Thermodynamic patterns of eukaryotic genes

As it reads the information encoded throughout the genome, RNA polymerase II can travel along the DNA template for thousands of nucleotides. In the process, it encounters the physical forces of DNA/DNA and RNA/DNA pairing that can vary depending on the local sequence composition. It has been shown that the 5′- and 3′-untranslated regions (UTRs), introns and exons have characteristic guanine/cytosine (GC) content, which could affect RNA transcription and processing. Nucleotide composition could influence protein recruitment, RNA secondary structure, transcription rate, DNA melting or RNA/DNA and DNA/DNA duplex stability. The free energy (ΔG) necessary to unwind polynucleotide duplexes with defined length can be calculated from the measured values of entropy (ΔS) and enthalpy (ΔH) for the 10 possible nearest-neighbour DNA/DNA interactions, and the 16 possible RNA/DNA interactions.

This work aims to study the role of thermodynamic stability in RNA processing by using appropriate thermodynamic parameters that allow comparison of mRNA/DNA with DNA/DNA duplex stability.

To map the regions in transcripts that contribute to differential stability, we calculated the thermodynamic profile of the 50 bp upstream and the 50 bp downstream of the transcript start sites, the 5′- and 3′- splice sites and the ends of the 3′-UTRs of all Human, Caenorhabditis elegans, Arabidopsis thaliana, Danio rerio, Drosophila melanogaster and Saccharomyces cerevisiae transcripts[1]. This approach allows alignment of the thermodynamic profile of the individual sequences with respect to the regions responsible for RNA processing. As the RNA polymerase maintains a 9-bp RNA/DNA duplex in the transcription bubble during elongation, we used a window size of 9 bp for ΔG calculations.

Methods

Genomes and annotations

Annotations and sequences were obtained from the Ensembl genome browser as follows: release 65 of H. sapiens genes (GRCh37.p5), release 68 of D. melanogaster genes (BDGP5), release 68 of D. rerio genes (Zv9) and release 68 of A. thaliana genes (TAIR 10). The full-length sequences of C. elegans transcripts were obtained from Wormbase (WB190). Sequences of 50 bp,

flanking the transcript start sites, end sites and splice sites of C. elegans, were obtained from Ensembl (WB220).

Calculation of thermodynamic stability

ΔG of the nearest-neighbour interactions was calculated by Perl-based software using Kowalski's sliding-window approach[2]. Published values of ΔH and ΔS (at 37 °C and 1 M salt concentration) for each nearest-neighbour interaction for DNA/DNA duplexes[3] and RNA/DNA duplexes[4] were used. Calculations were carried out with a step size of 1 bp and a window size of 9 bp.

1. Nedelcheva-Veleva, M. N., Sarov, M., Yanakiev, I., Mihailovska, E., Ivanov, M. P., Panova, G. C., & Stoynov, S. S. (2013). The thermodynamic patterns of eukaryotic genes suggest a mechanism for intron–exon recognition. Nature Communications, 4, 2101.
2. Huang, Y. & Kowalski, D. *WEB-THERMODYN: Sequence analysis software for profiling DNA helical stability*. *Nucleic Acids Res.* **31**, 3819–3821 (2003).
3. SantaLucia, J. Jr. *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*. *Proc. Natl Acad. Sci. USA* **95**, 1460–1465 (1998).
4. Sugimoto, N. *et al. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes*. *Biochemistry* **34**, 11211–11216 (1995).