PROMOTER MODELING APPROACHES
APPLIED TO THE INVESTIGATION
OF p63 UP- AND DOWNSTREAM PROMOTERS

Shelest E.S.¹, Wingender E.¹,²
¹ Dept. of Bioinformatics, UKG, University of Göttingen, Goldschmidtstr. 1, D-37077 Göttingen;
² BIOBASE GmbH, Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany

* Corresponding author: e-mail: ksh@bioinf.med.uni-goettingen.de

Key words: p63, promoter modeling, transcription factor binding sites, keratinocytes

SUMMARY

Motivation: The transcription factor p63 is a homolog of p53, the tumor suppressor in higher mammals. p63, but not p53, can be expressed from at least two alternative transcription start sites (TSS), yielding a full-length form from the upstream and truncated form (ΔN) from the downstream TSS. The ΔN form acts as p53 antagonist, which makes the understanding of its regulation an important task. To date, the regulators of the p63 promoters are still to be identified.

Results: Comparative analysis of the p63 promoter regions of several species revealed highly conserved combinations of transcription factor binding sites (TFBS), which are suggested as the models of the regulatory patterns for the up- and downstream promoters. The predicted involvement of RXR in the regulation of p63 is in agreement with the experimental data. The other predictions are presently under the experimental evaluation.

With this work, we demonstrate the applicability of the methods of promoter modeling previously developed in our group to a new kind of task: investigation of unknown regulatory patterns in a single promoter based on phylogenetic comparisons.

INTRODUCTION

p53, p63, and p73 constitute a family of DNA-binding proteins that share significant sequence homology. Both p63 and p73 can be expressed from at least two alternative transcription start sites (TSS), yielding full-length forms (transactivating, TA) from the upstream and truncated forms (ΔN) from the downstream TSS. The up- and downstream TSS are under control of two distinct promoters. In spite of the structural similarity, p63 and p73 demonstrate functional differences from p53 and between each other (Waltermann et al., 2003). Compared to p53, the roles of its homologs are more diverse. Although the TA forms of the factors can induce p53-responsive genes, the factors are not specifically assigned to tumor cells and cannot be defined as tumor-suppressors. p73 is frequently overexpressed in various malignancies, but also plays role in normal development of nervous and immune systems; p63 is known to be important for skin development, being specifically expressed in keratinocytes. Thus, the diversity of the functions and complexity of the transcription model makes the understanding of the transcription regulation of p63 and p73 a challenging task.

Little is known about the regulation of transcription of the two homologs of p53. It has been shown that the p73 promoters are regulated by E2F, p53 and (indirectly) by TGFβ (transforming growth factor β), but these factors have either no or negative effect on the
Computational structural and functional genomics and transcriptomics

p63 promoter. To date, there is no positive information about the p63 regulators. Thus, we decided to undertake a computational analysis of the promoters in order to supply the experimentalists with suggestions, which they could later confirm with their methods.

METHODS AND ALGORITHMS

Promoter sequences were extracted from ENSEMBL based on sequence homology to the corresponding regions in the human genome (AB055067 for deltaNp73 and AF124530 for dNp63). Homologous sequences were identified with NCBI Blast (http://www.ncbi.nlm.nih.gov). The set of the p63 upstream promoters contained 5 sequences (human, mouse, rat, cow, and dog). The downstream promoters were represented by 6 species for the (human, mouse, rat, cow, dog, and chicken). The length of the sequences in both sets was 1500bp (-1399/+100).

Negative training set consisted of all human promoter sequences from EPD database (1871 seq.). The set was checked for the absence of p63 promoters. The length of the sequences was 1500bp (-1399/+100).

Multiple alignments of orthologous promoter sequences were performed with the Multi-LAGAN tool (http://lagan.stanford.edu/lagan_web/index.shtml)

Search for potential binding sites was undertaken with the help of the Match™ tool (Kel et al., 2003) (http://www.biobase.de/cgi-bin/biobase/transfac/start.cgi). The thresholds for the matrix search were adjusted in such a way that the matrix (or set of matrices) for each factor could re-identify 80% of the true positive set (i.e., the set of genuine binding sites). The binding sites for p63 (p53) were predicted with a tool P53MH (Hoh et al., 2002), which searches for two p53 binding sites separated by a gap up to 13 bp.

The prediction of potentially functional TFBS pairs was performed by two independent methods: (i) set of approaches to promoter model construction as described in (Shelest, Wingender, 2005); the approach of distance distributions described in (Shelest, 2006).

Databases
Eukaryotic Promoter Database (http://www.epd.isb-sib.ch), release 77-1
Ensembl Genome Browser (http://www.ensembl.org/index.html)
TRANSFAC® Professional release 9.4 (http://www.biobase.de)

RESULTS AND DISCUSSION

The starting point of the analysis was the fact that the transcription of p63 is keratinocyte-specific; thus, it was reasonable to check in the first place the transcription factors active in these cells. The search for the “keratinocytes” in the field "cell specificity" in TRANSFAC database revealed 9 transcription factors (AP-2, Sp1, KRF-1, RXR-α, ESE-2, ESE-2b, POU2F3, ΔNp63, p63), from which only 6 possessed PWMs from the TRANSFAC matrix library. These 6 factors (AP-2, Sp1, RXR-α, POU2F3 (Oct-2), ΔNp63, p63) were taken for the analysis (ΔNp63 and p63 have the binding sites identical to p53, hence the same matrix).

After the identification of single potential binding sites with the help of the Match™ tool, we analyzed the occurrences of combinations of these TFBS. Up- and downstream promoters were considered separately.

The predictions for the TF binding site (TFBS) combinations were made by two independent approaches. The first (Shelest, Wingender, 2005) considers overrepresentation of sequences containing certain TFBS pairs in the investigated set compared with a negative set. We adjusted the parameters in such a way that the pairs were present in 100 % of the investigated sets, and less than in 10 % of the negative training set. The results are shown in Table 1, right column. The second approach, called
“distance distributions approach” (Shelest, 2006), considers the pairs which occur on “overrepresented” distances in comparison with analytically calculated profile of distance distribution in the random case (i.e., when the binding sites are distributed randomly). The results of the application of this method are shown in the left column of the Table 1. Note that both approaches identified practically the same pairs (Table 1).

On the next step of the analysis we looked whether the found combinations of TFBS were evolutionary conserved. For that we mapped the TFBS on the plots representing the conserved regions of the promoters (Fig. 1 and 2).

Table 1. Comparison of the predictions of TFBS pairs made by the two approaches

<table>
<thead>
<tr>
<th>Distance distributions approach</th>
<th>TFBS pairs approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. DOWNSTREAM PROMOTERS</strong></td>
<td></td>
</tr>
<tr>
<td>Oct-2 - Sp1 (84)</td>
<td>Oct-2-RXR (15)</td>
</tr>
<tr>
<td>Oct-2-RXR (12-15)</td>
<td>RXR-RXR (30)</td>
</tr>
<tr>
<td>RXR-RXR (30)</td>
<td>Sp1-RXR (39), (45) and (69)</td>
</tr>
<tr>
<td>Sp1-p53 (23-34)</td>
<td>Sp1-Sp1 (38-52) and (126-135)</td>
</tr>
<tr>
<td>Sp1-Sp1 (126-130)</td>
<td></td>
</tr>
<tr>
<td><strong>B. UPSTREAM PROMOTERS</strong></td>
<td></td>
</tr>
<tr>
<td>AP-2 – Sp1 (26), (32), (81-82)</td>
<td>AP-2 – Sp1 (80-82)</td>
</tr>
<tr>
<td>AP-2-Oct-2 (52), (121-125)</td>
<td>AP-2-Oct-2 (52)</td>
</tr>
<tr>
<td>AP-2-RXR (78-81)</td>
<td>AP-2-RXR (81)</td>
</tr>
<tr>
<td>Oct-Oct (30-31)</td>
<td>Oct-2 - Sp1 (78-84)</td>
</tr>
<tr>
<td>Oct-2-RXR (7-10)</td>
<td>Sp1-RXR (12), (72-82)</td>
</tr>
<tr>
<td>Sp1-RXR (72-82)</td>
<td>Sp1-RXR (81)</td>
</tr>
<tr>
<td>Sp1-Sp1 (5)</td>
<td></td>
</tr>
</tbody>
</table>

Notes. Marked with bold font are the coinciding pairs. In parentheses is shown the distance range (i.e., not less than the first number, not more than the last).

As it can be seen on the Fig. 1, the highly conserved regulatory module is constituted from TFBS for 3 factors: Oct-2 (POU2F3), RXR and Sp1. RXR sites can be used by retinoic acid, the involvement of which is in agreement with previously reported experimental data (Bamberger et al., 2002).

As it can be seen on the Fig. 1, the highly conserved regulatory module is constituted from TFBS for 3 factors: Oct-2 (POU2F3), RXR and Sp1. RXR sites can be used by retinoic acid, the involvement of which is in agreement with previously reported experimental data (Bamberger et al., 2002).

Figure 1. Distribution of the TF binding sites in the regions conserved between the human p63 downstream promoter sequence and five orthologs: cow (A), chicken (B), dog (C), mouse(D) and rat (E).

○ -Oct-2;  ■ - RXR;  △ - Sp1.
Figure 2. Distribution of the TF binding sites in the regions conserved between the human p63 upstream promoter sequence and four orthologs: cow (A), dog (B), mouse (C) and rat (D).

The distribution of the sites in the upstream promoters (Fig. 2) deserves more discussion. One can notice that there are three “islands” of high conservation: –1380 - –1000 (worse conserved between rodents and human), -600 - -500 (not present in cow) and –400 - +80. An interesting behavior demonstrates the combination of RXR-Oct-2 TFBS. It is conserved in the region –1380 - –1000 in cow and dog, but is not present in mouse and rat where this region is also not conserved. However, it appears now in the region –600- –400, in which it is not found in cow, but is detected in other species, appearing twice in dog. We can speculate that this combination was present in two copies in the common ancestors of these species and retained as such in the dog and human, whereas cow has lost one and rodents the other copy.

The predicted combinations are presently under experimental verification in the laboratory of Molecular Oncology headed by Prof. M. Dobbelstein (Göttingen).

ACKNOWLEDGEMENTS

The work was supported by grant from the German Federal Ministry of Education and Research to the Intergenomics Bioinformatics Competence Center (grant No. 031U110A).

REFERENCES


BGRS’2006