Heterochromatin-like domains silence harmful genes and promote transcriptional memory in bacteria

Haley Amemiya, Thomas Goss, Taylor Nye, Rebecca Hurto, Lyle Simmons, and Peter Freddolino

Review Timeline:		
	Transfer from Review Commons:	12th May 21
	Editorial Decision:	18th May 21
	Revision Received:	15th Sep 21
	Editorial Decision:	19th Oct 21
	Revision Received:	22nd Nov 21
	Accepted:	29th Nov 21

Editor: Stefanie Boehm

**Transaction Report: This manuscript was transferred to
The EMBO Journal following peer review at Review Commons.**

Dear Dr. Freddolino,

Thank you for transferring your manuscript from Review Commons to The EMBO Journal. I have now carefully read your study, as well as looked at the referee reports and your response, and discussed everything with the other members of the editorial team. Taking all this into consideration, we have decided to invite a revision at The EMBO Journal.

In the revised version it will be important to fully address the referees' main concerns and to ensure that all conclusions and more general models are either sufficiently supported by data or explicitly discussed as being more speculative. In your response to the Review Commons referee reports, you have indicated that you plan to provide further experimental support for the proposed transcriptional memory effect. We agree with the referees (Ref#1- page 8,9; ref#2- main point 2; ref #3- point 2) that this is an important point and additional experimental evidence must thus be included in the revised version. In addition, please also revise the manuscript and figures to address the more technical issues and questions the referees raise on specific points, as well as expanding the information on methods and in figure legends. When submitting the revised manuscript, please also remember to include a detailed point-by-point response listing the referee comments and your response individually for each issue.

Please also be aware that it is our policy to allow only a single round of major revision. We realize that addressing all referee points fully, in particular the key issues, may require substantial additional analyses with potentially unclear outcomes. As we are aware that lab work worldwide is currently still affected by the pandemic and that an experimental revision may be delayed, we can extend the revision time when needed and have extended our 'scooping protection policy' to cover the period required for a full revision. However, we will require strong support from all referees on the revised version of the manuscript in order to consider it further for publication at EMBO Journal. Should you encounter issues that significantly delay a revision or make certain experiments not feasible, please contact us to discuss potential alternatives as soon as possible.

Please also feel free to contact me should you have any other further questions. Thank you for the opportunity to consider your work, I look forward to receiving your revised manuscript.

Kind regards,

Stefanie Boehm

Stefanie Boehm
Editor
The EMBO Journal

Referee reports

Review #1 -

This is an interesting study addressing the molecular basis and functional consequences of chromatin organization in bacteria. The authors combine a chromatin enrichment assay IPOD-HR, transcriptomics and functional genomics technologies to show the presence of heterochromatin-like domains (EPODs) on different genomic islands in E. coli and B. subtilis. This study extends their pioneering work published in 2009 to a functional level and reveal exciting new features of bacterial genome organization

These interesting findings should interest a large community around dynamic fields such as chromosome organization and management, bacteriophage regulation, genome dynamics. The large amount of data accompanying this manuscript will also be beneficial to many researchers.
I found several minor points that should be addressed before publication:

• The EPOD calling method is quite complex. It could be helpful for the first figure to show on an example that include EPODs and non EPOD region, the raw data, the Z score and the EPOD calling with the different tresholds.

• The authors used two "different" MG1655 strains for their study, they claim that genetic differences might appear between these lab strains and influence EPODs (page 4). Figure 1A shows that the number of EPODs is substantially changed in these two stains (210 vs 280). The authors did not discuss that and did not use strain number 2 for the rest of the manuscript. I do not really see the interest of this data. Comparing different strains, pathogens and commensal for example can be really interesting, may be it is beyond the scope of this manuscript.

• Page 4 "the relaxed threshold EPODs ...." This notion was not defined , it makes this paragraph difficult to follow.

• Page 5. The authors show EPODs enrichment for the GO term cellular response to acid chemical is rather mysterious. In E. coli this GO term only contain 6 members with diverse functions (chemotaxis x2, amino acid transport, acid pH response, small RNA and an unknown protein) . I am surprised that this group present a significant EPODs enrichment. A table of EPODs coordinate and overlapping genes will be very helpful to explore the data.

• Page 5 and figure 1E: I don't understand the representation used for this figure. What does the X axis correspond to ? Kb ? Number? IP/input ? I don't understand why this show an enrichment of HNS IP inside EPODs? What does the red star mean?

• Figure 1D . The color code is not clear for me. What are the two green shades on each panel?

• Page 8-9: Increased lag phase upon medium changes is well known features on E.

coli microbiologist. Here the 40 h lag is particularly long. Did the authors check that they did not pick up mutants rather than epigenetic adaptation? It can explain the faster adaptation to the second shift. Moreover I do not see causality between the presence of the EPOD and the adaptation. Is lag phase changed in the H-NS or H-NS/stPA mutants?

• Page 9: "... to ensure sufficiently tight silencing of these genes except when they are needed (Fig 5B) " I do not see that on figure 5B.

• Figure 5A : could the author show the entire Rac prophage here.

• Figure 5B and page 9-10: Data presented on figure 5b do not sustain the claim that dps, hupAB, hfq , ... show a decrease in median occupancy across prophages . Medians are not changed, perhaps distribution are changed (*) . Clearly the representation is not adequate to illustrate this point. The number of EPODs used to make this plot should me mentioned. A representation with each EPOD might be more appropriate to reveal interesting outliers.

• Figure 5C. It would be interesting to include the information of the phage appurtenance for each EPOD. Here it is not clear for me if a given phage will present only one EPOD controlled my a single or multiple NAPs or if the prophage is split in different EPODs with different chromatin factors?

• Page 10: The authors need to clarify the text, it is not always evident to distinguish if they are talking about ChIP-seq data or IPOD data. Here, for example "This also aligned with a specific type of class of EPODs associated with high Fis and HFQ binding (Fig 6A). On figure 6A there is no ChiP seq data. I assume that.

• Figure 6A . The authors claim that prophage presented on figure 6A lost protein occupancy upon deletion of fis or hfq (page 10). I do not see the protein occupancy lost on the presented data. For some part of the prophage protein occupancy is even higher in the mutants than in the WT.

• Figure 6A and B . could the authors annotate the essD rpzD borD ybcV genes on fig 6A.

• Figure 6D: 1 hexbin (0,2.5) contains 10-14 EPODs while all others are below 5. Is it an artefact?

These interesting findings should interest a large community around dynamic fields such as chromosome organization and management, bacteriophage regulation, genome dynamics. The large amount of data accompanying this manuscript will also be beneficial to many researchers.

Review #2 –

The work by Amemiya et al "Distinct heterochromatin-like domains promote transcriptional memory and silence parasitic genetic elements in bacteria" describes the contribution of nucleoid associated proteins (NAP) to the transcriptionally inactive, kb-long EPODs region along the E. coli genome that were described by the same team in a former study.

The authors investigate EPODs in NAPs mutants. They show that in single mutants, the distribution of EPODs doesn't vary much when the cells are in the same growth conditions. Some NAPs have hardly any effect on those. This suggests that the global protein occupancy patterns remain somehow redundant and that the chromosomes don't like to stay "naked" (but, nature abhors vacuum, right). Interestingly, deletion of pairs of NAPs stpA and hns have a stronger impact on EPODs, with a strong decreased being observed in the double mutant, although both single mutants have little impact suggesting the two NAPs can compensate for each other's loss. Several evidences suggest that these regions could be relevant for both adaptation to metabolic changes, and that EPOD would maintain some kind of epigenetic memory over generations, and silencing of genomic elements otherwise detrimental to the cell.

Main concerns:

"ePOD patterns are conserved across laboratory evolution": the authors have tested only a single variant strain of MG1655! This is not enough to state the former, they should tune down the claim, just mention they used two 1655 background strains. And, I tend to disagree that the PODs are stable between the two WT, as they represent ~210 and 280 positions in rich medium? To me that looks like a relatively significant difference, it could be interesting to dig in a bit further to double check whether this is really only subtle differences from the EPODs calling, or a subset of genes/regions that behave differently. But given the close phylogenetically relationships between these strains, this is a bit puzzling.

The "memory" observation is very intriguing and an exciting observation. To me it is really one of the most interesting aspect of the work (although the use of the word "memory" could be discussed). How long would it last? Would adding additional doubling would diminish it?
I think the authors should validate their hypothesis that Idn proteins are diluted, as they suggest. And I think they could validate their hypothesis by adding only one or two experiments that would bring (hopefully) a lot. For instance, could add a couple of divisions and see if the response increases at 5kDG#2. They could also check whether entering stationary phase eliminates the "memory" by giving cells ample time to reassemble EPODs. Or any experiment that would give time to the cells in the population to do so.

Could the authors mention how many prophage EPODs they are looking at? 43 or so, according to panel 5C? Have they done a null model using random analysis of regions

covering a similar amount of DNA to double-check the variations they observe in all the conditions are significant (not just compared to wt) ?

The "negative" EPOD observed for B. subtilis result from the z-score computation. The authors actually don't show the negative y axis of the IPOD-HR for all the E. coli panels. They should be consistent, and display the negative scores as well for those. What does a "nEPOD" actually mean, with respect to the sequencing data? Given it seems to correlate with AT rich sequences, could it be a sequencing caveat, where the region would still contain proteins (Rok, etc.) but be less "visible" throughout the experiment? The B. subtilis genome contains several prophages that are highly AT rich. Could the authors double check that the reads align correctly in these regions as well? (it may be in the methods, in which case i missed it). I also think the authors should discuss or be more specific about the regions they are looking at (plot the GC%, for instance; annotate some genes on panel 7C, B; etc.)

Minor concerns

Figure 1D: state this is a representative region in the legend.

Figure 3B: I am not sure about the color labels. The author claim a loss of EPOD in 5KDG, but there is still quite an enrichment in protein occupancy in this condition, no? When is this enrichment large enough to be called an EPOD is a matter of parameters, but The analysis is in z-score, so what are the absolute values of the enrichment? Would it be possible to put the profil in supplementary figures, just to give an idea to the reader? I agree that the result shows a change upon growth in 5KDG, but I'd be interested to see more raw data. The Hns Chip profile could also be plotted along the IPOD-HR occupancy?

I think S4A and B panels should be reduced and included in the main figure 2D, and that the RNA expression in single Hns and StpA mutant should also be included.

Figure 5A. SieB is the prophage? I don't understand this panel. The EPODs barely seem to vary in the region, and it looks as if there are more proteins spread along the whole region in the hns single mutant. Could the authors improve the figure and the text to be more clear on what they are pointing at?

I have a logical issue with the comment "EPODs may be functional units that serve as hotspots for foreign DNA", as the causality links are unclear to me at this stage. I may have missed something though.

I find the study interesting, though sometimes a bit speculative, notably regarding the generalization of some experiments.

The "memory" part which, in my opinion, deserves a couple of additional experiments to be further investigated, is an exciting observation and hypothesis.

The study will be of interest of researchers interested in bacterial chromosomal structure, organization, and gene expression regulation. It should also be of interest to some people working in epigenetics and chromosome regulation in other organisms.

Review #3 -

This paper describes the role of Nucleoid Associated Proteins (NAPs) in the formation of Extended Protein Occupancy Domains (EPODs). These are features previously described in the E. coli chromosome, but whose function remain unclear. Furthermore, the contribution of individual NAPs in their formation is also unclear. Using the sequencing-based technique of in vivo protein occupancy display at high resolution (iPOD-HR), the authors test the individual contribution of NAPs to the formation of EPODs by deleting them. They then show that loss of EPODs correlate with expression of genes in certain metabolic pathways and of prophages in E. coli, suggesting a role for these protein-DNA features in silencing section of the chromosome. They also apply this approach in the Gramm positive Bacillus subtilis to demonstrate that the EPODs can also be found in phylogenetically distant bacteria. Based on these results, the authors propose that the EPODs are a conserved strategy for gene silencing in bacteria that has features similar to those of heterochromatin in eukaryotes.

The paper is thought-provoking and provides a new perspective on the role of NAPs and the regulation of gene expression in bacteria. My main criticism is that the paper often lacks enough details to understand it. The method section is brief, and the paper does not have any diagram describing the main method used, iPOD-HR. This also prevents the understanding of some of the analysis used in the paper.

Other comments:

1. The description of the symetrized overlap distance (page 4) is unclear, and could not find any other section where the authors extended this description. Similarly, is unclear what is a relaxed threshold EPOD
2. The observation of EPOD-mediated transcription memory is perhaps the most surprising result in the paper (Figure 4). However, the authors only tried one condition in these experiments, leaving multiple unanswered questions. If possible, I suggest to test whether the outcome of this transcriptional memory changes with the length of the period that cells are grown in glucose. Extending these observations would strengthen the transcriptional memory claim. It is also unclear in this section the functional advantage of such transcriptional memory (how frequently is E. coli expected to switch between carbon sources in the wild?)
3. In page 11 (Figure 6) it is unclear why did the authors choose a deletion of cspE as a control
4. I suggest not to abuse the use of heterochromatin as analogy, since it only goes so far. Although EPODs are linked to silencing, they contain no nucleosomes or other factors of eukaryotic chromatin. There is obviously also no ancestral relation between heterochromatin and EPODs. Finally, the phenomenon has its own merit as a purely bacterial mechanism. This is why I suggest that authors avoid terms such as 'Heterochromatin domains', which was used in the title of the last section of the results
5. Unclear sentence: "These findings implicate Hfq, a well documented RNA

chaperone and that has only been recently explored as a protein to compress dsDNA" (page 10)

The paper is thought-provoking and provides a new perspective on the role of NAPs and the regulation of gene expression in bacteria. It should be of interest to a broad audience (fields: molecular biology, cell biology, microbiology)

Response to Reviewers

We thank all three reviewers for their detailed readings of the manuscript and helpful feedback, which has enabled us to substantially strengthen the key points of the paper. In the test below, we provide all reviewer comments in *italic text*, our responses in normal text, and additions to the manuscript made directly in response to those comments as quotations or block quotes. The revised version of the manuscript that we have submitted highlights all textual changes in green text to facilitate identifying points that have changed.

Reviewer #1 (Evidence, reproducibility and clarity (Required)):

This is an interesting study addressing the molecular basis and functional consequences of chromatin organization in bacteria. The authors combine a chromatin enrichment assay IPOD-HR, transcriptomics and functional genomics technologies to show the presence of heterochromatin-like domains (EPODs) on different genomic islands in E. coli and B. subtilis. This study extends their pioneering work published in 2009 to a functional level and reveal exciting new features of bacterial genome organization

These interesting findings should interest a large community around dynamic fields such as chromosome organization and management, bacteriophage regulation, genome dynamics. The large amount of data accompanying this manuscript will also be beneficial to many researchers. I found several minor points that should be addressed before publication:

Comment:
• The EPOD calling method is quite complex. It could be helpful for the first figure to show on an example that include EPODs and non EPOD region, the raw data, the Z score and the EPOD calling with the different tresholds.

Response:
We thank the reviewer for their comment and have included a new panel, Fig. 1A, that shows an example of an EPODs called at different thresholds, both in an EPOD and non-EPOD area, and the robust Z scores that drive those EPOD calls. All raw data is publicly available through GEO as GSE164796 (reviewer key ozufkmmwddqhrix). We also note that the methods paper describing the EPOD calling methods in particular has recently been published (*PLoS Biology* 19(6):e3001306, 2021).

Comment:
• The authors used two "different" MG1655 strains for their study, they claim that genetic differences might appear between these lab strains and influence EPODs (page 4). Figure 1A shows that the number of EPODs is substantially changed in these two stains (210 vs 280). The authors did not discuss that and did not use strain number 2 for the rest of the manuscript. I do not really see the interest of this data. Comparing different strains, pathogens and commensal for example can be really interesting, may be it is beyond the scope of this manuscript.

Response:
We thank the reviewer for this comment, which provided us with the impetus to more deeply consider what aspects of the strain genotypes might give rise to the differences that we observed. As described in the new text incorporated into the section "Large-scale patterns of protein occupancy are highly maintained across conditions and laboratory isolates", we traced the likely cause of the differences in EPODs between these two laboratory strains to two mutations in the genes *crl* and *dgcJ*, both of which exert substantial control over motility and

adhesion genes. Indeed, we found upon inspection of the differences in EPOD locations between the two strains that there was an enrichment of differential EPODs on genes involved in motility, providing us with a connection between the known regulatory impacts of mutated genes and the corresponding pattern of EPOD occupancy. We believe that this new analysis provides a sharper justification for including the comparison involving WT(2).

We also agree with the reviewer that a broader survey of protein occupancy patterns across different strains would be of great interest; however, we believe it falls beyond the scope of the current manuscript, where we have focused instead on the effects of different nucleoid-associated proteins.

Comment:
• Page 4 "the relaxed threshold EPODs ...." This notion was not defined , it makes this paragraph difficult to follow.

Response:
A deeper explanation of the thresholds has been added to the beginning of the paragraph noted by the reviewer.

Comment:
• Page 5. The authors show EPODs enrichment for the GO term cellular response to acid chemical is rather mysterious. In E. coli this GO term only contain 6 members with diverse functions (chemotaxis x2, amino acid transport, acid pH response, small RNA and an unknown protein) . I am surprised that this group present a significant EPODs enrichment. A table of EPODs coordinate and overlapping genes will be very helpful to explore the data.

Response:
We have provided files of gene lists that fall within each EPOD category in Supplementary File 3.

Comment:
• Page 5 and figure 1E: I don't understand the representation used for this figure. What does the X axis correspond to ? Kb ? Number? IP/input ? I don't understand why this show an enrichment of HNS IP inside EPODs? What does the red star mean?

Response:
We have updated the corresponding captions to indicate that the displayed plots are kernel density plots, which are effectively smoothed histograms. The key piece of information being conveyed is that in all of the indicated conditions, there is an enrichment of H-NS ChIP signal in the EPOD regions (with significance indicated by the red star, as described in the corresponding caption). This information is consistent with previous findings, and serves to motivate the consideration of nucleoid-associated protein deletions in the remainder of the paper.

Comment:
• Figure 1D . The color code is not clear for me. What are the two green shades on each panel?

Response:
We put a color key to better indicate what is displayed. The two shades of green in one panel is just the border (darker green) of the robust z score plotted (light green) -- the color coding

serves primarily to link the EPOD calls under different conditions with the corresponding occupancy traces.

Comment:
• Page 8-9: Increased lag phase upon medium changes is well known features on E. coli microbiologist. Here the 40 h lag is particularly long. Did the authors check that they did not pick up mutants rather than epigenetic adaptation? It can explain the faster adaptation to the second shift. Moreover I do not see causality between the presence of the EPOD and the adaptation. Is lag phase changed in the H-NS or H-NS/stPA mutants?

Response:
We examined the input sequencing data from the IPOD data in the shift experiments and did not observe mutations in the Glu2 timepoint (which took place after KDG exposure). In addition, to provide an expanded view of the effects of pre-exposure to KDG on competitive fitness, we have replaced the second round of KDG lag time kinetics with a series of head-to-head competition experiments between KDG-exposed cells (after various passage times in glucose media) with KDG-naive cells. We found that the memory persisted after 12 hours of growth (sufficient for the cells to undergo eight doublings), but was mostly lost after 24 hours and completely gone after 48 hours in glucose minimal media (Figure 4); the results are detailed in newly added text in the section titled "Metabolic pressures induce changes in EPODs". The combination of a direct lack of mutations in the population of cells that grew in KDG, and the reversion of KDG-exposed cells to show no growth benefits in KDG after 24-48 hours in glucose media, all argue against a mutational origin for the observed memory.

The reviewer's suggestion to look at the behavior of H-NS/StpA cells in these experiments is well taken; however, in general the deletion of these genes yields such a strongly pleiotropic phenotype that we do not believe any changes could be usefully ascribed to effects on memory at the *idn* promoter rather than some other impact on cell growth.

Comment:
• Page 9: "... to ensure sufficiently tight silencing of these genes except when they are needed (Fig 5B) " I do not see that on figure 5B.

Response:
The original figure callout was incorrect and should have referred to Figure 4B; however, in the revised text this section has been substantially re-written anyway to focus on newly added head-to-head competition experiments.

Comment:
• Figure 5A : could the author show the entire Rac prophage here.

Response:
We increased the Rac prophage area displayed in Fig. 5A, which more clearly shows the distinction between the noted H-NS silenced operons and surrounding regions.

Comment:
• Figure 5B and page 9-10: Data presented on figure 5b do not sustain the claim that dps, hupAB, hfq , ... show a decrease in median occupancy across prophages . Medians are not changed, perhaps distribution are changed (*) . Clearly the representation is not adequate to illustrate this point. The number of EPODs used to make this plot should me mentioned. A

representation with each EPOD might be more appropriate to reveal interesting outliers.

Response:

We have addressed each of the points above in three different ways: 1) We've included a deeper description of the statistics displayed, including a more nuanced differentiation from the significance tests used in Fig. 5B vs Fig 5C; 2) Included the number of EPODs displayed, 3) added a supplemental figure that shows the changes displayed in Figure 5B per EPOD for each genotype compared to WT (Figure S6).

Comment:
• Figure 5C. It would be interesting to include the information of the phage appurtenance for each EPOD. Here it is not clear for me if a given phage will present only one EPOD controlled my a single or multiple NAPs or if the prophage is split in different EPODs with different chromatin factors?

Response:
We have included text on page 11("Interestingly, Δ*hns* and Δ*hfq* are clustered together, as are Δ*stpA* and Δ*fis,* suggesting prophages are silenced by multiple factors and NAPs. The association of multiple NAPs to specific prophage EPODs may indicate that they act cooperatively, for instance in the case of H-NS and StpA, or independently, which in many cases remains to be determined.") to explain that multiple NAPs can silence both in concert or independently but remains to be determined in the case of Fis and Hfq. Some prophages, especially large ones such as Rac, will indeed contain multiple EPODs. The interplay between different NAPs in establishing the several EPODs within a prophage (and indeed, even within a single EPOD) is an ongoing topic of investigation in our laboratory, and we believe that attempting to add this information to the present manuscript would prove to be overwhelming. The reviewer will likely be interested to know, however, that in our analysis there are indeed several different classes of EPODs present in some large prophages. For example, in both the Rac and DLP12 prophages, there are several regions dominated by class 2 (H-NS dependent) EPODs, directly flanking other regions dominated by class 5 (Fis/Hfq-dependent) EPODs. We expect in ongoing research to further characterize the phenomenology, and mechanistic interconnections, of these binding events, but that is (in our view) beyond the scope of the present manuscript.

Comment:
• Page 10: The authors need to clarify the text, it is not always evident to distinguish if they are talking about ChIP-seq data or IPOD data. Here, for example "This also aligned with a specific type of class of EPODs associated with high Fis and HFQ binding (Fig 6A). On figure 6A there is no ChiP seq data. I assume that.

Response:
More text was provided to indicate that there whether IPOD or ChIP-seq data was being described on page 10. We also added text to the Figure 6A axis labels and caption to further this explanation in this portion of the text.

Comment:
• Figure 6A . The authors claim that prophage presented on figure 6A lost protein occupancy upon deletion of fis or hfq (page 10). I do not see the protein occupancy lost on the presented

data. For some part of the prophage protein occupancy is even higher in the mutants than in the WT.

Response:
We added dashed lines to Figure 6A to show the level of occupancy in WT and applied the dashed lines to the rest of the panels for the mutants to better appreciate the loss of occupancy over that region. While we agree that the losses in occupancy are modest, there is a quantitative decrease throughout the genes of interest. We also note at the end of the same paragraph that since we see evidence that Fis and Hfq work together to silence expression at these loci, we might well expect that deletion of both factors would be needed to see complete loss of occupancy (as we showed for H-NS and StpA in FIgure 2).

Comment:
• Figure 6A and B . could the authors annotate the essD rpzD borD ybcV genes on fig 6A.

Response:
The gene annotations were added.

Comment:
• Figure 6D: 1 hexbin (0,2.5) contains 10-14 EPODs while all others are below 5. Is it an artefact?

Response:
The high density in this hexbin was a result of the identical expression value given from Rockhopper. While it is still valid, we decided to change the analysis to Kallisto followed by Sleuth to be consistent throughout the figure; the overplotting noted by the reviewer is absent in the revised analysis.

Reviewer #1 (Significance (Required)):

These interesting findings should interest a large community around dynamic fields such as chromosome organization and management, bacteriophage regulation, genome dynamics. The large amount of data accompanying this manuscript will also be beneficial to many researchers.

Reviewer #2 (Evidence, reproducibility and clarity (Required)):

The work by Amemiya et al "Distinct heterochromatin-like domains promote transcriptional memory and silence parasitic genetic elements in bacteria" describes the contribution of nucleoid associated proteins (NAP) to the transcriptionally inactive, kb-long EPODs region along the E. coli genome that were described by the same team in a former study.

The authors investigate EPODs in NAPs mutants. They show that in single mutants, the distribution of EPODs doesn't vary much when the cells are in the same growth conditions. Some NAPs have hardly any effect on those. This suggests that the global protein occupancy patterns remain somehow redundant and that the chromosomes don't like to stay "naked" (but, nature abhors vacuum, right). Interestingly, deletion of pairs of NAPs stpA and hns have a stronger impact on EPODs, with a strong decreased being observed in the double mutant, although both single mutants have little impact suggesting the two NAPs can compensate for each other's loss. Several evidences suggest that these regions could be relevant for both adaptation to metabolic changes, and that EPOD would maintain some kind of epigenetic

memory over generations, and silencing of genomic elements otherwise detrimental to the cell.

Main concerns:

Comment:
"ePOD patterns are conserved across laboratory evolution": the authors have tested only a single variant strain of MG1655! This is not enough to state the former, they should tune down the claim, just mention they used two 1655 background strains. And, I tend to disagree that the PODs are stable between the two WT, as they represent ~210 and 280 positions in rich medium? To me that looks like a relatively significant difference, it could be interesting to dig in a bit further to double check whether this is really only subtle differences from the EPODs calling, or a subset of genes/regions that behave differently. But given the close phylogenetically relationships between these strains, this is a bit puzzling.

Response:

We changed the language of the title of the section to "laboratory isolates" and removed references to evolution to soften the claims, and included text that briefly discusses these differences within that section. We agree that these differences seem large despite the close relationship between these strains. However, we believe that this is unlikely to be a feature of EPOD calling, as we used specific normalization methods (discussed in the supplementary text) and cut offs based on the WT strain that would account for any subtle differences in calling. To investigate this further, we performed another analysis to look at the differences in EPODs in both WT strains and show that the genetic differences, including the *dgcJ* mutation present in (1) but not (2), exhibit changes in protein occupancy that are in line with expectations due to this genetic variation (added to supplemental figure 1). A detailed description of the genetic differences between these strains, and the corresponding effects on EPOD locations, can be found in the text beginning "Important genetic differences have previously been observed...", which we do not quote here due to its length.

Comment:
The "memory" observation is very intriguing and an exciting observation. To me it is really one of the most interesting aspect of the work (although the use of the word "memory" could be discussed). How long would it last? Would adding additional doubling would diminish it?
I think the authors should validate their hypothesis that Idn proteins are diluted, as they suggest. And I think they could validate their hypothesis by adding only one or two experiments that would bring (hopefully) a lot. For instance, could add a couple of divisions and see if the response increases at 5kDG#2. They could also check whether entering stationary phase eliminates the "memory" by giving cells ample time to reassemble EPODs. Or any experiment that would give time to the cells in the population to do so.

Response:

We appreciate the reviewer's interest in our memory observations, and agree that further experiments to better characterize the memory effects would be useful. To address the key questions here, we performed a series of head-to-head competition experiments where the KDG exposed cells were competed with un-exposed cells either immediately after being taken out of KDG, or at several subsequent timepoints (12, 24, 48, or 72 hours later) having been grown in glucose minimal media in the interim. We found that the memory persisted after 12 hours of growth (sufficient for the cells to undergo eight doublings), but was mostly lost after 24 hours and completely gone after 48 hours in glucose minimal media (Figure 4). It is important to

note that in the IPOD-HR experiments, the EPODs and transcriptional regulatory state of the cells had returned to baseline within six doublings (Figure 3), so it appears that the memory effects persist longer than it takes for the EPODs to fully re-assemble, but not through prolonged stationary phase. We have added substantial additional text to the section titled "Metabolic pressures induce changes in EPODs" discussing these new findings.

Comment:
Could the authors mention how many prophage EPODs they are looking at? 43 or so, according to panel 5C? Have they done a null model using random analysis of regions covering a similar amount of DNA to double-check the variations they observe in all the conditions are significant (not just compared to wt) ?

Response:

We included the number of prophage EPODs (41) in the text and figure legend. We have also performed a permutation test as the Reviewer suggested and included new significance stars in panel 5C to examine the variation observed relative to what would be seen for randomly positioned genomic intervals.

Comment:
The "negative" EPOD observed for B. subtilis result from the z-score computation. The authors actually don't show the negative y axis of the IPOD-HR for all the E. coli panels. They should be consistent, and display the negative scores as well for those. What does a "nEPOD" actually mean, with respect to the sequencing data? Given it seems to correlate with AT rich sequences, could it be a sequencing caveat, where the region would still contain proteins (Rok, etc.) but be less "visible" throughout the experiment? The B. subtilis genome contains several prophages that are highly AT rich. Could the authors double check that the reads align correctly in these regions as well? (it may be in the methods, in which case i missed it). I also think the authors should discuss or be more specific about the regions they are looking at (plot the GC%, for instance; annotate some genes on panel 7C, B; etc.)

Response:

We thank the reviewer for these comments and questions and dug extensively into the data to ensure our confidence in our results. We've included a new supplementary figure (Fig. S8) where we examined the negative signal in *B. subtilis* and *E. coli*. First, we examined whether the negative peaks were a feature of reads aligning incorrectly at particular regions, and found that there was no change in density of input reads in any of the categories (LPODs, EPODs, Background) (**Fig. S8A**). We then looked at the LPODs called in *E. coli* and examined whether there were any trends in sequence composition (AT content), or correlation with known NAPs / TFs (**Fig. S8B**). We did not find any significant correlations between LPODs and any known factors to suggest them being regulatory in nature. We then examined whether LPODs in *E. coli* overlapped any pathways, and found that many were associated with highly expressed genes (**Fig. S8C**). This makes sense due to the data processing where we subtract RNA polymerase signal, indicating that these negative peaks found in *E. coli* are due to high RNA polymerase binding. Therefore, we next investigated whether the process of subtracting RNA polymerase signal impacted the location of LPODs called in both species. We would expect that if LPODs are a feature of RNA polymerase binding, that including RNA polymerase signal would greatly impact the location of LPODs. Indeed, we see that in the case of *B. subtilis*, LPODs vs. LPODs no RNA polymerase ChIP subtraction have a very low jaccard distance, indicating high similarity. While in *E. coli*, we see that including RNA polymerase binding leads to a high

Jaccard distance and low similarity. This supports the hypothesis that LPODs in *E. coli* are largely a feature of high expressed genes and RNA polymerase binding, while in *B. subtilis*, negative binding is a feature of a robust regulatory protein. The most likely explanation for this finding is that *B. subtilits* Rok, and potentially some other protein(s) as well, actually partition away from the interphase (likely into the organic phase) during the IPOD extraction, which gives rise to a negative IPOD-HR signal due to depletion of DNA in Rok-bound regions from the interphase. We thus did not include the negative signal in our *E. coli* plots. We have also included labels for the *B. subtilis* genes and information about the regions displayed.

*ptr*Minor concerns**

Comment:
Figure 1D: state this is a representative region in the legend.

Response:
We have included in the figure legend and in the text that the region is a representative one.

Comment:
Figure 3B: I am not sure about the color labels. The author claim a loss of EPOD in 5KDG, but there is still quite an enrichment in protein occupancy in this condition, no? When is this enrichment large enough to be called an EPOD is a matter of parameters, but The analysis is in z-score, so what are the absolute values of the enrichment? Would it be possible to put the profil in supplementary figures, just to give an idea to the reader? I agree that the result shows a change upon growth in 5KDG, but I'd be interested to see more raw data. The Hns Chip profile could also be plotted along the IPOD-HR occupancy?

Response:
The IPOD-HR occupancy score displayed is an overlay of all conditions. Notably, the two blue conditions are almost identical, while the red is decreased specifically at the start of the operon and binding sites of regulatory factors. We have included the H-NS ChIP profile plotted alongside the occupancy to better visualize the overlap of H-NS binding at this region where loss is observed. We have included a supplemental figure panel (Fig. S5A) that includes the raw data of IPOD vs. input and RNA polymerase ChIP vs. input $\log_{10}$ tracks for readers to get an idea of changes at the operon.

Comment:
I think S4A and B panels should be reduced and included in the main figure 2D, and that the RNA expression in single Hns and StpA mutant should also be included.

Response:
We have moved the panels from S4A and B to figure 2 in the main text.

Comment:
Figure 5A. SieB is the prophage? I don't understand this panel. The EPODs barely seem to vary in the region, and it looks as if there are more proteins spread along the whole region in the hns single mutant. Could the authors improve the figure and the text to be more clear on what they are pointing at?

Response:
We have included color labels to show that the entire region contains annotated prophages. We added text to specifically point to the changes in protein occupancy specifically at the promoter and gene body of sieB, especially at the *sieB* transcription start site.

Comment:
I have a logical issue with the comment "EPODs may be functional units that serve as hotspots for foreign DNA", as the causality links are unclear to me at this stage. I may have missed something though.

Response:
We have previously shown that Tn5 integration frequency is higher in EPODs compared to the rest of the genome (*Freddolino, 2021, PLoS Biology*). We added the following text to that sentence to further address the reviewer's point, which is well taken:

> ...although additional data using different mechanisms of horizontal gene transfer would be required to further investigate the extent to which EPODs show enhanced integration rates, vs. simply forming on horizontally acquired DNA once it has already integrated

Reviewer #2 (Significance (Required)):

I find the study interesting, though sometimes a bit speculative, notably regarding the generalization of some experiments.

The "memory" part which, in my opinion, deserves a couple of additional experiments to be further investigated, is an exciting observation and hypothesis.

The study will be of interest of researchers interested in bacterial chromosomal structure, organization, and gene expression regulation. It should also be of interest to some people working in epigenetics and chromosome regulation in other organisms.

Reviewer #3 (Evidence, reproducibility and clarity (Required)):

This paper describes the role of Nucleoid Associated Proteins (NAPs) in the formation of Extended Protein Occupancy Domains (EPODs). These are features previously described in the E. coli chromosome, but whose function remain unclear. Furthermore, the contribution of individual NAPs in their formation is also unclear. Using the sequencing-based technique of in vivo protein occupancy display at high resolution (iPOD-HR), the authors test the individual contribution of NAPs to the formation of EPODs by deleting them. They then show that loss of EPODs correlate with expression of genes in certain metabolic pathways and of prophages in E. coli, suggesting a role for these protein-DNA features in silencing section of the chromosome. They also apply this approach in the Gramm positive Bacillus subtilis to demonstrate that the EPODs can also be found in phylogenetically distant bacteria. Based on these results, the authors propose that the EPODs are a conserved strategy for gene silencing in bacteria that has features similar to those of heterochromatin in eukaryotes.

Comment:
The paper is thought-provoking and provides a new perspective on the role of NAPs and the regulation of gene expression in bacteria. My main criticism is that the paper often lacks enough details to understand it. The method section is brief, and the paper does not have any diagram describing the main method used, iPOD-HR. This also prevents the understanding of some of the analysis used in the paper.

Other comments:

Response:
We thank the reviewer for their comments about the need for more method descriptions. The definitive methods paper for IPOD-HR has recently been published in PLoS Biology (doi: 10.1371/journal.pbio.3001306). Accompanying that paper is a github repository including detailed descriptions, examples, and source code of the analysis methods used in our work. We believe that these additional resources will substantially aid in understanding the methods being used here.

Comment:
1. The description of the symetrized overlap distance (page 4) is unclear, and could not find any other section where the authors extended this description. Similarly, is unclear what is a relaxed threshold EPOD

Response:
We have included further description and provided resources to better understand the relaxed threshold EPOD (pg. 2) as well as included an example of a relaxed threshold EPOD ("Loose EPOD") in **Figure 1A**. We hope that this further explanation also helps with the understanding of the symmetrized overlap.

Comment:
2. The observation of EPOD-mediated transcription memory is perhaps the most surprising result in the paper (Figure 4). However, the authors only tried one condition in these experiments, leaving multiple unanswered questions. If possible, I suggest to test whether the outcome of this transcriptional memory changes with the length of the period that cells are grown in glucose. Extending these observations would strengthen the transcriptional memory claim. It is also unclear in this section the functional advantage of such transcriptional memory (how frequently is E. coli expected to switch between carbon sources in the wild?)

We appreciate the reviewer's interest in our memory observations, and agree that further experiments to better characterize the memory effects would be useful. To address the key questions here, we performed a series of head-to-head competition experiments where the KDG exposed cells were competed with un-exposed cells either immediately after being taken out of KDG, or at several subsequent timepoints (12, 24, 48, or 72 hours later) having been grown in glucose minimal media in the interim. We found that the memory persisted after 12 hours of growth (sufficient for the cells to undergo eight doublings), but was mostly lost after 24 hours and completely gone after 48 hours in glucose minimal media (Figure 4). It is important to note that in the IPOD-HR experiments, the EPODs and transcriptional regulatory state of the cells had returned to baseline within six doublings (Figure 3), so it appears that the memory effects persist longer than it takes for the EPODs to fully re-assemble, but not through prolonged stationary phase. We have added substantial additional text to the section titled "Metabolic pressures induce changes in EPODs" discussing these new findings, including a direct answer to the reviewer's inquiry regarding the expected switching times for *E. coli* to encounter different carbon sources in the wild.

Response:

Comment:
3. In page 11 (Figure 6) it is unclear why did the authors choose a deletion of cspE as a control

Response:

We chose cspE as a control as it did not interact with any known NAPs that we were testing; we note this in the revised manuscript.

Comment:
4. I suggest not to abuse the use of heterochromatin as analogy, since it only goes so far. Although EPODs are linked to silencing, they contain no nucleosomes or other factors of eukaryotic chromatin. There is obviously also no ancestral relation between heterochromatin and EPODs. Finally, the phenomenon has its own merit as a purely bacterial mechanism. This is why I suggest that authors avoid terms such as 'Heterochromatin domains', which was used in the title of the last section of the results

Response:
We appreciate the reviewer's point here; we believe that the analogy to eukaryotic heterochromatin has value up to a point, but that it should not be used to directly refer to the bacterial phenomenon that we are discussing here. We have re-titled the last section of the Results "Extended protein occupancy domains silence horizontally acquired DNA across diverse species", and ensured that all other occurrences of the term 'heterochromatin' are clearly used only in the context of analogies between EPODs and eukaryotic heterochromatin.

Comment:
5. Unclear sentence: "These findings implicate Hfq, a well documented RNA chaperone and that has only been recently explored as a protein to compress dsDNA" (page 10)

Response:
We have reworded the sentence above to hopefully provide more clarity:

Reviewer #3 (Significance (Required)):

The paper is thought-provoking and provides a new perspective on the role of NAPs and the regulation of gene expression in bacteria. It should be of interest to a broad audience (fields: molecular biology, cell biology, microbiology)

Re: EMBOJ-2021-108708R
Heterochromatin-like domains silence harmful genes and promote transcriptional memory in bacteria

Dear Dr. Freddolino,

Thank you for submitting your revised manuscript which had been reviewed at Review Commons. We sent the revised version back to the initial referees and have now received their reports (please see comments below). All referees appreciate that you have largely addressed their concerns, but still have a few issues that should be resolved in the next, final revision of the manuscript. In particular it will be important to add further information on the IPOD-HR approach as both referee #1 and referee #2 point out. In this revised manuscript, I would also ask you to please address a number of editorial and formatting issues that are listed in detail below.

Please feel free to contact me if you have further questions regarding this revision or any of the specific points listed below. Thank you again for giving us the chance to consider your manuscript for The EMBO Journal.

Kind regards,

Stefanie

Stefanie Boehm
Editor
The EMBO Journal

Referee #1:

In this paper, the authors describe the describes the role of Nucleoid Associated Proteins (NAPs) in the formation of Extended Protein Occupancy Domains (EPODs). These are features previously described in the E. coli chromosome, but whose function remain unclear. Furthermore, the contribution of individual NAPs in their formation is also unclear. Using the sequencing-based technique of in vivo protein occupancy display at high resolution (iPOD-HR), the authors test the individual contribution of NAPs to the formation of EPODs by deleting them. They then show that loss of EPODs correlate with expression of genes in certain metabolic pathways and of prophages in E. coli, suggesting a role for these protein-DNA features in silencing section of the chromosome. They also apply this approach in the Gramm positive Bacillus subtilis to demonstrate that the EPODs can also be found in phylogenetically distant bacteria. Based on these results, the authors propose that the EPODs are a conserved strategy for gene silencing in bacteria that has features similar to those of heterochromatin in eukaryotes.

This is a revision of the original paper. The new version is a clear improvement from the original version. The authors incorporated more text and data to clarify previous criticisms. The paper is interesting and thought provoking and is definitely worth publishing. However, I still have few comments that I think the authors should address before publication.

A first major criticism I have is that the paper still does not have sufficient information to be understood without having to read a technical paper on IPOD-HR. In my view, this is unacceptable. This is even more important since the last section in the paper is really a proof that this technique can be applied to organisms other than E. coli. Resolution of this limitation of the paper requires very little from the authors: minimal information on the technique, such as incorporation of a diagram summarizing the main steps of the protocol in the first figure, or a short paragraph describing the missing steps of the technique in a summarized form, would suffice. The authors can they refer to other papers with more detailed description of the technique.

A second criticism is the still significant degree of speculation in the section on memory. The authors extended this section from the previous version, and it is indeed stronger. But the current data is only suggestive that the EPOD is serving as a memory device. The authors can still provide more convincing or at least discuss alternative interpretations:
• It would help to show evidence (such as in OD readings) of the 8 doublings after transfer from KDG to Glucose. Despite that fewer doublings would represent a significant dilution of proteins, the presence of some key components, such as a KDG transporter, would make a significant impact in the utilization of KDG as carbon source even if present in a copy number of few copies per cell. A similar situation occurs for the metabolism of arabinose, where only cells carrying arabinose transporters, as

result of transcription noise, are able to metabolize this sugar
• Also relating to the experiments on the memory, it is puzzling why the authors do not report sustained changes in the EPODs around the idnDOTR operon after the return to M9-glucose from KDG. Indeed, Figure 3 shows that the IPOD pattern returns to the one observe originally for glucose when shifted from KDG to glucose, but their interpretation of the data in Figure 4 has to do with memory in protein occupancy at that locus
• The suggestion that the product of idnDOTR transcription may be toxic to explain long lag times is tenuous. Global changes in transcription and translation could also explain a long lag time. Indeed, a long lag time is likely observed with other poorly metabolized carbon sources, such as acetate, which may not have and EPOD linked to them. This is something easy to test, but the authors should at least leave this observation open to other explanations.

Minor comments:
• Page 9_ The section starting with "When comparing an existing H-NS ChIP-seq dataset to our observation..." seems to be redundant with respect to the beginning of the paragraph, where the authors also mention the ChIP data supporting H-NS binding to this operon
• Figure 4_ The description of times of growth in the legend of figure 4 is confusing. For example, the authors mention that the KDG adapted cells are incubate for 24hr in glucose, but they also have a timepoint at 12hrs. Some text editing would be helpful

Referee #2:

The revised manuscript addresses most of my comments. I think the authors did a relatively good job at responding to the concerns and suggestions for experiments. They could discuss a bit more on how this work differs from the one they just published in PloS biology, and schematize the IPOD-HR approach either in sup or main figure.

The discussion about the differences of the two wt strain is interesting. It would have been nice if the authors had validated their hypothesis with an experiment, but I understand this may have been a lot of work. That would validate their sentence "we show that IPOD-HR is able to detect changes in global protein occupancy due to genetic and environmental differences." I think at the moment, despite the strong correlation between the IS1 and IS5 insertions in dgcJ and Crl, respectively, the phenotype and the ePODs, it remains to be experimentally demonstrated, unless there are no other mutations in the WT(2) background.

The "memory" experiment brings another clue about the potential interest of EPODS.

I only have a comment about the Panel 1A which is supposed to be a representative EPOD / looseEPOD region: the IPOD-HR occupancy seems still quite high on the right side of the panel (nohA). It would be nice to add another representative region (you can put them side to side in smaller sizes), to have an idea of how fluctuating the data are. I understand that this is what the PLOS biology paper does, though. But still.

Referee #3:

The authors answered most of my previous comments. One important point remains, I appreciated that the authors included an illustration of EPOD calling on figure 1A, however i think it is not informative enough on the EPOD calling procedure. For exemple, I do not see the reason that excludes nohA from EPOD calling and do not see the reason to include part of ydfI in the EPOD. May be this exemple is not the best the authors could chose.

Response to Reviewers

We thank the Reviewers for their consideration of our revised manuscript, and we believe that the additional suggestions that they made have helped us to further strengthen the paper. In the point-by-point response below, we show Referee comments in *italic text*, and our replies in normal text.

Referee #1:

In this paper, the authors describe the describes the role of Nucleoid Associated Proteins (NAPs) in the formation of Extended Protein Occupancy Domains (EPODs). These are features previously described in the E. coli chromosome, but whose function remain unclear. Furthermore, the contribution of individual NAPs in their formation is also unclear. Using the sequencing-based technique of in vivo protein occupancy display at high resolution (iPOD-HR), the authors test the individual contribution of NAPs to the formation of EPODs by deleting them. They then show that loss of EPODs correlate with expression of genes in certain metabolic pathways and of prophages in E. coli, suggesting a role for these protein-DNA features in silencing section of the chromosome. They also apply this approach in the Gramm positive Bacillus subtilis to demonstrate that the EPODs can also be found in phylogenetically distant bacteria. Based on these results, the authors propose that the EPODs are a conserved strategy for gene silencing in bacteria that has features similar to those of heterochromatin in eukaryotes.

This is a revision of the original paper. The new version is a clear improvement from the original version. The authors incorporated more text and data to clarify previous criticisms. The paper is interesting and thought provoking and is definitely worth publishing. However, I still have few comments that I think the authors should address before publication.

We thank the reviewer for their extremely helpful comments, both on the initial submission and the revised version. We address their remaining questions and comments point-by-point below.

A first major criticism I have is that the paper still does not have sufficient information to be understood without having to read a technical paper on IPOD-HR. In my view, this is unacceptable. This is even more important since the last section in the paper is really a proof that this technique can be applied to organisms other than E. coli. Resolution of this limitation of the paper requires very little from the authors: minimal information on the technique, such as incorporation of a diagram summarizing the main steps of the protocol in the first figure, or a short paragraph describing the missing steps of the technique in a summarized form, would suffice. The authors can they refer to other papers with more detailed description of the technique.

We thank the reviewer for this comment, which we agree strengthens the overall ability of the reader to appreciate what has been done and what our results mean. We added a new first paragraph to the results section (beginning with "In order to investigate the contributions...") that both provides a summary of the IPOD-HR method (including a newly added schematic in Appendix Fig. S1).

A second criticism is the still significant degree of speculation in the section on memory. The authors extended this section from the previous version, and it is indeed stronger. But the current data is only suggestive that the EPOD is serving as a memory device. The authors can still provide more convincing or at least discuss alternative interpretations:

We appreciate the reviewer's comments on this point, and we are in complete agreement that at present we are at the beginning, rather than the end, of understanding precisely what is happening to the interplay

between local regulation, EPOD formation, and potential transcriptional memory at *idn*. In addition to direct responses to the detailed points raised by the reviewer below, we have attempted throughout this section of the manuscript to provide a balanced discussion of the various possible explanations for our data, for example including the text "Further investigation of the kinetics of EPOD re-formation after induction is lifted, and indeed the identities of proteins that bind to the idn region immediately after EPOD re-formation, will be required to distinguish between the possible mechanisms noted above…. While additional direct evidence is needed, our findings are consistent with the possibility that the structure of the EPOD in this region is such that transcriptional initiation is faster upon second induction within some time window after an initial induction, providing a transcriptional memory that facilitates responses to repeated stresses." in framing our discussion of the potential memory behavior.

• *It would help to show evidence (such as in OD readings) of the 8 doublings after transfer from KDG to Glucose. Despite that fewer doublings would represent a significant dilution of proteins, the presence of some key components, such as a KDG transporter, would make a significant impact in the utilization of KDG as carbon source even if present in a copy number of few copies per cell. A similar situation occurs for the metabolism of arabinose, where only cells carrying arabinose transporters, as result of transcription noise, are able to metabolize this sugar*

We agree that this is a critical point. Our observation of eight doublings having occurred after transfer from KDG to glucose arises directly from colony counts. We now provide the raw count data underlying the eight-doubling observation in Appendix Table S1.

• *Also relating to the experiments on the memory, it is puzzling why the authors do not report sustained changes in the EPODs around the idnDOTR operon after the return to M9-glucose from KDG. Indeed, Figure 3 shows that the IPOD pattern returns to the one observe originally for glucose when shifted from KDG to glucose, but their interpretation of the data in Figure 4 has to do with memory in protein occupancy at that locus*

We had already speculated that the transcriptional memory likely arose from changes in H-NS filament structure (eg, from unabridged to bridged), or from post-translational modifications, rather than the simple presence vs. absence of H-NS; this argument was present precisely because, as the reviewer noted, simple EPOD occupancy re-forms quickly on the *idn* promoter when induction is removed. We have expanded our discussion on this point to read as follows:

"One possible explanation for these findings would be that some of the gene products induced to metabolize 5KDG may themselves be detrimental under normal growth conditions, and thus another role of the EPOD at the idn promoter may be to ensure tight silencing of these genes until they are needed. It is also possible, however, that some other aspect of global physiological state (e.g., stress-induced changes in transcription or translation), rather than idn gene expression, are responsible for the observed lag."

We also note that, while not shown in the paper, we performed similar media-jump experiments to those shown in the paper on *hha* cells, and found that unlike the wild type cells, the *hha* cells show both weaker direct adaptation to 5KDG and no significant memory:
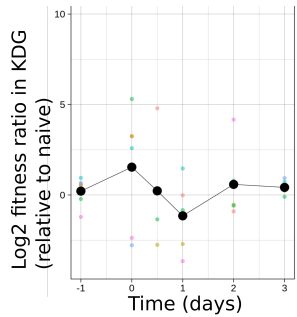
(the data above are for *hha* cells, and are comparable to Fig 4B in the text)

In addition, we found that naive *hha lacZ* cells were more fit than naive WT cells in a head to head competition (delta log2 fold changes of -1.5, -0.5, and -1.9 in three biological replicates), and that *hha* cells did not have as large a marginal drop in fitness in M9+glucose after KDG exposure. All of these preliminary data are consistent with the hypothesis that formation of bridged filaments by Hha is important in effective silencing of the *idn* genes by H-NS; the kinetics of formation of these structures *in vivo* may have an impact on the duration of transcriptional memory. At present, however, such a model is purely speculative; these findings on Hha drive ongoing work in our laboratory tracking the kinetics of re-formation of EPODs after induction, which will likely yield substantial new insight into both the mechanisms and effects of the changes in occupancy during/after induction of *idn* and similar loci. Particularly due to the pleiotropy of *hha* knockouts, we believe it is premature to present our findings on *hha* mutants in the manuscript itself; we provide the data here to provide the reviewer with a fuller picture of what may be occurring.

• The suggestion that the product of idnDOTR transcription may be toxic to explain long lag times is tenuous. Global changes in transcription and translation could also explain a long lag time. Indeed, a long lag time is likely observed with other poorly metabolized carbon sources, such as acetate, which may not have and EPOD linked to them. This is something easy to test, but the authors should at least leave this observation open to other explanations.

We have rephrased this portion of the text to read
"One possible explanation for these findings would be that some of the gene products induced to metabolize 5KDG may themselves be detrimental under normal growth conditions, and thus another role of the EPOD at the *idn* promoter may be to ensure tight silencing of these genes until they are needed. It is also possible, however, that some other aspect of global physiological state (e.g., stress-induced changes in transcription or translation), rather than *idn* gene expression, are responsible for the observed lag."

Minor comments:
• Page 9_ The section starting with "When comparing an existing H-NS ChIP-seq dataset to our observation..." seems to be redundant with respect to the beginning of the paragraph, where the authors also mention the ChIP data supporting H-NS binding to this operon

We have deleted this sentence and changed the section to: "The consistency with the previously reported H-NS binding [20], coupled with our own findings that protein occupancy is reduced

while there is an induction of *idnDOTR* operon genes in the *stpA* and *hns* double mutant (**Fig. EV2B**), further supports our hypothesis that H-NS is a major contributor to the suppression of the *idnDOTR* operon (**Fig. 3B**)."

• Figure 4_ The description of times of growth in the legend of figure 4 is confusing. For example, the authors mention that the KDG adapted cells are incubate for 24hr in glucose, but they also have a timepoint at 12hrs. Some text editing would be helpful

We have clarified this caption, most notably by detailing that the 0.5 day and 1 day timepoints were taken from parallel cultures that had been started after the 0 day timepoint.

Referee #2:

The revised manuscript addresses most of my comments. I think the authors did a relatively good job at responding to the concerns and suggestions for experiments. They could discuss a bit more on how this work differs from the one they just published in PloS biology, and schematize the IPOD-HR approach either in sup or main figure.

We thank the reviewer for their input on this point, as it allows us an opportunity to further highlight the key new insights from the present study. Our recent PLoS Biology paper was a presentation of the IPOD-HR method itself, and then gave its application focusing on genetic knockouts of three local regulators (*lexA*, *argR*, *purR*) and under different physiological conditions. In contrast, the emphasis of the present study is on the large panel of nucleoid associated protein deletions that we considered here (eight new genotypes that were not considered in our prior paper), and on the follow-up experiments that we performed to further elucidate the physiological importance of various types of large scale protein occupancy by NAPs. In the revised manuscript we now highlight these differences more thoroughly in the first paragraph of the Results section (beginning with "In order to investigate the contributions...").

The discussion about the differences of the two wt strain is interesting. It would have been nice if the authors had validated their hypothesis with an experiment, but I understand this may have been a lot of work. That would validate their sentence "we show that IPOD-HR is able to detect changes in global protein occupancy due to genetic and environmental differences." I think at the moment, despite the strong correlation between the IS1 and IS5 insertions in dgcJ and Crl, respectively, the phenotype and the ePODs, it remains to be experimentally demonstrated, unless there are no other mutations in the WT(2) background.

The reviewer raises an important point here. We enumerate all six genetic differences between the strains being considered both earlier in that paragraph and in the Methods section, and in order to acknowledge the ambiguity pointed out by the reviewer, we have added the following text immediately after the dgcJ/crl text: "As noted above, the strains considered here also differ in genotype at gatC, glpR, ybhJ, and mntP; although it appears unlikely based on the genes involved, we cannot rule out contributions from the differences at those loci in the observed changes in EPOD locations."

The "memory" experiment brings another clue about the potential interest of EPODS.

We thank the reviewer for this helpful comment, and we agree wholeheartedly -- this will be a highly fruitful area for future work.

I only have a comment about the Panel 1A which is supposed to be a representative EPOD / looseEPOD region: the IPOD-HR occupancy seems still quite high on the right side of the panel (nohA). It would be nice to add another representative region (you can put them side to side in smaller sizes), to have an idea of how fluctuating the data are. I understand that this is what the PLOS biology paper does, though. But still.

We appreciate the reviewer's comments here, which were shared with Reviewer #3. We have added an additional panel showing another prophage region to more clearly show the distinctions between EPOD and non-EPOD regions. We have also added an additional textual explanation to the Figure 1A caption clarifying why the called EPOD boundaries exist where they do -- notably, that an EPOD can be broken by a drop in the overall rolling average occupancy, or by a single location where occupancy drops below background levels. We also note explicitly why nohA is not called as an EPOD (*i.e.*, that the region of sustained occupancy there is not long enough). We also added to the Methods section a more detailed description of the EPOD calling algorithm (contained in the section "Rescaling of IPOD-HR occupancy tracks and subsequent EPOD calling"). We believe these additions will help put the displayed occupancy in context; the EPOD definition is what it is, and the EPOD calls themselves are made algorithmically based on our definition.

Referee #3:

The authors answered most of my previous comments. One important point remains, I appreciated that the authors included an illustration of EPOD calling on figure 1A, however i think it is not informative enough on the EPOD calling procedure. For exemple, I do not see the reason that excludes nohA from EPOD calling and do not see the reason to include part of ydfl in the EPOD. May be this exemple is not the best the authors could chose.

We appreciate the reviewer's comments here, which were shared with Reviewer #2. We have added an additional panel showing another prophage region to more clearly show the distinctions between EPOD and non-EPOD regions. We have also added an additional textual explanation to the Figure 1A caption clarifying why the called EPOD boundaries exist where they do -- notably, that an EPOD can be broken by a drop in the overall rolling average occupancy, or by a single location where occupancy drops below background levels. We also note explicitly why nohA is not called as an EPOD (*i.e.*, that the region of sustained occupancy there is not long enough). We also added to the Methods section a more detailed description of the EPOD calling algorithm (contained in the section "Rescaling of IPOD-HR occupancy tracks and subsequent EPOD calling"). We believe these additions will help put the displayed occupancy in context; the EPOD definition is what it is, and the EPOD calls themselves are made algorithmically based on our definition.

Thank you again for submitting the final revised version of your manuscript. I am pleased to inform you that we have now formally accepted it for publication in The EMBO Journal.

***

Corresponding Author Name: Peter Freddolino
Journal Submitted to: EMBO Journal (via transfer from Review Commons)
Manuscript Number: EMBOJ-2021-108708

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

➜ the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
➜ figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
➜ graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
➜ if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
➜ Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

➜ a specification of the experimental system investigated (eg cell line, species name).
➜ the assay(s) and method(s) used to carry out the reported observations and measurements
➜ an explicit mention of the biological and chemical entity(ies) that are being measured.
➜ an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

➜ the exact sample size (n) for each experimental group/condition, given as a number, not a range;
➜ a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
➜ a statement of how many times the experiment shown was independently replicated in the laboratory.
➜ definitions of statistical methods and measures:
  • common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  • are tests one-sided or two-sided?
  • are there adjustments for multiple comparisons?
  • exact statistical test results, e.g., P values = x but not P values < x;
  • definition of 'center values' as median or average;
  • definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

B- Statistics and general methods

Please fill out these boxes ⬇ (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	N/A
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	N/A
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	N/A
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	No
For animal studies, include a statement about randomization even if no randomization was used.	N/A
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	No
4.b. For animal studies, include a statement about blinding even if no blinding was done	N/A
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	All frequentist tests were nonparametric and without distributional assumptions; Bayesian model fitting was assessed as described in the text
Is there an estimate of variation within each group of data?	Yes

Is the variance similar between the groups that are being statistically compared?	Yes

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Catalog numbers are provided
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	N/A

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	N/A
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	N/A
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm	N/A

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	N/A
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	N/A
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	N/A
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	N/A
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	N/A
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	N/A
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	N/A

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules	Done (GEO link is provided)
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-	N/A
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited	N/A

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	No