LETTER

doi:10.1038/nature13709

# A structure–based mechanism for tRNA and retroviral RNA remodelling during primer annealing

Sarah B. Miller[1]†*, F. Zehra Yildiz[1]*, Jennifer A. Lo[1], Bo Wang[1] & Victoria M. D'Souza[1]

**To prime reverse transcription, retroviruses require annealing of a transfer RNA molecule to the U5 primer binding site (U5-PBS) region of the viral genome[1,2]. The residues essential for primer annealing are initially locked in intramolecular interactions[3–5]; hence, annealing requires the chaperone activity of the retroviral nucleocapsid (NC) protein to facilitate structural rearrangements[6]. Here we show that, unlike classical chaperones, the Moloney murine leukaemia virus NC uses a unique mechanism for remodelling: it specifically targets multiple structured regions in both the U5-PBS and $tRNA^{Pro}$ primer that otherwise sequester residues necessary for annealing. This high-specificity and high-affinity binding by NC consequently liberates these sequestered residues—which are exactly complementary—for intermolecular interactions. Furthermore, NC utilizes a step-wise, entropy-driven mechanism to trigger both residue-specific destabilization and residue-specific release. Our structures of NC bound to U5-PBS and $tRNA^{Pro}$ reveal the structure-based mechanism for retroviral primer annealing and provide insights as to how ATP-independent chaperones can target specific RNAs amidst the cellular milieu of non-target RNAs.**

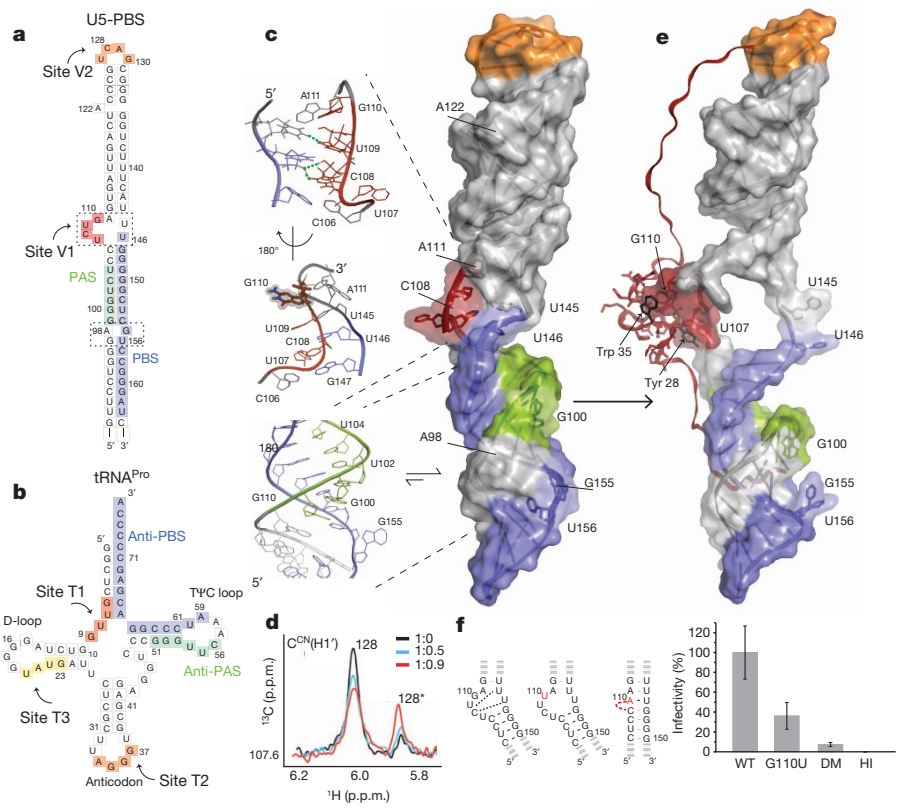



**Figure 1 | Structure, function and NC interaction of U5-PBS.** **a**, **b**, Secondary structures of U5-PBS (**a**) and $tRNA^{Pro}$ (**b**) with complementary PBS/anti-PBS and PAS/anti-PAS sequences shown in blue and green, respectively. The two internal loops in U5-PBS are boxed, and the first, second and third high-affinity NC binding site sequences are shown in red, orange and yellow, respectively. The $tRNA^{Pro}$ residue numbering reflects the canonical numbering scheme for all tRNAs and, hence, the number 17 is omitted. **c**, NMR structure of the free U5-PBS RNA. Insets: top, zoom-in view of the sequestration of $U_{145}$ and $^{+1}U_{146}$ residues by $U_{109}$ and $C_{108}$ of the $UCUG_{110}$ NC binding site; middle, $G_{110}$ aromatic ring is turned outside of the helix in a conformation poised for NC interaction; bottom, the minor, extruded conformation of $G_{155}$ is shown. PAS residues are sequestered by alternating G–C and G–U base pairs in both conformations. **d**, One-dimensional slice of a $^{1}H–^{13}C$ spectrum depicting the NC tail-mediated increase of the minor conformation (indicated by an asterisk) of the tetraloop $C_{128}$ residue. On the basis of the population distribution in the pre-bound state, we estimate the extruded conformation to have a free energy of ~1.4 kcal $mol^{-1}$ greater than the stacked conformation. **e**, Structure of NC bound to $UCUG_{110}$ via the zinc finger (the Trp 35 and Tyr 28 stacking interactions are in black) showing that the protein tails can extend to contact the $UCAG_{130}$ tetraloop and residues in and near the ($G_{97}A$, $G_{155}U$) internal loop. **f**, Left, secondary structures of the wild-type (WT), G110U and deletion mutant (DM). Right, reduction of viral infectivity is observed in mutants ($n = 6$). Error bars indicate standard deviations ($n = 6$ for both packaging and infectivity experiments). As a negative control, heat-inactivated (HI) virions were used for infection. A packaging assay was also done to confirm that the mutations do not affect genome encapsidation (see Extended Data Fig. 4g).

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. †Present address: Department of Biology, Georgetown University, Washington DC 20057, USA.
*These authors contributed equally to this work.

Retroviruses preferentially use specific host tRNAs as primers for the first step of reverse transcription; for example, human immunodeficiency virus requires $tRNA^{Lys\,3}$, while Moloney murine leukaemia virus (MLV) uses $tRNA^{Pro}$ [1,2]. Two distinct sequences in the tRNA anneal to complementary sequences in the retroviral U5-PBS domain to form the initiation complex: the 3′ end of the tRNA acceptor stem anneals to the 18-nucleotide PBS sequence[7], while a portion of the tRNA TΨC arm base-pairs with a primer activation signal (PAS)[8,9] (Fig. 1a, b). However, for primer annealing to occur, favourable intramolecular associations involving both the PBS and PAS in U5-PBS and the complementary anti-PBS and anti-PAS sequences in tRNA must first be disrupted by NC chaperone proteins. Mechanistically, all NC proteins have thus far been thought to function as classical, ATP-independent chaperones, using both their zinc-finger domain(s) and unstructured tails for this process[10,11] (Extended Data Fig. 1a). ATP-independent chaperones are known to permit RNA molecules to access higher energy conformations and then allow refolding by rapidly dissociating from the RNA during the process[12]. These transitory, low-affinity interactions generally necessitate the coating of an RNA with many molecules of chaperone[12] (Extended Data Fig. 1b). In addition, and in contrast, to the transient interactions of NC proteins[13–15], the NC zinc fingers are also capable of sequence-specific, high-affinity binding to RNA[16]. However, this mode of interaction has, until now, been thought to be used exclusively for the recognition of the genome via interaction with the Ψ-genome packaging signal during viral assembly (Supplementary Discussion 1). To gain insights into the mechanism of NC-mediated primer annealing in MLV, a prototypical retrovirus, we solved structures of both the genomic U5-PBS RNA and the $tRNA^{Pro}$ primer, both in the free form and in complex with MLV NC proteins, by NMR spectroscopy.

The free U5-PBS is a largely linear molecule capped by a structured tetraloop ($UCAG_{130}$) and contains one single-nucleotide bulge ($A_{122}$) and two internal loops (($UCUGA_{111}$, $UU_{146}$) and ($GA_{98}$, $GU_{156}$)) (Fig. 1a, c, Extended Data Figs 2, 3 and Extended Data Table 1). In the absence of NC, the ($UCUGA_{111}$, $UU_{146}$) internal loop maintains a distinct, folded configuration in which residue $C_{108}$ of the NC binding site sequesters the 5′ end of the PBS sequence ($^{+1}U_{146}$) via an intramolecular ribose zipper interaction[17,18] (superscript denotes the PBS position from 5′ to 3′) (Fig. 1c and Extended Data Fig. 3e–g). Residue $U_{109}$ of the NC binding site also base pairs with $U_{145}$, which is the first template residue read by reverse transcriptase. Continuous intramolecular base stacking interactions from $U_{144}$ to $^{+2}G_{147}$ further serve to tether the 5′ end of the PBS inside the internal loop. On the other side of the bulge, continuous nuclear Overhauser enhancements (NOEs) indicate the stacking of $C_{106}$ with $C_{108}$ and, hence, extrusion of residue $U_{107}$. Importantly, residue $G_{110}$ exhibits a *syn* glycosidic torsion angle and faces the major groove, making it poised for interaction with NC (Fig. 1c). In the NC-bound structure, the NC zinc finger binds the $UCUG_{110}$ sequence of the internal loop (dissociation constant ($K_d$) = 33 ± 3 nM; Extended Data Fig. 4a) in a mode similar to that previously described for $UCUG_{309}$ in the MLV Ψ-genome packaging signal[19–21] (see Supplementary Discussion 1 and 2). Notably, since NC-binding residues are initially involved in sequestering the first template ($U_{145}$) and the first PBS ($^{+1}U_{146}$) residues, the well-defined internal loop structure is mutually exclusive with NC binding. Thus, NC binding liberates $U_{145}$ and $^{+1}U_{146}$ for primer annealing and reverse transcription initiation (Fig. 1e).

Whereas the NC tails are not involved in binding the Ψ-packaging signal, in U5-PBS, the NC tails specifically remodel the ($GA_{98}$, $GU_{156}$) internal loop and the $UCAG_{130}$ capping tetraloop (Fig. 1e). In the absence of NC, the six PAS residues in the lower stem, $_{+1}G_{99}$ to $_{+6}U_{104}$ (subscript with plus sign denotes the PAS position from 5′ to 3′), form base pairs with PBS residues (Fig. 1c and Extended Data Fig. 2a), and hence are not available for primer annealing. The preceding ($GA_{98}$, $GU_{156}$) internal loop, however, exists in multiple conformations; in the major form, a continuous internal stacking of all residues leads to the formation of tandem $A_{98}$–$^{+10}G_{155}$ and $G_{97}$–$^{+11}U_{156}$ non-canonical base pairs, while in the minor, destabilized form, the $^{+10}G_{155}$ and $^{+11}U_{156}$ PBS residues are extra-helical (Fig. 1c and Extended Data Figs 2a, 3h–j). Similarly, the capping $UCAG_{130}$ tetraloop forms a YNMG-type structure[22] (Extended Data Fig. 3a), with the $C_{128}$ base either stacked on $A_{129}$ or, in a small population, extruded from the structure (Fig. 1d and Extended Data Fig. 3c). Interaction with the NC tails alters the equilibria between the two conformations in favour of the minor, destabilized conformations (Fig. 1d) and hence leads to release of the 10th and 11th ($^{+10}G_{155}$ and $^{+11}U_{156}$) PBS residues and the $C_{128}$ tetraloop residue. The destabilization of the latter is important for cooperative binding of a second NC to the tetraloop (Extended Data Fig. 4a; see Supplementary Discussion 3). Thus, the positively charged NC tails do not globally destabilize U5-PBS but instead specifically target residues inherently predisposed for destabilization. Furthermore, because PAS residues immediately follow the destabilized internal loop (Fig. 1a), the NC tails also specifically perturb residues $_{+1}G_{99}$, $_{+2}G_{100}$, and $_{+3}G_{101}$ (Extended Data Fig. 3k). Interestingly, both NC tails (Ala 1–Arg 17 and Arg 4–Leu 56) remain disordered, indicating that destabilization must occur via transient interactions that are nevertheless residue-specific owing to the inherent accessibility of the particular RNA residues and the orientation constraints imposed by the zinc finger binding to $UCUG_{110}$ (Fig. 1e and Extended Data Fig. 3l). In live viruses, a G110U mutant designed to liberate the $U_{145}$ and $^{+1}U_{146}$ residues and a deletion mutant designed to sequester them exhibited only

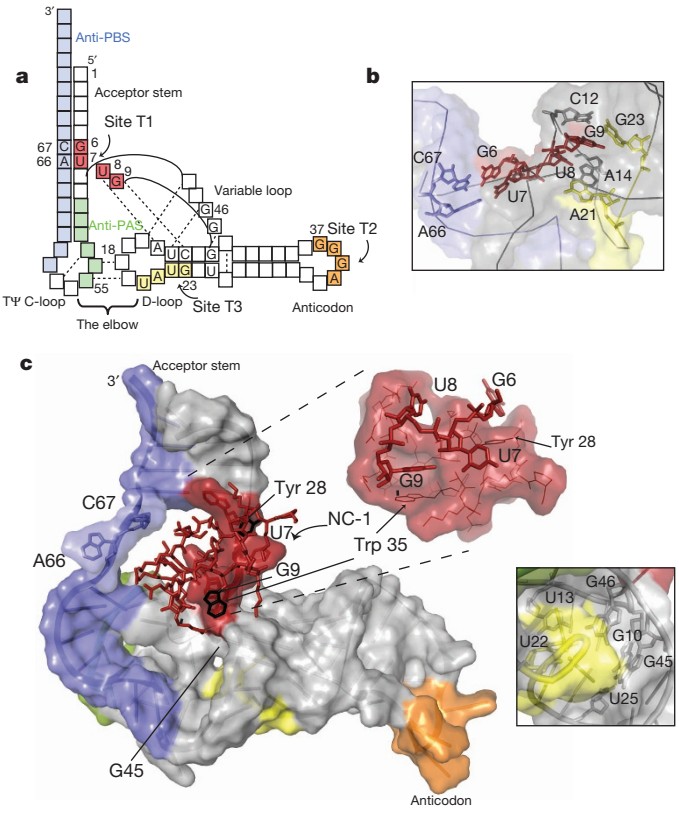



**Figure 2 | Structure and NC interaction of site T1 in $tRNA^{Pro}$.** **a**, Cartoon representation of the L-shaped $tRNA^{Pro}$ structure. Dashed lines show the long-range elbow and base triple interactions with the D-stem loop. Solid lines denote the covalent linkages. **b**, Three-dimensional model of the free $tRNA^{Pro}$ showing site T1 residues $GUUG_9$ (red) involved in intramolecular interactions. **c**, Structure of the NC–$tRNA^{Pro}$ complex. The aromatic NC residues Tyr 28 and Trp 35 are shown in black. The zinc finger interaction with the $GUUG_9$ binding site is shown in red. The tails are excluded for simplicity since they form random coils and do not specifically interact with the tRNA (Extended Data Fig. 9f). Top inset shows a close-up of the zinc finger interaction, with $G_9$ inserted into the hydrophobic pocket of the NC protein due to stacking interactions of Trp 35. Bottom inset shows the interaction of the variable loop with the core of the tRNA molecule that is maintained after interaction with NC-1.

56% and 7%, respectively, of wild-type MLV infectivity (Fig. 1f). While the severe infectivity defect of deletion mutant virions confirms the importance of high-affinity NC binding in releasing the 5′ end of the PBS, the partial defect observed for G110U virions indicates that tail-mediated interactions are also required for optimal function.

In $tRNA^{Pro}$, NC binding occurs first at $GUUG_9$ (site T1; 4 ± 2 nM) followed by the anticodon loop $AGGG_{37}$ (site T2; 13 ± 3 nM) and then the D-stem loop $UAUG_{23}$ (site T3; 834 ± 343 nM) (Extended Data Fig. 4b). Importantly, titration of MLV NC into the human immunodeficiency virus (HIV) primer, $tRNA^{Lys\,3}$, did not lead to NMR chemical shift perturbations, thus confirming the specificity of MLV NC for $tRNA^{Pro}$ (Extended Data Fig. 1f, g; for assignment strategy of $tRNA^{Pro}$ see Extended Data Figs 5–8 and Supplementary Discussion 4). The structure of the first NC bound to $tRNA^{Pro}$ shows the zinc finger making extensive contacts with the $GUUG_9$ sequence that links the acceptor stem with the D-stem loop, with $G_9$ stacking within the zinc finger pocket (Fig. 2, Extended Data Table 1 and Extended Data Fig. 9a). Importantly, before NC binding, all four $GUUG_9$ residues are involved in intramolecular interactions: $G_6$ and $U_7$ are part of the acceptor stem, and $U_8$ and $G_9$ are involved in core tertiary interactions (Fig. 2a, b and Extended Data Figs 6, 8a). Similar to U5-PBS, because these contacts are mutually exclusive with NC interactions, NC binding leads to major RNA remodelling events. First, since residues $G_6$ and $U_7$ are initially base paired with anti-PBS residues $^{-10}C_{67}$ and $^{-11}A_{66}$ (superscript with minus sign denotes the anti-PBS position from 3′ to 5′), respectively, NC binding releases these sequestered anti-PBS residues (Fig. 2c and Extended Data Fig. 9b, c). Second, because $U_8$ and $G_9$ are involved in core tertiary interactions via triple base formation with the D-stem loop ($U_8$:$A_{14}$–$A_{21}$, $G_9$:$C_{12}$–$G_{23}$) (Fig. 2a, b and Extended Data Fig. 8a), NC binding disrupts these core tertiary interactions (Extended Data Fig. 9c). As a result of this rearrangement, D-stem residues $A_{21}$ and $G_{23}$, which are part of the $UAUG_{23}$ site T3 sequence, are made partially available for the third NC binding event. Globally, while the helical arrangement between the TΨC-stem and the acceptor stem is lost, the helical stacking between the D-stem and anticodon stem is preserved (Extended Data Fig. 9d, e), as is the variable loop interaction and the TΨC-loop:D-loop interface (the 'elbow') (Fig. 2a, c and Fig. 3a).

In comparison with sites T1 and T3 (see later), there is a marked mechanistic difference in the remodelling activity of the NC that binds the second site, T2—this NC uses its tails to achieve residue-specific destabilization. After the first NC binding event, the second NC accesses the residual elbow structure by anchoring its zinc finger to the distal $AGGG_{37}$ sequence in the anticodon loop, with the $G_{37}$ base stacking inside the zinc finger (Fig. 3b and Extended Data Fig. 9g). While the elbow interaction is maintained, the NC tails specifically perturb D-loop and TΨC-loop residues $G_{16}$ and $A_{59}$, respectively (Fig. 3c, d), which are in close proximity to each other (Fig. 3b). Prior to NC binding, these residues are extruded out of their respective loops and are hence available for NC interaction. Thus, as in the interaction with U5-PBS, the NC tails do not cause global destabilization of $tRNA^{Pro}$ but instead target specific, already accessible residues for remodelling.

We also structurally characterized the third NC binding event using the $tRNA^{Pro}$–$T1_MT2_M$ construct (see Supplementary Discussion 5). Our structures show that NC zinc finger binding to $UAUG_{23}$ in the D-stem disrupts the entire helix and, because the D-stem architecture is crucial for the D-loop–TΨC interaction, eliminates the residual elbow tertiary structure (Fig. 3a, e and Extended Data Fig. 9h, i). Consequently, the interactions between D-loop residues $G_{18}$ and $G_{19}$ and TΨC-loop anti-PAS residues $_{-1}C_{56}$ and $_{-2}U_{55}$ are lost, leading to the release of these

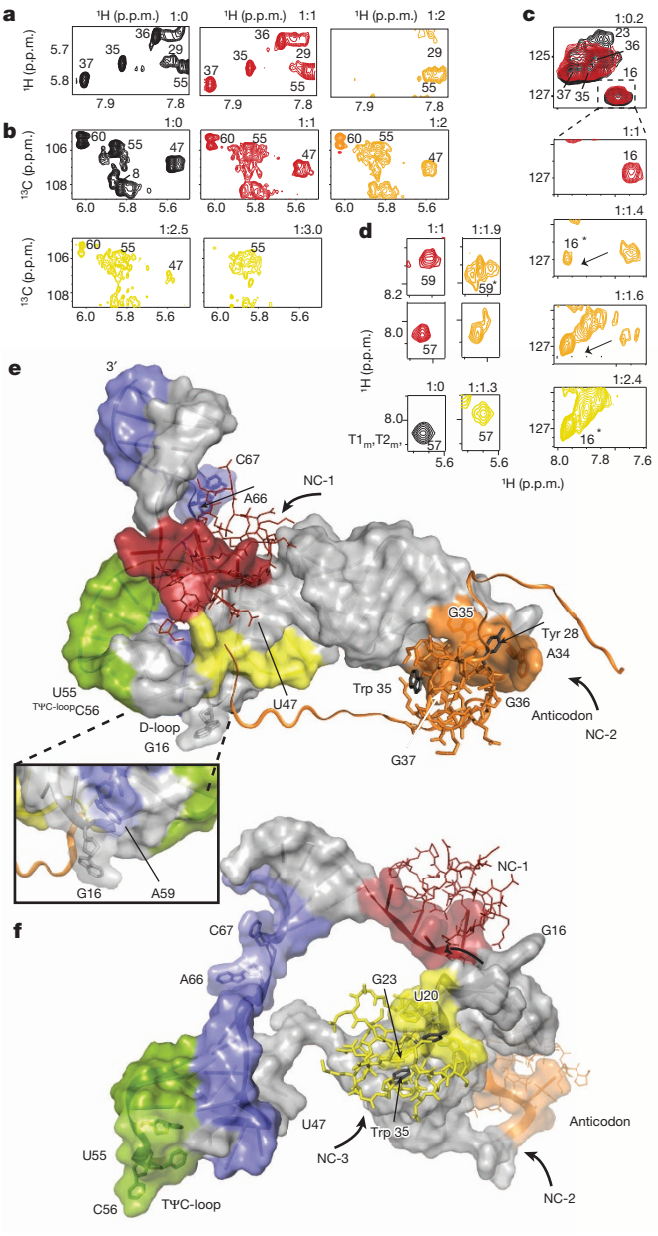



**Figure 3 | Structure and NC interaction with sites T2 and T3 in $tRNA^{Pro}$.** **a**, A portion of the two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectrum with H8/H1′ correlations for $tRNA^{Pro}$ showing the perturbation of anticodon residues only after titrations above 1.0 equivalent of NC, thus confirming the sequential binding mode. **b**, $^{1}H$–$^{13}C$ two-dimensional U-labelled heteronuclear multiple quantum coherence (HMQC) spectra showing that whereas the U8 resonance perturbation occurs upon the addition of one equivalent of NC, the perturbation of D-loop–TΨC signals occur only upon the third NC binding, thus indicating that the elbow contacts are maintained during the first two binding events. **c**, $^{1}H$–$^{13}C$ two-dimensional HMQC spectra showing selective perturbation of the extruded $G_{16}$ residue in the D-loop as evidenced by a marked chemical shift change is shown by asterisks. **d**, Regions of two-dimensional NOESY for NC complexes with $tRNA^{Pro}$ and $tRNA^{Pro}$–$T1_MT2_M$. Top and middle panels show that the protected $A_{57}$ in the TΨC loop is not perturbed, but the extruded $A_{59}$ is affected upon the second NC binding. In the $T1_MT2_M$ mutant (bottom panel), NC binding to the third site disrupts the TΨC–D-loop interaction, resulting in a chemical shift change for residue $A_{57}$. Thus, the lack of $A_{57}$ perturbation in the native $tRNA^{Pro}$ also demonstrates that the elbow region is maintained after the second NC binding. **e**, Structure of NC bound to $tRNA^{Pro}$ sites T1 and T2 via the zinc fingers. The structures show that the NC-2 protein tails can extend to contact the D-loop–TΨC-loop elbow region. Inset shows the proximity of the extruded $G_{16}$ and $A_{59}$ residues. For clarity, only the tails of the NC-2 protein are shown. **f**, Model of three NCs bound to $tRNA^{Pro}$ (based on the structure of NC bound to $T1_MT2_M$ $tRNA^{Pro}$; see Extended Data Fig. 9i), showing the loss of the elbow tertiary interactions upon binding of the third NC, which leads to the release of anti-PAS sequences. For clarity only the zinc-finger portion of the NC proteins are shown.

sequestered anti-PAS residues (subscript with minus sign denotes the anti-PAS position from 3′ to 5′) (Fig. 3d). NC binding to site T3 thus serves to dismantle the residual tRNA tertiary structure before primer annealing. Importantly, destabilization of the D- and TΨC-loop residues by the anticodon site T2 NC tails is maintained even after the elbow contacts are dismantled by the third NC binding event (Fig. 3c), prohibiting the freed TΨC-loop from forming an intrinsically stable structure[23] (see Extended Data Fig. 6) and ensuring that the released anti-PAS residues will remain accessible during primer annealing.

Our data demonstrate how MLV NC 'captures' specific portions of both the U5-PBS and $tRNA^{Pro}$ through high-affinity interactions with residues that are normally engaged in intramolecular stabilizing interactions and results in the subsequent 'release' of these sequestered residues, thereby reducing the energetic barrier for primer–template complex formation (Fig. 4). The combinations of liberated and pre-exposed residues within $tRNA^{Pro}$ and U5-PBS are exactly complementary and therefore poised for intermolecular base pairing. Furthermore, the complementarity of liberated sequences to regions that are already exposed in the counterpart RNA allows remodelling to occur with a limited number of NC molecules (Fig. 4). Indeed, the presence of four NC molecules is sufficient for formation of a functional U5-PBS–$tRNA^{Pro}$ complex (Extended Data Fig. 4c, d). Importantly, because the NC binding sites are perfectly positioned in close proximity to, but not overlapping with, the RNA-annealing sequences, subsequent dissociation of NC from the annealed complex is not required[24]. In fact, the presence of NC has been shown to be important for the elongation step of reverse transcription[25,26].

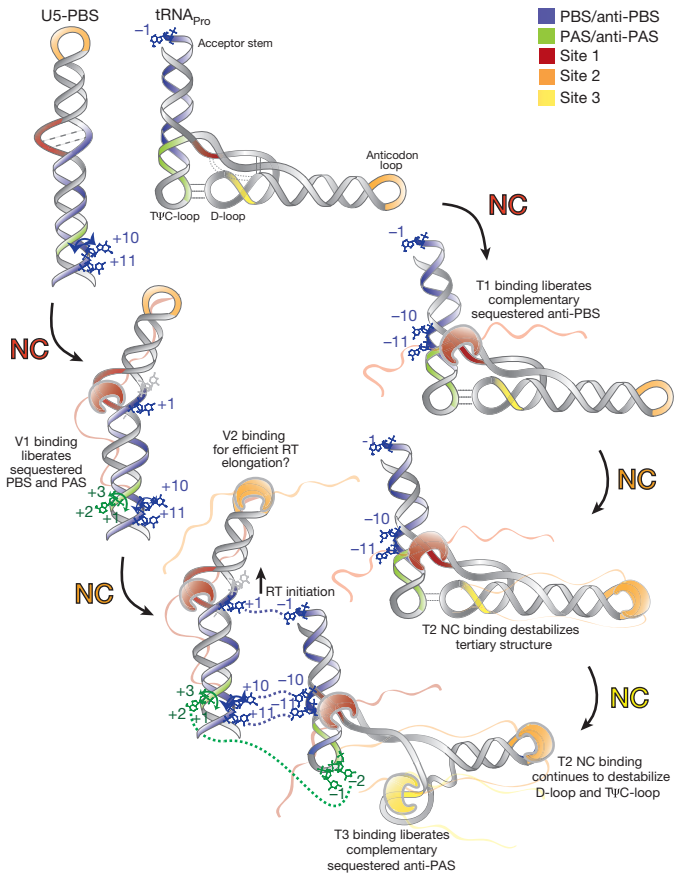



**Figure 4 | Capture-and-release mechanism for NC-mediated remodelling of MLV U5-PBS and $tRNA^{Pro}$.** A mechanistic model for NC zinc-finger- and tail-mediated remodelling via a step-wise, residue-specific release and residue-specific destabilization of sequestered PBS/PAS and anti-PBS/anti-PAS sequences in U5-PBS and $tRNA^{Pro}$, respectively. Upon NC binding to U5-PBS, the PBS residues $^{+1}U_{146}$, $^{+10}G_{155}$ and $^{+11}U_{156}$ are released, and the PAS residues $_{+1}G_{99}$, $_{+2}G_{100}$ and $_{+3}G_{101}$ are destabilized. Reciprocally, in $tRNA^{Pro}$, NC binding to site T1 frees anti-PBS residues $^{-10}C_{67}$ and $^{-11}A_{66}$ ($^{-1}A_{76}$ is already available in free $tRNA^{Pro}$), and binding to site T2 and site T3 renders the anti-PAS residues $_{-1}C_{56}$, $_{-2}U_{55}$ and $_{-3}U_{54}$ accessible for annealing. NC binding to the second site in U5-PBS may be important for primer extension by reverse transcriptase (RT) since stem loops with stable tetraloops have been shown to stall reverse transcription.

In addition to defining previously undiscovered roles for high-affinity NC binding events in the retroviral lifecycle, our study has implications for the location, timing and specificity of primer annealing (see Supplementary Discussion 6). Like MLV NC, some other RNA chaperones and remodellers bind with high affinity to their substrates; however, they typically require the input of additional energy for subsequent dissociation[27]. The MLV NC-mediated capture-and-release mechanism described here is distinct from mechanisms used by other known ATP-independent RNA chaperones[28,29]: During the capture-and-release RNA remodelling, NC uses high-affinity interactions to bind a limited number of sites with high specificity. Furthermore, unlike typical chaperones, which cause global destabilization to allow access to higher energy conformations, the mechanism of NC-mediated remodelling in primer annealing involves the formation of stable, lower energy complexes with RNA that cause strategic local destabilization of the regions important for annealing. Consistent with this, the thermodynamic analyses of all U5-PBS–NC and $tRNA^{Pro}$–NC interactions show high binding affinities with entropically driven profiles (see Supplementary Discussion 7). This entropy-driven, capture-and-release remodelling thus represents the first example, to our knowledge, of a new mechanism by which RNA chaperones can specifically select their specific targets from a sea of cellular RNAs.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 February; accepted 24 July 2014.**
**Published online 7 September 2014.**

1. Harada, F., Peters, G. G. & Dahlberg, J. E. The primer tRNA for Moloney murine leukemia virus DNA synthesis. Nucleotide sequence and aminoacylation of tRNAPro. *J. Biol. Chem.* **254,** 10979–10985 (1979).
2. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40,** 9–17 (1985).
3. Mougel, M. *et al.* Conformational analysis of the 5′ leader and the gag initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucleic Acids Res.* **21,** 4677–4684 (1993).
4. Paillart, J. C. *et al.* First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* **279,** 48397–48403 (2004).
5. Wilkinson, K. A. *et al.* High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **6,** e96 (2008).
6. Levin, J. G., Mitra, M., Mascarenhas, A. & Musier-Forsyth, K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* **7,** 754–774 (2010).
7. Cordell, B., Stavnezer, E., Friedrich, R., Bishop, J. M. & Goodman, H. M. Nucleotide sequence that binds primer for DNA synthesis to the avian sarcoma virus genome. *J. Virol.* **19,** 548–558 (1976).
8. Beerens, N., Groot, F. & Berkhout, B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.* **276,** 31247–31256 (2001).
9. Beerens, N. & Berkhout, B. Switching the *in vitro* tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA* **8,** 357–369 (2002).
10. Thomas, J. A. & Gorelick, R. J. Nucleocapsid protein function in early infection processes. *Virus Res.* **134,** 39–63 (2008).
11. Rein, A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol.* **7,** 700–705 (2010).
12. Woodson, S. A. Taming free energy landscapes with RNA chaperones. *RNA Biol.* **7,** 677–686 (2010).
13. De Rocquigny, H. *et al.* Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl Acad. Sci. USA* **89,** 6472–6476 (1992).
14. Hargittai, M. R., Mangla, A. T., Gorelick, R. J. & Musier-Forsyth, K. HIV-1 nucleocapsid protein zinc finger structures induce $tRNA^{Lys,3}$ structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **312,** 985–997 (2001).
15. Prats, A. C. *et al.* Viral RNA annealing activities of the nucleocapsid protein of Moloney murine leukemia virus are zinc independent. *Nucleic Acids Res.* **19,** 3533–3541 (1991).
16. Rein, A., Henderson, L. E. & Levin, J. G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* **23,** 297–301 (1998).
17. Tamura, M. & Holbrook, S. R. Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **320,** 455–474 (2002).

18. Nonin-Lecomte, S., Felden, B. & Dardel, F. NMR structure of the *Aquifex aeolicus* tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation. *Nucleic Acids Res.* **34,** 1847–1853 (2006).
19. D'Souza, V. & Summers, M. F. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431,** 586–590 (2004).
20. Dey, A., York, D., Smalls-Mantey, A. & Summers, M. F. Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus. *Biochemistry* **44,** 3735–3744 (2005).
21. D'Souza, V. *et al.* Identification of a high affinity nucleocapsid protein binding element within the Moloney murine leukemia virus ψ-RNA packaging signal: implications for genome recognition. *J. Mol. Biol.* **314,** 217–232 (2001).
22. Theimer, C. A., Finger, L. D. & Feigon, J. YNMG tetraloop formation by a dyskeratosis congenita mutation in human telomerase RNA. *RNA* **9,** 1446–1455 (2003).
23. de Smit, M. H. *et al.* Structural variation and functional importance of a D-loop–T-loop interaction in valine-accepting tRNA-like structures of plant viral RNAs. *Nucleic Acids Res.* **30,** 4232–4240 (2002).
24. Martin-Tumasz, S., Richie, A. C., Clos, L. J. II, Brow, D. A. & Butcher, S. E. A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. *Nucleic Acids Res.* **39,** 7837–7847 (2011).
25. Gonsky, J., Bacharach, E. & Goff, S. P. Identification of residues of the Moloney murine leukemia virus nucleocapsid critical for viral DNA synthesis *in vivo*. *J. Virol.* **75,** 2616–2626 (2001).
26. Liu, S., Harada, B. T., Miller, J. T., Le Grice, S. F. & Zhuang, X. Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nature Struct. Mol. Biol.* **17,** 1453–1460 (2010).
27. Fedorova, O., Solem, A. & Pyle, A. M. Protein-facilitated folding of group II intron ribozymes. *J. Mol. Biol.* **397,** 799–813 (2010).
28. Semrad, K. Proteins with RNA chaperone activity: a world of diverse proteins with a common task-impediment of RNA misfolding. *Biochem. Res. Int.* **2011,** 532908 (2011).
29. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **482,** 322–330 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We wish to thank the laboratory of S. Goff for the pNCS plasmid, C. Salguero for the artistic rendition of the model, as well as M. Summers and R. Gaudet for critical reading of the manuscript and the Merck Research Fellowship and Damon Runyon Cancer Research scholarship for funding.

**Author Contributions** V.M.D'S., S.B.M. and F.Z.Y. conceived and designed the experiments. J.A.L. and S.B.M. performed and analysed the isothermal titration calorimetry (ITC) experiments, V.M.D'S., S.B.M. and F.Z.Y. did the structural analysis, and B.W. and S.B.M. performed and analysed the virological experiments. S.B.M., F.Z.Y. and V.M.D'S. wrote the manuscript.

**Author Information** Coordinates and restraints for the final ensembles of the MLV U5-PBS and $tRNA^{Pro}$ structures have been deposited in the Protein Data Bank under accession numbers 2MQT, 2MQV, 2MS1 and 2MS0. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M.D'S. (dsouza@mcb.harvard.edu).

## METHODS

**Sample preparation.** The MLV NC protein was prepared as described previously[21]. The HIV NC protein was made in an analogous manner to that of the MLV NC protein; the NC sequence was amplified from pNL4-3 plasmid[30], cloned into pGEX-6p1 (GE Healthcare) vector, and the purified fusion protein was cleaved from glutathione *S*-transferase (GST) through the use of PreScission Protease. NMR experiments were used to confirm the correct folding of the proteins. The HIV and MLV U5-PBS RNA samples were also made using methods previously described[21]. $tRNA^{Pro}$ and $tRNA^{Lys\,3}$ DNA templates were constructed through annealing and ligation of oligonucleotides designed to contain a T7 promoter and two 2′-*O*-methoxy modified nucleotides at the 5′ end of the template strand to maintain 3′ end homogeneity[31]. For the U5-PBS construct used for NMR studies, a single base-pair swap (G96C: C157G) was made to ameliorate the spectral overlap problem that arose owing to five consecutive guanosine residues in a row. This swap does not change the secondary or tertiary structure of the U5-PBS domain (Extended Data Fig. 2b). The G110U mutation was structurally characterized to ensure that the mutation does not give rise to an alternate structure that sequesters the 5′ end of the PBS (data not shown).

**Isothermal titration calorimetry.** Prior to all ITC experiments, NC proteins and RNAs were exchanged into ITC buffer (25 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 10 mM Tris, pH 7.0). For each ITC experiment, reaction heats ($\mu Cal\ s^{-1}$) were measured for 1.5–2 μl titrations of 70–90 μM NC into 5 μM RNA at 30 °C using an iTC200 machine (MicroCal). Titration curves were analysed using ORIGIN (OriginLab).

**NMR data acquisition, resonance assignment and structure calculations.** For NMR experiments, RNA samples were resuspended in NMR buffer (10 mM Tris-HCl, pH 7.5, and 10 mM NaCl). NMR data for $tRNA^{Pro}$ were collected with and without the presence of 1 mM $MgCl_2$ and 100–150 mM NaCl. For NC complex studies, the samples were prepared in buffers containing 1 mM $MgCl_2$. NMR data were acquired using a Bruker 700 MHz spectrometer equipped with a cryoprobe. Spectra were recorded at 35 °C, with the exception of data for the imino region, which were also collected at 10 °C. Non-exchangeable assignments were made using two-dimensional NOESY, two-dimensional HMQC and three-dimensional HMQC-NOESY experiments using both unlabelled and nucleotide-specific labelled (G–, U–, A–, and C–$^{15}N$,$^{13}C$-labelled and C/U/A-deuterated) or protonated samples. Residual dipolar coupling (RDC) data for U5-PBS were collected using Pf1 phage ($12\ mg\ ml^{-1}$, ASLA Biotech) as has been described[32].

Initial structures were calculated as described[26] using manually assigned restraints in CYANA[28]. In contrast to all long range triple bases in $tRNA^{Pro}$, the two base pairs between the D-loop and TΨC-loop were not observable by NMR: these hydrogen bonds were modelled on the basis of our knowledge of tRNA structures. The crystal structure of $tRNA^{Phe}$ was used as a guide for soft phosphate–phosphate distance restraints within a particular stem[33]. For U5-PBS–NC and $tRNA^{Pro}$–NC complexes, the ten best CYANA models were then used for final structure calculations in AMBER similar to what has been described[34]. Specifically, the refinement included 50,000 steps, with temperature increasing from 0 K to 500 K over the first 12,500 steps, remaining at 500 K over the next 32,500 steps, and then decreasing to 0 K over the next 5,000 steps. These calculations incorporated all upper limit restraints used in CYANA but not the angle restraints. The individual structures generated were then used for tensor fitting, and the above structure calculation process was repeated with the RDC restraints (for U5-PBS) along with a final minimization that included 8,000 steps. In the final minimization step for the U5-PBS structure, loose hydrogen bond restraints used for the ribose zippers were removed. Several rounds of *in vacuo* AMBER calculations were done until the distance violations were less than 0.5 Å. Molecular images were generated with PyMOL (http://www.pymol.org).

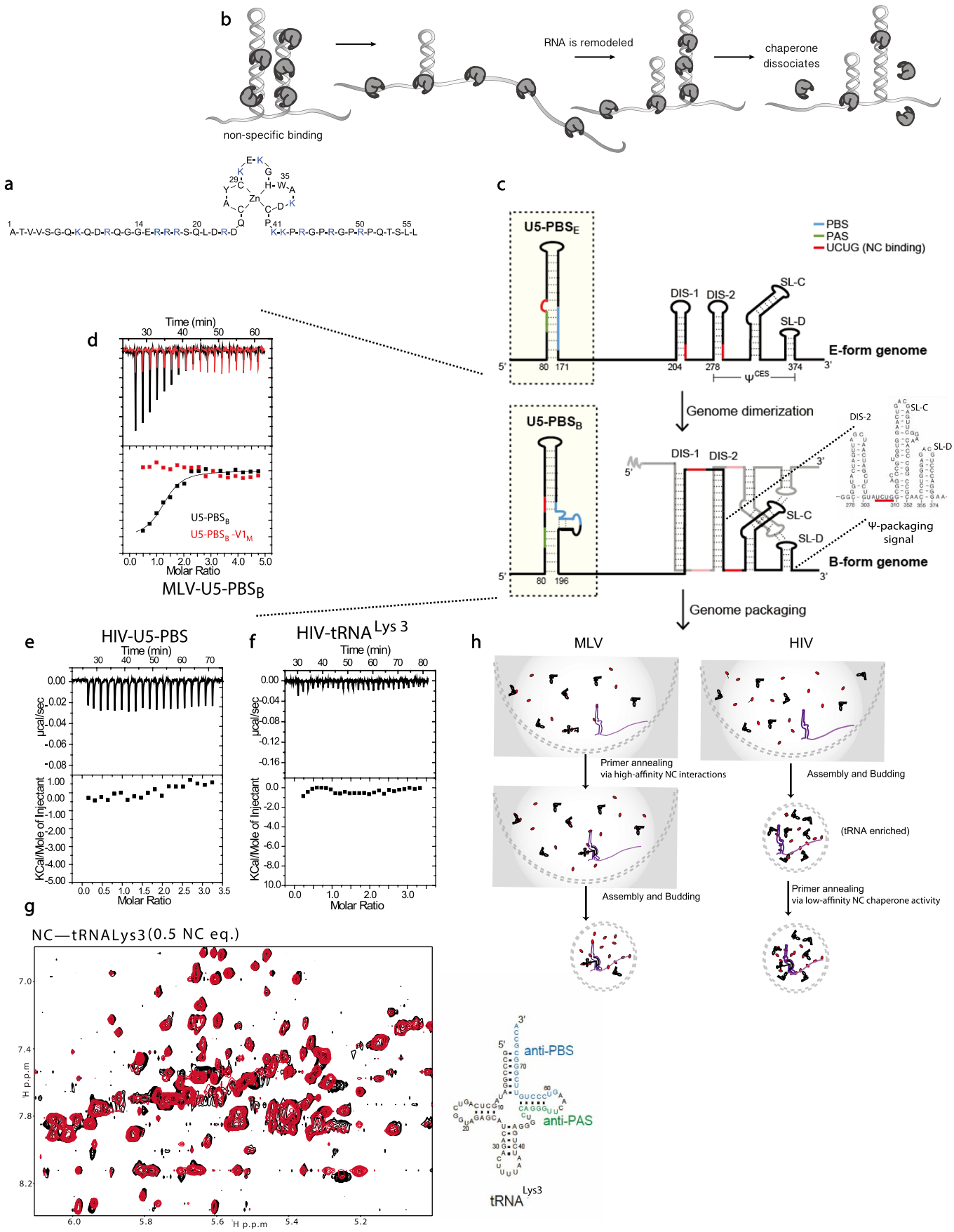
**Mutagenesis of pNCS vector for MLV virus production.** The pNCS plasmid encoding the MLV genome was a gift from the laboratory of S. Goff. The presence of the repeated U5 sequence at both ends of the proviral DNA required use of the NdeI restriction endonuclease to cleave the pNCS plasmid into two pieces, a 3 kb fragment and a 9 kb fragment. This allowed for selective introduction of the mutations only into the U5 sequence in the 5′ long terminal repeat (LTR). The 9 kb fragment was circularized by self-ligation and used as a template to introduce the G110U, deletion mutation or Ψ C331G mutation using a QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Mutant plasmids were then digested with NdeI and re-ligated with the 3 kb fragment to form the complete mutant pNCS. DNA sequencing confirmed both the orientation of the insert and the presence of the desired mutation.

**MLV virion infectivity and genome packaging assay.** 293T and Rat2 cells were grown at 37 °C and 5% $CO_2$ in DMEM supplemented with 10% fetal bovine serum (FBS) and 5% penicillin-streptomycin. The 293T cells were used for transfection-mediated viral production, while Rat2 cells were used for infectivity assays. Transfection of 293T cells with pNCS was performed at a cell confluency of ~80% using Fugene-6 (Roche). Virion-containing supernatant was harvested 48 h after transfection and filtered through a 0.45 μm filter. Viruses were quantified with the use of exogenous Luciferase RNA template (Promega) provided in excess during qPCR reactions. Purified, exogenous MLV reverse transcriptase (RT) (Promega) was used to generate a linear standard curve ($R^2 > 0.98$) for the calculation of RT activity. SYBR Green-based qPCR was carried out as described[35], and dissociation curve measurements were performed at the conclusion of each run to confirm primer specificity. Equal numbers of wild-type or mutant virions were used to infect 293T cells, as has been described[36,37]. As a negative control, separate aliquots of wild-type virions were incubated at 80 °C for 15 min to render the virus non-infectious. The RT-based virus quantification assay was used to quantify the viral yield resulting from infection.

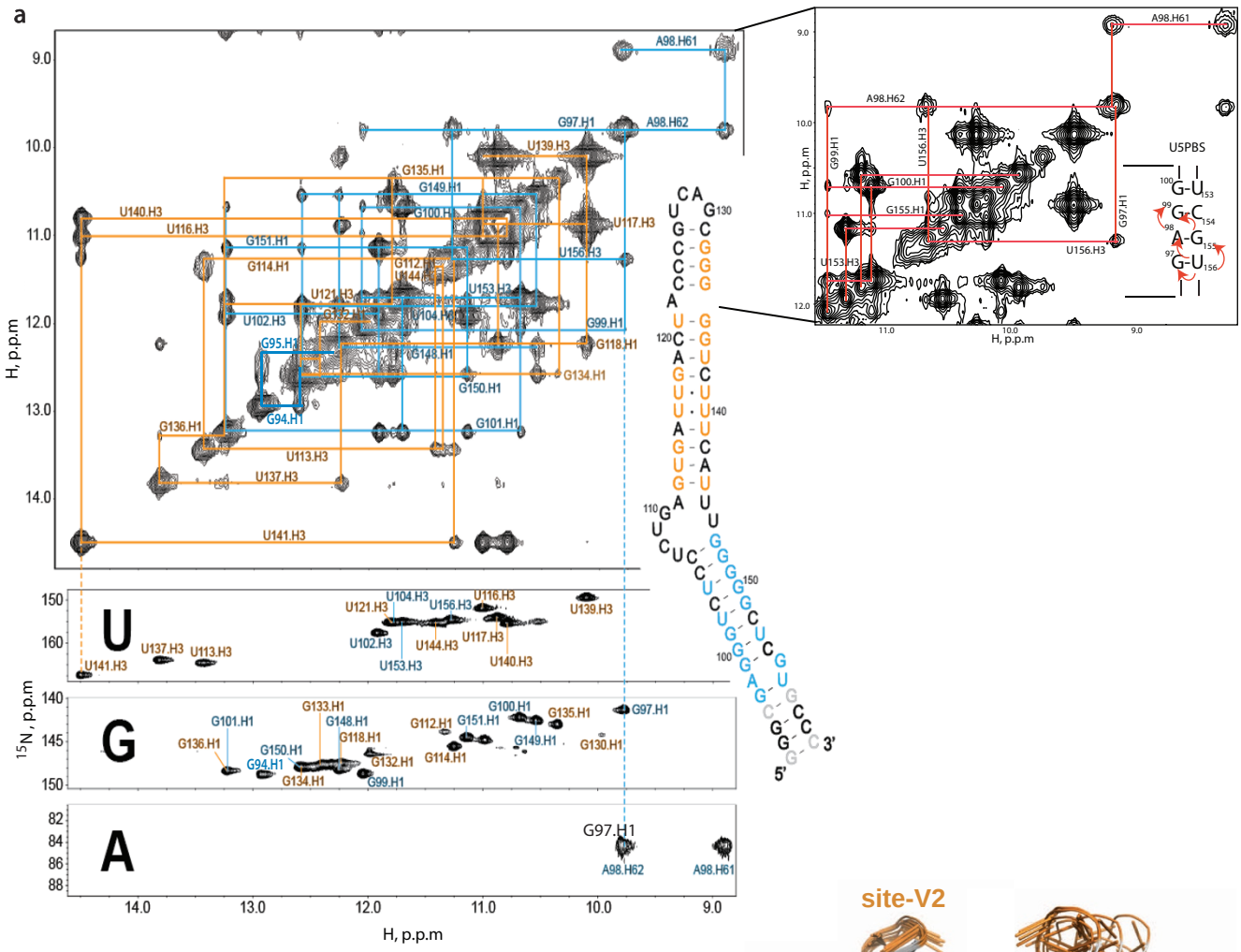
For the genome packaging assay, total virion RNA was extracted from viral supernatant with the QIAamp Viral RNA Mini Kit (Qiagen) using 50–150 μl of viral supernatant per RNA extraction column. Three independent extractions were pooled and viral genomic RNA was quantified by qPCR using a TaqMan probe-based assay system as has been described[38] with an AABI 7900 sequence detection apparatus (Applied Biosystems). Negative control reactions lacking viral RNA template yielded negligible signals, and standard curves were generated using threshold cycle (Ct) values from a range of different input viral DNA (pNCS) concentrations. A two-sample, one-tailed unpaired *t*-test was performed to compare mean values of RT activity for the using the statistical package Stata (StataCorp), assuming unequal variances and using a 99% confidence interval.

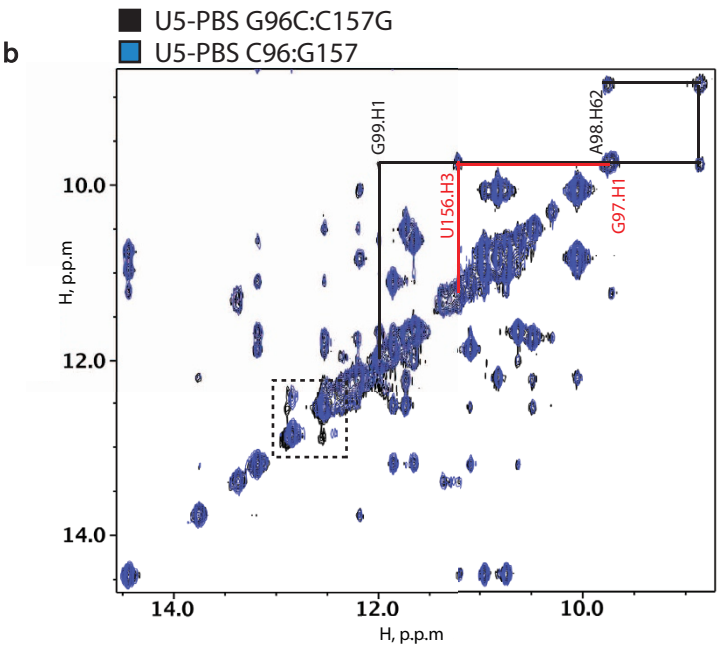
For heat-annealing, 250 pmol of MLV U5-PBS RNA was incubated with 250 pmol $tRNA^{Pro}$ at 50 °C for 5 min followed by 85 °C for 15 min in 50 mM Tris-HCl (pH 7.2), 50 mM KCl and 5 mM $MgCl_2$. For NC-mediated annealing, the same amounts of RNA were incubated with five molar equivalents of NC at 37 °C for 16 h. After annealing, 100 U RNase inhibitor and 750 U MLV reverse transcriptase were added to each sample, along with Cy3-labelled dideoxy CTP (ddCTP) to a final concentration of 0.4 nM. Following ethanol precipitation, the RNA was digested by Riboshredder RNase and resolved on a denaturing gel. DNA terminated with Cy3-ddCTP was visualized with an imager.

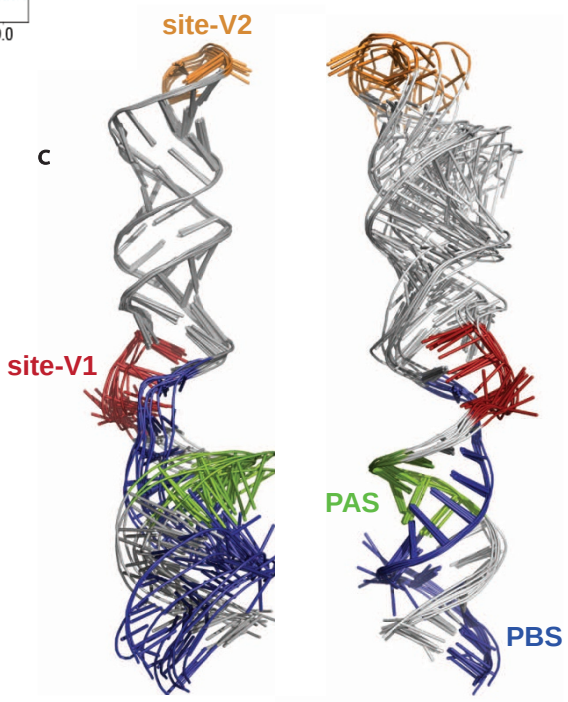

30. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* **59,** 284–291 (1986).
31. Miller, J. T., Khvorova, A., Scaringe, S. A. & Le Grice, S. F. Synthetic $tRNA^{Lys,3}$ as the replication primer for the HIV-$1_{HXB2}$ and HIV-$1_{Mal}$ genomes. *Nucleic Acids Res.* **32,** 4687–4695 (2004).
32. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5,** 1065–1074 (1998).
33. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* **337,** 427–442 (2004).
34. Spriggs, S., Garyu, L., Connor, R. & Summers, M. F. Potential intra- and intermolecular interactions involving the unique-5′ region of the HIV-1 5′-UTR. *Biochemistry* **47,** 13064–13073 (2008).
35. Pizzato, M. *et al.* A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. *J. Virol. Methods* **156,** 1–7 (2009).
36. Auerbach, M. R., Shu, C., Kaplan, A. & Singh, I. R. Functional characterization of a portion of the Moloney murine leukemia virus *gag* gene by genetic footprinting. *Proc. Natl Acad. Sci. USA* **100,** 11678–11683 (2003).
37. Lim, D., Orlova, M. & Goff, S. P. Mutations of the RNase H C helix of the Moloney murine leukemia virus reverse transcriptase reveal defects in polypurine tract recognition. *J. Virol.* **76,** 8360–8373 (2002).
38. Onafuwa-Nuga, A. A., King, S. R. & Telesnitsky, A. Nonrandom packaging of host RNAs in moloney murine leukemia virus. *J. Virol.* **79,** 13528–13537 (2005).



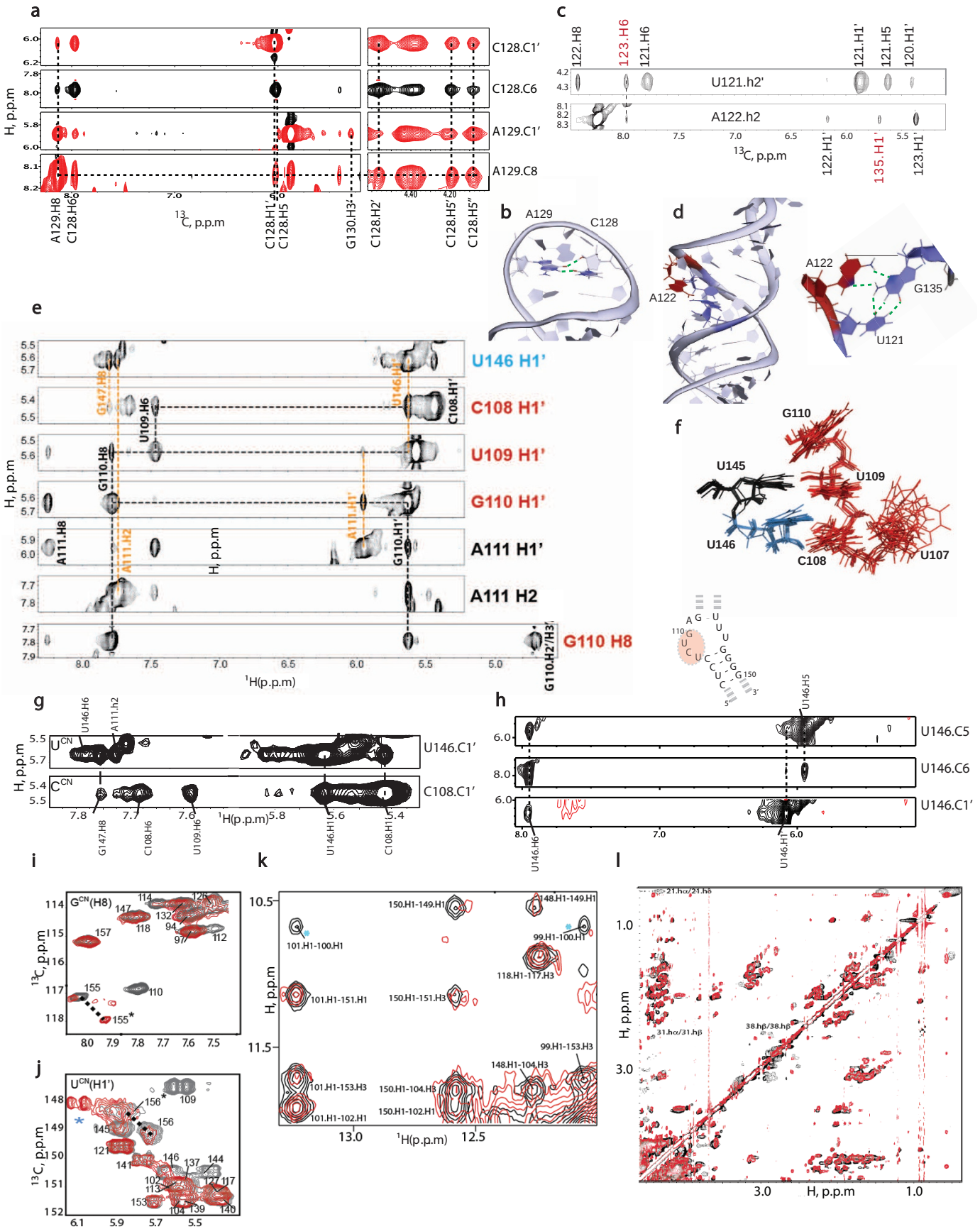
**Extended Data Figure 1 | Specific, higher-affinity interactions of MLV NC with monomeric MLV U5-PBS and tRNA$^{Pro}$ and its implications.** **a**, MLV NC sequence depicting the zinc finger moiety with amino- and carboxy-terminal tails. **b**, Cartoon representation displaying the general mechanism of ATP-independent chaperones. **c**, Cartoon representation of the E- and B-forms of U5-PBS genomic RNA, including the secondary structure of the Ψ-genome packaging signal, with the high-affinity sites coloured in red (see Supplementary Discussion 1). **d**, ITC data for MLV NC binding to native U5-PBS-B (black) and site 1 mutant ($V1_M$, red) forms. Top, raw ITC data for 1.5 μl titrations of 80 μM NC into 5 μM RNA at 30 °C. Bottom, data following peak integration, with continuous black line representing the fit for a one-site binding model. Elimination of site V1 in the U5-PBS B-form completely abolishes the binding. **e**, **f**, ITC data showing weak binding of HIV NC to HIV U5-PBS RNA and tRNA$^{Lys\,3}$ primer. **g**, Portions of two-dimensional NOESY spectra collected in $D_2O$ displaying an overlay of free tRNA$^{Lys\,3}$ primer (black) and 0.5 equivalents of MLV NC (red). Lack of any chemical shift perturbation indicates the absence of any interactions between the two molecules. **h**, Models for primer annealing in MLV and HIV-1: in MLV (left), high-affinity binding of the NC domain (red) of Gag to the U5-PBS region of the viral genome and to the tRNA$^{Pro}$ primer promotes tRNA annealing in the cytosol of the host cell before virion budding. Further supporting this model is the evidence that MLV virions are only slightly enriched for the tRNA$^{Pro}$ primer (see Supplementary Discussion 1). In HIV-1 (right), NC does not bind with high affinity to the U5-PBS domain or to the primer tRNA$^{Lys\,3}$. Furthermore, the tRNA$^{Lys\,3}$ primer is highly enriched in HIV-1 virions (see Supplementary Discussion 1), which leads to primer annealing after virion budding, mediated by weak, non-specific interactions of HIV-1 NC with the viral and primer RNAs. For the sake of simplicity, only one of the two packaged retroviral genome copies is shown.



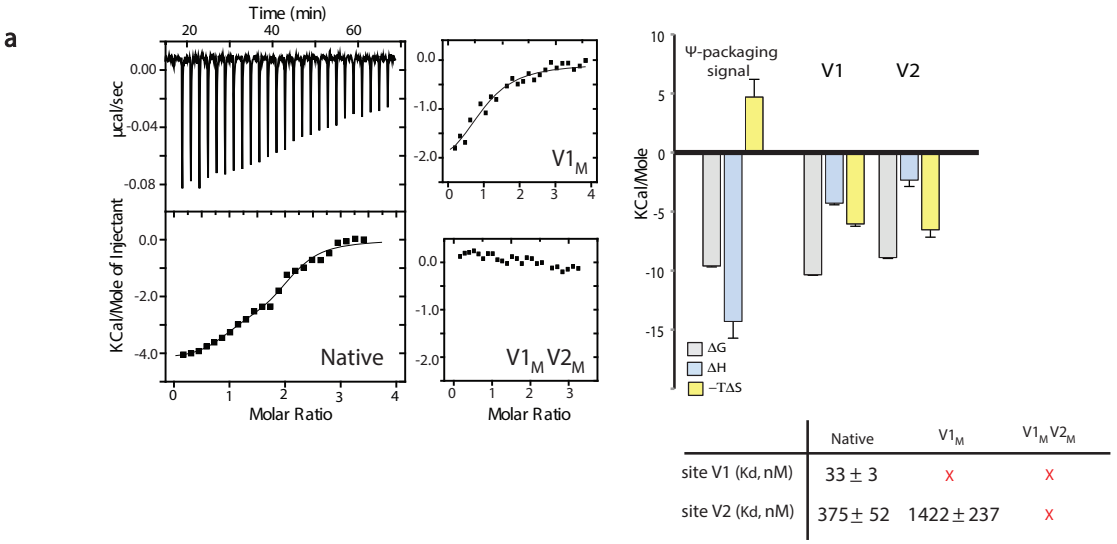



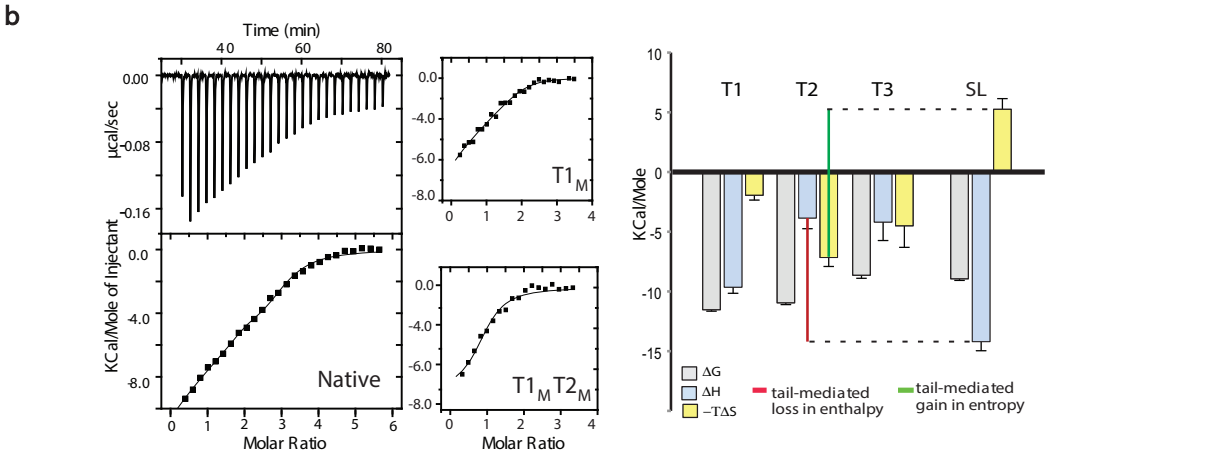
**Extended Data Figure 2 | Structural characterization of U5-PBS.** **a**, Top, portion of the $^{1}H$–$^{1}H$ two-dimensional NOESY spectrum showing imino-to-imino NOEs for the U5-PBS RNA. Data were collected at 10 °C in 10 mM NaCl and 10 mM Tris (pH 5.0). Imino-to-imino connectivities for the upper stem are shown in orange, while those for the lower stem are shown in blue. Inset, secondary structure of the U5-PBS construct used for structural and biochemical studies. Grey residues indicate non-native positions: the G96C:C157G base pair represents a base-pair swap alteration that was made to aid in unambiguous assignments of the otherwise five consecutive Gs and the terminal G–C base pair was added for the purposes of transcriptional efficiency and for SmaI digestion of the DNA template. Bottom, portions of the $^{1}H$–$^{15}N$ two-dimensional HSQC spectra for $^{15}N$, $^{13}C$-labelled U5-PBS. The imino resonances for U- and G-labelled samples are shown, as are amino resonances for the A-labelled sample. The $A_{98}$ amino resonances had unusually downfield chemical shifts due to interaction with the $G_{155}$ inside the helix. Formation of the stacking interactions of $G_{155}$ is shown on the right with a zoom-in inset. The stacking is also confirmed by a $G_{155}$ to $U_{156}$ walk in the $D_2O$ two-dimensional NOESY spectrum (data not shown). **b**, Portion of the $^{1}H$–$^{1}H$ two-dimensional NOESY imino spectrum overlay for U5-PBS constructs with and without the G96C:C157G base-pair swap. No change in the imino-to-imino connectivities were observed for the RNA except at the expected mutation site (shown in boxed region). In fact, the $GA_{98}$, $GU_{156}$ bulge immediately above the G–C swap has the exact chemical shifts, line widths and so on in both constructs, demonstrating that they have similar secondary structures. Importantly, similar binding affinities with the NC protein were obtained for both constructs. **c**, U5-PBS NMR ensemble alignments: the ensemble of the lowest-energy AMBER structures is shown. They are aligned by the top portion (left) or bottom portion (right) of the molecule. NC binding sites V1 ($UCUG_{110}$) and V2 ($UCAG_{130}$) are shown in red and orange, respectively. Left, alignment of residues 112–125 and 132–144, yielding a root mean squared deviation (r.m.s.d.) value of 0.6 ± 0.2 Å. Right, alignment of residues 94–106 and 147–159, excluding G155, which appears from NOE data to exhibit multiple conformations. The resulting r.m.s.d. for alignment of the bottom residues is 1.2 ± 0.5 Å.



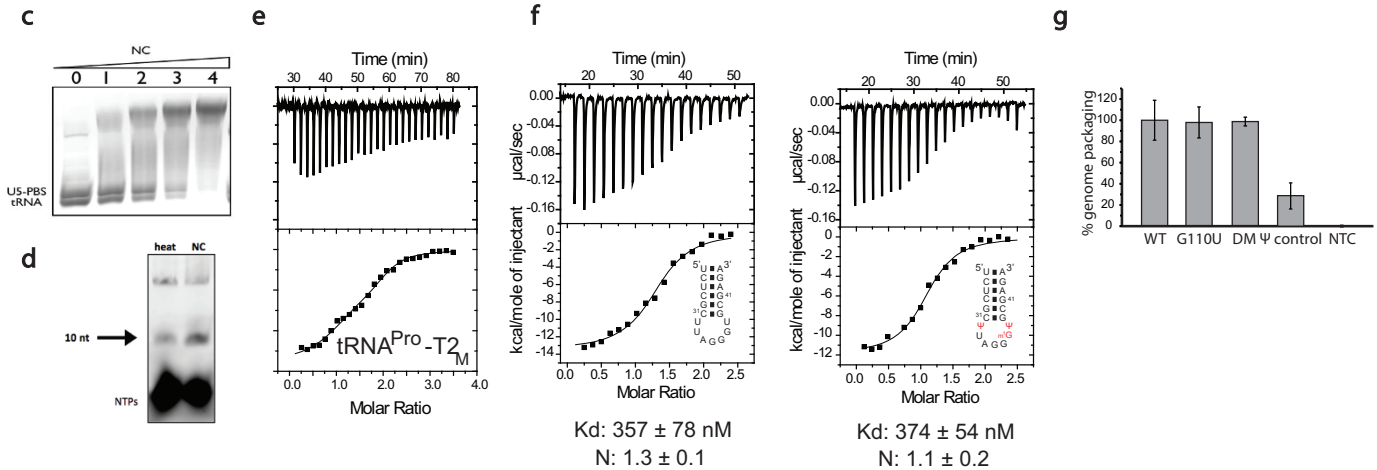
**Extended Data Figure 3 | Structural characterization of U5-PBS in the free and bound form.** **a**, Three-dimensional NOESY-HMQC strip plot of the $^{13}$C-edited $^{1}$H–$^{1}$H planes for $C_{128}$ and $A_{129}$ aromatic and H1′ protons. NOEs from $A_{129}$ H8 to the preceding $C_{128}$ ribose protons are indicative of stacking interaction between the two residues. In contrast, the minor conformation that is populated upon NC binding does not give rise to any inter-residue NOEs (data not shown). **b**, Structure of the UCAG tetraloop. The U–G hydrogen bonds, characteristic of UNCG tetraloops, are shown in green. **c**, Lack of NOEs between $U_{121}$ and $A_{122}$ indicate a break in regular stacking interactions between these two residues. Strip plots of $^{13}$C-edited $^{1}$H–$^{1}$H planes for $U_{121}$ H2′ and $A_{122}$ H2 from three-dimensional NOESY-HMQC spectra are shown. The $U_{121}$ H2′–$C_{123}$ H6 and $A_{122}$ H2–$G_{135}$ H1′ long-range NOEs are indicated in red. **d**, The $A_{122}$ bulge forms a triple base interaction. Hydrogen bonds are shown in green, and the $A_{122}$ H2–$G_{135}$ H1′ distance is indicated by a solid black line. **e**, Strip plots of $^{13}$C edited $^{1}$H–$^{1}$H planes from three-dimensional NOESY HMQC spectra of U5-PBS, showing interresidue NOE connectivities. Black, dashed lines show sequential inter-residue connectivities, while the orange lines represent long-range interactions. The strong $G_{110}$ H8-to-H1′ intraresidue NOE is indicative of a *syn* glycosidic torsion angle. Residues in the PBS and NC binding site are labelled with blue and red, respectively. **f**, Lowest-energy AMBER structures of the ($UCUGA_{111}$, $UU_{146}$) internal loop, showing the flexibility of $U_{107}$ due to lack of inter-residue NOEs. Alignment of residues 108–111 and 144–146 yields an r.m.s.d. of 0.8 ± 0.3 Å. **g**, Portions from three-dimensional NOESY-HMQC spectra collected for selectively labelled $^{13}$C,$^{15}$N-U5-PBS. As evidence for the ribose-zipper motif, intense H1′–H1′ inter-residue connectivities are observed between U146 and C108, which are located on opposite strands. **h**, Portions from three-dimensional NOESY-HMQC spectra collected for selectively labelled $^{13}$C,$^{15}$N-U5-PBS in complex with NC. Both the ribose and aromatic protons of $U_{146}$ do not give rise to any inter-residue NOEs, indicating the lack of $U_{146}$ interaction with neighbouring residues. An inset of the secondary structure of the bulge region is shown to visualize the availability of the released uracils. **i**, **j**, $^{1}$H–$^{13}$C two-dimensional HMQC spectra collected with 0 (black) and 0.9 (red) equivalents of NC. Perturbation of $G_{155}$, $U_{156}$ signals towards the minor (indicated by an asterisk) conformation are shown by dashed lines. Blue asterisk indicates the emergence of new freed uridine resonances after NC addition. **k**, Specific perturbation of $G_{100}$ is seen by selective loss of imino NOE to $G_{101}$. **l**, A portion of two-dimensional NOESY spectra collected in D2O for a 1:1 Ψ-packaging signal–NC complex (black) and U5-PBS–NC complex (red). The complete match of the two data sets indicates that, similar to the Ψ–NC interaction[19], the NC tails continue to exist in a random coil confirmation upon complex formation.



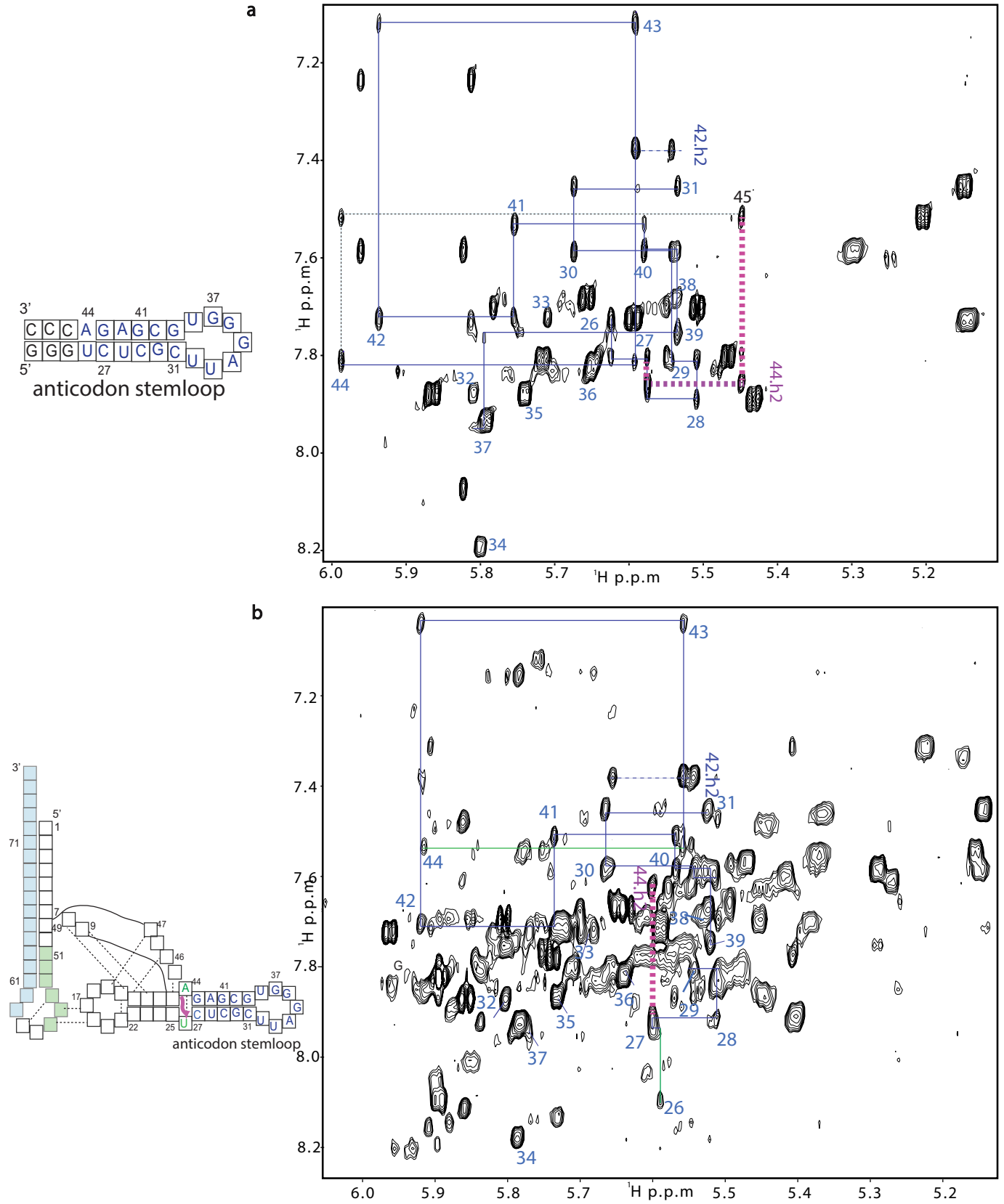

| | Native | $V1_M$ | $V1_MV2_M$ |
|---|---|---|---|
| site V1 (Kd, nM) | 33 ± 3 | x | x |
| site V2 (Kd, nM) | 375 ± 52 | 1422 ± 237 | x |



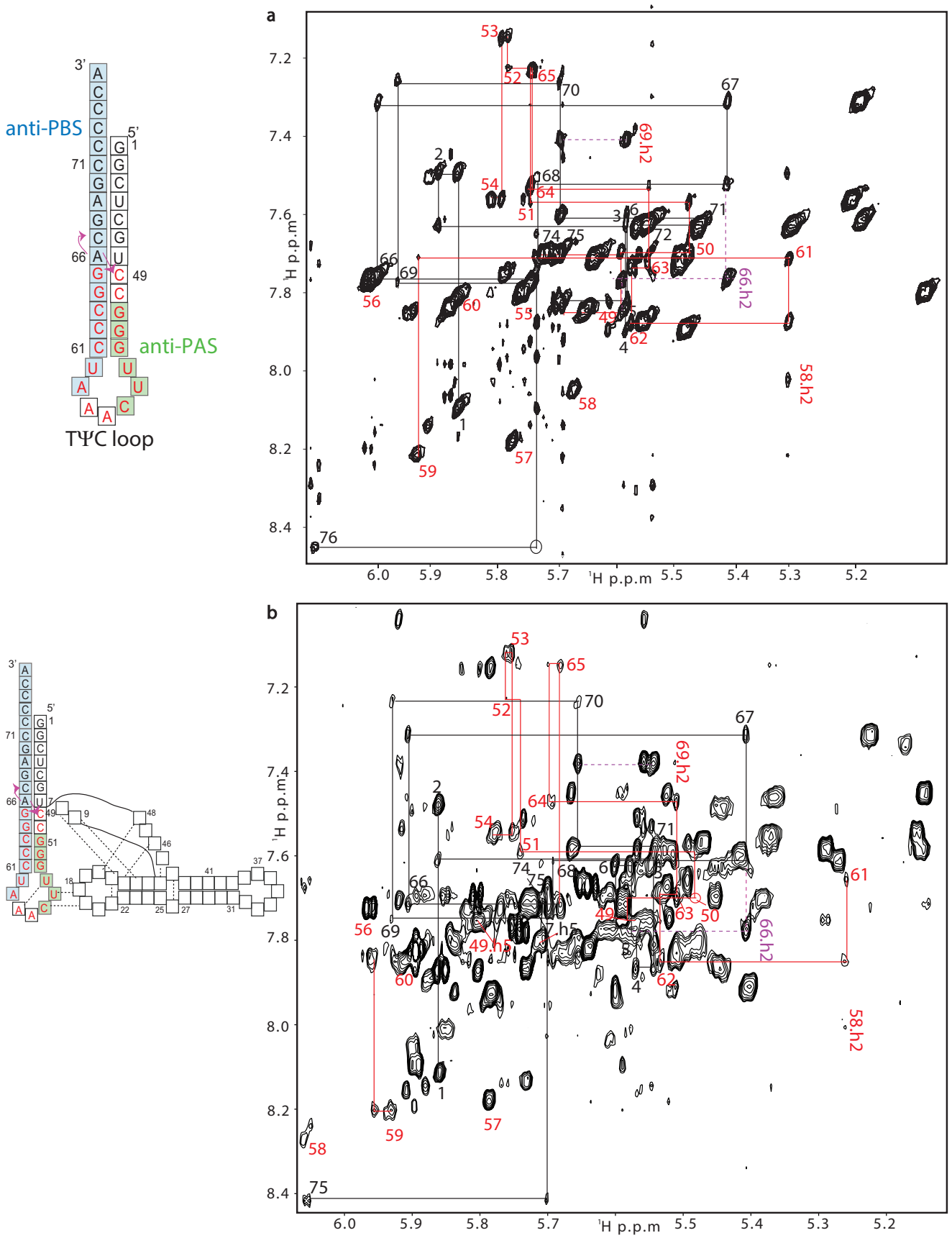

| | Native | $T1_M$ | $T2_M$ | $T1_MT2_M$ |
|---|---|---|---|---|
| site T1 (Kd, nM) | 4 ± 2 | x | 11 ± 7 | x |
| site T2 (Kd, nM) | 13 ± 3 | 153 ± 28 | x | x |
| site T3 (Kd, nM) | 834 ± 343 | 439 ± 55 | 315 ± 130 | 704 ± 85 |



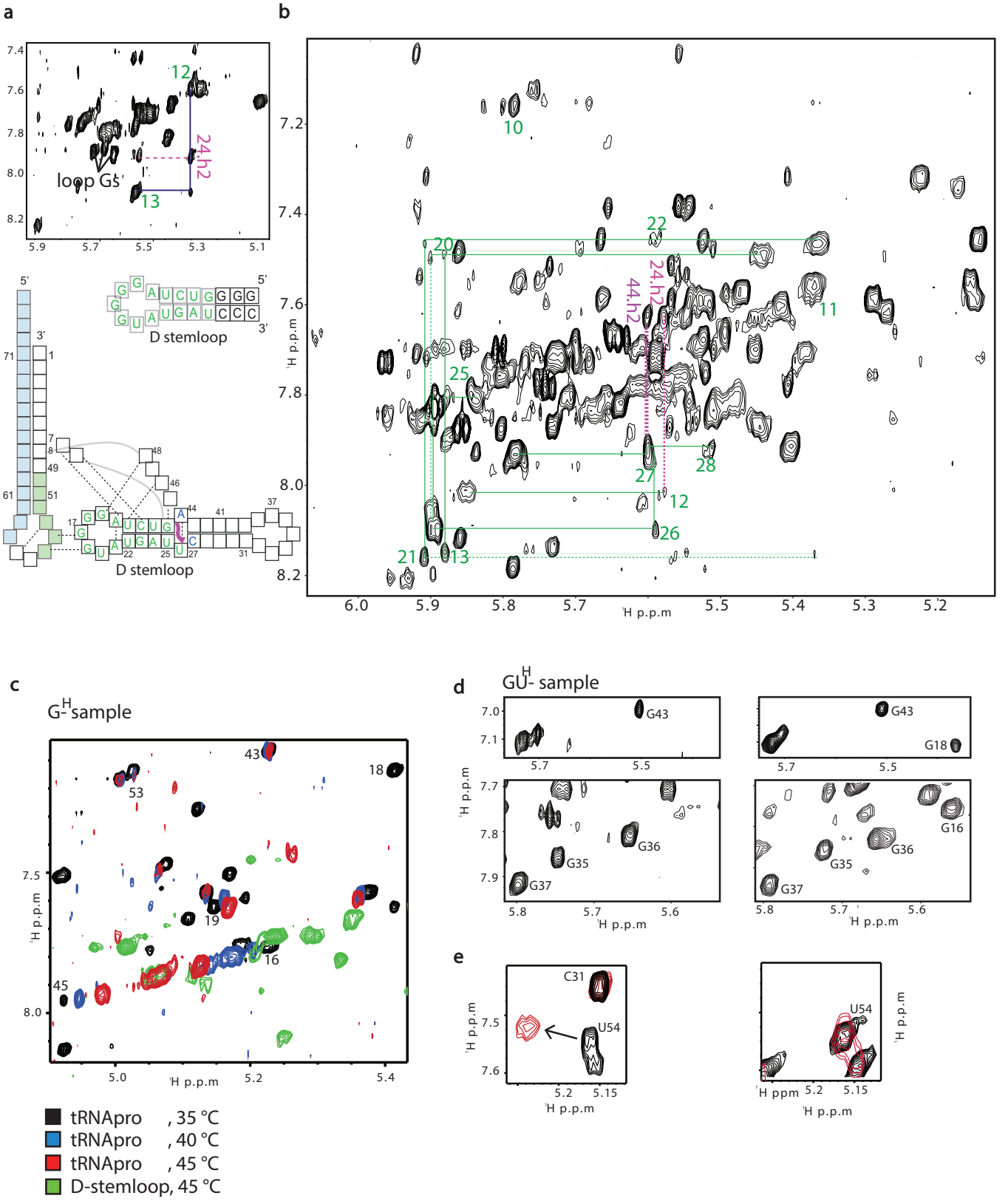
**Extended Data Figure 4 | Thermodynamic analysis of NC-U5-PBS and NC-tRNA$^{Pro}$ interaction.** **a**, **b**, ITC data for NC interactions with the U5-PBS domain (monomeric form; see Supplementary Discussion 1) (**a**) and tRNA$^{Pro}$ (**b**), along with their respective mutant RNAs ($V1_M$ = G110U; $V2_M$ = G130A; $T1_M$ = G9A; $T2_M$ = G35A,G36A,G37A; $T3_M$ was not made since G23 is part of a critical base triple). The $K_d$ values are the averages of three independent experiments, $n = 3$. Continuous lines represent the fit for two- and three-site binding models for U5-PBS and tRNA$^{Pro}$, respectively. As expected, $V1_M$ is fit with a one-site binding model, while $V1_MV2_M$ abolishes NC binding. Similarly, $T1_M$ and $T2_M$ are fit with two-site binding models, while $T1_MT2_M$ is fit with a one-site binding model. Entropy values were calculated using $T = 303$ K. The net gain in entropy (green) and loss in enthalpy (red) are shown for site T2 in full-length tRNA$^{Pro}$ in comparison with the isolated anticodon $SL_{AC}$ RNA. The favourable enthalpy resulting from zinc-finger interaction with the anticodon loop in tRNA$^{Pro}$ ($\Delta H = -14.2$ kcal mol$^{-1}$) is offset by loss of enthalpy in RNA elements beyond the anticodon stem (loss of 9.8 kcal mol$^{-1}$) and, consequently, a significant entropic compensation (gain of 11.5 kcal mol$^{-1}$) is observed (see Supplementary Discussion 7). **c**, Titration of NC into a mixture of tRNA$^{Pro}$ and U5-PBS. Lanes 1, 2, 3 and 4 correspond to 1, 2, 3 and 4 equivalents of NC. Addition of four equivalents of NC to a mixture of tRNA$^{Pro}$ and U5-PBS led to a near complete annealing of the molecule. For this experiment, the U5-PBS dimeric construct was used because, for a functional assay, the entire 18-nucleotide primer binding site is required. NMR studies on this construct, however, showed that the construct is in equilibrium with the monomeric form. This experiment demonstrates that only a few NC molecules are required to yield a functional complex. **d**, Reverse transcription of heat-annealed versus NC-annealed U5-PBS–tRNA$^{Pro}$ initiation complexes. The 10-nucleotide DNA product is obtained by reverse transcriptase terminating transcription after incorporating ddCTP at position 10 (and subsequent removal of the tRNA nucleotides by RNase treatment). **e**, ITC data for NC binding to tRNA$^{Pro}$ $T2_M$ ($AGGG_{37}$ is mutated to $AAAA_{37}$). **f**, ITC data confirming that base modifications are not required for NC binding. ITC data ($n = 3$) were fit well with a one-site binding model. **g**, A packaging assay confirms that the mutations do not affect genome encapsidation ($n = 6$). As a positive control, the observed decrease for **Ψ** (C331G) control is similar to that previously observed[37], and the non-template control (NTC) exhibits no viral RNA signal.



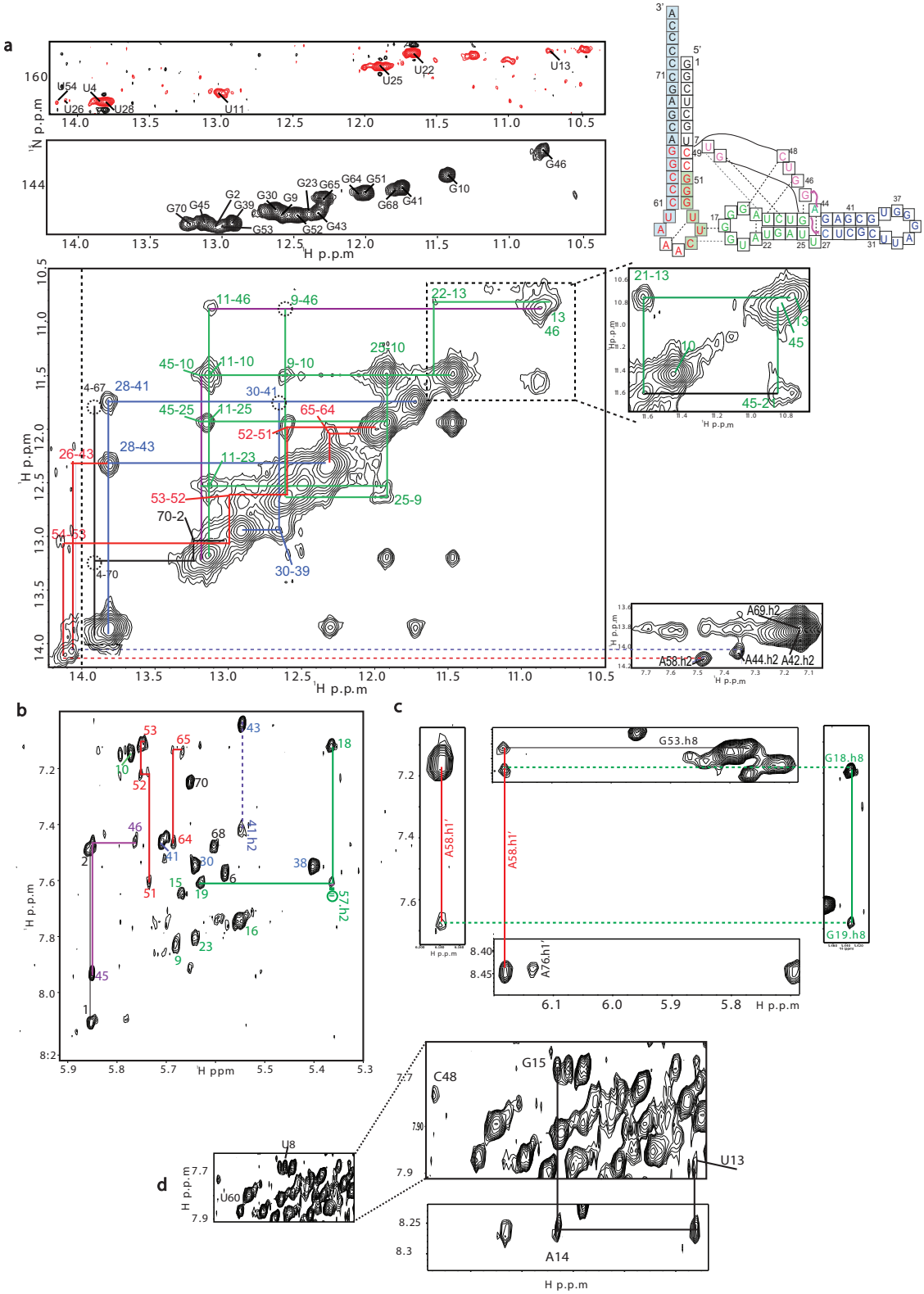

**Extended Data Figure 5 | Assignment of anticodon stem loop in tRNA$^{Pro}$.** **a**, **b**, Two-dimensional $^1$H–$^1$H NOESY spectra for the isolated anticodon stem loop (**a**) and full-length tRNA$^{Pro}$ (**b**) constructs, with blue lines indicating inter- and intra-residue aromatic-to-H1′ NOE connectivities (see also Supplementary Discussion 4).



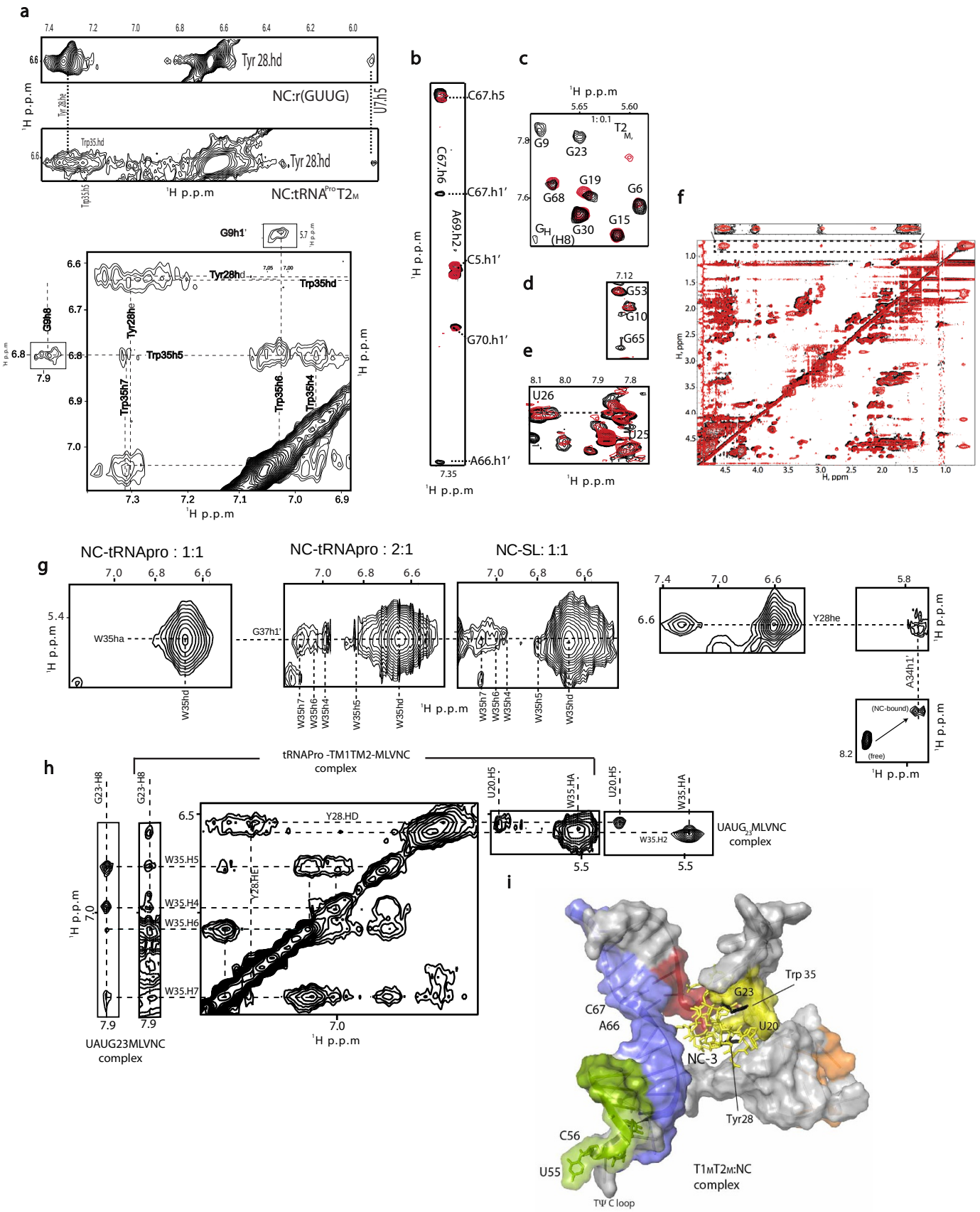

**Extended Data Figure 6 | Assignment of the acceptor–TΨC-stem loop construct confirms proper coaxial stacking in the context of the full-length tRNA$^{Pro}$ tertiary structure.** **a**, **b**, Two-dimensional $^1$H–$^1$H NOESY spectra for the isolated acceptor–TΨC-stem loop (**a**) and full-length tRNA$^{Pro}$ (**b**) constructs, with black and red lines indicating inter- and intra-residue aromatic-to-H1′ NOE connectivities in the acceptor stem and TΨC-stem loop, respectively (see also Supplementary Discussion 4).




**Extended Data Figure 7 | Assignment of the D-stem loop.** **a**, **b**, Spectra for the isolated D-stem loop (**a**) and full-length $tRNA^{Pro}$ (**b**) constructs, with lines indicating inter- and intra-residue aromatic-to-H1′ NOE connectivities. The chemical shifts of residues in the D-stem loop are drastically altered in the context of the full-length tRNA due to the formation of multiple long-range interactions between the D-stem loop and other regions in the tertiary structure of $tRNA^{Pro}$. **c**, Extreme sensitivity of the $tRNA^{Pro}$ elbow tertiary interaction to temperature. At higher temperature, the D-stem loop loses its long-range interaction with the TΨC-loop, and chemical shifts resemble those for the D-stem loop in the isolated form (see Supplementary Discussion 4). **d**, **e**, Sensitivity of the $tRNA^{Pro}$ elbow tertiary interaction to magnesium divalent salt. The data show the downfield shift of the $U_{54}$ base in the TΨC-loop and the characteristic signal for $G_{16}$, $G_{18}$ that arises as the D-loop becomes structured upon elbow interaction formation. In contrast, the bottom right panel shows the insensitivity of the isolated TΨC loop to divalent salts (see also Supplementary Discussion 4).




**Extended Data Figure 8 | Structural characterization of tRNA$^{Pro}$.** **a**, Imino resonances for full-length tRNA$^{Pro}$ indicate proper tertiary structure formation. Portions of the $^{1}$H–$^{15}$N two-dimensional HSQC spectra for U- (top) and G-labelled (centre) full-length tRNA$^{Pro}$ constructs, with imino assignments indicated. A portion of the $^{1}$H–$^{1}$H two-dimensional NOESY spectrum showing imino-to-imino NOEs is shown at the bottom. **b**, Two-dimensional NOESY of a G$^{H}$ sample of the T2$_{M}$ construct wherein only the guanosines were protonated and the other three nucleotides were deuterated. This mutant was chosen to unambiguously assign the G resonances from the D-loop and the variable loop by reducing the spectral complexity from the G-rich anticodon loop. **c**, Strips from fully protonated and G-protonated two-dimensional NOESY spectra showing direct evidence for the D-loop–TΨC interaction (see Supplementary Discussion 4). The left panel shows data collected at low temperature that allow us to confirm the long-range assignment because the A$_{58}$ H1′ NOE spin diffuses to the aromatic proton of G$_{19}$ via G$_{18}$. **d**, Strips from fully protonated sample showing the NOE walk from the D-stem to the loop residue G$_{15}$ that forms the critical Levitt base pair with C$_{48}$. This arrangement leads to an unusual downfield shift of the C$_{48}$ H1′ (see Supplementary Discussion 4).



**Extended Data Figure 9 | Structural characterization of NC-tRNA$^{Pro}$ interactions.** **a**, The packing of $U_7$ against the Tyr 28 side-chain protons leads to a downfield shift of the H5 and H1′ protons of $U_7$. Intermolecular NOEs characteristic of the fourth G of an NNNG motif entering the hydrophobic pocket of the NC protein zinc finger are also observed for $G_9$ of the $GUUG_9$. An r(GUUG) construct essentially gives the same NOE pattern as the tRNA$^{Pro}$ $GUUG_9$ upon interacting with NC, and was used as a guide for assignments. This technique has been used previously for the genome-Ψ-packaging–NC complex. The tRNA$^{Pro}$ $T2_M$ construct was used for assignments since it has no effect on site T1 binding but allows for unambiguous assignments; that is, it ensures that no NC is bound to the second site at 1:1 titration. Nevertheless, we have confirmed via HMQCs that identical site T1 binding chemical shifts are observed in the native and the $T2_M$ mutant. **b**, Portion of two-dimensional NOESY spectra collected with 0 (black) and 0.3 (red) NC equivalents. Selective perturbation of the helical walk between the 10th and 11th anti-PBS residues, $^{-10}C_{67}$ and $^{-11}A_{66}$, is shown. In contrast, the typical minor groove connectivity of the $A_{69}$ H2 shows the preservation of adjacent intramolecular associations of the seventh ($^{-7}G_{68}$), eighth ($^{-8}A_{69}$) and ninth ($^{-9}G_{70}$) anti-PBS residues. **c**, A portion of two-dimensional NOESY spectrum showing H8/H1′ correlations for the selectively protonated $T2_M$–$G^H$ tRNA$^{Pro}$ sample. Data for the $T2_M$ construct are shown to demonstrate that the observed structural changes are due solely to NC binding to site T1 rather than site T2. Addition of only 0.1 NC equivalent results in immediate perturbation of specific signals ($G_6$ and $G_9$), confirming the location of the first NC binding site. **d**, **e**, Portion of two-dimensional NOESY spectra collected with 0 (black) and 0.3 (red) NC equivalents, showing the preservation of the stacking interaction between the D-stem and anticodon stem (**e**), whereas the walk between the acceptor stem and TΨC stem is disrupted (**d**). **f**, A portion of two-dimensional NOESY spectra collected in $D_2O$ for a 1:1 Ψ-packaging-signal–NC complex (black) and tRNA$^{Pro}$ $T2_M$–NC complex (red). The complete match of the two data sets indicates that, similar to the Ψ–NC interaction, the NC tails continue to exist in a random coil conformation upon complex formation[19]. **g**, Characteristic intermolecular NOEs from the $G_{37}$ residue of the anticodon $AGGG_{37}$ binding site to the NC zinc finger. The isolated anticodon (SL) construct essentially gives the same NOE pattern as the site T2 tRNA$^{Pro}$ binding upon interacting with NC. However, this pattern is observed only upon titration of NC over 1:1 in the tRNA$^{Pro}$ construct, demonstrating the sequential binding involved in NC-mediated remodelling. **h**, Characteristic intermolecular NOEs from the $G_{23}$ residue from the $UAUG_{23}$ binding site to the NC zinc finger. The tRNA$^{Pro}$ $T1_MT2_M$ construct NOE pattern was an exact match with that of the NC–$r(UAUG)_{23}$ complex, explicitly confirming the third NC binding site, site T3. The delta proton of Tyr 28 proton of the NC protein displays an intermolecular NOE to the downfield-shifted H5 proton of $U_{20}$, which indicates packing of RNA protons against the Tyr 28 side-chain protons. Furthermore, the presence of characteristic intermolecular NOEs between the H8 proton of $G_{23}$ and the Trp 35 aromatic H5 and H6 protons indicates specific insertion of $G_{23}$ into the hydrophobic pocket of the zinc knuckle domain of the NC protein. **i**, Structure of NC bound to tRNA$^{Pro}$ $T1_MT2_M$.

**Extended Data Table 1 | Statistics for the NMR-derived structures for U5-PBS and tRNA$^{Pro}$ in the free form and in various NC-complexed states, respectively**

| | U5-PBS | | U5PBS:NC | | |
|---|---|---|---|---|---|
| | CYANA model | AMBER structure | U5PBS | NC | U5PBS :NC |
| **NMR-derived restraints** | | | | | |
| $^{1}$H-$^{1}$H distance restraints | 696 | 555 | 675 | 162 | 64 |
| Intraresidue | 324 | 324 | 333 | 20 | - |
| Interresidue | 372 | 372 | 342 | 115 | 42 |
| H-bond restraints (4/H-bond) | 282 | 141 | 272 | 27 | 22 |
| Torsion angle restraints | 316 | - | 302 | - | |
| RDC restraints | - | 32 | - | - | |
| **Statistics for AMBER structure** | | | | | |
| Mean AMBER energy (kcal mol-1) | 15,961 | | - | | |
| Mean constraint energy (kcal mol-1) | 36 | | - | | |
| Distance violations (> 0.5 Å) | 0 | | - | | |
| Dihedral angle violations (> 5°) | 0 | | - | | |
| RDC violations (> 0 Hz) | 0 | | | | |
| **Target function (Å$^2$)** | | | | | |
| Mean ± Std Dev | - | | 4.32 ± 0.56 | | |
| Minimum | - | | 3.35 | | |
| Maximum | - | | 5.29 | | |
| **Restraint violations**[2] | | | | | |
| Av. sum of upper distance viol. (Å) | - | | 27± 5 | | |
| Av. max. upper distance viol. (Å) | - | | 0.4 ± 0.04 | | |
| Av. sum of VDW viol. (Å) | - | | 13.9 ± 1.3 | | |
| Av. max. VDW viol. (Å) | - | | 0.61 ± 0.17 | | |
| Av. sum of torsion angle viol. (°) | - | | 3 ± 3 | | |
| Av max. torsion angle viol. (°) | - | | 4.84 ± 7.35 | | |
| Av. sum of lower distance viol. (Å) | - | | 3 ± 1 | | |
| Av. max. lower distance viol. (Å) | - | | 0.21 ± 0.02 | | |
| **Structure Convergence (Å)** | | | | | |
| Top (residues 114-142) | 0.6 ± 0.2 | | 0.75 ± 0.23 | | |
| bottom (residues 94-106,147-154,156-159) | 1.2 ± 0.5 | | 1.5 ± 0.15 | | |
| Internal loop ($UCUGA_{111}$, $UU_{146}$) | 0.8 ± 0.3 | | - | | |
| NC (Arg21-Pro43) | - | | 0.41 ± 0.12 | | |

| | tRNA:NC (CYANA) | | | tRNA:NC (AMBER) | | |
|---|---|---|---|---|---|---|
| | tRNA | NC1 | tRNA:NC | tRNA | NC1 | tRNA :NC |
| **NMR-derived restraints** | | | | | | |
| $^{1}$H-$^{1}$H distance restraints | 768 | 159 | 49 | 639 | 159 | 49 |
| Intraresidue | 342 | 20 | - | 342 | 20 | - |
| Interresidue | 426 | 139 | 36 | 297 | 139 | 36 |
| H-bond restraints (4/H-bond) | 261 | 22 | 13 | 130 | 22 | 13 |
| Torsion angle restraints | 270 | - | - | - | - | - |
| **Restraint violations** | | | | | | |
| Mean AMBER energy (kcal mol-1) | - | | | -19316 | | |
| Mean constraint energy (kcal mol-1) | - | | | 125.78 | | |
| Distance violations (> 0.5 Å) | - | | | 1 | | |
| Dihedral angle violations (> 5°) | - | | | 0 | | |
| Max. dihedral angle violation (°) | - | | | 0 | | |
| Max. distance constraint violation (Å) | - | | | 0.54 | | |
| **Structure Convergence (Å)** | | | | | | |
| Anticodon stem (residues 26-31,39-46) | - | | | 0.69 ± 0.15 | | |
| D-stem (residues 10-14,20-24) | - | | | 0.54 ± 0.14 | | |
| T ΨC (residues 49-53,61-65) | - | | | 0.34 ± 0.09 | | |
| NC (Arg21-Pro43) | - | | | 0.65 ± 0.15 | | |

| | tRNA:NC1:NC2 (CYANA) | | | tRNA:NC1:NC2 (AMBER) | | |
|---|---|---|---|---|---|---|
| | tRNA | NC1 NC2 | tRNA: NC1:NC2 | tRNA | NC1 NC2 | tRNA :NC1:NC2 |
| **NMR-derived restraints** | | | | | | |
| $^{1}$H-$^{1}$H distance restraints | 758 | 320 | 97 | 629 | 320 | 97 |
| Intraresidue | 339 | 40 | - | 339 | 40 | - |
| Interresidue | 419 | 280 | 71 | 290 | 280 | 71 |
| H-bond restraints (4/H-bond) | 261 | 46 | 26 | 132 | 46 | 26 |
| Torsion angle restraints | 270 | - | - | - | - | - |
| **Restraint violations** | | | | | | |
| Mean AMBER energy (kcal mol-1) | - | | | -22220 | | |
| Mean constraint energy (kcal mol-1) | - | | | 186.17 | | |
| Distance violations (> 0.5 Å) | - | | | 0 | | |
| Dihedral angle violations (> 5°) | - | | | 0 | | |
| Max. dihedral angle violation (°) | - | | | 0 | | |
| Max. distance constraint violation (Å) | - | | | 0.38 | | |
| **Structure Convergence (Å)** | | | | | | |
| Anticodon stem (residues 26-31,39-46) | - | | | 0.94 ± 0.35 | | |
| D-stem (residues 10-14,20-24) | - | | | 0.97 ± 0.20 | | |
| T ΨC (residues 49-53,61-65) | - | | | 0.69 ± 0.23 | | |
| NC (Arg21-Pro43) | - | | | 0.66 ± 0.24 | | |
| NC2 (Arg21-Pro43) | - | | | 0.89 ± 0.16 | | |

* Pairwise r.m.s.d. was calculated among ten refined structures.