THE EMBO JOURNAL

Manuscript EMBO-2012-81884

In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1 dependent transcription termination

Odil Porrua, Fruzsina Hobor, Jocelyne Boulay, Karel Kubicek, Yves d'Aubenton-Carafa, Rajani Kanth Gudipatti, Richard Stefl and Domenico Libri

Corresponding author: Domenico Libri, Centre de Génétique Moléculaire

Review timeline:

Submission date:	30 April 2012
Editorial Decision:	01 June 2012
Revision received:	04 July 2012
Editorial Decision:	16 July 2012
Revision received:	17 July 2012
Accepted:	26 July 2012

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.)

1st Editorial Decision 01 June 2012

Thank you for submitting your manuscript for consideration by The EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports all three referees are generally positive about your manuscript, but Referee #3 raises concerns about the significant novelty of your findings. However, given the positive recommendations from the other two referees we are willing to consider a revised version of the manuscript, provided that you discuss the strength of your findings relative to the existing literature further. Besides asking for modifications of the manuscript text, all three referees also have technical concerns that you will have to address in an adequate manner, before we can consider further steps towards publication. I would especially like to emphasize the suggestion from referee #2 to model the NMR shifts on Nab3 to visualize terminator motif recognition through the suggested induced-fit mechanism.

Given the referees' positive recommendations, we offer you the opportunity to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance or rejection of your manuscript will therefore depend on the completeness of your responses to the full satisfaction of the referees in this revised version.
Please do not hesitate to contact me if you have questions related to the review process and the requests made by the referees.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Peer-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

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed.

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

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #1

This manuscript describes a new data set that defines Nrd1-dependent terminators in S. cerevisiae. An artificially synthesised pool of random DNA fragments allowed the identification of unbiased sequences that trigger Nrd1-dependent transcription termination. This new data set clearly shows the existence of an Nrd1 terminator super motif that contains both Nrd1 and Nab3 binding sites as well as a newly characterized AT-rich sequence. The authors characterized each critical element of the "super motif" and their impact on the termination. Moreover, a detailed study of Nab3 binding that extends current knowledge of Nab3 binding site specificity is provided. Also it is demonstrated that these termination signals are universal and may be used in both Nrd1 and CPF dependent pathways. In my opinion the manuscript provides new and significant data that merits urgent publication.

A few issues should be addressed before publishication:
1. Clone 78 seems to be mainly Nab3 dependent. Consistently, there is considerable read through transcription in clone 78-INV. Therefore the question arises: Is the AU rich motif functional in the context of an Nrd1 binding site as well? Or does it only work efficiently when preceded by a Nab3 binding site? Can the authors show more results supporting the view that the AU motif isnt only Nab3-specific?
2. A figure summarizing examples of known Nrd1-dependent terminators, which are similarly organized and have AT-rich sequences would strongly improve the manuscript.
3. One minor issue - Different clones are employed in the Fig1 B and C. Why? It would be better to use the same clones to avoid confusion.

Referee #2

Libri and colleagues continue their study of the NNS termination pathway by selecting efficient terminators from randomized sequence. In the process, they gain some new insights into the process. The work is solid and convincing. I have some minor issues to address, but overall I think it's a nice paper that would be appropriate for EMBO Journal.

1. The one figure that seems weak is the NMR analysis. One thing that would help a great deal is to see the shifts shown in parts C and D mapped onto the three dimensional structure. Looking at the data, it's not obvious that this is an example of "induced fit" as stated in the text. The fact that resonances shift upon ligand binding doesn't necessarily mean rearrangements in the protein surface (I believe). The easiest way to convince readers is to show the structure(s), making it obvious whether parts of the protein are moving.

2. One point that could use some emphasis is the finding that the selected NNS terminators can also function as polyA type terminators when placed further downstream. The field is still debating whether the NNS and polyA pathways are variations of the same mechanism or are distinct. This paper echoes experiments from the Buratowski lab (Kim et al. Mol Cell 2006) showing that mutations in termination factors affect one or the other pathway, but not both. I would recommend that the sentence in the abstract stating that the same sequences function in both pathways also make it clear that this is apparently due to recognition by different factors.

3. Although I think their thinking is correct, I think the authors want to be careful about the wording in a couple of statements in the paper:

"artificial CUTs provide a unique opportunity to distinguish between these possibilities since they were selected outside of the complex natural genomic context." I don't think I would agree with this...because this was still an in vivo selection for the strongest terminators, it's very possible that "double" terminators (having sequences for both termination pathways) could emerge as the winners. The same selection pressures that operate at natural terminators could easily be at work in the genetic selection used here.

"Because most of our artificial terminators led to the production of unstable transcripts and were sensitive to the NNS pathway, it is unlikely that CPF-dependent terminators are significantly present in our winning set." If both NNS and CPF termination can both be triggered by a specific terminator at some level, you could still get unstable RNA due to NMD or other cytoplasmic degradation pathways degrading the species that are not degraded by exosome. This statement should be reworded so it's clear you mean the transcripts are Rrp6-sensitive, rather than the broader term "unstable".

4. The discussion could be trimmed a little by removing the sections that simply repeat the results.

Referee #3

Two recent papers have mapped the in vivo binding sites of the Nrd1/Nab3 complex, while previous analyses have addressed motifs involved in sequence-dependent termination. Here the authors take an inventive and complementary approach by isolating and characterizing multiple, artificial in vivo binding sites.
Selected inserts are clearly shown to confer Nrd1p-dependent termination and to generate short, Rrp6-dependent transcripts.

The motif analyses confirm the previously reported Nab3p binding site and, consistent with published data, indicate that the in vitro characterized Nrd1p binding sites are less reliably present. The potential weakness of this analysis is that the actual Nrd1p-Nab3p binding sites are not localized within the 2x 60 nt inserts.

The authors then perform mutational analysis on one of the identified motifs. This analysis suffers from a very common problem, that of trying to generalize conclusions from specific constructs. Previous in vivo mutational analysis on Nrd1p-Nab3p termination used an authentic terminator as a starting point, and it is unclear that starting from an artificial sequence helps the analysis. This is particularly the case when the construct is placed as an mRNA terminator. The signal recognized in this location appear to be at least partially distinct from early termination, but what features are being recognized remains unclear.

The biochemical analyses appear to have well preformed, but will of quite specialist interest.

Overall, the approach is interesting and the findings provide some advance in understanding of the role the Nrd1p-Nab3p complex. However, the authors overstate both the significance and novelty of the results.

Specific points:
1) In some places the description might leave the reader with the impression that the work is more

novel than is the case. The in vivo characterization of sites and motifs required for Nrd1p-Nab3p dependent termination was reported by Carroll et al. (2004) and this could have been referenced and discussed. Similarly, the preferred in vivo binding site for Nab3 was reported to be UCUUG (Woltzka et al. 2011), while Creamer et al. (2011) proposed related but more extended sites, and this could also have been reported and discussed.

2) It is not clear what makes a 5 nt binding site a "supermotif"?

3) The English usage should be addressed.

1st Revision - authors' response 04 July 2012

Response to the referees

We would like to thank the referees for their constructive criticisms that, we believe, contributed to significantly improve the manuscript.

Referee #1

This manuscript describes a new data set that defines Nrd1-dependent terminators in S. cerevisiae. An artificially synthesised pool of random DNA fragments allowed the identification of unbiased sequences that trigger Nrd1-dependent transcription termination. This new data set clearly shows the existence of an Nrd1 terminator super motif that contains both Nrd1 and Nab3 binding sites as well as a newly characterized AT-rich sequence. The authors characterized each critical element of the "super motif" and their impact on the termination. Moreover, a detailed study of Nab3 binding that extends current knowledge of Nab3 binding site specificity is provided. Also it is demonstrated that these termination signals are universal and may be used in both Nrd1 and CPF dependent pathways. In my opinion the manuscript provides new and significant data that merits urgent publication. A few issues should be addressed before publishication:

1. Clone 78 seems to be mainly Nab3 dependent. Consistently, there is considerable read through transcription in clone 78-INV. Therefore the question arises: Is the AU rich motif functional in the context of an Nrd1 binding site as well? Or does it only work efficiently when preceded by a Nab3 binding site? Can the authors show more results supporting the view that the AU motif isnt only Nab3-specific?

The referee is correct. Indeed, we had not shown that the AU-rich motif is functional also when associated to an Nrd1 site. Therefore we made an additional construct in which this motif is mutated in the context of the 78-INV clone. The result, shown in supplementary fig.5, clearly underscores the importance of the AU-rich motif even when associated to an Nrd1 site.

2. A figure summarizing examples of known Nrd1-dependent terminators, which are similarly organized and have AT-rich sequences would strongly improve the manuscript.

Following the suggestion of this referee, a few examples of natural CUTs and snRNAs containing supermotifs and/or AU-rich motifs were added to figure 2C. Remarkably, the very same AU-rich motif present in artificial CUT #78 is also found in 5 out of 10 natural NNS-terminators.

3. One minor issue - Different clones are employed in the Fig1 B and C. Why? It would be better to use the same clones to avoid confusion.

Actually, four of the clones used are the same (#8,9, 14,17), to which we added four for which we only verified the Rrp6p sensitivity and two that were only tested for Nrd1-sensitivity. We could remove the clones that are not present in both panels, but considering the scope of the figure, i.e. to show that a majority of clones of this class are Rrp6-sensitive and Nrd1-sensitive, using the same clones does not appear to be absolutely necessary.

Referee #2

Libri and colleagues continue their study of the NNS termination pathway by selecting efficient terminators from randomized sequence. In the process, they gain some new insights into the process. The work is solid and convincing. I have some minor issues to address, but overall I think it's a nice paper that would be appropriate for EMBO Journal.

1. The one figure that seems weak is the NMR analysis. One thing that would help a great deal is to see the shifts shown in parts C and D mapped onto the three dimensional structure. Looking at the data, it's not obvious that this is an example of "induced fit" as stated in the text. The fact that resonances shift upon ligand binding doesn't necessarily mean rearrangements in the protein surface (I believe). The easiest way to convince readers is to show the structure(s), making it obvious whether parts of the protein are moving.

We agree that the induced fit mode of binding may not be obvious from Figure 6C, and D. Therefore, we included a new Fig. 6E that displays the perturbed residues upon binding to AUCUUGA and AUCUUCA RNAs, mapped on the previously determined structure of the Nab3 RRM (Hobor et al., 2011). We modified the manuscript as follows:

"… Mapping these perturbations on the previously determined structure of the Nab3 RRM (Hobor et al, 2011) show that the binding of UCUUG provides additional changes at the β2-strand and α-helices when compared to the binding of UCUUC (Figure 6E). This strongly indicates that the recognition of the G-containing Nab3 termination site involves an induced fit mechanism, in which α_D helix and the flanking regions are rearranged on the canonical RRM upon the RNA binding."

2. One point that could use some emphasis is the finding that the selected NNS terminators can also function as polyA type terminators when placed further downstream. The field is still debating whether the NNS and polyA pathways are variations of the same mechanism or are distinct. This paper echoes experiments from the Buratowski lab (Kim et al. Mol Cell 2006) showing that mutations in termination factors affect one or the other pathway, but not both. I would recommend that the sentence in the abstract stating that the same sequences function in both pathways also make it clear that this is apparently due to recognition by different factors.

We agree with the referee. We modified the abstract as requested and clarified this point also in the results and discussion. The paper from the Buratowski lab was also cited.

3. Although I think their thinking is correct, I think the authors want to be careful about the wording in a couple of statements in the paper: "artificial CUTs provide a unique opportunity to distinguish between these possibilities since they were selected outside of the complex natural genomic context." I don't think I would agree with this...because this was still an in vivo selection for the strongest terminators, it's very possible that "double" terminators (having sequences for both termination pathways) could emerge as the winners. The same selection pressures that operate at natural terminators could easily be at work in the genetic selection used here.

We partially disagree with the referee. Indeed he/she is correct in stating that, overall, we might have isolated a significant fraction of "double" terminators and that some selection for this might be at work in our genetic system. However, the clone that we used for the mutational analysis (#78) was shown to be largely independent of Rna15p and Hrp1p, and presumably of the CPF pathway. Thus, sequence 78 was selected in the absence of a selective pressure linked to CPF termination and is therefore unlikely to contain "CPF-generated" information. Our point is that for this clone, an NNS-generated information can be used by the CPF. This is very different from what can be evicted from the study of a natural CUT. Indeed, the latter might contain both NNS- and CPF-termination information if the same sequence belongs to two overlapping transcription units with start sites located at different distance from the termination region. We did not intend to state about the general case, but to provide a proof of principle using one model case. We recognize, however, that the

wording we used was misleading, as well as the use of clones for which the CPF-independency was not clearly established. We have clarified this point in the revised version.

"Because most of our artificial terminators led to the production of unstable transcripts and were sensitive to the NNS pathway, it is unlikely that CPF-dependent terminators are significantly present in our winning set." If both NNS and CPF termination can both be triggered by a specific terminator at some level, you could still get unstable RNA due to NMD or other cytoplasmic degradation pathways degrading the species that are not degraded by exosome. This statement should be reworded so it's clear you mean the transcripts are Rrp6-sensitive, rather than the broader term "unstable".

The referee is correct, we have modified the sentence accordingly.

4. The discussion could be trimmed a little by removing the sections that simply repeat the results.

We have shortened the discussion as requested.

Referee #3

Two recent papers have mapped the in vivo binding sites of the Nrd1/Nab3 complex, while previous analyses have addressed motifs involved in sequence-dependent termination. Here the authors take an inventive and complementary approach by isolating and characterizing multiple, artificial in vivo binding sites. Selected inserts are clearly shown to confer Nrd1p-dependent termination and to generate short, Rrp6-dependent transcripts. The motif analyses confirm the previously reported Nab3p binding site and, consistent with published data, indicate that the in vitro characterized Nrd1p binding sites are less reliably present.

The potential weakness of this analysis is that the actual Nrd1p-Nab3p binding sites are not localized within the 2x 60 nt inserts.

We do not understand this criticism, the sites are all contained within the inserts as shown in several figures. The referee might have misunderstood some of our statements.

The authors then perform mutational analysis on one of the identified motifs. This analysis suffers from a very common problem, that of trying to generalize conclusions from specific constructs. Previous in vivo mutational analysis on Nrd1p-Nab3p termination used an authentic terminator as a starting point, and it is unclear that starting from an artificial sequence helps the analysis.

We disagree with the referee. As stated in the manuscript the interest of using artificial CUTs was to generate a homogeneous and large set of NNS terminators and perform a robust statistical analysis based on a reliable background model. As explained in the manuscript, this analysis cannot be easily performed with natural CUTs, due to the strong genomic sequence biases and overlapping selective pressures. From this analysis, we obtained sequence motifs and arrangements of motifs that are found in a large number of sequences. Then we performed a mutational analysis of these motifs on one of the terminators that happened to contain many of them. Therefore it seems to me that we did exactly the contrary of what the referee states. We did not start from a particular case to generalize to the whole. Rather, we started from general, statistical findings and verified them on a representative model case.

This is particularly the case when the construct is placed as an mRNA terminator. The signal recognized in this location appear to be at least partially distinct from early termination, but what features are being recognized remains unclear.

This could be a misunderstanding. In Fig. 7 we show that almost all the mutations that impair NNS termination, also impair CPF (mRNA) termination, indicating that the signals are largely overlapping, not “at least partially distinct” (see also the comments of the other referees). The signals required for CPF termination in these sequences are obviously defined by the mutational

analysis (i.e. the AU-rich region, the Nrd1 and Nab3 sites), although the factors that recognize these signals are different in early or late termination. These points are emphasized in the revised version of the paper, according to the request of referee #2.

The biochemical analyses appear to have well preformed, but will of quite specialist interest. Overall, the approach is interesting and the findings provide some advance in understanding of the role the Nrd1p-Nab3p complex. However, the authors overstate both the significance and novelty of the results.

We understand the concern of the referee, but we strongly disagree. Our work was not aimed at re-discovering the Nrd1 and Nab3 sites, but to contribute to the understanding of how the limited informational content of these sites leads to transcription termination. Whenever our findings were not novel, we cited the appropriate papers: a few citations were overlooked by this referee (see below) but were present in the original paper. In our opinion the following points underscore the novelty and significance of our work:

1) We define AU-rich motifs as bona fide NNS termination signals. These motifs appear to be as important as Nab3 sites and likely more important than Nrd1 sites for termination. Sparse mutations outside of canonical Nrd1 or Nab3 sites that impair termination (including in AU-rich regions) had been previously reported (cited in the paper), but the role of these sequences had not been studied. Note that, due to the strong AT bias in the yeast genome and in intergenic regions, these motifs escape detection if a set of genomic CUTs is used for the analysis.
2) We exclude, in the context of NNS termination, the role of Hrp1p and Rna15p in recognizing these sequences. Rather, we show that AU-rich sequences contribute to the binding of the Nrd1-Nab3 heterodimer to the RNA.
3) We show here that NNS terminators are more complex than thought and that the mere presence of the sites, or the strength of the interaction with the Nrd1p-Nab3p heterodimer alone cannot define a terminator. Notably, we show that the association with AU-rich motifs in supermotifs is important for termination.
4) We show that the G extension of the Nab3 site is important for termination and even essential depending on the context.
5) This nucleotide is also a major determinant of affinity for the interaction with Nab3p. We show that its recognition occurs specifically via an induced fit mechanism.
6) We describe the implication of a novel domain of Nab3 in the specific recognition of the G nucleotide extension
7) Our experiments strongly suggest that the two major termination pathways recognize largely overlapping signals. This is particularly remarkable in the light of the fact that these signals are recognized by different factors.

Specific points:

1) In some places the description might leave the reader with the impression that the work is more novel than is the case. The in vivo characterization of sites and motifs required for Nrd1p-Nab3p dependent termination was reported by Carroll et al. (2004) and this could have been referenced and discussed.

We had referenced and discussed a more recent paper from the Corden's lab concerning the sequence of the sites and their recognition by the Nrd1-Nab3 heterodimer (Carroll, 2007) together with additional papers: these references were appropriately cited when we discussed the binding of Nrd1-Nab3 to the RNA, which was not addressed in the 2004 Carrol paper. Nevertheless, we have now also included the Carrol (2004) reference as requested.

Similarly, the preferred in vivo binding site for Nab3 was reported to be UCUUG (Woltzka et al. 2011), while Creamer et al. (2011) proposed related but more extended sites, and this could also have been reported and discussed.

The referee might have overlooked that we reported and discussed both papers in the original version. Here is an example, from the discussion: *"The presence of a G extending the Nab3 site was recently noticed in experiments detecting RNAs associated with the NNS complex (Hogan et al, 2008) and in more recent crosslinking approaches (Creamer et al, 2011; Wlotzka et al, 2011)."* And

a bit later: *"These structural changes impact the strength and the geometry of the interaction and might explain why CUUG is strongly preferred as a Nab3p crosslinking site relative to UCUU (Wlotzka et al, 2011)".* We extended the discussion of the extended sites described in Creamer et al. and provided a plausible explanation for the differences

2) It is not clear what makes a 5 nt binding site a "supermotif"?

We think the referee misunderstood our statements: we propose that "supermotifs" are defined by the association of a Nrd1 or Nab3 site and an AU-rich region, not the extended Nab3 or Nrd1 sites. Here is the original sentence from the results section*:" Besides extended AU-rich regions, we observed a significant association between the UCUUG or the GUAA sequences and AU-motifs, suggesting that proximity between these sites is functionally relevant within larger termination supermotifs*". And a bit later: *"These findings confirm the importance of the supermotifs containing canonical NNS complex binding sites associated with AU-rich regions."* Similar statements were also reported in the discussion.

3) The English usage should be addressed.

We revised the English usage.

2nd Editorial Decision 16 July 2012

Thank you for submitting a revised version of your manuscript. It has now been seen by one of the original referees and this person finds that all issues raised by the referees have been sufficiently addressed. However, there are a few technical issues concerning text and figures that I need you to address before we can officially accept the manuscript for publication.

The figure legends for fig 5B and 6A and B should state what the error bars indicate and how many experimental replicas they were based on.
For the supplementary figures, the gel image in fig S1C appears to contain lanes that have been cropped and pasted together for clarity. The inclusion of this data is not a problem per se but all cropped lanes need to be framed to indicate that the image is compiled from lanes that were not flanking in the original gel. It should be stated in the figure legend that the lanes were cropped from the same gel, and we would also ask you to provide the original gel image as source data.

We now generally encourage the publication of source data, particularly for electrophoretic gels and blots, with the aim of making primary data more accessible and transparent to the reader. Would you be willing to provide files comprising the original, uncropped and unprocessed scans of all gels used in the figures? We would need 1 file per figure (which can be a composite of source data from several panels) in jpg, gif or PDF format, uploaded as "Source data files". The gels should be labelled with the appropriate figure/panel number, and should have molecular weight markers; further annotation would clearly be useful but is not essential. These files will be published online with the article as a supplementary "Source Data". Please let me know if you have any questions about this policy.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to your revision.

Sincerely yours,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #2

The authors have nicely addressed my comments and those of the other reviewers. I believe the paper is now ready for publication in EMBO.

2nd Revision - authors' response 17 July 2012

Thanks for your email and for taking care of our paper - we have submitted a revised version of the manuscript , including source files for figures S1C, 1 and 3, and introduced the text changes as requested.

