Independent Insertion of Alu Elements in the Human Ribosomal Spacer and Their Concerted Evolution

Iris Laudien Gonzalez, Richard Petersen, and James E. Sylvester
Pathology Department, Hahnemann University

A 2,700-bp segment of human ribosomal DNA (rDNA) spacer upstream of the rRNA promoter contains a set of four Alu elements, two in the direction of rRNA transcription and two in the opposite orientation. We report and compare the sequences of these Alu elements found in three rDNA clones and seek to determine the origin of the cluster, either from a single insertion followed by duplications or from multiple simultaneous or independent insertions. The high (20%-27%) divergence among members of a set and the lack of similarity/complementarity of sequences flanking different members of the set demonstrate the independent insertion of each of the four Alu elements into A-rich sequences on the appropriate strand of the rDNA. We also demonstrate that the Alu sets found in different rDNA repeats are subject to concerted evolution, yielding divergences of only 0.4%-3% between Alu elements in equivalent positions. However, the pairs of adjacent similarly oriented Alu elements do not show reduced divergence, indicating that there is no recombination or gene conversion between similarly oriented but not equivalently positioned Alu elements. Finally, crossing-over must occur in the rDNA junction region between Alu element 3 and the nonribosomal sequences at the telomere end of the acrocentric chromosome, so that the Alu elements of the terminal rDNA repeats and the terminal repeats themselves evolve in concert with the rDNA repeats located internally in the tandem array.

Introduction

Alu elements constitute the largest dispersed repetitive DNA family in the human genome. The availability of a large set of Alu sequence data has allowed classification of these elements into families and subfamilies, which probably arose during different rounds of amplification and dispersal (Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988). Comparison of individual Alu sequences to a derived consensus sequence yields an average difference of 17% (Jurka and Smith 1988). Alu elements in specific locations have been used as markers to study gene and genome evolution (Barsh et al. 1983; Ruffner et al. 1987) and have been implicated in frequent rearrangements and deletions in certain genes (Lehrman et al. 1987). The human ribosomal DNA (rDNA) spacer region contains several members of the Alu element family (Higuchi et al. 1981; Sylvester et al. 1986; Dickson et al., accepted). On the average, Alu elements appear every 5,000 bp, so it is not surprising that the ribosomal spacer contains such elements.

The 44-kb human rDNA is divided into four EcoRI-defined fragments, A, B, C and D [fig. 1(a)]. The 12-kb C fragment contains nontranscribed spacer sequences.

1. Key words: human rRNA, Alu elements, concerted evolution.

Address for correspondence and reprints: Iris L. Gonzalez, Pathology Department, MS 435, Hahnemann University, Broad and Vine, Philadelphia, Pennsylvania 19102; phone (215) 448-8646/8296.

© 1989 by The University of Chicago. All rights reserved.
0737-4038/89/0604-0008$02.00

413
The 3' end of the C fragment, a section bound by a BamHI site and by the EcoRI site, is designated as the CBE fragment and contains four Alu elements. Sylvester et al. (accepted) have sequenced and described two complete CBE fragments (clones C and F) which were cloned from different genomes and, presumably, from different chromosomes. In addition, a junction clone (clone T) containing C-fragment rDNA and nonribosomal sequences that belong to the telomere side of an acrocentric chromosome has been characterized by Worton et al. (1988) and sequenced by Sylvester et al. (accepted). A set of four Alu elements is found within a 2,700-bp stretch in each of the two complete CBE fragments, and a pair of Alu elements (half a set) is found in the junction clone [figs. 1(b), 1(c)]. We will show here that the four Alu elements were inserted separately into this part of the spacer. Further, our comparison of all the Alu elements in the three rDNA clones shows that the sets of four Alu elements are subject to concerted evolution, along with the rest of rDNA, and that these Alu elements do not recombine with similarly oriented but not equivalently positioned counterparts. Finally, we demonstrate frequent crossing-over which keeps the terminal rDNA repeat similar to the internal repeats, confirming the results of Worton et al. (1988) concerning the homogeneity of the p-arm termini of acrocentric chromosomes.

Material

Human rDNA fragments containing the Alu elements have been sequenced, and a composite is to be reported elsewhere (Sylvester et al., accepted). The region containing these elements is the 3' end of the EcoRI-defined "C" fragment of the rDNA repeat, bounded by a 5' BamHI site and by the 3' EcoRI site (CBE) [see fig. 1(a), 1(b)]. The sets of Alu elements designated as C1–C4 and F1–F4 were obtained from
CBE clones Cup and F1A described by Sylvester et al. (accepted). The set designated T3–T4 was obtained from the junction clone DJ15 described by Worton et al. (1988). The sequences of these previously unreported Alu elements were aligned with help of the MASE program from the NIH-supported Molecular Biology Computer Research Resource, Boston (Faulkner and Jurka 1988).

Results and Discussion

Figures 1(b) and 1(c) show the arrangement of the Alu elements within CBE and in the junction rDNA fragment. Alu 1 and Alu 2 are oriented in the direction of rDNA transcription, whereas Alu 3 and Alu 4 are in the opposite orientation. The junction rDNA fragment contains only Alu 3 and Alu 4 and, in fact, begins exactly at the border of Alu 2 (Worton et al. 1988; Sylvester et al., accepted). Figure 2 shows the sequence alignment of the Alu elements from our clones, together with the Alu consensus sequence of Jurka and Smith (1988).

Figure 3 contains the fraction of sites that differ between pairs of Alu elements, ordered by their position within the sets. It is immediately apparent that there is very little difference (0.4%–2.9%) between equivalently located Alu sequences belonging to different sets, whereas the differences between nonequivalents range from 17.1% to 27.8%.

The difference matrix allows calculation of a tree showing the relatedness among these sequences. This is shown in figure 4. The tree clearly shows the close relationship of equivalently positioned Alu elements and the separate branching of Alu elements 1 with Alu elements 2, and of the Alu elements 3 with Alu elements 4.

The arrangement of these Alu repeats could have originated in one of the following four ways:

1. Insertion of a single Alu element was followed by a nearby duplication and later was followed by an inverted duplication.
2. Insertion of a single source element, either into rDNA or elsewhere, was followed by transpositions into different sites in various rDNA repeats. The Alu elements were then brought together into a set within one rDNA repeat.
3. Four Alu elements were inserted into one rDNA repeat or were inserted into some other region and brought into rDNA as a unit.
4. The four distinct Alu elements were separately inserted at different sites within the same region of four different rDNA repeats. They were brought together by three crossing-over events.

All the models share the prediction that the set of four Alu elements spread to other rDNA repeats and was fixed by the mechanisms that lead to concerted evolution. These alternatives can be distinguished by examining similarities among Alu members and similarity/complementarity of the sequences that surround them.

Hypothesis 1 places the following requirements on the sequences: (a) All members of the set must be derived from each other by duplication after the original insertion and belong to the same Alu subfamily. (b) The sequences between Alu 1 and Alu 2 must be complementary to those between Alu 3 and Alu 4. (c) Some sequences flanking the pair Alu 1 and Alu 2 and the pair Alu 3 and Alu 4 must be similar, even if the duplication occurred long ago.

a. We can rule out hypothesis 1 on the basis of all these criteria. While all members of the set belong to the same larger S family as defined by Jurka and Smith (1988), Alu 1, Alu 3, and Alu 4 belong to the smaller Sd branch (extra A at position 264;
<table>
<thead>
<tr>
<th></th>
<th>CON</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-GCCGCGGCCGCGGGCTCGTACGCGCT</td>
<td>GGGCCCTCGCGGGCTCGTACGCGCT</td>
<td>GGGCCCTCGCGGGCTCGTACGCGCT</td>
<td>GGGCCCTCGCGGGCTCGTACGCGCT</td>
<td>GGGCCCTCGCGGGCTCGTACGCGCT</td>
</tr>
<tr>
<td>25</td>
<td>GTAATCCCAGCTACTTTGGGAGGCCG</td>
<td>GTAATCCCAGCTACTTTGGGAGGCCG</td>
<td>GTAATCCCAGCTACTTTGGGAGGCCG</td>
<td>GTAATCCCAGCTACTTTGGGAGGCCG</td>
<td>GTAATCCCAGCTACTTTGGGAGGCCG</td>
</tr>
<tr>
<td>50</td>
<td>GTCATCCCTCTATTTGGGAGGCCGCA</td>
<td>GTCATCCCTCTATTTGGGAGGCCGCA</td>
<td>GTCATCCCTCTATTTGGGAGGCCGCA</td>
<td>GTCATCCCTCTATTTGGGAGGCCGCA</td>
<td>GTCATCCCTCTATTTGGGAGGCCGCA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Cl</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>AGGGCCGGCTACCTGAG-GGTCGA</td>
<td>AGGGCCGGCTACCTGAG-GGTCGA</td>
<td>AGGGCCGGCTACCTGAG-GGTCGA</td>
<td>AGGGCCGGCTACCTGAG-GGTCGA</td>
<td>AGGGCCGGCTACCTGAG-GGTCGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>GGAGTTGGGAGACAGCCAGCAAC</td>
<td>GGAGTTGGGAGACAGCCAGCAAC</td>
<td>GGAGTTGGGAGACAGCCAGCAAC</td>
<td>GGAGTTGGGAGACAGCCAGCAAC</td>
<td>GGAGTTGGGAGACAGCCAGCAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>GAGGCCGGTGGACTGCAAGGCGT</td>
<td>GAGGCCGGTGGACTGCAAGGCGT</td>
<td>GAGGCCGGTGGACTGCAAGGCGT</td>
<td>GAGGCCGGTGGACTGCAAGGCGT</td>
<td>GAGGCCGGTGGACTGCAAGGCGT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Fl</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>ATGGTGAAACCCC--GTCTCTACT-</td>
<td>ATGGTGAAACCCC--GTCTCTACT-</td>
<td>ATGGTGAAACCCC--GTCTCTACT-</td>
<td>ATGGTGAAACCCC--GTCTCTACT-</td>
<td>ATGGTGAAACCCC--GTCTCTACT-</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>AAAAATACAAAA--ATTAG--CCGGGC</td>
<td>AAAAATACAAAA--ATTAG--CCGGGC</td>
<td>AAAAATACAAAA--ATTAG--CCGGGC</td>
<td>AAAAATACAAAA--ATTAG--CCGGGC</td>
<td>AAAAATACAAAA--ATTAG--CCGGGC</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>CACCACTCAGGCGCTACTCAAGC</td>
<td>CACCACTCAGGCGCTACTCAAGC</td>
<td>CACCACTCAGGCGCTACTCAAGC</td>
<td>CACCACTCAGGCGCTACTCAAGC</td>
<td>CACCACTCAGGCGCTACTCAAGC</td>
<td></td>
</tr>
</tbody>
</table>
Concerted Evolution of Alu Elements in rDNA

FIG. 2.—Alignment of Alu element sequences from the rDNA spacer. Cl–C4, Fl–F4, and T3–T4 are sets of Alu elements found in different rDNA repeats. F2, although completely present, was not sequenced entirely; however, it is identical to C2 in the 5' 121 bases, which include the region that contains most of the differences among Alu elements.

T244; A272), and Alu 2 belongs to the Se branch (C244; G272). Further, the phylogenetic tree that best fits the difference data (fig. 4) shows Alu 1 to branch together with Alu 2, and Alu 3 to branch with Alu 4, precluding a duplication of one pair to generate the other.

b. The sequences between Alu 1 and Alu 2 consist of (TAAA)₄–(TACA)₁₀(TAAA(A))₉ and are not complementary to those between Alu 3 and Alu 4 which consist largely of (GCTT)₉ and (TTTC)₉.

c. Although the Alu elements can be recognized as such after much divergence and although their sequences can be aligned, it is not possible to align the sequences
that flank both borders of the similarly oriented Alu pairs. This indicates that the sequences flanking similarly oriented Alu elements in the sets are not derived from one another by duplication.

We can conclude from these three points that the members of the set are not derived from each other by duplication after insertion into a single rDNA repeat.

Hypothesis 2 also requires that all Alu elements of a set be derived from the same source, and it can be ruled out on the basis of subfamily differences and on the basis of sequence divergence.

Hypothesis 3 implies that the four Alu elements were inserted as a set into a single rDNA repeat and that this rDNA arrangement was spread and fixed in all other repeats. The process responsible for this outcome is molecular drive (Dover 1982). In this case, one rDNA repeat of 200 in a haploid genome took over by chance. This hypothesis cannot be ruled out, although we will show that hypothesis 4 is more likely.

Hypothesis 4 proposes that the Alu elements could originally have inserted in-
dependently at different sites in A-rich regions of different rDNA repeats, possibly not all at the same time. This hypothesis does not require an initial high density of Alu insertions, although these are known, for example, in the human thymidine kinase and beta-tubulin genes (Lee et al. 1984; Flemington et al. 1987; Slagel et al. 1987). The set of four Alu elements had to be built up gradually by crossing-over. This was followed by spreading and fixation of this particular rDNA pattern. This is a more likely prospect than hypothesis 3, which requires simultaneous insertion of four Alu elements in the same rDNA repeat.

We will describe the facts and observations that are consistent with hypothesis 4 (see fig. 5).

There have been several waves of Alu element amplification and insertion (Willard et al. 1987; Britten et al. 1988). The human rDNA repeats, which probably numbered several hundred at the time of Alu insertion, contain A-rich sequences in the CBE fragment (Sylvester et al., accepted) and also in the nontranscribed spacer downstream from the 28S gene (La Volpe et al. 1985; Dickson et al., accepted). Alu elements inserted preferentially into such host regions (Daniels and Deininger 1985). The sequences in which these four Alu elements are embedded are as well conserved as the equivalently positioned Alu elements themselves, differing only in the numbers of short repetitive motifs, although the region between Alu 3 and Alu 4 varies in length among rDNA repeats, owing to different numbers of larger blocks (Sylvester et al., accepted). Thus, the whole region—Alu elements and surrounding sequences—evolves in concert in all rDNA repeats.

The rDNA contains hot spots for recombination. Recombination is known to occur in several parts of the spacer and is most easily recognized by spacer length variation involving repeated sequence blocks in specific locations (Erickson and Schmickel 1985). (Alu elements within such repeated sequence blocks could be duplicated along with them by unequal crossing-over, but we find no evidence of duplication in this set.) The tandem arrangement of rDNA repeats leads to recombination between different repeats, both on homologous and on nonhomologous chromosomes. Crossing-over can bring together Alu elements that were in different rDNA repeats (fig. 5). Certain crossing-over events may be favored by recombination hot spots. We have evidence for two such hot spots in the CBE fragment. One must be located in the region between Alu 2 and Alu 3, as demonstrated by the following two facts: (1) the terminal rDNA repeats, with their Alu elements, evolve in concert with the internal rDNA repeats on all acrocentric chromosomes (present paper; Sylvester et al., accepted); and (2) the nonribosomal sequences immediately adjacent to the terminal rDNA are found to be similar on all acrocentric chromosomes (Worton et al. 1988), an indication of exchanges among homologous and nonhomologous chromosomes.

A second site of favored recombination is located between Alu 3 and Alu 4. This region contains repeated blocks, 72–93 bp in length, that consist of simple repeat motifs separated by the tetramer CGTG. Unequal crossing-over would produce the observed length variation among the three clones described in the present paper: clone T has nine blocks, while clones C and F have five and six blocks, respectively (Sylvester et al., accepted). Southern blot analysis of genomic DNA also reveals this variation and shows that individuals differ in the proportions of the variants.

Our model requires that a third hot spot for recombination should exist in the region between Alu 1 and Alu 2, a region that consists of short repeats (TAMA), (M = A or C in the IUB code). The nature of this sequence would not allow us to distinguish between length variation that was due to either slipped-strand mispairing
1. Waves(s) of Alu element insertions at different sites within A-rich regions of different rDNA repeats. Can be different on all acrocentric chromosomes. May not be all at the same time.

2. Crossing-over between different rDNA repeats can bring together Alu elements from homologous and from non-homologous chromosomes.

During replication (Tautz et al. 1986; Levinson and Gutman 1987) or crossing-over (Smith 1973; Tartof 1973; Arnheim et al. 1980).

The Alu elements in CBE separate stretches of simple repetitive sequences. An Alu element may have inserted into such a stretch, splitting it into two "islands." These islands could then evolve independently, owing to geographical isolation. One now finds stretches of different simple motifs on either side of each Alu element; in other words, the non-equivalently located islands are less recognizable as once having been together than the Alu paralogues are recognized as being related. On the other hand, the equivalent islands differ only in (a) the number of repeats of the simple motifs possibly generated by slipped-strand mispairing during replication (Tautz et al. 1986; Levinson and Gutman 1987) and (b) the number of larger repeated blocks generated by recombination (Smith 1973; Tartof 1973; Arnheim et al. 1980).

The region surrounding Alu 1 is one segment where the motifs on both sides of the Alu element may still show relatedness: (TAMAA)_n is upstream and (TAMA)_n is downstream. The whole region could have consisted of either one of these motifs
3. Another round of crossing-over.

![Diagram](image)

**FIG. 5.**—Proposed origin and evolution of Alu elements in the CBE fragment of rDNA spacer. Only two rounds of crossing-over are shown. N-R = nonribosomal; HS = hot spot for recombination; CR = coding region. Alu elements are indicated by arrow heads and are labeled a–o.

before Alu insertion. Splitting, followed by a mutation in one island, was followed by homogenization within this island. It is possible that this relationship between flanking sequences indicates a relatively recent insertion of this Alu element.

The mechanisms responsible for concerted evolution of tandem genes eventually homogenized the rDNA repeats. Some Alu elements were lost in the process; others were preserved. We are seeing a set of survivors. Homogeneity is evidenced by the high similarity of equivalently positioned Alu elements and surrounding sequences in clones that were isolated from three different genomes.

Although recombination appears to be frequent in the CBE region, there is no apparent recombination or gene conversion among similarly oriented nonequivalent Alu elements in the set of four. The differences between the pairs Alu 1 and Alu 2 average 20%, and those between the pairs Alu 3 and Alu 4 average 17.8%, while the oppositely oriented pairs of Alu 1 and Alu 3 average 18.9%. The frequent recombination among paralogous Alu elements that leads to deletions in the low-density-lipoprotein receptor gene (Lehrman et al. 1987) is not seen in rDNA. This is probably due to preferred recombination at the rDNA hot spots.

**Conclusion**

We have demonstrated the independent insertion, in the rDNA spacer, of four Alu elements which are now subject to concerted evolution along with the rest of rDNA. The assembly of this set of Alu elements required crossing-over between different
rDNA repeats, which may have been favored by recombination hot spots that may still be active today and which may contribute to the maintenance of homogeneity among terminal and internal rDNA repeats. Whether selective pressure has helped to keep this arrangement for some unknown function remains to be determined. We expect to obtain more information about Alu-element insertions in rDNA as we characterize the remainder of the spacer, which is known to contain more of these elements.

**Sequence Availability**

These sequences have deposited in GenBank under accession numbers M22492-M22501.

**Acknowledgments**

We thank Walter Fitch for his very helpful suggestions in reworking the manuscript and for the phylogenetic tree. We also thank David Schlessinger for sharing unpublished material. This work was supported by the Pathology Department, Hahnemann University.

**LITERATURE CITED**


LEHRMAN, M. A., D. W. RUSSELL, J. L. GOLDSTEIN, and M. S. BROWN. 1987. Alu-Alu recom-
bination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. J. Biol. Chem. 262:3354-3361.


SYLVESTER, J. E., R. PETERSEN, and R. D. SCHMICKEL. Human ribosomal DNA: novel sequence organization in 4.5 kilobases upstream of the promoter. Gene (accepted).


WALTER M. FITCH, reviewing editor

Received December 13, 1988; revision received February 15, 1989