332–334 *Nucleic Acids Research, 2002, Vol. 30, No. 1*

TRANSCompel®: a database on composite regulatory elements in eukaryotic genes

Olga V. Kel-Margoulis[1,2,*], Alexander E. Kel[1,2], Ingmar Reuter[1], Igor V. Deineko[2] and Edgar Wingender[1,3]

[1]BIOBASE GmbH, Halchtersche Strasse 33, D-38304 Wolfenbüttel, Germany, [2]Institute of Cytology and Genetics SB RAN, 10 Lavrentyev pr., 630090, Novosibirsk, Russia and [3]AG Bioinformatik, Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-38124 Braunschweig, Germany

Received September 24, 2001; Accepted October 1, 2001

ABSTRACT

Originating from COMPEL, the TRANSCompel® database emphasizes the key role of specific interactions between transcription factors binding to their target sites providing specific features of gene regulation in a particular cellular content. Composite regulatory elements contain two closely situated binding sites for distinct transcription factors and represent minimal functional units providing combinatorial transcriptional regulation. Both specific factor–DNA and factor–factor interactions contribute to the function of composite elements (CEs). Information about the structure of known CEs and specific gene regulation achieved through such CEs appears to be extremely useful for promoter prediction, for gene function prediction and for applied gene engineering as well. Each database entry corresponds to an individual CE within a particular gene and contains information about two binding sites, two corresponding transcription factors and experiments confirming cooperative action between transcription factors. The COMPEL database, equipped with the search and browse tools, is available at http://www.gene-regulation.com/pub/databases.html#transcompel. Moreover, we have developed the program CATCH™ for searching potential CEs in DNA sequences. It is freely available as CompelPatternSearch at http://compel.bionet.nsc.ru/FunSite/CompelPatternSearch.html.

INTRODUCTION

Based on known examples, we define a composite element (CE) as a combination of transcription factor binding sites which, as such, and through protein–protein interactions between the transcription factors involved, provides a known regulatory feature (1–3). Thus, interacting factors may differ by the structure of DNA-binding, activation, oligomerization and other domains. Along with structural differences, functional properties of the transcription factors, and hence their specific contribution to the transcription regulation, may vary significantly. Co-operative action of the transcription factors within the CEs results in a new highly specific pattern of gene transcription that cannot be provided by the involved factors separately. CEs are structural–functional units that provide cross-coupling of gene regulatory pathways and, in particular, cross-coupling of signal transduction pathways (3).

There are two main types of CEs: synergistic and antagonistic. In synergistic CEs, simultaneous interactions of two factors with closely situated target sites results in a non-additive high level of a transcriptional activation. Within an antagonistic CE, two factors interfere with each other.

The TRANSCompel® database originated from the COMPEL database (1–8). Several new features have been introduced to improve representation of the composite regulatory elements, and content of the database has been significantly increased.

CONTENT OF THE DATABASE

During the last 2 years, the number of CEs described in the database has increased by ~30%. Two freely available versions of the database have been released, versions 4.4 and 6.0 (Table 1). Distribution of genes among species is as follows: 73 human, 40 mouse, 22 rat, 8 chick and 19 others.

Among recent entries there are CEs containing binding sites for the following transcription factors: Smads, 14 entries;

Table1. Content of the TRANSCompel® releases 4.4 and 6.0

	Number of entries	
	Release 4.4	Release 6.0
CEs	202	256
Genes	131	162
Links to EMBL	171	216
Transcription factors linked to TRANSFAC	171	216
Interactions	639	948
Evidences	602	846
References	207	281

*To whom correspondence should be addressed at present address: BIOBASE GmbH, Halchtersche Strasse 33, D-38304 Wolfenbuettel, Germany. Tel: +49 5331 858426; Fax: +49 5331 858470; Email: oke@biobase.de

steroidogenic factor 1, 11 entries; SREBP, 8 entries; AML/PEBP, 10 entries; PU.1, 19 entries; c-Ets-1,2, 39 entries.

We have extended a description of the experimental evidences confirming factor cooperation within CEs. In a separate field we provide information about protein domains involved in the functional co-operation and/or physical interactions between transcription factors.

Another new field in the database is devoted to the information about confirmed or suggested molecular mechanisms of cooperation between transcription factors.

CLASSIFICATION OF THE CEs

Starting with TRANSCompel 4.4, functional classification of the CEs is one of the important features of the database. For that, we have applied a classification of CEs according to the specific transcriptional regulation they provide due to co-operative action of transcriptional factors binding to their target sites (3). In release 6.0, 200 CEs have been classified into the following five main groups.

1. 'Inducible/inducible': 81 CEs are formed by binding sites for two inducible factors providing cross-coupling of signal transduction pathways. To this group, we have classified, for instance, 14 CEs within different mammalian genes consisting of binding sites for Ets and AP-1 transcription factors, providing cross-coupling of Ras/Raf- and PKC-dependent signalling pathways.
2. 'Inducible/constitutive': 39 CEs are composed of binding sites for an inducible and a constitutive ubiquitous factor providing some additional features of the inducible regulation. For instance, within Smad/TEF3 and Smad/Sp1 CEs, Smads are inducible by TGF-β signalling, and TEF3 and Sp1 are constitutive transcription factors. Thus, constitutive factors took an essential part in the regulation by TGF-β.
3. 'Tissue-restricted/ubiquitous': 30 CEs are formed by binding sites for a tissue-enriched and a constitutive ubiquitous factor providing some additional features of the tissue-specific transcriptional regulation. For example, steroidogenic cell-restricted transcription factor SF-1 and ubiquitous Sp1 are known to synergistically activate gene expression in steroidogenic cells.
4. 'Inducible/tissue-restricted': 27 CEs are constituted by binding sites for a tissue-enriched and an inducible factor providing tissue-specific responses to inducing signals. This group may be illustrated by Pit1/AP-1, Pit1/Ets CEs, where Pit1 is a pituitary-restricted transcription factor, whereas AP-1 and Ets are ubiquitous inducible factors. These CEs provide pituitary-restricted induction.
5. 'Tissue-restricted/tissue-restricted': 23 CEs comprise binding sites for two tissue-enriched factors providing particular aspects of tissue-specific regulation. For example, Ptx-1 is expressed in all pituitary lineages and SF-1 in pituitary gonadotropes only. CEs formed by binding sites for these two factors regulate expression of genes exclusively in gonadotropes where both factors are present.

AVAILABILITY

Being maintained internally as a relational database, TRANSCompel® is distributed as a single ASCII flat file. Public versions 4.4 and 6.0 are available at http://www.gene-regulation.com/pub/databases.html#transcompel; the current professional version can be obtained from BIOBASE (http://www.biobase.de; four updates per year). Release COMPEL 3.0 can be found at http://compel.bionet.nsc.ru/. A detailed description of the fields is given in the database documentation. Web-based search and browse options are available.

CONNECTED PROGRAM

The program CATCH™ for searching potential CEs in DNA sequences has been developed. A preliminary version of this program, CompelPatternSearch, is publicly available at http://compel.bionet.nsc.ru/FunSite/CompelPatternSearch.html. The current version can be obtained from BIOBASE. A sequence under study is scanned by this program using all CEs collected in the TRANSCompel® database as individual search patterns. Several parameters are available, restricting the search: maximal mismatches in the cores of site1 and site2 comprising the CEs, maximal variation of the distance between two sites, and composite score cut-off value (6,9). The composite score reflects how well the match coincides with the known examples of the CE in TRANSCompel®. This scoring function takes into account the number of mismatches in both sites and the distance between them. All found matches are directly linked to the TRANSCompel® entries containing the corresponding CEs.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

Authors are grateful to N. A. Kolchanov and A. G. Romaschenko for valuable discussions on the concept of composite regulatory elements. Part of this work has been funded by Volkswagen-Stiftung (I/75941).

REFERENCES

1. 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.
2. Kel,O.V., Romaschenko,A.G., Kel,A.E., Wingender,E. and Kolchanov,N.A. (1997) Composite regulatory elements: classification and description in the COMPEL database. *Mol. Biol.*, **31**, 498–512.
3. Kel-Margoulis,O.V., Romaschenko,A.G., Kolchanov,N.A., Wingender,E. and Kel,A.E. (2000) COMPEL: a database on composite regulatory elements providing combinatorial transcriptional regulation. *Nucleic Acids Res.*, **28**, 311–315.
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) TRANSFAC, TRRD and COMPEL: towards a federated database system on transcriptional regulation. *Nucleic Acids Res.*, **25**, 265–268.
5. Heinemeyer,T., Wingender,E., Reuter,I., Hermjakob,H., Kel,A.E., Kel,O.V., Ignatieva,E.V., Ananko,E.A., Podkolodnaya,O.A., Kolpakov,F.A. *et al.* (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.*, **26**, 362–367.
6. Kel-Margoulis,O.V., Kel,A.E., Frisch,M., Romaschenko,A.G., Kolchanov,N.A. and Wingender,E. (1998) COMPEL a database on composite regulatory elements. *Proceedings of the First International Conference on Bioinformatics of Genome Regulation and Structure*. ICG, Novosibirsk, Vol. 1, pp. 54–57.

7. Heinemeyer,T., Chen,X., Karas,H., Kel,A.E., Kel,O.V., Liebich,I., Meinhardt,T., Reuter,I., Schacherer,F. and Wingender,E. (1999) Expanding of the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.*, **27**, 318–322.
8. Kel-Margoulis,O.V., Romaschenko,A.G., Deineko,I.V., Kolchanov,N.A., Wingender,E. and Kel,A.E. (2000) Database on composite regulatory elements in eukaryotic genes (COMPEL). *Proceedings of the Second International Conference on Bioinformatics of Genome Regulation and Structure*. ICG, Novosibirsk, Vol. 1, pp. 45–48.
9. Kel,A., Kel-Margoulis,O., Babenko,V. and Wingender,E. (1999) Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells. *J. Mol. Biol.*, **288**, 353–376.