Volume 214, number 1, 87–91 FEB 04558 April 1987

Pecularities of immunoglobulin gene structures as a basis for somatic mutation emergence

N.A. Kolchanov, V.V. Solovyov and I.B. Rogozin

Institute of Cytology and Genetics, USSR Academy of Sciences, Siberian Department, Novosibirsk 630090, USSR

Received 16 January 1987

Using the method of contextual analysis, we have studied the collection of somatic mutations in immunoglobulin genes. It has been found that the emergence of somatic mutations can be based on the reparation of complementarity violations in the heteroduplexes corresponding to complementary palindromes or direct repeats of DNA.

Somatic mutation; Immunoglobulin gene; DNA repeat; Mismatch repair

1. INTRODUCTION

Somatic hypermutagenesis of immunoglobulin genes is one of the key processes providing variability of amino acid sequencies within the V-regions which is necessary for the antigen-binding specificity of antibodies. Somatic mutations are now well established for V-lambda, V-kappa and V-heavy genes [1]. They are restricted to the rearranged V-genes and their flanking regions and do not occur in C-genes [2]. However the mechanisms responsible for somatic hypermutagenesis are still unknown [1].

Since somatic mutation is an inherent feature of V-genes one can suggest that it is the consequence of certain features of the V-gene's primary structure organization missing in the genes coding for other proteins.

In order to reveal the mechanisms producing somatic mutations in variable genes we have studied by the method of contextual analysis [3] the primary structure organization in two gene groups. The former included 5 V-genes, and the latter 3 C-genes, 2 hemoglobin genes and c-myc gene [4–8]. V-genes appeared to have some pronounced features distinguishing them from the rest genes, that can provide the molecular basis for the emergence of somatic mutations.

Correspondence address: N.A. Kolchanov, Institute of Cytology and Genetics, USSR Academy of Sciences, Siberian Department, Novosibirsk 630090, USSR

2. METHODS

The method of contextual analysis [3] reveals groups of direct, tandem, inverted repeats or complementary palindromes of length l with k violations of homology (complementarity) (fig.1) in a real sequence of length N and compares their number $n(l,k)$ with the expected value $\bar{n}(l,k)$ for a random sequence of length N with the same frequencies of nucleotides q_A, q_T, q_G, q_C. To estimate the statistical significance of differences between the actual and the expected number of repeats the 95% confidence limit $n_0(l,k)$ was calculated. Then, if $n(l,k) \geqslant n_0(l,k)$ the difference was significant and the corresponding repeats or palindromes were designated as nonrandom.

The list of sequences studied is presented in table 1.

3. RESULTS AND DISCUSSION

The results of the contextual analysis of two groups of genes are presented in tables 1 and 2 with

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

Table 1

Characteristics of nonrandom complementary palindromes

Gene	Length	Total number of nonrandom complementary palindromes	Concentration of complementary palindromes for the group of genes
V1 – mouse V_λ	330	36	
V2 – mouse V_λ	330	35	
V3 – mouse $V_\varkappa$	285	32	0.088
V4 – human $V_\varkappa$	285	10	
V5 – mouse $V_\varkappa$	300	23	
C1 – mouse C_γ	990	0	
C2 – mouse C_α	1032	27	
C3 – mouse $C_\varkappa$	321	7	0.018
Hb1 – human α-globin	426	31	
Hb2 – mouse α-globin	489	9	
c-myc human	1000	3	

Concentration of complementary palindromes is calculated by the formula $C = (\Sigma n_i/\Sigma l_i)$ where n_i is the total number of nonrandom palindromes in the i-th gene of the group and l_i is the length

length of repeats or complementary palindromes varying from 7 to 45 bp and the number of homology or complementarity violations from 0 to 23 bp.

The V-genes appeared to have pronounced features distinguishing them from the rest of the genes.

The first striking peculiarity is their high saturation by complementary palindromes. The concentration of nonrandom complementary palindromes for the group of V-genes was 5-times as high as for the second group (table 1).

The second pronounced feature distinguishing V-genes from the others is the presence of nonrandom tandem repeats or direct repeats (or both of these simultaneously) placed in noncoinciding translation frames (table 2).

A large content of nonrandom direct repeats in coinciding translation frames is a characteristic feature of genes coding globular proteins. The presence of such repeats (e.g. see fig.1) is stipulated by the fact that globular proteins coded by these genes are saturated by regular secondary structures (α-helices, β-structures). These structures set limits on the composition and arrangement of amino acids in a primary structure of proteins. Taking into account the non-uniformity of codon usage in genes, these limits produce a high level of nonperfect direct or tandem repeats in coinciding translation frames [3].

It is this disposition of direct and tandem repeats that was revealed in the group including C-, Hb-genes and in c-myc gene (table 2). For V-genes another pattern of repeat distribution in translation frames was observed. Localization of nonrandom direct repeats in coinciding translation frames was a feature of two genes only (i.e. V1 and V2), while in the three other genes (V3, V4, V5) direct repeats were mainly observed in noncoinciding frames (table 2).

As for nonrandom tandem repeats, they occurred in V-genes either in noncoinciding frames (V1, V2, V4) or were uniformly distributed in every possible frame (V3 and V5).

The third striking peculiarity of V-genes revealed by the contextual analysis is a high concentration of nonrandom tandem repeats in noncoinciding translation frames (table 2). The concentration of such repeats for the group of V-genes was 3-times as high as for the second group.

We speculated that the revealed features of V-gene's primary structure organization ensure the

Table 2

Characteristics of nonrandom direct and tandem repeats

Gene	Direct repeats		Tandem repeats			
	Total number of nonrandom direct repeats	Share of nonrandom direct repeats in coinciding translation frames	Total number of nonrandom tandem repeats	Concentration of nonrandom tandem repeats for the group of genes	Share of nonrandom tandem repeats in coinciding translation frames	Concentration of nonrandom tandem repeats in noncoinciding translation frames for the group of genes
V1	92	0.96 (+)	9		0.00 (−)	
V2	67	0.93 (+)	9		0.00 (−)	
V3	11	0.09 (−)	15	0.037	0.60 (.)	0.030
V4	101	0.20 (−)	12		0.00 (−)	
V5	12	0.00 (−)	13		0.25 (.)	
C1	112	0.50 (+)	32		0.56 (+)	
C2	118	0.82 (+)	24		0.75 (+)	
C3	79	0.73 (+)	29	0.033	0.59 (+)	0.010
Hb1	93	0.93 (+)	10		0.80 (+)	
Hb2	40	0.63 (+)	16		0.81 (+)	
c-myc	19	0.79 (+)	32		0.75 (+)	

Statistical characteristics of arrangement of nonrandom repeats in possible translation frames in genes were made on the basis of the binomial distribution criterion under confidential level $p < 0.01$. Gene is marked by (+) if repeats were observed ($p < 0.01$) mainly in coinciding translation frames. They are denoted by (−) if repeats were observed ($p < 0.001$) mainly in different frames and genes are marked by (.) if repeats were uniformly distributed in all translation frames possible

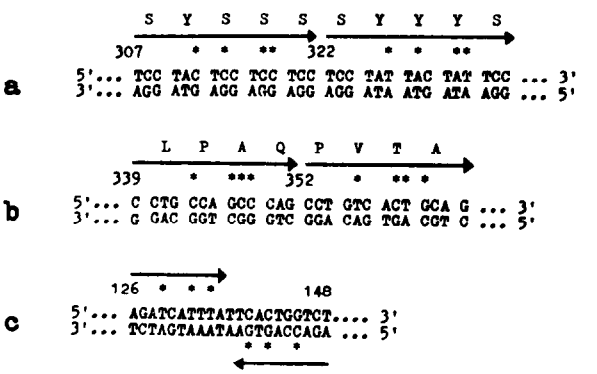

Fig.1. Tandem repeats and complementary palindrome in genes. Tandem repeats made up by pairs of fragments placed in coinciding translation frames (a) and in noncoinciding translation frames (b), respectively. The repeated fragments are indicated by arrows and violation of homology by asterisks. The amino acid fragments encoded by these repeated fragments are also shown. (c) Complementary palindrome.

molecular basis for the emergence of somatic mutations [9].

The first mechanism for emergence of somatic mutations is presented in fig.2. It includes the formation of a nonperfect DNA duplex as a result of mispairing of opposite complementary strands of direct repeats and the subsequent reparation of complementarity violations. After restoration of a normal double helix a recurrent correction of complementarity violations either restores the normal DNA primary structure or produces mutations.

The second mechanism of somatic mutation emergence is based on the reparation of complementarity violations in nonperfect hairpins corresponding to complementary palindromes (fig.3a).

The former mechanism of somatic mutagenesis is similar to some extent to the process of intergenic conversion [10] but the radical difference

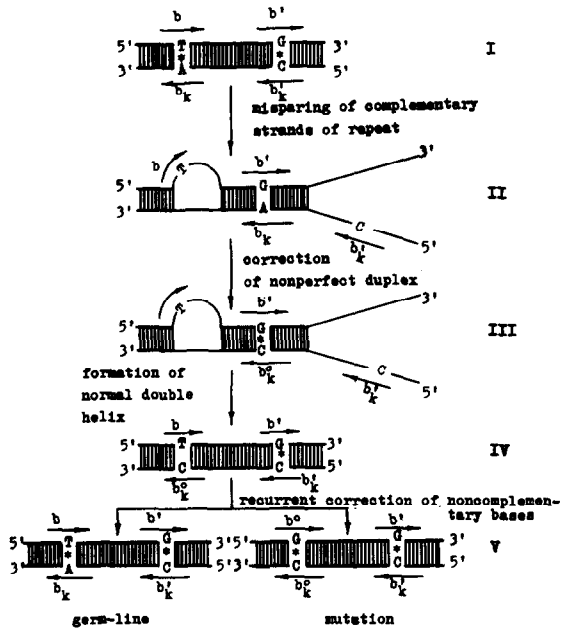

Fig.2. Emergence of the mutation T ⟶ G on the basis of misalignment of direct repeats. (I) Initial DNA structure containing two imperfect fragments of repeat (b and b_k are two complementary strands of repeat); (II) mispairing of complementary strands of repeat; (III) correction of imperfect duplex; (IV) formation of normal double helix; (V) recurrent correction of noncomplementary bases.

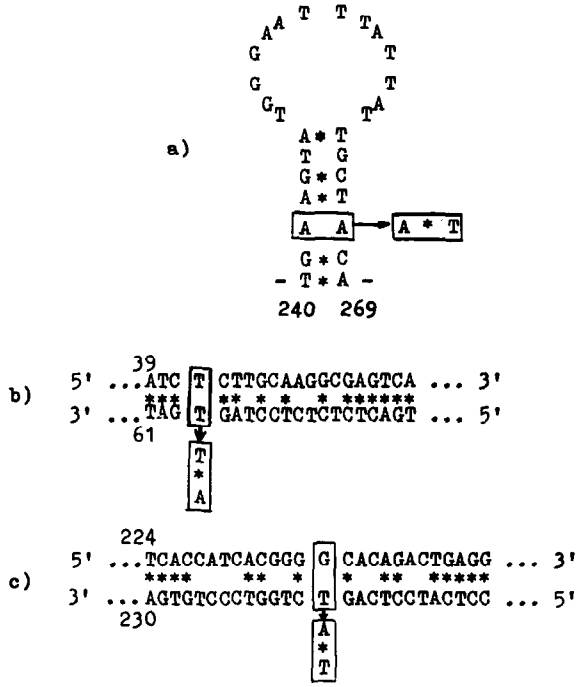

Fig.3. Emergence of mutation in immunoglobulin genes. (a) The mutation A ⟶ T in the 267th position of V3 gene has resulted from the correction of imperfect hairpin; (b,c) the mutations A ⟶ T and G ⟶ A in the 64th and 238th positions of V3 and V1 genes, respectively, resulted from the correction of imperfect duplexes formed on the basis of direct or tandem repeats.

is that two repeated fragments undergoing misalignment and repair correction belong to the same gene rather than to two different genes.

As to the latter mechanism, it was Ripley et al. [11] who for the first time interpreted the emergence of some mutations in *E. coli* and yeast genes on the basis of the assumption that nucleotide substitutions can arise as a result of mismatch repair of nonperfect DNA hairpins.

The analysis of a collection of 20 somatic mutations that have been revealed through comparison of the germ-line genes with the rearranged V1, V2, V3 and V5 genes [4,5,7] has proved the correctness of these mechanisms. It has been ascertained that for each of 20 mutations in the V1, V2, V3 and V5 genes there is at least one nonperfect duplex corresponding to a complementary palindrome or a direct repeat (tandem repeat) which can provide the emergence of a mutation on the basis of repair correction of the duplex.

Fig.3 shows examples of such mutations. The first mutation (A ⟶ T in the 267th position of V3 gene) has resulted from the correction of a nonperfect hairpin (fig.3a) and the other two (A ⟶ T and G ⟶ A in the 64th and 238th positions of V3 and V1 genes, respectively) from the correction of nonperfect duplexes formed on the basis of direct or tandem repeats (fig.3b,c).

The formation of nonperfect DNA duplexes in V-genes on the basis of misalignment of DNA segments corresponding to complementary palindromes or tandem and direct repeats has resulted, obviously, for two reasons.

The first is the high saturation of V-genes by complementary palindromes and tandem (direct) repeats which increases the probability of misalignment.

The second reason, in our opinion, is connected

with the disturbance of a normal chromatin structure as a result of rearrangements of V-genes in the course of B-lymphocyte differentiation, since somatic mutations occur only in the rearranged V-genes joined to J-segments [5]. It is known that release of DNA from nucleosomes or weakening of its interaction with histones results in torsional strain in the DNA [12], which can produce its local melting with further misalignment of complementary DNA strands and formation of nonperfect DNA duplexes providing the structural basis for somatic mutations.

A supplementary analysis was carried out for estimating the stability of the revealed distinctions among the V-gene group and other genes in relation to the sample size. The group of 13 V-genes ($3V_\lambda$, $5V_\varkappa$, $5V_H$) and 13 other genes (10 C-genes, 2 hemoglobin genes and c-myc gene) were taken into consideration. In this case the same differences in the characteristics of repeats were discovered as previously:

(i) the mean concentration of palindromes for the V-gene group (0.025) was 6-times higher than for the second group (0.004);

(ii) the mean concentration of tandem repeats beyond the framework of translation for the V-gene group (0.020) was 6.5-times higher than for the second group (0.003).

It is necessary to note that the detailed study of the collection of somatic mutations in 16 V-genes made it possible for us to show a certain connection between mutations and repeats ($P < 10^{-11}$) as well as a nonrandom connection between repeats and CDRs of V-genes. A more detailed account of these results will be given in our next contribution.

Since the proposed mechanism of somatic mutation emergence is based on the formation of heteroduplexes (and cruciform structures in DNA corresponding to complementary palindromes as well), one can experimentally verify their presence with DNases sensitive to single-stranded DNA fragments [13].

REFERENCES

[1] Tonegawa, S. (1983) Nature 302, 575–581.
[2] Gearhart, P.J. and Bogenhagen, D.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3439–3443.
[3] Solovyov, V.V., Zharkikh, A.A. and Kolchanov, N.A. (1985) Mol. Biol. 19, 524–535 (in Russian).
[4] Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1982) Nature 298, 380–382.
[5] Gorsci, J., Rollini, P. and Mach, B. (1983) Science 220, 1179–1181.
[6] Bank of Nucleotide Sequences of the Institute of Cytology and Genetics of the USSR Academy of Sciences.
[7] Gershenfeld, H.K., Tsukamoto, A., Weissman, I.L. and Joho, R. (1981) Proc. Natl. Acad. Sci. USA 78, 7674–7678.
[8] Rabbitts, T.H., Hamlyn, P.H. and Baer, R. (1983) Nature 306, 760–768.
[9] Kolchanov, N.A., Solovyov, V.V. and Rogozin, I.B. (1985) Dokl. USSR Acad. Sci. 281, 994–999 (in Russian).
[10] Baltimore, D. (1981) Cell 24, 592–594.
[11] Ripley, L.S. and Glickman, B.W. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 851–861.
[12] Lilley, D.M.J. (1983) Nature 305, 276–277.
[13] Lilley, D.M.J. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 101–113.