Nucleic Acids Research, 2025, **53**, gkae1244
https://doi.org/10.1093/nar/gkae1244
Advance access publication date: 27 December 2024
Molecular and Structural Biology

SspA is a transcriptional regulator of CRISPR adaptation in *E. coli*

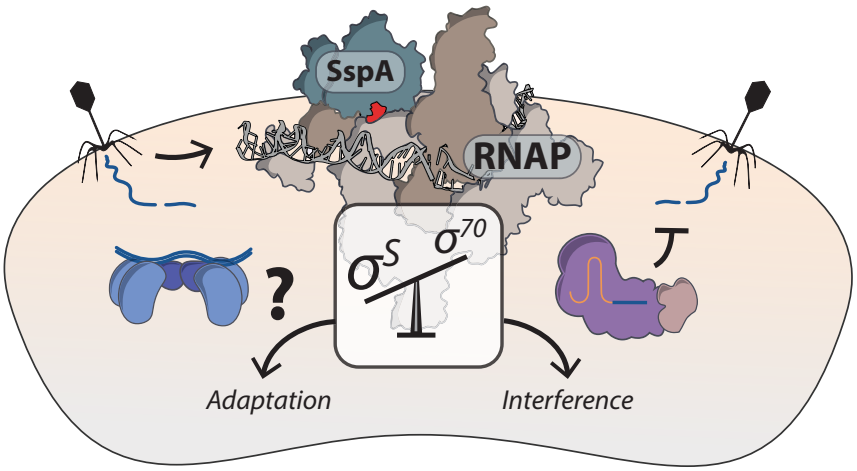
Santiago C. Lopez[1,2], Yumie Lee[1], Karen Zhang[1,2] and Seth L. Shipman[1,3,4,*]

[1]Gladstone Institute of Data Science and Biotechnology, 1650 Owens St, San Francisco, CA 94158, USA
[2]Graduate Program in Bioengineering, University of California, San Francisco and Berkeley, 1700 Fourth St, San Francisco, CA 94158, USA
[3]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, 600 16th Street, San Francisco, CA CA94158, USA
[4]Chan Zuckerberg Biohub San Francisco, , 499 Illinois St, San Francisco, CA 94158, USA
[*]To whom correspondence should be addressed. Tel: +1 415 734 4058; Email: seth.shipman@gladstone.ucsf.edu

Abstract

The CRISPR integrases Cas1-Cas2 create immunological memories of viral infection by storing phage-derived DNA in CRISPR arrays, a process known as CRISPR adaptation. A number of host factors have been shown to influence adaptation, but the full pathway from infection to a fully integrated, phage-derived sequences in the array remains incomplete. Here, we deploy a new CRISPRi-based screen to identify putative host factors that participate in CRISPR adaptation in the *Escherichia coli* Type I-E system. Our screen and subsequent mechanistic characterization reveal that SspA, through its role as a global transcriptional regulator of cellular stress, is required for functional CRISPR adaptation. One target of SspA is H-NS, a known repressor of CRISPR interference proteins, but we find that the role of SspA on adaptation is not H-NS-dependent. We propose a new model of CRISPR-Cas defense that includes independent cellular control of adaptation and interference by SspA.

Graphical abstract



Introduction

CRISPR-Cas is an adaptive immune system found in archaea and bacteria that is used to defend the host from foreign invaders, such as viruses or mobile genetic elements (1–5). This defence is mediated by Cas (CRISPR associated) proteins, which are capable of creating immune memories of invading nucleic acids and using those memories to mount RNA-guided degradation of invaders in the event of a future encounter (6–9). This process of storing immunological memory is known as CRISPR adaptation, and it is mediated by a phylogenetically conserved duo of proteins, Cas1 and Cas2, that form an integrase complex capable of inserting new DNA fragments (prespacers) into the cell's CRISPR array (10–13).

Studies spanning the past two decades have uncovered substantial mechanistic understanding of how CRISPR adaptation works and some of the key host factors that assist the CRISPR Cas1-Cas2 integrase complex in creating immune memories (14–24). Double-stranded DNA fragments are the preferred substrate for the CRISPR Cas1-Cas2 integrases and can arise from a variety of sources, such as foreign DNA

Received: May 27, 2024. **Revised:** November 23, 2024. **Editorial Decision:** November 29, 2024. **Accepted:** December 4, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

degradation by helicase-nuclease enzymatic complexes like the RecBCD complex (25) or AddAB (26) as well as from the replicating bacterial and phage genomes (27). Fragments captured by the CRISPR Cas1-Cas2 integrases can then undergo trimming by Cas4 (28), DnaQ (29) or other host exonucleases (30), generating free 3′ OH groups required as substrates for spacer integration (21,29). Cas1-Cas2 integrase docking at the Leader-Repeat junction of the CRISPR array requires the Integration Host Factor (IHF) (31–34), which generates a bend in the Leader sequence that accommodates the integrase complex and allows it to form stabilising contacts with the DNA (33). Docking enables the CRISPR integrase complex to catalyse a series of two nucleophilic attacks and add a new spacer at the Leader-Repeat junction (12,22,33).

Spacer integration creates staggered double strand breaks at either end of the duplicated repeat. Recent *in vitro* evidence suggests that host polymerases, in coordination with genome replication or transcription, could aid in repairing the CRISPR array (22) (Figure 1A). The expanded and repaired CRISPR array is capable of supporting further rounds of spacer acquisition.

Despite this knowledge, several open questions remain, including what host factors are responsible for regulation of CRISPR-Cas activity (35–39), and repair of the CRISPR array post-spacer integration (22,40), though some factors have been recently suggested play a role in this process (18,22). Furthermore, in contrast to noteworthy successes in heterologous reconstitution and harnessing of the CRISPR interference machinery across the tree of life, most notably CRISPR-Cas9, there are conspicuously few reports of successful heterologous expression of a CRISPR adaptation system outside of its native host (41,42) and no reports in eukaryotic systems. We, therefore, set out to discover additional host factors required for CRISPR adaptation.

Here, we develop and use a CRISPRi-based genetic screen to identify new host factors that participate in CRISPR adaptation in the Type I-E *Escherichia coli* system. We report that a novel host factor, SspA, acts as a transcriptional-level regulator of CRISPR adaptation through its role as a global transcriptional regulator of cellular stress. Transcript level-measurements of Cas1 and Cas2 expression in the ΔSspA background along with lac-inducible promoter-driven expression of the CRISPR integrases further confirmed that SspA does not directly regulate the expression of the CRISPR adaptation complex; rather, SspA must affect the expression of other coding or non-coding genes that in turn affect CRISPR adaptation. Lastly, we find that SspA regulation of CRISPR adaptation does not function via H-NS, a known regulator of CRISPR interference and a member of the SspA regulon. Our data support independent pathways for regulating the adaptation and interference components of CRISPR immunity, both downstream of SspA.

Material and methods

Bacterial strains and culturing

All strains used in this study can be found in Supplementary Table S2. Wild-type *E. coli* K-12 W3110 (BW25113) strain, generously provided by Joseph Bondy-Denomy, was used for all experiments in this study, unless specified. *Escherichia coli* K-12 MG1655 and LC-E75 (43) (derivative of MG1655, Addgene #115925) were used for the CRISPRi screen. *Escherichia coli* NEB-5-alpha (NEB C2987) was used for plasmid cloning. Keio collection (44) single-gene knock-out (KO) mutants, derivatives of BW25113, were generously provided by Carol Gross.

Additional deletions on Keio single-gene KO backgrounds were generated by λ_{Red} recombinase-mediated insertion of an FRT-flanked chloramphenicol (Cm^R) resistance cassette (45). This cassette was amplified from pKD3 (45) (Addgene #45604) with homology arms (50bp each) corresponding to the genomic sequences immediately up- and downstream of the intended deletion site. This amplicon was electroporated into the Keio strains expressing the λ_{Red} recombinase from pKD46 (45). Clones were isolated by selection on LB + chloramphenicol (10 μg/mL) plates. After polymerase chain reaction (PCR) genotyping and sequencing to confirm locus-specific insertion, the chloramphenicol and pre-existing kanamycin cassettes was excised by transient expression of FLP recombinase from pE-FLP (46) (Addgene #45978) to leave a single FRT scar, whenever specified in the text (i.e. Δ*gene*::*FRT*).

The *polA* ΔKlenow mutant was generated by λ_{Red} recombinase-mediated insertion of an FRT-flanked Cm^R resistance cassette into the Klenow fragment of *E. coli* BW25113 Polymerase I. This cassette was amplified from pKD3 with homology arms (50bp each), corresponding to the genomic regions flanking the Klenow fragment, as reported previously (47). This amplicon was electroporated into BW25113, expressing the λ_{Red} recombinase from pKD46. Clones were isolated by selection on LB + chloramphenicol (10 μg/mL) plates. PCR genotyping and sequencing confirmed the locus-specific insertion.

For the CRISPRi screen and CRISPR-Cas adaptation experiments, LB containing 1.5% w:v agar was used to grow strains on plates (growth at 37°C until single colonies became visible, usually ∼16 h). Strains were subsequently grown in LB broth at 37°C with 250 r.p.m. shaking, with appropriate inducers and antibiotics as described below.

For CRISPR-Cas adaptation experiments, strains were grown in LB broth at 37°C with 250 r.p.m. shaking, with appropriate antibiotics as described below. All experiments were performed without induction. We opted to not induce expression of Cas1-Cas2 as well as do rescue experiments on low-copy plasmids with genes in their native genomic architecture to avoid the cellular burden of potentially-toxic protein overexpression. Differences in spacer acquisition rates were observed with and without Cas1-Cas2 overexpression, but we observed more variability in cell growth and subsequent adaptation rates when overexpressing proteins, and thus opted against this strategy.

For CRISPR-Cas defence experiments, strains were grown in LB broth supplemented with 10 mM $MgSO_4$ and 0.2% maltose at 30°C with 250 r.p.m. shaking, with appropriate inducers and antibiotics as described below. For plaque assays, cells were mixed with top agar (0.5% w:v LB agar, supplemented with 10 mM $MgSO_4$ and 0.2% maltose and the appropriate antibiotics) poured over LB plates supplemented with the appropriate antibiotics and grown at 30°C overnight.

Inducers and antibiotics were used at the following working concentrations: 2 mg/mL L-Arabinose (GoldBio A-300), 1 mM IPTG (GoldBio I2481C), 1mM m-Toluic acid, 1 ug/mL anhydrotetracycline, 35 μg/mL kanamycin (GoldBio K-120), 25 μg/mL spectinomycin (GoldBio S-140), 100 μg/mL car-

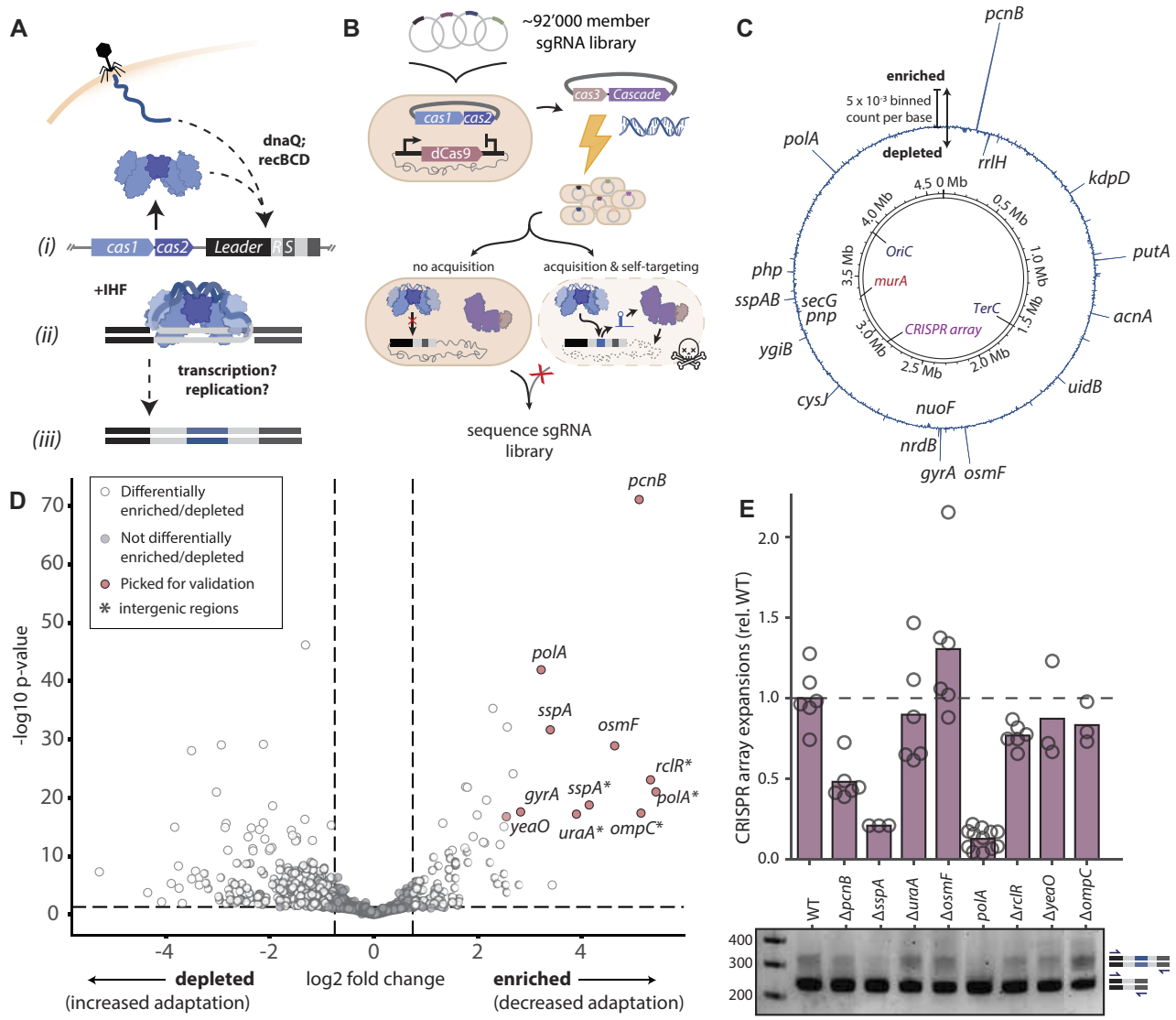


Figure 1. CRISPRi screen identifies adaptation host factors. (**A**) Overview of the CRISPR adaptation process, highlighting key known host factors. (**B**) Schematic of the CRISPRi adaptation host factor screen. (**C**) Binned coverage plot of sgRNAs across the *Escherichia coli* genome. sgRNA occupancy was calculated as the difference between the normalised (post/pre-screen) binned sgRNA counts per base of the experimental (+dCas9) and paired control (–dCas9) conditions. Regions of the genome with high ('enriched') sgRNA coverage are interpreted to be genomic loci that positively regulate CRISPR adaptation; regions of the genome with low (or negative, i.e. 'depleted') sgRNA coverage are interpreted to be genomic loci that negatively regulate CRISPR adaptation. The highest-ranking regions with attributable genes are labelled; other labelled loci are the *Ori* and *Ter* regions, the *murA* gene and the CRISPR-II array. $n = 9$ biological replicates. (**D**) Volcano plot showing $\log_2$ fold change for all genes versus their adjusted $-\log_{10}$*P*-values ($n = 9$ biological replicates). The horizontal dashed line represents an adjusted *P*-value of 0.05; the vertical lines represent $\log_2$ fold changes of –0.75 and 0.75. Genes targeted by sgRNAs differentially enriched that were selected for individual validation are labelled on the plot. (**E**) Top: deep-sequencing based measurement of the rates of new spacer acquisition in Keio knockouts harbouring pSCL565, after growth for 48h in liquid culture without induction of Cas1-Cas2 expression. Acquisition rates are shown relative to the wild-type parental strain. Open circles represent biological replicates ($n \geq 3$), bars are the mean (one-way ANOVA effect of strain $P < 0.0001$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*pcnB* $P = 0.00217$, Δ*sspA* $P = 0.000102$, *polA* ΔKlenow $P < 0.0001$; others ns). Bottom: representative agarose gel for the data shown. Expansions of the CRISPR array can be seen as higher sized bands above the parental array length. Additional statistical details in Supplemental Table S1.

benicillin (GoldBio C-103) and 25 μg/mL chloramphenicol (GoldBio C-105).

Phage strains and culturing

A virulent variant of phage Lambda (λ_{vir}) (48), generously provided by Luciano Marraffini, was used throughout this study. λ_{vir} was propagated on BW25113 grown in LB at 30°C based on previous studies (49). Briefly, overnights of *E. coli* BW25113 were grown at 30°C in 5mL LB + 10 mM $MgSO_4$ and 0.2% maltose. The next day, 300 uL of bacterial culture was infected with 10 uL of serial dilutions of λ_{vir} in LB + 10 mM $MgSO_4$ and 0.2% maltose, incubated at 30°C for 15 min, and added to 5 mL top agar, mixed gently and poured over LB agar plates. Plates were grown overnight at 30°C. Plates from the dilution series that showed evidence of confluent lysing of *E. coli* were covered in 5 mL LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose, placed on a shaker to agitate gently at room temperature for 2 h. Then, the lysate was transferred to a 15 mL conical tube, centrifuged at 4500 g × 15 min to remove the bacterial debris, and filtered through a 0.2 um filter. Phage titres were determined by

preparing 1:10 dilutions of λ_{vir} in LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose, and spotting 2.5 uL of the dilutions over top agar lawns of BW25113, which had been previously prepared by mixing 100 uL of the overnight culture with 5 mL of top agar (0.5% w:v LB agar, supplemented with 10 mM $MgSO_4$ and 0.2% maltose) and poured over LB agar plates. Serial dilutions of λ_{vir} were prepared in LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose, and 2.5uL of each dilution was spotted on the top agar using a multichannel pipette. Plates were tilted to allow phage spots to drip down the plate for easier quantification and left to dry completely at room temperature. Plates were incubated at 30°C overnight.

Plasmids

Plasmid pSCL565, encoding an IPTG-inducible *E. coli* Cas1-Cas2 cassette, spectinomycin resistance cassette and a pCDF ori, was constructed by PCR amplification of pCas1 + 2 (14) (Addgene #72676) to replace the T7 promoter by an IPTG-inducible Lac promoter.

Plasmid pSCL563, encoding an m-Tol-inducible *E. coli* Cas3-Cascade operon, carbenicillin resistance cassette and a pRSF ori was constructed by Gibson cloning.

The *sspAB* rescue set of plasmids, designed to rescue the loss of CRISPR adaptation phenotype of the *ΔsspA* mutant, were constructed by first Gibson cloning the *sspAB* operon (including 236bp upstream of *sspA*, containing the predicted promoter (50) between *rpsI* and *sspA*) into a low copy plasmid backbone (pSC101 ori) containing a carbenicillin resistance cassette. This yielded pSCL735 (*sspAB* rescue). Variants of the *sspAB* operon were generated by targeted PCRs to yield pSCL747 (*sspA* rescue, not tested because toxic in Δ*sspA* background); pSCL748 (*sspB* rescue); pSCL751 (*sspA* frameshifted $AN^{5\text{-}6}$ > $AQ^{5\text{-}6}$ GCC|AAC > GCT|*CAA*|C + *sspB*); pSCL752 (*sspA* + *sspB* frameshifted $PR^{9\text{-}10}$ > $PS^{9\text{-}10}$ CCA|CGT > CCA|*TCG*|T); pSCL753 (*sspA* frameshifted $AN^{5\text{-}6}$ > $AQ^{5\text{-}6}$ GCC|AAC > GCT|*CAA*|C + *sspB* frameshifted $PR^{9\text{-}10}$ > $PS^{9\text{-}10}$ CCA|CGT > CCA|*TCG*|T); and pSCL770 (*sspA* $PHP^{84\text{-}86}$ > $AAA^{84\text{-}86}$ + *sspB*).

The CRISPR defence set of plasmids, designed to pre-immunise *E. coli* strains against λ_{vir} by expressing an *E. coli* CRISPR-I array with a first spacer encoding either a Target (complementary to the λ genome (1,51) or a Non-Target (NT) spacer, were constructed by cloning a $spacer_1$-swapped *E. coli* CRISPR-I array (Cas-adjacent array in K-12 *E. coli*) into a high copy plasmid backbone (ColE1) containing a kanamycin resistance cassette. This yielded pSCL787 (Target, $spacer_1$ complementary to the $\lambda_{vir}R$ gene (1) and pSCL788 (Non-Target, $spacer_1$ complementary to the *S. cerevisiae ade2* gene).

The *sspAB-hns* rescue set of plasmids, designed to rescue the loss of CRISPR adaptation phenotype of the Δ*hns* and *ΔsspA* Δ*hns* mutants, were constructed by Gibson cloning the *hns* operon (including 419-bp upstream and 122-bp downstream of *hns*, containing the predicted promoter (50), regulatory and terminator regions contained between *tdk-hns-galU*, respectively) and/or *sspAB* operons (as above) into a low copy plasmid backbone (pSC101 ori) containing a carbenicillin resistance cassette. This yielded pSCL785 (*hns* rescue) and pSCL786 (*hns // sspAB* rescue).

pSCL832 was constructed from pSCL565 by swapping the *E. coli* Cas1-Cas2 CDS with an eGFP CDS via Gibson Assembly.

Additional plasmid information can be found in Supplementary Table S3.

CRISPRi adaptation host factor screen

LC-E75 (43), a derivative of MG1655 *E. coli* encoding a Tetracycline-inducible dCas9 cassette integrated at the Phage 186 *attB* site, and *E. coli* MG1655 were electroporated with pSCL565, and transformants were isolated on LB + spectinomycin after overnight growth at 37°C. Single colonies were inoculated into 5 mL LB + spectinomycin and grown overnight. Each experiment was repeated three times in triplicates for a total of nine paired LC-E75 (experiment) – MG1655 (control) screens.

The next day, cultures were electroporated with a library of 92, 919 sgRNAs (psgRNA (43) Pooled Library #115927, Addgene), targeting coding and non-coding regions across the *E. coli* genome, as described in (43,52). Briefly, 4 mL of the overnight cultures were diluted into 400 mL LB + spectinomycin and grown for 2 h at 37°C with shaking (250 r.p.m.). Cells were then subjected to an electroporation prep: cultures were split into 50 mL falcon tubes, chilled on ice for 10 min, and pelleted at 4000g for 15min at 4°C. The supernatants were discarded, and cells were washed with 30 mL of ice-cold ultra-pure, DNAse/RNAse free, pyrogen-free H_2O (updH_2O). The resuspended cultures were chilled on ice for another 10 min and then pelleted at 4000 g for 15 min at 4°C. These wash steps were repeated twice, for three total washes. After the last wash, cells were resuspended in 600 uL of 10% glycerol in updH_2O (∼800 uL final volume).

Then, 180 uL of cells were added to 0.2-cm gap electroporation cuvettes (BioRad #1652086), and ∼1 ug of the sgRNA library was mixed with the cells (total volume in electroporation cuvette < 200 uL). Cells were electroporated with the following settings: 2.5 kV, 25 uF, 200 Ω. After the pulse, cells were quickly recovered in 25 mL of pre-warmed LB + spectinomycin and placed in a shaking incubator for 1 h at 37°C. The cultures were then transferred into 75 mL of pre-warmed LB + spectinomycin + kanamycin, and dilutions were plated on LB + spectinomycin + kanamycin to estimate CFUs.

The next day, CFUs were estimated, and the experiments were continued only if the library coverage was estimated to be >1000×. If so, 20 mL of the overnight cultures were diluted in 1 L warmed LB + spectinomycin + kanamycin + 1 uM anhydrotetracycline (aTc); the remainder of the overnight cultures was collected by centrifugation for pre-experiment library quantification.

Cultures were grown for 3 h, after which the electroporation prep was performed as described above. After the last centrifugation step, each pellet was resuspended in 150 uL of a mix of *murA* targeting pre-spacer oligonucleotides and ∼1 ug of pSCL563 in updH_2O. The *murA* targeting prespacer mix was prepared by combining and annealing complementary single-stranded oligos that encode a prespacer targeting the essential gene *murA* (F and R sequences: AGGTTATGGCAACCGATCTGCGTGCATCAGCAAGC; GCTTGCTGATGCACGCAGATCGGTTGCCATAACCT), to a final concentration of 3.125 uM per oligo. After electroporation, cells were rescued with 5 mL of pre-warmed LB + carbenicillin + kanamycin + 1 mM m-Toluic acid +

1 uM aTc and placed in a shaking incubator for 1 h at 37°C. Then, these cultures were then transferred into 20 mL of pre-warmed LB + carbenicillin + kanamycin + 1 mM m-Toluic acid + 1 uM aTc and placed in a shaking incubator overnight at 37°C. Cultures (post-experiment library samples) were harvested the next day by centrifugation, 4000 g × 30 min, followed by plasmid extraction using the Qiagen Plasmid Plus Midi kit (cat. no. 12143).

Sequencing of the sgRNA libraries was performed as follows. Around 1 uL of the plasmid extractions were used as template in 50 uL PCR reactions, using 37 uL of updH_2O, 10 uL 5X Q5 reaction buffer, 1 uL 10 mM dNTPs, 1 uL Q5 Hot Start HiFi DNA polymerase and 0.25 uL 100 uM Forward and Reverse primers. The primers used contained Illumina adapters to make the amplicons compatible with our downstream sequencing prep, as well as 1–5 random nucleotides between the Illumina adapter and the annealing sequence to introduce diversity into the sequencing library. The PCR reaction was run using the standard recommended Q5 cycling conditions: 98°C initial denaturation × 30 s; 30 cycles of 98°C × 10 s, 62°C × 30s, 72°C × 30 s; final extension of 2 min at 72°C. Amplicons were then cleaned up using AMPure XP beads (A63880), indexed using custom indexing oligos and sequenced on an Illumina NextSeq instrument with ∼2 million reads per biological replicate. A list of primers can be found in Supplementary Table S4.

Fluorescence-based monitoring of the Lac promoter activity

Escherichia coli BW25113 (control) and Δ*sspA* strains were transformed with pSCL832 by electroporation, and transformants were isolated on LB + spectinomycin after overnight growth at 37°C. Single colonies ($n \geq 3$) were inoculated into 3 mL of LB + spectinomycin and grown overnight with 250 r.p.m shaking at 37°C. The next day, cultures were diluted 1:100 in 3 mL of LB + spectinomycin and grown to log phase (∼4 h). Subsequently, OD_{600} of the cultures was measured on a Spectramax i3 plate reader, and cultures were normalised to an $OD_{600} = 0.05$. Around 200 uL of cultures were placed on clear-bottom plate and incubated at 37°C on a Spectramax i3 plate reader, with fluorescence readings (wavelength = 508 nm) every 30 s for a total of 7.5 h.

qPCR

For experiments to estimate the relative plasmid copy number in Δ*sspA* cells versus the WT reference, *E. coli* BW25113 (control) and Δ*sspA* strains were transformed with pSCL565 by electroporation, and transformants were isolated on LB + spectinomycin after overnight growth at 37°C. Single colonies ($n \geq 3$) were inoculated into 3 mL of LB + spectinomycin and grown overnight with 250 r.p.m shaking at 37°C. The next day, cultures were diluted 1:100 in 3 mL of LB + spectinomycin and grown to log phase (∼4 h). Then, 1mL of cultures was harvested by centrifugation (21 000 g × 1 min), then resuspended in 250 uL of updH_2O. These samples were heated to 95°C for 15 min, then placed on ice to cool. Then, lysates were treated with two units of Proteinase K (NEB) for 30 min, followed by Proteinase K inactivation by incubation at 95°C for 10 min. Lastly, lysates were centrifuged at 21 000 g for 2 min, and supernatants were diluted 1:500 in updH_2O. Around 5 uL of the diluted supernatant was used in 20 uL qPCR reactions, set up using the NEB Luna Universal qPCR Master Mix following the manufacturer's instructions. Quantitative polymerase chain reaction (qPCR) primers were designed to target pSCL565's CDF ori and *cas1* regions, using the genomic *ompA* as a reference.

For experiments to investigate the possible YeaO-dependent mobilisation of the insG-containing IS4 element, single colonies ($n = 3$) of *E. coli* BW25113 (control) and Δ*yeaO* strains were inoculated into 3 mL of LB and grown overnight with 250 r.p.m shaking at 37°C. The next day, cultures were diluted 1:100 in 3 mL of LB and grown to log phase (∼4 h). Then, 1 mL of cultures was harvested by centrifugation (21 000 g × 1min). Then, 1 mL of cultures was harvested by centrifugation (21 000 g × 1 min), then resuspended in 250 uL of updH_2O. These samples were heated to 95°C for 15 min, then placed on ice to cool. Then, lysates were treated with two units of Proteinase K (NEB) for 30 min, followed by Proteinase K inactivation by incubation at 95°C for 10 min. Lastly, lysates were centrifuged at 21 000 g for 2 min, and supernatants were diluted 1:500 in updH_2O. Around 5 uL of the diluted supernatant was used in 20uL qPCR reactions, set up using the NEB Luna Universal qPCR Master Mix following the manufacturer's instructions. qPCR primers were designed to target three endogenous regions across the predicted IS4 element: the left end; the region from which most *insG*-derived CRISPR-acquired spacers had originated; and the right end of the IS4 element. The genomic *ompA* was used as a reference.

Primers are listed in Supplementary Table S4.

RT-qPCR

Single colonies ($n = 3$) of *E. coli* BW25113 (control) and Δ*sspA* strains were inoculated into 3 mL of LB and grown overnight with 250 r.p.m shaking at 37°C. The next day, cultures were diluted 1:100 in 3 mL of LB and grown to log phase (∼4 h). Then, 1 mL of cultures was harvested by centrifugation (21 000 g × 1 min). The pellets were resuspended in 1:2 parts TE:RNAprotect (Qiagen), and resuspended pellets were incubated at room temperature for 5 min. Then, samples were centrifuged for 10 min at 5000 g, the supernatant decanted, and the resulting pellets gently resuspended and incubated in 200 uL of TE containing 15 mg/mL lysozyme and 10 uL Proteinase K (NEB) at room temperature for 10 min. Around 700 uL of Buffer RLT was added to the samples and these were vortexed vigorously, followed by addition of 500 uL of 100% ethanol and additional mixing. RNA was then extracted using the Qiagen RNeasy Mini Kit, following manufacturer's instructions. Eluted in 100 uL of updH_2O, and DNase I (NEB) treated for 1 h at 37°C. Lastly, the DNase-treated RNA was cleaned up using the RNA clean & concentrate kit (Zymo), eluted in 150 uL of updH_2O, and normalised to a final concentration of 50 ng/uL in updH_2O. 100 ng of total RNA was used as input in 20 uL RT-qPCR reactions, set up using the NEB Luna Universal One-Step RT-qPCR kit following the manufacturer's instructions. qPCR was performed in technical triplicates of three biological replicates (three independent colonies). Relative gene expression was calculated using the ΔΔCt method after normalisation of genomic *cas1* (up- and downstream) gene, *cas2* gene expression to expression of the genomic *mreB* and *GAPDH* genes. Primers are listed in Supplementary Table S4.

RNAseq

Single colonies ($n = 3$) of *E. coli* BW25113 (control) and Δ*sspA* strains were inoculated into 3 mL of LB and grown overnight with 250 r.p.m shaking at 37°C. The next day, cultures were diluted 1:100 in 10 mL of LB and grown to log phase (∼4 h). Then, RNA was stabilized by adding 1.1 mL of stop solution (95% ethanol, 5% acid-buffered phenol) to the cultures and placing these in a dry ice – ethanol bath. Then, the cultures were split in 1mL aliquots for safekeeping, and cells were collected by spinning the samples at 21 000 g for 30 s. After removing the supernatant, cells were flash-frozen in by placing the tubes on the dry ice – ethanol slurry. Samples were then stored at 80°C until RNA extraction.

RNA was isolated from the frozen pellets through two rounds of extraction with 1 mL acid-buffered phenol-chloroform (Thermo Fisher Scientific), pre-heated to 67°C. RNA was further precipitated on a dry ice – ethanol bath for 30min after addition of 1 volume of isopropanol and 1/10 volume of 3 M sodium acetate (NaOAc, pH 5.5). Following incubation on a dry ice – ethanol, RNA was pelleted via centrifugation at 4°C and 21 000 g for 30 min. Pellets were washed twice with 500 uL of ice-cold 70% ethanol, air-dried and resuspended in 90 uL updH_2O. Then, to remove contaminant DNA, samples were treated with 4 uL Turbo DNase I (Invitrogen) and 10 uL 10 × Turbo DNase I buffer (Invitrogen) for 40 min at 37C. Samples were then mixed with 96 uL updH_2O, and the RNA was subsequently re-extracted with 200 uL buffered acid phenol-chloroform. RNA was precipitated by addition of 20 uL 3M NaOAc, and 600 uL of isopropanol, followed by placing the samples at −80°C for 4 h. RNA was pelleted by centrifugation at 4°C and 21 000 g for 30 min. Pellets were washed twice with 500 uL of ice-cold 70% ethanol, air-dried and resuspended in 30 uL updH_2O. Finally, the RNA concentration and integrity were assessed using Qbit RNA High Sensitivity and TapeStation RNA ScreenTape (Agilent) procedures, respectively.

Around 4 uL of total RNA in 11 uL updH_2O was used as starting material for RNAseq library preparation, using the NEBNext rRNA Depletion Kit (NEB #E7850) and NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB #E7770) kits, following the manufacturer's protocols (Section 5 of the NEBNext rRNA Depletion Kit (Bacteria) Instruction Manual, Version 5.0, 9/22). All steps were performed assuming degraded RNA (RIN ≤ 2), i.e. without additional RNA fragmentation. TapeStation RNA ScreenTapes were used to assess the rRNA-depleted RNA and cDNA integrity.

Following NEB adaptor ligation, PCR enrichment of adaptor-ligated DNA was performed using custom Illumina-compatible indexing primers that allowed for subsequent quantification and pooling of the samples, as described previously (53). Lastly, the indexed, pooled samples were sequenced on an Illumina NextSeq instrument, alloting 15 M reads per biological replicate. The corresponding data analysis is described below.

CRISPR-Cas adaptation experiments

Naïve CRISPR-Cas adaptation

E. coli BW25113 (control) and strains of interest were transformed with pSCL565 by electroporation, and transformants were isolated on LB + spectinomycin after overnight growth at 37°C. In the case of 'plasmid rescue' experiments, strains of interest we co-transformed with pSCL565 and the rescue plasmid by electroporation, and transformants were isolated on LB + spectinomycin + carbenicillin after overnight growth at 37°C.

Single colonies ($n \geq 3$) were inoculated into individual wells of a 96-well deep well plate containing 500 uL of LB + spectinomycin (and carbenicillin, if needed), and grown for 48 h with 1000 r.p.m shaking at 37°C. After 48 h of growth, 75 uL of the cultures were mixed with 75 uL of updH_2O, heated to 95°C for 10 min and spun-down. Around 0.5 uL of the supernatant was used as template for 25 uL PCR reactions (same recipe and cycling protocol as above). We designed primers to amplify a region of the *E. coli* CRISPR-II array, contained between the end of the Leader sequence and the second pre-existing spacer. To reduce the number of indices needed per sample, we designed three barcoded F primers (one per biological replicate) to amplify the CRISPR arrays – these would enable us to pool the samples post-CRISPR array amplification and de-multiplex the biological replicates during data analysis. A list of primers can be found in Supplementary Table S4.

In some cases, CRISPR array expansions are visible on an agarose gel as laddering caused by larger arrays (expanded) migrating slower than the shorter parental arrays. We visualised this by running 5 uL of the pooled PCR products on Invitrogen 2% Agarose SYBR safe E-Gels (A42135). Gels were re-stained with SYBR Gold before imaging.

Spacer electroporation assays

E. coli BW25113 (control) and strains of interest were transformed with pSCL565 by electroporation, and transformants were isolated on LB + spectinomycin after overnight growth at 37°C. The next day, single colonies ($n \geq 3$) were inoculated in 5 mL of LB + spectinomycin and grown for 2 h with 250 r.p.m shaking at 37°C. At this point, cells were harvested by centrifugation (4000 g for 15 min at 4°C), washed three times with 1 mL of updH_2O, and resuspended in a 150 uL mix of prespacer psAA33. This prespacer mix was generated by annealing two complementary 35 bp single-stranded oligos that encode a prespacer that has been previously reported to be acquired efficiently by the *E. coli* Type I-E CRISPR adaptation machinery (54) (final concentration: 3.125 uM per oligo). The cells were then transferred to a 0.2 cm gap electroporation cuvettes (BioRad #1652086), these were electroporated with the following settings: 2.5 kV, 25 uF, 200 Ω. After the pulse, cells were quickly recovered in 5 mL of pre-warmed LB + spectinomycin, and placed in a shaking incubator, to incubate for 3 h at 37°C. After 3 h of growth, the cultures were harvested by centrifugation (4000 r.p.m. × 15 min). The pellets were resuspended in 100 uL of updH_2O, heated to 95°C for 10 min and spun-down. Around 0.5 uL of the supernatant was used as template for 25 uL PCR reactions, using the same recipe, cycling protocol and CRISPR-II array-targeting primers as described above. A list of primers can be found in Supplementary Table S4.

CRISPR-Cas interference experiments

Phage plaque assays

BW25113 (control), Δ*sspA*::FRT, Δ*hns*::FRT and Δ*sspA*::FRT Δ*hns*::FRT strains were transformed with plasmids encoding either Target or Non-Target CRISPR-I arrays (pSCL787 and pSCL788, respectively), and transformants were isolated on LB + kanamycin after overnight growth at 37°C. Single colonies ($n \geq 3$) were inoculated into

3 mL of LB + kanamycin supplemented with 10 mM $MgSO_4$ and 0.2% maltose and grown overnight with 250 r.p.m shaking at 30°C. The next day, top agar lawns of each bacterial culture were prepared by mixing 100 uL of overnight cultures with 5 mL of top agar (0.5% w:v LB agar, supplemented with 10 mM $MgSO_4$ and 0.2% maltose and kanamycin). Top agar mixtures were poured over LB agar + kanamycin plates and left to dry at room temperature, partially open by a sterilizing flame. Serial dilutions of λ_{vir} were prepared in LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose, and 2.5 uL of each dilution was spotted on the top agar using a multichannel pipette, and left to dry completely at room temperature. Plates were incubated at 30°C overnight.

Efficiency of plating was calculated as the number of plaques formed by λ_{vir} on lawns of a strain harbouring pSCL787 (Target) divided by the plaques formed by λ_{vir} on lawns of a strain harbouring pSCL788 (Non-Target). Full plaque assay plates for all $n = 3$ biological replicates in Supplementary Figure S5.

Phage resistance infection growth curves

BW25113 (control), Δ*sspA*::FRT, Δ*hns*::FRT and Δ*sspA*::FRT Δ*hns*::FRT strains were transformed with plasmids encoding either Target or Non-Target CRISPR-I arrays (pSCL787 and pSCL788, respectively), and transformants were isolated on LB + kanamycin after overnight growth at 37°C. Single colonies ($n \geq 3$) were inoculated into 3 mL of LB + kanamycin supplemented with 10 mM $MgSO_4$ and 0.2% maltose and grown overnight with 250 r.p.m shaking at 30°C. The next day, cultures were diluted 1:100 in 3 mL of LB + kanamycin supplemented with 10 mM $MgSO_4$ and 0.2% maltose and grown to log phase (∼4 h). Subsequently, OD_{600} of the cultures was measured on a Spectramax i3 plate reader, and cultures were normalised to an $OD_{600} = 0.05$. Around 200 uL of cultures was infected with a range of MOIs ($10 \rightarrow 10^{-8}$), using serial dilutions of λ_{vir} prepared in LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose. Cultures were loaded on clear-bottom plate and incubated at 30°C on a Spectramax i3 plate reader, with OD_{600} readings every 2.5 min for a total of 16 h.

CRISPR-Cas primed adaptation after phage infection

BW25113 (control), Δ*sspA*::FRT, Δ*hns*::FRT and Δ*sspA*::FRT Δ*hns*::FRT strains were transformed with plasmids encoding either Target or Non-Target CRISPR-I arrays (pSCL787 and pSCL788, respectively), and transformants were isolated on LB + kanamycin after overnight growth at 37°C. Single colonies ($n \geq 3$) were inoculated into 3 mL of LB + kanamycin supplemented with 10 mM $MgSO_4$ and 0.2% maltose and grown overnight with 250 r.p.m shaking at 30°C. The next day, cultures were diluted 1:100 in 3 mL of LB + kanamycin supplemented with 10 mM $MgSO_4$ and 0.2% maltose and grown to log phase (∼4 h). Subsequently, OD_{600} of the cultures was measured on a Spectramax i3 plate reader, and cultures were normalised to an $OD_{600} = 0.05$. Around 200 uL of cultures was infected with λ_{vir} at an MOI of 0.1, and cultures were loaded on clear-bottom plate and incubated at 30°C on a Spectramax i3 plate reader, with OD_{600} readings every 1.5 min for a total of 3 h.

After 3 h, the cultures were harvested by centrifugation, resuspended in 100 uL of updH_2O, heated to 95°C for 10 min and spun-down. Around 1 uL of the supernatant was used as template for 25 uL PCR reactions (same recipe and cycling protocol as above). We designed primers to amplify a region of the *E. coli* CRISPR-II array contained between the end of the Leader sequence and the second pre-existing spacer. The barcoded F primer approach, described above, was used to pool PCRs and de-multiplex biological replicates during data analysis.

Protein model structures

Protein model coordinates were retrieved from the RSCB Protein Data Bank (codes 7DY6 and 8ET3). Figures were prepared using UCSF ChimeraX (55).

Data analysis

The data analysis for this project can be broken down into six modules: (1) processing of the sequencing reads to extract, count and group sgRNAs by gene/gene-adjacent regions; (2) generate binned coverage plots of sgRNAs across the *E. coli* genome; (3) identify the statistically enriched/depleted sgRNAs, using PyDESeq2 (56), a Python implementation of DEseq2 (57); (4) quantify the rates of CRISPR adaptation; (5) extract new spacers perform spacer analysis; and somewhat separately, (6) RNAseq analysis of differentially expressed genes between the WT and Δ*sspA* strains and subsequent GO enrichment analysis. All data analysis was performed in Jupyter Lab (58), and all code to replicate this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen.

Sequencing data processing: from reads to sgRNA counts

First, fastq reads were trimmed using sickle-trim (59). For each fastq, a counter of sgRNAs was generated by extracting the sgRNA from each read, provided that this sgRNA could be found in the original synthesised psgRNA library (43). Then, the sgRNAs were BLASTed (60) against the *E. coli* MG1655 genome and the top hit was saved. For each sample, a DataFrame of genomic_location–sgRNA–count was generated and used for downstream analysis. All data corresponding to the screens can be found in Supplementary Tables S5-S6. The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/blast_screen_hits_clean.ipynb.

Binned coverage plot of sgRNAs across the *E. coli* genome

We generated occupancy arrays for each sample, using counts generated above. These arrays contain cumulative counts of sgRNAs per base, i.e. occupancy $O = [c_1, c_2, \ldots, c_n]$, where n is the size of the *E. coli* MG1655 genome and c_i are the total sgRNA counts at that position. We then normalised the counts to the total sgRNA count in that sample, i.e. $O = [c_1/\text{sum\_sgRNAs}, c_2/\text{sum\_sgRNAs}, \ldots, c_n/\text{sum\_sgRNAs}]$, where sum_sgRNAs is total sgRNA count. Next, we calculated the mean occupancy for the experimental and control conditions, i.e. $O_{LC\text{-}E75} = (O_{biorep1} + O_{biorep2} + \ldots + O_{biorep9}) / 9$, where $O_{LC\text{-}E75}$ is the mean occupancy for the experimental condition, and O_{biorep_i} are the normalised counts for each biological replicate of the screen run in the experimental condition. Lastly, we calculated the delta occupancy, or difference between the mean sgRNA occupancies of the experimental

and control conditions, and posteriorly calculated the mean delta sgRNA occupancy in a sliding window, in the interest of interpretability. We used pyCirclize (61) to generate the final occupancy plot.

The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/plot_genome_coverage_clean.ipynb.

Identification of enriched/depleted sgRNAs

We performed statistical testing for enriched/depleted sgRNAs from binned sgRNA (sum of all sgRNAs per gene) count data generated in (1) using the PyDESeq2 package (56) and compared each experimental sample to its paired control, and controlled for pre-experimental variation in the relative sgRNA library composition by including the sgRNA counts from the pre-experiment library as an interactor factor (i.e. sgRNA_counts ~ input_lib_counts + knockdown (yes/no)). Genes that have less than a total of 10 reads for all of their sgRNAs in the dataset were removed from the analysis. The $\log_2$FoldChange ($\log_2$FC) value represents the enrichment or depletion of each gene. The lists of all genes, $\log_2$FC and adjusted *P*-values can be found in Supplementary Table S6. The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/deseq2_volcano_allhits_clean.ipynb.

Quantification of the rates of CRISPR adaptation

First, fastq reads were trimmed using sickle-trim. For each fastq, we filtered for reads containing the Leader-repeat junction of the *E. coli* CRISPR-II array. We then identified newly acquired spacers from the array sequences by recursive identification of CRISPR repeats and comparison of putative new spacers to pre-existing spacers in the array, using a lenient search algorithm allowing for a maximum of 3bp mismatches. We generated sums of new expansions in CRISPR arrays per condition, and used these to calculate the rate of CRISPR adaptation (100 * number of newly expanded CRISPR arrays / total number of arrays sequenced). Lastly, we normalised the rate of CRISPR adaptation for each condition by the wild-type rate CRISPR adaptation, so as to make inter-experiment comparisons feasible and more interpretable. All normalised rates corresponding to the CRISPR adaptation experiments, as well as the 'run' label (i.e. batch in which the experiment were run and sequenced) can be found in Supplementary Tables S7. The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/spacer_fishing_clean.ipynb.

Newly-acquired spacer analysis

Spacer analysis involves in several steps:

Extraction of new spacers

We began by using the same recursive new spacer search algorithm described above to extract new spacers. In parallel to extracting spacers, we also stored information regarding the total number of arrays sequenced and the fraction of those that were expanded, to use as normalisation for comparisons across samples that might vary in sequencing depth or quality.

Identification of spacer origin

Next, we generated a counter of newly acquired spacers and their frequencies. We used this to generate FASTA files of new spacers and their counts, which were subsequently BLASTed to two databases: the *E. coli* K-12 genome (taxid 511145) and pSCL565, to capture spacers derived from the Cas1-Cas2 expression plasmid. To identify the source of acquired spacers during CRISPR-Cas primed adaptation amid phage infection experiments, unique spacers extracted in steps described above were BLASTed to four databases: the *E. coli* K-12 genome (taxid 511145); the bacteriophage lambda genome (taxid 2681611); and pSCL787 or pSCL788, to capture spacers derived from the defence plasmids. In both cases, BLAST searches were performed with high stringency (≥90% identity, i.e. 30/33-bp match between spacer and reference query) to obtain unique matches to the reference maps. We then parsed the BLAST results and filtered the genome-matching spacers for LacI, Cas1 and Cas2, as we assumed that spacers from these sources were most likely plasmid derived.

Mapping spacers to reference genomes

Using the spacer genomic (lambda or *E. coli* K-12) or plasmidic (pSCL565, pSCL787 and pSCL788) location, target locus and counts, we generated coverage maps of the different genomes and plasmids where the spacers could have been sourced from, as well as spacer counts per location (i.e. counts of how many of the new spacers were *E. coli*, lambda or plasmid derived). Briefly, for each BLAST record, we first checked whether the BLAST record mapped to any of our reference genomes, and if so, added counts (from b.) to spacer origin and occupancy counters. The occupancy array is generated analogously to those used to estimate sgRNA coverage (see 2. above), and is genome-size aware (i.e. accounts for start-end junctions).

Spacer neighbourhood analysis

We also used the spacer ←→ genome information to look into the 15-bp up- and downstream of the genomic origin of the new spacer, in the hopes of capturing information regarding the PAM (canonically, AAG for this CRISPR adaptation system) and any other discernible motifs. This was done by mapping the spacer back to its reference genome, using the BLAST results and extracting 15 bases upstream and downstream of the spacer. These sequences were compiled and Logomaker (62) was used to generate sequence logos for the up and downstream region. This yielded Figure 2D.

Spacer origin distribution

Next, we used the spacer origin counters to obtain information about the breakdown of spacers by their origin (*E. coli*, lambda or plasmids). To do so, we first normalised the spacer count per location to the number of arrays sequenced. In parallel, we also normalised the spacer count per location to the total number of new spacers identified, converting this metric to the percent of spacers mapping to each location. This allowed us to then plot the new spacer count with respect to spacer origin and strain of interest (Figure 2B), in addition to the percent of new spacers belonging to each spacer origin and strain of interest (Figure 2C).

Coverage plots

Lastly, we generated genome coverage plots for *E. coli*, lambda, and the plasmids, as described in 2. above. For

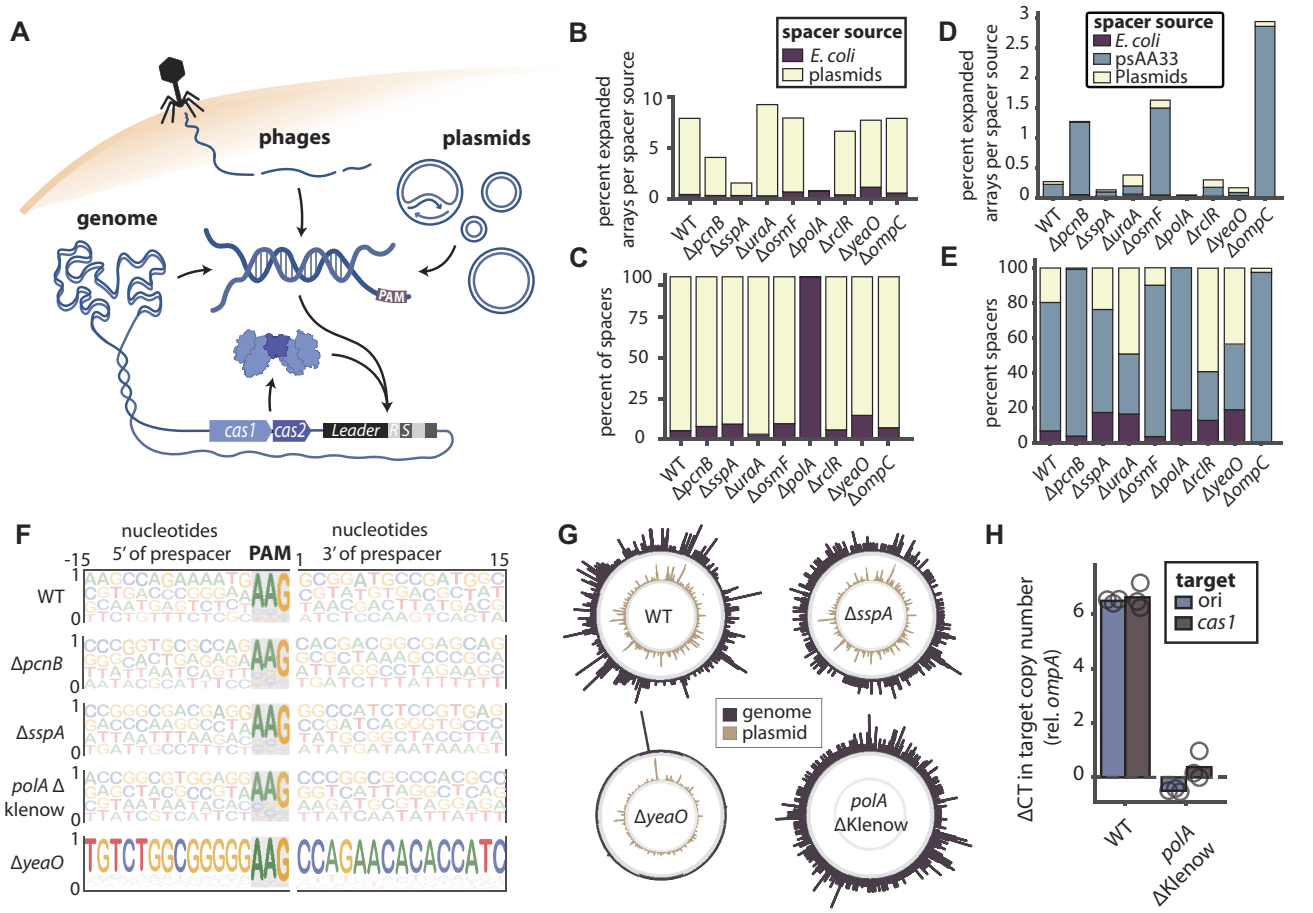


Figure 2. Features of spacers acquired in knockout strains. (**A**) Prespacer substrates for CRISPR adaptation arise from a variety of sources. (**B**) Breakdown of percent normalised spacer count (total number of new spacers/number of CRISPR arrays sequenced × 100) according to spacer origin (*E. coli* or plasmid) and strain of interest after naïve CRISPR adaptation assays. b-e: Δ*polA*: *polA* ΔKlenow fragment mutant. (**C**) Breakdown of percent of spacer attributable to each spacer origin (*E. coli* or plasmid) and strain of interest after naïve CRISPR adaptation assays. (**D**) Breakdown of percent normalised spacer count (total number of new spacers / number of CRISPR arrays sequenced × 100) according to spacer origin (*E. coli* or plasmid) and strain of interest after prespacer electroporation CRISPR adaptation assays. (**E**) Breakdown of percent of spacer attributable to each spacer origin (*E. coli* or plasmid) and strain of interest after prespacer electroporation CRISPR adaptation assays. (**F**) Motifs in the 15-bp up- and downstream of the newly acquired spacer in its source location. (**G**) Binned coverage plot of newly acquired spacer across the *E. coli* genome (outer, purple) and pSCL565 plasmid (inner, tan) for the wild-type strain (top-left) and derivatives. See Supplementary Figure S1 for the full set. h. qPCR-based measurement of the relative copy number of pSCL565 *Ori* and *cas1* sequences in the wild-type and *polA* ΔKlenow mutant. Delta CT values in Supplemental Table S8. Open circles represent biological replicates ($n \geq 3$), bars are the mean (one-way ANOVA effect of strain and target $P < 0.0001$; Sidak's corrected multiple comparisons for wild-type versus Δ*sspA*, CDF ori copy number $P < 0.0001$, *cas1* copy number $P < 0.0001$). Additional statistical details in Supplemental Table S1.

the *E. coli* and lambda genomes, we generated binned coverage plots by calculating the coverage as a sliding mean, or binned coverage. The spacer coverage for plasmids was generated without binning spacer occupancy. This analysis yielded Figure 2E–H, Figure 5C, Supplementary Figure S1 and Supplementary Figure S5.

The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/map_new_spacers_clean.ipynb.

RNAseq analysis

First, fastq reads were trimmed using sickle-trim. For each fastq, we then aligned the reads to an indexed reference *E. coli* BW25113 genome (GCF_000750555.1_ASM75055v1), using Bowtie2 (63). Following some SAM → BAM → sorting → index data finagling using Samtools (64), we used featureCounts (65) to generate a table of read counts per genomic feature across the *E. coli* genome (GCF reference file GCF_000750555.1). We then used the PyDESeq2 package (56) to statistically test for genes that were significantly under or overexpressed in the Δ*sspA* condition in comparison to the WT reference. Genes with less than a total of 10 reads in the dataset were removed from the analysis. The lists of all genes for which an adjusted *P*-value could be calculated along with their $\log_2$FC and adjusted *P*-values can be found in Supplementary Table S10. Subsequent GO enrichment analyses were performed using GOATOOLS (66). With a significance cutoff of 0.05 and Benjamini–Hochberg multiple correction test for follow-up multiple comparisons. The Jupyter Notebook for this analysis can be found here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/RNAseq_clean.ipynb.

Biological replicates

Biological replicates were taken from distinct samples, not the same sample measured repeatedly.

Data availability

All data supporting the findings of this study are available within the article and its supplementary information. Data used to generate all figures and perform statistical analysis, alongside a Jupyter Notebook to recreate our figures is available on GitHub here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen/blob/main/plot_run_stats_clean.ipynb. All sequencing data associated with this study is available on NCBI SRA (PRJNA1109382).

Code availability

All code used to process or analyse data from this study is available on GitHub here: https://github.com/Shipman-Lab/CRISPRi_host_factor_screen.

Results

CRISPRi screen identifies adaptation host factors

We designed a genome-wide CRISPRi screen to identify potential host factors that participate in Type I-E CRISPR adaptation (Figure 1A). This screen utilizes a library of 92 919 gRNAs that are distributed across a population of *E. coli*, each of which direct a catalytically dead Cas9 (dCas9) to knock down transcription at a single locus, with multiple redundant gRNAs per gene (43,52). We utilized this CRISPRi library in a negative selection scheme designed to deplete adaptation-competent cells. Specifically, we electroporated oligonucleotide prespacers that matched an essential gene into *E. coli* expressing a Type I-E CRISPR system. Integration of this prespacer into the CRISPR array would lead to the generation of a self-targeting crRNA and ultimately death of the adaptation-competent library members. CRISPRi knockdown of host factors involved in CRISPR adaptation or CRISPR interference would reduce self-targeting, leading to enrichment of host factor gRNAs in the population following selection (Figure 1B). While our screen will capture both adaptation and interference relevant genes, subsequent validation for this work focuses on adaptation host factors.

For this screen, we used *E. coli* LC-E75, a derivative of K-12 MG1655, which encodes a Tetracycline-inducible dCas9 cassette integrated at the Phage 186 attB site (43). We built an LC-E75 strain that carried a plasmid-encoded, IPTG-inducible Cas1-Cas2 cassette (plasmid hereon referred to as pSCL565). We then electroporated this strain and *E. coli* K-12 MG1655 (parental strain serving as a control) with a library of 92 919 plasmid-encoded sgRNAs, which target both coding and non-coding regions across the *E. coli* genome (43). The libraries were grown overnight, and subsequently passaged and grown to mid-log phase (∼3h) with dCas9 induction; the remainder of the overnight library cultures were harvested for sgRNA library sequencing (pre-screen library). Then, cells were co-electroporated with (1) a plasmid encoding an m-Toluic acid-inducible *E. coli* Cas3-Cascade, the effector of the Type I-E CRISPR interference system, and (2) a 35-bp dsDNA spacer targeting the essential gene *murA*. Cells were rescued in media containing inducers for Cas3-Cascade and dCas9 and antibiotics to select for their respective plasmids, cultured overnight, and harvested for sgRNA library sequencing (post-screen library). We extracted the sgRNA plasmid libraries and prepared samples for sequencing by amplifying the sgRNAs using a primer pool targeting the region upstream of the sgRNA promoter and downstream of the tracrRNA. The primers contained Illumina adapters to make the amplicons compatible with our downstream sequencing prep. Sequencing of the sgRNAs libraries yielded sgRNA counts for the dCas9-expressing LC-E75 and control dCas9-less parental strains, which allowed the calculation of the binned enrichment/depletion of sgRNAs across the *E. coli* genome (Figure 1C).

We found peaks of sgRNA enrichment that were distributed across the *E. coli* genome and did not cluster around the *murA* locus. Additionally, we identified *polA*, *priA* and *gyrA*, essential genes previously suggested to play a role in the CRISPR adaptation process. This highlights the advantage of a knock-down approach over transposon-based KO approaches, where essential genes would have been lost from the library altogether. We found several other regions of the *E. coli* genome where sgRNAs were strongly enriched, suggesting additional host factors (18).

We quantified differentially enriched or depleted sgRNAs from their cumulative sgRNA counts (sum of all sgRNAs per gene), by comparing each experimental sample (+dCas9) to its paired control (–dCas9) using PyDESeq2 package (43,52,56). We filtered out genes with less than 10 cumulative reads, and controlled for variation in relative sgRNA library composition by including pre-screen sgRNA counts as an interaction factor in the model. We found 571 differentially enriched/depleted genes and gene-adjacent regions, out of a total of 12 809 gene/gene-adjacent regions considered in our analysis (Figure 1D). Interestingly, a subset of the differentially enriched genes (i.e. CRISPR adaptation deficient when knocked-down) also had their gene-adjacent regions differentially enriched (shown with asterisked gene names).

We selected the top eight gene regions with highest $\log_2$ fold changes for individual validation using knockout mutants from the Keio collection (44) in a naive adaptation assay. One additional gene, *gyrA*, is essential and could not be validated with a knockout. Although polA knockouts are non-viable, a *polA* Klenow fragment deletion mutant is viable (47), and was thus used in validation assays alongside the other non-essential genes.

We electroporated wild-type and knockout strains with pSCL565 and grew them in liquid culture for 48h without inducers for Cas1-Cas2 to achieve a moderate level of expression from transcriptional leak, as employed previously (25). We then sequenced the CRISPR II array of these cells (i.e. endogenous CRISPR array flanked by the *ygcE* and *ygcF* genes (67), hereon referred to as CRISPR-II) and quantified the rate of CRISPR adaptation as the fraction of sequenced arrays that had acquired new spacers. Biological replicates run on different days were normalised to the CRISPR adaptation rate of the wild-type parental Keio strain (Figure 1E). We found that three mutants showed significantly decreased rates of CRISPR adaptation compared to the wild-type strain: *pcnB*, *sspA* and *polA* ΔKlenow.

Features of spacers acquired in knockout strains

Spacers captured by the CRISPR Cas1-Cas2 integrases come from a variety of sources. Defence associated sources include mobile genetic elements and phages. However, in the absence of interference machinery, spacers derived from the bacterial genome and plasmids accumulate (Figure 2A). We next tested whether any of the hits that we chose for validation modified the source of new spacers. We found that, consistent with pre-

vious findings (14,54), the majority of new spacers acquired in the wild-type strain were plasmid-derived (Figure 2B). This finding held for all mutants except *polA* ΔKlenow, which acquired spacers solely from the genome. The breakdown of spacer origin as a percent of all newly acquired spacers starkly illustrates this finding (Figure 2C).

Next, given that our screen's success hinges partly on the relative ability of a given knockdown to acquire the electroporated self-targeting spacer, it is conceivable that knockdowns that for whatever reason were defective in their ability to acquire exogenously-provided spacers would show up as hits in our screen. Thus, we performed spacer electroporation assays on our deletion strains to determine whether any of these are defective in acquiring electroporated spacers.

We found, broadly speaking, similar patterns in the rates of CRISPR adaptation to those observed when subjecting cells strictly to naïve CRISPR adaptation (Figure 2D, E), that is, reduced rates of CRISPR adaptation from most mutants. We also found that, overall, the electroporated spacer either added to or replaced part of the pool of spacers that were previously acquired from the plasmid: this is particularly noticeable in the case of Δ*osmF* and Δ*ompC*, where the electroporated spacers are acquired readily and replace almost entirely what were previously plasmid-derived spacers.

We also found some notable differences between the naïve and electroporated spacer acquisition patterns. For instance, although in naïve spacer acquisition assays, the spacer repertoire in the Δ*pcnB* strain was similar to the WT, both in relative makeup and relative distribution of spacers from *E. coli* genome and plasmid, the Δ*pcnB* mutant acquired electroporated spacers readily, suggesting that *pcnB* might be involved in the prespacer generation step. In contrast to the Δ*pncB* mutant, the Δ*polA* mutant remained significantly impaired in its ability to acquire new spacers, particularly from plasmids, though it was able to acquire electroporated spacers. The Δ*sspA* mutant remained comparably incapable of acquiring spacers, even those electroporated, suggesting that sspA's role on CRISPR adaptation is downstream of the prespacer processing step.

We next sought to determine whether the differences in new spacer acquisition could be explained by a change in PAM preference or other motifs up- or downstream of the spacer. We searched 15-bp up- and downstream of the newly acquired spacer in its source location, and found that all mutants showed similar PAM preferences to the wild-type strain, consistent with previous reports (68). Similarly, all mutants except the *yeaO* deletion mutant showed no additional up- and downstream motif preferences of preference, beyond the AAG PAM.

The *yeaO* mutant displayed strong motif preferences up- and downstream of the genome-derived spacers (Figure 2F), which prompted us to map all newly acquired spacers for each mutant to their respective source on either the *E. coli* genome or pSCL565 plasmid (Figures 2G; see Supplementary Figure S1 for an expanded view of these figures). We found that the distribution of new spacers from both sources were mostly consistent between the wild-type and the mutants tested (Supplementary Figure S1A–I) with two exceptions: the *yeaO* and *polA* ΔKlenow mutants.

We found that, as suggested by the prespacer neighbourhood motif analysis (Figure 2F), the *yeaO* mutant acquired spacers almost uniquely from one location in the genome, which maps to the gene *insG*, encoding an IS4 transposase (Figure 2G, bottom-left; Supplementary Figure S1H). We initially hypothesised that YeaO could be a transcriptional regulator of the InsG-mediated mobilisation of IS4, explaining the increased spacer acquisition from a transposable element. However, in a new set of experiments we did not detect this enrichment of insG in the *yeaO mutant* and qPCR assays from those follow-up experiments measuring the relative abundance of the left and right ends of the putative IS4 element in the Δ*yeaO* versus WT backgrounds failed to find evidence for mobilization (Supplementary Figure S2A, B). Thus, either the insG is not reliably mobilized across replicates or the single peak we observed initially was the result of a jackpot event during acquisition or sequencing preparation.

The *polA* ΔKlenow mutant had a similar distribution of prespacers originating from the genome when compared to the wild-type, but we were unable to map any prespacers to the plasmid, suggesting that the plasmid was unable to serve as a source of prespacers in this mutant (Figure 2G, bottom-right; Supplementary Figure S1f). Given the loss of CRISPR adaptation from the pSCL*565* plasmid but wild-type levels of CRISPR adaptation from the genome (Figure 2B), we hypothesised that the *polA* ΔKlenow mutant could be deficient in plasmid replication (69). To test this, we measured the relative number of copies of the pSCL*565 Ori* and *cas1* sequences (the latter also found in the genome) in the wild-type and *polA* ΔKlenow mutant. We found a nearly 80-fold difference in the relative number of copies of pSCL*565* that the *polA* ΔKlenow mutant contains compared to the wild-type strain, which would explain this strain's decreased ability to acquire new spacers from the plasmid (significantly decreased number of plasmid copies per cell) and also why it was identified as a hit in the initial screen (Figure 2H).

SspA is a transcriptional regulator of CRISPR adaptation

Our CRISPR adaptation assays and downstream analysis of acquired spacers revealed that Δ*sspA* was consistently and significantly defective in naïve CRISPR adaptation, despite no other noticeable differences in the features of its acquired spacers when compared to the wild-type parental strain (Figure 2B, C, E and F). Additionally, we found that the decrease in CRISPR adaptation in the Δ*sspA* background was not due to decreased levels of protein expression from the lac-inducible promoter used in our assays (Supplementary Figure S3). We thus selected *sspA* for further mechanistic characterisation.

E. coli SspA was discovered four decades ago during a screen for proteins induced by the stringent response (70). Over the years, its reported cellular functions have increased, and SspA has become particularly linked to global stress response (71,72) through its action as an RNA polymerase (RNAP)-associated protein (73,74). Crystal structures of *E. coli* RNAP-promoter open complex with SspA have revealed that SspA inhibits σ^{70} promoter escape through contacts with both RNAP and σ^{70} through a conserved PHP motif (71,74–76). This promoter escape inhibition induces a rewiring of the cellular transcriptomic landscape towards expression of σ^{S} genes, with implications on stress tolerance, motility and virulence (71,74–76). The *sspA* gene is encoded in a two-member operon, upstream of *sspB*. SspB acts as a specificity-enhancing factor for the ClpXP protease (77). It helps maintain protein homeostasis by escorting SsrA-tagged

peptides, resulting from stalled ribosomes, to the ClpXP protease and promoting their degradation (Figure 3A), thus simultaneously freeing ribosomes and replenishing the pool of amino-acids that can become a precious resource in conditions of starvation.

Though we found *sspA* and not *sspB* as a significant hit in our screen, we sought to confirm that the defects in CRISPR adaptation observed in the Δ*sspA* mutant were due strictly to the lack of SspA, and not due to polar effects of this mutation on the downstream *sspB* gene. We compared the rates of CRISPR adaptation in wild-type strains to those in Δ*sspA::kan*^{R} and Δ*sspB::kan*^{R} mutants carrying pSCL*565* (Figure 3B). We found that the Δ*sspA* mutant was deficient at new spacer acquisition, but the Δ*sspB* mutant acquired spacers at rates indistinguishable from the wild-type strain (Figure 3C). We attempted to deliver an *sspA* rescue plasmid into the Δ*sspA::kan*^{R} strain, but this yielded no transformants over multiple attempts. However, we found that we could rescue the CRISPR adaptation phenotype to wild-type levels when Δ*sspA::kan*^{R} carrying pSCL*565* were additionally electroporated with an *sspAB* cassette, encoding the SspA and SspB proteins under control of their native promoter and on a low-copy (∼5) plasmid. This suggests that lack of *sspA* alone is sufficient to cause the loss of adaptation phenotype, and that this can be rescued by supplying a copy of the *sspA* gene in *trans*, under its native regulation.

Given these findings, we next sought to determine which part of the SspA protein was responsible for the loss of adaptation phenotype. We were particularly interested in the SspA PHP$^{84\text{-}86}$ motif, which has been reported to be indispensable for stabilisation of interactions between SspA, σ^{70} and the RNAP complex. Via this interaction, SspA acts as a transcriptional repressor of σ^{70} promoters by inhibiting promoter escape (74). Triple-Alanine substitutions in this motif cause pleiotropic cellular effects such as increased swarming and defects in acid-resistance and phage P1 growth (72,78). Thus, given SspA's role as a transcriptional rewiring agent, we decided to test whether an SspA PHP$^{84\text{-}86}$ > AAA$^{84\text{-}86}$ mutant, deficient in σ^{70}-RNAP binding, would also phenocopy the Δ*sspA* mutant in terms of loss of CRISPR adaptation. To do this, we designed rescue plasmids, encoding variants of the *sspAB* operon under endogenous regulation, on low-copy (∼5) plasmids (Figure 3D).

We found that rescue plasmids encoding the full *sspAB* operon or an early frameshifted *sspB* could rescue CRISPR adaptation to levels comparable to wild-type. However, rescue plasmids encoding SspB, an early frameshifted *sspA*, early frameshifted *sspA* and *sspB*, and crucially, the SspA PHP$^{84\text{-}86}$ > AAA$^{84\text{-}86}$-SspB RNAP binding mutant were all deficient in CRISPR adaptation (Figure 3E).

Having performed our previous CRISPR adaptation assays with Cas1-Cas2 expression under control of a Lac-inducible promoter, we believed that SspA has an indirect transcriptional regulatory role on adaptation: rather than acting directly on the Cas1-Cas2 shared promoter, SspA would affect the expression of other coding or non-coding genes that in turn affect CRISPR adaptation, through direct or indirect effects on the Cas1-Cas2 integrase. To confirm that the Cas1-Cas2 transcript levels remained unchanged in the ΔsspA background, we quantified the expression levels of Cas1 and Cas2 in the WT and ΔsspA backgrounds by RT-qPCR. We found that cas1 and cas2 transcript levels were not significantly decreased in the ΔsspA background (Figure 3F).

In a further attempt to identify the differentially expressed genes that could help explain the loss of adaptation phenotype observed in the ΔsspA strain, we performed RNAseq experiments to compare the transcriptomes of wild-type and ΔsspA strains. We found nearly a thousand genes that were differentially expressed (Figure 3G). We clustered these genes through GO term enrichment analyses, and found that statistically significant genes groups included those involved in chemotaxis/cell motility, transmembrane transport, DNA damage response and repair, as well as regulation of transcription (Supplementary Figure S4A).

Lastly, we hypothesized that if other genes, transcriptionally affected by the loss of sspA, were to participate in CRISPR adaptation, there is a possibility that they also be significant hits in our CRISPRi screen. Following this hypothesis, we evaluated the set of genes that are at the overlap between CRISPRi screen hits and those found to be differentially expressed in our ΔsspA versus WT RNAseq experiment. The hope was to find a gene or set of genes that are both differentially expressed in the RNAseq experiments and are statistically significant in the CRISPRi screen.

We found nearly 250 genes that were significant hits in both the CRISPRi screen and differentially expressed in the ΔsspA KO versus WT (Supplementary Figure S4B). In an attempt to find patterns among this set of genes, we once again performed GO term enrichment analyses (Supplementary Figure S4C). Though there were statistically significant gene groups, notably those associated to ATP biosynthesis, protein translation and ribosomal function, we are unable to speculate further as to what the SspA interactor network is, in terms of its ability to affect CRISPR adaptation. Experiments to identify the next layer in this mechanism go beyond the scope of this manuscript, but this set of genes will help inform future work.

Taken together, our data is consistent with a model in which SspA's role as RNAP-σ^{70} interactor and transcriptional rewiring agent is required for functional CRISPR adaptation.

H-NS regulates CRISPR interference downstream of SspA

Having established SspA's role as a regulator of CRISPR adaptation, we sought to test whether it could also play a role in regulating CRISPR interference. As SspA has been shown to be rapidly and highly upregulated following lambda infection (79), it could serve as a link between phage infection and CRISPR defence generally. Given previous reports of the role of SspA in downregulating levels of H-NS (72,80), a repressor of the CRISPR interference machinery, we hypothesised that SspA could be acting on the CRISPR-Cas system via H-NS (51,81,82) (Figure 4A).

To assess the effects of SspA on CRISPR mediated anti-phage defence and the potential interactions between SspA and H-NS in regulating this defence, we constructed Δ*sspA::FRT*, Δ*hns::FRT* and Δ*sspA::FRT* Δ*hns::FRT E. coli* strains (Figure 4B). Previous studies have shown that H-NS is a strong repressor of CRISPR-Cas gene expression, but that this repression can be relieved by knocking out H-NS (51,81). This de-repression can result in defence against bacteriophages, provided that these cells' CRISPR arrays encode one or more spacers targeting the phage genome (hereinafter referred to as 'pre-immunised *E. coli*') (1,51,81).

We electroporated our mutant strains with a plasmid carrying a CRISPR array encoding a first spacer complementary

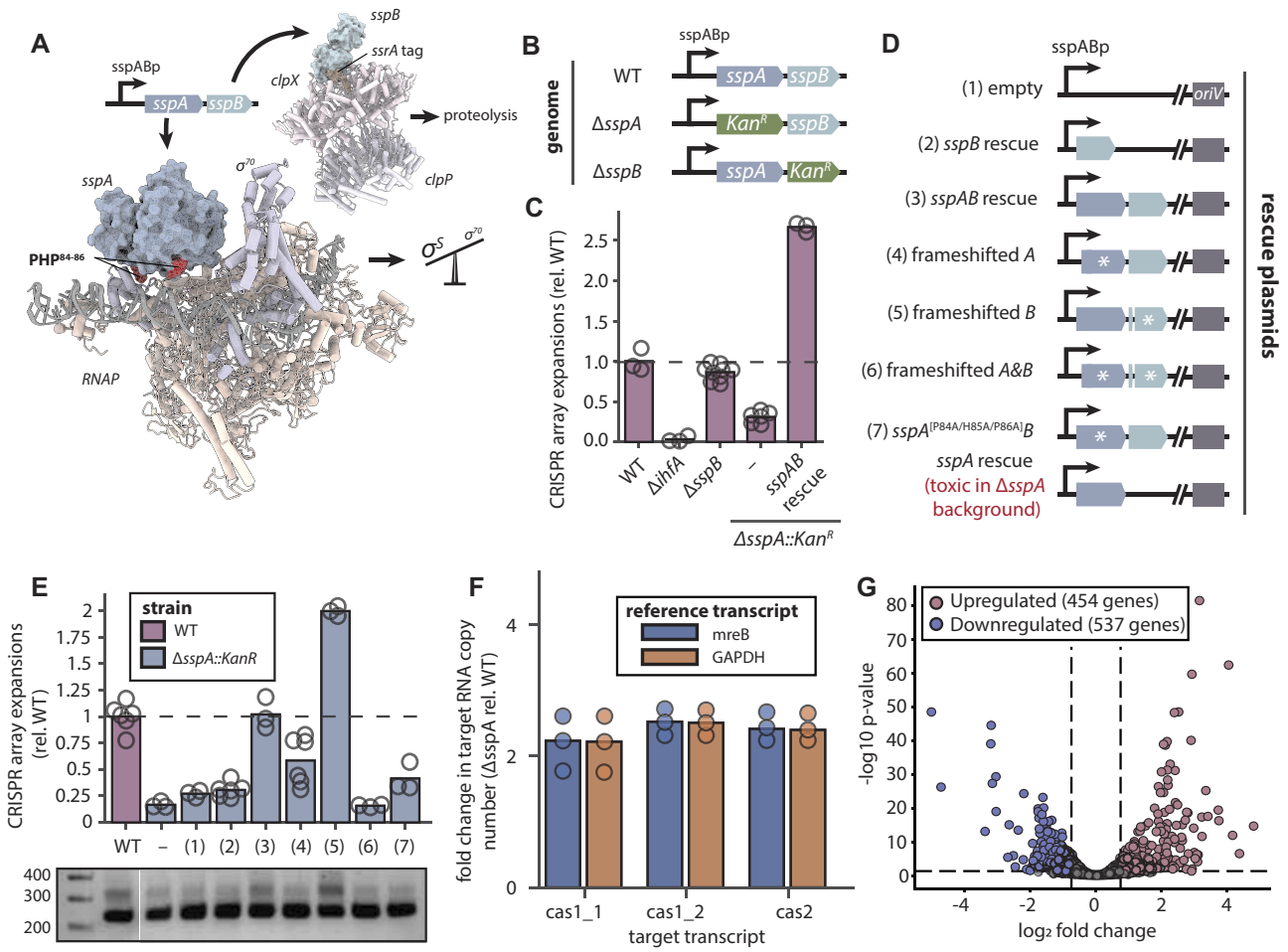


Figure 3. *sspA* is a transcriptional regulator of CRISPR adaptation. A. *sspAB* operon, proteins and function. Bottom left: crystal structure of an SspA dimer in complex with *E. coli* RNAP-promoter open complex, showing the conserved SspA PHP$^{84\text{-}86}$ residues interacting with RNAP and σ^{70} (PDB 7DY6 (74)). Top right: crystal structure of SspB escorting an SsrA-tagged substrate being delivered to the ClpXP protease complex (PDB 8ET3 (77)). B. Schematic of the *sspAB* operon of WT, Δ*sspA*::*kan*^{R} and Δ*sspB*::*kan*^{R} strains. *kan*^{R}: kanamycin resistance cassette. C. Deep-sequencing based measurement of the rates of new spacer acquisition in strains harbouring pSCL565 and, in the case of the Δ*sspA*::*kan*^{R}, either an empty plasmid or a low (~5) copy plasmid encoding the *sspAB* operon, after growth for 48 h in liquid culture. Adaptation rates are shown relative to the wild-type parental strain. The Δ*ihfA* strain was used as a negative control, as it is required for *in vivo* spacer acquisition in the *E. coli* type I-E CRISPR system (31,32). Open circles represent biological replicates ($n \geq 3$), bars are the mean. Horizontal dashed line represents the mean rate of spacer acquisition in the wild-type strain (one-way ANOVA effect of strain $P < 0.0001$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*sspA* $P < 0.0001$, Δ*sspB* $P = 0.109807$; Δ*sspA* versus Δ*sspB* $P < 0.0001$). D. Schematic of the *sspAB* operon variant rescue plasmids. All plasmids are low (~5) copy, and encode variants of the *sspAB* operon under its native regulation. Frameshift mutants of SspA (AN$^{5\text{-}6}$ > AQ$^{5\text{-}6}$ GCC|AAC > GCT|*CAA*|C) and SspB (PR$^{9\text{-}10}$ > PS$^{9\text{-}10}$ CCA|CGT > CCA|*TCG*|T) encode sequences with single base insertions to cause protein translation to terminate early. The SspA PHP$^{84\text{-}86}$ > AAA$^{84\text{-}86}$ mutant is RNAP-binding deficient and thus does not enable the shift in promoter use ($\sigma^{70} \rightarrow \sigma^S$) (74). A single *sspA* rescue plasmid yielded no transformants into the Δ*sspA*::*kan*^{R} strain over multiple attempts. E. Top: deep-sequencing based measurement of the rates of new spacer acquisition in strains harbouring pSCL565 and, in the case of the Δ*sspA*::*kan*^{R}, either an empty plasmid or a low (~5) copy plasmid encoding variants of the *sspAB* operon as described in d., after growth for 48 h in liquid culture. Adaptation rates are shown relative to the wild-type parental strain. Open circles represent biological replicates ($n \geq 3$), bars are the mean. Horizontal dashed line represents the mean rate of spacer acquisition in the wild-type strain (one-way ANOVA effect of strain $P < 0.0001$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*sspA* $P < 0.0001$, Δ*sspA* + empty plasmid $P < 0.0001$, Δ*sspA* + *sspAB rescue* $P = 1$, Δ*sspA* + *sspA** (PHP84-86 > AAA84-86) & *sspB* rescue $P < 0.0001$; Δ*sspA* versus rescues, Δ*sspA* + empty vector $P = 0.997758$, Δ*sspA* + *sspA** (PHP84-86 > AAA84-86) & *sspB* $P = 0.334315$, Δ*sspA* + *sspAB* $P < 0.0001$, Δ*sspA* + *sspB* $P = 0.892991$, Δ*sspA* + *sspA** & *sspB** (frameshifted) $P = 1$). Bottom: representative agarose gel for the data shown. Expansions of the CRISPR array can be seen as higher sized bands above the parental array length. F. RT-qPCR of the fold-change in RNA copy number of *cas1* and *cas2* in the Δ*sspA* versus WT strains, measured relative to *mreB* and *GAPDH*. g. Volcano plot showing the $\log_2$ fold change in expression of genes in the Δ*sspA* versus WT strains versus their adjusted $-\log_{10}P$-values ($n = 3$ biological replicates). The horizontal dashed line represents an adjusted *P*-value of 0.05; the vertical lines represent $\log_2$ fold changes of –0.75 and 0.75. We identified nearly a thousand genes that were differentially expressed. Additional statistical details in Supplemental Table S1.

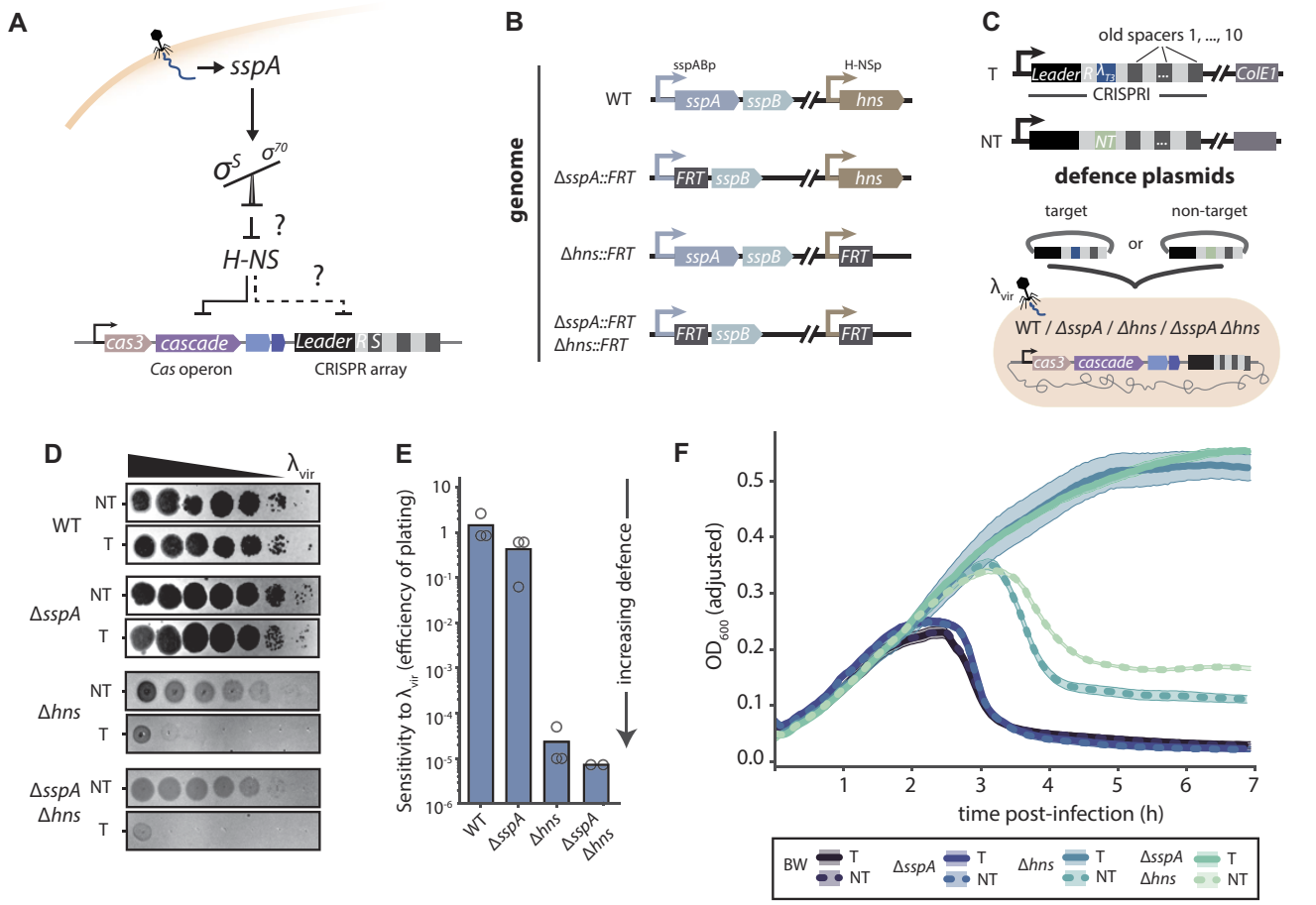


Figure 4. H-NS regulates CRISPR interference downstream of SspA. (**A**) Model for SspA-mediated regulation of CRISPR-Cas defence. Phage infection triggers upregulation of SspA (79), which in turn induces a global transcriptional shift towards σ^S-regulated promoters. This results in H-NS downregulation (72,80), induction of CRISPR-Cas mediated defence through de-repression Cas gene expression (51,81), leading to increased rates of CRISPR adaptation and interference. (**B**) Schematic of the *sspAB* and *hns* operons of WT, Δ*sspA*::*FRT*, Δ*hns*::*FRT* and Δ*sspA*::*FRT* Δ*hns*::*FRT* strains. *FRT*: flippase recognition target, a scar left after the removal of resistance cassettes. (**C**) Schematic of the CRISPR interference-mediated defence assays in pre-immunised *E. coli* strains. Top: schematic of the CRISPR-I immunisation (defence) plasmids. All plasmids are low (~5) copy and encode an *E. coli* CRISPR-I array with a first spacer encoding either a Target (complementary to the λ genome (1,51)), or a Non-Target (NT) spacer. Bottom: The experimental strains were electroporated with either the T or NT plasmid, and infected to varying titres of λ_{vir}. Note that the strains encode a complete endogenous *E. coli* Type I-E CRISPR-Cas system. (**D**) Representative plaque assays of λ_{vir} on experimental strains (described above) pre-immunised with either T or NT defence plasmids. Strains were infected with λ_{vir} and grown on plates at 30°C for 16 h. Full plaque assay plates for $n = 3$ biological replicates in Supplementary Figure S5. (**E**) Efficiency of plating of λ_{vir} on experimental strains; raw plaque counts and subsequent analysis in Supplemental Table S9. Open circles represent biological replicates ($n \geq 3$) of individual plaque assays, bars are the mean (one-way ANOVA effect of strain $P = 0.033454$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*sspA* $P = 0.181757$, Δ*hns* $P = 0.043319$, Δ*sspA* Δ*hns* $P = 0.043316$; for Δ*hns* versus Δ*sspA* Δ*hns* $P = 1$). E. Anti-phage defence and growth in overnight liquid culture of experimental strains, post λ_{vir} infection (MOI: 0.1). Hue around solid line (mean) represents the standard deviation across three biological replicates.

to the lambda genome (T: target (1,51)) or a control CRISPR array with a non-target first spacer (NT: non-target). Then, we infected these pre-immunised strains with varying titres of λ_{vir} and quantified phage defence (Figure 4C). Because of the pre-immunisation, this assay measures the ability of mutants to mount anti-phage defence via CRISPR interference and should be CRISPR adaptation-independent.

Plaque assays revealed that a wild-type strain was unable to mount defence against new rounds of infection even when pre-immunised with an anti-λ spacer, as reported previously (1,51,81) (Figure 4D). Pre-immunised Δ*sspA* mutants were similarly unable to defend against λ_{vir}. However, pre-immunised Δ*hns* mutants were capable of mounting considerable defence against λ_{vir}. Interestingly, we saw no differences in anti-λ_{vir} defence between the Δ*hns* and Δ*hns* Δ*sspA* pre-immunised mutants, suggesting that the CRISPR-Cas mediated anti-phage defence observed in the Δ*hns* Δ*sspA* mutants was determined solely by the lack of CRISPR interference repression by H-NS, and that Δ*sspA* has no additive effect on CRISPR interference-mediated anti-phage defence on the Δ*hns* background. Quantification of efficiency of plating confirmed these findings (Figure 4E), as did additional experiments measuring anti-phage defence in overnight liquid culture growth assays (Figure 4F). Together, these results suggest that H-NS and SspA are epistatic for CRISPR interference, with H-NS acting downstream of SspA on the regulation of CRISPR interference-mediated anti-phage defence.

SspA regulates CRISPR adaptation independently of H-NS

Given that SspA may regulate CRISPR interference via H-NS, we sought to determine whether SspA, in turn, regulates CRISPR adaptation via H-NS as well. To do so, we performed

deep sequencing of the CRISPR arrays from samples harvested 3 h post λ_{vir} infection in liquid cultures of wild-type, Δ*hns*, Δ*sspA* and Δ*hns* Δ*sspA* mutants harbouring either T or NT plasmids. We found no differences in the rates of CRISPR adaptation across conditions, except in the Δ*hns* + T cultures, which substantially increased rates of CRISPR adaptation (Figure 5A). These new spacers were primarily λ_{vir} derived (Figure 5B), and that the majority of the acquired spacers are found immediately downstream and on same strand as the immunising spacer, consistent with primed CRISPR adaptation (Figure 5C, D, Supplementary Figure S6A). Interestingly, we saw a substantial decrease in λ_{vir} derived spacers in the Δ*hns* Δ*sspA* + T conditions (Supplementary Figure S6B). Although the rates of CRISPR adaptation in the Δ*hns* + T condition were low (0.5% of CRISPR arrays expanded, i.e. five cells per thousand with a newly expanded array) and could not explain the defence demonstrated by the Δ*hns* + T cultures at the time of sample collection (Figure 4F), our results underscore the requirement for SspA for adequate primed CRISPR acquisition, in a closer-to-natural and defence-relevant setting (10,18,30).

We next sought to determine whether SspA modulates naïve CRISPR adaptation via H-NS. For this, we used the Δ*sspA::FRT*, Δ*hns::FRT* and Δ*sspA::FRT* Δ*hns::FRT E. coli* strains (Figure 5E), and assessed the mutants' ability to acquire new spacers after co-electroporation of pSCL565 alongside a low (∼5) copy rescue plasmid encoding the *sspAB* operon, *hns* operon, or both, under their native genomic contexts and regulation (Figure 5F). We found that Δ*sspA*, Δ*hns*, and Δ*sspA* Δ*hns* mutant strains all showed defects in CRISPR adaptation, with the double Δ*sspA* Δ*hns* mutant showing the strongest defect (Figure 5G). Complementation of the knockout strains with their respective rescue plasmids restored CRISPR adaptation to levels comparable to wild-type.

Since H-NS deletion de-represses CRISPR interference (Figure 4D–F), we hypothesised that its effect on CRISPR adaptation could be indirect, through the removal of cells that acquired genome-derived spacers via CRISPR interference-mediated self-targeting. To remove the confounding effect of increased self-targeting in the Δ*hns* background, we built Δ*cas3-cascade::cm*^{R} knockouts on top of the Δ*sspA* and Δ*hns* genetic backgrounds, and assessed the mutants' ability to acquire new spacers after electroporation with pSCL565. We found that although the Δ*sspA* Δ*cas3-cascade::cm*^{R} mutant still remained substantially CRISPR adaptation deficient, the Δ*hns* Δ*cas3-cascade::cm*^{R} mutant recovered CRISPR adaptation to levels comparable to wild-type (Figure 5H). This confirmed that the apparent CRISPR adaptation deficiency of the Δ*hns* mutant was caused by self-targeting through de-repression of CRISPR interference, and not additional effects on CRISPR adaptation. Taken together, our data supports a role for SspA in CRISPR adaptation that is independent of H-NS.

$P = 0.310715$; Δ*hns* versus Δ*hns* Δ*cas3-cascade* $P < 0.0001$; Δ*sspA* Δ*cas3-cascade* versus Δ*hns* Δ*cas3-cascade* $P < 0.0001$). Horizontal dashed line represents the mean rate of spacer acquisition in the wild-type strain. Additional statistical details in Supplemental Table S1.

Discussion

We developed a novel negative selection CRISPRi screen, designed around the concept of stimulated CRISPR self-immunity, to identify potential host factors that participate in CRISPR adaptation in *E. coli*. We identified a new host factor in our screen, SspA. In validation experiments, adaptation assays and downstream analysis of newly acquired spacers revealed that a *sspA* knockout mutant is consistently and significantly defective in naïve CRISPR adaptation, despite no other noticeable differences in the features of its acquired spacers when compared to the wild-type parental strain. Further, we found that mutations that abolish SspA's ability to bind to the RNAP complex cause a loss-of-adaptation phenotype, suggesting that SspA acts as a transcriptional-level regulator of CRISPR adaptation. A series of phage sensitivity and CRISPR adaptation assays revealed that SspA regulates CRISPR adaptation independently of H-NS, a known regulator of CRISPR interference-mediated anti-phage defence and a member of the SspA regulon. Taken together, our data support independent control of CRISPR adaptation and interference downstream of SspA.

Importantly, SspA does not regulate CRISPR adaptation through direct transcriptional regulation of the Cas1-Cas2 integrase expression (i.e. by regulation of their shared promoter), as rates of CRISPR adaptation are significantly decreased in ΔsspA strains that these proteins from non-native promoters. Rather, it is likely that SspA, by acting as a global transcriptional regulator of cellular stress, affects the expression of other coding or non-coding genes that in turn affect CRISPR adaptation.

We find that our data is consistent with a model where the immunisation and interference steps could occur separately, perhaps even temporally so. We speculate that phage infection could trigger the rapid accumulation of SspA (79), opening a window for the acquisition of new spacers; this window may close rapidly as the levels of SspA decline, but this sudden SspA accumulation may be enough to cause downregulation of H-NS (72), thus opening a second window for CRISPR interference to occur (Figure 6). However, more studies are required to determine whether the sudden accumulation of SspA in response to phage infection is a ubiquitous response beyond lambda, what phage element or phage-induced signal triggers this sudden spike in SspA levels, and whether this spike is indeed sufficient to significantly deplete levels of H-NS and open a window for CRISPR interference to occur. Further, though our data strongly suggest that SspA acts on CRISPR adaptation at a transcriptional level, additional work is needed to discover the target (s) of the SspA-mediated transcriptional rewiring.

Despite the success of our host factor screen in revealing SspA as a novel regulator of CRISPR adaptation, we have evidence that the screen also resulted in both false positive and false negatives, as is typical of genetic screens. For instance, genes that affect the copy number of plasmids (e.g. *pcnB*) can come through as false positives. Additionally, we were interested in host factors affecting adaptation, yet the screen we designed will capture genes that affect not only adaptation, but also CRISPR interference and even electroporation efficiency. Though we did assess our candidate hits' ability to acquire electroporated spacers, we did not assess each of the KOs for their ability to mediate CRISPR interference. An assay that is solely (or mostly) reliant on CRISPR-mediated spacer acquisition, such as those used recently to identify more active Cas1-Cas2 integrases (83), could help reduce these non-adaptation related hits.

False negatives might arise from at least two sources. First, from a screen design perspective, because we deliver pre-

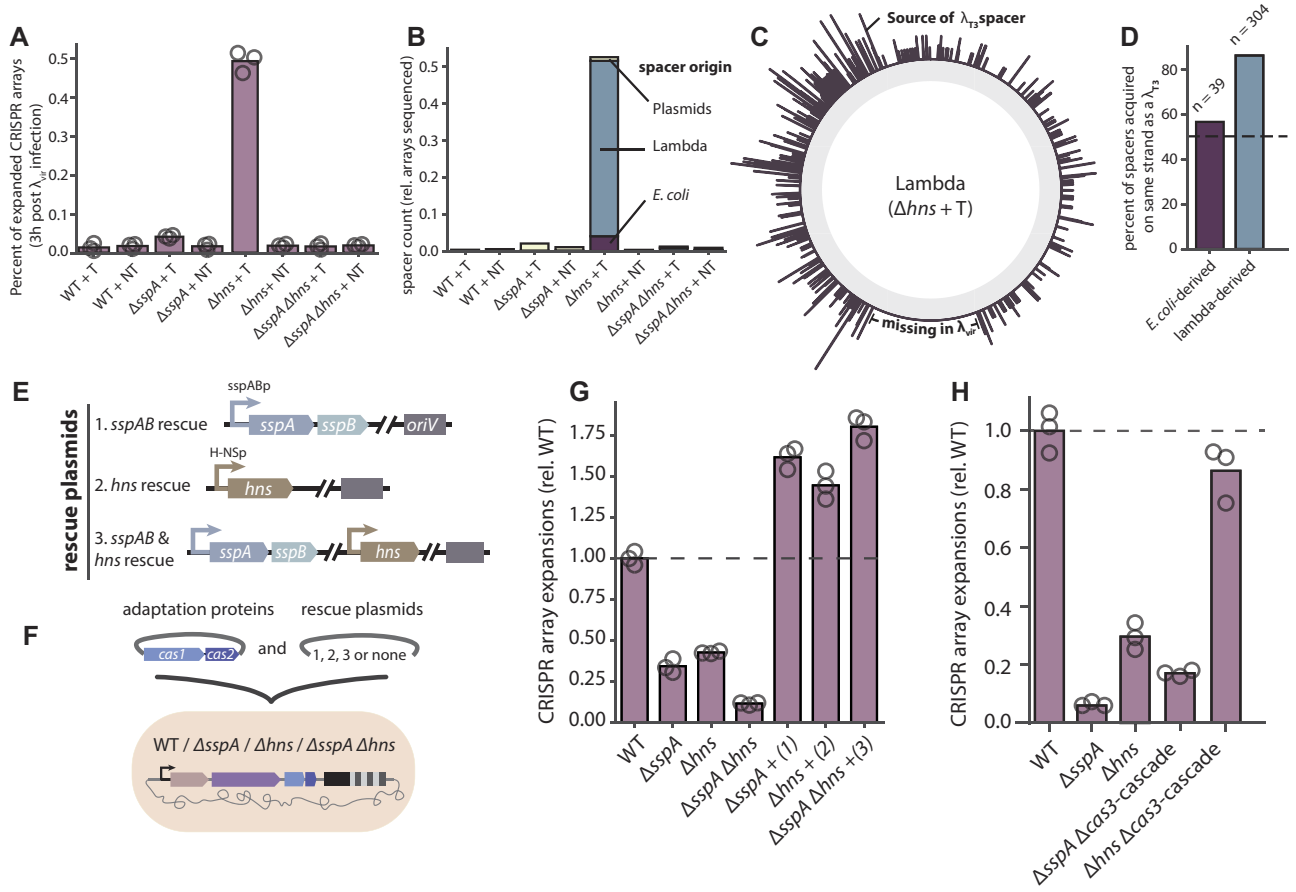


Figure 5. SspA regulates CRISPR adaptation independently of H-NS. (**A**) Deep-sequencing based measurement of the rates of new spacer acquisition in strains pre-immunised with either a T or NT defence plasmid, harvested 3 h post λ_{vir} infection in liquid culture and growth at 30°C. Open circles represent biological replicates ($n \geq 3$), bars are the mean (one-way ANOVA effect of strain $P < <0.0001$; Sidak's corrected multiple comparisons for wild-type + T versus knockouts + T, Δ*sspA* $P = 082553$, Δ*hns* $P < 0.0001$, Δ*sspA* Δ*hns* $P = 0.999999$; Δ*sspA* + T versus knockouts + T, Δ*hns* $P < 0.0001$, Δ*sspA* Δ*hns* $P = 0.154762$; Δ*hns* + T versus Δ*hns* + NT $P < 0.0001$; Δ*hns* + T versus Δ*sspA* Δ*hns* + T $P < 0.0001$). (**B**) Breakdown of normalised spacer count (total number of new spacers / number of CRISPR arrays sequenced) according to spacer origin (*E. coli*, lambda or plasmid) and strain of interest. (**C**) Binned coverage plot of Δ*hns* + T newly acquired spacers across the lambda genome (outer, purple). The location of the T immunisation spacer is shown on the lambda genome; 'missing in λ_{vir}' indicates a genomic region missing in our strain of λ_{vir}. (**D**) Percent of spacers acquired that are on the same strand as the T immunisation spacer, according to the spacer source (*E. coli* or lambda). (**E**) Schematic of the *sspAB* and *hns* operonic rescue plasmids. All plasmids are low (~5) copy, and encode either 1. The sspAB operon, 2. The hns operon, or 3. both, under their native regulation. (**F**) Schematic of the CRISPR adaptation assays in wild-type, *sspA* and/or *hns* mutant strains. Strains were electroporated with pSCL565 and rescue plasmids 1., 2. or 3. (see **E**), and assessed for their ability to acquire new spacers into the endogenous CRISPR I array. (**G**) PCR-based detection of new spacer acquisition into the CRISPR I array of wild-type, of WT, Δ*sspA*::*FRT*, Δ*hns*::*FRT* and Δ*sspA*::*FRT* Δ*hns*::*FRT* strains harbouring pSCL565 and rescue plasmids 1., 2. or 3. (see **E**), after growth for 48 h in liquid culture. Open circles represent biological replicates ($n \geq 3$), bars are the mean. Horizontal dashed line represents the mean rate of spacer acquisition in the wild-type strain (one-way ANOVA effect of strain $P < <0.0001$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*sspA* $P < 0.0001$, Δ*hns* $P < 0.0001$, Δ*sspA* Δ*hns* $P < 0.0001$; Δ*sspA* versus knockouts, Δ*hns* $P = 0.714182$, Δ*sspA* Δ*hns* $P = 0.002269$, Δ*sspA* + *sspAB* rescue $P < 0.0001$; Δ*hns* versus knockouts, Δ*sspA* Δ*hns* $P < 0.0001$, Δ*hns* + *hns* rescue $P < 0.0001$; Δ*sspA* Δ*hns* versus Δ*sspA* Δ*hns* + *sspA* & *hns* rescues $P < 0.0001$). (**H**) PCR-based detection of new spacer acquisition into the CRISPR I array of WT, Δ*sspA*::*FRT*, Δ*hns*::*FRT*, Δ*sspA*::*FRT* Δ*cas3-Cascade*::*Cm*^{R} or Δ*hns*::*FRT* Δ*cas3-Cascade*::*Cm*^{R} strains harbouring pSCL565 after growth for 48h in liquid culture. Open circles represent biological replicates ($n \geq 3$), bars are the mean (one-way ANOVA effect of strain $P < 0.0001$; Sidak's corrected multiple comparisons for wild-type versus knockouts, Δ*sspA* $P < 0.0001$, Δ*hns* $P < 0.0001$, Δ*sspA* Δ*cas3-cascade* $P < 0.0001$, Δ*hns* Δ*cas3-cascade* $P = 0.125466$; Δ*sspA* versus Δ*hns* $P = 0.004161$; Δ*sspA* versus Δ*sspA* Δ*cas3-cascade*

trimmed self-targeting prespacers to the library-carrying cells, we would not capture host factors involved in prespacer substrate generation. Nonetheless, this was a deliberate screen design choice for the kinds of host factors we hoped to capture (namely, genetic regulators of the process), and future iterations on our approach could be used to detect factors at earlier stages of the CRISPR adaptation process by delivering non-trimmed spacers to identify host factors involved in prespacer substrate generation. Additionally, future adaptations of our screen could harness a more targeted library to assess whether specific cellular pathways (i.e. DNA repair, stress responses) are involved in the regulation of CRISPR adaptation. Second, on a technical front, ineffective gRNAs that are partially or completely unable to target a given gene for repression and off-target gRNA effects that cause have been shown to cause pervasive toxicity in past CRISPRi screens in *E. coli* (43) can both contribute to false negatives, likely explaining the fact that we did not observe the known host factor IHF as a hit.

Another factor that could have influenced our signal / noise ratio was the choice of control condition used to contrast our screen hits to, to identify significantly enriched or depleted genes. We reasoned that using a –dCas9 strain carrying the sgRNA library would allow us to control for sgRNAs that were preferentially propagated or lost from the pool in ways

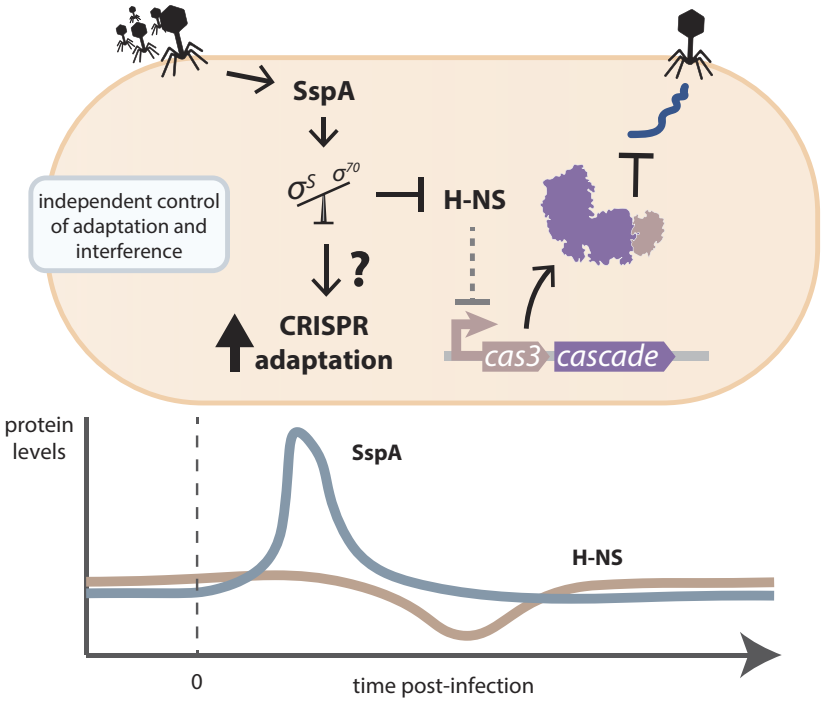


Figure 6. Proposed model for the independent control of CRISPR adaptation and interference. In both cases, the regulation of CRISPR immunity happens downstream of SspA, through its role as a global transcriptional rewiring agent.

not related to our screen. However, only controls for changes in sgRNA abundance that are not dCas9-dependent. Future uses of this screen should carefully consider the choice of additional appropriate controls and contrast conditions, such as the use of one or more non-targeting spacers, or uninduced dCas9 conditions.

One hit from our screen was *polA*, whose Klenow fragment was reported to be capable of repairing CRISPR arrays that have been cleared of the Cas1-Cas2 integrases *in vitro* (22). However, our results do not support this role. Though we found that CRISPR adaptation levels were significantly diminished in the *polA* ΔKlenow mutant, this decrease was attributable to the loss of acquisition of plasmid-derived spacers; The loss-of-adaptation phenotype seen in the Δ*pcnB* mutant is likely due to a similar effect, as *pcnB* has been shown to be required for copy number maintenance of ColE1 and other plasmids (84). We cannot, however, rule out a role for *polA* in array repair, though it is conceivable that there is redundancy in host factors capable of this task. Indeed, functional redundancy of host factors is a possible explanation for not capturing the comprehensive set of these proteins. We anticipate that more complex combinatorial knockdown and activation screens could be used to tackle this problem. Furthermore, we believe that pairing genetic screens such as our CRISPRi screen with orthogonal physical screens, such as proximity labelling and pull-down assays (85), will yield rich and informative datasets, which are likely to uncover a more comprehensive set of host factors required for CRISPR adaptation.

Data availability

All data supporting the findings of this study are available within the article and its supplementary information. Sequencing data associated with this study is available on NCBI SRA (PRJNA1109382). All code to replicate the analyses can be found in https://github.com/Shipman-Lab/CRISPRi_host_factor_screen and https://doi.org/10.5281/zenodo.14262115.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Kate Crawford and Alex González Delgado for experimental guidance, assistance and comments on the manuscript.

Author contributions: S.C.L. conceived the study, and with S.L.S., outlined the scope of the project and designed experiments. All experiments were performed and analysed by S.C.L., with technical assistance from Y.L. and K.A.Z for Figure 1E (Y.L.) and Figure 4f (K.A.Z.). S.C.L. and S.L.S. wrote the manuscript, with input from all authors.

Funding

National Science Foundation [MCB 2137692 to S.L.S.]; Pew Biomedical Scholars Program and the Chan Zuckerberg Biohub – San Francisco [to S.L.S.]; Berkeley Fellowship for Graduate Study [to S.C.L.].

Conflict of interest statement

None declared.

References

1. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S.,

Koonin,E.V. and Van Der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
2. 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.
3. Horvath,P. and Barrangou,R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
4. Terns,M.P. and Terns,R.M. (2011) CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.*, **14**, 321–327.
5. Wiedenheft,B., Sternberg,S.H. and Doudna,J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331–338.
6. Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR interference limits horizontal gene transfer in Staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
7. Semenova,E., Jore,M.M., Datsenko,K.A., Semenova,A., Westra,E.R., Wanner,B., Van Der Oost,J., Brouns,S.J.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.*, **108**, 10098–10103.
8. Marraffini,L.A. and Sontheimer,E.J. (2010) CRISPR interference: rNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181–190.
9. Sapranauskas,R., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.*, **39**, 9275–9282.
10. Jackson,S.A., McKenzie,R.E., Fagerlund,R.D., Kieper,S.N., Fineran,P.C. and Brouns,S.J.J. (2017) CRISPR-Cas: adapting to change. *Science*, **356**, eaal5056.
11. Sternberg,S.H., Richter,H., Charpentier,E. and Qimron,U. (2016) Adaptation in CRISPR-Cas Systems. *Mol. Cell*, **61**, 797–808.
12. Sasnauskas,G. and Siksnys,V. (2020) CRISPR adaptation from a structural perspective. *Curr. Opin. Struct. Biol.*, **65**, 17–25.
13. Barrangou,R. and Marraffini,L.A. (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*, **54**, 234–244.
14. Yosef,I., Goren,M.G. and Qimron,U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.*, **40**, 5569–5576.
15. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, **519**, 193–198.
16. Nuñez,J.K., Harrington,L.B., Kranzusch,P.J., Engelman,A.N. and Doudna,J.A. (2015) Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, **527**, 535–538.
17. Nuñez,J.K., Kranzusch,P.J., Noeske,J., Wright,A.V., Davies,C.W. and Doudna,J.A. (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.*, **21**, 528–534.
18. Ivančić-Baće,I., Cass,S.D., Wearne,S.J. and Bolt,E.L. (2015) Different genome stability proteins underpin primed and naïve adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.*, **43**, 10821–10830.
19. Killelea,T., Dimude,J.U., He,L., Stewart,A.L., Kemm,F.E., Radovčić,M., Ivančić-Baće,I., Rudolph,C.J. and Bolt,E.L. (2023) Cas1–Cas2 physically and functionally interacts with DnaK to modulate CRISPR Adaptation. *Nucleic Acids Res.*, **51**, 6914–6926.
20. Santiago-Frangos,A., Buyukyoruk,M., Wiegand,T., Krishna,P. and Wiedenheft,B. (2021) Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. *Curr. Biol.*, **31**, 3515–3524.
21. Wang,J.Y., Tuck,O.T., Skopintsev,P., Soczek,K.M., Li,G., Al-Shayeb,B., Zhou,J. and Doudna,J.A. (2023) Genome expansion by a CRISPR trimmer-integrase. *Nature*, **618**, 855–861.
22. Budhathoki,J.B., Xiao,Y., Schuler,G., Hu,C., Cheng,A., Ding,F. and Ke,A. (2020) Real-time observation of CRISPR spacer acquisition by Cas1–Cas2 integrase. *Nat. Struct. Mol. Biol.*, **27**, 489–499.
23. Malone,L.M., Hampton,H.G., Morgan,X.C. and Fineran,P.C. (2022) Type I CRISPR-Cas provides robust immunity but incomplete attenuation of phage-induced cellular stress. *Nucleic Acids Res.*, **50**, 160–174.
24. Fagerlund,R.D., Wilkinson,M.E., Klykov,O., Barendregt,A., Pearce,F.G., Kieper,S.N., Maxwell,H.W.R., Capolupo,A., Heck,A.J.R., Krause,K.L., *et al.* (2017) Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci.*, **114**, 5122–5128.
25. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505–510.
26. Modell,J.W., Jiang,W. and Marraffini,L.A. (2017) CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature*, **544**, 101–104.
27. Lee,H. and Sashital,D.G. (2022) Creating memories: molecular mechanisms of CRISPR adaptation. *Trends Biochem. Sci.*, **47**, 464–476.
28. Hu,C., Almendros,C., Nam,K.H., Costa,A.R., Vink,J.N.A., Haagsma,A.C., Bagde,S.R., Brouns,S.J.J. and Ke,A. (2021) Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature*, **598**, 515–520.
29. Kim,S., Loeff,L., Colombo,S., Jergic,S., Brouns,S.J.J. and Joo,C. (2020) Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature*, **579**, 141–145.
30. Shiriaeva,A.A., Kuznedelov,K., Fedorov,I., Musharova,O., Khvostikov,T., Tsoy,Y., Kurilovich,E., Smith,G.R., Semenova,E. and Severinov,K. (2022) Host nucleases generate prespacers for primed adaptation in the *E. coli* type I-E CRISPR-Cas system. *Sci. Adv.*, **8**, eabn8650.
31. Yoganand,K.N.R., Sivathanu,R., Nimkar,S. and Anand,B. (2017) Asymmetric positioning of Cas1–2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res.*, **45**, 367–381.
32. Nuñez,J.K., Bai,L., Harrington,L.B., Hinder,T.L. and Doudna,J.A. (2016) CRISPR immunological memory requires a host factor for specificity. *Mol. Cell*, **62**, 824–833.
33. Wright,A.V., Liu,J.-J., Knott,G.J., Doxzen,K.W., Nogales,E. and Doudna,J.A. (2017) Structures of the CRISPR genome integration complex. *Science*, **357**, 1113–1118.
34. Santiago-Frangos,A., Henriques,W.S., Wiegand,T., Gauvin,C.C., Buyukyoruk,M., Graham,A.B., Wilkinson,R.A., Triem,L., Neselu,K., Eng,E.T., *et al.* (2023) Structure reveals why genome folding is necessary for site-specific integration of foreign DNA into CRISPR arrays. *Nat. Struct. Mol. Biol.*, **30**, 1675–1685.
35. Patterson,A.G., Yevstigneyeva,M.S. and Fineran,P.C. (2017) Regulation of CRISPR–Cas adaptive immune systems. *Curr. Opin. Microbiol.*, **37**, 1–7.
36. Smith,L.M., Hampton,H.G., Yevstigneyeva,M.S., Mahler,M., Paquet,Z.S.M. and Fineran,P.C. (2024) CRISPR-Cas immunity is repressed by the LysR-type transcriptional regulator PigU. *Nucleic Acids Res.*, **52**, 755–768.
37. Smith,L.M., Jackson,S.A., Malone,L.M., Ussher,J.E., Gardner,P.P. and Fineran,P.C. (2021) The Rcs stress response inversely controls surface and CRISPR–Cas adaptive immunity to discriminate plasmids and phages. *Nat. Microbiol.*, **6**, 162–172.
38. Westra,E.R., Buckling,A. and Fineran,P.C. (2014) CRISPR–Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.*, **12**, 317–326.
39. Perez-Rodriguez,R., Haitjema,C., Huang,Q., Nam,K.H., Bernardis,S., Ke,A. and DeLisa,M.P. (2011) Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in Escherichia coli: envelope stress triggers CRISPR silencing. *Mol. Microbiol.*, **79**, 584–599.
40. Wright,A.V. and Doudna,J.A. (2016) Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.*, **23**, 876–883.

41. Schmidt,F., Cherepkova,M.Y. and Platt,R.J. (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, **562**, 380–385.
42. González-Delgado,A., Mestre,M.R., Martínez-Abarca,F. and Toro,N. (2019) Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR–Cas system in Vibrio vulnificus. *Nucleic Acids Res.*, **47**, 10202–10211.
43. Cui,L., Vigouroux,A., Rousset,F., Varet,H., Khanna,V. and Bikard,D. (2018) A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nat. Commun.*, **9**, 1912.
44. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
45. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci*, **97**, 6640–6645.
46. St-Pierre,F., Cui,L., Priest,D.G., Endy,D., Dodd,I.B. and Shearwin,K.E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.*, **2**, 537–541.
47. Zhang,J., Mahdi,A.A., Briggs,G.S. and Lloyd,R.G. (2010) Promoting and avoiding recombination: contrasting activities of the *Escherichia coli* RuvABC holliday junction resolvase and RecG DNA translocase. *Genetics*, **185**, 23–37.
48. Hossain,A.A., McGinn,J., Meeske,A.J., Modell,J.W. and Marraffini,L.A. (2021) Viral recombination systems limit CRISPR-Cas targeting through the generation of escape mutations. *Cell Host Microbe*, **29**, 1482–1495.
49. Su,M.-T., Venkatesh,T.V. and Bodmer,R. (1998) Large- and small-scale preparation of bacteriophage λ lysate and DNA. *BioTechniques*, **25**, 44–46.
50. Salgado,H. (2006) RegulonDB (version 5.0): escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.*, **34**, D394–D397.
51. Pougach,K., Semenova,E., Bogdanova,E., Datsenko,K.A., Djordjevic,M., Wanner,B.L. and Severinov,K. (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.*, **77**, 1367–1379.
52. Rousset,F., Cui,L., Siouve,E., Becavin,C., Depardieu,F. and Bikard,D. (2018) Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLoS Genet.*, **14**, e1007749.
53. Lear,S.K., Lopez,S.C., González-Delgado,A., Bhattarai-Kline,S. and Shipman,S.L. (2023) Temporally resolved transcriptional recording in E. coli DNA using a Retro-Cascorder. *Nat. Protoc.*, **18**, 1866–1892.
54. Shipman,S.L., Nivala,J., Macklis,J.D. and Church,G.M. (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science*, **353**, aaf1175.
55. 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.
56. Muzellec,B., Teleńczuk,M., Cabeli,V. and Andreux,M. (2022) PyDESeq2: a python package for bulk RNA-seq differential expression analysis. bioRxiv doi: https://doi.org/10.1101/2022.12.14.520412, 16 December 2022, preprint: not peer reviewed.
57. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
58. Granger,B.E. and Perez,F. (2021) Jupyter: thinking and Storytelling With Code and Data. *Comput. Sci. Eng.*, **23**, 7–14.
59. Joshi,N. and Fass,J. (2011) sickle - A windowed adaptive trimming tool for FASTQ files using quality. Available at: https://github.com/najoshi/sickle.
60. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.
61. Shimoyama,Y. (2022) pyCirclize: circular visualization in Python. Available at: https://github.com/moshi4/pyCirclize.
62. Tareen,A. and Kinney,J.B. (2020) Logomaker: beautiful sequence logos in Python. *Bioinformatics*, **36**, 2272–2274.
63. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
64. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
65. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
66. Klopfenstein,D.V., Zhang,L., Pedersen,B.S., Ramírez,F., Warwick Vesztrocy,A., Naldi,A., Mungall,C.J., Yunes,J.M., Botvinnik,O., Weigel,M., *et al.* (2018) GOATOOLS: a Python library for gene ontology analyses. *Sci. Rep.*, **8**, 10872.
67. Almendros,C., Mojica,F.J.M., Díez-Villaseñor,C., Guzmán,N.M. and García-Martínez,J. (2014) CRISPR-Cas functional module exchange in Escherichia coli. *mBio*, **5**, e00767-13.
68. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
69. Alexander,D.L., Lilly,J., Hernandez,J., Romsdahl,J., Troll,C.J. and Camps,M. (2014) Random mutagenesis by error-prone pol plasmid replication in Escherichia coli. In: Gillam,E.M.J., Copp,J.N. and Ackerley,D. (eds.) *Directed Evolution Library Creation, Methods in Molecular Biology*. Springer New York, NY, Vol. **1179**, pp. 31–44.
70. Reeh,S., Pedersen,S. and Friesen,J.D. (1976) Biosynthetic regulation of individual proteins in relA + and relA strains of Escherichia coli during amino acid starvation. *Mol. Gen. Genet. MGG*, **149**, 279–289.
71. Hansen,A.-M., Gu,Y., Li,M., Andrykovitch,M., Waugh,D.S., Jin,D.J. and Ji,X. (2005) Structural basis for the function of stringent starvation protein a as a transcription factor. *J. Biol. Chem.*, **280**, 17380–17391.
72. Hansen,A., Qiu,Y., Yeh,N., Blattner,F.R., Durfee,T. and Jin,D.J. (2005) SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. *Mol. Microbiol.*, **56**, 719–734.
73. Ishihama,A. and Saitoh,T. (1979) Subunits of RNA polymerase in function and structure. *J. Mol. Biol.*, **129**, 517–530.
74. Wang,F., Shi,J., He,D., Tong,B., Zhang,C., Wen,A., Zhang,Y., Feng,Y. and Lin,W. (2020) Structural basis for transcription inhibition by E. coli SspA. *Nucleic Acids Res.*, **48**, 9931–9942.
75. Travis,B.A., Ramsey,K.M., Prezioso,S.M., Tallo,T., Wandzilak,J.M., Hsu,A., Borgnia,M., Bartesaghi,A., Dove,S.L., Brennan,R.G., *et al.* (2021) Structural basis for virulence activation of francisella tularensis. *Mol. Cell*, **81**, 139–152.
76. Lou,J., Cai,J., Hu,X., Liang,Y., Sun,Y., Zhu,Y., Meng,Q., Zhu,T., Gao,H., Yu,Z., *et al.* (2022) The stringent starvation protein SspA modulates peptidoglycan synthesis by regulating the expression of peptidoglycan synthases. *Mol. Microbiol.*, **118**, 716–730.
77. Levchenko,I., Seidel,M., Sauer,R.T. and Baker,T.A. (2000) A specificity-enhancing factor for the ClpXP degradation machine. *Science*, **289**, 2354–2356.
78. Hansen,A., Lehnherr,H., Wang,X., Mobley,V. and Jin,D.J. (2003) *Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. *Mol. Microbiol.*, **48**, 1621–1631.
79. Drahos,D.J. and Hendrix,R.W. (1982) Effect of bacteriophage lambda infection on synthesis of groE protein and other Escherichia coli proteins. *J. Bacteriol.*, **149**, 1050–1063.

80. Hansen,A.-M. and Jin,D. (2012) SspA up-regulates gene expression of the LEE pathogenicity island by decreasing H-NS levels in enterohemorrhagic Escherichia coli. *BMC Microbiol.*, **12**, 231.
81. Westra,E.R., Pul,Ü., Heidrich,N., Jore,M.M., Lundgren,M., Stratmann,T., Wurm,R., Raine,A., Mescher,M., Van Heereveld,L., *et al.* (2010) H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.*, **77**, 1380–1393.
82. Pul,Ü., Wurm,R., Arslan,Z., Geißen,R., Hofmann,N. and Wagner,R. (2010) Identification and characterization of *E. coli* CRISPR- *cas* promoters and their silencing by H-NS. *Mol. Microbiol.*, **75**, 1495–1512.
83. Hofmann,R., Herman,C., Mo,C.Y., Mathai,J. and Marraffini,L.A. (2024) Deep mutational scanning identifies variants of Cas1 and Cas2 that increase spacer acquisition in type II-A CRISPR-Cas systems. bioRxiv doi: https://doi.org/10.1101/2024.10.10.617623, 10 October 2024, preprint: not peer reviewed.
84. Masters,M., Colloms,M.D., Oliver,I.R., He,L., Macnaughton,E.J. and Charters,Y. (1993) The pcnB gene of Escherichia coli, which is required for ColE1 copy number maintenance, is dispensable. *J. Bacteriol.*, **175**, 4405–4413.
85. Santin,Y.G., Doan,T., Lebrun,R., Espinosa,L., Journet,L. and Cascales,E. (2018) In vivo TssA proximity labelling during type VI secretion biogenesis reveals TagA as a protein that stops and holds the sheath. *Nat. Microbiol.*, **3**, 1304–1313.

Received: May 27, 2024. **Revised:** November 23, 2024. **Editorial Decision:** November 29, 2024. **Accepted:** December 4, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.