RNA molecules containing exons originating from different members of the cytochrome P450 2C gene subfamily (CYP2C) in human epidermis and liver

Peter G. Zaphiropoulos*

Department of Bioscience, Center for Nutrition and Toxicology, Karolinska Institute, Novum, 141 57 Huddinge, Sweden

Received April 22, 1999; Revised and Accepted May 17, 1999

ABSTRACT

Reverse transcription–PCR analysis in human epidermis, using primers from CYP2C18 and CYP2C19, revealed products containing combinations between canonically defined exons of these two genes. The major RNA species identified contained 2C18 exon 8 spliced with 2C19 exon 2. However, the terminal exons 1 and 9 were never detected in any of these composite molecules. When similar experiments were performed with liver RNA, exons 1 and 9 of both 2C18 and 2C19 were readily identified in composite 2C18/2C19 RNAs. Moreover, molecules containing 2C9 sequences spliced with 2C18 exons were also detected. These findings suggest that during the process of RNA splicing of the 2C transcripts, various exon juxtaposition events may occur, including combinations between exons of distinct genes. However, the frequency of these events is quite low and the levels of the composite RNA molecules are generally estimated at less than one molecule per cell. Since the order of these genes on chromosome 10q24 is CYP2C18–CYP2C19–CYP2C9, it is conceivable that the composite RNAs may result from multiple canonical and inverse splicing events of a long pre-mRNA that encompasses the three genes. However, these molecules could also be rationalized as being the products of trans splicing phenomena between distinct pre-mRNAs.

INTRODUCTION

The splicing process is a major event in eukaryotic gene expression allowing non-contiguous segments of pre-mRNAs, the exons, to be joined together. The mechanism underlying this phenomenon was originally thought to be rationalized by a scanning model mechanism, which, in its simplest form, suggests that the spliceosome moves along the pre-mRNA and excises the introns as it encounters them (1–3). However, the fact that within large introns there are sequences with a good match to a single splice site consensus that nevertheless remain non-utilized is difficult to explain by the scanning model. More recently it has been proposed that splice sites are initially recognized as closely spaced pairs, positioned across an exon (or possibly across small introns) and subsequent exon juxtaposition events allow the defined exons to be joined together (4,5). Although the detailed mechanisms of exon juxtaposition are not fully understood, a prominent role for the non-snRNP class of splicing factors, the SR proteins and the hnRNPs, is emerging (6–15).

In line with the proposed exon definition/exon juxtaposition model of splicing is the finding of RNA molecules that have an inverted order of exons (exon scrambling) (16,17). Once the exons are defined, it is plausible that some juxtaposition events may allow exons that are not in the same order as in genomic DNA to be brought together. It is also possible that splice sites across a single exon could be joined, resulting in exon circles (18–21). Moreover, the RNA species containing scrambled exons also have properties that are consistent with circular molecules, as exemplified by the lack of both a poly(A)+ tail and a cap structure. Furthermore, in some cases, the exons composing the scrambled RNAs were found to be absent in certain alternative spliced mRNA molecules (22,23). This observation suggested that the processes of exon skipping and circular RNA formation may be interrelated and could arise from the same mechanism.

To further investigate the variations that may occur during the process of exon juxtaposition, it was examined whether exons from distinct genes could be spliced together. As a model system, the cluster of the human CYP2C genes was selected. These are four single copy genes that have been mapped in chromosome 10q24, in a region of ~500 kb (24). The gene order and orientation in the cluster is Cen–CYP2C18–CYP2C19–CYP2C9–CYP2C8–Tel. The human CYP2C18 and CYP2C9 genes span >55 kb and possess nine exons, the intron–exon boundaries of which are conserved with those of CYP2C genes of other species (25).

Surprisingly, and contrary to expectations, epidermal RNA molecules containing various combinations of exons from the CYP2C18 and the CYP2C19 genes, joined at their canonical splice sites, were detected. Moreover, liver RNA molecules containing exon combinations between the CYP2C18 and either the CYP2C19 or the CYP2C9 gene, including the terminal exons 1 and 9, were also identified.

*Tel: +46 8 6089151; Fax: +46 8 6081501; Email: peza@cnt.ki.se
MATERIALS AND METHODS

Total human epidermal RNA was isolated by the guanidinium isothiocyanate method (26) as described before (23). Total liver RNA was purchased from Clontech. cDNA synthesis was performed using either random hexamers or oligo(dT) primers, following previously established protocols (23). Both the initial and the nested PCR amplifications were performed for 30 cycles, with 1 min at 94°C, 1 min at 52–54°C and 1 min at 72°C, using Perkin-Elmer model 480 and 2400 thermocyclers or a MJ Research model PTC-200 DNA engine. For the nested amplifications, 1 μl of the first amplification mix was used directly. When Taq polymerase was substituted for Expand (Boehringer), the extensions were performed for 3 min at 72°C, using the buffers supplied by the manufacturer. The PCR products were analyzed by gel electrophoresis and cloned using the pGEM-T vector system (Promega). Plasmid DNA was extracted from randomly chosen clones using the JetQuick kit (Genomed) and sequenced with dideoxynucleotides at the facilities of Cybergene AB. Sequencing comparisons were performed with the BLAST programs of NCBI.

RESULTS

Non-polyadenylated RNA molecules containing 2C19 exons spliced 5' of 2C18 exons

During the RT–PCR analysis of the 2C18 scrambled RNA molecules (23), some products were identified that contained segments of 2C19 exons. These preliminary findings prompted the initiation of a more thorough analysis of RNA species that might contain combinations of 2C19 and 2C18 exons. A set of 2C18 exon 5 antisense primers was used with sets of either 2C19 exon 4 or exon 5 sense primers (Table 1) in RT–PCR amplification of oligo(dT) and random primed cDNA from human epidermis. Several different sized products were detected from the random primed but not from the oligo(dT) primed cDNA with both sets of PCR primers. This finding suggests that the products obtained originate from non-polyadenylated RNAs, as oligo(dT) priming is unable to result in detectable amplification. Cloning and sequencing of these products established that they were specifically initiated with the primers used and contained several combinations of 2C19 exons spliced together with 2C18 exons at their canonical splice sites (Table 2).

Polyadenylated and non-polyadenylated RNA molecules containing 2C18 exons spliced 5' of 2C19 exons

Since in the experiment described above the design of the primers used was such that 2C19 sense primers were combined with 2C18 antisense primers, RT–PCR analysis was subsequently performed using sense primers from 2C18 and antisense primers from 2C19. Sets of PCR primers from 2C18 exon 7 and from 2C19 exon 3 were combined (Table 2) and used to amplify oligo(dT) and random primed cDNA (Fig. 1). Interestingly, the patterns of the PCR products obtained from both the oligo(dT) and the random primed cDNA were quite similar. Two major species were revealed (Fig. 1, lanes 5 and 6) and by cloning and sequencing were found to contain 2C18 exons spliced together with 2C19 exons at their canonical splice sites (Table 2).

To obtain additional evidence that the species detected originate from polyadenylated RNAs, the experiment described above was repeated using new cDNA preparations that were either primed or not primed by oligo(dT). Only the oligo(dT) primed cDNA resulted in detectable PCR products of the expected size, suggesting that an oligo(dT) priming event has to occur during reverse transcription in order to allow efficient amplification of these species (data not shown).

To further analyze the RNA molecules that contained 2C18 exons spliced upstream of 2C19 exons, the exon 7 sense primers of 2C18 were substituted with exon 4 sense primers (Table 2). Again, the pattern obtained with oligo(dT) primed cDNA had similarities to the one with random hexamers.

---

Table 1. Sequence of the PCR primers used in the RT–PCR amplification experiments

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19</td>
<td></td>
</tr>
<tr>
<td>1f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-TGG ATC CAG CTG TGG CT</td>
</tr>
<tr>
<td>1f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-GGT CTC TGT TGG CTC T</td>
</tr>
<tr>
<td>2f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-TGG GTA TTT TGG CCT GAA G</td>
</tr>
<tr>
<td>2f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-ATA TGA AGC AGT GAA GAG G</td>
</tr>
<tr>
<td>3f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-TGG GGA AGA GGA GAC TCA</td>
</tr>
<tr>
<td>3f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-GGA GGT GAG AAA AAC CAA T</td>
</tr>
<tr>
<td>4f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-ACT TGA TGG AAA AAT TCA ATG</td>
</tr>
<tr>
<td>4f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-CTC AGG ATT CTG AGC TCT</td>
</tr>
<tr>
<td>5f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-GGA GAG ATG AAA AGA ACA TC</td>
</tr>
<tr>
<td>5f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-AAG AAT CCC TGG AGA TGA A</td>
</tr>
<tr>
<td>6f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-CAT GAG TAC ATT GAC TCT C</td>
</tr>
<tr>
<td>6f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-GAC GTG TGA TGA ATT CAA A</td>
</tr>
<tr>
<td>7f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-ATA CAT AAC TTT TAA TGT AAC</td>
</tr>
<tr>
<td>7f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-TTT TAT TAT GAC TTC CTG G</td>
</tr>
<tr>
<td>9f&lt;sub&gt;2C18&lt;/sub&gt;</td>
<td>5'-GCA GCC ACC AAA CTA TCC GCC</td>
</tr>
<tr>
<td>9f&lt;sub&gt;2C19&lt;/sub&gt;</td>
<td>5'-CAG ACA GGA ATG AAC CAG</td>
</tr>
</tbody>
</table>

Primers are numbered according to the exons, with f and F representing sense and r and R antisense primers.
However, the latter was characterized by a higher complexity. This finding suggests that the products common in the oligo(dT) and the random primed cDNA preparations originate from polyadenylated RNAs. However, the products present exclusively in the random primed cDNA are likely to result from non-polyadenylated RNAs. Cloning and sequencing of these products established that they contained combinations of 2C18 and 2C19 exons spliced at their canonical splice sites (Table 2). Only the longest 2C18/2C19 species, containing exons (4–5–6–7–8)2C18 –(2–3) 2C19, was detected in both oligo(dT) and random primed cDNAs. Worth noting is that the first 2C19 exon present in these products was always found to be exon 2.

RT–PCR products containing repeats of 2C19 exons: evidence supporting a circular nature of these RNA molecules

To investigate whether RNA molecules containing combinations of 2C18 and 2C19 exons could be detected not only with sets of PCR primers from each of these two CYP2C genes, but also with primers originating from a single CYP2C gene, PCR was performed using primers in both orientations of 2C19 exon 5 (Table 1). It was anticipated that most of the products would represent typically scrambled molecules, containing exclusively 2C19 exons. Oligo(dT) primed cDNA did not result in specifically amplified PCR products, however, random primed cDNA resulted in numerous products that were cloned and sequenced (Fig. 2 and Table 2). As expected, most of these products contained exclusively 2C19 exons, with the most abundant species having a 5–4–5 order of exons. However, some products were found to contain a 2C18 exon, exon 8, that was flanked on both sides with 2C19 exons (three out of 50 independent clones; order of exons: 5–8 2C18 –2–3–4–5). Moreover, three additional independent clones were identified that contained repeats of 2C19 exons 5 and 4 (order of exons: 5–4–5–4–5), indicating that scrambling had occurred twice between exon 5 and exon 4. To rationalize the mechanism of generation of these molecules, one has to assume that not only one but two distinct trans splicing events had occurred. Alternatively, a simpler interpretation for the latter molecule is

White boxes represent 2C18 exons, shaded boxes 2C19 exons and hatched boxes 2C9 exons. The forward or reverse primers of individual exons are indicated by the number of the exon and the letters F or R, respectively. Note that some products contained sequences representing complete introns (intron), segments of introns (intronic) or were of unknown origin (?).
that a circular RNA template composed of exons 4 and 5 has been reverse transcribed several times during cDNA synthesis generating linear 5–4–5–4–5–4–5⋯ species. Indeed, the capability of polymerases to follow a rolling circle model of DNA/RNA synthesis is well documented (28,29). Moreover, when the same experiment was performed with Expand, a combination of thermostable polymerases that is optimized for the amplification of long DNA templates, a PCR product containing four repeats of the scrambled 5–4 core sequence could be detected (data not shown).

The polyadenylated 2C18/2C19 RNA molecules do not contain the expected terminal exons

When sense primers from exon 5 of 2C18 were combined with antisense primers from exon 5 of 2C19 (Table 1) a major product was detected from the undiluted cDNA, suggesting that the level of synthesis is well documented (28,29). Moreover, when the same experiment was performed with Expand, a combination of thermostable polymerases that is optimized for the amplification of long DNA templates, a PCR product containing four repeats of the scrambled 5–4 core sequence could be detected (data not shown).

mRNA, serial 10-fold dilutions of epidermal oligo(dT) cDNA were amplified with 2C18 exon 7 sense primers combined with either 2C19 exon 3 antisense or 2C18 exon 9 antisense primers. The 10⁻¹ dilution allowed the synthesis of equivalent amounts of the 2C18 product as the 2C18/2C19 product generated from the undiluted cDNA, suggesting that the level of expression of the 2C18/2C19 species is about an order of magnitude less than the 2C18 mRNA (data not shown).

Lack of RNA molecules containing combinations of 2C18 exons with either 2C8 or 2C9 exons in human epidermis

To determine whether the finding of RNA molecules containing exons from two genes of the CYP2C cluster might not be limited to the neighboring CYP2C18 and CYP2C19 genes, RT–PCR amplification was performed using primers from 2C18 exons combined with primers from exons of either 2C8 or 2C9. However, in no case could a specifically amplified PCR product be detected (data not shown). This finding suggests that the proximity of the CYP2C18 and CYP2C19 genes in chromosomal locus 10q24 might be an essential factor that allows the generation of epidermal RNA molecules containing exon combinations from these two genes.

2C18/2C19 RNA molecules containing 2C18 exon 1 or 2C19 exon 9 in human liver

The unexpected finding that in epidermis the 2C18 exon 1 and the 2C19 exon 9 are not present in the 2C18/2C19 RNA molecules prompted an investigation as to whether these exons might be present in composite RNAs from the liver. Indeed, by RT–PCR analysis using primer combinations either from 2C18 exon 1 and 2C19 exon 3 or from 2C18 exon 7 and 2C19 exon 9 (Table 1). 2C18/2C19 products containing the terminal exons were detected from both oligo(dT) and random primed cDNA (Table 2). Interestingly this experiment revealed that 2C18 exon 1 can also be spliced with exons from 2C9. This 2C18/2C9 product was generated because the exon 3 antisense primer used also amplifies 2C9 sequences (complementary except for the 5’ nucleotide to 2C9). What is worth noting is that the 2C9 exon spliced with 2C18 exon 1 is not a canonical exon but originates from the 5’ flanking sequence of the CYP2C9 gene. It is enclosed between positions –246 and –111 and has AG and GT dinucleotides at its borders (25).

2C18/2C19 RNA molecules containing 2C18 exon 9 or 2C19 exon 1 in human liver

Since a terminal exon from each of the CYP2C18 and CYP2C19 genes was found to be present in composite 2C18/2C19 RNA molecules in liver, it was investigated whether the remaining terminal exons, i.e. exon 9 from CYP2C18 or exon 1 from CYP2C19, could also be present in composite 2C18/2C19 species. For that purpose RT–PCR analysis was performed using primer combinations either from 2C19 exon 1 and 2C18 exon 5 or from 2C19 exon 5 and 2C18 exon 9 (Table 1). Indeed, RNA molecules containing exon combinations between 2C18 and 2C19 that included either 2C19 exon 1 or 2C18 exon 9 were detected (Table 2). Moreover, a species containing 2C9 exon 5 spliced to 2C18 exon 8 was also identified. This product was generated because the exon 5 sense primer used also amplifies 2C9 sequences (complementary except for position 4 to 2C9).
DISCUSSION

The presence of various combinations of 2C18 exons with either 2C19 or 2C9 exons in contiguous RNA molecules is intriguing and suggests that juxtaposition events of defined exons might not be limited to single genes. Splicing in *trans* is known to occur in lower eukaryotes, the trypanosomes and also in nematodes and flatworms (30,31). Moreover, recent experimental findings have suggested that this phenomenon may also be occurring in higher eukaryotes (32–34).

The RNA molecules identified in this report containing combinations of exons from different CYP2C genes are not abundant. Northern analysis, in our hands, is not sensitive enough to allow their detection and efforts to show their presence by RNase mapping were also unsuccessful. However, the fact that the exons present in these molecules are spliced at correct sites provides a strong argument that they pre-exist in the RNA samples analyzed and therefore do not result from any type of PCR artifact. Certainly these RNA species are not abundant and may, in some cases, be present at levels lower than one copy per cell. On the other hand it should be noted that in yeast, 80% of expressed genes exist at levels of 0.1–2 molecules per cell, implying that low abundance transcripts represent the majority of expressed RNAs in biological systems (35).

**Trans** splicing versus splicing from a multigene primary transcript

What is a mechanism that may account for the finding of RNA molecules containing exon combinations from tandemly arranged genes? At the genomic level, although the order of these genes has been determined as CYP2C18, CYP2C19 and CYP2C9, their relative orientation is unknown, i.e. whether after the end of the CYP2C18 gene there is the start of the CYP2C19 gene or vice versa. Moreover, it is not known if these human CYP2C genes are in the same phase, i.e. transcribed from the same DNA strand, although judging from the genomic organization of other CYP subfamilies, this is anticipated (36). However, whatever their relative orientation may be, as long as these CYP2C genes are in the same phase, it is conceivable that a primary RNA transcript could be generated that encompasses not only the 5’-most gene (CYP2C18 or CYP2C9), but also additional genes of the cluster. This would support the notion that a transcription unit in higher eukaryotes may encompass a neighboring gene, as is the case with *Caenorhabditis elegans* (37). Moreover, since splicing is known to have the capability to join exons flanking introns that are >100 kb, one may assume that, in this primary transcript, exon joining of distantly localized exons originating from different CYP2C genes could take place. In fact, the finding of a transcript containing 2C18 exon 1 joined to a 2C9 exon-like sequence present in the promoter of the CYP2C9 gene may support the postulation of splicing events from a putative multigene primary transcript. However, to account for the observation that 2C18 exons were found both upstream and downstream of 2C19 exons, one has to assume that inverse splicing events also take place in processing of the extended primary transcript. This type of splicing, with downstream exons joined 5’ of upstream exons, has been demonstrated to occur only within internal exons (22,38). Therefore, to rationalize the finding of 2C18/2C19 RNA molecules containing the terminal exons 1 and 9, the multigene primary transcript hypothesis has to include an additional novel element, namely that inverse splicing events are not limited to internal exons but may also encompass terminal exons.

Inclusion of the 2C18 and 2C19 terminal exons in composite 2C18/2C19 RNA molecules can also be interpreted as an argument that these species represent products of authentic *trans* splicing events. Canonical pre-mRNAs of individual CYP2C genes, at low frequency, undergo splicing events in *trans*, simply because of the proximity of the newly synthesized primary transcripts to the active splicesomal complex. In fact, this would be in agreement with the increasing body of evidence that suggests a tight linkage between transcription and pre-mRNA processing (39). However, the identified RT–PCR product that contains 2C18 exon 8 embedded within 2C19 exons, i.e. the exon 5–82C18–2–3–4–5 molecule, cannot be rationalized by a single *trans* splicing event. To account for this RNA species one has to assume that two distinct *trans* splicing events have occurred, one that joined exon 5 of 2C19 with exon 8 of 2C18 and another that joined exon 8 of 2C18 with exon 2 of 2C19. An alternative and simpler interpretation though is that this molecule was generated by inverse splicing events from an extended primary transcript that encompasses the two CYP2C genes. On the other hand, an extensive analysis of scrambled RNAs containing exclusively 2C8 exons revealed products that can only be rationalized by *trans* splicing events (C.Finta and P.G.Zaphiropoulos, unpublished observations).

The lack of the canonical terminal exons in the 2C18/2C19 RNA molecules in epidermis was not an anticipated finding. These observations suggest that in this tissue, alternative terminal exons are used, since it is difficult to assume that internal exons can substitute for terminal exons. In this context it is worth noting that the penultimate exon with a segment of the 3’ flanking intron can substitute for the last exon in certain transcripts of the putative tumor suppressor gene Patched 2 (40), providing some clues as to the variations in the definition of terminal exons that may occur. In liver, though, 2C18/2C19 RNA species containing terminal exons have been detected not only by the RT–PCR technology, but also by means of cDNA library screening (41). This 2C18/2C19 cDNA contained 2C19 exon 9 and had exon 8 of 2C18 joined to exon 2 of 2C19, i.e. the most frequent joining of 2C18 and 2C19 exons observed by RT–PCR.

In summary, the present work provides evidence that, at low frequency, exons from distinct CYP2C genes can be spliced together. Whatever the mechanism of generation of these RNA molecules may be, authentic *trans* splicing events, splicing from a multigene primary transcript or a combination of both, these findings substantiate the proposal that the splicing process is not restricted to the linear order of exons of single genes. At low frequency, events that may juxtapose exons from neighboring genes, irrespective of their relative orientations, can occur, implying an additional complexity in the expression of the informational content of genes.

ACKNOWLEDGEMENTS

Dr Csaba Finta is acknowledged for helpful discussions and critical reading of the manuscript. This work was supported by grants from the Swedish Natural Science Research Council, the foundations Lars Hiertas Minne and Åke Wiberg and the Karolinska Institute.
REFERENCES