Bidirectionality and transcriptional activity of the EWSR1 promoter region

EMELY MÖLLER1, NILS MANDAHL1, MARIOLA ILISZKO2, FREDRIK MERTENS1 and IOANNIS PANAGOPOULOS1

1Department of Clinical Genetics, Lund University Hospital, SE-221 85 Lund, Sweden; 2Department of Biology and Genetics, Medical University of Gdansk, Debinki 1, 80-210 Gdansk, Poland

Received August 28, 2008; Accepted October 15, 2008

DOI: 10.3892/or_00000267

Abstract. EWSR1 is involved in chimeric proteins which play crucial roles in the development of a variety of bone and soft tissue tumors. Many of the chimeric genes involving EWSR1 have been extensively studied, whereas less is known about the wild-type (wt) gene and its regulation. As the expression of the chimeric gene is driven by the EWSR1 promoter, it is of importance to study the mechanisms regulating wt EWSR1 expression. We estimated the transcriptional activity of the EWSR1 promoter through deletion fragments driving reporter gene expression. This assay identified the 100-bp region immediately downstream of the EWSR1 transcriptional start site (+1) and the downstream region from +100 to +300 as important regions for transcriptional regulation. We also found that EWSR1 and RHBDD3, a gene located directly upstream of EWSR1 that is likely to share regulatory elements with EWSR1, were co-expressed in the tissue panels, Ewing tumor biopsies and cell lines. Thus, our results show that the EWSR1 promoter functions in a bidirectional manner, thereby regulating also RHBDD3, and identifies specific regions that strongly influence promoter activity.

Introduction

The structurally and functionally related TET family of RNA-binding proteins includes EWSR1, FUS, TAF15, the fruitfly protein CAZ and the Zebrafish proteins EWSR1A and EWSR1B (1-5). These proteins are characterized by Arg-Gly-Gly (RGG)-repeats in the carboxy (COOH)-terminal part, an RNA-binding domain in the central region with a Cys2/Cys2-type zinc finger motif on the 3'-side and a Ser-Tyr-Gly-Gln-Gln-Ser (SYGQQS) repeat-rich region in the amino (NH2)-terminal part. In humans, the latter is known to function as a transcriptional activation domain and has been involved in a variety of chimeras associated with cancer (6-8). The TET proteins have been found to interact with distinct TFIID subpopulations, RNA polymerase II and proteins of the splicing machinery, indicating a role in transcription initiation and elongation as well as mRNA processing (reviewed in ref. 9).

Of the TET family members, EWSR1 is the gene most often found in neoplasia-associated fusions. The most extensively studied example is EWSR1-FLI1 which is created by the t(11;22)(q24;q12) present in >85 % of Ewing sarcoma family tumors (3,10). Fusions involving EWSR1 result in the formation of aberrant transcription factors through replacing the RNA-binding portions of EWSR1 with the COOH-terminal DNA-binding domain of the fusion partner. The ultimate consequence of the fusion is believed to be transcriptional deregulation of genes targeted by the transcription factor.

Since the expression of the chimeric gene is driven by the EWSR1 promoter, it is important to understand the mechanisms regulating the wild-type (wt) EWSR1 expression. However, little is known about the EWSR1 promoter, apart from the genomic sequence 600-bp upstream of the initiator ATG codon studied by Plougastel et al (11). This GC-rich region contained several putative transcriptional start sites, whereas it lacked consensus TATA or CCAAT sequences, thus resembling the promoters of housekeeping genes (11).

We extended the analysis of the EWSR1 regulatory region to include ~2 kbp and assayed the transcriptional activity of this region through measuring the ability of deletion constructs to drive the expression of the firefly luciferase gene. As bioinformatic analysis suggests that the EWSR1 promoter is bidirectional, we also estimated the relative expression of EWSR1 and RHBDD3 (also called PTAG), a gene which is located directly upstream of EWSR1. EWSR1 (accession no. NM_005243) and RHBDD3 (NM_012265) are oriented head-to-head with their transcriptional start sites situated within 400 bp of each other, and are thus likely to share regulatory elements.

Materials and methods

Bioinformatic analyses. The EWSR1 5'-flanking region (part of the sequence with accession no. Y08806) chosen for the
Table I. The sequences and combinations of primers used for the construction of reporter gene constructs.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’-3’)</th>
<th>Fragment (primer combination)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWSprom49F</td>
<td>acgtggtccagctttgcccccta</td>
<td>Template</td>
</tr>
<tr>
<td>EWSprom2222R</td>
<td>gggagcctcagccagacgcc</td>
<td>Template</td>
</tr>
<tr>
<td>EWSprom169F-KpnI</td>
<td>tgtcaagtacctctctgcgtccagg</td>
<td>pE1 (169F-KpnI + 1721R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE2 (169F-KpnI + 1811R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE3 (169F-KpnI + 2156R-BgI)</td>
</tr>
<tr>
<td>EWSprom964F-KpnI</td>
<td>gtaggggtatccaggattgtcttg</td>
<td>pE4 (964F-KpnI + 1721R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE5 (964F-KpnI + 1811R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE6 (964F-KpnI + 2156R-BgI)</td>
</tr>
<tr>
<td>EWSprom1324F-KpnI</td>
<td>caccacccgtgaccggccgtgc</td>
<td>pE7 (1324F-KpnI + 1721R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE8 (1324F-KpnI + 1811R-BgI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pE9 (1324F-KpnI + 2156R-BgI)</td>
</tr>
<tr>
<td>EWSprom1721R-BgII</td>
<td>gcagggccagatctccgacaa</td>
<td>pE1, pE4, pE7</td>
</tr>
<tr>
<td>EWSprom1811R-BgII</td>
<td>caccagatctacccctgcgacct</td>
<td>pE2, pE5, pE8</td>
</tr>
<tr>
<td>EWSprom2156B-BgII</td>
<td>tgccgaagattcctccctgaccc</td>
<td>pE3, pE6, pE9</td>
</tr>
<tr>
<td>EWSprom1324F-BgII</td>
<td>caccacccgatctccggtgacc</td>
<td>p7Erev (1324F-BgII + 1721R-KpnI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p9Erev (1324F-BgII + 2156R-KpnI)</td>
</tr>
<tr>
<td>EWSprom1721R-KpnI</td>
<td>gacggcagacggtacccctgacca</td>
<td>p7Erev</td>
</tr>
<tr>
<td>EWSprom2156R-KpnI</td>
<td>tgccgaagtcctccctcctc</td>
<td>p9Erev, p10Erev</td>
</tr>
<tr>
<td>EWSprom1427F-BgII</td>
<td>ctccccgcatcctgccgtcc</td>
<td>p10Erev (1427F-BgII + 2156R-KpnI)</td>
</tr>
</tbody>
</table>

The present study extends roughly 1600-bp upstream to 400-bp downstream of the EWSR1 transcriptional start site (accession no. X66899) (11). The CpG-content of the region was estimated with the CpGPlot software [(12), http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html]. The program GenomeVISTA [(13), http://genome.lbl.gov/vista/index.shtml] was used to search genome assemblies for candidate orthologous sequences to the EWSR1 5′-flanking region. The multiple sequence alignment was performed using the Multalin software [(14), http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html]. The weight-matrix based program Match public version (v) 1.0 (both available at www.gene-regulation.com) (cut-off; 0 mismatches), were used to search for putative transcription factor binding sites (TFBSs) utilizing the TRANSFAC v 6.0 database. The weight-matrix based program MatInspector [(16), www.genomatix.de] which utilizes the MatBase database of TFBSs was also used for this purpose (cut-off selections; core similarity 0.8, matrix similarity 0.8). Default settings were used for all analyses unless otherwise indicated.

Materials. The cell line HEK293 (human embryonic kidney, ICLC, Genova, Italy) and the human Ewing family tumor cell lines TC-71 (DSMZ no. ACC 516), SK-N-MC (DSMZ no. ACC 203), A-673 (ATCC no. CRL-1598), SK-PN-DW (ATCC no. CRL-2139), RH-1 (DSMZ no. ACC 493), MHH-ES-1 (DSMZ no. ACC 167), RD-ES (ATCC no. HTB-166), SK-ES-1 (ATCC no. HTB-86), SK-NEP-1 (ATCC no. HTB-48) and TTC-466 (kindly provided by Dr P.H. Sorensen, Vancouver) were used for transfection experiments and real-time PCR analysis. SK-N-MC cells have been described as derived from neuroblastoma, A-673 and RH-1 from rhabdomyosarcoma and SK-NEP-1 from Wilms tumor, but are now regarded as Ewing family tumor cell lines expressing EWSR1-FLI1 fusion transcripts (17-19). Cells were cultured according to the suppliers' instructions with the additional supplements of 4 mM L-glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin. Biopsy samples from four Ewing family tumors were also available for real-time PCR analysis.

RT-PCR analysis was performed on all the tumor samples and cell lines to confirm the presence of EWSR1-FLI1 type 1 or type 2 fusion transcripts. Total RNA was extracted from frozen tumor biopsies or cell lines using the TRIzol reagent (Invitrogen, Carlsbad, CA) and 2.5 μg RNA was reverse transcribed in a 20 μl reaction volume containing 1X First-strand buffer (Invitrogen), 10 mM DTT (Invitrogen), 0.5 mM of each dNTP, 25 mM random hexamers, 37 units of RNaseguard (GE Healthcare Biosciences, Little Chalfont, UK) and 200 units of M-MLV reverse transcriptase (Invitrogen). The reaction was carried out at 37°C for 60 min, 70°C at 10 min and then kept at 4°C. Nested PCR to detect EWSR1-FLI1 fusion transcripts was performed as described (20). PCR products were directly sequenced using the Big Dye terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City,
Construction of reporter gene plasmids. All the primer sequences are listed in Table I. The 2 kbp EWSR1 5'-flanking region was amplified from 300 ng DNA of the CTA-984G1 clone (accession no. AL031186) with AccuPrime Pfx PCR (Invitrogen) in a 50 μl reaction volume of 1X AccuPrime Pfx reaction mix, 1 unit of AccuPrimePfx DNA polymerase and 0.3 μM of each of the primers EWSprom 49F forward and EWSprom2222R reverse. The PCR was run on a PCT-200 DNA Engine (MJ Research, Waltham, MA) with the cycling profile of an initial denaturation for 2 min at 95°C followed by 30 cycles of 15 sec at 95°C, 30 sec at 58°C and 2 min at 68°C, and a final extension for 2 min at 72°C. Serially truncated fragments of the 5'-flanking region; pE1-E9, p7E, p9E and p10E (Fig. 1A), were amplified as described above using the first amplified fragment as template and the primer combinations according to Table I. The fragments were then digested with the restriction endonucleases BgII and KpnI and subcloned between the corresponding sites of the vector pHrL. In pHrL, the firefly and Renilla luciferase genes are transcribed in opposite directions and the expression of the internal control Renilla luciferase is driven by the SV40 early enhancer/promoter (21). Fragments p7E, p9E and p10E (Fig. 1B) were cloned in the reverse orientation. All the primer sequences are listed in Table I. The number of putative TFBSs within the 770-bp sequence was narrowed down utilizing three different prediction programs. All three programs identified a potential MYB-binding site (MRE-element), four SP-1 sites (GC-, CA-, or GC-), a YY1-binding site and an E-box (binding MYC, MAX or USF) (Fig. 2B). No apparent TATA- or CAAT-boxes were identified by these programs.

Transfection and luciferase assays. The cell lines HEK293, TC-71, MHH-ES-1, RD-ES and TTC-466 were used for the transcriptional activation assay. Cells were seeded at a density of 2-5x10^4 cells/well in 96-well plates and 24 h later transfected with the test constructs pE1-E9. The PolyFect or Effectene transfection reagent (Qiagen, Valencia, CA) was used to transfect HEK293 cells with 0.5 or 0.1 μg plasmid DNA, respectively, according to the manufacturers’ instructions. TC-71, MHH-ES-1 and TTC-466 cells were transfected using the FuGENE HD transfection reagent (Roche Applied Sience, Indianapolis, IN) at the FuGENE (μl):DNA (μg) ratio 5:2 (TC-71) or 6:2 (MHH-ES-1, RD-ES and TTC-466), according to the manufacturer’s instructions. HEK293 and MHH-ES-1 cells were also transfected with the constructs p7Erev, p9Erev and p10Erev (Fig. 1B). Each construct was measured in at least eight replicas and the results were presented as fold differences in gene expression levels between the test and calibrator samples.

Results

In silico analysis of the EWSR1 5'-flanking region. The 5'-flanking region of EWSR1 was found to have a high (≥68%) GC-content and a putative CpG-island, which covers the intergenic region between EWSR1 and RHBDD3 and the first exons of both genes, was identified (Fig. 2A). The GenomeVISTA alignment of candidate orthologous sequences suggests that the 770-bp region encompassing the intergenic portion, the first exon of EWSR1 and the sequence immediately downstream of it, has the highest sequence similarity across the species investigated. The more detailed multiple sequence alignment of the same region revealed smaller regions of high (≥90%) or low (≤50%) consensus sequence (Fig. 2B). The number of putative TFBSs within the 770-bp sequence was narrowed down utilizing three different prediction programs. All three programs identified a potential MYB-binding site (MRE-element), four SP-1 sites (GC-, CA-, or CT-boxes), a YY1-binding site and an E-box (binding MYC, MAX or USF) (Fig. 2B). No apparent TATA- or CAAT-boxes were identified by these programs.

Transcriptional activity of the EWSR1 promoter region. HEK 293 cells and Ewing sarcoma cell lines expressing the EWSR1-FLI1 type 1 (TC-71), EWSR1-FLI1 type 2 (MHH-ES-1) or EWSR1-ERG (TTC-466) fusion transcripts were used in the real-time analysis. The fragments pE3, pE6 and pE9 had the lowest activities in all the cell lines (P<0.001) (Fig. 1A). In all cell lines the expression of the type 2 transcript (data not shown) is consistent with those obtained from MHH-ES-1, TTC-466 and HEK293 cell transfections. The activities of fragments p7Erev, p9Erev and p10Erev (cloned in the reverse orientation) were also assessed. In MHH-ES-1 cells, the p10Erev fragment had the highest activity, followed by p9Erev, pE7 (P<0.001) and then pE9 and p7Erev. In HEK 293 cells, p10Erev had equally high activity as pE7 (Fig. 1B).

Gene expression by RT-PCR and real-time PCR. In the EWSR1-FLI1 type 1 transcript, exon 7 of EWSR1 is fused to exon 6 of FLI1 and in the type 2 transcript, the EWSR1 exon 7
is fused to exon 5 of FLI1. The RT-PCR analysis confirmed the presence of type 1 EWSR1-FLI1 fusion transcripts in cell lines TC-71, SK-N-MC, A-673, SK-PN-DW and RH-1 cells, and type 2 transcripts in MHH-ES-1, RD-ES, SK-ES-1 and SK-NEP-1 cells. Cell line TTC-466 expresses the EWSR1-ERG fusion transcript. EWSR1-FLI1 type 1 transcripts were also found in three of the tumor biopsies and the type 2 transcript in one tumor.

The real-time PCR data show that the EWSR1 promoter is active, and that the EWSR1 and RHBDL3 genes are co-expressed, in both normal tissue panels (Fig. 3A) and Ewing biopsies and cell lines (Fig. 3B).
Although *EWSR1* is well-known as the 5'-partner gene of numerous chimeras in human tumors, less is known about the gene itself and its regulation. Therefore, we investigated the regulatory region of *EWSR1* to delineate specific portions that prominently influence promoter activity. Moreover, the head-to-head arrangement of the *EWSR1* and *RHBDD3* genes...
with a short distance (<400 bp) between their transcriptional start sites suggested to us that they may share a bidirectional promoter region. Thus, we also investigated the bidirectional activity of the core promoter region and compared the relative gene expression levels of EWSR1 and RHBDD3 in tissue panels and Ewing tumor biopsies and cell lines.

The results from the transcriptional activation assay show that deleting different parts of the sequence from -1600 to -440 had little effect on the promoter activity. In contrast, deleting parts of the -440 to +400 portion had a considerable effect on activity, indicating that this region is important for transcriptional activity. The fact that it encompasses the transcriptional start sites of both EWSR1 and RHBDD3, together with the high degree of sequence similarity across species observed for this area, further implies an important role in regulation. The fragments containing the entire -440 to +400 portion (pE3, pE6 and pE9) had the lowest activities in all the cell lines, presumably because this specific region recruits negatively acting transcription factors. Deletion of the +65 to +400 part (fragments pE2, pE5 and pE8) notably increased activity. The pE2, pE5 and pE8 fragments had the highest activities in all cell lines used, except TC-71. The deleted part contains three putative binding sites for transcription factors of the SP-family and an E-box recognized by factors such as MYC, MAX and USF. Particularly the SP factors have been associated with the regulation of housekeeping genes and GC-rich promoters (23).

Further deletion of the 90 bp portion immediately downstream of the EWSR1 +1 site (fragments pE1, pE4 and pE7) decreased promoter activity, except in TC-71 cells. These activities were less than half of those achieved by the fragments retaining the 90-bp portion, indicating that this narrow region

Figure 3. The relative expression levels of EWSR1 and RHBDD3. Expression values have been normalized to RPLPO and are presented as fold differences in expression levels relative to the calibrator sample (ovary or SK-NEP-1, respectively). Light grey bars represent the level of EWSR1 expression and the dark grey bars that of RHBDD3. (A) Co-expression of EWSR1 and RHBDD3 in normal tissue panels. (B) EWSR1 and RHBDD3 are co-expressed also in Ewing tumors and cell lines. TB denotes the tumor biopsy samples; the presence of type 1 or 2 EWSR1-FLI1 fusion transcripts is indicated.
is important for transcriptional activation and most likely also for transcriptional initiation, due to its position close to the +1 site. This region contains a putative binding site for YY1, a ubiquitously expressed transcription factor which may act indirectly or directly to activate or repress gene expression (24). In TC-71 cells, the deletion had no effect on activity, suggesting cell-specific regulation depending on distinct transcription factor content.

The experiment with the fragments cloned in the reverse orientation (p7Erev, p9Erev and p10Erev) shows that the promoter region is capable of bidirectional activity, although to some degree in a cell-specific manner. It is noteworthy that in contrast to the above-mentioned results from the transcription activation assay, p9Erev which contains the entire -440 to +400 portion activated transcription more effectively than the p7Erev fragment. Fragment p10Erev differs from the others in that it only reaches to, and not downstream of, the RHBBDD3 +1 site. It had high activity in both cell lines, implying an important role for the intergenic area upstream of the RHBBDD3 +1 site. Furthermore, as shown by the real-time PCR data, EWSR1 and RHBBDD3 are co-expressed in tissue panels and Ewing tumors and cell lines. These results support the idea that these genes may be co-regulated by a common bidirectional promoter.

The bidirectional arrangement has previously been described for gene pairs arranged head-to-head on opposite strands with <1 kbp between their transcriptional start sites. Bidirectional promoters may be required to maintain a stoichiometric relationship between the gene products or for the co-expression of genes involved in the same pathway. Moreover, bidirectionality has been associated with a higher GC-content compared to non-bidirectional promoters. A CpG-island is often found between the start sites of the genes which also overlaps the first exons of each gene (25,26). The latter is true also for the region shared by EWSR1 and RHBBDD3. RHBBDD3 was originally identified as a pro-apoptotic gene with decreased expression in pituitary tumors. Loss of expression was associated with methylation of its CpG island (27). Recently, loss of RHBBDD3 expression was found in 16 out of 18 colorectal tumor samples. However, in this instance the tumors and normal colon had similar densities of CpG island methylation (28). The mechanism causing this loss of expression is currently unknown.

In conclusion, we have delineated specific regions of the EWSR1 promoter that seem to be important for promoter function and provide evidence that the EWSR1 promoter is capable of bidirectional activity. The regions we identified may be the subject of future studies which aim to characterize the transcription factor networks which are involved in the regulation of this gene (or rather, these genes). Moreover, the bidirectional activity raises questions concerning the mechanisms behind the loss of RHBBDD3 expression in colorectal tumors and gene expression status of EWSR1 in these tumors.

Acknowledgements

This study was supported by the Swedish Children’s Cancer Foundation, the Swedish Cancer Society, the Swedish Research Council, the Gunnar Nilsson’s Cancer Foundation and the Lund University Hospital.

References


