An Experimental Verification of the Predicted Effects of Promoter TATA-Box Polymorphisms Associated with Human Diseases on Interactions between the TATA Boxes and TATA-Binding Protein

Ludmila Savinkova1*, Irina Drachkova1, Tatyana Arshinova1, Petr Ponomarenko1, Mikhail Ponomarenko1, Nikolay Kolchanov1,2

1 Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, Russia, 2 Novosibirsk State University, Novosibirsk, Russia

Abstract

Human genome sequencing has resulted in a great body of data, including a stunningly large number of single nucleotide polymorphisms (SNPs) with unknown phenotypic manifestations. Identification and comprehensive analysis of regulatory SNPs in human gene promoters will help quantify the effects of these SNPs on human health. Based on our experimental and computer-aided study of SNPs in TATA boxes and the use of literature data, we have derived an equation for TBP/TATA equilibrium binding in three successive steps: TATA-binding protein (TBP) sliding along DNA due to their nonspecific affinity for each other ↔ recognition of the TATA box ↔ stabilization of the TBP/TATA complex. Using this equation, we have analyzed TATA boxes containing SNPs associated with human diseases and made in silico predictions of changes in TBP/TATA affinity. An electrophoretic mobility shift assay (EMSA)-based experimental study performed under the most standardized conditions demonstrates that the experimentally measured values are highly correlated with the predicted values: the coefficient of linear correlation, r, was 0.822 at a significance level of $a < 10^{-5}$ for equilibrium $K_D$ values, $(1 / K_D)$, and 0.785 at a significance level of $a < 10^{-3}$ for changes in equilibrium $K_D$ (6) due to SNPs in the TATA boxes $(\Delta \equiv - \ln[K_D(TATA)] - (- \ln[K_D(TATAMal)])$. It has been demonstrated that the SNPs associated with increased risk of human diseases such as α, β- and δ-thalassemia, myocardial infarction and thrombophlebitis, changes in immune response, amyotrophic lateral sclerosis, lung cancer and hemophilia B Leyden cause 2–4-fold changes in TBP/TATA affinity in most cases. The results obtained strongly suggest that the TBP/TATA equilibrium binding equation derived can be used for analysis of TATA-box sequences and identification of SNPs with a potential of being functionally important.

Introduction

Single nucleotide polymorphisms (SNPs) represent the commonest type of genetic variation in man. Most of more than 15 million annotated polymorphisms are located in DNA coding regions, which makes the mechanism of their action on the phenotype perfectly clear: a certain protein will be deficient [1]. Polymorphisms in regulatory regions and the effects of these polymorphisms on gene expression remain to be much less well studied. Many of non-coding SNPs lie within regulatory DNA motifs, altering their affinity for transcription factors and thus also altering the expression levels of genes cis-targeted by those motifs, which accounts for differences in responses to external and internal signals, propensity to certain diseases, and sensitivity to therapy, to mention a few. Thus, information on polymorphisms represents a huge resource for biomedical studies, diagnostics and drug development.

The transcription of any protein-coding gene by RNA polymerase II starts with assembly of the basal transcriptional machinery on the core promoter. The core promoter is the term for the DNA region spanning ~100 nucleotides to the left (in the 5’-region) and to the right (in the 3’-region) from the transcription start site, with a variable number of regulatory regions [2] such as the TATA box, BRE (TFIIB recognition element), Inr (initiator element), MTE (motif ten element), DPE (downstream promoter element), DCE (downstream core element), and XCPE1 (X core promoter element 1) [3] and others.

The TATA box is the most well-studied core promoter element, even though the TATA-containing promoters make up only 10–16% of the genes read by RNA polymerase II, of which only 30% contain the canonical TATA box, which provides evidence that...
the canonical TATA-box sequence, TATAAAA, is variable in natural promoters [4]. The formation of the basal transcription complex on the TATA-containing promoters of genes read by RNA polymerase II starts with the recognition and binding of the TATA box by the TATA-binding protein (TBP), a subunit of the transcription factor TFIIID.

It has been demonstrated that identical TATA-box mutations in different promoters have different effects on promoter activity [5]. Although TATA boxes are so much variable, some TATA-containing promoters can be very sensitive to mutations in the TATA boxes. This statement is supported by molecular-genetically and clinically identified single nucleotide polymorphisms associated with increased risk of hereditary diseases and found in the TATA boxes of human gene promoters, [6]. The importance of knowing more about SNPs in regulatory regions (in particular, the TATA box) is even more relevant since such SNPs make individual sensitivity to bacterial and viral infection as well as the therapeutic drugs used for treatment individual. It is especially important to know how to identify and study the SNPs that are associated with risk of widespread polygenic (complex) human diseases such as arthritis, hypertension, cancer, Alzheimer disease and others.

Obviously, no search for functionally important SNPs, their analysis or systematization are possible without the use of bioinformatic approaches. However, many of the approaches that have been developed to date are not used for the lack experimental verification. To fill in the gap, we are currently running a comprehensive experimental and computer-aided study of SNPs in TATA boxes and their effects on the interactions between TPB and TATA boxes. Based on our earlier experimental [7] and computer-aided study of TBP/TATA interactions and the literature data, we had previously developed a model [8] describing the binding of TPB to TATA boxes in three successive steps: TPB sliding along DNA [4,9] due to their nonspecific affinity for each other [7,8] ↔ recognition of the TATA box [10] ↔ stabilization of the TBP/TATA complex [11]. With the use of that model, we have analyzed interactions between TPB and TATA boxes in human gene promoters containing SNPs associated with diseases and also determined in situ changes in the affinity of TPB for those SNPs. In the current work, we have performed, under the most standardized experimental conditions, a study of interactions between recombinant human TBP (hTPB) and nucleotides identical to the TATA boxes in the promoters of the genes for z, b, and g-globins, tissue factor, triosephosphate isomerase, NO synthase, superoxide dismutase, cytochrome P450 2A6, mannose-binding lectin, coagulation factor IX and interleukin-1 beta without and with SNPs associated with various pathologies. The obtained equilibrium KD values, which characterize TBP/TATA affinity without and with SNPs in TATA boxes, are highly correlated with predicted values.

**Results and Discussion**

The table 1 presents the experimentally obtained values of equilibrium dissociation constants (KD), which characterize the affinity of hTPB for ODNs. The ODNs are identical to TATA boxes with the flanking regions of the promoters of the genes being studied in healthy individuals and patients with SNPs in the TATA boxes in the promoters of the genes being studied. Also, the table 1 presents negative natural logarithms of the predicted and experimentally determined values of KD and changes in KD when TATA boxes contained SNPs: δ = −ln(KD,TATAMut) − (− ln(KD,TATA)).

**SNPs in the TATA Boxes of the β-, δ- and ψ-globin Genes**

It has been demonstrated [12–20] that SNPs in the TATA boxes of the promoters of the β-, δ-globin genes and the ψ-globin pseudogene (HbB, HbD, and Hb ψ, respectively) in man lead to β-, δ- and ψ-thalassemias of varying severity due to disruption of the balance in the synthesis of structurally normal globin chains, which compose normal hemoglobins: A (z2b2), the commonest structural unit of hemoglobin in adult humans, HbA, with an amount of ~97%) and A2 (z2b2, with an amount of ~3%) [21]. Most commonly affected is the synthesis of ψ- and β-globin chains, which corresponds to ψ- and β-thalassemia, respectively. Imbalances of any of three globin chains cause their aggregation, hemolysis and failure in erythropoiesis. The forms of hemoglobinopathies are mild, moderate and severe. While patients with mild and moderate hemoglobinopathies may have asymptomatic anemia and a normal quality of life, patients with severe hemoglobinopathies may have serious disorders such as hemolytic anemia, skeletal abnormalities, poor growth, jaundice, be dependent on transfusions, to name a few [22]. δ-thalassemia has a lower prevalence because in healthy adult humans hemoglobin A2 (z2ψ2), which consists of two δ-globin chains, exists in an amount of ~3% of hemoglobin A (z2B2) [19].

As can be seen from the sequences presented in the table 1, the first T in the TATA box of the β- and δ-globin genes in a healthy individual is replaced by C; the TATA box appears as CATAAAA instead of canonical TATAAAA [16]. The effects of this substitution in the TATA box have been studied by a range of researchers. The use of TPB from the yeast S. cerevisiae [23] demonstrated that this substitution has little effect on binding. It has also been demonstrated [24] that the promoter containing the CATAAAA sequence is 1/40 as efficient in inducing transcription in vivo as the classical TATAAAA sequence. It has also been demonstrated [25] that transcription in HeLa is decreased threefold when the first T in the TATA box is replaced by C (TATAAAA→CATAAAA). As can be seen from these examples, a T to C substitution at position 1 in the TATA box variously affect transcription and, therefore, TBP/TATA interactions. It is possible that these differences are due to differences in the sequences that flank the TATA box and differences in experimental conditions. It has been demonstrated [26,7] that the affinity of TPB for oligonucleotides with different abundances of AT pairs in the sequences flanking the TATA box can be 25–30 times as different. Our estimate of the affinity of TPB for the TATA box in the β-globin gene of healthy individuals (WT in the table 1) is KD = 44 nM.

Analysis of the equilibrium KD values for the TATA boxes in the SNP-containing promoters of the β- and δ-globin genes demonstrates that a 2–2.6-fold decrease in affinity is associated with thalassemia intermedia or thalassemia minor. Affected individuals do not depend on red blood cell transfusions and normally have a good quality of life. Only the 28A>G mutation [17] found in two Kurdish individuals, brothers aged 1.5 years, is associated with thalassemia major; both patients were dependent on red blood cell transfusions. Analysis revealed partial or total lack of normal β-globin mRNAs in them. The use of a vector with a similar mutation (TATAAAA → TATACAA) in HeLa led to a 20-fold decrease in transcription as compared to normal levels [25], which is good agreement with the characteristic that we obtained for the interaction between hTPB and ODN containing this substitution: KD = 300 nM, which is 1/8 as much as the normal affinity. The -29A>G SNP led to a nearly 9-fold decrease in hTPB/TATA affinity (KD = 390 nM) and in that case β-thalassemia intermedia was detected in an individual with a practically normal quality of life. In a work devoted to artificial mutagenesis of the β-globin
Table 1. The experimental verification of predicted changes in $K_D$ for TBP/TATA when the TATA boxes contain SNPs associated with human diseases.

| Gene       | Sequences (5’ - 3’ strands) | Allele | Norm/disease association         | Prediction | Experiment |
|------------|-----------------------------|--------|-----------------------------------|------------|------------|
| Hb$^a$     | cagaggctgggCATAAAAGtcagggca | WT     | Norm [16]                         | -ln($K_D$) | $K_D$, nM   |
|            | cagaggctgggGTAAAAAGtcagggca | -31A>G | $\beta$-thalassemia [18]          | 18.47      | 44±3       | 16.94      |
|            | cagaggctgggCATAAAAGtcagggca | -30T>A | $\beta$-thalassemia [16]          | 17.87      | -0.60      | 16.13 -0.81 |
|            | cagaggctgggCATAAAAGtcagggca | -30T>C | $\beta$-thalassemia [14]          | 17.01      | -1.46      | 15.97 -0.97 |
|            | cagaggctgggCATAAAAGtcagggca | -29A>G | $\beta$-thalassemia [12]          | 17.49      | -0.98      | 16.01 -0.93 |
|            | cagaggctgggCATAAAAGtcagggca | -28A>G | $\beta$-thalassemia [15]          | 17.23      | -1.24      | 14.76 -2.18 |
|            | cagaggctgggCATAAAAGtcagggca | -28A>C | $\beta$-thalassemia [17]          | 17.76      | -0.71      | 16.26 -0.68 |
|            | cagaggctgggCATAAAAGtcagggca | -27A>G | $\beta$-thalassemia [13]          | 18.11      | -0.36      | 16.56 -0.38 |
| Hb$^b$     | acaggaccccCACATATCAaataaat  | WT     | Norm [19]                         | 18.94      | 46±5       | 16.89      |
|            | acaggaccccCACATATCAaataaat  | -31A>G | $\delta$-thalassemia [19]         | 18.33      | -0.61      | 15.94 -0.95 |
| Mb2        | catcatttctTATAAGccgctacc   | WT     | Norm [37]                         | 19.68      | 28±4       | 17.39      |
|            | catcatttctTATAAGccgctacc   | -35T>C | $\sigma$-thalassemia [20]         | 18.28      | +0.56      | 16.00 +0.30 |
| Tpi        | cgccgcttcTATAGAGgggcat     | WT     | Norm [32]                         | 20.11      | 4.83       | 19.15      |
|            | cgccgcttcTATAGAGgggcat     | -24T>G | Neurological and muscular disorders [32] | 19.08 | -1.03 | 15.71 -3.41 |
| FIX        | acagctcagcTTGACTTGTgtcaca  | WT     | Norm [39]                         | 18.24      | 510±60     | 14.49      |
|            | acagctcagcTTGACTTGTgtcaca  | -26G>C | hemophilia B Leyden [39]          | 17.75      | -0.49      | 150±100 145.1  |
| Il1b       | ttgtaagcgCATAAAACagcggg    | WT     | Norm [29]                         | 18.67      | 29±4       | 17.36      |
|            | ttgtaagcgCATAAAACagcggg    | -31C>T | lung cancer [29]                  | 19.85      | +1.18      | 17.88 +1.42 |
| Tf         | ggcgggccccTTTATGgctgggca   | WT     | Norm [41]                         | 18.91      | 72±9       | 16.45      |
|            | ggcgggccccTTTATGgctgggca   | -21C>T | thrombophlebitis and myocardial infarction [41] | 19.43 | +0.52 | 17.47 +1.02 |
| Nos2a      | atggggtgagTAAATAAAttcttg    | WT     | Norm [38]                         | 19.85      | 1.8±0.2    | 20.14      |
|            | atggggtgagTAAATAAAttcttg    | -21T>C | resistance to infections [38]     | 20.06      | +0.21      | 1.6±0.3 20.25 +0.11 |

$K_D$ (standard deviation); $\delta$ - the difference between the affinity of hTBP for ODNs with and without SNPs in their TATA boxes expressed as natural logarithms, $\delta = -\ln(K_D^{TATA}) - (-\ln(K_D^{TATA}))$.  

doi:10.1371/journal.pone.0054626.t001

SNPs in the TATA Boxes of the CYP2A6 (Cytochrome P450) and IL1B (Interleukin-1 Beta) Genes

The products of these genes, nicotine oxidase and proinflammatory cytokine, play a role in carcinogen activation, drug detoxification and the formation of inflammatory cell responses.

In some patients with lung cancer, the -48T>G polymorphism was detected in the TATA box of the CYP2A6 gene promoter [28], and in some, -31C>T in the TATA box of the IL1B gene promoter [29]. The -48T>G substitution destroys the TATA box in the gene encoding nicotine oxidase (CYP2A6) [28] and is associated with increased risk of inflammatory diseases and lung cancer in smokers: the hTBP/TATA affinity is decreased 4.7-fold. When the -31C>T polymorphism in the TATA box of the IL1B gene was the case, we demonstrated that the hTBP/TATA affinity was increased more than fourfold: $K_D$ = 29 nM in healthy people and $K_D$ = 7 nM in affected people. This increase is because the gene [27], the conclusion was made that 15–30% hTBP/TATA binding is enough for transcription at normal levels. Our quantitative estimates of TBP/TATA binding (see the table 1, $\beta$ globins) and the mRNA amounts provided in some works [12–18] ($K_D$ = 390 nM, 25% of mRNAs; $K_D$ = 300 nM, no mRNAs; $K_D$ = 87 nM, 10% of mRNAs; $K_D$ = 64 nM, 20% of mRNAs, $K_D$ = 116 nM, 8–13% of mRNAs; $K_D$ = 99 nM, 50% of mRNAs) fail to make us reach the same conclusion and suggest that regulation of $\beta$-globin gene transcription in vivo and in vitro occurs on a more integrated and individual basis. We have determined the affinity of hTBP for the TATA boxes in the $\delta$-globin gene and the $\alpha$-globin pseudogene with SNPs associated with $\delta$- and $\alpha$-thalassemia intermedi. The affinity of hTBP to the TATA box is decreased 2.6-fold in $\delta$-thalassemia and is increased 1.3-fold in $\alpha$-thalassemia.
SNPs in the TATA Boxes of the SOD1 (Superoxide Dismutase 1) and TPI (Triosephosphate Isomerase) Genes

TATA-box polymorphisms in these genes are associated with diseases that cause neurological and muscular disorders. Two patients with amyotrophic lateral sclerosis were reported [31] to have the −27A>G polymorphism in the TATA box (TGTTAA instead of TATAAA) of the SOD1 gene encoding the enzyme superoxide dismutase 1. We have demonstrated a 4.2-fold decrease in hTBP/TATA affinity: $K_D = 40 \text{ nM}$ in healthy people and $K_D = 170 \text{ nM}$ in affected people. The TPI gene is a housekeeping gene. Triosephosphate isomerase, which is the enzyme that this gene encodes, is involved in glycolysis and occurs in every organism. If an SNP causes its deficiencies, neuromuscular disorders and hemolytic anemia are expected [32]. Additionally, it has recently been demonstrated [33] that triosephosphate isomerase in stomach cancer can convert drug-resistant cells into drug-sensitive cells, which renders chemotherapy more effective and makes this enzyme appear as a candidate target for new drugs against stomach cancer. It has been demonstrated that mutations causing deficiencies in TPI are associated with chronic hematolytic anemia, degenerative neurological disorders, cardiomyopathy, infant mortality and more [32]. With the −24T>G SNP in the TATA box of this gene, a very dramatic (more than 30-fold) decrease in TBP/TATA affinity was demonstrated; that decrease being correlated with a low gene expression level [32]. As can be seen from the case with the SNP-containing TATA box in the TPI gene promoter, even a very strong (41-fold) change in TBP/TATA affinity is associated with a decrease in gene expression and enzyme activity in erythrocytes: the TPI activity is decreased by 80–98% in some patients and by 26–50% in vivo in heterozygous individuals [34].

SNPs in the TATA Boxes of the MBL2 (Mannose-binding Lectin) and NOS2A (NO Synthase) Genes

The products of these genes, mannose-binding lectin and NO synthase, are involved in many responses produced by the organism, including the immune response. As is known, genetically determined variation in MBL concentrations in human blood serum accounts for varying sensitivity to infections and predisposition to autoimmune, metabolic and cardiovascular diseases [35]. Low MBL levels are associated with increased risk of recurrent infections [36]. We have demonstrated a twofold decrease in TBP/TATA affinity when the TATA box of this gene contains the −35T>C SNP [37]. A slight increase in TBP/TATA affinity caused by the −21T>C polymorphism [38] located near the TATA box in the promoter of the NOS2A gene encoding NO synthase is associated with increased resistance to diseases such as malaria, acute respiratory and lung diseases.

SNPs in the TATA Boxes of the FIX and TF Genes

Polymorphisms in the coagulation factor IX and tissue factor gene promoters are associated with vascular diseases. The promoter of the wild-type clotting factor IX gene contains a region for binding to the hepatocyte nuclear factor HNF4 located in the TATA box. HNF4 is the main factor controlling coagulation factor IX expression in healthy individuals and when it is unable to effectively bind to an altered site, the individual will develop hemophilia B Leyden.

It has been found [39] that the −26G>C mutation reduces HNF4 binding to the background level. TBP binds to this site with a very low specificity: $K_D = 510 \text{ nM}$ without this polymorphism and 500 nM with this polymorphism.

TF is a transmembrane protein expressed in many tissues, including the outermost layer of the vessel walls, where it rapidly activates coagulation whenever integrity is compromised [40]. A 2.7-fold increase in affinity with −21C>T SNP in the TATA element of the tissue factor gene factor is consistent with the known enhancement in gene expression and increased risk of thrombophlebitis and myocardial infarction [41].

**Statistical Data Analysis**

A comparison of the experimentally obtained and previously predicted values of changes to TBP/TATA affinity when the TATA boxes contained mutations was performed using the standard software program package STATISTICA [42] and demonstrated that these values were well correlated to each other. Fig. 1 presents the 95% confidence intervals for the linear regressions built. As can be seen, some of the absolute values as predicted for TBP/TATA affinity, $K_D$, and mutational changes, $\delta$, do not fall within the respective confidence intervals. We found no significant correlation between excursions outside the 95% confidence intervals and the controlled parameters that we used for in silico predictions and experimental measurements made in vitro. This implies that some binding parameters remain to be included in the equation for TBP/TATA binding in three successive steps that we are verifying. The observed in silico underprediction of TBP affinity and mutational changes in TATA boxes suggests that damage events in DNA sequence context can produce a cooperative effect at the TBP/TATA binding site. Indeed, the universal molecular processes that have an influence on TBP/TATA binding include the specific packaging of the core promoter into nucleosomes between positions −70 and +30 relative the transcription start site [43], with position −43 being the one of the nucleosome center [44]. The binding site of the nucleosome center on the DNA is AT-rich and matches with the optimal localization of TATA boxes [45]. This suggests that eukaryotic promoters are likely to possess the composite element $\backslash (H3-H4)/(H2A-H2B)(H3-H4)$, which was detected experimentally [44,45], but has not yet been considered for use in tools developed for in silico analysis.

**Conclusions**

As is known, completion of the Human Genome Project has resulted in a wealth of new data and posed new challenges. For example, a large number of SNPs with unknown phenotypic manifestations have been found. Consequently, identification and analysis of regulatory SNPs in human genes will help quantify the effects of these SNPs on human health and sensitivity to drugs and environmental factors.

Transcriptional regulation of gene expression is performed by a large number of proteins and protein complexes, which interact with DNA and one another and cooperatively stimulate or inhibit gene expression in response to internal and external signals. Transcription factors are the key players in this process. The interaction between TBP and the TATA box on TATA-containing promoters is one of the rate-limiting stages of transcription initiation. Relationships between the quantitative characteristics of TBP/promoter interactions and gene transcription levels are very important for understanding the mechanisms of activation and inhibition of gene transcription and expression;
however, these data, especially on man, are desperately needed.
The existing literature data on the interaction between TBP and
the TATA box have been obtained using different TPBs: from S.
cerevisiae, D. melanogaster, mice, man, full-length TBP molecules
and its DNA-binding domain. The model systems used, too, were
different (yeast, flies, laboratory animals, human cell lines, to name
a few), and so were the experimental conditions, which prevents
comparisons of the results so obtained or making inferences about
the effect that a particular substitution in the TATA box has on
TBP/TATA affinity. In the current work, we have performed,
under the most standardized experimental conditions, a study of
interactions between recombinant human TBP (a full-length
molecule with the amino acid composition as in the natural
human TBP molecule) and 28 ODN identical to the TATA boxes
in gene promoters in healthy people and patients whose diseases
are associated with SNPs in TATA boxes. As can be seen from the
results provided herein, the wild-type natural promoters of four
genes studied (the \( \beta \)-globin, \( \delta \)-globin, \( \psi \)-globin and IL1B genes)
contain TATA boxes instead of TATA boxes, three genes (\( \text{NOS2A}, \text{SOD1} \)
and \( \text{CYP2A6} \)) contain classical TATA-box sequences, and
two genes (\( \text{TPI} \) and \( \text{MBL} \)) contain TATATA boxes. The highest
affinity \( K_D = 1.8 \text{ nM} \) and \( K_D = 4.8 \text{ nM} \) was observed for the
TBP/TATA interaction in the \( \text{NOS2A} \) and \( \text{TPI} \) genes, in which
the TATA-box sequences are classical, TATATAA and TA-
TATA, respectively. The lowest affinity \( K_D = 72 \text{ nM} \) was observed
for the \( \text{TF} \) gene, in which the TATA-box sequence has a
reduced sequence (TTTATA) in healthy people. It has now been
experimentally demonstrated for the first time that TBP/TATA
affinity in monogenic and polygenic pathologies is not beyond the
affinity range between \( J \) and \( K \) (as predicted \( \text{in silico} \) and measured \( \text{in vitro} \). The coefficient of
linear correlation, \( r \), is 0.785 at a significance level of \( a < 10^{-3} \). 95% confidence intervals for linear regression built using the package
\text{STATISTICA}. doi:10.1371/journal.pone.0054626.g001

![Figure 1. The experimentally measured affinity and affinity change are highly correlated with the predicted values.](Image)

The correlation between the TBP/TATA affinity as predicted \( \text{in silico} \) and measured \( \text{in vitro} \). The coefficient of linear correlation, \( r \), is 0.822 at a significance level of \( a < 10^{-7} \). (B) The correlation between SNP-caused TBP/TATA affinity change (\( \delta \)) as predicted \( \text{in silico} \) and measured \( \text{in vitro} \). The coefficient of linear correlation, \( r \), is 0.785 at a significance level of \( a < 10^{-5} \).

Materials and Methods

Protein Expression and Purification

Recombinant full-length human TBP containing only the native
amino acid sequence was overexpressed in \( \text{E. coli} \) BL21(DE3) cells
transformed with \( \text{pAR3038-hTBP} \) (plasmid \( \text{pAR3038-hTBP} \) was
the kind gift of Prof. B. Pugh, Pennsylvania State University).
Expression of hTBP was as described by Pugh [47] with
modifications (the IPTG concentration were 1 mM instead of
0.1 mM and the induction time was 3 h instead of 1.5 h); TBP was
purified to homogeneity using three-step procedure involving
polyethylenimine precipitation, phosphocellulose chromatogra-
phy, and ammonium sulfate precipitation as described by Pugh
[47]. Based on Coomassie Brilliant Blue R 250 stained SDS-
PAGE analyses coupled with scanning densitometry, the purity
of these TBP preparation was determined to be more than 98%.
The total protein concentration was determined by Bradford
[48]. The concentration of active TBP was determined by titrations of TBP
against known concentrations of TATA AdML, which were well
above the \( K_D \), and was about 50% from total protein
concentration.

Labeling Oligodeoxyribonucleotides with \( ^{32} \text{P} \)

Twenty-six base pair oligodeoxyribonucleotides (ODNs)
synthesized and additionally purified by electrophoresis in PAGE
(Biosset, Novosibirsk) were used. The ODN sequences used were
identical to the TATA boxes (with and without SNPs) with the
flanking sequences of the \( \beta \)-globin gene, \( \delta \)-globin gene, \( \psi \)-globin pseudogene, and the genes for tissue factor, triosephosphate
isomerase, NO synthase 2A, superoxide dismutase 1, cytochrome
P450 2A6, mannose-binding lectin, clotting factor IX and human
interferlin-1 beta cited from works referenced in the table 1.
ODNs quality was tested for with the use of MALDI TOF MS
(Bruker Daltonics). Labeled double-stranded ODNs were obtained
by labeling both strands with \( ^{32} \text{P}-\text{ATP} \) (Biosan, Novosibirsk) using
T4 polynucleotide kinase (SibEnzime, Novosibirsk), annealing at
95° C (at equinomolar concentration) and slowly (for not less than
3 h) cooling to room temperature. Duplexes were analyzed in 15% non-denaturing PAGE (1x TBE) [49], isolated and purified by electroelution.

Measurements of the Equilibrium Dissociation Constants for hTBP/TATA Complexes

The equilibrium dissociation constants (K_D) for the complexes of hTBP with TATA-containing double-stranded ODNs identical to wild-type and SNP-containing TATA-box variants were measured using a traditional approach, which included titration of a fixed amount of active TBP (typically 0.3 nM) with the increasing concentrations of TATA-containing ODN to reach equilibrium. The time to reach equilibrium was determined previously for each ODN. Each K_D value was determined following not less than 8 experimental runs.

Experiments on hTBP/ODN binding were run at 25°C in a buffer (20 mM HEPES-KOH (pH 7.6), 5 mM MgCl2, 70 mM KCl, 1 mM DTT, 100 µg/ml BSA, 0.01% NP-40, 5% glycerol) until equilibrated. The hTBP-ODN complexes were separated from the unbound ODN using a gel retardation assay (EMSA). Electrophoresis was performed using 5% PAGE in Tris-glycine buffer (pH 8.3) for 40 min at a temperature of 10°C and a field intensity of 25 V/cm. The gels were dried and exposed to an Imaging Screen-K (Kodak) for use with a Molecular Imager PhorosFX Plus phosphorimager (Bio-Rad). The screen was scanned by the phosphorimager and the radioautographs were quantitated using Quantity One 4.5.0 software (Bio-Rad). The equilibrium K_D values for the hTBP-ODN complexes, which characterize the affinity of TBP for TATA boxes, were determined using OriginPro 8 (for an example, see Fig. 2).

Prediction of hTBP/TATA-box Affinity

The affinity of hTBP to ODNs with and without SNP in their TATA boxes expressed as natural logarithms was calculated using the equation (equation obtained earlier [8] by determining K_D complexes TBP/TATA to eight independent human genes, one gene of rat and mouse one gene [7]) for TBP/TATA equilibrium binding in three successive steps: sliding [9] ↔ recognition [10] ↔ stabilization [11]:

\[- \ln[K_D,TATA](S^0) = 10.90 + 0.15 \times \text{PWM}_{TATA}(S^0) - 0.23 \times \ln[K_D,dsDNA](S^0) - 0.20 \times \ln[K_D,ssDNA](S^0)\]

where 10.90 is the non-specific TBP/DNA affinity [4]; PWM_{TATA} is the highest score of Bucher’s TATA-box weight matrix [10] from among all its 22 possible positions on the strand S^0 and its complementary strand, each 26 bp long; ln[K_D,dsDNA] is the mean of the regression of the affinity of TBP for the double-stranded DNA on the strand S^0 that has the highest score of Bucher’s TATA-box weight matrix (the sliding stage) [4,9,10]; ln[K_D,ssDNA] is the half-sum of the regressions of TBP affinity for each strand on the sequences of each of these strands with the highest scores of Bucher’s TATA-box weight matrix (the stabilization stage) [10,50,51]; 0.15, 0.23, 0.20 are the stoichiometric coefficients for three steps of TBP/TATA binding as calculated previously [8]. The difference (δ) between the affinity of hTBP for ODNs with and without SNPs in their TATA boxes expressed as natural logarithms:

\[\delta = - \ln[K_D,TATA_{Mut}] - (- \ln[K_D,TATA])\]

The coefficients of correlation (r) and their levels of significance (α) were calculated using the standard software program package STATISTICA [42].

Figure 2. Experimental measurements of the equilibrium K_D for hTBP/TATA complexes. The use of an electrophoretic mobility shift assay for measuring hTBP/TATA affinity: (A) a TATA-containing ODN in healthy people (the –21C allele); (B) an ODN with the TATA box containing the SNP associated with myocardial infarction and thrombophlebitis (the –21t allele); (C) TBP/TATA binding isotherms inferred from electrophoregrams using OriginPro 8.

doi:10.1371/journal.pone.0054626.g002
References

1. Bailey J (2009) Mining for SNPs and SSRs using SNPServer, dbSNP and SSR taxonomy tree. Methods Mol Biol 537: 303–321.
2. Baumann M, Pontiller J, Ernst W (2010) Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an update. Mol Biotechnol 45: 241–247.
3. Juve-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339: 225–229.
4. Hahn S, Buratowski S, Sharp F, Guarente L (1989) yeast TATA-binding protein TFID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc Natl Acad Sci USA 86: 5718–5722.
5. Wolner BS, Graaff JD (2001) TATA-flanking sequences influence the rate and stability of TRBP and TFIIB. J Biol Chem 276: 6260–6266.
6. Savinova LK, Ponomarenko MP, Ponomarenko PM, Drachkova IA, Lysova MV, et al. (2009) TATA box polymorphisms in human gene promoters and associated hereditary pathologies. Biochemistry 74: 117–129.
7. Savinova LK, Drachkova IA, Ponomarenko MP, Lysova MV, Arshinova TV et al. (2007) Interaction between the recombinant TATA-binding protein and the TATA-boxes of the mammalian gene promoters. Ecol Genetic 5: 44–49.
8. Ponomarenko PN, Savinova LK, Drachkova IA, Lysova MV, Ponomarenko MP, et al. (2008) A step-by-step model of TRBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. Doklady Biochem and Biophys 419: 88–92.
9. Coleman RA, Pugh BF (1995) Evidence for functional binding and stable shifting of the TATA protein on non-specific DNA. J Biol Chem 270: 13630–13639.
10. Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 302 unrelated promoter sequences. J Molecular Biology 212: 363–578.
11. Joo ZS, Chun TK, Leiberman PM, Baikalov I, Berk AJ, et al. (1996) How proteins recognize the TATA box. J Mol Biol 261: 239–254.
12. Antonarakis SE, Irkin SH, Cheng TC, Scott AF, Sexton JP, et al. (1984) Beta-thalassemia due to a T–A mutation in the ATA box. Biochem Biophys Res Commun 112–114.
13. Badens C, Jassim N, Martini N, Mattei JF, Elion J, et al. (1999) Characterization of a series of alleles in humans with reduced activity at the triosophosphate isomerase locus. Am J Hum Genet 56(2): 300–316.
14. Wang X, Lu Y, Yang J, Shi Y, Lan M, et al. (2008) Identification of triosophosphate isomerase as an anti-drug resistance agent in human gastric cancer cells using functional proteomic analysis. J Cancer Res Clin Oncol 134(9): 995–1003.
15. Humphries A, Atianon A, Lalloz MR, Layton DM (1999) Ancestral origin of variation in the triosophosphate isomerase gene promoter. Hum Genet 104(6): 486–91.
16. Eisen DP, Mischel TM (2005) Impact of mannose-binding lectin on susceptibility to infectious diseases. Clin Infect Dis 37(11): 1496–505.
17. Takahashi K, Ezenwokwu RA (2005) The role of the mannose-binding lectin in innate immunity. Clin Infect Dis 7: S40–4.
18. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
19. Ioshikhes I, Trifonov EN, Zhang MQ. (1999) Periodical distribution of transcription factor sites in promoter regions and connection with chromatin structure. J. Molecular Biology 293: 549–559.
20. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ et al. (2006). A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. Doklady Biochem and Biophys 419: 88–92.
21. Stewart JJ, Fischback JA, Chen X, Stargell LA (2006) Non optimal TATA box regulatory elements. J. Molecular Biology 212: 563–578.
22. Antoniou M, de Boer E, Spanopoulos E, Inam A, Grosfeld F (1995) TBP binding and the rate of transcription initiation from the human β-globin gene. Nucl Acid Res 23(17): 3473–3480.
23. Polkonec O, Rauhio A, Rauhio U, Paunen M (2000) CYP2A6: a human coumarin 7-hydroxylase. J Biol Chem 275: 139–147.
24. Zieroldlindin S, Ryberg D, Maggini V, Skauh V, Czanian F, et al. (2004) Polyorphism of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. Int J Cancer 109: 353–356.
25. Wang Y, Kato N, Hoshida H, Taniguchi H, Goto T, et al. (2003) Interleukin-1 beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection. Hepatol 37: 66–71.
26. Niemann S, Broom WJ, Brown RH (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.
27. Watanabe M, Zingg BC, Müllermüller HW (1996) Molecular analysis of a series of alleles in humans with reduced activity at the triosophosphate isomerase locus. Am J Hum Genet 56(2): 300–316.
28. Wang X, Lu Y, Yang J, Shi Y, Lan M, et al. Identification of triosophosphate isomerase as an anti-drug resistance agent in human gastric cancer cells using functional proteomic analysis. J Cancer Res Clin Oncol 134(9): 995–1003.
29. Humphries A, Atianon A, Lalloz MR, Layton DM (1999) Ancestral origin of variation in the triosophosphate isomerase gene promoter. Hum Genet 104(6): 486–91.
30. Eisen DP, Mischel TM (2005) Impact of mannose-binding lectin on susceptibility to infectious diseases. Clin Infect Dis 37(11): 1496–505.
31. Takahashi K, Ezenwokwu RA (2005) The role of the mannose-binding lectin in innate immunity. Clin Infect Dis 7: S40–4.
32. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
33. Ioshikhes I, Trifonov EN, Zhang MQ. (1999) Periodical distribution of transcription factor sites in promoter regions and connection with chromatin structure. J. Molecular Biology 293: 549–559.
34. Humphries A, Atianon A, Lalloz MR, Layton DM (1999) Ancestral origin of variation in the triosophosphate isomerase gene promoter. Hum Genet 104(6): 486–91.
35. Eisen DP, Mischel TM (2005) Impact of mannose-binding lectin on susceptibility to infectious diseases. Clin Infect Dis 37(11): 1496–505.
36. Takahashi K, Ezenwokwu RA (2005) The role of the mannose-binding lectin in innate immunity. Clin Infect Dis 7: S40–4.
37. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
38. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
39. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
40. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
41. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
42. Afifi A, Clark V, May S (2003) Computer-Aided Multivariate Analysis. New York: Cambridge University Press.
43. Ioshikhes I, Trifonov EN, Zhang MQ. (1999) Periodical distribution of transcription factor sites in promoter regions and connection with chromatin structure. J. Molecular Biology 293: 549–559.
44. Godde JS, Nakatani Y, Wolffe AP (1995) The amino-terminal tails of the core histones and the translational position of the TATA Box determine TBP/TFIIA association with macronuclei DNA. Nucleic Acid Research 23: 4537–4546.
45. Kel AE, Kel-Margoulis OV, Babenko V, Wingender E (1999) Recognition of regulatory elements by transcription factor sites. In: Computational molecular biology: an integrated approach. New York: Cambridge University Press.
46. Orosz F, Ola´h J, Ova´di J (2006) Triosephosphate isomerase deficiency: facts and hypotheses. J Bioenergot Biomembr 38(3): 337–347.
47. Godde JS, Nakatani Y, Wolffe AP (1995) The amino-terminal tails of the core histones and the translational position of the TATA Box determine TBP/TFIIA association with macronuclei DNA. Nucleic Acid Research 23: 4537–4546.
48. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–253.
49. Godde JS, Nakatani Y, Wolffe AP (1995) The amino-terminal tails of the core histones and the translational position of the TATA Box determine TBP/TFIIA association with macronuclei DNA. Nucleic Acid Research 23: 4537–4546.