Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 2037–2038, March 1998

Commentary

A base called J

Larry Simpson

Howard Hughes Medical Institute and Departments of Cell, Molecular, and Developmental Biology, and Medical Microbiology, Immunology, and Molecular Genetics, University of California at Los Angeles, Los Angeles, CA 90095

Nucleotide modifications in DNA are infrequent enough, and their role is of sufficient controversy and possible biological importance, that the discovery of a novel modification is a cause for celebration to modification aficionados and of great interest to everyone else. In previous publications from the Borst laboratory, a novel DNA base, β-D-glucosyl-hydroxymethyluracil, was found to replace approximately 0.5% of the thymine in the DNA of African trypanosomes (1–4). They designated this modified base by the somewhat unpretentious but nevertheless catchy term, J.

The African trypanosomes, which are the causal agents of African Sleeping Sickness in humans, undergo in the bloodstream of the mammalian host a selective expression of variant surface glycoprotein (VSG) genes as a mechanism to evade the immune response of the host (5). The trypanosomes have up to 20 different telomeric-localized expression sites and more than a hundred VSG genes, and are able to modulate VSG expression by several mechanisms, including replacing the VSG gene in the active expression site or activating a new expression site and silencing the old site (6). The mechanism for silencing all but one VSG sites and for the inheritance of these stable transcriptional states is unknown but is thought to involve an epigenetic mechanism because it is independent of promoter sequences. The original discovery of J stemmed from the observation that silenced telomeric genes could not be digested with restriction enzymes such as *Pvu*II, *Pst*I, *Hin*dIII, and *Sph*I (7, 8). Using ^{32}P-nucleotide postlabeling and a polyclonal antiserum that specifically recognizes J-containing DNA, it was shown that more than half of the J nucleotides in *Trypanosoma brucei* bloodstream cells were located at the GGGTTA telomeric repeats and repetitive sequences at the boundaries of silenced VSG expression sites, and were not present in actively transcribed telomeric VSG genes (4). Interestingly, J was not detected in procyclic *T. brucei*, which represent the insect midgut stage of the life cycle and which do not express VSG genes. These results led to the theory that J was somehow involved in expression site control in bloodstream *T. brucei*, either as a primary cause of expression site inactivation by modifying chromatin structure, or as a consequence of silencing and playing a role in the maintenance of the repressed state (4).

In the paper by van Leeuwen *et al.* in this issue of the *Proceedings* (9), an evolutionary test of this hypothesis was performed, and the results raised more questions than they provided answers. Using the J-specific antibody to enrich for modified molecules and ^{32}P-postlabeling to confirm the presence of J bases, the level of sensitivity was approximately 1 in 10^7 bases. By this method, the presence of the modified J base was established in representatives of all kinetoplastid genera, including both monogenetic (parasitic in one host) and digenetic (parasitic in two hosts) species, and also in a free-living flagellate, *Diplonema*. J was not detected in DNA from any other organisms, including the higher eukaryotes, humans, Drosophila, nematodes, and yeast; the deep-branching amitochondrial eukaryotes, *Trichomonas*, *Giardia*, and *Entamoeba*; the mitochondria-containing apicomplexa; and the eubacterium, *Escherichia coli*. Interestingly, dinoflagellates, which belong to the same Alveolate protist clade as the apicomplexa, did contain three modified nucleotides that may be analogs of J.

The presence of J in kinetoplastid species that do not undergo antigenic variation is at first glance at variance with the hypothesis that J is involved with repression of the telomeric VSG genes during antigenic variation in the African trypanosome. van Leeuwen *et al.* (9) suggest that J could be involved in a generalized repression of transcription of repetitive sequences in other kinetoplastids or in a repression of recombination between repetitive sequences. This would be consistent with the hypothesis of Yoder *et al.* (10) that the major role of methylation in the DNA of eukaryotes is not developmental gene regulation, as has been hypothesized (11, 12), but rather the control and transcriptional repression of intragenomic parasitic sequences such as retroviruses, Alu elements, and transposons. It is suggested that J may play a similar role in kinetoplastids. To explain the correlation of an absence of nucleotide modification in the procyclic phase of *T. brucei* (7) and an absence of VSG gene regulation, it is suggested that the African trypanosomes might have recruited the global repression mechanism mediated by J for induction or maintenance of silencing of the VSG gene expression sites. This hypothesis, of course, awaits the genetic test of the identification and knockout of enzymes involved with biosynthesis of J.

The finding that J appears to be a molecular marker for the kinetoplastid protozoa, including both the well studied human, animal, and plant parasites in the trypanosomatid group, and at least one parasitic species in the poorly studied sister group, the bodonids/cryptobiids, is intriguing. The presence of J in *Diplonema* is interesting and is consistent with a tentative phylogenetic reconstruction using rRNA sequences that places these cells as monophyletic with kinetoplastids (D. Maslov, S. Yasuhira, and L.S., unpublished results). The kinetoplastid protozoa are a large group of lower eukaryotic cells that diverged from the euglenoid flagellates approximately 500 million years ago (13–16). These cells represent one of the earliest extant eukaryotic lineages possessing mitochondria and, probably as a result of this ancient divergence, exhibit many unusual and even some novel biological phenomena. For example, all cytosolic mRNAs have a 39-nt spliced leader sequence that is added to the 5′ end by trans-splicing (17–21). All glycolytic enzymes are contained within a peroxysomal-related membrane-bound organelle, the glycosome, rather than being free in the cytosol (22). The mitochondrial genome is present as a giant network of thousands of catenated minicircles and maxicircles (23). Finally, a remarkable type of RNA editing occurs in the mitochondrion, in which the transcripts of more than half of the maxicircle genes are modified after transcription by the insertion and deletion of multiple uridine residues at specific sites, usually within coding regions, giving rise to translatable mRNAs (24, 25). Even more remarkable is that the information for these editing events is

contained within a class of small guide RNAs, which are mainly transcribed from the minicircles (26, 27). This existence of two genomes, one containing G-rich cryptogenes and the other, genes (28) whose transcripts encode the sequence information for the U insertions and deletions, is unique to the kinetoplastids. In spite of the fact that homologues (or analogs) of most of these unusual phenomena have been found to exist in other cells and that a deeper knowledge of more protista, which represent the deepest divisions of the eukaryotes (29) and about which we know very little, will almost certainly provide equally startling phenomena, nevertheless the plethora of novel biological structures and activities that occur in kinetoplastids is striking. Any molecular marker that ties together these various cells may prove very informative from an evolutionary or functional point of view, and J may provide just that marker.

1. Gommers-Ampt, J. H., van Leeuwen, F., de Beer, A. L., Vliegenthart, J. F., Dizdaroglu, M., Kowalak, J. A., Crain, P. F. & Borst, P. (1993) *Cell* **75,** 1129–1136.
2. Gommers-Ampt, J., Lutgerink, J. & Borst, P. (1991) *Nucleic Acids Res.* **19,** 1745–1751.
3. van Leeuwen, F., Wijsman, E. R., Kuyl-Yeheskiely, E., van der Marel GA, van, B. J. & Borst, P. (1996) *Nucleic Acids Res.* **24,** 2476–2482.
4. van Leeuwen, F., Wijsman, E. R., Kieft, R., van der Marel, G. A., van Boom, J. H. & Borst, P. (1997) *Genes Dev.* **11,** 3232–3241.
5. Cross, G. A. (1996) *BioEssays* **18,** 283–291.
6. Pays, E., Vanhamme, L. & Berberof, M. (1994) *Annu. Rev. Microbiol.* **48,** 25–52.
7. Pays, E., Delauw, M. F., Laurent, M. & Steinert, M. (1984) *Nucleic Acids Res.* **12,** 5235–5247.
8. Bernards, A., van Harten-Loosbroek, N. & Borst, P. (1984) *Nucleic Acids Res.* **12,** 4153–4170.
9. van Leeuwen, F., Taylor, M. C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R. & Borst, P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2366–2371.
10. Yoder, J. A., Walsh, C. P. & Bestor, T. H. (1997) *Trends Genet.* **13,** 335–340.
11. Holliday, R. & Pugh, J. E. (1975) *Science* **187,** 226–232.
12. Lengauer, C., Kinzler, K. W. & Vogelstein, B. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 2545–2550.
13. Maslov, D. A., Lukes, J., Jirku, M. & Simpson, L. (1996) *Mol. Biochem. Parasitol.* **75,** 197–205.
14. Maslov, D. A. & Simpson, L. (1995) *Parasitol. Today* **11,** 30–32.
15. Simpson, L. & Maslov, D. A. (1994) *Curr. Opin. Genet. Dev.* **4,** 887–894.
16. Simpson, L. & Maslov, D. A. (1994) *Science* **264,** 1870–1871.
17. Sather, S. & Agabian, N. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 5695–5699.
18. Campbell, D., Thornton, D. & Boothroyd, J. (1984) *Nature (London)* **311,** 350–355.
19. De Lange, T., Michels, P., Veerman, H., Cornelissen, A. & Borst, P. (1986) *Nucleic Acids Res.* **12,** 3777–3790.
20. LeBowitz, J. H., Smith, H. Q., Rusche, L. & Beverley, S. M. (1993) *Genes Dev.* **7,** 996–1007.
21. Ullu, E. & Tschudi, C. (1990) *Nucleic Acids Res.* **18,** 3319–3326.
22. Opperdoes, F. R. (1990) *Biochem. Soc. Trans.* **18,** 729–731.
23. Simpson, L. (1986) *Int. Rev. Cytol.* **99,** 119–179.
24. Alfonzo, J. D., Thiemann, O. & Simpson, L. (1997) *Nucleic Acids Res.* **25,** 3751–3759.
25. Simpson, L. & Thiemann, O. H. (1995) *Cell* **81,** 837–840.
26. Blum, B., Bakalara, N. & Simpson, L. (1990) *Cell* **60,** 189–198.
27. Sturm, N. R. & Simpson, L. (1990) *Cell* **61,** 879–884.
28. Simpson, L. (1997) *Mol. Biochem. Parasitol.* **86,** 133–141.
29. Sogin, M., Gunderson, J., Elwood, H., Alonso, R. & Peattie, D. (1989) *Science* **243,** 75–77.