 157
FANCF 4 9 56 163
PAPSS2 2 17 73 216
PTPRC-tgt2 4 25 146 350
PTPRC-tgt5 6 11 43 110
VEGFA 4 29 141 433
XRCC5 0 3 36 146
TRAC 11 50 147 247
Total sites 62 248 1226 3848
Total mismatches 253 1143 6448 21732
C
Number of off-target sites
Number of mismatches
1 31 234 807 991 747 405 335 198 84 15
D
Number of mismatches
Total Mismatches
E
Total Mismatches
position in spacer
2292 1988 1643 1466 1252 1041 1071 1081 1319 1035 755 898 827 475 604 642 612 865 895 971

Figure S2

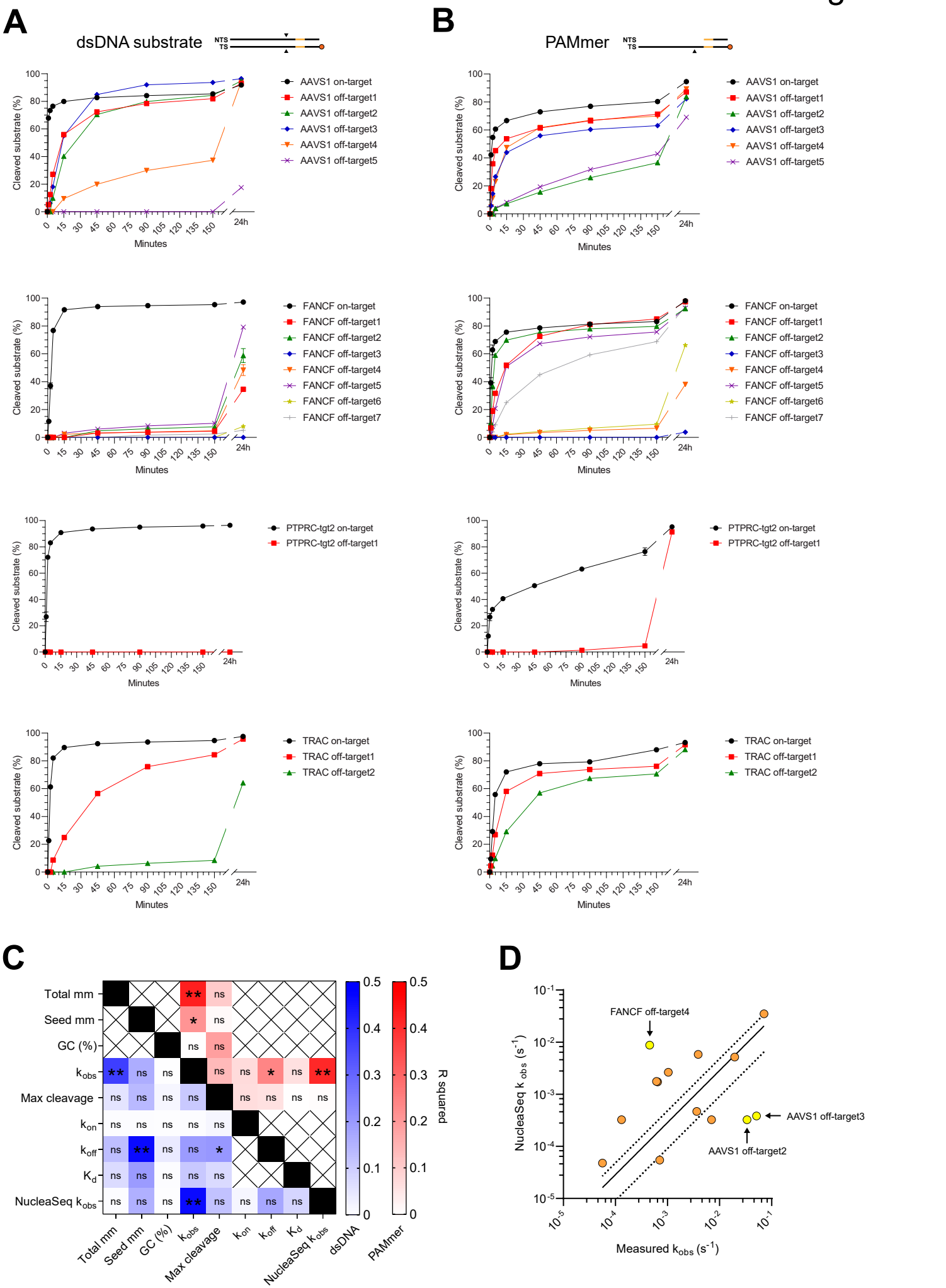
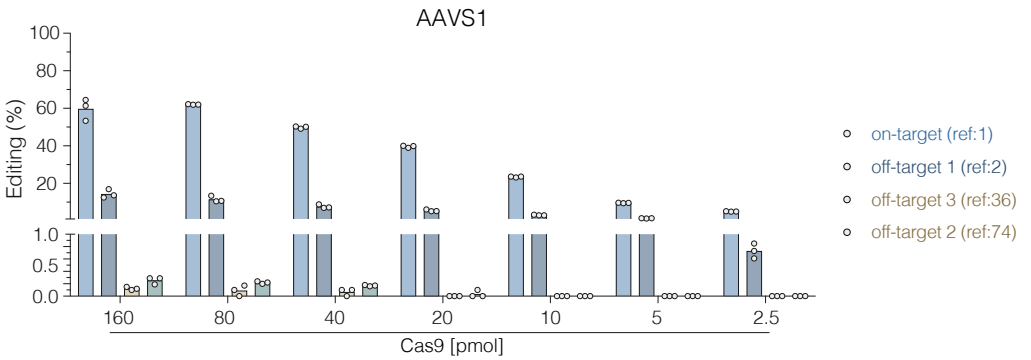

A
dsDNA substrate
NTS
TS
B
PAMmer
Cleaved substrate (%)
Minutes
24h
AAVS1 on-target
AAVS1 off-target1
AAVS1 off-target2
AAVS1 off-target3
AAVS1 off-target4
AAVS1 off-target5
FANCF on-target
FANCF off-target1
FANCF off-target2
FANCF off-target3
FANCF off-target4
FANCF off-target5
FANCF off-target6
FANCF off-target7
PTPRC-tgt2 on-target
PTPRC-tgt2 off-target1
TRAC on-target
TRAC off-target1
TRAC off-target2
C
Total mm
Seed mm
GC (%)
kobs
Max cleavage
kon
koff
Kd
NucleaSeq kobs
dsDNA
PAMmer
R squared
D
FANCF off-target4
AAVS1 off-target3
AAVS1 off-target2
NucleaSeq kobs (s-1)
Measured kobs (s-1)

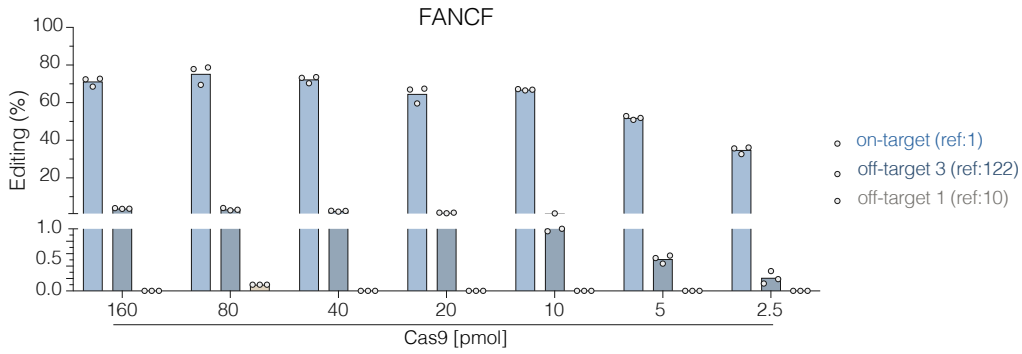
Figure S3

A

Gene	Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Editing (%)
AAVS1	guide RNA	G	G	G	G	C	C	A	C	T	A	G	G	G	A	C	A	G	G	A	T	
AAVS1	on-target	C	C	C	C	G	G	T	G	A	T	C	C	C	T	G	T	C	C	T	A	59.7
AAVS1	off-target 1	.	.	.	T	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	14.3
AAVS1	off-target 2	A	.	.	.	C	.	G	T	G	.	.	.	.	.	.	.	.	.	A	.	0.3
AAVS1	off-target 3	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	0.1
AAVS1	off-target 4	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	<0.1
AAVS1	off-target 5	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	G	<0.1
FANCF	guide RNA	G	G	A	A	T	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	Editing (%)
FANCF	on-target	C	C	T	T	A	G	G	G	A	A	G	A	C	G	T	C	G	T	G	G	71.2
FANCF	off-target 1	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	<0.1
FANCF	off-target 2	.	.	.	C	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	<0.1
FANCF	off-target 3	.	.	.	*	G	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	3.7
FANCF	off-target 4	T	G	G	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	<0.1
FANCF	off-target 5	A	.	.	.	.	.	.	.	T	.	.	.	G	.	.	.	.	.	.	.	<0.1
FANCF	off-target 6	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	<0.1
FANCF	off-target 7	.	.	.	.	.	.	T	.	.	.	A	.	T	.	.	.	.	.	.	.	<0.1
PTPRC -tgt2	guide RNA	C	G	T	T	T	T	G	A	G	T	T	G	G	G	A	T	G	G	G	G	Editing (%)
PTPRC -tgt2	on-target	G	C	A	A	A	A	C	T	C	A	A	C	C	C	T	A	C	C	C	C	84.0
PTPRC -tgt2	off-target 1	.	.	.	.	A	.	.	T	.	.	.	.	.	.	-	.	.	.	.	.	0.0
TRAC	guide RNA	A	G	A	G	T	C	T	C	T	C	A	G	C	T	G	G	T	A	C	A	Editing (%)
TRAC	on-target	T	C	T	C	A	G	A	G	A	G	T	C	G	A	C	C	A	T	G	T	77.1
TRAC	off-target 1	C	T	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	0.8
TRAC	off-target 2	.	T	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	0.4

B





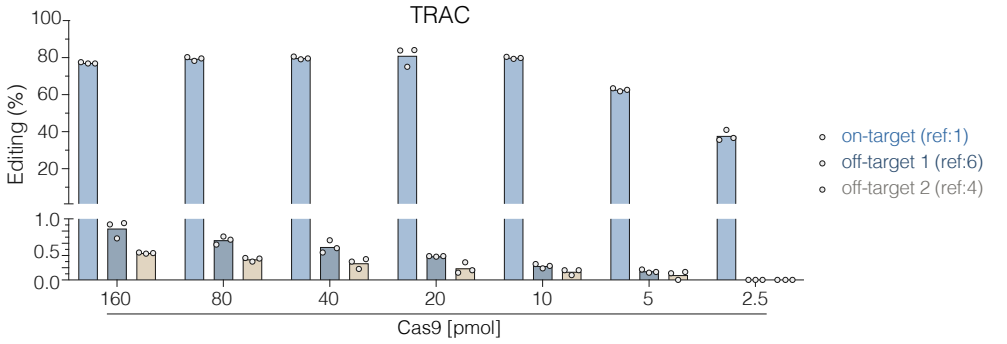
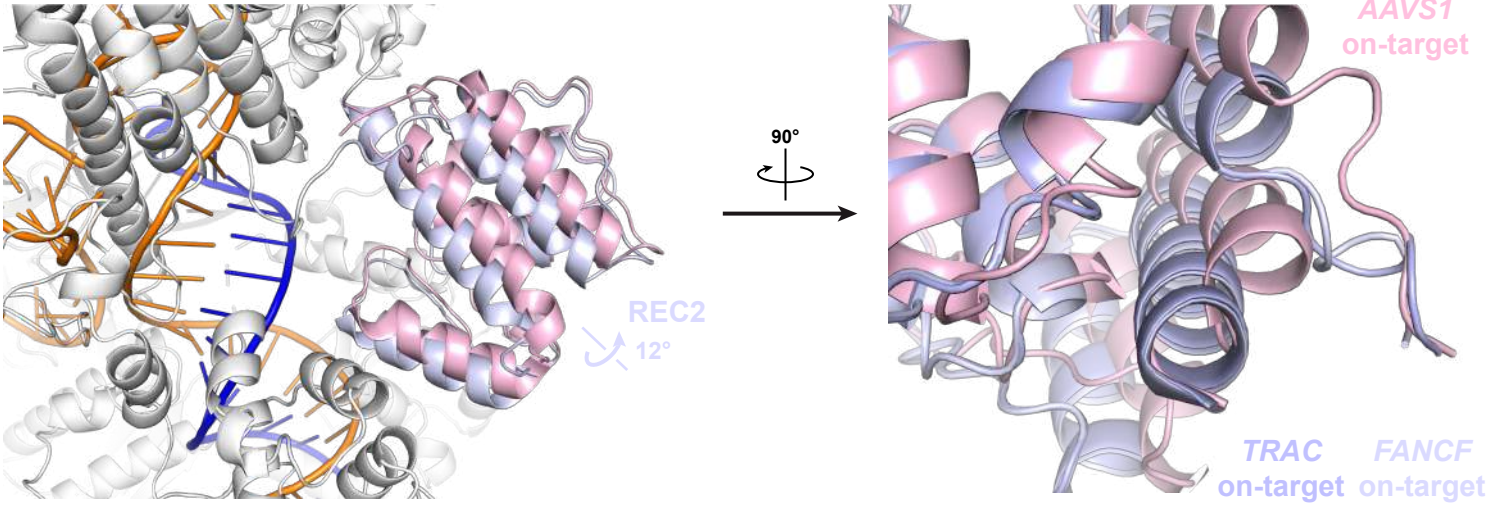


Figure S4

A


REC2
12°
90°
AAVS1
on-target
TRAC
on-target
FANCF
on-target

B

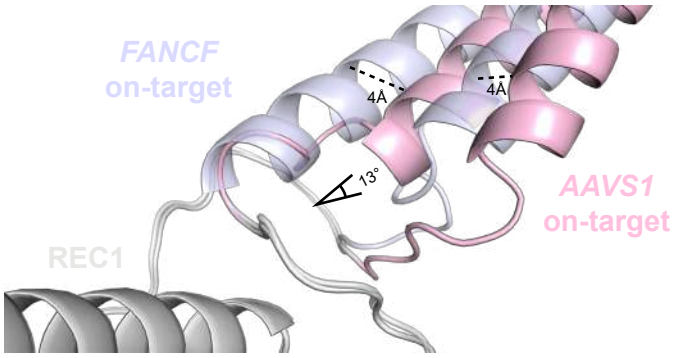

FANCF
on-target
4Å
4Å
13°
AAVS1
on-target
REC1

Figure S5

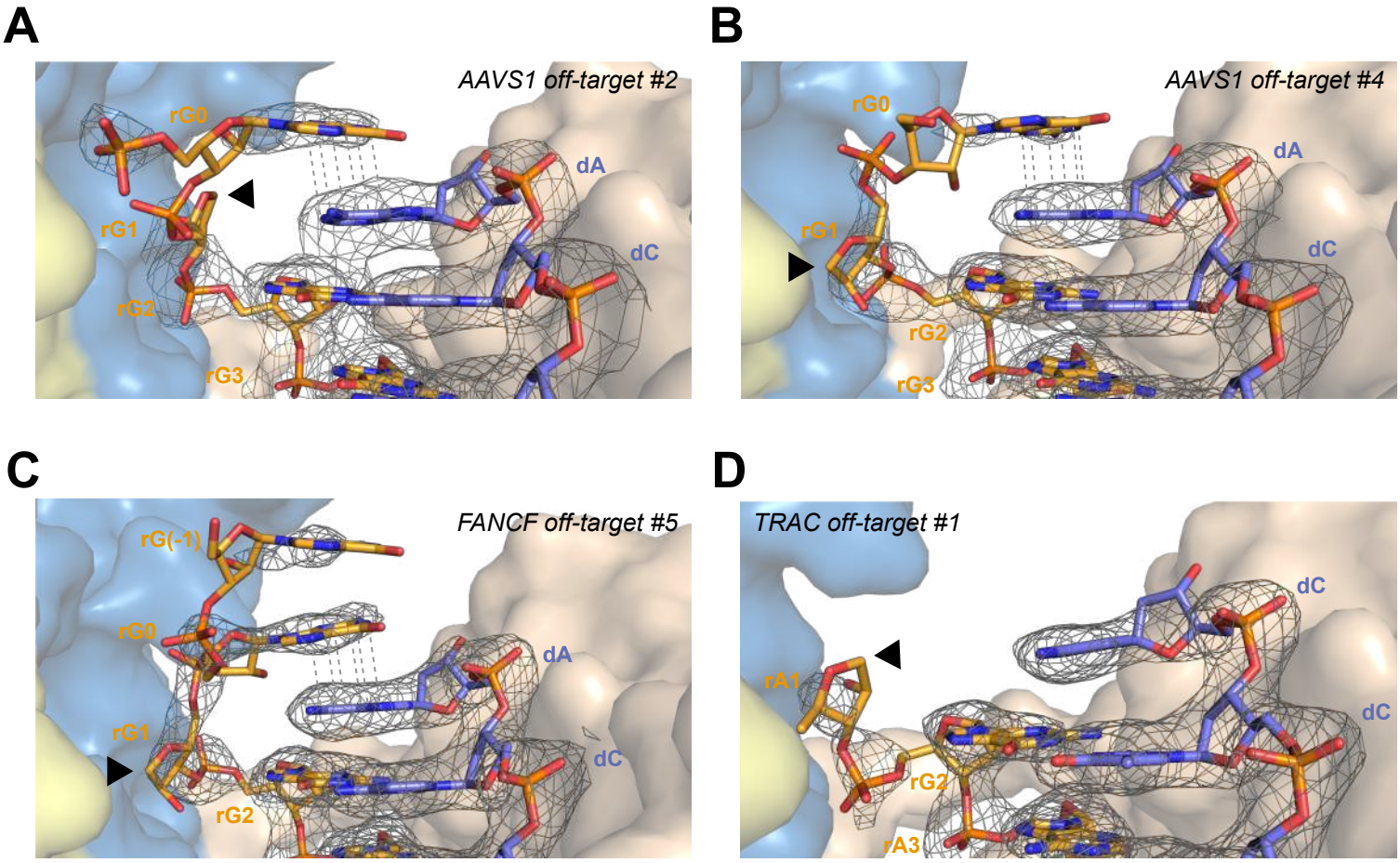

A
AAVS1 off-target #2
rG0
rG1
rG2
rG3
dA
dC
B
AAVS1 off-target #4
rG0
rG1
rG2
rG3
dA
dC
C
FANCF off-target #5
rG(-1)
rG0
rG1
rG2
dA
dC
D
TRAC off-target #1
rA1
rG2
rA3
dC
dC

Figure S6

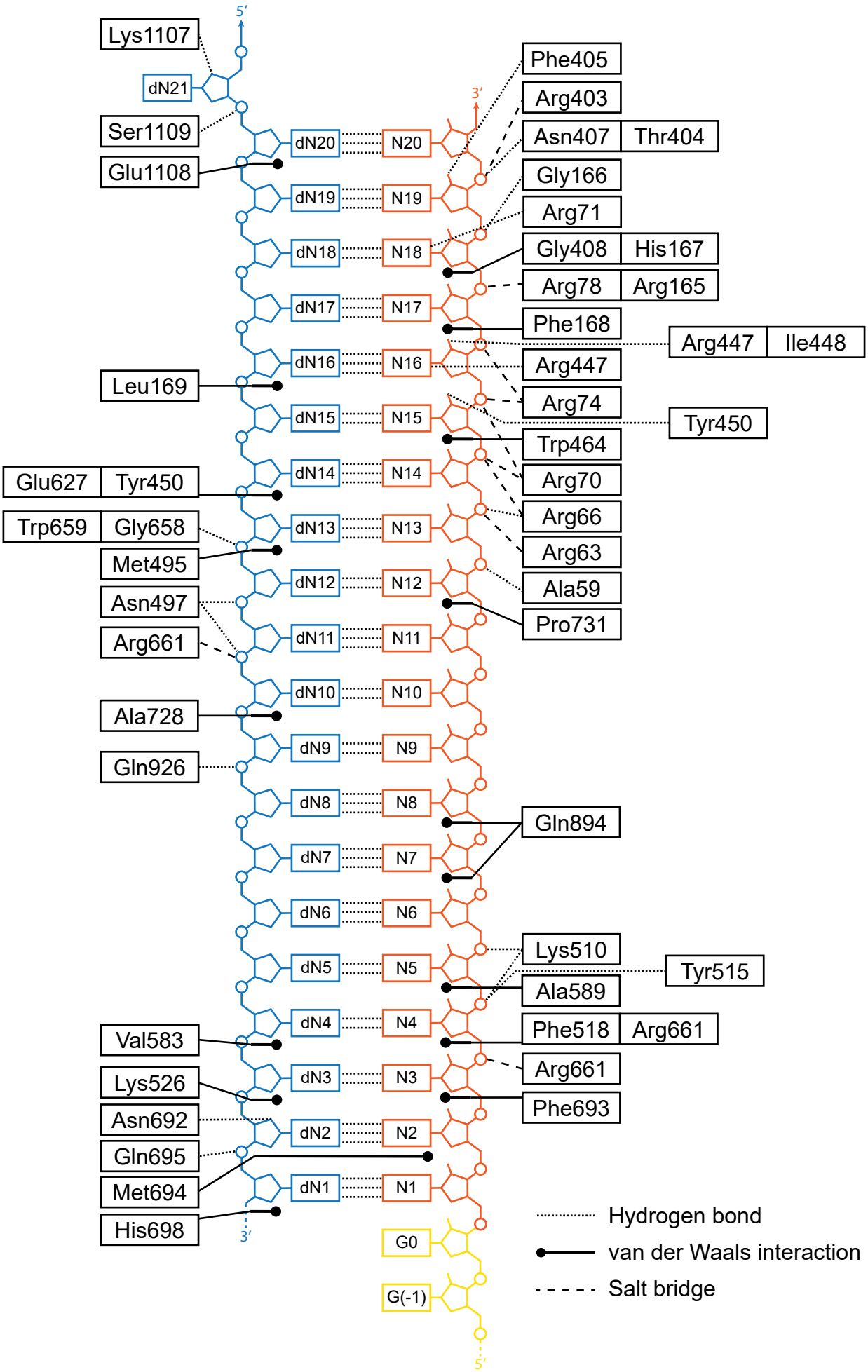

5′
3′
Lys1107
dN21
Ser1109
Glu1108
dN20
N20
dN19
N19
dN18
N18
dN17
N17
dN16
N16
Leu169
dN15
N15
dN14
N14
Glu627
Tyr450
Trp659
Gly658
dN13
N13
Met495
dN12
N12
Asn497
dN11
N11
Arg661
dN10
N10
Ala728
dN9
N9
Gln926
dN8
N8
dN7
N7
dN6
N6
dN5
N5
dN4
N4
Val583
dN3
N3
Lys526
Asn692
dN2
N2
Gln695
Met694
dN1
N1
His698
G0
G(-1)
Phe405
Arg403
Asn407
Thr404
Gly166
Arg71
Gly408
His167
Arg78
Arg165
Phe168
Arg447
Ile448
Arg447
Arg74
Tyr450
Trp464
Arg70
Arg66
Arg63
Ala59
Pro731
Gln894
Lys510
Tyr515
Ala589
Phe518
Arg661
Arg661
Phe693
Hydrogen bond
van der Waals interaction
Salt bridge

Figure S7

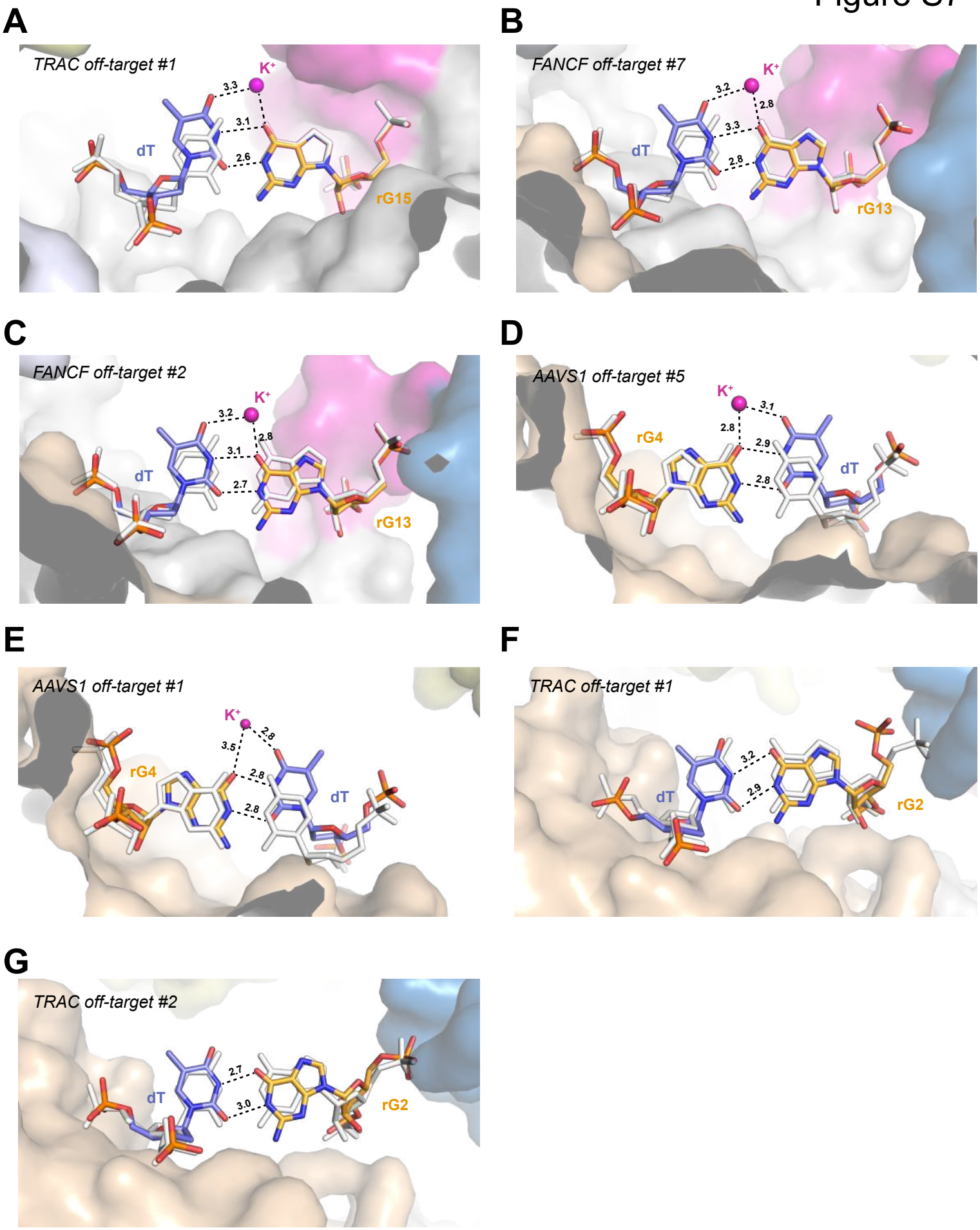
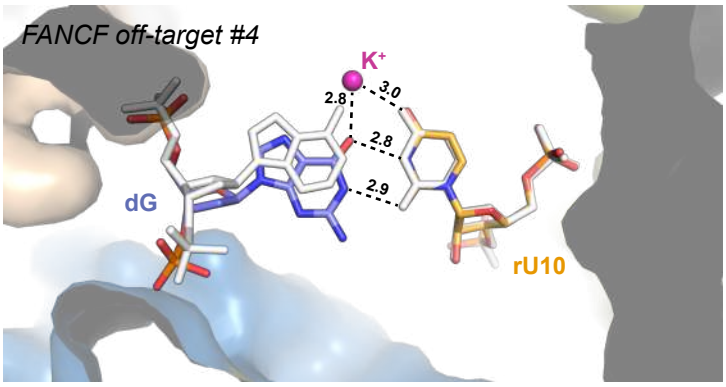
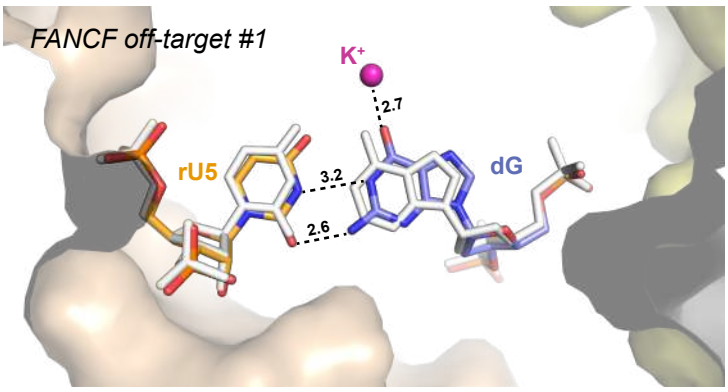

A
TRAC off-target #1
K+
3.3
3.1
2.6
dT
rG15
B
FANCF off-target #7
K+
3.2
2.8
3.3
2.8
dT
rG13
C
FANCF off-target #2
K+
3.2
2.8
3.1
2.7
dT
rG13
D
AAVS1 off-target #5
K+
3.1
2.8
2.9
2.8
rG4
dT
E
AAVS1 off-target #1
K+
2.8
3.5
2.8
2.8
rG4
dT
F
TRAC off-target #1
3.2
2.9
dT
rG2
G
TRAC off-target #2
2.7
3.0
dT
rG2

Figure S8

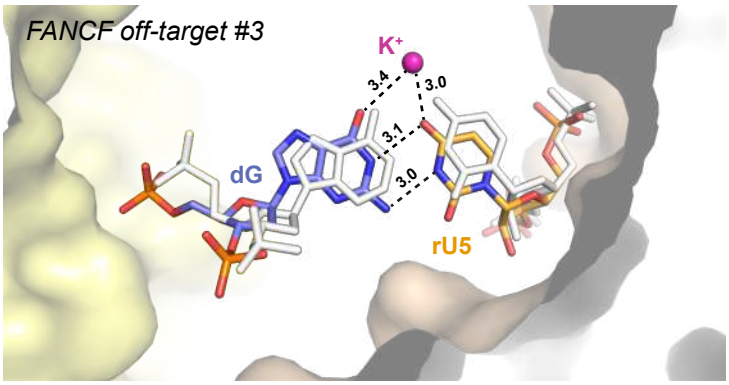
A



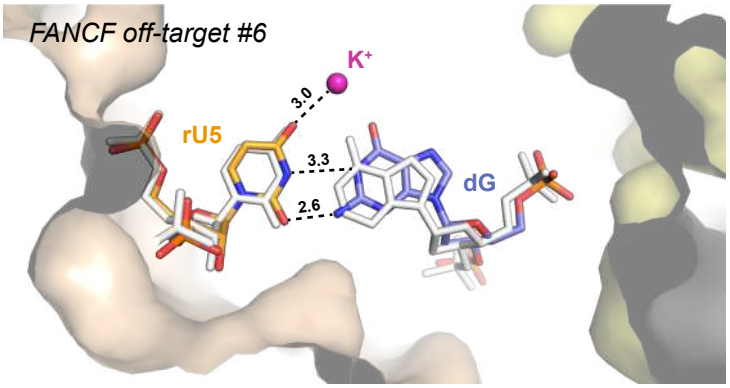
B



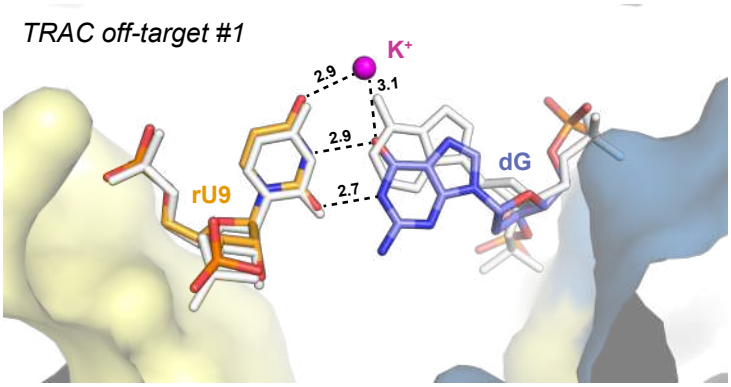
C



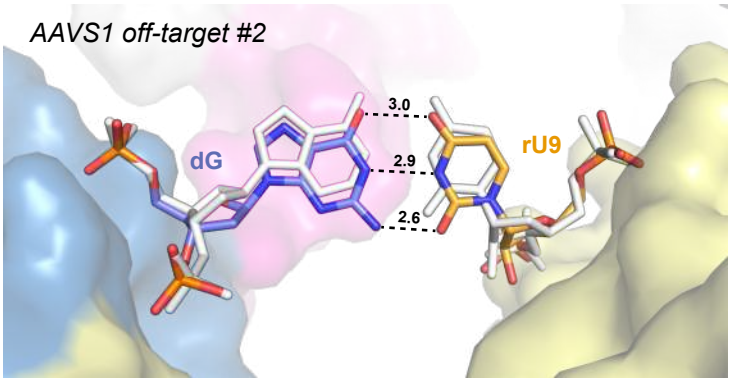
D



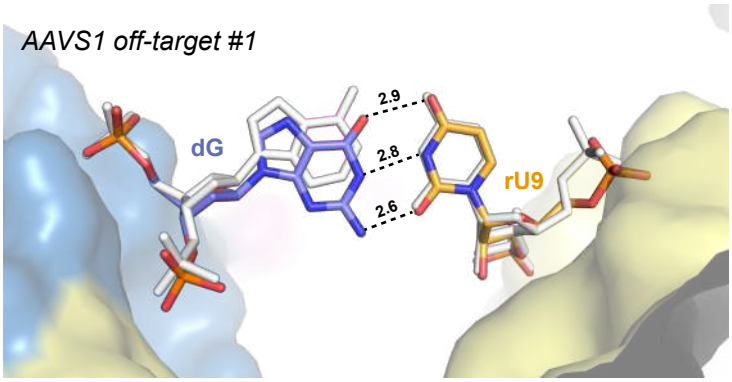
E



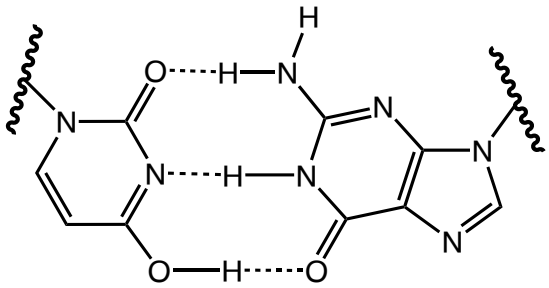
F



G



H



keto-enol-Uracil amino-dGuanine

Figure S9

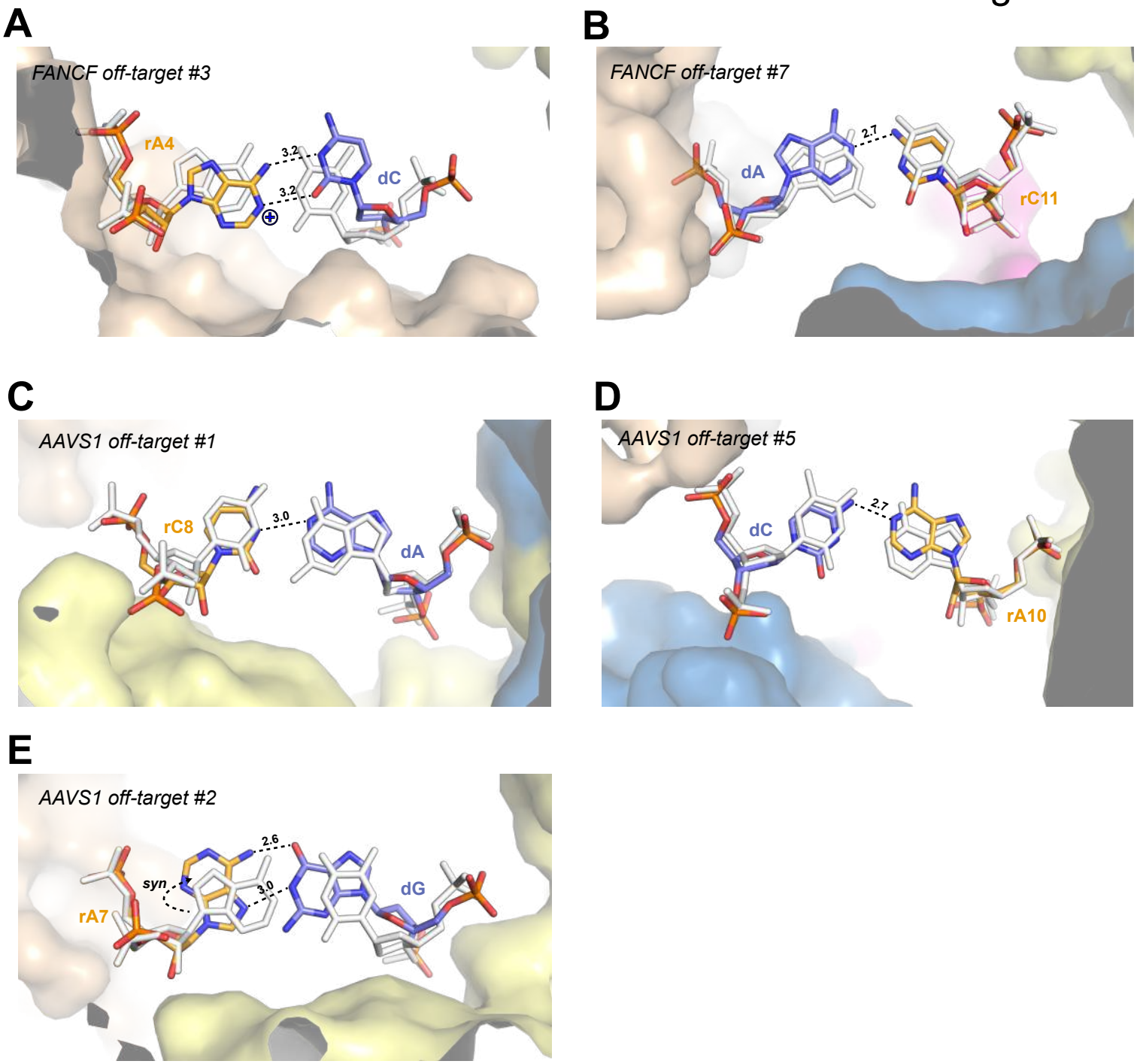

A
FANCF off-target #3
rA4
3.2
3.2
dC
B
FANCF off-target #7
dA
2.7
rC11
C
AAVS1 off-target #1
rC8
3.0
dA
D
AAVS1 off-target #5
dC
2.7
rA10
E
AAVS1 off-target #2
2.6
syn
3.0
dG
rA7

Figure S10

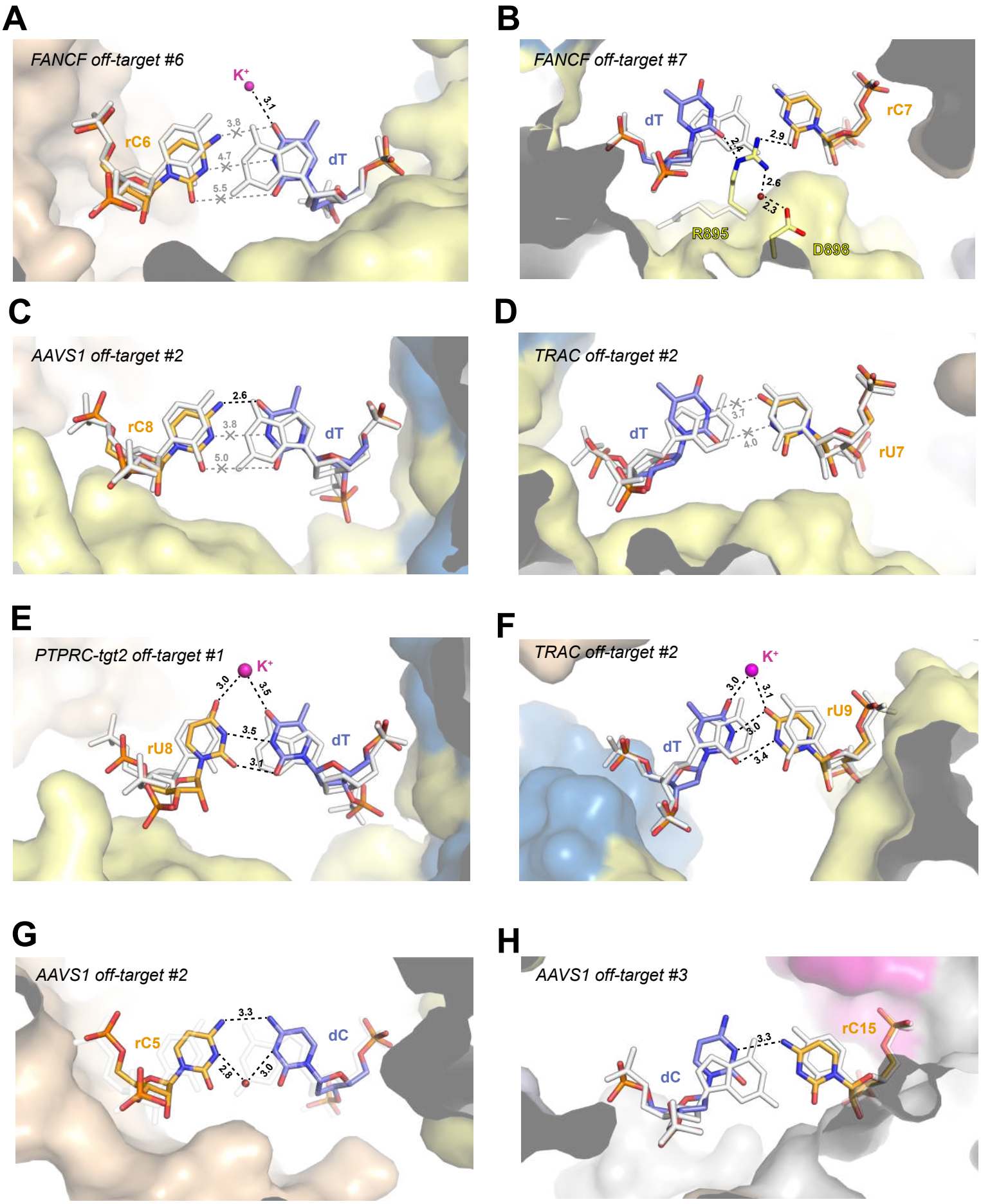

A
FANCF off-target #6
K+
3.1
3.8
4.7
5.5
rC6
dT
B
FANCF off-target #7
dT
rC7
2.4
2.9
2.6
2.3
R895
D898
C
AAVS1 off-target #2
2.6
3.8
5.0
rC8
dT
D
TRAC off-target #2
dT
3.7
4.0
rU7
E
PTPRC-tgt2 off-target #1
K+
3.0
3.5
3.5
3.1
rU8
dT
F
TRAC off-target #2
K+
3.0
3.1
3.0
3.4
rU9
dT
G
AAVS1 off-target #2
3.3
2.8
3.0
rC5
dC
H
AAVS1 off-target #3
3.3
rC15
dC

Figure S11

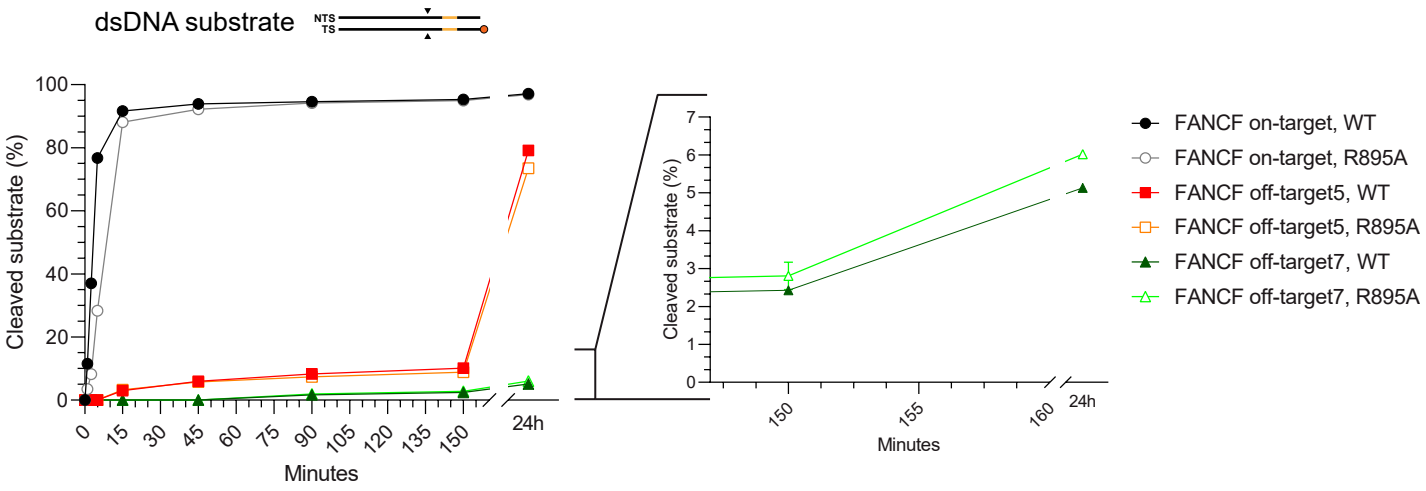
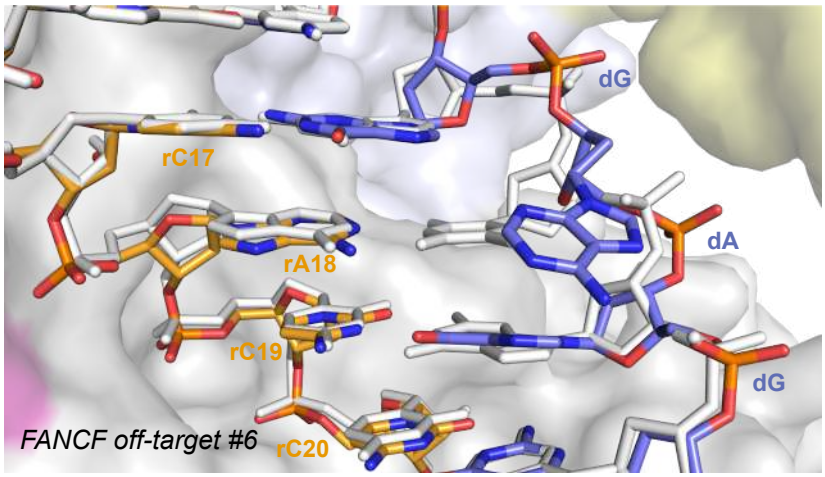

dsDNA substrate
NTS
TS
Cleaved substrate (%)
100
80
60
40
20
0
0
15
30
45
60
75
90
105
120
135
150
24h
Minutes
Cleaved substrate (%)
7
6
5
4
3
2
1
0
150
155
160
24h
Minutes
FANCF on-target, WT
FANCF on-target, R895A
FANCF off-target5, WT
FANCF off-target5, R895A
FANCF off-target7, WT
FANCF off-target7, R895A

Figure S12

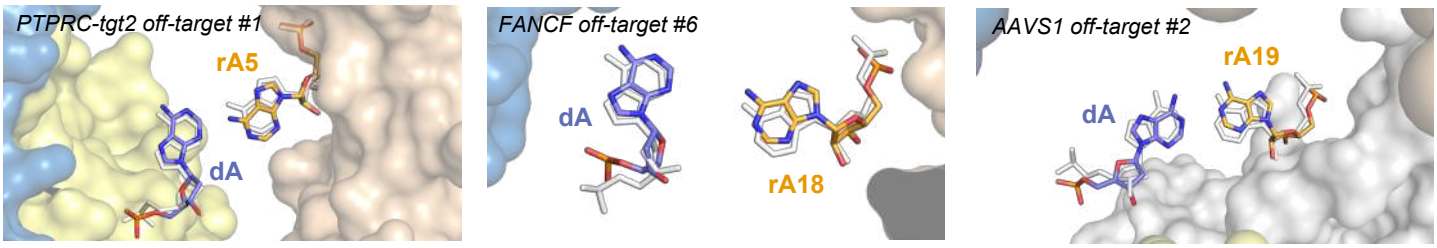
A



B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Total Mismatches	2029	1771	1453	1295	1100	882	888	924	1171	885	632	769	738	409	506	524	484	708	775	801
dA-rA mismatches	139	36	322	74	66	97	72	39	51	85	49	15	7	57	2	24	16	89	155	148
%dA-rA mismatches	10.8%	9.5%	7.8%	6.9%	5.9%	4.7%	4.7%	4.9%	6.3%	4.7%	3.4%	4.1%	3.9%	2.2%	2.7%	2.8%	2.6%	3.8%	4.1%	4.3%

C



D

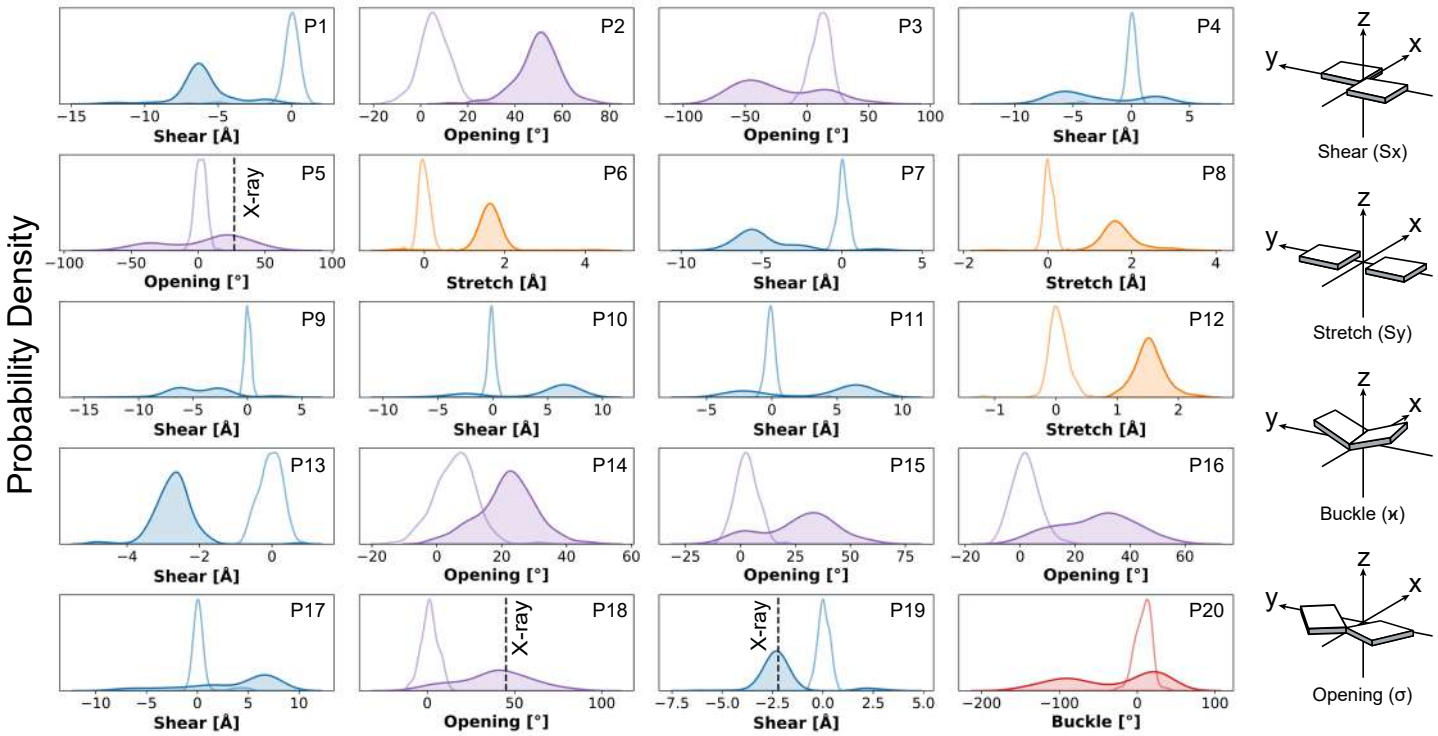


Figure S13

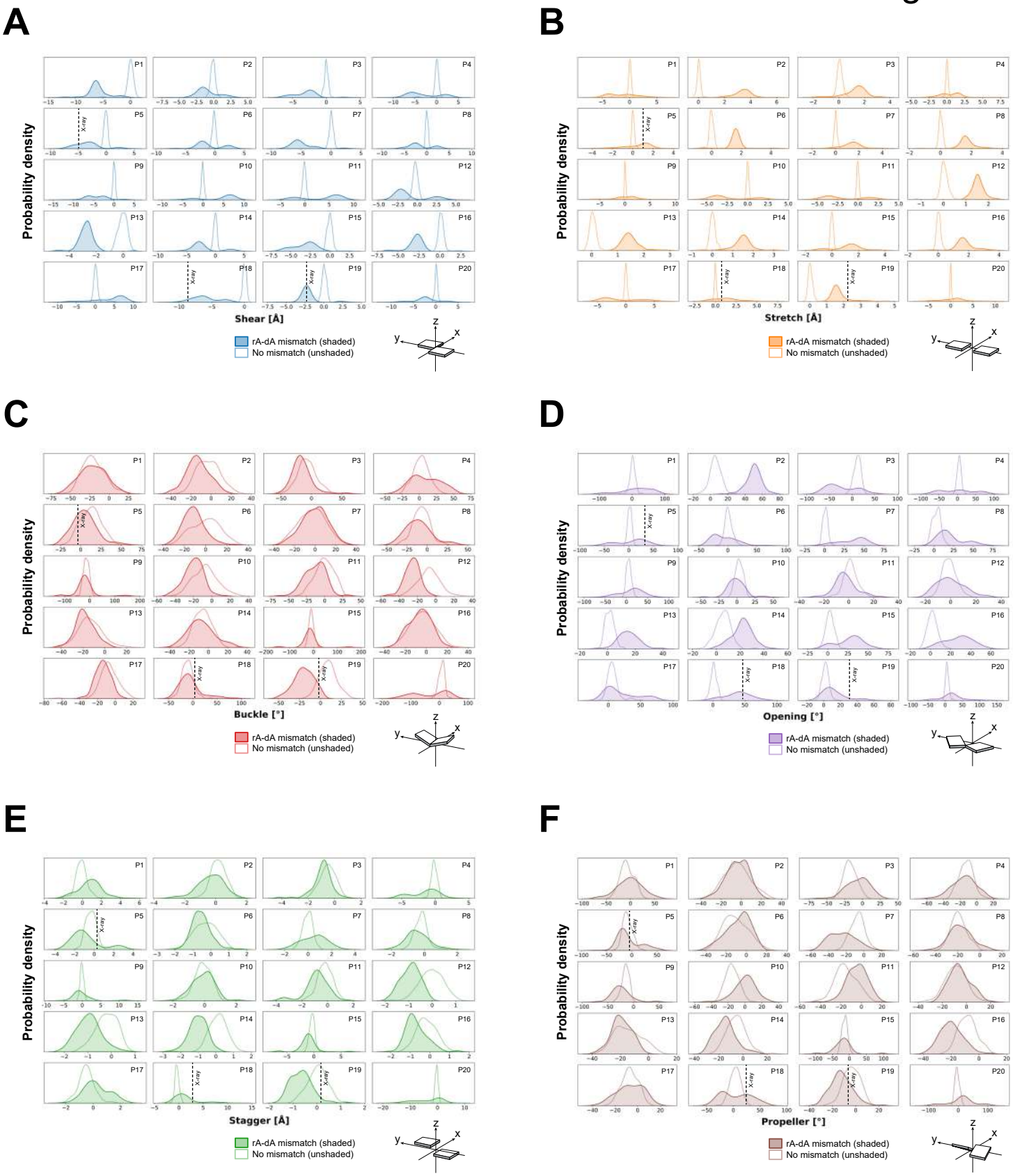
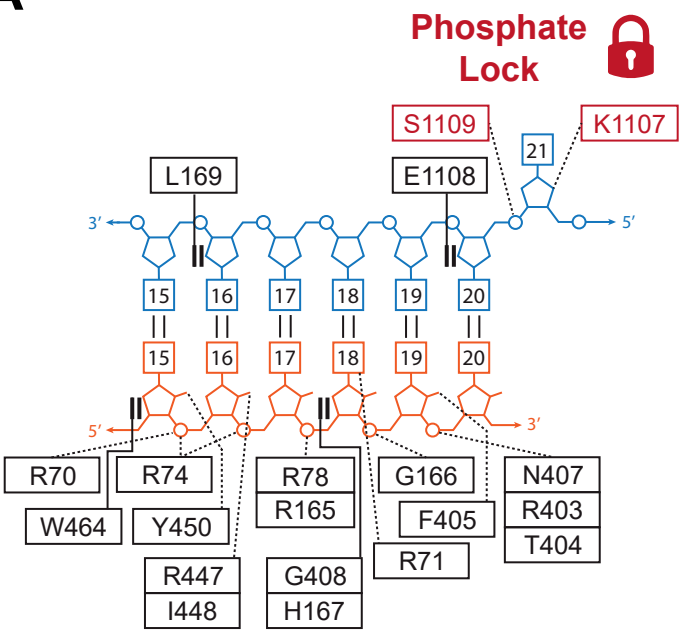
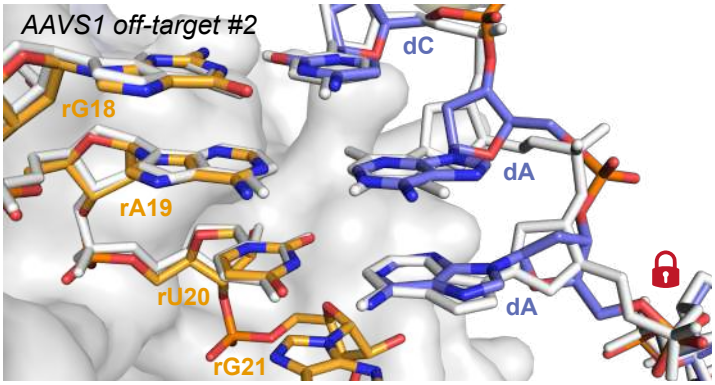

A
B
C
D
E
F
Probability density
Shear [Å]
Stretch [Å]
Buckle [°]
Opening [°]
Stagger [Å]
Propeller [°]
X-ray
rA-dA mismatch (shaded)
No mismatch (unshaded)
z
x
y

Figure S14

A



B



C

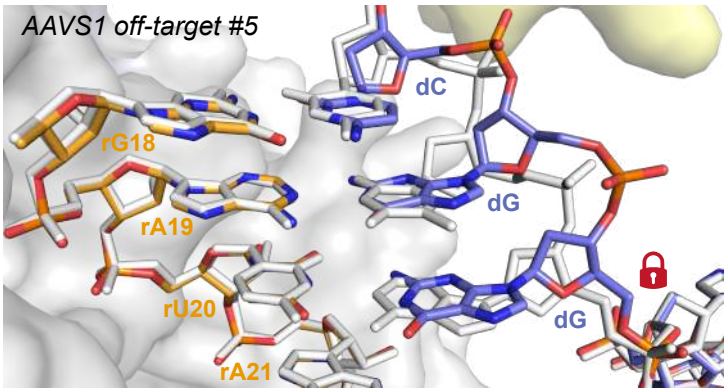


Figure S15

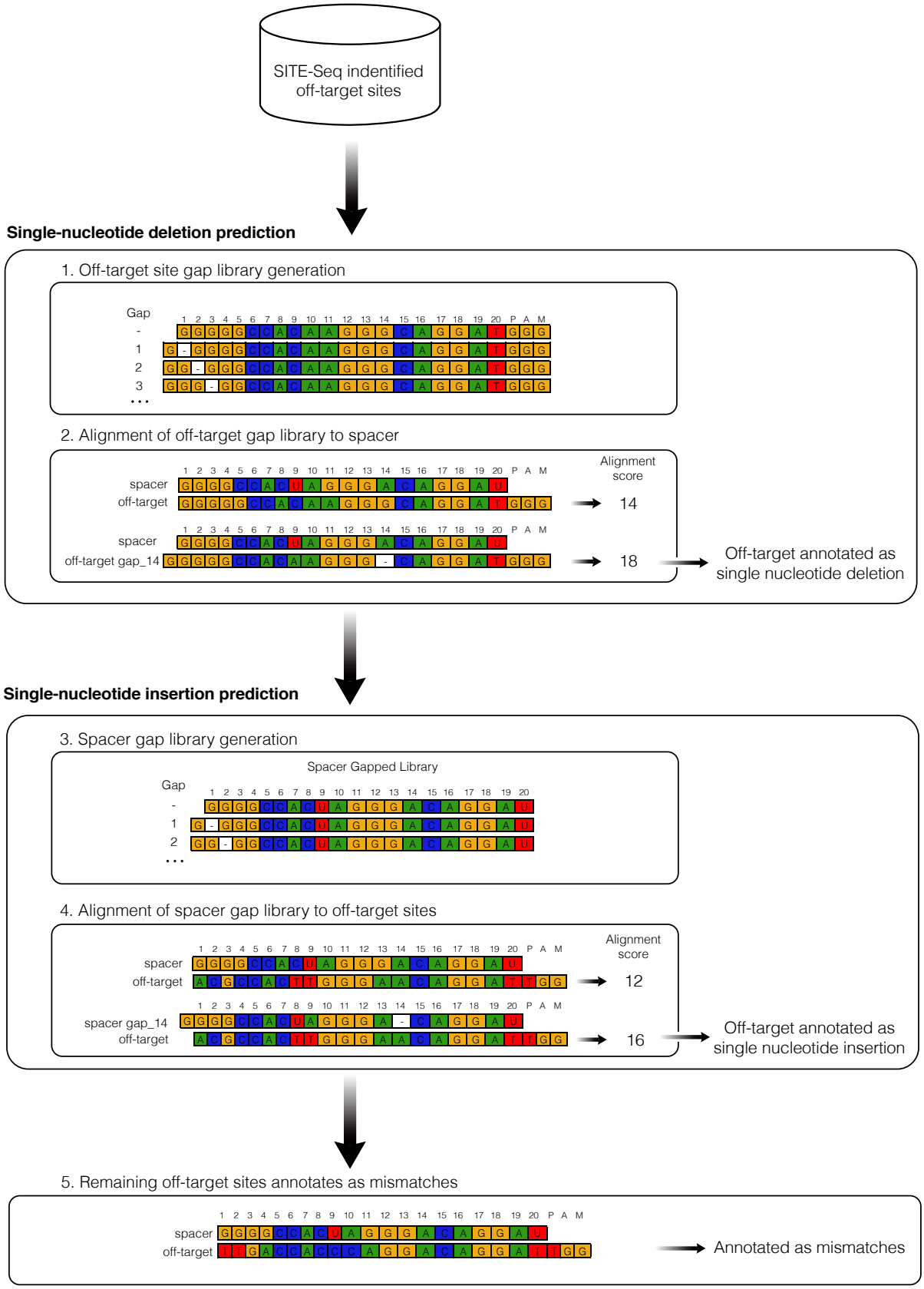

SITE-Seq indentified off-target sites
Single-nucleotide deletion prediction
1. Off-target site gap library generation
Gap
2. Alignment of off-target gap library to spacer
Alignment score
spacer
off-target
off-target gap_14
14
18
Off-target annotated as single nucleotide deletion
Single-nucleotide insertion prediction
3. Spacer gap library generation
Spacer Gapped Library
4. Alignment of spacer gap library to off-target sites
spacer gap_14
12
16
Off-target annotated as single nucleotide insertion
5. Remaining off-target sites annotates as mismatches
Annotated as mismatches

Figure S16

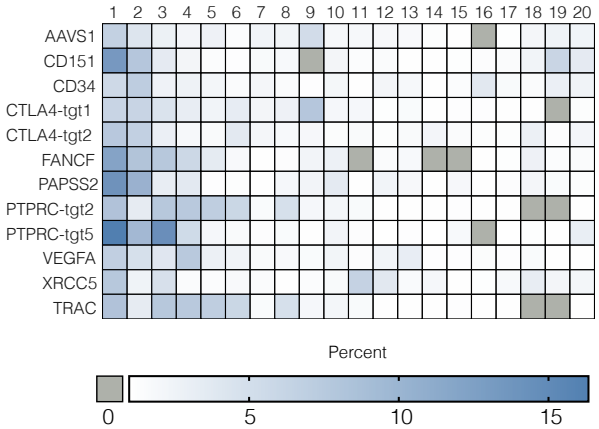
A

SITE-Seq Assay Recovered Sites

	4 nM			16 nM			64 nM			256 nM		
	Mismatches	Deletions	Insertions	Mismatches	Deletions	Insertions	Mismatches	Deletions	Insertions	Mismatches	Deletions	Insertions
AAVS1	18	1	0	49	5	1	244	20	5	916	89	20
CD151	2	0	0	6	0	0	22	3	3	70	5	4
CD34	3	0	0	9	1	1	134	14	5	601	68	21
CTLA4-tgt1	4	0	0	21	2	2	93	7	8	200	17	15
CTLA4-tgt2	3	0	0	7	0	0	24	2	0	131	20	6
FANCF	4	0	0	9	0	0	54	1	1	149	7	7
PAPSS2	2	0	0	16	0	1	62	6	5	178	27	11
PTPRC-tgt2	4	0	0	25	0	0	142	4	0	338	9	3
PTPRC-tgt5	5	1	0	10	1	0	39	3	1	98	5	7
VEGFA	4	0	0	29	0	0	138	2	1	390	35	8
XRCC5	0	0	0	3	0	0	26	5	5	106	30	10
TRAC	11	0	0	48	2	0	142	4	1	234	9	4
Total Mismatches	60	2	0	232	11	5	1120	71	35	3411	321	116

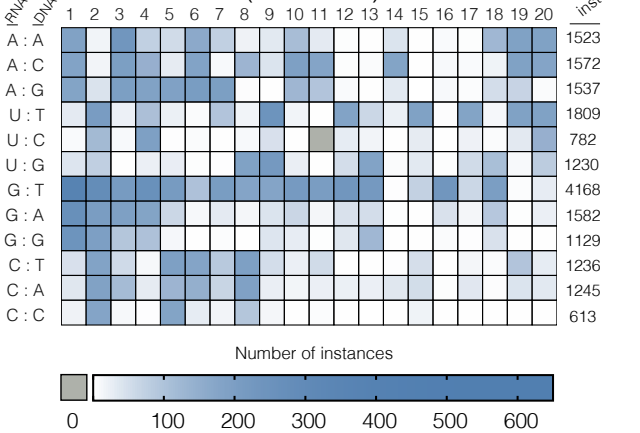
B

Mismatch Position
(n = 3411)



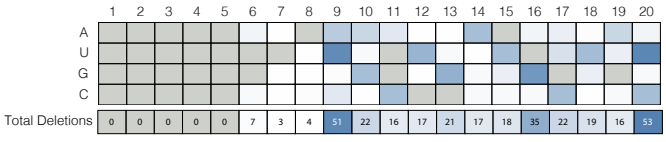
C

Nucleotide Mismatch Frequency
(n = 3411)



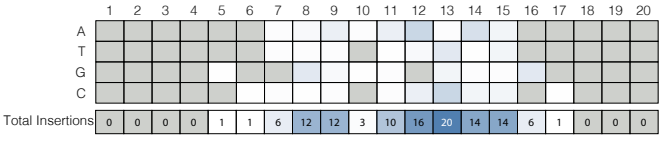
D

Single-nucleotide Deletion Frequence
(n =321)



E

Single-nucleotide Insertion Frequence
(n = 116)



Number of Sites

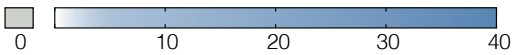
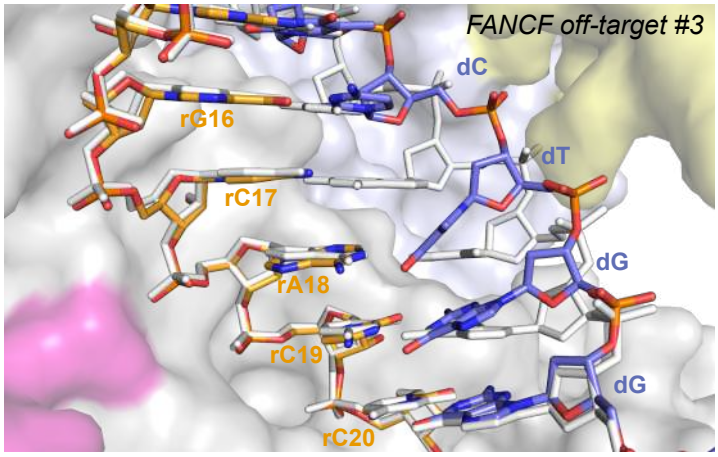
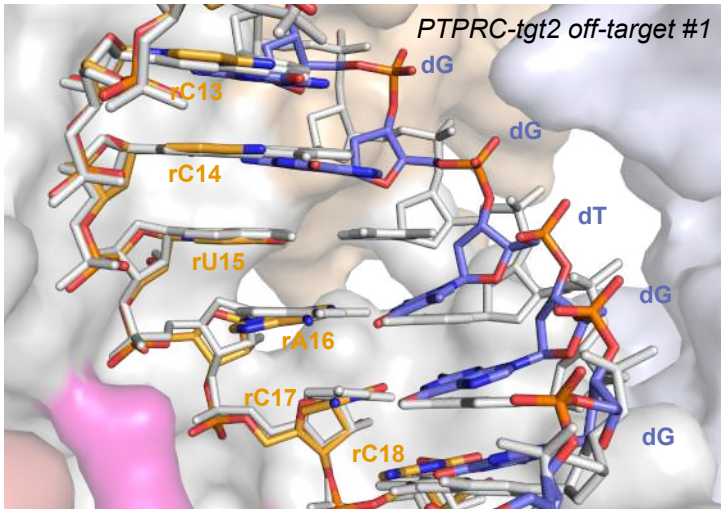


Figure S17

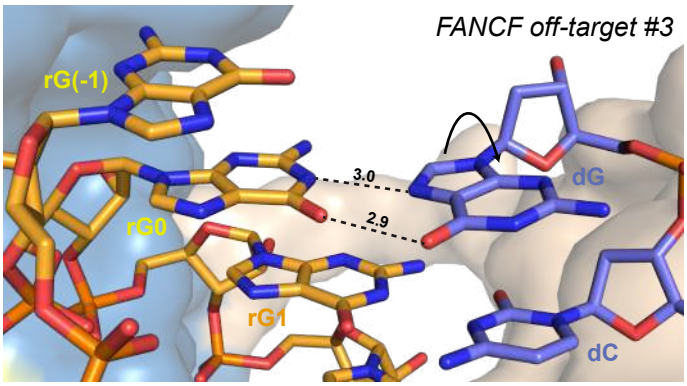
A



B



C



D

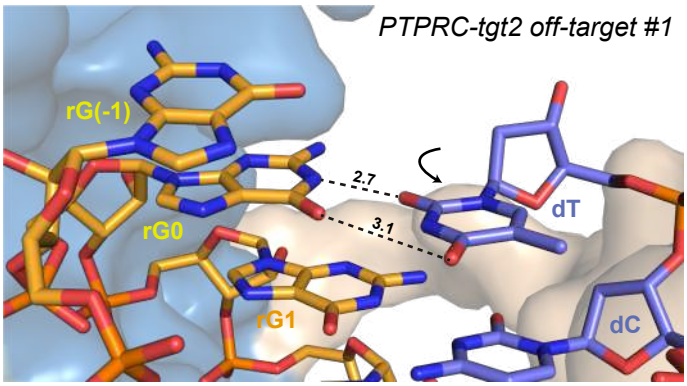
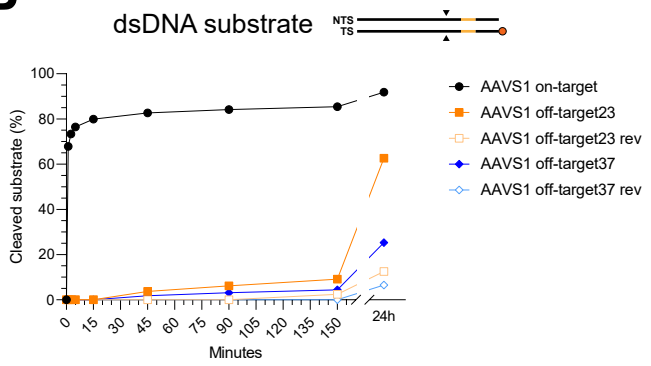


Figure S18

A

Gene	sequence	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	P	A	M
AAVS1	guide RNA	G	G	G	G	G	C	C	A	C	U	A	G	G	G	A	C	A	G	G	A	U			
AAVS1	on-target		C	C	C	C	G	G	T	G	A	T	C	C	C	T	G	T	C	C	T	A	A	C	C
AAVS1	off-target23-A	G	T	·	·	·	·	·	·	·	—	·	·	·	·	G	·	·	·	·	·	·	C	C	C
AAVS1	off-target23-B		·	T	·	·	C	·	G	T	G	·	·	·	·	G	·	·	·	·	·	·	C	C	C
AAVS1	off-target23-rev		·	T	·	·	C	·	G	·	G	·	·	·	·	G	·	·	·	·	·	·	C	C	C
AAVS1	off-target37-A	T	A	·	T	·	·	·	·	·	—	·	·	T	·	·	·	·	·	·	·	·	C	C	C
AAVS1	off-target37-B		T	A	·	T	C	·	G	T	G	·	·	T	·	·	·	·	·	·	·	·	C	C	C
AAVS1	off-target37-rev		T	A	·	T	C	·	G	·	G	·	·	T	·	·	·	·	·	·	·	·	C	C	C

B



C

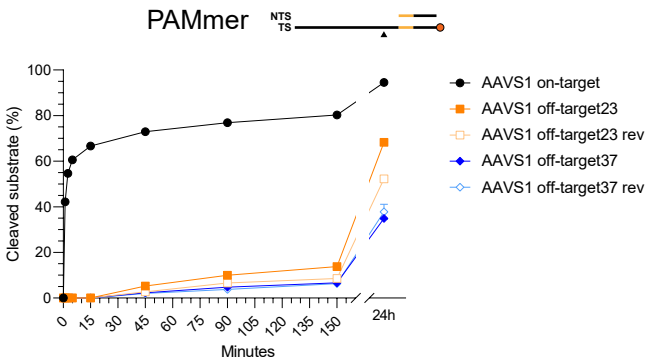
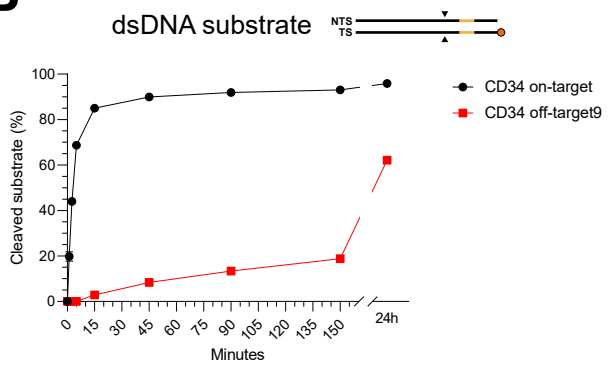


Figure S19

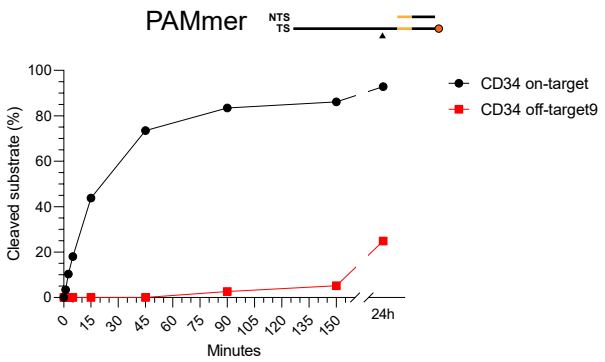
A

CD34	guide RNA	G	A	U	A	G	G	A	G	A	A	G	A	U	G	A	U	G	U	A	U	A		
CD34	on-target		T	A	T	C	C	T	C	T	T	C	T	A	C	T	A	C	A	T	A	T	C	C	C
CD34	off-target9-A	C	·	·	·	·	·	–	·	·	·	·	·	·	·	·	·	·	T	·	·	·	A	C	C
CD34	off-target9-B		C	T	A	T	·	C	·	·	·	·	·	·	·	·	·	·	T	·	·	·	A	C	C

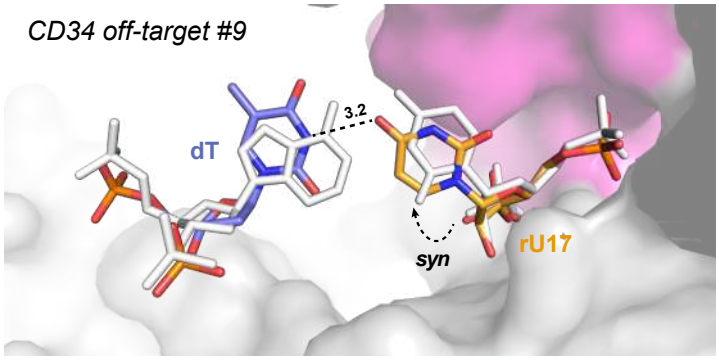
B



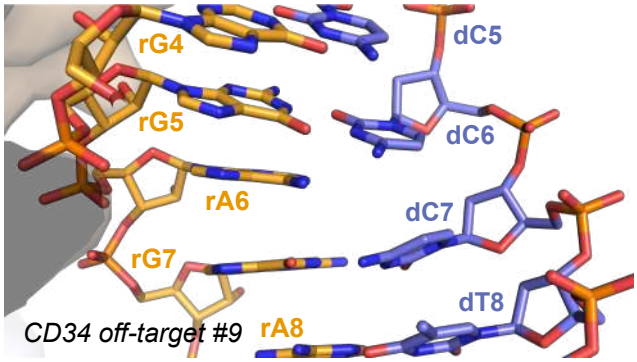
C



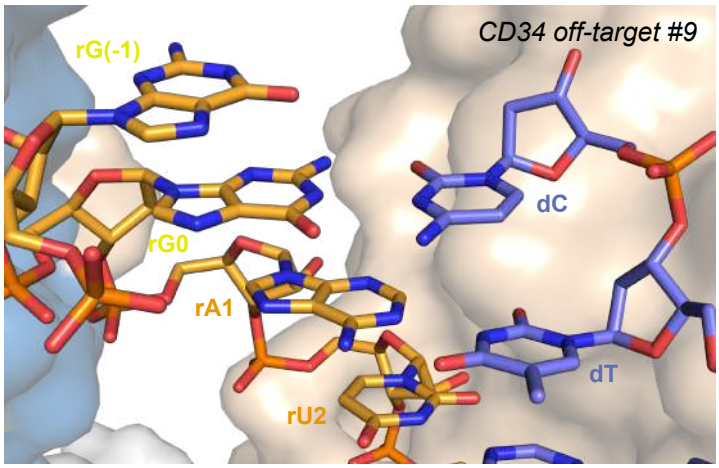
D



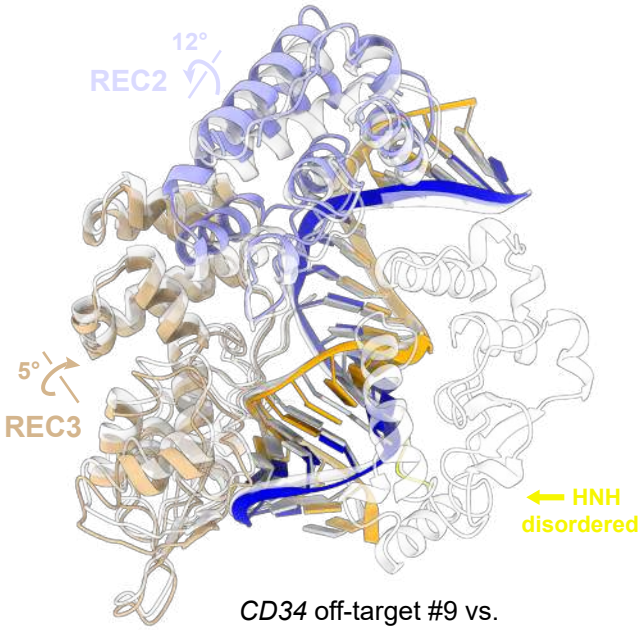
E



F



G



CD34 off-target #9 vs.
TRAC on-target

Figure S20

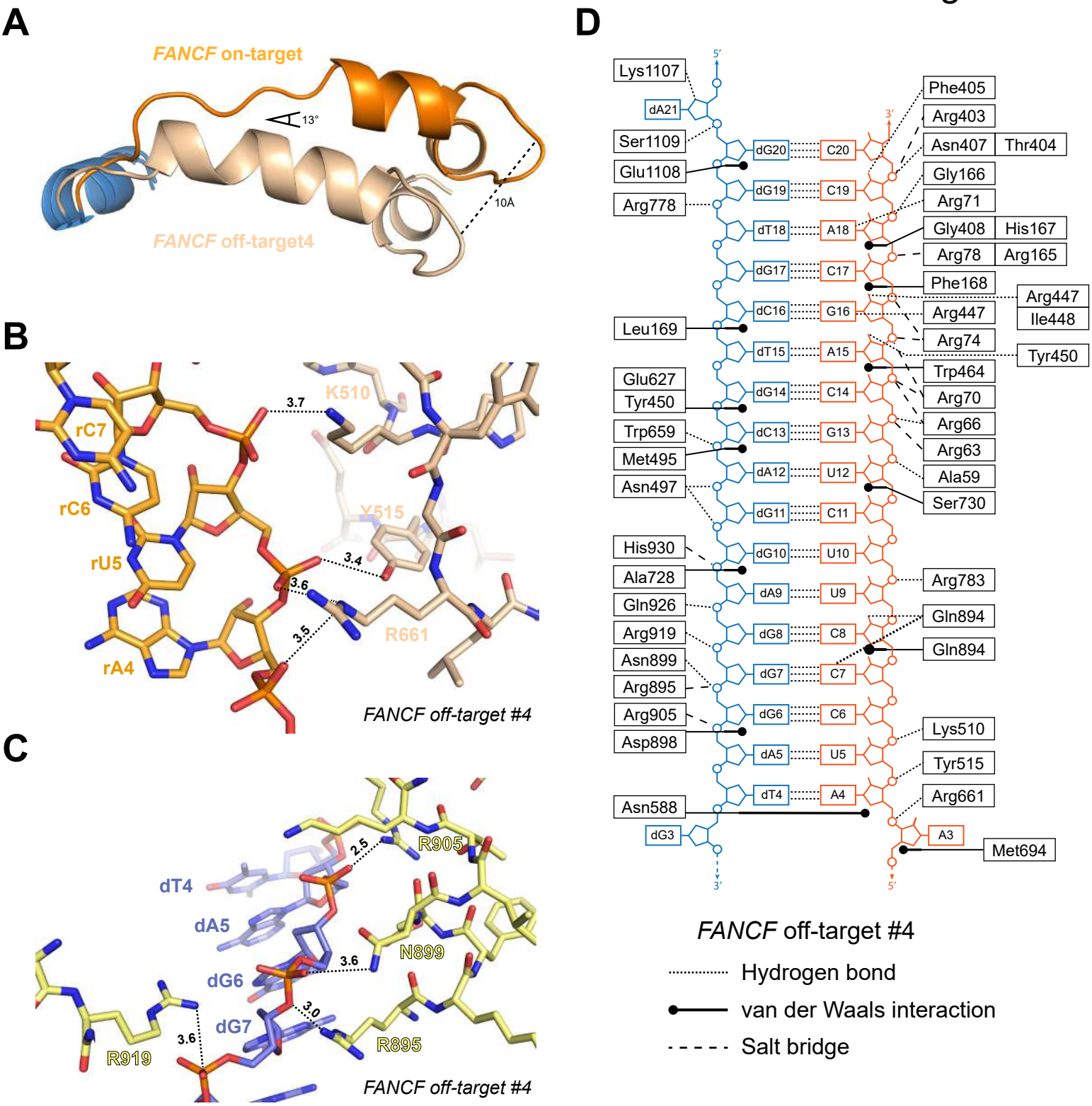
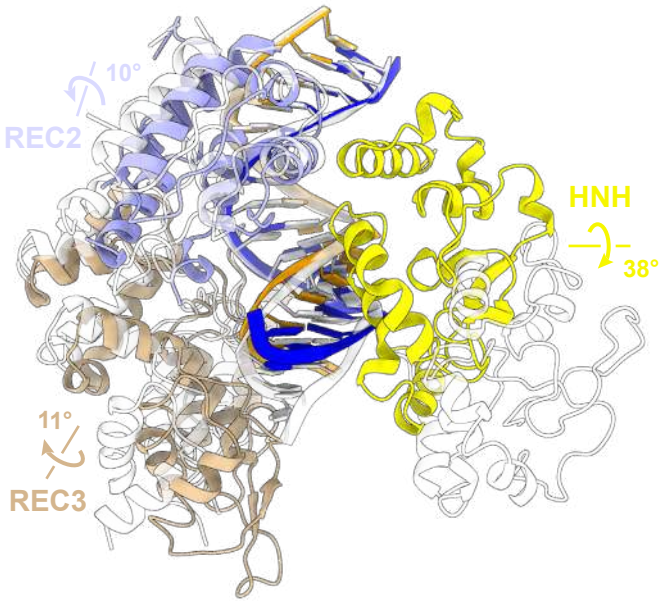

A
FANCF on-target
13°
10Å
FANCF off-target4
B
K510
3.7
rC7
rC6
rU5
rA4
Y515
3.4
3.6
3.5
R661
FANCF off-target #4
C
R905
2.5
dT4
dA5
N899
3.6
dG6
3.0
dG7
3.6
R919
R895
FANCF off-target #4
D
Lys1107
dA21
Ser1109
Glu1108
Arg778
Leu169
Glu627
Tyr450
Trp659
Met495
Asn497
His930
Ala728
Gln926
Arg919
Asn899
Arg895
Arg905
Asp898
Asn588
dG3
dG20
dG19
dT18
dG17
dC16
dT15
dG14
dC13
dA12
dG11
dG10
dA9
dG8
dG7
dG6
dA5
dT4
C20
C19
A18
C17
G16
A15
C14
G13
U12
C11
U10
U9
C8
C7
C6
U5
A4
A3
Phe405
Arg403
Asn407
Thr404
Gly166
Arg71
Gly408
His167
Arg78
Arg165
Phe168
Arg447
Arg447
Ile448
Arg74
Tyr450
Trp464
Arg70
Arg66
Arg63
Ala59
Ser730
Arg783
Gln894
Gln894
Lys510
Tyr515
Arg661
Met694
FANCF off-target #4
Hydrogen bond
van der Waals interaction
Salt bridge

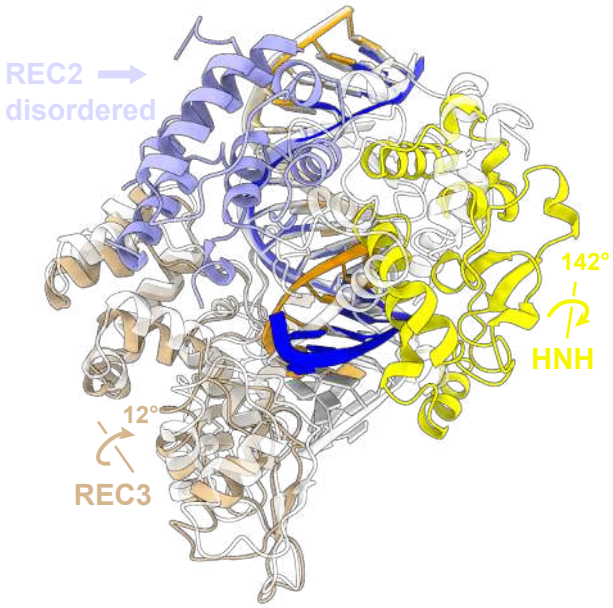
Figure S21

A



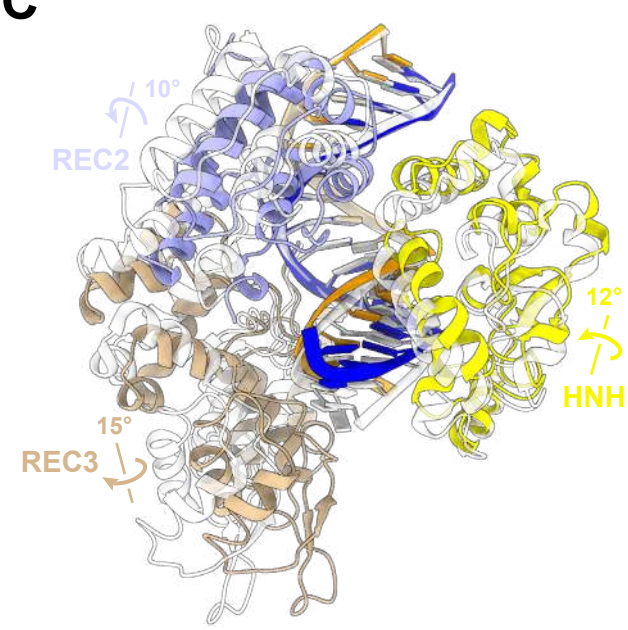
FANCF off-target4 vs.
pre-catalytic complex (PDB: 6O0Z)

B



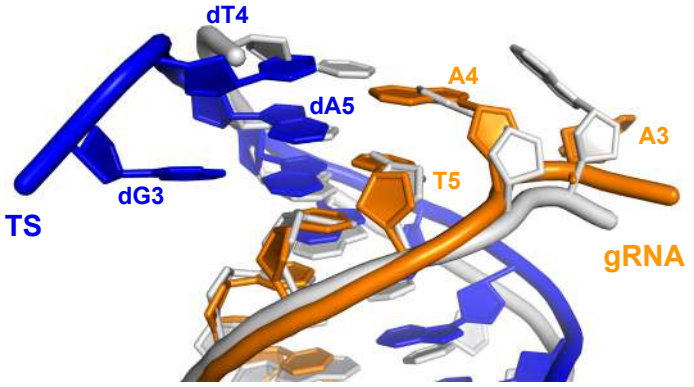
FANCF off-target4 vs.
post-catalytic complex (PDB: 6O0Y)

C



FANCF off-target4 vs.
xCas9 3.7 HF mutant (PDB: 6AEG)

D



FANCF off-target4 xCas9 3.7 (PDB: 6AEG)

Preflight Results

Document Information

Title: ilovepdf_merged_compressed (1).pdf
Author:
Creator:
Producer: iLovePDF

Preflight Information

Profile: Convert to PDF/A-2b
Version: Qoppa jPDFPreflight v2022R1.00
Date: Aug 24, 2022 4:11:32 PM

Legend: (X) - Can NOT be fixed by PDF/A-2b conversion.
(!X) - Could be fixed by PDF/A-2b conversion. User chose to be warned in PDF/A settings.

Page 68 Results
(X) Page uses transparency but does not have a device independent Blending Color Space

Page 70 Results
(X) Page uses transparency but does not have a device independent Blending Color Space

Page 79 Results
(X) Page uses transparency but does not have a device independent Blending Color Space

Page 83 Results
(X) Page uses transparency but does not have a device independent Blending Color Space
(X) Font widths must be the same in both the font dictionary and the embedded font file.