REVIEWS

Role of Retroposition in Autoregulation of Genomic Processes (Do Genes Program the Body and Retroposons Program the Genome?)

D. V. Bebikhov, A. Yu. Postnov, and T. A. Nikinenko

Translated from *Byulleten' Eksperimental'noi Biologii i Meditsiny*, Vol. 126, No. 7, pp. 4-14, July, 1998
Original article submitted November 17, 1997

Retroposition is the copying of genetic information from one locus of the genome to another locus. This process involves reverse transcriptase, one of the most ancient cellular enzymes. All retroposition elements are divided into retrotransposons (that code reverse transcriptase) and retroposons (that do not code this enzyme). There are specialized families of retroposition elements that selectively copy themselves into certain compartments or loci of the genome and ignore other loci and compartments. Retroposition can form the basis for regulation of genomic processes. The following facts are in agreement with this suggestion. Retroposition elements are present in each genome. The families of retroposition elements develop more rapidly than the genome as a whole. Retropositions are of frequent occurrence. Integration of retroposition elements is a signal that triggers certain processes in the genome at the integration site. The data on a particular localization of retroposition elements in the genome agree with the suggestion discussed here. The processes of retroposition and pathology are probably interrelated. Trisomy of chromosome containing many retroposition elements can increase the activity of retroposition processes. The copying of retroposition elements into a locus containing the oligonucleotide repeat cluster may trigger the dynamic mutation. The revealed interlinear genomic polymorphism in the organization of the ID family in normotensive and hypertensive rats suggests the existence of additional copies of this element in the genome of spontaneously hypertensive rats.

Key Words: *genome; genome evolution; autoregulation of genome; reverse transcriptase*

Retroposition

1. Retroposition is the copying of genetic information from one locus of the genome to another locus. Genetic information can be transferred from one locus of the genome to another locus. This process is designated as *transposition*. Information can be transposed from the initial locus to the target locus by the DNA transfer. This process may also involve an RNA intermediate. The former process occurs in prokaryotic and eukaryotic cells. The latter process is known to proceed only in eukaryotes. This pathway of transposition is prevalent in eukaryotic organisms. This is probably due to the fact that such a process includes two stages:

1) DNA $\xrightarrow{\text{transcription}}$ RNA (many copies);
2) RNA $\xrightarrow{\text{reverse transcriptase}}$ DNA (many copies).

The number of copies of the information carrier can be increased during each stage of this process.

Russian Cardiological Research-and-Production Center, Ministry of Health of Russia, Moscow

0007-4888/98/0007-0643$20.00

Therefore, the probability of integration into the target locus is increased.

Such a pathway includes the stage of reverse (retro-) transformation of genetic information. The term *retroposition* sets off the RNA-mediated transposition from other transposition processes. The fragment of genetic information transferred has been labeled as *retroposon* (RP) [31].

RPs autotransposed by such a pathway and containing the gene of reverse transcriptase (in the composition of their sequences) were termed *retrotransposons* [5]. Retrovirus (in the form of provirus, proretrovirus) is an example of such a sequence. RPs that are not retrotransposons use foreign reverse transcriptase for retroposition.

In infectious particles the retrotransposon-virus is represented in the form of a RNA copy. This particle can be transferred into another cell. Nonviral transposition is the transfer of genetic information within a cell. This process accounts for the formation of many large families of pseudogenes, as well as for a great diversity of genomic retroposition elements. The majority of their functions are unknown. Several authors assume that nonviral retrotransposons are "molecular parasites". The role of reverse transcription in the cell is limited to self-reproduction of "egoistic" DNA [28, 58] (Scheme 1).

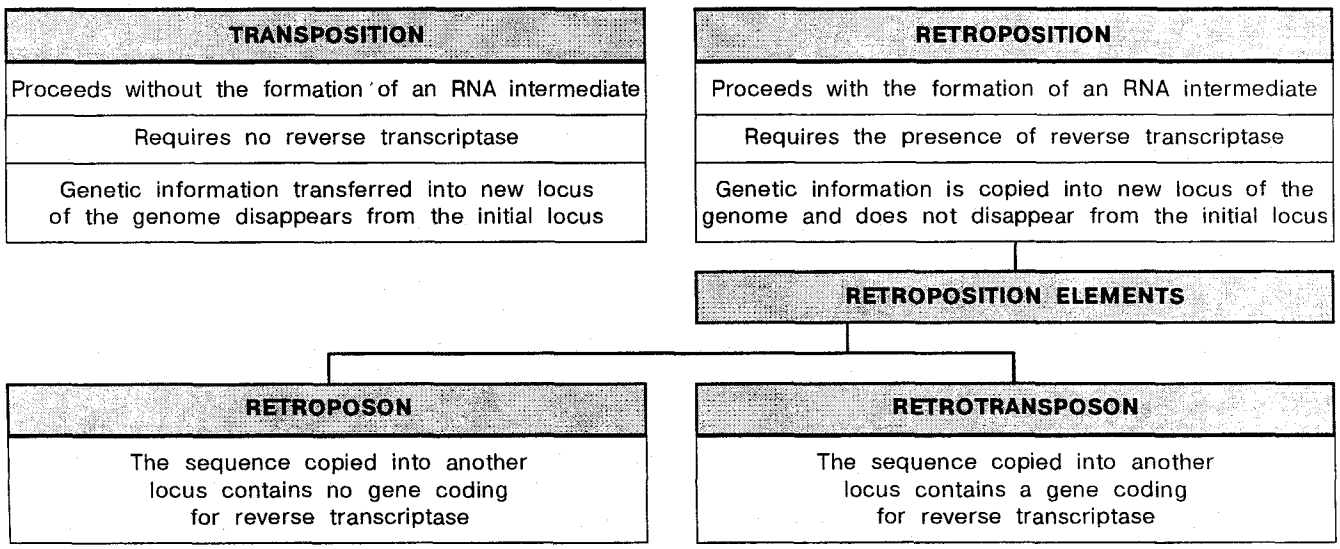
2. Reverse transcriptase is one of the most ancient cellular enzymes. Reverse transcriptase is an RNA-dependent DNA polymerase. This enzyme catalyzes replication of DNA on RNA matrix. Moreover, reverse transcriptase acts as an RNase H (RNA cleavage) and integrase (endonuclease). The latter enzyme mediates the integration of a retro-element DNA copy into the genome. Reverse transcriptase was in existence before the differentiation of live organisms into prokaryotes and eukaryotes and the appearance of retroviruses [68]. Telomerase (reverse transcriptase with the integrated matrix) is probably a more ancient enzyme [48]. Retroviral and retrotransposon reverse transcriptases are probably derived from cellular telomerase lacking the integrated matrix RNA [5].

Retroposition Elements (RE)

1. Retrotransposons. Genomes of all eukaryotes organisms contain many transposition elements. Most of them are retrotransposons, because they contain the sequences homologous to those coding for the retroviral reverse transcriptase. RPs can be classified into two families depending on the presence of long terminating repeats (LTR): LTR-containing and LTR-free (LINE-like) RPs [5] (Scheme 2).

Such retrotransposons as Ty elements of yeasts, Mys, IAP, and VL30 of mice, and human THE-1 belong to LTR-containing elements [74]. LTR-containing elements are similar to proretroviruses. They contain enhancer sequences interacting with host regulatory signals [72]. Expression of several yeast Ty retrotransposons is strongly coordinated with the life cycle of the host cell [19]. Sometimes LTR-containing elements are designated as copia-like retrotransposons. The copia retrotransposon was revealed in Drosophila [65]. In its most limited sense, copia-like retrotransposons constitute a group that belongs to the superfamily of LTR-containing elements.

LTR-free (LINE-like) mobile genetic elements are long moderately repetitive sequences of DNA



Scheme 1.

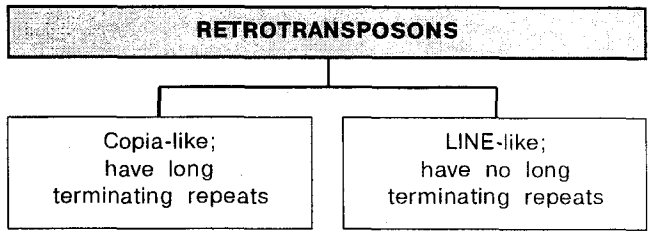
which were originally identified in mammalian cells [35]. These elements are characterized by the absence of LTR, duplication of the host DNA in integration sites, and the 3' flanking poly(A) sequence (which length is varied). The sequence coding for reverse transcriptase in LINE elements contains no domain coding for the RNase activity [54,61]. LINE-1 of humans [20], mice [51], and rats [36] and HeT-A of Drosophila [61] belong to LTR-free elements.

Reverse transcriptase, a product of expression of LINE elements, was obtained in experimental systems. For example, the expression of reverse transcriptase of the human LINE-1 element was observed in yeasts [50].

In addition to these two types of retrotransposons, the genome of vertebrates contains numerous copies of proretroviruses [11,23,45]. Presumably, they cause no hazard to the host because they are degraded and their progeny cannot leave the cell. Some of these retrotransposons produce reverse transcriptase [52]. Retroposition of such sequences was observed. In some instances this retroposition can be induced experimentally [39,62,64].

2. Expression of retrotransposon reverse transcriptases in eukaryotic cells. Retrotransposons can be expressed during certain periods of life in various types of cells. It can be stated that the expression and transposition of full-length functional copies of elements occur in embryonal cells of organisms of various taxonomic groups. Otherwise, it would be impossible to explain the evolutionary genetics (distribution) of these elements [5]. The endogenous ERV-3 proretrovirus represented by a single copy is expressed in the human placenta [42]. The protein LINE product was immunologically identified in embryonal and somatic cells of the mouse seminal vesicles [22]. Endogenous RNA-dependent DNA polymerase was found in the microsomal fraction of embryonic cells of human NTera2D1 carcinoma [26]. The presence of full-length RNA copies of LINE-1 elements in reverse transcriptase-containing fractions allows us to assume that this enzyme is coded by the LINE element. Endogenous reverse transcriptase is expressed by the brain cells of newborn rats. The expression is typical of this stage of ontogenesis [6].

Thus, reverse transcriptases were isolated from the cells of many tissues of higher animals. They are probably endogenous enzymes. Therefore, such tissue conditions are suitable for copying and transposition of retrotransposons and RPs over the genome. The latter elements include numerous families of moderately repetitive sequences of the mammalian genome.



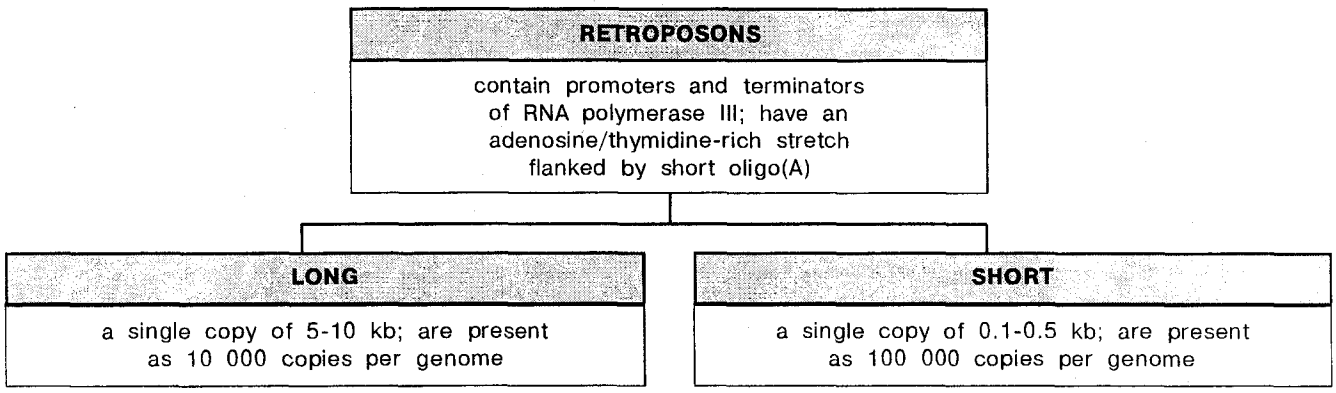
Scheme 2.

Regular expression of retrotransposons (but not the expression induced by several exogenous factors) provides the conditions necessary for the RP transposition. Thus, retroposition can affect evolution, ontogenesis, and pathogenesis.

3. Retroposons. Mammalian genome contains numerous families and subfamilies of repetitive sequences that are transposed with the involvement of reverse transcription. Divergence and conversion between these families make their systematics and classification difficult. It is conventional to distinguish long and short RPs (Scheme 3).

Short RPs are characterized by a 0.1-0.5-kb-long single copy, which is limited by short unidirectional repeats, and display a varied location in the genome. They present in 100 000 copies per genome. Short RPs contain an adenosine/thymidine (AT)-rich stretch, which is identical with the 3' end of cellular tran-scripts and terminates by short oligo(A). In addition, they contain polyadenylation and splicing signals, promoters and terminators of RNA polymerase III, and sequences homologous to recombination breakpoints and low-molecular-weight nuclear RNA involved in splicing. The human Alu element and mouse B1, B2, and R elements belong to this type of RPs [8,44]. Analysis of various RPs revealed that certain copies can be primarily integrated into AT-rich stretches of other copies with the formation of dimers (as in the case of the Alu sequence) or new elements (B2-OBY and Alu-Kpn1), which may be then amplified to individual families of longer RPs [63].

Long RPs are characterized by a 5-10-kb single copy and high heterogeneity of sequences within the family. They are usually flanked by short unidirectional repeats and display a varied localization in the genome. The family of long RPs may be represented by 10 000 copies per genome. The Kpn1 family of primates, MIF1 family of rodents, and other families were shown to belong to this type of RPs. Both long and short RPs contain an AT-rich stretch identical with the 3' end of cellular transcripts and terminated by short oligo(A). They may contain all other signal sequences typical of short RPs. Unsteady localization in the genome is characteristic of the copies of the mouse L1 family [73] and human Kpn1 family



Scheme 3.

[57]. The increased potency for reorganization of sequences is the specific feature of these elements. This is manifested in processes of obtaining DNA inserts and conversion of several sequences that increase the heterogeneity of the family [46].

Kpn1 and Alu are the most extensively studied RP families. The Alu family represents 3% of genomic DNA in humans. This family is represented by 500 000 copies per genome. On the average, each gene contains five Alu copies. These copies undergo amplification and distribution over the genome with the involvement of RNA intermediates and reverse transcriptase which is expressed by retrotransposons present in the genome [63]. The Kpn1 family is present as 20 000 copies per genome (one copy per each five genes). In the mean, one copy may be present in each 150 kb of genomic DNA [53]. Kpn1 are dispersed over the genome. They flank genes and are present in introns and inside of centromeric satellite DNA [47]. The Kpn1 family belongs to the class of long RPs. Their transposition was shown to induce the most pronounced reorganizations (including amplifications and diminutions) of clusters of repetitive sequences [46,60].

Specialization of Certain Families of Moderately Repetitive Sequences

1. Specialized families of repeats. Nonrandom localization in the genome is typical of many families of moderately repetitive sequences. Several RE families are selectively copied into certain compartments and loci of the genome ignoring other loci and compartments. A new polymorphic Alu subfamily, A25, has been recently found. Its copies are present in chromosome 8 [13]. The He-T family of moderately repetitive sequences differs in the localization in the heterochromatin and telomeric and pericentromeric regions. Thus, this family is referred to as a heterochromatin-telomeric family [69,70]. LINE-like retrotransposon G is found only in the pericentromeric chromatin [27]. RP Dm2103 is a telomere-specific mobile element of the drosophila genome. *In situ* hybridization showed that this element is present only in telomeric regions [7]. The family of hoppel mobile elements is present primarily in telomeres [7]. The LINE-like retrotransposon (a "He-TA box") has a very interesting specialization [16]. This 150-bp sequence containing poly(A) at the 3' end transposes to the broken ends of chromosomes the binding to the breakpoint region through poly(A). Thus, the breakpoint region is repaired, and a new telomere is formed. The "He-TA box" retrotransposon is selectively copied into the broken end of the chromosome. Finally, there are specialized mobile elements that can be characterized as gene-specific elements (for example, RI and RII retrotransposons). They are selectively copied into a certain site of genes of ribosomal RNA [37].

2. Probable role of specialized families. The data on compartmentalization ("polarization") of the interphase nucleus of drosophila embryonal cells were obtained in 1984 [71]. Telomeric and centromeric material was shown to form the poles of interphase nuclei. Comparison of these results with the data on the presence of repetitive sequences specific for a centromere [27] and telomere, as well as for both [70], would be worthwhile. The following scheme of polarization of the interphase nucleus is proposed. Homologous sequences tend to be closely positioned. Thus, two compartments are formed in the interphase nucleus: telomeric DNA (due to association of copies of telomere-specific RPs) and centromeric DNA (due to association of copies of centromere-specific repetitive sequences). The elements characterized by a wide specificity (centromere-telomere-specific elements) serve probably for association with the nuclear matrix. Thus, they fix polarization of the interphase nucleus. If so,

these repetitive sequences should contain a great number of signals of association with the matrix.

Retroposition Can Provide the Basis for Autoregulation of Genomic Processes

Tissue- and stage-specific expression of reverse transcriptases by retrotransposons, differentiation and specialization of RP families, and the fact that retroposition is a specific feature of eukaryotic genome (as compared with prokaryotes) suggests that the mechanisms of retroposition form the basis for autoregulation of genomic processes. If so, the following facts should be anticipated: 1) the number and position of RPs in the genome correlate with the number and position of genes (unique sequences); 2) the rate of evolution of RPs is not lower that that of the genome (as a whole); 3) retroposition is not a rare event in the population; and 4) the RE integrated is perceived as a signal, i.e., the element copy integration induces pronounced transforming effects on genomic processes (at the integration site). A question arises: as to whether our suggestions correlate with experimental results. Let us consider this problem.

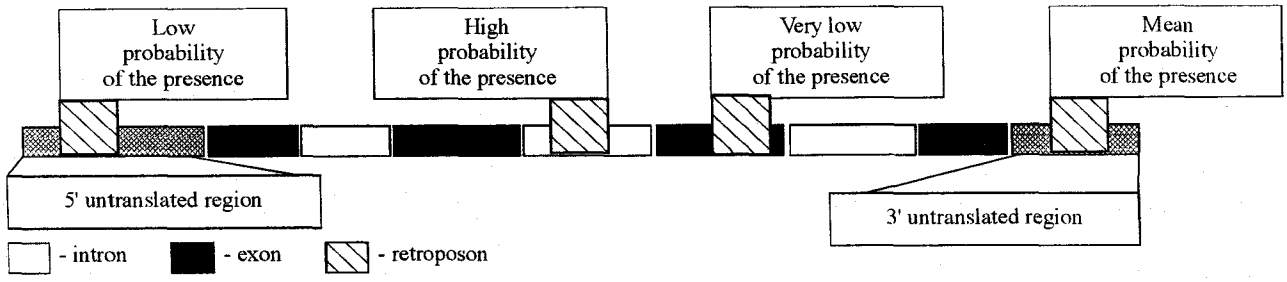
1. The presence of REs in each gene. The human genome is suggested to contain approximately 100 000 genes [67] and probably no less than 1000 families of moderate repeats [8]. Therefore, even if each family contains only 100 copies, each unique sequence of DNA in the genome (each gene in the genome) is located closely to a particular copy of the family of moderately repetitive sequences. Data obtained on cloned human genes agree with this conclusion. There is an increasing body of data on the primary sequence of certain genes and on sequenced genomic regions. Analysis of these data revealed a new specific feature of the structure of eukaryotic genes. This feature eluded scientists because of the absence of statistically significant data. *The specific eukaryotic gene contains (in the composition of its sequence) one or several RPs belonging to one or several families of genomic moderate repeats.* Some of the following facts indicate that the presence of RPs in the gene is probably responsible for a coordinated functioning of this gene in the genome. This structural feature of the specific eukaryotic gene received little attention due to the fact that all genomic REs were not considered to be typical of the genome, i.e., they were believed to be "the egoistic DNA" [28,58].

The presence of RPs in gene introns was shown [14,75]. Now, the RP presence in the gene is well documented [9,32,34]. The AE1 gene of the human anion exchanger is an interesting example. The primary sequence of this gene includes nine RPs [66].

Recent studies indicate that RPs are present not only in introns but also in other elements of the gene (exons [15,56], 3' untranslated region [9], and 5' untranslated region [77]). RPs of the Alu family are present in 5% of mRNA undergoing processing. Approximately 85%, 14%, and 1% of all Alu elements were found in the 3' region, 5' region, and translated region of mRNA, respectively (that is, in gene exons) [77]. Most short signal sequences (boxes) regulating the expression are found exactly in the 5' region. Thus, this region can be regarded as a genome-controlling site. Most RPs contain boxes [8,44] regulating the expression [51]. Therefore, the RP integration into the 5' region can affect gene functions.

Thus, the essential presence of one RP or RPs of one or several families in the specific eukaryotic gene is confirmed. These retroposition sequences belonging to the class of genomic moderate repeats can be positioned in any element of gene structure. Scheme 4 showing the structural organization of the specific eukaryotic gene illustrates these facts.

2. The rate of evolution of RPs is higher than that of the genome. The genome is known to evolve. DNA sequences undergo point mutations, deletions, amplifications, insertions, substitutions, and significant genomic reconstructions. The rate of changes in the nucleotide sequence of any genomic region induced by these factors is the rate of this region evolution. The evolution rates of various types of genomic DNA sequences are different. It was suggested that repetitive genomic sequences differ from



Scheme 4.

unique sequences by having a higher evolution rate [30]. This is confirmed by the data on the RP evolution. The evolution mechanism of the Alu family was studied in primates [41]. The oldest subfamilies are larger. However, young subfamilies contain less copies. The polymerase chain reaction used for evaluating the Sb2 RP (young Alu subfamily) revealed this element in genomes of the green marmoset, orangutan, gorilla, and chimpanzee, as well as in the human genome. However, the sequencing procedure showed that monkeys contain (at this position) an old and highly mutant Alu element typical of primates. Analysis of DNA sequences flanking integrated Alu elements showed that they correspond to the expected sequences typical of phylogenesis of primates. However, the comparison of sequences of these Alu elements revealed that the human RP envolves more rapidly. Presumably, a conversion event occurred within the human Alu element. This event transferred the element from the oldest subfamily into one of the youngest subfamilies. Obviously, there is an intense conversion between the members of subfamilies of interspersed repetitive sequences. The Alu RP belonging to the HS-1 subfamily and carrying a mutation typical of the CS subfamily is of interest. This Alu element is an intermediate form of evolution (between these two subfamilies) [40].

It can be suggested that the evolution rates of various types of genomic sequences are as follows. Unique sequences evolve most slowly. The rate of changes of the primary sequence of moderate repeats during the evolution is higher. The most rapid changes occur in the clusters of highly repetitive sequences. However, high repeats cannot contain any repetitive sequences. Some of them are only the substrates for DNA-binding proteins [10]. Thus, the evolution of high repeats should be a self-closed process (in most cases). However, it is not appropriate for moderate repeats. Families and subfamilies of these repeats carry various sets of signal sequences. The influence of the evolution of moderate repeats on the evolution as a whole was demonstrated by the relationship between the appearance of new strains and the formation of new families of moderate repeats [30]. Eliminating high repeats from consideration, it can be said with confidence that the evolution rate of RPs is higher then that of the genome. Thus, the evolution of moderate genomic repeats can be the determining factor (or one of these factors) in the evolution of the genome as a whole.

3. High frequency of retropositions. The RP evolution (even in the case of conversion between RPs) cannot be considered as the determining factor of genomic evolution if the frequency of retropositions is low. However, evolutional and clinical molecular and biological studies show that transpositions of moderate repeats in the genome and distribution of new RE families over the genome represent a common and quite intense "background" of genomic evolution.

Some of the youngest Alu elements are absent from the orthologous loci in most primate species. This indicates the presence of recent retropositions [41]. RPs of the human AE1 gene belong to four interrelated subfamilies distributed over the genome at various periods of evolution of primates. This example is a series of consecutive retropositions into the gene (in spite of the fact that this gene is one of the most conservative genes) [66]. The recently found members of the young Alu subfamily differ from the typical human Alu element by the presence of deletion at the 5' end of 34 nucleotides. The polymerase chain reaction revealed that this subfamily was located on human chromosome 8. This subfamily polymorphism was observed within the population [13].

Molecular and biological clinical studies indicate that retroposition is an abundant event. Investigations of a restriction polymorphism in 300 bp (which was previously described in humans) showed that this polymorphism results from the Alu element integration. Polymerase chain reaction showed that the percent of retroposition-carrying individuals varied considerably between three various groups [40]. Individual differences in the sites of Alu inserts are probably abundant. Studies of the human CC10 gene support this conclusion [34]. The Alu insert into intron 2 of this gene was revealed in 3% of 168 individuals included in the study.

Thus, retroposition is an abundant event. RPs contain numerous signal sequences. Therefore, the evolution of moderate repeats seems unlikely to be a self-closed process. However, the data on signal functions of integrated RP triggering several molecular and biological processes in the retroposition locus are necessary to consider the evolution of moderate repeats as the driving force for genomic evolution (as a whole).

4. Integration of RE is the signal. There is a large body of evidence that the RP integrated into a certain locus of the genome acts as the signal, i.e., triggers certain genomic processes at the integration site. There are several physical factors of such RP action. In most cases the signal sequence in the RP composition accounts for its effects. Sometimes, reciprocal orientation of RPs can be the signal. Moreover, the degree of the RP sequence degeneration can act as the signal. It is important that the signal (the RP integration) reads depending on the

"context" (genomic environment). This is illustrated by the following facts.

1) The RE contains the signal sequence. Studies of genomic environment of the signal element of the negative response to calcium illustrate the signaling effects of RP. This element (nCARE, type 2) is the DNA regulatory sequence, a palindrome, following several thymidine residues. nCARE inhibits the transcription in response to the increase in extracellular calcium. It was shown [51] that most RPs of the Alu family contain nCARE. Analysis of 7SL RNA (from which Alu elements are derived) revealed the nCARE palindrome positioned closely to the poly(A) tail. Reverse transcription of poly(A) adds several T. Reverse transcript (the Alu element containing the functioning nCARE) can be then integrated into the genome. Thus, some genes become associated with nCARE. The product of expression of these genes is inhibited by the increase in the concentration of extracellular calcium.

2) Reciprocal orientation of REs appears to be the signal. Molecular and biological studies of patients with chronic myeloid leukemia showed that reciprocal orientation of RPs in the integration locus can be the factor determining further events in this genomic region [55]. A complex translocation was revealed in the genome of such patients. The cloning of a recombinant product of this translocation (the BCR gene-containing fragment) showed that recombination occurred at the site between two opposite Alu elements. Both Alu elements are positioned close to the recombination point. Under certain conditions two opposite RPs can serve as the recombination signal (similarly to A- and T-boxes of the SAR/MAR type sequences) [1,12,21,33].

Common Principles of Position and Degeneration of REs

1. Intergenic asymmetry of RE degeneration. The Alu repeat of 820 bp was found in the 3' region of the tissue plasminogen activator (tPA) gene [9]. The sequence of this element shows 86-88% homology to sequences of Alu elements of 3' untranslated regions of the genes for cytochrome P-450, lysozyme, and protein p53, as well as of other Alu 3' flanking regions of human genes. By contrast, this Alu element located in the 3' region of the tPA gene was less than 70% homologous to other Alu elements dispersed in introns of the same gene (at a direct and reverse orientation). This phenomenon can be considered as an intergenic asymmetry of homologies of repetitive sequences. This asymmetry does not seem to be an accident. If analogous structural elements of genes contain REs, which display a higher degree of homology to each other than to the same REs in other structural gene elements, this reflects rather an unknown specific feature of genome function. Another possibility is that these REs perform some similar function in 3' regions of genes.

There is evidence [33] suggesting functional compartmentalization of the interphase nucleus. The sequences performing the same function at a certain period of time (transcribing sequences) appear to be closely located. This results in the formation of a specialized functional compartment of transcription characterized by optimal concentrations of various transcriptional factors and a high probability of the enzyme-substrate interaction. A question arises: which factors contribute to such a location? Association is probably typical of homologous sequences in the interphase nucleus. This is similar to association of large DNA loops from various chromosomes containing clusters of rRNA genes. Such an association results in the formation of a specialized nuclear compartment, the nucleolus [59]. Unique transcribing genes could associate because of the presence of homologous REs. The data suggest that this is not the association of REs of introns or 5' flanking regions, because these processes would inhibit the transcription or make its regulation difficult (respectively). Thus, this is the association of REs integrated into 3' regions of genes.

If the 3' regions of genes are closely located in the interphase nucleus, the selective homology of 3' flanking RPs of genes [9] could be maintained by conversion, because such a process between these RPs is more probable than that between other RPs of this family.

2. Interspecies genomic symmetry of RE location. The data on the interspecies asymmetry in the location of REs are of interest. For example, the 3' noncoding region of the mouse dehydrogenase subunit was shown to contain B1 element [24]. It was stressed that the same region of the same human gene contains three Alu copies. These two families of moderate repeats in various mammalian strains display homology to each other. Therefore, an identical gene contains the identical (homologous) repeat in the identical region. It can be suggested that symmetrical localization of REs is the signal programming the same process in two genomes. This signal probably provides a simultaneous synthesis of all subunits of dehydrogenase. Then, it would be possible to reveal the analogous element in other genes of subunits. Verification of the effect of intergenomic symmetry of REs will provide new approaches to the problem of molecular and biological principles of the "canalized evolution" and Vavilov's "homologous series of variability".

Retroposition and Pathology

1. Shift of the equilibrium between integrated and extrachromosomal genomic REs can stimulate retroposition. A dynamic equilibrium between extrachromosomal cyclic copies of REs and retroelements integrated into the genome was believed to be present in the cell nucleus [3]. This equilibrium state may be disrupted in the case of a sharp increase or decrease in the total number of integrated copies. The system then comes into equilibrium. This is manifested in the redistribution of REs integrated over the genome. In this respect, the results of molecular and biological studies of partial duplication of the HRX gene in the trisomy 11 genome resulting in acute leukemia are of interest [15]. Two variants of this partial duplication were investigated. In the first case, the Alu RP was integrated into exon 6 and intron 1 of this gene. In the second case, the Alu RP was integrated into exon 6 only. Therefore, integration of the mobile element into exon was observed in both cases. Retroposition into the coding region of the gene is a rare event. This study revealed its two variants. The probability of retroposition is sharply increased in trisomy (at all) or in trisomy of certain chromosomes. We assume that the retroposition processes are activated as a result of restored equilibrium between circular extrachromosomal and integrated intrachromosomal copies of certain RP family in the genome. A question arises: whether the increase in the total chromosome material (an additional chromosome) shifts the equilibrium? This is possible if: 1) the extra chromosome is the site of preferred localization of the RE family and 2) this chromosome contains a long RE-rich region. It was mentioned above that there are chromosome-specific REs [13]. Moreover, the region of chromosome 21 responsible for Down's syndrome was recently shown to contain a great number of moderately repetitive sequences [76].

2. Retroposition can initiate dynamic mutation. Dynamic mutation is a prolonged process resulting from a progressive increase (expansion) in the trinucleotide repeat cluster. The number of repeat copies in the cluster (n) increases from one generation to another. Once n reaches the critical value, the expression of a nearby gene is disrupted. However, the gene structure remains unchanged. Dynamic mutation was shown to be the genomic determinant of at least eight familial diseases. We assume that the RP integration into the cluster of a highly repetitive sequence may be the signal initiating dynamic mutation in this cluster [3]. This suggestion was indirectly confirmed by recent findings. A comparison of the 5' region sequences in human and bovine genes of von Willebrand factor (vWF), which regulates the transcription in blood cells (megakaryocytes), has shown that this region is extremely conservative [38]. However, the following difference was revealed. The human vWF gene contains a variable cluster of the poly(GT) dinucleotide repeat of 18-26 bp. At this site the bovine vWF gene contains a 523-bp insert. This insert is the Alu-type RP of 331 bp flanked by a simple 192-bp repeat. Microsatellite clusters of 20 bp and 200 bp flanking RPs were detected in these regions of human and bovine genes, respectively. The following scheme accounting for this difference has been proposed [3]. The RP is integrated into the trinucleotide repeat cluster. Such a genomic reconstruction triggers an order of magnitude increase in this cluster. Therefore, the bovine vWF gene RP is the factor triggering dynamic mutation or an analogous process. The RP integration and generation of dynamic mutation or analogous process were observed in the bovine genome. However, this study provides no answer to the following questions: 1) whether the increased microsatellite cluster is stabilized; and 2) whether this process is associated with any disease. Studies of the expression and structure of human CC10 gene probably responsible for the age-related predisposition to asthma revealed the RP located close to the variable satellite cluster [34]. The CC10 gene contains three short exons separated by the first long and the second short introns. Four RPs were found in the introns of this gene. Three Alu elements were positioned in intron 1; one Alu element was positioned in intron 2. One polymorphic region was identified within each intron. In the first case, it was a microsatellite with a varying number of copies of tetra- and pentanucleotide sequences. This microsatellite was the 5' flanking region of the third Alu element in the first intron. In the second case, 3% of 169 studied chromosomes contained the Alu insert into intron 2 (at a distance of 45 bp from the place where intron 2 was located to the 3' exon). So, the presence of a recent transposition was obvious. A high-level expression of the CC10 gene in the epithelium of the respiratory tract has been revealed. Considering the anti-inflammatory and immunomodulating properties of the gene product, impaired expression of this gene may be responsible for a persistent inflammation of respiratory pathways.

A hypervariable microsatellite was well characterized in the gene coding for angiotensin. Molecular variants of this gene are associated with arterial hypertension. This question has been widely discussed in scientific literature. This microsatellite is the cluster of the GT dinucleotide repeat [43]. Sequencing of the fragments of angiotensin gene was

not performed in this work. It is unknown whether the variable microsatellite (GT)*n* is flanked by any RP.

We assume that retroposition is the factor triggering essential hypertonia. Spontaneously hypertensive rats (SHR) have been widely used for studying essential hypertonia. If our suggestion is correct, genomic polymorphism of the RE organization occurs in Wistar rats (the ancestor of SHR and normotensive control WKY rats), i.e., the interlinear polymorphism of one of the RE families in WKY rats and SHR, would be expected. Our experiments showed that such polymorphism of the ID family can be present. It is likely that the genome of SHR contains additional copies of the ID repeat (which are absent in the genome of WKY rats). This ID retroposition can be used as a molecular marker of SHR. It is unknown whether there is any relationship between elevated arterial pressure in SHR and this retroposition. The correlation between high arterial pressure and genomic polymorphism will be the subject of further studies.

REFERENCES

1. A. P. Akif'ev, G. A. Khudolii, I. Ya. Belyaev, A. I. Potapenko, *Usp. sovr. gen.*, **18**, 163-194 (1993).
2. D. V. Bebikhov, *Genetika*, **29**, No. 3, 373-387 (1993).
3. D. V. Bebikhov, T. A. Nikonenko, A. Yu. Postnov, Yu. V. Postnov, *Kardiologiya*, **37**, No. 4, 80-86 (1997).
4. O. N. Danilevskaya, D. V. Bebikhov, E. V. Kurenova, B. A. Leibovich, *Izv. Sibirskogo otd. AN SSSR. Ser. Biol. nauki*, Vyp. 2, 7-7 (1989).
5. V. A. Ivanov and Yu. V. Il'in, *Molekul. biol.*, **29**, No. 2, 258-280 (1995).
6. V. A. Ivanov, P. I. Pakhotin, N. V. Bobkova, *Dokl. RAN*, **323**, 173-177 (1992).
7. E. V. Kurenova, B. A. Leibovich, I. A. Bass, *et al.*, *Genetika*, **26**, No. 10, 1701-1712 (1990).
8. A. P. Ryskov, *Genomic Organization* [in Russian], Moscow (1989).
9. A. G. Sarafanov, M. Ya. Timofeeva, V. M. Bannikov, *et al.*, *Molekul. biol.*, **29**, No. 2, 287-293 (1995).
10. P. A. Slominskii, M. I. Shadrina, S. A. Limborskaya, *Ibid.*, **31**, No. 1, 45-48 (1997).
11. N. V. Khaidarova, Kh. P. Nesmeyanova, V. Z. Tarantul, *Ibid.*, **31**, No. 1, 38-44 (1997).
12. A. P. Akifyev, *Physiol. Gen. Biol. Rev.*, **10**, 51-56 (1995).
13. S. S. Arcot, J. J. Fontius, P. L. Deininger, and M. A. Batzer, *Biochim. Biophys. Acta.*, **1263**, No. 1, 99-102 (1995).
14. *A.* Barta, R. I. Richards, J. D. Baxter, and J. Shine, *Proc. Natl. Acad. Sci. USA*, **78**, 4867-4871 (1981).
15. O. A. Bernard, S. P. Romana, S. A. Schichman *et al.*, *Leukemia*, **9**, 1487-1490 (1995).
16. H. Biessmann and S. Carter, J. Mason, *Proc. Natl. Acad. Sci. USA*, **87**, 1758-1765 (1990).
17. H. Biessmann and J. M. Mason, *EMBO J.*, **7**, 1081-1086 (1988).
18. H. Biessmann, J. M. Mason, K. Ferry *et al.*, *Cell*, **61**, 663-671 (1990).
19. V. W. Bilanchone, J. A. Claypool, P. T. Kinsey, and S. B. Sandmeyer, *Genetics*, **134**, 685-670 (1993).
20. J. D. Boeke and V. Corces, *Annu. Rev. Microbiol.*, **43**, 403-434 (1989).
21. T. J. Boulicast, *J. Cell. Biochem.*, **50**, 111-123 (1992).
22. D. Branciforte and S. L. Martin, *Mol. Cell. Biol.*, **14**, 2584-2592 (1994).
23. J. M. Coffin, in: *RNA Tumor Viruses*, J. Coffin (Ed.), New York (1982), **1**, pp. 1109-1203.
24. P. A. Costeas and L. A. Tonelli, J. M. Chinsky, *Biochim. Biophys. Acta*, **1305**, 25-28 (1996).
25. O. N. Danilevskaya, E. V. Kurenova, and M. N. Pavlova *et al. Chromosoma*, **100**, 118-124 (1991).
26. J.-M. Deragon, D. Sinnett, and D. Labuda, *EMBO J.*, **9**, 3363-3368 (1990).
27. P. DiNocera, F. Graziani, and G. Lavorgna, *Nucleic Acids Res.*, **14**, 675-681 (1986).
28. W. E. Doolitle and C. Sapienza, *Nature*, **284** (1980).
29. J. G. Douglas, M. Thibonnier, and J. T. Wright, *J. Assoc. Acad. Minor. Phys.*, **7**, No. 1, 16-21 (1996).
30. G. Dover and E. Coen, *Nature*, **290**, 731-739 (1981).
31. A. M. Einer, P. L. Deininger, and A. Efstratiadis, *Annu. Rev. Biochem.*, **55**, 631-661 (1986).
32. I. L. Flink and E. Morkin, *Gene*, **163**, No. 2, 279-282 (1995).
33. S. Gasser and U. Laemmli, *Trends Gen*, **3**, 16-22 (1987).
34. J. G. Hay, C. Danel, C. S. Chu, and R. G. Crystal, *Am. J. Physiol.*, **268**, No. 4, Pt. 1, 565-575 (1995).
35. C. A. Hutchinson, S. C. Hardies, D. D. Loeb *et al.*, in: *Mobile DNA*, D. E. Berg and M. M. Howe (Eds.), Washington (1975), pp. 593-617.
36. H. Ilves and M. Speek, *Mol. Cell. Biol.*, **12**, 4242-4248 (1992).
37. J. Jakubczak, Y. Xiong, and T. Eickbush, *J. Mol. Biol.*, **212**, 37-40 (1990).
38. N. Janel, J. L. Schwachtgen, M. R. Bakhshi *et al., Gene*, **167**, 291-295 (1995).
39. N. A. Jenkis and N. G. Copeland, *Cell*, **43**, 811-819 (1985).
40. D. H. Kass, C. Aleman, M. A. Batzer, and P. L. Deininger, *Genetica*, **94**, 1-8 (1994).
41. D. H. Kass, M. A. Batzer, and P. L. Deininger, *Mol. Cell. Biol.*, **15**, 19-25 (1995).
42. N. Kato, S. Pfeiffer-Ohlsson, M. Kato *et al., J. Virol.*, **61**, 2182-2191 (1987).
43. Y. V. Kotelevtsev, R. P. Lifton, S. Christopher *et al.* Cell, **71**, 169-180 (1992).
44. A. S. Krayev, T. V. Markusheva, D. A. Kramerov *et al., Nucleic Acids Res.* **10** 7461-7475 (1982).
45. E. Larsson, N. Kato, and M. Cohen, *Curr. Top. Microbiol. Immunol.*, **148**, 115-132 (1989).
46. M. I. Lerman, R. E. Thayer, and M. F. Singer, *Proc. Natl. Acad. Sci. USA*, **80**, 3966-3966 (1983).
47. Q. R. Liu and P. K. Chan, *DNA Cell Biol.*, **12**, 149-156 (1993).
48. V. Lundbland and E.H. Blackburn, *Cell*, **60**, 529-530, (1990).
49. S. L. Martin, *Mol. Cell. Biol.*, **11**, 4804-4807 (1991).
50. S. L. Mathias, A. F. Scott, H. H. Kazazian *et al., Science*, **254**, 1808-1810 (1991).
51. G. S. McHaffie and S. H. Ralston, *Bone*, **17**, 11-14 (1995).
52. J. A. Mietz, Z. Grossman, K. K. Lueders, and E. L. Kuff, *J. Virol.*, **61**, 3020-3029 (1987).
53. T. Miyake, K. Migita, and Y. Sakaki, *Nucleic Acids Res.*, **11**, 6837-6846 (1983).
54. L. J. Mizrokhi and A. M. Mazo, *Proc. Natl. Acad. Sci. USA*, **87**, 9216-9220 (1990).
55. C. Morris, A. Jeffs, T. Smith *et al.*, Oncogene., **12**, 677-685 (1996).
56. J. E. Mullersman and L. M. Pfeffer, *J. Interferon Cytokine Res.*, **15**, 815-817 (1995).
57. H. Nomiyama, T. Tsuzuki, S. Wakasugi *et al., Nucleic Acids Res.*, (1984), **12**, 5225-5234.
58. L. E. Orgel and F. H. C. Crick, *Nature*, **284**, 604-607 (1980).
59. R. P. Perry, *Annu. Rev. Biochem.*, **45**, 605-629 (1976).

60. S. S. Potter, *Proc. Natl. Acad. Sci. USA,* **81**, 1012-1016 (1984).
61. A. F. Priimagi, L. J. Mizrokhi, and Y. V. Ilyin, *Gene.*, **70**, 253-262 (1988).
62. W. Quint, H. van der Putten, F. Jananssen, and A. Berns, *J. Virol.*, **41**, 901-908 (1982).
63. J. H. Rogers, *Int. Rev. Cytol.*, Suppl. 17, S322 (1984).
64. W. P. Rowe and C. A. Kozak, *Proc. Natl. Acad. Sci. USA*, **77**, 4871-4874 (1980).
65. G. M. Rubin, in: *Mobile Genetic Elements*, J. Shapiro (Ed.), New York (1983), pp. 329-334.
66. K. E. Sahr, W. M. Taylor, B. P. Daniels, and P. Jarolim, *Genomics*, **24**, 491-501 (1994).
67. T. B. Shows, A. V. Sakaguchi, and S. L. Naylor, *Adv. Hum. Genet* Naylor, **12**, 341-452 (1982).
68. H. M. Temin, *Nature*, **339**, 254-255 (1989).
69. K. Traverse and M. Pardue, *Chromosoma*, **97**, 261-271 (1989).
70. K. Valgeirsdottir, K. Traverse, and M. Pardue, *Proc. Natl. Acad. Sci. USA*, **87**, 7998-8002 (1990).
71. L. Van der Ploeg, A. Lin, and P. Borst, *Cell*, **36**, 459-468 (1984).
72. H. Varmus and P. Brown, in: *Mobile DNA*, D. E. Berg and M. M. Howe(Eds.), Washington (1989), pp. 55-108.
73. C. Voliva, S. L. Martin, C. A. Hutchison, and M. H. Edgell, *J. Mol. Biol.*, **178**, 795-813 (1984).
74. Y. Xiong and T. H. Eickbush, *EMBO J.*, **9**, 3353-3362 (1990).
75. P. Young, R.W. Scott, D.H. Hamer et al. *Nucleic Acids Res.*, **10**, 3099-3116 (1982).
76. J. Yu, S. Tong, Y. Shen, and F.-T. Kao, *Proc. Natl. Acad. Sci. USA*, **94**, 6862-6867 (1997).
77. I.G. Yulug, A. Yulug, and E.M. Fisher, *Genomics*, **27**, 544-548 (1995).