

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

Corresponding Author: Professor Robert Landick

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The regulation of transcription elongation by NusG specialized paralogs is an important and understudied topic. Of the known categories of NusG specialized paralogs, the UpxY genes associated with the Bacteroidetes capsular biosynthesis operons are the most complicated. They activate their cognate operons while anti-NusG factors block non-cognate operons. This results in hierarchical expression of the CPS operons, thereby affecting the biology of these important gut bacteria.

This manuscript presents the first detailed mechanistic study of the UpxY/UpxZ factors, including their molecular relationship to the CPS operons and to the *Bacteroides fragilis* RNA polymerase. The authors identify pausing sites associated with the CPS leader regions, which are involved in UpxY recruitment. And they show that UpxZ factors bind to non-cognate UpxY proteins to inhibit their ability to activate their CPS operons. These data are important to the transcription field, as they demonstrate a new level of complexity for NusG-driven regulation of transcription elongation and pausing. This manuscript will be well-cited by this community for those reasons. These data also will be recognized and unquestionably cited by the community of researchers studying the biology of Bacteroidetes in the gut. Overall, the data is also thorough and substantial. For those reasons, this manuscript represents an important study worthy of publication in a broad readership journal.

Minor comments and questions:

1. Line 97. "strong Eco pauses" is not defined. Is that simply referring to *E. coli*? Also, the authors state that "eight CPS leader pauses exhibited an obvious consensus sequence typical of Eco pauses and appear to be type 1 pauses, meaning pauses stimulated by nascent RNA pause hairpins..." However, where are the data showing that CPS pauses match type 1 pauses that include an RNA hairpin? The RNA hairpin is dealt with in a later figure, but I searched for some consensus comparison and did not locate it.
2. Lines 113-114. The D and G operons did not demonstrate any pause sites in vitro? It is still worth showing their gel images alongside the identification of pause sites for the other operons.
3. While reading the main text is easy and clear, locating the actual data that supports the statements in the earlier part of the main text is unnecessarily difficult. As one example, the authors state on line 114 that PIVoT assays reveal that, "strikingly, the PSA, B, E, F and H leader segments encoded single prominent pause sites..." This sounds wonderful and I anxiously looked for these data. The authors cited Fig 1b, d to support their statement. However, Fig 1b is a summary graphic and Fig 1d shows gel images of the E site in the presence and absence of NusA and UpxYe, which is discussed later in the text. However, the authors also cited Ext Data Figs 2 and 3. Yet, Ext Data 2 indeed introduces the PIVoT assay but includes a panel on an unrelated claim (panel c), a panel with unclear purpose (panel d), and two more panels on unrelated claims (panels e and f). It is not that these data are not important. For example, I presume panel f is being shown as a representative demonstration for how the pausing locations were generally identified? And panels c and e correlate with other claims in the text. So, presumably, the pausing data that I am searching for is shown in Ext Data 3. However, some Ext Data 3 panels show gel slices that include both Run-off and pausing sites while other panels show only a narrower slice of a pause site, which is not marked by position or size marker. Some panels include NusA and others don't. Also, the pausing is not quantified, and the legend does not explicitly state that the experiment was performed in replicate. It was – the methods section says so. And the main text indicates that the data were quantified. Yet at this point, I've inspected three figures and still don't feel like I've seen the data that supports the original claim: "the PSA, B, E, F and H leader segments encoded single prominent pause sites." I recommend that the authors consider separating some of the extended data figures into

multiple items so that the data can be more easily found and understood, and so that more data, such as larger gel slices, or half-life measurements, can be shown.

4. Lines 123-124. It is stated that Y factors either inhibited or enhanced pause sites but figures are not cited. Presumably, this refers to Ext Data Fig 3. However, the data in this figure are not quantified, so, one can't quickly assess the effect of Y factors on pause sites. The authors should add half-life measurements and error measurements

5. Line 129-130. "The effects of Yx were specific to the NET-seq identified leader pauses..." It is difficult to independently assess this from Ext Data Fig 3 since a marker was not included for the NET-seq pause site. I think the authors have estimated the pause sites using DNA ladders, and I assume they are also correct in their conclusion, but since it literally can't be directly assessed, I would soften the statement to include "likely to be specific..."

6. Line 177. Include reference for statement that the antibodies were specific for PSE.

7. Line 207, Perhaps I'm confused, but the sentence cites Fig 1, which doesn't seem to be the correct figure.

8. Also regarding line 207, it doesn't seem meaningful to say there is variability in KOW domain. From the authors' data, there appears to be pockets of high and low conservation within the KOW domains. It is unclear what exactly is being said about the KOW domain.

9. General comment: it is substantially difficult to examine the gels shown in the figures as they are very small panel additions.

Reviewer #2

(Remarks to the Author)

Saba and colleagues studied the regulation of multiple capsular polysaccharides that facilitates capsule diversity in human gut *Bacteroides* species. Diverse capsules contribute to surface variability, which helps protect *Bacteroides* from phage infection and plays a role in processes like immune modulation, biofilm formation, antibiotic resistance, and inflammation. Each CPS operon is preceded by paralogs of the transcription elongation factor

NusG, called UpxY (Y), and an anti-Y UpxZ (Z). While Y proteins enhance transcript elongation Z inhibits noncognate Ys. Combined with promoter inversions this regulation limits CPS transcription to a single operon in most cells. In this manuscript, Saba and colleagues show that Y recognizes a paused RNA polymerase via sequences in both the exposed non-template DNA and the upstream duplex DNA and Z binds non-cognate Y proteins to inhibit their association with RNAP.

Their conclusions were supported by data showing that:

CPS leader regions encode pause sites for RNAP with similar but not identical sequences, as might be expected for Y recruitment sites that must distinguish among Y paralogs.

Purified Y proteins tested for their effect on specific pausing sites showed that Y of (A,B,E) inhibited the cognate leader pause, whereas Y(C,F,H) enhanced the cognate leader pause.

Measurements of binding constants by biolayer interferometry showed that Z proteins inhibit non-cognate Y proteins by direct interaction. Furthermore, Z blocked cognate Y action only at high concentrations

In vivo studies in which opsE segments were replaced with the corresponding opsA segments showed that Y(A) activated expression of PS(E), indicating that ops (X) pause sites function as Y (X) recruitment sites

To examine whether the variability in ops (X) sequences could be related to the variability in Y paralogs, the authors compared the apparent evolutionary relationships to sequence and structural alignments of Y (x) proteins RfaH, and NusGs, showing that both Y(X) protein and opsX DNA sequences clustered into two distinct classes (anti-pausing A, B, E and pro-pausing D, F, E). The outliers were PS G and C.

The authors conclude that capsule diversity results from intricate complex regulation that involves pause-dependent, locus-specific cognate Y activation and noncognate Z inhibition.

Major comments

Saba and colleagues have conducted an impressive yet highly intricate study. Those who aren't familiar with molecular specifics may find it easy to get lost in the complexity. The sheer volume of details and technical jargon can make it difficult to follow the narrative. Consequently, it's essential to simplify the presentation of the results by clearly outlining the questions, findings, and conclusions. It would be beneficial to incorporate some of the text from the results into the discussion. Additionally, consider shortening the manuscript, possibly by omitting the section on NusA, as it doesn't contribute significantly to the overall understanding. Furthermore, the last three chapters of the results, while interesting, are somewhat tangential and do not enhance overall comprehension. In this instance, less is more.

Reviewer #3

(Remarks to the Author)

The manuscript by Saba et al. investigates complex mechanisms that determine specificity of UpxY family of transcription elongation factors which control the expression of capsular polysaccharides (PS) in *Bacteroides fragilis*.

Unlike RfaH, their distant cousin from *E. coli*, UpxY proteins have to compete not only with the housekeeping NusG but also among themselves to achieve locus-specific expression of capsular polysaccharides. Surprisingly, different UpxYs use quite different strategies to control their productive recruitment to the transcription elongation complex (a combination of initial loading onto RNA polymerase and escape) and also rely on cross-inhibition by UpxZ proteins to ensure that (mostly) one PS is expressed at a time. The authors paint a very detailed picture of this regulation, with nicely flowing text and figures. A newly identified dual role of a nascent RNA hairpin in pause and escape and the use of upstream DNA duplex as an additional recognition determinant for NGN are very interesting.

The manuscript is well written, the figures are clear, and conclusions are supported by the data. The Methods section is

exemplary; all the expected details are present and information that is typically omitted (yet very helpful) is also included. That said, the manuscript is overloaded with information and is therefore quite dense; combined with very complicated and different mechanisms of UpxYs it could be difficult to follow at times, particularly by people who are not familiar with NusG and its specialized paralogs.

In figures, every detail matters, including colors of lines in cartoons. Even though I think that cartoons are very helpful and that the authors did make an excellent job in using them throughout the figures, I feel that sometimes there is just too much detail and not enough words to draw attention to them. Fonts are tiny in figures – sure, people can zoom in, but looking at the entire figure helps.

Comments on figures:

Figure 1d: the cartoon depicting the pause-escape pathway is great but is not necessarily helpful here: it is not well described in the legend or in the main text, in contrast to Figure 6 where it is. To simplify Fig. 1, this cartoon could be deleted and panel C can go into the supplement.

In Figure 2f, colored lines around "cells" presumably indicate if PSE is expressed – but the last, full ops, cell should have a green line around it as well; I completely missed this point. Also, lines that connect the cells to lanes on the Western blot are not needed, as long as the labels match on cartoons and on the blot – and they do not. All of these variants are promoter-locked, so the label on the leftmost cell is not needed. And perhaps draw a WT cell with promoters in different orientations?

Figure 3: in Yx effect column, "—" can be interpreted as no effect, maybe explain

Figure 4: a & c, "asRNA pairing position..." caption is difficult to read; maybe just say in the legend that it is marked relative to the pause position and make the font less compressed.

Figure 5 and 6: yellow font is tough to see on a white background

L63/73 abbreviate the nt DNA at the first mention

L69/75 Proteobacteria vs Pseudomonadota - I prefer the former, but only one is needed

L74 a reference or 2 should be added after the sentence describing RfaH; they are cited anyway

L129 "Importantly, the effects of YX were specific to the NET-seq identified leader pauses" – does this mean that Yxs do not reduce pauses anywhere else in the x operon? Do we know if they reduce polarity, if observed, in the cps operons?

L148 The first part of the first sentence does not belong here. Can even start from the second sentence.

L350 State that the Zx mechanism of interference is inferred from modeling that, even if high confidence, is still hypothetical. Although the authors present convincing evidence that Interactions between Z and Y are direct and the hierarchical inhibition makes logical sense, no functional data that support the mode of binding suggested by AF3 are presented. This is not the central point of the UpxY mechanism of action, and I do not think that such data are required, but the conclusion needs to be softened.

L478 source of GlycoBlue?

L711 something happened to ref 95

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have completed substantial changes to the manuscript, which together successfully accommodated the critiques and questions posed by myself and the other reviewers. I do not have any remaining criticisms for the manuscript and support its publication as currently written. These data represent an important contribution to the field and will be cited accordingly.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

The regulation of transcription elongation by NusG specialized paralogs is an important and understudied topic. Of the known categories of NusG specialized paralogs, the UpxY genes associated with the Bacteroidetes capsular biosynthesis operons are the most complicated. They activate their cognate operons while anti-NusG factors block non-cognate operons. This results in hierarchical expression of the CPS operons, thereby affecting the biology of these important gut bacteria.

This manuscript presents the first detailed mechanistic study of the UpxY/UpxZ factors, including their molecular relationship to the CPS operons and to the *Bacteroides fragilis* RNA polymerase. The authors identify pausing sites associated with the CPS leader regions, which are involved in UpxY recruitment. And they show that UpxZ factors bind to non-cognate UpxY proteins to inhibit their ability to activate their CPS operons. These data are important to the transcription field, as they demonstrate a new level of complexity for NusG-driven regulation of transcription elongation and pausing. This manuscript will be well-cited by this community for those reasons. These data also will be recognized and unquestionably cited by the community of researchers studying the biology of Bacteroidetes in the gut. Overall, the data is also thorough and substantial. For those reasons, this manuscript represents an important study worthy of publication in a broad readership journal.

Minor comments and questions:

1. Line 97. “strong Eco pauses” is not defined. Is that simply referring to *E. coli*? Also, the authors state that “eight CPS leader pauses exhibited an obvious consensus sequence typical of Eco pauses and appear to be type 1 pauses, meaning pauses stimulated by nascent RNA pause hairpins...” However, where are the data showing that CPS pauses match type I pauses that include an RNA hairpin? The RNA hairpin is dealt with in a later figure, but I searched for some consensus comparison and did not locate it.

We thank the reviewer for noting these areas of ambiguity. We have added a panel (Ex Data 1c) to illustrate how CPS pauses match type I pauses that include an RNA hairpin. We have also edited the text to direct the reader to Fig 1b for a definition of “strong Eco pauses”:

Line 95-99:

“NET-seq revealed single prominent pause sites in most CPS operon leader regions (Fig. 1a,b; Extended Data Fig. 1b)³⁷. Eight CPS leader pauses exhibited an obvious consensus sequence that resembles strong *E. coli* pauses (Fig. 1b) as well as apparent nascent RNA pause hairpins (PHs) that resemble those known to enhance pausing allosterically in concert with NusA in other bacteria (e.g., the so-called type-1 *E. coli his* and *B. subtilis trp* leader region pauses; Extended Data Fig. 1c)^{37,38,39,40}. Pausing in the PSC leader region (the only *Bfr* CPS operon with a non-invertible, constitutively ON promoter)⁴¹ occurred at multiple sites; weak pausing occurred at a site resembling the other seven in sequence and location (Fig. 1b and Extended Data Fig. 1b). We designated the CPS leader pause sites *ops_x* (‘X’ designates the CPS operon) based on analogy to the RfaH *ops* site.”

In Figure 4ab, we provide evidence for RNA hairpin stimulation of pausing at a representative Class 1 *ops_X* pause. Additionally, NusA strongly stimulates Class 1, Class 2, and *ops_C* pauses, consistent with predictions of RNA hairpins at these sites (Extended Data Fig. 3). NusA does not stimulate pausing at *ops_E* when the pause RNA hairpin is disrupted or when it is deleted (Fig. 4a, Extended Data Fig. 8a).

2. Lines 113-114. The D and G operons did not demonstrate any pause sites in vitro? It is still worth showing their gel images alongside the identification of pause sites for the other operons.

We did not perform in vitro experiments involving the D and G operons because we were unable to purify *Y_D* and *Y_G* proteins. We have added asterisks (*) to Figure 1b to clarify that the pause sites are known from the in vivo data only. We also have asterisks next to these operon letters in the 'Y_X effect' column in Fig 3b to indicate that the *Y_D* and *Y_G* effects are inferred.

3. While reading the main text is easy and clear, locating the actual data that supports the statements in the earlier part of the main text is unnecessarily difficult. As one example, the authors state on line 114 that PIVoT assays reveal that, “strikingly, the PSA, B, E, F and H leader segments encoded single prominent pause sites...” This sounds wonderful and I anxiously looked for these data. The authors cited Fig 1b, d to support their statement. However, Fig 1b is a summary graphic and Fig 1d shows gel images of the E site in the presence and absence of NusA and UpxYe, which is discussed later in the text. However, the authors also cited Ext Data Figs 2 and 3. Yet, Ext Data 2 indeed introduces the PIVoT assay but includes a panel on an unrelated claim (panel c), a panel with unclear purpose (panel d), and two more panels on unrelated claims (panels e and f). It is not that these data are not important. For example, I presume panel f is being shown as a representative demonstration for how the pausing locations were generally identified? And panels c and e correlate with other claims in the text. So, presumably, the pausing data that I am searching for is shown in Ext Data 3. However, some Ext Data 3 panels show gel slices that include both Run-off and pausing sites while other panels show only a narrower slice of a pause site, which is not marked by position or size marker. Some panels include NusA and others don't. Also, the pausing is not quantified, and the legend does not explicitly state that the experiment was performed in replicate. It was – the methods section says so. And the main text indicates that the data were quantified. Yet at this point, I've inspected three figures and still don't feel like I've seen the data that supports the original claim: “the PSA, B, E, F and H leader segments encoded single prominent pause sites.” I recommend that the authors consider separating some of the extended data figures into multiple items so that the data can be more easily found and understood, and so that more data, such as larger gel slices, or half-life measurements, can be shown.

We thank the reviewer for their suggestions to improve the manuscript and have made major changes to Ext Data Figs 2 and 3 to address their concerns. Ext Data 2ab introduces the PIVoT assay and Ext Data 2c shows results with the consensus pause. Ext Data 2d now shows all six assayed CPS leader regions in the absence of factors. Ext Data 3 shows factor effects on these pauses, some of which (i.e., *ops_C*, *ops_F*, and *ops_H*) are more obvious pause sites in the

presence of NusA or Y_X . We have moved details of quantitative pause mapping to Supplementary Fig. 1 and removed the old Supplementary Fig. 1 (a picture of scaffolds) since the information is redundant with what is provided in the Supplementary Materials (pS17).

With regard to the suggestion that not all gel panels include both run-off and pause RNAs and aren't well marked, we have edited Ext Data Fig 3 to have consistent gel slices (all gels have run-off and pausing sites). We have indicated pause band sizes on the gel, which were obtained by a variety of methods: i) running PSE-, PSF-, and PSC-transcribed samples alongside ladders and quantitative mapping; ii) subsequent comparison of ops_E alongside ops_A and ops_B (Source Data for ExFig 6b for ops_A , Source Data for Fig 5e for ops_B) to confirm sizes of these latter pauses; iii) ops_F and ops_H pauses were additionally verified by comparison to a sequencing ladder (Supplementary Fig 1c).

With regard to the concern that pausing is not quantified, we now have provided quantitative pausing data for representative Class 1 and Class 2 ops_X/Y_X in Fig. 1e. Quantitative pausing data for each Class 1 Y_X is provided in Figs 2d, 4bd, and 5cef.

4. Lines 123-124. It is stated that Y factors either inhibited or enhanced pause sites but figures are not cited. Presumably, this refers to Ext Data Fig 3. However, the data in this figure are not quantified, so, one can't quickly assess the effect of Y factors on pause sites. The authors should add half-life measurements and error measurements.

We thank the reviewer for noting this oversight. We provide quantitative data for this statement for representative classes in Fig 1d, and have edited the text accordingly to direct the reader as follows.

Line 125-126:

"Intriguingly, $Y_{A,B,E}$ inhibited the cognate leader pause, whereas $Y_{C,F,H}$ enhanced the cognate leader pause (Fig 1c,d, Extended Data Fig. 3)."

5. Line 129-130. "The effects of Y_X were specific to the NET-seq identified leader pauses..." It is difficult to independently assess this from Ext Data Fig 3 since a marker was not included for the NET-seq pause site. I think the authors have estimated the pause sites using DNA ladders, and I assume they are also correct in their conclusion, but since it literally can't be directly assessed, I would soften the statement to include "likely to be specific..."

We have edited the statement accordingly, as follows. Line 130-133:

"Importantly, the effects of Y_X were likely to be specific to the NET-seq identified leader pauses, consistent with ops_X sites functioning as specific Y_X -recruitment sites. Y_X specifically modulated pausing at cognate ops_X but not non-cognate ops_X or other positions (Extended Data Fig. 3)."

6. Line 177. Include reference for statement that the antibodies were specific for PSE.

The hybridoma producing this mAb was produced in the 1990s, a time when it was thought that *B. fragilis* produces only two capsular polysaccharides, PSA and PSB. With the subsequent demonstration that this organism produces eight capsular polysaccharides, we were able to determine that this mAb reacts to PSE. As these data are unpublished, we now provide a western immunoblot showing that this mAb reacts with all seven other polysaccharide mutants, but not with the PSE mutant (Supplementary Figure 5).

We have added a reference to this supplementary figure on Line 445-446:

“The blot was probed with a mouse monoclonal antibody specific to PSE (Supplementary Figure 5), washed with TBST, and probed with alkaline phosphatase conjugated goat-anti mouse IgG (Pierce).”

7. Line 207, Perhaps I’m confused, but the sentence cites Fig 1, which doesn’t seem to be the correct figure.

We thank the reviewer for identifying this typo. We have corrected the text to reference Fig 3a as follows. Line 210:

“...(iii) variability in the C-terminal KOW domain (Fig 3a, Extended Data Fig. 7).”

8. Also regarding line 207, it doesn’t seem meaningful to say there is variability in KOW domain. From the authors’ data, there appears to be pockets of high and low conservation within the KOW domains. It is unclear what exactly is being said about the KOW domain.

We have edited the paragraph to include a statement on the potential impact of KOW variability, as follows. Line 212-213:

“The variable Y_x sequences in contacts to the ntDNAhp, protrusion, and gate loop are consistent with Y_x recognition and potential effects on pausing^{27,56,57}, whereas variability in KOW may enable novel target specificity mechanisms or coupling of transcription to other cellular processes.”

9. General comment: it is substantially difficult to examine the gels shown in the figures as they are very small panel additions.

We have expanded Figure 1 and Figure 5 gels to aid their examination by the reader.

Reviewer #2 (Remarks to the Author):

Saba and colleagues studied the regulation of multiple capsular polysaccharides that facilitates capsule diversity in human gut *Bacteroides* species. Diverse capsules contribute to surface

variability, which helps protect *Bacteroides* from phage infection and plays a role in processes like immune modulation, biofilm formation, antibiotic resistance, and inflammation.

Each CPS operon is preceded by paralogs of the transcription elongation factor NusG, called UpxY (Y), and an anti-Y UpxZ (Z). While Y proteins enhance transcript elongation Z inhibits noncognate Ys. Combined with promoter inversions this regulation limits CPS transcription to a single operon in most cells. In this manuscript, Saba and colleagues show that Y recognizes a paused RNA polymerase via sequences in both the exposed non-template DNA and the upstream duplex DNA and Z binds non-cognate Y proteins to inhibit their association with RNAP.

Their conclusions were supported by data showing that:

CPS leader regions encode pause sites for RNAP with similar but not identical sequences, as might be expected for Y recruitment sites that must distinguish among Y paralogs.

Purified Y proteins tested for their effect on specific pausing sites showed that Y of (A,B,E) inhibited the cognate leader pause, whereas Y(C,F,H) enhanced the cognate leader pause.

Measurements of binding constants by biolayer interferometry showed that Z proteins inhibit non-cognate Y proteins by direct interaction. Furthermore, Z blocked cognate Y action only at high concentrations

In vivo studies in which opsE segments were replaced with the corresponding opsA segments showed that Y(A) activated expression of PS(E), indicating that ops (X) pause sites function as Y (X) recruitment sites

To examine whether the variability in ops (X) sequences could be related to the variability in Y paralogs, the authors compared the apparent evolutionary relationships to sequence and structural alignments of Y (x) proteins RfaH, and NusGs, showing that both Y(X) protein and opsX DNA sequences clustered into two distinct classes (anti-pausing A, B, E and pro-pausing D, F, E). The outliers were PS G and C.

The authors conclude that capsule diversity results from intricate complex regulation that involves pause-dependent, locus-specific cognate Y activation and noncognate Z inhibition.

Major comments

Saba and colleagues have conducted an impressive yet highly intricate study. Those who aren't familiar with molecular specifics may find it easy to get lost in the complexity. The sheer volume of details and technical jargon can make it difficult to follow the narrative. Consequently, it's essential to simplify the presentation of the results by clearly outlining the questions, findings, and conclusions. It would be beneficial to incorporate some of the text from the results into the discussion. Additionally, consider shortening the manuscript, possibly by omitting the section on NusA, as it doesn't contribute significantly to the overall understanding. Furthermore, the last three chapters of the results, while interesting, are somewhat tangential and do not enhance overall comprehension. In this instance, less is more.

We appreciate suggestions from the reviewer to improve our manuscript by removing results. In response, we have removed the data in Fig. 4b analyzing the combinatorial effects of NusA and the pause hairpin. The figure is now much improved in clarity, enabling the reader to focus

on differences between pause hairpins minus and plus escape duplexes. We also edited the text to accordingly, as follows (see ~~strikeout~~) (Line 246-247).

“asDNAs that disrupt the PSE PH by pairing with the 5' arm but not those that pair just upstream reduced pausing ~~both in the absence and presence of NusA~~ (Fig. 4b).”

We have also removed one gel panel in original Fig 5c that illustrated the quantitation, since it was redundant. This change enabled us to expand the size of the remaining gel panel to improve overall clarity. We think these changes greatly simplify the presentation of results and trust they sufficiently address the reviewer's concerns.

Reviewers 1 and 3 found the latter three chapters of the results very interesting, so we have opted to keep these data in the manuscript. Figure 4 highlights a novel RNA element (escape duplexes) associated with *ops_x* sites. Figures 5 and 6 demonstrate the UpxYs utilize upstream duplex DNA to enable target-specificity among multiple NusG paralogs. The escape duplexes are associated only with UpxYs that make particularly extensive contacts to their targets. Together, these adaptations enable *Bacteroides* to manage multiple NusG paralogs.

Reviewer #3 (Remarks to the Author):

The manuscript by Saba et al. investigates complex mechanisms that determine specificity of UpxY family of transcription elongation factors which control the expression of capsular polysaccharides (PS) in *Bacteroides fragilis*.

Unlike RfaH, their distant cousin from *E. coli*, UpxY proteins have to compete not only with the housekeeping NusG but also among themselves to achieve locus-specific expression of capsular polysaccharides. Surprisingly, different UpxYs use quite different strategies to control their productive recruitment to the transcription elongation complex (a combination of initial loading onto RNA polymerase and escape) and also rely on cross-inhibition by UpxZ proteins to ensure that (mostly) one PS is expressed at a time. The authors paint a very detailed picture of this regulation, with nicely flowing text and figures. A newly identified dual role of a nascent RNA hairpin in pause and escape and the use of upstream DNA duplex as an additional recognition determinant for NGN are very interesting.

The manuscript is well written, the figures are clear, and conclusions are supported by the data. The Methods section is exemplary; all the expected details are present and information that is typically omitted (yet very helpful) is also included.

That said, the manuscript is overloaded with information and is therefore quite dense; combined with very complicated and different mechanisms of UpxYs it could be difficult to follow at times, particularly by people who are not familiar with NusG and its specialized paralogs.

In figures, every detail matters, including colors of lines in cartoons. Even though I think that cartoons are very helpful and that the authors did make an excellent job in using them throughout the figures, I feel that sometimes there is just too much detail and not enough words to draw attention to them. Fonts are tiny in figures – sure, people can zoom in, but

looking at the entire figure helps.

We appreciate the suggestions from the reviewer to improve our manuscript and have made the following changes:

- We moved Fig 1c to Extended Data Fig 2b
- We removed the data in Fig. 4b analyzing the combinatorial effects of NusA and the pause hairpin (see response to Reviewer 2).
- We have added labels to most panels to improve overall clarity.
- We have removed a redundant gel panel (quantified data remain) from Figure 5 and rearranged remaining panels (see response to reviewer 2). The figure now has consistent fonts and the overall appearance of the figure is more balanced.
- We increased font sizes where possible and minimized font format variability.

Comments on figures:

Figure 1d: the cartoon depicting the pause-escape pathway is great but is not necessarily helpful here: it is not well described in the legend or in the main text, in contrast to Figure 6 where it is. To simplify Fig. 1, this cartoon could be deleted and panel C can go into the supplement.

We appreciate the suggestion to improve the figure and have moved panel C to Extended Data Fig 2b.

In Figure 2f, colored lines around "cells" presumably indicate if PSE is expressed – but the last, full ops, cell should have a green line around it as well; I completely missed this point. Also, lines that connect the cells to lanes on the Western blot are not needed, as long as the labels match on cartoons and on the blot – and they do not. All of these variants are promoter-locked, so the label on the leftmost cell is not needed. And perhaps draw a WT cell with promoters in different orientations?

We thank the reviewer for noting the lack of green line around the 'full ops' mutant and have corrected the error. We also have removed the label on the leftmost cell as suggested. In our figure, the single cells for each mutant represent the population because it is promoter-locked, but a single WT cell would represent only a single combination, so we opted to not show a WT cell.

Figure 3: in Yx effect column, "—" can be interpreted as no effect, maybe explain

We thank the reviewer for noting this point of ambiguity. We edited the superscript '#' description associated with 'Y_x Effect[#]' to clarify: "[#]Anti(–) or Pro(+) pausing effect of Y_x".

Figure 4: a & c, "asRNA pairing position..." caption is difficult to read; maybe just say in the legend that it is marked relative to the pause position and make the font less compressed.

We thank the reviewer for identifying the compressed font in this caption and have made appropriate changes to improve its clarity.

Figure 5 and 6: yellow font is tough to see on a white background

We have added black strokes to the yellow font to improve font clarity.

L63/73 abbreviate the nt DNA at the first mention

We have added the abbreviation at line 63.

L69/75 Proteobacteria vs Pseudomonadota - I prefer the former, but only one is needed

We thank the reviewer for identifying this inconsistency and have made the change on L69 to 'Proteobacteria'.

L74 a reference or 2 should be added after the sentence describing RfaH; they are cited anyway

We have added references as suggested and as follows (Line 73).

"RNAP pauses at the 12-nucleotide *ops*, allowing RfaH to associate via sequence-specific interactions with a non-template strand DNA hairpin (ntDNAhp) exposed by the paused RNAP^{22,23,32,33}."

L129 "Importantly, the effects of YX were specific to the NET-seq identified leader pauses" – does this mean that Yxs do not reduce pauses anywhere else in the x operon? Do we know if they reduce polarity, if observed, in the cps operons?

We do not observe obvious Yx-modulated pauses at other positions on our tested templates. Yx was previously shown to be required for expression of downstream genes and removal of the leader region caused polarity (Chatzidaki et al, *J. Bacteriology* 2009, 23 7288-7295, PMID: 19801412).

L148 The first part of the first sentence does not belong here. Can even start from the second sentence.

We have edited the sentence as follows (Line 151):

"We next sought to test whether Y_x binding requires sequence upstream of the putative ntDNAhp region using in vitro binding, in silico interaction, and in vivo gene expression assays."

L350 State that the Zx mechanism of interference is inferred from modeling that, even if high confidence, is still hypothetical. Although the authors present convincing evidence that Interactions between Z and Y are direct and the hierarchical inhibition makes logical sense, no functional data that support the mode of binding suggested by AF3 are presented. This is not the central point of the UpxY mechanism of action, and I do not think that such data are required, but the conclusion needs to be softened.

We have softened our conclusions about the Zx mechanism of action by specifying that the structure is inferred from modeling, as follows (Line 355):

“We found that Z_X inhibits Y_X recruitment to ops_X -PECs directly, likely by blocking Y_X interaction with the conserved β' clamp helices (CH) and the ops_X usDNA.”

L478 source of GlycoBlue?

We thank the reviewer for noting this missing detail. We have edited the manuscript as follows (Line 478, 486):

“The adenylation reaction was incubated for 4 hrs incubation at 65°C, inactivated at 85°C for 5mins, and precipitated overnight at -20°C with isopropanol and GlycoBlue (Invitrogen AM9516).”

L711 something happened to ref 95

We thank the reviewer for noting the formatting error and have corrected it as follows (Line 718):

“A model of Y_B was made using Modeller^{94,95} and fitted to 8PHK³³. Additional upstream and downstream DNA were modeled using Pymol.”