The HIV-2 Rev-response element: determining secondary structure and defining folding intermediates

Sabrina Lusvarghi[1], Joanna Sztuba-Solinska[1], Katarzyna J. Purzycka[1], Gary T. Pauly[2], Jason W. Rausch[1],* and Stuart F. J. Le Grice[1],*

[1]HIV Drug Resistance Program, Reverse Transcriptase Biochemistry Section, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA and [2]Chemical Biology Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

ABSTRACT

Interaction between the viral protein Rev and the RNA motifs known as Rev response elements (RREs) is required for transport of unspliced and partially spliced human immunodeficiency virus (HIV)-1 and HIV-2 RNAs from the nucleus to the cytoplasm during the later stages of virus replication. A more detailed understanding of these nucleoprotein complexes and the host factors with which they interact should accelerate the development of new antiviral drugs targeting *cis*-acting RNA regulatory signals. In this communication, the secondary structures of the HIV-2 RRE and two RNA folding precursors have been identified using the SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) chemical probing methodology together with a novel mathematical approach for determining the secondary structures of RNA conformers present in a mixture. A complementary chemical probing technique was also used to support these secondary structure models, to confirm that the RRE2 RNA undergoes a folding transition and to obtain information about the relative positioning of RRE2 substructures in three dimensions. Our analysis collectively suggests that the HIV-2 RRE undergoes two conformational transitions before assuming the energetically most favorable conformer. The 3D models for the HIV-2 RRE and folding intermediates are also presented, wherein the Rev-binding stem–loops (IIB and I) are located coaxially in the former, which is in agreement with previous models for HIV-1 Rev-RRE binding.

INTRODUCTION

RNA processing is an important facet of human immunodeficiency virus (HIV) replication. During early infection, HIV RNA transcribed from the integrated provirus is multiply spliced, exported from the nucleus via the Nxf1 export pathway and translated into viral accessory proteins, such as Nef, Tat and Rev (1,2). The Rev protein thus produced re-enters the nucleus where it binds the viral Rev response element (RRE), a conserved structural element embedded within the *env* gene in singly or unspliced HIV RNA (3–8). The resulting nucleoprotein complex is diverted from the cellular splicing machinery toward nuclear export via the Exportin-1 pathway (9). This process is critical for HIV replication, as singly spliced RNAs are translated into Env and/or select accessory proteins, whereas unspliced RNAs are translated into gag or gag-pol polyproteins, or are packaged into nascent virions (2,10). Rev-mediated RNA export is essential for both HIV-1 and HIV-2 infection, although the process has been studied to a much greater extent with the former virus (5,6,10,11).

Although the secondary structure of the HIV-1 RRE (RRE1) has been investigated by computational analysis and both chemical and enzymatic probing (12–16), a consensus structure has proven elusive. In the original model (12), RRE1 is comprised of an extended stem (SL I) capped by a five-way junction that joins it to four

smaller stem– or stem–loops (SL II-V). SL II is bifurcated, creating a short stem (SL IIA) at the base of stem loops SL IIB and SL IIC. An alternative structure determined by enzymatic probing is similar to the original in many respects, including the lengths of the stem loops and base pairing patterns within SL I, SL II and SL V. In this model, however, SL III and IV constitute a single, extended stem loop designated SL III/IV, which also creates a central four-way, rather than a five-way, junction (13,14). Differences between the two models may reflect variable probing and/or computational methodologies, or variations in SL I length among RRE1 constructs. Alternatively, it has been suggested that RRE1 assumes both the original and alternative form during the course of viral replication, and that the distribution of these forms among HIV-1 RNAs provides an additional level of regulatory control of RRE function. Although this is an intriguing possibility, a study of how mutations in RRE1 affect virus replication suggests that the length of SL I does not affect the structural arrangement of the other stem loops, nor is it conclusive regarding which form(s) is/are present *in vivo* (15). To further confound the issue, both models have been supported in independent studies using the SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) method of RNA probing (16,17).

To mediate nuclear export, HIV-1 Rev (Rev1) combines three functional domains: (i) an arginine-rich motif required for RNA binding, (ii) an oligomerization domain, the two faces of which facilitate Rev–Rev interactions and (iii) a disordered effector domain containing a nuclear export signal recognized by Crm-1, a component of the Exportin-1 nuclear export pathway (18–20). Current models indicate that Rev1 initially binds RRE1 at high-affinity binding sites at the base of SL IIB and within the purine-rich internal loop of SL I nearest the central junction (7,21–23). Additional Rev1 molecules then bind sequentially and cooperatively, presumably via a series of Rev–Rev and Rev–RNA interactions (19,20,24–26). RNA protection analysis suggests that Rev assembles along SL IIB and SL I, and that other RRE stem loops are not directly involved in Rev binding (12,13).

In contrast, comparatively little is known about HIV-2 Rev (Rev2) and the interaction with its homologous RRE (RRE2). Although predicted computationally, the secondary structure of RRE2 has not been probed by chemical or enzymatic methods. In comparative functional studies, it was demonstrated that although Rev1 facilitates nuclear export of RNAs containing either RRE1 or RRE2, Rev2 is non-functional in an HIV-1 context (5,6,11). Incomplete reciprocity between Rev1/RRE1 and Rev2/RRE2 systems may be explained by failure of Rev2 to adequately oligomerize along RRE1 RNA. It has been shown, however, that Rev2 is capable of inhibiting nuclear export in a Rev1/RRE1 system when provided *in trans*, suggesting that the HIV-2 protein is capable of interacting to some extent with Rev1, RRE1 or both (11). Finally, mutational analysis suggests that both Rev1 and Rev2 recognize their respective RRE counterparts on the basis of RNA secondary structure rather than primary nucleotide sequence (5,6).

In this work, complementary chemical and biochemical techniques have been used to determine the structure of the HIV-2 RRE. SHAPE analysis indicates that the low-energy form of RRE2 contains five peripheral stems/stem loops arranged in a manner similar to that originally reported for RRE1. Intriguingly, non-denaturing fractionation of RRE2 samples revealed two additional conformers that convert to the more stable low-energy form in a time-dependent fashion. This transition was verified by kinetic analysis, and the structures of the three forms were determined using a novel method for mathematically dissecting ensemble SHAPE profiles. Site-directed hydroxyl radical footprinting was used to confirm these structures, as well as examine the relative positioning of select motifs in 3D space. Finally, using RNAComposer modeling software (27), we have constructed 3D structural models of the open, intermediate and closed RRE2 conformers. Implications of these findings on the overall mechanism(s) of Rev–RRE interactions in both viruses are discussed.

MATERIALS AND METHODS

RNA preparation

RRE2 and RRE1, 5′-untranslated region (UTR)+1 and +15 transcription templates were polymerase chain reaction amplified from plasmids containing HIV-2$_{\text{ROD}}$ and HIV-1$_{\text{NL4-3}}$ and genomic DNA, respectively. Forward primers introduced a T7 promoter and, in the cases of RRE2 and RRE1, reverse primers introduced a 3′ non-viral structure cassette, as previously reported (28). No structure cassette was introduced into the 5′-UTR+1 and +15 constructs. Primer sequences are provided in Supplementary Data. Polymerase chain reaction products were precipitated, re-suspended in TE buffer [10 mM Tris–HCl (pH 7.4) and 0.1 mM ethylenediaminetetraacetic acid (EDTA)] and used as template for *in vitro* transcription reactions (MEGAshortscript T7 Kit, Life Technologies). *In vitro* synthesized RNAs were purified by denaturing gel electrophoresis (5% polyacrylamide and 7 M urea), electroeluted, precipitated, re-suspended in TE buffer and stored at −20°C. The 1M7 (1-methyl-7-nitroisatoic anhydride) was synthesized as previously described (29).

Renaturation and 1M7 treatment of RNAs

RNA (40 pmol) was diluted to a volume of 36 μl in water, heated to 95°C for 3 min and then immediately placed on ice. Once cool, the sample volume was increased to 40 μl by adding 4 μl of 10× RNA folding buffer [1× RNA folding buffer: 50 mM Tris–HCl (pH 8.0), 100 mM NaCl and 5 mM MgCl$_2$]. Samples were then divided into 20 μl experimental (+) and control (−) aliquots and incubated at 37°C for 5, 10, 20, 40 and 100 min, at which point 1 μl of 50 mM 1M7 [in anhydrous DMSO (+)] or DMSO alone (−) was added, and incubation at 37°C was continued for 5 min. Completed reactions were precipitated and re-suspended in 20 μl of TE buffer.

Primer extension

For detection of 2′-*O*-adducts, 10 pmol of (+) and (−) RNAs was mixed, respectively, with equimolar amounts of Cy5- or Cy5.5-labeled DNA primers complementary to the structure cassette. These mixtures were incubated at 85°C for 1 min, 60°C for 5 min, 35°C for 5 min and 50°C for 2 min to hybridize primers to RNA before initiation of reverse transcription. Reverse transcription was conducted according to manufacturer's instructions (Life Technologies; 50°C, 45 min, 0.5 μl/100 U Superscript III RT), after which RNA was hydrolyzed by incubating samples with NaOH (final concentration 400 mM) for 3 min at 95°C. Hydrolysis was terminated, and the solution was neutralized by addition of equimolar amount of HCl. Experimental and control reactions [(+) and (−)] were pooled, ethanol precipitated and re-suspended in deionized formamide. Transcription templates and equivalent primers labeled with WellRed D2 and Licor IR-800 dyes were used in accordance with manufacturer's instructions (Thermo Sequenase Cycle Sequencing kit, Affymetrix) to generate the DNA sequencing ladders used to identify reverse transcription products after capillary electrophoresis. Sequencing reactions were ethanol precipitated, re-suspended in deionized formamide and then pooled with the (+) and (−) reverse transcription products. Sample mixtures were analyzed using a CEQ800 Genetic Analysis System (Beckman-Coulter) programmed with the separation method described by Mitra *et al.* (30).

SHAPE data analysis

Electropherogram peaks were converted to SHAPE reactivity values using Shapefinder software as described previously (31). These values were normalized and outliers excluded as previously described (32). RNAstructure software (33,34) was used to predict RNA secondary structure(s) on the basis of pseudo-free energy constraints derived from SHAPE reactivity values. All reactivity data used in 2D structure analysis were averaged from at least two independent experiments. The contributions of individual RRE2 conformers to the ensemble SHAPE reactivity profile(s) were calculated as described in Supplementary Data.

Native gel electrophoresis

Six 10 pmol samples of RRE RNA were prepared in a total volume of 18 μl of aqueous solution, heated to 95°C for 2 min and then placed on ice for 2 min. Two microliters of 10× RNA buffer [500 mM Tris–HCl (pH 8.0), 1 M NaCl and ±50 mM MgCl₂] was subsequently added to each sample, which was then incubated at 37°C for 0, 5, 10, 20, 40 or 100 min, transferred to ice and then mixed with 2 μl of non-denaturing gel loading buffer (50% w/v glycerol, 0.05% xylene cyanol and 0.05% bromophenol blue). One picomoles of each sample was subjected to overnight non-denaturing polyacrylamide gel electrophoresis (PAGE) (~16 h; 8% polyacrylamide gel, 1× TBE, 4°C). Gels were stained with SYBR Green II RNA stain (Life Technologies) according to

manufacturer's instructions and quantified using a Typhoon Trio+ imager (GE Healthcare). Rate constants for transitions among RRE2 conformers were calculated as previously described (35) using Sigmaplot 11.0 (Systat Software, Inc.). This calculation is described in detail in Supplementary Data.

Methidiumpropyl EDTA–Fe(II) cleavage

Methidiumpropyl EDTA (MPE) was synthesized according to published procedures (26). RNA (20 pmol) was re-folded as described earlier in the text and evenly divided into 'experimental' and 'control' reactions. The MPE–Fe(II) complex was pre-formed by incubating MPE and Fe(SO₄)₂(NH₄)₂·6H₂O (1:0.4 ratio) for 2, 5, 10, 20 or 40 min at room temperature. The MPE–Fe(II) complex was added to the RNA solution to a final concentration of 1.0 μM and allowed to intercalate for 1 min at room temperature. Control reactions omitted MPE. The cleavage reaction was initiated by addition of sodium ascorbate and hydrogen peroxide (final concentrations 1 mM and 0.03%, respectively) and incubation at room temperature for 15 s. Reactions were quenched and precipitated by addition of 1/10 volume of 3 M sodium acetate, 1 μl of 20 μg/μl glycogen and 2.5 volumes ice-cold ethanol (95%). RNAs were re-suspended in 10 μl of water and subjected to reverse transcription as described earlier in the text. The resulting cDNAs were processed as described earlier in the text and separated by capillary electrophoresis.

Hydroxyl radical cleavage

RNA (20 pmol) was re-folded as described earlier in the text and divided equally into experimental and control reactions. Fe(II)–EDTA [25 mM Fe(SO₄)₂(NH₄)₂·6H₂O and 50 mM EDTA, pH 8.0], 0.15% hydrogen peroxide and 10 mM sodium ascorbate solutions were freshly prepared before each experiment. Hydroxyl radical cleavage was initiated by adding 2 μl of Fe(II)–EDTA complex, 1 μl of sodium ascorbate and 1 μl of hydrogen peroxide to the experimental reaction. A control reaction lacking Fe(II)–EDTA was performed in parallel. After incubation at 37°C for 1 min, reactions were quenched and precipitated by addition of 1/10 volume of 3 M sodium acetate, 1 μl of 20 μg/μl glycogen and 2.5 volumes ice-cold ethanol (95%). RNAs were re-suspended and reverse transcribed, and products were processed as described earlier in the text and separated by capillary electrophoresis.

RNA structure modeling

Molecular modeling was performed using RNAComposer (27), Discovery Studio v3.5 (Accelrys, San Diego, CA, USA) and PyMol v1.5.0.4 (The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA). Sequence and secondary structure data obtained by SHAPE for the closed form of RRE2 were submitted to the RNAComposer web server unedited and *in toto* (http://euterpe.man.poznan.pl/Home). The program returned the five lowest energy structures obtained by homology modeling and several rounds of

energy minimization. SHAPE reactivity values, however, are not used directly to either determine or validate RNA–RNA interactions in 3D. Notably, other than selecting from among multiple low-energy conformers generated by RNAComposer, no reformatting or manual intervention was required to generate the RRE2 closed form model presented in Figure 7C and E.

In contrast, because there are no structurally resolved high-order junctions in FRABASE (36) homologous to those reported here for the open and intermediate forms of RRE2, structures initially obtained by submission to RNAComposer were implausible. Consequently, sequence and secondary structure re-formatting was required both before and after submission to generate the 3D models of the RRE2 open and intermediate conformers. A detailed description of the procedure used is presented in Supplementary Data. The animation illustrating a potential RRE2 folding transition was generated with PDB coordinates generated for the open, intermediate and closed conformers together with the Morph module in Pymol v1.5.0.4.

RESULTS

Secondary structure mapping of RRE2 RNA

SHAPE (28,37,38) was used to determine the secondary structure of RRE2, probing RNAs after incubating in folding buffer at 37°C for 5, 10, 20, 40 or 100 min to (i) determine the lowest energy secondary structural form and (ii) identify potential conformational precursors that might exist during folding. Secondary structures generated after 5- or 40-min folding are shown in Figure 1A and B, respectively. SHAPE profiles derived from three independent 5-min folding experiments consistently produced an RRE2 secondary structural model resembling the original model reported for RRE1 (12). SL I, IIA, IIB, IIC, III, IV and V in the structure are designated accordingly (Figure 1A). In addition, the RRE2 structure depicted in Figure 1A contains a bulge within SL IIB and an internal loop within SL I that are homologous to the two purine-rich high-affinity Rev-binding sites reported for RRE1 (7,21–23).

In contrast, the RRE2 conformer predicted to form on prolonged incubation (40 min) contains internal base pairing between 117–119 and 186–188 nt (Figure 1B). Identical structures were also obtained after folding the RNA for 10, 20 or 100 min (data not shown). Formation of this internal 'bridge' converts the central five-way junction of Figure 1A into adjoining four- and three-way junctions, which correlates well with differences in SHAPE reactivity profiles produced from RNAs folded for 5- or 40-min (Figure 1C). Although differences in reactivity values for 117–119 and 186–188 nt transition from moderate/high and low/unreactive, respectively, differences in values for other nucleotides are either minimal or unlikely to reflect a transition from single- to double-stranded RNA (or *vice versa*). For example, the change in reactivity from 0.95 to 1.5 at 194 nt would not be expected to significantly affect secondary structure prediction, as

both values indicate a high propensity for these nucleotides to exist in an unpaired form.

### Identification of RRE2 folding intermediates by non-denaturing PAGE

Collectively, SHAPE data of Figure 1 strongly suggest that RRE2 undergoes a time-dependent conformational transition. To segregate RRE2 conformers and examine the time dependence of their distribution, RNAs were folded for 0, 5, 10, 20, 40 or 100 min, then rapidly cooled to 4°C and immediately subjected to non-denaturing PAGE at the same temperature. Three major RNA species (A, B and C) were resolved using this technique as shown in Figure 2A. A fourth, slower migrating, low-abundance species was not considered because (i) it seemed to be produced in part by transient dissolution of species C and (ii) its concentration remained essentially unchanged over the course of the experiment. Inclusion of $MgCl_2$ in the gels and/or running buffers had no effect on the distributions of conformers (data not shown).

Each major, independently migrating species was attributed to a separate RRE2 conformer, the amounts of which clearly change over time. The abundance of conformer A rapidly decreases during the first 20 min of folding. In contrast, the abundance of conformer B slightly increases after folding for 10 min, then slowly decreases with time. Conformer C continually accumulates until it constitutes the most abundant species at 100 min. Quantitative analysis of these data (Figure 2B) demonstrates that conversion of conformers A and B to C is consistent with consecutive first-order reaction kinetics. The rate constants for the A→B and B→C transitions are 0.089 ± 0.007 and $0.012 \pm 0.001 \, min^{-1}$, which translate to half-lives of 11 and 185 min for the A and B conformers, respectively.

Theoretical basis for determining the secondary structures of individual RNA conformers in a mixture

Despite compelling evidence for an RRE2 structural transition, assigning secondary structures to each conformer was challenging, as SHAPE requires that the sample assumes a uniform structure in solution (39). Moreover, as RRE2 RNA is not homogeneous after 5 or 40 min of folding, the data of Figure 1 represent averaged structures comprising contributions from all conformers in the respective mixtures. Consequently, although structures in Figure 1 likely resemble the most prevalent conformers, neither may perfectly match any specific conformation. Moreover, because species A is always in relatively low abundance, its conformation is likely not reflected in either structure. To address these issues, a mathematical method for extracting the contributions of individual conformers from ensemble reactivity values was devised.

We postulate that each ensemble reactivity value obtained from a mixture of RNA conformers equals the sum of reactivity values of the component conformers weighted according to the fractional contribution of each to the total RNA population. Moreover, if the ensemble reactivity values and fractional contributions of individual conformations change with differing
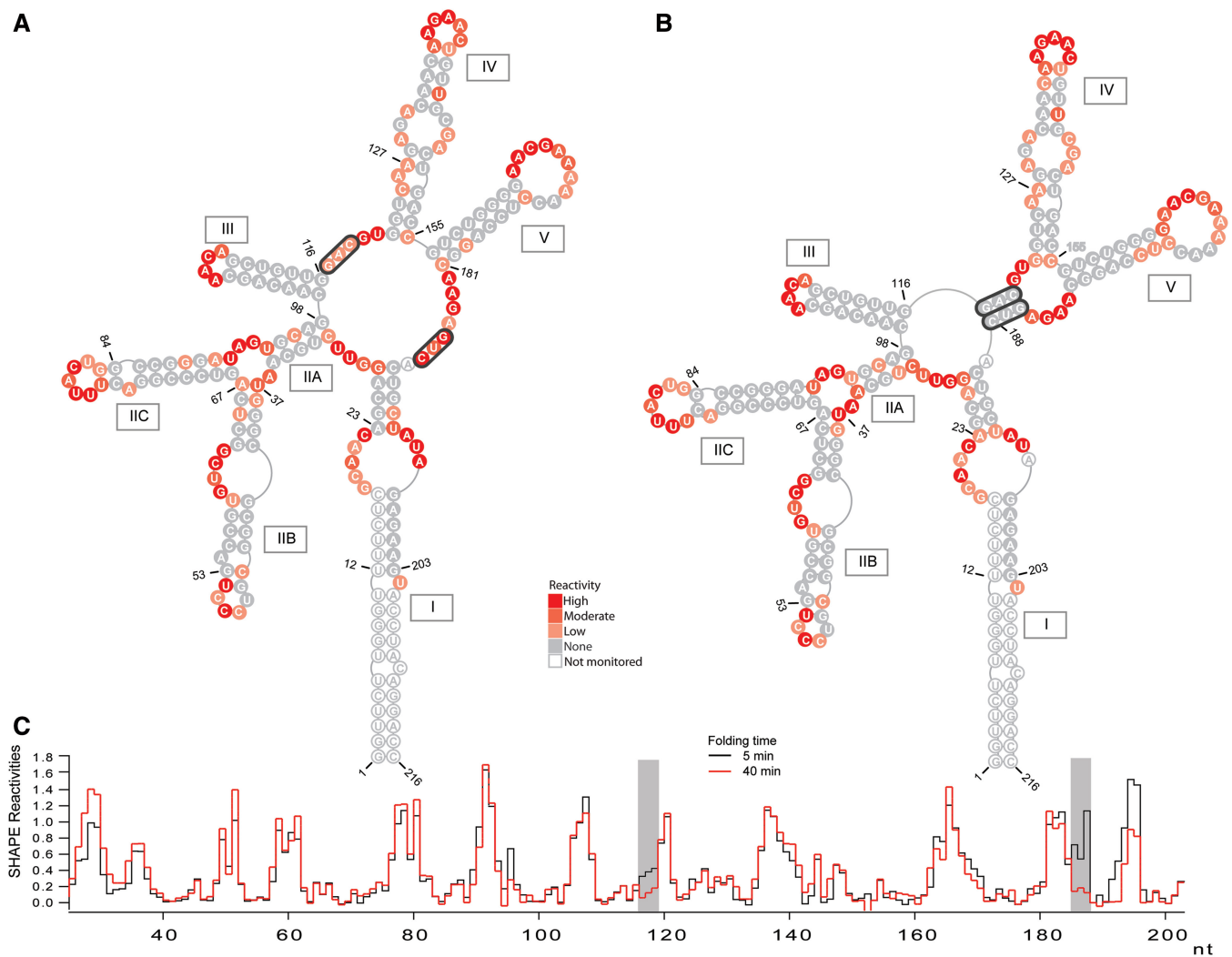
Figure 1. RRE2 structures obtained by SHAPE as a function of RNA folding time. RRE2 RNA assumes subtly different structures after folding for 5 min (**A**) or 40 min (**B**). Nucleotides are color-coded according to normalized SHAPE reactivity values. The most reactive (least conformationally constrained) nucleotides have high reactivities (>0.7 U) and are presented in red. Nucleotides with very low reactivity (<0.3 U) are highly constrained, most likely double-stranded and shown in gray. Nucleotides with intermediate reactivity values are presented in intermediate shades of red in increasing intensity from (0.3–0.5 U) to (0.5–0.7 U), respectively. Reactivities of the 5′ and 3′ RNA termini (represented in white) were not assigned because of the poor resolution near the primer and the final extension product. (**C**) Step plots for quantitative comparison of reactivity values obtained from RRE2 folded for 5 (black line) or 40 min (red line). Boxed regions correspond to the nucleotides involved in forming the bridge (i.e. 117–120 and 186–188 nt).

conditions (e.g. RNA folding time), and these values can be determined for a number of conditions equal to or greater than the total number of conformers in the mixture, then specific reactivity values can be mathematically calculated for each conformer (see Supplementary Data for algebraic formulae and additional details).

Validating the method for determining the secondary structures of individual RNA conformers in a mixture

To verify our method for deconvoluting ensemble reactivity values, SHAPE experiments were conducted on RNA constructs derived from the HIV-1 5′-UTR, the secondary structure of which has been well established (40). Specifically, the wild-type 5′-UTR+1 construct corresponds to the 5′ terminal 316 nt of HIV-1$_{NL4-3}$ RNA,

including the transactivation response (TAR) and Poly-A hairpins (Figure 3A, upper). The 5′-UTR+15 construct, in contrast, lacks the 5′-terminal 14 nt (Figure 3A, lower), which we predicted would free some or all of nt 46–57 from base pairing constraints, and that the reactivity values produced for these nucleotides in SHAPE experiments would increase accordingly.

Both RNAs were shown to be structurally homogeneous in solution (Figure 3B), and SHAPE data were consistent with the proposed structures (Figure 3A and C). Specifically, the reactivity values for 45–51 nt of the wild-type 5′-UTR+1 were relatively low, most likely indicating that the TAR stem is base paired and intact. Conversely, deleting the 14 5′-terminal nucleotides in the 5′-UTR+15 construct leaves nucleotides at the 3′-end of TAR largely unpaired, save for an unanticipated small internal hairpin
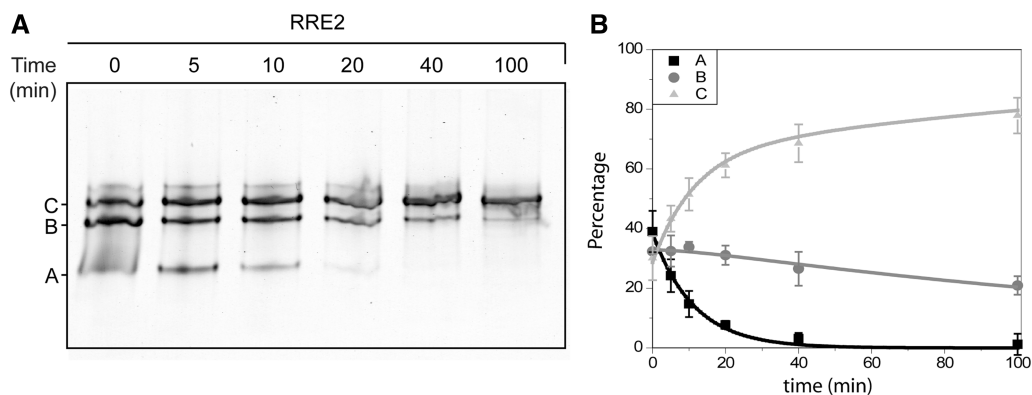
Figure 2. Time-dependent conformational analysis. (**A**) Native gel electrophoresis of RRE2 as a function of incubation time. RRE2 is a mixture of major conformers (A, B and C) at short incubation times, the relative ratio of which vary with time, with conformer C ultimately predominating. (**B**) Scatter plot representing percentage of each conformer as a function of time. Best-fit curves and rate constants k_1 and k_2 were calculated using SigmaPlot. Data points each represent the average of three independent measurements, and standard error is indicated.

(51–57 nt) predicted to form by RNAstructure software (Figure 3A, lower). As a consequence, reactivity values determined for 45–51 nt in this construct are considerably higher than those produced from the wild-type RNA (Figure 3C, gray box).

To recapitulate conditions where a single RNA was present in two conformations of different proportion in solution, SHAPE experiments were performed for 75:25% and 25:75% mixtures of 5′-UTR+1 and 5′-UTR+15 constructs, respectively. Ensemble reactivity values generated for these mixtures were then deconvoluted into component 'calculated' reactivity values for both 5′-UTR constructs, according to the algebraic transformation described earlier in the text and in Supplementary Data. Calculated values were then compared with the experimentally determined reactivity profiles, and the results are presented as step plots in Figure 4A and B, respectively. Overall, calculated and experimentally determined reactivity values are similar throughout the profiles of both constructs. Reactivity values for 45–51 nt were also accurately calculated, indicating that our methodology is sensitive enough to detect variations that might be present in structurally distinct RNA conformers. Finally, secondary structural models produced by inputting the calculated reactivity profiles into the RNAstructure software were identical to those produced using experimentally determined values (data not shown), further establishing the reliability of this methodology.

**Determining the secondary structures of individual RRE2 conformers**

Having validated the mathematical method for deconvoluting ensemble reactivity profiles, we used this approach to calculate reactivity profiles and determine the structures of RRE2 conformers A, B and C (Figure 5A–C, respectively). Interestingly, it seems that the structures determined by ensemble SHAPE and presented in Figure 1A and B are identical to those derived from the mathematically determined reactivity values of conformers B and C, respectively (Figure 5B and C). Conformer A, which could not be detected using simple

ensemble SHAPE, closely resembles conformer B, except that the nucleotides comprising SL IIA are unpaired, resulting in a six-way rather than a five-way central RNA junction (Figure 5A). The structures of conformers A, B and C (re-designated open, intermediate and closed, respectively), together with their sequential A→B→C conversion pathway, strengthen our postulate that RRE2 undergoes a progressive transition toward the most highly structured form.

Probing RRE2 with MPE–Fe(II)

MPE is a bifunctional reagent composed of methidiumpropyl, an intercalator motif reported to selectively bind certain sites within RNA, and EDTA, a chelator of metal ions, such as Fe(II) (41,42). Under oxidizing conditions, the Fe(II)–EDTA complex generates highly reactive hydroxyl radicals capable of damaging RNA and inducing strand scission. MPE–Fe(II) has been used to probe the 3D structure of tRNA by binding selectively at natural or engineered CpG steps in double-stranded RNA, especially when these motifs are adjacent to a single-nucleotide bulge (41,43). In a separate study, however, binding of MPE–Fe(II) to RNA was considered to be non-specific, and the reagent was used accordingly to map the solvent accessibility of RNA nucleotides (42). Here, MPE–Fe(II) has been used as a structural probe to validate the 2D models generated for RRE2 and provide information regarding the 3D organization of RRE2 substructures.

The RNA hydrolysis patterns produced by MPE–Fe(II) on RRE2 were distinct from those observed in previous reports (Figure 6A and B). The most consistent cleavage of RNA folded for 5 or 40 min occurred among nucleotides in both strands of the SL IIA and SL III stems, confirming the existence of these motifs. Moreover, hydrolysis of RRE2 was not widespread, indicating that MPE–Fe(II) binds to a limited number of sites on this RNA. It is also of particular significance that cleavage is not observed within the CpG steps predicted to exist in SL IIB and IIC, as RNAs were cleaved extensively within or adjacent to these putative high affinity MPE-binding sites in a previous study (41). Taken together, these data
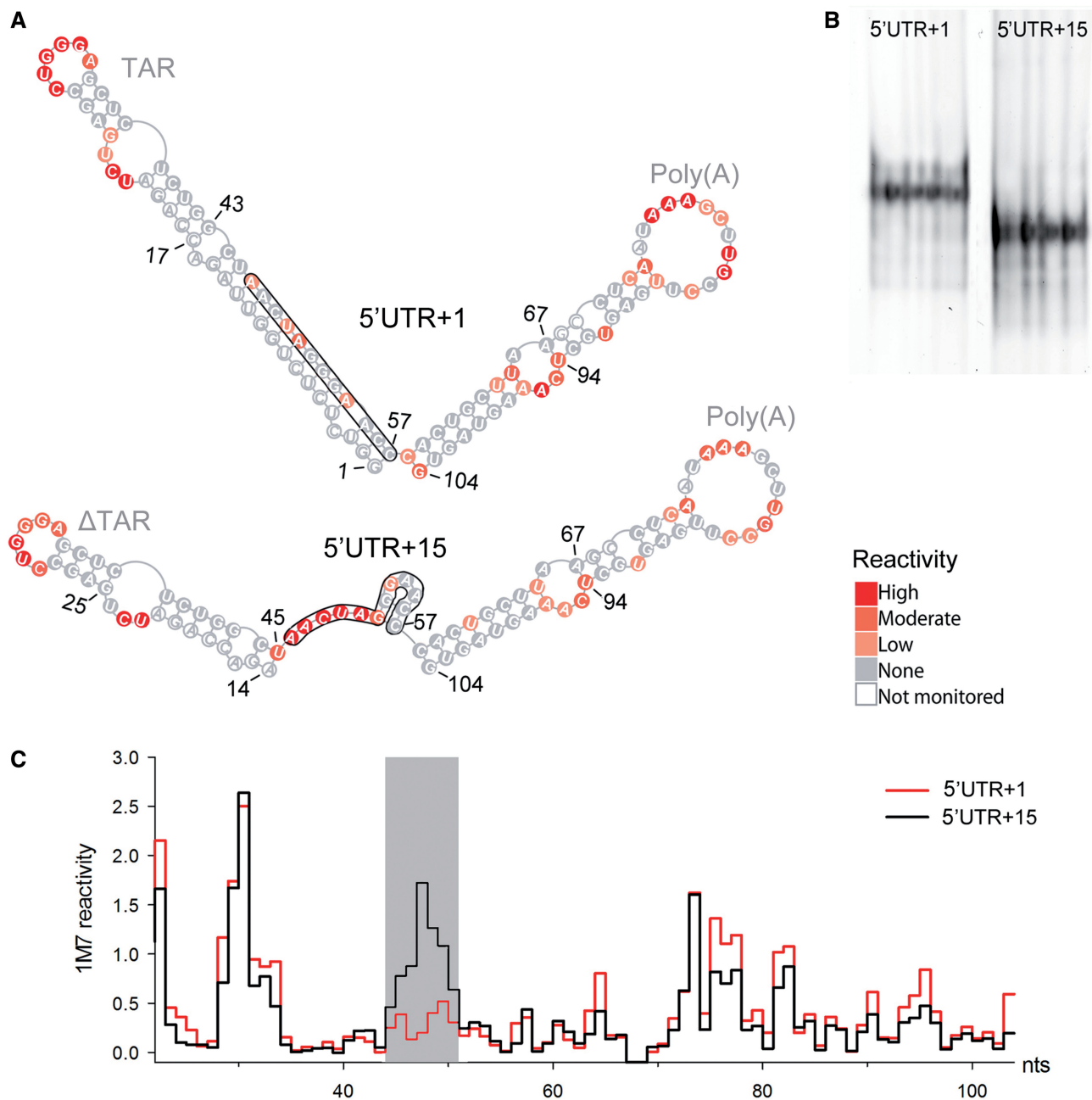
Figure 3. (A) Secondary structures of 5′-UTR+1 and 5′-UTR+15 RNAs obtained by SHAPE. Nucleotides originally predicted to be unpaired in the deletion mutant are highlighted. (B) Native gel electrophoresis confirms that 5′-UTR+1 and 5′-UTR+15 RNAs assume uniform conformations. (C) Step plots for quantitative comparison of reactivity values obtained from 5′-UTR+1 (red line) and 5′-UTR+15 (black line) RNAs. The boxed region highlights 45–51 nt, a subset of the sequence highlighted in (A), for which considerable differences in reactivity are observed, and which the RNA Structure software predicts will be unpaired in the mutant.

suggest that MPE binds to a select site(s) within RRE2 at or near the central junction region at the bases of SL IIA and III. This binding specificity was not anticipated and may indicate a dependence on the sequences and structures of individual RNAs.

The dependence of RRE2 cleavage on RNA folding time also confirms the structural transitions predicted by SHAPE. When probing RNA folded for 5 min, cleavage of 117–119 and 186–188 nt was relatively weak, suggesting the central bridge region had not yet formed in many of

the RNAs in the mixed population, and/or that, collectively, these nucleotides are not as close to the MPE-binding site(s) as during later stages of RRE2 folding. Such was not the case, however, with RRE2 RNA folded for 40 min, an extensive cleavage of bridge nucleotides was observed (Figure 6B). Not only does this cleavage pattern confirm that the base-paired bridge region in the closed RRE2 conformer exists and is proximal to both SL IIA and SL III but also that the population of RRE2 conformers changes over time.
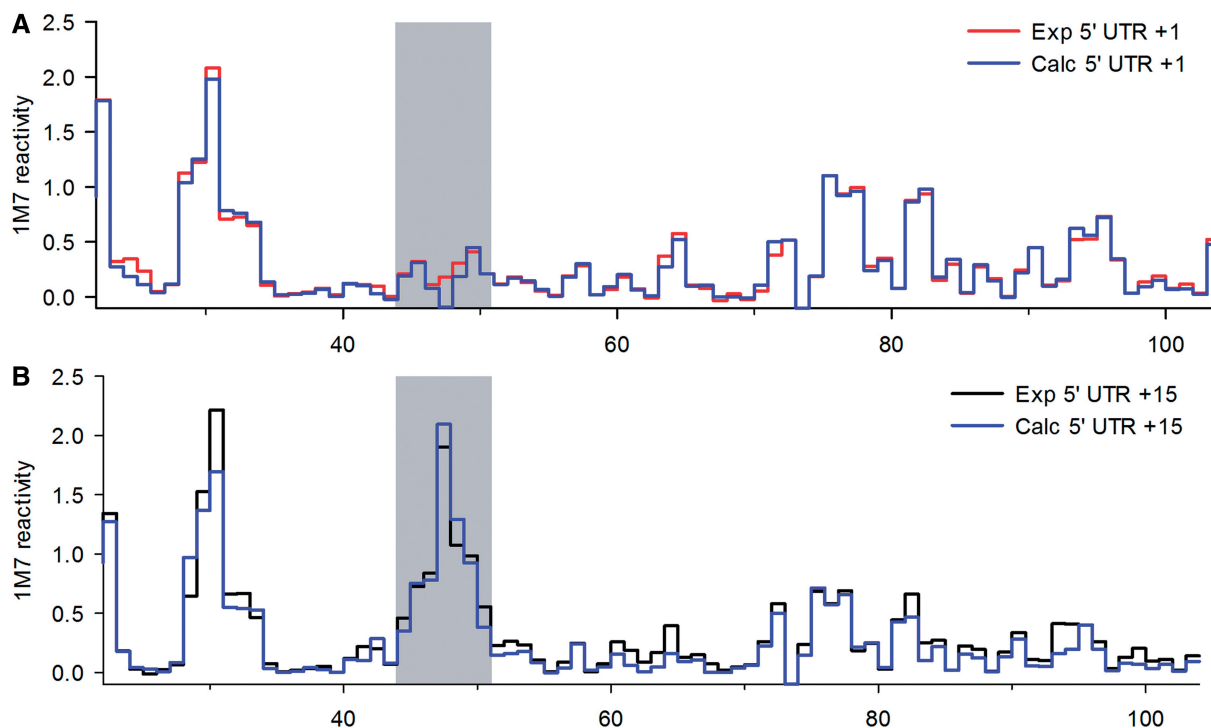
Figure 4. Step plots for quantitative comparison of reactivity values measured experimentally (black or red line) or determined by mathematical deconvolution of values obtained from RNA mixtures (blue lines). Values obtained for (**A**) 5′-UTR+1 RNA and (**B**) 5′-UTR+15 RNA nucleotides are indicated.
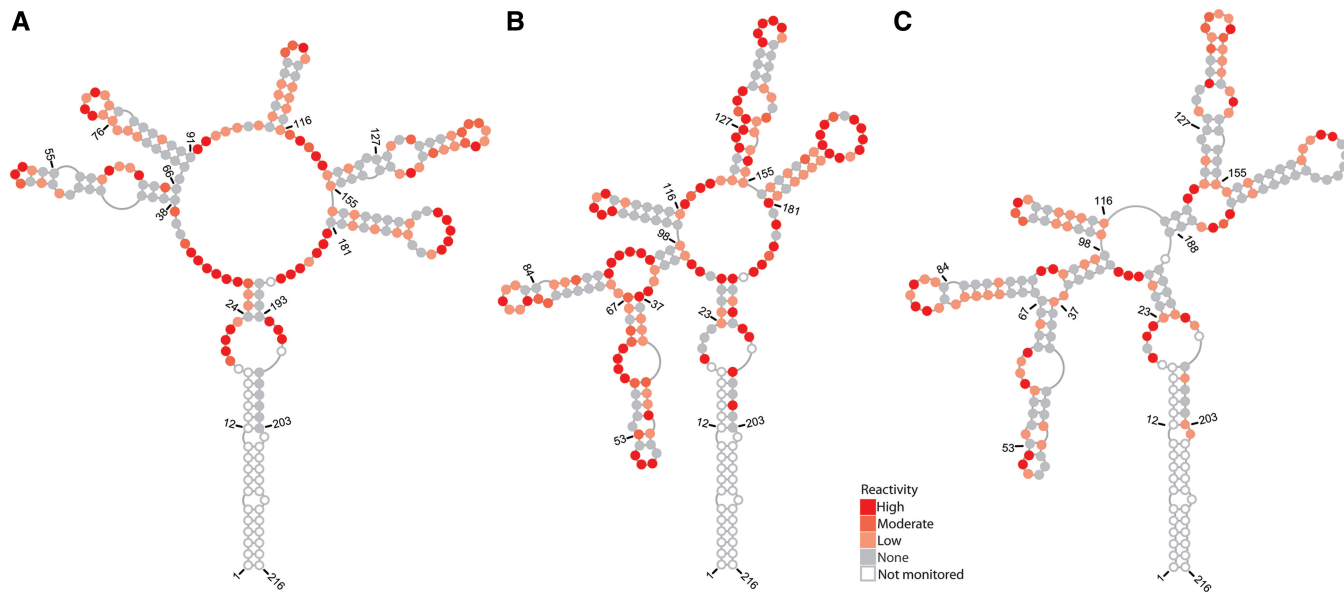
Figure 5. Secondary structures of the open (**A**), intermediate (**B**) and closed (**C**) conformers. The secondary structure of each conformer produced by RNA Structure software from reactivity values calculated mathematically as described in Supplementary Data.

Finally, parallel experiments were conducted using Fe(II)–EDTA as a probing reagent to confirm that the cleavage patterns depicted in Supplementary Figure S1A and B reflected the binding specificity of the intercalation moiety of MPE–Fe(II). This was indeed the case, as RRE2 RNA was cleaved much more extensively and with less specificity by Fe(II)–EDTA (compare Figure 6C, Supplementary Figure S1).

Modeling alternative RRE2 conformers

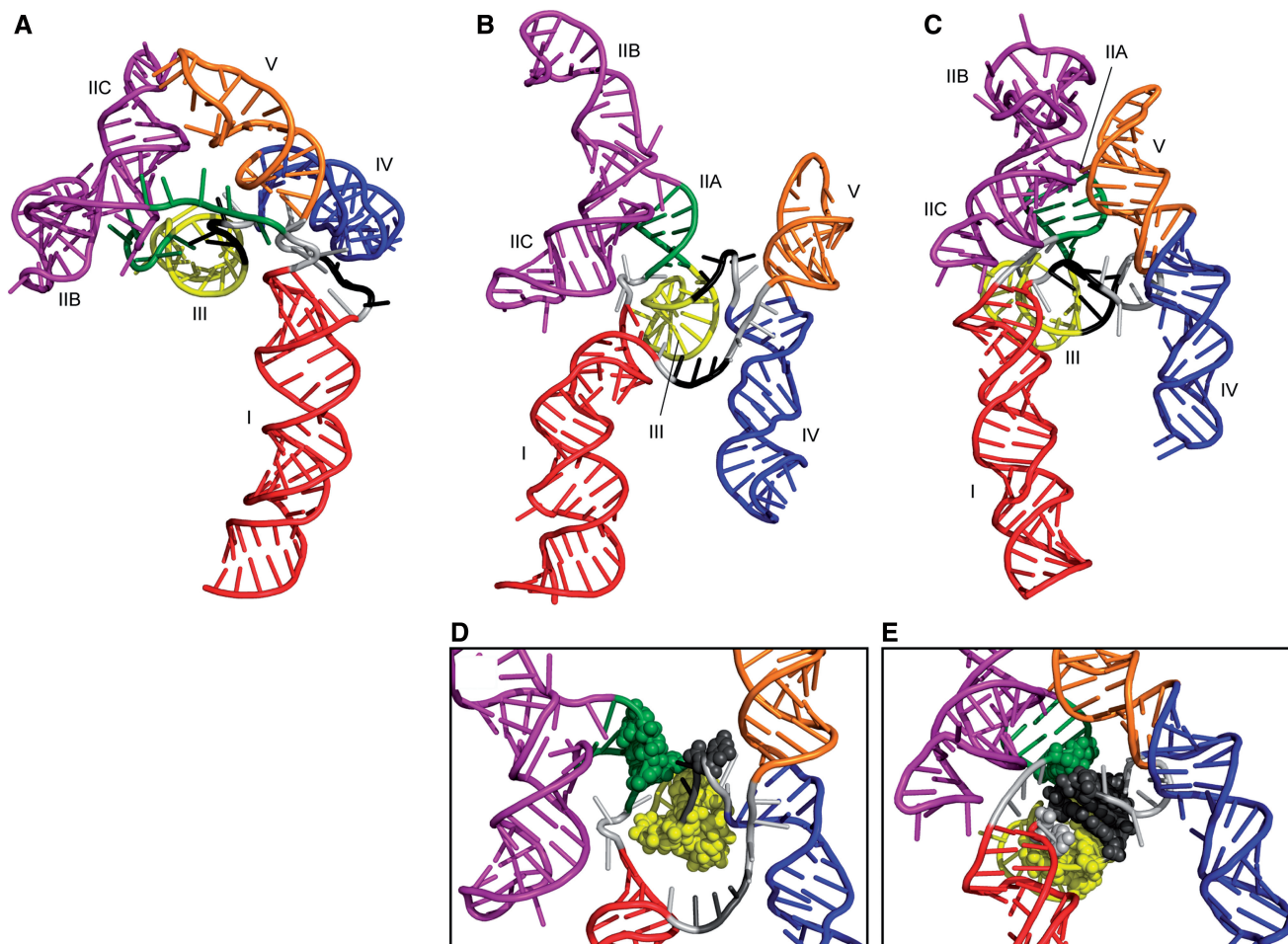The 3D models of the RRE2 open, intermediate and closed forms are presented in Figure 7A–C, respectively.
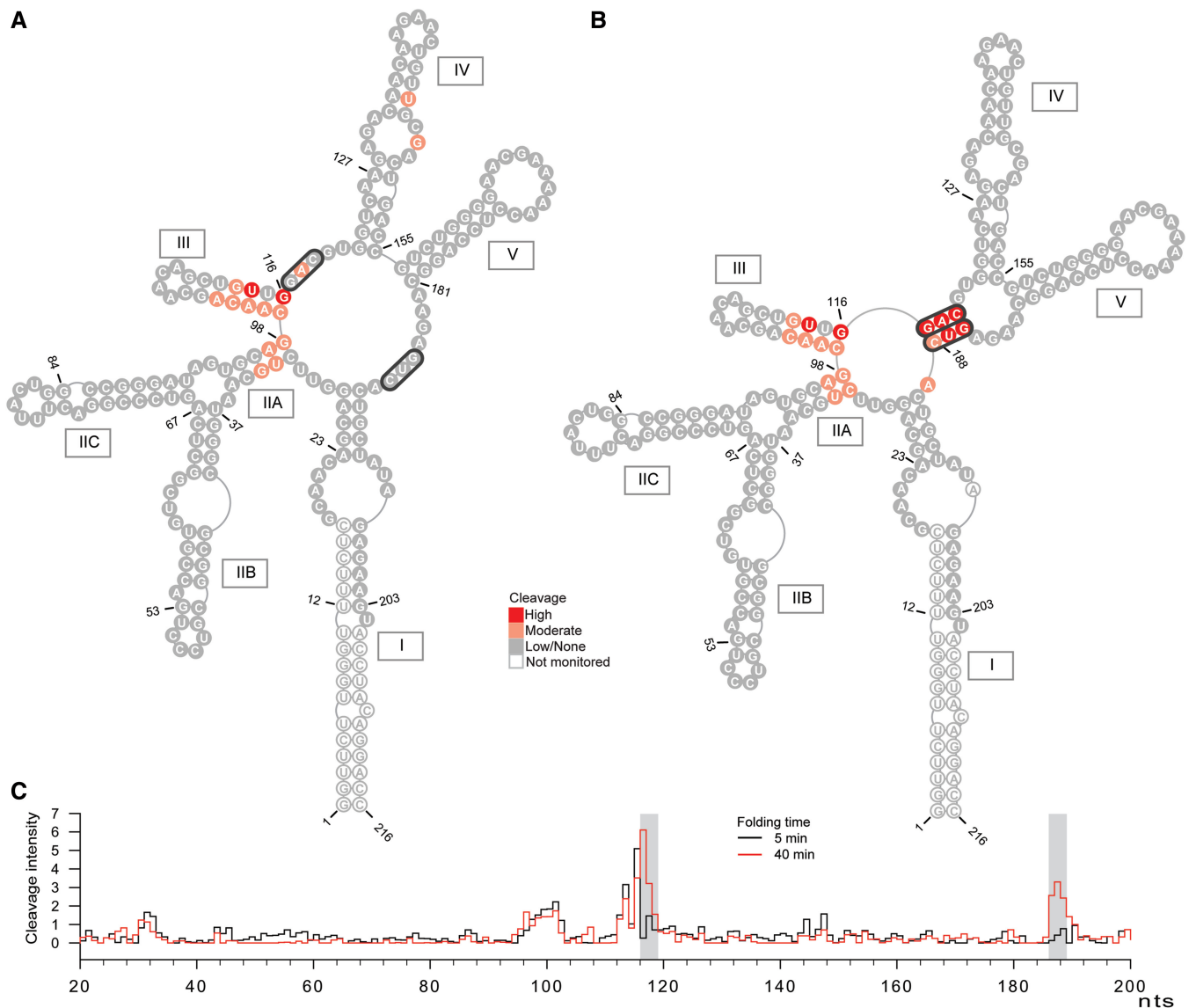
Figure 6. Probing RRE2 intermediate and close conformers with MPE–Fe(II) after folding for 5 min (**A**) or 40 min (**B**). Intermediate and closed secondary structural forms proposed to predominate during these folding stages are depicted. Nucleotides most frequently cleaved by MPE–Fe(II) are shown in shades of red. (**C**) Comparison of MPE–Fe(II) cleavage profiles obtained from RRE2 RNA folded for 5 (black line) or 40 min (red line). Boxed regions correspond to nucleotides involved in closing the structure.

Among the low-energy closed conformers predicted by RNAComposer (27), SL IIB was consistently oriented orthogonally to SL IIC and coaxially with SL I across the intervening four-way junction. SL IV and V were likewise arranged coaxially in most of the structures. Differences among conformers were mostly limited to the manner in which the SL IV–V substructure was rotated relative to the SL I–IIB axis around the 'bridge' helix linking the central four- and three-way junctions. The third lowest energy conformer was ultimately selected for presentation in Figure 7C because the face of the structure opposite the centrally located SL III formed a binding pocket amenable to consecutive and cooperative binding of Rev proteins to SL II, the central junction region and SL I, as has been proposed for HIV-1 (19).

In contrast to the closed conformation of RRE2 (Figure 7C), which was generated entirely and exclusively by RNAComposer using the inputted RRE2 sequence and secondary structure, creating the open and intermediate RRE2 conformations required manual intervention (Figures 6B and 7A). This is because no six- or five-way junctions homologous to those predicted to exist in the open and intermediate forms of RRE2 were available for comparison in the RNA structural database FRABASE (36). Perhaps most notable feature of all RRE2 conformations is the internal structural consistency of the four primary substructures: SL I, SL II, SL III and the SL IV–V axis. For example, in all predicted forms of RRE2, SL IIB and SL IIC are arranged orthogonally to each other, even though SL IIA is not formed in the open

Figure 7. RRE2 3D structural models. (**A–C**) Models of the open, intermediate and closed forms, respectively. Secondary structural motifs are indicated and color-coded as follows: SL I, red; SL IIA, dark green, SL IIB, IIC and adjacent connecting loops, magenta; SL III, yellow; SL IV, blue; SL V, orange. A near view of the (**D**) intermediate and (**E**) closed forms of RRE2 highlight nucleotides cleaved by MPE–Fe(II) with an intensity of 3 or higher according to the cleavage intensity scale of Figure 6C (atoms represented as spheres).

conformer. In addition, SL IV and SL V are arranged coaxially in conformers generated by RNAComposer. This is likely because there are no unpaired nucleotides separating the terminal base pairs of these stem loops to disrupt direct π-stacking between these elements.

The most significant differences among the structures involve the central junction (including, in this case, SL IIA and the bridge), and how changes within the junction affect positioning of RRE2 substructures relative to each other. An animation illustrating the transformations required for conversion from RRE2 open to intermediate to closed form is provided as a Supplementary Movie. Transition from the open to intermediate conformation requires several concerted translations/rotations of RRE2 substructures, the most notable of which is rotation of SL IIB/IIC caused by base pair formation in SL IIA (compare Figure 7A and B). Pairing of the bridge helix defines the subsequent transition from the intermediate to closed RRE2 form, which also involves (i) inward rotation of the SL IV–V substructure, (ii) folding of the apical loop of SL IIB toward the central junction and (iii) formation of mutually stabilizing

contacts between SL IIB and SL V. In most variants of the RRE2 closed form predicted by RNAComposer, SL IIB is arranged orthogonally to SL IIC and coaxially with SL I, consistent with current models of Rev assembly on the HIV-1 RRE (26), whereby the protein initially binds to high affinity sites located within purine-rich internal loops within SL IIB (7,44) and SL I (22), then oligomerizes in linear fashion along an SL IIB–SL I axis. Given the close contact between SL IIB and SL V in the RRE2 closed conformer, some rotational/translational flexibility of the SL IV–SL V substructure may be required to create space for Rev2 binding to the putative SL IIB site. Such flexibility is indeed likely, as rotational positioning of this SL IV–SL V substructure around the bridge helix was the most variable feature among closed form conformers.

Mapping of MPE cleavage patterns onto the 3D structures served to validate the intermediate and closed RRE2 conformations. The grouping of cleavage products (Figures 6, 7D and E) suggests preferential binding to a single site near the junction of SL IIA, III and 117–119 nt of the central junction(s) in both the intermediate and closed forms. In the intermediate form, as 186–188 nt are unpaired

and displaced relative to 117–119 nt, they are not cleaved by MPE (Figures 6 and 7D). In contrast, 117–119, 186–188 nt and two adjacent nucleotides are brought into proximity of the putative MPE-binding site on base pairing at the bridge region in the closed form and are, therefore, subject to hydrolysis by the chemical nuclease (Figures 6 and 7E). The patterns of MPE cleavage and the 3D models of both the intermediate and closed forms of RRE2 seem to be mutually supportive, i.e. hydrolysis patterns are consistent with the positioning of cleaved residues in both cases, and the models support the notion of a primary MPE-binding site near the base of SL III.

DISCUSSION

In this study, SHAPE was combined with a hydroxyl radical footprinting strategy to determine the secondary structure of RRE2. Collectively, these complementary approaches suggest RRE2 undergoes two conformational transitions before assuming the most energetically favorable conformer. Analysis of the intermediate conformer revealed a secondary structural organization similar to that observed for the five-way junction model proposed for RRE1 (12,17). In both RNAs, five-stem/stem–loop motifs are predicted to be arranged around a central five-way junction, one of which (SL II) bifurcates to form a three-way junction. Likewise, a purine-rich bulge within the SL I motif and proximal to the central junction region is common to both RNAs.

Although similar, important differences between the RREs are also apparent. For example, the lengths of the connecting loops vary significantly between RRE2 and RRE1, as do the sizes and locations of internal loops and bulges. These differences are particularly noteworthy within the SL II substructure(s) and the purine-rich bulge(s) of SL I, as these elements in RRE1 are bound by Rev1 with high affinity. Previous computational analysis of the SL II motif in RRE1 revealed that although there is some sequence diversity in this region, the predicted secondary structure of this element remains essentially the same across a wide variety of HIV-1 strains (45). The predicted structures of RRE2 variants are likewise similar, as shown by our analysis of the bifurcated SL II substructures of 23 independent HIV-2 RRE variants (Supplementary Figure S2). Taken together, these data suggest that structure, rather than sequence, is the principal determinant of RRE function.

RRE2 was found to assume three distinct conformations during folding, where the open and intermediate conformers are ultimately converted into the more stable, highly structured closed form, which predominated at 100 min. Whether this pathway reflects what occurs *in vivo* remains to be established, but it is mostly consistent with models of RNA folding where formation of stem–loops in which the two halves of the stem are in close proximity is favored. More significantly, mathematical determination of SHAPE reactivity values, and therefore structures, of individual conformers in a mixture represents a novel approach that allows secondary structures of individual forms of RNA in a heterogeneous sample to be determined. This method only requires that (i) the individual RNA conformers can be fractionated and quantified and (ii) the distribution of conformers within the population of RNAs changes over time (i.e. during folding) or with differing solution conditions.

In contrast to our results with RRE2, the structure of a homologous RRE1 construct was invariant over the course of the 100-min incubation period (data not shown), thus any potential folding precursors could be neither characterized nor evaluated. Conceivable, highly transient folding intermediates might be probed using fast-acting SHAPE reagents (39) if a means of arresting the folding process during its early stages can be devised.

The recently published RNAComposer modeling software proved extremely useful for developing a 3D model of RRE2 (27). We developed plausible models when necessary by sub-dividing the structure into smaller motifs and/or re-ordering nucleotide sequence as described in Supplementary Data. If, as predicted, the conformations of RRE2 substructures are generally independent of the global structural context, the key to obtaining a reliable closed form structure is determining the correct organization of the central junction region; we believe that recognition of the bridge motif constitutes an important step in this direction. Given the similarities in the predicted arrangements of RRE2 and RRE1 structural motifs, it seems likely that the central junction regions of the two elements are also similarly arranged. Despite the apparent absence of a base-paired bridge motif in RRE1, the SL III-SL IV and SL V-SL I connecting loops in this RNA may be brought into similarly close proximity by non-standard base pairing or other interactions not detectable by either SHAPE or the RNAstructure software. Such interactions might serve to stabilize what otherwise seems to be a highly flexible RRE1 central junction region, which may be of particular importance given the overall complexity of the motif (12).

Finally, the 3D model of the RRE2 closed form depicts a coaxial arrangement of SL IIB and SL I motifs, suggesting that Rev2 molecules might assemble cooperatively along the long axis of the RRE2 in a manner resembling that proposed for RRE1. We anticipate that such a nucleoprotein complex would be stabilized by both Rev–Rev and Rev–RNA interactions, the latter of which would involve Rev2 binding to the two high affinity sites on RRE2 homologous to those observed on RRE1, as well as several other sites. A structural model of the Rev2–RRE2 complex might be developed by combining the 3D RRE2 model presented here with the high-resolution nuclear magnetic resonance (NMR) and X-ray crystal structures of homologous Rev1–RNA and Rev1–Rev1 complexes, respectively (19,44). Although such a model would be speculative and require extensive experimental validation, it could also potentially provide critical insight into the structural basis for nuclear export of HIV RNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1 and 2 and Supplementary Movie.

FUNDING

REFERENCES

1. Hammarskjold,M.L. (1997) Regulation of retroviral RNA export. *Semin. Cell Dev. Biol.*, **8**, 83–90.
2. Cullen,B.R. (2003) Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.*, **28**, 419–424.
3. Sodroski,J., Goh,W.C., Rosen,C., Dayton,A., Terwilliger,E. and Haseltine,W. (1986) A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*, **321**, 412–417.
4. Emerman,M., Vazeux,R. and Peden,K. (1989) The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. *Cell*, **57**, 1155–1165.
5. Malim,M.H., Bohnlein,S., Fenrick,R., Le,S.Y., Maizel,J.V. and Cullen,B.R. (1989) Functional comparison of the Rev trans-activators encoded by different primate immunodeficiency virus species. *Proc. Natl Acad. Sci. USA*, **86**, 8222–8226.
6. Dillon,P.J., Nelbock,P., Perkins,A. and Rosen,C.A. (1990) Function of the human immunodeficiency virus types 1 and 2 Rev proteins is dependent on their ability to interact with a structured region present in env gene mRNA. *J. Virol.*, **64**, 4428–4437.
7. Malim,M.H., Tiley,L.S., McCarn,D.F., Rusche,J.R., Hauber,J. and Cullen,B.R. (1990) HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell*, **60**, 675–683.
8. Fernandes,J., Jayaraman,B. and Frankel,A. (2012) The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol.*, **9**, 6–11.
9. Fornerod,M., Ohno,M., Yoshida,M. and Mattaj,I.W. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90**, 1051–1060.
10. Pollard,V.W. and Malim,M.H. (1998) The HIV-1 Rev protein. *Annu. Rev. Microbiol.*, **52**, 491–532.
11. Garrett,E.D. and Cullen,B.R. (1992) Comparative analysis of Rev function in human immunodeficiency virus types 1 and 2. *J. Virol.*, **66**, 4288–4294.
12. Kjems,J., Brown,M., Chang,D.D. and Sharp,P.A. (1991) Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl Acad. Sci. USA*, **88**, 683–687.
13. Mann,D.A., Mikaelian,I., Zemmel,R.W., Green,S.M., Lowe,A.D., Kimura,T., Singh,M., Butler,P.J., Gait,M.J. and Karn,J. (1994) A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J. Mol. Biol.*, **241**, 193–207.
14. Charpentier,B., Stutz,F. and Rosbash,M. (1997) A dynamic *in vivo* view of the HIV-I Rev-RRE interaction. *J. Mol. Biol.*, **266**, 950–962.
15. Zhang,M.J. and Dayton,A.I. (1996) Two secondary structures for the RRE of HIV-1? *J. Acquir Immun. Defic. Syndr. Hum. Retrovirol.*, **13**, 403–407.
16. Legiewicz,M., Badorrek,C.S., Turner,K.B., Fabris,D., Hamm,T.E., Rekosh,D., Hammarskjold,M.L. and Le Grice,S.F. (2008) Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. *Proc. Natl Acad. Sci. USA*, **105**, 14365–14370.
17. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
18. Hammarskjold,M.H. and Rekosh,D. (2011) A long-awaited structure is rev-ealed. *Viruses*, **3**, 484–492.
19. Daugherty,M.D., Liu,B. and Frankel,A.D. (2010) Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. Mol. Biol.*, **17**, 1337–1342.
20. DiMattia,M.A., Watts,N.R., Stahl,S.J., Rader,C., Wingfield,P.T., Stuart,D.I., Steven,A.C. and Grimes,J.M. (2010) Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element. *Proc. Natl Acad. Sci. USA*, **107**, 5810–5814.
21. Olsen,H.S., Nelbock,P., Cochrane,A.W. and Rosen,C.A. (1990) Secondary structure is the major determinant for interaction of HIV rev protein with RNA. *Science*, **247**, 845–848.
22. Daugherty,M.D., D'Orso,I. and Frankel,A.D. (2008) A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol. Cell*, **31**, 824–834.
23. Huang,X.J., Hope,T.J., Bond,B.L., McDonald,D., Grahl,K. and Parslow,T.G. (1991) Minimal Rev-response element for type 1 human immunodeficiency virus. *J. Virol.*, **65**, 2131–2134.
24. Jain,C. and Belasco,J.G. (2001) Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell*, **7**, 603–614.
25. Pond,S.J., Ridgeway,W.K., Robertson,R., Wang,J. and Millar,D.P. (2009) HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc. Natl Acad. Sci. USA*, **106**, 1404–1408.
26. Daugherty,M.D., Booth,D.S., Jayaraman,B., Cheng,Y. and Frankel,A.D. (2010) HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc. Natl Acad. Sci. USA*, **107**, 12481–12486.
27. Popenda,M., Szachniuk,M., Antczak,M., Purzycka,K.J., Lukasiak,P., Bartol,N., Blazewicz,J. and Adamiak,R.W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.
28. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
29. Steen,K.-A., Siegfried,N.A. and Weeks,K.M. (2011) Synthesis of 1-methyl-7-nitroisatoic anhydride (1M7). *Nat. Protoc. Exch.*, doi:10.1038/protex.2011.255.
30. Mitra,S., Shcherbakova,I.V., Altman,R.B., Brenowitz,M. and Laederach,A. (2008) High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.*, **36**, e63.
31. Vasa,S.M., Guex,N., Wilkinson,K.A., Weeks,K.M. and Giddings,M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
32. Low,J.T. and Weeks,K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.
33. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
34. Mathews,D.H. (2006) RNA secondary structure analysis using RNAstructure. *Curr. Protoc. Bioinformatics*, **Chapter 12**, Unit 12.6.
35. Jones,C.T. (1970) Determination of the kinetic constants of two consecutive first-order reactions. *Biochem. J.*, **118**, 810–812.
36. Popenda,M., Szachniuk,M., Blazewicz,M., Wasik,S., Burke,E.K., Blazewicz,J. and Adamiak,R.W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics*, **11**, 231.
37. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension

(SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.

38. Sztuba-Solinska,J. and Le Grice,S.F. (2012) Probing Retroviral and Retrotransposon Genome Structures: The "SHAPE" of Things to Come. *Mol. Biol. Int.*, **2012**, 530754.

39. Weeks,K.M. and Mauger,D.M. (2011) Exploring RNA structural codes with SHAPE chemistry. *Accounts Chem. Res.*, **44**, 1280–1291.

40. Berkhout,B. and van Wamel,J.L. (2000) The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA*, **6**, 282–295.

41. Gherghe,C.M., Leonard,C.W., Ding,F., Dokholyan,N.V. and Weeks,K.M. (2009) Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. *J. Am. Chem. Soc.*, **131**, 2541–2546.

42. Han,H., Schepartz,A., Pellegrini,M. and Dervan,P.B. (1994) Mapping RNA regions in eukaryotic ribosomes that are accessible to methidiumpropyl-EDTA.Fe(II) and EDTA.Fe(II). *Biochemistry*, **33**, 9831–9844.

43. White,S.A. and Draper,D.E. (1989) Effects of single-base bulges on intercalator binding to small RNA and DNA hairpins and a ribosomal RNA fragment. *Biochemistry*, **28**, 1892–1897.

44. Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Muhandiram,D.R., Kay,L.E., Frankel,A.D. and Williamson,J.R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273**, 1547–1551.

45. Phuphuakrat,A. and Auewarakul,P. (2003) Heterogeneity of HIV-1 Rev response element. *AIDS Res. Hum. Retroviruses*, **19**, 569–574.