nature communications

Article

https://doi.org/10.1038/s41467-024-55215-9

Bacteroides expand the functional versatility of a conserved transcription factor and transcribed DNA to program capsule diversity

Received: 2 July 2024

Accepted: 2 December 2024

Published online: 30 December 2024

Check for updates

Jason Saba[1,2], Katia Flores[3,4], Bailey Marshall[1,5], Michael D. Engstrom[1], Yikai Peng[1], Atharv S. Garje[1,6], Laurie E. Comstock[3,4] & Robert Landick[1,7]

The genomes of human gut bacteria in the genus *Bacteroides* include numerous operons for biosynthesis of diverse capsular polysaccharides (CPSs). The first two genes of each CPS operon encode a locus-specific paralog of transcription elongation factor NusG (called UpxY), which enhances transcript elongation, and a UpxZ protein that inhibits noncognate UpxYs. This process, together with promoter inversions, ensures that a single CPS operon is transcribed in most cells. Here, we use in-vivo nascent-RNA sequencing and promoter-less in-vitro transcription (PIVoT) to show that UpxY recognizes a paused RNA polymerase via sequences in both the exposed non-template DNA and the upstream duplex DNA. UpxY association is aided by 'pause-then-escape' nascent RNA hairpins. UpxZ binds non-cognate UpxYs to directly inhibit UpxY association. This UpxY-UpxZ hierarchical regulatory program allows *Bacteroides* to generate subpopulations of cells producing diverse CPSs for optimal fitness.

Bacteroides are abundant and crucial members of the modern human gut microbiota. A key evolved feature of these bacteria is the ability of each strain to produce numerous (eight or more) distinct capsular polysaccharides (CPS)[1,2] that are tightly regulated so that only one CPS is typically produced per bacterial cell. This bet-hedging strategy generates *Bacteroides* populations with great surface variability that protect from phage[3–5] and mediate immune modulation, biofilm formation, antibiotic resistance, and inflammation[6–11].

CPS diversity is achieved by regulating both transcription initiation and elongation of CPS biosynthesis operons. *Bacteroides fragilis* (*Bfr*) has eight distinct CPS operons, producing PSA–PSH. All but PSC use invertible promoters and all encode *upxY* (Y_X) and *upxZ* (Z_X) paralogs as the first genes in each operon[12,13]. The fraction of each promoter oriented ON versus OFF varies with environmental conditions[14]. CPS promoter inversions are stochastic and multiple CPS promoters are oriented ON in most cells simultaneously[15–17]. *Bacteroides* prioritize expression of one promoter-ON CPS operon over others by regulating RNA polymerase (RNAP) elongation via the operon-specific Y_X elongation activator and Z_X inhibitor of non-cognate Y_X. Z_X inhibits a subset of non-cognate Y_X possibly via direct binding (e.g., $Z_{\underline{A}}$ from PS<u>A</u> may inhibit $Y_{\underline{E}}$ from PS<u>E</u>). *Bfr* Y_X paralogs must distinguish among eight target CPS loci to enable operon-specific regulation, but how this discrimination is accomplished is unknown.

Y_X family proteins are specialized (i.e., locus-specific) paralogs of NusG/Spt5, the only universal transcription factor found in archaea, eukaryotes, and bacteria[18]. NusG-family regulators bind RNAPs during

[1]Department of Biochemistry, University of Wisconsin–Madison, Madison, WI, USA. [2]Microbiology Doctoral Training Program, University of Wisconsin–Madison, Madison, WI, USA. [3]Department of Microbiology, University of Chicago, Chicago, IL, USA. [4]Duchossois Family Institute, University of Chicago, Chicago, IL, USA. [5]Cell and Molecular Biology Training Program, University of Wisconsin–Madison, Madison, WI, USA. [6]Genetics Training Program, University of Wisconsin–Madison, Madison, WI, USA. [7]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, USA.
e-mail: rlandick@wisc.edu

https://doi.org/10.1038/s41467-024-55215-9

transcript elongation and modulate RNAP activity through interactions with the RNAP and the surface-exposed ntDNA strand[19–21]. Globally acting *Escherichia coli* NusG and its single specialized paralog RfaH increase elongation rate and decrease pausing[22–25]. In contrast, *Bacillus subtilis*, *Mycobacterium tuberculosis*, and *Thermus thermophilus* NusGs enhance both pausing and intrinsic termination[26–30]. Pausing during transcript elongation is a universal regulatory feature of RNAPs that allows site-specific recruitment of transcription factors (TFs)[31] and guides RNA synthesis.

Among the known $NusG_{SP}$ families, RfaH of Proteobacteria is the best understood. RfaH targets operons that contain a DNA element called *ops* (operon polarity suppressor) in their leader regions (DNA between the transcription start site and the translation start codon of the first gene). RNAP pauses at the 12-nucleotide *ops*, allowing RfaH to associate via sequence-specific interactions with a non-template strand DNA hairpin (ntDNAhp) exposed by the paused RNAP[22,23,32,33]. Other $NusG_{SP}$ include LoaP in Firmicutes[21], TaA in Myxococcota[34], and plasmid-encoded ActX in Proteobacteria[35].

The CPS operon leader regions are required for Y-mediated regulation[12], consistent with sequence-specific Y_X recruitment to RNAP paused in this region (Fig. 1a). In principle, Y_X could recognize ntDNA (like RfaH), nascent RNA (like LoaP), or both to discriminate among

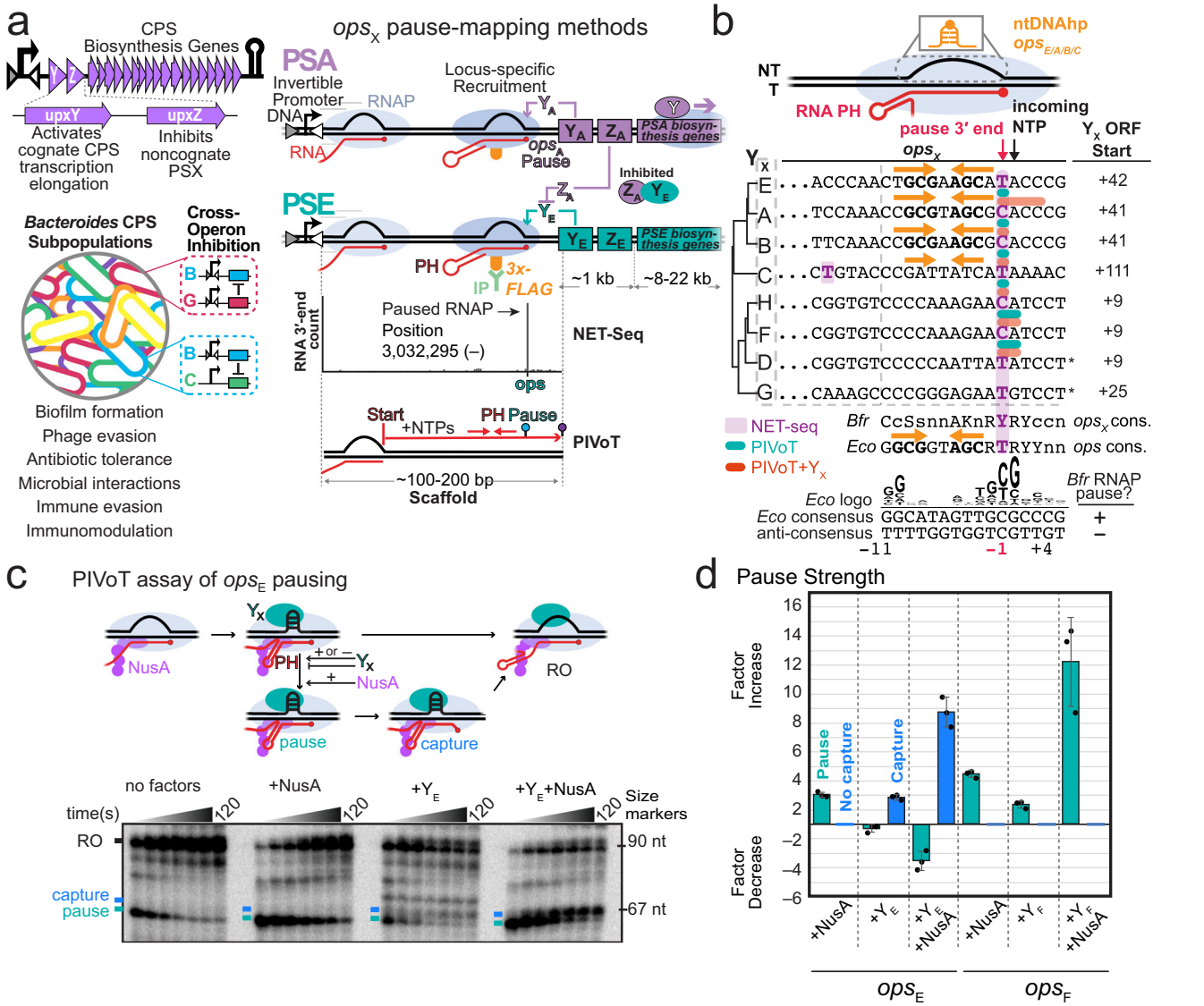

Fig. 1 | Bacteroides fragilis RNAP pauses in CPS operon leader regions in vivo and in vitro at candidate Y_X recruitment sites called ops_X. a Representative CPS operon diagram highlighting Y_X and Z_X, the first two genes in *B. fragilis* PSX operons. Horizontal triangles mark the inverted repeats recognized by Mpi recombinase for promoter inversion[17]. Proposed roles for CPS diversity in *B. fragilis* subpopulations (colored coats) are listed[3,6,8–10,99]. The schematics illustrate the proposed roles of Y_X activation and Z_X inhibition of noncognate Y_X in generating subpopulation CPS diversity[12,13]. Y_X is recruited to RNAP paused at cognate but not non-cognate ops_X sites that encode a pause hairpin (PH). Z_X directly binds Y_X from heterologous operons and inhibits its recruitment. In vivo (NET-seq) and in vitro (PIVoT) methods for identifying RNAP pause sites (ops_X) are illustrated. **b** Transcriptional pauses in CPS leader regions identified in this study are shown in comparison to the RfaH *ops* pause and the *E. coli* consensus elemental pause sequences[37]. T template strand, NT non-template strand. Fully conserved nucleotides are capitalized; largely conserved nucleotides are lowercase. Asterisks (*) indicate operons were only probed in vivo. **c** PIVoT assay of PSE promoter-distal leader regions. Assays included 1 μM NusA or 150 nM Y_E added concomitantly with NTPs as indicated. RNAs from a reaction time course were separated by 8% Urea-PAGE. **d** NusA and Y_X synergistic activities at cognate ops_X sites. Y_X association manifests as pause inhibition or pause enhancement (aqua bars), or capture (blue bars). Fold changes in pausing or capture (if applicable) after addition of factors are shown relative to baseline (no factors control). Data are presented as mean values ± SD from $n = 3$ independent experiments.

multiple, similar CPS operon targets. We used both in vivo and in vitro analyses to identify pauses in CPS operon leader regions, establish that these pause sites function as recruitment sites for Y, and discover $NusG_{SP}$–DNA interactions and mechanisms that mediate Y–CPS operon specificity. We found that Z directly binds noncognate Ys to block Y action and that differential Y_X–Z_X affinities enable CPS hierarchical control of transcript elongation. These results define mechanisms that explain the exquisite specificity of multiple $NusG_{SP}$ and that allow *Bacteroides* to program CPS diversity in the highly dynamic human gut environment.

Results

Bacteroides fragilis RNAP pauses in CPS operon leader regions in vivo and in vitro at candidate Y_X-recruitment sites (ops_X)

Specific Y_X recruitment sites likely exist in CPS leader regions because these leader sequences are variable and are required for Y_X activity[12]. Since *Eco*RfaH is recruited to RNAP at leader region *ops* pause sites, we first asked if *Bfr*RNAP pauses in the leader regions of CPS operons. To identify candidate Y_X-recruiting pause sites directly in vivo, we used nascent elongating transcript sequencing (NET-seq) (Fig. 1a, b and Supplementary Fig. 1a). NET-seq allows genome-scale identification of precise nascent RNA 3′ ends, which are enriched at pause sites[36,37].

NET-seq revealed single prominent pause sites in most CPS operon leader regions (Fig. 1a, b and Supplementary Fig. 1b)[37]. Eight CPS leader pauses exhibited an obvious consensus sequence that resembles strong *E. coli* pauses (Fig. 1b) as well as apparent nascent RNA pause hairpins (PHs) that resemble those known to enhance pausing allosterically in concert with NusA in other bacteria (e.g., the so-called type-1 *E. coli his* and *B. subtilis trp* leader region pauses; Supplementary Fig. 1c)[37–40]. Pausing in the PSC leader region (the only *Bfr* CPS operon with a non-invertible, constitutively ON promoter)[41] occurred at multiple sites; weak pausing occurred at a site resembling the other seven in sequence and location (Fig. 1b and Supplementary Fig. 1b). We designated the CPS leader pause sites ops_X ('X' designates the CPS operon) based on analogy to the RfaH *ops* site.

To test whether the ops_X pause recruits Y_X, we generated recombinant *Bacteroides fragilis* RNAP (r*Bfr*RNAP) and assayed CPS leader regions using promoter-less in vitro transcription (PIVoT) (Fig. 1a and Supplementary Fig. 2a, b)[42,43]. PIVoT bypasses the need for σ^A-dependent initiation. We first asked if r*Bfr*RNAP recognizes the consensus elemental pause signal defined for *Eco*RNAP (Fig. 1b)[37]. Signals resembling this consensus direct pausing by a wide variety of RNAPs from bacteria to human[37,44,45]. Bacterial pause sequences are reported to differ in some species[46,47] and have not been tested for Bacteroidota. We found that r*Bfr*RNAP pauses strongly at the consensus sequence but not anti-consensus sequence (Supplementary Fig. 2c), suggesting its pause signals resemble those of *Eco*RNAP and most other tested RNAPs.

We next assayed pausing in for six of the eight CPS leader regions (PSA, B, C, E, F, and H). Strikingly, the PSA, B, E, F, and H leader segments encoded prominent pause sites that corresponded exactly to the sites found by NET-seq (Supplementary Figs. 2d, 3). Pausing was less prominent but detectable at ops_C, consistent with the heterogeneous pausing observed the NET-seq. We conclude that CPS operon leader regions encode strong pause sites for RNAP with similar but not identical sequences, as might be expected for Y_X recruitment sites that must distinguish among Y_X paralogs.

To ask if the CPS leader pauses function as targets for Y_X recruitment and test whether they are modulated by regulators like NusA and Y_X, we purified recombinant *Bfr*NusA and Y_X for these six PSA operons (Y_A, Y_B, Y_C, Y_E, Y_F, and Y_H; Methods) and tested their effects on pausing using PIVoT. In *Eco* and *Bsu*, NusA stimulates pausing in part via contacts to PHs[37,39,43,48–50]. All six CPS leader pauses were greatly enhanced by NusA (Fig. 1c and Supplementary Fig. 4). Intriguingly, $Y_{A,B,E}$ inhibited the cognate leader pause, whereas $Y_{C,F,H}$ enhanced the cognate leader pause (Fig. 1c, d and Supplementary Fig. 4). Y_E additionally trapped a fraction of RNAP just downstream from the pause site, as seen previously with *Eco*RfaH ('capture' in Fig. 1c, d and Supplementary Fig. 4). Thus, Y_X association with paused elongation complexes (PECs) may manifest as either pro-pausing or anti-pausing activity.

Importantly, the effects of Y_X are likely to be specific to the NET-seq identified leader pauses, consistent with ops_X sites functioning as specific Y_X-recruitment sites. Y_X only modulated pausing at cognate ops_X but not non-cognate ops_X or other positions (Supplementary Fig. 4). We conclude that the NET-seq-identified leader pauses are bona fide target sites for Y_X association with *Bfr*RNAP. Notably, $ops_{A,B,E}$ encode putative ntDNAhps at [−11 to +1] that resemble the *ops* ntDNAhp known to recruit *Eco*RfaH (5′-GCG–AGC stems; Fig. 1b and Supplementary Fig. 1c). The *Bfr* ops_X ntDNAhp sequences differ, consistent with specific recruitment of cognate Y_X. However, $ops_{F,H}$ are identical in the ntDNAhp region, suggesting that some other element contributes to specificity.

RNAP capture by Y_X–ops_X interaction, which is evident by accumulation of RNAs a few nucleotides longer than the primary pause RNA for ops_E but not ops_A or ops_B (Supplementary Fig. 4), suggests some but not all ops_X sites exhibit pause cycling[31,33,51]. Pause cycling occurs when the ntDNA is captured by a regulator that also contacts RNAP (e.g., *Eco*σ^{70} or RfaH), anchoring the PEC and hindering extension beyond 2–3 nt[52,53]. Trapped PECs can be rescued by RNA cleavage factors GreA,B[33], creating a cycle that repeats until ntDNA contacts rearrange to allow normal elongation[51].

Importantly, even in the presence of globally acting *Bfr*NusG, Y_F still enhances ops_F pausing (Supplementary Fig. 5a). Thus, Y_X appears to outcompete *Bfr*NusG even though both NusG and its specialized paralog Y_X use the same primary binding site on RNAP (Supplementary Fig. 5b).

Z_X inhibits Y_X at ops_X through direct Z_X–Y_X interaction

We next sought to test whether Y_X binding requires sequence upstream of the putative ntDNAhp region using in vitro binding, in silico interaction, and in vivo gene expression assays. Y_E is predicted to be inhibited by Z_A but not by Z_E or Z_C in a strain with only the PSA, PSE, and PSC promoters oriented ON (expression hierarchy PSA > E > C)[13,17]. We call this strain $[AE]_{ON}$ for simplicity because the PSC promoter is constitutive[13]. To test our prediction, we measured Z_A–Y_E.and Z_E–Y_E binding constants by biolayer interferometry (BLI) (Fig. 2a, b). Z_A but not Z_E bound tightly to Y_E (K_D ~ 0.9 nM vs ~88 nM). We conclude that Z_X acts through direct Y_X binding.

To understand how Z_A might interact with Y_E, we predicted their association using AlphaFold 3[54] (Fig. 2c and Supplementary Fig. 6). The Z_A–Y_E complex, which was predicted with high confidence, placed Z_X on the RNAP-binding interface of Y_E. When modeled into an *Eco*RNAP-RfaH-*ops*-PEC (PDB 8PHK)[33] by alignment of the Y_E NGN domain with the RfaH NGN, Z_A clashed with two major PEC features: (*i*) the RNAP clamp helices (CH), which provide the primary RNAP binding site for all NusG-family regulators (Fig. 2c, orange); and (*ii*) the proximal upstream DNA duplex (usDNA). Thus, Z_X likely inhibits Y_X by preventing its recruitment to RNAP at ops_X pause sites.

We next used PIVoT to test whether Z_A or Z_E blocked Y_E inhibition of pausing at the candidate ops_E pause site as predicted by the AlphaFold model. Z_E blocked Y_E action only at high concentrations (K_I approximating the K_D measured by BLI; Fig. 2d). In contrast, Z_A inhibited Y_E at all tested concentrations. We conclude that differential Y_X–Z_X affinities enable CPS hierarchical control of transcript elongation (Fig. 2e).

Y_X targets extended ops_X sites in vivo

Using these insights into Z_X–Y_X interaction, we tested whether ops_X pause sites function as Y_X recruitment sites in vivo and which

https://doi.org/10.1038/s41467-024-55215-9

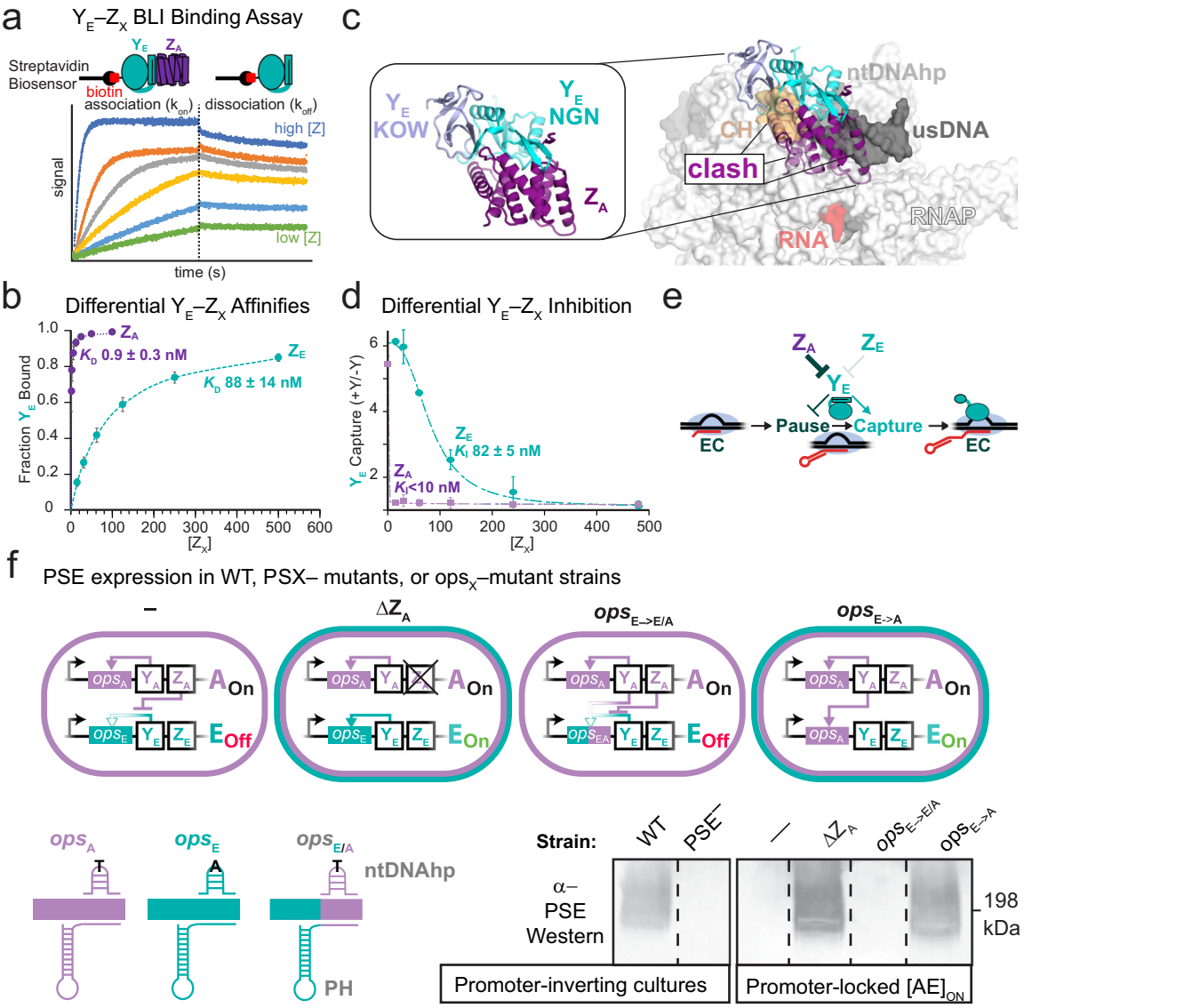

Fig. 2 | *ops*$_X$ pause sites are recruitment sites in vivo that enable Y_X-locus specificity, CPS hierarchical control, and can be re-wired to bypass direct inhibition by Z_X. **a** Z_X directly binds Y_X as revealed by biolayer interferometry (BLI[100]) over a range of Z_X concentrations yielding biotin-Y_E–Z_X on and off rates: Y_E–Z_A $k_{on} = 1.1 \times 10^6 \pm 2.9 \times 10^5$ $M^{-1}s^{-1}$, $k_{off} = 9.0 \times 10^{-4} \pm 2.4 \times 10^{-4}$ s^{-1}; Y_E–Z_E $k_{on} = 1.1 \times 10^5 \pm 2.5 \times 10^4$ $M^{-1}s^{-1}$, $k_{off} = 9.6 \times 10^{-3} \pm 7.3 \times 10^{-4}$ s^{-1}. Assays were performed in triplicate and globally fit to a 1:1 binding model (see Methods). Reported K_Ds are averages from three independent global fits and errors represent standard deviations. **b** Y_E–Z_A and Y_E–Z_E binding curves. The control group used buffer in place of Z_X and was subtracted from all binding curves shown (see Methods). Data are presented as mean values ± SD from $n = 3$ independent experiments. **c** AlphaFold3[54] model of Y_E–Z_A and steric clash evident when Y_E–Z_A is aligned to an RfaH-bound PEC[23]. **d** Fold changes in Y_E capture as a function of cognate Z_E or non-cognate Z_A concentration in PIVoT assays (Methods). Subsets of NusA, Y_E, variable [Z_X], and NTPs were added to initiate pause assays (50 nM Y_E, 15–480 nM Z_X, 1 μM NusA, final). The control group used buffer in place of Z_X and was subtracted from all binding curves shown (see Methods). Data are presented as mean values ± SD from $n = 3$ independent experiments. **e** Model for Z_X inhibition of Y_X recruitment. **f** Strain background used in *ops*$_X$ replacement experiments are depicted (Δ*mpi* M44 in each strain ensured only the PSA, PSC, and PSE promoters are oriented ON). In WT, promoter orientations are variable in single cells, but some cells express PSE. PSE$^-$ is an insertion mutant that abrogates PSE expression. In promoter-locked [AE_{ON}] strains, only Y_A-activated genes are expressed because of cross-operon inhibition of Y_E and Y_C by Z_A. Strains with partial [−10:−1]$_E$ or full [−38:−1]$_E$ segments of *ops*$_E$ were replaced with their *ops*$_A$ counterparts ([−10:−1]$_A$ and [−43:−1]$_A$) and assayed for their ability to rescue PSE expression by Western blot.

sequences govern cognate Y_X function. Using a constitutive [$AE]_{ON}$ strain[17], we replaced *ops*$_E$ segments with the corresponding *ops*$_A$ segments. We predicted that the *ops*$_E$–*ops*$_A$ swapped strain should activate PSE expression because Y_A should bind *ops*$_A$ in PSE. To ask if the PH-encoding region of *ops*$_X$ is required for Y_X recruitment, we also constructed a hybrid *ops*$_{E–A}$ strain in which only the ntDNAhp region corresponding to the RfaH *ops* but not the PH-encoding region of *ops*$_E$ was replaced with *ops*$_A$ sequence (Fig. 2f). Using antibodies confirmed to detect PSE in a WT strain but not in a PSE$^-$ mutant, we tested for PSE expression in [$AE]_{ON}$ and derivative strains: ΔZ_A, hybrid *ops*$_{E–A}$, and full *ops*$_{E→A}$ (Fig. 2f). PSE was (*i*) not expressed in [$AE]_{ON}$; (*ii*) expressed in ΔZ_A; (*iii*) not expressed in the hybrid *ops*$_{E–A}$ strain; and expressed in the full *ops*$_{E→A}$ swapped strain.

To confirm that the upstream PH-encoding region is required for Y_X action, we also tested Y_A and Y_E effects similarly using PIVoT (Supplementary Fig. 7a, b). Neither Y_A nor Y_E modulated pausing or PEC capture at WT levels unless the full cognate *ops*$_X$ including the upstream PH-encoding region was present. Thus, both in vivo and

https://doi.org/10.1038/s41467-024-55215-9

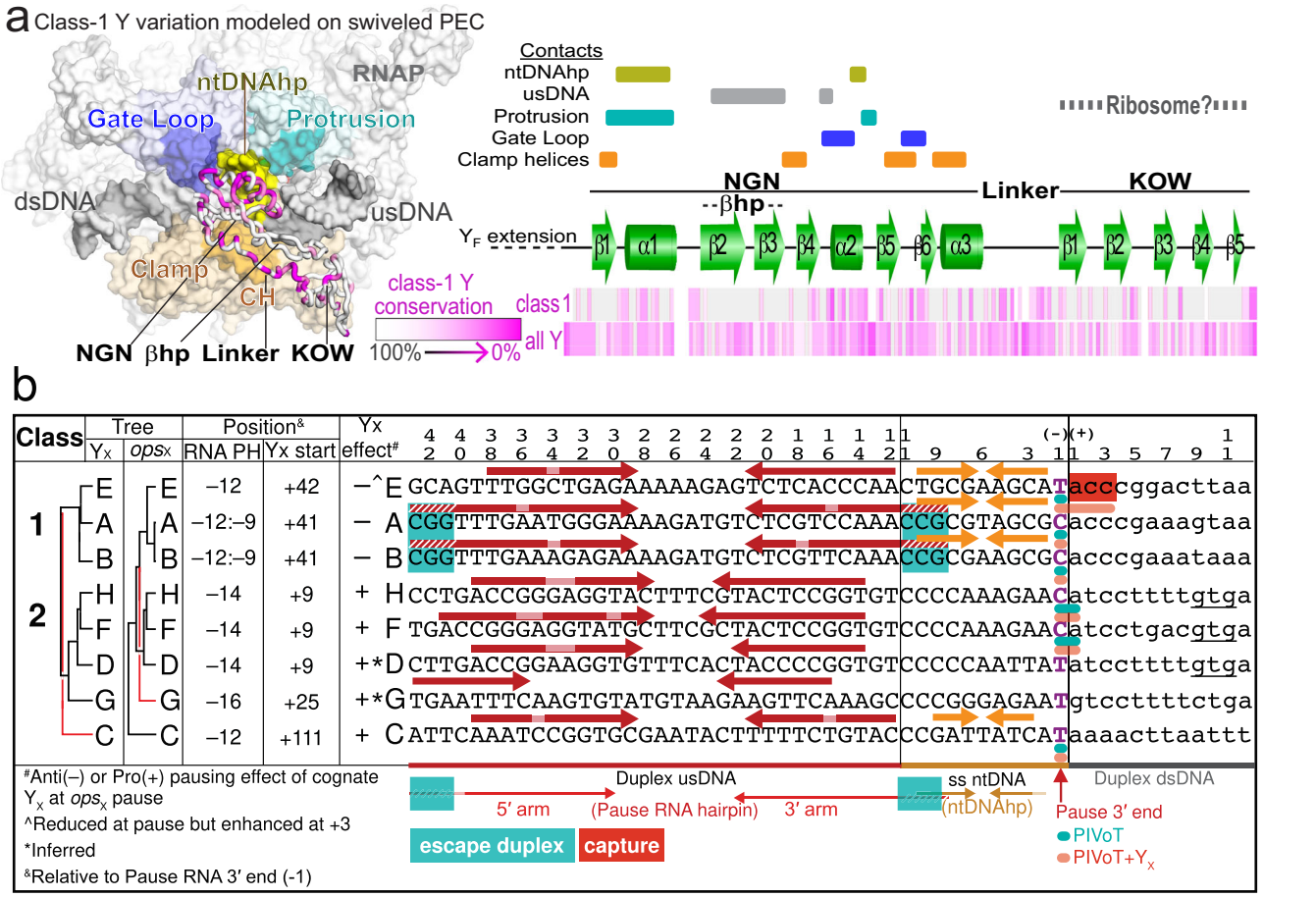

Fig. 3 | Y_X can be divided into distinct classes. a Analysis of sequence conservation and solvent accessibility of CPS operon Y_Xs from *B. fragilis* (Genbank accession NC_003228.3; strain NCTC 9343). Known contacts to RNAP modules or DNA are based on structures of *E. coli* RNAP in complex with NusG (PDB 6C6U)[23] or RfaH (PDB 6C6S)[23] (bars on right). The structural model, based on PDB 8PHK[33], depicts key $NusG_{SP}$-interacting modules of a PEC (gate loop, protrusion, clamp, ntDNAhp, upstream and downstream DNA (usDNA and dsDNA) and features of $NusG_{SP}$ (NGN, KOW, β hairpin). The extent of sequence conservation among Class-1Y_X is shown on a magenta color scale mapped to RfaH in the 8PHK model and also compared linearly to conservation among all Y_X proteins. **b** Sequence comparisons among ops_X annotated with features relevant to pausing and Y_X action compared to phylograms[101] of Y_X and ops_X shown on right. The red lines indicate the alternative clustering of ops_C and ops_G versus Y_C and Y_G relative to the uniform clustering of other Y_X and ops_X sequences into Class 1 and Class 2.

in vitro, the cognate upstream PH-encoding region is required for full Y_X activity.

We conclude that ops_X is comprised of both the ntDNAhp region and the upstream PH-encoding region. These regions are necessary and sufficient to program Y_X recruitment and enhancement of CPS-operon transcription. The inactivity of Y_X at hybrid sites establishes that the ~40 bp *Bacteroides* CPS ops_X sequences differ fundamentally from the RfaH *ops* that requires only a 12-bp ntDNAhp sequence. Additional recognition of the upstream PH-encoding region likely aids Y_X discrimination among target sites. However, determining whether these upstream sequences contact Y_X as a nascent RNA hairpin, as proposed for LoaP[55], or as duplex DNA required further experimentation.

Y_X–ops_X pairs can be divided into distinct classes

To ask if the variability in ops_X sequences could be related to variability in Y_X paralogs, we compared their apparent evolutionary relationships to sequence and structural alignments of Y_X, RfaH, and NusGs (Fig. 3a and Supplementary Fig. 8). Strikingly, both Y_X protein and ops_X DNA sequences clustered into two distinct classes with two outliers (anti-pausing Class-1, PSA,B,E; pro-pausing Class-2, PSD,F,H; Outliers PSG,C) (Fig. 3b and Supplementary Fig. 8). We use the ops_X pause site defined as position −1 as a reference in this analysis. Class-1 DNA–RNA sequences exhibited several key features: (*i*) an apparent ntDNAhp (orange arrows); (*ii*) an apparent PH that extends to −12 to −9 (red arrows; relative to −1 pause RNA 3′ nucleotide position); and (*iii*) the Y_X gene start codon is at +41, +42. Class-1 Y_X protein sequences (Fig. 3a) exhibited (*i*) an identical β2–β3 hairpin sequence in the NGN domain (LPTQFVIRQLYKRR[R/K]RVEVP); (*ii*) variable sequences (pink) in NGN α1 and α2 that contact the *ops* ntDNAhp (yellow), RNAP protrusion, and RNAP gate loop; and (*iii*) variability in the C-terminal KOW domain (Fig. 3a and Supplementary Fig. 8). The variable Y_X sequences in contacts to the ntDNAhp, protrusion, and gate loop are consistent with Y_X recognition and potential effects on pausing[27,56,57], whereas variability in KOW may enable target specificity or coupling of transcription to other cellular processes.

The Class-1 PSA,B PHs have greater potential to extend towards the pause RNA 3′ end (teal highlight) relative to the PSE PH. Extension of PHs past −10 is thought to destabilize PECs at intrinsic terminators[58], but we did not observe termination at these sites. An alternative role of PHs extending past −10 could be to aid PEC escape from pause cycles if auxiliary factors like GreA,B are insufficient. Thus, we postulated that base-pairing of the PSA,B PHs at −11, −10, and −9 could explain why, in contrast to Y_E, Y_A, and Y_B did not capture PECs in pause cycles

(Supplementary Fig. 4 and Fig. 3b red highlight) (see next section). Based on an apparent ability to prevent PEC capture by Y_X, we call this PH extension the escape duplex (ED).

Pro-pausing Class-2 (PSD,F,H) sequences exhibited features that differed from Class-1 (Fig. 3b and Supplementary Fig. 8). For Class-2 DNA-RNA: (*i*) ops_X lacks an obvious ntDNAhp; (*ii*) the apparent PH extends only to −14; and (*iii*) the Y_X gene start codon is at +9 relative to ops_X. For Class-2 Y_X: (*i*) the β2–β3 hairpin sequence is variable with pattern of basic residues distinct from Class-1; (*ii*) NGN α1 and α2 also are variable but distinct from Class-1 and thus consistent with differential recognition and different effects on pausing; and (*iii*) the KOW domain exhibits greatly increased positive charge relative to Class-1 (Supplementary Fig. 8).

PSC,G were outliers whose Y_X and ops_X clustered differently relative to Class-1,2. Their apparent PHs extended to −12 or −16, respectively. The Y_X start codons were at +111, +25 and both Y_X sequences were relatively divergent compared to Class-1,2. Y_C enhanced rather than inhibited the ops_C pause (Supplementary Fig. 4). Class-2 Y_X and PSC Y_C exhibited charge similarity to the LoaP KOW proposed to bind RNA hairpins (Supplementary Fig. 8).

We conclude that Y_X regulators diverged during evolution to form at least two distinct classes within which the interactions that determine Y_X–ops_X specificity and pro- vs. anti-pausing action appear to have followed different trajectories.

ops_X PHs stabilize PECs but also can aid escape of PECs captured by Y_X-DNA contacts

We next sought to assess the function of the putative ops_X PHs (Fig. 3b). We focused on Class-1 ops_X to investigate the impact of the PH and ED (Fig. 3b and Supplementary Fig. 9). The strong effect of NusA on Class-1 pauses (Fig. 1d and Supplementary Fig. 4) made it likely the PHs stimulate pausing[39,43,48–50,59]. Further, removal of the PH-encoding region from an ops_E scaffold eliminated NusA-stimulation of pausing (Supplementary Fig. 10a). To probe the functions of the conventional ops_E PH and the unconventional ops_B PH + ED, we used complementary antisense oligonucleotides (asDNAs or asRNAs) to progressively disrupt the 5′ arm of the PSE,B PHs (Fig. 4a, c).

asDNAs that disrupt the PSE PH by pairing with the 5′ arm, but not those that pair just upstream, reduced pausing (Fig. 4b). Thus, the PH alone stimulates pausing at ops_X and *Bfr*NusA significantly stimulates pausing in a PH-dependent manner. We conclude that ops_X sites are type-1 pauses that encode NusA-stabilized PHs, in notable contrast to the type-2 RfaH *ops* that lacks a PH[38].

To test the idea that the apparent ED could aid escape of PECs, we measured the effect on capture of antisense RNAs (asRNAs) that disrupt the ED by pairing to the distal bases of 5′ arm of the ops_B PH. ops_B but not ops_E encodes an ED, and Y_B does not cause PEC capture in contrast to Y_E (Fig. 4c, d and Supplementary Fig. 4). Addition of asRNAs that progressively disrupted the ED caused Y_B to capture PECs in pause cycles. Thus, ops_B, and by analogy ops_A, PHs not only stimulate ops_X pausing synergistically with NusA to allow time for Y_X recruitment, but also use an ED to drive forward translocation at the pause. The ED breaks extensive contacts by Y_X necessary for its initial recruitment but problematic for subsequent EC escape.

Y_X distinguishes PECs via multipartite NGN interactions with exposed ntDNA and upstream duplex DNA

We next sought to determine how Class-1 Y_X proteins distinguish cognate vs. non-cognate ops_X sites via the PH-encoding region (Fig. 2). Since the ntDNA of ops_E and ops_B are most similar, particularly at the key −6 ntDNAhp position (Fig. 3b and Supplementary Fig. 10b), we reasoned that the contribution of sequences upstream from the ntDNAhp might be most apparent by swapping regions between ops_E and ops_B. We used PIVoT to measure Y_X effects on NusA-stimulated pausing and capture using templates with $ops_{E–B}$ swapped sequences or Y_E–Y_B hybrid proteins that separate potential NGN vs. KOW contributions (Fig. 5a). To enable the direct comparison between ops_E and ops_B, we used a variant of ops_B that lacked the ED ($ops_{B,\ -ED}$).

To ask if Y_X recognizes the upstream DNA or the PH RNA encoded by it, we first tested whether the upstream DNA sequences affected Y_X action in the absence of a PH (Fig. 5b). With the PH removed, Y_B stimulated RNAP capture at the ops_B pause site by a factor of ~4.5 (Fig. 5c). When 3-bp segments of the ops_B usDNA were replaced with ops_E sequence, Y_B capture of RNAP decreased either modestly (substitutions 1 and 2) or nearly completely (substitution 3). We next asked if changing both the upstream DNA and the PH from ops_B to ops_E sequences had more effect on Y_B action than changing just the upstream DNA, as predicted if the PH functions in Y_B action. However, even combining the 1 + 2 + 3 substitutions in the upstream DNA and PH had no greater defect in Y_B action than introducing substitution 3 to the upstream DNA alone (Fig. 5d). We conclude that Y_B recognition of the extended ops_B site depends on the usDNA and not on the PH RNA.

We next investigated the contributions of the upstream sequences in progressively interconverted ops_E and ops_B to PEC capture by Y_X (Fig. 5e and Supplementary Fig. 11). To simplify this comparison, we used a variant of ops_B in which capture was activated by removing the ED (Supplementary Figs. 9, 11). Strikingly, Y_E continued to function even when the ops_E ntDNAhp was changed to the ops_B ntDNAhp. However, the Y_E effect was mostly lost and Y_B capture progressively increased as the usDNA was increasingly converted to ops_B sequence (Fig. 5e). Thus, multiple segments of usDNA contribute to Y_B recognition of ops_B. Consistent with our in vivo experiments (Fig. 2f), we conclude that ops_X sequences are multipartite ntDNA and usDNA signals of ~40 nucleotides whose constituent parts variably contribute to Y_X recruitment in different CPS operons.

We next asked if the NGN alone recognizes ops_X as it does for RfaH–*ops* interaction[23,32] or if the KOW domain might also participate, as proposed for LoaP[55]. Attempts to purify a Class-1 NGN alone yielded only insoluble protein. Instead, we compared NGN–KOW $Y_{E–B}$ hybrids to Y_E and Y_B on ops_E, ops_B, and an $ops_{E\text{-}B}$ hybrid scaffold (Fig. 5f). For both Y_E and Y_B, the effect on capture or pausing was determined completely by the NGN domains. We conclude that recognition of ops_X by at least Class-1 Y_X is mediated by the NGN and not the KOW domain.

Class-1 Y_X protects upstream DNA from exonucleolytic cleavage

For the Y_X NGN to contact upstream duplex DNA, the DNA must distort from a canonical B-form trajectory departing the PEC (Fig. 6a). Although protein interactions can easily bend duplex DNA[60], we sought direct physical evidence for usDNA–Y_X-NGN interaction. Exonuclease III (ExoIII) has been used extensively to detect PEC boundaries on DNA[61–63]. Since $Y_{E,B}$ variably depend on distal usDNA in our activity assays, we assayed $ops_{E,B}$ with cognate Y_X.

Over the full time course, $Y_{E,B}$ strongly stabilized a −21 footprint, 6–7 base pairs upstream of RNAP (Supplementary Fig. 12). However, Y_B but not Y_E also slowed ExoIII digestion at −24, and −31 to −34. Further, these same upstream protections were caused by a $Y_{B,E}$ NGN–KOW hybrid (Fig. 6b and Supplementary Fig. 12). We conclude that Y_B NGN likely contacts usDNA at least near −21 to −24, and −31 to −34.

As an additional test of the upstream Y_B contacts, we performed ExoIII assays on scaffolds containing ops_B-to-ops_E sequence changes to distal usDNA (−36 to −34 and −26 to −24) and proximal usDNA (−18 to −16). These substitutions strongly reduced upstream protection from ExoIII (Fig. 6c and Supplementary Fig. 13). Together, our results suggest a set of Y_X specificity determinants reflected in both physical contacts detected with ExoIII and sequence effects on Y_X activity.

To understand these contacts in a structural context, we modeled Y_E and Y_B into an RfaH-*ops*-PEC structure (PDB 8PHK)[33]. Both Y_E and Y_B are predicted to have a much larger positively charged surface approximately in the path of the usDNA (Supplementary Fig. 14). This charge is created largely by basic residues in the beta hairpin mini-

https://doi.org/10.1038/s41467-024-55215-9

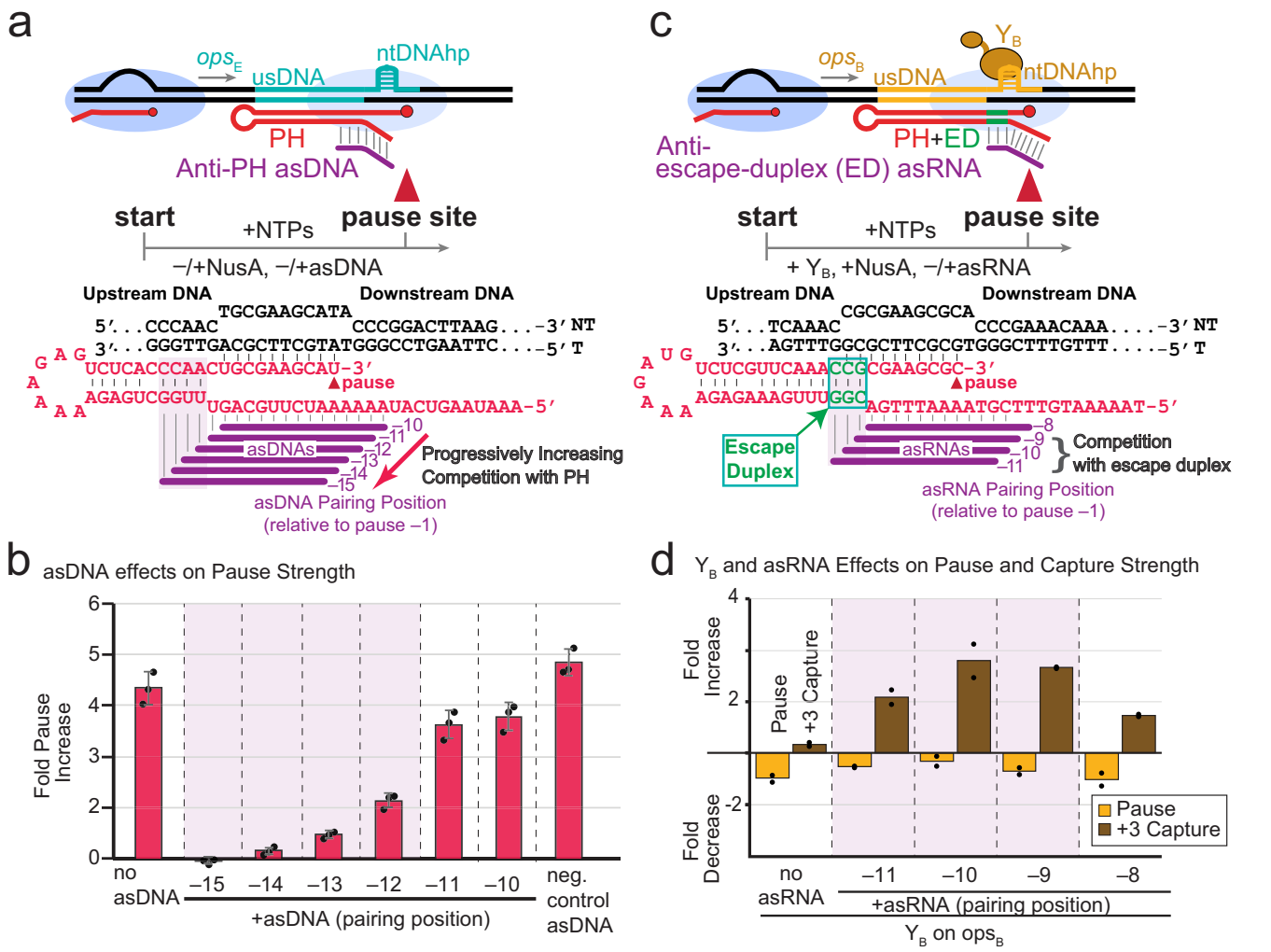

Fig. 4 | Nascent RNA hairpins promote pausing or pausing-then-escape at ops_X. **a** Experimental scheme. r*Bfr*RNAP was reconstituted upstream of the ops_E pause, enabling PH formation upon RNA extension. **b** Antisense DNA (asDNA; 10 μM final) effects on NusA enhancement of PH-stimulated ops_E pausing, where different asDNAs disrupt PH formation to different extents. asDNA oligonucleotides were added concomitantly with NusA (or storage buffer) and NTPs (1 μM and 100 μM each NTP, final). Data are presented as mean values ± SD from $n = 3$ independent experiments. **c** Experimental scheme. r*Bfr*RNAP was reconstituted upstream of ops_B pause, enabling PH formation. **d** Antisense RNAs (asRNAs; 0.5 μM final) pairing with an escape duplex (ED, green; the ED is unique to PSA and PSB). asRNAs inhibited PEC escape, leading to accumulation of a captured RNA (ops_B + 3 nt). Assays were performed in the presence of 1 μM NusA and 150 nM Y_B. The amount of RNA paused or captured 45 s after addition of NTPs is shown. Fold changes are relative to plus NusA and minus Y_B. Data are presented as mean values with data points from $n = 2$ independent experiments shown.

domain of $Y_{E,B}$ and could position the usDNA for sequence-specific readout by NGN.

Discussion

Human gut *Bacteroides* strains synthesize numerous surface CPS that are highly regulated to create subpopulations in which primarily a single PS locus is transcribed, providing phenotypic plasticity to environmental challenges. To coordinate CPS gene expression in a manner that maximizes CPS diversity, *Bacteroides* have developed a complex hierarchy involving locus-specific cognate Y_X activation and noncognate Z_X inhibition.

We have elucidated the biochemical mechanisms of *Bacteroides* CPS hierarchical control (Fig. 7): (*i*) *Bfr*RNAP pauses prominently at single CPS leader-region pause sites (ops_X); (*ii*) ops_X programs NusA-enhanced, RNA hairpin-stabilized transcriptional pauses that create time windows for Y_X recruitment; (*iii*) Z_X inhibits non-cognate Y_X directly via differential binding affinities, forming a heterodimer that precludes Y_X recruitment by steric clash of Z_X with RNAP and ops_X; (*iv*) Y_X locus-specific recruitment depends on multipartite interactions of the Y_X NGN domain with the exposed ops_X ntDNA and upstream duplex DNA; (*v*) Y_Xs evolved into functionally distinct classes; and (*vi*) Y_X-bound PECs use different mechanisms to escape ops_X. This combination of multiple functions at a single RNAP pause site has little precedent and may reflect the strong evolutionary pressure associated with the challenges of discriminating among multiple similar $NusG_{SP}$s.

Bacteroides belong to the greater phylum Bacteroidota, evolutionarily distant from the commonly studied model proteobacterium *E. coli* and firmicute *B. subtilis*. Despite the importance of these bacteria to human health, there is a limited understanding of *Bacteroides* transcription regulation. Our recombinant *Bfr*RNAP overexpression system enables facile production and genetic manipulation of *Bfr*RNAP. Multiple questions can now be addressed, including the roles of uncharacterized RNAP sequence insertions[64], the molecular interactions of RNAP with TFs (e.g., σ^A) and small molecules (e.g., ppGpp), and sequence-dependent effects on transcriptional activities (e.g., backtracking, translocation, etc.). Recombinant RNAPs enable studies of both lineage-specific transcription mechanisms and evolutionary comparisons. r*Bfr*RNAP will enhance mechanistic understanding in the

https://doi.org/10.1038/s41467-024-55215-9

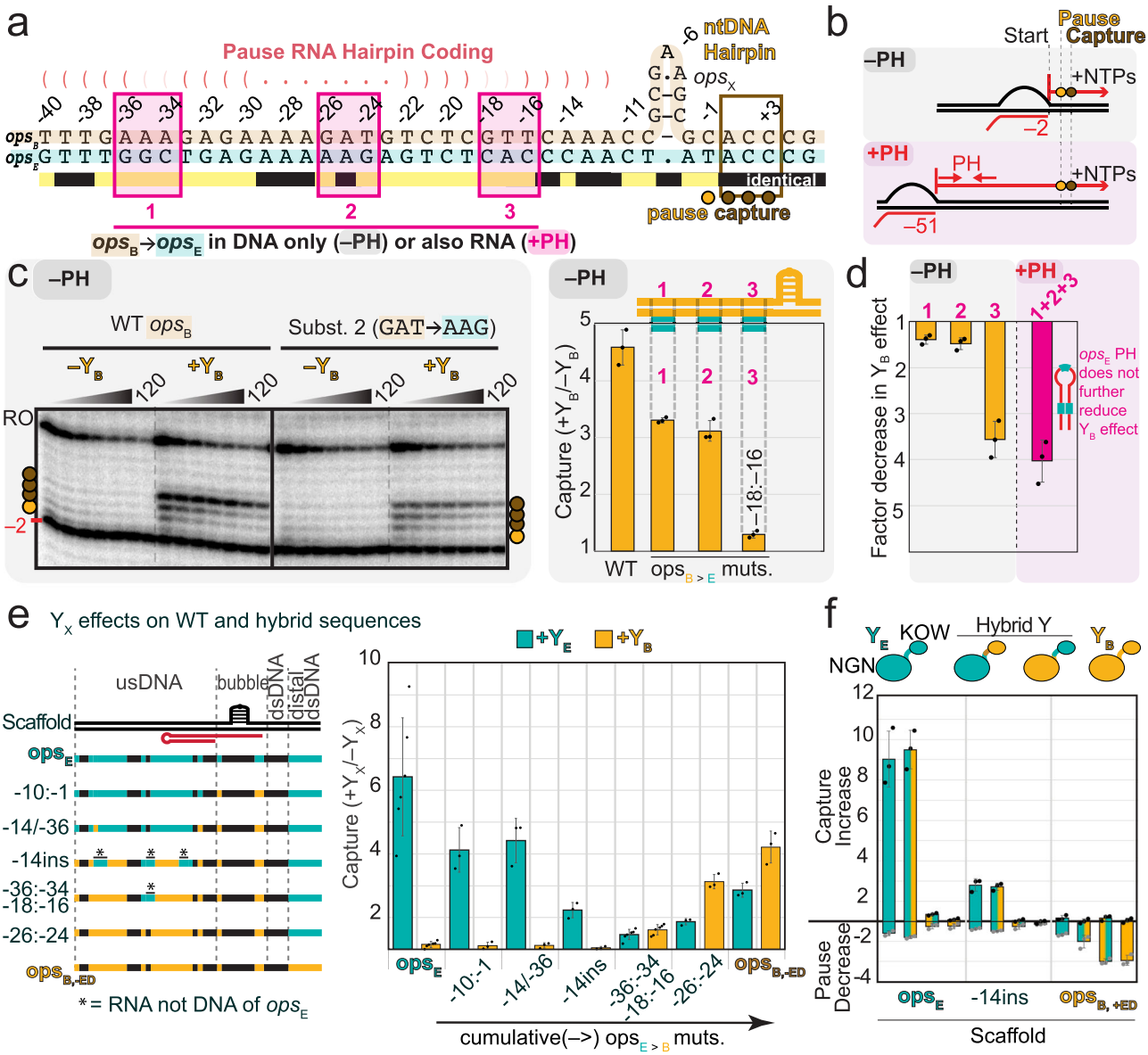

Fig. 5 | Y_X distinguish ops_X binding sites through variations in the non-template DNA and upstream duplex DNA. a Diagram comparing Class 1 ops_E and ops_B sequences. Regions varied in experiments shown in (**b**–**d**) are highlighted in magenta. **b** Experimental scheme to assay effects of the usDNA and PH. **c** PIVoT assays (1 μM NusA, 100 μM each NTP, 150 nM Y_X when added) comparing Y_B fold effects on capture on WT versus mutant scaffolds. Y_B causes capture in these experiments because native RNA structures are not permitted to form (including the escape duplex (ED)). The gel panel depicts time-courses of pausing on WT ops_B vs a mutant having substitution 2 from (**a**). Right: results from a representative single time-point (i.e., 10 s after NTP addition) comparison of Y_B effects across multiple mutant scaffolds. Data are presented as mean values ± SD from $n = 3$ independent experiments. **d** Comparison of Y_B effects in the absence or presence of a PH (the ops_B scaffold plus PH was designed to prohibit ED formation to enable this comparison). Data are presented as mean values ± SD from $n = 3$ independent experiments. **e** Y_X fold effect on pausing at the 45 s time point on WT ops_E or hybrid sequences progressively mutated from ops_E towards ops_B with the ED disrupted (see Supplementary Fig. 11 for scaffold sequence). PIVoT assays were performed at a single timepoint (45 s). Data are presented as mean values ± SD from $n = 3$ or $n = 6$ independent experiments. **f** PIVoT assay of pause and capture for WT vs Hybrid NGN–KOW Y_X (150 nM each) on WT vs hybrid (labeled '–14ins') ops_X sequences. Data are presented as mean values ± SD from $n = 3$ independent experiments.

entire field of transcription, as demonstrated by numerous recent studies in *M. tuberculosis*, *C. difficile*, and *B. subtilis*[26–28,65,66].

We found that ops_X recruitment sites for Y_X are ~40 bp multipartite DNA elements with both upstream duplex and transcription bubble ntDNA components, in striking contrast to the 12-nucleotide ntDNAhp (*ops*) necessary for RfaH-recruitment and the proposed nascent RNA hairpin necessary for LoaP recruitment[32,55]. The ntDNAhps formed by *ops* and ops_X differ in apparent structure and position relative to the pause site. All eight *ops* sites in *E. coli* targeted by the single RfaH encode the same ntDNAhp sequence: 5′-GCGGTAGC[67]. The longer *Bacteroides* 5′-YGCGNAGCR ntDNAhps exhibit both similar (GCG..AGC stems) and distinct (loop) features compared to *ops*. These differences highlight how *Bacteroides* evolved to manage numerous $NusG_{SP}$. Extensive Y_X–ops_X interactions may also accelerate *Bacteroides* adaptation by expanding the sequence space available for functional bifurcation following gene duplication.

We found that Z_X inhibits Y_X recruitment to ops_X–PECs directly, likely by blocking Y_X interaction with the conserved β′ clamp helices (CH) and the ops_X usDNA. Z_X could also tune heterologous operon PSX expression or limit self-expression through negative feedback. Ultimately, Y_X–Z_X interactions define the cell surface architecture of

https://doi.org/10.1038/s41467-024-55215-9

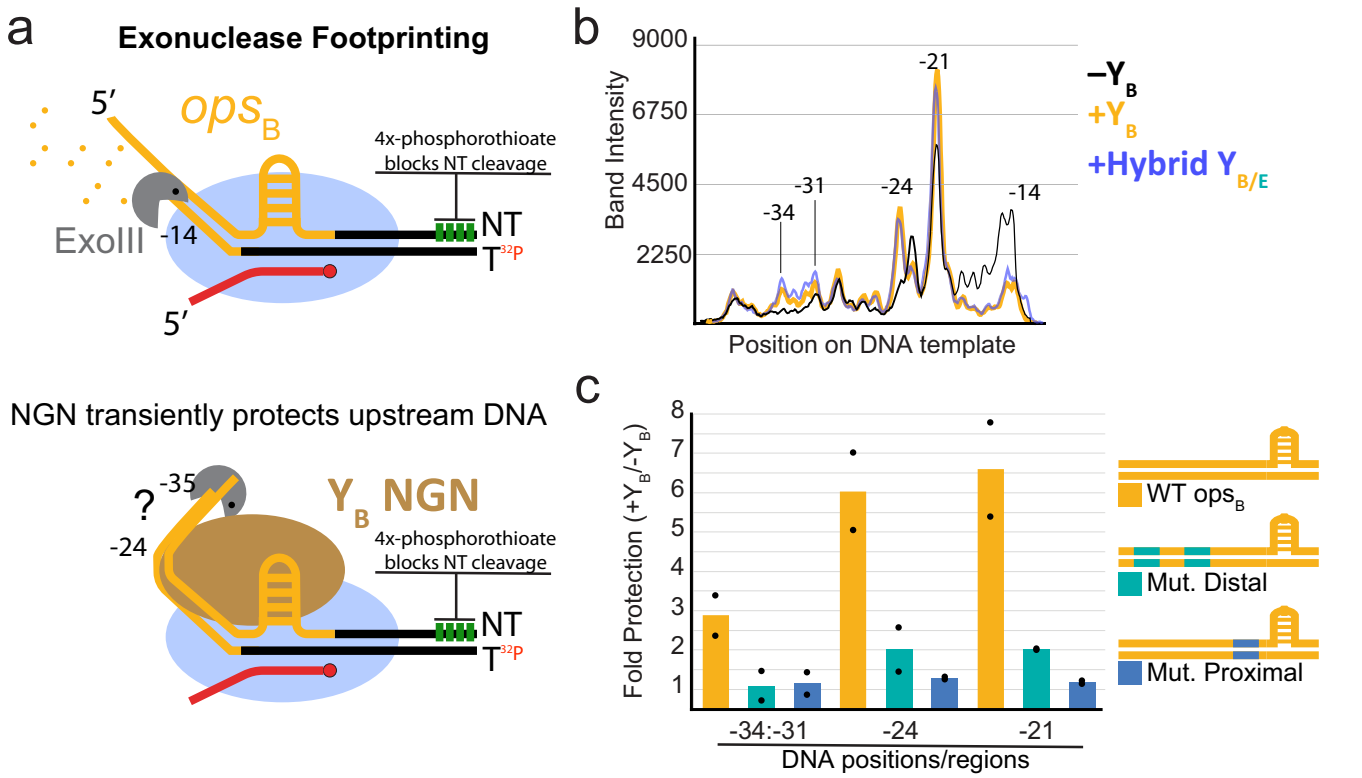

Fig. 6 | Y_X contacts sequences in the upstream duplex DNA and ntDNAhp to recognize cognate ops_X via a capture-then-escape mechanism. a Y_X protects the distal upstream DNA from exonucleolytic cleavage. Exonuclease III (ExoIII) cleaves in the 3′-to-5′ direction but temporarily halts when encountering obstacles such as DNA-bound proteins. Protection was assayed in the absence or presence of Y_X or hybrid NGN–KOW at various timepoints. **b** Pseudodensitometry traces of template DNA cleavage products separated by 8% Urea-PAGE. Band intensities reflect relative levels of cleavage products after 5 s of exonucleolytic cleavage. Traces are representative of experiments performed in at least duplicate (Supplementary Fig. 12). **c** Quantification of upstream DNA protection from exonucleolytic cleavage at various regions on WT or mutant ops_B scaffolds. Data are presented as mean values with data points from $n = 2$ independent experiments shown.

Bacteroides. Our findings provide a foundation for understanding them.

The closer Class-2 start-codon proximity to ops_X (9 bp) suggests that Class-2 Y_X may play a stronger role in ribosome association for coupled transcription–translation of the Y_X gene. Translation is not well studied in *Bacteroides*[68–71], but both the similarity of anti-pausing by *Bfr*NusG to *Eco*NusG (Supplementary Fig. 5) and the location of stop codons relative to intrinsic terminators[72] suggests transcription and translation may be coupled in *Bacteroides* – like *E. coli* but unlike *B. subtilis*[72–82]. RfaH is thought to recruit ribosomes for coupled translation in *E. coli*[83,84]. Start codon GUG is thought to initiate ribosomes 5–10 times more weakly than AUG in *E. coli*[85]. Taken together, these differences are consistent with evolution of Class-2 Y_X–ops_X pairs for tight linkage of Y_X and ribosome recruitment at ops_X sites immediately adjacent to the translation start site. Both these potential distinctions (relative to Class 1) in Class-2 Y_X–ops_X function and interesting differences evident for Y_C–ops_C and Y_G–ops_G require future experimental investigation.

We also discovered a regulatory RNA element—the ops_X PH ED—involved in the regulation of PSA and PSB. The conserved role of PHs at ops_X is to enhance pausing with NusA. The $ops_{A,B}$ ED provides a driving force to propel RNAP out of pause-cycling traps created by extensive interactions that occur at these sites. Possibly, ops_E does not encode an ED because Y_E interacts with less sequence (Supplementary Fig. 12) and Gre factor may be sufficient for its escape as it is for RfaH[33]. Alternatively, the strong kinetic difference in escape mechanisms could be exploited by *Bacteroides* in CPS expression control. We propose that the ED evolved in response to evolutionary pressure to expand Y_X specificity.

Our results provide new mechanistic insights into transcriptional regulation by a large class of $NusG_{SP}$, Y_X (UpxY). We find that determinants of transcriptional pausing in the phylum Bacteroidota resemble those found for other bacteria, but that recruitment sites for these $NusG_{SP}$s differ notably both in being multipartite and much more extensive (~40 bp) than found for *E. coli* RfaH (~12 bp). Two aspects of the Y_X recruitment mechanisms provide precedent for new types of transcriptional regulation: (1) the upstream DNA is a sequence-specific platform for PEC regulation, and (2) pause hairpins can include escape duplexes that can drive escape from regulator-stabilized pauses. These discoveries highlight the importance of studying transcriptional mechanisms in diverse bacteria.

Methods

Plasmids, oligonucleotides, and strains used in this study are listed in Supplementary Tables S1–4. Nucleic acid scaffolds used in PIVoT assays are organized by figure in Supplementary Fig. 16. All reported measurements were taken from distinct samples.

E. coli strain construction

E. coli strain RL3569 was created by P1 transduction of RL1674 with donor strain RL3570[86] harboring the rifampicin-resistance mutation S522F in *rpoB*. Briefly, 5 mL of donor strain RL3570 was grown to saturation (overnight) in LB + 5 mM $CaCl_2$. The next day, 50 µl of the donor strain was mixed with 100 µl of a 10^{-5} dilution (in LB + 5 mM $CaCl_2$) of a freshly made P1 stock, then incubated at 37 °C for 20 min without shaking. 2.5 mL of 45c-equilibrated R top agar (0.8 % agar, 1% tryptone, 0.8% NaCl, 0.1% yeast extract, supplementing to a final concentration of 2 mM $CaCl_2$ and 0.1% glucose after autoclaving) was

https://doi.org/10.1038/s41467-024-55215-9

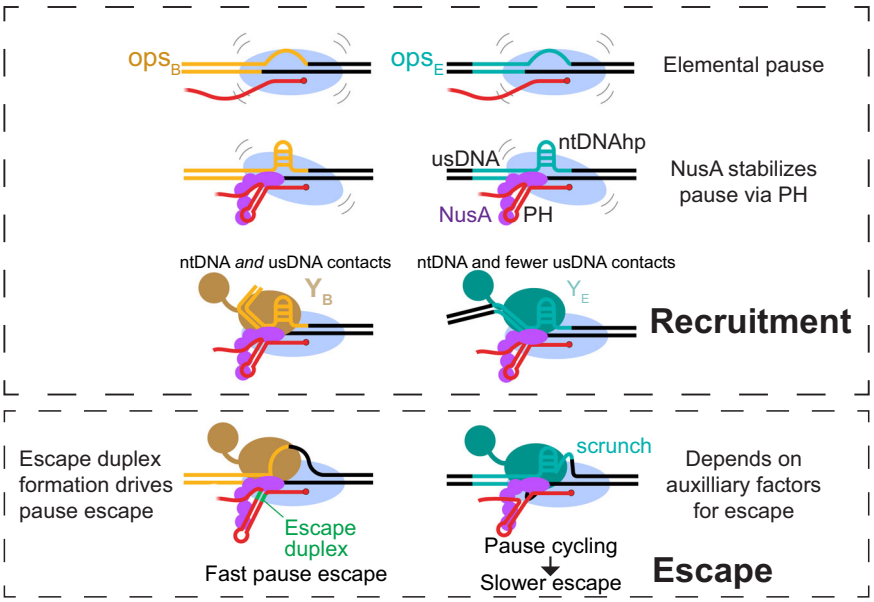

Fig. 7 | Model for Y_X-specific recruitment. *Bfr*RNAP pauses in the 5′ leader of CPS operons to provide time for Y_X recruitment. These pauses arise initially through RNA–DNA contacts to RNAP (elemental pause), then are stabilized synergistically by a PH and NusA. Y_X is recruited with high fidelity to cognate operons by multipartite ~40-bp ops_X elements, with variable influence of constituent elements depending on the CPS operon. Y_Xs that interact extensively with cognate ops_X (Y_A, Y_B) are associated with escape duplex (ED)–encoding PHs, which provides force in the form of base-pairing to drive forward translocation and inhibit backtracking. Differences among escape mechanisms (e.g., those with or without EDs) may aid differential regulation.

added to the bacteria-phage mixture, flicked to mix, then poured evenly onto a thick, moist, freshly-made R plate (1.2% agar, 1% tryptone, 0.8% NaCl, and 0.1% yeast extract, supplementing to a final concentration 2 mM $CaCl_2$ and 0.2% glucose after autoclaving). The plates were incubated at 37 °C overnight in a plastic bag with wet paper towels. The next day, the plate was transferred to a 4c room and overlayed with 5 mL of MC solution (10 mM $MgSO_4$ + 5 mM $CaCl_2$). After a 5 h incubation at 4 °C, the overlayed solution containing fresh P1 lysate was collected, 0.2 μm filter-sterilized, then stored in the dark at 4c until use. The recipient strain (RL1674) was grown to saturation (overnight) in LB + 5 mM $CaCl_2$ + 20 μg chloramphenicol/mL. The next day, 100 μl of donor P1 phage serial dilutions were separately mixed with 100 μl of recipient strain overnight culture, then incubated at 37 °C for twenty minutes with no shaking. The mixture was plated on LB agar + 20 μg chloramphenicol/mL + 100 μg rifampicin/mL. Candidates were sequence-verified.

B. fragilis strain construction

Bacterial growth. *B. fragilis* NCTC 9343 (ATCC25285; Genbank assembly ASM2598v1) strains were grown in basal medium[87] or on BHI plates supplemented with 5 mg hemin/liter and 2.5 μg vitamin K_1/L. Mutants Δ*mpi*M44[17], Δ*mpi*M44Δ*upaZ*[13] and ΩPSE[41] were previously constructed. For selection of cointegrants, gentamycin (200 μg/ml) and erythromycin (5 μg/ml) were added to the plates when indicated.

Construction of mutant PSE *ops* and HP-*ops* regions in 9343Δ*mpi*M44. Two different alterations to the PSE 5′ UTR were made in the Δ*mpi*M44 strain. In the first mutant, the *ops* sequence of the PSE locus (CTGCGAAGCATA) was replaced with the *ops* sequence of the PSA locus (ccgcgtagcgca). In the second mutant, a larger replacement was made and included the hairpin region adjacent to the *ops* sequence. The sequence from the PSE 5′ UTR (ttggctgagaaaaagagtctcacccaaCTGCGAAGCATA) was replaced with the sequence from the PSA 5′UTR (cggtttgaatgggaaaagatgtctcgtccaaaccgcgtagcgca). The recombinant plasmids were created by PCR amplifying two (*ops*) or three (HP-*ops*) DNA segments using Phusion polymerase (NEB) with Δ*mpi*M44 as template with the primers listed in Table S2. These segments were cloned into BamHI-digested pLGB13[88] using NEBuilder (NEB). Plasmids were sequenced to confirm the correct assembly of the segments. Plasmids were conjugally transferred from *E. coli* S17 λpir to Δ*mpi*M44 and after overnight co-incubation, were plated on BHIS with gentamycin and erythromycin. The resulting cointegrants were passaged in basal medium for several hours and plated on BHIS with 50 ng anhydrotetracycline to select for double cross-over recombinants. These strains were tested by PCR for replacement of the PSE sequences with the respective PSA sequences and the genomes of these two strains were sequenced to confirm the correct replacements.

Western immunoblot analysis

Bacterial strains were grown overnight to an apparent OD_{600} of ~1.2. Bacteria were pelleted and resuspended in 1× LDS loading buffer (Invitrogen) and boiled for 5 min. Cell lysates (equivalent to 3.5 μl of the original culture) were loaded onto 4–12% NuPAGE (Invitrogen) and run with MES buffer until the 17 kDa molecular weight standard had run to the bottom of the gel to allow for migration of the high molecular weight PSE further into the gel. The contents of the gel were transferred to PVDF and blocked with 5% skim milk in TBS with 0.5% tween (TBST). The blot was probed with a mouse monoclonal antibody specific to PSE (Supplementary Fig. 15) used at 1:100 dilution, washed with TBST, and probed with a 1:2000 dilution of alkaline phosphatase conjugated goat-anti rabbit IgG (Invitrogen Catalog # 31340 Lot YA366475). After washing with TBST, the blot was developed with BCIP/NBT (KPL).

NET-seq

B. fragilis NCTC 9343 rpoC-3xFLAG was streaked onto BHIS plates and incubated at 37 °C anaerobically for 2 days. A swab from a dense area on the plate was used to inoculate overnight cultures. The next day, 10 mL of the overnight culture was used to inoculate 500 mL SBM (starting apparent OD_{600} 0.04 as measured by a Denville® CO8000 Personal Cell Density Meter). When the apparent OD_{600} measured 0.65, cultures were removed from the anaerobic chamber and 300 mL was used for subsequent steps.

To harvest nascent transcripts for the NET-seq workflow, cultures were filtered between two vacuum filtration systems using a 0.45 μm pore nitrocellulose filter (GVS Micron Sep, 1215305). Cells were scraped off each filter using a spatula and plunged immediately into liquid nitrogen (i.e., cells from the same culture were combined into the same 50 mL conical tube containing ~25 mL liquid nitrogen). Collected cells were cryo-lysed using a RETSCH mixer mill (MM 400) as previously described[37], with the exception that 50 mL stainless steel canisters and a 25 mm stainless steel ball were used to perform the cryomilling.

To isolate nascent transcripts, we performed a modified 3xFLAG-IP protocol with previously described buffers[37]. Specifically, the thawed grindate volume was scaled to 5.5 mL with lysis buffer (1× lysis stock [20 mM Tris, pH 8.0, 0.4% Triton X-100, and 0.1% NP-40 substitute], 100 mM NH4Cl, 1× EDTA-free cOmplete Mini protease inhibitor cocktail [Roche Diagnostics GmbH, 11836170001], 10 mM MnCl2, and 50 μ/mL RNasin [Promega, N211B], and 0.4 mg/mL puromycin), DNA was partially digested for 20 min with RQ1 DNase (0.054 μ/mL [0.02 μ/mL for the *E. coli*-only NET-seq pilot experiment])[Promega, M6101], and digestion reactions were stopped by addition of EDTA to 28 mM (final concentration). RNAP-nascent transcript complexes were directly immunoprecipitated using Anti-FLAG M2 affinity gel (Sigma, A2220) (i.e., without buffer exchange), and the precipitated RNAP-nascent transcript complexes were subsequently washed four times (1× lysis stock, 100 mM NH4Cl, 300 mM KCl, 1 mM EDTA, and 50 μ/mL RNasin)[Promega, N2515]. RNAP-nascent transcript complexes were eluted twice with 3xFLAG peptide (Sigma, F4799) (1× lysis stock, 100 mM NH4Cl, 2 mg/mL 3xFLAG peptide, 1 mM EDTA, and 50 μ/mL RNasin). Nascent transcripts were purified using a miRNeasy kit [Qiagen, 217084] as previously described[37]. However, to reduce phenol and chaotropic salt contamination, nascent transcripts were subjected to an additional overnight isopropanol-GlycoBlue (Invitrogen, AM9516) precipitation at −20 °C.

For nascent transcript library generation, we followed a modification of a previous NET-seq workflow[36,37]. Specifically, our workflow included using custom adaptors compatible with an Illumina NovaSeq X instrument. Likewise, the DNA adapter used for nascent transcript 3′ end ligation was adenylated using components from a NEB 5′ DNA Adenylation kit (E2610; 6 μM DNA linker [RL15032], 80 μM ATP, 6 μM Mth RNA ligase, and 1× Adenylation Reaction Buffer). The adenylation reaction was incubated for 4 h incubation at 65 °C, inactivated at 85 °C for 5 min, and precipitated overnight at −20 °C with isopropanol and GlycoBlue (Invitrogen AM9516). The precipitated, adenylated DNA linker was ligated to 750 ng of precipitated nascent transcripts, in duplicate, using components of a NEB T4 RNA Ligase 2, truncated (T4 Rnl2tr) kit (M0242; 10% DMSO, 22% PEG8000, 3 μM adenylated DNA linker, T4 Rnl2tr [14.7 μ/μL], RNasin [2 μ/μL], and 1× T4 RNA Ligase Reaction Buffer). These ligation reactions were incubated at 37 °C for 4 h. After this incubation, T4 Rnl2tr was inactivated by incubation with Proteinase K (0.04 μ/μL) (NEB, P8107) at 37 °C for 1 h. RNAs were fragmented, resolved, gel extracted, and precipitated as previously described[36,37], with the exception that the gel extraction incubation at 70 °C was increased to 25 min. cDNAs were synthesized using a custom adapter (RL14637) and a previously described protocol[36,37], with the exception that the reaction time was increased to 1 h. Circularization of gel extracted and precipitated cDNAs was performed using a protocol previously described[36,37], with the exception that the circularization reaction incubation period was increased to 3 h and the gel extraction incubation period was increased as above. After circularization, cDNA libraries were PCR amplified using minimal cycles and custom adapters, gel extracted, and precipitated as previously described[36,37]. Library concentration and amplified product size distribution were determined using an Agilent TapeStation 4150. NET-seq libraries were sequenced by the University of Wisconsin-Madison Biotechnology Center on an Illumina NovaSeq X instrument.NET-seq data were processed using a combination of custom scripts and standard tools. Briefly, adapters, linker, and control oligos potentially contaminating each sample were trimmed from raw reads using cutadapt (v3.4). Reads with a minimum length of 14 nts were mapped to the *B. fragilis* genome (NC_003228.3) using Bowtie (v1.3.0) allowing both one mismatch and random assignment of reads mapping to multiple loci based on alignment stratum (Bowtie options --best -a -M 1 -v 1). Alignments were converted to BAM and BED files using samtools (v1.16.1) and bedtools (v2.30.0). The specific 3′ end counts for each genome position were determined using bedtools (options -d -strand -5 [plus strand] or -d -strand + −5 [minus strand]).

r*Bfr*RNAP cloning and purification

B. fragilis RNAP coding regions were codon-optimized using Gene Designer from DNA2.0 (now ATUM) using *E. coli* codon frequencies[89] and amplified from synthetic DNA (IDT) of *B. fragilis* NCTC 9343, then cloned into a pRM756 backbone[90], incorporating a His10-ppx tag at the C-terminus of β′ and a Strep tag at the N-terminus of β. RBS sites were optimized using denovodna.com[91,92]. This plasmid enables T7 overexpression of all subunits under IPTG control.

r*Bfr*RNAP was purified similarly to *E. coli* RNAP[93], with changes described below. Following transformation of RL3569 with pJS015, a colony was picked and inoculated into a 3 mL LB + 25 μg kanamycin/mL + 20 μg chloramphenicol/mL. Two milliliter of overnight culture was used to inoculate 2 L LB + 25 μg kanamycin/ml + 10 drops Sigma Antifoam Y-30 Emulsion in baffled Fernbach flasks and incubated at 37 °C. When the apparent OD_{600} reached 0.4, the temperature was dropped to 16 °C, overexpression was induced by addition of 200 μM IPTG, and incubation was continued with shaking at 200 RPM overnight (~18 h). Cell cultures were placed on ice for 20 min, then pelleted by centrifugation at 3000 × *g* for 15 min at 4 °C.

Moving forward, all steps were performed at 4 °C or on ice, and all buffers were filtered through 0.2 μm filters. Pellets were resuspended in 30 mL lysis buffer (50 mM Tris-HCl pH 8.0, 5% glycerol, 100 mM NaCl, 2 mM EDTA, 10 mM BME, 10 mM DTT, 0.1 mg/mL phenylmethylsulfonyl fluoride, with one dissolved tablet of Roche cOmplete™ ULTRA EDTA-Free Protease Inhibitor Cocktail). The resuspended cell solution was sonicated for 20 min total (alternating sonication on/off times of 5 min) with settings Power 8, Duty Cycle 20%. The lysate was then transferred to round-bottom polycarbonate tubes and spun at 27,000 × *g* for 15 min. The supernatant was transferred to a 100 mL beaker with stir bar, then 6.5% PEI was slowly added to a final concentration of 0.6% while stirring. The solution was stirred for one hour, then transferred to open-top, round-bottom polycarbonate tubes and spun at 11,000 × *g* for 15 min. After decanting supernatant, a tissue homogenizer was used to resuspend the pellet in 25 mL of TGEDZ (10 mM Tris-HCl pH 8.0, 5 % glycerol, 0.1 mM EDTA, 5 μM $ZnCl_2$, 1 mM dithiothreitol) with added 0.3 M NaCl. The solution was spun at 11,000 × *g* for 15 min. After decanting supernatant, a tissue homogenizer was used to resuspend the pellet in 25 mL of TGEDZ with added 1 M NaCl. The solution was spun at 11,000 × *g* for 15 min. The supernatant was transferred into a 100 mL beaker with stir bar, then finely-ground $AmSO_4$ was added to the stirring solution to a final concentration of ~0.37 g/mL and precipitated overnight. The solution was transferred to Oak Ridge round-bottom tubes and spun at 27,000 × *g* for 15 min.

The pellet was dissolved in 35 mL of HisTrap Binding Buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 5 mM beta-mercaptoethanol (BME), then spun at 27,000 × *g* for 15 min in the same Oak Ridge round-bottom tube. The supernatant was filtered through 0.2 μm filters and applied at 1 mL/min to a HisTrap HP 5 mL column, pre-equilibrated with HisTrap Binding Buffer. The column was washed with HisTrap Binding Buffer at 5 mL/min until A280 reached baseline, then washed at 5 mL/min with 2% HisTrap Elution Buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 M imidazole, 5 mM beta-mercaptoethanol

[BME]) until A_{280} reached baseline. rBfRNAP was eluted at 5 mL/min with a 2–50% gradient of HisTrap Elution Buffer (translating to a 20–500 mM imidazole gradient). Three milliliter elution fractions containing rBfRNAP were pooled, filtered through 0.2 μm filters, then the NaCl concentration was reduced to 150 mM for the following purification step by dilution with TGEDZ buffer.

HisTrap elution fractions were pooled then diluted with 100 mM Tris-HCl, pH 8.0, 1 mM EDTA, 10 mM DTT to adjust the salt concentration to 150 mM NaCl. The sample was then applied to a 5 mL Strep-Tactin® XT High Capacity column pre-equilibrated with 2 CV Buffer W (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 10 mM DTT) at 2 mL/min. The flow-through was reapplied to the column at 0.037 mL/min. The column was then washed with 5 CV of Buffer W. rBfRNAP was eluted with Buffer BXT (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 10 mM DTT, 50 mM D+Biotin (Acros Organics)).

Pooled fractions from the previous step were applied at 1.5 mL/min to a HiTrap HP column pre-equilibrated with TGEDZ + 200 mM NaCl. The column was then washed with TGEDZ + 200 mM NaCl until A280 reached baseline, then r*Bfr*RNAP was eluted with TGEDZ + 500 mM NaCl at 2.5 mL/min.

Pooled fractions from the previous step were dialyzed overnight in RNAP storage buffer (10 mM Tris-HCl, pH 8.0, 25% glycerol, 100 mM NaCl, 100 μM EDTA, 1 mM $MgCl_2$, 20 μM $ZnCl_2$, 10 mM DTT) using a 10 kDa MWCO cassette, then concentrated using Ultra-4, MWCO 100 kDa (Sigma-Aldrich Z648043-24EA) to a final concentration of 8 μM. The solution was then aliquoted, flash-frozen, and stored at −80 °C.

Cloning and purification of transcription factors

All TFs (NusG, NusA, Y_A, Y_B, Y_C, Y_E, Y_F, Y_H, Y_B(NGN)–Y_B(KOW), Y_E(NGN)– Y_B(KOW)) were cloned into a pTYB2 backbone (Addgene catalog N6702S) after PCR amplification from *Bacteroides fragilis* ATCC 25285 (NCTC 9343) genomic DNA by NEB HiFi DNA assembly (Gibson Assembly). This vector enables IPTG-inducible over-expression of proteins fused at the C-terminus to the *Saccharomyces cerevisiae* VMA intein and chitin-binding domain. Importantly, to ensure efficient self-cleavage via the intein, an Ala residue was incorporated at the C-terminus of all transcription factor coding sequences.

After plasmid sequence verification, RL1674 (*E. coli* BL21 Rosetta™ (DE3)) was transformed by electroporation with pTYB2-derived constructs, then plated on LB agar with 100 μg ampicillin/mL and 20 μg chloramphenicol/mL (for retention of pRARE2 plasmid). For each expression construct, a single colony was picked and used to inoculate a 3 mL overnight LB culture grown at 37 °C containing the same concentration of antibiotics. The next day, 1 mL of overnight culture was used to inoculate a 200 mL LB culture containing antibiotics (3% ethanol was added for all Y_X constructs) and grown at 37 °C. When the OD reached 0.2–0.3, the incubation temperature was dropped to 16 °C and shaking continued for 30 min. Subsequently, a final concentration of 200 μM IPTG was added and incubation continued overnight (16–18 h). The next day, cultures were placed on ice for 20 min, then pelleted at 3000 × *g* for 15 min at 4 °C.

Pellets were resuspended in 40 mL of Chitin Wash Buffer (CWB; 30 mM Tris-HCl, pH 7.5–8.0 depending on protein pI, 0.5 M NaCl, 1 mM EDTA, 0.05% Tween® 20) plus one dissolved tablet of Roche cOmplete™ ULTRA EDTA-Free Protease Inhibitor Cocktail. The cell suspension was sonicated 10 min at 20% duty cycle, Power 8. The lysate was pelleted at 30,000 × *g* for 30 min at 4 °C, then the supernatant was passed through 0.2 μm filters.

The subsequent steps were performed at room temperature closely following manufacturer's instructions. Briefly, 3 mL of a homogenous suspension of NEB Chitin Resin (Catalog S6651L) were loaded into a 25 mL Poly-Prep Gravity Chromatography Column (Biorad), washed with 5 mL of mQH_2O, then equilibrated by washing 3 times each with 10 mL of CWB. The lysate was subsequently loaded onto the column, then washed three times each with 10 mL of CWB. Cleavage Buffer (CB) was made by adding 500 μl of 1 M DTT (prepared fresh from solid reagent) to 10 mL of CWB, then a quick flush was performed by adding 3 mL of CB. SDS-PAGE revealed no premature elution in the quick flush fraction. Immediately after dripping stopped, the bottom and top of the column were capped, parafilmed, and the column was incubated at room temperature overnight (16–18 h) to allow sufficient time for cleavage. The next day, cleaved protein was eluted by addition of 1.5 mL CWB + 10 mM DTT, then dialyzed overnight in 10 mM Tris-HCl, pH 7.5–8.0 depending on pI, 2% glycerol, 100 mM NaCl, 100 μM EDTA, 10 mM DTT using a 10 K MWCO cassette. After removal from the dialysis cassette, additional glycerol was added to a final concentration of 25%. The solution was aliquoted, flash-frozen, then stored at −80 °C until use.

PIVoT assays

A direct reconstitution approach was used to assemble elongation complexes (ECs). Briefly, RNA and template DNA oligonucleotides were mixed at a ratio of 1:1.2 (5 μM: 6 μM) in transcription buffer (TB; 20 mM Tris-OAc, pH 7.7, 40 mM KOAc, 5 mM $Mg(OAc)_2$, 1 mM DTT), then annealed by slow cooling in a thermocycler. To assemble 10× ECs, first the annealed RNA:tDNA scaffold and RNAP were mixed in TB and incubated for 15 min at 37 °C. Then, non-template DNA oligonucleotide was added and incubation continued for an additional 15 min at 37 °C. The solution was diluted with TB to prepare 2× EC (subtracting volume of further additions) and incubated for 1 min at 37 °C. Then, 5 μCi of [α-^{32}P]NTP (depending on the scaffold) was added and incubated for 3 min at 37 °C. Additional GTP was added such that the final concentration of GTP in the solution was 10 μM, and incubation continued for 3 min at 37 °C.

2× ECs were aliquoted and all comparisons made were therefore performed with identically formed ECs. The assay was performed at 37 °C: transcription was restarted by addition of 2× NTPs minus/plus TFs or storage buffer. For Fig. 5c, Y_B was pre-incubated with halted ECs following reconstitution at −3 and incorporation labeling to −2 prior to restarting transcription. Timepoints were taken by mixing 5 μl reaction aliquots with 5 μl of 2× Stop Buffer (25 mM EDTA, 8 M Urea, 1× TBE, 0.1% bromophenol blue, 0.1% xylene cyanol). The ratio and concentrations of EC components in the 1× EC solution was 1:1.2:1.4:1.6 (R:T:RNAP:NT; 50 nM, 60 nM, 70 nM, 80 nM). The final reaction concentrations of TFs are indicated in each figure legend. Unless otherwise indicated, NTPs are added to a final reaction concentration of 100 μM. RNAs were resolved by 8% or 15% Urea-PAGE with 0.5× TBE running buffer until the leading dye ran off the gel. Gels were exposed to PhosphorImager screens and scanned using a Typhoon Phosphorimager. To quantify effects in ImageQuant, boxes were drawn around the pause band ops_X, the capture band(s) (if applicable), and beyond. After subtracting background, the fractions of RNA at ops_X or at capture positions were averaged and errors reflect standard deviation from at least three replicates (unless indicated otherwise).

For the Z-titration assay in Fig. 2d, data were fit in Kaleidagraph to a sigmoidal function of the form $y = a + (b\text{-}a) / (1 + (x/c)^d)$ where a= ymin, b is ymax, c is the Z_X concentration at mid-point, and d is slope at mid-point; and weighted by standard deviation (error bars) from three assays.

Biolayer interferometry

Preparation of biotinylated-Y_E: pJS060 was cloned similarly to other pTYB2-derived constructs (see above), with the exception that two oligos were included in the Gibson assembly to introduce the 16 codon Avi-tag™ onto the N-terminus of upeY. Expression, cell harvesting, and lysis conditions are as described above. Avi-Y_E was biotinylated on a gravity column as described below:

The subsequent steps were performed at room temperature closely following NEB instructions. Briefly, 3 mL of a homogenous suspension of NEB Chitin Resin (Catalog S6651L) were loaded into a 25 mL Poly-Prep Gravity Chromatography Column (Biorad), washed with 5 mL of mQH_2O, then equilibrated by washing 3 times each with 10 mL of CWB. The lysate was subsequently loaded onto the column, then washed three times each with 10 mL of CWB. The column was then washed with three times each with Avi Chitin Wash Buffer (AviCWB = 10 mM Tris 8.0, 0.5 M KGlu, 0.1% Tween20). Components from Avidity BirA500 Kit were used in the subsequent biotinylation reaction: a biotinylating solution (500 μL AviCWB, 70 μL of BiomixA, 70 μL Biomix B, 10 μL of 1 mg/mL BirA) was added to the column and the reaction was allowed to continue for 2.5 h. The column was subsequently washed three times each with 10 mL of CWB. Cleavage Buffer (CB) was made by adding 500 μl of 1 M DTT (prepared fresh from solid reagent) to 10 mL of CWB, then a quick flush was performed by adding 3 mL of CB. Immediately after dripping stopped, the bottom then the top of the column were capped, parafilmed, and the column was incubated at room temperature overnight (16–18 h) to allow sufficient time for cleavage. The next day, cleaved protein was eluted by addition of 1.5 mL CWB + 10 mM DTT, then dialyzed overnight in 10 mM Tris-HCl pH 7.5, 2% glycerol, 100 mM NaCl, 100 μM EDTA, 1 mM DTT using a 10 K MWCO cassette. After removal from the dialysis cassette, additional glycerol was added to a final concentration of 20%. The solution was aliquoted, flash-frozen, then stored at −80 °C until use. Importantly, Biotin-Y_E retained activity in vitro.

For each titration, 1 mL of 0.3 μM biotinylated-Y_E was prepared in Octet Binding Buffer 4.1 (OBB4.1 = PBS + 400 mM NaCl + 0.01% Triton X-100 + 0.25% BSA). Z_A solution was prepared at 100 nM in OBB4.1 with twofold serial dilutions down to 1.56 nM. Z_E solution was prepared at 500 nM in OBB4.1 with serial dilutions down to 31.3 nM. Plates were prepared for binding assays: in plate 1200 μL of OBB4.1 was placed in each well of column 1 containing a biosensor (up to 8 biosensors per experiment); plate 2 (containing 'half-area' wells permitting 100 μL volumes) column 1 contained 100 μL/well of OBB4.1, column 2 contained 100 μL/well of 0.3 μM biotinylated-Y_E, and column 3 contained 100 μl/well of Z_X serial dilutions or buffer (as a blank/reference) prepared above.

A basic kinetics assay was performed using standard acquisition rates at 30 °C on a ForteBio Octet RED96 system. Octet® Streptavidin (SA) Biosensors were pre-equilibrated for 10 min at 30c. Step times: Baseline (Plate 2 Column 1 (P2C1)) = 60 sec; Loading (P2C2 = 320 sec (or until 2 nm loading density reached); Baseline (P2C1) = 60 sec; Association (P2C3) = ≥ 300 sec; Dissociation (P2C1) = ≥ 300 sec.

Data were processed using Octet Data Analysis Software. The reference biosensor curve (bio-Y_E + buffer in place of Z_X) was subtracted from all binding curves. Traces were subsequently aligned along the *Y* axis at pre-association baseline with interstep correction performed at the dissociation step. Noise Filtering (Savitsky-Golay-Filtering, smoothingfunction) was performed. Data from each experiment were independently globally fit. For each binding pair tested, two out of three global fits have R^2 values around 0.95 or greater and chi-squared values less than 3 as recommended by ForteBio. Given the two orders of magnitude difference in binding constants, limited conclusions we are making, and parsimonious agreement of these constants among replicates and with our PIVoT assays, we deemed the fits overall acceptable. The average and standard deviation of the kinetic parameters from the global fits are reported. Equilibrium constants are calculated from models. The value 'Req/Rmax' is reported as fraction Y_E bound.

Exonuclease footprinting

Nucleic acid scaffolds used in exonuclease footprinting assays were each comprised of: (i) a ^{32}P-labeled template DNA oligo, (ii) a non-template DNA oligo with four consecutive phosphorothioate bonds at the 3′ end, and (iii) an RNA oligo with 3′ end at the position of pausing in ops_X and having noncomplementary bases upstream of the RNA-DNA hybrid to prohibit backtracking.

Template DNA oligo (20 μM) was labeled in a T4 PNK reaction with 1 μCi of [γ-^{32}P]ATP and allowed to proceed for 15 mins at 37 °C. ATP (1 μL of 1 mM) was subsequently added to the reaction and allowed to proceed for 30 min at 37 °C. Reactions were stopped by heating at 65 °C for 20 min and oligos were subsequently purified using G-50 columns pre-equilibrated with TE and following the manufacturer's instructions.

TECs were reconstituted essentially as described in in vitro transcription assays, except that the molar ratio of T:R:Pol:NT was 1:2:3:5 (50 nM ^{32}P-T: 100 nM R: 150 nM RNAP: 250 nM NT). TECs were subsequently split into 35 μL aliquots and incubated with either storage buffer or Y_X variants for 3 min at 37 °C. Tubes were shifted to 30 °C and allowed to incubate for 3 min before removing a 5 μl aliquot (time 0) and mixing with equal volume 2× Stop Buffer. Exonuclease reactions were initiated by adding 100 μ of ExoIII, and aliquots were removed from reactions and mixed with stop buffer at times indicated in figures.

To quantify both transient and stable protection from exonucleolytic cleavage, pseudodensitometry traces were generated for the first timepoint lane. Regions of interest were identified by comparison to a sequencing ladder. Areas under the peaks of these regions were determined by manual integration in Microsoft Excel, then divided by the sum of the areas under all peaks to the right of it. These values were determined in the absence or presence of Y_B, and their ratio is reported as fold change ($+Y_B/-Y_B$) for each sequence variant.

Structural models

A model of Y_B was made using Modeller[94,95] and fitted to 8PHK[33]. Additional upstream and downstream DNA were modeled using Pymol. The Y_E–Z_A complex structure was predicted using AlphaFold 3[54], yielding an interface predicted template modeling (iPTM) score of 0.89 and predicted template modeling (pTM) score of 0.9 (values above 0.8 represent confident high-quality predictions). Additional confidence metrics are illustrated in Supplementary Fig. 6. RNA secondary structures were predicted using RNAFold[95].

The *Bfr*RNA polymerase PEC model was generated using Modeller[96], the *M. tuberculosis* PEC formed on the *B. subtilis trpL* pause sequence (8E74)[27], NusA and NusG NGN models from SWISS-MODEL[97], and *Porphymonas gingevalis* RNAP (8DKC)[98].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The NET-seq data generated in this study have been deposited in the NCBI GEO database under accession code GSE281607. The Y_E–Z_A model generated by AlphaFold3 is available on Zenodo at https://doi.org/10.5281/zenodo.14110860. Source data are provided with this paper for all other experiments. Source data are provided with this paper.

Code availability

Scripts for analyzing NET-seq data are available on Zenodo at https://doi.org/10.5281/zenodo.14110860.

References

1. Deng, H. et al. Bacteroides fragilis prevents clostridium difficile infection in a mouse model by restoring gut barrier and microbiome regulation. *Front. Microbiol.* **9**, 2976 (2018).
2. Li, X. et al. A strain of Bacteroides thetaiotaomicron attenuates colonization of Clostridioides difficile and affects intestinal

microbiota and bile acids profile in a mouse model. *Biomed. Pharmacother.* **137**, 111290 (2021).
3. Carasso, S. et al. Inflammation and bacteriophages affect DNA inversion states and functionality of the gut microbiota. *Cell Host Microbe* **32**, 322–334.e329 (2024).
4. Hryckowian, A. J. et al. Bacteroides thetaiotaomicron-infecting bacteriophage isolates inform sequence-based host range predictions. *Cell Host Microbe* **28**, 371–379.e375 (2020).
5. Porter, N. T. et al. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. *Nat. Microbiol.* **5**, 1170–1181 (2020).
6. Bechon, N. et al. Capsular polysaccharide cross-regulation modulates bacteroides thetaiotaomicron biofilm formation. *mBio* **11** (2020). https://doi.org/10.1128/mBio.00729-20.
7. Jiang, X. et al. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* **363**, 181–187 (2019).
8. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).
9. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).
10. Porter, N. T., Canales, P., Peterson, D. A. & Martens, E. C. A subset of polysaccharide capsules in the human symbiont bacteroides thetaiotaomicron promote increased competitive fitness in the mouse gut. *Cell Host Microbe* **22**, 494–506.e498 (2017).
11. Ramakrishna, C. et al. Bacteroides fragilis polysaccharide A induces IL-10 secreting B and T cells that prevent viral encephalitis. *Nat. Commun.* **10**, 2153 (2019).
12. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of *Bacteroides fragilis*. *J. Bacteriol.* **191**, 7288–7295 (2009).
13. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. *Proc. Natl. Acad. Sci. USA* **107**, 11976–11980 (2010).
14. Troy, E. B., Carey, V. J., Kasper, D. L. & Comstock, L. E. Orientations of the Bacteroides fragilis capsular polysaccharide biosynthesis locus promoters during symbiosis and infection. *J. Bacteriol.* **192**, 5832–5836 (2010).
15. Lan, F. et al. Single-cell analysis of multiple invertible promoters reveals differential inversion rates as a strong determinant of bacterial population heterogeneity. *Sci. Adv.* **9**, eadg5476 (2023).
16. Lan, F. et al. Massively parallel single-cell sequencing of diverse microbial populations. *Nat. Methods* **21**, 228–235 (2024).
17. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. USA* **100**, 10446–10451 (2003).
18. Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27 (2012).
19. Bailey, M. J., Koronakis, V., Schmoll, T. & Hughes, C. Escherichia coli HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the rfaH (sfrB) locus required for expression of sex factor and lipopolysaccharide genes. *Mol. Microbiol.* **6**, 1003–1012 (1992).
20. Bies-Etheve, N. et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**, 649–654 (2009).
21. Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. *Nat. Microbiol.* **2**, 17003 (2017).
22. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).
23. Kang, J. Y. et al. Structural basis for transcript elongation control by NusG/RfaH universal regulators. *Cell* **173**, 1650–1662.e1614 (2018).
24. Leeds, J. A. & Welch, R. A. RfaH enhances elongation of Escherichia coli hlyCABD mRNA. *J. Bacteriol.* **178**, 1850–1857 (1996).
25. Yakhnin, A. V. et al. Robust regulation of transcription pausing in Escherichia coli by the ubiquitous elongation factor NusG. *Proc. Natl. Acad. Sci. USA* **120**, e2221114120 (2023).
26. Czyz, A., Mooney, R. A., Iaconi, A. & Landick, R. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**, e00931 (2014).
27. Delbeau, M. et al. Structural and functional basis of the universal transcription factor NusG pro-pausing activity in Mycobacterium tuberculosis. *Mol. Cell* **83**, 1474–1488.e1478 (2023).
28. Mandell, Z. F. et al. NusG is an intrinsic transcription termination factor that stimulates motility and coordinates gene expression with NusA. *Elife* **10** (2021). https://doi.org/10.7554/eLife.61880.
29. Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I. & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.* **1**, 15007 (2016).
30. Sevostyanova, A. & Artsimovitch, I. Functional analysis of Thermus thermophilus transcription factor NusG. *Nucleic Acids Res.* **38**, 7432–7445 (2010).
31. Landick, R. Transcriptional pausing as a mediator of bacterial gene regulation. *Annu. Rev. Microbiol.* **75**, 291–314 (2021).
32. Zuber, P. K. et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *Elife* **7** (2018). https://doi.org/10.7554/eLife.36349.
33. Zuber, P. K. et al. Concerted transformation of a hyper-paused transcription complex and its reinforcing protein. *Nat. Commun.* **15**, 3040 (2024).
34. Paitan, Y., Orr, E., Ron, E. Z. & Rosenberg, E. A NusG-like transcription anti-terminator is involved in the biosynthesis of the polyketide antibiotic TA of Myxococcus xanthus. *FEMS Microbiol. Lett.* **170**, 221–227 (1999).
35. Nunez, B., Avila, P. & de la Cruz, F. Genes involved in conjugative DNA processing of plasmid R6K. *Mol. Microbiol.* **24**, 1157–1168 (1997).
36. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
37. Larson, M. H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–1047 (2014).
38. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. Usa.* **97**, 7090–7095 (2000).
39. Guo, X. et al. Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* **69**, 816–827.e814 (2018).
40. Kang, J. Y. et al. RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Mol. Cell* **69**, 802–815.e805 (2018).
41. Krinos, C. M. et al. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555–558 (2001).
42. Daube, S. S. & von Hippel, P. H. Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes. *Science* **258**, 1320–1324 (1992).
43. Toulokhonov, I., Artsimovitch, I. & Landick, R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**, 730–733 (2001).

44. Bao, Y., Cao, X. & Landick, R. RNA polymerase SI3 domain modulates global transcriptional pausing and pause-site fluctuations. *Nucleic Acids Res.* (2024). https://doi.org/10.1093/nar/gkae209.
45. Gajos, M. et al. Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res.* **49**, 4402–4420 (2021).
46. Kireeva, M. L. & Kashlev, M. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. USA* **106**, 8900–8905 (2009).
47. Yakhnin, A. V. et al. NusG controls transcription pausing and RNA polymerase translocation throughout the *Bacillus subtilis* genome. *Proc. Natl. Acad. Sci. USA* **117**, 21628–21636 (2020).
48. Ha, K. S., Toulokhonov, I., Vassylyev, D. G. & Landick, R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J. Mol. Biol.* **401**, 708–725 (2010).
49. Jayasinghe, O. T., Mandell, Z. F., Yakhnin, A. V., Kashlev, M. & Babitzke, P. Transcriptome-wide effects of NusA on RNA polymerase pausing in *Bacillus subtilis*. *J. Bacteriol.* **204**, e0053421 (2022).
50. Kolb, K. E., Hein, P. P. & Landick, R. Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J. Biol. Chem.* **289**, 1151–1163 (2014).
51. Strobel, E. J. & Roberts, J. W. Two transcription pause elements underlie a sigma70-dependent pause cycle. *Proc. Natl. Acad. Sci. USA* **112**, e4374–e4380 (2015).
52. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
53. Roberts, J. W. Biochemistry. RNA polymerase, a scrunching machine. *Science* **314**, 1097–1098 (2006).
54. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* (2024). https://doi.org/10.1038/s41586-024-07487-w.
55. Elghondakly, A., Wu, C. H., Klupt, S., Goodson, J. & Winkler, W. C. A NusG Specialized Paralog That Exhibits Specific, High-Affinity RNA-Binding Activity. *J. Mol. Biol.* **433**, 167100 (2021).
56. Eckartt, K. A. et al. Compensatory evolution in NusG improves fitness of drug-resistant M. tuberculosis. *Nature* (2024). https://doi.org/10.1038/s41586-024-07206-5.
57. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. The beta subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 (2011).
58. You, L. et al. Structural basis for intrinsic transcription termination. *Nature* **613**, 783–789 (2023).
59. Zhu, C. et al. Transcription factors modulate RNA polymerase conformational equilibrium. *Nat. Commun.* **13**, 1546 (2022).
60. Harteis, S. & Schneider, S. Making the bend: DNA tertiary structure and protein-DNA interactions. *Int. J. Mol. Sci.* **15**, 12335–12363 (2014).
61. Landick, R. & Yanofsky, C. Isolation and structural analysis of the Escherichia coli trp leader paused transcription complex. *J. Mol. Biol.* **196**, 363–377 (1987).
62. Nedialkov, Y., Svetlov, D., Belogurov, G. A. & Artsimovitch, I. Locking the non-template DNA to control transcription. *Mol. Microbiol.* **109**, 445–457 (2018).
63. Samkurashvili, I. & Luse, D. S. Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. *J. Biol. Chem.* **271**, 23495–23505 (1996).
64. Lane, W. J. & Darst, S. A. Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* **395**, 671–685 (2010).
65. Cao, X. et al. Basis of narrow-spectrum activity of fidaxomicin on Clostridioides difficile. *Nature* **604**, 541–545 (2022).
66. Vishwakarma, R. K., Qayyum, M. Z., Babitzke, P. & Murakami, K. S. Allosteric mechanism of transcription inhibition by NusG-dependent pausing of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **120**, e2218516120 (2023).
67. Hustmyer, C. M., Wolfe, M. B., Welch, R. A. & Landick, R. RfaH counter-silences inhibition of transcript elongation by H-NS-StpA nucleoprotein filaments in pathogenic Escherichia coli. *mBio* **13**, e0266222 (2022).
68. Accetto, T. & Avgustin, G. Inability of Prevotella bryantii to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. *PLoS ONE* **6**, e22914 (2011).
69. Mastropaolo, M. D., Thorson, M. L. & Stevens, A. M. Comparison of Bacteroides thetaiotaomicron and Escherichia coli 16S rRNA gene expression signals. *Microbiology* **155**, 2683–2693 (2009).
70. Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst.* **1**, 62–71 (2015).
71. Wegmann, U., Horn, N. & Carding, S. R. Defining the bacteroides ribosomal binding site. *Appl. Environ. Microbiol.* **79**, 1980–1989 (2013).
72. Johnson, G. E., Lalanne, J. B., Peters, M. L. & Li, G. W. Functionally uncoupled transcription-translation in Bacillus subtilis. *Nature* **585**, 124–128 (2020).
73. Adhya, S. & Gottesman, M. Control of transcription termination. *Annu. Rev. Biochem.* **47**, 967–996 (1978).
74. Burmann, B. M. et al. A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010).
75. Byrne, R., Levin, J. G., Bladen, H. A. & Nirenberg, M. W. The in vitro formation of a DNA-Ribosome complex. *Proc. Natl. Acad. Sci. USA* **52**, 140–148 (1964).
76. Castro-Roa, D. & Zenkin, N. In vitro experimental system for analysis of transcription-translation coupling. *Nucleic acids Res.* **40**, e45 (2012).
77. Landick, R., Carey, J. & Yanofsky, C. Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc. Natl. Acad. Sci. USA* **82**, 4663–4667 (1985).
78. McGary, K. & Nudler, E. RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* **16**, 112–117 (2013).
79. Miller, O. L. Jr, Hamkalo, B. A. & Thomas, C. A. Jr Visualization of bacterial genes in action. *Science* **169**, 392–395 (1970).
80. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
81. Saxena, S. et al. Escherichia coli transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018).
82. Stevenson-Jones, F., Woodgate, J., Castro-Roa, D. & Zenkin, N. Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc. Natl. Acad. Sci. USA* **117**, 8462–8467 (2020).
83. Burmann, B. M. et al. An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291–303 (2012).
84. Lorber, C. G. [Differential diagnosis of maxillofacial neuralgia]. *ZWR* **85**, 514–518 (1976).
85. O'Donnell, S. M. & Janssen, G. R. The initiation codon affects ribosome binding and translational efficiency in Escherichia coli of cI mRNA with or without the 5′ untranslated leader. *J. Bacteriol.* **183**, 1277–1283 (2001).
86. Jin, D. J. & Gross, C. A. Mapping and sequencing of mutations in the Escherichia coli rpoB gene that lead to rifampicin resistance. *J. Mol. Biol.* **202**, 45–58 (1988).
87. Pantosti, A., Tzianabos, A. O., Onderdonk, A. B. & Kasper, D. L. Immunochemical characterization of two surface polysaccharides of Bacteroides fragilis. *Infect. Immun.* **59**, 2075–2082 (1991).
88. Garcia-Bayona, L. & Comstock, L. E. Streamlined genetic manipulation of diverse bacteroides and parabacteroides isolates from

the human gut microbiota. *mBio* **10** (2019). https://doi.org/10.1128/mBio.01762-19.
89. Welch, M. et al. Design parameters to control synthetic gene expression in Escherichia coli. *PLoS ONE* **4**, e7002 (2009).
90. Windgassen, T. A. et al. Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res.* **42**, 12707–12721 (2014).
91. Reis, A. C. & Salis, H. M. An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.* **9**, 3145–3156 (2020).
92. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
93. Saba, J. et al. The elemental mechanism of transcriptional pausing. *Elife* **8** (2019). https://doi.org/10.7554/eLife.40981.
94. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
95. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
96. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5 6 1–5 6 37 (2016).
97. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
98. Bu, F. et al. Cryo-EM structure of porphyromonas gingivalis RNA polymerase. *J. Mol. Biol.* **436**, 168568 (2024).
99. Coyne, M. J. et al. Polysaccharide biosynthesis locus required for virulence of Bacteroides fragilis. *Infect. Immun.* **69**, 4342–4350 (2001).
100. Sultana, A. & Lee, J. E. Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr. Protoc. Protein Sci.* **79**, 19 25 11–19 25 26 (2015).
101. Madeira, F. et al. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* (2024). https://doi.org/10.1093/nar/gkae241.

Acknowledgements

We thank members of the Landick and Comstock labs for helpful discussions and comments on the manuscript. This work was supported by NIH R01 GM038660 and USDA Hatch WIS05004 to R.L., NIH R01 AI093771 to L.C., the Duchossois Family Institute, and the DOE Office of Science, Biological and Environmental Research Program Great Lakes Bioenergy Research Center (DE-SC0018409). A.G. was supported by the NIH Predoctoral Training Program in Genetics (T32 GM007133). J.S. was supported by the NIH Biotechnology Training Grant (T32 GM135066 and T32 GM008349), an NIH F31 Graduate Fellowship (F31 GM142153), and a SciMed Graduate Research Scholars Fellowship from the UW–Madison Graduate School and Wisconsin Alumni Research Foundation.

Author contributions

R.L. and J.S. conceived of the study. J.S. conceived and developed assays, cloned most plasmids, purified all proteins, performed most experiments, and analyzed data. K.F. constructed plasmids for *B. fragilis* genetic manipulation, created *Bacteroides* strains and performed Western blots. M.E., B.M., and J.S. wrote custom scripts. J.S. and R.L. interpreted data. M.E., Y.P., and A.G. performed experiments. R.L. and J.S. constructed structural models. J.S. and R.L. wrote the original manuscript and designed figures. J.S., R.L., and L.C. revised the manuscript. R.L., L.C., and J.S., secured funding. R.L. and L.C. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55215-9.

Correspondence and requests for materials should be addressed to Robert Landick.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.