A scalable framework for the discovery of functional helicase substrates and helicase-driven regulatory switches

Mildred Delaleau, Eric Eveno, Isabelle Simon, Annie Schwartz, Marc Boudvillain

A scalable framework for the discovery of functional helicase substrates and helicase-driven regulatory switches

Mildred Delaleau,[1,4] Eric Eveno,[1,4] Isabelle Simon,[1,2] Annie Schwartz,[1] and Marc Boudvillain[1,2,3]

[1]: Centre de Biophysique Moléculaire, CNRS UPR4301, rue Charles Sadron, 45071 Orléans cedex 2, France ; affiliated with Université d'Orléans.

[2]: ED 549, Sciences Biologiques & Chimie du Vivant, Université d'Orléans, France.

[3]: Corresponding author (marc.boudvillain@cnrs.fr; +33 238 25 55 85)

[4]: These authors contributed equally to the work.

Competing interest statement: M.D., A.S., and M.B. are co-inventors in a patent application for H-SELEX filled by CNRS innovation, a subsidiary of CNRS

Running title: Discovery of helicase substrates and synthetic riboswitches by Helicase-SELEX

ABSTRACT

Helicases are ubiquitous motor enzymes that remodel nucleic acids (NA) and NA-protein complexes in key cellular processes. To explore the functional repertoire and specificity landscape of helicases, we devised a screening scheme –Helicase-SELEX- that enzymatically probes substrate and cofactor requirements at global scale. Using the transcription termination Rho helicase of *Escherichia coli* as a prototype for Helicase-SELEX, we generated the first, genome-wide map of Rho utilization (*Rut*) sites. The map reveals many new features, including promoter- and intrinsic terminator-associated *Rut* sites, bidirectional *Rut* tandems, and cofactor-dependent *Rut* sites with inverted G>C skewed compositions. We also implemented an H-SELEX variant where we used a model ligand, serotonin, to evolve synthetic *Rut* sites operating *in vitro* and *in vivo* in a ligand-dependent manner. Altogether, our data illustrate the power and flexibility of Helicase-SELEX to seek constitutive or conditional helicase substrates in natural or synthetic NA libraries for fundamental or synthetic biology discovery.

SIGNIFICANCE STATEMENT

Helicases are ubiquitous NTP-dependent enzymes that disrupt nucleic acid (NA) helices and NA-protein interactions. Despite the implication of helicases in many cellular processes and diseases, their target repertoires and determinants of functional specialization often remain uncertain. We developed Helicase-SELEX to combinatorically probe helicase substrate requirements and find natural or synthetic substrates in large NA sequence libraries. Using the transcription termination Rho helicase as prototype, we discovered ~3300 functional substrate sequences in *Escherichia coli*, thereby providing the first detailed map of *R*ho *ut*ilization (*Rut*) sites at genome scale. We also evolved synthetic *Rut* switches eliciting Rho activity only in presence of a selected inducer. Thus, Helicase-SELEX is a unique new approach to characterize or exploit helicases for fundamental or biotechnology purposes.

INTRODUCTION

Helicases are molecular motors that convert the energy of NTP hydrolysis into mechanical work to unwind NAs, translocate along NAs, or remodel NA-protein interactions (1-3). These ubiquitous enzymes are involved in every cellular event dealing with NAs, including the major steps of gene expression (1-3). For instance, the bacterial transcription termination Rho helicase dissociates transcription elongation complexes (TECs) at specific genomic loci (4-6) and protects the genome by clearing deleterious R-loops (7). The importance of helicases in key NA transactions makes them attractive pharmacological targets to treat various diseases, from pathogenic infections to cancers (8). Their unique NA remodeling capacities and regulatory functions could also prove useful in synthetic biology, a prospect that, to our knowledge, has not been explored yet.

Although evolved from the same ATPase core, helicases can display very distinct organizations (e.g. monomer vs. ring-shaped oligomer), interactions with NAs, or mechanisms of action (2, 3). For instance, DEAD-box RNA helicases only unwind short RNA helices *in vitro* using a non-directional local strand separation mechanism (1, 3) whereas specimens such as Rho, or eukaryotic Upf1-like helicases, are processive translocases that can unwind long duplexes in a 5'→3' directional manner (9, 10). This diversity of structures, functions, and mechanisms stems from specific amino acid motifs/domains decorating the helicase core, which alone or through the recruitment of/by cofactors also contribute to the specificity of each helicase towards its NA substrates (2, 3). In many cases, however, the molecular determinants of helicase specialization and the repertoire of functional helicase substrates are uncertain or were probed superficially with a few model NA constructs in minimalist assays. New approaches based on combinatorial screening of NA substrate libraries could help address this shortcoming.

NA aptamers that bind specifically to a target molecule can be isolated from large NA sequence libraries by Systematic Evolution of Ligands by EXponential enrichment (SELEX) (11). In a standard SELEX assay, the target of interest is immobilized on a carrier (usually a bead) to enable isolation of target-bound NA species after incubation with a NA sequence library. After elimination of the supernatant containing the non-binding sequences, the selected NA species are eluted from the carrier, amplified by (RT)-PCR, and converted into a new pool of NA sequences. This pool is subjected to another round of selection-amplification and the process is repeated iteratively until the NA pool is enriched enough with target-specific aptamers. SELEX can been adapted to new tasks such as isolation of natural aptamers from genomic libraries (12), discovery of synthetic ribozymes (13), or targeting of small-molecules with bead-immobilized NA libraries (14). The SELEX (or its variant) protocol usually yields three-dimensional NA motifs with specifications matching the desired requirements.

We surmised that a SELEX-like strategy could help discover functional helicase substrates. Indeed, the strand separation activity of helicases provides a simple means to isolate active sequences upon unwinding of bead-immobilized NA duplexes (**Fig. 1A, B**). A selection based on helicase activity should allow discrimination of NA sequences interacting productively with the enzyme (released in supernatant) from all unreactive species, i.e., non-interacting sequences and unproductive aptamers (trapped in bead-bound duplexes). Enrichment would be obtained by (RT)-PCR of the supernatant fraction followed by assembly of a new pool of NA duplexes for the next round of selection. This new iterative scheme, hereafter termed H-SELEX (Helicase-SELEX), combined with deep sequencing of the winner sequence libraries, should allow extensive, combinatorial exploration of helicase substrate requirements. Moreover, we surmised that cofactor(s) may be included in H-SELEX screens to probe helicase partner role(s) or, alternatively, to evolve orthogonally controlled helicase substrates for synthetic biology. The latter would represent a new class of regulatory NA

switches (e.g. riboswitches) that may work in various *in vitro* and *in vivo* settings exploiting the activities of helicases. H-SELEX would thus represent the first scalable method to evolve synthetic riboswitches from scratch rather than upon grafting of an 'expression platform' to an independently selected aptamer, as in traditional approaches.

Here, we demonstrate the power and versatility of H-SELEX using the transcription termination factor Rho helicase from *Escherichia coli* as model system. Rho disrupts transcription complexes halted along the *E. coli* chromosome in a manner than depends on the availability of *Rut* (Rho utilization) sites within nascent transcripts (**Fig. S1**) and on the action of cofactors (4-6). Natural riboswitches modulating access to *Rut* sites as function of the concentration of a protein (15) or metabolite (16, 17) ligand also represent to date the only known examples of riboswitches governing the function of an helicase.

First, we used H-SELEX and a library of RNA:DNA duplexes bearing sequences derived from the *E. coli* genome to explore the natural substrate repertoire of Rho. We show that enrichment in helicase substrates can be readily achieved by H-SELEX and that selected candidates include known *Rut* sites and many new instances spread along the *E. coli* genome. We also show that inclusion of a Rho cofactor (NusG) in the selection scheme can modulate the H-SELEX outcome and help probe specificity determinants at genome scale. Then, we used a library of synthetic RNA:DNA duplexes and serotonin (5-HT) as a model of orthogonal cofactor to evolve 5-HT-dependent riboswitches by H-SELEX. We show that the riboswitches can control the activity of Rho *in vitro* (duplex unwinding assay) an *in vivo* (reporter expression assay), thereby paving the way towards new, helicase-dependent regulatory modalities for synthetic biology and biotechnology.

RESULTS AND DISCUSSION

H-SELEX screening of Rho helicase substrates at genome scale

Bead-immobilized RNA:DNA duplexes containing a 5'-tailed *Rut* sequence can be unwound by Rho in an ATP-dependent manner (18). Immobilization on streptavidin-coated beads does not affect Rho efficiency or rate of duplex unwinding (18). We used these properties in a H-SELEX scheme aimed at selecting active *Rut* motifs for the Rho helicase from RNA sequence libraries. We designed DNA templates containing a variable sequence region framed by constant sequences allowing *in vitro* transcription and hybridization of the resulting transcripts with a biotinylated oligonucleotide (**Fig. 1A, B**). In a first setup, the variable sequence region was a library of DNA sequences (~90 base pairs [bp]) providing deep coverage of the *E. coli* genome. The resulting RNA:DNA constructs thus contained distinct *E. coli* sequences in their 5'-overhanging RNA tails (**Fig. 1A, B**), some of which likely are natural *Rut* sites. We also hybridized a complementary oligonucleotide (Block oligo in **Fig. 1A**) to the upstream constant region to limit risks of interference upon intramolecular pairing with the variable sequence. The 'reporter' DNA strand contained a 5'-biotin for immobilization of the RNA:DNA duplex library on streptavidin-coated magnetic beads (**Fig. 1A, B**). The high stability of the 57 bp-long, reporter DNA:RNA helix (ΔG^0 = -96.8 kcal/mol) ensures that RNA strands are released in the supernatant only upon ATP-dependent, Rho-directed unwinding of the bead-affixed duplexes (9), thereby providing for a robust means of functional selection based on helicase activity. Iterative rounds of such a tightly controlled selection coupled to RT-PCR amplification of winner sequences (H-SELEX assay; **Fig. 1B)** should thus yield natural substrate sequences optimally tailored to the enzyme requirements and reaction conditions.

To test this proposal, we performed parallel H-SELEX experiments with Rho in presence or absence of the NusG transcription factor. Since NusG stimulates Rho-dependent

transcription termination at suboptimal *Rut* sites (19), we wondered if this property would translate into selection of distinct $NusG^+$ vs $NusG^-$ pools of substrates.

Our starting library of RNA:DNA constructs (R_0) displayed a significant basal reactivity towards the Rho helicase (**Fig. 1C**), consistent with many *Rut* sites in the *E. coli* transcriptome (19-21). In the first H-SELEX round (R_1), ~4.10^{14} bead-immobilized R_0 constructs (coverage > 10^7 x) were incubated with Rho, either without or with NusG, before addition of Mg-ATP to initiate the helicase reaction (see **Table S1** for details on rounds). About 20% of the R_0 transcripts were recovered from the supernatant after 5 min of reaction at 37°C, irrespective of the presence/absence of NusG. In each case, the selected R_0 transcripts (i.e. unwound 'winners' from the reaction supernatant) were amplified by RT-PCR and resulting DNA templates were transcribed with T7 RNA polymerase to generate a pool of transcripts enriched in Rho-responsive sequences (**Fig. 1B**). This R_1^- or R_1^+ pool was then converted into bead-affixed RNA:DNA duplexes for the next round of H-SELEX (round 2, yielding pool R_2^- or R_2^+ for the $NusG^-$ or $NusG^+$ condition, respectively). This procedure was repeated iteratively while reducing the helicase reaction time to increase selection stringency (from as much as 300 s down to 20 s from round 6 and beyond). We observed a gradual increase in the reactivity of the pools of RNA:DNA duplexes, which started to plateau at round 9 (**Fig. 1C**). Moreover, the duplex pools obtained in presence of NusG were slightly more reactive than their $NusG^-$ counterparts (**Fig. 1C**). These data support that H-SELEX can readily drive selection of helicase-suited NA sequences in a manner that depends on selection conditions (e.g. ± NusG).

H-SELEX charts Rut sites along the E. coli genome

Genome-wide mapping of *Rut* sites has not been achieved yet. Identification of Rho-dependent termination loci by transcriptomic approaches (usually upon partial/transient Rho inactivation) is instructive but imprecise and biased by indirect effects (e.g. Rho-dependent deregulation of

transcription factors affects their respective regulons), by translation masking intragenic *Rut* sites in most conditions, and by essential 3'→5' exonucleases trimming transcript 3'-ends extensively and in a locus-specific manner (19, 22, 23). There are no such issues with the well-controlled conditions of H-SELEX.

To map the H-SELEX-enriched sequences to the *E. coli* genome, we deep-sequenced the R_{10}^{+} and R_{10}^{-} pools (respectively obtained with and without NusG) as well as the starting library R_0. Comparison of the $R_{10}^{\pm}$ and R_0 pools revealed hundreds of discrete H-SELEX-enriched peaks (hereafter called *Rut* peaks but not to be confused with actual *Rut* sites; see supplementary discussion and Fig. S2) spread along both strands of the *E. coli* genome (**Fig. S3A-B** and **Dataset S1**; 2-fold median enrichment [FE] cutoff). Clusters of *Rut* peaks were observed, in particular in prophages (**Fig. 2A**), as expected (24), and in regions encompassing CRISPR loci (**Fig. 2B**; note that CRISPR loci were shown to be targeted by Rho in *Salmonella* and *Vibrio* relatives (25)). By contrast, some regions hardly contain *Rut* peaks, such as the core genome region containing *nusG* and other essential genes encoding ribosome and RNA polymerase subunits (**Fig. 2C**).

About 20% more *Rut* peaks were detected in the $NusG^{+}$ condition (**Fig. 3A**), consistent with NusG stimulation of Rho activity at suboptimal *Rut* sites (19). Notwithstanding, peak length (P_L) distributions were comparable for the $R10^{-}$ and $R10^{+}$ pools (**Fig. S3C**) and a vast majority of *Rut* peaks were detected at similar locations in both $NusG^{+}$ and $NusG^{-}$ conditions (**Fig. 3A** and **Dataset S1;** median FE > 1 for overlap threshold).

Rut peaks were observed at expected locations for well-characterized terminators such as the Rho-dependent attenuator in the 5'UTR of the *rho* gene (**Fig. 3B**; P_L ~160 nt) (26) or the trpt' terminator at the end of the *trp* operon (**Fig. 3C**; P_L ~240 nt) (27). Two *Rut* peaks (P_L ~70-100 nt) at the start of the *trp* operon suggest that premature Rho-dependent termination may also contribute to the tryptophan-induced attenuation of the operon (28). Interestingly, the

upstream *Rut* peak overlaps with the binding site for sRNA RydC (**Fig. 3C**, pink arrow), which stimulates *trp* expression (29) and may thus control premature Rho-dependent termination in a manner akin to regulation of the *rpoS* gene by sRNAs DsrA, ArcZ, and RprA (30). These observations illustrate the wealth of information and new hypotheses generated by H-SELEX mapping.

We detected *Rut* peaks for 81% (34/42) of experimentally validated Rho-dependent termination loci (**Table S2** and **Fig. 3B-D, S4**), although some fall below the FE ≥ 2 threshold (**Fig. S5A-C**). Seven of the eight loci without detectable *Rut* peaks are located in 5'UTRs (**Table S2** and **Fig. S5D, E**). Most correspond to attenuation signals governed by riboswitch structures (*mgtA*, *lysC*, *ribB*, *sugE*, *thiC*) (16, 17, 31), which may thus not fold adequately for productive interaction with Rho in the conditions of our H-SELEX assay. In the case of *mgtA*, Rho-dependent attenuation may be specific to *Salmonella* (16) since it was also not detected in *E. coli* previously (17, 19). Notwithstanding, we found *Rut* peaks linked to riboswitch-dependent regulation in the 5'leader regions of *corA* and *thiB* (**Fig. S4A, B**) and in the upstream part of the *thiM* CDS (**Fig. 3D**), consistent with the reported contributions of Rho-dependent termination in these systems (17, 32). Moreover, ~30% of *Rut* peaks overlap with bicyclomycin sensitive transcripts (BSTs) identified by transcriptomics (19), in a manner often consistent with the greater precision of H-SELEX (**Fig. S6, Table S2**, and **Dataset S2**). Taken together, these data demonstrate that H-SELEX is a powerful means to identify genuine helicase interaction sites at genome/transcriptome scale.

A majority of *Rut* peaks are antisense to genes (**Fig. 3E** and **Dataset S2**), consistent with the prominent role of Rho in suppression of pervasive antisense transcription (19). The location of 7-10% of *Rut* peaks at/near promoters (**Fig. 3F** and **Dataset S3**) was not anticipated from previous work and may also contribute to suppress pervasive transcription and punctuate transcription units. Another unexpected, albeit rarer finding is the overlapping of *Rut* peaks

with intrinsic terminators (**Fig. 3F** and **Dataset S4**). This suggests backup or redundancy between both types of termination signals in the corresponding 3'UTRs and is reminiscent of the situation reported for some *Bacillus subtilis* transcripts (33). Similarly unanticipated is a small subset of *Rut* peaks that form overlapping tandems on opposite strands, often near divergent promoters (**Fig. 3F, S2A**). This particular arrangement could not be envisioned from anisotropic C>G 'bubble' models of Rho-dependent termination (20, 34) but is supported by *in vitro* transcriptional probing of bidirectional *Rut* tandems (**Fig. S7**).

Consistent with previous predictions (20, 30, 31), we found *Rut* peaks in many 5'UTRs (**Fig. 3E, S8** and **Dataset S2**) where they may contribute to Rho-dependent attenuation mechanisms (4). *Rut* peaks are also found between genes of the same operon (**Fig. S9A-C**) where they may also have conditional regulatory roles, as in the *gal* operon (35). In fact, many *Rut* peaks are located between genes, irrespective of the relative sense/antisense orientation of the genes, where they neatly punctuate the genome to either protect or regulate the downstream units (**Fig. 3B-C, S9**). These configurations are unlikely to occur by chance and further highlight the functional relevance of our H-SELEX assay.

We did not find consensus motifs associated to the *Rut* peaks that would easily explain Rho preference for the corresponding sequences. When compared to a control pool of *Rut*-less sequences, *Rut* peak sequences are usually enriched in U residues and YC dinucleotides, are poorer in G residues, have a lower propensity to form secondary structures, and display U>G and C>G biases (**Fig. 4A**). These features are generally consistent with studies of individual Rho-dependent terminators and with models locating Rho-dependent signals in C>G 'bubbles' (20, 34). Yet, $NusG^{+}$-specific *Rut* peaks can deviate from these rules, about half of them displaying an 'inverted' G >C bias (**Fig. S10**). We selected two of these G>C-biased *Rut* peaks for further characterization and indeed detected downstream *in vitro* Rho-dependent termination signals in presence of NusG (**Fig. 4B**).

The $NusG^+$-specific *Rut* peaks are generally shorter and smaller (i.e. lower FE values) than their NusG-independent counterparts (**Fig. 4C**). These characteristics (and the distinct sequence composition evoked above) may contribute to explain the $NusG^+$ requirement. Intriguingly, the $NusG^-$-specific *Rut* peaks are similarly shorter and smaller than NusG-independent peaks but are also characterized by unusually low R_0 read coverages (**Fig. 4C**). Low R_0 coverage may create signal artifacts (e.g. upon partition of the R_0 library for the ±NusG experiments) and urges caution about the significance of the $NusG^-$ dependence. Consistent with this limitation and contrary to $NusG^+$ peaks (**Fig. 4B**), we could not reproduce the $NusG^-$ condition upon testing representative $NusG^-$ peaks in *in vitro* transcription termination assays (**Fig. S11**).

The identification of discrete *Rut* peaks by H-SELEX and lack of strong consensus rules for the peak sequences may seem paradoxical but are in line with known structural features. Indeed, *Rut* recognition is mediated by the composite Primary Binding Site (PBS) of the Rho hexamer, which contains a 5'YC (Y is a pyrimidine) binding pocket on each monomer (36). Hence, various arrangements of 5'YC dinucleotides in the RNA chain may compose a *Rut* site if the spacing between the single-stranded 5'YC dinucleotides permits proper interaction with several Rho monomers (Fig. S1 and supplementary discussion). This may be achieved with short single-stranded RNA (ssRNA) stretches between 5'YC dinucleotides (36) or in the context of elaborate RNA structures ensuring an optimal spatial arrangement of 5'YC dinucleotides. In line with this idea is the discovery of natural riboswitches able to govern Rho activity upon structural *Rut* remodeling (16, 32).

A variant of H-SELEX to evolve synthetic riboswitches

In a distinct H-SELEX setup, we used a synthetic sequence library to seek conditional substrates able to elicit Rho activity in presence of a small-molecule inducer. We prepared a library of

RNA:DNA duplexes ($^{s}R_0$ library) wherein the variable sequence region (**Fig. 1A**) is a 80-nt long randomized region (N_{80}) derived from a synthetic oligonucleotide library. In principle, this fully randomized N_{80} region grants exploration of a much larger conformational space (1 nmole ~ 6.10^{14} diverse sequences) than does the library of *E. coli* origin (maximal diversity of ~10^7 sequences) and, thus, provides a better chance to find RNA motifs matching the desired requirements. Another important feature of the H-SELEX variant procedure is the mix of positive and negative selection rounds, respectively performed in presence and absence of the inducer, to favor the enrichment in inducible versus constitutively active sequences. In positive selection rounds (+ inducer), reactive ssRNA species are processed as described above (**Fig. 1B**) whereas in negative selection rounds (- inducer) unreactive RNA:DNA duplexes are instead harvested and used to prepare the next round library. We chose 5-HT, an important metabolite and disease biomarker (37) for which no RNA sensor has been developed yet, as model of orthogonal inducer. Using a control RNA:DNA duplex bearing a strong *Rut* site (38), we verified that 5-HT has no adverse effect on Rho helicase activity (**Fig. S12A, B**).

Preliminary experiments suggested that negative selections performed in early rounds of H-SELEX favor enrichment in poorly reactive sequences rather than in inducible sequences. We thus chose to start with positive selection rounds and to implement negative selections only once the sequence pool is maximally enriched in Rho-responsive species. We performed the first H-SELEX round ($^{s}R_1$) with ~10^{15} bead-affixed $^{s}R_0$ duplexes, and recovered the ssRNA species released by Rho in the supernatant after 2 min of reaction in presence of 5-HT (see **Table S3** for details). The ssRNA products were amplified by RT-PCR and the resulting DNA templates were used to generate the next-round library of bead-affixed RNA:DNA duplexes (**Fig. 1B**). After three rounds of positive selection, we tested the RNA:DNA duplexes ($^{s}R_3$ pool) with our standard helicase assay where all reactants evolve freely in solution. In presence of 5-HT, the $^{s}R_3$ duplexes were less reactive under these conditions than when immobilized on

streptavidin-coated beads (**Fig. 5A**, green curves). Such bead-tethering bias was not seen in absence of 5-HT (**Fig. 5A**, red curves) or with the genomic version of H-SELEX (**Fig. 1C**, black and gray R_{10} curves) but is sometimes observed in conventional SELEX experiments (39). To solve this issue, we performed positive selection rounds with libraries of untethered RNA:DNA duplexes and used polyacrylamide gel electrophoresis (PAGE) instead of bead fractionation to purify ssRNA products (rounds 4-10). After several rounds with this protocol, the reactivity of the duplex pool reached a maximum irrespective of the presence/absence of 5-HT (**Fig. 5B**, rounds 8-10). We thus performed negative selections to seek 5-HT-responsive candidates from the enriched pool of substrates (rounds 11 to 13). We observed a marked drop in duplex pool reactivity specifically in absence of 5-HT (**Fig. 5B**). The $^{s}R_{13}$ duplexes were much more efficiently unwound by Rho in presence than in absence of 5-HT (**Fig. 5C**), thereby demonstrating the interest of H-SELEX to evolve inducible NA switches.

In subsequent H-SELEX rounds, we combined negative (-5-HT) and positive (+5-HT) selection steps at every round to improve the pool of 5-HT-responsive candidates. We reverted to fractionation of bead-immobilized duplex libraries to streamline the process and used a 10-fold lower concentration of 5-HT to increase selection stringency. After several iterations of this mixed selection scheme (rounds 14-21), the reactivity of the duplex pool slightly declined (**Fig. 5B**, top graph) but remained strongly stimulated by 5-HT (bottom graph, 10 mM curve) and even evolved favorably at the lower inducer concentration (1 mM curve).

Characterization of 5-HT riboswitch candidates

To identify 5-HT-responsive sequences, we deep-sequenced the intermediate $^{s}R_{13}$ and final $^{s}R_{21}$ pools. Both pools displayed comparable nucleotide compositions with *Rut*-like C>G biases along most of the N_{80} region, but the enrichment in unique sequences was much higher for $^{s}R_{21}$ than for $^{s}R_{13}$ (30 vs 0.5%; **Fig. S13**). We did not detect consensus motifs apart from U/C-rich

stretches at non-conserved positions, including near sequence 3'ends (see examples in **Fig. S12A**). Yet, the two top-ranking $^{s}R_{21}$ sequences (21a and 21b) head large sequence clusters (**Fig. S14**) and can be tentatively folded into similar secondary structures (**Fig. 6A**).

Consistent with the H-SELEX scheme, both 21a and 21b sequences are poor Rho substrates in absence of 5-HT but elicit strong duplex unwinding responses in its presence (**Fig. 6B**). Among tested sequences, only 13b from the $^{s}R_{13}$ pool triggers a comparably strong inducible response (**Fig. S12C**). Sequences 21a, 21b, and 13b were thus characterized further. Sequence 21b elicits the strongest Rho helicase response, whichever the 5-HT concentration tested, but is also the least selective for 5-HT vs the analog tryptamine (**Fig. S15A, B**). All three riboswitch sequences bind Rho tighter in presence than in absence of 5-HT whereas Rho affinity for the control *aRut* sequence does not vary (**Fig. S15C**). Yet, this does not fully explain the performances of the riboswitches since Rho binding remains 10-20-fold weaker than for *aRut*, i.e. in the same range than for the unproductive *iRut* sequence (K_d > 40 nM) (38). This paradox may stem from the absence of ATP in the equilibrium binding reactions (to avoid duplex unwinding), although the non-hydrolysable AMPNP analog does not improve binding (**Fig. S15C**). Alternatively, 5-HT could stimulate a downstream, ATP-dependent step upon productive loading of the riboswitches, such as RNA entry within (and/or closure of) the Rho ring, as NusG does with suboptimal *Rut* substrates (40). Elucidation of the exact molecular/structural mechanisms of the 5-HT riboswitches will require further work, well beyond the present study.

To determine if the riboswitches can govern Rho activity *in vivo*, we used a plasmid reporter system wherein the $_{sf}$GFP gene is under control of a *Rut* sequence (*aRut*, 21a, 21b, or 13b sequence) located in its 5'-UTR (**Fig. 6C, S16**). The plasmids harbor a second reporter gene (dsRED-Express2) for signal normalization. We also prepared plasmid derivatives wherein the *Rut* sequence is absent ('None' construct) or replaced by the non-productive *iRut*

sequence (38). Plasmids were transformed in *E. coli* and normalized $_{sf}$GFP reporter signals were determined by flow cytometry after growth in the presence or absence of 5-HT. Signals for the three controls (None, *aRut, iRut*) were unaffected by 5-HT, whereas 5-HT induced decreases of the reporter signals with the 21a, 21b, and 13b sequences (**Fig. 6C, S17**). These observations are consistent with Rho-dependent attenuation mediated by the riboswitches in presence of 5-HT. Surprisingly, the lowest 5-HT effect was for the top-ranking 21a sequence, which may thus not fold optimally or fast enough *in vivo*. Similar inducible reporter responses were obtained with the analog tryptamine whereas N-acetyl-5-hydroxy-tryptamine could not regulate the riboswitches (**Fig. S17**). This suggests that serotonin and tryptamine use a similar binding mode, possibly involving protonated primary amine contacts to RNA phosphate groups and indole ring stacking to nucleobases (41). However, differences in intracellular concentrations of the analogs cannot be excluded and may mitigate the apparent specificity and contribution of individual functional groups to riboswitch function (e.g., compare **Fig. S15B** and **S17**).

Theophylline-sensing riboswitches, which represent to date the most developed class of synthetic riboswitches, operate *in vivo* at similar millimolar ligand concentrations and with comparable dynamic ranges for first-generation specimens (42). Optimization of our riboswitch sequences and reporter scaffold, mimicking strategies used for the theophylline riboswitches (42) may thus lead to second-generation, serotonin-sensing systems with improved properties. In any case, our data demonstrate that H-SELEX is a valuable tool to evolve riboswitches from scratch, thereby granting access to a new type of synthetic regulation based on helicase function *in vitro* or *in vivo*.

CONCLUSIONS

Helicases are NA remodeling enzymes that play critical roles in the biology of living organisms and viruses (1-3). Many specimens have broad cellular duties and NA target repertoires that vary with cellular status, external stimuli, intracellular localization, availability of cofactors, etc. (1-3). Although omics approaches can help untangle this biological complexity, they rely on *in vivo* setups not easily tuned to test individual parameters or rule out indirect effects. To help address this limitation, we developed H-SELEX, an enzymatic screening and amplification approach that yields functional NA substrates tailored to the tested helicase and experimental conditions. The starting library can contain natural or synthetic NA sequences and, in principle, can accommodate NA architectures distinct from the 5'-tailed RNA:DNA duplex designed for 5'→3' translocating RNA helicases such as Rho (**Fig. 1A**), although this remains to be tested (see supplementary discussion). Alternative NA designs may include 3'-tailed duplexes, RNA- or DNA-only constructs, short duplexes suitable for non-processive DEAD-box helicases (1, 3), elaborate NA structures containing a *trans*-releasable strand (for selection), and combinations thereof. Moreover, H-SELEX may be implemented with wild-type helicases or partially active mutants, with or without cofactor(s), and in many well-controlled setups allowing deep probing (or manipulation) of helicase specificity (see supplementary discussion).

We used H-SELEX to identify functionally active binding sites for the Rho helicase (*Rut* sites) within the full *E. coli* transcriptome. We discovered ~3300 *Rut* peaks (**Fig. 2B** and **Dataset S1**), i.e. about two orders more than sites known with some precision previously (**Table S2**). Transcriptomic studies support the existence of several thousands of Rho-dependent loci (19, 24, 30) but cannot locate them accurately (22, 23), thereby emphasizing the interest and complementarity of H-SELEX (**Fig. S6**). In addition, H-SELEX can identify *Rut* targets in translated or poorly transcribed sections of the genome that are hard to detect otherwise and thus provides a broad, yet detailed picture of the helicase substrate repertoire. A few known *Rut*

sites were not detected by H-SELEX (**Table S2**), probably because these sites require experimental conditions, e.g. specific salts/cofactors or a deeper coverage by the R_0 library, distinct from those we used. Although burdensome, replicating H-SELEX with distinct R_0 fragment size ranges might help address the limited R_0 coverage in some regions (**Fig. 3D**) while H-SELEX enrichment peaks observed in several conditions (e.g. overlap $NusG^+$/$NusG^-$ peaks in **Fig. 3A** and **Dataset 1**) identify core constituents of the natural substrate repertoire.

In another proof-of-concept H-SELEX trial, we evolved the first synthetic riboswitches able to govern the activity of a helicase (**Fig. 6**). The riboswitches are functional *in vitro* and *in vivo* at the 5-HT ligand concentration (10 mM) used during early selection rounds (**Fig. 5B**). Their performance is comparable to that of other first-generation riboswitches (14, 42), suggesting that their ligand affinity/specificity and dynamic range could be similarly improved, e.g. upon partial re-randomization and further screening. Alternatively, H-SELEX may be adapted to improve riboswitch features earlier, e.g. by performing initial positive selection rounds with lower ligand concentrations or negative selection rounds with ligand analogs.

Overall, we expect H-SELEX to become a method of choice to explore the functional repertoire of helicases or to evolve synthetic biosensors or regulatory circuits based on helicases. As for conventional SELEX (11), all H-SELEX steps (**Fig. 1B**) are in principle compatible with automation. Despite the investment cost, H-SELEX automation should increase throughput, allow sample multiplexing and easier exploration of experimental determinants for both the genomic and synthetic versions of H-SELEX, and should thus facilitate the deployment of H-SELEX for other helicases, including ones using distinct NA chemistries or strand separation mechanisms.

MATERIALS AND METHODS

Materials

Unless specified otherwise, chemicals and enzymes were from Sigma-Aldrich and New England Biolabs, respectively. Radionucleotides were obtained from PerkinElmer and synthetic oligonucleotides (**Table S4**) from Eurogentec. Rho (concentrations in hexamers) and NusG proteins were prepared as described (43). DNA templates for *in vitro* transcription (**Table S5**) were prepared by standard PCR (43). Plasmid pFACS-aRut (**Fig. S16**) was obtained by subcloning a synthetic DNA fragment containing the dsRED-Express2 and $_{sf}$GFP genes under control of divergent pTac promoters between the AatII and AvrII sites of plasmid pZE12luc (44). The DNA fragment also bears the highly-efficient *aRut* sequence (38) between the $_{sf}$GFP gene and its driving promoter. Plasmids pFACS-iRut, pFACS-13b, pFACS-21a, pFACS-21b, and pFACS-RutLess were obtained by subcloning PCR-amplified DNA fragments bearing, respectively, sequence *iRut*, 13b, 21a, 21b (**Fig. S12A**) or no sequence in place of *aRut* into the XhoI and BglII sites of plasmid pFACS-aRut.

RNA:DNA duplex library for 'genomic' H-SELEX

The library of DNA fragments was prepared as described (12) with minor modifications. A pellet of a 5 mL overnight LB culture of strain MG1655 was dissolved in 600 µL of lysis buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA, 0.6% SDS, 0.12 mg/ml Proteinase K) and incubated for 1h at 37°C. The mixture was extracted with phenol-chloroform and genomic DNA (gDNA) was recovered from the aqueous phase after precipitation with ethanol and resuspension in TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA). The gDNA was incubated with RNase A (50µg/ml) for 1h at 37°C, sonicated with a Vibracell 75115 apparatus, precipitated with ethanol and resuspended in 100 µL TE buffer. Then, 25µg of gDNA were mixed with 1.9 nmole of primer R45-ran-rev in Klenow buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$,

and 1 mM DTT) and incubated at 95°C for 5 min before addition of dNTPs (0.3 mM each, final concentration) and 0.5U/µL of Klenow polymerase. The mixture was incubated successively at 4°C for 5 min, 25°C for 25 min, and 50°C for 5 min before addition of EDTA **(**10 mM, final concentration) and incubation for 10 min at 75°C. DNA products were purified on GeneJET column **(**Thermo-Fisher Scientific) and mixed with 3.8 nmol of ARN107-ran-For primer in Klenow buffer. Primer extension was performed with 0.5U/µL of Klenow polymerase to yield second strands of gDNA fragments following the same steps described above for first strand synthesis with primer R45-ran-rev. The gDNA fragments in the desired size range were purified by 7% PAGE (12). The resulting gDNA fragment library (~50 ng) was PCR-amplified (13 cycles) with primers FWD and REV (0.5 µM each), dNTPs (0.2 mM each), and 0.02U/µL Vent polymerase in Vent buffer (20 mM Tris-Cl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2mM $MgSO_4$, 0.1% Triton X-100) to introduce the sequences of the T7 promoter and reporter pairing region (**Fig. 1A**). For the first PCR cycle, the mixture was incubated at 94°C for 1 min, 50°C for 1 min, and 70°C for 1 min. For subsequent cycles, it was incubated for 30 s at 94°C, 30s at 50°C, and 30s at 70°C. About 200 pmoles of the resulting R_0 library of DNA fragments were transcribed with T7 RNA polymerase, as described (43). The R_0 transcripts were purified on a RNA clean&concentrator column (Zymo research). Then, ~1.5 nmoles of R_0 transcripts were mixed with, respectively, 1.1 and 2 molar equivalents of the 5'-biotinylated SEL and BLOCK oligonucleotides in hybridization buffer I (20 mM HEPES, pH 7.5, 150 mM potassium acetate, 10 mM NaCl, 100 nM $CaCl_2$). The mixture was incubated at 70°C for 2 min and cooled to 25°C for 15 min before use in H-SELEX.

RNA:DNA duplex library for 'synthetic' H-SELEX

A library of single-stranded DNA (ssDNA) sequences containing 80 randomized nucleotides flanked by two primer-binding regions for PCR (**Table S4**) was purchased from Eurogentec.

About 1 nmole of the ssDNA library was converted in 10 cycles of PCR amplification (94°C for 1 min, 50°C for 1 min, 70°C for 1 min) into double-stranded DNA (dsDNA) templates for T7 transcription (**Fig. 1B**). The initial PCR mixture contained 0.5 µM of ssDNA library, 3 µM of FWD and REV primers, 0.3 mM dNTPs (each), and 50 U of Taq DNA polymerase in 2 mL of Taq buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$). The dsDNA library was purified on GeneJET columns and used directly for *in vitro* transcription with T7 RNA polymerase. The transcription reaction was performed as described (43) in parallel microtubes to accommodate the larger volume (2.5 mL). Following concentration by ethanol precipitation, the resulting ssRNA library was purified on custom-made, 1 mL Sephadex G50 spin columns, phenol extracted, ethanol precipitated, and stored in $M_{10}E_1$ (10 mM MOPS, 1 mM EDTA, pH 6.5) buffer at -20°C. To allow monitoring of reaction species during Helicase-SELEX, a fraction of the ssRNA library (~10 pmoles) was dephopshorylated with calf intestine phosphatase and ^{32}P-labeled with γ[P32]-ATP and T4 polynucleotide kinase (43). The ^{32}P-labeled ssRNA and 1 nmole of unlabeled ssRNA library were then mixed in hybridization buffer II (150 mM potassium acetate, 20 mM HEPES pH 7.5, 0.1 mM EDTA) before addition of 1.1 molar equivalent of 5'-biotinylated SEL oligonucleotide. The mixture was incubated at 70°C for 2 min and cooled to 25°C for 15 min before use in H-SELEX.

H-SELEX assays

The starting libraries of RNA:DNA duplexes were immobilized on streptavidin-coated magnetic beads (Dynabead M-280 or T1, Invitrogen) following a protocol described previously (18). Briefly, beads (1 µL of bead slurry per pmole of RNA:DNA duplexes) were washed with BW buffer (1M KCl, 5 mM Tris-Cl, pH 7.5, 0.5 mM EDTA) before addition to the crude mix of RNA:DNA hybrids (from section above) and incubation for 1h at 22°C. Beads were then separated from supernatant on a Magnum FLX magnet (Alpaqua), washed first with BW buffer

and then with hybridization buffer I or II (depending on the library; see above) to remove unbound RNA/DNA species. Bead-affixed duplexes were pre-incubated with Rho (± NusG for genomic H-SELEX; +5-HT for synthetic H-SELEX) in helicase buffer (hybridization buffer supplemented with 0.1 mg/mL of acetylated BSA) for 10 min at 37°C. Helicase reactions were initiated with $MgCl_2$ and ATP (1mM, final concentrations) and incubated for 300 s (genomic H-SELEX) or 120 s (synthetic H-SELEX) at 37°C under shacking (300 rpm) to homogenize the bead suspension. Reaction were quenched with 20 mM EDTA and the supernatants containing the released ssRNA strands were magnetically separated from the beads. The ssRNA strands were purified on a RNA clean&concentrator column and reverse transcribed with REV primer (1.2 molar equivalent) and Superscript III reverse transcriptase (Invitrogen; 2U per pmole of ssRNA) in the buffer supplied with the enzyme. Reaction mixtures were incubated for 1h at 50°C and for 15 min at 70°C. The ssDNA products were amplified by PCR (12 cycles) with Taq DNA polymerase using FWD and REV primers (0.6 µM, final concentrations) to generate the dsDNA template library for the next round of T7 transcription, assembly of the RNA:DNA duplex library, and functional selection by H-SELEX (performed as described in section above).

Variations in H-SELEX conditions (library size, concentrations of reactants, incubation time, etc.) were introduced in subsequent rounds to limit biases and increase selection stringency (see **Tables S1** and **S3**). Notably, an alternative selection approach was used for synthetic H-SELEX (rounds 4 to 13) to avoid immobilization biases. Here, the library of RNA:DNA duplexes was prepared with non-biotinylated REV oligonucleotide and purified by native 6% PAGE as described (43). Purified RNA:DNA duplexes (80 nM) were mixed with Rho (80 or 320 nM) and serotonin (0 or 10 mM) in helicase buffer and incubated for 10 min at 37°C. Helicase reaction was initiated with a mix of $MgCl_2$ (1mM, final concentrations) and oligonucleotide TRAP (400 nM, final concentration), which is complementary to the REV

oligonucleotide. Reaction was quenched with 0.5% SDS and 20 mM EDTA after 0.5 to 20 min of incubation at 37°C. The pool of released RNA strands (positive selection scheme) or unreactive RNA:DNA duplexes (negative selection scheme) was purified on a 9% PAGE gel containing 0.5% SDS (43). Reverse transcription and PCR amplification to generate the dsDNA template library for the next round of H-SELEX were performed as above.

Sequence processing and bioinformatics analysis

Bioinformatics analyses were performed with in-house Python scripts and software tools from the Galaxy server (45). The dsDNA libraries obtained by H-SELEX were analyzed by 2x150 base paired-end sequencing on a NextSeq (genomic H-SELEX) or Miseq (synthetic H-SELEX) Illumina instrument at the I2BC sequencing facility (CNRS, Gif-sur-Yvette). Starting, blunt-ended dsDNA pools (~2.5 µg each) were processed by the I2BC facility using standard procedures and the NextSeq 500/550 Mid Output Kit v2 or MiSeq reagent kit v2 (Illumina). Samples were supplemented with DNA from coliphage phiX174 to mitigate potentially low sequence diversity of H-SELEX enriched pools. Multiplexed sequencing of the pools resulted in 4.6 to 9.5 x 10^6 paired-end reads per pool after quality control filtering (FastQC v0.11.5) and adapter trimming (Cutadapt1.15). Paired-end reads (≥30 nt) were concatenated and coliphage phiX174 sequences were expunged from sequence libraries by selecting reads containing FWD and REV primer-binding regions using Fastaq-join v1.1.2 (Max % difference between matching segments: 8; Min length of matching segments: 6), Filter Fasta v2.1, and Cutadapt v1.16 tools.

For genomic H-SELEX, curated reads were mapped on the MG1655 genome (NC_00096.3) with Bowtie2 (v2.4.2, default options). Coverage and RPM normalization were performed with Bedtools v2.29.2. The $Log_{10}FE$ profiles were obtained with MACS2 bdgcmp v2.1.1 using the R_{10}^+ or R_{10}^- sample as input and R_0 sample as control. Peak calling was performed with MACS2 bdgpeakcall v2.1.1 (cutoff: 0; Min peak length: 30; Max gap: 30). We

used median R_0 read count per nucleotide position (≥ 1) and median $Log_{10}FE$ (≥ 0.3) values to select *Rut* peaks. Overlaps among peaks or between peaks and characteristic regions (e.g. genes, terminators) were obtained with Bedtools Intersect intervals v2.30.0. A control pool of 3,000 *Rut*-less sequences was built by sampling 130 nt-long regions of the MG1655 genome that were not intersecting with *Rut* peaks on either strand using the seqtk_sample utility of Galaxy.

For synthetic H-SELEX, curated reads were grouped using the collapse tool of Galaxy and clustered with BLASTN (0.001 e-value cutoff) using top-ranked sequences as queries. In parallel, raw reads were processed and clustered with FASTAptamer (46) using Levenshtein edit distance = 3 and Min number of sequences per cluster = 3. Both approaches yielded similar results for top-ranked clusters (we stopped FASTAptamer analysis before all bottom-ranking clusters were defined).

Duplex unwinding kinetics

Duplex substrates were prepared by hybridizing ^{32}P-labeled transcripts (from a given round library or corresponding to a single winner sequence) with the REV oligonucleotide and were purified by native 6% PAGE (18, 38). For genomic H-SELEX libraries, duplexes containing ^{32}P-labeled REV and unlabeled RNA strands were used instead. Helicase kinetics were determined with the purified ^{32}P-labeled RNA:DNA duplexes, as described (38). Briefly, duplexes (5 nM) were mixed with Rho hexamers (20 nM) (± 40 nM NusG for genomic H-SELEX libraries) in helicase buffer (supplemented with indicated concentration of 5-HT or analog) and incubated for 3 min at 37°C. Then, 1 mM $MgCl_2$, 1 mM ATP, and 400 nM oligo TRAP were added to the helicase mixture before incubation at 37°C. Reaction aliquots were taken at various times and mixed with two volumes of quench buffer (10 mM EDTA, 1.5% SDS, 300 mM sodium acetate, 6 % Ficoll-400) before loading on 9% PAGE gels that contained 1X TBE and 0.5% SDS. In control experiments, helicase reactions were performed with bead-

tethered duplexes. In this case, ^{32}P-labeled transcripts were hybridized with 5'-biotinylated SEL oligonucleotide and immobilized on beads as described above. Helicase reactions were performed and analyzed by PAGE as for the non-biotinylated duplexes. Detection and quantification of gel bands were performed by phosphorimaging with a Typhoon FLA-9500 imager (18, 38). Reaction time-courses obtained with unique duplex sequences were fitted to an equation describing pseudo-first order kinetics, as described for the Rho helicase (38).

Transcription termination experiments

Standard transcription termination experiments were performed as described (20). Briefly, DNA template (0.1 pmol), *E. coli* RNAP (0.3 pmol), Rho (1.4 pmol), NusG (0 or 2.8 pmol), and Superase-In (0.5U/µL; Ambion) were mixed in 18 µL of transcription buffer (40 mM Tris-HCl, pH 8.0, 50 mM KCl, 5 mM $MgCl_2$, 1,5 mM DTT) and incubated for 10 min at 37°C. Then, 2 µL of initiation mix (2 mM ATP, GTP, and CTP, 0.2 mM UTP, 2.5 µCi/µL of ^{32}P-αUTP, 250 µg/mL rifampicin) were added to reaction mixtures before incubation for 20 min at 37°C. Transcription reactions were stopped with 4 µL of EDTA (0.5 M), 6 µL of tRNA (0.25mg/mL), and 80µL of sodium acetate (0.42 M) before precipitation at -20°C with 330 µL of ethanol. Reaction pellets were dissolved in denaturing loading buffer (95% formamide, 5 mM EDTA), and analyzed by denaturing 8% PAGE and Typhoon FLA-9500 phosphorimaging.

In vivo reporter assays

Reporter plasmid pFACS-aRut (**Fig. S16**) or derivative pFACS-iRut, pFACS-13b, pFACS-21a, pFACS-21b, or pFACS-RutLess was electro-transformed into DH5α cells and plated on LB-agarose plates containing 50 µg/mL carbenicillin. Fresh colonies were used to inoculate 5 mL aliquots of Wless-carbenicillin medium (tryptophan-less version of the Neidhardt supplemented MOPS defined medium (47) containing 0.2% glucose and 50 µg/mL carbenicillin). Overnight

cultures were diluted 100-fold in 1 mL of fresh Wless-carbenicillin medium containing 0 or 10 mM 5-HT, tryptamine, or N-acetyl-5-hydroxy-tryptamine and were incubated at 37°C for 2 h. Samples were then analyzed by flow cytometry with an LSRFortessa X20 cell analyzer (Becton Dickinson) equipped with 488 nm and 570 nm lasers, and 530/30 nm and 586/15 nm band-path emission filters for sfGFP and dsREDexpress2, respectively. The sfGFP and dsRED-Express2 fluorescence signals of ~10^4 cells were measured for each sample. PMT settings were adjusted for each construct/experiment but were strictly identical for each pair of ±inducer samples. Per-cell sfGFP/dsRED-Express2 ratios were determined with Flowing Software 2.5.1 (http://flowingsoftware.btk.fi).

Data availability

NGS data have been deposited in the European Nucleotide Archive at EMBL-EBI under accession codes PRJEB50170 (R_0 and R_{10} libraries) and PRJEB50172 ($^sR_{13}$ and $^sR_{21}$ libraries). Flow cytometry data have been deposited in FlowRepository under accession codes FR-FCM-Z4Q9 and FR-FCM-Z53A.

AKNOWLEDGMENTS

We thank Emilie Soares for contributions to H-SELEX development, Laetitia Cobret for technical help with reporter assays, David Gosset for advice with flow cytometry, and Jérôme Bonnet for critical reading of the manuscript. We acknowledge the sequencing and bioinformatics expertise of the I2BC HTS facility, supported by France Génomique (funded by ANR-10-INBS-09). This work was supported by ANR grants to M.B. (ANR-15-CE11-0024-2 and ANR-19-CE44-0009-01).

REFERENCES

1. Linder P & Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12(8):505-516.
2. Brosh RM, Jr. & Matson SW (2020) History of DNA Helicases. *Genes (Basel)* 11(3).
3. Valentini M & Linder P (2021) Happy Birthday: 30 Years of RNA Helicases. *Methods Mol Biol* 2209:17-34.
4. Bossi L, Figueroa-Bossi N, Bouloc P, & Boudvillain M (2020) Regulatory interplay between small RNAs and transcription termination factor Rho. *Biochim Biophys Acta Gene Regul Mech* 1863(7):194546.
5. Roberts JW (2019) Mechanisms of Bacterial Transcription Termination. *J Mol Biol* 431(20):4030-4039.
6. Hao Z, Svetlov V, & Nudler E (2021) Rho-dependent transcription termination: a revisionist view. *Transcription* 12(4):171-181.
7. Raghunathan N, *et al.* (2018) Genome-wide relationship between R-loop formation and antisense transcription in Escherichia coli. *Nucleic Acids Res* 46(7):3400-3411.
8. Abdelkrim YZ, Banroques J, & Kyle Tanner N (2021) Known Inhibitors of RNA Helicases and Their Therapeutic Potential. *Methods Mol Biol* 2209:35-52.
9. Schwartz A, *et al.* (2009) A stepwise 2'-hydroxyl activation mechanism for the bacterial transcription termination factor Rho helicase. *Nat Struct Mol Biol* 16(12):1309-1316.
10. Kanaan J, *et al.* (2018) UPF1-like helicase grip on nucleic acids dictates processivity. *Nature communications* 9(1):3752.
11. Stoltenburg R, Reinemann C, & Strehlitz B (2007) SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* 24(4):381-403.
12. Zimmermann B, Bilusic I, Lorenz C, & Schroeder R (2010) Genomic SELEX: a discovery tool for genomic aptamers. *Methods* 52(2):125-132.
13. Bartel DP & Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* 261(5127):1411-1418.
14. Boussebayle A, *et al.* (2019) Next-level riboswitch development-implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. *Nucleic Acids Res* 47(9):4883-4895.
15. Figueroa-Bossi N, *et al.* (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28(11):1239-1251.

16. Hollands K, *et al.* (2012) Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A* 109(14):5376-5381.
17. Bastet L, *et al.* (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Res* 45(12):7474-7486.
18. Soares E, Schwartz A, Nollmann M, Margeat E, & Boudvillain M (2014) The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho helicase decrypted by time-resolved Nucleotide Analog Interference Probing (trNAIP). *Nucleic Acids Res* 42(14):9270-9284.
19. Peters JM, *et al.* (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev* 26(23):2621-2633.
20. Nadiras C, Eveno E, Schwartz A, Figueroa-Bossi N, & Boudvillain M (2018) A multivariate prediction model for Rho-dependent termination of transcription. *Nucleic Acids Res* 46(16):8245-8260.
21. Sedlyarova N, *et al.* (2017) Natural RNA Polymerase Aptamers Regulate Transcription in E. coli. *Mol Cell* 67(1):30-43 e36.
22. Dar D & Sorek R (2018) High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res* 46(13):6797-6805.
23. Wang X, *et al.* (2019) Processing generates 3' ends of RNA masking transcription termination events in prokaryotes. *Proc Natl Acad Sci U S A* 116(10):4440-4445.
24. Cardinale CJ, *et al.* (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320(5878):935-938.
25. Stringer AM, Baniulyte G, Lasek-Nesselquist E, Seed KD, & Wade JT (2020) Transcription termination and antitermination of bacterial CRISPR arrays. *eLife* 9:e58182.
26. Matsumoto Y, Shigesada K, Hirano M, & Imai M (1986) Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators. *J Bacteriol* 166(3):945-958.
27. Wu AM, Christie GE, & Platt T (1981) Tandem termination sites in the tryptophan operon of Escherichia coli. *Proc Natl Acad Sci U S A* 78(5):2913-2917.
28. Yanofsky C, *et al.* (1981) The complete nucleotide sequence of the tryptophan operon of Escherichia coli. *Nucleic Acids Res* 9(24):6647-6668.

29. King AM, Vanderpool CK, & Degnan PH (2019) sRNA Target Prediction Organizing Tool (SPOT) Integrates Computational and Experimental Data To Facilitate Functional Characterization of Bacterial Small RNAs. *mSphere* 4(1).
30. Sedlyarova N, *et al.* (2016) sRNA-Mediated Control of Transcription Termination in E. coli. *Cell* 167(1):111-121 e113.
31. Adams PP, *et al.* (2021) Regulatory roles of Escherichia coli 5' UTR and ORF-internal RNAs detected by 3' end mapping. *eLife* 10.
32. Kriner MA & Groisman EA (2015) The Bacterial Transcription Termination Factor Rho Coordinates Mg(2+) Homeostasis with Translational Signals. *J Mol Biol* 427(24):3834-3849.
33. Nicolas P, *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* 335(6072):1103-1106.
34. Alifano P, Rivellini F, Limauro D, Bruni CB, & Carlomagno MS (1991) A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* 64(3):553-563.
35. Wang X, Ji SC, Jeon HJ, Lee Y, & Lim HM (2015) Two-level inhibition of galK expression by Spot 42: Degradation of mRNA mK2 and enhanced transcription termination before the galK gene. *Proc Natl Acad Sci U S A* 112(24):7581-7586.
36. Skordalakes E & Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114(1):135-146.
37. Martin AM, *et al.* (2017) The Diverse Metabolic Roles of Peripheral Serotonin. *Endocrinology* 158(5):1049-1063.
38. Walmacq C, Rahmouni AR, & Boudvillain M (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *J Mol Biol* 342(2):403-420.
39. Ozer A, White BS, Lis JT, & Shalloway D (2013) Density-dependent cooperative non-specific binding in solid-phase SELEX affinity selection. *Nucleic Acids Res* 41(14):7167-7175.
40. Lawson MR, *et al.* (2018) Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol Cell* 71(6):911-922 e914.

41. Helene C, Dimicoli JL, & Brun F (1971) Binding of tryptamine and 5-hydroxytryptamine (serotonin) to nucleic acids. Fluorescence and proton magnetic resonance studies. *Biochemistry* 10(20):3802-3809.
42. Wrist A, Sun W, & Summers RM (2020) The Theophylline Aptamer: 25 Years as an Important Tool in Cellular Engineering Research. *ACS Synth Biol* 9(4):682-697.
43. Simon I, Delaleau M, Schwartz A, & Boudvillain M (2021) A Large Insertion Domain in the Rho Factor From a Low G + C, Gram-negative Bacterium is Critical for RNA Binding and Transcription Termination Activity. *J Mol Biol* 433(15):167060.
44. Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25(6):1203-1210.
45. Goecks J, Nekrutenko A, Taylor J, & Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
46. Alam KK, Chang JL, & Burke DH (2015) FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Molecular therapy. Nucleic acids* 4:e230.
47. Neidhardt FC, Bloch PL, & Smith DF (1974) Culture medium for enterobacteria. *J Bacteriol* 119(3):736-747.
48. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13):3406-3415.
49. Will S, Joshi T, Hofacker IL, Stadler PF, & Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* 18(5):900-914.

FIGURE LEGENDS

Figure 1: Helicase-SELEX (H-SELEX). **(A)** Starting library of RNA:DNA duplex substrates. FWD and REV primers are for RT-PCR amplification only. The BLOCK oligonucleotide is only incorporated in RNA:DNA hybrids derived from genomic fragment libraries. **(B)** Diagram depicting the principle of H-SELEX. **(C)** Reaction time-courses for Rho-dependent unwinding of selected RNA:DNA duplex libraries ($NusG^-$ conditions only in left graph). Helicase reactions were performed at 37°C with 5 nM ^{32}P-labeled duplexes, 20 nM Rho, and 1 mM Mg-ATP in helicase buffer 1. Each reaction was performed only once to limit library depletion except for the R_{10}^- library, which was also analyzed after immobilization of a small aliquot (~0.2 pmole) on streptavidin-coated beads (open gray circles).

Figure 2: Genomic H-SELEX detects putative targets for the Rho helicase within the *E. coli* genome. **(A)** Richness in *Rut* peaks in prophages as compared to the whole genome ($NusG^-$ condition). **(B, C)** Close-up views on genomic regions that are rich and poor in *Rut* peaks, respectively (different scales along x-axis). R_{10}/R_0 enrichment profiles (log_{10} scale) illustrate the presence of *Rut* peaks along both genome strands (~1 peak per 3000 nt on average). Values in the 0-2 range are in dark blue. The *Rut* peaks (cutoff: median $Log_{10}FE \geq 0.3$) collectively represent 3.7 and 4.4% of the genomic sequence (both strands) for $NusG^-$ and $NusG^+$ conditions, respectively.

Figure 3: *Rut* peaks within the *E. coli* genome. **(A)** Number of *Rut* peaks detected in ± NusG conditions ($Log_{10}FE \geq 0.3$ for primary condition and $Log_{10}FE > 0$ for overlapping *Rut* peak in other condition). **(B-D)** Close-up views on *rho*, *trpABCE*, and *thiMD* regions (different scales along x-axis) illustrate the presence of discrete *Rut* peaks at expected locations. Gold rectangles locate peaks with median $Log_{10}FE \geq 0.3$. Cyan arrows locate *Rut* sites of known Rho-dependent

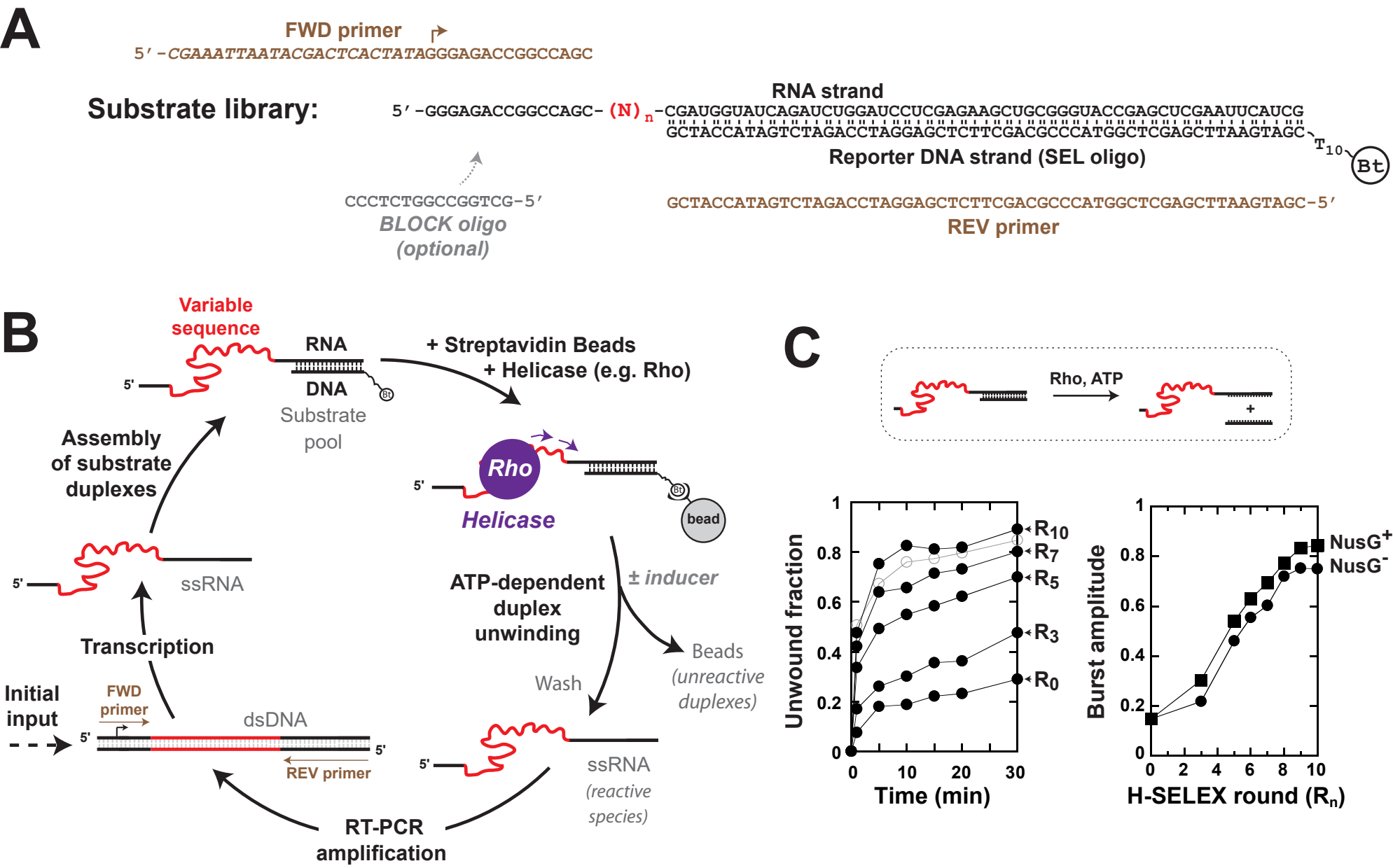
terminators. Pink arrow in panel D locates the binding site of sRNA RydC at the start of *trpE* CDS (29). **(E, F)** Classes of *Rut* peaks as function of genomic features (diagrams not to scale).

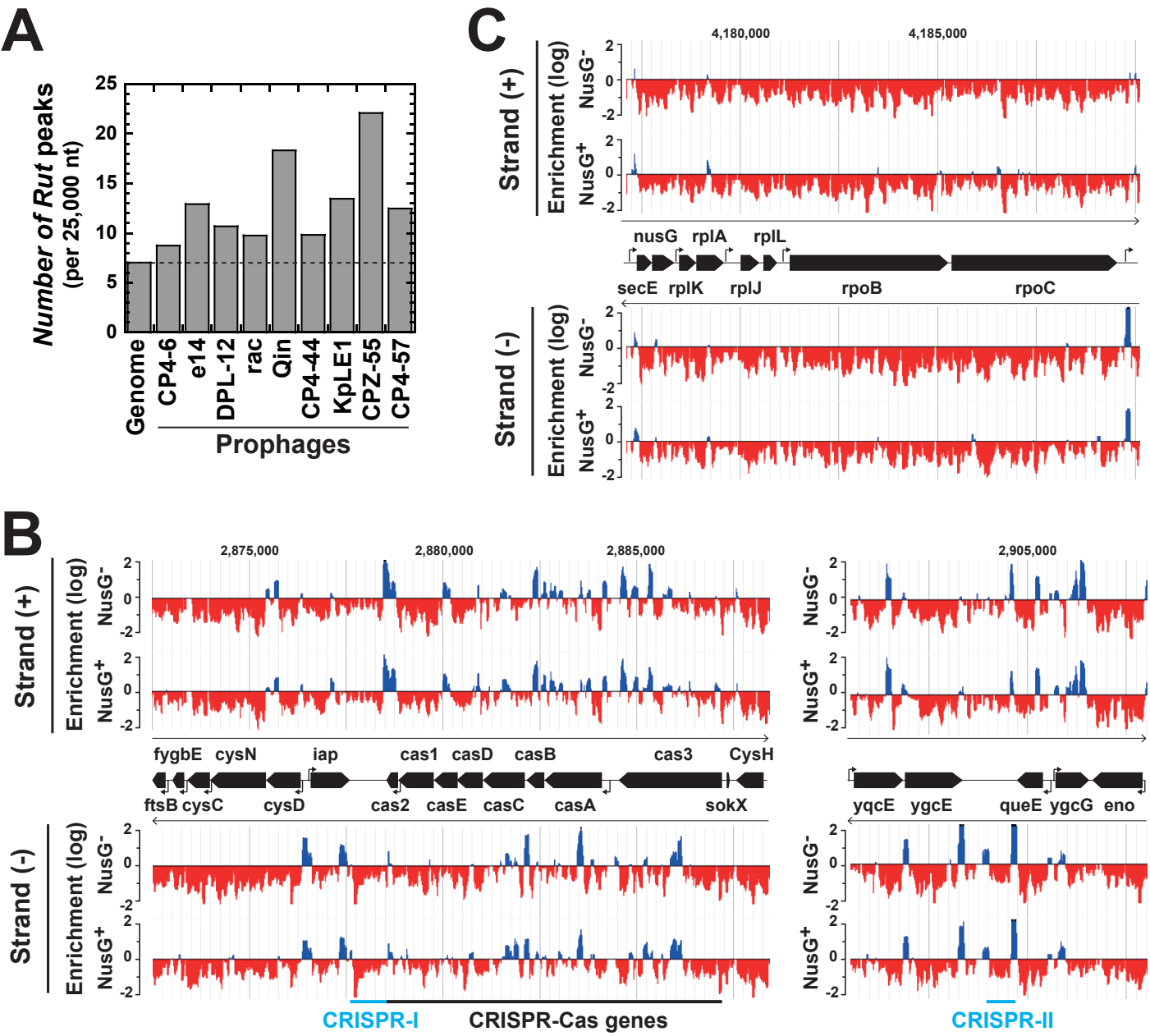
Figure 4: *Rut* peaks have distinct sequence compositions. **(A)** Parameter frequency profiles for the 'overlap' *Rut* peaks obtained in $NusG^-$ (blue curves) and $NusG^+$ (green curves) conditions. The control pool (magenta curves) contains 3000 sequences (130 nt-long) picked randomly in *Rut*-less genome regions. ΔG is the free energy of secondary structure formation computed with Mfold (48) and normalized to 1 kB. **(B)** Representative $NusG^+$-specific *Rut* peaks with G>C biases induce *in vitro* Rho-dependent termination in presence of NusG (green arrows). C%-G% contents of the tested *Rut* peaks (cyan arrows) are indicated below gels. **(C)** *Rut* peak categories display distinct features.

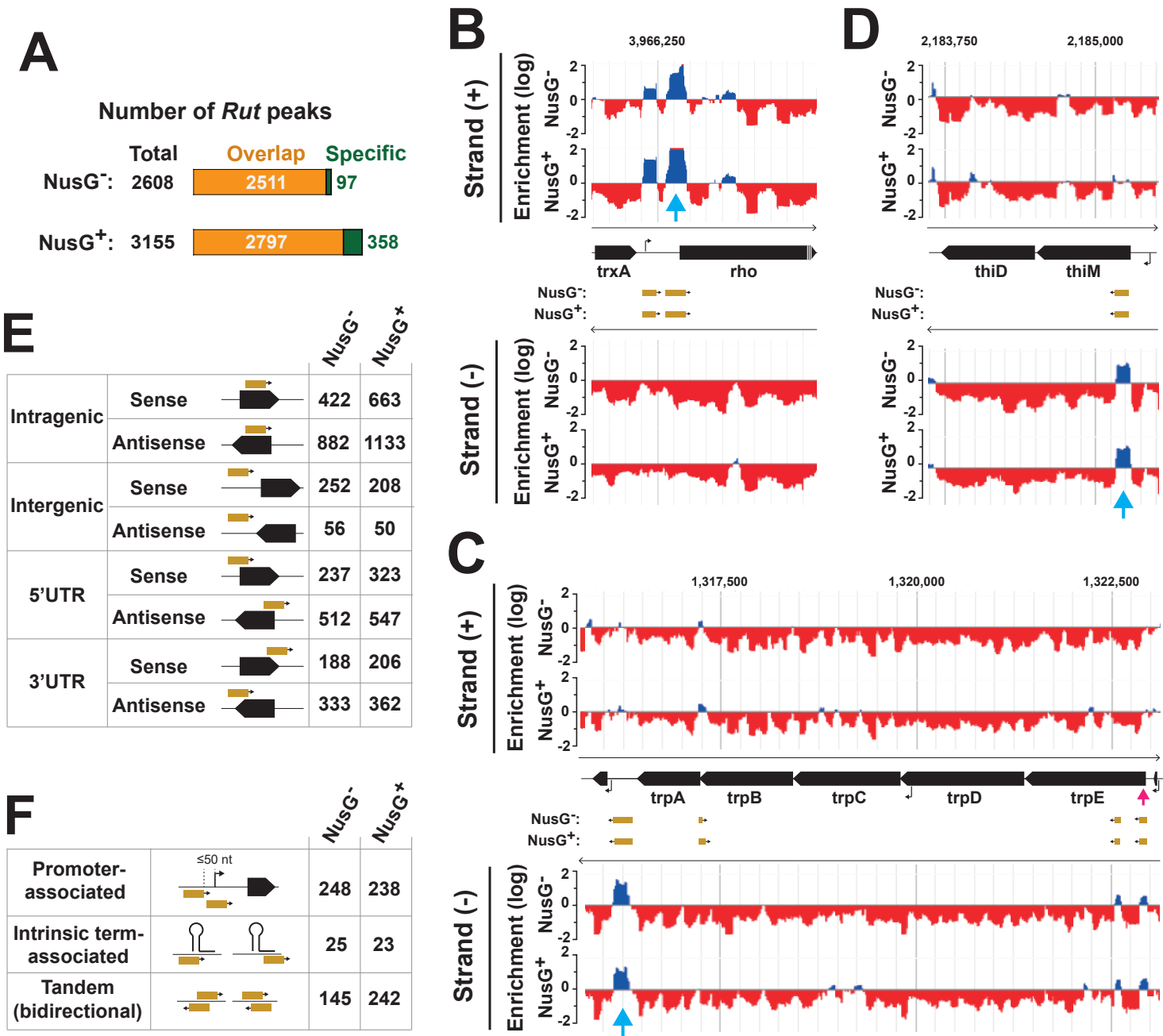
Figure 5: Synthetic H-SELEX yield riboswitches governing the activity of the Rho helicase as function of an inducer (± 10 mM 5-HT). **(A)** Reactivity of the early $^sR_3$ library is affected by bead tethering of the RNA:DNA duplexes. **(B)** Evolution of the reactivity of the libraries with/without 5-HT as function of H-SELEX rounds. F_{30} is the fraction of duplexes unwound after 30 min. The activation ratio is the ratio of F_{30} values determined in presence vs absence of 5-HT. **(C)** Representative PAGE gels and corresponding graph illustrate the differential reactivity of the intermediate $^sR_{13}$ library in presence vs absence of 5-HT. Each reaction was performed only once to limit libraries depletion.

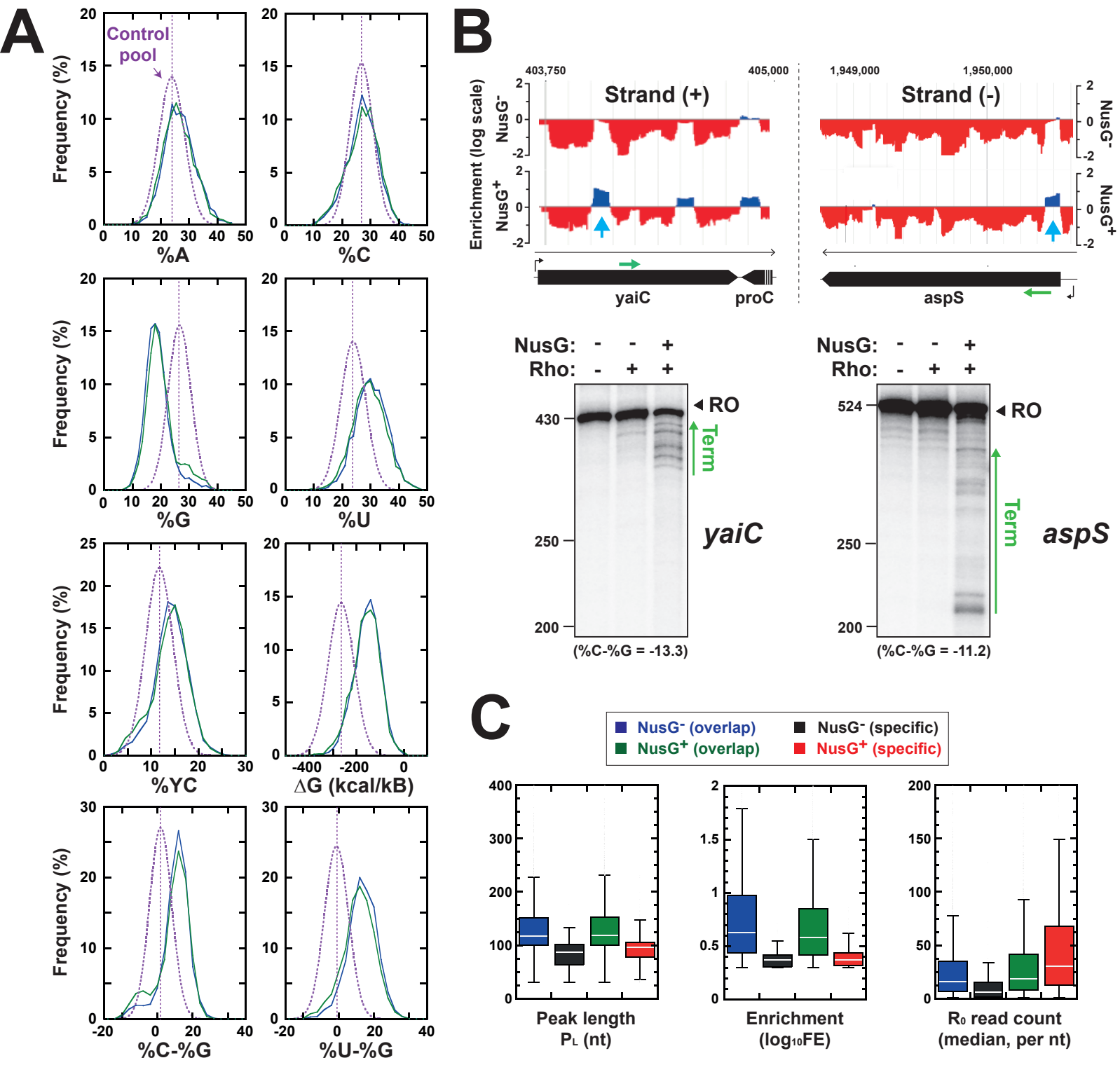
Figure 6: Top-ranking riboswitches from final $^sR_{21}$ library. **(A)** The 21a and 21b sequences can be tentatively folded into similar secondary structures using software such as LocARNA (49). Upstream constant nucleotides (**Fig. 1A**) are in blue. Nucleotides present at same positions in both structures are in red. **(B)** Rho-mediated duplex unwinding kinetics (±10 mM 5-HT) with

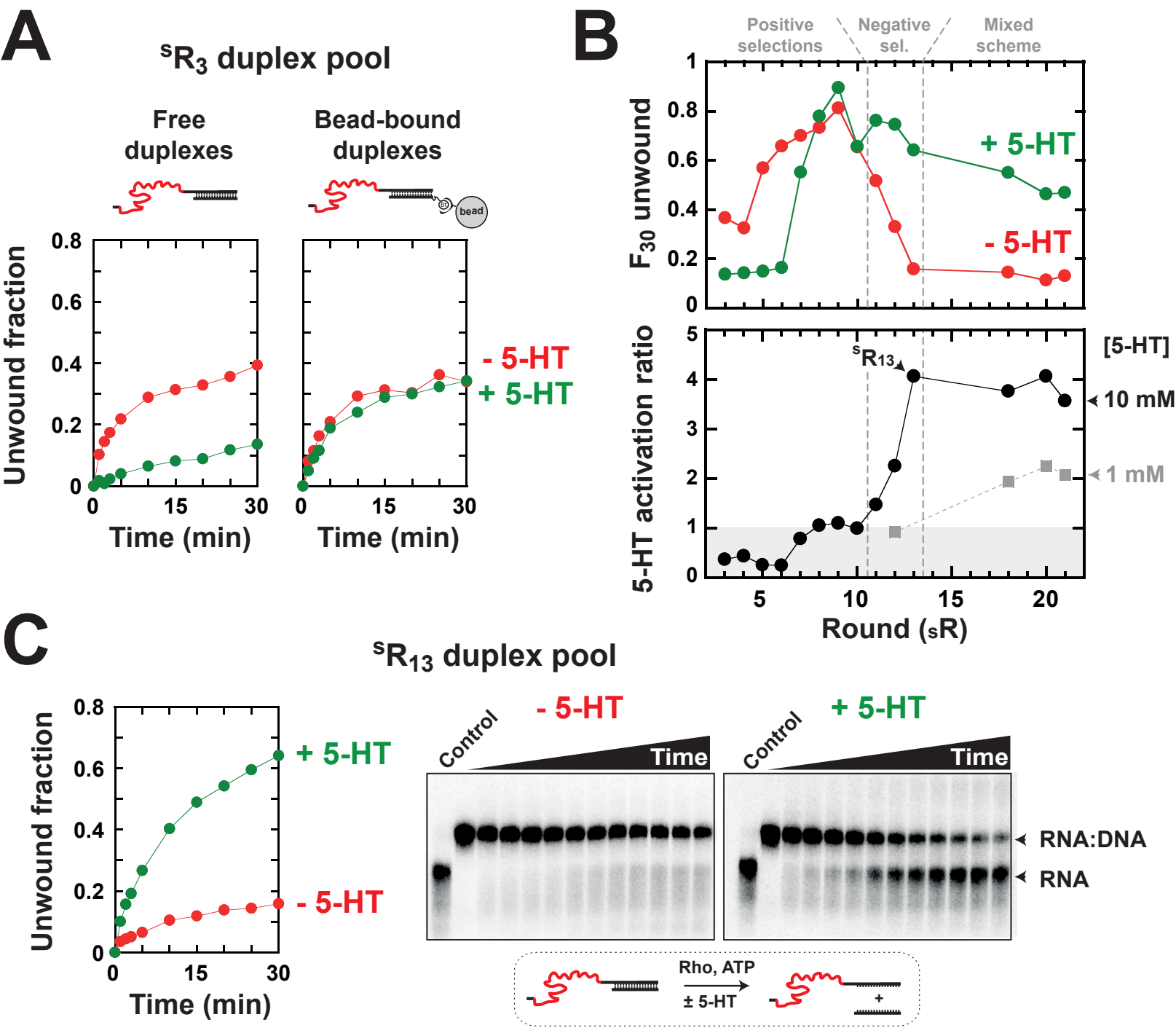
the 21a and 21b riboswitches and *aRut* and *iRut* control duplexes (mean ± SD from 2-3 independent experiments). Unwinding of the *aRut* duplex (+5-HT) is used as reference (dotted lines). **(C)** The 21a and 21b riboswitches can regulate the activity of a downstream reporter *in vivo* as function of ± 10 mM 5-HT. Histograms show per-cell sfGFP/dsRED-Express2 ratios measured by flow cytometry for *E. coli* cells transformed with relevant reporter plasmids (tested sequence [seq] indicated above each graph). See also **Fig. S17**.

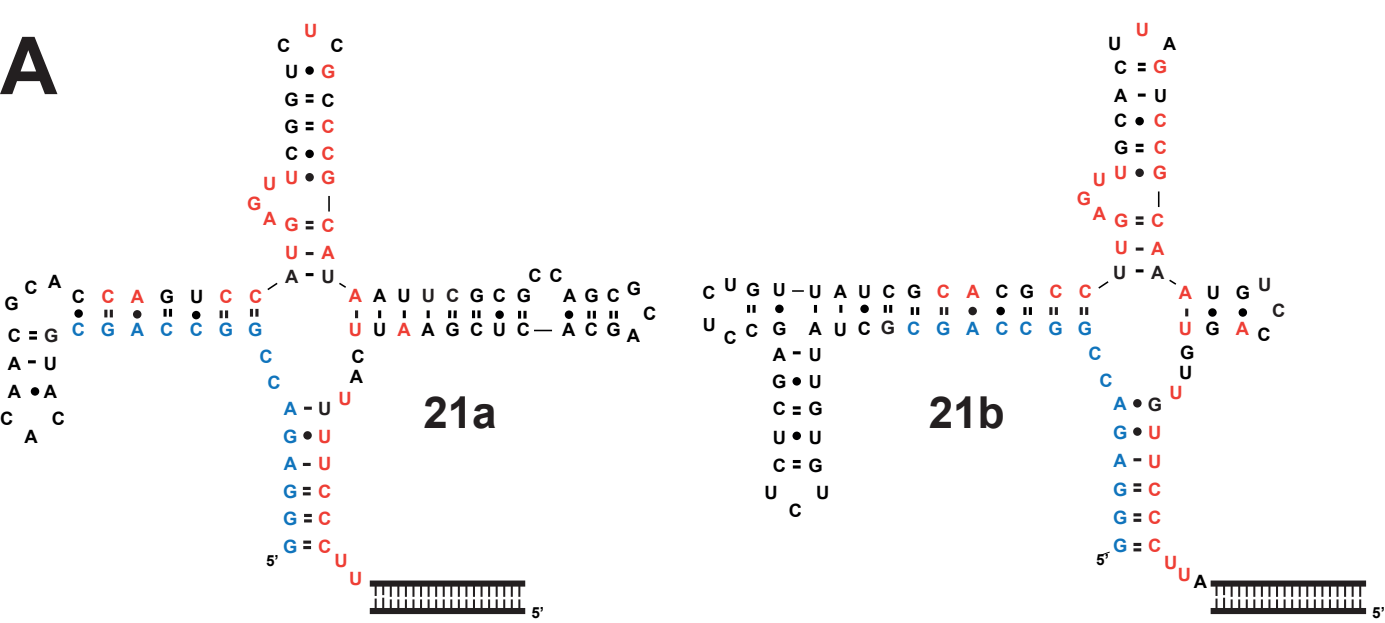

A
FWD primer
5′-CGAAATTAATACGACTCACTATAGGGAGACCGGCCAGC
Substrate library:
RNA strand
5′-GGGAGACCGGCCAGC-(N)n-CGAUGGUAUCAGAUCUGGAUCCUCGAGAAGCUGCGGGUACCGAGCUCGAAUUCAUCG
GCTACCATAGTCTAGACCTAGGAGCTCTTCGACGCCCATGGCTCGAGCTTAAGTAGC-T10-Bt
Reporter DNA strand (SEL oligo)
CCCTCTGGCCGGTCG-5′
BLOCK oligo
(optional)
GCTACCATAGTCTAGACCTAGGAGCTCTTCGACGCCCATGGCTCGAGCTTAAGTAGC-5′
REV primer
B
Variable sequence
RNA
DNA
Substrate pool
+ Streptavidin Beads
+ Helicase (e.g. Rho)
Rho
Helicase
bead
ATP-dependent duplex unwinding
± inducer
Beads
(unreactive duplexes)
Wash
ssRNA
(reactive species)
RT-PCR amplification
dsDNA
FWD primer
REV primer
Transcription
Assembly of substrate duplexes
ssRNA
Initial input
C
Rho, ATP
Unwound fraction
Time (min)
R10
R7
R5
R3
R0
Burst amplitude
H-SELEX round (Rn)
NusG+
NusG-

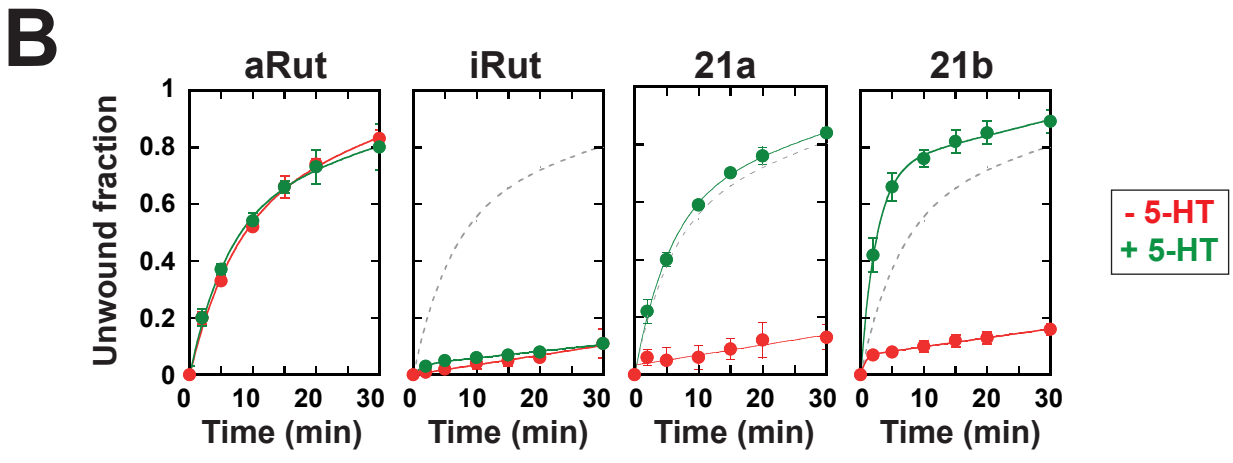

A
Number of Rut peaks
(per 25,000 nt)
25
20
15
10
5
0
Genome
CP4-6
e14
DPL-12
rac
Qin
CP4-44
KpLE1
CPZ-55
CP4-57
Prophages
C
4,180,000
4,185,000
Strand (+)
Enrichment (log)
NusG-
NusG+
nusG
rplA
rplL
secE
rplK
rplJ
rpoB
rpoC
Strand (-)
B
2,875,000
2,880,000
2,885,000
2,905,000
fygbE
cysN
iap
cas1
casD
casB
cas3
CysH
ftsB
cysC
cysD
cas2
casE
casC
casA
sokX
yqcE
ygcE
queE
ygcG
eno
CRISPR-I
CRISPR-Cas genes
CRISPR-II

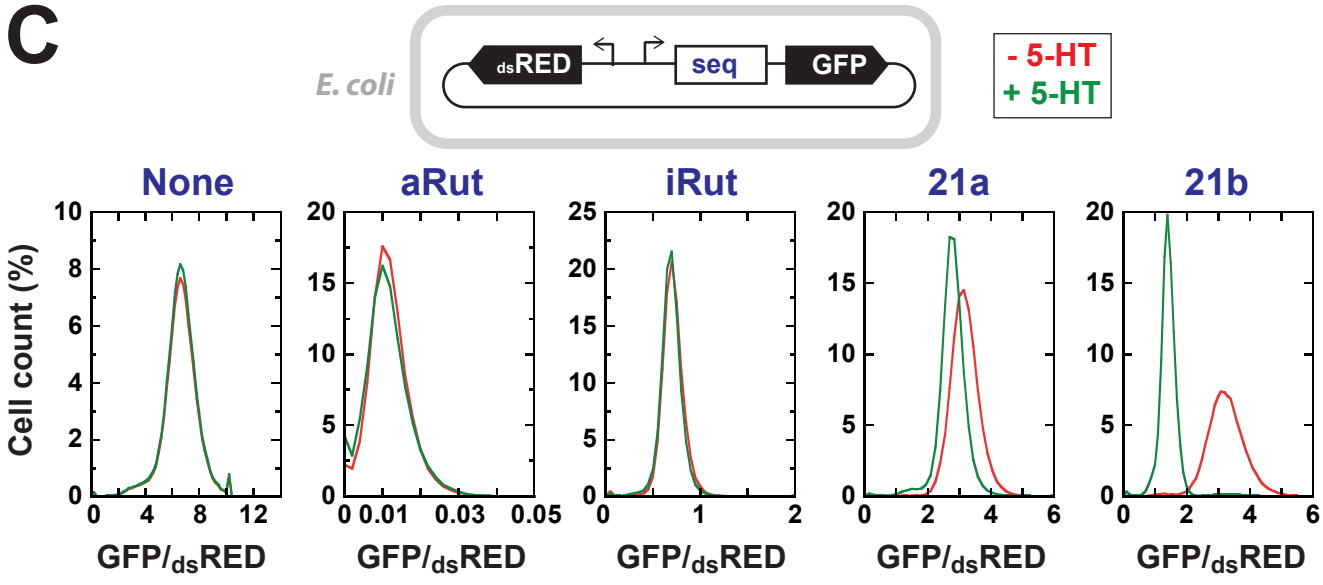

A
Number of Rut peaks
Total
Overlap
Specific
NusG⁻: 2608
2511
97
NusG⁺: 3155
2797
358
B
3,966,250
Strand (+)
Enrichment (log)
NusG⁻
NusG⁺
trxA
rho
NusG⁻:
NusG⁺:
Strand (-)
D
2,183,750
2,185,000
thiD
thiM
C
1,317,500
1,320,000
1,322,500
trpA
trpB
trpC
trpD
trpE
E
NusG⁻
NusG⁺
Intragenic
Sense 422 663
Antisense 882 1133
Intergenic
Sense 252 208
Antisense 56 50
5'UTR
Sense 237 323
Antisense 512 547
3'UTR
Sense 188 206
Antisense 333 362
F
NusG⁻
NusG⁺
Promoter-associated 248 238
≤50 nt
Intrinsic term-associated 25 23
Tandem (bidirectional) 145 242


A
Control pool
Frequency (%)
%A
%C
%G
%U
%YC
ΔG (kcal/kB)
%C-%G
%U-%G
B
Strand (+)
Strand (-)
403,750
405,000
1,949,000
1,950,000
Enrichment (log scale)
NusG-
NusG+
yaiC
proC
aspS
NusG: - - +
Rho: - + +
RO
Term
430
250
200
524
yaiC
aspS
(%C-%G = -13.3)
(%C-%G = -11.2)
C
NusG- (overlap)
NusG+ (overlap)
NusG- (specific)
NusG+ (specific)
Peak length
PL (nt)
Enrichment
(log10FE)
R0 read count
(median, per nt)


A
$^{s}R_3$ duplex pool
Free duplexes
Bead-bound duplexes
bead
Unwound fraction
Time (min)
- 5-HT
+ 5-HT
B
Positive selections
Negative sel.
Mixed scheme
F_{30} unwound
+ 5-HT
- 5-HT
5-HT activation ratio
$^{s}R_{13}$
[5-HT]
10 mM
1 mM
Round (sR)
C
$^{s}R_{13}$ duplex pool
Unwound fraction
Time (min)
+ 5-HT
- 5-HT
Control
Time
RNA:DNA
RNA
Rho, ATP
± 5-HT


A
21a
21b
5'
5'
5'
5'


B
aRut
iRut
21a
21b
Unwound fraction
Time (min)
- 5-HT
+ 5-HT


C
E. coli
dsRED
seq
GFP
- 5-HT
+ 5-HT
None
aRut
iRut
21a
21b
Cell count (%)
GFP/dsRED

A scalable framework for the discovery of functional helicase substrates and synthetic helicase-driven regulatory switches

Mildred Delaleau, Eric Eveno, Isabelle Simon, Annie Schwartz, and Marc Boudvillain

Supplementary information

(Tables S1 to S5; Figure S1 to S17)

SUPPLEMENTARY DISCUSSION

Rut peaks versus Rut sites

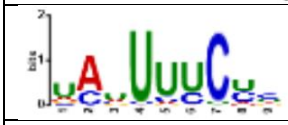
In the present work, we define *Rut* peaks as the single-stranded genomic regions where we detect H-SELEX enrichment ($R_{10}^{\pm}/R_0$) with the MACS2 peak caller (see methods). The sizes of the *Rut* peaks are constrained by the MACS2 settings (e.g. independent peaks are separated by ≥ 30 nt) and by the lengths of the reads in the R_0 and $R_{10}^{\pm}$ libraries:

Library	Length of reads (after trimming of constant regions)			
	mean	SD	shortest	longest
R_0	87	16	30	293
R_{10}^-	95	12	30	258
R_{10}^+	95	11	30	258

Rut peaks should not be confused with actual *Rut* sites, which are the RNA sequences recognized by Rho in the nascent transcripts and are necessary to activate the enzyme (**Fig. S1**). The size of a standard *Rut* site has been estimated to be 60-80 nt from a wealth of biochemical, biophysical, and structural studies (1-7). However, deletion analyses suggest that *Rut* sites can contain dispersed and/or redundant segments (8, 9) and that a minimal *Rut* segment, functional *in vitro* and *in vivo*, may be as short as 30-35 nt (10, 11), which is consistent with our detection of *Rut* peaks in this size range (**Dataset S1**). Assuming that ~10 nt are required to span two PBS clefts and an additional ~10 nt are necessary to reach and fill the SBS (**Fig. S1A**) (1, 12), this means that binding to only three out of the six PBS clefts might be sufficient to activate Rho (while binding to the full PBS+SBS would require at least ~60 nt). It follows that a single *Rut* peak may contain several *Rut* sites, while a single *Rut* site may contain one or several minimal *Rut* segments together with auxiliary elements (9).

Close inspection of read coverages for *Rut* peaks of average length (~130 nt; **Fig. S3C**) suggests that the sizes of the actual (embedded) *Rut* sites may be in the ~60 nt range (**Fig. S2A, B**), although we cannot exclude that shorter, minimal *Rut* segments exist. Similar inspection of longer *Rut* peaks support the existence of several *Rut* sites (**Fig. S2C**) within the same *Rut* peak. The existence of clusters of *Rut* sites where *Rut* density/redundancy along the transcribed strand will favor Rho-dependent termination is consistent with previous observations/models (8, 9, 13, 14).

Rut peaks being larger than actual *Rut* sites may complicate detection of consensus features. Yet, we did not find strongly conserved motifs but only degenerated YC-rich motifs in the *Rut* peak datasets –consistent with the structure-based Rho model of **Fig. S1A**- using software tools such as STREME (https://meme-suite.org/meme/), e.g.:

Motif logo[a]	P-value	E-value	number of occurrences in NusG⁻ overlap dataset
	4.9×10^{-20}	6.8×10^{-19}	1300
	1.9×10^{-6}	2.6×10^{-5}	672

[a] upon performing STREME with the NusG⁻ overlap dataset and the *Rut*-less sequence pool as control.

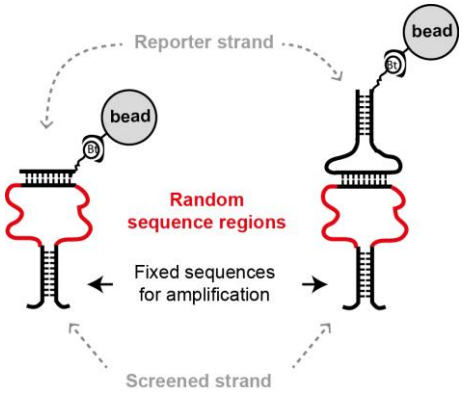
H-SELEX with other helicases

Although H-SELEX has not yet been performed with other helicases, knowledge gained with *E. coli* Rho suggests that the method could be successful and useful for other specimens. The easiest cases to start with are probably helicases that do not require substantial modifications of the current H-SELEX scheme, i.e. other 5'→3' translocases. This obviously includes Rho factors from phylodivergent bacteria but also helicases from superfamily 1 (SF1) such as coronavirus helicases (15) or eukaryotic Sen1. The latter triggers termination of non-coding transcription in a Rho-like manner (16, 17). Full-length Sen1 can be purified and is active in unwinding assays *in vitro* (18). Although it is believed that the intrinsic specificity of Sen1 is low and its target recognition mediated by cofactors Nrd1 and Nab3, this has not been tested extensively; H-SELEX might be a good way to do so. Since Sen1 seems to have a lower duplex unwinding processivity (<40 bp) (18) than *E. coli* Rho (~80 bp) (4), the length of the reporter duplex (**Fig. 1A**) may have to be shortened for efficient selection of H-SELEX libraries. A potentially higher rate of false positives should in turn be expected due to the higher risk of spontaneous dissociation of the shorter RNA:DNA duplex. This may be monitored with control helicase reactions with ADP instead of ATP (or maybe even by attempting a control H-SELEX experiment with ADP) and/or by performing and analyzing independent H-SELEX replicates. Such controls should also be useful for H-SELEX analyses of 'distributive' helicases such as DEAD-box specimens (19-21) (see below) but are labor intensive and would benefit from automation of H-SELEX.

Other SF1, Upf1-like helicases (Upf1 included) are highly processive 5'→3' translocases with likely multiple (yet poorly defined) physiological targets (21-23) and are, in principle, ideal candidates for H-SELEX. However, some specimens contain self-inhibitory domains and require specific cofactors for activation, and possibly for specific target recognition (24, 25). Notwithstanding, it might be useful to compare H-SELEX maps obtained with full-length enzymes (when active) and truncated versions in presence/absence of relevant cofactors. One potential advantage of active, truncated helicase domains such as the one of Upf1 (24, 25) is that their minimal recognition site are smaller than that of Rho, which could prove useful to seek shorter conditional NA substrates from synthetic libraries (e.g. starting with a N_{30} instead of N_{80} variable sequence). Since some Upf1-like helicases also translocate very efficiently on DNA (18, 22), these properties may be used to evolve synthetic DNA biosensors (e.g. for testing the presence of analytes in harsh media/samples unsuitable for RNA) by H-SELEX. In this case, the H-SELEX scheme will have to be adapted so that libraries of DNA duplexes can be prepared and tested iteratively. Methods to prepare and amplify DNA instead of RNA have been developed for standard SELEX (26) and should not be too hard to transpose to H-SELEX. A DNA version of H-SELEX might also prove useful to study 'true' DNA helicases. Although target accessibility/specificity is governed by the cellular DNA machinery (e.g. replicative or DNA repair complexes) and/or accessory factors that create the ssDNA loading platforms for the DNA helicases (or help helicases assemble onto their substrates), sequences particularly favorable to helicase action might still be uncovered by H-SELEX.

Converting H-SELEX for the study of helicases using a 3'→5' translocation mechanism, such as DExH-box SF2 helicases (27), should be straightforward since it only requires an inversion of substrate polarity:

DExH-box helicases bear auxiliary domains that are thought to contribute to substrate specificity, possibly via the recruitment of cofactors (27). However, information on these key features remains limited in most cases and H-SELEX could prove useful here. Finally, either type of substrate design above (5'→3' or 3'→5') may be used to investigate the substrate preference of DEAD-box SF2 helicases, as these enzymes rely on a distributive, non-directional mechanism but are stimulated by single-stranded RNA overhangs on either side of the duplex region (19-21). However, the duplex should be sufficiently short to permit unwinding in a single enzymatic round and probably made of RNA only, as (at least some) DEAD-box proteins can also contact and recognize the A-form duplex (20, 28). For these reasons, DEAD-box helicases might be the most difficult cases for H-SELEX. Yet, global exploration of their sequence specificity, alone or in presence of cofactors, could prove useful and may include screening of synthetic sequence libraries, as these would facilitate testing of other substrate configurations that may have biological relevance (e.g. kissing loop complexes; (29)), as exemplified here:



In any case, one should keep in mind that a good knowledge of the helicase under consideration (biochemistry, strand separation mechanism, etc.) is necessary to adapt H-SELEX adequately and to limit risks of over-interpretation or biases during selection. Similarly, utilization of highly pure helicase (and cofactor) preparations is important, as the powerful selection and exponential amplification process of SELEX-like approaches also makes them sensitive to the presence of contaminants (26). Hence, carefully chosen experimental conditions with highly pure reactants will help minimize enrichment of unspecific NA sequences and increase the specificity of helicase substrate selection.

Table S1 : Summary of conditions used during 'genomic' H-SELEX (*E. coli* genomic library)[a]

Round	Duplex library size (pmol)	Selection mode	RNA:DNA duplex (nM)	Rho (nM)	Reaction time (s)
R_1	704	beads	180	80	300
R_2	360	beads	180	80	120
R_3	720	beads	180	80	60
R_4	720[b] or 360[c]	beads	180	80	30
R_5	710[b] or 295[c]	beads	180[a] or 150[b]	80	30
R_6	650[b] or 630[c]	beads	170[a] or 160[b]	80	20
R_7	270[b] or 305[c]	beads	110[a] or 140[b]	80	20
R_8	620[b] or 520[c]	beads	160[a] or 130[b]	80	20
R_9	360	beads	180	80	20
R_{10}	360	beads	180	80	20

[a] Parallel selections were performed with 0 or 160 nM NusG (NusG$^-$ or NusG$^+$ condition, respectively). [b] NusG$^-$ condition. [c] NusG$^+$ condition.

Table S2: *Rut* peaks observed for known Rho-dependent terminators

Terminator locus (name)	Location	*Enrichment fold* [d]		*Rut* peak coordinates [c,e]	Length P_L (nt) [c]	Strand	Ref	BST [f]
		-NusG	*+NusG*					
topAI (yjhX)	5'UTR	347	170	4534172-4534339	168	-	(30)	yes
cspG	5'UTR	295	5	1051306-1051452	147	+	(30)	no
rho	5'UTR	148	29	3966310-3966469	160	+	(31)	no
tufB [a]	5'UTR	120	117	4175765-4175947	183	+	(32)	no
ydjL	upstream (intergenic)	33	36	1861452-1861678	227	-	(30)	yes
ytfL	5'UTR-promoter	31	5	4441253-4441359	107	-	(30)	no
trpA (trpt')	3'UTR	20	13	1316117-1316363	247	-	(33)	yes
thiM	intragenic	12	16	2185171-2185285	115	-	(30, 34)	no
yhiI	5'UTR	11	18	3630684-3630786	103	-	(30)	no
mdtJ	Upstream promoter	9	7	1673836-1673968	133	-	(30)	yes
tyrT	3'UTR	7	3	1287524-1287606	83	-	(35)	yes
eptB	5'UTR-promoter	5	5	3710472-3710703	232	-	(30)	no
racR	intragenic	5	5	1419754-1419849	96	-	(36)	no
ilvG	intragenic	4	4	3951960-3952063	104	+	(37)	no
fimE	3'UTR	4	3	4542807-4542961	155	+	(38)	yes
ssrS	3'UTR	3	4	3056151-3056293	143	+	(39)	yes
yhaM	upstream (intergenic)	3	3	3256642-3256766	125	-	(30)	no
thiB	5'UTR	3	2	75440-75536	97	-	(34)	no
crp-yhfK	intragenic	3	<2	3486155-3486221	67	+	(30)	no
cspB	5'UTR	2	3	1641484-1641687	204	-	(30)	no
srkA	5'UTR-promoter	2	3	4042356-4042451	96	+	(30)	no
tfaS	3'UTR	2	3	2471095-2471323	229	+	(40)	yes
corA [a]	5'UTR	2	3	4001232-4001415	184	+	(41)	no
yajO	5'UTR	n.d.	2	438132-438183	52	-	(30)	no
moaA [b]	5'UTR	<2	<2	816941-817025	85	+	(30)	no
argT	5'UTR-promoter	<2	<2	2427753-2428018	266	-	(30)	no
mnmG	3'UTR	<2	<2	3925717-3925770	54	-	(30)	no
rpoS [b]	5'UTR	<2	<2	2867649-2867817	169	-	(42, 43)	no
trmL [b]	5'UTR	<2	<2	3781072-3781153	82	+	(30)	no
pgaA [b]	5'UTR	<2	n.d.	1092318-1092442	125		(44)	no
rpoH	5'UTR	<2	n.d.	3600967-3601020	54	-	(30)	no
tnaA	5'UTR	<2	n.d.	3888525-3888621	96	+	(45)	no
add [b]	5'UTR-intragenic	n.d.	<2	1702219-1702347	129	+	(30)	no
galT	intercistronic	n.d.	<2	789955-790046	92	-	(46)	no
thiC [b]	5'UTR	n.d.	n.d.	n.d.	n.d.	-	(34)	yes
mgtA [a]	5'UTR	n.d.	n.d.	n.d.	n.d.	+	(47)	no
lysC	5'UTR	n.d.	n.d.	n.d.	n.d.	-	(34)	no
ribB	5'UTR	n.d.	n.d.	n.d.	n.d.	-	(34, 47)	no
sugE	5'UTR	n.d.	n.d.	n.d.	n.d.	+	(30)	yes
cfa	5'UTR	n.d.	n.d.	n.d.	n.d.	+	(30)	no
cyaA	5'UTR	n.d.	n.d.	n.d.	n.d.	+	(30)	no
chiP [a]	intragenic	n.d.	n.d.	n.d.	n.d.	+	(48)	yes

[a] evidence for Rho-dependent termination obtained in *Salmonella enterica*

[b] low/discontinuous coverage of corresponding region in R_0 library

[c] values for the longest observed peak among the ±NusG conditions

[d] n.d. stands for 'not detected'

[e] genome id: NC_000913.3

[f] Matching Bicyclomicyn sensitive transcript (BST) detect by transcriptomics in ref. (49), i.e. as in **Fig. S6**.

Table S3 : Summary of conditions used during 'synthetic' H-SELEX

Round	Duplex library size (pmol)	Selection mode	Selection[a]	5-HT inducer (mM)	RNA:DNA duplex (nM)	Rho (nM)	reaction time (s)
$^{s}R_1$	1000	beads	positive	1	200	600	120
$^{s}R_2$	450	beads	positive	10	200	600	30
$^{s}R_3$	450	beads	positive	10	200	600	30
$^{s}R_4$	50	PAGE	positive	10	80	320	900
$^{s}R_5$	50	PAGE	positive	10	80	320	120
$^{s}R_6$	50	PAGE	positive	10	80	320	30
$^{s}R_7$	50	PAGE	positive	10	80	80	30
$^{s}R_8$	50	PAGE	positive	10	80	80	30
$^{s}R_9$	50	PAGE	positive	10	80	80	30
$^{s}R_{10}$	50	PAGE	positive	10	80	80	30
$^{s}R_{11}$	50	PAGE	negative	0	80	80	600
$^{s}R_{12}$	50	PAGE	negative	0	80	80	900
$^{s}R_{13}$	50	PAGE	negative	0	80	80	1200
$^{s}R_{14}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) and 120 (p)
$^{s}R_{15}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) and 60 (p)
$^{s}R_{16}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) and 60 (p)
$^{s}R_{17}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) and 60 (p)
$^{s}R_{18}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) and 60 (p)
$^{s}R_{19}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) then 60 (p)
$^{s}R_{20}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) then 60 (p)
$^{s}R_{21}$	100	beads	mixed	0 (n) and 1 (p)	80	80	1200 (n) then 60 (p)

[a] : Reactive ssRNA strands or unreactive RNA:DNA duplexes were recovered after helicase reactions performed under positive (p, with 5-HT) or negative (n, without 5-HT) selection conditions, respectively. In mixed schemes, an helicase reaction is first performed without 5-HT (n), beads bearing the unreactive duplexes are separated from the supernatant and directly used in an helicase reaction with 5-HT (p), and then ssRNA strands are harvested and processed into the next H-SELEX round.

Table S4 : Oligonucleotides used in this work	
Synthetic library	5'GGGAGACCGGCCAGC(N_{80})CGATGGTATCAGATCTGGATCCTCGAGAAGCTGC
FWD	5'CGAAATTAATACGACTCACTATAGGGAGACCGGCCAGC
REV	5'CGATGAATTCGAGCTCGGTACCCGCAGCTTCTCGAGGATCCAGATCTGATACCATCG
SEL	5'biotin/TTTTTTTTTTCGATGAATTCGAGCTCGGTACCCGCAGCTTCTCGAGGATCCAGATCTGATACCATCG
BLOCK	5'GCTGGCCGGTCTCCC
TRAP	5'CGATGGTATCAGATCTGGATCCTCGAGAAGCTGCGGGTACCGAGCTCGAATTCATCG
IRUT-OLN	5'GGGAGACCGGCCAGCGGTCGAGGAAGAAGGAGACAGAGGAGAAGGAAGAAGGAGACAGAGGAGAAGTCGATGGTATCAGATCTGG
LESS-OLN	5'-GGGAGACCGGCCAGCCGATGGTATCAGATCTGG
FORWA	5'-CGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTGGGAGACCGGCCAGC
FORWB	5'-GCGGCCGCACTCGAGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGT
FACS-REV	5'-ACTGTCTTACACACCGGTAAGACAGCCCAGATCTGATACCATCG
C1-OLN	5'GGGAGACCGGCCAGCCCCATGTATCGTCGAGGGCAGTTCTTGGATCCTCTGTAAGAGATTACGGTTATCTCCGTATGAAACAGTTGTTTACCCTGCGATGGTATCAGATCTGG
R45-ran-Rev	5'AGATCTGATACCATCGNNNNNNNNN
ARN107-ran-For	5'GGGAGACCGGCCAGCNNNNNNNNN

Table S5 : DNA templates used for *in vitro* transcription termination experiments

Name	Pi[a]	Pf[a]	Strand[b]	forward PCR primer (first amplification[c])	Reverse PCR primer
yaiC	403,952	404,355	+	5'GTCTAACCTATAGGATACTTACAGCCAGC CTGGCAGATAGCGAGCAG	5'CGCCCGTCATGC CGTCGCGGG
ydbA	1,465,392	1,465,782	+	5'GTCTAACCTATAGGATACTTACAGCCATG CAAAGGAAAACTCT	5'CCGGATCGTCCG GTATTATC
holE	1,924,973	1,925,490	+	5'GTCTAACCTATAGGATACTTACAGCCATC TATTCCTTTTTGTAATAACTTTTTTACAGAGC	5'TGAGGTTTTACGA GGCTCATATTGCGC
yobA	1,924,868	1,925,270	-	5'GTCTAACCTATAGGATACTTACAGCCAGC TCGCGAAACCAGCTGC	5'ATACTGATGCGTT AAATGCGCATGTGC
aspS	1,950,204	1,950,701	-	5'GTCTAACCTATAGGATACTTACAGCCAGG CCCGATATAAAGTGG	5'CATCTGCGCGGTT GATGATAGTCAGC
yqgA	3,109,253	3,109,680	+	5'GTCTAACCTATAGGATACTTACAGCCAGA CATCAGTAAAAGCAGAAACGCTC	5'AGCGATGCCAGA CCAAAAATTGATGT CATG
speC	3,109,018	3,109,502	-	5'GTCTAACCTATAGGATACTTACAGCCATG GCTTGTTTGTTCGCAAAGTC	5'GCAAGAATGCCA CTGCGACTATCCGC
metR	4,012,314	4,012,635	-	5'GTCTAACCTATAGGATACTTACAGCC AGGGAGAAATCCTGTTGC	5'GCTAACACCAGAC GCACTTC

[a] Genomic coordinates (NC_000913.3) of the MG1655 sequence encoded by the DNA template

[b] Genome strand containing the MG1655 sequence encoded by the DNA template

[c] A second PCR amplification was performed with the listed reverse primer and the universal T7A1 primer (5'TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCC)

SUPPLEMENTARY REFERENCES

1. Skordalakes E & Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114(1):135-146.
2. McSwiggen JA, Bear DG, & von Hippel PH (1988) Interactions of Escherichia coli transcription termination factor rho with RNA. I. Binding stoichiometries and free energies. *J Mol Biol* 199(4):609-622.
3. Koslover DJ, Fazal FM, Mooney RA, Landick R, & Block SM (2012) Binding and translocation of termination factor rho studied at the single-molecule level. *J Mol Biol* 423(5):664-676.
4. Walmacq C, Rahmouni AR, & Boudvillain M (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *J Mol Biol* 342(2):403-420.
5. Morgan WD, Bear DG, Litchman BL, & von Hippel PH (1985) RNA sequence and secondary structure requirements for rho-dependent transcription termination. *Nucleic Acids Res* 13(10):3739-3754.
6. Bear DG*, et al.* (1988) Interactions of Escherichia coli transcription termination factor rho with RNA. II. Electron microscopy and nuclease protection experiments. *J Mol Biol* 199(4):623-635.
7. Hart CM & Roberts JW (1994) Deletion analysis of the lambda tR1 termination region. Effect of sequences near the transcript release sites, and the minimum length of rho-dependent transcripts. *J Mol Biol* 237(3):255-265.
8. Zalatan F, Galloway-Salvo J, & Platt T (1993) Deletion analysis of the Escherichia coli rho-dependent transcription terminator trp t'. *J. Biol. Chem.* 268(23):17051-17056.
9. Ciampi MS (2006) Rho-dependent terminators and transcription termination. *Microbiology* 152(Pt 9):2515-2528.
10. Graham JE & Richardson JP (1998) rut Sites in the nascent transcript mediate Rho-dependent transcription termination in vivo. *J Biol Chem* 273(33):20764-20769.
11. Guerin M, Robichon N, Geiselmann J, & Rahmouni AR (1998) A simple polypyrimidine repeat acts as an artificial Rho-dependent terminator in vivo and in vitro. *Nucleic Acids Res* 26(21):4895-4900.
12. Thomsen ND, Lawson MR, Witkowsky LB, Qu S, & Berger JM (2016) Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. *Proc Natl Acad Sci U S A* 113(48):E7691-E7700.
13. Nadiras C, Eveno E, Schwartz A, Figueroa-Bossi N, & Boudvillain M (2018) A multivariate prediction model for Rho-dependent termination of transcription. *Nucleic Acids Res* 46(16):8245-8260.
14. Alifano P, Rivellini F, Limauro D, Bruni CB, & Carlomagno MS (1991) A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* 64(3):553-563.
15. Spratt AN*, et al.* (2021) Coronavirus helicases: attractive and unique targets of antiviral drug-development and therapeutic patents. *Expert Opin Ther Pat* 31(4):339-350.
16. Porrua O & Libri D (2013) A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* 20(7):884-891.
17. Porrua O, Boudvillain M, & Libri D (2016) Transcription Termination: Variations on Common Themes. *Trends in genetics : TIG* 32(8):508-522.
18. Han Z, Libri D, & Porrua O (2017) Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res* 45(3):1355-1370.
19. Putnam AA & Jankowsky E (2013) DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim Biophys Acta* 1829(8):884-893.
20. Mallam AL, Del Campo M, Gilman B, Sidote DJ, & Lambowitz AM (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490(7418):121-125.

21. Grass LM*, et al.* (2021) Large-scale ratcheting in a bacterial DEAH/RHA-type RNA helicase that modulates antibiotics susceptibility. *Proc Natl Acad Sci U S A* 118(30).
22. Kanaan J*, et al.* (2018) UPF1-like helicase grip on nucleic acids dictates processivity. *Nature communications* 9(1):3752.
23. Kim YK & Maquat LE (2019) UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA* 25(4):407-422.
24. Fiorini F, Boudvillain M, & Le Hir H (2013) Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res* 41(4):2404-2415.
25. Chamieh H, Ballut L, Bonneau F, & Le Hir H (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol* 15(1):85-93.
26. Stoltenburg R, Reinemann C, & Strehlitz B (2007) SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* 24(4):381-403.
27. Ozgur S*, et al.* (2015) The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J* 282(5):850-863.
28. Yang Q, Del Campo M, Lambowitz AM, & Jankowsky E (2007) DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* 28(2):253-263.
29. Sakamoto T, Oguro A, Kawai G, Ohtsu T, & Nakamura Y (2005) NMR structures of double loops of an RNA aptamer against mammalian initiation factor 4A. *Nucleic Acids Res* 33(2):745-754.
30. Adams PP*, et al.* (2021) Regulatory roles of Escherichia coli 5' UTR and ORF-internal RNAs detected by 3' end mapping. *eLife* 10.
31. Matsumoto Y, Shigesada K, Hirano M, & Imai M (1986) Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators. *J Bacteriol* 166(3):945-958.
32. Brandis G, Bergman JM, & Hughes D (2016) Autoregulation of the tufB operon in Salmonella. *Mol Microbiol* 100(6):1004-1016.
33. Wu AM, Christie GE, & Platt T (1981) Tandem termination sites in the tryptophan operon of Escherichia coli. *Proc Natl Acad Sci U S A* 78(5):2913-2917.
34. Bastet L*, et al.* (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Res* 45(12):7474-7486.
35. Kupper H, Sekiya T, Rosenberg M, Egan J, & Landy A (1978) A rho-dependent termination site in the gene coding for tyrosine tRNA su3 of Escherichia coli. *Nature* 272(5652):423-428.
36. Cardinale CJ*, et al.* (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320(5878):935-938.
37. Wek RC, Sameshima JH, & Hatfield GW (1987) Rho-dependent transcriptional polarity in the ilvGMEDA operon of wild-type Escherichia coli K12. *J Biol Chem* 262(31):15256-15261.
38. Hinde P, Deighan P, & Dorman CJ (2005) Characterization of the detachable Rho-dependent transcription terminator of the fimE gene in Escherichia coli K-12. *J Bacteriol* 187(24):8256-8266.
39. Chae H*, et al.* (2011) Rho-dependent termination of ssrS (6S RNA) transcription in Escherichia coli: implication for 3' processing of 6S RNA and expression of downstream ygfA (putative 5-formyl-tetrahydrofolate cyclo-ligase). *J Biol Chem* 286(1):114-122.
40. Menouni R, Champ S, Espinosa L, Boudvillain M, & Ansaldi M (2013) Transcription termination controls prophage maintenance in Escherichia coli genomes. *Proc Natl Acad Sci U S A* 110(35):14414-14419.
41. Kriner MA & Groisman EA (2015) The Bacterial Transcription Termination Factor Rho Coordinates Mg(2+) Homeostasis with Translational Signals. *J Mol Biol* 427(24):3834-3849.
42. Sedlyarova N*, et al.* (2016) sRNA-Mediated Control of Transcription Termination in E. coli. *Cell* 167(1):111-121 e113.

43. Nadiras C, Schwartz A, Delaleau M, & Boudvillain M (2018) Evaluating the Effect of Small RNAs and Associated Chaperones on Rho-Dependent Termination of Transcription In Vitro. *Methods Mol Biol* 1737:99-118.
44. Figueroa-Bossi N*, et al.* (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28(11):1239-1251.
45. Konan KV & Yanofsky C (2000) Rho-dependent transcription termination in the tna operon of Escherichia coli: roles of the boxA sequence and the rut site. *J Bacteriol* 182(14):3981-3988.
46. Wang X, Ji SC, Jeon HJ, Lee Y, & Lim HM (2015) Two-level inhibition of galK expression by Spot 42: Degradation of mRNA mK2 and enhanced transcription termination before the galK gene. *Proc Natl Acad Sci U S A* 112(24):7581-7586.
47. Hollands K*, et al.* (2012) Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A* 109(14):5376-5381.
48. Bossi L, Schwartz A, Guillemardet B, Boudvillain M, & Figueroa-Bossi N (2012) A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* 26(16):1864-1873.
49. Peters JM*, et al.* (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev* 26(23):2621-2633.

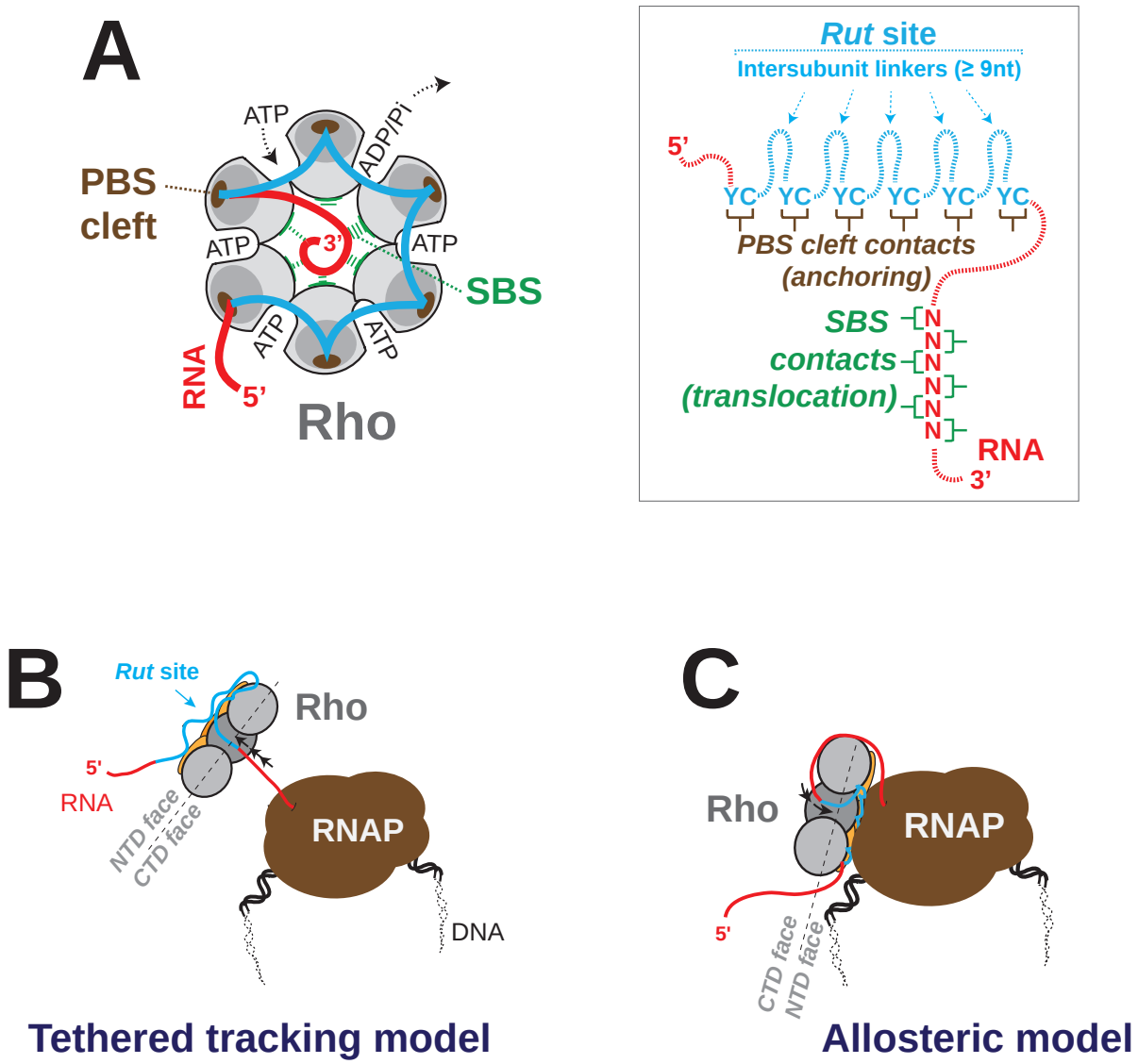


Figure S1: (A) Diagram showing the organization of the hexameric Rho helicase and its interaction with RNA. *Rut* sites are recognized by the composite Primary Binding Site (PBS) of Rho, which includes one 5'YC binding cleft per monomer and crowns the Rho hexamer (Skordalakes, Cell 2003, PMID: 12859904). RNA also binds the composite Secondary Binding Site (SBS) within Rho central channel. The SBS threads RNA through the Rho ring in a sequential, ATP-dependent manner (Thomsen, PNAS 2016; PMID: 27856760). **(B)** The classical 'tethered tracking' model of Rho-dependent transcription termination (Steinmetz, PNAS 1994, PMID: 7509071) stipulates that Rho binds nascent transcripts at ribosome-free *Rut* sites and then translocates along RNA to catch up with RNA polymerase (RNAP). In the process, RNA-PBS contacts are maintained while SBS-RNA contacts evolve in an ATP-dependent manner. Once in contact with RNAP, Rho disrupts the transcription complex through RNA:DNA hybrid shearing or by pushing RNAP forward. **(C)** The alternative allosteric model proposes that Rho first binds RNAP during transcription of non-coding DNA regions and then scans the nascent transcript in search of a *Rut* site. Recognition of a *Rut* site activates Rho, which then destabilizes the transcription complex by an allosteric mechanism (reviewed in Hao, Transcription 2021, PMID: 34705601). The two models are not necessarily mutually exclusive and hybrid versions are possible (Song, Nat Commun 2022, PMID: 35351884). The implication of the Rho helicase in the clearance of R-loops, i.e. transcriptional RNA:DNA duplexes (Leela, PNAS 2013, PMID: 23251031), is not depicted here.

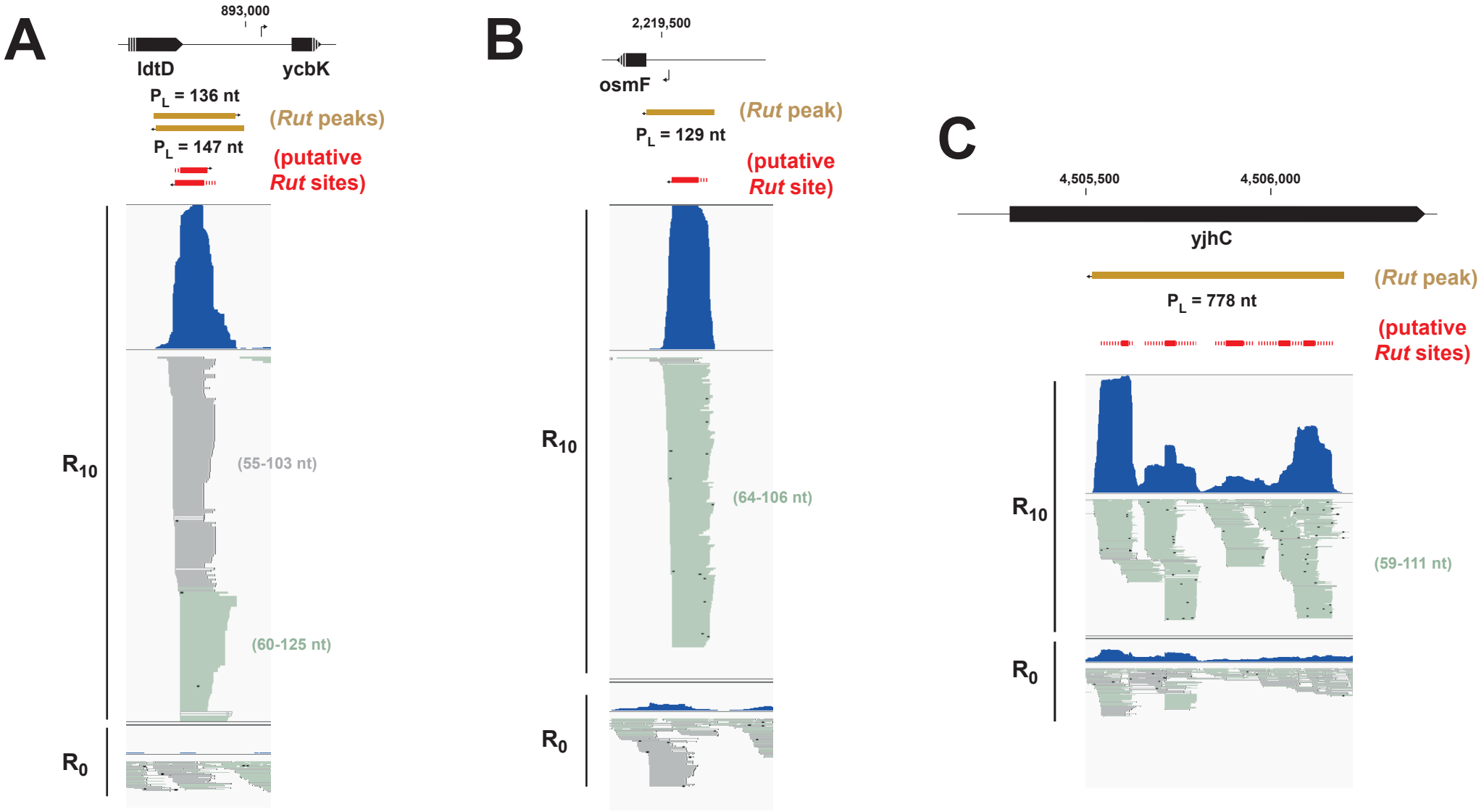


Figure S2: Integrative Genomics Viewer (IGV) snapshots comparing aligned R_0 and R_{10} reads for representative *Rut* peaks detected by H-SELEX (NusG$^-$ condition). Forward and reverse strand reads are shown in gray and green, respectively. Default autoscaled IGV coverage tracks are shown per library (in blue) and should not be confused with the JBrowse-depicted, per-strand R_{10}/R_0 enrichment profiles shown in other figures (e.g. Fig. 2B). Gold rectangles locate the *Rut* peaks detected with the MACS2 peak caller (median $\log_{10}$FE ≥ 0.3). Identification of individual putative *Rut* sites (red bars; see Fig. S1) within the *Rut* peaks is tentative as it was perfomed by visual inspection of the R_{10} read stacks and cannot be easily automated (due to the dependence on the matching R_0 reads and necessary normalization of the library depths).

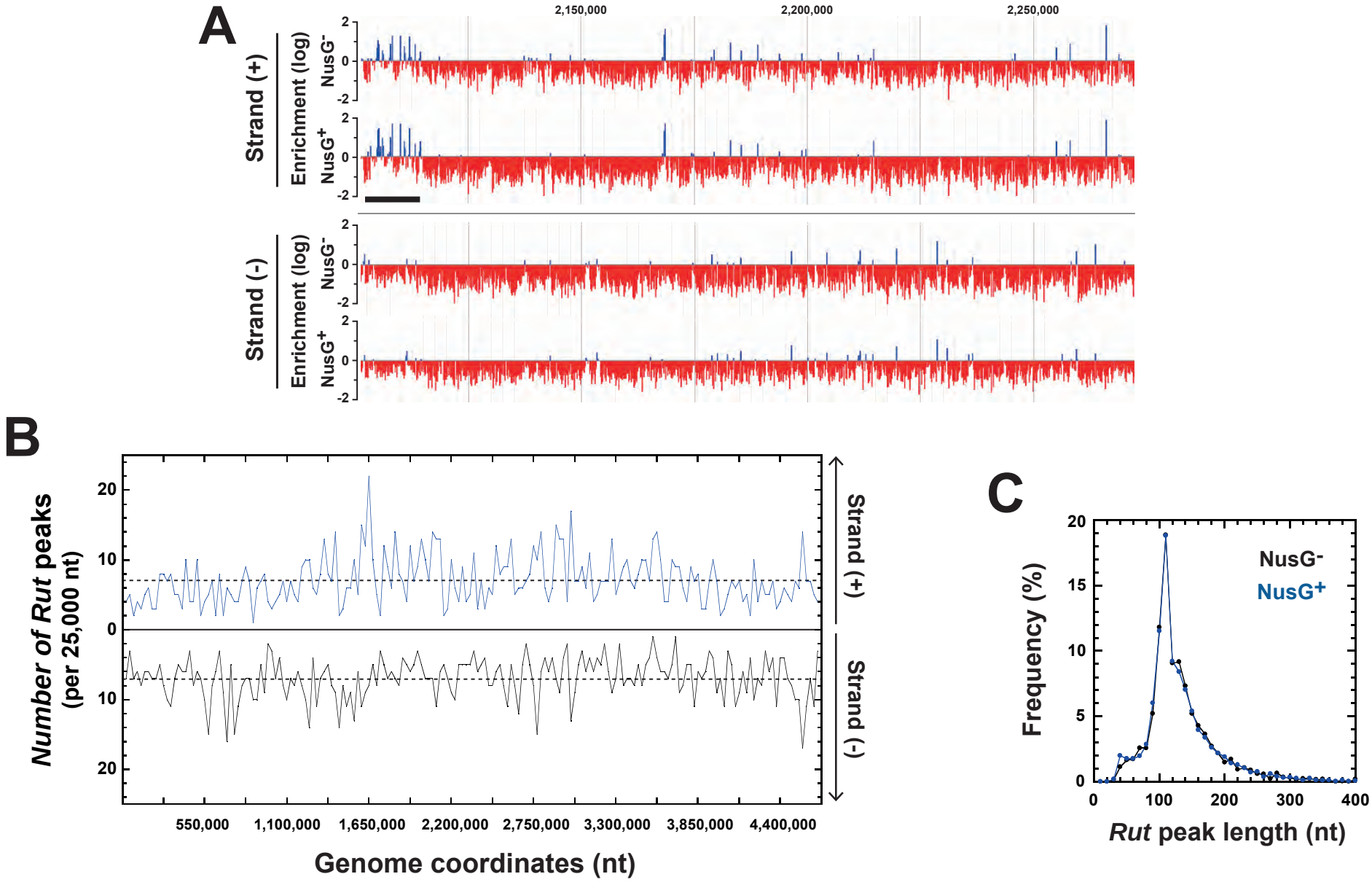


Figure S3: (A) H-SELEX R_{10}/R_0 enrichment profiles ($\log_{10}$ scale) for a representative genomic region with *Rut* peaks appearing in blue. The black bar locates a cluster of *Rut* sites such as the ones found in prophage and CRISPR regions (see Fig. 2A, B). **(B)** Frequency of the *Rut* peaks (cutoff: median $\log_{10}$FE ≥ 0.3) along the MG1655 genome for the NusG$^-$ condition (top and bottom strands in blue and black, respectively). Dotted lines indicate average frequency (bin size: 25,000 nt). **(C)** Distribution of the lengths of *Rut* peaks (median $\log_{10}$FE ≥ 0.3) detected upon processing sequence reads from the R_{10}^- (NusG$^-$) and R_{10}^+ (NusG$^+$) enrichment pools. The average peak length (±SD) is (130 ± 59) nt for R_{10}^- and (128 ± 58) nt for R_{10}^+. By comparison, the average read length (±SD) was (87 ± 16) nt for the R_0 library, (95 ± 12) nt for the R_{10}^- library and (95 ± 11) nt for the R_{10}^+ library.

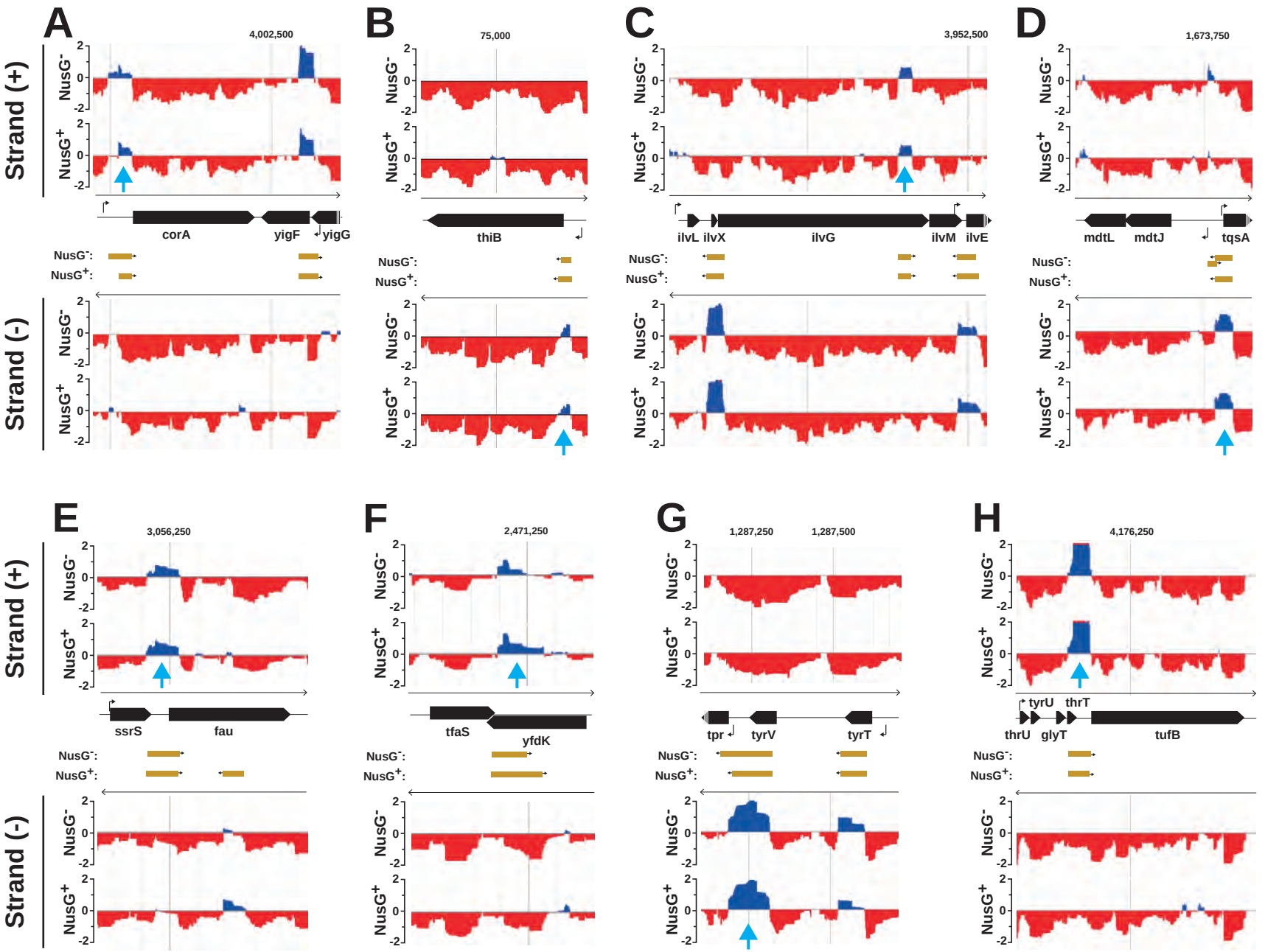


Figure S4: Examples of H-SELEX 'fold enrichment' (FE) profiles ($\log_{10}$ scale) observed for regions containing known Rho-dependent terminators. Gold rectangles locate *Rut* peaks with median $\log_{10}$FE ≥ 0.3 while cyan arrows mark the peaks best matching the known terminators. Although both the *ilv* (panel C) and *trp* (Fig. 3C) operons are regulated by transcriptional attenuation, only *trp* contains *Rut* peaks suggestive of participation of premature Rho-dependent termination. Note that different scales along the genome coordinates (x axis) are used for the various panels.

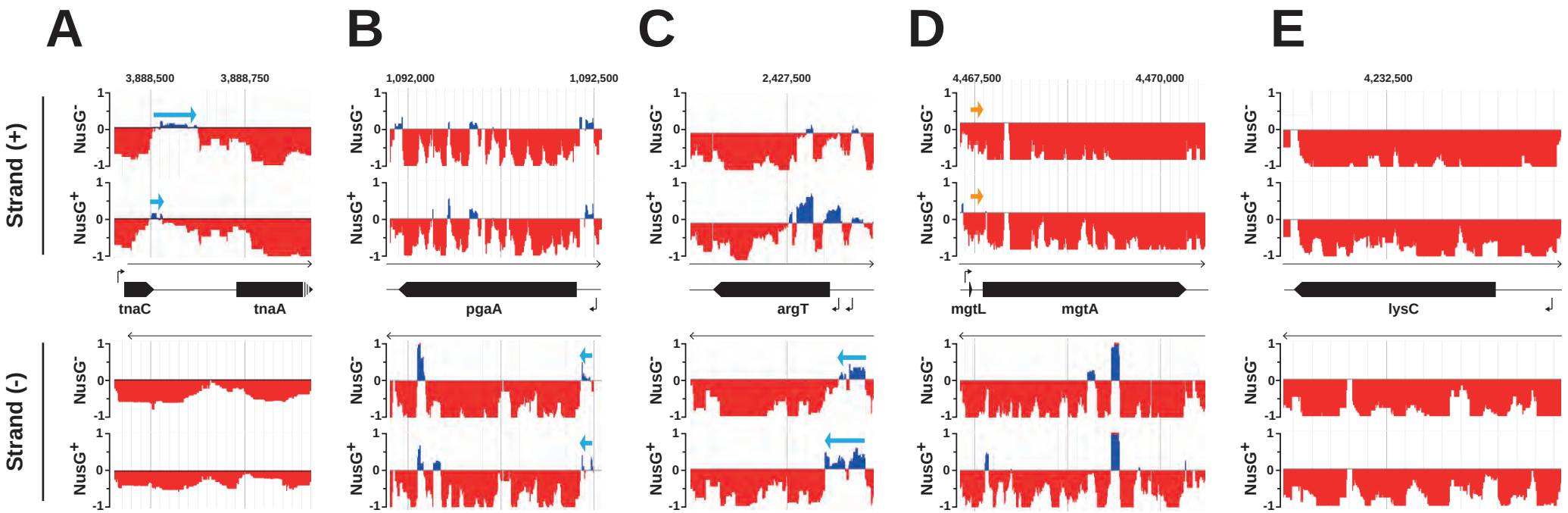


Figure S5: Representative H-SELEX fold enrichment profiles (log_{10} scale) for known Rho-dependent terminators where *Rut* peaks are below the detection threshold (A-C) or are not observed (D, E). Light blue arrows in pannels A-C identify the *Rut* peaks below the detection threshold (median $log_{10}FE < 0.3$). In pannel D, orange arrows mark the region in the 5'leader of *mgtA* where a *Rut* peak was expected. Location of the Rho-dependent signal in the *LysC* cistron (E) is not known. Note that different scales along the genome coordinates (x axis) are used for the various panels.

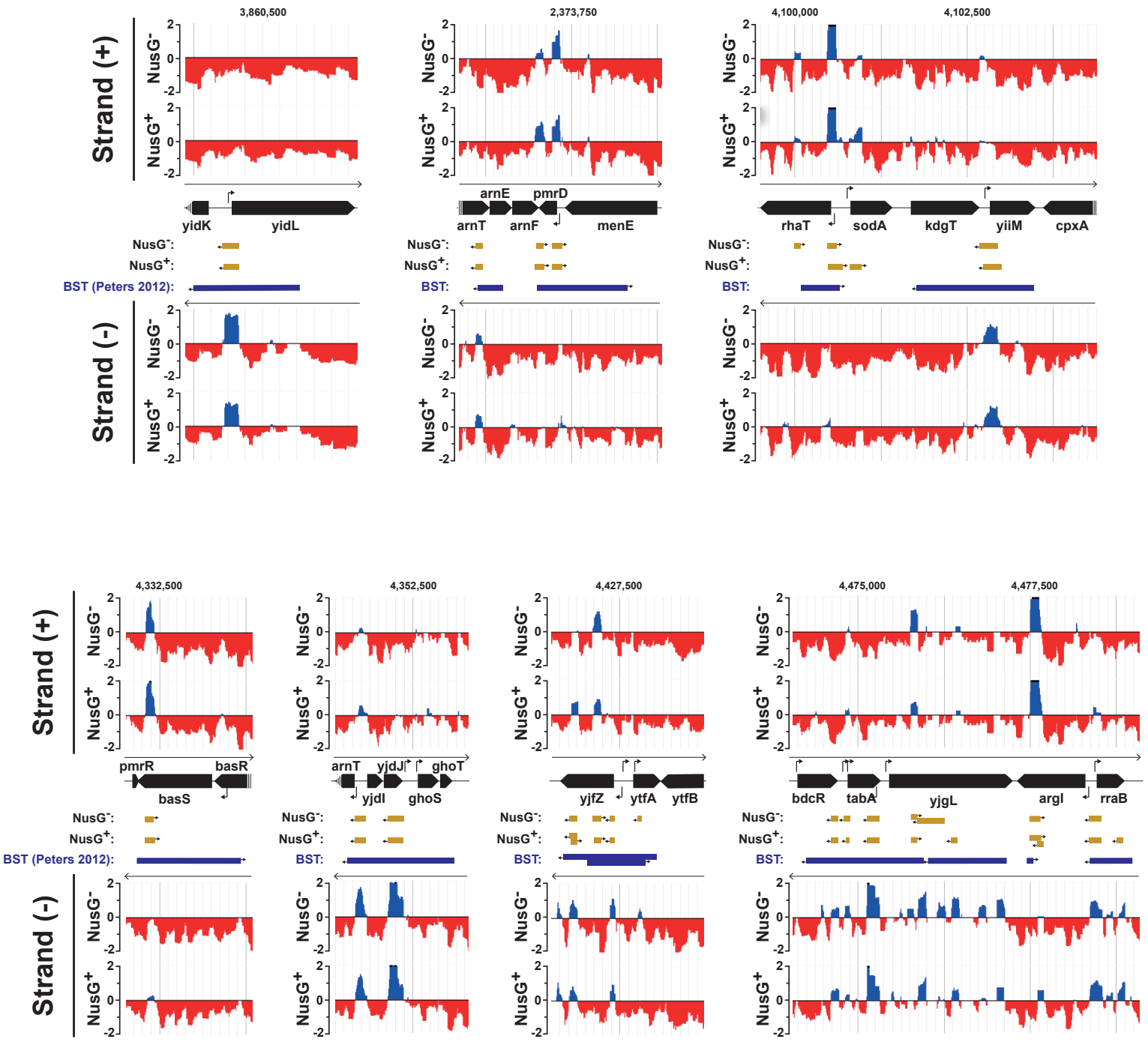


Figure S6: Examples of overlaping matches between H-SELEX *Rut* peaks (gold arrows) and Bicyclomycin Sensitive Transcript (BST) regions (blue arrows) detected by transcriptomics (Peters, Gene Dev. 2012, PMID: 23207917) provide further evidence of the relevance of the 'genomic' H-SELEX methodology. In most cases, the BST starts before the *Rut* peak, consistent with posttranscriptional 3'→5' trimming of the Rho-dependent transcripts (even beyond the actual *Rut* site) by exonucleases *in vivo*, and with the superiority of H-SELEX to map Rho targets precisely. In some instances (e.g. bottom right diagrams), several *Rut* peaks overlap with a single BST, so that it more difficult to predict which *Rut* peak(s) is (are) at the origin of the BST signal.

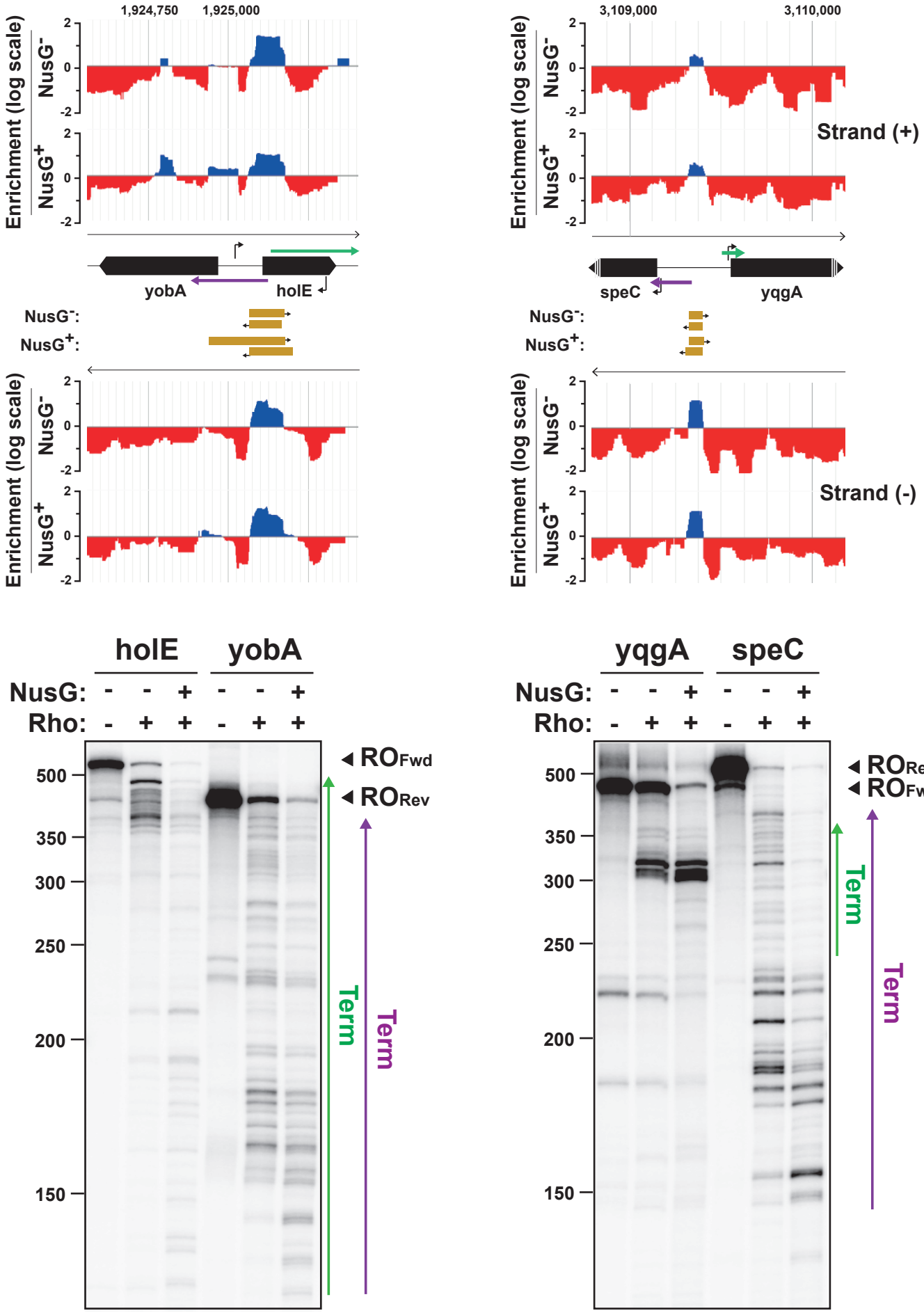


Figure S7: Representative examples of divergent *Rut* peak tandems (gold rectangles for peaks above the detection threshold, median $\log_{10}$FE ≥ 0.3). Rho-dependent termination signals were confirmed by *in vitro* transcription termination experiments with DNA templates holE, yobA, yqgA, and speC (see Table S5 for details). Green and purple arrows indicate the termination regions observed in the forward (strand +) and reverse (strand -) directions, respectively. 'RO' and 'Term 'locate runoff and termination transcript bands, respectively. Different x-axis scales are used for the left and right diagrams.

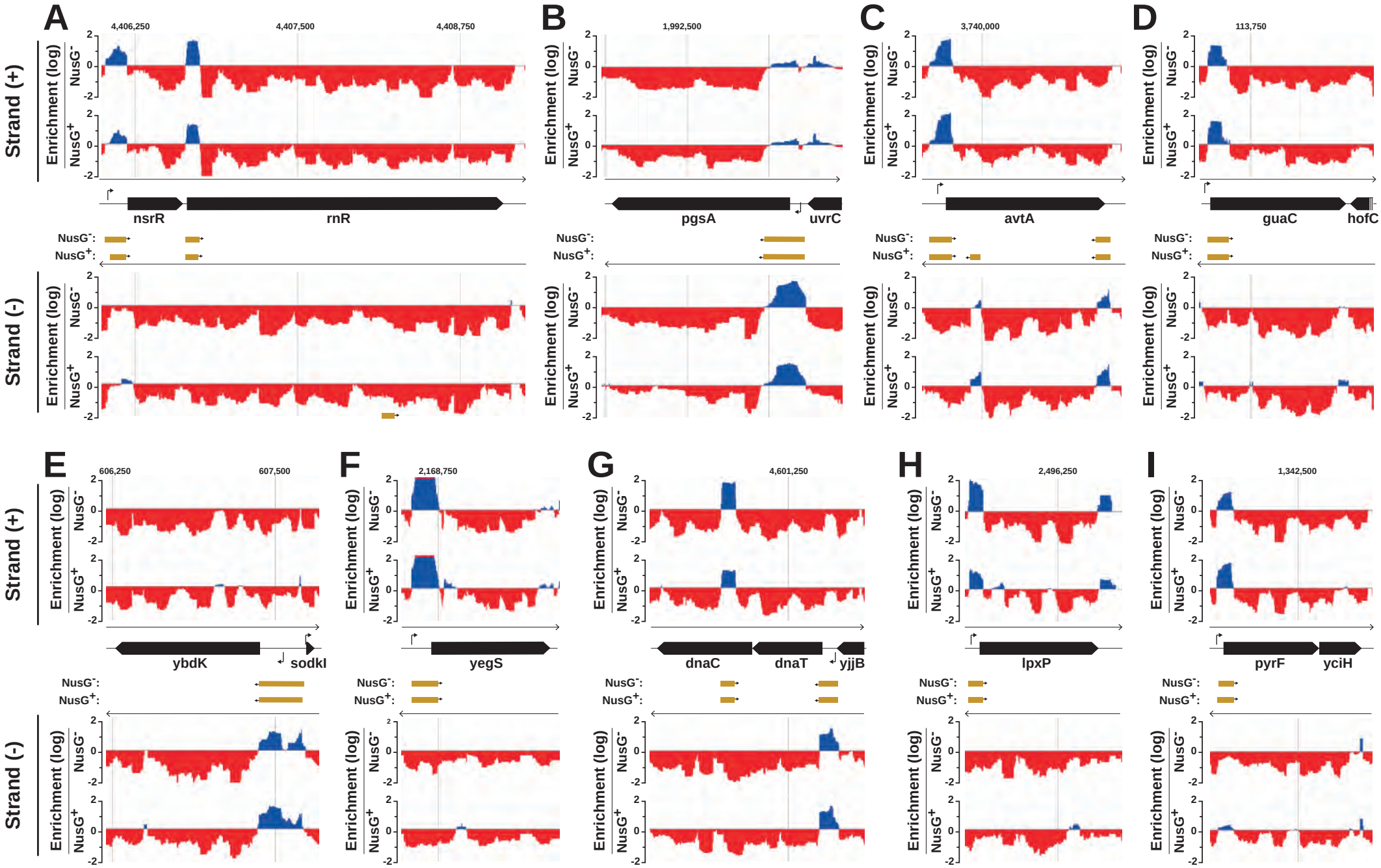


Figure S8: representative examples of *Rut* peaks found in 5'UTRs. Gold rectangles locate *Rut* sites above the detection threshold (median $\log_{10}$FE ≥ 0.3). Note that different scales along the genome coordinates (x axis) are used for the various panels.

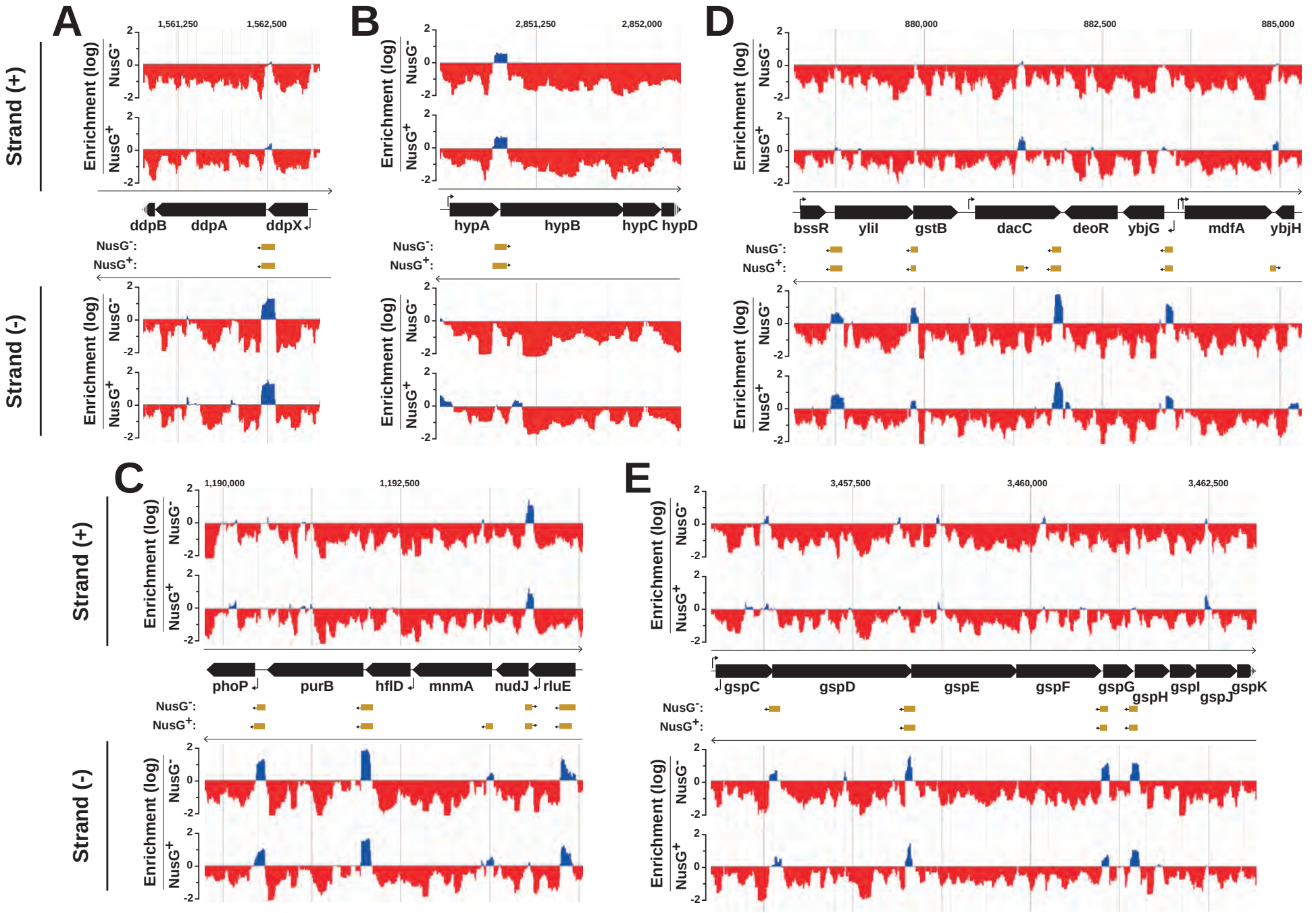


Figure S9: *Rut* peaks are frequently found at gene boundaries. Examples of *Rut* peaks located between genes of the same operon in **(A-C)** sense or **(D,E)** antisense orientation. Gold rectangles locate *Rut* sites above the detection threshold (median $\log_{10}$FE ≥ 0.3). Note that different scales along the genome coordinates (x axis) are used for the various panels.

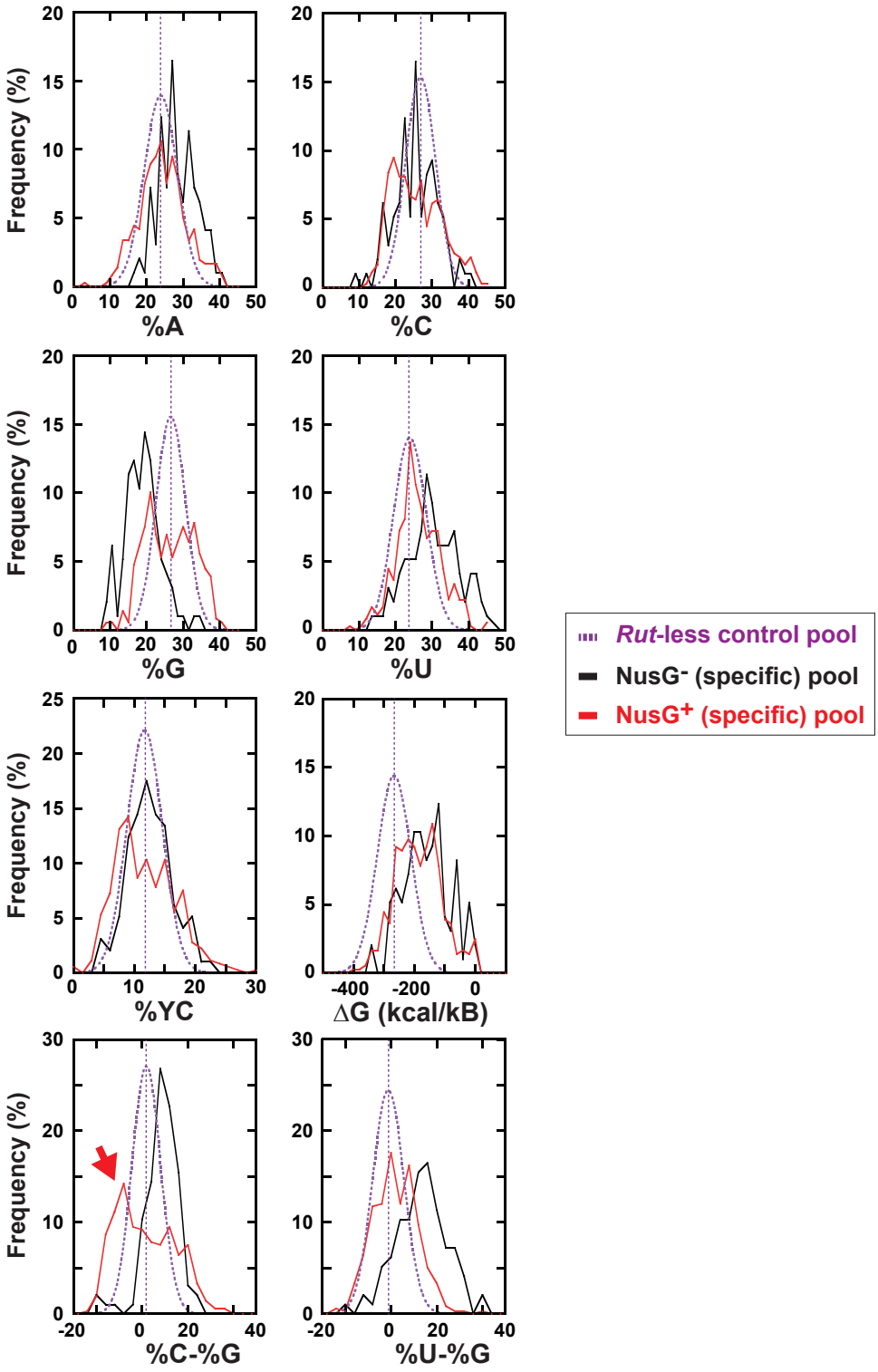


Figure S10: Parameter frequency profiles for the categories of *Rut* peaks listed in Dataset S1 (key is inset). The control pool (magenta curves) contains 3000 sequences (130 nt-long) that have been picked randomly in genomic regions devoid of *Rut* peaks (i.e in regions where median $Log_{10}FE < 0.3$ for both strands). The red arrow highlights the inverted C>G bias observed for some NusG$^+$-specific peaks.

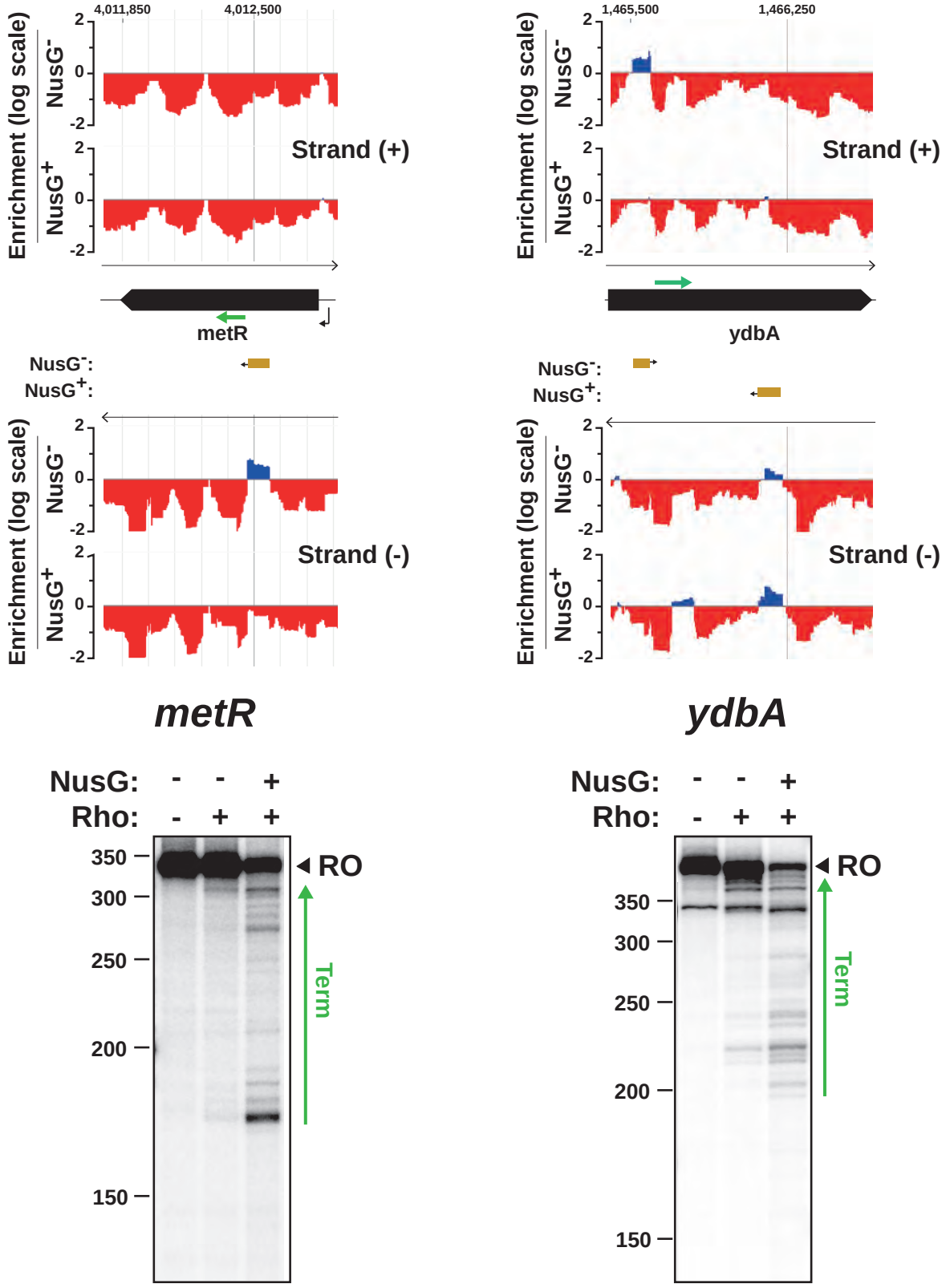
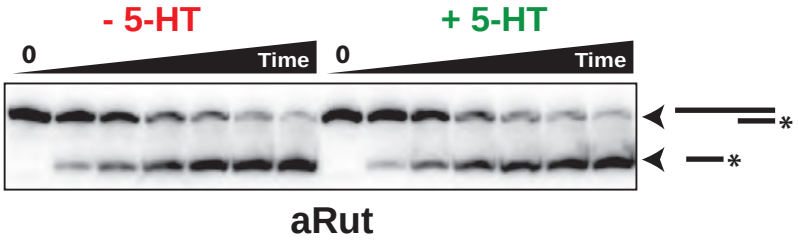


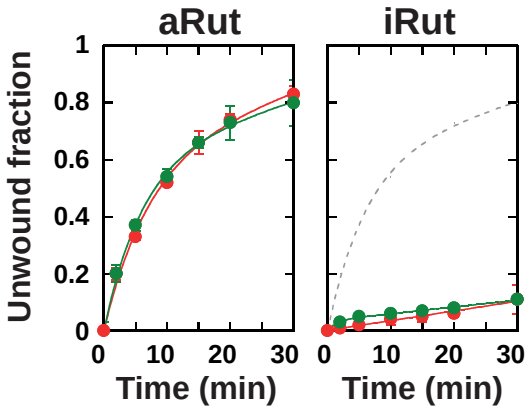
Figure S11: Representative examples of *Rut* peaks detected only in absence of NusG (median $\log_{10}$FE ≥ 0.3; gold rectangles). Weak *in vitro* termination signals (green arrows) were detected downstream from the *Rut* peaks in the presence of Rho and were stimulated by the presence rather than the absence of NusG.

A

Name	Sequence
aRut	5′ACUUCUCCUCUGUCUCCUUCUUCCUUCUCCUCUGUCUCCUUCUUCCUCGACC
iRut	5′GGUCGAGGAAGAAGGAGACAGAGGAGAAGGAAGAAGGAGACAGAGGAGAAGU
13a	5′CCCAUGUAUCGUCGAGGGCAGUUCUUGGAUCCUCUGUAAGAGAUUACGGUUAUCUCCGUAUGAAACAGUUGUUUACCCUG
13b	5′AUCAUGCAUAUUGGCCAUUGCGUUGCUCGCACCUGGGGACCUGUGCGUUCCCGCCCACCGUGUCUCCAAGGAUCCCUACA
13c	5′UUGGUCAACCAGCAACGGCACCCUGGUGCGCGCUCCCGCUUUCGACUCUCAGUCGCCUGCCAGCGCAUCGGGCAAUCCCA
13d	5′UCGACUCUAUCCCGGCCAACCACUACACCAUGUUUGGUGAACUGGUGUCGCUAUCUGGCAUCUUGACGUUUUCCACAGAA
21a	5′GUACACAACGCACCAGUCCAUGAGUUCGGUCUCGCCCGCAUAAUUCGCGCCAGCGCAGCACUCGAAUUCAUUUUCCCUU
21b	5′GCUAUUGUGUCUCUCGAGCCUCUGUUAUCGCACGCCUUGAGUUGCACUUAGUCCGCAAAUGUCCAGUGUUGUUCCCUUA
21c	5′UUCGUUCUGUUGGACUGUGUCACCUCGGGCACGAACAUGCAUGACCUCGUGCAGUCCUGUCCAGCACUAUCCAGCAAUCG
21d	5′UGAGUGAUGACCUGUGCCCGUCGCCUUACUGUCCGUAUCUCACGUUGAACUUGCCUACCUCGCGUGAGUGUUUUCCCGAU

B





C

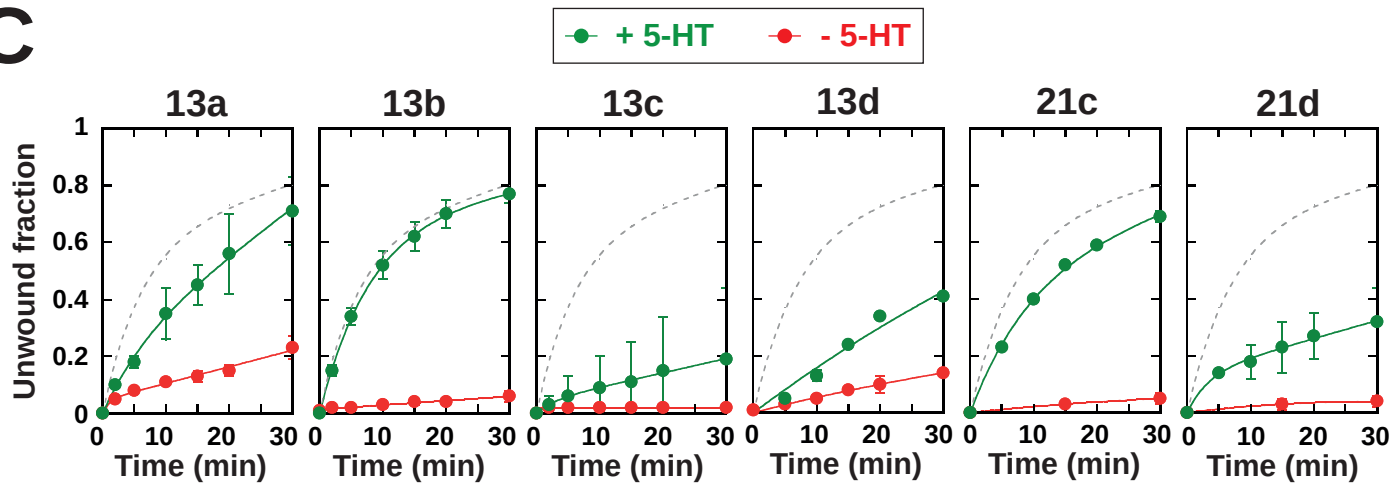


Figure S12: Ligand-dependent Rho helicase responses with selected RNA:DNA duplexes. **(A)** Schematic of the duplexes and list of ssRNA intervening sequences (in blue in the diagram). The upstream ssRNA and RNA:DNA duplex regions are the same as in Fig. 1A. **(B)** Serotonin (5-HT; 10 mM) does not affect Rho-directed unwinding of control RNA:DNA substrates bearing a strong *Rut* site (aRut) or the reverse complementary sequence (iRut). A PAGE gel for the aRut duplex (left) is shown as a representative example (the star depicts a ^{32}P label). **(C)** Rho-directed unwinding of the indicated duplexes in absence/presence of 10 mM 5-HT. Graph data points are means ± error from 2-3 independent experiments. Dotted gray curves represent the unwinding timecourse for the control aRut duplex in presence of 5-HT.

Pool	Total reads (N)	Unique reads (K)	Sequence enrichment (%) (1-K/N) x 100	%GC
sR13	4,505,433	4,484,277	0.5	53.2
sR21	2,912,590	2,061,635	29.2	54.7

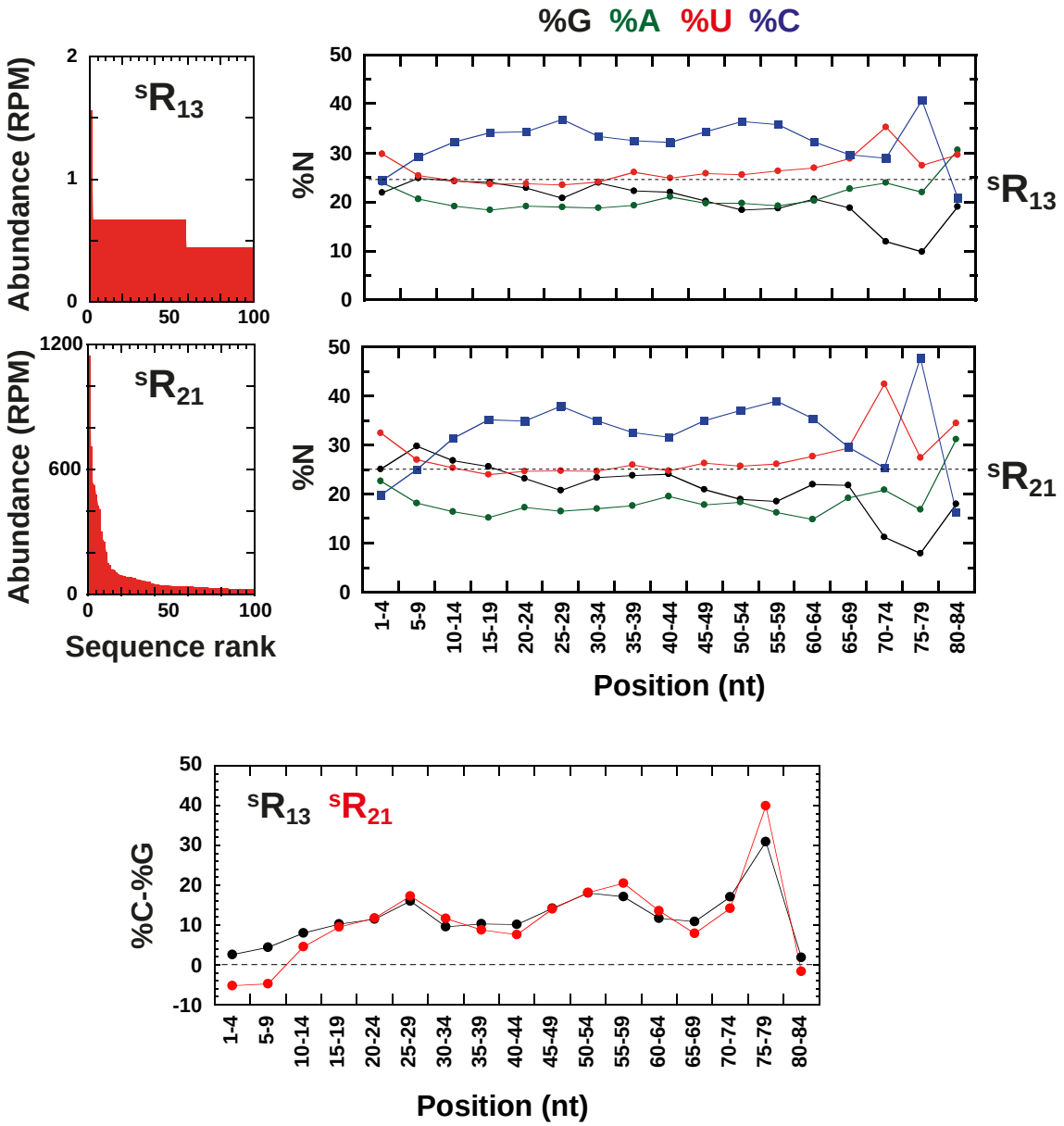
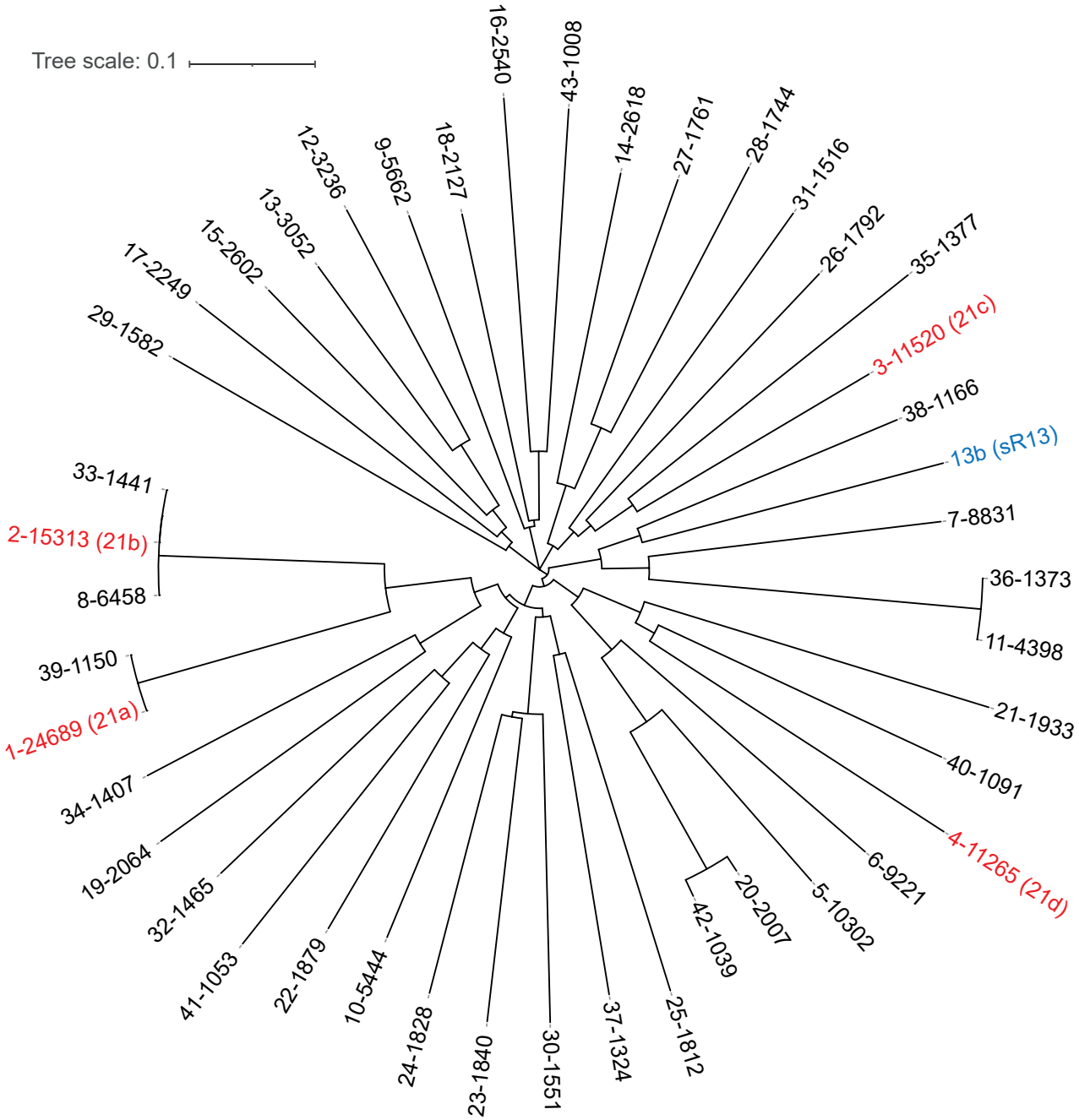
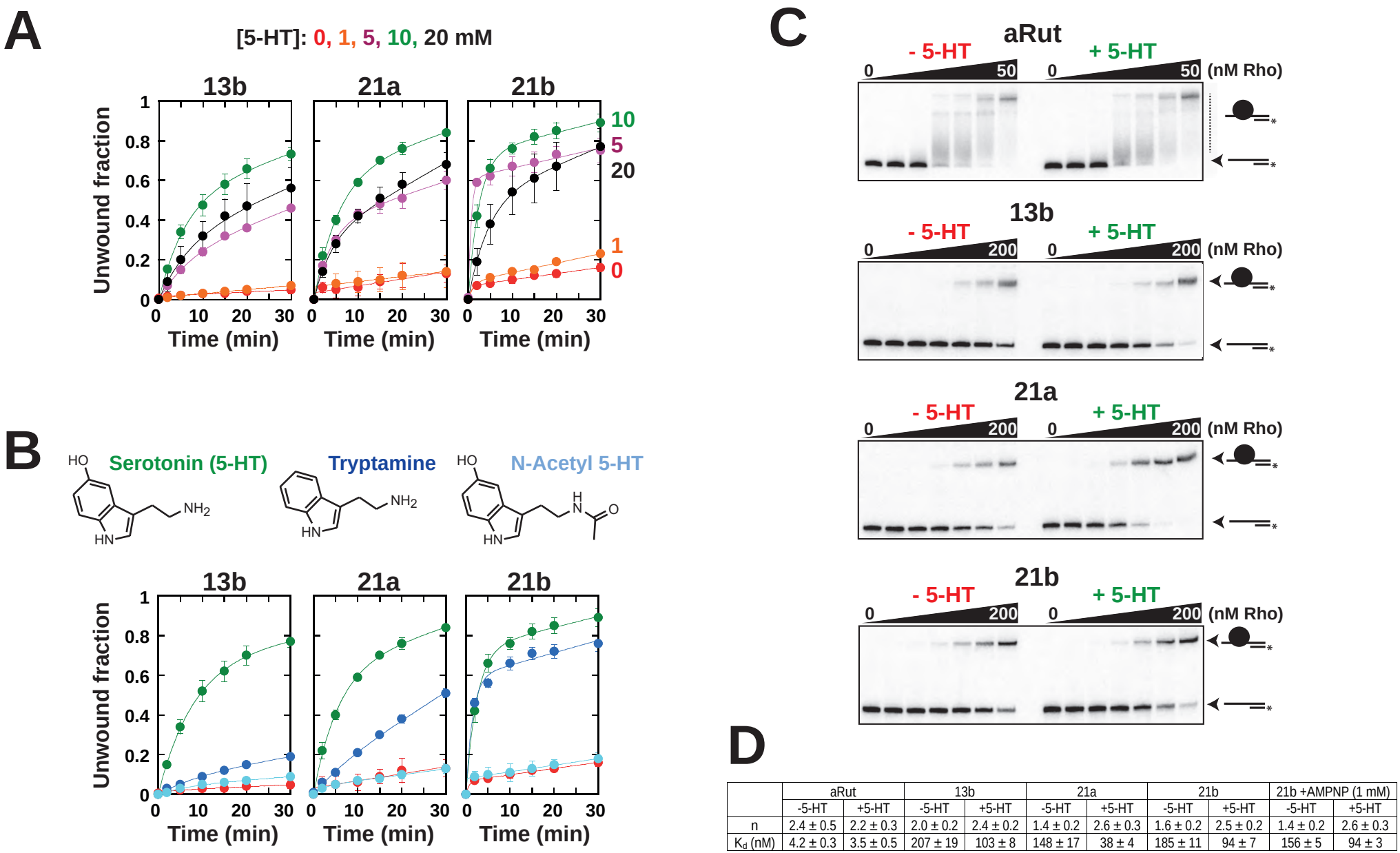


Figure S13: Compositions of the $^{s}R_{13}$ and $^{s}R_{21}$ sequence pools as determined by Illumina deep sequencing. The %N data were obtained by standard FastaQC analysis and thus represent the average %N per block of consecutive nucleotide residues along the N_{80} region for all sequences of the indicated library. Note the different scales on the abundance histograms indicating a much higher enrichment of the top-ranked sequences for the $^{s}R_{21}$ library.



FASTAptamer cluster			Top-ranked sequence		
Rank	Number of sequences	$^{s}R_{21}$ reads	Name	$^{s}R_{21}$ reads	Relative abundance in cluster (%)
1	305	36,078	21a	24,689	68.4
2	458	34,456	21b	15,313	44.4
3	164	16,018	21c	11,520	71.9
4	151	14,774	21d	11,265	76.2
717	6	164	13b	77	46.9

Figure S14: Unrooted phylogenetic tree for the top-ranked $^{s}R_{21}$ sequences (>1000 reads). The $^{s}R_{13}$ sequence 13b (in blue) has been added for comparison. The tree was built with Seaview v4.3.3 software using MUSCLE alignment (default parameters) and the BioNJ distance method (bootstrap = 100; gap sites ignored). Sequence names in tree follow the «rank-number of reads» format, as computed with the Galaxy collapse tool. The four top-ranked $^{s}R_{21}$ sequences (21a-d) also head the four top-ranked clusters identified with FASTAptamer (table). By contrast, sequence 13b is found in the 717th-ranked FASTAptamer cluster of $^{s}R_{21}$.



	aRut		13b		21a		21b		21b +AMPNP (1 mM)	
	-5-HT	+5-HT	-5-HT	+5-HT	-5-HT	+5-HT	-5-HT	+5-HT	-5-HT	+5-HT
n	2.4 ± 0.5	2.2 ± 0.3	2.0 ± 0.2	2.4 ± 0.2	1.4 ± 0.2	2.6 ± 0.3	1.6 ± 0.2	2.5 ± 0.2	1.4 ± 0.2	2.6 ± 0.3
K_d (nM)	4.2 ± 0.3	3.5 ± 0.5	207 ± 19	103 ± 8	148 ± 17	38 ± 4	185 ± 11	94 ± 7	156 ± 5	94 ± 3

Figure S15: Biochemical features of the 21a, 21b, and 13a duplexes. Rho-directed unwinding of the duplexes as a function of **(A)** serotonin concentration or **(B)** inducer (red curve: no inducer; otherwise: 10 mM inducer). **(C)** Representative EMSA gels showing the effect of ± 10 mM serotonin on the interaction of Rho (black circle) with the duplexes. **(D)** Estimates of the equilibrium binding parameters upon data fitting with the Hill equation are shown in the table. EMSA experiments were adapted from PMID:21673658 with samples containing 0.1 nM of ^{32}P-labeled duplex, 0 or 10 mM 5-HT, 30 µg/mL tRNA, and 0 to 200 nM Rho in helicase buffer. Samples were equilibrated at 37°C for 15 min before loading on a native 5% PAGE gel containing 0.1% Triton X-100.

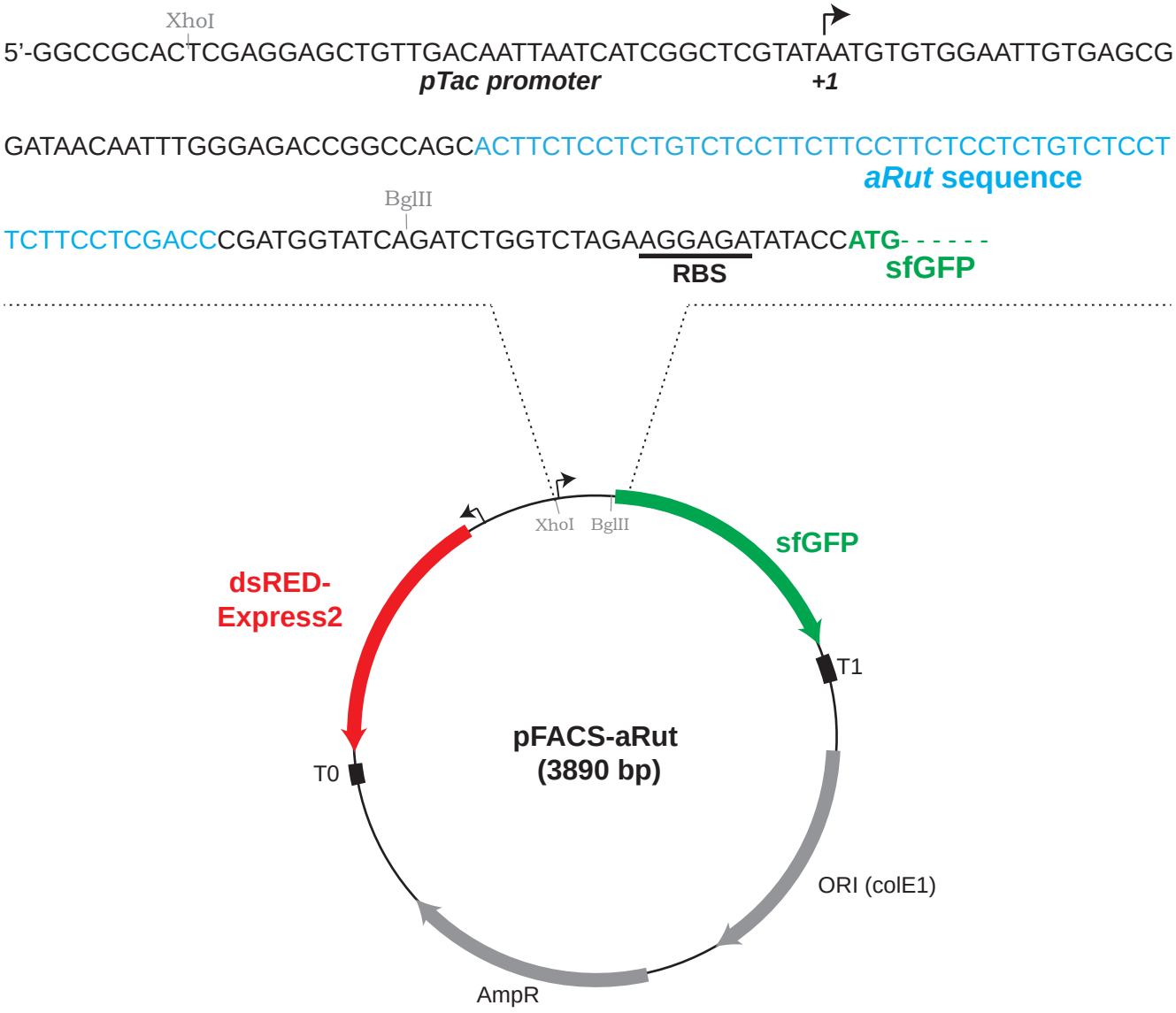


Figure S16: Schematic of the reporter pFACS-aRut plasmid. The sequence upstream from the sfGFP reporter gene is provided above the diagram. Plasmid derivatives pFACS-iRut, pFACS-13b, pFACS-21a, pFACS-21b, and pFACS-RutLess respectively contain the sequence *iRut*, 13b, 21a, 21b (listed in Figure S8A), or no sequence instead of the *aRut* sequence. The dsRED-Express2 reporter has a high maturation rate and low phototoxicity, as compared to other red fluorescent proteins (PMID: 18953349).

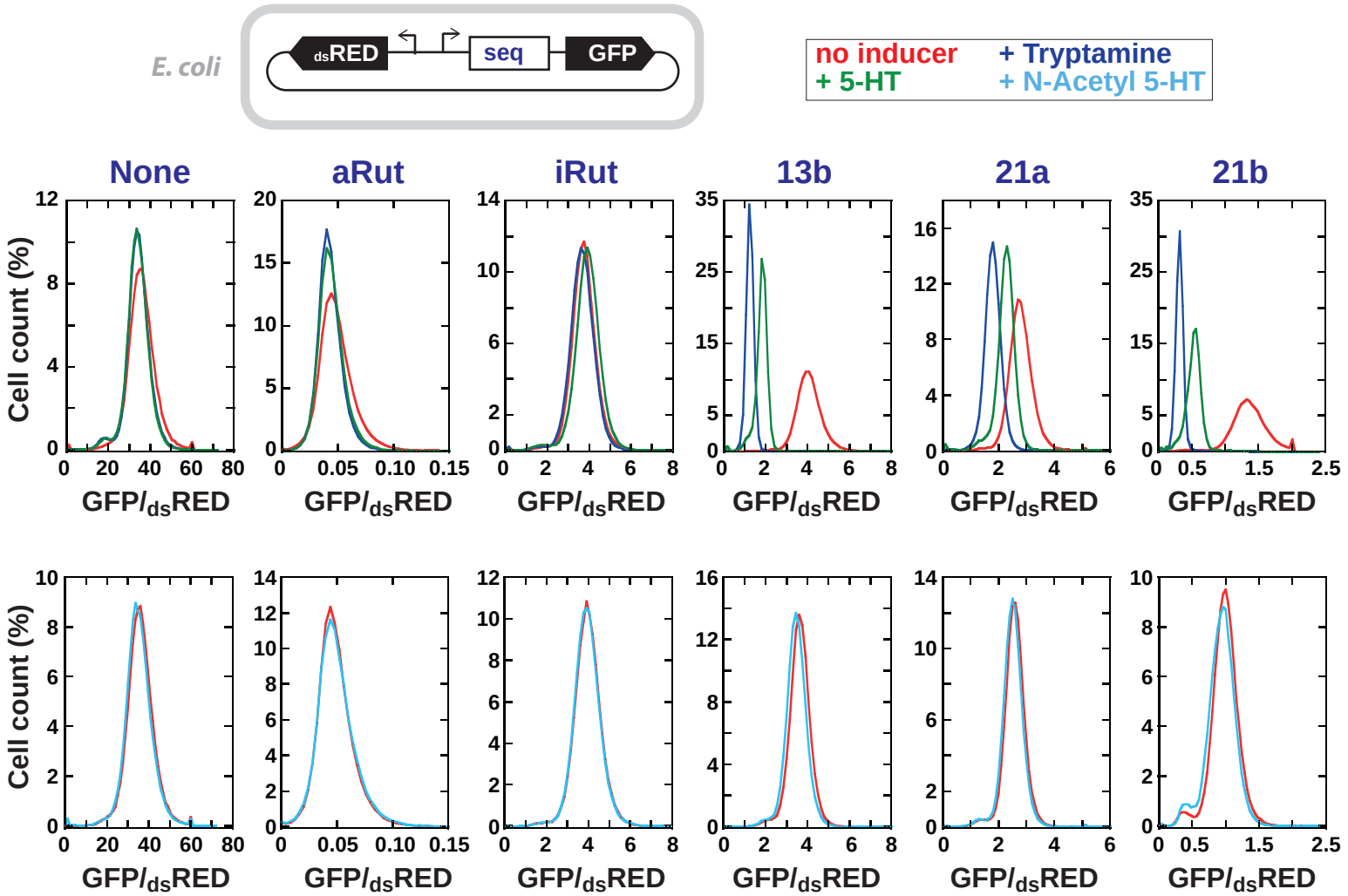


Figure S17: Normalized i*n vivo* reporter responses for the various inserted sequences (names in blue above graphs; sequences are listed in Figure S8A). The graphs show the distributions of per-cell $_{sf}$GFP/$_{ds}$RED$_{Express2}$ ratio responses obtained after 2h incubation in the presence of various inducers (key is inset) as determined by flow cytometry. All samples were prepared and analyzed the same day (samples shown in Figure 6C are from an independent experiment).