*Journal of Biomolecular Structure &
Dynamics, ISSN 0739-1102
Volume 15, Issue Number 4, (1998)
©Adenine Press (1998)*

Mammalian Retroposons Integrate at Kinkable DNA Sites

http://www.albany.edu/chemistry/sarma/jbsd.html

**Jerzy Jurka[1]*,
Paul Klonowski[1]
and Edward N. Trifonov[2]**

[1]Genetic Information Research Institute,
440 Page Mill Road, Palo Alto, CA
94306, U.S.A.

[2]Department of Structural Biology,
Weizmann Institute of Science,
Rehovot 76100, Israel

Abstract

Integration of retroposed RNA in mammals occurs at staggered breaks resulting from an enzyme-generated pair of nicks at opposite DNA strands, preferably within 15-16 bp. Although consensus sequences associated with the two nicks appear somewhat different from one another, both nicking sites are rich in TA, CA and TG dinucleotide steps which are known as specific DNA sites where kinks may occur under bending constraints. This suggests that during interaction with the endonucleolytic enzyme, or enzymes, DNA undergoes bending at the integration sites and kinks are formed, as initial steps in generating the nicks. Nicking at kinkable sites, particularly at TA steps, may also play a role in integration of other insertion elements.

Introduction

Non-viral retroposons, also referred to as non-LTR retrotransposons, originate via reverse transcription of cellular RNA. They are usually flanked by direct repeats, resulting from duplication of their target sites. The target site is defined as a stretch of DNA delimited by a pair of nicks generated at both strands of the double-stranded DNA. The sequences at which the nicks occur are nonrandom (1) and are believed to be recognized and cut by an endonucleolytic domain (EN) of the reverse transcriptase encoded by the ORF2 of L1 elements (1, 2). The sequence signal where the first of the two nicks is likely to be generated is represented by the 5′ TTAAAA/3′ AATTTT consensus sequence. This consensus is somewhat different and more robust statistically than the consensus sequence associated with the second nick. However, our study indicates that in spite of these apparent differences the underlying structures of both nicking sites seem to be essentially the same since both of them are compatible with the potential formation of DNA kinks, defined as abrupt deflections of the DNA axis leading to unstacking of two neighboring base pairs (3). This observation sheds a new light on the mechanism of target recognition and DNA cleavage in retroposon integration.

Materials and Methods

We have analyzed the dinucleotide composition of 344 perfectly duplicated Alu targets over 9 bp in length and their adjacent regions as described before (1). The sequences are available over the internet (http://www.girinst.org/~server/ publ). In the first of the two analyzed sets, all 5′ copies of duplicated targets start at the same position and are extended by an additional 20 bp towards their 5′ ends, wherever possible. In the second set all 3’ duplicated targets end up at the same position and are 3′ extended by 20 bp.

Results and Discussion

Figures 1a and 1b show cumulative frequencies of three selected groups of dinu-

*Author to whom correspondence should be addressed. Phone: (650) 326-5588; Fax: (650) 326-2001; E-mail: jurka@gnomic.stanford.edu cleotides around both ends of the duplicated targets extending to the adjacent regions. The three groups are: purine doublets - AA, AG, GA, GG, pyrimidine doublets - TT, TC, CT, CC, and dinucleotides associated with kinks - TA, CA, TG (3-5). Both figures show very similar patterns of dinucleotide frequencies with kink-associated doublets reaching maximum at both nicking sites, preceded by enhanced frequencies of di-pyrimidines and followed by peaks of di-purines.

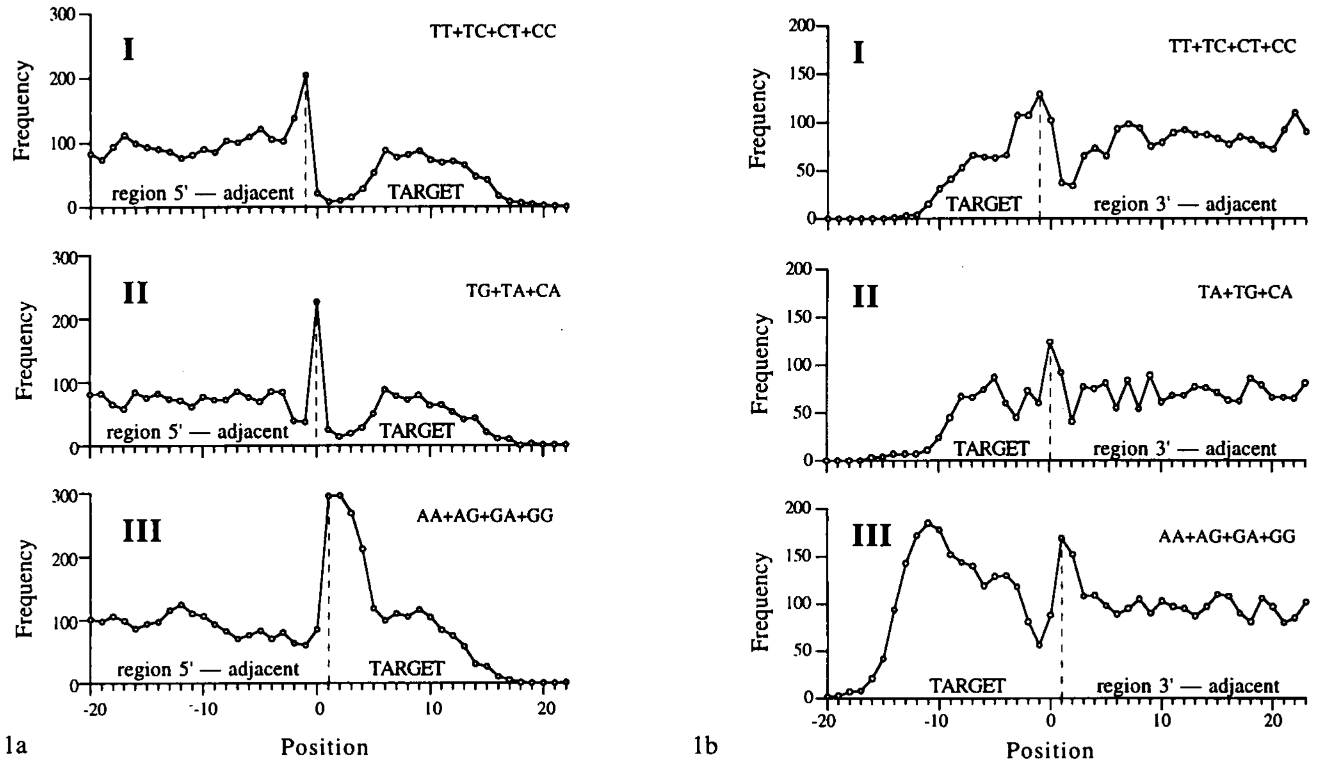

Figure 1: (a) Distribution of dinucleotides associated with the potential DNA kink around the 5′-ends of the duplicated targets and in the 5′-adjacent regions. (b) Distribution of dinucleotides associated with the potential DNA kink around the 3′-ends of the duplicated targets and in the 3′-adjacent regions. The dinucleotides corresponding to each section of the figure are indicated. The larger of the two peaks in Figure 1b (part III) corresponds to the A-rich target region near the first nicking site (Figure 1a, part III), which may be involved in interaction of poly(A) tails of the retroposed RNA (cf. 19).

The exact positions of the nicking sites associated with the kink signal from Figure 1a cannot be determined in most cases since the A-rich 5′ ends of the duplicated targets merge with the poly(A) tail of the retroposon as illustrated by the following consensus sequence of the integration target (1).

5′ TT|AAAANNNNNNNTYTN{retroposon}....AAAA|AAAANNNNNNNTYTN 3′.

In previous work (1) the nick was assumed to occur predominantly between 5′ TT and the following 5′AAAA. The second nick (Figure 1b), occurs primarily either at the kink site or downstream. Since the presence of the potential kink sites is the feature common to both ends of the target region, this suggests that in both cases the borderline is located at or within a few bases from the kink sites.

The two nicking sites are of different strength and contain different proportions of the kinkable YR dinucleotides. To illustrate this we listed in Tables I and II, the frequencies of dinucleotides around the strong and weak nicking sites, respectively, 7 bp in either direction. Kinkable YR dinucleotides dominate the first cutting site: 171 TA, 32 CA and 23 TG (Table I). At the second site there are 52 TA, 44 CA and 28 TG (Table II). The overall ratio of these dinucleotides in the first and the second site is almost 2:1 (226:124), confirming the relative weakness of the second site. In both cases, however, the dinucleotides TA, TG and CA are among those which in the cutting positions exceed their respective averages (not shown), in the -7 to +7 region (Tables I and II).

Another interesting observation is that CpG doublets are rare within the target as compared to the non-target region. This is particularly visible in Table II. Separate analysis (data not shown), indicates that the ratio of non-target to target CpG doublets is close to 5:1. This, however, may be related to either bendability or to CpG methylation which may interfere with target recognition by the endonuclease.

The pattern of di-pyrimidines followed by di-purines, 4-5 bases downstream (see Figure 1a and Table I), resembles distribution of the dinucleotides in the nucleosome DNA where TT and other YY dinucleotides are similarly located at outward positions of the nucleosome DNA, while AA and other RR dinucleotides are located preferentially at the DNA-histone interface, about 5 bases downstream from the YY dinucleotides (6, 7). Such a pattern would place the kinkable steps TA, CA and TG at the minor groove facing the protein. This suggests that the EN may bind the respective DNA site in a manner similar to histones and cause the kinking/nicking.

Table I

Numerical distribution of all dinucleotides in the immediate vicinity of the 5′-ends of the duplicated targets, corresponding to the strong nicking site. The second column indicates the most abundant (consensus) dinucleotide and an asterisk indicates potential DNA kink sites in both Tables.

From	Cons.	To	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
-7	AT	-6	32	28	22	29	34	15	25	2	35	16	24	15	22	8	16	21
-6	TT	-5	52	21	24	26	33	15	19	0	26	16	34	11	21	8	24	14
-5	TT	-4	49	19	27	36	25	12	22	1	31	19	36	15	20	12	12	7
-4	TT	-3	53	16	40	16	22	11	28	1	36	13	30	18	20	7	17	15
-3	TT	-2	83	15	14	19	35	5	6	1	54	8	37	16	35	6	3	7
-2	TT	-1	156	22	17	12	19	7	8	0	12	4	37	7	23	4	9	7
-1	*TA	0	15	1	171	23	5	0	32	0	6	1	57	7	4	1	19	2
0	AA	1	8	0	19	3	0	0	3	0	11	3	199	66	1	1	27	3
1	AA	2	5	3	8	4	2	0	2	0	12	4	168	64	6	2	60	4
2	AA	3	7	4	7	7	4	0	5	0	20	7	175	36	12	3	54	3
3	AA	4	16	5	9	13	5	2	6	1	32	22	148	39	12	9	21	4
4	AA	5	28	10	15	12	7	8	23	0	53	29	64	38	28	13	12	4
5	AA	6	35	19	35	27	17	17	26	0	26	16	49	23	19	8	17	10
6	AA	7	33	15	23	26	18	11	29	2	40	12	49	26	14	11	15	20

In the recently published model of retroposon integration (1), it has been proposed that, originally, the nicking enzyme is attracted to the stronger site with the consensus sequence 5′ TTAAAA/3′ AATTTT. The selection of the second, weaker site is not likely to be guided by the sequence signal alone, but also by the distance from the first site and it may be forced by the enzyme itself even at sites poorly resembling a perfect kinkable sequence. The possibilities include a single enzyme with two similar active sites, a dimer of two identical enzymes each contributing a single active site, or a single-site enzyme cutting two nearby targets sequentially. We favor the first two possibilities as most compatible with the relatively well-defined distance limits between the nicks (1).

Although the kinks occur exactly at TA/TA, CA/GT and GT/CA steps, the local unwinding of DNA actually involves several base pairs in either direction from the kink (DNA "breathing"). Therefore, the single strand cut is expected to occur any-

Table II

Numerical distribution of all dinucleotides in the immediate vicinity of the 3′-ends of the duplicated targets, corresponding to the weak nicking site. For further information see legend to Table I and the text.

From	Cons.	To	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
-7	AA	-6	21	13	20	26	21	9	28	0	21	23	52	22	16	27	35	10
-6	AA	-5	27	9	17	26	18	9	44	1	26	15	61	33	16	7	24	11
-5	AA	-4	31	15	24	17	11	9	19	1	35	21	60	30	15	16	21	19
-4	AA	-3	52	13	13	14	36	6	18	1	33	15	61	15	16	9	27	15
-3	TT	-2	47	36	24	30	14	10	19	0	40	14	44	21	17	12	9	7
-2	TT	-1	38	31	18	31	38	22	11	1	38	22	13	23	23	15	7	13
-1	*TA	0	43	14	52	28	24	21	44	1	9	7	16	17	9	4	37	18
0	AA	1	15	6	33	31	9	7	28	2	15	12	82	38	9	5	26	23
1	AA	2	12	6	19	11	10	6	11	3	47	24	79	19	23	17	40	14
2	AA	3	24	18	21	29	13	10	27	3	34	28	47	39	12	12	9	13
3	AA	4	32	9	18	23	21	11	34	2	28	16	36	24	17	16	27	22
4	AT	5	27	11	34	25	17	10	22	3	36	18	29	31	21	11	24	14
5	TT	6	44	17	17	23	21	11	15	3	25	23	35	25	26	18	14	15
6	TT	7	38	21	33	24	21	18	27	3	20	12	34	15	15	5	26	20

where within the unwound region. This is consistent with in vitro digestion experiments by the endonuclease portion of the L1-encoded reverse transcriptase (2), showing frequent nicking both at and near 5′ TA. Similar conclusions can be obtained from sequence data analysis of DNA targets for retroposition (1). The "breathing" may explain why the kinkable steps are not as frequent at the second cutting site: the initial kinking at this site could actually occur in its close vicinity rather than exactly at the site. Indeed, the peak in Figure 1b (II) is broader as compared to the main peak of the distribution of the kinkable steps around the first site (Figure 1a, II).

It would be important to reconstruct the sequence of events at the initial stages of the retroposition. Does sequence recognition precede the bending and kinking? Or are the kinks formed transiently due to thermal motion and then fixed by EN binding? The former scenario is more appealing since the sequence in close vicinity to the kinkable site suggests DNA bending similar to the one caused by histones, as discussed above.

Endonuclease cleavage at kinkable sites, particularly at TA steps, is likely to be a more general phenomenon not necessarily associated with retroposition. For example, the crystal structure of EcoRV restriction endonuclease and its complexes with GGGATATCCC shows a kink of approximately 50 degrees at the TA step (5). Although TA steps may be crucial for kink formation, the sequences preceding and following the kinkable steps appear to vary, pending particular enzymes involved. It is well known that TA doublets are common targets for a large variety of transposons from bacteria to primates (8-15). However, sequences around the TA targets have been shown to converge to significant consensus patterns only for some transposons (15). Given the overall picture, we may postulate that endonucleases mediating transposon integration, as well as, at least some restriction endonucleases, generate kinks primarily at TA and, less frequently, at CA and TG steps in a large variety of sequence contexts. These contexts may determine different energy barriers for different endonucleases to overcome in order to generate the kinks and to nick the DNA. It must also be pointed out that DNA kinking at non-YR dinucleotide steps is possible as well but it is thermodynamically unfavorable and, therefore, less frequent.

Bending of the target DNA, although not kinking, seems to be an intrinsic feature associated with retroviral integration (16-18), which preferrably occurs at nucleosomal sites. This may indicate fundamental similarities between the mechanisms of integration of retroviruses and non-LTR retrotransposons.

Acknowledgements

This work was supported in part by the U.S. Department of Energy, Grant DE-FG03955ER62139 and by Grant 1 P41 LM06252-01 from the National Institutes of Health. We thank Jolanta Walichiewicz for help with the preparation of this manuscript.

References and Footnotes

1. J. Jurka, *Proc. Nat. Acad. Sci. 94*, 1872-1877 (1997).
2. Q. Feng, J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke, *Cell 87*, 905-916 (1996).
3. P. T. McNamara, A. Bolshoy, E. N. Trifonov and R. E. Harrington, *J. Biomolecular Struct. & Dynam. 8(3)*, 529-538 (1990).
4. S. C. Schultz, G. C. Shields and T. A. Steitz, *Science 253*, 1001-1007 (1991).
5. F. K. Winkler, D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos and K. S. Wilson, *EMBO J. 12(5)*, 1781-1795 (1993).
6. I. Ioshikhes, A. Bolshoy, K. Derenshteyn, M. Borodovsky, M. and E. Trifonov, *J. Mol. Biol. 262*, 129-139 (1996)
7. G. Mengeritsky and E. N. Trifonov, *Nucl. Acids Res. 11(11)*, 3833 (1983).
8. M. Amutan, E. Nyyssonen, J. Stubbs, M. R. Diaz-Torres and N. Dunn-Coleman, *Curr. Genet. 29*, 468-473 (1996).
9. D. C. Glayzer, I. N. Roberts, D. B. Archer and R. P. Oliver, *Mol. Gen. Genet. 249*, 432-438

(1995).

10. P. Kachroo, S. A. Leong and B. B. Chattoo, *Mol. Gen. Genet. 245*, 339-348 (1994).
11. T. Tenzen, Y. Matsuda, H. Ohtsubo and E. Ohtsubo, *Mol. Gen. Genet. 245*, 441-448 (1994).
12. T. Tenzen, S. Matsutani and E. Ohtsubo, *J. Bacteriol. 172*, 3830-3836 (1990).
13. T. Tenzen and E. Ohtsubo, *J. Bacteriol. 173*, 6207-6212 (1991).
14. M. Tudor, M. Lobocka, M. Goodell, J. Pettitt and K. O'Hare, *Mol. Gen. Genet. 232*, 126-134 (1992).
15. H. G. A. M. van Luenen and R. H. A. Plasterk, *Nucl. Acids Res. 22(3)*, 262-269 (1994).
16. P. M. Pryciak and H. E. Varmus, *Cell 69*, 769-780 (1992).
17. H.-P. Muller and H. E. Varmus, *EMBO Journal 13(19)*, 4704-4714 (1994).
18. D. Pruss, R. Reeves, F. D. Bushman and A. P. Wolffe, *J. Biol. Chem. 269(40)*, 25031-25041 (1994).
19. J. Jurka, and P. Klonowski, *J. Mol. Evol. 43*, 685-689 (1996).

Date Received: October 12, 1997

Communicated by the Editor Zippora Shakked