Prediction and design of zinc finger target sites for an important human regulatory region (locus control region)

Sharma B. Sharan1,2 and Verma Ramtej J.2*

1. Life Sciences Research Division, Indrashil Institute of Science and Technology (IIST), Indrashil University (IU), Kadi, Mehsana – 382740, Gujarat, INDIA
2. Department of Human Genetics, Zoology and Biomedical Technology, University School of Sciences, Gujarat University, Navrangpura, Ahmedabad – 380009, Gujarat, INDIA
*rjvermasosgu@gmail.com

Abstract

In order to understand the regulation of a gene, it is important to understand the mechanism of DNA-protein interaction at the molecular level. Recognition of zinc finger domains is one approach to discern the target sites of DNA-protein interactions. Utilizing this information, zinc finger proteins (ZFPs), most common DNA binding proteins in mammals, can be synthetically designed and engineered for the modulation of endogenous genes expression and for targeted genome engineering. Here, we discovered zinc finger target sites in an important regulatory DNA sequence tyrosinase 5’ upstream regulatory sequence (TYR 5’URS) using online bioinformatics tool ‘Zinc Finger Tools’ (ZF Tools). TYR 5’URS acts as a transcriptional cis regulatory/locus control region (LCR) of a multifactorial tyrosinase (TYR) gene. A total of 79 top scoring zinc finger target sites could be detected in TYR 5’URS. Top score of the identified target sites ranged from 40.87 to 61.01.

Further, we generated protein coding sequences for the top scoring target sites which contained canonical amino acid linker between the zinc finger modules. Amino acid sequences of designed ZFPs, when subjected to identity analysis, did not correspond significantly to any known human protein indicating the specificity of the zinc fingers for the target sites. Information and data generated in this work will be further helpful to design synthetic ZFPs and to assemble the zinc finger-DNA binding domain coding sequence to target TYR 5’URS. Designing ZFPs specific to TYR gene regulatory elements will allow us to explore and understand the mutation spectrum of this gene in a population. This will further pave the way to therapeutic interventions.

Keywords: Locus control region, target sites, tyrosinase, ZF tools, zinc finger domain.

Introduction

Zinc finger proteins (ZFPs) are one of the most abundant protein groups with extensive range of biological functions. ZFPs are involved in the regulation of several cellular processes by interaction with DNA, RNA, PAR (poly-ADP-ribose) and other proteins1. Variety of zinc finger domains are widely utilised by transcription factors in eukaryotes to interact with ZFPs. ZFPs play diverse physiological roles through their ability to regulate gene expression to maintain tissue homeostasis2-3. Thus, role of ZFPs in diseases including tumorigenesis, neurodegenerative and other human diseases has been an interesting area of research.

Advances in understanding the mode of zinc finger-DNA interactions led to the ability to engineer synthetic ZFPs to target specific DNA site in the genome1. ZFPs are potential candidates for therapeutic interventions due to their small size and cell penetrating abilities. Studies in the recent past have shown that engineered ZFPs could be directly delivered into mammalian cells5,6. Design of novel ZFPs is a valuable tool for the modulation of gene expression and for targeted genome editing7. This will further pave the way to design drugs that will target specific ZFPs to avoid or restore abnormal expression of the particular gene.

In this study, we discovered zinc finger target sites of a regulatory DNA sequence tyrosinase 5’ upstream regulatory sequence (TYR 5’URS) using web-based online bioinformatics tool ‘Zinc Finger Tools’ (ZF Tools Ver 3.0) which utilizes knowledge from available and updated ZF-DNA interactions data set. TYR 5’URS is a regulatory element of an important multifunctional gene TYR, mutations in which cause oculocutaneous albinism type 1 (OCA1).

TYR 5’URS might be involved in some cases of OCA1 as not all TYR mutations lead to this disease. Therefore, targeting this regulatory region using ZFPs will allow us to explore the biological functions of this region. The identified zinc finger target sites in this work were further analysed to design the ZFPs which will be useful in future research to target TYR 5’URS.

Material and Methods

Gene sequence: In this study, we analysed the DNA sequence of human gene tyrosinase 5’ upstream regulatory sequence (TYR 5’URS) which was retrieved from GenBank.

Identification of contiguous (uninterrupted) target sites (zinc finger domain): The TYR 5’URS gene sequence was subjected to zinc finger domain analysis using ZF Tools8.
which is a web-based tool available online (https://www.scripps.edu/barbas/zfdesign/zfdesignhome.php). The sequence was supplied to the tool to identify contiguous zinc finger target sites within the gene with default parameters. In brief, all the triplets (GNG, ANN, TNN, CNN and GN(ACT)) from total ZF set were searched with minimum target size of 18 bp (default).

**Design of ZFPs for target sites identified in the sequence:** Predicted target sites after the primary search were analysed further to generate the protein coding sequence for a ZFP expected to recognize the input DNA target sequence. On clicking the ZFP hyperlink, adjacent to generated target sequence, output displayed the designed ZFP and corresponding amino acid sequence.

**Correspondence of the designed ZFPs with known sequences:** Amino acid sequences of designed ZFPs of the top scoring target sites were further subjected to protein blast analysis (for organism Homo sapiens) using NCBI’s online blastp suite to determine the correspondence of the ZFPs with known sequences.

**Results**

**Characteristics of gene sequence:** TYR 5'URS is a regulatory sequence of 2261 bp which acts as a transcriptional cis regulatory/locus control region (LCR). Features of this gene are summarized in table 1. Polymorphism in this sequence may lead to pathogenic conditions as per the information from ClinVar.

**Identified target sites:** ‘ZF Tools’ searched all the possible available zinc finger sites within the TYR 5'URS DNA sequence as per the given parameters. A screenshot of search output result highlighting coverage map is given in fig. 1. Total of 79 target sites were found in the forward strand at different positions of different lengths (bp) in the query sequence (Table 2). Top score of the identified target sites ranged from 40.87 to 61.01 indicating the predicted affinity and specificity for the corresponding ZFP.

The higher is the score, the better is the site. Top scoring sites having minimum score of 60 (14 sites) are represented in Table 2 with target sequence (5’ to 3’) along with other relevant information. Different subsites were possible for the target sites as reflected in table 2 (number of subsites) since length of target sites was longer than given default parameter of 18 bp. Subsites are the more specified locations of the target sites generated as per given parameters. Few of the target sites were positive for target site overlap (TSO) (Table 2). Presence of overlaps indicates the decreased modularity of the site which should be evaluated during the synthetic design of ZFP.

**Generated protein coding sequences for identified target sites:** ZFP hyperlink was provided adjacent to identified target sites to generate the protein coding sequence for a ZFP expected to recognize the input DNA target sequence. An example of designed ZFP and corresponding amino acid sequence for a target sequence is given in Fig. 2 which is a screenshot of output result. Output page displayed the default backbone and linker sequences along with each DNA triplet of the target site in question beside the amino acid sequence of the helix expected to recognize it (Fig. 2).

The amino acid sequence of the entire ZFP was also available in the output which contained canonical amino acid linker (TGEKP) between the zinc finger modules (Fig. 2). The amino acid sequences of the zinc finger array were provided by the tool which will be helpful to design oligonucleotides to assemble the zinc finger-DNA binding domain coding sequence to target TYR 5'URS in future studies.

**Correspondence of the ZFPs with known sequences:** No significant matches were found for the designed ZFPs of the top scoring target sites as listed in table 2 (with minimum score of 60) upon blastp analysis.

![Fig. 1: Output result of target site search for the TYR 5'URS DNA sequence highlighting coverage map](image-url)
Discussion
Identification of unique signatures laying in the DNA sequences especially regulatory sequences is instrumental to understand the important role they play in various biological processes and their relevance. We have analysed unique signatures like motifs, and also therapeutic vector capabilities of LCRs to discern biological significance of these regulatory regions\textsuperscript{9,10}. In a similar approach, in this work we discovered the ZFP target sites in an important regulatory region, TYR 5'URS gene. TYR 5'URS works as a locus control region (regulatory element) and can drive the expression of an important multifunctional TYR gene in cells.

A large number of mutations have been reported in the TYR gene\textsuperscript{11}. It is now known that not all the OCA1 cases are linked with mutation in TYR gene\textsuperscript{12} and some mutations in regulatory element might be responsible for the disease for which molecular basis is yet to be established. Designing ZFPs specific to TYR gene regulatory elements will allow us to explore the mutation spectrum gene in a population.

Zinc finger domains are important target regions for genome engineering and modulation of gene expression in analytical and/or therapeutic settings\textsuperscript{13}. In eukaryotic transcriptional factors, zinc finger domain is a frequently utilized DNA binding motif where identity of amino acids at the target site delineates the specificity of zinc fingers for the recognition of DNA sequence\textsuperscript{1}. To achieve higher degree of specificity, these amino acids can be changed to engineer protein modules containing multiple zinc-finger motifs which bind to specific target site\textsuperscript{14,15}.

This approach paved the way to introduce breaks at specific sites of genomic DNA towards genome engineering\textsuperscript{16}. Recognition of zinc finger target sites is an important aspect of genome engineering, especially gene correction using zinc finger nucleases. In this work, we used the web-based online tool ZF Tools\textsuperscript{8} to identify the zinc finger target sites in an important regulatory gene, TYR 5'URS. ZF Tools is a user-friendly bioinformatics tool for systematic analysis of target sites appropriate for either gene regulation or nuclease targeting.
Amino acid sequence for a ZFP expected to bind to any chosen target site can be generated using this tool which is one of its important features. ZFPs constructed using this tool generally has excellent affinity toward target sites. We could identify 79 top scoring zinc finger target sites in the TYR 5'URS DNA sequence using default parameters of ZF Tools. Discovered target sites were longer than set parameter and hence subsites were also generated for the target sites. Target sites can be refined by analysing the subsites. Scores of the identified sites give an indication of the relevance of the sequence. The higher the score, the better is the affinity it will have for the designed zinc finger. Target site overlap (TSO) were present in few of the sites which indicate that the decreased affinity of the triplet meaning zinc finger domain may specify 4 bp site. In this case, TSO incompatibility should be evaluated for the surrounding DNA sequence before the design of ZFP.

Zinc finger was designed for the discovered target sites in the last step. ZF tools design the optimal ZFP for the target site from all the available ZF sets in its database. Protein coding amino acid sequences of designed ZFPs were analysed further to determine whether designed ZFP is already available in the NCBI database. Designed ZFPs did not correspond significantly to any human protein/ZFP indicating that these designed ZFPs are unique and can be used to specifically target TYR 5'URS DNA sequence to understand the regulatory potential of the gene.

### Conclusion and Future Scope
Zinc finger target sites were discovered in an important regulatory DNA sequence TYR 5'URS in this study. Further, ZFPs were designed from the top scoring target sites consisting of amino acid sequences of the zinc finger array. Moreover, amino acid sequences of designed ZFPs did not match significantly to any known human protein/ZFP suggesting the exclusivity of the ZFPs.

### Table 1
Characteristics of tyrosinase 5' upstream regulatory sequence

<table>
<thead>
<tr>
<th>Gene</th>
<th>Tyrosinase 5' upstream regulatory sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>LOC107080646 (TYR 5'URS)</td>
</tr>
<tr>
<td>Map Location</td>
<td>11q14.3</td>
</tr>
<tr>
<td>Nucleotide Accession</td>
<td>NG_046561</td>
</tr>
<tr>
<td>Size (bp)</td>
<td>2261</td>
</tr>
<tr>
<td>Feature</td>
<td>Transcriptional cis regulatory/Locus control region (LCR)</td>
</tr>
<tr>
<td>Clinical Significance</td>
<td>Pathogenic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Position</th>
<th>Length (bp)</th>
<th>Subsites</th>
<th>Score</th>
<th>Target site overlap (TSO)</th>
<th>Target sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>36</td>
<td>7</td>
<td>60.4</td>
<td>Yes</td>
<td>CTT AAA GAG ... AAT TGA TGG</td>
</tr>
<tr>
<td>167</td>
<td>24</td>
<td>3</td>
<td>60.58</td>
<td>No</td>
<td>TGG CCA TAG ... AGA GGC ATG</td>
</tr>
<tr>
<td>228</td>
<td>27</td>
<td>4</td>
<td>60.3</td>
<td>No</td>
<td>GGG GAC CAC ... AAT GTA CTA</td>
</tr>
<tr>
<td>436</td>
<td>39</td>
<td>8</td>
<td>60.77</td>
<td>Yes</td>
<td>AAA GTG ATA ... AAG ATG GGT</td>
</tr>
<tr>
<td>958</td>
<td>36</td>
<td>7</td>
<td>60.1</td>
<td>Yes</td>
<td>ATA CAC AGA ... CTG GTG AGT</td>
</tr>
<tr>
<td>1130</td>
<td>21</td>
<td>2</td>
<td>60.33</td>
<td>No</td>
<td>ATA AAA AGC ... GAG TAG GCA</td>
</tr>
<tr>
<td>1160</td>
<td>27</td>
<td>4</td>
<td>60.28</td>
<td>No</td>
<td>GAT ATT ATT ... CCC TGG ACC</td>
</tr>
<tr>
<td>1214</td>
<td>33</td>
<td>6</td>
<td>60.22</td>
<td>No</td>
<td>TGG GCC CCA ... AGA ATT GCA</td>
</tr>
<tr>
<td>1406</td>
<td>24</td>
<td>3</td>
<td>60.48</td>
<td>No</td>
<td>AAG GCC ACA ... GCC AGC AGG</td>
</tr>
<tr>
<td>1607</td>
<td>27</td>
<td>4</td>
<td>60.1</td>
<td>No</td>
<td>CCT GCC AGG ... GAT ACC ACA</td>
</tr>
<tr>
<td>1888</td>
<td>30</td>
<td>5</td>
<td>60.84</td>
<td>No</td>
<td>TGA ACT CTG ... CAT ACT GGG</td>
</tr>
<tr>
<td>1960</td>
<td>30</td>
<td>5</td>
<td>61.01</td>
<td>No</td>
<td>GAA ATA AGA ... GAA AAG GAA</td>
</tr>
<tr>
<td>2186</td>
<td>36</td>
<td>7</td>
<td>60.55</td>
<td>Yes</td>
<td>AGG TAG CCT ... CAG AGA GAC</td>
</tr>
<tr>
<td>2214</td>
<td>48</td>
<td>11</td>
<td>60.65</td>
<td>Yes</td>
<td>AGA GAG ACT ... GTA TGG GAA</td>
</tr>
</tbody>
</table>
The generated data and information will be vital to synthetically engineer and design ZFPs to target TYR 5'URS DNA sequence in future studies which will allow us to understand the regulatory spectrum of this upstream regulatory sequence. The similar approach can be applied to discover and design zinc finger domains and ZFPs in other eukaryotic genes.

References


3. Tong Q., Tsai J., Tan G., Dalgin G. and Hotamisligil G.S., Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation, Molecular and Cellular Biology, 25, 706-715 (2005)


(Received 20th August 2019, accepted 19th October 2019)