

Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity

Andreas Klingenhoff¹, Kornelie Frech², Kerstin Quandt² and Thomas Werner^{1,2}

¹Institute of Mammalian Genetics, GSF-National Research Center for Environment and Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg and ²Genomatix Software GmbH, Karlstrasse 55, D-80333 Munchen, Germany

Received on October 6, 1998; accepted on November 26, 1998

Abstract

Motivation: Gene regulation often depends on functional modules which feature a detectable internal organization. Overall sequence similarity of these modules is often insufficient for detection by general search methods like FASTA or even Gapped BLAST. However, it is of interest to evaluate whether modules, often known from experimental analysis of single sequences, are present in other regulatory sequences.

Results: We developed a new method (FastM) which combines a search algorithm for individual transcription factor binding sites (MatInspector) with a distance correlation function. FastM allows fast definition of a model of correlated binding sites derived from as little as a single promoter or enhancer. ModelInspector results are suitable for evaluation of the significance of the model. We used FastM to define a model for the experimentally verified NF κ B/IRF1 regulatory module from the major histocompatibility complex (MHC) class I HLA-B gene promoter. Analysis of a test set of sequences as well as database searches with this model showed excellent correlation of the model with the biological function of the module. These results could not be obtained by searches using FASTA or Gapped BLAST, which are based on sequence similarity. We were also able to demonstrate association of a hypothetical GRE-GRE module with viral sequences based on analysis of several GenBank sections with this module.

Availability: The WWW version of FastM is accessible at: <http://www.gsf.de/cgi-bin/fastm.pl> and <http://genomatix.gsf.de/cgi-bin/fastm2/fastm.pl>

Contact: werner@gsf.de

Introduction

Large-scale genome sequencing projects like the Human Genome Project reveal huge amounts of anonymous sequence data containing regulatory regions. Similarity of these regulatory regions is often restricted to short discontinuous DNA stretches and therefore frequently escapes detection by sequence similar-

ity searches using, for example, FASTA (Pearson and Lipman, 1988) or Gapped BLAST (Altschul *et al.*, 1997). Regulatory units like promoters, enhancers or silencers contain functional modules, many of which consist of at least two transcription factor (TF) binding sites in conserved order, separated by a spacer. Methods for the identification of single TF binding sites already exist (e.g. Frech *et al.*, 1993, 1997a; Chen *et al.*, 1995; Quandt *et al.*, 1995; Prestridge, 1996). Recently, a database of so-called composite elements (COMPELs) of two TF binding sites was developed (Kel *et al.*, 1995; Heinemeyer *et al.*, 1998). These COMPELs were all shown to exhibit synergistic or antagonistic functions. Such combinations might be useful for the detection of target genes, as discussed in Lavorgna *et al.* (1998).

Here, we used a new method (FastM) which allows the definition of models representing such regulatory modules to define and analyze an NF κ B/IRF1 regulatory module described for the major histocompatibility complex (MHC) class I HLA-B gene promoter (Johnson and Pober, 1994). This module was taken from the COMPEL database (Heinemeyer *et al.*, 1998). Based on GenBank searches with the model, we were able to find promoter sequences of quite different genes which are regulated in a similar way. FASTA or Gapped BLAST failed to reflect the common function in their search results. FastM can also be used to define hypothetical modules which lack experimental data about synergistic binding, as demonstrated for a GRE pair from the Mouse Mammary Tumor Virus (MMTV) LTR. Therefore, FastM is a useful tool for the description and evaluation of regulatory modules.

Methods

FastM

FastM is a new method for the development of simple models of transcriptionally regulatory units. These models are composed from various individual elements, their sequential order and the distances between them. The parameters of a FastM model are solely defined by the user.

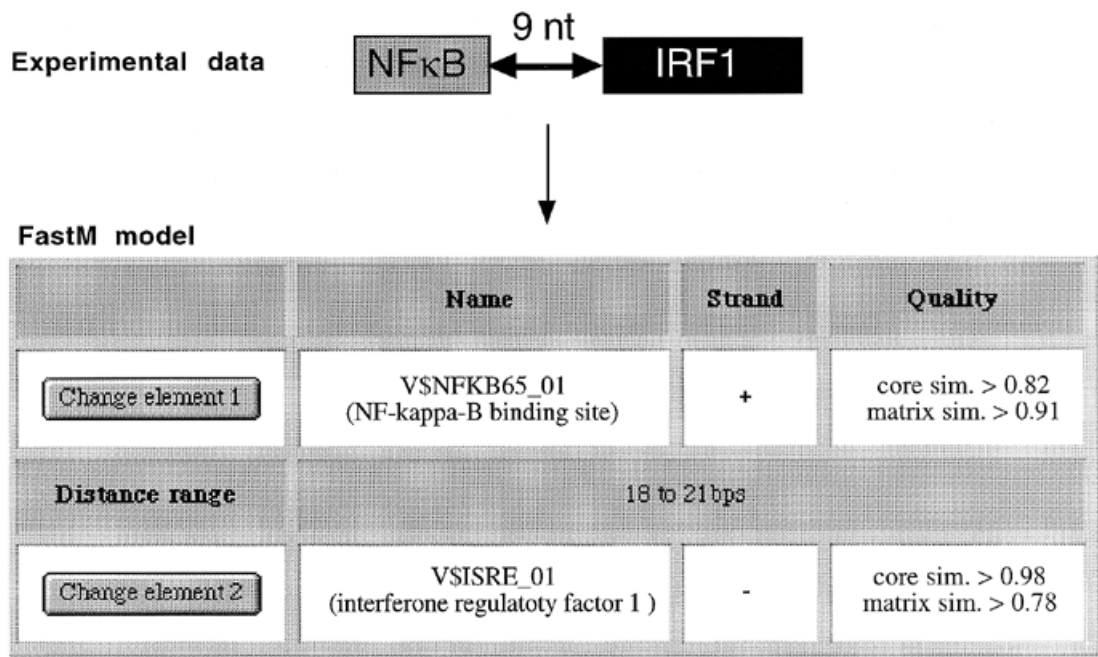


Fig. 1. FastM model of the MHC class I module. Top, a graphical representation of the experimental data used for the model is shown. Below, a screen shot of the final model is shown as displayed by FastM. The name column indicates the matrix used, strand indicates the relative strand orientation, and the quality column identifies the individually set core and matrix similarities. Between the elements, the allowed distance range is shown (minimum distance 18, maximum distance 21 nucleotides). Distances are calculated from the beginning of one matrix to the beginning of the next matrix, not between the end of the first and the beginning of the second matrix as shown for the experimental data. The actual distance range used by ModelInspector is 13–26 bp (see Methods for details).

Individual elements may be TF binding sites detected by MatInspector (Quandt *et al.*, 1995) or user-defined IUPAC sequences. Additional elements like hairpins, direct repeats, short multiple repeats and terminal repeats are supported by a commercial version (FastM professional, Genomatix Software GmbH). A more detailed description of these individual elements is presented in Frech *et al.* (1997b). Parameters for each element are the individual cut-off values (in the case of weight matrices, these are the core similarity and the matrix similarity; in the case of IUPAC sequences, the maximum number of mismatches allowed) and the strand orientation (+ strand, – strand, or both strands). Further features of a model are the allowed distance ranges (e.g. 10–20 bp) for adjacent individual elements. Distances are given from the beginning of an element to the beginning of the next element. The complete model for the MHC class I module described in this paper is shown in Figure 1 as it appears on screen of the World Wide Web (WWW) version of FastM.

Availability

An on-line FastM with an easy-to-use interface is available on the WWW (<http://www.gsf.de/cgi-bin/fastm.pl> or

<http://genomatix.gsf.de/cgi-bin/fastm2/fastm.pl>). This WWW version is restricted to models of two TF binding sites which can be selected from the MatInspector library or provided by the user as IUPAC strings. A strand-specific orientation of individual elements is possible, but not mandatory. The combination with ModelInspector allows immediate scanning of user sequences or GenBank/EMBL sections with the defined FastM model. The user is notified about completion of the analysis by e-mail.

ModelInspector

After a model is defined by FastM, ModelInspector can be used to scan sequences of unlimited length or database sections for matches to the model. ModelInspector first scans the sequences for individual elements which are present in the distance range defined by the model. ModelInspector automatically extends the distance range by $(\text{maxdist.} - \text{mindist.})/2$, at least by five nucleotides in each direction. Then, the score of a model match is calculated as the sum of the individual element scores. The program was described in detail in Frech *et al.* (1997b).

Table 1. Search results of different representations of the MHC class I COMPEL in the primate section of GenBank (Release 107)

Test sets	Positive				Negative	Database additional matches
	HLA-A	HLA-B	β -2-micro-globulin	β -IFN	HLA-C	
No. of sequences	16	42	1	5	24	83 853
ModelInspector	15	42	1	5	1	189
FASTA	0	42	0	0	24	1
Gapped Blast	0	25	0	0	17	7

Total of 227 066 571 nucleotides in 83 941 sequences.

FASTA analysis

The FASTA analysis (Pearson and Lipman, 1988) was carried out with the GCG version using the exact COMPEL sequence of the HLA-B promoter (X82915, 87-120) with a word size of two, a gap creation penalty of 16, a gap extension penalty of four and an E () value of <5.

Gapped BLAST analysis

Gapped BLAST (Altschul *et al.*, 1997) was carried out on the NCBI Blast Server (<http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=0>) with the COMPEL sequence. We used the Basic BLAST search which does not have parameters to be set by the user.

Results

Tumor necrosis factor (TNF) and interferons (β -IFN, γ -IFN) synergistically induce expression of the human MHC class I gene HLA-B. The synergism is mediated by binding of the TFs NF κ B and IRF1 to their respective binding sites which comprise a regulatory module of 34 bp (MHC class I COMPEL) located 198 bp upstream of the transcription start site of the HLA-B gene (Johnson and Pober, 1994).

Definition of the FastM model

The first step to build a FastM model for the HLA-B regulatory module was to assess detection of the binding sites by weight matrices. Therefore, the initial model consisted of a NF κ B matrix followed by an IRF1 matrix (weight matrices from the MatInspector matrix library; Quandt *et al.*, 1995) in a distance of ~10 nucleotides. This information was taken from Johnson and Pober (1994). The HLA-B sequence (Ac: X82915) served as a training sequence. We set the thresholds for the core similarity (CS) and the matrix similarity (MS) to 0.7 to account for the possibility of one weak binding site which might occur due to cooperative binding of the two TFs (Perkins *et al.*, 1993; Lee *et al.*, 1994). Any strand combination of two matrix matches was allowed provided the matches occurred within the defined distance range. If more than one match to this initial FastM model would have been found in the training sequence (which was not the case), selection of the correct match would have been poss-

ible based on the distance of the matches to the transcription start site (TSS) known from experimental data. The results of this analysis were used to optimize the parameters for strand orientation, distance and the matrix thresholds, ensuring that the model exactly fitted the regulatory module in the training sequence (Figure 1).

Analysis of GenBank primate sections with the HLA-B model

This model was used in a ModelInspector search in the GenBank primates section. We found 15 matches to HLA-A and 42 matches to HLA-B sequences, as well as one promoter of a β -2-microglobulin gene and 5 β -IFN promoter sequences (Table 1). Only one HLA-C sequence was found, although several such sequences are present in the database. We found 189 further matches to the FastM model, 138 of which were located in large anonymous genomic sequences and could not be evaluated further. The remaining matches were not analyzed further and were considered false positives.

Comparison with FASTA and Gapped BLAST

HLA promoter sequences are known to share considerable sequence similarity, raising the question whether the results from our FastM model could be achieved as well with a conventional similarity search. Therefore, we carried out FASTA (Pearson and Lipman, 1988) and Gapped BLAST (Altschul *et al.*, 1997) searches using the parameters detailed in the Methods for each program. To assess the quality of the results and to provide a basis for comparison of the different methods, a test set of 64 true-positive and 24 true-negative sequences was collected. Sixteen HLA-A, 42 HLA-B, 1 β -2-microglobulin and 5 β -IFN promoter sequences that are annotated in the primate sections of GenBank Release 107 and contain at least 200 bp upstream of the TSS were taken as a positive test set. The 24 annotated HLA-C promoter sequences containing this region comprised the negative training set because they are known not to contain the functional MHC class I module. The results obtained by ModelInspector, FASTA and Gapped BLAST are shown in Table 1. ModelInspector using the FastM model correctly predicted all except one of the sequences in the positive test set. FASTA and Gapped BLAST each correctly predicted significantly fewer sequences.

Table 2. Results of GenBank search for GRE-GRE modules

GenBank section (Release 107)	Number of base pairs	Number of GRE sites in 1 million bp	(abs.)	Number of GRE-GRE modules in 1 million bp	(abs.)
Primates	227 066 571	410	(93 041)	0.36	(82)
Other mammalian	14 270 603	405	(5773)	0.56	(8)
Other vertebrates	20 805 593	410	(8536)	0.43	(9)
Rodents	53 548 991	473	(25 339)	0.65	(35)
Viral	51 446 291	460	(23 665)	2.45	(126)

Table 3. GRE-GRE module matches in viral section

Virus	Classification	Number of matches
Mouse mammary tumor virus	Type B retrovirus	28
Thymotropic retrovirus	Type B retrovirus	4
Gibbon leukemia virus	Type C retrovirus	2
Murine leukemia virus	Type C retrovirus	31
Murine sarcoma virus	Type C retrovirus	21
Other type C retroviruses	Type C retrovirus	8
HIV-1	Lentivirus	23
Other viruses		7

To elucidate these large differences obtained by FastM/ModelInspector and sequence homology searches with FASTA or Gapped BLAST, we carried out a multiple alignment of sequences comprising ~220 bp upstream of the TSS of HLA-A (X55710), HLA-B, HLA-C (U03044), β -2-microglobulin (M17986) and β -IFN (J00218), all of which were taken from GenBank Release 107. The alignment of the sequences with the GCG program PileUp is shown in Figure 2. The three MHC class I sequences (HLA-A, HLA-B and HLA-C) show strong conservation over the complete length of the promoter sequence with a remarkable exception in the region where the IRF1 binding site is located. The β -2-microglobulin and the β -IFN promoter sequences exhibit no significant sequence similarity to any of the other sequences. Comparison of the HLA-B and HLA-C sequence showed that two point mutations in the potential NF κ B site are correlated with loss of responsiveness of the HLA-C promoter to TNF. The single HLA-C sequence found by ModelInspector was a revertant in which the NF κ B binding site was restored.

We also used the FastM model of the MHC class I COMPEL to search other GenBank sections (rodents, other mammalian, viral). The most remarkable finding was that 19 of the 22 matches in the viral section belonged to retroviral sequences (probably all within LTRs).

Analysis of a hypothetical module

However, for many sequences, synergy of binding sites has not been verified experimentally. For example, the promoter of

MMTV is known to contain three glucocorticoid responsive elements (GRE; Tanaka *et al.*, 1993). We chose two of these with a distance of 10–90 bp to define a hypothetical module with FastM. The question whether this putative module occurs only in MMTV, is specific for viral promoters or also occurs in cellular promoters can easily be answered using this FastM model and ModelInspector. First, we scanned GenBank Release 107 (sections primates, rodents, other mammalian, other vertebrates, viral) for matches to individual GRE sites. The frequency of GRE sites found by MatInspector is about the same in all sections (~440/1 million bp; see Table 2). Then, the same GenBank sections were analyzed with ModelInspector and the GRE-GRE model generated by FastM. The results show that the module is apparently virus associated as it is clearly overrepresented in viral sequences compared to all other sections analyzed (Table 2). A classification of the 126 matches detected in the viral section is presented in Table 3. This analysis shows that the GRE-GRE module does not only occur in MMTV, but also in other viruses, mostly retroviruses.

Discussion

We described a new method for the definition of regulatory elements that has been implemented in the program FastM. The WWW version of FastM is very easy to use and is combined with ModelInspector, allowing immediate evaluation of the defined models. A single sequence or even a hypothetical combination of individual elements taken from the literature is sufficient for definition of a model and subsequent database scanning to identify similarly organized se-

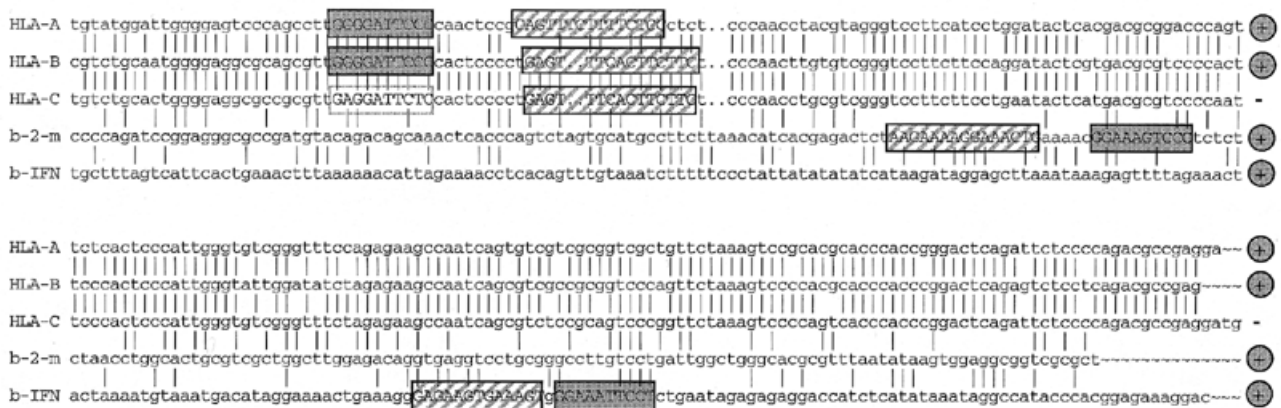


Fig. 2. Alignment of three HLA, one β -2-microglobulin (β -2-m) and one β -interferon (β -IFN) promoter sequences. The NF κ B binding sites are indicated by gray boxes and the IRF1 binding sites are indicated by boxes hatched in gray. Encircled plus signs indicate functionality of the regulatory module consisting of the NF κ B and the IRF1 site. The NF κ B site in HLA-C is rendered non-functional by two point mutations (indicated by a dotted box).

quences. FastM/ModelInspector can find organizational similarities among sequences which do not show significant sequence similarity, as has been demonstrated for the model of the MHC class I COMPEL which was based on a single HLA-B sequence. The database search with this model (Figure 1) also detects the HLA-A promoter sequences which are regulated in the same synergistic way (Johnson and Pober, 1994), although the nucleotide sequences of the IRF1 binding sites differ in a few positions (Hakem *et al.*, 1991). β -2-Microglobulin is expressed at the cell surface as a heterodimer with MHC class I antigens (David-Watine *et al.*, 1990) and shows a similar developmental regulation and tissue specificity (Chamberlain *et al.*, 1988). In the similar regulated β -IFN promoters (Leblanc *et al.*, 1990; Lonergan *et al.*, 1993), which are also found by the FastM model for the MHC class I COMPEL, the sequences of the regulatory modules show even less sequence homology. In the β -IFN promoter, the binding sites are separated by only 1 bp compared to 9 bp in the HLA-B promoter (Figure 2). The flexible spacing introduced by ModelInspector allows us to account even for such large variation. Furthermore, FastM/ModelInspector is able to distinguish between sequences that show high sequence similarity but differ in function because the functional module (e.g. TF binding sites) is not conserved. This could be shown for the HLA-B and HLA-C promoters which share an almost identical sequence in the region of the MHC class I COMPEL (Figure 2). However, due to the two point mutations in the NF κ B site, the HLA-C promoter does not respond to TNF and is therefore not synergistically inducible (Johnson and Pober, 1994).

The alignment in Figure 2 shows clearly that neither the location nor the orientation of the COMPEL module is conserved among those five sequences. However, the internal

organization of the module is conserved, suggesting that functional selection worked at the level of the module rather than at the sequence level.

FASTA (Pearson and Lipman, 1988) and Gapped BLAST (Altschul *et al.*, 1997), which also require only a single sequence for database searches, are restricted to finding sequences with a high overall sequence similarity. A search with the MHC class I COMPEL sequence from the HLA-B promoter neither finds the overall highly similar HLA-A promoter sequences due to local mismatches, nor the less conserved β -2-microglobulin and β -IFN promoter sequences (Table 1). In contrast, the HLA-C sequences were found which do not contain a functional MHC class I COMPEL despite the two mismatches in the NF κ B site. Therefore, FASTA and Gapped BLAST are insensitive to the few, but crucial, mismatches of the HLA-C promoter to the regulatory module of the HLA-B promoter. Using the test set (Table 1), FastM/ModelInspector correctly predicted 98% of the positive test sequences. FASTA and Gapped BLAST predicted 66 and 39% of the positive test sequences.

The more elaborate methods capable of detecting discontinuous organizational patterns, like GenomeInspector (Quandt *et al.*, 1996a,b), FunSiteP (Kondrakhin *et al.*, 1995) or ModelGenerator (Frech *et al.*, 1997b), require either very long or multiple preselected sequences and/or extensive a priori knowledge, which limits their applicability. Since FastM models are suitable for automatic analysis by ModelInspector, new models can be immediately verified by database searches. Thus, FastM is a useful addition to the other existing methods.

We tested the matrix family option of FastM professional with the same COMPEL. A matrix family contains several weight matrices for the same factor and reports the best

match only. The results obtained by this analysis revealed a potential molecular basis for a so far unexplained phenomenon. JC virus is known to replicate in brain in astrocytes and oligodendrocytes (Major *et al.*, 1992). MHC class I genes (HLA-A, HLA-B and HLA-C) and β -2-microglobulin, which are normally expressed at very low levels in the brain, are induced due to viral infection and cytokines (Achim and Wiley, 1992; Mavria *et al.*, 1998). The virus-induced expression of MHC class I genes is also known to depend on IFN (Njenga *et al.*, 1997). We found JC viruses to contain the MHC module, although in a slightly relaxed form. This suggests that JC virus replication as well as MHC class I expression after viral infection in brain might be regulated by this common NF κ B/IRF1 module.

We have so far developed a library of ~50 models, each representing an experimentally verified regulatory module. Apart from a few exceptions, the models show a high specificity, revealing <1 match/10 000 bp.

As shown for the putative GRE-GRE module from MMTV, FastM models can be used to generate and verify hypotheses in the absence of experimental data. The finding that this module is indeed a viral, more precisely, a retroviral module would have been hard to achieve without a method like FastM. The results from a search with a FastM model can also be used to select a training set of sequences suitable for more detailed analysis by ModelGenerator (Frech *et al.*, 1997b) or other higher level methods revealing additional common features to expand the models (e.g. to develop a more extensive promoter model).

FastM should be a very useful tool for initial analysis of regulatory regions for which experimental data are limited, like putative regulatory regions in new genomic sequences.

Acknowledgements

This work was supported by the BMBF project FANGREB 0311641 and by EU grant BI04-CT95-0226 (TRADAT).

References

- Achim, C.L. and Wiley, C.A. (1992) Expression of major histocompatibility complex antigens in the brains of patients with progressive multifocal leukoencephalopathy. *J. Neuropathol. Exp. Neurol.*, **51**, 257–263.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Chamberlain, J.W., Nolan, J.A., Conrad, P.J., Vasavada, H.A., Vasavada, H.H., Ganguly, S., Janeway, C.A., Jr and Weissmann, S.M. (1988) Tissue-specific and cell surface expression of human major histocompatibility complex class I heavy (HLA-B7) and light (β 2-microglobulin) chain genes in transgenic mice. *Proc. Natl Acad. Sci. USA*, **85**, 7690–7694.
- Chen, Q.K., Hertz, G.Z. and Stormo, G.D. (1995) MATRIX SEARCH 1.0: A computer program that scans DNA sequences for transcriptional elements using a database of weight matrices. *Comput. Applic. Biosci.*, **11**, 563–566.
- David-Watine, B., Israel, A. and Kourilsky, P. (1990) The regulation and expression of MHC class I genes. *Immunol. Today*, **11**, 286–292.
- Frech, K., Herrmann, G. and Werner, T. (1993) Computer-assisted prediction, classification and delimitation of protein binding sites in nucleic acids. *Nucleic Acids Res.*, **21**, 1655–1664.
- Frech, K., Dietze, P. and Werner, T. (1997a) ConsInspector 3.0: New library and enhanced functionality. *Comput. Applic. Biosci.*, **13**, 109–110.
- Frech, K., Danescu-Mayer, J. and Werner, T. (1997b) A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J. Mol. Biol.*, **270**, 674–687.
- Hakem, R., Le Bouteiller, P., Jezo-Bremont, A., Harper, K., Campese, D. and Lemonnier, F.A. (1991) Differential regulation of HLA-A3 and HLA-B7 MHC class I genes by IFN is due to nucleotide differences in their IFN response sequences. *J. Immunol.*, **147**, 2384–2390.
- Heinemeyer, T. *et al.* (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.*, **26**, 362–367.
- Johnson, D.R. and Pober, J.S. (1994) HLA Class I heavy-chain gene promoter elements mediating synergy between tumor necrosis factor and interferons. *Mol. Cell. Biol.*, **14**, 1322–1332.
- Kel, O.V., Romaschenko, A.G., Kel, A.E., Wingender, E. and Kolchanov, N.A. (1995) A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res.*, **23**, 4097–4103.
- Kondrakhin, Y.V., Kel, A.E., Kolchanov, N.A., Romashchenko, A.G. and Milanesi, L. (1995) Eukaryotic promoter recognition by binding sites for transcription factors. *Comput. Applic. Biosci.*, **11**, 477–488.
- Lavorgna, G., Bonicelli, E., Wagner, A. and Werner, T. (1998) Detection of potential target genes in silico? *Trends Genet.*, **14**, 375–376.
- Leblanc, J.-F., Cohen, L., Rodrigues, M. and Hiscott, J. (1990) Synergism between distinct enhancer domains in viral induction of the human beta interferon gene. *Mol. Cell. Biol.*, **10**, 3987–3993.
- Lee, Y.-H., Yano, M., Liu, S.-Y., Matsunaga, E., Johnson, P.F. and Gonzalez, F.J. (1994) A novel cis-acting element controlling the rat CYP2D5 gene and requiring cooperativity between C/EBP-b and Sp1 factor. *Mol. Cell. Biol.*, **14**, 1383–1394.
- Lonergan, M., Dey, A., Becker, K.G., Drew, P.D. and Ozata, K. (1993) A regulatory element in the beta2-microglobulin promoter identified by *in vivo* footprinting. *Mol. Cell. Biol.*, **13**, 6629–6639.
- Major, E.O., Amemiya, K., Tornatore, C.S., Houff, S.A. and Berger, J.R. (1992) Pathogenesis and molecular biology of progressive multifocal leukoencephalopathy, the JC virus-induced demyelinating disease of the human brain. *Clin. Microbiol. Rev.*, **5**, 49–73.
- Mavria, G., Hall, K.T., Jones, R.A. and Blair, G.E. (1998) Transcriptional regulation of MHC class I gene expression in rat oligodendrocytes. *Biochem. J.*, **330**, 155–161.
- Njenga, M.K., Pease, L.R., Wettstein, P., Mak, T. and Rodriguez, M. (1997) Interferon alpha/beta mediates early virus-induced expression of H-2D and H-2K in the central nervous system. *Lab. Invest.*, **77**, 71–84.

- Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
- Perkins,N.D., Edwards,N.L., Duckett,C.S., Agranoff,A.B., Schmid,R.M. and Nabel,G.J. (1993) A cooperative interaction between NF- κ B and Sp1 is required for HIV-1 enhancer activation. *EMBO J.*, **12**, 3551–3558.
- Prestridge,D.S. (1996) SIGNAL SCAN 4.0: additional databases and sequence formats. *Comput. Applic. Biosci.*, **12**, 157–160.
- Quandt,K., Frech,K., Karas,H., Wingender,E. and Werner,T. (1995) MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.*, **23**, 4878–4884.
- Quandt,K., Grote,K. and Werner,T. (1996a) GenomeInspector: A new approach to detect correlation patterns of elements on genomic sequences. *Comput. Applic. Biosci.*, **12**, 405–413.
- Quandt,K., Grote,K. and Werner,T. (1996b) GenomeInspector: Basic software tools for analysis of spatial correlations between genomic structures within megabase sequences. *Genomics*, **33**, 301–304.
- Tanaka,H., Dong,Y., McGuire,J., Okret,S., Poellinger,L., Makino,I. and Gustafsson,J.A. (1993) The glucocorticoid receptor and a putative repressor protein coordinately modulate glucocorticoid responsiveness of the mouse mammary tumor virus promoter in the rat hepatoma cell line m1.19. *J. Biol. Chem.*, **268**, 1854–1859.