

Electrophoretic Mobility Shift Assay: Analyzing Protein – Nucleic Acid Interactions

Carolina Alves and Celso Cunha
*Center for Malaria and Tropical Diseases
Institute of Hygiene and Tropical Medicine, New University of Lisbon
Portugal*

1. Introduction

Interactions between proteins and nucleic acids mediate a wide range of processes within a cell from its cycle to the maintenance of cellular metabolic and physiological balance. These specific interactions are crucial for control of DNA replication and DNA damage repair, regulation of transcription, RNA processing and maturation, nuclear transport, and translation.

The characterization of protein-nucleic acid interactions is essential not only for understanding the wide range of cellular processes they are involved in, but also the mechanisms underlying numerous diseases associated with the breakdown of regulatory systems. These include, but are far from being limited to, cell cycle disorders such as cancer and those caused by pathogenic agents that rely on or interfere with host cell machinery. More recently, it has been hypothesized that many neurological disorders such as Alzheimer's, Huntington's, Parkinson's, and polyglutamine tract expansion diseases are a consequence, at least in part, of aberrant protein-DNA interactions that may alter normal patterns of gene expression (Jiménez, 2010).

The electrophoretic mobility shift assay (EMSA), also known as gel retardation assay, is a regularly used system to detect protein-nucleic acid interactions. It was originally developed with the aim of quantifying interactions between DNA and proteins (Fried & Crothers, 1981; Garner & Revzin, 1981) and since then evolved to be suitable for different purposes including the detection and quantification of RNA-protein interactions. EMSA is most commonly used for qualitative assays including identification of nucleic acid-binding proteins and of the respective consensus DNA or RNA sequences. Under proper conditions, however, EMSA can also be used for quantitative purposes including the determination of binding affinities, kinetics, and stoichiometry.

EMSA is a commonly used method in the characterization of transcription factors, the most intensely studied DNA-binding proteins, and the largest group of proteins in humans, second only to metabolic enzymes. Their purification and identification is crucial in understanding gene regulatory mechanisms. Transcription factors are sequence specific DNA binding proteins that are usually assembled in complexes formed prior to transcription initiation. They bind discreet and specific DNA sequences in the promoter

region functioning either as an activator or repressor of expression of the targeted gene through protein-protein interactions (reviewed by Simicevic & Deplancke, 2010). Transcription factors play essential roles during development and differentiation. It is well established that disruption of normal function of tissue-specific transcription factors, as a result of mutations, is often associated with a number of diseases including most forms of cancer, neurological, hematological, and inflammatory diseases. Additionally, transcription factors are often found differentially expressed in different pathologies suggesting an at least indirect involvement on the onset or progression of diseases. One of the most prominent examples of the involvement of transcription factors in development and progression of diseases is perhaps the p53 protein. p53 is a transcription factor involved in the modulation of expression of several genes that regulate essential cellular processes such as cell proliferation, apoptosis, and DNA damage repair (reviewed by Puzio-Kuter, 2011). Mutations in p53 that cause loss of function were reported in about 50% of all cancers. It is believed that this loss of function makes cancer cells more prone to the accumulation of mutations in other genes thus facilitating and accelerating the formation of neoplasias (reviewed by Goh et al., 2011).

In our laboratory, research is mainly directed to the study of host-pathogen interactions during hepatitis delta virus (HDV) replication and infection. HDV is the smallest human pathogen so far identified and infects human hepatocytes already infected with the hepatitis B virus (HBV). Both viruses have the same envelope proteins that are coded by the HBV DNA genome. HDV is, thus, considered a satellite virus of HBV. The HDV genome consists of a single-stranded, circular, RNA molecule of about 1700 nucleotides. This genome contains only one open reading frame from which two forms of the same protein, the so-called delta antigen, are derived by an editing mechanism catalyzed by cellular adenosine deaminase I. Both forms, small and large delta antigen, were shown to play crucial roles during virus replication: the small delta antigen is necessary for virus RNA accumulation and the large delta antigen plays an important role during envelope assembly (reviewed by Rizzetto, 2009). However, neither protein seems to display any known enzymatic activity. Accordingly, HDV is highly dependent on the host cell machinery for virus replication. It has been shown through EMSA that the small delta antigen binds *in vitro* to RNA and DNA without any specificity, which is in agreement with one of the roles attributed to the protein as a chaperone (Alves et al., 2010). Making use of different experimental approaches it was possible to identify a number of cellular proteins that interact with HDV antigens or RNA (reviewed by Greco-Stewart & Pelchat, 2010). However, the precise role played by most host factors during the virus life cycle remains elusive. Furthermore, it is highly consensual among HDV researchers that many other cellular factors that interact with delta antigens or HDV RNA remain to be identified and it is crucial to find those that interact with HDV RNA for a better insight on its replication and as possible targets for new therapies.

In this chapter we will review the principles of EMSA and its advantages and limitations for the quantitative and qualitative analysis of protein-nucleic acid interactions. The key parameters influencing the quality of protein samples, binding to nucleic acids, complex migration in gels, and sensitivity of detection will be discussed. Finally, an overview of the principles, advantages and disadvantages of methods that are an alternative to gel retardation assays will be provided.

2. Advantages and limitations

Since its first publication, in 1981, several improvements and variant techniques of EMSA were reported. Originally described as a method to qualitatively detect protein-DNA interactions, gel retardation assays rapidly became one of the most popular methods to map interaction sequences and domains not only in DNA but in RNA-protein interactions as well. EMSA was also adapted in order to allow the determination of quantitative parameters including complex stoichiometry, binding kinetics and affinity.

Several features made EMSA one of the most popular methods among researchers that study protein-nucleic acid interactions. Probably, the main advantages of EMSA when compared to other methods, as we will further discuss in the next sections, may be considered as follows: (1) EMSA is a basic, easy to perform, and robust method able to accommodate a wide range of conditions; (2) EMSA is a sensitive method, using radioisotopes to label nucleic acids and autoradiography, it is possible to use very low concentrations (0.1nM or less) and small sample volumes (20 μ L or less; Hellman & Fried, 2007). Even though, less sensitive, non-radioactive labels are often used as well. These labels can further be detected using fluorescence, chemiluminescence or immunohistochemical approaches. Although less sensitive than radioisotopes, the wide variety of labels that can be used makes EMSA a very versatile method; (3) EMSA can also be used with a wide range of nucleic acid sizes and structures as well as a wide range of proteins, from small oligonucleotides to heavy transcription complexes; (4) Under the right conditions a gel retardation assay can separate the distribution of proteins between several nucleic acids within a single sample (Fried & Daugherty, 1998) or distinguish between complexes with different protein stoichiometry and/or binding site distribution (Fried & Crothers, 1981); (5) Finally, but not less important, it is possible to use both crude protein extracts and purified recombinant proteins enabling the identification of new nucleic acid-interacting proteins or characterization of specific proteins and its targets.

Despite its sensitivity, versatility and usually easy to perform protocols, EMSA is often considered to bear a number of limitations. Dissociation can occur during electrophoresis since samples are not at equilibrium during the run, thus preventing detection. Additionally, complexes that are not stable in solution may be stable in the gel requiring very short runs so that the observed pattern relates to what happens in solution. EMSA does not provide a straightforward measure of the weights or entities of the proteins as mobility in gels is influenced by several other factors. Also, EMSA does not directly provide information on the nucleic acid sequence the proteins are bound to. However, this problem may usually be overcome using footprinting approaches as described further ahead. Kinetic studies using EMSA are limited since the time resolution for a regular EMSA protocol consists of the time required to mix the binding reaction and for the electrophoretic migration to occur before the mix enters the gel. Only processes that have relaxation times larger than the interval required for solution handling are suitable for kinetic studies.

3. How complexes migrate in gels

In this section, we will start with a simple account of the characteristics of the electrophoretic mobility of nucleic acids alone, and afterwards we will discuss how the formation of protein-nucleic acid complexes alters these characteristics.

In a non-denaturing agarose or polyacrylamide gel and conventional buffer conditions the nucleic acids, being negatively charged, will migrate towards the anode when electric current is applied. The gel will then act as a sieve selectively impeding the migration in proportion to the nucleic acid molecular weight, which is generally proportional to its charge. Therefore, and as the weight is approximately related to chain length, the length of nucleic acid is estimated by its migration. There is though another property that affects gel migration that is the topology of the nucleic acid (conformation, circularity) making the molecules seem longer or shorter than they really are. Secondary and tertiary structures can be removed using denaturing agents (for example, formaldehyde, formamide and urea) allowing for the electrophoretic mobility to become a simple function of molecular weight. Obviously, this denaturing step cannot be applied in a gel retardation assay as it would impede the interaction between the protein and nucleic acid.

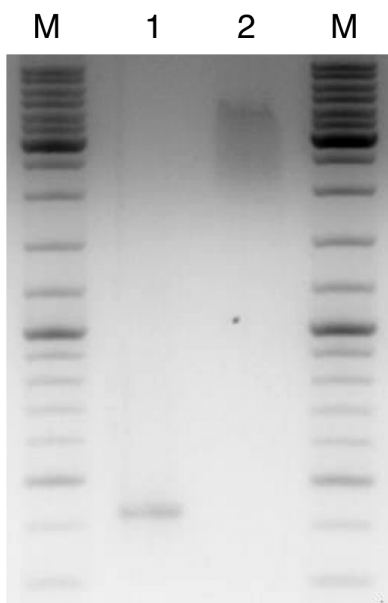


Fig. 1. Example of an electrophoretic mobility shift assay. An unlabeled DNA of 400 base pairs (bp) was incubated in a phosphate buffer (137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4 , 1.5mM KH_2PO_4 , pH 7.4) in the absence (1) or presence (2) of $2\mu\text{M}$ of small delta antigen. The samples were loaded onto a 1.5% agarose gel and after electrophoresis in TAE buffer (40mM Tris acetate, 1mM EDTA) the DNA was stained with ethidium bromide. (M) represents the molecular weight marker (GeneRuler DNA Ladder mix, Fermentas).

When a protein is added to the mix and interacts with the nucleic acid forming complexes it results in a change in gel migration relative to that of the free nucleic acid. This shift is mainly due to an obvious increase in the molecular weight, the adjustment of charge and eventual changes in the nucleic acid conformation. In figure 1 we give an example of an EMSA study where the small delta antigen was added to DNA. It is clear that the addition

of the small delta antigen (Fig.1. well 2) to a 400bp DNA fragment results in the formation of a complex with decreased gel mobility when compared with the unbound DNA (Fig.1. well 1). We can conclude that under our *in vitro* binding conditions, the small delta antigen interacts with the given 400bp DNA fragment causing a clear mobility shift.

It is expected that when protein binds a nucleic acid fragment there will be a decrease in relative mobility and if the protein doesn't induce any appreciable bend on the nucleic acid then the conformational contribution to the decrease is small. Although an increase in the protein molecular weight results in reduction of gel migration it has been reported that the increase of the nucleic acid length can have the opposite effect. This was reported for the Lac repressor bound to DNA fragments of increasing sizes, which resulted in an increase of relative mobility (Fried, 1989). This observation indicates that the ratio of protein and nucleic acid weights is more important in the migration than the absolute weight of the complex. Another interesting study reports that the binding of protein to a nucleic acid can accelerate mobility. This was observed for relatively large linear DNA binding to a protein from the hyperthermophilic *Methanothermus fervidus* that was shown to induce nucleic acid condensation (Sandman et al., 1990). In this case the conformational change of the DNA is a stronger factor than the weight increase, causing acceleration rather than a decrease in relative mobility.

Overall, the conformational features that influence gel migration of protein-nucleic acid complexes are not thoroughly studied and questions are only raised when exceptions emerge such as the ones mentioned above. Nowadays, the EMSA method is almost exclusively used to analyze the interaction between proteins and nucleic acids and to a lesser extent its conformations that can influence gel migration. When exceptions arise and the retardation pattern is not exactly as predicted, it can still point out clearly whether the molecules are interacting or not. In the end, the exact location of the resulting gel bands cannot be predicted but the answer is usually unambiguous.

External factors can also influence the separation of the bound or unbound nucleic acid such as the nature of the gel matrix and temperature during electrophoresis. Generally, the best resolution is obtained with the smallest pore diameter that allows the migration of unbound nucleic acid. However, if large complexes are expected there should be a compromise in pore size so that they can enter the gel matrix. As will be discussed below, polyacrylamide gels offer the best conditions for small complexes and nucleic acid fragments. On the other hand, agarose gels are more suitable for larger aggregates.

The detection of a protein-nucleic acid complex within a gel depends critically on the resolution obtained between unbound nucleic acid and the formed complexes as well as its stability within the gel matrix. In most cases, the gel matrix is expected to stabilize the preformed complex as it impedes the diffusion of dissociating components maintaining the concentration of protein and nucleic acid (and complex) at levels as high or higher than those achieved in the equilibrium binding reaction. This of course is compromised if for instance the salt concentration in the binding reaction differs largely from that in the electrophoresis/gel buffer, resulting in an adjustment in salt concentration that could disrupt the complexes formed. As the gel retardation method is an *in vitro* assay, when extrapolating to the *in vivo* conditions one must be careful as the former may provide favorable binding conditions that are not achieved at physiological concentrations.

4. The method

There are five focal steps in a conventional EMSA protocol that involve different variables susceptible to optimization: (1) preparation of protein sample; (2) synthesis and labeling of nucleic acid; (3) binding reaction; (4) non-denaturing gel electrophoresis and (5) detection of the outcome. In this segment we will discuss each step separately mentioning the key variables in each one and the options available for any given situation. Figure 2 represents schematically the regular steps in a gel retardation assay that will be discussed below. Whenever possible we will also refer to examples in the literature.

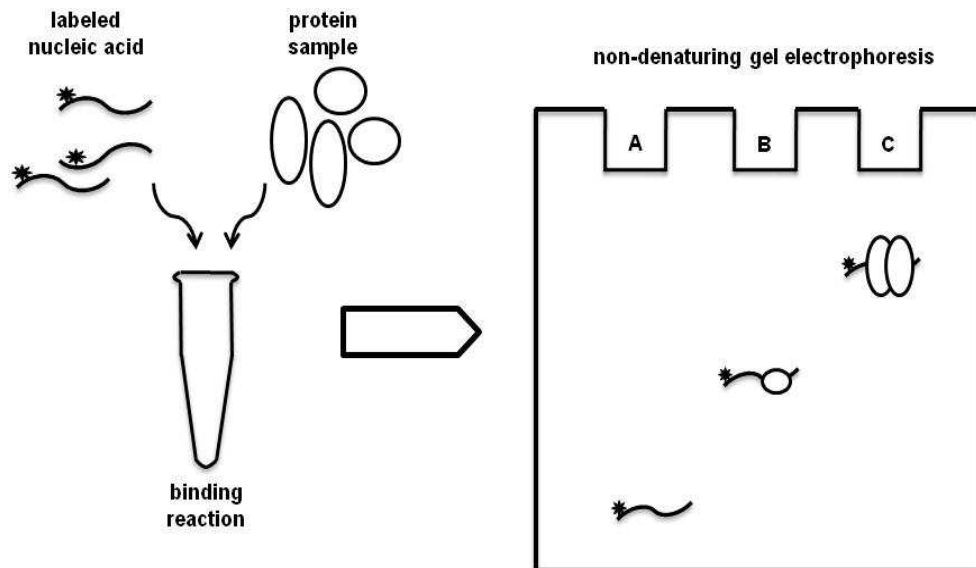


Fig. 2. A schematic representation of a conventional EMSA protocol. The labeled nucleic acid, simplified as lines with a star representing the label, is mixed with the protein sample, represented by the oval shapes, in a binding reaction and then loaded into a non-denaturing gel. After electrophoresis the result is detected according to the label in the nucleic acid. On the schematic gel (A) represents a well on which only the labeled nucleic acid was loaded. The free nucleic acid is expected to have more mobility than the bound molecules. In well (B) is symbolized a labeled nucleic acid binding to one small peptide and in well (C) is binding to two larger proteins. The heavier complex (in C) is expected to display the lowest mobility during electrophoresis and therefore is closer to the beginning of the gel.

4.1 Preparation of the protein sample

Regarding the protein sample, the EMSA can be divided into two categories based on whether the nucleic acid-interacting protein is known or not. Therefore, preparing the protein sample will depend on which category it falls, in order to obtain an optimal performance.

When faced with a putative nucleic acid-binding protein or complex of completely unknown subcellular origin, whole cell extracts must be used. If there is an educated guess

on the nature of the protein, it is advisable to isolate nuclear and cytoplasmic proteins from crude extracts improving the results. Particularly, if the binding protein is thought to be nuclear and in low abundance, the isolation of nuclear extracts will prevent the dilution that would occur if whole cell extracts were used, which could render the concentration too low for the protein to be even detected.

Cell extracts are easy and relatively fast to obtain and the methods are commonly derived from the protocol described by Dignam and collaborators almost three decades ago (Dignam et al., 1983). This method isolates both nuclear and/or cytoplasmic proteins suitable for later analysis using EMSA. One disadvantage in preparing cell extracts is its crudeness; they generally degrade faster than purer preparations due to the presence of cellular proteases. To limit protein degradation or alteration the protocol should be performed on ice or at 4°C and protease inhibitors should be added. A control test can easily be performed to assess the viability of the extract by using ubiquitous DNA probes (Kerr, 1995). If these fail then the cell extract might be “dead”. Despite its disadvantages cell extracts are needed when the interest lies in identifying new nucleic acid-binding proteins or when a complex of different proteins is needed to interact with the target nucleic acid as sometimes one recombinant protein cannot bind by itself. Tissue samples can also be a source of protein sample for these assays. The same care should be taken as in whole cell extracts to minimize the activity of proteases.

If the nucleic acid-binding protein is known then recombinant proteins can be expressed and purified. Recombinant or heterologous proteins are commonly expressed in bacteria or an eukaryotic cell line of interest. Fusion proteins of the target are generally constructed with a tag to facilitate purification. Common tags, such as glutathione-S-transferase (GST), tandem affinity purification tag (TAP tag), maltose binding protein (MBP) or 6xHistidine, are cloned in frame with the protein. Sometimes it is possible to include a protease cleavage site between the protein of interest and the tag so the latter can be easily removed after purification. Even though a tag can be very helpful, it should be taken into account that it can alter the recombinant protein conformation and even disrupt its binding ability. On the other hand they can be helpful in stabilizing the protein terminus they are close to. A careful study is needed when choosing the tag and usually small peptides are preferred to minimize its impact on the recombinant protein of interest.

There are several systems available for the production of heterologous proteins of which bacterial extracts of *Escherichia coli* are one of the most widely used. This Gram-negative bacterium remains an attractive host due to its ability to grow rapidly and with high density using inexpensive substrates. Its genetics has been well characterized for quite some time and there is a wide range of cloning vectors as well as mutant host strains that make it such a versatile system. Typically, the heterologous complementary DNA is cloned into a compatible plasmid which is then transfected into the bacteria to achieve a high gene dosage. This doesn't necessarily guarantee the accumulation of high levels of a full-length active form of the recombinant protein but other efforts can be made to improve that. To achieve high-level production in *E. coli* strong promoters should be used such as the bacteriophage T7 late promoter, and usually the T7 polymerase is also present under IPTG (isopropyl-β-D-1-thiogalactopyranoside)-induction. In the past years several strains have been engineered to improve the recombinant protein yields through efforts to increase mRNA stability as well as improve transcription

termination and translational efficiency (reviewed by Baneyx, 1999 and Makino et al., 2011). However, this extensively used system for protein overexpression has an important drawback when studying eukaryotic proteins. The bacterial systems are not able to perform post-translational modifications that would eventually happen *in vivo* in eukaryotic cells.

When working with recombinant nucleic acid-binding proteins it should be taken into account the importance of post-translational modifications on the protein's binding ability. A careful research of previous reports might hint if it is necessary to perform modifications prior to the binding reaction. In some cases post-translational modifications change the sequence-specificity of the binding. For example, genotoxic stress induces modifications on the C-terminus of the tumor suppressor protein p53 that modulate its DNA-binding specificity (Apella & Anderson, 2001). If the modifications are crucial, rather than using bacterial extracts a more biologically relevant host should be considered. Transient gene expression in mammalian cells has become a routine approach to express proteins in cell lines such as human embryonic kidney cells. The benefits are obvious for the production of eukaryotic proteins in mammalian cells as post-translational modifications will likely be native or near-native, solubility and correct folding are more likely to occur as well as expression of proteins in their proper intracellular compartments. These methods, however, tend to be more expensive as cells need a more complex growth media and there is a lower diversity in cloning vectors. To get out of the latter limitation an alternative approach uses baculovirus-infected insect cells. In this method a recombinant virus is produced either by site-specific transposition of an expression cassette into the shuttle vector or through homologous recombination (reviewed by Jarvis, 2009).

When expressing recombinant proteins, sometimes, the heterologous genes interfere severely with the survival of the host cell. For toxic proteins produced in *E. coli* strains there are some techniques available to get around this problem. A highly toxic gene can be defined as a gene that, when introduced into a cell, causes cell death or severe growth and maintenance defects even prior to expression induction. The best solution for expressing a highly toxic gene is to enable the host to tolerate it during the growth phase, so that after induction an efficient expression ensures a rapid and quantitative production of the toxic protein before the cell dies (reviewed by Saida et al., 2006). This can be achieved by different strategies such as manipulation of the gene's transcriptional and translational control elements, for example, by suppressing basal expression of the toxic protein from leaky inducible promoters. Managing the coding sequence to produce reversible inactive forms or controlling the plasmid copy number is also an option as well as selecting less susceptible *E. coli* strains or adding stabilizing sequences.

Cell-free systems are also available to express recombinant proteins including *in vitro* transcription\translation systems such as rabbit reticulocyte systems, wheat germ based systems or *E. coli* cell-free protein expression systems (reviewed by Endo & Sawasaki, 2006). Here, proteins can be expressed directly from cDNA templates obtained through PCR, avoiding subcloning which makes it a faster method by skipping this step, and eventually cheaper. It can also be used to express proteins that seriously interfere with the cell physiology such as the toxic proteins mentioned above. On the other hand these methods usually achieve smaller yields than for instance bacterial extracts approaches.

4.2 Synthesis and labeling of nucleic acids

One of the key advantages of EMSA is its versatility as it can be performed using a wide range of nucleic acid structures and sizes. This method can characterize both double- and single-stranded DNA as well as RNA, triplex and quadruplex nucleic acids or even circular fragments. The probe design and synthesis depends on the application or purpose of the study and is a significant aspect, as it will influence the detection and therefore the sensitivity of the results. There are two main aspects to consider in this step: the length of the nucleic acid and its labeling.

Unlabeled nucleic acids can be used in a gel retardation assay and be detected by post-electrophoretic staining with chromophores or fluorophores that bind nucleic acids or in the “classical way” using ethidium bromide. However the use of labeled nucleic acids is usually preferred as it can facilitate detection and add sensitivity to the method. The most common choice is radioisotope labeling as it offers the best sensitivity without interfering with the structure of the probe. A higher sensitivity makes it ideal for assays that have a limited amount of starting material. The radioisotope, usually ^{32}P , can be incorporated in the nucleic acid during its synthesis, by the use of labeled nucleotides, or afterwards via end labeling using a kinase or a terminal transferase. With a radioactive label the EMSA results can be easily detected by autoradiography. Even if radioisotope labeling confers high sensitivity to the method it implies handling hazardous radioactive material requiring extra safety measures that may not be available. Other labels can be used as alternatives that, even though are less sensitive, are a lot safer to manipulate and more stable such as fluorophores, biotin or digoxigenin (Holden & Tacon, 2011). When these molecules are used detection is achieved by chemiluminescence or immunohistochemistry. Although, in general radioisotope labeling achieves higher sensitivity there are some reports that similar results can be obtained with other labels such as Cyano dye Cy5 (Ruscher et al., 2000).

Although the most common approach is the labeling of the nucleic acid probe there are protocols available that employ protein labeling at the same time. For example, Adachi and co-workers suggest the use of an iodoacetamide derivative labeling of the thiol residue of cysteins (Adachi et al., 2005). Using radioisotope labeled DNA mixed with a nuclear protein extract they perform a conventional EMSA and after detection by autoradiography the complexes are eluted from excised gel bands and treated with 5-iodoacetamidofluorescein for protein labeling. The sample is then loaded onto a denaturing gel and after electrophoresis is transferred to a membrane and detected with anti-fluorescein antibody. This allows the characterization of the proteins in the complex giving information on how many proteins are present and their molecular weight. However it is not able to detect proteins without cystein residues.

Regarding the length of the nucleic acid probe, it depends on what is being studied. If one is looking for specific binding sites, small probes can be used to assess with each segment the protein will interact. The use of short nucleic acids has several advantages as they are easily synthesized and inexpensive to purchase; a small sequence has less non-specific binding sites (it should be particularly advantageous when a protein has low sequence-specificity); the electrophoretic resolution between complexes and free nucleic acid is higher so shorter electrophoresis times can be used. Nevertheless, in a short sequence the binding sites are closer to the molecular ends which can cause aberrant binding and it can be tricky to resolve the free nucleic acid from the complexes formed if these have a very high molecular weight.

On the other hand, the longer nucleic acid targets avoid these problems but will have more non-specific binding sites and the mobility shift is generally smaller requiring longer electrophoresis times as they run more slowly. A compromise needs to be reached depending on what the EMSA study is trying to achieve.

4.3 Binding reaction

The interaction between proteins and nucleic acid is sensitive to salt concentration and pH as it will influence the protein charge and conformation. However, the experimental conditions are very versatile in that different buffers can achieve good results. The most commonly used are Tris based buffers but other options include 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 3-(N-morpholino)-propanesulfonic acid (MOPS), and glycine or phosphate buffers. Naturally, it is advisable to provide an environment as close as possible to physiological conditions so the data obtained *in vitro* can be related to what happens *in vivo*.

Additives can be included in the binding reaction either if the interactions require the presence of co-factors or stabilizing agents, or as helpful components to minimize non-specific binding. Glycerol or other small neutral solutes, for example sucrose, can be added to the binding mixture to stabilize labile proteins or enhance the stability of the interaction (Vossen et al., 1997). These solutes are used at final concentrations of 2M or less, as higher concentrations might interfere with the sample's viscosity and complicate handling. Other assays may require the presence of co-factors for a correct interaction such as the presence of cAMP for the *E. coli* CAP protein (Fried & Crothers, 1984) or ATP for human recombinase Rad51 (Chi et al., 2006). Non-ionic detergents are used to maximize protein solubility. In this case, the concentrations used depend on the detergent and system under study. Nuclease and phosphatase inhibitors can be useful as well as protease inhibitors, which as mentioned before, are particularly important when the protein sample comes from cell extracts. These inhibitors are commercially available and the concentration depends on the manufacturer's instructions. Some of the additives mentioned, particularly those involved in stabilizing the formed complexes can be included not only in the binding mixture but also in the gel buffers.

To minimize non-specific loss of protein the addition of a carrier protein (less than 0.1mg/mL) such as bovine serum albumin can be very helpful. The addition of unlabeled competing nucleic acids is suitable when there are secondary binding activities that mask the relevant one. Of course this only works if the protein interacts with the target nucleic acid with greater affinity than its competitor and the secondary binding does not discriminate between the sequences. Since the presence of a competing nucleic acid will always reduce the amount of specific binding, testing different competitors and concentrations is needed to optimize the assay. Another option to circumvent the problem of non-specific binding is the addition of salt at concentrations that will disrupt non-specific ionic bonds but leave the more specific interactions unimpaired.

4.4 Non-denaturing gel electrophoresis

After the binding reaction the free nucleic acid is separated from the formed complexes by non-denaturing gel electrophoresis. EMSA can be performed on polyacrylamide or agarose

gels depending mainly on the size of the nucleic acid and desired resolution. The average pore size is estimated to be around 5 to 20nm in diameter for 10 and 4% acrylamide gels respectively (Lane et al., 1992). Typically the higher concentration gels are used for oligonucleotides and small RNAs and the lowest concentration for DNA fragments of around 100bp. A polyacrylamide gradient gel is sometimes preferred over linear gels as the gradient in pore size increases the range of molecular weight fractionated in a single run, which is particularly important when the complex has a much higher weight than the free nucleic acid (Walker, 1994). When complexes of different composition are formed, the gradient gels are also more likely to separate those with close molecular weight.

Agarose gels, on the other hand, have a pore size of around 70 to 700nm (Lane et al., 1992) in diameter and are therefore mostly used in assays with larger nucleic acid fragments or when large protein complexes are expected. Overall, polyacrylamide gels offer a better resolution for nucleic acid-protein complexes with a molecular weight of up to 500,000Da (Fried, 1989 as cited in Hellman & Fried, 2007).

Regarding the electrophoresis buffers, it should be taken into account the fact that the interaction between nucleic acids and proteins involves an ionic component. Therefore, the buffer's ionic strength and pH are important features that play a role in the complex stability. Although this is a very important factor there hasn't been, to our knowledge, any thorough study on the subject. The choice of electrophoresis buffers is varied and generally low ionic strength buffers are preferred and sometimes coincide with the buffer used in the binding reaction. Buffers with a medium salt concentration help stabilize the complexes, generate less heat during electrophoresis and also increase the speed of migration. High salt concentrations not only disrupt the complexes but also interfere with its movement into the gel matrix and lead to significant heating during the electrophoresis. Too low salt concentrations can also disrupt the stability of the preformed complexes as well as separate a double stranded DNA template (Kerr, 1995). The most common buffers are TBE (90mM Tris-Borate, 2mM EDTA, pH 8) and TAE (40mM Tris-Acetate, 1mM EDTA, pH 8). However, there are some complexes that cannot be detected with the classical buffers. For example the complexes formed between phage Mu repressor and its operators have an electrophoresis buffer-dependent stability and require Tris-glycine buffer at pH 9.4. (Alazard et al., 1992 as cited in Lane et al., 1992).

Particularly, in agarose gels it is important to monitor the temperature during electrophoresis to prevent the gel from heating up which could result in dissociation of the nucleic acid-protein complexes. Some cases may require that pre-cooling of the gel or even that the electrophoresis proceeds at lower than room temperatures, which can be achieved with special refrigeration devices.

4.5 Detection

The detection of an EMSA result will naturally depend on the labels used if any has been used. The results uncovered can involve the detection of the mobility shift between free nucleic acid and the complexed form or the detection of the mobility shift of free protein and the complexes.

Looking at the nucleic acid component without any label added the shift in mobility can be detected by staining with molecules that bind nucleic acids. Different products can be used

ranging from the classic but hazardous ethidium bromide to other chromophores or fluorophores such as RedSafe DNA Stain (ChemBio) or SYBR® Safe DNA gel stain (Invitrogen). When the nucleic acid has been previously labeled the detection methods depend on the nature of the label. A ^{32}P radioisotope is one of the easiest and most sensitive methods to detect nucleic acids but it's a hazardous material to work with. Other very common labels are biotin, digoxigenin or fluorophores. These labels are innocuous but usually give less sensitive results and the detection procedure can involve extra steps such as transfer to a membrane and incubation with primary and secondary antibodies as well as intermediate washing steps. The results in these cases can be observed by immunohistochemistry or chemiluminescence approaches.

The detection of protein mobility shift involves less direct methods, meaning, extra steps such as a denaturing step and electrotransfer onto a membrane, may be necessary as they are usually immunodetected. If the protein of interest is known, and a specific antibody is available, it can be used in detection. If not, a method such as the one discussed above, proposed by Adachi and colleagues that involves labeling the thiol group of cysteins and using an antibody against the label. Stepwise, the easier way to detect protein in an EMSA is by labeling it with radioisotope, a method designated by reverse EMSA that will be discussed ahead. This procedure has the disadvantage of working with radioactive material but the mobility shift can be visualized by autoradiography.

5. EMSA applications

The gel retardation assay has been used under different conditions in order to achieve specific results. The method is useful in studying not only the interaction between proteins and nucleic acids but also in assessing nucleic acid conformational characteristics. It can be used to characterize bends in the DNA double helix with polyacrylamide gels and comparative measurements (for an example Crothers & Drak, 1992) or to detect complexes formed with super coiled DNA being sometimes designated as topoisomer gel retardation (for examples see Palecek, 1997; Nordheim & Meese, 1988). In this section we mention how a gel retardation assay can help characterize protein-nucleic acid interactions.

5.1 Binding constants

Although EMSA is most commonly used as a qualitative assay it can, under certain conditions, provide quantitative data for relatively stable complexes. One of its earliest applications was in the measurement of kinetic and thermodynamic parameters. The association rates are determined by mixing the complex components at known concentrations and loading them in a running gel at precise intervals (for an example Spinner et al., 2002). For dissociation rates, a time course experiment is done by addition of competing nucleic acid to the preformed complexes (Fried & Crothers, 1981). The binding constant can be determined by the amount of complex formed as a function of protein concentration at equilibrium or as a ratio of the association and dissociation constants (for an example Demarse et al., 2009). An alternative method to measure kinetic and thermodynamic constants is the nitrocellulose filter binding assay that will be mentioned below.

As an example we show in figure 3 the titration of a DNA with the small delta protein to assess binding constants. The binding reaction was done by incubating the samples in a

phosphate buffer during the same period of time (10 minutes) and then loading them onto an agarose gel for electrophoresis. It is clear that when the protein is present at only $0.25\mu\text{M}$ it does not interfere with the DNA mobility (Fig.3. well 2) as the band covered the same distance as the first sample, in which the protein was not present (Fig.3. well 1). But when $1.5\mu\text{M}$ of the small delta antigen are present in the binding reaction there is almost no free DNA present and the majority of the molecules are bound in a complex (Fig3. well 5). In the intermediate concentrations it can be clearly observed the decreasing presence of free DNA and increasing DNA-protein complexes as the protein concentration raises. We can consider that the dissociation constant can be estimated by quantifying the disappearance of the free DNA band (Demarse et al., 2009). From figure 3 we can say that the apparent dissociation constant is between 1 and $1.5\mu\text{M}$.

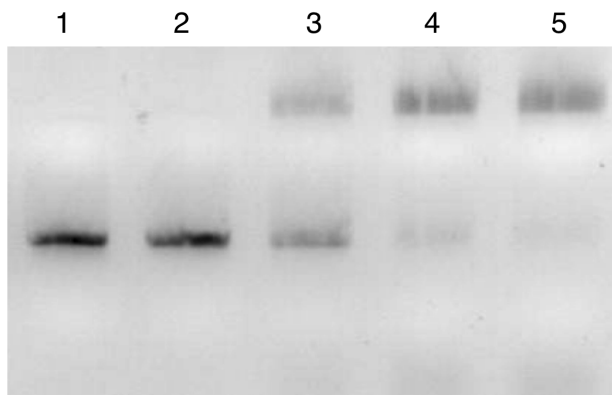


Fig. 3. Titration of a 500bp DNA fragment with the small delta antigen to estimate binding constants. An unlabeled 500bp DNA complementary to part of the HDV RNA was incubated, in a phosphate buffer (137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4 , 1.5mM KH_2PO_4 , pH 7.4), with increasing concentrations of small delta antigen of 0; 0.25; 0.5; 1; and $1.5\mu\text{M}$ and samples were loaded onto wells 1, 2, 3, 4 and 5, respectively. Electrophoresis was in a 1.5% agarose gel in TAE buffer and the DNA was stained with ethidium bromide.

5.2 Cooperativity

Proteins can bind nucleic acids in a cooperative manner, that is, the complexes formed involve the binding of more than one protein to a specific nucleic acid segment. These multiprotein complexes may be a consequence of direct protein-protein interaction needed for nucleic acid binding, or a protein-induced deformation of the nucleic acid is a prerequisite to facilitate the binding of a second protein, or it may result from the bringing together of molecules bound at distinct sites in the nucleic acid sequence. The cooperativity can be inferred in a gel retardation assay from the underrepresentation of intermediate complexes between the unbound and saturated states. Multiprotein complexes can be comprised of a single protein species forming a homomultimer or of different proteins. The latter can be easily characterized by EMSA by the stability of the complexes formed with one protein in the presence or absence of the other(s).

5.3 Stoichiometry

Determining the important parameter that is stoichiometry is not as easy a task as it seems. The apparent weight changes estimated from the complexes' gel mobility are not applicable in determining the stoichiometry due to complications of charges and conformational effects on gel migration. A different approach is needed. The presence of truncated or extended protein derived from the wild-type but with the same binding and multimerization capacity will originate new bands that can reflect the monomers bound to the nucleic acid (Hope & Struhl, 1987). A similar method that will be discussed in the next segment is the supershift EMSA that uses an antibody specific for the binding protein recognizing an epitope that is accessible while the protein is bound to the nucleic acid. The addition of the antibody to the preformed complex can provide an estimate of the number of proteins bound by the extent of increments in retardation (Michael N & Roizman B, 1991 as cited in Lane & Prentki, 1992).

A more complex approach has been proposed in 1988 to determine a complex's stoichiometry (Granger-Schnarr et al., 1988). After the separation of the free and the complexed nucleic acid on a non-denaturing gel, the proteins are transferred to a membrane after sodium dodecyl sulfate (SDS) denaturation. This then allows the detection of proteins directly or indirectly using a specific antibody. The protein bands as well as the nucleic acids autoradiograph are then quantified by densitometry and the relative stoichiometry can be determined. The need for a specific antibody limits this method to complexes formed by well known proteins with available antibodies.

6. EMSA variants

Over the years variations or coupling of the EMSA protocol with other methods has been proposed to enhance its results or obtain more information from one experiment. Some examples of these EMSA-based approaches will be presented.

6.1 Reverse EMSA (rEMSA)

A reverse EMSA consists in labeling the protein sample rather than the nucleic acid (Filion et al., 2006). This method shows the difference in mobility between the free protein and nucleic acid-bound protein. It is an approach that can facilitate the determination of the protein binding affinity using different nucleic acids. Because the label used is ^{35}S instead of ^{32}P it is less sensitive than the conventional EMSA due to the isotope's energy.

6.2 Supershift EMSA

The supershift EMSA uses the same protocol as a regular EMSA except in that an antibody against the binding protein is added. As a result there is a more marked mobility shift during electrophoresis because the antibody will increase the overall complex molecular weight, hence the term supershift. This method can help identify if the proteins present in the complex have a specific epitope and is also used to validate previously identified proteins. It can also improve resolution when the difference between free nucleic acid and the complex is very small.

6.3 Multiplexed competitor EMSA (MC-EMSA)

The multiplexed EMSA was developed in 2008 by Smith and Humphries to characterize nuclear protein and DNA interactions, namely with transcription factors. In this method the nuclear extract is incubated with a pool of unlabeled DNA consensus competitors prior to adding the labeled DNA probe. An initial EMSA run will determine which cocktail competes with the probe binding to nuclear proteins which will then run individually in another EMSA to determine the precise competitor (Smith & Humphries, 2008). It is a competition-based method to identify uncertain DNA binding proteins requiring only a prior knowledge of transcription factor consensus sequences.

6.4 Two-dimensional EMSA (2D-EMSA)

The two-dimensional EMSA is a process that combines EMSA with proteomic or sequencing techniques to identify the proteins or the nucleic acid sequences that are present in the formed complexes. Two slightly different protocols have been developed to identify the interacting proteins and another method aims at the target nucleic acid sequence.

An initial approach was proposed by Woo and colleagues as they tried to identify and characterize transcription factors (Woo et al., 2002). A crude nuclear extract is partially purified by gel filtration and the resulting fractions are then bound to the nucleic acid probe and analyzed by EMSA. Meanwhile, in parallel, the pI and molecular weight of the putative interacting protein(s) is estimated as the fractions are analyzed by isoelectric focusing or SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) in order to characterize possible candidates. Next, spots with the predetermined pI and molecular weight of the candidates are excised from a two-dimensional array of nuclear proteins and the proteins are eluted, renatured and tested for their binding ability through EMSA and the spots are afterwards analyzed by mass spectrometry for protein identification. This method is limited to proteins that can re-form into functional nucleic acid-binding conformations after the denaturing SDS-PAGE step, although EMSA can still show results even if renaturation efficiency is low. Because the final EMSA step that confirms the binding is performed with protein eluted from single spots it is only possible to identify proteins that interact with the nucleic acids as monomers or homomultimers. Proteins that only interact when complexed with other proteins will give a negative result on the validation EMSA.

A similar 2D-EMSA technique has since then been developed that incorporates EMSA into a two-dimensional proteomics approach by replacing the isoelectric focusing with EMSA as the first dimension of the 2D method (Stead et al., 2006). The protein sample, in the presence or absence of the nucleic acid, is separated by native PAGE as in a conventional EMSA. The protein bands from both conditions are then separated in a second dimension by denaturing SDS-PAGE. The proteins showing the nucleic acid dependent shift in mobility can be extracted from the gel for mass spectrometry identification. This approach does not require any previous knowledge of the chemical or physical properties of the binding protein and does not require protein renaturation after gel excision. It is also not limited to identify proteins that bind by themselves or as homomultimers and allows the characterization of complexes composed of different proteins.

These 2D approaches were developed by the two groups to study transcription factors, therefore, double stranded DNA is used as a nucleic acid probe but they can also adapted to

other nucleic acid probes making them quite versatile methods to identify nucleic acid-interacting proteins.

Chernov and collaborators have developed a similar protocol with two dimensions but instead of aiming to identify the interacting protein(s) it characterizes and maps the specific protein target sites in regions of the human genome (Chernov et al., 2006). This approach is also based on first separating the complexes from the free nucleic acid in a non-denaturing gel and afterwards separating it under denaturing conditions (Vetchinova et al., 2006). The group used a pool of radioisotope-labeled short DNA sequences covering the genome region of interest and mixed it with a nuclear extract from a specific cell line. The formed complexes were separated in a non-denaturing one-dimensional standard EMSA. The complexes were localized by autoradiography and the gel strip containing them was excised and treated with a denaturing agent, SDS, to disrupt the preformed complexes. The strip is then loaded onto the second-dimension denaturing gel and another electrophoresis is performed. The gel is autoradiographed to determine the location of the freed DNAs, which are afterwards cut from the gel to be analyzed. By pairing this method with high-throughput sequencing the authors were able to identify a multitude of specific protein binding sites within a given genomic region.

6.5 EMSA-three-dimensional-electrophoresis (EMSA-3DE)

A three dimensional approach has very recently emerged to purify nucleic acid binding proteins from complexes separated by EMSA (Jiang et al., 2011). This method focuses on recovering the protein in high yield for subsequent analysis and has been developed to study low abundant transcription factors. In this EMSA-based purification procedure the complexes formed are extracted after a native PAGE retardation assay and applied to two-dimensional electrophoresis, isoelectric focusing and SDS-PAGE. The EMSA conditions are systematically optimized to reduce non-specific binding and increase protein yield. After the three electrophoreses the sample can then be electrotransferred onto a nitrocellulose or polyvinylidene difluoride membrane for southwestern and western blotting analysis to further characterize the complexes. Spots of interest can be cut from the gel or the membrane for protein identification by mass spectrometry.

7. Alternatives to EMSA

There are several alternatives to EMSA used in the analysis of nucleic acid-protein interactions with its own advantages and disadvantages when compared to EMSA.

7.1 Footprinting

Footprinting is essentially a protection assay used to characterize the binding site recognized by a given protein. It relies on the fact that a protein bound to the nucleic acid will protect it and interfere with the modification of the sequence it is bound to. The modification can be chemical or enzymatic and it is usually the endonuclease cleavage of radioisotope-labeled nucleic acid previously mixed with the protein(s) of interest. After cleavage the resulting ladder is analyzed on denaturing polyacrylamide gel and visualized by autoradiography. The gaps in the ladder are indicative of sites protected by the protein or proteins in the mixture (reviewed by Hampshire et al., 2007). This method was originally

developed to characterize sequence selectivity but it is also helpful in estimating the binding strength through a footprinting reaction over a range of protein concentrations. For slow binding reactions footprinting can also be applied to assess the reaction kinetics estimating the association and dissociation rates. Although it is a widely used method, there are other approaches that provide higher throughput as the ones described ahead.

A variant on DNA footprinting is the *in vivo* approach, a technique that enables the detection of DNA-protein interactions as they occur in the cell. *In vivo* footprinting also relies on the fact that the bound protein protects the nucleic acid, at its binding site, from cleavage by endonucleases or modification by a chemical agent. The difference is that the cleavage of DNA is carried out within the nucleus following the *in vivo* binding of the proteins to chromatin. Footprints and endonuclease hypersensitive sites that are due to deformations of DNA in chromatin can be detected by this *in vivo* method. This method has been coupled with deep sequencing to identify DNaseI hypersensitive sites in the genome of different cell lines. It enabled the precise identification of a large number of specific cis-regulatory protein binding events with a single experiment (Boyle et al., 2011). Accordingly, the data obtained by this procedure may be more significant and representative of true events when compared with data obtained by the previously described *in vitro* footprinting.

7.2 Nitrocellulose filter binding

Nitrocellulose filter binding assays were developed in the 70s as a rapid enough method to allow kinetic as well as equilibrium studies of DNA-protein interactions (Riggs et al., 1968 and Riggs et al., 1970 as cited in Helwa & Hoheisel, 2010). The manipulation required is rapid enough to allow such measurements. The assay is based on the premise that proteins can bind to nitrocellulose without losing the ability to bind DNA. After the binding reaction the mixture is separated by electrophoresis and then blotted onto a nitrocellulose membrane. Only protein bound DNA remains on the membrane as the free double-stranded DNA will not be retained on nitrocellulose. The amount of DNA on the membrane can be quantified by measuring the label on the nucleic acid. However, this method has its limitations such as the fact that the proteins involved are not identified or the proportion in which they bind DNA. It also provides no information on the DNA sequence the protein interacts with unless well defined nucleic acid fragments are used and is limited to double stranded DNA as single stranded DNA can bind to nitrocellulose under certain conditions resulting in undesirable background.

7.3 Microfluidic mobility shift assay (MMSA)

The capillary microfluidic mobility shift assay (MMSA) is a method that uses fluorescence-based multi-well capillary electrophoresis to characterize protein-nucleic acid interactions. For example, it has been used effectively in characterizing RNA-protein binding in a study of the interaction between human immunodeficiency virus 1 transactivator of transcription and the transactivation-responsive RNA (Fourtounis et al., 2011). This technique requires only nanoliter amounts of sample that are introduced into microscopic channels and separated by pressure-driven flow and application of a potential difference. The free molecules or complexes are visualized by LED-induced fluorescence, discarding the need for hazardous radiolabeling. With the ability to perform 384-well screening this method has an increased capacity over regular EMSA to be compatible with high-throughput screenings.

7.4 Yeast hybrid systems

The yeast one-hybrid is an approach used to identify proteins that bind a given nucleic acid sequence as opposed to the methods that are suited to identify the nucleic acid sequences preferably recognized by a known protein. The protocol is based on a hybrid prey protein fused to a transcription activation domain that allows the expression of a reporter gene when the prey protein interacts with the DNA bait (reviewed by Deplancke et al., 2004). This method allows for a proteome-scale analysis depending on the prey protein library but only detects monomers that bind the target nucleic acid. Although it is an *in vivo* approach it is performed in yeast (*Saccharomyces cerevisiae*), which may not be the endogenous context, and is limited to DNA-protein interactions.

RNA-protein interactions can be studied with a yeast three-hybrid system that involves the expression in yeast cells of not one but three chimerical molecules, which assemble in order to activate two reporter genes (Kraemer et al., 2000). It represents a modification of the yeast two-hybrid system, widely used to identify protein-protein interactions, that was designed to allow high sensitivity *in vivo* detection of RNA-protein interactions. The yeast three-hybrid system includes: a fusion protein consisting of a DNA binding protein and a RNA-binding protein; a hybrid protein consisting of a transcription activating domain and a peptide thought to interact with a particular RNA; a RNA intermediate that promotes the interaction of the two hybrid proteins, this RNA includes the RNA that interacts with the system's RNA-binding protein and the RNA molecule to be investigated. The successful interaction of these 3 components allows the reconstitution of a transcription factor and subsequent activation of reporter genes (Hook et al., 2005 and Wurster & Maher, 2010)

7.5 ChIP assays

Chromatin immunoprecipitation (ChIP) is a commonly used method to study DNA-binding proteins *in vivo* and a standard method for the identification of transcription binding sites and histone modification locations (reviewed by Massie & Mills, 2008). In this method a cross-linking agent (e.g. formaldehyde) is added to cells to covalently bind proteins and chromatin that are in direct contact. Afterwards, the cells are lysed and chromosomal DNA is isolated and fragmented. Specific antibodies are used to immunoprecipitate the targeted proteins with the cross-linked DNA. The bound nucleic acid is released by reverting the cross-linking and then analyzed. Classically, the DNA was characterized by polymerase chain reaction (PCR) which required some previous knowledge of the candidate DNA regions. Nowadays, the DNA bound to protein is more commonly characterized through more powerful tools either coupled with microarrays that represent the genome (ChIP-chip) or state-of-the-art high-throughput sequencing (ChIP-seq). The improvements in DNA sequencing technology allow tens of millions of sequence reads, therefore ChIP-seq has a major advantage of increased sensitivity and resolution to add to the fact that it is not limited to predetermined probe sets as ChIP-chip. The major strength of the ChIP-based approaches is that they capture complexes *in vivo* and the binding reactions can be studied under different cellular conditions and at different time points. However it also has important limitations. The method requires high-quality antibodies that are available only for a limited number of proteins. To circumvent this, epitope-tagged proteins could be used although it usually implies the introduction of modified genes into the endogenous locus in order to obtain expression at physiological levels. This method does not distinguish between

proteins that bind directly to the genomic DNA and those that only interact with other proteins that do bind.

7.6 SELEX

The Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is a well established method that enables the selection of enriched sequences from a random library that bind recombinant proteins. This procedure starts with the synthesis of the oligonucleotide library and then incubating the generated sequences with the putative interacting protein(s). The sequences that bind are eluted, amplified by PCR and subjected to more rounds of selection with increasing stringency conditions. This allows the identification of the tightest-binding sequences. It is a widely used approach to obtain transcription factors binding motifs as it requires low amounts of purified proteins (Matys et al., 2006). This approach becomes very complicated to use when large numbers of nucleic acid-binding proteins are analyzed as it then requires multiple rounds of selection. Another limitation is the fact that it is aimed at the identification of the best binding DNA targets *in vitro* and does not allow the characterization of the exact *in vivo* selectivity.

7.7 Protein microarray

A protein microarray is a method that allows high-throughput analysis in which labeled nucleic acids are queried against proteins immobilized on a chip (reviewed by Hu et al., 2011). In a functional protein microarray, thousands of purified recombinant proteins can be immobilized in a glass slide in discrete locations forming a high-density protein matrix, providing a flexible platform to characterize different protein activities. It is a very versatile method as it can perform a semi-quantitative analysis of protein binding to a wide range of molecules (nucleic acids, other proteins, antibodies, lipids, glycans...). In theory, it is feasible to print arrays of all the annotated proteins of a given organism originating a whole proteome microarray. However, it implies the expression and purification of each individual protein and several conditions need to be optimized to render the proteins apt for this method. Since the protein is immobilized it is crucial to guarantee that its structural integrity remains intact especially the binding domains that are to be studied.

7.8 Nucleic acid microarrays

Nucleic acid microarrays can also be used for a direct analysis of protein-nucleic acid interactions. In this case it is the nucleic acid that is immobilized and not the protein. Nucleic acid chips are a powerful and versatile tool in biological research. They consist of high-density arrays of oligonucleotides or complementary DNA that can cover a whole genome (reviewed by Stoughton, 2005). For protein-interaction studies, the protein(s) of interest is expressed usually with an epitope tag, and purified. The tag serves two purposes; it helps to isolate the protein through affinity purification, and allows detection by an epitope-specific reporter antibody. After incubation of the protein with the nucleic acid chip the signal intensities at the several array spots can be measured.

7.9 Ribonucleoprotein Immunoprecipitation – Microarray (RIP-chip)

RNA immunoprecipitation and chip hybridization (RIP) is a protocol very similar to ChIP-chip except that it targets RNA-protein interactions rather than DNA-protein (Keene et al.,

2006). RIP-chip is an approach that consists on a microarray profiling of RNAs obtained from immunoprecipitated RNA-protein complexes. Genome-wide arrays are used to identify messenger RNAs (mRNAs) that are present in endogenous messenger ribonucleoprotein complexes making it a great tool to identify the physiological substrates of mRNAs. The endogenous complexes are immunoprecipitated from cell lysates which limits this study to kinetically stable interactions. Even though it can identify RNA-protein complexes with heteromultimers, at least one of the proteins has to be previously known to be the basis of immunoprecipitation and “fish out” the whole complex.

7.10 Crosslinking and Immunoprecipitation (CLIP) and Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)

The RIP-chip method that has just been described is limited to studies of very stable RNA-protein complexes; to remediate this problem another method is available to study RNA-binding proteins. The crosslinking and immunoprecipitation (CLIP) approach uses *in vivo* UV crosslinking prior to the complexes immunoprecipitation to identify less stable interactions (Ule et al., 2003). After immunoprecipitation RNA molecules are separated and cDNA sequencing is carried on. However, this method is not perfect as the commonly used UV 254nm RNA-protein crosslinking has low efficiency and it is difficult to distinguish between crosslinked RNAs from background non-crosslinked fragments that can be detected in the sample due to the presence of abundant cellular RNAs.

A more recent approach tries to further improve the CLIP method using photoreactive ribonucleoside analogs such as 4-thiouridine or 6-thioguanosine (Hafner et al., 2010). In this photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) protocol the photoreactive nucleosides are incorporated into nascent transcripts within living cells. The irradiation is performed with UV light of 365nm, which induces an efficient crosslink of the labeled cellular RNA to its interacting proteins. The labeled RNAs are isolated after co-immunoprecipitation, and converted into cDNA for deep sequencing. The precise crosslinking position can be identified by mutations in the sequenced cDNA making it possible to distinguish the crosslinked fragments from background.

7.11 High-Throughput Sequencing – Fluorescent Ligand Interaction Profiling (HiT-FLIP)

Very recently a new method was developed to characterize DNA-protein interactions using second-generation sequencing instruments (Nutiu et al., 2011). This method allows high throughput and quantitative measurement of DNA-protein binding affinity. This High-Throughput Sequencing - Fluorescent Ligand Interaction Profiling (HiT-FLIP) uses the optics of a high-throughput sequencer to visualize *in vitro* binding of a protein to the sequenced DNA in a flow cell. The new method was initially used on a *Saccharomyces cerevisiae* transcription factor. The fluorescently tagged protein was added at different concentrations to a flow cell containing around 88 million DNA clusters, the equivalent of over 160 yeast genomes. The traditional EMSA was used as an independent validation of the dissociation constants obtained and found a high correlation with values obtained with the new method and those from EMSA as reported in literature. This high-throughput method has an obvious advantage in the fact that it can provide hundreds of millions of measurements but is limited to DNA-protein interactions and requires expensive

equipment. Another advantage is that the sequencing instrument can measure multiple fluorescent wavelengths allowing hetero and homodimeric forms to be measured in the same run, using distinct tags on individual proteins.

8. Conclusion

Since the first report, 30 years ago, EMSA became one of the most popular methods for detection and characterization of protein-nucleic acid interactions. Hundreds of protocols have been published accommodating modifications in virtually every parameter influencing the experimental outcome. Improvements were made in all EMSA steps including the methods for preparation of protein samples and purification, synthesis and labeling of nucleic acids, and detection. This allowed enlarging and diversifying the applications of EMSA and resulting in a number of variants of the method.

However, despite the large amount of available literature and protocols trial and error will ultimately be the way to optimize the EMSA conditions for the nucleic acid-protein complex to be analyzed. The guidelines discussed above help to provide an initial protocol adjusted to each study but slight changes may be needed to improve binding and detection of the complexes.

In recent years, the use of highthroughput approaches to detect biologically relevant interactions, including those between proteins and nucleic acids, was reported. Development of these approaches was made possible, at least in part, by the availability of more sensitive and specific equipment and tools. Although EMSA cannot achieve a high throughput level it remains a valuable tool to confirm the detected interactions.

9. Acknowledgements

We are grateful to Dr. Cristina Branco for constructive comments. Work in the authors' laboratory is supported by Fundação para a Ciência e Tecnologia (PTDC/SAU-MII/098314/2008). CA is a recipient of a FCT PhD grant.

10. References

- Adachi, Y.; Chen, W.; Shang, W. & Kamata, T. (2005). Development of a direct and sensitive detection method for DNA-binding proteins based on electrophoretic mobility shift assay and iodoacetamide derivative labelling. *Analytical Biochemistry*, Vol.342, pp. 348-351
- Alves, C.; Cheng, H.; Roder, H. & Taylor, J. (2010). Intrinsic disorder and oligomerization of the hepatitis delta virus antigen. *Virology*, Vol. 370, pp 12-21
- Apella, E. & Anderson, C. (2001). Post-translational modifications and activation of p53 by genotoxic stresses. *European Journal of Biochemistry*, Vol.268, No.10, pp. 2764-2772
- Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, Vol.10, pp. 411-421
- Boyle, A.; Song, L.; Lee, B.; London, D.; Keefe, D.; Birney, E.; Iyer, V.; Crawford, G. & Furey, T. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Research*, Vol.21, No.3, pp. 456-464

- Chernov, I.; Akopov, S.; Nikolaev, L. & Sverdlov, E. (2006). Identification and mapping of DNA binding proteins target sequences in long genomic regions by two-dimensional EMSA. *BioTechniques*, Vol.41, pp. 90-96
- Chi, P.; Van Komen, S.; Sehorn, M.; Sigurdsson, S. & Sung, P. (2006). Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function. *DNA Repair*, Vol.5, No.3, pp. 381-391
- Crothers, D. & Drak, J. (1992). [3] Global features of DNA structure by comparative gel electrophoresis. *Methods in Enzymology*, Vol 212, pp. 46-71
- Demarse, N.; Ponnusamy, S.; Spicer, E.; Apohan, E.; Baatz, J.; Ogretmen, B. & Davies, C. (2009). Direct binding of glutaraldehyde 3-phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation. *Journal of Molecular Biology*, Vol.394, No.4, pp. 789-803
- Deplancke, B.; Dupuy, D.; Vidal, M. & Walhout, A. (2004). A gateway-compatible yeast one-hybrid system. *Genome Research*, Vol.1, 2093-2101
- Dignam, J.; Lebovitz, R. & Roeder, R. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research*, Vol.11, No.5, pp. 1457-1489
- Endo, Y. & Sawasaki, T. (2006). Cell-free expression systems for eukaryotic protein production. *Current Opinion in Biotechnology*, Vol.17, pp. 373-380
- Filion, G.; Fouvry, L. & Defosse, P.-A. (2006). Using reverse electrophoretic mobility shift assay to measure an compare protein-DNA binding affinities. *Analytical Biochemistry*, Vol.357, pp. 156-158
- Fourtounis, J.; Falgueyret, J.-P. & Sayegh, C. (2011). Assessing protein-RNA interactions using microfluidic capillary mobility shift assays. *Analytical Biochemistry*, Vol.411, pp. 161-163
- Fried, M. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis*, Vol. 10, pp. 366-376
- Fried, M. & Crothers, D. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acid Research*, Vol.9, No.23, pp. 6505-6525
- Fried, M. & Crothers, D. (1984). Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. *Journal of Molecular Biology*, Vol.172, No.3, pp. 241-262
- Fried, M. & Daugherty, M. (1998). Electrophoretic analysis of multiple protein-DNA interactions. *Electrophoresis*, Vol.19, pp. 1247-1253
- Garner, M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Research*, Vol.9, No.13, pp. 3047-3060
- Goh, A.; Coffil, C. & Lane, D. (2011). The role of mutant p53 in human cancer. *The Journal of Pathology*, Vol.223, No.2, pp. 116-126
- Granger-Schnarr, M.; Lloubes, R.; de Murcia, G. & Schnarr, M. (1988). Specific protein-DNA complexes: immunodetection of the protein component after gel electrophoresis and Western blotting. *Analytical Biochemistry*, Vol.174, No.1, pp. 235-238
- Greco-Stewart, V. & Pelchat, M. (2010). Interaction of host cellular proteins with components of the hepatitis delta virus. *Viruses*, Vol.2, No.1, pp 189-212
- Haffner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M.; Jungkamp, A.-C.; Munschauer, M.; Ulrich, A.; Wardle, G.; Dewell, S.; Zavola, M. & Tuschli, T. (2010). PAR-CLIP – A method to identify transcriptome-wide binding sites of RNA binding proteins. *Journal of Visualized Experiments*, Vol.41, pp.2034

- Hampshire, A.; Rusling, D.; Broughton-Head, V. & Fox, K. (2007). Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods*, Vol.42, pp. 128-140
- Hellman, L. & Fried, M. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, Vol.2, No.8, pp. 1849-1861
- Helwa, R. & Hoheisel, J. (2010). Analysis of DNA-protein interactions: from nitrocellulose filter binding assays to microarray studies. *Analytical & Bioanalytical Chemistry*, Vol.398, pp. 2551-2561
- Holden, N. & Tacan, C. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of Pharmacological and Toxicological Methods*, Vol.63, pp. 7-14
- Hope, I. & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *The EMBO Journal*, Vol.6, No.9, pp. 2781-2784
- Hook, B.; Bernstein, D.; Zhang, B. & Wickens, M. (2005). RNA-proteins interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA*, Vol.11, No.2, pp. 227-233
- Hu, S.; Xie, Z.; Qian, J.; Blackshaws, S. & Zhu, H. (2011). Functional protein microarray technology. *WIREs Systems Biology and Medicine*, Vol.3, pp. 255-268
- Jarvis, D. (2009). Baculovirus-insect cell expression systems. *Methods in Enzymology*, Vol.463, pp. 191-222
- Jiang, D.; Jia, Y. & Jarrett, H.W. (2011). Transcription factor proteomics: Identification by a novel gel mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography-electrospray-mass spectrometry analysis. *Journal of Chromatography A*, Vol.1218, No.39, pp. 7003-7015
- Jiang, S.; Macias, M. & Jarrett, H. (2010). Purification and identification of a transcription factor, USF2, binding to E-box elements in the promoter of human telomerase reverse transcriptase (hTERT). *Proteomics*, Vol.10, No.2, pp. 203-211
- Jimenez, J. (2010). Protein-DNA interaction at the origin of neurological diseases: a hypothesis. *Journal of Alzheimer's Disease*, Vol.22, No.2, pp. 375-391
- Keene, J.; Komisarow, J. & Friedersdorf, M. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature Protocols*, Vol.1, pp. 302-307
- Kerr, L. (1995). [42] Electrophoretic mobility shift assay. *Methods in Enzymology*, Vol.254, pp. 619-632
- Kraemer, B.; Zhang, B.; SenGupta, D.; Fields, S. & Wickens, M. (2000). Using the yeast three-hybrid system to detect and analyze RNA-protein interactions. *Methods in Enzymology*, Vol.328, pp. 297-321
- Lane, D.; Prentki, P. & Chandler, M. (1992). Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiological Reviews*, Vol.56, No.4, pp. 509-528
- Makino, T.; Skretas, G. & Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microbial Cell Factories*, Vol.10, No.1, pp. 32-42
- Massie, C. & Mills, I. (2008). ChIPping away at gene regulation. *EMBO Reports*, Vol.9, No.4, pp. 337-343
- Matys, V.; Kel-Margoulis, O.; Fricke, E.; Liebich, I.; Land, S.; Barre-Dirrie, A.; Reuter, I.; Checkmenev, D.; Krull, M.; Hornischer, K.; Voss, N.; Stegmaier, P.; Lewicki-Potapov, B.; Saxel, H.; Kel, A. & Wingender, E. (2006). TRANSFAC and its module TRANSCOMPLE: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research*, 1 34 Database issue D108-110

- Nordheim, A. & Meese, K. (1988). Topoisomer gel retardation: detection of anti-Z-DNA antibodies bound to Z-DNA within supercoiled DNA minicircles. *Nucleic Acids Research*, Vol.16, No.1, pp. 21-37
- Nutiu, R.; Friedman, R.; Luo, S.; Khrebtukova, I.; Silva, D.; Li, R.; Zhang, L.; Schroth, G. & Burge, C. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology*, Vol.29, No.7, pp. 659-664
- Palecek, E.; Vlk, D.; Stankova, V.; Brazda, V.; Vojtesek, B.; Hupp, T.; Schaper, A. & Jovin, T. (1997). Tumor suppressor protein p53 binds preferentially to supercoiled DNA. *Oncogene*, Vol.15, No.18, pp. 2201-2209
- Puzio-Kuter, A. (2011). The role of p53 in metabolic regulation. *Genes & Cancer*, Vol.2, No.4, pp. 385-391
- Rizzetto, M. (2009). Hepatitis D: Thirty years after. *Journal of Hepatology*, Vol.50, pp. 1043-1050
- Ruscher, K.; Reuter, M.; Kupper, D.; Trendelenburg, G.; Dirnagl, U. & Meisel, A. (2000). A fluorescence based non-radioactive electrophoretic mobility shift assay. *Journal of Biotechnology*, Vol.78, pp. 163-170
- Saida, F.; Uzan, M.; Odaert, B. & Bontems, F. (2006). Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Current Protein and Peptide Science*, Vol.7, pp. 47-56
- Sandman, K.; Krzycki, J.; Dobrinski, B.; Lurz, R. & Reeve, J. (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histone. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.87, No.15, pp. 5788-5791
- Simicevic, J. & Deplancke, B. (2010). DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. *Molecular BioSystems*, Vol.6, No.3, pp. 462-468
- Spinner, D.; Liu, S.; Wang, S-W. & Schmidt, J. (2002). Interaction of the myogenic determination factor myogenin with E12 and a DNA target: mechanism and kinetics. *Journal of Molecular Biology*, Vol.317, No.3, pp. 431-445
- Stead, J.; Keen, J. & McDowall, K. (2006). The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics*, Vol.5, No.9, pp. 1697-1702
- Stoughton, R. (2005). Applications of DNA microarrays in biology. *Annual Review of Biochemistry*, Vol.74, pp. 53-82
- Ule, J.; Jensen, K.; Ruggiu, M.; Mele, A.; Ule, A. & Darnell, R. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science*, Vol.302, No.5648, pp. 1212-1215
- Vetchinova, A.; Akopov, S.; Chernov, I.; Nikolaev, L. & Sverdlov, E. (2006). Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19. *Analytical Biochemistry*, Vol.354, pp. 85-93
- Vossen, K.; Wolz, R.; Daugherty, M. & Fried, M. (1997). Role of macromolecular hydration in the binding of the *Escherichia coli* cyclic AMP receptor to DNA. *Biochemistry*, Vol.36, No.39, pp. 11640-11647
- Walker, J. (1994). Gradient SDS polyacrylamide gel electrophoresis of proteins. *Methods in Molecular Biology*, Vol.32, pp. 35-38
- Woo, A.; Dods, J.; Susanto, E.; Ulgiati, D. & Abraham, L. (2002). A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics*, Vol.1, No.6, pp. 472-478
- Wurster, S. & Maher, L. (2010). Selections that optimize RNA display in the yeast three-hybrid system. *RNA*, Vol.16, No.2, pp. 253-258



Gel Electrophoresis - Advanced Techniques

Edited by Dr. Sameh Magdeldin

ISBN 978-953-51-0457-5

Hard cover, 500 pages

Publisher InTech

Published online 04, April, 2012

Published in print edition April, 2012

As a basic concept, gel electrophoresis is a biotechnology technique in which macromolecules such as DNA, RNA or protein are fractionated according to their physical properties such as molecular weight or charge. These molecules are forced through a porous gel matrix under electric field enabling uncounted applications and uses. Delivered between your hands, a second book of this Gel electrophoresis series (Gel Electrophoresis- Advanced Techniques) covers a part, but not all, applications of this versatile technique in both medical and life science fields. We try to keep the contents of the book crisp and comprehensive, and hope that it will receive overwhelming interest and deliver benefits and valuable information to the readers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Carolina Alves and Celso Cunha (2012). Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions, Gel Electrophoresis - Advanced Techniques, Dr. Sameh Magdeldin (Ed.), ISBN: 978-953-51-0457-5, InTech, Available from: <http://www.intechopen.com/books/gel-electrophoresis-advanced-techniques/electrophoretic-mobility-shift-assay-analyzing-protein-nucleic-acid-interactions>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.