362–367 Nucleic Acids Research, 1998, Vol. 26, No. 1

Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL

T. Heinemeyer, E. Wingender*, I. Reuter, H. Hermjakob, A. E. Kel[1], O. V. Kel[1], E. V. Ignatieva[1], E. A. Ananko[1], O. A. Podkolodnaya[1], F. A. Kolpakov[1], N. L. Podkolodny[1] and N. A. Kolchanov[1]

Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-38124 Braunschweig, Germany and [1]Institute of Cytology and Genetics SB RAS, pr. Lavrentyeva-10, 630090 Novosibirsk, Russia

Received September 30, 1997; Accepted October 3, 1997

ABSTRACT

TRANSFAC, TRRD (Transcription Regulatory Region Database) and COMPEL are databases which store information about transcriptional regulation in eukaryotic cells. The three databases provide distinct views on the components involved in transcription: transcription factors and their binding sites and binding profiles (TRANSFAC), the regulatory hierarchy of whole genes (TRRD), and the structural and functional properties of composite elements (COMPEL). The quantitative and qualitative changes of all three databases and connected programs are described. The databases are accessible via WWW: http://transfac.gbf.de/TRANSFAC or http://www.bionet.nsc.ru/TRRD

INTRODUCTION

As the international efforts to sequence complete genomes gain momentum, the availability of software for the interpretation of these sequences is of increasing importance. With regard to the regulatory potential of genomic regions, a bundle of programs has been developed and is partially available on the World Wide Web (WWW), others can be downloaded from several servers freely or upon request or are part of commercial solutions (see refs 1 and 2 for reviews). However, the application of these algorithms strictly depends on the availability of a reliable basis for the deduction of appropriate search patterns which may be individual and experimentally proven genomic sequence elements, consensus strings derived from compiled and aligned elements, or complete profiles such as positional weight matrices giving a statistical description of regulatory signals. However, to achieve a real interpretation of the meaning of a certain genomic sequence, we must be able to exceed the mere recognition of regulatory signals encoded by this sequence towards the biological context in which these signals gain their functionality.

Considering transcriptional regulation, the databases TRANSFAC (developed at GBF Braunschweig since 1988; 3,4), TRRD (Transcription Regulatory Region Database, developed at the Institute of Cytology and Genetics SB RAS since 1993; 4,5) and COMPEL (about Composite Elements, a common effort of both groups; 6) try to match these requirements. Their specific aims and present status as well as their linkages will be described subsequently.

Users are asked to cite this article when publishing results which have been obtained with the database tools described here.

TRANSFAC

Contents

The TRANSFAC database is maintained using a relational database management system (RDBMS). ASCII flat files have been released in March (3.1) and July (3.2), combining the contents of 48 tables of the RDBMS to six flat files (SITE, FACTOR, CLASS, MATRIX, CELL, GENE).

The contents of the six flat file tables are listed in Table 1. The SITE table comprises 4143 sequences with a total of 69 070 nucleotides. It gives information about the localization and sequence of individual regulatory elements within a gene, the method(s) by and the cellular context in which these elements have been identified, and the transcription factors which bind to them. It also contains a number of consensus strings in IUPAC 15 letter code. The interacting transcription factors are given by name and their TRANSFAC accession numbers, the latter as active hyperlink to the corresponding entry of the FACTOR table. Similarly, FACTOR entries display lists of linked SITEs they have been shown to bind to and which have been included in the TRANSFAC SITE table. Descriptions of individual transcription factors in the FACTOR table comprise physico-chemical, local and global structural features, as well as functional properties.

Until release 3.2, entries of the GENE table basically comprised the name of a gene/gene product, the classification number according to the scheme developed by P. Bucher (7), and active links to TRANSFAC, COMPEL and TRRD. Starting with release 3.3, the sites belonging to a gene will be listed explicitly and sorted by their order in 5′-3′ direction (direction of transcription).

* To whom correspondence should be addressed. Tel: +49 531 6181 427; Fax: +49 531 6181 266; Email: ewi@gbf.de

Table 1. Content of the TRANSFAC tables, release 3.2

Table	Entries
SITE	4401
GENE[a]	1095
FACTOR[b]	2166
CLASS	28
MATRIX	260
CELLS	857
METHOD	52
REFERENCE[c]	5462

[a]900 entries of which are connected to TRANSFAC, 426 to TRRD and 243 to both TRANSFAC and TRRD.
[b]Among the FACTOR entries, 1114 are assigned to one of the factor classes.
[c]Total number of articles cited in SITE, FACTOR, CLASS, and MATRIX, giving rise to >15 000 citations.

Transcription factor classification

As for previous releases, most transcription factors whose genes are known have been classified according to the structure of their DNA-binding domain (DBD). The characteristics of these DBD structures are explained in the CLASS table. First for release 3.2, we have extended this classification scheme to a more comprehensive transcription factor classification scheme by defining four so-called 'superclasses', each of them comprising several 'classes' which mainly correspond to the previously defined classes (Table 2). Several sublevels have been defined as well ('families', optional 'subfamilies', 'genera' and 'species') as has been described elsewhere in greater detail (8). The complete classification scheme is available on the TRANSFAC WWW server (http://transfac.gbf.de/TRANSFAC/cl/cl.html). The CL (class) line of FACTOR entries now gives, in addition to the previous CLASS assignment, the numerical classification code which is actively linked to the scheme.

Cross-referencing with external databases

TRANSFAC entries are cross-linked with a number of external databases. Thus, 2936 SITE entries are linked to 1143 entries of the EMBL data library thus creating a total of 4099 cross-links. Similarly, the cross-links between TRANSFAC and other external databases are listed in Figure 1. FlyBase links are provided by M. Ashburner (9).

Since August 1997, the references which are included in the tables SITE, FACTOR, CLASS and MATRIX are linked to PubMed through their bibliographic data.

Integration with external browsing tools

Using the Sequence Retrieval System Version 5 (SRS5, Thure Etzold, EBI; 10), TRANSFAC could be integrated in the network of other sequence databases and sequence analysis tools. SRS5 access to TRANSFAC is available at the URL http://transfac.gbf.de/srs5/ . SRS5 parsers for the TRANSFAC flat files are available via anonymous FTP at ftp://transfac.gbf.de/ . These parsers will be updated for each new release of TRANSFAC.

Properties of the WWW interface

At URL http://transfac.gbf.de/ , the TRANSFAC database can be accessed over the WWW. For users interested in technical information and data structure of TRANSFAC, an on-line documentation is available. Easy access to data is possible using the 'Search' and 'Extended Search' functions of the query tools. The 'Browse' option lists all available entries in large tables. In all search results, hyperlinks which are generated on the fly allow access to cross-linked databases and additional information.

As a new feature, a graphic visualisation of the local characteristics (such as *trans*-activating domain, leucine zipper, phosphorylation site) listed in the FACTOR feature table has been included. Starting with release 3.3 and the new format of the GENE table, the same algorithm will enable the distribution of individual transcription factor binding sites within a gene to be displayed graphically.

Connected programs

For scanning new DNA sequences for potential regulatory elements, the programs PatternSearch 1.1 (11) and MatInspector 2.1 (12) can be used at the TRANSFAC WWW site. As published previously, PatternSearch uses the sequence data of the SITE table, whereas the library connected with MatInspector has been selected from the MATRIX table of the TRANSFAC database. FastM, developed by T. Werner and his co-workers (13), is a program which allows scanning of new DNA sequences or the EMBL/GenBank databases with a combination of two matrices out of the matrix library. It thus enables complex analyses of regulatory genomic regions.

Table 2. Classification of transcription factors: superclasses and classes

Superclass:	Basic domain	Zinc finger domain	Helix–turn–helix domains	β-Scaffold with minor groove contacts[a]
Classes:	bZIP	Cys_2Cys_2	Homeo	REL
	bHLH	Cys_2His_2	Winged helix	MADS
	bHLH-ZIP	Cys_6 clusters	Trp-clusters	TBP
			TEA	HMG

[a]This superclass is non-transitive, since some of its class members may exert minor groove contacts but do not contain a β-scaffold within their DNA-binding domain (such as HMG proteins), or vice versa.

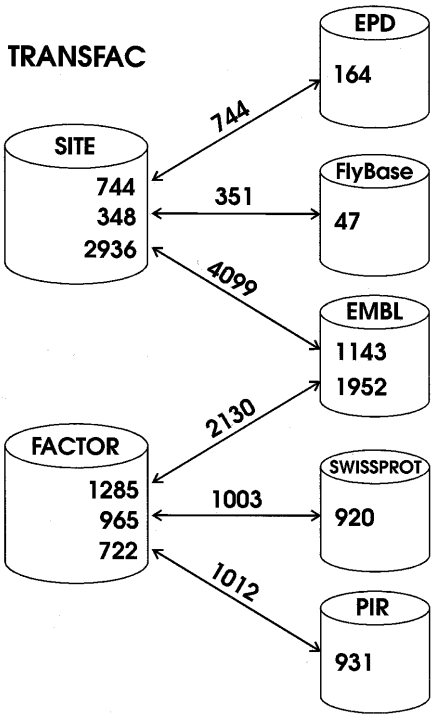

Figure 1. Cross-references between the TRANSFAC tables SITE and FACTOR and several external databases. Depicted are the number of entries of either TRANSFAC table which are connected to one of the external databases, the number of linked external database entries, and the number of links between these entries.

TRRD (TRANSCRIPTION REGULATORY REGION DATABASE)

Data structure

The TRRD database describes the structure of eukaryotic transcription regulatory regions (promoters, enhancers) and specific patterns of gene expression (5,14). The database schema represents the hierarchical structure of regulatory regions. The following regulatory units are described in the database: (i) *cis*-acting elements that provide interaction of transcription factors with DNA (15); (ii) composite elements that contain closely located *cis*-elements working in closely interrelated manner (see below about COMPEL database) (6); (iii) promoters, enhancers and silencers that usually occupy regions of several hundred base pairs and include a number of single and composite elements; (iv) extended transcription regulatory regions on 5′ and 3′ ends of genes or located within introns, that include the above-mentioned regulatory units; (v) the integral transcription regulatory system of a single gene.

Along with the structure of a regulatory region of a gene, the description of its expression pattern is an essential part of TRRD. The new release of TRRD 4.0 (February 1998) will comprise a new structure for describing the conditions under which a gene is expressed. These expression profiles consist of sets of vectors of expression data comprising information about: (i) cell cycle stage; (ii) developmental stage; (iii) cell type, tissue, organ; (iv) influence of signals external to the cell: chemicals, cytokines, hormones, growth factors, vitamins, etc.; and the level of gene expression under these conditions in qualitative terms.

The new schema allows more precise presentation of information about gene expression and in machine-readable form. One of the major advantages of release 4.0 is the presence of internal links between structural regulatory units and vectors of expression data. These links are of great importance for understanding the molecular mechanisms of the transcriptional regulation a gene is subjected to. Signal transduction pathways are indicated by these links that point to the *cis*-element (and binding factor) on one side and to the corresponding signal and its influence on gene expression described in the vector of expression data on the other side. Two examples of expression pattern descriptions are given in Table 3.

Content

The content of TRRD release 3.5 is summarized in Table 4. Starting from October 1997, the TRRD database is maintained using a commercial relational database management system. The relational TRRD consists of 24 tables linked by 1:n and m:n relations. Periodically, we release TRRD into one ASCII flat file combining contents of the relation tables.

Table 3. Representation of gene expression patterns in TRRD[a]

Field descriptor[b]	Content of the fields	Human $^{A}\gamma$ globin gene	Human cyclin A gene
RE	accession number of vector of expression data	G000261.001	G001155.011
RY	stage of cell cycle		G1/S boundary
RD	stage of development	fetus	
RO	organ	liver	
RU	tissue		
RN	cell type or cell line	definitive erythroid cells	primary fibroblasts
RI	signal		cAMP→PKA
FF	effect of the signal		induction
RL	qualitative level of expression	maximal level	
RS	link to site accession number		1850

[a]As an example, only one vector of expression data is shown for each of two genes.
[b]Additional fields will indicate whether the vector of expression data refers to either mRNA or protein, will give links to TRRD promoter accession number, comments, and reference to the original publication(s).

Table 4. Content of the TRRD database, release 3.5

Table	Number of entries
T_GENE	426
T_PROMOTER	596
T_SITE	2147
T_REFERENCE	1759

Table 5. Functional gene systems described in WWWTRRD

Functional gene system	Number of entries in TRRD	Ref.
Interferon-inducible genes	60	21
Erythroid-specific genes	33	22
Genes of lipid metabolism	50	16
Glucocorticoid controlled genes	35	23
Cell cycle dependent genes	20	24
Muscle-specific genes	25	14

There are 426 gene entries in the TRRD 3.5 compiling information which has been extracted from 1759 papers. Genes of the following organisms have been described in the database: human (179 entries); mouse (114 entries); rat (69 entries); chicken (27 entries); viruses (13 entries); frog (7 entries); rabbit (7 entries); other organisms (10 entries).

Information hypertext system WWWTRRD

By means of World Wide Web (WWW) we have developed an information hypertext system WWWTRRD that is to provide Internet access to the TRRD entries converted from TRRD flat file into html format. Retrieval of individual entries can be done by browsing the whole database or by an elaborate search routine that enables searching for gene or site entries by name of gene, name of transcription factor, by keyword or by regulation landmarks. An exhaustive text search over the whole database is also available. The entries retrieved are supplied by active links to the partner databases TRANSFAC and COMPEL and external databases such as EMBL or EPD. In addition, the WWWTRRD provides hypertext links to textual and graphical presentation of functional gene systems that are listed in Table 5. Group of genes involved in the lipid metabolism may serve as an interesting example of functional gene system (16). The proteins encoded by these genes participate in a variety of processes essential for normal lipid homeostasis. This group contains genes of transport proteins (apolipoproteins); enzymes of both lipid biosynthesis and degradation; transcription factors involved in the regulation of these genes.

Thus, the current release of WWWTRRD contains descriptions of the main types of transcription regulation in eukaryotic cells: tissue-specific regulation (e.g., of muscle-specific genes); gene responses to external signals (interferon-inducible genes and glucocorticoid-controlled genes); cell cycle-dependent gene regulation and gene regulation in development (erythroid-specific genes).

Visualization tools

We have developed a Java based tool for retrieving and visualizing the structure of regulatory regions as it is represented in TRRD. All transcription factor binding sites of a TRRD entry can be graphically displayed showing their relative localization, names of transcription factors, sequences (optionally) and some other information. The tool provides a zoom function for more detailed representation of different regulatory sub-regions of a gene. These routines have been implemented using a previously developed C++ and Java object library termed MGL (Molecular Genetic Language) (http://www.bionet.nsc.ru/MGL/) (17). It provides the instruments for managing data from molecular genetic databases and their graphic visualization over the Internet. A stand-alone version of the visualization program running under Windows provides additional options. For example, it can generate a Windows metafile suitable for making presentations. The visualization tools are available on the TRRD WWW server (http://www.bionet.nsc.ru/systems/Mgl/TRRD_Viewer.html). They enable the creation of overviews of the organization of different types and variants of transcription regulatory regions. An automatically generated diagram for the structure of the 5′-regulatory region of the glycoprotein hormone α-subunit gene (α-*GH*) shows that there are at least three overlapping complex regulatory units within a very short region (~150 bp; Fig. 2): the promoter, a placenta-specific enhancer and a steroid-dependent negative regulatory element, which together comprise 11 transcription factor binding sites. This example illustrates the complex hierarchical structure of transcription regulatory regions of some eukaryotic genes.

COMPEL

Data structure

COMPEL is a database on composite regulatory elements (CE) of vertebrate genes. Such elements are located in transcription regulatory regions and contain two closely situated binding sites for different transcription factors. They are essential for transcription regulation in a highly specific manner due to specific DNA–protein and protein–protein interactions (6).

Two main types of composite elements are described in COMPEL. Composite elements of synergistic type contain sites for transcription factors that simultaneously bind to DNA and synergistically activate transcription. For composite elements of antagonistic type, two transcription factors influence transcription in opposite directions.

Contents

COMPEL is distributed in a single flat file. It is electronically accessible via Internet by anonymous ftp (ftp://transfac.gbf.de/pub/databases/compel/) or by browsing the database through the Web (http://www.bionet.nsc.ru/COMPEL/ and via active links from TRANSFAC). Some new fields have been added to the database schema in the recent COMPEL release (COMPEL 2.1). In the flat file, these fields are represented as follows: the **TT** line gives information about cell types in the case of tissue-specific composite elements only; the lines **T1/T2** and **C1/C2** provide links to TRANSFAC FACTOR and CELL tables for the transcription factors that bind to the composite element; the **FU** line classifies functional types of the composite element (Table 6). This line has been introduced since in release 2.1, most of the

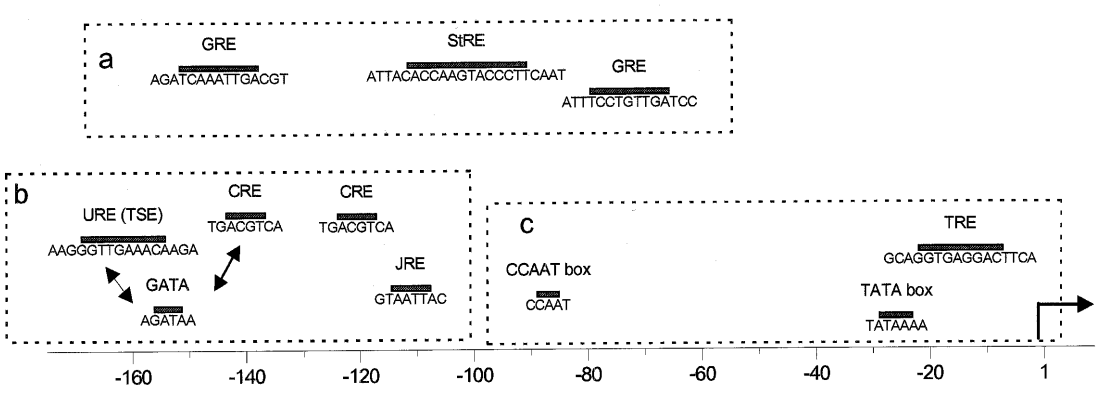

Figure 2. A fragment of the 5′ regulatory region of glycoprotein hormone α-subunit gene, α-*GH* (G000271, TRRD accession number). This 5′-regulatory region comprises: (**a**) a steroid-dependent negative regulatory element with glucocorticoid regulatory elements (GRE) and a steroid regulatory element (StRE); (**b**) a placenta-specific enhancer (URE, upstream regulatory element; TSE, trophoblast-specific element; GATA, GATA-factor binding site; CRE, cAMP-responsive element; JRE, junctional regulatory element); (**c**) a proximal promoter containing a CCAAT and a TATA box as well as a TRE (thyroid hormone regulatory element). Region b contains two composite elements which are indicated by arrows (URE/GATA and GATA/CRE).

CEs have been classified according to their function (18). There are (i) 67 composite elements that confer tissue-specific regulation among that liver (20 entries), T, B- and myeloid cells (30 entries), muscle (five entries), pituitary (six entries), other (six entries); (ii) 33 composite elements mediating induction by steroid hormones (two entries), interleukins (two entries), other molecular signals (14 entries), acute-phase and immune response (12 entries); (iii) three composite elements regulating cell-cycle dependent gene expression; (iv) six composite elements involved in developmental control.

Connected programs

On the base of the sequence and structural information stored in the COMPEL database we are creating a number of new methods for recognition of different CEs in genomic sequences. Being very specific transcription regulators, CEs are sensitive indicators of the regulatory function of the sequence under analysis. Therefore, a new program for searching potential composite elements of definite types is available now at the COMPEL and TRANSFAC WWW sites. We have applied a statistical approach based on fuzzy calculation that permits the application of structural features (distance between site pairs and their mutual orientation as retrieved from the database; see Fig. 3) and calculated binding affinity (19,20) of the CE for the corresponding transcription factors. The distribution of these structural features was computed on the basis of CE collected in the COMPEL database and individual sites collected in TRANSFAC and TRRD. Currently, search routines for two types of CE are available: NFATp/AP-1 composite elements (earlier called NFAT sites) for T-cell specific genes and NF-κB/NF-IL6 composite elements for acute-phase genes. We tested the program for recognition of NFATp/AP-1 composite elements in a large set of T-cell specific gene sequences. The test revealed a high specificity of the method to the regulatory regions of genes expressed in T-cells (Kel *et al.*, manuscript in preparation). The frequency of potential CEs found within regulatory regions and introns was 3–4 times higher than within coding regions. Clusters of potential NFATp/AP-1 CEs have been specifically found within promoter regions. This method may be used for searching potential composite elements of this type as well as other types collected in COMPEL.

Table 6. Functional types of composite elements[a]

Functional type	Functional property of F1	Functional property of F2	Function of the composite element F1/F2	No. of CE
T01	Tissue-specific	Inducible by signal ***A***[a]	Tissue-specific induction by signal ***A***	15
T02	Tissue-specific	Constitutive ubiquitous	Tissue-specific regulation triggered by a ubiquitous factor	14
T03	Inducible by signal ***A***	Constitutive ubiquitous	Inducible regulation for which a constitutive factor is essential	10
T04	Inducible by signal ***A***	Inducible by signal ***B***[a]	Cross-coupling of different signal transduction pathways	25
T05	Tissue-specific factor inducible by signal ***A***	Inducible by signal ***B***	Tissue-specific cross-coupling of different signal transduction pathways	9
T06	Cell-cycle specific	Cell-cycle independent	Cell-cycle specific regulation for which a cell cycle-independent factor is essential	1

[a]Composite elements confer combinatorial regulation of a gene by two different transcription factors (F1 and F2) binding to their individual sites.
[b]Signals ***A*** and ***B*** are external signals (hormones, cytokines, growth factors, etc.) causing gene induction or repression.

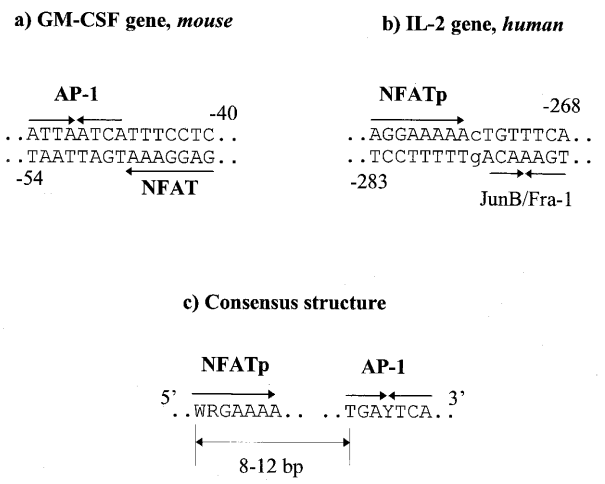

Figure 3. NFATp/AP-1 composite element specific for activated T cells. (**a**, **b**) Two examples from the COMPEL database; (**c**) consensus structure derived from COMPEL entries and used for recognition of the CE of this type. The sometimes imperfect palindromic AP-1 site is always located on the 3′-side of the asymmetric NFATp site WRGAAAA.

FEDERATION OF TRANSFAC, TRRD AND COMPEL

Some steps towards a further federation of the three databases described here have been taken. First of all, a TRANSFAC WWW mirror site has been established in Novosibirsk (http://www.bionet.nsc.ru/transfac/) and a TRRD WWW mirror site in Braunschweig (http://transfac.gbf.de/trrd/). A joint routine for extended search through all three databases has been simultaneously developed and is available now on both servers. Moreover, efforts are being made to include TRRD and COMPEL into SRS5 enabling complex queries over these three and additional databases. SRS5 will make use of the TRANSFAC/TRRD/COMPEL links as they appear in the jointly maintained GENE table. These joint retrieval capabilities are reinforced by providing additional links between, e.g., TRANSFAC FACTOR and TRRD SITE tables.

ACKNOWLEDGEMENTS

Different parts of this work were funded by the European Commission (BIO4 CT950226), by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (project no. X224.6), by the Russian Ministry of Sciences, the Russian National Program ‘Human Genome’ and the Russian Fundamental Research Foundation (grants: 94-04-13241-a, 94-04-12757-a, 97-04-49740, 96-04-50006), by the Siberian Branch of Russian Academy of Sciences, as well as by the North Atlantic Treaty Organization (grant no. 951149).

REFERENCES

1 Frech, K., Quandt, K. and Werner, T. (1997) *Trends Biochem. Sci.* **22**, 103–104.
2 Frech, K., Quandt, K. and Werner, T. (1997) *Comput. Appl. Biosci.* **13**, 89–97.
3 Wingender, E., Dietze, P., Karas, H. and Knüppel, R. (1996) *Nucleic Acids Res.* **24**, 238–241.
4 Wingender, E., Kel, A. E., Kel, O. V., Karas, H., Heinemeyer, T., Dietze, P., Knüppel, R., Romaschenko, A. G. and Kolchanov, N. A. (1997) *Nucleic Acids Res.* **25**, 265–268.
5 Kel, O. V., Romachenko, A. G., Kel, A. E., Naumochkin, A. N. and Kolchanov, N. A. (1995) Proceedings of the 28th Annual Hawaii International Conference on System Sciences [HICSS], Biotechnology Computing, IEE Computer Society Press, Los Alamitos, CA **5**, 42–51.
6 Kel, O. V., Romaschenko, A. G., Kel, A. E., Wingender, E. and Kolchanov, N. A. (1995) *Nucleic Acids Res.* **23**, 4097–4103.
7 Bucher, P. and Trifonov, E. M. (1986) *Nucleic Acids Res.* **14**, 10009–10026.
8 Wingender, E. (1997) *Mol. Biol.* **31**, 483–497.
9 Gelbart, W. M., Crosby, M., Matthews, B., Rindone, W. P., Chillemi, J., Russo Twombly, S., Emmert, D., Ashburner, M., Drysdale, R. A., Whitfield, E., *et al.* (1997) *Nucleic Acids Res.* **25**, 63–66. [See also this issue *Nucleic Acids Res.* (1998) **26**, 85–88.]
10 Etzold, T., Ulyanov, A. and Argos, P. (1996) *Methods Enzymol.* **266**, 114–128.
11 Wingender, E., Karas, H. and Knüppel, R. (1996) Pacific Symposium on Biocomputing ‘97 (PSB′97), R. B. Altman, A. K. Dunker, L. Hunter, T. E. Klein (eds). World Scientific, Singapore, New Jersey, London, Hong Kong. pp. 477–485.
12 Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884.
13 Frech, K., Danescu-Mayer, J. and Werner, T. (1997) *J. Mol. Biol.* **270**, 674–687.
14 Kel, A. E., Kolchanov, N. A., Kel, O. V., Romaschenko, A. G., Ananko, E. A., Ignatyeva, E. V., Merkulova, T. I., Podkolodnaya, O. A., Stepanenko, I. L., Kochetov, A. V., *et al.* (1997) *Mol. Biol.* **31**, 521–530.
15 Wingender, E. (1993) *Gene Regulation in Eukaryotes*. VCH., Weinheim.
16 Ignatieva, E. V., Merkulova, T. I., Vishnevsky, O. V. and Kel, A. E. (1997) *Mol. Biol.* **31**, 575–591.
17 Kolpakov, P. A. and Babenko, V. N. (1997) *Mol. Biol.* **31**, 540–547.
18 Kel, V. O., Kel, A. E., Romaschenko, A. G., Wingender, E. and Kolchanov, N. A. (1997) *Mol. Biol.* **31**, 498–512.
19 Kel, A. E., Kondrakhin, Y. V., Kolpakov, P. A., Kel, O. V., Romashenko, A. G., Wingender, E., Milanesi, L. and Kolchanov, N. A. (1995) *Proc. Third Int. Conf. Intell. Syst. Mol. Biol.* **3**, 197–205.
20 Berg, O. G. and von Hippel, P. H. (1987) *J. Mol. Biol.* **193**, 723–750.
21 Ananko, E. A., Bazhan, S. I., Belova, O. E. and Kel, A. E. (1997) *Mol. Biol.* **31**, 592–604.
22 Podkolodnaya, O. A. and Stepanenko, I. L. (1997) *Mol. Biol.* **31**, 562–574.
23 Merkulova, T. I., Merkulov, V. M. and Mitina, R. L. (1997) *Mol. Biol.* **31**, 605–615.
24 Kel, O. V. and Kel, A. E. (1997) *Mol. Biol.* **31**, 548–561.