nature biotechnology

Article

https://doi.org/10.1038/s41587-022-01492-y

Precise DNA cleavage using CRISPR-SpRYgests

In the format provided by the authors and unedited

Supplementary Material

Supplementary Tables

Note: *All Supplementary Tables are attached together as a single .xlsx file, separated by tabs.*

Supplementary Notes

Supplementary Figures

Supplementary References

Supplementary Notes:

Supplementary Note 1: Restriction enzymes

Restriction enzymes (REs) transformed the field of molecular cloning by enabling and accelerating the assembly of recombinant DNA fragments[1]. REs commonly utilized for cloning applications recognize sequence motifs typically 4-8 base pairs in length and generate DNA double strand breaks (DSB) (**Fig. 1a**). Despite a diverse catalog of REs, there remain challenges for molecular cloning since the availability of these motifs in DNA substrates is stochastic, and to be useful, the RE motif must be conveniently located and generally occur only once (so-called 'single cutters'). Aside from purposefully designed multiple cloning sites (MCSs) harboring several unique RE motifs (**Supplementary Fig. 1a**), most RE motifs either do not occur in a substrate or are found more than once, preventing specific digestion of a DNA substrate to generate the desired fragments. Although improved molecular cloning methods like isothermal assembly (i.e. Gibson Assembly[2]) can improve flexibility, these approaches have notable drawbacks (**Supplementary Figs. 1b**, **1c**, and **Supplementary Note 2**). Given these caveats, most laboratories purchase large repertoires of REs which can be costly but remain inadequate since REs cannot comprehensively address all sequences.

Supplementary Note 2: Isothermal assembly

One method that has been developed to circumvent some caveats of identifying unique restriction sites in a plasmid backbone is isothermal assembly (commonly known as Gibson Assembly[2]). Isothermal assembly enables the modification of DNA constructs through enzymatic assembly of multiple different DNA fragments. The protocol typically involves first cleaving the backbone at one or two sites to provide free DNA ends (usually by restriction digest), generating PCR products of the insert or modification of interest, generating PCR products of flanking sequences that bridge from the insert to the restriction site, and then performing the assembly reaction with an enzyme mixture comprised of an exonuclease, a polymerase, and a ligase (which rely on homology of the overlapping sequences between the fragments and the backbone; **Supplementary Fig. 1b**). Assembly reactions can contain several inserts depending on the complexity of the cloning objective. To initiate the reaction, it is advisable to liberate free DNA ends in the backbone using restriction enzymes (instead of performing around-the-backbone PCR; see more below). However, like most traditional cloning methods, this makes isothermal assembly beholden to unique restriction sites proximal to the site of intended insertion or modification. As the distance between the alteration of interest and unique restriction sites increases, so too does the size of the requisite bridging PCR products, which can reduce efficiency/fidelity of the assembly reaction. Furthermore, this might also necessitate the design of additional PCR fragments, and might require spanning low complexity regions (i.e. homopolymeric regions) or high/low GC-content. An additional area that can require extensive troubleshooting is the isothermal assembly flanks (regions of homology between the fragments to be assembled), as these regions are restricted to specific sequences either 5' or 3' of the restriction sites, which can mean that they are high or low in Tm, have poor sequence complexity, or have high hairpin propensity (which again, together can all reduce efficiency). Finally, certain methods try to circumvent these issues by performing around-the-backbone PCR to generate the linear backbone 'acceptor' molecule, instead of generating the backbone via restriction digestion (which would in principle make the assembly more efficient by requiring fewer fragments). However, this approach is fraught with issues, including then necessitating sequencing of the entire plasmid backbone of any derivative clonal constructs, since most polymerases don't have adequate fidelity to avoid introducing unwanted polymorphisms in the plasmid backbone. Until recent more affordable whole-plasmid sequencing methods were developed[3,4], it was expensive, cumbersome, and often unfeasible (due to low complexity or high GC-content regions) to sequence most plasmid backbones. Moreover, around-the-backbone PCR is infeasible for large backbones and plasmids carrying repetitive sequences, as these frequently lead to polymerase stalling or slippage. Thus, considering these potential pain points, a simpler way to decrease the cost, reduce the complexity, and improve the efficiency and fidelity of isothermal assembly reactions would be to open the plasmid backbone directly adjacent to the modification of interest (**Supplementary Fig. 1c**), which in the majority of cloning projects is not possible via conventional restriction enzyme digests.

<u>Supplementary Note 3</u>: Rapid synthesis of gRNAs by oligo extension and IVT

To expedite the production of gRNAs for custom SpRYgests, we compared several methods to generate a dsDNA template for IVT of gRNAs. First, we cloned duplexed oligonucleotides (oligos) corresponding to the spacer sequence into a BsaI-digested entry plasmid that harbors a T7 promoter and the SpCas9 gRNA scaffold (MSP3485). The resulting pT7-spacer-gRNA plasmid can be linearized using HindIII for run-off transcription that terminates several bases after the 3' end of the gRNA. Secondly, we generated DNA substrates encoding pT7-gRNAs via oligo extension using either Klenow Fragment (3'→5' exo-) or Q5 polymerase. Oligo extension is performed by incubating a target-specific oligo encoding the T7 promoter, the 20bp spacer and a portion of the SpCas9 scaffold, where the latter section permits annealing with a common bottom-strand oligo (oKAC682) encoding the remainder of the SpCas9 scaffold. The common oligo partially anneals to the target-specific oligo, allowing oligo extension to generate a full-length dsDNA template for IVT (**Supplementary Fig. 3a**). We observed that the three gRNA synthesis methods led to gRNAs that exhibited comparable cleavage efficiencies during SpRYgests (**Supplementary Fig. 3b**).

Notably, when performing the manufacturer recommended RQ1 DNase treatment to remove the *in vitro* transcription DNA template, we observed partial degradation of the target substrate molecule even when performing a bead clean-up of the gRNA prior to the SpRYgest reaction (**Supplementary Fig. 12a**). We also performed a heating step to re-fold the gRNA (5 min at 95 °C, which should at least partially heat inactivate the DNase), but still observed non-specific degradation of the substrates. We hypothesize that residual DNase carried through the clean-up can cause partial degradation of the substrate, which is supported by the complete degradation of the substrate in the presence of concentrated RQ1 DNase (**Supplementary Fig. 12b**). In subsequent optimized one-pot gRNA transcription reactions, we elected to omit the DNase treatment rather than performing a heat-inactivation step (when using thermolabile DNase) to avoid adding additional steps or complexity to the SpRYgest workflow.

We further optimized and streamlined the gRNA generation process by combining the dsDNA template generation and IVT steps (**Fig. 4a** and **Supplementary Fig. 17a**). We demonstrated that the gRNA yield from one-pot reactions performed for as little as one hour generate sufficient gRNA quantities for full-scale SpRYgests (**Supplementary Fig. 17**). The one-pot gRNA synthesis method only involves approximately 10 minutes of hands-on time and incubations between 1 to 4 hours (**Supplementary Figs. 17a**-**17c** and **18a**). This timing is in comparison to generating a gRNA using a cloned IVT template plasmid by RE digest and ligation, which can take at least 3 days. For each of these methods, gRNA synthesis can be performed in parallel to ordering PCR primers that are otherwise required for cloning, making the timing of SpRYgests not substantially different than when using REs (**Supplementary Fig. 18**).

Supplementary Note 4: Analysis of cleavage rate constants and target site features

Although a failure to reach digest completion using SpRYgest was rare (7 out of the 103 SpRYgests performed using protein from lysate (4 partial digests, 3 failed; **Fig. 1h**), we sought to identify putative sequence features of target sites that might explain these occurrences. We calculated the rate constants for substrate cleavage for all WT SpCas9, SpG, and SpRY reactions, and classified them based on reaction speed and completion (**Supplementary Fig. 7a** and **Supplementary Table 1**). While the PAM was a major determinant of activity for WT SpCas9 and SpG, in contrast, no striking sequence features emerged for SpRY (**Supplementary Fig. 7b**). The guanine preference in the 20th position of the spacer is due to the design of target sites that intentionally harbored a guanine in that position, necessary for transcription from the T7 promoter. We also performed a manual analysis of the gRNAs that failed to reach complete substrate cleavage with SpRY (< 80% cleavage at final timepoint) or that had a slow rate constant (< 0.01) (**Supplementary Table 1**). Consistent with previously described negative sequence features[5], 5 of 8 such gRNAs contained a cytosine in the most PAM-proximal position of the spacer, and 7 of 8 of the corresponding target sites encoded an NCN or NTN PAM (consistent with the NRN>NYN preference of SpRY[6]) (**Supplementary Table 1**). However, neither of these features were absolute in their impact on SpRYgest activity. Of the 21 target sites with a C in the first spacer position, 17 were cleaved to completion, and of the 49 sites with an NYN PAM, 43 reactions went to completion (**Supplementary Table 1**). There were 10 total sites that included both sequence features, 4 of which failed to reach completion. Thus, based on this initial analysis, the combination of both sequence features is moderately predictive of potential SpRYgest sites that might fail to reach completion or exhibit poor cleavage kinetics.

To look more comprehensively at sequence trends, we then developed a computational model to predict *in vitro* cleavage efficiency with WT SpCas9 or SpRY. We generated a model using XGBoost[7] trained on 24-nt target sequences (including the 20-nt protospacer + 4-nt PAM) and the WT SpCas9 and SpRY experimental rate constants (**Supplementary Figs. 8a**-**8c**). We evaluated the prediction performance of this model using leave-one-out cross-validation. For WT SpCas9, the model was highly predictive of the rates (Pearson's correlation coefficient (r)=0.946; **Supplementary Fig. 8a**) and the most important nucleotides for the performance of the model correspond to the PAM (with all five target sites with NGGN PAMs being predicted to be efficiently cleaved). This is consistent with the previously characterized NGG PAM preference of WT SpCas9[8–11]. However, the model had poor predictive ability for SpRY cleavage of substrates (Pearson's correlation coefficient (r)=0.086; **Supplementary Fig. 8b**), indicating that SpRY does not have a strong dependency on the PAM[6]. We then performed SHAP (SHapley Additive exPlanations[12]) analysis on the trained XGBoost data to investigate nucleotide importance across target sites. For WT SpCas9, we once again observed that an NGGN PAM was a strong positive predictor of cleavage efficiency (**Supplementary Fig. 8d**). With SpRY, although no sequence features were positively correlated, two moderately negative features were apparent: C or T bases were slightly disfavored in the 2nd position of the PAM, along with cytosines in the PAM-proximal position of the spacer being negatively correlated with activity (**Supplementary Fig. 8e**).

We also assessed whether internal gRNA interactions impacted SpRYgest efficiency[13,14], either via a published script[13] or Geneious Prime (version 2020.1.2). However no clear trends emerged, since although nearly all weak gRNAs with SpRY exhibited substantial intramolecular interactions, many efficient gRNAs did as well.

Although we identified two minor sequence features that can either together or alone, in some cases, be detrimental to SpRYgest efficiency, larger datasets will be required to extract additional informative features. Thus, given these results we recommend that users *ab initio* design a second strand targeting gRNA when the initial gRNA/target site harbors a combination of an NYN PAM and a 1st spacer C.

Supplementary Note 5: Target site and guide design considerations for TtAgo

It was previously demonstrated that *Thermus thermophilus* Argonaute (TtAgo) can be directed to generate a DSB in a dsDNA template using a pair of 5'-phosphorylated ssDNA guides and that it has a strong preference for AT-rich regions of substrates[15]. New England Biolabs has also characterized TtAgo guide requirements: https://www.neb.com/tools-and-resources/usage-guidelines/guidelines-for-optimization-of-tth-argonaute-ttago-reactions. Their recommendations for TtAgo guide design include: 1) to design ssDNA guides with a 5' matched thymidine, 2) to avoid adenosine in the 12th position of guides, 3) to select target sites with low GC content, 4) that the addition of ET SSB may aid in cutting higher GC content substrates under specific conditions, and 5) ThermoPol Reaction Buffer with 2mM magnesium(II) is optimal for most reactions.

All TtAgo reactions were conducted following the protocol outlined by NEB (Protocol for generating break in phosphodiester backbone with Tth Argonaute (TtAgo, NEB #M0665)). As a positive control we ordered a 5'-phosphorylated guide pair from Swarts *et al.*[15] (oKAC1347 and oKAC1348) and cloned 98 base pairs of the 17% GC content target region into pUC19. We then designed 5 pairs of ssDNA guides following the guidelines outlined above. While we found the positive control TtAgo guide could efficiently introduce a DSB in the linearized target plasmid (**Supplementary Fig. 9e**), otherwise, TtAgo was only able to cleave 2 of 5 sites to near-completion when following NEB guide design rules (**Fig. 2a** and **Supplementary Fig. 9d**). We found that increasing the $MgSO_4$ concentration from 2mM to 10mM slightly improved cleavage efficiency (**Supplementary Fig. 9c**), however the increase in magnesium(II) concentration did not rescue activity on failed sites (**Supplementary Fig. 9f**).

While our manuscript was under review, a manuscript describing molecular cleavage reactions using an improved *Pyrococcus furiosus* Argonaute (PfAgo)-based workflow was published[16,17]. Although this system offers advantages for scalable assembly of multiple fragments with user-definable overhangs, the efficiency of cleavage using PfAgo remains incomplete for most target sites and there are restrictive guide design requirements. Additionally, PCR-based methods have higher inherent error compared to SpRYgests, due to the fidelity of polymerases.

<u>Supplementary Note 6</u>: Optimization of a one-pot gRNA IVT reaction

To streamline the process of performing a SpRYgest, we sought to optimize a one-pot gRNA synthesis reaction that could simultaneously produce the dsDNA *in vitro* transcription template and *in vitro* transcribe the gRNA (**Supplementary Fig. 17a**). The one-pot gRNA product could then be directly combined with SpRY protein and the DNA substrate to assemble the complete SpRYgest reaction (**Fig. 4a**).

To explore the feasibility of a one-pot gRNA synthesis method, first we determined the gRNA yields at various reaction times for our previously established gRNA synthesis method (**Supplementary Note 3**) and a commercially available one-pot kit (EnGen sgRNA Synthesis Kit; NEB, # E3322S) (**Supplementary Fig. 17b**). While our standard IVT method and the one-pot EnGen kit generated sufficient quantities of RNA by 4 hours for two different gRNAs, our standard IVT method yielded approximately 4-fold more gRNA than the EnGen kit at each timepoint (**Supplementary Fig. 17b**). We then determined whether we could assemble a more efficient one-pot reaction that could generate comparable gRNA yields to our standard IVT method. To do so, we examined seven different combinations of reagents, including one-pot reactions comprised of Klenow Fragment (3'→5' exo-) (NEB, M0212L) to generate the IVT template, and RiboMAX T7 Express Enzyme Mix (Promega) to transcribe the gRNA (**Supplementary Fig. 17c**). Other variations included the additional of 3x more target-specific and common oligos (for both our homebrew one-pot and the EnGen kit), or scaling down the reagents used by our homebrew one-pot method by 1/4 to reduce cost. Of all combinations tested, we found that our new one-pot and scaled-down one-pot reactions consistently generated greater quantities compared to our previous standard 2-step IVT method or the EnGen kit (**Supplementary Fig. 17c**).

With each of the seven gRNA synthesis methods that we assessed for four different gRNAs, we then performed SpRYgest reactions using the unpurified and unquantified gRNA transcription products directly in SpRYgest reactions containing 859.5 fmol of plasmid substrate (**Supplementary Fig. 17d**). For 3 of 4 gRNAs, we observed complete cleavage for all seven gRNA synthesis methods. However, for the remaining gRNA (site NTGA-1), the homebrew one-pot and EnGen reactions containing 1x target-specific and common oligos failed to reach completion (with homebrew one-pot outperforming EnGen; **Supplementary Fig. 17d**). The addition of 3x both the target-specific and common oligos to either method rescued gRNA yield to sufficient quantities to permit SpRYgest reactions proceeding to completion. Furthermore, to identify methods that make gRNA IVT more cost-effective, we also tested scaling down the reagents used by our homebrew one-pot method by 1/4. The scaled-down one-pot reactions generated comparable yields for the four gRNAs per µL of transcription reaction relative to the full-scale one-pot (**Supplementary Fig. 17c**), and was also effective when used directly in SpRYgest reactions resulting in complete substrate digestion (**Fig. 4b**).

Thus, we recommend that gRNA synthesis reactions be performed with our optimized one-pot reactions, either containing Klenow Fragment and RiboMAX T7 Express Enzyme Mix with 3x template oligos (either at full-scale

or scaled-down), or containing the EnGen kit with 3x template oligos. We calculated the enzymatic cost for each gRNA synthesis method, based on current vendor list-prices, to be:

(a) One-pot standard, $8.68 per gRNA from:

- Klenow Fragment (NEB, M0212L), $0.63 per gRNA ($252 / 1000 U; 2.5 U / gRNA)
- T7 RiboMAX (Promega, P1320), $8.68 per gRNA ($434 / 50 reactions)

(b) One-pot scaled-down, $2.33 per gRNA from:

- Klenow Fragment (NEB, M0212L), $0.16 per gRNA ($252 / 1000 U; 0.625 U / gRNA)
- T7 RiboMAX (Promega, P1320), $2.17 per gRNA ($434 / 200 reactions)

(c) EnGen kit, $20.80 per gRNA from:

- EnGen sgRNA Synthesis Kit (NEB, E3322S), $416 / 20 reactions

Given the flexibility of modifying components in the one-pot gRNA synthesis reactions[18], we anticipate that further increases in gRNA yield along with reductions in cost are achievable. Furthermore, sufficient yields of most gRNAs can likely be generated with one-pot incubation times shorter than 4 hours (**Supplementary Fig. 15b**). Although we did not formally test scaling down the EnGen kit, it's possible that similar optimizations could be performed to comparatively reduce the cost per gRNA. Finally, most institutions have quotes from vendors that would further reduce these costs.

Supplementary Note 7: Design considerations and recommendations for SpRYgests

In human cells we previously observed a preference by SpRY to target sites with NRN PAMs more efficiently than those with NYN PAMs[6] (where R is A or G and Y is C or T). However, *in vitro* SpRY is essentially PAMless and can efficiently cleave sites with any PAM (**Figs. 1e**-**1h**, and **Supplementary Fig. 5c**). Therefore, when designing gRNAs to perform a SpRYgest, we recommend focusing solely on the exact position at which the DSB is to be introduced. Assuming that SpRY generates a blunt DSB 3 bp upstream of the PAM, gRNAs should be designed to position the intended DSB between position 3 and 4 of the guide sequence, counting from the PAM proximal end (**Supplementary Fig. 6a**). Finally, for each intended DSB two potential gRNAs can be designed, each targeting a different strand. Thus, two distinct gRNAs can be designed for each cut site, allowing the selection of the most efficient gRNA for DSB generation, and potentially avoiding rare off-target edits against closely matched sites (which can be identified and avoiding by using SpOT-check).

Once the gRNA target sites are identified, the next step is to design target-specific ~ 54 bp oligos required to generate the dsDNA template for IVT (see **Supplementary Note 3**). This workflow enables the rapid synthesis of new gRNAs and allows custom SpRYgest of any DNA sequence. Using a fixed amount of the SpRY and gRNA ribonucleoprotein (RNP) complex (with gRNAs generated as described in **Supplementary Notes 3** and **6**), we were able to introduce all intended DSBs with at least >78% cleavage efficiency for all unique sites that we sought to generate DSBs (**Fig. 4c**). Since it is trivial to include additional SpRY RNP in the molecular reaction, we envision that the total cleavage percentage of nearly any substrate could reach completion with minor alterations to the reaction setup.

To SpRYgest 800 fmol of a supercoiled plasmid that is ~8,000 bp (approximately 4 µg of DNA), we recommend the following starting conditions that can be appropriately modified: SpRY protein at a final concentration of 1 µM and IVT gRNA (prepared without DNase treatment; see **Supplementary Note 3**) at a final concentration of 2 µM in Buffer 3.1 (NEB), and allowing the reaction to proceed at 37 °C for approximately 3 hours. The digestion timeline can potentially be shortened for efficient sites, and when using higher concentrations of SpRY. Note, if generating gRNAs via the one-pot method, it is also possible to add 5 µL of unpurified and unquantified gRNA IVT reaction directly into the SpRYgest reaction mixture (which generally will lead to the inclusion of a large excess of the gRNA). If incomplete SpRYgestion is observed, we recommend performing a clean-up of the gRNA preparation, proceeding with quantification to ensure that adequate gRNA is being generated, and performing a transient heating and cooling step to denature and fold the gRNA . Finally, the timing of gRNA synthesis via the one-pot method can be varied and reduced to as little as 1 hour for most gRNAs (**Supplementary Fig. 17b** and **17c**).

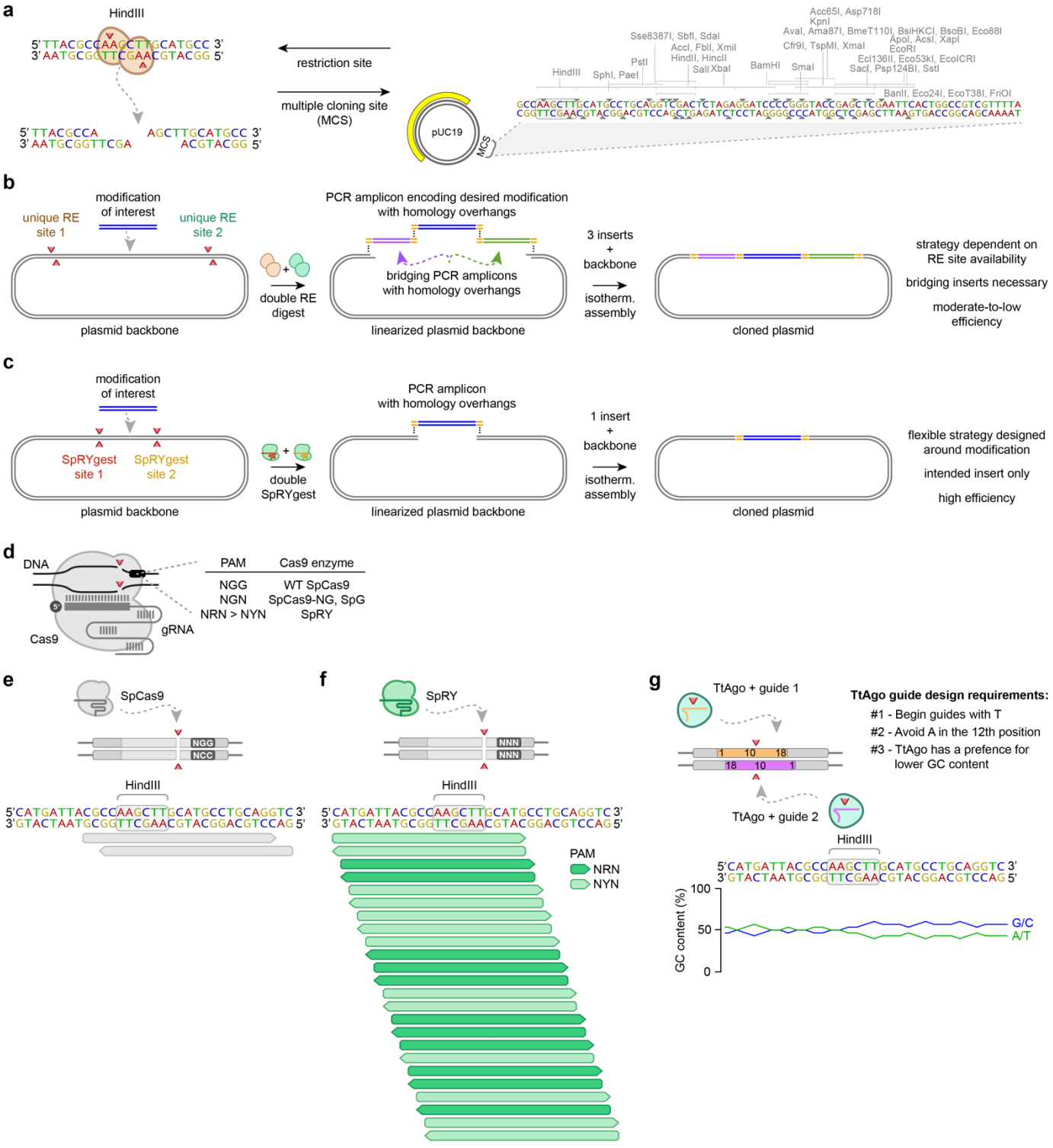
Supplementary Note 8: Annotation of substrate off-targets via SpOT-check

To identify potential gRNAs with potential closely matched off-target sites in prospective DNA substrates, we developed software called SpOT-check (SpRYgest off-target checker), which is freely available online at www.kleinstiverlab.org/spotcheck. Beyond enumerating the number and sequences of closely matched sites in the substrate, SpOT-check can also be used to design gRNA sequences to create a cut at a designated position (for both top and bottom strand, with oligos needed for IVT as output) in a provided plasmid sequence. Alternatively, pre-designed guide sequences can be used as the input. For each gRNA sequence, SpOT-check reports any potential off-target sites up to off-by-6 mismatches, permitting users to prioritize gRNAs targeting the strand with fewer potential off-targets. SpOT-check "batch mode" will also be available for download from GitHub, allowing users to input multiple guides and plasmid sequences at once in CSV format as well as specifying any maximum number of mismatches for which to report off-targets. Using SpOT-check, we computed the off-target profiles of all gRNAs used in this study (**Supplementary Table 7**).

Supplementary Note 9: Script to calculate SpRYgest rate constants

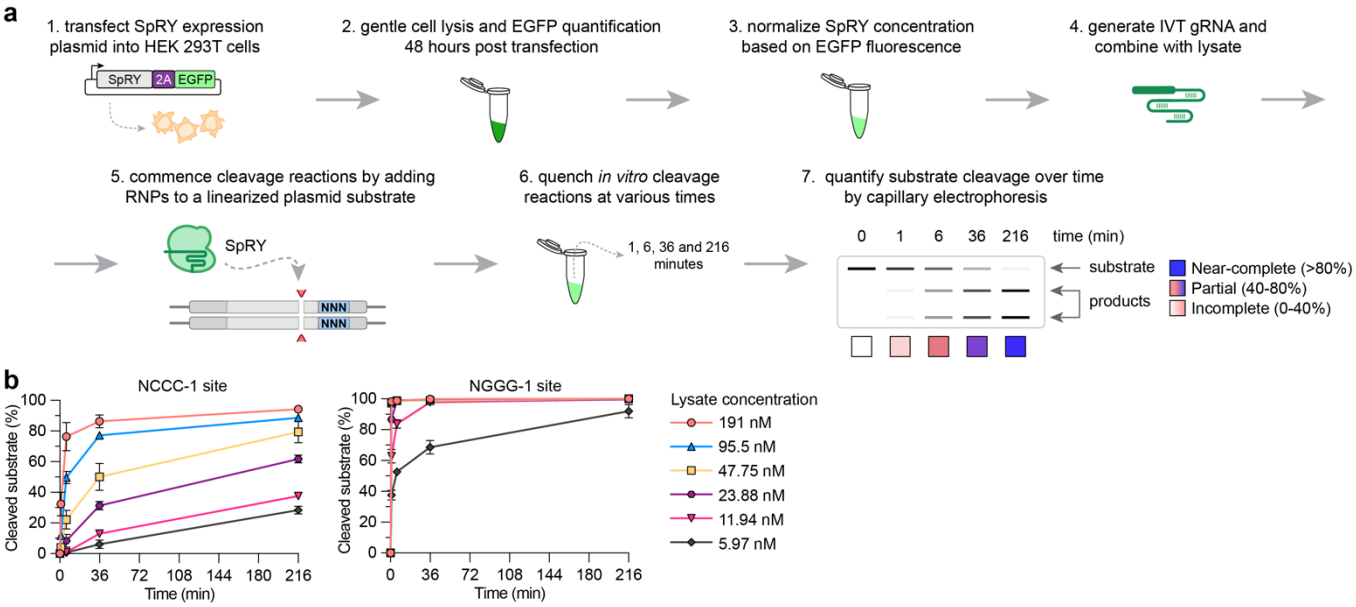
```
import numpy as np
import pandas as pd
from scipy.optimize import curve_fit


timepoints = [0, 1, 6, 36, 216]
df = pd.read_csv('depletion.csv')
"""
enter timepoints (i.e. min or sec) in list above
enter depletion values (% remaining substrate) in rows of csv file
"""


def func(x, a, b):
	return a * np.exp(-b * x)

def exp_decay_fit(timepoints, depletion_df,
                              init_rate_est=[0.0001, 0.001, 0.01]):
	"""
	timepoints: list of timepoints
	depletion: list of depletion values corresponding to the timepoints
	init_rate_est: initial rates estimates to iterate over in nonlinear fitting

	return: list of rate constants
	"""
	ks = []

	for index, row in df.iterrows():

		depletion = list(row)
		print(depletion)
		if len(depletion)==0:
			break
		# map initial conditions to fit error
		min_search = []

		for j in init_rate_est:
			p0 = [100, j]
			# nonlinear curve fit
			try:
				popt, pcov = curve_fit(func, timepoints, depletion, p0=p0)
				# prediction from outputted paramters
				pred = [func(x, popt[0], popt[1]) for x in timepoints]
				# error
				error = sum([(x[0] - x[1]) ** 2 for x in zip(pred, depletion)])
				# append into dictionary mapping initial conditions to fit error
				min_search.append([error, list(popt)])
			except:
				print('except')
				continue
		if len(min_search) != 0:
			ks.append(sorted(min_search)[0][1][1])
		else:
			ks.append('NaN')
	return ks


ks = exp_decay_fit(timepoints, df)
df['rates']=ks
df.to_csv('rates.csv')
```

Supplementary Figures:

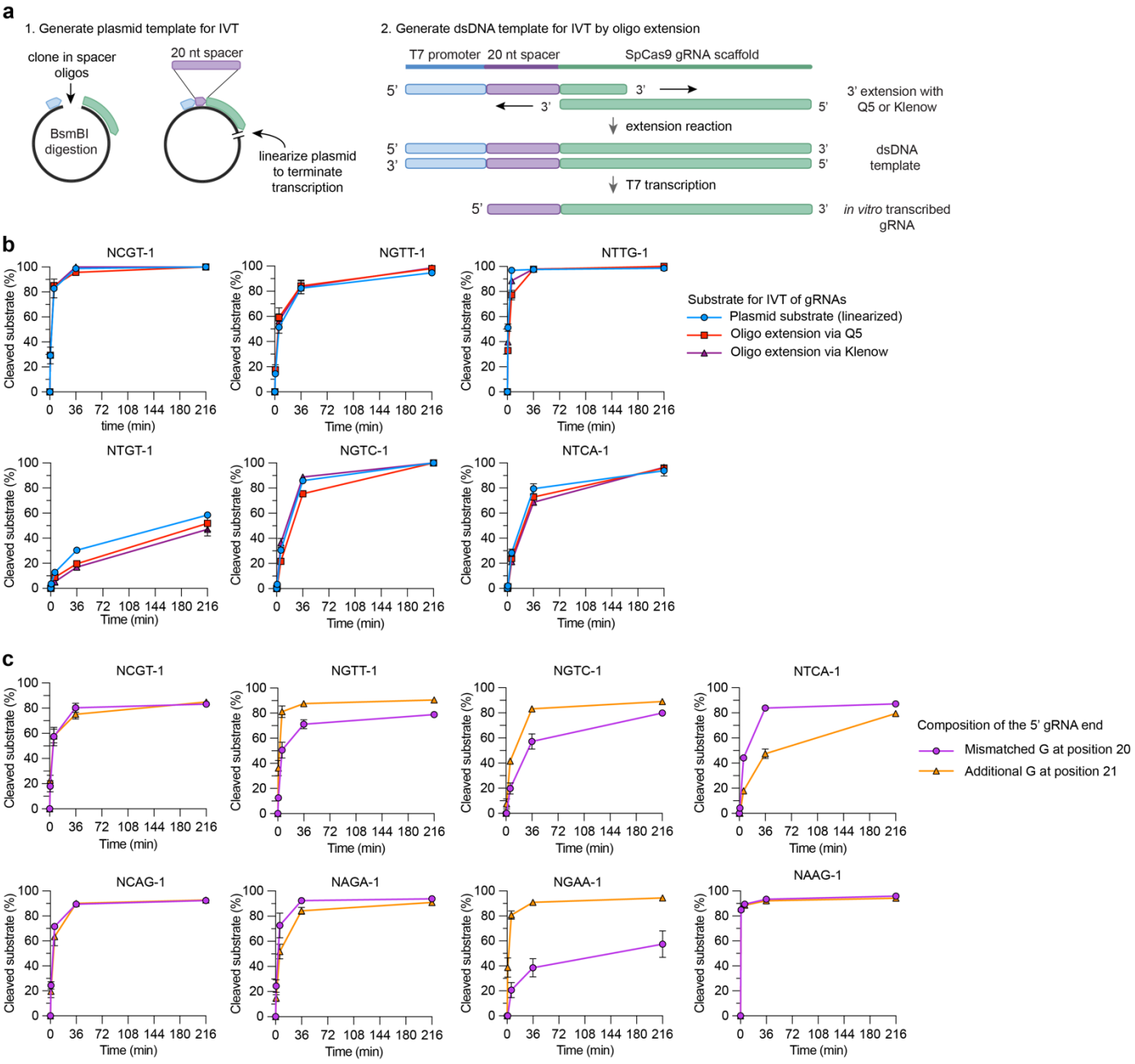


Supplementary Figure 1: Comparison of DNA-targeting endonucleases. (**a**) Illustration of target recognition and cleavage by HindIII (left panel) and of the multiple cloning site (MCS) in pUC19 with any single cutter restriction enzyme shown, including HindIII (right panel). (**b**,**c**) Schematics of isothermal assembly reactions. In **panel b**, the lack of proximal unique restriction sites necessitates the use of additional bridging PCR amplicons in the assembly mixture, which increases the complexity of the reaction while decreasing the success rate. The

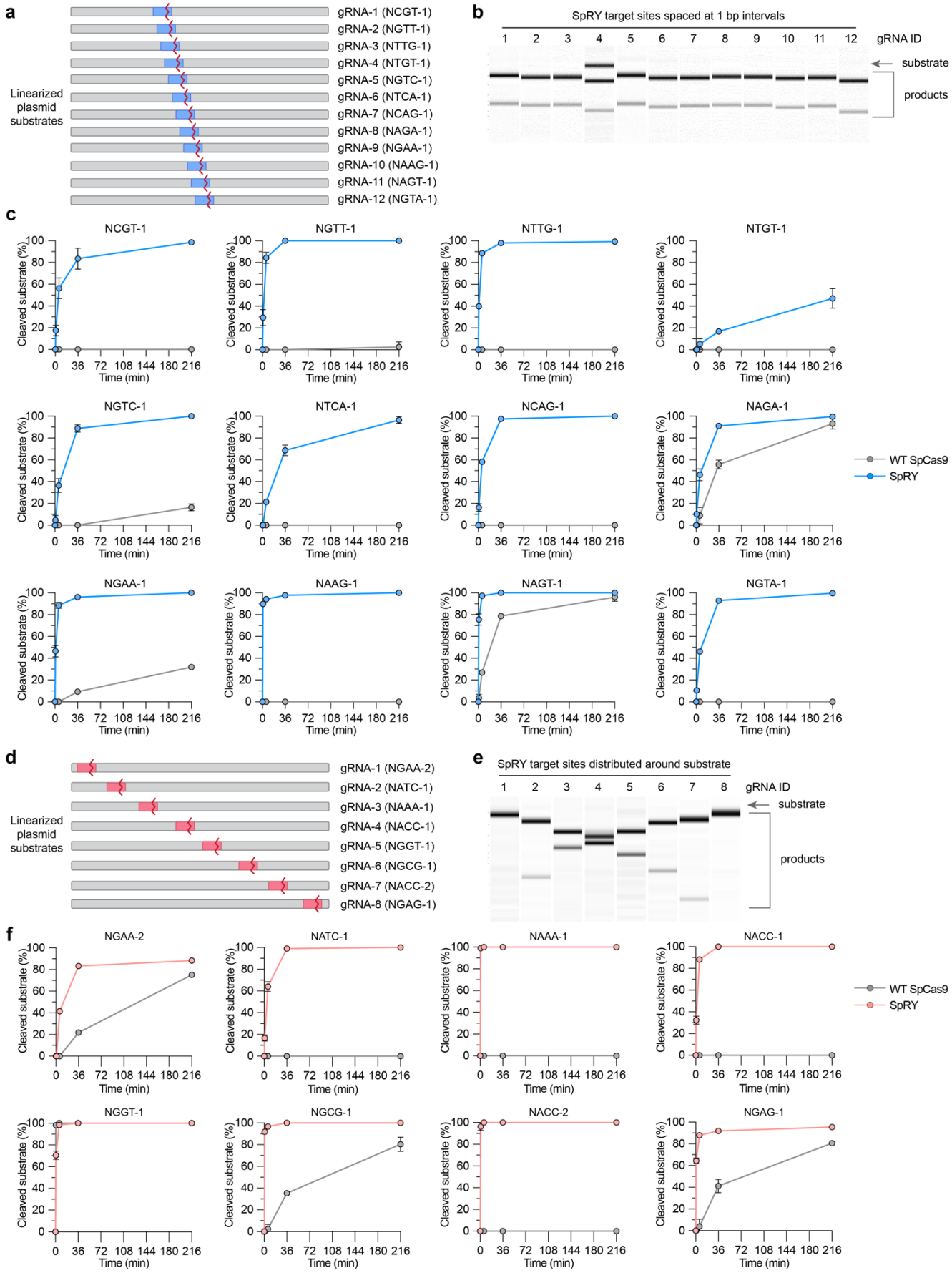
isothermal assembly reaction mixture contains the linearized plasmid backbone, the PCR amplicon encoding the modification of interest, the bridging PCR amplicons, and an enzyme mixture (exonuclease, polymerase, and ligase). In **panel c**, cleaving the plasmid backbone directly adjacent to the modification of interest causes the isothermal assembly reaction to require only a single PCR amplicon in the assembly mixture, which decreases the complexity and improves the success rate of the reaction. The isothermal assembly reaction mixture contains the linearized plasmid backbone, the PCR amplicon encoding the modification of interest, and the enzyme mixture. (**d**) Schematic of SpCas9 complexed with a gRNA and bound to a target DNA sequence bearing various PAMs, depending on the SpCas9 variant. gRNA, guide RNA; PAM, protospacer-adjacent motif. (**e**) The PAM requirement of wild-type (WT) SpCas9 restricts target site design and DNA cleavage to sites harboring NGG PAMs, limiting precise DNA break positioning in a DNA molecule. (**f**) In contrast to WT SpCas9, the near-PAMless SpRY variant might offer unconstrained targeting for *in vitro* DNA digests. For **panels e** and **f**, target sites (20 nt spacer with 3 nt PAM) are shown that target the top and bottom strand (with right and left facing points, respectively). (**g**) Schematic of guide design requirements to potentially generate a double strand break with TtAgo.



Supplementary Figure 2: Generation of SpRY protein from human cell lysate for SpRYgest cleavage reactions. (**a**) Schematic of SpRY production and SpRYgest workflow. SpRY protein is expressed in human cells and harvested by gentle lysis, with SpRY concentrations normalized by EGFP fluorescence versus a fluorescein standard curve. The SpRY protein from the lysate is complexed with a gRNA (produced by *in vitro* transcription or via chemical synthesis), and the resulting RNPs are combined with a linearized plasmid template to initiate *in vitro* cleavage reactions. Aliquots from the reaction are extracted at various timepoints, terminated, and the cleavage of DNA substrates over time is analyzed by capillary electrophoresis. (**b**) Titration of normalized lysate to vary the amount of SpRY protein as input into *in vitro* cleavage reactions, performed utilizing gRNAs targeted to NCCC-1 or NGGG-1 sites (left and right panels, respectively). Unless otherwise noted, all other small-scale SpRYgests using linearized plasmid substrates were performed using 180 nM fluorescein-normalized SpRY lysate. The cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3.



Supplementary Figure 3: gRNA production via *in vitro* transcription. (**a**) Illustration of three different methods to produce gRNAs via *in vitro* transcription (IVT), using templates generated either by (1) cloning a dsDNA plasmid template with a T7 promoter, the spacer sequence, and gRNA scaffold (left panel), or (2) extension of overlapping ssDNA oligonucleotides using either Q5 polymerase or Klenow (right panel). (**b**) Comparison of the effectiveness of SpRYgests when using gRNAs produced via IVT from the three different DNA substrates shown in **panel a** (run-off IVT from a cloned plasmid, or from a dsDNA fragment generated by oligonucleotide extension with Q5 or Klenow). (**c**) Comparison of the effectiveness of SpRYgests when using gRNAs transcribed with 20 or 21 nt spacers (containing a matched or mismatched 5'G to improve transcription from the T7 promoter). For **panels b** and **c**, the cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3.



Supplementary Figure 4: Initial SpRYgest cleavage experiments on 20 sites. (**a**) Schematic of the 12 gRNAs in a specific region of the plasmid that shift by 1 bp intervals. (**b**) Visualization of the cleavage products of

SpRYgest reactions at the final timepoint (216 minutes) when using the 12 1-nt interval gRNAs and lysates containing SpRY, run on a QIAxcel Fast Analysis Kit (Qiagen) and visualized using QIAxcel ScreenGel Software (v1.5.0.16, Qiagen). Representative image of a single replicate from n = 3 in **panel c**. (**c**) Plots of substrate cleavage over time with the 12 1bp-spaced gRNAs in the presence of lysates containing WT SpCas9 or SpRY. Cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3. (**d**) Schematic of the 8 gRNAs distributed at various positions in the substrate. (**e**) Visualization of the cleavage products of SpRYgest reactions at the final timepoint (216 minutes) when using the 8 distributed gRNAs and lysates containing SpRY, analyzed as described in **panel b**. Representative image of a single replicate from n = 3 in **panel f**. (**f**) Plots of substrate cleavage over time with the 8 distributed gRNAs in the presence of lysates containing WT SpCas9 or SpRY. Cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3.

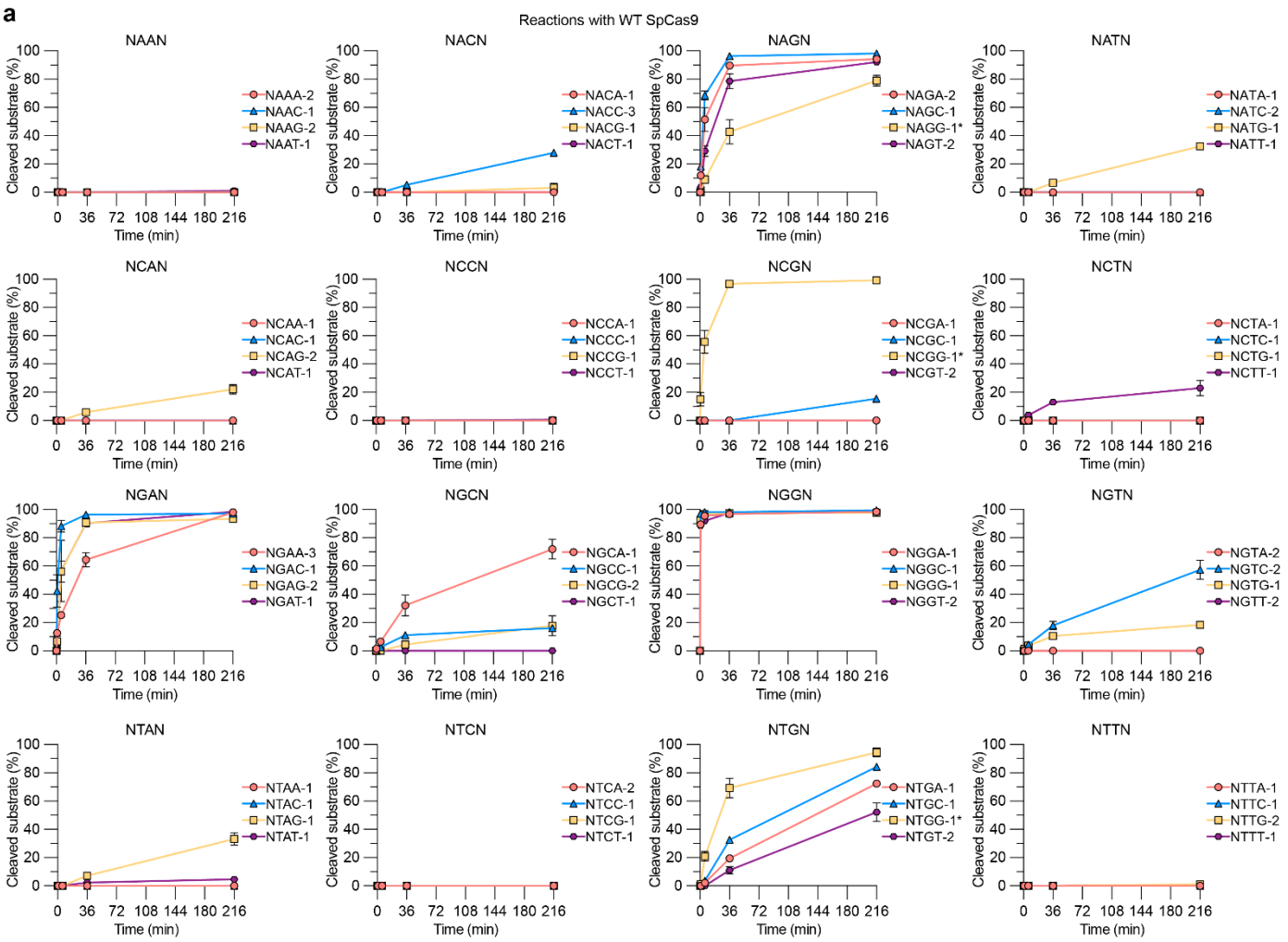


Figure 5a *(legend follows)*

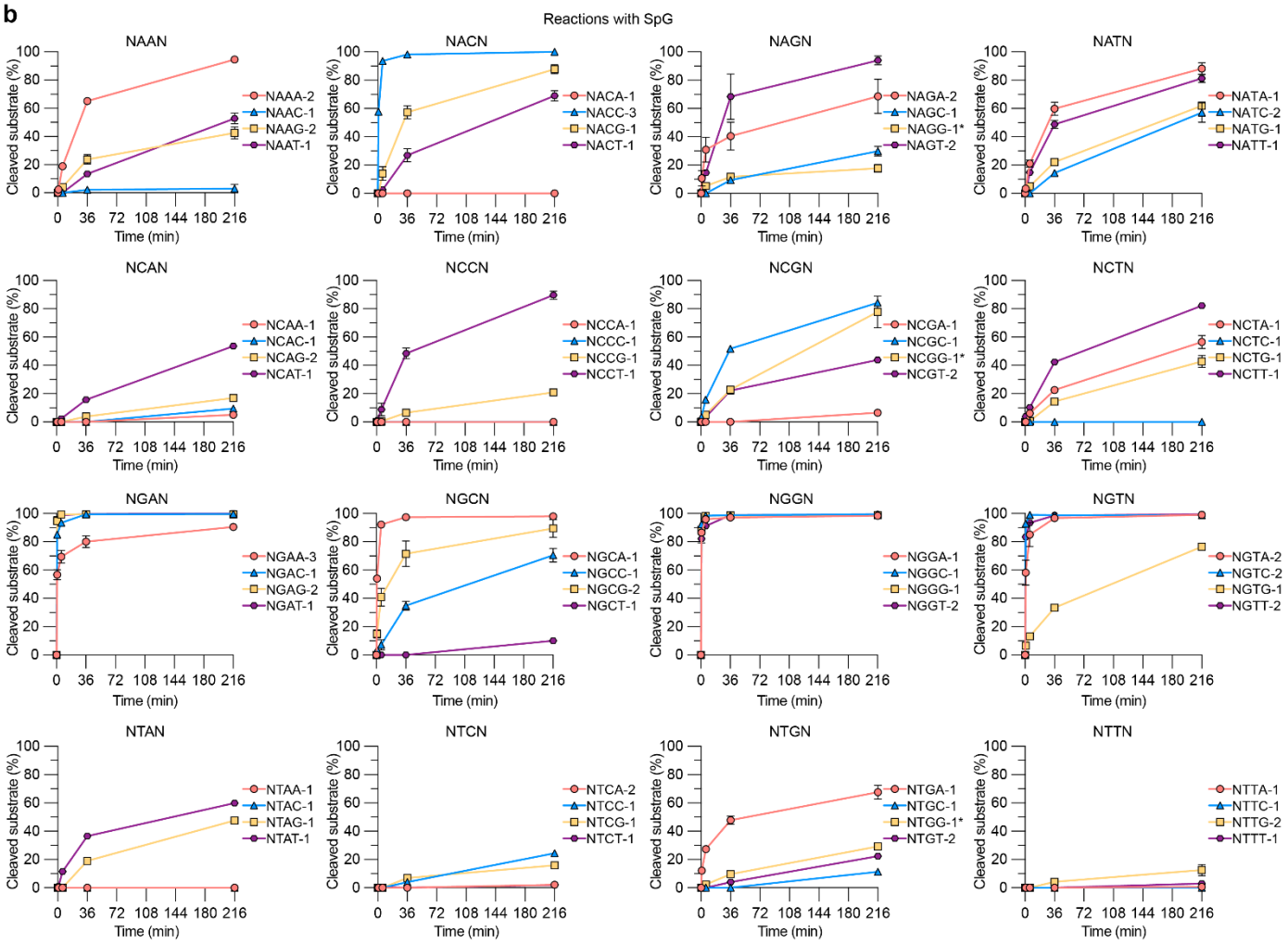
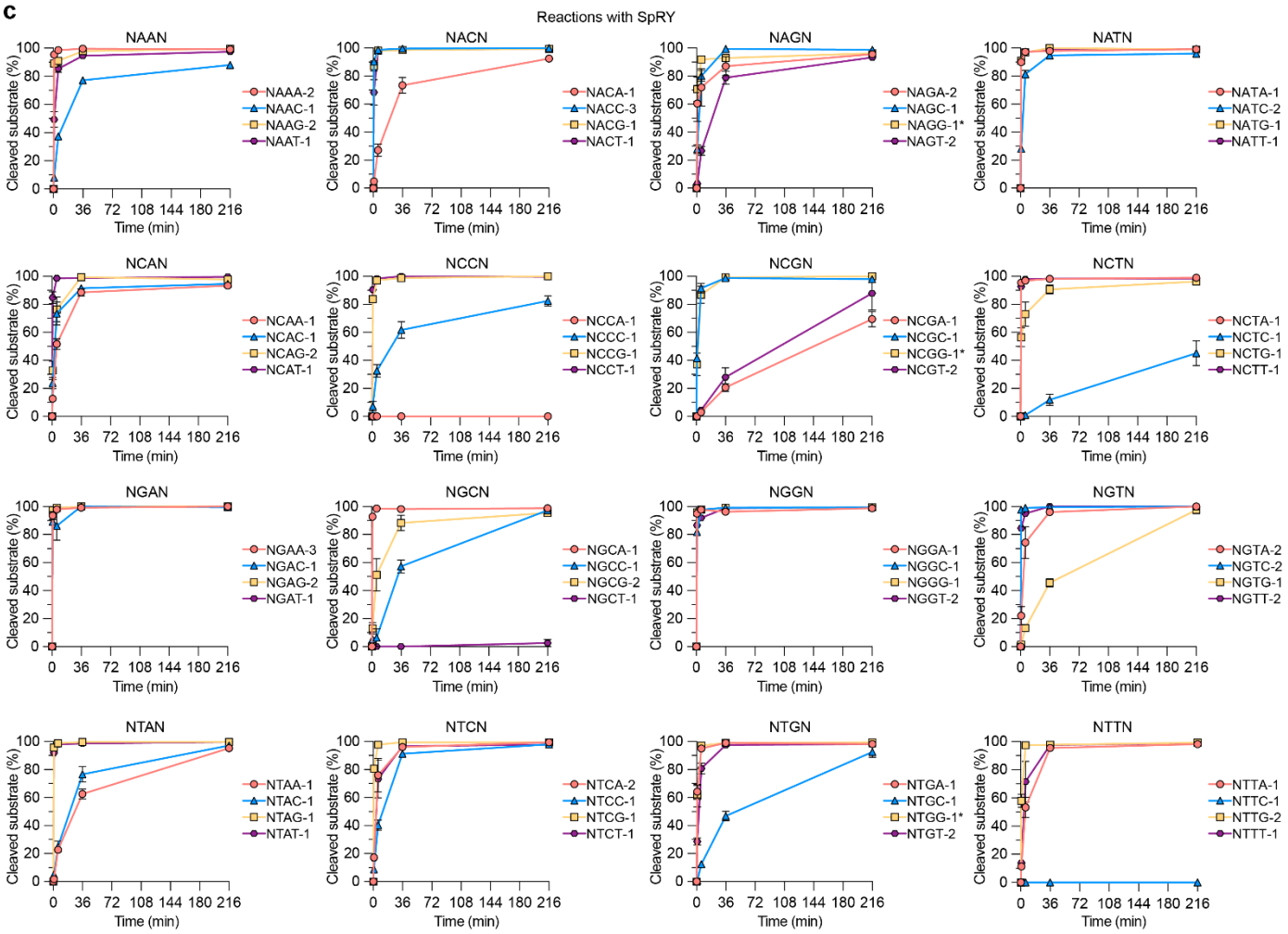
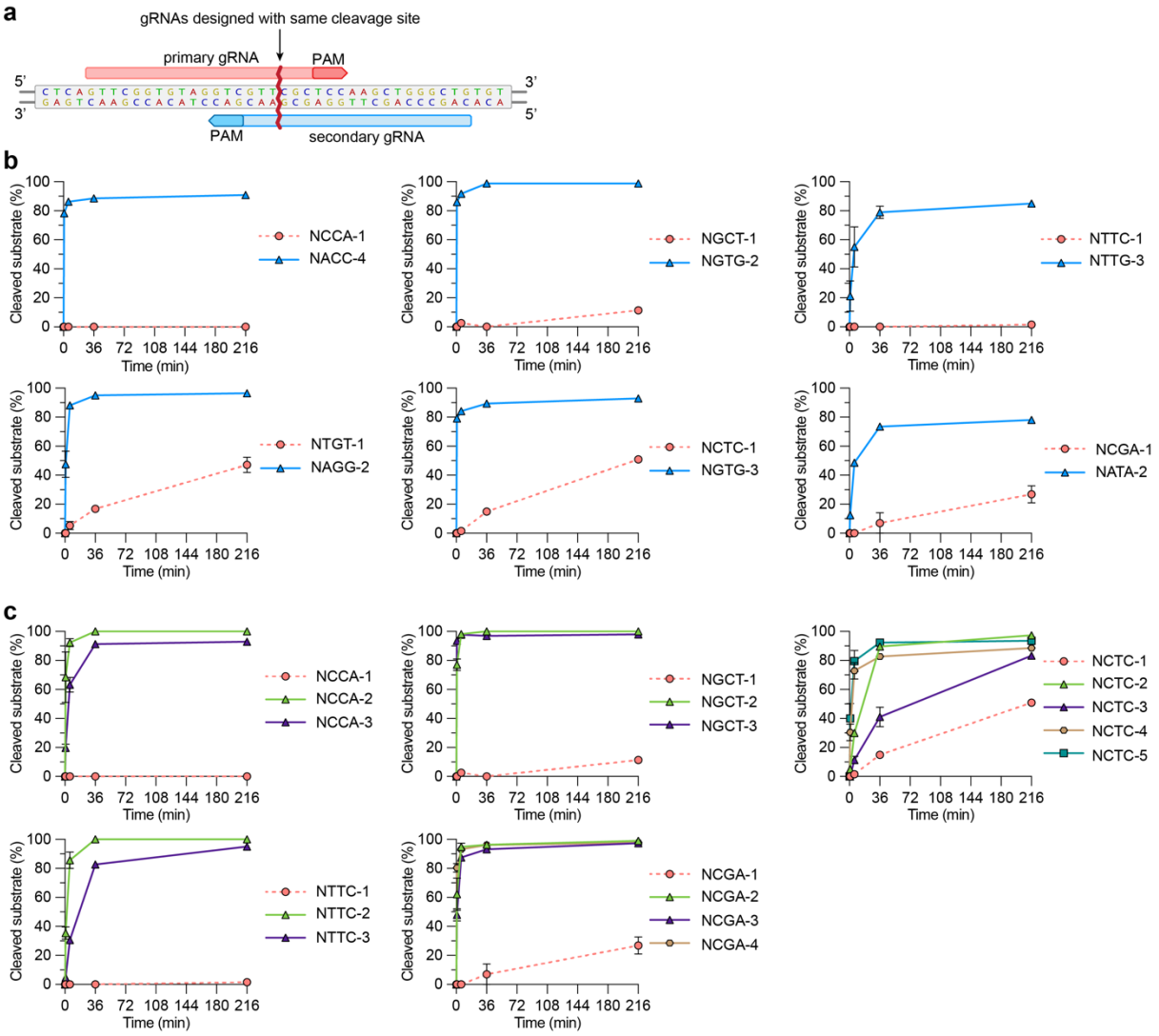


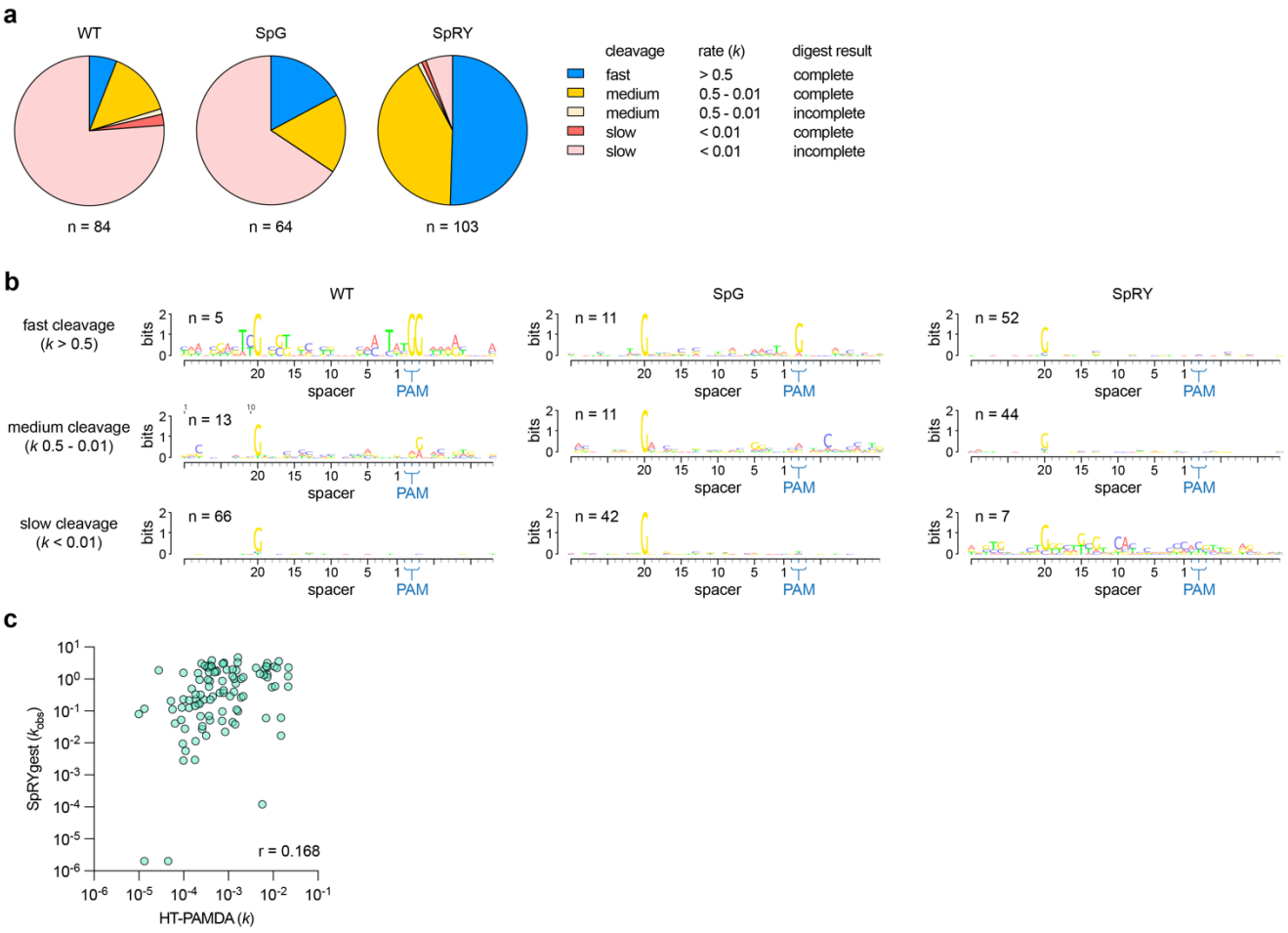
Figure 5b *(legend follows)*



Supplementary Figure 5: Cleavage activity of WT, SpG, and SpRY across 64 sites. (**a**-**c**) Time-course *in vitro* cleavage reactions were performed using wild-type (WT) SpCas9, SpG, or SpRY[6] protein generated from human cell lysates (**panels a**-**c**, respectively). Lysates were separately combined with 64 different gRNAs targeted to sites that sample each possible identity in the 2nd, 3rd, and 4th positions of an NNNN PAM, as well as a linearized plasmid substrate. Cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3.

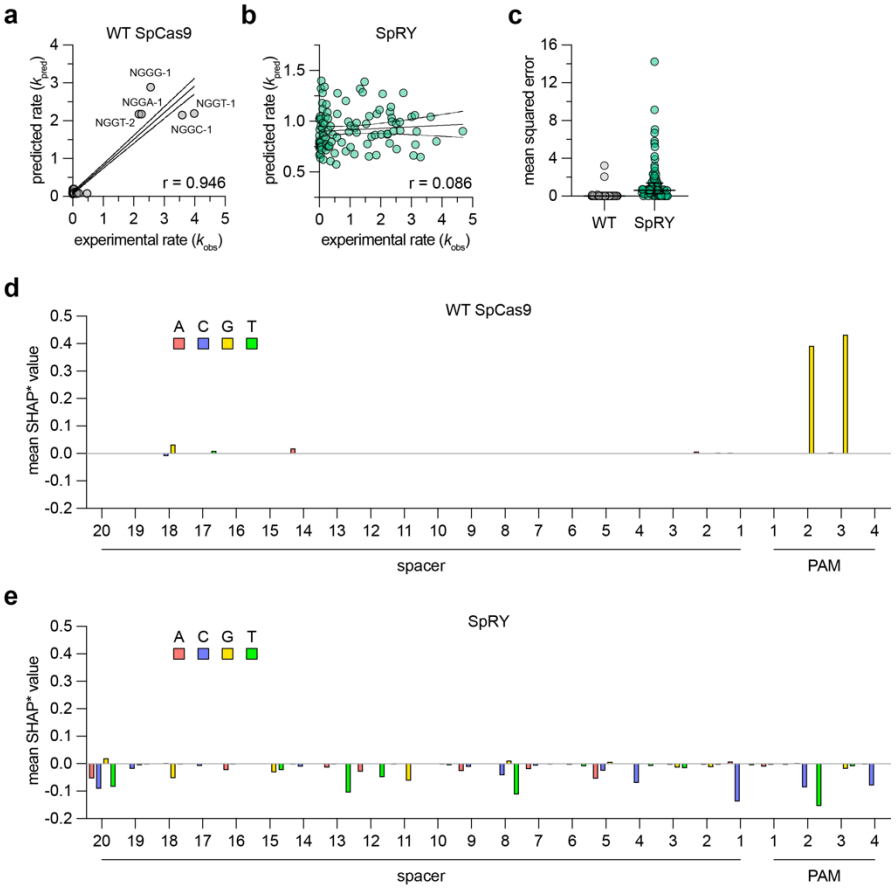


Supplementary Figure 6: Reconciliation of failed SpRYgest sites. (**a**) Schematic of utilizing a secondary gRNA to generate the same DSB as the primary gRNA by targeting the opposite strand, based on SpRY generating a blunt DSB 3 nt from the PAM. (**b**) Comparison of the cleavage efficiencies of primary gRNAs that exhibited partial or incomplete cleavage, and newly designed secondary gRNAs targeted to the opposite strand. (**c**) Cleavage efficiencies of supplementary gRNAs designed to target additional sites with the same PAM as the primary gRNAs that exhibited partial or incomplete cleavage. For **panels b** and **c**, the primary gRNA is indicated using a salmon-colored dashed line with salmon-colored filled circles; cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m. shown for n = 3.



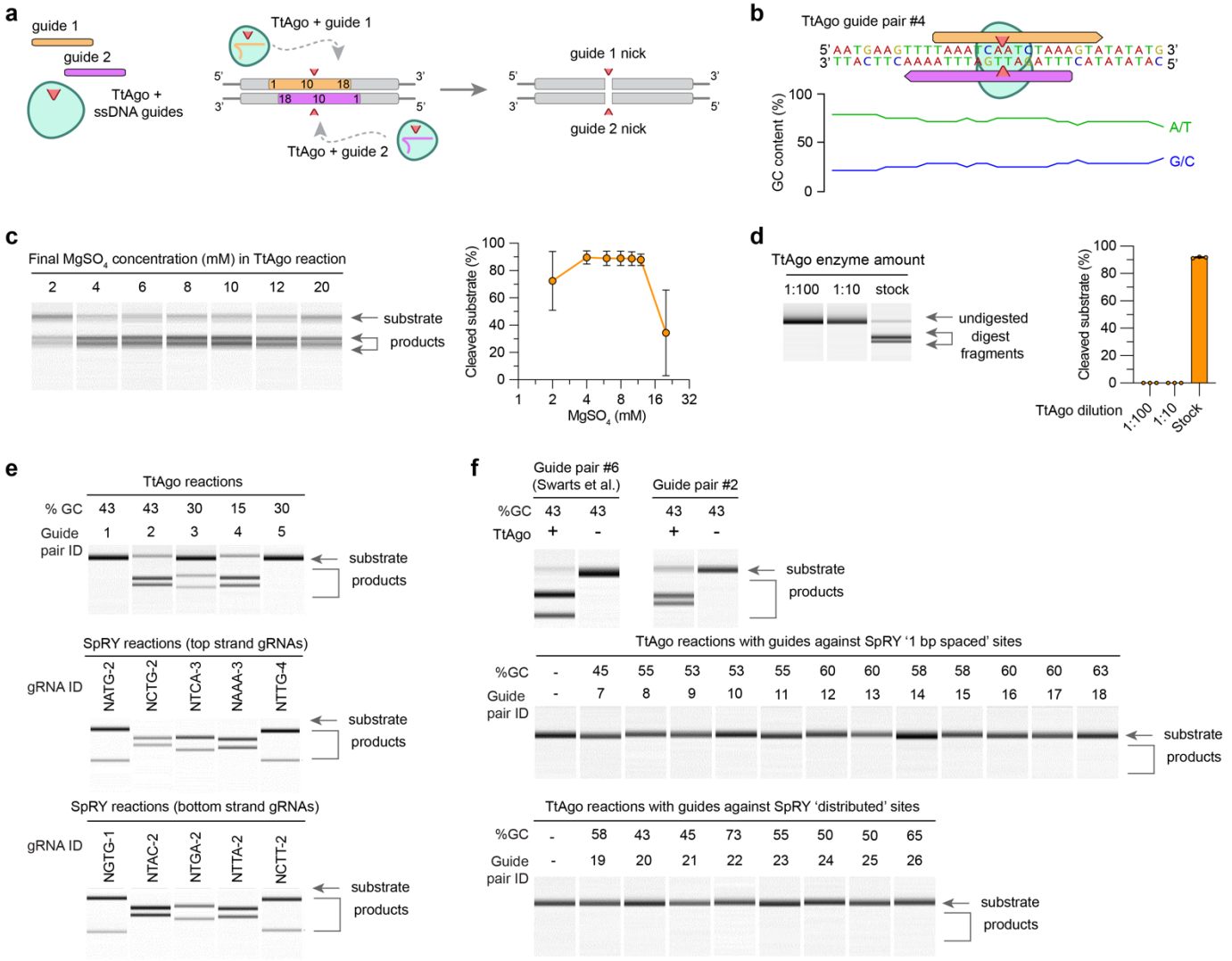
Supplementary Figure 7: Analysis of *in vitro* cleavage reaction rate constants and sequence motifs.

(**a**) Rate constants for *in vitro* cleavage reactions with WT SpCas9, SpG, and SpRY were calculated (**Supplementary Table 1** and **Supplementary Note 9**) and categorized based on the speed of the reaction and degree of reaction completion. (**b**) Sequence logos were generated to explore putative sequence features for each Cas9 variant and category of substrate cleavage speed (logos generated using Geneious Prime; version 2020.1.2). The guanine preference in the 20th position of the spacer is due to the design of target sites that intentionally harbored a guanine in that position, necessary for transcription from the T7 promoter. (**c**) Correlation between the experimental SpRYgest rate constants (k_{obs}) and our previously determined HT-PAMDA data for SpRY (k) against any given PAM[6]. Two SpRYgest k_{obs} datapoints at 2.05e-12 were plotted as 2.05e-6 to be visible on the same chart. The SpRYgest rate constants are the mean of 3 biological replicates; the HT-PAMDA rate constants are the mean of two biological replicates.



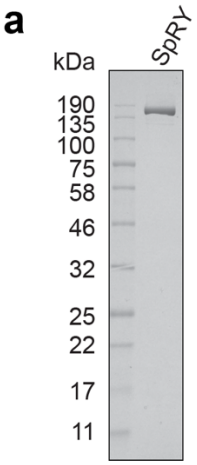
Supplementary Figure 8: Modeling target site features for WT and SpRY cleavage efficiency.

(**a**,**b**) Correlation between the predicted rate constants (k_{pred}) and the experimentally determined rate constants (k_{obs}) from *in vitro* cleavage reactions with WT or SpRY (panels **a** and **b**, respectively). Predicted rate constants were generated using XGBoost[7]. Solid line represents the best-fit line and the dashed lines the 95% confidence interval. Each experimental rate is the mean of 3 biological replicates. (**c**) Mean squared error between the predicted and experimentally determined rate constants for each gRNA/target pair used in the WT and SpRY models. Lines represent the mean and interquartile range. (**d**,**e**) SHapley Additive exPlanations (SHAP) values for nucleotide importance for the prediction of specific outcomes were calculated using the SHAP TreeExplainer[12] for WT and SpRY (panels **d** and **e**, respectively). Positive and negative values indicate beneficial or detrimental nucleotide preferences, respectively.

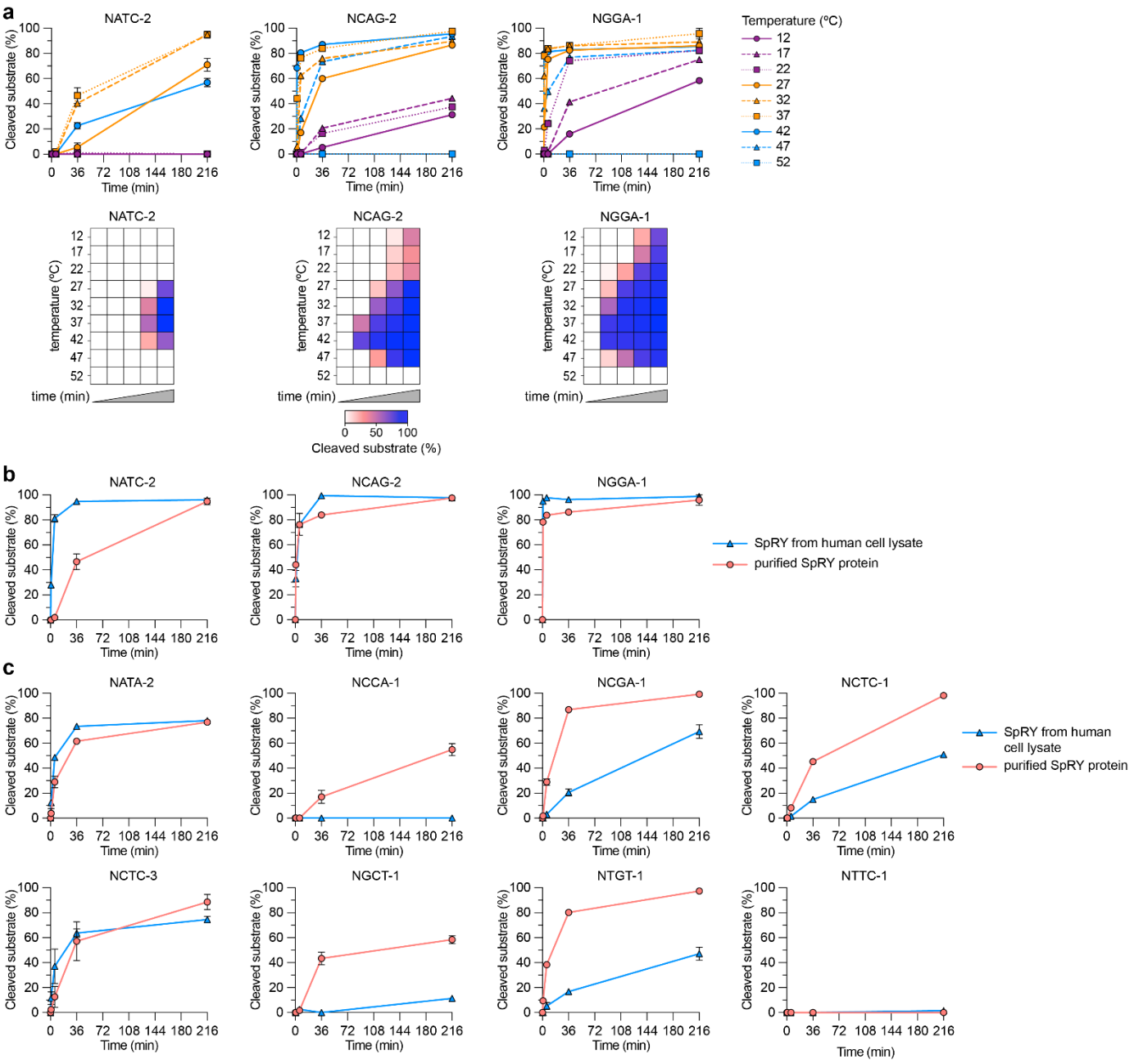


Supplementary Figure 9: Substrate cleavage reactions with TtAgo. (**a**) Schematic of TtAgo generating a DSB when programmed with two separate 5'P-ssDNA guides. (**b**) Schematic of TtAgo guide design for a site that was compatible with the design requirements (TtAgo guide pair #4; see **Fig. 2a** and **Supplementary Table 4**). (**c**) Titration of $MgSO_4$ concentration in TtAgo cleavage reactions using a linearized substrate with a pair of 5'P-ssDNA guides to generate a DSB (guide pair #2; see **Supplementary Table 4**). The left panel is a representative image from one replicate generated and the right panel is a graph showing data from n = 3 with mean and s.e.m. shown. (**d**) Titration of TtAgo enzyme concentration in cleavage reactions performed on a linearized substrate with a pair of 5'P-ssDNA guides (guide pair #2). The left panel is a representative image from one replicate and the right panel is a graph showing data from n = 3 with mean and s.e.m. shown. (**e**) Representative capillary electrophoresis traces comparing final timepoint aliquots of substrate cleavage reactions with TtAgo and SpRY on 5 sites (at 60 and 216 minutes, respectively), as shown in **Fig. 2a**. These 5 sites follow the TtAgo guide design requirements (see **Supplementary Note 5**). (**f**) Representative capillary electrophoresis traces showing final timepoint aliquots of substrate cleavage reactions with TtAgo and pairs of 5'P-ssDNA guides (analyzed after 60 minute reactions). Two guide pairs (with pair #6 from Swarts et al.[15] and

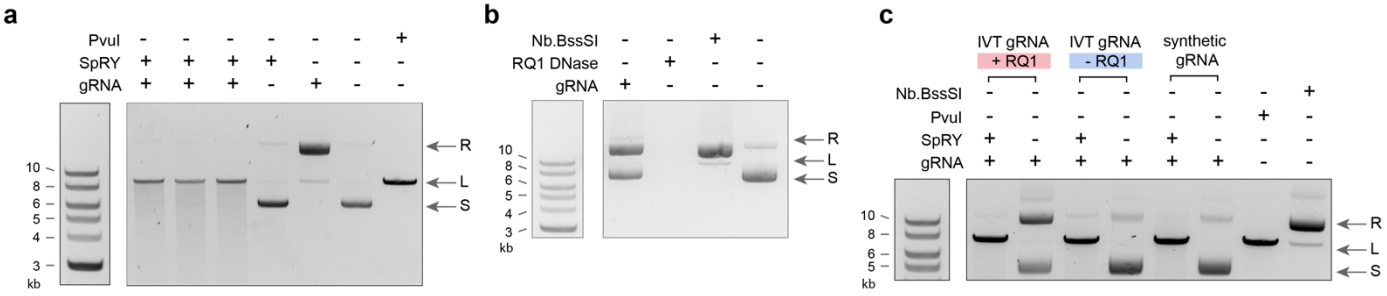
#2 from **Fig. 2a**) were previously shown to result in near-complete cleavage (top panel), and 20 new sites that do not follow the TtAgo guide design requirements (analogous to the SpRYgest sites from **Figs. 1c** and **1d** in the middle and bottom panels, respectively). For **panels c**-**f**, cleavage of DNA substrates was analyzed by capillary electrophoresis using QIAxcel ScreenGel Software v1.5.0.16 (Qiagen); with mean and s.e.m. shown for n = 3 in graphs; representative capillary electrophoresis traces are shown.



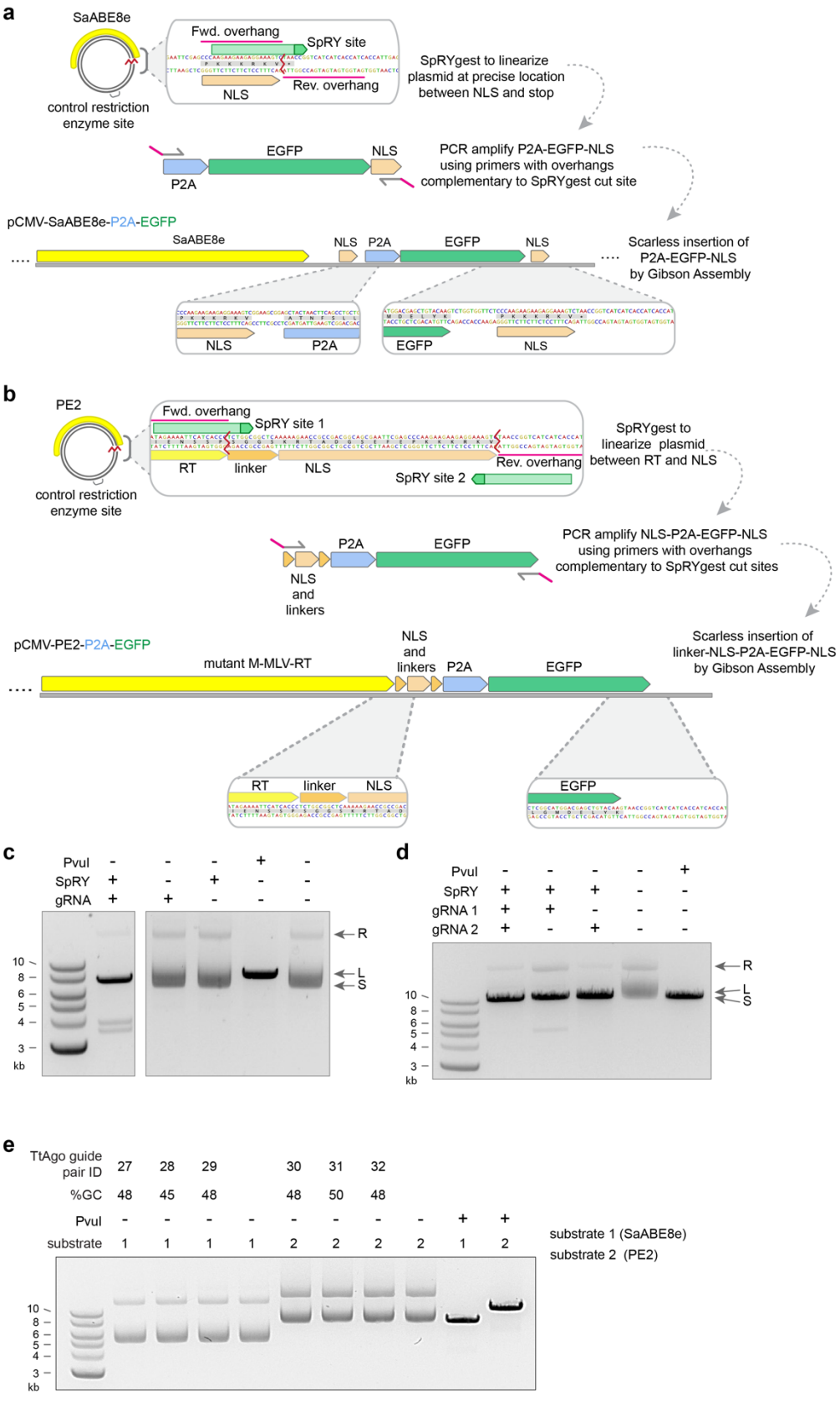
Supplementary Figure 10: Purification of SpRY. (**a**) Tris-glycine gel image of purified SpRY protein run alongside a protein standard (Broad Range; New England Biolabs P7712). Novex 10-20% Tris-glycine gel post stained with SimplyBlue Safe Stain (both purchased from ThermoFisher). The SpRY protein was purified using HiTrap DEAE Sepharose resin, a HisTrap HP column, and a HiTrap heparin HP column, followed by dialysis and a protein concentration step (for additional details, please see the **Methods** section). n = 1.



Supplementary Figure 11: Assessment of SpRYgest parameters. (**a**) Time-course SpRYgests using SpRY expressed from human cell lysates were performed across nine different temperatures, plotted as scatterplots or heatmaps (top and bottom panels, respectively). (**b**,**c**) Plots illustrating the rates of substrate cleavage when using SpRY from human cell lysates or purified SpRY protein, for gRNAs whose reactions with SpRY lysates either went to completion (**panel b**) or exhibited partial or incomplete substrate cleavage (**panel c**). Reactions were performed at 37ºC. For **panels a-c**, cleavage of DNA substrates was analyzed by capillary electrophoresis; mean and s.e.m., shown for n = 3 (s.e.m. not shown for heatmaps).

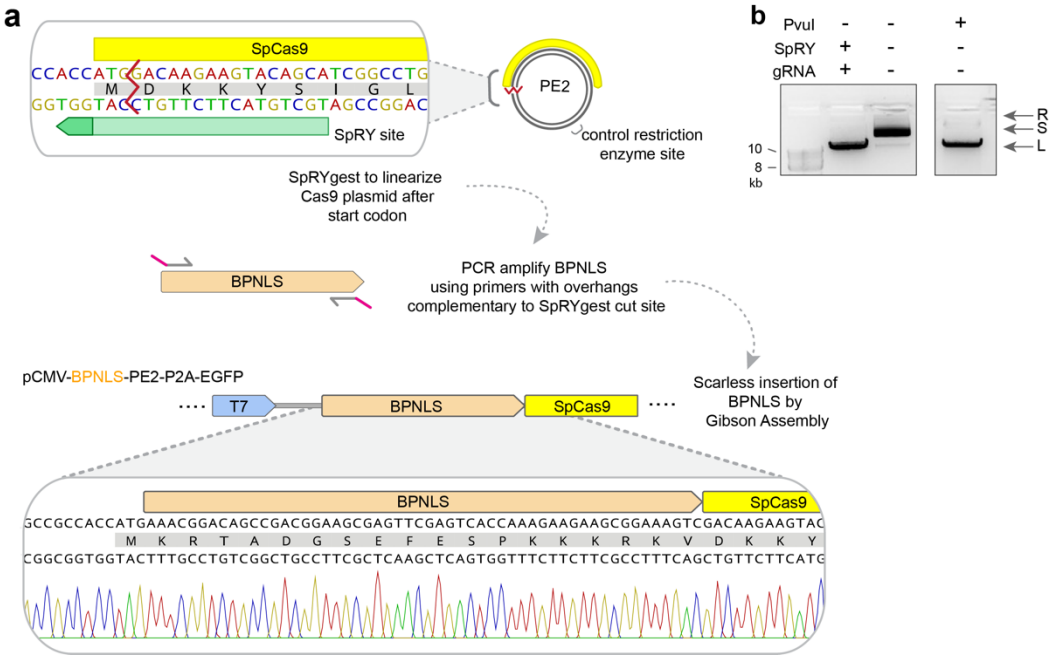


Supplementary Figure 12: DNase contamination during gRNA preparation can cause non-specific degradation of substrate. (**a**) Agarose gel illustrating degradation of plasmid substrates, resulting from a probable carry-forward of the RQ1 DNase present in the gRNA *in vitro* transcription protocol. The gRNA transcription reactions were cleaned up using beads following the RQ1 DNase digestion, and heated to 95 °C for 5 minutes for gRNA refolding. (**b**) Agarose gel illustrating plasmid conformations after incubation with the a gRNA only control, concentrated RQ1 DNase (Promega), or Nb.BssSI nickase (New England Biolabs). (**c**) Agarose gel comparing plasmid conformations after treatment with RNPs formed with *in vitro* transcribed gRNAs generated with and without the RQ1 DNase treatment, and commercial synthetic gRNAs. For all panels, plasmid conformations are: R, relaxed; L, linear; S, supercoiled. Nb.BssSI is a nickase; all agarose gels were stained with ethidium bromide. Gels images from n = 1 experiments. The 1 kb Plus DNA Ladder (NEB, N3200) is shown in kilobase sizes.

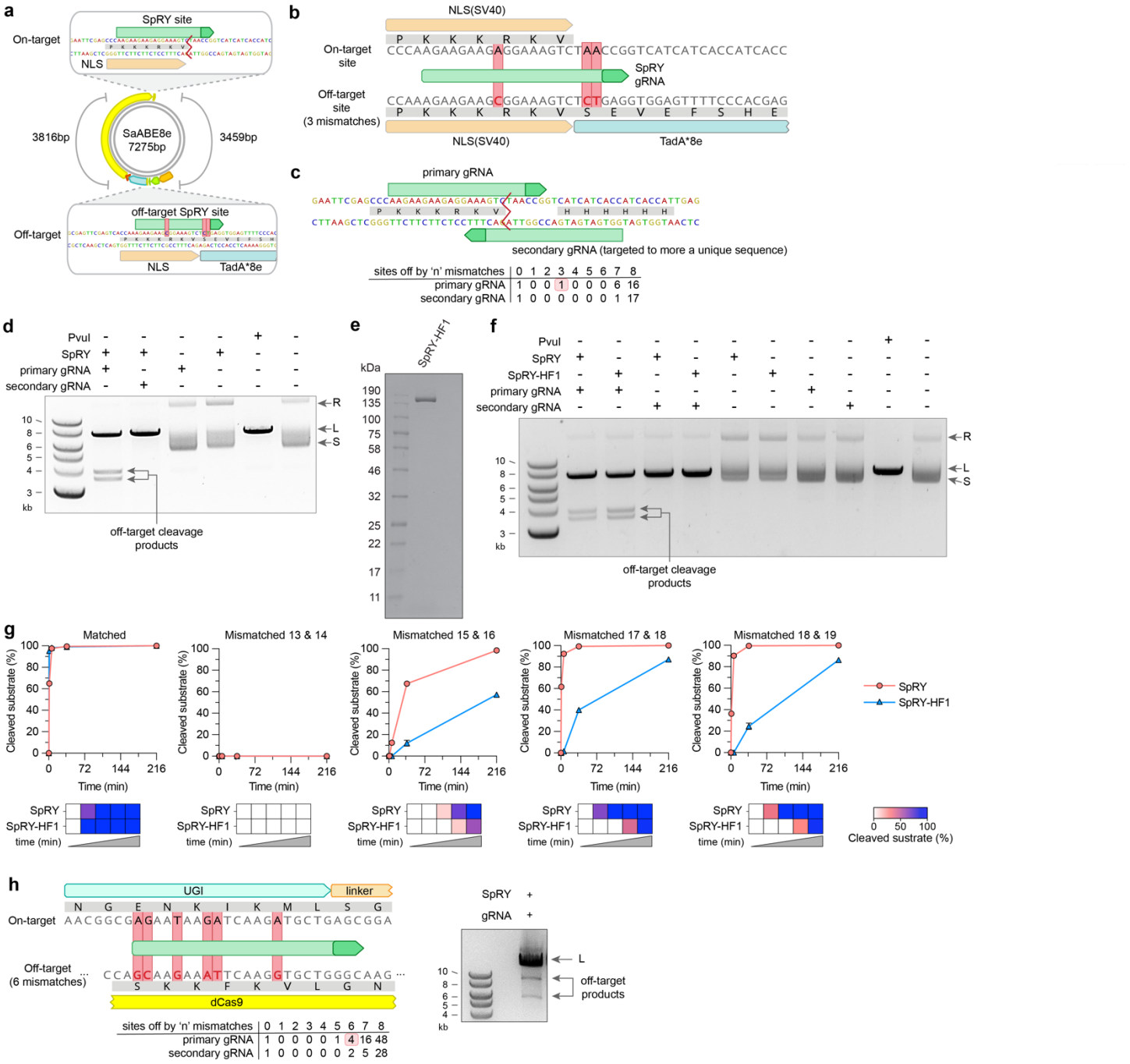


Supplementary Figure 13: Cloning reactions to generate large sequence insertions via SpRYgest. (**a**,**b**) Schematics of the SpRYgest cloning strategies to add P2A-EGFP sequences to SaCas9-ABE8e and PE2 plasmids via single and double SpRYgests, in **panels a** and **b**, respectively. (**c**,**d**) Agarose gel images for the

SpRYgest reactions of SaCas9-ABE8e and PE2 plasmids in **panels c** and **d**, respectively. (**e**) Agarose gel image of TtAgo reactions with pairs of 5'P-ssDNA guides targeted to the same SpRYgest sites in **panels a** and **b**. For **panels c**-**e** plasmid conformations are: R, relaxed; L, linear; S, supercoiled; agarose gels were stained with ethidium bromide; gels images from n = 1 experiments; the 1 kb Plus DNA Ladder (NEB, N3200) is shown in kilobase sizes.

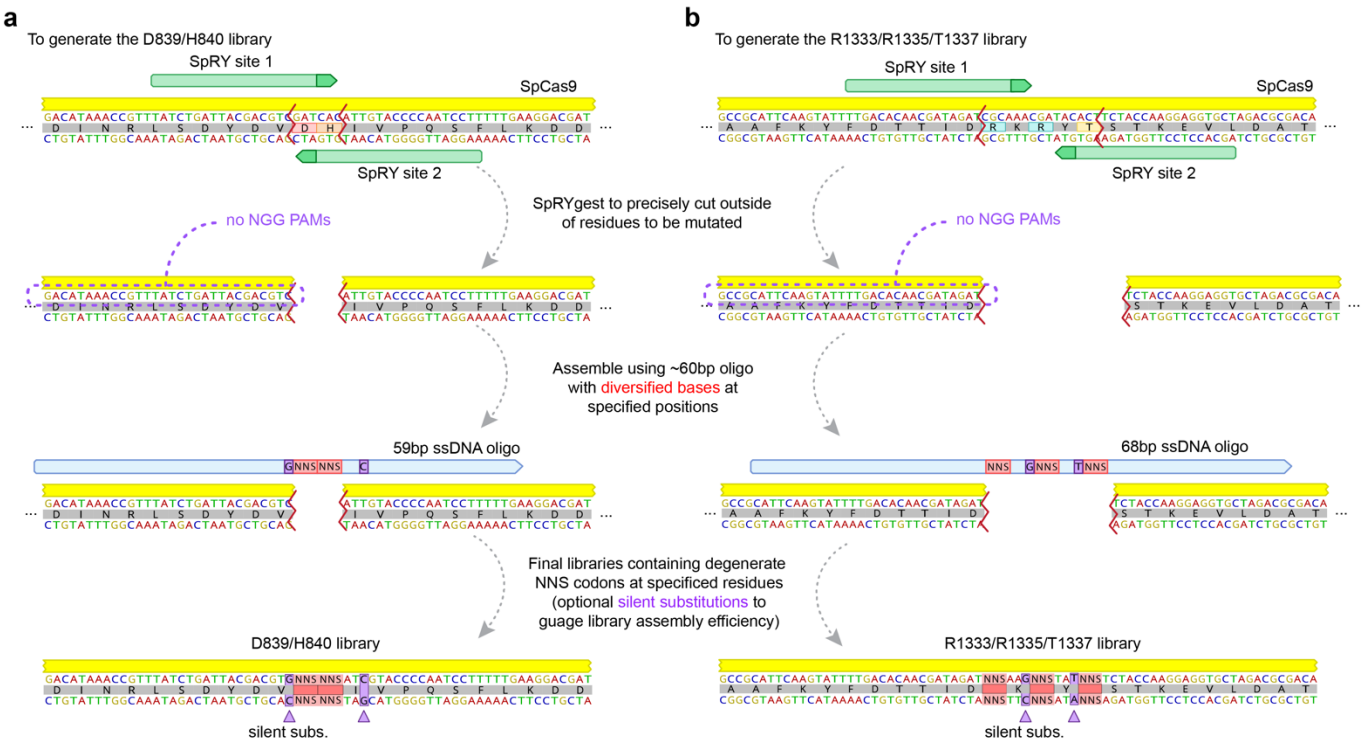


Supplementary Figure 14: Short sequence insertions via SpRYgest. (**a**) Schematic of the SpRYgest strategy to add a short sequence (a BPNLS) to the N-terminus SpCas9 in a human expression plasmid. The section of a representative Sanger sequencing reaction illustrates the cloning junction from a correct clone. (**b**) Agarose gel image for the SpRYgest reaction in **panels a**. Plasmid conformations are: R, relaxed; L, linear; S, supercoiled; agarose gel stained with ethidium bromide; image from n = 1 experiment; the 1 kb Plus DNA Ladder (NEB, N3200) is shown in kilobase sizes.

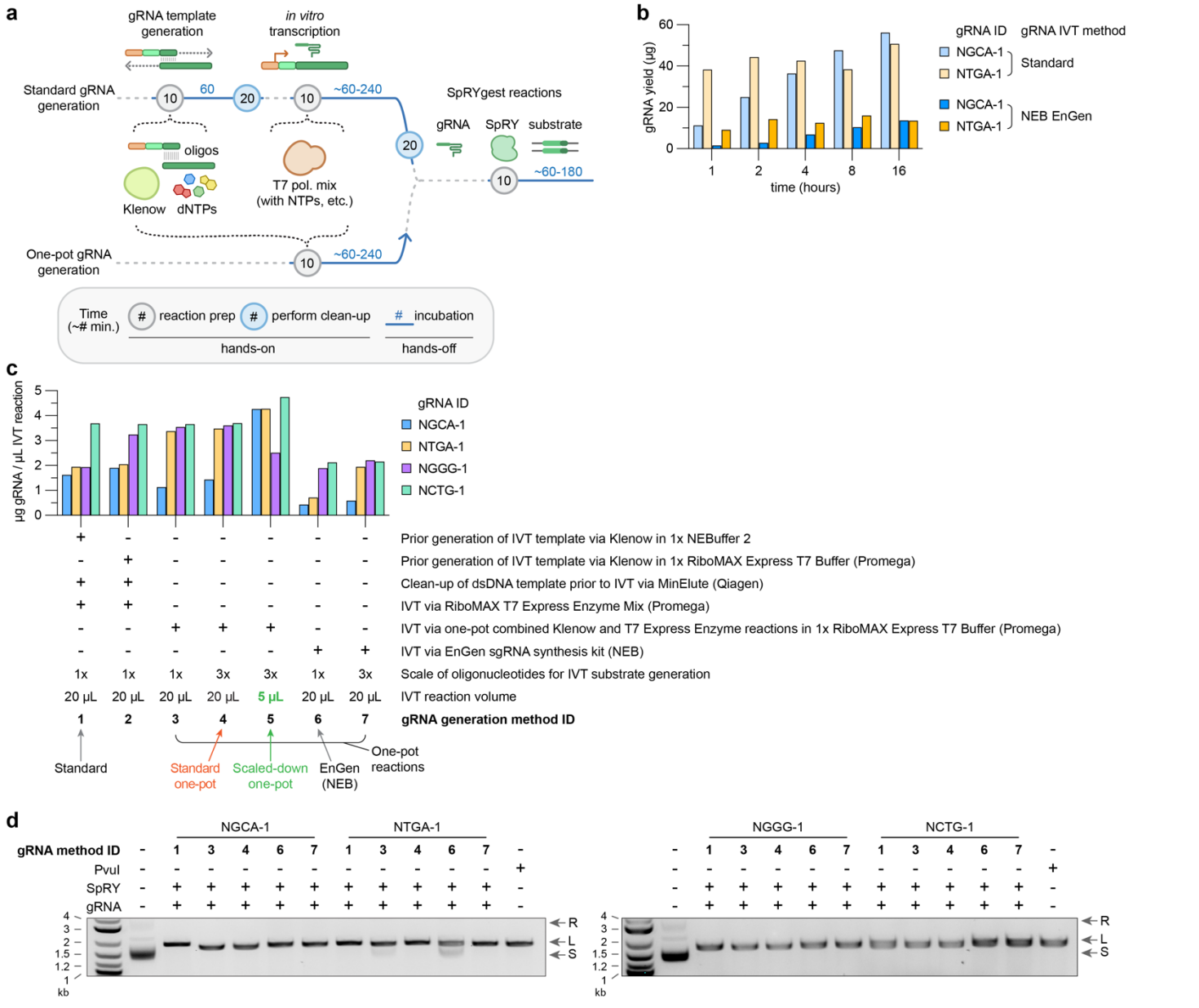


Supplementary Figure 15: Observation and mitigation of off-target cleavage in SpRYgest reactions. (**a**) Schematic of the SaCas9-ABE8e plasmid used in **Supplementary Fig. 11a**, with the intended on-target site and an off-target site containing 3 mismatched bases highlighted. (**b**) Sequence alignment illustrating the positions of mismatches between the SaCas9-ABE8e on-target and off-target sites. (**c**) Schematic showing the strategy of targeting the opposite strand with a secondary gRNA in the SpRYgest of the SaCas9-ABE8e plasmid. The bottom panel table shows the number of putative closely matches sites in the target plasmid, generated using SpOT-check. The red shading highlights the off-by-3 mismatch off-target that led to cleavage by SpRY. (**d**) Agarose gel image showing the mitigation of off-target SpRYgest cleavage of the SaCas9-ABE8e plasmid by targeting the opposite strand with a secondary gRNA. (**e**) Gel image of purified SpRY-HF1[6] protein run alongside

a protein standard (Broad Range; New England Biolabs P7712). Novex 10-20% Tris-glycine gel post stained with SimplyBlue Safe Stain (both purchased from ThermoFisher). n = 1. (**f**) Agarose gel cleavage reactions with SpRY and SpRY-HF1 when using the primary and secondary gRNAs for the SaCas9-ABE8e plasmid digest. (**g**) Comparison of the cleavage efficiency of purified SpRY and SpRY-HF1 when using gRNAs that are either perfectly matched or that harbor pairs of mismatches to the target spacer sequence. Cleavage of DNA substrates was analyzed by capillary electrophoresis; mean, and in some cases s.e.m., shown for n = 3. (**h**) Sequence alignment of on- and off-target sites for an additional gRNA for which low level off-target cleavage was observed in SpRYgest of site NGCG-4. The bottom panel table shows the number of putative closely matches sites in the target plasmid, generated using SpOT-check. The red shading highlights the off-by-6 mismatch off-target that led to cleavage by SpRY. For **panels d**, **f**, and **h**, plasmid conformations are: R, relaxed; L, linear; S, supercoiled; agarose gels stained with ethidium bromide; gels images from n = 1 experiments; the 1 kb Plus DNA Ladder (NEB, N3200) is shown in kilobase sizes.



Supplementary Figure 16: Molecular cloning of saturation mutagenesis libraries within regions of SpCas9 via SpRYgest. (**a**,**b**) Schematics of SpRYgest strategies to generate saturation mutagenesis libraries of residues of SpCas9 important for its catalytic activity (within the HNH domain; **panel a**) and its NGG PAM preference (within the PAM-interacting domain; **panel b**). Silent substitutions were included in the library oligo to enable assessment of library construction efficiency (marked in orange boxes). Note that neither cloning approach is possible with WT SpCas9 due to a lack of NGG PAMs, as indicated by the purple dashed rectangles.



Supplementary Figure 17: Optimization of a one-pot gRNA IVT reaction. (**a**) Schematic illustrating the hands-on and incubation times of our previously established Standard gRNA synthesis method (Klenow + T7 RiboMAX; see **panel c** and **Supplementary Note 3**) versus streamlined 'one-pot' reactions. Approximate times in minutes are shown; lines not drawn to scale. (**b**) Comparison of *in vitro* transcribed gRNA yield for our previously established Standard gRNA synthesis method and a commercially available one-pot kit (EnGen sgRNA Synthesis Kit; NEB, # E3322S). (**c**) Comparison of *in vitro* transcribed gRNA yield generated via different methods with 4 hour reaction incubations (see also **Supplementary Note 6**). The gRNA yields are corrected for the µL volumes of each transcription reaction. For **panels b** and **c**, the IVT products were cleaned up using paramagnetic beads prior to quantification; datasets are from n = 1. (**d**) Agarose gel of SpRYgest reactions performed with gRNAs taken directly from gRNA synthesis reactions without conducting a clean-up step. The gRNA method IDs refer to the method IDs from **panel c**. The quantities of gRNAs NGCA-1 and NTGA-1 used in SpRYgest reactions were normalized to final concentration of 2 µM in the 50 µL SpRYgest reaction based on

concentrations determined in duplicate reactions that were cleaned up and quantified (plotted in **panel c**). For SpRYgest reactions performed with gRNAs NGGG-1 and NCTG-1, 5 µL of the uncleaned and unquantified gRNAs from the one-pot reactions were used directly in the SpRYgest reaction. Plasmid conformations are: R, relaxed; L, linear; S, supercoiled; agarose gels were stained with ethidium bromide; gels images from n = 1 experiments; the 1 kb Plus DNA Ladder (NEB, N3200) is shown in kilobase sizes.

Supplementary Figure 18: Timelines for molecular cloning reactions. (**a**-**c**) Gantt charts and workflow details for performing standard cloning reactions with restriction enzyme digests (**panel a**), cloning reactions via SpRYgests (**panel b**), and for generating saturation mutagenesis library via SpRYgests (**panel c**). Note that the oligonucleotide ordering times and most of the other process steps also apply to restriction enzyme cloning. The timing of oligo ordering can be expedited if local DNA synthesis facilities or next-day synthesis and delivery services are available.

Supplementary References:

1. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research* **42**, 3–19 (2014).
2. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).
3. Emiliani, F. E., Hsu, I. & McKenna, A. Circuit-seq: Circular reconstruction of cut in vitro transposed plasmids using Nanopore sequencing. 2022.01.25.477550 Preprint at https://doi.org/10.1101/2022.01.25.477550 (2022).
4. Mumm, C. *et al.* On-Ramp: a tool for rapid, multiplexed validation of plasmids using nanopore sequencing. 2022.03.15.484480 Preprint at https://doi.org/10.1101/2022.03.15.484480 (2022).
5. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat Biotechnol* **32**, 1262–1267 (2014).
6. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
7. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016). doi:10.1145/2939672.2939785.
8. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
9. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
10. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233–239 (2013).
11. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
12. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020).
13. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* **7**, 11750 (2016).
14. Moreb, E. A. & Lynch, M. D. A Meta-Analysis of gRNA Library Screens Enables an Improved Understanding of the Impact of gRNA Folding and Structural Stability on CRISPR-Cas9 Activity. *The CRISPR Journal* **5**, 146–154 (2022).
15. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).
16. Enghiad, B. *et al.* PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. *Nat Commun* **13**, 2697 (2022).
17. Enghiad, B. & Zhao, H. Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synth. Biol.* **6**, 752–757 (2017).
18. Gurevich, V. V., Pokrovskaya, I. D., Obukhova, T. A. & Zozulya, S. A. Preparative in vitro mRNA synthesis using SP6 and T7 RNA polymerases. *Analytical Biochemistry* **195**, 207–213 (1991).