Supplementary Information for:

The Structure of an Entire HIV-1 RNA Genome

Joseph M. Watts, Kristen K. Dang, Robert J. Gorelick, Christopher W. Leonard, Julian W. Bess, Jr., Ronald Swanstrom, Christina L. Burch and Kevin M. Weeks*

* weeks@unc.edu

Contents:

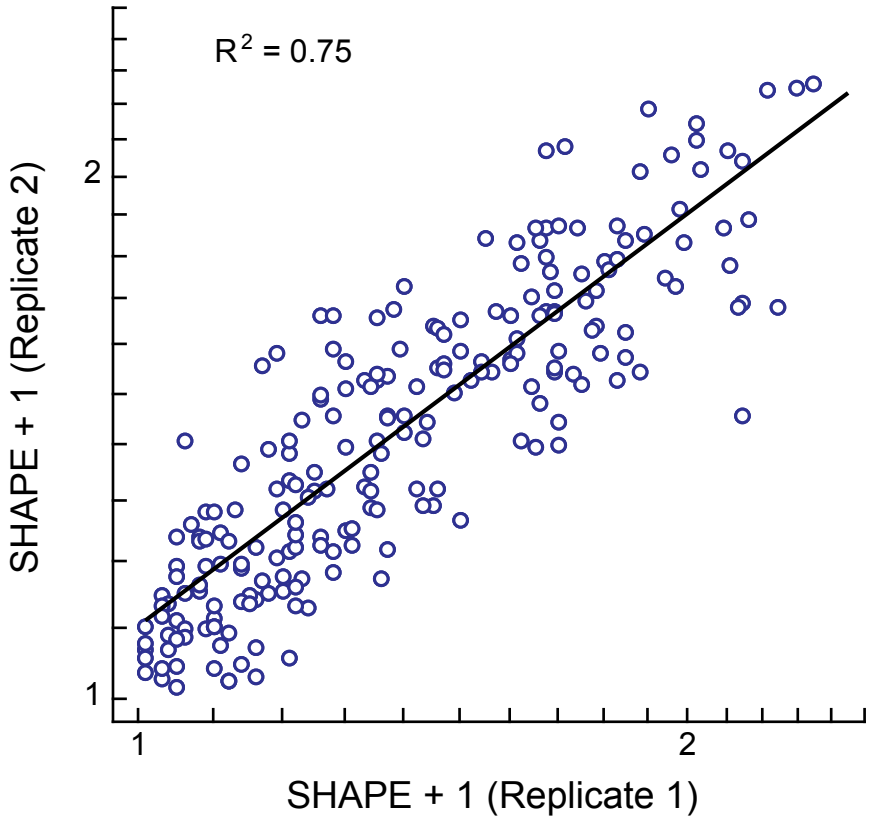


Figure S1. Reproducibility of SHAPE measurements between biological replicates. Reactivities corresponding to extension reactions performed with primer 9 are shown plotted on a logarithmic scale. Replicates were performed on independent HIV-1 RNA genome preparations, by different individuals (J.M.W. and C.W.L.), roughly one year apart.

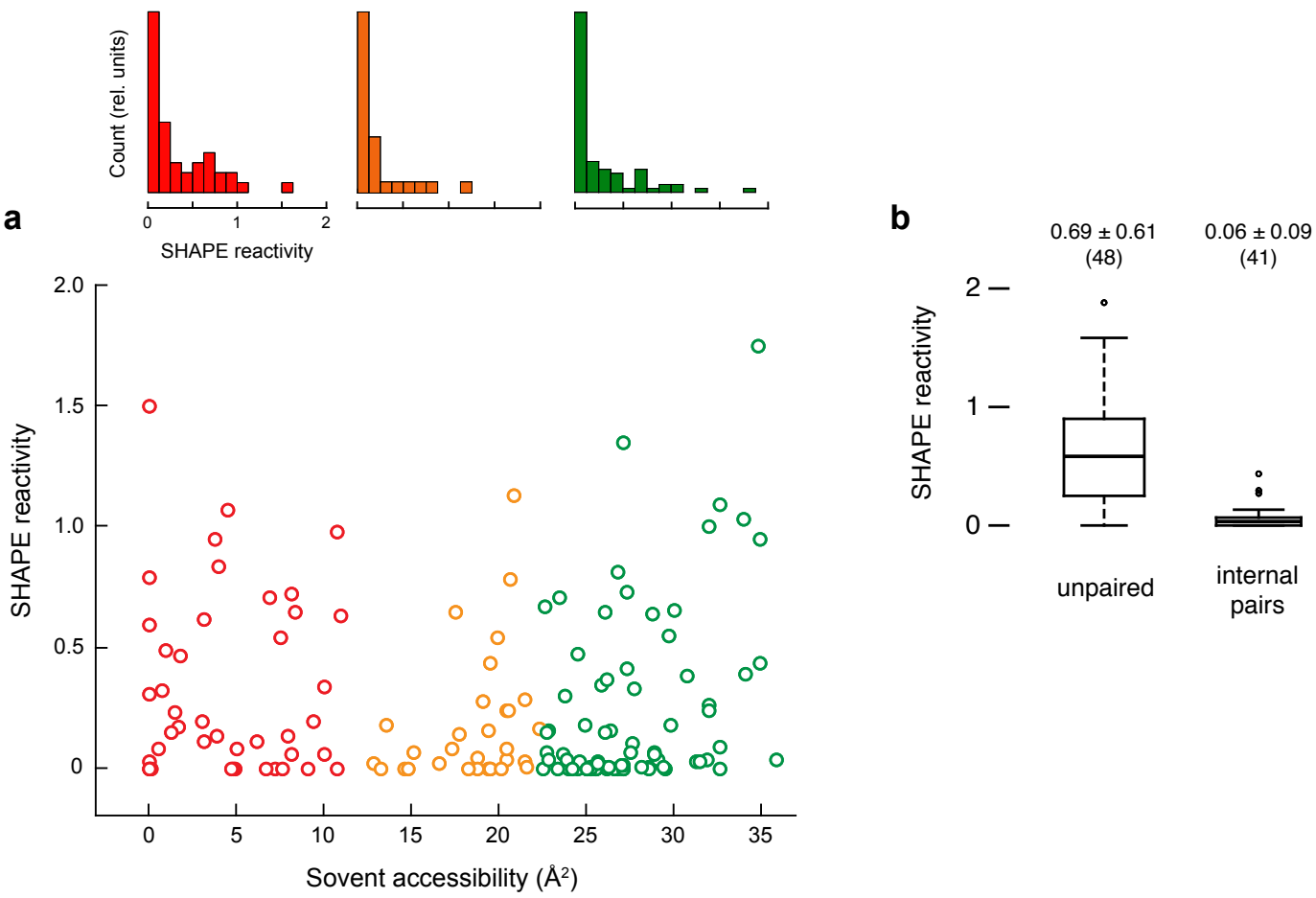


Figure S2. SHAPE reactivities are strongly sensitive to RNA secondary structure, but not to solvent accessibility. These SHAPE data are from the RNase P specificity domain RNA[10,11], which has a compact structure with significant, tightly packed, tertiary interactions[12]. Solvent accessibility was calculated using a 1.4 Å radius probe for the ribose 2'-oxygen atom. (a) SHAPE reactivities do not correlate with solvent accessibility (R^2 = 0.0004). Small panels at top give SHAPE reactivity distributions for nucleotides whose solvent accessibility are low, med or high (in red, orange and green, respectively). Distributions are similar and all regions contain nucleotides with both high and low reactivities. (b) SHAPE reactivities strongly discriminate between unpaired and base paired nucleotides. Box plot representations are shown. Numerical values give the mean ± standard deviation; the number of measurements is in parentheses. Internal base pairs are defined as positions that are paired as visualized crystallographically and are adjacent to other canonically paired nucleotides[11]. The strong predictive relationship between SHAPE reactivity and secondary structure is consistent with benchmarks showing SHAPE measures local disorder in RNA[13].

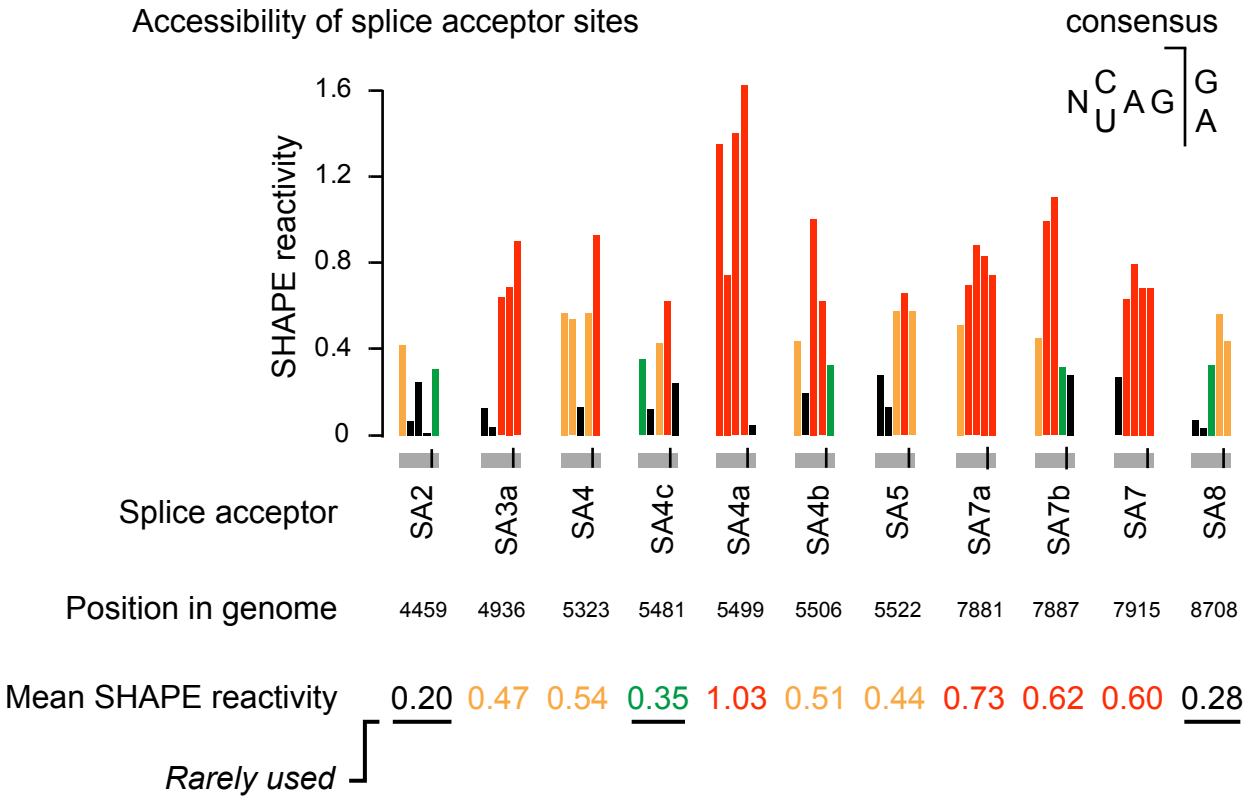


Figure S3. SHAPE reactivities at splice acceptor sites in the NL4-3 genome. Histograms show SHAPE reactivities at each splice acceptor site across the five nucleotide consensus motif. The mean SHAPE reactivity at each acceptor is colored according to the scale used throughout the manuscript. The mean SHAPE reactivity across all five nucleotide windows in the HIV genome is 0.34; thus, all splice acceptors have high SHAPE reactivities, except the three (underlined) characterized as rarely used sites[14].

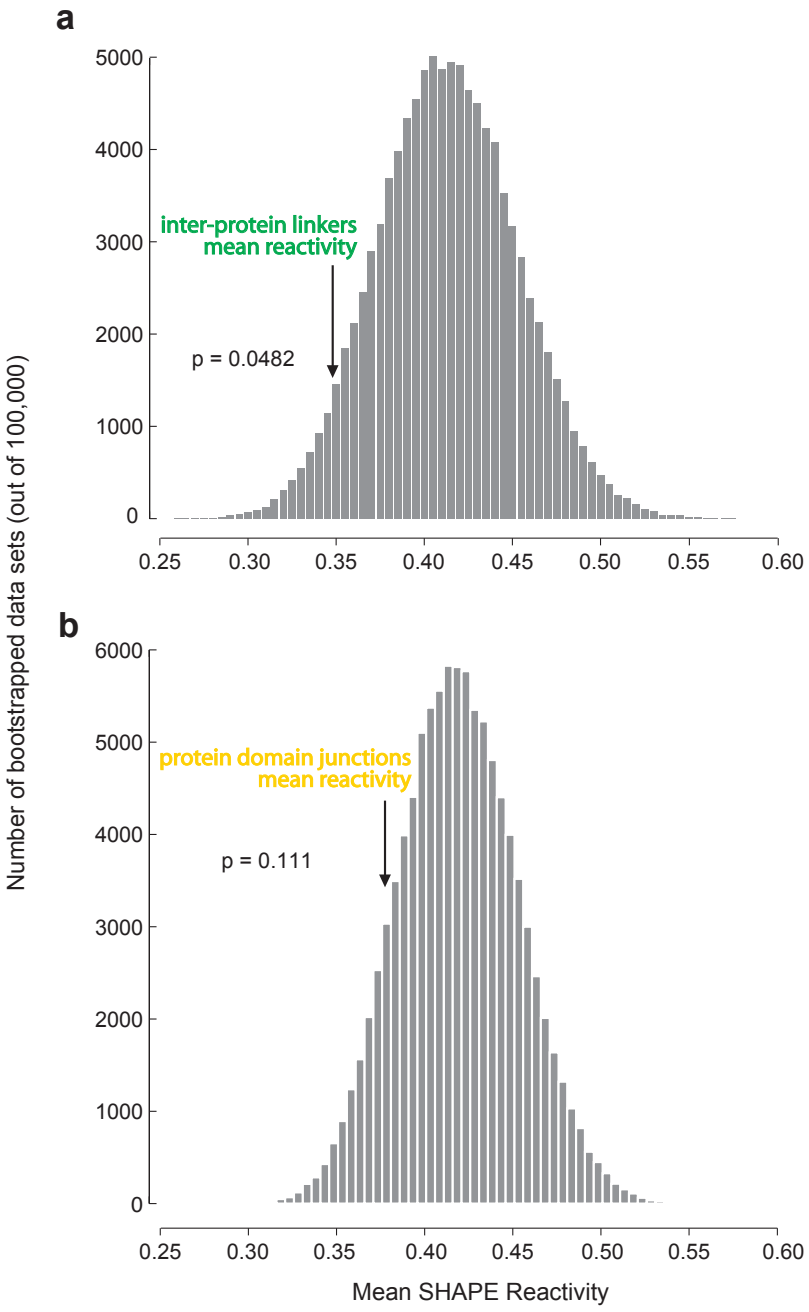


Figure S4. Histograms comparing mean SHAPE reactivities for inter-protein linker regions and protein domain junctions with distributions of equivalent-length sequences obtained from random regions in the HIV genome by a bootstrap statistical analysis. p-values give the probability that the low SHAPE reactivities in the collection of genome elements occurred by chance.

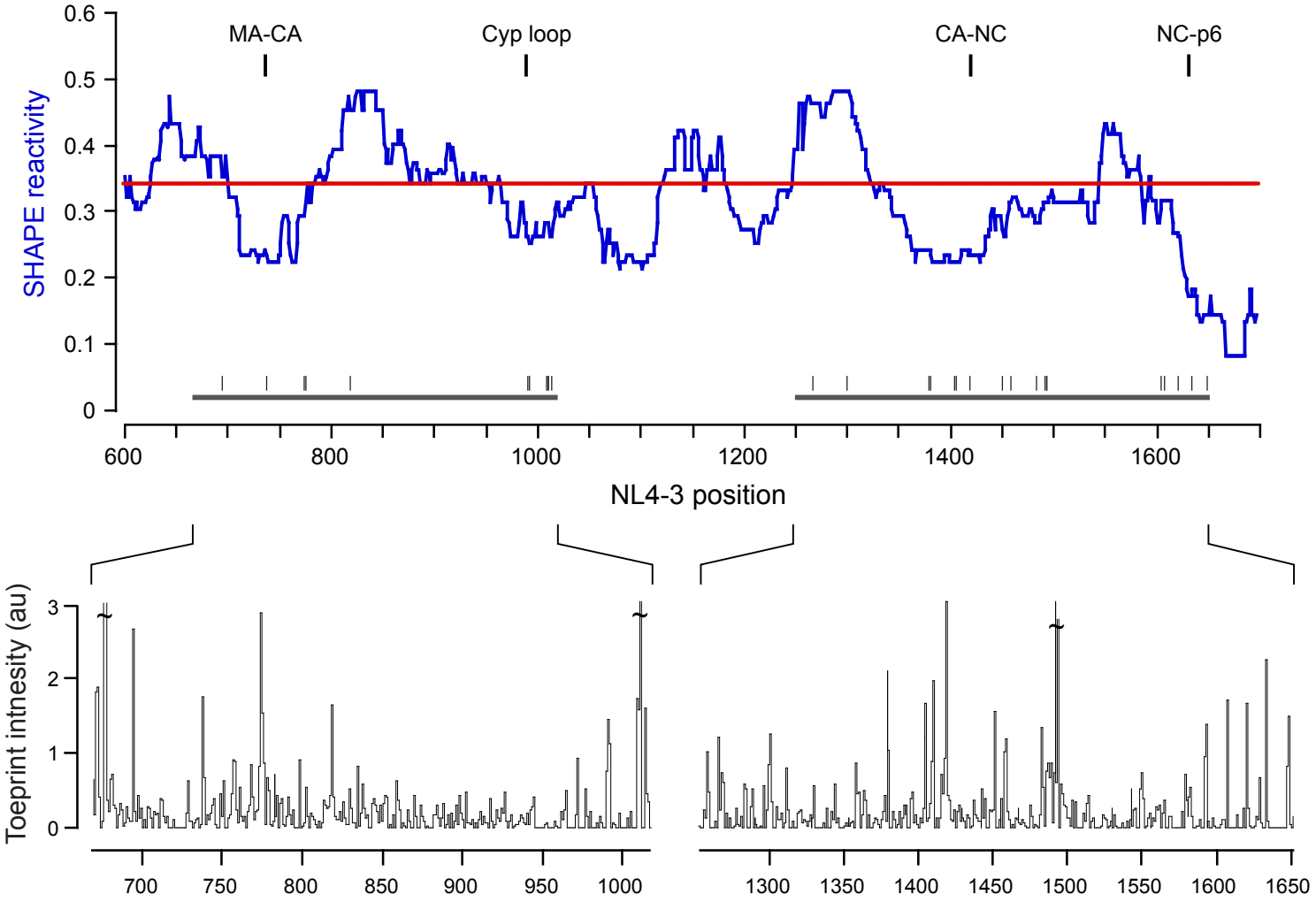


Figure S5. Distribution of ribosome pause sites in Gag at the MA-CA and CA-NC junctions. Pause sites were identified by inhibition of reverse transcriptase-mediated primer extension, after first inhibiting ribosome processivity with cycloheximide. (Top panel) Median SHAPE reactivities in Gag over a 75 nt widow, reproduced from Fig. 1 in the main text, are shown in blue. Red line shows median SHAPE reactivity over entire genome. Domain junctions and the cyclophilin loop are indicated explicitly. Gray bars indicate regions scanned in the toeprinting experiment; the strongest pause sites are indicated with vertical bars. (Lower panels) Toeprinting intensity as a function of genome position.

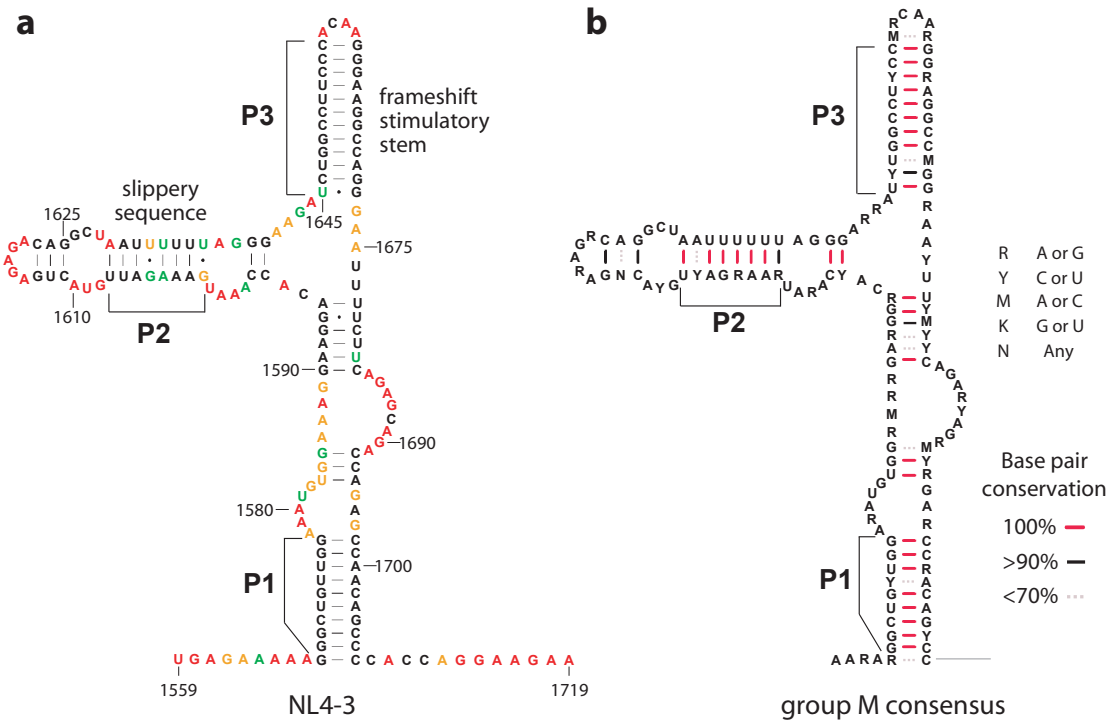


Figure S6. Structure of the HIV-1 gag-pol frameshift element. **(a)** SHAPE-constrained secondary structure. Nucleotides are colored according to their SHAPE reactivities using the scale shown in Fig. 3b. **(b)** Sequence and structural conservation for the 3-helix junction model across 37 HIV-1 group M reference sequences.

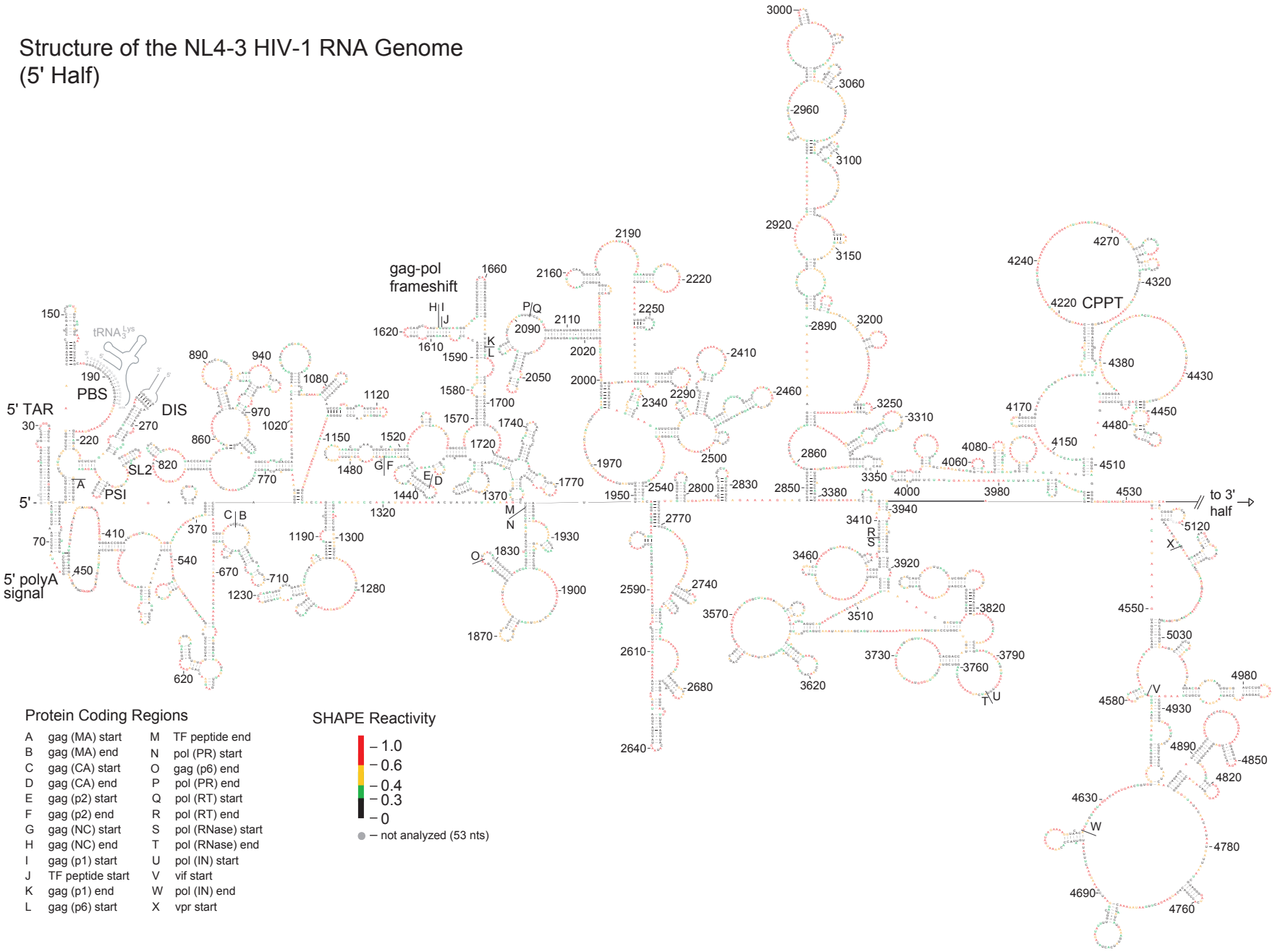
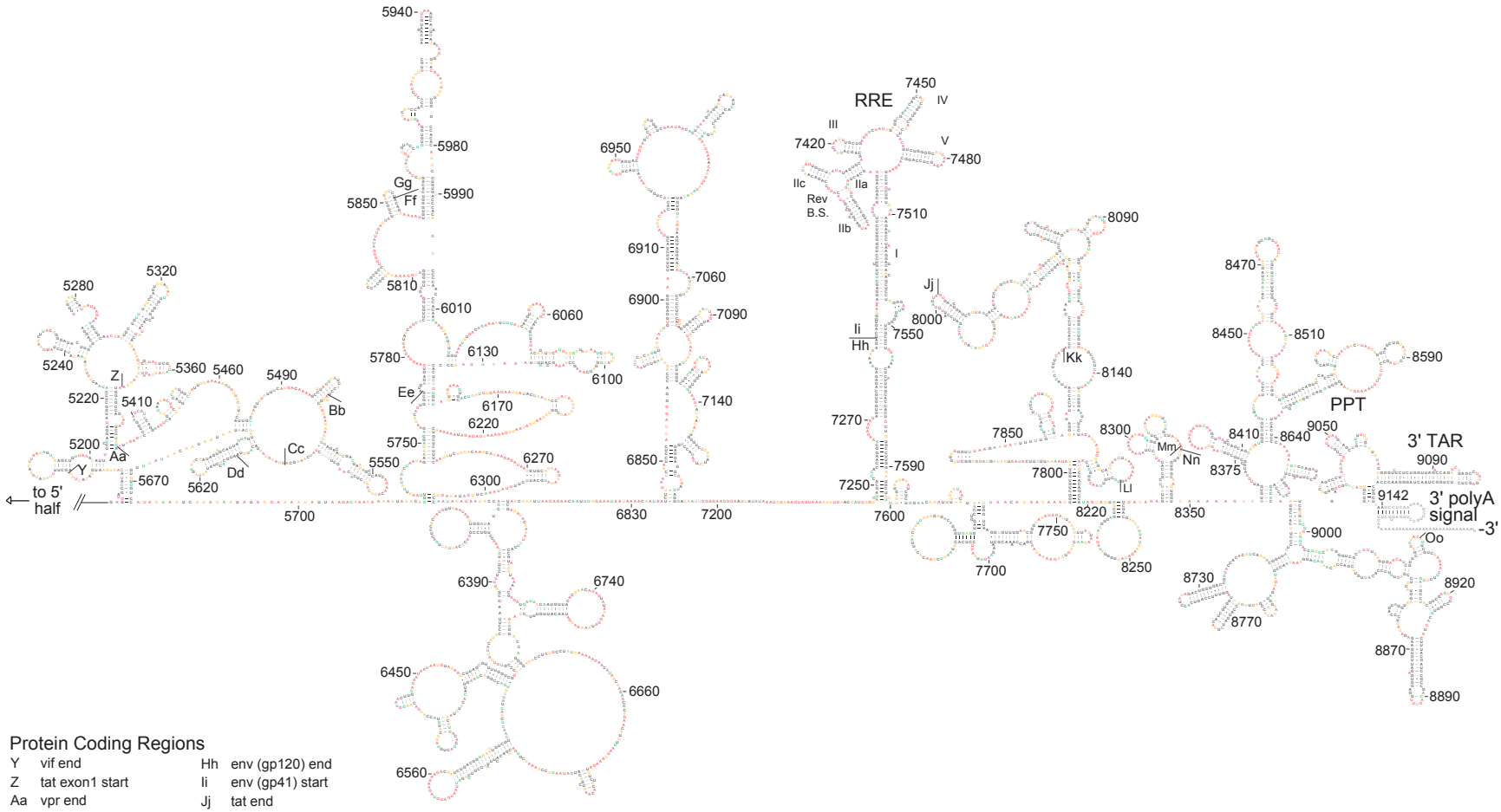


Figure S7. Detailed secondary structure for the NL4-3 HIV-1 genome, including nucleotide identities. Divided into two panels.

Structure of the NL4-3 HIV-1 RNA Genome (3' Half)



Protein Coding Regions

Y	vif end	Hh	env (gp120) end
Z	tat exon1 start	Ii	env (gp41) start
Aa	vpr end	Jj	tat end
Bb	rev exon 1 start	Kk	tat/rev exon 2 start
Cc	tat/rev exon 1 end	Ll	rev end
Dd	vpu start	Mm	env end
Ee	env signal peptide start	Nn	nef start
Ff	env (gp120) start	Oo	nef end
Gg	env signal peptide end		

Table S1: NL4-3 SHAPE primer sequences

Number	Primer Sequence	Binding Site in the NL4-3 Genome
1	GCTTAATACCGACGCTCTCGC	342-362
2	CGTTCTAGCTCCCTGCTTGCC	443-463
3	CCTGGCTGTTGTTTCCTGTGT	697-717
4	GTTCCTGCTATGTCACTTCCCC	1030-1051
5	GTGTCGCTCCTGGTCCCAATGC	1341-1362
6	CTCTCTTCTGGTGGGGCTGTTG	1699-1720
7	CAATTATGTTGACAGGTGTAGG	2033-2054
8	GCATCGCCCACATCCAGTACTG	2415-2436
9	CTGATGTTTTTTGTCTGGTGTG	2740-2761
10	GCCCCTGCTTCTGTATTTCTGC	3074-3095
11	GCTGCCCCATCTACATAGAAAG	3411-3432
12	CAGCACTGACCAACCCATCTAC	3737-3758
13	GCTGTTTCTTGCCCTGTCTCTG	4047-4068
14	CTTTCCCCTGCACTGTACCCCC	4350-4371
15	CAAACTGGATCTCTGCTGTCCCTG	4456-4479
16	CCCTGACCCAAATGCCAGTCTC	4816-4837
17	GCTCCCTCTGTGGCCCTTGGTC	5123-5144
18	GCTGTCTCCGCTTCTTCCTGCC	5517-5538
19	CCCCATTTCCACCCCCATCTCC	5798-5819
20	GTGGGGTTAATTTTACACATGG	6118-6139
21	GAATCGCAAAACCAGCCGGGGC	6415-6436
22	CATTTTGCTCTACTAATGTTAC	6773-6752
23	CATCTCTTGTTAATAGCAGCCC	7111-7132
24	TCTGGCCTGTACCGTCAGCGTC	7365-7386
25	CTCTGTCCCACTCCATCCAGGTC	7638-7660
26	CCTACCAAGCCTCCTACTATCA	7820-7841
27	CTATTCCTTCGGGCCTGTCGG	7940-7960
28	GCAAAATCCTTTCCAAGCCCTG	8305-8326
29	GTAGCCTTGTGTGTGGTAGATCC	8670-8692
30	GTACAGGCAAAAAGCAGCTGC	9053-9073
31	TTTTTTTTTTTTTTTTTTTTTGAAG	9170-9173 / poly(A)

Table S2: Protein Domains in Figure 1d

Protein	PDB (ref.)	Protein Residues	Color
MA/CA	2GOL [1]	1-105	Blue
		106-132	Green
		133-216	Red
		217-231	Yellow
		232-278	Pink
		278-279	Yellow
CA (C-terminal)	1A43 [2]	280-363	Gray
PR	3PHV [3]	1-99	Blue
RT/RNase H (p66)	1HMV [4]	2-210	Red
		211-253	Yellow
		254-310	Pink
		311-326	Yellow
		327-416	Gray
		417-437	Green
		438-556	Magenta
IN	1WJA [5]	1-47	Red
		48-55	Yellow
	1BIS [6]	56-209	Pink
		210-218	Yellow
	1QMC [7]	219-270	Magenta
gp120	1GC1 [8]	90-127	Blue
		180-200	Yellow
		201-492	Blue
gp41	1SZT [9]	35-68, 117-144	Red

Supplementary References

1. Kelly, B.N. *et al.* Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(N)(133-278). *Biochemistry* **45**, 11257-11266 (2006).
2. Worthylake, D.K., Wang, H., Yoo, S., Sundquist, W.I. & Hill, C.P. Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 85-92 (1999).
3. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Geoghegan, K.F., et al. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* **342**, 299-302 (1989).
4. Rodgers, D.W. *et al*. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **92**, 1222-1226 (1995).
5. Cai, M. et al. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **4**, 567-577 (1997).
6. Goldgur, Y. *et al*. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. USA* **95**, 9150-9154 (1998).
7. Eijkelenboom, A.P. *et al*. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* **2**, 807-810 (1995).
8. Kwong, P.D. *et al*. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659 (1998).
9. Tan, K., Liu, J., Wang, J., Shen, S. & Lu, M. Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. USA* **94**, 12303-12308 (1997).
10. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144-4145 (2007).
11. Wilkinson, K.A. *et al*. Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* **15**, in press (2009).
12. Krasilnikov, A.S., Yang, X., Pan, T. & Mondragon, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760-764 (2003).
13. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G. & Weeks, K.M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S^2) in RNA. *J. Am. Chem. Soc.* **130**, 12244-12245 (2008).
14. Purcell, D.F. & Martin, M.A. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* **67**, 6365-6378 (1993).

Supplementary Datasets

Dataset S1. Helix file for the complete NL4-3 RNA genome structure model. In tab-delimited ascii format. Columns are: first nucleotide in helix, last nucleotide in helix, number of base pairs in helix.

Dataset S2. All SHAPE reactivities and pairing probabilities for the NL4-3 HIV-1 RNA genome. In tab-delimited ascii format. Columns are: nucleotide position, nucleotide identity, SHAPE reactivity, pairing probability.