LOCUS CONTROL REGIONS: DESCRIPTION IN A DATABASE

*Podkolodnaya O.A. and Levitsky V.G.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: opodkol@bionet.nsc.ru
*Corresponding author

Keywords: transcription regulation, regulatory gene regions, LCR, databases

Resume

A database was created on the structure-functional organization of the locus-control regions (LCRs) in eukaryotic genomes. The database contains two sections including (i) formalized descriptions of LCRs and (ii) hypertextual description of the LCR structure and function. Both sections are linked by hypertexts between each other and with external databases such as TRRD, SWISS-PROT, EMBL, and Medline.

Availability:

The database is maintained under SRS and is available by the Internet via http://wwwmgs.bionet.nsc.ru/mgs/dbases/lcr/

Introduction

In the middle of 80-ties, a regulatory element of the novel type was detected that provides coordinated stage- and tissue-specific regulation of human β-globin locus genes transcription [Grosveld et al., 1987]. This element was named as the Locus Control Region (LCR). Later on, analogous regulatory regions were found in the other gene loci.

Some LCRs may differ significantly by the content of the constituent elements; by the content of genes controlled by these LCRs; and by the LCR disposition respectively regulated genes. LCR may regulate both expression of the whole gene cluster (e.g., human β-globin ) and the separate gene (e.g., human adenosine deaminase). LCRs regulating the clusters of α- and β-globin genes in mammals, chicken lysozyme, mouse glycophorin are located in the 5’ region. On the contrary, LCR regulating growth hormone and human CD2 genes are located in the 3’ region, whereas for adenosine deaminase gene, the corresponding regulatory region is situated in the first intron. Although it was found that LCR might regulate not only the whole gene cluster but also the separate gene, the traditional name of this regulatory element is not changed. Formally, as a locus controlling region (LCR) we denote the DNA fragment (or the grouped DNA fragments), such that in transgene experiments it provides the high level of tissue-specific expression of the linked gene integrated into transgene construction, proportionally the copy number of this construction and independently the place of its insertion in genome [Li et al., 1999; Grosveld, 1999].

LCRs provide coordinated tissue- and stage-specific expression of the genes entering the regulated cluster. Currently, the data are available on more than 30 LCRs and LCR-like elements. As a rule, the LCR structure is complex. Its particular elements are marked by sites with the high sensitivity to DNasease I (HSS), which could be tissue-specific or expressed in all the tissues [Grosveld et al., 1987; Chung et al., 1993]. Functionally, the elements entering different LCRs are represented by (a) boundary elements or insulators [Abruzzo and Reitman 1994; Jackson et al., 1996], (b) the chromatin domain opening elements [Ortiz et al., 1997; Festenstein et al., 1996], (c) facilitator elements [Aronow et al., 1995], and (d) enhancers. Enhancers are found almost in all the LCRs. Generally, the most part of the LCR enhancer activity is produced by the core sequences with the length of about 200 bp, where transcription factor binding sites are located. A special type of enhancers, in addition to the classic ones, was found among LCRs. Activity of these enhancers depends upon their orientation and is revealed during their integration into the chromatin, but is not detected in the experiments with the transient transfection [Terajima et al., 1995].

Each of the elements listed above makes its own impact into transcription regulation of the corresponding gene cluster, but only interaction of the whole integrity of elements provides the proper LCR functioning.

Results and discussion

In order to describe structure-functional LCR features, we have developed a special format and created a specialized database LCR-TRRD. Informational fields of this database produce the knowledge on the structure of the gene locus under regulation; the structure of the LCR itself; its individual components; along with information on experimental evidence supporting these facts. The database is linked by the hypertexts to the
table of genes (TRRDGENES) in the TRRD database [Kolchanov et al., 2000], and to the databases EMBL, SWISS-PROT, and MEDLINE.

In Figure 1, the schemes of some LCRs are given together with the gene loci regulated by them and described in the LCR-TRRD database.

Each LCR is supplied by an enhancer as a mandatory element. It was proved that enhancers in the LCRs are tissue-specific, as a rule, and, hence, exactly they are responsible for the LCR's tissue-specificity in the most cases. Tissue-specificity of enhancers themselves is produced by the presence of the tissue-specific transcription factors. So, for the erythroid-specific LCRs, the sites binding the factors NF-E2/AP1 and GATA-1 are typical. In enhancer of the lysozyme gene cluster LCR, the binding site for the myeloid-specific factor PU.1 is located, whereas within the enhancer of the growth hormone gene, the pituitary-specific factor Pit-1 was found.

The LCR-regulated transcription in T-lymphocytes (T-cell receptor α/δ, adenosine deaminase, CD2) is marked by the presence of SOX4 and LEF-1/TCF-1 transcription factors binding sites. It is necessary to note that both these factors are specifically expressed in T-lymphocytes, being referred to the HMG protein family, which bind to DNA and cause its bending. In the LCR-TRRD database, there are links to the other sections of TRRD database containing detailed description of transcription factor binding sites found in the corresponding LCR elements.

Enhancers are the best studied LCR elements, but they are not the only ones. The principles of organization and functioning of the other regulatory structures are not completely studied yet. It is supposed that the LCR's action is determined by its ability to establish and maintain the open chromatin domain [Dillon and Grosveld, 1993]. Relying on the fact that nucleosome organization of DNA is laid in the basic level of its packaging, we have performed the analysis and search in the locus control DNA sequences of the features responsible for the nucleosome binding site formation. For the nucleosome site recognition, we have used the computer program [Levitsky and Katokhin, 2000], based on the discriminant analysis method. If the recognition function value tends to +1, this fact evidences about the better ability of the sequence analyzed to nucleosome positioning. In Figure 2, the results of such analysis of the DNA sequence of the chicken β-globin gene cluster LCR are shown. It can be seen that in the HSS2 and HSS1 containing DNA regions, the deviation of the recognition function values from the basal level is detected, this pointing out to decrease of nucleosome potential in these regions. The HSSs shown in this Figure are the typical tissue-specific hypersensitive sites. They can be found in all the erythroid cell lines and are absent in the non-erythroid ones. Usually, hypersensitive to DNAase I sites are associated with the active regulatory regions that lack canonical nucleosomes. In this case, a nucleosome may be either missing or be partially destroyed. Probably, the contextual DNA features of these regulatory regions...
may be responsible for decrease in nucleosome binding potential, thus, favoring to tissue-specific pattern of transcription regulation.

Acknowledgments

The work was partially supported by the Russian Human Genome Program, Ministry of Science and Technology of Russian Federation, Russian Foundation for Basic Research (grants Nos. 98-04-49479, 98-07-91078, 99-07-90203, 00-07-90337, 00-04-49229). The authors are grateful to Belova O.E. for the database filling and helpful discussions; to N.L.Podkolodny, and D.A. Grigorovich for SRS and software installment and support; and to G.V. Orlova for translation of the paper into English.

References