Biophysics, Vol. 44, No. 4, 1999, pp. 636–643. Translated from Biofizika, Vol. 44, No. 4, 1999, pp. 655–663.

MOLECULAR BIOPHYSICS

Contribution of Signals and Antisignals to the Mutation Spectrum of the *td* Intron Insertion Site

M. P. Ponomarenko, Yu. V. Ponomarenko, and N. A. Kolchanov

Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

Received September 15, 1997

Abstract—Insertion sites of the *td* intron were studied. Tetranucleotides $W_{-30}A_{-29}W_{-28}$ Y_{-27}, $V_{-19}B_{-18}W_{-17}A_{-16}$ and $A_{11}W_{12}A_{13}W_{14}$ (signals) are more often found at normal sites, whereas tetranucleotides $Y_{-27}W_{-26}M_{-25}G_{-24}$, $V_7B_8W_9A_{10}$ and $Y_{15}W_{16}M_{17}G_{18}$ (antisignals) occur more frequently at defective sites. The antisignals destroy the site by inhibiting the signals. At a large number of random substitutions, the site becomes defective because of destruction of signals; at a low number of substitutions, because of the appearance of antisignals.

Key words: mutation spectrum, signals, antisignals, *td* intron insertion sites

INTRODUCTION

The basic genetic processes of replication, transcription, splicing and translation are regulated by functional sites in DNA and RNA. Mutations in these sites change the level of their activity by several orders of magnitude [1, 2]. Contextual features determining the site activity are important for site recognition [3, 4]. Conventionally it is assumed that site activity depends on its closeness to the consensus [5–8]. In [9] we have demonstrated that site activity can depend on local concentration of site-specific oligonucleotides.

Endonuclease I-TevI cuts the intronless allele of the thymidine kinase gene of phage T4. Repair of this region with the intron-containing allele as template inserts the *td* intron 23 bp upstream of the cut [10]. The region of the intronless allele from –30 through +18 is called "insertion site of the *td* intron" [11].

The authors of [10] synthesized the wild-type variant of this site and generated 83 mutant sites using random mutagenesis. For each of these 83 sites, the effciency of restriction by I-TevI was measured. The activity of 49 mutated sites was not lower than that of the wild-type site (normal sites); the activity of 34 sites was lower than the activity of the wild-type site (defective site). Analysis of these experimental data in [10] did not distinguish the particular nucleotides determining the site activity. This was caused by the high similarity between the mutated sites and the wild-type site (1 to 18 substitutions per site; 6 substitutions on the average) and their heterogeneity in the similarity level and positional variability (0 through 26 substitutions per position; 10 substitution on the average).

In this study we applied the computer system SITEVIDEO [12] and the binomial criterion [13] to analysis of the experimental data from [10]. This resulted in determination of the contextual features of normal and defective sites (signals and antisignals, respectively). If the number of substitutions is small, the wild type becomes defective because of appearance of antisignals; at the large number of substitutions, because of destruction of signals.

METHOD

The following notation will be used: $Z = \{z_j\}_{1 \le j \le 4}$ is a tetranucleotide in alphabet $z_j \in$ {A, T, G, C, W = A/T, R = A/G, M = A/C, K = T/G, Y = T/C, S = G/C, B = T/G/C, V = A/G/C, H = A/T/C, D = A/T/G, N = A/T/G/C}; $Z_i = \{s_{i+j-1} = z_j\}_{1 \le j \le 4}$ means that Z occupies

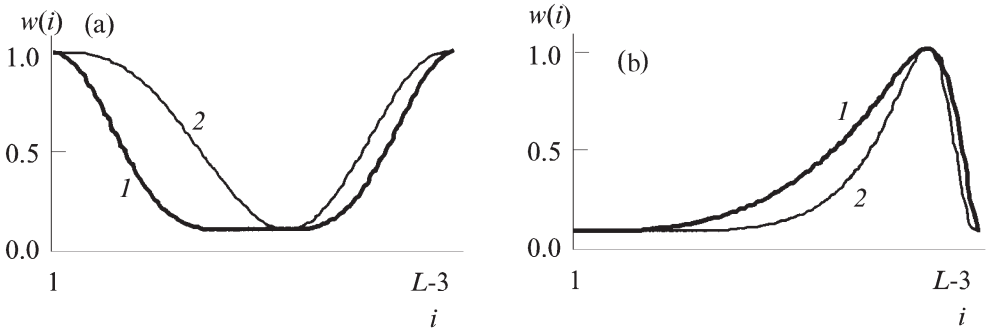

Fig. 1. Examples of weight functions $w(i)$ for terminal (a) and internal (b) important positions differing by the important region size (curves *1* and *2*).

position i in site S with sequence $\{s_i\}_{1\le i\le L}$ of length L in alphabet $s_i \in \{A, T, G, C\}$; N is the sample of sites; $n(Z_i)$ is the number of sites with tetranucleotide Z in position i; $f(Z_i) = n(Z_i)/N$ is the frequency of such sites.

In this notation the main idea of the method reduces to estimation of the significance $P(Z_i)$ of localization of tetranucleotide Z in position i using the binomial criterion. The hypothesis "Z_i is more frequent in normal sites than in defective sites" was verified using formula

$$P_+(Z_i) = \sum_{m=0}^{n_-(Z_i)} C_{N_-}^m \times f_+(Z_i)^m \times [1 - f_+(Z_i)]^{N_- - m} < 0.05, \quad (1)$$

where the indices "+" and "–" denote normal and defective sites, respectively.

Hypothesis "Z_i is more frequent in defective sites than in normal sites" was verified by

$$P_-(Z_i) = \sum_{m=0}^{n_+(Z_i)} C_{N_+}^m \times f_-(Z_i)^m \times [1 - f_-(Z_i)]^{N_+ - m} < 0.05. \quad (2)$$

Application of formulas (1) and (2) leads to the following problem. Since for an arbitrary site there is no information about significant locations of tetranucleotides $\{Z_i^*\}$, the formulas should be applied to all possible locations of tetranucleotides $\{Z_i\}$. On one hand, this allows one to determine the significant locations of tetranucleotides $\{Z_i^*\}$. On the other hand, for a site of length L these formulas are applied $15^4 \times (L - 3)$ times. Thus, in addition to the significant locations of tetranucleotides $\{Z_i^*\}$, another $2\text{x}15^4 \times (L - 3) \times 0.05$ random locations of tetranucleotides $\{Z_i\}$ are selected. To avoid this problem when formulas (1) and (2) are applied, we used the computer system SITEVIDEO [12] developed for identification of significant concentrations of tetranucleotides $\{Z_w^*\}$ (regardless of their exact location $\{Z_i^*\}$).

The system SITEVIDEO was used to analyze two sets of nucleotide sequences $\{S_{YES}\}$ and $\{S_{NO}\}$ that differ by the presence/absence of some feature. For a sequence S the weighted concentration of tetranucleotide Z was computed using

$$Z_w(S) = \sum_{i=1}^{L-3} \delta(s_i s_{i+1} s_{i+2} s_{i+3} = z_1 z_2 z_3 z_4) \times w(i), \quad (3)$$

where $\delta(x = y) = 1$ for $x = y$, $\delta(x = y) = 0$ for $x \ne y$; $w(i)$ is the weight of position i (the weight is high for important positions; $0 \le w(i) \le 1$).

Figure 1 gives examples of functions $w(i)$. The system uses 180 functions $w(i)$ differing by the size and localization of the most important positions.

Applying formula (3) at fixed Z and w to all sequences from $\{S_{YES}\}$ and $\{S_{NO}\}$, SITEVIDEO constructs their distribution $\{Z_w\}\{S_{YES}\}$ and $\{Z_w\}\{S_{NO}\}$. The significance of differences between these distributions is estimated depending on the means α_1, variations α_2, densities α_3, and interval of values α_4. It also computes the significance α_5 and α_6 of fit of $\{Z_w\}\{S_{YES}\}$ and $\{Z_w\}\{S_{NO}\}$ respectively to the Gaussian distribution. Dependent on $\{\alpha_t\}_{1\le t\le 6}$, SITEVIDEO ascribes to $\{Z_w\}$ the estimate of its utility for discrimination between the samples $\{S_{YES}\}$ and $\{S_{NO}\}$:

Table 1. Weighted concentrations of tetranucleotides significant for the *td* intron insertion site as identified by SITEVIDEO [12]

Sample		Weighted concentration, Z_w			Utility, $U(Z_w)$	Significance, α	Correlation, r
Sequence type	Size	Tetranucleotide $Z = z_1z_2z_3z_4$	Weight, $w(i)$, figure	Mean ± standard deviation			
Normal sites	49	YWMG	1a, curve *1*	0.09 ± 0.24	0.901	$<10^{-6}$	0.02
Defective sites	34			0.95 ± 0.73			
Normal sites	49	VBWA	1b, curve *1*	0.30 ± 0.39	0.894	$<10^{-6}$	
Defective sites	34			1.24 ± 0.59			
Normal sites	49	AWAW	1b, curve *2*	1.24 ± 0.80	0.996	$<10^{-7}$	–0.02
Random DNA	100			0.26 ± 0.48			
Normal sites	49	WAWY	1a, curve *2*	2.13 ± 0.54	0.992	10^{-40}	
Random DNA	100			0.66 ± 0.67			

Table 2. Contextual signals and antisignals of the *td* intron insertion sites

Tetranucleotide localization, $Z_i = z_iz_{i+1}z_{i+2}z_{i+3}$	Site type	Number of sites		Frequency $f(Z_i)$	Significance	Signal type
		total	have Z_i			
$V_{-19}B_{-18}W_{-17}A_{-16}$	normal	49	39	0.80	$P_+ < 0.025$	signal
	defective	34	21	0.61		
$A_{11}W_{12}A_{13}W_{14}$	normal	49	45	0.92	$P_+ < 10^{-4}$	signal
	defective	34	23	0.67		
$W_{-Ç0}A_{-29}W_{-28}Y_{-27}$	normal	49	27	0.55	$P_+ < 10^{-3}$	signal
	defective	34	9	0.26		
$V_7B_8W_9A_{10}$	normal	49	5	0.10	$P_- < 10^{-7}$	antisignal
	defective	34	16	0.47		
$Y_{-27}W_{-26}M_{-25}G_{-24}$	normal	49	0	0.00	$P_- < 10^{-3}$	antisignal
	defective	34	5	0.15		
$Y_{15}W_{16}M_{17}G_{18}$	normal	49	3	0.06	$P_- < 10^{-5}$	antisignal
	defective	34	12	0.35		

$$U_t(Z_w) = \begin{cases} 1, \text{ if } \alpha_t < 0.01; \\ 1.3 - 28.3 \times \alpha_t + 55.6 \times \alpha_t^2, \text{ if } 0.01 \le \alpha_t \le 0.1; \\ -1, \text{ if } \alpha_t > 0.1. \end{cases} \quad (4)$$

Formula (4) ascribes to Z_w the maximal utility $U_t(Z_w) = 1$ for significant t-difference between $Z_w\{S_{YES}\}$ and $Z_w\{S_{NO}\}$ ($a_t < 0.01$); minimal utility $U_t(Z_w) = -1$ for random difference ($a_t > 0.1$); intermediate utility $U_t(Z_w)$ from –1 to 1 for intermediate difference ($0.01 \le \alpha_t \le 0.1$). To decrease heterogeneity, SITEVIDEO applies this formula to 100 random half-samples $\{Z_w\{S_{YES}\}_g\}_{1 \le g \le 100}$ and $\{Z_w\{S_{NO}\}_g\}_{1 \le g \le 100}$, computes 600 estimates of utility Z_w, and takes the average

$$U(Z_w) = \frac{\sum_{t=1}^{6}\sum_{g=1}^{100} U_{tg}(Z_w)}{600}. \quad (5)$$

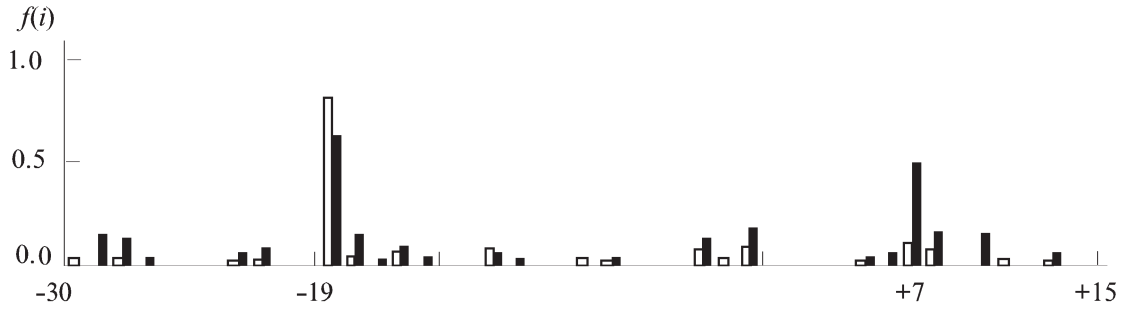

Fig. 2. Positional frequency $f(i)$ of tetranucleotide VBWA in the normal and defective sites (white and dark columns, respectively). Horizontal axis: position i.

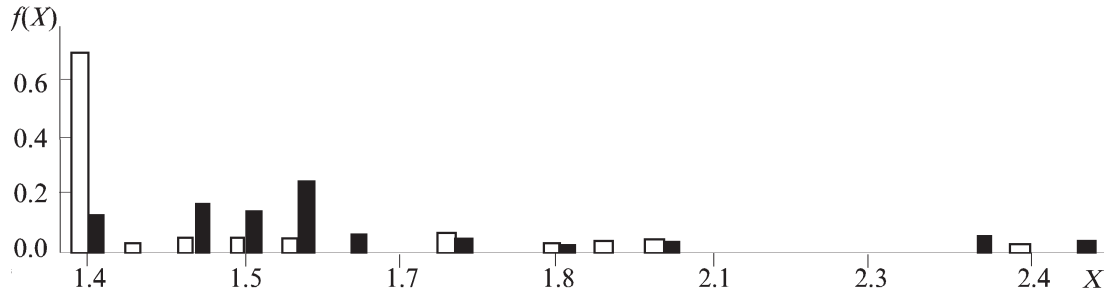

Fig. 3. Histogram of B-helical roll angle averaged over positions 9 through 18 of normal and defective sites (white and dark columns, respectively). Horizontal axis: average roll (X, roll in degrees). Vertical axis: frequency $f(x)$.

Using (5), SITEVIDEO calculates utilities $U(Z_w)$ for all $180 \times 15^4 \approx 10^7$ possible concentrations of tetranucleotides Z_w and selects tetranucleotides Z_w^* that have positive utility $U(Z_w^*) > 0$ and do not correlate with more effective Z_w. Formula (5) ascribes $U(Z_w) > 0$ if more than half of 600 differences between $Z_w\{S_{YES}\}$ and $Z_w\{S_{NO}\}$ are significant. According to the binomial criterion [13], the probability to pick a random concentration Z_w with $U(Z_w) > 0$ once is less than 10^{-40}; the probability to pick up one such concentration out of 10^{-7} possibilities is less than $10^7 \times 10^{-40} = 10^{-33}$.

Thus, the method applied is to use SITEVIDEO [12] in order to find significant concentrations of tetranucleotides $\{Z_w^*\}$, and to determine, using formulas (1) and (2), the significant locations $\{Z_i^*\}$ of these tetranucleotides $\{Z^*\}$.

RESULTS AND DISCUSSION

The described method was used to analyze the experimental data from [10]. The results obtained are presented in Tables 1 and 2. The use of SITEVIDEO allowed us to determine that the highest utility $U = 0.901$ for discrimination between 49 normal and 34 defective insertion sites of the *td* intron was observed for the concentration of tetranucleotides YWMG weighted by the function $w(i)$ corresponding to curve *1* in Fig. 1a. This concentration was 0.09 ± 0.24 for normal sites and 0.95 ± 0.73 for defective sites, and the difference between these values is highly significant ($\alpha < 10^{-6}$). The concentration of VBWA weighted by $w(i)$ corresponding to curve *1* in Fig. 1b also had high utility ($U = 0.894$). The low correlation $r = 0.02$ between the concentrations of tetranucleotides YWMG and VBWA show that they are independent (Table 1).

Following the conventional approach, we also compared 49 normal sites $\{S_{YES}\}$ and 100 random sites $\{S_{NO}\}$. The highest utilities $U = 0.996$ and $U = 0.992$ for discrimination of these samples were obtained for two independent ($r = -0.02$) concentrations of tetranucleotides AWAW and WAWY weighted by functions represented by curves *2* in Figs. 1a and 1b, respectively.

Thus, SITEVIDEO identified four tetranucleotides YWMG, VBWA, AWAW, and WAWY whose

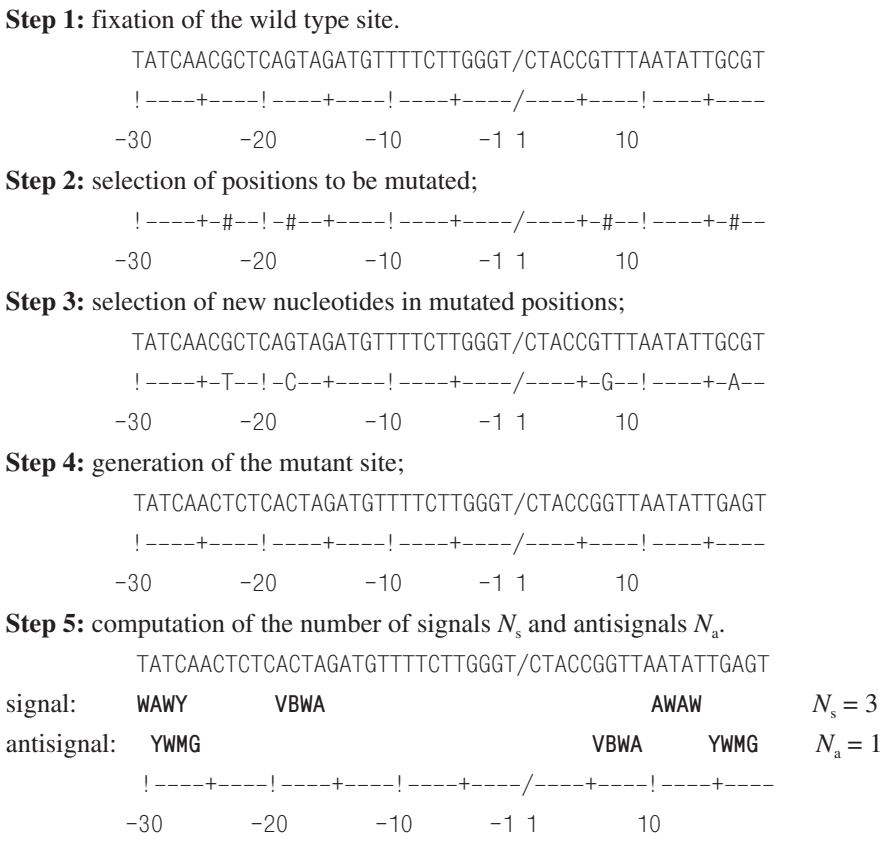

Fig. 4. Modeling of mutations in the wild-type *td* intron insertion site.

concentrations are significant for the normal insertion sites of the *td* intron (Table 1). Testing of these tetranucleotides with formulas (1) and (2), six significant locations were identified, three of which were more frequent in the normal sites, and the other three, in the defective sites (Table 2).

For example, consider tetranucleotide VBWA. Its positional frequency in normal and defective sites (white and dark columns) is shown in Fig. 2. One can see that in the normal sites it occurs almost exclusively in $V_{-19}B_{-18}W_{-17}A_{-16}$. This localization was observed in $n_+ = 39$ out of $N_+ = 49$ normal sites and in $n_- = 21$ out of $N_- = 34$ defective sites (frequencies $f_+ = 0.80$ and $f_- = 0.60$ respectively). According to formula (1), tetranucleotide $V_{-19}B_{-18}W_{-17}A_{-16}$ occurs more frequently in the normal sites than in the defective sites ($P_+ < 0.025$). This means that the wild-type site becomes defective if $V_{-19}B_{-18}$ $W_{-17}A_{-16}$ is damaged. Thus we call $V_{-19}B_{-18}W_{-17}A_{-16}$ signal of this site. In Fig. 2 one can see that the normal and defective sites differ by tetranucleotide $V_7B_8W_9A_{10}$ ($n_- = 16$, $N_- = 34$, $n_+ = 5$, $N_+ = 49$, $f_- = 0.47$ and $f_+ = 0.10$). According to formula (2), $V_7B_8W_9A_{10}$ occurs more frequently in the defective sites than in the normal sites ($P_- < 10^{-7}$). This means that a site becomes defective if in addition to "signal" tetranucleotide $V_{-19}B_{-18}W_{-17}A_{-16}$ it acquires "defective" tetranucleotide. Thus we call $V_7B_8W_9A_{10}$ an "antisignal." This result agrees with the statistical mechanics of protein-binding sites in DNA [14], according to which identical sites compete for the protein and inhibit each other. Previously we have observed that near splicing sites, TATA boxes, and AUG start codons, the alternative variants are observed [15].

Finally, in Fig. 2 one can see that tetranucleotide VBWA is practically absent from all other positions of the normal and defective sites. Therefore, significance estimates with formulas (1) and (2) are meaningless.

Analogous analyses have identified signals $W_{-30}A_{-29}W_{-28}Y_{-27}$, $A_{11}W_{12}A_{13}W_{14}$, and antisignals $Y_{-27}W_{-26}M_{-25}G_{-24}$, $Y_{15}W_{16}M_{17}G_{18}$ (Table 2; $P_+ < 10^{-4}$, $P_+ < 10^{-3}$, $P_- < 10^{-3}$ and $P_- < 10^{-5}$ respectively). One can observe the following regularity: antisignal $Y_{-27}W_{-26}M_{-25}G_{-24}$ is preceded by signal $W_{-30}A_{-29}W_{-28}Y_{-27}$; antisignal $Y_{15}W_{16}\,M_{17}G_{18}$ is preceded by signal $A_{11}W_{12}A_{13}W_{14}$. Since the antisignals do not damage the preceding signals, the remaining explanation is that they damage the DNA conformation. We have tested this conjecture using the known dinucleotide helical angles [16].

Figure 3 shows histograms of the roll angle [17] in positions 9 through 18 of the site containing signal $A_{11}W_{12}A_{13}W_{14}$ and antisignal $Y_{15}W_{16}M_{17}G_{18}$. The roll was $1.4 \pm 0.1°$ per nucleotide for 35 out of 49 normal sites and for 5 out of 34 defective sites (white and dark columns, respectively). According to formula (1) this difference is significant ($P_+ < 10^{-11}$). Thus antisignal $Y_{15}W_{16}M_{17}G_{18}$ can allosterically inhibit signal $A_{11}W_{12}A_{13}W_{14}$. This also is an evidence of significance of signal $A_{11}W_{12}A_{13}W_{14}$ and antisignal $Y_{15}W_{16}M_{17}G_{18}$ for the *td* intron insertion site.

For additional verification of the importance of the observed signals and antisignals, we have predicted the frequency of defective site $f_-(k)$ after k substitutions in the wild-type site. The following computational model was used (Fig. 4): 1, get the wild type site; 2, select at random k positions; 3, select random substitutions; 4, get the mutated site; 5, compute the number of signals N_s and antisignals N_a; 6, check whether the site is normal by the rule

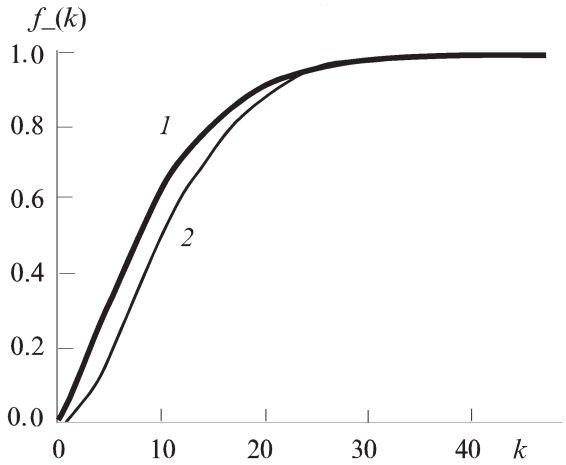

Fig. 5. Predicted frequency of defective sites and $f_s(k)$ for k random substitutions in the wild-type *td*-insertion site: curve *1* for $f_-(k)$ computed by (6); curve *2* for $f_s(k)$ computed by (8).

$$\text{If } \{N_s > 1;\ N_a < 1\} \text{ then } \{\text{site is normal}\} \text{ otherwise } \{\text{site is defective}\}. \quad (6)$$

Formula (6) checks for absence of all antisignals and allows for absence of one signal, since none of the signals was found in all 49 normal sites (Table 2). This formula was used to test 10^{-7} mutated sites for each number of substitutions k from 1 through 48, and to compute the frequency of defective sites $f_-(k)$ (Fig. 5, curve *1*). As expected, $f_-(k)$ increases with increasing k. Agreement between the predicted frequencies $f_-(k)$ and the experimental data was tested with the χ^2 criterion:

Table 3. Reliability of predicted frequencies of defective sites

Number of substitutions k	Experimental data from [10]			Predicted frequencies of defective sites	
	Number of sites		Frequency of defective sites	Formula (6) (signals and antisignals)	Formula (8) (signals)
	total	defective			
4	9	2	0.22	0.24	0.11
5	10	3	0.30	0.31	0.17
6	8	3	0.38	0.38	0.23
7	3	2	0.67	0.45	0.29
8	4	3	0.75	0.51	0.36
9	4	2	0.50	0.56	0.42
χ^2				1.57	8.12
Significance				$\alpha < 0.05$	$\alpha > 0.10$

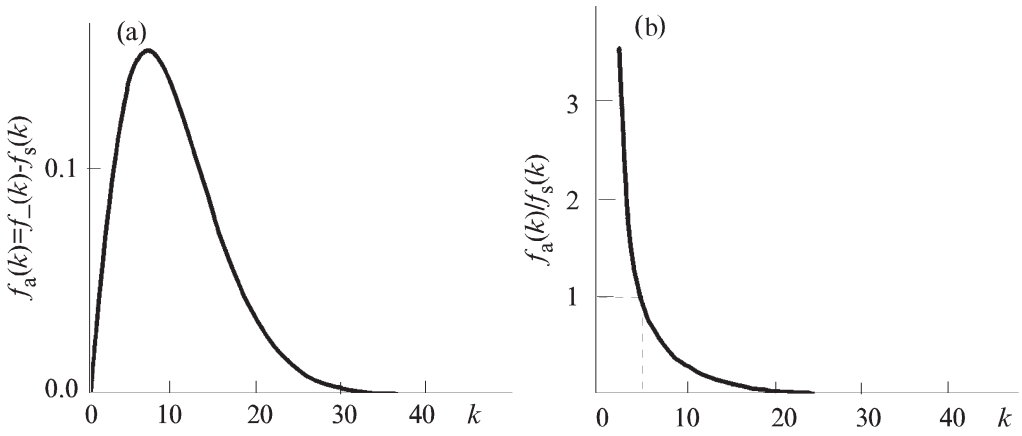

Fig 6. Contribution of antisignals to the mutation spectrum of the *td* intron insertion site: absolute (a); relative (b). Horizontal axis: the number of substitutions *k*.

$$\chi^2 = \sum_{k=a}^{b} \frac{[N_-(k) - N(k) \therefore \phi 6 f_-(k)]^2}{N(k) \times f_-(k) \times [1 - f_-(k)]}, \qquad (7)$$

where $N(k)$ and $N_-(k)$ are the numbers of all and defective sites with k substitutions in the experiment from [10].

Formula (7) was applied for the number of substitutions k from 4 through 9, since only in this interval both normal and defective sites were observed in the experiment [10]. This formula was inapplicable to the mutant sites with $k = \{1, 2, 3, 10, 11, 12, 13, 15, 18\}$ also observed in the experiment, since in these cases either only normal or only defective sites were observed. The comparison of the predicted and observed frequencies of defective sites $f_-(k)$ is given in Table 3. For example, for $k = 4$ substitutions, formula (6) predicts $f_-4 = 0.24$ defective sites, whereas in [10] $N_-(4) = 2$ sites out of $N(4) = 9$ were defective, yielding frequency $2/9 = 0.22$. In this case the predicted frequency of the defective sites was quite close to the experimental one. In Table 3 one can see that for k from 5 through 9 the predicted frequencies are close to the observed ones. The $c^2 = 1.57$ value computed by formula (7) shows good agreement between prediction and observation in this case ($\alpha < 0.05$). This means, that taking into account all signals and antisignals found above, one can predict the frequencies of defective *td* intron insertion sites $f_-(k)$ that agree with the experimental data [10]. This also demonstrates the significance of the observed signals and antisignals.

Since formula (6) successfully predicts the total contribution of the signals and antisignals to the mutation spectrum of the *td* intron insertion site, we have considered independent contribution of the signals and the antisignals. The above described algorithm was modified so as to substitute (6) for the following formula:

$$\text{If } \{Ns > 1\} \text{ then } \{\text{site is normal}\}, \text{ else } \{\text{site is defective}\} \qquad (8)$$

Formula (8), unlike (6), takes into account only the damage to the signals. Thus predicted frequencies of defective sites $f_-(k)$ with k substitutions (from 1 through 48) are shown in Fig. 5 (curve *2*). Comparison of these frequencies with the experimental data is presented in Table 3. One can see that restriction of the analysis to the signals only yields underestimated frequencies of defective sites ($\chi^2 = 8.12$, $\alpha > 0.1$). This means that the antisignals contribute significantly to the mutation spectrum:

$$f_a(k) = f_-(k) - f_s(k), \qquad (9)$$

where $f_-(k)$ and $f_s(k)$ are the frequencies of the defective sites computed by formulas (6) and (8) respectively.

The contribution of antisignals computed by (9) is shown in Fig. 6. The absolute value $f_a(k)$ reaches maximum $f_a(k) = 0.15$ at $k = 7$ (Fig. 6a). At $k \le 5$ the contribution of antisignals exceeds the contribution of signals and the relative value $f_a(k)/f_s(k) > 1$ (Fig. 6b). Thus at the low number of random substitutions $k \le 5$ the mutation spectrum of the *td* intron insertion sites

is determined by appearance of the antisignals, whereas at the large number of substitutions $k > 5$ it is determined by the damage to the signals. Thus the experimental results of [10] contained implicit information about both signals and antisignals of this site. This information was uncovered using SITEVIDEO [12] and the binomial criterion (formulas (1) and (2)).

Finally, one should note that the negative contextual features are used for recognition of some types of sites [15,18–21]. In particular, in previous studies we have observed that the use of antisignals in addition to the conventional consensus approach allowed for twice lower overprediction (the type 2 error) at a constant type 1 error in analysis of splicing sites [12] and eukaryotic promoters of transcription [21]. However, until now the biological relevance of this improvement was not clear. The present analysis shows that the presence of antisignals may inhibit a potential site even if there is almost no deviation from the consensus. This gives hope that identification and use of antisignals would improve the reliability of site recognition algorithms.

REFERENCES

1. Jonson, J., Norberg, T., Carlsson, L., *et al*., *Nucl. Acids Res*., 1993, vol. 21, p. 733.
2. Barrick, D., Villanueba, K., Childs, J., *et al*., *Nucl. Acids Res*., 1994, vol. 22, p. 1287.
3. Gelfand, M.S., *J. Comp. Biol*., 1995, vol. 2, p. 87.
4. Kraus, R.J., Murray, E.E., Wiley, S.R., *et al*., *Nucl. Acids Res*., 1996, vol. 24, p. 1531.
5. Mulligan, M., Hawley, D., Entriken, R., *et al*., *Nucl. Acids Res*., 1984, vol. 12, p. 789.
6. Berg, O.B. and von Hippel, P.H., *J. Mol. Biol*., 1987, vol. 193, p. 723.
7. Gelfand, M.S., *J. Mol. Evol*., 1992, vol. 35, p. 239.
8. Ulyanov, A.V. and Stormo, G.D., *Nucl. Acids Res*., 1995, vol. 23, pp. 1434–1440.
9. Ponomarenko, M., Kolchanova, A., Kolchanov, N., *J. Comput. Biol*., 1997, vol. 4, p. 83.
10. Bryk, M., Quirk, S., Mueller, J., *et al*., *EMBO J*., 1993, vol. 12, p. 2141.
11. Dujon, B., Belfort, M., Butow, B.A., *et al*., *Gene*, 1989, vol. 82, p. 115.
12. Kel. A., Ponomarenko, M., Likhachev, E., *et al*., *CABIOS*, 1993, vol. 9, p. 617.
13. Leman, E., *Proverka statisticheskikh gipotez* (Statistical Hypothesis Testing), Moscow: Nauka, 1979.
14. Berg, O.B., *J. Biomol. Struct. Dynam*., 1988, vol. 2, p. 265.
15. Kolchanov, N.A., *et al*., *Proc. 2nd Internat.Conf. on Bioinformatic, Supercomputing and Complex Genome*, St. Petersburg, 1992, p. 445.
16. Suzuki, M. and Yagi, N., *Nucleic Acids Res*., 1995, vol. 23, p. 2083.
17. Dickerson, R.E., Bansal, M., Calladine, C.R., *et al*., *EMBO J*., 1989, vol. 8, p. 1.
18. Rodier, F. and Sallantin, J., *Biochemie*, 1985, vol. 67, p. 533.
19. Quinqueton, J. and Moreau, J., *Biochemie*, 1985, vol. 67, p. 541.
20. Sallantin, J., *Biochemie*, 1985, vol. 67, p. 549.
21. Milanesi, L., Kolchanov, N.A., Rogozin, I.B., Kel, A.E., and Titov, I.I., *Guide to Human Genome Computing*, Bishop, M.J., Ed., Cambridge: Acad. Press, 1994, p. 249.