Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 364 (2007) 60–66
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Identification of tissue-specific DNA–protein binding sites by means of two-dimensional electrophoretic mobility shift assay display

Igor P. Chernov, Kira A. Timchenko, Sergey B. Akopov, Lev G. Nikolaev *, Eugene D. Sverdlov

Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia

Received 11 December 2006
Available online 2 February 2007

Abstract

We developed a technique of differential electrophoretic mobility shift assay (EMSA) display allowing identification of tissue-specific protein-binding sites within long genomic sequences. Using this approach, we identified 10 cell type-specific protein-binding sites (protein target sites [PTSs]) within a 137-kb human chromosome 19 region. In general, tissue-specific binding of proteins from different nuclear extracts by individual PTSs did not follow the all-or-nothing principle. Most often, PTS–protein complexes were formed in all cases, but they were different for different nuclear extracts used.

Keywords: DNA–protein interactions; Electrophoretic mobility shift assay (EMSA); Differential display

The publication of the draft human genome sequence [1,2] and sequences of other metazoan genomes [3] enabled positioning and analysis of various genome functional elements, first of all coding sequences [4,5]. However, complete functional annotation of sequenced eukaryotic genomes cannot be limited to RNA coding sequences and should also include positioning of all non-coding regulatory elements. Unfortunately, experimental data on genomic positions of the multitude of regulatory sequences, such as enhancers, promoters, transcription terminators, and replication origins, are very limited, especially at the whole genome level. In general, most genomic regulatory elements (e.g., enhancers) are gene specific and therefore have different structures, in particular because genes or group of genes usually are regulated independently. Therefore, prediction of these elements by computational methods is difficult and not always reliable.

Most of the genomic regulatory elements that determine structural and functional properties of metazoan organisms are known to act tissue and/or cell specifically [6]. The tissue specificity is in turn determined by binding of tissue-specific sets of proteins, in particular transcription factors, which is one of the main mechanisms of gene expression regulation [7]. The absence or presence of a specific transcription factor complex with the corresponding *cis* element might be the result of the absence or presence of this transcription factor in the cell, posttranslational modification of the factor facilitating or preventing the binding, or modification (e.g., methylation) of the *cis* element itself. Therefore, DNA sequences capable of tissue-specific binding of regulatory proteins play an important role in genomic regulatory networks, and their large-scale identification, cloning, and mapping is of great importance for the understanding of functional activity of genomes. Taking into account a huge quantity of *cis* regulatory elements in genomes, this task cannot be achieved using conventional approaches.

To identify tissue-specific protein-binding sites within long genomic sequences, we developed a differential

* Corresponding author. Fax: +7 495 330 6538.
E-mail address: lev@humgen.siobc.ras.ru (L.G. Nikolaev).

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.01.040

electrophoretic mobility shift assay (EMSA)[1] display based on the two-dimensional (2D) EMSA procedure proposed previously in our laboratory [8,9]. Using this approach, we identified 10 cell type-specific protein-binding sites (protein target sites [PTSs]) within a 137-kb human chromosome 19q13.1 region.

Materials and methods

Cells and nuclear extract

Jurkat cells (American Type Culture Collection [ATCC] TIB-152, human acute T-cell leukemia) were grown in suspension in RPMI 1640, PANC-1 (ATCC CRL-1469, human pancreatic epithelioid carcinoma) in a mixture of Dulbecco's modified Eagle's medium (DMEM)–F12 Ham/RPMI 1640 (1:1), and HepG2 (ATCC HB-8065, human hepatocellular carcinoma) in DMEM with the addition of 1.5% l-glutamine. All media were supplemented with 10% fetal calf serum. Nuclear extracts were isolated as described previously [10] with modifications [11].

Preparation of short-fragment libraries and selection of protein-binding DNA fragments

DNA of R30072, R28588, F19410, and R30879 cosmids that cover the 137-kb human chromosome 19 region (kindly provided by A. Olsen, Lawrence Livermore National Laboratory, USA) was isolated using a Wizard Plus Minipreps DNA Purification System (Promega, USA) according to the manufacturer's protocol. To remove residual *Escherichia coli* genomic DNA, the isolated cosmid DNA was treated with adenosine triphosphate (ATP)-dependent DNase (Epicentre, USA) as recommended by the manufacturer. To obtain a short-fragment library, the DNA was further digested exhaustively with *Sau*3A and *Alu*I restriction endonucleases and was ligated to the library primer (ACTGAGCTCGAGTATCCATGAACA) and corresponding adapters (GATCTGTTCATGG for *Sau*3A and TGTTCATGGAACTC for *Alu*I). The short-fragment library was polymerase chain reaction (PCR) amplified using the library primer and was ^{32}P labeled as described previously [9,12].

For 2D EMSA, 3 to 5×10^6 cpm (10–30 ng) of the labeled short-fragment mixture, 2 μg poly(dI-dC)*poly(dI-dC) (Sigma, USA), 1 μg Cot-1 DNA (Invitrogen, USA), and 4 μg Jurkat, PANC-1, or HepG2 cell nuclear extract protein were mixed in 15 μl (final volume) of 12 mM Hepes–KOH (pH 7.9), 12% glycerol, 60 mM KCl, 0.3 mM K–ethylenediaminetetraacetic acid (EDTA), and 0.6 mM dithiothreitol. The reaction mix was incubated for 20 min at room temperature and then separated by 2D electrophoresis as described previously [13].

The gel was autoradiographed overnight and the region containing PTSs (Fig. 1) was excised; cut into small pieces; placed in 10 ml of 0.5 M NH_4Ac, 1 mM EDTA, and 0.1% sodium dodecyl sulfate (SDS); and kept overnight at 37 °C to elute DNA. DNA in a 1-ml eluate was precipitated with ethanol, washed, and dissolved in 100 μl TE buffer. Then 2 μl of the solution obtained was used as template for PCR amplification with the library primer (20 s/94 °C, 60 s/60 °C, and 90 s/72 °C for 20 cycles). The PCR product was labeled, and the 2D EMSA selection procedure was repeated. After the second selection round, the DNA fragments were eluted and PCR amplified as described above to produce the short-fragment library of putative PTSs.

Differential display

Short-fragment libraries of putative PTSs prepared for three cell lines were labeled with ^{32}P as described previously [9,12]. Labeled DNA of each library (~ 100 ng) was mixed with 10 μl of formamide containing 0.1% bromophenol blue and 0.1% xylene cyanol, heated at 70 °C for 20 min, and resolved in a 6% denaturing polyacrylamide gel ($0.4 \times 200 \times 400$ mm) containing 7 M urea and 89 mM Tris–borate (pH 8.0) and 2 mM EDTA (TBE) buffer. The electrophoresis was conducted at a constant power of 40 W until the xylene cyanol ran approximately 350 mm, and the gel was dried and autoradiographed. Differential bands were excised from the gel, and DNA eluted in 30 μl of water for 2 h at 60 °C. A 3-μl volume of the eluate was PCR amplified with the library primer and cloned into a pGEM-T vector (Promega).

One-dimensional EMSA

For EMSA, the inserts of individual clones were labeled by PCR as described above and were purified by electrophoresis in a 5% polyacrylamide gel. EMSA was done essentially as described above with a 50,000-cpm probe, 1 μg nuclear extract protein, and 1 μg poly(dI-dC)*poly(dI-dC). For competition experiments, an excess of unlabeled probes was added.

Sequencing, computer analysis, and mapping

Sequencing was done with an ABI PRISM BigDye Terminator Kit (version 3.1, Applied Biosystems, USA) using an ABI PRISM 3100–Avant automated sequencer. The sequences obtained were mapped by comparison with those deposited in GenBank using the BLAST [14] server at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST). The data were further analyzed using the University of California, Santa Cruz (UCSC) Human Genome Browser [15] (http://genome.ucsc.edu).

[1] *Abbreviations used:* EMSA, electrophoretic mobility shift assay; 2D, two-dimensional; PTS, protein target site; ATCC, American Type Culture Collection; DMEM, Dulbecco's modified Eagle's medium; PCR, polymerase chain reaction; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; UCSC, University of California, Santa Cruz; L, LINE; LTR, long terminal repeat.