A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses

Shu-Yun Le*, Michael H.Malim[1], Bryan R.Cullen[1] and Jacob V.Maizel
Division of Cancer Biology and Diagnosis, Laboratory of Mathematical Biology, National Cancer Institute, NCI/FCRF, National Institutes of Health, Bldg. 469, Rm 151, Frederick, MD 21701 and [1]Howard Hughes Medical Institute and Departments of Medicine and Microbiology and Immunology, Duke University Medical Center, Durham, NC 27710, USA

Received September 18, 1989; Revised and Accepted February 11, 1990

ABSTRACT

A series of unusual folding regions (UFR) immediately 3′ to the cleavage site of the outer membrane protein (OMP) and transmembrane protein (TMP) were detected in the envelope gene RNA of the human immunodeficiency virus (HIV-1, HIV-2) and simian immunodeficiency virus (SIV) by an extensive Monte Carlo simulation. These RNA secondary structures were predicted to be both highly stable and statistically significant. In the calculation, twenty-five different sequence isolates of HIV-1, three isolates of HIV-2 and eight sequences of SIV were included. Although significant sequence divergence occurs in the *env* coding regions of these viruses, a distinct UFR of 234-nt is consistently located ten nucleotides 3′ to the cleavage site of the OMP/TMP in HIV-1, and a 216-nt UFR occurs forty-six and forty-nine nucleotides downstream from the OMP/TMP cleavage site of HIV-2 and SIV, respectively. Compensatory base changes in the helical stem regions of these conserved RNA secondary structures are identified. These results support the hypothesis that these special RNA folding regions are functionally important and suggest that the role of this sequence as the Rev response element (RRE) is mediated by secondary structure as well as primary RNA sequence.

INTRODUCTION

The lentiviruses are pathogenic retroviruses that induce chronic, degenerative diseases in their hosts. Human immunodeficiency virus (HIV-1, HIV-2) and simian immunodeficiency virus (SIV) are related lentiviruses, HIV-2 shares about 45% nucleotide sequence similarity with HIV-1 and 75% sequence similarity with SIV (macaque) (1). The genome of the virus family encodes not only the three virion proteins (Gag, Pol and Env) common to other known retroviruses, but also at least six other proteins required whose functions remain incompletely unknown (2). Two of these nonstructural regulatory proteins, Tat and Rev, are nuclear *trans*-activators essential for virus replication in culture (3–7).

It has been suggested that the Tat proteins may be directly involved in mediating the recognition of their respective viral *trans*-activation response (TAR) element (8). The TAR elements of HIV-1 and HIV-2 have been identified in the LTR R region, and coincide with a highly significant predicted RNA secondary structure (9–14). Also, we previously reported that the viral structural proteins are encoded by a population of incompletely spliced viral transcripts that are constituently expressed in the nucleus but excluded from the cytoplasm in the absence of Rev (15). This Rev response was shown to require a specific Rev response element (RRE) which coincides with a large, predicted RNA secondary structure present in the transmembrane protein (TMP) coding region just downstream from the boundary of the outer membrane protein (OMP) and TMP of Env in HIV-1 (15) and HIV-2 (16).

An interesting question raised by these experiments is whether the coincidence of the RRE with a predicted RNA secondary structure is significant. In this report, unusual folding regions (UFR) near the cleavage site of the OMP/TMP in HIV-1, HIV-2 and SIV RNA's were detected by a recently developed Monte Carlo simulation method (17–19) employing statistical scores described below.

Statistical analyses of the distributions of the scores in 1000 nt segments encompassing OMP/TMP junction region show that UFR's in this region are extremely stable and highly unusual relative to other regions of HIV-1, HIV-2 and SIV. Although significant sequence divergence occurs in *env* coding regions of these viruses, a distinct secondary structure of 234-nucleotide(nt) is consistently situated ten nucleotides downstream of the cleavage site of the OMP/TMP in HIV-1, and a 216-nt distinctive folding region forty-six and forty-nine nucleotides downstream of the OMP/TMP cleavage site of HIV-2 and SIV. Predictable

* To whom correspondence should be addressed at Biomolecular Modelling Section, Division of Biological Sciences, National Research Council of Canada, Bldg M-54, Rm 1109, Ottawa, Ontario K1A OR6, Canada

occurrence suggests a biological role for UFR's and extends our previous biochemical results showing that the Rev response requires a specific RRE that is associated with an RNA secondary structure near the boundary of the OMP and TMP in both HIV-1 (15) and HIV-2 (16). Moreover, the occurrence of compensatory base changes in the stem regions and conserved patterns in the hairpin loops of these HIV-1, HIV-2 and SIV sequences is established. The possible biological significance of each of these conserved features will be discussed.

METHODS

Significance (S_1) and stability (S_2) scores are defined as two standardized measures of the statistical significance and thermodynamic stability of local secondary structure in the sequence. As previously described (17, 18), the significance score of a RNA segment is defined as:

$$S_1 = (E - E_r) / SD_r$$

In the equation, E is the lowest free energy of the real biological sequence in the segment. E_r and SD_r are the mean and standard deviation of the lowest free energies from a random sample of large number of random permutations of the real sequence. The randomly shuffled sequences are generated by Monte Carlo simulation. Similarly, the stability score S_2 is defined as:

$$S_2 = (E - E_b) / SD_b$$

where E_b and SD_b are the mean and standard deviation of the sample which is constructed by sliding the same size window as the segment one base at a time from 5′ to 3′ (19). S_2 is a measure of local variation in predicted stability that additionally serves to provide a similar scale as S_1, facilitating comparison in combined plots of data. Lower score values represent places where predicted structures are either more stable than a random sequence of the same composition, S_1, or more stable than the average; over all segments of the natural sequence, S_2. Since we do not literally determine biological significance by these tests, we have adopted the accepted practice of referring to such features by the term 'unusual' folding regions (UFR).

The computer program SEGFOLD(19) is designed to calculate the significance and stability scores of the segment as a chosen-sized window steps one nucleotide at a step along the sequence in the 5′ to 3′ direction. The program is vectorized, and optimized to run on the Cray Operating System of a CRAY X-MP/24 supercomputer. In general, the following steps are often employed in finding UFR's. (i) To obtain an outline of distinct character of the statistical significance and thermodynamic stability of local secondary structures in the sequence, the distributions of the S_1 and S_2 of segments with selected sizes (e.g. 60, 100, 150 nucleotides etc.) are depicted along the sequence. If the sequence is long enough, the data are usually averaged and smoothed before plotting. Those UFR's where the lowest S_1 or S_2 occur are noted. (ii) The sizes of those UFR's are accurately assessed by an exhaustive Monte Carlo simulation (17) over window sizes ranging from 40 to 300 bases in increments of one or two nucleotides to the UFR identified in step (i). In the calculation, the global minima of S_1 and S_2 and the starting position of the segment are extracted for each window. The RNA secondary structure of the detected folding region is computed by the NEWFOLD program, which is a refinement of Zuker's FOLD program (20) using a version of the Turner energy rules with penalties of asymmetric loops and internal loops closed by AU pairs (23, 24). With sequences of known structure NEWFOLD gives better agreement than RNAFOLD.

All of the sequences used in the study come from the compilation and analysis of nucleic acid and amino acid sequences in Human Retroviruses and AIDS (21). The 25 isolates of HIV-1 in the study are HXB2, HXB3, BH10, BH8, BRU, PV22, PNL4-3, NY5, MAL, ELI, RF, CDC-451 (CDC), SF2, MN, SC, Brain (BRVA), Zaire-6 (Z6), Zaire-3 (Z3), Zaire HZ321 (HZ321), Z2, Z-84 (JY1), JH3, WMJ1, WMJ2, AND WMJ3. Three isolates ROD, HIV2FG (NIHZ) and SBLISY of HIV-2, six isolates MM142, MM251, BK28, K6W, K78 and PK190 of SIV (macaque), clones SMH3 and SMH4 (MMH4) of simian (sooty mangabey) immunodeficiency virus and isolate TYO-1 (TYO) of simian (African green monkey) immunodeficiency virus are also used in this study.

RESULTS

Detection of the UFR at the HIV-1 OMP/TMP junction

Here, we give an example for detecting and determining the size of the UFR with the present method using the UFR existing in the Env coding region of HIV-1 (isolates BH10 and ARV/SF2) at the boundary between OMP and TMP (12). In the Monte Carlo simulation the 500 nt OMP/TMP junction region (7659−8158) of HIV-1 (HXB2) is selected as a searching area, where the TMP start position is 7758. The relationships between the global minima of S_1 and S_2 and window size in each simulation are depicted in Fig.1(a). Starting positions of these UFR's identified in each simulation using various window sizes are delineated in Fig.1(b). The window size that corresponds to the deepest valley, or global minimum B, is 234 or 236 nt (see Fig.1(a)) with S_2 less than −3.2. We also detect a corresponding deep valley, A, with the same window size, with S_1 less than −5.0. Using that window size, the UFR is further located in Fig.1(b). In the case of 234-nt (or 236-nt) window the distribution of the lowest S_1 and S_2 coincide (Fig. 1(b)). Thus, the UFR (7768−8001) or (7767−8002) is both the most significant and most stable in the OMP/TMP junction domain (7659−8158). Figure 2 is a map of S_1 and S_2 over the folding domain (7768−8001) showing that it is also both most stable and most significant relative to others over a more extensive domain (7359−8358) extending from 400 nt 5′ and 600 nt 3′ of the cleavage site of OMP/TMP. This UFR coincides with a specific RNA target sequence which Rev response requires in the trans-activation of HIV-1 (BH10) (15).

Detection of the UFR's in other isolates of HIV-1, HIV-2 and SIV

Using the same procedure the UFR's near the cleavage site of OMP/TMP in other isolates of HIV-1, HIV-2 and SIV are also identified. The searching region for each isolate also includes 100 nt upstream and 400 nt downstream of the cleavage site. The detected distinctive RNA folding regions of these different isolates of HIV-1 are listed in Table 1, and those of HIV-2 and SIV are listed in Table 2. The thermodynamic stability and statistical significance of UFR's were examined again in the 1000-nt region that includes 400 nt upstream and 600 nt downstream of the cleavage site of OMP/TMP.

A continued statistical analysis of S_1 and S_2 distributions of stem-loop structures was made by averaging the stability and significance scores in each sequence position of the extended OMP/TMP region for these isolates of HIV-1, HIV-2 and SIV, respectively. The data were delineated in Fig.3(a)−(c). It is clear

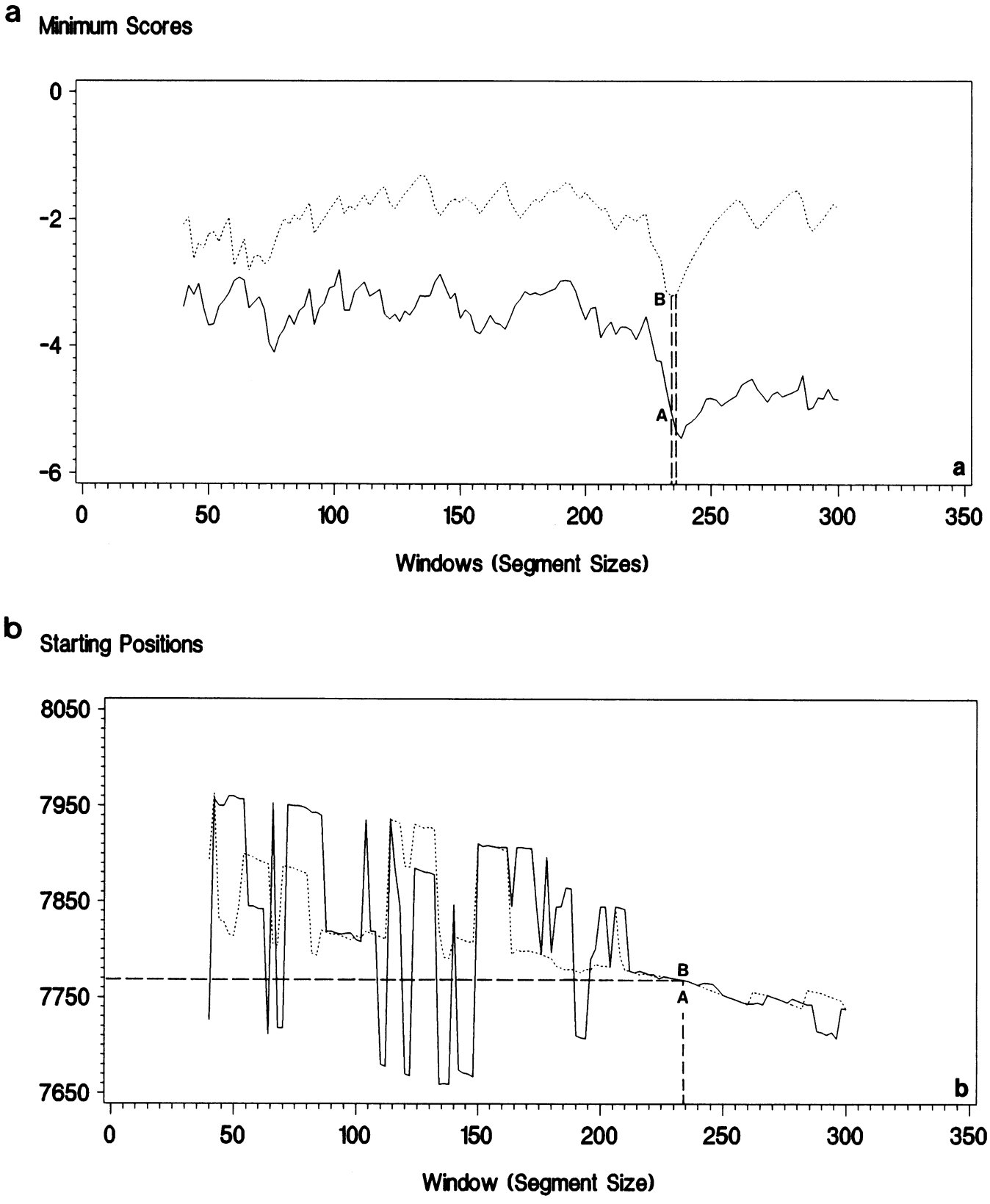


Figure 1. (a) Distributions of S_1, minimal significance score (solid line), and S_2, stability score (dotted line), in the OMP/TMP junction region (7659–8158) of HIV-1(HXB2) using different window sizes. The calculation was carried out repeatedly by increasing two-nt to the 3′ end of the segment window from 40 to 300-nt. For each window size, the simulation was carried out by sliding one-nt along the RNA sequence, and global minima of S_1 and S_2 were extracted and plotted against the window size. The deep trough B indicates that the global minima of S_2 is obtained when the window size is taken as 234 or 236-nt in the exhaustive simulation. (b) Distributions of S_1 (solid line) and S_2 (dotted line) with different window sizes. The map was obtained by plotting the starting positions of the UFR's against the window sizes. The starting position of TMP coding region is position 7758. When the window size is taken as 234 or 236-nt, the two distributions are coincident with each other. The chosen UFR is 7768–8001 (234-nt) or 7767–8002 (236-nt).

that a distinct stable RNA folding region (234-nt) consistently occurs 10 nucleotides downstream from the cleavage site of OMP/TMP of HIV-1, and a distinctive stable region (216-nt) occurs 46 and 49 nt 3′ to the cleavage site of OMP/TMP of HIV-2 and SIV, respectively. It should be mentioned that the sequences from isolates BH10, BH8, PV22, HXB3, which are virtually identical with the sequence of HXB2, and isolates WMJ1 and WMJ3, which are nearly identical with the sequence of WMJ2 (21), are not included in the sample of the statistical analysis of HIV-1 to avoid excessive bias from over represented sequences. Similarly, the sample of different isolates of SIV contains the four sequences of isolates MM142, MM251, TYO-1 and clones SMH3 and SMH4 of sooty mangabey (the sequences from isolates BK28, K6W, K78 and PK190 are not included).

Highly Conserved RNA Secondary Structures

The predicted stable secondary structure of the UFR (7768−8001) of HIV-1 (HXB2) is shown in Fig. 4. The structural model indicates that the complex RNA secondary structure consists of a long stem-loop A and a short stem B, as well as five hairpin structures B1a, B1b, C, D and E. Comparative sequence analyses reveal that these structural features are conserved in all of the 25 sequences of HIV-1. Figure 5 demonstrates the positions of each of these stems as conserved in all of these sequences. For example, the long stem-loop structure A has been represented by 14 boxes marked by A1-A7 and A1′-A7′ where the first base of the box (each of A1-A7) is base paired with the last base of the correspondence box (one of A1′-A7′). Compensatory base changes that occur in each stem and base mutations in the loops are displayed with annotations in Figure 5. Most compensatory base changes in the stems are base transitions from A-U to G-U or G-C to G-U. A few base transitions from A-U to G-C can be found out in isolates PNL4-3, MAL, Z3, JY1, and Z321. No base transversions are detected in the conserved stems. Moreover, the conserved pattern UUA, AACAAU, GCAU, and AUYAARCA are identified in the hairpins B1b, C1, D2 and E1.

The distinct structures of the special RNA folding regions of HIV-2 (ROD), SIV (K78) and SIV (TYO-1) are shown in Fig. 6(a)−(c). Their compensatory base changes and nucleotide mutations are all shown in the figures. Again, most of the compensatory base changes in the conserved helices are base transitions of A-U or G-C to G-U. Only one base transversion (from G-C to C-G) in a conserved stem was found in the different sequence isolates of SIV (Fig 6b, K6W). These structures show similar structural features to those of HIV-1, that is a long stem-loop, a short stem and five hairpins. Obviously, the structures derived from different isolates of HIV-2 are virtually identical to those derived from different isolates of SIV (macaque and sooty mangabey). The conserved patterns UURYU, AACA, AGAA and AACNAARAAY are found in the hairpin loops of derived from 10 different sequences of HIV-2 and SIV species. The RNA secondary structure of SIV (TYO-1) of African green monkey, however has some different structural features from those of HIV-2 and SIV of macaque and sooty mangabey (Fig. 6(c)).

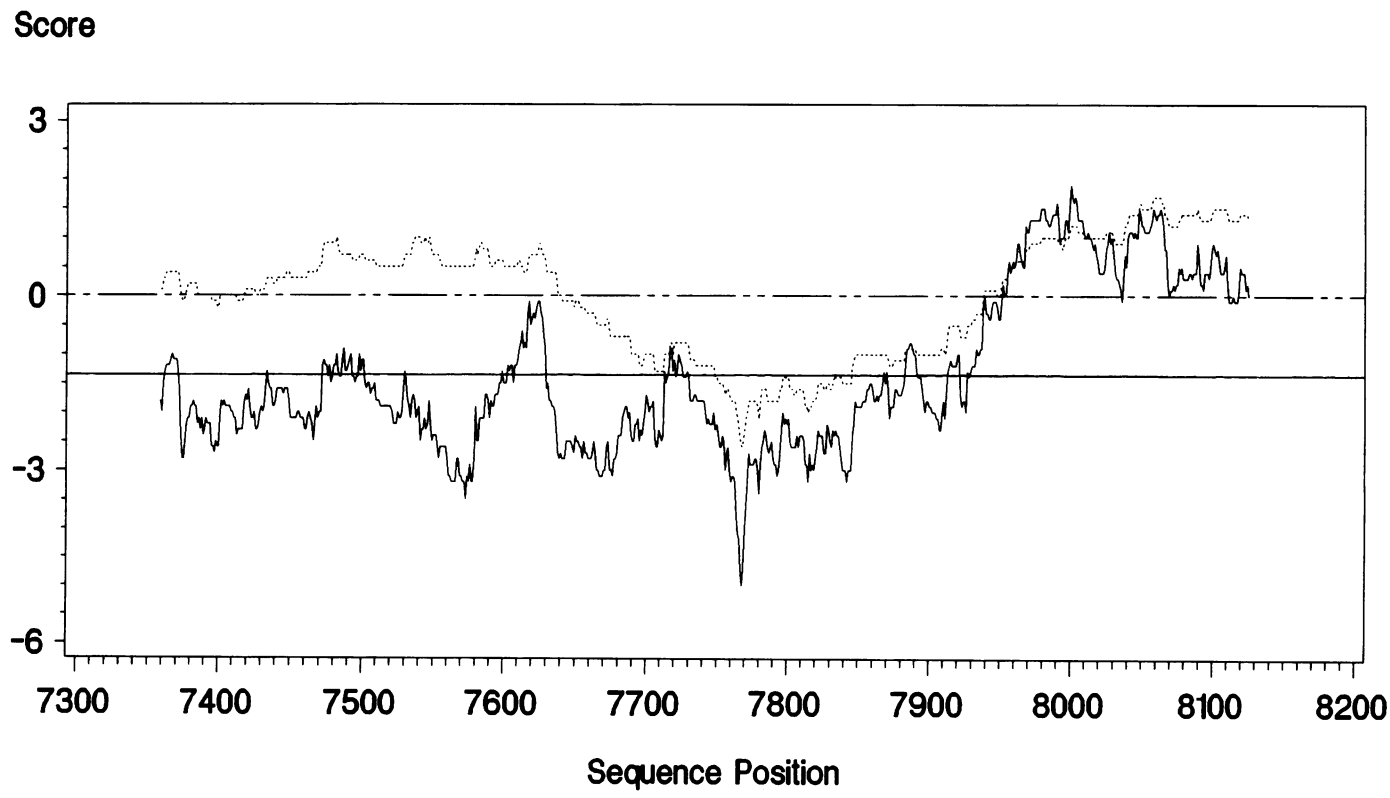


Figure 2. Distributions of S_1, (solid line), and S_2, (dotted line), of stem-loop structures in the extended OMP/TMP junction domain (7359−8358). The map was made by plotting the two scores against the starting position of each segment along the sequence. The window size is chosen as 234-nt. The same sequence position of global minima of two curves is at position 7768. A horizontal solid and dash line represent the average S_1 and S_2 scores in the junction region.

DISCUSSION

We described in a previous report (25) that HIV-1 envelope region sequence varies in independent isolates, and even in sequential viral isolates from the same patient, by alignments of both the envelope nucleotide and amino acid sequences position of 17 isolates (25). The three significant clusters F, G and H of the hypervariable site in envelope RNAs of 17 HIV isolates (25) are in the extended OMP/TMP junction region (1000-nt) used in this study. Although sequence divergence among different isolates occurs in this domain from a distinct RNA secondary structure of 234-nt consistently exists at ten nt downstream from the cleavage site of the OMP/TMP. There is clearly significantly greater sequence conservation in the UFR's than in the aligned total envelope coding regions of these 25 isolates. In fact, the unusual RNA folding regions include three significant clusters

Table 1. Unusual Folding Regions Detected in the TMP of HIV-1

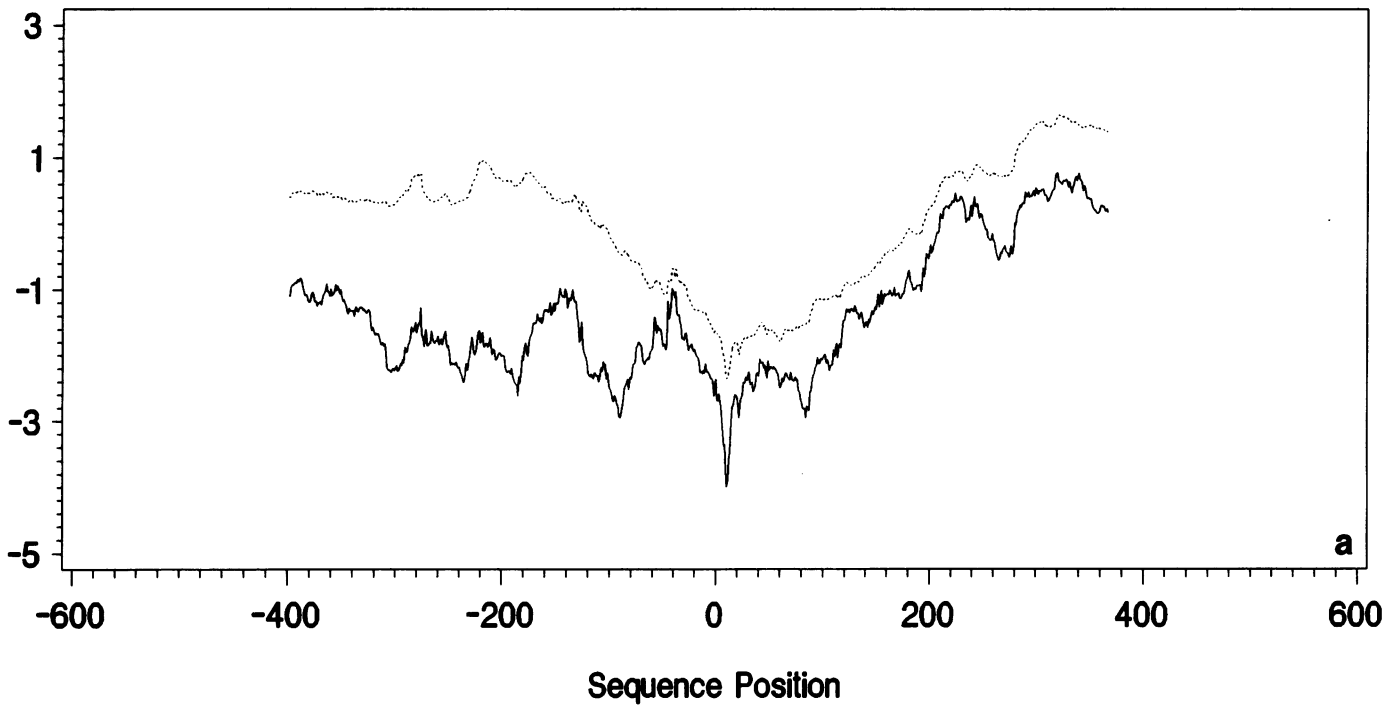
Isolate in HIV-1	Region in the Sequence	Distance from TMP Starting	S_1	S_2	Energy (kcal/mol)
BRU	7362-7595	10	-4.1	-2.5	-107.3
BRV	1758-1991	10	-4.7	-3.0	-111.1
CDC	2060-2293	13	-4.4	-2.2	-104.8
ELI	7305-7538	10	-4.5	-2.8	-106.5
HXB2	7768-8001	10	-5.0	-2.6	-111.8
JH3	1827-2060	10	-4.1	-2.6	-106.3
JY1	1598-1831	10	-3.8	-2.5	-102.9
MAL	7349-7582	10	-4.5	-2.9	-114.0
MN	7783-8016	7	-4.8	-2.7	-113.3
NY5	1586-1819	13	-5.4	-2.4	-114.5
PNL4-3	7759-7992	10	-4.7	-2.6	-108.4
RF	7317-7550	13	-4.5	-2.6	-109.0
SC	2482-2715	13	-4.1	-2.6	-107.8
SF2	7775-8008	11	-3.5	-1.9	-96.0
WMJ2	1545-1778	13	-5.2	-2.6	-115.0
Z2	7762-7995	10	-3.7	-2.5	-102.6
HZ321	1641-1874	10	-3.6	-2.2	-100.4
Z3	1588-1821	10	-4.3	-2.1	-112.3
Z6	3117-3350	10	-4.3	-2.6	-106.5

Table 2. Unusual Folding Regions Detected in the TMP of HIV-2 and SIV

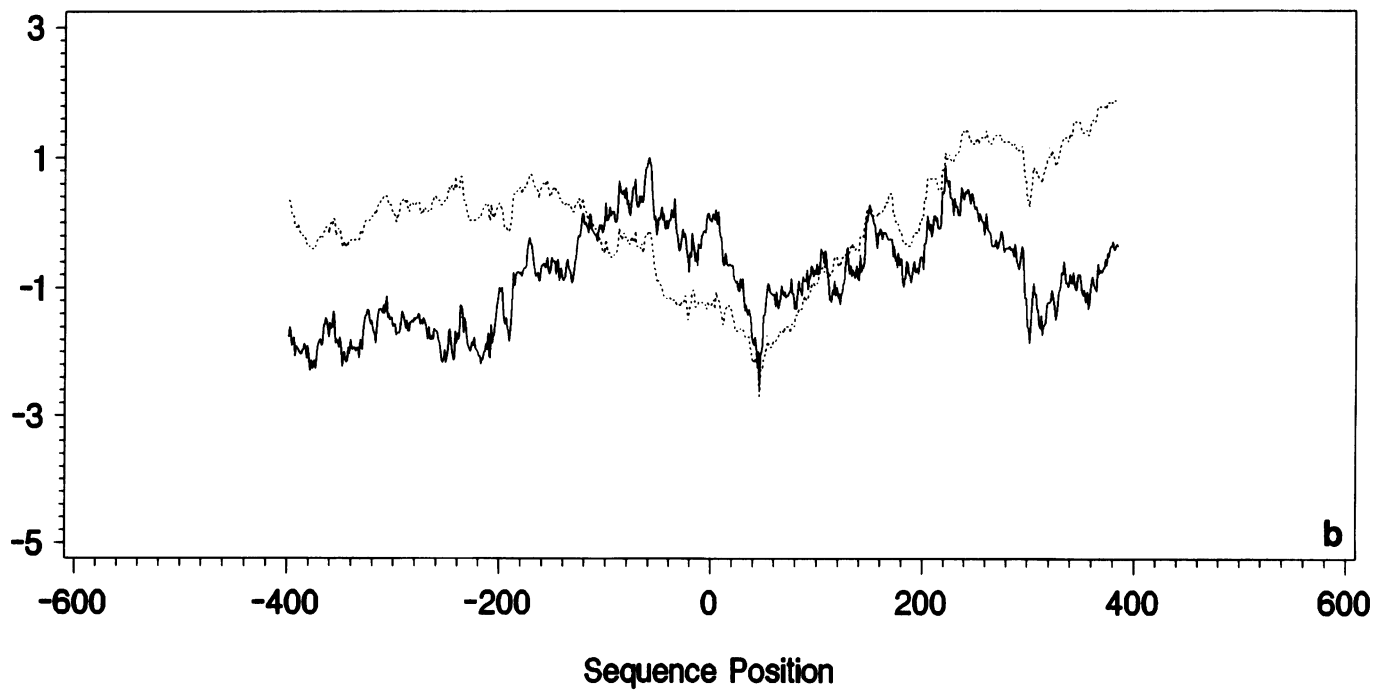
Isolate in HIV-2 and SIV	Region in the Sequence	Distance from TMP Starting	S_1	S_2	Energy (kcal/mol)
HIV2FG	7661-7876	46	-3.0	-2.8	-93.9
ROD	7696-7911	46	-2.5	-2.5	-101.1
SBLISY	7670-7885	46	-2.3	-2.8	-93.5
MM142	7690-7905	49	-2.9	-2.8	-99.6
MM251	8177-8392	49	-3.1	-2.8	-100.8
MMH4 (Sooty)	8160-8375	49	-2.3	-2.1	-94.5
TYO (Green Monkey)	7394-7609	46	-4.9	-2.3	-116.6

Table 1 and 2. The S_1 (significance) and S_2 (stability) scores and the lowest free energy of stem-loop structures are derived using the SEGFOLD program (19). In the calculation the Tinoco energy rules (26, 27) are used.

a


Scores
3
1
-1
-3
-5
-600
-400
-200
0
200
400
600
Sequence Position
a

b


Scores
3
1
-1
-3
-5
-600
-400
-200
0
200
400
600
Sequence Position
b

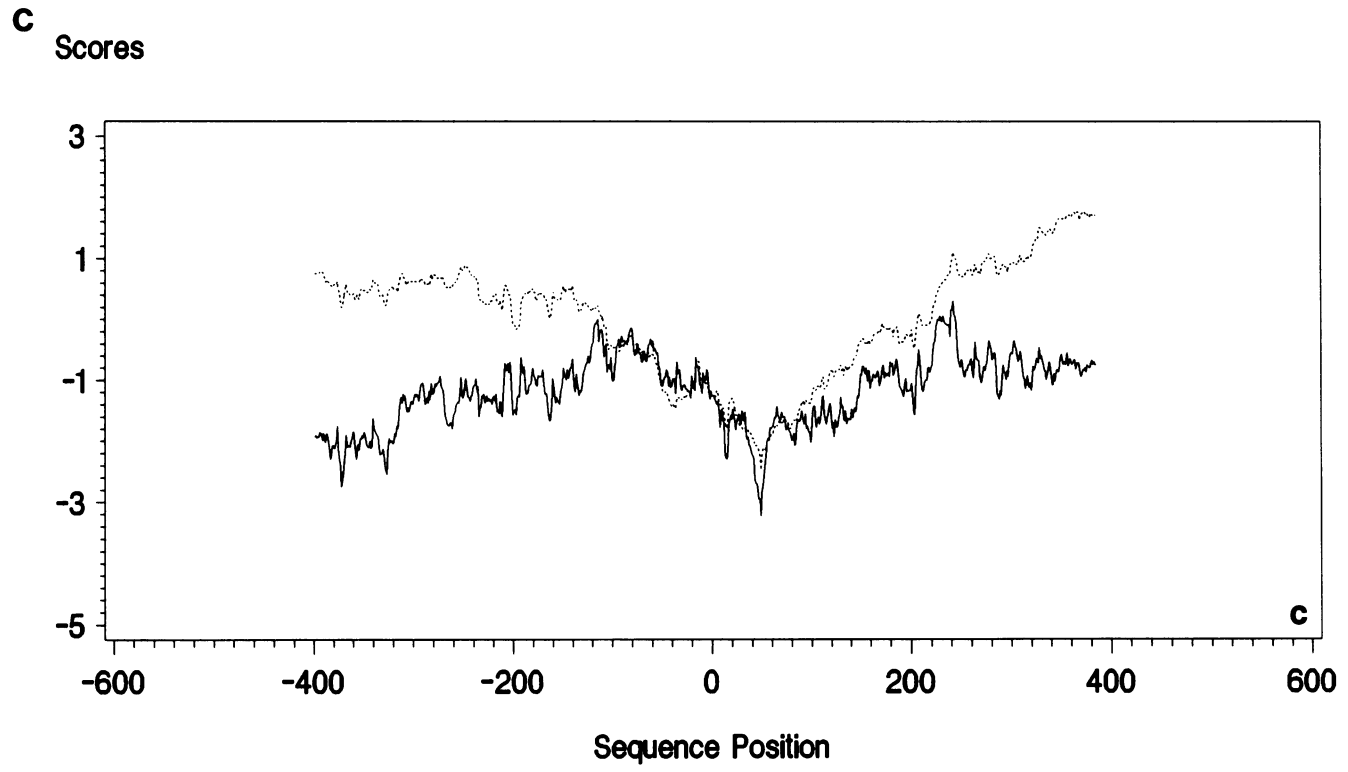


Figure 3. Combined maps of the distributions of S_1, and S_2, of stem-loop structures in the extended OMP/TMP region (1000-nt) of envelope genes of primate immunodeficiency viruses. (a) 19 isolate sequences of HIV-1. (b) 3 isolate sequences of HIV-2. (c) 4 isolate sequences of SIV. In the maps, the distributions of S_1 (solid line) and S_2 (dotted line) are made by averaging their scores in each position. The window size is taken as 234-nt in (a), 216-nt in (b) and (c), and the corresponding TMP start site is taken as position zero.

C8, C9 and C10 of the conserved sites detected in envelope coding regions of 17 HIV isolates (25).

The predicted structure of the UFR is also highly conserved in all isolates of HIV-1. Compensatory base changes in homologous stems of these related sequences allow the conservation of the secondary structure despite some sequence divergence. This suggests that the common UFR detected in all isolates of HIV-1 may play an important role in the life cycle of HIV-1. We have reported that the RRE must be at least 104 and possibly more than 210 nt in length by mutational analysis experiment (15). The minimal size of the RRE is fully contained in the UFR's described here.

In this manuscript, we also demonstrate that a 216-nt conserved distinct RNA folding region exists near to the cleavage site of OMP/TMP in all three isolates of HIV-2. Similar to the case of HIV-1, the HIV-2 envelope coding region shows more variability than other regions of the genome. The complete *env* gene displays about 85% nucleotide sequence identity among these three isolates (ROD, HIV2FG, and SBLISY), but there is about 92% sequence similarity among the UFR's of these three isolates. The existing genetic variation is 65% in the loops and compensatory base changes of conserved helical stems. The high conservation of HIV-2 UFR's further suggests that they have biological roles. Recently, we demonstrated that HIV-2 and SIV (macaque) also encode functional Rev *trans*-activators that can induce the cytoplasmic expression of the unspliced viral transcripts encoding the viral structural proteins (16). Mutational analyses (16) show that the RNA target sequence specificity of the HIV-2 Rev response is closely related by both primary sequence and predicted secondary structure to the RRE just downstream of the cleavage site of OMP/TMP in the *env* gene HIV-1. The UFR's detected from all isolates of HIV-2 in this

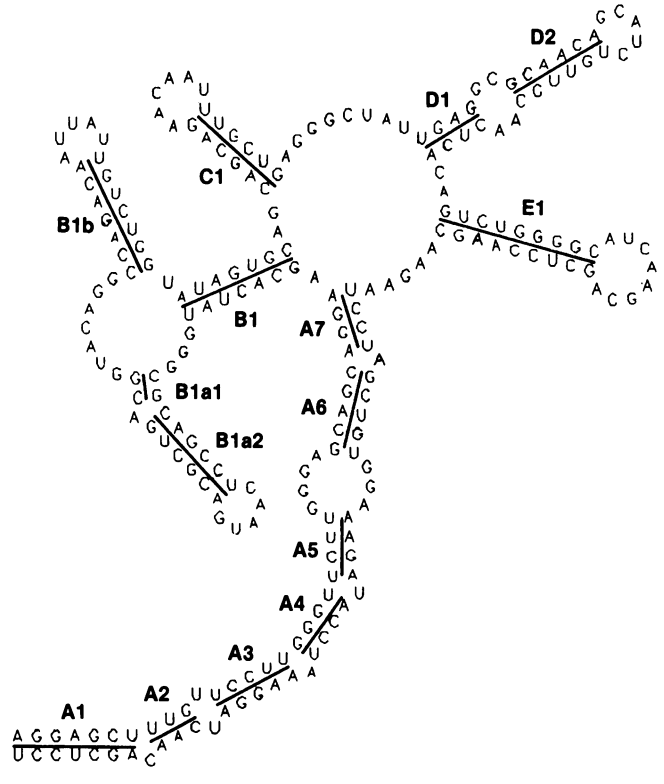
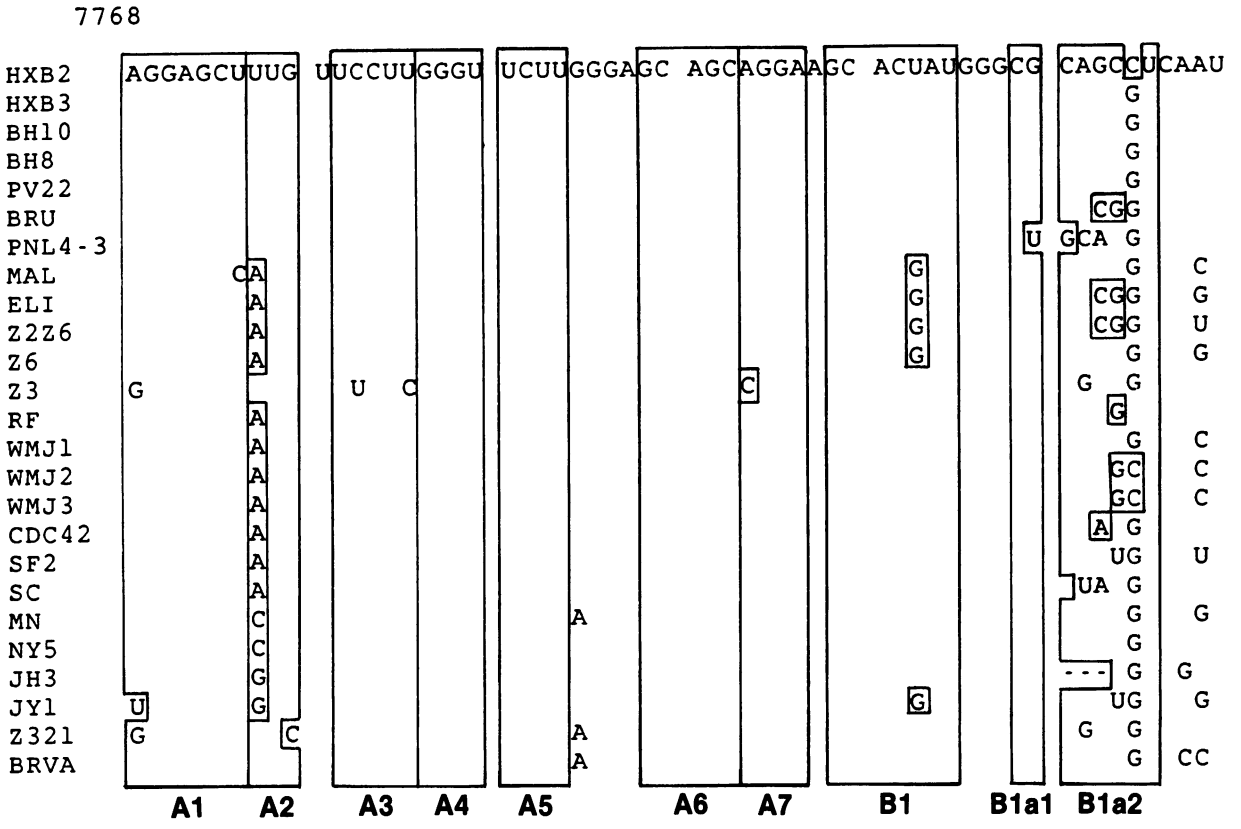
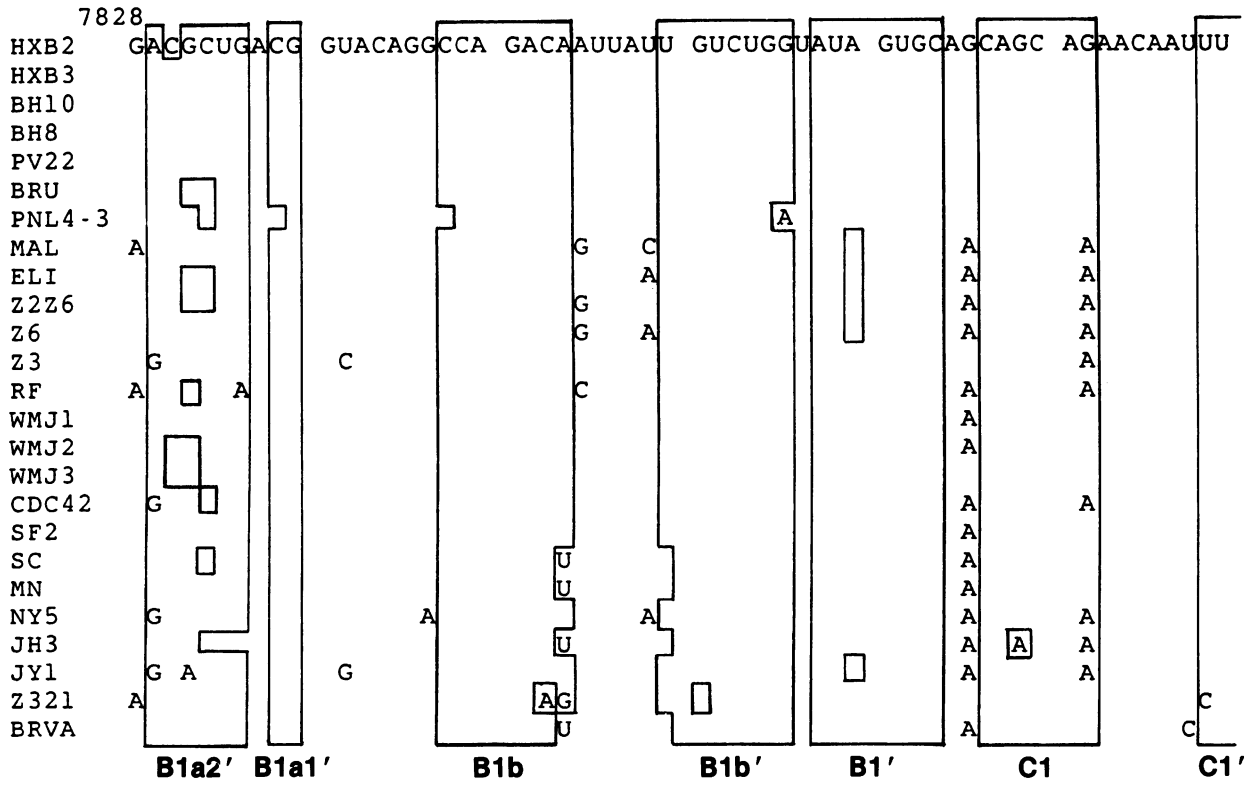


Figure 4. Predicted secondary structure of the unusual RNA folding region detected in HXB2 isolate of HIV-1. The structure is computed using the Turner energy rules (see text).


7768
HXB2 AGGAGCUUUG UUCCUUGGGU UCUUGGGAGC AGCAGGAAGC ACUAUGGGCG CAGCCUCAAU
HXB3
BH10
BH8
PV22
BRU
PNL4-3
MAL
ELI
Z2Z6
Z6
Z3
RF
WMJ1
WMJ2
WMJ3
CDC42
SF2
SC
MN
NY5
JH3
JY1
Z321
BRVA
A1
A2
A3
A4
A5
A6
A7
B1
B1a1
B1a2


7828
HXB2 GACGCUGACG GUACAGGCCA GACAAUUAUU GUCUGGUAUA GUGCAGCAGC AGAACAAUUU
HXB3
BH10
BH8
PV22
BRU
PNL4-3
MAL
ELI
Z2Z6
Z6
Z3
RF
WMJ1
WMJ2
WMJ3
CDC42
SF2
SC
MN
NY5
JH3
JY1
Z321
BRVA
B1a2'
B1a1'
B1b
B1b'
B1'
C1
C1'

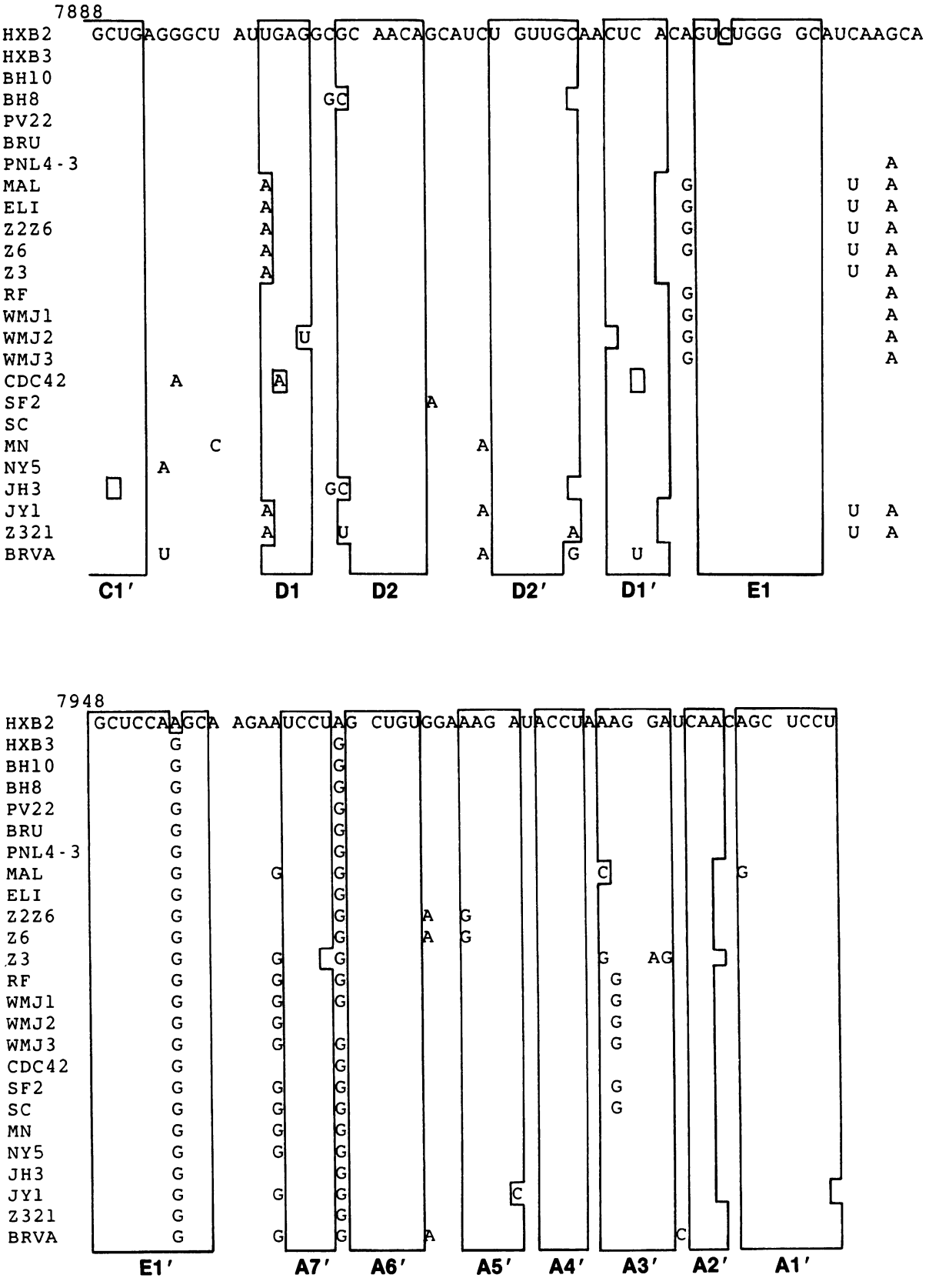


Figure 5. Alignment of the sequences of UFR's from 25 different isolates of HIV-1. Deletions are marked by ' − ' and a blank space indicates no change from the nucleotide in the isolate HXB2 sequence. The marked number is according to the sequence number of the isolate HXB2. Stem regions in the RNA secondary structure model shown in Fig. 4 are marked by boxes and labeled by letter A−E and A′−E'. Any bases within a box that are enclosed by another box fail to form an appropriate base pair.

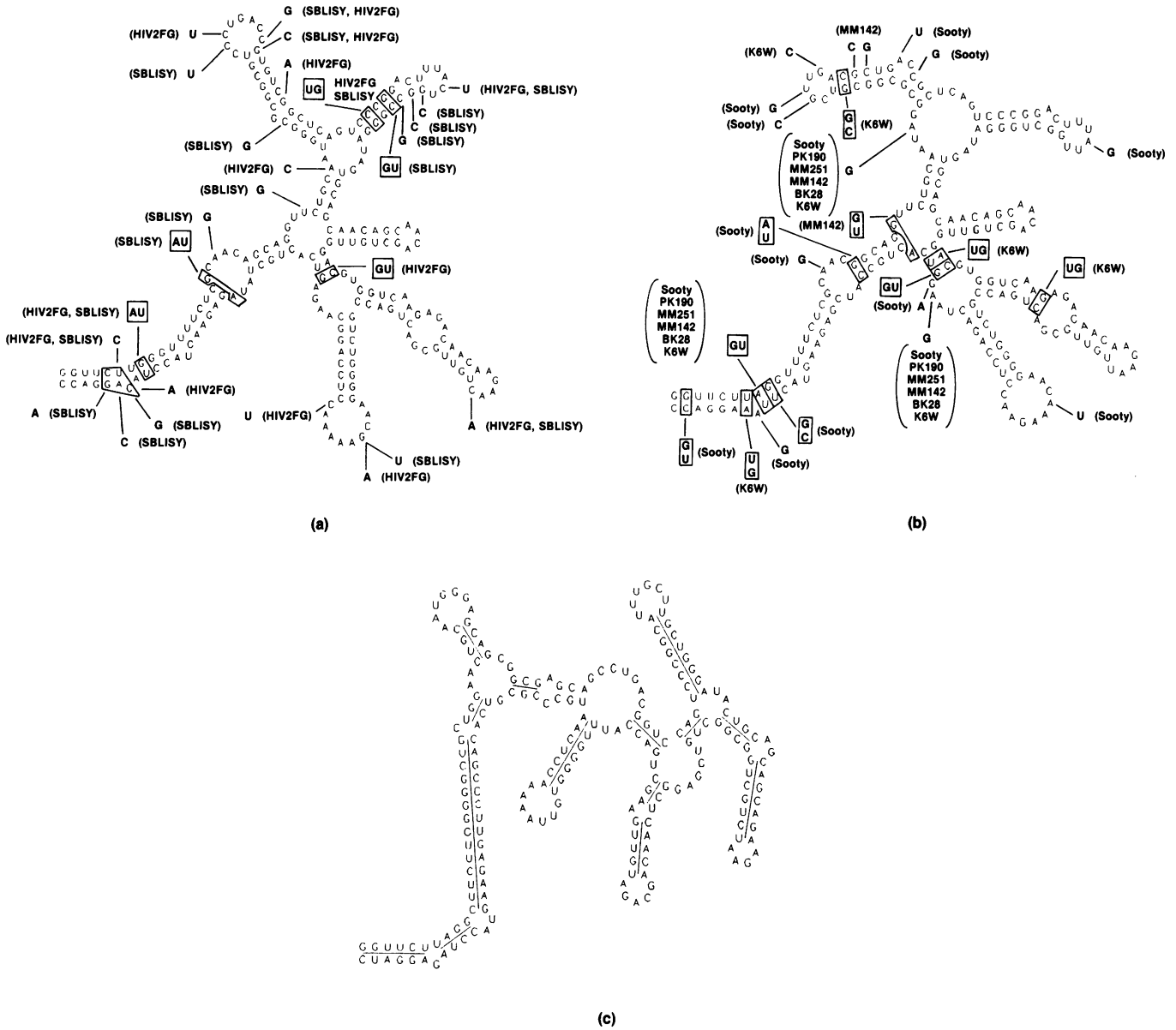


Figure 6. Predicted secondary structures of the UFR's detected in HIV-2 and SIV. The structure is computed using the Turner energy rules in program NEWFOLD. (a) Isolate ROD of HIV-2. Compensatory base changes and nucleotide mutations in the isolates HIV2FG and SBLISY are marked by the box. (b) Isolate K78 of SIV (macaque). The compensatory base changes and nucleotide mutations in the isolates of SIV (macaque and sooty mangabey) are marked by the box. (c) Isolate TYO-1 of SIV (African green monkey).

study agree with these experimental results.

The striking similarity in the complexity, overall appearance and location of conserved structures of HIV-2 and SIV (macaque and sooty mangabey), along with 97% average sequence identity among UFR's of seven isolates of SIV (macaque and sooty mangabey) and about 90% average identity among 10 sequences of SIV and HIV-2, contrasts with only 75% average overall sequence identity between HIV-2 and SIV (macaque), and argues that these UFR's play a common biological function in both SIV and HIV-2.

The UFR detected in SIV TYO-1 (from African green monkey) shows considerable greater variability in its sequence than other isolates of the macaque and sooty mangabey. It is only 64% similar to other. Nevertheless, the UFR predicted in the appropriate region shares a certain degree of similarity with those predicted from HIV-1, HIV-2 and SIV (macaque and sooty mangabey). Recently, we have reported the incomplete ability of the Rev proteins of HIV-1, HIV-2 and SIV (macaque) to cross-complement the expression of unspliced viral mRNA (16). The sequence divergence noted here suggests that a similar non-reciprocal complementation might be observed between the SIV (macaque) and SIV (African green monkey) systems.

ACKNOWLEDGEMENT

We thank Dr. Jih-H. Chen for the contribution of the NEWFOLD program.

REFERENCES

1. Smith, T.F., Srinivasan, A., Schochetman, G., Marcus, M. and Myers, G. (1988) Nature 333, 573–575.
2. Varmus, H. (1988) Genes and Development 2, 1055–1062.
3. Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C. and Haseltine, W.A. (1986) Cell 44, 941–947.
4. Fisher, A.G., Feinberg, M.B., Josephs, S.F., Harper, M.E., Marselle, L.M., Reyes, G., Gonda, M.A., Aldovini, A., Debouk, C., Gallo, R.C. and Wong-Stall, F. (1986) Nature 320, 367–371.
5. Terwilliger, E., Burghoff, R., Sia, R., Sodroski, J., Haseltine, W. and Rosen, C. (1988) Nature 320, 367–371.
6. Sadaie, M.R., Benter, T. and Wong-Staal, F. (1988) Science 239, 910–914.
7. Viglianti, G.A. and Mullins, J.I. (1988) J. Virol. 62, 4523–4532.
8. Emerman, M., Guyader, M., Montagnier, L., Baltimore, D. and Muesing, M.A. (1987) EMBO J. 6, 3755–3760.
9. Feng, S. and Holland, E.C. (1988) Nature 334, 165–167.
10. Hauber, J. and Cullen, B.R. (1988) J. Virol. 62, 673–679.
11. Jakobovits, A., Smith, D.H., Jakobovits, E.B. and Capon, D.J. Mol. Cell. Biol. 8, 2555–2561.
12. Le, S.-Y., Chen, Jih-H., Braun, M.J., Gonda, M.A. and Maizel, J.V. (1988) Nucleic Acids Res. 16, 5153–5168.
13. Muesing, M.A., Smith, D.H. and Capon, D.J. (1987) Cell 48, 691–701.
14. Fenrick, R., Malim, M.H., Hauber, J., Le, S.-Y., Maizel, J.V. and Cullen, B.R. (1989) J. Virol., in press.
15. Malim, M.H., Hauber, J., Le, S.-Y, Maizel, J. and Cullen, B.R. (1989) Nature 338, 254–257.
16. Malim, M.H., Bohnlein, S., Fenrick, R., Le, S.-Y, Maizel, J.V. and Cullen, B.R. (1989) Proc. Natl. Acad. Sci. USA, in press.
17. Le, S.-Y. and Maizel, J.V. (1989) J. Theor. Biol. 138, 495–510.
18. Le, S.-Y., Chen, Jih-H., Currey, K.M. and Maizel, J.V. (1988) CABIOS 4, 153–159.
19. Le, S.-Y., Chen, J.-H. and Maizel, J.V. (1989) Proceedings of the 6th Conversation in Biomolecular Stereodynamics, in press.
20. Zuker, M. and Stiegler, P. (1981) Nucleic Acids Res. 9, 133–148.
21. Myers, G., Rabson, A.B., Josephs, S.F., Smith, T.F. and Wong-Stall, F. (1989) A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. In Human Retroviruses and AIDS. Los Alamos, New Mexico, USA.
22. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson T. and Turner, D.H. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 9373–9377.
23. Turner, D.H., Sugimoto, N., Jaeger, J.A., Longfellow, C.E., Freier, S.M. and Kierzek, R. (1987) Cold Spring Harbor Symp. Quant. Biol. 52, 123–133.
24. Papanicolaou, C., Gouy, M. and Ninio, J. (1984) Nucleic Acids Res. 12, 31–44.
25. Le, S.-Y., Chen, Jih-H., Chatterjee, D. and Maizel, J.V. (1989) Nucleic Acids Res. 17, 3275–3288.
26. Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 985–1002.
27. Cech, T.R., Tanner, N.K., Tinoco, I.Jr., Weir, B.R., Zuker, M. and Perlman, P.S. (1983) Proc. Natl. Acad. Sci. USA 80, 3903–3907.