Two-dimensional gel electrophoresis for identifying nucleic acid-binding proteins

Jonathan A. Stead and Kenneth J. McDowall

Introduction

The electrophoretic mobility shift assay (EMSA) has been used extensively to analyse nucleic acid–protein interactions for two decades. It is based on the finding that when a fragment of nucleic acid (the 'probe') is bound by protein, its electrophoretic mobility is retarded when compared to its free state. Thus, the final position of a labelled probe when bound by protein appears shifted. Accordingly, the EMSA is also referred to as the gel shift or gel retardation assay. Even relatively labile nucleic acid–protein complexes (with apparent dissociation constants in the range of hundreds of nM) can be analysed using EMSAs as the gel matrices provide a stabilising 'caging effect'. Complexes are also stabilised by the use of electrophoresis buffers with relatively low ionic strength.

Although the EMSA is used widely to detect the presence of nucleic acid-interacting factors and to assess the specificity and affinity of interactions, for the purpose of identifying the actual polypeptide(s) that interacts with a particular probe, many laboratories, including our own, have resorted to purifying the binding factor. This is often a laborious and time-consuming process; thus, we developed an identification method that circumvents the need for extensive purification. The approach utilises the resolving power of SDS–polyacrylamide gel electrophoresis to allow the identification of protein that, in a prior EMSA step, has altered mobility because it binds a probe. In principle, the approach need not be limited to the identification of DNA-binding proteins as the EMSA can be used to study proteins that bind RNA. We were recently invited to provide a detailed protocol for our method.

Methods for the identification of DNA-binding proteins are important as they allow function to be studied *in vivo* using reverse genetic approaches. They also aid biochemical and structural studies by allowing comparison with database entries of known properties and structure, and permitting the overexpression and manipulation of the corresponding genes in the natural or a heterologous host.

Overview of technique

The protocol is applicable to cases where the following three conditions are met; (i) an EMSA has already revealed the existence of a nucleic acid-interacting factor of interest, (ii) the interaction with probe produces a tight band that can migrate to a central position within a gel and (iii) sufficient complex is formed to allow detection of the protein component by Coomassie staining. With regard to the latter, the rule-of-thumb is that if a protein band (spot) can be detected by a standard Coomassie-staining protocol, it should be readily identifiable by peptide mass fingerprinting (PMF).

The level of purification required before the identification of protein by our approach is dependent on the number of copies of the nucleic acid-binding factor per cell. In our experience, we are able to assay binding activity in bacterial samples containing approximately 100 μg of protein. This amount is equivalent to the total protein content of approximately 6×10^8 *Escherichia coli* cells. Although we have no experience of working with samples from eukaryotic sources, we estimate using published values that the same amount of protein could be obtained from 2×10^7 haploid yeast cells. The amount of protein that can be detected readily by Coomassie staining is 0.5 pmol or more, which is equivalent to 3×10^{11} molecules. Given the aforementioned estimates and assuming that the majority of the molecules of a nucleic acid-interacting factor can be shifted, factors present at 500 or 15,000 copies per cell of *E. coli* or yeast, respectively, should be identifiable in cell extracts

using our approach. While transcription factors such as the cAMP receptor protein of *E. coli* and Sp1 of yeast are present in numbers in excess of these levels, many are not. Indeed, a recent analysis of protein expression in yeast indicates that the average copy number of its transcription factors is approximately 1,000. While in many cases, protein purification will be required before the adoption of our approach to identify a nucleic acid-interacting factor, the level of purification required should be easily obtainable in a single step such as ion-exchange chromatography.

An EMSA is run within a tube gel as is a control that lacks probe. The tubes are then place on top of separate SDS-polyacrylamide gels and the proteins resolved. Gels corresponding to an experimental and mock analysis are shown in Figure 1.

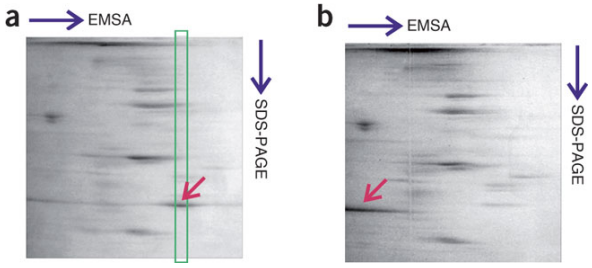

Fig. 1. Two-dimensional gel electrophoresis for identifying nucleic acid-binding proteins. Panel A, experimental sample. The green box indicates those polypeptides that are in line with the position of the complex at the end of running the first dimension, whereas the arrow indicates the position within this box of the unique band (spot) that corresponds to the nucleic acid-interacting factor. Panel B, mock sample. The arrow indicates the position of the free protein factor. Taken from Stead & McDowall (2007).

Publications

Stead, J.A., Keen J.N., & McDowall, K.J. (2007) Two-dimensional gel electrophoresis for identifying proteins that bind DNA or RNA. *Nature Prot.*, **2**, 1839-1848.

Stead, J.A. & McDowall, K.J. (2006) The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Mol. Cell. Proteomics*. **5**, 1697-1702.

Funding

This work was funded by a BBSRC grant and studentship to K. J. M.