PERSPECTIVES: TRANSCRIPTION

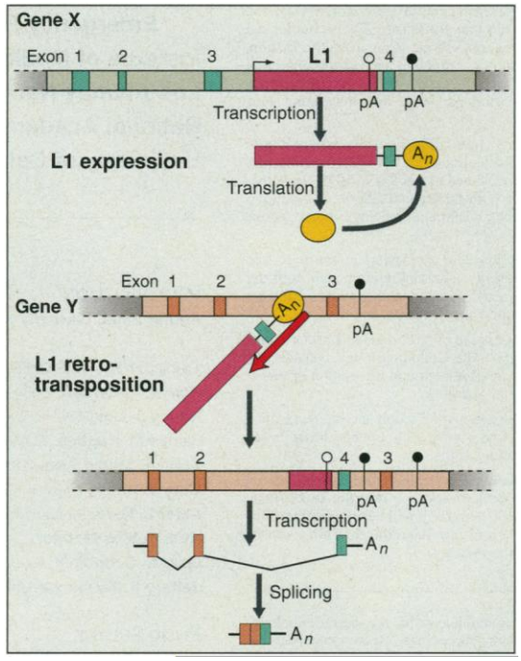
Exon Shuffling in Retrospect

Tom Eickbush

Retrotransposons are short sequences of DNA (ranging in length from a few hundred to tens of thousands of base pairs) that make new copies of themselves by reverse transcription of an RNA intermediate. These mobile DNA elements have had a profound influence in shaping eukaryotic genomes. For example, nuclear spliceosomal introns (which remove the introns from transcribed RNA) (*1*), retroviruses (*2*), and even telomerase (the reverse transcriptase that maintains the ends of eukaryotic chromosomes) (*3*) may all have once been retrotransposons. The very abundance of these mobile elements is remarkable. Thirty percent of the human genome results from this sort of reverse transcription (*4*), and it is likely that another significant fraction comprises older elements that have mutated into obscurity. Although these insertions of DNA from elsewhere in the genome occasionally give rise to deleterious mutations, it is not clear how much burden this excess DNA has placed on the overall functioning of the human genome. Indeed, the results published on page 1530 of this issue (*5*) suggest that retrotranspositions in the human genome can shuffle exons, a consequence that can be considered beneficial.

The most abundant retrotransposons in the human genome are the LINE-1 (L1) elements. They belong to the non–long-terminal repeat (non-LTR) class of retrotransposons, which integrate into the genome by an inherently sloppy mechanism. Insertion of non-LTR elements is typically accompanied by major 5′-end truncations, with only a polyadenylation [poly(A)] tract at the 3′ end common to all copies. Current models of non-LTR retrotransposition are largely based on the analysis of R2 elements (retrotransposons that insert specifically in the 28S ribosomal RNA genes of arthropods), made possible by the high specificity of this element's reverse transcriptase and endonuclease for its RNA template and insertion site (*6*). These studies suggest that non-LTR retrotransposition is initiated by cleavage of the target site with the 3′ end of one of the released DNA strands used to prime reverse transcription starting at the 3′ end of the element's transcript. How the element is attached upstream of the target site and the second strand of DNA then synthesized, is poorly understood.

Recently, Moran and his co-workers developed an efficient L1 retrotransposition assay in cultured human cells using an L1 element marked with a selectable *neo* gene stably expressed from an episome (*7*). The L1 insertions generated at various chromosomal sites by this expression system have all the hallmarks of authentic non-LTR retrotransposition events. An unexpected feature of these insertions was that the poly(A) sequences terminating the L1 insertions were not derived from the L1 poly(A) signal, but from an SV40 signal located on the episome a short distance downstream of the L1 element. Thus, the L1 element's own poly(A) signal is weak, but bypassing it has few consequences because the element can use downstream poly(A) signals. This weak specificity of the L1 retrotransposition machinery for its own 3′ end is responsible for the insertion of Alu sequences (reverse transcripts of an abundant, short cellular RNA) and processed pseudogenes (reverse transcripts of spliced mRNAs) in the human genome (*6*, *8*).



Exon retroshuffling. Transcription of an L1 element within gene X bypasses its own weak polyadenylation (pA) signal, instead terminating at the pA signal of the gene. The reverse transcriptase and endonuclease (yellow) encoded by this transcript bind to the poly(A) tail (A_n) and insert a cDNA copy of the element's RNA (red arrow) into gene Y. This step results in the transduction of exon 4 of gene X into gene Y, creating a new gene.

As reported in this issue (*5*), Moran and colleagues now take their finding a step further by demonstrating that L1 is capable of transducing downstream sequences. The *neo* reporter gene in the new assay was placed entirely downstream of L1, and both the promoter and initiation codon of the *neo* gene were replaced with a 3′ splice site. The only means by which this modified *neo* gene could be expressed after transduction was if it was inserted into an actively transcribed gene and spliced onto the transcript derived from that gene. Such events were readily detected and characterized as authentic L1 insertions in several genes. Indeed, the frequency at which the transduced *neo* gene was expressed was consistent with that expected from the fraction of the human genome occupied by genes (correcting for the proper orientation and reading frame needed for expression).

This dramatic result has two clear implications: L1 does not avoid becoming inserted into introns, and while doing so it can readily transduce a downstream exon (encoding part of the gene). This result suggests a simple new mechanism by which exons can be shuffled in the human genome (see the figure). Transcription of L1 within an intron of gene X will often bypass the weak poly(A) sequence of the L1 element. If retrotransposition of this transcript occurs in the intron of gene Y, a new gene will be created, which has the upstream exon sequences of gene Y spliced onto the transduced terminal exon of gene X. The remarkable aspects of this exon "retroshuffling" mechanism are that homologous sequences are not required, the relative positions of genes X and Y are irrelevant, and the donor gene remains unaffected. The transduction of downstream sequences has also been previously documented with endogenous L1 insertions of recent and ancient origin (*9*–*11*). Many additional transduction events have probably been missed because the extensive 5′ truncations accompanying most L1 insertion events may have eliminated traces of the L1 element responsible for the DNA shuffling.

The author is in the Biology Department, University of Rochester, Rochester, NY 14627–0211, USA. E-mail: eick@uhura.cc.rochester.edu

These new findings clearly add fuel to the debate over whether transposable elements can be a beneficial component of eukaryotic genomes. At first glance, selfish gene models for transposable element maintenance do not predict that an element would be so indiscriminate in its choice of 3′ ends for retrotransposition. Transducing downstream sequences would be of little benefit to the element itself. One should consider, however, why L1 might find it advantageous to avoid its own poly(A) site. As pointed out by Moran *et al.*, the likely reason L1 elements have weak poly(A) signals is that it enables them to reside in the introns of a gene with little effect; strong poly(A) signals would disrupt expression of the gene. Thus, L1 is under conflicting pressure to retain its own poly(A) site, without that site being too "powerful." Although the occasional shuffling of an exon may be an advantage to the host, the element itself is simply increasing its chances of survival.

This work proffers the prediction that non-LTR elements in genomes with a high percentage of intron sequences will have weak poly(A) signals. Elements in genomes with a low percentage of introns will have strong poly(A) signals and will seldom transduce downstream sequences. It also suggests that it is not only Alu sequences and processed pseudogenes that may owe their origin to L1, but just about any insertion ending with an A-rich tail. The fraction of the human genome attributable to retrotransposition just got larger.

References and Notes

1. P. A. Sharp, *Science* **254**, 663 (1991).
2. Y. Xiong and T. H. Eickbush, *EMBO J.* **9**, 3353 (1990).
3. T. M. Nakamura *et al.*, *Science* **277**, 955 (1997).
4. A. F. A. Smit, *Curr. Opin. Genet. Dev.* **6**, 743 (1996).
5. J. V. Moran, R. J. DeBerardinis, H. H. Kazazian Jr., *Science* **283**, 1530 (1999).
6. D. D. Luan, M. H. Korman, J. L. Jakubczak, T. H. Eickbush, *Cell* **72**, 595 (1993).
7. J. V. Moran *et al.*, *ibid.* **87**, 917 (1996).
8. H. H. Kazazian Jr. and J. V. Moran, *Nature Genet.* **19**, 19 (1998).
9. S. E. Holmes, B. A. Dombrowski, C. M. Krebs, C. D. Boehm, H. H. Kazazian Jr., *ibid.* **7**, 143 (1994).
10. J. C. McNaughton *et al.*, *Genomics* **40**, 294 (1997).
11. R. Rozmahel *et al.*, *ibid.* **45**, 554 (1997).
12. I thank H. Malik and B. Burke for helpful comments.

PERSPECTIVES: MOLECULAR SPECTROSCOPY

Ultrafast Glimpses at Water and Ice

Edwin J. Heilweil

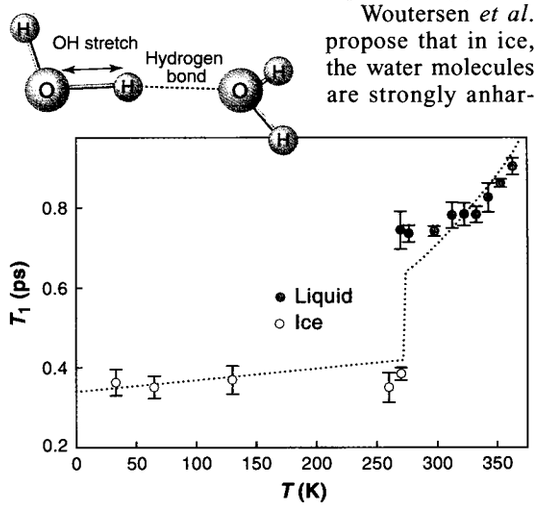
Even the properties of seemingly simple molecules can be rather elusive. Water—perhaps the most important medium for supporting chemical processes and biological functions—exhibits unusual properties, the origins of which remain a subject of debate. Compared with other molecular liquids and solids, water has a high dielectric constant and compressibility. Most notably, its density displays a maximum at several degrees above the freezing temperature (*1*). Now, modern spectroscopic techniques are beginning to decipher the origin of some of water's anomalous characteristics. Recent experimental and theoretical studies with transient infrared spectroscopy and molecular dynamics methods are providing interesting, although controversial, results on the microdynamics of water and ice that underlie the bulk properties.

Ultrafast mid-infrared spectroscopy (at wavelengths from 2 to 10 μm) has gained in popularity over the past 15 years because of its sensitivity to vibrational processes and orientation of molecules (*2*). A time-dependent infrared (IR) spectrum of a sample provides detailed information on molecular vibrational energy flow, conformational rearrangements, chemical reaction intermediates, and transitory tertiary structures of the system. For example, ultraviolet (UV) laser pulses with durations from picoseconds (1 ps = 10^{-12} s) to femtoseconds (1 fs = 10^{-15} s) tuned to a molecular electronic transition can cause that species to dissociate or react with other molecules. IR pulses of similar duration—either tuned to a discrete molecular vibration or broadband—can subsequently be used to probe the molecular absorptions originating from transient intermediates and stable products as a function of the delay between the earlier "pump" UV pulse and the subsequent "probe" IR pulse (pump-probe spectroscopy). Analysis of such time-dependent IR spectra allows one to assign the structure and monitor the time evolution of chemical species. Detailed mechanistic information and kinetic rates can thus be directly obtained for a dynamical system.

Such ultrafast IR technologies are now being used to study the microdynamics of liquid water and ice. Woutersen *et al.* recently published femtosecond IR measurements of the vibrational relaxation time (T_1) of the OH stretch in water and ice as a function of temperature (*3*). These results extended earlier spectroscopic and kinetics studies of room temperature water (*4, 5*).

Woutersen *et al.* reported vibrational population decay times (for the first excited, or $v = 1$ state) for the OH-stretching mode of HOD molecules diluted in D_2O (see figure) (*3, 4*). The decay time of vibrations depends on the molecular environment. For example, the OH vibrations may be coupled to other atomic motions in the system, such as hydrogen bonds between water molecules. The latter may take up the energy of vibrational modes. According to accepted theories (*6*), one would have expected the decay times to decrease with increasing temperature, because the number of lower, energy-accepting modes usually increases with temperature. However, Woutersen *et al.* observed a different behavior (see figure). They found that T_1 was nearly constant for the ice phase (30 to 270 K), abruptly increased at the solid-liquid phase-transition at 273 K, and then increased monotonically up to 363 K. Only a few inorganic condensed-phase systems are known to display such an inverted T_1 temperature dependence (*7*).

Woutersen *et al.* propose that in ice, the water molecules are strongly anhar-



Ultrafast vibrations. Vibrational lifetime (T_1) of the OH-stretching mode of dilute HOD:D_2O as a function of temperature. The dashed curve is calculated by using the hydrogen-bond frequency from the OH-stretch redshift (*11*). [Reprinted with permission of the authors (*3*)]

The author is in the Optical Technology Division, Physics Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8441, USA. E-mail: edwin.heilweil@nist.gov