The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome

Carine TISNÉ[1], Bernard P. ROQUES[2] and Frédéric DARDEL[1]

[1]Laboratoire de Cristallographie et RMN Biologiques, UMR 8015 CNRS and [2]Pharmacochimie Moléculaire et Structrale, INSERM U266, FRE2463 CNRS, Faculté de Pharmacie, 4 avenue de l'Observatoire, 75006 PARIS.

Running title : HIV-1 primer annealing

Correspondance should be addessed to F.D. *email: dardel@pharmacie.univ-paris5.fr*

Tel : +33 1 53 73 99 93

Fax : +33 1 53 73 99 25

SUMMARY

Reverse transcription of HIV-1 viral RNA uses human $tRNA^{Lys}_3$ as a primer. The first step of viral replication is thus the annealing of the primer tRNA onto the primer-binding site located in the 5' leader region of the viral RNA. This involves large rearrangements of both RNA structures and requires the chaperone activity of the viral nucleocapsid protein. We have developped a novel approach to analyse dynamically such RNA refolding events, using heteronuclear NMR spectroscopy of mixtures of ^{15}N-labelled and unlabelled large RNA fragments (up to 50 kDa). We have thus been able to characterise the detailed mechanisms of both heat- and nucleocapsid-mediated annealing, and to identify previously unknown key steps. The role played by the nucleocapsid is two-fold: it facilitates strand exchange at the level of the tRNA acceptor stem, presumably via its basic N- and C-terminal extensions and it unlocks the highly stable tertiary interactions at the level of the TΨC loop, most likely by specific interactions involving its two zinc knuckles.

INTRODUCTION

Structured RNAs have been found to be involved in an increasing number of fundamental biological functions. These structures are often quite plastic and rearrangements of RNA folds have been described to act as switches or recognition elements in a large number of biological processes : mRNA splicing, translational control of gene expression, ribosome function, ribosome recruitment, antisense gene silencing, viral replication, retrotransposition... In several cases, these structural rearrangements have been shown to require action of "helper" proteins such as RNA helicases or RNA chaperones (1-3). A large amount of data has been gathered on the factors required

for such RNA rearrangements, in terms of nucleotide sequence or associated proteic factors, and sometimes on the structure of the starting and final states. However, at a molecular level, very little is known regarding the actual dynamic mechanism of such refolding events. This is because it is technically difficult to track the sequence of fleeting events using classical structural biology techniques. So far, the most successful approaches have used various types of fluorescence spectroscopy, which can provide useful dynamical data, even for very short timescales (4,5). However, it is difficult to derive detailed structural information from such analyses, for regions of the molecule afar from the chromophore(s).

The present study addresses the major structural RNA rearrangements which occur when $tRNA^{Lys}_3$, the primer of HIV-1 reverse transcriptase, is annealed onto the viral genome. This system has been characterized extensively by numerous biochemical and virological studies. The 18 nucleotides at the 3' end of the tRNA are complementary to a segment of the 5'-UTR of viral genome, the primer binding site (PBS). Within the viral particle, the primer tRNA is already placed onto the viral RNA(6) with which it forms extended secondary and tertiary interactions. In addition to the tRNA/PBS duplex formation, additional interactions have indeed been proposed to take place, in particular between the U-rich anticodon of the tRNA

and an A-rich loop located immediately upstream from the PBS with the viral RNA (7-11). This complex RNA assembly appears to be involved in the recruitment of reverse transcriptase and is required for efficient initiation of viral replication *in vivo* (12). The formation of an initiation-proficient complex does not occur spontaneously at physiological temperature, but requires the action of the viral nucleocapsid protein (3), which appears to act as a nucleic acid chaperone by facilitating strand exchange (13).

The mechanism by which the nucleocapsid promotes annealing of the primer tRNA and the exact sequence of events remain obscure. It was originally observed that, in the absence of magnesium, excess NC was capable of denaturing tRNA, but that magnesium rendered tRNA resitant to this destabilisation (14). This suggested that NC could possibly act as an RNA helicase, at least under certain conditions. More recent studies using a variety of biophysical techniques have shown that this is probably not the case and *per se*, the NC cannot "open" $tRNA^{Lys}_3$ (15-17), although some degree of destabilisation of RNA structures by NC has been observed (5,16). It is thus likely that the viral RNA is also required to disrupt the primer structure, assisted by the chaperone activity of the viral nucleocapsid. The dynamical mechanism of this process is however still unknown as only the initial and final states have been characterised.

Using heteronuclear NMR of RNA/RNA/protein complexes under controlled temperature conditions, we have been able to observe the progressive formation of the HIV-1 reverse transcription initiation complex in a quasi-continuous fashion. In particular, we have identified the nucleation site where the viral RNA starts invading the $tRNA^{Lys}_3$ structure, we have characterised previously unexpected intermediates and differentially annealed states. We were also able to characterise the different roles of the nucleocapsid in these various steps and to propose a detailed model of the NC-mediated annealing process. The viral protein plays a key role, by "unlocking" stable tertiary interactions within the primer tRNA.

EXPERIMENTAL PROCEDURES

Samples

Recombinant ^{15}N-labelled $tRNA^{Lys}_3$ was expressed *in vivo* and purified as previously described (18). The PBS (18 nucleotides) and the HIV-1 A-rich loop (25 nucleotides) were purchased from either Dharmacon research or MWG Biotech. The sequences of RNA samples used in this study are indicated in figure 1. The HIV-1 viral RNA fragment (65 nucleotides) that contains both the PBS and the HIV-1 A-rich loop spans nucleotides 132 to 196 of the full length viral RNA (HIV-1_{Mal} isolate). The (1-55)NCp7 protein was obtained by chemical synthesis as already reported (19) and used as zinc-complexed form (20). Poly-L-lysine (molecular weight 4000-15000 g/mol) was from Sigma-Aldrich.

Samples were dissolved in 10mM potassium phosphate buffer pH 6.5 containing 10% ^2H_2O. The interaction between the $tRNA^{Lys}_3$ and the HIV-1 A-loop was carried out at 0.5 mM of each species in 100 mM NaCl and 5 to 10 mM $MgCl_2$ according to a previous study that measured the equilibrium binding constant of $tRNA^{Lys}_3$ complexed with the HIV-1 A-loop in various salt conditions (21). The interaction between the $tRNA^{Lys}_3$ and the PBS was performed in 300 mM KCl at an RNA concentration of 0.4 mM each for the temperature jump or 1 mM each for the progressively heated sample. The interaction between the $tRNA^{Lys}_3$ and the PBS in the presence of NCp7(1-55) protein was conducted as follows: $tRNA^{Lys}_3$ at a concentration of 0.35 mM was first mixed to two equivalents of the NC protein, and reference NMR experiments (NOESY and TROSY –see below) were first performed at both 15°C and 30°C, the sample was then taken out of the spectrometer and then one equivalent of PBS was added. Similar conditions were used for the study with two equivalents of the poly-L-lysine. For all experiments involving addition PBS oligoribonucleotides to the tRNA sample, the two RNA solutions were cooled on ice prior to the mixing and the sample

was immediately transferred to the NMR spectrometer, in order to minimise the start of the annealing process before data acquisition.

For the temperature jump, the $tRNA^{Lys}_3$/PBS mix was heated directly to 50°C, using the NMR spectrometer variable temperature unit. For the progressive heating procedure, the $tRNA^{Lys}_3$/PBS mix was heated from 15°C to 80°C by increments of 5°C or 10°C, and NMR spectra were recorded at each step. At higher temperatures, the exchange rate of imino protons with water is however increased and can severely broaden some peaks. Therefore, after recording experiments at each temperature step, the samples were subsequently cooled back to 15°C and NMR experiments were also recorded under conditions identical to the reference spectra.

For the study of the formation of the $tRNA^{Lys}_3$/PBS complex in the presence of two equivalents of NC (or poly-L-lysine), the NMR sample were heated from 5°C to 35°C by steps of 10°C in order to stay within physiological conditions and avoid denaturation of the protein. In all cases, when the $tRNA^{Lys}_3$/PBS annealing was completed, a stoichiometric amount of the A-rich loop RNA fragment was added to the sample and a TROSY experiment at 15°C was recorded, in the presence of 100 mM NaCl and 5 to 10 mM $MgCl_2$. The interaction between the larger HIV-1 viral fragment (65 nucleotides) and $tRNA^{Lys}_3$ was performed at 1 mM of each species, under the same buffer and temperature conditions as the $tRNA^{Lys}_3$/PBS complex.

NMR

2D spectra were recorded on a Bruker Avance DRX600 spectrometer equipped with a standard triple resonance probe. 2D NOESY experiments in 90% H_2O/10% ^2H_2O were recorded at various temperatures (see above) with mixing times of 150 ms when the $tRNA^{Lys}_3$/PBS complex was not formed, or 120 ms after formation of the complex. The spectral widths were 12600 Hz in both dimensions. 512 t1 increments were recorded each

with 2048 data points. The solvent signal suppression was performed using the 'jump and return' sequence (22). ^{1}H-^{15}N TROSY (23) experiments were carried out with 256 t1 increments and 1024 data points per increment. The spectral widths were 4500 Hz and 2494 Hz in the proton and nitrogen dimensions, respectively.

RESULTS

NMR analysis of large RNA complexes remains technically challenging. We therefore proceeded by a stepwise strategy, analysing smaller complexes first (see the RNA fragments used in fig. 1), and then using the assignments and data collected on these for the larger assemblies. Selective ^{15}N labelling of primer tRNA$^{Lys}_3$ within the complex also allowed editing of signal from the other partners and the use of transverse relaxation optimised spectroscopy (TROSY) improved the resolution on the large complexes.

Assignment of the tRNA$^{Lys}_3$/PBS complex

In order to be able to follow by NMR the appearance of the complex between the primer tRNA and the PBS, the first step was to assign the imino resonances of the RNA duplex corresponding to the PBS annealed to the acceptor and T arms of tRNA$^{Lys}_3$ (fig. 2a). For this purpose, we used the tRNA$^{Lys}_3$/PBS complex annealed in the presence of NC for 48 hours at 35°C, conditions which yield a unique, well-defined assembly (see below). In the annealed complex, most of the resonances corresponding to the native tRNA$^{Lys}_3$ fold have disappeared, including those specific of the tertiary structure, *i.e.* the T54-A58 reverse-Hoogsteen pair, the G18-Ψ55 interaction or the base triple C13-G22-m^{7}G46. These are replaced by a new set of peaks (fig. 2) which were assigned as follows: N1 of G and and N3 of U can be easily distinguished by their ^{15}N chemical shifts (fig. 2b,c). As the PBS is not ^{15}N-labelled, only the imino groups of tRNA$^{Lys}_3$ are observable in the TROSY experiments whereas imino signals of both tRNA$^{Lys}_3$ and PBS are observable in NOESY experiments. U66 and U67 were readily assigned as two NOE-connected ^{15}N-labelled uridines. U67 gives an NOE with an unlabelled guanine that therefore belongs to the PBS (G187), whereas U66 is connected to the ^{15}N-labelled guanine G65. This pattern of connections can only exist in the context of tRNA$^{Lys}_3$/PBS complex. From G65, the assignment walk down to G194 is unambiguous (fig.

2c), on the other side, G187 is weakly connected to G69 at 12.23 ppm (not shown). From G69, we could easily assign G70, G71 (^{15}N-labelled) and G183 (unlabelled). At this step, the only peaks that remained unassigned were a non-labelled G at 13.25 ppm, connected to a ^{15}N-labelled G at 12.96 ppm and to a non-labelled G at 12.37 ppm. This pattern corresponds to G181 (13.25 ppm), G73 (12.96ppm) and G180 at 12.37 ppm. U60 and G61 could not be assigned, but possibly correspond to some of the weak remaining G and U peaks seen in TROSY experiment. The assignments of most of the tRNA$^{Lys}_3$/PBS complex imino protons thus obtained were crucial as they demonstrate the existence of a fully annealed complex. The corresponding peaks were then used as a signature of the presence of the complex in the various NMR samples.

The heat-annealed tRNA$^{Lys}_3$/PBS complex

In order to characterize the dynamics and the limiting steps of the annealing process, the behaviour of a stoichiometric mixture of tRNA$^{Lys}_3$ and the 18 nucleotide PBS was analysed first, as a function of temperature. The two RNA were mixed on ice and the sample temperature was gradually increased (see Materials and Methods). The first significant event occurs at 15°C, as weak peaks corresponding to U66 and U67 imino protons in the complex context start appearing on the NOESY experiments (not shown) and on the TROSY experiment where additionally the peak corresponding to U64 in the tRNA$^{Lys}_3$/PBS complex can also be seen (fig. 3a). The lower part of the tRNA acceptor stem, which contains U66 and U67 thus appear to be the starting point of the formation of the tRNA$^{Lys}_3$/PBS complex. It is also possible that the unpaired GCCA 3'-terminus anneals with the PBS at this early stage, but this could not be unambiguously observed on our data, as only G73 carries a ^{15}N-attached imino proton, and its resonance overlaps with that of G65 in the free tRNA.

When the temperature was increased to 25°C, all resonances corresponding to the annealed acceptor stem in the tRNA$^{Lys}_3$/PBS complex could be observed and increased in intensity, but

the resonances corresponding to free $tRNA^{Lys}_3$ remained clearly observable, indicating that both species co-exist in the sample in comparable amounts (figure 4). The ratio between the two species did not vary over several days, indicating that this situation corresponded to an equilibrium.

A NOESY experiment recorded on the sample did not reveal any exchange peaks, but interestingly, all NOE crosspeaks originating from the acceptor stem imino protons had vanished in this experiment, including the strong G6-U67 intra base pair contact (not shown), whereas all NOE connections could still be observed for the D and T stem imino protons. This indicates that the acceptor stem undergoes a faster conformational exchange that broadens the corresponding peaks down to the noise level. When the sample was heated above 40°C and up to 60°C, the ratio of the peaks corresponding to the free $tRNA^{Lys}_3$ *vs.* those of the $tRNA^{Lys}_3$/PBS complex decreased. However, the resonances corresponding to the tertiary interactions involving T54 and Ψ55 remained observable. Heating to 80°C was required to obtain a complete disappearance of these peaks and a full conversion to the annealed complex form, as obtained above with the NC (fig. 3b-c).

Gradually heating the complex however resulted in a mixture of two forms of the annealed complex, with conformational heterogeneity at the level of U66 and U67 (peaks labelled a or b on fig. 3). Form a appears before form b upon heating. This heterogeneity is not observed when the sample is quickly heated to 50°C or when it is annealed in the presence of NC where only form a is present. Depending upon the heating protocol, some structural heterogeneity can therefore be observed. The spectra of the final forms obtained by temperature jump or NC annealing are extremely similar, confirming previous studies using chemical and enzymatic probing (24).

Nucleocapsid mediated annealing of the PBS

The formation of the initiation complex in the presence of the nucleocapsid protein was monitored in a similar fashion. Figure 5a shows the superimposition of TROSY experiments of a sample containing $tRNA^{Lys}_3$ with 2 equivalents of NCp7(1-55) protein before and after adding a stoichiometric amount of the PBS. Over the time course of the experiments, approximately 10 hours, we observed the appearance of imino resonances characteristic of the duplex formed between the $tRNA^{Lys}_3$ and the PBS. However, only a fraction of the primer tRNA molecules present in the sample were annealed to the PBS and all the resonances of unannealed $tRNA^{Lys}_3$ imino groups are still observable (fig. 5a). Interestingly, approximately 24 hours later, only the resonances of the complex remained visible, with the exception of traces of T54 and Ψ55 peaks (fig. 5b). Therefore, at 15°C, NC already mediates the slow formation of the initiation complex, even at a low ratio of protein per nucleotide. Heating the sample to 35°C resulted in the disappearance of the T54 and Ψ55 resonances, even after cooling the sample back to 15°C (see fig. 2b). Thus, at physiological temperatures, the NC allows a complete annealing of the PBS to the primer tRNA, by disrupting the stable TΨC/D loop interactions.

It has recently been reported that poly-L-lysine can substitute for the NC protein in annealing $tRNA^{Lys}_3$ to the PBS (17), suggesting that this reaction could be dominated by electrostatic effects. The formation of the $tRNA^{Lys}_3$/PBS complex in the presence of poly-L-lysine was thus investigated, under conditions identical to those used for the NC-mediated annealing. As for the NC, poly-L-lysine promoted rapid strand invasion of the tRNA by the PBS RNA fragment (fig. 6). At 35°C, all resonances corresponding to the acceptor and T-stems had disappeared (not shown). However, poly-L-lysine failed to unfold the tertiary interactions within the TΨC loop, as the Ψ55 and T54 resonances remained observable (see the T54 resonance in fig. 6b). Hence, the complex formed in the presence of poly-L-lysine is

structurally different from that obtained with NC, as the tRNA structure is incompletely melted.

Interaction between the A-rich loop and the $tRNA^{Lys}_3$.

In addition to the helix formed by the PBS and the 3' end of the primer, the viral RNA contacts the anticodon base of $tRNA^{Lys}_3$ via an A-rich loop located immediately upstream from the PBS (fig. 1). Post-transcriptional modification of nucleotides within the tRNA anticodon have been shown to be crucial for this loop-loop interaction (9,21). The labelled tRNA used in these studies was expressed *in vivo* and thus contains several of these modifications, including 2-thiolation of the uridine in position 34 (s^2U34) and threonyl-carbamoyl adenosine in position 37 (t^6A37) (18). Such a modified recombinant tRNA was shown to induce reverse transcription pauses at position +16 (18), a hallmark of the A-rich loop/anticodon interaction also observed with natural $tRNA^{Lys}_3$, but not with naked transcripts (25). We thus tried to find direct evidence for this interaction in samples containing both the ^{15}N-labelled primer $tRNA^{Lys}_3$ and an RNA oligonucleotide corresponding to the A-rich stem-loop (fig. 1). In TROSY experiments, the additional A-U pairs predicted to mediate this interaction should become visible, as the corresponding imino protons are carried by the anticodon loop uridines of the tRNA, which is ^{15}N-labelled. Whatever the conditions tested: free $tRNA^{Lys}_3$, $tRNA^{Lys}_3$ annealed to the PBS, either with or without NC, and with or without Mg^{2+}, we failed to detect any additional convincing uridine crosspeak in these experiments, even at low temperature (5°C). However, we consistently observed a strong reduction or even a complete disappearance of the NH peak corresponding to the threonyl- moiety of t^6A37 (not shown), which is located immediately next to these uridines within the anticodon (see fig. 1). It should be emphasized that this resonance does not correspond to an imino group, but to an amide group which is much less sensitive to chemical exchange with the solvent. It is in particular not sensitive to the addition of salt, NC or Mg^{2+} or to temperature. Hence, its

disappearance probably reflects a broadening resulting from conformational exchange at the level of the anticodon loop. It is therefore likely that this loop-loop interaction does take place within our NMR sample, but that it is weak and that the lifetime of the intermolecular A-U base pairs are too short to be directly detected (*i.e.* ≤ 1 ms).

Interaction with the whole PBS region of the viral RNA

Since $tRNA^{Lys}_3$ makes extended interactions with the PBS, it was of interest to investigate its interaction with a larger RNA fragment encompassing both the PBS and the A-rich loop. The larger 65 nucleotide fragment shown in fig. 1 was used, as it recapitulates most of the structured regions present in the 2D structure of the reverse transcription initiation complex reported by Isel *et al* (11), based on footprinting data. The results obtained with this larger RNA parallel those obtained with the smaller PBS fragment, with progressive strand invasion resulting in appearance of a new set of peaks. As the annealing proceeds, the linewidths of peaks in standard proton experiments (1D and 2D NOESY) became very broad, in keeping with the large size of the complex (141 nucleotides, *i.e.* ~50 Kd), however the heteronuclear TROSY experiments remained sufficiently well resolved for analysis (see fig. 7). The resonances corresponding to the annealed form appear at the same position as those observed for the minimal PBS fragment (compare fig. 7 to fig. 4). This indicates that the mechanism of annealing and the final structure of the tRNA are similar. There are however two major specific features which are worth pointing out: (i) Compared to the smaller PBS fragment, slightly higher temperatures (~5 to 10°C) are required to obtain the same level of melting of the tRNA, presumably because of an increased steric hindrance or because the viral RNA fragment folds by itself into a more stable structure. (ii) The t^6A37 resonance disappears immediately upon mixing the two RNAs, even before heating the sample. Thus, the A-loop/anticodon interaction seems to take place also with this larger viral RNA fragment. Furthermore, this event precedes the melting of the tRNA structure.

DISCUSSION

In the present work, we performed a detailed NMR analysis of the annealing between human $tRNA^{Lys}_3$, the primer of HIV-1 reverse transcription and the primer binding site on the viral genome. To our knowledge, it is one the first detailed mechanistic analysis of an RNA molecular rearrangement mechanism, at the base pair level. It allows us to decompose the annealing process into a defined sequence of events and to delineate the role of the viral nucleocapsid protein at different essential stages.

The melting process appears to start at the bottom of the acceptor and TΨC-stem, close to the four-way junction of the tRNA cloverleaf structure. This may appear counter-intuitive at first but becomes less surprising if one considers $tRNA^{Lys}_3$ nucleotide sequence. The tip of the acceptor- and TΨC-stem are very G/C-rich and thus are expected to be very stable. Furthermore, the TΨC-loop of the tRNA is strongly stabilised by an extended network of tertiary interactions. These two parts of the structure probably act as structural "locks" on the $tRNA^{Lys}_3$ fold. The cloverleaf junction, on the other hand, is closed by weaker pairs: A7-U66, G6-U67 and A50-U64, which correspond to the first nucleotides that switch to the annealed form at 15°C (fig. 3a). This is thus a "weak spot" through which the viral RNA can invade the primer structure, furthermore, we previously showed that NC further destabilises this region of the tRNA structure (16). Additionnally, PBS binding to the unpaired 3' end of the tRNA, which does not involve melting of the latter, could provide either help in maintaining the two RNAs in close proximity and/or provide a second route to acceptor stem invasion.

Interestingly, at close to physiological temperatures, a large part of the annealing takes place spontaneously in the absence of NC (fig. 3b and 4), although there is an equilibrium between the free and annealed forms that is progressively shifted toward the annealed form as temperature increases. Annealing then proceeds along the tRNA acceptor stem, as evidenced by the disappearance of the corresponding NOE crosspeaks at 25°C. On the other hand, the

TΨC loop, which is closed by 3 consecutive G-C and tightened by tertiary interactions with the D-loop, resists melting and strand invasion up to 60°C. This is in keeping with reports that heat-annealing at 60°C results in less efficient reverse transcription priming than heat-annealing at 85°C (26).

In this context, the role of the nucleocapsid in the annealing process appears to be twofold: (i) It greatly facilitates strand exchange at low temperature, thus favouring complete hybridisation of the acceptor stem under physiological conditions. This is consistent with previous data which show that HIV-1 NC is capable of enhancing the annealing of nucleic acid strands by as much as four orders of magnitude (27). (ii) It permits the complete melting of the TΨC "lock", required to obtain a full annealing. The first of these two activities can be mimicked by poly-L-lysine, suggesting that the energy barrier in this process is mostly electrostatic and thus can be effectively countered by a polycationic polymer. This observation is in agreement with several reports that indicate that the N- and C-terminal basic extensions of the nucleocapsid protein play a dominant role in this strand exchange process (17,20). However, the second TΨC "unlocking" activity is nucleocapsid-specific and cannot be reproduced using poly-L-lysine. This suggests that the zinc knuckles could alternatively play an important part in this essential step.

Interestingly, in the currently available structures of NC complexed with nucleic acids (28-30), most of the specific contacts between the protein and the nucleic acid occur via the zinc knuckles. Furthermore, the NC protein interacts preferentially with the acceptor stem and the TΨC loop of both human $tRNA^{Lys}_3$ (31) and yeast $tRNA^{Phe}$ (32) and this recognition is dependent upon the NC zinc knuckles structure (31). In addition, in the presence of a shortened NC containing only the zinc knuckles, NCp7(12-53), the apparent lifetime of the reverse-Hoogsteen T54-A58 pair within the TΨC loop is reduced by a factor of 2.5 (16), confirming the role of these knuckles in altering the stability of this part of the tRNA. It is

thus reasonable to assume that the nucleocapsid protein interacts with the tip of the TΨC stem/loop using a similar binding mode as that previously reported for two other stem-loops within the viral RNA, SL2 and the SL3 encapsidation signal (29,30), where the knuckles make key contacts with the loop. Such a binding could have two non-exclusive effects on the $tRNA^{Lys}_3$ structure: (i) it could directly destabilize the TΨC loop within the tertiary structure of the tRNA. Several of the important aromatic residues within the knuckles could insert themselves between the loop nucleotides, acting as opening wedges, as previously described (28). (ii) By binding specifically to an open form of the TΨC loop, it could displace the equilibrium toward the unfolded form of $tRNA^{Lys}_3$. The sequence of events that results from this model of the annealing mechanism is schematized in fig. 8.

The role of the A-rich loop/anticodon interaction in the annealing process is less clear. Our data shows that this interaction is weak and apparently independent of the hybridisation between the PBS sequence and the 3' end of the primer tRNA. It can precede the annealing, which is not surprising, given that this loop-loop interaction does not require any structural rearrangement in either of the two RNA partners. It could possibly serve two non-exclusive purposes: (i) it could act as a guide interaction, ensuring a pre-positioning of the two partners before annealing. (ii) it could provide an additional structural specificity element for recognition by the reverse transcriptase.

The HIV system studied here involves large RNA fragments by NMR standards, up to 141 nucleotides. The combined used of selective labelling and TROSY nevertheless allowed us to obtain high quality data. This shws that this technical approach is viable and therefore applicable to other RNAs, thereby opening the way to dynamic studies of other systems, such as those involved other retroviral replication events or in spliceosome function.

Up to now, most studies focussing on RNA refolding events have addressed the questions of what were the factors required for annealing (in terms of nucleotide sequence, chaperones,

biochemical conditions...) and what were the folds of the intial and final states. None of these earlier works addresses the question of the mechanism, which is how annealing actually proceeds. The present work provides such dynamical information at the molecular level, by dissecting the RNA rearrangement process into a discrete sequence of events. This provides key information in the context of retroviral replication, actually merging a large number of previous data into a rational framework. For instance, our model reconciles apparently conflicting data suggesting on one hand that annealing can be performed by NC variants containg only the basic N- and C-terminal extensions and lacking the zinc knuckles (17,20), and on the other hand that intact zinc knuckles are required for forming initiation proficient complexes (26). Our results show that annealing is indeed promoted by polycationic molecules such as poly-lysine, which mimics the knuckle-deleted NC, but that the complex formed is structurally different from that obtained with the native NC. In the former case, the persistence of some tertiary structure in the tRNA could interfere with the recruitment or placement of the reverse transcriptase. The identification of these partially annealed states and rate-limiting steps also offers another interesting opportunity : acting on these intermediates could provide to a new means to interfere with a very early step in the AIDS virus replication.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr. Nelly Morellet for many constructive discussions. This work was supported in part by grants from ANRS, “Ensemble contre le SIDA” and by the “Bonus Qualité Recherche” from the Université René Descartes - Paris 5.

REFERENCES

1. Staley, J. P., and Guthrie, C. (1999) *Mol. Cell* **3**, 55-64.
2. Raghunathan, P. L., and Guthrie, C. (1998) *Science* **279**, 857-860.
3. Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Grüninger-Leitch, F., Barré-Sinoussi, F., Le Grice, S. F. J., and Darlix, J. L. (1989) *EMBO J.* **8**, 3279-3285
4. Rist, M. J., and Marino, J. P. (2002) *Biochemistry* **41**, 14762-14770
5. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mely, Y. (2002) *J. Mol. Biol.* **317**, 385-399
6. Jiang, M., Mak, J., Ladha, A., Cohen, E., Klein, M., Rovinski, B., and Kleiman, L. (1993) *J. Virol.* **67**, 3246-3253
7. Beerens, N., Groot, F., and Berkhout, B. (2000) *Nucleic Acids Res.* **28**, 4130-4137.
8. Beerens, N., and Berkhout, B. (2000) *J. Biol. Chem.* **275**, 15474-15481.
9. Isel, C., Marquet, R., Keith, G., Ehresmann, C., and Ehresmann, B. (1993) *J. Biol. Chem.* **268**, 25269-25272
10. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B., and Marquet, R. (1995) *J. Mol. Biol.* **247**, 236-250
11. Isel, C., Westhof, E., Massire, C., Le Grice, S. F., Ehresmann, B., Ehresmann, C., and Marquet, R. (1999) *Embo J* **18**, 1038-1048.
12. Zhang, Z., Kang, S. M., and Morrow, C. D. (1998) *AIDS Res. Hum. Retroviruses* **14**, 979-988.
13. Tsuchihashi, Z., and Brown, P. O. (1994) *J. Virol.* **68**, 5863-5870
14. Khan, R., and Giedroc, D. P. (1992) *J. Biol. Chem.* **267**, 6689-6695
15. Gregoire, C. J., Gautheret, D., and Loret, E. P. (1997) *J. Biol. Chem.* **272**, 25143-25148
16. Tisne, C., Roques, B. P., and Dardel, F. (2001) *J. Mol. Biol.* **306**, 443-454.
17. Hargittai, M. R., Mangla, A. T., Gorelick, R. J., and Musier-Forsyth, K. (2001) *J. Mol. Biol.* **312**, 985-997
18. Tisne, C., Rigourd, M., Marquet, R., Ehresmann, C., and Dardel, F. (2000) *RNA* **6**, 1403-1412
19. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M. C., Darlix, J. L., and Roques, B. P. (1991) *Biochem. Biophys. Res. Commun.* **180**, 1010-1018

20. de Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M. C., Roques, B., and Darlix, J. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6472-6476
21. Bajji, A. C., Sundaram, M., Myszka, D. G., and Davis, D. R. (2002) *J. Am. Chem. Soc.* **124**, 14302-14303
22. Plateau, P., and Guéron, M. (1982) *J. Am. Chem. Soc.* **104**, 7310-7311
23. Weigelt, J. (1998) *J. Am. Chem. Soc.* **120**, 10778-10779
24. Brule, F., Marquet, R., Rong, L., Wainberg, M. A., Roques, B. P., Le Grice, S. F., Ehresmann, B., and Ehresmann, C. (2002) *RNA* **8**, 8-15
25. Isel, C., Lanchy, J. M., Le Grice, S. F. J., Ehresmann, C., Ehresmann, B., and Marquet, R. (1996) *EMBO J.* **15**, 917-924
26. Rong, L., Liang, C., Hsu, M., Kleiman, L., Petitjean, P., de Rocquigny, H., Roques, B. P., and Wainberg, M. A. (1998) *J Virol* **72**, 9353-9358
27. Dib-Hajj, F., Khan, R., and Giedroc, D. P. (1993) *Protein Sci* **2**, 231-243
28. Morellet, N., Demene, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournie-Zaluski, M. C., and Roques, B. P. (1998) *J. Mol. Biol.* **283**, 419-434
29. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science* **279**, 384-388
30. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) *J. Mol. Biol.* **301**, 491-511
31. Tisne, C., Roques, B. P., and Dardel, F. (2003) *Biochimie* **85**, 557-561
32. Khan, R., Chang, H. O., Kaluarachchi, K., and Giedroc, D. P. (1996) *Nucleic Acids Res* **24**, 3568-3575

Legends to the Figures

Figure 1: Sequences of RNA fragments used in this study.

The sequence of $tRNA^{Lys}_3$ is shown with the base modifications carried by the recombinant tRNA used in this study (18). The three viral RNA fragments used are shown on the lower panel with the numbering of the full-length viral genome. The two smaller fragments (A-loop and PBS) are indicated by grey boxes. In addition to the boxed nucleotides, the A-rich stem/loop fragment carried an additional G as the first nucleotide, in order to close the stem.

Figure 2: Assignment of the NC-annealed $tRNA^{Lys}_3$/PBS complex.

a) Predicted secondary structure of the $tRNA^{Lys}_3$/PBS complex. b) ^{1}H-^{15}N TROSY experiment of the complex. This experiment was carried out at 15°C after annealing the sample at 35°C in the presence of two equivalents of NC. c) NOESY experiment of the same sample as in b) with a mixing time of 120 ms. Only the imino proton window is shown.

Figure 3: Progressive heat annealing of the PBS to $tRNA^{Lys}_3$.

TROSY spectra of a stoichiometric $tRNA^{Lys}_3$/PBS mixture. Only the region corresponding to the ^{15}N uridine imino groups is shown. a) Spectrum recorded at 15°C. b) Spectrum recorded at 15°C after heating the complex at 45°C. c) Spectrum recorded at 15°C after heating the complex at 80°C.

Figure 4: Partial annealing of the PBS at room temperature.

TROSY experiment recorded at 25°C of the $tRNA^{Lys}_3$/PBS mixture. Three sets of peaks are labelled: (i) Red-labelled peaks correspond to the annealed $tRNA^{Lys}_3$/PBS form. (ii) Green-labelled peaks: "stable" free tRNA peaks. NOE connectivities from these imino protons can still be observed in a 2D NOESY experiment recorded at 25°C with a 120 ms mixing time. (iii) Black-labelled peaks: "unstable" free tRNA peaks. NOE connectivities from these imino protons are broadened down to baseline in a 2D NOESY, indicating rapid exchange. Inset: The "stable" imino protons are shown in green on the 2D tRNA structure.

Figure 5: Slow NC-mediated annealing at 15°C.

a) Superimposition of two TROSY experiments recorded at 15°C showing the imino groups of $tRNA^{Lys}_3$ with two equivalents of NCp7(1-55) (in black) or with two equivalents of NCp7(1-55) and the PBS (in red). b) TROSY experiment recorded at 15°C on the same the $tRNA^{Lys}_3$/PBS/NC mixture, 24 hours later. The dashed lines connect the imino peaks corresponding to the annealed form of the $tRNA^{Lys}_3$/PBS complex.

Figure 6: Poly-L-lysine annealing.

Comparison of the ^{15}N uridine imino region of the TROSY spectra recorded on the $tRNA^{Lys}_3$/PBS complex at 35°C a) in the presence of two equivalents of NCp7(1-55), b) in the presence of two equivalents of Poly-L-Lysine.

Figure 7: Annealing of the large (65 nt) PBS fragment onto $tRNA^{Lys}_3$. TROSY experiment recorded at 25°C of a stoichiometric $tRNA^{Lys}_3$/viral RNA mixture. Red: peaks corresponding to the annealed form. Black: peaks corresponding to the free tRNA.

Figure 8: Model for the mechanism of the NC-mediated annealing of the primer tRNA to the PBS.

Step 1: the PBS (red) invades the weak bases at the four-way junction of the cloverleaf of $tRNA^{Lys}_3$ (black). Step 2: annealing proceeds along the acceptor stem, facilitated by the basic extensions of the NC (green). Step 3: The nucleocapsid zinc knuckles form specific interactions with the TΨC loop which favour the opening of the tRNA tertiary structure. Step 4: complete annealing.

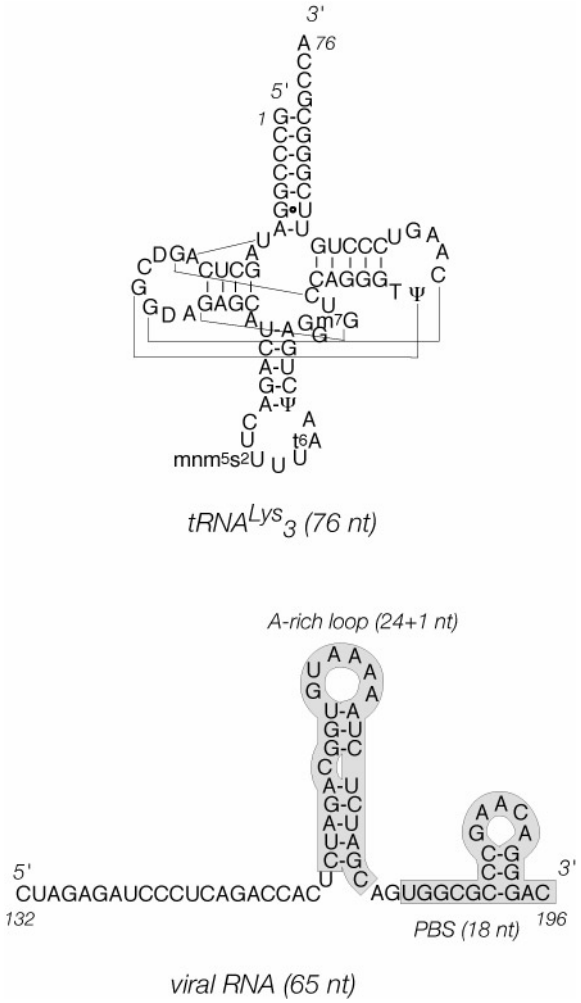
3'
76
5'
1
tRNA^Lys_3 (76 nt)
A-rich loop (24+1 nt)
5'
CUAGAGAUCCCUCAGACCAC
132
AGUGGCGC-GAC
PBS (18 nt)
196
3'
viral RNA (65 nt)

figure 1

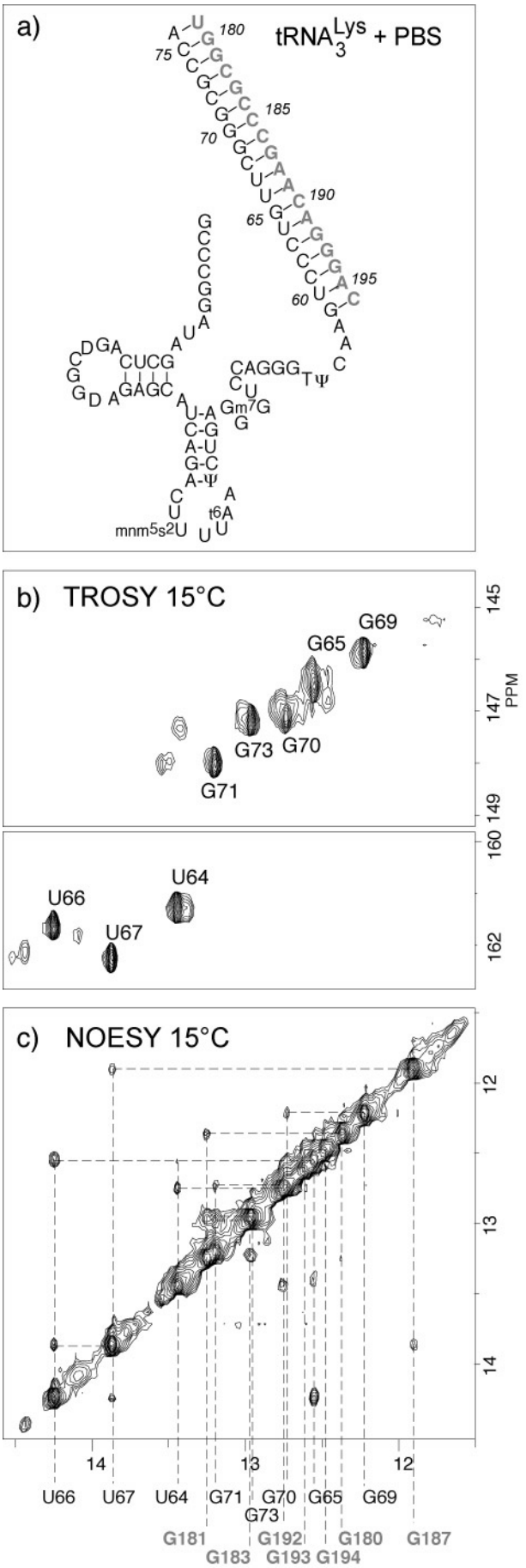
a)
tRNA3Lys + PBS
b) TROSY 15°C
G69
G65
G73
G70
G71
U64
U66
U67
PPM
c) NOESY 15°C
U66 U67 U64 G71 G70 G65 G69
G73
G181 G192 G180 G187
G183 G193 G194

figure 2

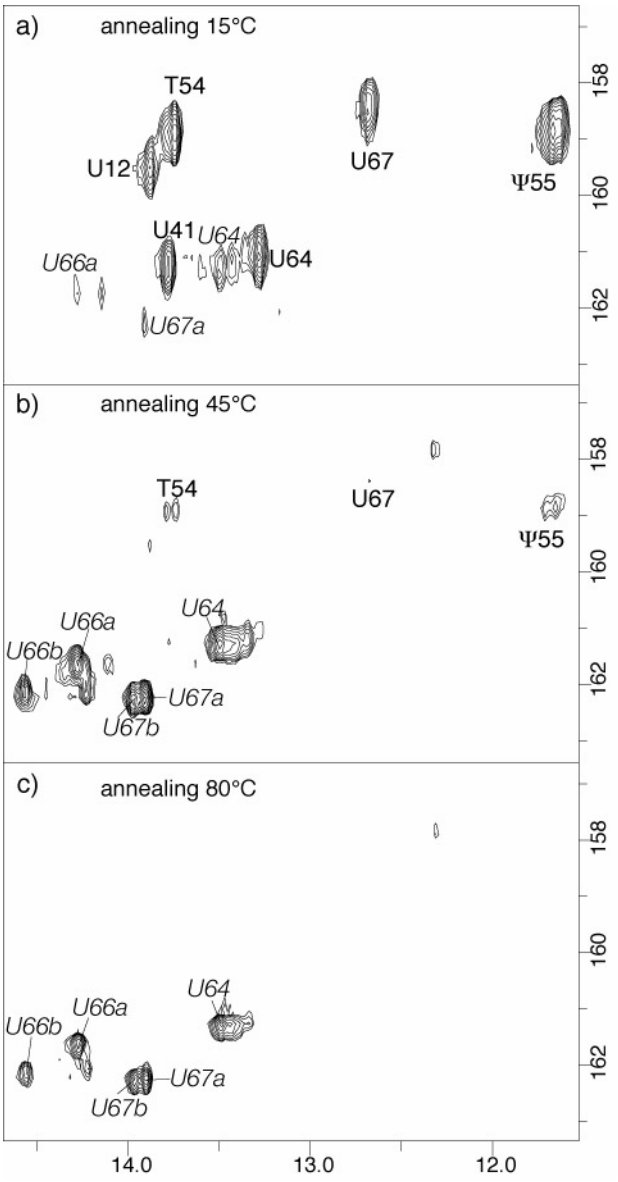
a)
annealing 15°C
T54
U12
U67
Ψ55
U41
U64
U64
U66a
U67a
b)
annealing 45°C
T54
U67
Ψ55
U64
U66a
U66b
U67a
U67b
c)
annealing 80°C
U64
U66a
U66b
U67a
U67b
158
160
162
14.0
13.0
12.0

figure 3

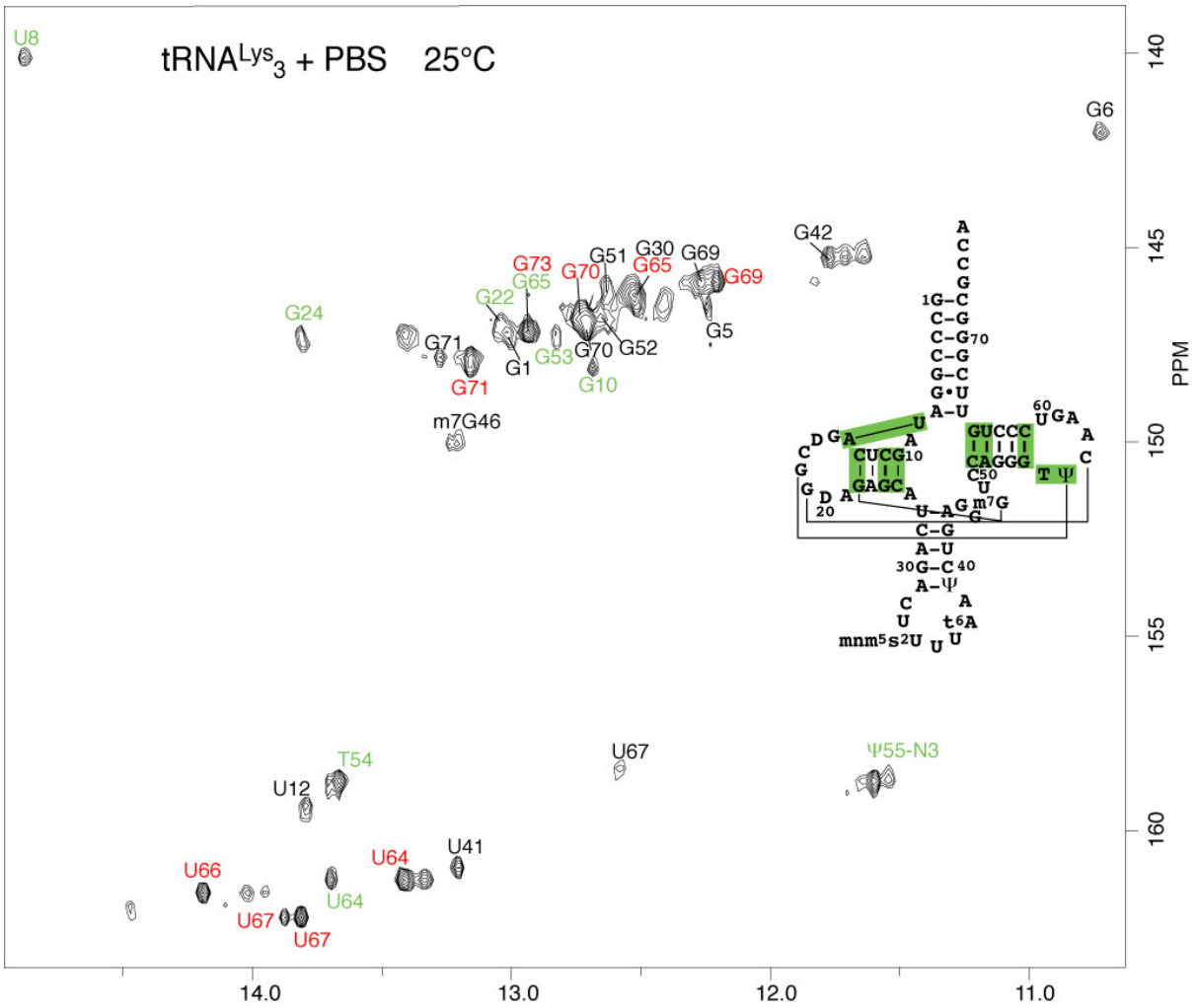
tRNA$^{Lys}_3$ + PBS 25°C
U8
G6
G42
G24
G73
G65
G70
G51
G30
G69
G65
G69
G22
G5
G71
G53
G70
G52
G1
G71
G10
m7G46
T54
U12
U67
Ψ55-N3
U41
U64
U66
U64
U67
U67
140
145
150
155
160
PPM
14.0
13.0
12.0
11.0

figure 4

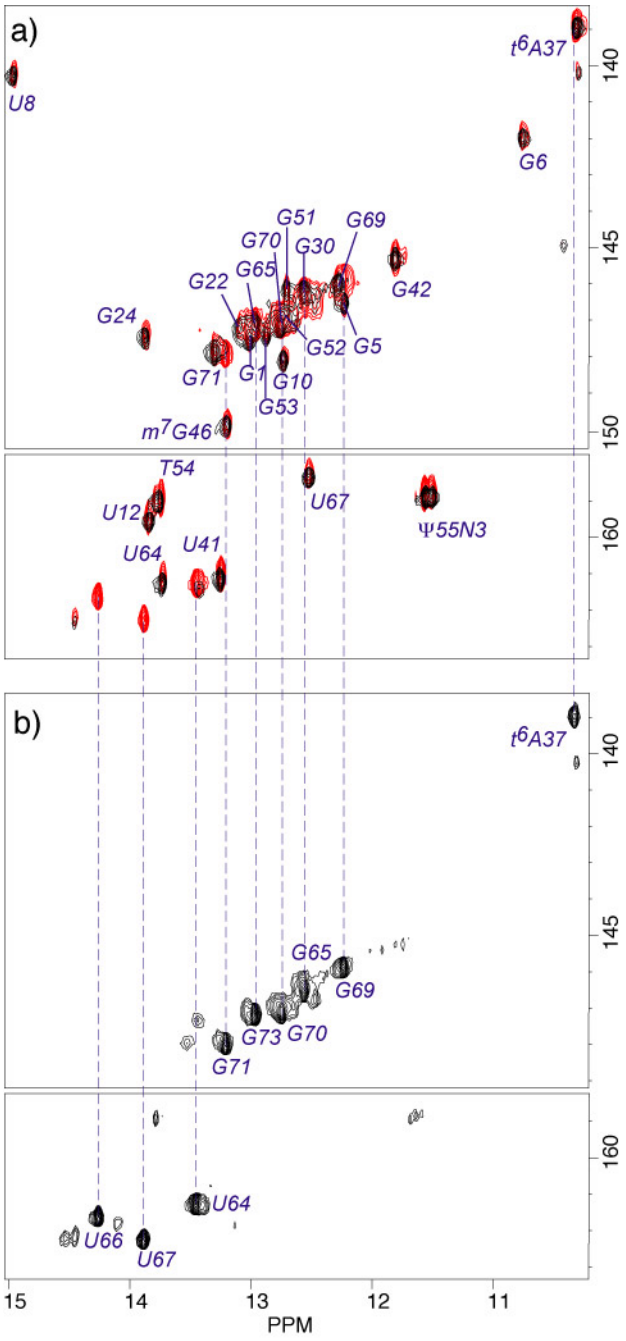
a)
U8
t6A37
G6
G51
G69
G70
G30
G65
G22
G42
G24
G52
G5
G71
G1
G10
G53
m7G46
T54
U67
U12
Ψ55N3
U64
U41
b)
t6A37
G65
G69
G73
G70
G71
U64
U66
U67
140
145
150
160
160
15
14
13
12
11
PPM

figure 5

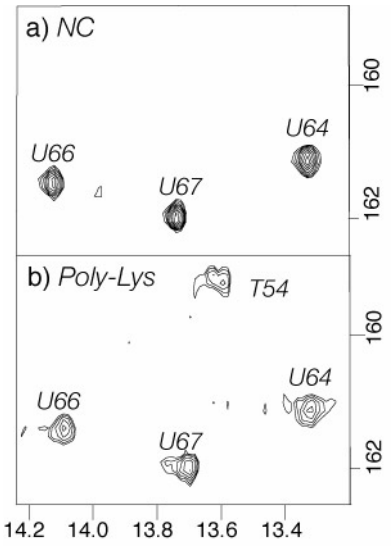
a) NC
U66
U64
U67
160
162
b) Poly-Lys
T54
U64
U66
U67
160
162
14.2
14.0
13.8
13.6
13.4

figure 6

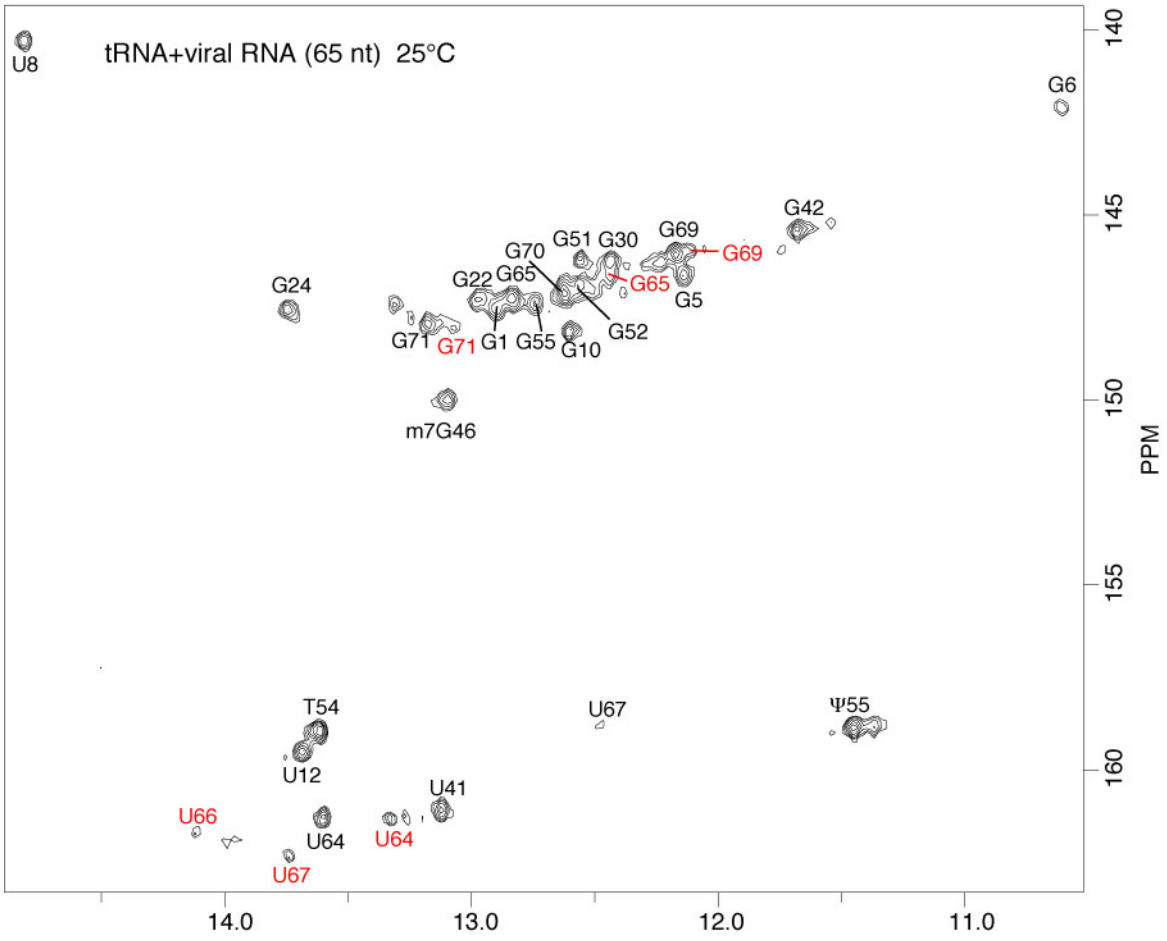
tRNA+viral RNA (65 nt) 25°C
U8
G6
G42
G69
G69
G51 G30
G70
G65
G22
G65
G5
G24
G71
G71
G1
G55
G10
G52
m7G46
T54
U67
Ψ55
U12
U41
U66
U64
U64
U67
140
145
150
155
160
PPM
14.0
13.0
12.0
11.0

figure 7

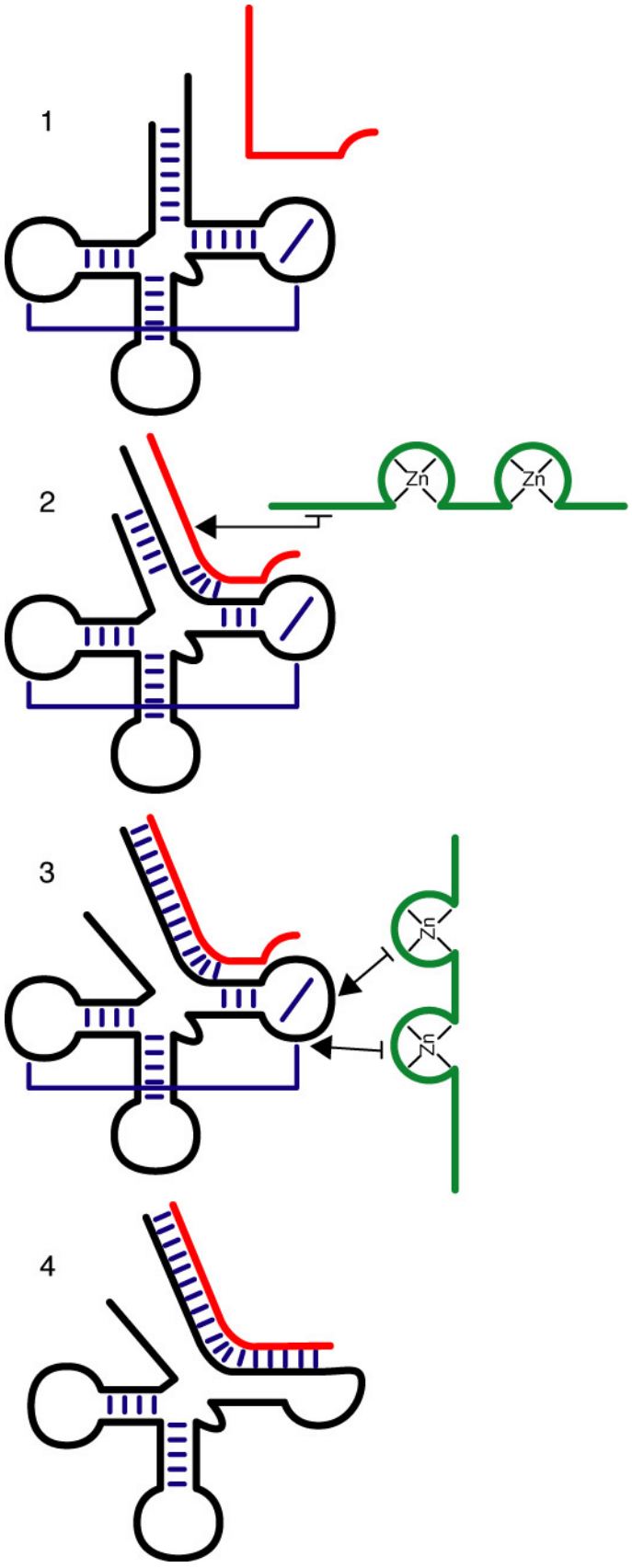

figure 8