Point mutations within 663–666 bp of intron 6 of the human *TDO2* gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site

Gennady V. Vasiliev, Vasily M. Merkulov, Viktor F. Kobzev, Tatyana I. Merkulova*, Mikhail P. Ponomarenko, Nikolay A. Kolchanov

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

Received 18 October 1999

Abstract Single base mutations G→A at position 663 and G→T at position 666 of intron 6 of the human tryptophan oxygenase gene (*TDO2*) are associated with a variety of psychiatric disorders [Comings, D.E. et al. (1996) Pharmacogenetics 6, 307–318]. Binding of rat liver nuclear extract proteins to synthetic double-strand oligonucleotides corresponding to three allelic states of the region between 651 bp and 680 bp of human *TDO2* intron 6 has been studied by gel shift assay. It has been demonstrated that to each allelic state of the region there corresponds a specific set of proteins that interacts with it. With the aid of computer analysis and using specific anti-YY-1 antibodies it has been shown that both mutations damage the YY-1 transcription factor binding site.

Key words: Tryptophan oxygenase gene; Binding site; YY-1 transcription factor; Mutation; Psychiatric disorder

1. Introduction

Tryptophan 2,3-dioxygenase (TDO2, EC 1.13.11.11) is the rate-limiting enzyme in the oxidative degradation of tryptophan [1], the serotonin precursor, which therefore controls serotonin level in the body. Defects in serotonin metabolism and abnormal serotonin/tryptophan levels have been reported for many behavioural disorders (Tourette syndrome, attention deficit hyperactivity disorder, alcoholism, depression, and others) [2–5]. This suggests the *TDO2* gene as a potential candidate gene in psychiatric genetics. Functional variants of this gene could account for the observed variations in both serotonin and tryptophan levels and contribute to the development of psychiatric disorders [4,5].

Molecular cloning and sequencing of over 9000 bp of human *TDO2* DNA, which included all its exons and introns 1, 5 and 6, allowed Comings et al. to identify four polymorphisms of this gene and to develop genetic tests for them [6,7]. Among them, G→A and G→T mutations 2 bp apart in the middle of intron 6 showed a significant positive association with drug dependence, Tourette syndrome and attention deficit hyperactivity disorder [7]. Single nucleotide substitutions reported in introns are often associated with various diseases [8–12]. However, the mechanisms by which such mutations could affect gene expression normally remain unclear, unless the mutations affect splicing signals. We have demonstrated that the region between 651 and 680 bp from the start of the intron 6 of the human *TDO2* gene, in which polymorphic sites associated with behavioural disorders are located, binds liver nuclear proteins, and that every allelic state the region can be in is accounted for a specific set of proteins that interact with it. The disturbance in the binding of the transcription factor YY-1 is one of the effects of the mutations.

2. Materials and methods

2.1. Oligonucleotides

Oligonucleotides corresponding to both strands of the fragment from 651 to 680 bp of intron 6 of *hTDO2* (WT) and its variants (M1 and M2) (Fig. 1A) were kindly provided by D.E. Comings (Department of Medical Genetics, Duarte, CA, USA). After annealing the oligonucleotides were labelled with Klenow enzyme and (α-^{32}P)dATP and purified by gel-electrophoresis in 10% non-denaturing PAAG.

The double-strand competing oligonucleotides were: HNF3, 5′-cagtCGAGTTGACTAAGTCAATAATCAGAATCAGTCG-3′ [13]; GATA, 5′-cagtGATCTCCGGCAACTGATAAGGATTCCCTG-3′ [14]; HNF1, 5′-cagTGGTTAGTGTGGTTAATGATCTACAGTT-3′ [15]; C/EBP, 5′-cagtGATCCATATTAAGGACATGCCG-3′ [16]; Sp1, 5′-cagtCGACTCTAGGCGGGGTTAAAGTTCT-3′ [17]; AP1, 5′-cagtAGCTTGATGAGTCAGCCGGATC-3′ [18]; SRE, 5′-cagtACAGGATGTCCATATTAGGACATCTGCGT-3′ [19]; and ANF, 5′-cagtGCTGGTGAGATTGTGCCACAGCTCTGCA-3′ [20]. The oligonucleotides were synthesised by the H-phosphonate method and purified by gel-electrophoresis [21].

2.2. Nuclear extract preparation and gel shift assay

Nuclear extracts from the rat liver were prepared according to Gorski et al. [22] with the modifications described in Shapiro et al. [23]. The binding reactions with DNA were carried out in 10 μl reaction volume containing 1–4 μg of nuclear extracts preincubated with 0.1–0.4 μg calf thymus DNA at 4°C for 10 min, 2 ng (^{32}P)-labelled (200–500 cps) DNA probe in a binding buffer (25 mM HEPES (pH = 7.6), 80 mM KCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol). For the competition analysis, the reaction mixture additionally contained 80 ng of the corresponding unlabelled oligonucleotide. For antibody-containing reactions, the extracts were preincubated with 1 μl of anti-YY-1 antibody (Santa-Cruz Biotechnology, Santa-Cruz, CA, USA) at 4°C for 10 min. After incubation for 10 min at room temperature the samples were analysed by electrophoretic fractionation on a 5% non-denaturing polyacrylamide gel in 0.5×TBE buffer at 180 V at 4°C. After electrophoresis gels were dried, protein DNA complexes were visualised by autoradiography.

2.3. Recognition of transcription factor binding sites

A method based on the averaging of oligonucleotide frequencies at each position of the functional sites of DNA was used for recognition of these sites [24]. The initial data for the method to start working are a set of N aligned DNA sequences of length L of the site in question $\{S_n = (s_{n,i})\}_{i=1;L;\, n=1;N}$. Oligonucleotide frequencies at each position of the site are calculated using the following alphabets: $E_0 = \{e_{01} = \text{A},$

*Corresponding author. Fax: (7) (3832) 356558.
E-mail: merkti@niboch.nsc.ru

PII: S0014-5793(99)01513-6

e_{02} = T, e_{03} = G, e_{04} = C}, E_1 = {e_{11} = W = A+T, e_{12} = S = G+C}, E_2 = {e_{21} = R = A+G, e_{22} = Y = T+C}, E_3 = {e_{31} = M = A+C; e_{32} = K = T+G}, E_4 = {e_{41} = WW, e_{42} = WS, e_{43} = SW, e_{44} = SS}, E_5 = {e_{51} = YY, e_{52} = YR, e_{53} = RY, e_{54} = RR} and E_6 = {e_{61} = MM, e_{62} = MK, e_{63} = KM, e_{64} = KK}:

$$F(e_{pj,i}) = \{\sum_{n=1;N} \ddot{I}_{q=1;Q}\delta(s_{n,i+q}\in e_{pj})\}/N; \quad (1)$$

here: p is the alphabet number, j is the number of the oligonucleotide in the alphabet; Q is the oligonucleotide length, 1 for E_0, E_1, E_2, E_3 or 2 for E_4, E_5, E_6; δ(TRUE) = 1 and δ(FALSE) = 0.

For each alphabet E_p, the partial similarity $D_p(S)$ between an arbitrary sequence S and the functional site in question is assessed by frequencies $F(e_{pj,i})$:

$$D_p(S) = [\{\sum_{i=1;L-Q+1} F(e_{pj,i})\times\ddot{I}_{q=1;Q}\delta(s_{n,i+q}\in e_{pj})\}-\alpha_p]/\beta_p; \quad (2)$$

here: a_p and b_p are normalising coefficients calculated in accordance with the rule: 'mean D_p is $\mathrm{Mean}_{\mathrm{SITE}}(D_p) = 1$ for known sites, and $\mathrm{Mean}_{\mathrm{RAND}}(D_p) = -1$ for random DNA'. The integral estimate of this similarity is the mean value of these partial estimates:

$$D(S) = \sum_{p=0;6} D_p(S)/7, \quad (3)$$

the recognition rule for which reads: IF {$D(S) > 0$} THEN {S is the site in question}.

3. Results

We used a mobility shift assay to see if nuclear factors bind to double-strand oligonucleotides that correspond to the different allelic states of the 651–680 bp region relative to start of intron 6 of the human *TDO2* gene (WT, M1 and M2 oligonucleotides, Fig. 1A). Since *TDO2* is a liver-specific gene a rat liver nuclear extract was used. The extract proteins formed three complexes with the WT oligonucleotide, which corresponds to the most frequent allele (Fig. 1B). Both nucleotide

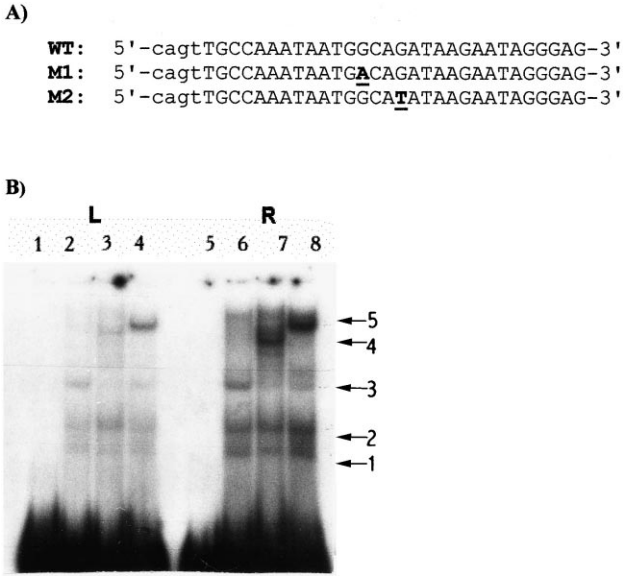

Fig. 1. Distinct nuclear proteins bind to double-strand oligonucleotides comprising the sequences of three alleles of 651–680 bp region of intron 6 of the human *TDO2* gene. (A) Coding strands of oligonucleotides used: WT most frequent allele, M1 and M2 mutant variants. Point mutations are underlined, lowercase letters correspond to nucleotides added to make 5′-overhangs. (B) Binding of nuclear proteins to WT (lanes 2, 6), M1 (lanes 3, 7), M2 (lanes 4, 8) oligonucleotides. Lanes 1, 5: no extract. Oligonucleotides were incubated with 1 (L) or 4 (R) μg of rat liver nuclear extract protein.

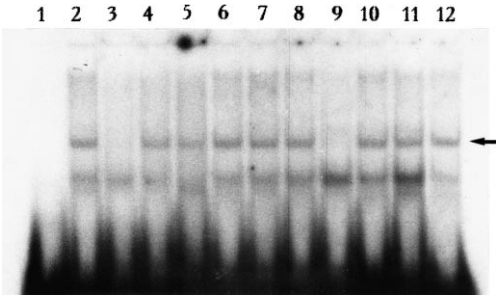

Fig. 2. Competition of WT oligonucleotide protein complex formation by various oligonucleotides. Lanes: 1, no extract; 2, no competitor; 3–12, a 40-fold excess of unlabelled oligonucleotides: 3, WT; 4, M1; 5, M2; 6, AP1; 7, HNF1; 8, GATA; 9, SRE; 10, ANF; 11, AP2; 12, HNF3. Arrow denotes the position of complex 3.

substitutions found in humans [10,11] caused a dramatic change in binding pattern. The substitution of G by A at position 663 bp (the M1 oligonucleotide) caused the band corresponding to complex 3 to abate, if not disappear, abruptly and a group of less mobile bands to occur ('complex 4'). In the case of the G→T substitution at position 666 bp (the M2 oligonucleotide) the intensity of the indicated band had been decreasing not so strongly, but there also appeared new slowly migrating bands ('complex 5').

To answer the question as to which proteins could bind to the WT, M1 and M2 oligonucleotides we did the search for regions homologous to the binding sites of various transcription factors using the TESS program package [25]. We found regions, similar to the binding sites for the GATA; HNF1 and Sp1 transcription factor families, and to the SRE element, to which several different factors can bind [26], and the ANF element, whose binding factor is unknown [20]. Double-strand oligonucleotides corresponding to such sites existing in real genes and previously used by other investigators for the same purposes, were synthesised and used as competitors in gel retardation experiments. The available oligonucleotides corresponding to the binding sites of the HNF3, C/EBP and AP1 families of transcription factors were taken as competitors.

As turns out, of the oligonucleotides used, only SRE (Fig. 2) of the mouse *c-fos* gene promoter (−318 to −289) was competing against WT: in the presence of its 40-fold excess complex 3 would disappear. As is known, transcription factors SRF, YY-1 and NF-IL6 bind to the mouse *c-fos* SRE [26] which suggests that these proteins might be candidates for binding to the region in question. Likely involvement of transcription factors YY-1 and SRF in the formation of complex 3 is also supported by the fact that the G→A transition (oligonucleotide M1), which abruptly decreases the intensity of band 3, occurs in a putative consensus (CCAATTT) of SRF and YY-1 binding sites [26–29].

A further theoretical analysis of the nucleotide sequences of the allelic states of the region in question was performed using a method developed for averaging nucleotide and dinucleotide frequencies at the positions of the site in question [24]. The original data were 27 DNA fragments experimentally found to contain the transcription factor YY-1 binding site [30]. The fragments were aligned using the standard method of multiple alignment maximising Gibbs potential [31]. With each position tried for the start of the putative YY-1 site (position '0'

of the site), the mean nucleotide and dinucleotide frequencies were calculated by Eqs. (2) and (3) in all positions of the variant being analysed. The profile of similarity between YY-1 sites and WT is presented in Fig. 3A, where dotted lines indicate the 95% boundaries of similarity with YY-1 sites (upper) and random DNA (lower). As can be seen, the peak of similarity between WT and YY-1 sites at position 661 is over the upper 95% boundary, which points (arrow) at region 661–667 bp of intron 6 of the human *TDO2* gene, which has a reliable similarity with the YY-1 sites.

As turned out, the G→A (663) and G→T (666) mutations are located within the likely YY-1 site. In the case of G→A (663) mutation (variant M1), the similarity of the YY-1 sites with region 661–667 bp is lower than the lower 95% boundary, which implies that there is no YY-1 site in. If the mutation is G→T (666), the similarity of the YY-1 sites with variant M2 is to over the upper 95% boundary (Fig. 3B). As can be seen, quantitatively the pattern of similarity between the YY-1 and WT (strong), M1 (none) and M2 (weak) behaves as band 3 does while the liver extracts bind to these three variants (Fig. 1).

As to the other two transcription factors, SRF and NF-IL6, which bind to the SRE element of the *c-fos* gene, [25], no binding site has been predicted for either WT, M1 or M2 of 661–668 bp of intron 6 of the human *TDO2* gene by a similar analysis (Fig. 3). This implies that among YY-1, SRF and NF-IL6, YY-1 is the most likely transcription factor to bind to region 651–680 bp of intron 6 of the human *TDO2* gene.

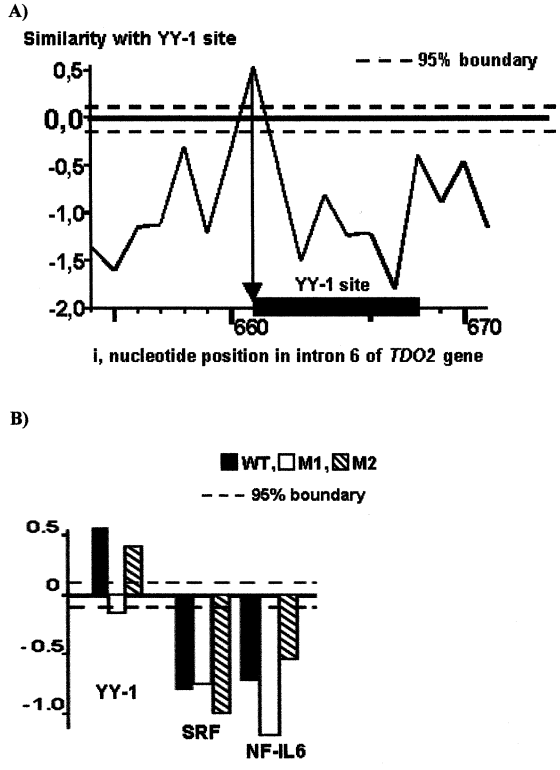

Fig. 3. Recognition of the YY-1 transcription factor binding site in region 651–680 bp of intron 6 of human *TDO2*. (A) The profile of similarity between YY-1 sites and WT as calculated by Eqs. (2) and (3). Dotted lines indicate the 95% boundaries of similarity with YY-1 sites (upper) and random DNA (lower). The peak of similarity at position 661, which is over the 95% boundary, points out (arrow) the location of the potential YY-1 site in region 661–668 bp (black rectangle). (B) Similarity of WT, M1 and M2 with YY-1, SRF and NF-IL6 transcription factor binding sites.

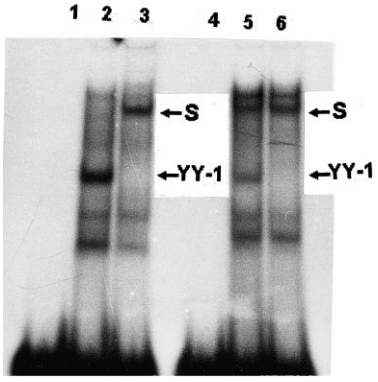

Fig. 4. Effect of anti-YY-1 antibodies on mobility shifts of WT and MII oligonucleotides. Lanes 1, 4: no extract. Lanes 2, 5: 3 μg of nuclear extract, no antibodies. Lanes 3, 6: 3 μg of nuclear extract was preincubated with 1 μl of specific anti-YY-1 antibodies. The positions of YY-1 containing complexes and supershift complexes (S) are indicated on the right.

To provide an independent control, variants WT, M1, and M2 were analysed using a program for recognising binding sites of ten random transcription factors T3R, c-Fos, OCT, GR, EN, RAR, GATA, MEF-2, AP-1, PR. Similar to the results of recognition of the SRF and NF-IL6 sites, no matches were found for the similarity patterns of ten control sites with variants WT, M1 and M2, and WT (strong), M1 (none) and M2 (weak) band 3 in the experimental binding of the corresponding oligonucleotides with the liver nuclear extract (Fig. 1B) and the pattern of similarity of the YY-1 sites with these variants (Fig. 3B). Thus the control tests confirmed independently that YY-1 is the most likely transcription factor with the binding site located within 651–680 bp of human *TDO2* intron 6.

Specific anti-YY-1 antibodies used in the band shift assay gave total support to this hypothesis. As is shown in Fig. 4, addition of antibodies to the nuclear extract totally eliminates the protein WT oligonucleotide complex represented by band 3, which is replaced by a slower migrating band. Complex 3 also disappears in the case of oligonucleotide M2, however no supershift can be seen, because this band is as mobile as one of the bands of the initial pattern.

4. Discussion

Higher or lower activity of the TO enzyme in the liver could be the cause of abnormal serotonin and tryptophan levels observed on many behavioural disorders. With *TDO2* as a potential major candidate gene in psychiatric genetics, Comings et al. screened all the exons of this gene and its 5′- and 3′-regulatory regions for polymorphisms, but this search had been essentially negative [6].

Over the recent years a great bulk of evidence has been accumulated that regulatory areas of genes may be located not only in their 5′- and 3′-regions but also within the body of genes [32,33]. In rat *TDO2* gene we have found the region of specific binding of glucocorticoid receptor extending from the end of intron 4 to the exon G. It was shown that the receptor binds to this region more efficiently than to the 5′-

flanking region [34] of this gene containing the known GREs [35].

And it was the corresponding region of the human *TDO2* gene that polymorphic sites, associated with such behavioural disorders as Tourette syndrome, attention deficit hyperactivity disorder and drug dependence, were found in [10]. These are two nucleotide substitutions located much central in intron 6 (G→A at position 663 bp and G→T at position 666 bp relative to start of the intron). In this work we have demonstrated that the region where the mutations are located (between 651 bp and 680 bp) is the binding site for liver nuclear extract proteins, at least part of which are transcription factors. It is conceived to be of special importance that these mutations are responsible for a dramatic change in the set of the proteins bound to the region. Among the changes, we have identified the disturbance of transcription factor YY-1 binding, which is due to the destruction of its binding site.

YY-1 is a polyfunctional protein, which was shown to either repress [33] or stimulate [34] gene expression depending on the context of the corresponding regulatory region. As a component of the nuclear matrix YY-1 may also be involved in chromatin organisation possibly by tethering DNA to nuclear matrix [35]. Whichever it is, it influences transcription intensity. Thus, both mutations damaging the YY-1 transcription factor binding site may result in a change in *TDO2* transcription level, which accounts for phenotypic changes.

All together, these facts suggest that the region encompassing introns 4–6 might be of regulatory importance for the expression of the *TDO2* gene; and the region in the middle of intron 6 of the human *TDO2* gene, where associations with psychiatric disorders were found, appears to be an important component of this regulatory region. If this is as it seems to be, then the observed alteration in the set of the protein factors that bind to this region may have important consequences for regulation of the expression of the *TDO2* gene and may also unambiguously indicate by which mechanisms the functional variants of the gene arise following single nucleotide substitutions within its intron.

Acknowledgements: The authors are thankful to O.A. Podkolodnaya and Yu.V. Ponomarenko for the set of the sequences of transcription factor YY-1 binding sites and Dr. D.E. Comings for fruitful discussion. The work was supported by the Russian Foundation for Basic Research (98-04-49654, 98-07-90126, 98-07-91078).

References

[1] Hayaishi, O. (1976) J. Biochem. 79, 13–21.
[2] Brown, S.-L. and van Praag, H.M. (1990) The Role of Serotonin in Psychiatric Disorders, Bruner/Mazel, New York.
[3] Virkkunen, M. and Linnoila, M. (1990) Ann. Med. 22, 327–331.
[4] Comings, D.E. (1990) Am. J. Med. Genet. 36, 418–430.
[5] Comings, D.E. (1993) Alcohol Alcohol 2 (Suppl.), 237–241.
[6] Comings, D.E., Muhleman, D., Dietz, G.W. and Donlon, T. (1995) Genomics 29, 390–396.
[7] Comings, D.E., Gade, R., Muchleman, D., Chiu, C., Shijuan, W., To, M., Spence, M., Dietz, G., Winn-Deen, E., Rosental, R. and Lesier, H. (1996) Pharmacogenetics 6, 307–318.
[8] Brockmoller, J., Cascorbi, I., Sachse, C. and Roots, I. (1998) Toxicol. Lett. 102, 173–183.
[9] Gelernter, J., Kranzler, H. and Lacobelle, J. (1998) Genomics 52, 289–297.
[10] Tsai, M.Y., Wong, P.W., Garg, U., Hanson, N.Q. and Schwichtenberg, K. (1997) Biochem. Mol. Med. 61, 9–15.
[11] Knoll, A., Schunkert, H., Reichwald, K., Danser, A.H., Bauer, D., Platzer, M., Stein, G. and Rosenthal, A. (1997) Hum. Mol. Genet. 6, 1527–1534.
[12] Wang, Y.C., Lin, M.L., Lin, S.J., Li, Y.C. and Li, S.Y. (1997) Hum. Mutat. 10, 393–399.
[13] Lai, E., Prezioso, V.R., Smith, E., Litvin, O., Costa, R.H. and Darnell, J.E. (1990) Genes Dev. 4, 1427–1438.
[14] Yamamoto, M., Ko, L.J., Leonard, M.V., Beug, H., Orkin, S.H. and Engel, J. (1990) Genes Dev. 4, 1650–1662.
[15] Cereghini, S., Raymondjen, M., Carranca, A., Herbomel, P. and Yaniv, M. (1987) Cell 50, 627–638.
[16] Bartsch, D., Ghirardi, M., Skehel, P.A., Karl, K.A., Herder, S.P., Chen, M., Bailey, C.H. and Kandel, E.R. (1995) Cell 83, 979–992.
[17] Ammendola, R., Gounari, F., Piaggio, G., De Simone, V. and Cortese, R. (1990) Mol. Cell. Biol. 10, 387–390.
[18] Jonat, C., Rahmsdorf, H., Park, K., Cato, A.C., Gebel, S., Ponta, H. and Herrlich, K.-K. (1990) Cell 62, 1189–1204.
[19] Ryan, W.A., Franza, B.R. and Gilman, M.L. (1989) EMBO J. 8, 1785–1792.
[20] Di Lorenzo, D., Williams, P. and Ringold, G. (1991) Biochem. Biophys. Res. Commun. 176, 1326–1332.
[21] Kumarev, V.P., Kobzev, V.F., Kuznedelov, K.D. and Sredin, Y.G. (1991) Nucleic Acids Res. Symp. Ser. 24, 234.
[22] Gorski, K., Carnero, M. and Schibler, U. (1986) Cell 47, 767–776.
[23] Shapiro, D.G., Sharp, P.A., Wahli, W.W. and Keller, M.J. (1988) DNA 7, 47–55.
[24] Ponomarenko, M.P., Ponomarenko, J.V., Frolov, A.S., Podkolodnaya, O.A., Vorobiev, D.G., Kolchanov, N.A. and Overton, G.C. (1999) Bioinformatics 15, 644–653.
[25] Schug, J. and Overton, G.C. (1997) Technical report CBIL-TR-1997-1001-v 0.0, Comp. Biology and Informatics Laboratory, School of Medicine, University of Pensylvania, PA.
[26] Treisman, R. (1992) TIBS 17, 423–426.
[27] Gualberto, A., LePage, D., Pons, G., Mader, S.L., Park, K., Atchison, M.L. and Walsh, K. (1992) Mol. Cell. Biol. 12, 4209–4214.
[28] Lee, T.-C., Chow, K-L., Fang, P. and Schwartz, R.J. (1991) Mol. Cell. Biol. 11, 5090–5099.
[29] Meier, V.S. and Groner, B. (1994) Mol. Cell. Biol. 14, 128–137.
[30] Vorobiev, D.G., Ponomarenko, J.V. and Podkolodnaya, O.A. (1998) in: Proc. First Int. Conf. On Bioinformatics of Genome Regulation and Structure, Vol. 1, pp. 58–61.
[31] Lawrence, C. (1994) Comput. Chem. 18, 255–258.
[32] Ye, J. and Young, H.A. (1997) FASEB J. 11, 825–833.
[33] Prada, A.E., Zahedi, K. and Davis, A.E. (1998) Immunology 199, 377–388.
[34] Merkulov, V.M. and Merkulova, T.I. (1992) Biochem. Biophys. Acta 1132, 100–102.
[35] Danesch, U., Closs, B., Schmid, W., Shutz, G., Schule, R. and Renkawitz, R. (1987) EMBO J. 6, 625–630.