

## THE STRUCTURE OF AN ENTIRE HIV-1 RNA GENOME

Joseph M. Watts<sup>1</sup>, Kristen K. Dang<sup>2</sup>, Robert J. Gorelick<sup>5</sup>, Julian W. Bess, Jr.<sup>5</sup>, Ronald Swanstrom<sup>3</sup>, Christina L. Burch<sup>4</sup>, Kevin M. Weeks<sup>1</sup>

<sup>1</sup>Department of Chemistry, <sup>2</sup>Department of Biomedical Engineering, <sup>3</sup>Lineberger Cancer Center, <sup>4</sup>Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3290; <sup>5</sup>AIDS and Cancer Virus Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702-1201

Retroviral replication is regulated at many levels, including using conserved genomic RNA structures to exploit and circumvent host biology. The vast majority of regions within viral RNA genomes are structurally uncharacterized. In order to understand how RNA structure influences retroviral biology, we have interrogated the structure of an entire ~9,200 nucleotide HIV-1 genome at single nucleotide resolution using SHAPE, a high-throughput RNA analysis technology. We used a statistical model of molecular evolution to show that many new RNA structures identified by SHAPE are broadly conserved among HIV-1 genomes. Some simple genome elements previously shown to be functionally important, including the ribosomal *gag-pro-pol* frameshift stem-loop, are components of large, highly complex, RNA motifs. Thirteen of fourteen RNA sequences that encode the peptide loops that link protein domains are highly structured, consistent with a model in which viral mRNA structure routinely functions to slow ribosome elongation to promote native protein folding. Other structures sequester hypervariable regions and likely enable the genome to tolerate high levels of sequence heterogeneity without disrupting neighboring RNA regulatory structures. This complete analysis of a viral RNA genome emphasizes that long RNAs are punctuated by numerous previously unrecognized, but conserved, RNA motifs and reveals an intimate connection between higher order RNA structure and the proteins it encodes. Many of these newly identified and conserved RNA structures represent promising targets for new antiretrovirals.