

Negative Sense Transcripts in HIV-1: The Regulation of Expression and Role of vpo/Vpo.

Michael S Barbagallo (Supervisors Nick Deacon and Jenny Mosse)
School of Applied Sciences and Engineering

Introduction

The positive sense strand of the HIV-1 genome encodes nine different proteins, as depicted in Figure 1 below. These include structural proteins (Gag, Pol and Env), regulatory proteins (Tat and Rev) and the accessory proteins (Vpu, Vif, Vif and Nef) (Schwartz et al., 1992). A tenth, negative sense, transcript encoded by the antisense gene, vpo, has also been identified, the function of which is unknown (Bukrinsky et al., 1990; Miller, 1988; Deacon et al., 2005). Transcription of the vpo gene is controlled by long terminal repeat (LTR) sequences and occurs early in infection, at the same time as regulatory protein gene transcription (Brentley et al., 2004). Negative sense transcripts have been reported in a number of other gene systems, including the Rev-ErbA, a member of the T3steroid hormone receptor family in rats (Lazar et al., 1990), ASM-1, complementary to the human c-myc protooncogene (Celano et al., 1992), and ebna from the Epstein-Barr virus (Prang et al., 1995).

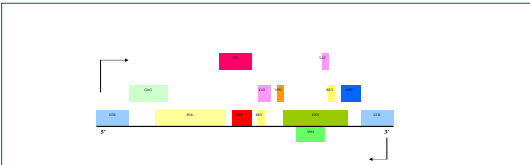


Figure 1. The organization of the HIV-1 proviral genome. Positive sense genes, gag, pol and env (structural proteins), tat and rev (regulatory proteins) and vpu, vpr, vif and nef (accessory proteins) indicated above the line. The negative sense gene, vpo is indicated below the line, its sequence complementary to part of env. (Adapted from Steffy and Wong-Staal, 1991).

The antisense protein, Vpo, is recognised by circulating antibodies of HIV-1 individuals (Vanhee-Brossollet, et al., 1995; Michael et al., 1994). It is primarily associated with various cellular membranes, including the enveloped viral particles released from infected cells (Briquet et al., 2002). Endogenously expressed vpo transcript has also been demonstrated to inhibit replication of HIV-1 (BRU, IIB, NDK), but not HIV-2 (Tageva and Vaquerio, 1997). The detection of Vpo in the membranes of HIV-1 infected cells and the envelope of the virus particle suggests that Vpo may play a pivotal role in the life cycle of the virus. A series of short open reading frames (sORFs) upstream of the vpo gene have been identified and have been shown to affect downstream vpo gene expression in preliminary experiments (Yap, Vardarli and Deacon, unpublished results).

Preliminary analysis of the predicted 189 amino acid vpo sequence predicts two highly hydrophobic transmembrane regions, a cysteine rich-region and a proline repeat motif. The proline repeat sequence motif is similar to the PxxP repeat sequence of the HIV-1 Nef protein (Picard et al., 2002) and the ORF-3 protein in Hepatitis E virus (Ray et al., 1992). In both of these proteins the PxxP region has been shown to interact with cellular protein kinases. Further investigations into the regulation and function of vpo/Vpo are essential to gain further understanding of the life cycle of the HIV-1 virus.

Objectives

This study will: (1) investigate the conservation of Vpo and its sORFs across the strains and subtypes of HIV-1, (2) investigate the mechanism of regulation of Vpo gene expression by its series of upstream sORFs and (3) investigate the targeting of the protein to sub-cellular compartments.

Materials and Methods

Conservation of Vpo Across Strains and Subtypes of HIV-1

A total of 37 HIV-1 complete genome sequences spanning the subtypes, A, B, C, D, E, F1, F2, G, H, J, K, N, O and U were randomly selected and downloaded from The HIV Sequence Database (www.hiv.lanl.gov). In addition, 5 strains of HIV-2 spanning the subtypes A, B and G, 8 sequences from the HIV₂222_3 from African Green Monkey, 2 from Mandrill and 2 from Simulium were also downloaded for comparison of the presence and similarity of Vpo and the vpo gene sequence.

All sequences were scanned for ORFs using the Translation Overview tool in DNAMAN (Lynnsoft BioSoft) in all six reading frames with a line length of 8, minimum ORF length of 6 amino acids and width of 4. The location of the env gene was noted and thus the presence of any large ORF opposite was designated Vpo. The reading frame, amino acid size and position of the ORF was noted and the sequences collected. The DNA sequences collected were translated into amino acid sequences using the Translation tool in DNAMAN. The resultant amino acid sequences were subjected to a multiple sequence alignment using Vector NTI (Invitrogen) with the following parameters: Gap opening penalty of 10, Gap extension penalty of 0.5, Gap separation penalty range of 8 and a % identity for alignment delay of 40 and the similar features conserved across the different strains observed. The consensus sequence generated for Vpo as a result of the multiple sequence alignment was analyzed by the Hydrophobicity profile tool in DNAMAN.

Conservation of sORFs Across Strains and Subtypes of HIV-1

All of the sequences upstream from the Vpo ORF were collected and scanned for ORFs using the Translation Overview tool in DNAMAN in all three minus reading frames with a line length of 8, minimum ORF length of 6 and width of 4. The number, position, reading frame and size of the sORFs were observed for each sequence.

Results and Discussion

Conservation of Vpo Across Strains and Subtypes of HIV-1

Of all the HIV-1 sequences selected only five did not display a sequence similar to the Vpo sequence proposed by Miller (1988) (Table 1A). While they displayed the env gene they did not display an ORF for vpo. Two of these were from subtype A, while all of the subtype O sequences and one from subtype U did not display any vpo like similarities. This indicates that over 85% of the 37 sequences tested display a sequence similar to Vpo. The reading frames vary across all three negative frames, while the size of Vpo also varies quite markedly from 108 to 190aa. In agreement with Miller (1988) in which only 12 sequences from HIV-1 were selected mainly from subtypes A and B, Vpo is quite highly conserved across the subtypes of HIV-1. The fact that Vpo is quite highly conserved (as indicated in Figure 2A and B), is most likely due to its position directly across a portion of the env gene, another highly conserved gene in HIV-1.

A number of amino acid sequence motifs are observed as a result of the multiple sequence alignment (Figure 2A). Firstly, there is an unusual cysteine rich motif located close to the N-terminal of the protein predicted. A similar cysteine motif has also been observed in Oct-4, functioning as a transcription repressor or activator binding site, thus regulating transcription of the gene (Nordhoff et al., 2001). The function of this cysteine rich motif is currently unknown and may represent some form of interaction point or play a role in an activity carried out or mediated by this protein (if it is fully translated into one). Secondly there is the presence of the PxxP motif, which has also been observed in HIV-1 Nef (Picard et al., 2002) and ORF-3 in Hepatitis E virus (Ray et al., 1992). In both of these proteins the PxxP region has been shown to interact with cellular protein kinases; perhaps this may also occur in Vpo. Lastly a hydrophobicity profile of the consensus observed one potential membrane spanning region in contrast to the two predicted by Miller (1988). As depicted in Figure 2C, this region is highly conserved amongst the sequences analyzed indicating its potential as a membrane associated protein.

Sequences from HIV-2 and some SIV sequences were also investigated. None of the HIV-2 sequences displayed any Vpo like similarities as did the major part of the SIV sequences. This may be associated with the higher infectivity of HIV-1 strains. However some of the SIV₂₂₂ strains displayed some sequence similarities to Vpo. This is consistent with SIV₂₂₂ being an ancestor of HIV-1, while the other SIV sequences are ancestors of HIV-2.

Conservation of sORFs Across Strains and Subtypes of HIV-1

Conservation of the sORFs upstream of Vpo was also assessed. As displayed in Table 1A and B, the number of sORFs varied from 1 through to 8. Subtypes B and C displayed a more consistent number of sORFs averaging from 5 to 7. An example (Figure 3) of the N1-43 strain (Subtype B) contains six sORFs upstream of Vpo varying in size from 6 to 48aa. The occurrence of sORFs or upstream AUGs (uAUGs) has been identified to affect the efficiency of translation in a number of gene systems. These include the suppressor of cytokine signalling 1 protein (SOCS-1) (Schuler et al., 2000), human β_1 -adrenergic receptor gene (Evanko et al., 1998), and cathelicidin (Wu et al., 2002). While in most cases the last HIV-1 Vpo sORF is quite well conserved amongst the sequences selected (data not shown) their role in the regulation of Vpo (if any) and the role of Vpo itself is yet to be determined.

Conclusions

Vpo is clearly quite conserved across the strains and subtypes of HIV-1 and some strains of its ancestor SIV₂₂₂. The fact that some strains of SIV₂₂₂ also displayed Vpo like ORFs suggests that Vpo may be introduced into viral genome via a chimpanzee pathway, and the pathway into which Vpo was introduced may play a role in unlocking the function of this gene. Finally a series of sORFs is present upstream of Vpo and may function as regulators of vpo expression.

References

Bentley, K., Deacon, N., Soria, S., Zachar, S., and Chuchmi, M. (2004) "Mutational analysis of the HIV-1 LTR as a promoter of negative sense transcription." *Archives of Virology* 148, 2277-2284

Briquet, S., and Vaquerio, C. (2002) "Immunoblotting Studies of an Antisense Protein in HIV-1-Infected Cells and Viral Particles." *Virology* 292, 177-184.

Bukrinsky, M.I., and Ellis, A.J. (1990) "Pox Strand of the HIV Provirus DNA is Expressed at Early Stages of Infection." *AIDS Res Human Retroviruses* 6, 425-435.

Calvez, P., Barthelet, C.H., Kiser, D.L., Wessinger, A., Rubin, R.D., Beale, S.R., and Cohen, J.S. (1993) "Characterization of an Endogenous RNA Transcript with Homology to the Antisense Strand of the Human α -Globin Gene." *The Journal of Biological Chemistry* 268, 1502-1506.

Escueta, D.S., Ellis, C.E., Vancanach, V., and Friauf, T. (1998) "Regulation of the Transcriptional Repression of the Human β_1 -Adrenergic Receptor Gene." *Biochem. and Biophys. Res. Commun.* 244, 385-402.

Lezar, M.A., Rodin, R.A., Cardona, G., and Chen, W.Y. (1990) "Gene Expression from the c-Ha-Ras-ErbA Oncogene Locus." *The Journal of Biological Chemistry* 265, 12059-12063.

Michael, N., Valdez, M.T., Dancy, L., Ehrenberg, P.M., Blaise, J.D., Rappaport, J., and Redfield, R.R. (1994) "Negative-Strand RNA Transcripts Are Produced in Human Immunodeficiency Virus Type 1-Infected Cells and Persist by a Novel Proteinase Cleavage Pathway." *The Journal of Virology* 68, 970-977.

Miller, H.H. (1988) "Human Immunodeficiency Virus May Encode a Novel Protein on the Genomic DNA Plus Strand." *Science* 253, 1420-1422.

Nordhoff, V., Habben, K., Bauer, A., Oltros, V., Malespina, A., and Scholer, H.R. (2001) "Comparative Analysis of Human, Bovine, and murine Oct-4 upstream promoter sequences." *Maternal and Child Health* 7, 303-317.

Prang, G., Gorenstein, A., Holmsky, E., Cline, D., and Galbraith, J. (2002) "Interaction with Human T_H1 Tryptase Kinase Reveals Convergent Evolution of the Nef Protein from Simian and Human Immunodeficiency Virus: Deep Divergence Molecular Surface Fusion." *Virology* 295, 320-327.

Prang, N., Weil, H., and Schramm, F. (1995) "Epstein-Barr Virus Lytic Replication is Controlled by Phosphorylation of the Nuclear Protein EBNA1." *The Journal of Virology* 69, 2464-2468.

Ray, R., James, S., Murrell, V., and Ray, S. (1992) "Indian Hepatitis E Virus Shows a Major Deletion in the Small Open Reading Frame." *Virology* 189, 203-206.

Schuler, G., Brandt, S., and Ray, S. (2000) "Regulation of Transcriptional Repression of the SOCS1 Major Open Reading Frame by Its Upstream Open Reading Frame." *Biochem. and Biophys. Res. Commun.* 268, 205-210.

Schwartz, L.F., Taylor, B.K., and Pringle, C.H. (1992) "Nucleotide Sequence and Immunoreactivity of the HIV-1 α -Globin Gene." *Journal of Virology* 66, 970-977.

Steffy, K., and Wong-Staal, F. (1991) "Genetic Regulation of Human Immunodeficiency Virus." *Molecular Biology Reviews* 55, 193-205.

Yap, N.H., Vaquerio, C. (1997) "Expression of Naturally Occurring Antisense RNA Inhibits Human Immunodeficiency Virus Type 1 Helicase Protein Expression." *The Journal of Virology* 71, 2023-2031.

Yap, N.H., Vaquerio, C. (1997) "Expression of naturally occurring antisense RNA inhibits human immunodeficiency virus type 1 helicase protein expression." *Journal of General Virology* 78, 2023-2031.

Yap, N.H., Vaquerio, C., Thomas, H., Sanyal, N., D'Amico, L., Lany, J., and Vaquerio, C. (1995) "A Novel Antisense RNA Derived from the HIV-1 α -Globin Gene Encodes a Protein Which is Recognized by Circulating Antibodies of HIV-1 Individuals." *Virology* 208, 198-202.

Wu, H., Ross, C.R., and Blacha, R. (2002) "Characterization of an upstream open reading frame in the 5' untranslated region of PR-3, a cathelicidin antimicrobial peptide." *Immunology* 25, 5-16.

Yap, D.H., Vardarli, N., and Deacon, N.J. (2005) Unpublished Results.

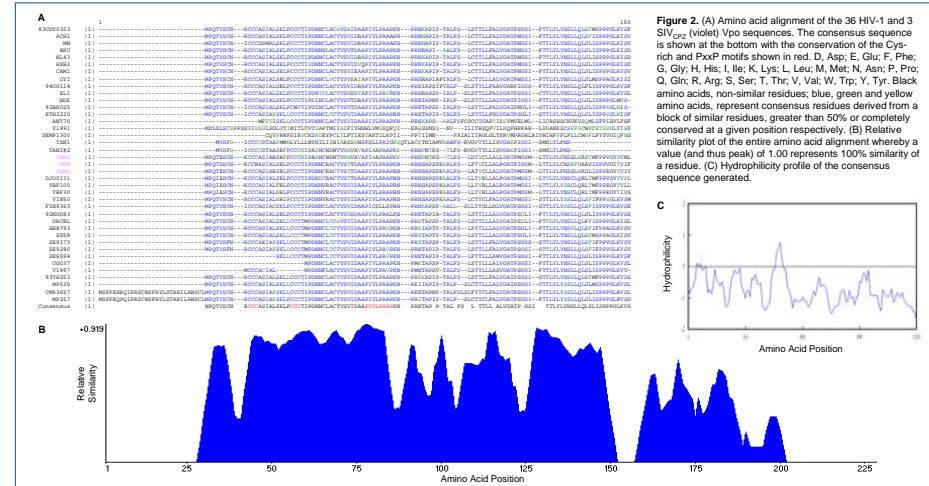


Figure 2. (A) Amino acid alignment of the 36 HIV-1 and 3 SIV₂₂₂ (violet) Vpo sequences. The consensus sequence is shown at the bottom with the conservation of the Cys-rich and PxxP motifs shown in red. D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Black amino acids, non-similar residues; blue, green and yellow amino acids, represent consensus residues derived from a block of similar residues, greater than 50% or completely conserved at a given position respectively. (B) Relative similarity plot of the entire amino acid alignment whereby a value (and thus peak) of 1.00 represents 100% similarity of a residue. (C) Hydrophobicity profile of the consensus sequence generated.

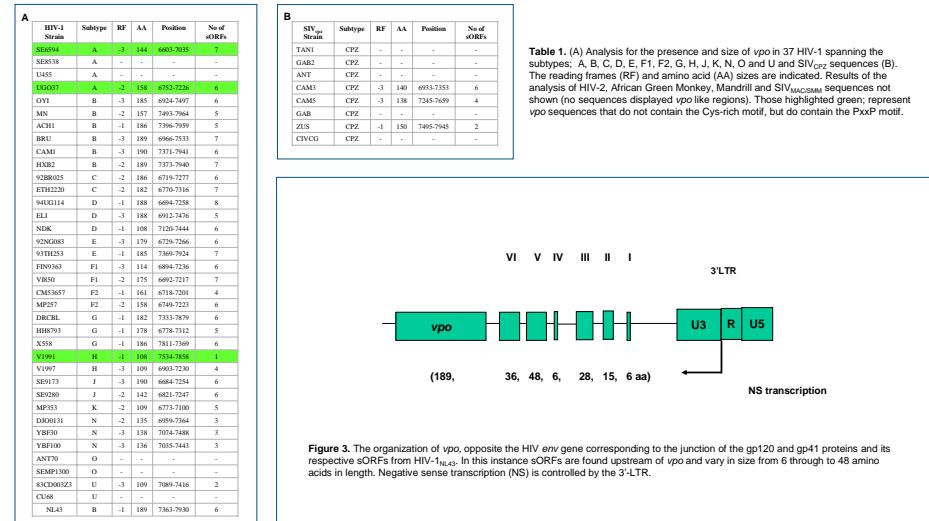


Table 1. (A) Analysis for the presence and size of vpo in 37 HIV-1 spanning the subtypes; A, B, C, D, E, F1, F2, G, H, J, K, N, O and U and SIV₂₂₂ sequences (B). The reading frames (RF) and amino acid (AA) sizes are indicated. Results of the analysis of HIV-2, African Green Monkey, Mandrill and SIV₂₂₂ sequences are not shown (no sequences displayed vpo like regions). Those highlighted green, represent vpo sequences that do not contain the Cys-rich motif, but do contain the PxxP motif.

HIV-1 Strain	Subtype	RF	AA	Position	No. of sORFs
SE0994	A	-	144	6602-7021	7
SE8358	A	-	-	-	-
U485	A	-	-	-	-
SE8097	B	-	189	6753-7226	8
OY1	B	-	185	6054-7497	4
MN	B	-	157	7493-7964	5
AC31	B	-	186	7396-7959	5
BRU1	B	-	189	6966-7533	7
CAM1	B	-	190	7371-7661	6
H2B3	B	-	189	7373-7640	7
SE88025	C	-	186	6719-7277	6
ETH220	C	-	182	6770-7316	7
945G114	D	-	188	6694-7258	8
EL1	D	-	188	6822-7476	5
NDK	D	-	108	7220-7444	6
92N0683	E	-	179	6726-7266	6
93T0251	E	-	185	7396-7979	7
FR7623	F1	-	186	6753-7226	8
SE815	F1	-	152	6895-7217	7
CMS1827	F2	-	181	6738-7200	7
MF257	F2	-	158	6748-7223	6
DRCHL	G	-	182	7333-7879	6
H8B76	G	-	178	6746-7322	6
X558	G	-	186	7281-7569	6
V1991	H	-	108	7512-7624	1
V1997	H	-	109	6903-7230	3
SE9173	J	-	190	6884-7254	6
SE9136	K	-	182	6822-7427	6
MF953	K	-	109	6773-7000	5
DX0013	N	-	135	6659-7364	3
Y8F30	N	-	138	7074-7488	3
Y8F100	N	-	136	7055-7443	3
AN767	O	-	-	-	-
SEMP100	O	-	-	-	-
KX00323	U	-	109	7089-7416	2
CV108	U	-	-	-	-
NE-43	B	-	189	7363-7930	6

SIV Strain	Subtype	RF	AA	Position	No. of sORFs
TANI	CPZ	-	-	-	-
GAB2	CPZ	-	-	-	-
ANF	CPZ	-	-	-	-
CAM1	CPZ	3	148	4913-7133	6
CAM2	CPZ	-3	138	7245-7639	4
GAB	CPZ	-	-	-	-
ZLS	CPZ	-1	150	7495-7945	2
CTVCG	CPZ	-	-	-	-