Available online at www.sciencedirect.com
ScienceDirect
Antiviral Research 75 (2007) 167–172
Antiviral Research
www.elsevier.com/locate/antiviral

Short communication

Identification of Herpes TATT-binding protein

Lucjan S. Wyrwicz [a,b,*], Leszek Rychlewski [a,c]

[a] *BioInfoBank Institute, Limanowskiego 24A/16, 60-744 Poznan, Poland*
[b] *Department of Gastroenterology, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Roentgena 5, 02-781 Warsaw, Poland*
[c] *Faculty of Physics, A. Mickiewicz University, Umultowska 85, 61-614 Poznan, Poland*

Received 3 November 2006; accepted 5 March 2007

Abstract

The regulation of viral gene expression is a compilation of virus and host factors influencing the transcription machinery. In Epstein-Barr Virus (EBV) a distinct regulatory element utilizing the TATT-box was described. The motif is present in promoters of lytic cycle genes and resembles a crucial host genome motif (TATA-box). Since the binding specificity of eukaryotic proteins recognizing TATA-box (TBP) was determined and no specific preference for interaction with TATT motif was found, we performed a genome-wide fold recognition search to identify viral proteins potentially recognizing the TATT-box. By applying profile–profile comparisons and homology-based protein structure prediction we identified a protein of unknown function from *Gammaherpesvirinae* (BcRF1 of EBV) and their *Betaherpesvirinae* homologs (UL87 of CMV) as proteins encoding TBP fold. Although overall sequence identity is very low (circa 10%), the saddle-like fold and presence of important residues on a surface of DNA–protein interface marked both proteins as distantly related to TBP and permitted the characterization of a putative molecular basis of selective recognition of TATT-motif by BcRF1.

Keywords: *Herpesviridae*; Regulation of transcription; Promoter; Protein structure prediction; Bioinformatics

Herpesviridae constitute a diverse family of viruses characterized by a large dsDNA genome containing between 80 and 150 genes and the ability to enter a latent phase of infection in specific host cells (Subak-Sharpe and Dargan, 1998). Species of Herpes are classified into three different evolutionary subfamilies—*Alphaherpesvirinae* (HSV1, HSV2, VZV), *Betaherpesvirinae* (CMV, HHV-6, HHV-7) and *Gammaherpesvirinae* (EBV, HHV8).

As viruses utilize several host cell mechanisms in replication, the regulation of gene expression in dsDNA viruses arise as a complex result of interplay between host and viral protein factors. Previously a specific sequence motif (TATT) of the core promoter of *Gammaherpesvirinae* lytic cycle genes was identified (Serio et al., 1998; Tang et al., 2004). The motif is highly similar to the most common regulatory element of eukaryotic promoters—TATA-box (Bucher, 1990). The compilation of TATA- and TATT-box signals from EBV genome and difference between these two elements are shown in Table 1.

TATA-binding protein (TBP) is one of the most conserved proteins of eukaryotic and archaeal genomes. As the molecule has been a subject of thorough crystallographic studies the mechanism of its action is well known (Bewley et al., 1998) and a systematic analysis of TATA-box motif mutants suggests that TBP can bind to both TATA and TATT element. However, interaction with TATT-motif was established to be less effective than binding to the consensus TATA-box motif (Patikoglou et al., 1999). The TBP is one of the essential cellular proteins its expression is constitutive rather than induced. Moreover, the existence of an overrepresented motif in a group of functionally related genes, *per analogiam* to regulation of gene expression in humans (Kel et al., 2001), suggests the presence of an additional transcription regulator specifically interacting with the described motif.

Using Meta-BASIC (http://basic.bioinfo.pl) (Ginalski et al., 2004), a highly sensitive method aimed at the identification of distant similarity that applies a comparison of sequence profiles (protein family sequence signatures) (Jaroszewski et al.,

Abbreviations: TBP, TATA-binding protein; PDB, Protein Data Bank

* Corresponding author at: BioInfoBank Institute, ul. Limanowskiego 24A, 60-744 Poznan, Poland. Tel.: +48 61 865 35 20; fax: +48 61 864 33 50.
E-mail address: lucjan@bioinfo.pl (L.S. Wyrwicz).

0166-3542/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.antiviral.2007.03.002

Table 1
Sequence characteristics of TATA- and TATT-box motifs extracted from EBV genome (GenBank: NC_007605)

Position	TATA box					TATT box				
	A	G	C	T	consensus	A	G	C	T	consensus
-5	7	8	15	10		8	10	6	6	
-4	11	12	11	6		8	7	6	9	
-3	7	7	14	12		7	7	7	9	
-2	16	15	5	4		13	8	7	2	
-1	5	17	10	8		8	10	3	9	
+1	0	4	6	30	T	0	0	1	29	T
+2	39	0	0	1	A	30	0	0	0	A
+3	0	0	3	37	T	0	0	0	30	T
+4	39	1	0	0	**A**	0	0	0	30	**T**
+5	31	0	0	9	**A > t**	12	0	0	18	**T > a**
+6	32	3	3	2	A	27	1	1	1	A
+7	19	10	1	10	A > g, t	23	3	2	2	A
+8	8	13	8	11		13	7	7	3	
+9	11	15	10	4		10	5	8	7	
+10	5	17	15	3		6	7	11	6	
+11	5	12	9	14		2	8	12	8	

2005) enriched by predicted secondary structure, we identified the weak similarity between BcRF1 of EBV and TBP (proteins from thermophilic archea *Pyrococcus woesi* – Protein Data Bank – PDB accession: 1pcz and *Sulfolobus acidocaldarius* – 1mp9). The domain was identified in a central part of BcRF1 as well as in homologous proteins from *Gammaherpesvirinae* and their homologs in *Betaherpesvirinae* including UL87 of CMV.

In order to investigate the proper sequence–structure relationship, we subsequently submitted the mentioned above sequences of Herpes virus TBP-like proteins to the Protein Structure Prediction Meta-Server (http://bioinfo.pl/meta) (Bujnicki et al., 2001), which combines several different prediction methods to enable the collection of most similar structural templates from crystallographically solved protein structures deposited in the Protein Data Bank (Berman et al., 2000). The protein fold assignment was confirmed with confident scores by various prediction methods, including FFAS3 (Jaroszewski et al., 2005), Orfeus (Ginalski et al., 2003b), and INUB (Fischer, 2003).

The prediction quality assessment method—3D-Jury (Ginalski et al., 2003a) supported the identification of a saddle-like TATA-binding protein fold with EBV's BcRF1 having a 3D-Jury score of 68.00 and CMV's UL87 a somewhat lower score of 61.12. Previous benchmarks demonstrated that 3D-Jury scores above 50 are associated with fold assignment error rate in less than <5% of total cases (Ginalski et al., 2003a). The combination of Meta-BASIC and Meta-Server methodologies (von Grotthuss et al., 2003) was used to create the structural alignment of Herpes proteins and selected TBPs (Fig. 1). The structure and function of flanking protein sequences, as well as *Betaherpesvirinae* N-terminal domain DUF587 (PFAM accession PF04532) remain unknown as these could not be reliably identified with the described bioinformatic methodology.

The overall sequence identity between the investigated Herpes virus proteins and TBPs is very low (9–14%). To confirm that apart from sharing the general fold the TBP homologs also retained the proper function, we analyzed the critical 15 residues involved in interaction with minor groove of TATA-box (Bewley et al., 1998). Their location and conservation throughout the sequence alignment is shown in Fig. 1.

The TATA-binding protein is a symmetrical domain folding into a saddle-like structure, where each subunit of TBP consists of a five-stranded, curved antiparallel β-sheet and two α-helices (one long and one short), as depicted in Fig. 2 (Bewley et al., 1998). The central eight strands form a curved β-sheet interacting with the distorted minor groove of the TATA-box, while the longer α-helix nearly orthogonal to the β-sheet provides a structural scaffold for the curved β-sheet (together with the shorter helix).

Apart from binding to the minor groove of the double-stranded DNA molecule containing the TATA-box motif, the TBP bends the nucleic acid and facilitates dissociation of the two strands of DNA. TBP utilizes two phenylalanine residues (F_3, F_A; residues marked with "3" and "A" in the alignment in Fig. 1) to kink DNA in the concave structure (Patikoglou et al., 1999). The structural relationship between these residues was mapped onto the schematic representation of TATA-motif according to the crystallographically solved structure of the TBP/TATA-box complex (PDB: 1pcz). The result is shown in Fig. 3. The asparagine (N_1, N_8) and valine (V_2, V_9) residues are located in the very center and symmetrically present in both parts of this duplicated fold. The peripherally located basic amino acids interacting with phosphate groups are not involved in recognition of specific nucleic acid bases. In TBP, atypically for the size of its interaction surface, there are only six hydrogen bonds involved in the interaction between nucleic acid bases and the concave interior of the protein (four bonds from asparagines N_1, N_8; two from threonines T_7, T_F, respectively), The majority of TATA-box/TBP interactions arise as the result of hydrophobic and van der Waals interactions (Bewley et al., 1998).

The comparison of 15 residues (1–9, A–F) involved in interaction with minor groove of DNA (Fig. 3) reveals that six

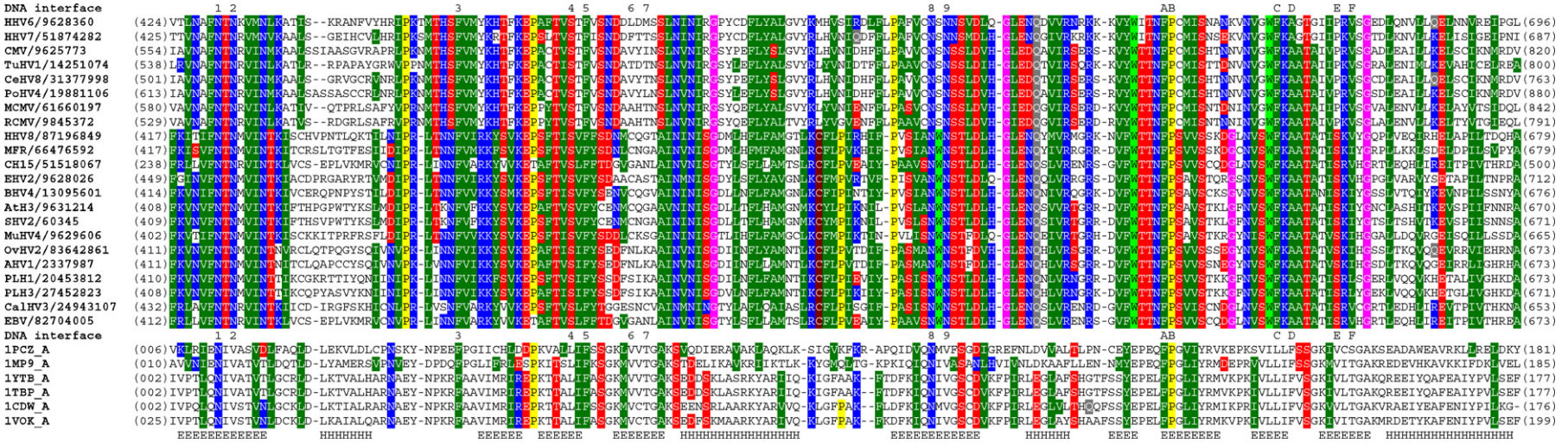

Fig. 1. Sequence alignment of BcRF1 and UL87 homologs and selected TATA-binding proteins. Sequences are marked with either the organism code and GenBank gene identifier or their corresponding PDB database entry. The homologs of EBV and CMV proteins were collected with MetaBasic from "nr90" database (the NCBI NR database filtered at similarity level of 90% to remove redundancy). Sequences included in the alignment represents *Betaherpesvirinae* (HHV6—Human herpesvirus 6; HHV7—Human herpesvirus 7; CMV—Human herpesvirus 5; TuHV1—Tupaia herpesvirus; CeHV8—Cercopithecine herpesvirus 8; PoHV4—Pongine herpesvirus 4; MCMV—Murid herpesvirus 1; RCMV—Murid herpesvirus 2) and *Gammaherpesvirinae* (EBV—Human herpesvirus 4; HHV8—Human herpesvirus 8; MFR—Macaca fuscata rhadinovirus; CH15—Cercopithecine herpesvirus 15; EHV2—Equid herpesvirus 2; BHV4—Bovine herpesvirus 4; AtH3—Ateline herpesvirus; SHV2—Saimiriine herpesvirus; MuHV4—Murid herpesvirus 4; OvHV2—Ovine herpesvirus 2; AHV1—Alcelaphine herpesvirus 1; PLH1—Porcine lymphotropic herpesvirus 1; PLH3—Porcine lymphotropic herpesvirus 3; CalHV3—Callitrichine herpesvirus 3). Residues localized at the DNA–protein interface were marked (1–9, A–F). Secondary structure was coded with letters (E—extended, H—Helix) according to the structure of 1PCZ.

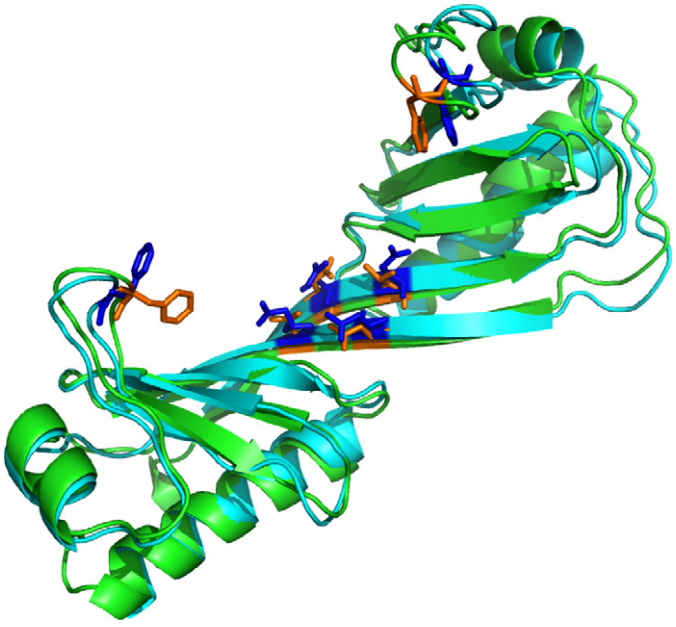

Fig. 2. Superimposed structures of TBP (green and orange; PDB: 1pcz) and EBV TATT-binding protein (cyan and blue) modeled with Modeller 6.2, according to the structural alignment shown in Fig. 1. Note the side chains of two conserved phenylalanines (F_3, F_A) and asparagine/valine cluster responsible for TATA/TATT differentiation.

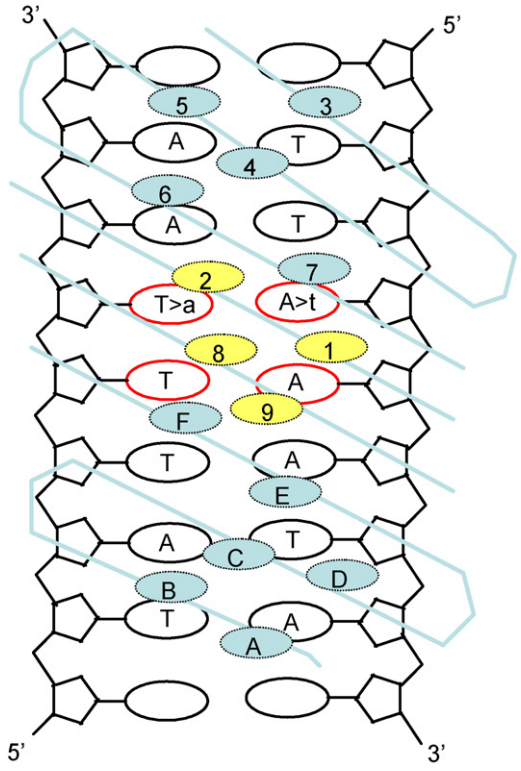

Fig. 3. The schematic view of minor groove of DNA of TATT motif and location of the amino acids interacting with the nucleic acid bases. The consensus sequence of TATT motif was shown. The residues creating DNA–protein interface are numbered according to positions in Fig. 1. The location of four asparagines is shown in yellow (conserved—N_1, N_8; herpes specific—N_2, N_9).

residues are seemingly preserved throughout both TBPs and the investigated *Herpesviridae* homologues (asparagines—N_1, N_8; phenylalanines—F_3, F_5, F_A; proline—P_B). For two pairs of amino acids localized within the close spatial proximity to each other (C-D and E-F), we observed compensatory mutations in form of a joint change of leucine–phenylalanine (L_C–F_D) in TBP to phenylalanine–alanine (F_C–A_D); and an exchange of valine–threonine (V_E–T_F) or valine–serine (V_E–S_F) in TBP with serine–valine (S_E–V_F) present in *Gammaherpesvirinae* and proline–valine (P_E–V_F) in *Betaherpesvirina*e.

Of the multiple sequence alignment positions corresponding to residues located on the interaction surface of the Herpes virus TBP-like protein, the strongest variation occurred at positions 6 (TBP's valine V_6) and 7 (TBP's threonine T_7). There is a higher incidence ratio of valine–threonine consensus among *Betaherpesvirinae*, but no general rule for the amino acid composition seems to exist and the positions differ among all human pathogens (*Gammaherpesvirinae*: EBV—G_6–N_7, HHV8—C_6–G_7; *Betaherpesvirinae*: CMV—V_6–I_7; HHV6—L_6–M_7; HHV7—F_6–T_7). Contrary to the observed variability of amino acid usage at positions 6 and 7, serine (S_4) corresponding to TBP's leucine (L_4) is completely conserved in Herpes virus, which is consistent to spatially close phenylalanines (F_3 and F_5). Side chains of the phenylalanine residues are involved in recognition of distal part of the motif (+7 and +8 in Table 1).

Perhaps, the most intriguing aspect of the emerging active site is the presence of the conserved asparagines N_2 and N_9 in place of valines (V_2, V_9) localized in the center of the TBP concave (Fig. 3). This results in a creation of a cluster of four asparagines in the region responsible for differentiation between TATA and TATT motifs (thymines at positions +4 and +5 of TATT consensus—compare with Table 1 and Fig. 3).

To investigate the interaction of EBV BcRF1, the protein model was created using *Archea* PDB template 1pcz (TBP molecule bound to the consensus TATT oligonucleotide—TATTTAAG). The resulting model was obtained with Modeller 6.2 (Sanchez and Sali, 2000) according to the structural alignment shown in Fig. 1. To study the protein's binding mode we used a dsDNA particle merged from solved structures of TPB and TATA-box mutants (PDB entries: 1qn8—GCTATTAAAGGGCA; 1qn7—GCTATATAAGGGCA) (Patikoglou et al., 1999). The minimal conformation of asparagines cluster (N_1, N_2, N_8, N_9 and additional EBV-specific N_7) in complex with TATT was calculated by application of Dreiding potential (Discovery Studio ViewerPro Suite; Accelrys Software Inc). Results of these calculations are shown in Fig. 4. Therefore, we conclude that asparagine N_7 (via interaction with N_2) creates an additional hydrogen bond to thymine at position +4 of TATT motif. Formation of this complex is augmented by interaction between N_1 and thymine +5, as well as N_8 and complement adenine (A′ + 5), however – as the rotation of amide groups at N_1 and N_8 is possible – this particular interaction cannot be responsible for the specific recognition of thymine at position +5 (compare with data shown in Table 1). Notably, both N_1 and N_8 are conserved between TBP and the investigated TATT-binding protein, while N_1 and N_9 are specific for Herpes virus TBP.

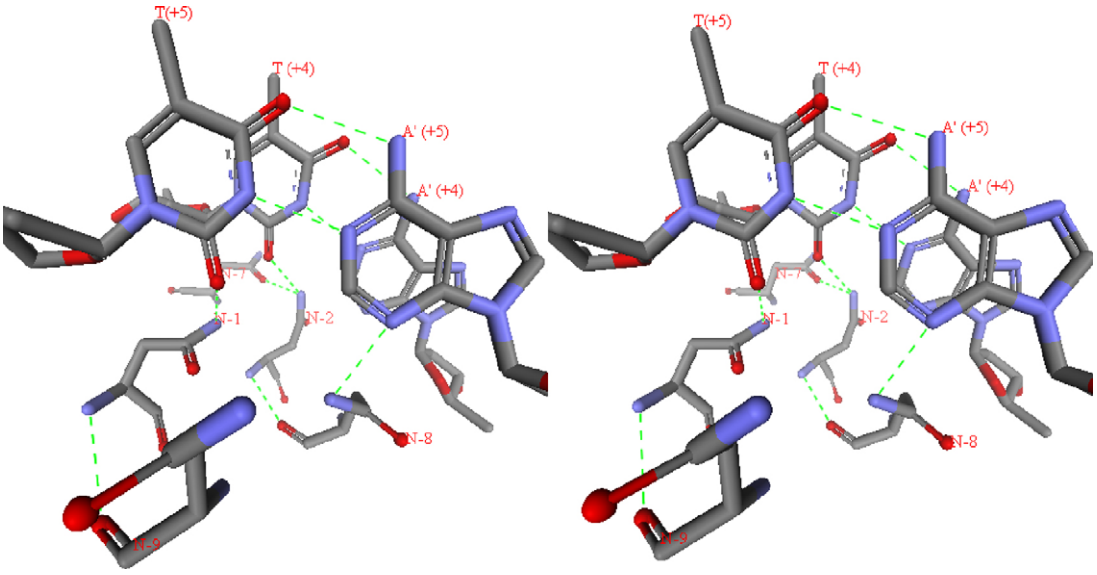

Fig. 4. The stereo-view of potential interaction between asparagines residues (N_1, N_2, N_7, N_8, N_9) in EBV TATT-binding protein and the fragment of TATT motif (base pairs +4 and +5).

It is important to note the existing discrepancy in the naming of EBV genes. The EBV gene described here (BcRF1) is not the same as IL-10 homolog encoded by similarly named BCRF1. Indeed, the two proteins are unrelated and encoded in different genomic loci located at positions 125073–128686 (BcRF1) and 9631–10262 (BCRF1) of the EBV genome reference sequence (GenBank entry NC_007605).

In summary, the application of bioinformatic methodology permitted the identification previously unknown, specific TATT-binding proteins in *Beta-* and *Gammaherpesvirinae.* The subsequent structural study described above, confirms the fold assignment and allows for identification of the common residues localized on the surface of the DNA–protein interface. We infer that the novel TATT-binding protein utilizes a highly similar mechanism of DNA–protein interaction to eukaryotic TATA-binding protein. In particular, while the conservation of essential residues present on the surface of saddle-like structure is by no means perfect, the presence of critical phenylalanine residues responsible for kinking DNA coupled with the higher number of hydrogen-bond donors and acceptors supports this functional assignment. The question whether the TATT-binding protein evolved from TBP or represents a product of convergent evolution cannot be answered easily.

The presence of asparagine residues N_2 and N_9 creates additional possibilities for specific recognition of bases at positions +4 and +5 of the motif. Although these residues are present throughout the whole family of Herpes virus TBP (Fig. 1), the lack of EBV-specific N_7 could modulate the specificity of the recognized TATT-motif. If this is the case—the high variability at positions 6 and 7, corresponding to conserved V_6 and T_7 in TBP, leaves a potential for the further engineering of proteins interacting with the minor grove of TATA-like motifs. Since TATA-binding protein is active only in complex with other protein agents (Johnson et al., 2003), the identification of proteins interacting with viral TATT-binding protein emerges as an important issue. At present, we are unable to provide any details on the function of the flanking regions of BcRF1 and its homologs, but it is probable that their protein products possess several other functions associated with TBP's activity (e.g.: TFIIA, TFIIB).

Our analysis apart from providing an insight into the action of TATT-binding protein allows for further exploitation of discovery in a field of gene therapy. Since—TATT motif is not a specific regulatory motif of human genes, and Herpes viruses represent a potential gene therapy vector, the further optimization of TATT-binding proteins can lead to creation of vectors harboring genes regulated with modifiable activators of its transcription.

Acknowledgements

This work has been supported by MNiSW and the European Commission (LSHG-CT-2003-503265, LSHG-CT-2004-512035) and MNiSW (2P05A00130 and PBZ-MNiI-2/1/2005). LSW is a fellow of the Young Scientist Award from The Foundation for Polish Science. Authors would like to thank Dariusz Plewczynski for support in molecular dynamics calculations and Grzegorz Koczyk for critical reading of the manuscript.

References

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P., 2000. The Protein Data Bank. Nucleic Acids Res. 28 (1), 235–242.

Bewley, C., Gronenborn, A., Clore, G., 1998. Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu. Rev. Biophys. Biomol. Struct. 27, 105–131.

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.

Bujnicki, J., Elofsson, A., Fischer, D., Rychlewski, L., 2001. Structure prediction meta server. Bioinformatics 17, 750–751.

Fischer, D., 2003. 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. Proteins 51, 434–441.

Ginalski, K., Elofsson, A., Fischer, D., Rychlewski, L., 2003a. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19, 1015–1018.

Ginalski, K., Pas, J., Wyrwicz, L., von Grotthuss, M., Bujnicki, J., Rychlewski, L., 2003b. ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. Nucleic Acids Res. 31, 3804–3807.

Ginalski, K., von Grotthuss, M., Grishin, N., Rychlewski, L., 2004. Detecting distant homology with Meta-BASIC. Nucleic Acids Res. 32, W576–W581.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., Godzik, A., 2005. FFAS03: a server for profile–profile sequence alignments. Nucleic Acids Res. 33, W284–W288.

Johnson, S., Dubeau, L., White, R., Johnson, D., 2003. The TATA-binding protein as a regulator of cellular transformation. Cell Cycle 2, 442–444.

Kel, A., Kel-Margoulis, O., Farnham, P., Bartley, S., Wingender, E., Zhang, M., 2001. Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors. J. Mol. Biol. 309, 99–120.

Patikoglou, G., Kim, J., Sun, L., Yang, S., Kodadek, T., Burley, S., 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. 13, 3217–3230.

Sanchez, R., Sali, A., 2000. Comparative protein structure modeling. Introduction and practical examples with modeller. Methods Mol. Biol. 143, 97–129.

Serio, T.R., Cahill, N., Prout, M.E., Miller, G., 1998. A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. J. Virol. 72, 8338–8343.

Subak-Sharpe, J., Dargan, D., 1998. HSV molecular biology: general aspects of herpes simplex virus molecular biology. Virus Genes 16, 239–251.

Tang, S., Yamanegi, K., Zheng, Z.M., 2004. Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. J. Virol. 78, 2609–2614.

von Grotthuss, M., Pas, J., Wyrwicz, L., Ginalski, K., Rychlewski, L., 2003. Application of 3D-Jury, GRDB, and Verify3D in fold recognition. Proteins 53 (Suppl. 6), 418–423.