Overlapping coding regions and transcripitional units of two essential chromosomal genes (CCT8, TRP1) in the fungal pathogen Candida albicans

Michaela Gerads and Joachim F. Ernst*

Institut für Mikrobiologie and Biologisch-Medizinisches Forschungszentrum, Heinrich-Heine-Universität, Universitätstraße 1/26, 12, D-40225 Düsseldorf, Germany

ABSTRACT

Sequencing of the 3′-untranslated region of the CCT8 gene of the fungal pathogen Candida albicans revealed that the CCT8 coding region overlaps 13 bp with the coding region of the convergently orientated TRP1 gene. The same overlap was found in three strains with different genetic backgrounds. 3′-RACE was used to determine that the CCT8 and TRP1 transcripts extended significantly into the coding region of the adjacent gene, which also contained sequences encoding the poly(A) addition site. A strain retaining one wild-type CCT8/TRP1 locus on one chromosome and a deletion on the other homologous chromosome contained both CCT8 and TRP1 transcripts; this result indicates that both transcripts are synthesized from the same gene locus. The CCT8/TRP1 gene pair of C. albicans constitutes an extreme natural case of trancripitional overlap in a eukaryote. The results confirm that convergent overlapping transcription units are compatible with expression of the overlapping genes.

INTRODUCTION

The coding regions of chromosomal genes in eukaryotes are typically separated by non-coding regions of variable lengths, in which transcription initiation and termination processes proceed. Transcriptional interference is a plausible reason for the maintenance of this separation of genes in evolution. For transcriptional units orientated in the same direction transcriptional interference has been clearly shown and is manifested by repression of a downstream promoter by an upstream transcription unit (1–4). Overlap at transcript 5′ ends (in the case of neighboring divergent promoters) may be selected against by a similar mechanism. For convergent transcription units transcriptional interference may arise by colliding RNA polymerase II molecules reading the same template; alternatively, hybridization of complementary transcripts leading to their degradation may also occur in this case. In a recent report the first alternative has been made less likely by the demonstration that RNA polymerase II is detectable far beyond the poly(A) adenylation site and within the coding region of a convergently transcribed gene without alteration of its expression (5).

Although the complete genome of the yeast Saccharomyces cerevisiae is known and considerable information exists on its functional content, no case of the overlap of two genes that are functional has been reported. Overlap of genes would be expected to occur especially in yeast because of the tight packaging of genes on their chromosomes. Some S. cerevisiae genes contain open reading frames on the antisense strand, but they do not appear to be expressed (6). The overlap of reading frames in the Ty1 element and the overlap of the capsid and polymerase reading frames encoded by the L-A dsRNA (killer), which are resolved by translational frameshifting (7,8), are not typical examples of a yeast chromosomal gene.

In this report we demonstrate that two chromosomal genes considered essential for the human pathogenic yeast Candida albicans, TRP1 and CCT8, overlap in their coding regions and consequently in their transcriptional units. TRP1 encodes phosphoribosylanthranilate synthase and deletion of TRP1 does not allow growth in the absence of added tryptophan (9). CCT8 encodes subunit 8 of the cytoplasmic Cct chaperonin complex (10); its complete disruption in C. albicans is not possible (11) suggesting that is as essential as individual CCT genes in S. cerevisiae (10). The TRP1/CCT8 gene pair represents a unique natural example of two convergently transcribed genes, whose overlap is permitted in a yeast genome, thus confirming recent results on the presence of RNA polymerase II in adjacent coding regions (5).

MATERIALS AND METHODS

Strains

The C. albicans strains ATCC10231 (12), SC5314 and its derivatives CAI4 and CAI8 (13), and SGY-243 (14) have been described. The C. albicans strain FR1-1 (11) corresponds to strain CAI4 and is deleted in one of its CCT8 alleles between positions 808 and 1684 of the CCT8 genomic sequence (DDBJ/EMBL/GenBank accession no. U37371, revised; A of ATG start sequence is in position 13); between the deleted sequence the

*To whom correspondence should be addressed. Tel/Fax: +49 211 811 5176; Email: joachim.ernst@uni-duesseldorf.de
modified ‘URA blaster’ module containing cat repeats surrounding the URA3 gene is inserted (11,13).

**Sequencing**

The 2.8 kb BamHI–EcoRI fragment in pRS426U1, which was derived from ATCC10231 genomic DNA, contained the 5′-deleted CCT8 gene and the TRP1 promoter and coding regions (11). To verify sequences at the 3′ end of the CCT8 coding region (TRP1 overlap) a subclone was constructed (pRFG-Stop) containing the 489 bp EcoRV–NcoI fragment encompassing this region (positions 1483–1972 in the genomic CCT8 sequence) inserted into pUC21 (15). The insert of pRFG-Stop was sequenced by extension from both ends using M13 universal and reverse primers. Genomic sequences of strains CAI8 and SGY-243 were obtained by PCR on isolated genomic DNAs using primers p3'RACE/CCT8 (5′-TG-CCTGGTGCCGGTGCAGTTGA; positions 1229–1250 in the CCT8 genomic sequence) and p3'RACE/TRP/CCT8gen (SGY-243) were sequenced from both ends as described above. The SGY-243 genomic sequence was confirmed by subcloning a 458 bp EcoRV fragment of pPCR-TRP1/CCT8gen into pUC-21 (pPCR-RFG/Stop) or by deleting a 248 bp CiaI fragment of pPCR-TRP1/CCT8gen (PCR-T/CarV), thus facilitating sequencing of the overlap region.

**3′-RACE**

Total RNA of strain SC5314 was isolated as described (16) and poly(A) RNA was selected using the Oligotex mRNA Mini Kit (Qiagen). The following procedures for the generation of cDNA and 3′-RACE were according to a commercial protocol (5′3′ RACE Kit, Boehringer Mannheim). Briefly, cDNA was synthesized using the oligo dT anchor primer (5′-GAC-CACCGGTATCGATGTCGACTTTTTTTTTTTTTTTTV) and AMV reverse transcriptase. Amplification of specific mRNA 3′ ends was performed using the PCR anchor primer (5′-GAC-CACCGGTATCGATGTCGAC) and a gene-specific primer. For CCT8 primers p3'RACE/CCT8 or p3'RACE/CCT8neu (5′-ATGCTGCCCACCGCTGAAGACACTG; positions 1403–1426 in the CCT8 genomic sequence) were used for 3′-RACE (parameters 35 cycles; denaturation at 94°C for 30 s; annealing at 70°C for 1 min; elongation at 72°C for 1 min). For TRP1 primers p3'RACE/TRP or p3'RACE/TRPneu (5′-CGCAAAATGGTGAGTTTGGCCGCTG; positions 1867–1844) were used (parameters 28 cycles: denaturation at 94°C for 30 s; annealing at 70°C for 1 min; elongation at 72°C for 90 s). The PCR products were inserted into the Smal site of pUC18 (SureClone Ligation Kit, Pharmacia).

**Southern and northern blots**

DNA of strains SC5314 and FR1-1 was isolated as described (17), digested by various restriction enzymes and separated by agarose gel electrophoresis (0.8% gel). Separated fragments were transferred to nylon membranes and blots were probed using digoxigenin-labeled fragments using standard procedures (protocol by Boehringer Mannheim). The CCT8 probe was a 0.6 kb BamHI–SalI fragment encompassing the first half of the CCT8 coding sequence; the BamHI site had been inserted by PCR immediately upstream of the ATG translational start sequence (11). The TRP1 probe was a 0.6 kb HindIII fragment containing 5′-untranslated sequences and the 5′-half of the TRP1 coding region. Hybridizing probes were visualized by staining with nitroblue tetrazolium salt/5-bromo-4-chloro-3-indolyl phosphate toluidinium salt (protocol of Boehringer Mannheim).

Total RNA of strains SC5314 and FR1-1 was isolated and 15 µg RNA were separated on a denaturing 1.2% agarose gel containing formaldehyde (16). After transfer to a nylon membrane the blot was hybridized with a 248 bp Smal fragment of plasmid pUC-C/T. This plasmid was constructed by insertion of a 5′-ATT CCC GGG TAT CCA TAT AGA TTC and C/T-Smal2 (5′-ATT CCC GGG TAA AAG ATT CAT TG) (Smal sites, bold) into the Smal site of pUC18. The sequence of the 248 bp Smal probe is missing in the deleted cct8 allele of strain FR1-1.

**RESULTS**

**Overlap of CCT8 and TRP1 coding regions**

Further analysis of the CCT8 sequence (10) revealed that segments of its 3′-untranslated region (3′-UTR) were almost identical to the antisense strand of TRP1 3′ end; conversely, the published TRP1 sequence (9) had segments of identity in its 3′ region to the antisense strand of the CCT8 3′ end. Therefore, we decided to resequence this genomic region to elucidate the nature of the presumed CCT8/TRP1 overlap. Our original clone of the C. albicans CCT8 gene, pU1, was obtained from genomic DNA of strain ATCC10231 (12) and contained 1.5 kb of 3′-UTR downstream from the CCT8 coding region. This region was resequenced and revealed that the CCT8 and TRP1 genes are convergently orientated and that their coding regions overlap by 13 bp (including TGA stop sequences in both
genes). The genomic arrangement is shown in Figure 1 (wild-type) and the sequence of the overlap region is shown in Figure 2. Because it appeared possible that the result was specific for strain ATCC10231 alone we decided to sequence the CCT8/TRP1 overlap from C. albicans strains of different genetic backgrounds. To this effect we used two primers (p3’RACE/CCT8 and p3’RACE/TRP) to amplify the CCT8/TRP1 overlap region by PCR. A single 0.65 kb PCR product was obtained using DNA of strains CAI8 (derivative of SC5314) (13) and SGY-243 (14), which were sequenced. The sequence obtained showed 100% identity to the sequence obtained for strain ATCC10231 suggesting that the overlap of the CCT8/TRP1 coding regions is a general characteristic of C. albicans strains.

Our revised sequence corrects the deduced protein sequences of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p. The revised Cct8p sequence has a length characteristic of NcoI protein, which also ends in two acidic residues: W D Q previously described data (shown below) we think that a deletion occurred in the sequence predicts the 3′′S′′ end of both Cct8p and Trp1p.
detected. In genomic DNA of strain FR1-1, besides all fragments detected in the wild-type, the expected additional fragments due to the deleted CCT8 allele were seen: a 0.85 kb fragment (EcoRI), a 1.2 kb fragment (NcoI) and a 4.3 kb fragment (HpaI). Since the 5’ region of CCT8 is not affected in the FR1-1 deletion no additional fragment was discovered in a SalI digest of FR1-1 genomic DNA.

Likewise, using the TRP1 probe the expected fragments were discovered in the wild-type and the disruption strain. One fragment hybridized with the TRP1 probe of SC5314-DNA after digestion with EcoRI (3.2 kb), HpaI (0.85 kb), while two fragments were seen after digestion with NcoI (3.3 and 2 kb) and SalI (1.95 and 2.7 kb). The result obtained for the SalI digest was surprising, because only the 1.9 kb fragment corresponds to the known TRP1 sequence, suggesting that only one allele of TRP1 contains the SalI site shown in Figure 1. The allele with the 1.9 kb SalI fragment is disrupted in strain FR1-1, because the 1.9 kb fragment increases to 4.8 kb due to the ‘URA blaster’. The expected additional bands are detected in DNA of mutant FR1-1 after digestion with EcoRI (2.5 kb) and NcoI (0.5 kb; weak signal in Fig. 4), while the HpaI fragment, which is situated outside the deletion, is not changed.

Thus, the Southern results provide no evidence for major genomic rearrangements in any of the CCT8 or TRP1 alleles or in surrounding DNA sequences.

**CCT8 and TRP1 are expressed on a single chromosome**

Although the Southern results had excluded any major rearrangements in any of the CCT8 and TRP1 alleles, which could lead to their inactivation, it appeared still possible that the CCT8 or TRP1 genes were transcribed on different homologous chromosomes, with one allele silenced by more subtle mechanisms. Alternatively, convergent transcription of both genes on a single chromosome had to be considered. To decide among these possibilities we analyzed transcripts of strain FR1-1 (11) containing the described CCT8 deletion on one chromosome. Total RNA of the mutant was separated on a denaturing agarose gel and blots were probed using a specific probe (CT-probe) designed (i) to hybridize to both CCT8 and TRP1 transcripts and (ii) is missing in the deleted allele of FR1-1. Using this probe only alleles on the undeleted chromosomes can be detected.

Figure 5 demonstrates that indeed the 1.7 kb CCT8 transcript, as well as the 0.9 kb TRP1 transcript is detectable in RNA of mutant FR1-1 as in the wild-type strain SC5314. The ratio of the TRP1 to the CCT8 transcript appears greater in the mutant FR1-1 compared with the wild-type strain, which is consistent with the deletion of one CCT8 copy. It must be considered that the deletion in strain FR1-1 only deletes the extreme 3’ end of the TRP1 coding region; thus, a modified TRP1 transcript containing ‘URA blaster’ sequences at its 3’ end and having the same size as the authentic TRP1 transcript, is still a possibility. Therefore, alteration of the TRP1/CCT8 transcript ratio may be either due to
a relief of transcriptional interference, or more simply, by differences in allele copy numbers.

Thus, these results demonstrate unequivocally that the overlapping CCT8 and TRP1 genes are convergently transcribed on a single chromosome.

**DISCUSSION**

Overlap of coding regions of genes is common in viral genomes, but the exception in eukaryotes including fungi. In this report we present a unique natural case of gene overlap in fungi, in which the coding regions of two essential genes of the fungal pathogen *C. albicans* overlap in their 3′ regions by 13 bp. We have excluded the possibility that transcriptional interference is avoided in the diploid *C. albicans* by silencing one allele of CCT8 or TRP1 on each homologous chromosome by gross chromosomal alteration, which could allow unperturbed expression of the other gene on a different homologous chromosome. By deleting the CCT8/TRP1 locus on one chromosome we have shown that both CCT8 and TRP1 transcripts are derived from the same chromosome. Thus, transcription of both genes is compatible with their convergent orientation and the overlap of their coding regions.

Overlap of the 3′ ends of coding regions could imply that signals required for the generation of transcript 3′ ends are situated in the antisense strand of the coding region of the neighboring gene. We have shown this directly for the poly(A) addition sites of the CCT8 and TRP1 genes. In yeast the formation of transcript 3′ ends depends on the formation of a primary transcript which terminates probably within 400 bp of the poly(A) addition site (termination being defined as the dissociation of the RNA polymerase II from its template DNA) (18,19). Recently, it has been shown that RNA polymerases II can transcribe in excess of 2 kb downstream of the poly(A) addition site (5). Efficient processing at the poly(A) site appears directly related to early termination, possibly because an exonuclease ‘catches up’ with the RNA polymerase and initiates its release (5). In *S. cerevisiae* the polyadenylation sites comprise an upstream efficiency element (consensus UAUAUA), a positioning element (consensus AAA/UA/AA) and the actual poly(A) addition site [consensus Y(A)n] 15–30 nt downstream (20,21). Although transcript termination sequences for *Calicibon* have not yet been defined we note that sequences resembling the yeast efficiency element (UAUAUA and UAUAU) and the positioning element (AAUA-UAA) precede the poly(A) site of the CCT8 poly(A) addition site, one of which has the structure CAA (Fig. 2). Likewise, sequences resembling the efficiency element (UAUAGAUA) and the positioning element (AACAG) preceed the poly(A) addition site of the TRP1 transcript, which is situated at a UA sequence (Fig. 2).

The described CCT8/TRP1 gene pair provides an extreme case for transcriptional overlap in the case of convergent transcriptional units in that the coding regions of the genes involved are overlapping. Other examples (in which the respective coding regions are not overlapping) include the nmt1/gut2 and nmt2/avn2 transcription units in fission yeast (5), as well as the RHO1/MRP2 transcription unit in *S. cerevisiae* (22). In the latter case it was shown that expression and polyadenylation was not affected by overproduction or elimination of the complementary natural antisense transcript. Only if either the RHO1 or the MRP2 transcript was transcribed by a strong promoter was the overlapping gene inactivated (22). In the case of the *Calicibon* CCT8/TRP1 gene pair the deletion of one CCT8 allele did not significantly increase TRP1 transcript levels. Thus, it appears that overlapping converygent orientated transcription units do not affect each other significantly unless one (or both) units are expressed at very high levels. In contrast, overlapping transcription units orientated in the same direction clearly interfere with each other (1–4).

If convergent transcription units are functionally compatible it can be speculated why cases like the overlapping coding regions of the CCT8/TRP1 genes in *C. albicans* do not occur more frequently, i.e. in the compact genome of its close relative *S. cerevisiae*. Open reading frames are occasionally observed on the antisense strand of coding regions, but they do not appear to be expressed (6). As discussed above, highly expressed genes (e.g. encoding glycolytic proteins) are not expected to overlap with other genes. Gene regulation via antisense transcripts has been described in Dictyostelium (23), but in *S. cerevisiae* antisense regulation mechanisms leading to dsRNA degradation, impaired nuclear transport or reduced translation are not general mechanisms of gene regulation (24). The configuration of the CCT8/TRP1 genes suggests that also in *C. albicans* antisense regulation as a general mechanism of gene regulation does not occur. On the other hand, once generated in evolution, convergent coding regions will be evolutionarily ‘immobilized’, because (i) informational content is located both on the sense and the antisense strands and (ii) translocation of only one gene of the pair to new genomic sites is prohibited unless both genes are mobilized functionally intact. It remains to be determined at which point in fungal evolution the CCT8/TRP1 overlap occurred and if it is specific for all or only a subset of *Candida* species. Potentially, the evolutionarily constrained sequences of the CCT8/TRP1 pair are advantageous in that they can be used diagnostically to identify *Calicibon* subtypes with a common probe sequence, e.g. by PCR methods.

**Figure 5.** CCT8 and TRP1 transcripts in *C. albicans* strains. Strain SC5314 (lanes 1 and 2) and mutant strain FR1-1 (lanes 3 and 4) were grown in YPD (lanes 1 and 3) or in SD minimal medium (lanes 2 and 4) and total RNA was prepared. Fifteen micrograms RNA were separated by denaturing gel electrophoresis and the RNA blot was probed using the ‘CT probe’ (Fig. 1), able to detect both CCT8 and TRP1 transcripts. Note that the probe sequence is missing in the deleted CCT8 allele of the FR1-1 mutant strain.
ACKNOWLEDGEMENTS

This work was supported by grant BMH4-CT96-0310 of the European Community and by the Deutsche Forschungsgemeinschaft.

REFERENCES