ISSN 0026-8933, Molecular Biology, 2009, Vol. 43, No. 3, pp. 472–479. © Pleiades Publishing, Inc., 2009.
Original Russian Text © P.M. Ponomarenko, M.P. Ponomarenko, I.A. Drachkova, M.V. Lysova, T.V. Arshinova, L.K. Savinkova, N.A. Kolchanov, 2009, published in Molekulyarnaya Biologiya, 2009, Vol. 43, No. 3, pp. 512–520.

MATHEMATICAL AND SYSTEM BIOLOGY

UDC 577.113.3:57.087.1

Prediction of the Affinity of the TATA-Binding Protein to TATA Boxes with Single Nucleotide Polymorphisms

P. M. Ponomarenko[a], M. P. Ponomarenko[b], I. A. Drachkova[b], M. V. Lysova[b], T. V. Arshinova[b], L. K. Savinkova[b], and N. A. Kolchanov[a,b]

[a] *Novosibirsk State University, Novosibirsk, 630090 Russia*

[b] *Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

e-mail: savinkl@mail.ru

Received July 17, 2008
Accepted for publication September 9, 2008

Abstract—The TATA-binding protein (TBP) is the subunit of basal transcription factor TFIID that recognizes the TATA boxes of TATA-containing promoters in class II genes, binds to them, and starts the assemblage of the RNA polymerase II basal transcription complex. The sequence of the TATA box with its flanking regions affects the levels of basal and activated transcription. The association of polymorphic TATA boxes with human hereditary diseases supports the hypothesis that TBP–TATA interaction modulates gene expression in vivo. The objective of this work is to determine changes in the TBP/TATA affinity related to polymorphisms in TATA boxes of human promoters. Changes in TBP/TATA affinities were assessed in silico by using our equation for equilibrium TBP/TATA binding in four consecutive steps: nonspecific binding ↔ sliding ↔ braking (stopping) ↔ stabilization. Our predictions agree with known examples of TATA-box polymorphisms and human hereditary diseases associated with them.

DOI: 10.1134/S0026893309030157

Key words: TATA-binding protein, TATA box, polymorphism, affinity alteration

INTRODUCTION

Transcription by RNA polymerase II involves the assemblage of a basal transcription complex (BTC) on the core promoter and the interaction of numerous transcription factors with one another and the DNA promoter region. By a core promoter is meant a 40–50 bp long DNA region around a transcription start site. It determines the accurate positioning of the transcription complex relative to the start nucleotide [1]. A core promoter is also considered to be a complex of functional elements, which includes a set of instructions specific for each promoter in the form of certain DNA sequences [2]. The core promoter structure includes the most widespread elements: the TATA box, located approximately 30 bp upstream from the transcription start [1, 2]; the Inr (initiator) element, located at the transcription start; BRE elements, which are located on the left and right of the TATA box and are recognizable by the basal transcription factor TFIIB; DPE, a downstream promoter element located 28 bp downstream from the transcription site; and MTE, a 10-bp long sequence on the left of DPE [2]. It should be noted that there is no universal set of core promoter elements, and each promoter has its own set of such elements [2].

At present, it is thought that TATA-containing promoters constitute no more than 20–25% of the total number of class II genes [1, 2]. Binding of the TATA-binding protein (TBP) within TFIID (consisting of TBP and auxiliary factors) with such promoters induces the assemblage of the RNA polymerase II transcription complex. The recruitment of TBP is a rate-limiting step in transcriptional initiation in vivo [3]. The binding of TBP with the TATA box induces a sharp DNA strand kink (80–100°) [4, 5], which is stabilized by interaction with TFIIB and TFIIA [6]. Unlike most DNA-binding proteins, which specifically interact with the major DNA groove to form hydrogen bonds between amino acid residues and nucleotides, TBP interacts with the minor groove and recognizes DNA features: flexibility or deformability [7]. Specific TBP binding is controlled to a large extent by the recognition of the DNA structure, because the symmetric location of donor and acceptor groups in the minor groove hampers discrimination between AT and TA (as well as GC and CG), whose combination forms a TATA box.

It has been shown that TBP specifically binds not only the canonical TATA sequence (TATAAAA) but also sequences notably different from it [8]. Identical mutations in TATA boxes in the context of different promoters differently influence the promoter function.

The replacement of a nucleotide in the promoter of the *hsp70* gene (TATA ⟶ TGTA) reduces the transcription rate nearly tenfold, and corresponding replacements in the *his3* promoter are reduced nearly hundredfold [9, 10]. Different effects of mutations in the $T_1A_2T_3A_4A_5A_6A_7$ consensus were observed in a study of the activities of 25 promoters [11]. Mutations at positions 1 and 6 and 3 and 6 ($T_1 \longrightarrow C$, $T_3 \longrightarrow C$, $A_6 \longrightarrow G$) reduced promoter activities by 70% in comparison with the wild-type sequences. Mutations yielding the sequences TAAAAA and TATGAA reduced promoter activities to the undetectable level. An important role of the sequences adjacent to the TATA box has been revealed: in spite of identical mutations, TATA boxes can be good or bad substrates depending on context. Sequences flanking the TATA box can also affect the affinity of TBP to TATA. They have been found to influence the basal transcription level and response of the promoter to activators [12]. The replacement of flanking regions of the strong adenoviral *MLP* promoter with those of the weak adenoviral basal *E4* promoter highly responsive to activators yields an *MLP* promoter with stronger response to activators and vice versa. The interaction between TBP and the TATA box virtually equally depends on the flanking sequences and the TATA boxes proper. The sequence motifs in flanking regions that influence TBP/TATA-box interaction are usually unique to a certain TATA-box sequence [13].

In spite of the broad variability of TATA elements and TBP tolerance of it, single nucleotide polymorphisms (SNPs) in TATA elements can cause severe disorders, as shown by human hereditary diseases associated with the aforementioned examples of TATA box polymorphisms.

In our earlier experimental and in silico studies of SNPs in eukaryotic TATA boxes, we designed a step-by-step model of TBP binding to TATA boxes [14, 15]. Here, we apply this model to the analysis of TATA elements of human promoters bearing SNPs associated with some diseases. We also present the in silico prediction of altered TBP affinity to TATA.

Our results point to the potential of this model for the detection and annotation of SNPs with unknown functions, which may influence important physiological processes in humans and other organisms.

EXPERIMENTAL

The affinities of TBP to TATA, $-\mathrm{Ln}[K_{D,\,TATA}](S^0)$, were determined for each gene in the first column of the table from the $S^0 = \{s_1, s_2, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_{25}, s_{26}\}$ sequences of 26 bp in length from the second column (normal TATA box of the gene) according to the equation deduced in our previous work [14]:

$$-\mathrm{Ln}[K_{D,\,TATA}](S^0) = 10.9 + 0.15\mathrm{PWM}_{TATA}(S^0) - 0.2\mathrm{Ln}[K_{D,\,ssDNA}](S^0) - 0.23\mathrm{Ln}[K_{D,\,dsDNA}](S^0). \quad (1)$$

Here 10.90 is the experimentally measured affinity of TBP to random DNA [8], $\mathrm{PWM}_{TATA}(S^0)$ is the maximum value of the Bucher weight matrix criterion for the TATA box [16] of all variants of its application to sequence S^0 and the complementary strand; $-\mathrm{Ln}[K_{D,\,ssDNA}](S^0)$ is the mean TBP affinity to each of the single strands (ssDNA) at the Bucher maximum position, evaluated according [17]; $-\mathrm{Ln}[K_{D,dsDNA}](S^0)$ is the mean TBP affinity to the double-stranded DNA (dsDNA), evaluated according [18], for each stretch of the DNA strand, direct or reverse, where the maximum Bucher criterion was detected; 0.15, 0.2, and 0.23 are stoichiometric coefficients of the three steps of TBP binding to TATA [14].

The affinity of TBP to TATA in the sequence $S^{\#} = \{s_1, s_2, .., s_{i-1}, t_i, s_{i+1}, \ldots, s_{25}, s_{26}\}$ of the allelic TATA box variant with the mutation $s_i \longrightarrow t_i$ (column 4) was also evaluated according to Eq. (1). Correspondingly, the change of the affinity in this case,

$$\delta\{-\mathrm{Ln}[K_{D,\,TATA}](S^0; s_i \longrightarrow t_i)\} = -\mathrm{Ln}[K_{D,\,TATA}](S^{\#}) - \{-\mathrm{Ln}[K_{D,\,TATA}](S^0)\}, \quad (2)$$

is shown in column 5 of the table. The phenotypic manifestation and association with diseases are shown in the next two columns, 6 and 7.

To obtain a reliable assessment of changes $\delta\{-\mathrm{Ln}[K_{D,\,TATA}](S^0; s_i \longrightarrow t_i)\}$, we calculated the confidence boundaries of δ at $\alpha < 0.05$ according to the Student *t* test with ν degrees of freedom assuming the hypothesis "H_0: $\delta\{-\mathrm{Ln}[K_{D,\,TATA}](S^0; s_i \longrightarrow t_i)\} = 0$":

$$\pm\delta_{5\%}\{-\mathrm{Ln}[K_{D,\,TATA}](S^0)\} = \tau_{\alpha<0.05,\,\upsilon=3\times26-1=77} \times \sqrt{\frac{\sum_{i=1}^{26}\sum_{k=1}^{3}\delta\{-\mathrm{Ln}[K_{D,\,TATA}](S^0; s_i \longrightarrow t_{i,k})\}^2}{3\times26\times(3\times26-1)}}, \quad (3)$$

where $t_{i,\,k=1}$, $t_{i,\,k=2}$, $t_{i,\,k=3}$ are all the three possible replacements of nucleotide s_i at position i of the normal sequence S^0.

Column 3 presents confidence boundary values $\pm\delta_{5\%}$ for nucleotide substitutions in each sequence S^0 of column 2. It is important that the $\pm\delta_{5\%}$ values were within 0.06–0.22 Napierian log units, $-\mathrm{Ln}(K_D)$, the average value is 0.088. This corresponds to ±10% accuracy of measurement of the equilibrium dissociation constant K_D in our experiment [15]. Equation (1) was deduced from the results of our K_D analysis in

Predictions of changes of TBP affinity (δ) to human TATA boxes with substitutions, deletions (Δ), and insertions

Gene	Normal nucleotide sequence	$\pm\delta_{5\%}$	Mutation	δ	Associated trait	Disorder	Reference
1	2	3	4	5	6	7	8
Reduced TBP affinity to TATA							
β-Globin	cagggctggg**CATAAA**Agtcagggca	0.08	–30**T** ⟶ A	–1.46*	Deficiency of β-globin	β-Thalassemia	[19]
β-Globin	cagggctggg**CATAAA**Agtcagggca		–29**A** ⟶ G	–1.24*	Deficiency of β-globin	β-Thalassemia	[20]
″	″		–30**T** ⟶ C	–0.98*	″	″	[21]
″	″		–28**A** ⟶ G	–0.71*	″	″	[22]
″	″		–31**A** ⟶ G	–0.60*	″	″	[23]
″	″		–28**A** ⟶ C	–0.53*	″	″	[24]
″	″		–27**A** ⟶ T	–0.36*	″	″	[25]
δ-Globin	acaggaccagC**A**TAAAAggcagggca	0.08	–31**A** ⟶ G	–0.61*	Deficiency of δ-globin	δ-Thalassemia	[26]
Cytochrome P450 2A6	tttcaggcagTA**T**AAAggcaaaccac	0.10	–48**T** ⟶ G	–1.49*	Nicotine oxidase deficiency	Lung carcinoma	[27]
Superoxide dismutase 1, *SOD1*	aggtctggccT**A**TAAAgtagtcgcgg	0.09	–27**A** ⟶ G	–1.17*	SOD1 deficiency	Amyotrophic lateral sclerosis	[28]
Mannan-binding lectin	catctatttcTA**T**ATAgcctgcaccc	0.11	–35**T** ⟶ C	–1.11*	Reduced pathogen opsonization	Secondary infections	[29]
Triosephosphate isomerase, *TPI*	cgcggcgctcTATA**T**AAgtgggcagt	0.09	–24**T** ⟶ G	–1.03*	TPI deficiency (2%)	Neurosis, anemia	[30]
17β-Hydroxysteroid dehydrogenase	cgaagcaggTGATATC**A**Agcccagag	0.08	–27A ⟶ C	–0.60*	Deficiency of the estrone ⟷ estradiol equilibrium regulator (55%)	Breast cancer	[31]
Coagulation factor IX	acagctcagcTT**G**TACTT**T**ggtacaa	0.11	–26**G** ⟶ C	–0.49*	Coagulation factor deficiency	Leiden hemophilia B	[32]
			–20**T** ⟶ A	–0.23*			[32]
Coagulation factor VII	ggaggc**A**GAGAACTTTgcccgtcagt	0.07	–32**A** ⟶ C	–0.15*		Moderate coagulation	[33]
Estrogen receptor	cctctcggTC**T**TTAAAAggaagaagg	0.07	–59**T** ⟶ G	–0.31*	Deficiency (50%) of estrogen receptor β	Arterial hypertension	[34, 35]
A1 transport protein	aatcTATAAAAggaacta**g**tctcggc	0.06	–17**g** ⟶ C	–0.28*	A1 protein deficiency	Lower risk of myocardial infarction	[36]
UDP glucuronosyl transferase (*UGT 1A7*)	tccac**t**tacTATATTATAggagctta	0.09	–57**t** ⟶ G	–0.17*	Lower rate of exocarcinogen detoxication	Orolaryngeal cancer	[37]
Growth hormone 1	gccag**g**TATAAAAAgggcccacaaga	0.09	–31**g** ⟶ Δ	–0.13*	Growth hormone deficiency	Short stature	[38]
Connexin 40	aa**g**gcgacagatacGATTAAAAAgag	0.09	–44**g** ⟶ A	–0.12*	Connexin 40 deficiency (35%)	Atrial arrhythmia	[39]
Glutathione-S-transferase	ttatgt**a**gggTATAAAgcccctcccg	0.10	–63**a** ⟶ C	–0.11*	Glutathione-S-transferase deficiency (12%)	DNA damage	[40]
UCP3 transport protein	agcc**c**gtgtgTATAAgaccagtgcca	0.11	–55**c** ⟶ T	–0.05	Low metabolic rate	Obesity	[41]
Kallikrein	gataagggcTTTTAAAAg**c**ctcccca	0.06	–19**c** ⟶ G	–0.05	Kallikrein deficiency	Renal disorders	[42]
StAR	cttcag**c**ggggGACATTTAAgacgca	0.06	–33**c** ⟶ T	–0.05	Aldosterone deficiency (40%)	Arterial hypertension	[43]

Table 1. (Contd.)

Gene	Normal nucleotide sequence	$\pm\delta_{5\%}$	Mutation	δ	Associated trait	Disorder	Reference
1	2	3	4	5	6	7	8
Thrombomodulin	ggcc**g**ggcacTTATAAactcgagccc	0.10	–33**g** ⟶ A	–0.04	Thrombomodulin deficiency	Thrombophlebitis	[44]
Apolipoprotein A1	tggctgcagaCATA**A**ATAggccctgc	0.08	–27**A** ⟶ C	–0.03	Cholesterol accumulation	Stenocardia	[45]
UDP glucuronosyl transferase 1A1, *UGT1A1*	gcca**TA**TATATATATATAAgtaggag	0.06	–38**TA** ⟶ TA**TA**	–0.01	Reduced catabolism of endogenous substances in blood	Icterus, Gilbert's syndrome	[46]
			–38**TA** ⟶ TATA**TA**	–0.01			
			–38**TA** ⟶ Δ	+0.01	Rare allele	Norm	
Increase in the affinity of TBP toTATA							
Endoribonuclease, *RMRP*	tctttagggcTATAAAATA**c**tactct	0.08	–24**c** ⟶ G	+0.07	RMRP deficiency, cartilage hair hypoplasia	Norm	[48]
Transferrin	ccgg**g**aatggAATAAAgggacgcggg	0.07	–34**g** ⟶ T	+0.07	Iron transport disorder	Anemia in women	[49]
Angiotensinogen	acccctcagcTATAAATAgggc**A**tcg	0.09	–20**A** ⟶ C	+0.07	Excess of the regulator of vascular tone	Small vessel disease	[50]
Myeloblastin	gggcTATAAGAggagcttga**c**cgtgg	0.08	–7**c** ⟶ T	+0.09*	Excess of myeloblastin	Leukemia	[51]
NOS2A	atggggtgagTATAAATAc**t**tcttgg	0.10	–21**t** ⟶ C	+0.21*	Excess of NO synthase	Sclerosis, diabetes mellitus	[52]
BCR (breakpoint cluster region)	c**g**gcgcggccccTTTAAGA**g**gcccgc	0.08	–48**A** ⟶ Δ	+0.50*	Elevated production of hyperactive tyrosine kinase	More severe leukemia	[53]
Tissue factor	gccggcccTTTATAg**c**gcgcggggca	0.09	–21**c** ⟶ T	+0.52*	Excess of tissue factor	Myocardial infarction	[55]
γE-Crystallin pseudogene	gccctcctgctata**c**agccccgccgc	0.12	–28**c** ⟶ T	+0.84*	Excess of the protein pseudogene product	Cataract	[56]
Interleukin-1β	ttttgaaagc**C**ATAAAAAcagcgagg	0.08	–31**C** ⟶ T	+1.18*	Excess of interleukin 1β	Lung carcinoma, high risk of cancer	[57]
Change of the affinity in the cases of polymorphisms resulting in TATA box appearance/disappearance							
Daffy antigen/receptor for chemokines	ccttggcTCTTA**T**CTTggaagcacag	0.07	–33**T** ⟶ C	–0.78*	Absence of receptors for chemokines used by *Plasmodium vivax* to invade erythrocytes and cause malaria	Malaria resistance	[58]
Cytochrome P4502B6	gatgaaattt**T**ATAACAgggtgcaga	0.08	–82**T** ⟶ C**	–0.59*	Excess of the enzyme	Liver cancer	[59]
Progesterone receptor	aaagtcgggagataaa**g**gagccgcgt	0.07	+331**g** ⟶ A	+0.30*	Excess of progesterone receptor-B	Endometrial cancer	[60]
ψζ-Globin pseudogene	ctgccacacccacattatt**t**agaaaat	0.22	–70**t** ⟶ C	+0.56*	Transcription of the pseudogene suppresses the neighboring μ, α1, α2 globin genes	α-Thalassemia	[61]

Notes: * Significant ($\alpha < 0.05$) affinity change (Eq. (3)).

** Nucleotides in the promoter of the cytochrome P4502B6 gene are numbered with reference to the translation start [59].

[14]. Correspondingly, significant changes of TBP affinity to TATA are shown in the table with asterisks.

RESULTS AND DISCUSSION

Prediction of Lower TBP Affinity to a TATA Box with Nucleotide Substitutions

Examples of such predictions are shown in the table. The greatest decrease is predicted for SNPs in the promoters of the following human genes: β-globin (δ = 1.46 and –1.24), cytochrome P450 2A6 (*CYP2A6*), Cu/Zn superoxide dismutase (*SOD1*), mannan-binding lectin, and triosephosphate isomerase (*TPI*) (δ = –1,49, –1.17, –1.11, and –1.03, respectively).

It is known that damage of the β-globin gene cluster may cause the reduction of the level of normal β-globin chains or the production of defective β-globin. Such abnormalities are associated with human β- and δ-thalassemias. When –30T in the TATA box is replaced by A, the β-globin mRNA production constitutes as little as 13% of the normal rate [19], and the –29A > G substitution reduces the mRNA level to 25% [20]. Moreover, this substitution impairs splicing to cause the production of aberrant mRNA unable to encode β-globin. The carriers of both SNPs showed β-thalassemia.

Cytochrome P450 2A6 (*CYP2A6*) is regarded as the principal human nicotine oxidase [27]. The –48T > G substitution, which disrupts the natural TATA box [27], was predicted to reduce the affinity of TBP (δ = –1.49). In experiments, it reduced the expression of a reporter gene twofold in comparison with the wild-type promoter. This polymorphism is associated with the elevated risk of cancer in smokers.

The substitution –27A > G in the TATA box of the *SOD1* gene reduces the expression rate of this gene. It is associated with amyotrophic lateral sclerosis, a neurodegenerative disease, which causes progressive muscle weakness and atrophy throughout the body [28].

The *MBL2* gene codes for C-type mannan-binding lectin 2, which acts as a soluble receptor. Its binding to mannose residues in cell membranes is involved in the initiation and regulation of innate immune response [29]. The deficiency of MBL2 in blood serum causes the total deficiency of opsonin (blood plasma protein which binds to bacteria and macrophages and favors their coagulation) associated with risk of recurrent infections. The substitution –35T > C in the TATA box of the *MBL2* gene promoter [29] δ = –1.11) impairs the binding of TFIID to the TATA box and reduces the gene expression and pathogen opsonization.

Triosephosphate isomerase, encoded by the *TPI* gene, converts D-glyceraldehyde-3-phosphate to dihydroxyacetone phosphate. The *TPI* gene is expressed in all cell types, as it belongs to housekeeping genes. Each human tissue contains multiple TPI forms, produced by posttranslational modifications of one gene product. The mutation –24T > G (δ = –1.03) in the TATA-like sequence of the gene causes the reduction of its mRNA level. As a result, the total activity of the enzyme is 2–10% of its normal level. Carriers of this mutation show neurological and muscular disorders and hemolytic anemia [30].

The least decrease in the affinity (δ = –0.04 and –0.03) is expected for SNPs in the TATA boxes of the human genes for thrombomodulin (*TM*) and apolipoprotein A1 (*APO A1*). These SNPs are associated with thrombophlebitis (*TM*) [44] and a higher risk of stenocardia (*APO A1*) [45]. Whereas the *APO A1* polymorphism touches the sequence of the TATA box proper, the SNP in *TM* is located in the region upstream of TATA; nevertheless, the predicted degrees of affinity loss are close. It is known that the 5'-flanking regions of TATA boxes are more conservative and that they exert a greater influence on TBP binding than the downstream sequence [13]. These examples demonstrate that the efficiency of TBP interaction with TATA boxes in vivo is determined by flanking sequences in no lesser a degree than by the TATA boxes themselves [12, 13].

Borderline examples are shown at the end of the first section of the table and at the beginning of the second one. Inserts of TA into the TATA box of the gene for UDP-glucuronosyltransferase 1A1 [46] reduce gene expression and result in the predisposition to jaundice and Gilbert's syndrome, whereas the deletion is not related to any disorder. Apparently, the affinity change inferred by us (δ = –0.01 and +0.01, respectively) is in agreement with these data. Another example is the decrease of the expression of the gene for the RNA component of mitochondrial RNA processing endoribonuclease (*RMRP*) related to variation of the distance between the TATA box and transcription start owing to short deletions and insertions. It is associated with cartilage-hair hypoplasia [47]. A polymorphic variant of the TATA box was detected in a healthy person from the control group [48]. The prediction (δ = +0.07) is in agreement with the data reported in [48].

Prediction of Higher TBP Affinity to a TATA Box

The second section of the table contains examples of predictions of higher TBP affinity to a mutant TATA box. A significant increase was predicted for SNPs in the promoters of the genes for NO synthase (*NOS2A*), *BCR*, and the tissue factor (TF): δ = +0.21, +0.50, and +0.52, respectively. All these SNPs are located in regions adjacent to the TATA boxes. The prediction of increased TBP affinity to the TATA of the *NOS2A* promoter (+0.21) agrees with the elevated expression of

the gene, which is associated with an elevated risk of disseminated sclerosis, diabetes mellitus, and a predisposition to acute respiratory diseases [52].

We calculated the increase in the TBP affinity to TATA (+0.50) for the known deletion –48A at the last position of the TATA box of the fused *BCR/ABL* gene, coding for hyperactive tyrosine kinase [53]. This prediction agrees with elevated gene expression and a more severe form of chronic myelogenous leukemia [54].

Elevated affinity (+0.52) was predicted for an SNP in the TATA box of the gene for TF. It agrees with elevated gene expression and the risk of thrombophlebitis and myocardial infarction [55].

The greatest increase of the TBP affinity to TATA was predicted for promoters of the interleukin 1β-gene and the γ-crystallin cluster. In the γ-crystallin cluster (*CRYG*), two genes encode the full-size protein, and four, its fragments [56]. The mutation within the TATA box "improves" its sequence: TATACA is replaced by TATATA. The prediction is +0.84. The promoter activity increases tenfold, and the expression of the mutant gene for ψγE-crystallin, by 30% in comparison with the wild-type gene. The proteinaceous product of the pseudogene is the N-terminal fragment of γ-crystallin. Its molecular weight is 6 kDa. The elevated content of this fragment causes cataract.

The –31C > T substitution improving the gene for interleukin 1β ($\delta = +1.18$) is associated with elevated risk of non-small cell lung carcinoma [57] and hepatocellular carcinoma, as well as with the elevated sensitivity to *Helicobacter pylori*.

Prediction of Change of the TBP Affinity to a DNA That Has Either Acquired or Lost a TATA Box

The last section of the table contains predictions (Eqs. (1) and (2)) of change of TBP affinity to TATA, when a TATA either appears or disappears. The most common examples are polymorphisms in the promoters of the genes for the Daffy antigen receptor for chemokines ($\delta = –0.78$) and cytochrome P450 2B6 ($\delta = –0.59$). The malarian parasite *Plasmodium vivax* uses this chemokine receptor to invade erythrocytes and cause malaria [58]. The substitution –33T > C destroys the TATA/GATA box of the *DARC* gene, preventing the appearance of the receptor on the erythrocyte surface. Therefore, native West African carriers of this polymorphism are resistant to malaria.

The decrease of the affinity ($\delta = –0.59$) predicted for the substitution 82T > C in the *CYP2B6* gene agrees with experimental data on the promoter activity. This polymorphism causes a decrease of TBP binding to the TATA box and an increase in the affinity of C/EBP to the resulting binding site; i.e., the original TATA box is converted to a potential binding site for transcription factor C/EBP [59]. The rate of *CYP2B6* transcription increases threefold, and the content of mature mRNAs, more than twofold. It is known that cytochrome CYP2B6 can activate procarcinogens; this is why its excess is associated with the elevated risk of hepatocellular carcinoma.

We applied Eqs. (1) and (2) to 43 predictions of altered TBP affinities to TATA in all polymorphic human TATA boxes known by now. These predictions were statistically significant ($\alpha < 0.05$) in 32 cases of 43 (74%). Thus, changes of TBP affinities to TATA predicted by our equation for the equilibrium binding of TBP to TATA boxes agrees with known human TATA box polymorphisms and associated hereditary disorders. The quantitative proof of these predictions and the effects of polymorphisms on the interaction between TBP and TATA demands experimental studies, which are scheduled for the nearest future.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, projects nos. 08-04-01048 and 09-04-00653.

REFERENCES

1. Juven-Gershon T., Hsu J.Y., Kadonaga J.T. 2006. Perspectives of the RNA polymerase II core promoter. *Biochem. Soc. Trans.* **34**, 1047–1050.
2. Juven-Gershon T., Hsu J.Y., Theisen J.W.M., et al. 2008. The RNA polymerase II core promoter: The gateway to transcription. *Curr. Opin. Cell Biol.* **20**, 253–259.
3. Stewart J.J., Stargell L.A. 2001. The stability of the TFIIA–TBP–DNA complex is dependent on the sequence of the TATAAA element. *J. Biol. Chem.* **276**, 30078–30084.
4. Kim J.L., Nikolov D.B., Burley S.K. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*. **365**, 520–527.
5. Kim Y., Geiger J.H., Hahn S., Sigler P.B. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature*. **365**, 512–520.
6. Lagrange T., Kapanidis A.N., Tang H., et al. 1998. New core promoter element in RNA Polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44.
7. Faiger H., Ivanchenko M., Haran T.E. 2007. Nearest-neighbor non-additivity versus long-range non-additivity in TATA-box structure and its implication for TBP-binding mechanism. *Nucleic Acids Res.* **35**, 4409–4419.
8. Hahn S., Buratowski S., Sharp P.A., Guarente L. 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non-nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. USA*. **86**, 5718–5722.
9. Caron C., Rousset R., Beraud C., et al. 1993 Functional and biochemical interaction of the HTLV-I Tax1 transactivator with TBP. *EMBO J.* **12**, 4269–4278.

10. Perier R.C., Praz V., Junier T., Bonnard C., Bucher P. 2000. The eukaryotic promoter database (EPD). *Nucleic Acids Res*. **28**, 302–303.
11. Colgan J., Manley J.L. 1995. Cooperation between core promoter elements influences transcriptional activity *in vivo*. *Proc. Natl. Acad. Sci. USA*. **92**, 1955–1959.
12. Wolner B.S., Gralla J.D. 2001. TATA-flanking sequences influence the rate and stability of TATA-binding protein and TFIIB binding. *J. Biol. Chem.* **276**, 6260–6266.
13. Faiger H., Ivanchenko M., Cohen I., Haran T.E. 2006. TBP flanking sequences: Asymmetry of binding, long-range effects and consensus sequences. *Nucleic Acids Res*. **34**, 104–119.
14. Ponomarenko P.M., Savinkova L.K., Drachkova I.A., et al. 2008. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single-nucleotide polymorphisms. *Dokl. Akad. Nauk.* **419**, 828–832.
15. Savinkova L.K., Drachkova I.A., Ponomarenko M.P., et al. 2007. Interaction of recombinant TATA-binding protein with TATA boxes of mammalian gene promoters. *Ekol. Genet*. **5**, 44–49.
16. Bucher P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol*. **212**, 563–578.
17. Ponomarenko M.P., Savinkova L.K., Kel' A.E., Kolchanov N.A. 1997. Computer modeling of TATA box sequences of eukaryotic promoters. *Dokl. Akad. Nauk*. **355**, 557–561.
18. Ponomarenko M.P., Ponomarenko J.V., Frolov A.S., et al. 1999. Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins. *Bioinformatics*. **15**, 687–703.
19. Fei Y.J., Stoming T.A., Efremov G.D., et al. 1988. Beta-thalassemia due to a T-A mutation within the ATA box. *Biochem. Biophys. Res. Commun*. **153**, 741–747.
20. Antonarakis S.E., Irkin S.H., Cheng T.C., et al. 1984. Beta-thalassemia in American Blacks: Novel mutations in the "TATA" box and an acceptor splice site. *Proc. Natl. Acad. Sci. USA*. **81**, 1154–1158.
21. Cai S.P., Zhang J.Z., Doherty M., Kan Y.W. 1989. A new TATA box mutation detected at prenatal diagnosis for beta-thalassemia. *Am. J. Hum. Genet*. **45**, 112–114.
22. Orkin S.H., Sexton J.P., Cheng T.C., et al. 1983. ATA box transcription mutation in beta-thalassemia. *Nucleic Acids Res*. **11**, 4727–4734.
23. Takihara Y., Nakamura T., Yamada H., et al. 1986 A novel mutation in the TATA box in a Japanese patient with beta(+)-thalassemia. *Blood*. **67**, 547–550.
24. Poncz M., Ballantine M., Solowiejczyk D., et al. 1982. Beta-thalassemia in a Kurdish Jew. Single base changes in the T-A-T-A box. *J. Biol. Chem*. **257**, 5994–5996.
25. Baden C., Jassim N., Martini N., et al. 1999. Characterization of a new polymorphism, IVS-I-108 (T⟶C), and a new beta-thalassemia mutation, –27 (A⟶T), discovered in the course of a prenatal diagnosis. *Hemoglobin*. **23**, 339–344.
26. Frischknecht H., Dutly F. 2005. Two new delta-globin mutations: Hb A2-Ninive [delta133 (H11) Val-Ala] and a delta(+)-thalassemia mutation [–31 (A⟶G)] in the TATA box of the delta-globin gene. *Hemoglobin*. **29**, 151–154.
27. Pitarque M., von Richter O., Oke B., et al. 2001. Identification of a single nucleotide polymorphism in the TATA box of the *CYP2A6* gene: Impairment of its promoter activity. *Biochem. Biophys. Res. Commun*. **292**, 455–460.
28. Niemann S., Broom W.J., Brown R.H., Jr. 2007. Analysis of a genetic defect in the TATA box of the *SOD1* gene in a patient with familial amyotrophic lateral sclerosis. *Muscle Nerve*. **36**, 704–707.
29. Boldt A.B., Culpi L., Tsuneto L.T., et al. 2006. Diversity of the *MBL2* gene in various Brazilian populations and the case of selection at the mannose-binding lectin locus. *Hum. Immunol*. **67**, 722–734.
30. Watanabe M., Zingg B.C., Mohrenweiser H.W. 1996. Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. *Am. J. Hum. Genet*. **58**, 308–316.
31. Peltoketo H., Piao Y., Mannermaa A., et al. 1994. A point mutation in the putative TATA box, detected in nondiseased individuals and patients with hereditary breast cancer, decreases promoter activity of the 17 beta-hydroxysteroid dehydrogenase type 1 gene 2 (*EDH17B2*) *in vitro*. *Genomics*. **23**, 250–252.
32. Reijnen M.J., Sladek F.M., Bertina R.M., Reitsma P.H. 1992. Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proc. Natl. Acad. Sci. USA*. **89**, 6300–6303.
33. Kavlie A., Hiltunen L., Rasi V., Prydz H. 2003. Two novel mutations in the human coagulation factor VII promoter. *Thromb. Haemost*. **90**, 194–205.
34. Hilips S., Richter A., Oesterreich S., et. al. 2007. Functional activity of a genetic variant in the estrogen receptor beta TATA box. *Clin. Pharmacol. Ther*. **81**, S87.
35. Schubert C., Pencalla A., Fliegner D., et al. 2006. In: *Berliner Symposium Geschlechterforschung in der Medizin*. Berlin, p. 36.
36. Zwarts K.Y., Clee S.M., Zwinderman A.H., et al. 2002. ABCA1 regulatory variants influence coronary artery disease independent of effects on plasma lipid levels. *Clin. Genet*. **61**, 115–125.
37. Lankisch T.O., Vogel A., Eilermann S., et al. 2005. Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. *Mol. Pharmacol*. **67**, 1732–1739.
38. Horan M., Millar D.S., Hedderich J., et al. 2003. Human growth hormone 1 (GH1) gene expression: Complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. *Hum. Mutat*. **21**, 408–423.
39. Groenewegen W.A., Firouzi M., Bezzina C.R., et al. 2003. A cardiac sodium channel mutation cosegregates with a rare connexin40 genotype in familial atrial standstill. *Circ. Res*. **92**, 14–22.
40. Liu X., Campbell M.R., Pittman G.S., et al. 2005. Expression-based discovery of variation in the human glutathione S-transferase M3 promoter and functional analysis in a glioma cell line using allele-specific chromatin immunoprecipitation. *Cancer Res*. **65**, 99–104.
41. Otabe S., Clement K., Dina C., et al. 2000. A genetic variation in the 5' flanking region of the UCP3 gene is

associated with body mass index in humans in interaction with physical activity. *Diabetology.* **43**, 245–249.

42. Slim R., Torremocha F., Moreau T., et al. 2002. Loss-of-function polymorphism of the human kallikrein gene with reduced urinary kallikrein activity. *J. Am. Soc. Nephrol.* **13**, 968–976.
43. Casal A.J., Sinclair V.J., Capponi A.M., et al. 2006. A novel mutation in the steroidogenic acute regulatory protein gene promoter leading to reduced promoter activity. *J. Mol. Endocrinol.* **37**, 71–80.
44. Le Flem L., Picard V., Emmerich J., et al. 1999. Mutations in promoter region of thrombomodulin and venous thromboembolic disease. *Arterioscler. Thromb. Vasc. Biol.* **19**, 1098–1104.
45. Matsunaga A., Sasaki J., Hakamata H., et al. 1999. Compound heterozygosity for an apolipoprotein A1 gene promoter mutation and a structural nonsense mutation with apolipoprotein A1 deficiency. *Arterioscler. Thromb. Vasc. Biol.* **19**, 348–355.
46. Kaniwa N., Kurose K., Jinno H., et al. 2005. Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686C > T (P229L) found in an African-American. *Drug Metab. Dispos.* **33**, 458–465.
47. Ridanpaa M., van Eenennaam H., Pelin K., et al. 2001. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage–hair hypoplasia. *Cell.* **104**, 195–203.
48. Bonafe L., Schmitt K., Eich G., et al. 2002. RMRP gene sequence analysis confirms a cartilage-hair hypoplasia variant with only skeletal manifestations and reveals a high density of single-nucleotide polymorphisms. *Clin. Genet.* **61**, 146–151.
49. Lee P.L., Halloran C., Beutler E. 2001. Polymorphisms in the transferrin 5' flanking region associated with differences in total iron binding capacity: possible implications in iron homeostasis. *Blood Cells Mol. Dis.* **27**, 539–548.
50. Schmidt H., Aulchenko Y.S., Schweignofer N., et al. 2004. Angiotensinogen promoter B-haplotype associated with cerebral small vessel disease enhances basal transcriptional activity. *Stroke.* **35**, 2592–2597.
51. Zimmer M., Medcalf R.L., Fink T.M., et al. 1992. Three human elastase-like genes coordinately expressed in the myelomonocyte lineage are organized as a single genetic locus on 19pter. *Proc. Natl. Acad. Sci. USA.* **89**, 8215–8219.
52. Burgner D., Rockett K., Ackerman H., et al. 2003. Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes Immunity.* **4**, 506–514.
53. Lifshitz B., Fainstein E., Marcelle C., et al. 1988. BCR genes and transcripts. *Oncogene.* **2**, 113–117.
54. Tian H., Zheng W.Y., Fu Y.G., et al. 2004. Polymorphism analysis of 5' promotor region of BCR gene. *Ai Zheng.* **23**, 812–815.
55. Arnaud E., Barbalat V., Nicaud V., et al. 2000. Polymorphisms in the 5' regulatory region of the tissue factor gene and the risk of myocardial infarction and venous thromboembolism: The ECTIM and PATHROS studies (Etude Cas-Témoins de l'Infarctus du Myocarde and Paris Thrombosis Case-Control Study). *Arterioscler. Thromb. Vasc. Biol.* **20**, 892–898.
56. Brakenhoff R.H., Henskens H.A., van Rossum M.W., et al. 1994. Activation of the gamma E-crystallin pseudogene in the human hereditary Coppock-like cataract. *Hum. Mol. Genet.* **3**, 279–283.
57. Zienolddiny S., Ryberg D., Maggini V. 2004. Polymorphisms of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. *Int. J. Cancer.* **109**, 353–356.
58. Tournamille C., Colin Y., Cartron J.P., Le Van Kim C. 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.* **10**, 224–228.
59. Zukunft J., Lang T., Richter T., et al. 2005. A natural CYP2B6 TATA box polymorphism (–82T⟶C) leading to enhanced transcription and relocation of the transcriptional start site. *Mol. Pharmacol.* **67**, 1772–1782.
60. Liu Z., Wong J., Tsai S.Y., et al. 2001. Sequential recruitment of steroid receptor coactivator-1 (SRC-1) and p300 enhances progesterone receptor-dependent initiation and reinitiation of transcription from chromatin. *Proc. Natl. Acad. Sci. USA.* **98**, 12426–12431.
61. De Gobbi M., Viprakasit V., Hughes J.R., et al. 2006. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science.* **312**, 1215–1217.