

High frequency *trans*-splicing in a cell line producing spliced and polyadenylated RNA polymerase I transcripts from an rDNA-*myc* chimeric gene

Célia Chen, Nicole Fossar, Dominique Weil², Marine Guillaud-Bataille¹, Gisèle Danglot¹, Brigitte Raynal, François Dautry², Alain Bernheim¹ and Olivier Brison*

Laboratoire de Génétique Oncologique, ¹Laboratoire de Génomique Cellulaire des Cancers, UMR 8125, Institut Gustave Roussy, Villejuif, France and ²Génétique Moléculaire des Fonctions Cellulaires, UPR 1983, Institut André Lwoff, Villejuif, France

Received October 10, 2004; Revised March 15, 2005; Accepted April 4, 2005

DDBJ/EMBL/GenBank accession nos AJ749659–AJ749668

ABSTRACT

The 2G1MycP2Tu1 cell line was obtained following transfection of human colon carcinoma cells from the SW613-S cell line with a plasmid carrying a genomic copy of the human *MYC* gene. 2G1MycP2Tu1 cells produce *MYC* mRNAs and proteins of abnormal size. In order to analyze the structure of these abnormal products, a cDNA library constructed using RNA isolated from these cells was screened with a *MYC* probe. Fifty clones were studied by DNA sequencing. The results indicated that a truncated copy of the *MYC* gene had integrated into an rDNA transcription unit in 2G1MycP2Tu1 cells. This was confirmed by northern blot analysis, PCR amplification on genomic DNA and fluorescent *in situ* hybridization (FISH) experiments on metaphase chromosomes. 2G1MycP2Tu1 cells produce hybrid rRNA-*MYC* RNA molecules that are polyadenylated and processed by splicing reactions involving natural and cryptic splice sites. These transcripts are synthesized by RNA polymerase I, as confirmed by actinomycin D sensitivity experiments, suggesting that 3' end processing and splicing are uncoupled from transcription in this case. 2G1MycP2Tu1 cells also produce another type of chimeric mRNAs consisting of correctly spliced exons 2 and 3 of the *MYC* gene fused to one or more extraneous 5' exons by proper splicing to the acceptor sites of *MYC* exon 2. These foreign exons belong to 33 different genes, which are located on 14 different chromosomes. These observations and the results of FISH and

Southern blotting experiments lead us to conclude that *trans*-splicing events occur at high frequency in 2G1MycP2Tu1 cells.

INTRODUCTION

Eukaryotic gene expression is a complex multi-step process requiring several multi-component cellular machines. Each machine carries out a separate step in the gene expression pathway. For example, the maturation of eukaryotic mRNAs that are synthesized by RNA polymerase II consists of the addition of a cap structure at the 5' end of the mRNA, splicing out of the introns, cleavage and polyadenylation at the 3' end (1). Splicing and 3' end polyadenylation have been traditionally viewed as post-transcriptional events. However, it has emerged recently that these processing events are tightly linked to transcription (2). The C-terminal domain of the largest subunit of RNA polymerase II plays an important role in coupling transcription with RNA processing. This domain specifically binds the capping enzyme, splicing factors and 3' end processing factors, thus targeting them to the nascent transcripts. Nevertheless, a controversy persists as to the requirement of RNA polymerase II for polyadenylation. It has been reported that RNA polymerase I-generated transcripts are polyadenylated in mouse 3T3 cells (3). Furthermore, it seems that the polyadenylation machinery is able to function independently of RNA polymerase II transcription in yeast (4). In contrast, it has been shown that herpes simplex virus *tk* mRNAs synthesized by RNA polymerase I are not polyadenylated in monkey COS-7 cells (5). Thus, it is still unclear whether RNA polymerase II transcription is necessary for polyadenylation and, more generally, for mRNA processing. Although splicing reactions can occur independently of transcription in reconstituted systems *in vitro* (albeit with a low

*To whom correspondence should be addressed at UMR 7147, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. Tel: +33 1 42 34 66 68; Fax: +33 1 42 34 66 74; Email: olivier.brison@curie.fr

efficiency), there is almost no data documenting the possible uncoupling of splicing from RNA polymerase II-mediated transcription *in vivo*. Splicing of RNA polymerase I transcripts has been reported for protein-coding genes in *Trypanosoma brucei* (6).

In the vast majority of cases, maturation of pre-mRNAs occurs by splicing in *cis*. However, cases of maturation by *trans*-splicing have been described, especially in trypanosomes (7), nematodes (8) and *Drosophila* (9,10). In this latter organism, *trans*-splicing is very important in generating protein diversity. In mammals, there are few examples of pre-mRNA maturation by *trans*-splicing [(11–13) and references therein]. For example, in the rat, the carnitin octanoyltransferase (*COT*) gene produces three different transcripts through *cis*- and *trans*-splicing reactions (11). In any case, it seems that *trans*-splicing always occurs in parallel with, and at a far lower frequency than, *cis*-splicing (12). Furthermore, the existence of proteins encoded by *trans*-spliced mRNAs has not been reported in mammals, except in the case of the rat *COT* gene. Therefore, the physiological function, if any, of these maturation pathways by 'alternative' *trans*-splicing (as opposed to 'specialized' *trans*-splicing which is a normal mRNA maturation pathway in some organisms such as trypanosomes and nematodes) remains to be established. Nevertheless, the potentiality for *trans*-splicing does exist in mammalian cells and RNA reprogramming strategies based on it and aimed at therapeutic interventions have been recently described (14).

We are interested in the role of *MYC* gene amplification in tumor cells. We have isolated two types of sub-clones from the SW613-S human colon carcinoma cell line that exhibit either a high or a low level of *MYC* gene amplification and whose cells disclose marked phenotypic differences (15,16). In order to demonstrate that the level of amplification of the *MYC* gene plays a role in these phenotypic differences, stable transfectants were established following transfection of cells displaying a low level of amplification with a plasmid containing a *MYC* gene. The structure of the abnormal *MYC* mRNAs produced by one of these transfectants was analyzed in detail. During the course of this analysis, we uncovered very unusual pathways of RNA maturation in these cells, involving splicing and polyadenylation on RNA polymerase I-generated *MYC* transcripts as well as *trans*-splicing at a high frequency.

MATERIALS AND METHODS

Construction and analysis of the cDNA library

The cDNA library was constructed in a vector derived from bacteriophage lambda using the ZAP-cDNA[®] Gigapack[®] III Gold Cloning Kit (Stratagene). Briefly, cDNA first strand was synthesized in the presence of methyl-dCTP on polyA⁺ RNA prepared from 2G1MycP2Tu1 cells, using the Moloney murine leukemia virus reverse-transcriptase and a poly-dT primer containing a XhoI site. The RNA template was partially digested with RNase H and the second strand of the cDNAs was synthesized with *E.coli* DNA polymerase I. An adaptor oligonucleotide containing an EcoRI site was ligated onto the cDNA ends. After digestion with XhoI, the cDNAs were size-fractionated by chromatography on a Sepharose CL-2B column. The selected cDNAs were ligated to Uni-Zap XR

vector DNA digested with EcoRI and XhoI, and *in vitro* packaging was carried out. The cDNA library contained 1.5×10^6 independent recombinants.

Screening of the library was performed on 200 000 recombinants by *in situ* hybridization as described by Sambrook and Russel (17) using XL1-blue *E.coli* bacteria and an EcoRI/ClaI fragment of the human *MYC* gene (exon 3) as a probe. More than 500 positive clones were detected by the probe and 50 of them were selected for further purification and analysis. Phage clones were purified by two additional rounds of *in situ* hybridization. Phagemids were then excised from each phage clone by recombination *in vivo* in the SolR *E.coli* bacteria and were used to infect XL1-blue *E.coli* bacteria.

Plasmids were extracted using a minipreparation technique based on alkaline lysis (17) and the cDNAs were analyzed by sequencing. Three primers were used: a 20mer oligonucleotide corresponding to the promoter of phage T3 RNA polymerase which is located upstream of the cDNA insert (AATTAACCCTCACTAAAGGG); a 22mer antisense primer corresponding to the promoter of phage T7 RNA polymerase which is located downstream of the cDNA insert (GTAATACGACTCACTATAGGGC); a 20mer antisense primer whose sequence is located at the beginning of exon 2 of the human *MYC* gene (TCCTCCTCGTCGCAGTAGAA). This last primer allowed the determination of the cDNA sequences located upstream of *MYC* exon 2. The sequences were analyzed by comparison with the sequence of human *MYC* mRNA (Align software from Scientific and Educational Software) and by comparison with nucleic acid sequences in public database libraries (program Blast).

Nucleic acid extraction and analysis

Nuclear, cytoplasmic and total cellular RNAs were prepared as described by Weil *et al.* (18). Polyadenylated RNA was purified by chromatography on an oligo(dT)-cellulose column (Oligo(dT)-cellulose type 7 from Amersham Pharmacia Biotech Inc) as previously described (19). Analysis by agarose gel electrophoresis and northern blotting were performed as previously reported (20). Genomic DNA was extracted and analyzed by Southern blotting as described previously (21).

Amplification by the PCR was performed on 1 µg of genomic DNA with 2.5 U of *Taq* polymerase (Bioprobe) and 100 pmol of each primer in a 100-µl volume using the reaction buffer provided by the supplier. The sense primer (CCAGGTACCTAGCGGTT) was designed from the sequence of the first internal transcribed spacer (ITS1) of the human rDNA transcription unit and the antisense primer (CTCCCATCTTGACAAGTC) from that of the first intron of the human *MYC* gene. The conditions for PCR were 1 min at 95°C, 1 min at 60°C and 1 min at 72°C for 30 cycles. Reverse transcription reactions were carried out as follows: polyA⁺ RNA (5 µg) and random hexamers (400 ng) were mixed in the presence of 100 mM Tris-HCl pH 8.0 and 50 mM NaCl (final volume 10 µl), heated for 3 min at 70°C, and left to anneal by cooling for 30 min at room temperature. The mixture was then adjusted to a volume of 30 µl containing 50 mM Tris-HCl pH 8.0, 16.6 mM NaCl, 37.5 mM KCl, 5 mM MgCl₂, 10 mM DTT, 125 µM each dNTP, 100 µg/ml bovine serum albumin, 1500 U/ml rRNasin (Promega), and 1000 U of M-MLV reverse transcriptase (Promega) and incubated for

1 h at 37°C. One-tenth of the reaction mixture was used for each PCR assay which was performed in a 100- μ l volume under the conditions described above but for 40 cycles. The sense (ACGGAGATCACCATCGTCAA) and antisense (TGCTGCTGCTGGTAGAAGTT) primers were designed from the sequence of exon 2 of the human *MSF* and *MYC* genes, respectively. Aliquots (40 μ l) of the amplification reactions were analyzed by agarose gel electrophoresis.

Fluorescent *in situ* hybridization

Chromosome metaphase spreads were prepared from cultured SW613-2G1 and 2G1MycP2Tu1 cells, using standard methods. Lymphocytes from normal individuals were used as a control. Fluorescent *in situ* hybridization (FISH) experiments were performed as previously described (22). The *MYC* probe was prepared from cosmid K880, which contains a 35 kb human genomic insert encompassing the *MYC* gene (23). Purified cosmid DNA was labeled by nick translation in the presence of digoxigenin-dUTP, mixed with Cot1 human DNA and sonicated salmon sperm DNA, and used for hybridization. After washing, the slides were incubated with a Rhodamine-labeled anti-digoxigenin antibody (Qbiogene), yielding a red fluorescent signal. This probe reveals the expected locus (8q24) on normal chromosome metaphases. The rDNA genomic probe was derived from BAC clone CIT-HSP 2505H6 obtained from the Genoscope (Evry, France). This BAC clone contains several copies of the whole 43 kb rDNA unit (18S-5.8S-28S-intergenic spacer). Purified BAC DNA was labeled by random priming in the presence of Alexa 488-dUTP (Molecular Probes), producing a green fluorescent signal. On normal metaphases, this probe hybridized to the short arm of all acrocentric chromosomes (13,14,15,21 and 22), yielding signals of variable intensities, depending on the locus. For each hybridization experiment, ~20 metaphase images were acquired with a Vysis station (Downers Grove, IL, USA) using the Quips Smart Capture FISH Imaging software.

RESULTS

Chimeric transcripts in 2G1MycP2Tu1 cells

Stable transfectants were obtained after transfection of cells from clone 2G1 (displaying a low level of amplification of *MYC*) with a plasmid carrying an 8 kb fragment of genomic DNA encompassing the *MYC* gene. Two of these transfectants produce large amounts of *MYC* mRNAs and proteins of abnormal size. We wondered whether these abnormal mRNAs were responsible for the abnormal pattern of *MYC* proteins observed in these cells, and whether this unusual pattern of expression of the *MYC* gene has a role in the phenotypic properties acquired by these cells. To address these questions, we analyzed in detail the structure of the abnormal *MYC* mRNAs produced by one of the two transfectants, the 2G1MycP2Tu1 cell line.

A cDNA library constructed using RNA isolated from 2G1MycP2Tu1 cells was screened with a *MYC* probe (third exon). Fifty positive clones were sampled, purified and analyzed by DNA sequencing. The isolated *MYC* cDNAs could be classified into three groups. The cDNAs of the first group

(11 clones) consist of *MYC* exons 2 and 3 fused to upstream rDNA sequences. The second group (35 clones) contains chimeric cDNAs composed of exons 2 and 3 of the *MYC* gene fused in 5' to one or several exons from another gene. The third group is represented by cDNAs containing a truncated *MYC* exon 2 of variable length spliced to *MYC* exon 3. The latter were probably derived from mRNAs that had not been fully retro-transcribed and, as such, were not studied further.

2G1MycP2Tu1 cells harbor a chimeric rDNA-*MYC* gene

The rDNA-*MYC* fusion cDNAs were of two different types. One type is represented by clones 13 and 21 (Figure 1A). The rDNA sequences found upstream of the *MYC* exons are composed of a truncated ITS1 (Internal Transcribed Spacer 1) region and, in the case of clone 21, of an almost complete 18S region. The 18S and ITS1 sequences of clone 21 are contiguous like in a normal rDNA transcription unit. These rDNA sequences are fused to a truncated *MYC* exon 1, which is itself spliced to exon 2. The latter is properly spliced to *MYC* exon 3. A likely explanation for these observations is that integration of a copy of the *MYC* plasmid occurred in the ITS1 region of one of the rDNA transcription units of 2G1MycP2Tu1 cells. The junction points, localized at position 6163 of the rDNA transcription unit and at position 2855 of the *MYC* gene, have no noticeable feature, suggesting that the fusion occurred by illegitimate recombination following transfection.

To confirm this hypothesis, we ought to identify the expected junction in the genomic DNA of 2G1MycP2Tu1 cells. An amplification by PCR was carried out using as a template genomic DNA extracted from cells of the 2G1MycP2Tu1 and 2G1 cell lines. The sense primer was located in the ITS1 region and the antisense primer hybridized to *MYC* intron 1 (Figure 2A). As predicted, a fragment of 273 nt was amplified from 2G1MycP2Tu1 DNA but not from 2G1 DNA (Figure 2B). As a positive control, a fragment corresponding to the *APC* gene could be amplified from both DNAs with the appropriate primers. The PCR fragment obtained with 2G1MycP2Tu1 DNA was cloned and sequenced. Its sequence was identical to that of cDNA clones 13 and 21 in the region encompassing the junction point (ITS1 and *MYC* exon 1) (Figure 2A). To confirm that a copy of the *MYC* gene did integrate into one of the rRNA gene clusters, FISH experiments were performed on metaphase chromosomes from 2G1 and 2G1MycP2Tu1 cells (Figure 3). In the parental 2G1 cells, a *MYC* probe detects a unique copy of the *MYC* gene carried on a single chromosome 8 and amplified copies of the gene located in an HSR (homogeneously staining region). 2G1MycP2Tu1 cells have *MYC* copies integrated into two additional chromosomal sites: one on a marker chromosome (mkB) and one on the short arm of a derivative of chromosome 13 (13p+). The identity of this chromosome was checked by chromosome painting experiments using a chromosome 13-specific probe (data not shown). This last site of integration is one of the known locations of rRNA genes in human cells (24). FISH experiments carried out with an rDNA probe confirm the co-localization of *MYC* gene copies and rRNA genes at the 13p+ site (Figure 3). The *MYC* and rDNA signals also co-localize in FISH experiments carried out with interphase nuclei (data not shown), confirming the molecular proximity of the corresponding sequences.

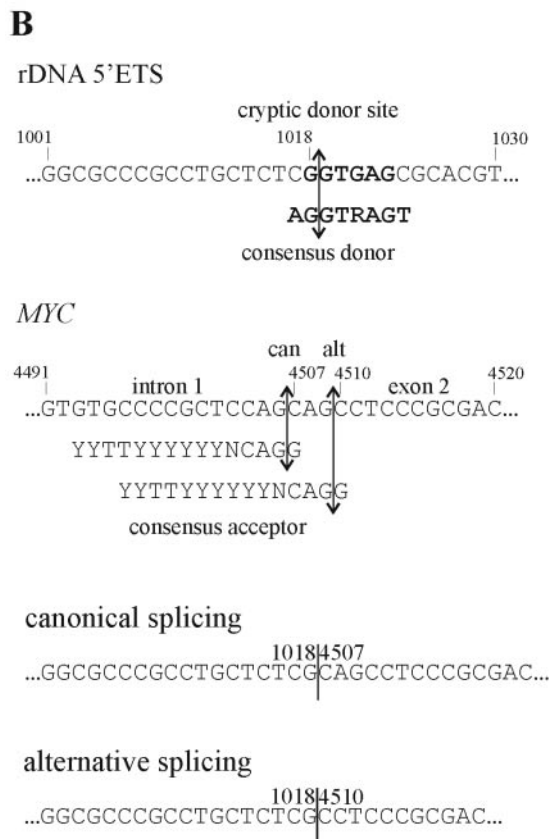


Figure 1. Structure of chimeric rDNA/MYC cDNAs. (A) Schematic representation of the structure of 11 cDNA clones deduced from sequencing analysis. Boxes in light and dark gray represent rDNA and MYC sequences, respectively. ITS1, internal transcribed spacer 1; 5'ETS, 5' external transcribed sequence; EX, exon; can and alt, canonical and alternative splice acceptor site of MYC exon 2. 'Δ' indicates that the corresponding sequence is truncated. MYC exon1Δ is only 27 nt long. Base pair coordinates are given relative to the transcription unit for rDNA sequences (accession no. U13369) and to the genomic sequence (accession no. D10493) for MYC. The dotted line limiting the 5'ETS box between positions 430 and 665 delineates the beginning of the longest and the shortest of the nine cDNA clones. (B) Sequences of the rDNA 5'ETS and of the MYC gene encompassing the cryptic splice donor site and the acceptor site of exon 2, respectively. Consensus sequences for donor and acceptor sites are shown below, with double-headed arrows indicating the cleavage sites. The lower part of the figure shows the sequences of hybrid mRNAs after either canonical or alternative splicing.

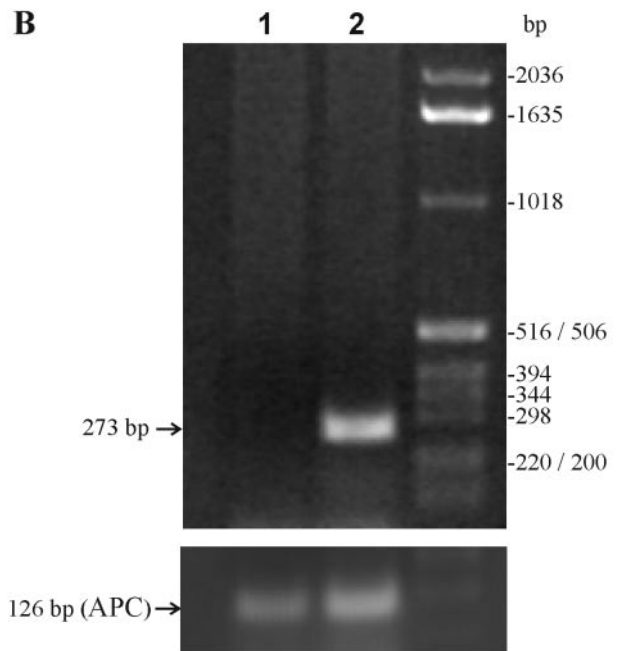
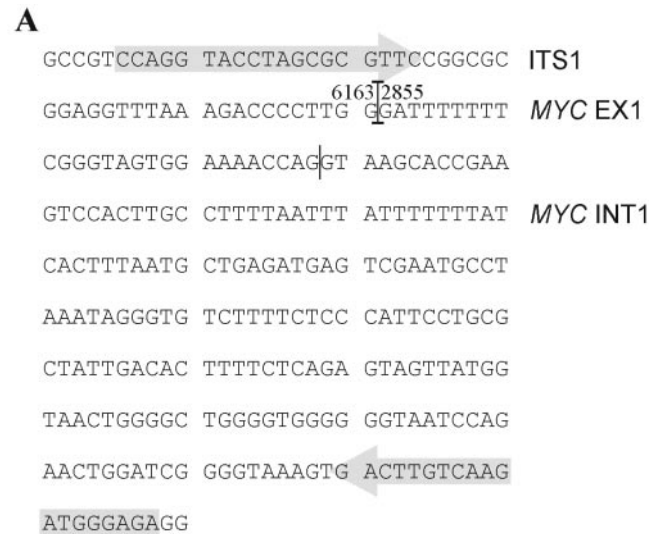


Figure 2. Detection of the MYC-rDNA junction by genomic PCR. (A) DNA sequence encompassing the junction point between the MYC gene and the rDNA transcription unit. Gray arrows indicate the position of the primers used for genomic PCR. Symbols and coordinates are as in Figure 1. The junction point between ITS1 and MYC exon 1 is indicated by an I-shaped line and the border between MYC exon 1 and intron 1 by a vertical line. (B) Genomic PCR was performed on DNA extracted from 2G1 (lane 1) or 2G1MycP2Tu1 cells (lane 2). Arrows point to the 273 bp amplified fragment obtained using the primers shown in (A) and to the 126 bp fragment amplified from the APC gene. The size of the DNA fragments used as molecular weight markers is indicated on the right.

We conclude from all these observations that copies of the MYC gene are inserted into an rDNA transcription unit.

Structure of rRNA-MYC fusion transcripts

The second type of rDNA-MYC cDNAs is represented by nine different clones (Figure 1A). The 5' region of these cDNAs consists of sequences of various length derived from the 5'ETS

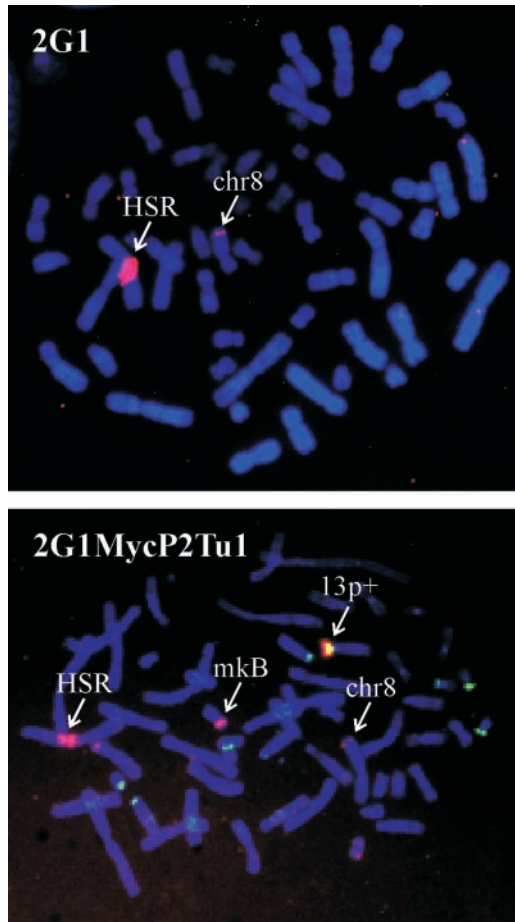


Figure 3. Chromosomal localization of *MYC* gene copies in 2G1 and 2G1MycP2Tu1 cells. FISH experiments were performed with a *MYC* probe (red) on 2G1 (parental cells) and 2G1MycP2Tu1 (transfectant) metaphases and with an rDNA probe (green) on 2G1MycP2Tu1 metaphases. chr8: resident *MYC* gene; HSR: amplified copies of the *MYC* gene localized in an HSR; 13p+: co-localization of the exogenous copies of the *MYC* gene and of rRNA genes on a rearranged chromosome 13 short arm; mkB: insertion of exogenous copies of the *MYC* gene into an unidentified marker chromosome. Note that the rDNA probe reveals, as expected, additional rDNA copies on other acrocentric chromosomes (green spots).

(5' external transcribed spacer) region of rDNA. In the different clones, this 5'ETS sequence is more or less truncated in 5' (starting at position 430–665) but, in all clones, it ends at position 1018 and is fused to *MYC* exon 2. The latter is properly spliced to *MYC* exon 3. The human *MYC* gene has two splice acceptor sites at the 5' end of exon 2: one so-called 'canonical' site (at position 4507) and one 'alternative' site (at position 4510). Exon 1 is spliced to the canonical versus alternative splice site of exon 2 in ~70% and 30% of *MYC* mRNAs, respectively (25). For 8 out of the 9 cDNA clones, the 5'ETS sequence is exactly joined to the canonical site of *MYC* exon 2 and for the remaining one, it is joined to the alternative site. This strongly suggests that the corresponding transcripts have been generated by a splicing event occurring between a cryptic splice donor site in the 5'ETS region (at position 1018) and one of the *MYC* splice acceptor sites of exon 2 (Figure 1B). Indeed, the sequence of the 5'ETS region around position 1018 does match the consensus sequence for a

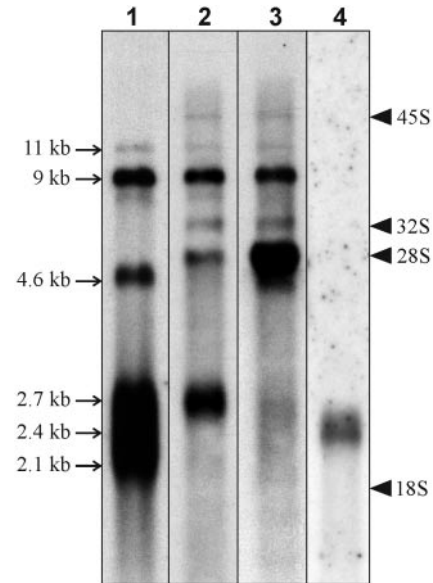


Figure 4. Abnormal *MYC* mRNAs in 2G1MycP2Tu1 cells. Polyadenylated RNA from 2G1MycP2Tu1 (lanes 1–3) or 2G1 (lane 4) cells was analyzed by northern blotting using a *MYC* exon 3 probe (lanes 1 and 4), a 5'ETS probe (lane 2) or an ITS1 probe (lane 3). Exposure time for lane 4 was 9 times longer than for the other lanes. Arrows on the left point to *MYC* mRNAs and arrowheads on the right indicate the position of rRNAs. Polyadenylated RNA preparations were not completely pure so that 45S, 32S and 28S rRNAs were visible in these samples with the rDNA probes. Detection of the 32S rRNA by the 5'ETS probe and of the 28S rRNA by the 5'ETS and ITS1 probes is due to cross-hybridization.

splice donor site (26) including the hallmark GT dinucleotide at positions 1019 and 1020.

Cytoplasmic polyadenylated RNA from 2G1MycP2Tu1 cells was analyzed by northern blotting using a *MYC* exon 3 probe (Figure 4). As expected, RNAs of abnormal size are detected (11, 9, 4.6 kb and a smear between 2.7 and 2.1 kb as compared to 2.4 kb for *MYC* mRNAs in parental 2G1 cells). On short exposures (data not shown), two clusters of RNAs at 2.1 and 2.7 kb can be distinguished within the smear. The 11 and 9 kb RNAs are revealed by both a 5'ETS and an ITS1 probe. The 4.6 and 2.7 kb bands are detected only by an ITS1 or a 5'ETS probe, respectively. A probe corresponding to *MYC* intron 1 hybridizes exclusively to the 11 kb RNA (Figure 5). These results indicate that chimeric rRNA-*MYC* RNA molecules are indeed synthesized in 2G1MycP2Tu1 cells and that these species are polyadenylated. This was confirmed by comparative quantitative analysis of total RNA, polyA⁺ RNA and polyA⁻ RNA from 2G1MycP2Tu1 cells by northern blotting with a *MYC* probe. The polyA⁺ fraction was enriched 12-fold in chimeric RNAs as compared to total RNA whereas these species were 2-fold less abundant in the polyA⁻ fraction than in total RNA. The corresponding figures for the keratin 18 mRNA, used as an internal control, were 13-fold and 3-fold, respectively (data not shown). The nucleocytoplasmic distribution of the chimeric RNAs was studied on northern blots prepared with cytoplasmic and nuclear RNA (Figure 5). The high molecular weight species (11, 9 and 4.6 kb) predominate in the nuclear fraction. In contrast, the 2.1–2.7 kb RNAs are more abundant in the cytoplasm.

Maturation pathways for rRNA-MYC fusion transcripts

From all the results presented above, we infer a model that can explain how the rRNA-MYC chimeric transcripts are

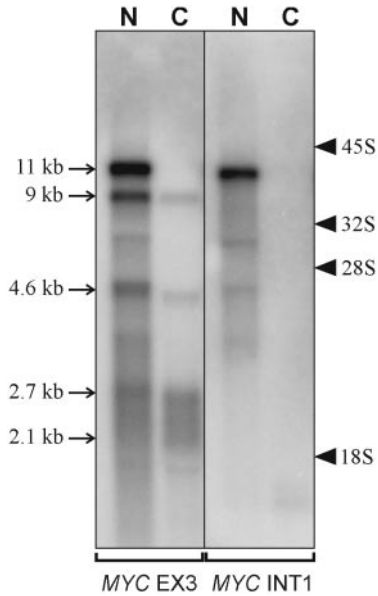


Figure 5. Nucleocytoplasmic distribution of chimeric MYC mRNAs, Northern blot analysis was performed on 10 µg of nuclear (lanes N) or cytoplasmic (lanes C) RNA extracted from 2G1MycP2Tu1 cells. An MYC exon 3 or intron 1 probe was used, as indicated. Symbols are as in Figure 4.

produced in 2G1MycP2Tu1 cells. A copy of the *MYC* gene has integrated into one of the rDNA transcription units of the genome (Figure 6A) at position 6163 within the ITS1 region and the *MYC* gene was truncated at position 2855, with deletion of the 5' upstream sequences (Figure 6B, first line). This resulted in a chimeric transcription unit whose primary transcript is ~11 000 nt long. This primary transcript could correspond to the 11 kb RNA observed on northern blots, which is detected by the 5'ETS, ITS2, MYC intron1 and MYC exon 3 probes and which is essentially located in the nucleus. From this primary transcript, two maturation pathways are possible. In the first one, both introns of the *MYC* gene are spliced out and a cleavage occurs between the 5'ETS and 18S regions, as it is the case during maturation of the normal 45S ribosomal RNA precursor (Figure 6B, pathway 1). This leads to the production of a chimeric 4.3 kb RNA from which the cDNA clones 13 and 21 would be derived. This RNA could correspond to the 4.6 kb band observed on northern blots which is revealed by both the ITS1 and MYC exon 3 probes (Figure 4). This RNA species is inefficiently transported to the cytoplasm. It is also probably very inefficiently translated, if at all, since its 5' untranslated region is very long and composed of 18S and ITS1 sequences which are known to fold into many stem-loop secondary structures (27).

In the second maturation pathway, the primary transcript is spliced between a cryptic donor site located in the 5'ETS region (at position 1018) and MYC exon 2 acceptor sites (canonical or alternative). The MYC intron 2 is spliced out in a normal way (Figure 6B, pathway 2). The resulting chimeric RNA is composed of the 5' half of the 5'ETS region

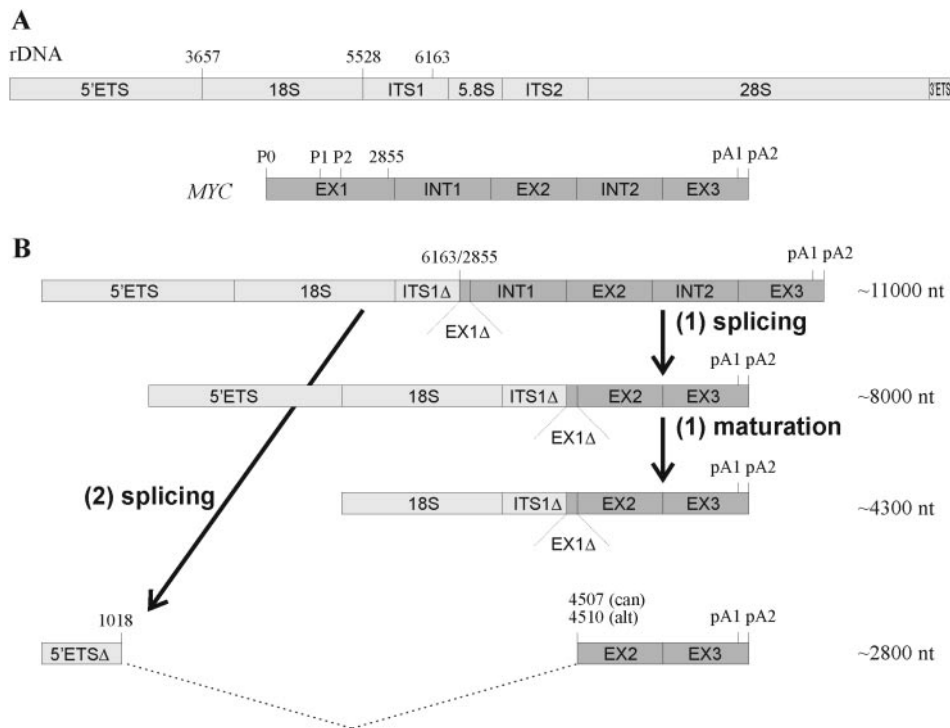


Figure 6. Possible maturation pathways for the hybrid rRNA/MYC transcripts. (A) The structures of the rDNA transcription unit and of the *MYC* gene are represented respectively by light and dark gray boxes. The position of the *MYC* gene promoters (P0, P1 and P2) and polyadenylation sites (pA1 and pA2) is indicated. (B) Two possible maturation pathways (1 and 2) are indicated by bold arrows. The approximate calculated size of the different transcripts is given on the right. 3'ETS, 3' external transcribed spacer. Other symbols and coordinates are as in Figure 1.

spliced to MYC exons 2 and 3 and is roughly 2.8 kb long. The nine cDNAs with such a structure (Figure 1A) would be derived from this transcript. This RNA could correspond to the 2.7 kb band revealed by both the 5'ETS and MYC exon 3 probes on northern blots (Figure 4). This species is efficiently exported to the cytoplasm.

Transcription of the chimeric rDNA-MYC gene by RNA polymerase I

The existence of a chimeric rDNA-MYC gene in 2G1MycP2Tu1 cells raises the question of the nature of the RNA polymerase that transcribes it. From the structure of this chimeric gene (Figure 6B), one would expect that RNA polymerase I drives its expression. To address this question, the sensitivity to actinomycin D of the polymerase responsible for the synthesis of abnormal MYC mRNAs was assayed. 2G1MycP2Tu1 cells were treated for seven hours with the drug and the level of MYC mRNAs was analyzed on northern blots (Figure 7A). The synthesis of these mRNAs was inhibited by 80–95% in the presence of 0.03 $\mu\text{g}/\text{ml}$ of actinomycin D. Such a sensitivity to low doses of this drug is characteristic of RNA polymerase I transcription (28). Accordingly, a similar dose–response curve was observed for the 45S ribosomal RNA precursor (Figure 7B). In contrast, RNA polymerase II transcription requires higher doses of actinomycin D for inhibition, as exemplified by *FGF3* (fibroblast growth factor) gene expression for which an inhibition of 85% is reached at no less than 1 $\mu\text{g}/\text{ml}$ of actinomycin D. It is also noteworthy that 2G1MycP2Tu1 cells produce a small amount of apparently normal MYC mRNAs (2.4 kb) whose accumulation is insensitive to low doses of actinomycin D. These mRNAs are presumably derived from the resident gene copies which are transcribed by RNA polymerase II. From all these results, we conclude that the RNAs derived from the chimeric rDNA-MYC gene in 2G1MycP2Tu1 cells are synthesized by RNA polymerase I.

Trans-spliced mRNAs in 2G1MycP2Tu1 cells

The second group of chimeric MYC cDNAs comprises 35 different clones. They are composed of one or more exons coming from another gene, spliced to MYC exon 2 which is itself normally spliced to exon 3. Such a gene was dubbed 'exon-donor'. Out of the 35 cDNAs analyzed, we identified 33 different exon-donor genes which are located on 14 different chromosomes (Table 1). The non-MYC exons are spliced either to the canonical or to the alternative acceptor splice site of MYC exon 2. For each exon-donor gene, the genomic sequence was examined and, in all cases, the sequence at the junction point matched the consensus sequence for a splice donor site (data not shown). These observations indicate that the mRNAs corresponding to the 33 different cDNA clones were generated by a bona fide splicing mechanism.

A possible explanation would be that extensive integration of introduced MYC gene copies occurred at many chromosomal sites in 2G1MycP2Tu1 cells. At each of these sites, the MYC gene copy would have integrated in such a way that a chimeric exon-donor-MYC gene would be created, similar to the chimeric rDNA-MYC gene described above. Chimeric transcripts could then be generated by *cis*-splicing of the primary transcripts produced by such genes. This hypothesis

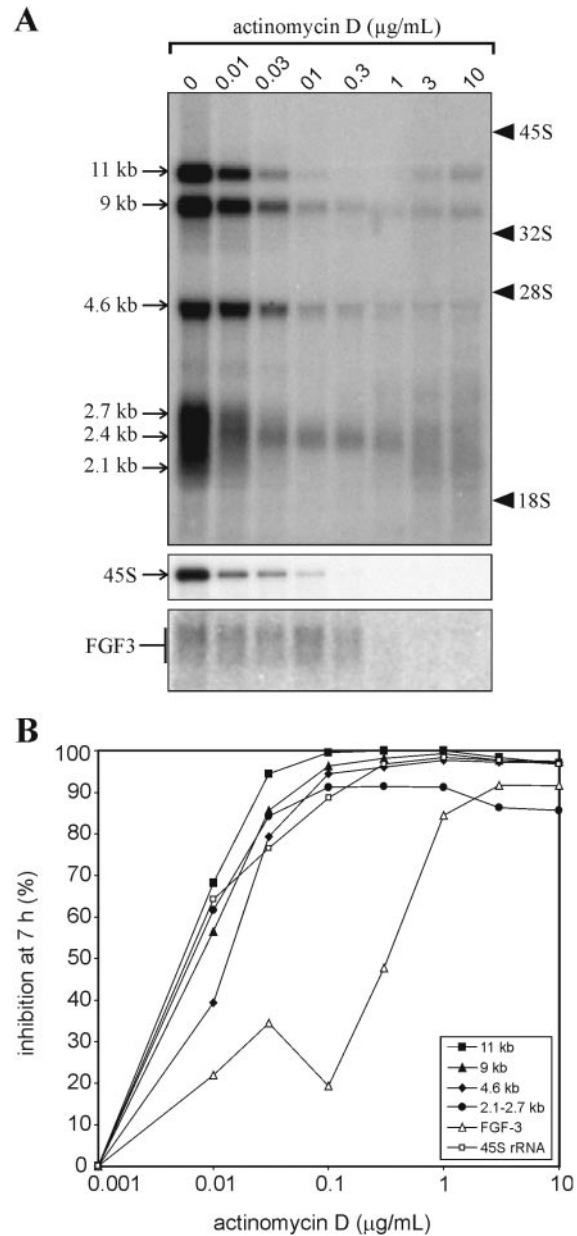


Figure 7. Sensitivity of MYC mRNA synthesis to actinomycin D in 2G1MycP2Tu1 cells. (A) 2G1MycP2Tu1 cells were incubated for 7 h in the presence of the indicated concentrations of actinomycin D. Total cellular RNA was extracted and analyzed by northern blotting using an MYC exon 3 probe (upper panel), an rDNA 5'ETS probe which reveals the 45S ribosomal RNA (middle panel) or an *FGF3* probe (lower panel). Symbols are as in Figure 4. (B) Quantification of the results shown in (A). Hybridization signals were quantified as described (30), and the results are expressed as the percentage of inhibition of the synthesis of each RNA species relative to the zero time point.

seemed highly unlikely to us but experiments were devised to formally rule it out. If a copy of the MYC gene has integrated into an exon-donor gene downstream of the exon which is spliced to MYC exon 2, this should result in a rearrangement of one allele of the gene. The structure of three 'exon-donor' genomic loci in 2G1 and 2G1MycP2Tu1 cells was analyzed by Southern blotting: the laminin gamma C3 gene (*LAMC3*), the DNA polymerase α gene (*POLA2*) and a gene coding for a putative protein (*FLJ14775*) (Figure 8A). In each case,

Table 1. Features of exon-donor genes

Chromosomal localization ^a	Gene name ^b	Spliced exons ^c	MYC splicing site ^d
1p34.2	NFYC	ex 2	can
1p36.13	AJ749659	One exon	can
1p36.31	ESPEN	ex 7b	can
1p36.33			
2p11.2	AJ749662	One exon	alt
2q11.2	MAP4K4	ex 2	can
2q14.2	RALB	ex 1	can
3p22.2	ACVR2B	ex 1	can
3p23	GLB1	ex 1	can
4p16.3	KIAA1530	ex 1 to 7	can
4q24	AJ749663	One exon	can
5p15.33	BF508938 (EST)	ex 1	alt
5q31.1	TCF7	ex 2	can
5q31.3	AJ749665	One exon	alt
7q21.2	CDK6	ex 1	alt
7q21.2	CDK6	ex 1	can
7q22.1	FLJ12298	ex 1+2+4+5	can
9q22.2	CKS2	ex 1	alt
9q34.13	LAMC3	ex 1	alt
11q13.1	POLA2	ex 1+9 to 17	can
11q13.2	LRP5	ex 6	can
14q32.11	CHES1	ex 1, 1b and 2	alt
14q32.33	BF309232 (EST)	ex 1	can
16p12.1	AJ749661	Two exons	can
16p13.3	FLJ14154	ex 1	can
16p13.3	PKD1	ex 1	can
17q25.3	MSF	ex 2	can
17q25.1	FLJ14775	ex 1	can
19p13.3	BC014506 (mRNA)	ex 2+3	can
21q22.3	PDXK	ex 1b+ex 2	can
21q22.3	D21S2056E	ex 9 to 12	can
22q11.1	FLJ 31573	ex 7	alt
Repeated sequence		Four exons	alt
Repeated sequence		One exon	can

^aAccording to the UCSC Genome Browser, July 2003 assembly (<http://genome.ucsc.edu>). The distance between some locations is given in megabases (Mb) and as a percentage of the length of the corresponding chromosome (NCBI built 34).

^bFor exons corresponding to ESTs or mRNAs previously described but not presently assigned to a known gene, the accession number is indicated. Accession numbers for the sequences determined in this work are EST BF508938; AJ749664; EST BF309232; AJ749660. Accession numbers in italics refer to exons that were not previously found in an EST or mRNA and are localized in a region of the human genome with no recorded gene (July 2003 assembly).

^cExons (ex) denoted by b have not been reported previously for the corresponding gene. Their number is derived from that of the preceding exon. Accession numbers for the corresponding sequences are: ESPEN: AJ749667; CHES1: AJ749666; PDXK: AJ749668.

^dcan: canonical; alt: alternative. Note that for the CDK6 gene two different cDNAs were isolated, corresponding to canonical or alternative splicing on MYC exon 2. Two cDNA clones with an identical structure were isolated for the *NFYC* gene.

the restriction enzyme was chosen so as to generate a restriction fragment encompassing the whole intron containing the putative site where insertion of the *MYC* gene should have occurred. In all three cases, the alleles of the exon-donor gene are in a germ-line configuration in the region analyzed, in both 2G1 and 2G1MycP2Tu1 cells. These results do not support the integration hypothesis, at least for these three genes. Furthermore, FISH experiments performed with a *MYC* probe indicated that there are only two new *MYC* gene integration sites

in 2G1MycP2Tu1 cells, as compared to 2G1 cells (Figure 3). This is not consistent with our finding that the non-*MYC* exons originate from 33 different exon-donor genes.

An alternative possibility would be that recombination events leading to cellular heterogeneity take place in the 2G1MycP2Tu1 cell line, resulting in the appearance of subpopulations of cells bearing genomic fusions of the *MYC* gene. If these subpopulations are minority, they would not be detected by Southern or FISH analysis. To address this question, we derived sub-clones from the 2G1MycP2Tu1 cell line and attempted to detect the presence of the MSF-*MYC* fusion transcript (Table 1) in these cells. Reverse transcription-PCR reactions carried out with polyA⁺ RNA prepared from three sub-clones revealed that they all produce this chimeric transcript (Figure 8B). This result does not support the hypothesis that this transcript is produced by a minority subpopulation of 2G1MycP2Tu1 cells harboring an exogenous *MYC* gene integrated in the resident *MSF* gene. The most likely explanation for all these observations is that the chimeric *MYC* mRNAs with extraneous exons were generated by splicing events occurring in *trans*, between pre-mRNAs derived from one of the exon-donor gene and from a *MYC* gene copy. Therefore, we conclude that *trans*-splicing occurs at high frequency in cells of the 2G1MycP2Tu1 cell line.

DISCUSSION

Uncoupling of pre-mRNA processing from transcription

We have found that the 2G1MycP2Tu1 cell line, isolated following transfection of 2G1 cells with a plasmid carrying the *MYC* gene, harbors a truncated copy of this gene integrated into an rDNA transcription unit. This results in the production of chimeric rRNA-*MYC* transcripts that are synthesized by RNA polymerase I. These transcripts are of two different types, depending on the rRNA sequences found at their 5' end (5'ETS or 18S/ITS1). We propose that they are derived from the chimeric primary transcript by two different maturation pathways (Figure 6) involving normal splicing of the *MYC* introns and maturation by cleavage of the rRNA sequences (pathway 1) or splicing between a cryptic donor site in the 5'ETS sequence and one of the acceptor sites of *MYC* exon 2 (pathway 2). Since the putative end-products of these two pathways (4.6 and 2.7 kb mRNAs, respectively) accumulate to comparable amounts in the cells, the 5'ETS cryptic donor site appears to be used as efficiently as the donor site of *MYC* exon 1 during maturation of the chimeric transcript. Although the intrinsic efficiency of the 5'ETS site is unknown, it is possible that the efficiency of the natural donor site of *MYC* exon 1 is impaired because of the proximity (27 nt) of the junction point with rRNA sequences which might fold into highly ordered secondary structures. Another non-exclusive possibility is that splicing is not coupled with transcription during the synthesis of these transcripts because the splicing machinery cannot associate with RNA polymerase I. Thus, splicing would necessitate relocation of the primary transcript, presumably from the nucleolus to the splicing machinery in the nucleoplasm and the use of the cryptic site could be enhanced during this process.

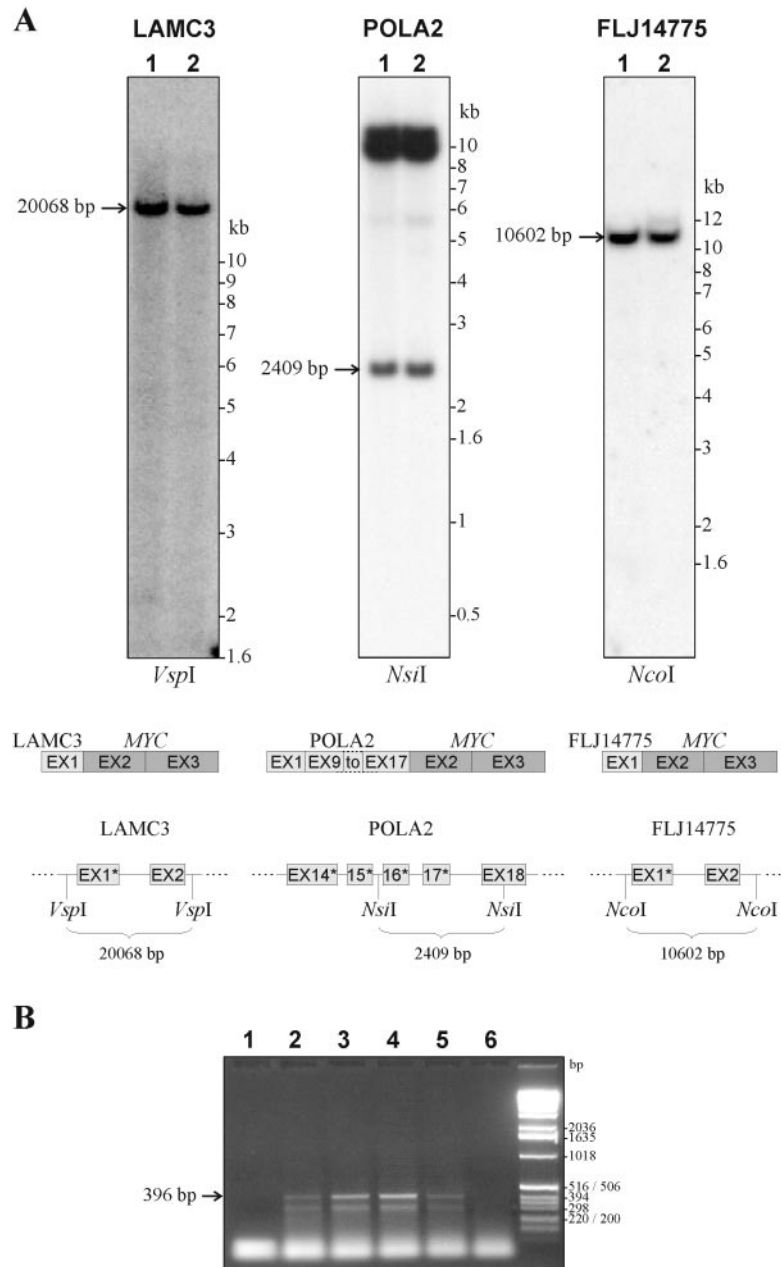


Figure 8. Analysis of four genomic loci in 2G1 and 2G1MycP2Tu1 cells. **(A)** Genomic DNA from 2G1 (lane 1) or 2G1MycP2Tu1 (lane 2) cells was analyzed by Southern blotting using a laminin gamma 3 (LAMC3) probe, a DNA polymerase alpha subunit B (POLA2) probe or a probe corresponding to the gene encoding the putative protein FLJ14775. The restriction enzyme used in each case is indicated at the bottom of each panel. Arrows point to the hybridizing fragment of interest whose size is indicated in base pairs. Note that the POLA2 probe reveals additional genomic DNA fragments since it contains several exons. The position of the DNA fragments used as molecular weight markers is indicated on the right. The structure of the corresponding chimeric cDNAs is shown with MYC exons 2 and 3 symbolized by dark gray boxes and exons from the exon-donor genes (LAMC3, POLA2 or FLJ14775) by light gray boxes. The genomic structure of these three genes in the vicinity of the exons found in the chimeric cDNAs is schematized at the bottom of the figure with the position of the restriction enzyme sites of interest and the size of the resulting fragments indicated below. The exons marked with a star are those included in the cDNA fragment used as a probe. **(B)** PolyA⁺ RNA prepared from 2G1MycP2Tu1 (lane 2), 2G1MycP2Tu1-clone 2, -clone 3, -clone 4 (lanes 3–5), and 2G1 (lane 6) cells was reverse-transcribed in the presence of random hexamers. Amplification by PCR was carried out in the absence (lane 1) or in the presence (lanes 2–6) of the corresponding cDNAs. Arrow points to the 396 bp amplified fragment obtained using primers located in exon 2 of the human *MSF* and *MYC* genes. The size of the DNA fragments used as molecular weight markers is indicated on the right.

It is still unclear whether polyadenylation can be uncoupled from transcription *in vivo*. Few data are available and contradictory results have been reported (see Introduction). Since the chimeric rRNA-MYC transcripts produced by 2G1MycP2Tu1 cells are polyadenylated and since there is no evidence that

RNA polymerase I can associate with 3' end processing factors, our results are in favor of a possible uncoupling of the two processes. It is worth mentioning that these results were obtained in a system where the studied gene is stably integrated into a chromosomal site whereas previously

reported results (3,5) were obtained with transient expression assays.

As far as we know, there is no data available on the possible uncoupling of splicing from transcription *in vivo*. Such a situation may exist in trypanosomes since splicing of RNA polymerase I transcripts has been reported in these organisms [reviewed in (6)] and since this polymerase is presumably unable to recruit the splicing machinery but this concerns a peculiar situation of *trans*-splicing. We provide here evidence that the chimeric rRNA-MYC transcripts synthesized by RNA polymerase I in 2G1MycP2Tu1 cells are processed in these human cells through bona fide splicing events. Thus, our results indicate that splicing may be uncoupled from RNA polymerase II transcription in mammalian cells.

High frequency *trans*-splicing in human cells

Our analysis of the structure of cDNA clones reveals the presence in 2G1MycP2Tu1 cells of many different hybrid transcripts consisting of one or more 5' exons, originating from more than 30 different exon-donor genes and correctly spliced to MYC exon 2, itself properly spliced to exon 3. The donor site of the foreign exon is joined to one or the other of the two functional alternative acceptor sites of exon 2 of the human MYC gene. We propose that these molecules are generated by *trans*-splicing events occurring at a relatively high frequency in these cells. Indeed, we have ruled out the possibility that numerous different integration events took place in these cells. The cDNA clones corresponding to mRNAs generated by *trans*-splicing were as abundant in the library as the chimeric rDNA-MYC clones. Since the hybrid rRNA-MYC mRNAs accumulate to a high level in 2G1MycP2Tu1 cells (see Figure 4), *trans*-splicing events should occur at a high frequency in these cells. On northern blots, a MYC probe detects a group of abundant mRNAs with an average size of 2.1 kb (Figure 4). These molecules are not revealed by any of the rDNA probes used. They probably correspond to at least part of the hybrid mRNAs generated by *trans*-splicing. Can these hybrid mRNAs be translated into chimeric proteins when open reading frames are joined in phase? It is likely because western blots prepared with protein extracts from these cells and probed with an anti-MYC antibody disclose the presence of a group of abnormal MYC proteins of higher molecular weights (C. Chen, unpublished results). These proteins are not present in the parental 2G1 cell line.

Are the *trans*-spliced MYC mRNA molecules derived from the same MYC gene copy and primary transcripts as the hybrid rRNA-MYC mRNAs? This seems most likely to us since the synthesis of the 2.1 kb mRNAs is as sensitive to actinomycin D inhibition as that of the rRNA-MYC transcripts (Figure 7A). In this case, *trans*-splicing would also be uncoupled from transcription in 2G1MycP2Tu1 cells, as postulated above for 5'ETS-MYC mRNA *cis*-splicing. As mentioned above, one may speculate that this uncoupling is a consequence of a relocalization of the nucleolar primary transcript to the splicing machinery, a process that could favor the occurrence of splicing events in *trans*. It is also possible that the inefficiency of the *cis*-splicing reaction involving the MYC exon 1 donor site, due to the proximity of the junction point with rRNA sequences, favors splicing in *trans*. The high level of chimeric rRNA-MYC primary transcripts in 2G1MycP2Tu1 cells, due

to the strength of the rDNA promoter may also contribute to increase the frequency of *trans*-splicing. Finally, it is also possible that a cryptic sequence element present in the rRNA moiety of the chimeric rRNA-MYC primary transcripts enhances the efficiency of *trans*-splicing. Sequence elements playing a role in *trans*-splicing have been described (11,29). A better knowledge of the elements of the chimeric rDNA-MYC gene that are responsible for the high efficiency of *trans*-splicing could help in the design of a vector aimed at RNA reprogramming strategies, in particular those using spliceosome-mediated RNA *trans*-splicing (14).

The foreign exons spliced to MYC exon 2 are most probably derived from primary transcripts synthesized by RNA polymerase II on the exon-donor genes and, therefore, are capped at their 5' end. The situation in 2G1MycP2Tu1 cells is reminiscent of that in trypanosomes. In these organisms, the maturation of uncapped transcripts synthesized by RNA polymerase I naturally occurs by *trans*-splicing (6). This allows these transcripts to have a capped 5' end which is necessary for efficient translation. It has been reported that uncapped mRNAs synthesized by RNA polymerase I in yeast are very inefficiently translated (4). We have identified more than 30 exon-donor genes and many more must exist since most of them were identified through the isolation of a single cDNA clone. Are *trans*-splicing events occurring at random in 2G1MycP2Tu1 cells and, if so, is the selection of foreign exons only governed by the probability of encounter, i.e. the abundance of the exon-donor gene primary transcripts? Most probably not. Indeed, we did not isolate any hybrid cDNAs containing exons from genes known to be expressed at a high level in our cells such as the keratin 8 and 18 genes, the β -actin gene or the ferritin-H gene. In contrast, we isolated cDNA clones containing exons from genes that are likely to be expressed at much lower levels: transcription factors NFYC or TCF7, DNA polymerase α (POLA2), for example. An important parameter for the selection of the source of foreign exons could be the regional intranuclear localization of the genes. Indeed, in eight cases we identified exon-donor genes which are close to one another on a chromosomal scale: they are separated by a distance representing from <2% to ~5% of the total length of the chromosome (Table 1). Furthermore, it is noteworthy that three exon-donor genes have a pericentromeric localization and eight of them a subtelomeric one. It is therefore possible that the regional organization of the genes inside the nucleus, which governs the localization of the primary transcripts, plays a role in the selection of the genes that can be donors of exons in *trans*-splicing reactions. Horiuchi *et al.* (10) proposed that, in *Drosophila*, *trans*-splicing occurs locally and is dependent on chromosome proximity.

ACKNOWLEDGEMENTS

We thank Hugues Roest Crolius and Gabor Gyapay (Genoscope, Evry, France) for advice and for providing us with the rDNA genomic BAC clone. We are grateful to Christian Lavielle (Laboratoire de Génétique Oncologique) for gift of materials and for critical reading of the manuscript, and to Joseph Mautner (Munich, Germany) for the gift of cosmid K880. This work was supported by grants from the Comité de Recherche de l'Institut Gustave Roussy and from

the CNRS (Action Incitative Puces à ADN) to O.B. Funding to pay the Open Access publication charges for this article was provided by the Centre National de la Recherche Scientifique.

Conflict of interest statement. None declared.

REFERENCES

- Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
- Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.
- Grummt, I. and Skinner, J.A. (1985) Efficient transcription of a protein-coding gene from the RNA polymerase I promoter in transfected cells. *Proc. Natl Acad. Sci. USA*, **82**, 722–726.
- Lo, H.J., Huang, H.K. and Donahue, T.F. (1998) RNA polymerase I-promoted HIS4 expression yields uncapped, polyadenylated mRNA that is unstable and inefficiently translated in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **18**, 665–675.
- Smale, S.T. and Tjian, R. (1985) Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Mol. Cell. Biol.*, **5**, 352–362.
- Lee, M.G.-S. and Van der Ploeg, L.H.T. (1997) Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu. Rev. Microbiol.*, **51**, 463–489.
- Liang, X.H., Haritan, A., Uliel, S. and Michaeli, S. (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell*, **2**, 830–840.
- Nilsen, T.W. (1993) Trans-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.*, **47**, 413–440.
- Dorn, R., Reuter, G. and Loewendorf, A. (2001) Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in *Drosophila*. *Proc. Natl Acad. Sci. USA*, **98**, 9724–9729.
- Horiuchi, T., Giniger, E. and Aigaki, T. (2003) Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, *lola*. *Genes Dev.*, **17**, 2496–2501.
- Caudevilla, C., Codony, C., Serra, D., Plasencia, G., Roman, R., Graessmann, A., Asins, G., Bach-Elias, M. and Hegardt, F.G. (2001) Localization of an exonic splicing enhancer responsible for mammalian natural trans-splicing. *Nucleic Acids Res.*, **29**, 3108–3115.
- Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R. and Maniatis, T. (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell*, **10**, 21–33.
- Zhang, C., Xie, Y., Martignetti, J.A., Yeo, T.T., Massa, S.M. and Longo, F.M. (2003) A candidate chimeric mammalian mRNA transcript is derived from distinct chromosomes and is associated with nonconsensus splice junction motifs. *DNA Cell Biol.*, **22**, 303–315.
- Mansfield, S.G., Chao, H. and Walsh, C.E. (2004) RNA repair using spliceosome-mediated RNA trans-splicing. *Trends Mol. Med.*, **10**, 263–268.
- Donzelli, M., Bernardi, R., Negri, C., Prosperi, E., Padovan, L., Lavielle, C., Brison, O. and Scovassi, I. (1999) Apoptosis-prone phenotype of human colon carcinoma cells with a high level amplification of the *c-myc* gene. *Oncogene*, **18**, 439–448.
- Lamonerie, T., Lavielle, C., Haddada, H. and Brison, O. (1995) IGF-2 autocrine stimulation in tumorigenic clones of a human colon carcinoma cell line. *Int. J. Cancer*, **61**, 587–592.
- Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Weil, D., Brosset, S. and Dautry, F. (1990) RNA processing is a limiting step for murine tumor necrosis factor beta expression in response to interleukin-2. *Mol. Cell Biol.*, **10**, 5865–5875.
- Modjtahedi, N., Lavielle, C., Poupon, M.F., Landin, R.M., Cassingena, R., Monier, R. and Brison, O. (1985) Increased level of amplification of the *c-myc* oncogene in tumors induced in nude mice by a human breast carcinoma cell line. *Cancer Res.*, **45**, 4372–4379.
- Galdemard, C., Brison, O. and Lavielle, C. (1995) The proto-oncogene FGF-3 is constitutively expressed in tumorigenic, but not in non-tumorigenic, clones of a human colon carcinoma cell line. *Oncogene*, **10**, 2331–2342.
- Galdemard, C., Yamagata, H., Brison, O. and Lavielle, C. (2000) Regulation of FGF-3 gene expression in tumorigenic and non tumorigenic clones of a human colon carcinoma cell line. *J. Biol. Chem.*, **275**, 17364–17373.
- Regnier, V., Meddeb, M., Lecointre, G., Richard, F., Duverger, A., Nguyen, V.C., Dutrillaux, B., Bernheim, A. and Danglot, G. (1997) Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. *Hum. Mol. Genet.*, **6**, 9–16.
- Mautner, J., Joos, S., Werner, T., Eick, D., Bornkamm, G.W. and Polack, A. (1995) Identification of two enhancer elements downstream of the human *c-myc* gene. *Nucleic Acids Res.*, **23**, 72–80.
- Francke, U. (1981) High-resolution ideograms of trypsin-Giemsa banded human chromosomes. *Cytogenet. Cell Genet.*, **31**, 24–32.
- Bodescot, M. and Brison, O. (1996) Characterization of new human *c-myc* mRNA species produced by alternative splicing. *Gene*, **174**, 115–120.
- Burset, M., Seledtsov, I.A. and Solov'yev, V.V. (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.*, **28**, 4364–4375.
- Noller, H.F. (1984) Structure of ribosomal RNA. *Annu. Rev. Biochem.*, **53**, 119–162.
- Lindell, T.J. (1980) Inhibitors of mammalian RNA polymerases. In P.S., Sarin and R.C., Gallo (eds), *Inhibitors of DNA and RNA Polymerases*. Pergamon Press, Oxford, pp. 111–141.
- Liu, Y., Kuersten, S., Huang, T., Larsen, A., MacMorris, M. and Blumenthal, T. (2003) An uncapped RNA suggests a model for *Caenorhabditis elegans* polycistronic pre-mRNA processing. *RNA*, **9**, 677–687.
- Prochasson, P., Delouis, C. and Brison, O. (2002) Transcriptional deregulation of the keratin 18 gene in human colon carcinoma cells results from an altered acetylation mechanism. *Nucleic Acids Res.*, **30**, 3312–3322.