THE
EMBO
JOURNAL

Manuscript EMBO-2012-84297

Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor

Jian Li and David S. Gilmour

Corresponding author: David S. Gilmour, Penn State University

Review timeline:

Submission date:	20 December 2012
Editorial Decision:	29 January 2013
Revision received:	22 April 2013
Accepted:	24 April 2013

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

Editor: Anke Sparmann

1st Editorial Decision 29 January 2013

Thank you for submitting your manuscript (EMBOJ-2012-84297) to our editorial office. Please find enclosed the comments of two of the three reviewers whom we had asked to evaluate your research for The EMBO Journal. Despite several e-mails, I have unfortunately not heard back from the third referee, which explains the delay in reaching a decision. At this stage, I do not think I will still receive her/his comments and have decided to move forward with the two reports on hand.

Both reviewers clearly appreciate your study and are supportive of publication in The EMBO Journal. Nevertheless, referee #1 expresses some concerns that should be addressed based on the reviewer's constructive suggestions. Given the positive comments provided, I would like to invite you to submit a suitably revised manuscript to The EMBO Journal. I should add that it is our policy to allow only a single major round of revision and that it is therefore important to address the raised concerns at this stage.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website:
http://www.nature.com/emboj/about/process.html

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE COMMENTS

Referee #1

This manuscript by Li and Gilmour reports the identification of Motif1-binding protein, a factor that interacts with previously described Motf1 sequences upstream of many Drosophila promoters. The occurrence of Motif1 is a very good predictor of M1BP binding, and the authors suggest that this sequence 'hard-wires' genes for M1BP association and thus Pol II binding and gene activity. In agreement with this, depletion of M1BP by RNAi caused a decreased expression of ~600 genes, ~66% of which appeared to be direct targets. Pol II ChIP confirmed that this resulted from a loss of promoter Pol II levels in the absence of M1BP.

The authors show that genes bound by M1BP are involved in basic cellular processes and show less variability in their expression in different tissues and through development than genes with GAF or TATA sequences. This supports their suggestion that M1BP may direct gene activity in a 'stand-alone' fashion. Also supporting this idea, they find that the presence of GAF and TATA sequences correlates with binding sites for a number of other transcription factors, but that genes with Motif1 tend not to have binding motifs for other factors.

Overall, this manuscript tells a nice story indicating that the Drosophila GAF, TATA and M1BP transcription factors might work in different ways to stimulate transcription of different kinds of genes. Despite the lack of homologs, one would presume that there are mammalian factors that might function similarly to GAF and M1BP, making this a generally applicable story. That said, I do have some concerns about over-reach in specific claims, as outlined below.

Primary concerns:

1) The findings concerning a potential role of the +1 nucleosome in the transient pausing observed at M1BP-associated genes are potentially quite interesting, but would need to be further developed prior to publication. As the manuscript stands, the authors confirm prior reports that the most highly paused genes (apparently those with GAF bound) lack much in the way of a +1 nucleosome, and agree that this long-lived pausing is likely nucleosome independent. They then find that a +1 nucleosome is present at M1BP-associated genes and speculate that this nucleosome may play a role in the lower level of Pol II pausing seen at these genes.
A model wherein a +1 nucleosome causes pausing, even at a subset of genes, is appealing but what the authors show here is simply a co-occurrence of a +1 and a transiently paused Pol II at a subset of genes, without demonstrating a trend, cause/effect, or anything to take this finding beyond mere speculation. Can they further develop this idea to show a trend? For example:
1) Can they show that the M1BP-less, GAF-less, paused genes that display lower +1 nucleosome occupancy have lower pausing indices than those with M1BP and higher levels of +1 nucleosome? This would suggest that the nucleosome is playing an active role in the pausing process and allow them to begin to define a relationship between the level of a +1 nucleosome and the appearance or duration of pausing.
2) More importantly, the authors should show us the nucleosome occupancy of genes that don't undergo pausing of Pol II. If their model holds, then genes with very efficient movement of Pol II into the gene should have much lower levels of the +1 nucleosome. Figure 7 currently lacks any comparison of paused genes with genes that don't show pausing, which severely undercuts the authors' arguments about the role of the +1 nucleosome in 'causing' paused Pol II.
I feel strongly that these straightforward analyses should be included in a revised manuscript.

2) The authors state in the abstract and main text that the majority of TATA box-containing promoters are controlled at PIC formation, but I don't actually see any data supporting this claim. Figure 1C does show that TATA is under-represented at genes in the 2nd and 3rd decile of pausing indices, but it appears to be equally present (not enriched, not excluded) from the most paused genes (top decile pausing index) and the least paused genes (lower half of the deciles). If the authors wish to make a claim such as this, it will need to be much better substantiated. As a start, they could include the distribution of pausing indices of TATA-containing genes in the figure 7A where they give this information for GAF and M1BP associated genes. If TATA genes have statistically

significantly lower pausing indices than genes with the GAF or Motif1 sequences, this would help support this claim.

Minor points:
1) the authors state in the abstract that GAF and Motif1 are 'mutually exclusive on the genome', but this is an overstatement. There are >150 of several thousand genes analyzed that have both motifs, as stated within the body of the text. I know they are working to make the abstract brief, but insertion of the word 'nearly' would make this much more accurate.

2) The authors cite their own unpublished work numerous times in this manuscript where published papers would be more appropriate. For example, Core et al., Cell Reports 2012 shows that PICs are short lived and advance rapidly to a paused state (p.3); GAF and its GAGA motif have been shown previously to be enriched at the most paused genes, including by the Gilmour lab, and Levine group (Hendrix et al., PNAS 2008) as mentioned on p.5.

Referee #2

The manuscript by Li and Gilmour presents a clear cut demonstration that there are two classes of pol II paused promoters/genes. In the first instance they are characterized by being bound by either GAGA or a novel transcription factor that the authors have purified and cloned, M1BP. But these two classes of genes, fascinatingly, differ in many other ways too. They show very different variances in expression levels, nucleosome architectures, GO terms, and likely the sequence specific transcription factors that regulate them. The authors nicely extend studies showing that TATA containing promoters represent a third class of gene by similar criteria.

Wow! This is a real breakthrough. It untangles what was a muddy and complex appreciation that polymerase paused genes were somewhat different from other genes, and now shows dramatic differences between three classes of gene. While M1BP is not conserved, one imagines now that the signatures of these three classes are clear, it should be relatively straightforward to see, in some other publication, if they are also found in other animal phyla, e.g. vertebrates.

While we often for convenience try to divide genes up into different regulatory groupings, usually the differences are not clear cut or compelling. In this case they are. It seems that we are looking at fundamentally different architectures that very much look to be associated with differing regulatory strategies for genes regulated to different degrees. I see this as profound and believe this paper would be as at home in Cell, Science or Nature as the EMBO J.

I truly have no concerns or criticisms. I look forward to seeing this published.

1st Revision - authors' response 22 April 2013

We thank the referees for their supportive evaluation. Since referee #2 had no criticisms, we've only addressed the concerns of referee #1. My responses are in italics.

Referee #1
Primary concerns:

1) The findings concerning a potential role of the +1 nucleosome in the transient pausing observed at M1BP-associated genes are potentially quite interesting, but would need to be further developed prior to publication. As the manuscript stands, the authors confirm prior reports that the most highly paused genes (apparently those with GAF bound) lack much in the way of a +1 nucleosome, and agree that this long-lived pausing is likely nucleosome independent. They then find that a +1 nucleosome is present at M1BP-associated genes and speculate that this nucleosome may play a role in the lower level of Pol II pausing seen at these genes.
A model wherein a +1 nucleosome causes pausing, even at a subset of genes, is appealing but what the authors show here is simply a co-occurrence of a +1 and a transiently paused Pol II at a subset of

genes, without demonstrating a trend, cause/effect, or anything to take this finding beyond mere speculation. Can they further develop this idea to show a trend? For example:
1) Can they show that the M1BP-less, GAF-less, paused genes that display lower +1 nucleosome occupancy have lower pausing indices than those with M1BP and higher levels of +1 nucleosome? This would suggest that the nucleosome is playing an active role in the pausing process and allow them to begin to define a relationship between the level of a +1 nucleosome and the appearance or duration of pausing.
2) More importantly, the authors should show us the nucleosome occupancy of genes that don't undergo pausing of Pol II. If their model holds, then genes with very efficient movement of Pol II into the gene should have much lower levels of the +1 nucleosome. Figure 7 currently lacks any comparison of paused genes with genes that don't show pausing, which severely undercuts the authors' arguments about the role of the +1 nucleosome in 'causing' paused Pol II.
I feel strongly that these straightforward analyses should be included in a revised manuscript.

We have focused on the second point because the M1BP-less, GAF-less paused genes are likely to be a mixture of genes, some similar to M1BP genes and other similar to GAF genes (although being orchestrated by factors other than M1BP or GAF). In a new supplemental figure, Figure S4, we have divided the M1BP associated genes into ones with high efficiency of pausing and ones with low efficiency of pausing. The former group exhibits higher nucleosome occupancy in the +1 position, thus supporting our model that the +1 nucleosome contributes to the efficiency of pausing.

2) The authors state in the abstract and main text that the majority of TATA box-containing promoters are controlled at PIC formation, but I don't actually see any data supporting this claim. Figure 1C does show that TATA is under-represented at genes in the 2nd and 3rd decile of pausing indices, but it appears to be equally present (not enriched, not excluded) from the most paused genes (top decile pausing index) and the least paused genes (lower half of the deciles). If the authors wish to make a claim such as this, it will need to be much better substantiated. As a start, they could include the distribution of pausing indices of TATA-containing genes in the figure 7A where they give this information for GAF and M1BP associated genes. If TATA genes have statistically significantly lower pausing indices than genes with the GAF or Motif1 sequences, this would help support this claim.

Because the majority of TATA genes don't have Pol II levels in the promoter or body of the gene above background, we don't feel that presenting the pausing index of the TATA genes is meaningful. Instead, we calculated the percentages of M1BP, GAF, and TATA genes that have paused Pol II and these numbers are: 91%, 85%, and 26% respectively. Hence, the majority of TATA genes don't associate with Pol II in the Drosophila S2 cells indicating that they will require assembly of a preinitiation complex to become active. This information has been added to the end of the discussion. We also softened our statement at the end of the abstract.

Minor points:
1) the authors state in the abstract that GAF and Motif1 are 'mutually exclusive on the genome', but this is an overstatement. There are >150 of several thousand genes analyzed that have both motifs, as stated within the body of the text. I know they are working to make the abstract brief, but insertion of the word 'nearly' would make this much more accurate.

We have incorporated the word "nearly" into the 4th line of the abstract.

2) The authors cite their own unpublished work numerous times in this manuscript where published papers would be more appropriate. For example, Core et al., Cell Reports 2012 shows that PICs are short lived and advance rapidly to a paused state (p.3); GAF and its GAGA motif have been shown previously to be enriched at the most paused genes, including by the Gilmour lab, and Levine group (Hendrix et al., PNAS 2008) as mentioned on p.5.

We have incorporated these references as recommended by the referee.

