Multifunctional viral protein R of human immunodeficiency virus-1 as a potential drug target

Vivek Darapaneni*

Department of virology and computational biochemistry, Saket Institute for Biomolecular Research, Visakhapatnam, India 530022

ABSTRACT

The viral protein R (Vpr) plays a pivotal role in the infectious lifecycle of human immunodeficiency virus-1. The objective of this study is to find the degree of conservation of Vpr and to detect conserved binding sites, which might be used as target sites for potential anti-Vpr drugs. The conservation analysis was based on 5301 amino acid sequences identified novel conserved and highly conserved sites. The novel conserved sites which have been identified are: Leu42, Gly43 and Val57; Arg73 and Cys76; Glu24, His33, Cys76 and Ser79. The outcome of this study provide the foundation for developing anti-Vpr drugs which have abridged potential to induce drug resistance through mutations.

ARTICLE HISTORY: Submitted: 13/8/14; Revised: 29/9/14; Accepted: 8/10/14

KEY WORDS: Human immunodeficiency virus, Viral Protein R, Drugs, Mutation, Resistance, Binding sites

Abbreviations: HIV, Human Immunodeficiency Virus; Vpr, Viral Protein R; HAART, Highly Active Antiretroviral Therapy; MSA, Multiple Sequence Alignment

*: Correspondence author, Tel: +91-986-68 90 432
E-mail: darapanenivivek.sibr@gmail.com
Human immunodeficiency virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS), affecting approximately 2.5 million people and resulting in 1.5 million deaths annually [1]. There are two types of HIV namely, human immunodeficiency virus type 1 (HIV-1) and human immunodeficiency virus type 2 (HIV-2) [2, 3]. Since the discovery of HIV-1 (which was way back in 1983) up till now, HIV-1 remains to be foremost concern regarding public health globally [4]. By comprehending the devastating nature of HIV-1, the researchers worldwide are unified to find solutions which will help in fighting this elusive virus. As a result of robust and concerted efforts, the highly active antiretroviral therapy (HAART) was developed [5-7]. HAART has drastically enhanced and increased the lifespan of HIV patients; nevertheless it is unable to eliminate the virus totally from the HIV patients. This is due to the fact that infected cells are resistant to the apoptosis which is conferred by HIV. Without eliminating these cells it is impossible to eradicate the virus totally from an infected individual. Therefore is of paramount importance to develop antiretroviral drugs against those proteins which are responsible for conferring apoptosis resistance to the infected cell.

HIV-1 is a RNA virus which belong to lentivirus class of retrovirus whose genome encodes three structural proteins (gag, pol and env), two regulatory proteins (tat and rev) and four accessory proteins (vpu, vif, nef and vpr) [8, 9]. The viral protein R (Vpr) is a 14KDa multifunctional, basic protein of length 96 amino acids [10]. Vpr is known to exist as dimers, trimers, tetramers and high order multimers [11]. The multiple roles of Vpr identified are: (i) Vpr modulates the viral genome transcription [12] (ii) Vpr also play critical role in apoptosis induction, cell cycle control disruption and introduction of defects in mitosis [13] (iii) Vpr is responsible for nuclear import of pre-integration complex [14] (iv) Vpr facilitates the reverse transcription of HIV-1 genome [15] (v) Vpr suppresses the activation of host immune [16] (vi) Vpr helps in alters the mutation rate of HIV-1 [17] (vii) Vpr inhibits normal cell growth by arresting G2/M transition [18].

The emergence of HIV-1 mutants which are drug resistant and the spreading of these mutants globally are serious cause of concern. To overcome this problem one has to exploit accessory proteins as a potential drug targets and understand the conservation pattern of these accessory proteins. At present, there are no licensed drugs available targeting the Vpr protein of HIV-1. The aims of this study is to identify the degree of conservation of Vpr in order to detect the conserved region which can be potential drug target sites and identify potential ligand binding sites which can act as foundation for future drug development which target Vpr.
Methods

1. Sequence analysis and protein structure retrieval

The full length sequences of the Vpr protein were obtained from UniProt [19]. For alignment of the collected proteins sequences, MUSCLE version 3.8 [20] was used with default parameters. Multiple refinements of the obtained alignment were carried out resulting in 26-30 iterations, until no further improvement was attained. For sequence conservation plot, plotcon program (EMBOSS package) was used with comparison matrix EBLOSUM62 (default) and window size of 10 residues [21]. The structure of Vpr protein (residues 1-96) was obtained from Protein Data Bank (PDB) [22] with the entry 1M8L [23].

2. Conservation and binding site analysis

By giving multiple sequence alignment and protein structure file as an input, conserved regions in VPR protein were identified and mapped onto the Vpr protein structure using ConSurf server. (http:// consurf.tau.ac.il/) [24-27]. By taking evolutionary relationships among protein sequences into account, ConSurf algorithm produces meaningful conservation scores. The conservation score given by ConSurf server is divided into scale of nine grades which are given for the purpose of visualization. Most variable positions in the protein are placed in grade one (turquoise), intermediately conserved positions are placed in grade five (white), and most conserved positions are placed in grade nine (maroon) [24-27]. Ligand binding sites (LBS) on Vpr protein structure were identified using COFACTOR (http://zhanglab.ccmb.med.umich.edu/COFACTOR/), identifies the LBS using both global and local with templates from PDB and match local motifs of the identified template with that of query structure [28], TM-site (http://zhanglab.c cmb.med.umich.edu/COACH/), identifies the LBS using intermediary approach if both local and global alignments [29, 30], S-site (http://zhanglab.c cmb.med.umich.edu/COACH/) identifies the LBS by explicitly comparing binding site specific sequence profiles [29,30] and SiteHound (http://scbx.mssm.edu/sitehound/sitehound-web/Input.html) which uses energy based method to find regions with high potential for ligand interactions. The structure of the protein is characterized by different probes for identification of binding sites of different types [31].

Results

1. Multiple alignments of protein sequences and Vpr protein structure

For the HIV-1 Vpr protein 5301 sequences were obtained from UniProt [19]. The overview of multiple sequence alignment of Vpr protein sequences is shown in the figure 1. The residues 1-51 form an N-terminal domain and residues 52-96 form C-terminal domain. The obtained structure of Vpr protein (shown in figure 2) contains three α-helices: helix 1...
Helix 1 and helix 2 are located in N-terminal domain while, helix 3 is located in C-terminal domain. Helix 1 and helix 2 are joined by a short inter domain loop which is five amino acids in length. The two domains are connected by a short linker region which is four amino acids in length.

2. Conserved and variable residues

The variable and conserved residues in the Vpr protein were identified using ConSurf server [24-27] and are illustrated in figure 2. The variable residues of grades 1-3 and conserved residues of grades 7-8 are clustered together in Table 1. The conservation scores were projected onto the spacefill models of the Vpr protein in figure 2. Conservation scores for the Vpr protein were obtained between -0.777 (maximum conservation) and 5.111 (maximum variability) by the ConSurf server [24-27]. In general the Vpr protein is conserved with

Figure 1: Outline of the multiple sequence alignment for the Vpr protein. The line graph illustrates the degree of conservation with the window size of ten residues obtained using plotcon program [10].
Figure 2: The spacefill representation of the Vpr protein with conservation grades mapped onto the structure. Important residues position which showed highest conservations and variations are labelled.
approximately 59% of the residues belong to grades 7-9 (conserved), while 24% of the residues belong to grades 1-3 (variable). Altogether, twenty six residue positions (approximately 27% of total residues) were found to be highly conserved (grade 9) and are shown in the table 1. The residue positions 1, 2, 27, 33, 53, 57, 71, 73 and 78 showed highest conservation among all the sequences analyzed. In total, sixteen residues (16.7% of total residues) were found to be highly variable (grade 1). The residue positions 15, 37, 41, 45 and 84 showed highest variation among all the sequences analyzed. The N-terminal loop (residues 1-17) was found to be variable in nature with only 41% of the residues were conserved. Helix 1 and helix 2 were found to be conserved with 75% and 63% of the total residues present in these helices belong to grades 7-9 respectively. The inter domain loop connecting helix 1 and helix 2 was found to be conserved with residues Phe34, Pro35 and Trp38 belonging to grades 8-9, residue Arg36 belonging to grade 6 (intermediately conserved) while, residue Ile37 was found to be highly variable (grade 1). The linker region joining the two domains was found to be completely conserved. The C-terminal loop (residues 78-96) was found to be intermediately conserved.

3. Ligand binding sites

Putative ligand binding sites were identified using COFACTOR (detects LBS by global and local alignments with template structures in PDB), TM-site (detects LBS by transitional approach balancing global and local alignments), S-site (detects LBS by comparing sequence profiles which are binding site specific) and SiteHound (detects LBS by positive interaction between a chemical probe and protein structure) algorithms [28-31]. The detected LBS are shown in table 2.

<table>
<thead>
<tr>
<th>Residue Classification</th>
<th>Vpr protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly conserved (grade 9)</td>
<td>Met1, Glu2, Pro5, Arg12, Leu23, Glu24, Lys27, Glu29, Ala30, His33, Phe34, Pro35, Leu39, Gly43, Tyr47, Thr53, Gly56, Val57, Leu64, Gln65, His71, Arg73, His78, Ser79, Ile81, Lys95</td>
</tr>
<tr>
<td>Conserved (grades 7-8)</td>
<td>Gln8, Gly9, Pro14, Trp18, Leu20, Glu21, Leu26, Val31, Arg32, Thr38, Leu42, Gln44, Ile46, Thr49, Tyr50, Gly51, Asp52, Arg62, Gln66, Leu68, Phe69, Ile74, Gly75, Cys76, Arg80, Gly82, Ile83, Arg88, Arg90, Asn91, Gly92</td>
</tr>
<tr>
<td>Variable (grades 1-3)</td>
<td>Ala4, Glu6, Tyr15, Asp17, Thr19, Asn28, Ile37, Ser41, His45, Glu48, Thr55, Glu58, Leu60, Ile61, Ile63, Ile70, Arg77, Ile84, Gln85, Gln86, Thr89, Ala93, Ser94</td>
</tr>
</tbody>
</table>

Table 1: Variable and conserved residues of Vpr proteins of HIV-1 identified using the ConSurf server [24-27]
The COFACTOR identified ten LBS, namely C1-C10 ('c' denotes binding sites identified by COFACTOR). The binding sites C1, C5, C6, C7 and C9 identified by COFACTOR were found to span both N-terminal and C-terminal domains. Further C8 and C2 were found to be located in N-terminal and C-terminal domains respectively. Further binding sites C3, C4 and C10 were found to be located at the interface between N-terminal and C-terminal domains. TM-site identified five LBS, namely TM1-TM5 ('TM' denotes binding sites identified by TM-site). Except for TM5 (located in the N-terminal domain), all the other binding sites are located in the C-terminal domain. S-site identified five LBS, namely S1-S5 ('S' denotes binding sites identified by S-site). The binding site S2 and S3 were found to be located at the interface between the two domains, while S4 and S5 are located in the N-terminal regions.

Further binding site S1 was found to span the two domains. SiteHound identified forty LBS, of which only ten were selected based on the number of conserved residues present in the binding site. The identified LBS are SH1-SH10 ('SH' denotes binding sites identified by SiteHound). Except for SH5 and SH8 which are located in the N-terminal region, all the remaining binding sites were found to span the two domains. As shown in the table 2 binding sites C10, TM4 and S1 were found to be completely conserved, while binding site TM5 was found to be variable. The remaining binding sites showed varying degree of residue conservation and are shown in the table 2.

**Discussion**

1. **Sequence conservation and novel binding sites of the Vpr protein**

The objectives of this study was to establish the degree of conservation of Vpr and to identify binding sites in the conserved regions, which might form prospective target sites for anti-HIV drug discovery in the future. Protein sequence conservation crop up, owing to sites which are either functionally important (for instance protein-protein interaction or signals) or structurally important \[^{[32]}\]. In contrast variable sites emerge as a result of either adaptation or evolutionary pressure to evade the immune system.

The N-terminal domain was found to be more conserved when compared to C-terminal domain. Previously it was shown that mutation Phe72Leu resulted in diminished nuclear accumulation and lessened Vpr incorporation into newly formed virions \[^{[33]}\]. The position Phe72 was found to be intermediate conserved belonging to grade 5. Likewise mutation Arg77Gln was shown to diminish Vpr cytopathicity \[^{[34]}\]. The position Arg77 was found to be highly variable belonging to grade 1. It was observed that deletion of residues 83-89 resulted in Vpr losing its ability to induce G2 arrest and cell
Table 2: The putative ligand binding sites identified using COFACTOR, TM-site, S-site and SiteHound algorithms [28-31].

<table>
<thead>
<tr>
<th>Site</th>
<th>Amino acid residues positions on protein structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42, 56*, 59, 63, 66</td>
</tr>
<tr>
<td>3</td>
<td>40, 43*, 56*</td>
</tr>
<tr>
<td>4</td>
<td>44, 45, 47*, 51, 56*</td>
</tr>
<tr>
<td>5</td>
<td>38, 42, 45, 64*, 72, 75, 86</td>
</tr>
<tr>
<td>6</td>
<td>43*, 44, 56*, 60, 63, 64*, 67, 68, 71*, 74, 83</td>
</tr>
<tr>
<td>7</td>
<td>28, 43*, 61, 62</td>
</tr>
<tr>
<td>8</td>
<td>19, 26, 30*</td>
</tr>
<tr>
<td>9</td>
<td>27*, 28, 42, 43*, 44, 46, 53*, 60, 61, 64*, 65*</td>
</tr>
<tr>
<td>10</td>
<td>42, 43*, 57*</td>
</tr>
</tbody>
</table>

Note: Conserved residues (grade 7-8) are shown in bold face while highly conserved residues (grade 9) are shown in bold face with asterisk mark. Intermediately conserved residues (grade 5-6) are underlined.

dead [35, 36]. Residue positions 83 and 88 were found to be conserved residues (grades 7-8), residue position 87 was found to be intermediately conserved (grade 6), residue positions 84, 85 and 89 were found to be highly variable (grade1) and residue position 86 was found to be variable (grade 2). Previously it was reported that Vpr-Cyclophilin A interaction was mediated by Pro14 and Pro35 [37]. While Pro14 was conserved at grade 8, Pro35 was found to be highly conserved at grade 9. Mutation of Arg80, Arg87 and Arg88 to alanine was shown to result in loss of Vpr’s ability to induce G2 arrest [38]. Residues Arg80 and Arg88 were found to
be conserved residues (grade 8), while Arg87 was found to be intermediately conserved (grade 6).

Recently it was shown that residues Trp54 and His71 play a vital role in Vpr dimer formation \[^{[39]}\]. Trp54 was found to be intermediately conserved (grade 6), while His71 was found to be highly conserved (grade 9). Zhou and Ratner showed that phosphorylation of residue Ser79 is associated with G2 arrest \[^{[40]}\]. Ser79 was found to be highly conserved (grade 9). Position 54 play an decisive role in Vpr-uracil DNA glycosylase interaction \[^{[41]}\]. Position 54 was found to be intermediately conserved (grade 6).

22LLEEL26 and 64LQQLL68 motifs are required for coactivation (Vpr mediated) of glucocorticoid receptors \[^{[42]}\]. Residue positions 23, 24, 64 and 65 were found to be highly conserved (grade 9); residue positions 26, 66 and 68 were found to be conserved (grades 7-8); positions 25 and 67 were found to be intermediately conserved (grade 5) and residue 22 was found to be variable (grade 4). Mutations Phe34Ile and His71Arg resulted in lost ability of Vpr to localize in perinuclear area \[^{[14]}\]. Both Phe34 and His71 were found to be highly conserved (grade 9).

Most of the predicted LBS identified on the Vpr protein are surrounded by conserved residues. Targeting these LBS using small molecule ligands will most probably result in disruption of Vpr function. Among the predicted LBS identified by different methods, few novel binding sites were detected. These novel sites may be either involved in new protein-protein interactions or sites of known protein-protein interactions. The novel binding sites identified are C3, C10, TM4, S1, S4 and SH5. These novel sites are completely conserved and the functions of these sites are not known from previous studies. Targeting these sites using small ligand molecule would inhibit the functions of the Vpr protein. Apart from the completely conserved sites, novel sites which are mostly conserved were also identified. These sites are C4, C8, TM1, TM2, TM3, S2, S3, S5, SH1, SH3, SH4, SH6, SH8 and SH10. These sites might be functionally important which can be explained by the degree of conservation found in these sites and the functions of these sites are not elucidated yet.

In conclusion the study of 5301 Vpr protein sequences divulge an elevated level of sequence conservation pattern that intersects with prospective LBS rendering the VPR protein an exceptional drug target. Targeting the conserved binding sites identified in this study using small drug molecule will effectively diminish the activity of the Vpr protein. Moreover, anti-Vpr drugs targeting these conserved
sites are less likely to become ineffective due to drug resistance in the future.

References


