In Silico Analysis of Regulatory Elements of the Vitamin D Receptor

Shirin Farivar 1, Raya Amirinejad 2, Seyyedeh Batool Hassani 4, Zeinab Shirvani-Farsani 1*
Bahar Naghavi gargari 3

Received 26/2/2019, Accepted 5/1/2020, Published 1/6/2020

Abstract

Vitamin D receptor (VDR) is a nuclear transcription factor that controls gene expression. Its impaired expression was found to be related to different diseases. VDR also acts as a regulator of different pathways including differentiation, inflammation, calcium and phosphate absorption, etc. but there is no sufficient knowledge about the regulation of the gene itself. Therefore, a better understanding of the genetic and epigenetic factors regulating the VDR may facilitate the improvement of strategies for the prevention and treatment of diseases associated with dysregulation of VDR. In the present investigation, a set of databases and methods were used to identify putative functional elements in the VDR locus. Histone modifications, CpG Islands, epigenetic marks at VDR locus were indicated. In addition, repeated sequences, enhancers, insulators, transcription factor binding sites and targets of the VDR gene, as well as protein-protein interactions with bioinformatics tools, were reported. Some of these genetic elements had overlapped with CpG Islands. These results revealed important new insight into the molecular mechanisms of the VDR gene regulation in human cells and tissues.

Key words: CpG Islands, Epigenetics, In silico analysis, Regulation elements, VDR

Introduction

Vitamin D is a steroid hormone that balances calcium and phosphate levels and is crucial for bone structure. In addition, vitamin D has a key role in the immune response and inflammatory processes (1). Vitamin D acts through its receptor, the vitamin D receptor (VDR) (2).

Vitamin D receptor is a nuclear transcription factor that regulates gene transcription directly by binding to the Vitamin D responsive element (VDRE) present in the target genes (3). Also, its impaired expression was found to be related to different diseases. It is shown that VDR mRNA and protein expression decrease in a small group of human adenocortical carcinomas (ACCs) (4, 5), therefore; VDR plays important roles in immune system (6), lung (7), prostate and mammmary cancer (8), Alzheimer’s disease (9).

1 Department of Cell and Molecular Biology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University G.C., Tehran, IR Iran.
2 Genetics Department, Breast cancer Research center, Motamed Center Institute, ACECR, Tehran, IR Iran
3 Department of Basic Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran.
4 Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University G.C., Tehran, IR Iran.
5 Corresponding author: z.shirvani@sbu.ac.ir
6 ORCID ID: https://orcid.org/0000-0003-2071-1208

multiple sclerosis (10), Rheumatoid Arthritis and Systemic Lupus Erythematosus (11), cardiovascular system (12).

VDR silencing may result from methylation of CpG island or genetic variation of VDR promoter (13). Polymorphisms in the VDR gene, such as the restriction site of the Taql and Bsml polymorphisms have been associated with the different inflammatory diseases and could affect individual responses to vitamin D treatment (14). Moreover, the VDR promoter has a GC-rich region from −790 bp to 380 bp (15). However, disclosing the genetic variation and CpG island methylation complexity of the VDR gene remains challenging. This is partly because of the lack of a perfect understanding of the cis-regulatory elements that control gene expression. The present study investigated some properties of the VDR gene sequence and introduced some mechanisms that may control the expression of this gene since the comprehensive study of the genetic and epigenetic features of the human VDR gene may uncover new aspects of its regulation. In addition, simultaneous investigation of several databanks and approaches will provide a better comprehensive understanding of the VDR gene, and will help improve the clinical management, treatment, and responses to vitamin D.
Materials and Methods
In this investigation, the desired locus is the Homo sapiens VDR gene (NM_000376) that is located on 12q13-14 and is composed of 11 exons (16). Some databases were used for identifying regulatory elements in this VDR gene. The databases and software used for analyzing the VDR locus were presented in Table 1.

Results:
Analysis of CpG Islands
Analyzing CpG islands by the UCSC browser indicated the presence of 4 CpG islands in the VDR. CpG Islands 71 and 77 have a length greater than 300 bp, while, CpG Islands 24, and 16 have length of less than 300 bp (Fig. 1). Further analysis performed by “bona fide” represented that two CpG Islands on (1060 and 1062,) VDR gene overlapped with CpG Island 71 predicted by the UCSC browser (Fig. 1).

In addition, analysis of the VDR gene using “Weizmann Evolutionary CpG islands” pointed five different CpG Islands that overlap with exons except for CpG Island 0.9 (Fig. 1). Moreover, the “CpGProD” program detects six putative CpG Islands (CpG1, CpG2, CpG3, CpG4, CpG5, and CpG6) that all have length more than 500 bp but they have lower CpGo/e ratio compared with UCSC browser (Fig. 1).

Table 1. Databases and software used for analysis VDR locus.

<table>
<thead>
<tr>
<th>Application</th>
<th>Program/database</th>
<th>Web sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpG Islands Detection</td>
<td>UCSC</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td></td>
<td>Bona fide CGIs</td>
<td><a href="http://epigraph.mpiinf.mpg.de/download/CpG_islands_revisited/">http://epigraph.mpiinf.mpg.de/download/CpG_islands_revisited/</a></td>
</tr>
<tr>
<td></td>
<td>CpGProD</td>
<td>edhttp://pbil.univ-lyon1.fr/software/cpgprod.html</td>
</tr>
<tr>
<td></td>
<td>Weizmann Evolutionary</td>
<td>UCSC (<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>)</td>
</tr>
<tr>
<td>determination of the CGI's methylation status</td>
<td>Bona fide CGIs</td>
<td><a href="http://epigraph.mpiinf.mpg.de/download/CpG_islands_revisited/">http://epigraph.mpiinf.mpg.de/download/CpG_islands_revisited/</a></td>
</tr>
<tr>
<td>identifying histone modifications</td>
<td>UCSC</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td>Finding repeated sequences</td>
<td>Sequence-based</td>
<td><a href="http://www.igs.cnrs-mrs.fr/SERV/">http://www.igs.cnrs-mrs.fr/SERV/</a></td>
</tr>
<tr>
<td></td>
<td>Estimation of Repeat</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variability</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td></td>
<td>Repeat masker</td>
<td></td>
</tr>
<tr>
<td>DNase I hypersensitive clusters</td>
<td>UCSC</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td>Transcription factor binding sites</td>
<td>PReMode</td>
<td><a href="http://genomequebec.mcgill.ca/PReMod/">http://genomequebec.mcgill.ca/PReMod/</a></td>
</tr>
<tr>
<td>Identification of transcriptional targets</td>
<td>CisRed</td>
<td><a href="http://www.cisred.org/">www.cisred.org/</a></td>
</tr>
<tr>
<td>Identifying transcriptional enhancers</td>
<td>TRRUST</td>
<td><a href="http://www.grnpedia.org/trrust/">http://www.grnpedia.org/trrust/</a></td>
</tr>
<tr>
<td></td>
<td>Struct2Net</td>
<td><a href="http://cb.csail.mit.edu/cb/struct2net/webserver/">http://cb.csail.mit.edu/cb/struct2net/webserver/</a></td>
</tr>
<tr>
<td></td>
<td>EnhancerAtlas</td>
<td>(<a href="http://enhanceratlas.org/">http://enhanceratlas.org/</a>)</td>
</tr>
<tr>
<td>Prediction of CTCF sites</td>
<td>CTCFBSDB</td>
<td><a href="http://insulatordb.uthsc.edu">http://insulatordb.uthsc.edu</a></td>
</tr>
</tbody>
</table>
Figure 1. Organization human VDR gene. Numbered exons are indicated by orange boxes. The different promoters are represented by red boxes. The CpG Island positions that result from different programs and databases are represented by colored boxes.

Analysis of CpG Methylation

There is a remarkable homology between certain DNA properties of CpG islands and their epigenetic statues, which were used to score all CpG islands according to their strength. High CpG island strength is indicated by the absence of DNA methylation, frequent promoter activity and open chromatin structure (16). Based on what was presented, we found one bona fide CpG island with a combined epigenetic score > 0.67 (CpG Island 1062) and another with a score > 0.33 (CpG Island 1060)(Fig. 1). In addition, CCGC motif analysis shown that there are several CCGC motifs within CpG Island 1062 (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg18 &lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A4652158746585081&hgsid=603289241_vDNe68Zs8Ty78FrjURMU1aMldoYD).

Histone Modifications analysis

The presence of Histone Marks (H3K4Me1, H3K4Me3, and H3K27m3) within the VDR gene sequence was investigated to analysis histone methylation. The H3K4me1 histone mark was the mono-methylation of lysine 4 of the H3 histone protein, and it is associated with enhancers in addition to DNA regions downstream of transcription starts (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=603289241_vDNe68Zs8Ty78FrjURMU1aMldoYD&c=chr12&g=wgEncodeRegMarkEnhH3k4me1).

The H3K4Me3 histone mark was the tri-methylation of lysine 4 of the H3 histone protein, and it is associated with promoters that are active or poised to be activated. Results revealed the presence of regions that had H3K4Me1, H3K4Me3, and H3K27m3 in different cell lines of the ENCODE group. These regions were mostly overlapped with the promoter, CpG Island 1062 and the other CpG islands (Fig. 2). Also, studying the acetylation of H3K27 and H3K9 obtained using ChIP-seq by ENCODE project indicated that H3K27ac, its promoter as well as CpG Islands were in VDR exon 1, while H3K9ac was in exons 1, 3, 9 and introns 2, 8 (Fig. 2).
Investigation of different classes of repeats

UCSC browser (the RepeatMasker software) were used for studying repeated sequences in VDR locus. Results showed that there were different classes of repeats include several long interspersed nuclear elements (LINE) that overlap with CpG Island 1060. In addition, many short interspersed nuclear elements (SINE), mostly family Alu, a few long terminal repeat elements (LTR) were found which include retroposons, some DNA repeat elements (13), several simple repeats (microsatellites), few low complexity repeats (Supplementary Fig. 1). It was found that none of these repeats overlaps with any CpG Island. Moreover, tandem repeats were analyzed by "Sequence-based Estimation of Repeat Variability" and indicated that there are over ten tandem repeats in VDR locus (supplementary Table 1).

In silico Transcription Factor Binding Sites analysis

PReMode and CisRed databases were used for analyzing Transcription Factor Binding Sites. PReMode database describes more than 100,000 computational predicted cis-regulatory modules within the human genome (17). The results of PReMode indicated that there were four cis-regulatory modules (transcription factor binding sites) with different lengths in VDR locus. These modules have binding sites for some transcription factors. However, one module was overlapped with the CpG Islands in the VDR gene (Fig. 3a). For instance, one of these modules has been presented in Figure 3b. Also, the CisRed database identified two conserved DNA sequence motifs in the VDR gene (in promoter and exon 1, Fig. 3c).

Figure 2. Schematic diagram of Histone modifications patterns of ENCODE Consortium for different cell lines. Bars indicate histone modification regions and their height represents repeated times of these modifications.

Figure 3. Transcription Factor Binding Sites analysis using PReMode and CisRed databases. (a) PReMode and CisRed identified four (Green) and two (red) conserved DNA sequence motifs in the VDR gene respectively that two motifs were overlaped with CpG Islands (purple) in the VDR gene. (b) One of the PReMode modules is shown in detail. Colored rectangles indicate different transcription factors binding sites. (c) DNA sequences of two CisRed modules.
Transcription factors regulating VDR and transcriptional targets

VDR downstream transcriptional targets and VDR regulators were investigated from TRRUST v2 (Table 2). Those were activators and repressors but some of them have unknown functions. TRRUST v2 (Transcriptional Regulatory Relationships Unraveled by Sentence based Text mining) is a database consisting of 844 regulatory interactions for 800 TFs in humans (18). Also, a functional network of VDR target genes and a protein-protein interaction network of VDR regulator genes were indicated in supplementary Figure 2 (http://www.grnpedia.org/trrust/).

<table>
<thead>
<tr>
<th>Mode of Regulation</th>
<th>VDR Target genes</th>
<th>VDR regulator genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td>CALB1, CD14, CDH1, CDKN1B, CST5, CYP2A1, CYP3A4, HOXA10, NCOR1, TNFRSF21, TPT1</td>
<td>JUN, POU1F1, SMAD3, TP63, WT1, ZNF224</td>
</tr>
<tr>
<td>Repression</td>
<td>CYP7A1, DKK4, EGFR, IGFBP3, IL10, IL12B, IL2, IL6, PTH, SKP2</td>
<td>KLF4, SNAI1, SNAI2, STAT3</td>
</tr>
<tr>
<td>Unknown</td>
<td>AMH, ASAP2, BGLAP, BHLHE40, CAMP, CXCL8, CYP1A1, CYP3A, CYP3A5, CYP3A7, DDIT4, ERBB2, FOXP3, HLA-DRB1, PHB, PLD1, PODXL, PRL, TLR10, TRPV6, UCP1</td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, Struct2Net was used for finding VDR protein-protein interactions (Structure-based computational predictions of protein-protein interactions). This algorithm used logistic regression and computed a single score (between 0 and 1). A score of 0 indicated minimal confidence in the possibility of an interaction between the two proteins while a score of 1 indicated maximum confidence (19). It was found that a total of 15 predicted protein-protein interactions, were not yet experimentally observed (supplementary Table 2).

Identifying transcriptional enhancers

VDR transcriptional enhancers were investigated by EnhancerAtlas. The present results showed that there were several enhancers affecting the VDR gene. Some of these enhancers were located near transcription start sites (TSSs) of the VDR gene, while the others were at significant distances from the gene’s TSS. The number of these enhancers in different kinds of cells was more than three enhancers (supplementary Fig. 3).

Searching CTCF sites in VDR locus

The UCSC Browser (ENCODE) and the “CTCFBSDB” database were used for finding CCCTC-binding factor (CTCF) binding sites. Results indicated the presence of eleven CTCF binding sites in the VDR gene from ENCODE. Some of these binding sites were located in CpGI1, CpGI3, CpGI5, CpGI6 identified by CpG ProD, and mostly overlapped with CpGI62 as predicted by Bona Fide. In addition, some of the CTCF binding sites overlap with the VDR promoter (Fig. 1). However, only one CTCF binding site, in the VDR gene located in the nucleotide position 9491, was identified in the VDR gene using the “CTCFBSDB” database was identified. The sequence of this insulator was AGTACCAGCAGGTGGCACAC.

Discussion:

Recent studies have shown that the biological actions of vitamin D are mediated through VDR (20). Unfortunately, precise control and molecular mechanisms of this receptor expression are still not fully understood. In this study databases and in silico tools were used for identifying putative regulatory elements and gene specifications that might mediate regulation of VDR. The computational identification of these properties and elements presented new insights into the understanding of the control of human VDR expression and its association with health and disease. This bioinformatics analysis supports the idea that different genetic and epigenetic elements are involved in the regulation of the activity and expression of the VDR gene.

An in silico study reported by Halsall et al. (21) shown that exon 1a in VDR was within a strong CpG island and thus, transcription from exon 1a may be regulated by methylation. Mark et al. (22) found that some CpGIs are in a region from −790 bp to +380 bp downstream of the VDR transcription start site. They observed that these CpG islands were methylated in breast cancer. In addition, promoter methylation of the VDR gene was found in adrenocortical carcinoma (13). It was not in parathyroid adenomatous tissues (23).

This investigation showed that there were additional CpG islands in the VDR gene. The association of these CpG Islands with other information including transcription factor binding sites, histone modifications, enhancers, and genomic
insulators was studied. In the other words, because of the overlapping regulatory elements with CpG islands at VDR locus, DNA methylation could be modulated TF binding sites at the VDR gene, interacts with histone modification of VDR promoter and controls enhancers/insulators (24).

In addition, because in this study, it was found that some repeated elements at VDR locus overlap with CpG Island, in the other hand, number of these repeats are variable in different individuals (25), so DNA methylation events of VDR promoter are variable, and this could be used to personalized medicine (26).

The present results demonstrated that histone marks overlapped with CpG Island at VDR locus. The number and distribution patterns of histone modifications regulate the expression of genes. The overlapping two epigenetic events at VDR locus may increase, decrease or counteract the effects of each other that further investigations are needed to study these effects.

A comprehensive realization of the actions of vitamin D and its receptor VDR needs analysis of VDR binding sites. The VDR recognizes a specific DNA element comprised of two hexameric nucleotide half-sites separated by three base pairs (27). By studying the common VDR binding site (VDRE) in the regulatory sequences of VDR targets, the results indicated the presence of 50 proteins that may be targeted for VDR. According to these results and the other researches (28, 29), VDR activates some genes and represses others, but the effects of VDR on some genes are unknown. Therefore, functional studies need to investigate the action of VDR precisely and further. Furthermore, it was presented that VDR is a target for many transcription factors. In these cases, too, their effects on VDR expression are unknown. Thus, the roles of VDR in this regulatory network should be examined.

Through the investigation of the regulatory elements such as enhancers and insulators on VDR locus. Several enhancers located at distances from the VDR promoter region were found. The previous studies suggest that regulatory enhancers around promoters may be more usual than once believed and is presenting the important role of chromatin looping and chromatin rearrangement in transcription control (30). In addition, modification at some of these enhancer regions may be significant to following changes in the level of VDR transcription (31). The precise function of some enhancer regions at VDR locus has not yet been determined due to their unknown existence in several human cells and tissues.

The results also revealed that a few insulators in VDR locus, located in VDR CpG islands, overlapped with histone modifications. CTCF is a factor that plays important roles in genomic processes including imprinting, transcription, chromatin rearrangements and chromatin interactions (32). CTCF be able to bind to a wide range of sequences and regulate expression of genes via its functions as a repressor or activator. In addition, its binding site has an insulator role (33). Wang et al. (34) reported the presence of several CpGs at the CTCF consensus motif in different cell types. They also showed that methylation is indeed a global feature of the regulatory diversity of CTCF. Therefore, our results show crosstalk among CTCF binding sites (insulator), CpG methylation and histone modifications that may affect VDR gene regulation and these effects could be varied in different cell types and tissues.

In conclusion, these findings from the bioinformatics sequence analyses and databases showed that there is crosstalk among specific regulatory motifs, various set of CGIs methylation, transcription factors and histone modifications which have combinatorial influences on VDR gene expression. These can provide targets for further and more precise functional analysis, even a promising therapeutic strategy for different diseases. However, these regulations should be examined experimentally to help disease treatment and prevention. Finally, our data provide new insight into the action of VDR as a modular component in regulatory pathways that control the different processes.

Acknowledgments

This study was supported as a project in Shahid Beheshti University, Tehran and it was not received any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' declaration:
- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for republication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Shahid Beheshti University G.C..

References:
2. Dursun E, Gezen-Ak D. Vitamin D receptor is present on the neuronal plasma membrane and is co-

468

تحليل في السليكون للعناصر التنظيمية لجينات مستقبلات فيتامين دي
1
stein Fazel 1
روبة أمير نجاد 2
بهار نقوی کرکری 3
سیده بتول حسنی 4
ژنیب شیرفانی فارسی 1

الخلاصة:
مستقبل فيتامين دي (VDR) هو عامل النسخ النووي الذي يتحكم بالتنفس الجيني. تم التوصل إلى أن الضعف في التنفس في الجين VDR يرتبط بأمراض مختلفة. يقوم VDR بمهمة تنظيم مسارات مختلفة مثل التثبيت، الاندماج، والتعبير عن الكلاسيكو وأدوات الفوسفات. إلخ. مع ذلك لا توفر معلومات كافية حول تنظيم الجين نفسه. لذلك فإن فهم أكثر وضوحاً للعوامل التنظيمية لـ VDR يمكن من خلال استخدام استراتيجيات القناعات من الأمراء المتعلقة بـ VDR وعلاقة النصوص أو الدراسات على مجموعة من قواعد المعلومات المتوزعة في VDR. وتم تصميم عناصر وقواعد DNA المتماثلة في VDR. بالإضافة إلى تفاعلات بين البروتينات في VDR، طهران، إيران.

الكلمات المفتاحية: جزر CpG، علم التخلق، تحليل في السليكون، عناصر التنظيم، VDR