Hpr1 Is Preferentially Required for Transcription of Either Long or G+C-Rich DNA Sequences in *Saccharomyces cerevisiae*

SEBASTIÁN CHÁVEZ, MARÍA GARCÍA-RUBIO, FÉLIX PRADO, AND ANDRÉS AGUILERA*

Departamento de Genética, Facultad de Biología, Universidad de Sevilla, 41012 Seville, Spain

Received 21 May 2001/Returned for modification 26 June 2001/Accepted 24 July 2001

Hpr1 forms, together with Tho2, Mr1, and Thp2, the THO complex, which controls transcription elongation and genome stability in *Saccharomyces cerevisiae*. Mutations in genes encoding the THO complex confer strong transcription-impairment and hyperrecombination phenotypes in the bacterial lacZ gene. In this work we demonstrate that Hpr1 is a factor required for transcription of long as well as G+C-rich DNA sequences. Using different lacZ segments fused to the GAL1 promoter, we show that the negative effect of lacZ sequences on transcription depends on their distance from the promoter. In parallel, we show that transcription of either a long LYS2 fragment or the *S. cerevisiae* YAT1 G+C-rich open reading frame fused to the GAL1 promoter is severely impaired in hpr1 mutants, whereas transcription of LAC4, the Kluyveromyces lactis ortholog of lacZ but with a lower G+C content, is only slightly affected. The hyperrecombination behavior of the DNA sequences studied is consistent with the transcriptional defects observed in hpr1 cells. These results indicate that both length and G+C content are important elements influencing transcription in vivo. We discuss their relevance for the understanding of the functional role of Hpr1 and, by extension, the THO complex.

The control of genome stability is essential to ensure maintenance of genetic information in all cells of a living organism. Dysfunction of this control causes mutations and chromosomal aberrations that can give rise to loss of gene function, cell death, or irreversible changes in the cell program.

Genetic recombination is required for mitotic DNA repair and for proper meiotic chromosome segregation. In addition, it may also be responsible for processes of genetic instability. A number of animal diseases, including cancer, originate by events of mitotic recombination between repeats that lead to chromosomal aberrations (34). Several elements have been described to enhance mitotic recombination, including DNA damage, replication defects, alteration of chromatin structure, and transcriptional activity (reviewed in reference 3). Ikeda and Matsumoto (26) first described the influence of transcription on recombination showing that recombination of phage λ was stimulated by transcription. In yeast, the first example of transcription-associated recombination was the finding that a hotspot of ribosomal DNA (rDNA) recombination, *HOT1*, was dependent on RNA polymerase I-driven transcription (55, 60). Thomas and Rothstein (56) extended transcription-induced recombination to sequences transcribed by RNA polymerase II (RNAPII). Additional examples of RNAPII-dependent recombination have been subsequently described in yeast (21, 36, 50) and mammalian cells (37, 57). Special mention must be made of the modulation of recombination at the immunoglobulin loci, as both V(D)J recombination (7, 31, 38) and class switching (15) are positively controlled by transcription. Switching (15) are positively controlled by transcription.

Copyright © 2001, American Society for Microbiology. All Rights Reserved.
As relevance of the THO complex in transcription depends on the transcribed DNA sequence, investigation of the features that make transcription of a particular DNA sequence dependent on Hpr1 can provide some clues to understanding its precise function. We have found that transcriptional impairment in hpr1 occurs primarily in long transcription units as well as in DNA sequences with a high G+C content fused to the GAL1 promoter. The relevance of these results for understanding the functional role of the THO complex is discussed.

MATERIALS AND METHODS

Yeast strains and plasmids. The two isogenic yeast strains used in this study were W303-1A (MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1) and U768-4C (MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 hpr1Δ: HIS5). All plasmids used are monocyte CEN-based plasmids and are listed in Table 1.

Analysis of gene expression and recombination. For the analysis of GAL1-driven expression, mid-log phase cells were inoculated with 3% glycerol—driven expression, mid-log phase cells were inoculated with 3% glycerol

As relevance of the THO complex in transcription depends on the transcribed DNA sequence, investigation of the features that make transcription of a particular DNA sequence dependent on Hpr1 can provide some clues to understanding its precise function. We have found that transcriptional impairment in hpr1 occurs primarily in long transcription units as well as in DNA sequences with a high G+C content fused to the GAL1 promoter. The relevance of these results for understanding the functional role of the THO complex is discussed.

RESULTS

Transcription impairment through lacZ caused by hpr1Δ is not dependent on particular lacZ sequences but on their distance from the promoter. Transcription of the Escherichia coli lacZ gene in S. cerevisiae is severely impaired in hpr1 mutants

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRS416</td>
<td>YCp vector based on the URA3 gene</td>
<td>54</td>
</tr>
<tr>
<td>p16GAL1lacZ</td>
<td>pRS416 containing the lacZ region fused to the GAL1 promoter</td>
<td>35</td>
</tr>
<tr>
<td>pSch202</td>
<td>pRS416 containing the PHO5 region fused to the GAL1 promoter</td>
<td>11</td>
</tr>
<tr>
<td>pSch212</td>
<td>pSch202 with lacZ transcriptionally fused to the 3’ end UTR of PHO5</td>
<td>11</td>
</tr>
<tr>
<td>pRS314-L</td>
<td>YCp vector pRS314, based on the TRP1 gene and containing two 598-bp LEU2 internal fragments</td>
<td>43</td>
</tr>
<tr>
<td>pSch211</td>
<td>pSch212 with lacZ in opposite orientation</td>
<td>This study</td>
</tr>
<tr>
<td>pSch212A17-1</td>
<td>pSch202 with a 0.4-kb fragment of the 3’ end of lacZ fused to the 3’ end UTR of PHO5</td>
<td>This study</td>
</tr>
<tr>
<td>pSch229</td>
<td>pSch202 with the first 439 bp of lacZ ORF fused to the 3’ end UTR of PHO5</td>
<td>This study</td>
</tr>
<tr>
<td>pSch226</td>
<td>pSch202 with the HpaI-BstIII 447-bp fragment of lacZ inserted in the 3’ end UTR of PHO5</td>
<td>This study</td>
</tr>
<tr>
<td>pSch251</td>
<td>pSch202 with a 393-bp fragment of GAL1 (from position 843 to 1236 of the ORF) inserted in the 3’ end UTR of PHO5</td>
<td>This study</td>
</tr>
<tr>
<td>pSch215</td>
<td>pRS416 containing a 439-bp fragment of the 5’ end of lacZ fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch213</td>
<td>pRS416 containing the 447-bp HpaI-BstIII fragment of lacZ fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch216</td>
<td>pRS416 containing the 352-bp PvuII-EcoRI fragment of lacZ fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch218</td>
<td>pSch202 containing the first 439-bp fragment of the 5’ end of lacZ inserted between the GAL1 promoter and the PHO5 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>pSch219</td>
<td>pSch202 containing the 447-bp HpaI-BstIII fragment of lacZ between the GAL1 promoter and the PHO5 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>pSch220</td>
<td>pSch202 containing the 352-bp PvuII-EcoRI fragment of lacZ inserted between the GAL1 promoter and the PHO5 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>pSch221</td>
<td>pRS314-L containing the first 439 bp of lacZ ORF inserted between the repeats</td>
<td>This study</td>
</tr>
<tr>
<td>pSch222</td>
<td>pSch214 containing the 447-bp HpaI-BstIII fragment of lacZ inserted between the repeats</td>
<td>This study</td>
</tr>
<tr>
<td>pSch223</td>
<td>pSch314-L containing the 352-bp PvuII-EcoRI fragment of lacZ inserted between the repeats</td>
<td>This study</td>
</tr>
<tr>
<td>pSch205</td>
<td>pSch314-L containing the entire lacZ gene inserted between the repeats</td>
<td>11</td>
</tr>
<tr>
<td>pSch227</td>
<td>pRS416 containing the 3.7-kbp EcoRV fragment of LYS2 fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch230</td>
<td>pSch314-L containing the 3.7-kbp EcoRV fragment of LYS2 inserted between the repeats</td>
<td>This study</td>
</tr>
<tr>
<td>pSch255</td>
<td>pRS416 containing the entire LAC4 coding region fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch264</td>
<td>pSch314-L containing the entire LAC4 coding region inserted between the repeats</td>
<td>This study</td>
</tr>
<tr>
<td>pSch247</td>
<td>pRS416 containing the entire YAT1 coding region fused to the GAL1 promoter</td>
<td>This study</td>
</tr>
<tr>
<td>pSch248</td>
<td>pSch314-L containing the entire YAT1 coding region inserted between the repeats</td>
<td>This study</td>
</tr>
</tbody>
</table>

a UTR, untranslated region.

Mapping of MNase cleavage sites. Yeast spheroplasts and micrococcal nuclease (MNase) digestions were performed according to Fedor and Kornberg (19) with the modifications of Chávez et al. (13). Spheroplasts prepared from mid-log phase cultures transformed with p16GAL1lacZ and grown in the appropriate selective medium containing 2% glucose or 2% galactose were lysed and immediately digested with 6.25 to 800 M of MNase. For naked DNA controls, genomic DNA was extracted as previously described (28) and digested with 0.003 to 1.6 M of MNase under the same conditions.

MNase-cleaved genomic DNA was digested with EcoRI (for the endogenous GAL1 gene) or Clal (for the GAL1::lacZ fusion) and resolved in 1.5% agarose. As internal size markers, we used genomic DNA digested with SacI or XbaI (for the GAL1 promoter fused to lacZ). For the analysis of the endogenous GAL1 gene, the probe used was the 196-bp GAL1 fragment located immediately downstream of the EcoRI site and obtained by PCR with the oligonucleotides ATTCGACAGGTATACGAC and TTAACACTTCACGTCC. For the analysis of GAL1::lacZ the probe used was the 202-bp lacZ fragment immediately upstream of the Clal site and obtained by PCR with the oligonucleotides TCTGCCGACGCTATCG and TGAAGAGCCAGTGGGC. Miscellaneous. Serial deletions of the PHO5::lacZ fusion constructs were constructed using a double-stranded nested deletion kit from Amersham Pharmacia. Published methods were used for RNA and DNA hybridizations (13, 44).
at the elongation level. Transcription through PHO5 is not appreciably affected in hpr1 cells, but it becomes sensitive to hpr1 when lacZ is fused to PHO5 in a single transcription unit (11). In order to find out which structural elements or sequence motifs present in lacZ are responsible for this transcriptional elongation impairment, we constructed serial deletions of the lacZ gene in the PHO5::lacZ transcriptional fusion under control of the GAL1-regulated promoter. The resulting deletions were introduced into both wild-type and hpr1 cells. Yeast transformants grown in galactose-containing medium were then assayed for acid phosphatase activity. Expression was clearly lower in the transformants harboring PHO5-lacZ fusions than in those containing only PHO5, in both wild-type and hpr1 cells (Fig. 1A). The presence of a lacZ fragment as short as 1 kb downstream of PHO5 reduced PHO5 expression to 40% in wild-type cells. However, the reduction was considerably stronger in hpr1, reaching transcription values below 10% of those of PHO5 alone (Fig. 1A). Even the shortest fusion, encompassing \( \sim 0.4 \) kb of the 5' end of lacZ, showed reduced levels of phosphatase activity in the wild-type (65%) and, to a greater degree, hpr1 (35%) cells (Fig. 1A).

Serial deletions were also made in a PHO5::lacZ fusion carrying lacZ in an opposite orientation. A similar profile of phosphatase activities was obtained (Fig. 1B). Although all fusions showed lower expression levels than PHO5 alone, the negative transcriptional effect was clearly stronger in hpr1 than in wild-type cells. A \( \sim 0.4 \)-kb lacZ fragment, for example, was enough to reduce the phosphatase activity to under 15% of the level shown by PHO5 alone in hpr1 strains (Fig. 1B). Thus, the two end fragments of lacZ were able to impair transcription in hpr1 cells.

The previous results suggest that there is not a particular lacZ sequence responsible for the transcriptional elongation impairment caused by hpr1Δ. On the contrary, transcriptional impairment could occur through any lacZ region. To confirm this and to show that the negative effect of hpr1Δ on acid phosphatase expression really takes place at the transcriptional rather than posttranscriptional level, we performed Northern analyses of selected PHO5::lacZ-fragment constructs. We inserted three different \( \sim 0.4 \)-kb fragments of lacZ corresponding to the two ends and the center of the gene (plasmids pSCh229, pSCh226, and pSCh251; Table 1) immediately downstream of a PHO5 gene under GAL1 control. Galactose-induced transcription of the resulting fusion constructs was analyzed in wild-type and hpr1 cells. The results shown in Fig. 2 indicate a substantial decrease in the accumulation of full-length mRNA of the three fusion constructs in hpr1 cells (13 to 25% of the wild-type levels), whereas only a minor effect was observed with PHO5 alone. In addition to the full-length PHO5-lacZ mRNA, a shorter transcript exhibiting the same size as PHO5 was detected in hpr1. The presence of this shorter transcript suggests that hpr1 cells transcribe poorly through lacZ sequences, downstream of the PHO5 open reading frame (ORF). The same results were obtained when the lacZ fragments were located in the opposite orientation (data not shown). To confirm that this phenomenon was due to lacZ itself and not to the 3' end of PHO5, we replaced the lacZ segment with a 416-bp fragment of the 5' end of GAL1, a gene whose expression is not affected in hpr1 cells (67) (see Fig. 9A). Northern analysis of the resulting PHO5::GAL1-lacZ fusion was carried out in wild-type and hpr1 cells (Fig. 2). A weaker reduction in accumulation of full-length mRNA was measured in hpr1 (60% of wild-type levels), and no short transcript was detected. This confirms that the transcriptional defects of the PHO5-lacZ fusion constructs were mainly due to the presence of lacZ fragments in the transcription units.

Altogether, these results suggest that the longer distance between the promoter and lacZ sequences the greater the transcriptional elongation impairment. To test this possibility, we inserted upstream of PHO5 the same three lacZ fragments.
used previously. The resulting transcriptional fusions exhibited similar transcription levels and patterns in wild-type and hpr1 cells (Fig. 3). Therefore, the distance between lacZ and the promoter can modulate the negative effect of the lacZ fragments on transcription in hpr1 cells.

Transcription through long DNA sequences is negatively affected by hpr1Δ. The data shown in Fig. 1 indicate that the longer the constructs containing lacZ fragments are, the lower the transcriptional yield exhibited in both wild-type and hpr1 cells. To evaluate the influence of transcript length on the transcriptional effect of hpr1, we put the same three above-mentioned lacZ fragments immediately downstream of the GAL1 promoter. The kinetics of accumulation of these three short lacZ fragments was identical in wild-type and hpr1 cells (Fig. 4A). Full-length mRNA accumulated shortly after induction in both strains and at similar levels in all three constructs, in contrast with the clear difference between the wild type and hpr1 shown by the entire lacZ (Fig. 4A). The same results were obtained when the lacZ fragments were cloned in the opposite orientation (data not shown). These results strongly suggest that short transcription units are not impaired by hpr1, even if they contain DNA fragments that hinder transcription in a different context.

We have previously shown that in hpr1 and other mutants affected in the THO complex, the ability of a given DNA segment, like lacZ, to impair transcriptional elongation correlates with hyperrecombination of a direct-repeat system containing that segment. This hyperrecombination is transcription dependent (11, 42, 44). We tested, therefore, the recombination frequency of direct-repeat systems containing either one of the three lacZ fragments used in the previous experiments flanked by two leu2 repeats. In agreement with the absence of effect of the hpr1 mutation on transcription of such lacZ fragments, we did not detect a significant stimulation of recombination when the fragments were located between the leu2 repeats (Fig. 4B). Again, in this case there was a clear difference between the short lacZ fragments and the entire lacZ, which stimulates recombination between direct repeats up to 200-fold in hpr1 cells (11) (Fig. 4B).

The previous results suggest that transcript length is an important feature in determining the requirement of Hpr1 in transcription. To confirm this, we constructed comparable transcription and recombination systems containing only yeast long DNA sequences. We randomly chose a fragment from a long yeast ORF, a 3.7-kb fragment of the S. cerevisiae LYS2 gene. We either fused it to the GAL1 promoter or inserted it...
between the leu2 repeats for transcriptional and recombinational analyses, respectively. The new constructs were introduced in wild-type and hpr1 cells (Fig. 5). Full-length mRNA from the GAL1pr::lacZΔ::PHO5 fusion constructs in wild-type (W303-1A) and hpr1 (U768-4C) cells. (A) Northern blot analyses of PHO5-containing mRNAs driven from the GAL1 promoter. Plasmids used were pSc1218, pSc1220, and pSc1219 (carrying an 0.4 kb of the S′ end, middle part, and S′ end of lacZ fused to PHO5, respectively). As a control we used pSc1202 (carrying the PHO5 gene; data not shown), which gave identical results as those shown in Fig. 2. The transcript levels of each construct with respect to PHO5 were similar to those of the wild type shown in Fig. 2. Other details were as described for Fig. 2. (B) Quantification of Northern analyses.

FIG. 3. Transcription analysis of three different GAL1pr::lacZΔ::PHO5 fusion constructs in wild-type (W303-1A) and hpr1 (U768-4C) cells. (A) Northern blot analyses of PHO5-containing mRNAs driven from the GAL1 promoter. Plasmids used were pSc1218, pSc1220, and pSc1219 (carrying an 0.4 kb of the S′ end, middle part, and S′ end of lacZ fused to PHO5, respectively). As a control we used pSc1202 (carrying the PHO5 gene; data not shown), which gave identical results as those shown in Fig. 2. The transcript levels of each construct with respect to PHO5 were similar to those of the wild type shown in Fig. 2. Other details were as described for Fig. 2. (B) Quantification of Northern analyses.

The most evident difference between lacZ and the bulk of S. cerevisiae genes is the G+C content. The majority of yeast genes show a G+C content of around 40%, whereas that of lacZ is 56.2%. In order to investigate whether the G+C content influences the transcriptional impairment of lacZ in hpr1 cells, we used the Kluyveromyces lactis LAC4 gene, a yeast homologue of lacZ with 40% G+C (1). We placed LAC4 under GAL1 control, creating a GAL1pr::LAC4 fusion similar to those previously used in this work. Transformants of wild-type and hpr1 isogenic strains were used to determine the kinetics of accumulation of mRNA. The results presented in Fig. 7A show that accumulation of LAC4 full-length mRNA was only moderately diminished in hpr1 cells (50% of the wild-type level after 90 min of induction). Nonetheless, in the same background and after an identical induction time, lacZ full-length mRNA was almost absent (Fig. 4A) (11). The overall comparison of LAC4 and lacZ transcriptional behaviors showed that transcription through LAC4 was at least fivefold more efficient than that through lacZ in hpr1 cells. Thus, two transcription units, identical in length and differing in G+C content, were differentially affected by hpr1. This indicates that, in addition to transcript length, the G+C content of a DNA sequence may be an important feature influencing transcriptional impairment by hpr1.

To determine the effect of LAC4 in recombination, we placed the entire LAC4 ORF between the leu2 direct repeats.
The resulting L-LAC4 system exhibited a high frequency of recombination in hpr1 cells (90-fold above wild-type levels [Fig. 7B]). This frequency was eightfold lower than that shown by L-lacZ but was comparable to the frequency reached by L-LYS2 (Fig. 3B and 4B). An increase in transcription efficiency is accompanied therefore by a lower recombination frequency. It is important that the L-LAC4 system shows a hyperrecombination phenotype, because the size of the full transcript in this system is approximately 5 kb. Indeed, and consistent with our hypothesis, we have shown by Northern analysis (Fig. 7C) that transcription of the L-LAC4 system is impaired in hpr1 cells.

To further investigate the influence of G+C content on transcription of a DNA sequence, regardless of whether coming from bacteria or yeast, we decided to analyze a yeast gene with a G+C content comparable to that of lacZ. YAT1, a 2-kb long ORF, is the gene in the S. cerevisiae genome with the highest G+C content (58%). We placed YAT1 under GAL1 control in a transcriptional system similar to those used in the previous experiments. Northern analysis showed that high levels of YAT1 full-length mRNA were reached after galactose induction in the wild type, but only a minimal accumulation was detected in hpr1 cells (Fig. 8A). In agreement with this transcriptional impairment, a recombination system bearing YAT1 as intervening sequence (L-YAT1) displayed an extremely high frequency of recombination in hpr1 (12%) that was 173 times the frequency reached in the wild type (Fig. 8B).
This level of hyperrecombination is comparable to that of L-lacZ (Fig. 4B) and higher than the levels of L-L-lacZ. This level of hyperrecombination is comparable to that of CDC48, EGT2, C-rich gene content can modulate the Hpr1 dependency of gene transcription. Nucleosome positioning is only limited to the GAL1 promoter (Fig. 9C). Indeed, a lack of nucleosome positioning is also observed over the bacterial sequences upstream of the GAL1 promoter, through which transcription has also been shown to be impaired in hpr1 mutants (44). Our results, therefore, show that lacZ adopts a random nucleosomal organization in yeast and that hpr1 has no effect on nucleosome positioning.

**DISCUSSION**

In this work we have investigated why transcription of DNA sequences like E. coli lacZ is especially sensitive to hpr1. We have shown that 0.4-kb lacZ fragments fused to PHO5 under the GAL1 promoter are sufficient to increase the Hpr1 dependency of transcription, but not when they are transcribed alone. Such an effect is position dependent: the longer the distance between the lacZ sequence and the GAL1 promoter, the stronger the impairment of transcription caused by hpr1. In addition, we see transcription of long yeast DNA sequences like LYS2 fused to the GAL1 promoter is negatively affected by hpr1. We have also shown that transcription of K. lactis LAC4, a eukaryotic homologue of lacZ that is equal in length but with a much lower G+C content, exhibits a milder Hpr1 dependency in S. cerevisiae, whereas YAT1, an S. cerevisiae G+C-rich gene shorter than lacZ, is dramatically affected by hpr1. Taken together, our results indicate that transcription of DNA sequences such as lacZ is mainly affected by the Hpr1 gene, and that the presence of a C-rich sequence in these DNA sequences is sufficient to reduce transcription levels.

FIG. 6. Transcription analyses of five yeast endogenous genes, EGT2, CDC48, KAR2, OLE1, and GOG5, having high levels of expression and different transcript sizes, in wild-type and hpr1 cells. Total RNA was isolated from mid-log phase cells, grown in YEPD broth, and used for Northern analyses. Internal fragments of each gene and of the 23S rDNA, obtained by PCR, were used as DNA probes. The hpr1: wild-type transcript ratio was obtained from the mRNA levels that were quantified in a Fuji FLA3000 and normalized with respect to the rRNA levels.

**FIG. 7.** Transcription and recombination analyses of LAC4 in wild-type and hpr1 cells. (A) Northern analyses of LAC4 mRNAs in strains transformed with the plasmid pSCh255, which contains the entire LAC4 coding sequence under the control of the GAL1 promoter. (B) Recombination frequencies of cells transformed with plasmid pSCh254, which harbors the leu2-based direct-repeat L-LAC4 construct containing LAC4 as the intervening region. (C) Northern analyses of the L-LAC4 repeat construct in wild-type and hpr1 cells. Other details are as described for Fig. 4.
together, these results indicate that both length and G+C content are important elements influencing gene transcription in vivo and that Hpr1 is an important factor controlling transcription of either long or G+C-rich DNA sequences fused to a strong promoter such as GAL1pr.

**Hpr1 is required for proper transcription of long DNA sequences.** We have shown that transcription of long DNA sequences is compromised in hpr1 cells, whereas shorter sequences are either unaffected or mildly influenced. This conclusion is supported by the inappreciable effect of hpr1 on transcription of short lacZ fragments directly fused to the GAL1 promoter, whether or not upstream of PHO5, whereas the same lacZ fragments do confer hpr1 dependency when fused downstream of PHO5 (Fig. 2 and 3). This conclusion is also supported by the marked negative effect of hpr1 on transcription of the 3.7-kb-long LYS2 fragment under control of the GAL1 promoter (Fig. 5). The transcriptional analysis of five highly transcribed chromosomal genes showed a negative effect of hpr1 on transcription of the three longest ones, EGT2 (3.3 kb), CDC48 (2.7 kb), and KAR2 (2.1 kb), but no negative effect on the other two, OLE1 (1.6 kb) and GOG5 (1.1 kb) (Fig. 6). Further experiments would be required to know whether these results are also valid for poorly expressed genes.

Processivity defects of an RNA polymerase can be more easily detected with transcription of long rather than short DNA templates. Comparison between long and short transcripts is in fact a common method to quantify the effect of transcription factors on RNAPII-mediated elongation (66). Consequently, a length-dependency effect of hpr1 on transcription is expected if Hpr1 controls elongation. The longer the transcription unit, the higher the probability of RNAPII reaching a DNA region requiring the function of Hpr1. Even in the wild-type strain, long transcription units are less efficiently expressed than short ones, as we have observed by comparing

FIG. 8. Transcription and recombination analyses of YAT1 in wild-type and hpr1 cells. (A) Northern blot analyses of YAT1 mRNAs in cells transformed with plasmid pSCh247 containing the entire YAT1 coding sequence under the control of the GAL1 promoter. (B) Recombination analyses of cells transformed with plasmid pSCh248, which harbors the leu2-based direct-repeat system containing the entire YAT1 gene as intervening sequence. Other details are as described for Fig. 4.

FIG. 9. MNase digestion pattern of the GAL1 gene and the GAL1pr::lacZ fusion in wild-type and hpr1 strains under repression and activation conditions. (A) Northern blot analysis of GAL1 mRNAs. (B) Nucleosome positioning over the GAL1 gene. (C) MNase digestion pattern of the GAL1pr::lacZ fusion. A scheme of the analyzed regions of GAL1 and lacZ indicating the position of nucleosomes and the most relevant regulatory elements is shown. Asterisks indicate the MNase hypersensitive sites associated with the activation of transcription of GAL1.
the expression levels of several PHO5::lacZ fusion constructs that differ in the size of the lacZ fragment (Fig. 1). Thus, it is possible that the absence of Hpr1 enhances elongation defects already present in the wild type, the occurrence of such defects being more likely as the DNA sequence to be transcribed becomes longer.

**Transcription of G+C-rich DNA sequences is Hpr1 dependent.** Our results support a correlation between G+C content and HPR1 function. This conclusion is based on three sets of data. First, although most tested genes are either slightly or not affected by hpr1 (Fig. 7) (11, 67), transcription of lacZ (56% G+C) in *Saccharomyces* is strongly affected and direct-repeat systems containing lacZ sequences that are transcribed exhibit hyperrecombination in hpr1 cells (11). Second, transcription of LAC4, a 40% G+C-rich lacZ orthologue of *K. lactis*, is at least five times more efficient than lacZ in hpr1 cells (Fig. 7A), and direct repeats flanking LAC4 recombine at frequencies eightfold lower than those containing lacZ (Fig. 7B). Finally, YAT1, a 2-kb-long gene from *Saccharomyces* with a 58% G+C content, is strongly affected by hpr1 both in transcription and in recombination (Fig. 8).

As far as we know, neither a direct influence of the G+C content of the template on transcription elongation efficiency nor the requirement of auxiliary factors for transcription of G+C-rich genes has been described. The only sequence features that have been shown to affect elongation are those of specific transcriptional pausing sites (59). Some pause sites identified in bacteria are G+C rich, like the ops signals in *E. coli* (4), but others are not. It is, therefore, unlikely that G+C-rich genes exhibit, in general, a higher probability of containing a pause signal. Nonetheless, a high G+C content might affect transcriptional elongation by stabilizing secondary structures in the nascent RNA that can function as pausing signals. At least in the case of T7 phage, a lower number of hydrogen bonds in the RNA hairpins eliminates some pauses (32), suggesting that a high G+C content might contribute to stronger RNA-mediated elongation impairments.

A G+C-rich nascent RNA might also form more stable RNA-DNA hybrids within the template. Twelve-nucleotide-long RNA-DNA hybrids negatively affect RNAPII processivity in vitro (29). Another class of RNA-DNA hybrids are R-loops, produced by the association of nascent mRNA with upstream template DNA. It has been proposed that these R-loops are formed during transcription of the G+C-rich human immunglobulin switch region in vivo (15), and they have been detected after in vitro transcription (58). The cleavage of R-loops by specific nucleases might initiate class-switch recombination (58), providing a mechanism to explain transcription-associated recombination. Nothing is known about the influence of RNA-DNA structures on RNAPII-dependent transcription (20). However, it has been proposed that R-loops formed during rRNA transcription elongation in *E. coli* constitute roadblocks for the next transcribing RNA polymerase (25).

**Molecular features making transcription elongation Hpr1 dependent.** Unless Hpr1 plays more than one function, we should expect a common mechanistic requirement during transcription of long versus G+C-rich genes requiring its action. It is not evident which kind of transcription-impairing signal might link long and G+C-rich DNA sequences. One possibility would be the existence of shorter G+C-rich regions hindering transcription within long genes. However, we can exclude this possibility with the results of this work. Considering a 300-bp-long window (the minimal lacZ fragment conferring an effect of hpr1 in transcription), the maximum G+C content is 46% for LYS2, 45.5% for LAC4, and 59% for lacZ. Shorter or longer windows show similar results. In addition, no difference in the G or C content is found on the cDNA strands. As LYS2 is no more G+C-rich than LAC4, length is the most likely reason the two genes behave differently in hpr1 mutants. Consistently, when LAC4 is located between the leu2 repeats in the L-LAC4 construct, the transcription unit containing LAC4 becomes longer and transcription becomes significantly affected in hpr1 cells (Fig. 7).

Alternatively, the link between long and G+C-rich sequences might be unrelated to the DNA sequence itself. Transcription of long and G+C-rich sequences may produce some kind of transcriptional event that would be overcome by the action of Hpr1. Genetic analyses have provided some hints as to the nature of this kind of event. Several mutants have been described that display synthetic phenotypes with hpr1 (2, 67). The fact that topoisomerase mutants (top1, top2, and top3) become sick in an hpr1 background may establish a link with DNA topology (2, 48). For example, the accumulation of negative supercoiling impairs transcriptional elongation of bacterial genes in vitro (30), and positive supercoiling diminishes RNAPII-dependent transcription in yeast cells. R-loops are formed in the absence of DNA topoisomerase I in *E. coli* (16), and they have been proposed to impair transcription elongation (25). Elongation by RNAPII alters template topology (8), producing an accumulation of positive and negative supercoiling ahead of and behind the RNAPII, respectively (33). It is, therefore, expected that positive supercoiling will be stronger at the 3′ region of long transcription units. Examples of DNA sequences that impair transcription elongation more efficiently in distal locations have been reported (63). The strong effect of hpr1 on transcription of long DNA sequences agrees with this view.

Chromatin structure is another source of stress affecting RNAPII-mediated transcription elongation that requires the action of specific auxiliary factors (reviewed in reference 39). Some mutations resulting in a poor growth phenotype with hpr1 are in fact related to chromatin structure, like SIN1-2 or those causing histone imbalance (67). The organization of DNA in a proper, nucleosome-positioned chromatin structure is favored by some A+T-rich motifs (27) and is prevented by some G+C-rich sequences (62). A G+C-rich sequence might, therefore, be biased against a proper chromatin structure. The lacZ gene is in fact unable to support stably organized chromatin in *S. cerevisiae* (Fig. 9). Transcription of G+C-rich genes might then be impaired in hpr1 due to an aberrant chromatin structure.

If these hypotheses are true, the location of a G+C-rich region in the 3′ region of a gene should produce a stronger effect, since it would combine two sources of transcriptional stress in the same place: superhelicity and aberrant chromatin structure. Indeed, negative supercoiling can induce changes of DNA structure in CG sequences, with this altered structure being able to produce transcriptional elongation blocks (41). Our results with the PHO5::lacZ fusion constructs support this view, since the location of the G+C-rich lacZ fragments at the
3′ region produced a clear transcripational effect in hpr1 mutants (Fig. 2), whereas their location at the 5′ region had no effect (Fig. 2 and 3).

Finally, we cannot exclude the possibility that the size and G+C content of the nascent RNA, and not of its DNA template, are what determines the requirement of Hpr1 in transcription. This would imply that Hpr1, and by extension the THO complex, might also control RNA metabolism beyond transcription elongation. Such a possibility is consistent with the observation that transcription and RNA processing are coupled (reviewed in references 6, 23, 47) and could explain the observed RNA export defects of hpr1 mutants (51). In this respect, the recent observations that hpr1 is suppressed by overexpression of the putative RNA helicase SUB2 and that sub2 mutants are also hyperrecombinant (18) are noteworthy.

Hpr1 is stably associated in the cell nucleus with Tho2, Mtf1, and Thp2, forming the THO complex (12). Mutations affecting any of the four proteins cause the same phenotypes in transcription and transcription-associated recombination, although the quantitative effect of each mutation is different (12). Since the THO complex is a functional unit, the information obtained studying the gene spectrum affected by the THO complex is a functional unit, the information obtained studying the THO complex sheds light on the in vivo functional role of the complex. Characterization of the biochemical and functional properties of the THO complex in vitro will help in understanding why transcription of long as well as G+C-rich genes preferentially requires a specific cellular function. It remains to be seen whether these conclusions can be extended to endogenous yeast genes or to DNA sequences transcribed from poorly expressed promoters.

The existence of Tho2 homologues as well as of long genes and G+C-rich DNA sequences transcribed from poorly expressed promoters.

ACKNOWLEDGMENTS

We thank J. Polaina for providing the LAC4 gene, M. Beato for providing his facilities (IMT, Université de Marburg) for part of the chromatin analyses presented in this study, E. Santero for critical reading of the manuscript, and D. Haun for style correction.

This project was supported by grants from the Spanish Ministry of Science and Culture (PB96–1350) and the Human Frontier Science Program (RG0075/1999–M).

REFERENCES


62. Yankulov, K., J. Blau, T. Purton, S. Roberts, and D. L. Bentley. 1994. Transcriptional elongation by RNA polymerase II is stimulated by transac-